



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ
ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

ΤΟΜΕΑΣ ΣΗΜΑΤΩΝ, ΕΛΕΓΧΟΥ ΚΑΙ ΡΟΜΠΟΤΙΚΗΣ

**Συνέργεια Κατάτμησης Εικόνων
και Αναγνώρισης Αντικειμένων
με Γεωμετρικές και Στατιστικές Τεχνικές
της Όρασης Υπολογιστών**

ΔΙΔΑΚΤΟΡΙΚΗ ΔΙΑΤΡΙΒΗ

ΤΟΥ

ΙΑΣΟΝΑ ΚΟΚΚΙΝΟΥ

Διπλωματούχου Ηλεκτρολόγου Μηχανικού &
Μηχανικού Υπολογιστών Ε.Μ.Π.

Αθήνα, Ιούνιος 2006



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ & ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

ΤΟΜΕΑΣ ΣΗΜΑΤΩΝ, ΕΛΕΓΧΟΥ ΚΑΙ ΡΟΜΠΟΤΙΚΗΣ

Συνέργεια Κατάτμησης Εικόνων και Αναγνώρισης Αντικειμένων με Γεωμετρικές και Στατιστικές Τεχνικές της Όρασης Υπολογιστών

ΔΙΔΑΚΤΟΡΙΚΗ ΔΙΑΤΡΙΒΗ

ΤΟΥ

ΙΑΣΟΝΑ ΚΟΚΚΙΝΟΥ

Διπλωματούχου Ηλεκτρολόγου Μηχανικού &
Μηχανικού Υπολογιστών Ε.Μ.Π.

Συμβουλευτική Επιτροπή:

Καθ. Πέτρος Μαραγκός (Επιβλέπων)
Καθ. Ανδρέας Γεώργιος Σταφυλοπάτης
Καθ. Στέφανος Κόλλιας

Επαμελής εξεταστική επιτροπή

...
Π. Μαραγκός
Καθηγητής Ε.Μ.Π.

...
Σ. Κόλλιας
Καθηγητής Ε.Μ.Π.

...
Α. Γ. Σταφυλοπάτης
Καθηγητής Ε.Μ.Π.

...
Σ. Τζαφέστας
Καθηγητής Ε.Μ.Π.

...
Τ. Σελλής
Καθηγητής Ε.Μ.Π.

...
Ν. Μαρατός
Καθηγητής Ε.Μ.Π.

...
Ν. Γαλατσάνος
Καθηγητής
Παν. Ιωαννίνων

Αθήνα, Ιούνιος 2006

Περιεχόμενα

1	Εισαγωγή	1
1.1	Το Πρόβλημα της Συνέργειας	1
1.2	Προσέγγιση στο πρόβλημα της Συνέργειας	2
1.3	Συνεισφορές και Δομή Διδακτορικής Διατριβής	4
2	Θεωρητικό Υπόβαθρο και Επισκόπηση Ερευνητικού Πεδίου	7
2.1	Κατάτμηση Εικόνων	7
2.2	Αναγνώριση Αντικειμένων	19
2.3	Πιθανοτικά Μοντέλα για Προβλήματα Ανάλυσης και Αναγνώρισης Προ- τύπων	32
3	Ανάλυση και Εκμάθηση ενός Βιολογικά Εμπνευσμένου Μοντέλου Χα- μηλού και Μέσου Επιπέδου Λειτουργιών Όρασης	45
3.1	Συστήματα Βασισμένα στο Βιολογικό Σύστημα Όρασης	46
3.2	Ένα Μοντέλο για Χαμηλού και Μέσου Επιπέδου Λειτουργίες της Όρασης	49
3.3	Ανάλυση του Μοντέλου με Τεχνικές Όρασης Υπολογιστών	56
3.4	Εκμάθηση των Παραμέτρων του Μοντέλου	61
3.5	Πειραματικά Αποτελέσματα	66
3.Α'	Υπολογισμός Κάτω Φράγματος για τον Αλγόριθμο Εκμάθησης Βαρών .	71
4	Ανάλυση και Κατάτμηση Εικόνων Υφής	73
4.1	Εισαγωγή	73
4.2	Το Μοντέλο AM-FM	75
4.3	Τοπικά Αναγεννητικά Μοντέλα για Υφή	80
4.4	Διάκριση Υφής και Ακμών	86
4.5	Κατάτμηση Εικόνων Υφής με Χαρακτηριστικά Διαμόρφωσης και Σταθ- μισμένη Εξέλιξη Καμπυλών	91
5	Συνέργεια μεταξύ Αναγνώρισης Αντικειμένων και Κατάτμησης Εικό- νων μέσω του Αλγόριθμου Expectation Maximization	101
5.1	Η προσέγγιση του EM για το Πρόβλημα της Συνέργειας	102
5.2	Βήμα E : Κατάτμηση Βάσει Αντικειμένων	106
5.3	Εκτίμηση Παραμέτρων: Βήμα M	113
5.4	Συνεργετική Ανίχνευση Αντικειμένων	120
5.5	Επισκόπηση Προηγούμενης Εργασίας και Συζήτηση	129
5.Α'	Σχέση Ανανέωσης Παραμέτρων AAM Βάσει του EM	134
5.Β'	Περιγραφή Μοντέλου Ανίχνευσης	135

6	Ανίχνευση μέσω Σημείων Ενδιαφέροντος και Γραφικών Δικτύων	137
6.1	Εισαγωγή	137
6.2	Σημεία Ενδιαφέροντος για Ανίχνευση Αντικειμένων.	137
6.3	Χαρακτηριστικά Τύπου Primal Sketch	139
6.4	Κάτωθεν Ανίχνευση Αντικειμένων	144
6.5	Εισαγωγή Άνωθεν Πληροφορίας	149
6.6	Κατασκευή Γραφικών Μοντέλων	155
6.7	Πειραματικά Αποτελέσματα	156
6.A'	Αναλυτικές Σχέσεις Μηνυμάτων Μεταξύ Γειτόνων	160
7	Συμπεράσματα και Κατευθύνσεις Μελλοντικής Έρευνας	163
	Βιβλιογραφία	167

Κατάλογος Σχημάτων

1.1	Αναγκαιότητα της από κοινού αντιμετώπισης λειτουργιών όρασης χαμηλού και υψηλού επιπέδου [165]: (α) Εικόνα που περιέχει ένα αντικείμενο (β) Αποτελέσματα ανίχνευσης ακμών (γ) Μεγέθυνση της περιοχής του προσώπου. Η εικόνα δεν είναι κατάλληλη για την εφαρμογή τυπικών μεθόδων κατάτμησης και ανίχνευσης αντικειμένων, λόγω της χαμηλής ανάλυσης, της έντονης σκίασης και της περίπλοκης δομής της. Χρησιμοποιώντας όμως το μοντέλο ενός ανθρώπου η ερμηνεία της διευκολύνεται, επιτρέποντας την αποτελεσματική εκτέλεση των επιμέρους λειτουργιών της κατάτμησης και ανίχνευσης.	2
1.2	Μοντέλο της άνωθεν όρασης [190]: παραμετρικά μοντέλα που υπάρχουν για τα οπτικά αντικείμενα χρησιμοποιούνται για την ερμηνεία μίας σκηνής μέσω της ανακατασκευής της.	3
2.1	Χαμηλού και Υψηλού επιπέδου προβλήματα της κατάτμησης: Πάνω γραμμή: Τα κριτήρια με τα οποία αποκαλούμε τις περιοχές της εικόνας ομοιογενείς δεν είναι κοινά για όλες τις εικόνες, ούτε μπορούν να ποσοτικοποιηθούν εύκολα. Μόνο στην εικόνα (α) είναι χρήσιμη η τιμή της φωτεινότητας ως χαρακτηριστικό για κατάτμηση· για τις υπόλοιπες το χρησιμοποιούμενο χαρακτηριστικό είναι η διασπορά του θορύβου (β) η σκίαση (παράγωγος της φωτεινότητας) (γ) η ομοιότητα των κηλίδων (δ) και η προσήνειά τους (ε). Κάτω γραμμή: Αναλόγως της κλίμακας και του κύριου ενδιαφέροντος του θεατή αλλάζει και η κατάτμηση μίας σκηνής.	8
2.2	(α): Γεωμετρικές ποσότητες που εμπλέκονται στην εξέλιξη μίας καμπύλης (β) Διαμερισμός μίας εικόνας.	12
2.3	Παραμετρική και έμμεση αναπαράσταση μίας καμπύλης.	14
2.4	Εξέλιξη ενός Geodesic Active Contour: η τοπολογία της καμπύλης αλλάζει αβίαστα ενώ οι γωνίες των αντικειμένων εντοπίζονται με ακρίβεια.	16
2.5	Πολυδιακριτική αναπαράσταση (α) της εικόνας και (β) των ακμών της.	19
2.6	(α)-(β): Αρχική και τελική θέση των ορίων της περιοχής που αντιστοιχεί στην φωτεινό τμήμα της εικόνας και (γ)-(δ) στο σκοτεινό τμήμα.	19
2.7	Αναγνώριση μέσω αντιστοίχισης μοντέλων [149]: ορισμένα από τα χαρακτηριστικά σημεία της εικόνας, όπως γωνίες και διασταυρώσεις μπορούν να αντιστοιχιστούν σε αυτά προαποθηκευμένων αντικειμένων. Αν καθορίζοντας κατάλληλα τις παραμέτρους προβολής του αντικειμένου μπορούν να ερμηνευθούν περισσότερα χαρακτηριστικά, αυτό θεωρείται ενδεικτικό της ύπαρξης του αντικειμένου.	21

2.8	Αναγνώριση μέσω τεχνικών νευρωνικών δικτύων/αναγνώρισης προτύπων [192]: τα αντικείμενα ανιχνεύονται από μία καταναμημένη αρχιτεκτονική μέσω μίας αλληλουχίας εξαγωγής περιπλοκότερων και ακριβέστερων χαρακτηριστικών	22
2.9	Αναγνώριση αντικειμένων μέσω τμηματικών αναπαραστάσεων [65]: η αναπαράσταση ενός αντικειμένου ανάγεται στην μοντελοποίηση των επιμέρους τμημάτων του και των εξαρτήσεων μεταξύ τους	23
2.10	Αναγνώριση μέσω παραμορφώσιμων μοντέλων των αντικειμένων [157]: μοντελοποιώντας την παραμόρφωση στην εμφάνιση και στο σχήμα του αντικειμένου, μπορεί να αναπαρασταθεί μία κατηγορία αντικειμένων χρησιμοποιώντας ένα χαμηλοδιάστατο παραμετρικό μοντέλο	24
2.11	(α) Επιπεδοσύνολα του πραγματικού μέρους της απεικόνισης Ikeda (β) Σημεία τα οποία χρησίμευσαν ως σύνολο εκπαίδευσης (γ-ε) Προσεγγίσεις μέσω ενός δικτύου RBF, ενός mixture-of-experts και ενός SVM αντίστοιχα.	28
2.12	Από [246]: Πέρα από τις τυπικά θεωρούμενες πηγές αλλοίωσης (φωτισμός και θόρυβος) στην δημιουργία μίας εικόνας από ένα αντικείμενο υπεισέρχονται και άλλες πηγές παραμόρφωσης, όπως η χωρική παραμόρφωση και οι επικαλύψεις.	34
2.13	Γραφικό δίκτυο για την αναπαράσταση των αλληλεξαρτήσεων της κατανομής (2.38).	37
2.14	Ανίχνευση προσώπου χρησιμοποιώντας γραφικά μοντέλα: Πάνω γραμμή: Εικόνα στην είσοδο του προσώπου, και τοπικά καθοριζόμενα δυναμικά παρατήρησης $\log \Phi_i(X_i)$, όπου X_i η θέση του δεξιού ματιού, της μύτης και του δεξιού χείλους, αντίστοιχα. Κάτω γραμμή: Γραφικό δίκτυο για την αναπαράσταση και ανίχνευση προσώπου, και αποτελέσματα αλγορίθμου Belief Propagation για τις ύστερες κατανομές των θέσεων των σημείων, $\log B_i(X_i)$	39
3.1	(α) Διάγραμμα της αρχιτεκτονικής BCS/FCS και (β) αντίστοιχες περιοχές του οπτικού συστήματος.	48
3.2	(α) Διάγραμμα του συστήματός μας και (β) αντίστοιχες περιοχές του βιολογικού συστήματος όρασης	50
3.3	Η αρχιτεκτονική του μοντέλου μας. Κάτω αριστερά φαίνεται η αρχιτεκτονική του συστήματος και οι αλληλεπιδράσεις των δομικών μονάδων του. Στο υπόλοιπο σχήμα φαίνεται ο υπολογισμός περιγραμμάτων σε μία συγκεκριμένη κλίμακα: η ανίχνευση χαρακτηριστικών σε πολλαπλές κατευθύνσεις στο Στάδιο I ακολουθείται από τον σχηματισμό περιγραμμάτων στο Στάδιο II. Εκεί η κάτωθεν πληροφορία συνδιάζεται αναδρομικά με το άνωθεν σήμα saliency που υπολογίζεται από το Στάδιο III. Τα Στάδια I και II υπακούουν αναδρομική δυναμική που καθορίζεται από τις αλληλεπιδράσεις τους.	51
3.4	(α) Εικόνα στην είσοδο του συστήματος, (β)-(γ) Κατευθυντική Ενέργεια για την κάθετη κατεύθυνση χωρίς και με κανονικοποίηση, αντίστοιχα, (δ)-(ε) μέγιστο ως προς κατεύθυνση της κατευθυντικής ενέργειας χωρίς και με κανονικοποίηση, αντίστοιχα.	52
3.5	Οι λοβοί που δίνει το μοντέλο καμπύλων Elastica (3.13) για παραμέτρους (α) $\lambda = 10, \sigma = 0.4$, (β) $\lambda = 10, \sigma = 0.2$, (γ) $\lambda = 20, \sigma = 0.2$	54

3.6	Ανίχνευση φανταστικών ακμών: στις πάνω εικόνες αντιλαμβανόμαστε ανύπαρκτες ακμές σε λευκές περιοχές. Στις κάτω εικόνες φαίνονται τα αποτελέσματα ανίχνευσης ακμών του δικτύου μας, το οποίο χρησιμοποιώντας το υποσύστημα ανίχνευσης saliency μοντελοποιεί αυτή τη συμπεριφορά του ανθρώπινου οπτικού συστήματος.	55
3.7	Σύγκριση αποτελεσμάτων ανίχνευσης ακμών τύπου Canny [28] και του ιεραρχικού δικτύου μας [126]. Ορίζοντας τα κατώφλια και για τις δύο μεθόδους ώστε να επιστρέφουν το ίδιο πλήθος ακμών, από τη μέθοδο Canny παίρνουμε ένα μεγάλο πλήθος ακμών σε περιοχές με υφή. Αντίθετα, το δίκτυό μας χρησιμοποιώντας μία υψηλότερου επιπέδου διαδικασία ανίχνευσης ομαλών ακμών εστιάζει σε ομαλές και έντονες ακμές. Η διασπορά της Gaussian και τα πάνω και κάτω κατώφλια της μεθόδου Canny που αναγράφονται σε κάθε εικόνα ορίστηκαν ξεχωριστά, ώστε να δίνουν την οπτικά καλύτερη απόδοση.	56
3.8	Τύπος συνδέσεων για μία BM (β) Το σύστημά μας ως μία BM.	61
3.9	Διαφορά μεταξύ των προκαθορισμένων βαρών και αυτών που έχουν εκμαθηθεί. Στις περιττές γραμμές δείχνουμε το προκαθορισμένα βάρη και στις άρτιες αυτά που εκμαθεύτηκαν· μία υψηλή τιμή φωτεινότητας δείχνει μία ισχυρή ανασταλτική σύνδεση μεταξύ του νευρώνα στο κέντρο και του νευρώνα στην αντίστοιχη θέση. Οι πρώτες δύο γραμμές δείχνουν βάρη για το στάδιο I και οι κάτω για το στάδιο II. Οι στήλες (α)/(β) αντιστοιχούν σε βάρη μεταξύ κάθετων/οριζόντιων νευρώνων σε μία τραχιά κλίμακα και οι (γ)/(δ) σε λεπτή κλίμακα.	67
3.10	Εκμάθηση ανίχνευσης ακμών: (α) Εικόνα στην είσοδο του συστήματος, (β)-(δ) πιθανότητα ύπαρξης ακμών, χρησιμοποιώντας ευρετικές εκφράσεις για τα βάρη, σε αυξανόμενες αναλύσεις της εικόνας (ε) πιθανότητα ακμής βάσει κατατιμήσεων ανθρώπων (στ)-(η) πιθανότητα ακμής με βάση τα βάρη που προέκυψαν από τον αλγόριθμο εκμάθησης.	68
3.11	Ανίχνευση ακμών σε μία εικόνα που δεν ανήκει στο σύνολο εκπαίδευσης: (α) Εικόνα στην είσοδο του συστήματος, (β)-(δ) πιθανότητα ύπαρξης ακμών, χρησιμοποιώντας ευρετικές εκφράσεις για τα βάρη, σε αυξανόμενες αναλύσεις της εικόνας (ε) πιθανότητα ακμής βάσει κατατιμήσεων ανθρώπων (στ)-(η) πιθανότητα ακμής με βάση τα βάρη που προέκυψαν από τον αλγόριθμο εκμάθησης.	69
3.12	Συγκριτικά αποτελέσματα για ανίχνευση ακμών, για καθιερωμένες μεθόδους ανίχνευσης ακμών και το δίκτυό μας, σε διαφορετικές κλίμακες· οι συντομεύσεις {F,M,C}S σημαίνουν Fine-, Mid- και Coarse-Scale αντίστοιχα. Παρατηρούμε ότι το σύστημά μας έχει καλύτερες επιδόσεις από τις περισσότερες μεθόδους, ενώ η μέθοδος του Brightness & Texture Gradient των [154] που έχει καλύτερες επιδόσεις χρησιμοποιεί και χαρακτηριστικά υφής στην ανίχνευση ακμών. . . .	70
4.1	Τοπικά ζωνοπερατές εικόνες υφής.	76

4.2	Εξαγωγή χαρακτηριστικών με τη μέθοδο DCA: ένα σύνολο απο ζωνοπερατά φίλτρα Gabor χρησιμοποιείται αρχικά για να απομονωθούν και αποδιαμορφωθούν οι μεμονωμένες συνιστώσες της (4.3). Το κυριάρχο κανάλι επιλέγεται σε κάθε σημείο της εικόνας και οι παράμετροί του χρησιμοποιούνται για την περιγραφή της υφής. Η κυρίαρχη δομή του σήματος υφής συλλαμβάνεται έτσι από τις παραμέτρους της DCA.	79
4.3	Χαρακτηριστικά υφής μέσω DCA και αποδιαμόρφωσης. (α) Συνθετικό σήμα υφής, (β) Κωδικοποίηση κυριάρχου καναλιού i_{DCA} μέσω φωτεινότητας. (ς) Εκτίμηση πλάτους a_i του κυριάρχου καναλιού και (δ) Εκτίμηση πλάτους A_i βάσει της σχέσης (4.8). Το εκτιμώμενο πλάτος κωδικοποιείται μέσω της τιμής της φωτεινότητας της εικόνας.	80
4.4	Τα στοιχεία της βάσης (συμπαγής γραμμή) και συνάρτηση εμπιστοσύνης (διακεκομμένη) για τις 1Δ τομές των τριών θεωρούμενων κλάσεων: (α) Υφή (β) Ακμή (γ) Ομαλά Τμήματα.	82
4.5	Ανακατασκευή της μη-ομαλής συνιστώσας της εικόνας (α) χρησιμοποιώντας (β) τις παραμέτρους του μοντέλου ακμών, (4.27) και (γ) τις παραμέτρους του μοντέλου υφής, (4.12). Η ανακατασκευή στην εικόνα (γ) αναδεικνύει τις περιοδικές πτυχές της τοπικής δομής, αντίθετα με την εικόνα (β) όπου απότομες, μη-περιοδικές μεταβάσεις είναι πιο έντονες.	87
4.6	Διάκριση περιοχών υφής από ακμές, χρησιμοποιώντας τοπικά αναγεννητικά μοντέλα: Εικόνα (α) και πλάτη των μοντέλων υφής (β) και ακμής (γ) $\sqrt{A_1^2 + A_2^2}$, με τα A_1, A_2 εκτιμώμενα βάσει του τύπου (4.20). Τα πλάτη στις εικόνες (β) και (γ) είναι ανάλογα με την πιθανοφάνεια των δεδομένων, δεδομένης της κλάσης του σήματος. Οι ύστερες πιθανότητες των κλάσεων ομαλού σήματος (δ), σήματος υφής (ε) και ακμής (στ) προκύπτουν από τον τύπο του Bayes. Η φωτεινότητα της εικόνας κωδικοποιεί την πιθανότητα της αντίστοιχης υπόθεσης.	91
4.7	Φυσικές εικόνες (α) και ύστερες πιθανότητες των μοντέλων υφής (β) και ακμής (γ) αντίστοιχα. Η φωτεινότητα της εικόνας κωδικοποιεί την πιθανότητα της αντίστοιχης υπόθεσης.	92
4.8	Αποτελέσματα κατάτμησης χρησιμοποιώντας διαφορετικά χαρακτηριστικά και αλγόριθμους εξέλιξης καμπυλών: (α) Είσοδος (β) Κατάτμηση χρησιμοποιώντας χαρακτηριστικά DCA και την φωτεινότητα [129]. (γ) κατάτμηση με τα χαρακτηριστικά διάχυσης των [193]. (δ) Κατάτμηση χρησιμοποιώντας χαρακτηριστικά διάχυσης και σταθμισμένη εξέλιξη καμπυλών (4.39).	97
4.9	Αποτελέσματα κατάτμησης χρησιμοποιώντας χαρακτηριστικά διαμόρφωσης και σταθμισμένη εξέλιξη καμπυλών.	98
5.1	Επισκόπηση της προσέγγισής μας.	102
5.2	Εικόνα στην είσοδο (α) και επαναλήψεις του EM (β)-(δ). Πάνω γραμμή: εξέλιξη του χωρίου της υπόθεσης του αντικειμένου σε επίπεδο pixels (βήμα E). Κάτω γραμμή: αποτελέσματα σύνθεσης, χρησιμοποιώντας τα παραπάνω βάρη (βήμα M).	104
5.3	Χειρισμός επικαλύψεων: Η σύνθεση στην εικόνα (γ) χρησιμοποιεί την κατάτμηση στην εικόνα (β), ενώ αυτή της εικόνας (δ) δεν έχει κάποια πληροφορία σχετικά με το χωρίο της περιοχής του προσώπου.	105

- 5.4 Το αναγεννητικό μοντέλο της προσέγγισής μας: οι παράμετροι σχήματος και εμφάνισης s, t συνθέτουν την εμφάνιση του αντικειμένου μέσω ενός μοντέλου, που στη συνέχεια συνδυάζεται με το τμηματικά σταθερό μοντέλο του φόντου. Ο συνδυασμός επιτυγχάνεται μέσω ενός πίνακα κατάτμησης (κρυφές μεταβλητές), που καθορίζει ποια τμήματα της εικόνας ανατίθενται στην υπόθεση του αντικειμένου. Για την εκτίμηση του πίνακα κατάτμησης συνδυάζεται η πιστότητα του μοντέλου στις παρατηρήσεις με πληροφορία σχετικά με το χωρίο του αντικειμένου. Για εκτίμηση παραμέτρων η πιθανοφάνεια των παρατηρήσεων που βρίσκονται μέσα στη 'μάσκα' μεγιστοποιείται. 106
- 5.5 Αποτελέσματα άνωθεν κατάτμησης για εικόνες αυτοκινήτων και προσώπων, χρησιμοποιώντας το βήμα E σε επίπεδο τμημάτων. Χρησιμοποιώντας ένα σταθερό κατώφλι για τα αποτελέσματα του βήματος E, τα τμήματα της εικόνας που ανήκουν στην υπόθεση του αντικειμένου της ανατίθενται. Αν και η κατάτμηση του αντικειμένου είναι χοντρικά σωστή, εμφανίζονται προβλήματα λεπτής κλίμακας, όπως μία ατυχής κατάτμηση χαμηλού επιπέδου να έχει ως αντίκτυπο στην άνωθεν κατάτμηση τον ανακριβή εντοπισμό των ορίων και γωνιώδη περιγράμματα. 109
- 5.6 Σύγκριση μεταξύ των δύο μεθόδων άνωθεν κατάτμησης: Η εξέλιξη καμπυλών παρέχει καλύτερα αποτελέσματα από τη μέθοδο που υλοποιεί το βήμα E σε επίπεδο τμημάτων, δίνοντας ομαλά περιγράμματα, που εντοπίζουν σωστά τα όρια των αντικειμένων. 114
- 5.7 Διαφορές στην προσαρμογή χρησιμοποιώντας εναλλακτικά κριτήρια AAM: (α) Εικόνα προς ανάλυση, (β) ανακατασκευή μέσω του απλού κριτηρίου ελάχιστων τετραγώνων, (γ) εισάγοντας πληροφορία κατάτμησης μέσω του EM, (δ) τα αποτελέσματα του βήματος E. Εισάγοντας τον EM καθίσταται εύρωστη η προσαρμογή των AAM καθώς αγνοούνται περιοχές με μη μοντελοποιούμενη εμφάνιση, όπως το τρίτο παράθυρο ή τα μαλλιά στις κάτω γραμμές. 119
- 5.8 Πάνω γραμμή: Η άνωθεν πληροφορία μπορεί να βελτιώσει τις επιδόσεις ανίχνευσης περιορίζοντας τις εσφαλμένες θετικές αποφάσεις (false positives): η περίπλοκη δομή στο φόντο, προκαλεί μία εσφαλμένη πρόταση για τη θέση του αντικειμένου, που φαίνεται με την κόκκινη διακεκομμένη γραμμή στην εικόνα (α). Αυτή η υπόθεση απορρίπτεται τόσο λόγω της χαμηλής πιθανοφάνειας των παραμέτρων, που οδηγούν στη σύνθεση μίας περίεργης μορφής στην εικόνα (β) όσο και στα χαμηλότερα αποτελέσματα του βήματος E, που φαίνονται με μία χαμηλότερη φωτεινότητα στην εικόνα (γ). Κάτω γραμμές: αποτελέσματα άνωθεν/κάτωθεν ανίχνευσης βάσει αυτής της προσέγγισης. Οι κόκκινες διακεκομμένες γραμμές δείχνουν υποθέσεις που απορρίφθηκαν από το άνωθεν σύστημα. 122
- 5.9 Αποτελέσματα ανίχνευσης στο σύνολο δεδομένων της αναφοράς [62]: παρατηρούμε ότι ο συνδυασμός άνωθεν και κάτωθεν πληροφορίας οδηγεί σε καλύτερες επιδόσεις ανίχνευσης. 123

5.10	Καμπύλες Ακρίβειας-Ανάκλησης για την ανίχνευση προσώπων με επικάλυψη χρησιμοποιώντας (α) απλή προσαρμογή AAM και (β) τα αποτελέσματα του EM. Εδώ η επίδραση του συνδυασμού κατάτμησης και ανίχνευσης γίνεται πιο εμφανής, καθώς η χρήση του EM εισάγει ευρωστία στην προσαρμογή των AAM. Χωρίς τον EM ο ταξινομητής που χρησιμοποιεί τα χαρακτηριστικά του AAM αποτυγχάνει οδηγώντας σε χειρότερες επιδόσεις.	124
5.11	Μέσοι πίνακες ακμών για ένα αυξανόμενο πλήθος επαναλήψεων του αλγόριθμου εκμάθησης AAM: (α) Χωρίς αντιστοίχιση, (β) ενδιάμεσα αποτελέσματα και (γ) μοντέλο που χρησιμοποιείται στα πειράματα ανίχνευσης. Ο μέσος πίνακας ακμών αποκτάται από το μέτρο της παραγωγής των εικόνων του συνόλου εκπαίδευσης, μετά την αντιστοίχισή τους με το μοντέλο AAM. Ξεκινώντας από ένα ασαφή μέσο, σταδιακά προκύπτουν ξεκάθαρες, περίπλοκες δομές όπως το εσωτερικό των ματιών και των τροχών, ή ακμές πάνω στη μύτη και τα παράθυρα. . . .	126
5.12	Αποτελέσματα εκμάθησης μοντέλου AAM: Αν και το σύνολο εκπαίδευσης περιέχει μεγάλη ποικιλία στο σχήμα και στην εμφάνιση, περίπλοκη δομή στο φόντο και επικαλύψεις των αντικειμένων, οι συνθέσεις είναι καθαρές και συλλαμβάνουν την εμφάνιση των αντικειμένων της κατηγορίας.	128
5.13	Καμπύλες Ακρίβειας-Ανάκλησης για ανίχνευση αυτοκινήτων και προσώπων, χρησιμοποιώντας τα αυτόματα κατασκευασμένα μοντέλα. Και στις δύο περιπτώσεις, η χρήση του αλγόριθμου EM (στήλης (α),(γ)) οδηγεί σε βελτιωμένες επιδόσεις, τόσο λόγω των καθαρότερων παραμέτρων AAM όσο και της πληροφορίας κατάτμησης.	129
5.14	Καθολικές (α) και τμηματικές (β) αρχιτεκτονικές για το πρόβλημα της συνέργειας	129
6.1	Gaussian scale space της εικόνας, που προκύπτει συνελίσσοντας την εικόνα με Gaussian πυρήνες αυξανόμενης διασποράς σ . Σε μεγαλύτερες κλίμακες επικρατούν οι κυριότερες δομές της εικόνας, αλλά μετακινούνται και θολώνουν τα όριά τους.	140
6.2	Πολυκλιμακωτή εξαγωγή χαρακτηριστικών ακμών, χρησιμοποιώντας τον τελεστή \mathcal{G}_γ (6.1) για διαφορετικές τιμές της διασποράς του Gaussian πυρήνα· σκοτεινές περιοχές δείχνουν μία υψηλή τιμή του διαφορικού τελεστή. Η κλίμακα στην οποία μεγιστοποιείται η απόκριση του κανονικοποιημένου τελεστή καθορίζεται από το εύρος της ακμής. . . .	140
6.3	Πολυκλιμακωτή εξαγωγή ράβδων, χρησιμοποιώντας τον τελεστή \mathcal{A}_γ (6.3) για διαφορετικές τιμές της διασποράς του Gaussian πυρήνα. Η κλίμακα στην οποία μεγιστοποιείται η απόκριση του τελεστή καθορίζεται από το πάχος της δομής τύπου ράβδου.	142
6.4	Πολυκλιμακωτή εξαγωγή κηλίδων, με τον τελεστή \mathcal{H}_γ (6.3)· σκοτεινές περιοχές δείχνουν μία υψηλή τιμή του διαφορικού τελεστή. Τα μέγιστα εμφανίζονται σε σημεία όπου η εικόνα μπορεί τοπικά να προσεγγιστεί μέσω μίας κυκλικής δομής	143

- 6.5 Χαρακτηριστικά ράβδων και ακμών για τέσσερις κατηγορίες αντικειμένων. Αριστερή στήλη: μέγιστα στο χώρο και στην κλίμακα της έντασης ακμών και ράβδων (πράσινο και γαλάζιο χρώμα, αντίστοιχα). Μέση και δεξιά στήλη: ευθείες ράβδοι και ακμές εξαγόμενες μέσω του ανεξάρτητου- κλίμακας κριτηρίου προσαρμογής ακμών. Το πάχος των ράβδων και η ευρύτητα των ακμών εκτιμώνται από την τοπική δομή της εικόνας και καθορίζουν τους λόγους των αξόνων των ελλείψεων και τα μήκη των κάθετων γραμμών στα άκρα των ακμών, αντίστοιχα. 145
- 6.6 Επιλογή καταχωρήσεων codebook: από τις συστάδες που προκύπτουν από την ομαδοποίηση των χαρακτηριστικών σημείων επιλέγεται το υποσύνολό που είναι χρησιμότερο για την ανίχνευση του αντικειμένου. . . 147
- 6.7 Αναπαραστάσεις για πρόσωπα, άλογα, αυτοκίνητα και αγελάδες που προκύπτουν από τη διαδικασία επιλογής καταχωρήσεων. Οι κόκκινοι κύκλοι συμβολίζουν χαρακτηριστικά κηλίδων, οι πράσινες γραμμές ακμές και οι γαλάζιες ελλείψεις ράβδους. 148
- 6.8 Μηχανισμός εκτίμησης του $\sum_{\mathcal{P}_i} \Phi_i(\mathcal{P}_i) \prod_j m_{i,j}(\mathcal{P}_i)$ για έξι διαφορετικούς κόμβους: Ράβδοι στο μπροστά και πίσω πόδι, ακμές στην πλάτη και στο δεύτερο πόδι και κηλίδες σε δύο διακλαδώσεις). (α) Δείγματα από το μήνυμα που έρχεται από το υπόλοιπο γραφικό δίκτυο $\prod_j m_{i,j}(\mathcal{P}_i)$. (β) Πέντε ισχυρότερα δείγματα, βάσει του όρου $\Phi_i(\mathcal{P}_i)$ για (γ) Δείγματα από την κατανομή $\sum_{\mathcal{P}_i} \Phi_i(\mathcal{P}_i) \prod_j m_{i,j}(\mathcal{P}_i)$ που προκύπτει προσεγγίζοντας την $\sum_{\mathcal{P}_i} \Phi_i(\mathcal{P}_i) \prod_j m_{i,j}(\mathcal{P}_i) \simeq \sum_{i=1}^N l_n \delta(P - S_n)$ με μία Gaussian. Ο υπολογισμός του αθροίσματος πάνω σε όλα τα \mathcal{P}_i αποφεύγεται, εξετάζοντας μόνο δείγματα από την υποπεριοχή του χώρου θέσης που 'προτείνει' το γραφικό δίκτυο. 152
- 6.9 Τεμπλέτες για χαρακτηριστικά τύπου (α) ακμών (β) ράβδων και (γ) κηλίδων, και εφαπτόμενα διανύσματα στα σύνορα των υποτιμημάτων της τεμπλέτας. Επιφανειακά ολοκληρώματα πάνω στα υποτιμήματα μπορούν να αντικατασταθούν με επικαμπύλια, ελατώνοντας τον χρόνο υπολογισμού από $O(N^2)$ σε $O(N)$, όπου N η μία διάσταση της τεμπλέτας. Η πυκνότητα των διανυσμάτων είναι ανάλογη του βήματος ολοκλήρωσης και επιτρέπει τον έλεγχο του συμβιβασμού ταχύτητας/ακρίβειας 153
- 6.10 Σημεία τύπου ράβδων, ακμών και κηλίδων που ανιχνεύονται χρησιμοποιώντας άνωθεν γνώση, κατωφλιώνοντας την πιθανότητα παρατήρησης του τμήματος του αντικειμένου. 154
- 6.11 Δομή γραφικού μοντέλου για τα χαρακτηριστικά που εξήγαγε το άνωθεν σύστημα για 40 καταχωρήσεις. 156
- 6.12 Μηχανισμός περιορισμού θετικών αποκρίσεων: Το κάτωθεν σύστημα ανίχνευσης προτείνει δύο δυνατές τοποθεσίες της εικόνας ως πιθανές να περιέχουν ένα πρόσωπο. Ενεργοποιώντας το άνωθεν μοντέλο το οποίο αναζητεί περιοχές στην εικόνα που να αντιστοιχούν στη δομή του, για την ψευδή υπόθεση βρίσκονται λίγες αντιστοιχίες, και αυτές συμπτωματικά: αντίθετα για την αληθή υπόθεση οι περισσότερες καταχωρήσεις του codebook έχουν βρεί μία αντίστοιχη δομή στην εικόνα. Έτσι, εξασθενεί η ψευδής υπόθεση και ισχυροποιείται η αληθής. . . . 157

- 6.13 Αποτελέσματα συνεργετικής ανίχνευσης αντικειμένων: το κάτωθεν σύστημα προτείνει ένα σύνολο πιθανών τοποθεσιών για αντικείμενα, οι οποίες επαληθεύονται από το άνωθεν σύστημα, όπως περιγράφεται στο κείμενο. Για μία συγκεκριμένη τιμή ενός κατωφλιού, οι τοποθεσίες με κόκκινη διακεκομμένη γραμμή απορρίπτονται, ενώ με την πράσινη συνεχή εγκρίνονται. Γενικότερα έτσι περιορίζονται οι ψευδείς ανιχνεύσεις, με το κόστος ορισμένων χαμένων θετικών (βλ. δεξιά στήλη). . . . 158
- 6.14 Πάνω γραμμή: (α) Βελτιώσεις στην επίδοση που εισάγονται σε διαφορετικά στάδια επεξεργασίας: BU: bottom-up, NB: naive Bayes, TD: Top-down, GM: Graphical model, GMTD: Graphical model & Top-down. (β) Συμβιβασμός ακρίβειας / ταχύτητας: η επίδοση του άνωθεν ανιχνευτή βελτιώνεται αυξάνοντας το πλήθος των υποθέσεων (particles) και ελαττώνοντας το βήμα ολοκλήρωσης. (γ) Σύγκριση με αποτελέσματα άλλων μεθόδων, μετά την καταπίεση μη-μεγίστων. Κάτω γραμμή: Συγκριτικά αποτελέσματα για τις άλλες κατηγορίες αντικειμένων. . . . 159

Κατάλογος Πινάκων

- 2.1 Ταξινόμηση προβλημάτων που θίγονται στα επόμενα κεφάλαια της διατριβής. 43
- 4.1 Ανακατασκευή σήματος μέσω διαφορετικών αναγεννητικών μοντέλων: Για κάθε γραμμή του πίνακα ένα σήμα από μία διαφορετική κλάση έχει συντεθεί (έντονη γραμμή), και επιδεικνύεται η ανακατασκευή του μέσω των διαθέσιμων αναγεννητικών μοντέλων (αχνή γραμμή), χρησιμοποιώντας τη συνάρτηση εμπιστοσύνης που φαίνεται με διακεκομμένη γραμμή. Από πάνω προς τα κάτω, το σήμα έχει προκύψει από το μοντέλο ακμής, υφής και ομαλών σημάτων. Το σφάλμα ανακατασκευής είναι μικρότερο όταν χρησιμοποιείται το αντίστοιχο του σήματος αναγεννητικό μοντέλο· εξαίρεση αποτελεί η περίπτωση του ομαλού μοντέλου, όπου το μικρότερο πλήθος παραμέτρων οδηγεί πάντα σε μεγαλύτερο σφάλμα ανακατασκευής. 89



ΥΠΟΥΡΓΕΙΟ ΕΘΝΙΚΗΣ ΠΑΙΔΕΙΑΣ ΚΑΙ ΘΡΗΣΚΕΥΜΑΤΩΝ
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ ΕΠΕΑΕΚ
ΕΥΡΩΠΑΪΚΗ ΕΝΩΣΗ
ΣΥΓΧΡΗΜΑΤΟΔΟΤΗΣΗ
ΕΥΡΩΠΑΪΚΟ ΚΟΙΝΩΝΙΚΟ ΤΑΜΕΙΟ
ΕΥΡΩΠΑΪΚΟ ΤΑΜΕΙΟ ΠΕΡΙΦΕΡΕΙΑΚΗΣ ΑΝΑΠΤΥΞΗΣ



Η ΠΑΙΔΕΙΑ ΣΤΗΝ ΚΟΡΥΦΗ
Επιχειρησιακό Πρόγραμμα
Εκπαίδευσης και Αρχικής
Επαγγελματικής Κατάρτισης

Η παρούσα διδακτορική διατριβή αποτελεί υποέργο του προγράμματος:
“Ηράκλειτος: Υποτροφίες έρευνας με προτεραιότητα στην βασική έρευνα”

Το Πρόγραμμα “ΗΡΑΚΛΕΙΤΟΣ” συγχρηματοδοτείται από το
Ευρωπαϊκό Κοινωνικό Ταμείο (75%) και από Εθνικούς Πόρους (25%).

The Project "HRAKLEITOS" is co-funded by the
European Social Fund (75%) and National Resources (25%).

ΠΡΟΛΟΓΟΣ

Η όραση υπολογιστών είναι ένα ευρύ ερευνητικό πεδίο με ενδιαφέρον τόσο θεωρητικό όσο και πρακτικό· πέρα από τις ιατρικές, βιομηχανικές και εμπορικές εφαρμογές της, η κατανόηση και μοντελοποίηση των μηχανισμών της όρασης είναι ένα γοητευτικό όσο και δύσκολο πρόβλημα.

Παρά τις μεγάλες προσδοκίες και την αυξανόμενη υπολογιστική ισχύ, μετά από έρευνα περίπου τριών δεκαετιών οι πρακτικές εφαρμογές της όρασης υπολογιστών παραμένουν περιορισμένες σε πολύ συγκεκριμένα προβλήματα. Πέρα από τη δεδομένη δυσκολία του προβλήματος, αυτό θα μπορούσε να αποδοθεί και στην τύπου 'διαίρει και βασίλευε' τακτική που έχει ακολουθηθεί από τις πρώτες δεκαετίες. Στα πλαίσια αυτής της διατριβής επιδιώκεται μία συνθετική προσέγγιση σε προβλήματα της όρασης υπολογιστών, έχοντας ως βασικό στόχο την από κοινού αντιμετώπιση δύο κύριων περιοχών της όρασης, την κατάτμηση εικόνων και την αναγνώριση αντικειμένων.

Για την εμπιστοσύνη και τη στήριξη του κατά την ενασχόλησή μου με ένα θέμα τόσο συναρπαστικό ευχαριστώ τον επιβλέποντά μου, Καθ. Πέτρο Μαραγκό. Από τις πρώτες ημέρες της συνεργασίας μας με 'πίεσε' να οργανώσω την εργασία και τον τρόπο σκέψης μου, χωρίς ποτέ να περιορίσει την επιστημονική μου ελευθερία, και πιστεύω ότι αυτή η διατριβή δεν θα ήταν εφικτή αν δεν είχα παραδειγματιστεί από την επιμονή, υπομονή και τον ενθουσιασμό του για την έρευνα.

Θα ήθελα να ευχαριστήσω και τα άλλα δύο μέλη της τριμελούς μου επιτροπής, Καθ. Στέφανο Κόλλια και Καθ. Ανδρέα Σταφυλοπάτη οι οποίοι ήταν πάντα πρόθυμοι να με βοηθήσουν και να με στηρίζουν όποτε τους το ζήτησα. Ευχαριστώ επίσης τους Καθ. Σπυρίδωνα Τζαφέστα, Καθ. Νικόλαο Μαράτο, Καθ. Τιμολέοντα Σελλή, και Καθ. Νικόλαο Γαλατσάνο, οι οποίοι δεχτήκαν να είναι μέλη της επταμελούς μου επιτροπής.

Ωφελήθηκα σημαντικά από την παραμονή μου στην ερευνητική ομάδα Odyssee του INRIA στην Sophia Antipolis και από την ερευνητική μου ενασχόληση με το βιολογικό σύστημα όρασης που μου πρότειναν οι Dr. Rachid Deriche και Καθ. Olivier Faugeras. Επίσης, ευχαριστώ τον Καθ. Alan Yuille για την συνεργασία που είχαμε κατά τη σύντομη αλλά συναρπαστική περίοδο της παραμονής μου στο UCLA.

Ήταν μεγάλη μου τύχη που κατά την διάρκεια του διδακτορικού μου εργαστηρίου σε ένα περιβάλλον με σπάνιους ανθρώπους, με τους οποίους ελπίζω να παραμείνουμε δεμένοι και στο μέλλον όπως τώρα. Ιδιαίτερα θα ήθελα να ευχαριστήσω τον Βασίλη Πιτσικάλη, ο οποίος για πάνω από πέντε χρόνια αποτέλεσε υπόδειγμα συντροφικότητας, τη Χάρης Παπασάικα που σε τόσες περιστάσεις 'έσωσε' την κατάσταση, το πιθανοτικό cluster του Γιώργου Παπανδρέου και Νάσσου Κατσαμάνη για συζητήσεις, σχόλια, αναφορές και την βοήθειά τους, τον Γιώργο Ευαγγελόπουλο για τη συνεργασία που είχαμε στην ανάλυση υφής και την Νατάσα Σοφού για τα πάντα

εμπυχωτικά της σχόλια. Στον παλαιότερο του εργαστηρίου μας, Δημήτρη Δημη-
τριάδη εύχομαι καλή σταδιοδρομία, και στους νεότερους, Σταμάτη Λευκιμιάτη και
Τάσο Ρούσσο καλή συνέχεια. Τέλος, θα ήθελα να ευχαριστήσω τους φίλους μου που
συνέχισαν να με ψάχνουν, παρά την πολυετή εξαφάνισή μου.

Αυτή η διατριβή, όπως και ό,τι άλλο οφείλεται στους γονείς μου, Μαριέττα και
Στέφανο. Ελπίζω να αποτελεί μία μικρή ανταπόδοση για τη στοργή, φροντίδα και
υπομονή τους για όλα αυτά τα χρόνια.

Αθήνα, Ιούνιος 2006

ΠΕΡΙΛΗΨΗ

Αυτή η διατριβή ασχολείται με τα προβλήματα της αναγνώρισης αντικειμένων και της κατάρτησης εικόνων, έχοντας ως βασικό στόχο τον συνδυασμό τους. Το πλαίσιο μέσα στο οποίο διατυπώνονται οι σημαντικότερες συνεισφορές είναι η πιθανοτική, Bayesian, προσέγγιση σε προβλήματα όρασης υπολογιστών, που μπορεί να συνοψιστεί στην ιδέα της Ανάλυσης-μέσω-Σύνθεσης: έχοντας στη διάθεσή μας ένα μοντέλο σύνθεσης της εικόνας, η ανάλυση μίας νέας εικόνας ανάγεται στην προσαρμογή του μοντέλου.

Υιοθετώντας την πιθανοτική προσέγγιση προβλημάτων όπως η ανίχνευση ακμών, η μοντελοποίηση υφής, η κατάρτηση και η αναγνώριση αντικειμένων διατυπώνονται σε ένα κοινό πλαίσιο, μέσω αναγεννητικών (generative) και γραφικών μοντέλων. Κοινός άξονας για όλες τις συνεισφορές της διατριβής αποτελεί η διατύπωση και χρήση μοντέλων που να ανταποκρίνονται στις ιδιαιτερότητες του προβλήματος της όρασης. Η καινοτομία που εισάγει αυτή η διατριβή μπορεί να συνοψιστεί στα παρακάτω σημεία:

- Εισαγωγή του αλγόριθμου Expectation Maximization για τη διατύπωση του προβλήματος της Συνέργειας.
- Συνδυασμός Μοντέλων Ενεργούς Εμφάνισης (Active Appearance Models - AAM) και Τεχνικών Επιπεδοσυνόλων για τη Συνέργεια Κατάρτησης και Αναγνώρισης.
- Χρήση Γραφικών Μοντέλων και Σημείων Ενδιαφέροντος για την από κοινού Άνωθεν και Κάτωθεν Ανίχνευση Αντικειμένων.
- Τοπικά Αναγεννητικά Μοντέλα για την Ανάλυση και Κατάρτηση Εικόνων Υφής, διάκριση Υφής/Ακμών και τη Σταθμισμένη Εξέλιξη Καμπυλών.
- Πιθανοτική Διατύπωση και Εκμάθηση ενός Μοντέλου Χαμηλού και Μέσου Επιπέδου Λειτουργιών του Βιολογικού Συστήματος Όρασης.

ABSTRACT

This dissertation deals with the problems of object recognition and image segmentation, focusing primarily on their combination. The major contributions of this dissertation build on the probabilistic, Bayesian approach to computer vision problems, whose essence is captured in the Analysis-by-Synthesis framework: once an image synthesis model is available, image analysis can be phrased as model fitting.

Adopting the probabilistic approach, problems like edge detection, texture modeling, image segmentation and object recognition are phrased in a common setting, using generative and graphical models. The common theme for all the contributions of this dissertation is the formulation and usage of probabilistic models that account for the specific aspects of the vision problem considered. The novelty introduced in this thesis can be summarized in the following points:

- Introduction of the Expectation Maximization algorithm for the formulation of the Synergy problem.
- Combination of Active Appearance Models and Level Set techniques for the Synergy between Segmentation and Recognition.
- Joint Bottom-up and Top-down Object Detection using Graphical Models and Interest Point Features.
- Local Generative Models for Texture Analysis and Segmentation, Edge/Texture Discrimination and Weighted Curve Evolution.
- Probabilistic Formulation and Learning of a Biologically Motivated Model of Low- and Mid- Level tasks.

Κεφάλαιο 1

Εισαγωγή

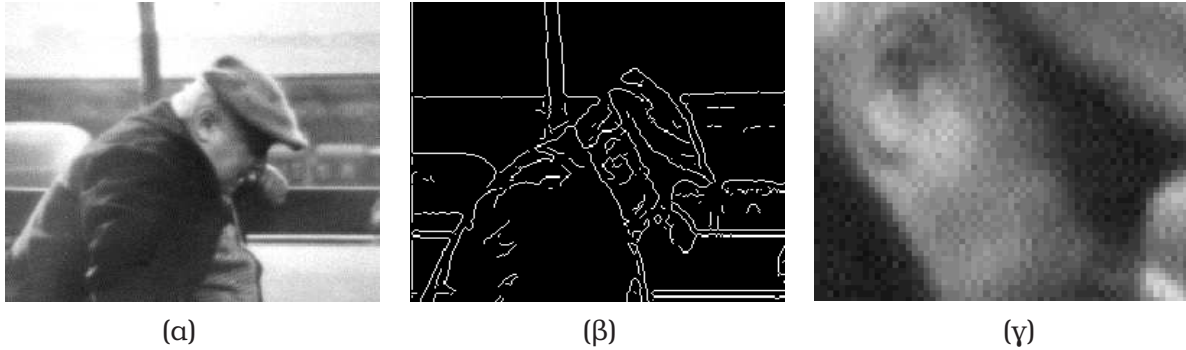
1.1 Το Πρόβλημα της Συνέργειας

Το κυρίαρχο παράδειγμα της όρασης υπολογιστών μπορεί να συνοψιστεί στο μοντέλο που παρουσιάζεται στην αναφορά [153], σύμφωνα με το οποίο η επεξεργασία της οπτικής πληροφορίας ακολουθεί μία κάτωθεν (bottom-up) πορεία, με διαδικασίες αυξανόμενης πολυπλοκότητας να χρησιμοποιούν τα αποτελέσματα των προηγούμενων, απλούστερων σταδίων. Αυτό το μοντέλο είναι δημοφιλές καθώς επιτρέπει να εξετάσει κανείς το κάθε υποπρόβλημα χωριστά, αποφεύγοντας ένα αυξημένης πολυπλοκότητας μοντέλο. Βάσει αυτής της προσέγγισης διαφορετικές κοινότητες της όρασης υπολογιστών έχουν εστιάσει σε μεμονωμένα προβλήματα, με την προσδοκία ότι κάποτε με αρκετή υπολογιστική ισχύ και βελτιώσεις των υπάρχοντων μοντέλων και τεχνικών θα επιτευχθεί μηχανιστικά η ίδια επίδοση με το βιολογικό σύστημα όρασης. Αν και ουσιαστική πρόοδος έχει σημειωθεί έτσι στα επιμέρους προβλήματα, ελάχιστα συστήματα έχουν κατασκευαστεί που να είναι λειτουργικά σε μη-τετριμμένα περιβάλλοντα ενώ οι πρακτικές εφαρμογές της όρασης υπολογιστών παραμένουν περιορισμένες.

Η αποτυχία αυτή μπορεί να αποδοθεί στις απλουστεύσεις για τις υπόλοιπες συνιστώσες του προβλήματος που χρησιμοποιεί η κάθε υποκοινότητα. Για παράδειγμα, οι περισσότεροι αλγόριθμοι μέσω λειτουργιών της όρασης, όπως κατάτμηση εικόνων, εκτίμηση κίνησης ή αντίληψη βάθους θεωρούν ότι τα αντικείμενα στο εσωτερικό τους μπορούν να μοντελοποιηθούν ως τμηματικά συνεχείς συναρτήσεις φωτεινότητας και διατυπώνουν μαθηματικά προβλήματα που βασίζονται στην παραδοχή αυτή. Σε ένα υψηλότερο επίπεδο, αρκετοί αλγόριθμοι αναγνώρισης αντικειμένων θεωρούν ότι το αντικείμενο έχει διαχωριστεί από το περιβάλλον του, δηλαδή ότι το πρόβλημα της κατάτμησης έχει ήδη λυθεί. Κατά τη διάρκεια αυτής της δεκαετίας έχει αρχίσει να διαφαίνεται ότι πολυπλοκότερες προσεγγίσεις είναι απαραίτητες, που να αντιμετωπίζουν το πρόβλημα της όρασης σε περισσότερα του ενός επίπεδα, επιτρέποντας την επίλυση δύσκολων προβλημάτων αντίληψης συνδυάζοντας έτσι τα επιμέρους υποπρόβλημά τους (συνέργεια). Στηρίζοντας αυτήν την προσέγγιση, αποτελέσματα από το χώρο της Γνωστικής Ψυχολογίας [171] και της Ψυχοφυσιολογίας [143, 142] καταδεικνύουν ότι και στο βιολογικό σύστημα όρασης η ροή της πληροφορίας ακολουθεί ένα σύστημα ανάδρασης, όπου η συνεργασία μεταξύ χαμηλότερων και υψηλότερων σταδίων επεξεργασίας είναι συνεχής και αμφίδρομη.

Ένα παράδειγμα που αναδεικνύει την χρησιμότητα μίας τέτοιας προσέγγισης φαίνεται στο σχήμα Σχ. 1.1 [165]. Ενώ εμείς αντιλαμβανόμαστε άμεσα την ύπαρ-

ξη ενός ατόμου μέσα στην εικόνα και θεωρούμε την υπόλοιπη εικόνα ως φόντο (background), όταν ένας υπολογιστής προσπαθήσει να κατατμήσει την εικόνα και να κάνει αναγνώριση αντικειμένων (εν προκειμένω πλαϊνών όψεων προσώπου), δύσκολα θα δώσει σωστά αποτελέσματα. Σε ότι αφορά την κατάτμηση, όπως φαίνεται στην εικόνα Σχ. 1.1 (β) οι μεταβολές της φωτεινότητας σπάνε την περιοχή του προσώπου σε περισσότερα μικρότερα τμήματα ενώ αφήνουν το πίσω μέρος του κεφαλιού να ενωθεί με το φόντο. Όσο για την αναγνώριση, η ανάλυση του προσώπου είναι τόσο χαμηλή που δεν επιτρέπει την ανίχνευσή του με κάποια σιγουριά.

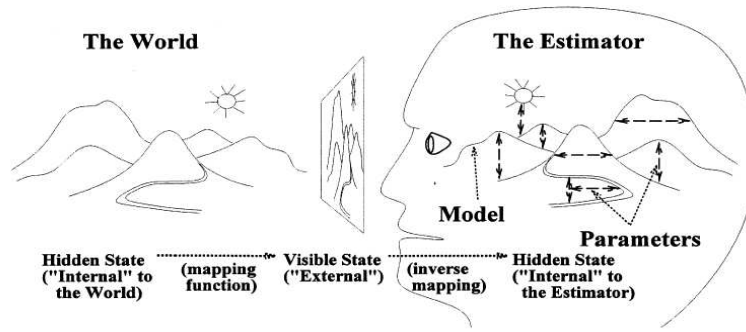


Σχήμα 1.1: Αναγκαιότητα της από κοινού αντιμετώπισης λειτουργιών όρασης χαμηλού και υψηλού επιπέδου [165]: (α) Εικόνα που περιέχει ένα αντικείμενο (β) Αποτελέσματα ανίχνευσης ακμών (γ) Μεγέθυνση της περιοχής του προσώπου. Η εικόνα δεν είναι κατάλληλη για την εφαρμογή τυπικών μεθόδων κατάτμησης και ανίχνευσης αντικειμένων, λόγω της χαμηλής ανάλυσης, της έντονης σκίασης και της περίπλοκης δομής της. Χρησιμοποιώντας όμως το μοντέλο ενός ανθρώπου η ερμηνεία της διευκολύνεται, επιτρέποντας την αποτελεσματική εκτέλεση των επιμέρους λειτουργιών της κατάτμησης και ανίχνευσης.

Η ιδέα της συνέργειας είναι ότι αν, έστω και με μικρή βεβαιότητα, αρχικά θεωρήσουμε ότι υπάρχει ένα πρόσωπο στην εικόνα, μπορούμε να ερμηνεύσουμε οπότε και να αγνοήσουμε τις ακμές που προκύπτουν από την σκίαση, ενώ γνωρίζοντας το σχήμα ενός προσώπου καταλαβαίνουμε ότι δεν ενώνεται με το υπόλοιπο φωτεινό τμήμα της εικόνας. Από την άλλη, έχοντας βρει μία περιοχή της εικόνας που να συμφωνεί με τις προβλέψεις του μοντέλου του προσώπου μπορούμε να αυξήσουμε την εμπιστοσύνη μας στην υπόθεση αυτή, η οποία αρχικά μπορεί να βασίστηκε μόνο σε τοπικά χαρακτηριστικά, όπως η μύτη.

1.2 Προσέγγιση στο πρόβλημα της Συνέργειας

Το γενικότερο πλαίσιο της άνωθεν (top - down) όρασης στο οποίο έχουν διατυπωθεί οι παραπάνω ιδέες μπορεί να συνοψιστεί στην ιδέα της *Ανάλυσης - μέσω - Σύυθσεως*. Η προσέγγιση αυτή βασίζεται στην χρήση μοντέλων για τα αντικείμενα του οπτικού μας περιβάλλοντος, και δεδομένης μίας νέας σκηνής, ανάγει την ερμηνεία της στην ανακατασκευή της μέσω των μοντέλων που έχουμε στη διάθεσή μας, π.χ. για πρόσωπα, αυτοκίνητα, κτήρια κ.ο.κ. (Σχ. 1.2). Έτσι, η ερμηνεία της σκηνής δεν ακολουθεί μία καθαρά κάτωθεν πορεία αλλά καθορίζεται και από την πρότερη γνώση (prior knowledge) που διαθέτουμε για τον τρόπο με τον οποίον μπορούν να έχουν προκύψει οι παρατηρήσεις από τα αντικείμενα. Η κάτωθεν ροή πληροφορίας μπορεί να θεωρηθεί ότι επιτυγχάνει την επίσπευση αυτής της διαδικασίας, παρέχοντας



Σχήμα 1.2: Μοντέλο της άνωθεν όρασης [190]: παραμετρικά μοντέλα που υπάρχουν για τα οπτικά αντικείμενα χρησιμοποιούνται για την ερμηνεία μίας σκηνής μέσω της ανακατασκευής της.

ενδείξεις στο άνωθεν σύστημα όρασης σχετικά με το ποιες υποθέσεις είναι ωφέλιμο να εξεταστούν.

Ένα μαθηματικό πλαίσιο στο οποίο έχει διατυπωθεί αυτή η προσέγγιση είναι η Θεωρία Προτύπων (Pattern Theory) την οποία αρχικά εισήγαγε ο Ulf Grenander [78] και αργότερα ανέλυσαν, εφάρμοσαν και επέκτειναν εργασίες όπως [75, 79, 166, 218]. Όπως αναλύεται στην αναφορά [166] η θεωρία αυτή αντιμετωπίζει το πρόβλημα της κατανόησης προτύπων το οποίο αποτελεί ένα υπερσύνολο της αναγνώρισης προτύπων· στόχος είναι να εκμαθευτούν, αναλυθούν, και αναπαραχθούν τα πρότυπα που εμφανίζονται στον φυσικό κόσμο σε όλη τους την ποικιλία και πολυπλοκότητα κατά τέτοιο τρόπο που να επιτρέπει την εξαγωγή των φυσικών αιτίων που τα προκάλεσαν.

Χρησιμοποιώντας το πιθανοτικό πλαίσιο το οποίο ακολουθείται σε αυτήν τη διατριβή για τη διατύπωση του προβλήματος, θεωρούμε ότι το μοντέλο $\mathcal{M}(k)$ του κάθε αντικειμένου k χρησιμοποιεί ένα σύνολο παραμέτρων $\mathcal{A}_{\mathcal{M}(k)}$ για να συνθέσει το υποσύνολο των παρατηρήσεων της σκηνής που του αντιστοιχούν· για τη σύνθεση μίας σκηνής περισσότερα του ενός αντικείμενα συνδυάζονται, ενώ εισάγονται επιπλέον τυχαίες μεταβλητές που περιγράφουν την σκηνή, όπως ο φωτισμός, το βάθος στο οποίο βρίσκονται τα αντικείμενα, η θέση της κάμερας, κ.ο.κ. Συνδυάζοντας τα παραπάνω με ένα μοντέλο του σφάλματος έχουμε μία κατανομή πιθανότητας πάνω στις παρατηρήσεις I μίας εικόνας $P(I|\mathcal{A})$, όπου το σύνολο παραμέτρων $\mathcal{A} = \{\mathcal{M}, \mathcal{A}_{\mathcal{M}(1)}, \dots, \mathcal{A}_{\mathcal{M}(K)}, \mathcal{A}_{sc}\}$ συνοψίζει την ερμηνεία της σκηνής μέσω του συνόλου μοντέλων \mathcal{M} , των παραμέτρων των επιμέρους μοντέλων και των παραμέτρων της σκηνής \mathcal{A}_{sc} .

Εισάγοντας μέσω μίας πρότερης κατανομής (prior distribution) πάνω στις παραμέτρους $P(\mathcal{A})$ τη γνώση μας σχετικά με τους τρόπους με τους οποίους παραμορφώνονται και συνδυάζονται τα αντικείμενα, η ερμηνεία της σκηνής ανάγεται στην εφαρμογή του τύπου του Bayes

$$P(\mathcal{A}|I) = \frac{P(I|\mathcal{A})P(\mathcal{A})}{P(I)} \propto P(I|\mathcal{A})P(\mathcal{A}) \quad (1.1)$$

Συνοπτικά, η αναγνώριση προτύπων αναζητά να προσεγγίσει την ποσότητα $P(\mathcal{A}|I)$ ενώ η κατανόηση προτύπων επιδιώκει την κατασκευή και χρήση των όρων $P(I|\mathcal{A})$, $P(\mathcal{A})$ βάσει των οποίων υπολογίζεται και ο όρος $P(\mathcal{A}|I)$.

Με βάση αυτή την προσέγγιση μπορούμε να διαχωρίσουμε την κατασκευή μον-

τέλων από την αντιμετώπιση συγκεκριμένων λειτουργιών της όρασης· έτσι για παράδειγμα το ίδιο μοντέλο για ένα πρόσωπο μπορεί να χρησιμοποιηθεί για την παρακολούθηση της κίνησής του, την ανίχνευση και την κατάτμησή του· εκτιμώντας τις παραμέτρους του μοντέλου επιλύονται ταυτόχρονα όλα τα προβλήματα, ενώ βελτιώνοντας το μοντέλο του προσώπου βελτιώνεται η επίδοση του συστήματος στα επιμέρους προβλήματα.

Επίσης, χρησιμοποιώντας μοντέλα για αντικείμενα, τα προβλήματα της όρασης χαμηλού επιπέδου καθίστανται καλώς ορισμένα, καθώς ο χώρος λύσεών τους περιορίζεται. Η διαφορά με τη συνηθισμένη μαθηματική προσέγγιση της κανονικοποίησης (regularization) είναι ότι οι περιορισμοί πλέον δεν βασίζονται στη μαθηματική διαίσθηση, αλλά μπορούν να εκμαθευτούν από το περιβάλλον.

Έχοντας θέσει το πλαίσιο αυτό, τα κύρια προβλήματα είναι η κατασκευή του μοντέλου σύνθεσης $P(I|\mathcal{A})$ και της πρότερης κατανομής $P(\mathcal{A})$ (στάδιο εκμάθησης), και η εκτίμηση των παραμέτρων του μοντέλου δεδομένης μίας εικόνας (στάδιο ανάλυσης). Τα μαθηματικά εργαλεία για την εφαρμογή της προσέγγισης αυτής αποτελούν τα αναγεννητικά και γραφικά μοντέλα, στα οποία βασίστηκε και η έρευνά που έγινε στα πλαίσια αυτής της διατριβής.

1.3 Συνεισφορές και Δομή Διδακτορικής Διατριβής

Στόχος αυτής της διατριβής είναι να εξεταστεί η δυνατότητα διατύπωσης και εκμετάλλευσης της συνέργειας μεταξύ των διεργασιών της αναγνώρισης αντικειμένων και κατάτμησης εικόνων. Τα δύο αυτά προβλήματα είναι ευρύτατα, και συνδυάζουν πτυχές σχεδόν από όλο το φάσμα της όρασης υπολογιστών, όπως ανίχνευση ακμών, ανάλυση υφής, αναγνώριση προτύπων, εκμάθηση μοντέλων και ανίχνευση αντικειμένων. Κατά την εξέλιξη και ωρίμανση της έρευνας που έγινε στα πλαίσια της διδακτορικής διατριβής ασχοληθήκαμε και με τις επιμέρους συνιστώσες του προβλήματος, ώστε αποκτώντας μία αίσθηση των προβλημάτων που εμφανίζουν να βρεθεί πού και πώς θα μπορούσε να χρησιμοποιηθεί ο συνδυασμός τους. Το ενοποιητικό πλαίσιο για την έρευνα που έγινε στα πλαίσια της διατριβής αυτής αποτέλεσε η διατύπωση, εκμάθηση και χρήση πιθανοτικών μοντέλων που να ανταποκρίνονται στις γεωμετρικές ιδιαιτερότητες των οπτικών αντικειμένων.

Η δομή της διατριβής αντικατοπτρίζει τα στάδια της έρευνας που εκπονήθηκε στα πλαίσιά της: στο 2ο κεφάλαιο παρουσιάζονται σύντομα οι κύριες τεχνικές αναγνώρισης αντικειμένων και κατάτμησης εικόνων καθώς και τα βασικά στοιχεία των αναγεννητικών και γραφικών μοντέλων. Στο 3ο κεφάλαιο παρουσιάζεται η εργασία που έγινε στα πλαίσια της ανάλυσης της συνέργειας μεταξύ χαμηλού και μέσου επιπέδου λειτουργιών της όρασης, όπως ανίχνευση ακμών, ομαλοποίηση εικόνας και σύνδεση περιγραμμάτων. Το 4ο κεφάλαιο περιέχει τα αποτελέσματα της έρευνας που έγινε στην κατάτμηση εικόνων υφής, όπου ο συνδυασμός γεωμετρικών και στατιστικών μεθόδων κατέστη δυνατός στο πλαίσιο των αναγεννητικών μοντέλων. Στο 5ο κεφάλαιο εισάγεται μία στατιστική προσέγγιση στην από κοινού κατάτμηση και ανίχνευση αντικειμένων, βάσει του αλγορίθμου Expectation Maximization, ενώ στο 6ο κεφάλαιο προτείνεται μία μέθοδος που συνδέει αναγεννητικά και γραφικά μοντέλα για τον συνδυασμό άνωθεν και κάτωθεν πληροφορίας για την ανίχνευση αντικειμένων. Οι ερευνητικές συνεισφορές που έγιναν στα πλαίσια της διατριβής αυτής εντοπίζονται στις παρακάτω περιοχές:

Κεφάλαιο 2: Μη-γραμμική προσέγγιση συναρτήσεων με τεχνικές εκμάθησης μηχανών

Ξεκινώντας από ένα πρωταρχικό πρόβλημα της αναγνώρισης προτύπων, την προσέγγιση συναρτήσεων βάσει ενός συνόλου δεδομένων εκπαίδευσης, υλοποιήθηκαν, συγκρίθηκαν και εφαρμόστηκαν τεχνικές από το χώρο της μάθησης μηχανών, όπως δίκτυα τύπου Radial Basis Function, αρχιτεκτονικές Takagi-Sugeno-Kang/Mixture-of-Experts και Support Vector Machines.

Κεφάλαιο 3: Εκμάθηση μοντέλων για λειτουργίες όρασης χαμηλού και μέσου επιπέδου.

Έχοντας ως στόχο την κατασκευή ενός μοντέλου για λειτουργίες όρασης χαμηλού και μέσου επιπέδου (ανίχνευση ακμών, ομαλοποίηση και κατάτμηση εικόνων) που να βασίζεται στη λειτουργία του βιολογικού συστήματος όρασης, ασχοληθήκαμε με την μεταβολική και πιθανοτική του διατύπωση και την εκμάθηση των παραμέτρων του. Χρησιμοποιώντας την μεταβολική προσέγγιση η λειτουργία του μοντέλου ερμηνεύτηκε ως υλοποίηση μίας συγκεκριμένη μορφής συμπερασματολογίας (Mean Field Approximation) πάνω σε ένα γραφικό δίκτυο. Βάσει αυτής της θεώρησης, η εκμάθηση του μοντέλου επιτεύχθηκε μέσω τεχνικών εκμάθησης παραμέτρων γραφικών δικτύων, δίνοντας συστηματικά καλύτερα αποτελέσματα ανίχνευσης ακμών.

Κεφάλαιο 4: Ανάλυση, κατάτμηση και ανίχνευση υφής με χαρακτηριστικά διαμόρφωσης, αναγεννητικά μοντέλα και σταθμισμένη εξέλιξη καμπυλών

Για το πρόβλημα της κατάτμησης εικόνων αρχικά προτάθηκε ο συνδυασμός χαρακτηριστικών υφής τύπου AM-FM με τεχνικές εξέλιξης καμπυλών, που έδωσε ικανοποιητικά αποτελέσματα στην κατάτμηση φυσικών εικόνων που περιείχαν έντονη υφή. Ακολουθώντας μία πιθανοτική προσέγγιση, ερμηνεύτηκε η διαδικασία εξαγωγής χαρακτηριστικών μέσω ενός αναγεννητικού μοντέλου, καθιστώντας έτσι δυνατή την εισαγωγή εναλλακτικών μοντέλων για ακμές και ομαλές περιοχές. Μία πρακτική εφαρμογή της ιδέας αυτής είναι ο συνδυασμός διαφορετικών πηγών πληροφορίας στην διαδικασία της κατάτμησης, μέσω της εισαγωγής των ύστερων πιθανοτήτων τους στον αλγόριθμο εξέλιξης καμπυλών. Ξεκινώντας από μία πιθανοτική μέθοδο σύντηξης πληροφορίας αυτό επιτυγχάνεται μέσω ενός νέου σχήματος, που ονομάζουμε σταθμισμένη εξέλιξη καμπυλών. Έτσι αντιμετωπίζεται το πρακτικό πρόβλημα της διάκρισης περιοχών υφής και ακμών απαλείφοντας την επίδραση των λανθασμένων χαρακτηριστικών υφής πάνω σε ακμές/ομαλές περιοχές και των έντονων ακμών σε περιοχές υφής.

Κεφάλαιο 5: Συνέργεια αναγνώρισης αντικειμένων και κατάτμησης εικόνων μέσω αναγεννητικών μοντέλων και του αλγορίθμου Expectation Maximization. (EM)

Βάσει της προσέγγισης των αναγεννητικών μοντέλων, προτάθηκε η εφαρμογή του αλγορίθμου EM στο πρόβλημα της σύζευξης. Αντιμετωπίζοντας το πρόβλημα της κατάτμησης εικόνων ως την ανάθεση περιοχών της εικόνας σε μοντέλα για αντικείμενα και την αναγνώριση/ανάλυση αντικειμένων ως την εκτίμηση των παραμέτρων ενός μοντέλου, το πρόβλημα ανάγεται στην εκτίμηση παραμέτρων με ελλειπή δεδομένα,

καθώς η κατάτμηση είναι άγνωστη. Ο αλγόριθμος EM αποτελεί μία καθιερωμένη προσέγγιση για προβλήματα με ελλιπή δεδομένα που εγγυημένα καταλήγει σε μία (τοπικά) βέλτιστη λύση ενώ η ταχύτητά του καθιστά δυνατή την πρακτική εκμετάλλευση της συνέργειας. Η σαφήνεια της διατύπωσης μέσω του αλγορίθμου EM έχει επιτρέψει τον συνδυασμό διαφορετικών τεχνικών της όρασης υπολογιστών, όπως μοντέλα ενεργούς εμφάνισης και εξέλιξη καμπυλών, επιτρέποντας επίσης την εισαγωγή πληροφορίας κατάτμησης στις σχέσεις προσαρμογής των αναγεννητικών μοντέλων. Συστηματικά καλύτερα αποτελέσματα σε ανίχνευση και κατάτμηση αντικειμένων έχουν επιτευχθεί μέσω της συνέργειας των δύο προβλημάτων, ενώ επεκτείνοντας την εφαρμοσιμότητα της μεθόδου προτάθηκε μία τεχνική για την αυτοματοποιημένη κατασκευή αναγεννητικών μοντέλων.

Κεφάλαιο 6: Συνδυασμός άνωθεν και κάτωθεν πληροφορίας για ανίχνευση αντικειμένων μέσω γραφικών μοντέλων και πρωταρχικών χαρακτηριστικών εικόνας

Η προσέγγιση της εξαγωγής σημείων ενδιαφέροντος για ανίχνευση αντικειμένων μέσω τμηματικών αναπαραστάσεων αντιμετωπίστηκε στο πλαίσιο των αναγεννητικών γραφικών μοντέλων, εστιάζοντας στον συνδυασμό της άνωθεν πληροφορίας που παρέχει ένα αναγεννητικό μοντέλο με την κάτωθεν πληροφορία που προέρχεται από την εικόνα. Χαρακτηριστικά τύπου ράβδων, ακμών και κηλίδων χρησιμοποιήθηκαν για την εκμάθηση απλών γραφικών μοντέλων τύπου τα οποία εφαρμόζονται στην κάτωθεν ανίχνευση αντικειμένων. Η εισαγωγή άνωθεν πληροφορίας με τεχνικές γραφικών δικτύων κατέστη δυνατή μέσω μίας ταχείας μεθόδου για εκτίμηση της πιθανοφάνειας υποθέσεων εξειδικευμένη σε προβλήματα που περιλαμβάνουν εικόνες. Πάνω σε γνωστά σύνολα δεδομένων αυτή η προσέγγιση έχει δώσει αποτελέσματα ισάξια με τα καλύτερα των τελευταίων ετών, χρησιμοποιώντας σημαντικά απλούστερες αναπαραστάσεις για αντικείμενα.

Κεφάλαιο 2

Θεωρητικό Υπόβαθρο και Επισκόπηση Ερευνητικού Πεδίου

2.1 Κατάτμηση Εικόνων

Το πρόβλημα της κατάτμησης εικόνων μπορεί να οριστεί σε ένα γενικό επίπεδο ως η αναζήτηση ενός διαχωρισμού της εικόνας σε ομοιογενείς περιοχές, κατά έναν τρόπο που να αντιστοιχεί στην ανθρώπινη αντίληψη. Η δυσκολία του προβλήματος βρίσκεται τόσο σε υψηλό επίπεδο, καθώς διαφορετικοί άνθρωποι μπορούν να κατατμήσουν την ίδια σκηνή με διαφορετικούς τρόπους, όσο και σε χαμηλότερο, καθώς τα κριτήρια με τα οποία ορίζεται η ομοιογένεια των περιοχών ποικίλουν από εικόνα σε εικόνα, ενώ είναι συχνά δύσκολη η διατύπωση και ποσοτικοποίησή τους (π.χ. Σχ. 2.1).

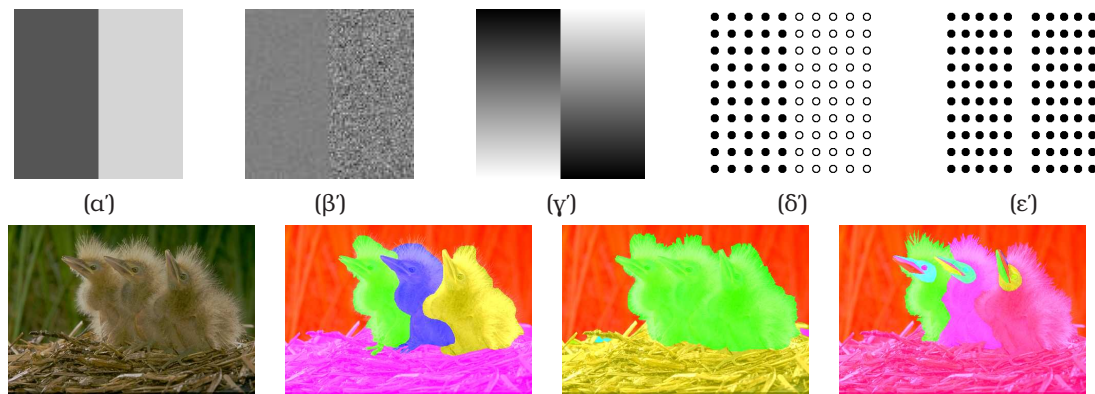
Η συνηθισμένη προσέγγιση θεωρεί ότι τα τμήματα της εικόνας μπορούν να ομαδοποιηθούν βάσει κάποιου χαρακτηριστικού χαμηλού επιπέδου, όπως η φωτεινότητα της εικόνας, το χρώμα, ή η έξοδος μίας συστοιχίας φίλτρων. Τα κριτήρια που χρησιμοποιούνται στη διατύπωση των αλγορίθμων κατάτμησης βασίζονται τότε στην ομοιογένειά των χαρακτηριστικών στο εσωτερικό των περιοχών ή/και την ασυνέχειά τους εκατέρωθεν των ορίων τους, ενώ οι επιθυμητές λύσεις μπορούν να περιοριστούν εισάγοντας επιπλέον όρους που αφορούν το μήκος των διαχωριστικών γραμμών και την έκταση των περιοχών.

2.1.1 Κριτήρια και μέθοδοι κατάτμησης εικόνων

Τα κριτήρια που χρησιμοποιούνται για την κατάτμηση εικόνων μπορούν να ομαδοποιηθούν στις παρακάτω κύριες κατηγορίες:

Άνωθεν κριτήρια

Ως ποιοτική κατάτμηση θεωρείται αυτή όπου μέσα σε κάθε τμήμα της εικόνας μπορεί να κατασκευαστεί ένα απλοποιημένο μοντέλο των παρατηρήσεων που να τα μοντελοποιεί ικανοποιητικά. Χαρακτηριστικά παραδείγματα είναι το τμηματικά ομαλό μοντέλο της εικόνας [168, 15, 217], η τμηματικά σταθερή προσέγγισή του [248, 32] και γενικότερα τα αναγεννητικά μοντέλα για κατάτμηση [219, 218]. Η χρήση ενός μοντέλου για την εικόνα είναι συγχρόνως πρόβλημα και προτέρημα, καθώς αφενός είναι δύσκολο να κατασκευαστεί ένα μοντέλο που να συνδυάζει εκφραστικότητα



Σχήμα 2.1: Χαμηλού και Υψηλού επιπέδου προβλήματα της κατάτμησης: Πάνω γραμμή: Τα κριτήρια με τα οποία αποκαλούμε τις περιοχές της εικόνας ομοιογενείς δεν είναι κοινά για όλες τις εικόνες, ούτε μπορούν να ποσοτικοποιηθούν εύκολα. Μόνο στην εικόνα (α) είναι χρήσιμη η τιμή της φωτεινότητας ως χαρακτηριστικό για κατάτμηση· για τις υπόλοιπες το χρησιμοποιούμενο χαρακτηριστικό είναι η διασπορά του θορύβου (β) η σκίαση (παράγωγος της φωτεινότητας) (γ) η ομοιότητα των κηλίδων (δ) και η προσήνειά τους (ε). Κάτω γραμμή: Αναλόγως της κλίμακας και του κύριου ενδιαφέροντος του θεατή αλλάζει και η κατάτμηση μίας σκηνής.

και απλότητα, αφετέρου ένα άνωθεν κριτήριο επιτρέπει την κατανόηση των αποτελεσμάτων της κατάτμησης μέσω της ανασύνθεσης των παρατηρήσεων και την άμεση προσέγγιση του προβλήματος της συνέργειας.

Κριτήρια δυαδικής ομαδοποίησης (pairwise clustering)

Το κριτήριο επιζητεί την ομοιότητα των χαρακτηριστικών στο εσωτερικό των περιοχών και τη διαφορετικότητά τους σε σχέση με αυτά των υπόλοιπων. Παραδείγματα είναι το κριτήριο κανονικοποιημένων τομών [202, 238, 150] και οι τεχνικές παραγοντοποίησης πινάκων συνοχής [185]. Η διαφορά με την παραπάνω μέθοδο είναι ότι δεν χρειάζεται η χρήση κάποιου υψηλότερου επιπέδου μοντέλου ώστε να διατυπωθεί το επιζητούμενο κριτήριο· αφενός έτσι απλοποιείται το πρόβλημα, αφετέρου η όποια λύση είναι δύσκολο να ενσωματωθεί σε μία συνθετότερη προσέγγιση, που να περιλαμβάνει αντικείμενα, ή να συνεργάζεται με συστήματα που βασίζονται σε άλλες πηγές πληροφορίας, όπως κίνηση και βάθος.

Κριτήρια ομαδοποίησης δεδομένων

Τα χαρακτηριστικά της εικόνας θεωρούνται ως απλά δεδομένα, και εφαρμόζονται κοινός αλγόριθμοι ομαδοποίησης, ενώ η όποια γεωμετρική πληροφορία εισάγεται θεωρώντας τις συντεταγμένες των σημείων της εικόνας επίσης ως χαρακτηριστικά [7, 36]. Θα μπορούσαμε να θεωρήσουμε τα κριτήρια αυτά ως ειδικές περιπτώσεις των παραπάνω δύο, όταν αφαιρείται η όποια πρότερη πληροφορία σχετικά με την κατάτμηση. Τέτοια κριτήρια συχνά χρησιμοποιούνται για προβλήματα όπου ζητείται να βρεθεί γρήγορα μία συνοπτική περιγραφή των δεδομένων της εικόνας, όπως για παράδειγμα στην ανάκληση εικόνων βάσει περιεχομένου· είναι όμως περιορισμένων δυνατοτήτων, καθώς δεν λαμβάνουν υπόψη την ιδιαιτερότητα του προβλήματος της όρασης.

Κριτήρια συνεκτικότητας εικόνας

Στην περίπτωση αυτή, ξεκινώντας από την ανάλυση της εικόνας σε συνεκτικά τμήματα βάσει της γειννίασης των παρατηρήσεων και της ομοιότητας των χαρακτηριστικών ή της απουσίας ακμών, ζητείται να βρεθεί μία διαμέριση της εικόνας που να προκύπτει από την συσσώρευση τμημάτων. Τέτοιοι αλγόριθμοι, όπως για παράδειγμα ο αλγόριθμος watershed [10] έχουν προκύψει από την περιοχή της μαθηματικής μορφολογίας, ενώ η γενικότερη ιδέα απαντά και σε τεχνικές τύπου region growing και θεωρίας γράφων [59]. Αν και μαθηματικά σαφείς και αποδοτικοί αλγόριθμοι μπορούν να κατασκευαστούν βάσει αυτής της προσέγγισης, κρίσιμα θέματα όπως π.χ. η επιλογή του κριτηρίου συνεκτικότητας ή η μέθοδος συσσώρευσης τμημάτων δεν έχουν πάντα κάποια προφανή λύση.

Κριτήρια ακμών/περιγράμματος

Η προσέγγιση αυτή ανάγει την κατάτμηση στην ομαδοποίηση της εξόδου ενός συστήματος ανίχνευσης ακμών με κλειστά περιγράμματα. Έτσι αφενός παρακάμπτεται το πρόβλημα της μοντελοποίησης των χαρακτηριστικών, χρησιμοποιώντας τα αποτελέσματα ενός αλγορίθμου ακμών ως μία σύνοψή τους, αφετέρου εισάγονται νέα προβλήματα όπως η ανίχνευση των ακμών και η ομαδοποίησή τους. Ικανοποιητικά αποτελέσματα βάσει αυτού της προσέγγισης μπορούν να επιτευχθούν σε σχετικά απλές εικόνες [118, 31]· για το γενικότερο πρόβλημα της κατάτμησης συνήθως συνδυάζονται με αλγορίθμους που χρησιμοποιούν κριτήρια που εκφράζονται βάσει των χαρακτηριστικών των περιοχών [150, 178].

Έχοντας διατυπώσει ένα κριτήριο για την καταλληλότητα μίας κατάτμησης, απομένει η βελτιστοποίησή του· για το στόχο αυτό η κοινότητα της όρασης υπολογιστών αντλεί μαθηματικές μεθόδους από τις παρακάτω κυρίως περιοχές:

Μέθοδοι Λογισμού Μεταβολών

Εκφράζοντας την κατάτμηση ως μία συνάρτηση, το κριτήριο που εκφράζει την ποιότητά της μπορεί να γραφεί ως ένα συναρτησιακό που ποσοτικοποιεί τις ιδιότητες της κατάτμησης, όπως το μήκος των ορίων, τη σύμπτωσή τους με μεταβολές της φωτεινότητας και την ομοιογένεια των χαρακτηριστικών στο εσωτερικό των περιοχών. Χρησιμοποιώντας τεχνικές του Λογισμού Μεταβολών, ξεκινώντας από μία αρχική κατάσταση η συνάρτηση εξελίσσεται προς την κατεύθυνση η οποία μειώνει ταχύτερα το χρησιμοποιούμενο συναρτησιακό. Από την συνθήκη ελαχιστοποίησης του συναρτησιακού προκύπτουν αλγόριθμοι Γεωμετρικών Μερικών Διαφορικών Εξισώσεων ανάμεσα στους οποίους κατατάσσονται οι τεχνικές Εξέλιξης Καμπυλών [31, 248] και Ανισοτροπικής Διάχυσης [187, 213]. Συνηθέστερα τέτοιες τεχνικές εφαρμόζονται σε συνδυασμό με τα τοπικά κριτήρια, αλλά μπορούν να επεκταθούν και σε κριτήρια δυαδικής ομαδοποίησης [211].

Μέθοδοι στατιστικής συμπερασματολογίας

Θεωρώντας την κατάτμηση S μίας εικόνας ως την κατάσταση ενός φυσικού συστήματος, το κριτήριο που ελαχιστοποιείται μπορεί να ερμηνευθεί ως η ενέργειά του

$E(S, O)$: οι παρατηρήσεις της εικόνας, O εισάγουν εξωτερικούς περιορισμούς στο σύστημα ενώ κριτήρια όπως το μήκος ακμών λειτουργούν ως εσωτερικοί όροι ενέργειας. Μπορεί τότε να χρησιμοποιηθεί ο τύπος του Boltzmann για να ορίσει μία κατανομή πιθανότητας πάνω στις κατατμήσεις

$$P(S|O) = \frac{\exp(-E(S, O))}{Z} \quad Z = \sum_S \exp(-E(S, O)) \quad (2.1)$$

που θεωρεί πιθανότερες καταστάσεις με μικρότερη ενέργεια. Το πρόβλημα της κατάτμησης ανάγεται έτσι στην εκτίμηση της ύστερης (posterior) κατανομής πιθανότητας στον χώρο των κατατμήσεων. Μία αμεσότερη οπτική γωνία είναι αυτή των αναγεννητικών μοντέλων, όπου ζητείται η ύστερη κατανομή της κατάτμησης της εικόνας χρησιμοποιώντας το Bayesian πλαίσιο, οπότε έχουμε $P(S|O) = \frac{\exp(-\log P(O|S) - \log(P(S)))}{P(O)}$

Για την εκτίμηση της ύστερης κατανομής χρησιμοποιούνται τόσο στοχαστικές τεχνικές δειγματοληψίας Markov Chain Monte Carlo για Τυχαία Μαρκοβιανά Πεδία (Markov Random Fields-MRFs)[75, 219] καθώς και μεταβολικές τεχνικές [245], που παρέχουν μία σημειακή εκτιμήτρια αντί της ύστερης κατανομής. Για συγκεκριμένες επιλογές του πιθανοτικού μοντέλου και του αλγόριθμου υπολογισμού της ύστερης κατανομής μπορούμε να δούμε τα μεταβολικά μοντέλα στο πλαίσιο της στατιστικής συμπερασματολογίας, θεωρώντας ότι ξεκινώντας από μία αρχική κατάσταση με έναν άπληστο (greedy) τρόπο αναζητούν μία κορυφή της συνάρτησης ύστερης πιθανότητας.

Μέθοδοι θεωρίας γράφων

Θεωρώντας κάθε pixel της εικόνας ως έναν κόμβο ενός γράφου και την ομοιότητα μεταξύ δύο pixels ως τα βάρη των ακμών του γράφου, το πρόβλημα της κατάτμησης ανάγεται στην εύρεση των τομών που ελαχιστοποιούν μία συνάρτηση κόστους πάνω στον γράφο. Αναλόγως του κριτηρίου που εκφράζει το κόστος της τομής έχουμε συναρτησιακά όμοια με αυτά των στατιστικών/μεταβολικών μεθόδων, που ελαχιστοποιούνται χρησιμοποιώντας αλγόριθμους Επιστήμης των Υπολογιστών [24, 23, 59] και καθολικά συναρτησιακά που ελαχιστοποιούνται με μεθόδους της Φασματικής Θεωρίας Γράφων (Spectral Graph Theory) [202, 238].

Μία σημαντική πτυχή είναι η αποδοτικότητα των τεχνικών με τις οποίες γίνεται η βελτιστοποίηση των κριτηρίων· μία απλή ιδέα είναι να γίνει αρχικά η κατάτμηση σε μία μικρότερη ανάλυση και αργότερα να χρησιμοποιηθούν τα αποτελέσματα για να αρχικοποιηθεί η κατάτμηση στην πλήρη ανάλυση, εξοικονομώντας χρόνο και εξασφαλίζοντας ευρωστία. Η ιδέα αυτή μπορεί να διατυπωθεί και να επεκταθεί στο πλαίσιο των πολυπλεγματικών (multigrid) μεθόδων, τόσο για μεταβολικές [176] όσο και για τις στατιστικές [5] και γραφικές μεθόδους [74].

Σημειώνουμε εδώ ότι τα σύνορα μεταξύ των μεθόδων και κριτηρίων αυτών δεν είναι αυστηρά· για παράδειγμα μεταβολικές μέθοδοι εξέλιξης καμπυλών έχουν εφαρμοστεί για την εύρεση τοπικών βέλτιστων στα πλαίσια της εύρεσης της Maximum-a-Posteriori εκτίμησης της κατάτμησης [248, 219], αλγόριθμοι τύπου δειγματοληψίας/θεωρίας γράφων για την ελαχιστοποίηση συναρτησιακών που διατυπώθηκαν στα

πλαίσια των μεταβολικών μεθόδων, και αλγόριθμοι εξέλιξης καμπυλών ως προσεγγιστικές λύσεις του προβλήματος των κανονικοποιημένων τομών [201].

Στα πλαίσια της έρευνάς μας ασχοληθήκαμε με μεταβολικές μεθόδους καθώς συνδυάζουν κομψότητα, σαφήνεια και αποδοτικότητα. Επίσης ένα σημαντικό τους χαρακτηριστικό είναι ότι είναι δυναμικές, οπότε μπορούν να χρησιμοποιηθούν παράλληλα με άλλες διαδικασίες όπως η αναγνώριση αντικειμένων, σε αντίθεση με άλλες τεχνικές [162, 238] όπου αυτή η σύζευξη είναι σχεδόν τεχνητή.

2.1.2 Μεταβολική Κατάτμηση Εικόνων

Σύμφωνα με τις τεχνικές κατάτμησης που βασίζονται στη θεωρία του Λογισμού Μεταβολών [230, 42] η επίλυση του προβλήματος της κατάτμησης ανάγεται στην ελαχιστοποίηση ενός συναρτησιακού που ποσοτικοποιεί την ποιότητά της. Βάσει της μεταβολικής προσέγγισης η ανάλυση των προβλημάτων γίνεται στο συνεχές πεδίο βοηθώντας την διαίσθηση και την σαφήνεια ενώ τα προβλήματα διακριτοποίησης εμφανίζονται όταν πλέον αναλαμβάνουν οι αριθμητικοί αλγόριθμοι την ελαχιστοποίηση του κριτηρίου.

Το γενικότερο πλαίσιο των μεταβολικών μεθόδων είναι το εξής:

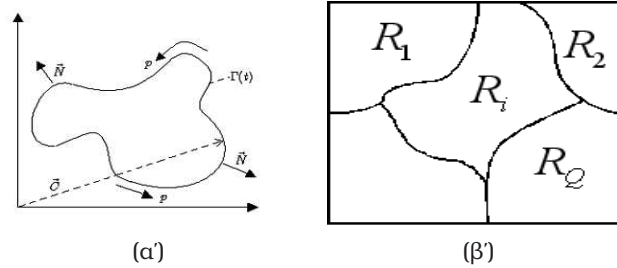
- Θεωρούμε την κατάτμηση της εικόνας ως μία συνάρτηση. Η έκφρασή της μπορεί να είναι είτε μέσω του labelling της εικόνας είτε μέσω των συνόρων των περιοχών.
- Σχηματίζουμε μία έκφραση που ποσοτικοποιεί την ποιότητα μίας κατάτμησης. Καθώς η έκφραση ορίζεται πάνω σε συναρτήσεις (κατατμήσεις) είναι ένα συναρτησιακό.
- Βρίσκουμε τη συνθήκη για ένα ελάχιστο της έκφρασης, χρησιμοποιώντας τη θεωρία του Λογισμού Μεταβολών.
- Εξελίσσουμε την κατάτμηση προς την κατεύθυνση που θα ικανοποιήσει τη συνθήκη ελάχιστου.

Από τα πλέον γνωστά συναρτησιακά είναι αυτό που εισήγαγαν οι Mumford και Shah [168], το οποίο ουσιαστικά ξεκίνησε την έρευνα στο πεδίο των μεταβολικών τεχνικών [161]:

$$E(U, \Gamma) = \iint_R (I - U)^2 dx dy + \alpha \iint_{R-\Gamma} \|\nabla U\|^2 dx dy + \beta |\Gamma| \quad (2.2)$$

Οι όροι που εμπλέκονται στο συναρτησιακό αυτό είναι: η παρατηρούμενη εικόνα, I , μια τμηματικά ομαλή προσέγγιση της εικόνας U , τα όρια των περιοχών, Γ και το πεδίο ορισμού της συνάρτησης, R . Σημειώνουμε ότι μπορεί τα σύνορα Γ να μην απομονώνουν πάντα περιοχές της εικόνας, περιέχοντας ανοικτές καμπύλες. Περιοριζόμενοι στην κατάτμηση εικόνων θεωρούμε ότι η βελτιστοποίηση του συναρτησιακού (2.2) γίνεται μόνο πάνω σε διαμερίσεις της εικόνας, οπότε αν η εικόνα χωρίζεται στις περιοχές $R_i, i = 1 \dots M$ (Σχ. 2.2β), έχουμε $\Gamma_i = \partial R_i$ $\Gamma = \bigcup_{i=1}^M \Gamma_i$.

Η ποιότητα μίας κατάτμησης θεωρείται ότι συνδέεται αντίστροφα με την ποσότητα $E(U, \Gamma)$ (2.2), καθώς οι όροι που αθροίζονται εκφράζουν ανεπιθύμητες ποσότητες. Συγκεκριμένα, έχουμε:



Σχήμα 2.2: (α): Γεωμετρικές ποσότητες που εμπλέκονται στην εξέλιξη μίας καμπύλης (β) Διαμερισμός μίας εικόνας.

- $\iint_R (I - U)^2 dx dy$: ένα διπλό ολοκλήρωμα πάνω σε όλο το πεδίο της εικόνας που τιμωρεί την απόσταση της ανακατασκευής U από την εικόνα I .
- $\iint_{R-\Gamma} \|\nabla U\|^2 dx dy$: ένα διπλό ολοκλήρωμα σε όλο το πεδίο ορισμού εκτός από τα σύνορα των περιοχών που τιμωρεί τις ανωμαλίες της ανακατασκευής. Στα σύνορα επιτρέπεται η U να έχει ασυνέχειες, οπότε προκύπτει μία τμηματικά ομαλή προσέγγιση.
- $|\Gamma| = \int_{\Gamma} ds$: ένα επικαμπύλιο ολοκλήρωμα που εκτιμά το μήκος των ορίων των περιοχών και οδηγεί σε λύσεις με ομαλά σύνορα, αποφεύγοντας περιττές καμπυλώσεις. Για την παραμετροποίηση της καμπύλης χρησιμοποιείται η εγγενής γεωμετρική ποσότητα του μήκους τόξου, s .

Καθώς η ελαχιστοποίηση του συναρτησιακού αποσυνδέεται από το σχηματισμό του, εναλλακτικές προσεγγίσεις και απλουστεύσεις έχουν οδηγήσει σε πλήθος διαφορετικών αλγόριθμων [15, 139, 32, 217].

Παρόμοια κριτήρια έχουν εμφανισθεί σε προηγούμενες και επόμενες εργασίες [75, 15, 245], με τη διαφορά ότι είναι ορισμένα στο διακριτό πεδίο και οι αλγόριθμοι που προτείνονται για την ελαχιστοποίησή τους είναι εγγενώς διακριτοί. Αντιθέτως, οι συνθήκες ελάχιστου της (2.2) οδηγούν στην επίλυση Μερικών Διαφορικών Εξισώσεων-ΜΔΕ όποτε θεωρητικές πτυχές, αριθμητικές μέθοδοι επίλυσης και η διαίσθηση που έχουν αναπτυχθεί γύρω από αυτά τα προβλήματα μπορούν να αξιοποιηθούν για τους σκοπούς της όρασης υπολογιστών.

Γνωρίζουμε [230] ότι η αναγκαία συνθήκη για την ελαχιστοποίηση ενός συναρτησιακού J που ορίζεται πάνω στην συνάρτηση $f(x, y)$ ως

$$J(f) = \int g(f, f_x, f_y) dx dy \quad (2.3)$$

δίνεται από τον μηδενισμό της παραγώγου Euler του συναρτησιακού :

$$[J]_f = \frac{\partial g}{\partial f} - \frac{\partial}{\partial x} \left(\frac{\partial g}{\partial f_x} \right) - \frac{\partial}{\partial y} \left(\frac{\partial g}{\partial f_y} \right) \quad (2.4)$$

Η ύπαρξη μίας συνοριακής συνθήκης της μορφής

$$H(f(x, y))|_{\Gamma} = 0 \quad (2.5)$$

στην αρχική διατύπωση του προβλήματος οδηγεί στην συνοριακή συνθήκη της ΜΔΕ :

$$\frac{\partial H}{\partial f_x} \mathcal{N}_x + \frac{\partial H}{\partial f_y} \mathcal{N}_y = 0 \quad (2.6)$$

Από τις σχέσεις αυτές προκύπτει ότι για σταθερό Γ η συνθήκη ελάχιστου στην οποία εμπλέκεται η συνάρτηση U εκφράζεται από το παρακάτω ελλειπτικό πρόβλημα με συνοριακές συνθήκες :

$$\Delta U = c(U - I), \quad \left. \frac{\partial U}{\partial \mathcal{N}} \right|_{\partial R_i} = 0 \quad (2.7)$$

όπου $\frac{\partial}{\partial \mathcal{N}}$ είναι η κατευθυντική παράγωγος κάθετα στο σύνορο ∂R_i της περιοχής i .

Τέλος, η ελαχιστοποίηση του συναρτησιακού $J(\Gamma) = \int_{\Gamma} ds$ δίνεται από την σχέση εξέλιξης

$$\frac{\partial \Gamma(s)}{\partial t} = \kappa(s) \mathcal{N} \quad (2.8)$$

όπου \mathcal{N} είναι το ορθογώνιο διάνυσμα (Σχ. 2.2(α)) και $\kappa(s)$ είναι η καμπυλότητα (curvature) της καμπύλης Γ : $\kappa(s) = \text{div}(\mathcal{N}(s))$.

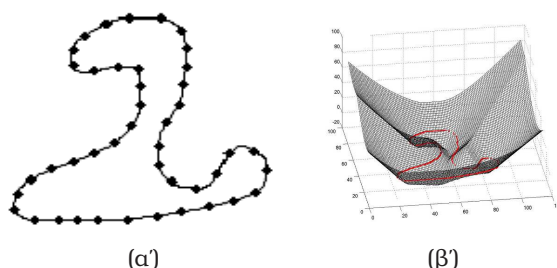
Ένα άλλο σημαντικό βήμα στην καθιέρωση των μεταβολικών μεθόδων ήταν το έργο των Kass, Witkin και Terzopoulos [118], από το οποίο πήρε ώθηση το πεδίο της Ενεργής Όρασης [14]. Η ιδέα είναι να αρχικοποιείται μία παραμετρική, παραμορφώσιμη καμπύλη (snake) που αναπαριστά το σύνορο μίας περιοχής κάπου μέσα στην εικόνα, θεωρώντας ότι βρίσκεται σχετικά κοντά στα όρια των αντικειμένων. Ο εντοπισμός των ορίων επιτυγχάνεται εξελίσσοντας την καμπύλη προς την κατεύθυνση προς στην οποία αυξάνεται η ένταση των ακμών, διατηρώντας παράλληλα τα όριά τους ομαλά. Το συναρτησιακό που χρησιμοποιείται είναι :

$$J(\Gamma) = \alpha \int_0^1 E_{int}(\Gamma(p)) dp + \beta \int_0^1 E_{img}(\Gamma(p)) dp + \gamma \int_0^1 E_{con}(\Gamma(p)) dp \quad (2.9)$$

Ο πρώτος όρος αντιστοιχεί σε γενική πρότερη γνώση για την καμπύλη σχετικά με την ομαλότητά της, ο δεύτερος όρος εξαρτάται από την εικόνα και ωθεί τα περιγράμματα σε περιοχές με μεγάλη ένταση ακμών ενώ ο τρίτος περιορίζει την καμπύλη αναλόγως της όποιας πληροφορίας εισάγει ο χρήστης.

Αυτή η εργασία αποτελεί και την πρώτη πρακτική εφαρμογή της ιδέας του συνδυασμού κάτωθεν πληροφορίας, όπως είναι η ισχύς ενός τελεστή ανίχνευσης ακμών με πρότερη γνώση που αφορά την ομαλότητα των ορίων και άνωθεν περιορισμούς που εισάγει ένα εξωτερικό σύστημα, όπως ένας χρήστης, ή στην περίπτωση που μας ενδιαφέρει ένα σύστημα αναγνώρισης. Ως επεκτάσεις του μοντέλου αυτού μπορούν να θεωρηθούν πλήθος τεχνικών της Όρασης Υπολογιστών, όπως τα Παραμορφώσιμα Περιγράμματα (Deformable Contours) [208, 14, 13] οι Ενεργές Καμπύλες (Active Contours) [248, 31, 32, 178] οι Παραμορφώσιμες Τεμπλέτες (Deformable Templates) [247, 90] και τα μοντέλα ενεργού σχήματος/εμφάνισης [38, 37, 226, 112].

Τα παραπάνω μοντέλα ξεκινούν από μία αρχική εκτίμηση των ορίων των αντικειμένων/περιοχών και μέσω μίας διαδικασίας βελτιστοποίησης τα προσαρμόζουν στα δεδομένα της εικόνας. Περιορίζοντας τον χώρο αναζήτησης σε μία περιοχή γύρω από την τρέχουσα εκτίμηση της θέσης των περιγραμμάτων, αυτή η ` ενεργή ` μέθοδος αντιπαρέρχεται προβλήματα της κλασσικής προσέγγισης που θα απαιτούσε αρχικά



Σχήμα 2.3: Παραμετρική και έμμεση αναπαράσταση μίας καμπύλης.

την επίλυση του προβλήματος της ανίχνευσης ακμών στην πλήρη έκταση του, και κατόπιν την εκτέλεση πολυπλοκότερων διαδικασιών.

Από τη δεκαετία του 80 οπότε και πρωτοεμφανίστηκαν οι τεχνικές τύπου snakes έχουν εξελιχθεί τόσο τα μαθηματικά μοντέλα όσο και τα κριτήρια με τα οποία αντιμετωπίζεται το πρόβλημα της κατάτμησης στο πλαίσιο των μεταβολικών μεθόδων. Κύρια σημεία της ωρίμανσης αυτής είναι η εισαγωγή των μεθόδων επιπεδοσυνόλων (level-set methods) [174, 200, 180] για την αριθμητική επίλυση των Μερικών Διαφορικών Εξισώσεων που προκύπτουν και τα κριτήρια των Geodesic Active Contours[31] και Region Competition [248] που παρουσιάζονται στο επόμενο εδάφιο.

2.1.3 Τεχνικές Επιπεδοσυνόλων

Ένα κύριο ζήτημα στις μεθόδους ενεργών περιγραμμάτων είναι η παραμετροποίηση της καμπύλης: αν και μαθηματικά το μήκος τόξου δίνει την προφανή παραμετροποίηση, υπολογιστικά δεν είναι πρακτικό να υπολογίζεται από την καμπύλη για να χρησιμοποιείται για την εξέλιξή της. Συνεπώς χρειάζεται η εισαγωγή κατάλληλων αναπαραστάσεων που να επιτρέπουν τον χειρισμό ποσοτήτων όπως η καμπυλότητα και το ορθογώνιο διάνυσμα της καμπύλης, καθώς και να προσφέρονται για την ενσωμάτωσή τους σε αλγορίθμους βελτιστοποίησης. Μία λύση για παράδειγμα είναι η χρήση splines για την αναπαράσταση της καμπύλης [13], όπου θεωρούμε ότι κάποια σημεία αναφοράς σε συνδυασμό με μία συνθήκη ομαλότητας αρκούν για να ορίσουν την καμπύλη. Στα πλεονεκτήματα των μεθόδων αυτών προσμετράται η υπολογιστική αποδοτικότητα, που συνεπάγεται και την δυνατότητα κατασκευής βιώσιμων μοντέλων αυξημένης πολυπλοκότητας [13]. Πέρα όμως από προβλήματα σχετικά με την αυτόματη επιλογή των σημείων, εγγενή προβλήματα αυτής της μεθόδου είναι η αδυναμία χειρισμού σημείων όπου δεν ορίζεται η καμπυλότητα (π.χ. γωνίες) καθώς και η τοπολογική τους δυσκαμψία: ένα περίγραμμα αντιστοιχεί πάντα σε μία περιοχή και δεν μπορεί να σπάσει σε δύο, ακόμα και αν υπάρχει κενό ανάμεσα σε δύο αντικείμενα.

Τα προβλήματα αυτά μπορούν να αντιμετωπιστούν χρησιμοποιώντας έμμεσες (implicit) μεθόδους για την αναπαράσταση καμπυλών και συγκεκριμένα τεχνικές επιπεδοσυνόλων (level-set techniques) [174, 200, 180]. Οι τεχνικές αυτές αντιμετωπίζουν τα προαναφερθέντα προβλήματα χωρίς να καταφεύγουν σε αριθμητικά τεχνάσματα τα οποία χρησιμοποιούσαν άλλες μέθοδοι, ενώ μπορούν εύκολα να γενικευθούν σε δεδομένα υψηλότερων διαστάσεων, αντίθετα με μεθόδους που χρησιμοποιούν παραμετρικές αναπαραστάσεις της καμπύλης.

Η ιδέα των μεθόδων επιπεδοσυνόλων είναι να εκφραστεί η εξελισσόμενη καμπύλη ως το μηδενικό επιπεδοσύνολο μίας συνάρτησης μεγαλύτερης διάστασης και η εξέλιξη της καμπύλης να εκφραστεί μέσω της εξέλιξης της συνάρτησης αυτής (Σχ.

2.3). Συγκεκριμένα, χρησιμοποιείται μία Lipschitz συνάρτηση εμβύθισης Φ τέτοια ώστε ανά πάσα στιγμή το μηδενικό επιπεδοσύνολό της να μας δίνει την καμπύλη Γ :

$$\Phi(x, y, t) : \{\Phi(x, y, t) = 0\} = \Gamma(t) \quad (2.10)$$

Τα περισσότερα μεταβολικά προβλήματα που σχετίζονται με την κατάτμηση καταλήγουν στην εξέλιξη μίας καμπύλης $\Gamma(t) = (x(t), y(t))$ σύμφωνα με μία σχέση της μορφής:

$$\frac{\partial \Gamma}{\partial t} = V \mathcal{N} \quad (2.11)$$

όπου το V είναι ένα πεδίο ταχύτητας που ορίζεται πάνω στην καμπύλη, και εξαρτάται τόσο από την καμπύλη όσο και τις ιδιότητες της εικόνας που χρησιμοποιούνται στην εξέλιξη, όπως ένταση ακμών.

Από την σχέση $\Phi(\Gamma(t), t) = 0$, τον κανόνα αλυσίδας και την (2.11) προκύπτει:

$$\begin{aligned} \frac{d\Phi(\Gamma(t), t)}{dt} &= 0 \\ \frac{\partial \Phi(\Gamma(t), t)}{\partial t} + \nabla \Phi(\Gamma) \cdot \frac{\partial \Gamma}{\partial t} &= 0 \\ \frac{\partial \Phi(\Gamma(t), t)}{\partial t} &= -\nabla \Phi(\Gamma) V \cdot \mathcal{N} \end{aligned} \quad (2.12)$$

Γεωμετρικοί όροι που εμπλέκονται στην παραπάνω εξίσωση, ή στον όρο V όπως η καμπυλότητα και το κάθετο διάνυσμα στην καμπύλη προκύπτουν άμεσα από την συνάρτηση εμβύθισης Φ :

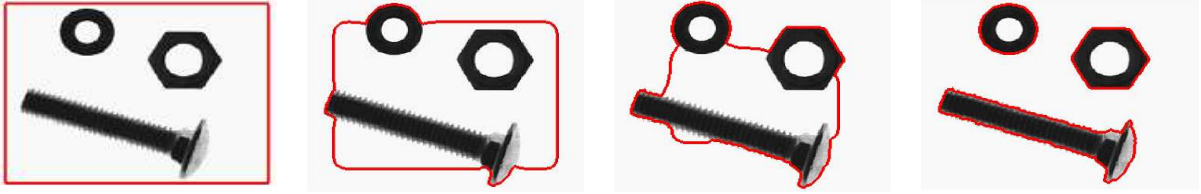
$$\kappa = \frac{\Phi_{xx}\Phi_y^2 - 2\Phi_{xy}\Phi_x\Phi_y + \Phi_{yy}\Phi_x^2}{(\Phi_x^2 + \Phi_y^2)^{\frac{3}{2}}}, \quad \mathcal{N} = \frac{\nabla \Phi}{|\nabla \Phi|} \quad (2.13)$$

όπου το κάθετο διάνυσμα \mathcal{N} δείχνει προς το εξωτερικό της καμπύλης Γ , στο εσωτερικό της οποίας η συνάρτηση Φ παίρνει αρνητικές τιμές. Χρησιμοποιώντας λοιπόν τη σχέση $\nabla \Phi \mathcal{N} = \frac{\nabla \Phi}{|\nabla \Phi|} \nabla \Phi = |\nabla \Phi|$ και τη σχέση (2.12) παίρνουμε τον νόμο εξέλιξης του Φ :

$$\frac{\partial \Phi}{\partial t} = V |\nabla \Phi| \quad (2.14)$$

Η παραπάνω σχέση ισχύει υποχρεωτικά στα σημεία του μηδενικού επιπεδοσυνόλου της συνάρτησης και μπορεί να επεκταθεί και στο υπόλοιπό της [200].

Το πλεονέκτημα αυτής της διατύπωσης είναι ότι η σχέση (2.14) μπορεί να χρησιμοποιηθεί για την εξέλιξη της καμπύλης χωρίς να επιβάλλεται ο υπολογισμός της θέσης της σε κάθε επανάληψη, η εκτίμηση των γεωμετρικών όρων από το περίγραμμά της κ.ο.κ. ενώ η τοπολογία της καμπύλης μπορεί να αλλάζει δυναμικά, χωρίς να χρειάζεται κάποιος ιδιαίτερος χειρισμός ή κάποια απόφαση για το αν πρέπει να αλλάξει ή όχι. Από την άλλη, ένα βασικό μειονέκτημά της είναι ότι ανάγει την εξέλιξη μίας καμπύλης στην μεταβολή μίας επιφάνειας, επιβαρύνοντας υπολογιστικά το πρόβλημα· για το λόγο αυτό έχει προταθεί η τεχνική narrow-band που περιορίζει την εξέλιξη της συνάρτησης Φ μόνο γύρω από την τρέχουσα θέση του μηδενικού επιπεδοσυνόλου της, καθώς και πολυκλιμακωτές [178], πολυπλεγματικές τεχνικές [176] και έμμεσες τεχνικές ολοκλήρωσης [229] που μπορούν να επισπεύσουν την εξέλιξη των καμπυλών.



Σχήμα 2.4: Εξέλιξη ενός Geodesic Active Contour: η τοπολογία της καμπύλης αλλάζει αβίαστα ενώ οι γωνίες των αντικειμένων εντοπίζονται με ακρίβεια.

2.1.4 Geodesic Active Contours

Ένα βασικό πρόβλημα με τις τεχνικές τύπου snakes είναι το πώς μπορεί να μην επηρεάζει την εξέλιξη της καμπύλης η οποιαδήποτε παραμετροποίησή της. Η λύση στο πρόβλημα αυτό δόθηκε ανεξάρτητα από δύο ομάδες [31, 120], όπου η ελαχιστοποίηση μίας απλουστευμένης εκδοχής του συναρτησιακού (2.9) των [118] αποδείχθηκε ότι ισοδυναμεί με την εύρεση γεωδαισικών καμπυλών πάνω στην εικόνα, χρησιμοποιώντας μία Riemannian απόσταση, η οποία καθορίζεται από την ίδια την εικόνα. Αναθέτοντας μικρότερο μήκος σε περιοχές της εικόνας που υπάρχει μεγάλη ένταση ακμών και μεγαλύτερο μήκος σε περιοχές που η εικόνα είναι ομαλή, οι γεωδαισικές μεταξύ δύο σημείων θα διέρχονται από τις ακμές της εικόνας. Συγκεκριμένα αν το Riemannian μήκος του διαστήματος ds είναι ίσο με $g(|\nabla I|)ds$, όπου g είναι μία φθίνουσα συνάρτηση της έντασης ακμών, $|\nabla I|$, τότε το μεταβολικό πρόβλημα διατυπώνεται ως η ελαχιστοποίηση του συναρτησιακού:

$$J[\Gamma] = \int_0^L g(|\nabla I(\Gamma(s))|) ds = \int_0^1 g(|\nabla I(\Gamma(p))|) \left| \frac{\partial \Gamma}{\partial p}(p) \right| dp \quad (2.15)$$

Από τις εξισώσεις Euler - Lagrange προκύπτει η σχέση εξέλιξης:

$$\frac{\partial \Gamma}{\partial t} = g(|\nabla I|)\kappa\mathcal{N} - (\nabla g(|\nabla I|) \cdot \mathcal{N})\mathcal{N} \quad (2.16)$$

Ο πρώτος από τους δύο όρους εξομαλύνει την καμπύλη, λόγω του όρου καμπυλότητας, εκτός από περιοχές όπου υπάρχει μεγάλη ένταση ακμών, οπότε η χαμηλή τιμή του $g(|\nabla I|)$ ακυρώνει τη δράση του. Ο δεύτερος όρος έλκει την καμπύλη προς την κατεύθυνση όπου αυξάνεται η ένταση ακμών, εξασφαλίζοντας ότι η εξέλιξη της καμπύλης θα σταματήσει σε ένα μέγιστό της. Παρατηρούμε ότι στις εξισώσεις αυτές δεν υπεισέρχεται πλέον η παραμετροποίηση της καμπύλης καθώς οι όροι \mathcal{N}, κ που έχουν σχέση με την εξελισσόμενη καμπύλη Γ είναι γεωμετρικές ποσότητες που δεν εξαρτώνται από την παραμετροποίησή της.

Ένα βασικό πρόβλημα της μεθόδου αυτής είναι ότι καθώς χρησιμοποιείται μόνο πληροφορία σχετικά με την ένταση ακμών, είναι ακατάλληλη για περιοχές όπου υπάρχει θόρυβος, υφή, κτλ. οπότε και παγιδεύεται σε υποβέλτιστες λύσεις. Επίσης, αν η καμπύλη αρχικοποιηθεί σε περιοχές όπου δεν υπάρχουν ακμές, ο μόνος όρος που υπεισέρχεται στην εξέλιξη της είναι η καμπυλότητά της, οπότε σταδιακά συρρικνώνεται προς ένα σημείο. Εισάγοντας έναν σταθερό όρο (balloon) [34] στη σχέση εξέλιξης καμπυλών:

$$\frac{\partial \Gamma_i}{\partial t} = g(|\nabla I|)\kappa\mathcal{N} - [(\nabla g(|\nabla I|) \cdot \mathcal{N}) + c]\mathcal{N} \quad (2.17)$$

αντιμετωπίζεται αυτό το πρόβλημα, αλλά προκύπτουν νέα· συγκεκριμένα, το περιγράμμα πρέπει να βρίσκεται πλέον είτε στο εξωτερικό του αντικειμένου είτε στο εσωτερικό του, οπότε ανάλογα εισάγεται καθορίζεται και το πρόσημο του c . Από τα παραπάνω φαίνεται ότι αν και κομψές, οι τεχνικές των Geodesic Active Contours/Snakes σπάνια επαρκούν για επίλυση προβλημάτων κατάτμησης σε φυσικές εικόνες, πέρα από πολύ απλές περιπτώσεις, όπως αυτές στο Σχ. 2.4

2.1.5 Region Competition

Μία σημαντική συνεισφορά στον χώρο των μεταβολικών αλγόριθμων ήταν η εργασία [248] των Zhu & Yuille όπου προτάθηκε ένας μεταβολικός αλγόριθμος για την επίλυση του προβλήματος της κατάτμησης χρησιμοποιώντας μία στατιστική διατύπωση.

Συγκεκριμένα, το πεδίο ορισμού της εικόνας χωρίζεται σε M περιοχές $R_i, i = 1 \dots M$ με σύνορα $\Gamma_i = \partial R_i$, και μέσα σε κάθε μία από αυτές χρησιμοποιείται μία παραμετρική κατανομή $P(\cdot; a_i)$ για τις παρατηρήσεις της. Το πρόβλημα της κατάτμησης εκφράζεται ως η ελαχιστοποίηση του συναρτησιακού:

$$J(\Gamma, \{a\}) = \sum_{i=1}^M \frac{\lambda}{2} \int_{\Gamma_i} ds - \iint_{R_i} \log P(I; a_i) + \nu, \quad \Gamma = \{\Gamma_1, \dots, \Gamma_M\} \quad (2.18)$$

Ο πρώτος όρος τιμωρεί μεγάλου μήκους σύνορα μεταξύ των περιοχών, προτιμώντας ομαλά περιγράμματα, ο δεύτερος επιβάλλει τα δεδομένα να μοντελοποιούνται ικανοποιητικά από τις αντίστοιχες κατανομές ενώ ο τελευταίος τιμωρεί το πλήθος των περιοχών, αποφεύγοντας την υπερκατάτμηση της εικόνας. Στην περίπτωση όπου οι κατανομές μέσα σε κάθε περιοχή μοντελοποιηθούν με Gaussians ίσων διασπορών σ , το συναρτησιακό γίνεται

$$J(\Gamma, \{a\}) = \sum_{i=1}^M \frac{\lambda}{2} \int_{\Gamma_i} ds - \iint_{R_i} \frac{(I - \mu_i)^2}{2\sigma_i^2} + c \quad (2.19)$$

όπου το μ_i είναι η μέση τιμή της i -οστής Gaussian κατανομής, ενώ η ποσότητα c περιλαμβάνει τόσο τους όρους ποινής ν όσο και το άθροισμα $\sum_i |R_i| - \log \sqrt{2\pi}\sigma$, όπου με $|\cdot|$ συμβολίζουμε την πληθικότητα. Πέρα από τον όρο c , η παραπάνω έκφραση προκύπτει από το συναρτησιακό (2.2) όταν πάρουμε το όριο του για $\alpha \rightarrow \infty$, οπότε αποκλείονται οι οποιεσδήποτε διακυμάνσεις της U εκτός των συνόρων των περιοχών. Η U τότε είναι τμηματικά σταθερή και η τιμή της που ελαχιστοποιεί τον πρώτο όρο ποινής του (2.2) είναι η μέση τιμή της εικόνας μέσα στην περιοχή i , δηλαδή η μ_i .

Το πρακτικό πλεονέκτημα αυτής της πιθανοτικής διατύπωσης είναι ότι επιτρέπει μία μεγαλύτερη ποικιλία κατανομών για κάθε περιοχή, που μπορούν να καθορίζονται δυναμικά, με βάση τις παρατηρήσεις. Έτσι για παράδειγμα μπορούν να χρησιμοποιηθούν Gaussian κατανομές πάνω σε πολυδιάστατα χαρακτηριστικά, με μη-διαγώνιους πίνακες συνδιακύμανσης, κατανομές τύπου Mixture of Gaussians κ.ο.κ.[219, 193]. Ενώ αυτά δεν προβλέπονται από το αρχικό μοντέλο Mumford-Shah, έχοντας κατά νου την πιθανοτική του ερμηνεία μπορούν να χρησιμοποιηθούν αβίαστα. Ένα θεωρητικό πλεονέκτημα είναι ότι επιτρέπει να δούμε το πρόβλημα της κατάτμησης στο γενικότερο πλαίσιο της Ανάλυσης-μέσω-Σύνθεσης, θεωρώντας ότι οι τιμές των μ_i παρέχουν ένα μοντέλο σύνθεσης για τα τμήματα της εικόνας.

Για την ελαχιστοποίηση του συναρτησιακού (2.18) χρησιμοποιούνται τεχνικές του Λογισμού Μεταβολών από τις οποίες προκύπτει η παρακάτω σχέση για την εξέλιξη των ορίων της περιοχής i που συνορεύει με την περιοχή j :

$$\frac{\partial \Gamma_i}{\partial t} = -\lambda \kappa \mathcal{N} + \log \frac{P_i(I; a_i)}{P_j(I; a_j)} \mathcal{N} \quad (2.20)$$

Ο πρώτος όρος της εξέλιξης προκύπτει από τον όρο ποινής στο μήκος του Γ_i και εμφανίζεται στις περισσότερες εξισώσεις εξέλιξης καμπυλών. Η καινοτομία έγκειται στον δεύτερο όρο που οδηγεί την περιοχή i στην κατάληψη των παρατηρήσεων της εικόνας που μοντελοποιεί καλύτερα από την περιοχή j : αν $\frac{P(I; a_i)}{P(I; a_j)} > 0$ η καμπύλη i προελαύνει, διαφορετικά οπισθοχωρεί. Από την συμπεριφορά του αυτή ο αλγόριθμος ονομάζεται Region Competition καθώς οι καμπύλες ανταγωνίζονται για την κατοχή των παρατηρήσεων της εικόνας.

Πέρα από την θέση των συνόρων, χρειάζεται να εκτιμηθούν και οι παράμετροι των κατανομών μέσα στις περιοχές: οι παράμετροι τίθενται ίσες με τις εκτιμήτριες μέγιστης πιθανοφάνειας (Maximum Likelihood) των Gaussians μέσα σε κάθε περιοχή, όπως αυτή ορίζεται από την τρέχουσα θέση των ορίων της. Αυτό γίνεται σε εναλλαγή με την εξέλιξη των καμπυλών, δίνοντας μία μη επιβλεπόμενη μέθοδο για την ελαχιστοποίηση του συναρτησιακού (2.18).

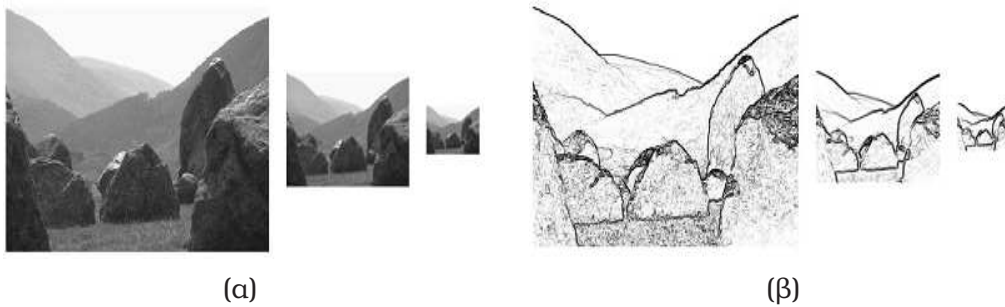
Η αρχική προσέγγιση έχει επεκταθεί [219] ώστε να χρησιμοποιεί μία μεγαλύτερη οικογένεια κατανομών από τις Gaussians ενώ συνδυάζοντας τις ντετερμινιστικές εξισώσεις εξέλιξης καμπυλών με αλγόριθμους στοχαστικής αναζήτησης αποφεύγονται τα τοπικά ελάχιστα στα οποία μπορούν να καταλήξει ο αλγόριθμος αναζήτησης του Region Competition.

Εμπλουτίζοντας τον αρχικό σχήμα εξέλιξης στην αναφορά [179] προτάθηκε ο συνδυασμός των όρων των συναρτησιακών των Geodesic Active Contours και του Region Competition σε έναν ενιαίο όρο ποινής, από την ελαχιστοποίηση του οποίου προκύπτει η σχέση εξέλιξης των Geodesic Active Regions:

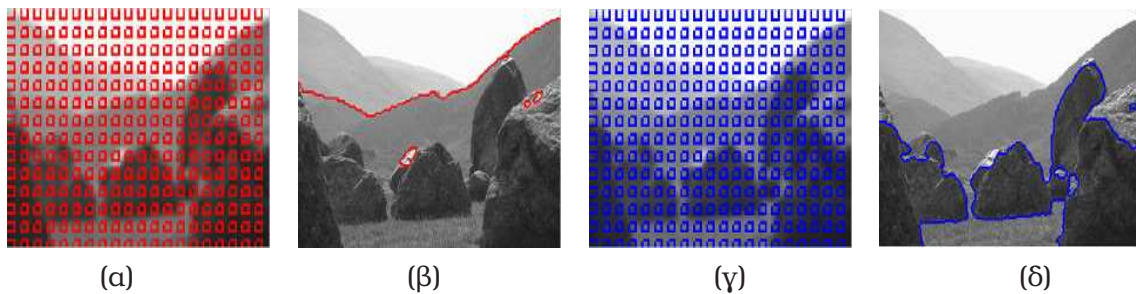
$$\frac{\partial \Gamma_i}{\partial t} = \log \frac{P_i(I; a_i)}{P_j(I; a_j)} \mathcal{N} + \lambda (g(|\nabla I|) \kappa \mathcal{N} - (\nabla g(|\nabla I|) \cdot \mathcal{N}) \mathcal{N}) \quad (2.21)$$

Για την υλοποίηση αυτής της σχέσης εξέλιξης εισήχθησαν τεχνικές επιπεδοσυνόλων ενώ ένα πολυδιακριτικό σχήμα χρησιμοποιήθηκε για να εξασφαλίσει ευρωστία και ταχύτητα. Στη δική μας υλοποίηση χρησιμοποιούμε τις ρυθμίσεις της αναφοράς αυτής όπου αναλύονται και θέματα πρακτικής σημασίας, όπως η εγγυημένη κάλυψη όλης της εικόνας από τα επιπεδοσύνολα και ο χειρισμός των πολλαπλών επικαλύψεων: εναλλακτικές προσεγγίσεις που κινούνται σε ένα παρόμοιο πλαίσιο παρουσιάζονται στις αναφορές [32, 241].

Στις παρακάτω εικόνες δείχνουμε ένα παράδειγμα εφαρμογής του αλγόριθμου Region Competition υλοποιημένου μέσω τεχνικών επιπεδοσυνόλων, σύμφωνα με τις ρυθμίσεις των [178]: παρατηρούμε ότι δεν χρειάζεται η αρχικοποίηση των καμπυλών με συγκεκριμένο τρόπο, ενώ τα σύνορα των περιοχών εντοπίζονται με μεγάλη ακρίβεια.



Σχήμα 2.5: Πολυδιακριτική αναπαράσταση (α) της εικόνας και (β) των ακμών της.



Σχήμα 2.6: (α)-(β): Αρχική και τελική θέση των ορίων της περιοχής που αντιστοιχεί στην φωτεινό τμήμα της εικόνας και (γ)-(δ) στο σκοτεινό τμήμα.

2.2 Αναγνώριση Αντικειμένων

2.2.1 Εισαγωγή

Τα δύο κύρια προβλήματα της όρασης υψηλού επιπέδου είναι: η ανίχνευση αντικειμένων (object detection), στόχος της οποίας είναι να ληφθεί η απόφαση αν υπάρχει ένα αντικείμενο, όπως ένα πρόσωπο σε μία εικόνα και η αναγνώριση αντικειμένων (object recognition), όπου υπάρχουν πολλαπλά αντικείμενα και πρέπει να αποφασίσουμε με βάση τις παρατηρήσεις μας υπέρ ενός από αυτά. Συνήθως δε με τον όρο αναγνώριση αντικειμένων εννοείται και όλο το φάσμα των προβλημάτων της όρασης υψηλού επιπέδου.

Δύο συμπληρωματικές πτυχές του προβλήματος, που εισάγουν επιπλέον απαιτήσεις είναι

- Η χρήση κατηγοριών αντικειμένων αντί συγκεκριμένων αντικειμένων, οπότε πρέπει να ληφθεί υπόψη η ποικιλία που υπάρχει μέσα στην ίδια κατηγορία.
- Ο εντοπισμός των αντικειμένων (object localization), όπου στόχος είναι πέρα από την λήψη της απόφασης περί της ύπαρξης/κατηγορίας του αντικειμένου να βρεθεί στην εικόνα η τοποθεσία του.

Γενικότερα αυτό το οποίο έχει κανείς κατά νου λέγοντας 'αναγνώριση αντικειμένων' μπορεί να εκφραστεί με βάση τους παραπάνω όρους ως ο εντοπισμός των θεωρούμενων κατηγοριών σε μία εικόνα και η αναγνώριση των μεμονωμένων αντικειμένων των κατηγοριών στις οποίες εστιάζουμε· π.χ. μετά τον εντοπισμό ενός αυτοκινήτου να βρούμε και αν είναι το δικό μας.

Σε αντίθεση με την εντύπωση που μπορεί να έχει κανείς διαισθητικά, στην αναγνώριση αντικειμένων η δυσκολία ενός προβλήματος είναι αντιστρόφως ανάλογη της

εξειδίκευσής του. Για παράδειγμα, ενώ χρησιμοποιώντας ακόμα και απλοϊκές μεθόδους μπορεί να είναι εφικτός ο εντοπισμός ελαττωματικών εργοστασιακών εξαρτημάτων βάσει ελάχιστων διαφορών αποτελεί μία μεγάλη πρόκληση να αναπτυχθούν μέθοδοι που να αποφαινόνται για το αν υπάρχει ένας άνθρωπος σε μία εικόνα.

Οι παράγοντες που δυσχεραίνουν την επίλυση του προβλήματος μπορούν να συνοψιστούν στα παρακάτω σημεία :

- Ποικιλία της εμφάνισης μέσα στην ίδια κατηγορία.
- Μεταβολή του φωτισμού.
- Εσφαλμένες ανιχνεύσεις λόγω περίπλοκη δομής (clutter).
- Αλληλοεπικαλύψεις με διαφορετικά αντικείμενα.
- Πολλαπλές κλίμακες/κατευθύνσεις.
- Αλλαγές στην εμφάνιση λόγω περιστροφής του αντικειμένου.

Ο κύριος όγκος της έρευνας στην αναγνώριση αντικειμένων έχει εστιάσει στα τέσσερα πρώτα προβλήματα, θεωρώντας ότι τα τελευταία δύο μπορούν να αντιμετωπιστούν μέσω εξαντλητικής αναζήτησης και ξεχωριστά μοντέλα για διαφορετικές όψεις αντίστοιχα, που σημαίνει όμως μεγαλύτερες υπολογιστικές απαιτήσεις. Πρόσφατες εργασίες έχουν προτείνει μεθόδους για να απλοποιηθούν και επισπεύσουν την αντιμετώπισή τους [215, 62], αλλά αυτές οι πτυχές του προβλήματος είναι συμπληρωματικές του αντικείμενου της διατριβής, οπότε θα εστιάσουμε στα υπόλοιπα προβλήματα.

2.2.2 Επισκόπηση Τεχνικών Αναγνώρισης Αντικειμένων

Τις πρώτες δύο δεκαετίες της έρευνας στην αναγνώριση αντικειμένων υπήρχαν δύο κυρίαρχα ρεύματα, αντίστοιχα της άνωθεν και κάτωθεν προσέγγισης στην όραση : τα κάτωθεν μοντέλα, π.χ. [11, 183], βασιζόμενα στο μοντέλο του Marr και σε ιδέες από την συντακτική αναγνώριση αντικειμένων [72] θεωρώντας ότι μία εικόνα μπορεί να αναλυθεί σε κάποια πρωτογενή σχήματα (π.χ. κύλινδροι, παραλληλεπίπεδα, σφαίρες) αναζητούν να βρουν πώς μπορούν να συνδεθούν τα ανιχνευθέντα σχήματα μέσω περιπλοκότερων μοντέλων για αντικείμενα. Τα άνωθεν μοντέλα [48, 80], χρησιμοποιώντας τριδιάστατες αναπαραστάσεις για τα αντικείμενα ανάγουν το πρόβλημα της αναγνώρισης στην εκτίμηση των παραμέτρων της προβολής του μοντέλου στην εικόνα. Όπως αναλύθηκε και στο κεφάλαιο της εισαγωγής, οι προσεγγίσεις αυτές δεν ανταποκρίνονται στη δυσκολία του προβλήματος, καθώς η εξαγωγή των πρωτογενών σχημάτων είναι προβληματική στην πράξη, ενώ δεν είναι βιώσιμη η κατασκευή και χρήση τριδιάστατων μοντέλων για κάθε μεμονωμένο αντικείμενο που υπάρχει στο περιβάλλον μας.

Στις αρχές τις προηγούμενης δεκαετίας, καθώς οι περιορισμοί των παραπάνω προσεγγίσεων είχαν γίνει αντιληπτοί, ξεκίνησαν οι πρώτες προσπάθειες αφενός για να ληφθεί υπόψη η ποικιλία στην εμφάνιση που μπορούν να παρουσιάζουν τα αντικείμενα μίας κατηγορίας και αφετέρου για την κατασκευή τέτοιων μοντέλων με μικρότερη ανθρώπινη παρέμβαση. Σύντομα η αναγνώριση αντικειμένων στράφηκε προς στατιστικά μοντέλα για αναπαράσταση αντικειμένων, αντίστοιχα αυτών που έχουν επικρατήσει στην τεχνολογία φωνής και τεχνικές εκμάθησης μηχανών για την αυτόματη κατασκευή τους από δεδομένα εκπαίδευσης. Καθώς η προσπάθεια αυτή



Σχήμα 2.7: Αναγνώριση μέσω αντιστοίχισης μοντέλων [149]: ορισμένα από τα χαρακτηριστικά σημεία της εικόνας, όπως γωνίες και διασταυρώσεις μπορούν να αντιστοιχιστούν σε αυτά προαποθηκευμένων αντικειμένων. Αν καθορίζοντας κατάλληλα τις παραμέτρους προβολής του αντικειμένου μπορούν να ερμηνευθούν περισσότερα χαρακτηριστικά, αυτό θεωρείται ενδεικτικό της ύπαρξης του αντικειμένου.

έχει αποφέρει καρπούς, δυσκολότεροι στόχοι φαίνονται πλέον εφικτοί, όπως η εκμάθηση μοντέλων χωρίς επίβλεψη, και ο συνδυασμός των μοντέλων των αντικειμένων με άλλες λειτουργίες της όρασης.

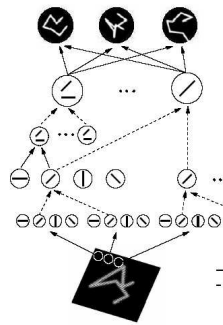
Παρακάτω παρουσιάζονται αρχικά συνοπτικά τα κύρια ρεύματα της αναγνώρισης αντικειμένων και στη συνέχεια αναλυτικότερα οι περιοχές με τις οποίες ασχοληθήκαμε στα πλαίσια της διατριβής αυτής. Θα μπορούσαμε να ταξινομήσουμε τις τεχνικές αναγνώρισης αντικειμένων βάσει του εύρους του προβλήματος που καλύπτουν, των μοντέλων των αντικειμένων που χρησιμοποιούν και των μαθηματικών/αλγοριθμικών τεχνικών στις οποίες βασίζονται. Προτιμήθηκε να χρησιμοποιηθεί ένας συνδυασμός αυτών των κριτηρίων, ώστε να διαφανούν οι 'σχολές' που υπάρχουν στο πεδίο. Ορισμένες βιβλιογραφικές αναφορές που καλύπτουν σε μεγαλύτερο βάθος διάφορες πτυχές του αυτού του πεδίου είναι [60, 66, 221, 239].

Αναγνώριση μεμονωμένων αντικειμένων μέσω μοντέλων και αντιστοίχισης

Με βάση την προσέγγιση αυτή [48, 80], τα αντικείμενα που βρίσκονται σε μία διδιάστατη εικόνα αναγνωρίζονται χρησιμοποιώντας αποθηκευμένες αναπαραστάσεις τριδιάστατων αντικειμένων ή συνδυασμούς διδιάστατων όψεών τους. Οι βασικές συνιστώσες του προβλήματος είναι η κατασκευή των αναπαραστάσεων, η παραγωγή ενός περιορισμένου πλήθους υποθέσεων δεδομένης μίας εικόνας και η εκτίμηση της πιστότητας των παρατηρήσεων στις προβλέψεις του μοντέλου.

Για την αναπαράσταση των μοντέλων έχουν χρησιμοποιηθεί αρχικά τριδιάστατα μοντέλα όπως μοντέλα CAD και μοντέλα πλέγματος (wire framed objects) ενώ στην αναφορά [222] οι γραμμικοί συνδυασμοί από διδιάστατες όψεις εισήχθησαν, περιορίζοντας την πολυπλοκότητα του προβλήματος.

Για την αναζήτηση της θέσης του αντικειμένου μπορούν να χρησιμοποιηθούν άνωθεν τεχνικές pruning για να περιορίσουν τις δυνατές ερμηνείες των παρατηρήσεων βάσει των περιορισμών που εισάγει η γεωμετρία του αντικειμένου [80], η κάτωθεν τεχνικές όπως ο μετασχηματισμός Hough [149], όπου ένα σύνολο από σημεία ενδιαφέροντος 'ψηφίζει' για τις δυνατές θέσεις του αντικειμένου. Έχοντας κάποιες υποψήφιες τοποθεσίες, η εκτίμηση της πιστότητας των παρατηρήσεων επιτυγχάνεται βάσει κάποιου κατάλληλου μέτρου της απόστασης μεταξύ των χαρακτηριστικών των παρατηρήσεων και του μοντέλου, όπως η απόσταση Hausdorff μεταξύ των ακ-



(a)

Σχήμα 2.8: Αναγνώριση μέσω τεχνικών νευρωνικών δικτύων/αναγνώρισης προτύπων [192]: τα αντικείμενα ανιχνεύονται από μία κατανεμημένη αρχιτεκτονική μέσω μίας αλληλουχίας εξαγωγής περιπλοκότερων και ακριβέστερων χαρακτηριστικών

μών τους [103] ή το κριτήριο ελάχιστων τετραγώνων μεταξύ της πρόβλεψης και των μετρήσεων των θέσεων των σημείων ενδιαφέροντος.

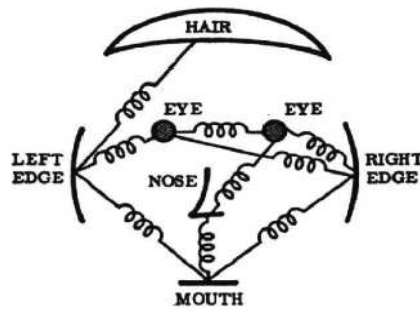
Αν και οι προσεγγίσεις αυτές έχουν δώσει ικανοποιητικά αποτελέσματα σε πρακτικές εφαρμογές (π.χ. επισκόπηση εξαρτημάτων σε βιομηχανικές γραμμές συναρμολόγησης, ανάκληση αντικειμένων από μία βάση δεδομένων), το πεδίο εφαρμογών τους είναι περιορισμένο σε συγκεκριμένα αντικείμενα αντί για κατηγορίες αντικειμένων. Όμως, η εμπειρία που έχει συσσωρευθεί έχει αποδειχθεί ωφέλιμη για την διατύπωση και επίλυση νέων προβλημάτων που προκύπτουν σε σύγχρονες τεχνικές αναγνώρισης αντικειμένων.

Αναγνώριση αντικειμένων με τεχνικές αναγνώρισης προτύπων & νευρωνικών δικτύων

Οι τεχνικές αυτής της σχολής ξεκίνησαν από την έρευνα στο χώρο των νευρωνικών δικτύων, όπου το πρόβλημα της ανίχνευσης αντιμετωπίζεται μέσω της παράλληλης, κατανεμημένης επεξεργασίας της πληροφορίας της εικόνας (προσέγγιση 'κυλιόμενου παράθυρου'). Ξεκινώντας από την εξαγωγή ενός συνόλου χαρακτηριστικών τα οποία είναι αναλλοίωτα ως προς ανεπιθύμητες πηγές ποικιλίας στην εμφάνιση, όπως ο φωτισμός, η ανίχνευση ενός αντικειμένου εκφράζεται ως ένα πρόβλημα αναγνώρισης προτύπων: δεδομένου ενός συνόλου χαρακτηριστικών μέσα στο εξεταζόμενο παράθυρο, ζητείται να ληφθεί η απόφαση για το εάν υπάρχει ένα αντικείμενο.

Η εξέλιξη των ιδεών στην περιοχή αυτή ακολουθεί ως ένα βαθμό την πρόοδο στο πεδίο της αναγνώρισης προτύπων και αποτυπώνεται στις αναφορές [189, 196, 212, 175, 227, 95, 3, 199]. Τα προβλήματα που καλούνται να αντιμετωπίσουν τέτοιες τεχνικές αφορούν την εξαγωγή κατάλληλων χαρακτηριστικών για τα αντικείμενα [220, 175, 95, 199], την αποδοτική υλοποίηση των αλγορίθμων ανίχνευσης [227] καθώς και όλα τα προβλήματα του πεδίου της αναγνώρισης προτύπων [224], όπως η εκμάθηση από ένα περιορισμένο πλήθος δεδομένων, η αποφυγή της υπερεκπαίδευσης σε συνδυασμό με την κατασκευή ενός ευέλικτου ταξινομητή κ.ο.κ.

Από τα βασικά πλεονεκτήματα των μεθόδων αυτών είναι η ωριμότητα των διαθέσιμων τεχνικών από την περιοχή της μηχανικής μάθησης, όπως τα Support Vector Machines [224, 175], αλγόριθμοι τύπου Boosting [227], καθώς και η ταχύτητα με την οποία μπορεί να γίνει η διαδικασία της ανίχνευσης αν χρησιμοποιηθούν κατάλληλες τεχνικές [227]. Υπάρχουν όμως δύο κύρια προβλήματα: το βασικότερο είναι



(a')

Σχήμα 2.9: Αναγνώριση αντικειμένων μέσω τμηματικών αναπαραστάσεων [65]: η αναπαράσταση ενός αντικειμένου ανάγεται στην μοντελοποίηση των επιμέρους τμημάτων του και των εξαρτήσεων μεταξύ τους

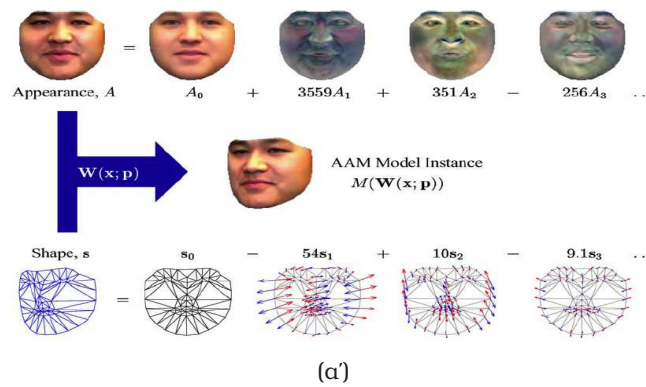
ότι για την εκμάθησή τους χρειάζεται συνήθως ένα σύνολο εκπαίδευσης της τάξης των μερικών χιλιάδων εικόνων, καθώς περιλαμβάνεται ένα μεγάλο πλήθος παραμέτρων που πρέπει να εκμαθευτούν. Αντίθετα, σε αναφορές όπως [57] επιτυγχάνονται ικανοποιητικά αποτελέσματα ακόμα και με μία ή δύο εικόνες, ενώ εμείς ως άνθρωποι αρκούμαστε σε μία εικόνα, όταν βλέπουμε για παράδειγμα για πρώτη φορά ένα νέο εξάρτημα υπολογιστή. Επίσης μία τέτοια προσέγγιση δεν είναι εύκολο να αντιμετωπίσει επικαλύψεις με άλλα αντικείμενα και γενικότερα προβλήματα που δεν έχουν συμπεριληφθεί στο σύνολο εκπαίδευσης. Αντίθετα, χρησιμοποιώντας το πλαίσιο της ανάλυσης μέσω σύνθεσης μπορούν να εντοπιστούν και να αγνοηθούν οι μη-προβλέψιμες περιοχές της εικόνας.

Αναγνώριση αντικειμένων με τμηματικές αναπαραστάσεις

Τα μοντέλα που ανήκουν στην κατηγορία των τμηματικών ταξινομητών [58, 63, 57, 62, 3, 27, 233, 95, 160, 237] βασίζουν την ανίχνευση των αντικειμένων στην ανίχνευση χαρακτηριστικών σημείων (keypoints) των αντικειμένων. Χρησιμοποιώντας ανεξάρτητα μοντέλα για την ανίχνευση κάθε ενός από αυτά απλουστεύεται σημαντικά το πρόβλημα, καθώς η ποικιλία που υπάρχει στην εμφάνιση του καθενός μοντελοποιείται ανεξάρτητα και δεν χρειάζεται να ληφθούν υπόψη όλοι οι συνδυασμοί του. Επίσης είναι εύκολο να χρησιμοποιηθεί στατιστική πληροφορία για την χωρική διάταξη των τμημάτων, ώστε να εισαχθεί και γεωμετρική γνώση στην διαδικασία ανίχνευσης.

Αν και η δημοσίευση που παρουσίασε αυτό το μοντέλο ήταν το άρθρο των Fischler & Erschlanger [65], η πρώτη εφαρμογή τους σε πρακτικά προβλήματα παρουσιάζεται μετά από είκοσι έτη στην αναφορά [237]. Διαφορετικές μαθηματικές διατυπώσεις του ίδιου μοντέλου έχουν προταθεί κατά καιρούς [65, 237, 95, 223, 3, 144], αλλά η επικρατέστερη πλέον είναι η στατιστική [27, 233, 62, 135, 57, 58, 45] που επιτρέπει αβίαστα να συνδυαστούν όροι όπως η πληροφορία σχετικά με την εμφάνιση και η γεωμετρία του αντικειμένου, ενώ επιτρέπει την διατύπωση της εκμάθησής τους στο πλαίσιο των αλγορίθμων για γραφικά και αναγεννητικά μοντέλα.

Η προσέγγιση αυτή συνδυάζει την αποδοτικότητα του κάτωθεν μοντέλου της όρασης, όπου από απλά επιμέρους τμήματα αναγνωρίζονται περιπλοκότερα αντικείμενα, με τους περιορισμούς που εισάγει η πρότερη γνώση των άνωθεν μοντέλων, απλοποιώντας την αναζήτηση σε συνδυασμούς που θα μπορούσαν να είχαν προκύψει από



Σχήμα 2.10: Αναγνώριση μέσω παραμορφώσιμων μοντέλων των αντικειμένων [157]: μοντελοποιώντας την παραμόρφωση στην εμφάνιση και στο σχήμα του αντικειμένου, μπορεί να αναπαρασταθεί μία κατηγορία αντικειμένων χρησιμοποιώντας ένα χαμηλοδιάστατο παραμετρικό μοντέλο

το αντικείμενο. Το βασικό της μειονέκτημα είναι ότι αντί για το αντικείμενο καθαυτό μοντελοποιούνται τα τμήματά του και οι αλληλεξαρτήσεις τους που θεωρούνται χρήσιμες για την ανίχνευσή του, δίνοντας ένα μοντέλο περιορισμένων δυνατοτήτων για άλλες λειτουργίες, όπως για παράδειγμα για την κατάτμηση ενός αντικειμένου ή την αναγνώριση της ταυτότητάς του.

Αναγνώριση αντικειμένων με παραμορφώσιμα μοντέλα

Η προσέγγιση αυτή βασίζεται στην χρήση ενός παραμετρικού μοντέλου των παραμορφώσεων της εμφάνισης και γεωμετρίας ενός πρωτότυπου αντικειμένου, το οποίο να καλύπτει όλη την ποικιλία που χαρακτηρίζει μία κατηγορία αντικειμένων. Δεδομένης μίας νέας εικόνας και μίας αρχικοποίησης, το πρόβλημα της αναγνώρισης/ανίχνευσης ανάγεται στην μεταβολή των παραμέτρων του μοντέλου ώστε αυτό να αναπαράγει τις παρατηρήσεις. Οι δύο βασικές συνιστώσες της προσέγγισης αυτής είναι η μοντελοποίηση της παραμόρφωσης και η εκτίμηση των παραμέτρων των μοντέλων.

Ως πρώτες προσπάθειες στην κατεύθυνση της μοντελοποίησης της παραμόρφωσης του σχήματος μπορούν να θεωρηθούν αφενός τεχνικές όπως τα παραμορφώσιμα πρότυπα (Deformable Templates) [247], αφετέρου η αναγνώριση μέσω συνδυασμού των όψεων [222]. Στην πρώτη περίπτωση ένα μοντέλο για ορισμένες περιοχές του αντικειμένου επαναληπτικά παραμορφώνεται εωσότου υπάρξει αντιστοίχιση των περιοχών του μοντέλου με αυτές τις εικόνες, ενώ στην δεύτερη προσέγγιση [222] έχοντας κάποιες όψεις ενός αντικειμένου υπολογίζεται ο συνδυασμός τους που μπορεί να ερμηνεύσει μία νέα όψη του αντικειμένου από μία διαφορετική οπτική γωνία. Μία συμπληρωματική κατεύθυνση έρευνας στις αρχές και στα μέσα της προηγούμενης δεκαετίας ήταν η προσπάθεια για την έκφραση όλης της ποικιλίας στην εμφάνιση ενός αντικειμένου μέσω της τεχνικής PCA [121, 220, 160]: θεωρώντας ότι η φωτεινότητα μίας κατηγορίας αντικειμένων μπορεί να εκφραστεί μέσω ένα γραμμικού αναπτύγματος, η τεχνική PCA μπορεί να χρησιμοποιηθεί για να βρεθούν τα στοιχεία βάσης για τα οποία ελαχιστοποιείται το σφάλμα ανακατασκευής.

Ο συνδυασμός των ιδεών αυτών οδήγησε ανεξάρτητα στα ουσιαστικά ισοδύναμα (Active Appearance Models- AAM) [37, 41, 157] και Morphable Models [226, 112, 16]. Τα μοντέλα αυτά θεωρούν ότι οι παρατηρήσεις που προέρχονται από το

αντικείμενο μπορούν να μοντελοποιηθούν από την χωρική παραμόρφωση του αντικειμένου, η οποία ακολουθείται από την παραμόρφωση της φωτεινότητάς του. Στην χωρική παραμόρφωση συμπεριλαμβάνονται τόσο παραμορφώσεις από απλούς μετασχηματισμούς (π.χ. μετατόπιση, περιστροφή, αφινικοί μετασχηματισμοί) όσο και από περιπλοκότερες παραμορφώσεις που είναι χαρακτηριστικές της συγκεκριμένης κατηγορίας αντικειμένων, όπως είναι η έκφραση ενός προσώπου ή το μήκος ενός αυτοκινήτου.

Τα μοντέλα αυτά αποτελούν μία σημαντική επιτυχία της άνωθεν όρασης κυρίως λόγω της μαθηματικής τους απλότητας σε συνδυασμό με το πλήθος εφαρμογών τους σε πρακτικά προβλήματα. Το βασικό πρόβλημα με τα μοντέλα αυτά είναι ότι καθώς χρησιμοποιείται ένας επαναληπτικός αλγόριθμος δεν είναι εφικτό να ερευνηθεί κανείς για όλες τις θέσεις, κλίμακες της εικόνας κ.ο.κ. αν υπάρχει ένα πρόσωπο. Συνεπώς τέτοιες μέθοδοι δεν είναι κατάλληλες για ανίχνευση αντικειμένων εξαρχής, αλλά μπορούν να χρησιμοποιηθούν για την λήψη της τελικής απόφασης, χρησιμοποιώντας τις εξόδους ενός άλλου ταξινομητή για αρχικοποίηση. Για αυτό χρησιμοποιούνται κυρίως για αναγνώριση αντικειμένων θεωρώντας ότι είναι γνωστό ότι στην θέση που αρχικοποιούνται υπάρχει ένα αντικείμενο κάποιας συγκεκριμένης κατηγορίας. Μία άλλη εφαρμογή τους είναι η καταγραφή της εξέλιξης (tracking) αντικειμένων, όπου χρησιμοποιείται η πρόβλεψη βάσει της προηγούμενης θέσης του μοντέλου για να ξεκινήσει η αναζήτηση της τρέχουσας.

2.2.3 Μοντέλα από το χώρο της Αναγνώρισης Προτύπων

Σύμφωνα με την προσέγγιση της αναγνώρισης προτύπων μία περιοχή της εικόνας θεωρείται ως ένα σύνολο από δεδομένα, για τα οποία ζητείται από έναν ταξινομητή να αποφασίσει αν περιέχουν ένα αντικείμενο. Οι περισσότερες από τις τεχνικές που χρησιμοποιούνται αντιμετωπίζουν μία ειδική περίπτωση του προβλήματος της προσέγγισης συναρτήσεων: η είσοδος του συστήματος είναι το όρισμα της συνάρτησης και η απόφαση που παίρνει η τιμή της. Στη διάρκεια της εκπαίδευσης παρέχονται ζεύγη εξόδων-εισόδων της συνάρτησης και ζητείται να βρεθεί μία προσέγγιση της συνάρτησης που να μπορεί να γενικεύσει τη συμπεριφορά του συστήματος όταν του παρουσιάζονται εισοδοί τις οποίες δεν έχει συναντήσει στο στάδιο εκπαίδευσης. Στα πλαίσια της διατριβής αυτής αρχικά εξετάστηκαν αρκετές μέθοδοι για το πρόβλημα της προσέγγισης συναρτήσεων [132], ανάμεσα στις οποίες συμπεριλαμβάνονται μοντέλα τύπου SVM[224], Radial Basis Function Networks-RBFs[76] και Mixture of Experts-MoEs [115]. Στη συνέχεια παρουσιάζονται συνοπτικά οι σημαντικότερες και αποδοτικότερες τεχνικές από αυτές που δοκιμάστηκαν.

Δίκτυα Radial Basis Function

Τα δίκτυα τύπου RBF συνδυάζουν απλούς αλγόριθμους για εκπαίδευση με μία διαισθητικά σαφή ερμηνεία με επιθυμητές ιδιότητες για προσέγγιση συναρτήσεων, όπως ομαλότητα, διαφορισιμότητα και τοπικότητα της προσέγγισης. Για την προσέγγιση μίας συνάρτησης f αθροίζονται K όροι, ο καθένας από τους οποίους συμβάλλει στην τελική έκφραση μόνο σε μία υποπεριοχή του πεδίου ορισμού της συνάρτησης:

$$f(X) \simeq F(X) = \sum_{i=1}^K C_i G_i(X - X_i) \quad (2.22)$$

Η συνάρτηση G_i είναι μία φθίνουσα συνάρτηση της απόστασης του X από το σημείο X_i : μία συνηθισμένη επιλογή είναι μία Gaussian συνάρτηση ή μία συνάρτηση σπλήνας (spline).

Το βασικό τους πλεονέκτημα σε σχέση με παλαιότερους ταξινομητές, π.χ. τύπου Multi Layer Perceptrons είναι ότι η έκφραση που χρησιμοποιείται είναι διαισθητικά σαφής, ενώ οι σχέσεις τους με τεχνικές regularization [76] τους έχει παρέχει ένα σαφές θεωρητικό υπόβαθρο.

Για προβλήματα όρασης εξετάστηκαν απλές αρχιτεκτονικές π.χ. [189] που κύριο σκοπό είχαν να εξετάσουν αν και πώς μπορεί να εκμαθηθεί μία συνάρτηση ταξινόμησης f βάσει γνωστών όψεων των αντικειμένων, X_i . Συγκεκριμένα για κάθε όψη i του αντικειμένου μία Gaussian με κέντρο X_i τοποθετείται στην έκφραση (2.22), και για μία νέα όψη X ζητείται να βρεθεί κατά πόσο η $f(X) \simeq F(X)$ παίρνει υψηλή τιμή, οπότε μπορεί να θεωρηθεί ότι η νέα όψη προήλθε από το αντικείμενο.

Mixture of Experts

Στις αρχιτεκτονικές τύπου Mixture of Experts [115] η προσέγγιση της συνάρτησης βασίζεται στον συνδυασμό τοπικών εκφράσεων κατά έναν τρόπο που προσφέρεται σε μία πιθανοτική ερμηνεία, επιτρέποντας τη χρήση αποτελεσματικών αλγορίθμων εκτίμησης παραμέτρων. Προσεγγίσεις της ίδιας μορφής είχαν προταθεί και νωρίτερα από το χώρο της Ασαφούς Λογικής, όπου είναι γνωστές ως Takagi Sugeno Kang Networks [214], όπου το ίδιο μοντέλο παρουσιάζεται με την διαφορά ότι όπου στη συνέχεια λέμε πιθανότητα, στην αναφορά [214] χρησιμοποιείται ο όρος δυνατότητα.

Δεδομένης μίας εισόδου X , έχουμε K τοπικά συστήματα προσέγγισης l_i , το καθένα από τα οποία δίνει μία εκτίμηση της εξόδου του συστήματος, $l_i(X)$. Οι έξοδοι των συστημάτων αυτών σταθμίζονται με όρους που εξαρτώνται από την τιμή του X , $\mu_i(X)$, και η έξοδος του συστήματος προκύπτει ως ένας σταθμισμένος μέσος των επιμέρους εξόδων:

$$F(X) = \frac{\sum_{i=1}^K l_i(X)\mu_i(X)}{\sum_{i=1}^K \mu_i(X)} \quad (2.23)$$

Μία τέτοια αρχιτεκτονική ανάγει το πρόβλημα της προσέγγισης της f σε K απλούστερα υποπροβλήματα περιορίζοντας το πεδίο στο οποίο γίνεται η προσέγγιση μέσω των $\mu_i(X)$, επιτρέποντας τη χρήση απλών εκφράσεων $l_i(X)$ για την κατασκευή πολύπλοκων συναρτήσεων $F(X)$. Ο όρος $\mu_i(X)$ είναι συνήθως μία Gaussian και χρησιμοποιείται για να επιλέξει ποια από τις τοπικές προσεγγίσεις χρησιμοποιείται (μηχανισμός gating), ενώ ο όρος $l_i(X)$ είναι μία απλή έκφραση της μορφής $l_i(X) = A \cdot X + b$.

Η πιθανοτική ερμηνεία της αρχιτεκτονικής αυτής προκύπτει αβίαστα: θεωρούμε ότι έχουμε K εναλλακτικές αποκρίσεις του συστήματος στην είσοδο X , με μέση τιμή $E(U|X, i) = l_i(X)$, $i = 1, \dots, K$. Δεδομένης της εισόδου X η πιθανότητα να έχουμε την i -οστή απόκριση δίνεται από τον όρο $P(i|X) = \frac{1}{Z}\mu_i(X)$, όπου $Z = \sum_i \mu_i(X)$ ώστε να ισχύει $\sum_i P(i|X) = 1$. Δεδομένης μίας παρατήρησης X η προσδοκώμενη τιμή της εξόδου είναι:

$$P(Y|X) = \sum_i P(Y, h_i|X) = \sum_i P(Y|i, X)P(h_i|X) = \frac{\sum_i l_i(X)\mu_i(X)}{\sum_i \mu_i(X)} = F(X) \quad (2.24)$$

Το πλεονέκτημα της ερμηνείας αυτής έναντι του μοντέλου TSK είναι ότι επιτρέπει

την χρήση αποδοτικών αλγόριθμων τύπου Expectation Maximization [52] για την εκμάθηση των παραμέτρων των l_i, μ_i ενώ μπορούμε να εκτιμήσουμε και τις διασπορές των εξόδων κάθε συνάρτησης, οπότε εκτός από τη μέση τιμή της πρόβλεψης για την έξοδο του συστήματος παίρνουμε και μία εκτίμηση του πόσο ακριβής είναι αυτή. Αυτό επιτρέπει για παράδειγμα τον συνδυασμό της εξόδου αυτού του συστήματος με τις εξόδους άλλων ταξινομητών.

Support Vector Machines

Το θεωρητικό υπόβαθρο των δικτύων τύπου SVM [224, 26] στηρίζεται στην εργασία των Vapnik & Chervonenkis βάσει της οποίας μπορούν να εκφραστούν άνω όρια του σφάλματος γενίκευσης μίας συνάρτησης ταξινόμησης. Το πρόβλημα της εκπαίδευσης διατυπώνεται ως η ελαχιστοποίηση ενός όρου που εκφράζει το λάθος εκπαίδευσης και ενός όρου ποινής στην πολυπλοκότητα του ταξινομητή. Η ιδέα αυτή είναι αρκετά παλιά, αλλά βασιζόταν κυρίως σε εμπειρικές θεωρήσεις, ενώ η θεωρία της στατιστικής μάθησης μηχανών επέτρεψε να αποσαφηνιστεί το πώς θα εκφραστούν οι όροι αυτοί.

Συγκεκριμένα, θεωρούμε ότι έχουμε διαθέσιμα N δεδομένα εκπαίδευσης που αποτελούνται από ζεύγη διανυσμάτων εισόδου \mathbf{x}_i και αποφάσεων εξόδου $y_i \in \{-1, 1\}$, $i = 1 \dots N$. Στην απλούστερη περίπτωση, επιθυμούμε να βρούμε ένα υπερεπίπεδο ορισμένο από τη σχέση $y = \mathbf{w}^T \mathbf{x} + b$ τέτοιο ώστε να διαχωρίζει όλα τα δεδομένα. Στην περίπτωση που τα δεδομένα είναι γραμμικά διαχωρίσιμα, αλλάζοντας το $|\mathbf{w}|$ μπορούμε να βρούμε ένα \mathbf{w} τέτοιο ώστε να ισχύει

$$\begin{aligned} \mathbf{w}^T \mathbf{x}_i + b &\geq 1 & \text{αν } y_i &= 1 \\ \mathbf{w}^T \mathbf{x}_i + b &\leq -1 & \text{αν } y_i &= -1 \end{aligned} \quad (2.25)$$

Τα δεδομένα από τις δύο κλάσεις για τα οποία ισχύει η ισότητα αποτελούν τα δυσκολότερα διαχωρίσιμα σημεία, και καθορίζουν το διάστημα (margin) του ταξινομητή, το οποίο ισούται με $2/|\mathbf{w}|$. Από την πλευρά της αρχιτεκτονικής του ταξινομητή, είναι επιθυμητό να έχει ένα μεγάλο διάστημα, καθώς αυτό συνεπάγεται ότι διαχωρίζει τα δεδομένα με μεγάλη ασφάλεια, οπότε μπορεί άμεσα να διατυπωθεί το πρόβλημα της μεγιστοποίησης του διαστήματος υπό τις συνθήκες (2.25). Η ιδέα της μεγιστοποίησης του διαστήματος προέρχεται από την εργασία των Vapnik & Chervonenkis όπου προκύπτει ότι ταξινομητές με μεγάλο margin έχουν και μικρότερα άνω όρια σφάλματος γενίκευσης.

Καθώς τα δεδομένα μας μπορεί να μην είναι διαχωρίσιμα ή μπορεί και να μην επιθυμούμε τον διαχωρισμό τους εις βάρος της απλότητας της συνάρτησης ταξινόμησης, εισάγονται μεταβλητές χαλαρότητας (slack variables) στις συνθήκες (2.25) οπότε προκύπτει το παρακάτω πρόβλημα ελαχιστοποίησης:

$$\begin{aligned} &\text{Ελαχιστοποίησε} && \frac{1}{2} \mathbf{w}^T \mathbf{w} + C \sum_{i=1}^N \xi_i \\ &\text{Υπό τις συνθήκες} && \left\{ \begin{array}{ll} \mathbf{w}^T \mathbf{x}_i + b \geq +1 - \xi_i, & y_i = +1 \\ \mathbf{w}^T \mathbf{x}_i + b \leq -1 + \xi_i, & y_i = -1 \\ \xi_i, \xi_i^* \geq 0 \end{array} \right. \end{aligned}$$

όπου C είναι ένας όρος στάθμισης που εκφράζει πόσο σημαντικά θεωρούμε τα όποια λάθη του ταξινομητή. Το παραπάνω πρόβλημα είναι ένα πρόβλημα τετραγωνικού προγραμματισμού και γνωρίζουμε ότι μπορούμε να βρούμε το καθολικό ελάχιστο

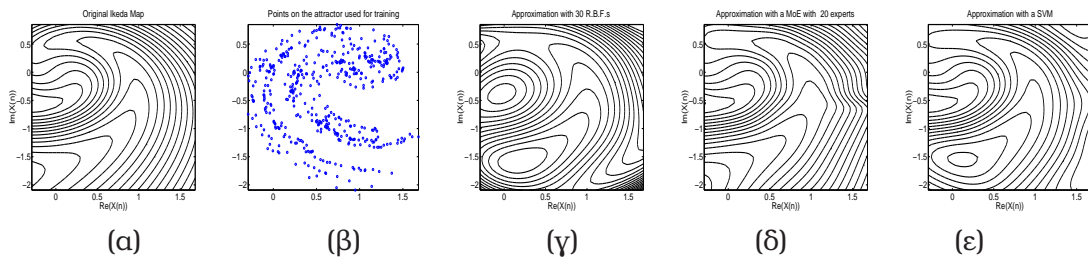
της συνάρτησης κόστους (κάτι που δεν ισχύει για τα νευρωνικά δίκτυα). Το δυικό του προβλήματος αυτού είναι:

$$\begin{aligned} \text{Μεγιστοποίηση} & - \frac{1}{2} \sum_{i,j=1}^N \alpha_i \alpha_j y_i y_j \mathbf{w}_i^T \mathbf{w}_j + \sum_{i=1}^N \alpha_i \\ \text{Υπό τις συνθήκες} & \begin{cases} \sum_{i=1}^N \alpha_i y_i = 0 \\ \alpha_i \in [0, C] \end{cases} \end{aligned} \quad (2.26)$$

Από τις συνθήκες Karush-Kuhn-Tucker προκύπτει ότι για το βέλτιστο διάνυσμα \mathbf{w} ισχύει:

$$\mathbf{w} = \sum_{i=1}^N \alpha_i \mathbf{x}_i \rightarrow f(\mathbf{x}) = \sum_{i=1}^N \alpha_i \mathbf{x}^T \mathbf{x}_i + b \quad (2.27)$$

Παρατηρούμε ότι μπορούμε αντί του εσωτερικού γινομένου $\mathbf{x}^T \mathbf{x}_i$ δύο σημείων να χρησιμοποιήσουμε οποιαδήποτε συνάρτηση έχει ιδιότητες εσωτερικού γινομένου (πυρήνες Mercer). Τη συνθήκη αυτή ικανοποιούν οι περισσότερες συναρτήσεις που χρησιμοποιούνται στην αναγνώριση προτύπων, όπως η σιγμοειδής, η Gaussian σπλήνες κ.α. επιτρέποντας την επαναδιατύπωση του προβλήματος μάθησης για ταξινομητές τύπου RBF ή Multi-Layer Perceptrons στο πλαίσιο των SVMs.



Σχήμα 2.11: (α) Επιπεδοσύνολα του πραγματικού μέρους της απεικόνισης Ickeda (β) Σημεία τα οποία χρησίμευσαν ως σύνολο εκπαίδευσης (γ-ε) Προσέγγισεις μέσω ενός δικτύου RBF, ενός mixture-of-experts και ενός SVM αντίστοιχα.

Στο Σχ. 2.11 οπτικοποιούμε τις δυνατότητες προσέγγισης μίας διδιάστατης συνάρτησης χρησιμοποιώντας αυτά τα μοντέλα [132]: παρατηρούμε ότι τόσο τα δίκτυα MoE όσο και τα SVM δίνουν πολύ καλά αποτελέσματα, για μία αρκετά πολύπλοκη συνάρτηση όπως αυτή της απεικόνισης Ickeda. Το πλεονέκτημα των SVMs είναι ότι δεν χρειάζεται να καθοριστεί εξαρχής το πόσοι όροι χρησιμοποιούνται στο ανάπτυσμά τους, των δε MoE η ταχύτητα με την οποία εκπαιδεύονται. Εκτενέστερα αναφερόμαστε σχετικά στην εργασία [132].

Στην αναγνώριση αντικειμένων στο τέλος της περασμένης δεκαετίας είχε επικρατήσει η άποψη ότι αν έχει κανείς αρκετά δεδομένα εκπαίδευσης και υπολογιστική ισχύ, ένας ταξινομητής τύπου SVM είναι η ενδεικνυόμενη 'λύση', π.χ. [175, 95]. Αν και τα SVMs εξακολουθούν να δίνουν τα καλύτερα αποτελέσματα στην κατηγορία των αλγορίθμων αναγνώρισης προτύπων, προβλήματα όπως η αποδοτικότητα της ανίχνευσης και το μέγεθος του συνόλου εκπαίδευσης έχουν φέρει στο προσκήνιο άλλες τεχνικές της αναγνώρισης προτύπων, όπως τη μέθοδο Boosting [227] ή τα Relevance Vector Machines. Καθώς, όμως, στην πορεία της έρευνας που έγινε στην διάρκεια αυτής της διατριβής διαπιστώθηκε ότι πέρα από την ανίχνευση πρέπει το μοντέλα

για τα αντικείμενα να καθιστούν δυνατή και τη σύνθεσή τους, στραφήκαμε προς τις άλλες δύο κατηγορίες τεχνικών που παρουσιάζονται στη συνέχεια.

2.2.4 Τμηματικοί (Part-Based) Ταξινομητές

Η χρήση τμηματικών ταξινομητών επιτρέπει την απλοποίηση του προβλήματος της ανίχνευσης ενός αντικειμένου ανάγοντάς την στην ανίχνευση απλούστερων τμημάτων του, τα οποία συνδυάζονται για να δώσουν την πιθανότητα της ύπαρξης του αντικειμένου. Πέρα από τη δυνατότητα χρήσης αποτελεσματικών αλγορίθμων για ανίχνευση, αυτή η προσέγγιση εξασφαλίζει επιθυμητές ιδιότητες όπως την ευρωστία ως προς την επικάλυψη, συμπαγή αναπαράσταση των αντικειμένων και τη δυνατότητα αποφυγής της εξαντλητικής αναζήτησης πάνω σε κλίμακες.

Η στατιστική προσέγγιση συνδυάζει την πληροφορία σχετικά με την εμφάνιση με αυτήν για τη γεωμετρία, θεωρώντας τις τοποθεσίες $X = X_{1...N}$ των τμημάτων του αντικειμένου ως τυχαίες μεταβλητές και χρησιμοποιώντας μία από κοινού κατανομή πιθανότητας για τις σχετικές τους θέσεις $P(X)$, σε συνδυασμό με τις επιμέρους κατανομές $P_i(I_i|X_i)$ για την πιθανοφάνεια των παρατηρήσεων δεδομένης της θέσης του τμήματος i . Χρησιμοποιώντας τον τύπο του Bayes έχουμε για την ύστερη πιθανότητα των θέσεων των τμημάτων του αντικειμένου X δεδομένης της εικόνας I

$$P(X|I) = \frac{P(I|X)P(X)}{P(I)} \propto P(I|X)P(X) = \prod_i P_i(I_i|X_i)P(X) \quad (2.28)$$

Στην παραπάνω σχέση ο όρος $P(X)$ εισάγει πληροφορία για τις πιθανές σχετικές τοποθεσίες των τμημάτων, η περιοχή I_i της εικόνας γύρω από τη θέση X_i του χαρακτηριστικού σημείου i μοντελοποιείται μέσω της $P_i(I_i|X_i)$, ενώ υποθέτουμε ότι οι παρατηρήσεις I_i είναι ανεξάρτητες μεταξύ τους, όταν πλέον δεσμεύουμε πάνω στη θέση του τμήματος.

Για να ληφθεί η απόφαση για το εάν υπάρχει ένα αντικείμενο χρειάζεται μία επιπλέον κατανομή που να ερμηνεύει τις παρατηρήσεις στις θέσεις αυτές μέσω εναλλακτικών υποθέσεων· σχηματίζοντας τον λόγο της πιθανοφάνειας των παρατηρήσεων βάσει της υπόθεσης του αντικειμένου και των εναλλακτικών παίρνουμε μία ποσότητα που μπορεί να χρησιμοποιηθεί για την ταξινόμησή του.

Ξεκινώντας από αυτήν την γενική προσέγγιση, διάφορες παραλλαγές της προκύπτουν από τους ορισμούς των τμημάτων των αντικειμένων και από τους αλγόριθμους που χρησιμοποιούνται για να βρεθούν τα μέγιστα των παραπάνω εκφράσεων.

Ορισμός Τμημάτων Αντικειμένου

Η απλούστερη προσέγγιση χρησιμοποιεί ένα προκαθορισμένο σύνολο από τμήματα για το κάθε αντικείμενο, και θεωρεί ότι στο σύνολο εκπαίδευσης έχουν εντοπιστεί χειρωνακτικά οι θέσεις τους στην εικόνα. Αν και αυτή η μέθοδος μας δίνει πλήρη έλεγχο πάνω στην κατασκευή του μοντέλου και συνήθως δίνει και καλύτερα αποτελέσματα [45], είναι δύσκολο να οριστούν και να εντοπιστούν τα τμήματα για όλα τα οπτικά αντικείμενα που αντιλαμβανόμαστε, όπως ένα καλώδιο, ένας τοίχος ή ένα σύννεφο.

Μία εναλλακτική προσέγγιση που απαιτεί ελάχιστη ή και καθόλου επίβλεψη βασίζεται στην χρήση σημείων ενδιαφέροντος, που εξάγονται χρησιμοποιώντας για παράδειγμα έναν τελεστή ανίχνευσης γωνιών [158] ή τις τοποθεσίες των μέγιστων

κάποιας διαφορικής ποσότητας, όπως η Laplacian-of-Gaussian [148, 149]. Έτσι μπορούν να εντοπιστούν χαρακτηριστικά όπως μάτια, ρόδες, πλήκτρα, γωνίες στο μέτωπο κ.ο.κ. χωρίς την ανθρώπινη παρέμβαση.

Η βασική θετική συνιστώσα των προσεγγίσεων αυτών είναι το μικρό πλήθος σημείων που πρέπει να χειριστεί κανείς· αυτό καθιστά δυνατή την εξοντωτική (exhaustive) αναζήτηση πάνω σε κλίμακες για την αυτόματη εκμάθηση των μοντέλων ενώ επιτρέπει την εισαγωγή τεχνικών εκμάθησης μηχανών. Ξεκινώντας από τη μοντελοποίηση των χαρακτηριστικών των τμημάτων μέσω απλών Gaussian [233, 62] συναρτήσεων, τα τελευταία χρόνια έχουν εισαχθεί ιδέες από μεταβολική Bayesian εκτίμηση [62, 57], και από την κατηγοριοποίηση κειμένου [207, 210], καθιστώντας εφικτή την αυτοματοποιημένη κατασκευή μοντέλων για ένα μεγάλο πλήθος κατηγοριών.

Πέρα από αυτό, η ανεξάρτητη από την κλίμακα ανίχνευση αυτών των σημείων μπορεί να αυτοματοποιήσει την εκτίμηση της κλίμακας του αντικειμένου, παρακάμπτοντας την εξαντλητική αναζήτηση πάνω σε κλίμακες που θα χρειαζόταν διαφορετικά.

Μοντελοποίηση Γεωμετρίας

Σε αντίθεση με την μοντελοποίηση της εμφάνισης, όπου όλη η επεξεργασία γίνεται τοπικά, στην περίπτωση αυτή χρειάζεται να ληφθούν υπόψη οι συνδυασμοί των θέσεων των χαρακτηριστικών σημείων του αντικειμένου. Η πολυπλοκότητα του προβλήματος στην περίπτωση που εξετάζουμε όλους τους δυνατούς συνδυασμούς θέσεων είναι μεγάλη, οπότε χρειάζεται κάποια προσέγγιση.

Η απλούστερη είναι αυτή των [27], όπου από όλη την εικόνα απομονώνονται τα σημεία εκείνα για τα οποία τοπικά υπολογίσιμοι όροι πιθανοφάνειας ξεπερνάνε ένα κατώφλι. Στη συνέχεια εκτιμάται η πιθανότητα της συνεύρεσης (constellation) των χαρακτηριστικών σημείων στις συγκεκριμένες θέσεις, για όλους τους δυνατούς συνδυασμούς μεταξύ των χαρακτηριστικών σημείων. Το βασικό μειονέκτημα αυτής της μεθόδου είναι ότι πολλά σημεία τα οποία είναι κάτω από το κατώφλι μπορεί να αντιστοιχούν σε τμήματα του αντικειμένου, οπότε η διαδικασία αυτή είναι δυνατό να οδηγήσει στην πρόωρη απόρριψή του.

Η προσέγγιση των [58] θεωρεί ότι η κατανομή $P(X|O)$ μπορεί να γραφεί σε μία απλούστερη μορφή, η οποία εκμεταλλεύεται τις όποιες ανεξαρτησίες υπάρχουν μεταξύ των σημείων. Για παράδειγμα για το μοντέλο ενός αυτοκινήτου, μπορεί κανείς να υποθέσει ότι η θέση της πίσω ρόδας είναι ανεξάρτητη από τη θέση της μπροστινής, όταν γνωρίζει κανείς που βρίσκεται η πόρτα ενδιάμεσά τους. Η έκφραση τέτοιων ιδιοτήτων γίνεται αβίαστα μέσω γραφικών μοντέλων [182]: οι όποιες εξαρτήσεις υπάρχουν μεταξύ των τυχαίων μεταβλητών (εν προκειμένω οι θέσεις των τμημάτων) εκφράζονται μέσω ενός γράφου, ο οποίος επιτρέπει την απλούστευση των υπολογισμών των στατιστικών ποσοτήτων που μας ενδιαφέρουν. Έτσι είναι δυνατόν με έναν κατανομημένο αλγόριθμο να υπολογιστεί η βέλτιστη τοποθεσία του αντικειμένου χρησιμοποιώντας απλές εκφράσεις· περισσότερες λεπτομέρειες σχετικά θα δοθούν στο επόμενο τμήμα του κεφαλαίου αυτού.

2.2.5 Παραμορφώσιμα Μοντέλα

Το μοντέλα AAM χειρίζονται ξεχωριστά τις χωρικές παραμορφώσεις από τις παραμορφώσεις στην εμφάνιση, και τις συνδυάζουν κατά έναν μη-γραμμικό τρόπο.

Συγκεκριμένα, η εμφάνιση (ύφή) \mathcal{T} του αντικειμένου συντίθεται σε ένα πρωτότυπο σύστημα συντεταγμένων στο οποίο δεν υπεισέρχονται χωρικές αλλοιώσεις ως ένα ανάπτυγμα σε μία γραμμική βάση:

$$\mathcal{T}(\mathbf{x}; \mathbf{t}) = \sum_{i=0}^{N_T} \mathbf{t}_i T_i(\mathbf{x}), \quad (2.29)$$

όπως και στην μέθοδο των eigenfaces [121, 220]. Παραπάνω με $\mathbf{t} = (\mathbf{t}_1, \dots, \mathbf{t}_{N_T})$ συμβολίζουμε το διάνυσμα συντελεστών του αναπτύγματος, T τις συναρτήσεις βάσης, ενώ T_0 είναι η μέση τιμή της φωτεινότητας του αντικειμένου και $\mathbf{t}_0 = 1$. Από το ανάπτυγμά αυτό μπορεί να εκφραστεί η ποικιλία στην εμφάνιση που προκύπτει από τον φωτισμό και το χρώμα του αντικειμένου, π.χ. στα μαλλιά η στα μάτια.

Στη συνέχεια η σύνθεση αυτή παραμορφώνεται χωρικά, ώστε να έρθει σε αντιστοιχία με την παρατήρηση που έχουμε στην εικόνα, χρησιμοποιώντας ένα πεδίο παραμόρφωσης, \mathcal{S} το οποίο πάλι εκφράζεται ως ένα γραμμικό ανάπτυγμα:

$$\mathcal{S}(\mathbf{x}; \mathbf{s}) \equiv (\mathcal{S}_x(\mathbf{x}; \mathbf{s}), \mathcal{S}_y(\mathbf{x}; \mathbf{s})) = \sum_{i=1}^{N_S} \mathbf{s}_i S_i(\mathbf{x}) \quad (2.30)$$

Παραπάνω χρησιμοποιούμε τον συμβολισμό $\mathbf{s} = (\mathbf{s}_1, \dots, \mathbf{s}_{N_S})$, για τους συντελεστές του αναπτύγματος της παραμόρφωσης, ενώ αργότερα χρησιμοποιούμε εναλλακτικά τα πεδία $\mathcal{S} (\mathcal{S}_x, \mathcal{S}_y)$ αναλόγως του ποιο είναι απλούστερο. Το πεδίο αυτό οποίο φέρνει το σημείο $(\mathcal{S}_x(\mathbf{x}; \mathbf{s}), \mathcal{S}_y(\mathbf{x}; \mathbf{s}))$ της εικόνας σε αντιστοιχία με το σημείο του πρωτότυπου αντικειμένου $\mathbf{x} = (x, y)$, οπότε σύμφωνα με το μοντέλο των AAM θεωρούμε ότι $I(\mathcal{S}(\mathbf{x}; \mathbf{s})) \simeq \mathcal{T}(\mathbf{x}; \mathbf{t})$, όπου I είναι η θεωρούμενη εικόνα και το \mathbf{x} ανήκει στο εσωτερικό του αντικειμένου.

Η προσαρμογή των μοντέλων AAM τυπικά επιτυγχάνεται ελαχιστοποιώντας επαναληπτικά ως προς τις παραμέτρους \mathbf{s} και \mathbf{t} ένα κριτήριο που ορίζεται στο σύστημα συντεταγμένων της τεμπλέτας:

$$E(\mathbf{s}, \mathbf{t}) = \sum_{\mathbf{x} \in \mathcal{R}} (I(\mathcal{S}(\mathbf{x}; \mathbf{s})) - \mathcal{T}(\mathbf{x}; \mathbf{t}))^2 \quad (2.31)$$

όπου \mathcal{R} είναι το σύνολο των στοιχείων στο εσωτερικό του αντικειμένου. Αυτό έχει το πλεονέκτημα ότι χρησιμοποιούμε μία στατική βάση για τα στοιχεία εμφάνισης και σχήματος, τα οποία διαφορετική θα έπρεπε σε κάθε επανάληψη να παραμορφώνονται στο σύστημα συντεταγμένων της εικόνας.

Για την ελαχιστοποίηση της παραπάνω ποσότητας τρεις κυρίως προσεγγίσεις έχουν προταθεί:

- Στις αναφορές [226, 112] χρησιμοποιείται ένας αλγόριθμος στοχαστικής βελτιστοποίησης, ο οποίος αποτιμά την παράγωγο της συνάρτησης κόστους χρησιμοποιώντας σε κάθε επανάληψη ένα υποσύνολο των παρατηρήσεων και χρησιμοποιεί gradient descent.
- Η επίσπευση της διαδικασίας gradient descent επιτυγχάνεται στην αναφορά [37] μέσω μίας συνάρτησης παλινδρόμησης που συνδέει το σφάλμα σύνθεσης με τις βέλτιστες ανανεώσεις των παραμέτρων· αυτή η συνάρτηση μαθαίνεται κατά τη διάρκεια της εκπαίδευσης και μπορεί να αποτιμηθεί αποδοτικά κατά την προσαρμογή σε μία νέα εικόνα.

- Τέλος, στην αναφορά [157] χρησιμοποιείται μία διατύπωση του προβλήματος της προσαρμογής που επιτρέπει την χρήση του αλγορίθμου Newton-Raphson για την εκτίμηση των παραμέτρων χωρίς να απαιτεί τον επαναυπολογισμό της Hessian της συνάρτησης κόστους σε κάθε βήμα.

Η ύπαρξη αποδοτικών αλγορίθμων για την προσαρμογή τους καθιστά δυνατή τη χρήση τους σε πραγματικό χρόνο [157], παρακάμπτοντας το βασικό εμπόδιο στην εφαρμογή τους.

Περίπλοκότερα μοντέλα όπως τα 3-Δ παραμορφώσιμα μοντέλα των Blanz & Vetter [16] έχουν δώσει τα καλύτερα αποτελέσματα των τελευταίων ετών στην αναγνώριση προσώπων, ενώ πρόσφατα έχουν προταθεί επεκτάσεις των μοντέλων αυτών σε δυσκολότερα προβλήματα, όπως η μοντελοποίηση αντικειμένων με πολλά επίπεδα [111].

2.3 Πιθανοτικά Μοντέλα για Προβλήματα Ανάλυσης και Αναγνώρισης Προτύπων

Αν και η ποικιλία μεθόδων που υπάρχουν στην όραση υπολογιστών έχει ωφελήσει το πεδίο εισάγοντας ιδέες και τεχνικές από διαφορετικά ώριμα πεδία, καθιστά δύσκολη την επικοινωνία και την από κοινού χρήση ιδεών σε διαφορετικά προβλήματα.

Κατά τη διάρκεια της τελευταίας δεκαετίας όμως είναι αισθητή μία στροφή ανάλογη της τεχνολογίας αναγνώρισης φωνής προς την στατιστική διατύπωση και ενοποίηση προβλημάτων της όρασης και στην χρήση τεχνικών στατιστικής εκμάθησης μηχανών, τόσο σε χαμηλού/μέσου επιπέδου προβλήματα, όπως ομαλοποίηση εικόνας, εκτίμηση κίνησης/βάθους, π.χ. [67, 70, 71] όσο και σε υψηλού επιπέδου προβλήματα, όπως tracking και αναγνώριση [13, 209, 210].

Στη βάση της διατριβής αυτής βρίσκεται η χρήση των αναγεννητικών μοντέλων, δηλαδή πιθανοτικών μοντέλων με ένα μικρό σύνολο από παραμέτρους που μπορούν να χρησιμοποιηθούν για να ανακατασκευάσουν τις παρατηρήσεις της εικόνας και των γραφικών μοντέλων, όπου οι εξαρτήσεις μεταξύ των τυχαίων μεταβλητών που περιλαμβάνονται σε ένα μοντέλο μπορούν να αναπαρασταθούν μέσω ενός γράφου. Σημειώνουμε ότι συνηθίζεται ως αναγεννητικά μοντέλα να αναφέρονται όλα τα μοντέλα που μπορούν να παράσχουν μία έκφραση της πιθανοφάνειας των παρατηρήσεων, οπότε περιλαμβάνονται και τα περισσότερα γραφικά μοντέλα. Για την παρουσίαση αυτής της διατριβής χρησιμοποιούμε τη διάκριση της αναφοράς [249], που βοηθάει κυρίως στην παρουσίασή τους, τονίζοντας ότι τα αναγεννητικά μοντέλα, όπως τα εννοούμε εδώ, έχουν μία τριτομμένη γραφική δομή και τα γραφικά μοντέλα συνήθως πολύ απλά αναγεννητικά μοντέλα στα δυναμικά παρατήρησή τους.

Τα υβρίδια μεταξύ τους προκύπτουν αβίαστα, ενώ συνηθίζεται με τον όρο αναγεννητικά μοντέλα να εννοούνται και οι δύο προσεγγίσεις, έχοντας κατά νου το γενικότερο πλαίσιο της θεωρίας προτύπων.

Στη συνέχεια παρουσιάζονται στοιχεία των δύο πεδίων που θα βοηθήσουν στην ενιαία παρουσίαση των συνεισφορών που έγιναν στα πλαίσια της διατριβής αυτής: εκτενέστερες παρουσιάσεις περιλαμβάνονται στις αναφορές [78, 167, 90, 113, 249, 71].

2.3.1 Αναγεννητικά Μοντέλα

Τα αναγεννητικά μοντέλα για προβλήματα της όρασης υπολογιστών ταιριάζουν στο γενικότερο πλαίσιο της Bayesian αναγνώρισης προτύπων. Η προσέγγιση συνίσταται στην κατασκευή για την κάθε μοντελοποιούμενη κατηγορία εικόνων ενός μοντέλου σύνθεσης με ένα μικρό πλήθος παραμέτρων το οποίο μπορεί, να εκφράζει την ποικιλία των παρατηρήσεων που προέρχονται από την κλάση και να παρέχει μία έκφραση για την πιθανοφάνεια μίας παρατηρούμενης εικόνας δεδομένης της σύνθεσης του μοντέλου, βάσει ενός μοντέλου σφάλματος.

Ήδη συναντήσαμε δύο από τις πλέον συνηθισμένες σχέσεις σύνθεσης

- Για το εσωτερικό μίας περιοχής στο πλαίσιο της κατάτμησης:

$$I(x) = c, \quad \mathcal{A} = c \quad (2.32)$$

- Για την προσέγγιση μίας εικόνας μέσω ενός γραμμικού συνδυασμού στοιχείων βάσης, όπως π.χ. στην ανάλυση Fourier, ή στις μεθόδους Principal/Independent Components Analysis:

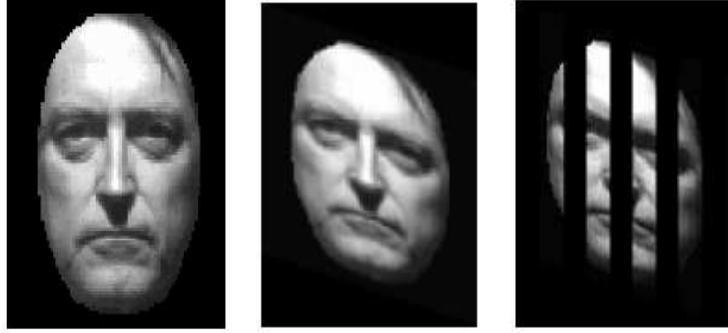
$$I(x) = \sum_{k=1}^K c_k B_k(x), \quad \mathcal{A} = [c_1, \dots, c_K] \quad (2.33)$$

Σύμφωνα με την αναφορά [166] ανάμεσα στα αντικείμενα και την εμφάνισή τους που αποτυπώνεται σε μία εικόνα παρεμβάλλονται γενικότερα τέσσερις κύριοι τύποι παραμορφώσεων, που πρέπει να μοντελοποιηθούν στα πλαίσια ενός αναγεννητικού μοντέλου:

- Θόρυβος κατά τη λήψη της εικόνας.
- Υπέρθωση με άλλα σήματα.
- Χωρική παραμόρφωση.
- Επικαλύψεις με άλλα αντικείμενα.

Οι πρώτοι δύο τύποι παραμορφώσεις οδηγούν στη χρήση κάποιου αναπτύγματος τύπου PCA και στην εισαγωγή ενός μοντέλου σφάλματος, τυπικά τύπου λευκού Gaussian θορύβου (White Gaussian Noise- WGN). Οι χωρικές παραμορφώσεις μπορούν είτε να μοντελοποιηθούν μέσω ενός πεδίου παραμόρφωσης (deformation field), όπως στα μοντέλα AAM είτε μέσω ενός γραφικού δικτύου, όπως στα τμηματικά μοντέλα. Τέλος, για την μοντελοποίηση της διαδικασίας επικάλυψης για την πλήρης λύση θα χρειαζόταν να συνδυαστούν περισσότερα του ενός αντικείμενα, να υπολογιστεί η προβολή τους για την (άγνωστη) θέση της κάμερας κ.ο.κ. Αντί αυτού, μπορούμε να δούμε την κατάτμηση όπως εφαρμόζεται στο κεφάλαιο 5 ως μία παράκαμψη αυτής της διαδικασίας, όπου απλά αποφασίζεται ποια μέρη του αντικειμένου είναι ορατά, χωρίς να χρειάζεται μία ερμηνεία όλης της σκηνής.

Η αναγνώριση προτύπων μέσω αναγεννητικών μοντέλων βασίζεται στον νόμο του Bayes που επιλέγει την κλάση με την μέγιστη-εκ-των-υστερών (Maximum-a-Posteriori - MAP) πιθανότητα. Συγκεκριμένα, αν υποθέσουμε ότι έχουμε K εναλλακτικές τάξεις $\mathcal{C} = \{c_1, \dots, c_i, \dots, c_K\}$ και μία παρατήρηση I , το μοντέλο i χρησιμοποιεί ένα σύνολο παραμέτρων \mathcal{A}_i στην *σχέση σύνθεσης* του $I_i(\cdot; \mathcal{A}_i)$ για να προσεγγίσει τις παρατηρήσεις. Θεωρώντας τις παραμέτρους του μοντέλου αυτές άγνωστες



Σχήμα 2.12: Από [246]: Πέρα από τις τυπικά θεωρούμενες πηγές αλλοίωσης (φωτισμός και θόρυβος) στην δημιουργία μίας εικόνας από ένα αντικείμενο υπεισέρχονται και άλλες πηγές παραμόρφωσης, όπως η χωρική παραμόρφωση και οι επικαλύψεις.

αποκτούμε την μαρτυρία μοντέλου (model evidence) ολοκληρώνοντας πάνω σε αυτές:

$$P(I|c_i) = \int_{\mathcal{A}_i} P(I|\mathcal{A}_i, c_i)P(\mathcal{A}_i|c_i)d\mathcal{A}_i \quad (2.34)$$

όπου $P(\mathcal{A}_i)$ είναι η πρότερη κατανομή των παραμέτρων του μοντέλου. Παραπάνω χρησιμοποιήθηκε η σχέση $\int_b P(a, b|c) = P(a|c)$ σε συνδυασμό με τον κανόνα αλυσίδας $P(a, b|c) = P(a|b, c)P(b|c)$.

Συνήθως το παραπάνω ολοκλήρωμα πάνω στις παραμέτρους παρακάμπτεται, χρησιμοποιώντας απευθείας μία σημειακή εκτιμήτρια, θεωρώντας ότι η μάζα της ολοκληρούμενης ποσότητας είναι προσεγγιστικά μια συνάρτηση Dirac γύρω από αυτήν. Οι δύο συνηθέστερες εκτιμήτριες είναι οι εκτιμήτριες Μεγίστης Πιθανοφάνειας (Maximum-Likelihood - ML) και Μέγιστου-εκ-των-Υστέρων (Maximum-a-Posteriori - MAP)

$$\begin{aligned} \mathcal{A}_i^{ML} &= \arg \max_{\mathcal{A}_i} P(I|\mathcal{A}_i, c_i) \\ \mathcal{A}_i^{MAP} &= \arg \max_{\mathcal{A}_i} P(I|\mathcal{A}_i, c_i)P(\mathcal{A}_i|c_i) \end{aligned} \quad (2.35)$$

Για παράδειγμα, για την περίπτωση που έχουμε ένα ανάπτυγμα πάνω σε μία βάση, σε συνδυασμό με την υπόθεση του λευκού θορύβου μπορούμε να γράψουμε

$$\log P(I; \mathcal{A}) = -\frac{\sum_{x \in P} \left[\sum_{k=1}^K \mathcal{A}_k B_k(x) - I(x) \right]^2}{2\sigma^2} + \frac{|P| \log 2\pi\sigma^2}{2},$$

όπου P είναι το σύνολο των παρατηρήσεων στην εικόνα που μοντελοποιείται και το $|\cdot|$ συμβολίζει την πληθικότητα. Ο λογάριθμος της πιθανοφάνειας των δεδομένων είναι τότε ανάλογος του αθροίσματος του τετραγώνου των λαθών, οπότε η εκτίμηση των παραμέτρων μέγιστης πιθανοφάνειας ανάγεται στην αντιστροφή του πίνακα που προκύπτει από το κριτηρίου Ελάχιστων Τετραγώνων. Ας σημειωθεί εδώ ότι αυτός είναι και ο βασικός λόγος που εισάγεται η παραδοχή του WGN· καθώς πρόκειται για σφάλμα μοντελοποίησης, πρακτικά το σφάλμα ανακατασκευής είναι συσχετισμένο, αλλά λαμβάνοντας αυτό υπόψη οδηγούμαστε σε περίπλοκες σχέσης εκτίμησης παραμέτρων.

Εισάγοντας τις εκτιμήτριες αυτές στον τύπο του Bayes μπορούμε να πάρουμε προσεγγίσεις των ύστερων πιθανοτήτων των κλάσεων, βάσει των οποίων μπορούμε να κάνουμε ταξινόμηση χρησιμοποιώντας ένα Τεστ Λόγου Γενικευμένης Πιθανοφάνειας (Generalized Likelihood Ratio Test (GLRT)) [119]:

$$P(c_i|I) = \frac{P(c_i)P(I|c_i)}{\sum_{c_j \in \mathcal{C}} P(c_j)P(I|c_j)} \simeq \frac{P(I|\hat{\mathcal{A}}_i, c_i)}{\sum_{c_j \in \mathcal{C}} P(I|\hat{\mathcal{A}}_j, c_j)} \quad (2.36)$$

Κρυφές μεταβλητές

Καθώς συχνά στην διαδικασία παραγωγής των παρατηρήσεων υπεισέρχονται παράγοντες οι οποίοι δεν καταγράφονται μαζί με τις παρατηρήσεις, εκφράζοντας την πιθανοφάνεια των παρατηρήσεων βάσει των σχέσεων σύνθεσης συχνά εισάγονται άγνωστες μεταβλητές που πρέπει να ληφθούν υπόψη κατά την εκτίμηση των παραμέτρων και την χρήση των μοντέλων. Για παράδειγμα όταν έχουμε μετρήσεις φωτεινότητας και δεν γνωρίζουμε της πηγές φωτισμού, οι θέσεις τους είναι κρυφές μεταβλητές οι οποίες πρέπει να απαλειφθούν όταν κάνουμε φωτομετρική στερέωση. Αντίστροφα, όταν θέλουμε να εκτιμήσουμε την θέση των πηγών φωτισμού το τριδιάστατο σχήμα των αντικειμένων είναι μία άγνωστη ποσότητα που πρέπει να λάβουμε υπόψη κατά τη διάρκεια της συμπερασματολογίας.

Μία άλλη περίπτωση όπου οι κρυφές μεταβλητές προκύπτουν άμεσα είναι στην εκμάθηση των βάσεων που χρησιμοποιούνται για τα αναγεννητικά μοντέλα. Στην περίπτωση αυτή οι παρατηρήσεις είναι το σύνολο εκπαίδευσης, οι παράμετροι που θέλουμε να εκτιμήσουμε τα στοιχεία βάσης και οι κρυφές μεταβλητές οι συντελεστές στα αναπτύγματα. Τότε, η εκτίμηση παραμέτρων ισοδυναμεί με την χρήση μίας τεχνικής όπως η PCA, ICA, TCA ενώ οι συντελεστές του αναπτύγματος πλέον θεωρούνται ως περιττές τυχαίες μεταβλητές (nuisance terms). Τέλος, όπως θα δούμε και στο κεφάλαιο 5 μπορούμε να δούμε και την κατάτμηση της εικόνας ως ένα πεδίο κρυφών μεταβλητών, που πρέπει να απαλειφεί από τη διαδικασία εκτίμησης παραμέτρων των αναγεννητικών μοντέλων.

Η μέθοδος αντιμετώπισης των κρυφών μεταβλητών είναι η απαλοιφή τους μέσω ολοκλήρωσης (integrate-out): έχοντας ένα πιθανοτικό μοντέλο $P(I, Z, \mathcal{A})$ που συνοψίζει τη διαδικασία σύνθεσης των παρατηρήσεων I μέσω του συνόλου παραμέτρων \mathcal{A} που θέλουμε να εκτιμήσουμε και ενός συνόλου κρυφών μεταβλητών Z που παρεμβάλλονται στη διαδικασία σύνθεσης, έχουμε

$$P(\mathcal{A}|I) \propto P(I|\mathcal{A})P(\mathcal{A}) = \int_Z P(I, Z|\mathcal{A})P(\mathcal{A}) \quad (2.37)$$

Καθώς το ολοκλήρωμα αυτό μπορεί να είναι δύσκολο να υπολογιστεί, προσεγγιστικές μέθοδοι μεγιστοποίησης του όρου $P(I|\mathcal{A})$ συχνά χρησιμοποιούνται, όπως για παράδειγμα ο αλγόριθμος Μεγιστοποίησης Προσδοκίας Expectation Maximization-EM, ο οποίος χρησιμοποιεί ιδιότητες της συνάρτησης του λογαρίθμου.

Περιορισμοί αναγεννητικών μοντέλων

Τα αναγεννητικά μοντέλα για προβλήματα όρασης είναι διαισθητικά ελκυστικά, καθώς χρησιμοποιώντας απλά παραμετρικά μοντέλα μπορούν να ερμηνεύσουν ενός μεγάλο πλήθος παρατηρήσεων. Ένας βασικός περιορισμός τους είναι ότι δεν εκ-

μεταλλεύονται τη δομή των αλληλεξαρτήσεων μεταξύ των τυχαίων μεταβλητών για αποδοτικότερους αλγόριθμους και πιο συμπαγή μοντέλα. Για παράδειγμα, μοντελοποιώντας ένα πρόσωπο, ένα καθολικό μοντέλο μπορεί να χρησιμοποιεί έναν επαναληπτικό αλγόριθμο προσαρμογής μοντέλου βασισμένο σε όλες τις παρατηρήσεις. Θα ήταν επιθυμητό αν η διαδικασία προσαρμογής έχει καταλήξει σε μία καλή ανακατασκευή μίας περιοχής του προσώπου, να 'παγώσει' στην περιοχή αυτή και να εστιάσει στις υπόλοιπες περιοχές. Καθώς όμως τα αναγεννητικά μοντέλα τύπου AAM είναι καθολικά, δεν μπορούν να συσχετίσουν τις παραμέτρους τους με συγκεκριμένες περιοχές, οπότε και να εκμεταλλευτούν στην εκτίμηση των παραμέτρων την όποια ανεξαρτησία μπορεί να υπάρχει μεταξύ τους.

Πέρα από αυτό η αυτόματη εκμάθηση αναγεννητικών μοντέλων από δεδομένα είναι συχνά προβληματική, καθώς εμπλέκονται οι παράμετροί τους ως κρυφές μεταβλητές, οπότε αυξάνοντας την εκφραστικότητά τους (μέσω περισσότερων παραμέτρων), μεγαλώνει και η πιθανότητα η εκμάθηση να σταματήσει σε μία τοπικά βέλτιστη λύση.

2.3.2 Γραφικά Μοντέλα

Στα περισσότερα προβλήματα όρασης υπολογιστών χρειάζεται να εκμεταλλευτούμε τις όποιες ανεξαρτησίες υπάρχουν μεταξύ των τυχαίων μεταβλητών για την κατασκευή απλούστερων και αποδοτικότερων μοντέλων. Η έκφραση τέτοιων ιδιοτήτων γίνεται αβίαστα μέσω γραφικών μοντέλων [182, 240, 113, 44, 114], καθώς επιτρέπουν την μοντελοποίηση και χρήση κατανομών πάνω σε ένα μεγάλο πλήθος τυχαίων μεταβλητών, εκμεταλλεύόμενα τις όποιες ανεξαρτησίες υπάρχουν μεταξύ τους. Για την αναπαράσταση της από κοινού κατανομής τους χρησιμοποιείται η δομή ενός γράφου, οι κόμβοι του οποίου είναι οι τυχαίες μεταβλητές και οι ακμές του εκφράζουν τις εξαρτήσεις μεταξύ τους. Έτσι μπορούν να κατασκευαστούν αλγόριθμοι συμπερασματολογίας που εκμεταλλεύονται τις όποιες ανεξαρτησίες βάσει της δομής του και όχι από τις όποιες ευρετικές εφευρεθούν για το εκάστοτε πρόβλημα.

Συγκεκριμένα, ας θεωρήσουμε ότι οι αλληλεξαρτήσεις των τυχαίων μεταβλητών που εμπλέκονται στο πρόβλημά μας μπορούν να μοντελοποιηθούν μέσω του μη-κατευθυντικού γράφου στο Σχ. 2.13 και ότι θέλουμε να εκτιμήσουμε την περιθωριακή κατανομή της μεταβλητής X_1 . Από τη δομή του γράφου μπορούμε να γράψουμε την έκφραση της από κοινού κατανομής του διανύσματος τυχαίων μεταβλητών $X = \{X_1, X_2, X_3, X_4, X_5\}$:

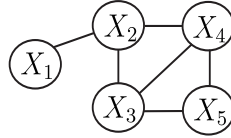
$$P(X) = \prod_{C \in \mathcal{C}} P(X_C) = P(X_1, X_2)P(X_2, X_3, X_4)P(X_2, X_5)P(X_3, X_5) \quad (2.38)$$

όπου \mathcal{C} είναι το σύνολο των μέγιστων κλικών του γράφου, και $P(X_C)$ είναι θετικές συναρτήσεις κανονικοποιημένες ώστε $\sum_X \prod_{C \in \mathcal{C}} P(X_C) = 1$.

Η απευθείας χρήση του τύπου της περιθωριοποίησης απαιτεί τον υπολογισμό της έκφρασης

$$P(X_1) = \sum_{X_2, X_3, X_4, X_5} P(X_1, X_2, X_3, X_4, X_5) \quad (2.39)$$

το οποίο αν για παράδειγμα έχουμε διακριτές μεταβλητές με N δυνατές τιμές είναι πολυπλοκότητας $O(N^4)$. Εκμεταλλεύόμενοι όμως τη συγκεκριμένη μορφή του γι-



Σχήμα 2.13: Γραφικό δίκτυο για την αναπαράσταση των αλληλεξαρτήσεων της κατανομής (2.38).

νομένου, που προκύπτει από τη δομή του γράφου, μπορούμε να χρησιμοποιήσουμε την αντιμεταθετική ιδιότητα και να γράψουμε

$$\begin{aligned}
 P(X_1) &= \sum_{X_2} P(X_1, X_2) \sum_{X_3, X_4} P(X_2, X_3, X_4) \underbrace{\sum_{X_5} P(X_2, X_3, X_5)}_{\mu_5(X_2, X_3)} \\
 &= \sum_{X_2} P(X_1, X_2) \underbrace{\sum_{X_3, X_4} P(X_2, X_3, X_4)}_{\mu_{3,4}(X_2)} \mu_5(X_2, X_3) \\
 &= \sum_{X_2} P(X_1, X_2) \mu_{3,4}(X_2) \\
 &= \mu(X_1)
 \end{aligned} \tag{2.40}$$

όπου πλέον το μεγαλύτερο άθροισμα είναι πολυπλοκότητας $O(N^2)$. Αντίστοιχα αν είχαμε συνεχείς τυχαίες μεταβλητές θα είχαμε να υπολογίσουμε πλέον ένα διπλό ολοκλήρωμα αντί ενός τετραπλού. Αυτή η ιδέα βρίσκεται πίσω από καθιερωμένες τεχνικές όπως τα φίλτρα Kalman στην θεωρία ελέγχου τα Hidden Markov Models - HMMs στην επεξεργασία φωνής, τα Belief Networks στα έμπειρα συστήματα και τα δίκτυα τύπου Markov Random Field σε προβλήματα όρασης.

Μία συνηθισμένη απλούστευση είναι να τίθενται όλοι οι όροι που αντιστοιχούν σε κλίκες με περισσότερους από δύο όρους ίσοι με την μονάδα, οπότε απομένουν μόνο τα λεγόμενα δυναμικά ζεύγους (pairwise potentials), που συμβολίζονται με $\Psi(X_i, X_j)$. Επίσης, για τις παρατηρήσεις I_i που είναι διαθέσιμες εισάγεται ο όρος του δυναμικού παρατήρησης $\Phi_i(X_i) = P(I_i|X_i)$, όπου X_i είναι η κρυφή τυχαία μεταβλητή που ερμηνεύει την παρατήρηση I_i , μέσω ενός αναγεννητικού υπομοντέλου. Για παράδειγμα για το φίλτρο Kalman τα ζεύγη δυναμικού εκφράζουν τους περιορισμούς που εισάγει η δυναμική του συστήματος και τα δυναμικά παρατηρήσεων το θόρυβο που υπεισέρχεται στις μετρήσεις. Για ένα πρόβλημα αναγνώρισης αντικειμένων τα δυναμικά ζεύγους εκφράζουν τους γεωμετρικούς περιορισμούς στους οποίους υπόκεινται τα τμήματα του αντικειμένου και τα δυναμικά παρατήρησης την πιθανοφάνεια του τμήματος της εικόνας που μοντελοποιείται από το τμήμα του μοντέλου.

Συνδυαζόμενοι αυτοί οι όροι μας δίνουν την έκφραση πιθανοφάνειας για το σύνολο των παρατηρήσεων I βάσει του γραφικού δικτύου του μοντέλου του αντικειμένου O :

$$P(I|O) = \sum_X P(I, X|O) = \sum_X P(I|X, O)P(X|O) = \sum_X \prod_i \Phi_i(X_i) \prod_{\{i,j\} \in \mathcal{E}} \Psi(X_i, X_j) \tag{2.41}$$

Έτσι μπορούν περίπλοκες δομές στις παρατηρήσεις μας να ερμηνευτούν μέσω απλών αναγεννητικών μοντέλων (όροι $\Phi_i(X_i) = P(I_i|X_i)$) και ενός συνόλου κρυφών τυχαίων μεταβλητών X που έχουν μία συγκεκριμένη δομή αλληλεξαρτήσεων. Τα προβλήματα που τίθενται είναι αφενός η εκτίμηση της ύστερης κατανομής πάνω στις τυχαίες μεταβλητές X δεδομένης μίας παρατήρησης I , (συμπερασματολογία) και η εκμάθηση των δυναμικών παρατήρησης και ζεύγους χρησιμοποιώντας δεδομένα εκπαίδευσης (εκτίμηση παραμέτρων και δομής). Παρακάτω δίνουμε μία σύντομη παρουσίαση των αλγορίθμων συμπερασματολογίας για γραφικά δίκτυα καθώς εμφανίζονται σε διάφορα σημεία του κειμένου, ενώ παραπέμπουμε στις αναφορές [113, 188, 251] για θέματα εκμάθησης παραμέτρων σε γραφικά δίκτυα.

Αλγόριθμοι συμπερασματολογίας για γραφικά δίκτυα

Για το παράδειγμα στο Σχ. 2.13 είδαμε ένα παράδειγμα όπου χρησιμοποιώντας τη δομή του γράφου επιταχύνεται ο υπολογισμός περιθωριακών κατανομών τυχαίων μεταβλητών. Γενικότερα, για γράφους χωρίς βρόχους μπορεί να χρησιμοποιηθεί ένας κατανεμημένος αλγόριθμος από τον οποίο υπολογίζονται συγχρόνως οι ύστερες κατανομές όλων των μεταβλητών, χωρίς να γίνεται η διαδικασία απαλοιφής για κάθε μία ξεχωριστά.

Συγκεκριμένα, στον αλγόριθμο Belief Propagation/Sum Product [182] κάθε κόμβος i στέλνει μηνύματα στους γείτονές του $\{\mathcal{N}(i)\}$, στα οποία χρησιμοποιεί

- Τα μηνύματα που έχει λάβει από όλους τους υπόλοιπους γείτονές του, $\{\mathcal{N}(i) \setminus j\}$,
- Το δυναμικό παρατήρησης, Φ_i
- Το δυναμικό ζεύγους $\Psi_{i,j}$

Το μήνυμα από τον κόμβο i στον κόμβο j γράφεται:

$$m_{i \rightarrow j}(X_j) = \sum_{X_i} \Phi_i(X_i) \prod_{k \in \{\mathcal{N}(i) \setminus j\}} m_{k \rightarrow i}(X_i) \quad (2.42)$$

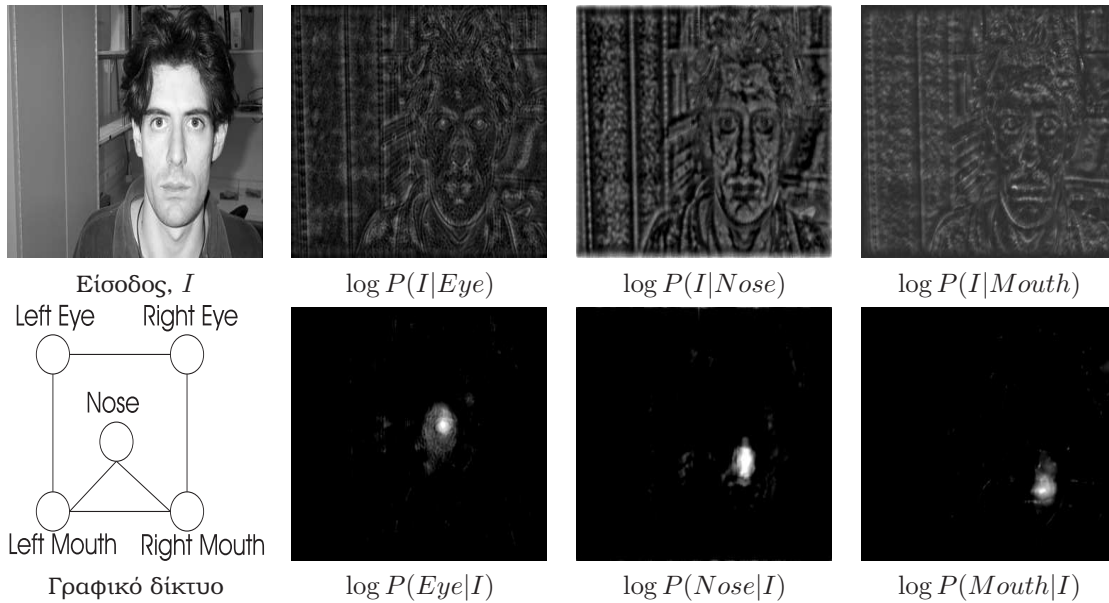
ενώ τα beliefs στον κόμβο i εκτιμώνται ως το γινόμενο των εισερχόμενων μηνυμάτων με τη συνάρτηση δυναμικού Φ_i :

$$B_i(X_i) = \Phi_i(X_i) \prod_{j \in \mathcal{N}(i)} m_{i \rightarrow j}(X_i) \quad (2.43)$$

Διαισθητικά, με το μήνυμα $m_{i \rightarrow j}$ ο κόμβος i περιορίζει τις πιθανές τιμές του κόμβου j συνδυάζοντας την αντίστοιχη πληροφορία που έχει πάρει από τους υπόλοιπους γειτονικούς κόμβους με τους δικούς του περιορισμούς για τις σχετικές τους θέσεις.

Όταν η διαδικασία μεταβίβασης μηνυμάτων έχει συγκλίνει, μπορεί να αποδειχθεί ότι εάν το γραφικό μοντέλο έχει τη δομή ενός δέντρου, η ποσότητα $B_i(X_i)$ θα ισούται πραγματικά με $P(X_i|I)$.

Ο αλγόριθμος αυτός έχει εμφανιστεί ως Belief Propagation στα Bayesian δίκτυα [182] forwards-backwards στα HMMs, και smoothing στα φίλτρα Kalman [177]. Στο άρθρο [134] ενοποιούνται οι παραπάνω διατυπώσεις ενώ εδώ χρησιμοποιούμε τη διατύπωση της αναφοράς [240]. Ένα παράδειγμα της εφαρμογής αυτού του αλγορίθμου στην ανίχνευση προσώπου φαίνεται στο Σχ. 2.14. Η ανίχνευση σημείων



Σχήμα 2.14: Ανίχνευση προσώπου χρησιμοποιώντας γραφικά μοντέλα: Πάνω γραμμή: Εικόνα στην είσοδο του προσώπου, και τοπικά καθοριζόμενα δυναμικά παρατήρησης $\log \Phi_i(X_i)$, όπου X_i η θέση του δεξιού ματιού, της μύτης και του δεξιού χείλους, αντίστοιχα. Κάτω γραμμή: Γραφικό δίκτυο για την αναπαράσταση και ανίχνευση προσώπου, και αποτελέσματα αλγορίθμου Belief Propagation για τις ύστερες κατανομές των θέσεων των σημείων, $\log B_i(X_i)$.

με βάση τοπικά χαρακτηριστικά δίνει πολλές εσφαλμένες ανιχνεύσεις, όπως φαίνεται στην πάνω γραμμή· εισάγοντας όμως τις αλληλεξαρτήσεις μεταξύ των σημείων του αντικειμένου γίνονται εντονότερες οι πραγματικές τοποθεσίες τους.

Γενικότερα μπορεί να μην υπάρχει μία πορεία απαλοιφής των τυχαίων μεταβλητών όπως αυτή του παραδείγματος που να εξασφαλίζει ότι το άθροισμα γίνεται πάνω σε ένα μικρό πλήθος τυχαίων μεταβλητών. Στην περίπτωση αυτή υπάρχουν δύο προσεγγίσεις: η στοχαστική και η μεταβολική.

Στην στοχαστική προσέγγιση οι στατιστικές ποσότητες που μας ενδιαφέρουν προσεγγίζονται με την τεχνική Monte Carlo· η μέση τιμή μίας συνάρτησης f των μεταβλητών X μπορεί να προσεγγιστεί μέσω N δειγμάτων της κατανομής ως:

$$\int_X f(X)P(X)dX \simeq \frac{1}{N} \sum_{i=1}^N f(X_i), \quad X_i \sim P(X) \quad (2.44)$$

Οι ιδιότητες των γραφικών μοντέλων εδώ είναι χρήσιμες για την εξαγωγή δειγμάτων από την κατανομή, καθώς επιτρέπουν την χρήση τεχνικών δειγματοληψίας Gibbs [75] και Metropolis για την κατασκευή μίας αλυσίδας Markov, όπου το προηγούμενο δείγμα παράγει έναν νέο χρησιμοποιώντας μόνο τοπική πληροφορία που καθορίζεται από τη δομή του δικτύου. Αυτό το οποίο εγγυώνται οι αλγόριθμοι Gibbs και Metropolis είναι ότι τα δείγματα αυτά όντως ακολουθούν την κατανομή $P(X)$ οπότε μπορεί να χρησιμοποιηθούν για εκτίμηση Monte-Carlo. Ο συνδυασμός αυτός είναι ευρύτερα γνωστός ως Markov Chain Monte Carlo [169], και χαρακτηρίζεται αφενός από την ευελιξία του, καθώς επιτρέπει την δειγματοληψία περίπλοκων κατανομών, π.χ. [251], αφετέρου από την βραδύτητά του.

Η μεταβολική προσέγγιση στη συμπερασματολογία [107, 228] υποθέτει ότι οι

κατανομές που θέλουμε να εκτιμήσουμε ανήκουν σε μία συγκεκριμένη οικογένεια \mathcal{Q} , και ανάγει την συμπερασματολογία στην αναζήτηση μίας κατανομής $Q \in \mathcal{Q}$ που να μεγιστοποιεί ένα κατάλληλο κριτήριο $J(Q)$. Ένα κατάλληλο μεταβολικό κριτήριο για να μεγιστοποιούμε για τον υπολογισμό της $P(X|Y)$ είναι η έκφραση:

$$J(Q) = \log P(Y) - KL(Q(X)||P(X|Y)) \quad (2.45)$$

όπου $KL(p(X)||q(X)) = -\sum_X p(X) \log \frac{p(X)}{q(X)}$ είναι η απόσταση Kullback-Leibler [43] μεταξύ των κατανομών p και q που ελαχιστοποιείται όταν $p(X) = q(X)$, οπότε και ισούται με το μηδεν. Συνεπώς το κριτήριο (2.45) μεγιστοποιείται όταν $Q(X) = P(X|Y)$ οπότε και ισούται με $\log P(Y)$.

Αν και η $P(X|Y)$ μπορεί να μην ανήκει στην οικογένεια \mathcal{Q} , το κριτήριο $J(Q)$ μας παρέχει ένα μέσο για την επιλογή ενός μέλους της \mathcal{Q} που να είναι κοντά στην $P(X|Y)$. Χρησιμοποιώντας τις ταυτότητες $P(Y)P(X|Y) = P(X, Y)$ και $\sum_X Q(X) = 1$, η σχέση (2.45) γράφεται ως:

$$\begin{aligned} J(Q) &= \sum_X Q(X) \log P(Y) - \sum_X Q(X) \log \frac{Q(X)}{P(X|Y)} \\ &= \sum_X Q(X) \log P(X, Y) - \sum_X Q(X) \log Q(X) = \\ &= \sum_X Q(X) \log P(X, Y) + S \end{aligned} \quad (2.46)$$

όπου η ποσότητα $S = -\sum_X Q(X) \log Q(X)$ ισούται με την εντροπία της κατανομής Q . Γράφοντας την κατανομή $P(X, Y)$ ως ένα γινόμενο πάνω στις κλίκες του γράφου, $P(X, Y) = \prod_i \Phi_i(X_i) \prod_{\{i,j\} \in \mathcal{E}} \Psi(X_i, X_j)$, όπου $\Phi_i(X_i) = P(Y_i|X_i)$ έχουμε

$$J(Q) = \sum_i \sum_{X_i} Q(X_i) \log \Phi_i(X_i) + \sum_{\{i,j\} \in \mathcal{E}} Q(X_i, X_j) \log \Psi(X_i, X_j) + S \quad (2.47)$$

όπου τα αθροίσματα πάνω στο διάνυσμα X έχει συρρικνωθεί σε επιμέρους αθροίσματα πάνω σε μικρότερους όρους.

Ο όρος που δυσχεραίνει την απευθείας ελαχιστοποίηση αυτού του όρου είναι η εντροπία της κατανομής, S , η οποία και απλοποιείται με τις μεταβολικές τεχνικές. Για παράδειγμα, η προσέγγιση μέσου πεδίου (Mean Field Approximation -MFA) θεωρεί ότι η από κοινού κατανομή του συνόλου των τυχαίων μεταβλητών $X = \{X_1, \dots, X_n\}$ μπορεί να προσεγγιστεί μέσω των κατανομών Q_i , $i = 1, \dots, N$ που ορίζονται πάνω σε ανεξάρτητες τυχαίες μεταβλητές:

$$P(X) \simeq Q(X) = \prod_i Q_i(X_i) \quad (2.48)$$

Στην περίπτωση αυτή ο όρος της εντροπίας γράφεται ως $S_{MF} = \sum_i S_i$, $S_i = -\sum_{X_i} Q_i(X_i) \ln(Q_i(X_i))$, καθώς η εντροπία ενός συνόλου ανεξάρτητων τυχαίων μεταβλητών ισούται με το άθροισμα της εντροπίας των μεμονομένων μεταβλητών. Πλέον όλοι οι όροι περιέχουν είτε μεμονωμένες μεταβλητές ή ζεύγη τους, οπότε μπορούμε να χρησιμοποιήσουμε έναν επαναληπτικό αλγόριθμο ο οποίος ανανεώνοντας τις τιμές των Q_i αυξάνει σταθερά το κριτήριο και η διαδικασία συγκλίνει σε ένα (τοπικό) μέγιστο του J .

Μία άλλη προσέγγιση χρησιμοποιεί την έκφραση για την εντροπία που θα προέκυπτε αν ο γράφος είχε τη δομή ενός δέντρου· στην περίπτωση αυτή η κατανομή πάνω στο γράφο θα μπορούσε να γραφεί ως

$$Q(X) = \frac{\prod_{i,j} Q(X_i, X_j)}{\prod_i Q(X_i)^{(q_i-1)} \quad (2.49)$$

όπου q_i είναι το πλήθος των γειτόνων του κόμβου i . Από την έκφραση αυτή προκύπτει η προσέγγιση Bethe της εντροπίας

$$S_{Bethe} = - \sum_{i,j} \sum_{X_i, X_j} Q(X_i, X_j) \log Q(X_i, X_j) + \sum_i (q_i - 1) \sum_{X_i} Q(X_i) \log Q(X_i) \quad (2.50)$$

Σημειώνουμε εδώ ότι η έκφραση (2.49) εν γένει δεν αντιστοιχεί σε κάποια πραγματική δομή γραφικού δικτύου, καθώς όταν υπάρχουν βρόχοι στο δίκτυο οδηγεί σε ασυνέπειες, μη ικανοποιώντας τις σχέσεις περιθωριοποίησης που περιλαμβάνουν παραπάνω από δύο όρους. Αντίθετα, η MFA που παρουσιάστηκε προηγουμένως αντιστοιχεί σε ένα συγκεκριμένο γραφικό δίκτυο όπου δεν υπάρχουν καθόλου ακμές μεταξύ των τυχαίων μεταβλητών. Συνεπώς δεν μπορούμε να ερμηνεύσουμε τον όρο που προκύπτει ελαχιστοποιώντας ως προς $Q(X)$ ως ένα κάτω όριο του κριτηρίου, καθώς η τιμή του δεν προκύπτει χρησιμοποιώντας ένα υποσύνολο των δυνατών κατανομών.

Μία ειδική περίπτωση είναι όταν ο γράφος του δικτύου έχει όντως τη δομή ενός δέντρου. Τότε πλέον η προσέγγιση (2.50) στην εντροπία γίνεται ακριβής, οπότε ελαχιστοποιώντας το κριτήριο που προκύπτει παίρνουμε τη σωστή κατανομή. Την ίδια κατανομή παρέχει όμως και ο αλγόριθμος Belief Propagation, οπότε πρέπει να υπάρχει κάποια σχέση μεταξύ τους. Ξεκινώντας από αυτήν την παρατήρηση αποδείχτηκε [240] ότι ο αλγόριθμος Belief Propagation μπορεί να ερμηνευθεί ως μία μέθοδος ελαχιστοποίησης της προσέγγισης Bethe στην ελεύθερη ενέργεια, δικαιολογώντας έτσι και θεωρητικά τα καλά αποτελέσματα που δίνει ο αλγόριθμος όταν εφαρμόζεται σε γραφικά μοντέλα με κύκλους (Loopy Belief Propagation). Η θεώρηση αυτή έχει επιτρέψει αφενός τη γενίκευση του αλγορίθμου χρησιμοποιώντας υψηλότερης πολυπλοκότητας όρους [240], αφετέρου την εισαγωγή εναλλακτικών [244] που έχουν εγγυημένη σύγκλιση.

2.3.3 Σχέσεις γραφικών και αναγεννητικών μοντέλων

Γενικότερα, τα αναγεννητικά μοντέλα χαρακτηρίζονται αφενός από την απλότητα με την οποία μπορεί να παραχθεί και να ερμηνευθεί ένα δείγμα τους, αφετέρου από την περιττή πολυπλοκότητα που εισάγει στους αλγόριθμους εκμάθησης και προσαρμογής μοντέλων η καθολική τους φύση. Από την άλλη τα γραφικά μοντέλα εκφράζουν συνοπτικά τις αλληλεξαρτήσεις μεταξύ των τυχαίων μεταβλητών που μοντελοποιούν, αλλά η σύνθεση δειγμάτων και η εκμάθηση παραμέτρων είναι χρονοβόρες ή γίνονται μέσω προσεγγίσεων.

Ένα παράδειγμα που αναδεικνύει τις διαφορές μεταξύ γραφικών και αναγεννητικών μοντέλων είναι η μοντελοποίηση του θορύβου Brown [167, 249]. Για μία διακριτοποιημένη στοχαστική ανέλιξη θορύβου Brown οι διαφορές μεταξύ των γειτονικών παρατηρήσεων είναι Gaussian τυχαίες μεταβλητές με μηδενική μέση τιμή και διασπορά σ . Συνεπώς μπορούμε να εκφράσουμε την πιθανοφάνεια του συνόλου

παρατηρήσεων ως ένα γινόμενο πάνω σε δυαδικά δυναμικά :

$$P(X) = \prod_{(i,j) \in \mathcal{C}} P(X_i, X_j) = \prod_{(i,j) \in \mathcal{C}} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(X_i - X_j)^2}{\sigma^2}\right)$$

$$\log P(x) \propto \sum_{(i,j) \in \mathcal{C}} (X_i - X_j)^2 \quad (2.51)$$

που μας δίνει το συνηθισμένο πρότερο μοντέλο για προβλήματα όρασης υπολογιστών χαμηλού/μέσου επιπέδου, την τετραγωνική νόρμα πάνω στις παραγώγους του σήματος. Εναλλακτικά, μπορούμε να χρησιμοποιήσουμε το γεγονός ότι το φάσμα του σήματος ακολουθεί τον κανόνα $1/f$ [216] που σημαίνει ότι μπορούμε να γράψουμε μία σχέση σύνθεσης χρησιμοποιώντας τον αντίστροφο μετασχηματισμό Fourier ως ένα αναγεννητικό μοντέλο [167].

$$X(x, y) = \sum_k \sum_m \frac{1}{\sqrt{2c(k^2 + m^2)}} c_{k,m} \cos(2\pi(kx + my) + \phi_{k,m})$$

$$c_{k,m} \sim N(0, 1), \quad \phi_{k,m} \sim U[0, 1] \quad (2.52)$$

όπου N είναι η Gaussian κατανομή και U η ομοιόμορφη. Αν θεωρήσουμε ότι λείπει η παρατήρηση j και θέλουμε να υπολογίσουμε την ύστερη κατανομή της βάσει των υπόλοιπων παρατηρήσεων, από τη σχέση (2.51) προκύπτει ότι χρειαζόμαστε μόνο τις γειτονικές παρατηρήσεις $X_i, i \in \mathcal{N}(j)$. Αντίθετα, η σχέση (2.52) απαιτεί να εκτιμήσουμε τις παραμέτρους $c_{k,m}$ του αναπτύγματος που εμπλέκουν όλες τις υπόλοιπες παρατηρήσεις.

Αν τώρα λείπει μία περιοχή της εικόνας και θέλουμε να συνθέσουμε το εσωτερικό της, οι παράμετροι του αναπτύγματος $c_{k,m}$ αρκούν για να μας δώσουν μία σύνθεση, ενώ χρησιμοποιώντας ένα γραφικό δίκτυο θα έπρεπε να καταφύγουμε στη χρήση ενός αλγόριθμου δειγματοληψίας.

Όπως προαναφέρθηκε, η διάκριση μεταξύ γραφικών δικτύων και αναγεννητικών μοντέλων δεν είναι αυστηρή και χρησιμοποιήθηκε κυρίως για να μπορέσουμε στη συνέχεια να ξεκαθαρίσουμε τη φύση είναι τα προβλήματα που αντιμετωπίζονται σε κάθε κεφάλαιο. Έτσι για παράδειγμα υπάρχουν γραφικά μοντέλα όπου στα δυναμικά παρατηρήσεων χρησιμοποιούνται αναγεννητικά μοντέλα για ένα υποσύνολο των παρατηρήσεων [249, 88], η αναγεννητικά μοντέλα σε ένα υψηλότερο επίπεδο από αυτό των κόμβων των γραφικών δικτύων, που ερμηνεύουν τις αλληλεξαρτήσεις τους χρησιμοποιώντας ένα παραμετρικό μοντέλο, όπως η Ανάλυση Παραγόντων [138].

Έχοντας αυτό κατά νου, μπορούμε να κάνουμε την παρακάτω συνοπτική ταξινόμηση των περιεχομένων κάθε κεφαλαίου σε μία η περισσότερες από τις πτυχές που θίχτηκαν σύμφωνα με τον πίνακα 2.1· αν και υπάρχουν επιμέρους πτυχές τους, όπως η επεξεργασία εικόνας, τα βιολογικά μοντέλα όρασης, η πολυκλιμακωτή ανάλυση εικόνων, και άλλες που δεν μπορούν να καταταγούν σε μία από αυτές τις περιοχές, το ενοποιητικό πλαίσιο για όλη την εργασία αυτή ήταν η αναγωγή των προβλημάτων σε ένα πιθανοτικό πλαίσιο.

	Προσέγγιση	Πτυχές Προβλήματος
Κεφάλαιο 3	Γραφικό Μοντέλο	Εκμάθηση παραμέτρων (IIS) Εκτίμηση κρυφών παραμέτρων (MFA)
Κεφάλαιο 4	Αναγεννητικό Μοντέλο	Εκτίμηση Παραμέτρων (WLSE) Ταξινόμηση
Κεφάλαιο 5	Αναγεννητικό Μοντέλο	Εκτίμηση Παραμέτρων (EM) Ταξινόμηση Εκμάθηση μοντέλου (AAM)
Κεφάλαιο 6	Γραφικό Μοντέλο	Εκτίμηση κρυφών παραμέτρων (Monte Carlo) Εκμάθηση Δομής
	Αναγεννητικό Μοντέλο	Αποτίμηση Πιθανοφάνειας

Πίνακας 2.1: Ταξινόμηση προβλημάτων που θίγονται στα επόμενα κεφάλαια της διατριβής.

Κεφάλαιο 3

Ανάλυση και Εκμάθηση ενός Βιολογικά Εμπνευσμένου Μοντέλου Χαμηλού και Μέσου Επιπέδου Λειτουργιών Όρασης

Το βιολογικό σύστημα όρασης αποτελεί μία πλούσια πηγή έμπνευσης για αλγόριθμους όρασης υπολογιστών. Αν και η δομή του είναι περίπλοκη, ορισμένες απλουστεύσεις επιτρέπουν την κατασκευή μοντέλων που καλύπτουν ένα μεγάλο μέρος των πρώτων στάδιων της λειτουργίας του. Χρησιμοποιώντας τις σχέσεις που υπάρχουν μεταξύ προβλημάτων της όρασης υπολογιστών, τη θεωρία νευρωνικού υπολογισμού και του βιολογικού συστήματος όρασης, στο κεφάλαιο αυτό παρουσιάζεται μία υπολογιστική και στατιστική ανάλυση ενός μοντέλου για χαμηλού και μέσου επιπέδου λειτουργίες της όρασης. Χρησιμοποιήθηκαν ιεραρχικά μοντέλα με αναδρομικές συνδέσεις σε κάθε επίπεδο, μέσω των οποίων καθίσταται δυνατή και μία χαμηλότερου επιπέδου διατύπωση και εφαρμογή της συνέργειας, όπου η κάτωθεν μαρτυρία περί ύπαρξης ακμών συνδυάζεται με άνωθεν πληροφορία σχετικά με την ομαλότητα και συνέχειά τους. Συγκεκριμένα, εστίασαμε στο Boundary Contour System / Feature Contour System - BCS / FCS του S. Grossberg και των συνεργατών του [85, 83, 84, 86] που σε μεγάλο βαθμό βρίσκεται πίσω από σύγχρονες νευρωνικές αρχιτεκτονικές για προβλήματα όρασης [242, 140, 172].

Η φύση αυτού του κεφαλαίου είναι συνθετική, οπότε ως έναν βαθμό παρουσιάζει πτυχές των διαφορετικών περιοχών που συνδυάστηκαν. Οι καινοτομικές συνεισφορές συνοψίζονται στα παρακάτω σημεία :

- Διατηρώντας ακέραιες τις λειτουργίες του αρχικού συστήματος BCS/FCS προτείνεται ένα απλούστερο μοντέλο το οποίο προσφέρεται σε μία μαθηματικά σαφή ανάλυση.
- Βάσει της χρήσης συναρτήσεων Lyapunov για την ανάλυση της λειτουργίας αναδρομικών νευρωνικών δικτύων διατυπώνεται η λειτουργία του κύριου τμήματος του δικτύου ως η βελτιστοποίηση ενός μεταβολικού κριτηρίου.
- Μέσω της μεταβολικής προσέγγισης στη συμπερασματολογία παρέχεται μία στατιστική ερμηνεία της λειτουργίας του, στην οποία βασίζεται η διατύπωση ενός αλγόριθμου για την εκμάθηση των βαρών του δικτύου.
- Αποτιμάται συστηματικά η επίδοση του δικτύου στην ανίχνευση ακμών χρησιμοποιώντας περίπλοκες φυσικές εικόνες.

3.1 Συστήματα Βασισμένα στο Βιολογικό Σύστημα Όρασης

Στην ενότητα αυτή σε συντομία παρουσιάζονται οι βασικές έννοιες πίσω από τη διατύπωση του μοντέλου μας, και οι κύριες αναφορές στις οποίες βασίστηκε αυτή η εργασία. Καθώς σε ορισμένα σημεία του κειμένου χρησιμοποιούνται όροι από ετερόκλητες περιοχές όπως η θεωρία νευρωνικού υπολογισμού (Neural Computation)[96, 50], η νευροφυσιολογία [123] και η επιστήμη της όρασης [101] που δεν είναι δυνατόν να παρουσιαστούν στα πλαίσια του κειμένου της διατριβής, παραπέμπουμε στις παραπάνω αναφορές καθώς και στην αναφορά [127] όπου έχουμε επιχειρήσει να παρουσιαστεί η εργασία αυτή σε ένα αυτόνομο κείμενο.

3.1.1 Στοιχεία από το Βιολογικό Σύστημα Όρασης

Εν συντομία το βιολογικό σύστημα όρασης χρησιμοποιεί μία ιεραρχική αρχιτεκτονική, όπου πρώιμες διαδικασίες εξαγωγής χαρακτηριστικών λαμβάνουν χώρα στον αμφιβληστροειδή, στον Lateral Geniculate Nucleus - LGN και στην περιοχή V1, και στη συνέχεια περιπλοκότερες λειτουργίες επιτελούνται στις περιοχές V1, V2, V4 και IT. Αν και στο γενικότερο πλαίσιο της συνέργειας αναζητούμε τις αλληλεπιδράσεις των λειτουργιών που εκτελούνται σε όλες τις περιοχές, στο κεφάλαιο αυτό εστιάζουμε στις περιοχές V1 και V2, που τυπικά θεωρούνται ότι πραγματοποιούν την ένωση περιγραμμάτων.

Τύποι νευρώνων στις περιοχές V1, V2

Στην περιοχή V1 απαντούν κυρίως ` απλά ` και ` περίπλοκα ` κύτταρα (simple / complex cells). Τα πρώτα αποκρίνονται σε συγκεκριμένες περιοχές της εικόνας, εστιάζοντας σε συγκεκριμένες συχνότητες και κατευθύνσεις, ενώ τα δεκτικά τους πεδία (receptive fields) μπορούν να μοντελοποιηθούν ικανοποιητικά μέσω 2-διάστατων συναρτήσεων Gabor [49, 50]. Συγκεκριμένα, ένα μιγαδικό φίλτρο Gabor έχει μία κρουστική απόκριση της μορφής

$$g(x, y) = \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right) \exp(ju_c x + jv_c y), \quad (3.1)$$

που ισούται με μία Gaussian η οποία διαμορφώνεται από ένα ημιτονοειδές. Η διασπορά σ της Gaussian καθορίζει την χωρική του επικέντρωση ενώ οι γωνιακές συχνότητες u_c, v_c την κατεύθυνση και κλίμακα των σημάτων στα οποία αποκρίνεται εντονότερα. Το πραγματικό μέρος αυτής της έκφρασης μας δίνει έναν άρτιο πυρήνα, ενώ το φανταστικό της μέρος έναν περιττό, που αντιστοιχούν και στα ονομαζόμενα άρτια και περιττά απλά κύτταρα (even- / odd- symmetric simple cells).

Τα περίπλοκα κύτταρα συνδυάζουν τις αποκρίσεις απλών κυττάρων παρέχοντας μία ανεξάρτητη φάση απόκριση που μπορεί να εξαχθεί υπολογιστικά παίρνοντας το μέτρο απόκρισης του μιγαδικού φίλτρου Gabor, ή γενικότερα με τη χρήση ενός Ορθογωνικού Ζεύγους Φίλτρων (Quadrature Filter Pair-QFP) [2, 94, 50]. Περιπλοκότερα μοντέλα της λειτουργίας αυτών των δύο κατηγοριών κυττάρων [29, 30, 77] λαμβάνουν υπόψη τις οριζόντιες αλληλεπιδράσεις μεταξύ τους, όπου πλέον μοντελοποιούνται ως μονάδες ενός συνόλου με το οποίο αλληλεπιδρούν δυναμικά.

Στην εργασία [97] ανακαλύφθηκαν νευρώνες στην περιοχή V2 που είναι δραστήριοι ακόμα και όταν δεν έχουν κάποιο ερέθισμα στο δεκτικό τους πεδίο, εάν οι γειτονικοί τους νευρώνες είναι ενεργοί. Η συνηθισμένη ερμηνεία της λειτουργίας αυτών των κυττάρων είναι ότι ενισχύουν τις αποκρίσεις ομάδων από νευρώνες που βρίσκονται σε αμοιβαία 'συνεπείς' τοποθεσίες, δηλαδή πάνω σε ομαλές καμπύλες με έντονη αντίθεση (salient). Στην αναφορά [64] περιγράφονται ψυχοφυσικά δεδομένα που αναλύουν βαθύτερα τη λειτουργία τους, προτείνοντάς τα ως έναν νευρωνικό μηχανισμό για λειτουργίες ομαδοποίησης.

Μοντελοποίηση Λειτουργίας νευρώνων

Αν και υπάρχουν μοντέλα της λειτουργίας μεμονωμένων νευρώνων μεγάλης πολυπλοκότητας και ακρίβειας [123], όταν πλέον μοντελοποιούνται ολόκληρες περιοχές του οπτικού φλοιού το μεγάλο πλήθος των νευρώνων απαιτεί τη χρήση χονδρικών προσεγγίσεων. Η προσέγγιση του Μέσου Ρυθμού Εκπυροσκόρησης (Mean Firing Rate -MFR) συνοψίζει τη συμπεριφορά ενός νευρώνα μέσω του μέσου ρυθμού εκπομπής παλμών (spikes) [123, 50]. Βάσει αυτής της προσέγγισης οι αλληλεπιδράσεις ενός δικτύου νευρώνων μπορούν να περιγραφούν μέσω διαφορικών εξισώσεων· η απλούστερη τέτοια εξίσωση είναι του προσθετικού μοντέλου, όπου ο MFR U_i του νευρώνα i επηρεάζεται από τους MFR U_j των γειτόνων του $j \in \mathcal{N}(i)$ ως:

$$\begin{aligned} \frac{dV_i}{dt} &= -aV_i + \sum_{j \in \mathcal{N}(i)} w_{i,j} U_j \\ U_i &= g(V_i) \end{aligned} \quad (3.2)$$

Το δυναμικό V_i του νευρώνα i σχετίζεται με το MFR U_i μέσω μίας μη-γραμμικής συνάρτησης g που διατηρεί την ποσότητα U_i θετική, ενώ συνήθως φτάνει σε κορεσμό για υψηλές τιμές του V . Στην περίπτωση που ο νευρώνας i επηρεάζει τους ρυθμούς εκπυροσκόρησης και των γειτόνων του, οδηγούμαστε σε μία αναδρομική αρχιτεκτονική, που μπορεί να έχει μη τετριμμένη δυναμική [96, 127].

Ένα περιπλοκότερο μοντέλο αλληλεπιδράσεων που χρησιμοποιήθηκε από τον S. Grossberg αντικαθιστά τη σχέση (3.2) με:

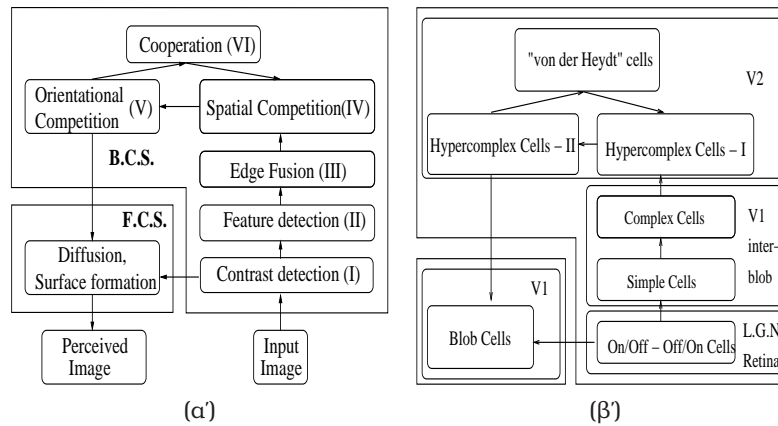
$$\begin{aligned} \frac{dV_i}{dt} &= -AV_i + (V_{max} - V_i)E_i + (V_{min} - V_i)I_i \\ \text{όπου } E_i &= \sum_{j \in \mathcal{N}(i)} w_j^E U_j, \quad I_i = \sum_{j \in \mathcal{N}(i)} w_j^I U_j \end{aligned} \quad (3.3)$$

Σύμφωνα με αυτή τη σχέση το δυναμικό V στη σταθερή του κατάσταση παραμένει φραγμένο στην περιοχή $[V_{min}, V_{max}]$ και δεν φτάνει σε κορεσμό για οσοδήποτε υψηλά ερεθιστικές (excitatory)- E ή ανασταλτικές (inhibitory)- I εισόδους:

$$\frac{dV_i}{dt} = 0 \rightarrow V_i = \frac{V_{max}E_i + V_{min}I_i}{A + E_i + I_i} \quad (3.4)$$

Εισάγοντας έναν όρο απωθητικής αναστολής (shunting inhibition) $-V_i G_{sh}$ στην σχέση (3.3),

$$\frac{dV_i}{dt} = -AV_i + (V_{max} - V_i)E_i + (V_{min} - V_i)I_i - V_i G_{sh}, \quad (3.5)$$



Σχήμα 3.1: (α) Διάγραμμα της αρχιτεκτονικής BCS/FCS και (β) αντίστοιχες περιοχές του οπτικού συστήματος.

μπορεί υπολογιστικά να επιτευχθεί η διαιρετική κανονικοποίηση· οι αλληλεπιδράσεις μεταξύ νευρώνων στις περιοχές V1, V2 έχουν μοντελοποιηθεί βάσει αυτής της σχέσης [29, 30, 205], παρέχοντας ακριβή αποτελέσματα και προβλέψεις.

3.1.2 Νευρωνικές Αρχιτεκτονικές για προβλήματα Όρασης

Προβλήματα που αφορούν τη πολυπλοκότητα, ακρίβεια και αποδεκτότητα των μοντέλων προκύπτουν όταν διατυπώνονται μοντέλα για ολόκληρες περιοχές του οπτικού φλοιού αντί για μεμονωμένους νευρώνες. Στην αναζήτηση τέτοιων μοντέλων τα ψυχοφυσικά φαινόμενα υποδεικνύουν τη λειτουργία των υπολογιστικών μονάδων ενώ η αρχιτεκτονική του συστήματος όρασης περιορίζει τα μοντέλα που μπορούν να προταθούν για την ερμηνεία των φαινομένων.

Καθώς η γνώση αυτή είναι μερική, για τα περισσότερα προβλήματα έχουν προταθεί διαφορετικά μοντέλα· για παράδειγμα η αντίληψη ψευδών περιγραμμάτων (illusory contours) έχει μοντελοποιηθεί στη αναφορά [146] ως μία διαδικασία που παραμένει στο επίπεδο του V1, στην αναφορά [172] μέσω ενός πολλαπλασιαστικού σήματος gating από την περιοχή V2 και στις εργασίες [82, 84] ως ένας προσθετικός όρος. Γενικότερα, λίγα μοντέλα έρχονται με δεδομένα φυσιολογίας που να στηρίζουν τη δομή τους, ενώ το κύριο μέλημα είναι συνήθως να αναπαραχθούν ψυχοφυσικά δεδομένα. Στην συνέχεια θα παρουσιάσουμε συνοπτικά το μοντέλο BCS / FCS το οποίο αποτέλεσε το σημείο εκκίνησης για την έρευνά μας στην περιοχή αυτή· εναλλακτικά μοντέλα [140, 146] που σχετίζονται με την εργασία αυτή παρουσιάζονται στις αναφορές [127, 125].

Το μοντέλο BCS - FCS

Το μοντέλο FACADE (Form And Colour and DEpth) που προτάθηκε από τον S. Grossberg και τους συνεργάτες του [83, 84, 86] ασχολείται σχεδόν με όλα τα προβλήματα χαμηλού και μέσου επιπέδου όρασης, ξεκινώντας από την ανίχνευση ακμών και καταλήγοντας στην στερέωση. Εμείς εστίασαμε στα υποσυστήματα BCS και FCS που σχετίζονται με τις λειτουργίες της ανίχνευσης περιγραμμάτων και κατάτμησης εικόνων.

Συνοπτικά η λειτουργία του BCS σχετίζεται με την αντίληψη ορίων, ξεκινώντας από την ανίχνευση χαρακτηριστικών και καταλήγοντας σε έναν μηχανισμό εκτίμησης

saliency ενώ το FCS αντιστοιχεί σε λειτουργίες που έχουν σχέση με περιοχές της εικόνας.

Τα στάδια του BCS είναι τα ακόλουθα (βλ. Σχ. 3.1):

Στάδιο I: Ανίχνευση αντίθεσης, όπου μοντελοποιείται η συμπεριφορά των κυττάρων On / Off που αποκρίνονται σε κηλίδες (blobs) της εικόνας, μέσω διαφορών από Gaussian συναρτήσεις διαφορετικής κλίμακας.

Στάδιο II: Ανίχνευση Στοιχειωδών Χαρακτηριστικών όπου μοντελοποιείται η συμπεριφορά των απλών κυττάρων, συνελίσσοντας την εικόνα με επιμηκυμένες Gaussian συναρτήσεις.

Στάδιο III: Ολοκλήρωση πηγών πληροφορίας, όπου θεωρείται ότι τα περίπλοκα κύτταρα (complex cells) συνδυάζουν πληροφορία από απλά κύτταρα, την άνωθεν πληροφορία από τη διαδικασία ομαδοποίησης περιγραμμάτων και το παράλληλο σύστημα FCS.

Στάδια IV -V: διαδικασίες ανταγωνισμού. Εδώ πλέον μία διαδικασία επιλογής μέγιστων δίνει έναν λεπτό, επικεντρωμένο πίνακα ακμών από τις ευρείες αποκρίσεις των προηγούμενων σταδίων.

Στάδιο VI: Ένωση ακμών, Ανίχνευση saliency Στο στάδιο αυτό κάθε νευρώνας συγκεντρώνει πληροφορία από τη γειτονιά του σχετικά με την πιθανότητα μία ακμή να διέρχεται από την περιοχή του, χρησιμοποιώντας μία δομή συνδέσεων της μορφής του Σχ. 3.5. Η έξοδος του σταδίου αυτού επιστρέφει στο στάδιο IV, δίνοντας έτσι μία αναδρομική διαδικασία που ανιχνεύει και ενισχύει ομαλές καμπύλες.

Ερχόμενοι στο σύστημα FCS, η λειτουργία του είναι η δημιουργία συνεχών εικόνων φωτεινότητας από τις παραγωγούς της που παρέχει το στάδιο I, οδηγώντας έτσι στην αντίληψη ομαλών επιφανειών. Η κατασκευή συνεχών επιφανειών επιτυγχάνεται μέσω μίας διαδικασίας που διαχέει τη δραστηριότητα των νευρώνων με εξαίρεση τις περιοχές όπου ένα σήμα από το BCS καταδεικνύει την παρουσία μίας ακμής, διατηρώντας έτσι διακριτές τις δραστηριότητες των νευρώνων εκατέρωθέν της.

3.2 Ένα Υπολογιστικά Εύχρηστο και Βιολογικά Αποδεκτό Μοντέλο για Χαμηλού και Μέσου Επιπέδου Λειτουργίες της Όρασης

Το μοντέλο FACADE μπορεί να ερμηνεύσει ικανοποιητικά πολλά ψυχοφυσικά φαινόμενα [84], ενώ οι περισσότερες ιδέες του είναι απλές και διαισθητικά ικανοποιητικές· ως σύνολο όμως το σύστημα είναι εξαιρετικά περίπλοκο σε ότι αφορά την ανάλυσή και υλοποίησή του, ενώ οι βελτιώσεις στις επιδόσεις που υπόσχεται δεν έχουν αποτιμηθεί σε σχέση με κλασσικούς αλγορίθμους όρασης υπολογιστών.

Για το λόγο αυτό εισάγουμε στην ενότητα αυτή ένα απλούστερο μοντέλο, το οποίο εκτελεί όλες τις παραπάνω λειτουργίες, ενώ προσφέρεται σε μία σαφή μαθηματική ανάλυση. Πέρα από το θέμα της απλότητας, οι περισσότερες αλλαγές μας βασίστηκαν σε ορισμένες καθιερωμένες προσεγγίσεις της όρασης υπολογιστών, που μπορούν να συνοψιστούν στα ακόλουθα σημεία :

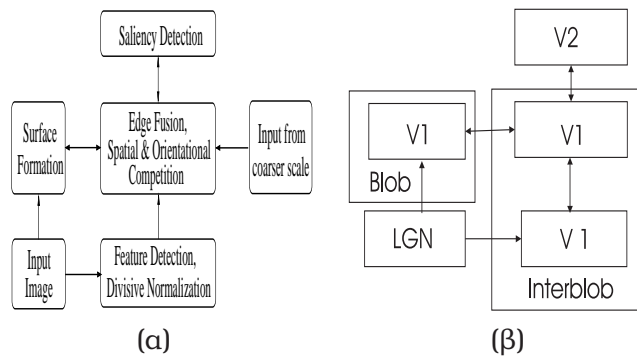
- Η λέπτυνση ακμών είναι μία εγγενώς αναδρομική διαδικασία. Ο όρος ανάδρασης από το τελευταίο στάδιο δεν βοηθάει στην λέπτυνση ακμών, αν και οδηγεί στην δημιουργία νέων· η ανάδραση θα έπρεπε να υπάρχει μέσα σε κάθε επίπεδο, χρησιμοποιώντας οριζόντιες συνδέσεις μεταξύ των νευρώνων.

- Το FCS θα έπρεπε να αλληλεπιδρά με το BCS, χρησιμοποιώντας πληροφορία σχετικά με τις περιοχές για να οδηγήσει τη δημιουργία ακμών.
- Στην αρχική αρχιτεκτονική [82, 86] τα αποτελέσματα από διαφορετικές κλίμακες απλά προστίθενται, ενώ γενικότερα στην όραση υπολογιστών χρησιμοποιούνται τα αποτελέσματα από χονδρικές κλίμακες για να οδηγήσουν αυτά σε λεπτότερες.

Η αρχιτεκτονική του μοντέλου που προτείνουμε και η αντιστοίχισή της με περιοχές του συστήματος όρασης φαίνονται στο Σχ. 3.2.

Οι λειτουργίες που εκτελούν τα τμήματα του μοντέλου μας είναι

- Στάδιο I: Εξαγωγή χαρακτηριστικών σχετικών με την ύπαρξη ακμών.
- Στάδιο II: Συνδυασμός διαφορετικών πηγών πληροφορίας, λήψη απόφασης για ύπαρξη ακμών.
- Στάδιο III: Εκτίμηση ποιότητας και σημαντικότητας (saliency) ακμών.
- Σύστημα Επιφάνειας: Ομαλοποίηση εικόνας χρησιμοποιώντας ανισοτροπική διάχυση.



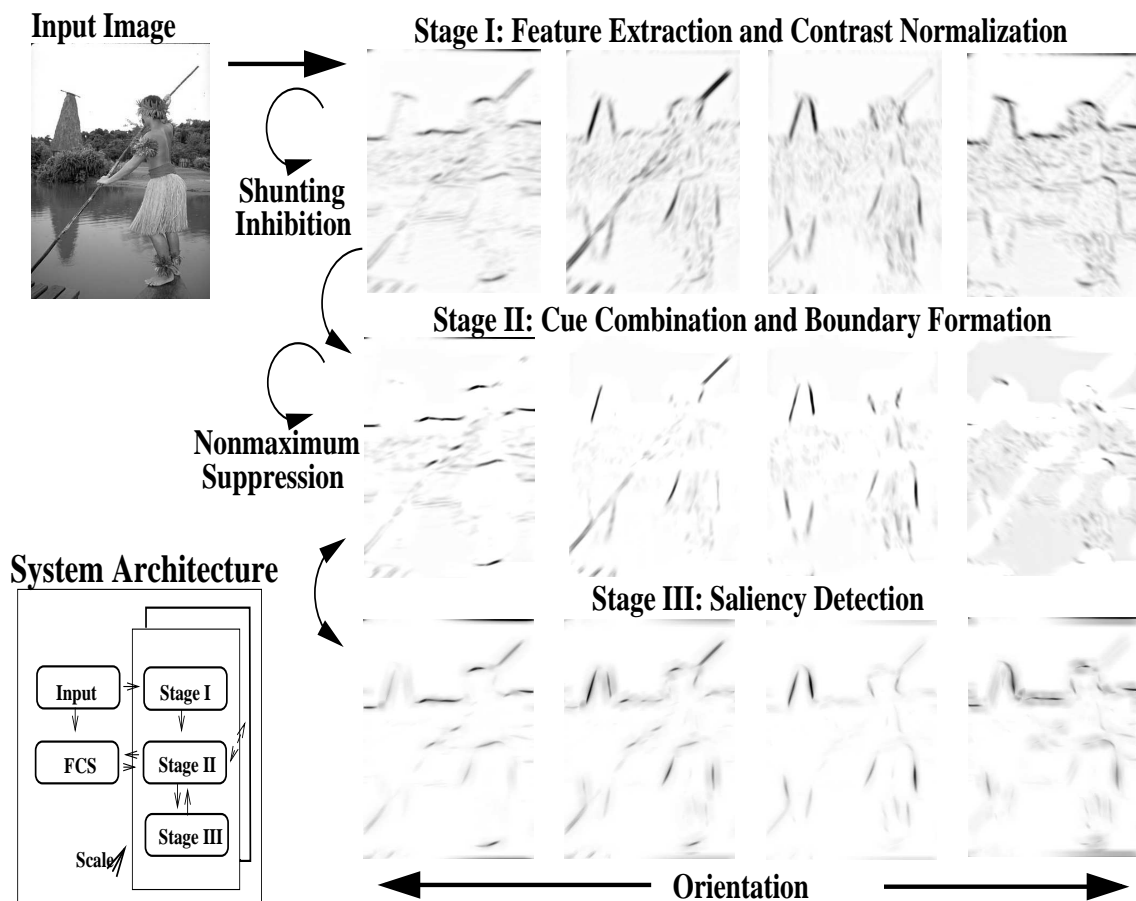
Σχήμα 3.2: (α) Διάγραμμα του συστήματός μας και (β) αντίστοιχες περιοχές του βιολογικού συστήματος όρασης

Τα κυριότερα αρχιτεκτονικά χαρακτηριστικά του συστήματός μας είναι η χρήση πολλαπλών κλιμάκων, συνεργασία ανίχνευσης ακμών/ομαλοποίησης εικόνας και η αλληλεπίδραση με έναν υψηλού-επιπέδου μηχανισμό για την ενίσχυση ομαλών περιγραμμάτων.

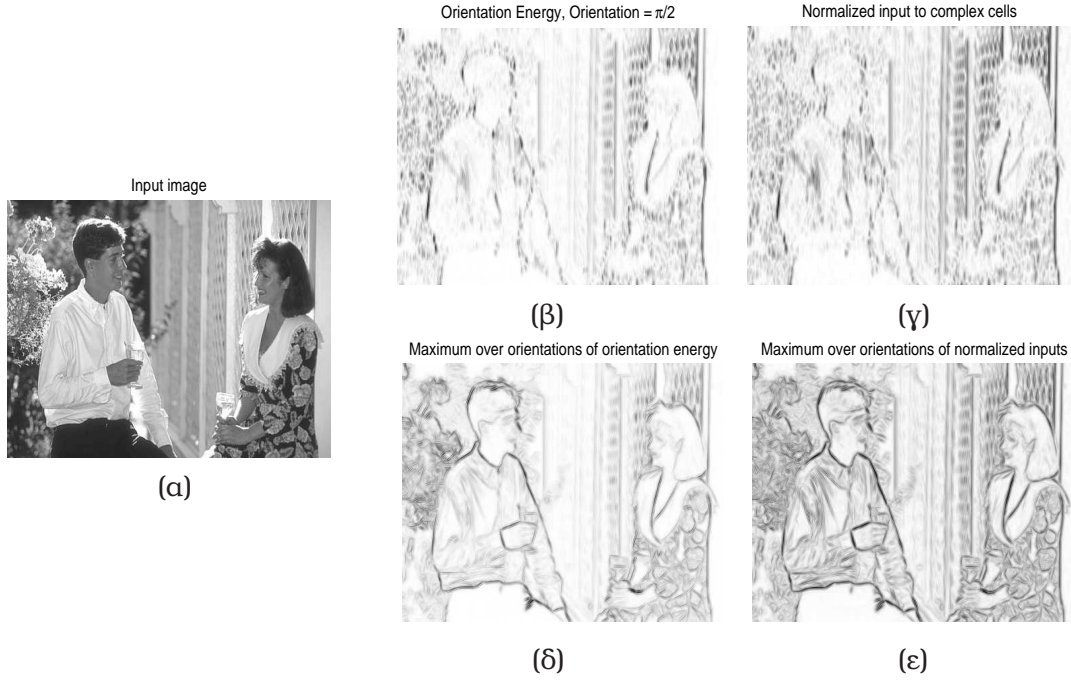
3.2.1 Στάδιο I: Εξαγωγή Χαρακτηριστικών, Διαιρετική Κανονικοποίηση

Τα πρώτα δύο στάδια του συστήματος BCS ενοποιούνται εδώ σε ένα, όπου η είσοδος στα απλά κύτταρα εκτιμάται μέσω φίλτρων Gabor· η ακολουθιακή εφαρμογή ενός φίλτρου τύπου difference - of - Gaussians με ένα φίλτρο difference -of -offset - Gaussians μπορεί να προσεγγιστεί με τη χρήση ενός φίλτρου Gabor [127], οπότε χρησιμοποιούνται εδώ λόγω των ιδιοτήτων βελτιστότητάς τους.

Για να λάβουμε υπόψη την κανονικοποίηση αντίθεσης που λαμβάνει χώρα στα απλά κύτταρα, χρησιμοποιούμε εδώ μία εξίσωση τύπου (shunting inhibition) [29]:



Σχήμα 3.3: Η αρχιτεκτονική του μοντέλου μας. Κάτω αριστερά φαίνεται η αρχιτεκτονική του συστήματος και οι αλληλεπιδράσεις των δομικών μονάδων του. Στο υπόλοιπο σχήμα φαίνεται ο υπολογισμός περιγραμμάτων σε μία συγκεκριμένη κλίματα: η ανίχνευση χαρακτηριστικών σε πολλαπλές κατευθύνσεις στο Στάδιο I ακολουθείται από τον σχηματισμό περιγραμμάτων στο Στάδιο II. Εκεί η κάτωθεν πληροφορία συνδιάζεται αναδρομικά με το άνωθεν σήμα saliency που υπολογίζεται από το Στάδιο III. Τα Στάδια I και II υπακούουν αναδρομική δυναμική που καθορίζεται από τις αλληλεπιδράσεις τους.



Σχήμα 3.4: (α)Εικόνα στην είσοδο του συστήματος, (β)-(γ) Κατευθυντική Ενέργεια για την κάθετη κατεύθυνση χωρίς και με κανονικοποίηση, αντίστοιχα, (δ)-(ε) μέγιστο ως προς κατεύθυνση της κατευθυντικής ενέργειας χωρίς και με κανονικοποίηση, αντίστοιχα.

πέρα από την είσοδο στον νευρώνα που καθορίζεται εξαρχής, ο όρος shunting μεταβάλλεται δυναμικά, βάσει της δραστηριότητας των γειτονικών νευρώνων. Η σχέση που καθορίζει το δυναμικό V_i του νευρώνα i γράφεται:

$$\frac{dV_i}{dt} = -AV_i + (C - V_i)I - V_i \sum_{j \in \mathcal{N}(i)} W_{i,j}U_j \quad (3.6)$$

$$U_i = \max(V_i, 0)$$

$$V_i^\infty = \frac{CI}{A + I + \sum_{j \in \mathcal{N}(i)} W_{i,j}U_j} \quad (3.7)$$

όπου U_j είναι ο MFR του νευρώνα j και με $W_{i,j}$ συμβολίζεται η ένταση της αλληλεπίδρασής τους. Για απλότητα, σε αυτήν και τις παρακάτω σχέσεις χρησιμοποιούμε τον μονοδιάστατο δείκτη i για έναν νευρώνα στην θέση k, l και με κατεύθυνση θ .

Παρατηρούμε ότι η σχέση αυτή μας παρέχει έναν βιολογικά αποδεκτό μηχανισμό για το νόμο του Weber: το σύστημά ακυρώνει τις όποιες τοπικές και καθολικές αλλοιώσεις της έντασης καθώς η έξοδος των νευρώνων προκύπτει από τη διαίρεση εκφράσεων που είναι αύξουσες συναρτήσεις της αντίθεσης της εικόνας. Εκτός από την κανονικοποίηση ως προς την αντίθεση, αποφεύγεται η δραστηριότητα του συστήματος σε περιοχές με θόρυβο ή υφή: καθώς όλοι οι νευρώνες είναι ενεργοί ταυτόχρονα σε τέτοιες περιοχές, καταλήγουν να αλληλοακυρώνουν τη δραστηριότητά τους.

Στο Σχ. 3.2.1 επιδεικνύουμε την λειτουργία του σταδίου αυτού: συγκρίνοντας τα αποτελέσματα στην εικόνα 3.2.1(δ) με αυτά στην 3.2.1(ε) παρατηρούμε ότι μέσω της διαιρετικής κανονικοποίησης σε περιοχές χαμηλής αντίθεσης όπως ο φράχτης στο φόντο ή οι πτυχώσεις του πουκάμισου οι ακμές γίνονται εντονότερες. Αντιθέτως,

περιοχές όπως ο θάμνος στα αριστερά καταπιέζονται, καθώς υπάρχει μία γενικότερη δραστηριότητα που οδηγεί στην αμοιβαία καταστολή των νευρώνων.

3.2.2 Στάδιο II: Σχηματισμός Περιγραμμάτων, Λέπτυνση Ακμών

Η λειτουργία αυτού του σταδίου είναι η βασικότερη της διαδικασίας ανίχνευσης περιγραμμάτων: όπως φαίνεται στο Σχ. 3.2, στο στάδιο αυτό λαμβάνει χώρα ο συνδυασμός πληροφορίας από κάτωθεν (Στάδιο I) και άνωθεν (Στάδιο III) πηγές, από τραχύτερες κλίμακες και από το σύστημα επιφάνειας, καθώς και η λέπτυνση ακμών.

Για τον *κάτωθεν όρο* οι εισοδοί άρτιων και περιπτών απλών κυττάρων συνδυάζονται σε ένα όρο τύπου κατευθυντικής ενέργειας [2, 163, 187]:

$$O = \sqrt{U_o^2 + U_e^2} \quad (3.8)$$

όπου U_e, U_o είναι οι αποκρίσεις σταθερής κατάστασης για τα απλά κύτταρα. Η ποσότητα O δεν εξαρτάται από την φάση του σήματος, που σημαίνει ότι είναι σταθερή αν στην είσοδο του συστήματος υπάρχει μία ημιτονοειδής μορφή: μία τέτοια έκφραση έχει χρησιμοποιηθεί για την μοντελοποίηση των περίπλοκων κυττάρων [2] και για την αντίληψη ορίων των αντικειμένων [163], ενώ τα δεδομένα φυσιολογίας που την υποστηρίζουν παρουσιάζονται στην αναφορά [50], σελ. 74-76. Η διαφορά με την αναφορά [187] είναι ότι πλέον οι έξοδοι των απλών κυττάρων έχουν προηγουμένως κανονικοποιηθεί, δίνοντας αποτελέσματα ανίχνευσης ακμών που είναι ανεξάρτητα της αντίθεσης της εικόνας.

Η λέπτυνση ακμών επιτυγχάνεται μέσω οριζόντιων συνδέσεων μεταξύ νευρώνων του ίδιου σταδίου. Αυτές οι αλληλεπιδράσεις είναι ανασταλτικής φύσης, οδηγώντας σε ένα δίκτυο τύπου winner - take - all [243], όπου η δραστηριότητα του πλέον ενεργού νευρώνα ενισχύεται εις βάρος των λιγότερο ενεργών γειτόνων του.

Πληροφορία σχετικά με την *saliency* και τα αποτελέσματα ανίχνευσης περιγραμμάτων σε *τραχύτερες κλίμακες* παρέχονται από άλλα στάδια επεξεργασίας, ενισχύοντας την δραστηριότητα των νευρώνων που βρίσκονται σε έντονα όρια ή σε τραχύτερες κλίμακες, ακόμη και αν η κάτωθεν είσοδος που λαμβάνουν δεν είναι αρκετά ισχυρή.

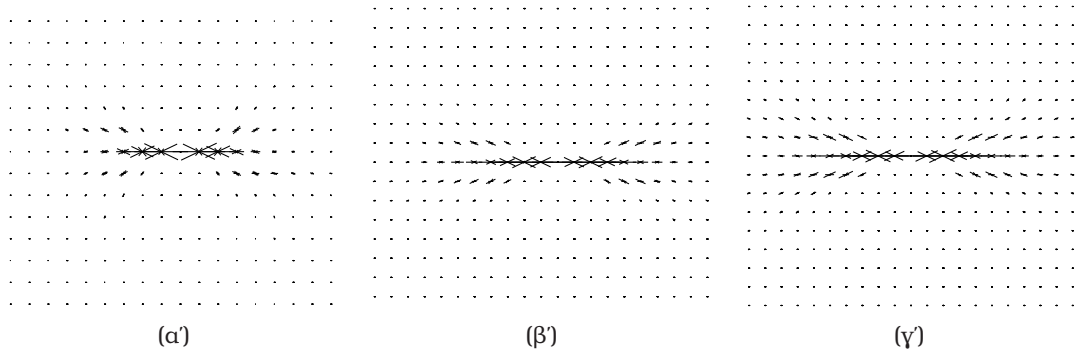
Η *διαδικασία επιφάνειας - surface process*, S , προωθεί τη δημιουργία ακμών σε περιοχές με έντονες ασυνέχειες σε χαρακτηριστικά. Για έναν νευρώνα που αποκρίνεται σε όρια με κατεύθυνση θ το μέτρο της κατευθυντικής παραγωγού κάθετα στο θ , $|\nabla S_{\theta\perp}|$, παρέχεται από το FCS στο στάδιο II.

Η σχέση εξέλιξης του δυναμικού ενός νευρώνα i στο στάδιο αυτό γράφεται ως:

$$\frac{dV_i}{dt} = -AV_i + (C - V_i)I - (V_i + D) \sum_{j \in \mathcal{N}(i)} W_{i,j} U_j \quad (3.9)$$

$$\text{όπου} \quad I = [c_1 O + c_2 C + c_3 T + c_4 |\nabla S_{\theta\perp}|].$$

Παραπάνω $\mathcal{N}(i)$ είναι η γειτονιά του νευρώνα i , το $W_{i,j}$ καθορίζει τις οριζόντιες αλληλεπιδράσεις των νευρώνων στο στάδιο II, T είναι το άνωθεν (*top-down*) σήμα που παρέχει το στάδιο III, C είναι ο όρος που προέχεται από ένα τραχύτερο σύστημα που λειτουργεί σε μία μεγαλύτερη κλίμακα (*coarse-scale*), O είναι ο κάτωθεν όρος (*oriented-energy*) και $U_j = g(V_j)$ είναι ο MFR του γειτονικού νευρώνα j .



Σχήμα 3.5: Οι λοβοί που δίνει το μοντέλο καμπύλων *Elastica* (3.13) για παραμέτρους (α) $\lambda = 10, \sigma = 0.4$, (β) $\lambda = 10, \sigma = 0.2$, (γ) $\lambda = 20, \sigma = 0.2$.

Για το στάδιο αυτό χρησιμοποιούμε τη σιγμοειδή συνάρτηση

$$U = g(V) = \frac{1}{1 + \exp(-(\frac{V}{\beta} - \gamma))} \quad (3.10)$$

για να συνδέσουμε το δυναμικό του νευρώνα με το MFR του, όπου το β καθορίζει την κλίση του καμπύλης εισόδου εξόδου, και το γ λειτουργεί ως παράμετρος κατωφλίου.

Σημειώνουμε ότι όλες οι παράμετροι που εμπλέκονται σε αυτές τις σχέσεις έχουν εκμαθηθεί χρησιμοποιώντας δεδομένα εκπαίδευσης σύμφωνα με τη μέθοδο της ενότητας 3.4.

3.2.3 Στάδιο III: Υπολογισμός Saliency

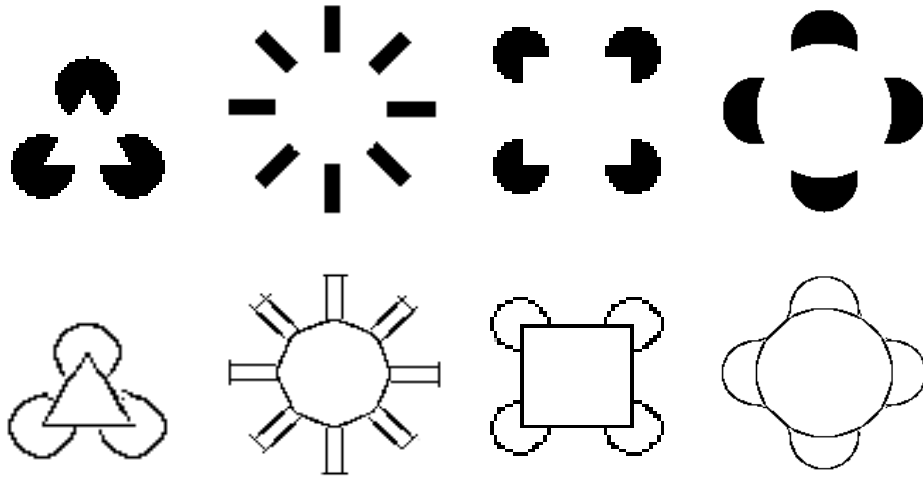
Σε αυτό το στάδιο μοντελοποιείται η λειτουργία των νευρώνων που περιγράφονται στην αναφορά [97], οι οποίοι αποκρίνονται και χωρίς να έχουν ένα σήμα στο δεκτικό τους πεδίο, επηρεαζόμενοι μόνο από τη δραστηριότητα της γειτονιάς τους. Σε ένα τέτοιο στάδιο επιτελούνται λειτουργίες αντίληψης φανταστικών περιγραμμάτων *illusory contours* όπου ενώ δεν υπάρχει καμία αντίθεση στην ένταση της εικόνας, αντιλαμβανόμαστε την ύπαρξη μίας ακμής (βλ. Σχ. 3.6).

Έτσι στο στάδιο αυτό έντονα περιγράμματα ανιχνεύονται και ανατροφοδοτούνται στο Στάδιο II, ενεργοποιώντας έτσι μία συνεργετική διαδικασία που ωθεί το σύστημα προς την ανίχνευση ομαλών και μεγάλων περιγραμμάτων. Ο υπολογισμός της *saliency* σε ένα σημείο μπορεί να επιτευχθεί χρησιμοποιώντας ένα σταθμισμένο άθροισμα των αποκρίσεων των γειτονικών νευρώνων, όπου τα βάρη λαμβάνουν υπόψη τις σχετικές τους τοποθεσίες. Για έναν οριζόντιο νευρώνα συνήθως χρησιμοποιούνται βάρη όπως στο Σχ. 3.5, που δίνουν έμφαση σε περιγράμματα χαμηλής καμπυλότητας.

Για κάθε μία από τις δύο πλευρές από τις οποίες μπορεί να έρθει η γραμμή (δεξιά/ αριστερά στην εικόνα) ένα σταθμισμένο άθροισμα υπολογίζεται ξεχωριστά και τα δύο αθροίσματα πολλαπλασιάζονται, ώστε να εξασφαλιστεί ότι υπάρχει μαρτυρία και από τις δύο πλευρές του νευρώνα.

Η εξίσωση που περιγράφει την εξέλιξη ενός νευρώνα στο στάδιο αυτό είναι

$$\frac{dV_i}{dt} = -AV_i + \left(\sum_{j \in N^+(i)} W_{i,j}^+ U_j \right) \left(\sum_{j \in N^-(i)} W_{i,j}^- U_j \right) \quad (3.11)$$



Σχήμα 3.6: Ανίχνευση φανταστικών ακμών: στις πάνω εικόνες αντιλαμβανόμαστε ανύπαρκτες ακμές σε λευκές περιοχές. Στις κάτω εικόνες φαίνονται τα αποτελέσματα ανίχνευσης ακμών του δικτύου μας, το οποίο χρησιμοποιώντας το υποσύστημα ανίχνευσης saliency μοντελοποιεί αυτή τη συμπεριφορά του ανθρώπινου οπτικού συστήματος.

όπου N^+ , N^- είναι τα υποσύνολα των γειτόνων που βρίσκονται στις δύο πλευρές του i , W^+ , W^- είναι τα αντίστοιχα βάρη σύνδεσης και U_j είναι ο MFR του νευρώνα j , που προκύπτει από την ανόρθωση του V_j .

Αρκετά περίπλοκες εκφράσεις έχουν χρησιμοποιηθεί [86, 172] για να εκφράσουν τέτοιους τύπους συνδέσεων, περιλαμβάνοντας ένα μεγάλο πλήθος παραμέτρων· βάσει της σχέσης που έχει καθιερωθεί στην αναφορά [235] με το μοντέλο Elastica που παρουσιάζεται αργότερα, χρησιμοποιούμε μόνο δύο απλές παραμέτρους, όπου η μία αντιστοιχεί στο μέγεθος και η άλλη στο εύρος των δύο λοβών, όπως φαίνεται στο Σχ. 3.5.

Η ικανότητα του σταδίου αυτού στην ανίχνευση φανταστικών περιγραμμάτων φαίνεται στο Σχ. 3.6 όπου δείχνουμε μερικά σχήματα τύπου Kanizsa για τα οποία το σύστημά μας έχει εντοπίσει τις φανταστικές ακμές που αντιλαμβανόμαστε. Φυσικά υπάρχουν πιο πολύπλοκα και ακριβή μοντέλα των λειτουργιών αυτών, αλλά τα αποτελέσματα που παίρνουμε είναι ικανοποιητικά.

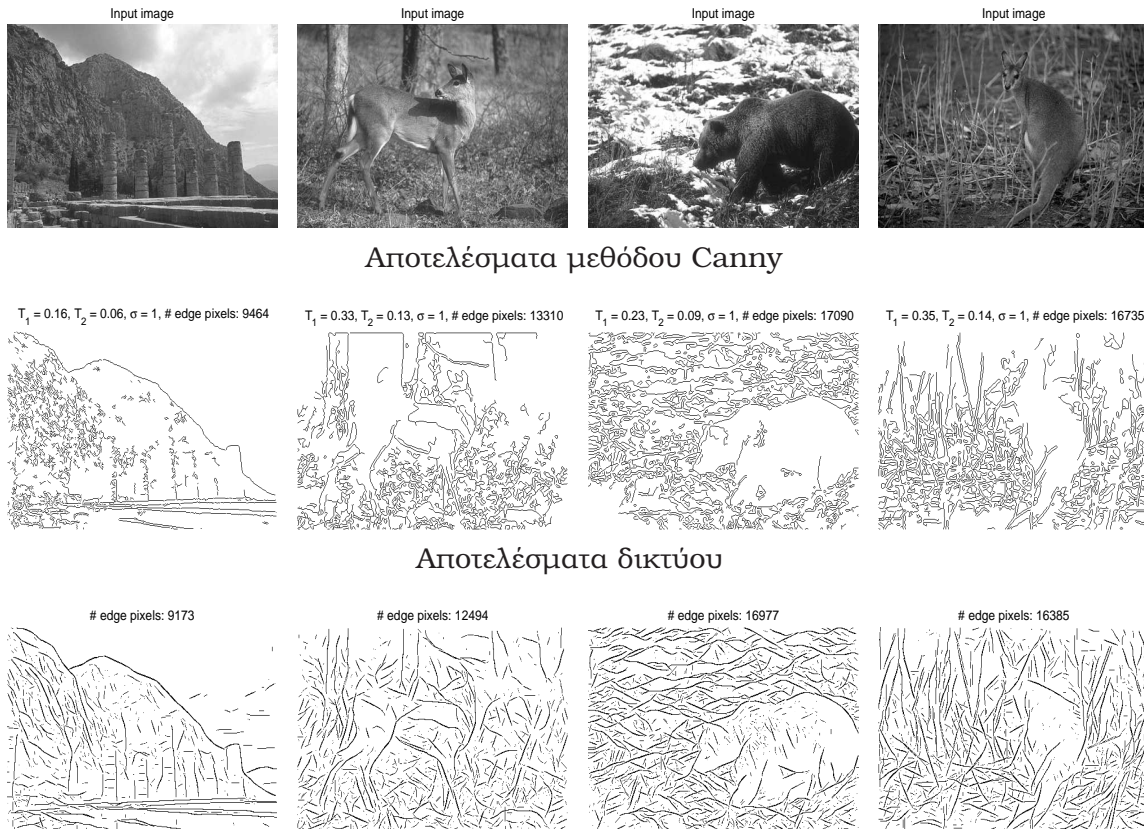
3.2.4 Σύστημα Επιφάνειας

Στο στάδιο αυτό ομαλοποιούνται τα χαρακτηριστικά που εξάγονται από το εσωτερικό των αντικειμένων, ενώ τα σημεία ασυνέχειάς τους προτείνονται στο σύστημα BCS ως πιθανές τοποθεσίες ακμών. Αν και στο αρχικό μοντέλο BCS - FCS χρησιμοποιούνται διαφορετικά χαρακτηριστικά, για απλότητα χρησιμοποιούμε εδώ την φωτεινότητα της εικόνας, όπως γίνεται και στις περισσότερες εφάμιλλες τεχνικές [187, 213, 12].

Η εξέλιξη του δυναμικού ενός νευρώνα του συστήματος αυτού, μπορεί να γραφεί ως:

$$\frac{d}{dt}S = \sum_{\theta} |\nabla_{\theta^{\perp}} S| (1 - (U^{\theta})) \quad (3.12)$$

όπου U^{θ} είναι ο MFR του νευρώνα του σταδίου II με κατεύθυνση θ , ο οποίος βρίσκεται μεταξύ των γειτόνων που χρησιμοποιήθηκαν για τον υπολογισμό του $\nabla_{\theta^{\perp}} S$, ενώ όπως



Σχήμα 3.7: Σύγκριση αποτελεσμάτων ανίχνευσης ακμών τύπου Canny [28] και του ιεραρχικού δικτύου μας [126]. Ορίζοντας τα κατώφλια και για τις δύο μεθόδους ώστε να επιστρέφουν το ίδιο πλήθος ακμών, από τη μέθοδο Canny παίρνουμε ένα μεγάλο πλήθος ακμών σε περιοχές με υφή. Αντίθετα, το δίκτυό μας χρησιμοποιώντας μία υψηλότερου επιπέδου διαδικασία ανίχνευσης ομαλών ακμών εστιάζει σε ομαλές και έντονες ακμές. Η διασπορά της Gaussian και τα πάνω και κάτω κατώφλια της μεθόδου Canny που αναγράφονται σε κάθε εικόνα ορίστηκαν ξεχωριστά, ώστε να δίνουν την οπτικά καλύτερη απόδοση.

στο [187] η τιμή S του νευρώνα i είναι η αφαιρούμενη ποσότητα στην έκφραση $\nabla_{\theta \perp} S$. Ο όρος $(1 - (U^\theta))$ εμποδίζει τη διάχυση όταν υπάρχει κάποια έντονη ακμή με γωνία θ , ενώ όταν απουσιάζουν οι ακμές ($U^\theta = 0$) αφήνει την τιμή του S να έρθει κοντά σε αυτή των γειτόνων του.

3.3 Ανάλυση του Μοντέλου με Τεχνικές Όρασης Υπολογιστών

Πολλές από τις λειτουργίες του μοντέλου μας όσο και του αρχικού συστήματος BCS/FCS αντιστοιχούν σε ορισμένες γνωστές τεχνικές της Όρασης Υπολογιστών. Αποσαφηνίζοντας αυτές τις σχέσεις μπορούμε να αποκτήσουμε μία καλύτερη αίσθηση της λειτουργίας του, και να εισάγουμε ωριμότερες τεχνικές της όρασης υπολογιστών. Αρχικά εξετάζουμε τις μεμονωμένες λειτουργίες του δικτύου και συνεχίζουμε με τη συνολική του λειτουργία.

3.3.1 Καταπίεση μη-μέγιστων

Η τεχνική της καταπίεσης μη-μέγιστων (non-maximum suppression) χρησιμοποιείται συχνά στην ανίχνευση ακμών για την εξαγωγή ενός λεπτού και καθαρού πίνακα ακμών από τις ασαφείς εξόδους που προκύπτουν από τη συνέλιξη με χωρικά εκτεταμένα φίλτρα [28, 53].

Οι τύπου Winner-Take-All [245, 243] συνδέσεις στο Στάδιο II εκτελούν μία ανάλογη λειτουργία εισάγοντας, όπως θα δούμε στη συνέχεια, έναν όρο ποινής στο εύρος των ορίων των αντικειμένων, κατά έναν τρόπο ανάλογο των αναφορών [12, 75].

3.3.2 Μοντέλο Elastica και ομαδοποίηση ακμών

Το γενικότερο πρόβλημα της ομαδοποίησης ακμών σε ενιαίες καμπύλες έχει μελετηθεί εκτενώς στην όραση υπολογιστών, καθώς είναι ένα σημαντικό ενδιάμεσο βήμα για την μετάβαση από απλά χαρακτηριστικά σε αναπαραστάσεις για αντικείμενα.

Το συγκεκριμένο σχήμα διασυνδέσεων μεταξύ των νευρώνων αρχικά προτάθηκε από τον S. Grossberg και στη συνέχεια χρησιμοποιήθηκε από πολλούς άλλους ερευνητές, όπως π.χ. [89, 181, 64, 235, 146, 172]. Αυτό είναι φυσικό, καθώς το σχήμα τους ευνοεί καμπύλες χαμηλής καμπυλότητας που εμφανίζονται συχνά στη φύση και εισάγει χρήσιμη πρότερη γνώση στην ανίχνευση περιγραμμάτων. Μία σαφής ερμηνεία αυτού του τύπου συνδέσεων παρουσιάστηκε στην αναφορά [235] όπου καθιερώθηκε μία σχέση με το μοντέλο καμπυλών Elastica [164].

Σύμφωνα με το μοντέλο αυτό μία καμπύλη Γ μοντελοποιείται ως η τροχιά ενός σωματιδίου στο χώρο (x, y, θ) όπου x, y είναι η θέση του σωματιδίου και θ η κατεύθυνσή του. Η κίνηση του σωματιδίου περιγράφεται τότε από την στοχαστική διαφορική εξίσωση:

$$\frac{d}{dt}x = \cos(\theta) \quad \frac{d}{dt}y = \sin(\theta) \quad \frac{d}{dt}\theta = \kappa \sim \mathcal{N}(0, \sigma^2; t) \quad (3.13)$$

όπου κ , η καμπυλότητα της τροχιάς, θεωρείται μία διαδικασία λευκού θορύβου. Το μήκος των καμπυλών περιορίζεται σταματώντας τη διαδικασία σε μία χρονική στιγμή που ακολουθεί μία κατανομή Laplace με παράμετρο λ .

Χρησιμοποιώντας το μοντέλο αυτό η πιθανότητα μίας καμπύλης Γ μπορεί να συσχετιστεί με την ενέργεια Elastica του Euler που εκφράζεται ως ένα επικαμπύλιο ολοκλήρωμα κατά μήκος της καμπύλης Γ :

$$E(\Gamma) = \int_{\Gamma} (\alpha k^2 + \beta) ds \quad (3.14)$$

Η ομαλότητα των καμπυλών καθορίζεται από την παράμετρο α ενώ η β σχετίζεται με το αναμενόμενο μήκος μίας καμπύλης. Συγκεκριμένα αποδεικνύεται [164] ότι από τη σχέση (3.13) έχουμε

$$P(\Gamma) \propto e^{-\int_{\Gamma} (\alpha k^2 + \beta ds)}, \quad \beta = \lambda, \quad \alpha = \frac{1}{2\sigma^2} \quad (3.15)$$

Συνεπώς μπορούμε να δούμε τα ελάχιστα της ενέργειας Elastica (3.14) ως κορυφές της κατανομής (3.15) πάνω στις τροχιές των σωματιδίων.

Η σχέση μεταξύ αυτού του μοντέλου ακμών με το σχήμα των λοβών καθιερώθηκε στην αναφορά [235]: ξεκινώντας από το μοντέλο Elastica η ύστερη πιθανότητα να

περνάει μία καμπύλη μέσα από το σημείο $X_0 = (x_0, y_0, \theta_0)$ εκφράζεται μετά από τις διακριτοποιήσεις ως το γινόμενο δύο σταθμισμένων αθροισμάτων:

$$P(X_0) = \left(\sum_{i \in r(X_0)} g(X_0, X_i) P(X_i) \right) \left(\sum_{j \in l(X_0)} g(X_0, X_j) P(X_j) \right)$$

Στην σχέση αυτή το g είναι μία εξαδιάστατη συνάρτηση πάνω στα $(x, y, \theta, x', y', \theta')$ που ισούται με την πιθανότητα ένα σωματίδιο να περάσει από το σημείο (x', y', θ') δεδομένου ότι την $t = 0$ βρίσκεται στη θέση x, y, θ , $r(X_0)$ και $l(X_0)$ είναι οι γειτονίες στις δύο πλευρές το σημείου X_0 και $P(X_i)$ είναι η πιθανότητα μία καμπύλη να περνάει μέσα από το X_i . Στην αναφορά [235] χρησιμοποιώντας προσεγγίσεις Monte Carlo το σχήμα της g βρέθηκε ότι δίνει τη μορφή συνδέσεων στο Σχ. 3.5. Μπορεί έτσι λοιπόν κανείς να συνδέσει τον όρο ανάδρασης με την ύστερη πιθανότητα διέλευσης μία καμπύλης μέσω ενός σημείου δεδομένου του περιβάλλοντός του. Πέρα από μία καλύτερη κατανόηση των τύπου των συνδέσεων, αυτή η σχέση απλοποιεί σημαντικά τον σχεδιασμό τους, χρησιμοποιώντας μόνο τις τιμές των λ και σ .

3.3.3 Μεταβολική ανάλυση του Σταδίου II

Η αρχιτεκτονική BCS/FCS ταιριάζει με τη χρήση των διαδικασιών ακμών/επιφάνειας (line/surface processes) που χρησιμοποιούνται στην Όραση Υπολογιστών, π.χ. [75, 245, 140, 12]. Η ιδέα είναι η ίδια, δηλαδή να υπάρχει ένα σύστημα για ανίχνευση ακμών (διαδικασία ακμών), ένα για την ομαλοποίηση της εικόνας (διαδικασία επιφάνειας) και αυτά τα δύο να συνεργάζονται. Στις παραπάνω αναφορές υπάρχει πάντα με τον έναν ή τον άλλο τρόπο κάποια σύνδεση με το ενεργειακό Mumford-Shah [168] (2.2) το οποίο ελαττώνεται καθώς το σύστημα εξελίσσεται.

Στην παράγραφο αυτή εισάγουμε για το σύστημά μας μία συνάρτηση Lyapunov, δηλαδή μία θετικά ορισμένη συνάρτηση της κατάστασης του συστήματος, η οποία ελαττώνεται καθώς αυτό εξελίσσεται χρονικά. Η ύπαρξη συναρτήσεων Lyapunov για αναδρομικά νευρωνικά δίκτυα [35, 100] έχει εκμεταλλευτεί στο παρελθόν [124, 140, 242, 245] για τον σχεδιασμό νευρωνικών δικτύων που επιλύουν μεταβολικά προβλήματα όρασης. Ο σκοπός μας εδώ είναι να δούμε πώς η αναδρομική αρχιτεκτονική για το στάδιο II που προτάθηκε εδώ μπορεί να ερμηνευτεί με μεταβολικούς όρους και πώς μπορούμε να ερμηνεύσουμε τη συμπεριφορά του βάσει του συναρτησιακού που κατασκευάζεται.

Ως σύστημά μας θεωρούμε το σύνολο των νευρώνων που βρίσκονται σε μία συγκεκριμένη κλίμακα στο στάδιο II και τους νευρώνες του Συστήματος Επιφάνειας στην ίδια κλίμακα ενώ οι τιμές των υπόλοιπων νευρώνων θεωρούνται σταθερές. Αντί της σχέσης (3.9) θεωρούμε μία απλοποιημένη μορφή της:

$$\frac{dV_i}{dt} = -AV + CI - D \sum_{j \in \mathcal{N}(i)} W_{i,j} U_j \quad (3.16)$$

όπου αντί της συναπτικής αλληλεπίδρασης μεταξύ νευρώνων [123] χρησιμοποιήσαμε το συνηθισμένο μοντέλο του αθροίσματος των εισόδων· αν και μπορεί κανείς να γράψει μία συνάρτηση Lyapunov χρησιμοποιώντας την αρχική σχέση (3.9) [35], οι εκφράσεις γίνονται πολύπλοκες και δεν βοηθούν τη διαίσθηση.

Θεωρούμε αρχικά την είσοδο I σταθερή, οπότε μία συνάρτηση Lyapunov του

συστήματος είναι:

$$\mathcal{E} = \sum_i [A \underbrace{\int_{1/2}^U g^{-1}(u) du}_{P(U): \text{Penalty for values of } U} - CIU] + \frac{D}{2} \underbrace{\sum_{i,j \in \mathcal{N}(i)} U_i W_{i,j} U_j}_{C(U): \text{cost of configurations of } U} \quad (3.17)$$

Παρακάτω εξετάζεται η περίπτωση που και το I μεταβάλλεται χρονικά οπότε δεν αποδεικνύουμε τον ισχυρισμό αυτό εδώ.

Εξετάζοντας τους όρους που εμπλέκονται στην έκφραση (3.17) βλέπουμε για την πρώτη έκφραση ότι μπορεί να γραφεί και ως:

$$P(U) = [U \log(U) + (1 - U) \log(1 - U)] / \beta + \gamma U - \gamma^2$$

Ο πρώτος όρος, $[U \ln(U) + (1 - U) \ln(1 - U)]$, είναι ένας όρος αρνητικής εντροπίας και τιμωρεί δυαδικές αποκρίσεις. Ο δεύτερος όρος, γU , τιμωρεί τις υψηλές αποκρίσεις γενικότερα. Ο πρώτος όρος οφείλεται στην χρήση της σιγμοειδούς στην σχέση (3.10) και ο δεύτερος στην μετατόπισή της κατά γ . Ο παράγοντας β σχετίζεται με την κλίση της σιγμοειδούς: αν το β είναι μεγάλο, δηλαδή η σιγμοειδής έχει μεγάλη κλίση η ποινή στις δυαδικές αποκρίσεις είναι μικρή και αντίστροφα. Η έκφραση αυτή συνδυάζεται με τον όρο $-IU$, ο οποίος ελαττώνει το κόστος μίας υψηλής τιμής του U , επιτρέποντας έτσι την εμφάνιση ακμών.

Οι όροι αλληλεπίδρασης $W_{i,j}$ εκφράζουν το βάρος σύνδεσης μεταξύ των νευρώνων i και j που βρίσκονται στις θέσεις/κατευθύνσεις (x_i, y_i, θ_i) και (x_j, y_j, θ_j) αντίστοιχα. Σκοπός των αλληλεπιδράσεων αυτών είναι να εξαχθεί ένας λεπτός πίνακας ακμών από τις ευρείες εισόδους του συστήματος. Στη συνέχεια χρησιμοποιούμε την αναλυτικότερη έκφραση για τέτοια βάρη $W_{i,j} = W^{\theta,\phi}(x_i - x_j, y_i - y_j)$ όπου χρησιμοποιούμε την χωρική συμμετρία των συνδέσεων. Από τις αρχικές μας προσπάθειες για την υπολογισμό των βαρών του δικτύου [126, 127] εξετάσαμε την περίπτωση όπου τα βάρη σύνδεσης μεταξύ όμοιων κατευθύνσεων εκφράζονται ως η διαφορά ενός ισοτροπικού Gaussian φίλτρου G_s και ενός φίλτρου με επιμήκυνση με κύριο άξονα κατά τη γωνία θ G_s^θ , ενώ για διαφορετικές κατευθύνσεις εκφράζεται μέσω μίας ισοτροπικής Gaussian G_d . Για τους πυρήνες αυτούς μπορούμε πάντα να βρούμε έναν άλλο πυρήνα g τέτοιο ώστε $g * g = G$ επίσης χρησιμοποιώντας το πεδίο Fourier είναι εύκολο να δείξουμε ότι τα g και G αυτά έχουν τις ίδιες ιδιότητες ισοτροπίας/κατευθυντικότητας.

Χρησιμοποιώντας τέτοιους πυρήνες ο όρος ενέργειας για τη διαδικασία γραμμών μπορεί να απλουστευθεί ως [127]:

$$C(U) = \sum_{i,j,\theta} [g_s * U^\theta]^2 - \sum_{i,j,\theta} [g_s^\theta * U^\theta]^2 \quad (3.18)$$

$$+ 1/2 \sum_{i,j,(\theta,\phi):\theta \neq \phi} [g_d * U^\theta] [g_d * U^\phi] \quad (3.19)$$

Στην παραπάνω σχέση με U^θ συμβολίζουμε το σύνολο των νευρώνων με κατεύθυνση θ και g_s είναι ένας ισοτροπικός πυρήνας, οπότε ο όρος $[g * U^\theta]^2$ τιμωρεί γενικότερα ευρείες ακμές, ανεξαρτήτως του αν τα στοιχεία ακμών βρίσκονται στην ίδια ευθεία ή είναι διασκορπισμένα. Το g^θ είναι μία Gaussian συνάρτηση με κύριο άξονα κατά την προτιμώμενη κατεύθυνση των νευρώνων, θ ο όρος $-[g_s^\theta * U^\theta]^2$ είναι συνεπώς ένα αρνητικό δυναμικό -έναν όρος 'αντίδρασης' (reaction) που οδηγεί στο σχηματισμό

ευθυγραμμισμένων περιγραμμάτων. Ο τελευταίος όρος τιμωρεί νευρώνες με διαφορετικές κατευθύνσεις που είναι συγχρόνως ενεργοί στις ίδιες γειτονιές, αφήνοντας έτσι μία μόνο κατεύθυνση να επικρατήσει.

Συνοψίζοντας οι όροι ποινή $[g_s * U_\theta]^2 [g_d * U^\theta] \cdot [g_d * U^\phi]$ καταπιέζουν ευρείες δομές, ενώ ο όρος αντίδρασης $-[g_s^\theta * U^\theta]^2$ οδηγεί στην εμφάνιση μεμονωμένων ακμών.

Αυτό το οποίο απομένει είναι να εισάγουμε και την αλληλεπίδραση με το σύστημα επιφάνειας. Επεκτείνοντας τα επιχειρήματα [100, 35, 242] μπορούμε να αποδείξουμε ότι βάσει της (3.12) μία συνάρτηση Lyapunov του συστήματος είναι:

$$\begin{aligned} \mathcal{E} = & \sum_i c_4 \underbrace{(1-U)|\nabla_{\theta^\perp} S|^2}_{\text{Line - Surface interaction}} - U \underbrace{[c_1 O + c_2 T + c_3 C]}_{\text{External inputs}} \\ & + a \underbrace{\int_{1/2}^U g^{-1}(u) du + \frac{d}{2} C(U)}_{\text{Cost for line process}}. \end{aligned} \quad (3.20)$$

όπου οι σταθερές c_1, \dots, c_4 έχουν απορροφήσει τη σταθερά c της (3.16).

Συγκεκριμένα, αυτή η συνάρτηση έχει ένα κάτω φράγμα, καθώς οι έξοδοι των νευρώνων στους αφαιρούμενους όρους δεν μπορούν να ξεπεράσουν τη μονάδα. Επίσης έχουμε ορίσει $\nabla_{\theta^\perp} S = S' - S_i$, όπου S' είναι το γειτονικό σημείο του i κατά μήκος μίας γραμμής με γωνία θ^\perp . Συνεπώς έχουμε ότι

$$\frac{\partial \mathcal{E}}{\partial S_i} = -2c_4 \sum_\theta (1 - U_i) \nabla_{\theta^\perp} S^\theta$$

ενώ από την (3.12) έχουμε:

$$\frac{dS_i}{dt} = \sum_\theta (1 - U_i) \nabla_{\theta^\perp} S^\theta$$

Καθώς $c_4 > 0$, για κάθε i έχουμε $\frac{\partial \mathcal{E}}{\partial S_i} \frac{dS_i}{dt} < 0$ οπότε η εξέλιξη της διαδικασίας επιφάνειας οδηγεί σε μία μείωση του \mathcal{E} .

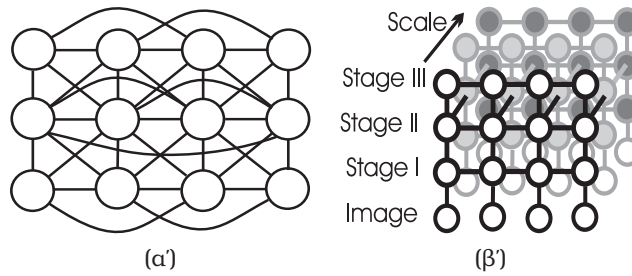
Για τη διαδικασία γραμμών, γράφοντας $cI = [c_1 O + c_2 T + c_3 C + c_4 |\nabla_{\theta^\perp} S|^2]$ έχουμε

$$\begin{aligned} \frac{\partial \mathcal{E}}{\partial U_i} &= aV - cI + d \sum_{j \in \mathcal{N}(i)} W_{i,j} U_j \quad \text{ka}j'wc \\ \frac{\partial C(U)}{\partial U_i} &= \sum_{j \in \mathcal{N}(i)} W_{i,j} U_j + U_j W_{j,i} = 2 \sum_{j \in \mathcal{N}(i)} W_{i,j} U_j \end{aligned}$$

εάν $W_{i,j} = W_{j,i}$, το οποίο επιβάλλεται εξ αρχής στις συνδέσεις των νευρώνων από την απαίτηση της συμμετρίας. Από τις (3.10) και (3.16) έχουμε για τον MFR του νευρώνα i

$$\frac{dU_i}{dt} = \left. \frac{dg(V)}{dV} \right|_{V=V_i} \frac{dV_i}{dt} = a \left[-aV_i + cI - d \sum_{j \in \mathcal{N}(i)} W_{i,j} U_j \right]$$

όπου η a είναι μία θετική ποσότητα, δεδομένου ότι η g είναι αύξουσα συνάρτηση.



Σχήμα 3.8: Τύπος συνδέσεων για μία BM (β) Το σύστημά μας ως μία BM.

Συμπεραίνουμε λοιπόν πάλι ότι οι $\frac{\partial \mathcal{E}}{\partial U_i}$ και $\frac{dU_i}{dt}$ έχουν αντίθετο πρόσημο για όλα τα i , οπότε και η εξέλιξη της διαδικασίας γραμμών οδηγεί επίσης στη ελάττωση της \mathcal{E} .

3.4 Εκμάθηση των Παραμέτρων του Μοντέλου

Στα αρχικά μας πειράματα με το μοντέλο [126, 127] χρησιμοποιήθηκαν αναλυτικές εκφράσεις για τα βάρη των συνδέσεων του μοντέλου, οι οποίες καθορίστηκαν με βασικό κριτήριο τη διαίσθησή και την εμπειρία. Είναι όμως αμφίβολο κατά πόσο οι δύο παράγοντες αυτοί επαρκούν για τον καθορισμό ενός μοντέλου, όταν εμπλέκονται τόσες αντιφατικές απαιτήσεις. Για το λόγο αυτό προτείνεται εδώ ένας αλγόριθμος για την μάθηση των παραμέτρων του από ένα σύνολο δεδομένων εκπαίδευσης, αναθέτοντας πλέον μόνο το σχεδιασμό του δικτύου στην δική μας κρίση. Για το σκοπό αυτό χρειάζεται μία πιθανοτική ερμηνεία της λειτουργίας του δικτύου, που επιτρέπει την διατύπωση του προβλήματος της εκτίμησης παραμέτρων σε ένα στέρεο πλαίσιο.

Συγκεκριμένα, θεωρώντας ότι η έξοδος του κάθε νευρώνα δίνει την πιθανότητα ύπαρξης ακμής στη θέση και διεύθυνσή του, το πρόβλημα της εκμάθησης μπορεί να εκφραστεί ως η ελαχιστοποίηση της απόστασης Kullback Leibler μεταξύ των εξόδων του δικτύου και των πραγματικών πιθανοτήτων ακμών. Ως σύνολο εκπαίδευσης χρησιμοποιούνται οι χειρωνακτικά καθορισμένες καταταμήσεις εικόνων του συνόλου δεδομένων του Berkeley [155] οι οποίες χρησιμοποιούνται ως ιδανικές έξοδοι ενός ανιχνευτή ακμών.

Καθώς το μοντέλο μας είναι αρχιτεκτονικά παρόμοιο με ένα γραφικό δίκτυο τύπου μηχανής Boltzmann (Boltzmann Machine- (BM)), παρουσιάζουμε αρχικά τα βασικά στοιχεία για την εκτίμηση παραμέτρων σε Boltzmann Machines. Στη συνέχεια σχετίζουμε την Mean Field Approximation με τη λειτουργία του μοντέλου μας και διατυπώνουμε έναν αλγόριθμο εκμάθησης των παραμέτρων του δικτύου.

3.4.1 Η μηχανή Boltzmann

Η μηχανή Boltzmann [1, 98, 96] (βλ. Σχ. 3.8) είναι ένα πιθανοτικό νευρωνικό δίκτυο από συμμετρικά συνδεδεμένους δυαδικούς νευρώνες, που διαχωρίζονται σε ορατές και κρυφές μονάδες, X και Y αντίστοιχα.

Ο διαχωρισμός των μονάδων στις ομάδες αυτές δεν είναι σταθερός, αλλά εξαρτάται από τα διαθέσιμα δεδομένα.

Ένα ενεργειακό της μορφής

$$E(X, Y) = - \left(\frac{1}{2} Y^T V Y + \frac{1}{2} X^T W X + \frac{1}{2} Y^T J X \right) \quad (3.21)$$

χρησιμοποιείται συνήθως για τις BM, όπου οι συμμετρικοί πίνακες V, W, J καθορίζουν τις ένδο/δια τμηματικές αλληλεπιδράσεις. Χρησιμοποιώντας αυτή τη συνάρτηση μία κατανομή πιθανότητας τύπου Boltzmann - Gibbs μπορεί να οριστεί πάνω στην κατάσταση του συστήματος:

$$P_{BM}(X, Y) = \frac{1}{Z} \exp(-E(X, Y)/T), \quad Z = \sum_{X, Y} \exp(-E(X, Y)/T), \quad (3.22)$$

όπου Z η (partition function) της κατανομής· για απλότητα θα θεωρήσουμε $T = 1$ στη συνέχεια. Η συνάρτηση αυτή μπορεί να γραφτεί με την τυπική μορφή των MRFs ως

$$P_{BM}(X, Y) = \prod_{(i,j) \in \mathcal{E}(X,Y)} \Phi(X_i, Y_j) \prod_{(i,j) \in \mathcal{E}(X,X)} \Psi(X_i, X_j) \prod_{(i,j) \in \mathcal{E}(Y,Y)} \Psi(Y_i, Y_j) \quad (3.23)$$

όπου $\mathcal{E}_{\{(X,Y),(X,X),(Y,Y)\}}$ συμβολίζουν τα ζεύγη αλληλεπιδρώντων μονάδων και τα Φ, Ψ, Ξ καθορίζονται από τις σχέσεις (3.21),(3.22). Ένας συμβολισμός σχετικός με τον \mathcal{E} είναι ο $\mathcal{N}(j) = \{i : (i, j) \in \mathcal{E}\}$ που συμβολίζει τη γειτονιά του j .

Το πρόβλημα της εκμάθησης για την Boltzmann Machine τίθεται [98] ως η ελαχιστοποίηση της απόστασης Kullback-Leibler (KL) μεταξύ της πραγματικής κατανομής $P(Y)$ των παρατηρήσιμων μονάδων και της κατανομής τους $P_{BM}(Y) = \sum_X P_{BM}(Y, X)$ που προκύπτει από τη σχέση (3.22):

$$\begin{aligned} KL(P(Y) \| P_{BM}(Y)) &= \sum_Y P(Y) \log \frac{P(Y)}{P_{BM}(Y)} \\ &= \sum_Y P(Y) \log P(Y) - \sum_Y P(Y) \log P_{BM}(Y) \end{aligned}$$

Στην παραπάνω εξίσωση μας ενδιαφέρει μόνο ο δεύτερος όρος, καθώς ο πρώτος είναι ανεξάρτητος από την Boltzmann Machine και ισούται με την εντροπία των Y . Η εκμάθηση μπορεί λοιπόν να θεωρηθεί ως η τροποποίηση των βαρών του δικτύου στη σχέση (3.21) ώστε να κατασκευαστεί μία κατανομή πιθανότητας που να ακολουθεί πιστά την κατανομή των παρατηρήσιμων μονάδων του δικτύου. Σημειώνουμε εδώ ότι για ένα διακριτό σύνολο δεδομένων η ελαχιστοποίηση της απόκλισης KL ανάγεται έτσι στο απλό κριτήριο μέγιστης πιθανοφάνειας.

Ο αλγόριθμος για την εκπαίδευση μηχανών Boltzmann [98] ανανεώνει τις τιμές των βαρών σύμφωνα με τη σχέση $\Delta W_{i,j} \propto \langle v_i v_j \rangle^+ - \langle v_i v_j \rangle^-$ όπου $\langle v_i v_j \rangle^+$ είναι η συσχέτιση των κόμβων v_i, v_j όταν οι παρατηρούμενοι κόμβοι είναι σταθεροί στις παρατηρούμενες τιμές τους και $\langle v_i v_j \rangle^-$ όταν το δίκτυο δεν έχει κάποια παρατήρηση και 'τρέχει ελεύθερο'. Η εκτίμηση αυτών των μέσων απαιτεί τη στοχαστική προσομοίωση του δικτύου που είναι υπολογιστικά απαιτητική και συνεπώς μη πρακτική για ένα δίκτυο τόσο μεγάλο όσο αυτό του μοντέλου μας. Προκύπτει λοιπόν η ανάγκη για λιγότερο ακριβείς αλλά πιο αποδοτικούς αλγόριθμους για συμπερασματολογία [232, 117] και όπως θα δούμε είναι ακριβώς αυτό που επιτελείται και από τους αλγόριθμους εξέλιξης που περιγράφησαν προηγουμένως.

3.4.2 Προσέγγιση Μέσου Πεδίου

Χρησιμοποιούμε εδώ την προσέγγιση μέσου πεδίου που παρουσιάστηκε στο κεφάλαιο 2· και εδώ έχουμε ένα σύνολο από παρατηρήσιμες μονάδες Y και κρυφές μονάδες X και θέλουμε να εκτιμήσουμε την κατανομή $P(X|Y)$. Ακολουθώντας τα ίδια βήματα καταλήγουμε στην ελαχιστοποίηση του κριτηρίου

$$J(Q) = \sum_i \sum_{X_i} Q(X_i) \log \Phi_i(X_i) + \sum_{\{i,j\} \in \mathcal{E}} Q(X_i, X_j) \log \Psi(X_i, X_j) + \sum_i Q_i(X_i) \log Q_i(X_i) \quad (3.24)$$

που προκύπτει από τον περιορισμό της αναζήτησης πάνω σε κατανομές της μορφής $Q(X) = \prod_i Q_i(X_i)$. όπου τα αθροίσματα πάνω στο διάνυσμα X έχουν συρρικνωθεί σε επιμέρους αθροίσματα πάνω σε μικρότερους όρους. Για να απλουσιεύσουμε τον συμβολισμό παραπάνω γράψαμε \mathcal{E} αντί του $\mathcal{E}_{(X,X)}$, αφήσαμε την σταθερά c και συνοψίσαμε όλες τις αλληλεπιδράσεις μεταξύ παρατηρήσιμων-κρυφών μονάδων για τον κόμβο i ως:

$$\sum_{j \in \mathcal{N}(i)} \sum_{X_j} Q_i(X_i) \log \Phi_{i,j}(X_i, X_j) = \sum_{X_i} Q_i(X_i) \log \Phi_i(X_i) \quad (3.25)$$

Αν οι τιμές των $Q_j(X_j)$, $j \in \mathcal{N}(i)$ θεωρηθούν σταθερές είναι εύκολο να εκτιμήσουμε τις τιμές των $Q_i(X_i = 0)$, $Q_i(X_i = 1)$ που μεγιστοποιούν το $J(Q)$. Η συνθήκη ακροτάτου για το $J(Q)$ είναι:

$$\begin{aligned} \frac{\partial J(Q)}{\partial Q_i(1)} &= 0 \rightarrow \\ \log(Q_i(1)) - \log(1 - Q_i(1)) &= \log \Phi_i(1) - \log \Phi_i(0) + \\ &\quad \sum_{j \in \mathcal{N}(i)} \sum_{X_j} Q_j(X_j) (\log \Psi_{i,j}(X_j, 1) - \log \Psi_{i,j}(X_j, 0)) \end{aligned}$$

Παραπάνω χρησιμοποιήσαμε τους όρους $Q_i(1)$, $\Phi_i(1)$ ως συντομεύσεις των $Q_i(X_i = 1)$, $\Phi_i(X_i = 1)$ και αντικαταστήσαμε το $Q_i(X_i = 0)$ με το $1 - Q_i(1)$. Στην περίπτωση που τα δυναμικά είναι της μορφής $\Psi_{i,j}(X_i, X_j) = \exp(C_{i,j} X_i X_j)$, $\Phi_i(X) = \exp(C_i X + b_i)$ η παραπάνω σχέση απλουσιεύεται ως εξής:

$$\begin{aligned} \log \frac{Q_i(1)}{1 - Q_i(1)} &= \sum_{j \in \mathcal{N}(i)} Q_j(1) C_{i,j} + C_i Q_i(1) + b_i \rightarrow \\ Q_i(1) &= \frac{1}{1 + \exp(-\sum_{j \in \mathcal{N}(i)} C_{i,j} Q_j(1) - C_i Q_i(1) - b_i)} \end{aligned} \quad (3.26)$$

Οι διαφορικές εξισώσεις για το στάδιο II (3.16) οδηγούν σε σταθερές καταστάσεις της μορφής (3.26), όπως φαίνεται θέτοντας την χρονική παράγωγο ίση με μηδέν και χρησιμοποιώντας την σχέση (3.10) για να εκφράσουμε την έξοδο των νευρώνων. Οι εξισώσεις εξέλιξης φέρνουν λοιπόν την κατανομή $Q(X) = \prod_i Q_i(X_i)$ πιο κοντά στην $P(X|Y)$ βάσει της απόστασης KL , όπου X είναι οι μονάδες του Σταδίου II και η ποσότητα $Q(X_i = 1)$ ισούται με $U_i = g(V_i)$.

Για το πρώτο στάδιο του δικτύου μας δεν είναι δυνατό να καταλήξουμε σε μία

τόσο άμεση ερμηνεία των διαφορικών εξισώσεων μέσω της θεωρίας MFA, καθώς η συνάρτηση Lyapunov που προκύπτει βάσει της εργασίας των [35] δεν μπορεί να γραφεί στην μορφή (3.21). Μπορούμε πάραυτα να θεωρούμε τις εξόδους των νευρώνων ως τις αναμενόμενες τιμές δυαδικών τυχαίων μεταβλητών που αντιστοιχούν στην παρουσία ή την απουσία μίας ακμής, οπότε και να εφαρμόσουμε το ίδιο σκεπτικό που εφαρμόζουμε για την εκπαίδευση στο στάδιο I.

3.4.3 Εκτίμηση των βαρών του Δικτύου

Σύμφωνα με την ερμηνεία του δικτύου μας ως μία μηχανή Boltzmann θεωρούμε τις ακόλουθες δύο περιπτώσεις:

- Διαδικασία μάθησης, όπου οι παρατηρήσιμες μονάδες είναι οι εντάσεις των pixels και οι εξοδοί του σταδίου II στην λεπτότερη κλίμακα, όπως αυτές καθορίζονται από τις κατατμήσεις των εικόνων από ανθρώπους.
- Διαδικασία δράσης, όπου οι παρατηρήσιμες μονάδες είναι μόνο οι εντάσεις των pixels.

Όπως προαναφέρθηκε, η εκπαίδευση του δικτύου επιζητά την ελαχιστοποίηση της απόστασης KL μεταξύ της εμπειρικής κατανομής των παρατηρήσεων (εικόνες και αντίστοιχες ακμές) και της κατανομής τους βάσει του δικτύου:

$$KL(P(Y)|P_{BM}(Y)) = \sum_Y P(Y) \log P(Y) - \sum_Y P(Y) \log P_{BM}(Y) \quad (3.27)$$

Αγνοώντας τον όρο $\sum_Y P(Y) \log P(Y)$ που δεν επηρεάζεται από την διαδικασία μάθησης εστιάζουμε στη μεγιστοποίηση της ποσότητας

$$\begin{aligned} L_0 &= \sum_{I,E} P(I, E) \log P_{BM}(I, E) \\ &= \sum_I \sum_E P(E|I)P(I) \log(P_{BM}(E|I)P_{BM}(I)) \end{aligned}$$

Ο δεύτερος όρος που προκύπτει σπάζοντας το γινόμενο μέσα στον λογάριθμο γράφεται ως

$$\sum_I \sum_E P(E|I)P(I) \log(P_{BM}(I)) = \sum_I P(I) \log(P_{BM}(I)) \quad (3.28)$$

Αυτός είναι ένας όρος ο οποίος όταν μεγιστοποιείται επιτρέπει στο δίκτυο να μοντελοποιήσει σωστά την κατανομή των εισόδων του. Στο γενικότερο πλαίσιο της έρευνάς μας ο όρος αυτός είναι σημαντικός, καθώς οδηγεί στη δημιουργία του μοντέλου σύνθεσης [51]· δεν μας ενδιαφέρει όμως αυτή η πτυχή του προβλήματος όταν κάνουμε ανίχνευση ακμών, οπότε τον αγνοούμε [117, 232]. Ο όρος που απομένει, $L_1 = \sum_I \sum_E P(E|I)P(I) \log(P_{BM}(E|I))$ δεν μπορεί να μεγιστοποιηθεί άμεσα, καθώς έχουμε μόνο ένα συγκεκριμένο σύνολο ζευγών εικόνας-ακμών. Για αυτό αντικαθιστούμε το άθροισμα πάνω σε όλα τα δυνατά ζεύγη εικόνων-ακμών με ένα πάνω στα διαθέσιμα:

$$L_1 \simeq L_2 = \sum_{I_{obs}} \sum_E P_{I_{obs}}(E) \log(P_{BM}(E|I)) \quad (3.29)$$

Με I_{obs} συμβολίζουμε μία εικόνα που περιλαμβάνεται στο σύνολο εκπαίδευσης και με $P_{I_{obs}}(E)$ την αντίστοιχη πιθανότητα ακμών.

Οι ποσότητες E, I που χρησιμοποιήθηκαν ως εδώ αντιστοιχούν σε ολόκληρες εικόνες και όχι σημεία τους. Καθώς όμως χρησιμοποιούμε την MFA αντικαθιστούμε την $P_{BM}(E|I_{obs})$ με την Q όταν έχει επιτευχθεί σύγκλιση $Q(E) = \prod_i Q_i(E_i)$, που μας δίνει την ακόλουθη απλούστευση:

$$L_2 \simeq L = \sum_{I_{obs}} \sum_i \sum_{E_i} P_{I_{obs}}(E_i) \log(Q_i(E_i)) \quad (3.30)$$

Αυτή την φορά όλοι οι όροι που περιλαμβάνονται στο κριτήριο που πρέπει να βελτιστοποιηθεί είναι διαθέσιμοι: οι εκφράσεις για τα $Q_i(E_i)$ καθώς και οι τοπικά ορισμένες πιθανότητες $P_{I_{obs}}(E_i)$.

Κατά τη διάρκεια της εκπαίδευσης γνωρίζουμε μόνο την επιθυμητή έξοδο του Σταδίου II στην λεπτότερη κλίμακα. Δεν γνωρίζουμε τις τιμές των μονάδων του σταδίου I ή των άλλων μονάδων του σταδίου II σε τραχύτερες κλίμακες. Κανονικά αυτό θα απαιτούσε την εφαρμογή σε κύκλους της διαδικασίας εκμάθησης, ξεκινώντας κάθε φορά από την τραχύτερη κλίμακα, κάνοντας όλα τα βήματα μέχρι να υπολογιστούν οι εκτιμήσεις για τις εξόδους του σταδίου II στην λεπτότερη κλίμακα και μετά να αλλάζουν οι συνδέσεις όλων των ενδιάμεσων κόμβων χρησιμοποιώντας έναν αλγόριθμο τύπου Back-Propagation [117]. Καθώς μία τέτοια λύση είναι εξαιρετικά πολύπλοκη για ένα ήδη πολύπλοκο δίκτυο όπως το δικό μας, μία απλούστερη εναλλακτική είναι να κατασκευάσουμε τις επιθυμητές εξόδους των ενδιάμεσων σταδίων, χρησιμοποιώντας την επιθυμητή έξοδο του δικτύου. Συγκεκριμένα για τις επιθυμητές εξόδους του σταδίου I χρησιμοποιούμε τους ιδανικούς χάρτες ακμών αφού έχουν φαρδυνθεί αρκετά ώστε να αντιστοιχούν στην κλίμακα στην οποία λειτουργεί το δίκτυο και για τις εξόδους του σταδίου II σε άλλες κλίμακες χρησιμοποιήσαμε τους ιδανικούς χάρτες ακμών καθώς υποτίθεται ότι το στάδιο αυτό δίνει λεπτούς χάρτες ακμών.

Για το Στάδιο I χρησιμοποιούμε τις τιμές σταθερής κατάστασης της (3.6) αντί της $Q_i(1)$ στην (3.30). Θεωρώντας την είσοδο του σταδίου I , και την επιθυμητή έξοδο P το κριτήριο που πρέπει να ελαχιστοποιηθεί γίνεται:

$$\begin{aligned} L(C) &= \sum_i P_i(1) \log \frac{I_i}{a + I_i + \sum_j C_j U_j} + P_i(0) \log \frac{a + \sum_j C_j U_j}{a + I_i + \sum_j C_j U_j} \\ &= \sum_i P_i(0) \log \left(a + \sum_j C_j U_j \right) \\ &\quad - (P_i(1) + P_i(0)) \log \left(a + I_i + \sum_j C_j U_j \right) + P_i(1) \log(I_i) \\ &= \sum_i P_i(0) \log \left(a + \sum_j C_j U_j \right) - \log \left(a + I_i + \sum_j C_j U_j \right) + c \end{aligned}$$

οπότε προκύπτει η εξής απλή σχέση για την εκμάθηση των βαρών του δικτύου:

$$\frac{\partial L(C)}{\partial C_j} = \sum_i P_i(0) \frac{U_j}{a + \sum_j C_j U_j} - \frac{U_j}{a + I_i + \sum_j C_j U_j} \quad (3.31)$$

Χάριν συντομίας, παραπάνω και στη συνέχεια χρησιμοποιούμε το συμβολισμό C_j αντί του C_{i-j} .

Για το Στάδιο II μπορούμε να υπολογίσουμε ένα κάτω φράγμα της αύξησης του κριτηρίου L που προκαλεί μία αλλαγή στο C το οποίο μπορούμε σε κάθε επανάληψη να το μεγιστοποιούμε. Η απόδειξη βασίζεται στον αλγόριθμο Improved Iterative Sampling [8] ο οποίος χρησιμοποιείται για εκτίμηση παραμέτρων σε Random Fields και περιέχεται στο τέλος του κεφαλαίου καθώς είναι μάλλον τεχνική. Το φράγμα δίνεται από την έκφραση:

$$\Delta L \geq - \sum_i P_i(0) \left(\sum_j \Delta C_j Q_j(1) \right) + 1 - Q_i(1) - Q_i(0) \sum_j \frac{Q_j(1)}{\Sigma Q_i} \exp(-\Sigma Q_i \Delta C_j)$$

η οποία έχει το χαρακτηριστικό ότι οι μερικές παράγωγοι του δεξιού μέρους ως προς τις συνιστώσες ΔC_j δεν είναι συζευγμένες, οπότε μπορούμε απευθείας να μεγιστοποιήσουμε το φράγμα κάνοντας gradient ascent:

$$\frac{\partial \Delta C_j}{\partial t} = \frac{\partial \Delta L'}{\partial \Delta C_j} = - \sum_i P_i(0) Q_j(1) + Q_i(0) Q_j(1) \exp(-\Sigma Q_i \Delta C_j) \quad (3.32)$$

3.5 Πειραματικά Αποτελέσματα

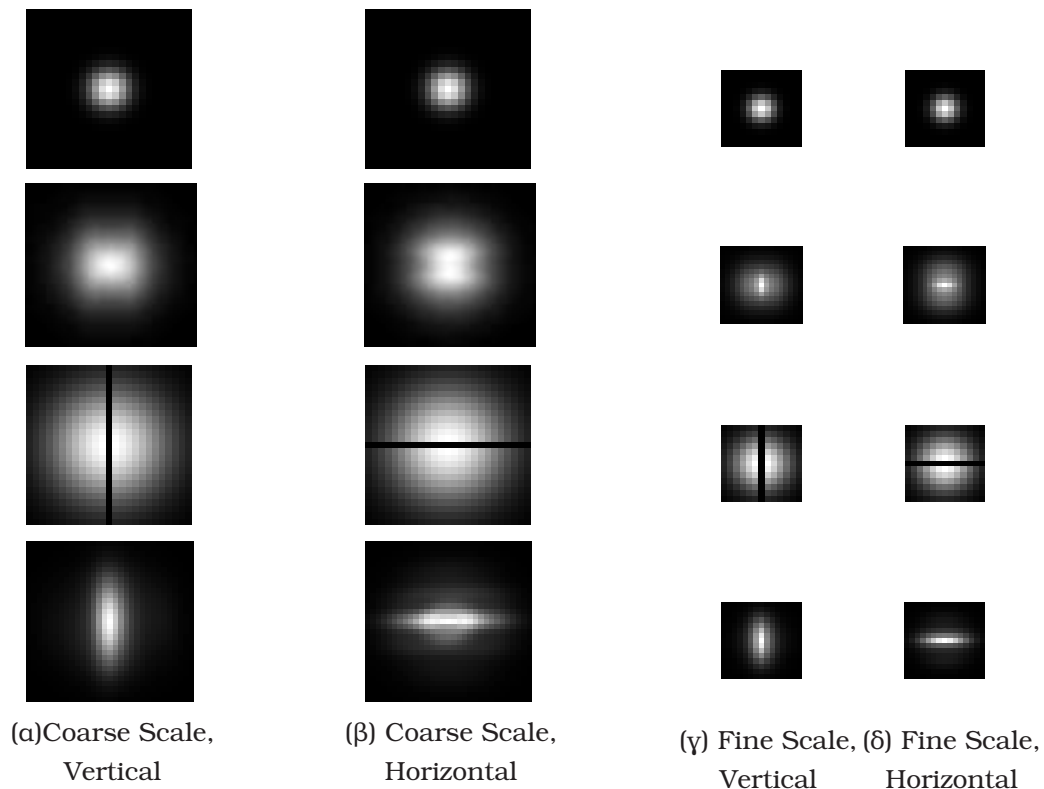
3.5.1 Αποτελέσματα Εκμάθησης Βαρών

Στο Σχ. 3.9 συγκρίνουμε τα βάρη που είχαν αρχικά επιλεγεί με τις ευρετικές επιλογές που είχαν γίνει αρχικά [126, 127] με αυτά που εκμαθεύτηκαν με βάση την μέθοδο της προηγούμενης παραγράφου. Από αυτά φαίνεται ότι αρκετές από τις επιλογές που έδειχναν διαισθητικά σωστές δεν είχαν αντίκρισμα στην πραγματικότητα. Για παράδειγμα, ενώ αρχικά θεωρήθηκε ότι νευρώνες που βρίσκονται κατά μήκος της ίδιας ευθείας δεν έχουν συνδέσεις αναστολής, από την εκμάθηση των βαρών βλέπουμε το αντίθετο. Ερμηνεύοντας εκ των υστέρων τον τύπο των συνδέσεων, μπορούμε να πούμε ότι το σήμα ανάδρασης που έρχεται από ένα άλλο στάδιο εξασφαλίζει την συνέχεια και την ομαλότητα των καμπυλών, οπότε πλέον οι ανασταλτικές συνδέσεις αποφεύγουν τις δυαδικές αποφάσεις. Καθώς με τα βάρη που έχουν εκμαθηθεί παρατηρείται στη συνέχεια μία ουσιαστική βελτίωση στις επιδόσεις, συμπεραίνουμε ότι ο αλγόριθμος εκμάθησης είναι απαραίτητος, δεδομένης της πολύπλοκης φύσης του δικτύου, που δεν επιτρέπει τη χρήση ευρετικών και 'διαίσθησης'.

Στο Σχ. 3.10 δείχνουμε τα αποτελέσματα του αλγόριθμου εκμάθησης χρησιμοποιώντας μία εικόνα που ανήκει στο σύνολο εκπαίδευσης. Παρατηρούμε ότι το δίκτυο αποφεύγει της απόλυτες αποφάσεις που έπαιρνε προηγουμένως και δίνει λιγότερα false alarms σε περιοχές με έντονη υφή. Επίσης βλέπουμε τα αποτελέσματα του αλγόριθμου για μία εικόνα που δεν ανήκει στο σύνολο εκπαίδευσης.

3.5.2 Αποτελέσματα Ανίχνευσης Ακμών

Για να συγκρίνουμε τις επιδόσεις του μοντέλου μας σε σχέση με κλασσικούς αλλά και πρόσφατους αλγόριθμους κατάτμησης εικόνων χρησιμοποιήσαμε τις εικόνες από το σύνολο δεδομένων του Berkeley [155] όπου φυσικές εικόνες έχουν γίνει διαθέσιμες μαζί με χειρωνακτικά καθορισμένες κατατμήσεις.

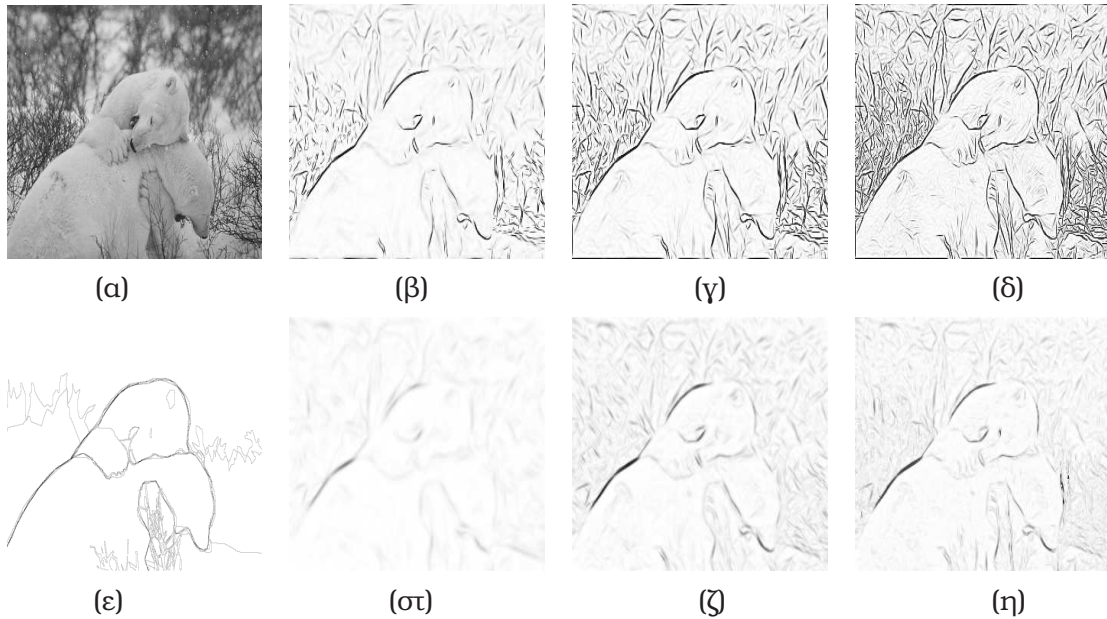


Σχήμα 3.9: Διαφορά μεταξύ των προκαθορισμένων βαρών και αυτών που έχουν εκμαθηθεί. Στις περιπτώσεις γραμμών δείχνουμε το προκαθορισμένο βάρος και στις άρτιες αυτά που εκμαθεύτηκαν· μία υψηλή τιμή φωτεινότητας δείχνει μία ισχυρή ανασταλτική σύνδεση μεταξύ του νευρώνα στο κέντρο και του νευρώνα στην αντίστοιχη θέση. Οι πρώτες δύο γραμμές δείχνουν βάρη για το στάδιο I και οι κάτω για το στάδιο II. Οι στήλες (α)/(β) αντιστοιχούν σε βάρη μεταξύ κάθετων/οριζόντιων νευρώνων σε μία τραχιά κλίμακα και οι (γ)/(δ) σε λεπτή κλίμακα.

Οι ποσότητες που χρησιμοποιούνται για να περιγράψουν επίδοση ενός ανιχνευτή ακμών είναι η ακρίβειά του (precision), P και η ανάκλησή του (recall), R . Εναλλακτικές όπως οι Receiver Operating Characteristic - ROC Curves δεν είναι κατάλληλες για την εκτίμηση της ποιότητας ανίχνευσης ακμών. Η ανάκληση ισούται με τον λόγο των σωστά εντοπισμένων στοιχείων ακμής προς το πλήθος στοιχείων ακμής που καθορίζουν οι ανθρώπινες κατατιμήσεις ($\simeq P(Detection|True)$) και η ακρίβεια ισούται με τον λόγο των σωστά ανιχνευθέντων pixels προς το πλήθος των ανιχνευθέντων pixels ($\simeq P(True|Detection)$). Για έναν ιδανικό ανιχνευτή και οι δύο αυτές ποσότητες θα έπρεπε να ισούνται με μονάδα, καθώς θα έπρεπε να ανιχνεύει τα πάντα $R = 1$, χωρίς να κάνει λάθη $P = 1$. Αυτές οι δύο ποσότητες συνδέονται αντίστροφα, και αλλάζοντας το κατώφλι του ανιχνευτή παίρνουν διαφορετικές τιμές, που μας δίνουν μία καμπύλη Precision-Recall - PR. Όταν δύο καμπύλες Precision-Recall δεν μπορούν να συγκριθούν εύκολα, π.χ. όταν τέμνονται, μία χρήσιμη έκφραση για να συνοψίσει την επίδοση των αντίστοιχων ανιχνευτών είναι η F -μέτρηση, που ορίζεται ως:

$$F = \frac{2}{(\alpha)P^{-1} + (1 - \alpha)R^{-1}}$$

όπου ο παράγοντας α τίθεται συνήθως ίσος με 0.5. Η μέγιστη F μέτρηση της καμπύ-



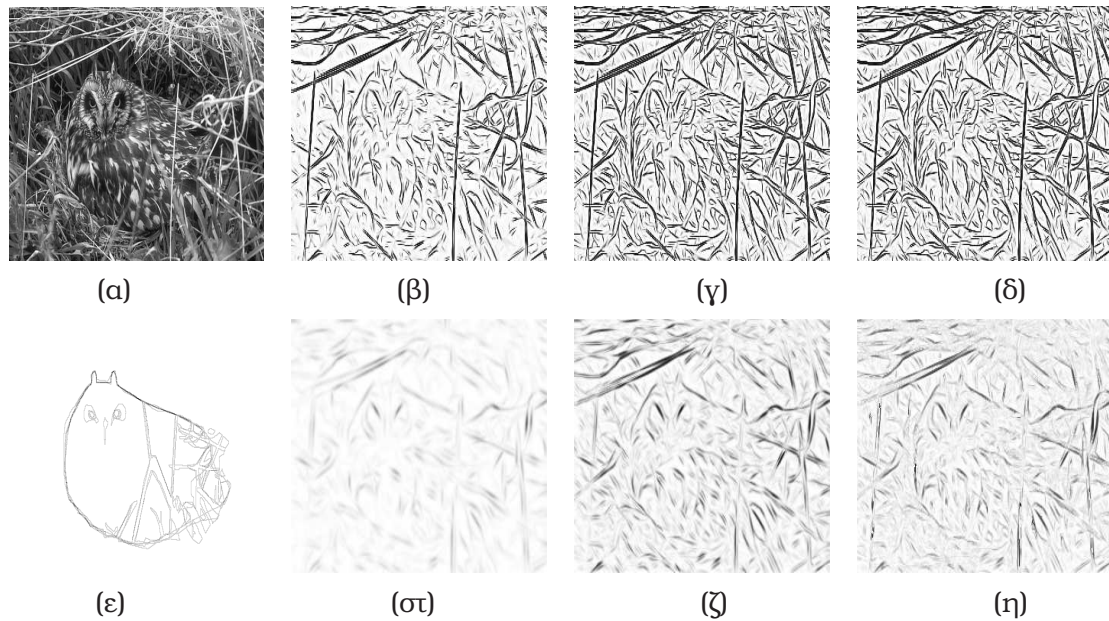
Σχήμα 3.10: Εκμάθηση ανίχνευσης ακμών: (α) Εικόνα στην είσοδο του συστήματος, (β)-(δ) πιθανότητα ύπαρξης ακμών, χρησιμοποιώντας ευρετικές εκφράσεις για τα βάρη, σε αυξανόμενες αναλύσεις της εικόνας (ε) πιθανότητα ακμής βάσει καταμήσεων ανθρώπων (στ)-(η) πιθανότητα ακμής με βάση τα βάρη που προέκυψαν από τον αλγόριθμο εκμάθησης.

λης μπορεί να χρησιμοποιηθεί για την σύγκριση δύο ανιχνευτών, καθώς μεγαλύτερες τιμές είναι πιο επιθυμητές.

Στο Σχ. 3.12 φαίνονται οι καμπύλες Precision-Recall για διάφορους αλγόριθμους ανίχνευσης ακμών. Στην λεζάντα κάθε εικόνας φαίνεται η μέγιστη F -μέτρηση κάθε καμπύλης καθώς και οι συντεταγμένες του μέγιστου. Συγκεκριμένα βλέπουμε:

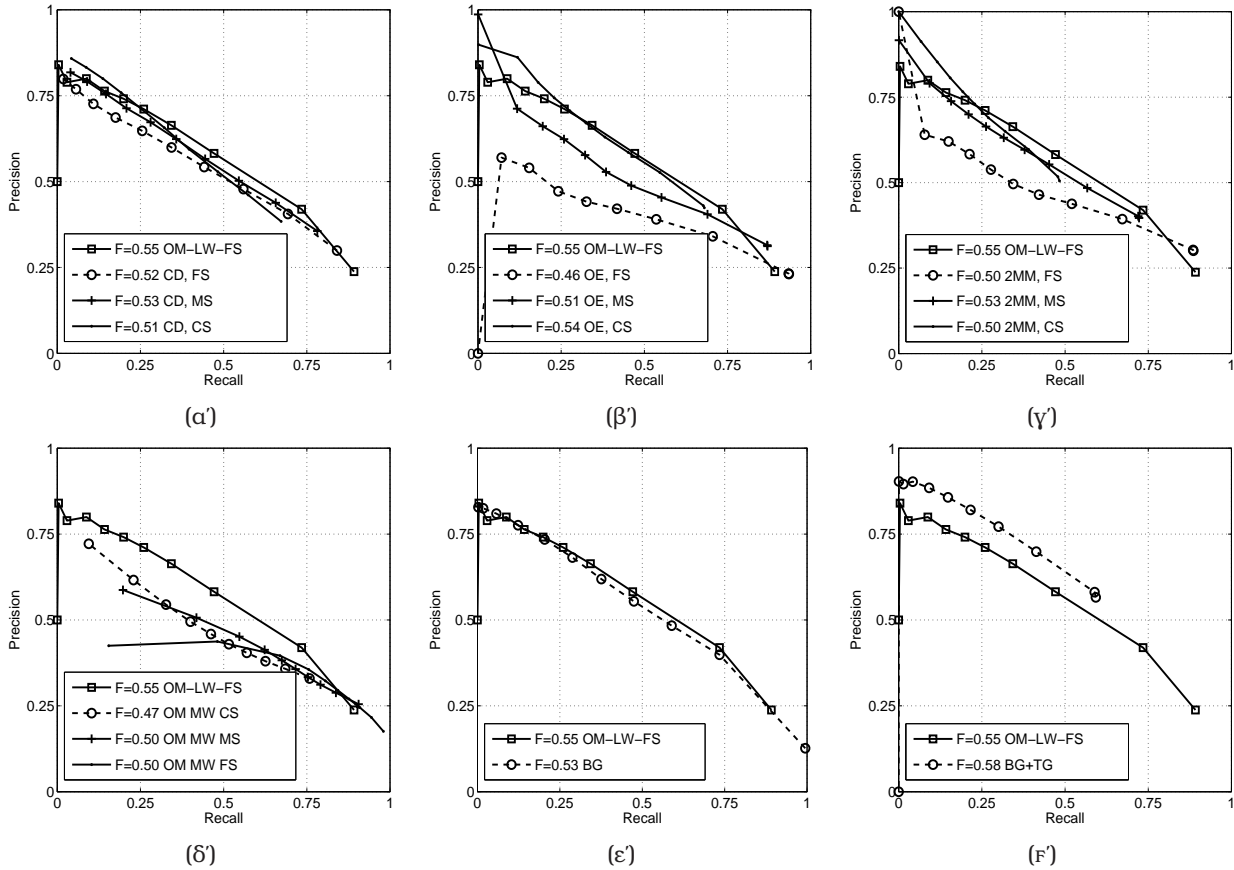
- Το σύστημά μας σε σύγκριση με την ανίχνευση ακμών Canny σε πολλαπλές κλίμακες[28].
- Το σύστημά μας σε σύγκριση με την μέθοδο ενέργειας-σε-κατεύθυνση (Oriented Energy- (OE)) [186] σε πολλαπλές κλίμακες.
- Το σύστημά μας σε σύγκριση με την μέθοδο πίνακα δεύτερων ροπών (Second moment matrix) σε πολλαπλές κλίμακες.
- Την επίδοση του συστήματός μας πριν και μετά από την εκμάθηση των παραμέτρων του.
- Την επίδοση του συστήματός σε σχέση με τη μέθοδο παραγώγου φωτεινότητας [154].
- Το σύστημά μας σε σύγκριση με την μέθοδο παραγώγου φωτεινότητας και υψής [154].

Παρατηρούμε ότι το σύστημά μας ξεπερνάει σε επιδόσεις όλους τους κλασσικούς αλγόριθμους ανίχνευσης ακμών, που είναι αναμενόμενο, καθώς συνδυάζει πληροφορία από πολλαπλές κλίμακες, region-based πληροφορία καθώς και πληροφορία



Σχήμα 3.11: Ανίχνευση ακμών σε μία εικόνα που δεν ανήκει στο σύνολο εκπαίδευσης: (α) Εικόνα στην είσοδο του συστήματος, (β)-(δ) πιθανότητα ύπαρξης ακμών, χρησιμοποιώντας ευρετικές εκφράσεις για τα βάρη, σε αυξανόμενες αναλύσεις της εικόνας (ε) πιθανότητα ακμής βάσει κατατιμήσεων ανθρώπων (στ)-(η) πιθανότητα ακμής με βάση τα βάρη που προέκυψαν από τον αλγόριθμο εκμάθησης.

σχετική με την saliency των ακμών. Παρατηρούμε επίσης ότι λειτουργεί αρκετά καλά για ένα μεγάλο διάστημα τιμών ανάκλησης, ενώ οι περισσότεροι άλλοι ανιχνευτές λειτουργούν σε περιορισμένες περιοχές. Αυτό μπορεί να ερμηνευθεί ως συνέπεια της λειτουργίας του συστήματος σε πολλαπλές κλίμακες. Παρατηρούμε όμως ότι δεν παίρνουμε τόσο καλά αποτελέσματα όσο με τη μέθοδο των [154] που συνδυάζει φωτεινότητα με πληροφορία υψής. Και αυτό όμως είναι αναμενόμενο, καθώς το σύστημά μας δεν χρησιμοποιεί πληροφορία σχετικά με υφή, παρά μόνο στο στάδιο I, όπου απλά αποφεύγεται ο εντοπισμός ακμών μέσα στην υφή μέσω της shunting inhibition.



CD	Canny Detection	OE	Oriented Energy
2MM	Second Moment Matrix	OM	Our Model
LW	Learned Weights	MW	Manually set Weights
BG	Brightness Gradient	BG+TG	Brightness & Texture Gradient

Σχήμα 3.12: Συγκριτικά αποτελέσματα για ανίχνευση ακμών, για καθιερωμένες μεθόδους ανίχνευσης ακμών και το δίκτυό μας, σε διαφορετικές κλίμακες· οι συντομεύσεις {F,M,C}S σημαίνουν Fine-, Mid- και Coarse- Scale αντίστοιχα. Παρατηρούμε ότι το σύστημά μας έχει καλύτερες επιδόσεις από τις περισσότερες μεθόδους, ενώ η μέθοδος του Brightness & Texture Gradient των [154] που έχει καλύτερες επιδόσεις χρησιμοποιεί και χαρακτηριστικά υψής στην ανίχνευση ακμών.

3.Α' Υπολογισμός Κάτω Φράγματος για τον Αλγόριθμο Εκμάθησης Βαρών

Για απλότητα δεν χρησιμοποιούμε τους όρους δυναμικού παρατήρησης που εμφανίζονται στη σχέση (3.26)-τα παρακάτω μένουν απαράλλαχτα στην περίπτωση που συμπεριληφθούν. Χρησιμοποιώντας τη σταθερή τιμή της εξίσωσης (3.16) αντί του $Q_i(1)$ έχουμε:

$$Q_i(1) = \frac{1}{1 + \exp(-\sum_{j \in \mathcal{N}(i)} C_j Q_j(1))}$$

$$Q_i(0) = \frac{\exp(-\sum_{j \in \mathcal{N}(i)} C_j Q_j(1))}{1 + \exp(-\sum_{j \in \mathcal{N}(i)} C_j Q_j(1))}$$

Συνεπώς, ο όρος που θέλουμε να μεγιστοποιηθεί γράφεται:

$$L(C) = \sum_i P_i(1) \log \frac{1}{1 + \exp(-\sum_{j \in \mathcal{N}(i)} C_j Q_j(1))} + P_i(0) \log \frac{\exp(-\sum_{j \in \mathcal{N}(i)} C_j Q_j(1))}{1 + \exp(-\sum_{j \in \mathcal{N}(i)} C_j Q_j(1))} \quad (3.33)$$

Μεταβάλλοντας κατά ΔC τις παραμέτρους του δικτύου έχουμε:

$$\begin{aligned} \Delta L(\Delta C) &= L(C + \Delta C) - L(C) \\ &= \sum_i P_i(1) \log \frac{1 + \exp(-\sum_{j \in \mathcal{N}(i)} C_j Q_j(1))}{1 + \exp(-\sum_{j \in \mathcal{N}(i)} (C_j + \Delta C_j) Q_j(1))} \\ &\quad + P_i(0) \log \left(\frac{\exp(-\sum_{j \in \mathcal{N}(i)} (C_j + \Delta C_j) Q_j(1))}{\exp(-\sum_{j \in \mathcal{N}(i)} C_j Q_j(1))} \right. \\ &\quad \left. \frac{1 + \exp(-\sum_{j \in \mathcal{N}(i)} C_j Q_j(1))}{1 + \exp(-\sum_{j \in \mathcal{N}(i)} (C_j + \Delta C_j) Q_j(1))} \right) \\ &= - \sum_i P_i(0) \sum_j \Delta C_j Q_j(1) - (P_i(0) + P_i(1)) \log \frac{1 + \exp(-\sum_{j \in \mathcal{N}(i)} (C_j + \Delta C_j) Q_j(1))}{1 + \exp(-\sum_{j \in \mathcal{N}(i)} C_j Q_j(1))} \\ &= - \sum_i P_i(0) \sum_j \Delta C_j Q_j(1) - \log \left[Q_i(1) + Q_i(0) \exp(-\sum_{j \in \mathcal{N}(i)} \Delta C_j Q_j(1)) \right] \end{aligned}$$

Στο πρώτο βήμα χρησιμοποιήσαμε την έκφραση του $L(C)$ στην (3.33), στο δεύτερο βασικές ιδιότητες της συνάρτησης \log και στο τέλος εκμεταλευόμαστε τις εκφράσεις για τα $Q_i(1)$ και $Q_i(0)$ που χρησιμοποιήσαμε και αρχικά.

Χρησιμοποιώντας τα βήματα του αλγόριθμου Improved Iterative Scaling [8], θα βρούμε ένα κάτω φράγμα του J : το πρώτο κάτω φράγμα προκύπτει εφαρμόζοντας της ανισότητα $-\log(a) \geq 1 - a$ στον τελευταίο όρο του J , που δίνει

$$\Delta L \geq - \sum_i P_i(0) \left(\sum_j \Delta C_j Q_j(1) \right) + 1 - Q_i(1) - Q_i(0) \exp(-\sum_j \Delta C_j Q_j(1))$$

Το δεύτερο φράγμα βασίζεται στην ανισότητα του Jensen, σύμφωνα με την οποία αν $\sum p_i = 1$, $p_i \geq 0 \quad \forall i$ έχουμε $\exp(\sum p_i q_i) \leq \sum p_i \exp(q_i)$. Εισάγουμε το

συμβολισμό $\Sigma Q_i = \sum_{j \in \mathcal{N}(i)} Q_j(1)$ οπότε έχουμε:

$$\exp\left(\sum_j \Delta C_j Q_j(1)\right) = \exp\left(-\Sigma Q_i \sum_j \frac{Q_j(1)}{\Sigma Q_i} \Delta C_j\right) \leq \sum_j \frac{Q_j(1)}{\Sigma Q_i} \exp\left(-\Sigma Q_i \sum_j \Delta C_j\right)$$

Συνεπώς έχουμε

$$\Delta L \geq - \sum_i P_i(0) \left(\sum_j \Delta C_j Q_j(1) \right) + 1 - Q_i(1) - Q_i(0) \sum_j \frac{Q_j(1)}{\Sigma Q_i} \exp\left(-\Sigma Q_i \Delta C_j\right)$$

Κεφάλαιο 4

Ανάλυση και Κατάτμηση Εικόνων Υφής Μέσω Χαρακτηριστικών Διαμόρφωσης, Αναγεννητικών Μοντέλων και Σταθμισμένης Εξέλιξης Καμπυλών

4.1 Εισαγωγή

Η υφή αποτελεί μία ισχυρή πηγή πληροφορίας για προβλήματα όρασης υπολογιστών όπως κατάτμηση, εκτίμηση σχήματος/βάθους/κίνησης, ανάκληση εικόνων από βάσεις δεδομένων, και αναγνώριση αντικειμένων. Το πρόβλημα που εμφανίζεται με τις εικόνες υφής είναι ότι καθώς οι περισσότεροι αλγόριθμοι της όρασης υπολογιστών απαιτούν τα αντικείμενα να έχουν ομαλή δομή στο εσωτερικό τους, η φωτεινότητα της εικόνας, που είναι αυτό που έχουμε διαθέσιμο, δεν είναι κατάλληλη ως χαρακτηριστικό της.

Απουσία ενός σαφούς μαθηματικού φορμαλισμού της υφής, κατά τη διάρκεια των προηγούμενων δεκαετιών αναπτύχθηκαν προσεγγίσεις που αντλούν ιδέες και τεχνικές από μία ποικιλία πεδίων. Κάνοντας μία γενική ταξινόμηση της έρευνας των προηγούμενων δεκαετιών, το πρόβλημα αντιμετωπίστηκε αναζητώντας είτε χαρακτηριστικά υφής τα οποία προσφέρουν μία συμπαγή και ομαλή περιγραφή της εικόνας π.χ. [21, 109, 22, 141, 248, 93, 178]. ή μοντέλα υφής, όπου το πρόβλημα της ανάλυσης υφής ανάγεται στην εκτίμηση των παραμέτρων ενός μοντέλου το οποίο μπορεί να ερμηνεύσει την δομή της εικόνας, π.χ. [47, 251, 93].

Δημοφιλέστερες σε συνδυασμό με το πρόβλημα της κατάτμησης εικόνων είναι οι τεχνικές που βασίζονται στην εξαγωγή χαρακτηριστικών. Η συνήθης προσέγγιση χρησιμοποιεί συστοιχίες γραμμικών φίλτρων τα οποία συνελισσόμενα με την εικόνα παρέχουν σήματα που προσφέρονται ευκολότερα στην ανάλυση, π.χ. [21, 109, 22, 93]. Ορισμένες επιθυμητές ιδιότητες τέτοιων χαρακτηριστικών είναι

- Χωρική επικέντρωση (spatial localization)
- Φασματική επικέντρωση (spectral localization)
- Χρήση πολλαπλών κλιμάκων.

Τα φίλτρα Gabor [73] ικανοποιούν με ένα βέλτιστο τρόπο τις πρώτες δύο απαιτήσεις καθώς μεγιστοποιούν ταυτόχρονα την χωρική και φασματική επικέντρωση [49] ενώ η λειτουργία των νευρώνων στα πρώτα στάδια του οπτικού συστήματος μπορεί να

μοντελοποιηθεί ικανοποιητικά ως συνέλιξη με ένα διδιάστατο φίλτρο Gabor. Για να καλυφθούν όλες οι δυνατές κλίμακες και κατευθύνσεις μία συστοιχία φίλτρων χρησιμοποιείται συνήθως, με το κάθε φίλτρο να εξειδικεύεται σε μία υποπεριοχή του πεδίου συχνότητων.

Από την δεύτερη κατηγορία γνωστότερα είναι τα μοντέλα τύπου MRF, που ερμηνεύουν την παρατηρούμενη εικόνα ως ένα δείγμα από την κατανομή πιθανότητας που ορίζει ένα MRF πάνω στο σύνολο παρατηρήσεων της εικόνας. Η τυπική προσέγγιση θεωρεί ότι το MRF χρησιμοποιεί ως κόμβους παρατήρησης τα μεμονωμένα pixels της εικόνας και τα δυναμικά ζεύγους ορίζονται βάσει π.χ. των διαφορών της φωτεινότητας. Η εκφραστική δυνατότητά τους έχει αναδειχθεί στην εργασία [251], όπου χρησιμοποιούνται συστοιχίες φίλτρων για να ορίσουν εμμέσως τα δυαδικά δυναμικά του MRF, εισάγοντας μία ενιαία θεώρηση των μεθόδων χαρακτηριστικών και μοντέλων υψής. Επεκτείνοντας αυτήν την ενοποίηση των μοντέλων υψής, στην εργασία [88] προτάθηκε ο συνδυασμός μοντέλων τύπου MRF με απλά αναγεννητικά μοντέλα όπως τοπικές βάσεις PCA, παρουσιάζοντας ένα ενιαίο πλαίσιο για την μοντελοποίηση και ανάλυση υψής. Το μοντέλο MRF μπορούν να μοντελοποιούν με ενιαίο τρόπο στοχαστικά και περιοδικά σήματα διαφορετικής φύσης πληροφορία [206] μπορεί να ενσωματωθεί άμεσα. Οι δυνατότητές τους όμως έχουν μείνει σε μεγάλο βαθμό ανεκμετάλλευτες καθώς εν γένει απαιτούν τεχνικές δειγματοληψίας τόσο για την εκτίμηση των παραμέτρων τους όσο και για τη σύνθεση δειγμάτων από τις κατανομές τους.

Μία διαφορετική προσέγγιση στην αντιμετώπιση του προβλήματος έχει ξεκινήσει από την θεωρία των textons του B. Julesz [116], όπου τέθηκε το πρόβλημα της εύρεσης μίας καθολικής βάσης χαμηλού επιπέδου για την αναπαράσταση της υψής. Ξεκινώντας από έναν έμμεσο ορισμό των textons ως συστάδες στον χώρο απόκρισης μίας συστοιχίας φίλτρων, στις εργασίες [150, 154] επιχειρείται η σύμπτυξη της ποικιλίας εμφάνισης της υψής σε ένα διακριτό σύνολο από συστάδες και κατόπιν η χρήση κατανομών για διακριτά δεδομένα. Ακολουθώντας μία κομψότερη και πιο μακρόπνοη προσέγγιση, στην εργασία [250] τα textons διατυπώθηκαν μέσω αναγεννητικών μοντέλων, θεωρώντας ότι κάθε texton μπορεί να ερμηνεύσει ένα σύνολο από απλούστερα πρωταρχικά στοιχεία, όπως ράβδους ή ακμές, συνοψίζοντας την απόκριση πολλών φίλτρων σε μία κοινή αιτία. Έτσι τα textons προκύπτουν ως ένας κώδικας που αναπτύσσεται αυτόματα για την αναπαράσταση της εικόνας με ένα περιεκτικό σε πληροφορία 'λεξιλόγιο', κατά έναν ανάλογο τρόπο με αυτόν όπου μία λέξη ερμηνεύει περισσότερα στοιχειώδη φωνήματα.

Ερχόμενοι στο πρόβλημα της κατάτμησης εικόνων, το κύριο πρόβλημα είναι η εξαγωγή μίας συμπαγούς περιγραφής βάσει της οποίας να μπορούν να ομαδοποιηθούν σωστά οι περιοχές της εικόνας. Για αυτόν τον στόχο οι τεχνικές που βασίζονται σε συστοιχίες φίλτρων είναι προβληματικές, καθώς εξάγουν χαρακτηριστικά μεγάλης διάστασης, τα οποία εύκολα οδηγούν σε υποβέλτιστες κατατμήσεις. Οι τεχνικές που βασίζονται σε μοντέλα υστερούν καθώς αν η κατάτμηση εκτιμάται δυναμικά, οι παράμετροι του μοντέλου πρέπει να εκτιμώνται παράλληλα με την κατάτμηση για κάθε περιοχή, καταλήγοντας σε μία υπολογιστικά ασύμφορη λύση. Τέλος τα textons εντοπίζονται σε διακριτές θέσεις στην εικόνα, ενώ οι περισσότεροι αλγόριθμοι κατάτμησης απαιτούν ένα πεδίο παρατηρήσεων ορισμένο σε όλη την έκταση της εικόνας.

Στα πλαίσια της ερευνάς αυτής της διατριβής αναζητήθηκαν χαμηλής διάστασης χαρακτηριστικά, τα οποία να χρησιμεύουν ταυτόχρονα στην αναπαράσταση της

υφής και στην κατάτμησή της. Το μοντέλο πάνω στο οποίο εργαστήκαμε διαφέρει από την συνηθισμένη ιδέα της εξαγωγής χαρακτηριστικών υφής μέσω φιλτραρίσματος και βασίζεται στο μοντέλο AM-FM διαμορφώσεων για την ανάλυση εικόνων υφής και συγκεκριμένα, στην μέθοδο Dominant Components Analysis. Η βασική ιδέα της μεθόδου αυτής είναι να αναπαριστά τοπικά την υφή μέσω ενός ημιτονοειδούς σήματος, του οποίου οι παράμετροι εκτιμώνται με τεχνικές αποδιαμόρφωσης. Έτσι, αφενός έχουμε μία συμπαγή περιγραφή της υφής, μέσω των παραμέτρων του μοντέλου, αφετέρου έχουμε ένα μοντέλο του οποίου οι παράμετροι εκτιμώνται με ταχύτητα και ακρίβεια, και το οποίο μπορούμε να χρησιμοποιήσουμε για την εξαγωγή όρων πιθανοφάνειας.

4.1.1 Συνεισφορές

Η εργασία αυτή έγινε σε συνεργασία με τον ΥΔ Γιώργο Ευαγγελόπουλο στα πλαίσια της διδακτορικής του έρευνας· οι ερευνητικές συνεισφορές του γράφοντος είναι οι παρακάτω:

- Εισαγωγή μίας προσέγγισης *τοπικών αναγεννητικών μοντέλων* για την ερμηνεία λειτουργιών ανάλυσης τμημάτων της εικόνας με χωρικά επικεντρωμένα φίλτρα.
- Χρήση του παραπάνω πλαισίου για την διατύπωση της διαδικασίας επιλογής καναλιού της μεθόδου DCA ως μία διαδικασία εξέτασης πολλαπλών υποθέσεων, την ερμηνεία της συνέλιξης με φίλτρα Gabor ως εκτίμηση παραμέτρων και την πιθανοτική ερμηνεία της ενέργειας Teager.
- Διαχωρισμός μεταξύ περιοχών ακμών, υφής και ομαλής φωτεινότητας, χρησιμοποιώντας μία ενιαία πιθανοτική προσέγγιση για την μοντελοποίηση και των δύο κλάσεων σημάτων στο παραπάνω πλαίσιο.
- Κατάτμηση εικόνων χρησιμοποιώντας χαρακτηριστικά DCA και συνδυασμός πληροφορίας από ετερογενείς πηγές στην διαδικασία κατάτμησης μέσω *σταθμισμένης εξέλιξης καμπυλών*, με βάρη που καθορίζονται αυτόματα βάσει των παραπάνω μοντέλων.

4.2 Το μοντέλο AM-FM

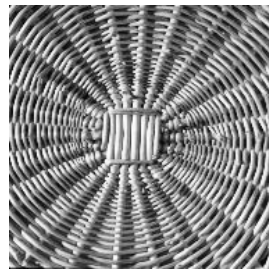
Μία μεγάλη ποικιλία σημάτων που απαντούν στον φυσικό κόσμο, όπως αυτά στο Σχ. 4.1 χαρακτηρίζονται από τοπικά ζωνοπερατή δομή· τέτοια σήματα μπορούν να μοντελοποιηθούν ικανοποιητικά μέσω σημάτων διαμόρφωσης πλάτους και συχνότητας (Amplitude Modulation-Frequency Modulation - AM-FM):

$$I(x, y) = a(x, y) \cos(\phi(x, y)) \quad (4.1)$$

όπου $a(x, y)$ είναι το *στιγμιαίο πλάτος* και $\phi(x, y)$ είναι η φάση του σήματος. Το στιγμιαίο πλάτος εκφράζει της διακυμάνσεις της περιβάλλουσας (envelope) του σήματος, ποσοτικοποιώντας την αντίθεση (contrast) της υφής, ενώ η φάση του περιγράφει την τοπική κυματοειδή δομή του σήματος δίνοντας πληροφορία για την κατεύθυνση και



(α)



(β)



(γ)

The number of possible intensity images notes the number of allowable gray levels. direct search, even for small ($m = 64$), t. Consequently, one is usually obliged to assumptions about the image and degradat as compromises at the computational si putational problem is overcome by expl servation that the posterior distribution i approximately the same neighborhood nal image, together with a sampling in the Gibbs Sampler. Indeed, our princ tribution is a general, practical, and mat approach for investigating MRF's by sai and by computing modes (Theorem

(δ)

Σχήμα 4.1: Τοπικά ζωνοπερατές εικόνες υφής.

την κλίμακα της υφής. Αντί της φάσης μία εναλλακτική ποσότητα που μπορεί να εξαχθεί ευκολότερα μέσω αποδιαμόρφωσης είναι η *στιγμαιαία συχνότητα*

$$\vec{\omega}(x, y) = \nabla\phi(x, y) = \left(\frac{\partial\phi}{\partial x}, \frac{\partial\phi}{\partial y} \right) \quad (4.2)$$

στον υπολογισμό της οποίας δεν υπεισέρχονται προβλήματα όπως η αναδίπλωση της φάσης στο 2π . Γενικότερα, τα σήματα διαμόρφωσης $a(x, y)$, $\vec{\omega}(x, y)$ θεωρούνται βαθυπερατά, επιτρέποντας την ερμηνεία απότομων μεταβολών της φωτεινότητας μέσω ομαλών συναρτήσεων που προσφέρονται ευκολότερα σε ανάλυση και στατιστική μοντελοποίηση.

Το σύνολο των σημάτων που μπορεί να εκφραστεί μέσω συναρτήσεων AM-FM μπορεί να επεκταθεί [93] σε όλα τα σήματα περιορισμένου εύρους ζώνης χρησιμοποιώντας αθροίσματα σημάτων AM-FM :

$$I(x, y) = \sum_{k=1}^K a_k(x, y) \cos(\phi_k(x, y)). \quad (4.3)$$

Η αποσύνθεση της εικόνας σε συνιστώσες AM-FM κατά την παραπάνω σχέση είναι ένα πολύπλοκο ανάστροφο πρόβλημα το οποίο έχει αντιμετωπιστεί μόνο μερικώς [93]: μία κοινή και πρακτική απλοποίηση είναι να χρησιμοποιείται ένα σύνολο από ζωνοπερατά φίλτρα όπως μία συστοιχία φίλτρων Gabor και να θεωρείται ότι υπάρχει μία αντιστοιχία συνιστωσών της (4.3) και εξόδων φίλτρων, οπότε η αποδιαμόρφωση γίνεται ξεχωριστά για την έξοδο κάθε φίλτρου (Σχ. 4.2). Η προσέγγιση αυτή έχει το πλεονέκτημα ότι χειρίζεται τα ομαλά σήματα που παίρνουμε από τις εξόδους των φίλτρων και γίνεται σημειακά, αποφεύγοντας πολύπλοκους αλγόριθμους για tracking των συνιστωσών. Το βασικό της πρόβλημα, το οποίο αντιμετωπίζει μερικώς η μέθοδος DCA είναι ότι δεν μπορεί να καθορίζει αυτόματα πόσες και ποιές συνιστώσες είναι ουσιαστικά ενεργές σε μία περιοχή της εικόνας, δίνοντας ένα μεγάλο πλήθος παραμέτρων.

4.2.1 Εκτίμηση Παραμέτρων

Έχοντας απομονώσει μία συνιστώσα της έκφρασης (4.3), το πρόβλημα της αποδιαμόρφωσης επιζητά την εξαγωγή συναρτήσεων $a(x, y)$, $\phi(x, y)$ που να συνδιάζουν πιστότητα στα δεδομένα, ανοχή στο θόρυβο και ομαλότητα. Για το πρόβλημα της αποδιαμόρφωσης εικόνων έχουν εφαρμοστεί τρεις διαφορετικές προσεγγίσεις.

A) Ο διδιάστατος αλγόριθμος Διαχωρισμού Ενέργειας [152](Energy Separation

Algorithm - ESA) βασίζεται στον 2-D ενεργειακό τελεστή Teager-Kaiser:

$$\Psi(I) \triangleq \|\nabla I\|^2 - I\nabla^2 I \quad (4.4)$$

ο οποίος εφαρμοζόμενος σε ένα σήμα AM-FM δίνει στην έξοδό του την ενέργεια διαμόρφωσης:

$$\Psi[a \cos(\phi)] \simeq a^2 |\vec{\omega}|^2 \quad (4.5)$$

Το σφάλμα στην προσέγγιση είναι αμελητέο αν θεωρηθεί ότι τα σήματα διαμόρφωσης πλάτους και συχνότητας δεν μεταβάλλονται γρήγορα στο χώρο και πολύ έντονα σε σχέση με το φέρον σήμα.

Εφαρμόζοντας την (4.4) στις παραγώγους του σήματος, παίρνουμε εκφράσεις από τις οποίες μπορούμε να εκτιμήσουμε τα σήματα διαμόρφωσης:

$$\frac{\Psi(f)}{\sqrt{\Psi(f_x) + \Psi(f_y)}} \approx |a(x, y)|, \quad \sqrt{\frac{\Psi(f_x)}{\Psi(f)}} \approx |\omega_1(x, y)|, \quad \sqrt{\frac{\Psi(f_y)}{\Psi(f)}} \approx |\omega_2(x, y)|, \quad (4.6)$$

Η απλότητα και κομψότητα αυτού του αλγορίθμου συνδυάζεται με την υπολογιστική του αποδοτικότητα, καθώς στην υλοποίησή του χρειάζεται μόνο διαφορές μεταξύ σημείων. Σημειώνουμε εδώ ότι έτσι υπεισέρχονται προβλήματα διακριτοποίησης των παραγώγων· τέτοια προβλήματα μπορούν να αντιμετωπιστούν χρησιμοποιώντας τις συνεχείς παραγώγους των φίλτρων Gabor αντί για τις διακριτές διαφορές των εξόδων τους [54, 56, 130].

B) Η μέθοδος Αναλυτικής Εικόνας [92] υπολογίζει τις AM-FM συνιστώσες του σήματος χρησιμοποιώντας μετασχηματισμούς στο πεδίο της συχνότητας. Συγκεκριμένα, η αναλυτική εικόνα J της I δίνεται από την έκφραση

$$J = I + j\hat{I} \quad (4.7)$$

όπου \hat{I} είναι ο μετασχηματισμός Hilbert της I . Στην περίπτωση που το σήμα I είναι τοπικά ημιτονοειδές, το μέτρο της αναλυτικής εικόνας ισούται με το πλάτος του ημιτόνου. Με βάση την αναλυτική εικόνα μπορούν να υπολογιστούν επίσης τα μέτρα και τα πρόσημα των κατευθυντικών παραγώγων της φάσης [92].

Έναλλακτικά, για την αποδιαμόρφωση της εξόδου ενός Ορθογωνικού Ζεύγους Φίλτρων (Quadrature Filter Pairs-QFP) [2, 94, 68, 21] μπορούμε να παρακάμψουμε την κατασκευή της εικόνας Hilbert. Ένα QFP αποτελείται από ένα ένα άρτιο και ένα περιττό φίλτρο που έχουν μηδενική απόκριση συχνότητας σε σταθερά σήματα και κοινό πλάτος απόκρισης συχνότητας, δηλαδή ένα ζεύγος Hilbert. Βάσει της γραμμικότητας του μετασχηματισμού Fourier, οι έξοδοι του ζεύγους αποτελούν ένα ζεύγος Hilbert, οπότε από αυτές μπορούμε απευθείας να εξάγουμε τις μετρήσεις φάσης και πλάτους. Τα ζεύγη άρτιων/περιττών φίλτρων Gabor μπορούν να θεωρηθούν προσεγγιστικά ως ένα QFP, αν θεωρήσουμε προσεγγιστικά μηδενικές τις αρνητικές συνιστώσες του φάσματος του μιγαδικού φίλτρου $h_e + jh_o$. Το πρόβλημα αυτής της προσέγγισης είναι ότι μας παρέχει μία μέτρηση φάσης σε κάθε σημείο, οπότε για τις στιγμιαίες συχνότητες χρειάζεται τόσο η διακριτοποίηση των παραγώγων όσο και ο χειρισμός του προβλήματος αναδίπλωσης της φάσης.

Η μεταβολική μέθοδος [22] εισάγει μέσω ενός μεταβολικού κριτηρίου στην διαδικασία αποδιαμόρφωσης απευθείας τον περιορισμό τα σήματα $\vec{\omega}(x, y)$, $a(x, y)$ να είναι

ομαλά, οπότε το πρόβλημα της αποδιαμόρφωσης ανάγεται στην αριθμητική επίλυση ενός συστήματος διαφορικών εξισώσεων. Αυτή η προσέγγιση είναι πιο ευέλικτη και προσφέρεται για εισαγωγή επιπλέον γνώσης, όπως όταν γνωρίζουμε ότι για ένα μέρος της εικόνας χρειάζεται μικρότερη πιστότητα στα δεδομένα, οπότε η εκτίμηση των παραμέτρων βασίζεται περισσότερο στην ομαλότητα. Για το πρόβλημα της αποδοτικής εξαγωγής χαρακτηριστικών υψής που ενδιαφερόμαστε όμως η μέθοδος αυτή είναι ακατάλληλη καθώς απαιτεί έναν επαναληπτικό αλγόριθμο ενώ οι υπόλοιπες παρέχουν απευθείας μία λύση.

Στα αποτελέσματα που φαίνονται στη συνέχεια του κεφαλαίου έχει χρησιμοποιηθεί ο αλγόριθμος ESA οπουδήποτε χρειάζεται η εξαγωγή χαρακτηριστικών διαμόρφωσης. Στις εικόνες που δείχνουν της ανακατασκευή των σημάτων, όπου χρειάζεται και η εκτίμηση της φάσης, χρησιμοποιούμε τη μέτρησή της που μας δίνει το ζεύγος φίλτρων Gabor, σύμφωνα με την μέθοδο αναλυτικής εικόνας.

Για την αποσύνθεση της εικόνας σε συνιστώσες AM-FM όπως προαναφέρθηκε θεωρούμε ότι η έξοδος κάθε φίλτρου Gabor απομονώνει μία συνιστώσα του σήματος η οποία εν συνεχεία αποδιαμορφώνεται. Ένα πρακτικό πρόβλημα είναι ότι καθώς η φέρουσα συχνότητα της συνιστώσας k τους σήματος μπορεί να μην συμπίπτει με την προκαθορισμένη κεντρική συχνότητα του φίλτρου Gabor, είναι αναμενόμενο το πλάτος A_k της συνιστώσας να συμπιέζεται, δίνοντας μία χαμηλότερη εκτίμησή του, a_k . Μπορούμε μερικώς να αναιρέσουμε αυτήν τη συμπεριφορά, υπολογίζοντας την απόκριση συχνότητας του φίλτρου στην εκτιμώμενη στιγμιαία συχνότητα του ημιτονοειδούς, $\vec{\omega}_k$: η σχέση

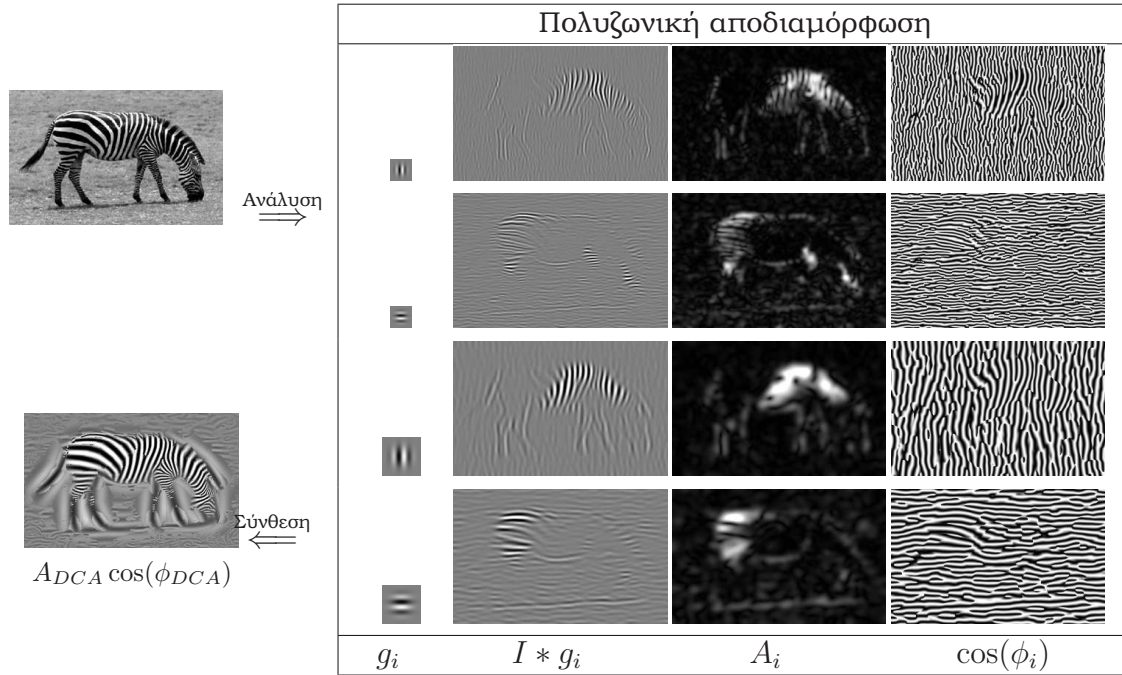
$$A_k = \frac{a_k}{|G_k(\vec{\omega}_k)|}, \quad (4.8)$$

όπου G_k είναι η απόκριση συχνότητας του αντίστοιχου φίλτρου Gabor, μας δίνει μία ποσότητα που παραμένει σχετικά ανεπηρέαστη από τις αποκλίσεις της στιγμιαίας συχνότητας από την κεντρική συχνότητα του φίλτρου.

Ανάλυση Κυρίαρχων Συνιστωσών

Από τη διαδικασία αποδιαμόρφωσης που περιγράφηκε παραπάνω κάθε συνιστώσα k της (4.3) προσφέρει ένα τριδιάστατο διάνυσμα χαρακτηριστικών (A_k, ω_k) σε κάθε σημείο της εικόνας προκύπτει ένα διάνυσμα χαρακτηριστικών διάστασης $3 \times K$, όπου K είναι το πλήθος των φίλτρων Gabor.

Για εικόνες υψής όπως αυτές στο Σχ. 4.1 τοπικά το φάσμα του σήματος περιέχεται σε μία μικρή περιοχή, και μπορεί να αναπαρασταθεί μέσω μίας ή έστω λίγων συνιστωσών AM-FM · ενώ τα χαρακτηριστικά που εξάγουμε περιγράφουν και όλες τις υπόλοιπες μηδενικές συνιστώσες της (4.3). Προκύπτει λοιπόν η ανάγκη για τον εισαγωγή ενός αλγόριθμου ο οποίος να εντοπίζει και να απομονώνει τις σημαντικές συνιστώσες του σήματος. Η μέθοδος της Ανάλυσης Κυρίαρχων Συνιστωσών [93, 92] (Dominant Components Analysis- DCA) προτείνει την χρήση ενός τοπικού κριτηρίου για την επιλογή μίας συγκεκριμένης συνιστώσας, από την αποδιαμόρφωση της οποίας παίρνουμε ένα τριδιάστατο διάνυσμα χαρακτηριστικών AM-FM . Θεωρώντας ότι μία μόνο συνιστώσα της σχέσης (4.3) αρκεί για την περιγραφή της εικόνας, η DCA σε κάθε σημείο (x, y) βάσει ενός κριτηρίου $\Gamma_k(x, y)$ επιλέγει το κανάλι $i(x, y)$ που είναι πλησιέστερο στην συνιστώσα, αποδιαμορφώνει την έξοδό του και χρησιμο-



Σχήμα 4.2: Εξαγωγή χαρακτηριστικών με τη μέθοδο DCA: ένα σύνολο απο ζωνοπερατά φίλτρα Gabor χρησιμοποιείται αρχικά για να απομονωθούν και αποδιαμορφωθούν οι μεμονωμένες συνιστώσες της (4.3). Το κυριάρχο κανάλι επιλέγεται σε κάθε σημείο της εικόνας και οι παράμετροί του χρησιμοποιούνται για την περιγραφή της υφής. Η κυρίαρχη δομή του σήματος υφής συλλαμβάνεται έτσι από τις παραμέτρους της DCA.

ποιεί τα χαρακτηριστικά AM-FM για την αναπαράσταση της υφής:

$$i(x, y) = \arg \max_{1 \leq k \leq K} \{\Gamma_k(x, y)\}, \quad (4.9)$$

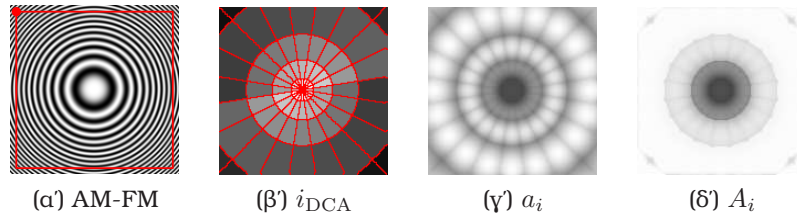
$$A_{DCA}(x, y) = A_{i(x,y)}(x, y), \quad \vec{\omega}_{DCA}(x, y) = \vec{\omega}_{i(x,y)}(x, y). \quad (4.10)$$

Το κριτήριο που χρησιμοποιείται τυπικά είναι το αποδιαμορφωμένο πλάτος του k -οστού καναλιού:

$$\Gamma_k(x, y) = |a_k(x, y)|, \quad (4.11)$$

το οποίο όπως παρουσιάζουμε στην επόμενη ενότητα μπορεί να αιτιολογηθεί πιθανοτικά. Στις εργασίες [128, 130] εξετάστηκε η δυνατότητα χρήσης της ενέργειας Teager ως ένα εναλλακτικό κριτήριο, το οποίο συνδυάζει πληροφορία για το πλάτος με πληροφορία για τη συχνότητα. Χρησιμοποιώντας το κριτήριο αυτό παίρνουμε μία εναλλακτική της Amplitude-based DCA - ADCA, την οποία ονομάζουμε Energy-based DCA-EDCA. Εφαρμόζοντας την EDCA σε φυσικές εικόνες παρατηρήσαμε [129] ότι δίνει καλύτερα αποτελέσματα, καθώς εντοπίζει με ακρίβεια απότομες αλλαγές στην εικόνα.

Οι διαφορές της DCA με τις υπόλοιπες μεθόδους εξαγωγής χαρακτηριστικών υφής μπορούν να αναδειχθούν χρησιμοποιώντας το χαρακτηριστικό σήμα στο Σχ. 4.3. Το σήμα αυτό έχει σταθερό πλάτος και περιέχει όλες τις χωρικές συχνότητες (ω_x, ω_y) μέσα στο φάσμα της εικόνας. Μία μέθοδος που χρησιμοποιεί textons θα έσπαγε την εικόνα σε τμήματα, κωδικοποιώντας ποιο texton τοπικά ταιριάζει καλύτερα με το σήμα της εικόνας, δίνοντας έναν διαμερισμό της εικόνας όπως στην εικόνα



Σχήμα 4.3: Χαρακτηριστικά υφής μέσω DCA και αποδιαμόρφωσης. **(α)** Συνθετικό σήμα υφής, **(β)** Κωδικοποίηση κυριάρχου καναλιού i_{DCA} μέσω φωτεινότητας. **(γ)** Εκτίμηση πλάτους a_i του κυριάρχου καναλιού και **(δ)** Εκτίμηση πλάτους A_i βάσει της σχέσης (4.8). Το εκτιμώμενο πλάτος κωδικοποιείται μέσω της τιμής της φωτεινότητας της εικόνας.

(β). Χρησιμοποιώντας μία συστοιχία φίλτρων θα διατηρούσαμε όλες τις αποκρίσεις φίλτρων σε κάθε σημείο, ακόμα και αν τα περισσότερα δεν παρέχουν καμία επιπλέον πληροφορία επιπλέον αυτής του πιο ενεργού φίλτρου.

Εξάγοντας τα χαρακτηριστικά της DCA αυτόματα επιλέγουμε μία μεμονωμένη συνιστώσα και εξάγουμε από αυτήν έναν διάνυσμα χαρακτηριστικών χαμηλής διάστασης που μεταβάλλεται ομαλά· συγκρίνοντας τις εικόνες (γ) και (δ) στο Σχ. 4.3 βλέπουμε ότι μετά την κανονικοποίηση των εξόδων των φίλτρων Gabor βάσει της (4.8) αποκτούμε μία εκτίμηση πλάτους πλησιέστερη στην πραγματική τιμή, που είναι παντού ίση με μονάδα. Ο αλγόριθμος αποδιαμόρφωσης μας επιτρέπει να μεταβούμε από το κβαντισμένο σύνολο κατευθύνσεων και κλιμάκων που χρησιμοποιούνται από την συστοιχία φίλτρων σε μία συνεχή αναπαράσταση. Πρακτικά μπορούμε να δούμε τη λειτουργία της μεθόδου DCA ως μία επέκταση των ιδεών των κατευθύνσιμων, κλιμακώσιμων (steerable, scalable) φίλτρων [68, 184] στα φίλτρα Gabor, καθώς από ένα πεπερασμένο σύνολο φίλτρων παίρνουμε μία ακριβέστερη, συνεχή περιγραφή.

4.3 Τοπικά Αναγεννητικά Μοντέλα για Υφή

Στην ενότητα αυτό διατυπώνουμε στο πλαίσιο των αναγεννητικών μοντέλων λειτουργίες που εμπλέκονται στην ανάλυση υφής. Συγκεκριμένα, η συνέλιξη με φίλτρα Gabor ερμηνεύεται ως εκτίμηση των παραμέτρων ενός μοντέλου που λαμβάνει υπόψη την τοπικότητα της ανάλυσης/μοντελοποίησης, ενώ το κριτήριο επιλογής καναλιού της μεθόδου DCA ανάγεται σε μία διαδικασία εξέτασης πολλαπλών υποθέσεων. Τέλος, βάσει αυτής της θεώρησης αιτιολογούμε τη χρήση της ενέργειας Teager θεωρώντας ότι στα δεδομένα έχει προστεθεί θόρυβος Brown. Η προσέγγιση αυτή συνδυάζει την αποδοτικότητα με την οποία μπορούν να γίνουν οι συνέλιξεις με την σαφήνεια των μεθόδων ανάλυσης μέσω μοντέλων, ενώ επιτρέπει τον συνδυασμό των μοντέλων υφής AM-FM με εναλλακτικά αναγεννητικά μοντέλα.

Η έρευνα στην κατεύθυνση της κατασκευής μοντέλων για γενικές εικόνες αποτυπώνεται σε αναφορές όπως [151, 6, 69, 88, 71], όπου το ενδιαφέρον έχει εστιάσει κυρίως στην αυτοματοποιημένη εκμάθηση μοντέλων που να ερμηνεύουν κατά έναν βέλτιστο τρόπο τις στατιστικές ιδιότητες των εικόνων ώστε να προσφέρονται για μία συμπαγή αναπαράσταση εικόνων και λειτουργίες όπως συμπίεση και αποθρομβοποίηση. Στα πλαίσια της εργασίας μας χρησιμοποιήσαμε αναλυτικώς ορισμένα αναγεννητικά μοντέλα, αφήνοντας την αυτόματη εκμάθησή τους από δεδομένα για

μελλοντική εργασία.

4.3.1 Τοπικά Αναγεννητικά Μοντέλα και Επιλογή Καναλιού

Ξεκινώντας από την υπόθεση που στηρίζει τη μέθοδο Dominant Components Analysis, ότι ένα σήμα υφής μπορεί τοπικά να μοντελοποιηθεί μέσω ζωνοπερατών σημάτων, διαιρούμε το πρόβλημα της μοντελοποίησης σε ένα σύνολο από απλούστερα προβλήματα. Θεωρούμε έναν διαμερισμό του πεδίου συχνότητας αντίστοιχο με αυτόν μίας συστοιχίας φίλτρων Gabor και για κάθε συνδυασμό κλίμακας και κατεύθυνσης που αντιστοιχεί στην κεντρική συχνότητα $\vec{\omega}$ του φίλτρου Gabor κατασκευάζουμε ένα αντίστοιχο μοντέλο υφής.

Απλοποιώντας τον συμβολισμό, κάνουμε την ανάλυσή μας κατά μήκος της εγκάρσιας τομής του σήματος κατά μήκος του $\vec{\omega}$, οπότε χρησιμοποιούμε και τη μονοδιάστατη συχνότητα $\omega = |\vec{\omega}|$, ενώ θεωρούμε ότι το μοντελοποιούμενο τμήμα είναι γύρω από το σημείο $x_0 = 0$. Για πιο περίπλοκα χαρακτηριστικά υφής, όπως διασταυρώσεις, κηλίδες, και γωνίες θα χρειαζόνταν διδιάστατα μοντέλα, αλλά αυτό θα μας απομάκρυνε από την DCA που μοντελοποιεί μονοδιάστατα, ημιτονοειδή χαρακτηριστικά.

Τοπική Μοντελοποίηση Υφής

Το μοντέλο που χρησιμοποιούμε θεωρεί ότι ένα σήμα υφής που επικεντρώνεται σε μία κλίμακα και κατεύθυνση μπορεί τοπικά να περιγραφεί μέσω μίας αρμονικής συνιστώσας, με την αντίστοιχη συχνότητα ω αλλά άγνωστη φάση ϕ , πλάτος A και σταθερή τιμή B :

$$I(x) \simeq I_T(x; \{A, \phi, B\}) = A \cos(\omega x + \phi) + B \quad (4.12)$$

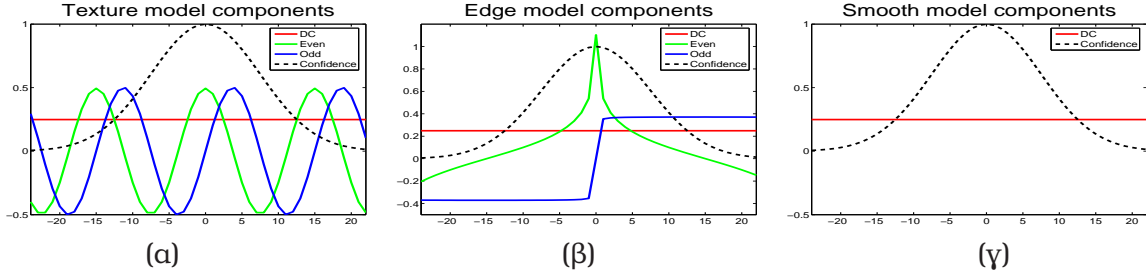
όπου ο δείκτης T συμβολίζει τη σύνθεση μέσω του μοντέλου υφής (Texture). Μπορούμε να εκφράσουμε τη σχέση σύνθεσης (4.12) ως ένα ανάπτυγμα σε μία γραμμική βάση:

$$I_T(x; \mathcal{A}) = \sum_{i=1}^{D_T} E_i B_{T,i}(x), \quad \text{όπου } D_T = 3 \quad (4.13)$$

$$\begin{aligned} E_1 &= A \cos(\phi), & E_2 &= -A \sin(\phi), & E_3 &= B \\ B_{T,1}(x) &= \cos(\omega x), & B_{T,2}(x) &= \sin(\omega x), & B_{T,3}(x) &= 1 \end{aligned} \quad (4.14)$$

Θα αναφερόμαστε στις παραπάνω συναρτήσεις ως το άρτιο, περιττό και σταθερό στοιχείο βάσης, αντίστοιχα.

Η ιδέα πίσω από τα *τοπικά αναγεννητικά μοντέλα* που προτείνουμε είναι να μοντελοποιηθεί άμεσα η εξάρτηση της ποιότητας ανακατασκευής στη σχέση (4.12) από την απόσταση x από το κεντρικό σημείο $x_0 = 0$. Αυτό αγνοείται σε αναφορές όπως [151, 88], όπου βάσει της υπόθεσης του λευκού Gaussian θορύβου, οι παράμετροι μέγιστης πιθανοφάνειας (ML) προκύπτουν από την μία προβολή ελαχίστων τετραγώνων του σήματος πάνω σε μία γραμμική βάση. Είναι όμως αντιδιδαισθητικό το ότι οι παρατηρήσεις στο όριο του τμήματος P παίζουν διακριτά διαφορετικό ρόλο από τις γειτονικές τους παρατηρήσεις, που δεν λαμβάνονται υπόψη στο κριτήριο ανακατασκευής.



Σχήμα 4.4: Τα στοιχεία της βάσης (συμπαγής γραμμή) και συνάρτηση εμπιστοσύνης (διακεκομμένη) για τις 1Δ τομές των τριών θεωρούμενων κλάσεων: (α) Υφή (β) Ακμή (γ) Ομαλά Τμήματα.

Στην προσέγγισή μας συνυπολογίζουμε την τοπικότητα της διαδικασίας απόφασης, χρησιμοποιώντας μία τιμή βεβαιότητας $G(x)$, π.χ. μία κανονικοποιημένη Gaussian συνάρτηση, που συνδέεται με τις προβλέψεις του μοντέλου στο σημείο x . Η ποσότητα $G(x)$ φθίνει με την απόσταση από το σημείο $x_0 = 0$ όπως φαίνεται στο Σχ. 4.4, και μία υπόθεση background εισάγεται για να αναλάβει την ερμηνεία της εμφάνισης της εικόνας μακριά από αυτό

Αυτό μπορεί να διατυπωθεί με σαφήνεια με την εισαγωγή μίας κρυφής δυαδικής τυχαίας μεταβλητής z_x η οποία καθορίζει το εάν η παρατήρηση στο σημείο x προέρχεται από την υπόθεση του foreground ($z_x = 1$) ή του background ($z_x = 0$) και ερμηνεύοντας την τιμή της συνάρτησης εμπιστοσύνης ως ίση με την αναμενόμενη τιμή του z_x : $P(z_x = 1|x) = G(x)$. Αναπαριστώντας της κατανομές του foreground και background ως $P_{T,f}, P_{T,b}$ αντίστοιχα, η πιθανοφάνεια της παρατήρησης $I(x)$ στο σημείο x για ένα σύνολο παραμέτρων \mathcal{A} γράφεται:

$$\begin{aligned}
 P_T(I(x)|x, \mathcal{A}) &= \sum_{z_x \in \{0,1\}} P_T(I(x), z_x|x, \mathcal{A}) \\
 &= \sum_{z_x \in \{0,1\}} P_T(I(x)|z_x, x, \mathcal{A})P(z_x|x) \\
 &= \underbrace{G(x)P_{T,f}(I(x)|I_T(x|\mathcal{A}))}_{z_x=1} + \underbrace{(1 - G(x))P_{T,b}(I(x))}_{z_x=0}. \tag{4.15}
 \end{aligned}$$

Η $P_{T,f}$ χρησιμοποιεί την σύνθεση του μοντέλου $I_T(x; \mathcal{A}_i)$ για να ερμηνεύσει τις παρατηρήσεις, ενώ η εναλλακτική υπόθεση $P_{T,b}$ μοντελοποιεί την φωτεινότητα της εικόνας ως μία ομοιόμορφη κατανομή, $P_{T,b}(I(x)) = c$. Παραπάνω χρησιμοποιήσαμε τη σχέση $P_T(z_x|x; \mathcal{A}) = P(z_x|x)$ καθώς η χωρική ελάττωση της εμπιστοσύνης στην πρόβλεψη του μοντέλου θεωρείται ότι δεν εξαρτάται από τις παραμέτρους του συγκεκριμένου μοντέλου και την κλάση του, εν προκειμένω, υφή.

Συγκρίνοντας με την προσέγγιση άλλων άρθρων, [151, 88], βλέπουμε ότι η παραπάνω διατύπωση αποσυνδέει διαισθητικά τις εκφράσεις σύνθεσης και πιθανοφάνειας, και εισάγει μεγαλύτερη ελαστικότητα και σαφήνεια, επιτρέποντας την κατασκευή απλών βάσεων για εξιδανικευμένα σήματα, όπως ημιτονοειδή. Αντιθέτως, στις εργασίες [151, 88] η τοπικότητα και η επιλεκτικότητα σε συγκεκριμένες κυματομορφές περιπλέκονται μέσω της εισαγωγής μίας χωρικά φθίνουσας περιβάλλουσας για τα στοιχεία βάσης.

Έκφραση Πιθανοφάνειας και Εκτίμηση Παραμέτρων

Θεωρώντας ότι το μοντέλο σφάλματος είναι μία διαδικασία λευκού θορύβου μπορούμε να γράψουμε:

$$\log P_T(I) = \sum_x \log P_T(I(x)|x). \quad (4.16)$$

Η έκφραση αυτή είναι δύσκολο να μεγιστοποιηθεί ως προς τις παραμέτρους του μοντέλου, καθώς το άθροισμα στη σχέση (4.15) είναι μέσα στον λογάριθμο. Για να φέρουμε το άθροισμα εκτός του λογαρίθμου χρησιμοποιούμε την κυρτότητα της συνάρτησης του λογαρίθμου και εφαρμόζουμε την ανισότητα του Jensen για να πάρουμε το παρακάτω κάτω φράγμα στο $\log P_T(I)$:

$$LB_T(I) = \sum_x G(x) \log P_{T,f}(I(x)|x) + \sum_x (1 - G(x)) \log c \quad (4.17)$$

Παρακάτω θα χρησιμοποιούμε αυτό το κάτω φράγμα αντί της αρχικής έκφρασης για την πιθανοφάνεια των δεδομένων. Όπως και στη μέθοδο του GLRT (2.36) αντί να ολοκληρώνουμε πάνω στις παραμέτρους, χρησιμοποιούμε την σημειακή εκτιμήτρια που ελαχιστοποιεί αυτό το κάτω φράγμα. Υπό την υπόθεση του Gaussian θορύβου με διασπορά σ^2 , ο πρώτος όρος της (4.17) μπορεί να γραφεί ως:

$$\sum_x G(x) \log P_{T,f}(I(x)|x) = -\frac{1}{2\sigma^2} \mathcal{R} - \sum_x G(x) \log(\sqrt{2\pi}\sigma), \quad (4.18)$$

$$\mathcal{R} = \sum_x G(x) [I(x) - I_T(x)]^2 = [\mathbf{I} - \mathbf{B}\mathcal{A}]^T \mathbf{G} [\mathbf{I} - \mathbf{B}\mathcal{A}], \quad (4.19)$$

$$\mathbf{G} : N \times N, \mathbf{I} : N \times 1, \mathbf{B} : N \times D_T, \mathcal{A} : D_T \times 1$$

όπου ο \mathbf{G} είναι ένας διαγώνιος πίνακας, με $\mathbf{G}_{i,i} = G(i)$, N το πλήθος των παρατηρήσεων, και χρησιμοποιούμε τους όρους $\mathbf{I}, \mathbf{B}, \mathcal{A}$ για να εκφράσουμε την (4.13) μέσω πινάκων.

Πέρα από το \mathcal{R} οι υπόλοιποι όροι είναι σταθεροί, οπότε η συνθήκη μέγιστου για την (4.18) δίνει τη λύση Σταθμισμένων Ελάχιστων Τετραγώνων (Weighted Least Squares Error-WLSE):

$$\hat{\mathcal{A}} = \mathbf{D}^{-1} (\mathbf{B}^T \mathbf{G} \mathbf{I}), \quad \mathbf{D} = \mathbf{B}^T \mathbf{G} \mathbf{B} \quad (4.20)$$

ενώ εισάγοντας την προκύπτουσα εκτιμήτρια στην (4.19) παίρνουμε:

$$\mathcal{R} = \mathbf{I}^T \mathbf{G} \mathbf{I} - \hat{\mathcal{A}}^T \mathbf{D} \hat{\mathcal{A}}. \quad (4.21)$$

Παρατηρούμε εδώ ότι η μείωση στο σφάλμα ανακατασκευής συνοψίζεται στις τιμές των παραμέτρων $\hat{\mathcal{A}}$: ο πρώτος όρος της (4.21) είναι ανεξάρτητος από το συγκεκριμένο μοντέλο και ο δεύτερος λαμβάνει περιλαμβάνει μόνο τις παραμέτρους $\hat{\mathcal{A}}$ και έναν σταθερό πίνακα.

Αν η συνάρτηση εμπιστοσύνης $G(x)$ είναι Gaussian, ο πολλαπλασιασμός των τριών στοιχείων βάσης της (4.12) με την $G(x)$ μας δίνει τις κρουστικές αποκρίσεις ενός άρτιου και ενός περιττού φίλτρου Gabor και ενός Gaussian φίλτρου. Αυτό σημαίνει ότι η συνέλιξη με τα φίλτρα αυτά μας παρέχει σε κάθε pixel τις τιμές του 3×1 διανύσματος $\mathbf{B}^T \mathbf{G} \mathbf{I}$. Πέρα από αυτό, θεωρώντας αμελητέα την απόκριση του

άρτιου φίλτρου Gabor σε ένα σταθερό σήμα έχουμε

$$\sum_x G(x) \cos(x) \simeq 0 \rightarrow \mathbf{D} = \begin{bmatrix} D_1 & 0 & 0 \\ 0 & D_2 & 0 \\ 0 & 0 & D_3 \end{bmatrix} \quad (4.22)$$

$$\sum_x G(x) \cos(2x) \simeq 0 \rightarrow D_1 = D_2 = \frac{1}{2} \sum_x G(x), \quad (4.23)$$

που σημαίνει ότι η αντιστροφή του πίνακα (4.20) ανάγεται στην κανονικοποίηση των κρουστικών αποκρίσεων των Gabor/Gaussian φίλτρων.

Εισάγοντας την έκφραση

$$G'(x) = \frac{G(x)}{\sum_x G(x)} = \frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\frac{x^2}{2\sigma_{GB}^2}\right), \quad (4.24)$$

όπου σ_{GB} είναι η διασπορά του φίλτρου Gabor έχουμε τις εκτιμήσεις παραμέτρων:

$$\begin{aligned} \hat{E}_1 &= 2 \sum_x G'(x) \sin(x) I(x) \\ \hat{E}_2 &= 2 \sum_x G'(x) \cos(x) I(x) \\ \hat{E}_3 &= \sum_x G'(x) I(x). \end{aligned}$$

Μπορούμε λοιπόν να ερμηνεύσουμε την συνέλιξη με φίλτρα Gabor ως την εκτίμηση της σταθμισμένης προβολής ελάχιστων τετραγώνων του σήματος σε μία ημιτονοειδή βάση· η στάθμιση εγγυάται ότι τα σημεία που είναι πιο κοντά στο κέντρο της Gaussian του φίλτρου ανακατασκευάζονται ακριβέστερα από τα πιο μακρινά. Κατά τον ίδιο τρόπο, η συνέλιξη με μία Gaussian μπορεί να ερμηνευθεί ότι εκτελεί μία προβολή ελάχιστων τετραγώνων σε μία βάση που αποτελείται από το σταθερό στοιχείο. Σημειώνουμε εδώ ότι χωρίς το πλαίσιο που εισάγουμε παραπάνω, η ερμηνεία μίας συνέλιξης ως μία προβολή σε μία βάση είναι εφικτή όταν το σύνολο φίλτρων είναι ορθογώνιο, οπότε και η βάση αποτελείται από τους πυρήνες των φίλτρων. Αντίθετα, εδώ οι πυρήνες των φίλτρων είναι συναρτήσεις Gabor, ενώ οι βάσεις ημιτονοειδή.

Η ελάττωση του σφάλματος ανακατασκευής βάσει του μοντέλου (4.12) μπορεί να εκφραστεί ως

$$\hat{\mathbf{A}}^T \mathbf{D} \hat{\mathbf{A}} = \sum_x G(x) \left[\frac{\hat{E}_1^2 + \hat{E}_2^2}{2} + \hat{E}_3^2 \right] = \sum_x G(x) \left[\frac{\hat{A}^2}{2} + \hat{B}^2 \right]. \quad (4.25)$$

Καθώς η παράμετρος $\hat{E}_3 = \hat{B}$ καθορίζεται από το Gaussian φίλτρο, ο μόνος όρος που εξαρτάται από το συγκεκριμένο υπομοντέλο είναι ο $\hat{E}_1^2 + \hat{E}_2^2 = \hat{A}^2$, ο οποίος ισούται με το τετράγωνο του κριτήριου επιλογής καναλιού βάσει πλάτους. Το κάτω όριο (4.17) μπορεί λοιπόν να γραφεί ως:

$$LB_T(I) = c_1 \sum_x G(x) \hat{A}^2 + c_2 \quad (4.26)$$

με τα c_1, c_2 να είναι σταθερές ανεξάρτητες από το συγκεκριμένο μοντέλο. Μπο-

ρούμε λοιπόν να ερμηνεύσουμε την επιλογή καναλιού ως μία διαδικασία ελέγχου πολλαπλών υποθέσεων [119], καθώς επιλέγει το κανάλι που μεγιστοποιεί ένα κάτω φράγμα του λογάριθμου της πιθανοφάνειας των παρατηρήσεων υπό την υπόθεση του μοντέλου (4.12).

Πέρα από τη θεωρητική του σαφήνεια, αυτό το γενικότερο πλαίσιο αντιμετωπίζει εύκολα θέματα όπως προβλήματα στο όριο της εικόνας και ελλειπή δεδομένα, καθώς και την μη-μηδενική απόκριση του άρτιου φίλτρου σε ένα σταθερό σήμα. Για το πρώτο πρόβλημα μπορούμε να θέσουμε ίσο με μηδέν το βάρος του σφάλματος ανακατασκευής στα σημεία πέρα από το όριο της εικόνας· αγνοώντας το σφάλμα ανακατασκευής εκεί παίρνουμε αποκρίσεις που έχουν ελάχιστα προβλήματα στο σύνορο, με έναν τρόπο εφάμιλλο της κανονικοποιημένης διαφορικής συνέλιξης [122, 234]. Για το δεύτερο πρόβλημα η λύση πάλι είναι άμεση: ο πίνακας $D = [\mathbf{B}^T \mathbf{G} \mathbf{B}]$ δεν είναι πλέον διαγώνιος, και ο αντίστροφός του πρέπει να χρησιμοποιηθεί για να εξάγουμε τις βέλτιστες παραμέτρους. Παρατηρούμε ότι το φίλτρο που παίρνουμε έτσι είναι ισοδύναμο με αυτό που προτείνεται στο άρθρο [140] για να κατασκευαστεί ένα άρτιο φίλτρο Gabor με μηδενική απόκριση σε σταθερά σήματα. Όλες οι αντιστροφές πινάκων μπορούν να προϋπολογιστούν, οπότε έχουμε ελάχιστη υπολογιστική επιβάρυνση.

Θόρυβος Brown & Ενέργεια Teager

Μία υπόθεση που στηρίζει την προηγούμενη ανάλυση είναι ότι η διαφορά μεταξύ της πρόβλεψης του μοντέλου και του παρατηρούμενου σήματος μπορεί να μοντελοποιηθεί ως λευκός Gaussian θόρυβος. Για συσχετισμένο θόρυβο οι εκφράσεις εκτίμησης παραμέτρων περιλαμβάνουν τη διαγωνιοποίηση του πίνακα συνδιακύμανσης και γίνονται αρκετά περίπλοκες [119]. Ειδικά όμως για τον θόρυβο Brown ο οποίος μπορεί να θεωρηθεί ως το ολοκλήρωμα Gaussian θορύβου μπορούμε να δουλέψουμε πάνω στην παράγωγο των παρατηρήσεων οι οποίες πλέον μπορούν να μοντελοποιηθούν μέσω των παραγώγων των στοιχείων της βάσης στις οποίες θεωρούμε ότι προστίθεται λευκός Gaussian θόρυβος.

Συγκεκριμένα, για το μοντέλο υψής I_T το οποίο περιλαμβάνει ένα μεμονωμένο ημιτονοειδές συχνότητας ω , αυτό ισοδυναμεί με την προβολή της παραγώγου του σήματος O' στην παράγωγο των στοιχείων της βάσης, $B'_{T,1} = \omega B_{T,2}(x)$ και $B'_{T,2} = -\omega B_{T,1}$. Από τις σχέσεις (4.20,4.21), ο πολλαπλασιασμός των στοιχείων της βάσης με μία σταθερά ω δεν θα έχει επίδραση στο σφάλμα ανακατασκευής, οπότε μπορούμε να την αγνοήσουμε και να εκτιμήσουμε το πλάτος A' της παραγώγου του σήματος I' χρησιμοποιώντας τα ίδια στοιχεία βάσης:

$$\begin{aligned} A' &= \sqrt{(O' * GB_{T,1})^2 + (O' * GB_{T,2})^2} \\ &= \sqrt{(O * GB'_{T,1})^2 + (O * GB'_{T,2})^2} \\ &= \omega \sqrt{(O * GB_{T,2})^2 + (O * GB_{T,1})^2} = \omega A. \end{aligned}$$

Επαναλαμβάνοντας της προηγούμενη ανάλυση και θεωρώντας ως δεδομένα μας την παράγωγο του σήματος, η πιθανοφάνεια των δεδομένων υπό την κλάση των σημάτων υψής εκφράζεται μέσω της ποσότητας $(A')^2 = \omega^2 A^2$, όπου A είναι η εκτίμηση του πλάτους για το σήμα πριν την παραγωγήσή του. Η ποσότητα $\omega^2 A^2$ ισούται με την

απόκριση του Τελεστή Ενέργειας Teager που είχαμε παρατηρήσει εμπειρικά [128, 130] ότι παρέχει καλύτερα χαρακτηριστικά υφής, δικαιολογώντας πλέον την χρήση του σε ένα πιθανοτικό πλαίσιο.

4.4 Διάκριση Υφής και Ακμών

Ένα σημαντικό πρακτικά και ενδιαφέρον θεωρητικά πρόβλημα είναι η διάκριση μεταξύ περιοχών που περιέχουν ακμές και περιοχών υφής. Τα φίλτρα που χρησιμοποιούνται για ανίχνευση ακμών και ανάλυση υφής έχουν παρόμοιες αποκρίσεις συχνότητας, η είναι ακόμη και τα ίδια, οπότε εύκολα συγχέουν τους δύο αυτούς τύπους περιοχών. Αυτό οδηγεί αφενός σε εσφαλμένες ανιχνεύσεις ακμών σε περιοχές υφής, αφετέρου σε χαρακτηριστικά υφής κατά μήκος των ορίων των αντικειμένων που δεν έχουν νόημα. Προηγούμενες εργασίες που ασχολούνται με το πρόβλημα αυτό περιλαμβάνουν τις αναφορές [150, 88], όπου χρησιμοποιώντας περίπλοκες και εν μέρει εμπειρικές μεθόδους επιχειρείται ο αυτόματος καθορισμός του ποιες περιοχές της εικόνας περιέχουν υφή.

Διατυπώνοντας την ανίχνευση ακμών στο πλαίσιο των τοπικών αναγεννητικών μοντέλων συγκρίνουμε τις δύο υποθέσεις σε μία κοινή βάση, χρησιμοποιώντας την ελάττωση του σφάλματος ανακατασκευής που εισάγουν. Βάσει αυτού του κριτηρίου μπορούμε να αναθέσουμε ύστερες πιθανότητες στις εναλλακτικές υποθέσεις και να αποφασίσουμε ποια δομή κυριαρχεί τοπικά.

4.4.1 Σχέσεις Σύνθεσης Ακμών

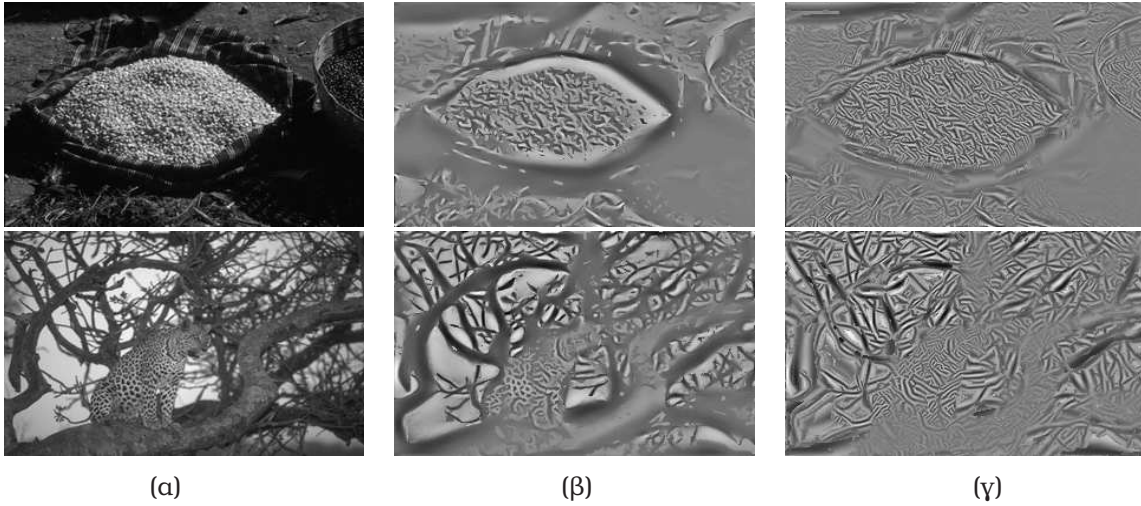
Στην εργασία [163] παρατηρήθηκε ότι αντιλαμβανόμαστε τις ακμές σε τοποθεσίες με σύμπτωση φάσης *-phase congruency*, δηλαδή περιοχές όπου το σήμα μπορεί να μοντελοποιηθεί τοπικά από ένα ανάπτυγμα σε σειρά Fourier της μορφής:

$$I(x) \simeq I_E(x) = A \sum_k a_k \cos(\omega_0 kx + \phi) + B. \quad (4.27)$$

Στη σχέση αυτή I_E είναι η προσέγγιση στο σήμα μέσω του μοντέλου της κατηγορίας των ακμών και ω_0 είναι η βασική συχνότητα της σειράς Fourier. Η διαφορά φάσης ϕ είναι κοινή για όλες τις αρμονικές συνιστώσες, οπότε στο σημείο $x_0 = 0$ έχουμε σύμπτωση φάσης, με διαφορετικές τιμές του ϕ να δίνουν διαφορετικούς τύπους ακμών: π.χ. για $\phi = 0$ και $\phi = \pi/2$ αντιλαμβανόμαστε ακμές τύπου ράβδου και τύπου βηματικής συνάρτησης αντίστοιχα.

Η χρήση ορθογωνικών ζευγών φίλτρων (Quadrature Filter Pairs-QFPs) σε ανίχνευση ακμών [163, 186, 68, 20] πήρε ώθηση από το μοντέλο αυτό. Είναι εύκολο να αποδειχθεί ότι συνελίσσοντας ένα τέτοιο ζεύγος φίλτρων με ένα σήμα της μορφής (4.27) και συνδυάζοντας τις εξόδους σε μία μέτρηση *τοπικής ενέργειας* [163] $LE(x) = (h_e * I)^2 + (h_o * I)^2$ αποκτούμε μία ποσότητα ανεξάρτητη της φάσης ϕ . Στην αναφορά [186] αποδεικνύεται ότι η προσέγγιση αυτή υπερέρχει της μεθόδου του Canny όταν εφαρμόζεται σε ακμές που εμφανίζονται σε συνδυασμό με σκίαση, λόγω των καλύτερων ιδιοτήτων εντοπισμού τους.

Μπορούμε άμεσα να διατυπώσουμε αυτή τη σχέση στο πλαίσιο των τοπικών μον-



Σχήμα 4.5: Ανακατασκευή της μη-ομαλής συνιστώσας της εικόνας (α) χρησιμοποιώντας (β) τις παραμέτρους του μοντέλου ακμών, (4.27) και (γ) τις παραμέτρους του μοντέλου υφής, (4.12). Η ανακατασκευή στην εικόνα (γ) αναδεικνύει τις περιοδικές πτυχές της τοπικής δομής, αντίθετα με την εικόνα (β) όπου απότομες, μη-περιοδικές μεταβάσεις είναι πιο έντονες.

τέλων μας, γράφοντας την (4.27) ως ένα ανάπτυγμα σε γραμμική βάση:

$$I_E(x; \mathcal{A}) = \sum_{i=1}^{D_E} E_i B_{E,i}(x), \quad \text{όπου } D_E = 3 \quad (4.28)$$

$$E_1 = A \cos(\phi), \quad E_2 = -A \sin(\phi), \quad E_3 = B \quad (4.29)$$

$$B_{E,1}(x) = \sum_k a_k \cos(\omega_0 k x), \quad B_{E,2}(x) = \sum_k a_k \sin(\omega_0 k x), \quad B_{E,3}(x) = 1$$

Προφανώς το σήμα που ορίζεται από την σχέση (4.27) είναι περιοδικό που δεν ανταποκρίνεται στην έννοια μίας ακμής. Όμως, όπως φαίνεται και στο Σχ. 4.4 η χωρική έκταση της σχετιζόμενης συνάρτησης εμπιστοσύνης είναι αρκετά μικρή ώστε να λαμβάνει υπόψη λιγότερο από το μισό της περιόδου του σήματος, όπου τα στοιχεία της βάσης συμπεριφέρονται σαν τυπικές ακμές τύπου βήματος και ράβδου. Συνεπώς τα φίλτρα $h_e = G(x)B_{E,1}(x)$, $h_o = G(x)B_{E,2}(x)$ που εμπλέκονται στη σχέση που αντιστοιχεί στην (4.20) αποκρίνονται σε σήματα ακμών, αγνοώντας την όποια περιοδικότητα μπορεί να υπάρχει σε μεγαλύτερη κλίμακα από αυτήν στην όποια μοντελοποιείται το σήμα.

Συγκεκριμένα, συνδέουμε με κάθε υπομοντέλο για υφή της μορφής (4.12) ένα αντίστοιχο μοντέλο ακμών, που εκφράζεται βάσει της (4.27): η βασική συχνότητα στην (4.27) τίθεται εμπειρικά ίση με $\omega_0 = 0.4\omega$, όπου ω είναι η συχνότητα στο μοντέλο υφής. Οι συνιστώσες a_k της σειράς Fourier στην (4.27) τίθενται ίσες με αυτές ενός περιοδικού τετραγωνικού συρμού.

Καθώς η προσέγγισή μας βασίζεται σε ένα μοντέλο του σήματος, αντί για την εξαγωγή χαρακτηριστικών από αυτό, η έκφραση για το σφάλμα ανακατασκευής (4.21) δεν περιορίζει κατά οποιονδήποτε τρόπο τα σήματα βάσης που χρησιμοποιούνται. Συνεπώς δεν χρειαζόμαστε να ασχοληθούμε με το λεπτό ζήτημα του να εγγυηθούμε ότι τα φίλτρα $h_e = G(x)B_{E,1}(x)$, $h_o = G(x)B_{E,2}(x)$ συνιστούν ένα QFP, όπως π.χ. στην αναφορά [163] όπου χρησιμοποιούνται τεχνικές στο πεδίο συχνότητας. Αν πάλι

θεωρήσουμε ότι το φάσμα του μιγαδικού φίλτρου $h_e + jh_o$ για αρνητικές συχνότητες είναι αρκετά μικρό ώστε να μπορεί να αγνοηθεί, μπορούμε να το δούμε ως ένα QFP. Η συνέλιξη με τα φίλτρα αυτά τότε αντιστοιχεί στην εκτίμηση της βέλτιστης προβολής στην βάση του μοντέλου ακμών $B_{E,1}, B_{E,2}$ και η τοπική του ενέργεια χρησιμεύει ως ένα μέτρο της εγγύτητας του μοντέλου ακμών στα παρατηρούμενα δεδομένα.

Συγκρίνοντας τα αναγεννητικά μοντέλα για ακμές και υφή, η σχέση (4.12) μοντελοποιεί σήματα τα οποία είναι ιδανικά εντοπισμένα στη συχνότητα, ενώ η (4.27) μπορεί να μοντελοποιήσει απότομες αλλαγές στην φωτεινότητα, χρησιμοποιώντας υψηλότερους όρους συχνότητας. Επίσης, στην περίπτωση του μοντέλου υφής, όπως φαίνεται στο (4.4) η συνάρτηση εμπιστοσύνης επιτρέπει για αρκετές ταλαντώσεις του σήματος, καθιστώντας το μοντέλο καλύτερα προσαρμοσμένο σε περιοδικά σήματα αντί για μεμονωμένες αλλαγές στην φωτεινότητα.

Μπορούμε να δούμε τι 'βλέπει' το κάθε μοντέλο στην εικόνα ανακατασκευάζοντας μία εικόνα μέσω των τοπικά εκτιμώμενων παραμέτρων των μοντέλων. Στο Σχ. 4.5 βλέπουμε ότι το μοντέλο της σχέσης (4.27) ερμηνεύει την μεταβολή της εικόνας μέσω απότομων μεταβάσεων στην φωτεινότητα, ενώ το μοντέλο της σχέσης (4.12) εστιάζει σε δομές τύπου κυματώσεων. Η αντίθεση (contrast) των συντιθέμενων εικόνων είναι ανάλογη του πλάτους του μοντέλου, που με τη σειρά του εκφράζει την δυνατότητα του μοντέλου να ανακατασκευάσει τις παρατηρήσεις. Παρατηρούμε ότι το μοντέλο ακμών έχει μεγαλύτερη αντίθεση κατά μήκος των ορίων των αντικειμένων ενώ το μοντέλο για την υφή στο εσωτερικό των περιοχών με υφή.

Μηδενική Υπόθεση: Ομαλά Σήματα

Συμπληρώνοντας το σύνολο μοντέλων, οι ομαλές περιοχές (smooth -S) μοντελοποιούνται ως σταθερά σήματα :

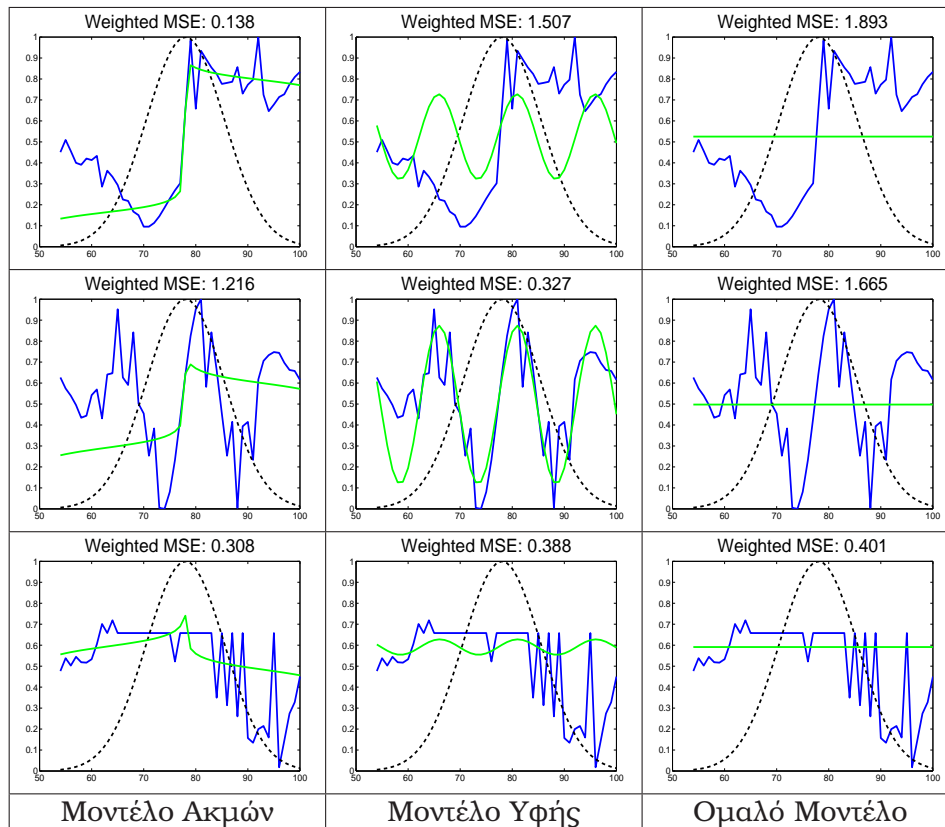
$$I(x) \simeq I_S(x) = B. \quad (4.30)$$

Αυτό το μοντέλο ελαττώνει την αξιοπιστία των χαρακτηριστικών ακμής/υφών σε ομαλές περιοχές της εικόνας. Χρησιμοποιώντας το ίδιο σκεπτικό με τα προηγούμενα δύο μοντέλα, μπορούμε να δούμε το αποτέλεσμα της συνέλιξης με μία Gaussian συνάρτηση ως την σταθμισμένη βέλτιστη προβολή των παρατηρήσεών μας στο σταθερό στοιχείο βάσης 1.

Αυτό το μοντέλο είναι εξ αρχής γνωστό ότι θα μας δώσει μεγαλύτερο σφάλμα ανακατασκευής από τα άλλα δύο, καθώς προκύπτει ως μια ειδική περίπτωση των σχέσεων (4.12,4.27) θέτοντας τον περιορισμό $A = 0$. Αυτό αντιμετωπίζεται μέσω ενός όρου ποινής τύπου Ελάχιστου Μήκους Περιγραφής - Minimum Description Length -MDL [119] στις εκφράσεις της πιθανοφάνειας για τα δύο προηγούμενα μοντέλα, ίσο με $MDL = \frac{2}{\sigma^2} \log(\sum_x G(x))$. Το 2 ισούται με τον επιπλέον αριθμό παραμέτρων, ενώ ο όρος $\log(\sum_x G(x))$ προκύπτει προσαρμόζοντας στην περίπτωσή μας την απόδειξη στο παράρτημα F του κεφάλαιου 6 από την αναφορά [119].

Διάκριση Ακμών/Υφής σε Πολλαπλές Κλίμακες

Έως τώρα εμμέσως θεωρούσαμε ότι και τα τρία μοντέλα επιχειρούν να ερμηνεύσουν μία περιοχή της εικόνας σε μία σταθερή κλίμακα. Καθώς επιλέγουμε ανάμεσα στις θεωρούμενες κλάσεις συνδυάζοντας πληροφορία από όλες τις κλίμακες, χρειαζόμαστε μία μέθοδο που να καθιστά τους όρους πιθανοφάνειας που προκύπτουν από



Πίνακας 4.1: Ανακατασκευή σήματος μέσω διαφορετικών αναγεννητικών μοντέλων: Για κάθε γραμμή του πίνακα ένα σήμα από μία διαφορετική κλάση έχει συντεθεί (έντονη γραμμή), και επιδεικνύεται η ανακατασκευή του μέσω των διαθέσιμων αναγεννητικών μοντέλων (αχνή γραμμή), χρησιμοποιώντας τη συνάρτηση εμπιστοσύνης που φαίνεται με διακεκομμένη γραμμή. Από πάνω προς τα κάτω, το σήμα έχει προκύψει από το μοντέλο ακμής, υφής και ομαλών σημάτων. Το σφάλμα ανακατασκευής είναι μικρότερο όταν χρησιμοποιείται το αντίστοιχο του σήματος αναγεννητικό μοντέλο· εξαίρεση αποτελεί η περίπτωση του ομαλού μοντέλου, όπου το μικρότερο πλήθος παραμέτρων οδηγεί πάντα σε μεγαλύτερο σφάλμα ανακατασκευής.

έναν διαφορετικό αριθμό παρατηρήσεων συγκρίσιμους. Για να εισάγουμε έναν όρο που είναι ανεπηρέαστος από την πρόσθεση σταθερών τιμών στην φωτεινότητα, θεωρούμε την ποσότητα $\mathcal{G}_{\mathcal{H}} = \log \frac{P(I|\mathcal{H})}{P(I|S)}$, όπου \mathcal{H} είναι ένα από τα E, T και S η υπόθεση ομαλού σήματος και αντικαθιστώντας τους λογαρίθμους με τα κάτω φράγματα τους που μας δίνει:

$$\mathcal{G}_{\mathcal{H}} = \frac{1}{2\sigma^2} (\mathcal{A}_{\mathcal{H}} D_{\mathcal{H}} \mathcal{A}_{\mathcal{H}}^T - \mathcal{A}_S D_S \mathcal{A}_S^T) + \text{MDL}(\mathcal{H}). \quad (4.31)$$

Ο πρώτος όρος ποσοτικοποιεί την αύξηση στον λογάριθμο της πιθανοφάνειας που οφείλεται στην εισαγωγή του πολυπλοκότερου μοντέλου \mathcal{H} , ενώ ο όρος $\text{MDL}(\mathcal{H})$ στηρίζει την ομαλή υπόθεση σε περιοχές όπου αυτή η αύξηση δεν είναι σημαντική. Η ποσότητα $\mathcal{G}_{\mathcal{H}}$ είναι ανάλογη του ολοκληρώματος της συνάρτησης εμπιστοσύνης $G_{\mathcal{H}}$. συνεπώς, διαιρώντας την με τον όρο $c \sum_x G_{\mathcal{H}}(x)$ αποκτούμε μία προσεγγιστικά ανεξάρτητη κλίμακας ποσότητα, το κέρδος σε λογάριθμο πιθανοφάνειας ανά pixel :

$$\mathcal{E}_{\mathcal{H}} = \frac{\mathcal{G}_{\mathcal{H}}}{c \sum_x G_{\mathcal{H}}(x)} \quad (4.32)$$

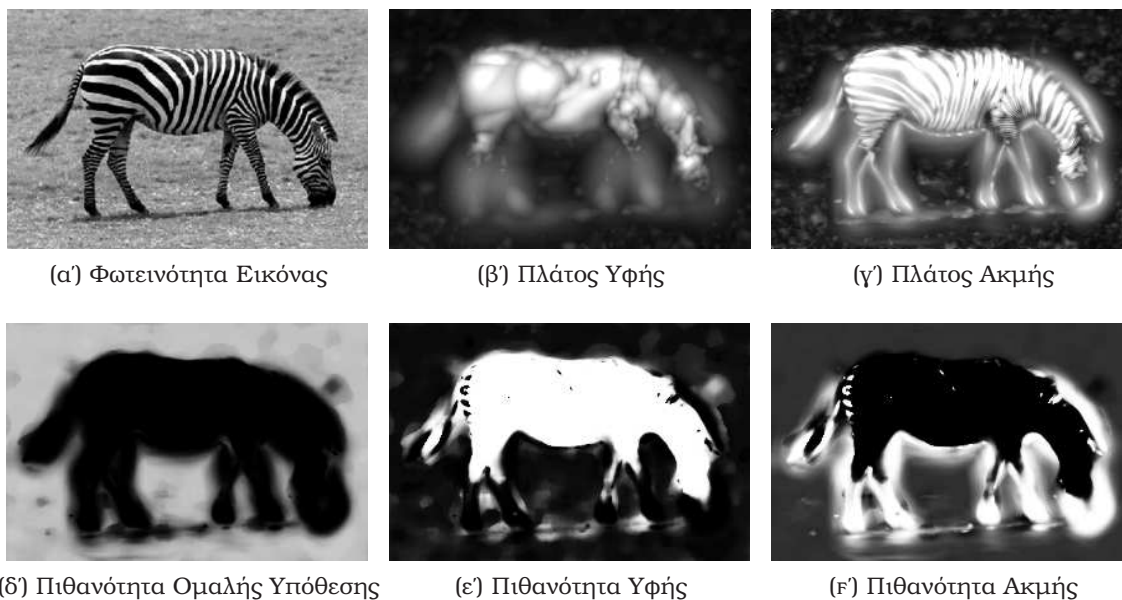
Η απόκλιση από την ανεξαρτησία κλίμακας οφείλεται στον όρο $-\frac{\log(\sum_x G_{\mathcal{H}}(x))}{\sum_x G_{\mathcal{H}}(x)}$ που προκύπτει από το MDL, και δείχνει μία προτίμηση σε υποθέσεις που μοντελοποιούν μεγαλύτερες περιοχές της εικόνας, που είναι διαισθητικά ικανοποιητικό. Ο όρος c καθορίζει την αυστηρότητα των ακόλουθων αποφάσεων ενώ σ^2 είναι η θεωρούμενη διασπορά της διαδικασίας θορύβου στο αναγεννητικό μοντέλο. Αν και ιδανικά αυτοί οι όροι θα έπρεπε να εκτιμηθούν από δεδομένα εκπαίδευσης, ικανοποιητικά αποτελέσματα σε μία μεγάλη ποικιλία εικόνων έχουν παρατηρηθεί χρησιμοποιώντας τις σταθερές τιμές $\sigma = 10^{-1}$ και $c = 10$.

Έχοντας καταλήξει σε ένα ανεξάρτητο κλίμακας κριτήριο μπορούμε να αποσυνδέσουμε την απόφαση ανάμεσα στο υπομοντέλο για την κάθε τάξη και την ταξινόμηση μιας γειτονιάς σε μία από τις τρεις τάξεις. Συγκεκριμένα, επιλέγουμε χωριστά τον συνδυασμό κλίμακας και κατεύθυνσης του υπομοντέλου που μεγιστοποιεί το κριτήριο για κάθε μία από τις δύο τάξεις και στη συνέχεια τις συγκρίνουμε με βάση εν γένει διαφορετικές επιλογές κλίμακας και κατεύθυνσης. Χρησιμοποιώντας τις μέγιστες τιμές των κριτηρίων $\mathcal{E}_{\mathcal{H}}$ μπορούμε να γράψουμε για τις ύστερες κατανομές των τριών κλάσεων:

$$\begin{aligned} P(T|O) &= \frac{P(O|T)}{P(O|T) + P(O|E) + P(O|S)} = \frac{R_T}{R_T + R_E + 1}, \\ P(E|O) &= \frac{R_E}{R_T + R_E + 1}, \\ P(S|O) &= \frac{1}{R_T + R_E + 1}, \\ \text{όπου } R_T &= \frac{P(I|T)}{P(I|S)} = \frac{1}{1 + \exp(-\mathcal{E}_T)} \\ R_E &= \frac{P(I|E)}{P(I|S)} = \frac{1}{1 + \exp(-\mathcal{E}_E)} \end{aligned} \quad (4.33)$$

Αυτή η προσέγγιση έχει εφαρμοστεί σε ποικιλία φυσικών εικόνων, όπως αυτές στα Σχ. 4.6,4.7 δίνοντας ικανοποιητικά αποτελέσματα. Σε αυτές τις εικόνες η πιθανό-

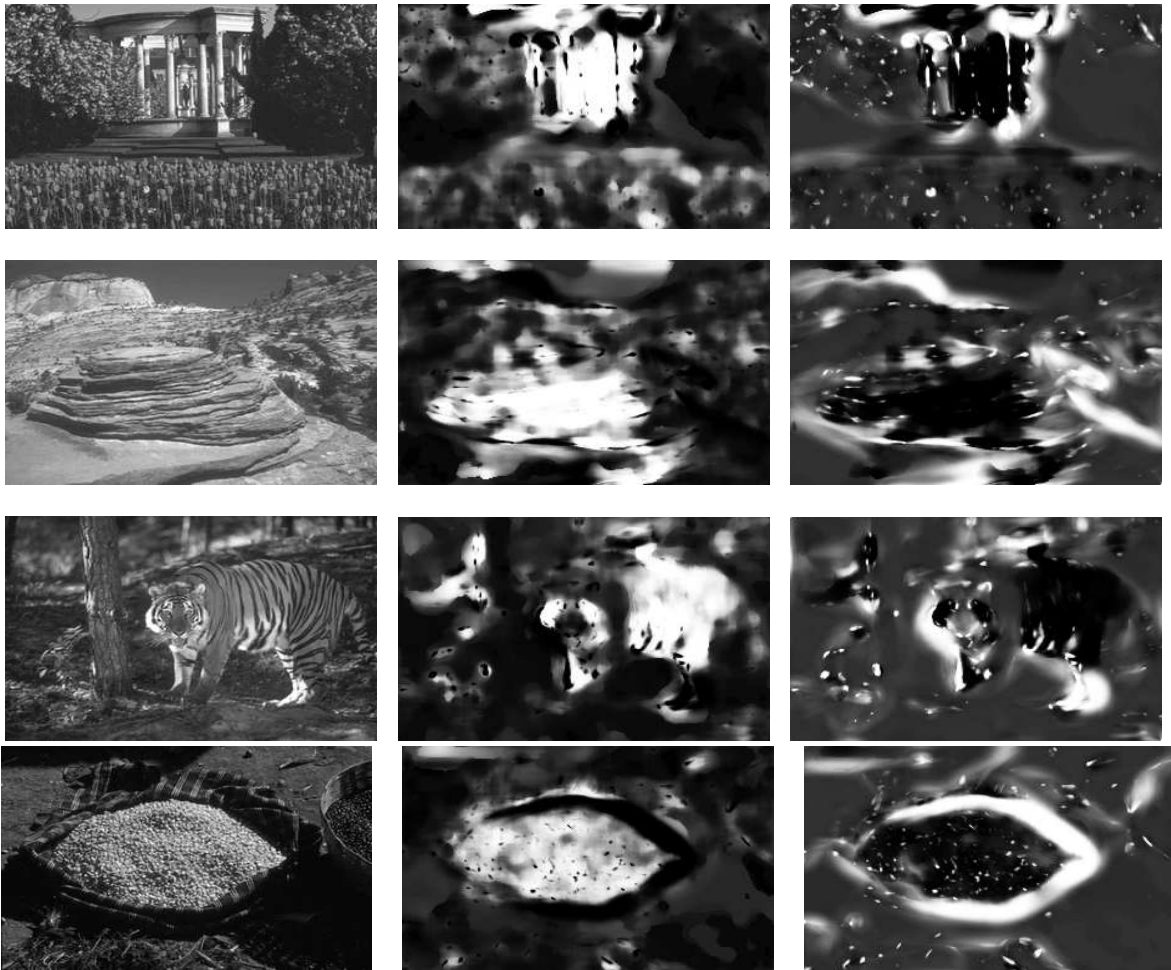
τητα μία ακμής είναι τυπικά υψηλότερη πάνω στα σύνορα των αντικειμένων, ενώ σε περιοχές με υφή η πιθανότητά της είναι χαμηλότερη από αυτήν που θα υποδείκνυε η εφαρμογή ενός φίλτρου ανίχνευσης ακμών. Παρατηρούμε επίσης ότι κατά μήκος των ορίων των αντικειμένων υπάρχει μία ελάττωση στην πιθανότητα της υφής, καθώς το μοντέλο για ακμές αντιπροτείνει μία καλύτερη ερμηνεία για την μεταβολή στην φωτεινότητα.



Σχήμα 4.6: Διάκριση περιοχών υφής από ακμές, χρησιμοποιώντας τοπικά αναγεννητικά μοντέλα: Εικόνα (α) και πλάτη των μοντέλων υφής (β) και ακμής (γ) $\sqrt{A_1^2 + A_2^2}$, με τα A_1, A_2 εκτιμώμενα βάσει του τύπου (4.20). Τα πλάτη στις εικόνες (β) και (γ) είναι ανάλογα με την πιθανοφάνεια των δεδομένων, δεδομένης της κλάσης του σήματος. Οι ύστερες πιθανότητες των κλάσεων ομαλού σήματος (δ), σήματος υφής (ε) και ακμής (στ) προκύπτουν από τον τύπο του Bayes . Η φωτεινότητα της εικόνας κωδικοποιεί την πιθανότητα της αντίστοιχης υπόθεσης.

4.5 Κατάτμηση Εικόνων Υφής με Χαρακτηριστικά Διαμόρφωσης και Σταθμισμένη Εξέλιξη Καμπυλών

Στην ενότητα αυτή αντιμετωπίζουμε το πρόβλημα της κατάτμησης εικόνων που περιέχουν υφή συνδυάζοντας τρεις συνιστώσες: τα χαρακτηριστικά διαμόρφωσης για την εξαγωγή μίας συμπαγούς περιγραφής της υφής, τα αναγεννητικά μοντέλα για την εκτίμηση των ύστερων πιθανοτήτων των υποθέσεων των τριών θεωρούμενων κατηγοριών και της εξέλιξης καμπυλών για την εύρεση της κατάτμησης. Ο συνδυασμός αυτών των τριών τεχνικών προκύπτει αβίαστα, δίνοντας ένα κομψό και αποτελεσματικό πλαίσιο για την αντιμετώπιση του προβλήματος.



(α) Εικόνα

(β) Πιθανότητα Υφής

(γ) Πιθανότητα Ακμής

Σχήμα 4.7: Φυσικές εικόνες (α) και ύστερες πιθανότητες των μοντέλων υφής (β) και ακμής (γ) αντίστοιχα. Η φωτεινότητα της εικόνας κωδικοποιεί την πιθανότητα της αντίστοιχης υπόθεσης.

4.5.1 Χρήση Χαρακτηριστικών Διαμόρφωσης για Κατάτμηση Εικόνων

Οι περισσότεροι αλγόριθμοι κατάτμησης εικόνων υψής έχουν να αντιμετωπίσουν το πρόβλημα του μεγάλου αριθμού χαρακτηριστικών που προκύπτουν από την χρήση μίας συστοιχίας φίλτρων. Όταν χρησιμοποιείται ένα στάδιο εκμάθησης για τις διαφορετικές τάξεις υψής [178] μπορούμε να πάρουμε ικανοποιητικά αποτελέσματα επιλέγοντας κάποια από τα χαρακτηριστικά για τον διαχωρισμό των κλάσεων υψής. Στην γενικότερη περίπτωση της κατάτμησης υψής χωρίς επίβλεψη ευρετικές που έχουν προταθεί, π.χ. [197] μπορούν να θεωρηθούν ως τεχνικές αναζήτησης προβολών - Projection Pursuit [102], που είναι από μόνο του ένα δύσκολο πρόβλημα.

Σε μία πρόσφατα απόπειρα να αντιμετωπιστεί αυτό το πρόβλημα οι Brox, Rousson, & Deriche [193] χρησιμοποίησαν μία μη-γραμμική διαδικασία διάχυσης διανυσματικών χαρακτηριστικών, τα οποία εξάγονται αρχικά από τις παραγωγούς τις εικόνας:

$$\begin{aligned} \frac{\partial u_i}{\partial t} &= \operatorname{div} \left(g \left(\sum_{k=1}^3 |\nabla u_k| \right) \nabla u_i \right), \quad g(x) = \frac{1}{|\nabla u| + 10^{-3}}, \quad (4.34) \\ u_1(x, y, 0) &= I_x^2(x, y), \\ u_2(x, y, 0) &= I_x(x, y) I_y(x, y), \\ u_3(x, y, 0) &= I_y^2. \quad (4.35) \end{aligned}$$

Χρησιμοποιώντας τα χαρακτηριστικά που προκύπτουν από την επίλυση αυτού του συστήματος ΜΔΕ σε μία χρονική στιγμή t σε συνδυασμό με την φωτεινότητα της εικόνας επιτεύχθηκαν ικανοποιητικά αποτελέσματα κατάτμησης εικόνων υψής χωρίς επίβλεψη. Ένα πρόβλημα είναι ότι τα χαρακτηριστικά αυτά δεν φέρουν πληροφορία για την κλίμακα της υψής, καθώς η χρονική στιγμή t , που ελέγχει την κλίμακα στην οποία εξάγονται τα χαρακτηριστικά είναι καθορισμένη εκ των προτέρων· σχετικά με το πρόβλημα αυτό, σε μία πρόσφατη εργασία τους οι Brox & Weickert [25] έχουν εισάγει μία μέθοδο εκτίμησης κλίμακας βάσει διάχυσης για να επεκτείνουν το αρχικό τους διάνυσμα χαρακτηριστικών. Παρά τα ικανοποιητικά αποτελέσματα που επιτυγχάνονται, δεν είναι ξεκάθαρο τι ακριβώς εκπροσωπούν τα χαρακτηριστικά αυτά, καθώς προκύπτουν ως ένα παραπροϊόν μίας διαδικασίας διάχυσης.

Η χρήση χαρακτηριστικών DCA για κατάτμηση εικόνων [191, 129, 56, 130] ενδείκνυται για τρεις κυρίως λόγους:

- Πληροφορία σχετικά με την κλίμακα, κατεύθυνση και αντίθεση της υψής αναπαρίσταται μέσω ενός χαμηλοδιάστατου διανύσματος χαρακτηριστικών.
- Τα χαρακτηριστικά DCA εξάγονται βάσει ενός μαθηματικά σαφούς μοντέλου για την αναπαράσταση εικόνας και παρέχουν ένα εύκολα ερμηνεύσιμο διάνυσμα χαρακτηριστικών.
- Η ερμηνεία μέσω αναγεννητικών μοντέλων επιτρέπει την εκτίμηση των ύστερων πιθανοτήτων των τάξεων υψής, ακμών και ομαλών σημάτων, που εκμεταλλευόμαστε αργότερα για να καθορίσουμε αυτόματα την επιρροή των χαρακτηριστικών υψής στην διαδικασία εξέλιξης καμπυλών.

Το διάνυσμα χαρακτηριστικών που χρησιμοποιούμε αποτελείται από την φωτεινότη-

τα της εικόνας και τα χαρακτηριστικά DCA (4.10), συγκεκριμένα το πλάτος διαμόρφωσης και το μέτρο και κατεύθυνση της στιγμιαίας συχνότητας.

Η κατανομή $P(\cdot; a_i)$ μέσα στην περιοχή i μοντελοποιείται ως το γινόμενο μίας Gaussian για τα τρία πρώτα χαρακτηριστικά και μίας κατανομής von-Mises για το χαρακτηριστικό της κατεύθυνσης θ . Η τελευταία είναι ανάλογη μίας Gaussian κατανομής για δεδομένα κατεύθυνσης:

$$P_{VM}(\theta; \theta_0, \kappa) = \frac{\exp(\kappa \cos(\theta - \theta_0))}{2\pi I_0(\kappa)} \quad (4.36)$$

όπου I_0 είναι μία συνάρτηση Bessel πρώτου τύπου με φανταστικό όρισμα. Η εκτιμήτρια της παραμέτρου θ_0 δίνεται από την έκφραση [87]:

$$\theta_0 = \tan^{-1} \left(\frac{\sum_k \sin(\theta_k)}{\sum_k \cos(\theta_k)} \right), \quad (4.37)$$

ενώ η κ μπορεί να εκτιμηθεί αριθμητικά ως η λύση της

$$I'_0(\kappa) = \hat{\alpha} I_0(\kappa), \quad \hat{\alpha} = \frac{1}{K} \sqrt{\left(\sum_{k=1}^K \sin(\theta_k) \right)^2 + \left(\sum_{k=1}^K \cos(\theta_k) \right)^2},$$

όπου K είναι το πλήθος των παρατηρήσεων μέσα στην περιοχή. Καθώς οι εκτιμήσεις κατεύθυνσης βρίσκονται μέσα στο διάστημα $[0, \pi)$, και όχι στο $[0, 2\pi)$ διπλασιάζουμε την τιμή τους πριν τις εισάγουμε στις εκφράσεις εκτίμησης παραμέτρων και πιθανοφάνειας.

Στην υλοποίησή μας που ακολουθεί την αναφορά [179] χρησιμοποιούμε μία ξεχωριστή συνάρτηση εμπύθισης για κάθε ελίσσόμενη καμπύλη, και επιλύουμε την εξίσωση εξέλιξης χρησιμοποιώντας ένα ευθύ σχήμα ολοκλήρωσης, όπως περιγράφεται στην αντίστοιχη εισαγωγική ενότητα.

Η κατανομή των δεδομένων μέσα σε κάθε περιοχή εκτιμάται παράλληλα με την διαδικασία εξέλιξης, επιτρέποντας την ανεπίβλεπτη κατάτμηση εικόνων υφής, ξεκινώντας από τυχαίες αρχικές συνθήκες για τα σύνορα των περιοχών. Αντιθέτως, στην αρχική εργασία [191] όπου συνδυαζόταν εξέλιξη καμπυλών με χαρακτηριστικά DCA η εξέλιξη καμπυλών περιοριζόταν σε ένα στάδιο βελτίωσης των αποτελεσμάτων, όπου εισήγαγε πληροφορία για την καμπυλότητα, ομαλοποιώντας τα περιγράμματα των περιοχών.

4.5.2 Σταθμισμένη Εξέλιξη Καμπυλών

Το κύριο πρόβλημα που αντιμετωπίσαμε στην εφαρμογή της μεθόδου DCA σε προβλήματα κατάτμησης υφής ήταν η ποιότητα των χαρακτηριστικών σε περιοχές που δεν χαρακτηρίζονται από υφή. Συγκεκριμένα, στα σύνορα των αντικειμένων το πλάτος που εξάγεται με τη μέθοδο DCA είναι υψηλό, περιγράφοντας ταλαντώσεις, ενώ θα έπρεπε να έχει μία μικρή, αδιάφορη τιμή. Επίσης, σε ομαλές περιοχές η κατεύθυνση του διανύσματος συχνότητας μεταβάλλεται κατά έναν απρόβλεπτο τρόπο, καθώς δεν υπάρχει κάποια δομή να την περιορίσει. Πέρα από αυτό, η χρήση όρων που προκύπτουν από την ανίχνευση ακμών μπορεί να εμποδίσει την εξέλιξη καμπυλών και να την περιορίσει στο εσωτερικό περιοχών με έντονη υφή.

Σε αυτό το τμήμα εισάγουμε μία μέθοδο που απαλλάσσει την διαδικασία κα-

τάτμησης από τέτοια προβλήματα, επιλέγοντας αυτόματα σε ποια χαρακτηριστικά της εικόνας πρέπει να βασιστούν οι εξισώσεις εξέλιξης. Βασιζόμενοι σε μία προσέγγιση συνδυασμού ταξινομητών, παρουσιάζουμε μία τροποποίηση του αλγορίθμου Region Competition που εκμεταλλεύεται της ύστερες πιθανότητες που ανατίθενται στις κλάσεις βάσει των οποίων έγινε η εξαγωγή χαρακτηριστικών, επιτρέποντας την αυτοματοποιημένη, χωρικά μεταβαλλόμενη επιλογή των πιο αξιόπιστων χαρακτηριστικών.

Οι supra - Bayesian τεχνικές σύντηξης [108] συνδυάζουν τις εξόδους πολλαπλών ταξινομητών (ειδικών) χειριζόμενες τις ως τυχαίες μεταβλητές και θεωρώντας την από κοινού κατανομή τους. Συγκεκριμένα, για δύο κατηγορίες i, j και ένα σύνολο χαρακτηριστικών F , κάθε πιθανοτικός ταξινομητής μεταξύ των δύο κλάσεων αυτών παρέχει την ποσότητα

$$\mathcal{L}_F = \log \frac{P_i(F; a_i)}{P_j(F; a_j)}.$$

Αυτή η ποσότητα που θα ονομάζουμε *log-odd* (λογάριθμος απόδοσης) ποσοτικοποιεί τον βαθμό στον οποίο η παρατήρηση F είναι πιο πιθανή υπό την υπόθεση i από ότι υπό την j . Η προσέγγισή που ακολουθούμε θεωρεί την ποσότητα \mathcal{L}_F ως μία τυχαία μεταβλητή που ακολουθεί μία Gaussian κατανομή δεδομένης της κατηγορίας i, j των δεδομένων F :

$$P(\mathcal{L}_F|i) \propto N(\mu_i, \sigma^2), \quad P(\mathcal{L}_F|j) \propto N(\mu_j, \sigma^2)$$

Έτσι μπορούμε να ποσοτικοποιήσουμε την βεβαιότητα την οποία αντιστοιχούμε στην απόφαση που κάνει ένας συγκεκριμένος ταξινομητής: μία μεγάλη διασπορά σ , δηλαδή μία χαμηλή εμπιστοσύνη στην ακρίβεια της ποσότητας *log-odd* μπορεί να ελαττώσει το αποτέλεσμα μίας μεγάλης τιμής των \mathcal{L}_F . Από έναν καλό ταξινομητή γενικότερα περιμένουμε ότι αν τα δεδομένα F είναι από την κλάση i είναι μεγάλη η πιθανότητα (χαμηλό σ) το \mathcal{L}_F να πάρει μεγάλη τιμή ($\mu_i \gg \mu_j$) και αντίστροφα.

Με βάση την παραπάνω προσέγγιση, τα αποτελέσματα N ταξινομητών που χρησιμοποιούν διαφορετικές τεχνικές ταξινόμησης ή διαφορετικά χαρακτηριστικά μπορούν να συνδυαστούν [108]: οι *log-odds* τους $\mathcal{L} = [\mathcal{L}_1, \dots, \mathcal{L}_K]$ θεωρούνται ως μία πολυδιάστατη τυχαία μεταβλητή που ακολουθεί μία Gaussian κατανομή δεδομένης της τάξης των δεδομένων, δηλαδή

$$P(\mathcal{L}|i) \propto N(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}), \quad P(\mathcal{L}|j) \propto N(\boldsymbol{\mu}_j, \boldsymbol{\Sigma})$$

Ο ύστερος λογάριθμος του λόγου πιθανοτήτων δεδομένων των *log-odds* όλων των ταξινομητών είναι τότε:

$$\log \frac{P(i|\mathcal{L})}{P(j|\mathcal{L})} = (\mathcal{L} - \frac{\boldsymbol{\mu}_i + \boldsymbol{\mu}_j}{2})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)$$

Παίρνουμε λοιπόν μία απόφαση χρησιμοποιώντας ως 'χαρακτηριστικά' τις εξόδους των ταξινομητών. Για την ειδική περίπτωση που οι έξοδοι των ταξινομητών είναι ασυσχέτιστες, έχουμε έναν διαγώνιο πίνακα συνδιακύμανσης· επιπλέον, προσθέτοντας και πολλαπλασιάζοντας κατάλληλες σταθερές, μπορούμε να εγγυηθούμε ότι $\boldsymbol{\mu}_i = -\boldsymbol{\mu}_j$ και όλα τα στοιχεία του διανύσματος $\boldsymbol{\mu}_i$ ισούνται με μονάδα, οπότε θα

έχουμε :

$$\log \frac{P(i|\mathcal{L})}{P(j|\mathcal{L})} = \sum_{k=1}^K \frac{\mathcal{L}_k}{\sigma_k^2}$$

Αυτή η σχέση εκφράζει μία διαισθητικά προφανή ιδέα : όταν ένας ταξινομητής k δίνει θορυβώδη αποτελέσματα, δηλαδή έχει ένα μεγάλο σ_k , ένα χαμηλότερο βάρος θα έπρεπε να συνδέεται με την απόφασή του και ανάποδα. Η χρήση των log-odds είναι ιδιαίτερα πρακτική, καθώς μπορούμε να εκφράσουμε αυτήν την πράξη συνδυασμού ως μία πρόσθεση.

Σε αυτό το πλαίσιο, ο πιθανοτικός όρος $\log \frac{P_i(F;a_i)}{P_j(F;a_j)}$ μπορεί να θεωρηθεί ότι αναθέτει τα χαρακτηριστικά F στις περιοχές i και j και εξελίσσει τα όρια των περιοχών ώστε να ταξινομήσει σωστά τις παρατηρήσεις. Αυτό που προτείνουμε για το πρόβλημα που αντιμετωπίζουμε είναι να χειριστούμε τις ύστερες πιθανότητες των κλάσεων ως ενδεικτικές της ακρίβειας ενός ταξινομητή που χρησιμοποιεί το αντίστοιχο υποσύνολο παρατηρήσεων. Θεωρούμε ότι τρεις ταξινομητές χρησιμοποιούνται για να καθορίσουν την εξέλιξη μίας καμπύλης, ο πρώτος βάσει της φωτεινότητας, ο δεύτερος βάσει των χαρακτηριστικών υφής και ο τρίτος ο οποίος ομαδοποιεί τις παρατηρήσεις εκατέρωθεν των δεδομένων.

Χρησιμοποιώντας τον προηγούμενο συμβολισμό έχουμε για τους πρώτους δύο :

$$\mathcal{L}_T = \log \frac{P(F_T; \theta_{T,i})}{P(F_T; \theta_{T,j})}, \quad \mathcal{L}_S = \log \frac{P(F_S; \theta_{S,i})}{P(F_S; \theta_{S,j})}, \quad (4.38)$$

όπου με $P(F_M; a_{M,i})$ συμβολίζουμε την πιθανοφάνεια του συνόλου χαρακτηριστικών F_M που εξάγεται από την κλάση M (υφή, $T - texture$ ή ομαλή, $S - smooth$) υπό την κατανομή P της συγκεκριμένης κλάσης για την περιοχή i , της οποίας η παράμετροι είναι $\theta_{M,i}$. Για την τελική απόφαση, η απόφαση του κάθε ταξινομητή σταθμίζεται με την πιθανότητα της αντίστοιχης υπόθεσης, που εκτιμάται σύμφωνα με τη σχέση (4.33). Έτσι για περιοχές υφής τα χαρακτηριστικά υφής έχουν μεγαλύτερη επίδραση στην εξέλιξη της καμπύλης από ότι τα χαρακτηριστικά φωτεινότητας, και αντίστροφα για ομαλές περιοχές. Αντί της σχέσης (2.21) μπορούμε τώρα να γράψουμε τη σχέση *σταθμισμένης εξέλιξης καμπυλών*:

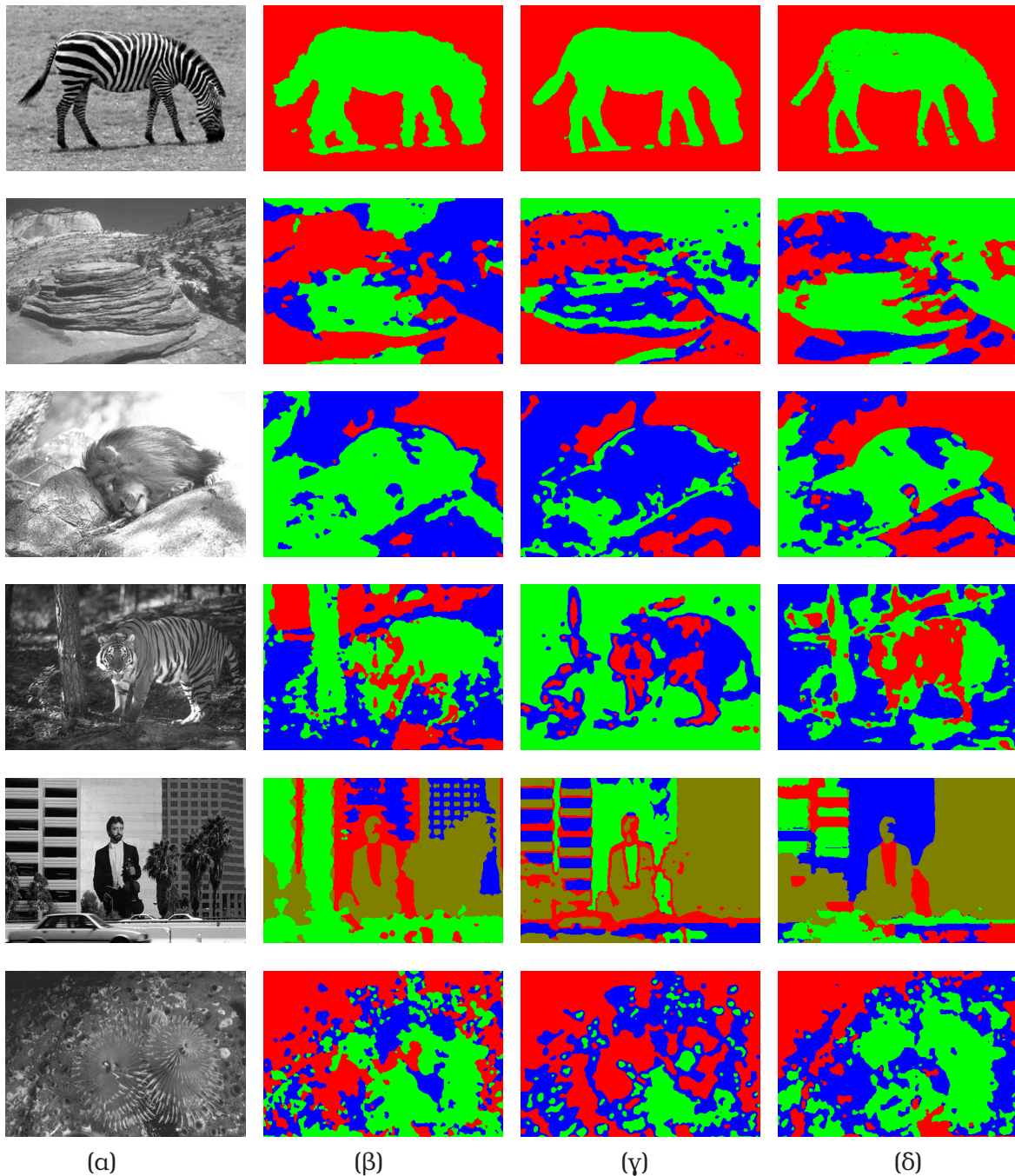
$$\frac{\partial \Gamma_i}{\partial t} = \left[\sum_{c \in T,S} w_c \log \frac{P(F_c; \alpha_{c,i})}{P(F_c; \alpha_{c,j})} - w_E [g \kappa \mathcal{N} + (\nabla g \cdot \mathcal{N}) \mathcal{N}] \right] \mathcal{N}, \quad (4.39)$$

όπου ως βάρη w_E, w_T, w_S χρησιμοποιούνται οι ύστερες πιθανότητες των μοντέλων ακμής, υφής και ομαλής φωτεινότητας. Επίτηδες αποφεύχθηκε εδώ η χρήση του συμβολισμού των πιθανοτήτων, ώστε να φανεί ότι πέρα από την διαισθητική προσέγγιση που παρουσιάστηκε νωρίτερα, δεν υπονοήθηκε κάποια άλλη αμεσότερη σχέση μεταξύ των ύστερων πιθανοτήτων και κάποιου βέλτιστου τρόπου συνδυασμού των ταξινομητών. Κάθε άλλο μέτρο έντασης υφής η ακμών, όπως π.χ. στις αναφορές [154, 150, 225] θα μπορούσε να χρησιμοποιηθεί στις παραπάνω σχέσεις, όπως και τα δικά μέτρα στο πλαίσιο των αναφορών αυτών.

Τέλος, σημειώνουμε ότι το w_E δεν σχετίζεται με την διαδικασία ανίχνευσης ακμών, που χρησιμοποιείται για τον υπολογισμό των g και ∇g , καθώς παρέχει μόνο μία ένδειξη του πόσο σημαντική πρέπει να θεωρηθεί η πληροφορία που σχετίζεται με

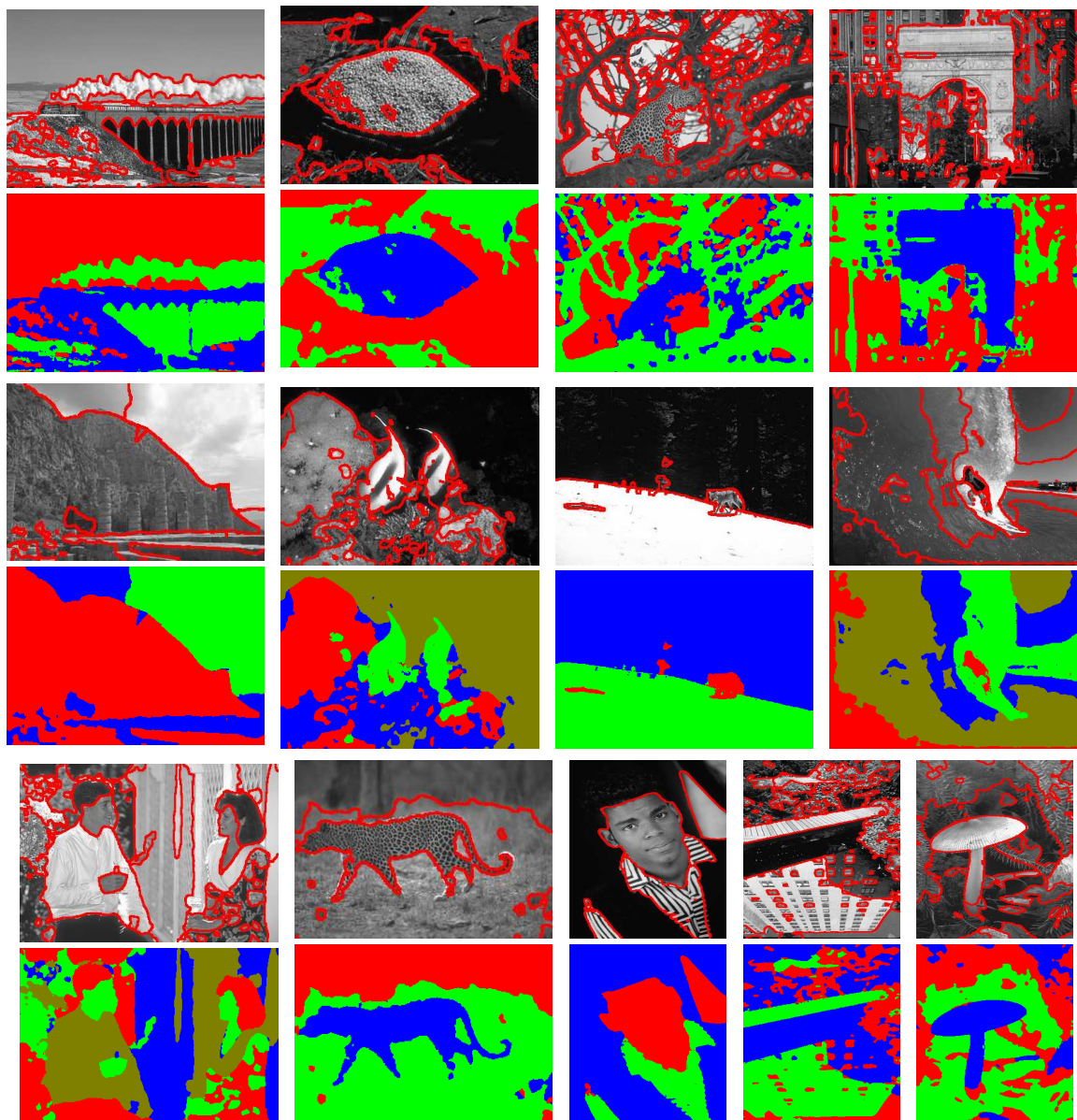
ακμές.

4.5.3 Πειραματικά Αποτελέσματα



Σχήμα 4.8: Αποτελέσματα κατάτμησης χρησιμοποιώντας διαφορετικά χαρακτηριστικά και αλγόριθμους εξέλιξης καμπυλών: (α)Είσοδος (β) Κατάτμηση χρησιμοποιώντας χαρακτηριστικά DCA και την φωτεινότητα [129]. (γ) κατάτμηση με τα χαρακτηριστικά διάχυσης των [193]. (δ) Κατάτμηση χρησιμοποιώντας χαρακτηριστικά διάχυσης και σταθμισμένη εξέλιξη καμπυλών (4.39).

Στο Σχ. 4.8 παρουσιάζουμε αποτελέσματα κατάτμησης χρησιμοποιώντας τρεις διαφορετικές μεθόδους: Στην δεύτερη στήλη παρουσιάζουμε τα αποτελέσματα της



Σχήμα 4.9: Αποτελέσματα κατάτμησης χρησιμοποιώντας χαρακτηριστικά διαμόρφωσης και σταθμισμένη εξέλιξη καμπυλών.

εξελιξης καμπυλών όπως στο άρθρο [129] χρησιμοποιώντας το τριδιάστατο διάνυσμα χαρακτηριστικών της μεθόδου Dominant Components Analysis για την περιγραφή της υφής. Στην επόμενη στήλη χρησιμοποιούμε τα χαρακτηριστικά των [193], ενώ στην τελευταία γραμμή παρουσιάζουμε αποτελέσματα με τη μέθοδο συνδυασμού πληροφορίας που παρουσιάστηκε στην προηγούμενη παράγραφο

Σε ότι αφορά την αρχική μέθοδο, [129] παρατηρούμε ότι αν και δίνει ικανοποιητικά αποτελέσματα σε περιοχές με έντονη υφή, π.χ. στις εικόνες με τους τίγρεις, σε ομαλές περιοχές το χαρακτηριστικό της κατεύθυνσης της υφής δίνει περίπλοκα όρια που δεν ανταποκρίνονται στα όρια των αντικειμένων. Επιπλέον, για παράδειγμα στην εικόνα με την ζέβρα, το μεγάλο πλάτος στα όρια των αντικειμένων διαστέλλει την περιοχή της εικόνας που ανατίθεται στο τμήμα της ζέβρας, καθώς η μεγάλη εκτίμηση του πλάτους στα όρια των αντικειμένων δεν μπορεί να εξηγηθεί ικανοποιητικά από την ανταγωνιζόμενη υπόθεση του πίσω μέρους της εικόνας το οποίο γενικότερα έχει

χαμηλό πλάτος διαμόρφωσης.

Τα χαρακτηριστικά που εξάγονται με τη μέθοδο των [193] είναι γενικότερα ομαλά και δίνουν αποτελέσματα συγκρίσιμα με τα χαρακτηριστικά που εξάγονται με τη μέθοδο Dominant Components Analysis, έχουν όμως το πρόβλημα ότι δεν φέρουν πληροφορία σχετικά με την κλίμακα της υψής και εμφανίζουν τα ίδια προβλήματα με τα χαρακτηριστικά τύπου Dominant Components Analysis σε ομαλές περιοχές. Σε πολλές εικόνες (π.χ. στο Σχ. 4.8 οι εικόνες με τον τίγρη, το λιοντάρι και τους βράχους) τα όρια των αντικειμένων ομαδοποιούνται σε μία ξεχωριστή περιοχή. Αυτό οφείλεται στο ότι οι μεγάλες τιμές της παραγώγου της εικόνας στα όρια ερμηνεύονται ως χαρακτηριστικά υψής, και απαιτούνε μία ξεχωριστή περιοχή της εικόνας για να τα ερμηνεύσει.

Τα αποτελέσματα της μεθόδου συνδυασμού χαρακτηριστικών γενικότερα δεν υποφέρουν από την σύγχυση μεταξύ πληροφορίας σχετικής με ακμή και υφή. Παίρνουμε γενικότερα καλύτερα αποτελέσματα κατάτμησης, με τα όρια των περιοχών να εντοπίζονται με ακρίβεια τα όρια των αντικειμένων. Παρατηρούμε επίσης ότι οι επιπτώσεις της ανώμαλης συμπεριφοράς του χαρακτηριστικού της κατεύθυνσης αντιμετωπίζεται με βάση το μικρότερο βάρος που έχει η υπόθεση υψής στις περιοχές αυτές. Μπορούμε όμως να εντοπίσουμε κάποιες εξαιρέσεις, όπως στην έκτη εικόνα με τα κτήρια όπου η υπόθεση της υψής για το αριστερό κτήριο δεν είναι αρκετά ισχυρή ώστε να οδηγήσει την κατάτμηση, με αποτέλεσμα την υπερκατάτμηση του στην περίπτωση του συνδυασμού πληροφορίας. Όμως, για τις περισσότερες εικόνες που εμφανίζονται εδώ, όπως και για τις περισσότερες άλλες εικόνες της βάσης [154], παρατηρούμε μία βελτίωση στην κατάτμηση χρησιμοποιώντας το νόμο της σταθμισμένης εξέλιξης καμπυλών, που είναι ιδιαίτερα εμφανή σε εικόνες που περιέχουν τόσο υφή όσο και ομαλές περιοχές.

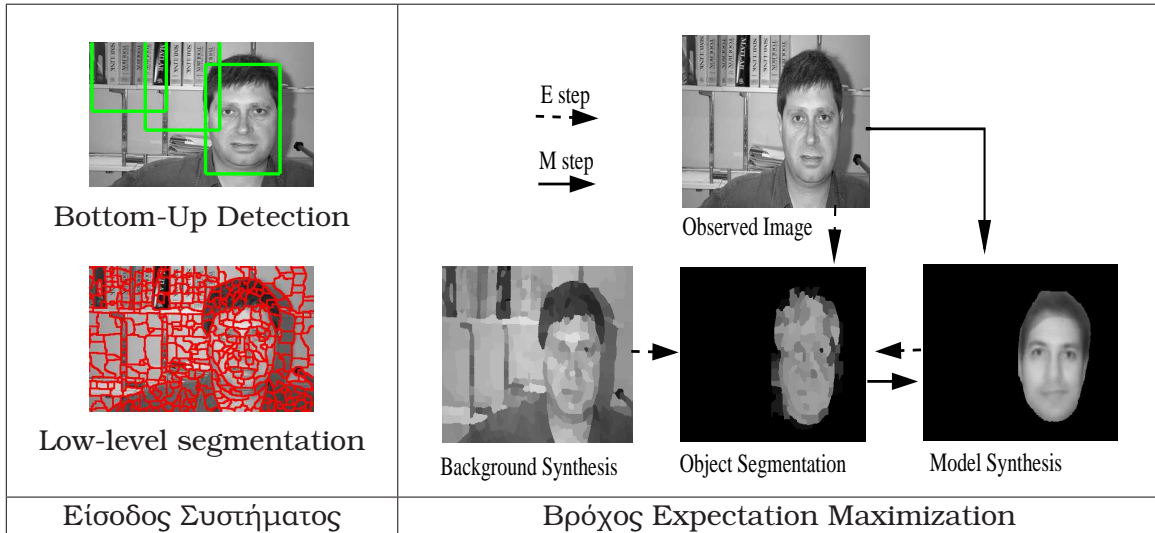
Κεφάλαιο 5

Συνέργεια μεταξύ Αναγνώρισης Αντικειμένων και Κατάτμησης Εικόνων μέσω του Αλγόριθμου Expectation Maximization

Στο κεφάλαιο αυτό εισάγουμε τον αλγόριθμο Expectation Maximization- EM ως ένα πλαίσιο για την διατύπωση και επίλυση του προβλήματος της συνέργειας μεταξύ της αναγνώρισης αντικειμένων και κατάτμησης εικόνων. Χρησιμοποιώντας αναγεννητικά μοντέλα για κατηγορίες αντικειμένων, η κατάτμηση διατυπώνεται ως το βήμα Expectation, όπου οι παρατηρήσεις ανατίθενται στις υποθέσεις των αντικειμένων και η ανάλυση αντικειμένων ως το βήμα Maximization, όπου τα μοντέλα προσαρμόζονται στις παρατηρήσεις.

Οι βασικές συνεισφορές αυτού του κεφαλαίου συνοψίζονται στα παρακάτω σημεία:

- Η εισαγωγή του αλγορίθμου EM για το πρόβλημα της συνέργειας επιτρέπει αφενός την διατύπωσή του με έναν μαθηματικά σαφή τρόπο, αφετέρου την χρήση διαφορετικών τεχνικών της όρασης υπολογιστών για την αντιμετώπισή του.
- Συνδέουμε τα μοντέλα ενεργούς εμφάνισης (Active Appearance Models - AAMs) με την κατάτμηση εικόνων, και επιτρέπουμε έτσι τον χειρισμό επικαλύψεων στην προσαρμογή μοντέλων AAM.
- Βάσει της μεταβολικής ερμηνείας του αλγορίθμου EM χρησιμοποιούμε τεχνικές εξέλιξης καμπυλών για την υλοποίηση του βήματος E, εισάγοντας έτσι τα μοντέλα AAM ως φορείς πρότερης γνώσης (shape priors) για την μεταβολική κατάτμηση εικόνων.
- Διερευνάται η χρησιμότητα της πληροφορίας κατάτμησης για την διαδικασία ανίχνευσης αντικείμενων. Χρησιμοποιώντας γνωστά σύνολα δεδομένων αποτιμώνται τα οφέλη της από κοινού κάτωθεν και άνωθεν αντιμετώπισης του προβλήματος, που αποδίδει συστηματικά καλύτερα αποτελέσματα.
- Επιδιώκοντας τη διεύρυνση της εφαρμοσιμότητας της προσέγγισής μας, ασχολούμαστε με την αυτοματοποιημένη κατασκευή μοντέλων χρησιμοποιώντας ένα σύνολο εκπαίδευσης με ελάχιστα σημεία αναφοράς (landmarks).



Σχήμα 5.1: Επισκόπηση της προσέγγισής μας.

Μετά από μία συνοπτική παρουσίαση του αλγορίθμου EM στην ενότητα 5.1 παρουσιάζουμε την προσέγγισή μας. Το βήμα E και η σχέση του με μεθόδους εξέλιξης καμπυλών αναλύονται στην ενότητα 5.2 ενώ στην ενότητα 5.3 παρουσιάζουμε τη διατύπωση του βήματος M στο πλαίσιο των AAM. Στην ενότητα 5.4 παρουσιάζονται τα πειράματα ανίχνευσης αντικειμένων, καθώς και η συνεισφορά μας στην αυτοματοποιημένη κατασκευή μοντέλων AAM, ενώ στην ενότητα 5.5 παρουσιάζουμε τη σχέση της προσέγγισής μας με τις κύριες αναφορές του πεδίου. Τεχνικές πτυχές καλύπτονται στα δύο παραρτήματα του κεφαλαίου.

5.1 Η προσέγγιση του EM για το Πρόβλημα της Συνέργειας

Επιδιώκουμε την επίλυση του προβλήματος της από κοινού κατάτμησης και ανίχνευσης αντικειμένων στο πλαίσιο της ανάλυσης-μέσω-σύνθεσης, όπου η ανάλυση των αντικειμένων μέσω ενός αναγεννητιού μοντέλου αποτελεί το μέσο και για τις δύο λειτουργίες. Αυτή η προσέγγιση οδηγεί σε ένα διαπλεκόμενο πρόβλημα, καθώς γνωρίζοντας μέσω της κατάτμησης τις παρατηρήσεις της εικόνας που ανατίθενται σε ένα αντικείμενο μπορούμε να εκτιμήσουμε τις παραμέτρους του και αντίστροφα. Ο EM είναι ένας επαναληπτικός αλγόριθμος για εκτίμηση παραμέτρων με ελλιπή δεδομένα, ο οποίος ταιριάζει ιδανικά με το πρόβλημά μας, παρέχοντας ένα εναργές και καλά θεμελιωμένο πλαίσιο για την διατύπωση του προβλήματος.

Πρωτού παρουσιάσουμε την προσέγγισή μας, εισάγουμε τον συμβολισμό και τις βασικές έννοιες με μία επισκόπηση της χρήσης του αλγορίθμου EM για το πρόβλημα της εκτίμησης παραμέτρων για μία κατανομή μίξης (mixture distribution).

5.1.1 Αλγόριθμος EM για μοντελοποίηση μέσω μίξης κατανομών

Θεωρούμε ότι έχουμε ένα σύνολο N ανεξάρτητων παρατηρήσεων $I = \{I_1, \dots, I_N\}$ που προκύπτουν από μία από τις K παραμετρικές κατανομές $P_k(I|\theta_k)$, που με

τη σειρά της επιλέγεται ανεξάρτητα για κάθε παρατήρηση με πρότερη πιθανότητα π_k . Η κατανομή l από την οποία η παρατήρηση I_n έχει προκύψει είναι άγνωστη και αναπαρίσταται χρησιμοποιώντας το διάνυσμα *κρυφών μεταβλητών* $Z_n = (z_{n,1}, \dots, z_{n,K})$, όπου $z_{n,j} = \delta(j - l)$ με δ την διακριτή συνάρτηση Dirac. Η σχέση που συνδέει την πιθανοφάνεια των παρατηρήσεων με το σύνολο παραμέτρων $\mathcal{A} = (\theta_1, \dots, \theta_K, \pi_1, \dots, \pi_K)$ προκύπτει αθροίζοντας πάνω στις κρυφές μεταβλητές:

$$\log P(I|\mathcal{A}) = \sum_{n=1}^N \log P(I_n|\mathcal{A}) = \sum_{n=1}^N \log \sum_{Z_n} P(I_n, Z_n|\mathcal{A}) \quad (5.1)$$

$$\text{όπου } P(I_n, Z_n|\mathcal{A}) = P(I_n|Z_n, \mathcal{A})P(Z_n|\mathcal{A}) = \prod_{k=1}^K [P_k(I_n|\theta_k)\pi_k]^{z_{n,k}} \quad (5.2)$$

Η μεγιστοποίηση αυτού του αθροίσματος ως προς το σύνολο παραμέτρων \mathcal{A} δεν είναι εύκολη, καθώς το άθροισμα πάνω στις κρυφές μεταβλητές Z_n εμφανίζεται μέσα στον λογάριθμο. Θεωρώντας όμως ότι το διάνυσμα κρυφών μεταβλητών Z_n είναι γνωστό, μπορούμε να εκφράσουμε τον λογάριθμο της πιθανοφάνειας των πλήρων παρατηρήσεων (full observations):

$$\log P(I, Z) = \sum_n \log P(I_n, Z_n) = \sum_n \sum_k z_{n,k} \log P_k(I_n|\theta_k)\pi_k \quad (5.3)$$

όπου το άθροισμα εμφανίζεται πλέον έξω από τον λογάριθμο, οπότε οι παράμετροι μπορούν να εκτιμηθούν απευθείας. Η ποσότητα που εισάγει ο EM για να εκμεταλλευτεί αυτήν την παρατήρηση είναι η προσδοκώμενη τιμή της (5.3) ως προς τις κρυφές μεταβλητές $z_{n,k}$, όπου η κατανομή πάνω στις $z_{n,k}$ αποκτάται βάσει των τρέχουσων εκτιμήσεων για τις παραμέτρους. Οδηγούμαστε έτσι σε ένα επαναληπτικό σχήμα:

- Βήμα E: υπολογισμός της ύστερης κατανομής $P_{\mathcal{A}^*}$ πάνω στα Z δεσμεύοντας στην προηγούμενη εκτίμηση των παραμέτρων, \mathcal{A}^* και τις παρατηρήσεις:

$$E_{n,k} \equiv P_{\mathcal{A}^*}(z_{n,k} = 1) = P(z_{n,k} = 1|I_n, \mathcal{A}^*) = \frac{\pi_k^* P_k(I_n|\theta_k^*)}{\sum_j \pi_j^* P_j(I_n|\theta_j^*)} \quad (5.4)$$

και σχηματισμός της προσδοκίας (expectation) του λογάριθμου πιθανοφάνειας των πλήρων παρατηρήσεων:

$$\langle \log P(I, Z) \rangle_{P_{\mathcal{A}^*}} = \sum_n \sum_k E_{n,k} \log \pi_k P_k(I_n|\theta_k) \quad (5.5)$$

Παραπάνω με τον συμβολισμό $\langle \cdot \rangle_{P_{\mathcal{A}^*}}$ συμβολίζουμε την προσδοκία της ποσότητας \cdot υπό την κατανομή $P_{\mathcal{A}^*}$.

- Βήμα M: μεγιστοποίηση της παραπάνω ποσότητας ως προς τις παραμέτρους των κατανομών:

$$\pi_k = \sum_n E_{n,k}/N, \quad \theta_k = \operatorname{argmax} \sum_n E_{n,k} \log P_k(I_n|\theta_k) \quad (5.6)$$

Διαισθητικά, στο βήμα E οι κρυφές δυαδικές μεταβλητές της (5.3) αντικαθίστανται με μία εκτίμηση της 'υπευθυνότητας' της κάθε κατανομής για τις παρατηρήσεις, που αργότερα χρησιμοποιείται για να απλοποιήσει την εκτίμηση παραμέτρων στο βήμα M.

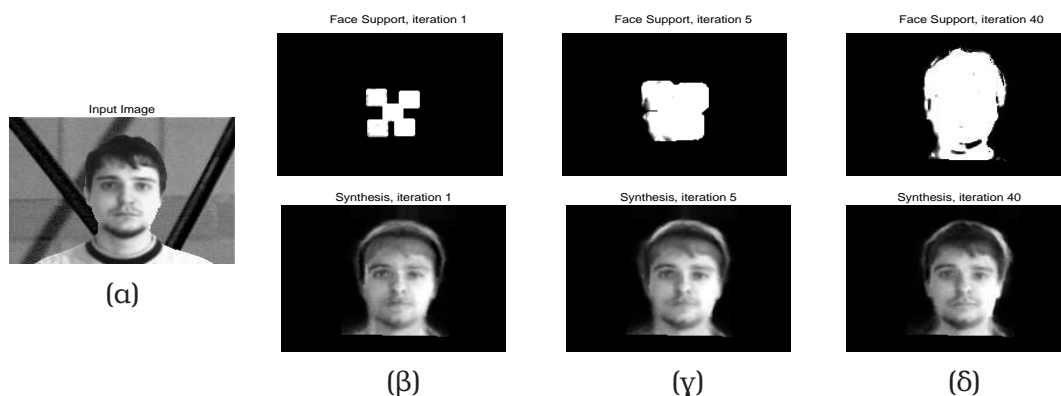
Αυτή η διαδικασία συστηματικά αυξάνει τον λογάριθμο της πιθανοφάνειας [52] και καταλήγει σε ένα τοπικό μέγιστο της (5.1).

5.1.2 Προσέγγιση στη συνέργεια βάσει του αλγορίθμου EM

Εφαρμόζουμε τον αλγόριθμο EM στο πρόβλημα της από κοινού κατάτμησης και ανάλυσης/ανίχνευσης αντικειμένων διατυπώνοντας την κατάτμηση ως το βήμα E, όπου οι παρατηρήσεις της εικόνας ανατίθενται κατά έναν πιθανοτικό τρόπο στην υπόθεση του αντικειμένου, και την ανάλυση του αντικειμένου ως το βήμα M, όπου οι παράμετροι του μοντέλου εκτιμώνται ώστε να μεγιστοποιηθεί η πιθανοφάνεια του τμήματος της εικόνας που έχει καταλάβει το αντικείμενο.

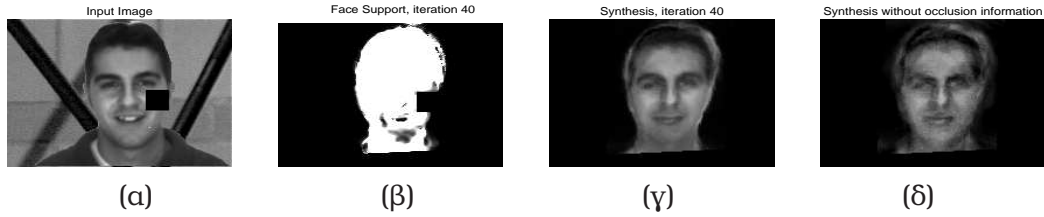
Η ιδέα πίσω από αυτήν την προσέγγιση είναι να θεωρηθεί η κατάτμηση ως ένα παράθυρο μέσα από το οποίο παρατηρείται το αντικείμενο, με δυαδικές μεταβλητές σε κάθε σημείο να καθορίζουν αν το αντικείμενο είναι ορατό η επικαλύπτεται. Η άνωθεν κατάτμηση τότε ανάγεται στην λήψη της απόφασης για το πού είναι καλύτερο, με την έννοια ότι αυξάνει την πιθανοφάνεια, να 'ανοίξουμε το παράθυρο' ενώ η προσαρμογή του μοντέλου εστιάζει μόνο στις περιοχές της εικόνας που φαίνονται μέσα από το παράθυρο.

Ένα απλό παράδειγμα φαίνεται στο Σχ. 5.2, όπου δείχνουμε τα αποτελέσματα της επαναληπτικής εφαρμογής των βημάτων E και M για μία απλοϊκή εικόνα· ξεκινώντας από μία τοποθεσία της εικόνας που προτείνεται από ένα αρχικό σύστημα ανίχνευσης, η σύνθεση και η κατάτμηση σταδιακά βελτιώνονται, συγκλίνοντας σε μία λύση που μοντελοποιεί ικανοποιητικά μία περιοχή της εικόνας μέσω του αντικειμένου.



Σχήμα 5.2: Εικόνα στην είσοδο (α) και επαναλήψεις του EM (β)-(δ). Πάνω γραμμή: εξέλιξη του χωρίου της υπόθεσης του αντικειμένου σε επίπεδο pixels (βήμα E). Κάτω γραμμή: αποτελέσματα σύνθεσης, χρησιμοποιώντας τα παραπάνω βάρη (βήμα M).

Ένα περιπλοκότερο παράδειγμα φαίνεται στο Σχ. 5.3, όπου μία σκοτεινή περιοχή επικαλύπτει ένα τμήμα του αντικειμένου. Η άνωθεν πρόβλεψη δεν είναι ακριβής στην επικαλυπτόμενη περιοχή, και οι παρατηρήσεις αυτές δεν ανατίθενται στο αντικείμενο κατά το βήμα E. Στο βήμα M συνεπώς τα σημεία αυτά δεν λαμβάνονται υπόψη κατά την εκτίμηση των παραμέτρων και έτσι επιτυγχάνεται η ευρωστία ως προς την



Σχήμα 5.3: Χειρισμός επικαλύψεων: Η σύνθεση στην εικόνα (γ) χρησιμοποιεί την κατάτμηση στην εικόνα (β), ενώ αυτή της εικόνας (δ) δεν έχει κάποια πληροφορία σχετικά με το χωρίο της περιοχής του προσώπου.

επικάλυψη του αντικειμένου· αντιθέτως, αγνοώντας τα αποτελέσματα του βήματος E τα αποτελέσματα της σύνθεσης χειροτερεύουν.

Σημαντικό για μία επιτυχή εφαρμογή είναι το συγκεκριμένο αναγεννητικό μοντέλο που χρησιμοποιείται με τον EM· για τα προηγούμενα σχήματα χρησιμοποιήθηκε μία βάση PCA για την φωτεινότητα προσώπων [220], όπου δεν υπάρχει κάποια διαδικασία αντιστοίχισης των προσώπων. Συνέπεια αυτού είναι τα διάφορα ανεπιθύμητα φαινόμενα που εμφανίζονται στη σύνθεση, για παράδειγμα γύρω από τα μαλλιά, ενώ δεν μπορούμε να αναφερόμαστε σε κάτι όπως το χωρίο του αντικειμένου, καθώς δεν υπάρχει σύμπτωση των χωρίων των αντικειμένων βάσει των οποίων κατασκευάστηκε το μοντέλο.

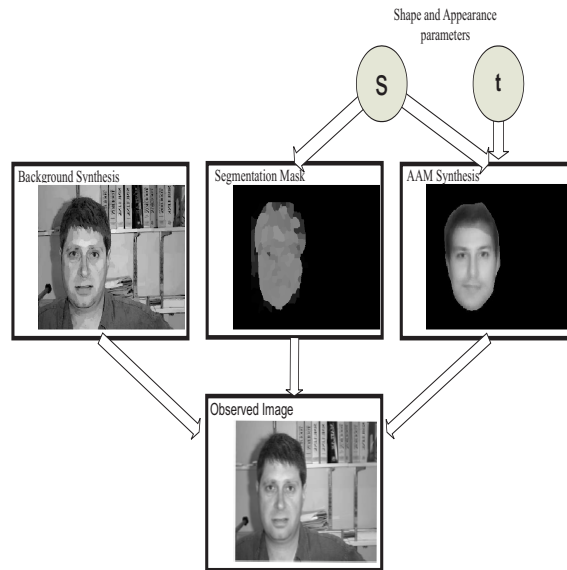
Τα προβλήματα αυτά μπορούν να αντιμετωπισθούν χρησιμοποιώντας μοντέλα ενεργούς εμφάνισης - AAM [112, 37, 157] που παρουσιάστηκαν στην ενότητα 2.2.5 ως αναγεννητικά μοντέλα που λαμβάνοντας υπόψη τις χωρικές παραμορφώσεις των αντικειμένων μπορούν να χειριστούν τόσο τις διαδικασίες της ανάλυσης/ανίχνευσης όσο και της κατάτμησης.

Διατυπώνοντας τα AAM πιθανοτικά η τυπική υπόθεση που χρησιμοποιείται [39] είναι ότι το σφάλμα μεταξύ της πρόβλεψης και των παρατηρήσεων μπορεί να μοντελοποιηθεί ως μία διαδικασία λευκού Gaussian θορύβου, τυπικής απόκλισης σ . Ο λογάριθμος της πιθανοφάνειας της παρατήρησης στο σημείο \mathbf{x} υπό το μοντέλο AAM μπορεί τότε να γραφεί:

$$\log P(I(\mathbf{x})|\mathbf{s}, \mathbf{t}) = \left[\frac{(I(\mathbf{x}) - \mathcal{T}(\mathcal{S}^{-1}(\mathbf{x})))^2}{2\sigma^2} - \frac{\log 2\pi\sigma^2}{2} \right] \quad (5.7)$$

όπου η συνάρτηση \mathcal{S}^{-1} φέρνει από το σύστημα συντεταγμένων της τεμπλέτας την πρόβλεψη της εμφάνισης $\mathcal{T}(\mathcal{S}^{-1}(\mathbf{x}))$ που αντιστοιχεί στην τιμή φωτεινότητας $I(\mathbf{x})$. Θεωρώντας ότι η μεγέθυνση/σμίκρυνση της περιοχής γύρω από το σημείο \mathbf{x} είναι αμελητέα, μπορούμε τότε να γράψουμε για το κριτήριο (2.31) $P(I) \propto \exp(-E(\mathbf{s}, \mathbf{t}))$, οπότε η προσαρμογή του AAM μπορεί να θεωρηθεί ως η εκτίμηση των παραμέτρων μέγιστης πιθανοφάνειας.

Μία μικρή βελτίωση της παραπάνω προσέγγισης στην οποία αναφερόμαστε στην ενότητα 5.3 συνίσταται στην χρήση της Jacobian της παραμόρφωσης στο κριτήριο που χρησιμοποιείται, ώστε να ληφθούν υπόψη οι χωρικές αλλοιώσεις του αντικειμένου στο κριτήριο πιθανοφάνειας. Μία σημαντικότερη βελτίωση που είναι δυνατή στα πλαίσια του EM είναι ο χειρισμός επικαλύψεων ή απροσδόκητων καταστάσεων φωτισμού χρησιμοποιώντας μοντέλα AAM. Βλέποντας τα μοντέλα AAM στο πλαίσιο των αναγεννητικών μοντέλων και του EM αντιμετωπίζουμε αυτό το πρόβλημα επιτρέποντας εναλλακτικές υποθέσεις να ερμηνεύσουν τις παρατηρήσεις, και απαλάσσοντας τα



Σχήμα 5.4: Το αναγεννητικό μοντέλο της προσέγγισής μας: οι παράμετροι σχήματος και εμφάνισης s, t συνθέτουν την εμφάνιση του αντικειμένου μέσω ενός μοντέλου, που στη συνέχεια συνδυάζεται με το τμηματικά σταθερό μοντέλο του φόντου. Ο συνδυασμός επιτυγχάνεται μέσω ενός πίνακα κατάτμησης (κρυφές μεταβλητές), που καθορίζει ποια τμήματα της εικόνας ανατίθενται στην υπόθεση του αντικειμένου. Για την εκτίμηση του πίνακα κατάτμησης συνδυάζεται η πιστότητα του μοντέλου στις παρατηρήσεις με πληροφορία σχετικά με το χωρίο του αντικειμένου. Για εκτίμηση παραμέτρων η πιθανοφάνεια των παρατηρήσεων που βρίσκονται μέσα στη 'μάσκα' μεγιστοποιείται.

AAM από την ερμηνεία των παρατηρήσεων, χωρίς αυτό να απαιτεί μία τροποποίηση της νόρμας σφάλματός τους. Αντίθετα, η τυπική προσέγγιση, π.χ. [81] είτε θεωρεί τις περιοχές που επικαλύπτονται ως γνωστές εκ των προτέρων, ή βασίζεται σε μία ευρετικά ορισμένη εύρωστη νόρμα για να αποφύγει την προσαρμογή του μοντέλου στα δεδομένα με επικάλυψη.

5.2 Βήμα Ε: Κατάτμηση Βάσει Αντικειμένων

Ένα πρακτικό πρόβλημα που προκύπτει όταν θεωρούμε το πρόβλημα της κατάτμησης ενός αντικειμένου είναι πώς μπορεί να οριστεί η συμπληρωματική υπόθεση του φόντου· αυτό το οποίο χρειάζεται γενικότερα είναι να είναι διαθέσιμη μία μηδενική υπόθεση ώστε να μπορεί να συγκριθεί με τις προβλέψεις του μοντέλου του αντικειμένου. Μία απλοϊκή προσέγγιση που χρησιμοποιήθηκε για το Σχ. 5.2 χρησιμοποιεί μία μη παραμετρική κατανομή για όλες τις παρατηρήσεις του φόντου· όμως για εικόνες με μεγάλη μεταβολή της εμφάνισης στο φόντο αυτή η κατανομή γίνεται σχεδόν ομοιόμορφη, και το μοντέλο του αντικειμένου προβλέπει την φωτεινότητα της εικόνας καλύτερα ακόμα και όταν δεν υπάρχει στην πραγματικότητα κάποιο αντικείμενο.

Ένα συμπληρωματικό πρόβλημα είναι η κανονικοποίηση του βήματος Ε: για το Σχ. 5.2 χρησιμοποιήθηκε πληροφορία από την γειτονιά του κάθε pixel, πολλαπλασιάζοντας τα αποτελέσματα του βήματος Ε με τον μέσο της γειτονιάς των αποτελεσμάτων του προηγούμενου βήματος, δίνοντας έτσι μία σταδιακή εξέλιξη του χωρίου του αντικειμένου προς τα έξω. Αυτή η ευρετική η περιπλοκότερες τεχνικές μπορούν

να δαπανήσουν χρόνο εωσότου ένα τμήμα της εικόνας μεταβεί από την μία υπόθεση στην άλλη, ενώ είναι ευκολότερο να κολλήσουν σε τοπικά ελάχιστα.

Παρακινούμενοι από τα παραπάνω παρουσιάζουμε στη συνέχεια δύο διαφορετικές προσεγγίσεις στην υλοποίηση του βήματος Ε.

5.2.1 Υλοποίηση του βήματος Ε σε επίπεδο τμημάτων

Η πρώτη προσέγγιση που ακολουθήθηκε είναι να χρησιμοποιηθεί εξαρχής μία υπερκατάτμηση της εικόνας και να αναπαρασταθεί η υπόθεση του φόντου μέσω ενός ξεχωριστού μοντέλου για κάθε τμήμα. Μία σχετιζόμενη κανονικοποίηση του βήματος είναι να θεωρήσουμε ότι οι κρυφές μεταβλητές δεν δείχνουν πλέον αν μεμονωμένα pixels ανήκουν στην υπόθεση του αντικειμένου, αλλά τμήματα της εικόνας. Η υπόθεση του φόντου μοντελοποιεί τις παρατηρήσεις I_R μέσα στην περιοχή R μέσω μίας σταθερής τιμής φωτεινότητας, που έχει αλλοιωθεί από Gaussian θόρυβο· οι παρατηρήσεις I_R μπορούν τότε να ανατεθούν στην υπόθεση \mathcal{O} του αντικειμένου ή \mathcal{B} του φόντου βάσει του τύπου του Bayes συγκρίνοντας τις πιθανοφάνειες των παρατηρήσεων:

$$E_{R,\mathcal{O}} = \frac{P(I_R|\mathcal{O})}{P(I_R|\mathcal{O}) + P(I_R|\mathcal{B})} \quad P(I_R|\mathcal{H}) \simeq P(I_R|F_{\mathcal{H}}^*, \mathcal{H})P(F_{\mathcal{H}}^*|\mathcal{H}), \quad \mathcal{H} \in \{\mathcal{O}, \mathcal{B}\} \quad (5.8)$$

Στην παραπάνω έκφραση η $P(F_{\mathcal{H}}^*|\mathcal{H})$ είναι μία πρότερη κατανομή πάνω στις παραμέτρους του μοντέλου και $F_{\mathcal{H}}^*$ είναι οι βέλτιστες με το κριτήριο MAP παράμετροι για την υπόθεση \mathcal{H} . Για το μοντέλο του φόντου θεωρούμε ότι έχουμε μία ομοιόμορφη πρότερη κατανομή πάνω στην σταθερή τιμή και την διασπορά του θορύβου, οπότε καταλήγουμε απλά στις εκτιμήτριες μέγιστης πιθανοφάνειας (το μέσο και τη διασπορά των παρατηρήσεων), ενώ για τα μοντέλα AAM χρησιμοποιούμε τα αποτελέσματα του βήματος M, το οποίο παρουσιάζεται στην επόμενη ενότητα.

Εφαρμόζοντας την παραπάνω σχέση στην πράξη, παρατηρούμε ότι το μέγεθος του τμήματος της εικόνας έχει μία ανεπιθύμητη επίδραση στην αυστηρότητα με την οποία λαμβάνεται η απόφαση για την ανάθεση του τμήματος. Αυτό οφείλεται στην απλοποιητική υπόθεση ότι το σφάλμα ανακατασκευής μπορεί να μοντελοποιηθεί ως μία διαδικασία ανεξάρτητου θορύβου. Τόσο για τα AAM όσο και για το μοντέλο του φόντου εξετάζοντας προσεκτικότερα το σφάλμα παρατηρούμε ότι είναι έντονα συσχετισμένο, οπότε ουσιαστικά υπερεκτιμούμε την μαρτυρία που μας προσφέρουν οι παρατηρήσεις μας. Για να αντιμετωπισθεί αυτό διαιρούμε τον λογάριθμο της πιθανοφάνειας των παρατηρήσεων I_R με ένα πολλαπλάσιο του πλήθους των παρατηρήσεων μέσα στην περιοχή $|R|$, παίρνοντας έτσι την ανεξάρτητη-κλίμακας ποσότητα: $\log P_{R,\mathcal{H}} = \alpha \frac{\sum_{i \in R} \log P(I_i|F_{\mathcal{H}}^*, \mathcal{H})}{|R|}$ όπου α είναι μία ελεύθερη παράμετρος. Μία κομψότερη προσέγγιση θα ήταν να εισάγουμε την υπόθεση του θορύβου Brown ώστε να μοντελοποιήσουμε τη συσχέτιση του σφάλματος ανακατασκευής· αυτό όμως θα περιέπλεκε τις σχέσεις εκτίμησης παραμέτρων.

Ένα δεύτερο σημείο είναι ότι το μοντέλο του φόντου γενικότερα είναι λεπτομερέστερο από ένα AAM, με την έννοια ότι το τελευταίο τυπικά μοντελοποιεί μία περιοχή που είναι 20 έως 30 φορές μεγαλύτερη από αυτήν ενός τμήματος που προκύπτει από την υπερκατάτμηση της εικόνας. Συνεπώς, το σφάλμα ανακατασκευής του μοντέλου σταθερής φωτεινότητας θα είναι σχεδόν πάντα μικρότερο, ή συγκρίσιμο με αυτό ενός AAM. Για το λόγο αυτό εισάγουμε έναν όρο ποινής τύπου MDL, ο οποίος 'τιμωρεί' το

μοντέλο του φόντου και εξισορροπεί την ανάθεση των παρατηρήσεων.

Τέλος, αν και απαιτείται η εισαγωγή της πιθανοφάνειας των παραμέτρων στην τελική απόφαση, αυτό περιπλέκει την αποτίμηση των πειραματικών αποτελεσμάτων που γίνονται στη συνέχεια, στην ενότητα 5.4. Συγκεκριμένα, εκεί διερευνάται το κατά πόσο τα αποτελέσματα κατάτμησης προσφέρουν συμπληρωματική πληροφορία ως προς αυτήν που παρέχουν οι παράμετροι του AAM· συνεπώς το να εισάγουμε τις τελευταίες μέσα στη διαδικασία κατάτμησης θα περιέπλεκε την ανάλυση των αποτελεσμάτων. Για αυτό χρησιμοποιούμε μόνο τον όρο πιστότητας δεδομένων $\log P_{R,\mathcal{H}}$ και παρουσιάζουμε εδώ τα αποτελέσματα άνωθεν κατάτμησης που αποκτούνται στις τοποθεσίες που ανταποκρίνονται σε αντικείμενα, αφήνοντας την από κοινού κατάτμηση και ανίχνευση αντικειμένων για την ενότητα 5.4 όπου το πρόβλημα αντιμετωπίζεται διεξοδικότερα.

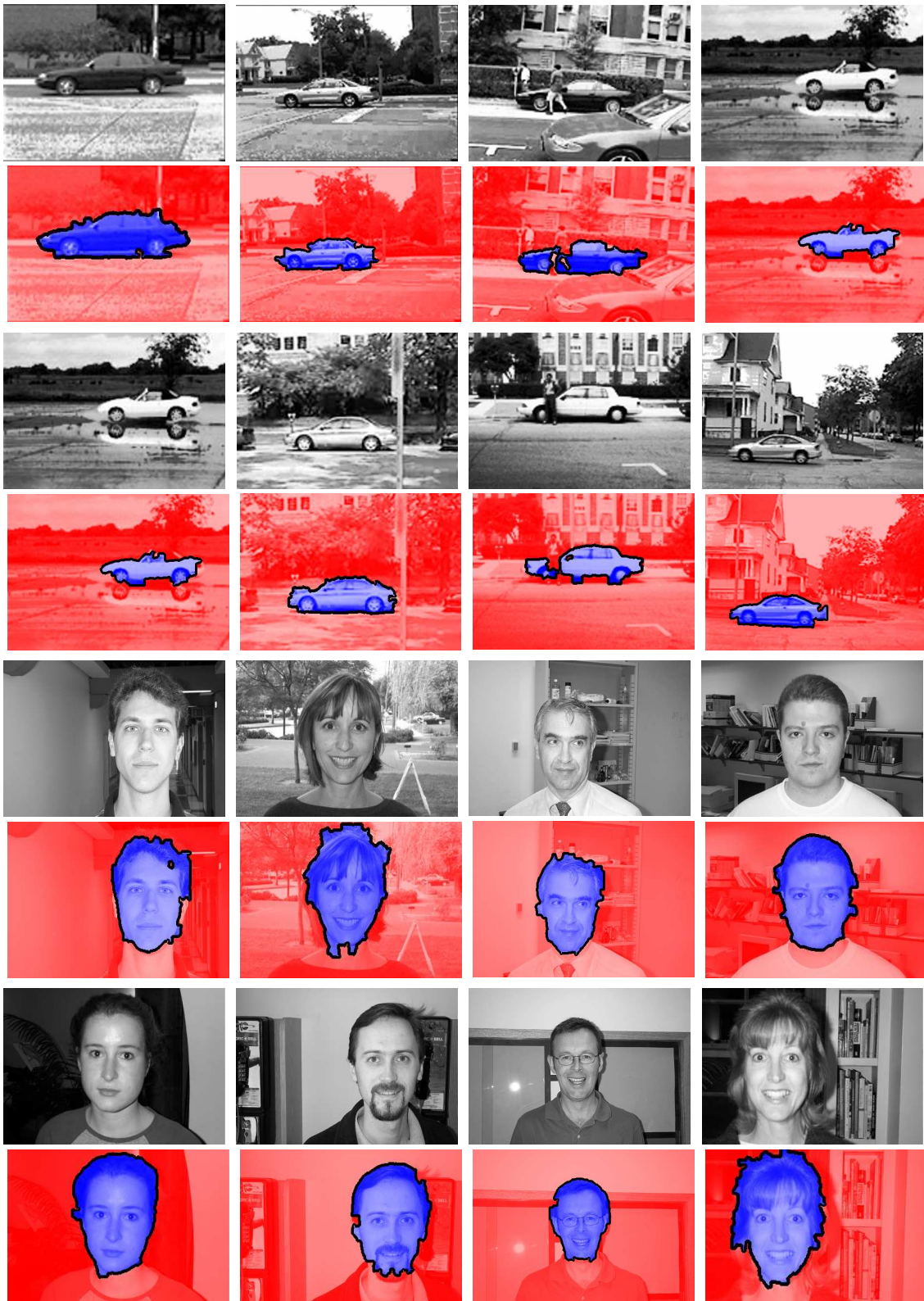
Έχοντας εισάγει όλους τους κανονικοποιητικούς όρους στο παραπάνω κριτήριο, η ύστερη πιθανότητα τα I_R να ανήκουν στην υπόθεση \mathcal{O} μπορεί να γραφεί συναρτήσει του όρου:

$$\log \frac{P(I_R|\mathcal{O})}{P(I_R|\mathcal{B})} = \alpha \left(\frac{\sum_{i \in R} \log P(I_i|F_{\mathcal{O}}^*, \mathcal{O}) - \log P(I_i|F_{\mathcal{B}}^*, \mathcal{B})}{|R|} \right) + \beta \quad (5.9)$$

Βλέποντας την διαδικασία του βήματος E ως την πιθανοτική ταξινόμηση των τμημάτων βάσει της σχέσης $E_{R,\mathcal{O}} = \frac{1}{1 + \exp(\log \frac{P(I_R|\mathcal{O})}{P(I_R|\mathcal{B})})}$ μπορούμε να εκτιμήσουμε τις σταθερές α, β με τεχνικές εκμάθησης μηχανών, αν έχουμε στη διάθεσή μας χειρωνακτικά κατετημημένες εικόνες στο σύνολο εκπαίδευσης. Αν και αρχικά ακολουθήθηκε αυτή η προσέγγιση, [131], στη συνέχεια παρατηρήθηκε ότι για ένα ευρύ φάσμα τιμών τα αποτελέσματα ήταν ποιοτικά όμοια. Αποφεύγοντας έτσι την σχετικά περίπλοκη παραπάνω λύση, τελικά επιλέξαμε δύο τιμές ρυθμίζοντας το σύστημά για λίγες εικόνες, και χρησιμοποιήσαμε στη συνέχεια τις ίδιες τιμές και για τις δύο κατηγορίες αντικειμένων.

Μετά από πειραματισμό με απλές και αποτελεσματικές τεχνικές κατάτμησης εικόνας καταλήξαμε στον αλγόριθμο watershed· οι ακμές της εικόνας εκτιμώνται μέσω του τελεστή παραγώγου φωτεινότητας [154] που συνδυάζει ταχύτητα με αποδοτικότητα, ενώ τα τοπικά ελάχιστα της έντασης ακμών χρησιμοποιήθηκαν ως σημεία εκκίνησης (markers).

Στο Σχ. 5.5 δείχνουμε το σύνορο της περιοχής που προκύπτει κατωφλιώνοντας τα αποτελέσματα του βήματος E για το αντικείμενο που αντιστοιχεί στο ισχυρότερο αποτέλεσμα κάτωθεν ανίχνευσης. Από τα αποτελέσματα φαίνεται η δυνατότητα της προσέγγισής μας να κατατμεί αντικείμενα μίας κατηγορίας με μεγάλη ποικιλία στην εμφάνιση και στο σχήμα. Τα αποτελέσματα είναι γενικότερα ικανοποιητικά, καθώς η κατάτμηση των αντικειμένων που παρουσιάζουν μεγάλη ποικιλία στο σχήμα και στην εμφάνιση εκτιμάται σωστά, ενώ αποκλείονται μη προβλέψιμες περιοχές, όπως γένια για πρόσωπα ή πεζοί για αυτοκίνητα. Όμως, η κατωφλίωση δεν μας επιτρέπει να επιβάλουμε την ομαλότητα των ορίων που προκύπτουν, με αποτέλεσμα να προκύπτουν γωνιώδη όρια όποτε τα αποτελέσματα του βήματος E για κάποιο μικρό τμήμα είναι κάτω από το κατώφλι. Επίσης, οι ανακρίβειες του αρχικού συστήματος κατάτμησης μεταφέρονται και στην υψηλού επιπέδου κατάτμηση, όπως φαίνεται έντονα για τις εικόνες των αυτοκινήτων, όπου η χαμηλού επιπέδου πληροφορία είναι θορυβώδης.



Σχήμα 5.5: Αποτελέσματα άνωθεν κατάτμησης για εικόνες αυτοκινήτων και προσώπων, χρησιμοποιώντας το βήμα Ε σε επίπεδο τμημάτων. Χρησιμοποιώντας ένα σταθερό κατώφλι για τα αποτελέσματα του βήματος Ε, τα τμήματα της εικόνας που ανήκουν στην υπόθεση του αντικειμένου της ανατίθενται. Αν και η κατάτμηση του αντικειμένου είναι χοντρικά σωστή, εμφανίζονται προβλήματα λεπτής κλίμακας, όπως μία ατυχής κατάτμηση χαμηλού επιπέδου να έχει ως αντίκτυπο στην άνωθεν κατάτμηση τον ανακριβή εντοπισμό των ορίων και γωνιώδη περιγράμματα.

5.2.2 Υλοποίηση του βήματος E χρησιμοποιώντας Εξέλιξη Καμπυλών

Παρά την απλότητα και αποδοτικότητα της η προηγούμενη προσέγγιση δεν επιτρέπει την εισαγωγή περιορισμών όπως η ομαλότητα των συνόρων, ή την σύμπτωση των συνόρων των αντικειμένων με τις ακμές της εικόνας. Η χρήση τέτοιων πηγών πληροφορίας είναι συνηθισμένη στην κατάτμηση εικόνων, οπότε θα ήταν επιθυμητό να συνδυαστεί η άνωθεν πληροφορία που παρέχει ένα AAM με κάποιον άλλο ευέλικτο αλγόριθμο κατάτμησης, παραμένοντας στο πλαίσιο του EM. Βάσει της μεταβολικής διατύπωσης του αλγορίθμου EM στη συνέχεια συνδυάζουμε την προσέγγισή μας με τεχνικές εξέλιξης καμπυλών, εισάγοντας έτσι τα AAM ως φορείς πρότερης γνώσης σχετικά με το σχήμα και την εμφάνιση για την κατάτμηση εικόνων.

Ερμηνεία αλγορίθμου Region Competition μέσω του EM

Υπάρχει μία σαφής διαισθητική σχέση μεταξύ του αλγορίθμου EM και του Region Competition: θεωρώντας την κατάτμηση της εικόνας ως ένα πεδίο κρυφών μεταβλητών, το βήμα E αντιστοιχεί στην εξέλιξη καμπυλών, και το βήμα M στην επανεκτίμηση των παραμέτρων των κατανομών. Εδώ καθιερώνουμε και μία μαθηματικά σαφή διατύπωση της σχέσης μεταξύ τους χρησιμοποιώντας την μεταβολική ερμηνεία του EM [170, 107].

Συγκεκριμένα, το βήμα E μπορεί να ερμηνευθεί επίσης ως ένα βήμα μεγιστοποίησης: εισάγοντας μία οποιαδήποτε κατανομή $Q(Z)$ πάνω στις κρυφές μεταβλητές, ένα κάτω φράγμα του $\log P(I|\mathcal{A})$ αποκτάται εφαρμόζοντας την ανισότητα του Jensen στην συνάρτηση του λογαρίθμου:

$$\log \left(\sum_Z \frac{P(I, Z|\mathcal{A})}{Q(Z)} Q(Z) \right) \geq \sum_Z Q(Z) \log \left(\frac{P(Z, I|\mathcal{A})}{Q(Z)} \right) \quad (5.10)$$

$$\begin{aligned} &= \sum_Z Q(Z) \log P(I|\mathcal{A}) + \sum_Z Q(Z) \log \frac{P(Z|I, \mathcal{A})}{Q(Z)} \\ &= \log P(I|\mathcal{A}) - KL(Q(Z)||P(Z|I, \mathcal{A})) \end{aligned} \quad (5.11)$$

όπου για να πάμε στη δεύτερη γραμμή χρησιμοποιήθηκε ο κανόνας αλυσίδας $P(I, Z|\mathcal{A}) = P(Z|I, \mathcal{A})P(I|\mathcal{A})$. Ο πρώτος όρος δεν επηρεάζεται από την επιλογή της Q , ενώ ο όρος KL συμβολίζει την απόκλιση Kullback-Leibler μεταξύ των κατανομών $Q(Z)$ και $P(Z|I, \mathcal{A})$, που είναι μία μη-αρνητική ποσότητα. Παρατηρούμε ότι για την επιλογή που γίνεται κατά το βήμα E, $Q(Z) = P(Z|I, \mathcal{A})$, η απόκλιση KL ισούται με μηδέν, οπότε αυτό ισοδυναμεί με την ελαχιστοποίηση του κάτω ορίου (5.10) ως προς την $Q(Z)$. Η απευθείας μεγιστοποίηση της ποσότητας $\log(P(I|\mathcal{A}))$ αντικαθίσταται λοιπόν με την διαδοχική μεγιστοποίηση αυτού του κάτω ορίου, θέτοντας την $Q(Z)$ ίση με $P(Z|I, \mathcal{A})$ (βήμα -E) και ακολούθως κάνοντας πιο στενό το όριο τροποποιώντας τις παραμέτρους (βήμα-M).

Συνεπώς και τα δύο βήματα του EM μπορούν να θεωρηθούν ως ελαχιστοποιήσεις

του κάτω ορίου της πιθανοφάνειας των παρατηρήσεων(5.10):

$$\sum_Z Q(Z) \log \left(\frac{P(I, Z|\mathcal{A})}{Q(Z)} \right) = \sum_Z Q(Z) \log P(I|Z, \mathcal{A})P(Z|\mathcal{A}) + S$$

$$Q(Z)=\prod_i Q_i(Z_i) \sum_n \sum_{Z_n} Q_n(Z_n) \log P(I_n|Z_n, \mathcal{A}) + \sum_Z Q(Z) \log P(Z|\mathcal{A}) + S, \quad (5.12)$$

όπου $S = -\sum_Z Q(Z) \log(Q(Z))$ είναι η εντροπία της κατανομής Q . Καθώς το Z_n είναι ένα K -διάστατο διάνυσμα με στοιχεία $z_{n,i} = \delta(i-l)$ το οποίο κωδικοποιεί την ανάθεση της παρατήρησης I_n στην υπόθεση l , έχουμε $P(I_n|Z_n, \mathcal{A}) = \prod_k P_k(I_n|\mathcal{A}_k)^{z_{n,k}}$, οπότε μπορούμε και να γράψουμε:

$$\sum_{Z_n} Q_n(Z_n) \log P(I_n|Z_n, \mathcal{A}) = \sum_k Q_{n,k} \log P_k(I_n|\mathcal{A}_k) \quad (5.13)$$

Παραπάνω $Q_{n,k}$ είναι η προσδοκώμενη τιμή της δυαδικής μεταβλητής $z_{n,k}$. Στον αλγόριθμο Region Competition η $Q_{n,k}$ επιτρέπεται να λάβει μόνο δυαδικές τιμές, καθώς δεν υπάρχουν παρατηρήσεις που να ανήκουν συγχρόνως σε παραπάνω από μία περιοχές. Μπορούμε λοιπόν να δούμε τον πρώτο όρο ως μία διακριτοποίηση του όρου πιστότητας δεδομένων του συναρτησιακού Region Competition, καθώς αθροίζει τον λογάριθμο της πιθανοφάνειας των παρατηρήσεων υπό την υπόθεση στην οποία έχουν ανατεθεί:

$$\sum_k \sum_{n \in R_k} \log P_k(I_n|\mathcal{A}_k) = \sum_n \sum_k Q_{n,k} \log P_k(I_n|\mathcal{A}_k), \quad R_k = \{n : Q_{n,k} = 1\} \quad (5.14)$$

Ο δεύτερος όρος εισάγει πρότερη γνώση στην κατάτμηση, χρησιμοποιώντας είτε παραμετρικές εκφράσεις είτε γενική πληροφορία. Στο πρόβλημα της μοντελοποίησης μέσω μίας μίξης κατανομών, αυτό αντιστοιχεί στην χρήση των πρότερων πιθανοτήτων για να ανατεθούν ευκολότερα παρατηρήσεις στις μεγαλύτερες συνιστώσες της μίξης. Στον αλγόριθμο Region Competition αυτός ο όρος δεν εξαρτάται από παραμέτρους και εισάγει μία γενικής φύσης ποινή σε μεγάλα σύνορα περιοχών· μεταβαίνοντας από τη συνεχή στην διακριτή διατύπωση, μπορούμε να γράψουμε τον όρο του φράγματος του EM που συνδέεται με το μήκος καμπύλης ως

$$P(Z|\mathcal{A}) = P(Z) = \prod_i \exp(-\partial Z_i) \quad (5.15)$$

όπου το ∂Z_i καταδεικνύει αν στην τοποθεσία i συμβαίνει μία μετάβαση στο label της κρυφής μεταβλητής.

Τέλος, ο όρος εντροπίας γενικότερα επιβάλλει ομαλές αναθέσεις των παρατηρήσεων στις διαθέσιμες υποθέσεις· καθώς στον Region Competition οι αναθέσεις γίνονται κατά έναν δυαδικό τρόπο, αυτός ο όρος είναι πάντα μηδενικός και δεν επηρεάζει το φράγμα.

Καταλήγοντας, αυτό που διαφοροποιεί τον EM για μοντελοποίηση μέσω μίξης κατανομών από τον αλγόριθμο Region Competition είναι η χρήση δυαδικών αποφάσεων για κάθε παρατήρηση, αντί ομαλών ύστερων πιθανοτήτων, η εισαγωγή ενός γεωμετρικού όρου που επιβάλλει ομαλά όρια, και η μέθοδος εξέλιξης των ορίων που εστιάζει στην περιφέρεια των περιοχών αντί να χρησιμοποιεί όλο το χωρίο τους. Μπο-

ρούμε να δούμε τις ιδιαιτερότητες αυτές του Region Competition στο πλαίσιο του EM ως εξής

Region Competition	EM
Εξέλιξη καμπυλών	Βήμα E
Ανανέωση παραμέτρων	Βήμα M
Χρήση δυαδικών αποφάσεων	Winner-Take-All EM [170]
(Τοπική) εξέλιξη καμπυλών	Freezing [170]
Όρος ποινής για μήκος ορίων	MRF πάνω στις κρυφές μεταβλητές

Έτσι αντί για ένα σχήμα που θα αγνοούσε τη γεωμετρική φύση του προβλήματος και θα χειριζόταν την εικόνα ως ένα απλό σύνολο παρατηρήσεων, ο Region Competition εισάγει χρήσιμη γεωμετρική πληροφορία, συνδυάζοντας στατιστικές και γεωμετρικές τεχνικές εξέλιξης καμπυλών.

Χρήση μοντέλων AAM για μεταβολική κατάτμηση εικόνων

Για να ενσωματώσουμε τα μοντέλα AAM στο παραπάνω πλαίσιο, θεωρούμε ότι η περιοχή k αντιστοιχεί στην υπόθεση \mathcal{O} ενός μοντέλου AAM, και εστιάζουμε στους όρους του φράγματος του EM που εμπλέκονται. Αυτοί περιλαμβάνουν τον όρο πιστότητας δεδομένων που σχετίζει την πρόβλεψη του μοντέλου με τις παρατηρήσεις:

$$\sum_n Q_{n,\mathcal{O}} \log P_{\mathcal{O}}(I_n | \mathbf{s}, \mathbf{t})$$

και τους περιορισμούς πάνω στις δυνατές αναθέσεις μέσω των κρυφών μεταβλητών που εκφράζονται ως

$$\sum_Z Q(Z) \log P(Z | \mathcal{A}) = \underbrace{\sum_Z Q(Z) \log P(Z)}_{\text{Όρος ομαλότητας}} + \underbrace{\sum_n Q_{n,\mathcal{O}} \log P(z_{n,\mathcal{O}} | \mathbf{s})}_{\text{Σχήμα AAM}} \quad (5.16)$$

Παραπάνω συνδυάζουμε τον όρο ποινής για ομαλά όρια $\sum_Z Q(Z) \log P(Z)$ με τον όρο που εξαρτάται από την τρέχουσα εκτίμηση του σχήματος του αντικειμένου. Θεωρώντας ότι n είναι ο δείκτης της παρατήρησης στο σημείο \mathbf{x} θα μπορούσαμε να θέσουμε $P(z_{n,\mathcal{O}} | \mathbf{s}) = H(\mathcal{S}^{-1}(\mathbf{x}, \mathbf{s}))$, περιορίζοντας τις παρατηρήσεις που ανατίθενται στο αντικείμενο να βρίσκονται στο εσωτερικό του. Για να αποφύγουμε αριθμητικά προβλήματα ομαλοποιούμε πρώτα τη χαρακτηριστική συνάρτηση του χωρίου του αντικειμένου συνελίσσοντας την με ένα Gaussian φίλτρο και χρησιμοποιούμε την τιμή που προκύπτει.

Μπορούμε να εισάγουμε απευθείας τους νέους όρους αυτούς καθώς σπάνε σε αθροίσματα πάνω στο πλέγμα της εικόνας· μεταγράφοντας την σχέση (5.17) για την περίπτωση μας καταλήγουμε στη σχέση εξέλιξης για το σύνολο του αντικειμένου στην περιοχή \mathbf{x}

$$\frac{\partial \Gamma_k}{\partial t} = -\mu \kappa \mathcal{N} + \log \frac{P_k(I | \mathbf{s}, \mathbf{t}, \mathcal{O}) P_k(\mathcal{S}^{-1}(\mathbf{x}, \mathbf{s}) | \mathcal{O})}{P_j(I | \mathcal{A}_j)} \mathcal{N} \quad (5.17)$$

Σε κάθε σημείο λοιπόν εξετάζουμε πρώτα αν στην περιοχή της τεμπλέτας $\mathcal{S}^{-1}(\mathbf{x})$ θα έπρεπε να εμπιστευθούμε την πρόβλεψη του AAM και ύστερα χρησιμοποιούμε την πιθανοφάνεια για να καθορίσουμε αν η υπόθεση j μοντελοποιεί καλύτερα τις παρατηρήσεις στην περιοχή \mathbf{x} . Επίσης, αποκλίνοντας από τον αρχικό κανόνα εξέλιξης

Region Competition εισάγουμε και πληροφορία σχετικά με ακμές κατά την αναφορά [178], ωθώντας τα σύνορα των τμημάτων ώστε να συμπίπτουν με έντονες μεταβολές στην φωτεινότητα.

Επίσης ο MDL-τύπου όρος στο μοντέλο του φόντου που είχε εισαχθεί για την υλοποίηση του βήματος E σε επίπεδο τμημάτων μεταφράζεται σε έναν όρο τύπου balloon που διαστέλλει το όρια του αντικειμένου, ενώ όπως στη αναφορά [178] ένας όρος ακμών εισάγεται για να ωθήσει τα όρια των τμημάτων προς περιοχές έντονης μεταβολής στη φωτεινότητα. Αυτές οι αλλαγές δίνουν την τελική σχέσεις εξέλιξης:

$$\frac{\partial \Gamma_k}{\partial t} = -\mu_k \mathcal{N} + \left[\log \frac{P_k(I|\mathbf{s}, \mathbf{t}, \mathcal{O}) P_k(\mathcal{S}^{-1}(\mathbf{x}, \mathbf{s})|\mathcal{O})}{P_j(I|\mathcal{A}_j)} + B \right] \mathcal{N} + \nabla G(|\nabla I|) \mathcal{N}$$

όπου B είναι η δύναμη balloon, και $G(|\nabla I|)$ είναι μία φθίνουσα συνάρτηση της δύναμης ακμών, $|\nabla I|$. Η εξέλιξη των καμπυλών έχει υλοποιηθεί χρησιμοποιώντας τεχνικές επιπεδοσυνόλων, οι οποίες καθώς επιτρέπουν το σχηματισμό οπών στο εσωτερικό των τμημάτων μπορούν να χειριστούν θέματα όπως επικαλύψεις.

Δύο ανταγωνιστικές υποθέσεις εισάγονται για το φόντο που τυπικά δημιουργούν δύο μεγάλες συστάδες για φωτεινές και σκοτεινές περιοχές: η αρχικοποίηση των ορίων και των παραμέτρων των κατανομών είναι τυχαία, πέρα από το μοντέλο AAM που χρησιμοποιεί τα αποτελέσματα του κάτωθεν συστήματος ανίχνευσης για την αρχικοποίηση της θέσης του.

Στο Σχ. 5.6 συγκρίνουμε τα αποτελέσματα άνωθεν κατάτμησης που μας παρέχουν οι δύο προσεγγίσεις: Σε όλα τα παραδείγματα τα σύνορα των αντικειμένων εντοπίζονται με ακρίβεια, ενώ οι όποιες μερικές αποτυχίες αποδίδονται κυρίως στις περιορισμένες δυνατότητες σύνθεσης του AAM. Αν και τα αποτελέσματα είναι καλύτερα από αυτά της προηγούμενης μεθόδου, δεν χρησιμοποιείται εξέλιξη καμπυλών για τα επόμενα πειράματα καθώς αφενός έχει μεγαλύτερες υπολογιστικές απαιτήσεις σε σχέση με την προηγούμενη προσέγγιση, και περιπλέκεται η ανάλυση μίας εικόνας όταν υπάρχουν πολλές ψευδείς ανιχνεύσεις.

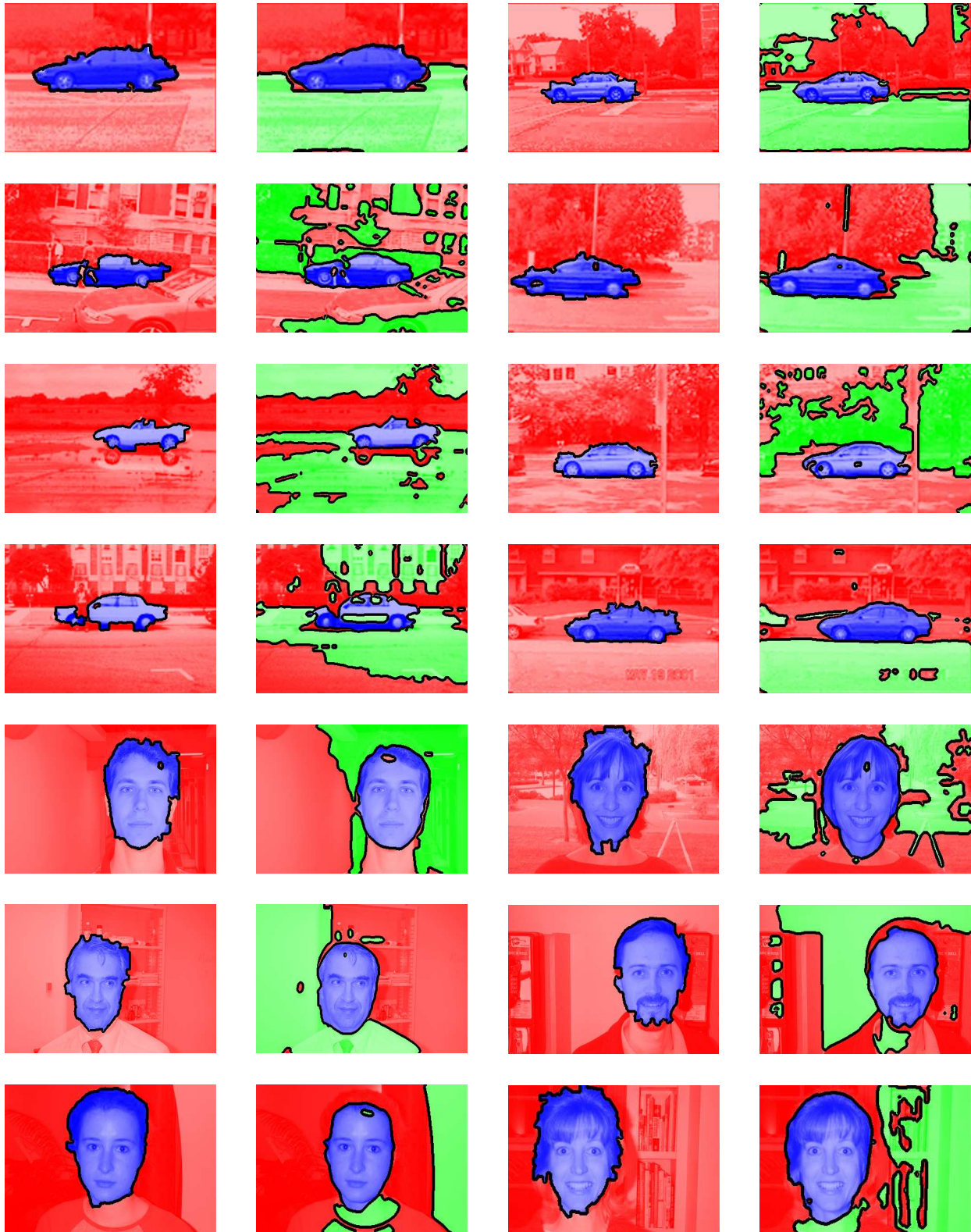
5.3 Εκτίμηση Παραμέτρων: Βήμα M

Στο βήμα M η πιθανοφάνεια των παρατηρήσεων που έχουν ανατεθεί στις παρατηρήσεις μεγιστοποιείται ως προς τις παραμέτρους των μοντέλων. Όπως δείχνουμε στην ενότητα αυτή, μεγιστοποιώντας το κριτήριο του EM προκύπτουν νέες σχέσεις ανανέωσης των AAM. Μπορούν έτσι να αντιμετωπιστούν προβλήματα που ανακύπτουν κατά την εφαρμογή μοντέλων τύπου AAM στο πρόβλημα της κατάτμησης, όπως ο χειρισμός των επικαλύψεων με άλλα αντικείμενα και ο εντοπισμός των ορίων του αντικειμένου.

5.3.1 Σχέσεις προσαρμογής παραμέτρων Σχήματος

Θεωρώντας για απλότητα μόνο τις υποθέσεις του αντικειμένου και του φόντου, το κριτήριο του EM γράφεται:

$$\sum_{\mathbf{x}} P_{\mathcal{O}}(\mathbf{x}) \log P(I(\mathbf{x})|\mathcal{O}) + (1 - P_{\mathcal{O}}(\mathbf{x})) \log P(I(\mathbf{x})|\mathcal{B}) \quad (5.18)$$



Τμηματικό Βήμα Ε Εξέλιξη Καμπυλών

Τμηματικό Βήμα Ε Εξέλιξη Καμπυλών

Σχήμα 5.6: Σύγκριση μεταξύ των δύο μεθόδων άνωθεν κατάτμησης: Η εξέλιξη καμπυλών παρέχει καλύτερα αποτελέσματα από τη μέθοδο που υλοποιεί το βήμα Ε σε επίπεδο τμημάτων, δίνοντας ομαλά περιγράμματα, που εντοπίζουν σωστά τα όρια των αντικειμένων.

όπου $P_{\mathcal{O}}(\mathbf{x})$ είναι η πιθανότητα η παρατήρηση στο σημείο \mathbf{x} να ανατίθεται στο αντικείμενο. Παρατηρούμε στο σημείο αυτό ότι η εκτίμηση της παραμόρφωσης του αντικειμένου κατά το βήμα M συνδέεται και με την ανάθεση παρατηρήσεων στο αντικείμενο, που έως τώρα την είχαμε διατυπώσει ως το βήμα E. Για παράδειγμα, μία παρατήρηση της εικόνας που είχε ανατεθεί στο αντικείμενο βάσει της προηγούμενης εκτίμησης της παραμόρφωσης μπορεί εν γένει να μην ανήκει στο χωρίο του για ένα νέο διάνυσμα παραμέτρων, οπότε η πιθανότητα να ανατεθεί στο αντικείμενο θα ισούται πλέον με μηδέν. Αυτή παρατήρηση οδηγεί στο να εκφράσουμε την ποσότητα $P_{\mathcal{O}}(\mathbf{x})$ ως:

$$P_{\mathcal{O}}(\mathbf{x}) = E(\mathbf{x})H(\mathcal{S}^{-1}(\mathbf{x}; \mathbf{s})) \quad (5.19)$$

όπου $E(\mathbf{x})$ είναι τα σταθερά αποτελέσματα του βήματος E ενώ το $H(\mathcal{S}^{-1}(\mathbf{x}; \mathbf{s}))$ μπορεί να μεταβάλλεται κατά το βήμα M· μπορούμε πλέον να ξαναγράψουμε το κριτήριο (5.18) ως

$$\sum_{\mathbf{x}} E(\mathbf{x})H(\mathcal{S}^{-1}(\mathbf{x}; \mathbf{s})) \log \frac{P(I(\mathbf{x})|\mathcal{O})}{P(I(\mathbf{x})|\mathcal{B})} + c \quad (5.20)$$

όπου ο όρος $c = \sum_{\mathbf{x}} \log P(I(\mathbf{x})|\mathcal{B})$ είναι μία σταθερά. Καθώς η βελτιστοποίηση αυτού του κριτηρίου υλοποιείται ευκολότερα όταν ορίζεται στο σύστημα συντεταγμένων της τεμπλέτας, μπορούμε να το γράψουμε ως:

$$\sum_{\mathbf{x}} E(\mathcal{S}(\mathbf{x}; \mathbf{s}))H(\mathbf{x})D(\mathbf{x}; \mathbf{s}) \log \frac{P(I(\mathcal{S}(\mathbf{x}; \mathbf{s}))|\mathcal{O})}{P(I(\mathcal{S}(\mathbf{x}; \mathbf{s}))|\mathcal{B})} \quad (5.21)$$

όπου τώρα το άθροισμα είναι πάνω στην τεμπλέτα και $D(\mathbf{x}; \mathbf{s})$ είναι η οριζουσα της Jacobian της απεικόνισης $\mathbf{x} \rightarrow \mathcal{S}(\mathbf{x})$ που ποσοτικοποιεί την τοπική μεγέθυνση/σμίκρυνση του πεδίου της τεμπλέτας. Απλοποιώντας το κριτήριο, θεωρούμε ότι το σφάλμα ανακατασκευής και για τις δύο υποθέσεις μοντελοποιείται ως λευκός Gaussian θόρυβος με μία κοινή διασπορά σ , οπότε έχουμε:

$$c \sum_{\mathbf{x}} E(\mathcal{S}(\mathbf{x}))H(\mathbf{x})D(\mathbf{x}; \mathbf{s}) [(I(\mathcal{S}(\mathbf{x})) - \mathcal{T}(\mathbf{x}))^2 - (I(\mathcal{S}(\mathbf{x})) - B(\mathcal{S}(\mathbf{x})))^2] \quad (5.22)$$

όπου \mathcal{T} είναι η σύνθεση βάσει του μοντέλου του αντικειμένου, B η ανακατασκευή της εικόνας χρησιμοποιώντας το μοντέλο για το φόντο και $c = \frac{1}{2\sigma^2}$ μία πολλαπλασιαστική σταθερά που αγνοείται στη συνέχεια.

Το αρχικό κριτήριο για προσαρμογή μοντέλων AAM (2.31) μπορεί να γραφεί με βάση τον παραπάνω συμβολισμό ως:

$$E(\mathbf{s}) = \sum_{\mathbf{x}} H(\mathbf{x}) (I(\mathcal{S}(\mathbf{x}; \mathbf{s})) - \mathcal{T}(\mathbf{x}))^2. \quad (5.23)$$

Συγκρίνοντας τα κριτήρια (5.22) και (5.23) παρατηρούμε τρία σημεία στα οποία το τελευταίο είναι ανακριβές:

- Η πληροφορία σχετικά με την κατάτμηση που παρέχει το πεδίο $E(\mathcal{S}(\mathbf{x}))$ αγνοείται, οδηγώντας έτσι σε προβλήματα σε σημεία με επικαλύψεις. Καθώς επιβάλλεται η εύρεση μίας αντιστοιχίας για κάθε σημείο της τεμπλέτας, το μοντέλο προσαρμόζεται σε οτιδήποτε μπορεί να επικαλύπτει το αντικείμενο.

- Η πιστότητα του μοντέλου του φόντου δεν συγκρίνεται με αυτήν του αντικειμένου, επιτρέποντας στη σύνθεση του αντικειμένου να περιοριστεί στο εσωτερικό ή να ξεπεράσει τα όρια του πραγματικού αντικειμένου, αν δεν υπάρχουν έντονες ακμές. Αντίθετα, αν υπάρχει ένα μοντέλο για το φόντο μπορεί να πιέσει προς τα πίσω/μπροστά το αντικείμενο και να εντοπίσει ορθά τα όριά του στην εικόνα, όπως φαίνεται στο Σχ. 5.7.
- Η μεγέθυνση/σμίκρυνση των σημείων της τεμπλέτας εξαιτίας της παραμόρφωσης αγνοείται· ουσιαστικά το κριτήριο (5.23) απαιτεί οι παρατηρήσεις να ταιριάζουν με το μοντέλο, αντί να το χρησιμοποιεί για να ερμηνεύσει τις παρατηρήσεις, όπως θα έπρεπε και όπως κάνει το κριτήριο του EM.

Στη συνέχεια δίνονται οι σχέσεις ανανέωσης των μοντέλων AAM που προκύπτουν μεταβαίνοντας από την (5.23) στη (5.22), εισάγοντας σταδιακά περιπλοκότερη πληροφορία· για να διατηρηθεί η συνέχεια, οι αποδείξεις περιέχονται στο παράρτημα 5.A'. Όπως και στην αναφορά [157] θεωρούμε τη μέθοδο Newton Raphson που χρησιμοποιεί μία προσέγγιση δεύτερης τάξης στην συνάρτηση κόστους γύρω από τις τρέχουσες παραμέτρους για να υπολογίσει την βέλτιστη ανανέωση των παραμέτρων. Μεταβάλλοντας τις παραμέτρους κατά Δs έχουμε:

$$I(\mathcal{S}(\mathbf{x}; \mathbf{s} + \Delta \mathbf{s})) \simeq I(\mathcal{S}(\mathbf{x}; \mathbf{s})) + \sum_{i=1}^{N_S} \frac{dI}{ds_i}(\mathbf{x}; \mathbf{s}) \Delta s_i \quad (5.24)$$

$$\frac{dI}{ds_i}(\mathbf{x}; \mathbf{s}) = \frac{\partial I(\mathcal{S}(\mathbf{x}; \mathbf{s}))}{\partial x} \frac{\partial \mathcal{S}_x}{\partial s_i} + \frac{\partial I(\mathcal{S}(\mathbf{x}; \mathbf{s}))}{\partial y} \frac{\partial \mathcal{S}_y}{\partial s_i} \quad (5.25)$$

όπου N_S είναι το πλήθος των στοιχείων της βάσης που χρησιμοποιούμε για να εκφράσουμε την χωρική παραμόρφωση του αντικειμένου. Θεωρώντας την ποσότητα I ως μία $R^{N_S} \rightarrow R^N$ απεικόνιση από τον χώρο των παραμέτρων στο πεδίο της εικόνας, συμβολίζουμε με $\frac{dI}{ds}$ την Jacobian της απεικόνισης αυτής. Για διακριτά δεδομένα τα I και $\frac{dI}{ds}$ μπορούμε να τα χειριστούμε ως $N \times 1$ και $N \times N_S$ πίνακες αντίστοιχα και να γράψουμε συνοπτικά την παραπάνω σχέση ως $\mathbf{I}_{s+\Delta s} = \mathbf{I}_s + \frac{dI}{ds} \Delta \mathbf{s}$, όπου $\Delta \mathbf{s}$ είναι το $N_S \times 1$ διάνυσμα μεταβολής.

Χρησιμοποιώντας αυτήν τη γραμμικοποίηση γύρω από το \mathbf{s} και εισάγοντας συμβολισμό πινάκων η τιμή της συνάρτησης κόστους (5.23) για τη νέα τιμή των παραμέτρων μπορεί να προσεγγιστεί ως:

$$C(\mathbf{s} + \Delta \mathbf{s}) = \sum_{\mathbf{x}} H(\mathbf{x}) (I(\mathbf{x}; \mathbf{s} + \Delta \mathbf{s}) - \mathcal{T}(\mathbf{x}))^2$$

$$\simeq \left[\mathbf{H} \odot \left(\mathbf{I}_s + \frac{dI}{ds} \Delta \mathbf{s} - \mathcal{T} \right) \right]^T \left[\mathbf{I}_s + \frac{dI}{ds} \Delta \mathbf{s} - \mathcal{T} \right] \quad (5.26)$$

$$= [\mathbf{H} \odot \boldsymbol{\mathcal{E}}]^T \boldsymbol{\mathcal{E}} + 2 [\mathbf{H} \odot \boldsymbol{\mathcal{E}}]^T \frac{dI}{ds} \Delta \mathbf{s} + \Delta \mathbf{s}^T \left[\mathbf{H} \odot \frac{dI}{ds} \right]^T \frac{dI}{ds} \Delta \mathbf{s} \quad (5.27)$$

όπου $\boldsymbol{\mathcal{E}} = \mathbf{I}_s - \mathcal{T}$ και με \odot συμβολίζουμε τον πολλαπλασιασμό πινάκων κατά στοιχείο. Για να απλοποιηθεί ο συμβολισμός, για όρους όπως ο $\mathbf{H} \odot \frac{dI}{ds}$ όπου η ποσότητα \mathbf{H} είναι $N \times 1$ και $\frac{dI}{ds}$ είναι $N \times N_S$, εννοείται ότι ο πίνακας \mathbf{H} πολλαπλασιάζεται προηγουμένως με το $1 \times N_S$ μοναδιαίο διάνυσμα.

Η σχέση (5.27) μας δίνει την τετραγωνική προσέγγιση στη συνάρτηση κόστους

$$C(\mathbf{s} + \Delta\mathbf{s}) = C(\mathbf{s}) + \mathcal{J}\Delta\mathbf{s} + \Delta\mathbf{s}^T \mathcal{H}\Delta\mathbf{s}$$

$$\mathcal{J} = 2 [\mathbf{H} \odot \boldsymbol{\mathcal{E}}]^T \frac{d\mathbf{I}}{ds}, \quad \mathcal{H} = \left[\left(\mathbf{H} \odot \frac{d\mathbf{I}}{ds} \right)^T \frac{d\mathbf{I}}{ds} \right] \quad (5.28)$$

με τις \mathcal{H} και \mathcal{J} να είναι οι Hessian και Jacobian μήτρες της συνάρτησης κόστους. Η βέλτιστη ανανέωση του διανύσματος παραμέτρων $\Delta\mathbf{s}^*$ δίνεται πλέον από τη σχέση

$$\Delta\mathbf{s}^* = -\frac{1}{2} [\mathcal{J}\mathcal{H}^{-1}]^T \quad (5.29)$$

που αποτελεί τον ευθύ προσθετικό (forwards additive) κανόνα για προσαρμογή μοντέλων AAM σύμφωνα με την ορολογία των [157].

Εισάγοντας τις πιθανότητες ανάθεσης παρατηρήσεων παίρνουμε ένα διαφορετικό κριτήριο:

$$\sum_{\mathbf{x}} E(\mathcal{S}(\mathbf{x}))H(\mathbf{x}) (I(\mathcal{S}(\mathbf{x})) - \mathcal{T}(\mathbf{x}))^2 = [(\mathbf{E} \odot \mathbf{H}) \odot \boldsymbol{\mathcal{E}}]^T [\boldsymbol{\mathcal{E}}] \quad (5.30)$$

για το οποίο οι αντίστοιχοι πίνακες μπορούν να γραφούν ως:

$$\mathcal{J} = \left[(\mathbf{H}' \odot \boldsymbol{\mathcal{E}})^T \frac{d\mathbf{I}}{ds} + \left(\frac{d\mathbf{E}}{ds} \odot \mathbf{H} \odot \boldsymbol{\mathcal{E}} \right)^T \boldsymbol{\mathcal{E}} \right]$$

$$\mathcal{H} = \left[\left(\mathbf{H}' \odot \frac{d\mathbf{I}}{ds} \right)^T \frac{d\mathbf{I}}{ds} + 2 \left(\mathbf{H} \odot \frac{d\mathbf{E}}{ds} \odot \boldsymbol{\mathcal{E}} \right)^T \left(\frac{d\mathbf{I}}{ds} \Delta\mathbf{s} \right) \right] \quad (5.31)$$

όπου $\mathbf{H}' = \mathbf{H} \odot \mathbf{E}$. Η προφανής τροποποίηση που εισάγεται έτσι είναι ο πολλαπλασιασμός της χαρακτηριστικής συνάρτησης του χωρίου του αντικειμένου, H με τα αποτελέσματα κατάτμησης E , επιβάλλοντας έτσι στον αλγόριθμο προσαρμογής να εστιάσει στις περιοχές του αντικειμένου που μοντελοποιούνται καλύτερα. Οι επιπλέον όροι που εμπλέκουν την Jacobian της κατάτμησης ως προς τις παραμέτρους σχήματος προκύπτουν από το γεγονός ότι μετατοπίζοντας το αντικείμενο αλλάζει και το βάρος με το οποίο ανατίθενται περιοχές της εικόνας σε συγκεκριμένες περιοχές του.

Χρησιμοποιώντας τον όρο $\mathcal{E}_B = \mathbf{I}_s - \mathbf{B}_s$ μπορούμε να γράψουμε για το κριτήριο που προκύπτει εισάγοντας το σφάλμα του μοντέλου του φόντου:

$$C(\mathbf{s}) = [\mathbf{H} \odot \mathbf{E}_s \odot (\boldsymbol{\mathcal{E}})]^T [\boldsymbol{\mathcal{E}}] - [\mathbf{H} \odot \mathbf{E}_s \odot (\boldsymbol{\mathcal{E}}_B)]^T [\boldsymbol{\mathcal{E}}_B] \quad (5.32)$$

από το οποίο παίρνουμε τους πίνακες

$$\mathcal{J} = \mathcal{J}_{\mathcal{E}} - \mathcal{J}_{\mathcal{E}_B}, \quad \mathcal{H} = \mathcal{H}_{\mathcal{E}} - \mathcal{H}_{\mathcal{E}_B} \quad (5.33)$$

Παραπάνω οι $\mathcal{J}_{\mathcal{E}}, \mathcal{H}_{\mathcal{E}}$ αντιστοιχούν στους πίνακες της (5.31) και $\mathcal{J}_{\mathcal{E}_B}, \mathcal{H}_{\mathcal{E}_B}$ είναι οι αντίστοιχοι πίνακες που προκύπτουν εισάγοντας τους όρους από το μοντέλο του φόντου. Καθώς ο όρος ο οποίος ελαχιστοποιείται δεν είναι πλέον μία κυρτή συνάρτηση μπορούν να προκύψουν αστάθειες· ένα μικρότερο βήμα συνεπώς επιλέγεται όποτε προτείνονται μεγάλες ανανεώσεις, εξασφαλίζοντας ότι οι νέες τιμές του διανύσματος

δεν θα ξεπεράσουν το τριπλάσιο της διασποράς τους στο σύνολο εκπαίδευσης.

Τέλος, εισάγοντας την κλιμάκωση της παραμόρφωσης παίρνουμε το αρχικό κριτήριο του EM, όπου η έκφραση για την ανανέωση λαμβάνει υπόψη όλους τους παραπάνω όρους, καθώς και την επιρροή της σμίκρυνσης/μεγέθυνσης του αντικειμένου στον όρο πιθανοφάνειας. Οι σχέσεις ανανέωσης δίνονται στο παράρτημα 5.Α'.

5.3.2 Εξισώσεις προσαρμογής εμφάνισης

Οι παράμετροι εμφάνισης μπορούν να ανανεωθούν χρησιμοποιώντας λιγότερο περιπλοκές σχέσεις, καθώς μεταβάλλοντάς τες δεν επηρεάζεται η κατάτμηση ή η παραμόρφωση. Στο σύστημα συντεταγμένων της τεμπλέτας το κριτήριο του EM μπορεί να γραφεί ως:

$$C(\mathbf{t}) = \sum_{\mathbf{x}} W(\mathbf{x}) \left[(I(\mathcal{S}(\mathbf{x})) - T_0(\mathbf{x})) - \sum_{i=1}^{N_T} \mathbf{t}_i T_i(\mathbf{x}) \right]^2$$

$$W(\mathbf{x}) = D(\mathbf{x}; \mathbf{s}) H(\mathbf{x}) E(\mathcal{S}(\mathbf{x})) \quad (5.34)$$

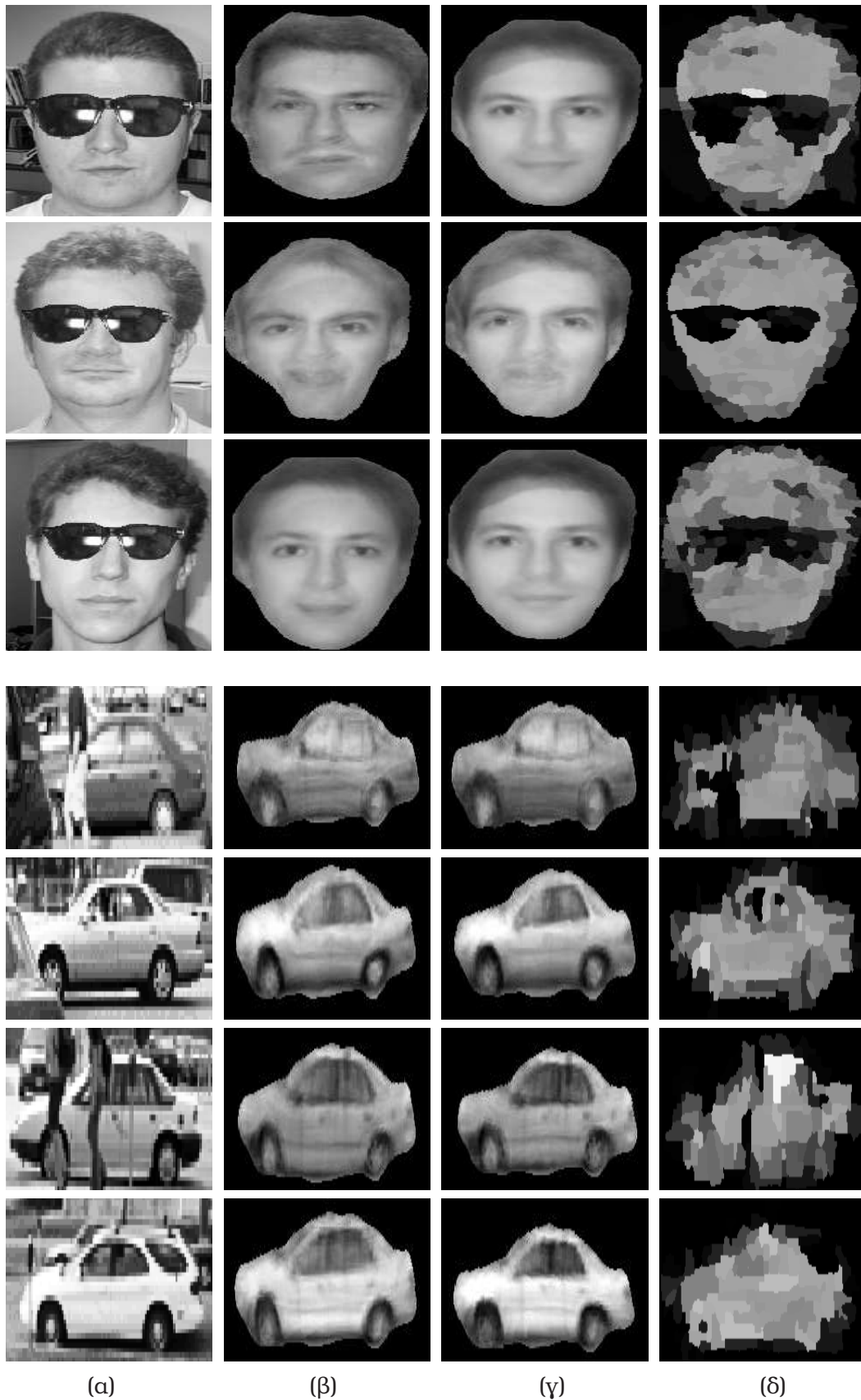
όπου N_T είναι το πλήθος των στοιχείων της βάσης εμφάνισης και ο όρος $W(\mathbf{x})$ συνδυάζει όλους τους παράγοντες της παραμόρφωσης. Αυτό μας δίνει τη λύση των σταθμισμένων ελάχιστων τετραγώνων:

$$\mathbf{t}^* = [\mathbf{W} \odot \mathcal{E}] [\mathbf{T}^T (\mathbf{W} \odot \mathbf{T})]^{-1} \quad (5.35)$$

όπου $\mathcal{E}(\mathbf{x}) = I(\mathcal{S}(\mathbf{x})) - T_0(\mathbf{x})$ και \mathbf{T} είναι ο $N \times N_T$ πίνακας που σχηματίζεται ενώνοντας τα στοιχεία της βάσης για την εμφάνιση.

Σημειώνουμε εδώ ότι στις σχέσεις ανανέωσης των παραμέτρων εμφάνισης και σχήματος μπορεί να εισαχθεί άμεσα η πιθανοφάνειά τους βάσει της πρότερης κατανομής τους, αν αυτή μοντελοποιείται μέσω μίας Gaussian κατανομής. Οι πίνακες που εμπλέκονται στις σχέσεις ανανέωσης τροποποιούνται θέτοντας $\mathcal{J}_i = \mathcal{J}_i - \lambda p_i / \sigma_i^2$, $\mathcal{H}_{i,i} = \mathcal{H}_{i,i} + \lambda \sigma_i^2$, όπου p_i είναι η προηγούμενη εκτίμηση της παραμέτρου p_i , σ_i η απόκλιση του στο σύνολο εκπαίδευσης και μέσω του λ ελέγχεται ο συμβιβασμός μεταξύ πρότερης γνώσης και πιστότητας στα δεδομένα. Στην υλοποίησή μας χρησιμοποιήσαμε τις τιμές $\lambda = .1$ για την εμφάνιση και $\lambda = 1$ για το σχήμα που έδωσαν ικανοποιητικά αποτελέσματα και για τις δύο κατηγορίες αντικειμένων.

Κάποια αποτελέσματα που δείχνουν τις βελτιώσεις που εισάγονται από τα εναλλακτικά κριτήρια προσαρμογής φαίνονται στο Σχ. 5.7. Για την πάνω γραμμή τα αποτελέσματα είναι σε επίπεδο λεπτομερειών, εισάγοντας μία συστηματική, αλλά όχι κριτικής σημασίας βελτίωση. Για τις τελευταίες δύο, όπου η κατάτμηση είναι απαραίτητη, η χρήση των αποτελεσμάτων του βήματος E βοηθάει να διατηρηθεί η ποιότητα της προσαρμογής του AAM, ενώ οι υπόλοιποι όροι εισάγουν πάλι μικρές, αλλά αισθητές βελτιώσεις. Αν και οι τελευταίοι δύο όροι προκύπτουν από την προσέγγισή μας, θα μπορούσε να θεωρηθεί ότι οι βελτιώσεις που εισάγουν (στήλη δ) δεν δικαιολογούν το επιπλέον υπολογιστικό κόστος που απαιτούν. Όμως, συγκρίνοντας τις εικόνες στην δεύτερη και τρίτη στήλη, φαίνεται ότι η πληροφορία κατάτμησης είναι απαραίτητη, ενώ και τα πειραματικά αποτελέσματα της επόμενης ενότητας καταδεικνύουν μία σταθερή βελτίωση στις επιδόσεις ανίχνευσης.



Σχήμα 5.7: Διαφορές στην προσαρμογή χρησιμοποιώντας εναλλακτικά κριτήρια AAM: (α) Εικόνα προς ανάλυση, (β) ανακατασκευή μέσω του απλού κριτηρίου ελάχιστων τετραγώνων, (γ) εισάγοντας πληροφορία κατάτμησης μέσω του EM, (δ) τα αποτελέσματα του βήματος E. Εισάγοντας τον EM καθίσταται εύρωστη η προσαρμογή των AAM καθώς αγνοούνται περιοχές με μη μοντελοποιούμενη εμφάνιση, όπως το τρίτο παράθυρο ή τα μαλλιά στις κάτω γραμμές.

5.4 Συνεργετική Ανίχνευση Αντικειμένων

Το ότι η άνωθεν γνώση μπορεί να βοηθήσει τη διαδικασία της κατάτμησης είναι ως ένα βαθμό σαφές, τόσο από προηγούμενες εργασίες όσο και από τα αποτελέσματα που παρουσιάστηκαν στις προηγούμενες ενότητες. Ο στόχος μας σε αυτήν την ενότητα είναι να εξετάσουμε την άλλη όψη του προβλήματος, δηλαδή κατά πόσο η συνεργασία μεταξύ της κατάτμησης και της αναγνώρισης μπορεί να βοηθήσει την αναγνώριση. Για τον λόγο αυτό χρησιμοποιήσαμε αρχικά καθαρά δεδομένα, με τα χαρακτηριστικά σημεία του αντικειμένου να έχουν εντοπιστεί χειρωνακτικά, ώστε να αποφευχθεί η σύγχυση μεταξύ διαφορετικών προβλημάτων, όπως η κατασκευή του μοντέλου και η ενέργεια των δύο διαδικασιών. Στη συνέχεια, θέτοντας ως στόχο την επέκταση της εφαρμοσιμότητας της προσέγγισής μας σε περισσότερες κατηγορίες αντικειμένων, χρησιμοποιήθηκε ένα σύστημα κάτωθεν ανίχνευσης αντικειμένων βάσει σημείων ενδιαφέροντος σε συνδυασμό με έναν νέο αλγόριθμο για την αυτοματοποιημένη εκμάθηση μοντέλων AAM χωρίς χαρακτηριστικά σημεία. Χρησιμοποιώντας τις ίδιες ρυθμίσεις για δύο κατηγορίες αντικειμένων, καταλήξαμε στα ίδια ποιοτικά αποτελέσματα σχετικά με την βελτιωμένη επίδοση που επιτυγχάνεται βάσει του αλγορίθμου EM.

5.4.1 Πλήρως Annotated Δεδομένα

Όπως και στις αναφορές [218, 144], θεωρούμε ότι το κάτωθεν σύστημα ανίχνευσης αντικειμένων προτείνει τις ενδιαφέρουσες τοποθεσίες της εικόνας από τις οποίες ξεκινάει η εφαρμογή των αναγεννητικών μοντέλων. Για να περιορίσουμε το πλήθος των υποθέσεων που εξετάζονται χρησιμοποιούμε ένα κατώφλι για τις τιμές του κάτωθεν συστήματος ανίχνευσης, το οποίο τίθεται ίσο με τη μέγιστη τιμή για την οποία ανιχνεύονται όλα τα υπάρχοντα αντικείμενα. Έτσι εξετάζονται μαζί με τις σωστές τοποθεσίες και πλήθος από λάθος υποθέσεις, οπότε ο ρόλος της ενέργειας είναι πλέον να ακυρώσει τις εσφαλμένες θετικές ανιχνεύσεις (false positives).

Η άνωθεν πληροφορία μπορεί να εξαχθεί τόσο από τις παραμέτρους των AAM όσο και από την κατάτμηση του αντικειμένου. Θα μπορούσε κανείς να ισχυριστεί ότι θα μπορούσαμε να πάρουμε την άνωθεν πληροφορία που εισάγεται από ένα AAM χωρίς τον EM, δηλαδή με απλή προσαρμογή AAM, οπότε πλέον δεν θα ήταν απαραίτητος ο από κοινού χειρισμός των δύο διαδικασιών. Όπως δείχνουμε αργότερα, αυτός ο ισχυρισμός εξασθενεί όταν τα αντικείμενα είναι μερικώς επικαλυπτόμενα, οπότε οι παράμετροι που προκύπτουν με απλή προσαρμογή AAM μειώνουν αισθητά τις επιδόσεις ανίχνευσης. Προς το παρόν όμως θεωρούμε ότι ο ισχυρισμός αληθεύει, και εξετάζουμε την πληροφορία που εισάγουν τα αποτελέσματα της κατάτμησης όχι πλέον στο πλαίσιο της ενέργειας αλλά θεωρούμενα ως μία μέτρηση που χρησιμοποιείται για την ανίχνευση αντικειμένων.

Έχουμε τρεις διαφορετικές πηγές πληροφορίας για την ανίχνευση αντικειμένων: Πρώτον, τον όρο πιθανοφάνειας που μας παρέχει το σύστημα που χρησιμοποιείται για κάτωθεν ανίχνευση C_{BU} . Δεύτερον, τις παραμέτρους εμφάνισης και σχήματος των μοντέλων AAM, τις οποίες χειριζόμαστε ως χαρακτηριστικά: Μοντελοποιώντας την κατανομή τους μέσω Gaussian συναρτήσεων για εικόνες που περιέχουν το αντικείμενο καθώς και εικόνες που περιέχουν μόνο το φόντο κατασκευάζουμε τον όρο

ταξινόμησης:

$$C_{AAM} = \log \frac{P(\mathbf{s}, \mathbf{t}|\mathcal{O})}{P(\mathbf{s}, \mathbf{t}|\mathcal{B})} \quad (5.36)$$

Τέλος, τα αποτελέσματα της άνωθεν κατάτμησης συνοψίζονται στο ποσοστό του χωρίου του αντικειμένου το οποίο τελικά του ανατίθεται κατά το βήμα Ε· η ποσότητα αυτή εκτιμάται παίρνοντας τον μέσο των αποτελεσμάτων του βήματος Ε πάνω στην περιοχή της εικόνας που αντιστοιχίζεται στο αντικείμενο:

$$C_{EM} = \frac{\sum_{\mathbf{x} \in \mathcal{R}} H(\mathcal{S}^{-1}(\mathbf{x}))E(\mathbf{x})}{\sum_{\mathbf{x} \in \mathcal{R}} H(\mathcal{S}^{-1}(\mathbf{x}))}, \quad (5.37)$$

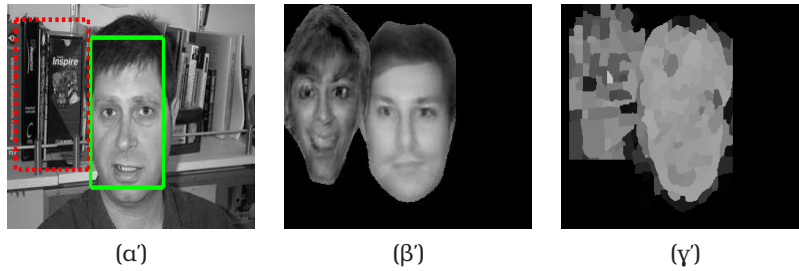
Συνδυάζουμε την πληροφορία που παρέχουν αυτές οι τρεις πηγές στο πλαίσιο των supra-Bayesian τεχνικών συνδυασμού πληροφορίας [108] που είδαμε και στο προηγούμενο κεφάλαιο. Και εδώ η έξοδος C_k του ταξινομητή k θεωρείται ως μία τυχαία μεταβλητή που ακολουθεί της κατανομής $P(C_k|\mathcal{O})$ και $P(C_k|\mathcal{B})$ υπό τις υποθέσεις του φόντου και του αντικειμένου, αντίστοιχα. Θεωρώντας το σύνολο των εξόδων των ταξινομητών ως ένα διάνυσμα από ανεξάρτητες τυχαίες μεταβλητές, $\mathbf{C} = (C_1, \dots, C_k)$ μπορούμε να χρησιμοποιήσουμε τις μεμονωμένες κατανομές τους για τον συνδυασμό των ταξινομητών:

$$\frac{P(\mathcal{O}|\mathbf{C})}{P(\mathcal{B}|\mathbf{C})} = c \frac{P(\mathbf{C}|\mathcal{O})}{P(\mathbf{C}|\mathcal{B})} = c \prod_{k=1}^K \frac{P(C_k|\mathcal{O})}{P(C_k|\mathcal{B})} \quad (5.38)$$

όπου $c = P(\mathcal{O})/P(\mathcal{B})$. Για απλότητα δεν λαμβάνουμε υπόψη τις αλληλεξαρτήσεις μεταξύ των εξόδων των ταξινομητών, αν και αυτό θα μπορούσε να εισάγει κάποια βελτίωση στις επιδόσεις.

Για την αποτίμηση της απόδοσης των διάφορων ταξινομητών χρησιμοποιούμε τόσο καμπύλες τύπου χαρακτηρισμού λειτουργίας δέκτη (Receiver Operating Characteristic- ROC), όπου τα αποτελέσματα ανίχνευσης συνοψίζονται στην απόφαση για το εάν η εικόνα περιέχει ένα αντικείμενο, και τις καμπύλες Ακρίβειας-Ανάκλησης (Precision Recall-PR) όπου το ποσοστό των ανιχνευθέντων αντικειμένων (Ανάκληση) συγκρίνεται με το ποσοστό των σωστών ανιχνεύσεων (Ακρίβεια). Οι καμπύλες ROC μπορούν να συνοψιστούν με το Ποσοστό Ίσου Σφάλματος (Equal Error Rate - EER), που ισούται με το σημείο όπου το ποσοστό ψευδών ανιχνεύσεων ισούται με το ποσοστό εσφαλμένων αρνητικών, ενώ οι καμπύλες PR από τη μέγιστη F -μέτρησή τους, που ορίζεται ως $F = \frac{2}{\frac{1}{R} + \frac{1}{P}}$. Σημειώνουμε εδώ ότι οι καμπύλες ROC συχνά είναι παραπλανητικές, καθώς ένα σύστημα ανίχνευσης που δίνει πολλαπλές ψευδείς ανιχνεύσεις θα δώσει ευκολότερα μία θετική απάντηση για μία μεγάλη εικόνα. Συνεπώς, επειδή στη βάση των [62] οι εικόνες που περιέχουν αντικείμενα είναι μεγαλύτερες από αυτές του φόντου, φαίνεται να λειτουργεί καλύτερα ένα σύστημα με χαμηλή ακρίβεια. Για το λόγο αυτό περιλαμβάνουμε τις καμπύλες ROC μόνο για σύγκριση με άλλες αναφορές, και θεωρούμε τις καμπύλες PR ως ενδεικτικές των επιδόσεων.

Όπως φαίνεται από τις τελευταίες, τα μεμονωμένα συστήματα ανίχνευσης συμπεριφέρονται με έναν συμπληρωματικό τρόπο σε διαφορετικές περιοχές, ενώ τα αποτελέσματα του συνδυασμού τους είναι συστηματικά καλύτερα. Σημειώνουμε συγκεκριμένα ότι τα αποτελέσματα κατάτμησης που παίρνουμε συνδυάζοντας το άνωθεν



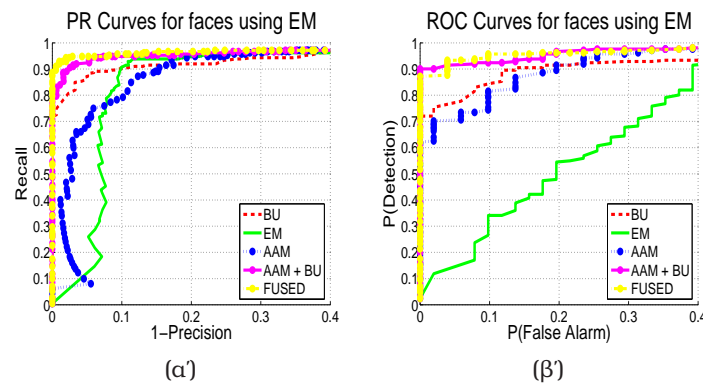
Σχήμα 5.8: Πάνω γραμμή: Η άνωθεν πληροφορία μπορεί να βελτιώσει τις επιδόσεις ανίχνευσης περιορίζοντας τις εσφαλμένες θετικές αποφάσεις (false positives): η περίπλοκη δομή στο φόντο, προκαλεί μία εσφαλμένη πρόταση για τη θέση του αντικειμένου, που φαίνεται με την κόκκινη διακεκομμένη γραμμή στην εικόνα (α). Αυτή η υπόθεση απορρίπτεται τόσο λόγω της χαμηλής πιθανοφάνειας των παραμέτρων, που οδηγούν στη σύνθεση μίας περίεργης μορφής στην εικόνα (β) όσο και στα χαμηλότερα αποτελέσματα του βήματος E, που φαίνονται με μία χαμηλότερη φωτεινότητα στην εικόνα (γ). Κάτω γραμμές: αποτελέσματα άνωθεν/κάτωθεν ανίχνευσης βάσει αυτής της προσέγγισης. Οι κόκκινες διακεκομμένες γραμμές δείχνουν υποθέσεις που απορρίφθηκαν από το άνωθεν σύστημα.

σύστημα ανίχνευσης μόνο με τα αποτελέσματα του AAM βελτιώνονται όταν εισάγεται πληροφορία κατάτμησης, αποσαφηνίζοντας ότι αυτή είναι μία συμπληρωματική πηγή πληροφορίας.

Ανίχνευση Καλυμμένων Αντικειμένων

Ένα βασικό επιχείρημα εναντίον των καθολικών αναγεννητικών μοντέλων είναι ότι δεν μπορούν να χειριστούν το πρόβλημα της επικάλυψης και έτσι αποτυγχάνουν σε περιπτώσεις που μπορούν να αντιμετωπιστούν από τμηματικά μοντέλα. Το πρόβλημα αυτό όμως μπορεί να αντιμετωπιστεί εισάγοντας πληροφορία σχετική με την κατάτμηση στην ανάλυση των αντικειμένων, επιτρέποντας έτσι να αγνοούνται οι περιοχές που δεν μπορούν να παρατηρηθούν.

Καθώς στο αρχικό σύνολο δεδομένων όλα τα πρόσωπα παρατηρούνται πλήρως, το κέρδος που εισάγεται είναι δύσκολο να αποτιμηθεί. Για αυτό επαναλαμβάνουμε το προηγούμενο πείραμα ανίχνευσης, προσθέτοντας τώρα γυαλιά ηλίου στα πρόσωπα Σχ. 5.7. Για να απλοποιήσουμε την ανάλυση των αποτελεσμάτων, θεωρούμε ότι



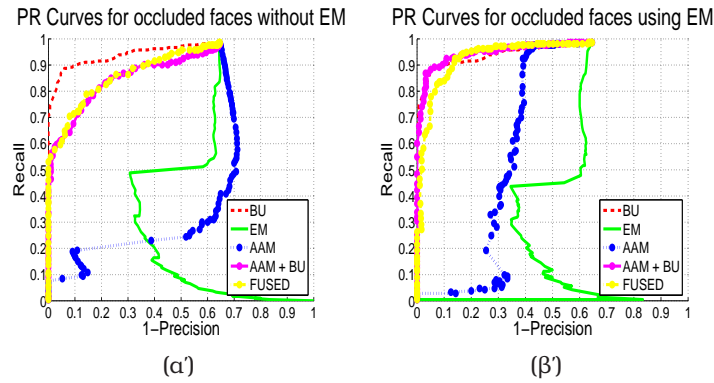
Σχήμα 5.9: Αποτελέσματα ανίχνευσης στο σύνολο δεδομένων της αναφοράς [62]: παρατηρούμε ότι ο συνδυασμός άνωθεν και κάτωθεν πληροφορίας οδηγεί σε καλύτερες επιδόσεις ανίχνευσης.

το κάτωθεν σύστημα δεν επηρεάζεται από την επικάλυψη και χρησιμοποιούμε τα αποτελέσματα που δίνει με τις πλήρως παρατηρούμενες εικόνες. Επίσης, χρησιμοποιούνται οι ταξινομητές και η μέθοδος συνδυασμού των εξόδων τους που εκμαθεύτηκαν με καθαρά δεδομένα, οπότε η επικάλυψη εμφανίζεται πλέον μόνο στο σύνολο εξέτασης.

Σε αυτό το πλαίσιο αποτιμήθηκε η επίδοση των άνωθεν ταξινομητών σε δύο εναλλακτικά σενάρια, (α) θέτοντας τα βάρη κατάτμησης $E(x)$ στην (5.22) ίσα με μονάδα παντού, αγνοώντας την πληροφορία σχετικά με την κατάτμηση και (β) θέτοντας τα $E(x)$ ίσα με τα αποτελέσματα που παρέχει ο EM. Οι εκτιμώμενες παράμετροι εισάγονται στους ταξινομητές που κατασκευάζονται με καθαρά δεδομένα για το αντίστοιχο σενάριο, ενώ η μέτρηση κατάτμησης και στις δύο περιπτώσεις εκτιμάται με βάση τις παραμέτρους του AAM μετά την σύγκλιση.

Όπως φαίνεται στο Σχ. 5.10, οι εκτιμήσεις των παραμέτρων από το σενάριο (α) δίνουν σημαντικά χειρότερες επιδόσεις, και προκαλούν την απότομη πτώση της καμπύλης Ακρίβειας-Ανάκλησης. Αντίθετα, στη δεύτερη περίπτωση οι επιδόσεις χειροτερεύουν σταδιακά, με την βασική πηγή λάθους να είναι ανακρίβειες του βήματος E που αναθέτουν μέρος των γυαλιών στο πρόσωπο. Επίσης, οι επιδόσεις του ταξινομητή που χρησιμοποιεί πληροφορία κατάτμησης είναι μόνο ελαφρά χειρότερες στο σενάριο (β) και είναι οι αναμενόμενες, δεδομένου ότι μέρος του αντικειμένου είναι κρυμμένο και συνεπώς είναι μικρότερη η έκταση της εικόνας που του ανατίθεται. Αντίθετα, στο σενάριο (α) οι λανθασμένες συνθέσεις των προσώπων δίνουν θορυβώδη αποτελέσματα και εμφανώς χειρότερες επιδόσεις. Αυτή η συμπεριφορά επηρεάζει και τα αποτελέσματα συνδυασμού ταξινομητών, όπου παρατηρείται η συστηματικά καλύτερη απόδοση που επιτυγχάνεται με την δεύτερη προσέγγιση.

Συνοψίζοντας, αυτά τα αποτελέσματα καταδεικνύουν τη σημασία του κοινού χειρισμού των προβλημάτων της ανίχνευσης και κατάτμησης αντικειμένων σε δύσκολες περιπτώσεις, όταν π.χ. πρόσωπα που επικαλύπτονται από γυαλιά, γένια ή κασκόλ. Το κέρδος στην περίπτωση αυτή δεν οφείλεται μόνο στην πληροφορία επαλήθευσης που παρέχεται από μία άνωθεν κατάτμηση, αλλά και από την εύρωστη προσαρμογή του μοντέλου που διατηρεί υψηλή την απόδοση του ταξινομητή που χρησιμοποιεί τις παραμέτρους του AAM.



Σχήμα 5.10: Καμπύλες Ακρίβειας-Ανάκλησης για την ανίχνευση προσώπων με επικάλυψη χρησιμοποιώντας (α) απλή προσαρμογή AAM και (β) τα αποτελέσματα του EM. Εδώ η επίδραση του συνδυασμού κατάτμησης και ανίχνευσης γίνεται πιο εμφανής, καθώς η χρήση του EM εισάγει ευρωστία στην προσαρμογή των AAM. Χωρίς τον EM ο ταξινομητής που χρησιμοποιεί τα χαρακτηριστικά του AAM αποτυγχάνει οδηγώντας σε χειρότερες επιδόσεις.

5.4.2 Ελάχιστα Annotated Δεδομένα

Ένα βασικό θέμα για να καταστεί η προσέγγισή μας κατάλληλη για την ανίχνευση πολλών κλάσεων αντικειμένων είναι το μέγεθος της ανθρώπινης παρέμβασης που χρειάζεται για την κατασκευή των μοντέλων. Για τον λόγο αυτό επιδιώκουμε την κατασκευή μοντέλων με ένα ελάχιστο πλήθος από χαρακτηριστικά σημεία χρησιμοποιώντας ως μόνη πληροφορία το κουτί που περιέχει το αντικείμενο (δύο σημεία).

Για την κάτωθεν ανίχνευση χρησιμοποιούμε το σύστημα που παρουσιάζεται στο επόμενο κεφάλαιο, το οποίο παρακάμπτει τη χειρωνακτική εξαγωγή χαρακτηριστικών σημείων χρησιμοποιώντας τελεστές ανίχνευσης σημείων ενδιαφέροντος· εδώ εξετάζουμε την πτυχή του προβλήματος που σχετίζεται με την αυτοματοποιημένη κατασκευή ενός AAM.

Αυτοματοποιημένη Κατασκευή AAM

Στην αναφορά [226] χρησιμοποιείται ένας αλγόριθμος τύπου bootstrapping όπου εναλλάσσει μεταξύ της προσαρμογής ενός AAM στο σύνολο εκπαίδευσης, την χρήση αλγορίθμων οπτικής ροής για την καλύτερη προσαρμογή της σύνθεσης στις παρατηρήσεις, και κατόπιν την ανανέωση του μοντέλου χρησιμοποιώντας τα συνδυασμένα αποτελέσματα από τις δύο παραμορφώσεις. Στην περίπτωση που εξετάζουμε εμείς, όπου δεν έχουμε προκατετμημένες εικόνες ενώ υπάρχουν και επικαλύψεις με άλλα αντικείμενα, ο υπολογισμός της οπτικής ροής δίνει συχνά εσφαλμένα αποτελέσματα, παρά τις προηγμένες μεθόδους που εξετάστηκαν [137]. Αφετέρου, αυτή η προσέγγιση στερείται σαφήνειας, καθώς δεν υπάρχει κάποια θεωρητική εγγύηση ότι αυτή η διαδικασία βελτιώνει την ποιότητα του μοντέλου, παρά μόνο η εμπειρική παρατήρηση ότι όλες οι παρατηρήσεις κάποτε μοντελοποιούνται επαρκώς.

Στην αναφορά [4] μία ξεκάθαρη διατύπωση του προβλήματος κατασκευής ενός AAM παρουσιάζεται, εκφράζοντάς το ως την εκτίμηση μίας βέλτιστης βάσης εμφάνισης και σχήματος για την αναπαράσταση μίας κατηγορίας αντικειμένων, ενώ προτείνεται ένας επαναληπτικός αλγόριθμος για την επίλυσή του. Στο αρχικό βήμα κάθε επανάληψης εκτιμώνται για κάθε εικόνα τα πεδία εμφάνισης και παραμόρφωσης

χρησιμοποιώντας το προηγούμενο μοντέλο AAM. Στη συνέχεια ανανεώνονται τα στοιχεία της βάσης του μοντέλου κάνοντας μία προβολή PCA στα πεδία παραμόρφωσης και τις εικόνες, μετά από την αντιστοίχισή τους με την τεμπλέτα του αντικειμένου.

Αν και για την βάση εμφάνισης η PCA όντως παρέχει τα στοιχεία για τα οποία ελαχιστοποιείται το σφάλμα ανακατασκευής των δεδομένων, δεν ισχύει το ίδιο και με την βάση σχήματος. Συγκεκριμένα παρατηρούμε ότι στην εκτίμηση της βάσης σχήματος η φωτεινότητα της εικόνας και ως εκ τούτου το κριτήριο που ελαχιστοποιείται δεν λαμβάνεται υπόψη, παρά μόνο εμμέσως, μέσω της επίδρασής τους στα μεμονωμένα πεδία παραμόρφωσης. Επίσης, αυτή η προσέγγιση είναι εφαρμόσιμη μόνο στον αλγόριθμο συνδυαστικής προσαρμογής (compositional fitting) που χρησιμοποιείται [157, 4], όπου συνδυασμοί από παραμορφώσεις δίνουν παραμορφώσεις που βρίσκονται εκτός του γραμμικού χώρου που ορίζουν τα διανύσματα βάσης. Στην περίπτωσή μας όμως, όπου χρησιμοποιείται ένα αθροιστικός αλγόριθμος προσαρμογής, όλες οι παραμορφώσεις εκφράζονται απευθείας ως γραμμικοί συνδυασμοί των συναρτήσεων βάσης· συνεπώς η επανεκτίμηση της βάσης μπορεί να δώσει πάλι μόνο γραμμικούς συνδυασμούς των αρχικών συναρτήσεων βάσης, καθιστώντας τετριμμένη την ανανέωσή της.

Για αυτό προτείνουμε την απευθείας ελαχιστοποίηση του κριτηρίου που χρησιμοποιείται από τους [4] ως προς τα στοιχεία της βάσης παραμορφώσεων στο πλαίσιο του αλγορίθμου EM-PCA [195]. Χειριζόμενοι τις συνιστώσες του αναπτύγματος ως κρυφές μεταβλητές και τα στοιχεία της βάσης ως παραμέτρους, μπορούμε πλέον να δούμε την διαδικασία προσαρμογής του AAM ως το βήμα E που ακολουθείται από το βήμα M που χρησιμοποιεί τις εκτιμηθείσες συνιστώσες για να ανανεώσει τα στοιχεία βάσης (παραμέτρους) ελαχιστοποιώντας το κριτήριο κόστους. Η βασική διαφορά με την αναφορά [195] έγκειται στο ότι το κριτήριο που ελαχιστοποιείται δεν είναι γραμμικό ως προς τα στοιχεία βάσης, οπότε χρησιμοποιούμε μία επαναληπτική διαδικασία. Συγκεκριμένα, μετά το βήμα E το κριτήριο των [4] γράφεται:

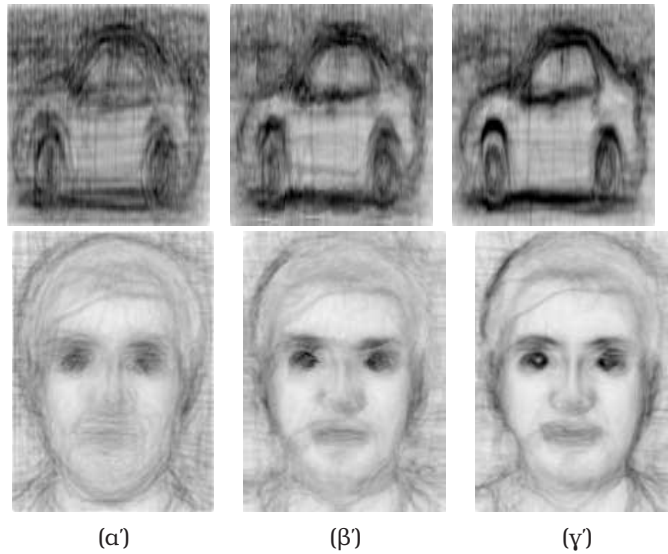
$$\sum_{k=1}^K \sum_{\mathbf{x}} [I_k(\mathcal{S}(\mathbf{x}; \mathbf{s}_k)) - \mathcal{T}_k(\mathbf{x})]^2 \quad (5.39)$$

όπου K είναι το πλήθος των χρησιμοποιούμενων εικόνων, \mathcal{T}_k είναι η -θεωρούμενη σταθερή- σύνθεση της εμφάνισης στο σύστημα συντεταγμένων της τεμπλέτας και \mathbf{s}_k είναι η τρέχουσα εκτίμηση των παραμέτρων που χρησιμοποιείται για την έκφραση της παραμόρφωσης της k -οστής εικόνας σύμφωνα με τη σχέση (2.30): $\mathcal{S}(\mathbf{x}; \mathbf{s}_k) = \sum_{i=1}^{N_S} \mathbf{s}_{k,i} S_i(\mathbf{x})$. Παίρνοντας την παράγωγο της παραπάνω έκφρασης ως προς τα οριζόντια και κάθετα στοιχεία του i -στού διανύσματος της βάσης παραμόρφωσης $S_{x,i}(\mathbf{x}), S_{y,i}(\mathbf{x})$ στο σημείο \mathbf{x} δίνει τη σχέση ανανέωσης

$$S'_{x,i}(\mathbf{x}) = S_{x,i}(\mathbf{x}) - c \sum_{k=1}^K \mathbf{s}_{k,i} \left. \frac{\partial I}{\partial x} \right|_{\mathcal{S}_k} [I_k(\mathcal{S}(\mathbf{x}; \mathbf{s}_k)) - \mathcal{T}_k(\mathbf{x})] \quad (5.40)$$

$$S'_{y,i}(\mathbf{x}) = S_{y,i}(\mathbf{x}) - c \sum_{k=1}^K \mathbf{s}_{k,i} \left. \frac{\partial I}{\partial y} \right|_{\mathcal{S}_k} [I_k(\mathcal{S}(\mathbf{x}; \mathbf{s}_k)) - \mathcal{T}_k(\mathbf{x})] \quad (5.41)$$

Παραπάνω με $\left. \frac{\partial I}{\partial x} \right|_{\mathcal{S}_k}$ συμβολίζουμε την παράγωγο του I μετά από την παραμόρφωσή του με το πεδίο $\mathcal{S}(\mathbf{x}; \mathbf{s}_k)$ και το βήμα ανανέωσης c εκτιμάται σε κάθε επανάληψη με αναζήτηση γραμμής.



Σχήμα 5.11: Μέσοι πίνακες ακμών για ένα αυξανόμενο πλήθος επαναλήψεων του αλγόριθμου εκμάθησης AAM: (α) Χωρίς αντιστοίχιση, (β) ενδιάμεσα αποτελέσματα και (γ) μοντέλο που χρησιμοποιείται στα πειράματα ανίχνευσης. Ο μέσος πίνακας ακμών αποκτάται από το μέτρο της παραγώγου των εικόνων του συνόλου εκπαίδευσης, μετά την αντιστοίχισή τους με το μοντέλο AAM. Ξεκινώντας από ένα ασαφή μέσο, σταδιακά προκύπτουν ξεκάθαρες, περίπλοκες δομές όπως το εσωτερικό των ματιών και των τροχών, ή ακμές πάνω στη μύτη και τα παράθυρα.

Μετά από κάθε επανάληψη τα νέα διανύσματα S' ορθοκανονικοποιούνται και οι συντελεστές των αναπτυγμάτων επανεκτιμώνται. Εναλλακτικά θα μπορούσαμε να εισάγουμε τον περιορισμό τα διανύσματα να είναι ορθογώνια απευθείας στο ελαχιστοποιούμενο κριτήριο, αλλά αυτό θα προκαλούσε την σύζευξη των σχέσεων ανανέωσης βάσης, που οδηγεί σε ένα δυσκολότερο πρόβλημα.

Οι σχέσεις ανανέωσης (5.40),(5.41) μπορούν να θεωρηθούν ότι εκτελούν μαζικά τον υπολογισμό οπτικής ροής που χρησιμοποιείται στην αναφορά [226]: η μεταβολή της παραμόρφωσης καθορίζεται από το γινόμενο μεταξύ των παραγώγων της εικόνας και της διαφοράς της με την πρόβλεψη του μοντέλου, όπως και στον υπολογισμό οπτικής ροής, με τη διαφορά ότι προκύπτει αθροίζοντας πάνω σε όλο το σύνολο εκπαίδευσης. Έτσι εισάγεται απευθείας πληροφορία στην εκτίμηση της βάσης παραμόρφωσης, ελαττώνοντας σταθερά το σφάλμα ανακατασκευής.

Πειραματικά παρατηρήσαμε ότι κάποια από τα προβλήματα που είχαμε αρχικά εντοπίσει χρησιμοποιώντας απευθείας οπτική ροή δεν αποφευχθήκαν· συγκεκριμένα, λεπτές παραμορφώσεις του αντικειμένου όπως οι κινήσεις των ματιών ερμηνεύονταν ως αλλαγές στην εμφάνιση και αντίστροφα. Αν και η χρήση μίας πολυδιακριτικής μεθόδου και η σταδιακή εισαγωγή ιδιοδιανυσμάτων [226] απέδωσαν κάποιες βελτιώσεις, σημαντικά προβλήματα παρέμεναν.

Για αυτό, βασιζόμενοι στα επιτυχημένα αποτελέσματα της αναφοράς [236] επιλέχτηκε να χρησιμοποιηθεί η ένταση ακμών αντί της φωτεινότητας, καθιστώντας έτσι εύρωστη την εκτίμηση της βάσης σχήματος. Έτσι, χρησιμοποιούμε το απόλυτο του τελεστή Laplacian-of-Gaussian ως ένταση ακμών ως εμφάνιση της εικόνας στη τυπική διατύπωση των AAM, ενώ πέρα από τη βάση του σχήματος, μόνο ο μέσος πίνακας ακμών ανανεώνεται σε κάθε επανάληψη. Εισάγοντας ιδιοδιανύσματα για τις ακμές

χειροτερεύουν τα αποτελέσματα, επιτρέποντας σε τμήματα ακμών να ` εξαφανίζονται, καθιστώντας έτσι την προσαρμογή ανεκτικότερη σε σφάλματα. Τέλος, σε κάθε επανάληψη, για να επιτευχθεί η εύρωστη εκτίμηση των παραμέτρων χρησιμοποιείται μία προσέγγιση coarse-to-fine συνελίσσοντας τις εικόνες με Gaussian φίλτρα σταδιακά μειούμενης διασποράς.

Μετά από τη σύγκλιση αυτής της διαδικασίας, εφαρμόζεται PCA στις παραμορφωμένες εικόνες φωτεινότητας, που βρίσκονται πλέον σε αντιστοιχία, δίνοντάς μας τη βάση εμφάνισης που χρησιμοποιείται μετά στον αλγόριθμο EM.

Πειράματα Ανίχνευσης

Ακολουθώντας την ίδια διαδικασία με τα πλήρως annotated δεδομένα, χρησιμοποιούμε και συνδυάζουμε τρεις ταξινομητές βασιζόμενους σε κάτωθεν πληροφορία κατάτμησης, τις παραμέτρους του AAM και πληροφορία περί της άνωθεν κατάτμησης, και εξετάζουμε την επίδραση που έχει η χρήση αλγορίθμου EM κατά την προσαρμογή των μοντέλων. Όπως φαίνεται από τις καμπύλες επιδόσεων στο Σχ. 5.13, και ιδιαίτερα για τα πειράματα ανίχνευσης αυτοκινήτων, η ίδια συμπεριφορά παρατηρείται: τόσο για τους μεμονωμένους άνωθεν ταξινομητές όσο και για τα συνδυασμένα αποτελέσματά τους, μία καλύτερη επίδοση επιτυγχάνεται συνδυάζοντας την κατάτμηση με τη διαδικασία προσαρμογής μοντέλων. Αξίζει επίσης να σημειωθεί ότι τα αποτελέσματα ανίχνευσης προσώπου είναι καλύτερα από αυτά του προηγούμενου πειράματος, που εώς ένα βαθμό είναι αναμενόμενο, καθώς προηγουμένως το AAM είχε κατασκευαστεί χρησιμοποιώντας εικόνες από ένα διαφορετικό σύνολο δεδομένων.

Εξάγουμε δε τα ίδια συμπεράσματα σχετικά με την χρησιμότητα του συνδυασμού του EM με την ανίχνευση αντικειμένων, ιδιαίτερα για το σύνολο δεδομένων των αυτοκινήτων, όπου άλλα αντικείμενα όπως πεζοδρόμια, θάμνοι, πεζοί, κολώνες κ.α. επικαλύπτουν μέρος των αντικειμένων που πρέπει να ανιχνευθούν.

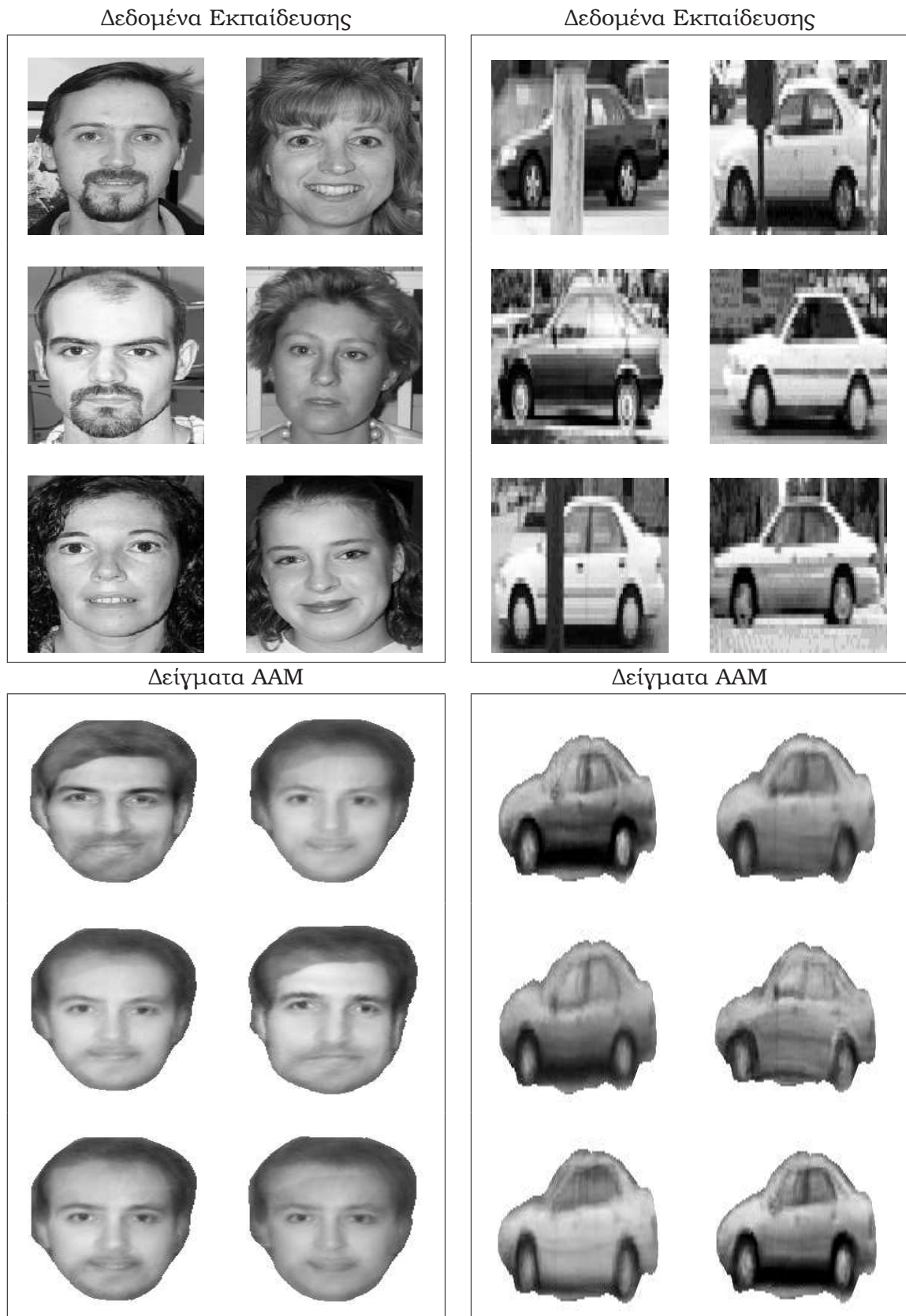
Σημειώνουμε εδώ ότι η χρήση προσώπων σταθερής κλίμακας μπορεί να έχει εισάγει μία μικρή βελτίωση στις επιδόσεις, οπότε δεν ισχυριζόμαστε ότι έχουμε υποχρεωτικά και τα καλύτερα αποτελέσματα. Το θεωρούμε πιο σημαντικό ότι παρατηρούμε μία συστηματική βελτίωση στις επιδόσεις ανίχνευσης, συνδυάζοντας άνωθεν και κάτωθεν πληροφορία μέσω του αλγορίθμου EM.

Πειραματικές ρυθμίσεις.

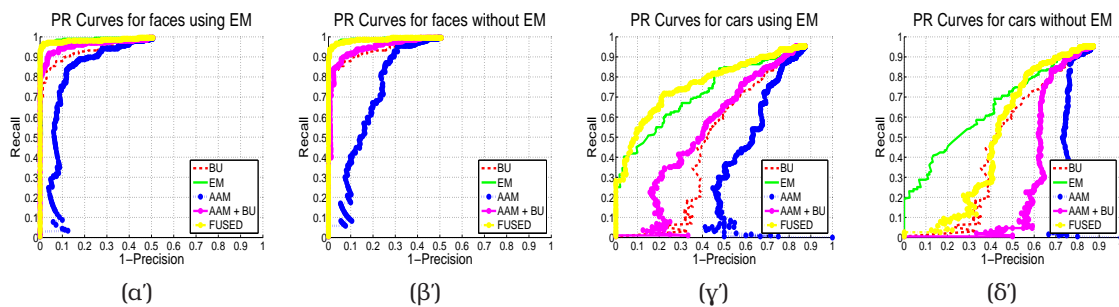
Για το πρώτο σύνολο πειραμάτων χρησιμοποιούμε το τύπου MRF σύστημα των [58] το οποίο περιγράφηκε στο κεφάλαιο 2, με τις ρυθμίσεις που αναφέρονται στο παράρτημα 5.Β'. Για την κατασκευή του AAM χρησιμοποιήθηκαν 100 πρόσωπα από τη βάση [110], ενώ βασιστήκαμε στις τεχνικές που περιγράφονται στην αναφορά [40].

Για το δεύτερο σύνολο πειραμάτων για την κατασκευή των κάτωθεν μοντέλων ανίχνευσης χρησιμοποιούμε τις ρυθμίσεις του επόμενου κεφαλαίου, ενώ τα AAM κατασκευάζονται με 40 εικόνες από κάθε κατηγορία.

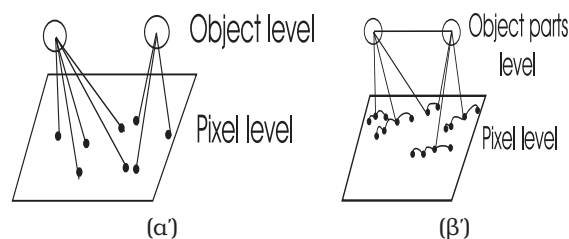
Για να συγκρίνουμε τα αποτελέσματά μας με αυτά άλλων ερευνητών, έχουμε χρησιμοποιήσει τις ρυθμίσεις της αναφοράς [62] για πρόσωπα και της [3] για αυτοκίνητα. Επιπλέον των ρυθμίσεων αυτών, μεγεθύνουμε τις εικόνες που περιέχουν αυτοκίνητα μεγεθύνονται κατά 2.5, καθώς χρειάζεται για το κάτωθεν σύστημα που παρουσιάζεται στο επόμενο κεφάλαιο. Τροποποιούμε ανάλογα το κριτήριο της αναφοράς [3] θεωρώντας ότι η προτεινόμενη θέση για ένα αυτοκίνητο είναι αληθής αν



Σχήμα 5.12: Αποτελέσματα εκμάθησης μοντέλου AAM: Αν και το σύνολο εκπαίδευσης περιέχει μεγάλη ποικιλία στο σχήμα και στην εμφάνιση, περίπλοκη δομή στο φόντο και επικαλύψεις των αντικειμένων, οι συνθέσεις είναι καθαρές και συλλαμβάνουν την εμφάνιση των αντικειμένων της κατηγορίας.



Σχήμα 5.13: Καμπύλες Ακρίβειας-Ανάκλιση για ανίχνευση αυτοκινήτων και προσώπων, χρησιμοποιώντας τα αυτόματα κατασκευασμένα μοντέλα. Και στις δύο περιπτώσεις, η χρήση του αλγορίθμου EM (στήλες (α),(γ)) οδηγεί σε βελτιωμένες επιδόσεις, τόσο λόγω των καθαρότερων παραμέτρων AAM όσο και της πληροφορίας κατάτμησης.



Σχήμα 5.14: Καθολικές (α) και τμηματικές (β) αρχιτεκτονικές για το πρόβλημα της συνέργειας

βρίσκεται στο εσωτερικό μίας έλλειψης γύρω από το κέντρο ενός αυτοκινήτου με οριζόντιο, μεγάλο και μικρό άξονα μήκους 2×62.5 και 2×25 pixels αντίστοιχα.

Επίσης, τα πρόσωπα κανονικοποιούνται ώστε η απόσταση μεταξύ των ματιών να είναι 50 pixels και η προτεινόμενη τοποθεσία θεωρείται ως ένα πρόσωπο εάν βρίσκεται στο εσωτερικό μίας έλλειψης διαστάσεων με οριζόντιο και κάθετο άξονα μήκους 2×30 και 2×50 pixels, αντίστοιχα.

5.5 Επισκόπηση Προηγούμενης Εργασίας και Συζήτηση

Έχοντας παρουσιάσει την δική μας συνεισφορά, εδώ πλέον τη συγκρίνουμε με τις υπάρχουσες εργασίες πάνω στο πρόβλημα που αντιμετωπίζουμε, ώστε να μπορούμε πλέον άμεσα να σχολιάζουμε τις σχέσεις τους με την εργασία μας.

5.5.1 Προηγούμενες εργασίες στην από κοινού ανίχνευση και κατάτμηση

Μπορούμε να ταξινομήσουμε τις περισσότερες από τις υπάρχουσες εργασίες στην από κοινού ανίχνευση και κατάτμηση αντικειμένων σε μία από τις αρχιτεκτονικές στο σχήμα Σχ. 5.14, με βάση το κατά πόσο χρησιμοποιούνε μία καθολική ή μία τμηματική αναπαράσταση του αντικειμένου. Οι αρχιτεκτονικές που βασίζονται σε καθολικές αναπαραστάσεις [69, 218, 131] τυπικά θεωρούν ότι ένα ενιαίο, μονολιθικό μοντέλο του αντικειμένου μπορεί να αναπαράγει την ποικιλία στο σχήμα και στην

εμφάνιση που χαρακτηρίζει την κατηγορία και να καταλάβει έτσι όλο το χωρίο του αντικειμένου. Αυτή η προσέγγιση έχει όλα τα προτερήματα αλλά και όλους τους περιορισμούς των αναγεννητικών μοντέλων που αναφέρθηκαν στο κεφάλαιο 2: αν και η ιδέα είναι διαισθητικά ελκυστική και μαθηματικά σαφής, συχνά είναι δύσκολη η κατασκευή ενός τέτοιου μοντέλου. Επίσης σε αυτό το πλαίσιο ο χειρισμός αρθρωτών κινήσεων και αυτοεπικαλύψεων γίνεται προβληματικός λόγω των απλουστευτικών υποθέσεών τους.

Από την άλλη, τα τμηματικά μοντέλα των αντικειμένων [19, 144, 238, 136] παρέχουν μία αναπαράσταση του αντικειμένου που προσφέρεται για τον υπολογισμό της κατάτμησης κατά έναν υβριδικό τρόπο, χρησιμοποιώντας άνωθεν πληροφορία όπου γίνεται διαθέσιμη από ένα τμήμα του αντικειμένου και χαμηλού επιπέδου χαρακτηριστικά για να ενώσουν το υπόλοιπο του αντικειμένου [238, 136]. Έτσι μπορούν να κατατμηθούν αποτελεσματικά αρθρωτά αντικείμενα, ενώ στο άρθρο [144] αναφέρονται άριστα αποτελέσματα στην ανίχνευση αυτοκινήτων. Αρκετά ερωτήματα όμως προκύπτουν σχετικά με την ορθότητα αυτών των προσεγγίσεων, ιδιαίτερα καθώς χρησιμοποιούνται με πολλές απλουστεύσεις, όπως να αγνοούνται οι συσχετίσεις μεταξύ των τμημάτων του αντικειμένου στην διαδικασία συμπερασματολογίας και να χρησιμοποιούνται ταυτόχρονα πολλαπλά τμήματα για την ερμηνεία των ίδιων παρατηρήσεων [144, 19], ή να χρησιμοποιούνται απλές κατανομές για να μοντελοποιήσουν τη φωτεινότητα μέσα σε όλο το χωρίο του αντικειμένου [136]. Πέρα από αυτό, ελλείπει ενός απτού παραμετρικού μοντέλου του αντικειμένου γίνεται δύσκολος ο συνδυασμός αυτών των μοντέλων με προβλήματα όπως η παρακολούθηση της κίνησης ενός αντικειμένου, ή η αναγνώριση της έκφρασης/κίνησής του.

Ερχόμενοι σε ένα πιο συγκεκριμένο επίπεδο, εξετάζουμε τώρα τις σχέσεις της προσέγγισής μας με τις βασικότερες αναφορές στο πεδίο. Η προσέγγιση των [252, 218] εκτελεί μία στοχαστική αναζήτηση στο χώρο των περιοχών και των υποθέσεων, επεκτείνοντας την προσέγγιση του Data-Driven MCMC [219] ώστε να εισάγει μοντέλα για τα αντικείμενα στο σύνολο των χρησιμοποιούμενων αναγεννητικών μοντέλων. Χρησιμοποιώντας τις κάτωθεν προτάσεις από καθιερωμένους αλγορίθμους [227, 28], κατά την αναζήτηση οι περιοχές και υποθέσεις των μοντέλων γεννώνται, ενώνονται ή απορρίπτονται ενώ τα σύνορά τους εντοπίζονται μέσω του Region Competition. Ο βασικός ισχυρισμός της ομάδας αυτής είναι ότι η κατάτμηση μίας εικόνας πρέπει να παρέχεται από μία συνεχή διαδικασία αναζήτησης στον χώρο λύσεων και όχι από μία μεμονωμένη λύση· συνεπώς βάσει του αλγορίθμου DDMCMC παράγονται δυναμικά πολλαπλές ερμηνείες της εικόνας, ενώ τα πιθανότερα αντικείμενα αναμένεται να είναι παρόντα στις περισσότερες από αυτές. Αν η υπόθεση κάποιου αντικειμένου προσωρινά αποκλεισθεί από την λύση, μετά από κάποιες επαναλήψεις θα επανεμφανισθεί, αυξάνοντας έτσι την πιθανότητα ανίχνευσής του. Παρά την κομψότητα και το δυναμικό αυτής της προσέγγισης, για ένα πρακτικό πρόβλημα ανίχνευσης αντικειμένων είναι πιθανό να μην χρειάζονται δείγματα από την ύστερη κατανομή πάνω σε ερμηνείες της σκηνής, και να αρκεί ένα παράθυρο το οποίο να περιέχει την υποψήφια τοποθεσία του αντικειμένου. Η δράση της συνέργειας στην περίπτωση αυτή μπορεί να συνοψιστεί στην εκτίμηση της πιθανότητας να είναι παρόν ένα αντικείμενο, που όπως δείχνουμε εδώ μπορεί να επιτευχθεί αποτελεσματικά μέσω του EM αντί στοχαστικής αναζήτησης.

Ακολουθώντας μία μη-αναγεννητική προσέγγιση, στις αναφορές [19, 18] και [144] αναπαραστάσεις μέσω codebooks χρησιμοποιούνται για την από κοινού ανίχνευση και κατάτμηση αντικειμένων, όπως φαίνεται στην αρχιτεκτονική στο Σχ.

5.14β. Χρησιμοποιώντας αποθηκευμένους πίνακες κατάτμησης για τα τμήματα του αντικειμένου, που καταδεικνύουν ποιο μέρος της εικόνας του ανήκει, η κατάτμησή του σχηματίζεται κατά έναν τρόπο σαν ένα παζλ, συνδυάζοντας της κατατμήσεις των επιμέρους τμημάτων του. Αν και αυτή η μέθοδος έχει αποδώσει πρακτικά δίνοντας ικανοποιητικά αποτελέσματα, τέτοιες αναπαραστάσεις είναι τυπικά περίπλοκες, χρησιμοποιώντας π.χ. 500 ή και περισσότερα τμήματα για το κάθε αντικείμενο, ενώ η κατάτμηση εξαρτάται από τη δυνατότητα να καλυφθεί μία μεγάλη έκταση του αντικειμένου με προαποθηκευμένα τμήματά του. Σε μία άλλη προσέγγιση που χρησιμοποιεί την αρχιτεκτονική του Σχ. 5.14β, στην αναφορά [238] εισάγεται ένα μέτρο προσήνειας (affinity measure) μεταξύ των pixels της εικόνας που μπορεί να ενσωματώνει πληροφορία σχετικής με τα αντικείμενα, οπότε χρησιμοποιώντας τεχνικές δυαδικής ομαδοποίησης βρίσκεται ένα ελάχιστο του κόστους διαμερισμού δεδομένων. Το μέτρο προσήνειας που χρησιμοποιείται οδηγεί σε μία ομαδοποίηση των pixels χρησιμοποιώντας τόσο χαμηλού επιπέδου πληροφορία (απουσία ακμών, ομοιότητα) όσο και υψηλού επιπέδου γνώση. Όμως, η έλλειψη μίας πιθανοτικής ερμηνείας εμποδίζει την συνεργασία τους με άλλες διεργασίες, ενώ το σύστημά τους τελικά δεν καθορίζει αν υπάρχει ένα αντικείμενο στην εικόνα. Η προσέγγισή μας χρησιμοποιεί ένα χαμηλοδιάστατο παραμετρικό μοντέλο τόσο για την κατάτμηση όσο και για την ανίχνευση ενός αντικειμένου είναι κατά τη γνώμη μας κομψότερη ενώ μπορεί να επεκταθεί σε δυσκολότερα προβλήματα, όπως να υπάρχουν πολλαπλά αντικείμενα, ή να εισαχθούν και άλλες πηγές πληροφορίας, όπως η κίνηση ή το βάθος.

Ερχόμενοι σε εργασίες που περιέχουν τον αλγόριθμο EM, σημειώνουμε αρχικά ότι η χρήση του αλγορίθμου EM για προβλήματα κατάτμησης σαφώς δεν είναι πρόσφατη· έχει χρησιμοποιηθεί ήδη για προβλήματα χαμηλού επιπέδου, όπως κατάτμηση εικόνων βάσει χαρακτηριστικών π.χ. [7, 17] ή εκτίμηση επιπέδων κίνησης [231]. Για το πρόβλημα της από κοινού ανίχνευσης και κατάτμησης αντικειμένων ο EM έχει προταθεί ανεξάρτητα στην αναφορά [136], όπου μία τμηματική αναπαράσταση που χρησιμοποιείται για ανίχνευση αντικειμένων συνδυάζεται με τον αλγόριθμο τομής γράφων [24] δίνοντας ακριβή αποτελέσματα στην κατάτμηση αρθρωτών αντικειμένων. Και εκεί χρησιμοποιείται ο αλγόριθμος EM, αλλά σε ένα πλαίσιο βελτιστοποίησης παρά εκτίμησης παραμέτρων: οι παράμετροι του σχήματος αντιμετωπίζονται ως κρυφές μεταβλητές, οπότε η δειγματοληψία της κατανομής τους διατυπώνεται ως το βήμα E. Το βήμα M τότε αντιστοιχίζεται στην μεγιστοποίηση μέσω τεχνικών τομής γράφων ενός όρου που ποσοτικοποιεί την ποιότητα κατάτμησης, ο οποίος ενσωματώνει τα δείγματα που λαμβάνονται κατά το βήμα E. Αυτή η προσέγγιση παρεκκλίνει από την χρήση του EM στο πλαίσιο των αναγεννητικών μοντέλων, όπου η εκτίμηση των παραμέτρων αντιστοιχίζεται στο βήμα M, και τα κρυφά δεδομένα χρησιμοποιούνται για να συνδέσουν τις παρατηρήσεις με τις παραμέτρους. Είναι δύσκολο να φανταστεί κανείς πώς η προσέγγιση αυτή θα μπορούσε να συνδυαστεί με τις τεχνικές που χρησιμοποιήθηκαν στη δική μας εργασία, όπως εξέλιξη καμπυλών και AAM, ή με ένα πιο εκφραστικό αναγεννητικό μοντέλο, όπου η εκτίμηση των παραμέτρων μπορεί να είναι περίπλοκη και η δειγματοληψία μπορεί να μην αποδίδει λόγω της μεγάλης διάστασης του διανύσματος παραμέτρων.

Τέλος, στις εργασίες των [69, 70, 236] ο EM χρησιμοποιείται για να κατατμήσει την εικόνα βάσει πληροφορίας σχετικά με τα αντικείμενα, με το μοντέλο των αντικειμένων και επιπέδων (sprites and layers). Διαισθητικά η προσέγγιση αυτή είναι εφάμιλλη της δικής μας, αλλά η αλληλεπίδραση μεταξύ των δύο διεργασιών

δεν ερευνάται: το μοντέλο του φόντου εκτιμάται από ένα γνωστό σύνολο εικόνων, εισάγοντας πρότερη γνώση που σπάνια είναι διαθέσιμη γενικότερα, ενώ τελικά δεν αποφασίζεται αν υπάρχει ένα αντικείμενο στην εικόνα. Επίσης, στις χωρικές παραμορφώσεις που χρησιμοποιούνται δεν εκμεταλλεύονται την πρότερη γνώση που υπάρχει σχετικά με την ποικιλία στο σχήμα, καθώς είτε θεωρούν απλούς μετασχηματισμούς [69] ή χρησιμοποιούν έναν γενικό όρο ποινής τύπου MRF πάνω στις τιμές ενός τοπικά σταθερού πεδίου παραμόρφωσης [236]. Αντιθέτως, τα μοντέλα τύπου AAM που χρησιμοποιούμε στην εργασία μας αποτελούν μία ισχυρή κλάση αναγεννητικών μοντέλων, που συλλαμβάνει την ποικιλία στο σχήμα μίας κατηγορίας αντικειμένων χρησιμοποιώντας μία απλή παραμετρική έκφραση.

5.5.2 Προηγούμενες εργασίες στην κατάτμηση με πληροφορία σχήματος

Συμπληρωματικά με την προσέγγιση της από κοινού αντιμετώπισης των υψηλού και χαμηλού επιπέδου λειτουργιών της όρασης, οι γενικής φύσης πρότεροι όροι ποινής που τυπικά χρησιμοποιούνται για το πρόβλημα της κατάτμησης [248, 168] έχουν σταδιακά επεκταθεί ενσωματώνοντας υψηλού επιπέδου γνώση: συγκεκριμένα πληροφορία σχετικά με το σχήμα μίας κατηγορίας αντικειμένων έχει συνδυαστεί με την εξέλιξη καμπυλών χρησιμοποιώντας τεχνικές επιπεδοσυνόλων, π.χ. στις αναφορές [145, 194, 46], δίνοντας πρακτικούς και αποδοτικούς αλγορίθμους. Αυτές οι εργασίες διαφέρουν από την τυπική προσέγγιση στο ότι η μοντελοποίηση του αντικειμένου δεν εισάγεται απευθείας στη διαδικασία κατάτμησης, αλλά ένας επιπλέον όρος προστίθεται στο ελαχιστοποιούμενο συναρτησιακό, που εκφράζει την γνώση για το σχήμα του αντικειμένου μέσω της συνάρτησης εμβύθισης της ελίσσόμενης καμπύλης. Οι προσεγγίσεις αυτές υποφέρουν από διάφορες πτυχές, που μπορούν να αντιμετωπιστούν χρησιμοποιώντας ένα AAM για να εκφράσει τη γνώση σχετικά με το σχήμα.

Στην αναφορά [145] μία ανάλυση τύπου PCA χρησιμοποιείται κατά τη διάρκεια της εκμάθησης πάνω στις συναρτήσεις των περιγραμμάτων που αποτελούν το σύνολο εκπαίδευσης ώστε να κωδικοποιηθεί η μεταβλητότητα στο σχήμα χρησιμοποιώντας μία χαμηλοδιάστατη βάση. Η βάση αυτή στη συνέχεια χρησιμοποιείται για να οδηγήσει τη συνάρτηση εμβύθισης προς πιθανότερες τοποθεσίες κρατώντας το σχήμα της πλησιέστερο σε αυτό των σχημάτων που χρησιμοποιούνται κατά την εκπαίδευση. Η καταλληλότητα της PCA για ανάλυση σχημάτων μέσω συναρτήσεων απόστασης δεν είναι όμως δεδομένη. Θεωρητικά, οι συναρτήσεις απόστασης δεν είναι μέλη ενός γραμμικού χώρου συναρτήσεων, οπότε η PCA δεν μπορεί να δικαιολογηθεί καθώς στην βάση αυτής της μεθόδου βρίσκεται η χρήση διαφορών. Πρακτικά, τα μοντέλα που κατασκευάζονται από την προσέγγιση μέσω της PCA δεν είναι πάντα κατάλληλα. Για παράδειγμα μπορούμε να θεωρήσουμε ένα σύνολο εκπαίδευσης αποτελούμενο από σχήματα που αντιστοιχούν στα ενδιάμεσα στάδια της παραμόρφωσης ενός κύκλου σε μία έλλειψη. Το μοντέλο αυτό θα χρειαστεί μία βάση της ίδιας διάστασης με το πλήθος του συνόλου εκπαίδευσης για να ανακατασκευάσει όλα τα σχήματά του, καθώς καμία από τις ενδιάμεσες συναρτήσεις απόστασης δεν μπορεί να εκφραστεί ως γραμμικός συνδυασμός των υπόλοιπων. Γνωρίζουμε όμως ότι υπάρχει μόνο μία μορφή παραμόρφωσης, που μπορεί να εκφραστεί με ακρίβεια χρησιμοποιώντας ένα AAM με ένα μόνο στοιχείο βάσης.

Εναλλακτικά, στην εργασία [194] μία Gaussian κατανομή χρησιμοποιείται σε

κάθε σημείο για να εκφράσει την πιθανότητα η παρατήρηση να ανήκει στο εσωτερικό της ελισσόμενης καμπύλης, που εκφράζεται ως η συνάρτηση *Heavyside* της εμβύθισης της καμπύλης. Η πρότερη κατανομή αυτή επιτρέπει τοπικές παραμορφώσεις, καθώς ορίζεται σε επίπεδο σημείων, αλλά δεν μπορεί να συλλάβει με ακρίβεια την κατανομή του συνόλου εκπαίδευσης. Θεωρώντας και πάλι το παράδειγμα της έλλειψης, δεν θα ήταν δυνατό να περιορίσουμε την καμπύλη από το να είναι στρογγυλή από την μία πλευρά και ελλειπτική από την άλλη. Αυτό το μοντέλο εσφαλμένα εισάγει N ανεξάρτητες μεταβλητές, όπου N είναι το πλήθος των σημείων της τεμπλέτας, ενώ υπάρχει μόνο μία, δηλαδή ο παράγοντας παραμόρφωσης.

Ένα τελευταίο θέμα αφορά την χρήση πρότερης γνώσης σχετικά με την κατανομή των παραμέτρων του AAM· κάθε αλγόριθμος ο οποίος χρησιμοποιεί πρότερη γνώση για το σχήμα θα δώσει εσφαλμένα αποτελέσματα κατάτμησης απουσία ενός αντικειμένου, ενισχύοντας τα -ανύπαρκτα- όριά του. Με την προσέγγισή μας οι ψευδείς ανιχνεύσεις καταρρέουν τόσο λόγω της ανακριβούς μοντελοποίησης των παρατηρήσεων, όσο και εξαιτίας της χαμηλής πιθανοφάνειας των παραμέτρων του μοντέλου.

Σε ένα γενικότερο επίπεδο, η άνωθεν προσέγγιση των αναγεννητικών μοντέλων στην κατάτμηση συνίσταται στον καθορισμό των αναθέσεων παρατηρήσεων της εικόνας σε μοντέλα αντικειμένων. Δύο σχετιζόμενα θεωρητικά σημεία γίνονται έτσι εμφανή:

Πρώτον, οι περιορισμοί στην κατάτμηση θα έπρεπε να είναι γενικής φύσης όροι, μοντελοποιώντας τον τρόπο με τον οποίο ένα αντικείμενο μπορεί να επικαλύπτεται από άλλα. Για παράδειγμα το μικρό μήκος ορίων εγγυάται ότι δεν εισάγουμε την υπόθεση της επικάλυψης για κάθε μεμονωμένη παρατήρηση που δεν μοντελοποιείται ικανοποιητικά από το μοντέλο του αντικειμένου. Ομοίως, η επιδιωκόμενη σύμπτωση των ορίων των περιοχών με τις ακμές της εικόνας μπορεί να δικαιολογηθεί από τις τυπικά διαφορετικές τιμές φωτεινότητας των αντικειμένων που βρίσκονται μπροστά ή πίσω από το αντικείμενο που κατατμείται.

Δεύτερον, τα όρια ενός αντικειμένου και της περιοχής που του ανατίθεται δεν πρέπει υποχρεωτικά να σχετίζονται, αν και αυτό συμβαίνει όταν δεν βρίσκεται τίποτα μπροστά στο αντικείμενο. Συνεπώς πληροφορία υψηλότερου επιπέδου, π.χ. σχετικά με την πληροφορία σχήματος θα έπρεπε να εκφράζεται από το μοντέλο του αντικειμένου, και όχι από κάποιον πρότερο όρο κατάτμησης, όπως γίνεται στις εργασίες που ακολουθούν αυτή την προσέγγιση, όπου επαναεφευρίσκεται ο τροχός, χρησιμοποιώντας τη συνάρτηση εμβύθισης για την μοντελοποίηση σχήματος.

5.Α' Σχέση Ανανέωσης Παραμέτρων AAM Βάσει του EM

Για έναν πίνακα F που επηρεάζεται από τις παραμέτρους s θεωρούμε την γραμμικοποίησή του γύρω από το s ως $F_{s+\Delta s} = F_s + \frac{dF}{ds} \Delta s$, όπου ο F είναι $N \times 1$, ο $\frac{dF}{ds}$ $N \times p$ και οι πίνακες s και Δs είναι $p \times 1$. Η συνάρτηση κόστους προσεγγίζεται τότε κρατώντας τους όρους πρώτης και δεύτερης τάξης, δίνοντας μία έκφραση της μορφής $C(s + \Delta s) = C(s) + \mathcal{J} \Delta s + \Delta s^T \mathcal{H} \Delta s$, με τον \mathcal{J} να είναι η $1 \times p$ Jacobian της συνάρτησης κόστους και η \mathcal{H} η $p \times p$ Hessian. Για το κριτήριο που περιλαμβάνει την πληροφορία κατάτμησης $C(s) = [H \odot E \odot \mathcal{E}]^T [\mathcal{E}]$, μπορούμε να γράψουμε

$$\begin{aligned} C(s + \Delta s) &= \left[\underbrace{H}_{a_0} \odot \left(\underbrace{E}_{b_0} + \underbrace{\frac{dE}{ds} \Delta s}_{b_1} \right) \odot \left(\underbrace{\mathcal{E}}_{c_0} + \underbrace{\frac{d\mathcal{E}}{ds} \Delta s}_{c_1} \right) \right]^T \left(\underbrace{\mathcal{E}}_{d_0} + \underbrace{\frac{d\mathcal{E}}{ds} \Delta s}_{d_1} \right) \\ &= \underbrace{[a_0 \cdot b_0 \cdot c_0]^T}_{\text{erase}} d_0 + 2 [a_0 \cdot b_0 \cdot c_0]^T d_1 + [a_0 \cdot b_0 \cdot c_1]^T d_1 \\ &\quad + [a_0 \cdot b_1 \cdot c_0]^T d_0 + 2 [a_0 \cdot b_1 \cdot c_0]^T d_1 + \underbrace{[a_0 \cdot b_1 \cdot c_1]^T}_{\text{erase}} d_1 \end{aligned}$$

όπου ο υποδείκτης δηλώνει την τάξη στο ανάπτυγμα Taylor των μεμονωμένων όρων, ενώ οι όροι που απαλείφονται είναι είτε σταθεροί ή μεγαλύτερης τάξης και συνεπώς αγνοούνται. Ενώνοντας τους όρους πρώτης και δεύτερης τάξης έχουμε λοιπόν:

$$\begin{aligned} \mathcal{H} &= \left[H \odot E \odot \left(\frac{d\mathcal{E}}{ds} \Delta s \right) \right]^T \frac{d\mathcal{E}}{ds} \Delta s + 2 \left[H \odot \left(\frac{dE}{ds} \Delta s \right) \odot \mathcal{E} \right]^T \frac{d\mathcal{E}}{ds} \Delta s \\ \mathcal{J} &= 2 [H \odot E \odot \mathcal{E}]^T \frac{d\mathcal{E}}{ds} \Delta s + \left[H \odot \frac{dE}{ds} \Delta s \odot \mathcal{E} \right] \mathcal{E} \end{aligned} \quad (5.42)$$

που μας δίνει την αντίστοιχη σχέση ανανέωσης για αυτό το κριτήριο.

Για να εισάγουμε την οριζουσα της παραμόρφωσης χρησιμοποιούμε τη σχέση (2.30) και την εκφράζουμε συναρτήσει του διανύσματος s και των οριζόντιων και κάθετων παραγώγων των στοιχείων της βάσης παραμόρφωσης:

$$D(x; s) = \left(\sum_k s_k \frac{\partial S_{k,x}}{\partial x} \right) \left(\sum_j s_j \frac{\partial S_{j,y}}{\partial y} \right) - \left(\sum_k s_k \frac{\partial S_{k,y}}{\partial x} \right) \left(\sum_k s_y \frac{\partial S_{k,x}}{\partial y} \right)$$

Μπορούμε τότε να σχηματίσουμε την πρώτη τάξης προσέγγιση:

$$\begin{aligned} D(s + \Delta s) &= [(s + \Delta s) \mathcal{S}_x^x] \odot [(s + \Delta s) \mathcal{S}_y^y] - [(s + \Delta s) \mathcal{S}_y^x] \odot [(s + \Delta s) \mathcal{S}_x^y] \\ &= D + \Delta s [\mathcal{S}_x^x \odot s \mathcal{S}_y^y + \mathcal{S}_y^y \odot s \mathcal{S}_x^x + \mathcal{S}_y^x \odot s \mathcal{S}_x^y + \mathcal{S}_x^y \odot s \mathcal{S}_y^x] + O(\Delta s^2) \\ &\simeq D + \frac{dD}{ds} \Delta s \end{aligned}$$

και να την εισάγουμε στο αντίστοιχο κριτήριο:

$$C(\mathbf{s} + \Delta\mathbf{s}) = \left[\mathbf{H} \odot \left(\mathbf{D} + \frac{d\mathbf{D}}{ds} \Delta\mathbf{s} \right) \odot \left(\mathbf{E} + \frac{d\mathbf{E}}{ds} \Delta\mathbf{s} \right) \odot \left(\boldsymbol{\varepsilon} + \frac{d\boldsymbol{\varepsilon}}{ds} \Delta\mathbf{s} \right) \right]^T \left(\boldsymbol{\varepsilon} + \frac{d\boldsymbol{\varepsilon}}{ds} \Delta\mathbf{s} \right)$$

Χρησιμοποιώντας μία αντίστοιχη διαδικασία με την προηγούμενη, δηλαδή διατηρώντας μόνο τους όρους πρώτης και δεύτερης τάξης καταλήγουμε στις παρακάτω εκφράσεις για την Jacobian και Hessian της συνάρτησης κόστους:

$$\begin{aligned} \mathcal{H} &= \left[\mathbf{H}' \odot \mathbf{D} \odot \left(\Delta\mathbf{s} \frac{d\boldsymbol{\varepsilon}}{ds} \right) \right]^T \Delta\mathbf{s} \frac{d\boldsymbol{\varepsilon}}{ds} + 2 \left[\mathbf{H} \odot \mathbf{D} \odot \left(\Delta\mathbf{s} \frac{d\mathbf{E}}{ds} \odot \boldsymbol{\varepsilon} \right) \right]^T \Delta\mathbf{s} \frac{d\boldsymbol{\varepsilon}}{ds} \\ &+ 2 \left[\mathbf{H}' \odot \left(\Delta\mathbf{s} \frac{d\mathbf{D}}{ds} \odot \boldsymbol{\varepsilon} \right) \right]^T \Delta\mathbf{s} \frac{d\boldsymbol{\varepsilon}}{ds} \\ \mathcal{J} &= 2 \left[\mathbf{H}' \odot \mathbf{D} \odot \boldsymbol{\varepsilon} \right]^T \Delta\mathbf{s} \frac{d\boldsymbol{\varepsilon}}{ds} + \left[\mathbf{H} \odot \mathbf{D} \odot \left(\Delta\mathbf{s} \frac{d\mathbf{E}}{ds} \right) \odot \boldsymbol{\varepsilon} \right]^T \boldsymbol{\varepsilon} \\ &+ \left[\mathbf{H}' \odot \left(\frac{d\mathbf{D}}{ds} \Delta\mathbf{s} \right) \odot \boldsymbol{\varepsilon} \right]^T \boldsymbol{\varepsilon} \end{aligned} \quad (5.43)$$

Σημειώνουμε εδώ ότι για απλότητα θεωρούμε τις παραμορφώσεις ως ομαλές 1-1 συναρτήσεις (διφεομορφισμούς) που σημαίνει ότι η Jacobian δεν μπορεί να αλλάξει πρόσημο, καθώς διαφορετικά θα είχαμε αναδίπλωση του αντικειμένου. Έτσι δικαιολογείται ότι παραπάνω χρησιμοποιούμε την ορίζουσα αντί της απόλυτης τιμής.

5.Β' Περιγραφή Μοντέλου Ανίχνευσης

Όπως στην αναφορά [58] χρησιμοποιούμε μία κατευθύνσιμη - διαχωρίσιμη (steerable - separable) πολυκλιμακωτή συστοιχία φίλτρων σε τέσσερις κλίμακες και κατευθύνσεις για να εξάγουμε αποτελεσματικά μία τοπική περιγραφή της εικόνας: για το λόγο αυτό χρησιμοποιούμε τον πυρήνα διαφορών από Gaussians και τα ορθογωνικά ζεύγη φίλτρων της αναφοράς [68] που ως ένα βαθμό μιμούνται τη λειτουργία των νευρώνων στα πρώτα στάδια του βιολογικού συστήματος όρασης. Σύμφωνα με την αναφορά [58], οι κατανομές των χαρακτηριστικών μοντελοποιούνται χρησιμοποιώντας Gaussian κατανομές και εισάγονται σε ένα κριτήριο τύπου MRF ως δυναμικά παρατήρησης $\Phi(l_n) = P_n(\mathbf{F}_{1_n})$. Η ποσότητα αυτή εκφράζει την πιθανοφάνεια των εξόδων της συστοιχίας φίλτρων \mathbf{F}_{1_n} στο σημείο l_n υπό την κατανομή του τμήματος n .

Εισάγοντας τα δυναμικά ζεύγους $\Psi_{n,m}(l_n, l_m)$ μεταξύ των τοποθεσιών των τμημάτων του αντικειμένου n, m μπορούμε πλέον να συνδυάσουμε την πρότερη γνώση σχετικά με τους δυνατούς συνδυασμούς τους με την πιθανοφάνεια των μεμονωμένων παρατηρήσεων και να εκφράσουμε έτσι την ύστερη κατανομή πιθανότητας πάνω στο διάνυσμα θέσεων: $L = (l_1, \dots, l_N)$:

$$P(L|\mathbf{F}) \propto P(\mathbf{F}|L)P(L) = \prod_n \Phi_n(l_n) \prod_{n,m \in V} \Psi_{n,m}(l_n, l_m) \quad (5.44)$$

όπου V είναι το σύνολο των ακμών του MRF.

Οι ενδιαφέρουσες τοποθεσίες για το αντικείμενο μπορούν να εξαχθούν από τα μέγιστα της ύστερης κατανομής, ο υπολογισμός των οποίων επισπεύδεται από την

διαχωρισιμότητα των Gaussian συναρτήσεων που εκφράζουν τα δυναμικά ζεύγους και τον αλγόριθμο Belief Propagation ο οποίος συνδυάζεται με ένα γραφικό μοντέλο τύπου δέντρου για αποτελεσματική συμπερασματολογία. Αποκλίνοντας ελαφρά από την προσέγγιση των [58] επιτυγχάνουμε μία βελτίωση στην επίδοση εισάγοντας την πιθανοφάνεια των εξόδων της συστοιχίας φίλτρων υπό την εναλλακτική υπόθεση του φόντου \mathcal{B} $P_{\mathcal{B}}(\mathbf{F}_{1_n})$ και μεγιστοποιώντας ως προς L την ποσότητα :

$$\frac{P(\mathbf{F}, L|\mathcal{O})}{P(\mathbf{F}, L|\mathcal{B})} = \frac{P(\mathbf{F}|L, \mathcal{O})P(L|\mathcal{O})}{P(\mathbf{F}|L, \mathcal{B})P(L|\mathcal{B})} = \prod_n \Phi'_n(l_n) \prod_{n,m \in V} \Psi_{n,m}(l_n, l_m) \quad \Phi'_n(l_n) = \frac{P_f(\mathbf{F}_{1_n})}{P_b(\mathbf{F}_{1_n})}. \quad (5.45)$$

Ο όρος $P(L|\mathcal{B})$ θεωρείται σταθερός, καθώς το μοντέλο του φόντου έχει μία ομοιόμορφη κατανομή πάνω στο σύνολο των τοποθεσιών. Η κατανομή $P_b(\mathbf{F}_{1_n})$ κατασκευάζεται από τυχαίες εικόνες, χρησιμοποιώντας σημεία μέγιστου της ενέργειας χαρακτηριστικών τύπου ακμών ράβδων και κηλίδων, που παρουσιάζονται στο επόμενο κεφάλαιο. Η ποσότητα $\frac{P(\mathbf{F}|L)}{P(\mathbf{F}|\mathcal{B})}$ ανιχνεύει έτσι με μεγαλύτερη ακρίβεια της δομές της εικόνας που αντιστοιχούν στα χαρακτηριστικά σημεία του αντικειμένου, επιτρέποντας συγχρόνως τη χρήση αποδοτικών διακριτικών τεχνικών για την εκτίμησή της. Στην περίπτωση μας χρησιμοποιούμε την προβολή των δεδομένων στο διαχωριστικό υπερεπίπεδο Fisher [55], που μπορεί να αποτιμηθεί αποδοτικά.

Κεφάλαιο 6

Συνεργετική Ανίχνευση Αντικειμένων μέσω Σημείων Ενδιαφέροντος και Γραφικών Δικτύων

6.1 Εισαγωγή

Στο κεφάλαιο αυτό προτείνεται μία μέθοδος για την συνέργεια μεταξύ υψηλού και χαμηλού επιπέδου πληροφορίας για ανίχνευση αντικειμένων που χρησιμοποιεί γραφικά μοντέλα και σημεία ενδιαφέροντος σε ένα πιθανοτικό πλαίσιο. Ο συνδυασμός των δύο τεχνικών επιτυγχάνεται με την χρήση απλών αναγεννητικών μοντέλων στο επίπεδο των δυναμικών παρατήρησης του γραφικού δικτύου.

Οι συνεισφορές μας συνοψίζονται σε τρία σημεία

- Εξαγωγή σημείων ενδιαφέροντος τύπου ράβδων και ακμών χρησιμοποιώντας τεχνικές από την πολυκλιμακωτή ανάλυση εικόνων, και χρήση τους για την αναπαράσταση και ανίχνευση αντικειμένων.
- Συνδυασμός άνωθεν και κάτωθεν πληροφορίας μέσω μίας αποδοτικής μεθόδου για την εκτίμηση πιθανοφάνειας υποθέσεων.
- Εκμάθηση γραφικών μοντέλων για την αναπαράσταση αντικειμένων.

6.2 Σημεία Ενδιαφέροντος για Ανίχνευση Αντικειμένων.

Η χρήση σημείων ενδιαφέροντος για προβλήματα όρασης υπολογιστών αντιμετωπίζει δύο από τα βασικά προβλήματα της όρασης, το πολύ μεγάλο πλήθος παρατηρήσεων που περιέχονται σε μία εικόνα και την συνύπαρξη αντικειμένων που βρίσκονται σε διαφορετικές κλίμακες.

Εξάγοντας ένα σύνολο χαρακτηριστικών τύπου κηλίδας (blob) [148, 149], διασταύρωσης (junction) ή γωνίας [148, 158, 198], αποκτούμε ένα σύνολο αξιόπιστων χαρακτηριστικών που προσφέρεται για λειτουργίες όπως αντιστοίχιση, εκτίμηση κίνησης κτλ.

Τα σημεία αυτά εντοπίζουν τις ενδιαφέρουσες περιοχές της εικόνας, με την έννοια ότι περιέχουν μη-τετριμμένη δομή· για παράδειγμα από μία εικόνα που περιέχει έναν φωτεινό κύκλο σε ένα σκοτεινό φόντο, ένα σημείο ενδιαφέροντος είναι το κέντρο

του κύκλου και αν χρησιμοποιείται ένα ανεξάρτητο κλίμακας κριτήριο, εντοπίζεται στην κλίμακα του κύκλου. Στο ακραίο αυτό παράδειγμα συνοψίζεται με ένα σημείο η δομή μίας εικόνας που μπορεί να περιέχει $O(10^4)$ παρατηρήσεις.

Ξεκινώντας από τις εργασίες των [233, 149, 198] μία σημαντική εφαρμογή αυτών των μεθόδων έχει προκύψει, η εφαρμογή τους ως περιγραφείς της εικόνας για τη λειτουργία της ανίχνευσης αντικειμένων. Συγκεκριμένα, από την περιοχή της εικόνας που καταδεικνύει ο τελεστής ενδιαφέροντος εξάγεται μία χαμηλοδιάστατη περιγραφή μέσω PCA [62] ή, συχνότερα πλέον, με τον μετασχηματισμό SIFT - Scale Invariant Feature Transform [149] η οποία περιγράφει την τοπική δομή της εικόνας. Το σχετικά μικρό πλήθος σημείων που εξάγονται από μία εικόνα καθιστά δυνατή την εξαντλητική αναζήτηση πάνω σε κλίμακες τόσο για εκμάθηση όσο και για ανίχνευση, και επιτρέπει την εισαγωγή τεχνικών μάθησης μηχανών. Πέρα από αυτό, η ανεξάρτητη κλίμακας ανίχνευση αυτών των σημείων μπορεί να αυτοματοποιήσει την εκτίμηση της κλίμακας του αντικειμένου [62], παρακάμπτοντας την εξαντλητική αναζήτηση που θα χρειαζόταν διαφορετικά.

Όμως, δύο προβληματικές πτυχές της χρήσης μίας τέτοιας προσέγγισης είναι:

- Οι δομές τύπου κηλίδας και διακλάδωσης στις οποίες αποκρίνονται οι περισσότερες μέθοδοι [198, 148, 149, 158] δεν είναι κατάλληλες για την περιγραφή επίμηκων δομών, όπως χείλη, πόδια κτλ., που παρέχουν χρήσιμη πληροφορία για τις περισσότερες κατηγορίες αντικειμένων.
- Πολλές φορές κάποια χρήσιμα σημεία δεν θεωρούνται αρκετά σημαντικά, λόγω χαμηλής έντασης, μερικής παρατηρησιμότητας, θορύβου κοκ. Ως εκ τούτου κάποια σημεία χάνονται, με αποτέλεσμα την προβληματική ανίχνευση των αντικειμένων.

Στο κεφάλαιο αυτό για το πρώτο πρόβλημα εισάγουμε μία μέθοδο που ανιχνεύει και εξάγει περιγραφείς για χαρακτηριστικά τύπου ακμών και ράβδων. Για το δεύτερο πρόβλημα χρησιμοποιούμε τις μεθόδους των γραφικών μοντέλων για να καθοδηγήσουμε την αναζήτηση για τα χαρακτηριστικά που δεν ανιχνεύθηκαν, σε συνδυασμό με μία αποτελεσματική μέθοδο για την εκτίμηση της πιθανοφάνειας των παραγόμενων υποθέσεων τοποθεσίας των σημείων.

Ένας από τους κύριους ισχυρισμούς μας είναι ότι η γεωμετρική πληροφορία και γενικής φύσης πρότυπα, όπως κηλίδες, ευθείες και ράβδοι επαρκούν για μία ικανοποιητική επίδοση σε ανίχνευση κατηγοριών αντικειμένων *χωρίς* την χρήση περαιτέρω πληροφορίας για την εμφάνιση. Εάν και η πληροφορία που μπορεί να παρέχουν περιγραφείς τύπου SIFT ή οι συντελεστές ενός αναπτύγματος σε PCA μπορεί να είναι χρήσιμη για να ξεχωρίσουν δύο αντικείμενα της ίδιας κατηγορίας, η ανίχνευση ενός αντικειμένου μπορεί να γίνει βάσει αναπαραστάσεων γενικότερης φύσης.

Στο πρώτο μέρος περιγράφουμε τη μέθοδο εξαγωγής σημείων τύπου primal-sketch (ακμές, ράβδοι και κύκλοι) βάσει των τεχνικών του T. Lindeberg. Στο δεύτερο μέρος παρουσιάζουμε αφενός μία μέθοδο για την εκμάθηση ενός γραφικού δικτύου που μοντελοποιεί τις χωρικές εξαρτήσεις μεταξύ των σημείων του αντικειμένου, αφετέρου μία αποτελεσματική μέθοδο για την συμπλήρωση των σημείων ενδιαφέροντος που μπορεί να μην ανιχνεύθηκαν από το αρχικό σύστημα.

6.3 Χαρακτηριστικά Τύπου Primal Sketch

Η αναπαράσταση της εικόνας μέσω πληροφορίας που εξάγεται γύρω από τέτοιες δομές μπορεί να δίνει καλά αποτελέσματα, π.χ. [233, 62], αλλά αυτός δεν είναι υποχρεωτικά ο μόνος τρόπος να εξαχθεί μία σποραδική αναπαράσταση της εικόνας. Οι περιγραφείς τύπου ακμών και σκελετών έχουν χρησιμοποιηθεί εκτενώς για την μοντελοποίηση και αναγνώριση αντικειμένων [99, 203], ενώ ενδείξεις από το ανθρώπινο οπτικό σύστημα [143] υποδεικνύουν ότι διαδικασίες ανίχνευσης τέτοιων χαρακτηριστικών λαμβάνουν χώρα στα πρώτα στάδια επεξεργασίας της οπτικής πληροφορίας.

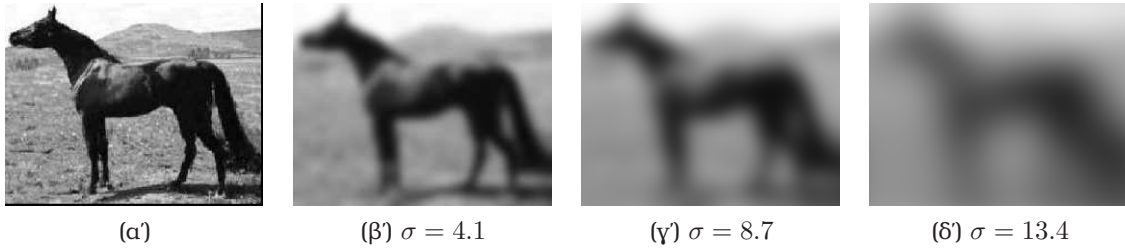
Ένα πρακτικό πρόβλημα που προκύπτει όταν χειριζόμαστε πληροφορία τύπου ακμής και σκελετών είναι ότι δεν υπάρχει μία προφανής περιγραφή μίας συνεχούς καμπύλης που να προσφέρεται σε τεχνικές μάθησης μηχανών· αντιθέτως τα χαρακτηριστικά τύπου κηλίδας μπορούν να περιγραφούν μόνο μέσω της θέσης τους, που είναι διακριτή. Επίσης, η εξαγωγή τέτοιων χαρακτηριστικών από συνεχείς εικόνες απαιτεί την εισαγωγή ενός κατωφλιού, που μπορεί να οδηγήσει στον κατακερματισμό των καμπυλών.

Στη σχετική εργασία των [159] η εικόνα ακμών περιγράφεται χρησιμοποιώντας σημειακά χαρακτηριστικά, που επιλέγονται με ένα ανεξάρτητο κλίμακας κριτήριο· αυτή η αναπαράσταση μπορεί να παράγει ένα υπερβολικά μεγάλο πλήθος από σημεία, ενώ δεν χρησιμοποιεί άμεσα την πληροφορία που σχετίζεται με το μήκος της καμπύλης, που είναι γενικά χρήσιμη για αναγνώριση. Στη αναφορά [63] οι ακμές εξάγονται χρησιμοποιώντας ανίχνευση ακμών τύπου Canny, ένωση ακμών, ανίχνευση σημείων στα οποία η καμπυλότητα αλλάζει πρόσημο και ακολούθως προβολή σε μία βάση PCA. Το γεγονός ότι η κλίμακα επηρεάζει την διαδικασία ανίχνευσης μπορεί να αντιμετωπιστεί χρησιμοποιώντας την προσέγγιση του [147] που χρησιμοποιούμε και εμείς στη συνέχεια· όμως η χρησιμοποιούμενη ακολουθία από λειτουργίες είναι ευαίσθητη στο θόρυβο, στις επικαλύψεις, αυτοεπικαλύψεις και παραμορφώσεις του αντικειμένου. Τέλος στην αναφορά [156] περιοχές που είναι αφινικά συνδιακυμαινόμενες (affine covariant) εξάγονται χρησιμοποιώντας ένα κριτήριο βασισμένο σε περιοχές, το οποίο όμως δεν προσφέρεται σε μία σαφή αναγεννητική ερμηνεία όπως τα χαρακτηριστικά ράβδου που χρησιμοποιούνται εδώ.

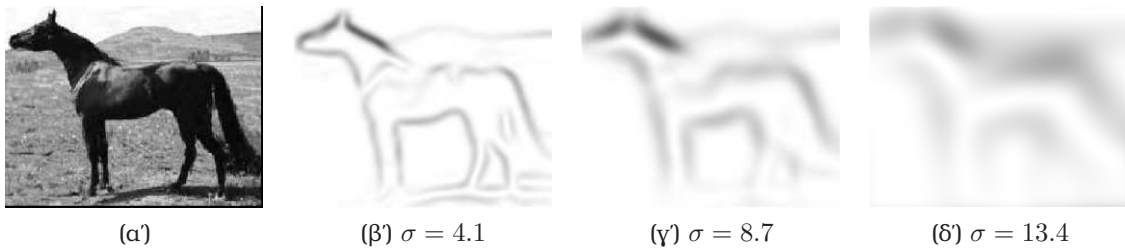
Παρακάτω, βασισμένοι στο ανεξάρτητο κλίμακας πλαίσιο ανίχνευσης χαρακτηριστικών του D. Lindeberg [148, 147] παρουσιάζουμε μία μέθοδο που εξάγει ευθεία τμήματα ακμών και ράβδων κατά έναν τρόπο που δεν επηρεάζεται από την κλίμακα, παρέχοντας απλούς περιγραφείς της ίδιας πολυπλοκότητας με ένα χαρακτηριστικό τύπου κηλίδας. Η συνεισφορά μας εδώ συνίσταται στην εξαγωγή από τις συνεχείς καμπύλες των ανιχνευθέντων χαρακτηριστικών απλών περιγραφέων που ορίζονται σε μεμονωμένα σημεία, καθώς και η πρόταση της χρήσης τους σε ένα σύστημα ανίχνευσης αντικειμένων.

6.3.1 Χαρακτηριστικά Ακμών

Το πρόβλημα της επιλογής κλίμακας στην ανίχνευση ακμών προκύπτει αφενός από την συνύπαρξη απότομων και σταδιακών μεταβάσεων στην φωτεινότητα, αφετέρου στην παρουσία θορύβου άγνωστης διασποράς, από την οποία καθορίζεται η βέλτιστη κλίμακα του φίλτρου ανίχνευσης ακμών. Μία λύση στο πρόβλημα είναι να χρησιμοποιούνται ταυτόχρονα φίλτρα σε πολλές κλίμακες και να συνδυάζονται οι αποκρίσεις



Σχήμα 6.1: Gaussian scale space της εικόνας, που προκύπτει συνελίσσοντας την εικόνα με Gaussian πυρήνες αυξανόμενης διασποράς σ . Σε μεγαλύτερες κλίμακες επικρατούν οι κυριότερες δομές της εικόνας, αλλά μετακινούνται και θολώνουν τα όριά τους.



Σχήμα 6.2: Πολυκλιμακωτή εξαγωγή χαρακτηριστικών ακμών, χρησιμοποιώντας τον τελεστή \mathcal{G}_γ (6.1) για διαφορετικές τιμές της διασποράς του Gaussian πυρήνα· σκοτεινές περιοχές δείχνουν μία υψηλή τιμή του διαφορικού τελεστή. Η κλίμακα στην οποία μεγιστοποιείται η απόκριση του κανονικοποιημένου τελεστή καθορίζεται από το εύρος της ακμής.

τους· προκύπτουν έτσι όμως τα ερωτήματα σε ποιες κλίμακες θα βρίσκονται τα φίλτρα και πώς θα συνδυάζονται οι αποκρίσεις τους. Τεχνικές μάθησης μηχανών μπορούν να χρησιμοποιηθούν για το στόχο αυτό όπως για παράδειγμα στο κεφάλαιο 3 ή στην αναφορά [154], όμως ουσιαστικά έτσι παρακάμπτεται το πρόβλημα της επιλογής κλίμακας.

Η προσέγγιση που εισηγήθηκε ο D. Lindeberg χρησιμοποιεί τον Gaussian χώρο κλίμακας (scale space) ως ένα αρχικό σύστημα για την πολυκλιμακωτή ανάλυση, πάνω στο οποίο χρησιμοποιούνται κανονικοποιημένοι διαφορικοί τελεστές, όπου η κανονικοποίηση αναιρεί τη μείωση στην έξοδο του τελεστή που προκύπτει από την ομαλοποίηση της εικόνας.

Συγκεκριμένα για ανίχνευση ακμών έχει προταθεί ο ανεξάρτητος κλίμακας τελεστής ανίχνευσης ακμών

$$\mathcal{G}_\gamma = t^\gamma \left(\left(\frac{\partial L_t}{\partial x} \right)^2 + \left(\frac{\partial L_t}{\partial y} \right)^2 \right), \quad \gamma = \frac{1}{2} \quad (6.1)$$

ως ένα μέσο για την αυτόματη επιλογή της κλίμακας στην οποία ανιχνεύεται μία ακμή. Στην παραπάνω σχέση L_t είναι το αποτέλεσμα της συνέλιξης της εικόνας I με ένα Gaussian φίλτρο σταθερής απόκλισης $\sigma = \sqrt{t}$ και ο παράγοντας t^γ χρησιμοποιείται για να κανονικοποιήσει ως προς την κλίμακα την διαφορική αυτή ποσότητα. Αυτό καθιστά τις αποκρίσεις σε διαφορετικές κλίμακες συγκρίσιμες και επιτρέπει την ανίχνευση μέγιστων ως προς την κλίμακα του κανονικοποιημένου μέτρου ισχύος ακμών.

Η συγκεκριμένη τιμή του γ προκύπτει από αναλυτικά μοντέλα ακμών και απαιτώντας το κριτήριο \mathcal{G}_γ να μεγιστοποιείται στην χαρακτηριστική κλίμακα της ακμής. Επίσης, στην αναφορά [148] η συνθήκη για το μέγιστο από κοινού στο χώρο και στην κλίμακα εκφράζεται μέσω ενός συνδυασμού από παραγώγους της ομαλοποιημένης εικόνας, καθιστώντας την ανίχνευσή τους ξεκάθαρη και αποτελεσματική.

Ο μηχανισμός επιλογής κλίμακας καθορίζει έτσι αυτόματα αν μία ακμή είναι απότομη η σταδιακά μεταβαλλόμενη, επιλέγοντας την κλίμακα κάθετα στην κατεύθυνση της ακμής που βελτιστοποιεί τον εντοπισμό της. Στο Σχ. 6.2 βλέπουμε τα αποτελέσματα του κανονικοποιημένου τελεστή ανίχνευσης ακμών σε μία εικόνα που περιέχει ακμές σε πολλαπλές κλίμακες· παρατηρούμε ότι για απότομες μεταβάσεις στη φωτεινότητα οι μικρές κλίμακες επικρατούν, όπως στον λαιμό και στα πόδια, ενώ ομαλές μεταβάσεις δίνουν μέγιστα σε μεγαλύτερες κλίμακες, όπως στην πλάτη και στα βουνα.

6.3.2 Χαρακτηριστικά Ράβδων

Τα χαρακτηριστικά ράβδων συμπληρώνουν τα χαρακτηριστικά ακμών στην ανίχνευση επίμηκων δομών, ιδιαίτερα σε περιοχές με θολές ακμές αλλά έντονες μεταβάσεις μεγάλης κλίμακας στην φωτεινότητα. Διαισθητικά μπορούμε να ερμηνεύσουμε την ανίχνευση καμπυλών τύπου ράβδων ως την εκτίμηση ενός σκελετού της εικόνας χρησιμοποιώντας τις συνεχείς τιμές της φωτεινότητας (gray-scale skeleton) [147].

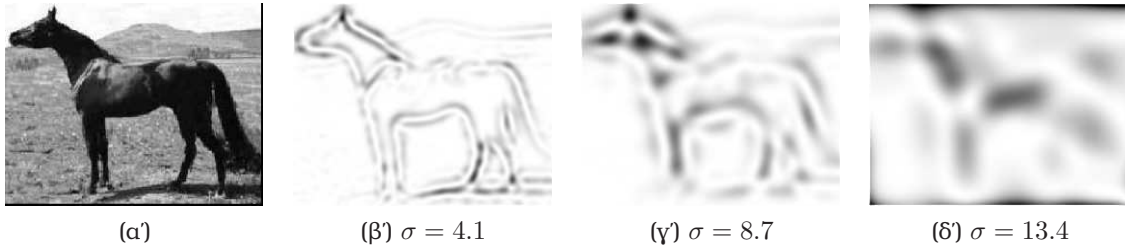
Για το σκοπό αυτό ο υπολογισμός του σκελετού η απλούστερα η προσαρμογή παραλληλογράμμων στην εικόνα συνήθως υλοποιείται μέσω μίας ακολουθίας ανίχνευσης ακμών, αναζήτησης αξόνων συμμετρίας, και συνδυασμού των ακμών όπως π.χ. στην εργασία [105]. Κάθε ένα από αυτά τα στάδια εισάγει σφάλματα και θόρυβο, εξαιτίας της φύσης των λειτουργιών που εμπλέκονται, ενώ απαιτώντας ανεξαρτησία από την κλίμακα θα περιέπλεκε περισσότερο το πρόβλημα.

Αντί να χρησιμοποιείται πρώτα ένα σύστημα ανίχνευσης ακμών στην προσέγγιση του [147] χρησιμοποιούνται οι κανονικοποιημένες ως προς την κλίμακα ιδιοτιμές της Hessian της ομαλοποιημένης εικόνας

$$\begin{aligned} L_{pp} &= \frac{t^\gamma}{2} \left(L_{xx} + L_{yy} - \sqrt{(L_{xx} - L_{yy})^2 + 4(L_{xy})^2} \right) \\ L_{qq} &= \frac{t^\gamma}{2} \left(L_{xx} + L_{yy} + \sqrt{(L_{xx} - L_{yy})^2 + 4(L_{xy})^2} \right) \end{aligned} \quad (6.2)$$

για να εντοπιστούν τα μέγιστα της έντασης των χαρακτηριστικών τύπου ράβδου. Συγκεκριμένα, καθώς η Hessian της ομαλοποιημένης εικόνας χρησιμοποιείται για την τοπική προσέγγισή της μέσω ενός αναπτύγματος Taylor δεύτερης τάξης, οι ιδιοτιμές του μπορούν να χρησιμοποιηθούν για να δώσουν μία συνοπτική περιγραφή της τοπικής δομής της εικόνας: για παράδειγμα δύο θετικές ιδιοτιμές σημαίνουν μία κυρτή επιφάνεια, διαφορετικά πρόσημα μία δομή σάγματος ενώ ίσες ιδιοτιμές αντιστοιχούν σε μία δομή τύπου κηλίδας.

Για τις δομές που ενδιαφερόμαστε τα μέτρα των ιδιοτιμών πρέπει να είναι σημαντικά διαφορετικά μεταξύ τους, ώστε η καμπυλότητα στην μία κατεύθυνση (κάθετα στη ράβδο) να είναι πιο έντονη από την όποια καμπυλότητα στην άλλη (παράλληλα στη ράβδο). Η απαίτηση αυτή μπορεί να διατυπωθεί για παράδειγμα μέσω του κριτηρίου $(L_{pp}^2 - L_{qq}^2)^2$ του οποίου τα μέγιστα ως προς την κλίμακα και το χώρο εν-



Σχήμα 6.3: Πολυκλιμακωτή εξαγωγή ράβδων, χρησιμοποιώντας τον τελεστή \mathcal{A}_γ (6.3) για διαφορετικές τιμές της διασποράς του Gaussian πυρήνα. Η κλίμακα στην οποία μεγιστοποιείται η απόκριση του τελεστή καθορίζεται από το πάχος της δομής τύπου ράβδου.

τοπίζουν δομές τύπου ράβδου· Εναλλακτικά κριτήρια παρουσιάζονται στην αναφορά [147]. Στο σύστημα συντεταγμένων $x - y$ η παραπάνω ποσότητα γράφεται:

$$\mathcal{A}_\gamma = t^{2\gamma} \left(\left(\frac{\partial^2 L_t}{\partial x^2} - \frac{\partial^2 L_t}{\partial y^2} \right)^2 + 4 \frac{\partial^2 L_t}{\partial x \partial y} \right), \quad \gamma = \frac{3}{4} \quad (6.3)$$

όπου πάλι η κανονικοποίηση με $t^{2\gamma}$ καθιστά τη διαδικασία ανίχνευσης χαρακτηριστικών ανεπηρέαστη από την κλίμακα. Από τις αποκρίσεις του τελεστή αυτού στο Σχ. 6.3 βλέπουμε την σχέση της κλίμακας των δομών στην εικόνα με την απόκριση του τελεστή· έχουμε έτσι μεγάλες αποκρίσεις αρχικά στα πόδια και στη σκιά του αλόγου, που είναι λεπτές δομές ενώ αυξάνοντας την κλίμακα εντείνονται οι αποκρίσεις σε δομές όπως ο λαιμός και ο κορμός του.

Τέλος, σημειώνουμε ότι το κριτήριο για την μεγιστοποίηση της ισχύος της ράβδου εκφράζεται χρησιμοποιώντας συνδυασμούς από τελεστές παραγωγίσεως από Gaussian συναρτήσεις, καθιστώντας έτσι εύκολη την εξαγωγή καμπυλών με μέγιστη ισχύ χαρακτηριστικών.

6.3.3 Χαρακτηριστικά Κηλίδων

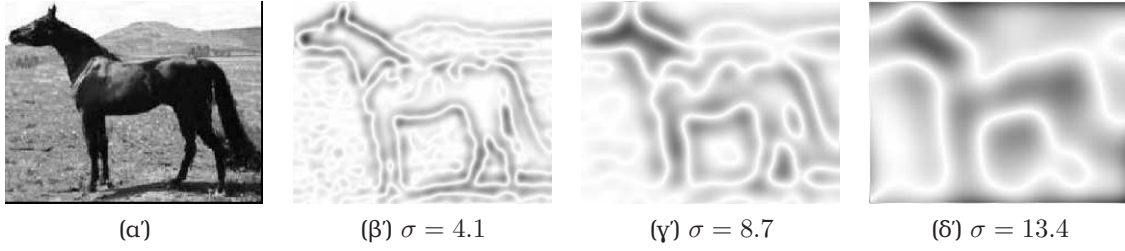
Συμπληρώνουμε την αναπαράστασή μας για την εικόνα χρησιμοποιώντας δομές τύπου κηλίδων που ανιχνεύονται στα μέγιστα χώρου και κλίμακας του κανονικοποιημένου ως προς την κλίμακα τελεστή Laplacian της Gaussian [148]:

$$\mathcal{H}_\gamma = t^\gamma \left(\frac{\partial^2 I_t}{\partial x^2} + \frac{\partial^2 I_t}{\partial u^2} \right)^2 \quad \gamma = 1 \quad (6.4)$$

Ο τελεστής αυτός συνδέεται με τον τελεστή διαφοράς από Gaussians που χρησιμοποιείται για την εξαγωγή χαρακτηριστικών σημείων τύπου SIFT [149]. Τέτοια χαρακτηριστικά αποκρίνονται σε δομές όπως μάτια, ρόδες, σπλές κτλ. και είναι ιδιαίτερα χρήσιμα όταν η κατηγορία δεν έχει κάποια άλλη ξεκάθαρη δομή.

6.3.4 Εξαγωγή Χαρακτηριστικών Ευθείων Γραμμών

Από τη φύση των κριτηρίων που χρησιμοποιούνται τα μέγιστα της ισχύος των χαρακτηριστικών εμφανίζονται κατά μήκος συνεχών καμπυλών, παρέχοντας έτσι κατάλληλες εισόδους για λειτουργίες σύνδεσης σημείων. Αποφεύγοντας περίπλοκες τεχνι-



Σχήμα 6.4: Πολυκλιμακωτή εξαγωγή κηλίδων, με τον τελεστή \mathcal{H}_γ (6.3): σκοτεινές περιοχές δείχνουν μία υψηλή τιμή του διαφορικού τελεστή. Τα μέγιστα εμφανίζονται σε σημεία όπου η εικόνα μπορεί τοπικά να προσεγγιστεί μέσω μίας κυκλικής δομής

κές, όπως αυτές που αναφέρονται στο κεφάλαιο 3 χρησιμοποιήσαμε μία κλασσική μέθοδο [173] για να κατασκευάσουμε αλυσίδες από ακμές και ράβδους, οι οποίες κατόπιν λειτουργούν ως ανεξάρτητες δομές δεδομένων. Καθώς αφενός η παρουσίαση της μεθόδου αυτή θα διέκοπτε τη συνέχεια του κειμένου, αφετέρου η μέθοδος εφαρμόστηκε ως έχει, παραπέμπουμε στην παραπάνω αναφορά για μία παρουσίασή της. Μερικά αποτελέσματα φαίνονται στην πάνω γραμμή του σχήματος Σχ. 6.5, όπου παρατηρούμε ότι παίρνουμε ομαλές καμπύλες που γενικότερα ταιριάζουν με την αντίληψη μας για τις ακμές και τις ράβδους της εικόνας.

Οι αλυσίδες ακμών που κατασκευάζονται στη συνέχεια κατατμούνται σε ευθεία τμήματα εικόνας, χρησιμοποιώντας έναν επαυξητικό αλγόριθμο προσαρμογής ευθειών [66]. Βάσει αυτού του αλγορίθμου ανατίθενται σημεία σταδιακά σε μία ευθεία γραμμή εωσότου ένα κριτήριο για την ποιότητα της ανακατασκευής τους από την ευθεία πέσει κάτω από ένα όριο. Τότε εισάγεται μία νέα ευθεία, η οποία αναλαμβάνει τα υπόλοιπα σημεία κ.ο.κ. εωσότου για όλες τις παρατηρήσεις να υπάρχει και μία ευθεία που να τις ερμηνεύει.

Για να ποσοτικοποιηθεί η ποιότητα ανακατασκευής χρησιμοποιείται το κριτήριο Ολικών Ελάχιστων Τετραγώνων (Total Least Squares-TLS) το οποίο υπολογίζει τις αποστάσεις των σημείων από τη ευθεία που τα προσεγγίζει χρησιμοποιώντας και τις δύο συντεταγμένες τους, και όχι μόνο την y_i , όπως το κριτήριο ελάχιστων τετραγώνων. Τα σημεία που αντιστοιχίζονται σε ένα ευθύγραμμο τμήμα θεωρείται ότι ικανοποιούν προσεγγιστικά μία σχέση της μορφής $ax_i + by_i + c = 0$, όπου (x, y) οι συντεταγμένες του σημείου και a, b, c οι παράμετροι της ευθείας, που υπόκεινται στον περιορισμό $a^2 + b^2 = 1$. Η απόσταση του σημείου x_i, y_i από την ευθεία αυτή ισούται με $ax_i + by_i + c$, οπότε η ελαχιστοποίηση του κριτηρίου TLS υπό τη συνθήκη $a^2 + b^2 = 1$ δίνει τις σχέσεις

$$c = -a\bar{x} - b\bar{y}$$

$$\begin{pmatrix} \bar{x}^2 - \bar{x}^2 & \bar{x}\bar{y} - \bar{x}\bar{y} \\ \bar{x}\bar{y} - \bar{x}\bar{y} & \bar{y}^2 - \bar{y}^2 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \mu \begin{pmatrix} a \\ b \end{pmatrix} \quad (6.5)$$

Από τα δύο ιδιοδιανύσματα του προβλήματος αυτού επιλέγουμε αυτό για το οποίο ελαχιστοποιείται το σφάλμα ανακατασκευής.

Προσαρμόζοντας τη μέθοδο αυτή στην περίπτωση που μας ενδιαφέρει, χρησιμοποιούμε το παρακάτω κριτήριο για να σταματήσουμε την επέκταση μίας γραμμής:

$$C = \frac{1}{l} \sqrt{\frac{1}{N} \sum_{i=1}^N (I_x(i) - M_x(i))^2 + (I_y(i) - M_y(i))^2}, \quad (6.6)$$

όπου N είναι το πλήθος των σημείων της καμπύλης που αντιστοιχίζονται στην ευθεία γραμμή και l το μήκος της καμπύλης. Ο δεύτερος όρος του γινόμενου εκτιμά την διασπορά του σφάλματος ανακατασκευής μεταξύ της πρόβλεψης του μοντέλου ευθείας γραμμής (M_x, M_y) και της τοποθεσίας του χαρακτηριστικού (I_x, I_y) , ποσοτικοποιώντας την ακρίβεια του μοντέλου. Η διαίρεση με το μήκος l μας παρέχει μία ανεξάρτητη κλίμακας ποσότητα, καθώς τόσο το μήκος της γραμμής όσο και η εκτίμηση τις διασποράς είναι ανάλογες της μεγέθυνσης/σμίκρυνσης της εικόνας.

Πέρα από την κλίμακα, τοποθεσία και κατεύθυνση, χρησιμοποιούμε και τον βαθμό επιμήκυνσης των χαρακτηριστικών (η λόγο για συντομία), που εκτιμάται από την τοπική δομή της εικόνας στην κλίμακα των χαρακτηριστικών. Συγκεκριμένα, ο λόγος των απόλυτων των ιδιοτιμών $\frac{L_{pp}}{L_{qq}}$ της Hessian μήτρας της ομαλοποιημένης εικόνας L είναι μία ποσότητα που δεν εξαρτάται την αντίθεση της εικόνας, και εκφράζει την επιμήκυνση των χαρακτηριστικών. Παίρνοντας το μέσο του λόγου αυτού κατά μήκος της καμπύλης στην οποία έχει ταιριαχτεί η ευθεία γραμμή αποκτούμε μία εκτίμηση του πάχους των ράβδων και του εύρους των ακμών.

Επίσης, οι ποσότητες στις σχέσεις 6.1,6.3 μπορούν να χρησιμοποιηθούν για να ποσοτικοποιήσουν την ένταση ενός τμήματος ακμών/ράβδου: παίρνοντας τον μέσο τους κατά μήκος της εξαχθείσας καμπύλης αποκτούμε μία ποσότητα η οποία μας επιτρέπει να προσαρμόσουμε την ευαισθησία του συστήματός μας.

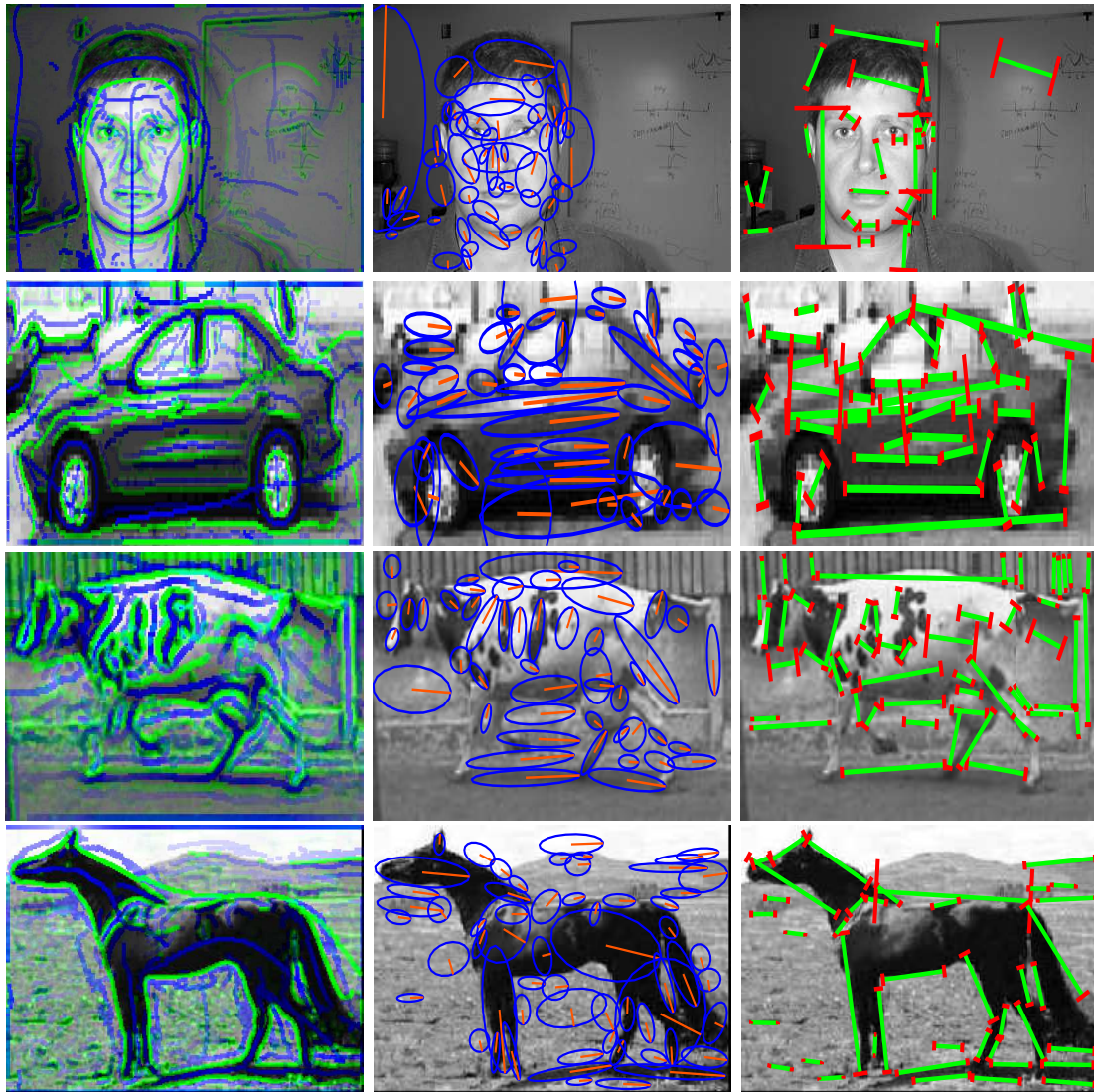
Όπως φαίνεται στο σχήμα Σχ. 6.5 οι καμπύλες και οι ευθείες που εξάγονται σε μεγάλο βαθμό συνοψίζουν την κυρίαρχη δομή της εικόνας, που καθορίζεται τόσο από την κλίμακα όσο και από την αντίθεση των χαρακτηριστικών. Παρατηρούμε όμως ότι σε ορισμένα σημεία η εξαγωγή των χαρακτηριστικών δεν είναι ακριβής· κύριες πηγές σφάλματος είναι οι εξής:

- Η χαμηλή αντίθεση της εικόνας προκαλεί χαμηλή ένταση των χαρακτηριστικών, οπότε για μία σχετικά υψηλή κατωφλίωση αγνοούνται.
- Οι ευθείες γραμμές δεν αρκούν για να περιγράψουν γενικές καμπύλες.
- Ο αλγόριθμος για το συνδυασμό σημείων σε καμπύλες δεν μπορεί να συμπληρώσει χαμένα σημεία, οδηγώντας σε αλυσίδες σπασμένες σε τμήματα.
- Η σχετικά αραιή δειγματοληψία των κλιμάκων που χρησιμοποιείται μπορεί να οδηγήσει σε ανακριβή μέγιστα κλίμακας και συνεπώς σε ανεπιτυχή εξαγωγή ομαλών ευθειών

Τα θέματα αυτά τα αφήνουμε για μελλοντική εργασία, καθώς δεν είναι κριτικής σημασίας για τη λειτουργία του υπόλοιπου συστήματος.

6.4 Κάτωθεν Ανίχνευση Αντικειμένων

Όπως και στο προηγούμενο κεφάλαιο, για να ενεργοποιηθούν τα μοντέλα που χρησιμοποιούν άνωθεν πληροφορία χρησιμοποιούμε ένα απλό και αποδοτικό σύστημα που καταδεικνύει περιοχές της εικόνας όπου θα μπορούσε να βρίσκεται ένα αντικείμενο. Ο ρόλος αυτού του τμήματος μέσα στο όλο σύστημα μπορεί να θεωρηθεί ως ένας μηχανισμός προσοχής, ο οποίος άμεσα εντοπίζει όλα τα αντικείμενα, με το κόστος ενός μεγάλου πλήθους ψευδών ανιχνεύσεων.



Σχήμα 6.5: Χαρακτηριστικά ράβδων και ακμών για τέσσερις κατηγορίες αντικειμένων. Αριστερή στήλη: μέγιστα στο χώρο και στην κλίμακα της έντασης ακμών και ράβδων (πράσινο και γαλάζιο χρώμα, αντίστοιχα). Μέση και δεξιά στήλη: ευθείες ράβδοι και ακμές εξαγόμενες μέσω του ανεξάρτητου- κλίμακας κριτηρίου προσαρμογής ακμών. Το πάχος των ράβδων και η ευρύτητα των ακμών εκτιμώνται από την τοπική δομή της εικόνας και καθορίζουν τους λόγους των αξόνων των ελλείψεων και τα μήκη των κάθετων γραμμών στα άκρα των ακμών, αντίστοιχα.

6.4.1 Κατασκευή μοντέλου

Αναπαράσταση μέσω codebook

Όπως στις αναφορές [3, 144] χρησιμοποιούμε μία αναπαράσταση τύπου codebook για να συνοψίσουμε την ποικιλία στην εμφάνιση του αντικειμένου. Στις προσεγγίσεις αυτές οι καταχωρήσεις του codebook (Codebook-Entries - CEs) κατασκευάζονται μέσω της ομαδοποίησης των χαρακτηριστικών σημείων που εξάγονται από το σύνολο εκπαίδευσης βάσει της φωτεινότητας της περιοχής γύρω από τα σημεία ενδιαφέροντος. Αντίθετα με τις προσεγγίσεις αυτές, δεν χρησιμοποιούμε τη φωτεινότητα καθόλου· θεωρούμε ότι το μεγαλύτερο μέρος της πληροφορίας της εικόνας περιλαμβάνεται στην θέση και στον τύπο του χαρακτηριστικού που ανιχνεύεται. Αυτό μας δίνει ένα πολύ μικρό codebook, με λιγότερες από 70 καταχωρήσεις σε σχέση με πάνω από 500 σε προηγούμενες εργασίες, το οποίο μας επιτρέπει να μάθουμε σχέσεις μεταξύ των καταχωρήσεων του codebook και να μάθουμε τη δομή της κατηγορίας αντικειμένων χρησιμοποιώντας λίγες εικόνες για εκπαίδευση. Αν και ενδεχομένως απλοϊκή, δείχνουμε ότι μία τέτοια αναπαράσταση μπορεί να πετύχει ικανοποιητικά αποτελέσματα ανίχνευσης για διαφορετικές κατηγορίες αντικειμένων.

Για να μάθουμε ένα μοντέλο χρησιμοποιούμε εικόνες όπου το αντικείμενο εμφανίζεται σε σταθερή τοποθεσία και κλίμακα· για κάθε χαρακτηριστικό σημείο εξάγουμε το διάνυσμα θέσης του $\mathcal{K} = (\mathbf{x}, s, r, \theta)^T$ που περιέχει την τοποθεσία $\mathbf{x} = (x, y)$, την κλίμακα s , το βαθμό επιμήκυνσης r και τη γωνία θ . Για να εφαρμόσουμε αλγόριθμους ομαδοποίησης που χρησιμοποιούν Ευκλείδιες αποστάσεις μεταξύ των περιγραφέντων θέσεων τους μετασχηματίζουμε παίρνοντας τον λογάριθμο της κλίμακας σ , χρησιμοποιώντας την εμβύθιση $\theta \rightarrow (\cos(2\theta), \sin(2\theta))$ της κατεύθυνσης στον \mathbb{R}^2 και κατόπιν κανονικοποιούμε τα δεδομένα διαιρώντας τα με την διασπορά τους.

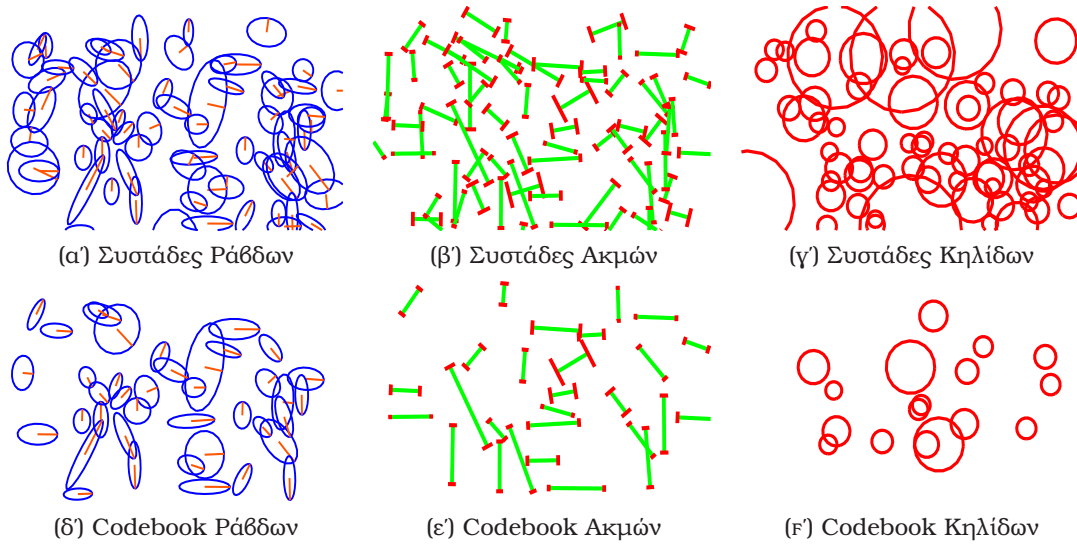
Οι μετασχηματισμένοι περιγραφείς θέσεως ομαδοποιούνται ξεχωριστά για κάθε τύπο χαρακτηριστικών, χρησιμοποιώντας αρχικά τον αλγόριθμο $k - means$ και στη συνέχεια τον EM για να εκτιμηθούν τα κέντρα και οι διασπορές των κατανομών που αντιστοιχούν σε κάθε CE. Εμπειρικά θέσαμε τον αριθμό των συστάδων ίσο με 60 για όλες τις κατηγορίες αντικειμένων, ενώ άλλες τιμές δεν έδωσαν σημαντικά διαφορετικά αποτελέσματα. Για να αποφύγουμε προβλήματα που προκύπτουν όταν ένα μικρό πλήθος παρατηρήσεων είναι διαθέσιμο (π.χ. εκτιμώμενοι πίνακες συνδιασποράς με πολύ μικρή ορίζουσα) θεωρούμε ότι τα στοιχεία του διανύσματος θέσης είναι ανεξάρτητα μεταξύ τους, οπότε η κατανομή P_k που αντιστοιχεί στη συστάδα k γράφεται:

$$P_k(\mathcal{K}) = P_{k,\mathbf{x}}(\mathbf{x})P_{k,s}(s)P_{k,r}(r)P_{k,\theta}(\theta) \quad (6.7)$$

και κάθε ένας από τους όρους του γινομένου μοντελοποιείται ως μία Gaussian κατανομή.

Επιλογή υποσυνόλου codebook

Όπως στην αναφορά [19] ένα υποσύνολο από τις καταχωρήσεις του codebook επιλέγονται βάσει της επίδοσής τους στην ανίχνευση αντικειμένων όταν λειτουργούν μεμονωμένα. Η λειτουργία της ανίχνευσης προσομοιώνεται, όπου η τοποθεσία του αντικειμένου είναι άγνωστη και το περιορισμένο διάνυσμα θέσης $\mathcal{K}'_i = (s_i, r_i, \theta_i)$ ενός εξαγόμενου χαρακτηριστικού χρησιμοποιείται για να αποφασιστεί αν μπορεί να ανατεθεί σε μία καταχώρηση.



Σχήμα 6.6: Επιλογή καταχωρήσεων codebook: από τις συστάδες που προκύπτουν από την ομαδοποίηση των χαρακτηριστικών σημείων επιλέγεται το υποσύνολό που είναι χρησιμότερο για την ανίχνευση του αντικειμένου.

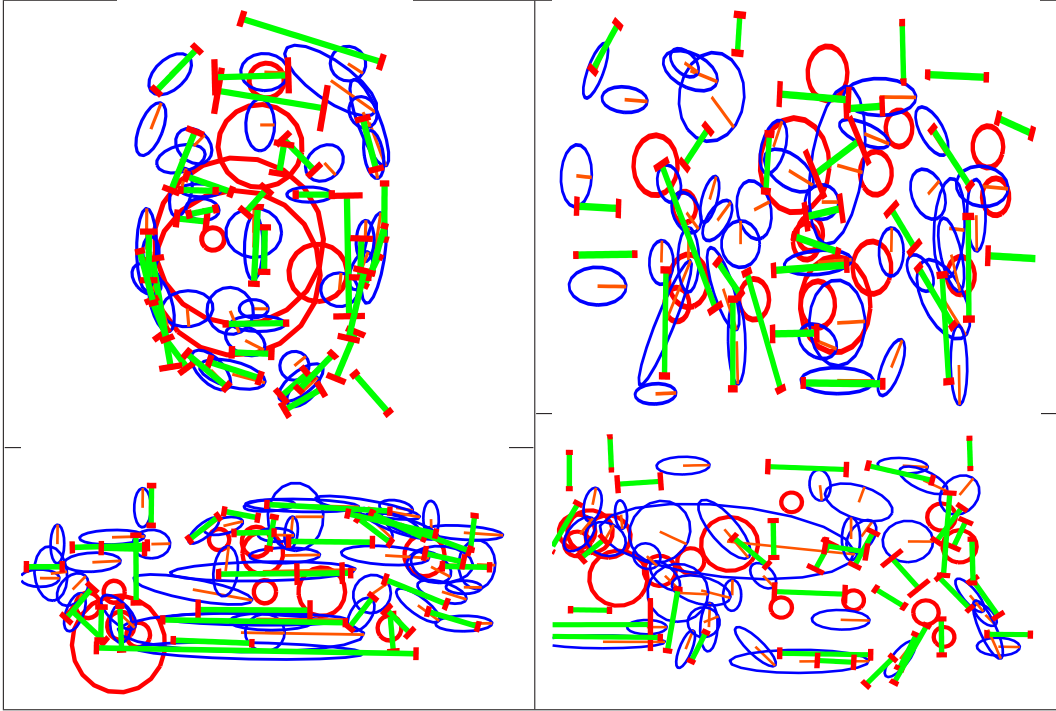
Μια κατανομή P_B κατασκευάζεται για να ληφθούν υπόψη οι στατιστικές ιδιότητες γενικών φυσικών εικόνων και η συμπεριφορά του συστήματος ανίχνευσης σημείων ενδιαφέροντος. Για παράδειγμα, αν το σύστημά μας λειτουργεί μέσα σε μία πόλη οριζόντιες και κάθετες ράβδοι, που αντιστοιχούν π.χ. σε πεζοδρόμια, κτήρια, πεζούσι και δρόμους εμφανίζονται συχνά, δεν είναι οι καλύτερες για την ανίχνευση ενός αντικειμένου όπως μία μοτοσυκλέτα· αντίθετα, μία κυκλική δομή όπως η ρόδα απαντά σχετικά σπανιότερα, οπότε είναι κατάλληλη.

Η κατανομή αυτή κατασκευάζεται παίρνοντας τα χαρακτηριστικά όλων των σημείων ενδιαφέροντος και κατασκευάζοντας μία κατανομή $P_B(\mathcal{K}) = P_{B,s}(s)P_{B,r}(r)P_{B,\theta}(\theta)$ χρησιμοποιώντας μη παραμετρικές κατανομές με Gaussian πυρήνες [104], όπου το εύρος ζώνης τους εκτιμάται μέσω cross-validation. Το χαρακτηριστικό i μπορεί λοιπόν να ανατεθεί στο CE k κατωφλιώνοντας τον λόγο $L_i = \frac{P_{k,s}(s)P_{k,r}(r)P_{k,\theta}(\theta)}{P_B(\mathcal{K}_i)}$ που εμφανίζεται στον νόμο του Bayes. Ο λόγος των πρότερων πιθανοτήτων που θα έπρεπε να εμφανίζεται παραπάνω ενσωματώνεται στο χρησιμοποιούμενο κατώφλι, το οποίο μεταβάλλεται για να βρεθεί η βέλτιστη τιμή του.

Για μία συγκεκριμένη τιμή του κατωφλιού, k από τα χαρακτηριστικά που ανατίθενται σε μία CE προέρχονται από εικόνες που δεν περιέχουν αντικείμενα ή είναι συμπτωματικά κάπου αλλού πάνω στο αντικείμενο, όπως ένα χέιλος που αντιστοιχίζεται σε ένα φρύδι, ενώ τα υπόλοιπα n προέρχονται όντως από το αντίστοιχο CE. Για να καθοριστεί αυτό εισάγεται εκ των υστέρων το πλήρες διάνυσμα θέσης και τα χαρακτηριστικά για τα οποία $P_{k,\mathbf{x}}(\mathbf{x}) > e^{-2}$ διατηρούνται ως πραγματικές ανιχνεύσεις.

Μπορούμε έτσι να συσχετίσουμε σε κάθε CE μία τιμή ακρίβειας (precision) και ανάκλησης (recall) $P = \frac{n}{n+k}$ $R = \frac{n}{N}$, με το N να είναι το πλήθος των αντικειμένων στο σύνολο εκπαίδευσης. Οι τιμές της ακρίβειας και ανάκλησης συνδυάζονται στην F -μέτρηση: $F = \frac{2}{P^{-1}+R^{-1}}$ μεταβάλλοντας το κατώφλι η μέγιστη τιμή της ποσότητας αυτής εκτιμάται και χρησιμοποιείται για την επιλογή των καταχωρήσεων που είναι καταλληλότερες για ανίχνευση, καθορίζοντας συγχρόνως μία τιμή για το κατώφλι (βλ. Σχ. 6.6).

Στο σχήμα Σχ. 6.7 φαίνονται οι καταχωρήσεις που επιλέχθηκαν βάσει της δια-



Σχήμα 6.7: Αναπαραστάσεις για πρόσωπα, άλογα, αυτοκίνητα και αγελάδες που προκύπτουν από τη διαδικασία επιλογής καταχωρήσεων. Οι κόκκινοι κύκλοι συμβολίζουν χαρακτηριστικά κηλίδων, οι πράσινες γραμμές ακμές και οι γαλάζιες ελλείψεις ράβδους.

δικασίας αυτής για τέσσερις κατηγορίες αντικειμένων. Βλέπουμε ότι οι πιο χαρακτηριστικές δομές χρησιμοποιούνται στην αναπαράσταση, ενώ συχνά αρκετές καταχωρήσεις αντιστοιχούν σε διαφορετικές θέσεις του ίδιου τμήματος του αντικειμένου, όπως τα πόδια του αλόγου.

6.4.2 Παραγωγή Υποθέσεων Αντικειμένων

Οι αποφάσεις των μεμονωμένων καταχωρήσεων μπορούν να συνδυαστούν χρησιμοποιώντας ένα πιθανοτικό μοντέλο το οποίο λαμβάνει υπόψη την από κοινού κατανομή τους. Το απλούστερο τέτοιο μοντέλο δεν θεωρεί καθόλου τις μεταξύ τους αλληλεπιδράσεις, οπότε συνδέονται όλες οι καταχωρήσεις απευθείας με τον κόμβο που αναπαριστά την τοποθεσία του αντικειμένου O . Για μία συγκεκριμένη υποψήφια τοποθεσία αντικειμένου η ανίχνευση βασίζεται στον λόγο $R = \frac{P(I|O_{x,y})}{P(I|B)}$ των πιθανοτήτων των υποθέσεων η εικόνα I να προέρχεται από ένα αντικείμενο O στην τοποθεσία (x, y) και την γενική υπόθεση (background) B . Για μία δεδομένη ανάθεση \mathcal{H} των χαρακτηριστικών σε CEs το συγκεκριμένο μοντέλο που χρησιμοποιούμε επιτρέπει την παραγοντοποίηση του όρου πιθανοφάνειας πάνω στις μεμονωμένες καταχωρήσεις οπότε μπορούμε να γράψουμε:

$$R_{\mathcal{H}} = \prod_{k \in H} \frac{P(f_k^{\mathcal{H}} | h_k, O_{x,y}) P_k^O}{P(f_k^{\mathcal{H}} | h_k, B) P_k^B} \prod_{k \in M} \frac{1 - P_k^O}{1 - P_k^B}. \quad (6.8)$$

Παραπάνω H είναι το υποσύνολο των καταχωρήσεων που έχουν βρει μία αντιστοιχισή, M τα υπόλοιπα, $f_k^{\mathcal{H}}$ τα διάνυσμα θέσης για το χαρακτηριστικό που ανατίθεται στην καταχώρηση k από το \mathcal{H} και $P_k^{\{O,B\}}$ είναι η πρότερη πιθανότητα η καταχώρηση

k να βρει μία αντιστοίχιση σε εικόνες που περιέχουν το αντικείμενο ή αυτό απουσιάζει, αντίστοιχα. Η εφαρμογή μίας τέτοιας έκφρασης πρακτικά εμποδίζεται από την ανάγκη να υπολογιστεί το άθροισμα πάνω σε όλες τις δυνατές αναθέσεις \mathcal{H} για να εντοπισθούν οι περιοχές υψηλού ενδιαφέροντος. Αντί για αυτό, αντιστρέφουμε την άνωθεν δομή της εξάρτησης, και χρησιμοποιούμε τα ανιχνευθέντα σημεία για να προτείνουν περιοχές της εικόνας που θα μπορούσαν να περιέχουν το αντικείμενο.

Για μία εικόνα εξάγονται τα σημεία ενδιαφέροντος και τα διανύσματα θέσης τους αντιστοιχίζονται στις καταχωρήσεις του μοντέλου, όπως περιγράφεται στην προηγούμενη παράγραφο· για κάθε καταχώρηση που χρησιμοποιείται στην αναπαράσταση του αντικειμένου αποφασίζεται ποια χαρακτηριστικά μπορούν να της ανατεθούν χρησιμοποιώντας την τιμή κατωφλίου που δίνει την καλύτερη μέτρηση F .

Στη συνέχεια, για κάθε δυνατή αντιστοίχιση ενός χαρακτηριστικού f σε μία καταχώρηση k , η πλήρης κατανομή $P_k(f)$ χρησιμοποιείται για να καθοριστούν οι τοποθεσίες της εικόνας (x, y) για τις οποίες ο λόγος $\frac{P(f_k|h_k, O_{x,y})P(h_k|O_{x,y})}{P(f_k|h_k, B)P(h_k|B)}$ είναι μεγαλύτερος της μονάδας, δηλαδή να βρεθεί σε ποια σημεία της εικόνας μπορούμε να θεωρήσουμε ότι ένα αντικείμενο στη θέση (x, y) παρήγαγε το χαρακτηριστικό f .

Η μαρτυρία υπέρ ενός αντικειμένου στα σημεία αυτά πολλαπλασιάζεται με τον παράγοντα $\frac{P(f_k|h_k, O_{x,y})P(h_k|O_{x,y})}{P(f_k|h_k, B)P(h_k|B)}$ και οι δυνατές τοποθεσίες αντικειμένων προτείνονται στις τοπικά μέγιστες αποκρίσεις που εντοπίζονται συσσωρεύοντας τις συνεισφορές από όλες τις δυνατές αντιστοιχίσεις.

Η μέθοδος ανίχνευσης αυτή είναι παρόμοια με αυτήν της αναφοράς [144], με τη διαφορά ότι οι Gaussian κατανομές που χρησιμοποιούμε για τις τοποθεσίες των τμημάτων μας απαλλάσσουν από τον αλγόριθμο ομαδοποίησης τύπου Mean-Shift που χρησιμοποιείται εκεί, ενώ χρησιμοποιώντας ένα αναγεννητικό μοντέλο για τα χαρακτηριστικά μπορούμε να κατασκευάσουμε εκφράσεις πιθανοφάνειας. Συγκεκριμένα, καθώς ο περιορισμός ότι ένα χαρακτηριστικό αντιστοιχίζεται σε μία μόνο καταχώρηση δεν ικανοποιείται υποχρεωτικά κατά τη διάρκεια της ανίχνευσης, μπορεί η ίδια πληροφορία να μετρηθεί ξανά οδηγώντας σε υποβέλτιστες αποφάσεις. Αυτή η προσέγγιση εξασφαλίζει την ταχύτητα με την οποία εκτελείται αυτό το στάδιο, αλλά μπορεί εύκολα να βελτιωθεί μόλις το μοντέλο του αντικειμένου ενεργοποιηθεί στην προτεινόμενη τοποθεσία. Αντιστοιχίζοντας το κάθε χαρακτηριστικό σε μία μόνο καταχώρηση, αυτήν για την οποία επιτυγχάνεται η μέγιστη αύξηση στο κριτήριο (6.8), δίνει άμεσα μία αύξηση στην απόδοση, όπως φαίνεται αργότερα στην παράγραφο των πειραματικών αποτελεσμάτων.

6.5 Εισαγωγή Άνωθεν Πληροφορίας

Οι υποθέσεις για αντικείμενα που παράγονται τυπικά βασίζονται σε ένα μικρό σύνολο από ανιχνευθείσες καταχωρήσεις του codebook, με τις υπόλοιπες να θεωρούνται ότι χάθηκαν. Αυτό οφείλεται κυρίως στους παρακάτω παράγοντες:

- Ένα codebook εγγενώς παρέχει μία υπερπλήρη αναπαράσταση.
- Η διαδικασία ανίχνευσης χαρακτηριστικών εμφανίζει από μόνη της προβλήματα, όπως αναλύθηκε στην ενότητα 6.3, με αποτέλεσμα ορισμένα χαρακτηριστικά να μην ανιχνεύονται.
- Τα εξαγόμενα χαρακτηριστικά μπορεί να μην βρίσκουν κάποια καταχώρηση στο codebook που να τα μοντελοποιεί επαρκώς.

Τα δύο τελευταία προβλήματα μπορούν να αντιμετωπιστούν χρησιμοποιώντας άνωθεν πληροφορία για να οδηγήσουν την αναζήτηση για τα χαμένα χαρακτηριστικά: οι καταχωρήσεις που έχουν ανιχνευθεί μπορούν να προτείνουν δυνατές τοποθεσίες για αυτές που έχουν χαθεί, οι οποίες στη συνέχεια αποτιμώνται με βάση την υπάρχουσα πληροφορία που υπάρχει στην εικόνα.

Η χρήση αυτής της ιδέας έχει αρχικά επιδιωχθεί στις αναφορές [19, 144], χρησιμοποιώντας τμήματα από δεδομένα εκπαίδευσης (exemplars) αντί ενός αναγεννητικού μοντέλου· διατυπώνοντας το πρόβλημα στο πλαίσιο των αναγεννητικών μοντέλων ελαττώνει το υπολογιστικό κόστος και την πολυπλοκότητα αναπαράστασης της άνωθεν πληροφορίας, καθώς δεν χρειάζονται τμήματα/περιγράμματα για μία συγκεκριμένη κατηγορία αντικειμένων. Στην αναφορά [91] μία συντακτική προσέγγιση χρησιμοποιείται σε ένα παρόμοιο πλαίσιο, αλλά εκεί χρησιμοποιείται ένα στατικό σύνολο από περιγραφείς τύπου *primal-sketch* [88], ενώ η εργασία εστιάζει στην ερμηνεία γενικής φύσης δομών, παρά στην ανίχνευση αντικειμένων.

Εδώ διατυπώνουμε το πρόβλημα στο πλαίσιο των γραφικών μοντέλων και εστιάζουμε στην έκφραση και αποτίμηση των δυναμικών παρατήρησης. Χρησιμοποιώντας τη σχέση $\Phi_i(\mathcal{P}_i) = P(I|\mathcal{P}_i)$ για να σχετίσουμε την πιθανοφάνεια των παρατηρήσεων I δεδομένης της θέσης του \mathcal{P}_i -στού κόμβου, η βέλτιστη διάταξη $\mathcal{P} = (\mathcal{P}_1, \dots, \mathcal{P}_N)$ μπορεί να βρεθεί ως το μέγιστο της ποσότητας:

$$P(I|O) = \prod_i \Phi_i(\mathcal{P}_i) \prod_{(i,j) \in \mathcal{C}} \Psi_{i,j}(\mathcal{P}_i, \mathcal{P}_j) \quad (6.9)$$

όπου ο δεύτερος όρος είναι πάνω στις κλίκες της κατανομής του γράφου και για το δίκτυο που χρησιμοποιούμε εδώ συνδέει απευθείας τη θέση του τμήματος με τη θέση του αντικειμένου: $\Psi_{i,j}(\mathcal{P}_i, \mathcal{P}_j) = P(\mathcal{P}_i|\mathcal{P}_j)$ όπου \mathcal{P}_j είναι η θέση του κόμβου που αντιστοιχεί στο κέντρο του αντικειμένου.

Η απλή προσέγγιση στο πρόβλημα είναι η χρήση του αλγόριθμου Belief Propagation: τα μηνύματα από τα γνωστά σημεία ενδιαφέροντος δεσμεύουν τις θέσεις των χαμένων και σε συνδυασμό με τις παρατηρήσεις της εικόνας μπορούν να ορίσουν τις ύστερες κατανομές τους [58]. Το πρακτικό εμπόδιο βρίσκεται στην εκτίμηση των αθροισμάτων

$$m_{i,j}(\mathcal{P}_j) = \sum_{\mathcal{P}_i} \Phi_i(\mathcal{P}_i) \Psi_{i,j}(\mathcal{P}_i, \mathcal{P}_j) \prod_{k \in \{\mathcal{N}(i) \setminus j\}} m_{k,i}(\mathcal{P}_k, \mathcal{P}_i) \quad (6.10)$$

στα μηνύματα μεταξύ των γειτόνων. Με λόγια η έκφραση αυτή ζητάει για κάθε τιμή του διανύσματος θέσης \mathcal{P}_j του 'δέκτη' j να υπολογιστεί το κατά πόσο είναι συμβατή με τη θέση του αποστολέα i (όρος $\Psi_{i,j}(\mathcal{P}_i, \mathcal{P}_j)$) και την πληροφορία που έχει αυτός από την εικόνα (όρος $\Phi_i(\mathcal{P}_i)$) και τους υπόλοιπους κόμβους (όρος $\prod_{k \in \{\mathcal{N}(i) \setminus j\}} m_{k,i}(\mathcal{P}_k, \mathcal{P}_i)$) για όλες τις δυνατές κλίμακες, τοποθεσίες, γωνίες και λόγους. Ο χώρος των δυνατών διανυσμάτων θέσης είναι τυπικά τεράστιος, και ακόμα και όταν αποτελεσματικοί αλγόριθμοι χρησιμοποιηθούν για την επίσπευση των αθροισμάτων ο αλγόριθμος παραμένει αργός.

Μία προσέγγιση που είναι κάπου στο ενδιάμεσο είναι αυτή της εργασίας [135], όπου ένα σύνολο χαρακτηριστικών εξάγεται αρχικά, και μετά αντί για εξαντλητική αναζήτηση ο αλγόριθμος belief propagation χρησιμοποιείται για να προτείνει τις πιο πιθανές τοποθεσίες για τα αντικείμενα. Αν και μειώνει το υπολογιστικό φόρτο, η μέθοδος αυτή δεν είναι σε θέση να ανακάμψει από τις απώλειες της διαδικασίας

ανίχνευσης χαρακτηριστικών.

Στην περίπτωση μας, αφενός γνωρίζουμε εξαρχής τα χαρακτηριστικά εκείνα που οδήγησαν στην θέση του αντικειμένου που εξετάζεται· αυτό οδηγεί σε κλειστές εκφράσεις για τα μηνύματα που στέλνονται σε γειτονικούς κόμβους, καθώς χρησιμοποιούνται Gaussian συναρτήσεις δυαδικού δυναμικού και μία κατανομή Dirac χρησιμοποιείται για το δυναμικό παρατήρησης στους γνωστούς κόμβους. Αφετέρου, το άθροισμα πάνω σε όλες τις δυνατές τοποθεσίες των απολεσθέντων χαρακτηριστικών σημείων είναι υπολογιστικά απαιτητικό, καθώς απαιτεί βασικά την εξαγωγή ενός όρου πιθανοφάνειας για κάθε δυνατή τιμή της τοποθεσίας του απολεσθέντος σημείου. Μπορούμε να επιταχύνουμε αυτή τη διαδικασία παρατηρώντας ότι η ποσότητα που αθροίζεται θα πάρει τιμές διαφορετικές από το μηδέν μόνο σε περιοχές που τα εισερχόμενα μηνύματα είναι πάνω από το μηδέν. Αυτή είναι μία περίπτωση που ταιριάζει με την εφαρμογή τεχνικών τύπου Monte Carlo / particle filtering [169, 13], που αποφεύγουν την εκτίμηση της συνάρτησης πιθανοφάνειας πάνω σε όλο τον χώρο τοποθεσίας.

Αντί για μία προσέγγιση που βασίζεται στο particle filtering σε όλα τα στάδια της [204, 106, 209], προσεγγίζουμε ποσότητες της μορφής $\sum_{\mathcal{P}_i} \Phi_i(\mathcal{P}_i) \prod_j m_{i,j}(\mathcal{P}_i)$ ως εξής (βλ. Σχ. 6.8):

- Υπολογίζουμε αναλυτικά την ποσότητα $P_{in}(\mathcal{P}_i) = \prod_j m_{i,j}(\mathcal{P}_i)$ χρησιμοποιώντας κλειστές εκφράσεις.
- Παίρνουμε N δείγματα S_n από το $P_{in}(\mathcal{P}_i)$
- Σχηματίζουμε την προσέγγιση Monte Carlo:

$$\sum_{\mathcal{P}_i} \Phi_i(\mathcal{P}_i) \prod_j m_{i,j}(\mathcal{P}_i) \simeq \sum_{i=1}^N l_n \delta(P - S_n) \quad (6.11)$$

χρησιμοποιώντας συναρτήσεις Dirac γύρω από τα δείγματα και τις πιθανοφάνειες $l_n = \Phi(S_n) = P(I|S_n)$ ως βάση.

- Προσεγγίζουμε την κατανομή μέσω της Gaussian που ελαχιστοποιεί την απόκλιση Kullback Leibler από την προσέγγιση Monte Carlo,

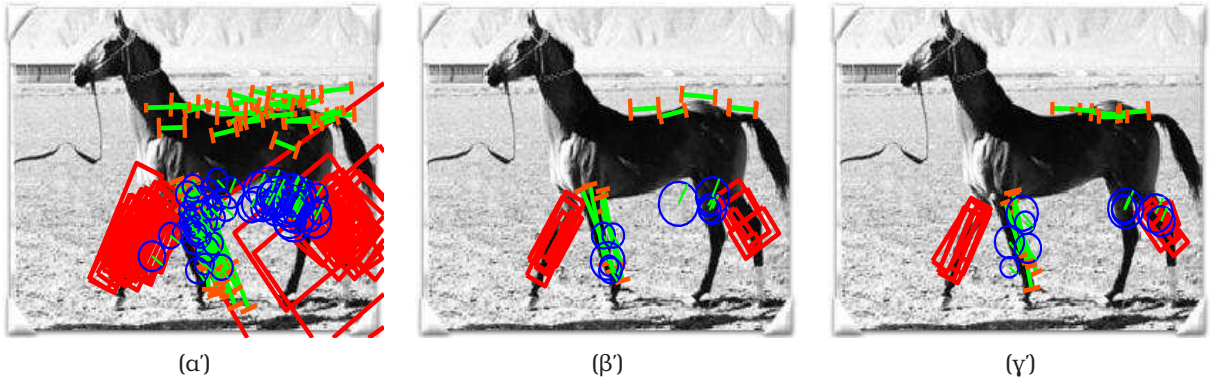
$$- \sum_n l_n \log \frac{P(S_n)}{l_n} \quad (6.12)$$

Αυτό απλά σημαίνει ότι χρησιμοποιούμε τους σταθμισμένους μέσους και διασπορές

$$\mu_i = \frac{\sum_i l_i S_i}{\sum_i l_i}, \quad \Sigma_i = \frac{\sum_i l_i (S_i - \mu_i)(S_i - \mu_i)^T}{\sum_i l_i} \quad (6.13)$$

- Εκτιμούμε αναλυτικά τα μηνύματα που στέλνονται στους γείτονες

Πέρα από το τρίτο βήμα, όλες οι άλλες λειτουργίες μπορούν να υλοποιηθούν αποτελεσματικά. Στην παρακάτω παράγραφο προτείνουμε μία νέα μέθοδο για τον αποτελεσματικό υπολογισμό των όρων πιθανοφάνειας, που επιτρέπει την πρακτική εφαρμογή της προσέγγισης αυτής.



Σχήμα 6.8: Μηχανισμός εκτίμησης του $\sum_{\mathcal{P}_i} \Phi_i(\mathcal{P}_i) \prod_j m_{i,j}(\mathcal{P}_i)$ για έξι διαφορετικούς κόμβους: Ράβδοι στο μπροστά και πίσω πόδι, ακμές στην πλάτη και στο δεύτερο πόδι και κηλίδες σε δύο διακλαδώσεις). (α) Δείγματα από το μήνυμα που έρχεται από το υπόλοιπο γραφικό δίκτυο $\prod_j m_{i,j}(\mathcal{P}_i)$. (β) Πέντε ισχυρότερα δείγματα, βάσει του όρου $\Phi_i(\mathcal{P}_i)$ για (γ) Δείγματα από την κατανομή $\sum_{\mathcal{P}_i} \Phi_i(\mathcal{P}_i) \prod_j m_{i,j}(\mathcal{P}_i)$ που προκύπτει προσεγγίζοντας την $\sum_{\mathcal{P}_i} \Phi_i(\mathcal{P}_i) \prod_j m_{i,j}(\mathcal{P}_i) \simeq \sum_{i=1}^N l_n \delta(P - S_n)$ με μία Gaussian. Ο υπολογισμός του αθροίσματος πάνω σε όλα τα \mathcal{P}_i αποφεύγεται, εξετάζοντας μόνο δείγματα από την υποπεριοχή του χώρου θέσης που 'προτείνει' το γραφικό δίκτυο.

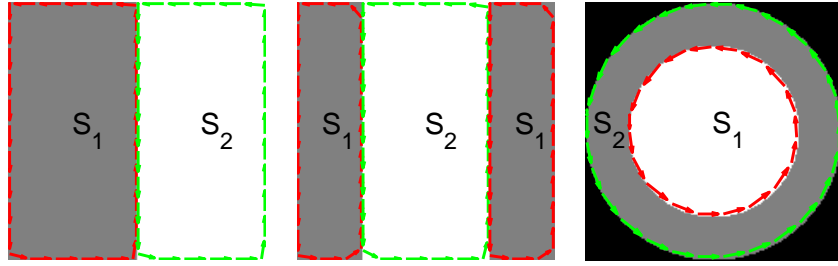
6.5.1 Αποτελεσματική Εκτίμηση της πιθανοφάνειας

Για την εκτίμηση της ποσότητας $l_n = P(I|S_n)$ υπάρχουν δύο επιλογές:

- Να εξάγουμε κατά έναν ευρετικό τρόπο χαρακτηριστικά \mathbf{F} που να υποδεικνύουν μία υποπεριοχή του χώρου τοποθεσίας ως πιο πιθανή και να χρησιμοποιήσουμε την προσέγγιση $P(I|S_n) \simeq P(\mathbf{F}|S_n)$. Η επιλογή αυτή απαιτεί την εξαγωγή διαφορετικών χαρακτηριστικών για κάθε τύπο υπόθεσης, σε πολλαπλές κλίμακες, κατευθύνσεις κ.ο.κ. ενώ βασίζεται στην εκμάθηση μίας κατανομής $P(\mathbf{F}|S_n)$ χρησιμοποιώντας δεδομένα εκπαίδευσης.
- Να ορίσουμε ένα αναγεννητικό μοντέλο, να ανακατασκευάσουμε την εικόνα μόνο στα σημεία που προτείνονται από τις υποθέσεις S_n , να εκτιμήσουμε το σφάλμα ανακατασκευής και να εξάγουμε μία αντίστοιχη έκφραση πιθανοφάνειας. Το κύριο πρόβλημα εδώ είναι αφενός η εκτίμηση των παραμέτρων του αναγεννητικού μοντέλου, αφετέρου ο υπολογισμός των όρων πιθανοφάνειας.

Η συνεισφορά μας έγκειται στο να καταστήσουμε δυνατή την εφαρμογή της δεύτερης επιλογής. Συγκεκριμένα, κατασκευάζουμε τις εκφράσεις για την πιθανοφάνεια των παρατηρήσεων χρησιμοποιώντας ένα απλό πρότυπο (τεμπλέτα) για κάθε τύπο χαρακτηριστικών, που μοντελοποιεί την εικόνα γύρω από το τμήμα του χαρακτηριστικού μέσω δύο (ή γενικότερα K) σταθερών τιμών έντασης, όπως φαίνεται στο σχήμα Σχ. 6.9.

Το αναγεννητικό μοντέλο που χρησιμοποιούμε θεωρεί ότι για τα σύνολο σημείων T_{S_n} στο εσωτερικό της τεμπλέτας που αντιστοιχεί στην υπόθεση S_n η εικόνα προκύπτει αλλοιώνοντας την τεμπλέτα με λευκό Gaussian θόρυβο. Δεδομένων των παραμέτρων του μοντέλου, F , δηλαδή τις σταθερές τιμές φωτεινότητας στο εσωτερικό των τμημάτων της τεμπλέτας και τη διασπορά σ της διαδικασίας θορύβου, μπορούμε



Σχήμα 6.9: Τεμπλέτες για χαρακτηριστικά τύπου (α) ακμών (β) ράβδων και (γ) κηλίδων, και εφαπτόμενα διανύσματα στα σύνορα των υπομημάτων της τεμπλέτας. Επιφανειακά ολοκληρώματα πάνω στα υπομήματα μπορούν να αντικατασταθούν με επικαμπύλια, ελαττώνοντας τον χρόνο υπολογισμού από $O(N^2)$ σε $O(N)$, όπου N η μία διάσταση της τεμπλέτας. Η πυκνότητα των διανυσμάτων είναι ανάλογη του βήματος ολοκλήρωσης και επιτρέπει τον έλεγχο του συμβιβασμού ταχύτητας/ακρίβειας

να γράψουμε την πιθανοφάνεια των παρατηρήσεων ως :

$$P(I|S_n) = \prod_{i \in T_{S_n}} P(I_i|F), \quad P(I_i|F) = N(I_i - c_i, \sigma) \quad (6.14)$$

όπου c_i είναι η πρόβλεψη του μοντέλου της τεμπλέτας για το σημείο i . Παίρνοντας τον λογάριθμο της παραπάνω έκφρασης και θέτοντας την παράγωγό του ίση με μηδέν μας δίνει τις εκτιμήτριες μέγιστης πιθανοφάνειας :

$$c_k = \frac{\sum_{i \in S_k} I_i}{|S_k|}, \quad k = 1 \dots K \quad (6.15)$$

$$\sigma^2 = \frac{\sum_{k=1}^K \sum_{i \in S_k} (I_i - c_k)^2}{\sum_{k=1}^K |S_k|} = \frac{\sum_i I_i^2 - \sum_{k=1}^K c_k^2 |S_k|}{\sum_{k=1}^K |S_k|} \quad (6.16)$$

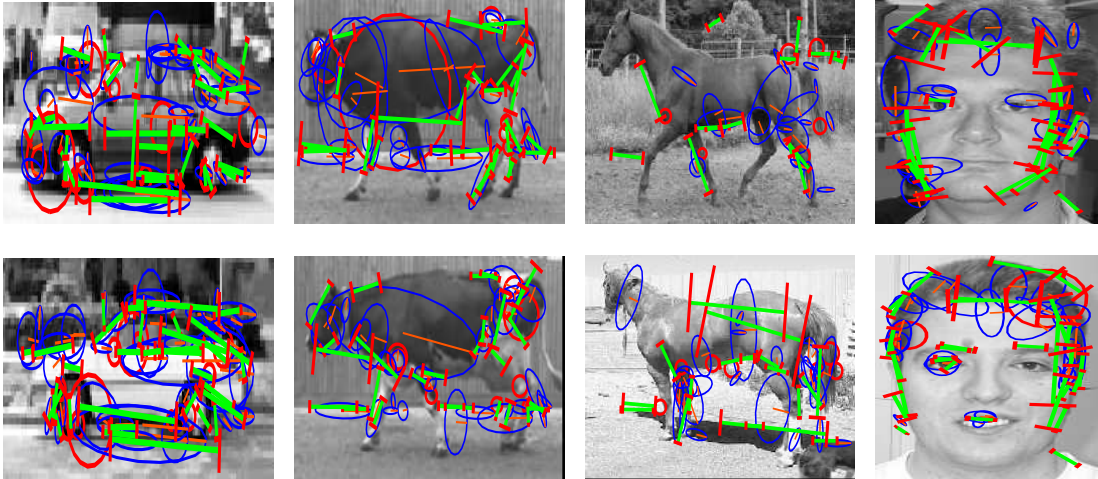
όπου K είναι το πλήθος των σταθερών υπομοντέλων, S_k είναι το σύνολο των παρατηρήσεων και το $|\cdot|$ δηλώνει την πληθικότητα ενός συνόλου.

Όλα τα αθροίσματα στις παραπάνω εκφράσεις μπορούν να θεωρηθούν ως διακριτοποιήσεις επιφανειακών ολοκληρωμάτων των συναρτήσεων $I(x, y)$, $I^2(x, y)$, $I(x, y)$ πάνω στα πεδία που αντιστοιχούν στα υπομήματα της τεμπλέτας. Η μέθοδός μας παρακάμπτει τα επιφανειακά ολοκληρώματα χρησιμοποιώντας το θεώρημα του Stokes:

$$\iint_S f(x, y) dx dy = \int_{\partial S} P dx + Q dy = \int_0^l (P, Q) \cdot T ds. \quad (6.17)$$

T είναι το μοναδιαίας νόρμας εφαπτόμενο διάνυσμα στην καμπύλη, l το μήκος της καμπύλης, ενώ Q και P είναι ένα ζεύγος συναρτήσεων τέτοιο ώστε

$$\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} = f(x, y) \quad (6.18)$$



Σχήμα 6.10: Σημεία τύπου ράβδων, ακμών και κηλίδων που ανιχνεύονται χρησιμοποιώντας άνωθεν γνώση, καταφλιώνοντας την πιθανότητα παρατήρησης του τμήματος του αντικειμένου.

. Ένα προφανές ζεύγος είναι οι ' ολοκληρωτικές εικόνες' [227]

$$Q(x, y) = \frac{1}{2} \int_0^x f(x, y) dx, \quad P(x, y) = -\frac{1}{2} \int_0^y f(x, y) dy \quad (6.19)$$

Καθώς όλες οι ποσότητες που εμπλέκονται στις εκφράσεις πιθανοφάνειας μπορούν να αντικατασταθούν από τις αντίστοιχες επικαμπύλιες εκφράσεις, αυτό μας επιτρέπει να υπολογίζουμε ταχύτατα την πιθανοφάνεια κάθε παραχθείσας θέσης χρησιμοποιώντας αθροίσματα πάνω στα σύνορα των υποτιμημάτων της τεμπλέτας αντί για πάνω στο χωρίο της. Πέρα από αυτό, το διάστημα μεταξύ των σημείων πάνω στα οποία γίνεται η άθροιση για να προσεγγιστεί το επικαμπύλιο ολοκλήρωμα μπορεί να ρυθμιστεί κατά βούληση, επιτρέποντας τον έλεγχο του ισοζυγίου μεταξύ ταχύτητας και ακρίβειας.

Μερικά ενδεικτικά αποτελέσματα φαίνονται στο Σχ. 6.10: χρησιμοποιώντας την άνωθεν γνώση που μας παρέχει η κατανομή του γραφικού δικτύου πληροφορία μπορούμε να ανιχνεύσουμε τμήματα του αντικειμένου που αρχικά χάνονται από το κάτωθεν σύστημα λόγω χαμηλής έντασης.

Πρακτικά σε περιοχές χωρίς έντονα χαρακτηριστικά παρατηρούνται χαμηλά λάθη ανακατασκευής και συνεπώς μεγάλα βάρη στη σχέση (6.11). Η μία λύση θα ήταν να χρησιμοποιηθεί κάποια πρότερη κατανομή πάνω στις παραμέτρους των μοντέλων της τεμπλέτας, ώστε να τιμωρούνται ομαλές περιοχές. Καθώς όμως αυτή η προσέγγιση θα απαιτούσε την ευρετική κατασκευή μίας τέτοιας κατανομής, χρησιμοποιούμε μία ιδέα αντίστοιχη αυτής που βασιστήκαμε στο κεφάλαιο της υφής. Συγκεκριμένα, θεωρούμε μία 'μηδενική υπόθεση' που μοντελοποιεί την φωτεινότητα μέσω ενός σταθερού σήματος, ώστε να ελαττώνει το βάρος των χαρακτηριστικών σε ομαλές περιοχές. Έτσι ως βάρος αντί του $P(I|S_i)$ χρησιμοποιούμε την ποσότητα $p_i = \frac{P(I|S_i)}{P(I|S_i)+P(I|C)}$, όπου $P(I|C)$ είναι η πιθανοφάνεια των παρατηρήσεων υπό το συμπληρωματικό μοντέλο.

6.6 Κατασκευή Γραφικών Μοντέλων

Βάσει του απλοϊκού γραφικού μοντέλου που χρησιμοποιούμε μέχρι τώρα θεωρείται ότι η θέση της κάθε καταχώρησης εξαρτάται μόνο από την θέση του αντικειμένου· αυτό μπορεί να είναι ένα εμπόδιο για την κατασκευή ακριβέστερων μοντέλων, ιδιαίτερα για αρθρωτά αντικείμενα, όπου η τύπου δέντρου δομή των αλληλεξαρτήσεων μένει ανεκμετάλλευτη [105, 204, 58]. Κατά τη διάρκεια της κατασκευής ενός μοντέλου χρησιμοποιώντας μόνο την κάτωθεν ροή πληροφορίας η αναζήτηση τέτοιων αλληλεξαρτήσεων είναι ανέφικτη, εκτός εάν ένα μεγάλο πλήθος δεδομένων είναι διαθέσιμο: τα περισσότερα ζεύγη, τριάδες, κ.ο.κ. καταχωρήσεων δεν θα είναι συγχρόνως ενεργά, καθιστώντας την κατασκευή συναρτήσεων από κοινού κατανομής προβληματική.

Χρησιμοποιώντας όμως την άνωθεν μέθοδο συμπλήρωσης μπορούμε να ανακτήσουμε τα περισσότερα από τα τμήματα που έχουν χαθεί και έτσι να εμπλουτίσουμε το αρχικό σύνολο παρατηρήσεών μας. Επίσης, οι ομαλές πιθανότητες παρατήρησης για τα τμήματα του αντικειμένου μπορούν να ενσωματωθούν στην εκτίμηση των δυναμικών ζεύγους και την επιλογή των κλικών κατά την εκμάθηση δομής.

Για κατανομές με δομή δέντρου όπου μόνο δυαδικές σχέσεις λαμβάνονται υπόψη, ο αλγόριθμος ελάχιστου εκτεινόμενου δέντρου Minimum Spanning Tree - MST μπορεί να χρησιμοποιηθεί για να βρει τον βέλτιστο γράφο που συνδέει όλους τους κόμβους, όπως στις αναφορές [33, 58]. Συγκεκριμένα, χρησιμοποιώντας Gaussian συναρτήσεις για τις μεμονωμένες και από κοινού κατανομές, το κέρδος στον λογάριθμο της πιθανοφάνειας που επιτυγχάνεται θεωρώντας την από κοινού κατανομή των κόμβων i και j είναι:

$$C_{i,j} = \sum_{(\mathcal{P}_i, \mathcal{P}_j) \in \mathcal{P}_{ij}} \log \frac{P(\mathcal{P}_i, \mathcal{P}_j)}{P(\mathcal{P}_i)P(\mathcal{P}_j)} = cN \log \frac{|\Sigma_i||\Sigma_j|}{|\Sigma_{i,j}|} \quad (6.20)$$

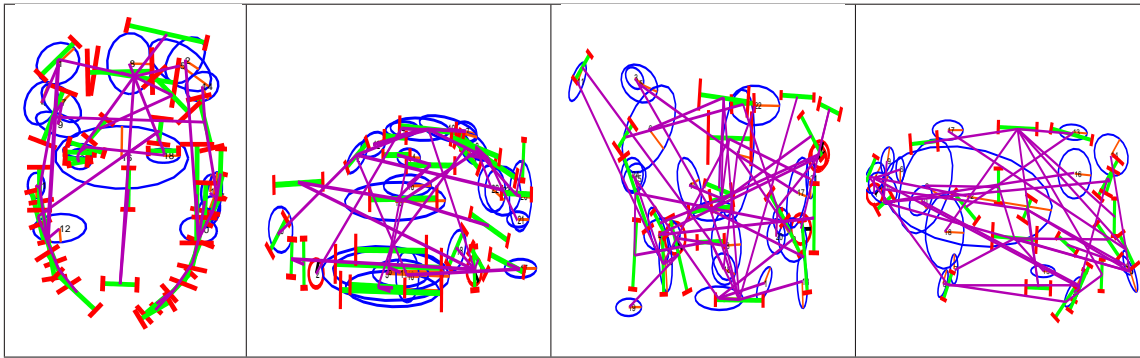
όπου \mathcal{P}_{ij} είναι το σύνολο των παρατηρούμενων ζευγών κόμβων (i, j) , N η πληθικότητά του, $\Sigma_{i,j}$ και Σ_i, Σ_j οι εκτιμήτριες μέγιστης πιθανοφάνειας των από κοινού και μεμονωμένων πινάκων διασποράς αντίστοιχα, και c μία σταθερά.

Καθώς δεν χρησιμοποιούμε δεδομένα που να έχουν επιλεγεί χειρωνακτικά δύο θέματα που προκύπτουν είναι ότι, πρώτον οι πίνακες συνδιασποράς εκτιμώνται από δεδομένα με διαφορετική ποιότητα, εξαιτίας των πιθανώς ανακριβών τοποθεσιών τμημάτων και δεύτερον η άνωθεν ανίχνευση τμημάτων παρέχει μόνο μία εκτίμηση της πιθανότητας να είναι εκεί ένα χαρακτηριστικό και όχι μία διακριτή απόφαση. Παρατηρήσαμε ότι αποκτούμε καλύτερα αποτελέσματα χρησιμοποιώντας τις πιθανότητες απευθείας στις εκτιμήσεις των παραμέτρων και του πλήθους ζευγών, αντί της κατωφλίωσής τους.

Συγκεκριμένα, οι παράμετροι των κατανομών εκτιμώνται μεγιστοποιώντας τον σταθμισμένο λογάριθμο της πιθανοφάνειας των παρατηρούμενων διανυσμάτων θέσης, $\sum_i p_i \log P(\mathcal{P}_i)$ που μας δίνει:

$$\mu_i = \frac{\sum_i p_i \mathcal{P}_i}{\sum_i p_i}, \Sigma_i = \frac{\sum_i p_i (\mathcal{P}_i - \mu_i)(\mathcal{P}_i - \mu_i)^T}{\sum_i p_i} \quad (6.21)$$

ενώ η ποσότητα N στη σχέση (6.20) αντικαθίσταται από την $\sum_i p_i$. Για κατανομές ζευγών η ποσότητα \mathcal{P}_i αντιστοιχεί στο διάνυσμα χαρακτηριστικών που σχηματίζεται ενώνοντας τα διανύσματα θέσεως των δύο τμημάτων και όπου l_i θέτουμε το γινόμενο των μεμονωμένων πιθανοτήτων να παρατηρηθεί κάθε μέλος του ζεύγους. Σημειώ-



Σχήμα 6.11: Δομή γραφικού μοντέλου για τα χαρακτηριστικά που εξήγαγε το άνωθεν σύστημα για 40 καταχωρήσεις.

νουμε εδώ ότι όπως και για την παραδοχή που κάναμε αρχικά, ότι τα στοιχεία του διανύσματος θέσης είναι ανεξάρτητα, θεωρούμε και εδώ ότι οι από κοινού κατανομές εκφράζονται ως ένα γινόμενο κατανομών πάνω στην γωνία, στην τοποθεσία και την κλίμακα των χαρακτηριστικών.

Όπως φαίνεται στο σχήμα Σχ. 6.11, η δομή του γραφικού μοντέλου που εξάγεται είναι διαισθητικά ικανοποιητική, καθώς οι περισσότερες ακμές είναι μεταξύ γειτονικών τμημάτων, συλλαμβάνοντας ορθά την αλληλεξάρτηση των χαρακτηριστικών θέσης τους.

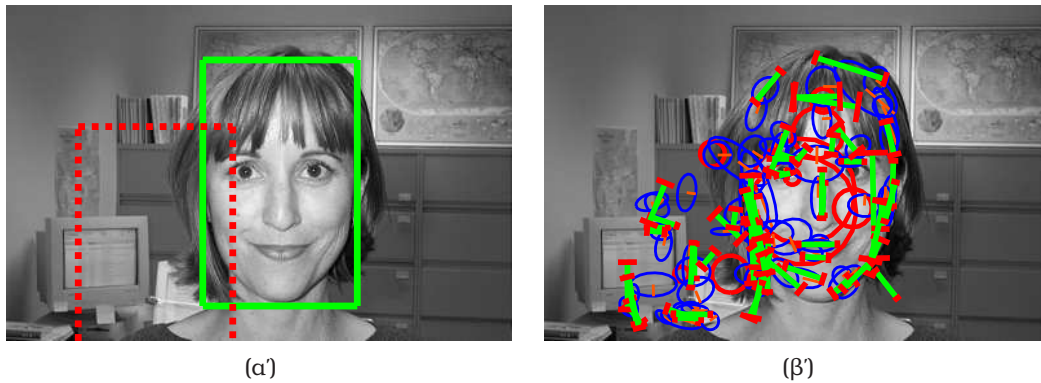
Για τον στόχο της ανίχνευσης, μετά από την εξαγωγή της δομής του γράφου εκτιμώνται οι παράμετροι των δυναμικών ζεύγους για δύο περιπτώσεις: η πρώτη αντιστοιχεί στις τοποθεσίες των τμημάτων που ανακτώνται σε εικόνες που περιέχουν το αντικείμενο (όπως προηγουμένως) και η δεύτερη σε εικόνες όπου το αντικείμενο απουσιάζει. Έτσι μπορούμε να εξετάσουμε κατά πόσο είναι αληθής μία υπόθεση αντικειμένου βάσει της πιθανοφάνειας της διάταξης των τμημάτων που εξάγονται από αυτήν.

Για το γενικότερο πλαίσιο του συνδυασμού κάτωθεν και άνωθεν πληροφορίας, θα μπορούσαμε να χρησιμοποιήσουμε αυτό το γραφικό δίκτυο αντί του προηγούμενου για να ανακτηθούν τα χαμένα τμήματα του αντικειμένου. Αναμενόμενο είναι ότι εκμεταλλευόμενοι την ακριβέστερη γνώση της από κοινού κατανομής των τμημάτων, επαρκούν λιγότερες υποθέσεις για την λειτουργία του άνωθεν συστήματος, καθιστώντας την εφαρμογή του συστήματος οικονομικότερη.

6.7 Πειραματικά Αποτελέσματα

Για να εξετάσουμε την βελτίωση που εισάγει το κάθε ένα από τα προτεινόμενα στάδια εστιάζουμε στην ανίχνευση αυτοκινήτων από τη βάση δεδομένων που παρουσιάζεται την αναφορά [3]. Καθώς σε όλες τις υπόλοιπες εργασίες, π.χ. [3, 144, 63, 9] παρέχονται καμπύλες ακρίβειας-ανάκλησης (Precision - Recall - PR) για τη συγκεκριμένη βάση δεδομένων μπορούμε να συγκρίνουμε απευθείας τα αποτελέσματά μας με τα δικά τους· για τις υπόλοιπες κατηγορίες αντικειμένων παίρνουμε ποιοτικά παρόμοια αποτελέσματα, αλλά δεν μπορούμε να τα συγκρίνουμε με προηγούμενες εργασίες, καθώς εκεί χρησιμοποιούνται καμπύλες ROC, που είναι ανεξάρτητες από την τοποθεσία του αντικειμένου.

Για τα αποτελέσματα που παρουσιάζονται στο σχήμα Σχ. 6.14 (α) αποσυνδέουμε την επίδοση του συστήματος ανίχνευσης από τη διαδικασία καταπίεσης μη-μέγιστων



Σχήμα 6.12: Μηχανισμός περιορισμού θετικών αποκρίσεων: Το κάτωθεν σύστημα ανίχνευσης προτείνει δύο δυνατές τοποθεσίες της εικόνας ως πιθανές να περιέχουν ένα πρόσωπο. Ενεργοποιώντας το άνωθεν μοντέλο το οποίο αναζητεί περιοχές στην εικόνα που να αντιστοιχούν στη δομή του, για την ψευδή υπόθεση βρίσκονται λίγες αντιστοιχίες, και αυτές συμπτωματικά· αντίθετα για την αληθή υπόθεση οι περισσότερες καταχωρήσεις του codebook έχουν βρεί μία αντίστοιχη δομή στην εικόνα. Έτσι, εξασθενεί η ψευδής υπόθεση και ισχυροποιείται η αληθής.

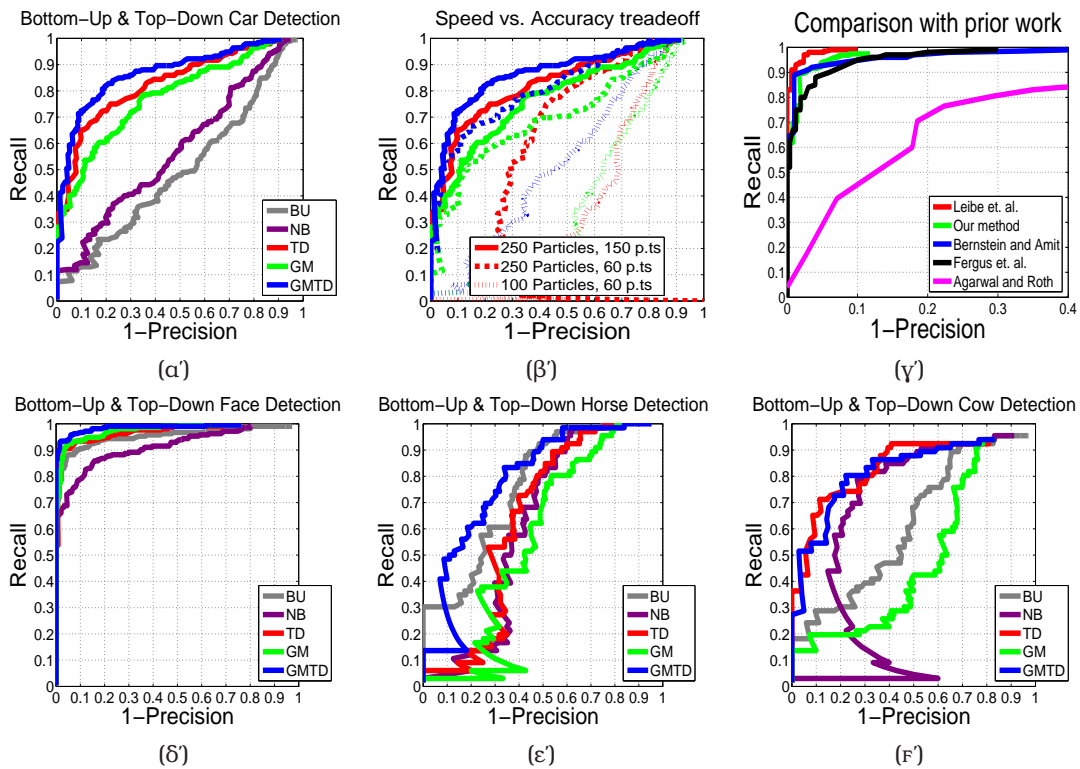
(nonmaximum suppression) [3] που εφαρμόζεται σε ένα ύστερο στάδιο χρησιμοποιώντας ένα μικρό παράθυρο διαστάσεων 5×5 · έτσι πολλές ψευδείς ανιχνεύσεις που βρίσκονται σε γειτονικές περιοχές μετρώνται ως ανεξάρτητες και δίνουν έναν φαινομενικά χαμηλό βαθμό ακρίβειας. Συνδυάζοντας τις εξόδους των μεμονωμένων ταξινομητών όπως στο προηγούμενο κεφάλαιο, παρατηρούμε ότι κάθε ένα από τα προτεινόμενα στάδια εισάγει μία βελτίωση στην επίδοση, με τα τελικά αποτελέσματα που συνδυάζουν τις άνωθεν και κάτωθεν μεθόδους ανίχνευσης να είναι αισθητά καλύτερα από αυτά της καθαρά κάτωθεν.

Η βελτίωση στην επίδοση που επιτυγχάνεται εισάγοντας ένα μεγαλύτερο πλήθος από υποθέσεις και μειώνοντας το βήμα ολοκλήρωσης φαίνεται στο σχήμα Σχ. 6.14(β)· για τα τελικά μας αποτελέσματα χρησιμοποιούμε τα αποτελέσματα της μεθόδου με την μεγαλύτερη ακρίβεια, που κοστίζει 40 δεύτερα ανά υπόθεση σε ένα υπολογιστή ταχύτητας 1.4 Ghz. ενώ οι άλλες δύο περιπτώσεις χρειάζονται περίπου 7 και 15 δεύτερα. Μία άμεση βελτίωση θα μπορούσε να επιτευχθεί χρησιμοποιώντας μία ακολουθία από ελέγχους [227], εξετάζοντας αρχικά αν βρέθηκαν τα πιο χαρακτηριστικά σημεία (π.χ. το μέτωπο για ένα πρόσωπο, οι ρόδες για ένα αυτοκίνητο) και κατόπιν, αν έχουν βρεθεί να συνεχίζεται ο έλεγχος.

Στο σχήμα Σχ. 6.14(γ) όπου συγκρίνουμε τα αποτελέσματά μας με αυτά πρόσφατων εργασιών, η μέθοδός μας κατατάσσεται ανάμεσα στις καλύτερες, και είναι συστηματικά κατώτερη μόνο από τα αποτελέσματα της εργασίας [144], όπου πληροφορία για την φωτεινότητα, ένα codebook μεγαλύτερης διάστατης και ένας τελεστής ενδιαφέροντος που λειτουργεί σε προκαθορισμένη κλίμακα χρησιμοποιούνται. Σημειώνουμε εδώ ότι καθώς περιστρέφουμε τα αντικείμενα κατά τη διάρκεια της εκπαίδευσης ώστε να έχουν την ίδια κατεύθυνση, μπορεί να έχει εισαχθεί κάποια μικρή βελτίωση στην επίδοση, καθώς στις υπόλοιπες εργασίες δεν αναφέρεται αυτό το σημείο. Η διαφορά αυτή είναι αμελητέα, καθώς τα αυτοκίνητα στην βάση [3] και τα μοντέλα που μαθαίνονται είναι σχεδόν συμμετρικά. Αυτό το οποίο θεωρούμε πιο σημαντικό είναι ότι άριστα αποτελέσματα ανίχνευσης έχουν επιτευχθεί συνδιάζοντας την κάτωθεν με την άνωθεν πληροφορία χωρίς περίπλοκους περιγραφείς εμφάνισης



Σχήμα 6.13: Αποτελέσματα συνεργετικής ανίχνευσης αντικειμένων: το κάτωθεν σύστημα προτείνει ένα σύνολο πιθανών τοποθεσιών για αντικείμενα, οι οποίες επαληθεύονται από το άνωθεν σύστημα, όπως περιγράφεται στο κείμενο. Για μία συγκεκριμένη τιμή ενός κατωφλιού, οι τοποθεσίες με κόκκινη διακεκομμένη γραμμή απορρίπτονται, ενώ με την πράσινη συνεχή εγκρίνονται. Γενικότερα έτσι περιορίζονται οι ψευδείς ανιχνεύσεις, με το κόστος ορισμένων χαμένων θετικών (βλ. δεξιά στήλη).



Σχήμα 6.14: Πάνω γραμμή: (α) Βελτιώσεις στην επίδοση που εισάγονται σε διαφορετικά στάδια επεξεργασίας: BU: bottom-up, NB: naive Bayes, TD: Top-down, GM: Graphical model, GMTD: Graphical model & Top-down. (β) Συμβιβασμός ακρίβειας / ταχύτητας: η επίδοση του άνωθεν ανιχνευτή βελτιώνεται αυξάνοντας το πλήθος των υποθέσεων (particles) και ελαττώνοντας το βήμα ολοκλήρωσης. (γ) Σύγκριση με αποτελέσματα άλλων μεθόδων, μετά την καταπίεση μη-μεγίστων. Κάτω γραμμή: Συγκριτικά αποτελέσματα για τις άλλες κατηγορίες αντικειμένων.

όπως στις αναφορές [149, 63], χρησιμοποιώντας ένα μικρό λεξικό και αποδοτικούς αλγορίθμους.

Στην κάτω γραμμή καμπύλες $P - R$ παρουσιάζονται για τις υπόλοιπες κατηγορίες, όπου πάλι βλέπουμε τη βελτίωση που επιτυγχάνεται συνδυάζοντας την άνωθεν με την κάτωθεν πληροφορία. Για πρόσωπα επιτυγχάνεται άριστος εντοπισμός των αντικειμένων, ενώ για τις άλλες δύο κατηγορίες τα αποτελέσματα είναι κάπως χειρότερα, κυρίως λόγω του συνδυασμού ενός μικρού συνόλου εκπαίδευσης και της πιο δύσκολης δομής των αντικειμένων.

Πειραματικές Ρυθμίσεις

Για αυτοκίνητα χρησιμοποιούμε τη βάση δεδομένων και τη μέθοδο αποτίμησης της αναφοράς [3], με τις ίδιες πειραματικές ρυθμίσεις που περιγράφονται στο προηγούμενο κεφάλαιο. Κατά τη διάρκεια της εκπαίδευσης 100 εικόνες χρησιμοποιούνται οι οποίες αν χρειαστεί αντιστρέφονται ώστε τα αυτοκίνητα να έχουν την ίδια κατεύθυνση. Για άλογα και αγελάδες χρησιμοποιούμε τα δεδομένα των [19, 144] αντίστοιχα και 50 εικόνες για εκπαίδευση και 50 για εξέταση. Οι αντίστοιχες ελλείψεις καλύπτουν προσεγγιστικά το 1/2 της επιφάνειας του κορμού των ζώων: συγκεκριμένα οι διαστάσεις των οριζώντιων/ κάθετων αξόνων είναι 20 / 45 pixels για αγελάδες και 35 / 50 pixels για άλογα.

6.A' Αναλυτικές Σχέσεις Μηνυμάτων Μεταξύ Γειτόνων

Κατά τη διάρκεια της εκπαίδευσης, όταν κατασκευάζονται οι συστάδες του codebook χρησιμοποιείται για κάθε συστάδα i η κατανομή $P_i(\mathcal{P})$. ακριβέστερα, αυτή είναι η μοντελοποίηση της κατανομής των χαρακτηριστικών, *δεδομένης* της θέσης του αντικειμένου: $P_i(\mathcal{P}|\mathcal{O})$. Κατά τη διάρκεια της ανίχνευσης των αντικειμένων η θέση του αντικειμένου είναι εν γένει άγνωστη· για αυτό όταν θέλουμε να εκφράσουμε το μήνυμα που στέλνει ο κόμβος i στο κόμβο j οπότε και εμπλέκεται η κατανομή $P(\mathcal{P}_i|\mathcal{P}_j)$ πρέπει να απαλοίσουμε την θέση του αντικειμένου μέσω ολοκλήρωσης. Θεωρώντας το διδιάστατο διάνυσμα θέσης \mathbf{x} , για κατανομές που εξασφαλίζουν ανεξαρτησία από μετατόπιση έχουμε (όλες οι σχέσεις γράφονται αντίστοιχα και για τα υπόλοιπα χαρακτηριστικά)

$$P(\mathbf{x}_i|\mathbf{x}_o) = \int_{\mathbf{x}_c} P(\mathbf{x}_i, \mathbf{x}_c|\mathbf{x}_o) = \int_{\mathbf{x}_c} P(\mathbf{x}_i|\mathbf{x}_c, \mathbf{x}_o)P(\mathbf{x}_c|\mathbf{x}_o) \quad (22)$$

$$= \int_{\mathbf{x}_c} \frac{P(\mathbf{x}_i, \mathbf{x}_c, \mathbf{x}_o)}{P(\mathbf{x}_c, \mathbf{x}_o)} P(\mathbf{x}_c|\mathbf{x}_o) = \int_{\mathbf{x}_c} \frac{P(\mathbf{x}_i, \mathbf{x}_o|\mathbf{x}_c)}{P(\mathbf{x}_o|\mathbf{x}_c)} P(\mathbf{x}_c|\mathbf{x}_o) \quad (23)$$

$$= \int_{\mathbf{x}_c} \frac{P(\mathbf{x}_i, \mathbf{x}_o|\mathbf{x}_c)}{P(\mathbf{x}_o|\mathbf{x}_c)} P(\mathbf{x}_o|\mathbf{x}_c) \quad (24)$$

όπου $P(\mathbf{x}_o|\mathbf{x}_c) = P(\mathbf{x}_c|\mathbf{x}_o)$ καθώς μοντελοποιούμε και τις δύο κατανομές μέσω των διαφορών των $\mathbf{x}_c, \mathbf{x}_o$. Στην τελική έκφραση όλοι οι όροι είναι διαθέσιμοι καθώς η θέση του αντικειμένου θεωρείται γνωστή, οπότε μπορούμε να χρησιμοποιήσουμε τις παραμέτρους που εκτιμώνται κατά την εκπαίδευση. Απομένει λοιπόν ο υπολογισμός των ολοκληρωμάτων, ο οποίος είναι αναλυτικά εφικτός καθώς οι συναρτήσεις είναι Gaussian.

Οι όροι που εμπλέκονται στην πρώτη ολοκληρούμενη ποσότητα γράφονται:

$$P(\mathbf{x}_o|\mathbf{x}_c) = \frac{1}{2\pi|\Sigma_{o,o}|^{1/2}} \exp\left(-[(\mathbf{x}_o - \mathbf{x}_c) - \mu_o]^T \Sigma_{o,o} [(\mathbf{x}_o - \mathbf{x}_c) - \mu_o]\right) \quad (25)$$

$$P(\mathbf{x}_o, \mathbf{x}_i|\mathbf{x}_c) = \frac{1}{(2\pi)^2|\Sigma_{oi}|^{1/2}} \exp\left(-\begin{bmatrix} (\mathbf{x}_o - \mathbf{x}_c) - \mu_o \\ (\mathbf{x}_i - \mathbf{x}_c) - \mu_i \end{bmatrix}^T \Sigma_{oi}^{-1} \begin{bmatrix} (\mathbf{x}_o - \mathbf{x}_c) - \mu_o \\ (\mathbf{x}_i - \mathbf{x}_c) - \mu_i \end{bmatrix}\right)$$

$$\Sigma_{oi} = \begin{bmatrix} \Sigma_{o,o} & \Sigma_{o,i} \\ \Sigma_{i,o} & \Sigma_{i,i} \end{bmatrix} \quad (26)$$

όπου για την μοντελοποίηση της κατανομής της θέσης των σημείων των codebooks $\mathbf{x}_o, \mathbf{x}_i$ αφαιρούμε τη θέση του κέντρου του αντικειμένου \mathbf{x}_c : με μ, Σ συμβολίζουμε τις εκτιμήτριες του μέσου και του πίνακα διασποράς των αντίστοιχων διαφορών. Μπορούμε λοιπόν να γράψουμε

$$\frac{P(\mathbf{x}_o, \mathbf{x}_i|\mathbf{x}_c)}{P(\mathbf{x}_o|\mathbf{x}_c)} = \frac{1}{2\pi|\Sigma_1|^{1/2}} \exp\left(-(\mathbf{x}_i - \mu'_i)^T \Sigma_1^{-1} (\mathbf{x}_i - \mu'_i)\right)$$

$$\mu'_i = \mu_i + \mathbf{x}_c + \Sigma_{i,o}^T \Sigma_{o,o}^{-1} (\mathbf{x}_o - (\mu_o + \mathbf{x}_c))$$

$$\Sigma_1 = \Sigma_{i,i} - \Sigma_{i,o}^T \Sigma_{o,o}^{-1} \Sigma_{i,o} \quad (27)$$

Καθώς το επόμενο βήμα περιλαμβάνει την ολοκλήρωση ως προς την άγνωστη ποσότητα \mathbf{x}_c , ξαναγράφουμε τη διαφορά μεταξύ των όρων μέσα στο εκθετικό ως μία έκφραση του \mathbf{x}_c :

$$\begin{aligned} \mathbf{x}_i - \mu'_i &= \mathbf{x}_i - \mu_i - \mathbf{x}_c - \Sigma_{i,o}^T \Sigma_{o,o}^{-1} (\mathbf{x}_o - (\mu_o + \mathbf{x}_c)) \\ &= \underbrace{[\Sigma_{i,o}^T \Sigma_{o,o}^{-1} - \mathbf{I}]}_{\Sigma_c} \mathbf{x}_c - \underbrace{[\mu_i + \Sigma_{i,o}^T \Sigma_{o,o}^{-1} (\mathbf{x}_o - \mu_o) - \mathbf{x}_i]}_{\mu_c} \end{aligned} \quad (28)$$

Συνεπώς μπορούμε να γράψουμε την παραπάνω σχέση ως:

$$\begin{aligned} \frac{P(\mathbf{x}_o, \mathbf{x}_i | \mathbf{x}_c)}{P(\mathbf{x}_o | \mathbf{x}_c)} &= \frac{|\Sigma_2|^{1/2}}{|\Sigma_1|^{1/2}} \frac{1}{2\pi |\Sigma_2|^{1/2}} \exp(-(\mathbf{x}_c - \mu_c)^T \Sigma_2^{-1} (\mathbf{x}_c - \mu_c)) \\ \Sigma_2 &= (\Sigma_c^T \Sigma_1^{-1} \Sigma_c)^{-1} \end{aligned}$$

Ο άλλος όρος που μένει προς ολοκλήρωση είναι ο $P(\mathbf{x}_o | \mathbf{x}_c)$ ο οποίος δίνεται από την (25), οπότε πέρα από τη σταθερά $\frac{|\Sigma_2|}{|\Sigma_1|}$ έχουμε να υπολογίσουμε το ολοκλήρωμα του γινομένου δύο Gaussian συναρτήσεων:

$$\begin{aligned} \int_{\mathbf{x}_c} &\frac{1}{2\pi |\Sigma_2|^{1/2}} \exp(-(\mu_c - \mathbf{x}_c)^T \Sigma_2^{-1} (\mu_c - \mathbf{x}_c)) \cdot \\ &\frac{1}{2\pi |\Sigma_{o,o}|^{1/2}} \exp\left(-[(\mathbf{x}_o - \mathbf{x}_c) - \mu_o]^T \Sigma_{o,o} [(\mathbf{x}_o - \mathbf{x}_c) - \mu_o]\right) \end{aligned} \quad (29)$$

το οποίο δίνει ξανά μία Gaussian συνάρτηση με πίνακα διασποράς:

$$\Sigma_3 = (\Sigma_2^{-1} + \Sigma_{o,o}^{-1})^{-1} \quad (30)$$

και μέσο:

$$\mu_3 = \Sigma_3 \Sigma_2^{-1} \mu_c + \Sigma_3 \Sigma_{o,o}^{-1} (\mathbf{x}_o - \mu_o) \quad (31)$$

Κεφάλαιο 7

Συμπεράσματα και Κατευθύνσεις Μελλοντικής Έρευνας

Στα πλαίσια της διατριβής αυτής επιζητήθηκε η επίλυση προβλημάτων της όρασης χρησιμοποιώντας μία πιθανοτική προσέγγιση, με κύριο στόχο την ενοποίηση και από κοινού αντιμετώπιση των προβλημάτων της κατάτμησης εικόνων και ανίχνευσης αντικειμένων. Από την έρευνα που έγινε έχουν προκύψει συνεισφορές τόσο στο χειρισμό της συνέργειας μεταξύ χαμηλού και υψηλού επιπέδου λειτουργιών της όρασης, όσο και στα επιμέρους προβλήματα.

Τα κύρια καινοτομικά αποτελέσματα και οι μελλοντικές τους επεκτάσεις μπορούν να συνοψιστούν στα παρακάτω σημεία:

Ανάλυση και Εκμάθηση ενός Βιολογικά Εμπνευσμένου Μοντέλου Χαμηλού και Μέσου Επιπέδου Λειτουργιών Όρασης

Η εργασία αυτή αποτέλεσε ένα πρώτο βήμα για την ολοκλήρωση μοντέλων βιολογικής και υπολογιστικής όρασης· ξεκινώντας από ένα απλοποιημένο μοντέλο για λειτουργίες του βιολογικού συστήματος όρασης, η λειτουργία του διατυπώθηκε ως η ελαχιστοποίηση ενός μεταβολικού κριτηρίου, ενώ η μεταβολική προσέγγιση στη συμπερασματολογία επέτρεψε την πιθανοτική του ανάλυση. Έτσι κατέστη δυνατή η εκμάθηση των βαρών του δικτύου από δεδομένα εκπαίδευσης, που οδήγησε σε συστηματικές βελτιώσεις στις επιδόσεις του δικτύου

Βλέποντας αυτήν την προσέγγιση ξανά στο μεταβολικό πλαίσιο, μπορούμε να πούμε ότι εκμαθήθηκε ένα κριτήριο που εμπλέκει όρους αντίστοιχους του συναρτησιακού Mumford-Shah και το οποίο ποσοτικοποιεί την ποιότητα μίας διαδικασίας ανίχνευσης ακμών. Θα ήταν ενδιαφέρον να εξεταστεί κατά πόσο αυτή η προσέγγιση μπορεί να εφαρμοστεί και σε άλλα μεταβολικά κριτήρια, π.χ. για εκτίμηση κίνησης, που συνήθως ορίζονται κατά έναν αυθαίρετο τρόπο. Η ιδέα της εκμάθησης κριτηρίων για προβλήματα όρασης είναι σχετικά νέα, π.χ. [67] και θα μπορούσε να οδηγήσει σε βελτιωμένους αλγόριθμους που να εκμεταλλεύονται τα χαρακτηριστικά του οπτικού μας περιβάλλοντος.

Ανάλυση και Κατάτμηση Εικόνων Υφής Μέσω Χαρακτηριστικών Διαμόρφωσης, Αναγεννητικών Μοντέλων και Σταθμισμένης Εξέλιξης Καμπυλών

Στα πλαίσια αυτής της ερευνητικής κατεύθυνσης επιδιώχθηκε η χρήση των μοντέλων διαμόρφωσης για τα προβλήματα της ανάλυσης υφής και κατάτμησης εικό-

νων. Εισάγοντας τα τοπικά αναγεννητικά μοντέλα ερμηνεύτηκε η διαδικασία εξαγωγής χαρακτηριστικών με φίλτρα Gabor ως η προσαρμογή ενός μοντέλου και η επιλογή καναλιού της μεθόδου DCA ως μία διαδικασία εξέτασης πολλαπλών υποθέσεων. Επεκτείνοντας αυτή την προσέγγιση, η ανίχνευση ακμών διατυπώθηκε στο ίδιο πλαίσιο των τοπικών αναγεννητικών μοντέλων, επιτρέποντας τον υπολογισμό ύστερων πιθανοτήτων για τρεις γενικές κλάσεις τοπικής δομής (υφή, ακμές και ομαλές περιοχές). Τέλος, βάσει μίας τεχνικής συνδυασμού ταξινομητών εισήχθη η μέθοδος Σταθμισμένης Εξέλιξης Καμπυλών που επιτρέπει τον συνδυασμό διαφορετικών πηγών πληροφορίας χρησιμοποιώντας της ύστερες πιθανότητες των τριών κλάσεων. Επιτεύχθηκε έτσι η κατάτμηση μίας ποικιλίας φυσικών εικόνων, δίνοντας συστηματικά καλύτερα αποτελέσματα.

Σε μελλοντική εργασία θα ήταν ενδιαφέρον να διατυπωθεί μία σύνδεση με την προσέγγιση του D. Lindeberg που παρουσιάστηκε στο τελευταίο κεφάλαιο. Η βασική διαφορά είναι ότι ενώ εδώ βασιζόμαστε στην ποιότητα ανακατασκευής της εικόνας, ενώ στην εργασία του χρησιμοποιούνται διαφορικές ποσότητες. Η επίτευξη μίας σύνδεσης θα μπορούσε αφενός να εισάγει χαρακτηριστικά τύπου υφής ως περιγραφείς για ανίχνευση αντικειμένων, αφετέρου την χρήση όρων πιθανότητας για τα χαρακτηριστικά primal sketch.

Συνέργεια μεταξύ Αναγνώρισης Αντικειμένων και Κατάτμησης Εικόνων μέσω του Αλγορίθμου Expectation Maximization

Σε αυτήν την πτυχή της έρευνας της διατριβής αντιμετωπίστηκε το πρόβλημα της από κοινού κατάτμησης, ανάλυσης και ανίχνευσης αντικειμένων, θέτοντάς το στο πλαίσιο του αλγορίθμου EM. Η σαφήνεια και η απλότητα είναι τα βασικά πλεονεκτήματα της προσέγγισης αυτής, που έχουν επιτρέψει την ενσωμάτωση διαφορετικών τεχνικών της όρασης υπολογιστών κατά έναν ξεκάθαρο τρόπο. Συγκεκριμένα συνδυάστηκαν μοντέλα AAM με τεχνικές εξέλιξης καμπυλών για κατάτμηση αντικειμένων, εισάγοντας πληροφορία κατάτμησης για τα πρώτα και στατιστική πληροφορία σχήματος για την εξέλιξη καμπυλών. Βάσει του EM κατασκευάστηκε ένα σύστημα το οποίο μπορεί να κατατμεί εικόνες αντικειμένων που ανήκουν σε κατηγορίες με μεγάλη ποικιλία στην εμφάνιση, βελτιώνοντας παράλληλα τις επιδόσεις ανίχνευσης.

Μία κατεύθυνση μελλοντικής έρευνας αποτελεί η εκμάθηση μοντέλων για γενικής φύσης αντικείμενα· σχετικά πρόσφατες εργασίες, π.χ. [233, 19, 62] έχουν δώσει το έναυσμα για μία έντονη δραστηριότητα γύρω από την αυτοματοποιημένη κατασκευή τμηματικών αναπαραστάσεων [3, 144, 63, 133] αλλά πολύ λίγες εργασίες ασχολούνται με την κατασκευή καθολικών αναγεννητικών μοντέλων [236]. Επίσης, ένα σημείο το οποίο έχει μόνο μερικώς αντιμετωπιστεί στην υπάρχουσα βιβλιογραφία [238, 136, 133] είναι ο συνδυασμός χαρακτηριστικών χαμηλού επιπέδου (π.χ. ακμές) με τμηματικά ή/και καθολικά αντικείμενα για την κατάτμηση εικόνων.

Συνεργετική Ανίχνευση Αντικειμένων μέσω Σημείων Ενδιαφέροντος και Γραφικών Δικτύων

Στα πλαίσια της εργασίας αυτής απλές γεωμετρικές δομές, όπως ευθείες ακμές/ράβδοι χρησιμοποιήθηκαν για την αναπαράσταση κατηγοριών αντικειμένων χρησιμοποιώντας ένα μικρό codebook για τα τμήματα του αντικειμένου. Η εισαγωγή μίας αποτελεσματικής μεθόδου για την αποτίμηση των όρων που εμπλέκονται στις εκφράσεις πιθανοφάνειας κατέστησε δυνατή την εισαγωγή άνωθεν γνώσης στο πρόβλημα της

ανίχνευσης μέσω τεχνικών γραφικών δικτύων, οδηγώντας σε συστηματικά καλύτερα αποτελέσματα.

Είναι ενδιαφέρον να εισαχθεί περιπλοκότερη πληροφορία για την εμφάνιση από τα χρησιμοποιούμενα εδώ τοπικά σταθερά μοντέλα, διατηρώντας παράλληλα τους υπολογισμούς απλούς και αποτελεσματικούς. Επίσης, διδιάστατα χαρακτηριστικά όπως γωνίες, διασταυρώσεις και γενικότερα περιπλοκότερα χαρακτηριστικά τύπου *primal sketch* μπορούν επίσης να ενσωματωθούν στο μοντέλο μας κατασκευάζοντας απλές παραμετρικές τεμπλέτες, επιτρέποντας τη χρήση τους σε άνωθεν διαδικασίες, βελτιώνοντας παράλληλα την επίδοση και των δύο μεθόδων ανίχνευσης.

Καταλήγοντας, για όλα τα προβλήματα που αντιμετωπίστηκαν, το μέσο για την διατύπωση και επίλυση προβλημάτων της όρασης αποτέλεσε η χρήση πιθανοτικών μοντέλων. Ο κύριος περιορισμός αυτής της προσέγγισης είναι η μεγάλη ποικιλία των οπτικών αντικειμένων που υπάρχουν στο περιβάλλον: αν και για συγκεκριμένες περιπτώσεις, π.χ. πρόσωπα, πεζοί ή αυτοκίνητα μπορούν να κατασκευαστούν μοντέλα εντοπίζοντας χειρωνακτικά τα σημεία τους, θα ήταν αδύνατο να γίνει το ίδιο για κάθε αντικείμενο που αντιλαμβανόμαστε.

Για το στόχο αυτό θα μπορούσαν να επεκταθούν οι συνεισφορές της παρούσας εργασίας προς την αυτοματοποιημένη κατασκευή μοντέλων μεγαλύτερης πολυπλοκότητας, τα οποία να συνδυάζουν τα προτερήματα των γραφικών και των αναγεννητικών μοντέλων. Για παράδειγμα, μπορεί κανείς να φανταστεί τις αναπαραστάσεις των μοντέλων του έκτου κεφαλαίου ως έναν σκελετό του αντικειμένου, που συλλαμβάνει συνοπτικά την κυρίαρχη δομή του αντικειμένου και τα AAMs του πέμπτου ως ένα μοντέλο υψηλότερης ακρίβειας, που 'χτίζεται' πάνω στον σκελετό του αντικειμένου. Καταλήγουμε έτσι στο γενικότερο πρόβλημα του συνδυασμού αναγεννητικών και γραφικών μοντέλων.

Επίσης, μία προσέγγιση που είχε μείνει σε αδράνεια για ένα μεγάλο χρονικό διάστημα και έχει πρόσφατα επανέρθει είναι η χρήση συντακτικών αναπαραστάσεων. Για την περίπτωση μας, η εκμάθηση των συντακτικών κανόνων για τους τρόπους με τους οποίους μία καμπύλη παράγει ευθεία τμήματα θα μπορούσε να δώσει συμπαγείς αναπαραστάσεις των αντικειμένων, που θα περιελάμβαναν για παράδειγμα για την κατηγορία των προσώπων δομές όπως 'μέτωπο', 'στόμα', 'μαλλιά' κ.ο.κ. καθώς και τους τρόπους με τους οποίους αυτές σχετίζονται.

Πέρα από τις σχετικά φιλόδοξες παραπάνω ιδέες, ένας απτός στόχος είναι η περαιτέρω εκμετάλλευση της πλούσιας μαθηματικής βάσης και διαισθητικής σαφήνειας του πιθανοτικού πλαισίου. Καθώς η ροή ιδεών από το χώρο της μάθησης μηχανών και της πιθανοτικής συμπερασματολογίας έχει αρχίσει να αποδίδει καρπούς π.χ. [57, 61, 210], φαίνεται πλέον εφικτή η κατασκευή επιτυχημένων συστημάτων όρασης υπολογιστών για πρακτικές εφαρμογές.

Βιβλιογραφία

- [1] D. Ackely, G. Hinton, and T. Sejnowski. A Learning Algorithm for Boltzmann Machines. *Cognitive Science*, 9:147–169, 1985.
- [2] E. Adelson and J. Bergen. Spatiotemporal Energy Models for the Perception of Motion. *J. Opt. Soc. Amer.*, 2(2):284–299, 1985.
- [3] S. Agrawal and D. Roth. Learning a Sparse Representation for Object Detection. In *Proc. Eur. Conf. on Comp. Vision*, volume 4, pages 113–130, 2002.
- [4] S. Baker, I. Matthews, and J. Schneider. Automatic Construction of Active Appearance Models as an Image Coding Problem. *IEEE Trans. PAMI*, 26:1380–1384, 2004.
- [5] A. Barbu and S. Zhu. Graph partition by Swendsen-Wang cuts. In *Proc. Intl Conf. on Comp. Vision*, 2003.
- [6] A. J. Bell and T. J. Sejnowski. The "Independent Components" of Natural Scenes are Edge Filters. *Vision Research*, 37(23):3327–3338, 1997.
- [7] S. Belongie, C. Carson, H. Greenspan, and J. Malik. Color- and texture-based image segmentation using EM and its application to content-based image retrieval. In *Proc. Intl Conf. on Comp. Vision*, 1998.
- [8] A. Berger. The Improved Iterative Scaling Algorithm: A Gentle Introduction, 1997. "<http://citeseer.ist.psu.edu/berger97improved.html>".
- [9] E. J. Bernstein and Y. Amit. Part-Based Statistical Models for Object Classification and Detection. In *Proc. IEEE Conf. on Comp. Vision and Pat. Recogn.*, 2005.
- [10] S. Beucher and F. Meyer. The Morphological Approach to Segmentation: The Watershed Transformation. In E. R. Dougherty, editor, *Mathematical Morphology in Image Processing*, pages 433–481. Marcel Dekker, New York, 1993.
- [11] I. Biederman. Human Image Understanding: Recent Research and a Theory. *Computer Vision, Graphics, and Image Processing*, 32:29–73, 1985.
- [12] M. Black, G. Sapiro, D. Marrimont, and D. Heeger. Robust Anisotropic Diffusion. *IEEE Trans. Im. Proc.*, 7(3):421–432, 1998.

- [13] A. Blake and M. Isard. *Active Contours*. Springer Verlag, 1998.
- [14] A. Blake and A. Yuille, editors. *Active Vision*. MIT Press, 1992.
- [15] A. Blake and A. Zisserman. *Visual Reconstruction*. MIT Press, 1987.
- [16] V. Blanz, T., and Vetter. Face Recognition Based on Fitting a 3D Morphable Model. *IEEE Trans. PAMI*, 25(9):1063 – 1074, 2003.
- [17] K. Blekas, A. Likas, N. P. Galatsanos, and I. E. Lagaris. A Spatially-Constrained Mixture Model for Image Segmentation. *IEEE Trans. Neur. Netw.*, 16(2):494–498.
- [18] E. Borenstein, E. Sharon, and S. Ullman. Combining Top Down and Bottom-Up Segmentation. In *Perceptual Organization Workshop, CVPR*, volume 4, 2004.
- [19] E. Borenstein and S. Ullman. Class-Specific, Top-Down Segmentation. In *Proc. Eur. Conf. on Comp. Vision*, 2002.
- [20] D. Boukerroui, J. Noble, and M. Brady. On the Choice of Band-Pass Quadrature Filters. *Journal of Mathematical Imaging and Vision*, 21:53–80, 2004.
- [21] A. Bovik, M. Clark, and W. Geisler. Multichannel Texture Analysis using Localized Spatial Filters. *IEEE Trans. PAMI*, 12(1):55–73, 1990.
- [22] A. C. Bovik, N. Gopal, T. Emmoth, and A. Restrepo. Localized Measurement of Emergent Image Frequencies by Gabor Wavelets. *IEEE Trans. Information Theory*, 38:691–712, 1992.
- [23] Y. Boykov and V. Kolmogorov. Computing geodesics and minimal surfaces via graph cuts. In *Proc. Int.l Conf. on Comp. Vision*, 2003.
- [24] Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *IEEE Trans. PAMI*, 23(11):1222–1239, 2001.
- [25] T. Brox and J. Weickert. A TV Flow based Local Scale Measure for Texture Discrimination. In *Proc. Eur. Conf. on Comp. Vision*, 2004.
- [26] C. Burges. A Tutorial on Support Vector Machines. 1998.
- [27] M. Burl, M. Weber, and P. Perona. A Probabilistic Approach to Object Recognition using Local Photometry and Global Geometry. In *Proc. Eur. Conf. on Comp. Vision*, 1998.
- [28] J. Canny. A Computational Approach to Edge Detection. *IEEE Trans. PAMI*, 8(6):679–698, 1986.
- [29] M. Carandini and D. Heeger. Summation and Division by Neurons in Visual Cortex. *Science*, 264:1333–1336, 1994.
- [30] M. Carandini, D. Heeger, and J. A. Movshon. Linearity and Normalization of Simple Cells of the Macaque Primary Visual Cortex. *Journal of Neuroscience*, 17:8621–8644, 1997.

- [31] V. Caselles, R. Kimmel, and G. Sapiro. Geodesic Active Contours. *Int.l. J. of Comp. Vision*, 22(1):61–79, 1997.
- [32] T. Chan and L. Vese. Active Contours Without Edges. *IEEE Trans. Im. Proc.*, 10(2):266–277, 2001.
- [33] C. Chow and C. Liu. Approximating discrete probability distributions with dependence trees. *IEEE Trans. Information Theory*, 14(3):462–467, 1968.
- [34] L. Cohen. On Active Contour Models and Balloons. *Computer Vision, Graphics, and Image Processing: Image Understanding*, 53(2):211–218, 1991.
- [35] M. Cohen and S. Grossberg. Absolute Stability of Global Pattern Formation and Parallel Memory Storage by Competitive Neural Networks. *IEEE Trans. Systems, Man, and Cybernetics*, 13(5):815–826, 1983.
- [36] D. Comaniciu and P. Meer. Mean shift: A Robust Approach toward Feature Space Analysis. *IEEE Trans. PAMI*, 24(5):603–619, 2002.
- [37] T. Cootes, G. J. Edwards, and C. Taylor. Active Appearance Models. *IEEE Trans. PAMI*, 23(6):681–685, 2001.
- [38] T. Cootes and C. Taylor. Active Shape Models- Smart Snakes. In *Proc. Brit. Mach. Vis. Conf*, 1992.
- [39] T. Cootes and C. Taylor. Constrained Active Appearance Models. In *Proc. Intl Conf. on Comp. Vision*, 2001.
- [40] T. F. Cootes. Statistical Models of Appearance for Computer Vision. Technical report, University of Manchester, 2000.
- [41] T. F. Cootes, G.V.Wheeler, K.N.Walker, and C. J. Taylor. View-Based Active Appearance Models. *Image and Vision Computing*, 20:657–664, 2002.
- [42] R. Courant and D. Hilbert. *Methods of Mathematical Physics*, volume I. Wiley Interscience, 1953.
- [43] Cover and Thomas. *Elements of Information Theory*. Wiley, 1993.
- [44] R. G. Cowell, A. P. Dawid, S. L. Lauritzen, and D. J. Spiegelhalter. *Probabilistic networks and expert systems*. Springer Verlag, 1999.
- [45] D. Crandall, P. Felzenszwalb, and D. Huttenlocher. Spatial Priors for Part-Based Recognition using Statistical Models. In *Proc. IEEE Conf. on Comp. Vision and Pat. Recogn.*, 2005.
- [46] D. Cremers, N. Sochen, and C. Schnorr. Towards Recognition Based Variational Segmentation Using Shape Priors and Dynamic Labelling. In *Proc. Intl. Conf. Scale Space*, 2003.
- [47] G. C. Cross and A. K. Jain. Markov random field texture models. *IEEE Trans. PAMI*, 5(1):25–39, Jan. 1983.

- [48] D. P. Huttenlocher and S. Ullman. Recognizing Solid Objects by Alignment with an Image. *Int.l. J. of Comp. Vision*, 5(2):195 – 212, 1990.
- [49] J. Daugman. Uncertainty Relation for Resolution in Space, Spatial Frequency and Orientation Optimized by Two-Dimensional Visual Cortical Filters. *J. Opt. Soc. Amer. (A)*, 2(7):160–169, 1985.
- [50] P. Dayan and L. Abbott. *Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems*. MIT Press, 2001.
- [51] P. Dayan, G. Hinton, R. Neal, and R. Zemel. The Helmholtz Machine. *Neural Computation*, 7:889–904, 1995.
- [52] A. Dempster, N. Laird, and D. Rudin. Maximum Likelihood from Incomplete Data via the EM algorithm. *Journal of The Royal Statistical Society, Series B*, 1977.
- [53] R. Deriche. Using Canny’s Criteria to Derive a Recursively Implemented Optimal Edge Detector. *Int.l. J. of Comp. Vision*, 1(2):167–187, 1987.
- [54] D. Dimitriadis and P. Maragos. Robust Energy Demodulation Based on Continuous Models with Application to Speech Recognition. In *Eurospeech*, 2003.
- [55] R. Duda, P. Hart, and D. Stork. *Pattern Classification*. Wiley-Interscience, New York, 2001.
- [56] G. Evangelopoulos, I. Kokkinos, and P. Maragos. Advances in Variational Image Segmentation using AM-FM models: Regularized Demodulation and Probabilistic Cue Integration. In *Proc. 3rd. Int.l Workshop on Variational and Level Set Methods (VLSM)*, pages 121–136, 2005.
- [57] L. Fei-Fei, R. Fergus, and P. Perona. One-Shot Learning of Object Categories. *IEEE Trans. PAMI*, 28(4):594–611, 2006.
- [58] P. Felzenswalb and D. Huttenlocher. Pictorial Structures for Object Recognition. *Int.l. J. of Comp. Vision*, 2005.
- [59] P. Felzenswalb and D. Huttenlocher. Efficient Graph Based Image Segmentation. *Int.l. J. of Comp. Vision*, 59(2), 2004.
- [60] R. Fergus. *Visual Object Category Recognition*. PhD thesis, 2005.
- [61] R. Fergus, L. Fei-Fei, P. Perona, and A. Zisserman. Learning Object Categories from Google’s Image Search. In *Proc. Int.l Conf. on Comp. Vision*, 2005.
- [62] R. Fergus, P. Perona, and A. Zisserman. Object Class Recognition by Unsupervised Scale-Invariant Learning. In *Proc. IEEE Conf. on Comp. Vision and Pat. Recogn.*, 2003.
- [63] R. Fergus, P. Perona, and A. Zisserman. A sparse object category model for efficient learning and exhaustive recognition. In *Proc. IEEE Conf. on Comp. Vision and Pat. Recogn.*, 2005.

- [64] D. Field, A. Hayes, and R. Hess. Contour Integration by the Human Visual System: Evidence for a Local 'Association Field'. *Vision Research*, 33(2):173–193, 1993.
- [65] M. Fischler and R. Erschlinger. The Representation and Matching of Pictorial Structures. *IEEE Trans. Computers*, 22(1):67–92, 1973.
- [66] D. Forsyth and J. Ponce. *Computer Vision: A Modern Approach*. Prentice Hall, 2003.
- [67] W. Freeman, E. Pasztor, and O. T. Carmichael. Learning Low-Level Vision. *Int.l. J. of Comp. Vision*, 40(1):25–47, 2000.
- [68] W. T. Freeman and E. H. Adelson. The Design and Use of Steerable Filters. *IEEE Trans. PAMI*, 13(6):891–906, 1991.
- [69] B. Frey and N. Jojic. Estimating Mixture Models of Images and Inferring Spatial Transformations Using the EM Algorithm. In *Proc. IEEE Conf. on Comp. Vision and Pat. Recogn.*, 1999.
- [70] B. Frey and N. Jojic. Transformation-Invariant Clustering Using the EM Algorithm. *IEEE Trans. PAMI*, 25(1):1–17, 2003.
- [71] B. Frey and N. Jojic. A Comparison of Algorithms for Inference and Learning in Probabilistic Graphical Models. *IEEE Trans. PAMI*, 27(9), 2005.
- [72] K. S. Fu. *Syntactic Pattern Recognition*. Prentice-Hall, 1974.
- [73] D. Gabor. Theory of Communication. *Journal of the IEE*, 93(III):429–457, 1946.
- [74] M. Galun, E. Sharon, R. Basri, and A. Brandt. Texture Segmentation by Multiscale Aggregation of Filter Responses and Shape Elements. In *Proc. Int.l Conf. on Comp. Vision*, 2003.
- [75] S. Geman and D. Geman. Stochastic Relaxation, Gibbs Distributions, and the Bayesian Treatment of Images. *IEEE Trans. PAMI*, 66(6):721–741, 1984.
- [76] F. Girosi, M. Jones, and T. Poggio. Regularization Theory and Neural Network Architectures. *Neural Computation*, 7:219–269, 1995.
- [77] N. Graham and A. Sutter. Normalization: Contrast-Gain Control in Simple (Fourier) and Complex (non-Fourier) Pathways of Pattern Vision. *Vision Research*, 40:2737–2761, 2000.
- [78] U. Grenander. *General Pattern Theory*. Oxford University Press, 1993.
- [79] U. Grenander and M. I. Miller. Representation of Knowledge in Complex Systems. *Journal of The Royal Statistical Society, Series B*, 56(4):549–603, 1994.
- [80] W. E. L. Grimson. *Object Recognition by Computer*. MIT Press, 1991.

- [81] R. Gross, I. Matthews, and S. Baker. Constructing and Fitting Active Appearance Models With Occlusion. In *IEEE Workshop on Face Processing in Video*, 2004.
- [82] S. Grossberg. *The Adaptive Brain I: Cognition, Learning, Reinforcement, and Rhythm*. North-Holland, 1987.
- [83] S. Grossberg. *Neural Networks and Natural Intelligence*. MIT Press, 1988.
- [84] S. Grossberg. 3-D Vision and Figure-Ground Separation by Visual cortex. *Perception and Psychophysics*, 55:48–121, 1994.
- [85] S. Grossberg and E. Mingolla. Neural Dynamics of Surface Perception: Boundary Webs, Illuminants, and Shape from Shading. *Computer Vision, Graphics, and Image Processing*, 37:116–165, 1987.
- [86] S. Grossberg, E. Mingolla, and J. Williamson. Synthetic Aperture Radar Processing by a Multiple Scale Neural System for Boundary and Surface Representation. *Neural Networks*, 8(7-8):1005–1028, 1995.
- [87] E. Gumbel, A. Greenwood, and D. Durand. The Circular Normal Distribution: Theory and Tables. *Journal of the American Statistical Association*, 48(261):131–152, 1953.
- [88] C. Guo, S. Zhu, and Y. Wu. A Mathematical Theory of Primal Sketch and Sketchability. In *Proc. Int.l Conf. on Comp. Vision*, 2003.
- [89] G. Guy and G. Medioni. Inferring Global Perceptual Contours from Local Features. *Int.l. J. of Comp. Vision*, 20(1):113–133, 1996.
- [90] P. Hallinan, G. Gordon, A. Yuille, and D. Mumford. *Two- and Three- Dimensional Patterns of the Face*. A. K. Peters, 1999.
- [91] F. Han and S. C. Zhu. Bottom-Up/Top-Down Image Parsing by Attribute Graph Grammar. In *Proc. Int.l Conf. on Comp. Vision*, 2005.
- [92] J. Havlicek and A. Bovik. Image Modulation Models. In A. Bovik, editor, *Handbook of Image and Video Processing*, pages 305–316. 2000.
- [93] J. Havlicek, D. Harding, and A. Bovik. Multidimensional Quasi-Eigenfunction Approximations and Multicomponent AM-FM Models. *IEEE Trans. Im. Proc.*, 9(2):227–242, 2000.
- [94] D. Heeger. Optical Flow Using Spatiotemporal Filters. *Int.l. J. of Comp. Vision*, 1:279–302, 1988.
- [95] B. Heisele, T. Serre, M. Pontil, T. Vetter, and T. Poggio. Categorization by Learning and Combining Object Parts. In *Advances in Neural Information Processing Systems*, 2001.
- [96] J. Hertz, A. Krogh, and R. G. Palmer. *Introduction to the Theory of Neural Computation*. Addison-Wesley, 1989.

- [97] R. Heydt, E. Peterhans, and G. Baumgarthner. Illusory Contours and Cortical Neuron Responses. *Science*, 224(4654):1260–1262, 1984.
- [98] G. Hinton and T. Sejnowski. Learning and Relearning in Boltzmann Machines. In Rumelhart and McClelland, editors, *Parallel Distributed Processing*, volume 1, chapter 7. MIT Press, 1986.
- [99] D. Hoffman and W. Richards. Parts of Recognition. *Cognition*, 18:65–96, 1985.
- [100] J. Hopfield. Neurons with Graded Response have Collective Computational Properties like those of Two-State Neurons. *Proc. Nat.l Academy of Sciences of USA*, 81:3088 – 3092, 1984.
- [101] D. Hubel. *Eye, Brain, and Vision*. Scientific American Library, 1988.
- [102] P. J. Huber. Projection Pursuit. *Ann. Statist.*, 13:435–525, 1985.
- [103] D. Huttenlocher, G. Klanderman, W., and Rucklidge. Comparing Images Using the Hausdorff Distance. *IEEE Trans. PAMI*, 15(9):850–863, 1993.
- [104] A. Ihler, E. Sudderth, W. Freeman, and A. Willsky. Efficient Sampling of Multi-Scale Gaussian Distributions. In *Advances in Neural Information Processing Systems*, 2004.
- [105] S. Ioffe and D. A. Forsyth. Probabilistic Methods for Finding People. *Int.l. J. of Comp. Vision*, 43(1):45–68, 2001.
- [106] M. Isard. Pampas: Real Valued Graphical Models for Computer Vision. In *Proc. IEEE Conf. on Comp. Vision and Pat. Recogn.*, 2003.
- [107] T. Jaakkola. Tutorial on Variational Approximation Methods. In *Advanced Mean Field Methods: Theory and Practice*. MIT Press, 2000.
- [108] R. Jacobs. Methods for Combining Experts' Probability Assesements. *Neural Computation*, 7:867–888, 1995.
- [109] A. K. Jain and F. Farrokhnia. Unsupervised texture segmentation using Gabor filters. *Pattern Recognition*, 24(12):1167–1186, 1991.
- [110] O. Jesorsky, K. Kirchberg, and R. Frischholz. Robust Face Detection Using the Hausdorff Distance. In *AVBPA*, 2001.
- [111] E. Jones and S. Soatto. Layered Active Appearance Models. In *Proc. Int.l Conf. on Comp. Vision*, 2005.
- [112] M. Jones and T. Poggio. Multidimensional Morphable Models: A Framework for Representing and Matching Object Classes. *Int.l. J. of Comp. Vision*, 29(2):107–131, 1998.
- [113] M. Jordan. *Learning in Graphical Models*. MIT Press, 1998.
- [114] M. Jordan. Graphical Models. *Statistical Science*, 19:140–155, 2004.

- [115] M. Jordan and R. Jacobs. Hierarchical Mixtures of Experts and the EM Algorithm. *Neural Computation*, 6(2):181-214, 1994.
- [116] B. Julesz. Textons, the Elements of Texture Perception and their Interactions. *Nature*, 290:91-97, 1981.
- [117] H. J. Kappen. Deterministic Learning Rules for Boltzmann Machines. *Neural Networks*, 8:537-548, 1995.
- [118] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active Contour Models. In *Proc. Int.l Conf. on Comp. Vision*, 1987.
- [119] S. M. Kay. *Fundamentals of Statistical Signal Processing: Estimation & Detection Theory*, volume I&II. Prentice Hall, 1993.
- [120] S. Kichenassamy, A. Kumar, P. Olver, A. Tannenbaum, and A. Yezzi. Gradient Flows and Geometric Active Contour Models. In *Proc. Int.l Conf. on Comp. Vision*, 1995.
- [121] M. Kirby and L. Sirovich. Applications of the Karhunen-Loeve Procedure for the Characterization of Human Faces. *IEEE Trans. PAMI*, 12:103-108, 1990.
- [122] H. Knutsson and C. F. Westin. Normalized Convolution and Differential Convolution: Methods for the Interpolation and Filtering of Incomplete and Uncertain Data. In *Proc. IEEE Conf. on Comp. Vision and Pat. Recogn.*, 1993.
- [123] C. Koch. *Biophysics of Computation: Information Processing in Single Neurons*. Oxford University Press, 1999.
- [124] C. Koch, J. Marroquin, and A. Yuille. Analog Neuronal Networks in Early Vision. Technical Report 751, MIT A.I. Lab, 1985. <ftp://publications.ai.mit.edu/ai-publications/pdf/AIM-751.pdf>.
- [125] I. Kokkinos, R. Deriche, O. Faugeras, and P. Maragos. Computational Analysis and Learning for a Biologically Motivated Model of Boundary Detection and Image Segmentation. *Under preparation*, 2006.
- [126] I. Kokkinos, R. Deriche, P. Maragos, and O. Faugeras. A Biologically Motivated and Computationally Tractable Model of Low and Mid-Level Vision Tasks. In *Proc. Eur. Conf. on Comp. Vision*, 2004.
- [127] I. Kokkinos, R. Deriche, T. Papadopoulo, O. Faugeras, and P. Maragos. Towards Bridging the Gap Between Computational and Biological Segmentation. Technical report, INRIA, 2006. Under preparation.
- [128] I. Kokkinos, G. Evangelopoulos, and P. Maragos. Advances in Texture Analysis: Energy Dominant Components and Multiple Hypothesis Testing. In *Proc. Int.l Conf. on Image Processing*, 2004.
- [129] I. Kokkinos, G. Evangelopoulos, and P. Maragos. Modulation-Feature based Textured Image Segmentation Using Curve Evolution. In *Proc. Int.l Conf. on Image Processing*, 2004.

- [130] I. Kokkinos, G. Evangelopoulos, and P. Maragos. Modulation Features, Generative Models and Weighted Curve Evolution for Texture Analysis and Segmentation. *IEEE Trans. PAMI*, 2006. Submitted for publication.
- [131] I. Kokkinos and P. Maragos. An Expectation Maximization Approach to the Synergy between Object Categorization and Image Segmentation. In *Proc. Int.l Conf. on Comp. Vision*, 2005.
- [132] I. Kokkinos and P. Maragos. Nonlinear Speech Modeling Using Models for Chaotic Systems. *IEEE Trans. Speech and Audio Processing*, 2005.
- [133] I. Kokkinos, P. Maragos, and A. Yuille. Bottom-Up and Top-Down Object Detection Using Primal Sketch Features and Graphical Models. In *Proc. IEEE Conf. on Comp. Vision and Pat. Recogn.*, 2006.
- [134] F. R. Kschischang, B. J. Frey, and H.-A. Loeliger. Factor Graphs and the Sum-Product Algorithm. *IEEE Trans. Information Theory*, 47(2):498–519, 2001.
- [135] M. P. Kumar, P. H. S. Torr, and A. Zisserman. Extending pictorial structures for object recognition. In *Proceedings of the British Machine Vision Conference*, 2004.
- [136] M. P. Kumar, P. H. S. Torr, and A. Zisserman. Obj cut. In *Proc. IEEE Conf. on Comp. Vision and Pat. Recogn.*, 2005.
- [137] J. S. L. Alvarez, J. Weickert. Reliable estimation of dense optical flow fields with large displacements. *Int.l. J. of Comp. Vision*, 39(1):41–56, 2000.
- [138] X. Lan and D. Huttenlocher. Beyond Trees: Common-Factor Models for 2D Human Pose Recovery. In *Proc. Int.l Conf. on Comp. Vision*, 2005.
- [139] Y. Leclerc. Constructing Simple Stable Descriptions for Image Partitioning. *Int.l. J. of Comp. Vision*, 3(1):73–102, 1991.
- [140] T. S. Lee. A Bayesian Framework for Understanding Texture Segmentation in the Primary Visual Cortex. *Vision Research*, 35(18):2643–2657, 1995.
- [141] T. S. Lee. Image Representation Using 2D Gabor Wavelets. *IEEE Trans. PAMI*, 18(10):959–971, 1996.
- [142] T. S. Lee and D. Mumford. Hierarchical Bayesian Inference in the Visual Cortex. *J. Opt. Soc. Amer. (A)*, 20(7), 2003.
- [143] T. S. Lee, D. Mumford, R. Romero, and V. A. Lamme. The Role of the Primary Visual Cortex in Higher Level Vision. *Vision Research*, 38:2429–2454, 1998.
- [144] B. Leibe, A. Leonardis, and B. Schiele. Combined Object Categorization and Segmentation with an Implicit Shape Model. In *Proc. Eur. Conf. on Comp. Vision*, 2004. SLCV workshop.
- [145] M. Leventon, O. Faugeras, and E. Grimson. Statistical Shape Influence in Geodesic Active Contours. In *Proc. IEEE Conf. on Comp. Vision and Pat. Recogn.*, 2000.

- [146] Z. Li. Visual Segmentation by Contextual Influences via Intracortical Interactions in Primary Visual Cortex. *Network: Computation in Neural Systems*, 10(2):187–212, 1999.
- [147] T. Lindeberg. Edge Detection and Ridge Detection with Automatic Scale Selection. *Int.l. J. of Comp. Vision*, 30(2), 1998.
- [148] T. Lindeberg. Feature Detection with Automatic Scale Selection. *Int.l. J. of Comp. Vision*, 30(2), 1998.
- [149] D. Lowe. Distinctive Image Features from Scale-Invariant Keypoints. *Int.l. J. of Comp. Vision*, 60(2), 2004.
- [150] J. Malik, S. Belongie, T. Leung, and J. Shi. Contour and texture analysis for image segmentation. *Int.l. J. of Comp. Vision*, 43(1):7–27, 2001.
- [151] D. Mallat and Z. Zhang. Matching Pursuit in a Time-Frequency Dictionary. *IEEE Trans. Signal Processing*, 41:3397–3415, 1993.
- [152] P. Maragos and A. Bovik. Image Demodulation Using Multidimensional Energy Separation. *J. Opt. Soc. Amer. (A)*, 12(9):1867–1876, 1995.
- [153] D. Marr. *Vision*. W.H. Freeman, 1982.
- [154] D. Martin, C. Fowlkes, and J. Malik. Learning to Detect Natural Image Boundaries Using Local Brightness, Color, and Texture Cues. *IEEE Trans. PAMI*, 26(5):530–549, 2004.
- [155] D. Martin, C. Fowlkes, D. Tal, and J. Malik. A Database of Human Segmented Natural Images and its Application to Evaluating Segmentation Algorithms and Measuring Ecological Statistics. In *Proc. Int.l Conf. on Comp. Vision*, 2001.
- [156] J. Matas, O. Chum, M. Urban, and T. Pajdla. Robust Wide Baseline Stereo from Maximally Stable Extremal Regions. In *Proc. Brit. Mach. Vis. Conf*, 2002.
- [157] I. Matthews and S. Baker. Active Appearance Models Revisited. *Int.l. J. of Comp. Vision*, 60(2):135–164, 2004.
- [158] K. Mikolajczyk and C. Schmid. Scale and Affine Invariant Interest Point Detectors. *Int.l. J. of Comp. Vision*, 60(1), 2004.
- [159] K. Mikolajczyk, A. Zisserman, and C. Schmid". Shape recognition with edge-based features. In *Proc. Brit. Mach. Vis. Conf*, pages 779–788, 2003.
- [160] B. Moghaddam and A. Pentland. Probabilistic Visual Learning for Object Representation. *IEEE Trans. PAMI*, 19(7):693–710, 1997.
- [161] J.-M. Morel and S. Solimini. *Variational Methods in Image Segmentation*. Birkhauser, Boston, 1995.

- [162] G. Mori, X. Ren, A. Efros, and J. Malik. Recovering Human Body Configurations: Combining Segmentation and Recognition. In *Proc. IEEE Conf. on Comp. Vision and Pat. Recogn.*, 2004.
- [163] C. Morrone and D. Burr. Feature Detection in Human Vision: a Phase-Dependent Energy Model. *Proceedings of the Royal Society of London B*, 235:221–245, 1988.
- [164] D. Mumford. Elastica and Computer Vision. In B. J., editor, *Algebraic Geometry and its applications*, pages 507–518. Springer Verlag, 1993.
- [165] D. Mumford. Neuronal Architectures for Pattern Theoretic Problems. In *Large Scale Theories of the Cortex*. MIT Press, 1994.
- [166] D. Mumford. Pattern Theory: A Unifying Approach. In *Perception as Bayesian Inference*. 1996.
- [167] D. Mumford. The Statistical Description of Visual Signals. In *Proc. ICIAM*, 1996.
- [168] D. Mumford and J. Shah. Optimal Approximations by Piecewise Smooth Functions and Associated Variational Problems. *Communications on Pure and Applied Mathematics*, 42(5):577–685, 1989.
- [169] R. Neal. Probabilistic Inference Using Markov Chain Monte Carlo Methods. Technical report, Univ. of Toronto, 1993.
- [170] R. Neal and G. Hinton. A View of the EM Algorithm that Justifies Incremental, Sparse and Other Variants. In M. Jordan, editor, *Learning in Graphical Models*. 1998.
- [171] A. Needham and R. Baillargeon. Effects of prior experience in 4.5-month-old infants object segregation. *Infant Behaviour and Development*, 21:1–24, 1998.
- [172] H. Neumann and W. Sepp. Recurrent V1-V2 Interaction in Early Visual Boundary Processing. *Biological Cybernetics*, 91:425–444, 1999.
- [173] R. Nevatia and K. Babu. Linear feature extraction and description. *Computer Graphics and Image Processing*, 13(3):257–269, July 1980.
- [174] S. Osher and J. Sethian. Fronts Propagating with Curvature-Dependent Speed: Algorithms Based on Hamilton-Jacobi Formulations. *Journal of Computational Physics*, 79:12–49, 1988.
- [175] C. Papageorgiou and T. Poggio. A Trainable System for Object Detection. *Int.l. J. of Comp. Vision*, 38(1):15–33, 2000.
- [176] G. Papandreou and P. Maragos. Multigrid Geometric Active Contour Models. *IEEE Trans. Im. Proc.*, 2006. to appear.
- [177] A. Papoulis. *Probability, Random Variables and Stochastic Processes*. McGraw-Hill, 1988.

- [178] N. Paragios and R. Deriche. Geodesic Active Regions: A new Paradigm to Deal with Frame Partition Problems in Computer Vision. *Journal of Visual Communication and Image Representation*, 3:249–268, 2002.
- [179] N. Paragios and R. Deriche. Geodesic Active Regions and Level Set Methods for Supervised Texture Segmentation. *Int.l. J. of Comp. Vision*, 46(3):223–247, 2002.
- [180] N. Paragios and S. Osher. *Geometric Level Set Methods in Imaging, Vision & Graphics*. Springer, 2002.
- [181] P. Parent and S. Zucker. Trace Inference, Curvature Consistency, And Curve Detection. *IEEE Trans. PAMI*, 11(8):823–839, 1989.
- [182] J. Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufman Publishers, 1988.
- [183] A. Pentland. Automatic Exctraction of Deformable Part Models. *Int.l. J. of Comp. Vision*, pages 107–126, 1990.
- [184] P. Perona. Deformable Kernels for Early Vision. *IEEE Trans. PAMI*, 17(5), 1995.
- [185] P. Perona and W. Freeman. A Factorization Approach to Grouping. In *Proc. Eur. Conf. on Comp. Vision*, pages 655–670, 1998.
- [186] P. Perona and J. Malik. Detecting and Localizing Edges Composed of Steps, Peaks and Roofs. In *Proc. Int.l Conf. on Comp. Vision*, pages 52–57, 1990.
- [187] P. Perona and J. Malik. Scale-Space and Edge Detection Using Anisotropic Diffusion. *IEEE Trans. PAMI*, 12(7):629–639, 1990.
- [188] S. D. Pietra, V. D. Pietra, , and J. Lafferty. Inducing Features of Random Fields. *IEEE Trans. PAMI*, 19(4), 1997.
- [189] T. Poggio and S. Edelman. A Network That Learns to Recognize Three-Dimensional Objects. *Nature*, 343:263–266, 1990.
- [190] R. P. Rao. An Optimal Estimation Approach to Visual Perception and Learning. *Vision Research*, 39:1963–1989, 1999.
- [191] N. Ray, J. Havlicek, S. Acton, and M. Pattichis. Active Contour Segmentation Guided by AM-FM DCA. In *Proc. Int.l Conf. on Image Processing*, 2001.
- [192] M. Riesenhuber and T. Poggio. Hierarchical Models of Object Recognition in Cortex. *Nature Neuroscience*, 2(11):1019–1025, 1999.
- [193] M. Rousson, T. Brox, and R. Deriche. Active Unsupervised Texture Segmentation on a Diffusion Based Space. In *Proc. IEEE Conf. on Comp. Vision and Pat. Recogn.*, 2003.
- [194] M. Rousson and N. Paragios. Shape Priors for Level Set Representations. In *Proc. Eur. Conf. on Comp. Vision*, 2002.

- [195] S. Roweis. EM Algorithms for PCA and SPCA. In *Advances in Neural Information Processing Systems*, 1998.
- [196] H. Rowley, S. Baluja, and T. Kanade. Neural Network-Based Face Detection. *IEEE Trans. PAMI*, 20(1):23–38, 1998.
- [197] B. Sandberg, T. Chan, and L. Vese. A Level-Set and Gabor-based Active Contour Algorithm for Segmenting Textured Images. Technical Report 02-39, UCLA CAM, 2002.
- [198] C. Schmid and R. Mohr. Local Grayvalue Invariants for Object Retrieval. *IEEE Trans. PAMI*, 19(5), 1997.
- [199] H. Schneiderman and T. Kanade. Object Detection Using the Statistics of Parts. *Int.l. J. of Comp. Vision*, 56(3):151–177, 2004.
- [200] J. Sethian. *Level Set Methods*. Cambridge University Press, 1996.
- [201] J. Shah. Riemannian Drums, Curve Evolution and Segmentation. *Journal of Visual Communication and Image Representation*, 11:142–153, 2000.
- [202] J. Shi and J. Malik. Normalized Cuts and Image Segmentation. *IEEE Trans. PAMI*, 22(8):888–905, 2001.
- [203] K. Siddiqi and B. Kimia. Parts of Visual Form: Computational Aspects. *IEEE Trans. PAMI*, 17:239–251, Mar. 1995.
- [204] L. Sigal, M. Isard, R. Sigelman, and M. Black. Attractive people. In *Advances in Neural Information Processing Systems*, 2003.
- [205] E. P. Simoncelli and D. J. Heeger. A Model of Neuronal Responses in Visual Area MT. *Vision Research*, 38(5):743–761, 1998.
- [206] K. Sivakumar and J. Goutsias. Morphologically Constrained GRFs: Applications to Texture Synthesis and Analysis. *IEEE Trans. PAMI*, 21(2), 1999.
- [207] J. Sivic, B. Russell, A. Efros, A. Zisserman, and W. Freeman. Discovering Object Categories in Image Collections. In *Proc. Int.l Conf. on Comp. Vision*, 2005.
- [208] L. Staib and J. Duncan. Boundary Finding with Parametrically Deformable Models. *IEEE Trans. PAMI*, 14(11):1061–1075, 1992.
- [209] E. Sudderth, A. Ihler, W. Freeman, and A. Willsky. Nonparametric belief propagation. In *Proc. IEEE Conf. on Comp. Vision and Pat. Recogn.*, 2003.
- [210] E. Sudderth, A. Torralba, W. Freeman, and A. Willsky. Learning Hierarchical Models of Scenes, Objects, and Parts. In *Proc. Int.l Conf. on Comp. Vision*, 2005.
- [211] B. Sumengen and B. Manjunath. Graph Partitioning Active Contours for Image Segmentation. *IEEE Trans. PAMI*, 28(4):509–521, 2006.

- [212] K. K. Sung and T. Poggio. Example-Based Learning for View-Based Human Face Detection. *IEEE Trans. PAMI*, 20(1):39–51, 1998.
- [213] B. t. H. Romeny, editor. *Geometry Driven Diffusion in Computer Vision*. Kluwer Academic Publishers, 1994.
- [214] T. Takagi and M. Sugeno. Fuzzy identification of systems and its applications to modelling and control. *IEEE Trans. Systems, Man, and Cybernetics*, 15(1):116–132, 1985.
- [215] A. Torralba, K. P. Murphy, and W. T. Freeman. Sharing Visual Features for Multiclass and Multiview Object Detection. In *Proc. IEEE Conf. on Comp. Vision and Pat. Recogn.*, 2004.
- [216] H. V. Trees. *Detection Estimation and Modulation Theory, Part I*. Wiley, 1971.
- [217] A. Tsai, A. Yezzi, and A. Willsky. Curve Evolution Implementation of the Mumford-Shah Functional for Image Segmentation, Denoising, Interpolation, and Magnification. *IEEE Trans. Im. Proc.*, 10(8):1169–1186, 2001.
- [218] Z. Tu, X. Chen, A. Yuille, and S. Zhu. Image Parsing: Unifying Segmentation, Detection, and Recognition. *Int.l. J. of Comp. Vision*, 63(2):113–140, 2005.
- [219] Z. Tu and S. Zhu. Image Segmentation by Data-Driven MCMC. *IEEE Trans. PAMI*, 24(5):657–673, 2002.
- [220] M. Turk and A. Pentland. Eigenfaces for Recognition. *Journal of Cognitive Neuroscience*, 3(1):71–86, 1991.
- [221] S. Ullman. *High-Level Vision: Object Recognition and Visual Cognition*. MIT Press, 1996.
- [222] S. Ullman and R. Basri. Recognition by Linear Combinations of Models. *IEEE Trans. PAMI*, 13:992–1005, 1991.
- [223] S. Ullman, M. Vidal-Naquet, and E. Sali. Visual Features of Intermediate Complexity and their Use in Classification. *Nature Neuroscience*, pages 682–687, 2002.
- [224] V. Vapnik. *Statistical Learning Theory*. Springer Verlag, 1995.
- [225] L. Vese and S. J. Osher. Modeling Textures with Total Variation Minimization and Oscillating Patterns in Image Processing. *Journal of Scientific Computing*, 19(1-3):553–572, 2003.
- [226] T. Vetter, M. Jones, and T. Poggio. A Bootstrapping Algorithm for Learning Linear Models of Object Classes. In *Proc. IEEE Conf. on Comp. Vision and Pat. Recogn.*, 1997.
- [227] P. Viola and M. Jones. Rapid Object Detection using a Boosted Cascade of Simple Features. In *Proc. IEEE Conf. on Comp. Vision and Pat. Recogn.*, 2001.

- [228] M. Wainwright and M. Jordan. Graphical Models, Exponential Families and Variational Inference. Technical report, UC Berkeley, 2003.
- [229] J. Weickert, B. t. H. Romeny, and M. A. Viergever. Efficient and Reliable Schemes for Nonlinear Diffusion Filtering. *IEEE Trans. Im. Proc.*, 7(3):398–410, 1998.
- [230] R. Weinstock. *Calculus of Variations*. Dover, 1974.
- [231] Y. Weiss and E. Adelson. Perceptually Organized EM: a Framework for Motion Estimation that Combines Information about Form and Motion. In *Proc. Int.l Conf. on Comp. Vision*, 1995.
- [232] M. Welling and G. E. Hinton. A New Learning Algorithm for Mean Field Boltzmann Mahines. Technical Report 2001-002, Gatsby Computational Neuroscience Unit, 2001.
- [233] M. Welling, M. Weber, and P. Perona. Unsupervised Learning of Models for Recognition. In *Proc. Eur. Conf. on Comp. Vision*, 2000.
- [234] C. F. Westin, K. Nordberg, and H. Knutsson. On the Equivalence of Normalized Convolution and Normalized Differential Convolution. In *Proc. IEEE Conf. on Comp. Vision and Pat. Recogn.*, 1994.
- [235] L. Williams and D. Jacobs. Stochastic Completion Fields: A Neural Model of Illusory Contour Shape and Saliency. *Neural Computation*, 9(4):837–858, 1997.
- [236] J. Winn and N. Jojic. LOCUS: Learning Object Classes with Unsupervised Segmentation. In *Proc. Int.l Conf. on Comp. Vision*, 2005.
- [237] L. Wiskott and C. Malsburg. A Neural System for the Recognition of Partially Occluded Objects in Cluttered Scenes. *Int.l Journal of Pattern Recognition and Artificial Intellingence*, 7(4):934–948, 1993.
- [238] S. Xu and J. Shi. Object Specific Figure-Ground Segregation. In *Proc. IEEE Conf. on Comp. Vision and Pat. Recogn.*, 2003.
- [239] M. Yang, D. Kriegman, and N. Ahuja. Detecting Faces in Images : A Survey. *IEEE Trans. PAMI*, 24(1):34–58, 2002.
- [240] J. Yedidia, W. Freeman, and Y. Weiss. Understanding Belief Propagation and Its Generalizations. In *Exploring Artificial Intelligence in the New Millenium*, pages 239–263. Elsevier Science & Technology Books, 2003. <http://www.merl.com/reports/docs/TR2001-22.pdf>.
- [241] A. Yezzi, A. Tsai, and A. Willsky. A Statistical Approach to Snakes for Bimodal and Trimodal Imagery. In *Proc. Int.l Conf. on Comp. Vision*, 1999.
- [242] A. Yuille. Energy Functions for Early Vision and Analog Networks. Technical Report 987, MIT A.I. Lab, 1987. <ftp://publications.ai.mit.edu/ai-publications/500-999/AIM-987.ps>.

- [243] A. Yuille. Winner-Take-All Networks. In M. Arbib, editor, *The Handbook of Brain Theory and Neural Networks*. MIT Press, 1995.
- [244] A. Yuille. CCCP Algorithms to Minimize the Bethe and Kikuchi Free Energies: Convergent Alternatives to Belief Propagation. *Neural Computation*, 14:1691-1772, 2001.
- [245] A. Yuille and D. Geiger. A Common Framework for Image Segmentation. *Int.l. J. of Comp. Vision*, 6:227-243, 1991.
- [246] A. L. Yuille, J. M. Coughlan, and D. Kersten. *Computational Vision: Principles of Perceptual Inference*. <http://gandalf.psych.umn.edu/~kersten/kersten-lab/courses/Psy8970/psy8993.html>. Book under preparation.
- [247] A. L. Yuille, P. W. Hallinan, and D. S. Cohen. Feature Extraction from Faces Using Deformable Templates. *Int.l. J. of Comp. Vision*, 8(2):99-111, 1992.
- [248] S. Zhu and A. Yuille. Region Competition: Unifying Snakes, Region Growing and Bayes/MDL for Multiband Image Segmentation. *IEEE Trans. PAMI*, 18(9):884-900, 1996.
- [249] S. C. Zhu. Statistical Modeling and Conceptualization of Visual Patterns. *IEEE Trans. PAMI*, 25(6):1-22, 2003.
- [250] S. C. Zhu, C. E. Guo, Y. Wang, and Z. Xu. What are Textons? *Int.l. J. of Comp. Vision*, 62(1):121-143, 2005.
- [251] S. C. Zhu, Y. N. Wu, and D. Mumford. Filters, Random Fields and Maximum Entropy (FRAME): Towards a Unified Theory for Texture Modeling. *Int.l. J. of Comp. Vision*, 27(2):107-126, 1998.
- [252] S. C. Zhu, R. Zhang, and Z. W. Tu. Integrating Top-Down/Bottom-Up for Object Recognition by Data-Driven Markov Chain Monte Carlo. In *Proc. IEEE Conf. on Comp. Vision and Pat. Recogn.*, 2000.

Βιογραφικό Σημείωμα

Σπουδές

- Διδακτορικό, Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών Εθνικό Μετσόβιο Πολυτεχνείο, 2006.
Επιβλέπων: Καθ. Πέτρος Μαραγκός.
Τίτλος Διδακτορικής Διατριβής: *Συνέργεια μεταξύ Κατάτμησης Εικόνων και Αναγνώρισης Αντικειμένων με Γεωμετρικές και Στατιστικές Τεχνικές της Όρασης Υπολογιστών.*
- Δίπλωμα Μηχανικού, Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών Εθνικό Μετσόβιο Πολυτεχνείο, 2001.
Τίτλος Διπλωματικής Εργασίας: *Πρόβλεψη και Ανάλυση Σημάτος Φωνής με Μεθόδους Ανάλυσης Χαστικών Χρονοσειρών.*

Ερευνητικά Ενδιαφέροντα

- **Όραση Υπολογιστών και Αναγνώριση Προτύπων:**
Στατιστικά Μοντέλα για Ανίχνευση και Αναγνώριση Αντικειμένων.
Πιθανοτική Συμπερασματολογία και Εκμάθηση για Γραφικά Μοντέλα.
Εξαγωγή Χαρακτηριστικών και Ανίχνευση Σημείων Ενδιαφέροντος.
Βιολογικά Μοντέλα για Προβλήματα Όρασης Χαμηλού και Μέσου Επιπέδου.
- **Επεξεργασία Εικόνας και Σημάτων:**
Κατάτμηση Εικόνων με Μεταβολικές Μεθόδους.
Ανάλυση και Κατάτμηση Υφής με Μοντέλα Διαμόρφωσης.
Μη Γραμμικά Δυναμικά Μοντέλα για Επεξεργασία Φωνής και Σημάτων.
Θεωρία Ανίχνευσης και Εκτίμησης, Μη-παραμετρική Ιχνηλάτηση.

Διακρίσεις-Βραβεία

- Βραβείο Πάρη Κανελλάκη για υψηλότερες επιδόσεις σε μαθήματα κατεύθυνσης πληροφορικής.
- Απέκτησε σε 4 έτη το 5-ετές Δίπλωμα της Σχολής ΗΜΜΥ του ΕΜΠ, κατατασσόμενος 7^{ος} μεταξύ 330. Βαθμός Αποφοίτησης: 9.34/10.
- Υποτροφία Ιδρύματος Μποδοσάκη ως μεταπτυχιακός φοιτητής.
- Υποτροφία ΙΚΥ ως προπτυχιακός βάσει ακαδημαϊκών επιδόσεων.

Ερευνητική Εμπειρία

- 2001-2006 Μεταπτυχιακός Ερευνητικός Βοηθός, Σχολή ΗΜΜΥ, ΕΜΠ.
- Συμμετοχή στο Πρόγραμμα Βασικής Έρευνας ΗΡΑΚΛΕΙΤΟΣ.
- Συμμετοχή στο Ερευνητικό Δίκτυο Αριστείας MUSCLE.
- Επίσκεψη στην ομάδα Odyssee, INRIA, Sophia Antipolis, για συνεργασία σε βιολογικά μοντέλα για προβλήματα όρασης.
- Επίσκεψη στο Center for Image and Vision Sciences του UCLA για συνεργασία σε προβλήματα αναγνώρισης.

Διδακτική Εμπειρία

2004-2006 Μεταπτυχιακός Βοηθός Διδασκαλίας, 'Σήματα και Συστήματα'.

1999-2000 Βοηθός Εργαστηρίου, 'Εισαγωγή στον Προγραμματισμό'.

Επαγγελματικές Δραστηριότητες

Κριτής Περιοδικών:

- IEEE Transactions on Image Processing.
- IEEE Transactions on Pattern Analysis and Machine Intelligence.

Προσωπικά

- Ημερομηνία και Τόπος Γεννήσεως: 8 Ιανουαρίου 1980, Χολαργός Αττικής.
- Εθνικότητα: Ελληνική.
- Οικογενειακή Κατάσταση: Άγαμος.
- Μέλος της Διεθνούς Ένωσης Ηλεκτρολόγων (IEEE) και του Τεχνικού Επιμελητήριου Ελλάδας.
- Ξένες Γλώσσες:

Αγγλικά: Certificate of Proficiency in English.

Γαλλικά: Diplome Approfondi de Langue Francaise.

Γερμανικά: Kleines Deutches Sprachdiplom.

Κατάλογος Δημοσιεύσεων

Περιοδικά

- 1 I. Kokkinos and P. Maragos, 'Nonlinear Speech Analysis Using Models for Chaotic Systems.', **IEEE Trans. on Speech and Audio Processing**, Vol. 13(6), pp. 1098-1109, 2005.
- 2 I. Kokkinos, G. Evangelopoulos and P. Maragos, 'Texture Analysis and Segmentation Using Modulation Features, Generative Models and Weighted Curve Evolution.', **IEEE Trans. on Pattern Analysis and Machine Intelligence**, revised & resubmitted, 2006.
- 3 I. Kokkinos and P. Maragos, 'Synergy Between Image Segmentation and Object Recognition Using the Expectation Maximization Algorithm.', submitted.
- 4 I. Kokkinos, R. Deriche, O. Faugeras and P. Maragos, 'Computational Analysis and Learning for a Biologically Motivated Model of Boundary Detection.', submitted.

Κύρια Συνέδρια Όρασης Υπολογιστών.

- 5 I. Kokkinos, P. Maragos and A. Yuille, 'Bottom-Up and Top-Down Object Detection Using Primal Sketch Features and Graphical Models.', In Proc. **IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)**, 2006.
- 6 I. Kokkinos and P. Maragos, 'An Expectation Maximization Approach to the Synergy Between Image Segmentation and Object Categorization.', In Proc. **IEEE Int'l. Conf. on Computer Vision (ICCV)**, 2005.
- 7 I. Kokkinos, R. Deriche, P. Maragos and O. Faugeras, 'A Biologically Motivated and Computationally Tractable Model of Low- and Mid- Level Vision Tasks.', In Proc. **European Conference on Computer Vision (ECCV)**, 2004.

Διεθνή Συνέδρια με κριτές

- 8 I. Kokkinos and P. Maragos, 'A Detection-Theoretic Approach to Texture and Edge Discrimination.', In Proc. **4th Int'l. Workshop on Texture Analysis and Synthesis**, in conjunction with ICCV 2005.
- 9 G. Evangelopoulos, I. Kokkinos and P. Maragos, 'Advances in Variational Image Segmentation using AM-FM models: Regularized Demodulation and Probabilistic Cue Integration.', In Proc. **3rd IEEE Variational and Level-Set Methods (VLSM) Workshop**, in conjunction with ICCV 05.

- 10 I. Kokkinos, G. Evangelopoulos and P. Maragos, '*Advances in Texture Analysis: Energy Dominant Component & Multiple Hypothesis Testing.*', In Proc. **IEEE Int'l. Conf. on Image Processing**, 2004.
- 11 I. Kokkinos, G. Evangelopoulos and P. Maragos, '*Modulation-Feature based Textured Image Segmentation Using Curve Evolution.*', In Proc. **IEEE Int'l. Conf. on Image Processing**, 2004.
- 12 V. Pitsikalis, I. Kokkinos and P. Maragos, '*Nonlinear Analysis of Speech Signals: Generalized Dimensions and Lyapunov Exponents.*', In Proc. **European Conference on Speech Communication and Technology (EURO-SPEECH)**, 2003.
- 13 P. Maragos, A. Dimakis and I. Kokkinos. '*Some Advances in Nonlinear Speech Modeling Using Modulations Fractals and Chaos.*' In Proc. **IEEE Int'l. Conf. on Digital Signal Processing**, 2002.