



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ
ΥΠΟΛΟΓΙΣΤΩΝ
ΤΟΜΕΑΣ ΣΗΜΑΤΩΝ, ΕΛΕΓΧΟΥ ΚΑΙ ΡΟΜΠΟΤΙΚΗΣ

Υπολογιστική Μοντελοποίηση
της Οπτικής Προσοχής

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

ΤΟΥ

Άλκη Μ. Χατζηιωσήφ

Επιβλέπων: Πέτρος Α. Μαραγκός
Καθηγητής Ε.Μ.Π.

Αθήνα, Ιούλιος 2008



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ
ΥΠΟΛΟΓΙΣΤΩΝ
ΤΟΜΕΑΣ ΣΗΜΑΤΩΝ, ΕΛΕΓΧΟΥ ΚΑΙ ΡΟΜΠΟΤΙΚΗΣ

Υπολογιστική Μοντελοποίηση
της Οπτικής Προσοχής

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

του

Άλκη Μ. Χατζηιωσήφ

Επιβλέπων: Πέτρος Α. Μαραγκός
Καθηγητής Ε.Μ.Π.

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή στις 22 Ιουλίου 2008

.....
Πέτρος Μαραγκός
Καθηγητής Ε.Μ.Π.

.....
Κωνσταντίνος Κόλλιας
Καθηγητής Ε.Μ.Π.

.....
Κωνσταντίνος Τζαφέστας
Λέκτορας Ε.Μ.Π.

Αθήνα, Ιούλιος 2008.

.....
Άλκης Μ. Χατζηιωσήφ

Διπλωματούχος Ηλεκτρολόγος Μηχανικός και Μηχανικός Υπολογιστών Ε.Μ.Π.

Copyright© Άλκης Μ. Χατζηιωσήφ, 2008.

Με επιφύλαξη παντός δικαιώματος.All rights reserved.

Απαγορεύεται η αντιγραφή, αποθήκευση και διανομή της παρούσας εργασίας, εξ ολοκλήρου ή τμήματος αυτής, για εμπορικό σκοπό. Επιτρέπεται η ανατύπωση, αποθήκευση και διανομή για σκοπό μη κερδοσκοπικό, εκπαιδευτικής ή ερευνητικής φύσης, υπό την προϋπόθεση να αναφέρεται η πηγή προέλευσης και να διατηρείται το παρόν μήνυμα. Ερωτήματα που αφορούν τη χρήση της εργασίας για κερδοσκοπικό σκοπό πρέπει να απευθύνονται προς το συγγραφέα.

Οι απόψεις και τα συμπεράσματα που περιέχονται σε αυτό το έγγραφο εκφράζουν το συγγραφέα και δεν πρέπει να ερμηνευθεί ότι αντιπροσωπεύουν τις επίσημες θέσεις του Εθνικού Μετσόβιου Πολυτεχνείου.

Περίληψη

Θέμα της παρούσας διπλωματικής εργασίας αποτελεί η μελέτη και μοντελοποίηση της οπτικής προσοχής. Οι μηχανισμοί προσοχής κατέχουν καίριο ρόλο στη λειτουργία και επιβίωση των ζωντανών οργανισμών, ανάμεσα σε αυτούς και του ανθρώπου, καθ' ότι σμικρύνουν δραστικά τη διάσταση της πληροφορίας που δέχεται από το περιβάλλον του, κρατώντας μόνο το σημαντικότερό της κομμάτι και επιτρέποντας την περαιτέρω της επεξεργασία. Εμπνεόμενοι από τέτοια βιολογικά πρότυπα, η δομή και οι μηχανισμοί των οποίων αποκαλύπτονται όλο και σε μεγαλύτερη λεπτομέρεια από επιστήμες όπως η νευροφυσιολογία, οι επιστήμονες - μηχανικοί των υπολογισμών και της Όρασης Υπολογιστών αναπτύσσουν διάφορα μοντέλα με σκοπό να προσεγγίσουν την αντίστοιχη βιολογική συμπεριφορά.

Βασιζόμενοι πάνω σε ένα από τα πλέον επιτυχημένα υπολογιστικά μοντέλα προσέγγισης της οπτικής προσοχής, την αρχιτεκτονική των Itti, Koch και Ullman, διερευνούμε, προτείνουμε και αξιολογούμε δυνατότητες αλλαγών, προσθηκών και επεκτάσεων πάνω σε αυτήν. Ξεκινούμε αρχικά εξετάζοντας την υλοποίηση της αρχιτεκτονικής αυτής χρησιμοποιώντας μορφολογικές αντί γραμμικές πολυκλιμακωτές αναπαραστάσεις. Στη συνέχεια, κινητοποιημένοι απο ενδείξεις για τη σημασία των γωνιών και συνδέσμων ως ελκυστών της οπτικής προσοχής, προτείνουμε την προσθήκη ενός σχετικού ανιχνευτή μέσα στο μοντέλο. Σε κάθε βήμα αξιολογούμε την προσπάθεια στη βάση πραγματικών ανθρώπινων αποκρίσεων. Τέλος, επεκτείνοντας την αρχική αρχιτεκτονική των Itti et al. σε τρεις διαστάσεις, εξετάζουμε την εύρεση χωροχρονικών, αντί απλά χωρικών, σημείων και περιοχών ενδιαφέροντος, θεωρώντας το χρόνο ως την τρίτη διάσταση. Παράλληλα, προτείνουμε ένα νέο γενικό μοντέλο βασισμένο στην έκπληξη που επιφέρουν απροσδόκητες αλλαγές στον χάρτη οπτικής προσοχής που εκφράζει την σκηνή που παρατηρούμε κάθε στιγμή.

Σκοπός της εργασίας είναι αφενός η συνοπτική αλλά συγκροτημένη παρουσίαση της έννοιας της οπτικής προσοχής στον αναγνώστη, αφετέρου η παρουσίαση και τεκμηρίωση, κατά το δυνατόν, ιδεών πάνω στην υπολογιστική της προσέγγιση. Ανάμεσα σε αυτές τις ιδέες υπάρχουν κάποιες οι οποίες τυγχάνουν

βαθύτερης διερεύνησης, άλλες δε, επιδέχονται μελλοντικό ραφινάρισμα το οποίο δεν μπορεί να λάβει χώρα στα στενά χρονικά περιθώρια μιας διπλωματικής εργασίας ενώ, τέλος, κάποιες αναφέρονται ως, ελπίζουμε δόκιμη, τροφή για σκέψη.

Λέξεις—κλειδιά

όραση υπολογιστών, οπτική προσοχή, βιολογικά εμπνεόμενοι υπολογιστικοί αλγόριθμοι, ανοδικοί και καθοδικοί μηχανισμοί προσοχής, χωρικά και χωρο-χρονικά σημεία ενδιαφέροντος, πολυκλιμακωτή ανάλυση, ανίχνευση γωνιών, αντίθεση, χρωματικός ανταγωνισμός

Abstract

The present diploma thesis deals with the study and modelling of visual saliency and attention. Attention mechanisms in general play a crucial role in the survival of living creatures, including humans. By drastically reducing the size of the information incoming, through the senses, from their surroundings, living organisms become able for its further, more detailed processing. Inspired by these biological mechanisms, neurophysiologists and computational scientists alike strive to develop architectures which model their behaviour.

We initially begin with studying one of the probably the most successful computational model of visual attention, the one developed by Itti, Koch and Ullman. Several possibilities exist for the slight alteration, enhancement or expansion of this model. Such modifications include the use of morphological, rather than linear, multiresolution analysis schemes or the addition of a corner detector. We outline the utility of the latter by citing relevant results from human experiments which indicate a strong connection between corners and perceptually salient regions of an image. All results are being evaluated on the basis of model human responses. Finally, we explore the expansion of the Itti et al. model to include time as a third dimension. In parallel, we present a new space-time general saliency model based on surprise resulting from unexpected changes on the saliency map.

It is the intention of this work to offer a comprehensive presentation of visual saliency, but also to present and, wherever possible, to justify a number of ideas regarding its computational modelling. Some of them are explored in detail, while others remain ripe for refinement — a process not always able to take place within the short timeframe of a diploma thesis. For these, we hope that they provide valuable food for thought.

Keywords

computer vision, visual attention, saliency, biologically inspired systems, computational algorithms, bottom-up and top-down attention mechanisms,

spatial and space-time interest points, multiscale image analysis, corner detection, contrast, color opponency

Ευχαριστίες

Η διπλωματική εργασία αποτελεί το προϊόν της ενασχόλησής μου με το αντικείμενο της Όρασης Υπολογιστών, η οποία ξεκίνησε τον Μάιο του 2007 όταν είχα την ευκαιρία να παρακολουθήσω το σχετικό μάθημα του καθηγητή Πέτρου Μαραγκού. Παρά τη στενότητα του χρόνου – απόρροια της ιδιόρρυθμης κατάστασης του ελληνικού πανεπιστημίου το τότε ακαδημαϊκό έτος – ο ζήλος και ο ενθουσιασμός με τον οποίο διαπραγματευόταν το αντικείμενο με έλκυσαν τόσο ώστε να το επιλέξω ως τομέα της διπλωματικής εργασίας. Θα ήθελα να τον ευχαριστήσω θερμά για αυτό όσο και για το χρόνο που αφιέρωσε κατά τη διάρκεια της εκπόνησής της.

Οι εκτενής συζήτηση του θέματος με τον Κώστα Ραπαντζίκο, ο οποίος πρόσφατα ολοκλήρωσε τη διδακτορική του διατριβή πάνω σε ένα κοντινό θέμα, όπως και με τον Γιώργο Ευαγγελόπουλο με βοήθησαν να αντιληφθώ πολλά καίρια σημεία του αντικειμένου. Επίσης, πολύτιμη στάθηκε και η βοήθεια των παιδιών στο εργαστήριο.

Θα ήθελα να ευχαριστήσω επίσης τον μεταδιδακτορικό ερευνητή του εργαστηρίου TCTS της Πολυτεχνικής της Mons, στο Βέλγιο, ο οποίος παρείχε το λογισμικό και την τεχνική υποστήριξη για το μηχανισμό εξαγωγής των πρότυπων ανθρώπινων αποκρίσεων οπτικής προσοχής για σύγκριση με τις αντίστοιχες υπολογισμένες τιμές. Βέβαια, τίποτα από αυτό δε θα είχε γίνει χωρίς τη βοήθεια 30 και πλέον συμφοιτητών και φίλων, οι οποίοι πρόθυμα συμμετείχαν στην έρευνα αυτή.

Δράττω την παρούσα ευκαιρία για να εκφράσω τις ευχαριστίες μου προς τους καθηγητές που με βοήθησαν στην επιλογή μου να συνεχίσω τις μεταπτυχιακές μου σπουδές στο εξωτερικό. Ανάμεσα σε αυτούς, τον επιβλέποντα της διπλωματικής μου εργασίας, καθηγητή Πέτρο Μαραγκό, τόσο για τις συμβουλές όσο και για τις συστάσεις του· τους κκ. Τρύφωνα Κουσιουρή και Σπύρο Τζαφέστα, για τον ίδιο λόγο· τον κ. Κωνσταντίνο Τζαφέστα, για τις πολλές ώρες που αφιέρωσε στις συζητήσεις μας για το θέμα.

Τίποτα από όλα αυτά δε θα ήταν πραγματικότητα χωρίς την αμέριστη υποστήριξη που μου προσέφερε και προσφέρει η οικογένειά μου, καθ' όλη τη ζωή και μέχρι τώρα ακαδημαϊκή μου πορεία. Η καλή θέληση, η υπομονή, η

κατανόηση και η αγάπη που τόσο γενναιόδωρα μου προσέφεραν αποτέλεσε εχέγγυο για την δημιουργία της παρούσας εργασίας αλλά και για την επίτευξη κάθε στόχου που έχω θέσει στη ζωή μου. Γι αυτό, τους χρωστάω αιώνια ευγνωμοσύνη και ως ελάχιστη τιμή, αφιερώνω το παρόν πόνημα.

Περιεχόμενα

1	Εισαγωγή	24
1.1	Γενικά για την Όραση Υπολογιστών ¹	24
1.2	Περί Προσοχής ¹	25
1.3	Οπτική Προσοχή	26
1.4	Εφαρμογές	27
1.5	Διάρθρωση της Διπλωματικής Εργασίας	30
2	Βιολογικό και μαθηματικό υπόβαθρο	32
2.1	Βιολογικά μοντέλα οπτικής προσοχής	32
2.1.1	Το Ανθρώπινο Σύστημα Όρασης	33
2.1.2	Ανίχνευση Χαρακτηριστικών	36
2.1.3	Προσεκτική (attentive) και Πρώιμη (preattentive) Οπτική Αντίληψη	38
2.2	Ανάλυση Εικόνων σε Πολλαπλές Κλίμακες	41
2.2.1	Γκαουσιανή Πυραμίδα	44
2.2.2	Adjunction Πυραμίδα	45
2.2.3	Open-Close Πυραμίδα	47
2.2.4	Haar - Wavelet Πυραμίδα	50
2.2.5	Leveling Πυραμίδα	53
3	Υπολογιστικά Μοντέλα Οπτικής Προσοχής	63
3.1	Γενικά	63
3.2	Το μοντέλο των Itti, Koch και Niebur	66
3.2.1	Εξαγωγή Πρωταρχικών Οπτικών Χαρακτηριστικών	68
3.2.2	Συγκρότηση του Χάρτη Οπτικής Προσοχής	70
3.3	Μεθοδολογία Σύγκρισης Αποκρίσεων	74
3.3.1	Παρακολούθηση της Ανθρώπινης Οπτικής Προσοχής	77
3.3.2	Μέτρο της ομοιότητας μεταξύ των αποκρίσεων	80
3.4	Χρήση Διαφορετικών Πυραμίδων	81

3.5	Ανίχνευση Γωνιών	91
3.5.1	Ο Ανιχνευτής Harris	92
3.5.2	Σύγκριση Ανιχνευτή Harris και του Μοντέλου Οπτικής Προσοχής	93
3.5.3	Διερεύνηση της Ενσωμάτωσης του Ανιχνευτή στο Μοντέλο Οπτικής Προσοχής	96
3.6	Σχετικά με την Κανονικοποίηση	97
4	Χωροχρονικά Μοντέλα	106
4.1	Ανοδικά Χωροχρονικά Μοντέλα	106
4.1.1	Χωροχρονική Επέκταση του Μοντέλου των Itti, Koch και Niebur από τους Rapantzikos et al.	107
4.1.2	Απευθείας Χωροχρονική Επέκταση του Μοντέλου των Itti et al.	109
4.1.3	Χρήσεις, Περιορισμοί και Προβλήματα	111
4.1.4	Περαιτέρω Επεκτάσεις	114
4.2	Εντοπισμός Ενδιαφερόντων Πλαισίων σε Ακολουθίες	115
4.2.1	Φιλοσοφία του Αλγορίθμου	116
4.2.2	Διάφορες Μεθοδολογίες Αξιολόγησης των Αλγορίθμων	117
4.3	Αλγόριθμοι Εύρεσης Ενδιαφέροντος Κάθε Πλαισίου	119
4.3.1	Προσδιορισμός Οπτικής Ροής Μεταξύ Πλαισίων	119
4.3.2	Μεθοδολογίες προς Σύγκριση	121
4.3.3	Σχόλια	125
5	Συμπεράσματα	126
5.1	Συμβολή της Διπλωματικής Εργασίας	126
5.2	Μελλοντικές Κατευθύνσεις	127
	Α' Εικόνες που Εξετάστηκαν	130

Κατάλογος Σχημάτων

- 1.1 Παραδείγματα εφαρμογών της Όρασης Υπολογιστών. Από αριστερά: οπτικά καθοδηγούμενο ρομπότ· βιοιατρική εικόνα· εύρεση αντικειμένων (εδώ ανθρώπινων μορφών, από την εργασία στο [24])· αποκατάσταση χαμένων μερών από αρχαίες τοιχογραφίες, από το [37]. 25
- 1.2 Αρχική εικόνα (πάνω αριστερά) και καταγεγραμμένες κινήσεις του βλέμματος υπό διαφορετικές οδηγίες – 1: ελεύθερη εξερεύνηση της εικόνας, 2: εξακριβώστε την κοινωνική θέση των ατόμων, 3: εξακριβώστε την ηλικία των ατόμων, 4: σκεφτείτε τι συνέβαινε πριν την άφιξη του επισκέπτη, 5: απομνημονεύσετε τα ρούχα των ατόμων στη σκηνή, 6: απομνημονεύσετε τη θέση των ατόμων και αντικειμένων, 7: υπολογίστε πόσο χρόνο ο επισκέπτης ήταν μακριά από την οικογένεια. Από το [47]. 28
- 2.1 Τομή του ανθρώπινου οφθαλμού 34
- 2.2 Διάγραμμα του ανθρώπινου οπτικού συστήματος. Φαίνονται οι οφθαλμοί (πάνω), ο έξω γονατώδης πυρήνας (περίπου στο κέντρο), ο οπτικός φλοιός (στο κάτω μέρος) και το νωτιαίο και πρόσθιο ρεύμα. Από το Gray's Anatomy, έκδοση 1918. 35

- 2.3 Αντιστοίχιση δισδιάστατων φίλτρων Gabor με αποκρίσεις κυττάρων του οπτικού φλοιού, εικόνα από Daugman, 1985 (αναπαραγωγή από το [1]). Ο Daugman αναφέρει σχετικά: “Πάνω σειρά: Πειραματικά καταγεγραμμένα 2Δ προφίλ των δεκτικών πεδίων (σημ. receptive fields) τριών απλών κυττάρων από ραβδωτό φλοιό γάτας, από τους L. A. Palmer και J. P. Jones της Ιατρικής Σχολής του Πανεπιστημίου της Πεννσυλβάνια. Κάθε σχήμα δείχνει το ενεργοποιητικό ή ανασταλτικό αποτέλεσμα ενός μικρού φωτός ή σκοτεινού σημείου που αναβοσβήνει, με ρυθμό πυροδότησης ίσο με αυτόν του κυττάρου, σαν μια συνάρτηση της (x, y) θέσης του ερεθίσματος, υπολογισμένη με αντίστροφη συσχέτιση του 2Δ ερεθίσματος με την ακολουθία των νευρικών αποκρίσεων. Η δεύτερη σειρά δείχνει τη βέλτιστη 2Δ συνάρτηση Gabor για το προφίλ του δεκτικού πεδίου κάθε κυττάρου, με τις παραμέτρους υπολογισμένες με τη μέθοδο των ελαχίστων τετραγώνων. Η τρίτη σειρά δείχνει το σφάλμα - υπόλοιπο μεταξύ την καταγεγραμμένη απόκριση κάθε κυττάρου και την αντίστοιχη συνάρτηση Gabor. Σε τυποποιημένους στατιστικούς ελέγχους, τα υπόλοιπα δεν διακρίνονταν από τυχαία σφάλματα σε 33 από τα 36 απλά κύτταρα που ελέγχθηκαν”. . . . 37
- 2.4 Πάνω σειρά (από τους Healy et al., 1996 [13]) – αριστερά: εικόνα στην οποία ο πρώιμος μηχανικός προσοχής δίνει άμεσο αποτέλεσμα για το αξιοπρόσεκτο σημείο. Δεξιά: παράδειγμα προσεκτικής οπτικής προσοχής – στην εικόνα οι distractors εμποδίζουν αποτελεσματικά την άμεση εύρεση του αξιοπρόσεκτου αντικειμένου αφού αυτό δε διαθέτει μοναδικά χαρακτηριστικά ώστε να ξεχωρίζει από αυτούς, απαιτώντας έτσι συνειδητή προσπάθεια. Κάτω σειρά – αριστερά: εικόνα όπου οι περισσότεροι παρατηρητές μπορούν, αρκετά εύκολα (αν και όχι τελείως αστραπιαία) να διακρίνουν το εξέχον αντικείμενο βάσει των πρωτογενών – πρώιμων του χαρακτηριστικών (από το [30]). Δεξιά: παρόμοια εικόνα, όπου η εύρεση του καθίσταται πιο δύσκολη και αρχίζει να απαιτεί κάποια συνειδητή προσπάθεια, καθ’ ότι τα δευτερεύοντα ερεθίσματα (εδώ, χρωματικές διαφορές) αποσπούν μεγάλο μέρος της αρχικής προσοχής. 39
- 2.5 Στην εικόνα αυτή του Rubin διακρίνουμε είτε ένα λευκό βάζο είτε δύο πρόσωπα σε προφίλ, αλλά όχι και τα δύο ταυτόχρονα – στρατηγική “ όλα - για - το - νικητή ” (winner-take-all) 41

- 2.6 Γκαουσιανή (αριστερά) και Λαπλασιανή (δεξιά) πυραμίδα (και πρώτα τέσσερα επίπεδα) της εικόνας *harbor* την αρχική μορφή της οποίας βλέπουμε στο επίπεδο 0 της Γκαουσιανής πάνω αριστερά. Η Λαπλασιανή πυραμίδα έχει κανονικοποιηθεί μεταξύ των ελάχιστων και μέγιστων τιμών κάθε επιπέδου ώστε να είναι πιο ευδιάκριτη. Βλέπουμε ότι καθώς ανεβαίνουμε στα επίπεδα, τα όρια των αντικειμένων γίνονται λιγότερο σαφή. 46
- 2.7 Adjunction (αριστερά) και αντίστοιχη “Λαπλασιανή” L^A (δεξιά) πυραμίδα (και πρώτα τέσσερα επίπεδα) της εικόνας *harbor* την αρχική μορφή της οποίας βλέπουμε στο επίπεδο 0 της Adjunction πυραμίδας πάνω αριστερά. Η L^A πυραμίδα έχει κανονικοποιηθεί μεταξύ των ελάχιστων και μέγιστων τιμών κάθε επιπέδου ώστε να είναι πιο ευδιάκριτη. Είναι εμφανής η διατήρηση των κύριων γεωμετρικών χαρακτηριστικών στις μεγαλύτερες κλίμακες σε αντίθεση με την αντίστοιχη Γκαουσιανή (συγκρίνετε με την εικόνα 2.6) 48
- 2.8 Παράδειγμα εφαρμογής των τελεστών διάβρωσης, διαστολής, ανοίγματος και κλεισίματος πάνω σε δυαδική εικόνα με σχήμα νησιού (Αστυπάλαια). Πάνω: αρχική εικόνα, μορφοποιητικό στοιχείο S και διάβρωσή της από αυτό (erosion). Κάτω: κατά σειρά διαστολή (dilation), άνοιγμα (opening) και κλείσιμο (closing) της αρχικής εικόνας με το S 49
- 2.9 Πυραμίδα ανοίγματος - κλεισίματος (open-close) (αριστερά) και αντίστοιχη “Λαπλασιανή” L^M (δεξιά) μαζί με τα πρώτα τέσσερα επίπεδα από πάνω προς τα κάτω, για την εικόνα *harbor* την αρχική μορφή της οποίας βλέπουμε στο επίπεδο 0 της (open-close) πάνω αριστερά. Η L^M πυραμίδα έχει κανονικοποιηθεί μεταξύ των ελάχιστων και μέγιστων τιμών κάθε επιπέδου ώστε να είναι πιο ευδιάκριτη. Εδώ, όπως και στην Adjunction πυραμίδα βλέπουμε τη διατήρηση των γεωμετρικών χαρακτηριστικών στις υψηλές κλίμακες. 51
- 2.10 Βήματα υπολογισμού του δισδιάστατου μετασχηματισμού κυματιδίων 52
- 2.11 Ζώνες του μετασχηματισμού κυματιδίων για την γκρίζα εικόνα *glass*. Αριστερά: αρχική εικόνα. Στη συνέχεια, κατά σειρά: ζώνη LL, ζώνη LH, ζώνη HL και ζώνη HH 53

- 2.12 Αριστερά: πυραμίδα κυματιδίων Haar (Haar wavelet), Δεξιά: αντίστοιχη “ Λαπλασιανή ” L^H , μαζί με τα πρώτα τέσσερα επίπεδα από πάνω προς τα κάτω, για την εικόνα *harbor* την αρχική μορφή της οποίας βλέπουμε στο επίπεδο 0 της πυραμίδας κυματιδίων (Haar) πάνω αριστερά. Η L^M πυραμίδα έχει κανονικοποιηθεί μεταξύ των ελάχιστων και μέγιστων τιμών κάθε επιπέδου ώστε να είναι πιο ευδιάκριτη. 54
- 2.13 Σημαδευτές και αντίστοιχα levelings της εικόνας *krumlov*. Αριστερά, πάνω και κάτω: Αρχική εικόνα. Πάνω σειρά: σημαδευτές – γκαουσιανά φιλτραρίσματα της εικόνας, τυπικής απόκλισης $\sigma = 4, 16$ και 64 αντίστοιχα. Κάτω: levelings που προκύπτουν από τον κάθε σημαδευτή. Η υλοποίηση έγχρωμων σημαδευτών και levelings έγινε ξεχωριστά σε κάθε χρωματικό κανάλι. 60
- 2.14 Πυραμίδα leveling Q (αριστερά) και πυραμίδες κοιλάδων $L^{Q,V}$ (κέντρο) και κορυφών $L^{Q,P}$ (δεξιά) με τα πρώτα τέσσερα επίπεδα από πάνω προς τα κάτω, για την εικόνα *harbor* την αρχική μορφή της οποίας βλέπουμε στο επίπεδο 0 της (leveling) πυραμίδας πάνω αριστερά. Οι $L^{Q,V}$ και $L^{Q,P}$ πυραμίδες έχει κανονικοποιηθεί μεταξύ των ελάχιστων και μέγιστων τιμών κάθε επιπέδου ώστε να είναι πιο ευδιάκριτες. Βλέπουμε την σχεδόν άριστη διατήρηση των περιγραμμάτων υψηλές κλίμακες, χωρίς τη δημιουργία ανεπιθύμητων ψευδών μεγίστων. 62
- 3.1 Γενική αρχιτεκτονική του μοντέλου των Itti, Koch και Niebur. Από το [16] 67
- 3.2 Διάγραμμα της αποδόμησης της εικόνας σε Γκαουσιανές και Λαπλασιανές πυραμίδες, και επακόλουθο φιλτράρισμα με πυρήνες Gabor. Από το [10] 70
- 3.3 Αριστερά: Σύνολο προσανατολισμένων φίλτρων για τις τέσσερις κατευθύνσεις - πάνω: πραγματικά μέρη, κάτω: φανταστικά μέρη. Δεξιά: Χαρακτηριστικά φάσματος ισχύος του συνόλου των φίλτρων (μαζί με τα αντίστοιχα συζυγή). Από το [10]. 71
- 3.4 Η λειτουργία του τελεστή $N(\cdot)$. Από το [16] 72
- 3.5 Σχηματική αναπαράσταση της δομής και λειτουργίας του μοντέλου των Itti et al.. Διακρίνουμε την πρόβλεψη για τη δυνατότητα ύπαρξης άλλων ενδιάμεσων χαρτών, όπως συνδέσμων, γωνιών, κίνησης κλπ. – η δυνατότητα προσθήκης ενός χάρτη γωνιών διερευνάται στην ενότητα 3.5.3. Εικόνα από [17]. 75

- 3.6 Εφαρμογή του υπολογιστικού μοντέλου των Itti et al. στην εικόνα cars. Πάνω – αριστερά: αρχική εικόνα, δεξιά: τελικός χάρτης οπτικής προσοχής. Δεύτερη σειρά: ενδιάμεσοι χάρτες χαρακτηριστικών. Κάτω – αριστερά: εξέλιξη των διαδοχικών θέσεων του επικέντρου της προσοχής (ΕτΠ). Δεξιά: αντίστοιχη εξέλιξη του WTA νευρωνικού δικτύου. Χρόνοι πυροδότησης της κάθε μίας από τις τρεις πρώτες θέσεις του ΕτΠ που απεικονίζονται αντιστοιχά: $108.2ms$, $176.8ms$ και $232.5ms$. Το σχήμα ετοιμάστηκε με τη βοήθεια του Saliency Toolbox το οποίο διατίθεται στην ιστοσελίδα του εργαστηρίου Koch (Caltech). . . . 76
- 3.7 Αριστερά: εικόνα με σαφή και ξεκάθαρα σημεία ενδιαφέροντος. Δεξιά: εικόνα όπου είναι πολύ δύσκολο να ορίσουμε αντικειμενικά μια ιεραρχία περιοχών ενδιαφέροντος. Οι εικόνες έχουν παρθεί από βάσεις δεδομένων δημόσιας κυριότητας (public domain databases). 77
- 3.8 Εικόνες (αριστερά) και αντίστοιχοι φιλτραρισμένοι μέσοι όροι αποκρίσεων (δεξιά). Από πάνω προς κάτω: εικόνα cars.jpg – αποτελέσματα από 23 διαφορετικές αποκρίσεις, εικόνα post.jpg – 25 διαφορετικές αποκρίσεις, εικόνα peacock.jpg – 28 διαφορετικές αποκρίσεις. 79
- 3.9 Από αριστερά: εικόνα προς έλεγχο· ενδεικτικό αποτέλεσμα του mouse tracking από ένα χρήστη· μέσος όρος αποτελεσμάτων από 55 χρήστες (από την ιστοσελίδα mouse tracking του TCTS)· μέσος όρος μετά από βαθυπερατό φιλτράρισμα με Γκαουσιανό φίλτρο τυπικής απόκλισης $\sigma = 6$ 80
- 3.10 Εικόνες με διαφορετικές τιμές συσχέτισης μεταξύ τους. Αριθμώντας από αριστερά, εικόνες 1-3: πλαίσια 1,2 και 36 από την ακολουθία *Hamburg taxi* — 4η: γκριζα εικόνα cars στο ίδιο μέγεθος. Τιμές των δισδιάστατων συσχετίσεων c_{ij} : $c_{12} = 0.9879$, $c_{13} = 0.7962$, $c_{14} = -0.1560$, $c_{23} = 0.7972$, $c_{24} = -0.1556$ και $c_{34} = -0.1568$ 81

- 3.11 Επάνω – αριστερά: αρχική εικόνα *rhox*. Δεξιά: τελικός χάρτης οπτικής προσοχής ο οποίος προέκυψε υλοποιώντας τον αλγόριθμο των Itti et al. με την αρχική Γκαουσιανή πυραμίδα. Οι χάρτες υπολογίστηκαν στο μέγεθος του επιπέδου 2 της πυραμίδας (αντί 4) για μεγαλύτερη ευκρίνεια. Μεσαία σειρά: ενδιάμεσοι χάρτες οπτικής προσοχής \bar{I} , \bar{C} και \bar{O} (αντίθεσης, χρωματικού ανταγωνισμού και προσανατολισμών). Κάτω: αρχική εικόνα με σημειωμένες τις περιοχές οι οποίες δίνουν μεγάλες αποκρίσεις στον τελικό χάρτη - με πορτοκαλί περίγραμμα $> 90 \%$ του μεγίστου, με κίτρινο $> 64 \%$ του μεγίστου. Τα όρια των περιοχών αυτών λήφθηκαν παίρνοντας αρχικά ένα δυαδικό σηματοδότη τα σημεία που μας ενδιαφέρουν. Στη συνέχεια, βρήκαμε μια “ μάσκα ” γύρω από αυτά, ως τη διαφορά των dilations του σηματοδότη με μορφοποιητικά στοιχεία παραπλήσιας ακτίνας. Στη μάσκα αυτή θέσαμε η εικόνα να παίρνει τιμές κίτρινου ή πορτοκαλί αντίστοιχα. 83
- 3.12 Επάνω – αριστερά: αρχική εικόνα *rhox*. Δεξιά: τελικός χάρτης οπτικής προσοχής ο οποίος προέκυψε υλοποιώντας τον αλγόριθμο των Itti et al. με την Adjunction πυραμίδα (οι χάρτες υπολογίστηκαν στο μέγεθος του επιπέδου 2 της πυραμίδας). Μεσαία σειρά: ενδιάμεσοι χάρτες οπτικής προσοχής \bar{I} , \bar{C} και \bar{O} (αντίθεσης, χρωματικού ανταγωνισμού και προσανατολισμών). Κάτω: αρχική εικόνα με σημειωμένες τις περιοχές οι οποίες δίνουν μεγάλες αποκρίσεις στον τελικό χάρτη - με πορτοκαλί περίγραμμα $> 90 \%$ του μεγίστου, με κίτρινο $> 64 \%$ του μεγίστου. 84
- 3.13 Επάνω – αριστερά: αρχική εικόνα *rhox*. Δεξιά: τελικός χάρτης οπτικής προσοχής ο οποίος προέκυψε υλοποιώντας τον αλγόριθμο των Itti et al. με την πυραμίδα Ανοίγματος - Κλεισίματος (οι χάρτες υπολογίστηκαν στο μέγεθος του επιπέδου 2 της πυραμίδας). Μεσαία σειρά: ενδιάμεσοι χάρτες οπτικής προσοχής \bar{I} , \bar{C} και \bar{O} (αντίθεσης, χρωματικού ανταγωνισμού και προσανατολισμών). Κάτω: αρχική εικόνα με σημειωμένες τις περιοχές οι οποίες δίνουν μεγάλες αποκρίσεις στον τελικό χάρτη - με πορτοκαλί περίγραμμα $> 90 \%$ του μεγίστου, με κίτρινο $> 64 \%$ του μεγίστου. 85

- 3.14 Επάνω – αριστερά: αρχική εικόνα *rhox*. Δεξιά: τελικός χάρτης οπτικής προσοχής ο οποίος προέκυψε υλοποιώντας τον αλγόριθμο των Itti et al. με την πυραμίδα κυματιδίων Haar (οι χάρτες υπολογίστηκαν στο μέγεθος του επιπέδου 2 της πυραμίδας). Μεσαία σειρά: ενδιάμεσοι χάρτες οπτικής προσοχής \bar{I} , \bar{C} και \bar{O} (αντίθεσης, χρωματικού ανταγωνισμού και προσανατολισμών). Κάτω: αρχική εικόνα με σημειωμένες τις περιοχές οι οποίες δίνουν μεγάλες αποκρίσεις στον τελικό χάρτη - με πορτοκαλί περίγραμμα > 90 % του μεγίστου, με κίτρινο > 64 % του μεγίστου. 86
- 3.15 Επάνω – αριστερά: αρχική εικόνα *rhox*. Δεξιά: τελικός χάρτης οπτικής προσοχής ο οποίος προέκυψε υλοποιώντας τον αλγόριθμο των Itti et al. με την πυραμίδα Leveling (οι χάρτες υπολογίστηκαν στο μέγεθος του επιπέδου 2 της πυραμίδας). Μεσαία σειρά: ενδιάμεσοι χάρτες οπτικής προσοχής \bar{I} , \bar{C} και \bar{O} (αντίθεσης, χρωματικού ανταγωνισμού και προσανατολισμών). Κάτω: αρχική εικόνα με σημειωμένες τις περιοχές οι οποίες δίνουν μεγάλες αποκρίσεις στον τελικό χάρτη - με πορτοκαλί περίγραμμα > 90 % του μεγίστου, με κίτρινο > 64 % του μεγίστου. 88
- 3.16 Αριστερά: δύο εικόνες στις οποίες επιτεύχθηκε μεγάλη ομοιότητα των αποκρίσεων του υπολογιστικού μοντέλου με το πρότυπο – μέτρα ομοιότητας $c_i = 0.4017$ και 0.4014 αντίστοιχα (και τα δύο με leveling πυραμίδα). Δεξιά: εικόνες στις οποίες η καλύτερη προσέγγιση (σε αυτές τις δύο με gaussian πυραμίδα) απέτυχε – $c_i = -0.0763$ και -0.0292 91
- 3.17 Αποτέλεσμα της ανίχνευσης γωνιών με τον ανιχνευτή Harris πάνω στη γκρίζα εικόνα stairs. Αριστερά: με χρήση του αρχικού μέτρου των Harris και Stephens. Δεξιά: με χρήση του προτεινόμενου από την Noble στο [36] μέτρου. 93
- 3.18 Παράδειγμα του τρόπου με τον οποίο εργαστήκαμε για τη σύγκριση των αποκρίσεων του ανιχνευτή με τις αντίστοιχες του προτύπου και του υπολογιστικού αλγορίθμου. Από αριστερά: αρχική εικόνα *turbine*· περιοχές της εικόνας με $C_h > 1\% \max\{C_h\}$ (με μαύρο χρώμα)· θετικά τοπικά μέγιστα των περιοχών αυτών· πρότυπο ανθρώπινης οπτικής προσοχής· αναμενόμενη από τον αλγόριθμο των Itti et al. οπτική προσοχή. 94
- 3.19 Εικόνες με μεγάλους λόγους m_{ref} , για τις αποκρίσεις περιοχών με $C_h > 10\% \max(C_h)$. Από αριστερά: $m_{ref} = 18.6753, 5.0117, 10.6732$ και 8.6271 96

- 3.20 Αποτελέσματα του τελεστή $N(\cdot)$ για εικόνες με διαφορετικό αριθμό και ιεραρχική σχέση γωνιών. Πάνω – αριστερά: εικόνα με μια κύρια γωνία, κέντρο: εικόνα με αριθμό μη-ισοδύναμων γωνιών, δεξιά: εικόνα με σχεδόν ακριβώς ισοδύναμες γωνίες. Κάτω: αντίστοιχες αποκρίσεις του ανιχνευτή Harris με το αρχικό προτεινόμενο μέτρο. Ο συντελεστής πολλαπλασιασμού του τελεστή $N(\cdot)$ για κάθε εικόνα (με $M = 10$ και συνεπώς μέγιστη θεωρητική τιμή $M^2 = 100$) ήταν αντίστοιχα 70.72, 21.53 και 1.40. Ως μέγιστα θεωρήθηκαν οι τιμές που είναι οι μεγαλύτερες από την γειτονιά τους μεγέθους - 8, κατά παράγοντα $1 + 10^{-3}$, ενώ αγνοήθηκαν οι άκρες των χαρτών απόκρισης του ανιχνευτή καθ' ότι έδιναν ψηλές τιμές στα σημεία όπου οι γραμμές έβγαιναν από την εικόνα. Οι επιπλέον παραμέτροι αυτοί του τελεστή $N(\cdot)$ διερευνούνται περαιτέρω στην ενότητα 3.6. 98
- 3.21 Επάνω – αριστερά: αρχική εικόνα *rhox*. Δεξιά: τελικός χάρτης οπτικής προσοχής ο οποίος προέκυψε υλοποιώντας τον αλγόριθμο των Itti et al. με την προσθήκη της απόκρισης του ανιχνευτή Harris ως τέταρτου ενδιάμεσου χάρτη. Μεσαίες σειρές: ενδιάμεσοι χάρτες οπτικής προσοχής \bar{I} , \bar{C} , \bar{O} και \bar{J} (αντίθεσης, χρωματικού ανταγωνισμού, προσανατολισμών και γωνιών). Κάτω: αρχική εικόνα με σημειωμένες τις περιοχές οι οποίες δίνουν μεγάλες αποκρίσεις στον τελικό χάρτη - με πορτοκαλί περίγραμμα > 90 % του μεγίστου, με κίτρινο > 64 % του μεγίστου. Συγκρ. με την εικόνα 3.11. 99
- 3.22 Παράδειγμα του φαινομένου του στραγγαλισμού (σχετικά) ισοδύναμων σημείων. Πάνω – αριστερά: αρχική εικόνα, δεξιά: τελικός χάρτης οπτικής προσοχής. Κάτω βλέπουμε τους τρεις ενδιάμεσους χάρτες, κατά σειρά: διαφορών έντασης I , χρωματικού ανταγωνισμού C και διαφορών προσανατολισμού O . Τα τρία κύρια αντικείμενα του χάρτη γίνονται αντιληπτά στο χάρτη C αλλά η συνεισφορά του στον κύριο χάρτη προκύπτει σχεδόν ισοδύναμη με αυτή των υπολοίπων (συντελεστές πολλαπλασιασμού 0–100: 58.45, 77.30 και 60.68 αντίστοιχα). 102
- 3.23 Μορφή της συνάρτησης $\alpha(r, d)$ όπως περιγράφεται στην εξίσωση 3.2. 103
- 3.24 Γειτονίες μεγέθους - 4 και μεγέθους - 8 για το κεντρικό εικονοστοιχείο ο. 104
- 3.25 Χάρτες οπτικής προσοχής της αρχικής εικόνας *balloon* (αριστερά) χρησιμοποιώντας άθροιση με τον τελεστή N (κέντρο) και επαναληπτική άθροιση (δεξιά). Τα αποτελέσματα λήφθηκαν με τη χρήση του Saliency Toolbox. 104

- 4.1 Τέσσερα παραδείγματα εντοπισμού χωρο-χρονικών σημείων ενδιαφέροντος από απλές συνθετικές ακολουθίες εικόνων για την επέκταση του ανιχνευτή Harris από τον Laptev. Από αριστερά: μια κινούμενη γωνία, η ένωση μιας μπάλας και ενός τοίχου, η σύγκρουση δύο μπάλων με σημεία ενδιαφέροντος εντοπισμένα σε σχετικά λεπτή κλίμακα και, τέλος, το ίδιο με το προηγούμενο αλλά με τα σημεία ενδιαφέροντος εντοπισμένα σε πιο αδρή κλίμακα. Από το [23]. 107
- 4.2 Αριστερά: διαδοχικές τομές ακολουθίας που εμπεριέχει κίνηση. Δεξιά: ισομετρική επιφάνεια η οποία εμπερικλείει τις σημαντικότερες περιοχές της ακολουθίας, βάσει ενός μοντέλου από το [2]. Βλέπουμε τον τρισδιάστατο προσανατολισμό της κινούμενης δομής κατά μια γωνία ως προς το χώρο. Εικόνες από το [2]. . . 108
- 4.3 Πάνω: πλαίσια από μέρος της ταινίας 'As good as it gets'. Από τη δημόσια διαθέσιμη βάση ακολουθιών του [25]. Η ακολουθία που εξετάσαμε περιλάμβανε 382 πλαίσια μεγέθους 133×222 . Δεύτερη σειρά: αντίστοιχες τομές του ενδιάμεσου χάρτη - όγκου αντίθεσης (contrast). Τρίτη σειρά: αντίστοιχες τομές για τον ενδιάμεσο όγκο χρωματικού ανταγωνισμού. Τέταρτη σειρά: τομές για τον ενδιάμεσο όγκο προσανατολισμών, στον οποίο γίνεται αντιληπτή η είσοδος στη σκηνή. Τελευταία σειρά: τομές του τελικού όγκου οπτικής προσοχής. 111
- 4.4 Χρήση της επέκτασης του μοντέλου των Itti et al. στις τρεις διαστάσεις. Πρώτη σειρά: αρχικό πλαίσιο και τομή του όγκου οπτικής προσοχής στη στιγμή αυτή. Επόμενες σειρές, από αριστερά: πλαίσια κάθε χρονικής στιγμής· τομές του όγκου οπτικής προσοχής για την αντίστοιχη χρονική στιγμή· (απόλυτη) διαφορά της τομής από την αρχική· (απόλυτη) διαφορά του τρέχοντος πλαισίου από το αρχικό. Παρουσιάζεται ακολουθία από τη βάση LISBON. 113
- 4.5 Αποτέλεσμα του αλγορίθμου των Lukas - Kanade μεταξύ των πλαισίων 47 και 48 της ακολουθίας jogging. Αριστερά: πλαίσιο 47 της ακολουθίας· Κέντρο: κίνηση προς το επόμενο πλαίσιο· Δεξιά: μέτρο της κίνησης, κανονικοποιημένο στη μέγιστή του τιμή. 120
- 4.6 Τιμές του μέτρου D_{I_1, I_2} για ένα μεγάλο αριθμό (6900) πλαισίων από την ταινία " 300 ". Αριστερά: τιμές για όλο το μήκος της ακολουθίας. Οι πολύ μεγάλες εξάρσεις αντιστοιχούν κατά πολύ μεγάλο ποσοστό σε αλλαγές σκηνής. Δεξιά: τιμές μέσα σε μια σκηνή. Βλέπουμε ότι εδώ οι τιμές κινούνται μέσα σε μια άλλη τάξη μεγέθους. 122

- 4.7 Διαγραμματική αναπαράσταση των διάφορων μεθόδων που χρησιμοποιήσαμε: εύρεση σημαντικότητας D βάσει (α) σταθμισμένων μεταβολών του χάρτη οπτικής προσοχής, (β) πρόβλεψης του επόμενου πλαισίου και (γ) πρόβλεψης του επόμενου χάρτη οπτικής προσοχής. 123
- 4.8 Πάνω: πλαίσια από μέρος της ταινίας 'As good as it gets'. Από τη δημόσια διαθέσιμη βάση ακολουθιών του [25]. Σημειώνουμε τα πλαίσια τα οποία δίνουν μεγάλες τιμές σημαντικότητας βάσει των διάφορων μέτρων που περιγράφονται. Δεύτερη γραμμή: τιμές σημαντικότητας με διαφορές των πλαισίων – μπλε συνεχής γραμμή: απλή διαφορά, κόκκινη στικτή γραμμή, διαφορά με πρόβλεψη του επόμενου πλαισίου. Τρίτη γραμμή: τιμές με απλές διαφορές των πλαισίων βεβαρημένες από την οπτική προσοχή. Τέταρτη γραμμή: τιμές με τις σταθμισμένες με την ταχύτητα διαφορές των χαρτών οπτικής προσοχής. Πέμπτη γραμμή: τιμές με πρόβλεψη του χάρτη οπτικής προσοχής του επόμενου πλαισίου. Κάτω (δύο σειρές): πρώτο πλαίσιο και "περίληψη" βάσει του παραπάνω μέτρου με κατώφλι 0.0475. 124

Κατάλογος Πινάκων

2.1	Κύρια χαρακτηριστικά των πρώιμων και προσεκτικών μηχανισμών προσοχής. Από το [30].	38
3.1	Αρχιτεκτονική του μοντέλου τριών επιπέδων (3L) και αντίστοιχες μεταβλητές κάθε επιπέδου. Από το [30].	65
3.2	Αποτελέσματα της αξιολόγησης για τις υλοποιήσεις του αλγόριθμου των Itti et al. με τις πέντε διαφορετικές πυραμίδες. Φαίνονται αναλυτικά τα αποτελέσματα για τις 26 φυσικές σκηνές που ελέγξαμε μαζί με τις επιπλέον 14 εικόνες που παρείχε η ιστοσελίδα της εφαρμογής validattention.	89
3.3	Σύγκριση των μέσων αποκρίσεων, τόσο του μοντέλου των Itti et al. όσο και του προτύπου, στις συνολικά 40 εικόνων που αναλύσαμε, για περιοχές με αυξανόμενο δείκτη ύπαρξης γωνιών. Συγκεκριμένα, παρατίθενται οι μέσες αποκρίσεις τόσο για περιοχές με μέτρο Harris αυξανόμενης έντασης (0.1, 1, 10, και 50 % του μεγίστου) όσο και για περιοχές – θετικά μέγιστα του μέτρου αυτού, με την αντίστοιχη ένταση.	95
3.4	Αποτελέσματα της αξιολόγησης για τις υλοποιήσεις του αλγόριθμου των Itti et al. με και χωρίς την ενσωμάτωση της απόκρισης Harris C_h ως τέταρτου ενδιάμεσου χάρτη οπτικής προσοχής. . .	98
3.5	Αποτελέσματα της αξιολόγησης για τις υλοποιήσεις του αλγόριθμου των Itti et al. με τις πέντε διαφορετικές πυραμίδες με την ενσωμάτωση της απόκρισης Harris C_h ως τέταρτου ενδιάμεσου χάρτη οπτικής προσοχής. Συγκρίνοντας με τον αντίστοιχο Πίνακα 3.2 βλέπουμε βελτίωση στη μεγάλη πλειοψηφία των περιπτώσεων.	100

Κεφάλαιο 1

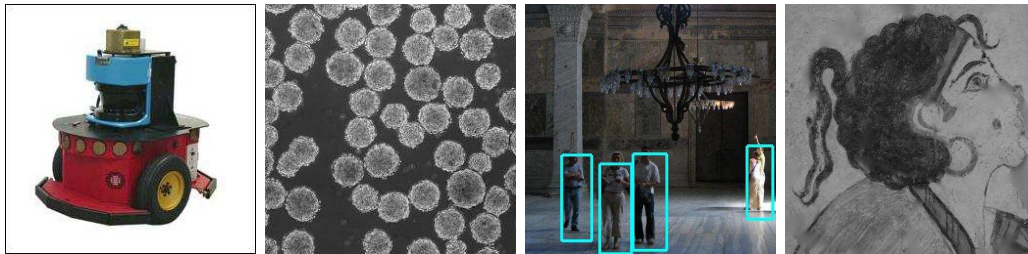
Εισαγωγικά Στοιχεία για την Όραση Υπολογιστών και την Οπτική Προσοχή

1.1 Γενικά για την Όραση Υπολογιστών¹

Η Όραση Υπολογιστών (αγγλικός όρος: Computer Vision) έχει ως βασικό στόχο την εύρεση μιας, κατά το δυνατόν πληρέστερης, συμβολικής περιγραφής των αντικείμενων του πραγματικού κόσμου που ενυπάρχουν μέσα σε μία σκηνή, βάσει μιας ή περισσοτέρων εικόνων που αναφέρονται σε αυτή [1]. Καλείται να απαντήσει σε ερωτήματα σχετικά με την ταυτότητα και τα χαρακτηριστικά των αντικειμένων μέσα σε μια σκηνή, τη θέση στην οποία βρίσκονται και τις σχέσεις μεταξύ τους. Η εξακρίβωση των πιο πάνω συνήθως αποτελεί σχετικά εύκολη δουλειά για ένα προηγμένο σύστημα βιολογικής όρασης όπως π.χ. του ανθρώπου. Η εύρεσή τους όμως με υπολογιστικές μεθόδους, όπως είναι αυτές της Όρασης Υπολογιστών, αποτελεί πρόκληση και, μέχρι σήμερα, ακόμα και η πιο προηγμένες (state-of-the-art) μεθοδολογίες απέχουν πολύ από το να προσεγγίσουν την ανθρώπινη ικανότητα αντίληψης του φυσικού, ορατού κόσμου.

Οι ρίζες της Όρασης Υπολογιστών (Ο.Υ.) ξεκινούν από τη δεκαετία του 1960, οπότε άρχισε να αναδεικνύεται ως επιστημονικό πεδίο μέσα από τη συμβολή τριών κύριων κλάδων – της Τεχνητής Νοημοσύνης (Artificial Intelligence), της Επεξεργασίας Σημάτων (Signal Processing) και της Αναγνώρισης Προτύπων (Pattern Recognition). Από τη δεκαετία του 1980 και έπειτα έχει ανδρωθεί σε μια γρήγορα εξελισσόμενη και εύρωστη ερευνητικά περιοχή η οποία βρίσκει εφαρμογές σε πολλούς τομείς, όπως ευρύτερα η ρομποτική, η επεξερ-

¹Η παράγραφος αυτή βασίζεται κυρίως στο [1].



Σχήμα 1.1: Παραδείγματα εφαρμογών της Όρασης Υπολογιστών. Από αριστερά: οπτικά καθοδηγούμενο ρομπότ· βιοιατρική εικόνα· εύρεση αντικειμένων (εδώ ανθρώπινων μορφών, από την εργασία στο [24])· αποκατάσταση χαμένων μερών από αρχαίες τοιχογραφίες, από το [37].

γασία εικόνων και βίντεο, βιοϊατρική, συστήματα παρακολούθησης, συστήματα κατηγοριοποίησης οπτικού υλικού κλπ., ή ακόμα και τις ψηφιακές εικαστικές τέχνες. Στο Σχήμα 1.1 βλέπουμε μερικά παραδείγματα των εφαρμογών αυτών. Πρέπει, τέλος, να σημειώσουμε μια απαραίτητη διάκριση της Ο.Υ. από συγγενείς επιστήμες, τις λεγόμενες επιστήμες εικόνων (*image sciences*): η Ο.Υ. προσπαθεί να εξάγει το συμβολικό περιεχόμενο μιας εικόνας, ενώ

- η επεξεργασία εικόνων (*image processing*) ασχολείται, όπως σαφώς δηλώνει ο όρος, με την εξαγωγή μιας εικόνας από μετασχηματισμό μιας άλλης,
- η γραφική υπολογιστών (*computer graphics*) με την απόδοση ενός συμβολικού περιεχομένου πάνω σε μια εικόνα (δηλαδή, κατά κάποιο τρόπο, την αντίστροφη διεργασία από την Ο.Υ.),
- η απεικόνιση (*imaging*) έχει ως αντικείμενο τη σύλληψη των φυσικών σκηνών πάνω σε εικόνες.

1.2 Περί Προσοχής¹

Ο άνθρωπος, όπως και όλοι οι ζωντανοί οργανισμοί, βομβαρδίζεται συνεχώς, μέσω των αισθητηρίων οργάνων του, από τεράστιες ποσότητες πληροφορίας. Στον εγκέφαλο φτάνουν, αδιάλειπτα, σήματα που μεταφέρουν στοιχεία για τα χαρακτηριστικά του γύρω κόσμου – οπτικά, ακουστικά, χαρακτηριστικά της οσμής, της αφής και υφής των αντικειμένων, της γεύσης. Ακόμα και για το πιο περίπλοκο γνωστικό όργανο που γνωρίζουμε, όπως είναι ο ανθρώπινος εγκέφαλος, η ροή αυτή της πληροφορίας είναι υπερβολικά μεγάλη ώστε να αναλυθεί

¹Κάποιες πληροφορίες για την παράγραφο αυτή, όπως και αυτή που ακολουθεί, πήραμε από τα [2] και [30]

στην αρχική, ακατέργαστη της μορφή. Έτσι, τα μέλη του ζωικού βασιλείου έχουν αναπτύξει δικούς τους μηχανισμούς φιλτραρίσματος της πληροφορίας αυτής, ώστε να εστιάζουν μόνο στα κομμάτια της τα οποία περιέχουν ενδιαφέρον και τους είναι χρήσιμα. Το σύνολο των μηχανισμών αυτών αποκαλούμε, γενικά και μάλλον ασαφώς, προσοχή. Λέμε ασαφώς, γιατί δεν υπάρχει ένας μοναδικός ορισμός της προσοχής λόγω της ποικιλίας των επιστημονικών κλάδων που την αναλύουν. Αρχικά αντικείμενο κάποιων φιλοσόφων – για παράδειγμα, κάποιες σκέψεις πάνω σε έννοιες σχετικές με την προσοχή βρίσκονται στο έργο του Καρτέσιου – δε μελετήθηκε συστηματικά και σε βάθος παρά με τις απαρχές της επιστήμης της ψυχολογίας. Η μεγάλη της σημασία τράβηξε το ενδιαφέρον – ή την προσοχή, αν θέλετε – επιστημόνων από κλάδους όπως η ανατομία, φυσιολογία, νευρολογία και, πιο πρόσφατα, την υπολογιστική επιστήμη. Η τελευταία αποτελεί και τη σκοπιά από την οποία η παρούσα εργασία θα προσεγγίσει το θέμα.

Μέσα στην πλειάδα των επιστημών που αγκαλιάζουν ως αντικείμενό τους την προσοχή, έχει προταθεί από τους Tsotsos et al. [18] ότι *ανεξαρτήτως του κλάδου, της μεθοδολογίας ή σκοπιμότητας ο κοινός παρονομαστής που καθορίζει την προσοχή είναι η μείωση του όγκου της πληροφορίας*. Αν και κάποιες φορές η προσοχή είναι προϊόν συνειδητής προσπάθειας, συνήθως εδράζεται πάνω σε αστραπιαίες, μη συνειδητές διαδικασίες. Βάσει αυτών γίνεται μια γρήγορη επιλογή των καίριων σημείων και εστιών ενδιαφέροντος του γύρω κόσμου και αποτελούν αντανακλαστικό απαραίτητο για την επιβίωση. Στις έννοιες αυτές θα ασχοληθούμε περισσότερο στο κεφάλαιο 2, όσον αφορά μια από τις πιο περίπλοκες μορφές προσοχής, την οπτική προσοχή – αντίστοιχος αγγλικός όρος visual attention – πάνω στην οποία έχουν ήδη αναπτυχθεί διάφορα υπολογιστικά μοντέλα.

1.3 Οπτική Προσοχή

Μέσω της οπτικής προσοχής (visual saliency) μπορούμε λοιπόν να εξάγουμε το σημαντικότερο κομμάτι της περιγραφής του περιβάλλοντος όπως το αντιλαμβανόμαστε μέσω των οφθαλμών και γενικότερα του οπτικού συστήματος. Η γρήγορη αλίευση της χρήσιμης πληροφορίας παρέχει τα εχέγγυα για καλύτερη αντίληψη του γύρω κόσμου, παραδείγματος χάρη ανίχνευση πιθανών κινδύνων, αποφυγή παγίδων και εμποδίων, ασφαλή πλοήγηση. Αρκεί μόνο να σκεφτούμε ότι τα σήματα τροχαίας έχουν σχεδιαστεί ώστε να παρέχουν ερεθίσματα που τραβάνε άμεσα την προσοχή – βαμμένα με έντονα χρώματα, γυαλιστερά, κάποια φωσφορίζοντα για τη νύκτα κλπ. με σκοπό να παρέχουν αντίθεση με το οπτικό τους περιβάλλον. Τα χρώματα πράσινο και κόκκινο, τα οποία χρησιμοποιούνται στα φανάρια των διασταυρώσεων, αποτελούν ζεύγος χρωματικού ανταγωνισ-

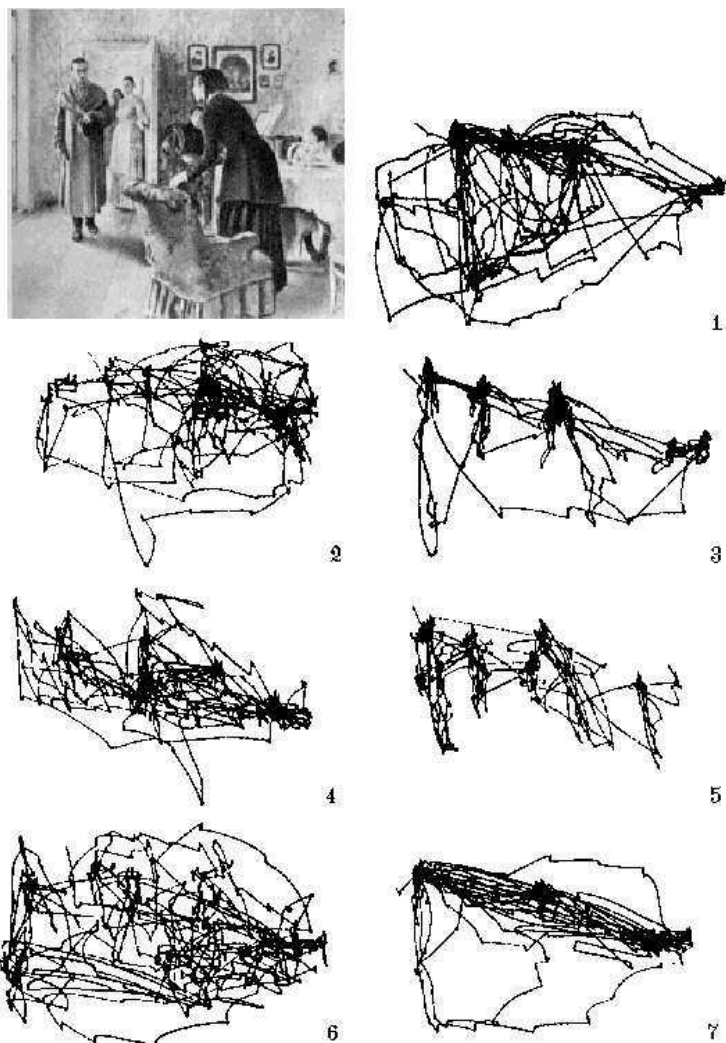
μού, ένα ιδιαίτερα αξιοπρόσεκτο χαρακτηριστικό όπως θα δούμε σε επόμενα κεφάλαια.

Πέρα από την άμεση, αντανακλαστική αντίληψη της οπτικής προσοχής (λειτουργία την οποία χαρακτηρίζουμε ως πρώιμη και περιγράφουμε αναλυτικότερα στην ενότητα 2.1.3), μπορούμε να δούμε ότι αυτή επηρεάζεται επίσης από τα διαφορετικά βιώματα και τις συνειδητή επεξεργασία της εικόνας που προκύπτει βάσει αυτών. Σε μια κλασική εικόνα από την έρευνα του Yarbus από το 1967 [47], την οποία βλέπουμε στο Σχήμα 1.2, είναι εμφανής η αλλαγή της φύσης των εστιάσεων των παρατηρητών όταν αυτοί δρουν βάσει διαφορετικών οδηγιών. Η συμπεριφορά αυτή, την οποία αποκαλούμε καθοδική (top-down) ή προσεκτική οπτική προσοχή, είναι δύσκολη στη μοντελοποίηση, πολύ περισσότερο από την άμεση, ανοδική (bottom-up) οπτική προσοχή, για την οποία έχουν αναπτυχθεί πολλά μοντέλα.

Ένα από τα πιο γνωστά τέτοια μοντέλα αποτελεί αυτό των Itti et al. το οποίο παρουσιάστηκε το 1998 και παραμένει η βάση πολλών από τις έρευνες που γίνονται στον τομέα. Υπάρχουν διάφοροι τρόποι αξιολόγησης των μοντέλων αυτών – το ένα είναι το πόσο προσομοιάζουν τα αποτελέσματά του με αυτά του πραγματικού, βιολογικού μοντέλου, τα οποία εξάγουμε με ψυχοφυσικά πειράματα. Μια τέτοια προσέγγιση χρησιμοποιήσαμε και εμείς στη διάρκεια της συγκρότησης της διπλωματικής εργασίας. Ένα δεύτερο σημείο αξιολόγησης είναι η ικανότητα του υπολογιστικού μοντέλου να επιλύει προβλήματα μηχανικής όρασης [45]. Τα τελευταία αποτελούν και μια από τις μεγαλύτερες εφαρμογές αυτών των μοντέλων, όπως θα περιγράψουμε στην αμέσως επόμενη ενότητα.

1.4 Εφαρμογές της Μοντελοποίησης της Οπτικής Προσοχής

Καθώς λοιπόν η βιολογική οπτική προσοχή κατέχει εξέχουσα σημασία στη λειτουργία των εξελιγμένων οργανισμών, η μοντελοποίησή της μπορεί να παρέχει σημαντικές υπηρεσίες και εφαρμογές σε διάφορους επιστημονικούς τομείς, με πρώτους αυτούς που ασχολούνται με τη μοντελοποίηση πιο γενικών ευφυών συστημάτων. Η χρήση της οπτικής προσοχής στην **ρομποτική** είναι μια από αυτές, καθ' ότι ένας μεγάλος αριθμός ρομποτικών συστημάτων βασίζεται σε αισθητήρες όρασης, αποκλειστικά ή σε συνδυασμό με άλλα συστήματα, αφής, ακοής κλπ. . Εξάγοντας τις πιο ενδιαφέρουσες περιοχές του οπτικού του πεδίου, το ρομποτικό σύστημα μπορεί να εστιάσει σε αυτές ώστε να τις αναλύσει με μεγαλύτερη λεπτομέρεια, γλιτώνοντας από το να κάνει την ίδια, πιθανόν απαιτητική, διεργασία σε όλο το εύρος της εικόνας που έχει μπροστά του. Με τη βοήθεια μηχανισμών προσοχής, ένα κινούμενο ρομπότ μπορεί να αντιληφθεί



Σχήμα 1.2: Αρχική εικόνα (πάνω αριστερά) και καταγεγραμμένες κινήσεις του βλέμματος υπό διαφορετικές οδηγίες – 1: ελεύθερη εξερεύνηση της εικόνας, 2: εξακριβώστε την κοινωνική θέση των ατόμων, 3: εξακριβώστε την ηλικία των ατόμων, 4: σκεφτείτε τι συνέβαινε πριν την άφιξη του επισκέπτη, 5: απομνημονεύστε τα ρούχα των ατόμων στη σκηνή, 6: απομνημονεύστε τη θέση των ατόμων και αντικειμένων, 7: υπολογίστε πόσο χρόνο ο επισκέπτης ήταν μακριά από την οικογένεια. Από το [47].

καίρια σημεία (landmarks) του χώρου δράσης του, να ενισχύσει παράλληλους μηχανισμούς αποφυγής εμποδίων ή, αν έχει σκοπό να εξερευνήσει ελεύθερα το χώρο, να καθοδηγηθεί στη διεύθυνση στην οποία εντοπίζει τα πιο αξιοπρόσεκτα (salient) αντικείμενα. Μια ιδιαίτερη κατηγορία καίριων σημείων είναι τα σήματα τροχαίας, τα οποία είναι συνήθως, λόγω και του σχεδιασμού τους, ελκυστές των πιο πρώιμων στοιχείων της οπτικής προσοχής. Με την αυτόματη οδήγηση οχημάτων να αποτελεί ένα από τα μεγάλα σχέδια της τρέχουσας έρευνας πάνω στη ρομποτική, η χρήση μηχανισμών οπτικής προσοχής μπορεί να αποδειχθεί μια χρήσιμη προσθήκη.

Η χρήση της οπτικής προσοχής με σκοπό την **αναγνώριση ανθρώπινων πράξεων** μέσα σε ακολουθίες εικόνων (video). Από αυτές εξάγονται χωροχρονικά σημεία ενδιαφέροντος τα χαρακτηριστικά των οποίων χρησιμεύουν στην κατηγοριοποίηση των πράξεων. Ξεκινώντας από βάσεις ακολουθιών με διάφορες πρότυπες ανθρώπινες δραστηριότητες (περπάτημα, τρέξιμο, αλλά και πιο περίπλοκες όμως χειραψίες, εναγκαλισμούς κλπ.) και διαχωρίζοντάς τις σε σύνολα εκμάθησης και ελέγχου, διάφοροι ερευνητές όπως οι Dollar et al. [6], Laptev et al. [23], [25] και Ραπαντζίκος [2] πρότειναν και αξιολόγησαν διάφορες μεθόδους αναγνώρισης των δραστηριοτήτων. Η τελευταία μάλιστα αποτελεί μια από τις πιο πρόσφατες και έχει επιτύχει τα μικρότερα ποσοστά σύγχυσης μεταξύ των πράξεων πάνω σε πρότυπη βάση ακολουθιών.

Ένα άλλο πρόσφορο πεδίο εφαρμογών αποτελεί η σύντμηση του τεράστιου όγκου οπτικής πληροφορίας που δημιουργείται και αποθηκεύεται καθημερινά υπό τη μορφή video. Αυτός ο όγκος προβάλλει επιτακτική την ανάγκη ύπαρξης μεθόδων οι οποίες μπορούν να εξάγουν από τις ακολουθίες αυτές τα κυριότερα σημεία, χωρίς να χάνεται το ουσιαστικό περιεχόμενο. Η δημιουργία τέτοιων σύντομων και περιεκτικών περιλήψεων (video (skims) εμπεριέχει την οπτική ανάλυση ως ένα κύριο συνιστόν στοιχείο, με τα υπόλοιπα να είναι η ηχητική και η λεκτική / σημασιολογική ανάλυση. Βασίζεται σε διάφορες μεθοδολογίες, όπως η εξαγωγή χωροχρονικών σημείων ενδιαφέροντος, η θεωρία πληροφορίας με την εύρεση των πιο "σπάνιων" περιοχών της ακολουθίας κ.ά. Μάλιστα, μια εφαρμογή με σκοπό τη **σύντμηση ακολουθιών εικόνων** έχει αναπτυχθεί πρόσφατα από τα εργαστήρια Όρασης Υπολογιστών, και Εικόνων, Βίντεο και Πολυμεσικών Συστημάτων του Εθνικού Μετσόβιου Πολυτεχνείου [8].

Μια παραπλήσια εφαρμογή, η οποία συνδυάζει στοιχεία από τις δύο προηγούμενες, όσον αφορά όμως δισδιάστατες εικόνες, είναι η **επιλογή εικόνων βάσει του ενδιαφέροντος** που τις χαρακτηρίζει. Αυτή είναι διαφορετική εφαρμογή από τη δημιουργία περιλήψεων ακολουθιών καθ' ότι οι εικόνες εδώ δεν αποτελούν πλαίσια μιας κοινής ακολουθίας αλλά (δυσνητικά) ανεξάρτητες λήψεις διαφορετικών σκηνών η κάθε μια. Ένα ενδιαφέρον παράδειγμα της εφαρμογής αυτής αποτελεί μια έρευνα της NASA η οποία είχε σκοπό να ιεραρχήσει τις εικόνες που λάμβανε το όχημα Mars Rover το οποίο εξερευνούσε την επιφάνεια

του πλανήτη Άρη. Καθώς το ότι το εύρος ζώνης για την αποστολή των εικόνων αυτών πίσω στη γη ήταν περιορισμένο, η αυτόματη ιεράρχηση (image ranking) των εικόνων θα μπορούσε να ξεχωρίσει αυτές με το μεγαλύτερο επιστημονικό ενδιαφέρον [46]. Η ανάλυση αυτή μπορεί να βασίζεται σε ένα σύνολο ευριστικών κανόνων που αξιολογούν τέτοιες εικόνες. Μια εναλλακτική υλοποίηση θα μπορούσε να είναι η εξαγωγή κάποιων πρωτογενών χαρακτηριστικών των εικόνων ή/και γενικά σημείων ενδιαφέροντος (salient points) και η ακόλουθη εκμάθηση ενός νευρωνικού συστήματος ώστε να αντικατοπτρίζει τις διαφορές αυτές τιμές σε ένα μέτρο σημαντικότητας της κάθε εικόνας, στη βάση πρότυπων απαντήσεων από ειδικούς του είδους.

Η **συμπύεση των εικόνων**, με διατήρηση των πιο αξιοπρόσεκτων περιοχών σε υψηλή ανάλυση και σχετική υποβάθμιση αυτών με χαμηλές τιμές προσοχής αποτελεί μια υποσχόμενη μέθοδο μείωσης του μεγέθους τους. Μια άλλη εφαρμογή αφορά τον κόσμο της διαφήμισης (και γενικά της σκηνοθεσίας): προβλέποντας τη θέση των αξιοπρόσεκτων περιοχών μέσα σε μια διαφημιστική καταχώρηση, εικόνα ή βίντεο, είναι δυνατή η αξιολόγηση της αποτελεσματικότητάς της να κατευθύνει τον παρατηρητή προς το προϊόν. Ένας σκηνοθέτης ο οποίος θέλει σε μια *σκηνή* να κατευθύνει τους θεατές προς ένα συγκεκριμένο σημείο, *θέτει*, αποδίδει σε αυτό πρώιμα αξιοπρόσεκτα χαρακτηριστικά.

Άλλες πιθανές χρήσεις θα ήταν η υποβοήθηση μηχανισμών αναγνώρισης αντικειμένων, κατάτμησης εικόνων κ.ά. — υπάρχει αναμφίβολα μια πληθώρα εφαρμογών της οπτικής προσοχής, οι οποίες είτε δοκιμάζονται είτε μας παραμένουν ακόμη άγνωστες. Καθώς βελτιώνουμε την κατανόηση των δικών μας, ανθρώπινων, αλλά και γενικότερα βιολογικών, μηχανισμών οπτικής προσοχής, και ανακαλύπτουμε τη χρήση και σημασία τους, αναμενόμενο είναι να εμπνεόμαστε για ανάλογες νέες εφαρμογές.

1.5 Διάρθρωση της Διπλωματικής Εργασίας

Στο **κεφάλαιο 2** παρουσιάζουμε αρχικά το βιολογικό υπόβαθρο πίσω από τα υπολογιστικά μοντέλα οπτικής προσοχής. Περιγράφουμε συνοπτικά το ανθρώπινο σύστημα όρασης και εμβαθύνουμε στα σημεία που αφορούν την οπτική προσοχή στις κύριες της μορφές, παραθέτοντας διάφορα παραδείγματα. Στη συνέχεια περιγράφουμε το μαθηματικό υπόβαθρο των μοντέλων και των επεκτάσεών τους) όπως θα εξετάσουμε στα επόμενα κεφάλαια. Συγκεκριμένα αναφερόμαστε σε έκταση πάνω στην ανάλυση εικόνων σε πολλαπλές κλίμακες (multiscale image analysis) ενώ παρουσιάζουμε και ορίζουμε πέντε διαφορετικούς τύπους πυραμιδικής αναπαράστασης εικόνων.

Στο **κεφάλαιο 3** αναφερόμαστε σε υπολογιστικά μοντέλα ανάλυσης της

οπτικής προσοχής δισδιάστατων εικόνων. Δίνουμε αρχικά μια σύντομη, γενική άποψη επί του θέματος και στη συνέχεια εστιάζουμε πάνω στο μοντέλο των Itti et al., το οποίο περιγράφουμε διεξοδικά. Το υπόλοιπο του κεφαλαίου ασχολείται με την παρουσίαση και αξιολόγηση κάποιων προτεινόμενων αλλαγών και επεκτάσεων πάνω στο μοντέλο αυτό. Ξεκινώντας με την περιγραφή του τρόπου σύγκρισης και αξιολόγησης μεταξύ των διάφορων αποκρίσεων, ο οποίος βασίστηκε στον ανθρώπινο παράγοντα, συνεχίζουμε ελέγχοντας την υλοποίηση του αλγόριθμου των Itti et al. με τις πέντε διαφορετικές πυραμίδες πολυκλιμακωτής αναπαράστασης του κεφαλαίου 2. Έπειτα προτείνουμε την προσθήκη ενός ανιχνευτή γωνιών στο μοντέλο, δείχνοντας αρχικά την αναγκαιότητα μιας παρόμοιας προσθήκης βάσει των ανθρώπινων αποκρίσεων προσοχής πάνω σε ένα σύνολο εικόνων. Τέλος, αναφέρουμε κάποιες σκέψεις πάνω στην κανονικοποίηση των αποκρίσεων που αφορούν τα επιμέρους συστατικά της οπτικής προσοχής.

Στο **κεφάλαιο 4** εξετάζουμε την οπτική προσοχή όσον αφορά και το χρόνο. Στο πρώτο μέρος θεωρούμε το χρόνο ως τρίτη διάσταση και προτείνουμε την επέκταση του μοντέλου των Itti et al. στις τρεις διαστάσεις, αναφέροντας παράλληλα και παρόμοιες προσπάθειες. Δίνουμε επίσης διάφορα παραδείγματα με τη χρήση μικρών ακολουθιών εικόνων (video). Στο δεύτερο μέρος θέτουμε τις βάσεις για ένα μοντέλο που βασίζεται στην έκπληξη, στο οποίο κάθε πλαίσιο μέσα σε μια ακολουθία γίνεται περισσότερο ή λιγότερο αξιοπρόσεκτο βάσει της αναμενόμενης από τα προηγούμενα μορφής του.

Το **κεφάλαιο 5** αποτελεί μια κατακλείδα της διπλωματικής εργασίας, αναφέροντας συνοπτικά τα συμπεράσματα, στοιχεία κάποιες άλλες ιδέες που δοκιμάστηκαν και τις υλοποιήσεις των αλγορίθμων που πραγματοποιήθηκαν στα πλαίσια της διπλωματικής εργασίας. Παρατίθενται επίσης κάποιες δυνατές μελλοντικές κατευθύνσεις.

Κεφάλαιο 2

Βιολογικό και μαθηματικό υπόβαθρο

Το κεφάλαιο αυτό αποτελείται από δύο (σχεδόν) διακριτά μέρη. Στο πρώτο παρουσιάζουμε το βιολογικό υπόβαθρο πίσω από τα μοντέλα προσέγγισης των μηχανισμών οπτικής προσοχής. Πολλά μοντέλα που υπάρχουν στη φύση, ιδιαίτερα τα συστήματα όρασης του ανθρώπου και άλλων πρωτεύοντων θηλαστικών, έχουν μελετηθεί εκτενώς στην ιατρική επιστήμη. Ο πλούτος αυτός της έρευνας που έχει συντελεστεί για το θέμα, ιδιαίτερα στη νευροφυσιολογία και την ψυχολογία, αποτέλεσε έμπνευση για πολλά υπολογιστικά μοντέλα τα οποία αναπτύχθηκαν με συνεργασία επιστημόνων και μηχανικών της Όρασης Υπολογιστών (Ο.Υ.) με επιστήμονες των κλάδων αυτών. Γι αυτό κρίνουμε σκόπιμο να παραθέσουμε μια γενική επισκόπηση των βιολογικών μηχανισμών οι οποίοι περιγράφουν την οπτική προσοχή και να ενδιατρίψουμε στα στοιχεία τα οποία αφορούν τα μοντέλα που θα παρουσιάσουμε.

Το δεύτερο μέρος αφορά το μαθηματικό υπόβαθρο και τις τεχνικές της Ο.Υ. πίσω από την ανάλυση και τα μοντέλα που θα εξετάσουμε. Αναφέρουμε την καταλληλότητα των φίλτρων Gabor ως προσεγγίσεις οπτικών αποκρίσεων βιολογικών συστημάτων. Επίσης περιγράφουμε διάφορους τύπους πυραμίδων, αναπαραστάσεων δηλαδή εικόνων σε πολλαπλά επίπεδα, τους οποίους θα χρησιμοποιήσουμε στο κεφάλαιο 3 στη μελέτη μας για βελτίωση υπάρχοντων υπολογιστικών μοντέλων οπτικής προσοχής.

2.1 Βιολογικά μοντέλα οπτικής προσοχής

Τα βιολογικά συστήματα όρασης έμπνευση και πρότυπο για πολλές από τις μεθοδολογίες που αναπτύχθηκαν πάνω στην Όραση Υπολογιστών. Η μελέτη των μηχανισμών όρασης, ιδιαίτερα στον άνθρωπο, αλλά και σε άλλα πρωτεύοντα

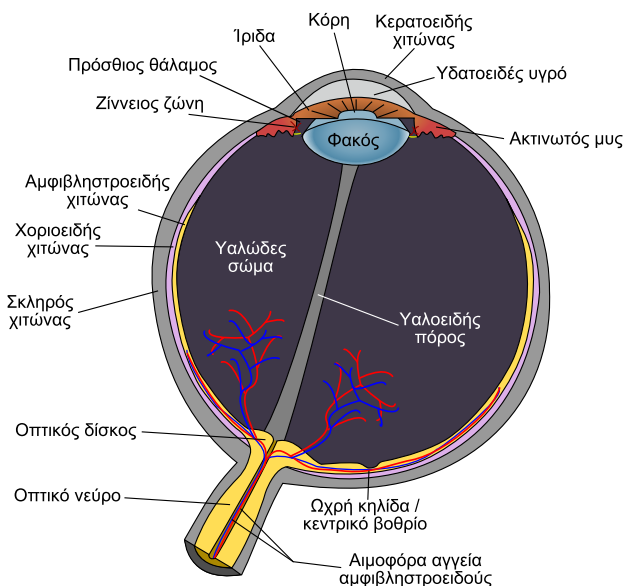
θηλαστικά (π.χ. πίθηκος), είχε ως αποτέλεσμα πολλοί αλγόριθμοι επεξεργασίας οπτικής πληροφορίας στην Ο.Υ. να αναπαράγουν την δομή των αντίστοιχων βιολογικών συστημάτων. Επίσης, πολύ συχνά οι αλγόριθμοι αυτοί χρησιμοποιούν τις αποκρίσεις των τελευταίων ως σύνολα επαλήθευσης, τη λεγόμενη *ground truth*, τεχνική την οποία θα εφαρμόσουμε και εμείς στα πλαίσια της διπλωματικής εργασίας.

Πολλά από τα ακαδημαϊκά πονήματα που αφορούν την Ο.Υ. ξεκινούν περιγράφοντας τη δομή του πιο γνωστού και διεξοδικά μελετημένου βιολογικού οπτικού συστήματος: αυτού του ανθρώπου. Το παρόν δε θα αποτελέσει εξαίρεση, λόγω του ότι πολλές σχετικές έννοιες και στοιχεία ορολογίας αναφέρονται σε ύστερο στάδιο. Η περιγραφή όμως θα παραμείνει λιτή και συνοπτική, καθ' όσον δεν αποτελεί σκοπό της διπλωματικής εργασίας. Στη συνέχεια, θα εμβαθύνουμε λίγο περισσότερο στους βιολογικούς μηχανισμούς αντίληψης χαρακτηριστικών και καθοδήγησης της οπτικής προσοχής στο πρώιμο (*preattentive*) και προσεκτικό (*attentive*) επίπεδο.

2.1.1 Το Ανθρώπινο Σύστημα Όρασης

Η όραση του ανθρώπου αποτελεί ένα περίπλοκο σύστημα το οποίο συγκροτείται από διάφορους μηχανισμούς πρόσληψης και επεξεργασίας οπτικής πληροφορίας οι οποίοι κατανέμονται στον οφθαλμό και σε ορισμένες περιοχές του εγκεφάλου. Ως κύρια μέρη του συστήματος αυτού θεωρούμε τον οφθαλμό και τον αμφιβληστροειδή χιτώνα (*retina*), τον έξω γονατώδη πυρήνα (*lateral geniculate nucleus*) και τον οπτικό φλοιό (*visual cortex*).

Ο **οφθαλμός** εμπερικλείει τόσο λειτουργίες πρόσληψης πληροφορίας όσο και επεξεργασίας της. Μια απλοποιημένη τομή του φαίνεται στο Σχήμα 2.1. Ειδικότερα ο αμφιβληστροειδής χιτώνας λειτουργεί ως υπολογιστής ο οποίος εξάγει την χρήσιμη πληροφορία της εικόνας μέσω χωροχρονικού και χρωματικού φιλτραρίσματος. Ο τελευταίος καλύπτει το εσωτερικό στρώμα του οφθαλμού και περιέχει τους φωτοαισθητήρες μαζί με στρώματα νευρικών κυττάρων. Το προσπίπτον φως εστιάζεται πάνω στον αμφιβληστροειδή, ο οποίος περιέχει δύο ειδών φωτοαισθητήρες, τα *ραβδία* (*rods*) και τα *κωνία* (*cones*), οι οποίοι κατανέμονται ανομοιόμορφα στην επιφάνειά του [1]. Τα ραβδία, περίπου μια τάξη μεγέθους περισσότερα στον αριθμό από τα κωνία, είναι ευαίσθητα σε πολύ χαμηλά επίπεδα φωτισμού και ευθύνονται για τη λεγόμενη σκοτοπική όραση. Τα κωνία είναι ενεργά σε κανονικά επίπεδα φωτισμού – οπότε αντίστοιχα μιλάμε για φωτοπική όραση – και ευθύνονται για την ικανότητα όρασης μεγάλης ακρίβειας και την αντίληψη του χρώματος. Υπάρχουν τρία είδη κωνίων τα οποία είναι ευαίσθητα σε τρεις επικαλυπτόμενες φασματικές ζώνες τις οποίες,

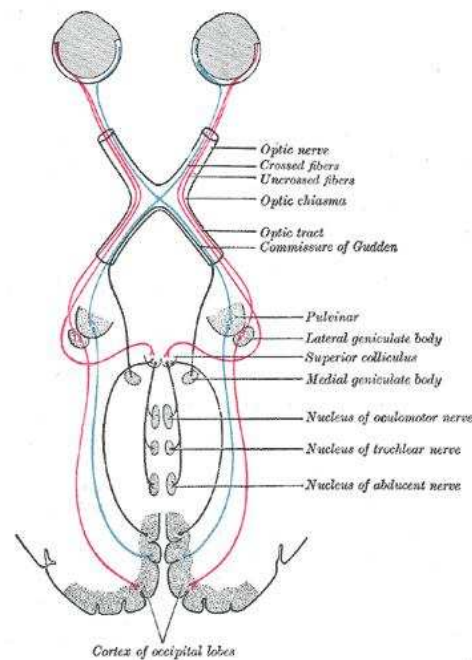


Σχήμα 2.1: Τομή του ανθρώπινου οφθαλμού

μάλλον καταχρηστικά¹, αντιστοιχούμε στα χρώματα μπλε, πράσινο και κόκκινο (με αυξάνουσα σειρά μήκους κύματος στο οποίο η κάθε ζώνη έχει τη μέγιστη απορροφητική ευαισθησία).

Οι νευρώνες του αμφιβληστροειδούς έχει βρεθεί ότι στη λειτουργία τους προσομοιάζουν συνελίξεις με αποκρίσεις νευρώνων με ερεθιστικές (excitatory) και ανασταλτικές (inhibitory) περιοχές στα δεκτικά τους πεδία, τους λεγόμενους μηχανισμούς κέντρου - περιφέρειας (center - surround). Αυτό είναι μια σημαντική αρχή η οποία κατέχει εξέχουσα θέση στα μοντέλα οπτικής προσοχής τα οποία θα εξετάσουμε στο κεφ. 3. Η ίδια αρχή περιγράφει, γενικά, αντίστοιχες λειτουργίες τόσο και στα άλλα δύο κύρια μέρη του συστήματος της ανθρώπινης όρασης τα οποία αναφέραμε. Ο **έξω γονατώδης πυρήνας** βρίσκεται μέσα στο θάλαμο του εγκεφάλου και είναι ο κύριος επεξεργαστής των ερεθισμάτων που προέρχονται από την αμφιβληστροειδή. Αποτελείται από διαδοχικά στρώματα νευρώνων και νευροπίλης (neuropil), σε μια δομή παρόμοια με αυτή ενός γλυκίσματος. Αν και μεγάλο μέρος των λειτουργιών του δεν έχει πλήρως προσδιοριστεί, έχει δείχθει ότι σε αυτόν πραγματοποιείται χρονική αποσυσχέτιση της οπτικής πληροφορίας με μηχανισμούς κέντρου - περιφέρειας,

¹Αυτό είναι ανακριβές καθ' ότι αφ' ενός μιλάμε για ζώνες και όχι για μονοχρωματικές αποκρίσεις, τις οποίες μπορούμε να φανταστούμε σαν μία κρουστική συνάρτηση στο πεδίο των συχνοτήτων και, αφ' ετέρου, οι ίδιες οι συχνότητες μέγιστης ευαισθησίας της κάθε ζώνης στην πραγματικότητα αντιστοιχούν στα χρώματα βιολετί, πράσινο - μπλε και πράσινο - κίτρινο.



Σχήμα 2.2: Διάγραμμα του ανθρώπινου οπτικού συστήματος. Φαίνονται οι οφθαλμοί (πάνω), ο έξω γονατώδης πυρήνας (περίπου στο κέντρο), ο οπτικός φλοιός (στο κάτω μέρος) και το νωτιαίο και πρόσθιο ρεύμα. Από το Gray's Anatomy, έκδοση 1918.

σε αντιστοιχία με τη χωρική αποσυσχέτιση την οποία είδαμε ότι πραγματοποιούν παρόμοιοι μηχανισμοί στον αμφιβληστροειδή.

Ο έξω γονατώδης πυρήνας προβάλλει την επεξεργασμένη πληροφορία πάνω στον πρωταρχικό οπτικό φλοιό, από τον οποίο δέχεται ισχυρή ανάδραση μέσω νευρωνικών συνάψεων. Ο πρωταρχικός ή ραβδωτός **οπτικός φλοιός** (primary / striate visual cortex) διατηρεί μια απεικόνιση της χωρικής οπτικής πληροφορίας, ενώ μέσω δύο κύριων οδών, του νωτιαίου (dorsal) και του πρόσθιου (ventral) ρεύματος, στέλνει πληροφορία προς άλλα μέρη του εγκεφάλου. Το νωτιαίο ρεύμα μεταφέρει πληροφορία σχετική με κίνηση και τις θέσεις των αντικειμένων, ενώ το κοιλιακό σχετική με αναγνώριση μορφής και απεικόνιση των αντικειμένων. Ο πρώτος δρόμος καταλήγει ειδικότερα στον ύστερο πλευρικό φλοιό (posterior parietal cortex), στον οποίο πιστεύεται ότι δημιουργείται ένας χάρτης ιεράρχησης του αξιοπρόσεκτου κάθε περιοχής του οπτικού πεδίου, ένας “χάρτης προσοχής” (saliency map) [9]. Ένα διάγραμμα του ανθρώπινου οπτικού συστήματος δίνεται στο Σχήμα 2.2. Η μελέτη και η προσομοίωση τέτοιων χαρτών αποτελεί το κύριο σκοπό της διπλωματικής εργασίας. Στη συνέχεια παραθέτουμε μια σύντομη πραγμάτευση επί μέρους

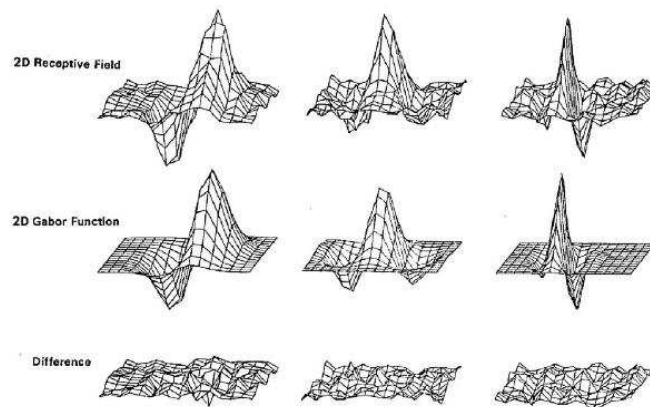
αποτελεσμάτων ερευνών πάνω στο ανθρώπινο, αλλά και άλλα βιολογικά οπτικά συστήματα, τα οποία συνδέονται με τα μοντέλα που θα εξετάσουμε.

2.1.2 Ανίχνευση Χαρακτηριστικών

Η ανίχνευση οπτικών χαρακτηριστικών (feature detection) έχει βρεθεί, από νευροβιολογικές και ψυχοφυσικές μελέτες, ότι αποτελεί μέρος της πρώιμης επεξεργασίας της οπτικής πληροφορίας στα βιολογικά συστήματα όρασης. Η επεξεργασία αυτή επιτρέπει στους οργανισμούς να αξιολογούν, με ένα γρήγορο τρόπο, τον όγκο της οπτικής πληροφορίας που δέχονται, τόσο χωρικά όσο και χωροχρονικά. Παραδείγματα τέτοιων χαρακτηριστικών του τρισδιάστατου κόσμου αποτελούν τα όρια των αντικειμένων, τα σχήματα των επιφανειών, η υφή (texture), η απόσταση από τον παρατηρητή, η κίνηση κ.ά. [1]. Τόσο τα βιολογικά όσο και τα τεχνητά συστήματα οδηγούνται στην εξαγωγή των χαρακτηριστικών αυτών χρησιμοποιώντας τις δισδιάστατες εικόνες που προσλαμβάνουν μέσω των ματιών ή μιας φωτογραφικής μηχανής αντίστοιχα. Η οπτική ερμηνεία του περιβάλλοντος, και η ανίχνευση πραγματικών, τρισδιάστατων χαρακτηριστικών, ανάγεται δηλαδή στην εξαγωγή κατάλληλων χαρακτηριστικών σε δισδιάστατες εικόνες.

Στην προσπάθεια να προσεγγιστεί το μοντέλο της πρώιμης αυτής επεξεργασίας της οπτικής πληροφορίας στα συστήματα των πρωτεύοντων θηλαστικών με γραμμικούς τελεστές, στο βαθμό που αυτό είναι δυνατό, αναπτύχθηκαν δύο εκ πρώτης άποψης αντίθετες απόψεις. Η πρώτη αντιμετωπίζει το μοντέλο αυτό ως χωρικούς, τοπικούς ανιχνευτές χαρακτηριστικών και η δεύτερη ως δέσμες φίλτρων στο πεδίο των συχνοτήτων. Ο διαχωρισμός αυτός, είναι, από μαθηματικής άποψης επιφανειακός, καθ' ότι, δεχόμενοι τη (σχεδόν έστω) γραμμική συμπεριφορά του μοντέλου, υπάρχει ένας δυϊσμός μεταξύ της χωρικής και της συχνοτικής αναπαράστασης γραμμικών χρονικά αμετάβλητων τελεστών, ως συνελίξεων και φίλτρων επιλογής συχνοτήτων αντίστοιχα. Η δυαδικότητα αυτή μπορεί να γίνει αντιληπτή με από κοινού ανάλυση στα πεδία του χώρου και της συχνότητας βάσει των ιδεών του Gabor². Βάσει των ιδεών αυτών, ο Daugman (1985) πρότεινε δισδιάστατα κατευθυνόμενα φίλτρα Gabor ως βέλτιστα μοντέλα των αποκρίσεων των απλών κυττάρων του ραβδωτού φλοιού (striate cortex). Στην εικόνα 2.3 επιδεικνύεται η πειραματική επιβεβαίωση αυτής της ιδέας, όπου οι κρουστικές αποκρίσεις δισδιάστατων φίλτρων Gabor φαίνονται να προσεγγίζουν τις αποκρίσεις των απλών κυττάρων [1].

²Ο Dennis (Dénes) Gabor(1900-1979) ήταν Ούγγρος ηλεκτρολόγος μηχανικός ο οποίος τιμήθηκε με το βραβείο Νόμπελ Φυσικής το 1971 για τη συμβολή του στην ανάπτυξη της ολογραφίας. Ασχολήθηκε κυρίως με την ηλεκτρονική οπτική και και μία από τις βασικές του ιδέες ήταν ότι, στην αναπαράσταση των εικόνων απαιτείται η χρήση όχι μόνο του πλάτους, αλλά και της φάσης.



Σχήμα 2.3: Αντιστοίχιση διδιάστατων φίλτρων Gabor με αποκρίσεις κυττάρων του οπτικού φλοιού, εικόνα από Daugman, 1985 (αναπαραγωγή από το [1]). Ο Daugman αναφέρει σχετικά: “Πάνω σειρά: Πειραματικά καταγεγραμμένα 2Δ προφίλ των δεκτικών πεδίων (σημ. receptive fields)τριών απλών κυττάρων από ραβδωτό φλοιό γάτας, από τους L. A. Palmer και J. P. Jones της Ιατρικής Σχολής του Πανεπιστημίου της Πεννσυλβάνια. Κάθε σχήμα δείχνει το ενεργοποιητικό ή ανασταλτικό αποτέλεσμα ενός μικρού φωτός ή σκοτεινού σημείου που αναβοσβήνει, με ρυθμό πυροδότησης ίσο με αυτόν του κυττάρου, σαν μια συνάρτηση της (x, y) θέσης του ερεθίσματος, υπολογισμένη με αντίστροφη συσχέτιση του 2Δ ερεθίσματος με την ακολουθία των νευρικών αποκρίσεων. Η δεύτερη σειρά δείχνει τη βέλτιστη 2Δ συνάρτηση Gabor για το προφίλ του δεκτικού πεδίου κάθε κυττάρου, με τις παραμέτρους υπολογισμένες με τη μέθοδο των ελαχίστων τετραγώνων. Η τρίτη σειρά δείχνει το σφάλμα - υπόλοιπο μεταξύ την καταγεγραμμένη απόκριση κάθε κυττάρου και την αντίστοιχη συνάρτηση Gabor. Σε τυποποιημένους στατιστικούς ελέγχους, τα υπόλοιπα δεν διακρίνονταν από τυχαία σφάλματα σε 33 από τα 36 απλά κύτταρα που ελέγχθηκαν”.

	Πρώιμος	Προσεκτικός
Δομή	Παράλληλη	Σειριακή
Ταχύτητα	Γρήγορη ($< 200ms$)	Πιο αργή ($> 200ms$)
Δυνατότητα Ελέγχου	Καμία	Ολική
Επίπεδο Συνείδησης	Χαμηλό	Υψηλό
Διανοητική Προσπάθεια	Χαμηλή	Υψηλή

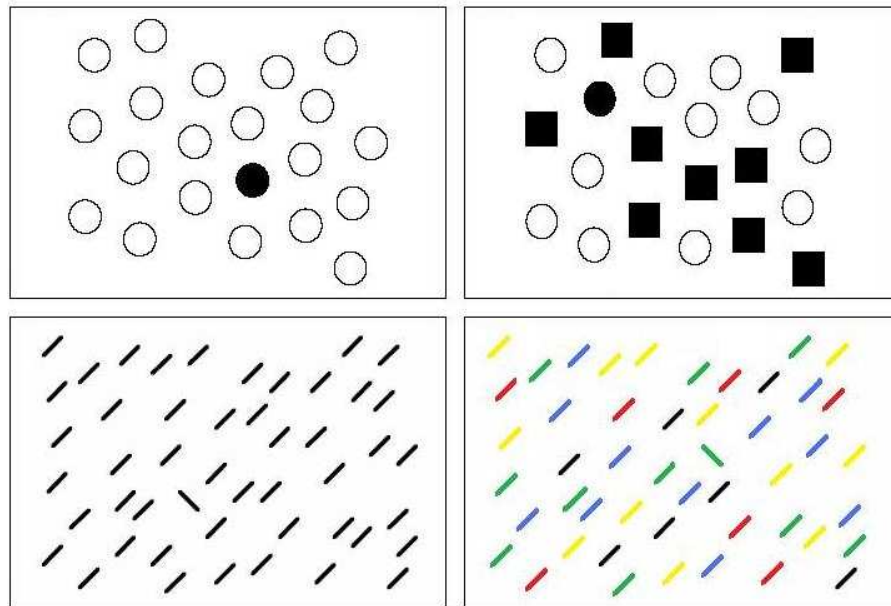
Πίνακας 2.1: Κύρια χαρακτηριστικά των πρώιμων και προσεκτικών μηχανισμών προσοχής. Από το [30].

2.1.3 Προσεκτική (attentive) και Πρώιμη (preattentive) Οπτική Αντίληψη

Από διάφορες μελέτες της συμπεριφοράς της ανθρώπινης προσοχής, οι οποίες διεξήχθησαν κυρίως στο δεύτερο μισό του 20ού αιώνα, έχει προκύψει το αποτέλεσμα ότι η προσοχή μπορεί να μοντελοποιηθεί με δύο διακριτούς μηχανισμούς. Οι μηχανισμοί αυτοί βρίσκουν εφαρμογή όχι μόνο στη χωρική οπτική προσοχή, αλλά και σε ακολουθίες εικόνων ή ηχητικές διεγέρσεις. Ένα σύντομο ιστορικό της εύρεσης αυτών των μηχανισμών δίνεται στο [30], από το οποίο παραθέτουμε ορισμένα ενδιαφέροντα στοιχεία.

Στη μελέτη τους το 1977 [40], οι Sneider και Schiffrin κατέδειξαν ότι ο χρόνος απόκρισης σε οπτικά ερεθίσματα εξαρτάται από τον αριθμό τους και το βαθμό στον οποίο διαφέρουν από δευτερεύοντα ερεθίσματα τα οποία τους αποσπούν την προσοχή. Ο χρόνος αντίδρασης βρέθηκε να κινείται σε δύο περιοχές τιμών: μια αστραπιαία απόκριση και πιο αργές αποκρίσεις οι οποίες αποτελούσαν προϊόν σε βάθους έρευνας της εικόνας. Βάσει αυτού, οι Sneider και Schiffrin συμπέραναν την ύπαρξη ενός αυτόματου, ασυνείδητου και ενός ελεγχόμενου, συνειδητού μηχανισμού προσοχής. Μετέπειτα ερευνητές επιβεβαίωσαν τα αποτελέσματα αυτά, καλώντας τους μηχανισμούς αυτούς εξωγενείς και ενδογενείς αντίστοιχα ή προσεκτικούς (attentive) και πρώιμους (preattentive), ορολογία που οφείλεται στους Treisman και Gelade και είναι η πλέον δημοφιλής σήμερα. Ο Πίνακας 2.1 συνοψίζει τα κύρια χαρακτηριστικά των δύο μηχανισμών αυτών.

Όσον αφορά ειδικότερα την οπτική προσοχή, στην εργασία της Treisman [44] διαπιστώθηκε ότι πρώιμα – preattentive – χαρακτηριστικά όπως η φωτεινότητα, το χρώμα και η κατεύθυνση των γραμμών δημιουργούν σαφή όρια μεταξύ αντικειμένων στο οπτικό πεδίο. Σε περιπτώσεις μάλιστα όπου τα όρια αυτά αποτελούνται από ιδιαίτερα εξέχοντα από το περιβάλλον τους αντικείμενα, ο χρόνος απόκρισης μπορεί να είναι τόσο μικρός όσο και $50ms$. Η διαφοροποίηση ενός από αυτά τα χαρακτηριστικά ανιχνεύεται αστραπιαία ως όριο, περίγραμμα ή ένδειξη ύπαρξης μιας διακριτής περιοχής. Αντίθετα, πιο σύνθετες διαφορές



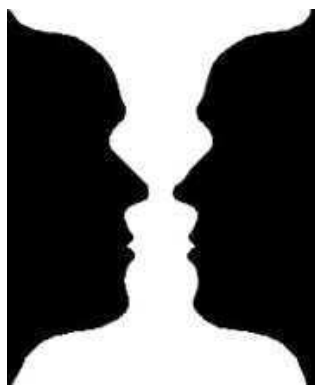
Σχήμα 2.4: Πάνω σειρά (από τους Healy et al., 1996 [13]) – αριστερά: εικόνα στην οποία ο πρώτος μηχανικός προσοχής δίνει άμεσο αποτέλεσμα για το αξιοπρόσεκτο σημείο. Δεξιά: παράδειγμα προσεκτικής οπτικής προσοχής – στην εικόνα οι distractors εμποδίζουν αποτελεσματικά την άμεση εύρεση του αξιοπρόσεκτου αντικειμένου αφού αυτό δε διαθέτει μοναδικά χαρακτηριστικά ώστε να ξεχωρίζει από αυτούς, απαιτώντας έτσι συνειδητή προσπάθεια. Κάτω σειρά – αριστερά: εικόνα όπου οι περισσότεροι παρατηρητές μπορούν, αρκετά εύκολα (αν και όχι τελείως αστραπιαία) να διακρίνουν το εξέχον αντικείμενο βάσει των πρωτογενών – πρώιμων του χαρακτηριστικών (από το [30]). Δεξιά: παρόμοια εικόνα, όπου η εύρεση του καθίσταται πιο δύσκολη και αρχίζει να απαιτεί κάποια συνειδητή προσπάθεια, καθ' ότι τα δευτερεύοντα ερεθίσματα (εδώ, χρωματικές διαφορές) αποσπούν μεγάλο μέρος της αρχικής προσοχής.

δεν ανιχνεύονται σε αυτό το στάδιο [2]. Η προσεκτική – attentive – διαδικασία, η οποία ενεργοποιείται σε βάθος χρόνου (βλ. Πίνακα 2.1), βασίζεται στην πρότερη γνώση για τον οπτικό κόσμο και σε συνειδητή επεξεργασία του οπτικού ερεθίσματος. Ένα παράδειγμα στο οποίο είναι ευδιάκριτη η ενεργοποίηση του καθ' ενός από τους δύο αυτούς μηχανισμούς περιλαμβάνεται στις εικόνες του Σχήματος 2.4. Πρέπει να σημειώσουμε βέβαια ότι τα όρια μεταξύ των μορφών απόκρισης που τους δεν είναι σε απόλυτο βαθμό καθορισμένα — ακόμα και για πρώιμα οπτικά χαρακτηριστικά, ήδη αναφέραμε ότι σημαντικό ρόλο παίζει ο αριθμός τους, το πόσο διαφέρουν από δευτερεύοντα ερεθίσματα — “ αποσπαστές προσοχής ” (distractors) αλλά και από τον αριθμό και τη μοναδικότητα των τελευταίων. Μια πρόσφατη μελέτη όπου γίνεται προσπάθεια να εξακριβωθεί πιο συστηματικά η επίδραση των χαρακτηριστικών των δευτερευόντων ερεθισμάτων είναι η [38]. Στο Σχήμα 2.4 δίνουμε μερικά παραδείγματα εικόνων - συνόλων ερεθισμάτων οι οποίες είτε αξιολογούνται οριακά στο ποιος μηχανισμός από τους δύο τείνει να κυριαρχήσει είτε οι distractors ουσιαστικά απενεργοποιούν σε αυτές το πρώιμο στάδιο οπτικής προσοχής.

Τα αποτελέσματα αυτά σχετικά με την πρώιμη ή ανοδική (bottom-up) οπτική προσοχή οδήγησαν την Treisman στην υπόθεση της ύπαρξης ενός κύριου χάρτη που περιέχει τις σημαντικές περιοχές της εικόνας και λαμβάνει πληροφορία σχετικά με το καθένα από τα πρώιμα οπτικά χαρακτηριστικά. Αυτή η υπόθεση αποτελεί και τη βασική ιδέα πίσω από πολλά μοντέλα που προσεγγίζουν την ανοδική διαδικασία, με ένα από αυτά το μοντέλο των Itti, Koch και Niebur, το οποίο εξετάζουμε διεξοδικά στην ενότητα 3.2 και η διερεύνηση της δυνατότητας βελτίωσής του αποτελεί ένα από τα κύρια σημεία της διπλωματικής εργασίας.

Η επεξεργασία της οπτικής πληροφορίας που λαμβάνεται από το φυσικό κόσμο δεν είναι βέβαια μια στατική διαδικασία, καθ' ότι η ροή της είναι συνεχής. Όσο εξοικειωνόμαστε με μια εικόνα, η καλύτερη και υψηλότερου επιπέδου κατανόηση που αναπτύσσουμε γι αυτήν οδηγεί στην κυριαρχία καθοδικών (top-down) προσεκτικών μηχανισμών, όπως, για παράδειγμα, ανάκληση γνώσης από τη μνήμη για παρόμοια αντικείμενα, προηγούμενες εικόνες από την ίδια περιοχή στην οποία αναφέρεται η σκηνή κλπ. Για παράδειγμα, για τα σχήματα από την εικόνα 2.4 στα οποία το αξιοπρόσεκτο αντικείμενο δεν είναι εξ' αρχής προφανές, βλέποντάς τα για δεύτερη φορά, διαπιστώνουμε ότι μπορούμε να βρούμε πολύ πιο εύκολα τις περιοχές που διαφέρουν. Στη συνεχή ροή πληροφορίας, η συνειδητή αυτή κατανόηση των εικόνων οδηγεί σταδιακά την προσεκτική διαδικασία να κυριαρχήσει πάνω στα πρώιμα ερεθίσματα, ωστόσο βρεθούμε σε μια τελείως διαφορετική σκηνή κ.ο.κ., έχουμε δηλαδή ένα διαρκή ανταγωνισμό.

Ο ανταγωνισμός ενυπάρχει επίσης μεταξύ των συστατικών της ανοδικής οπτικής προσοχής, και αντίστοιχα μεταξύ διαφορετικών ερμηνειών μιας εικόνας στην καθοδική οπτική προσοχή. Στην πρώτη περίπτωση, η περιοχή - νικητής του ανταγωνισμού “ καταργεί ” για ένα χρονικό διάστημα όλες τις υπόλοιπες



Σχήμα 2.5: Στην εικόνα αυτή του Rubin διακρίνουμε είτε ένα λευκό βάζο είτε δύο πρόσωπα σε προφίλ, αλλά όχι και τα δύο ταυτόχρονα – στρατηγική “ όλα - για - το - νικητή ” (winner-take-all)

και συγκεντρώνει όλη την προσοχή μας. Η ορολογία που εκφράζει αυτό το φαινόμενο ονομάζεται στρατηγική “ όλα - για - το - νικητή ” (winner-take-all, WTA). Στη δεύτερη περίπτωση, μόνο μια συνειδητή αντίληψη της εικόνας είναι συνήθως δυνατή. Ένα τέτοιο παράδειγμα αποτελεί η πασίγνωστη εικόνα η οποία οφείλεται στον ψυχολόγο Edgar Rubin (εικόνα 2.5) και σε αυτήν μπορούμε εναλλακτικά να δούμε δύο σκούρα πρόσωπα σε προφίλ το ένα απέναντι από το άλλο ή ένα λευκό βάζο — είναι όμως σχεδόν αδύνατον να δούμε και τα δύο ταυτόχρονα. Ο ανταγωνισμός και η στρατηγική “ όλα - για - το - νικητή ” αποτελεί κύριο στοιχείο του μοντέλου των Itti et al. (ενότητα 3.2).

2.2 Ανάλυση Εικόνων σε Πολλαπλές Κλίμακες

Η χρήση τεχνικών ανάλυσης σε πολλαπλές κλίμακες (multiscale image analysis) έχει βρεθεί να είναι πολύ χρήσιμη για σκοπούς εύρεσης χαρακτηριστικών και αντικειμένων μέσα σε εικόνες³. Σε πολλές καταστάσεις, τα πιο ενδιαφέροντα αντικείμενα και χαρακτηριστικά αυτά ανήκουν σε μία κλίμακα της εικόνας, και τα υπόλοιπα σε μία άλλη. Σε κάποιες όμως περιπτώσεις η δυαδική αυτή διάκριση δεν είναι δυνατή, καθ’ ότι η πληροφορία που μας ενδιαφέρει παρουσιάζεται σε διάφορες κλίμακες. Έτσι έχουν αναπτυχθεί κλιμακωτές μέθοδοι ανάλυσης, βάσει των οποίων εξάγεται μια σειρά από όλο και πιο αδρομερείς αναπαραστάσεις της αρχικής εικόνας, με σκοπό στη συνέχεια να

³Η εισαγωγή στην ανάλυση εικόνων σε πολλαπλές κλίμακες βασίζεται σε μεγάλο βαθμό στην αντίστοιχη που γίνεται στο [33]

χρησιμοποιήσουμε όλες ώστε να αλιεύουμε τη ζητούμενη πληροφορία.

Μια τέτοια πολυκλιμακωτή αναπαράσταση ορίζεται από τους μετασχηματισμούς από μια λεπτή στην επόμενη πιο αδρομερή κλίμακα. Πρακτικά, στην ανάλυση την οποία θα υλοποιήσουμε, αυτό ισοδυναμεί με το φιλτράρισμα της εικόνας (με ένα γραμμικό, π.χ. γκαουσιανό ή ένα μη γραμμικό / μορφολογικό φίλτρο) — επίσης συνήθως προβαίνουμε και στην υποδειγματοληψία του υψηλότερου, πιο αδρομερούς επιπέδου. Στην τελευταία περίπτωση το σύνολο αυτό των αναπαραστάσεων καλείται *πυραμίδα* για λόγους τους οποίους εύκολα μπορούμε να φανταστούμε αν θεωρήσουμε μια τέτοια αναπαράσταση με τα πιο αδρομερή, υποδειγματολημμένα επίπεδα πάνω από τα αντίστοιχα, μεγαλύτερα σε μέγεθος, λεπτομερή.

Είναι δυνατόν να οριστεί ένας μεγάλος αριθμός μετασχηματισμών πολυκλιμακωτής ανάλυσης. Περιορίζουμε τις επιλογές αυτές καθορίζοντας κάποιες επιθυμητές ιδιότητες των μετασχηματισμών αυτών [33], ξεκινώντας αρχικά από τις γενικές ιδιότητες που προσδίδουν ανεξαρτησία σε συγκεκριμένες αλλαγές, όπως:

- Χωρική ανεξαρτησία – ανεξαρτησία από τη μετατόπιση.
- Ισοτροπική ανεξαρτησία – ανεξαρτησία από την περιστροφή.
- Ανεξαρτησία από τη μεταβολή της φωτεινότητας: ο μετασχηματισμός πρέπει να προσαρμόζεται με κάθε *αναμόρφωση* της φωτεινότητας, δηλαδή κάθε αλλαγή που την αυξάνει σημείο-προσ-σημείο.

Μπορούν να προστεθούν και άλλες απαιτήσεις οι οποίες αφορούν το ίδιο το αποτέλεσμα του μετασχηματισμού:

- Ο μετασχηματισμός πρέπει να αποτελεί μια πραγματική απλοποίηση της εικόνας – αυτό σημαίνει ότι είναι μη αναστρέψιμος, καθ' όσον κάποια πληροφορία χάνεται από τη μία κλίμακα στην άλλη.
- Μια συγκεκριμένη μορφή απλοποίησης εκφράζεται από την αρχή του μεγίστου: μεταξύ οποιωνδήποτε διαφορετικών κλιμάκων, η μέγιστη φωτεινότητα στην πιο αδρομερή κλίμακα είναι πάντα χαμηλότερη από την αντίστοιχη μέγιστη στην λεπτομερή κλίμακα. Το αντίστροφο ισχύει για την ελάχιστη φωτεινότητα, η οποία θα πρέπει να είναι πάντα υψηλότερη.
- Αιτιατότητα (causality): οι αδρομερείς κλίμακες καθορίζονται αποκλειστικά από το τι συμβαίνει στις πιο λεπτομερείς.
- Ο μετασχηματισμός συνεπώς δεν μπορεί να δημιουργεί νέες δομές στις αδρότερες κλίμακες — συνήθως απαιτείται ότι δε θα δημιουργεί νέα τοπικά ακρότατα.

Επιπλέον, αν στεχεύουμε σε κατάτμηση της εικόνας – και γενικότερα σε ένα καλό προσδιορισμό των ορίων των αντικειμένων, χαρακτηριστικό χρήσιμο στην ανάλυση που θα πραγματοποιήσουμε στο κύριο μέρος της διπλωματικής εργασίας – ένα επιπλέον προαπαιτούμενο θα ήταν η διατήρηση της ευκρίνειας και σχετικής θέσης των περιγραμμάτων.

Οι γραμμικές αναπαραστάσεις, βάσει των περιορισμών που περιγράψαμε πιο πάνω, οδηγούν στους γραμμικούς χώρους κλίμακας (linear scale-spaces). Σε αυτούς, ο τελεστής με τον οποίο ανεβάνουμε σε πιο αδρή κλίμακα είναι ένα Γκαουσιανό φιλτράρισμα — μια τέτοια αναπαράσταση θα εξετάσουμε στην ενότητα 2.2.1. Δυστυχώς όμως αυτό οδηγεί σε θόλωμα της εικόνας και απώλεια της ευκρίνειας των ορίων των αντικειμένων, η οποία είναι ιδιαίτερα επιθυμητή, όπως αναφέραμε. Ένα άλλο πρόβλημα είναι η εμφάνιση ψευδών μεγίστων στις πολύ υψηλές (αδρές) κλίμακες, τα οποία, μαζί με το θόλωμα κάνουν αδύνατη την αναγνώριση των αντικειμένων. Αν και έχουν προταθεί αρκετές λύσεις για αυτό το φαινόμενο, δεν έγινε δυνατόν να εξαλειφθεί πλήρως.

Οι μορφολογικές αναπαραστάσεις περιλαμβάνουν ως στοιχειώδη συστατικά διαστολές (dilations) και διαβρώσεις (erosions) αυξανόμενου μεγέθους — περισσότερα για τους τελεστές αυτούς αναφέρουμε στην ενότητα 2.2.2. Ένας τελεστής ψ καλείται μορφολογικό φίλτρο εάν είναι [33]

- Επεκτατικός: $g > h \Rightarrow \psi g > \psi h$. Αυτό υπονοεί ότι $\psi(h \wedge k) < \psi h \wedge \psi k$ και $\psi(h \vee k) > \psi h \vee \psi k$.
- Αυτοδύναμος (idempotent): $\psi\psi = \psi$. Αυτό σημαίνει ότι ο τελεστής είναι σταθερός – επαρκεί μόνο μια εφαρμογή του ώστε να πάρουμε το τελικό αποτέλεσμα.

Κάποιες πρώτες προσεγγίσεις αφορούσαν τις κοκκιδομετρίες (granulometries) που αφορούσαν σε μια συνεχούς κλίμακας οικογένεια ανοιγμάτων (openings) και κλεισιμάτων (closings) – τα τελευταία είναι τελεστές που προέρχονται από διαστολές και διαβρώσεις. Τέτοιες μέθοδοι δεν δημιουργούν ψευδή μέγιστα και διατηρούν καλύτερα τις άκρες που περιέχονται στην εικόνα. Μια ακόμα πιο ισχυρή κλάση μορφολογικών φίλτρων είναι τα ανοίγματα και κλεισίματα με ανακατασκευή (reconstruction), τα οποία επιπλέον διατηρούν τα οριζόντια περιγράμματα. Τέλος, μια γενική κλάση τέτοιων φίλτρων είναι τα levelings, τα οποία επιλύουν τα πλείστα από τα προβλήματα που αναφέρουμε και έχουν κάποιες πολύ χρήσιμες ιδιότητες, όπως θα δούμε στην ενότητα 2.2.5.

Στη συνέχεια ακολουθεί μια σύντομη περιγραφή των διαφόρων τύπων πυραμίδων που αντιστοιχούν στις κύριες γραμμικές και μορφολογικές προσεγγίσεις που αναφέραμε. Πολλά στοιχεία δίνονται στο [3], από τα οποία αναφέρουμε κυρίως όσα χρησιμοποιήσαμε στην παρούσα διπλωματική εργασία.

2.2.1 Γκαουσιανή Πυραμίδα

Η Γκαουσιανή πυραμίδα προτάθηκε από τους Burt και Adelson στο [4] το 1983 και αποτελεί τον πλέον διαδεδομένο τύπο γραμμικής πυραμίδας, τον οποίο μάλιστα επέλεξαν οι Itti, Koch και Niebur στο μοντέλο οπτικής προσοχής που δημιούργησαν, όπως θα δούμε και στο επόμενο κεφάλαιο.

Η δημιουργία της πυραμίδας συνίσταται στο βαθυπερατό φιλτράρισμα του επιπέδου $n - 1$ με έναν γκαουσιανό δισδιάστατο πυρήνα και την υποδειγματοληψία (κατά παράγοντα δύο) ώστε να προκύψει το επίπεδο n . Συμβολίζοντας με $G_i, i = [1..n]$ το i -οστό επίπεδο της πυραμίδας και K τον Γκαουσιανό πυρήνα που χρησιμοποιούμε, έχουμε

$$G_0 = I \text{ (αρχική εικόνα)}$$

$$G_i = (\text{υποδ}) [G_{i-1} * K], \quad i = [1..n - 1]$$

οπότε προκύπτει μια πυραμίδα n -επιπέδων. Ο Γκαουσιανός πυρήνας K που χρησιμοποιούμε στην παρούσα εργασία προκύπτει από τη συνέλιξη δύο μονοδιάστατων πυρήνων μήκους 5 με στοιχεία $[1/16, 1/4, 3/8, 1/4, 1/16]$, είναι δηλαδή ο 5×5 πυρήνας

$$K = \begin{bmatrix} 0.0039 & 0.0156 & 0.0234 & 0.0156 & 0.0039 \\ 0.0156 & 0.0625 & 0.0938 & 0.0625 & 0.0156 \\ 0.0234 & 0.0938 & 0.1406 & 0.0938 & 0.0234 \\ 0.0156 & 0.0625 & 0.0938 & 0.0625 & 0.0156 \\ 0.0039 & 0.0156 & 0.0234 & 0.0156 & 0.0039 \end{bmatrix}$$

Η διαχωριστικότητα του πυρήνα μας επιτρέπει να αντικαταστήσουμε το ένα δισδιάστατο φιλτράρισμα με δύο φιλτραρίσματα κατά μήκος της κάθε διάστασης, εκμεταλλευόμενοι την επιμεριστικότητα της συνέλιξης:

$$K = K_1 * K_2 \Rightarrow I * K = I * (K_1 * K_2) = (I * K_1) * K_2$$

Επιτυγχάνεται δηλαδή η μείωση της υπολογιστικής πολυπλοκότητας από $O(n^2)$ σε $O(2n)$, που μπορεί να είναι πολύ σημαντική για πυρήνες μεγάλου μεγέθους. Ακόμα πιο σημαντική γίνεται στις περιπτώσεις όπου χρησιμοποιείται τρισδιάστατη συνέλιξη, όπως στις εφαρμογές που θα συναντήσουμε στο κεφάλαιο 4.

Κατά τη διαδικασία που περιγράψαμε πιο πάνω μπορούμε επίσης να προσεγγίσουμε τις αποκρίσεις φίλτρων Λαπλασιανής - της - Γκαουσιανής (Laplacian-of-Gaussian, LoG) πάνω σε κάθε επίπεδο της πυραμίδας. Προκύπτει έτσι η λεγόμενη Λαπλασιανή πυραμίδα, η οποία αντιπροσωπεύει τα ζωνοπερατά συνιστώμενα μέρη της εικόνας, (κατ' αντιστοιχία, η Γκαουσιανή πυραμίδα αντιπροσωπεύει τα βαθυπερατά συνιστώμενα μέρη), προκύπτει από τη διαφορά δύο διαδοχικών επιπέδων της Γκαουσιανής πυραμίδας. Πιο κάτω διατηρούμε τον προηγούμενο

συμβολισμό και επίσης συμβολίζουμε με $L_i, i = [1..n]$ το i -οστό επίπεδο της Λαπλασιανής πυραμίδας. Ο ολοκληρωμένος αλγόριθμος υπολογισμού και των δύο πυραμίδων έχει ως εξής:

$$G_{i+1,0} = K * G_n; L_i = G_i - G_{i+1,0}$$

$$G_{i+1} = (G_{i+1,0}) \downarrow_2, i = [1..n - 1]$$

Η Λαπλασιανή πυραμίδα περιέχει συνήθως πολύτιμη πληροφορία για τα αντικείμενα και άλλα στοιχεία που υπάρχουν σε κάθε εικόνα, καθώς και την κλίμακα η οποία τα χαρακτηρίζει. Στο Σχήμα 2.6 παρουσιάζουμε τα πρώτα τέσσερα επίπεδα των δύο πυραμίδων, για την εικόνα harbor.

2.2.2 Adjunction Πυραμίδα

Η Adjunction πυραμίδα είναι η πρώτη από της μη-γραμμικές πυραμίδες που θα εξετάσουμε. Κάθε επίπεδο της πυραμίδας προκύπτει από *διάβρωση* (erosion) του κατώτερου επιπέδου με ένα μορφοποιητικό στοιχείο S και υποδειγματοληψία του, κατ' αντιστοιχία με τη μέθοδο που ακολουθήσαμε στη Γκαουσιανή πυραμίδα. Κατά την αναδόμηση των κατώτερων (πιο λεπτομερών) επιπέδων από τα ανώτερα – διαδικασία την οποία δε θα χρησιμοποιήσουμε εδώ – χρησιμοποιείται η συμπληρωματική της διάβρωσης πράξη, η διαστολή (dilation). Η χρήση των δύο αυτών τελεστών δίνει και στην πυραμίδα το όνομά της, καθ' ότι adjunction καλούμε κάθε ζεύγος τελεστών, έστω (ϵ, δ) για τους οποίους ισχύει

$$\delta(x) \leq Y \Leftrightarrow X \leq \epsilon(Y)$$

για κάθε σύνολα - σήματα X, Y τα οποία ανήκουν στο ίδιο πλέγμα (lattice). Πλέγμα είναι ένα μερικώς ταξινομημένο σύνολο στο οποίο κάθε δύο στοιχεία του έχουν ένα μέγιστο και ένα ελάχιστο. Ο αναγνώστης που επιθυμεί να εμβαθύνει στο μαθηματικό υπόβαθρο που έχει αναπτυχθεί για να στηρίξει τους μορφολογικούς τελεστές μπορεί να ανατρέξει στο [1].

Η διάβρωση μιας εικόνας I , από το S , η οποία μαθηματικά συμβολίζεται με ϵ_S ορίζεται μαθηματικά ως εξής:

$$\epsilon_S(I)(x) = \max_{y \in S, x+y \in F} I(x+y)$$

όπου F είναι το πεδίο ορισμού της I . Αντίστοιχα σημειώνουμε ότι η διαστολή μιας εικόνας I από ένα μορφοποιητικό στοιχείο S συμβολίζεται με δ_S και ορίζεται ομοίως ως

$$\delta_S(I)(x) = \max_{y \in S, x-y \in F} I(x-y)$$



Σχήμα 2.6: Γκαουσιανή (αριστερά) και Λαπλασιανή (δεξιά) πυραμίδα (και πρώτα τέσσερα επίπεδα) της εικόνας *harbor* την αρχική μορφή της οποίας βλέπουμε στο επίπεδο 0 της Γκαουσιανής πάνω αριστερά. Η Λαπλασιανή πυραμίδα έχει κανονικοποιηθεί μεταξύ των ελάχιστων και μέγιστων τιμών κάθε επιπέδου ώστε να είναι πιο ευδιάκριτη. Βλέπουμε ότι καθώς ανεβαίνουμε στα επίπεδα, τα όρια των αντικειμένων γίνονται λιγότερο σαφή.

Οι τελεστές erosion και dilation, σε απλά λόγια, αντικαθιστούν κάθε τιμή της εικόνας από την ελάχιστη ή τη μέγιστη αντίστοιχα, σε μια περιοχή που ορίζεται από το δομικό στοιχείο S . Παράδειγμα εφαρμογής των τελεστών αυτών δίνεται στο Σχήμα 2.8. Ο αναγνώστης που επιθυμεί να εξετάσει πιο λεπτομερώς στο μαθηματικό υπόβαθρο μπορεί να ανατρέξει στο [1].

Παρόμοια με τη Λαπλασιανή πυραμίδα η οποία προκύπτει από την Γκαουσιανή, υπολογίζουμε και εδώ την αντίστοιχη ποσότητα στην οποία και πάλι θα αναφερόμαστε (μάλλον καταχρηστικά) ως Λαπλασιανή και θα συμβολίζουμε με L^A , ενώ με A θα συμβολίζουμε την Adjunction πυραμίδα. Ο αλγόριθμος κατασκευής των δύο πυραμίδων είναι συνεπώς ο εξής:

$$\begin{aligned} A_0 &= I \text{ (αρχική εικόνα)} \\ A_{i+1,0} &= \epsilon_S(A_i); L_i^A = A_i - A_{i+1,0} \\ A_{i+1} &= (A_{i+1,0}) \downarrow_2, i = [1..n - 1] \end{aligned}$$

Ως μορφοποιητικό στοιχείο S στην παρούσα εργασία χρησιμοποιούμε το μοναδιαίο δίσκο,

$$S = \begin{array}{c} \bullet \\ \bullet \bullet \bullet \\ \bullet \end{array}$$

Καθότι τα μη - γραμμικά μορφολογικά φίλτρα όπως το πιο πάνω διατηρούν τα γεωμετρικά χαρακτηριστικά των εικόνων καλύτερα από ότι τα γραμμικά εξομαλυντικά φίλτρα, όπως το Γκαουσιανό, σε πολλές περιπτώσεις υπερτερούν σε εφαρμογές πολυκλιμακωτής ανάλυσης [43]. Μια τέτοια εφαρμογή είναι και η εύρεση ενδιαφερόντων (salient) σημείων και περιοχών σε διάφορες κλίμακες τις εικόνας — στο επόμενο κεφάλαιο θα εξετάσουμε αν αυτό ισχύει και στη δική μας περίπτωση. Προς το παρόν, στο Σχήμα 2.7 βλέπουμε τα αντίστοιχα με το Σχήμα 2.6 αποτελέσματα για την Adjunction πυραμίδα. Αν και η πιο καλή διατήρηση των χαρακτηριστικών όπως τα σχήματα και η μεγαλύτερη σαφήνεια των ορίων των αντικειμένων είναι εμφανής, οι διαστολές και οι διαβρώσεις από μόνες τους δεν είναι οι πλέον κατάλληλες για να αναπαριστούν τις διαδοχικές κλίμακες καθ' ότι αλλάζουν τη θέση των ορίων της εικόνας [19]. Μια εναλλακτική μορφολογική αναπαράσταση περιγράφεται στην επόμενη ενότητα.

2.2.3 Open-Close Πυραμίδα

Η πυραμίδα Ανοίγματος - Κλεισίματος (open-close) προτάθηκε από τους Sun και Maragos στο [43] και ο τρόπος κατασκευής της είναι όμοιος με αυτόν της Adjunction πυραμίδας, με τη διαφορά ότι, κατά την κατασκευή του επόμενου



Σχήμα 2.7: Adjunction (αριστερά) και αντίστοιχη “Λαπλασιανή” L^A (δεξιά) πυραμίδα (και πρώτα τέσσερα επίπεδα) της εικόνας *harbor* την αρχική μορφή της οποίας βλέπουμε στο επίπεδο 0 της Adjunction πυραμίδας πάνω αριστερά. Η L^A πυραμίδα έχει κανονικοποιηθεί μεταξύ των ελάχιστων και μέγιστων τιμών κάθε επιπέδου ώστε να είναι πιο ευδιάκριτη. Είναι εμφανής η διατήρηση των κύριων γεωμετρικών χαρακτηριστικών στις μεγαλύτερες κλίμακες σε αντίθεση με την αντίστοιχη Γκαουσιανή (συγκρίνετε με την εικόνα 2.6)



Σχήμα 2.8: Παράδειγμα εφαρμογής των τελεστών διάβρωσης, διαστολής, ανοίγματος και κλείσιματος πάνω σε δυαδική εικόνα με σχήμα νησιού (Αστυπάλαια). Πάνω: αρχική εικόνα, μορφοποιητικό στοιχείο S και διάβρωσή της από αυτό (erosion). Κάτω: κατά σειρά διαστολή (dilation), άνοιγμα (opening) και κλείσιμο (closing) της αρχικής εικόνας με το S .

επιπέδου, χρησιμοποιούμε τον τελεστή ανοίγματος (opening) αντί του τελεστή διάβρωσης (erosion). Το άνοιγμα, α , μιας εικόνας, I , με ένα μορφοποιητικό στοιχείο S ισοδυναμεί με τη διαδοχική διάβρωση και διαστολή της με το στοιχείο αυτό:

$$\alpha_S(I) = \delta_S(\epsilon_S(I))$$

Αναφέρουμε ότι αντίστοιχα το κλείσιμο β μιας εικόνας ισοδυναμεί με πρώτα διαστολή και έπειτα διάβρωση της εικόνας:

$$\beta_S(I) = \epsilon_S(\delta_S(I))$$

Αποτέλεσμα της εφαρμογής των τελεστών σε μια εικόνα είναι γενικά η απλοποίηση του σχήματος. Τα openings ουσιαστικά ελαττώνουν τις κορυφές ενός σχήματος, ενώ τα closings τις κοιλάδες του - σε μια εικόνα αυτό ισοδυναμεί με εξάλειψη των στενών εξάρσεων - "χερσονήσων" της εικόνας για το πρώτο και σε γέμισμα των στενών "κολπίσκων" για το δεύτερο. Ένα ενδεικτικό παράδειγμα, για δυαδική εικόνα, βλέπουμε στο Σχήμα 2.8.

Συμβολίζοντας με M την open-close (ή Sun - Μαραγκός) πυραμίδα ενώ με L^M την αντίστοιχη "Λαπλασιανή", ο αλγόριθμος κατασκευής των δύο πυραμίδων είναι ο

$$\begin{aligned}
 M_0 &= I \text{ (αρχική εικόνα)} \\
 M_{i+1,0} &= \alpha_S(M_i); L_i^M = M_i - M_{i+1,0} \\
 M_{i+1} &= (M_{i+1,0}) \downarrow_2, i = [1..n - 1]
 \end{aligned}$$

Ο συγκεκριμένος τύπος πυραμίδας έχει δείχθει στο [43] ότι οδηγεί σε αναπαραστάσεις της εικόνας με μικρότερη εντροπία από τις αντίστοιχες γραμμικές πυραμίδες. Χρησιμοποιώντας τον τελεστή ανοίγματος στην πυραμίδα αναμένουμε μια σχετική απλοποίηση των γεωμετρικών στοιχείων τα οποία είναι αναντίστοιχα λεπτά για τη συγκεκριμένη κλίμακα, με ταυτόχρονη όμως διατήρηση των κύριων στοιχείων που αντιστοιχούν σε αυτήν. Στο Σχήμα 2.9 βλέπουμε, αντίστοιχα με τα προηγούμενα, αποτελέσματα για την open-close πυραμίδα, με εμφανή τη διατήρηση των κύριων χαρακτηριστικών, όπως τα όρια των αντικειμένων, της εικόνας και στα μεγαλύτερα επίπεδα.

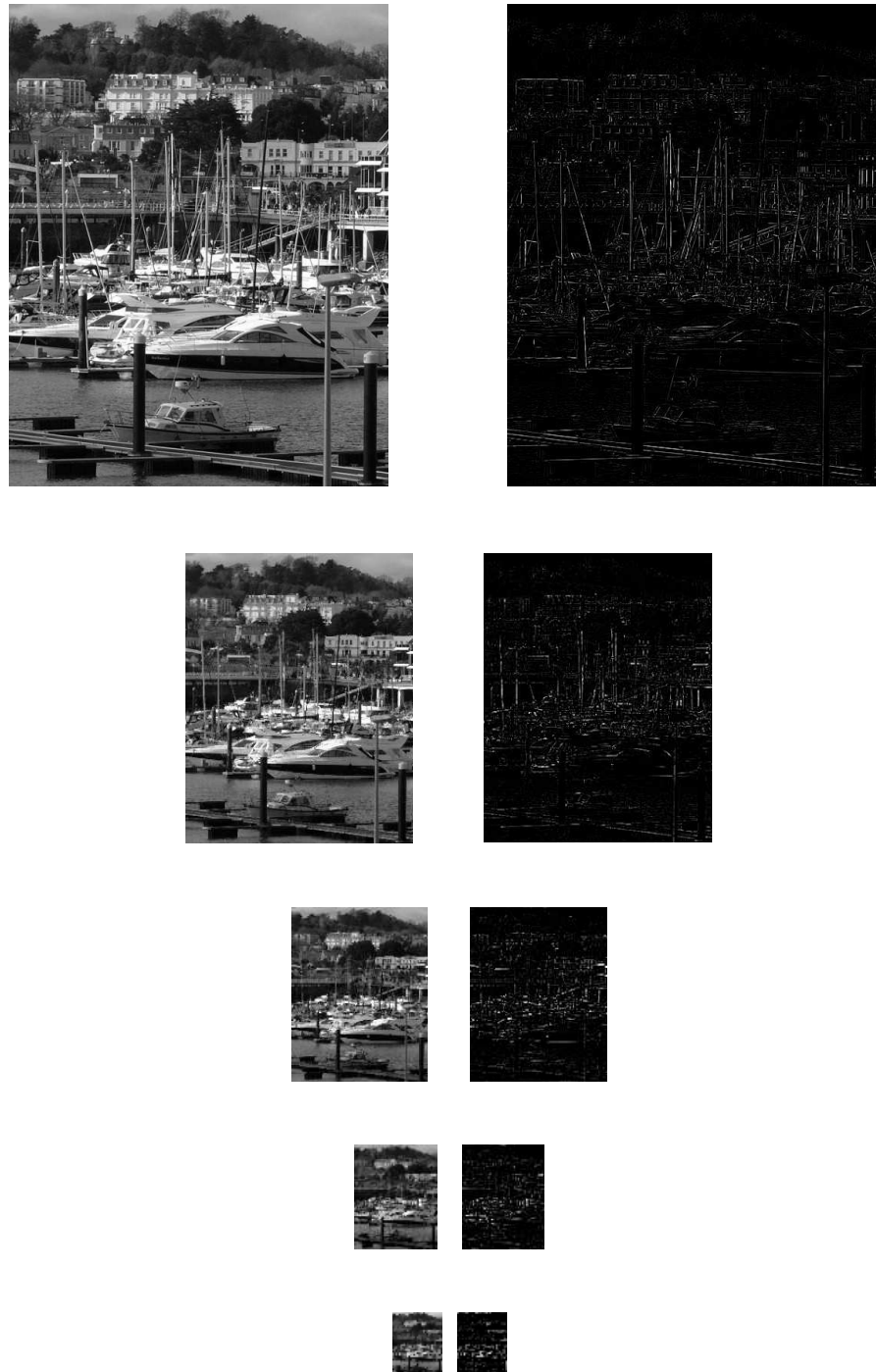
2.2.4 Haar - Wavelet Πυραμίδα

Η αποδόμηση σημάτων μέσω κυματιδίων (wavelets) είναι ένα σχετικά νέο εργαλείο [3], γραμμικό στην αρχική του μορφή. Έχουν όμως προταθεί αρκετές μη γραμμικές επεκτάσεις του, μία από τις θα χρησιμοποιήσουμε στην υλοποίηση της πυραμίδας. Τα σχήματα τόσο της γραμμικής όσο και της μη γραμμικής αναπαράστασης σε πολλαπλές κλίμακες με τη βοήθεια κυματιδίων υπάγονται σε ένα ενιαίο αξιωματικό πλαίσιο. Τα επιμέρους σχήματα γραμμικής ή μη αποδόμησης αποτελούν υποπεριπτώσεις του τελευταίου. Το πλαίσιο διακρίνει τα σχήματα αυτά ως συζευγμένα (coupled) ή μη συζευγμένα (uncoupled) — περισσότερες πληροφορίες υπάρχουν στο [14].

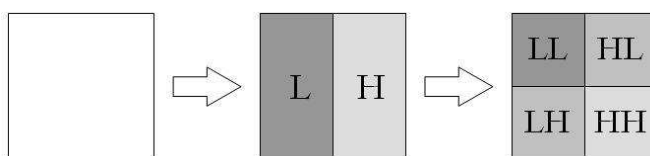
Σε κάθε σχήμα διακρίνουμε το *χώρο σήματος* και το *χώρο λεπτομέρειας*, οι οποίοι περιλαμβάνουν ο πρώτος το επιλεγμένο κομμάτι του σήματος εισόδου ως πληροφορία και ο δεύτερος το υπόλοιπο. Συγκρατώντας τόσο το σήμα όσο και τη λεπτομέρεια, καθώς μεταβαίνουμε από το ένα επίπεδο στο άλλο, είναι δυνατόν να ανακατασκευάσουμε την αρχική εικόνα. Στην περίπτωσή μας βέβαια σκοπός είναι μόνο η δημιουργία της πυραμίδας, η οποία όμως θα περιλαμβάνει και το χώρο λεπτομέρειας κατ' αντιστοιχία της Λαπλασιανής της Γκαουσιανής πυραμίδας, όπως θα δούμε πιο κάτω.

Μορφολογικό κυματίδιο Haar

Η μορφολογική εκδοχή του μονοδιάστατου κυματιδίου Haar υλοποιεί ένα μη γραμμικό τελεστή για την ανάλυση του σήματος, ο οποίος μπορεί να επιλέγει, μεταξύ δυο δειγμάτων το ελάχιστο (erosion) ή το μέγιστο (dilation). Αντίστοιχα, σε δύο ή περισσότερες διαστάσεις ο μετασχηματισμός κυματιδίων μπορεί να προκύψει από διαδοχική εφαρμογή του μονοδιάστατου μετασχηματισμού κατά



Σχήμα 2.9: Πυραμίδα ανοίγματος - κλεισίματος (open-close) (αριστερά) και αντίστοιχη “Λαπλασιανή” L^M (δεξιά) μαζί με τα πρώτα τέσσερα επίπεδα από πάνω προς τα κάτω, για την εικόνα *harbor* την αρχική μορφή της οποίας βλέπουμε στο επίπεδο 0 της (open-close) πάνω αριστερά. Η L^M πυραμίδα έχει κανονικοποιηθεί μεταξύ των ελάχιστων και μέγιστων τιμών κάθε επιπέδου ώστε να είναι πιο ευδιάκριτη. Εδώ, όπως και στην Adjunction πυραμίδα βλέπουμε τη διατήρηση των γεωμετρικών χαρακτηριστικών στις υψηλές κλίμακες.

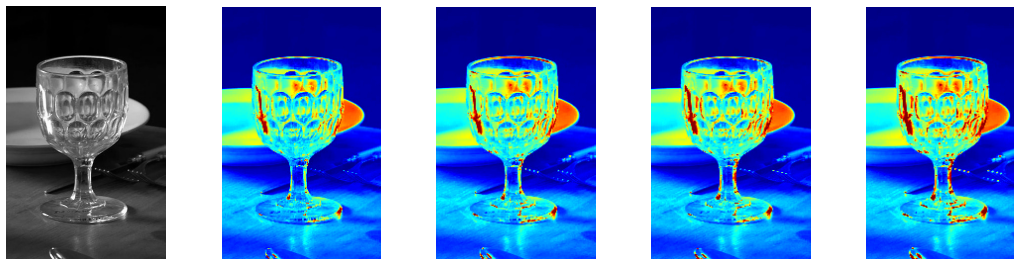


Σχήμα 2.10: Βήματα υπολογισμού του δισδιάστατου μετασχηματισμού κυματιδίων

μήκος κάθε μίας από τις διαστάσεις – για παράδειγμα, σε μια εικόνα, εφαρμόζοντας το μετασχηματισμό πρώτα κατά μήκος των γραμμών και μετά κατά μήκος των στηλών. Μπορούμε έτσι να πάρουμε ως έξοδο τα δείγματα που αντιστοιχούν στη μέγιστη ή την ελάχιστη τιμή μέσα σε γειτονίες 2×2 , όπως φαίνεται και στο Σχήμα 2.10. Διακρίνουμε τη δυνατότητα επιλογής μεταξύ τεσσάρων διαφορετικών καναλιών: από τον πρώτο μονοδιάστατο μετασχηματισμό προκύπτουν τα κανάλια H και L, με τα οποία συμβολίζουμε την επιλογή του μέγιστου ή του ελάχιστου αντίστοιχα από δύο δείγματα. Επαναλαμβάνοντας κατά μήκος της άλλης διάστασης, προκύπτουν τα τέσσερα κανάλια HH, HL, LH και LL. Αντίστοιχα, στις τρεις διαστάσεις προκύπτουν 8 τέτοια κανάλια, τα οποία αντιστοιχούν σε 8 διαφορετικές ζώνες συχνοτήτων κ.ο.κ. . Η χρήση του τρισδιάστατου μετασχηματισμού κυματιδίων ερευνάται στο [2] όπου χρησιμοποιείται στην ανάλυση βίντεο, το οποίο είναι ένα θέμα στο οποίο θα επανέλθουμε στο κεφάλαιο 4. Γενικά, στο παρόν σημείο ανάλυσης εικόνων, σημειώνουμε ότι το κανάλι LL αντιστοιχεί σε αργά μεταβαλλόμενες περιοχές κατά μήκος και των δύο αξόνων, τα HL, LH σε περιοχές γρήγορα μεταβαλλόμενες κατά τον ένα μόνο άξονα και το κανάλι HH σε περιοχές με γρήγορες μεταβολές ως και προς τους δύο άξονες. Στο Σχήμα 2.11 βλέπουμε ένα χαρακτηριστικό παράδειγμα της απόκρισης κάθε καναλιού, το οποίο επιλέγει διαφορετικό εύρος συχνοτήτων, σε μία εικόνα. Οι κόκκινες περιοχές αντιστοιχούν σε ψηλές τιμές απόκρισης ενώ οι μπλε περιοχές σε χαμηλές. Μπορούμε να δούμε ότι το κάθε κανάλι αναδεικνύει διαφορετικά στοιχεία της εικόνας. Σημειώνουμε, τέλος, ότι είναι δυνατός ο ορισμός και μη διαχωρίσιμων εκδοχών του μορφολογικού μετασχηματισμού κυματιδίων Haar.

Κατασκευή της Πυραμίδας

Για το σχηματισμό της μορφολογικής πυραμίδας κυματιδίων Haar χρησιμοποιούμε τις αποκρίσεις του καναλιού LL. Αυτό έρχεται σε αντιστοιχία με την επιλογή ενός Γκαουσιανού φίλτραρίσματος στην πρώτη πυραμίδα: κρατούμε τις βαθυπερατές συνιστώσες της εικόνας προς τις δύο κατευθύνσεις. Στη συνέχεια υποδειγματοληπτούμε κατά δύο – ουσιαστικά, σε κάθε γειτονιά



Σχήμα 2.11: Ζώνες του μετασχηματισμού κυματιδίων για την γκρίζα εικόνα glass. Αριστερά: αρχική εικόνα. Στη συνέχεια, κατά σειρά: ζώνη LL, ζώνη LH, ζώνη HL και ζώνη HH

2×2 εικονοστοιχείων αντιστοιχούμε αυτό με την ελάχιστη τιμή ώστε να την αντιπροσωπεύσει στο επόμενο επίπεδο της πυραμίδας. Συμβολίζουμε τη διαδικασία μετασχηματισμού κυματιδίων και επιλογής του καναλιού LL με h_{sLL} , ενώ με H συμβολίζουμε την προκύπτουσα πυραμίδα και με L^H την αντίστοιχη “Λαπλασιανή” – τα κομμάτια τα οποία απορρίπτονται από την έξοδο του καναλιού LL, και αντιπροσωπεύουν χαρακτηριστικά σε μεγαλύτερες συχνότητες. Η L^H ουσιαστικά συγκρατεί το λεγόμενο “σήμα λεπτομέρειας” κάθε επιπέδου καθώς μεταβαίνουμε στο επόμενο. Ο αλγόριθμος έχει συνεπώς ως εξής:

$$\begin{aligned} H_0 &= I \text{ (αρχική εικόνα)} \\ H_{i+1,0} &= h_{sLL}(H_i); L_i^H = H_i - H_{i+1,0} \\ H_{i+1} &= (H_{i+1,0}) \downarrow_2, i = [1..n - 1] \end{aligned}$$

Στο Σχήμα 2.12 βλέπουμε την αποδόμηση σε πυραμίδα της εικόνας harbor που χρησιμοποιήσαμε και στην επισκόπηση των υπόλοιπων πυραμίδων, μαζί με την L^H .

2.2.5 Leveling Πυραμίδα

Μια σχετικά καινούργια μορφολογική αναπαράσταση αποτελούν τα levelings, τα οποία προτάθηκαν αρχικά από τον Meyer το 1998 [32] ενώ μελετήθηκαν από τους Maragos [33], Matheron [31] και Serra [41]. Σε αυτά βασίζεται μια γενική και πολύ ισχυρή μορφή πολυκλιμακωτής αναπαράστασης η οποία διατηρεί τα περιγράμματα σε κάθε κλίμακα, ενώ εξασφαλίζει την μη εμφάνιση ψευδών μεγίστων (spurious extrema) στα υψηλότερα επίπεδα. Από μια κλίμακα στην άλλη, οι δομές της εικόνας φθίνουν σταδιακά, μετασχηματιζόμενες σε επίπεδες ή οιωνεί-επίπεδες ζώνες — διατηρώντας όμως, εφ’ όσον είναι ορατές, τις ίδιες σχετικές θέσεις όπως και στην αρχική εικόνα.



Σχήμα 2.12: Αριστερά: πυραμίδα κυματιδίων Haar (Haar wavelet), Δεξιά: αντίστοιχη “ Λαπλασιανή ” L^H , μαζί με τα πρώτα τέσσερα επίπεδα από πάνω προς τα κάτω, για την εικόνα *harbor* την αρχική μορφή της οποίας βλέπουμε στο επίπεδο 0 της πυραμίδας κυματιδίων (Haar) πάνω αριστερά. Η L^M πυραμίδα έχει κανονικοποιηθεί μεταξύ των ελάχιστων και μέγιστων τιμών κάθε επιπέδου ώστε να είναι πιο ευδιάκριτη.

Τα levelings αποτελούν μετασχηματισμούς της μορφής $\Lambda(f; m)$ οι οποίοι εξαρτώνται από δύο σήματα, την αναφορά – εικόνα f και το σηματοδευτή m . Αν ο σηματοδευτής είναι ο ίδιος αυτο-δυνατός (self-dual) δηλαδή προέρχεται από ένα τελεστή ο οποίος επιδρά συμμετρικά τόσο στην εικόνα όσο και στο φόντο, τότε και το αντίστοιχο leveling είναι ένα αυτοδυνατό φίλτρο το οποίο αντιμετωπίζει συμμετρικά τόσο τα φωτεινά όσο και τα σκοτεινά μέρη της εικόνας. Ως σηματοδευτής – αρχική συνθήκη (seed) για κάθε επίπεδο χρησιμοποιείται πολύ συχνά το αντίστοιχο του Γκαουσιανού χώρου κλίμακας, τακτική την οποία θα υιοθετήσουμε στη δημιουργία της leveling πυραμίδας. Η ανάλυσή των levelings σε χώρους κλίμακας είναι αλγεβρική και βασίζεται στη θεωρία πλεγμάτων (lattices). Κάποια βασικά στοιχεία από την ανάλυση αυτή, προερχόμενα από το [33] παρατίθενται αμέσως πιο κάτω. Υπάρχουν επίσης και διαφορετικές μαθηματικές αναπαραστάσεις, οι οποίες μοντελοποιούν τα levelings ως λύσεις μη γραμμικών μερικών διαφορικών εξισώσεων, με αρχική συνθήκη το σηματοδευτή m .

Μαθηματικός Ορισμός των levelings

Θεωρούμε γκριζες εικόνες οι οποίες ορίζονται πάνω σε ένα ψηφιακό πλέγμα, αποτελούμενο από εικονοστοιχεία (pixel). Καλούμε $N_G(p)$ το σύνολο των γειτόνων ενός εικονοστοιχείου p . Η μέγιστη (αντ. ελάχιστη) τιμή μιας συνάρτησης f μέσα στο $N_G(p)$ αντιπροσωπεύει τη στοιχειώδη διαστολή δf (αντ. διάβρωση ϵf) της συνάρτησης f στο εικονοστοιχείο p .

Τα απλά levelings επεκτείνουν τις επίπεδες ζώνες της εικόνας και δεν δημιουργούν νέα περιγράμματα. Πιο γενικά levelings επεκτείνουν οιωνεί επίπεδες (quasi-flat) ζώνες, οι οποίες ορίζονται όπως πιο κάτω:

Ορισμός 2.1 . Δύο εικονοστοιχεία x, y ανήκουν στην ίδια R -επίπεδη ζώνη μιας συνάρτησης f αν, και μόνο αν υπάρχει μια n -άδα εικονοστοιχείων (p_1, p_2, \dots, p_n) τέτοια ώστε $p_1 = x$ και $p_n = y$, και για κάθε i , τα (p_i, p_{i+1}) είναι γείτονες και ικανοποιούν τη συμμετρική σχέση $f_{p_i} R f_{p_{i+1}}$.

Η απλούστερη συμμετρική σχέση R είναι η ισότητα, $f_{p_i} = f_{p_{i+1}}$, στην οποία οι οιωνεί-επίπεδες ζώνες είναι επίπεδες. Μία πιο περίπλοκη σχέση R , για δύο γειτονικά εικονοστοιχεία p και q θα μπορούσε να είναι η $f_p \approx f_q$ η οποία ορίζεται ως $|f_p - f_q| \leq \lambda$, η οποία είναι συμμετρική και ορίζει οιωνεί-επίπεδες ζώνες με μια μέγιστη κλίση ίση με λ .

Στη συνέχεια το [33] προχωρεί στον ορισμό του leveling. Μία εικόνα g θα καλείται leveling μιας άλλης αρχικής εικόνας f εάν ικανοποιεί τον παρακάτω ορισμό:

Ορισμός 2.2 . Μια εικόνα g αποτελεί leveling μιας άλλης αρχικής εικόνας f αν, και μόνο αν, $\forall(x, y)$ γειτονικά εικονοστοιχεία ισχύει

$$g_p > g_q \Rightarrow \begin{cases} f_p \geq g_p \\ g_q \geq f_q \end{cases}$$

Συμβολίζουμε τη σχέση “ η εικόνα g είναι ένα leveling της f ” με $g < f$.

Παρατηρούμε ότι, εάν η συνάρτηση g είναι σταθερή, δεν υπάρχει κανένα ζεύγος γειτονικών εικονοστοιχείων (x, y) για το οποίο να ισχύει $g_p > g_q$. Συνεπώς η συνθήκη $\{g_p > g_q \Rightarrow f_p \geq g_p \ \& \ g_q \geq f_q\}$ είναι πάντοτε αληθής, δείχνοντας ότι μια επίπεδη, σταθερή συνάρτηση αποτελεί leveling κάθε άλλης συνάρτησης.

Ιδιότητες των levelings

Μία σημαντική αλγεβρική ιδιότητα των levelings είναι η εξής: Αν δύο συναρτήσεις g_1 και g_2 είναι και οι δύο levelings της ίδιας συνάρτησης f , τότε οι $g_1 \wedge g_2$ και $g_1 \vee g_2$ αποτελούν επίσης levelings της f . Αυτή η ιδιότητα μας επιτρέπει να αντιστοιχίσουμε νέα levelings με μια οικογένεια από levelings. Στη συγκεκριμένη περίπτωση, εάν η g_i είναι μια οικογένεια από levelings της f , το μορφολογικό κέντρο $(f \vee \wedge g_i) \wedge \vee g_i$ της οικογένειας αυτής είναι επίσης ένα leveling της f .

Επίσης, τα levelings ικανοποιούν και τις τρεις βασικές απαιτήσεις που περιγράψαμε στην εισαγωγή της ενότητας 2.2 – ανεξαρτησία από τη μετατόπιση, την περιστροφή και την μεταβολή του φωτισμού. Οι μαθηματικές απόδειξεις για αυτό παρέχονται λεπτομερώς στο [33]. Προκύπτουν επίσης τα εξής ενδιαφέροντα λήμματα:

Λήμμα . Η σχέση $\{g < f, g \text{ είναι ένα leveling της } f\}$ είναι συμμετρική και μεταβατική, δηλαδή μια σχέση προ-ιεραρχίας(preorder).

Λήμμα . Η οικογένεια των levelings, από την οποία εξαιρούμε τις τετριμμένες (trivial) σταθερές συναρτήσεις, επιβεβαιώνουν την αντι-συμμετρική σχέση: αν f είναι μια μη σταθερή συνάρτηση και leveling της g , και ταυτόχρονα η g είναι ένα leveling της f , τότε $f = g$.

Όντας μια αντι-συμμετρική προ-ιεραρχική σχέση, η $g < f$ είναι μια ιεραρχική σχέση, εκτός από παντού σταθερές συναρτήσεις. Το χαρακτηριστικό αυτό μας επιτρέπει την κατασκευή πολυκλιμακωτών αναπαραστάσεων μιας εικόνας σαν μια σειρά από levelings $(g_0 = f, g_1, g_2, \dots, g_n)$, όπου g_k είναι ένα leveling του g_{k+1} και, ως συνέπεια της μεταβατικής ιδιότητας, η g_k είναι επίσης ένα leveling κάθε συνάρτησης g_j για $j < k$.

Κατασκευή των levelings

Θεωρούμε τα levelings ως την τομή δύο μεγαλύτερων κλάσεων, των κατώτερων και ανώτερων levelings τα οποία ορίζονται όπως πιο κάτω:

Ορισμός 2.3 . Μια συνάρτηση g αποτελεί κατώτερο leveling μιας άλλης συνάρτησης f αν, και μόνο αν, για κάθε ζεύγος (p, q) γειτονικά εικονοστοιχεία ισχύει $g_p > g_q \Rightarrow q_q \geq f_q$.

Ορισμός 2.4 . Μια συνάρτηση g αποτελεί ανώτερο leveling μιας άλλης συνάρτησης f αν, και μόνο αν, για κάθε ζεύγος (p, q) γειτονικά εικονοστοιχεία ισχύει $g_p > g_q \Rightarrow q_p \leq f_p$.

Ο χαρακτηρισμός “ ανώτερο leveling ” προέρχεται από το γεγονός ότι όλες οι συνδεδεμένες συνιστώσες όπου $g > f$ είναι επίπεδες: για κάθε ζεύγος γειτονικών εικονοστοιχείων (p, q) ισχύει

$$\left. \begin{array}{l} g_q > f_q \\ g_p > f_p \end{array} \right\} \Rightarrow g_p = g_q.$$

Όμοια αν η g αποτελεί κατώτερο leveling της f όλες οι συνδεδεμένες συνιστώσες όπου $g < f$ είναι επίπεδες.

Μία συνάρτηση g αποτελεί leveling μιας συνάρτησης f αν, και μόνο αν αποτελεί ταυτόχρονα ανώτερο και κατώτερο leveling της f . Προσπαθώντας να βρούμε μια ισοδύναμη διατύπωση για τα κατώτερα levelings παρατηρούμε ότι τα εικονοστοιχεία που ικανοποιούν την απαίτηση του ορισμού 2.3 είναι και αυτά για τα οποία η διαστολή, δ , θα τους αυξήσει την τιμή, δηλαδή $g_q < \delta_q g$. Έτσι οδηγούμαστε σε ένα νέο κριτήριο:

Κριτήριο . Μια συνάρτηση g αποτελεί κατώτερο leveling μιας συνάρτησης f αν, και μόνο αν, ισχύει $g_q < \delta_q g \Rightarrow g_q \geq f_q$.

Οι σχέσεις της μορφής $[A \Rightarrow B]$ μπορούν να παρασταθούν σε διαζευκτική μορφή ως $[\neg A \vee B]$ οπότε μπορούμε να παραστήσουμε την $g_q < \delta_q g \Rightarrow g_q \geq f_q$ ως $g_q \geq \delta_q g \vee g_q \geq f_q$ ή, ισοδύναμα $g_q \geq f_q \wedge \delta_q g$. Όμοίως μπορούμε να εργαστούμε και για τα ανώτερα levelings, οπότε προκύπτουν τα εξής δύο κριτήρια:

Κάτω Κριτήριο . Μια συνάρτηση g αποτελεί κατώτερο leveling μιας συνάρτησης f αν, και μόνο αν, ισχύει $g \geq f \wedge \delta g$.

Άνω Κριτήριο . Μια συνάρτηση g αποτελεί ανώτερο leveling μιας συνάρτησης f αν, και μόνο αν, ισχύει $g \leq f \vee \epsilon g$.

Συνενώνουμε τα πιο πάνω σε ένα ολικό κριτήριο για τα levelings:

Ολικό Κριτήριο . Μια συνάρτηση g αποτελεί leveling μιας συνάρτησης f αν, και μόνο αν, ισχύει $f \wedge \delta g \leq g \leq f \vee \epsilon g$.

Γενικεύουμε το πιο πάνω ως εξής:

Γενικευμένο Κριτήριο . Μια συνάρτηση g αποτελεί leveling μιας συνάρτησης f αν, και μόνο αν, ισχύει

$$f \wedge \alpha g \leq g \leq f \vee \beta g. \quad (2.1)$$

όπου ο τελεστής α είναι επεκτατικός ($\alpha g \geq g$) και ο τελεστής β μη επεκτατικός ($\beta g \leq g$)

Χρησιμοποιώντας τη σχέση 2.1 μπορούμε να μετατρέψουμε κάθε συνάρτηση, λεγόμενη *σημαδευτή*, g σε leveling μιας συνάρτησης *αναφοράς* f . Αν η g δεν είναι ήδη ένα leveling της f , τότε το κριτήριο $f \wedge \alpha g \leq g \leq f \vee \beta g$ δεν ισχύει για ένα τουλάχιστον εικονοστοιχείο p . Διακρίνουμε έτσι δύο περιπτώσεις:

- $g_p < f_p \wedge \alpha_p g$. Συνεπώς η μικρότερη μεταβολή του g_p για την οποία το κριτήριο ισχύει είναι $g'_p = f_p \wedge \alpha_p g$. Παρατηρούμε ότι $g_p \leq g'_p \leq f_p$.
- $g_p > f_p \vee \beta_p g$. Συνεπώς η μικρότερη μεταβολή του g_p για την οποία το κριτήριο ισχύει είναι $g'_p = f_p \vee \beta_p g$. Παρατηρούμε ότι $g_p \geq g'_p \geq f_p$.

Παρατηρούμε ότι για $\{g_p = f_p\}$ το κριτήριο ικανοποιείται πάντοτε, οπότε προκύπτει μια άλλη διατύπωση του αλγορίθμου:

- lev^- : Για $\{g < f\}$ εφάρμοσε $g = f \wedge \alpha g$.
- lev^+ : Για $\{g > f\}$ εφάρμοσε $g = f \vee \beta g$.

Εύκολα βλέπουμε ότι ο αλγόριθμος αυτός ισοδυναμεί με την αντικατάσταση της τιμής του g παντού από τη νέα τιμή $g = (f \wedge \alpha g) \vee \beta g = (f \vee \beta g) \wedge \alpha g$. Επαναλαμβάνουμε τον αλγόριθμο ωσότου η σχέση 2.1 ισχύει παντού. Είμαστε βέβαιοι για τη σύγκλιση του αλγορίθμου καθ' ότι οι μεταβολές της g που περιγράφονται είναι, για κάθε σημείο μονότονες: οι διαδοχικές τιμές της g έρχονται όλο και πιο κοντά στην f μέχρι τη σύγκλιση. Στο [33] αναφέρονται επίσης βήματα που επιταχύνουν την εκτέλεση του αλγορίθμου καθώς και απόδειξη του γεγονότος ότι τα levelings αποτελούν ισχυρά μορφολογικά φίλτρα.

Πολυκλιμακωτά levelings

Έχοντας υπόψιν την πολυκλιμακωτή προοπτική της εφαρμογής των πιο πάνω, βλέπουμε ότι τα levelings μας επιτρέπουν την κατασκευή διάφορων πολυκλιμακωτών οικογενειών. Κάποιες από αυτές περιγράφονται στο [33], τις οποίες αναφέρουμε:

- Levelings σχετιζόμενα με μονότονες οικογένειες διαστολών (dilations) και διαβρώσεων (erosions).
- Levelings σχετιζόμενα με μια αυθαίρετη οικογένεια συναρτήσεων - σημαδευτών.
- Levelings σχετιζόμενα με μια εναλλασόμενη οικογένεια συναρτήσεων - σημαδευτών.

Η μέθοδος που θα χρησιμοποιήσουμε στην υλοποίηση της leveling πυραμίδας είναι η δεύτερη. Σε αυτήν θεωρούμε ένα leveling Λ μαζί με μια γενικευμένη διαστολή α και την αντίστοιχη συμπληρωματική της (adjunct) διάβρωση β . Χρησιμοποιούμε μια αυθαίρετη οικογένεια συναρτήσεων - σημαδευτών $g_i, i = 1, 2, 3, \dots$, και θεωρούμε τα αντίστοιχα levelings: με $\Lambda_{g_i} f$ συμβολίζουμε το leveling της f που σχετίζεται με το σημαδευτή g_i . Έτσι μπορούμε να κατασκευάσουμε μια οικογένεια levelings που σχετίζεται με την οικογένεια των σημαδευτών βάσει του ακόλουθου μηχανισμού:

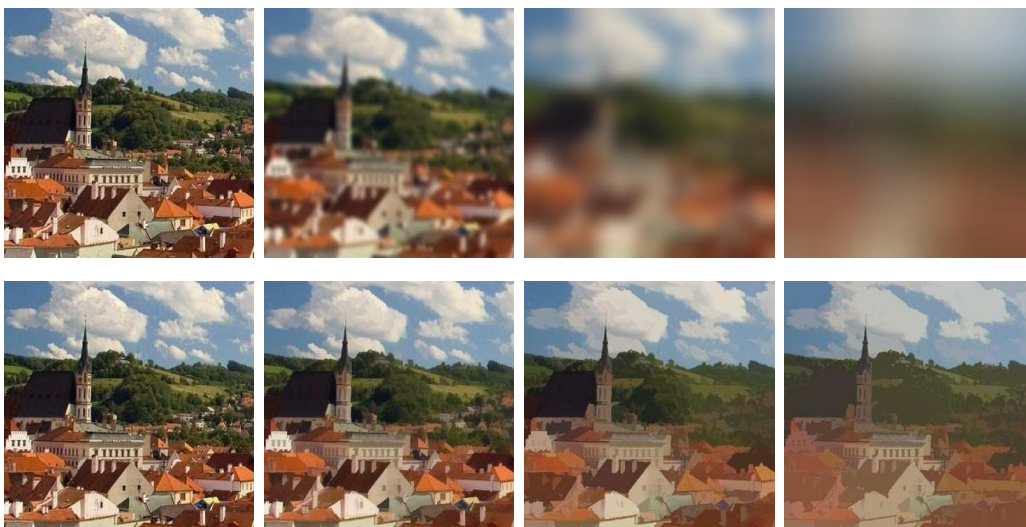
$$l_1 = \Lambda_{g_1} f, l_2 = \Lambda_{g_2} \Lambda_{g_1} f, \dots, l_n = \Lambda_{g_n} \Lambda_{g_{n-1}} \dots \Lambda_{g_2} \Lambda_{g_1} f. \quad (2.2)$$

Μπορούμε να εκφράσουμε την 2.2 ως τον αλγόριθμο που ακολουθεί. Έστω η αρχική εικόνα f και οι σημαδευτές g_i που σχετίζονται με μια αυξανόμενη παράμετρο κλίμακας i . Τα levelings l_i υπολογίζονται ως εξής:

$$\begin{aligned} l_1 &= \Lambda_{g_1} f \\ l_2 &= \Lambda_{g_2} l_1 \\ l_3 &= \Lambda_{g_3} l_2 \\ &\vdots \\ l_n &= \Lambda_{g_n} l_{n-1} \end{aligned}$$

Βάσει της πιο πάνω ακολουθίας διασφαλίζουμε ότι l_i είναι leveling του l_j για κάθε $j > i$.

Ως σημαδευτές μπορούμε, όπως είπαμε, να χρησιμοποιήσουμε οποιεσδήποτε αυθαίρετες συναρτήσεις. Συνήθως χρησιμοποιούνται ως σημαδευτές συνελιξίες της αρχικής εικόνας f με Γκαουσιανούς πυρήνες. Ως παράμετρος κλίμακας χρησιμοποιείται η τυπική απόκλιση σ του κάθε πυρήνα, με μεγαλύτερο σ να αντιστοιχεί σε πιο απλοποιημένη αναπαράσταση και συνεπώς υψηλότερη κλίμακα. Σε αυτή την περίπτωση, τα levelings l_i που προκύπτουν διατηρούν πολύ



Σχήμα 2.13: Σημαδευτές και αντίστοιχα levelings της εικόνας *krumlou*. Αριστερά, πάνω και κάτω: Αρχική εικόνα. Πάνω σειρά: σημαδευτές – γκαουσιανά φιλτραρίσματα της εικόνας, τυπικής απόκλισης $\sigma = 4, 16$ και 64 αντίστοιχα. Κάτω: levelings που προκύπτουν από τον κάθε σημαδευτή. Η υλοποίηση έγχρωμων σημαδευτών και levelings έγινε ξεχωριστά σε κάθε χρωματικό κανάλι.

καλύτερα τα περιγράμματα από ότι οι αντίστοιχοι γκαουσιανοί σημαδευτές τους – αρχικές συνθήκες κάθε βήματος του αλγορίθμου – g_i . Απορρίπτονται δηλαδή τα αντικείμενα που δεν ανήκουν στην κλίμακα χωρίς να αλλάζει το περίγραμμα αυτών που παραμένουν. Στην εικόνα 2.13 βλέπουμε ένα παράδειγμα εφαρμογής των levelings σε μια εικόνα με τρεις Γκαουσιανούς σημαδευτές διαφορετικών τυπικών αποκλίσεων.

Υλοποίηση της leveling πυραμίδας

Βάσει του πιο πάνω πολυκλιμακωτού αλγορίθμου υλοποιούμε την leveling πυραμίδα, προσθέτοντας υποδειγματοληψία κατά παράγοντα δύο στη μετάβαση από τη μια κλίμακα στην άλλη. Σαν σημαδευτές χρησιμοποιούμε τα αντίστοιχα επίπεδα της Γκαουσιανής πυραμίδας – εδώ, η αυξανόμενη παράμετρος κλίμακας δεν είναι μια ακολουθία όλο και μεγαλύτερων τιμών της τυπικής απόκλισης του Γκαουσιανού πυρήνα που φιλτράρει την αρχική εικόνα, αλλά η διαδοχική εφαρμογή του ίδιου φίλτρου καθώς πάμε από το ένα επίπεδο στο άλλο.

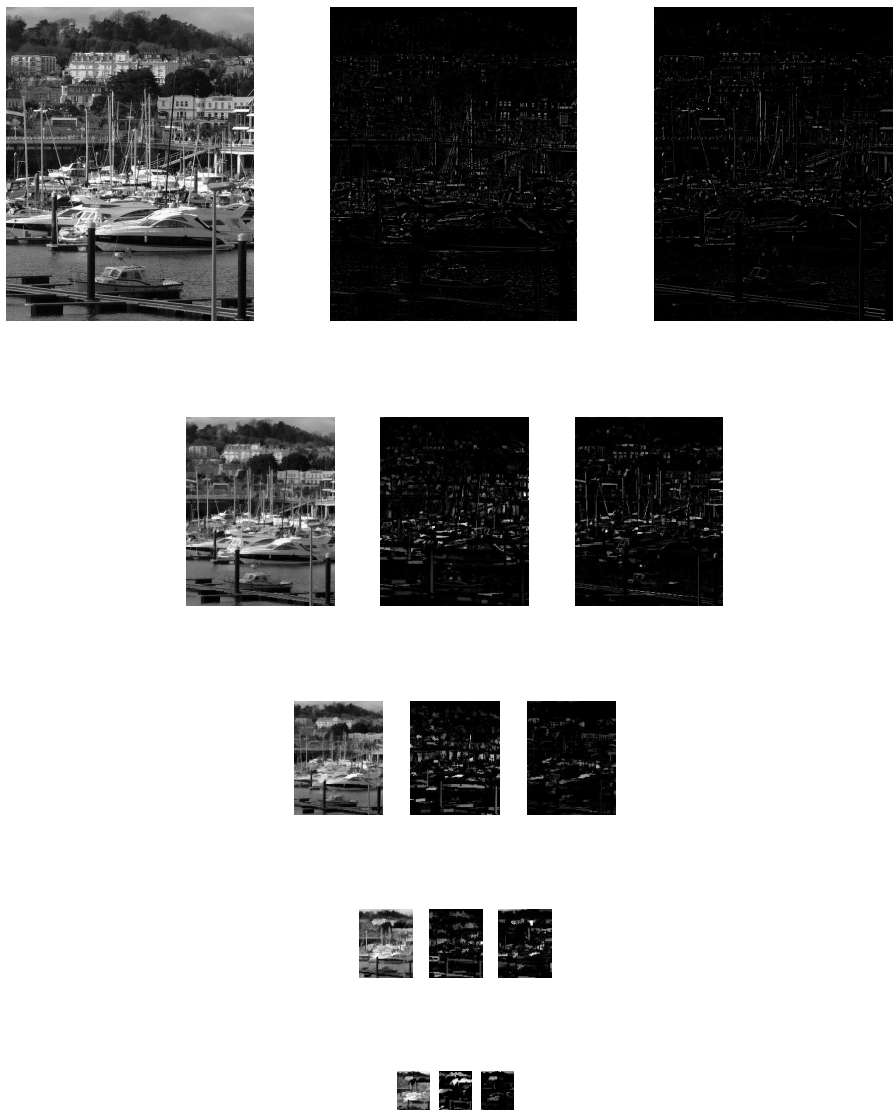
Στη leveling πυραμίδα θα χρησιμοποιήσουμε, σε ρόλο “ Λαπλασιανής ”, δύο διαφορετικές πολυκλιμακωτές αναπαραστάσεις. Κατά το leveling μιας εικόνας, κάποιες περιοχές, τις οποίες καλούμε κοιλάδες, γεμίζουν και ανυψώνονται, ενώ

άλλες οι οποίες εξέχουν, τις οποίες καλούμε κορυφές, υφίστανται καθίζηση ώστε να έχουμε τη δημιουργία των οιωνεί - επίπεδων περιοχών. Τις κοιλάδες και κορυφές που απορρίπτονται σε κάθε βήμα δημιουργίας της πυραμίδας τις κρατάμε σε δύο ξεχωριστές πυραμίδες, την $L^{Q,P}$ για τις κορυφές (peaks) και την $L^{Q,V}$ για τις κοιλάδες (valleys). Την ίδια τη leveling πυραμίδα τη συμβολίζουμε με Q . Ο αλγόριθμος κατασκευής των τριών αυτών πυραμίδων είναι ο εξής (K ο Γκαουσιανός πυρήνας):

$$\begin{aligned}
 Q_0 &= I, \quad g_0 = I \text{ (αρχική εικόνα)} \\
 g_{i,0} &= g_{i-1} * K, \text{ (σημαδευτής κάθε επιπέδου)} \\
 g_i &= (g_{i,0}) \downarrow_2, \quad i = [1..n - 1] \text{ (δημιουργία Γκαουσιανής πυραμίδας)} \\
 Q_{i+1,0} &= \Lambda(Q_i | g_{i,0}) \text{ (leveling)} \\
 L_i^{Q,P} &= \begin{cases} Q_i - Q_{i+1,0}, & Q_i - Q_{i+1,0} \geq 0 \\ 0 & Q_i - Q_{i+1,0} < 0 \end{cases} \\
 L_i^{Q,V} &= \begin{cases} Q_{i+1,0} - Q_i, & Q_{i+1,0} - Q_i \geq 0 \\ 0 & Q_{i+1,0} - Q_i < 0 \end{cases} \\
 Q_{i+1} &= (\text{υποδ})Q_{i+1,0}, \quad i = [1..n - 1] \text{ (δημιουργία Leveling πυραμίδας)}
 \end{aligned}$$

Θα μπορούσαμε να κρατήσουμε τις $L^{Q,P}$ και $L^{Q,V}$ σαν μια ενιαία πυραμίδα με θετικές και αρνητικές τιμές, όμως θέλοντας να διατηρήσουμε την υλοποίηση των επιπέδων με τιμές 0-255 (UTF - 8) όπως την αρχική εικόνα επιλέγουμε την ξεχωριστή αποθήκευσή τους. Αφαίρεση μεταξύ αριθμών σε αυτή την αναπαράσταση ισοδυναμεί με αφαίρεση, εαν το αποτέλεσμα της δίνει αριθμό μεγαλύτερο του μηδέν, ή μηδέν εαν το αποτέλεσμα είναι μικρότερο, όπως σημειώσαμε και στον αλγόριθμο.

Στην εικόνα 2.14 βλέπουμε τα πρώτα τέσσερα επίπεδα για τις πυραμίδες Q , $L^{Q,P}$ και $L^{Q,V}$. Η βέλτιστη διατήρηση των περιγραμμάτων είναι εμφανής - αυτό μπορεί να φανεί ακόμη περισσότερο όσο ανεβαίνουμε στα επίπεδα. Οι $L^{Q,P}$ και $L^{Q,V}$, οι οποίες κρατούν την πληροφορία που "ισοπεδώνεται" κατά τις μεταβάσεις σε ανώτερα επίπεδα, θα χρησιμοποιηθούν στη συνέχεια - όπως και οι υπόλοιπες πυραμίδες L που θεωρήσαμε στο κεφάλαιο αυτό - για ανίχνευση προσανατολισμών μέσα στις εικόνες.



Σχήμα 2.14: Πυραμίδα leveling Q (αριστερά) και πυραμίδες κοιλάδων $L^{Q,V}$ (κέντρο) και κορυφών $L^{Q,P}$ (δεξιά) με τα πρώτα τέσσερα επίπεδα από πάνω προς τα κάτω, για την εικόνα *harbor* την αρχική μορφή της οποίας βλέπουμε στο επίπεδο 0 της (leveling) πυραμίδας πάνω αριστερά. Οι $L^{Q,V}$ και $L^{Q,P}$ πυραμίδες έχει κανονικοποιηθεί μεταξύ των ελάχιστων και μέγιστων τιμών κάθε επιπέδου ώστε να είναι πιο ευδιάκριτες. Βλέπουμε την σχεδόν άριστη διατήρηση των περιγραμμάτων υψηλές κλίμακες, χωρίς τη δημιουργία ανεπιθύμητων ψευδών μεγίστων.

Κεφάλαιο 3

Υπολογιστικά Μοντέλα Οπτικής Προσοχής

3.1 Γενικά

Τα υπολογιστικά μοντέλα για την οπτική προσοχή (saliency) προσπαθούν να προβλέψουν τις πιο αξιοπρόσεκτες (salient) περιοχές μιας εικόνας και συνήθως αποτυπώνουν την πρόβλεψή τους αυτή σε έναν αντίστοιχο χάρτη στον οποίο η ένταση σε κάθε σημείο είναι ανάλογη της σημαντικότητας και του βαθμού στον οποίο αυτό είναι αξιοπρόσεκτο. Έμπνευση για τη δομή και τη λειτουργία των μοντέλων αυτών αποτελούν κυρίως σχετικές ανακαλύψεις από τους τομείς της βιολογίας και της νευροφυσιολογίας, μια γενική παρουσίαση των οποίων κάναμε στο κεφάλαιο 2. Πολλές εισαγωγικές πληροφορίες για τα υπολογιστικά μοντέλα υπάρχουν στα [2] και [30].

Υπάρχουν δύο κύριες μεθοδολογίες εξαγωγής του ενδιαφέροντος. Η πρώτη αφορά τοπικά χαρακτηριστικά της περιοχής της εικόνας, όπως αντίθεση (contrast), χρωματικό ανταγωνισμό, προσανατολισμούς κλπ. Τα μοντέλα που προκύπτουν με αυτή τη φιλοσοφία είναι κυρίως ανοδικά (bottom - up). Ένα κλασικό παράδειγμα αποτελεί η αρχιτεκτονική των Koch και Ullman (1985) [20]. Βασισμένη στη θεωρία συμψηφισμού χαρακτηριστικών (feature integration theory), ενσωματώνει όλα τα χαρακτηριστικά που περιγράψαμε σε ένα κύριο χάρτη οπτικής προσοχής (saliency map). Πάνω στο στη δουλειά των Koch και Ullman βασίζεται και το πιο γνωστό μοντέλο οπτικής προσοχής, αυτό των Itti et al., το οποίο περιγράφουμε εκτενώς στην επόμενη ενότητα. Επεκτάσεις των μοντέλων αυτών αφορούν την οριζόντια αλληλεπίδραση μεταξύ των διαδοχικών σταδίων εξαγωγής και επεξεργασίας των χαρακτηριστικών, όπως αυτό των Hamker et al. [11] ή συνεχή αλληλεπίδραση του κύριου χάρτη με τα κατώτερα στάδια επεξεργασίας, όπως στο μοντέλο των Tsotsos et al. [45]. Αυτό αποτελείται

από ένα μηχανισμό εμπροσθόδοτης (feed-forward) ανοδικής εύρεσης χαρακτηριστικών μαζί με ένα μηχανισμό επιλεκτικού συντονισμού (selective tuning) τους με ανάδραση.

Μια δεύτερη μεθοδολογία αφορά την εύρεση ενδιαφερόντων σημείων βάσει των ολικών (global) ιδιοτήτων τους. Μια ολική ιδιότητα είναι η σπανιότητα κάθε συστατικού των αντικειμένων μέσα στην εικόνα. Η ιδέα αυτή έχει προταθεί από τον Mudge το 1987 και εκφράζει ότι το αξιοπρόσεκτο κάθε συστατικού είναι αντιστρόφως ανάλογο με τον αριθμό των φορών που επαναλαμβάνεται μέσα στην εικόνα [34]. Άλλα μοντέλα πρότειναν κατάτμηση της εικόνας και επίδοση ενός μέτρου σημαντικότητας σε κάθε περιοχή – τέτοιοι μηχανισμοί όμως ήταν ιδιαίτερα ευαίσθητοι σε λάθη της κατάτμησης. Αντιμετωπίσεις ανάλογες με αυτήν του Mudge επεκτείνουν την ιδέα προσεγγίζοντας την προσοχή με βάση το αντίστροφο της πιθανότητας ύπαρξης - εμφάνισης ενός εικονοστοιχείου. Τέτοιες ιδέες τελικά εκφράστηκαν σε πιθανοτικά μοντέλα ή στα πλαίσια της θεωρίας πληροφορίας. Ο ίδιος ο Itti ενσωμάτωσε μια τέτοια πιθανοτική προσέγγιση στο αρχικό μοντέλο, δίνοντας γενικά καλύτερα αποτελέσματα [15].

Πρόσφατες προσεγγίσεις αποσκοπούν στην περιγραφή μηχανισμών που περιέχουν και τις δύο αυτές μεθοδολογίες, ενώ μοντελοποιούν στοιχεία τόσο πρώιμης (preattentive) όσο και προσεκτικής (attentive) οπτικής προσοχής. Τα πρώτα στοιχεία συνιστούν το μεγαλύτερο μέρος των μοντέλων που έχουν προταθεί μέχρι σήμερα, κάνοντας την υλοποίηση και προσθήκη προσεκτικών, καθοδικών (top-down) μεθόδων μια λογική εξέλιξη. Κάποια από αυτά τα στοιχεία υπάρχουν στα μοντέλα των Hamker et al. και Tsotsos et al. που αναφέραμε, ενώ ένα άλλο μοντέλο προτάθηκε πρόσφατα από τον Mancas. Εμπεριέχει τρία επίπεδα: ένα χαμηλό, το οποίο αφορά μόνο απλά ανοδικά χαρακτηριστικά, ένα μεσαίο επίπεδο το οποίο χρησιμοποιεί τα χαρακτηριστικά του χαμηλού επιπέδου για να αποφανθεί για το επίκεντρο της προσοχής αν ο χάρτης δε δίνει ξεκάθαρα αποτελέσματα και τέλος ένα υψηλό επίπεδο όπου κομμάτια της εικόνας που επελέγησαν από μια ακολουθία εστιάσεων της προσοχής συγκρίνονται μεταξύ τους και με τη μνήμη. Τα κύρια χαρακτηριστικά του μοντέλου αυτού των τριών επιπέδων (3L) φαίνονται στον Πίνακα 3.1.

Σαν βάση διερεύνησης – κυρίως πρώιμων – χαρακτηριστικών της οπτικής προσοχής θα χρησιμοποιήσουμε το προαναφερθέν μοντέλο των Itti et al., το οποίο παραμένει σημείο αναφοράς στην υπολογιστική υλοποίηση της προσοχής. Μετά την εκτενή περιγραφή του στην ενότητα 3.2, θα περιγράψουμε αρχικά μια μεθοδολογία αξιολόγησης των αποκρίσεων του μοντέλου στη βάση του ανθρώπινου παράγοντα (ενότητα 3.3). Στη συνέχεια θα εξετάσουμε και θα συγκρίνουμε διάφορες πιθανές προσθήκες ή παραλλαγές της αρχικής υλοποίησης. Στην ενότητα 3.4 εξετάζουμε την υλοποίηση του αλγορίθμου με τις μη γραμμικές πυραμίδες που περιγράψαμε στην ενότητα 2.2. Στην ενότητα 3.5 διερευνούμε την προσθήκη της απόκρισης ενός ανιχνευτή γωνιών ως επιπλέον

Πρώιμη Οπτική Προσοχή ▷ ▷ Προσεκτική Οπτική Προσοχή		
▷ Χαμηλό Επίπεδο ▷	▷ Μεσαίο Επίπεδο ▷	▷ Υψηλό Επίπεδο ▷
<p>Στόχος:</p> <ul style="list-style-type: none"> • Εύρεση πρώιμων, ανοδικών περιοχών ενδιαφέροντος <p>Χαρακτηριστικά:</p> <ul style="list-style-type: none"> • Αντανακλαστικά • Παράλληλος υπολογισμός • Πρώιμα οπτικά στοιχεία <p>Μεταβλητές της αρχικής εικόνας εισόδου:</p> <ul style="list-style-type: none"> • Φωτεινότητα • Χρώμα 	<p>Στόχος:</p> <ul style="list-style-type: none"> • Διαχωρισμός των χαμηλού επιπέδου παρόμοιων περιοχών οπτικής προσοχής • Δημιουργία της φθίνουσας ιεραρχίας των αξιοπρόσεκτων περιοχών (βραχυπρόθεσμη μνήμη) • Ετοιμασία του υψηλού επιπέδου ελέγχου των αξιοπρόσεκτων περιοχών <p>Χαρακτηριστικά:</p> <ul style="list-style-type: none"> • Αντανακλαστικά / Συνειδητό • Παράλληλος / Σειριακός υπολογισμός • Στοιχεία συμπεριφοράς <p>Μεταβλητές των αξιοπρόσεκτων περιοχών χαμηλού επιπέδου:</p> <ul style="list-style-type: none"> • Μέγεθος • Προσανατολισμός 	<p>Στόχος:</p> <ul style="list-style-type: none"> • Σύγκριση της περιοχής εστίασης με άλλες περιοχές μέσα στο ίδιο σήμα (βραχυπρόθεσμη μνήμη) ή την μακροπρόθεσμη μνήμη • Έλεγχος και επανακατεύθυνση της καθοδικής πληροφορίας στο χαμηλό και μεσαίο επίπεδο <p>Χαρακτηριστικά:</p> <ul style="list-style-type: none"> • Συνειδητό • Σειριακός υπολογισμός • Περίπλοκα στοιχεία <p>Μεταβλητές των εστιάσεων της εικόνας:</p> <ul style="list-style-type: none"> • Μέσα στις εστιάσεις, στην ίδια εικόνα: ομοιογένεια, συμμετρία • Μεταξύ εστιάσεων της ίδιας εικόνας (βραχυπρόθεσμη μνήμη): κανονικότητα, συμμετρία • Μεταξύ δομών διαφορετικών εικόνων (μακροπρόθεσμη μνήμη): αναγνώριση αντικειμένων

Πίνακας 3.1: Αρχιτεκτονική του μοντέλου τριών επιπέδων (3L) και αντίστοιχες μεταβλητές κάθε επιπέδου. Από το [30].

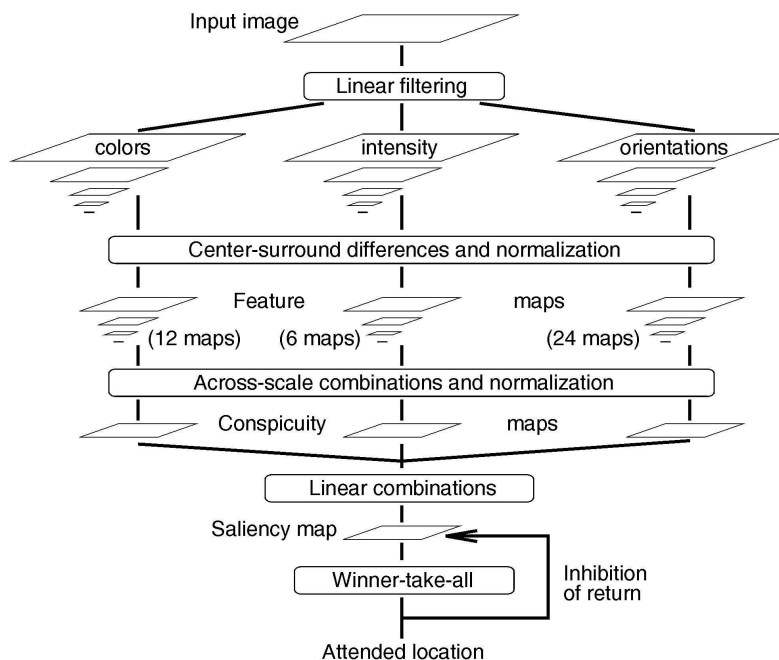
ενδιάμεσου χάρτη οπτικής προσοχής. Τέλος, στην ενότητα 3.6 αναφέρουμε κάποιες ιδέες για βελτίωση του τελεστή $N(\cdot)$.

3.2 Το μοντέλο των Itti, Koch και Niebur

Το μοντέλο αυτό προτάθηκε το 1998 στο [16] και αποτελεί ένα από τα πλέον πετυχημένα του είδους. Βασίζεται στη μελέτη αντίστοιχων βιολογικών συστημάτων, τα οποία υλοποιούν την επιλογή της κύριας, πιο αξιοπρόσεκτης (salient) περιοχής του οπτικού πεδίου, του λεγόμενου "έπικέντρου της προσοχής" (ΕτΠ - focus of attention, FOA). Αυτή η επιλογή γίνεται από ενός επεξεργάζοντας τη σκηνή τόσο με ένα πολύ γρήγορο, από κάτω προς τα πάνω (bottom-up), ανεξάρτητο συγκεκριμένων στόχων και βασισμένο στην οπτική προσοχή μηχανισμό και από έτερου με ένα πιο αργό, από πάνω προς τα κάτω και στη βάση συγκεκριμένων στόχων (top-down) μηχανισμό [35]. Αυτοί οι μηχανισμοί μπορούμε να πούμε ότι αντιστοιχούν χονδρικά στα γενικά μοντέλα πρώιμης προ-επισκόπησης και προσεκτικής επισκόπησης αντίστοιχα που παρουσιάσαμε στην ενότητα 2.1.3.

Το προτεινόμενο μοντέλο ουσιαστικά υλοποιεί το μη συνειδητό, από κάτω προς τα πάνω κομμάτι της προεπισκόπησης επεκτείνοντας μια άλλη βιολογικά εμπνευσμένη αρχιτεκτονική, αυτή των Koch και Ullman. Σχετίζεται με τη λεγόμενη "θεωρία συμψηφισμού χαρακτηριστικών" (feature integration theory) και αρχικά αποδομεί το οπτικό ερέθισμα σε έναν αριθμό τοπογραφικών χαρτών χαρακτηριστικών. Στη συνέχεια διαφορετικά χωρικά σημεία ανταγωνίζονται μεταξύ τους για προσοχή μέσα σε κάθε χάρτη, έτσι ώστε μόνο σημεία τα οποία εξέχουν ιδιαίτερα από το περιβάλλον τους διατηρούνται. Τελικά, αυτοί οι χάρτες σχηματίζουν ένα κύριο χάρτη οπτικής προσοχής, 'saliency map' ο οποίος αποτελεί το μέτρο της σημασίας κάθε σημείου. Το μοντέλο στη συνέχεια επιλέγει δυναμικά την ιεραρχία των σημαίνοντων περιοχών μέσα στην εικόνα. Ο αλγόριθμος έχει το πλεονέκτημα ότι είναι ιδανικός για παράλληλη υλοποίηση η οποία μπορεί να δίνει αστραπιαία αποτελέσματα (στα πλαίσια της διπλωματικής εργασίας, όπως και στο λογισμικό που διατίθεται από τους εμπνευστές της μεθόδου αυτής, η υλοποίηση γίνεται βέβαια σειριακά, καθ' ότι απευθύνεται σε συστήματα ενός επεξεργαστή). Αυτό μπορεί να γίνει διακριτό βλέποντας και τη σχηματική απεικόνιση της διαδικασίας στο Σχήμα 3.1.

Το μοντέλο δημιουργεί μια γκαουσιανή πυραμίδα εννέα επιπέδων, με το πρώτο επίπεδο να αποτελείται από την αρχική εικόνα, και κάθε ανώτερο επίπεδο να προκύπτει με την επεξεργασία του κατώτερου του με ένα βαθυπερατό, γκαουσιανό φίλτρο και την υποδειγματοληψία του σε κλίμακα 1:2. Αυτό έχει ως αποτέλεσμα κλίμακες από 1:1 (αρχική εικόνα, μηδενική κλίμακα) ως 1:256 (8η κλίμακα).



Σχήμα 3.1: Γενική αρχιτεκτονική του μοντέλου των Itti, Koch και Niebur. Από το [16]

Κάθε χαρακτηριστικό υπολογίζεται στη βάση γραμμικών λειτουργιών οι οποίες προσομοιάζουν οπτικά δεκτικά πεδία, όπως είδαμε και στο κεφάλαιο 2. Αυτή η αρχιτεκτονική είναι η πλέον κατάλληλη για τον εντοπισμό σημείων τα οποία κατέχουν εξέχουσα θέση από το περιβάλλον τους καθώς είναι ευαίσθητη σε τοπικές χωρικές ασυνέχειες. Αυτού του τύπου η ευαισθησία είναι μια γενική αρχή στον αμφιβληστροειδή χιτώνα (retina), ραχιαίο έξω γονατώδη πυρήνα (lateral geniculate nucleus) και στον πρωταρχικό οπτικό φλοιό (primary visual cortex), όπως αναφέραμε και στην ενότητα 2.1.3. Υλοποιείται στο μοντέλο ως η διαφορά μεταξύ λεπτομερών και αδρομερών κλιμάκων: το κέντρο, ή εστία, είναι ένα εικονοστοιχείο στην κλίμακα $c \in \{2, 3, 4\}$, και το περιβάλλον, ή περιφέρεια, είναι το αντίστοιχο εικονοστοιχείο στην κλίμακα $s = c + \delta, \delta \in \{3, 4\}$. Η διαφορά μεταξύ δυο χαρτών οι οποίοι βρίσκονται σε διαφορετική κλίμακα, η οποία θα συμβολίζεται με “ \ominus ” στη συνέχεια, προκύπτει από παρεμβολή στην πιο λεπτομερή κλίμακα και αφαίρεση σημείου - προς - σημείο, δηλαδή αν έχουμε τους δύο χάρτες f_p και g_q στις κλίμακες p και q αντίστοιχα, η διαφορά αυτή θα προκύπτει ως

$$f_p \ominus g_q = f_p \uparrow^r - g_q \uparrow^r, \quad r = \min\{p, q\}$$

Η χρήση διαφορετικών κλιμάκων τόσο για το c όσο και για το $\delta = s - c$ παρέχει

μια εύρεση χαρακτηριστικών πραγματικά πολλαπλών επιπέδων.

3.2.1 Εξαγωγή Πρωταρχικών Οπτικών Χαρακτηριστικών

Αρχικά υπολογίζεται η ένταση, I , της εικόνας σε κάθε pixel ως $I = (r + g + b)/3$, όπου r , g και b το κόκκινο, πράσινο και μπλε κανάλι της εικόνας εισόδου αντίστοιχα. Από την ένταση δημιουργείται μια γκαουσιανή πυραμίδα, $I(\sigma)$, όπου $\sigma \in [0..8]$ είναι η κλίμακα και $I(0) = I$. Στη συνέχεια τα τρία χρωματικά κανάλια κανονικοποιούνται από το I με σκοπό να αποσυζευχθεί η απόχρωση, hue, από την ένταση. Επίσης, λόγω του ότι οι διαφορές στην απόχρωση δεν γίνονται αντιληπτές σε πολύ μικρή ένταση, η κανονικοποίηση αυτή εφαρμόζεται μόνο σε σημεία όπου η ένταση παίρνει τιμές μεγαλύτερες από το 1/10 του μεγίστου της. Οι τιμές των υπόλοιπων σημείων (αυτών δηλαδή με ένταση φωτεινότητας μικρότερη του 1/10 της μέγιστης) των καναλιών r , g και b τίθενται ίσες με μηδέν.

Στη συνέχεια δημιουργούνται τέσσερα νέα χρωματικά κανάλια, $R = r - (g + b)/2$ για το κόκκινο, $G = g - (r + b)/2$ για το πράσινο, $B = b - (r + g)/2$ για το μπλε και $Y = (r + g)/2 - |r - g|/2 - b$ για το κίτρινο. Από αυτά κατασκευάζονται τέσσερις νέες γκαουσιανές πυραμίδες $R(\sigma)$, $G(\sigma)$, $B(\sigma)$ και $Y(\sigma)$ με τον ίδιο τρόπο όπως και η $I(\sigma)$.

Έπειτα υπολογίζονται οι διαφορές εστίας - περιφέρειας, οι οποίες συμβολίζονται με Θ , μεταξύ μιας εστιακής λεπτής κλίμακας c και μιας περιφερειακής πιο αδρής κλίμακας s . Από αυτές προκύπτουν οι αντίστοιχοι χάρτες χαρακτηριστικών. Η πρώτη ομάδα χαρτών αυτού του είδους αφορά τις διαφορές έντασης μεταξύ των διαφόρων κλιμάκων. Αυτές έχει βρεθεί ότι, στα θηλαστικά, γίνονται αντιληπτές μέσω νευρώνων ευαίσθητων είτε σε σκοτεινά κέντρα σε φωτεινό φόντο είτε σε φωτεινά κέντρα σε σκοτεινό φόντο [27]. Τα δύο αυτά είδη ευαισθησίας υπολογίζονται ταυτόχρονα σε ένα σύνολο έξι χαρτών $I(c, s)$ όπου $c \in \{2, 3, 4\}$ αντιπροσωπεύει την κεντρική, εστιακή κλίμακα και $s = c + \delta$, $\delta \in \{3, 4\}$ αντιπροσωπεύει την αδρή, περιφερειακή κλίμακα:

$$I(c, s) = |I(c) - I(s)|$$

Ένα δεύτερο σύνολο χαρτών κατασκευάζεται και για τα χρωματικά κανάλια τα οποία βιολογικά αντιπροσωπεύονται με ένα σύστημα “διπλού ανταγωνισμού”: στο κέντρο των πεδίων πρόσληψης οπτικής πληροφορίας, οι νευρώνες ενεργοποιούνται από ένα χρώμα - π.χ. το κόκκινο - και καταστέλλονται από ένα άλλο - π.χ. το πράσινο. Το αντίστροφο συμβαίνει στην περιφέρεια. Αυτός ο χωρικός και χρωματικός ανταγωνισμός έχει δειχθεί να ισχύει για τα ζεύγη χρωμάτων κόκκινο/πράσινο, πράσινο/κόκκινο, μπλε/κίτρινο και κίτρινο/μπλε

[7]. Σε αντιστοιχία με τους χάρτες πολυκλιμακωτών διαφορών έντασης, $I(c, s)$, υπολογίζουμε τους δύο χάρτες, $RG(c, s)$ για το διπλό ανταγωνισμό κόκκινου/πράσινου και πράσινου/κόκκινου και $BU(c, s)$ για το διπλό ανταγωνισμό μπλε/κίτρινου και κίτρινου/μπλε:

$$\begin{aligned} RG(c, s) &= |(R(c) - G(c)) \ominus (G(s) - R(s))| \\ BY(c, s) &= |(B(c) - Y(c)) \ominus (Y(s) - B(s))| \end{aligned}$$

Τέλος, η πληροφορία για τον τοπικό προσανατολισμό εξάγεται από την I χρησιμοποιώντας προσανατολισμένες πυραμίδες Gabor $O(\sigma, \theta)$ όπου και εδώ $\sigma \in [0..8]$ αντιπροσωπεύει την κλίμακα και $\theta \in \{0^\circ, 45^\circ, 90^\circ, 135^\circ\}$ είναι ο επιθυμητός προσανατολισμός. Οι πυραμίδες αυτές περιγράφονται στο [10] και προκύπτουν με φιλτράρισμα με προσανατολισμένους Gabor πυρήνες των επιπέδων της Λαπλασιανής πυραμίδας της εικόνας. Συμβολίζοντας δηλαδή με ψ_θ έναν τέτοιο πυρήνα, προσανατολισμένο σε γωνία θ , έχουμε

$$O(\sigma, \theta) = L_\sigma * \psi_\theta$$

όπου L η Λαπλασιανή πυραμίδα. Ο αλγόριθμος ταυτόχρονου υπολογισμού της τελευταίας μαζί με τη Γκαουσιανή, G , πυραμίδα που χρησιμοποιείται είναι αυτός που περιγράψαμε και στο κεφάλαιο 2, δηλαδή ο

$$\begin{aligned} G_{i+1,0} &= W * G_n; L_i = G_i - G_{i+1,0} \\ G_{i+1} &= (G_{i+1,0}) \downarrow \end{aligned}$$

με G_0 την αρχική εικόνα. Σχηματικά η όλη διαδικασία φαίνεται στο Σχήμα 3.2.

Οι πυρήνες Gabor ψ_θ , με τους οποίους συνελίσσουμε τα διάφορα επίπεδα της Λαπλασιανής πυραμίδας για την εξαγωγή της πληροφορίας προσανατολισμού για κάθε γωνία θ , έχουν τη γενική μορφή

$$\psi_\theta = LPF[m_\theta]$$

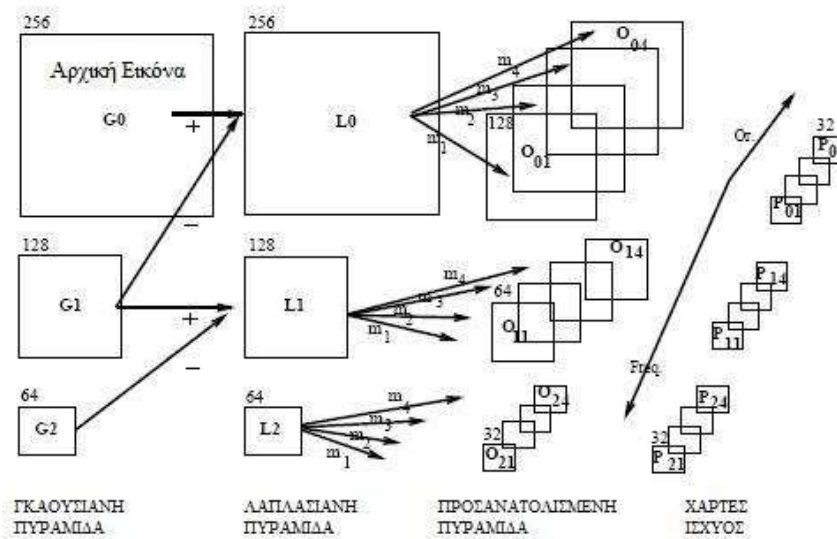
όπου m_θ είναι ένα μιγαδικό ημιτονοειδές προσανατολισμένο σε γωνία θ , ως εξής:

$$m_\theta(x, y) = e^{j(\pi/2)(\cos\theta \times x + \sin\theta \times y)}$$

Οι πυρήνες Gabor έχουν συνεπώς τη μορφή

$$\psi_\theta = \frac{1}{2\pi} e^{(j\pi/2)(x^2/\sigma_x^2 + y^2/\sigma_y^2)} \times m_\theta$$

Στο [10], όπως και στον αλγόριθμο εύρεσης ενδιαφερόντων σημείων που περιγράφουμε χρησιμοποιούνται τέτοια φίλτρα στις γωνίες $\theta = 0^\circ, 45^\circ, 90^\circ$ και 135° . Τα τέσσερα αυτά φίλτρα φαίνονται και στο Σχήμα 3.3. Χρησιμοποιώντας



Σχήμα 3.2: Διάγραμμα της αποδόμησης της εικόνας σε Γκαουσιανές και Λαπλασιανές πυραμίδες, και επακόλουθο φιλτράρισμα με πυρήνες Gabor. Από το [10]

τέτοιες ομάδες φίλτρων εύρους προσανατολισμού 45° , στο [10] επιδεικνύεται ότι μπορεί να επιτευχθεί κάλυψη 99% στο σύνολο του χώρου προσανατολισμού.

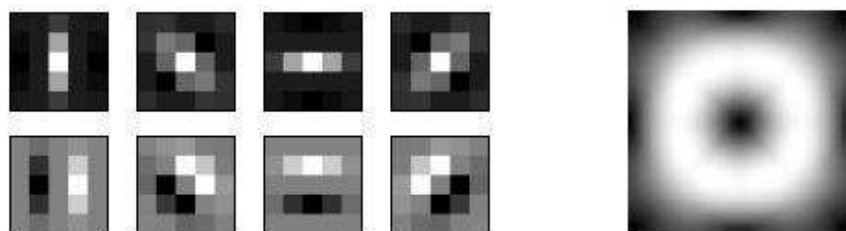
Χρησιμοποιώντας τα πιο πάνω, οι χάρτες των χαρακτηριστικών προσανατολισμού, κωδικοποιούν, σαν σύνολο, τη διαφορά του προσανατολισμού μεταξύ των κλιμάκων εστιών και περιφερειών με όμοιο τρόπο όπως και οι άλλοι δύο τύποι χαρτών:

$$O(c, s, \theta) = |O(c, \theta) \ominus O(s, \theta)|$$

Υπολογίζονται συνολικά 42 χάρτες χαρακτηριστικών: 6 για την ένταση, 12 για τον χρωματικό ανταγωνισμό, και 24 για τους προσανατολισμούς.

3.2.2 Συγκρότηση του Χάρτη Οπτικής Προσοχής

Μια σημαντική δυσκολία στο συνδυασμό των διαφόρων χαρτών χαρακτηριστικών είναι το ότι αντιπροσωπεύουν εγγενώς μη συγκρίσιμα μεγέθη, τα οποία έχουν διαφορετικά πεδία τιμών και έχουν εξαχθεί με διαφορετική μεθοδολογία το κάθε ένα, όπως περιγράφηκε πιο πάνω. Επίσης, στο συνδυασμό και των 42 χαρτών, αντικείμενα που διακρίνονται σημαντικά σε ένα μικρό αριθμό χαρτών μπορεί να καλυφθούν με θόρυβο ή λιγότερο σημαντικά αντικείμενα μέσα σε ένα



Σχήμα 3.3: Αριστερά: Σύνολο προσανατολισμένων φίλτρων για τις τέσσερις κατευθύνσεις - πάνω: πραγματικά μέρη, κάτω: φανταστικά μέρη. Δεξιά: Χαρακτηριστικά φάσματος ισχύος του συνόλου των φίλτρων (μαζί με τα αντίστοιχα συζυγή). Από το [10].

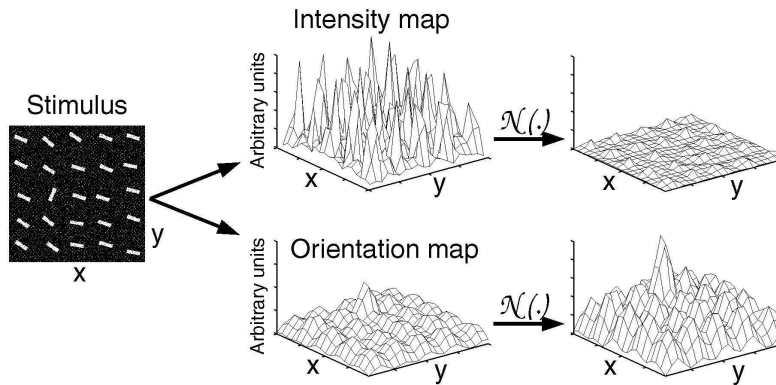
μεγάλο αριθμό χαρτών. Πιθανές λύσεις στο πρόβλημα αυτό της βέλτιστης σύντηξης των επιμέρους χαρτών χαρακτηριστικών περιλαμβάνουν απλή άθροιση, άθροιση με βάρη βάσει κάποιου τελεστή κανονικοποίησης ή ένα επαναληπτικό άθροισμα μεταξύ κάθε χάρτη χαρακτηριστικού και του φιλτραρισμένου χάρτη διαφορών των Γκαουσιανών¹.

Οι Itti, Koch και Niebur προτείνουν, στο αρχικό τους μοντέλο, ένα μηχανισμό κανονικοποίησης των χαρτών προς αποφυγή των πιο πάνω προβλημάτων (μεταγενέστερα βρέθηκε και η τρίτη, επαναληπτική μέθοδος να δίνει πολύ καλά αποτελέσματα). Ο μηχανισμός αυτός, τον οποίο αντιπροσωπεύουμε από εδώ και στο εξής ως τον τελεστή $N(\cdot)$, ενισχύει χάρτες στους οποίους υπάρχει ένας μικρός αριθμός από εξέχουσες και διακριτές κορυφές – αντιπροσωπεύοντας σημεία με υψηλό δείκτη ενδιαφέροντος. Αντίθετα, καταπνίγει χάρτες στους οποίους υπάρχει ένα μεγάλο πλήθος κορυφών με συγκρίσιμες τιμές μεταξύ τους. Ο τελεστής κανονικοποίησης $N(\cdot)$ δρα ως εξής:

1. Κανονικοποιεί τις τιμές του χάρτη σε μια σταθερή ακτίνα $[0..M]$ με σκοπό να εξαλειφθούν διαφορές της απόκρισης οι οποίες εξαρτώνται από τη φύση του κάθε χαρακτηριστικού,
2. Ξεχωρίζει το ολικό μέγιστο M και υπολογίζοντας το μέσο όρο, \bar{m} όλων των υπόλοιπων τοπικών μεγίστων και, τέλος
3. Πολλαπλασιάζει όλο το χάρτη με $(M - \bar{m})^2$.

Θεωρώντας τα τοπικά μέγιστα ως αποκρίσεις “ σημείων ενεργοποίησης ” μέσα στο χάρτη, ο τελεστής $N(\cdot)$ αγνοεί ομογενείς περιοχές, πράγμα που

¹Οι τρεις αυτές μεθόδοι παρέχονται επίσης ως επιλογές κανονικοποίησης στην εφαρμογή Saliency Toolbox που αναπτύχθηκε στο Koch Lab του Caltech.



Σχήμα 3.4: Η λειτουργία του τελεστή $\mathcal{N}(\cdot)$. Από το [16]

του επιτρέπει να υπολογίσει το μέτρο του μέσου επιπέδου ενεργοποίησης σε ολόκληρο το χάρτη, το οποίο αντιπροσωπεύεται από το \bar{m} . Εάν η διαφορά του μέσου αυτού επιπέδου από την απόκριση στο σημείο της μέγιστης ενεργοποίησης είναι μεγάλη, το σημείο αυτό κατέχει εξέχουσα θέση και ο χάρτης ενισχύεται σημαντικά. Αν, αντίθετα, η διαφορά αυτή είναι μικρή, ο χάρτης δεν περιέχει τίποτα το εξέχον και μοναδικό και έτσι η συμβολή του καταπνίγεται. Το κίνητρο των Itti, Koch και Niebur πίσω από το σχεδιασμό αυτό προέρχεται από τη βιολογία: σε γενικές γραμμές προσομοιάζει ανασταλτικούς μηχανισμούς στον πλευρικό φλοιό, στους οποίους γειτονικά όμοια χαρακτηριστικά αλληλοαναστέλλονται μέσω συγκεκριμένων, ανατομικά ορισμένων συνδέσεων [5].

Οι χάρτες χαρακτηριστικών συνδυάζονται σε τρεις ενδιάμεσους χάρτες οπτικής προσοχής ('conspicuity maps'): \bar{I} για την ένταση, \bar{C} για το χρώμα και \bar{O} για τους προσανατολισμούς, στην κλίμακα $\sigma = 4$. Προκύπτουν από άθροιση σε όλες τις κλίμακες, \otimes , δηλαδή προβολή των κανονικοποιημένων χαρτών στην κλίμακα 4 και άθροιση σημείο - προς - σημείο — συμβολίζοντας έναν οποιοδήποτε χάρτη με T_i , η λειτουργία του τελεστή ισοδυναμεί με

$$\otimes_i = \sum_i T_i \downarrow_4$$

Έτσι οι ενδιάμεσοι χάρτες υπολογίζονται όπως πιο κάτω:

$$\bar{I} = \otimes_{c=2}^4 \otimes_{s=c+3}^{c+4} N(I(c, s))$$

$$\bar{C} = \otimes_{c=2}^4 \otimes_{s=c+3}^{c+4} [N(RG(c, s)) + N(BY(c, s))]$$

Για τους προσανατολισμούς δημιουργούνται αρχικά τέσσερις ενδιάμεσοι χάρτες συνδυάζοντας τους έξι χάρτες χαρακτηριστικών που αντιστοιχούν στην ίδια

γωνία θ και στη συνέχεια συνδυάζονται σε ένα μοναδικό ενδιάμεσο χάρτη προσανατολισμού:

$$\bar{O} = \sum_{\theta \in \{0^\circ, 45^\circ, 90^\circ, 135^\circ\}} N \left(\bigotimes_{c=2}^4 \bigotimes_{s=c+3}^{c+4} N(O(c, s, \theta)) \right)$$

Το κίνητρο της δημιουργίας των τριών ξεχωριστών καναλιών \bar{I} , \bar{C} και \bar{O} και η ξεχωριστή κανονικοποίησή τους ήταν η υπόθεση ότι όμοια χαρακτηριστικά ανταγωνίζονται έντονα για οπτική προσοχή, ενώ διαφορετικά είδη χαρακτηριστικών συνεισφέρουν ανεξάρτητα στον τελικό χάρτη. Οι τρεις ενδιάμεσοι χάρτες ενδιαφέροντος κανονικοποιούνται και αθροίζονται στον τελικό χάρτη οπτικής προσοχής (saliency map) S :

$$S = \frac{1}{3} (N(\bar{I}) + N(\bar{C}) + N(\bar{O}))$$

Κάθε δεδομένη στιγμή, το μέγιστο του χάρτη οπτικής προσοχής ορίζει την πιο σημαντική τοποθεσία μέσα στην εικόνα, το σημείο δηλαδή που αποτελεί το επίκεντρο της προσοχής. Αντί απλά να επιλέγουν το σημείο αυτό, οι Itti, Koch και Niebur επεκτείνουν την ιδέα τους σε μια εφαρμογή που προσομοιάζει καλύτερα τον τρόπο με τον οποίο ένα βιολογικό σύστημα, χρησιμοποιώντας τους νευρώνες του, επιλέγει τη σειρά των σημείων στα οποία θα εστιάσει την προσοχή του μέσα στο οπτικό του πεδίο. Ο χάρτης οπτικής προσοχής (ΧΟΠ) μοντελοποιείται ως ένα διδιάστατο πλέγμα νευρώνων ολοκλήρωσης - και - πυροδότησης (integrate-and-fire). Αυτοί οι νευρώνες αποτελούνται από μια απλή χωρητικότητα η οποία ολοκληρώνει το φορτίο που δέχεται σε μια είσοδο, και ένα κατώφλι τάσης. Μόλις φτάσει σε αυτό το κατώφλι, ο νευρώνας πυροδοτείται ενώ το φορτίο του μηδενίζεται [21]. Ο ΧΟΠ τροφοδοτεί ένα διδιάστατο "όλα - για - το - νικητή" (WTA, OFN) νευρωνικό δίκτυο στην ίδια κλίμακα ($\sigma = 4$) στο οποίο οι συνάψεις μεταξύ των στοιχείων του εξασφαλίζουν ότι μόνο το πιο ενεργό σημείο παραμένει, ενώ τα υπόλοιπα καταπνίγονται.

Οι νευρώνες στον ΧΟΠ λαμβάνουν ενεργοποιήσεις από το S και είναι όλοι ανεξάρτητοι μεταξύ τους. Έτσι, το δυναμικό των νευρώνων που αντιστοιχούν σε πιο αξιοπρόσεκτες περιοχές αυξάνεται γρηγορότερα. Οι νευρώνες του ΧΟΠ απλά δρουν ως ολοκληρωτές για το φορτίο που δέχονται και δεν πυροδοτούνται. Δρουν όμως ως εισόδοι των αντίστοιχων νευρώνων του OFN δικτύου, οι οποίοι εξελίσσονται ανεξάρτητα ο ένας από τον άλλον ωστόσο ένας, ο "νικητής", ξεπεράσει πρώτος το κατώφλι και πυροδοτήσει. Αυτό με τη σειρά του πυροδοτεί τρεις ταυτόχρονους μηχανισμούς:

1. Το επίκεντρο της προσοχής (ΕπΠ) μετακινείται στη θέση του νευρώνα νικητή,
2. ενεργοποιείται η ολική απαγόρευση του OFN δικτύου, η οποία μηδενίζει το φορτίο όλων των OFN νευρώνων,

3. η τοπική απαγόρευση ενεργοποιείται μεταβατικά στο χάρτη οπτικής προσοχής, σε μια περιοχή του μεγέθους και της θέσης του ΕΤΠ· αυτό όχι μόνο προκαλεί δυναμικές μεταβάσεις του ΕτΠ, επιτρέποντας στο επόμενο πιο σημαντικό σημείο να λάβει τη θέση του νικητή, αλλά επίσης αποτρέπει το ΕτΠ από το να επιστρέψει άμεσα σε ένα σημείο όπου βρισκόταν μια πρόσφατη χρονική στιγμή.

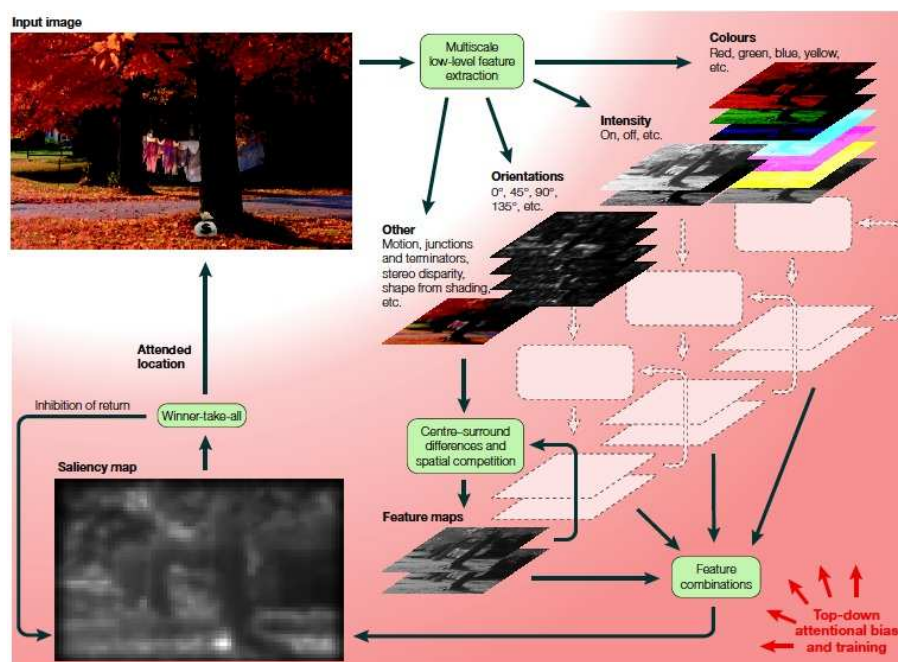
Μια τέτοια “απαγόρευση επιστροφής” έχει επιδειχθεί ως χαρακτηριστικό της λειτουργίας της ανθρώπινης όρασης [39]. Οι χρονικές σταθερές, χωρητικότητες και κατώφλια του νευρωνικού δικτύου επελέγησαν βάσει ψυχοφυσικών προτύπων έτσι ώστε το ΕτΠ μεταπηδά από μια αξιοπρόσεκτη περιοχή σε μια άλλη σε περίπου 30–70ms, ενώ η περιοχή ενός πρώην ΕτΠ αναστέλλεται για περίπου 500–900ms. Το μοντέλο εμπεριέχει επίσης μια ελαφρά προκατάληψη (bias) υπέρ αξιοπρόσεκτων περιοχών πιο κοντά στο τρέχον ΕτΠ, αναπαριστώντας τον κανόνα της “κοντινής προτίμησης” (proximity preference) ο οποίος αναφέρεται στην αρχική εργασία των Koch και Ullman [20], η οποία εν πολλοίς αποτέλεσε τη βάση του μοντέλου.

Στην αρχική υλοποίηση του αλγορίθμου, καθ’ ότι δεν μοντελοποιήθηκε κάποιο μοντέλο προσοχής από πάνω προς τα κάτω, το ΕΤΠ αποτελείται από ένα δίσκο με ακτίνα 1/6 της μικρότερου από το μήκος και πλάτος της εικόνας εισόδου. Σε επόμενη υλοποίηση, όπως αυτή που διατίθεται στο διαδίκτυο στη σελίδα του εργαστηρίου Koch, υπάρχει η επιλογή τα όρια της περιοχής προσοχής να προσεγγίζονται στη βάση των ενδιάμεσων χαρτών, όπου η συνεχής περιοχή με αρκετά μεγάλες τιμές οπτικής προσοχής γύρω από το ΕΤΠ θεωρείται ότι ανήκει στο ίδιο “αξιοπρόσεκτο” (salient) αντικείμενο. Επίσης, όπως αναφέραμε και στην εισαγωγή του κεφαλαίου, έχει πρόσφατα προστεθεί μια πιθανοτική προσέγγιση του στοιχείου της έκπληξης με σχετική επιτυχία [15]. Στην εικόνα 3.6 βλέπουμε ένα πλήρες παράδειγμα εφαρμογής του μοντέλου σε μια εικόνα, το οποίο περιέχει τόσο την εξαγωγή των επιμέρους χαρτών χαρακτηριστικών και χάρτη προσοχής, όσο και τη χρονική εξέλιξη της απόκρισης του χάρτη προσοχής ως ένα νευρωνικό δίκτυο.

Το Σχήμα 3.5 δείχνει παραστατικά τη δομή και λειτουργία του μοντέλου που μόλις περιγράψαμε. Σημειώνουμε ότι σημειώνεται και η πιθανότητα ύπαρξης ακόμη ενός (ή και περισσότερων) ενδιάμεσων χαρτών οπτικής προσοχής (‘conspicuity maps’), που να αντιστοιχούν σε κίνηση, συνδέσμους, γωνίες κλπ. Έναν χάρτη που να αντιστοιχεί στο τελευταίο θα εξετάσουμε στην ενότητα 3.5.

3.3 Μεθοδολογία Σύγκρισης Αποκρίσεων

Μια από τις σημαντικότερες δυσκολίες στην διαδικασία της εύρεσης των ενδιαφέροντων σημείων και περιοχών σε μια εικόνα είναι αυτή της επαλήθευσης



Σχήμα 3.5: Σχηματική αναπαράσταση της δομής και λειτουργίας του μοντέλου των Itti et al.. Διακρίνουμε την πρόβλεψη για τη δυνατότητα ύπαρξης άλλων ενδιάμεσων χαρτών, όπως συνδέσμων, γωνιών, κίνησης κλπ. – η δυνατότητα προσθήκης ενός χάρτη γωνιών διερευνάται στην ενότητα 3.5.3. Εικόνα από [17].

του κάθε υπολογιστικού μοντέλου. Εκτός ίσως από κάποιες εικόνες όπου η πληροφορία και το οπτικό ενδιαφέρον συγκεντρώνονται σαφώς και αποκλειστικά σε ορισμένα σημεία, σε πολλές περιπτώσεις φυσικών σκηνών είναι δύσκολο να αποδώσουμε ένα μέτρο σημαντικότητας σε όλα τα σημεία της εικόνας. Δυο τέτοια παραδείγματα φαίνονται στο Σχήμα 3.7. Έτσι οδηγούμαστε σε δύο κύρια ζητήματα: το πρώτο, που μόλις αναφέραμε, συνίσταται στην εύρεση ενός προτύπου, του λεγόμενου "ground truth". Το δεύτερο ζήτημα είναι, δεδομένου του προτύπου, η εύρεση ενός μέτρου ομοιότητας μεταξύ του μοντέλου και αυτού.

Μια δυνατή προσέγγιση – και αυτή η οποία θα χρησιμοποιήσουμε – απαντά στο πρώτο ερώτημα θέτωντας ως πρότυπο την ανθρώπινη συμπεριφορά. Αν και υπάρχουν εφαρμογές στις οποίες η ανθρώπινη συμπεριφορά μπορεί να μην είναι το καταλληλότερο πρότυπο – όπως συστήματα στα οποία ενυπάρχει προκατάληψη (bias) προς τον εντοπισμό συγκεκριμένων χαρακτηριστικών, π.χ. κάποια συστήματα παρακολούθησης – το ανθρώπινο παράδειγμα αποτελεί ένα καλό σχετικά μέτρο σύγκρισης στις πλείστες περιπτώσεις φυσικών ή τεχνητών



Σχήμα 3.6: Εφαρμογή του υπολογιστικού μοντέλου των Itti et al. στην εικόνα cars. Πάνω – αριστερά: αρχική εικόνα, δεξιά: τελικός χάρτης οπτικής προσοχής. Δεύτερη σειρά: ενδιάμεσοι χάρτες χαρακτηριστικών. Κάτω – αριστερά: εξέλιξη των διαδοχικών θέσεων του επικέντρου της προσοχής (ΕτΠ). Δεξιά: αντίστοιχη εξέλιξη του WTA νευρωνικού δικτύου. Χρόνοι πυροδότησης της κάθε μίας από τις τρεις πρώτες θέσεις του ΕτΠ που απεικονίζονται αντιστοίχα: $108.2ms$, $176.8ms$ και $232.5ms$. Το σχήμα ετοιμάστηκε με τη βοήθεια του Saliency Toolbox το οποίο διατίθεται στην ιστοσελίδα του εργαστηρίου Koch (Caltech).



Σχήμα 3.7: Αριστερά: εικόνα με σαφή και ξεκάθαρα σημεία ενδιαφέροντος. Δεξιά: εικόνα όπου είναι πολύ δύσκολο να ορίσουμε αντικειμενικά μια ιεραρχία περιοχών ενδιαφέροντος. Οι εικόνες έχουν παρθεί από βάσεις δεδομένων δημόσιας κυριότητας (public domain databases).

σκηνών.

Διάφορα ζητήματα προκύπτουν καθώς έχει βρεθεί ότι ο κάθε ανθρώπινος παρατηρητής μπορεί να επιδεικνύει σημαντικά διαφορετική συμπεριφορά καθώς το οπτικό του πεδίο σαρώνει μια εικόνα. Συγκεκριμένα, ενώ πολλοί παρατηρητές συγκεντρώνονται στο ίδιο συνήθως αντικείμενο τις πρώτες στιγμές που βλέπουν μια εικόνα, στη συνέχεια οι συμπεριφορές μπορεί να ποικίλλουν. Αυτό εξηγείται από ένα σχεδόν όμοιο στον καθένα ανοδικό (bottom - up), ασυνείδητο μηχανισμό οπτικής προσοχής ο οποίος επικρατεί αρχικά, για να παραχωρήσει σε μετέπειτα στιγμές ένα μέρος της λειτουργίας σε καθοδικούς (top-down) μηχανισμούς οι οποίοι εξαρτούνται έντονα από τον παρατηρητή [30]. Στα πειράματα των Stentiford και Le Meur έχει δειχθεί ότι, αν και υπάρχουν αποκλίσεις από άτομο σε άτομο, η μέση απόκρισή τους είναι καλή ένδειξη του πώς παρατηρούν το περιβάλλον οι άνθρωποι, και συνεπώς ένα καλό μέτρο για την οπτική προσοχή που επιδεικνύουν. Παράλληλα το ανθρώπινο πρότυπο είναι σχετικά εύκολο στην καταγραφή — το πόσο σχετικά αναλύουμε συνοπτικά αμέσως πιο κάτω.

3.3.1 Παρακολούθηση της Ανθρώπινης Οπτικής Προσοχής

Οι μέθοδοι που χρησιμοποιήθηκαν στα περισσότερα πειράματα που οδήγησαν στα συμπεράσματα που αναφέραμε πιο πάνω έκαναν χρήση τεχνικών απευθείας παρακολούθησης της κίνησης του ματιού (eye tracking). Από τη δεκαετία του '60 μέχρι σήμερα έχουν αναπτυχθεί διάφορες τεχνικές για το σκοπό αυτό, οι πλείστες από τις οποίες είναι ιδιαίτερα παρεμβατικές προς τα ανθρώπινα

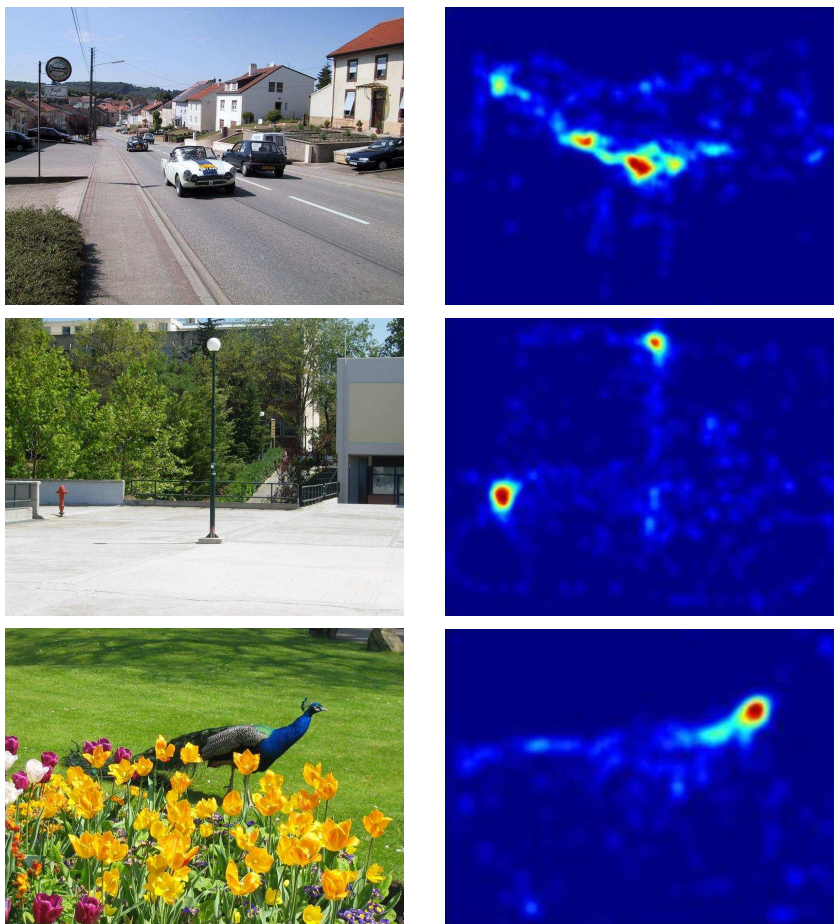
υποκείμενα, καθώς απαιτούν ακινησία του κεφαλιού, τοποθέτηση φακών επαφής με επιπλέον επαγωγικά πηνία στα μάτια κ.ο.κ. Μια σύγχρονη μέθοδος που έχει αναπτυχθεί από τον Le Meur εμπλέκει την παρακολούθηση της αντανάκλασης κατευθυνόμενου υπέρυθρου (και συνεπώς μη παρεμβαλλόμενου στο οπτικό πεδίο) φωτός πάνω στον κερατοειδή χιτώνα [26].

Σε κάθε περίπτωση, τέτοια ακριβή αποτελέσματα παρακολούθησης της ανθρώπινης όρασης είναι αρκετά δύσκολο να ληφθούν, τόσο σε κόστος, εξοπλισμό αλλά και εύρεση μεγάλου αριθμού πρόθυμων παρατηρητών. Μια προσέγγιση τους μπορεί να επιτευχθεί με τη χρήση τεχνικών παρακολούθησης της κίνησης του ποντικιού επάνω στην οθόνη (mouse tracking). Σε αυτή την απλή τεχνική αρχικά γίνεται η παρουσίαση στην οθόνη της προς έλεγχο εικόνας. Στη συνέχεια ζητείται από τον παρατηρητή να κοιτάζει την εικόνα για ένα συγκεκριμένο χρονικό διάστημα (συνήθως 10 - 15 δευτερόλεπτα) και να μετακινεί τον κέρσορα στο σημείο της εικόνας στο οποίο πέφτει το βλέμμα του κάθε στιγμή. Μέσω λογισμικού γίνεται καταγραφή της πορείας του κέρσορα πάνω στην εικόνα. Οι τελικοί πρότυποι χάρτες οπτικής προσοχής προκύπτουν συγκεντρώνοντας αποτελέσματα από πολλούς χρήστες (απαιτούνται συνήθως δείγματα από 30 περίπου άτομα), μαζί με βαθυπερατό φιλτράρισμα του μέσου όρου των αποκρίσεων. Η μέθοδος αυτή είναι ιδιαίτερα φθηνότερη από τις αντίστοιχες μεθόδους απευθείας παρακολούθησης του βλέμματος, καθώς αντί ειδικευμένου υλικού εξοπλισμού απαιτείται απλά κατάλληλο λογισμικό, ενώ παράλληλα είναι πολύ πιο φιλική προς τους παρατηρητές [30].

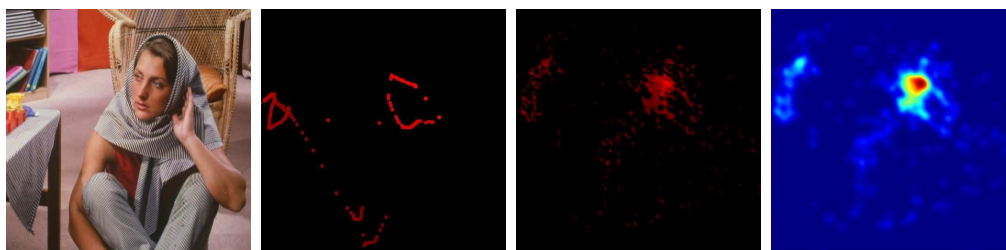
Μια τέτοια μέθοδος, υλοποιημένη σαν διαδικτυακή εφαρμογή, έχει αναπτυχθεί στο εργαστήριο Θεωρίας Κυκλωμάτων και Επεξεργασίας Σήματος (TCTS) του πανεπιστημίου της Mons του Βελγίου από τον μεταδιδακτορικό ερευνητή M. Mancas με βοήθεια φοιτητών. Απαιτεί απλά τη μετακίνηση του κέρσορα στις ενδιαφέρουσες περιοχές της εικόνας, όπως είπαμε και πιο πάνω ενώ η θέση του καταγράφεται κάθε 125ms σε ένα αρχείο. Η παρουσίαση της εικόνας στους παρατηρητές διαρκεί 10 δευτερόλεπτα, ενώ μεγαλύτερο βάρος δίνεται στις πρώτες εστιώσεις. Η διαδικτυακή εφαρμογή βρίσκεται στην ιστοσελίδα <http://tcts.fpms.ac.be/mousetrack/pageAccueil.php> και έχει υλοποιηθεί με ένα συνδυασμό κωδικοποίησης php και javascript. Στο Σχήμα 3.9 βλέπουμε τα αποτελέσματα του mouse tracking πάνω σε μια εικόνα, τόσο μια χαρακτηριστική απόκριση ενός μεμονωμένου χρήστη όσο και το μέσο όρο των αποτελεσμάτων από 55 χρήστες. Την εφαρμογή αυτή χρησιμοποιήσαμε σε συνεργασία με τους δημιουργούς της, ελέγχοντας 26 δικές μας εικόνες και καταγράφοντας τις αποκρίσεις από 21 μέχρι 31 παρατηρητών στην κάθε μια. Το σύνολο των παρατηρητών αποτελούσαν τόσο προπτυχιακοί και μεταπτυχιακοί φοιτητές όσο και άνθρωποι διαφόρων επαγγελμαμάτων και ηλικιών, σε μια προσπάθεια να αποτελέσουν ένα όσο δυνατόν πιο αντιπροσωπευτικό δείγμα. Επίσης, άλλες 14 εικόνες με τα αντίστοιχα αποτελέσματα παραχωρήθηκαν από την ιστοσελίδα.

Στην εικόνα 3.8 βλέπουμε τα αποτελέσματα σε τέσσερις διαφορετικές εικόνες. Οι αρχικές αποκρίσεις φιλτραρίστηκαν με γκαουσιανό φίλτρο τυπικής απόκλισης $\sigma = 6$ και παριστάνονται σε χρωματική κλίμακα στην οποία υψηλές τιμές παριστάνονται με κιτρινο - κόκκινο και χαμηλές τιμές με αποχρώσεις του μπλε (χάρτης χρώματος *jet*). Σκοπός του φιλτραρίσματος είναι η καλύτερη συνένωση των τροχιών προσοχής σε μορφές που αντιστοιχούν περισσότερο σε αντικείμενα, αλλά και η επίτευξη της απαιτούμενης ομοιότητας με αντίστοιχα αποτελέσματα από τα πειράματα απευθείας παρακολούθησης του βλέμματος.

Με την καταγραφή της οπτικής προσοχής μέσω ποντικιού προκύπτουν αρκετά ζητήματα [30], όπως



Σχήμα 3.8: Εικόνες (αριστερά) και αντίστοιχοι φιλτραρισμένοι μέσοι όροι αποκρίσεων (δεξιά). Από πάνω προς κάτω: εικόνα cars.jpg – αποτελέσματα από 23 διαφορετικές αποκρίσεις, εικόνα post.jpg – 25 διαφορετικές αποκρίσεις, εικόνα peacock.jpg – 28 διαφορετικές αποκρίσεις.



Σχήμα 3.9: Από αριστερά: εικόνα προς έλεγχο· ενδεικτικό αποτέλεσμα του mouse tracking από ένα χρήστη· μέσος όρος αποτελεσμάτων από 55 χρήστες (από την ιστοσελίδα mouse tracking του TCTS)· μέσος όρος μετά από βαθυπερατό φιλτράρισμα με Γκαουσιανό φίλτρο τυπικής απόκλισης $\sigma = 6$.

- Η αρχική θέση του δείκτη - κέρσορα μπορεί να προκαταλαμβάνει τα αποτελέσματα του μοντέλου, καθότι ο δείκτης αποτελεί, ιδιαίτερα για τους συνηθισμένους στη χρήση υπολογιστή παρατηρητές, το πρώτο σημείο το οποίο ψάχνουν στην οθόνη. Διαφορετικές προσεγγίσεις αποτελούν η τοποθέτηση του εκτός ή στο κέντρο της εικόνας. Βάσει του λογισμικού που χρησιμοποιήσαμε, ο δείκτης ήταν αρχικά τοποθετημένος πάνω αλλά όχι κατ' ανάγκη στο κέντρο της εικόνας.
- Με τη μέθοδο αυτή επιλέγονται μόνο περιοχές οι οποίες προκαλούν συνειδητά το ενδιαφέρον του παρατηρητή – αν και αυτό δεν είναι κατ' ανάγκη μειονέκτημα καθ' ότι κυρίως αυτές προσπαθούμε να προβλέψουμε.
- Ο ίδιος ο δείκτης κρύβει μέρος της εικόνας, με αποτέλεσμα να μη βρίσκεται ακριβώς επάνω στις ενδιαφέρουσες περιοχές αλλά πολύ κοντά σε αυτές. Το φαινόμενο αυτό μετριάζεται αισθητά με το βαθυπερατό φιλτράρισμα που γίνεται στο τέλος.

Παρ' όλα αυτά οι αποκρίσεις των μεθόδων παρακολούθησης της οπτικής προσοχής μέσω της κίνησης του ποντικιού δίνουν μια πολύ καλή προσέγγιση των αντίστοιχων που λαμβάνονται με πιο επίπονες και παρεμβατικές μεθόδους απευθείας παρακολούθησης της κίνησης του βλέμματος. Πριν αναφέρουμε τα σχετικά αποτελέσματα που δείχνουν αυτή την ομοιότητα, απαραίτητο είναι να ορίσουμε αρχικά ένα μέτρο για αυτήν.

3.3.2 Μέτρο της ομοιότητας μεταξύ των αποκρίσεων

Ως μέτρο της ομοιότητας μεταξύ των αποκρίσεων, θα χρησιμοποιήσουμε τη γραμμική δισδιάστατη συσχέτιση (εξίσωση 3.1). Η συσχέτιση μεταξύ δύο εικόνων λαμβάνει τιμές από το -1, όταν η μια εικόνα αποτελεί αναστροφή της



Σχήμα 3.10: Εικόνες με διαφορετικές τιμές συσχέτισης μεταξύ τους. Αριθμώντας από αριστερά, εικόνες 1-3: πλαίσια 1,2 και 36 από την ακολουθία *Hamburg taxi* — 4η: γκριζα εικόνα *cars* στο ίδιο μέγεθος. Τιμές των δισδιάστατων συσχετίσεων c_{ij} : $c_{12} = 0.9879$, $c_{13} = 0.7962$, $c_{14} = -0.1560$, $c_{23} = 0.7972$, $c_{24} = -0.1556$ και $c_{34} = -0.1568$.

άλλης – και γενικά με τιμές κάπως μικρότερες από το 0 δε θεωρούμε κάποια ομοιότητα – ως το 1, το οποίο δείχνει ότι οι εικόνες έχουν μια γραμμική σχέση μεταξύ τους. Στην παρακάτω σχέση θεωρούμε τις εικόνες - αποκρίσεις I_1 και I_2 με αντίστοιχες μέσες τιμές μ_1 και μ_2 .

$$C = \frac{\sum_i \sum_j (I_1(i, j) - \mu_1) \times (I_2(i, j) - \mu_2)}{\sqrt{\sum_i \sum_j (I_1(i, j) - \mu_1)^2 \times \sum_i \sum_j (I_2(i, j) - \mu_2)^2}} \quad (3.1)$$

Προσπαθώντας να δώσουμε μια οπτική άποψη του μέτρου αυτού, στο Σχήμα 3.10 δίνουμε τέσσερις εικόνες και τις αντίστοιχες μεταξύ τους τιμές συσχέτισης.

Όπως αναφέραμε στην αμέσως προηγούμενη ενότητα, χρησιμοποιώντας το μέτρο αυτό έχειδει ότι υπάρχει μεγάλη ομοιότητα μεταξύ των αποκρίσεων από απευθείας παρακολούθηση του βλέμματος (*eye tracking*) με παρακολούθηση με τη βοήθεια του ποντικιού του υπολογιστή (*mouse tracking*). Συγκεκριμένα, ελέγχθηκαν 14 εικόνες και οι αντίστοιχες αποκρίσεις βρέθηκαν να έχουν κατά μέσο όρο συσχέτιση 0.66 με τυπική απόκλιση 0.11 [30]. Οι αριθμοί αυτοί δείχνουν μια πολύ κοντινή σχέση μεταξύ των αποκρίσεων των δύο μεθόδων, στο βαθμό που μπορούμε να βασιστούμε στο (*mouse tracking*) ως αξιόπιστο μέτρο της προσοχής που αντιστοιχεί σε κάθε σημείο των εικόνων που θα ελέγξουμε.

3.4 Διερεύνηση Υλοποίησης του Μοντέλου των Itti et al. με Διαφορετικές Πυραμίδες

Σαν μια πρώτη απόπειρα εξέτασης της λειτουργίας και απόδοσης της αρχιτεκτονικής των Itti et al. υπό διάφορες παραλλαγές, θα δοκιμάσουμε την

αντικατάσταση της αρχικής γκαουσιανής πυραμίδας με κάποια μορφολογική πυραμίδα. Οι εναλλακτικές πυραμίδες που χρησιμοποιήσαμε ήταν αυτές που περιγράφησαν στην ενότητα 2.2, δηλαδή η *adjunction*, η πυραμίδα ανοίγματος - κλεισίματος (*open-close*), η πυραμίδα κυματιδίων *Haar* και η *leveling* πυραμίδα. Το κύριο στοιχείο που μας ώθησε αρχικά στη διερεύνηση αυτή είναι η καλύτερη διατήρηση των περιγραμμάτων η οποία χαρακτηρίζει τις μορφολογικές πυραμίδες. Ξεκινάμε αρχικά δίνοντας κάποια αποτελέσματα χρησιμοποιώντας την αρχική Γκαουσιανή πυραμίδα, για σκοπούς άμεσης σύγκρισης μεταξύ τους. Στη συνέχεια, αφού παρουσιάσουμε αντίστοιχες αποκρίσεις με χρήση και των υπόλοιπων πυραμίδων, θα τις αξιολογήσουμε χρησιμοποιώντας ως μέτρο την ανθρώπινη αντίληψη με τη μέθοδο που παρουσιάσαμε στην ενότητα 3.3.1. Σημειώνουμε ότι τα αποτελέσματα που ακολουθούν λήφθηκαν με δική μας υλοποίηση της αρχιτεκτονικής των Itti et al., ώστε να υπάρχει πιο εύκολη πρόσβαση στον καθορισμό των προς έλεγχο παραμέτρων, με πρώτη παράμετρο των τύπο της πυραμίδας, αλλά και άλλες που θα ελέγχουμε σε επόμενες ενότητες.

Η υλοποίηση του αλγορίθμου με την αρχική *Γκαουσιανή πυραμίδα* φαίνεται στην εικόνα 3.11. Θα παρουσιάσουμε τόσο τους ενδιάμεσους όσο και τον τελικό χάρτη οπτικής προσοχής και για τις πέντε πυραμίδες, χρησιμοποιώντας την ίδια εικόνα ώστε να υπάρχει μέτρο σύγκρισης. Για τους αλγόριθμους υλοποίησης των πυραμίδων μπορεί κάποιος να ανατρέξει στην ενότητα 2.2. Σημειώνουμε ότι, για μεγαλύτερη ευκρίνεια, οι χάρτες αθροίστηκαν στο επίπεδο 2 (και όχι 4) της πυραμίδας, πράγμα που ουσιαστικά δεν επηρεάζει τη λειτουργία των υπόλοιπων μηχανισμών του αλγορίθμου. Στην εικόνα 3.12 βλέπουμε τα αντίστοιχα αποτελέσματα για την *Adjunction* πυραμίδα. Το τελικό αποτέλεσμα δεν προσεγγίζει πολύ καλά τα αξιοπρόσεκτα αντικείμενα όσο η Γκαουσιανή πυραμίδα, αν και τα κυριότερα από αυτά παραμένουν τα ίδια - π.χ. το μεγάλο κόκκινο ταχυδρομικό κιβώτιο στην εικόνα, και η λευκή επιγραφή πάνω του. Τα περιγράμματα όμως διατηρούνται κάπως καλύτερα στους ενδιάμεσους και τον τελικό χάρτη, χωρίς να θολώνουν όσο πριν.

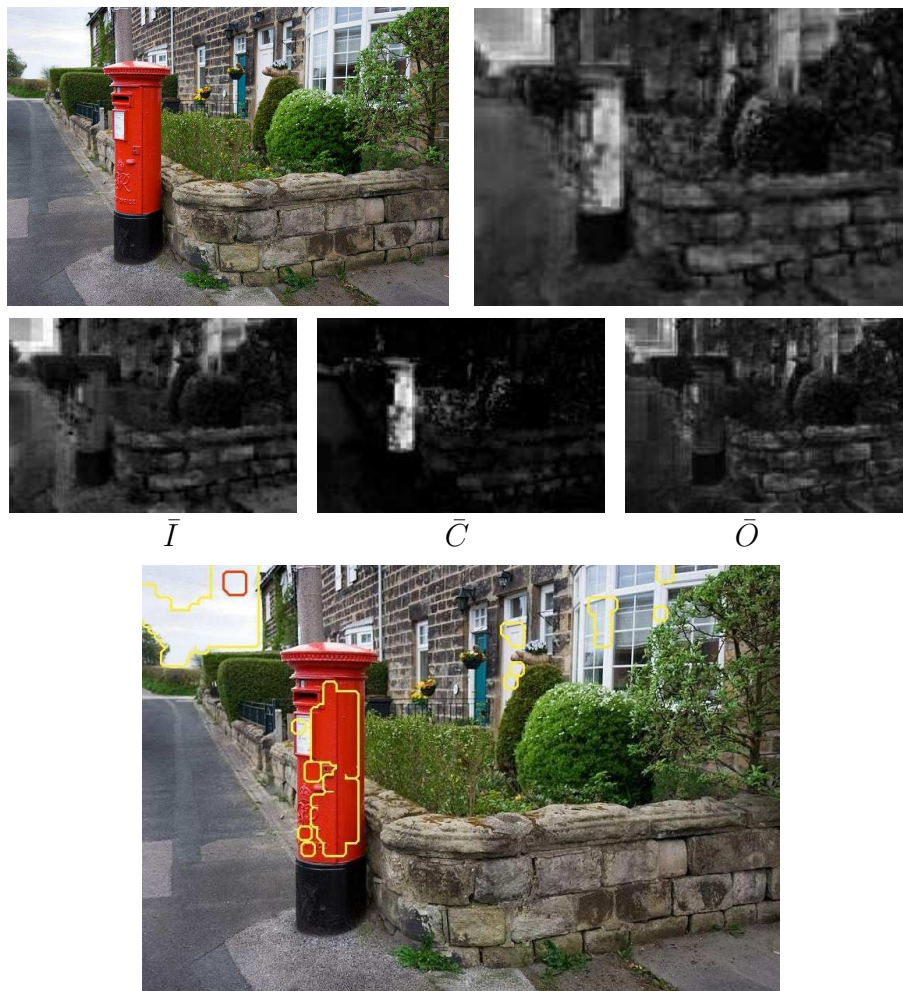
Συνεχίζοντας, στο Σχήμα 3.13 βλέπουμε τα αποτελέσματα της υλοποίησης του αλγορίθμου χρησιμοποιώντας την *πυραμίδα ανοίγματος - κλεισίματος*. Υπάρχει και εδώ μια σχετική ομοιότητα με τα αρχικά της Γκαουσιανής πυραμίδας, με τα αξιοπρόσεκτα αντικείμενα διατηρούν καλύτερα το περίγραμμά τους. Παράλληλα, εντοπίζονται περισσότερα αξιοπρόσεκτα αντικείμενα από ότι με την *adjunction* πυραμίδα.

Στην *πυραμίδα κυματιδίων Haar*, τα αποτελέσματα της χρήσης της βλέπουμε στο Σχήμα 3.14, τα αποτελέσματα μοιάζουν περισσότερο με τα αντίστοιχα της *adjunction*, όπου ισχύουν σε αδρές γραμμές τα ίδια σχόλια που κάναμε για εκείνη την περίπτωση.

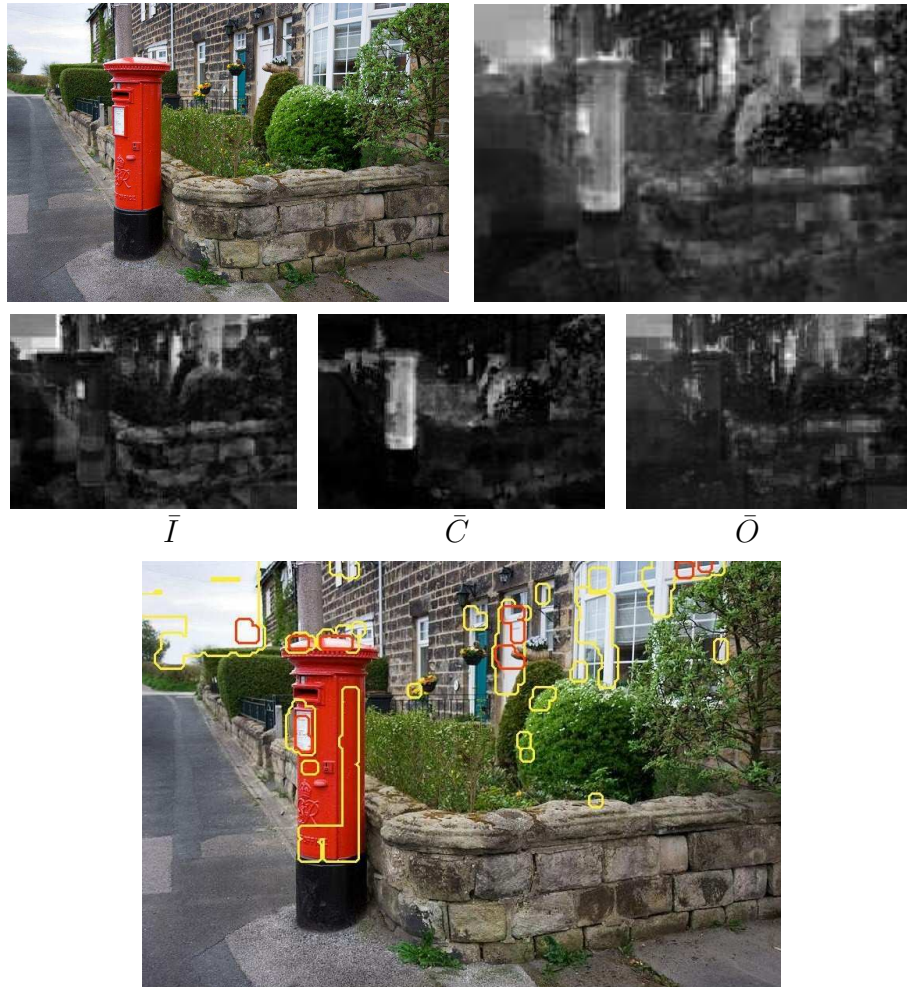
Για την *leveling* πυραμίδα, σημειώνουμε ότι δεν εξάγουμε μια μόνο “Λαπλασιανή” πυραμίδα ως διαφορά κάθε δύο διαδοχικών επιπέδων της πυραμίδας, αλλά δύο



Σχήμα 3.11: Επάνω – αριστερά: αρχική εικόνα *proa*. Δεξιά: τελικός χάρτης οπτικής προσοχής ο οποίος προέκυψε υλοποιώντας τον αλγόριθμο των Itti et al. με την αρχική Γκαουσιανή πυραμίδα. Οι χάρτες υπολογίστηκαν στο μέγεθος του επιπέδου 2 της πυραμίδας (αντί 4) για μεγαλύτερη ευκρίνεια. Μεσαία σειρά: ενδιάμεσοι χάρτες οπτικής προσοχής \bar{I} , \bar{C} και \bar{O} (αντίθεσης, χρωματικού ανταγωνισμού και προσανατολισμών). Κάτω: αρχική εικόνα με σημειωμένες τις περιοχές οι οποίες δίνουν μεγάλες αποκρίσεις στον τελικό χάρτη - με πορτοκαλί περίγραμμα $> 90\%$ του μεγίστου, με κίτρινο $> 64\%$ του μεγίστου. Τα όρια των περιοχών αυτών λήφθηκαν παίρνοντας αρχικά ένα δυαδικό σημαδεύτη τα σημεία που μας ενδιαφέρουν. Στη συνέχεια, βρήκαμε μια “μάσκα” γύρω από αυτά, ως τη διαφορά των dilations του σημαδευτή με μορφοποιητικά στοιχεία παραπλήσιας ακτίνας. Στη μάσκα αυτή θέσαμε η εικόνα να παίρνει τιμές κίτρινου ή πορτοκαλί αντίστοιχα.



Σχήμα 3.12: Επάνω – αριστερά: αρχική εικόνα *proa*. Δεξιά: τελικός χάρτης οπτικής προσοχής ο οποίος προέκυψε υλοποιώντας τον αλγόριθμο των Itti et al. με την Adjunction πυραμίδα (οι χάρτες υπολογίστηκαν στο μέγεθος του επιπέδου 2 της πυραμίδας). Μεσαία σειρά: ενδιάμεσοι χάρτες οπτικής προσοχής \bar{I} , \bar{C} και \bar{O} (αντίθεσης, χρωματικού ανταγωνισμού και προσανατολισμών). Κάτω: αρχική εικόνα με σημειωμένες τις περιοχές οι οποίες δίνουν μεγάλες αποκρίσεις στον τελικό χάρτη - με πορτοκαλί περίγραμμα $> 90\%$ του μεγίστου, με κίτρινο $> 64\%$ του μεγίστου.



Σχήμα 3.13: Επάνω – αριστερά: αρχική εικόνα *mailbox*. Δεξιά: τελικός χάρτης οπτικής προσοχής ο οποίος προέκυψε υλοποιώντας τον αλγόριθμο των Itti et al. με την πυραμίδα Ανοίγματος - Κλεισίματος (οι χάρτες υπολογίστηκαν στο μέγεθος του επιπέδου 2 της πυραμίδας). Μεσαία σειρά: ενδιάμεσοι χάρτες οπτικής προσοχής \bar{I} , \bar{C} και \bar{O} (αντίθεσης, χρωματικού ανταγωνισμού και προσανατολισμών). Κάτω: αρχική εικόνα με σημειωμένες τις περιοχές οι οποίες δίνουν μεγάλες αποκρίσεις στον τελικό χάρτη - με πορτοκαλί περίγραμμα $> 90\%$ του μεγίστου, με κίτρινο $> 64\%$ του μεγίστου.



Σχήμα 3.14: Επάνω – αριστερά: αρχική εικόνα *proa*. Δεξιά: τελικός χάρτης οπτικής προσοχής ο οποίος προέκυψε υλοποιώντας τον αλγόριθμο των Itti et al. με την πυραμίδα κυματιδίων Haar (οι χάρτες υπολογίστηκαν στο μέγεθος του επιπέδου 2 της πυραμίδας). Μεσαία σειρά: ενδιάμεσοι χάρτες οπτικής προσοχής \bar{I} , \bar{C} και \bar{O} (αντίθεσης, χρωματικού ανταγωνισμού και προσανατολισμών). Κάτω: αρχική εικόνα με σημειωμένες τις περιοχές οι οποίες δίνουν μεγάλες αποκρίσεις στον τελικό χάρτη - με πορτοκαλί περίγραμμα > 90 % του μεγίστου, με κίτρινο > 64 % του μεγίστου.

ξεχωριστές πυραμίδες, $L^{Q,V}$ και $L^{Q,P}$ οι οποίες αντιστοιχούν στις κοιλάδες που γεμίζονται και στις κορυφές που γκρεμίζονται, αντίστοιχα, κατά τη μετάβαση σε αδρότερη κλίμακα (βλέπε ενότητα 2.2.5). Τις δύο αυτές προσεγγίσεις της Λαπλασιανής που προκύπτουν τις φιλτράρουμε ξεχωριστά με φίλτρα Gabor, όπως στην ενότητα 3.2.1, παίρνοντας αρχικά δυο ξεχωριστές πυραμίδες προσανατολισμών, τις οποίες συμβολίζουμε με O_V και O_P :

$$\begin{aligned} O_V(\sigma, \theta) &= L_{\sigma}^{Q,V} * \psi_{\theta} \\ O_P(\sigma, \theta) &= L_{\sigma}^{Q,P} * \psi_{\theta} \end{aligned}$$

από τις οποίες σχηματίζουμε τους χάρτες διαφορών προσανατολισμού κέντρου - περιφέρειας:

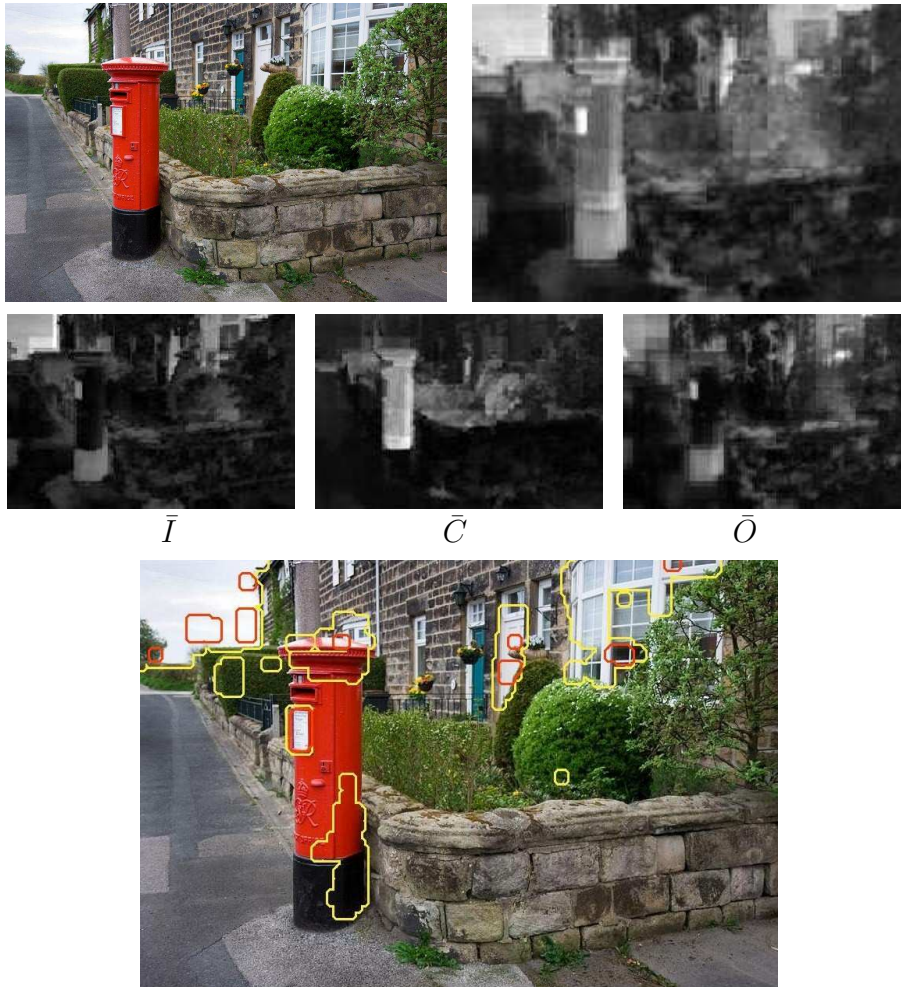
$$O_{VorP}(c, s, \theta) = |O_{VorP}(c, \theta) \ominus O_{VorP}(s, \theta)|$$

Ο ενδιάμεσος χάρτης προσανατολισμών, \bar{O} , προκύπτει δημιουργώντας αρχικά οκτώ ενδιάμεσους χάρτες συνδυάζοντας τους έξι χάρτες χαρακτηριστικών που αντιστοιχούν στην ίδια γωνία θ τόσο για τα O_V όσο και για τα O_P και στη συνέχεια κανονικοποιώντας και αθροίζοντας σε ένα μοναδικό ενδιάμεσο χάρτη προσανατολισμού, όπως πιο κάτω:

$$\bar{O} = \sum_{\theta \in \{0^\circ, 45^\circ, 90^\circ, 135^\circ\}} \left\{ N \left(\sum_{c=2}^4 \sum_{s=c+3}^{c+4} N(O_V(c, s, \theta)) \right) + N \left(\sum_{c=2}^4 \sum_{s=c+3}^{c+4} N(O_P(c, s, \theta)) \right) \right\}$$

Στο Σχήμα 3.15 βλέπουμε τα αποτελέσματα του αλγορίθμου με χρήση της leveling πυραμίδας. Φαίνεται, τόσο από το τελικό, όσο και από τα ενδιάμεσα αποτελέσματα, μια ομοιότητα με τα αντίστοιχα που προκύπτουν με την Γκαουσιανή πυραμίδα – και με την ανοίγματα - κλεισίματος – με τη διαφορά ότι τα περιγράμματα έχουν διατηρηθεί σαφώς καλύτερα. Π.χ., συγκρίνοντας με την εικόνα 3.11, βλέπουμε ότι και έχουν ανιχνευθεί διακριτές αξιοπρόσεκτες περιοχές, όπως το μικρό κομμάτι του ουρανού, η επιγραφή πάνω στο ταχυδρομικό κιβώτιο, η πόρτα, το μεγάλο παράθυρο κλπ. — επίσης, το περίγραμμα του ίδιου του ταχυδρομικού κιβωτίου ξεχωρίζει πλέον και στον τελικό χάρτη.

Παράγοντας παρόμοιους χάρτες και αποτελέσματα από όλες τις εικόνες που είχαμε στη διάθεσή μας, συγκεντρώσαμε και παραθέτουμε αναλυτικά στον Πίνακα 3.2 μια σύγκριση μεταξύ των αποκρίσεων των διάφορων υλοποιήσεων του αλγορίθμου με την κάθε μια από τις πέντε πυραμίδες. Ο Πίνακας 3.2 περιέχει τα μέτρα της ομοιότητας μεταξύ των αποκρίσεων αυτών και των προτύπων που εξήγαμε βάσει της ανθρώπινης συμπεριφοράς. Ως μέτρο ομοιότητας χρησιμοποιήσαμε τη δισδιάστατη ετεροσυσχέτιση, όπως στην εξίσωση 3.1. Παρατίθενται τα αποτελέσματα από κάθε μια από τις 26 εικόνες φυσικών σκηνών που αναλύσαμε, μαζί με τις επιπλέον 14 εικόνες που παραχωρήθηκαν από την ιστοσελίδα οπτικής προσοχής του TCTS, (αυτές κατά κατηγορία). Τέλος, υπολογίζεται η μέση τιμή και η διασπορά των αποκρίσεων για κάθε πυραμίδα.



Σχήμα 3.15: Επάνω – αριστερά: αρχική εικόνα *proa*. Δεξιά: τελικός χάρτης οπτικής προσοχής ο οποίος προέκυψε υλοποιώντας τον αλγόριθμο των Itti et al. με την πυραμίδα Leveling (οι χάρτες υπολογίστηκαν στο μέγεθος του επιπέδου 2 της πυραμίδας). Μεσαία σειρά: ενδιάμεσοι χάρτες οπτικής προσοχής \bar{I} , \bar{C} και \bar{O} (αντίθεσης, χρωματικού ανταγωνισμού και προσανατολισμών). Κάτω: αρχική εικόνα με σημειωμένες τις περιοχές οι οποίες δίνουν μεγάλες αποκρίσεις στον τελικό χάρτη - με πορτοκαλί περίγραμμα $> 90\%$ του μεγίστου, με κίτρινο $> 64\%$ του μεγίστου.

Εικόνα	G Γκαουσιανή	A Adjunction	M Open-Close	H κρυμ. Haar	Q Leveling
<i>balloon</i>	0.1726	-0.0275	0.0563	-0.0196	0.1182
<i>cars</i>	0.2245	-0.1534	0.1111	-0.0804	0.1654
<i>cows</i>	0.0937	-0.1343	0.0651	-0.0585	0.0536
<i>drive</i>	0.1675	-0.1132	0.1038	-0.0321	0.1220
<i>entrance</i>	0.2743	-0.0391	0.1288	0.0355	0.2266
<i>gardens</i>	0.0198	-0.1027	0.0181	-0.0515	-0.0369
<i>home</i>	-0.0365	-0.2146	-0.0922	-0.1771	-0.1612
<i>horses</i>	0.3608	-0.1591	0.1822	-0.1281	0.4017
<i>mechelec</i>	0.2754	0.0399	0.2948	0.1618	0.1644
<i>old building</i>	0.1094	-0.1287	-0.0392	-0.0845	0.1167
<i>parked cars</i>	-0.0292	-0.1278	-0.0516	-0.0908	-0.1131
<i>pbox</i>	0.3383	0.2130	0.3318	0.2722	0.3159
<i>peacock</i>	0.2453	-0.0415	0.0819	-0.0193	0.2439
<i>post</i>	0.2633	-0.0420	0.1867	0.0407	0.2150
<i>punchbowl</i>	0.1798	0.0210	0.1100	0.0378	0.0763
<i>road</i>	0.2445	-0.0494	0.1492	-0.0096	0.1581
<i>salesigns</i>	0.3941	0.0198	0.2795	0.0913	0.4014
<i>ship</i>	0.3639	-0.2824	0.2532	-0.1651	0.3605
<i>sign</i>	0.2861	0.0602	0.2016	0.1247	0.1622
<i>square</i>	0.0797	-0.0113	0.0896	0.0361	0.0395
<i>stairview</i>	0.0824	-0.2006	-0.0187	-0.1426	0.0098
<i>statue - krumlov</i>	0.0004	-0.1519	-0.0865	-0.1079	-0.0403
<i>street</i>	0.2177	-0.2776	0.0777	-0.0901	0.1924
<i>turbine</i>	0.1888	0.0645	0.1566	0.1903	0.1679
<i>wait</i>	0.1220	-0.0314	0.0674	0.0017	0.0428
<i>walking</i>	0.1119	-0.0873	-0.0300	-0.0846	0.1846
5 ειχ. φυσικών σκηνών	0.1143	-0.1035	0.0386	-0.0834	0.0552
3 ειχ. νυχτερινού φωτισμού	0.2883	0.0414	0.2330	0.0993	0.2454
3 ειχ. διαφημίσεων	0.0704	0.0524	0.0740	0.1281	0.0078
3 ειχ. ιστοσελίδων	0.1666	-0.0215	0.1091	0.0095	0.1857
Μέσος Όρος Αξιολόγησης $\in [-1, 1]$	0.1724	-0.0565	0.1017	-0.0014	0.1295
Τυπ. Απόκλιση	0.1196	0.1200	0.1332	0.1248	0.1463

Πίνακας 3.2: Αποτελέσματα της αξιολόγησης για τις υλοποιήσεις του αλγόριθμου των Itti et al. με τις πέντε διαφορετικές πυραμίδες. Φαίνονται αναλυτικά τα αποτελέσματα για τις 26 φυσικές σκηνές που ελέγξαμε μαζί με τις επιπλέον 14 εικόνες που παρέιχε η ιστοσελίδα της εφαρμογής validattention.

Παρατηρήσεις

Βλέποντας συνολικά τα αποτελέσματα στον Πίνακα 3.2, εξάγουμε το συμπέρασμα ότι η χρήση της Γκαουσιανής πυραμίδας, η οποία προτείνεται αρχικά, δίνει και τα πιο κοντινά στις ανθρώπινες αποκρίσεις αποτελέσματα, με αμέσως επόμενη, σε μικρή απόσταση, την leveling πυραμίδα. Γενικά, στις πλείστες από τις αποκρίσεις που εξετάσαμε, η leveling πυραμίδα δίνει παρόμοια αποτελέσματα με την Γκαουσιανή, με τη διαφορά ότι είναι πιο σαφή τα περιγράμματα των αντικειμένων. Κάπου ενδιάμεσα βρίσκεται και η απόκριση με τις open - close πυραμίδες, ενώ οι άλλες δύο συνήθως δεν δίνουν καλύτερα αποτελέσματα, όπως φαίνεται από τους μέσους όρους και από το γεγονός ότι μόνο σε 3 από τις 40 περιπτώσεις η πυραμίδα κυματιδίων Haar έδωσε τη μεγαλύτερη ομοιότητα με το πρότυπο (η adjunction σε καμία).

Ένα άλλο σημείο το οποίο μπορούμε να θίξουμε είναι ότι, το βαθυπερατό φιλτράρισμα των ανθρώπινων αποκρίσεων φυσιολογικά οδηγεί στην απώλεια των πιο απότομων περιγραμμάτων των αντικειμένων πάνω στους πρότυπους χάρτες οπτικής προσοχής, κάνοντας τους να μοιάζουν περισσότερο με αποκρίσεις στις οποίες χρησιμοποιήθηκαν Γκαουσιανά φίλτρα. Αυτό όμως δεν μπορούμε να πούμε ότι αποτελεί σφάλμα της μεθοδολογίας — το βλέμμα, όταν πεφτει πάνω στο όριο ενός αξιοπρόσεκτου αντικειμένου, δεν περιορίζεται αυστηρά από αυτό — γι αυτό, η σύγκριση με αποτελέσματα τα οποία βασίζονται στον ανθρώπινο παράγοντα, λογικό είναι να έχουν μια πιο θολή εικόνα της οπτικής προσοχής πάνω στα περιγράμματα. Εξ' άλλου, η μεθοδολογία αυτή βρέθηκε να δίνει πολύ κοντινά αποτελέσματα με τεχνικές απευθείας eye - tracking (αν και έχει προταθεί ότι η ομοιότητα αυτή οφείλεται στη διαδικασία ρύθμισης του eye - tracking η οποία γίνεται στο κέντρο της οθόνης παρατήρησης [26]). Συνεπώς, η κάθε μια από τις Gaussian και leveling πυραμίδες μπορεί να χρησιμοποιηθεί ανάλογα με τις απαιτήσεις του χρήστη ή ενός άλλου συστήματος το οποίο χρησιμοποιεί τα αποτελέσματα. Για παράδειγμα, αν απαιτείται από τις επόμενες εφαρμογές η εξαγωγή πληροφορίας περιγράμματος και από τον χάρτη οπτικής προσοχής (η οποία πιθανόν να βοηθήσει στην αναγνώριση του αντικειμένου κ.ο.κ.), θα προτιμηθεί η υλοποίηση με leveling πυραμίδες. Αν αυτό δεν είναι αναγκαίο, θα προτιμηθεί κατά κύριο λόγο η Γκαουσιανή πυραμίδα, λόγω ταχύτητας — οι διαρκείς επαναλήψεις του αλγορίθμου εύρεσης των levelings (βλ. ενότητα 2.2.5) προσδίδουν μεγαλύτερη υπολογιστική πολυπλοκότητα στην άλλη υλοποίηση. Αν και έχουμε την κάθε επιφύλαξη να αναφέρουμε ακριβή χρονικά διαστήματα, λόγω του μη πλήρη ελέγχου βελτιστοποίησης των αλγορίθμων που χρησιμοποιήσαμε, σε αδρές γραμμές η χρήση της leveling απαιτούσε περίπου 10 φορές περισσότερο χρόνο.

Τέλος, βλέποντας συνολικά τα αποτελέσματα, μπορούμε να προβούμε στην παρατήρηση ότι υπάρχουν κάποιες εικόνες στις οποίες συστηματικά — για



Σχήμα 3.16: Αριστερά: δύο εικόνες στις οποίες επιτεύχθηκε μεγάλη ομοιότητα των αποκρίσεων του υπολογιστικού μοντέλου με το πρότυπο – μέτρα ομοιότητας $c_i = 0.4017$ και 0.4014 αντίστοιχα (και τα δύο με leveling πυραμίδα). Δεξιά: εικόνες στις οποίες η καλύτερη προσέγγιση (σε αυτές τις δύο με gaussian πυραμίδα) απέτυχε – $c_i = -0.0763$ και -0.0292 .

όλες τις πυραμίδες – οι υπολογιστικοί αλγόριθμοι δίνουν πολύ κοντινά στο πρότυπο αποτελέσματα, και άλλες οι οποίες δείχνουν μεγάλες διαφορές με αυτά. Αξίζει να δούμε κάποιες από αυτές τις εικόνες (Σχήμα 3.16) και να εξηγήσουμε το φαινόμενο αυτό στη βάση της πρώιμης και προσεκτικής οπτικής προσοχής, έννοιες τις οποίες συζητήσαμε στην ενότητα 2.1.3. Στις πρώτες δύο από τις εικόνες έχουμε αποτελέσματα πολύ κοντινά στο πρότυπο. Παρατηρούμε ότι αυτές έχουν ένα αριθμό σαφώς ορισμένων περιοχών ενδιαφέροντος οι οποίες χαρακτηρίζονται από πρώιμα εξέχοντα χαρακτηριστικά – χρώμα, αντίθεση, προσανατολισμός – τα οποία προσελκύουν το βλέμμα. Στις άλλες δύο, στις οποίες υπάρχει μεγάλη διαφορά των αποτελεσμάτων των αλγορίθμων από το πρότυπο, βλέπουμε τα σημεία ενδιαφέροντος, να καθορίζονται από προσεκτικές κυρίως διαδικασίες και μη πρώιμα ερεθίσματα – π.χ., σε αυτές τις περιπτώσεις, το βλέμμα αντιλαμβάνεται την ύπαρξη ανθρώπων ή γνωστής φύσης αντικειμένων στη σκηνή και εστιάζει σε αυτούς ενώ τα πρώιμα χαρακτηριστικά, όπως η μεγάλη αντίθεση που υπάρχει στον ορίζοντα και πάνω στη γέφυρα στην τρίτη από αριστερά εικόνα ή στην λάμψη και τις χρωματικές διαφορές στην τέταρτη. Η εισχώρηση των προσεκτικών μηχανισμών οπτικής προσοχής (οι οποίοι ελάχιστα μοντελοποιούνται στις παρούσες εφαρμογές) στις ανθρώπινες αποκρίσεις ενισχύεται καθώς περνάει ο χρόνος παρατήρησης αφού, όπως είδαμε στην ενότητα 2.1.3, η πρώιμη οπτική προσοχή κυριαρχεί μόνο στα πρώτα κλάσματα του δευτερολέπτου ($< 200ms$).

3.5 Εύρεση Σημείων Ενδιαφέροντος με Χρήση Ανιχνευτή Γωνιών

Στην παράγραφο αυτή θα εξετάσουμε τη χρήση ανιχνευτών ενδιαφερόντων σημείων στην εύρεση των αξιοπροσεκτων περιοχών μιας εικόνας. Παράλληλα

θα εξετάσουμε τη σχέση της απόκρισής τους με αυτή του αλγορίθμου των Itti et al. που παρουσιάσαμε προηγουμένως - κατά πόσο ενεργές περιοχές του ανιχνευτή αντιστοιχούν ή όχι σε περιοχές με μεγάλες τιμές του χάρτη οπτικής προσοχής. Απώτερος σκοπός είναι να εξετάσουμε και κατά πόσο μπορεί να βελτιωθεί η λειτουργία του τελευταίου ενσωματώνοντας σε αυτόν έναν τέτοιο ανιχνευτή, διερεύνηση η οποία γίνεται στην ενότητα 3.5.3.

3.5.1 Ο Ανιχνευτής Harris

Ο ανιχνευτής ενδιαφερόντων σημείων - γωνιών που θα χρησιμοποιήσουμε, προτάθηκε για πρώτη φορά στο [12] από τους Harris και Stephens. Βασίζεται στην υπόθεση ότι, κοντά σε γωνίες, η φωτεινότητα της εικόνας έχει έντονες μεταβολές προς πολλαπλές κατευθύνσεις. Φιλτράροντας την αρχική εικόνα, I , με γκαουσιανά φίλτρα διαφορετικής μεταβλητότητας, σ_l^2 μπορούμε να υπολογίσουμε τις γωνίες σε διαφορετικές κλίμακες. Για το σκοπό αυτό αρχικά δημιουργούμε τη μήτρα ροπών δεύτερης τάξης

$$M = \begin{bmatrix} (L_x^i)^2 & (L_x^i)(L_y^i) \\ (L_x^i)(L_y^i) & (L_y^i)^2 \end{bmatrix}$$

όπου L_x^i και L_y^i είναι οι μερικές παραγώγοι της φιλτραρισμένης εικόνας ως εξής:

$$\begin{aligned} L_x^i(x, y; \sigma_l^2) &= \partial_x (g^i(x, y; \sigma_l^2)) * I \\ L_y^i(x, y; \sigma_l^2) &= \partial_y (g^i(x, y; \sigma_l^2)) * I \end{aligned}$$

με g^i τον δισδιάστατο Γκαουσιανό πυρήνα

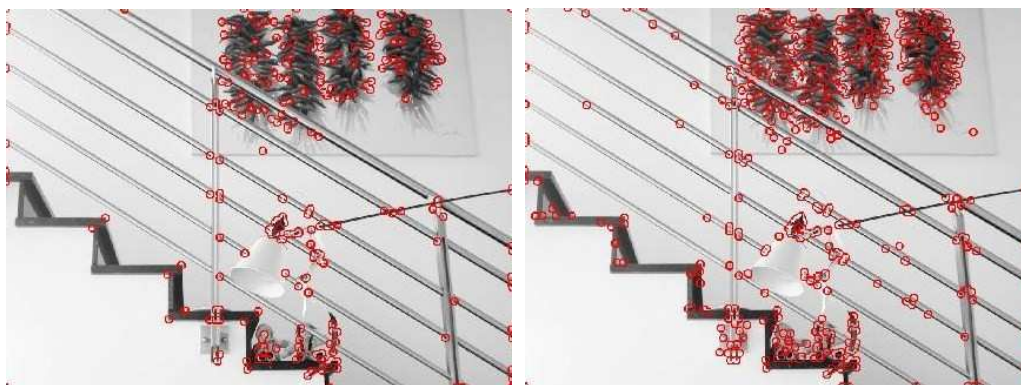
$$g^i(x, y; \sigma_l^2) = \frac{1}{2\pi\sigma_l^2} \exp(-(x^2 + y^2)/2\sigma_l^2)$$

Την ίδια τη μήτρα M μπορούμε να περάσουμε από ένα γκαουσιανό παράθυρο σε κλίμακα $\sigma_i^2 = s\sigma_l^2$, παίρνοντας τον τελικό πίνακα

$$\mu^i = g^i(x, y; \sigma_i^2) * M$$

Καθώς τώρα οι ιδιοτιμές λ_1, λ_2 , με $\lambda_1 \leq \lambda_2$, του πίνακα μ^i αντιπροσωπεύουν χαρακτηριστικές διακυμάνσεις της αρχικής εικόνας και στις δύο κατευθύνσεις της, δύο αρκετά μεγάλες τιμές των λ_1, λ_2 αποτελούν ένδειξη ύπαρξης σημείων ενδιαφέροντος. Για τον εντοπισμό τέτοιων σημείων, οι Harris και Stephens προτείνουν τον υπολογισμό των θετικών μεγίστων της συνάρτησης γωνιών

$$C_h = \det(\mu^i) - k \text{trace}^2(\mu^i) = \lambda_1\lambda_2 - k(\lambda_1 + \lambda_2)$$



Σχήμα 3.17: Αποτέλεσμα της ανίχνευσης γωνιών με τον ανιχνευτή Harris πάνω στη γκριζα εικόνα stairs. Αριστερά: με χρήση του αρχικού μέτρου των Harris και Stephens. Δεξιά: με χρήση του προτεινόμενου από την Noble στο [36] μέτρου.

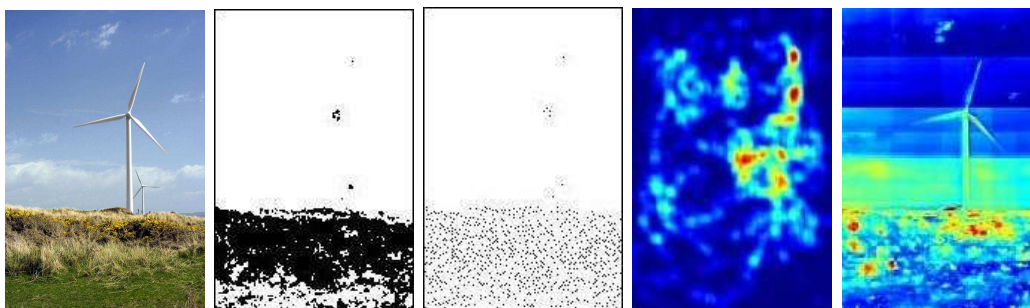
όπου k είναι μια σταθερά που εξαρτάται και από τις ιδιοτιμές. Ένα άλλο μέτρο, που αναφέρει η Noble στο [36], είναι το

$$C_n = \det(\mu^i) / \text{trace}(\mu^i) = \lambda_1 \lambda_2 / (\lambda_1 + \lambda_2)$$

με μεγάλες τιμές του πιο πάνω να αποτελούν ένδειξη ύπαρξης γωνιών. Στο Σχήμα 3.17 βλέπουμε ένα παράδειγμα των δύο αυτών υλοποιήσεων του ανιχνευτή Harris σε γκριζα εικόνα.

3.5.2 Σύγκριση Ανιχνευτή Harris και του Μοντέλου Οπτικής Προσοχής

Θα εξετάσουμε την ομοιότητα στις αποκρίσεις του ανιχνευτή γωνιών με αυτές του αρχικού μοντέλου οπτικής προσοχής. Ουσιαστικά θα ελέγξουμε δύο υποθέσεις: η πρώτη αφορά το αν υψηλές τιμές εξόδου του ανιχνευτή αντιστοιχούν σε μεγάλες τιμές οπτικής προσοχής τόσο βάσει του αρχικού μοντέλου. Η δεύτερη αφορά το αν αυτές οι τιμές αντιστοιχούν και σε “πραγματικές” τιμές μεγάλης οπτικής προσοχής, βάσει του πρότυπου που εξήγαμε με τη μέθοδο παρακολούθησης με βοήθεια ποντικιού υπολογιστή (ενότητα 3.3.1). Σημειώνουμε ότι στο σημείο αυτό δε θα ελέγξουμε την ολική ομοιότητα μεταξύ των τριών αυτών χαρτών — η υπόθεσή μας είναι εάν ισχυρές γωνίες αντιστοιχούν σε σημεία που επιστούν μεγάλη οπτική προσοχή, και όχι κατ’ ανάγκη το αντίστροφο, ότι δηλαδή κάθε ενδιαφέρουσα περιοχή αντιστοιχεί σε γωνία. Το τελευταίο είναι προφανώς λανθασμένο, καθ’ ότι υπάρχει μια πληθώρα άλλων αξιοπρόσεκτων χαρακτηριστικών, σε αντίθεση, χρώμα, προσανατολισμούς κ.ά., όπως είδαμε στην ενότητα 3.2.1.



Σχήμα 3.18: Παράδειγμα του τρόπου με τον οποίο εργαστήκαμε για τη σύγκριση των αποκρίσεων του ανιχνευτή με τις αντίστοιχες του προτύπου και του υπολογιστικού αλγορίθμου. Από αριστερά: αρχική εικόνα *turbine*· περιοχές της εικόνας με $C_h > 1\% \max\{C_h\}$ (με μαύρο χρώμα)· θετικά τοπικά μέγιστα των περιοχών αυτών· πρότυπο ανθρώπινης οπτικής προσοχής· αναμενόμενη από τον αλγόριθμο των Itti et al. οπτική προσοχή.

Πριν συνεχίσουμε, είναι αναγκαίο να προβούμε στις δύο πιο κάτω διευκρινήσεις:

- Ως σημεία με σχετική γωνιακή ένδειξη παίρνουμε αυτά στα οποία ο ανιχνευτής Harris δίνει απόκριση μεγαλύτερη από ένα ποσοστό της μέγιστης – ενδεικτικές τιμές του ποσοστού αυτού τις οποίες θα εξετάσουμε είναι 0.1%, 1%, 10% και 50% (π.χ. αρκετά μεγάλη θα ήτανε μια απόκριση τουλάχιστον 1% της μέγιστης). Βάσει της λειτουργίας του ανιχνευτή, όπως την είδαμε στην προηγούμενη ενότητα, αναφερόμαστε απλά σε περιοχές με ισχυρές μεταβολές και στις δύο διευθύνσεις.
- Ως γωνίες μιας (σχετικής) ισχύος, παίρνουμε τα σημεία τα οποία αποτελούν θετικά μέγιστα της απόκρισης του ανιχνευτή Harris, η τιμή της οποίας στο συγκεκριμένο σημείο - γωνία είναι μεγαλύτερη από ένα συγκεκριμένο ποσοστό της μέγιστης, όπως είδαμε και πιο πάνω.

Αν αναφερόμασταν μόνο στις γωνίες, θα αρκούσε μόνο το δεύτερο σημείο. Όμως επειδή στη συνέχεια θα χρησιμοποιήσουμε και απευθείας την ίδια την απόκριση - μέτρο του ανιχνευτή, παραθέτουμε και τα δύο. Η μεθοδολογία που ακολουθήσαμε είναι η εξής:

1. Σμίκρυνση του χάρτη απόκρισης γωνιών του ανιχνευτή στο μέγεθος του χάρτη οπτικής προσοχής, καθότι βρίσκονται σε διαφορετική κλίμακα. Ομοίως προβαίνουμε σε σμίκρυνση της πρότυπης απόκρισης οπτικής προσοχής, βάσει των αποτελεσμάτων που εξήχθησαν με τη μέθοδο της ενότητας 3.3.1.

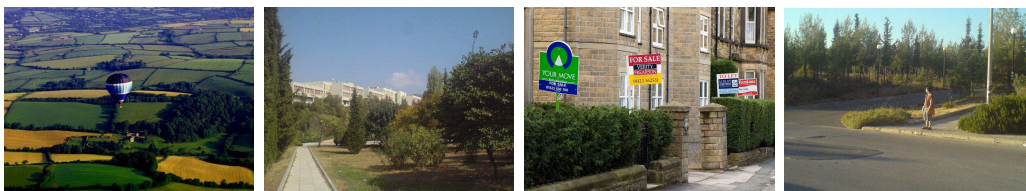
	Αποκρίσεις Μοντέλου	m_{mod}	Αποκρίσεις Προτύπου	m_{ref}
μέση τιμή οπτικής προσοχής \bar{s}	0.2552	–	0.0508	–
τιμές οπτικής προσοχής $\bar{s} \in [0, 1]$ για γωνίες (τοπικά μέγιστα ανιχνευτή με απόκριση C_h ως ποσοστό της μέγιστης):				
με $C_h > 0.1\% \max$	0.2625	1.0379	0.0816	1.5744
με $C_h > 1\% \max$	0.2792	1.1126	0.1056	2.0945
με $C_h > 10\% \max$	0.3085	1.2375	0.1398	3.1595
με $C_h > 50\% \max$	0.3484	1.4033	0.2066	5.6535
τιμές οπτικής προσοχής $\bar{s} \in [0, 1]$ για περιοχές με απλή απόκριση ανιχνευτή C_h ως ποσοστό της μέγιστης:				
με $C_h > 0.1\% \max$	0.2693	1.0662	0.0876	1.7108
με $C_h > 1\% \max$	0.2856	1.1406	0.1125	2.3128
με $C_h > 10\% \max$	0.3141	1.2552	0.1361	3.2746
με $C_h > 50\% \max$	0.3571	1.4257	0.1579	4.4053

Πίνακας 3.3: Σύγκριση των μέσων αποκρίσεων, τόσο του μοντέλου των Itti et al. όσο και του προτύπου, στις συνολικά 40 εικόνων που αναλύσαμε, για περιοχές με αυξανόμενο δείκτη ύπαρξης γωνιών. Συγκεκριμένα, παρατίθενται οι μέσες αποκρίσεις τόσο για περιοχές με μέτρο Harris αυξανόμενης έντασης (0.1, 1, 10, και 50 % του μεγίστου) όσο και για περιοχές – θετικά μέγιστα του μέτρου αυτού, με την αντίστοιχη ένταση.

2. Κανονικοποίηση των δύο χαρτών οπτικής προσοχής, αυτού που προκύπτει από το μοντέλο και του προτύπου, στην ίδια κλίμακα (π.χ. 0 - 1).
3. Επιλογή των σημείων με μια δεδομένη ισχύος γωνιακή απόκριση (πάνω από ένα ελάχιστο ποσοστό της μέγιστης), καθώς και τις αντίστοιχες γωνίες. Εύρεση της μέσης οπτικής προσοχής που αντιστοιχεί στα σημεία αυτά, βάσει των προτύπων. Υπολογισμός της μέσης οπτικής προσοχής που αντιστοιχεί σε όλα τα σημεία του χάρτη.
4. Εύρεση των αντίστοιχων ποσοτήτων βάσει του μοντέλου των Itti et al. και σύγκριση με τις προηγούμενες. Συγκεκριμένα συγκρίνουμε τους λόγους m_{ref} και m_{mod} , οι οποίοι αφορούν τις αποκρίσεις, s , των προτύπων και του μοντέλου, αντίστοιχα:

$$m_{ref} = \frac{\bar{s}_{ref, corners}}{\bar{s}_{ref, total}}$$

$$m_{mod} = \frac{\bar{s}_{mod, corners}}{\bar{s}_{mod, total}}$$



Σχήμα 3.19: Εικόνες με μεγάλους λόγους m_{ref} , για τις αποκρίσεις περιοχών με $C_h > 10\%max(C_h)$. Από αριστερά: $m_{ref} = 18.6753, 5.0117, 10.6732$ και 8.6271 .

Στο Σχήμα 3.18 βλέπουμε την εφαρμογή της μεθόδου σε μια εικόνα, μαζί με τους αντίστοιχους χάρτες οπτικής προσοχής τόσο βάσει του μοντέλου των Itti et al. όσο και του ανθρώπινου προτύπου. Κάνουμε το ίδιο και για τις 40 εικόνες που έχουμε στη διάθεσή μας, και, παίρνοντας τους μέσους όρους, συγκεντρώνουμε τα αποτελέσματά μας στον Πίνακα 3.3.

Από τα πιο πάνω φαίνεται ότι, ενώ οι ιδιαίτερα ενεργές περιοχές, σε γωνιακή απόκριση, μιας εικόνας έχουν αρκετή σχέση με την οπτική προσοχή που πέφτει στην περιοχή τους, το κλασικό μοντέλο των Itti et al. στις πλείστες περιπτώσεις δεν έχει ανάλογη απόκριση στα σημεία αυτά. Μάλιστα από τον Πίνακα 3.3 μπορούμε να δούμε ότι, όσο πιο έντονη είναι η ένδειξη ύπαρξης γωνίας, και γενικά η μεταβλητότητα της εικόνας προς αμφότερες διευθύνσεις στην περιοχή αυτή, τόσο πιο σημαντική αυτή γίνεται σε σχέση με τις γύρω περιοχές για έναν ανθρώπινο παρατηρητή. Αυτό προκύπτει από τον διαρκώς αυξανόμενο λόγο m_{ref} σε περιοχές με αυξανόμενη γωνιακή απόκριση. Εξετάζοντας και τα επί μέρους αποτελέσματα για κάθε εικόνα, βλέπουμε ότι για κάποιες εικόνες ο ανθρώπινος παράγοντας δίνει πολύ μεγαλύτερες τιμές οπτικής προσοχής στις γωνίες παρά αλλού, με λόγους m_{ref} μέχρι και 10-20 για τις αρκετά μεγάλες ($C_h > 10\%max(C_h)$). Στο Σχήμα 3.19 βλέπουμε ενδεικτικά κάποιες από αυτές τις εικόνες.

Συμπεραίνουμε, σαν κατακλείδα, ότι οι γωνίες παίζουν σημαντικό ρόλο στην οπτική προσοχή, αν και μάλλον όχι τον ίδιο σε κάθε εικόνα. Στην επόμενη ενότητα, συνεπώς, θα προσπαθήσουμε να μετασχηματίσουμε το μοντέλο ώστε η επίδραση των γωνιών να λαμβάνεται και αυτή υπ' όψιν στο σχηματισμό του χάρτη οπτικής προσοχής.

3.5.3 Διερεύνηση της Ενσωμάτωσης του Ανιχνευτή στο Μοντέλο Οπτικής Προσοχής

Με σκοπό να μελετήσουμε την επίδραση της προσθήκης του ανιχνευτή Harris στο μηχανισμό εξακρίβωσης του ενδιαφέροντος των σημείων της εικόνας

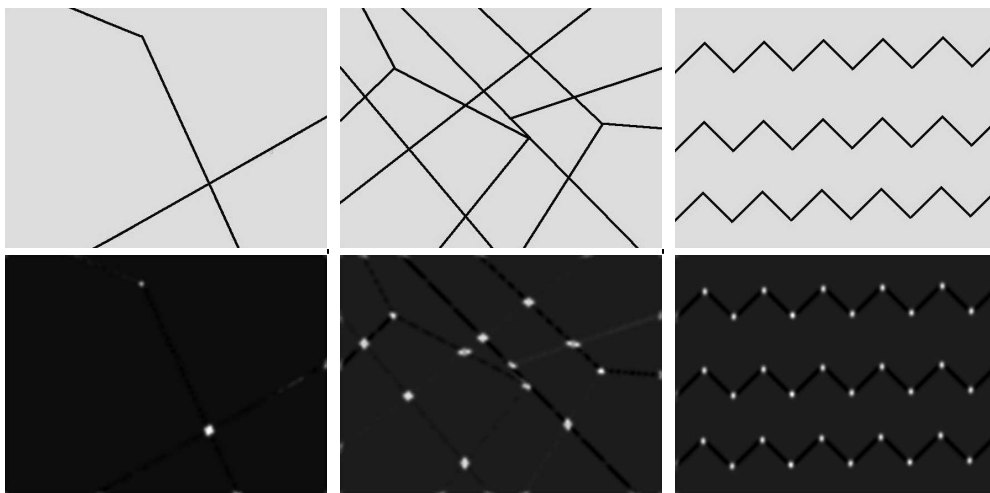
των Itti, Koch και Niebur, θα χρησιμοποιήσουμε τις τιμές του μέτρου C_h ως ένα ακόμη χάρτη χαρακτηριστικού (Conspicuity Map). Τον ονομάζουμε χάρτη ένδειξης γωνιών, έστω \bar{J} και θα τον αντιμετωπίσουμε όπως και τους τρεις αρχικούς – διαφοράς έντασης, χρωματικού ανταγωνισμού και διαφοράς προσανατολισμών – κανονικοποιώντας τον με τον τελεστή $N(\cdot)$. Αυτό θα έχει ως αποτέλεσμα ισχυρές και συγκεντρωμένες σε λίγα σημεία αποκρίσεις του ανιχνευτή να οδηγούν τον χάρτη ένδειξης γωνιών να συνεισφέρει καθοριστικά στο τελικό αποτέλεσμα, ενώ με πληθώρα ισοδύναμων αποκρίσεων ο χάρτης υποβαθμίζεται. Η επιλογή της αντιμετώπισης αυτής υπακούει στη γενική αρχή - υπόθεση των Koch και Ullman που αναφέραμε και σε προηγούμενο σημείο – ότι δηλαδή, τα όμοια χαρακτηριστικά ανταγωνίζονται μεταξύ τους για προσοχή. Σαν ένα διαισθητικό παράδειγμα της εφαρμογής της αρχής αυτής στην αντιστοίχιση των γωνιών με σημεία ενδιαφέροντος παραθέτουμε στο Σχήμα 3.20. Η μία έχει μια κύρια γωνία, η δεύτερη αρκετές μη-ισοδύναμες ενώ η τρίτη έναν αριθμό ισοδύναμων γωνιών. Ο βαθμός ενίσχυσης του τελεστή $N(\cdot)$ στην απόκριση του ανιχνευτή Harris για την κάθε μία εικόνα είναι αρκετά διαφορετικός.

Διερευνήσαμε την προσθήκη του χάρτη ένδειξης γωνιών, έστω τόσο στο αρχικό μοντέλο με χρήση Γκαουσιανής πυραμίδας, όσο και σε συνδυασμό με την υλοποίηση με τις υπόλοιπες τέσσερις πυραμίδες που περιγράψαμε στην ενότητα 2.2. Ένα παράδειγμα υλοποίησης, αντίστοιχο με αυτά που δώσαμε για την κάθε διαφορετική πυραμίδα στην ενότητα 3.4, φαίνεται στο Σχήμα 3.21. Για το αρχικό μοντέλο παραθέτουμε ξεχωριστά τις μέσες τιμές ομοιότητας μεταξύ των αποκρίσεων και του προτύπου στον Πίνακα 3.4. Βλέπουμε μια σχετικά καλύτερη προσέγγιση του μοντέλου με το πρότυπο — μάλιστα η εγγύτητα μεταξύ τους λαμβάνοντας υπόψιν τον χάρτη ένδειξης γωνιών ήταν μεγαλύτερη σε 34 από τις 40 εικόνες που εξετάσαμε, παρά με το αρχικό μοντέλο των τριών ενδιάμεσων χαρτών.

Στη συνέχεια, χρησιμοποιώντας την ίδια μεθοδολογία και στατιστική επεξεργασία όπως στην ενότητα 3.4, καταγράφουμε αναλυτικά τα αποτελέσματα στον Πίνακα 3.5. Βλέπουμε ότι η προσθήκη του ανιχνευτή Harris βελτιώνει σε κάποιο βαθμό τα αποτελέσματα.

3.6 Σχετικά με τον Τελεστή Κανονικοποίησης $N(\cdot)$

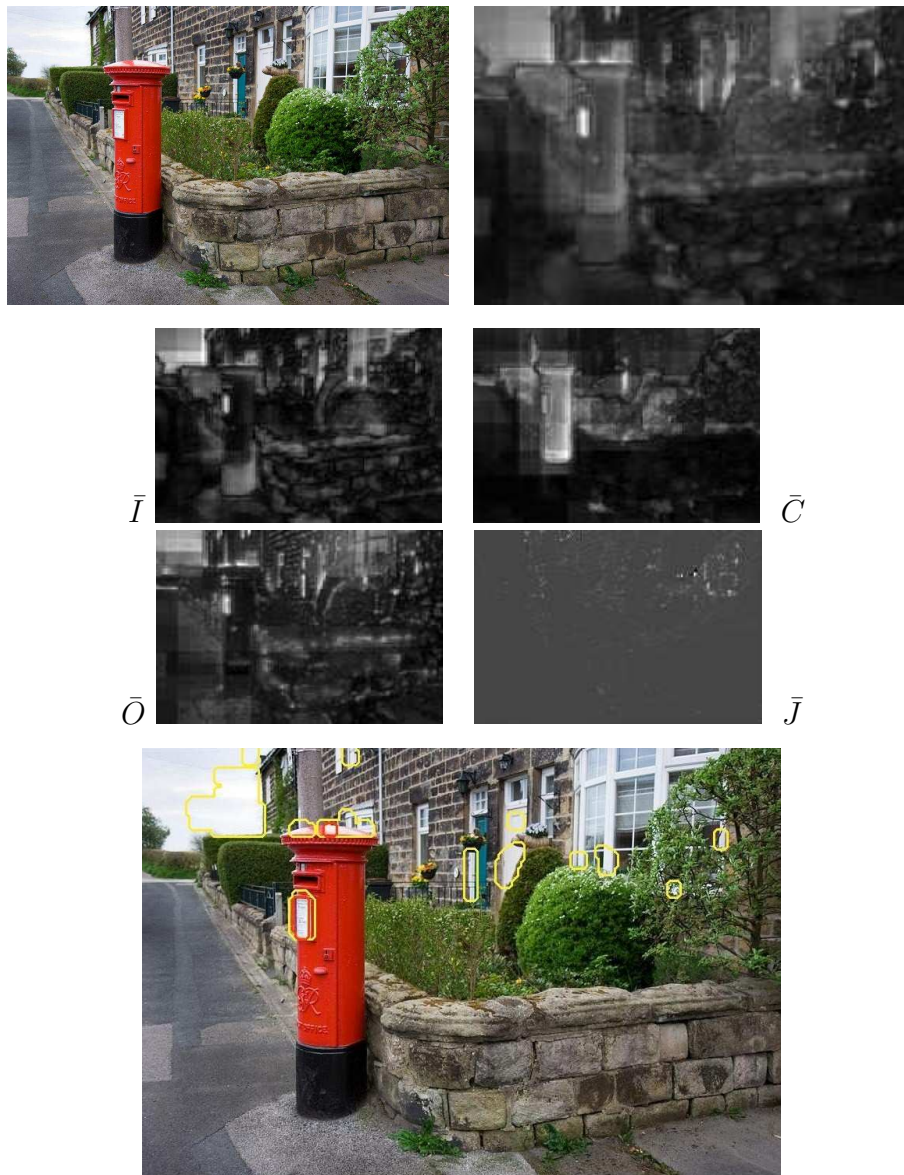
Ο υφιστάμενος τελεστής κανονικοποίησης των χαρτών χαρακτηριστικών, $N(\cdot)$, επιτρέπει αρκετά αποτελεσματικά την ανάδειξη των χαρτών με σημεία που ξεχωρίζουν. Προκύπτουν όμως και περιπτώσεις όπου ο τελεστής δεν λειτουργεί



Σχήμα 3.20: Αποτελέσματα του τελεστή $N(\cdot)$ για εικόνες με διαφορετικό αριθμό και ιεραρχική σχέση γωνιών. Πάνω – αριστερά: εικόνα με μια κύρια γωνία, κέντρο: εικόνα με αριθμό μη-ισοδύναμων γωνιών, δεξιά: εικόνα με σχεδόν ακριβώς ισοδύναμες γωνίες. Κάτω: αντίστοιχες αποκρίσεις του ανιχνευτή Harris με το αρχικό προτεινόμενο μέτρο. Ο συντελεστής πολλαπλασιασμού του τελεστή $N(\cdot)$ για κάθε εικόνα (με $M = 10$ και συνεπώς μέγιστη θεωρητική τιμή $M^2 = 100$) ήταν αντίστοιχα 70.72, 21.53 και 1.40. Ως μέγιστα θεωρήθηκαν οι τιμές που είναι οι μεγαλύτερες από την γειτονιά τους μεγέθους -8 , κατά παράγοντα $1 + 10^{-3}$, ενώ αγνοήθηκαν οι άκρες των χαρτών απόκρισης του ανιχνευτή καθ' ότι έδιναν ψηλές τιμές στα σημεία όπου οι γραμμές έβγαιναν από την εικόνα. Οι επιπλέον παραμέτροι αυτοί του τελεστή $N(\cdot)$ διερευνούνται περαιτέρω στην ενότητα 3.6.

Κατηγορία Εικόνων	Αρχικός αλγόριθμος	Προσθήκη χάρτη ένδειξης γωνιών
26 αρχικές φυσικές σχημές	0.1827	0.1922
5 φυσικές σχημές που δόθηκαν	0.1143	0.1358
3 εικόνες νυχτερινού φωτισμού	0.2883	0.2917
3 διαφημίσεις	0.0704	0.0760
3 ιστοσελίδες	0.1666	0.1714
Ολικός μέσος όρος	0.1724	0.1823
Τυπική απόκλιση	0.1196	0.1181

Πίνακας 3.4: Αποτελέσματα της αξιολόγησης για τις υλοποιήσεις του αλγόριθμου των Itti et al. με και χωρίς την ενσωμάτωση της απόκρισης Harris C_h ως τέταρτου ενδιάμεσου χάρτη οπτικής προσοχής.



Σχήμα 3.21: Επάνω - αριστερά: αρχική εικόνα *pbox*. Δεξιά: τελικός χάρτης οπτικής προσοχής ο οποίος προέκυψε υλοποιώντας τον αλγόριθμο των Itti et al. με την προσθήκη της απόκρισης του ανιχνευτή Harris ως τέταρτου ενδιάμεσου χάρτη. Μεσαίες σειρές: ενδιάμεσοι χάρτες οπτικής προσοχής \bar{I} , \bar{C} , \bar{O} και \bar{J} (αντίθεσης, χρωματικού ανταγωνισμού, προσανατολισμών και γωνιών). Κάτω: αρχική εικόνα με σημειωμένες τις περιοχές οι οποίες δίνουν μεγάλες αποκρίσεις στον τελικό χάρτη - με πορτοκαλί περίγραμμα $> 90\%$ του μεγίστου, με κίτρινο $> 64\%$ του μεγίστου. Συγκρ. με την εικόνα 3.11.

Εικόνα	G	A	M	H	Q
	Γκαουσιανή	Adjunction	Open-Close	κυμ. Haar	Leveling
<i>balloon</i>	0.1845	-0.0186	0.0671	-0.0078	0.1308
<i>cars</i>	0.2509	-0.1328	0.1371	-0.0583	0.1883
<i>cows</i>	0.1022	-0.1265	0.0741	-0.0529	0.0624
<i>drive</i>	0.1768	-0.1017	0.1126	-0.0199	0.1298
<i>entrance</i>	0.2852	-0.0330	0.1377	0.0426	0.2377
<i>gardens</i>	0.0369	-0.0911	0.0343	-0.0385	-0.0224
<i>home</i>	-0.0016	-0.1963	-0.0624	-0.1537	-0.1256
<i>horses</i>	0.3611	-0.1645	0.1802	-0.1328	0.4027
<i>mechelec</i>	0.3051	0.0609	0.3177	0.1865	0.1980
<i>old building</i>	0.1125	-0.1270	-0.0368	-0.0824	0.1193
<i>parked cars</i>	-0.0237	-0.1252	-0.0467	-0.0856	-0.1085
<i>pbox</i>	0.3377	0.2152	0.3320	0.2742	0.3156
<i>peacock</i>	0.2422	-0.0415	0.0818	-0.0186	0.2422
<i>post</i>	0.2594	-0.0424	0.1817	0.0420	0.2123
<i>punchbowl</i>	0.1915	0.0335	0.1236	0.0502	0.0950
<i>road</i>	0.2483	-0.0469	0.1532	-0.0069	0.1616
<i>salesigns</i>	0.4426	0.0706	0.3367	0.1446	0.4464
<i>ship</i>	0.3701	-0.2748	0.2616	-0.1528	0.3658
<i>sign</i>	0.2812	0.0595	0.1962	0.1237	0.1645
<i>square</i>	0.0876	-0.0034	0.0985	0.0451	0.0474
<i>stairview</i>	0.0913	-0.1961	-0.0103	-0.1369	0.0187
<i>statue - krumlov</i>	0.0092	-0.1470	-0.0756	-0.1011	-0.0320
<i>street</i>	0.2244	-0.2685	0.0874	-0.0795	0.1990
<i>turbine</i>	0.1752	0.0645	0.1437	0.1885	0.1549
<i>wait</i>	0.1285	-0.0256	0.0735	0.0066	0.0489
<i>walking</i>	0.1180	-0.0840	-0.0237	-0.0804	0.1911
5 εικ. φυσικών σχημών	0.1358	-0.0868	0.0587	-0.0660	0.0759
3 εικ. νυχτερινού φωτισμού	0.2917	0.0452	0.2362	0.1023	0.2493
3 εικ. διαφημίσεων	0.0760	0.0578	0.0794	0.1332	0.0126
3 εικ. ιστοσελίδων	0.1714	-0.0164	0.1150	0.0153	0.1905
Μέσος Όρος Αξιολόγησης $\in [-1, 1]$	0.1823	-0.0479	0.1115	0.0079	0.1395
Τυπ. Απόκλιση	0.1181	0.1194	0.1324	0.1236	0.1446

Πίνακας 3.5: Αποτελέσματα της αξιολόγησης για τις υλοποιήσεις του αλγόριθμου των Itti et al. με τις πέντε διαφορετικές πυραμίδες με την ενσωμάτωση της απόκρισης Harris C_h ως τέταρτου ενδιάμεσου χάρτη οπτικής προσοχής. Συγκρίνοντας με τον αντίστοιχο Πίνακα 3.2 βλέπουμε βελτίωση στη μεγάλη πλειοψηφία των περιπτώσεων.

όπως ακριβώς θα θέλαμε. Στην ενότητα αυτή θα εξετάσουμε διερευνητικά κάποια από αυτά τα ζητήματα και θα προτείνουμε κάποιες απλές αλλαγές και προσθήκες τις οποίες στο τέλος θα προσπαθήσουμε να αξιολογήσουμε.

Στραγγαλισμός Ολιγάριθμων Ισοδύναμων Σημείων. Από την κατασκευή του ο τελεστής $N(\cdot)$ ουσιαστικά θεωρεί μόνο ένα κύριο σημείο προσοχής σε κάθε ενδιαμέσο χάρτη. Αντιστοιχίζει, δηλαδή, την ποιότητα του χάρτη – όπου με τον όρο αυτό εννοούμε το πόσο ξεχωριστό είναι το μέγιστο σημείο προσοχής από το περιβάλλον του – στην ποιότητα αυτού του ενός και μόνο σημείου. Αυτό μπορεί να δώσει ανεπιθύμητα αποτελέσματα – για παράδειγμα, ας σκεφτούμε δύο ακραίες περιπτώσεις. Στην πρώτη, έστω ότι ο προς κανονικοποίηση ενδιαμέσος χάρτης διαθέτει ακριβώς δύο (ή κάποιο πολύ μικρό αριθμό) μέγιστα, τα οποία έχουν παραπλήσιες τιμές, ενώ βρίσκονται σε πολύ διακριτά σημεία του χάρτη - π.χ. το ένα στο αριστερό και το άλλο στο δεξιό μέρος της εικόνας. Ο μέσος όρος των υπόλοιπων τοπικών μεγίστων (εδώ δηλαδή η τιμή του δεύτερου μεγίστου) είναι πολύ μεγάλος, με αποτέλεσμα ο χάρτης να καταπνιγεί. Στη δεύτερη περίπτωση, έστω ότι τα δύο αυτά μέγιστα βρίσκονται υπερβολικά κοντά – πιθανόν να προήλθαν από το ίδιο αντικείμενο και, για τον παρατηρητή αντιστοιχούν στην ίδια αξιοπρόσεκτη περιοχή. Λειτουργώντας με ακριβώς τον ίδιο τρόπο, ο τελεστής καταπνιγεί το χάρτη. Ένα παράδειγμα που περιγράφει, εν μέρει, το φαινόμενο αυτό σε μία φυσική σκηνή δίνεται στο Σχήμα 3.22.

Στην πράξη, τέτοιες ακραίες περιπτώσεις όπως τις περιγράψαμε στην προηγούμενη παράγραφο είναι ουσιαστικά αδύνατο να συμβούν καθ' ότι το φαινόμενο αυτό μετριάζεται αρκετά από την ύπαρξη “ θορυβωδών ” περιοχών σε κάθε χάρτη οι οποίες διαθέτουν πλήθος τοπικών μεγίστων χαμηλής τιμής. Έτσι, η τιμή της ενίσχυσης καθορίζεται σε μεγάλο βαθμό από από τα μικρά αυτά τοπικά μέγιστα. Αυτό αποτελεί ένα δεύτερο ζήτημα, το οποίο αναφερόμαστε ως **Καθορισμός Παράγοντα Βάρους από Δευτερεύοντα Χαρακτηριστικά**. αφού έχει ως αποτέλεσμα οι τιμές της ενίσχυσης στις πλείστες φυσικές εικόνες να εμφανίζουν σχετικά μικρή διασπορά μεταξύ τους, όπως βλέπουμε στο Σχήμα 3.22 αλλά και από το ότι στις 26 εικόνες οι οποίες προέρχονται από φυσικές σκηνές (αναλυτικά το σύνολο των εικόνων παρουσιάζεται στο παράρτημα ;;), οι τυπικές αποκλίσεις των πολλαπλασιαστικών παραγόντων των τριών ειδών χαρτών \bar{I} , \bar{C} και \bar{O} ήταν 6.77, 8.51 και 8.07 με αντίστοιχους μέσους όρους 56.87, 61.56 και 57.04. Έτσι η ποιότητα του κάθε ενδιαμέσου χάρτη οφείλεται κατά πολύ στο βαθμό στον οποίο αυτός χαρακτηρίζεται από τέτοιες περιοχές με πολλά και χαμηλής τιμής τοπικά μέγιστα. Πολύ συχνά, αυτά μπορεί να προέρχονται από χαρακτηριστικά υψής στην αρχική εικόνα και όχι από αξιοπρόσεκτα (salient) αντικείμενα.

Κάποιες ιδέες για επίλυση ή μετριασμό αυτών των δύο ζητημάτων που αναφέραμε θα μπορούσαν να είναι οι εξής:



Σχήμα 3.22: Παράδειγμα του φαινομένου του στραγγαλισμού (σχετικά) ισοδύναμων σημείων. Πάνω – αριστερά: αρχική εικόνα, δεξιά: τελικός χάρτης οπτικής προσοχής. Κάτω βλέπουμε τους τρεις ενδιάμεσους χάρτες, κατά σειρά: διαφορών έντασης I , χρωματικού ανταγωνισμού C και διαφορών προσανατολισμού O . Τα τρία κύρια αντικείμενα του χάρτη γίνονται αντιληπτά στο χάρτη C αλλά η συνεισφορά του στον κύριο χάρτη προκύπτει σχεδόν ισοδύναμη με αυτή των υπολοίπων (συντελεστές πολλαπλασιασμού 0 – 100: 58.45, 77.30 και 60.68 αντίστοιχα).

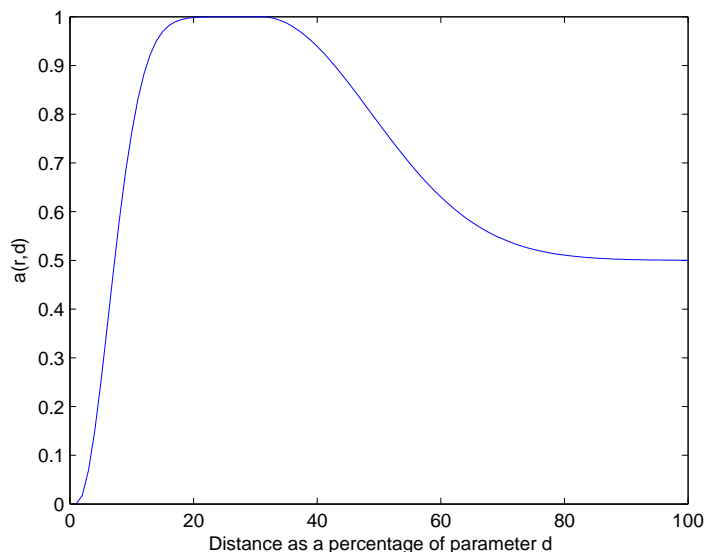
Σχετικά με το στραγγαλισμό ισοδύναμων σημείων σε πολύ κοντινές ή πολύ μακρινές αποστάσεις στο χάρτη, οι οποίες δε θα θέλαμε να θεωρούνται ανταγωνιστικά σημεία στην κανονικοποίηση του χάρτη, μπορούμε να λαμβάνουμε υπ' όψιν και την απόσταση του κάθε τοπικού μεγίστου από το ολικό. Μια υλοποίηση αυτής της ιδέας θα ήταν, αντί να χρησιμοποιήσουμε ως παράγοντα κανονικοποίησης τον αρχικό

$$w = (M - \bar{m})^2 = (M - (1/n) \sum_{i=1}^n m_i)^2$$

να χρησιμοποιούσαμε ένα παράγοντα της μορφής

$$w^R = (M - \bar{m})^2 = (M - (1/n) \sum_{i=1}^n \alpha(r_i^M, d) \times m_i)^2$$

όπου m_i είναι η τιμή του κάθε τοπικού μεγίστου και r_i^M η απόστασή του από το ολικό. Η $\alpha(r_i^M, d)$ είναι μια συνάρτηση που καθορίζεται από τη θέση του τοπικού μεγίστου ως προς το ολικό και μια χαρακτηριστική απόσταση, d , της εικόνας – αυτή μπορεί να είναι π.χ. το μήκος της ελάχιστης της διάστασης ή της διαγωνίου της. Ένα σκαρίφημα μιας τέτοιας συνάρτησης δίνουμε στο Σχήμα



Σχήμα 3.23: Μορφή της συνάρτησης $\alpha(r, d)$ όπως περιγράφεται στην εξίσωση 3.2.

3.23. Το σχεπτικό είναι ότι, αφ' ενός αγνοούμε τα πολύ κοντινά τοπικά μέγιστα και επίσης αγνοούμε, σε κάποιο βαθμό, τα τοπικά μέγιστα που βρίσκονται πολύ μακριά. Τα τελευταία αποφεύγουμε να αγνοήσουμε πλήρως καθ' ότι έτσι ο παραλλαγμένος τελεστής θα μπορούσε να ενισχύσει κατά υπερβολικό βαθμό χάρτες με πολυπληθείς ομάδες ισοδύναμων τοπικών μεγίστων με σχετικά ψηλές τιμές, μακριά από το ολικό μέγιστο: αυτό θα είχε τον κίνδυνο να αποκρυβούν στοιχεία των περιοχών αυτών που προέρχονται από τους υπόλοιπους χάρτες. Η ενδεικτική συνάρτηση του Σχήματος 3.23, την οποία θα χρησιμοποιήσουμε στην σύντομη μελέτη μας για το θέμα είναι η

$$\alpha(r, d) = \left\{ \begin{array}{ll} 1 - \exp[-(r^2)/(0.075d)^2], & r < 0.3d \\ 0.5 + 0.5\exp[-(r - 0.3d)^2/(0.25d)^2], & r \geq 0.3d \end{array} \right\} \quad (3.2)$$

Θέτωντας ως διάσταση d το μήκος της διαγωνίου, έχουμε ότι $r \in (0, d)$. Χρησιμοποιώντας τις τυπικές παραμέτρους του μοντέλου των Itti et al. (γχαουσισιανές πυραμίδες, μη χρήση ανιχνευτή γωνιών κλπ.) και αλλάζοντας απλά τον τελεστή $N(\cdot)$ κατά τον πιο πάνω τρόπο, εξετάσαμε τις αποκρίσεις για τις 40 εικόνες που είχαμε διαθέσιμες. Είδαμε ότι, σαν μέσος όρος, η εγγύτητα των αποκρίσεων με αυτές των προτύπων δεν μεταβλήθηκε. Εφ' όσον είδαμε ότι δεν αλλάζει τη μέση απόδοση του αλγορίθμου, απαιτείται να εξεταστεί η αλλαγή αυτή του τελεστή $N(\cdot)$ πάνω σε ένα προσεκτικά επιλεγμένο σύνολο εικόνων που εμφανίζουν το πρόβλημα του στραγγαλισμού χαρτών με ισοδύναμα αξιοπρόσεκ-



Σχήμα 3.24: Γειτονίες μεγέθους - 4 και μεγέθους - 8 για το κεντρικό εικονοστοιχείο ο.

τα σημεία. Μείωση του μεγάλου αριθμού των ανεπιθύμητων χαμηλής τιμής τοπικών ελαχίστων μπορεί να επιτευχθεί κάνοντας πιο δύσκολο το χαρακτηρισμό ενός σημείου ως τοπικό ελάχιστο. Η αρχική υλοποίηση του μοντέλου των Itti et al. χρησιμοποιεί γειτονιά μεγέθους-4 για προσδιορισμό των μεγίστων – ζητούμε δηλαδή το εικονοστοιχείο να είναι το μέγιστο μέσα στην αντίστοιχη γειτονιά – ενώ ένα πιο απαιτητικό κριτήριο θα ήταν η χρήση γειτονιάς μεγέθους-8 (Σχήμα 3.24). Επίσης, αντί απλά να απαιτούμε το υποψήφιο τοπικό ελάχιστο να είναι απλά το μέγιστο μέσα στη γειτονιά του, μπορούμε να απαιτήσουμε να είναι μεγαλύτερο κατά ένα κατώφλι $1 + \beta$ από τα γύρω του. Με προσεκτική επιλογή της παραμέτρου β μπορούμε εύκολα να διαλέξουμε τοπικά ελάχιστα τα οποία είναι αρκετά ισχυρά ώστε να είμαστε βέβαιοι ότι δεν προέκυψαν τυχαία, όπως π.χ. σε μια περιοχή που αντιστοιχεί σε ελαφριά υφή της αρχικής εικόνας. Η χαρακτηριστική τιμή που χρησιμοποιήσαμε στον έλεγχο της ιδέας αυτής είναι $\beta = 10^{-3}$. Ένα παράδειγμα όπου χρησιμοποιείται το κριτήριο αυτό είναι στην έξοδο ενός ανιχνευτή Harris στο Σχήμα ;; (όπως είδαμε στην ενότητα 3.5.1 ο ανιχνευτής αυτός καταλήγει να αναζητά τα τοπικά μέγιστα μιας συνάρτησης).

Πέρα από τη χρήση του τελεστή $N(\cdot)$, υπάρχουν επίσης πολλές άλλες τεχνικές που υλοποιούνται για την συγκρότηση του τελικού χάρτη οπτικής προσοχής από τους επί μέρους ενδιάμεσους χάρτες. Για παράδειγμα, οι ίδιοι οι δημιουργοί του αρχικού αλγορίθμου πρότειναν μετέπειτα ως εναλλακτική λύση την επαναληπτική άθροιση μεταξύ κάθε χάρτη χαρακτηριστικών και του χάρτη διαφορών - των - Γκαουσιανών - στην εικόνα 3.25 δίνουμε ένα τέτοιο



Σχήμα 3.25: Χάρτες οπτικής προσοχής της αρχικής εικόνας *balloon* (αριστερά) χρησιμοποιώντας άθροιση με τον τελεστή N (κέντρο) και επαναληπτική άθροιση (δεξιά). Τα αποτελέσματα λήφθηκαν με τη χρήση του Saliency Toolbox.

παράδειγμα. Άλλες τεχνικές άθροισης των επι μέρους χαρτών αφορούν ανάλυση κύριων παραγόντων (principal component analysis), πιθανόν ανάλυση κυρίαρχων παραγόντων (dominant component analysis) κ.ά. .

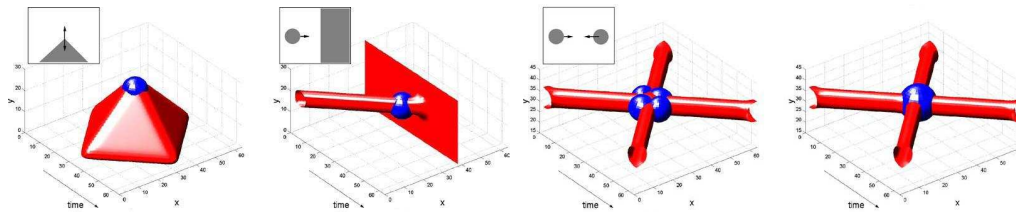
Κεφάλαιο 4

Μοντέλα Χωροχρονικής Οπτικής Προσοχής

4.1 Ανοδικά Χωροχρονικά Μοντέλα

Η εύρεση ενδιαφέροντων χωρο-χρονικών (αντί απλά χωρικών, όπως είδαμε στις δισδιάστατες εικόνες) σημείων και περιοχών αποτελεί, όπως είδαμε στο κεφάλαιο 1 μια από τις πιο συναρπαστικές εφαρμογές και συνάμα προκλήσεις της Όρασης Υπολογιστών, ενώ παράλληλα βρίσκει μια πληθώρα εφαρμογών όπως συντόμευση - περίληψη ακολουθιών, παρακολούθηση κ.ά. . Εξετάζοντας αντίστοιχες μεθόδους που υπάρχουν για δισδιάστατες εικόνες, οδηγούμαστε στο να διερευνήσουμε το βαθμό και τον τρόπο με τον οποίο είναι δυνατή η επέκτασή τους ώστε να συμπεριλάβουν και τη διάσταση του χρόνου. Στις προσεγγίσεις αυτές, αντιμετωπίζουμε το χρόνο σαν μια τρίτη διάσταση με αποτέλεσμα να παριστάνουμε την ακολουθία σαν ένα στερεό στον τρισδιάστατο Ευκλείδειο χώρο.

Μια από τις πρώτες τέτοιες εφαρμογές ήταν η επέκταση του χωρικού ανιχνευτή γωνιών Harris στις τρεις διαστάσεις από τον Laptev [23]. Η ιδέα αυτή βασιζόταν στο ότι πολλά ενδιαφέροντα γεγονότα μέσα σε ακολουθίες χαρακτηρίζονται από ισχυρές μεταβολές των δεδομένων τόσο στις χωρικές όσο και στις χρονικές διαστάσεις – π.χ. η είσοδος ενός ατόμου στο δωμάτιο ή η σύγκρουση ενός αυτοκινήτου – και θα επισήμαινε σημεία τα οποία αντιστοιχούσαν σε στιγμές μη-σταθερής κίνησης. Τέτοια σημεία αντιστοιχούν συνήθως σε επιταχυνόμενες δομές μέσα στην εικόνα και έτσι αρκετά πιθανόν να αντιστοιχούν και σε επιταχυνόμενα αντικείμενα του πραγματικού κόσμου. Μια σειρά από απλά παραδείγματα, παρμένη από τη σχετική εργασία, φαίνεται στο Σχήμα 4.1. Χρησιμοποιώντας την περιγραφή αυτή είναι δυνατός ο χαρακτηρισμός και κατηγοριοποίηση γεγονότων – το οποίο είναι και το πιο συνηθισμένο κριτήριο



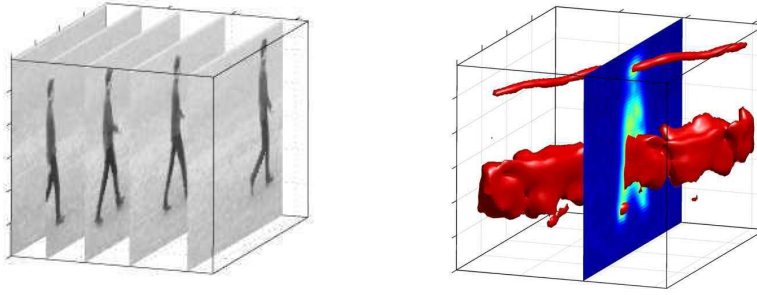
Σχήμα 4.1: Τέσσερα παραδείγματα εντοπισμού χωρο-χρονικών σημείων ενδιαφέροντος από απλές συνθετικές ακολουθίες εικόνων για την επέκταση του ανιχνευτή Harris από τον Laptev. Από αριστερά: μια κινούμενη γωνία, η ένωση μιας μπάλας και ενός τοίχου, η σύγκρουση δύο μπάλων με σημεία ενδιαφέροντος εντοπισμένα σε σχετικά λεπτή κλίμακα και, τέλος, το ίδιο με το προηγούμενο αλλά με τα σημεία ενδιαφέροντος εντοπισμένα σε πιο αδρή κλίμακα. Από το [23].

αξιολόγησης της κάθε μεθόδου.

Η αρχιτεκτονική των Itti et al., την οποία περιγράψαμε στην ενότητα 3.2, είναι δυνατόν και αυτή να επεκταθεί στις τρεις διαστάσεις. Η επέκταση αυτή είναι μάλλον ξεκάθαρη στους τομείς της αντίθεσης και του χρωματικού ανταγωνισμού - δημιουργούμε τρισδιάστατες πυραμίδες τις οποίες φιλτράρουμε με της ίδιας διάστασης Γκαουσιανούς πυρήνες. Κάποιες διευκρινήσεις είναι απαραίτητες στον τομέα της εύρεσης των προσανατολισμών. Στο χωροχρόνο οι προσανατολισμοί αυτοί σχετίζονται με την κίνηση που επιτελείται στην ακολουθία [23]. Προσανατολισμός π.χ. μιας χωροχρονικής δομής κατά μια γωνία ως προς τον χώρο αντιστοιχεί σε κίνηση, όπως βλέπουμε και στα δείγματα του Σχήματος 4.2.

4.1.1 Χωροχρονική Επέκταση του Μοντέλου των Itti, Koch και Niebur από τους Rapantzikos et al.

Στο ;; επιτελείται στην επέκταση του μοντέλου των Itti et al. με κάποιες διαφοροποιήσεις. Αποδομεί αρχικά τον κύριο όγκο στους αρχικούς όγκους χαρακτηριστικών, τους F_1, R, G, B, Y και F_3 από τους οποίους ο πρώτος αντιστοιχεί στην φωτεινότητα, οι επόμενοι τέσσερις στο κάθε ένα από τα χρωματικά κανάλια κόκκινου, πράσινου, μπλε και κίτρινου που περιγράψαμε στην ενότητα 3.2 ενώ ο τελευταίος αντιστοιχεί γενικά στους προσανατολισμούς. Η εξαγωγή του τελευταίου - η οποία είναι και η κύρια διαφορά με το αρχικό μοντέλο, στο οποίο υπολογίζονται ξεχωριστά ένας αρχικός χάρτης προσανατολισμού για κάθε γωνία - γίνεται χρησιμοποιώντας ανάλυση κύριων παραγόντων (principal



Σχήμα 4.2: Αριστερά: διαδοχικές τομές ακολουθίας που εμπεριέχει κίνηση. Δεξιά: ισομετρική επιφάνεια η οποία εμπερικλείει τις σημαντικότερες περιοχές της ακολουθίας, βάσει ενός μοντέλου από το [2]. Βλέπουμε τον τρισδιάστατο προσανατολισμό της κινούμενης δομής κατά μια γωνία ως προς το χώρο. Εικόνες από το [2].

components analysis) για την συγχώνευση των χωροχρονικών όγκων που προκύπτουν από τις αποκρίσεις για κάθε χωρο-χρονική γωνία, αντί ξεχωριστή κανονικοποίησή τους. Οι αποκρίσεις αυτές προκύπτουν με τη χρήση καθοδηγήσιμων φίλτρων (steerable filters) τα οποία προκύπτουν από το γραμμικό συνδυασμό ενός συνόλου φίλτρων βάσης, οδηγούμενα από δύο γωνίες θ και ϕ . Οι γωνίες αυτές ορίζουν τρεις δυνατές κατευθύνσεις α, β και γ σε σφαιρικές συντεταγμένες όπως πιο κάτω:

$$\begin{aligned}\alpha &= \cos(\theta)\sin(\phi) \\ \beta &= \sin(\theta)\sin(\phi) \\ \gamma &= \cos(\phi)\end{aligned}$$

Βάσει αυτών υπολογίζεται η προσανατολισμένη ενέργεια $E_{3d}(\theta, \phi)$ σε κάθε συνδυασμό γωνιών (στο ;; επιλέγονται 20 τέτοιοι συνδυασμοί). Ο όγκος προσανατολισμών F_3 , όπως αναφέραμε, συγκροτείται μέσω ανάλυσης σε κύριους παράγοντες, δηλαδή

$$F_3 = \Phi_{PCA}[E_{3d}(\theta, \phi)]$$

Στη συνέχεια υπολογίζονται οι όγκοι διαφοράς κέντρου c - περιφέρειας s , $F_{i,c/s}, i = 1, 2, 3$ σχηματίζοντας τις αντίστοιχες Γκαουσιανές πυραμίδες όπως και στον αλγόριθμο των Itti et al., δηλαδή

$$\begin{aligned}F_{1,c/s} &= |F_{1,c} \ominus F_{1,s}| \\ F_{2,c/s} &= |(R_c - G_c) \ominus (G_s - R_s)| + |(B_c - Y_c) \ominus (Y_s - B_s)| \\ F_{3,c/s} &= |F_{3,c} \ominus F_{3,s}|\end{aligned}$$

Παρατηρούμε επίσης τη συγχώνευση των δύο χαρτών χρωματικού ανταγωνισμού σε ένα. Οι όγκοι διαφορών κέντρου - περιφέρειας που προκύπτουν κανονικοποιούνται με τον τελεστή $N(\cdot)$ και συνθέτουν τους τρεις ενδιάμεσους χάρτες (conspicuity maps) αντίθεσης, χρωματικού ανταγωνισμού και προσανατολισμών, οι οποίοι, όπως και στο δισδιάστατο μοντέλο, κανονικοποιούνται ξεχωριστά και προστίθενται σε έναν κύριο χάρτη - όγκο οπτικής προσοχής (saliency map).

4.1.2 Απευθείας Χωροχρονική Επέκταση του Μοντέλου των Itti et al.

Στην παράγραφο αυτή προτείνουμε μια εναλλακτική επέκταση της αρχιτεκτονικής των Itti et al. η οποία ακολουθεί πλήρως – εξ' ου και ο χαρακτηρισμός “ απευθείας ” – όσο είναι δυνατόν, την αντίστοιχη που υπάρχει στις δύο διαστάσεις. Αποδομούμε αρχικά την ακολουθία στους όγκους (σε αντιστοιχία με τους χάρτες) I, R, G, B και Y , από τους οποίους δημιουργούμε τις αντίστοιχες τρισδιάστατες Γκαουσιανές πυραμίδες. Οι όγκοι κέντρου c - περιφέρειας s είναι συνεπώς οι

$$I(c, s) = |I(c) - I(s)|$$

και

$$\begin{aligned} RG(c, s) &= |(R(c) - G(c)) \ominus (G(s) - R(s))| \\ BY(c, s) &= |(B(c) - Y(c)) \ominus (Y(s) - B(s))| \end{aligned}$$

Για τους όγκους προσανατολισμού, πρέπει να λάβουμε υπ' όψιν περισσότερες χωρο-χρονικές κατευθύνσεις από ότι στις δύο διαστάσεις. Εργαζόμενοι με σφαιρικές συντεταγμένες, προσθέτουμε μια ακόμη γωνία ϕ μαζί με τη θ ώστε να σχηματίσουμε τρισδιάστατους πυρήνες Gabor $5 \times 5 \times 5$ με τους οποίους θα φιλτράρουμε την Λαπλασιανή πυραμίδα του όγκου φωτεινότητας I . Η γενική μορφή των πυρήνων αυτών θα είναι

$$\psi_{\theta, \phi} = LPF[m_{\theta, \phi}]$$

όπου $m_{\theta, \phi}$ είναι ένα μιγαδικό ημιτονοειδές προσανατολισμένο στις γωνίες θ και ϕ ως εξής:

$$m_{\theta, \phi}(x, y) = e^{j(\pi/2)(\cos\theta\sin\phi \times x + \sin\theta\sin\phi \times y + \cos\phi \times t)}$$

Με τους πυρήνες Gabor έχουν συνεπώς τη μορφή

$$\psi_{\theta, \phi}(x, y, t) = \frac{1}{2\pi} e^{(j\pi/2)(x^2/\sigma_x^2 + y^2/\sigma_y^2 + t^2/\sigma_t^2)} \times m_{\theta, \phi}$$

Χρησιμοποιούμε τέτοια φίλτρα στις γωνίες $\theta, \phi = 0^\circ, 45^\circ, 90^\circ$ και 135° , σε ένα σύνολο $4 \times 4 = 16$ δυνατών προσανατολισμών. Οι όγκοι των χαρακτηριστικών προσανατολισμού, οι οποίοι εκφράζουν τη διαφορά του προσανατολισμού μεταξύ των κλιμάκων κέντρων και περιφερειών υλοποιούνται ως εξής:

$$O(c, s, \theta, \phi) = |O(c, \theta, \phi) \ominus O(s, \theta, \phi)|$$

Διατηρώντας τις ίδιες κλίμακες κέντρου και περιφέρειας που χρησιμοποιεί ο αρχικός, δισδιάστατος αλγόριθμος, $c \in 2, 3, 4$ και $s = c + d, d \in 3, 4$, προκύπτουν 96 όγκοι για τους προσανατολισμούς, οι οποίοι μαζί με άλλους 6 για την ένταση και 12 για το χρωματικό ανταγωνισμό απαρτίζουν ένα τελικό σύνολο 114 όγκων.

Στη συνέχεια συγκροτούμε τους ενδιάμεσους όγκους οπτικής προσοχής \bar{I}, \bar{C} και \bar{O} , χρησιμοποιώντας τον τρισδιάστατο πλέον τελεστή $N(\cdot)$ με αντίστοιχο τρόπο με τον δισδιάστατο αλγόριθμο. Σημειώνουμε ότι εδώ ο τελεστής \otimes συμβολίζει παρεμβολή όλων των όγκων σε μια συγκεκριμένη κλίμακα (συνήθως 4) και πρόσθεση ογκοστοιχείο - προς - ογκοστοιχείο (voxel):

$$\begin{aligned} \bar{I} &= \otimes_{c=2}^4 \otimes_{s=c+3}^{c+4} N(I(c, s)) \\ \bar{C} &= \otimes_{c=2}^4 \otimes_{s=c+3}^{c+4} [N(RG(c, s)) + N(BY(c, s))] \end{aligned}$$

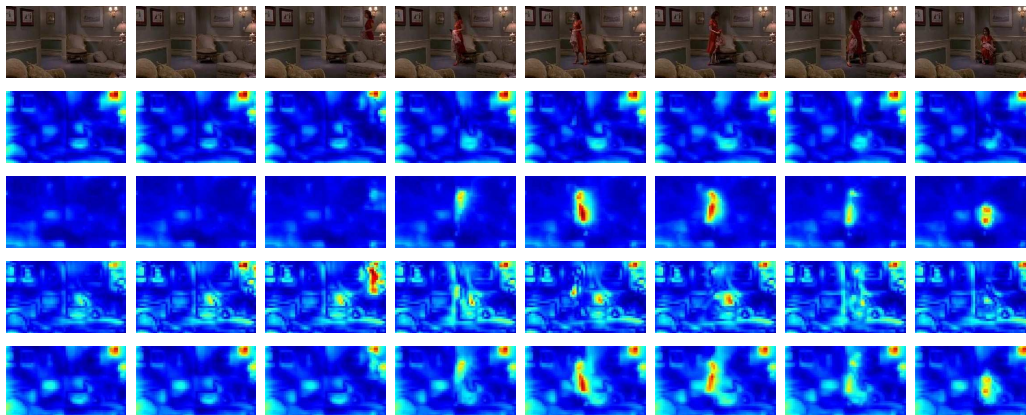
Για τους προσανατολισμούς δημιουργούνται τώρα αρχικά δεκαέξι ενδιάμεσοι όγκοι συνδυάζοντας τους έξι χάρτες χαρακτηριστικών που αντιστοιχούν στον ίδιο συνδυασμό γωνιών θ και στη συνέχεια συνδυάζονται σε ένα μοναδικό ενδιάμεσο όγκο προσανατολισμού:

$$\bar{O} = \sum_{\theta, \phi \in \{0^\circ, 45^\circ, 90^\circ, 135^\circ\}} N \left(\otimes_{c=2}^4 \otimes_{s=c+3}^{c+4} N(O(c, s, \theta, \phi)) \right)$$

Οι τρεις ενδιάμεσοι όγκοι συνδυάζονται στον τελικό όγκο οπτικής προσοχής (saliency volume) S ακριβώς όπως και στο δισδιάστατο μοντέλο, δηλαδή

$$S = \frac{1}{3} (N(\bar{I}) + N(\bar{C}) + N(\bar{O}))$$

Μια ενδιαφέρουσα συνέχεια της επέκτασης αυτής θα ήταν να τροφοδοτήσουμε τον προκύπτοντα όγκο οπτικής προσοχής σε ένα όλα - για - το - νικητή (winner-take-all) νεωρωνικό δίκτυο, αντίστοιχο με αυτό που περιγράφεται στη δισδιάστατη υλοποίηση. Με διαρκή τροφοδότηση του δικτύου από τον όγκο προσοχής και προβαίνοντας σε απαγόρευση των περιοχών των προηγούμενων αξιοπρόσεκτων χωροχρονικών σημείων κ.ο.κ. (βλ. την εκτενή περιγραφή του δισδιάστατου δικτύου στην ενότητα 3.2.2) αυτό θα κατέληγε στη δημιουργία μιας ιεραρχίας σημείων εστίασης μέσα στην ακολουθία. Η αλυσίδα σημείων ενδιαφέροντος που προκύπτει θα μπορούσε, μεταξύ άλλων, να χρησιμοποιηθεί



Σχήμα 4.3: Πάνω: πλαίσια από μέρος της ταινίας 'As good as it gets'. Από τη δημόσια διαθέσιμη βάση ακολουθιών του [25]. Η ακολουθία που εξετάσαμε περιλάμβανε 382 πλαίσια μεγέθους 133×222 . Δεύτερη σειρά: αντίστοιχες τομές του ενδιάμεσου χάρτη - όγκου αντίθεσης (contrast). Τρίτη σειρά: αντίστοιχες τομές για τον ενδιάμεσο όγκο χρωματικού ανταγωνισμού. Τέταρτη σειρά: τομές για τον ενδιάμεσο όγκο προσανατολισμών, στον οποίο γίνεται αντιληπτή η είσοδος στη σκηνή. Τελευταία σειρά: τομές του τελικού όγκου οπτικής προσοχής.

και απευθείας ακόμα για τη δημιουργία περίληψης της ακολουθίας — αν και τέτοια υλοποίηση δεν έχει δοκιμαστεί.

Στα πλαίσια της διπλωματικής εργασίας, υλοποιήσαμε την τρισδιάστατη εκδοχή του αλγόριθμου των Itti et al. η οποία περιγράφεται στην παρούσα παράγραφο. Στην εικόνα 4.3 βλέπουμε τα αποτελέσματα που προκύπτουν για μια ακολουθία παρμένη από την ταινία As Good As It Gets. Παρατηρούμε ότι οι όγκοι χαρακτηριστικών, όντας ευαίσθητοι και στις χρονικές μεταβολές, ενεργοποιούνται περισσότερο από κινούμενα αντικείμενα. Το κύριο μέρος της κίνησης, και ιδιαίτερα απότομων μεταβολών όπως π.χ. είσοδος και έξοδος ενός ατόμου από τη σκηνή, αντικατοπτρίζεται στον όγκο προσανατολισμών.

4.1.3 Χρήσεις, Περιορισμοί και Προβλήματα

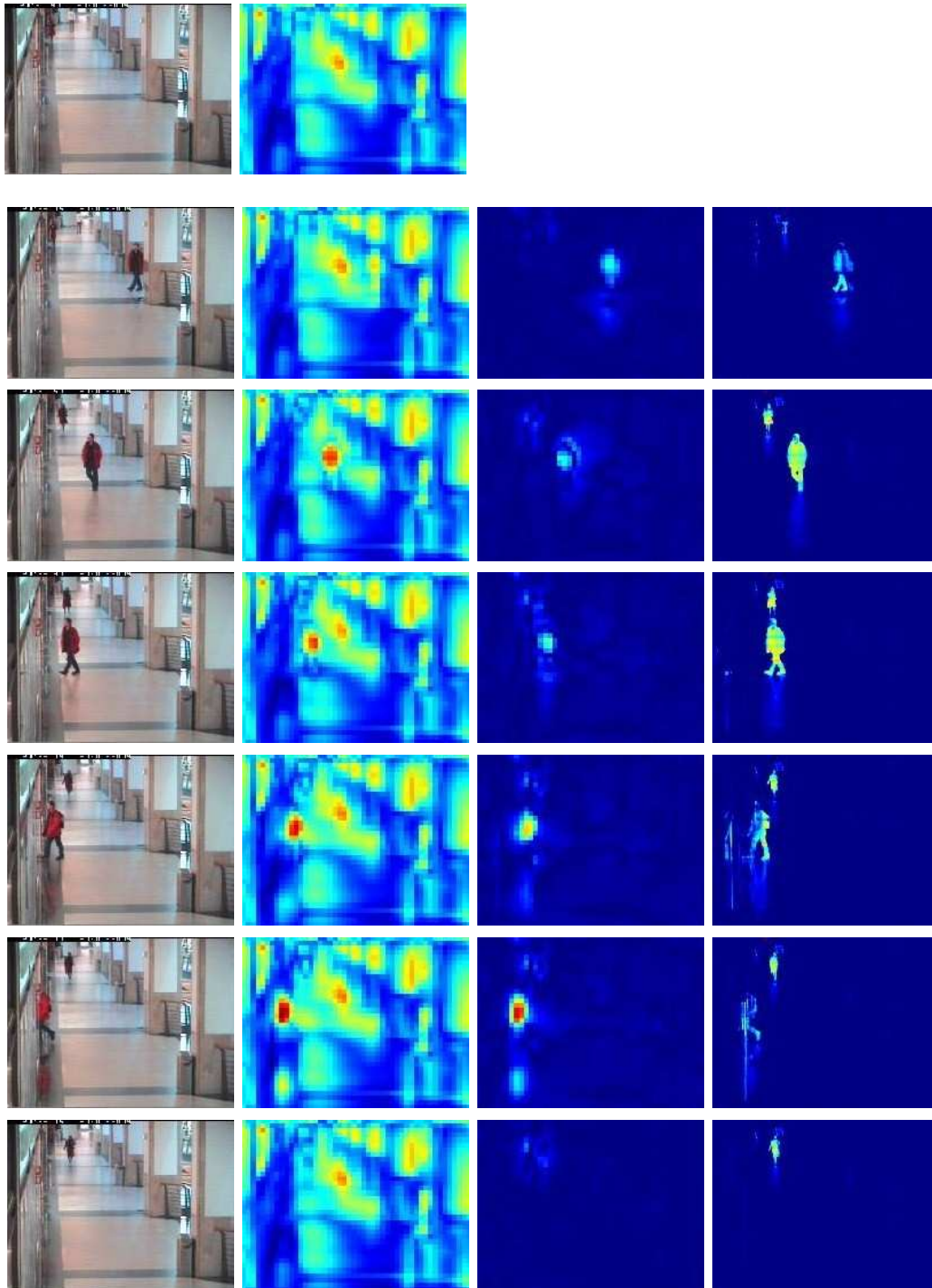
Τα χωροχρονικά μοντέλα οπτικής προσοχής χρησιμοποιούνται πολύ συχνά στην κατηγοριοποίηση ανθρώπινων πράξεων, κάτι στο οποίο έχει γίνει εκτενής δουλειά στα [2], [6] και [23]. Συνήθως εξάγεται ένας αριθμός αντιπροσωπευτικών χωροχρονικών σημείων για κάθε ακολουθία μαζί με τα χαρακτηριστικά των περιοχών γύρω από αυτά και στη συνέχεια συγκροτείται ένα αντίστοιχο διάνυσμα χαρακτηριστικών. Οι διάφορες ανθρώπινες δραστηριότητες ταξινομούνται στη συνέχεια στη βάση των διανυσμάτων αυτών, με μέτρο την

ευκλείδιός τους απόσταση από κάποια πρότυπα διανύσματα που αντιστοιχούν σε κάθε μια δραστηριότητα. Τα πρότυπα προκύπτουν από ένα σύνολο εκπαίδευσης — περισσότερα στο [2].

Μια άλλη χρήση είναι η σύντηξη της πληροφορίας των ακολουθιών, κρατώντας τις σημαντικές χωροχρονικές περιοχές και υποβαθμίζοντας τις υπόλοιπες, ή, σε επίπεδο πλαισίων, διατήρηση μόνο αυτών με τη σημαντικότερη πληροφορία — δηλαδή, η δημιουργία περιλήψεων. Ένας άλλος τομέας, ο οποίος ίσως αξίζει να διερευνηθεί περισσότερο, είναι η χρήση των χωροχρονικών μοντέλων σε εφαρμογές παρακολούθησης. Προς το τέλος της εργασίας αυτής, δοκιμάσαμε τη χρήση της επέκτασης του μοντέλου των Itti et al. πάνω σε ακολουθίες από τη δημόσια διαθέσιμη βάση CAVIAR, οι οποίες αποτελούν τις εξόδους από στατικές κάμερες σε ένα εμπορικό κέντρο στη Λισαβόνα ή στην είσοδο των εργαστηρίων INRIA στην Γκρενόμπλ. Ξεκινώντας από ένα αρχικό καρέ με την εικόνα του άδειου χώρου, υπολογίζουμε προς σύγκριση, δύο διαφορετικές ποσότητες: τη διαφορά των ίδιων των πλαισίων (σε γκρίζα μορφή) με το αρχικό — η οποία είναι η απλούστερη αλλά αρκετά αποτελεσματική μέθοδος ανίχνευσης της κίνησης — και την αντίστοιχη διαφορά των τομών του χάρτη - όγκου οπτικής προσοχής που αντιστοιχούν στις ίδιες χρονικές στιγμές. Ένα παράδειγμα εφαρμογής της διαδικασίας αυτής φαίνεται στο Σχήμα 4.4. Το απλό μοντέλο της διαφοράς του τρέχοντος από το αρχικό πλαίσιο βρίσκει εύκολα τα κινούμενα αντικείμενα, όμως, η διαφορά των τομών της οπτικής προσοχής δίνει πολύ μεγάλες αποκρίσεις στα γεγονότα όπως π.χ. εδώ η στιγμή στην οποία ο άνθρωπος στο κέντρο μπαίνει στην πόρτα και φεύγει από τη σκηνή. Στην αδυναμία του συστήματος να κρατήσει με την ίδια ακρίβεια το περίγραμμα των αντικειμένων πρέπει βέβαια να λάβουμε υπ' όψιν το ότι λειτουργεί σε υποδειγματολημμένες παραστάσεις της αρχικής ακολουθίας, λόγω απαιτήσεων χώρου στη μνήμη.

Περιορισμοί των Επεκτάσεων

Όπως αναφέρεται και στο [2], η χωροχρονική ανάλυση των γεγονότων μιας ακολουθίας έχει νόημα μόνο όταν αυτά εξελίσσονται σχετικά ομαλά μέσα σε αυτήν. Απουσιάζουν δηλαδή απότομες μεταβολές οι οποίες οφείλονται σε αλλαγή σκηνής ή γωνίας λήψης της κάμερας. Απαιτείται λοιπόν πρώτα η κατάτμηση της ακολουθίας σε σκηνές και πλάνα. Υπάρχουν διάφορες τεχνικές ανίχνευσης σκηνών μέσα σε μια ακολουθία οι οποίες έχουν αναπτυχθεί από διάφορες ερευνητικές ομάδες. Μια τεχνική είναι η χρήση των κατανομών των χρωμάτων ή της φωτεινότητας σε αντίστοιχα ιστογράμματα, με τις απότομες κορυφώσεις της διαφορών τους από πλαίσιο σε πλαίσιο να αντιστοιχούν σε αλλαγή σκηνής. Οι κορυφώσεις ανιχνεύονται θέτοντας ένα εμπειρικό κατώφλι στις τιμές. Για παράδειγμα, στο [42], χρησιμοποιώντας συγκριτικά ιστογράμματα



Σχήμα 4.4: Χρήση της επέκτασης του μοντέλου των Itti et al. στις τρεις διαστάσεις. Πρώτη σειρά: αρχικό πλαίσιο και τομή του όγκου οπτικής προσοχής στη στιγμή αυτή. Επόμενες σειρές, από αριστερά: πλαίσια κάθε χρονικής στιγμής· τομές του όγκου οπτικής προσοχής για την αντίστοιχη χρονική στιγμή· (απόλυτη) διαφορά της τομής από την αρχική· (απόλυτη) διαφορά του τρέχοντος πλαισίου από το αρχικό. Παρουσιάζεται ακολουθία από τη βάση LISBON.

χρωματικών διαφορών, έγινε δυνατή η επίτευξη ακρίβειας 91% στην κατάτμηση σε σκληρές πάνω σε ένα σύνολο ακολουθιών ολικής διάρκειας περίπου 5 ώρες.

Προβλήματα των Επεκτάσεων

Ένα από τα κύρια προβλήματα τα οποία αφορούν τη χρήση τέτοιων τρισδιάστατων μοντέλων είναι η ανάγκη αποθήκευσης ενός πολύ μεγάλου, ακόμα και βάσει των δυνατοτήτων των σύγχρονων υπολογιστών, όγκου δεδομένων. Για παράδειγμα, θεωρούμε μια ταινία ενός λεπτού με σχετικά μικρό μέγεθος πλαισίων, π.χ. 200×300 και τυπικό ρυθμό 30 πλαίσια ανά δευτερόλεπτο, με την κάθε τιμή να αποθηκεύεται σε μορφή *UTF-8* (1Byte). Η αποθήκευση της ταινίας θα απαιτεί 108MB στη μνήμη ταχείας προσπέλασης (RAM) για κάθε κανάλι, και κάτι περισσότερο για κάθε μια από τις αντίστοιχες πυραμίδες. Η αποθήκευση κάθε μιας από τους ενός από τους αρχικούς όγκους - χάρτες αποκρίσεων κέντρου - περιφέρειας, ο οποίος βρίσκεται στην κλίμακα 2, σε αριθμούς διπλής ακρίβειας, θα χρειάζεται τον ίδιο αριθμό — έχουμε 32 τέτοιους χάρτες για τους προσανατολισμούς (οι υπόλοιποι 64 αντιστοιχούν σε κλίμακες 3 και 4). Βλέπουμε δηλαδή, αν και όχι αδύνατο να εκπληρωθούν, οι απαιτήσεις σε χώρο της μνήμης ταχείας προσπέλασης είναι αρκετά μεγάλες. Απαιτείται λοιπόν βελτιστοποίηση της υλοποίησης των αλγορίθμων ως προς το χώρο που δεσμεύουν, πράγμα που συνήθως γίνεται σε βάρος του χρόνου εκτέλεσης. Στην παρούσα φάση παρακάμψαμε το πρόβλημα αυτό εξετάζοντας μόνο μικρού μεγέθους ακολουθίες, επιλέγοντας ακολουθίες μικρής διάρκειας ή υποδειγματοληπτώντας τις. Ένα ενδεικτικό μέγεθος που χρησιμοποιήσαμε π.χ. για την ακολουθία LISBON (Σχήμα 4.4) είναι $144 \times 192 \times 300$.

4.1.4 Περαιτέρω Επεκτάσεις

Στην ενότητα 3.5.3 είδαμε την επέκταση του δισδιάστατου μοντέλου των Itti et al. με την προσθήκη ενός ανιχνευτή γωνιών Harris. Μια δυνατή προσθήκη όσον αφορά το χωροχρονικό μοντέλο θα ήταν η προσθήκη ενός τρισδιάστατου ανιχνευτή Harris, ο οποίος δεν είναι παρά ο ανιχνευτής που προτείνουν οι Laptev et al. στο 3.4. Το γεγονός ότι ο τελευταίος ήδη χρησιμοποιείται για εύρεση χωροχρονικών σημείων ενδιαφέροντος μας προϋποθέτει ότι και οι δύο μηχανισμοί μαζί θα αποτελέσουν ένα ισχυρότερο εργαλείο από ότι ο κάθε ένας ξεχωριστά.

Μια άλλη ιδέα θα ήταν η υλοποίηση του μοντέλου με μορφολογικές τρισδιάστατες πυραμίδες, σε αντιστοιχία με την ενότητα 3.4. Αυτό βέβαια προϋποθέτει την ύπαρξη τρισδιάστατων μορφολογικών τελεστών, υλοποίηση η οποία πρέπει να κάνει κάποιος εξ' αρχής καθ' ότι δεν υπάρχει αντίστοιχη οικογένεια συναρτήσεων στο MATLAB. Η διατήρηση των τρισδιάστατων περιγραμμάτων - επι-

φαινοίων η οποία γίνεται δυνατή με αυτούς τους τελεστές μπορεί να οδηγήσει σε νέες δυνατότητες, όπως π.χ. η περιγραφή μιας πράξης βάσει του τρισδιάστατου σχήματος που αντιστοιχεί σε αυτήν. Στα αδύνατα σημεία της μεθόδου αυτής θα πρέπει βέβαια να καταλογιστεί και η απαιτούμενη υπολογιστική πολυπλοκότητα αν, απαιτώντας βέλτιστη διατήρηση του σχήματος, χρησιμοποιήσουμε leveling πυραμίδα — η ανάλυση σε τρεις αντί δύο διαστάσεις θα αυξήσει περαιτέρω το κόστος αυτό.

4.2 Εντοπισμός Ενδιαφερόντων Πλαισίων σε Ακολουθίες Βάσει της Χωρικής Οπτικής Προσοχής και της Κίνησης

Ο εντοπισμός ενδιαφερόντων πλαισίων μέσα σε ακολουθίες (video) βρίσκει μεγάλες εφαρμογές στην σύντμηση και δημιουργία περιλήψεων των ακολουθιών αυτών, όπως είδαμε και στο κεφάλαιο 1. Συνήθως επιθυμούμε να κρατήσουμε τα πλαίσια τα οποία συγκρατούν το μεγαλύτερο ποσοστό της πληροφορίας, απορρίπτοντας πλαίσια τα οποία δεν προσθέτουν κάτι νέο στην ιδέα που έχουμε για το περιεχόμενο της ακολουθίας. Δραστικές εναλλαγές, όπως έντονες κινήσεις, είσοδος και έξοδος προσώπων και αντικειμένων στην εικόνα, απότομες εκλάμψεις συνιστούν οπτικά χαρακτηριστικά τα οποία θεωρούμε, γενικά, ότι εμπεριέχουν ενδιαφέρον μέσα σε μια ακολουθία εικόνων.

Υπάρχουν βέβαια και πολλά στοιχεία τα οποία είναι αδύνατον να λάβει υπ' όψιν μια καθαρά οπτική αντίληψη. Γεγονότα που αφορούν αλλαγές στο ηχητικό κομμάτι της εικόνας, έναρξη και παύση μιας ομιλίας, κρότοι, διαρκώς αυξανόμενη ένταση του ήχου κλπ. μπορούν να αναλυθούν μέσω υπολογιστικών μοντέλων πάνω στο ηχητικό σήμα όπως για παράδειγμα γίνεται στο [8]. Στοιχεία που απαιτούν την εννοιολογική (semantic) κατανόηση της ακολουθίας είναι αδύνατον να βρεθούν μόνο με έναν υπολογιστικό αλγόριθμο σαν αυτούς που χρησιμοποιεί η επεξεργασία σήματος και εικόνας. Κάποιο μέρος όμως της διαδικασίας επιλογής των πλαισίων μπορεί να αυτοματοποιηθεί μέσω της υλοποίησης κάποιων ευριστικών κανόνων που προκύπτουν από την εμπειρία της “χειροκίνητης” κατασκευής περιλήψεων από ακολουθίες. Ένα τέτοιο σύνολο ευριστικών κανόνων, το οποίο έχει καταγραφεί σε συνεργασία με επαγγελματίες παραγωγούς και τεχνικούς, περιγράφεται στο [42] και περιέχει κυρίως ιδέες πάνω στις συνήθεις τακτικές σχηνοθεσίας που χρησιμοποιούνται στη δημιουργία των βίντεο. Σαν μια γενική ιδέα, κανόνες που αναφέρονται είναι π.χ. ότι η κίνηση αντικειμένων είναι αρκετά σημαντική ή ότι υψηλής έντασης ήχος συνήθως προηγείται σημαντικών σχημών, ενώ χαμηλής έντασης μουσική συνήθως αποτελεί μεταβατικό εργαλείο το οποίο σπάνια συνοδεύει σημαντικά

πλαίσια.

Στο κεφάλαιο αυτό ασχολούμαστε μόνο με το οπτικό κομμάτι της δημιουργίας περιλήψεων των ακολουθιών. Συγκεκριμένα, θα παρουσιάσουμε διάφορες μεθοδολογίες οι οποίες αποδίδουν ένα μέτρο σημαντικότητας σε κάθε πλαίσιο της ακολουθίας και θα προβούμε σε συγκρίσεις μεταξύ τους. Στη συνέχεια, θα εξηγήσουμε πώς τα μέτρα αυτά μας βοηθούν στην προσπάθεια να δημιουργήσουμε περιλήψεις ακολουθιών.

4.2.1 Φιλοσοφία του Αλγορίθμου

Αναζητούμε σημεία της ακολουθίας τα οποία μας επιφέρουν έκπληξη: αν διαφορετικά συμπεριφέρονταν με τρόπο που θα μπορούσαμε να προβλέψουμε, σημαίνει ότι ήδη έχουμε συγκεντρώσει αρκετή πληροφορία ώστε να καλύψουμε το οπτικό (και, γενικότερα, νοηματικό) τους περιεχόμενο. Θέλουμε δηλαδή να βρούμε τα πλαίσια στα οποία καινούργια στοιχεία εισάγονται στην εικόνα που έχουμε για το περιεχόμενο της ακολουθίας.

Εφ' όσον γνωρίζουμε τα χαρακτηριστικά της παρόντος πλαισίου, μαζί με το προφίλ της κίνησης που έχει τελεστεί από το προηγούμενο πλαίσιο, μπορούμε να έχουμε μια καλή ιδέα του πώς μπορεί να μοιάζει το επόμενο πλαίσιο. Όπως μαθαίνουμε από το Γυμνάσιο, σύμφωνα με τον πρώτο νόμο του Νεύτωνα, την αρχή της αδράνειας, έχουμε

Κάθε σώμα, που βρίσκεται μέσα σε ένα αδρανειακό σύστημα, διατηρεί την κατάσταση ηρεμίας, ή ευθύγραμμης και ομαλής κίνησής του, εφόσον καμία εξωτερική δύναμη δεν επιδρά για τη μεταβολή της.

Συνεπώς, εφ' όσον δεν εξασκούνται δυνάμεις στη σκηνή που περιγράφουν τα πλαίσια της ακολουθίας, τα επόμενα καρέ τείνουν να γίνουν όλο και πιο προβλέψιμα. Οι δυνάμεις είναι αυτές που εισάγουν κατά κύριο λόγο πληροφορία μέσα στη σκηνή, και μεγαλύτερη αποτυχία της πρόβλεψής μας αυτής σημαίνει ότι όντως κάτι σημαντικό και αξιοπρόσεκτο έχει τελεστεί. Το σταμάτημα ή η εκκίνηση ενός αντικειμένου, η είσοδος και η έξοδος του από την εικόνα, η απότομη επιτάχυνση ή η σύγκρουσή του με κάποιο άλλο αποτελούν συμβάντα τα οποία θα δώσουν μεγάλες αποκλίσεις από αυτό που αναμένουμε και συνεπώς μεγάλες τιμές ενδιαφέροντος στα πλαίσια τα οποία τα περιλαμβάνουν. Στην ενότητα 4.1 είδαμε ότι σε πολλές περιπτώσεις, οι χωροχρονικές επεκτάσεις των μοντέλων οπτικής προσοχής μπορούν να συλλάβουν τις κινήσεις αυτές. Εδώ θα εξετάσουμε το θέμα από την προοπτική της εύρεσης της οπτικής προσοχής σε κάθε πλαίσιο ξεχωριστά, και, βάσει αυτής, σε συνδυασμό την κίνηση της πραγματικής εικόνας, εύρεση της αναμενόμενης μορφής του χάρτη οπτικής προσοχής του επόμενου πλαισίου. Όσο πιο πολύ διαφέρει η πρόβλεψη αυτή από την πραγματικότητα, τόσο περισσότερο μπορούμε να πούμε ότι έχει εισαχθεί

από το επόμενο πλαίσιο ένα στοιχείο που μας προκαλεί έκπληξη. Η έκπληξη αυτή προσδίδει σημαντικότητα στο γεγονός.

Πέρα από την χρήση του προφίλ της κίνησης, μπορούμε να βασιστούμε και σε άλλες μεθόδους πρόβλεψης της επικείμενης εικόνας, όπως λ.χ. γραμμική πρόβλεψη ή κάποιου είδους στατιστική επεξεργασία βάσει των κατανομών των πλαισίων που προηγήθηκαν. Σε κάθε περίπτωση, το ενδιαφέρον κάθε πλαισίου θα είναι αντιστρόφως ανάλογο της επιτυχίας με την οποία αυτό έχει προβλεφθεί βάσει των προηγούμενων.

4.2.2 Διάφορες Μεθοδολογίες Αξιολόγησης των Αλγορίθμων

Η αξιολόγηση της επιλογής των κυριότερων πλαισίων στις ακολουθίες, και γενικά κάποιων περιλήψεών τους, δεν είναι ξεκάθαρη υπόθεση λόγω της υποκειμενικότητας που εμπεριέχει. Λόγω αυτού, δεν υπάρχει μια τυποποιημένη και ευρέως δεκτή μέθοδος υπολογισμού της ποιότητας μιας περίληψης βίντεο (video skim). Αυτοματοποιημένες μέθοδοι αξιολόγησης, μέσω προγραμμάτων ή προσομοιώσεων, είναι πολύ δύσκολο να δώσουν ακριβή αποτελέσματα, καθ' όσον είναι (προς το παρόν τουλάχιστον) ασύμβατες με την ανθρώπινη αντίληψη.

Εξ' αιτίας των πιο πάνω, η πιο συνηθισμένη μέθοδος είναι η αξιολόγηση από τους χρήστες με σχετικές μελέτες (user-study), τεχνική πολύ συνηθισμένη στον έλεγχο και προώθηση προϊόντων κ.α. Κοινός παρονομαστής των μελετών αυτών είναι η συγκέντρωση ενός ικανού αριθμού ατόμων τα οποία θα αξιολογήσουν τις περιλήψεις. Το [29] περιγράφει δύο τέτοια κριτήρια αξιολόγησης. Το σημαντικότερο είναι το κατά πόσο το συντομευμένο βίντεο εμπεριέχει αρκετή πληροφορία (informativeness). Ένα δεύτερο κριτήριο είναι το κατά πόσο αυτό παραμένει ευχάριστο στην παρακολούθηση όπως και το αρχικό (enjoyability). Η ευχαρίστηση που μπορεί να παρέχει η περίληψη εξαρτάται από την ομαλότητα (smoothness) της ακολουθίας, καθώς και από ηχητικά χαρακτηριστικά, όπως η ομιλία και η μουσική. Η προτεινόμενη προσέγγιση είναι η παρουσίαση στον κάθε χρήστη πρώτα των περιλήψεων και στη συνέχεια της αρχικής ακολουθίας - ταινίας, ώστε να μην γνωρίζει εξ' αρχής το πλήρες περιεχόμενο. Το αντίστροφο είναι και αυτό δυνατό. Η μεθοδολογία που περιγράφεται πιο πάνω χρησιμοποιήθηκε και σε ανάλογο πρόγραμμα του Εθνικού Μετσόβιου Πολυτεχνείου [8].

Στην πιο πάνω μέθοδο αξιολόγησης έχουμε να προτείνουμε και μια άλλη προσθήκη ως μέτρο σύγκρισης. Πολλές φορές οι παρατηρητές δίνουν απαντήσεις οι οποίες παρουσιάζουν αρκετές διαφορές στις περιοχές τιμών μεταξύ τους, ενώ δεν είναι σαφές τι σημαίνει "μηδέν" πληροφοριακό περιεχόμενο ή "πλήρες" για τον κάθε ένα. Για την περαιτέρω εξέταση του θέματος, και

με σκοπό τη σύγκριση μεταξύ αποκρίσεων οι οποίες εμπεριέχουν ασάφεια, προτείνουμε μια ιδέα εμπνευσμένη από τις κλινικές μελέτες που διεξάγει η ιατρική. Αναφερόμαστε στο γνωστό *placebo* – ελληνική μετάφραση: φαρμαχείκελο – που χρησιμοποιείται κάποιες φορές στην αξιολόγηση νέων φαρμακευτικών σκευασμάτων: από το σύνολο των ατόμων στα οποία ελέγχεται η επίδραση του νέου σκευάσματος, σε κάποια από αυτά χορηγείται κάποιο ουδέτερο στοιχείο ή ένα παλαιότερο σκεύασμα. Έτσι αποκτείνεται μια καλύτερη ιδέα για την αποτελεσματικότητα του νέου φαρμάκου. Σε αναλογία στις μελέτες αξιολόγησης των περιλήψεων, ένα τέτοιο *placebo* αποτελεί η εισαγωγή ως ακολουθιών προς αξιολόγηση περιλήψεων οι οποίες προκύπτουν από τυχαία επιλογή των περιεχόμενων πλαισίων. Για παράδειγμα, αν θέλουμε να συγκρίνουμε με μια περίληψη με μέγεθος $1/n$ της ακολουθίας, μπορούμε να παίρνουμε ένα πλαίσιο κάθε n ή να παίρνουμε όλα τα πλαίσια με πιθανότητα $1/n$ το καθένα.

Ένα εναλλακτικό σχήμα αξιολόγησης, το οποίο αναφέρεται στο [42] αποτελεί η εξέταση της αποτελεσματικότητας των περιλήψεων ως εργαλείο περιήγησης μέσα σε μια βιβλιοθήκη από ακολουθίες βίντεο. Σε αυτή την περίπτωση, οι χρήστες καλούνται να περιηγηθούν μέσα στη βιβλιοθήκη αυτή και να απαντήσουν μια σειρά ερωτήσεων. Η αποτελεσματικότητα της κάθε περιήγησης βασίζεται στο χρόνο που απαιτήθηκε για την ολοκλήρωση της εργασίας αυτής και στον αριθμό των ορθών απαντήσεων.

Στην παρούσα εργασία θα παραθέσουμε κάποια ποιοτικά αποτελέσματα και συμπεράσματα. Η διεξοδική αξιολόγηση των σχημάτων περιήγησης που θα προτείνουμε μπορεί να αποτελέσει μελλοντική εργασία. Ποιοτικά, όταν παίρνουμε σχετικά μικρές και σαφούς περιεχομένου ακολουθίες, μια καλή περιήγηση μπορούμε να πούμε ότι είναι αυτή η οποία εμπεριέχει όλα τα σημαντικά γεγονότα, όπως είναι, για παράδειγμα

- Είσοδος και έξοδος από την εικόνα
- Εκκίνηση ή στάση
- Απότομες κινήσεις
- Επιτάχυνση ή επιβράδυνση, αλλαγή φοράς κίνησης κλπ.

Σαν γενικός κανόνας επιλογής των πλαισίων μπορεί να θεωρηθεί ότι επιθυμούμε να επιλέγονται πλαίσια βάσει των οποίων μπορούμε να προβούμε σε πλήρη νοηματική περιγραφή της κάθε δραστηριότητας. Για παράδειγμα, μια περιγραφή της ακολουθίας του σχήματος 4.8 θα ήταν “ μια κυρία μπαίνει στη σκηνή, προχωρεί ως την πόρτα ενώ στη συνέχεια επιστρέφει στην καρέκλα και κάθεται, ανασηκώνοντας την κουβέρτα που κρατάει ”.

4.3 Αλγόριθμοι Εύρεσης Ενδιαφέροντος Κάθε Πλαίσιο

Θα παρουσιάσουμε και θα συγκρίνουμε μεταξύ τους πέντε διαφορετικές προσεγγίσεις, από πιο απλές – ως σημείο αναφοράς – ως πιο σύνθετες με τις οποίες αποδίδουμε ένα δείκτη ενδιαφέροντος σε κάθε πλαίσιο της ακολουθίας. Καθ' ότι σε κάποιες από αυτές κάνουμε χρήση της οπτικής ροής μεταξύ διαδοχικών πλαισίων, είναι σκόπιμο πρώτα να περιγράψουμε τη μέθοδο που χρησιμοποιήσαμε για το προσδιορισμό της.

4.3.1 Προσδιορισμός Οπτικής Ροής Μεταξύ Πλαισίων

Η μέθοδος που χρησιμοποιήσαμε πρόκειται για έναν από τους πλέον γνωστούς, πετυχημένους αλλά και απλούς αλγόριθμους υπολογισμού της οπτικής ροής και προτάθηκε το 1981 από τους Lucas και Kanade στο [28].

Ο αλγόριθμος συνίσταται στον προσδιορισμό του πεδίου οπτικής ροής $-\vec{d}(\vec{x}) = (d_x, d_y)$, έτσι ώστε οι διαδοχικές εικόνες να αντιστοιχίζονται ως

$$I_n(\vec{x}) \approx I_{n-1}(\vec{x} + \vec{d})$$

Θεωρώντας ότι το $-\vec{d}$ είναι σταθερό σε μια μικρή περιοχή - παράθυρο γύρω από το σημείο, η μέθοδος επιδιώκει να ελαχιστοποιήσει το τετραγωνικό σφάλμα

$$J_{\vec{x}}(\vec{d}) = \int_{\vec{x}' \in \mathbb{R}^2} G_{\rho}(\vec{x} - \vec{x}') \left(I_n(\vec{x}') - I_{n-1}(\vec{x}' + \vec{d}) \right)^2 d\vec{x}' \quad (4.1)$$

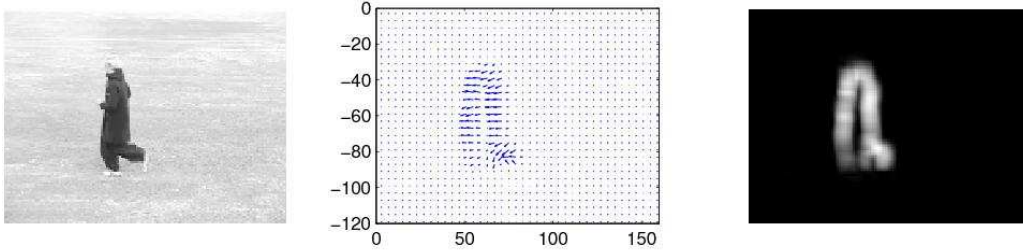
Ξεκινώντας από μια αρχική εκτίμηση, έστω \vec{d}_i , προσπαθούμε να βρούμε μια καλύτερη, έστω την $\vec{d}_{i+1} = \vec{d}_i + \vec{u}_i$, όπου \vec{u}_i αποτελεί τη βελτίωση που κάνουμε στο συγκεκριμένο βήμα. Αναπτύσσοντας κατά Taylor το $I_{n-1}(\vec{x} + \vec{d})$ γύρω από το $\vec{x} + \vec{d}_i$ βρίσκουμε

$$I_{n-1}(\vec{x} + \vec{d}) = I_{n-1}(\vec{x} + \vec{d}_i + \vec{u}_i) \approx I_{n-1}(\vec{x} + \vec{d}_i) + u_i^T \nabla I_{n-1}(\vec{x} + \vec{d}_i)$$

Αντικαθιστώντας το πιο πάνω στην εξίσωση 4.1 μπορεί ναδειχθεί ότι η λύση ελάχιστων τετραγώνων που βελτιώνει την προσέγγιση του πεδίου οπτικής ροής σε κάθε σημείο είναι

$$\vec{u}_i(\vec{x}) = \begin{bmatrix} (G_{\rho} * w_1^2)(\vec{x}) + \mu & G_{\rho} * (w_1 w_2)(\vec{x}) \\ G_{\rho} * (w_1 w_2)(\vec{x}) & (G_{\rho} * w_2^2)(\vec{x}) + \mu \end{bmatrix}^{-1} \cdot \begin{bmatrix} (G_{\rho} * (w_1 z))(\vec{x}) \\ G_{\rho} * (w_2 z)(\vec{x}) \end{bmatrix} \quad (4.2)$$

όπου



Σχήμα 4.5: Αποτέλεσμα του αλγορίθμου των Lukas - Kanade μεταξύ των πλαισίων 47 και 48 της ακολουθίας jogging. Αριστερά: πλαίσιο 47 της ακολουθίας. Κέντρο: κίνηση προς το επόμενο πλαίσιο. Δεξιά: μέτρο της κίνησης, κανονικοποιημένο στη μέγιστή του τιμή.

$$W(\vec{x}) = \begin{bmatrix} w_1(\vec{x}) & w_2(\vec{x}) \end{bmatrix} = \begin{bmatrix} \frac{\partial I_{n-1}}{\partial x}(\vec{x} + \vec{d}_i) & \frac{\partial I_{n-1}}{\partial y}(\vec{x} + \vec{d}_i) \end{bmatrix}$$

$$z(\vec{x}) = I_n(\vec{x}) - I_{n-1}(\vec{x} + \vec{d}_i)$$

Ο επαναπροσδιορισμός του πεδίου οπτικής ροής επαναλαμβάνεται ωστόσο πετύχουμε ικανοποιητική σύγκλιση, συνήθως με ένα κριτήριο του τύπου

$$|\vec{u}_i| < \epsilon$$

το οποίο διακόπτει τις επαναλήψεις όταν η διαφορά μεταξύ των διαδοχικών προσεγγίσεων γίνει μικρότερη από ϵ . Στο Σχήμα 4.5 βλέπουμε το αποτέλεσμα του αλγορίθμου πάνω σε δύο διαδοχικά πλαίσια της ακολουθίας jogging από τη δημόσια διαθέσιμη βάση ακολουθιών του ΚΤΗ που διατίθεται στο διαδίκτυο [22].

Ένα από τα χαρακτηριστικά του αλγορίθμου των Lucas-Kanade, όπως ισχύει και γενικά για τους αλγόριθμους τοπικού υπολογισμού της οπτικής ροής, είναι ότι δε δίνει μεγάλη πυκνότητα διανυσμάτων ροής, δηλ. η πληροφορία της οπτικής ροής σβήνει γρήγορα ανάμεσα στα όρια της κίνησης και τα εσωτερικά μέρη μεγάλων ομογενών περιοχών δείχνουν μικρή κίνηση (ενώ θα μπορούσε να ανήκουν σε μεγάλα κινούμενα αντικείμενα). Ο αλγόριθμος όμως υπερέρχει στην αντοχή και ρωμαλεότητά του απέναντι στο θόρυβο. Υπάρχουν αντίστοιχα αλγόριθμοι που δε βασίζονται σε τοπικούς υπολογισμούς, με πιο γνωστό τον αλγόριθμο των Horn και Schunck, ο οποίος δίνει μεγάλη πυκνότητα διανυσμάτων οπτικής ροής (δηλ. η πληροφορία της ροής η οποία λείπει από το εσωτερικό μεγάλων ομογενών περιοχών συμπληρώνεται από τα όρια της κίνησης). Ο αλγόριθμος όμως αυτός είναι πιο ευαίσθητος στον θόρυβο.

Πολυκλιμακωτή Μορφή του Αλγόριθμου των Lucas-Kanade

Για τους σκοπούς της διπλωματικής εργασίας χρησιμοποιήσαμε την πολυκλιμακωτή εκδοχή του αλγόριθμου των Lukas - Kanade. Ο αλγόριθμος σχηματίζει την γκαουσιανή πυραμίδα κάθε πλαισίου και ξεκινάει να υπολογίζει, με τον τρόπο που περιγράψαμε στον αλγόριθμο μονής κλίμακας, την οπτική ροή ξεκινώντας από τις πιο αδρές και προχωρώντας στις πιο λεπτομερείς κλίμακας. Η λύση της πιο αδρής κλίμακας χρησιμοποιείται ως λύση για τη πιο λεπτομερή κλίμακα, δίνοντας μεγαλύτερη ακρίβεια και ταχύτητα σύγκλισης στον αλγόριθμο. Η διαδικασία συνεχίζεται μέχρι να φτάσουμε στο επίπεδο 0, το οποίο αντιστοιχεί στην αρχική μορφή των πλαισίων.

4.3.2 Μεθοδολογίες προς Σύγκριση

Διαφορά μεταξύ των ίδιων των πλαισίων

Σαν ένα απλό σημείο αναφοράς, θα εξετάσουμε ως μέτρο της σημαντικότητας κάθε πλαισίου το βαθμό στον οποίο αυτό διαφέρει από το προηγούμενό του. Το μέτρο ομοιότητας το οποίο θα χρησιμοποιήσουμε είναι αυτό που αναφέραμε και στην ενότητα 3.3.2 (εξίσωση 3.1), δηλαδή η ετεροσυσχέτιση μεταξύ δισδιάστατων σημάτων - εικόνων, το οποίο παραθέτουμε ξανά για ευκολία:

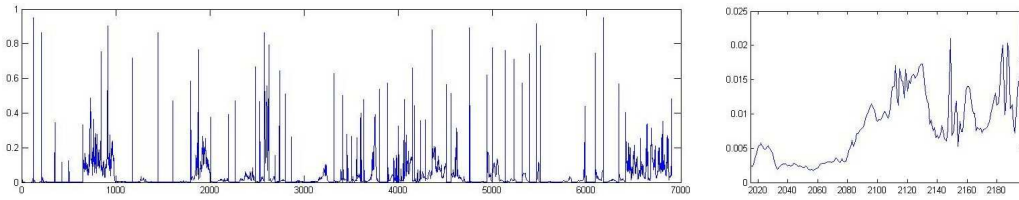
$$C = \frac{\sum_i \sum_j (I_1(i, j) - \mu_1) \times (I_2(i, j) - \mu_2)}{\sqrt{\sum_i \sum_j (I_1(i, j) - \mu_1)^2 \times \sum_i \sum_j (I_2(i, j) - \mu_2)^2}}$$

το οποίο λαμβάνει τιμές μεταξύ -1 και 1. Για να εκφράσουμε τη διαφορά, αντί την ομοιότητα, μεταξύ των πλαισίων, θα ορίσουμε το μέτρο D ως

$$D_{I_1, I_2} = 1 - C_{I_1, I_2}$$

το οποίο λαμβάνει τιμές από 0 (ταύτιση των πλαισίων) ως 2 (το ένα πλαίσιο αποτελεί αντιστροφή του άλλου). Βλέποντας και από το Σχήμα 4.6, το μέτρο αυτό κινείται κυρίως σε δύο διακριτές περιοχές τιμών. Η μια περιλαμβάνει πολύ μικρές τιμές, π.χ. ως 0.1 - 0.2 και η άλλη τιμές περίπου μια τάξη μεγέθους μεγαλύτερες. Η πρώτη περιοχή αντιστοιχεί με μεγάλη πιθανότητα σε πλαίσια που ανήκουν στην ίδια σκηνή, ενώ μεγάλες διαφορές προκύπτουν όταν γίνεται αλλαγή σκηνής. Μια χρήση του μέτρου αυτού θα ήταν η εύρεση των διαφορετικών σκηνών και αλλαγών πλάνων μέσα σε μια μεγάλη ακολουθία.

Μια εναλλακτική υλοποίηση αφορά τον υπολογισμό του χάρτη οπτικής προσοχής για κάθε πλαίσιο και χρήση του ως βάρη για κάθε εικονοστοιχείο των πλαισίων, με παρεμβολή στην ίδια κλίμακα και πολλαπλασιασμό σημείο - προς - σημείο. Έτσι εστιάζουμε περισσότερο στις διαφορές της εικόνας οι οποίες αντιστοιχούν σε σημαντικότερες περιοχές. Παράδειγμα δίνεται στο Σχήμα 4.8.



Σχήμα 4.6: Τιμές του μέτρου D_{I_1, I_2} για ένα μεγάλο αριθμό (6900) πλαισίων από την ταινία “ 300 ”. Αριστερά: τιμές για όλο το μήκος της ακολουθίας. Οι πολύ μεγάλες εξάρσεις αντιστοιχούν κατά πολύ μεγάλο ποσοστό σε αλλαγές σκηνής. Δεξιά: τιμές μέσα σε μια σκηνή. Βλέπουμε ότι εδώ οι τιμές κινούνται μέσα σε μια άλλη τάξη μεγέθους.

Προσδιορισμός Ενδιαφέροντος Βάσει Μεταβολών του Χάρτη Οπτικής Προσοχής

Το ίδιο με το πιο πάνω μπορούμε να επαναλάβουμε και με τη χρήση, αντί των ίδιων των πλαισίων, των χαρτών οπτικής προσοχής που αντιστοιχούν σε αυτά. Τους χάρτες αυτούς κανονικοποιούμε πριν τη σύγκριση. Σε αυτή την περίπτωση όμως αγνοούμε ότι κάποιες κινήσεις του χάρτη οπτικής προσοχής μπορεί να οφείλονται σε αντίστοιχες κινήσεις των αντικειμένων της πραγματικής εικόνας – ένας παρατηρητής που αντιλαμβάνεται την κίνηση των αντικειμένων, δεν εκπλήσσεται όταν ο χάρτης οπτικής προσοχής κινείται μαζί τους. Για το σκοπό αυτό, υπολογίζουμε την οπτική ροή μεταξύ των συγκεκριμένων πλαισίων και, θεωρώντας ότι ο χάρτης αναμένεται να κινηθεί με όμοιο τρόπο, την εφαρμόζουμε πάνω του ώστε να βρούμε πώς περιμένουμε να μοιάζει ο αντίστοιχος χάρτης του πιο πρόσφατου από τα αναφερόμενα πλαίσια. Δηλαδή, αν \vec{d} είναι το διάνυσμα οπτικής ροής που βρήκαμε μεταξύ των πλαισίων I_1 και I_2 , ώστε να

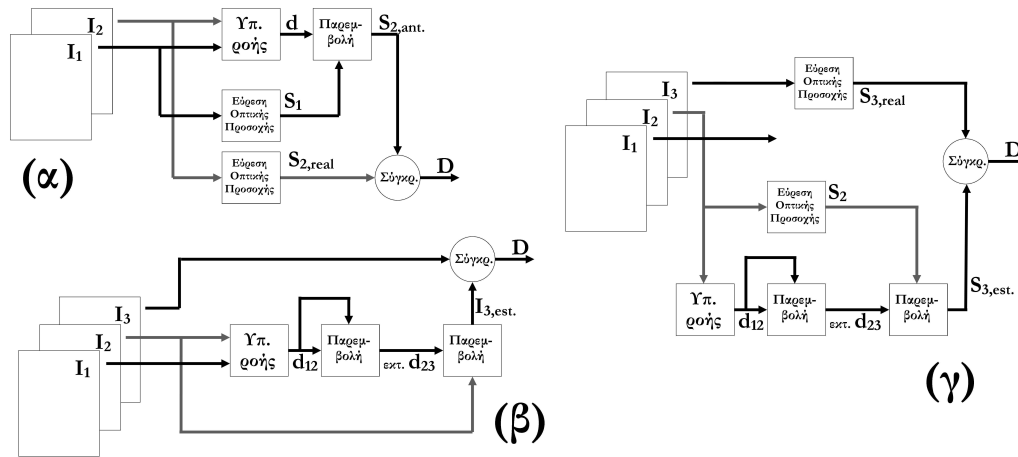
$$I_2(\vec{x}) \approx I_1(\vec{x} + \vec{d})$$

η αναμενόμενη μορφή του χάρτη οπτικής προσοχής του δεύτερου πλαισίου, έστω S_2^a (δείκτης a : ‘anticipated’), θα ισούται αντίστοιχα με

$$S_2^a(\vec{x}) \approx S_1(\vec{x} + \vec{d})$$

όπου προβαίνουμε σε όποιες αλλαγές κλίμακας είναι απαραίτητες λόγω της διαφοράς του μεγέθους των πλαισίων από το χάρτη. Στη συνέχεια υπολογίζουμε την πραγματική μορφή του χάρτη οπτικής προσοχής του δεύτερου πλαισίου, την S_2^r (δείκτης r : ‘real’) και εκφράζουμε τη “σημαντικότητα” του I_2 ως $D_{S_2^a, S_2^r}$. Το Σχήμα 4.7(α) δείχνει σχηματικά τη μεθοδολογία.

Στη φάση αυτή ουσιαστικά δεν προβαίνουμε σε καθαρή πρόβλεψη, δηλαδή έκφραση της αναμενόμενης μορφής του χάρτη πριν λάβουμε υπ’ όψιν το επόμενο



Σχήμα 4.7: Διαγραμματική αναπαράσταση των διάφορων μεθόδων που χρησιμοποιήσαμε: εύρεση σημαντικότητας D βάσει (α) σταθμισμένων μεταβολών του χάρτη οπτικής προσοχής, (β) πρόβλεψης του επόμενου πλαισίου και (γ) πρόβλεψης του επόμενου χάρτη οπτικής προσοχής.

πλαίσιο (εδώ χρησιμοποιήσαμε εκ των προτέρων το I_2 για την εύρεση της οπτικής ροής). Μπορούμε πιο επακριβώς να το περιγράψουμε ως υπολογισμό της διαφοράς μεταξύ των χαρτών με *στάθμιση της κίνησης* (movement compensation). Ένα ενδεικτικό παράδειγμα εφαρμογής βλέπουμε στο Σχήμα 4.8.

Προσδιορισμός Ενδιαφέροντος Βάσει Πρόβλεψης των Επόμενων Πλαισίων

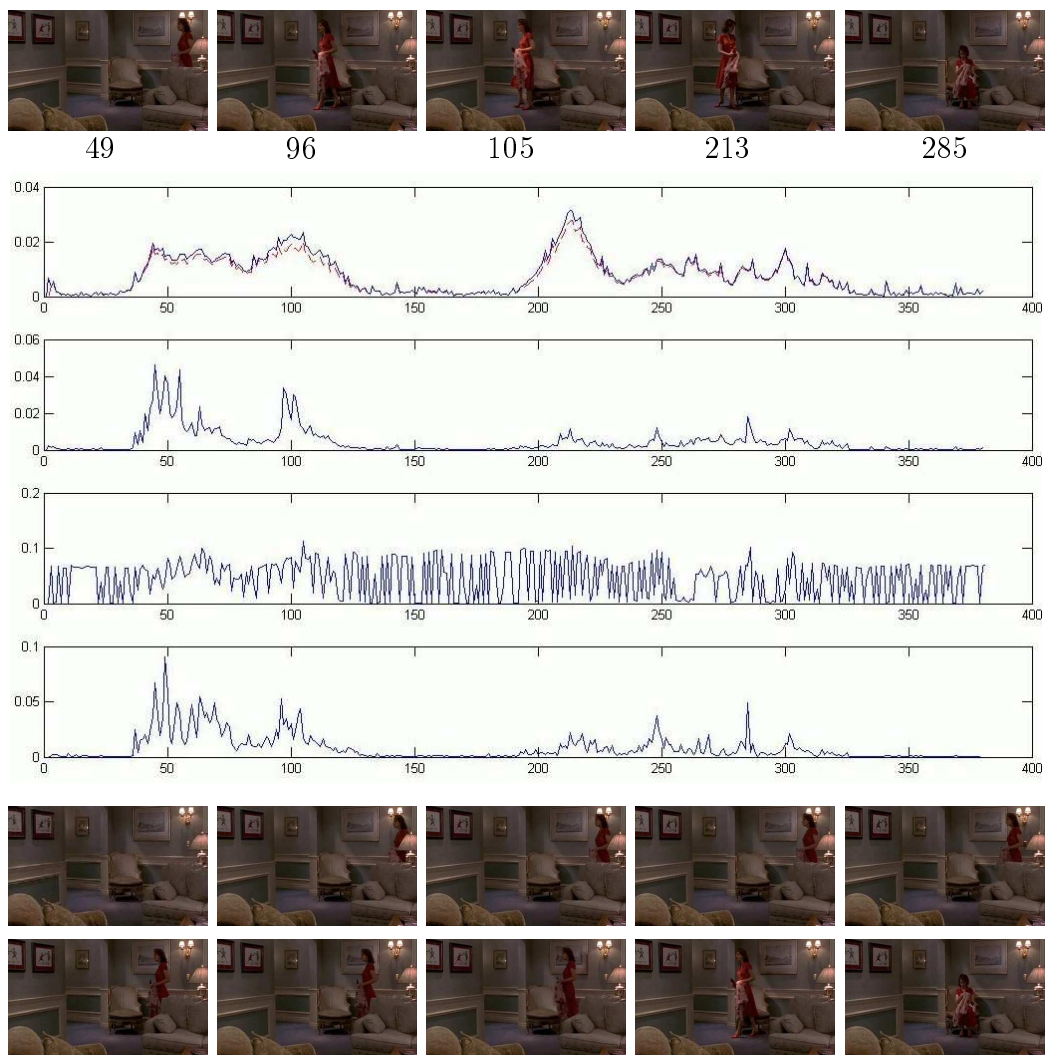
Προχωρούμε ένα βήμα πιο πέρα και στοχεύουμε στο να προβλέψουμε τη μορφή του επόμενου πλαισίου βάσει των δύο προηγούμενων. Θεωρούμε τρία διαδοχικά πλαίσια I_1, I_2 και I_3 . Βρίσκουμε την ροή \vec{d}_{12} μεταξύ των δύο πρώτων πλαισίων. Μπορούμε να θεωρήσουμε ότι η ροή μεταξύ των πλαισίων 2 και 3 θα παραμείνει περίπου η ίδια, η να την υπολογίσουμε εφαρμόζοντας την \vec{d}_{12} πάνω στον εαυτό της, δηλαδή

$$\vec{d}_{23}(\vec{x}) \approx \vec{d}_{12}(\vec{x} + \vec{d}_{12})$$

στη συνέχεια προβλέπουμε το πλαίσιο 3 ως

$$I_3^{est}(\vec{x}) \approx I_2(\vec{x} + \vec{d}_{23})$$

με μέτρο σημαντικότητας του το $D_{I_3^{est}, I_3^{real}}$. Ένα παράδειγμα της μεθόδου αυτής δίνουμε στο Σχήμα 4.8, ενώ στο Σχήμα 4.7(β) βλέπουμε σχηματικά τη λειτουργία της.



Σχήμα 4.8: Πάνω: πλαίσια από μέρος της ταινίας 'As good as it gets'. Από τη δημόσια διαθέσιμη βάση ακολουθιών του [25]. Σημειώνουμε τα πλαίσια τα οποία δίνουν μεγάλες τιμές σημαντικότητας βάσει των διάφορων μέτρων που περιγράφονται. Δεύτερη γραμμή: τιμές σημαντικότητας με διαφορές των πλαισίων – μπλε συνεχής γραμμή: απλή διαφορά, κόκκινη στικτή γραμμή, διαφορά με πρόβλεψη του επόμενου πλαισίου. Τρίτη γραμμή: τιμές με απλές διαφορές των πλαισίων βεβαρημένες από την οπτική προσοχή. Τέταρτη γραμμή: τιμές με τις σταθμισμένες με την ταχύτητα διαφορές των χαρτών οπτικής προσοχής. Πέμπτη γραμμή: τιμές με πρόβλεψη του χάρτη οπτικής προσοχής του επόμενου πλαισίου. Κάτω (δύο σειρές): πρώτο πλαίσιο και “ περιλήψη ” βάσει του παραπάνω μέτρου με κατώφλι 0.0475.

Προσδιορισμός Ενδιαφέροντος Βάσει Πρόβλεψης του Χάρτη Οπτικής Προσοχής

Το ίδιο με το πιο πάνω μπορούμε να εφαρμόσουμε και στην περίπτωση του χάρτη οπτικής προσοχής. Θεωρώντας τους χάρτες οπτικής προσοχής S_1, S_2 και S_3 των εικόνων I_1, I_2 και I_3 αντίστοιχα, εργαζόμενοι με όμοιο τρόπο προβλέπουμε το χάρτη 3 ως

$$S_3^{est}(\vec{x}) \approx S_2(\vec{x} + \vec{d}_{23})$$

και το μέτρο σημαντικότητας του πλαισίου 3 ως $D_{S_3^{est}, S_3^{real}}$. Το Σχήμα ;; δείχνει ένα αποτέλεσμα της μεθόδου, όπως και των υπολοίπων που αναφέραμε, πάνω σε μια ακολουθία από πραγματική ταινία. Επίσης, στο Σχήμα 4.7(γ) δίνουμε παραστατικά τα διαγράμματα λειτουργίας της τελευταίας μεθόδου.

Με την εξαγωγή του μέτρου σημαντικότητας κάθε πλαισίου, είναι εύκολη η δημιουργία περίληψης μιας ακολουθίας θέτοντας ένα κατώφλι βάσει του οποίου κρατάμε τα πλαίσια τα οποία το υπερβαίνουν. Σε κάποιες περιπτώσεις προστίθενται και κάποιοι περιορισμοί οι οποίοι στοχεύουν στη συνοχή των συνιστώντων μερών της περίληψης (smoothness).

4.3.3 Σχόλια

Η αξιολόγηση κάθε προσέγγισης του θέματος είναι, όπως αναφέραμε και στην εισαγωγή της ενότητας, επιρρεπής σε υποκειμενικά κριτήρια. Απαιτούνται διάφορες και εκτενείς έρευνες χρηστών (user studies) οι οποίοι θα αξιολογήσουν τις διάφορες περιλήψεις, οι οποίες απαιτούν αρκετό χρόνο και εμπλέκουν μεγάλο αριθμό ατόμων. Στις λίγες σχετικά ακολουθίες που εξετάσαμε, φάνηκε ότι γενικά τα ενδιαφέροντα γεγονότα, ιδίως αυτά που αναφέραμε στην εισαγωγή της ενότητας (είσοδος - έξοδος από τη σκηνή κλπ.) δίνουν μεγαλύτερες τιμές στα μέτρα (με εξαίρεση την διαφορά των χαρτών, σταθμισμένων με ταχύτητα όπως βλέπουμε στο σχήμα 4.8). Οι τεχνικές με χρήση πρόβλεψης δίνουν γενικά πιο σαφή αποτελέσματα, όπως μπορούμε να διακρίνουμε από τις πιο έντονες και διακριτές κορυφώσεις των αντίστοιχων διαγραμμάτων του ίδιου σχήματος. Σε κάθε περίπτωση, είναι ορατή η ανάγκη περαιτέρω διερεύνησης.

Κεφάλαιο 5

Συμπεράσματα

5.1 Συμβολή της Διπλωματικής Εργασίας

Το κύριο μέρος της διπλωματικής εργασίας απετέλεσε η μελέτη του δισδιάστατου μοντέλου των Itti et al.. Η διερεύνηση της χρήσης μορφολογικών πυραμίδων σε αυτό, στην καλύτερη γνώση μας, δεν έχει εξεταστεί προηγουμένως. Η χρήση ανίχνευσης γωνιών και συνδέσμων στην εικόνα, αν και αναφέρεται ως μια πιθανότητα από τους ίδιους τους Itti et al. στην αρχική δημοσίευση του 1998, δε φαίνεται να έχει διερευνηθεί εκτενώς στη βιβλιογραφία.

Ιδιαίτερα χρήσιμη ήταν η ενσωμάτωση της εξαγωγής της έρευνας πάνω στην ανθρώπινη αντίληψη της οπτικής προσοχής, στην οποία συμμετείχε μεγάλος αριθμός παρατηρητών και οδήγησε στην εύρεση ενός μέτρου αξιολόγησης των αποκρίσεων του κάθε μοντέλου. Η συλλογή των στοιχείων αυτών, πέρα από ενδιαφέρονσα, αποτελεί έναν σημαντικό πυλώνα υποστήριξης της δουλειάς που προτείναμε και διεξήγαμε.

Στα πλαίσια της εργασίας έγινε υλοποίηση διάφορων αλγορίθμων, όπως του ίδιου του μοντέλου των Itti et al., της δημιουργίας πολυκλιμακωτών πυραμίδων και της τρισδιάστατης επέκτασης του πρώτου. Στην τελευταία επιλύθηκαν αρκετά ζητήματα τεχνικής φύσης τα οποία οφείλονταν στην απουσία έτοιμων συναρτήσεων χειραγώγησης τρισδιάστατων δομών στο MATLAB, στο βαθμό τουλάχιστον στον οποίο υπάρχουν για δομές δύο διαστάσεων.

Αναλυτικά, η εργασία προβαίνει σε μια συνοπτική αλλά πλήρη κατά το δυνατόν παρουσίαση του θέματος της οπτικής προσοχής, περιέχοντας:

- Εισαγωγή στην έννοια και σημασία της προσοχής εν γένει και ειδικότερα της οπτικής προσοχής
- Περιγραφή των βιολογικών μηχανισμών οπτικής προσοχής, με εστίαση στα στοιχεία τα οποία μοντελοποιούνται υπολογιστικά

- Παρουσίαση του μοντέλου των Itti et al. και διερεύνηση των μετατροπών και επεκτάσεών του βάσει εργαλείων όπως μορφολογικές μορφές πολυκλιμακωτής ανάλυσης ή ανιχνευτές γωνιών, τα οποία ορίζονται και περιγράφονται.
- Σύγκριση των αποτελεσμάτων, για το σκοπό της οποίας διεξήχθη μια έρευνα οπτικής αντίληψης με πάνω από 30 συμμετέχοντες. Παρουσιάζεται τόσο η σύγκριση όσο και τα ίδια τα αποτελέσματα των ανθρώπινων αποκρίσεων.
- Παρουσίαση δυνατών επεκτάσεων του μοντέλου στις τρεις διαστάσεις, στη βάση τόσο δικής μας υλοποίησης όσο και πρότερης δουλειάς διάφορων ερευνητών, με προτάσεις σύντηξης στοιχείων από τα διάφορα αυτά μοντέλα.
- Περιγραφή εναλλακτικών μεθόδων βασιζόμενες σε καθοδικές (top-down) προσεγγίσεις της οπτικής προσοχής, παρουσίαση και πρόταση για βελτίωση διάφορων μορφών αξιολόγησης.

5.2 Μελλοντικές Κατευθύνσεις

Πιστεύουμε ότι η διπλωματική εργασία έθιξε έναν σημαντικό αριθμό πιθανών μελλοντικών κατευθύνσεων. Αν και υλοποίηση της προσθήκης ενός χάρτη γωνιακών ενδείξεων ενίσχυσε το μοντέλο οπτικής προσοχής, όπως είδαμε στο κεφάλαιο 3, ο ρόλος των γωνιών στην οπτική προσοχή φαίνεται ακόμη μεγαλύτερος ώστε να αναζητηθεί εναλλακτική μέθοδος ενσωμάτωσής του. Αυτή η μέθοδος πιθανόν και να ξεφεύγει από την έννοια του ενός ακόμη χάρτη.

Μια ελπιδοφόρα προσέγγιση για την δισδιάστατη αρχιτεκτονική οπτικής προσοχής αποτελεί η ανάλυση κύριων παραγόντων DCA – Dominant Component Analysis και η δυνατότητα χρήσης της Teager - Kaiser ενέργειας σε κάθε εικονοστοιχείο ως μέτρο της απόκρισης προσανατολισμού. Αντικαθιστώντας τέτοιους μηχανισμούς στη θέση των υφιστάμενων για την εύρεση των διακλιμακωτών διαφορών προσανατολισμού, είδαμε κάποια ενδιαφέροντα προκαταρκτικά αποτελέσματα η περαιτέρω μελέτη των οποίων δυστυχώς ξεπερνούσε το χρονικό πλαίσιο της εργασίας. Μια άλλη πιθανή εφαρμογή της DCA βρίσκεται στη συγχώνευση των ενδιάμεσων χαρτών.

Η χρήση της τρισδιάστατης επέκτασης του αλγορίθμου των Itti et al. σε συνδυασμό με την αντίστοιχη εκδοχή του ανιχνευτή Harris των Laptev et al., μπορεί να εξεταστεί διεξοδικά ως προς τις επιδόσεις της σε αναγνώριση σχημής και κατηγοριοποίηση ανθρώπινων πράξεων.

Τέλος, η περαιτέρω σύγκριση και αξιολόγηση των μοντέλων του δεύτερου μέρους του κεφαλαίου 4, πάνω σε μεγάλες βάσεις ακολουθιών, μπορεί να οδηγήσει σε επέκταση και ραφινάρισμά τους ώστε να αποδώσουν ενδιαφέροντα αποτελέσματα.



"Piled Higher and Deeper" by Jorge Cham
www.phdcomics.com

Παράρτημα Α΄

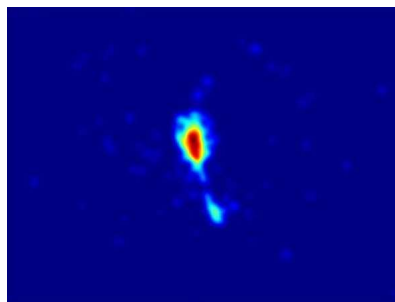
Εικόνες που Εξετάστηκαν - Αποτελέσματα Ιχνηλάτησης Κέρσορα

Στις σελίδες που ακολουθούν παρατίθενται οι 26 εικόνες φυσικών σκηνών (natural scenes) οι οποίες ελέγχθηκαν ως προς την προσοχή μέσω της διαδικτυακής εφαρμογής Validattention του εργαστηρίου TCTS, η οποία βρίσκεται στη σελίδα <http://tcts.fpms.ac.be/mousetrack/pageAccueil.php>. Χρησιμοποιήθηκε η τεχνική παρακολούθησης μέσω του δείκτη του ποντικιού (mouse tracking) ή "ιχνηλάτησης κέρσορα". Οι εικόνες συνοδεύονται με το αντίστοιχο, φιλτραρισμένο με ένα Γκαουσιανό φίλτρο τυπικής απόκλισης $\sigma = 6$ και μεγέθους 32×32 , αποτέλεσμα των μέσων αποκρίσεων από 21-31 χρήστες — ο αριθμός των χρηστών που συμμετείχαν στην κάθε εικόνα αναφέρεται πάνω από αυτή.

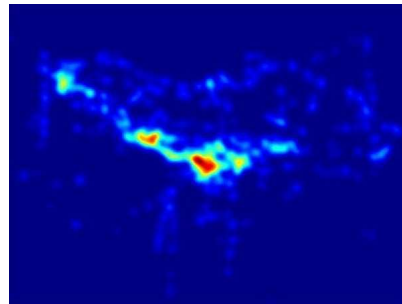
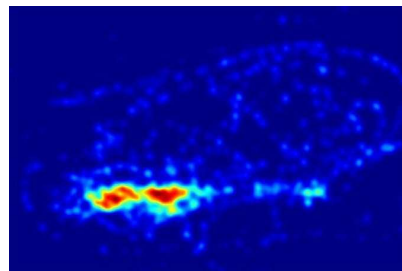
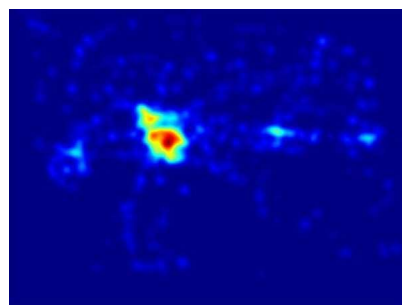
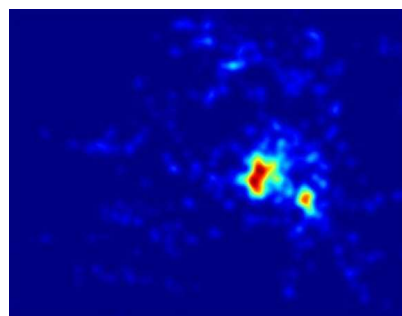
Εικόνα 1: *balloon.jpg* (27 διαφορετικές αποκρίσεις)



αρχική εικόνα

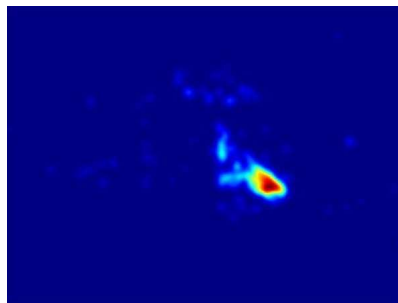
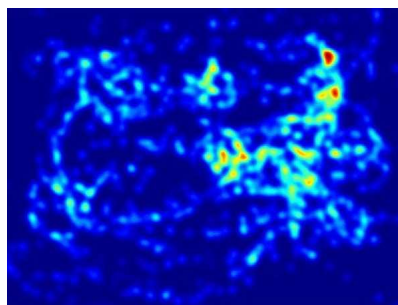
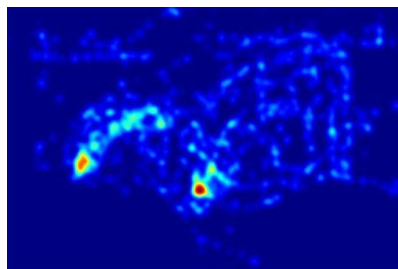
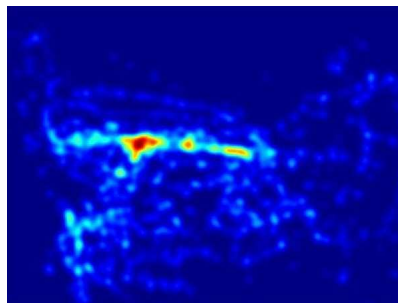


μ. ό. αποκρίσεων οπτικής προσοχής

Εικόνα 2: *cars.jpg* (23 διαφορετικές αποκρίσεις)Εικόνα 3: *cows.jpg* (25 διαφορετικές αποκρίσεις)Εικόνα 4: *drive.jpg* (24 διαφορετικές αποκρίσεις)Εικόνα 5: *entrance.jpg* (26 διαφορετικές αποκρίσεις)

αρχική εικόνα

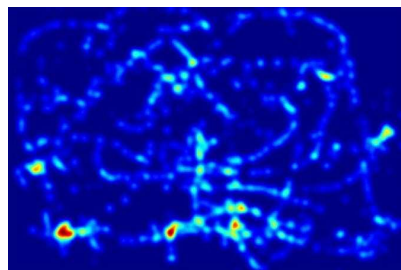
μ. ό. αποκρίσεων οπτικής προσοχής

Εικόνα 6: *gardens.jpg* (21 διαφορετικές αποκρίσεις)Εικόνα 7: *home.jpg* (29 διαφορετικές αποκρίσεις)Εικόνα 8: *horses.jpg* (29 διαφορετικές αποκρίσεις)Εικόνα 9: *mechelec.jpg* (22 διαφορετικές αποκρίσεις)

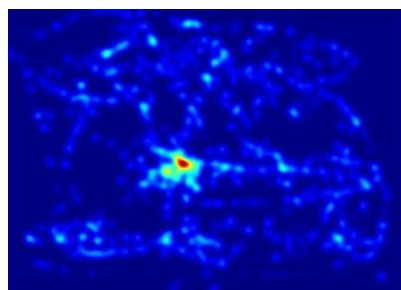
αρχική εικόνα

μ. ό. αποκρίσεων οπτικής προσοχής

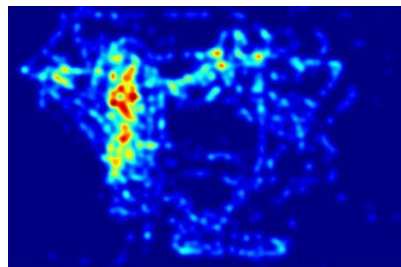
Εικόνα 10: *old building.jpg* (29 διαφορετικές αποκρίσεις)



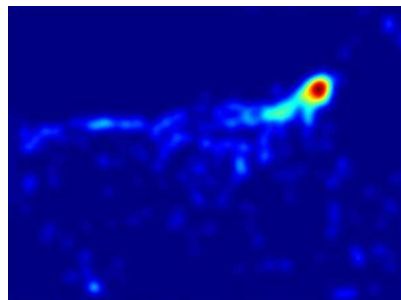
Εικόνα 11: *parked cars.jpg* (25 διαφορετικές αποκρίσεις)



Εικόνα 12: *rhox.jpg* (29 διαφορετικές αποκρίσεις)

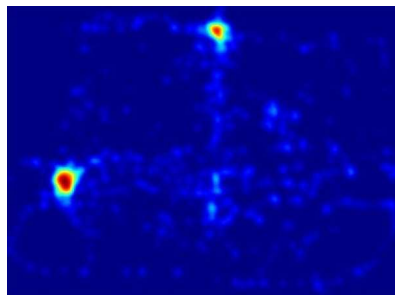
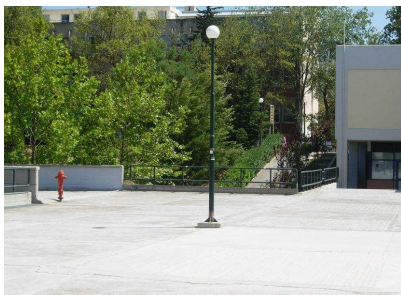
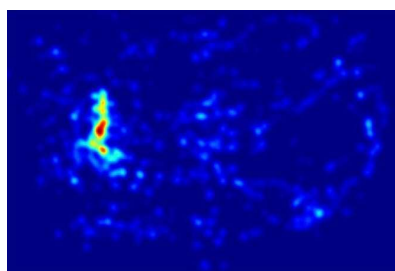
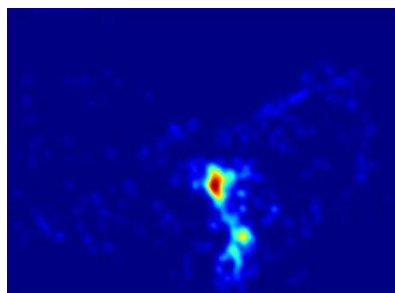
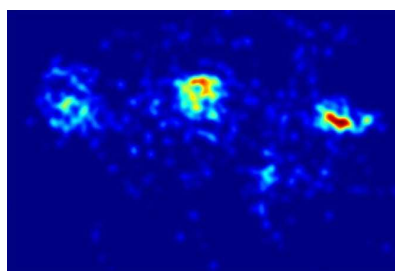


Εικόνα 13: *peacock.jpg* (28 διαφορετικές αποκρίσεις)



αρχική εικόνα

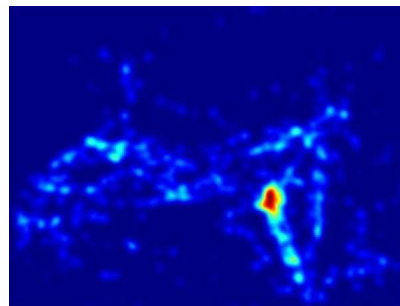
μ. ό. αποκρίσεων οπτικής προσοχής

Εικόνα 14: *post.jpg* (25 διαφορετικές αποκρίσεις)Εικόνα 15: *punchbowl.jpg* (26 διαφορετικές αποκρίσεις)Εικόνα 16: *road.jpg* (24 διαφορετικές αποκρίσεις)Εικόνα 17: *salesigns.jpg* (25 διαφορετικές αποκρίσεις)

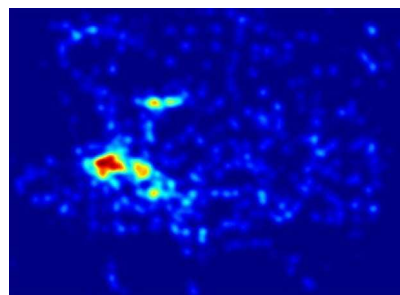
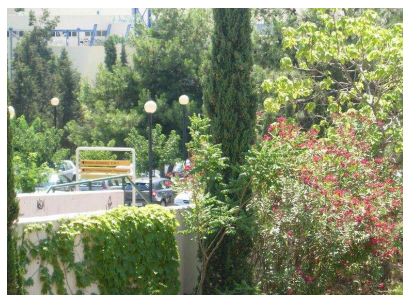
αρχική εικόνα

μ. ό. αποκρίσεων οπτικής προσοχής

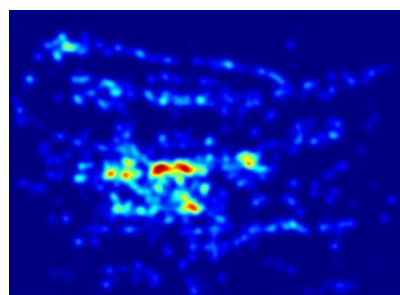
Εικόνα 18: *ship.jpg* (31 διαφορετικές αποκρίσεις)



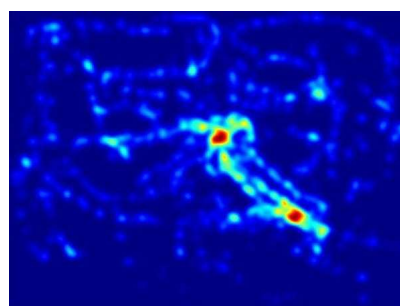
Εικόνα 19: *sign.jpg* (23 διαφορετικές αποκρίσεις)



Εικόνα 20: *square.jpg* (22 διαφορετικές αποκρίσεις)

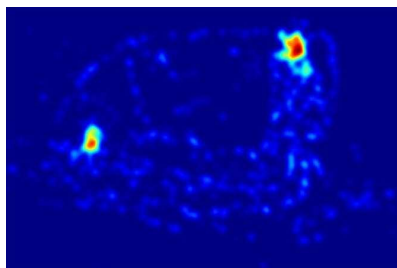
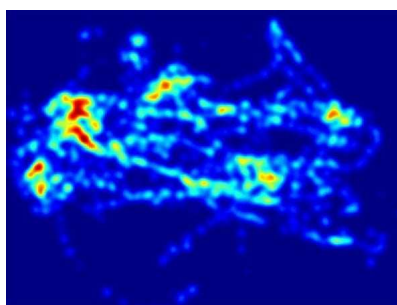
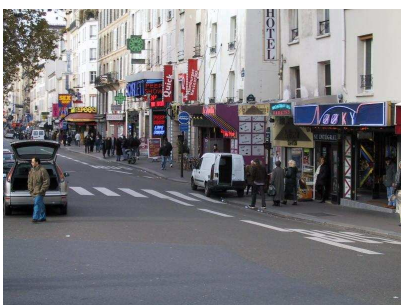
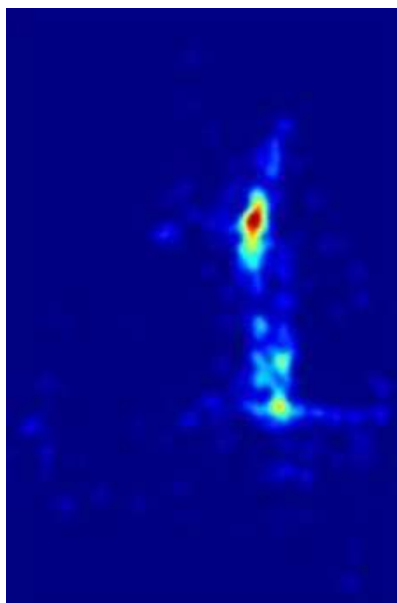


Εικόνα 21: *stairview.jpg* (23 διαφορετικές αποκρίσεις)



αρχική εικόνα

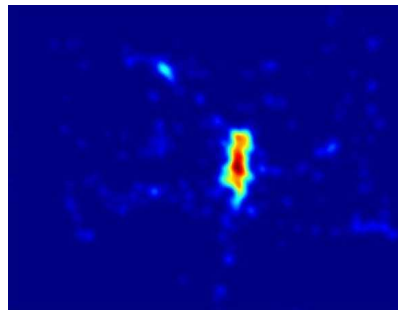
μ. ό. αποκρίσεων οπτικής προσοχής

Εικόνα 22: *statue - krumlov.jpg* (26 διαφορετικές αποκρίσεις)Εικόνα 23: *street.jpg* (21 διαφορετικές αποκρίσεις)Εικόνα 24: *turbine.jpg* (24 διαφορετικές αποκρίσεις)

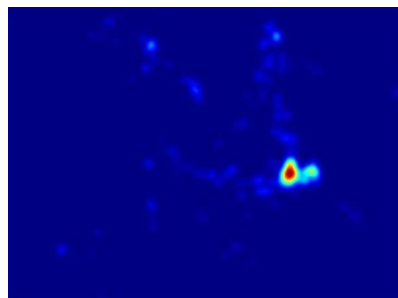
αρχική εικόνα

μ. ό. αποκρίσεων οπτικής προσοχής

Εικόνα 25: *wait.jpg* (23 διαφορετικές αποκρίσεις)



Εικόνα 26: *walking.jpg* (21 διαφορετικές αποκρίσεις)



αρχική εικόνα

μ. ό. αποκρίσεων οπτικής προσοχής

Βιβλιογραφία

- [1] Π. Μαραγκός, “ Ανάλυση Εικόνων και Όραση Υπολογιστών ”, Ιούνιος 2005.
- [2] Κ. Ραπαντζίκος, “ Χρήση υπολογιστικών μοντέλων οπτικής προσοχής για την ανάλυση στατικών και κινούμενων εικόνων ”, διδακτορική διατριβή, Απρίλιος 2008.
- [3] Κ. Τσιβερδής, “ Διερεύνηση πιθανότητας βελτίωσης της απόδοσης του προτύπου συμπίεσης ψηφιακών εικόνων JPEG, με τη χρήση σχημάτων ανάλυσης σε πολλαπλές κλίμακες ”, διπλωματική εργασία, Μάρτιος 2003.
- [4] Burt, P. and Adelson, E., ”The Laplacian Pyramid as a Compact Image Code”, IEEE Transactions on Communications, vol.31, issue 4, pp.532–540, Apr 1983.
- [5] M.W. Cannon and S.C. Fullenkamp, “A Model for Inhibitory Lateral Interaction Effects in Perceived Contrast,” Vision Res., vol. 36, no. 8, pp. 1,115–1,125, Apr. 1996.
- [6] P. Dollar, V. Rabaud, G. Cottrell, S. Belongie, ”Behavior Recognition via Sparse Spatio-Temporal Features”, VS-PETS, pp. 65–72, Oct 2005.
- [7] S. Engel, X. Zhang, and B. Wandell, “Colour Tuning in Human Visual Cortex Measured With Functional Magnetic Resonance Imaging,” Nature, vol. 388, no. 6,637, pp. 68–71, July 1997.
- [8] G. Evangelopoulos, K. Rapantzikos, A. Potamianos, P. Maragos, A. Zlatintsi and Y. Avrithis, ”Movie Summarization Based On Audio-Visual Saliency Detection”, Proc. Int’l Conference on Image Processing (ICIP-2008), San Diego, California, Oct. 2008.
- [9] J.P. Gottlieb, M. Kusunoki, and M.E. Goldberg, ”The Representation of Visual Saliency in Monkey Parietal Cortex”, Nature, vol. 391, no. 6,666, pp. 481–484, Jan. 1998.

-
- [10] H. Greenspan, S. Belongie, R. Goodman, P. Perona, S. Rakshit, and C.H. Anderson, "Overcomplete Steerable Pyramid Filters and Rotation Invariance," Proc. IEEE Computer Vision and Pattern Recognition, pp. 222–228, Seattle, Wash., June 1994.
- [11] F.H. Hamker, "Modeling Attention: From Computational Neuroscience to Computer Vision", Neurobiology of Attention, L. Itti, G. Rees, J. Tsotsos (editors), Academic Press, 2005.
- [12] C. Harris, M. Stephens, "A combined corner and edge detector, Alvey Vision Conference, pp. 147–152, 1988.
- [13] Healey, C. G., Booth, K. S., and Enns, J. T. "High-Speed Visual Estimation Using Preattentive Processing". ACM Transactions on Human Computer Interaction 3(2), pages 107–135, 1996.
- [14] H. J. A. M. Heijmans and J. Goutsias, "Nonlinear multiresolution signal decomposition schemes – Part II: Morphological Wavelets" , IEEE Transactions on Image Processing, Vol. 9, No. 11, November 2000.
- [15] Itti, L., Baldi, P., "Bayesian Surprise Attracts Human Attention", Advances in Neural Information Processing Systems, Vol. 19 (NIPS 2005), pp. 1–8, Cambridge, MA:MIT Press, 2006.
- [16] Itti, L., Koch, C., Niebur, E., "Model of saliency-based visual attention for rapid scene analysis", IEEE Trans. on Pattern Analysis and Machine Intelligence, 20(11) pp. 1254–1259, 1998.
- [17] L. Itti, C. Koch, Computational Modeling of Visual Attention, Nature Reviews Neuroscience, Vol. 2, No. 3, pp. 194–203, Mar 2001.
- [18] Itti, L., Rees, G., Tsotsos, J.K., "Neurobiology of Attention", Elsevier Academic Press, ISBN: 0-12-375731-2, 2005.
- [19] P. T. Jackway and M. Deriche, "Scale-space properties of multiscale morphological dilation-erosion", IEEE Trans. Pattern Anal. Mach. Intell. 18(1), 1996, 38–51.
- [20] C. Koch and S. Ullman, "Shifts in Selective Visual Attention: Towards the Underlying Neural Circuitry," Human Neurobiology, vol. 4, pp. 219–227, 1985.
- [21] C. Koch, Biophysics of Computation: Information Processing in Single Neurons. New York: Oxford Univ. Press, 1998.

- [22] Human action dataset, <http://www.nada.kth.se/cvap/actions/>
- [23] I. Laptev, "On Space-Time Interest Points", *International Journal of Computer Vision* 64(2/3), pp. 107–123, 2005.
- [24] I. Laptev, "Improvements og Object Detection Using Boosted Histograms", *Proc. BMVC '06*, Edinburgh, UK, pp. III:949–958.
- [25] I. Laptev, M. Marszalek, C. Schmid and B. Rozenfeld, "Learning realistic human actions from movies", 2008.
- [26] Le Meur, O., "Attention selective en visualisation d' images fixes et animees affichees sur ecran : Modeles et evaluation des performances - Applications", PhD Thesis, University of Nantes, 2005.
- [27] A.G. Leventhal, "The Neural Basis of Visual Function: Vision and Visual Dysfunction", vol. 4. Boca Raton, Fla.: CRC Press, 1991.
- [28] Lucas, B.D. and Kanade, T., "An iterative image registration technique with an application to stereo vision", *Proceedings of Imaging understanding workshop*, pp.121–130.
- [29] Yu F. Ma, L. Lu, H. J. Zhang and M. Li, "A User Attention Model for Video Summarization", *ACM Multimedia '02*, Dec 2002.
- [30] M. Mancas, "Computational Attention: Towards attentive computers", doctoral dissertation, 2007.
- [31] G. Matheron, "Les Nivellements", Technical Report, Centre de Morphologie Mathématique, 1997.
- [32] F. Meyer, "The levelings, in *Mathematical Morphology and Its Applications to Image and Signal Processing*" (H. Heijmans and J. Roerdink, Eds.), pp. 199–207, Kluwer Academic, Dordrecht/ Norwell, MA, 1998.
- [33] F. Meyer and P. Maragos, "Nonlinear Scale-Space Representation with Morphological Levelings", *Journal of Visual Communications and Image Representation*, vol.11, pp.245–265, 2000.
- [34] Mudge, T. N., Turney, J. L., and Volz "Automatic generation of salient features for the recognition of partially occluded parts", *Robotica*, Vol 5, pp 117–127, 1987.

-
- [35] E. Niebur and C. Koch, "Computational Architectures for Attention", R. Parasuraman, ed., *The Attentive Brain*, pp. 163–186. Cambridge, Mass.: MIT Press, 1998.
- [36] Alison Noble, "Descriptions of Image Surfaces", PhD thesis, Department of Engineering Science, Oxford University 1989.
- [37] G. Papandreou, P. Maragos and A. Kokaram, "Image Inpainting With a Wavelet Domain Hidden Markov Tree Model", *Proceedings of International Conference on Acoustics, Speech, and Signal Processing (ICASSP-08)*, Las Vegas, USA, April 2008.
- [38] Pöder, E., "Crowding, feature integration, and two kinds of 'attention'", *Journal of Vision*, 6(2):7, 163–169, 2006.
- [39] M.I. Posner and Y. Cohen, "Components of Visual Orienting," H. Bouma and D.G. Bouwhuis, eds., *Attention and Performance*, vol. 10, pp. 531–556. Hilldale, N.J.: Erlbaum, 1984.
- [40] Schneider, W. & R. M. Shiffrin, "Controlled and automatic human information processing: Detection, search, and attention", *Psychological Review*, 84, 1–66, 1977
- [41] J. Serra, "Quelques Propriétés des Nivellements", Technical Report 30/98/MM, Centre de Morphologie Mathématique, 1998.
- [42] M. Smith and T. Kanade, "Video Skimming and Characterization through the Combination of Image and Language Understanding Techniques", 1997.
- [43] F.-K. Sun and P. Maragos, "Experiments on Image Compression Using Morphological Pyramids", *Visual Communications and Image Processing IV, Proc. of SPIE*, vol.1141, pp.1303–1312, Philadelphia, PA, Nov. 1989.
- [44] Treisman, A.M., and Gelade, G., "A feature integration theory of attention", *Cognitive Psychology*, 12(1), 97–136, 1980.
- [45] J.K. Tsotsos, S.M. Culhane, W.Y.K. Wai, Y. Lai, N. Davis, F. Nuo, "Modeling visual attention via selective tuning", *Artificial Intelligence*, vol. 78, pp. 507–545, 1995.
- [46] Image Saliency Mapping and Ranking Using an Extensible Visual Attention Model Based on MPEG-7 Feature Descriptors, Dec 2005.

- [47] A. Yarbus, "Eye movements and vision", Plenum Press, New York, 1967.