



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ
ΥΠΟΛΟΓΙΣΤΩΝ
ΤΟΜΕΑΣ ΣΗΜΑΤΩΝ, ΕΛΕΓΧΟΥ ΚΑΙ ΡΟΜΠΟΤΙΚΗΣ
ΕΡΓΑΣΤΗΡΙΟ ΟΡΑΣΗΣ ΥΠΟΛΟΓΙΣΤΩΝ, ΕΠΙΚΟΙΝΩΝΙΑΣ ΛΟΓΟΥ ΚΑΙ
ΕΠΕΞΕΡΓΑΣΙΑΣ ΣΗΜΑΤΩΝ

Μοντέλα Inpainting για Αυτόματη Αποκατάσταση
Περιεχομένου - Υφής και Απομάκρυνση Κειμένου από
Εικόνες

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Ευτύχιος Αριστόδημος
Πνευματικάκης

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

Αθήνα, Ιούλιος 2004



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ
ΥΠΟΛΟΓΙΣΤΩΝ
ΤΟΜΕΑΣ ΣΗΜΑΤΩΝ, ΕΛΕΓΧΟΥ ΚΑΙ ΡΟΜΠΟΤΙΚΗΣ
ΕΡΓΑΣΤΗΡΙΟ ΟΡΑΣΗΣ ΥΠΟΛΟΓΙΣΤΩΝ, ΕΠΙΚΟΙΝΩΝΙΑΣ ΛΟΓΟΥ ΚΑΙ
ΕΠΕΞΕΡΓΑΣΙΑΣ ΣΗΜΑΤΩΝ

Μοντέλα Inpainting για Αυτόματη Αποκατάσταση
Περιεχομένου - Υφής και Απομάκρυνση Κειμένου από
Εικόνες

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Ευύχιος Αριστόδημος
Πνευματικάκης

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

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή στις 14 Ιουλίου 2004.

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

.....
Στέφανος Κόλλιας
Καθηγητής Ε.Μ.Π.

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

Αθήνα, Ιούλιος 2004.

.....
Ευτύχιος Αριστόδημος Πνευματικάκης
Διπλωματούχος Ηλεκτρολόγος Μηχανικός και Μηχανικός Υπολογιστών Ε.Μ.Π.

Copyright© Ευτύχιος Αριστόδημος Μ. Πνευματικάκης, 2004.
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

Η παρούσα εργασία ασχολείται με το Inpainting πρόβλημα, δηλαδή με την τεχνική επεξεργασίας μιας εικόνας με έναν μη αναγνωρίσιμο τρόπο. Η τεχνική αυτή στο Inpainting πρόβλημα αποσκοπεί στη διόρθωση λαθών που υπάρχουν στην εικόνα ή στην απομάκρυνση κάποιου αντικειμένου από την εικόνα. Αρχικά γίνεται μια θεμελίωση του Inpainting προβλήματος, και των βασικών τρόπων αντιμετώπισής του, καθώς και ο διαχωρισμός των δύο βασικών εννοιών που περιέχει μία εικόνα: Του περιεχομένου και της υφής. Στη συνέχεια παρουσιάζονται αναλυτικά οι τρόποι επίλυσης του προβλήματος πρώτα για εικόνες περιεχομένου, όπου η επίλυση γίνεται με μεταβολικές μεθόδους που οδηγούν σε επαναληπτικά σχήματα μερικών διαφορικών εξισώσεων που συγκλίνουν στην λύση, και στη συνέχεια για εικόνες υφής, όπου χρησιμοποιούνται τεχνικές σύνθεσης υφής, ενώ παρουσιάζεται και ο συνδιασμός των δύο κατηγοριών για ταυτόχρονο Inpainting περιεχομένου και υφής. Μετά την ολοκλήρωση της παρουσίασης των μεθόδων Inpainting, δίνεται έμφαση στην αυτόματη επιλογή της Inpainting περιοχής. Παρουσιάζεται ένας αλγόριθμος με μεθόδους της μορφολογικής επεξεργασίας εικόνας, για τον αυτόματο εντοπισμό κειμένου σε εικόνες και τέλος ο συνδιασμός των αλγορίθμων που οδηγεί σε έναν νέο αλγόριθμο για αυτόματο εντοπισμό και απομάκρυνση κειμένου από εικόνες.

Λέξεις—κλειδιά

όραση υπολογιστών, επεξεργασία εικόνας, Inpainting, αποκατάσταση εικόνας, περιεχόμενο, υφή, μεταβολικές μέθοδοι, μερικές διαφορικές εξισώσεις, σύνθεση υφής, εντοπισμός κειμένου σε εικόνες, μαθηματική μορφολογία

Abstract

This diploma thesis deals with the inpainting problem, namely the technique of modifying an image in a non detectable way. In the inpainting problem, this technique aims either in the restoration of damaged domains of the image or in the removal of selected objects. In the beginning, a theoretical foundation of the inpainting problem is presented, including the methods that are used for its solution, as well as the discrimination between the two major components of an image: Structure and texture. What follows, is an in depth analysis of the methods that solve the inpainting problem. For structure inpainting, where we use variational methods that yield to partial differential equations models which converge to the solution, and then for texture inpainting, where texture synthesis techniques are adopted. We also present the combination of these two methods for simultaneous structure and texture inpainting. After the presentation of the inpainting methods, emphasis is given on the automatic selection of the inpainting domain. We present an algorithm, with methods of morphological image processing, for automatic text detection in images. At the end, the combination of these algorithms leads to a new algorithm for automatic text detection and removal from pictures.

Keywords

computer vision, image processing, inpainting, image restoration, texture, variational methods, partial differential equations, texture synthesis, detection of text in images, mathematical morphology

Πρόλογος

Το Περιεχόμενο της Παρούσας Εργασίας

Η συγκεκριμένη εργασία ασχολείται με το Inpainting πρόβλημα το οποίο είναι μία πολύ πρόσφατη εφαρμογή της όρασης υπολογιστών. Τα κίνητρα για την ενασχόληση με το συγκεκριμένο πρόβλημα ήταν από τη μία οι σημαντικές και όμορφες εφαρμογές του Inpainting προβλήματος στην επεξεργασία εικόνας και στις ψηφιακές τέχνες και αφέτερου η εξοικείωση με σύγχρονες μεθόδους της όρασης υπολογιστών (μερικές διαφορικές εξισώσεις) που χρησιμοποιούνται για την επίλυση του Inpainting προβλήματος. Έτσι η εργασία ξεκίνησε με τη μελέτη των PDE μοντέλων που χρησιμοποιούνται για το Inpainting πρόβλημα αλλά και των αριθμητικών σχημάτων που χρησιμοποιούνται για την επίλυση τους. Η μελέτη του προβλήματος γρήγορα ανέδειξε και μια δεύτερη διαφορετική (αλλά και συμπληρωματική ως προς την πρώτη) κατηγορία μεθόδων για το Inpainting πρόβλημα, αυτή της σύνθεσης υφής. Έτσι στη συνέχεια κινηθήκαμε στη μελέτη και υλοποίηση αλγορίθμων σύνθεσης υφής και συσχέτισης τους με το Inpainting πρόβλημα. Τα στενά χρονικά όρια της διπλωματικής εργασίας δεν επέτρεψαν την κατεύθυνση της έρευνας προς κάποια νέα μέθοδο Inpainting, ή προς τη βελτίωση κάποιας υπάρχουσας. Έτσι η έρευνα κινήθηκε προς την προσπάθεια της αυτοματοποίησης του Inpainting προβλήματος, για τον αυτόματο εντοπισμό και αφαίρεση κειμένου από κάποια εικόνα. Όπως φαίνεται και από το περιεχόμενο της εργασίας, υπήρξε πολύ σημαντική προσπάθεια για την υλοποίηση των αλγορίθμων που θα παρουσιαστούν αλλά και για τον εντοπισμό της κατάλληλης βιβλιογραφίας καθώς το Inpainting πρόβλημα είναι πολύ πρόσφατο και δεν υπάρχει συγκεντρωμένη βιβλιογραφία.

Η εργασία είναι οργανωμένη σε πέντε κεφάλαια. Η παρουσίαση είναι κατά το δυνατόν αυτόνομη, αφού παρουσιάζονται όλα τα στοιχεία θεωρίας που θα χρησιμοποιούνται. Πάντως, θεωρούμε ότι ο αναγνώστης χρειάζεται μια σχετική γνώση και εμπειρία με κάποιες βασικές μεθόδους και κατευθύνσεις της όρασης υπολογιστών.

Στο **Κεφάλαιο 1**, επιχειρείται ένας ορισμός του Inpainting προβλήματος καθώς και μία αναφορά στις βασικές μεθόδους που χρησιμοποιούνται για την

επίλυσή του. Γίνεται ο διαχωρισμός μεταξύ περιεχομένου (structure), και υφής (texture) σε μία εικόνα καθώς το Inpainting πρόβλημα επιλύεται με διαφορετικούς τρόπους για εικόνες υφής και με διαφορετικούς για εικόνες περιεχομένου. Δείχνεται ότι το Inpainting πρόβλημα είναι ένα μη καλά ορισμένο πρόβλημα και ότι συνδέεται άμεσα με το θεμελιώδες πρόβλημα της επεξεργασίας εικόνας. Τέλος δείχνεται ότι οι βασικοί τρόποι επίλυσης του Inpainting προβλήματος μπορούν να αναχθούν σε διαφορετικούς τρόπους θεώρησης και μοντελοποίησης της έννοιας της εικόνας, γεγονός που δίνει ένα θεωρητικό έρεισμα στις μεθόδους Inpainting.

Στο **Κεφάλαιο 2**, γίνεται εκτενής αναφορά στα μοντέλα για Inpainting περιεχομένου. Τα μοντέλα αυτά υλοποιούνται από σχήματα μερικών διαφορικών εξισώσεων, (PDE σχήματα). Υπάρχουν δύο διαφορετικές κατηγορίες μεθόδων. Τα απευθείας PDE μοντέλα που μοντελοποιούν φυσικές διαδικασίες και τα PDE μοντέλα που προκύπτουν από μεταβολικές μεθόδους, (variational methods), και προέρχονται από θεωρήσεις της ενέργειας της εικόνας σε έναν κατάλληλο χώρο συναρτήσεων. Γίνεται αριθμητική υλοποίηση πολλών μοντέλων, παρουσιάζονται αποτελέσματα καθώς και η σύνδεση των δύο βασικών κατηγοριών σε ένα Inpainting μοντέλο. Τέλος παρουσιάζονται και άλλες εφαρμογές των Inpainting μοντέλων όπως η συμπίεση εικόνας και η αύξηση της ανάλυσης, (ψηφιακό zoom-in), μιας εικόνας.

Το **Κεφάλαιο 3** ασχολείται με την έννοια της υφής. Αρχικά γίνεται μια εισαγωγή στην έννοια της υφής και στα βασικά προβλήματα της ανάλυσής της. Από αυτά, ενδιαφέρον για την παρούσα εργασία παρουσιάζει το πρόβλημα της σύνθεσης υφής. Αφού εξετάζονται αλγόριθμοι σύνθεσης υφής δείχνεται ότι αυτοί μπορούν να χρησιμοποιηθούν και στο Inpainting πρόβλημα ενώ παράλληλα παρουσιάζονται και τροποποιήσεις αυτών, ειδικά για το Inpainting πρόβλημα. Στο τέλος του κεφαλαίου 3 αναφέρεται η μέθοδος του ταυτόχρονου Inpainting περιεχομένου και υφής, η οποία αποτελεί ίσως τη γενικότερη και αποτελεσματικότερη μέθοδο Inpainting μέχρι σήμερα.

Το **Κεφάλαιο 4** ασχολείται με το πρόβλημα του αυτόματου εντοπισμού και απομάκρυνσης κειμένου από εικόνες. Αρχικά παρουσιάζονται τα πλεονεκτήματα που έχει ο αυτόματος προσδιορισμός της Inpainting περιοχής. Στη συνέχεια παρουσιάζονται περιληπτικά οι βασικοί τρόποι αντιμετώπισης του προβλήματος του εντοπισμού της θέσης κειμένου σε μία εικόνα. Γίνεται η υλοποίηση μιας τέτοιας μορφολογικής μεθόδου και παρουσιάζονται αποτελέσματα. Τέλος παρουσιάζονται τα ιδιαίτερα χαρακτηριστικά του προβλήματος εντοπισμού της θέσης των γραμμάτων όταν αυτά πρέπει να απομακρυνθούν, και προτείνεται ένας νέος αλγόριθμος για την αυτόματη απομάκρυνση κειμένου από εικόνες.

Τέλος στο **Κεφάλαιο 5** γίνεται μια σύνοψη για την προσφορά της διπλωματικής εργασίας, ενώ παρουσιάζονται και κατευθύνσεις για μελλοντική έρευνα.

Ευχαριστίες

Η ενασχόλησή μου με το συγκεκριμένο θέμα της διπλωματικής μου εργασίας, καθώς και γενικότερα η ενασχόλησή μου με την όραση υπολογιστών οφείλεται στον καθηγητή Ε.Μ.Π. Πέτρο Μαραγκό. Θα ήθελα να τον ευχαριστήσω θερμά, για τη συμβολή του στην επιλογή του θέματος και στη διαμόρφωση των ερευνητικών μου ενδιαφερόντων καθώς και για την πολύτιμη βοήθειά του κατά τη διάρκεια εκπόνησης της διπλωματικής εργασίας.

Επίσης θα ήθελα να ευχαριστήσω τον υποψήφιο διδάκτορα Ε.Μ.Π. Γιώργο Παπανδρέου, για την πρόθυμη, άμεση και ουσιαστική του βοήθεια όποτε του ζητήθηκε κάτι τέτοιο.

Τέλος ιδιαίτερες ευχαριστίες οφείλω στην οικογένειά μου και σε όλα τα άτομα που βρίσκονται κοντά μου, για τη στήριξη που μου παρέχουν όλα αυτά τα χρόνια.

Περιεχόμενα

1	Θεμελίωση του Inpainting Προβλήματος και Σύνδεση του με το Θεμελιώδες Πρόβλημα της Επεξεργασίας Εικόνας	16
1.1	Το Inpainting Πρόβλημα	16
1.2	Διαχωρισμός Περιεχομένου - Υφής	17
1.3	Κατηγορίες Μεθόδων για το Inpainting Πρόβλημα	18
1.4	Το Θεμελιώδες Πρόβλημα της Επεξεργασίας Εικόνας	21
1.5	Δύο Μοντέλα Εικόνων για το Inpainting Πρόβλημα	22
1.5.1	Μοντέλο Εικόνας Φραγμένης Μεταβολής (Bounded Variation (BV) Image Model)	22
1.5.2	Μοντέλο Εικόνας Mumford-Shah	23
2	Inpainting Περιεχομένου με Μεθόδους Μερικών Διαφορικών Εξισώσεων	24
2.1	Χρήση των Μερικών Διαφορικών Εξισώσεων στην Ψηφιακή Επεξεργασία Εικόνας	24
2.2	Μοντέλο Ολικής Μεταβολής - Total Variation Model	25
2.2.1	Εισαγωγή	25
2.2.2	Χρήση του TV Φίλτρου για Inpainting	26
2.2.3	Αριθμητική Επίλυση	28
2.2.4	Αποτελέσματα	30
2.2.5	Άλλες Εφαρμογές του TV Φίλτρου	33
2.3	Επεκτάσεις του Μοντέλου TV Inpainting	38
2.3.1	Inpainting με Διαχύσεις Οδηγούμενες από την Καμπυλότητα	38
2.3.2	Το Μοντέλο Euler's Elastica Inpainting	40
2.4	Το Μοντέλο BCBS	44
2.4.1	Εξαγωγή του Μοντέλου - Ο Αλγόριθμος	45
2.4.2	Αριθμητική Υλοποίηση	46
2.4.3	Αποτελέσματα	48
2.5	Σύνδεση των Μοντέλων BCBS, CDD, Euler's Elastica Inpainting	51
2.6	Άλλα Μοντέλα για inpainting	52

2.6.1	Navier-Stokes, Μηχανική των Ρευστών και inpainting Εικόνας	52
2.6.2	Inpainting με Εξέλιξη Επιπεδοσυνόλων	54
2.6.3	Inpainting με Μεθόδους Κατάτμησης Εικόνας	55
3	Υφή	58
3.1	Εισαγωγή	58
3.2	Σύνθεση Υφής	60
3.2.1	Εισαγωγή - Πρώτες Προσπάθειες	60
3.2.2	Ο Αλγόριθμος των Efros-Leung	62
3.2.3	Επεκτάσεις του Αλγορίθμου Efros-Leung	65
3.3	Inpainting με Μεθόδους Σύνθεσης Υφής	66
3.3.1	Ο Αλγόριθμος Efros-Leung για Inpainting Εικόνων Υφής	67
3.3.2	Παραλλαγή του Αλγορίθμου Efros-Leung για Inpainting Υφής	71
3.3.3	Σύγκριση Μεταξύ Inpainting Υφής και Inpainting Πε- ριεχομένου	74
3.4	Ταυτόχρονο Inpainting Περιεχομένου και Υφής	76
3.4.1	Διάσπαση της Εικόνας σε Συνιστώσες Περιεχομένου και Υφής	77
3.4.2	Συσχέτιση του Inpainting Περιεχομένου και Υφής με τις Μεθόδους Συντήρησης Έργων Τέχνης	79
3.4.3	Ταυτόχρονο Inpainting Περιεχομένου και Υφής - Απο- τελέσματα	81
4	Αυτόματος Εντοπισμός και Απομάκρυνση Κειμένου από Εικό- νες	85
4.1	Αυτόματη Ανίχνευση Κειμένου σε Εικόνες	86
4.1.1	Συνοπτική Ανασκόπηση Μεθόδων	87
4.2	Εύρεση Κειμένου σε Εικόνες με Μεθόδους Μαθηματικής Μορ- φολογίας	89
4.2.1	Στοιχεία Μαθηματικής Μορφολογίας	89
4.2.2	Ανίχνευση Κειμένου	91
4.3	Αυτόματη Απομάκρυνση Κειμένου από Εικόνες	99
4.3.1	Αποτελέσματα	103
5	Συμπεράσματα	107
5.1	Συμβολή της Διπλωματικής Εργασίας	107
5.2	Κατευθύνσεις για Μελλοντική Έρευνα	108

Κατάλογος Σχημάτων

1.1	Διαχωρισμός περιεχομένου και υψής	18
1.2	Ο τρόπος Inpainting που ακολουθείται από τους συντηρητές έργων	19
2.1	Τοπολογία του inpainting προβλήματος με χρήση του TV μοντέλου	26
2.2	Ένα pixel και τα γειτονικά του σημεία	28
2.3	TV Inpainting και αφαίρεση θορύβου κατά μήκος ακμής	30
2.4	TV Inpainting σε πλατιές μπάρες	31
2.5	TV inpainting σε διαφορετικές περιπτώσεις	31
2.6	TV inpainting σε ακμή με καμπυλότητα	31
2.7	Αφαίρεση κειμένου με TV inpainting	33
2.8	Παράδειγμα TV zoom-in	35
2.9	Σύγκριση TV zoom-in με άλλες μεθόδους	35
2.10	Παράδειγμα κωδικοποίησης και συμπίεσης εικόνας	37
2.11	Το TV inpainting δεν ικανοποιεί την αρχή της συνεκτικότητας	38
2.12	Το CDD inpainting επανενώνει αρμονικά ακμές με καμπυλότητα	40
2.13	Τα διανύσματα \vec{n} και \vec{t}	42
2.14	Το μοντέλο elastic inpainting ικανοποιεί την αρχή της συνεκτικότητας.	43
2.15	Επίδραση του όρου b/a	43
2.16	Τα μοντέλα CDD και Euler's Elastica Inpainting δεν έχουν μοναδική λύση	44
2.17	Το μοντέλο BCBS διατηρεί την καμπυλότητα των ακμών	49
2.18	Το μοντέλο BCBS δεν λειτουργεί καλά σε περιοχές με υφή και μεγάλες περιοχές	49
2.19	TV και BCBS inpainting	50
2.20	Απομάκρυνση κειμένου με BCBS inpainting	50
2.21	Ομοιότητα μοντέλων TV, Mumford-Shah Inpainting	56

2.22	Παράδειγμα Mumford-Shah-Euler Inpainting: Αριστερα: Αρχική εικόνα, Κέντρο: Τελική εικόνα, Δεξιά: Τελική εικόνα της συνάρτησης z	57
3.1	Παραδείγματα εικόνων υφής	58
3.2	Παράδειγμα εικονας με υφή στον αληθινό κόσμο	59
3.3	Παράδειγμα σύνθεσης υφής	61
3.4	Ο αλγόριθμος των Efros-Leung	64
3.5	Παραδείγματα σύνθεσης υφής με τον αλγόριθμο των Efros-Leung	64
3.6	Σύνθεση υφής για διαφορετικές τιμές του παραθύρου	64
3.7	Επικάλυψη των δύο εικόνων	65
3.8	Παράδειγματα Inpainting υφής με τον αλγόριθμο Efros-Leung	68
3.9	Inpainting εικόνας υφής και περιεχομένου με τον αλγόριθμο Efros-Leung	69
3.10	Απομάκρυνση κειμένου με τον αλγόριθμο Efros-Leung	70
3.11	Συμβολισμός των διανυσμάτων	71
3.12	Το μοντέλο ικανοποιεί την αρχή της συνεκτικότητας	72
3.13	Απομάκρυνση μεγάλου αντικειμένου από εικόνα	73
3.14	Το Inpainting υφής χρειάζεται κατάλληλη περιοχή δείγμα για να αντιγράψει την κατάλληλη πληροφορία	75
3.15	Η έλλειψη παρόμοιας περιοχής για αντιγραφή πληροφορίας δεν δίνει καλά αποτελέσματα	75
3.16	Άλλο ένα παράδειγμα	75
3.17	Ο τρόπος Inpainting που ακολουθείται από τους συντηρητές έργων	80
3.18	Σχηματική αναπαράσταση του ταυτόχρονου Inpainting περιεχομένου και υφής	81
3.19	Παράδειγμα Inpainting περιεχομένου και υφής	82
3.20	Παράδειγμα αφαίρεσης κειμένου από έγχρωμη εικόνα με Inpainting περιεχομένου και υφής	83
3.21	Σύγκριση ταυτόχρονου Inpainting περιεχομένου και υφής με μεθόδους Inpainting μόνο υφής και περιεχομένου	84
4.1	Εικόνα με δύσκολο προσδιορισμό της Inpainting περιοχής από τον χρήστη	86
4.2	Αναλυτική παρουσίαση της μεθόδου εντοπισμού κειμένου	95
4.3	Αποτελέσματα για διαφορετικά μορφολογικά στοιχεία B	97
4.4	Παράδειγμα σε εικόνα με σκούρα γράμματα	98
4.5	Παράδειγμα σε εικόνα με έντονα μεταβλητό background	99
4.6	Αποτελέσματα χωρίς τη διαδικασία θόλωσης (blurring)	100

4.7	Η μορφή των γραμμάτων φαίνεται ξεκάθαρα, αλλά δεν έχει εντοπισθεί τελείως η θέση τους	100
4.8	Παράδειγμα reconstruction opening $\rho^-(M X)$	102
4.9	Παράδειγμα αυτόματου εντοπισμού και απομάκρυνση γραμμάτων με Inpainting υφής	104
4.10	Παράδειγμα αυτόματου εντοπισμού και απομάκρυνση γραμμάτων με ταυτόχρονο Inpainting περιεχομένου και υφής	105
4.11	Παράδειγμα αυτόματου εντοπισμού και απομάκρυνση γραμμάτων με Inpainting περιεχομένου	106

Κατάλογος Πινάκων

1.1	Βασικά προβλήματα της επεξεργασίας εικόνας	21
2.1	Αντιστοίχια μεγεθών κίνησης υγρού και εξέλιξης ισόφωτων καμ- πύλων	53

Κεφάλαιο 1

Θεμελίωση του Inpainting Προβλήματος και Σύνδεση του με το Θεμελιώδες Πρόβλημα της Επεξεργασίας Εικόνας

1.1 Το Inpainting Πρόβλημα

Με τον όρο inpainting εννοούμε την επεξεργασία μιας εικόνας με τρόπο τέτοιο ώστε ένας παρατηρητής που δεν ξέρει την αρχική εικόνα να μη μπορεί να παρατηρήσει ότι η εικόνα έχει υποστεί επεξεργασία. Η χρησιμότητα του inpainting είναι πολύ μεγάλη όπως φαίνεται και από τις παρακάτω εφαρμογές του:

- Συντήρηση αρχαίων πινάκων ζωγραφικής και ανάκτηση της αρχικής μορφής τους.
- Επεξεργασία φωτογραφιών με σκοπό την επιδιόρθωση ορισμένων λαθών (π.χ. ρωγμές, κόκκινα μάτια).
- Επιδιόρθωση λαθών που προκύπτουν από ασύρματη μετάδοση εικόνας (π.χ. χαμένα blocks σε εικόνες ή video frames).
- Απομάκρυνση κειμένου από εικόνες αλλά και αντικειμένων για τη δημιουργία ειδικών εφέ (special effects).
- Ψηφιακό zoom-in, υπερανάλυση εικόνας καθώς και συμπίεση εικόνας.

Ιστορικά, η επεξεργασία εικόνας για τον σκοπό της συντήρησης αρχαίων πινάκων ζωγραφικής υπήρχε από πολύ παλιά (π.χ. δεσ [18]). Στην σύγχρονη ψηφιακή επεξεργασία εικόνας το inpainting πρόβλημα ορίζεται ως εξής:

Έστω Ω η περιοχή μιας εικόνας u και $D \subset \Omega$ μια περιοχή της εικόνας όπου η πληροφορία λείπει ή είναι λάθος. Ο στόχος του inpainting είναι η ανάκτηση όλης της αρχικής εικόνας u βασιζόμενοι μόνο στην αρχική εικόνα που έχουμε $u_0|_{\Omega \setminus D}$. Η περιοχή D ονομάζεται και μάσκα του προβλήματος¹, ενώ η ανάκτηση της εικόνας πρέπει να γίνει με τρόπο που να μη γίνεται αντιληπτός από έναν εξωτερικό παρατηρητή. Η μάσκα D επιλέγεται από το χρήστη καθώς το τι θεωρείται λάθος σε μια εικόνα δεν είναι κάτι που μπορεί να οριστεί. Πάντως για συγκεκριμένα λάθη όπως ρωγμές σε φωτογραφίες υπάρχουν κάποιες αυτόματοι μέθοδοι εντοπισμού, π.χ. [4], ενώ συγκεκριμένα αντικείμενα όπως γράμματα μπορούν να εντοπισθούν με μεθόδους της αναγνώρισης προτύπων ή και άλλες απλούστερες μεθόδους.

Πριν προχωρήσουμε στην αναφορά των μεθόδων που χρησιμοποιούνται για την αντιμετώπιση του Inpainting προβλήματος, πρέπει να προχωρήσουμε σε έναν διαχωρισμό που θα διατηρήσουμε στο σύνολο της εργασίας.

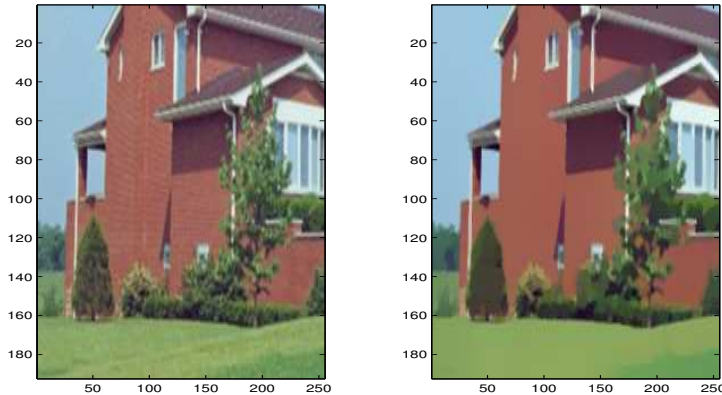
1.2 Διαχωρισμός Περιεχομένου - Υφής

Οι περισσότερες εικόνες του φυσικού κόσμου αποτελούνται από αντικείμενα τα οποία έχουν και περιεχόμενο και υφή. Με τον όρο περιεχόμενο, (*structure*), εννοούμε τη βασική μορφή του αντικειμένου, όπου η επιφάνεια του είναι ομογενής χωρίς να περιέχει λεπτομέρειες. Στο φυσικό κόσμο όμως, τα αντικείμενα όπως απεικονίζονται σε μία εικόνα σπάνια έχουν ομογενείς επιφάνειες. Οι λεπτομέρειες που υπάρχουν στις επιφάνειες των αντικειμένων αποτελούν την υφή, (*texture*), της εικόνας. Από τα παραπάνω γίνεται φανερό ότι διαχωρισμός των εννοιών περιεχομένου - υφής δεν είναι σαφής και καλά ορισμένος και έτσι γίνεται αντιληπτός από το ανθρώπινο μάτι.

Στην εικόνα 1.1 βλέπουμε ένα παράδειγμα. Στην αριστερή εικόνα βλέπουμε ότι οι επιφάνειες των αντικειμένων της εικόνας (σπίτι, γρασίδι) δεν είναι ομογενείς αλλά περιέχουν λεπτομέρειες που αντιστοιχούν στην υφή. Δεξιά βλέπουμε την απλοποιημένη εικόνα που προκύπτει από την αφαίρεση της υφής. Από την αφαίρεση των δύο εικόνων προκύπτει μία εικόνα που θεωρούμε ότι περιέχει μόνο υφή και ονομάζεται εικόνα υφής.

Κάτι άλλο που πρέπει να σημειώσουμε είναι ότι ο διαχωρισμός περιεχομένου και υφής σε μία εικόνα, εξαρτάται από την κλίμακα από την οποία παρατηρούμε την εικόνα. Λεπτομέρειες που τις χαρακτηρίζουμε ως υφή σε μία κλίμακα

¹Γίνεται αντιληπτό ότι για την απομάκρυνση ενός αντικειμένου η μάσκα πρέπει να οριστεί πάνω στο αντικείμενο.



Σχήμα 1.1: Διαχωρισμός περιεχομένου και υφής: Αριστερά έχουμε την πραγματική εικόνα που περιέχει και υφή ενώ δεξιά την απλοποιημένη εικόνα που έχει μόνο περιεχόμενο

παρουσιάζονται πλούσιες και σε περιεχόμενο αν παρατηρηθούν από πιο κοντά και γενικά ένα pixel δεν μπορεί να χαρακτηριστεί ως υφής ή περιεχομένου. Ο διαχωρισμός όμως υφής και περιεχομένου είναι πολύ σημαντικός για το Inpainting πρόβλημα καθώς υπάρχει διαφορετική αντιμετώπιση για εικόνες υφής και για εικόνες περιεχομένου όπως θα δούμε στη συνέχεια.

1.3 Κατηγορίες Μεθόδων για το Inpainting Πρόβλημα

Για την αντιμετώπιση του Inpainting προβλήματος για εικόνες υφής, χρησιμοποιούνται αλγόριθμοι που κατασκευάστηκαν για την επίλυση του συγγενούς προβλήματος της σύνθεσης υφής, (*texture synthesis*). Οι αλγόριθμοι αυτοί εκμεταλλεύονται γενικά χαρακτηριστικά της έννοιας της υφής (στοχαστικότητα, περιοδικότητα) και θα εξεταστούν αναλυτικά στο κεφάλαιο 3.

Για την αντιμετώπιση του Inpainting προβλήματος για εικόνες περιεχομένου μπορούμε να διακρίνουμε δύο διαφορετικές προσεγγίσεις:

1. Την προσομοίωση των τεχνικών που χρησιμοποιούνται από επαγγελματίες ζωγράφους που ασχολούνται με τη συντήρηση έργων τέχνης.
2. Μια Bayesian προσέγγιση και αντιμετώπιση του προβλήματος.

Πριν προχωρήσουμε αναλυτικά στην παρουσίαση Inpainting αλγορίθμων ας δούμε κάποια χαρακτηριστικά των 2 αυτών προσεγγίσεων

Τεχνικές συντήρησης πινάκων ζωγραφικής

Οι τεχνικές που εφαρμόζουν οι συντηρητές έργων τέχνης για την ανάκτηση κάποιου παλιού πίνακα συνοψίζονται στα εξής 4 βήματα[38].

1. Η ολική εικόνα καθορίζει το πως θα συμπληρωθεί και η Inpainting περιοχή D έτσι ώστε να διατηρείται η ενότητα του έργου.
2. Οι γραμμές - ακμές της εικόνας που κόβονται συνεχίζονται μέσα στην εικόνα.
3. Οι διαφορετικές περιοχές μέσα στην περιοχή D , όπως προέκυψαν από το προηγούμενο βήμα, συμπληρώνονται ανάλογα με το περιεχόμενο των ορίων τους.
4. Προστίθενται οι μικρές λεπτομέρειες της εικόνας που μπορούν να θεωρηθούν σαν την υφή της εικόνας.

Η διαδικασία αυτή φαίνεται και στην εικόνα 1.2



Σχήμα 1.2: Ο τρόπος Inpainting που ακολουθείται από τους συντηρητές έργων. Από αριστερά προς δεξιά: Αρχική εικόνα, συμπλήρωση ακμών, συμπλήρωση περιεχομένου, συμπλήρωση υφής.

Για την επίλυση του Inpainting προβλήματος με βάση τις παραπάνω τεχνικές μπορούμε να παρατηρήσουμε τα εξής:

- Για το πρώτο βήμα, η υπόλοιπη εικόνα δείχνει πως πρέπει να συμπληρωθεί η ζητούμενη περιοχή μόνο σε περιπτώσεις Inpainting υφής, όπως θα δούμε στο κεφάλαιο 3. Για Inpainting περιεχομένου (structure inpainting) η ανάκτηση της ζητούμενης περιοχής γίνεται κυρίως από πληροφορία που βρίσκεται σε μια μικρή περιοχή γύρω από την Inpainting περιοχή. Όμως πρόσφατα έχουν αναπτυχθεί μέθοδοι για Inpainting περιεχομένου που βασίζονται σε πληροφορία που λαμβάνεται από ολόκληρη την εικόνα (π.χ. δεσ [55], [2]). Οι μέθοδοι αυτές δίνουν ενθαρρυντικά αποτελέσματα και θα αποτελέσουν σημείο μελλοντικής έρευνας.

- Όσο αφορά τα δύο επόμενα βήματα, στην προσπάθεια προσομοίωσης αυτών βασίζεται και ο πρώτος Inpainting αλγόριθμος [38], γνωστός και ως BCBS μοντέλο. Το μοντέλο αυτό παρουσιάζεται στην ενότητα 2.4.
- Το τέταρτο βήμα αναφέρεται αποκλειστικά στην υφή και αντιμετωπίζεται με μεθόδους σύνθεσης ύφης καθώς οι Inpainting αλγόριθμοι περιεχομένου δίνουν λεία αποτελέσματα και συνεπώς, αποτυγχάνουν στο Inpainting υφής. Οι αλγόριθμοι για σύνθεση υφής όπως έχουμε αναφέρει θα παρουσιαστούν στο κεφάλαιο 3.

Bayesian προσέγγιση του Inpainting προβλήματος²

Στη φύση όταν παρατηρούμε κάτι, σπάνια βλέπουμε τα αντικείμενα ολόκληρα. Τις περισσότερες φορές ένα αντικείμενο μπαίνει μπροστά από κάποιο άλλο (οσζλυσιον). Παρόλα αυτά, το ανθρώπινο μάτι μπορεί να καταλάβει τα αντικείμενα των οποίων κάποιο μέρος κρύβεται, ενώ ταυτόχρονα αντιλαμβάνεται την εικόνα ως μία οντότητα παρά ως το ένα σύνολο διακριτών και ανεξάρτητων αντικειμένων. Το γεγονός αυτό σύμφωνα με τους επιστήμονες της ψυχολογίας της όρασης και της νόησης δίνει βάση στην άποψη ότι οι άνθρωποι διαρκώς εφαρμόζουν τον κανόνα του Bayes και παίρνουν αποφάσεις σύμφωνα με αυτόν.

Με μια Bayesian θεώρηση, το Inpainting πρόβλημα μπορεί να θεωρηθεί ως η προσπάθεια μεγιστοποίησης της παρακάτω a-posteriori πιθανότητας

$$p(u|u_0, D) = \frac{p(u_0|u, D)p(u|D)}{p(u_0|D)} \quad (1.1)$$

Στις περισσότερες περιπτώσεις μπορούμε να θεωρήσουμε ότι η περιοχή D στην οποία υπάρχουν λάθη που θέλουμε να διορθώσουμε είναι ανεξάρτητη του περιεχομένου της εικόνας, δηλαδή $p(u|D) = p(u)$, ενώ η ποσότητα $p(u_0|D)$ είναι μια σταθερά κανονικοποίησης. Με αυτά υπόψη και εισάγωντας μεγέθη εντροπίας, $E = -1/\beta \ln p$ η ποσότητα που θέλουμε να μεγιστοποιήσουμε γίνεται

$$E[u|u_0, D] = E[u_0|u, D] + E[u] \quad (1.2)$$

Αν $\beta = 1/(kT)$ με k την σταθερά Boltzmann και T την απόλυτη θερμοκρασία, τότε προκύπτει μία κατανομή Gibbs [21] από την οποία μπορούμε να συνδέσουμε την εντροπία με ένα είδος ενέργειας.

Το κλειδί για το Inpainting πρόβλημα είναι ένας κατάλληλος ορισμός του a-priori μοντέλου $p(u)$ ή $E(u)$. Το πρόβλημα αυτό συνδέεται άμεσα με το θεμελιώδες πρόβλημα της επεξεργασίας εικόνας (Fundamental problem of Image Processing).

²Η παράγραφος αυτή έχει βασιστεί στο [57]

T	U_0	U
Αφαίρεση θορύβου και θολώματος	$U_0 = Ku + n$	$U = u$
Inpainting	$u_0 _{\Omega \setminus D}$	$u _{\Omega}$
κατάτμηση	u_0	αντικείμενα $[u_k, \Omega_k]$, $k = 1, 2, \dots$
ανάλυση σε πολλαπλές κλίμακες	u_0	εικόνες σε πολλές κλίμακες $\{u_{\lambda_1}, u_{\lambda_2}, \dots\}$
εκτίμηση κίνησης	εικόνες $\{u_0^1, u_0^2, \dots\}$	ροές κίνησης $\{\vec{v}_1, \vec{v}_2, \dots\}$

Πίνακας 1.1: Βασικά προβλήματα της επεξεργασίας εικόνας

1.4 Το Θεμελιώδες Πρόβλημα της Επεξεργασίας Εικόνας

Το θεμελιώδες πρόβλημα της επεξεργασίας εικόνας αναζητεί το τι πραγματικά είναι μια εικόνα, με μαθηματικούς όρους, πως δημιουργείται και πως επεξεργάζεται. Ένα πρόβλημα επεξεργασίας εικόνας μπορεί γενικά να θεωρηθεί ως ένα σύστημα T εισόδου-εξόδου:

$$U = T[U_0] \quad (1.3)$$

όπου U_0 είναι η αρχική εικόνα ή μια σειρά από εικόνες, T είναι το σύστημα που επιτελεί την επιθυμητή επεξεργασία και U είναι η έξοδος του συστήματος. Στον πίνακα 1.1 φαίνονται ορισμένα από τα βασικά προβλήματα της επεξεργασίας εικόνας (Από το [59]).

Για την επίλυση των προβλημάτων του πίνακα 1.1 είναι σημαντικό να ορίσουμε έναν τρόπο μοντελοποίησης της εικόνας έτσι ώστε να ορίσουμε το κατάλληλο σύστημα T .

Για τη μοντελοποίηση της εικόνας υπάρχουν 3 προσεγγίσεις, που αναφέρονται παρακάτω μαζί με το πως συνδέονται με το Inpainting πρόβλημα που μας ενδιαφέρει.

- **Φυσική Προσομοίωση** Σύμφωνα με αυτή τη θεώρηση, οι εικόνες μοντελοποιούνται ως προσομοιώσεις φυσικών, χημικών ή βιολογικών διαδικασιών (π.χ. από τις εξισώσεις Navier-Stokes). Στην προσέγγιση αυτή υπάγεται η μέθοδος Navier-Stokes inpainting που θα δούμε στην ενότητα 2.6.1, ενώ μπορούμε να θεωρήσουμε ότι εκεί υπάγεται και η μέθοδος που προσομοιώνει τις τεχνικές των συντηρητών τέχνης αν θεωρήσουμε αυτή τη διαδικασία ως φυσική.

- **Τυχαία Πεδία (Random Fields)** Σύμφωνα με αυτή τη θεώρηση, οι εικόνες δημιουργούνται από δείγματα που παίρνονται από διάφορα τυχαία πεδία (Markov, Gibbs) τα οποία ακολουθούν κάποια συνάρτηση πυκνότητας πιθανότητας. Η θεώρηση αυτή είναι πολύ χρήσιμη για τη μοντελοποίηση εικόνων υφής για τις οποίες θα αναφερθούμε στο κεφάλαιο 3, ενώ μπορεί να χρησιμοποιηθεί και για μεθόδους Inpainting με στατιστική επεξεργασία εικόνας [2].
- **Χώροι Συναρτήσεων (Function Spaces)** Με αυτή τη θεώρηση, χρησιμοποιούνται κάποιοι κατάλληλοι χώροι συναρτήσεων για να μοντελοποιήσουν τις εικόνες με κάποιο μέτρο ενέργειας $E[u]$. Η θεώρηση αυτή, όπως φαίνεται και από τα παραπάνω βοηθάει πολύ για τη μοντελοποίηση του Inpainting προβλήματος ως ένα πρόβλημα απόφασης Bayes, και θα την εξετάσουμε αναλυτικότερα.

1.5 Δύο Μοντέλα Εικόνων για το Inpainting Πρόβλημα

Επειδή οι ακμές παίζουν πολύ σημαντικό ρόλο στο Inpainting πρόβλημα, αλλά και γενικότερα στην ανθρώπινη αντίληψη της όρασης, οι συναρτησιακοί χώροι που χρησιμοποιούνται για το Inpainting πρόβλημα τονίζουν ιδιαίτερα τις ακμές.

1.5.1 Μοντέλο Εικόνας Φραγμένης Μεταβολής (Bounded Variation (BV) Image Model)

Αρχικά και για λόγους πληρότητας δίνουμε τον ορισμό των BV συναρτήσεων. Για περισσότερες πληροφορίες ο αναγνώστης μπορεί να δει το [34] ή κάποιο βιβλίο λογισμού μεταβολών (variation calculus).

Ορισμός 1.1. Έστω Q ένα ανοικτό σύνολο. Μία συνάρτηση $u \in L^1(Q)$ λέγεται συνάρτηση φραγμένης μεταβολής, (*Bounded Variation*), αν οι μερικές της παράγωγοι ανήκουν σε κατανομές με πεπερασμένη ολική μεταβολή, (*Total Variation*), στο Q . Η τάξη αυτών των συναρτήσεων συμβολίζεται $BV(Q)$.

Μαθηματικά για μία συνάρτηση u είναι

$$u \in BV(Q) \Leftrightarrow \exists \mu_1, \dots, \mu_N : \int_Q u D_i \phi dx = - \int_Q \phi d\mu_i \quad \forall \phi \in C_0^\infty(Q), i = 1, \dots, N$$

όπου μ_1, \dots, μ_N είναι μέτρα Radon, (Radon measures), ορισμένα στο Q με πεπερασμένη μάζα στο Q .

Η ενέργεια μιας εικόνας σύμφωνα με το BV μοντέλο είναι [36]

$$E[u] = \int_{\Omega} |\nabla u| dx dy \quad (1.4)$$

Το Inpainting μοντέλο που προκύπτει από αυτή τη θεώρηση, ονομάζεται μοντέλο ολικής μεταβολής (Total Variation (TV) Inpainting Model) και είναι

$$E_{tv}[u|u_0, D] = \int_{\Omega} |\nabla u| dx dy + \frac{1}{2} \int_{\Omega \setminus D} \lambda_D (u - u^0)^2 dx dy \quad (1.5)$$

με λ_D έναν πολλαπλασιαστή Lagrange για τον οποίο ισχύει

$$\lambda_D = \begin{cases} \lambda, & (x, y) \in \Omega \setminus D \\ 0, & (x, y) \in D \end{cases} \quad (1.6)$$

1.5.2 Μοντέλο Εικόνας Mumford-Shah

Για να τονισθεί περισσότερο ο ρόλος των ακμών, οι Mumford, Shah προτείνουν ένα μοντέλο για την ενέργεια μιας εικόνας που αντιμετωπίζει ξεχωριστά τις ακμές (free boundary model)[47].

Αν Γ το σύνολο των ακμών της εικόνας τότε η εξίσωση ενέργειας - εντροπίας (1.2) γίνεται

$$E[u, \Gamma|u_0, D] = E[u|\Gamma] + E[\Gamma] + E[u_0|u, D], \quad (1.7)$$

από την οποία προκύπτει το Inpainting μοντέλο (Mumford - Shah Inpainting Model)

$$J[u, \Gamma|u^0, D] = \frac{1}{2} \int_{\Omega} \lambda_D (u - u^0)^2 dx + \frac{\gamma}{2} \int_{\Omega \setminus \Gamma} |\nabla u|^2 dx + \alpha H^1(\Gamma) \quad (1.8)$$

όπου H^1 είναι το μέτρο Hausdorff το οποίο για συνήθης καμπύλες είναι ίσο με το μήκος των καμπυλών.

Η αριθμητική επίλυση των παραπάνω μοντέλων Inpainting γίνεται με αρχές του λογισμού μεταβολών (calculus variation). Παίρνοντας την πρώτη Euler-Lagrange μεταβολική παράγωγο προκύπτουν υπολογιστικά σχήματα μερικών διαφορικών εξισώσεων τα οποία συγκλίνουν στην επιθυμητή λύση. Τα μοντέλα αυτά θα εξεταστούν αναλυτικά στο κεφάλαιο 2.

Κεφάλαιο 2

Inpainting Περιεχομένου με Μεθόδους Μερικών Διαφορικών Εξισώσεων

2.1 Χρήση των Μερικών Διαφορικών Εξισώσεων στην Ψηφιακή Επεξεργασία Εικόνας

Η χρήση των μερικών διαφορικών εξισώσεων (Partial Differential Equations - PDEs), στην όραση υπολογιστών είναι μία σχετικά σύγχρονη τάση, καθώς η επεξεργασία εικόνας ήταν παραδοσιακά συνδεδεμένη με μεθόδους της ψηφιακής επεξεργασίας σημάτων (μορφολογική επεξεργασία, διακριτοί μετασχηματισμοί, διακριτά πεδία Markov κ.λ.π.).

Τα τελευταία χρόνια όμως υπάρχει μια ευρεία χρήση των μερικών διαφορικών εξισώσεων στην όραση υπολογιστών και στην επεξεργασία εικόνας¹.

Ένα βασικό χαρακτηριστικό των PDEs είναι ότι αυτά αποτελούν συνεχή και όχι διακριτά μοντέλα. Έτσι η μοντελοποίηση των προβλημάτων γίνεται σε συνεχείς χώρους και η διακριτοποίηση εισέρχεται μόνο κατά την αριθμητική αναπαράσταση και επίλυση στον υπολογιστή.

Ορισμένα από τα πλεονεκτήματα που παρουσιάζουν τα συνεχή μοντέλα και ειδικότερα τα PDEs είναι τα εξής:

- Απλότητα στην επεξεργασία των μοντέλων καθώς αυτά δεν εξαρτώνται από κάποιο διακριτό πλέγμα.
- Σύνδεση των προβλημάτων με άλλα κλασσικά προβλήματα μερικών διαφορικών εξισώσεων π.χ. της φυσικής, και χρησιμοποίηση έτοιμων θεω-

¹Για ιστορικά στοιχεία της χρήσης των μερικών διαφορικών εξισώσεων στην όραση υπολογιστών ο αναγνώστης μπορεί να δει τα [41], [62]

ρητικών αποτελεσμάτων (π.χ. ύπαρξης και μοναδικότητας λύσης), καθώς τέτοια προβλήματα έχουν μελετηθεί για πολλά χρόνια.

- Καλύτερη προσέγγιση πραγματικών φαινομένων καθώς η διακριτοποίηση (που εισάγει κάποια προσέγγιση), γίνεται μόνο στο τέλος της επεξεργασίας.
- Χρήση φυσικών μεγεθών και εκμετάλλευση φυσικών νόμων για δημιουργία νέων μοντέλων της όρασης υπολογιστών.
- Εκμετάλλευση αποτελεσμάτων της αριθμητικής ανάλυσης για τη διακριτοποίηση των συνεχών μοντέλων.

Η χρήση των συνεχών μοντέλων των μερικών διαφορικών εξισώσεων αποδείχθηκε ότι είναι πολύ χρήσιμη για τη μη-γραμμική επεξεργασία εικόνας και έδωσε νέες δυνατότητες στην όραση υπολογιστών. Χαρακτηριστικά παραδείγματα της μη-γραμμικής επεξεργασίας εικόνας με χρήση PDEs είναι η μη-γραμμική εξίσωση ανισοτροπικής διάχυσης και καθώς και η ανάλυση εικόνας σε χώρους κλιμακας, (scale-spaces), ενώ όπως θα δούμε και σε αυτό το κεφάλαιο το Inpainting εικόνας περιεχομένου γίνεται με τη χρήση μη-γραμμικών μοντέλων των μερικών διαφορικών εξισώσεων και του λογισμού μεταβολών (variational calculus).

2.2 Μοντέλο Ολικής Μεταβολής - Total Variation Model

2.2.1 Εισαγωγή

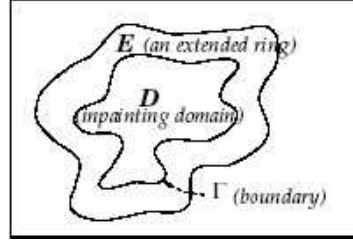
Το μοντέλο της ολικής μεταβολής Total Variation – TV για inpainting αποτελεί μια γενίκευση του αντίστοιχου φίλτρου total variation TV filter που χρησιμοποιείται με επιτυχία για την αποκατάσταση θορυβώδους εικόνας [61],[52],[36].

Στην θορυβώδη εικόνα υποθέτουμε ότι ο θόρυβος $n(x)$ είναι τυχαίος με μηδενική μέση τιμή και διασπορά σ^2

$$\mathbf{E}[n(x)] = 0, \quad \mathbf{E}[n^2(x)] = \sigma^2 \quad (2.1)$$

Έστω $u^0(x)$ η θορυβώδης εικόνα και $u(x)$ η εικόνα χωρίς θόρυβο. Τότε:

$$u^0(x) = u(x) + n(x) \quad (2.2)$$



Σχήμα 2.1: Τοπολογία του inpainting προβλήματος με χρήση του TV μοντέλου

Η διαφορά του TV φίλτρου με τις μεθόδους ελαχίστων τετραγώνων για αποκατάσταση εικόνας (π.χ. [23]) είναι ότι επιχειρείται η ελαχιστοποίηση της ολικής μεταβολής (νόρμας L^1 αντί για τη νόρμα L^2).

$$TV[u] = \int_{\Omega} |\nabla u| dx dy, \quad (2.3)$$

όπου Ω είναι το πεδίο ορισμού του σήματος – εικόνας, ∇u είναι η παράγωγος και $dx dy$ η στοιχειώδης επιφάνεια του Ω . Η μη γραμμικότητα που εισάγεται στο μοντέλο από την L^1 νόρμα, κάνει το μοντέλο πιο πολύπλοκο υπολογιστικά αλλά το κέρδος από τα αποτελέσματα είναι τεράστιο. Από τις υποθέσεις που κάναμε για το θόρυβο εξ. (2.1) προκύπτει ότι για την ελαχιστοποίηση της TV νόρμας πρέπει να ικανοποιούνται οι παρακάτω συνθήκες:

$$\int_{\Omega} u dx dy = \int_{\Omega} u^0 dx dy, \quad \frac{1}{|\Omega|} \int_{\Omega} (u - u^0)^2 dx dy = \sigma^2, \quad (2.4)$$

όπου $|\Omega|$ είναι το εμβαδόν του πεδίου ορισμού της εικόνας. Έτσι, οι εξισώσεις (2.3) και (2.4) όριζουν ένα πρόβλημα βελτιστοποίησης με συνθήκες. Εισάγοντας έναν πολλαπλασιαστή Laplace λ , χρειάζεται να ελαχιστοποιήσουμε το παρακάτω συναρτησιακό ενέργειας:

$$J[u] = \int_{\Omega} |\nabla u| dx dy + \frac{\lambda}{2} \int_{\Omega} (u - u^0)^2 dx dy. \quad (2.5)$$

2.2.2 Χρήση του TV Φίλτρου για Inpainting

Το συναρτησιακό της παραπάνω εξίσωσης μπορεί να χρησιμοποιηθεί και για inpainting σε μια μικρή περιοχή χωρίς υφή. Έστω D η περιοχή inpainting με όριο την τμηματικά λεία καμπύλη Γ και E μια καθορισμένη κλειστή περιοχή στο συμπλήρωμα D^c , έτσι ώστε $\Gamma \in E \cup D$ (Σχήμα 2.1)².

²Η τοπολογία αυτή θα διατηρηθεί σε όλη την εργασία για Inpainting περιεχομένου. Επίσης για τις τιμές φωτεινότητας των pixels θα χρησιμοποιείται ο συμβολισμός $u(i, j)$.

Η λύση του προβλήματος TV inpainting, σε αντιστοιχία με την αποθορυβοποίηση, (εξ. (2.5)), ισοδυναμεί με την ελαχιστοποίηση του παρακάτω συναρτησιακού: [14]

$$J_\lambda[u] = \int_{E \cup D} |\nabla u| dx dy + \frac{\lambda}{2} \int_E |u - u^0|^2 dx dy. \quad (2.6)$$

Για την επίλυση του παραπάνω προβλήματος ελαχιστοποίησης, χρειάζεται να βρούμε που μηδενίζεται η Euler-Lagrange παράγωγος του συναρτησιακού $J_\lambda[u]$.

Η Euler-Lagrange παράγωγος για το συναρτησιακό J_λ είναι

$$\frac{\partial J_\lambda[u]}{\partial u} = -\nabla \left(\frac{\nabla u}{|\nabla u|} \right) + \lambda_e (u - u^0) \quad (2.7)$$

για όλα τα σημεία $z = (x, y) \in E \cup D$ συν μια συνοριακή συνθήκη τύπου Neumann. Ο πολλαπλασιαστής Lagrange δίνεται απο τη σχέση

$$\lambda_e = \begin{cases} \lambda & z \in E \\ 0 & z \in D \end{cases}$$

Η μερική διαφορική εξίσωση που προκύπτει είναι

$$\frac{\partial u}{\partial t} = -\nabla \left(\frac{\nabla u}{|\nabla u|} \right) + \lambda_e (u - u^0) \quad (2.8)$$

Στην αριθμητική επίλυση της εξίσωσης (2.8), για να αποφύγουμε τη διαίρεση με το μηδέν σε περιοχές με μηδενική παράγωγο αντικαθιστούμε τον όρο

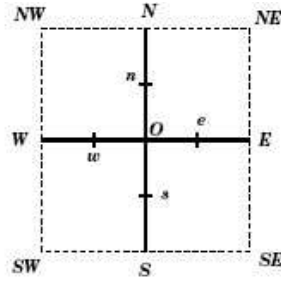
$$\nabla \left(\frac{\nabla u}{|\nabla u|} \right) \quad \text{με} \quad \nabla \left(\frac{\nabla u}{|\nabla u|_\alpha} \right) \quad \text{όπου} \quad |s|_\alpha := \sqrt{s^2 + \alpha^2} \quad (2.9)$$

και α ένας μικρός θετικός αριθμός.

Η παράμετρος α πέρα από τη βοήθεια στους υπολογισμούς παίζει και έναν ρόλο κατωφλίου. Σε λείες περιοχές όπου $|\nabla u| \ll \alpha$ είναι $|\nabla u|_\alpha \simeq \alpha$ και το μοντέλο προσομοιώνει τα γραμμικά μοντέλα νόρμας L^2 ενώ όταν $|\nabla u| \gg \alpha$, δηλαδή κατά μήκος των ακμών, το μοντέλο εφαρμόζεται κανονικά.

Η παράμετρος λ εκφράζει ποσοτικά (όπως θα δούμε στην παράγραφο 2.2.3) το κατά πόσο θα αλλάζουν τιμές τα pixels που ανήκουν στην περιοχή E (γύρω από την inpainting περιοχή). Το TV φίλτρο όπως είπαμε στην αρχή είναι φίλτρο αφαίρεσης θορύβου όποτε ταυτόχρονα με το inpainting μπορούμε να έχουμε και αποθορυβοποίηση στην περιοχή E . Σε περίπτωση απουσίας θορύβου θέτουμε $\lambda = \infty$ που σημαίνει ότι η περιοχή E παραμένει αναλλοίωτη ενώ αν υπάρχει και θόρυβος η παράμετρος λ παίρνει μια πεπερασμένη τιμή.

Για την απόδειξη της ύπαρξης λύσης του TV Inpainting μοντέλου διατυπώθηκαν τα δύο παρακάτω θεωρήματα.



Σχήμα 2.2: Ένα pixel και τα γειτονικά του σημεία

Θεώρημα 2.1. Υπαρξη βέλτιστης λύσης σε εικόνα με απουσία θορύβου
Έστω u_0 μία εικόνα στο χώρο $BV(\Omega)$ που παίρνει τιμές στο διάστημα $[0 \dots 1]$.
Τότε το *TV Inpainting* μοντέλο $TV(u) = \int_{\Omega} |\nabla u|$ με $u|_{\Omega \setminus D} = u_0|_{\Omega \setminus D}$ έχει
τουλάχιστον μία βέλτιστη λύση στο διάστημα $[0 \dots 1]$.

Θεώρημα 2.2. Υπαρξη βέλτιστης λύσης σε εικόνα με θορύβο
Έστω μία εικόνα u_0 στο $\Omega \setminus D$. Τότε υποθέτουμε ότι υπάρχει μία εικόνα u_c στο
 Ω που ανήκει στον χώρο $BV(\Omega)$ και η οποία ικανοποιεί την συνθήκη

$$\frac{1}{|\Omega \setminus D|} \int_{\Omega \setminus D} (u_c - u_0)^2 dx dy = \sigma^2$$

(δηλαδή η εικόνα που προκύπτει με καθαρισμό του θορύβου). Τότε υπάρχει
τουλάχιστον μία εικόνα u που κάνει βέλτιστο *Inpainting* στην περιοχή D και
αφαίρεση του θορύβου στην περιοχή $\Omega \setminus D$.

Η απόδειξη των παραπάνω θεωρημάτων υπάρχει στο [60].

2.2.3 Αριθμητική Επίλυση

Ένα από τα πλεονεκτήματα του μοντέλου ολικής μεταβολής είναι ότι υπάρχει
λύση, όχι όμως μοναδική, μέσω ενός επαναληπτικού αλγορίθμου. Η λύση αυτή
είναι πιο γρήγορη από τη λύση με το PDE της (2.8), η οποία συνήθως αργεί
λόγω της τιμής της χρονικής παραμέτρου που συνήθως είναι πολύ μικρή λόγω
των περιορισμών που εισέρχονται για ευστάθεια του προβλήματος.

Θεωρούμε ένα pixel O , E , N , W , S τα 4 γειτονικά του σημεία, και e , n , w , s
τα 4 αντίστοιχα ενδιάμεσα σημεία τα οποία δεν υπάρχουν στο διακριτό πλέγμα
της εικόνας αλλά θα μας διευκολύνουν στους υπολογισμούς. (Σχήμα 2.2)

Έστω $\mathbf{v} = (u^1, u^2) = \frac{\nabla u}{|\nabla u|}$. Τότε η απόκλιση σύμφωνα και με το παραπάνω
σχήμα μπορεί να θεωρηθεί κατά προσέγγιση

$$\nabla \cdot \mathbf{v} = \frac{\partial u^1}{\partial x} + \frac{\partial u^2}{\partial y} \simeq \frac{u_e^1 - u_w^1}{h} + \frac{u_n^2 - u_s^2}{h}, \quad (2.10)$$

όπου h είναι η διακριτική ικανότητα πάντα ίση με 1. Οι προσεγγίσεις στα ενδιάμεσα σημεία που δεν αντιστοιχούν σε πραγματικά σημεία της εικόνας είναι ως εξής

$$u_e^1 = \frac{1}{|\nabla u_e|} \left[\frac{\partial u}{\partial x} \right]_e \simeq \frac{1}{|\nabla u_e|} \frac{u_E - u_O}{h}, \quad (2.11)$$

$$|\nabla u_e| \simeq \frac{1}{h} \sqrt{(u_E - u_O)^2 + [(u_{NE} + u_N - u_S - u_{SE})/4]^2}, \quad (2.12)$$

$$\left[\frac{\partial u}{\partial x} \right] \simeq \frac{u_E - u_O}{h}, \quad (2.13)$$

$$\left[\frac{\partial u}{\partial y} \right] \simeq \frac{u_{NE} - u_{SE} + u_N - u_S}{4h}. \quad (2.14)$$

Πάρομοιες εκφράσεις με τις (2.11), (2.12), (2.13) και (2.14) μπορούν να εξαχθούν και για τα άλλα ενδιάμεσα σημεία n, w, s .

Έτσι για ένα σημείο O η εξ. (2.7) γίνεται

$$0 = \sum_{P \in \Lambda_O} \frac{1}{|\nabla u_p|} (u_O - u_P) + \lambda_e(O) (u_O - u_O^0), \quad (2.15)$$

όπου, π.χ., αν $P = E$, τότε $p \rightarrow e$. Αν ορίσουμε

$$w_P = \frac{1}{|\nabla u_p|}, \quad P \in \Lambda_O, \quad (2.16)$$

$$h_{OP} = \frac{w_P}{\sum_{P \in \Lambda_O} w_P + \lambda_e(O)}, \quad (2.17)$$

$$h_{OO} = \frac{\lambda_e(O)}{\sum_{P \in \Lambda_O} w_P + \lambda_e(O)}, \quad (2.18)$$

τότε η εξ. (2.15) γίνεται

$$u_O = \sum_{P \in \Lambda_O} h_{OP} u_P + h_{OO} u_O^0, \quad (2.19)$$

όπου ισχύει ότι

$$\sum_{P \in \Lambda_O} h_{OP} + h_{OO} = 1 \quad (2.20)$$

Η μη γραμμική εξίσωση (2.19) σε συνδιασμό με τη συνθήκη (2.20) είναι σε μορφή ενός βαθυπερατού φίλτρου. Για την αριθμητική επίλυση επιλέγουμε την επαναληπτική μέθοδο Gauss-Jacobi:

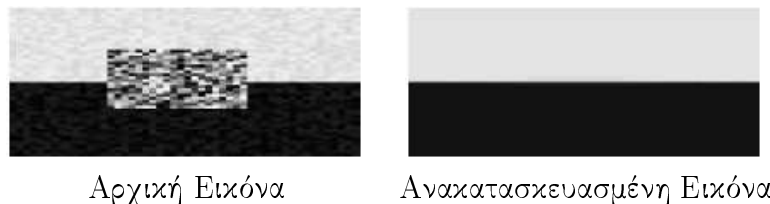
$$u_O^{(n)} = \sum_{P \in \Lambda_O} h_{OP}^{(n-1)} u_P^{(n-1)} + h_{OO}^{(n-1)} u_O^{(n-1)}, \quad (2.21)$$

όπου $h^{(n-1)} = h(u^{(n-1)})$. Αφού το h είναι βαθυπερατό φίλτρο, ο επαναληπτικός αλγόριθμος είναι ευσταθής. Συγκεκριμένα, το σύνολο τιμών $[0, 1]$ διατηρείται καθ' ολη την επαναληπτική διαδικασία.

2.2.4 Αποτελέσματα

Το μοντέλο TV inpainting είναι αποτελεσματικό όταν εφαρμόζεται σε μικρές περιοχές, χωρίς υφή. Μπορεί να επανενώσει ακμές, αλλά λόγω του ότι το συναρτησιακό ενέργειας δεν περιέχει κάποιον όρο μεγαλύτερης τάξης που να αντιστοιχεί στην καμπυλότητα curvature δεν επανενώνει σωστά ακμές με καμπυλότητα. Επεκτάσεις του μοντέλου με υψηλότερους όρους καμπυλότητας δίνονται στην ενότητα 2.3.

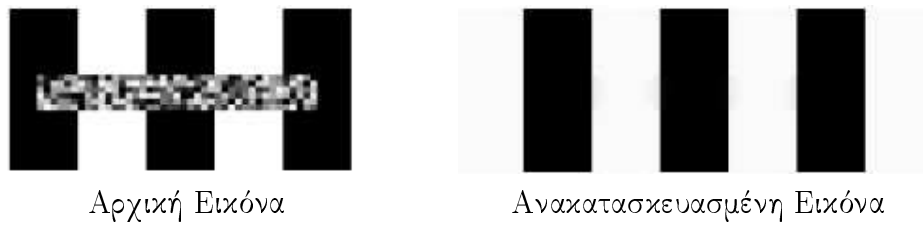
Στα παραδείγματα που ακολουθούν επιλέξαμε $\alpha = 0.01$. Γενικά παρατηρήσαμε ότι όσο μεγαλύτερη είναι η τιμή της παραμέτρου α τόσο πιο γρήγορα συγκλίνει ο αλγόριθμος. Όπως όμως αναφέραμε στην ενότητα 2.2.2 η παράμετρος α παίζει και τον ρόλο ενός κατωφλίου κατώ από το οποίο το μοντέλο συμπεριφέρεται σαν ένα μοντέλο νόρμας L^2 που δεν είναι ικανοποιητικά. Έτσι μια μεγάλη τιμή της α οδηγεί σε μη επιθυμητά αποτελέσματα. Στα περισσότερα παραδείγματα που ακολουθούν οι εικόνες δεν είχαν θόρυβο. Για αυτές τις εικόνες η παράμετρος λ ήταν πολύ μεγάλη (10^6). Όπου υπήρχε θόρυβος η παράμετρος λ επιλέχθηκε πολύ πιο μικρή (10).



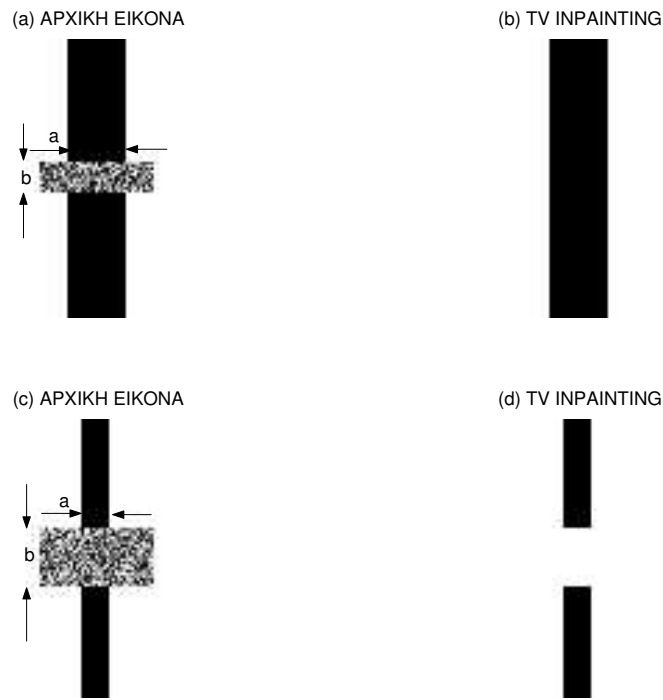
Σχήμα 2.3: TV Inpainting και αφαίρεση θορύβου κατά μήκος ακμής

Στα σχήματα 2.3 και 2.4 παρατηρούμε ότι το TV inpainting επανενώνει τις ακμές (που είναι ευθείες και πολύ έντονες) ενώ παράλληλα αφαιρεί και του θόρυβο που υπάρχει γύρω από την περιοχή όπου γίνεται inpainting (2.3).

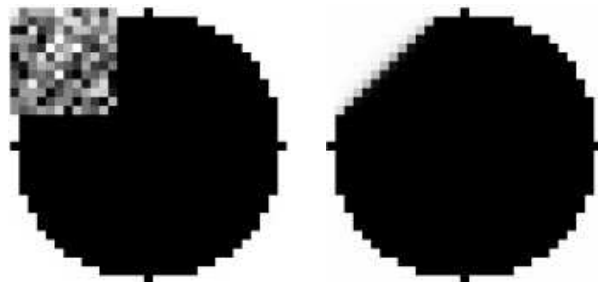
Στην εικόνα 2.5 παρατηρούμε ότι στο πρώτο παράδειγμα (εικόνες (a) και (b)) η μαύρη ακμή επανενώνεται κανονικά ενώ στο δεύτερο παράδειγμα (εικόνες



Σχήμα 2.4: TV Inpainting σε πλατιές μπάρες



Σχήμα 2.5: TV inpainting σε διαφορετικές περιπτώσεις



Σχήμα 2.6: TV inpainting σε ακμή με καμπυλότητα

(c) και (d)) η μαύρη ακμή δεν επανενώνεται. Το αν θα γίνει η επανένωση εξαρτάται από τη σχέση του πάχους a της μαύρης ακμής και του πάχους b της υπό inpainting περιοχής. Στην πρώτη περίπτωση έχουμε $a > b$ και η ακμή επανενώνεται ενώ στη δεύτερη περίπτωση έχουμε $b < a$ και η ακμή δεν επανενώνεται. Θεωρητικά αυτό εξηγείται από την εξίσωση (2.3)

$$TV[u] = \int_{\Omega} |\nabla u| dx dy,$$

Αν θεωρήσουμε ότι τα μαύρα pixels αντιστοιχούν σε τιμή 0 και τα λευκά σε τιμή 1 τότε αν επανενωθεί η μαύρη ακμή έχουμε $TV[u] = 2b$ ενώ αν όχι έχουμε $TV[u] = 2a$.³ Επειδή το μοντέλο επιδιώκει την ελαχιστοποίηση της TV νόρμας, η ακμή επανενώνεται ανάλογα με το πάχος της σε σχέση με την υπό inpainting περιοχή. Το ότι η ακμή στην δεύτερη περίπτωση δεν επανενώνεται έρχεται σε αντίθεση με τη λεγόμενη *αρχή συνεκτικότητας* (Connectivity Principle).

Η αρχή της συνεκτικότητας βασίζεται πάνω σε 2 κύριους παράγοντες.

1. Τα αποτελέσματα της *ψυχολογίας της όρασης*, (vision psychology), σύμφωνα με τα οποία το ανθρώπινο μάτι τείνει να επανενώνει τις ακμές, ανεξάρτητα από το λόγο b/a που είδαμε [31],[40].
2. Στις περισσότερες εικόνες υπάρχουν αντικείμενα σε πολλές κλίμακες. Ακόμα και τα αντικείμενα σε μικρές κλίμακες, λεπτές ακμές κ.λ.π. προσθέτουν πλούσια πληροφορία στην εικόνα και αντιμετωπίζονται ανάλογα από το ανθρώπινο μάτι, σε αντίθεση με το TV Inpainting μοντέλο που κάνει ανάλυση μόνο σε μία κλίμακα

Επίσης από την εικόνα 2.5, εύλογα ανακύπτει το ερώτημα τι γίνεται στην περίπτωση όπου $b = a$. Σε αυτήν την περίπτωση το μοντέλο TV Inpainting δεν μπορεί να δώσει μοναδική απάντηση καθώς είτε η μαύρη μπάρα επανενωθεί είτε όχι, η TV νόρμα παίρνει την ίδια τιμή. Το πρόβλημα που ανακύπτει οφείλεται στο ίδιο το Inpainting πρόβλημα αφού αυτό είναι *μη καλά ορισμένο*, (ill posed problem), και όχι στο TV μοντέλο, αφού όποια και από τις 2 λύσεις διαλέξει ικανοποιείται η ελαχιστοποίηση της νόρμας και το κριτήριο της απόφασης κατά Bayes.

Στην εικόνα 2.6 βλέπουμε ότι το μοντέλο TV inpainting αποτυγχάνει να συμπληρώσει αρμονικά το κομμάτι του κύκλου και απλώς συνδέει τις δύο κορυφές της inpainting περιοχής με τον μικρότερο σε μήκος τρόπο, δηλαδή με ένα ευθύγραμμο τμήμα έτσι ώστε να ελαχιστοποιηθεί η TV νόρμα⁴.

Παρά τις αδυναμίες του το μοντέλο TV inpainting μπορεί να δώσει ικανοποιητικά αποτελέσματα όπως φαίνεται και στην εικόνα 2.7

³Οι τιμές αυτές αντιστοιχούν στο συνολικό μήκος των ακμών στην Inpainting περιοχή.

⁴Μοντέλα inpainting που βασίζονται πάνω στο TV μοντέλο και αντιμετωπίζουν την καμπυλότητα και την αρχή της συνεκτικότητας αναφέρονται παρακάτω 2.3



Σχήμα 2.7: Αφαίρεση κειμένου με TV inpainting

Παρόλο που η ανακατασκευή της εικόνας⁵ περιέχει πολλά λάθη (θόλωμα, μη επανένωση των ακμών, απώλεια υφής και καμπύλων) το περιεχόμενο της εικόνας έχει ανακατασκευασθεί ικανοποιητικά με τη χρήση μόνο του απλού και γρήγορου TV μοντέλου.

2.2.5 Άλλες Εφαρμογές του TV Φίλτρου

Στην ενότητα αυτή θα ασχοληθούμε με άλλες δύο εφαρμογές του TV φίλτρου που όπως θα δούμε συνδέονται στενά με το inpainting.

Ψηφιακό Zoom-In Το ψηφιακό zoom-in ανήκει στη γενικότερη κατηγορία των προβλημάτων της ανακατασκευής εικόνας, με εφαρμογές στην συμπίεση δεδομένων, αύξηση της ανάλυσης ψηφιακής εικόνας. Σκοπός είναι όταν έχουμε μία ψηφιακή εικόνα διαστάσεων m επί n να κατασκευάσουμε μια άλλη εικόνα διαστάσεων $p \cdot m$ επί $p \cdot n$ όπου $p > 1$ (αν $p < 1$ αλλιώς αναφερόμαστε στο ψηφιακό zoom-out). Ένα παρόμοιο πρόβλημα μελετήθηκε από τους Chan, Osher, Shen [61] όπου μελετήθηκε πρώτη φορά το TV με σκοπό την αφαίρεση θορύβου.

⁵Η επεξεργασία έγχρωμων εικόνων σε όλη την εργασία γίνεται στον χώρο RGB και για κάθε χρωματικό κανάλι ξεχωριστά. Για επεξεργασία σε διαφορετικούς χρωματικούς χώρους (CB, HSV) δες π.χ. το [30].

Στην περίπτωση που από μια εικόνα m επί n θέλουμε να κατασκευάσουμε μια εικόνα $2m$ επί $2n$. Η διαδικασία που ακολουθούμε είναι η εξής: Για κάθε pixel της αρχικής εικόνας αντιστοιχεί ένα μπλοκ 2×2 της ζητούμενης. Ένα pixel του μπλοκ παίρνει την τιμή του pixel της αρχικής εικόνας και τα υπόλοιπα 3 μια τυχαία τιμή. Στη συνέχεια έστω Ω το σύνολο των pixel της ζητούμενης εικόνας και Ω_0 το σύνολο των pixel της εικόνας που έχουν πάρει τις τιμές τους από την αρχική εικόνα ($\Omega_0 \subset \Omega$). Για κάθε pixel $\alpha \in \Omega$ θεωρούμε την γειτονιά του N_α . Εδώ η γειτονιά ενός pixel θεωρούνται τα 4 γειτονικά του pixel. Για κάθε pixel α ορίζουμε την τοπική διασπορά του ως

$$|\nabla_\alpha u| = \sqrt{\sum_{\beta \in N_\alpha} (u_\beta - u_\alpha)^2} \quad (2.22)$$

Επίσης, σε αναλογία με το μοντέλο του inpainting ορίζουμε και το πολλαπλασιαστή Lagrange λ_e ως

$$\lambda_e(\alpha) = \begin{cases} \lambda, & \alpha \in \Omega_0 \\ 0, & \alpha \in \Omega \setminus \Omega_0 \end{cases}$$

Το συναρτησιακό που ζητούμε την ελαχιστοποίηση του είναι (σε αναλογία με την εξ. (2.6))

$$J_\lambda[u] = \sum_{\alpha \in \Omega} |\nabla_\alpha u| + \sum_{\alpha \in \Omega} \lambda_e(\alpha)(u_\alpha - u_\alpha^0)^2. \quad (2.23)$$

Για την ελαχιστοποίηση του συναρτησιακού ακολουθούμε ανάλογη πορεία με αυτή για αυτή για το TV inpainting [61].

Στο σχήμα 2.8 βλέπουμε ένα παράδειγμα του μοντέλου TV zoom-in για υπερανάλυση εικόνας. Τα αποτελέσματα της συγκεκριμένης μεθόδου είναι καλύτερα από αυτά που δίνει η απλή μέθοδος *nearest neighbor* και παρόμοια με αυτά της μεθόδου *bicubic interpolation*⁶ όπως φαίνεται από τις εικόνες του σχήματος 2.9.

Συμπύεση εικόνας με κωδικοποίηση ακμών Ο ρόλος των ακμών στην επεξεργασία εικόνας είναι πολύ σημαντικός καθώς εκεί συγκεντρώνεται πολύ πληροφορία σχετικά με τις μεταβολές που παρουσιάζει η εικόνα. Ειδικότερα σε εικόνες οι οποίες δεν έχουν υφή, οι ακμές συγκεντρώνουν το πιο σημαντικό μέρος της πληροφορίας της εικόνας καθώς εκφράζουν σχεδόν όλες τις μεταβολές

⁶Ουσιαστικά, η σύγκριση μεταξύ των δύο μεθόδων πρέπει να γίνει για ακόμη μεγαλύτερες αναλύσεις της εικόνας



Σχήμα 2.8: Παράδειγμα TV zoom-in



Σχήμα 2.9: *a*: Λεπτομέρεια της αρχικής εικόνας, *b*: Λεπτομέρεια του μοντέλου TV zoom-in, *c*: Λεπτομέρεια της μεθόδου nearest neighbor, *d*: Λεπτομέρεια της μεθόδου bicubic interpolation.

που υπάρχουν σε αυτή⁷. Έτσι η αποτελεσματικότητα ενός μοντέλου κωδικοποίησης και συμπίεσης εικόνας κρίνεται σε μεγάλο βαθμό από τη συμπεριφορά του ως προς τις ακμές.

Όπως είδαμε, το μοντέλο TV inpainting είναι ιδιαίτερα αποτελεσματικό. Τα προβλήματα που παρουσιάζει αφορούν την αρχή της συνεκτικότητας καθώς και ότι δεν λαμβάνει υπόψη την καμπυλότητα. Σε περίπτωση όμως που υπάρχει καλή κωδικοποίηση ακμών, τέτοια προβλήματα δεν παρουσιάζονται και το TV inpainting μπορεί να χρησιμοποιηθεί για τη συμπίεση εικόνας.

Το στάδιο της κωδικοποίησης έχει ως εξής:

- *Εύρεση ακμών* Εφαρμόζουμε έναν αρκετα ευαίσθητο ανιχνευτή ακμών π.χ. *Canny edge detector* έτσι ώστε να βρούμε όσο το δυνατόν περισσότερες ακμές της προς κωδικοποίηση εικόνας. Στις ακμές συμπεριλαμβάνουμε και το περίγραμμα της εικόνας.
- *Διαστολή* Στην δυαδική εικόνα ακμών που παίρνουμε εφαρμόζουμε έναν μορφολογικό τελεστή διαστολής (*morphological dilation*). Αυτό γίνεται έτσι ώστε να έχουμε τιμές των pixels εκατέρωθεν των ακμών της εικόνας.
- *Κωδικοποίηση* Οι τιμές που χρειαζόμαστε είναι μόνο οι τιμές στα μη μηδενικά pixels της εικόνας των ακμών μετά τη διαστολή. Αυτές κωδικοποιούνται με αρκετά *bits* έτσι ώστε να έχουμε ακριβή τιμή της εικόνας πάνω και γύρω από τις ακμές.

Για την αποκωδικοποίηση της εικόνας εφαρμόζουμε το μοντέλο του TV inpainting. Στην αρχική πρόβλεψη της εικόνας έχουμε τις κωδικοποιημένες τιμές για τα pixels πάνω και γύρω από τις ακμές ενώ τα υπόλοιπα pixels παίρνουν μια τυχαία τιμή. Στη συνέχεια εφαρμόζουμε το μοντέλο TV inpainting μέχρι αυτό να συγκλίνει στην τελική ανακατασκευασμένη εικόνα.

Το συναρτησιακό που ελαχιστοποιούμε είναι

$$J_\lambda[u] = \sum_{\alpha \in \Omega} |\nabla_\alpha u| + \frac{\lambda_T}{2} \sum_{\alpha \in \Omega} (u_\alpha - u_\alpha^0)^2 \quad (2.24)$$

όπου Ω το σύνολο των pixels της εικόνας T το σύνολο των μη μηδενικών pixels της εικόνας ακμών μετά τη διαστολή ενώ ο πολλαπλασιαστής Lagrange δίνεται από τη σχέση

$$\lambda_T(\alpha) = \begin{cases} \lambda, & \alpha \in T \\ 0, & \alpha \in \Omega \setminus T \end{cases} \quad (2.25)$$

⁷ Από το 1986 ο Marr [43] είχε τονίσει τον πολύ σημαντικό ρόλο των ακμών στην όραση υπολογιστών και είχε ερευνήσει την πιθανότητα ανακατασκευής της εικόνας μόνο από τις ακμές της

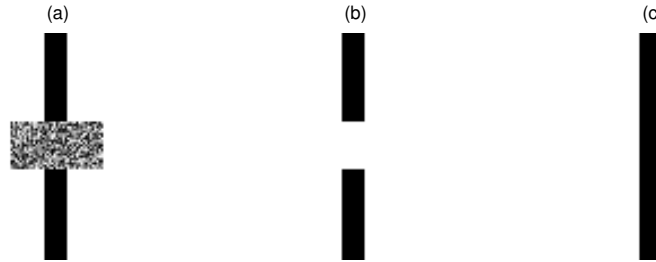


Σχήμα 2.10: Παράδειγμα κωδικοποίησης και συμπίεσης εικόνας. (a): αρχική εικόνα, (b): Εύρεση ακμών με τη μέθοδο Canny, (c): αρχική πρόβλεψη, (d): Ανακατασκευή με TV inpainting.

Στο σχήμα 2.10 φαίνεται η κωδικοποίηση και η ανάκτηση μιας εικόνας σύμφωνα με τη μέθοδο που περιγράψαμε.

Η συγκεκριμένη μέθοδος κωδικοποίησης διαφέρει από άλλα δημοφιλή πρότυπα π.χ. JPEG στο ότι δεν υπάρχει καμία εξάρτηση από τη διεύθυνση των ακμών. Επίσης επειδή η ανακατασκευή γίνεται με πληροφορία που παίρνεται από τις ακμές, και όχι βάσει κάποιου αντιστρόφου μετασχηματισμού, ακόμη και οι απαλές ακμές κατά κάποιο τρόπο ενισχύονται δίνοντας έτσι πιο εκφραστικά αποτελέσματα.

Το μοντέλο κωδικοποίησης που αναφέραμε είναι αρκετά καλό για συμπίεση με απώλειες και συνδιασμό με κάποιο μοντέλο συμπίεσης υψής (π.χ. με χρήση *Markov-Gibbs Random Fields*, *Granulometries*, *Fractals* θα μπορούσαν να



Σχήμα 2.11: Το TV inpainting δεν ικανοποιεί την αρχή της συνεκτικότητας. (a): Αρχική εικόνα, (b): Αποτέλεσμα TV inpainting, (c): Απάντηση από το ανθρώπινο μάτι.

δώσουν ένα αποτελεσματικό τρόπο κωδικοποίησης εικόνας. Αυτό αποτελεί αντικείμενο μελλοντικής έρευνας.

2.3 Επεκτάσεις του Μοντέλου TV Inpainting

Το μεγαλύτερο μειονέκτημα του μοντέλου TV inpainting που μελετήθηκε στην προηγούμενη ενότητα είναι ότι δεν επανενώνει τις ακμές ενός αντικειμένου όταν αυτές βρίσκονται σχετικά μακριά (αρχή της συνεκτικότητας). Το πρόβλημα αυτό γίνεται πιο οξύ σε εικόνες πολλαπλών κλιμάκων, αφού στις μικρές κλίμακες οι λεπτές γραμμές για παράδειγμα πολύ δύσκολα θα επανενωθούν, όπως φαίνεται από την εικόνα (2.5). Επίσης δεν λαμβάνεται υπόψη η καμπυλότητα και όλες οι σπασμένες ακμές τείνουν να επανενωθούν με γραμμικές προσεγγίσεις.

Στην ενότητα αυτή αναφερόμαστε σε δύο μοντέλα τα οποία αποτελούν επεκτάσεις του TV inpainting μοντέλου και αντιμετωπίζουν τα συγκεκριμένα προβλήματα.

2.3.1 Inpainting με Διαχύσεις Οδηγούμενες από την Καμπυλότητα

Το 2000 οι Chan και Shen [15] βασιζόμενοι στο αρχικό τους TV inpainting μοντέλο προχώρησαν σε μια επέκταση του με σκοπό την ικανοποίηση της αρχής της συνεκτικότητας.

Η έξοδος του TV inpainting μοντέλου (Εικόνα 2.11) δίνει τέσσερις γωνίες στις οποίες η καμπυλότητα έχει τιμή $\kappa = \pm\infty$. Αντίθετα η απάντηση που γενικά δίνει το ανθρώπινο μάτι, επανενώνει τις ακμές και ελαχιστοποιεί την καμπυλότητα. Άρα στο συναρτησιακό που οδηγεί στο αντίστοιχο PDE μοντέλο πρέπει να συμπεριληφθεί και ένας όρος καμπυλότητας.

Το καινούριο Inpainting μοντέλο εμπνεύστηκε από την εξίσωση της ανισοτροπικής διάχυσης.

Η γενική μορφή της εξίσωσης της μη γραμμικής ανισοτροπικής διάχυσης (Nonlinear anisotropic diffusion) είναι (Perona & Malik [50])

$$\frac{\partial u}{\partial t} = \nabla(D\nabla u) \quad (2.26)$$

όπου D ο συντελεστής διάχυσης (diffusion coefficient). Η εξ.(2.26) είναι ίδια με το PDE μοντέλο της εξ.(2.8) μέσα στην περιοχή inpainting με συντελεστή διάχυσης

$$D = \frac{1}{|\nabla u|}. \quad (2.27)$$

Οι Chan και Shen θεωρώντας το TV inpainting μοντέλο ως μία ειδική μορφή διάχυσης, το εμπλούτισαν αλλάζοντας τον συντελεστή διάχυσης D έτσι ώστε να περιέχει κάποιο όρο καμπυλότητας

$$D = \frac{|\nabla \kappa|^p}{|\nabla u|} \quad (2.28)$$

όπου $p \geq 1$ και κ η καμπυλότητα η οποία δίνεται από τον τύπο

$$\kappa = \nabla \left[\frac{\nabla u}{|\nabla u|} \right]$$

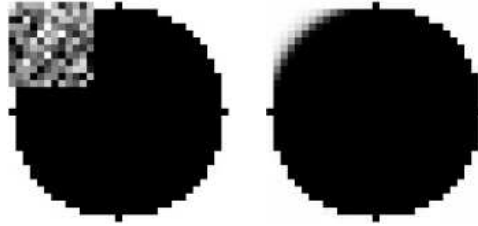
Διατηρώντας την αρχική τοπολογία του σχήματος 2.1 το PDE μοντέλο που προκύπτει είναι

$$\frac{\partial u}{\partial t} = \nabla \left[\frac{G(x, |\kappa|)}{|\nabla u|} \nabla u \right] + \lambda_\epsilon(x)(u - u_0) \quad (2.29)$$

όπου ο συντελεστής διάχυσης και ο πολλαπλασιαστής Lagrange είναι

$$G(x, s) = \begin{cases} 1, & x \in E \\ s^p, & x \in D \end{cases} \quad \lambda_\epsilon(x) = \begin{cases} \lambda, & x \in E \\ 0, & x \in D \end{cases} \quad (2.30)$$

Όπως είπαμε, το παραπάνω μοντέλο επιχειρεί να κάνει inpainting ικανοποιώντας την αρχή της συνεκτικότητας. Για να το πετύχει αυτό συνεχίζει τις κομμένες ακμές με στόχο την ελαχιστοποίηση της καμπυλότητας. Επειδή ο τρόπος που εξελίσσονται οι ακμές αποτελεί στην ουσία εξίσωση ανισοτροπικής διάχυσης το συγκεκριμένο μοντέλο ονομάστηκε *inpainting* με διαχύσεις οδηγούμενες από την καμπυλότητα (Curvature Driven Diffusions - CDD).



Σχήμα 2.12: Το CDD inpainting επανενώνει αρμονικά ακμές με καμπυλότητα

Η αριθμητική υλοποίηση του μοντέλου οδηγεί σε ένα επαναληπτικό σχήμα της μορφής

$$u^{(n+1)} = u^{(n)} - \Delta t \nabla \cdot j^{(n)} \quad (2.31)$$

Το χρονικό βήμα Δt στην εξίσωση (2.31) επηρεάζει την ταχύτητα σύγκλισης αλλά και την ευστάθεια του επαναληπτικού σχήματος. Στη δική μας αριθμητική υλοποίηση το χρονικό βήμα έπρεπε να επιλεγεί πολύ μικρό έτσι ώστε να υπάρχει ευστάθεια⁸, πράγμα που έκανε τον αλγόριθμο πολύ αργό. Μία λύση ήταν να εφαρμόσουμε πρώτα το γρήγορο και ευσταθές μοντέλο του TV inpainting και στη συνέχεια το CDD μοντέλο.

Παρόλες τις αδυναμίες του το CDD inpainting μοντέλο δίνει καλύτερα αποτελέσματα από το TV inpainting μοντέλο, όπως φαίνεται στην εικόνα 2.12 καθώς εισάγει και τον όρο της καμπυλότητας.

2.3.2 Το Μοντέλο Euler's Elastica Inpainting

Μία καμπύλη γ λέγεται ότι είναι *Euler's elastica* αν είναι η καμπύλη ισοροπίας (equilibrium curve) της ενέργειας ελαστικότητας:

$$E_2[\gamma] = \int_{\gamma} (a + b\kappa^2) ds, \quad (2.32)$$

όπου ds είναι το στοιχειώδες μήκος της καμπύλης, κ είναι η καμπυλότητα και a, b θετικές σταθερές. Η ονομασία προέρχεται από τον Euler που υπολόγισε την ενέργεια μια λεπτής ράβδου, χωρίς περιορισμούς ως προς τη στρέψη, κάτω από εξωτερικές δυνάμεις. Στην όραση υπολογιστών χρησιμοποιήθηκε πρώτα από τον Mumford [46] ως ένα μοντέλο καμπύλης. Οι Masnou και Morel [44] πρότειναν ένα μοντέλο inpainting με χρήση της Euler's elastica ενώ οι Chan,

⁸Γενικά το inpainting πρόβλημα είναι ένα όχι καλά ορισμένο πρόβλημα (ill-posed problem). Η εύρεση γρήγορων και σωστών μαθηματικών μοντέλων για inpaintings είναι μία από τις σπουδαιότερες προκλήσεις του προβλήματος.

Kang, Shen [60] μελέτησαν τις ιδιότητες του μοντέλου αυτού καθώς και το μαθηματικό υπόβαθρο.

Εισάγοντας τον όρο της Euler's Elastica μπορούμε να λάβουμε ένα νέο μοντέλο ενέργειας για την εικόνα, σύμφωνα με τη Bayesian θεώρηση για το Inpainting πρόβλημα (κεφάλαιο 1)

$$E[u] = \int_{\Omega} (a + b\kappa^2) |\nabla u| dx dy \quad (2.33)$$

Το συναρτησιακό ενέργειας που επιδιώκεται η ελαχιστοποίηση του, (σύμφωνα με το κριτήριο βέλτιστης απόφασης), με το συγκεκριμένο μοντέλο είναι

$$J_2^\lambda[u] = \int_{E \cup D} (a + b\kappa^2) |\nabla u| dx + \frac{\lambda}{2} \int_E (u - u_0)^2 dx dy \quad (2.34)$$

Για την εύρεση της εξίσωσης Euler - Lagrange χρησιμοποιήθηκε το παρακάτω θεώρημα (για απόδειξη δες [60])

Θεώρημα 2.3. Αν ϕ μια δεδομένη συνάρτηση με $\phi \in C^1(\mathbb{R}, [0, \infty])$ και

$$R[u] = \int_{E \cup D} \phi(\kappa) |\nabla u| dx$$

τότε ισχύει ότι

$$\frac{\partial R}{\partial u} = -\nabla \cdot \vec{V},$$

όπου το πεδίο \vec{V} δίνεται από τον τύπο

$$\vec{V} = \phi(\kappa) \vec{n} - \frac{\vec{t}}{|\nabla u|} \frac{\partial(\phi'(\kappa)) |\nabla u|}{\partial \vec{t}}. \quad (2.35)$$

και $\vec{n} \parallel \frac{\nabla u}{|\nabla u|}$ και \vec{t} το εφαπτόμενο διάνυσμα (σχήμα 2.13)

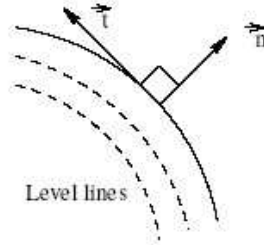
Από το παραπάνω θεώρημα προκύπτει η παρακάτω επαναληπτική μέθοδος για το elastica inpainting.

$$\frac{\partial u}{\partial t} = \nabla \cdot \vec{V} - \lambda_E (u - u_0), \quad (2.36)$$

με

$$\vec{V} = (a + b\kappa^2) \vec{n} - \frac{2b}{|\nabla u|} \frac{\partial \kappa |\nabla u|}{\partial \vec{t}} \vec{t},$$

$$\lambda_E(x) = \lambda \cdot 1_E(x) \quad \text{δείκτης του } E.$$



Σχήμα 2.13: Τα διανύσματα \vec{n} και \vec{t}

Οι Marquina και Osher [42] πρότειναν ότι η σύγκλιση του παρακάτω ελαφρώς τροποποιημένου επαναληπτικού μοντέλου είναι πιο γρήγορη.

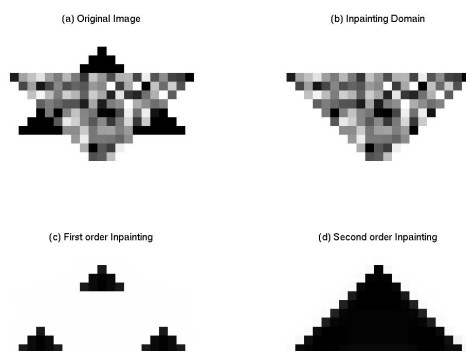
$$\frac{\partial u}{\partial t} = |\nabla u| \nabla \cdot \vec{V} - |\nabla u| \lambda_E (u - u_0), \quad (2.37)$$

Μια αναλυτική αριθμητική υλοποίηση του elastica inpainting (εξ. (2.37)) δίνεται στο [60]. Όπως και στο CDD inpainting (ενοτ. 2.3.1) έτσι και εδώ το χρονικό βήμα της επαναληπτικής μεθόδου πρέπει να επιλεγθεί ώστε είναι η μέθοδος να είναι ευσταθής. Στην αριθμητική υλοποίηση που επιχειρήσαμε το χρονικό βήμα ήταν πολύ μικρό, για λόγους ευστάθειας, πράγμα όμως που έκανε τη σύγκλιση πολύ αργή. Επίσης οι τιμές των παραμέτρων a και b παίζουν ρόλο στα αποτελέσματα του μοντέλου. Εκείνο που έχει σημασία είναι ο λόγος τους b/a αφού όταν αυτός είναι μηδέν έχουμε το γνωστό TV inpainting μοντέλο ενώ όσο μεγαλώνει τόσο λαμβάνεται περισσότερο υπόψη η καμπυλότητα. Μια λογική τιμή του λόγου είναι $b/a = 20$.

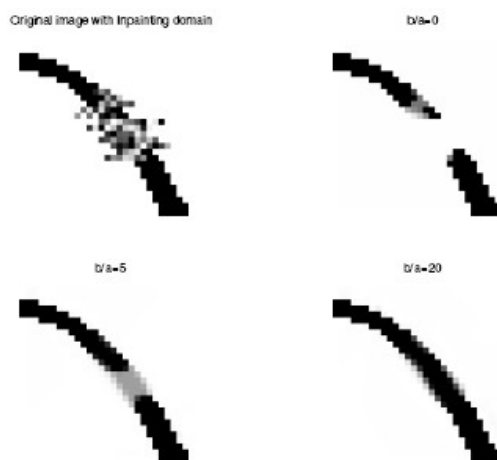
Στις εικόνες 2.14 και 2.15 δείχνουμε κάποια ενδεικτικά παραδείγματα από το [60] που δείχνουν ότι το elastic inpainting μοντέλο ικανοποιεί την αρχή της συνεκτικότητας και αντιμετωπίζει καλά ακμές με καμπυλότητα.

Τα μοντέλα CDD Inpainting και Euler's Elastica Inpainting βελτιώνουν το μοντέλο TV Inpainting και φαίνεται ότι λύνουν κάποια από τα προβλήματά του (αρχή της συνεκτικότητας, καμπυλότητα). Όμως η ύπαρξη λύσης και η ευστάθεια δεν είναι εξασφαλισμένες ενώ εύκολα μπορεί να δειχθεί η μη μοναδικότητα της λύσης.

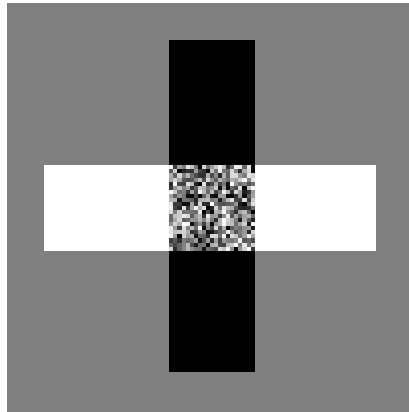
Στην εικόνα 2.16 η Inpainting περιοχή βρίσκεται στην τομή της μαύρης και της λευκής μπάρας, ενώ οι δύο μπάρες έχουν ακριβώς το ίδιο πάχος. Το Inpainting μοντέλο δεν μπορεί να αποφασίσει πια από τις δύο μπάρες να επανενώσει καθώς λόγω της συμμετρικότητας που υπάρχει και οι δύο επιλογές αποτελούν βέλτιστη λύση του προβλήματος.



Σχήμα 2.14: Το μοντέλο elastic inpainting ικανοποιεί την αρχή της συνεκτικότητας. Πάνω: Αρχική εικόνα και περιοχή inpainting. Κάτω: Αριστερά, TV inpainting και δεξιά elastic inpainting.



Σχήμα 2.15: Το μοντέλο elastic inpainting λαμβάνει περισσότερο υπόψη του την καμπυλότητα όσο μεγαλώνει ο όρος b/a : Πάνω: Αριστερά αρχική εικόνα, δεξιά ανακατασκευή με $b/a = 0$ (TV inpainting). Κάτω: Αριστερά, ανακατασκευή με $b/a = 5$, δεξιά, ανακατασκευή με $b/a = 20$. Παρατηρούμε ότι όσο μεγαλώνει ο λόγος $b/a = 0$ τόσο περισσότερο λαμβάνεται υπόψη η καμπυλότητα.



Σχήμα 2.16: Τα μοντέλα CDD και Euler's Elastica Inpainting δεν έχουν μοναδική λύση

Από αυτό το παράδειγμα φαίνεται πλέον ξεκάθαρα ότι το Inpainting πρόβλημα είναι ένα μη καλά ορισμένο πρόβλημα. Με βάση το κριτήριο της Bayesian απόφασης και οι δύο λύσεις είναι σωστές, ενώ πλέον δεν υπάρχουν στοιχεία και από την επιστήμη της ψυχολογίας της όρασης αφού και το ανθρώπινο μάτι σε μία τέτοια περίπτωση δεν μπορεί να αποφασίσει με βεβαιότητα. Επομένως η αδυναμία των μοντέλων οφείλεται και πάλι στο ίδιο το Inpainting πρόβλημα.

2.4 Το Μοντέλο BCBS

Στις ενότητες 2.2 και 2.3 αναφερθήκαμε στο μοντέλο TV inpainting καθώς και στις επεκτάσεις του. Τα μοντέλα αυτά προκύπτουν από τα αντίστοιχα μοντέλα ενέργειας της εικόνας που επιλέγουμε, έτσι ώστε να τονίσουμε ιδιαίτερα τα σημαντικά σημεία του Inpainting προβλήματος (ακμές, καμπυλότητα) αλλά και να ικανοποιηθούν συμπεράσματα της ψυχολογίας της όρασης (αρχή της συνεκτικότητας). Η λύση του προβλήματος αντιστοιχεί σε βέλτιστη απόφαση σύμφωνα με τον κανόνα του Bayes.

Πέρα όμως από αυτήν την αντιμετώπιση του Inpainting προβλήματος, που σχετίζεται άμεσα με το θεμελιώδες πρόβλημα της επεξεργασίας εικόνας, υπάρχει και η προσέγγιση σύμφωνα με τους κανόνες και τις τεχνικές των συντηρητών τέχνης.

Η έννοια επιδιόρθωσης εικόνας υπήρχε από την εποχή Αναγέννησης, με τη μορφή της συντήρησης μεσαιωνικών πινάκων ζωγραφικής. Η ιδέα αυτή επεκτάθηκε και στην επιδιόρθωση φωτογραφίας και video καθώς και στην απομάκρυνση αντικειμένων από εικόνες και video.

Το 2000 οι Bertalmio, Sapiro, Caselles, Ballester⁹ [38] εισήγαγαν τον όρο *inpainting* και προσπάθησαν να μιμηθούν τις μεθόδους που εφαρμόζονται από επαγγελματίες ζωγράφους που ασχολούνται με τη συντήρηση έργων τέχνης, όσον αφορά το περιεχόμενο της εικόνας.

2.4.1 Εξαγωγή του Μοντέλου - Ο Αλγόριθμος

Έστω D η περιοχή που πρέπει να επιδιορθωθεί και Γ το όριό της. Η τεχνική που πρότειναν οι Bertalmio, Sapiro, Caselles, Ballester συνεχίζει τις γραμμές με την ίδια φωτεινότητα, (ισόφωτες), (*isophote lines*) που φτάνουν στο όριο Γ ενώ παράλληλα διατηρεί τη γωνία άφιξής τους. Επίσης λαμβάνεται υπόψη και η καμπυλότητα των γραμμών έτσι ώστε αυτές καθώς προχωρούν στην περιοχή D να μην τέμνονται μεταξύ τους.

Για την υλοποίηση της παραπάνω διαδικασίας χρησιμοποιήθηκε ένα επαναληπτικό PDE μοντέλο. Η γενική μορφή ενός τέτοιου μοντέλου έχει τη μορφή

$$u^{(n+1)}(i, j) = u^{(n)} + \Delta t u_t^{(n)}(i, j), \quad \forall (i, j) \in \Omega \quad (2.38)$$

όπου $u^{(n)}$ η εικόνα κατά τη n -οστή επανάληψη και $u_t^{(n)}$ είναι η μεταβολή της εικόνας $u^{(n)}$ στη n -οστή επανάληψη του PDE μοντέλου.

Αν κατά τη n -οστή επανάληψη είναι $L^n(i, j)$ η πληροφορία που θέλουμε να μεταφέρουμε και $\vec{N}^{(n)}(i, j)$ η κατεύθυνση της εξέλιξης της πληροφορίας τότε μπορούμε να εκφράσουμε τη μεταβολή $u_t^{(n)}$ ως

$$u_t^{(n)} = \overrightarrow{\delta L^{(n)}}(i, j) \cdot \vec{N}^{(n)}(i, j),^{10} \quad (2.39)$$

Όταν το επαναληπτικό σχήμα συγκλίνει τότε θα έχουμε ότι $\overrightarrow{\delta L^{(n)}}(i, j) \cdot \vec{N}^{(n)}(i, j) = 0$ που σημαίνει ότι η πληροφορία L , των ισόφωτων καμπύλων θα έχει διαδοθεί κατά τη διεύθυνση \vec{N} , το διάνυσμα μεταβολής είναι κάθετο στο διάνυσμα των ισόφωτων καμπύλων.

Μία απαίτηση που έχουμε, είναι η πληροφορία να διαδίδεται με λείο τρόπο. Έτσι επιλέγουμε την πληροφορία να είναι ίση με μία διακριτή προσέγγιση του λαπλασιανού τελεστή: $L^{(n)} = \Delta u = u_{xx}^{(n)} + u_{yy}^{(n)}$.

⁹ Από τα αρχικά των οποίων προέρχεται και το όνομα του συγκεκριμένου μοντέλου

¹⁰ Χρησιμοποιώντας το συμβολισμό από το συνεχές πεδίο μπορούμε να γράψουμε το $\overrightarrow{\delta L^{(n)}}$ ως ∇L

Για την τον προσδιορισμό της κατεύθυνσης διάδοσης πληροφορίας $\overrightarrow{N^{(n)}}$ πρέπει να επιλέξουμε την κατεύθυνση των γραμμών με την ίδια φωτεινότητα.¹¹ Για κάθε pixel (i, j) της εικόνας η κλίση $\nabla u^{(n)}(i, j)$ εκφράζει την κατεύθυνση όπου υπάρχει η μεγαλύτερη μεταβολή. Αφού θέλουμε μεταφορά με κατεύθυνση κατά μήκος των γραμμών με ίδια φωτεινότητα επιλέγουμε την κάθετη σε αυτή κατεύθυνση $\nabla^\perp u^{(n)}(i, j) := (-\partial_y, \partial_x)$ που εκφράζει την κατεύθυνση στην οποία έχουμε τη μικρότερη μεταβολή.

Το PDE μοντέλο που προκύπτει με τις παραπάνω θεωρήσεις είναι

$$\frac{\partial u}{\partial t} = \nabla^\perp u \cdot \nabla (\Delta u) \quad (2.40)$$

όπου u η φωτεινότητα της εικόνας, $\nabla = (\partial_x, \partial_y)$, $\nabla^\perp = (-\partial_y, \partial_x)$ η κάθετη κλίση και Δ ο τελεστής Laplace $\partial_x^2 + \partial_y^2$.

Για να είναι η εξέλιξη των γραμμών με την ίδια φωτεινότητα ομαλή καθώς προχωράμε στην inpainting περιοχή D , μετά από κάποιο μικρό αριθμό επαναλήψεων του PDE μοντέλου εκτέλουμε και κάποιες επαναλήψεις κάποιου αλγόριθμου ανισοτροπικής διάχυσης, έτσι ώστε να τονίσουμε τις διαφορές μεταξύ των γραμμών με την ίδια φωτεινότητα και να εμποδίσουμε την τομή μεταξύ τους σε κάποιο σημείο. Συγκεκριμένα εφαρμόζουμε την εξίσωση των Perona Malik [50]

$$\frac{\partial u}{\partial t}(x, y, t) = g(x, y) \kappa(x, y, t) |\nabla u(x, y, t)|, \quad \forall (x, y) \in E \quad (2.41)$$

όπου E είναι η διαστολή του D (ως συνόλου σημείων) με έναν μικρό δίσκο, ώστε να πάρουμε υπόψη και τα σημεία (περιοχή E) λίγο έξω από την περιοχή D , κ είναι η καμπυλότητα και $g(x, y)$ είναι μια λεία συνάρτηση με $g(x, y) = 0 \quad \forall (x, y) \in \partial E$ και $g(x, y) = 1 \quad \forall (x, y) \in D$.¹²

2.4.2 Αριθμητική Υλοποίηση

Για την διακριτοποίηση και επίλυση του παραπάνω σχήματος θεωρούμε τη γενική εξίσωση ενός PDE μοντέλου:

$$u^{(n+1)}(i, j) = u^{(n)} + \Delta t u_t^{(n)}(i, j), \quad \forall (i, j) \in E \quad (2.42)$$

¹¹Μια άλλη επιλογή θα ήταν να επιλέξουμε σε κάθε σημείο το κάθετο διάνυσμα στο όριο της inpainting περιοχής. Η επιλογή αυτή δεν δίνει καλά αποτελέσματα [38].

¹²Με τη συνάρτηση g εισάγουμε και μια οριακή συνθήκη τύπου Dirichlet στην εξ. (2.41)

και χρησιμοποιούμε τις παρακάτω σχέσεις

$$u_t^{(n)}(i, j) = \left(\overrightarrow{\delta L^{(n)}}(i, j) \cdot \frac{\overrightarrow{N}(i, j, n)}{|\overrightarrow{N}(i, j, n)|} \right) |\nabla u^{(n)}(i, j)|, \quad (2.43)$$

$$\overrightarrow{\delta L^{(n)}}(i, j) = (L^{(n)}(i+1, j) - L^{(n)}(i-1, j), L^{(n)}(i, j+1) - L^{(n)}(i, j-1)) \quad (2.44)$$

$$L^{(n)}(i, j) = u_{xx}^{(n)}(i, j) + u_{yy}^{(n)}(i, j) \quad (2.45)$$

$$\frac{\overrightarrow{N}(i, j, n)}{|\overrightarrow{N}(i, j, n)|} := \frac{(-u_y^{(n)}(i, j), u_x^{(n)}(i, j))}{\sqrt{(u_x^{(n)}(i, j))^2 + (u_y^{(n)}(i, j))^2}} \quad (2.46)$$

$$\beta^{(n)}(i, j) = \overrightarrow{\delta L^{(n)}}(i, j) \cdot \frac{\overrightarrow{N}(i, j, n)}{|\overrightarrow{N}(i, j, n)|} \quad (2.47)$$

ενώ

$$|\nabla u^{(n)}| = \begin{cases} \sqrt{(u_{x_{bm}}^{(n)})^2 + (u_{x_{fM}}^{(n)})^2 + (u_{y_{bm}}^{(n)})^2 + (u_{y_{fM}}^{(n)})^2}, & \text{για } \beta^{(n)} > 0 \\ \sqrt{(u_{x_{bM}}^{(n)})^2 + (u_{x_{fm}}^{(n)})^2 + (u_{y_{bM}}^{(n)})^2 + (u_{y_{fm}}^{(n)})^2}, & \text{για } \beta^{(n)} < 0 \end{cases} \quad (2.48)$$

για κάθε pixel της περιοχής Ω όπου οι δείκτες f, b δηλώνουν εμπρόσθιο και οπίσθιο σχήμα αφαίρεσης (forward, backward) και οι δείκτες m, M το μέγιστο ή ελάχιστο αντίστοιχα μεταξύ της διαφοράς αυτής και του μηδέν π.χ.

$$u_{x_{fM}}^{(n)}(i, j) = \max [(u^{(n)}(i+1, j) - u^{(n)}(i, j)), 0].$$

Οι εξισώσεις (2.42) έως (2.48) εφαρμόζονται σε όλα τα pixels της inpainting περιοχής E . Αυτό γίνεται γιατί οι τιμές των παραγώγων στα σημεία κοντά στο όριο Γ χρειάζονται και τιμές των σημείων εκτός της D . Οι νέες όμως τιμές που προκύπτουν, εφαρμόζονται μόνο στα σημεία της περιοχής D αφού μόνο εκεί υπάρχει το πρόβλημα που πρέπει να επιδιορθωθεί.

Η σειρά των σημείων στα οποία εφαρμόζονται οι εξισώσεις ποροκύπτει ως εξής: Αρχίσουμε απο τα σημεία που βρίσκονται στο όριο της περιοχής. Μετά σε αυτό το σύνολο εφαρμόζουμε έναν τελεστή μορφολογικής διαστολής με ένα μοναδιαίο δίσκο και τα σημεία που προκύπτουν από τη διαφορά των δύο συνόλων είναι τα επόμενα σημεία που θα εφαρμοστούν οι εξισώσεις. Συνεχίζουμε την διαδικασία μέχρι να εκτελεστούν οι εξισώσεις σε όλα τα σημεία της περιοχής¹³. Με αυτό τον τρόπο πετυχαίνουμε να εφαρμόζονται οι εξισώσεις από έξω προς τα μέσα.

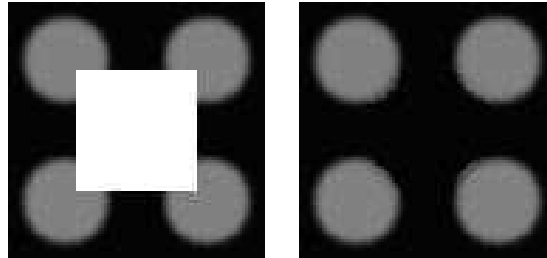
¹³Η μέθοδος αυτή είναι γνωστή και σαν onion peel method

2.4.3 Αποτελέσματα

Παρουσιάζουμε αποτελέσματα inpainting που ελήφθησαν με την παραπάνω μέθοδο BCBS. Αρχικά αναφέρουμε κάποιες γενικές παρατηρήσεις για τα αποτελέσματα της μεθόδου.

- Το BCBS inpainting μοντέλο δίνει καλά αποτελέσματα σε απλές περιπτώσεις χωρίς υφή και πολλές απότομες μεταβολές ενώ λαμβάνει υπόψη την καμπυλότητα (εικόνα 2.17).
- Το μοντέλο BCBS, όπως και τα μοντέλα CDD inpainting, elastica inpainting υλοποιούνται από PDE's τα οποία είναι 3ου βαθμού και τα οποία είναι πολύ ευαίσθητα ως προς την ευστάθεια. Η χρονική σταθερά Δt πρέπει να επιλεγεί ώστε να υπάρχει ευστάθεια. Η τιμή που επιλέχθηκε ήταν $\Delta t = 0.003$.
- Η κατάλληλη εναλλαγή μεταξύ επαναλήψεων του PDE σχήματος και της εξίσωσης ανισοτροπικής διάχυσης, ώστε η μέθοδος να δώσει καλά αποτελέσματα, δεν είναι σταθερή αλλά ανάλογα με την εικόνα αλλάζει ο αριθμός των επαναλήψεων, ενώ ακόμα και στην ίδια εικόνα, διαφορετικές περιοχές χρειάζονται αριθμό επαναλήψεων (εικόνα 2.20). Μια επιλογή είναι μετά από κάθε 100 επαναλήψεις του PDE σχήματος να ακολουθούν 15 επαναλήψεις της εξίσωσης ανισοτροπικής διάχυσης¹⁴.
- Η μέθοδος BCBS, όπως και όλες που έχουμε αναφέρει μέχρι τώρα, δεν αναπαράγει την υφή της εικόνας (εικόνες 2.18, 2.20). Μέθοδοι για inpainting υφής αναφέρονται στο επόμενο κεφάλαιο.
- Η ταχύτητα της μεθόδου εξαρτάται έντονα από το μέγεθος της inpainting περιοχής. Γενικά η μέθοδος είναι αργή ενώ για μεγάλες inpainting περιοχές δεν δίνει καλά αποτελέσματα. Πάντως για την αντιμετώπιση μεγάλων περιοχών μπορεί είτε πρώτα να εφαρμοστεί κάποιος άλλος πιο γρήγορος αλγόριθμος TV inpainting και στη συνέχεια το μοντέλο BCBS (εικόνα 2.19) ή να γίνει ανάλυση σε πολλαπλές κλίμακες μέσω λαπλασιανής πυραμίδας.

¹⁴Στο [38] προτείνεται μετά από 15 επαναλήψεις του PDE σχήματος να ακολουθούν 2 επαναλήψεις της εξίσωσης ανισοτροπικής διάχυσης



Σχήμα 2.17: Το μοντέλο BCBS διατηρεί την καμπυλότητα των ακμών



Αρχική Εικόνα



Ανακατασκευασμένη Εικόνα

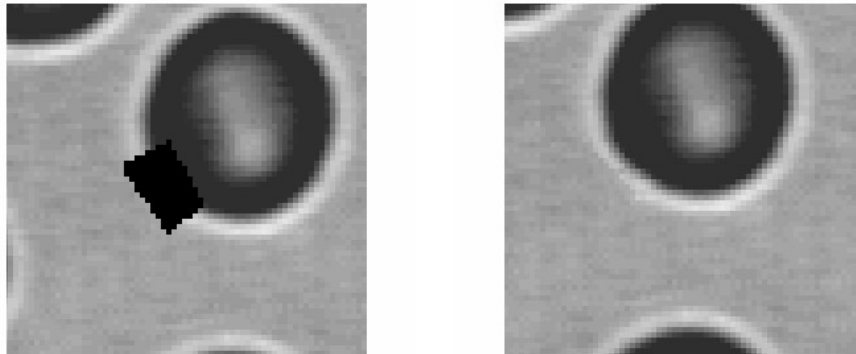


Λεπτομέρεια Αρχικής Εικόνας



Λεπτομέρεια Ανακατασκευασμένης

Σχήμα 2.18: Το μοντέλο BCBS δεν λειτουργεί καλά σε περιοχές με υφή και μεγάλες περιοχές



Σχήμα 2.19: TV και BCBS inpainting



Σχήμα 2.20: Απομάκρυνση κειμένου με BCBS inpainting. Κάποιες περιοχές (π.χ. φωτιστικό) ανακτώνται καλά ενώ κάποιες άλλες (π.χ. γλάστρα) δεν ανακτώνται καλά. Επίσης γίνεται εμφανές ότι δεν υπάρχει inpainting υφής.

2.5 Σύνδεση των Μοντέλων BCBS, CDD, Euler's Elastica Inpainting

Το μοντέλο BCBS inpainting στηρίζεται όπως είδαμε στην λεία μεταφορά της πληροφορία κατά μήκος των ισόφωτων καμπύλων της εικόνας και υλοποιείται από το PDE

$$\frac{\partial u}{\partial t} = \nabla^\perp u \cdot \nabla L(u) \quad (2.49)$$

όπου $L(u)$ εκφράζει το πόσο λεία είναι η εικόνα και στο συγκεκριμένο μοντέλο έχει επιλεγθεί η λαπλασιανή Δu . Διατηρώντας το συμβολισμό της εικόνας 2.13 το μοντέλο φτάνει στην τελική του κατάσταση όταν

$$\vec{t} \cdot \nabla L(u) = 0 \quad \text{και} \quad \frac{\partial L(u)}{\partial \vec{t}} = 0 \quad (2.50)$$

δηλαδή όταν η εικόνα γίνει λεία κατα μήκος των ισόφωτων καμπύλων.

Από την άλλη το μοντέλο CDD inpainting βασίζεται στην εξίσωση ανισοτροπικής διάχυσης

$$\frac{\partial u}{\partial t} = \nabla \cdot \left(\frac{g(\kappa)}{|\nabla u|} \nabla u \right), \quad (2.51)$$

όπου $g : \mathbb{R} \rightarrow [0, +\infty)$ με $g(0) = 0$, $g(\pm\infty) = \pm\infty$, η συνάρτηση που χρησιμοποιείται για να εμποδίσει τις μεγάλες καμπυλότητες και να ενθαρρύνει αποτελέσματα με μικρή καμπυλότητα ώστε να ικανοποιείται η αρχή της συνεκτικότητας. Τα δύο αυτά μοντέλα φαίνονται ορθογώνια μεταξύ τους αφού το πρώτο μεταφέρει την πληροφορία κατά μήκος των ισόφωτων ενώ το δεύτερο διαχέει την πληροφορία εγκάρσια ως προς τις ακμές.

Τα μοντέλα αυτά συνδέονται μέσω του elastica inpainting. Το αντίστοιχο PDE για το συγκεκριμένο μοντέλο είναι

$$\frac{\partial u}{\partial t} = \nabla \cdot \vec{V} \quad (2.52)$$

με

$$\vec{V} = \phi(\kappa) \vec{n} - \frac{1}{|\nabla u|} \frac{\partial(\phi'(\kappa)|\nabla u|)}{\partial \vec{t}} \vec{t}.$$

Από τις παραπάνω εξισώσεις βλέπουμε ότι η συνιστώσα $V_{\vec{n}} = \phi(\kappa)$ αντιστοιχεί στο μοντέλο CDD inpainting με $g(\kappa) = \phi(\kappa)$. Από την άλλη πλευρά η εφαπτόμενη συνιστώσα $V_{\vec{t}}$ μπορεί να γραφεί και ως

$$V_{\vec{t}} = -\frac{1}{|\nabla u|^2} \frac{\partial(\phi'(\kappa)|\nabla u|)}{\partial \vec{t}} \nabla^\perp u.$$

από όπου προκύπτει ότι η κλίση της είναι

$$\nabla \cdot V_{\vec{t}} = \nabla^\perp u \cdot \nabla \left(-\frac{1}{|\nabla u|^2} \frac{\partial(\phi'(\kappa)|\nabla u|)}{\partial \vec{t}} \right). \quad (2.53)$$

Από τις εξισώσεις (2.49), (2.53) παρατηρούμε ότι όρος $V_{\vec{t}}$ αντιστοιχεί στο μοντέλο BCBS με μέτρο λειότητας

$$L(u) = -\frac{1}{|\nabla u|^2} \frac{\partial(\phi'(\kappa)|\nabla u|)}{\partial \vec{t}} \quad (2.54)$$

Από τα παραπάνω βλέπουμε ότι το μοντέλο Euler's Elastica Inpainting συνδιάζει το BCBS και το CDD Inpainting μοντέλο. Η σύνδεση αυτή, παρόλο που από πρακτικής άποψης δεν είναι πολύ σημαντική καθώς και τα 3 μοντέλα δίνουν παραπλήσια μεταξύ τους αποτελέσματα, είναι πολύ σημαντική από τη θεωρητική πλευρά.

Από τη σκοπιά του θεμελιώδους προβλήματος της επεξεργασίας εικόνας, τόσο το BCBS μοντέλο, που μιμείται τις τεχνικές συντηρητών έργων τέχνης, όσο και το CDD μοντέλο, που προκύπτει από μια γενικευμένη διαδικασία ανισοτροπικής διάχυσης, μπορούν να υπαχθούν στην κατηγορία της εικόνας ως προσομοίωση κάποιας φυσικής διαδικασίας.

Το γεγονός ότι η μέθοδος του Elastica Inpainting, που προκύπτει από το κατάλληλο μοντέλο ενέργειας της εικόνας, αποτελεί συνδιασμό των δύο μεθόδων, δίνει τις μεθόδους μεταξύ τους και ενισχύει την ορθότητά τους για ένα πρόβλημα όχι καλά ορισμένο αλλά και υποκειμενικό, όπως είναι το Inpainting πρόβλημα.

2.6 Άλλα Μοντέλα για inpainting

Παρακάτω αναφέρουμε συνοπτικά άλλες μεθόδους για inpainting. Δίνουμε περιγραφικά τις ιδέες για το κάθε μοντέλο και τις βασικές εξισώσεις. Για περισσότερες πληροφορίες, ο αναγνώστης μπορεί να δει τις αντίστοιχες αναφορές.

2.6.1 Navier-Stokes, Μηχανική των Ρευστών και inpainting Εικόνας

Το μοντέλο BCBS, στο οποίο αναφερθήκαμε πριν, προσπαθεί να μιμηθεί τις τεχνικές που χρησιμοποιούνται από συντηρητές έργων τέχνης. Τα αποτελέσματα που δίνει είναι αρκετά καλά, όμως υλοποιείται από ένα τρίτης τάξης

Navier-Stokes	inpainting
συνάρτηση ροής Ψ	φωτεινότητα εικόνας u
ταχύτητα υγρού $v = \nabla^\perp \Psi$	διεύθυνση ισοφώτεινων $\nabla^\perp u$
στροβιλισμός $\omega = \Delta \Psi$	$w = \Delta u$
ιζώδες d	ανισοτροπική διάχυση v

Πίνακας 2.1: Αντιστοιχία μεγεθών κίνησης υγρού και εξέλιξης ισόφωτων καμπύλων

PDE μοντέλο του οποίου η ευστάθεια και η σύγκλιση δεν έχουν θεμελιωθεί μαθηματικά.

Το 2001 οι Bertalmio, Bertozzi, Sapiro[37], πρότειναν ένα νέο inpainting μοντέλο το οποίο ακολουθεί την ίδια τεχνική με το BCBS αλλά είναι βασισμένο πάνω στις εξισώσεις Navier-Stokes της ρευστοδυναμικής. Ο λόγος που εξετάστηκε ένα τέτοιο μοντέλο είναι ότι οι εξισώσεις αυτές είναι πολύ χρήσιμες στη ρευστοδυναμική και έχει γίνει βαθειά μελέτη ως προς τη σύγκλιση και την ευστάθειά τους.

Για ένα ασυμπιεστο υγρό που έχει ταχύτητα v , πίεση p και ιζώδες d ισχύουν οι παρακάτω εξισώσεις Navier-Stokes

$$v_t + v \cdot \nabla v = -\nabla p + d\Delta v, \quad \nabla \cdot v = 0 \quad (2.55)$$

Στις 2 διαστάσεις η συνάρτηση ροής του υγρού (Stream Function) Ψ ικανοποιεί την εξίσωση $\nabla^\perp \Psi = v$ ενώ ο στροβιλισμός της ταχύτητας (vorticity) $\omega = \nabla \times v$ προκύπτει από την (2.55) ότι ικανοποιεί την εξίσωση

$$\omega_t + v \nabla \cdot \omega = d\Delta \omega \quad (2.56)$$

ενώ ισχύει και ότι $\Delta \Psi = \omega$.

Από τις εξισώσεις αυτές προκύπτει ότι στην σταθερή κατάσταση ισχύει η σχέση

$$\nabla^\perp \Psi \cdot \nabla \Delta \Psi = 0 \quad (2.57)$$

Η αναλογία των εξισώσεων (2.57) με την (2.40) φανερώνει και την αναλογία μεταξύ της κίνησης ενός ασυμπιέστου υγρού και των ισοφώτεινων γραμμών μέσα στην inpainting περιοχή. Εισάγοντας τις αντιστοιχίες του πίνακα 2.1 προκύπτει το παρακάτω PDE inpainting μοντέλο

$$\frac{\partial w}{\partial t} + v \cdot \nabla w = v \nabla \cdot (g(|\nabla w|) \nabla w) \quad (2.58)$$

Παρατηρούμε ότι το παραπάνω επαναληπτικό σχήμα εφαρμόζεται στη λαπλασιανή της εικόνας $w = \Delta I$. Κάθε φορά η φωτεινότητα της εικόνας υπολογίζεται λύνοντας ένα πρόβλημα συνοριακών τιμών τύπου Poisson

$$\Delta u = w, \quad u|_{\partial\Omega} = u_0 \quad (2.59)$$

Το πλεονέκτημα της συγκεκριμένης μεθόδου είναι ότι έχει μοναδική λύση όπως προκύπτει από την ανάλυση των εξισώσεων Navier-Stokes. Η αριθμητική επίλυση, η θεωρητική θεμελίωση, άποτελέσματα και άλλες χρήσεις¹⁵ του Navier-Stokes inpainting παρουσιάζονται στο [37].

2.6.2 Inpainting με Εξέλιξη Επιπεδοσυνόλων

Το 2001 οι Ballester, Bertalmio, Caselles, Sapiro, Verdera [12] πρότειναν ένα μοντέλο inpainting το οποίο αποσκοπεί στην ομαλή εξέλιξη των ισοφωτεινων γραμμών μέσα στην inpainting περιοχή. Για το διαχωρισμό των ισοφωτεινων γραμμών η εικόνα αρχικά διασπάται στις εικόνες των επιπεδοσυνόλων, (level sets) για όλες τις στάθμες φωτεινότητας της εικόνας και ο αλγόριθμος εφαρμόζεται σε όλα τα επιπεδοσύνολα. Στη συνέχεια η τελική εικόνα παίρνεται από την ένωση των ανακατασκευασμένων επιπεδοσυνόλων.

Αρχικά δίνουμε τα παρακάτω στοιχεία από τη θεωρία επιπεδοσυνόλων που είναι απαραίτητα για την κατανόηση της μεθόδου.

Ορισμός 2.1. Για ένα σήμα f με πεδίο ορισμού το E το επιπεδοσύνολο $X_u(f)$ για τις διάφορες τιμές u ορίζονται ως

$$X_u(f) = [x \in E : f(x) \geq u], \quad -\infty < u < +\infty \quad (2.60)$$

Θεώρημα 2.4. Η ανακατασκευή ενός σήματος αν γνωρίζουμε όλα τα επιπεδοσύνολά του $X_u(f)$ γίνεται από τον τύπο

$$f(x) = \sup\{u : u \in X_u(f)\} \quad (2.61)$$

Στο μοντέλο inpainting η εικόνα¹⁶ διασπάται αρχικά στις εικόνες που προκύπτουν από τα επιπεδοσύνολα. Για κάθε τέτοια εικόνα επιδιώκεται η εύρεση των συναρτήσεων u, θ ώστε να ελαχιστοποιηθεί το παρακάτω συναρτησιακό

$$E_\Omega[\theta, u] = \int_\Omega |\nabla\theta|^p (a + b|\nabla k * u|) dx + \alpha \int_\Omega (|\nabla u| - \theta \nabla u) dx \quad (2.62)$$

¹⁵Το μοντέλο Navier-Stokes inpainting χρησιμοποιείται και για ψηφιακό zoom-in με πολύ καλά αποτελέσματα

¹⁶Θεωρούμε greyscale εικόνα με διακριτές στάθμες φωτεινότητας $[0, \dots, 255]$

όπου a, b, α θετικές σταθερές, k κάποιος πυρήνας εξομάλυνσης (π.χ. γκαουσιανός), u η συνάρτηση του επιπεδοσυνόλου και θ το διανυσματικό πεδίο της παραγωγού των επιπεδοσυνόλων σε μια ζώνη έξω από την περιοχή inpainting, δηλαδή μία συνάρτηση για την οποία έξω από την περιοχή Ω ισχύει $\theta(x) \cdot \nabla u(x) = |\nabla u(x)|$, $x \in \Omega^c$.

Για την εύρεση των συναρτήσεων u, θ παίρνουμε τις Euler παραγωγούς της ενέργειας E ως προς u, θ οπότε προκύπτουν τα 2 ακόλουθα επαναληπτικά PDE σχήματα.

$$\frac{\partial \theta}{\partial t} = -\nabla_{\theta} E(\theta, u) \quad (2.63)$$

$$\frac{\partial u}{\partial t} = -\nabla_u E(\theta, u) \quad (2.64)$$

Το πλεονέκτημα αυτής της μεθόδου είναι ότι τα PDE που προκύπτουν είναι 2ου βαθμού και αποδεικνύεται ότι για $p > 1$ υπάρχουν μοναδικές συναρτήσεις που ελαχιστοποιούν την εξ. (2.62). Πληροφορίες για την απόδειξη αυτή και γενικότερα την μαθηματική θεμελίωση της μεθόδου, την αριθμητική υλοποίηση καθώς και αποτελέσματα της μεθόδου αναφέρονται στα [12],[13]. Τέλος αξίζει να αναφερθεί ότι η συγκεκριμένη μέθοδος χρησιμοποιείται και για inpainting σε περισσότερες διαστάσεις *surface inpainting* [27].

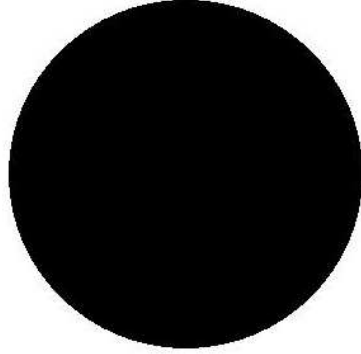
2.6.3 Inpainting με Μεθόδους Κατάτμησης Εικόνας

Το πρόβλημα της κατάτμησης εικόνας (segmentation) είναι ένα από τα πιο σημαντικά προβλήματα της όρασης υπολογιστών με πλήθος εφαρμογών¹⁷. Οι Mumford, Shah [47] πρότειναν ένα PDE μοντέλο για κατάτμηση εικόνων χωρίς υφή. Στο μοντέλο αυτό η εικόνα θεωρείται ως ένα σύνολο από λεία αντικείμενα (2 διαστάσεων) τα οποία ενώνονται μεταξύ τους κατά μήκος των ακμών τους. Αν Ω είναι το πεδίο της εικόνας και Γ το σύνολο των ακμών της εικόνας τότε το συναρτησιακό ενέργειας που προκύπτει είναι

$$E[u, \Gamma] = \frac{\gamma}{2} \int_{\Omega \setminus \Gamma} |\nabla u|^2 dx + \alpha H^1(\Gamma) \quad (2.65)$$

όπου H^1 είναι το μέτρο Hausdorff το οποίο για συνήθεις καμπύλες είναι ίσο με το μήκος τους. Οι Chan, Shen [14] και οι Tsai, Yezzi, Willsky [3] βασιζόμενοι στο μοντέλο αυτό πρότειναν το παρακάτω συναρτησιακό για inpainting (Mumford-Shah Inpainting Model)

¹⁷Εισαγωγή στο πρόβλημα της κατάτμησης, τρόποι επίλυσής του και πλήθος άλλων αναφορών μπορούν να βρεθούν σε πολλά βιβλία όρασης υπολογιστών (π.χ. [41]).



Σχήμα 2.21: Ομοιότητα μοντέλων TV, Mumford-Shah Inpainting

$$J_{MS}[u, \Gamma|u^0, D] = \frac{1}{2} \int_{\Omega} \lambda_D(u - u^0)^2 dx + \frac{\gamma}{2} \int_{\Omega \setminus \Gamma} |\nabla u|^2 dx + \alpha H^1 \quad (2.66)$$

Για την επίλυση του μοντέλου γίνεται χρήση της προσέγγισης Γ -σύγκλισης που χρησιμοποιήθηκε από τους Ambrosio, Tortotelli [5],[6] και η οποία οδηγεί σε έναν επαναληπτικό αλγόριθμο με πολύ γρήγορη σύγκλιση. Σύμφωνα με την προσέγγιση αυτή το σύνολο των ακμών Γ προσεγγίζεται από τη συναρτησή z_ϵ για την οποία ισχύει

$$z_\epsilon = \begin{cases} 0, & \text{κοντά στις ακμές (γειτονιά πάχους } O(\epsilon)) \\ 1, & \text{αλλού} \end{cases} \quad (2.67)$$

Εισάγωντας την παραπάνω προσέγγιση και θεωρώντας τις Euler παράγωγους ως προς τις u, z παίρνουμε το Euler-Lagrange σύστημα

$$\lambda_D(x)(u - u^0) - \gamma \nabla \cdot (z^2 \nabla u) = 0 \quad (2.68)$$

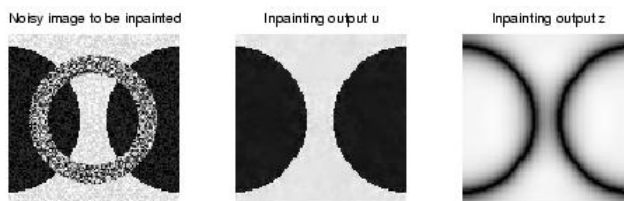
$$(\gamma |\nabla u|^2) z + \alpha (-2\epsilon \Delta z + \frac{z-1}{2\epsilon}) = 0 \quad (2.69)$$

το οποίο συγκλίνει ταχύτατα [19].

Το Mumford-Shah Inpainting Model έχει πολλές ομοιότητες με το TV inpainting μοντέλο. Ας εξετάσουμε ένα πολύ απλό παράδειγμα.

Στην εικόνα 2.21 έχουμε ένα μαύρο κύκλο ακτίνας r . Η TV νόρμα για αυτήν την εικόνα είναι

$$TV[u] = \int_{\Omega} |\nabla u| dx dy = 2\pi \int_0^\infty |\nabla u_r| dr = 2\pi r |u_1 - u_0| \quad (2.70)$$



Σχήμα 2.22: Παράδειγμα Mumford-Shah-Euler Inpainting: Αριστερα: Αρχική εικόνα, Κέντρο: Τελική εικόνα, Δεξιά: Τελική εικόνα της συνάρτησης z

όπου u_1, u_0 οι τιμές της εικόνας μέσα και έξω από τον κύκλο.

Για τη νόρμα του μοντέλου Mumford-Shah έχουμε

$$MS[u, \Gamma] = \frac{\gamma}{2} \int_{\Omega \setminus \Gamma} |\nabla u|^2 dx dy + \alpha H^1(\Gamma) = \alpha 2\pi r \quad (2.71)$$

Βλέπουμε ότι οι τιμές των δύο νορμών διαφέρουν μόνο κατά μία πολλαπλασιαστική σταθερά. Αυτό το αποτέλεσμα γενικεύεται και για πιο πολύπλοκες εικόνες.

Ποιοτικά, παρατηρώντας την (2.66) βλέπουμε ότι η ελαχιστοποίηση του συναρτησιακού γίνεται με την ελαχιστοποίηση του μήκους των ακμών όπως και στην περίπτωση του TV inpainting.

Οπότε επιβεβαιώνεται αυτό που λαμβάνεται και πειραματικά, ότι το Mumford-Shah Inpainting Model δεν λαμβάνει υπόψη του την καμπυλότητα αλλά ούτε ικανοποιεί την αρχή της συνεκτικότητας. Οι Esedoglu, Chen [19] εισήγαγαν στο μοντέλο γι' αυτό το σκοπό την Euler's elastica. Το τροποποιημένο συναρτησιακό προκύπτει (Mumford-Shah-Euler Inpainting Model)

$$J_{MS}[u, \Gamma | u^0, D] = \frac{1}{2} \int_{\Omega} \lambda_D (u - u^0)^2 dx + \frac{\gamma}{2} \int_{\Omega \setminus \Gamma} |\nabla u|^2 dx + \int_{\gamma} (\alpha + \beta \kappa^2) ds. \quad (2.72)$$

Για την αριθμητική επίλυση του παραπάνω συναρτησιακού στο [19] παρουσιάζονται 2 τεχνικές.

- Η μέθοδος των επιπεδοσυνόλων από τους Osher, Sethian
- Η μέθοδος προσέγγισης της Γ -σύγκλισης από τον De Giorgi

Τα αποτελέσματα του Mumford-Shah-Euler Inpainting Model είναι παρόμοια με αυτά του Euler's Elastica inpainting. Ενδεικτικά δείχνουμε μία εικόνα από το [19].

Περισσότερες πληροφορίες για τα inpainting μοντέλα βασισμένα πάνω σε αλγόριθμους κατάτμησης μπορούν να βρεθούν στα [19],[14],[59]

Κεφάλαιο 3

Υφή

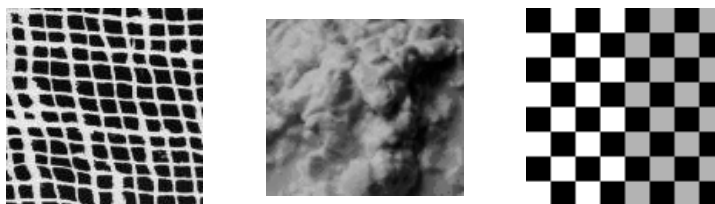
Οι αλγόριθμοι Inpainting που παρουσιάστηκαν στο κεφ. 2, αφορούν όπως είδαμε το περιεχόμενο της εικόνας και δεν δίνουν καλά αποτελέσματα σε εικόνες με υφή, αφού συμπληρώνουν τα διαφορετικά μέρη της Inpainting περιοχής με λείο και ομογενή τρόπο που συνεπάγεται την απουσία της υφής. Όμως στις περισσότερες εικόνες της φύσης υπάρχει υφή, οπότε μια ολοκληρωμένη αντιμετώπιση του Inpainting προβλήματος θα πρέπει να αντιμετωπίζει τις εικόνες που περιέχουν και υφή, εξίσου αποτελεσματικά.

3.1 Εισαγωγή

Πριν προχωρήσουμε στην προσέγγιση του Inpainting προβλήματος για περιπτώσεις εικόνων με υφή, ας δούμε κάποια εισαγωγικά περί υφής.

Η υφή (*texture*) είναι μια πολύ κεντρική έννοια στην όραση υπολογιστών ενώ παράλληλα υπάρχουν καθημερινά παραδείγματα υφής στον πραγματικό κόσμο. Όμως παρόλο που είναι πολύ εύκολο να αντιληφθούμε την υφή σε μία εικόνα με το ανθρώπινο μάτι, είναι πολύ δύσκολο να δωθεί ένας γενικός ορισμός της. Στην εικόνα 3.1 φαίνονται κάποια παραδείγματα εικόνων υφής.

Αυτό που μπορούμε να σημειώσουμε είναι ότι οι εικόνες υφής αποτελούνται από πρότυπα, (patterns), που επαναλαμβάνονται στην εικόνα σε μία ή περισσό-



Σχήμα 3.1: Παραδείγματα εικόνων υφής



Σχήμα 3.2: Παράδειγμα εικονας με υφή στον αληθινό κόσμο. Η εικόνα περιέχει πολλά διαφορετικά αντικείμενα όπου το καθ'ένα έχει τη δική του υφή.

τερες κλίμακες και που εμφανίζουν κάποια χωρική περιοδικότητα είτε σταθερή είτε όχι. Τα πρότυπα αυτά, που αποτελούν τις δομικές μονάδες μίας εικόνας υφής, ονομάζονται *texels* και η μορφή τους στο σύνολο της εικόνας εμφανίζει κάποιο στοχαστικό χαρακτήρα. Ανάλογα με αυτόν το στοχαστικό χαρακτήρα η υφή μπορεί να χαρακτηριστεί από τελείως στοχαστική (π.χ. λευκός θόρυβος), έως τελείως κανονική (π.χ. σκακιέρα).

Η υφή, σε αντίθεση με τη φωτεινότητα και το χρώμα, είναι μια ιδιότητα που αφορά μια ομάδα από pixel. Ένα pixel δεν έχει υφή, ενώ για κάθε επιφάνεια υπάρχει μια κλίμακα όπου η επιφάνεια δεν έχει υφή.

Εκτός από τις εικόνες υφής, που μπορούν να θεωρηθούν οι εικόνες της 3.1 υπάρχουν και οι εικόνες που δεν είναι αποκλειστικά υφής αλλά έχουν και περιεχόμενο, (structure), και οι οποίες είναι οι πιο συνηθισμένες εικόνες στη φύση (εικ. 3.2).

Η ύπαρξη υφής στο σύνολο σχεδών των εικόνων, την έχει αναγάγει σε πολύ σημαντική έννοια της όρασης υπολογιστών και έχει οδηγήσει στην ανάπτυξη του ξεχωριστού κλάδου της *ανάλυσης υφής*, (texture analysis) που ασχολείται αποκλειστικά με θέματα υφής.

Τα κυριότερα προβλήματα του κλάδου της ανάλυσης υφής είναι:

1. **Σύνθεση Υφής (Texture Synthesis):** Δεδομένης μίας εικόνας υφής ζητείται η κατασκευή μιας άλλης εικόνας με παρόμοια υφή αλλά όχι ακριβώς την ίδια. Το πρόβλημα της σύνθεσης υφής αποτελεί και τη βάση για το Inpainting υφής, καθώς η αντιμετώπιση του τελευταίου γίνεται με αλγόριθμους σύνθεσης υφής, και θα εξεταστεί αναλυτικά στην ενότητα 3.2. Άλλες εφαρμογές βρίσκονται στη γραφική με υπολογιστές αλλά και στη συμπίεση εικόνας.
2. **Ταξινόμηση υφής (Texture Classification):** Το πρόβλημα εδώ είναι η ταξινόμηση μίας εικόνας υφής σε μία ομάδα εικόνων υφής που είναι ήδη γνωστές και σχετίζεται με τον κλάδο της αναγνώρισης προτύπων. Οι

εφαρμογές της ταξινόμησης υφής είναι πολύ σημαντικές στην επεξεργασία ιατρικής εικόνας καθώς και στην ταξινόμηση στοιχείων του εδάφους (χώρα, βλάστηση) από εικόνες που λαμβάνονται από αεροφωτογραφίες.

- 3. Κατάτμηση υφής (Texture Segmentation):** Το πρόβλημα εδώ είναι ο διαχωρισμός των διαφορετικών ειδών υφής που υπάρχουν σε μία εικόνα και σχετίζεται με το πρόβλημα της ταξινόμησης υφής αφού μπορεί να αποτελεί το αρχικό στάδιο αυτού, ενώ μπορεί να εφαρμοστεί και γενικότερα για κατάτμηση εικόνας.

Αυτό που έχει ενδιαφέρον να σημειώσουμε, είναι ότι αφού δεν υπάρχει κάποιος ακριβής μαθηματικός ορισμός για την υφή, δεν υπάρχουν και μαθηματικές μεθόδους που να αξιολογούν τις διάφορες μεθόδους για τα παραπάνω προβλήματα. Η αξιολόγηση γίνεται κυρίως με το ανθρώπινο μάτι. Στην περίπτωση των προβλημάτων της ταξινόμησης και λιγότερο της κατάτμησης υφής κάποια αξιολόγηση μπορεί να γίνει με βάση κάποια σύγκριση με υπάρχουσες βάσεις δεδομένων για εικόνες υφής, όμως στην περίπτωση της σύνθεσης υφής, που είναι και το πρόβλημα που θα μας απασχολήσει, η αξιολόγηση γίνεται αποκλειστικά από το ανθρώπινο μάτι.

Τέλος, σε αυτήν τη μικρή εισαγωγή που κάνουμε, αξίζει να πούμε ότι ο τρόπος με τον οποίο το ανθρώπινο μάτι αντιλαμβάνεται την υφή αποτελεί αντικείμενο έρευνας πολλών επιστημονικών κλάδων όπως η ψυχοφυσική και η νευροφυσιολογία. Για περισσότερες πληροφορίες δες π.χ. [28], [26].

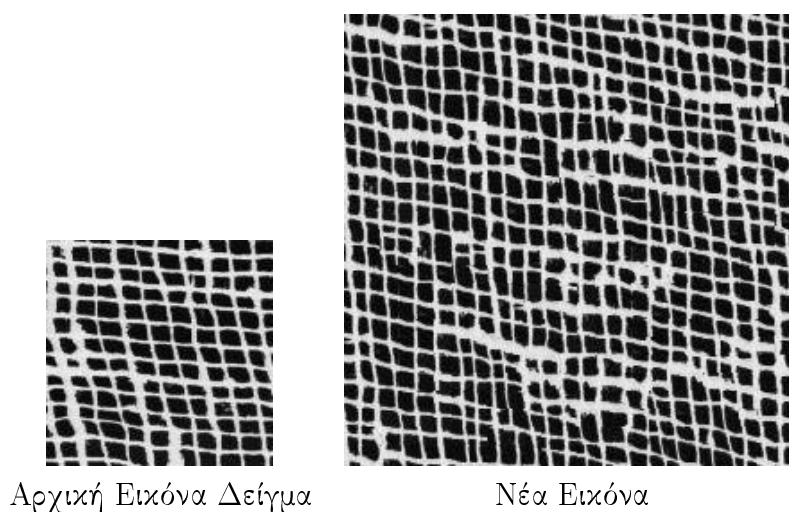
3.2 Σύνθεση Υφής

3.2.1 Εισαγωγή - Πρώτες Προσπάθειες

Το πρόβλημα της σύνθεσης υφής μπορεί να διατυπωθεί ως εξής: Δεδομένης μιας εικόνας υφής, ο στόχος είναι η σύνθεση μιας άλλης εικόνας με την ίδια υφή. Η μόνη υπόθεση που κάνουμε για την αρχική εικόνα, είναι ότι αυτή είναι τέτοιου μεγέθους ώστε να φαίνεται η χωρική περιοδικότητά της, δηλαδή να υπάρχουν ακέραιες οι δομικές μονάδες της texels. Σχηματικά το πρόβλημα της σύνθεσης υφής φαίνεται στο σχήμα 3.3.

Γενικά, όπως θα φανεί και στη συνέχεια, όλοι οι αλγόριθμοι σύνθεσης υφής είναι βασισμένοι σε μεθόδους ανάλυσης και μοντελοποίησης υφής, αφού η σύνθεση αποτελεί κατά κάποιο τρόπο την αντίστροφη διαδικασία της ανάλυσης.

Η πρώτη προσπάθεια σύνθεσης υφής ανήκει στον Garber το 1981 [20]. Στο μοντέλο του, η τιμή για κάθε pixel υπολογιζόταν ως ένας γραμμικός συνδιασμός των γειτονικών του pixel συν κάποιος στατικός θόρυβος. Το μοντέλο του, που θυμίζει το μοντέλο γραμμικής πρόβλεψης οδηγούμενης από λευκό θόρυβο



Σχήμα 3.3: Παράδειγμα σύνθεσης υφής

(white noise driven, linear prediction model) από την ψηφιακή επεξεργασία σήματος, που όμως χρησιμοποιείται και στην ανάλυση υφής, έδωσε σχετικά ικανοποιητικά αποτελέσματα ως μία πρώτη προσπάθεια.

Το 1995 οι Heeger, Bergen [25] πρότειναν ένα μοντέλο επηρεασμένο από τις θεωρίες της εποχής για την αντίληψη της υφής από το ανθρώπινο μάτι. Σύμφωνα με τις αντιλήψεις αυτές, (δες π.χ. [7]), το ανθρώπινο μάτι δύσκολα διακρίνει διαφορές σε 2 δείγματα υφής εάν αυτά έχουν παρόμοιες αποκρίσεις σε μια σειρά από γραμμικά φίλτρα. Έτσι, θεώρησαν ότι όλες οι ιδιότητες της υφής, εντοπίζονται μέσω των αποκρίσεων της εικόνας στα φίλτρα αυτά. Τα φίλτρα αυτά αποτελούσαν μία **Laplacian πυραμίδα** ή ακόμα καλύτερα μια **Steerable πυραμίδα**. Τα αποτελέσματα της μεθόδου των Heeger, Bergen ήταν πολύ καλά για εικόνες υφής με στοχαστικό χαρακτήρα, αλλά όχι ικανοποιητικά για εικόνες υφής με σταθερή περιοδικότητα.¹

Το 1997 ο De Bonet [11], χρησιμοποίησε και αυτό την ανάλυση σε πολλαπλές κλίμακες. Η μέθοδος αποτελείται από 2 στάδια. Πρώτα γίνεται η ανάλυση της εικόνας υφής σε πολλαπλές κλίμακες έτσι ώστε να εξαχθούν τα χαρακτηριστικά των texels. Κατά τη σύνθεση η νέα εικόνα αρχικοποιείται με τα texels και στη συνέχεια περνάει από μία διαδικασία τυχαιοποίησης (randomization), που δεν αλλάζει τα χαρακτηριστικά των texels ώστε αυτά να ενώνονται αρμονικά μεταξύ τους. Η μέθοδος αυτή, όπως και η προηγούμενη δίνει καλά αποτελέσματα μόνο σε περιπτώσεις υφής με ισχυρά στοχαστικό χαρακτήρα, αφού σε σταθερά περιοδικές περιπτώσεις υφής, η διαδικασία τυχαιοποίησης αλλοιώνει τα

¹Περισσότερα στοιχεία για τη συγκεκριμένη μέθοδο, αποτελέσματα αλλά και στοιχεία για τις Laplacian και Steerable πυραμίδες μπορούν να βρεθούν στο [25]

χαρακτηριστικά. Ακόμα, η αρχικοποίηση της νέας εικόνας με το ίδιο δείγμα, υποθέτει ότι η αρμονική ένωση των δειγμάτων μεταξύ τους είναι δυνατή, πράγμα που δεν είναι απαραίτητα σωστό.

Σημαντική, ήταν και η προσφορά των Simoncelli, Portilla [51, 58], όπου η σύνθεση υφής βασίζεται στα χαρακτηριστικά των συντελεστών wavelets πρώτης και δεύτερης τάξης. Η μέθοδος δίνει πολύ καλά αποτελέσματα για περιπτώσεις υφής και με στοχαστικό χαρακτήρα και με σταθερή περιοδικότητα, με κάποια λάθη στην πληροφορία υψηλής συχνότητας για περιπτώσεις υφής με σταθερή περιοδικότητα.

Ο πιο σημαντικός αλγόριθμος σύνθεσης υφής είναι αυτός των Efros, Leung [17]. Ο αλγόριθμος αυτός είναι εξαιρετικά απλός, ενώ δίνει πολύ εντυπωσιακά αποτελέσματα και γι'αυτό τον εξετάζουμε αναλυτικά στη συνέχεια.

3.2.2 Ο Αλγόριθμος των Efros-Leung

Στον αλγόριθμο των Efros-Leung η υφή μοντελοποιείται ως ένα *Τυχαίο Μαρκοβιανό Πεδίο*, (*Markov Random Field - MRF*).² Αυτό σημαίνει ότι η συνάρτηση κατανομής της φωτεινότητας ενός pixel, δεδομένων των τιμών φωτεινότητας των pixel που βρίσκονται στη γειτονιά του είναι ανεξάρτητη από την υπόλοιπη εικόνα. Ως γειτονιά ενός pixel θεωρούμε ένα τετράγωνο περιτήσ διάστασης με κέντρο το συγκεκριμένο pixel. Η διάσταση του τετραγώνου επιλέγεται μεγαλύτερη ή ίση από τη μεγαλύτερη διάσταση της δομικής μονάδας texel της εικόνας της υφής έτσι ώστε μέσα στη γειτονιά ενός pixel να περιέχονται όλα τα χαρακτηριστικά του δείγματος υφής.

Η σύνθεση υφής γίνεται pixel προς pixel. Έστω I_s η αρχική εικόνα ή δείγμα υφής που έχουμε και I η εικόνα που θέλουμε να συνθέσουμε.

Ας υποθέσουμε αρχικά ότι έχουμε συνθέσει όλη την εικόνα I εκτός από ένα pixel p . Ορίζουμε $\omega(p) \subset I$ την τετραγωνική εικόνα της γειτονιάς του p με κέντρο το p . Επίσης ορίζουμε ένα μέτρο διαφοράς $d(\omega_1, \omega_2)$ μεταξύ δύο τετραγωνικών εικόνων. Τέλος θεωρούμε μία τετραγωνική εικόνα $\omega \subset I_s$.

Η εικόνα ω που ταιριάζει καλύτερα με την εικόνα ω_p δίνεται από την σχέση $\omega_{best} = \arg \min_{\omega} d(\omega(p), \omega)$. Εμείς επιλέγουμε το σύνολο των εικόνων ω έτσι ώστε $d(\omega(p), \omega) < (1 + \varepsilon)d(\omega(p), \omega_{best})$ ³. Από το σύνολο αυτό επιλέγουμε τυχαία μια εικόνα ω_{can} . Η τιμή της εικόνας I στο pixel p θα είναι η τιμή του pixel που βρίσκεται στο κέντρο της επιλεγμένης εικόνας ω_{can} . Αυτό που μένει να ορίσουμε είναι τη διαφορά d μεταξύ δύο εικόνων. Η διαφορά αυτή ορίζεται ως το άθροισμα των τετραγώνων των διαφορών των δύο εικόνων, για τα pixel όπου

²Η μοντελοποίηση υφής ως Markov Random Field - MRF ή ως Gibbs Random Field χρησιμοποιείται ευρύτατα στην ανάλυση υφής.

³Το ε είναι μία μικρή θετική σταθερά π.χ. $\varepsilon = 0.05$.

οι εικόνες έχουν τιμές. Επίσης αν θέλουμε να κανονικοποιήσουμε τη διαφορά δίνοντας περισσότερο βάρος στα pixel που είναι κοντά στο κέντρο μπορούμε να εφαρμόσουμε συνέλιξη της εικόνας που προκύπτει από την τετραγωνική διαφορά των δύο εικόνων με έναν γκαουσιανό πυρήνα G ίδιου μεγέθους. Τα παραπάνω για τη διαφορά μπορούν να εκφραστούν με την εξίσωση.

$$d(\omega(p), \omega) = \sum_i \sum_j [M(i, j) \cdot (\omega(p)(i, j) - \omega(i, j))^2] * G(i, j) \quad (3.1)$$

όπου

$$M(i, j) = \begin{cases} 1, & \text{Αν υπάρχει το } \omega(p)(i, j) \\ 0, & \text{Το συγκεκριμένο pixel της εικόνας } I \text{ δεν έχει συνθεθεί ακόμα} \end{cases} \quad (3.2)$$

Όταν πρέπει να συνθεθούν περισσότερα του ενός pixel θα πρέπει να υπολογισθούν τα από κοινού μεγέθη σύμφωνα με την μοντελοποίηση της υψής ως MRF. Κάτι τέτοιο όμως γίνεται υπολογιστικά αδύνατο για εικόνες πραγματικών διαστάσεων. Αντί αυτού οι Efros, Leung εμπνευσμένοι από ένα άρθρο του Shannon το 1948 [56]⁴, ακολούθησαν διαφορετική προσέγγιση.

Αρχικά τοποθετούμε το δείγμα υψής μέσα στην εικόνα I . Η σύνθεση θα γίνει pixel προς pixel αρχίζοντας από το όριο όπου βρίσκεται το δείγμα I_s και πηγαίνοντας προς το όριο της εικόνας I , ακολουθώντας μια σπειροειδή διαδρομή.

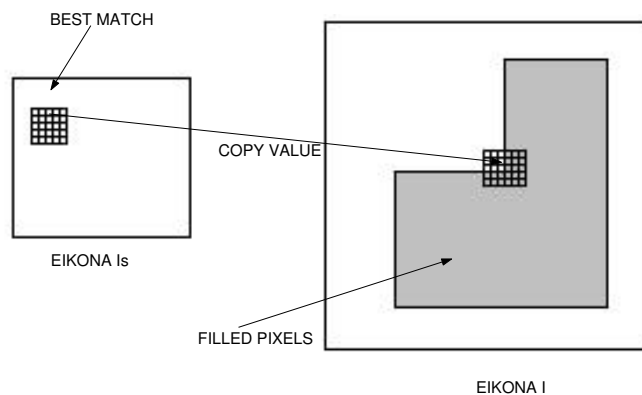
Για κάθε pixel, ακολουθούμε την διαδικασία που περιγράψαμε παραπάνω. Η διαφορά εδώ είναι ότι το κάθε pixel δεν έχει όλα τα γειτονικά του στοιχεία συμπληρωμένα. Αυτό ενσωματώνεται στην εξ. (3.2) μέσω της μάσκας M .

Η διαδικασία που ακολουθεί ο αλγόριθμος Efros-Leung φαίνεται στο σχήμα 3.4

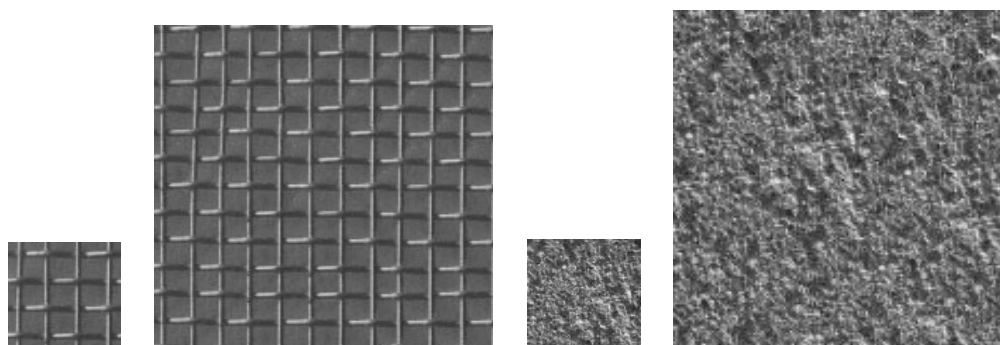
Ο αλγόριθμος των Efros-Leung δίνει πολύ καλά αποτελέσματα τόσο για εικόνες υψής ισχυρά στοχαστικού χαρακτήρα, όσο και για εικόνες υψής με σταθερή περιοδικότητα όπως φαίνεται και από την εικόνα 3.5

Αυτό που πρέπει να τονιστεί, είναι ότι η μόνη παράμετρος που θέτει ο χρήστης είναι το μήκος της γειτονιάς του κάθε pixel που αντιστοιχεί και στο μήκος της τετραγωνικής εικόνας (παράθυρο) που διατρέχει την εικόνα - δείγμα I_s για να βρει το κατάλληλο pixel. Αναφέραμε πριν ότι το μήκος αυτό πρέπει να είναι τουλάχιστον ίσο με τη μεγαλύτερη διάσταση της δομική μονάδας της υψής, $texel$, ώστε η σύνθεση της νέας εικόνας να γίνει σωστά. Αυτό φαίνεται καθαρά και από την εικόνα 3.6.

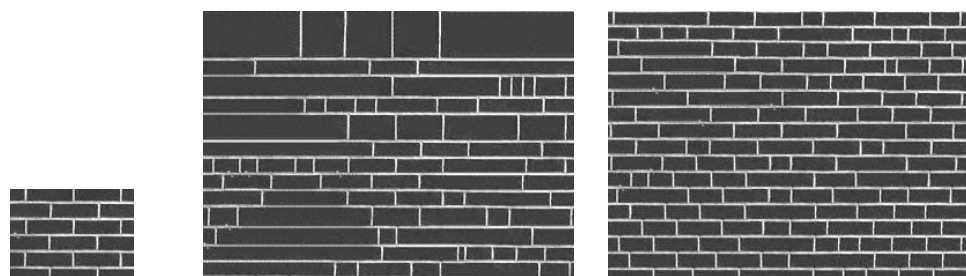
⁴ Στο άρθρο αυτό ο Shannon περιέγραψε τη δημιουργία τυχαίων λέξεων της αγγλικής με n γράμματα (ή προτάσεων με n λέξεις), παίρνοντας ως δείγμα ένα βιβλίο και μοντελοποιώντας την ακολουθία των γραμμάτων ως μια αλυσίδα Markov. Η ιδέα αυτή βρίσκει εφαρμογή και σε προηγμένες μεθόδους αναγνώρισης φωνής με μεγάλο λεξιλόγιο.



Σχήμα 3.4: Ο αλγόριθμος των Eros-Leung



Σχήμα 3.5: Παραδείγματα σύνθεσης υφής με τον αλγόριθμο των Eros-Leung

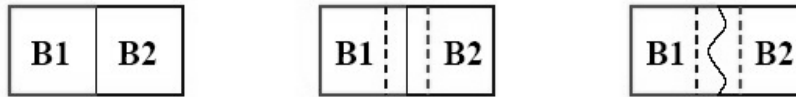


Εικόνα Δείγμα

Σύνθεση με μήκος 11

Σύνθεση με μήκος 23

Σχήμα 3.6: Σύνθεση υφής για διαφορετικές τιμές του παραθύρου. Από το [17].



Σχήμα 3.7: Επικάλυψη των δύο εικόνων

3.2.3 Επεκτάσεις του Αλγορίθμου Efros-Leung

Ο αλγόριθμος Efros-Leung δίνει πολύ καλά αποτελέσματα για τις εικόνες υψής που είδαμε στην προηγούμενη ενότητα. Το μόνο μειονέκτημα που έχει ο συγκεκριμένος αλγόριθμος είναι ότι είναι πολύ αργός. Αυτό οφείλεται στο γεγονός ότι ο αλγόριθμος συνθέτει την εικόνα pixel προς pixel και ακολουθεί την ίδια διαδικασία για κάθε pixel. Αν M_1, N_1, M_2, N_2, W οι διαστάσεις της εικόνας προς σύνθεση, της εικόνας δείγμα και η διάσταση του τετραγωνικού παραθύρου τότε η πολυπλοκότητα του αλγορίθμου είναι της τάξης του $O(M_1 N_1 M_2 N_2 W^2)$.

Επειδή για εικόνες προς σύνθεση με μεγάλες διαστάσεις, η διάρκεια του αλγορίθμου γίνεται πολύ μεγάλη (πάνω από 1 ώρα) θα αναφέρουμε μερικές μεθόδους που τον επιταγχύνουν.

Μια πολύ απλή, αλλά αποτελεσματική, διαδικασία για την επιτάχυνση του αλγορίθμου είναι να επιλέγει ο χρήστης την εικόνα δείγμα από όπου θα αντιγραφεί η υψή. Έτσι όταν η αρχική εικόνα είναι πολύ μεγάλη, δηλαδή περιέχει πάρα πολλές επαναλήψεις των texels, μπορούμε να επιλέξουμε ένα μικρό τμήμα της εικόνας δείγμα (αρκετά μεγάλο όμως ώστε να φαίνεται ξεκάθαρα η δομή της υψής), μειώνοντας σημαντικά το συνολικό χρόνο του αλγορίθμου χωρίς απώλειες στην ποιότητα του αλγορίθμου. Ένα μειονέκτημα της συγκεκριμένης μεθόδου είναι ότι η διαδικασία επιλογής της εικόνας δείγμα από τον χρήστη, εισάγει μία ακόμα παράμετρο στην διαδικασία, και έτσι μειώνεται το ποσοστό της αυτοματοποίησης της μεθόδου, σε περίπτωση που κάτι τέτοιο είναι επιθυμητό.

Αλγόριθμος των Efros-Freedman

Το 2001 οι Efros-Freedman [16] πρότειναν μια βελτίωση του αλγορίθμου Efros-Leung από πλευράς ταχύτητας, χωρίς να μειώνεται η ποιότητα των αποτελεσμάτων. Η βασική ιδέα είναι κάθε φορά που βρίσκεται η κατάλληλη τετραγωνική εικόνα από όπου αντιγράφεται το κεντρικό pixel να μην αντιγράφεται μόνο αυτό, αλλά όλα τα pixel της εικόνας που αντιστοιχούν σε μη συμπληρωμένα pixel. Επίσης η τετραγωνική εικόνα δεν τοποθετείται με κέντρο το προς σύνθεση pixel, αλλά επικαλύπτει την εικόνα προς σύνθεση ώστε να βρεθεί το σημείο που αυτές ταιριάζουν καλύτερα. Σχηματικά αυτό φαίνεται από την εικόνα 3.7.

Αν $B1$ η εικόνα προς σύνθεση και $B2$ η τετραγωνική εικόνα, τότε αρχικά

η εικόνα $B2$ ολισθαίνει μέσα στην $B1$ κατά ένα συγκριμένο μέγεθος στηλών. Στη συνέχεια κάθε γραμμή της εικόνας που προκύπτει από την ένωση των $B1$ και $B2$ βρίσκουμε πόσα pixel από την κάθε εικόνα θα κρατήσουμε ώστε οι δύο εικόνες να ταιριάζουν καλύτερα. Περισσότερες πληροφορίες καθώς και αποτελέσματα παρουσιάζονται στο [16] ενώ στο [35] παρουσιάζεται ένας ακόμα πιο γρήγορος αλγόριθμος (πραγματικού χρόνου) που βασίζεται στην ίδια ιδέα.

Σύνθεση υφής με ανάλυση σε πολλαπλές κλίμακες

Ένας ακόμα αλγόριθμος για σύνθεση υφής, παρόμοιος με τον αλγόριθμο των Efros-Leung, αλλά αρκετά πιο γρήγορος προτάθηκε από τους Wei, Levoy το 2000 [64]. Με τον αλγόριθμο αυτό η εικόνα προς σύνθεση αρχικοποιείται με τυχαίες τιμές (λευκός θόρυβος). Στη συνέχεια κατασκευάζονται 2 πυραμίδες (π.χ. γκαουσιανές) μία για την εικόνα δείγμα και μία για την εικόνα σύνθεσης. Αρχικά συνθέτεται η εικόνα που έχει τη μικρότερη διάσταση (κορυφή πυραμίδας) με αλγόριθμο παρόμοιο με αυτό των Efros-Leung. Στη συνέχεια η σύνθεση για κάθε επίπεδο της πυραμίδας γίνεται χρησιμοποιώντας στοιχεία και από την ίδια εικόνα αλλά και από το προηγούμενο επίπεδο της πυραμίδας. Η ανάλυση σε πολλαπλές κλίμακες επιταγχύνει τον αλγόριθμο γιατί η τιμή του παραθύρου μπορεί να κρατηθεί μικρή αφού αυτή η τιμή για κάποιο επίπεδο της πυραμίδας θα μπορεί να ενσωματώσει τα χαρακτηριστικά του texel.

Πέρα από αυτή την επιτάχυνση, προτείνεται και η εφαρμογή τεχνικών διανυσματικού κβαντισμού σε μορφή δένδρου (*Tree Structured Vector Quantization*). Η τεχνική αυτή αποτελεί μια συνηθισμένη τεχνική για συμπίεση δεδομένων [22], και μπορεί να οδηγήσει σε επιτάχυνση της διαδικασίας κατά 2 τάξεις μεγέθους με μικρή απώλεια (θόλωμα) στα τελικά αποτελέσματα. Αναλυτικότερες πληροφορίες για τον αλγόριθμο των Wei, Levoy αναφέρονται στο άρθρο τους [64]

3.3 Inpainting με Μεθόδους Σύνθεσης Υφής

Η έννοια της σύνθεσης υφής, υπήρχε στον κλάδο της όρασης υπολογιστών πολύ πριν από την εισαγωγή του Inpainting προβλήματος. Όταν εισήχθη το Inpainting πρόβλημα, είχαν ήδη αναπτυχθεί αλγόριθμοι για σύνθεση υφής που έδιναν πολύ καλά αποτελέσματα. Επίσης η συγκαίνια των 2 προβλημάτων, αφού ο στόχος είναι η παραγωγή νέας εικόνας, καθώς και η αδυναμία των Inpainting αλγορίθμων να αναπαραγάγουν την υφή της εικόνας, οδήγησε στην εφαρμογή των αλγορίθμων σύνθεσης υφής στο Inpainting πρόβλημα.

3.3.1 Ο Αλγόριθμος Efros-Leung για Inpainting Εικόνων Υφής

Τα πολύ καλά αποτελέσματα του αλγορίθμου των Efros-Leung οδηγούν στην υπόθεση ότι μπορούν να εφαρμοστούν και για Inpainting σε εικόνες υφής. Ο αλγόριθμος παραμένει ο ίδιος με κάποιες μικρές αλλαγές.

Αρχικά έχουμε την εικόνα I , και μία περιοχή $D \subset I$ όπου θέλουμε να εφαρμόσουμε το Inpainting. Η εικόνα I_s από την οποία παίρνουμε τα δείγματα είναι $I_s = \{(x, y) : (x, y) \in I \setminus D\}$. Επίσης αυτό που αλλάζει είναι η σειρά σύνθεσης των pixel. Η σειρά σύνθεσης των pixel μοιάζει πολύ με τη σειρά onion peel που χρησιμοποιήσαμε στους αλγόριθμους για Inpainting περιεχομένου στο κεφάλαιο 2 και περιγράφεται αναλυτικά στη συνέχεια.

Έστω W η δυαδική συνάρτηση με

$$W(i, j) = \begin{cases} 1, & \text{το σημείο } (i, j) \text{ έχει συνθεθεί} \\ 0, & \text{το σημείο } (i, j) \text{ δεν έχει συνθεθεί} \end{cases} \quad (3.3)$$

και έστω B το μορφολογικό στοιχείο

$$B = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 0 \end{bmatrix},$$

δηλαδή ένας δίσκος ακτίνας 1.

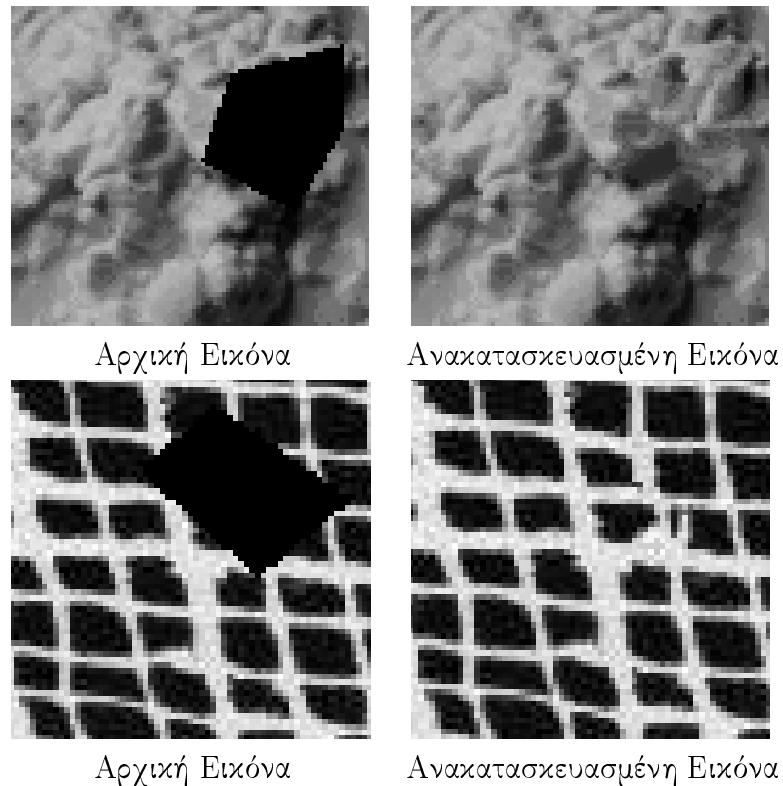
Τα επόμενα σημεία που θα συνθεθούν είναι το σύνολο

$$P = \{(x, y) : (x, y) \in (W \oplus B) \setminus W\},$$

όπου $W \oplus B$ είναι η μορφολογική διαστολή, (morphological dilation), του συνόλου W με το στοιχείο B .

Τα pixel του συνόλου P ανακατεύονται με τυχαίο τρόπο και μετά ταξινομούνται με φθίνουσα σειρά ως προς τον αριθμό των γειτονικών τους pixel που έχουν συνθεθεί, και γειτονικά σημεία μπορούν να θεωρηθούν ή όλα τα σημεία του παραθύρου ή απλώς τα σημεία της γειτονιάς των 8 σημείων. Η ταξινόμηση γίνεται γιατί όσο πιο πολλά γειτονικά συμπληρωμένα pixel έχει ένα σημείο τόσο περισσότερη πληροφορία έχουμε για το τι μέρος της δομικής μονάδας texel αντιπροσωπεύει, οπότε επιλέγουμε καλύτερη τιμή.

Από την ταξινόμηση αυτή προκύπτει μία νέα λίστα. Για κάθε pixel της νέας λίστας ακολουθούμε την ίδια διαδικασία με την περίπτωση της σύνθεσης. Βρίσκουμε το σύνολο των εικόνων $\omega \subset I_s$ που ικανοποιούν τη σχέση $d(\omega(p), \omega) < (1 + \varepsilon)d(\omega(p), \omega_{best})$ και επιλέγουμε 1 από αυτές ω_{can} , με τυχαίο τρόπο. Οι διαφορές γίνονται μόνο στα pixel όπου οι εικόνες έχουν τιμή και αυτό εξασφαλίζεται μέσω της μάσκας M εξ. (3.2). Ο λόγος που δεν επιλέγουμε



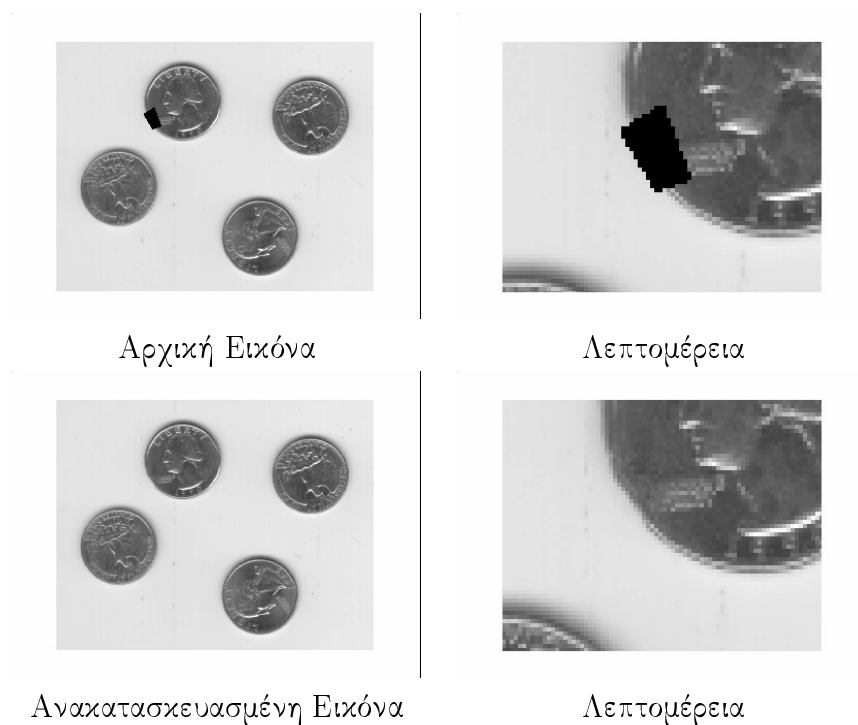
Σχήμα 3.8: Παράδειγματα Inpainting υφής με τον αλγόριθμο Efros-Leung

πάντα την καλύτερη εικόνα ω_{best} είναι για να μην έχουμε αντιγραφή της εικόνας ω_{best} στην εικόνα I αλλά μία παρόμοια εικόνα με την ίδια υφή. Το κεντρικό pixel της ω_{can} αντιγράφεται στη θέση του ζητούμενου pixel της εικόνας I .

Η απλή αυτή διαδικασία δίνει πολύ καλά αποτελέσματα όπως φαίνεται από τις εικόνες 3.8.

Όπως στη σύνθεση υφής, έτσι και εδώ η μόνη παράμετρος που πρέπει να επιλέξουμε είναι το μήκος του παραθύρου. Όπως έχουμε σημειώσει, για καλά αποτελέσματα το μήκος πρέπει να είναι τουλάχιστον ίσο με την μεγαλύτερη διάσταση της δομικής μονάδας *texel*. Στο πρώτο παράδειγμα της εικόνας 3.8 η εικόνα υφής έχει έντονα στοχαστικό χαρακτήρα χωρίς να μπορεί να προσδιοριστεί η μονάδα *texel*. Η διάσταση του παραθύρου επιλέχθηκε ίση με 5. Στο δεύτερο όμως παράδειγμα η εικόνα υφής έχει μια καθαρά περιοδική δομή. Το μήκος του παραθύρου επιλέχθηκε εδώ ίσο με 13. Το παράθυρο με τη διάσταση αυτή ενσωματώνει τις ιδιότητες της μονάδας *texel* και έτσι μπορεί να τις μεταφέρει με σώστο τρόπο (συνέχιση των ακμών, διατήρηση της δομής) στην ανακατασκευασμένη εικόνα.

Τα πολύ καλά αποτελέσματα που δίνει ο αλγόριθμος για Inpainting εικόνων



Σχήμα 3.9: Inpainting εικόνας υφής και περιεχομένου με τον αλγόριθμο Efron-Leung

υφής, δίνουν το ερέθισμα για εφαρμογή του αλγορίθμου και σε άλλες εικόνες που έχουν και υφή και περιεχόμενο. Τα αποτελέσματα είναι και εκεί εξίσου εντυπωσιακά.

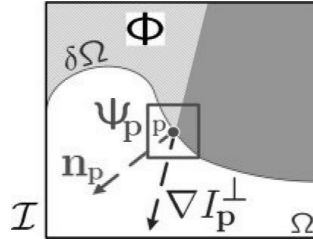
Στις εικόνες του σχήματος 3.9 βλέπουμε ότι η εικόνα έχει ανακατασκευασθεί, ενώ σε πολύ μεγάλο βαθμό έχει διατηρηθεί και η καμπυλότητα του σχήματος.

Στις εικόνες του σχήματος 3.10, το κείμενο έχει απομακρυνθεί με τέτοιο τρόπο ώστε να μην είναι δυνατό το ανθρώπινο μάτι να καταλάβει, ότι η τελική εικόνα είναι παράγωγο κάποιας επεξεργασίας. Η υφή που έχει αντιγραφεί προέρχεται κυρίως από περιοχές γειτονικές των γραμμάτων του κειμένου. Αυτό δείχνει τη σημασία που έχει ο ακριβής προσδιορισμός της Inpainting περιοχής. Στο θέμα αυτό θα αναφερθούμε εκτενώς στο κεφάλαιο 4.

Τέλος, αξίζει να σημειώσουμε ότι εκτός από τον αλγόριθμο Efron-Leung, μπορούμε να εφαρμόσουμε έναν από τους πιο γρήγορους της ενότητας 3.2.3, παίρνοντας παρόμοια αποτελέσματα.



Σχήμα 3.10: Απομάκρυνση κειμένου με τον αλγόριθμο Efros-Leung



Σχήμα 3.11: Συμβολισμός των διανυσμάτων

3.3.2 Παραλλαγή του Αλγόριθμου Efros-Leung για Inpainting Υφής

Ο αλγόριθμος Efros-Leung δίνει, όπως είδαμε, πολύ καλά αποτελέσματα για Inpainting εικόνων με υφή. Όμως, ο συγκεκριμένος αλγόριθμος βγήκε πολύ πριν από την εισαγωγή του προβλήματος του Inpainting στον κλάδο της επεξεργασίας εικόνας και γενικά η χρήση του ήταν για σύνθεση υφής. Ως εκ τούτου, δε λαμβάνει υπόψη του κάποια χαρακτηριστικά σημεία του Inpainting προβλήματος και τον σπουδαίο ρόλο που παίζουν σε αυτό στοιχεία όπως οι ακμές, η καμπυλότητα κ.ο.κ.

Έναν τέτοιο αλγόριθμο πρότειναν το 2003 οι Criminisi, Perez, Toyama [1]. Ο αλγόριθμος που πρότειναν είναι μια παραλλαγή του αλγορίθμου Efros-Leung. Η ουσιαστική αλλαγή είναι η σειρά με την οποία συνθέτονται τα pixel.

Στον αλγόριθμο Efros-Leung τα pixel συνθέτονται ανάλογα με το πιο έχει τους πιο πολλούς γείτονές του συμπληρωμένους, έτσι ώστε να είμαστε πιο σίγουροι ότι η τιμή που θα αντιγραφεί θα είναι σωστή. Στον αλγόριθμο των Criminisi, Perez, Toyama η σειρά εξαρτάται όχι μόνο από τους συμπληρωμένους γείτονες κάθε pixel, αλλά και από την κατεύθυνση των ισοφώτεινων, (isophotes), γραμμών αλλά και της παραγωγού της εικόνας στο σημείο αυτό.

Έτσι σε κάθε σημείο p του ορίου της Inpainting περιοχής δίνεται μία τιμή προτεραιότητας, (priority) $P(p)$. Για την τιμή προτεραιότητας ισχύει ότι

$$P(p) = C(p) \cdot D(p) \quad (3.4)$$

$$C(p) = \frac{\sum_{q \in \Psi_p \cap \Omega} C(q)}{|\Psi_p|} \quad (3.5)$$

$$D(p) = \frac{|\nabla I_p^\perp \cdot n_p|}{\alpha} \quad (3.6)$$

όπου $|\Psi_p|$ είναι το εμβαδόν του τετραγωνικού παραθύρου, $C(p)$ είναι ο όρος εμπιστοσύνης, (confidence) που αρχικά είναι 0 για κάθε pixel μέσα στην In-



Αρχική Εικόνα

Inpainting με τον
αλγόριθμο των Efron-LeungInpainting
Criminisi-Perez-Toyama

Σχήμα 3.12: Το μοντέλο ικανοποιεί την αρχή της συνεκτικότητας (Από το [1]).

painting περιοχή και 1 για κάθε pixel εκτός αυτής⁵ και α είναι μια σταθερά κανονικοποίησης (π.χ. 255 για μια 8 bit grey-scale εικόνα). Επίσης από το σχήμα 3.11 βλέπουμε ότι το διάνυσμα n_p είναι το κάθετο διάνυσμα στο όριο της Inpainting περιοχής στο σημείο p και το $\nabla I p^{\perp}$ ⁶ είναι το διάνυσμα των ισοφώτεινων γραμμών στην εικόνα στο σημείο p . Ο όρος $D(p)$ εκφράζει τη δύναμη με την οποία οι ισοφώτεινες γραμμές φθάνουν στο όριο της Inpainting περιοχής και δίνεται προτεραιότητα στις περιοχές με έντονες μεταβολές.

Παρουσιάζουμε αποτελέσματα της συγκεκριμένης μεθόδου και τα συγκρίνουμε με τον αλγόριθμο των Efron-Leung.

Στην εικόνα 3.12 παρατηρούμε ότι με τον αλγόριθμο Efron-Leung η ακμή που αντιστοιχεί στον στύλο κόβεται από κάποιο σημείο και πέρα, ενώ με το μοντέλο των Criminisi, Perez, Toyama συνεχίζεται κανονικά. Μπορούμε να πούμε ότι το συγκεκριμένο μοντέλο ικανοποιεί την αρχή της συνεκτικότητας που αναφέραμε στο προηγούμενο κεφάλαιο και ο λόγος που ισχύει αυτό είναι ότι δίνεται προτεραιότητα στα pixel του στύλου να συνθεθούν επειδή υπάρχει μεγάλη μεταβολή γύρω από αυτά. Έτσι αυτά συνθέτονται κατά μήκος της ακμής και επομένως την συνεχίζουν. Αντίθετα με τον αλγόριθμο Efron-Leung η σειρά που εφαρμόζεται η σύνθεση προκύπτει με την onion-peel μέθοδο. Παρατηρούμε ότι στα σημεία του στύλου που η οριζόντια απόσταση τους από το όριο της

⁵Ο όρος εμπιστοσύνης $C(p)$ εκφράζει τη σειρά με την οποία συνθέτονται τα pixel με την κλασσική onion peel μέθοδο.

⁶Θυμίζουμε ότι: $\nabla^{\perp} = (-\partial_y, \partial_x)$



Αρχική Εικόνα



Περιοχή Inpainting (Μάσκα)



Αλγόριθμος Efros-Leung



Criminisi-Perez-Toyama

Σχήμα 3.13: Απομάκρυνση μεγάλου αντικειμένου από εικόνα

Inpainting περιοχής είναι μεγαλύτερη από την κατακόρυφη η ακμή συνεχίζεται γιατί η σύνθεση γίνεται κατά μήκος αυτής. Αντίθετα η ακμή αποκόπτεται εκεί που η κατακόρυφη απόσταση γίνεται μεγαλύτερη γιατί τότε η σύνθεση γίνεται κάθετα στην ακμή και τα μόνα γειτονικά pixel που υπάρχουν αντιστοιχούν σε θάλασσα.

Στην εικόνα 3.13 βλέπουμε πάλι ότι το μοντέλο των Criminisi, Perez, Toyama δίνει καλύτερα αποτελέσματα από τον αλγόριθμο των Efros-Leung. Βλέπουμε ότι το συγκεκριμένο μοντέλο (κάτω δεξιά) επιτρέπει την καλή συνέχεια των γραμμών της εικόνας ενώ διαχωρίζει τα διαφορετικά είδη της υφής (βουνό, βλάστηση, θάλασσα, παραλία) πολύ καλά μεταξύ τους. Αντίθετα με τον αλγόριθμο των Efros-Leung τα αποτελέσματα είναι ικανοποιητικά αλλά με κάποια λάθη (σταμάτημα της βλάστησης, μη αρμονική συνέχιση της ακτογραμμής).

Ο αλγόριθμος των Criminisi, Perez, Toyama δίνει τα καλύτερα αποτελέσματα από όλους τους αλγορίθμους για Inpainting υφής που εξετάσαμε μέχρι τώρα. Ενδιαφέρον θα είχε αν στον πίνακα της προτεραιότητας βάζαμε και άλλον έναν όρο που να σχετίζεται με την καμπυλότητα. Αυτό θα αποτελέσει αντικείμενο μελλοντικής έρευνας.

3.3.3 Σύγκριση Μεταξύ Inpainting Υφής και Inpainting Περιεχομένου

Τα πολύ καλά αποτελέσματα που δίνει το Inpainting υφής με τον αλγόριθμο των Efros-Leung αλλά κυρίως των Criminisi, Perez, Toyama ακόμα και σε εικόνες με πλούσιο περιεχόμενο, οδηγούν σε μία σύγκριση του Inpainting υφής με τους αλγόριθμους Inpainting περιεχομένου που εξετάσαμε στο κεφάλαιο 2.

Στην πραγματικότητα όμως, μία τέτοια σύγκριση είναι αδόκιμη καθώς αυτές οι δύο μεγάλες κατηγορίες Inpainting έχουν τελείως διαφορετικές μεθόδους αλλά και στόχους.

Το Inpainting περιεχομένου όπως είδαμε ανάλογα με το περιεχόμενο της εικόνας κοντά στην Inpainting περιοχή υπολογίζει νέα πληροφορία και την οποία μεταφέρει μέσα στην Inpainting περιοχή. Επίσης, η διαδικασία γίνεται μέσω ενός PDE επαναληπτικού σχήματος, και ο στόχος είναι αυτό το σχήμα να συγκλίνει σε μία σταθερή κατάσταση, που θα ικανοποιεί κάποιο κριτήριο, (ελαχιστοποίηση κάποιας συνάρτησης ενέργειας κ.λ.π.).

Αντίθετα το Inpainting υφής δεν δημιουργεί νέα πληροφορία αλλά κάθε pixel της Inpainting περιοχής προκύπτει από την αντιγραφή ενός pixel από την εικόνα δείγμα, που γίνεται με τρόπο τέτοιο ώστε να διατηρηθεί η υφή της εικόνας. Επίσης δεν τίθενται θέματα σύγκλισης, ύπαρξης και μοναδικότητας λύσης, αφού ο αλγόριθμος σταματάει όταν εφαρμοστεί πάνω σε όλα τα pixel της Inpainting περιοχής και λόγω του στοχαστικού χαρακτήρα που έχει τις περισσότερες φορές η υφή, περισσότερα του ενός τελικά αποτελέσματα μπορούν να είναι αποδεκτά από το ανθρώπινο μάτι (που είναι και το μοναδικό κριτήριο αξιολόγησης της τελικής εικόνας).

Αυτό που πρέπει να τονιστεί όμως ιδιαίτερα, είναι ότι το Inpainting υφής αντιγράφει την πληροφορία μέσα στην Inpainting περιοχή, πράγμα που σημαίνει χρειάζεται να υπάρχει μια περιοχή στην εικόνα δείγμα (ή στο υπόλοιπο μέρος της εικόνας που θέλουμε να φτιάξουμε) με παρόμοια υφή ώστε η πληροφορία να αντιγραφεί από εκεί με αποτελεσματικό τρόπο. Αυτό φαίνεται σχηματικά από την εικόνα 3.14.

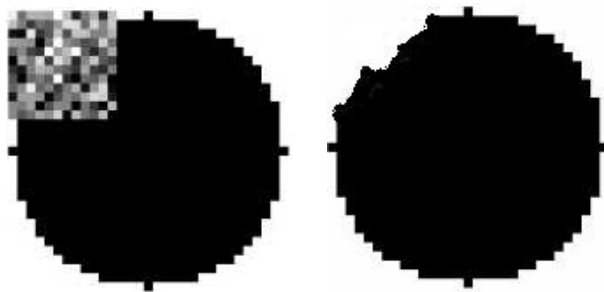
Σε περίπτωση όμως που δεν υπάρχει τέτοια περιοχή τότε το Inpainting υφής δεν δίνει καλά αποτελέσματα. Στην εικόνα 3.15 βλέπουμε ένα τέτοιο παράδειγμα. Το πάνω αριστερά κομμάτι που πρέπει να συμπληρωθεί, δεν έχει κάποιο ανάλογο κομμάτι ώστε να αντιγράψει πληροφορία από εκεί. Έτσι το αποτέλεσμα του Inpainting δεν είναι καθόλου ικανοποιητικό.

Το μοντέλο των Criminisi, Perez, Toyama μπορεί να δώσει καλύτερα αποτελέσματα αλλά και πάλι η έλλειψη της κατάλληλης εικόνας δείγματος δεν μας επιτρέπει να πάρουμε τα επιθυμητά αποτελέσματα (εικόνα 3.16)⁷.

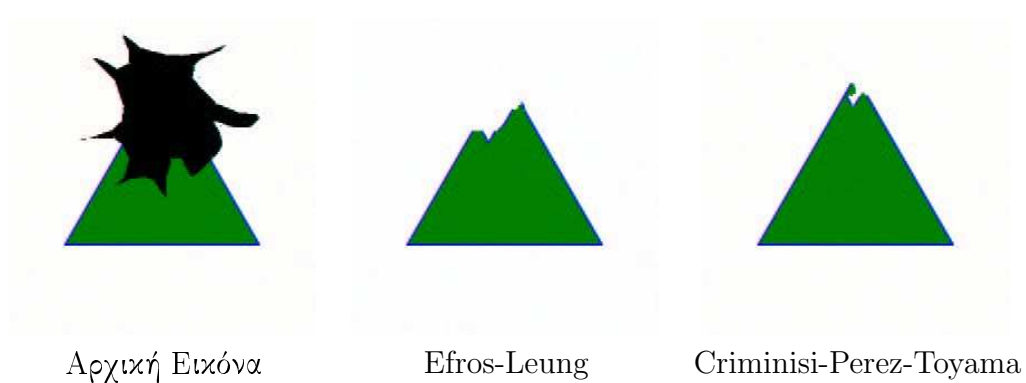
⁷Πεποίθησή μας είναι ότι η κορυφή του τριγώνου δεν μπορεί ποτέ να συμπληρωθεί σωστά



Σχήμα 3.14: Το Inpainting υφής χρειάζεται κατάλληλη περιοχή δείγμα για να αντιγράψει την κατάλληλη πληροφορία



Σχήμα 3.15: Η έλλειψη παρόμοιας περιοχής για αντιγραφή πληροφορίας δεν δίνει καλά αποτελέσματα. Αριστερά: αρχική εικόνα, Δεξιά: Inpainting με τον αλγόριθμο Efros-Leung.



Αρχική Εικόνα

Efros-Leung

Criminisi-Perez-Toyama

Σχήμα 3.16: Άλλο ένα παράδειγμα

Αυτό που πρέπει να σημειωθεί είναι ότι όπως η υφή και το περιεχόμενο συνυπάρχουν στις περισσότερες εικόνες, έτσι και οι δύο κατηγορίες αλγορίθμων θα μπορούσαν να συνδιαστούν για ακόμα καλύτερα αποτελέσματα. Προς τα εκεί άλλωστε οδηγούν και οι μέθοδοι που ακολουθούν οι συντηρητές τέχνης που αντιμετωπίζουν διαφορετικά υφή και περιεχόμενο. Αυτόν τον συνδιασμό των αλγορίθμων περιγράφουμε στην επόμενη ενότητα.

3.4 Ταυτόχρονο Inpainting Περιεχομένου και Υφής

Στα προηγούμενα κεφάλαια εξετάσαμε πολλές μεθόδους για το Inpainting πρόβλημα. Είδαμε μεθόδους για Inpainting περιεχομένου που βασίζονται σε PDE σχήματα, οι οποίες δίνουν καλά αποτελέσματα αλλά όπως είδαμε δεν αναπαράγουν την υφή μέσα στην Inpainting περιοχή D . Για Inpainting υφής εξετάσαμε διάφορους αλγόριθμους που χρησιμοποιούνται για σύνθεση υφής και οι οποίοι όπως είδαμε μπορούν να δώσουν ιδιαίτερα εντυπωσιακά αποτελέσματα σε αρκετά πολύπλοκες εικόνες. Πάντως οι αλγόριθμοι αυτοί δεν γεννούν πληροφορία, αλλά απλώς αντιγράφουν πληροφορία από άλλα κομμάτια της εικόνας και έτσι αν δεν υπάρχει η κατάλληλη περιοχή δειγματοληψίας, οι αλγόριθμοι αποτυγχάνουν να δώσουν αποδεκτά αποτελέσματα. Αυτό το πρόβλημα δεν υπάρχει με τα μοντέλα Inpainting περιεχομένου αφού σε αυτά η πληροφορία δεν αντιγράφεται αλλά κατασκευάζεται με βάση την πληροφορία κοντά στο όριο της Inpainting περιοχής.

Από τα παραπάνω φαίνεται, ότι όπως σε μια συνιτισμένη εικόνα συνυπάρχουν το περιεχόμενο και η υφή, έτσι και οι δύο κατηγορίες μεθόδων Inpainting αλληλοσυμπληρώνονται, και θα μπορούσαν να δώσουν καλύτερα αποτελέσματα σε περίπτωση που χρησιμοποιηθούν με ένα τρόπο που να τις συνδιάζει και τις 2. Έτσι το 2002 οι Bertalmio, Vese, Sapiro Osher [39] παρουσίασαν ένα μοντέλο για ταυτόχρονο Inpainting περιεχομένου και υφής (Simultaneous Structure and Texture Inpainting).

Η ιδέα του μοντέλου είναι πολύ απλή. Αν f η αρχική εικόνα τότε αυτή διασπάται σε 2 εικόνες u που αντιπροσωπεύει το περιεχόμενο της εικόνας και v που αναπαριστά την υφή της εικόνας (**$u + v$ decomposition**). Στη συνέχεια εκτελούμε κάποιον αλγόριθμο για Inpainting περιεχομένου (κεφ. 2) στην εικόνα περιεχομένου, ενώ στην εικόνα υφής εφαρμόζουμε κάποιον αλγόριθμο για Inpainting υφής. Στη συνέχεια οι 2 εικόνες που προκύπτουν επανενώνονται ώστε να προκύψει η ανακατασκευασμένη αρχική εικόνα.⁸

με έναν αλγόριθμο Inpainting υφής

⁸Στο [39] οι συγγραφείς χρησιμοποίησαν για Inpainting περιεχομένου το μοντέλο BCBS

Το μόνο καινούριο στοιχείο που πρέπει να εισάγουμε εδώ είναι πως γίνεται η διάσπαση της αρχικής εικόνας στις 2 αυτές συνιστώσες.

3.4.1 Διάσπαση της Εικόνας σε Συνιστώσες Περιεχομένου και Υφής

Γενικά στην επεξεργασία εικόνας, πολλές φορές θέλουμε από μια πραγματική εικόνα f να εξάγουμε την χρήσιμη πληροφορία που έχει. Το πρόβλημα αυτό θα μπορούσε να διατυπωθεί ως: *Δεδομένης μιας εικόνας f , να βρεθεί μια εικόνα u παρόμοια με την f , ώστε η u να αποτελεί την απλούστευσή της f .* Γενικά η μορφή της εικόνας u θα είναι μια εικόνα με ομογενείς, λείες περιοχές που διαχωρίζονται μεταξύ τους από έντονες και απότομες ακμές. Η συνιστώσα που απομένει v έτσι ώστε $f = u + v$ περιέχει κυρίως την υφή της εικόνας ή τον θόρυβο που ενδεχομένως υπήρχε, ενώ δε γίνεται κανένας διαχωρισμός μεταξύ υφή και θορύβου. Τέτοιες μέθοδοι διάσπασης (decomposition) της εικόνας σε 2 συνιστώσες, είναι η μέθοδος των Rudin, Osher, Fatemi [36], που θα αναφέρουμε παρακάτω, των Perona, Malik [50], όπου το περιεχόμενο u ανακτάται αν εφαρμόσουμε στην αρχική εικόνα την εξίσωση της ανισοτροπικής διάχυσης καθώς και άλλες πολλές, π.χ. των Mumford, Shah [47].

Το μοντέλο των Rudin-Osher-Fatemi

Το μοντέλο αυτό χρησιμοποιήθηκε για τον καθαρισμό της εικόνας από θόρυβο (denoising) και έτσι η συνιστώσα u που προέκυπτε ήταν η τελική εικόνα ενώ η συνιστώσα v απλώς αναπαριστούσε τον θόρυβο και δεν είχε καμία χρησιμότητα. Η μαθηματική έκφραση του μοντέλου, που θυμίζει το μοντέλο TV inpainting ήταν δεδομένης της αρχικής εικόνας f να βρεθεί η εικόνα περιεχομένου u ώστε να έχουμε

$$\inf_{u \in L^2} F(u) = \int_{\Omega} |\nabla u| dx dy + \lambda \int_{\Omega} |f - u|^2 dx dy \quad (3.7)$$

Ύπαρξη και μοναδικότητα της λύσης εξασφαλίζονται θεωρώντας ότι η εικόνα u ανήκει στο χώρο των εικόνων φραγμένης μεταβολής (bounded variation BV). Παίρνοντας την παράγωγο Euler-Lagrange προκύπτει η μεταβολική επαναληπτική μέθοδος για τον προσδιορισμό της συνιστώσας u

$$u^{n+1} = f + \frac{1}{2\lambda} \nabla \cdot \left(\frac{\nabla u^n}{|\nabla u^n|} \right) \quad (3.8)$$

με $u^0 = f$.

και για τη σύνθεση υφής τον πρώτο αλγόριθμο των Efros-Leung

Περισσότερες πληροφορίες για το παρών μοντέλο μπορούν να βρεθούν στο [36]

Θεώρηση της υφής κατά Meyer και επέκταση του ROF μοντέλου

Το μοντέλο ROF που περιγράψαμε δεν κάνει διαχωρισμό μεταξύ υφής και θορύβου. Αποδεικνύεται, μάλιστα ότι για μικρές τιμές της παραμέτρου λ η υφή αφαιρείται από την συνιστώσα u και έτσι τα αποτελέσματά του είναι αποδεκτά μόνο σε εικόνες που δεν έχουν θόρυβο.

Για να εξάγουμε από την αρχική εικόνα ταυτόχρονα και τη συνιστώσα u του περιεχομένου αλλά και την συνιστώσα v της υφής, ο Meyer [45] πρότεινε την μοντελοποίηση της υφής ως μια ταλαντευόμενη συνάρτηση (oscillating function) με μηδενική μέση τιμή ενώ παράλληλα πρότεινε ένα χώρο συναρτήσεων, δυαδικό ως προς τον Bounded Variation χώρο για τις συναρτήσεις της υφής.

Ορισμός 3.1. Έστω G ο χώρος Banach που αποτελείται από όλες τις συναρτήσεις $v(x, y)$ που μπορούν να γραφούν ως

$$v(x, y) = \partial_x g_1(x, y) + \partial_y g_2(x, y), \quad \mu\epsilon \quad g_1, g_2 \in L^\infty(\mathbb{R}^2), \quad (3.9)$$

και $\|v\|_*$ μία νόρμα στο χώρο αυτό, ορισμένη ως το κάτω φράγμα, για όλες τις δυνατές αναπαραστάσεις του v , των νόρμων L^∞ για τις συναρτήσεις $|\vec{g}|$ με $\vec{g} = (g_1, g_2)$, $|\vec{g}| = \sqrt{g_1^2 + g_2^2}$.

Με βάση αυτόν τον ορισμό ο Meyer πρότεινε το παρακάτω μοντέλο διαχωρισμού της εικόνας

$$\inf_u \left\{ E(u) = \int_{\Omega} |\nabla u| dx dy + \lambda \|u\|_*, \quad f = u + v \right\} \quad (3.10)$$

Για την αριθμητική υλοποίηση του μοντέλου του Meyer οι Vese, Osher χρησιμοποίησαν την προσέγγιση

$$\|\sqrt{g_1^2 + g_2^2}\|_{L^\infty} = \lim_{p \rightarrow \infty} \|\sqrt{g_1^2 + g_2^2}\|_{L^p} \quad (3.11)$$

και κατέληξαν στο παρακάτω πρόβλημα ελαχιστοποίησης

$$\inf_{g_1, g_2, u} \left\{ G_p(u, g_1, g_2) = \int_{\Omega} |\nabla u| + \lambda \int_{\Omega} |f - u - \partial_x g_1 - \partial_y g_2|^2 + \mu \left[\int_{\Omega} \left(\sqrt{g_1^2 + g_2^2} \right)^p \right]^{1/p} \right\} \quad (3.12)$$

Για $p \rightarrow \infty$ το μοντέλο προσεγγίζει αυτό του Meyer. Πάντως για διάφορες τιμές του p το μοντέλο δίνει ουσιαστικά τα ίδια αποτελέσματα οπότε θεωρούμε $p = 1$ για ταχύτητα στους υπολογισμούς αλλά και απλότητα στις πράξεις.

Για την ελαχιστοποίηση της (3.12), θεωρούμε σύμφωνα με τις αρχές του μεταβολικού λογισμού (Variational Calculus) τις Euler-Lagrange παραγώγους οπότε παίρνουμε το παρακάτω σύστημα (από το μηδενισμό των παραγώγων).

$$u = f - \partial_x g_1 - \partial_y g_2 + \frac{1}{2\lambda} \nabla \cdot \left(\frac{\nabla u}{|\nabla u|} \right) \quad (3.13)$$

$$\mu \frac{g_1}{\sqrt{g_1^2 + g_2^2}} = 2\lambda \left[\frac{\partial}{\partial x} (u - f) + \partial_{xx}^2 g_1 + \partial_{xy}^2 g_2 \right] \quad (3.14)$$

$$\mu \frac{g_2}{\sqrt{g_1^2 + g_2^2}} = 2\lambda \left[\frac{\partial}{\partial y} (u - f) + \partial_{xy}^2 g_1 + \partial_{yy}^2 g_2 \right] \quad (3.15)$$

με οριακές συνθήκες

$$r \frac{\nabla u}{|\nabla u|} (n_x, n_y) = 0 \quad (3.16)$$

$$(f - u - \partial_x g_1 - \partial_y g_2) n_x = 0 \quad (3.17)$$

$$(f - u - \partial_x g_1 - \partial_y g_2) n_y = 0 \quad (3.18)$$

Η αριθμητική επίλυση του παραπάνω συστήματος, καθώς και πολλά θεωρητικά στοιχεία και παραδείγματα, δίνεται αναλυτικά στο [63]⁹. Σημειώνουμε μόνο ότι οι αρχικές τιμές του επαναληπτικού αλγορίθμου είναι:

$$u^0 = f, g_1^0 = -\frac{1}{2\lambda} \frac{f_x}{|\nabla f|}, g_2^0 = -\frac{1}{2\lambda} \frac{f_y}{|\nabla f|}$$

Το επαναληπτικό σχήμα συγκλίνει μετά από t επαναλήψεις και οι συνιστώσες που παίρνουμε είναι $u = u^t$ και $v = \nabla \cdot (g_1^t, g_2^t)$.

3.4.2 Συσχέτιση του Inpainting Περιεχομένου και Υφής με τις Μεθόδους Συντήρησης Έργων Τέχνης

Πριν παρουσιάσουμε αποτελέσματα της μεθόδου ταυτόχρονου Inpainting περιεχομένου και υφής (Simultaneous Structure and Texture Inpainting), αξίζει να δούμε πως σχετίζεται το συγκεκριμένο μοντέλο με τις μεθόδους που χρησιμοποιούνται από ζωγράφους για την συντήρηση και αναπαλαίωση πινάκων ζωγραφικής.

Είδαμε στο πρώτο κεφάλαιο την διαδικασία που ακολουθείται για τη συντήρηση παλαιών πινάκων ζωγραφικής. Για ευκολία ξαναδείχνουμε την εικόνα 1.2 που εξηγεί σχηματικά αυτή τη διαδικασία.

⁹Στη συγκεκριμένη αναφορά, εξηγείται και ότι το παραπάνω μοντέλο αλλάζοντας τις τιμές των παραμέτρων λ και μ μπορεί να χρησιμοποιηθεί και για διαδικασίες κατάτμησης υφής (texture segmentation) και διαχωρισμού υφής (texture discrimination).



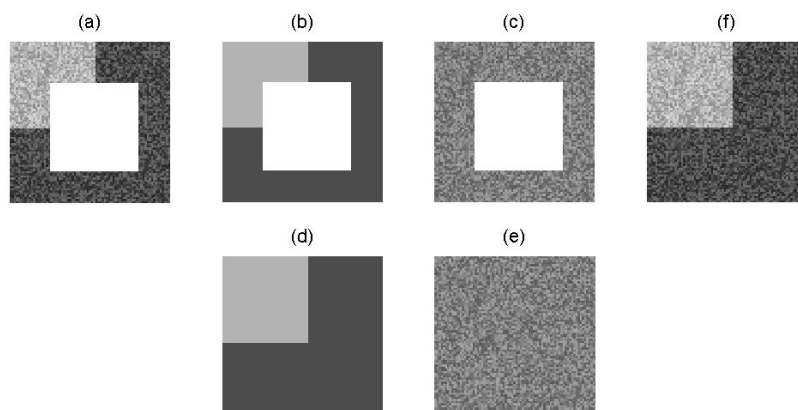
Σχήμα 3.17: Ο τρόπος Inpainting που ακολουθείται από τους συντηρητές έργων. Από αριστερά προς δεξιά: Αρχική εικόνα, συμπλήρωση ακμών, συμπλήρωση περιεχομένου, συμπλήρωση υφής.

Στο ταυτόχρονο Inpainting περιεχομένου και υφής η διαδικασία που ακολουθείται είναι η εξής:

1. **Διαχωρισμός περιεχομένου - υφής:** Διάσπαση της αρχικής εικόνας σε μία συνιστώσα περιεχομένου u , και σε μία συνιστώσα υφής v .
2. **Συμπλήρωση περιεχομένου:** Inpainting της συνιστώσας περιεχομένου με κάποιον από τους αλγόριθμους Inpainting π.χ. (TV inpainting, BCBS)
3. **Συμπλήρωση υφής:** Συμπλήρωση της Inpainting περιοχής της συνιστώσας v με κάποιον αλγόριθμο σύνθεσης υφής (π.χ. Efros-Leung)
4. **Επανένωση:** Πρόσθεση των δύο συμπληρωμένων συνιστωσών και απόκτηση της τελικής εικόνας.

Σχηματικά η μέθοδος μπορεί να παρουσιαστεί από τις εικόνες του σχήματος 3.18

Αυτό που πρέπει να σημειωθεί είναι ότι η τεχνική διαχωρισμού του περιεχομένου της εικόνας από την υφή της και στη συνέχεια η ξεχωριστή επεξεργασία τους προσεγγίζει κατά πολύ τον τρόπο με τον οποίο γίνεται η συντήρηση των έργων τέχνης. Το μόνο βήμα που δεν έχει που δεν έχει προσεγγισθεί ακόμα πλήρως είναι ότι η πληροφορία της όλης εικόνας καθορίζει και το πως θα συμπληρωθούν οι ζητούμενες περιοχές, καθώς κάτι τέτοιο γίνεται μόνο για τη συνιστώσα της υφής. Στην περίπτωση του περιεχομένου, όπου η πληροφορία αντλείται μόνο από τη γειτονία της Inpainting περιοχής, η έρευνα όπως προαναφέραμε βρίσκεται ακόμα σε αρχικό στάδιο αν και έχουν γίνει κάποια σημαντικά βήματα [55], [2].



Σχήμα 3.18: Σχηματική αναπαράσταση του ταυτόχρονου Inpainting περιεχομένου και υφής: (a): Αρχική εικόνα, (b),(c): Εικόνες περιεχομένου και υφής, (d),(e): Inpainting περιεχομένου και υφής, (f): Τελική εικόνα.

3.4.3 Ταυτόχρονο Inpainting Περιεχομένου και Υφής - Αποτελέσματα

Έχοντας αναλύσει το μοντέλο του ταυτόχρονου Inpainting περιεχομένου και υφής δίνουμε κάποια πειραματικά αποτελέσματα που ελήφθησαν με το συγκεκριμένο μοντέλο.

Στην εικόνα 3.19 βλέπουμε ένα παράδειγμα της συγκεκριμένης μεθόδου. Παρατηρούμε ότι ο διαχωρισμός του περιεχομένου και της υφής είναι πολύ καλός έτσι ώστε το Inpainting της κάθε συνιστώσας να είναι πολύ αποτελεσματικό, πράγμα που φαίνεται και από την τελική εικόνα¹⁰.

Στην εικόνα 3.20 δίνουμε ένα ακόμα παράδειγμα της μεθόδου για απομάκρυνση κειμένου σε μία έγχρωμη εικόνα. Η περιοχή Inpainting είναι αρκετά μεγάλη γεγονός που κάνει τον καλό διαχωρισμό περιεχομένου - υφής αναγκαίο για το σωστό Inpainting περιεχομένου. Από το παράδειγμα φαίνεται ότι το πρόβλημα του Inpainting σε έγχρωμες εικόνες είναι πολύ πιο δύσκολο καθώς κάποιο λάθος σε ένα από τα κανάλια μπορεί να δημιουργήσει ένα οδηγήσει σε λάθος χρώμα που φαίνεται έντονα (π.χ. πράσινο χρώμα). Επίσης από τα αποτελέσματα φαίνονται και οι αδυναμίες του αλγορίθμου σύνθεσης υφής καθώς τα γράμματα Y , D βρίσκονται πάνω σε ακμή και δεν υπάρχει αντίστοιχη περιοχή στην εικόνα υφής για να μεταφερθεί η πληροφορία.

Τέλος, στην εικόνα 3.21 βλέπουμε τα πλεονεκτήματα της συγκεκριμένης

¹⁰ Στα παραδείγματα αυτά, οι συνιστώσες υφής έχουν πολύ μικρές τιμές γύρω από το μηδέν. Για την απεικόνισή τους προσθέτουμε στις εικόνες μια γκριζα (dc), τιμή φωτεινότητας, ώστε η υφή να απεικονίζεται με τις μεταβολές της γκριζας επιφάνειας.



Σχήμα 3.19: Παράδειγμα Inpainting περιεχομένου και υφής: 1η γραμμή: Αρχική και τελική εικόνα, 2η γραμμή: Διαχωρισμός περιεχομένου και υφής, 3η γραμμή: Inpainting περιεχομένου και υφής. (Από το [39])



Αρχική Εικόνα



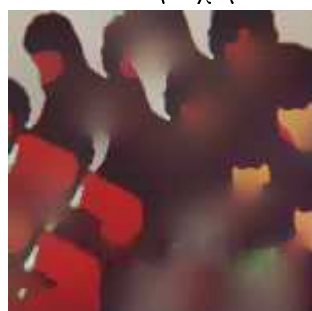
Εικόνα Περιεχομένου



Εικόνα Υφής



Περιοχή Inpainting



Inpainting Περιεχομένου



Inpainting Υφής



Τελική Εικόνα

Σχήμα 3.20: Παράδειγμα αφαίρεσης κειμένου από έγχρωμη εικόνα με Inpainting περιεχομένου και υφής



Σχήμα 3.21: Σύγκριση ταυτόχρονου Inpainting περιεχομένου και υφής με μεθόδους Inpainting μόνο υφής και περιεχομένου. Πάνω αριστερά: Αρχική Εικόνα. Πάνω δεξιά: Αποτέλεσμα ταυτόχρονου Inpainting περιεχομένου και υφής. Κάτω αριστερά: Αποτέλεσμα με Inpainting μόνο υφής. Αποτέλεσμα με Inpainting μόνο περιεχομένου. (Από το [39])

μεθόδου. Με το ταυτόχρονο Inpainting περιεχομένου και υφής η απομάκρυνση του αντικειμένου έχει γίνει πολύ αποτελεσματικά. Αντίθετα στην περίπτωση που έγινε μόνο Inpainting υφής στην αρχική εικόνα υπάρχουν κάποιες ασυνέχειες στην εικόνα (χέρι), ενώ με Inpainting μόνο περιεχομένου δεν προστίθεται υφή στην Inpainting περιοχή, πράγμα που φαίνεται από το θόλωμα που υπάρχει.

Κεφάλαιο 4

Αυτόματος Εντοπισμός και Απομάκρυνση Κειμένου από Εικόνες

Μέχρι τώρα στην εργασία, η Inpainting περιοχή επιλέγεται κάθε φορά από τον χρήστη. Η επιλογή αυτή φαίνεται λογική αφού η περιοχή που πρέπει να επιδιορθωθεί σε κάποια εικόνα, ειδικά όταν ο σκοπός είναι η απομάκρυνση κάποιου αντικειμένου, είναι κάτι το υποκειμενικό και δεν υπάρχει κάποιος γενικός κανόνας.

Όμως ορισμένες κατηγορίες λαθών και αντικειμένων μπορούν να τυποποιηθούν και να κατηγορηποιηθούν με βάση κάποια κοινά χαρακτηριστικά ως προς τη δομή τους. Παραδείγματα τέτοιων χαρακτηριστικών είναι το κείμενο που υπάρχει σε κάποιες εικόνες, οι ρωγμές σε φωτογραφίες και τα χαμένα πακέτα κατά την μετάδοση ασύρματων εικόνων. Για αυτές τις κατηγορίες αντικειμένων έχει ενδιαφέρον να ερευνηθεί το κατά πόσον μπορούν να εντοπισθούν αυτόματα σε μία εικόνα.

Πέρα από το ερευνητικό ενδιαφέρον, μια τέτοια αυτόματη ανίχνευση είναι χρήσιμη γιατί μπορεί να απαλλάξει τον χρήστη από την διαδικασία επιλογής που μπορεί να είναι επίπονη (π.χ. σε εικόνες που έχουν πολύ κείμενο πάνω τους), ενώ παράλληλα η ανίχνευση πρέπει να γίνει με μεγάλη λεπτομέρεια, η οποία για να επιτευχθεί από τον χρήστη χρειάζεται κάποιο εξειδικευμένο λογισμικό.

Η επιλογή της Inpainting περιοχής με λεπτομέρεια είναι πολύ σημαντική, όπως φαίνεται και από το παρακάτω παράδειγμα.

Στην εικόνα 4.1 παρατηρούμε ότι υπάρχει κείμενο σε όλη την επιφάνεια της εικόνας. Ο προσδιορισμός της θέσεως του κειμένου είναι πολύ δύσκολο να γίνει από το χρήστη λόγω του μεγάλου αριθμού των γραμμάτων (ή λέξεων) που πρέπει να επιλεγθούν μία μία, αλλά και της λεπτομέρειας στο σχήμα του κάθε γράμματος που πρέπει να διατηρηθεί. Αν ο χρήστης κάνει μία πρόχειρη επιλογή



Σχήμα 4.1: Εικόνα με δύσκολο προσδιορισμό της Inpainting περιοχής από τον χρήστη

της Inpainting περιοχής (ορθογώνιες περιοχές που περιέχουν κάθε γραμμή) τότε τα κυριότερα μειονεκτήματα που προκύπτουν είναι ότι η Inpainting είναι μεγαλύτερη από ότι χρειάζεται, γεγονός που επιβαρύνει υπολογιστικά την εφαρμογή και καθιστά μεθόδους όπως το Inpainting περιεχομένου μη εφαρμόσιμες, αλλά κυρίως ότι στην Inpainting περιοχή περιλαμβάνονται οι γειτονικές περιοχές του κειμένου που περιέχουν πλούσια πληροφορία για το πως πρέπει να γίνει το Inpainting.

Επομένως γίνεται αντιληπτό ότι υπάρχουν και άλλοι λόγοι που οδηγούν σε μία προσπάθεια για αυτοματοποίηση της εύρεσης της Inpainting περιοχής, πέρα από τα γενικά πλεονεκτήματα της έννοιας της αυτοματοποίησης. Στο κεφάλαιο αυτό εξετάζουμε το πρόβλημα της αυτόματης εύρεσης κειμένου σε εικόνες και απομάκρυνσής του με κάποια από τις Inpainting τεχνικές. Παρόλο που και τα δύο προβλήματα υπάρχουν ανεξάρτητα το ένα από το άλλο ο συνδιασμός τους, που παρουσιάζει και κάποια ιδιαίτερα χαρακτηριστικά όπως θα παρουσιάσουμε, δεν υπάρχει, από όσο ξέρουμε τουλάχιστον, στη βιβλιογραφία.

4.1 Αυτόματη Ανίχνευση Κειμένου σε Εικόνες

Το πρόβλημα της αυτόματης ανίχνευσης κειμένου σε εικόνες δεν είναι ένα καινούριο πρόβλημα στην επεξεργασία εικόνας. Αντιθέτως υπάρχει πολλά χρόνια, καθώς συνδέεται άμεσα με ένα άλλο σημαντικό πρόβλημα, αυτό της αναγνώρισης χαρακτήρων. Από την εποχή που άρχισαν να βρίσκουν εφαρμογή τα συστήματα οπτικής αναγνώρισης χαρακτήρων, (*Optical Character Recognition - OCR*), άρχισε μια προσπάθεια αποθήκευσης του τεράστιου όγκου πληρο-

φορίας μέχρι εκείνη την εποχή, σε ψηφιακή μορφή με σκοπό την δημιουργία των πιο εύχρηστων βάσεων δεδομένων και ψηφιακών βιβλιοθηκών. Τα συστήματα OCR παίρνουν μία σελίδα κειμένου ως εικόνα και στην συνέχεια με μεθόδους της αναγνώρισης προτύπων, (pattern recognition), αναγνωρίζουν το κάθε γράμμα της εικόνας, μετατρέποντάς το σε έναν ASCII χαρακτήρα. Τα συστήματα OCR δίνουν πολύ καλά αποτελέσματα αλλά υπόκεινται σε δύο σημαντικούς περιορισμούς.

1. Βρίσκουν κείμενο μόνο όταν αυτό είναι τυπωμένο σε καθαρό και ομοιογενές background.
2. Αντιμετωπίζουν όλη την σελίδα που έχουν προς αναγνώριση ως κείμενο, δηλαδή δεν κάνουν διάκριση μεταξύ αντικειμένων κειμένου και αντικειμένων μη κειμένου σε μία εικόνα.

Επομένως πριν δωθεί μια εικόνα σε ένα σύστημα OCR για αναγνώριση του κειμένου που υπάρχει σε αυτήν, η εικόνα πρέπει πρώτα να επεξεργαστεί έτσι ώστε αρχικά να εντοπισθούν οι περιοχές της εικόνας που περιέχουν κείμενο και το κείμενο να απομονωθεί από το background που το περιβάλλει. Η επεξεργασία αυτή γίνεται με το πρόβλημα της ανίχνευσης κειμένου σε εικόνες, το οποίο υπάγεται στον κλάδο της *ανάλυσης κειμένου* (document analysis).

Εκτός, από την παραπάνω εφαρμογή, το πρόβλημα της αυτόματης ανίχνευσης κειμένου σε εικόνες βρισκεί εφαρμογή και σε ορισμένες περιπτώσεις περίληψης video. Για παράδειγμα από το video ενός αγώνα ποδοσφαίρου μπορούμε να βρούμε τις στιγμές όπου έγιναν αλλαγές στο σκόρ, ανιχνεύοντας και αναγνωρίζοντας το σκόρ όταν εμφανίζεται στην εικόνα.

Υπάρχουν πολλοί τρόποι προσέγγισης του προβλήματος την ανίχνευσης κειμένου σε εικόνες ή video. Η αναφορά σε όλους αυτούς τους τρόπους είναι πολύ δύσκολη καθώς η βιβλιογραφία είναι τεράστια. Πάντως θα επιχειρήσουμε μία σύντομη αναφορά στις βασικές κατηγορίες μεθόδων και σε αυτές που θα φανούν χρήσιμες στις μεθόδους που θα παρουσιάσουμε. Πρέπει να σημειωθεί, ότι η ποικιλία των εικόνων και των γραμμάτων όσον αφορά διάφορα χαρακτηριστικά του προβλήματος (background, μέγεθος, γραμματοσειρά και κατεύθυνση γραμμάτων, ανάλυση εικόνας) καθιστούν το πρόβλημα της ανίχνευσης χαρακτήρων επίκαιρο αφού τελείως γενικές λύσεις δεν έχουν βρεθεί ακόμα.

4.1.1 Συνοπτική Ανασκόπηση Μεθόδων

Όπως προαναφέρθηκε, το πρόβλημα της εύρεσης κειμένου σε εικόνες, ξεκίνησε μαζί με την ανάπτυξη των συστημάτων οπτικής αναγνώρισης χαρακτήρων OCR. Τα πρώτα κείμενα που έπρεπε να αναγνωριστούν από τα συστήματα OCR

και να αποθηκευτούν σε ψηφιακή μορφή ήταν κείμενα από βιβλία, εγκυκλοπαίδειες, εφημερίδες κ.λ.π. Τα κύρια χαρακτηριστικά τέτοιων κειμένων, είναι ότι το κείμενο είναι συνήθως μαύρο με άσπρο background και ότι τα γράμματα έχουν το ίδιο μέγεθος (εκτός από τις επικεφαλίδες) και διατάσσονται οριζόντια. Το πρόβλημα ήταν ο διαχωρισμός των κομματιών κειμένου από τις εικόνες που είχε η κάθε σελίδα. Τα συγκεκριμένα και απλά χαρακτηριστικά του κειμένου σε αυτήν την περίπτωση καθιστούσαν το πρόβλημα πολύ απλό. Ο Bloomberg [8] χρησιμοποιώντας απλούς μορφολογικούς τελεστές open, close, rank filters σε πολλαπλές αναλύσεις της εικόνας, έδειξε ότι σε τέτοιες απλές δυαδικές εικόνες το κείμενο μπορεί να απομονωθεί πολύ εύκολα από την υπόλοιπη εικόνα και στη συνέχεια να αναγνωριστεί με ακρίβεια. Επίσης με μορφολογικούς τελεστές *Hit-Miss operators*, *Thickening*, *Thinning*¹ έδειξε πως με πολύ απλό τρόπο μπορούν να εξαχθούν από ένα κείμενο οι λέξεις που είναι γραμμένες με *italics* ή **bold** αντίστοιχα [9], πράγμα που είναι χρήσιμο για την αυτόματη ανίχνευση των λέξεων κλειδιά σε ένα κείμενο.

Οι ιδέες αυτές με τη μορφολογική επεξεργασία ήταν πολύ απλές και επιτυχημένες και χρησιμοποιήθηκαν ευρύτατα. Για το σκοπό αυτό μάλιστα δημιουργήθηκαν και σχετικό hardware το οποίο ήταν πολύ γρήγορο λόγω του δυαδικού χαρακτήρα των δεδομένων [8]. Το μειονέκτημα όμως ήταν να εφαρμοστούν μόνο σε δυαδικές εικόνες. Για εικόνες greyscale, όπου τα γράμματα βρίσκονται σε ανομοιογενές background έπρεπε να ακολουθηθεί διαφορετική διαδικασία. Για τη λύση του προβλήματος σε αυτήν την περίπτωση έπρεπε να ληφθούν υπόψη διάφορα χαρακτηριστικά που έχουν οι γραμματικοί χαρακτήρες. Ένα χαρακτηριστικό που χρησιμοποιείται για την εύρεση των γραμμάτων σε μία εικόνα είναι ότι αυτά συνήθως αντιπροσωπεύουν τις έντονες ακμές μίας εικόνας. Έτσι μία απλή, αλλά όχι πολύ αποτελεσματική μέθοδος είναι να βρεθεί πρώτα η παράγωγος της εικόνας και στη συνέχεια με κάποια ειδική κατωφλίωση *thresholding* να απομονωθούν οι ακμές που αντιστοιχούν σε χαρακτήρες[49]. Γενικά οι κατηγορίες των μεθόδων για εύρεση χαρακτήρων σε μία εικόνα μπορούν να χωρισθούν στις παρακάτω κατηγορίες:

- **Top-Down τεχνικές:** Οι Top-Down τεχνικές επεξεργάζονται αρχικά ολόκληρη την εικόνα και την χωρίζουν σε περιοχές που υπάρχει κείμενο. Χρειάζονται μία a-priori γνώση της δομής του κειμένου (μέγεθος γραμμάτων, οριζόντια ή κάθετη διάταξη). Ο χωρισμός γίνεται με διάφορους τρόπους όπως με ανάλυση ιστογράμματος (για την εύρεση της στάθμης των γραμμάτων), με οριζόντια ή κατακόρυφη προβολή (για την εύρεση γραμμών κειμένου) ή με χωρικούς μετασχηματισμούς (π.χ. Fourier για

¹Στοιχεία για τους τελεστές Hit-Miss operators καθώς και για τις εφαρμογές τους σε ανάλυση εικόνας κειμένου document image υπάρχουν στο [10]

την εκμετάλευση περιοδικοτήτων που εμφανίζουν συνεχόμενοι χαρακτήρες).

- **Bottom-Up τεχνικές:** Οι τεχνικές αυτές επεξεργάζονται αρχικά πολύ μικρά τμήματα της εικόνας (μεμονομένα pixels ή μικρά μπλοκ) και στην συνέχεια ομαδοποιούν τέτοιες μονάδες για τη δημιουργία των περιοχών όπου υπάρχει κείμενο. Στην κατηγορία αυτή των μεθόδων ανήκουν και οι τεχνικές της μαθηματικής μορφολογίας, όπου θα αναφερθούμε αναλυτικότερα παρακάτω.
- **Τεχνικές κατάτμησης υφής (Texture Segmentation):** Με τις τεχνικές αυτές, γίνεται η υπόθεση ότι οι χαρακτήρες παρουσιάζουν διαφορετική υφή από την υπόλοιπη εικόνα. Σε μία τυπική σελίδα κάποιου βιβλίου, οι χαρακτήρες εμφανίζουν περιοδική υφή και στην οριζόντια κατεύθυνση (τα γράμματα και οι λέξεις μίας γραμμής ενός κειμένου) και στην κατακόρυφη (οι γραμμές ενός κειμένου), οπότε με κάποιον κατάλληλο τρόπο κατάτμησης υφής μπορεί να γίνει ο επιθυμητός διαχωρισμός. Μία συνοπτική παρουσίαση τέτοιων μεθόδων γίνεται στο [48].
- **Τεχνικές της αναγνώρισης προτύπων:** Οι γραμματικοί χαρακτήρες παρά την ποικιλία που μπορούν να έχουν, σε μέγεθος, γραμματοσειρά, έχουν μια συγκεκριμένη. Τη δομή αυτή εκμεταλεύεται και ο κλάδος της αναγνώρισης προτύπων για την αναγνώρισή τους. Τέτοιες μέθοδοι μπορούν να εφαρμοστούν και για τον εντοπισμό της θέσης τους με κάποια εκπαιδευμένη μηχανή π.χ. Support Vector Machines [33].²

Στη συνέχεια παρουσιάζονται αναλυτικά μία μορφολογική μέθοδος που χρησιμοποιήσαμε για την εύρεση κειμένου σε εικόνες.

4.2 Εύρεση Κειμένου σε Εικόνες με Μεθόδους Μαθηματικής Μορφολογίας

4.2.1 Στοιχεία Μαθηματικής Μορφολογίας

Η μαθηματική μορφολογία [54] είναι μία μη γραμμική μεθοδολογία για την επεξεργασία εικόνας, που βασίζεται στην θεωρία συνόλων (set theory) και στην άλγεβρα πλεγμάτων (lattice algebra) και στοχεύει στην ποσοτική περιγραφή της γεωμετρικής δομής των σχημάτων σε μία εικόνα. Η πρώτη της χρήση ήταν στην δεκαετία του 60 για την ανάλυση δυαδικών εικόνων και για εφαρμογές της

²Περισσότερα για τα χαρακτηριστικά των γραμμάτων καθώς και για τη διαδικασία μάθησης μηχανών μπορούν να βρεθούν στο [32].

αναγνώρισης προτύπων σε δυαδικά σήματα. Στη συνέχεια επεκτάθηκε και στην επεξεργασία greyscale εικόνων και γενικότερα σε μη γραμμικές εφαρμογές της ψηφιακής επεξεργασίας σήματος. Τα βασικά της πλεονεκτήματα είναι το πλούσιο θεωρητικό υπόβαθρο που υπάρχει, η απλότητα των ιδεών της και η εύκολη υλοποίησή τους και η καταλληλότητά της για προβλήματα που σχετίζονται με το σχήμα των αντικειμένων. Αυτά τα πλεονεκτήματα την έχουν καθιερώσει σαν έναν από τους βασικότερους κλάδους της όρασης υπολογιστών, τη μορφολογική επεξεργασία εικόνας (morphological image processing), με σημαντικές εφαρμογές όπως η ανάλυση σχήματος, μεγέθους και άλλων γεωμετρικών χαρακτηριστικών ενός αντικειμένου, μη γραμμικό φιλτράρισμα, ανάλυση σε πολλαπλές κλίμακες, κατάτμηση εικόνας και αναγνώριση αντικειμένων.

Οι δύο βασικότεροι μορφολογικοί τελεστές είναι αυτοί της διαστολής (*dilation*), και της συστολής (*erosion*).

Σε περίπτωση δυαδικών εικόνων X , ο τελεστής της διαστολής αντιστοιχεί στην Minkowski πρόσθεση του συνόλου X με ένα σύνολο (παράθυρο) B

$$X \oplus B \equiv \{x + y : x \in X, y \in B\} = \bigcup_{y \in B} X_{+y} \quad (4.1)$$

όπου $X_{+y} \equiv \{x + y : x \in X\}$ η μετάθεση του συνόλου X κατά το διάνυσμα y .

Ανάλογα ο τελεστής της συστολής αντιστοιχεί στην Minkowski αφαίρεση του συνόλου X με το συμμετρικό του συνόλου (παράθυρο) B , $B^r \equiv \{x : -x \in B\}$

$$X \ominus B \equiv \{x : B_{+x} \subseteq X\} = \bigcap_{y \in B} X_{-y} \quad (4.2)$$

Οι τελεστές dilation και erosion σε περίπτωση greyscale εικόνων (ή γενικά μη δυαδικών σημάτων $f(x)$) γενικεύονται ως

$$(f \oplus B)(x) \equiv \bigvee_{y \in B} f(x - y) \quad (4.3)$$

$$(f \ominus B)(x) \equiv \bigwedge_{y \in B} f(x + y) \quad (4.4)$$

όπου \bigvee , και \bigwedge συμβολίζουν το supremum και infimum.

Από τους μορφολογικούς τελεστές της διαστολής και της συστολής προκύπτουν και οι τελεστές opening \circ , και closing \bullet , σύμφωνα με τις παρακάτω εξισώσεις

$$X \circ B = (X \ominus B) \oplus B \quad (4.5)$$

$$X \bullet B = (X \oplus B) \ominus B \quad (4.6)$$

Τέλος για λόγους πληρότητας δίνονται και οι ορισμοί για δύο ακόμα μορφολογικούς τελεστές που έχουν αναφερθεί στην παρούσα εργασία, τον τελεστή rank και το Hit-Miss φίλτρο.

Για ένα σύνολο $X \subseteq \mathbb{Z}^m$ και ένα πεπερασμένο παράθυρο W με n σημεία, ($|W| = n$), το k -th rank σύνολο ορίζεται ως

$$X \square_k W \equiv \{y \in \mathbb{Z}^m : |X \cap (W^r)_{+y}| \geq k\} \quad (4.7)$$

Το φίλτρο Hit-Miss ωρίζεται ως

$$X \otimes (A, B) \equiv (X \ominus A) \cup (X^c \ominus B) \quad (4.8)$$

όπου X^c το συμπλήρωμα του συνόλου X , ενώ το φίλτρο έχει νόημα μόνο όταν τα σύνολα A, B είναι ξένα μεταξύ τους. Πολλά περισσότερα στοιχεία για τη μορφολογική επεξεργασία εικόνας μπορούν να βρεθούν σε πολλά βιβλία όρασης υπολογιστών π.χ. [41],[54].

4.2.2 Ανίχνευση Κειμένου

Η μέθοδος που παρουσιάζουμε προτάθηκε το 2000 από τους Hasan, Karam [24]. Το μοντέλο επεξεργάζεται greyscale εικόνες με στάθμες $[0, 1, \dots, 255]$. Σε περιπτώσεις εγχρωμών εικόνων παίρνουμε τη συνιστώσα της φωτεινότητας³. Αρχικά γίνεται η υπόθεση ότι οι χαρακτήρες έχουν τιμές φωτεινότητας μεγαλύτερες τουλάχιστον κατά 30 από το background, δηλαδή αποτελούν ακμές της εικόνας. Αυτή η υπόθεση εισάγει περιορισμούς στο μοντέλο όπως θα φανεί και στην συνέχεια.

Δεδομένου ότι τα γράμματα αποτελούν κάποιες από τις πιο ισχυρές ακμές της εικόνας, αρχικά επιχειρείται καταπίεση των δευτερευόντων ακμών της εικόνας. Αυτό επιτυγχάνεται με το θόλωμα της εικόνας σύμφωνα με το παρακάτω φιλτράρισμα

$$Y_{bl} = \frac{(Y \circ B) \bullet B + (Y \bullet B) \circ B}{2} \quad (4.9)$$

³Η συνιστώσα φωτεινότητας λαμβάνεται από το χρωματικό μοντέλο RGB σύμφωνα με τον τύπο $Y = 0,299R + 0,587G + 0,114B$.

όπου Y η εικόνα φωτεινότητας που έχουμε, ενώ B κάποιο συμμετρικό μορφολογικό στοιχείο π.χ. $B = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$.

Στην συνέχεια θεωρούμε την μορφολογική παράγωγο (morphological gradient), της θολωμένης εικόνας Y_{bl}

$$e_s = Y_{bl} \oplus B - Y_{bl} \ominus B \quad (4.10)$$

όπου B είναι όπως πριν.

Η εικόνα της παραγώγου που παίρνουμε, κατωφλιοποιείται έτσι ώστε να καταπιεσθούν περαιτέρω οι δευτερεύουσες ακμές και να προκύψει μία δυαδική εικόνα με τις κύριες ακμές. Το επίπεδο του κατωφλίου γ δίνεται από τον τύπο⁴

$$\gamma = 1.5 \cdot \frac{\sum (e_s \cdot s)}{\sum s} \quad (4.11)$$

όπου ο πολλαπλασιασμός στον αριθμητή γίνεται pixel προς pixel ενώ η εικόνα s δίνεται από τον τύπο

$$s = \max(|g_1 * e_s|, |g_2 * e_s|) \quad (4.12)$$

όπου $g_1 = [-1, 0, 1]$ και $g_2 = [-1, 0, 1]^T$ είναι δύο πολλοί απλοί τελεστές πρώτης παραγώγου, και $*$ συμβολίζει τη γραμμική συνέλιξη.

Έτσι για τη δυαδική εικόνα ακμών e έχουμε

$$e = \begin{cases} 1, & \text{αν } e_s > \gamma \\ 0, & \text{αλλιώς} \end{cases} \quad (4.13)$$

Η δυαδική εικόνα e , φαίνεται ξεκάθαρα από τον τύπο της ότι παίρνει την τιμή 1 στα σημεία όπου βρίσκονται πάνω στις ακμές της θολωμένης εικόνας Y_{bl} , όπως αυτές προκύπτουν από τη μορφολογική παράγωγο.

Εφόσον έχουμε υποθέσει ότι τα γράμματα αντιστοιχούν στις κύριες ακμές της εικόνας, περιμένουμε ότι στην εικόνα e , υπάρχει το περίγραμμά τους. Για τα περιγράμματα των γραμμάτων μπορούμε να παρατηρήσουμε τα ακόλουθα:

1. Το περίγραμμα του κάθε γράμματος αναμένουμε να είναι μία κλειστή καμπύλη
2. Τα περιγράμματα των γραμμάτων βρίσκονται κοντά μεταξύ τους, ειδικά των γραμμάτων που ανήκουν στην ίδια λέξη

⁴Στο [24] για το κατώφλι δεν γίνεται ο πολλαπλασιασμός με 1.5. Στη δική μας υλοποίηση όμως με αυτό το κατώφλι, ελήφθησαν καλύτερα αποτελέσματα.

3. Τα περιγράμματα των γραμμάτων έχουν παρόμοιο μέγεθος, εφόσον υποθέσουμε ότι τα γράμματα έχουν το ίδιο μέγεθος

Από τα παραπάνω, συμπεραίνεται ότι τα γράμματα μπορούν να ομαδοποιηθούν μεταξύ τους είτε σε λέξεις είτε σε ακόμα μεγαλύτερες μονάδες όπως π.χ. γραμμές. Η ομαδοποίηση αυτή γίνεται με το μορφολογικό closing της δυαδικής εικόνας με κάποιο μορφολογικό στοιχείο. Με τη διαστολή (dilation), που είναι το πρώτο μέρος του closing, τα γράμματα που βρίσκονται το ένα δίπλα στο άλλο ομαδοποιούνται και έτσι με τη συστολή, erosion, ως το δεύτερο μέρος του closing, οι ομάδες γραμμάτων που έχουν δημιουργηθεί έχουν μεγάλη επιφάνεια και αποτελούν κλειστά σύνολα οπότε δεν αφαιρούνται. Ιδιαίτερη σημασία παίζει εδώ το μορφολογικό στοιχείο που επιλέγεται. Το μέγεθος του (εμβადόν) εκφράζει και το μέγεθος των ομάδων που θέλουμε να δημιουργήσουμε ενώ για το σχήμα του και η σχέση των διαστάσεων του χρειάζεται εκ των προτέρων γνώση για τη διάταξη των γραμμάτων. Για παράδειγμα, αν τα γράμματα είναι διατεταγμένα οριζόντια το μορφολογικό στοιχείο θα πρέπει να είναι ορθογώνιο με αναλογία οριζόντιας προς κατακόρυφης διάστασης περίπου 2.5 : 1. Αυτά θα γίνουν ξεκάθαρα από τα αποτελέσματα που θα παρουσιαστούν παρακάτω.

Από την εφαρμογή του μορφολογικού closing, προκύπτει μία νέα εικόνα e_c που περιέχει διάφορες περιοχές οι οποίες μπορούν να ονομαστούν ως πιθανές περιοχές κειμένου. Η κάθε τέτοια περιοχή πρέπει να επεξεργαστεί διαφορετικά. Ο διαχωρισμός των περιοχών μεταξύ τους, γίνεται βρίσκοντας τις συνεκτικές συνιστώσες, (connected components), της (δυαδικής) εικόνας e_c και χρησιμοποιούμε γειτονιά των 8 σημείων. Ο διαχωρισμός γίνεται δίνοντας στα pixel της κάθε συνιστώσας την ίδια τιμή και σε pixel διαφορετικών συνιστωσών, διαφορετικές τιμές. Ο ψευδοκώδικας του συγκεκριμένου αλγορίθμου δίνεται στο [24].

Η επεξεργασία της κάθε συνεκτικής συνιστώσας της εικόνας e_c συντελείται με μία ξεχωριστή κατωφλίωση της κάθε συνιστώσας παρόμοια με την πρώτη κατωφλίωση της εξίσωσης (4.13).

Έστω C_1, C_2, \dots, C_n οι συνεκτικές συνιστώσες της e_c ως δυαδικές εικόνες της ίδιας διάστασης και Y_1, Y_2, \dots, Y_n οι αντίστοιχες περιοχές της αρχικής εικόνας Y , δηλαδή $Y_i = Y \cdot C_i$.

Τότε με την κατωφλίωση προκύπτουν οι n δυαδικές εικόνες e_1, e_2, \dots, e_n με

$$e_i = \begin{cases} 1, & \text{αν } Y_i > \gamma_i \\ 0, & \text{αλλιώς} \end{cases} \quad (4.14)$$

με

$$\gamma_i = 1.5 \cdot \frac{\sum(Y_i \cdot s_i)}{\sum s_i}, \quad (4.15)$$

$$s_i = \max(|g_1 * Y_i|, |g_2 * Y_i|) \quad (4.16)$$

και $g_1 = [-1, 0, 1]$ και $g_2 = [-1, 0, 1]^T$ όπως πριν.

Από το άθροισμα (ένωση) αυτών των δυαδικών εικόνων προκύπτει μία νέα δυαδική εικόνα e

$$e = e_1 + e_2 + \dots + e_n \quad (4.17)$$

Στην εικόνα e τώρα υπάρχουν οι χαρακτήρες, διαχωρισμένοι μεταξύ τους, αλλά πιθανότατα υπάρχουν και άλλες συνιστώσες που δεν αντιστοιχούν σε χαρακτήρες. Για την εξάκλειψη αυτών των συνιστωσών προτείνεται στο [24] το εξής: Αν A, W, L το εμβαδόν, πλάτος και ύψος της κάθε συνιστώσας τότε μια συνιστώσα μηδενίζεται αν ισχύει μία εκ των συνθηκών:

$$\frac{A}{W \times L} < \frac{1}{4.5}, \quad \min\left(\frac{W}{L}, \frac{L}{W}\right) < 0.2$$

Αποτελέσματα της μεθόδου

Η παραπάνω μέθοδος γίνεται πιο ξεκάθαρη από τα αποτελέσματα που ακολουθούν.

Στην εικόνα 4.2 βλέπουμε ένα παράδειγμα της συγκεκριμένης μεθόδου. Οι γραμματικοί χαρακτήρες, φαίνεται ξεκάθαρα ότι έχουν τιμές φωτεινότητας αρκετά μεγαλύτερες από το background τους και έτσι παρά το θόλωμά τους, το περίβλημά τους παίρνει από το κατώφλι της παραγωγού. Από το closing με το μορφολογικό στοιχείο, το οποίο ήταν ένας 8×16 , παίρνουμε τέσσερις διαφορετικές περιοχές (συνεκτικές συνιστώσες τις οποίες χρωματίζουμε και με διαφορετικές αποχρώσεις του γκρι). Οι συνιστώσες αυτές φιλτράρονται, με διαφορετικό όμως κατώφλι η κάθε μία και έτσι προκύπτει η τελική εικόνα με το κείμενο. Παρατηρούμε ότι οι γραμματικοί χαρακτήρες έχουν εντοπιστεί πολύ καλά, αλλά έχουν εντοπιστεί και κάποιες άλλες περιοχές οι οποίες δεν εξαλήφθησαν από το μορφολογικό closing. Αυτό οφείλεται σε δύο λόγους: Πρώτον στο ότι στις περιοχές αυτές υπάρχουν έντονες ακμές και δεύτερον στο μορφολογικό στοιχείο που επιλέχθηκε για το closing, όπως θα δούμε και στην εικόνα 4.3.

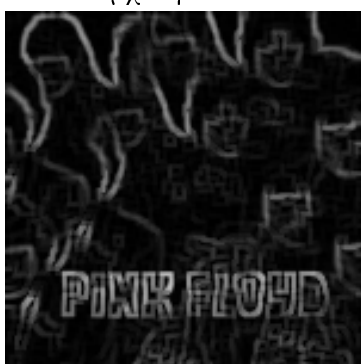
Στο σχήμα 4.3 παρατηρούμε ότι τα αποτελέσματα αλλάζουν ανάλογα με το μορφολογικό στοιχείο που επιλέγεται για το closing κάθε φορά. Τα καλύτερα αποτελέσματα όπως φαίνεται δίνονται για στοιχείο διαστάσεων 10×20 , όπου το τελικό αποτέλεσμα περιλαμβάνει σχεδόν μόνο το κείμενο. Όπως φαίνεται και από την πρώτη στήλη, το συγκεκριμένο στοιχείο είναι το κατάλληλο έτσι ώστε να έχουμε μόνο μία συνεκτική συνιστώσα, αυτή που περιέχει τα γράμματα. Αυτό επίσης που πρέπει να τονιστεί ιδιαίτερα είναι η σχέση του σχήματος



Αρχική Εικόνα



Θολωμένη (blurred), εικόνα



Μορφολογική παράγωγος



Κατωφλίωση (thresholding), παραγώγου



Διαστολή (Dilation)



Συστολή (Erosion)



Συνεκτικές συνιστώσες



2η κατωφλίωση - Εξαγωγή κειμένου

Σχήμα 4.2: Αναλυτική παρουσίαση της μεθόδου εντοπισμού κειμένου

του μορφολογικού στοιχείου με τη διάταξη των γραμμάτων. Στη συγκεκριμένη εικόνα έχουμε μόνο μία οριζόντια γραμμή κειμένου. Τα στοιχεία με μεγαλύτερη την οριζόντια διάσταση ομαδοποιούν τα γράμματα μεταξύ τους με τη διαστολή, ως πρώτο μέρος του closing, και έτσι παίρνανε ανέπαφα και από τη συστολή. Βλέπουμε στις 3 πρώτες γραμμές, ότι παρά τα όποια λάθη, τα γράμματα βγαίνουν όλα ακέραια. Αντίθετα στην τελευταία περίπτωση το μορφολογικό στοιχείο έχει μεγαλύτερη την κατακόρυφη διάσταση. Τα γράμματα προσπαθούν να ομαδοποιηθούν κατακόρυφα, όπως φαίνεται και από τις πολλές συνεκτικές συνιστώσες που δημιουργούνται. Αυτό όμως δε γίνεται καθώς υπάρχει μόνο μία γραμμή κειμένου και έτσι κάποια γράμματα δεν ανιχνεύονται καθόλου αφού δεν ομαδοποιούνται και κόβονται κατά τη συστολή erosion.

Στο πρώτο παράδειγμα (εικόνα 4.2), οι χαρακτήρες ικανοποιούν την παραδοχή που αναφέραμε στην αρχή της περιγραφής της μεθόδου, δηλαδή ότι έχουν τιμές φωτεινότητας αρκετά μεγαλύτερες από το background τους. Τι γίνεται στην περίπτωση που δεν ισχύει κάτι τέτοιο; Στις εικόνες του σχήματος 4.4 οι γραμματικοί χαρακτήρες έχουν αρκετά πιο μικρές τιμές φωτεινότητας από το background τους. Στην τελική εικόνα βλέπουμε ένα αντίθετο από τα προηγούμενα αποτέλεσμα. Δηλαδή τα pixel που ανήκουν σε γράμματα είναι και αυτά που έχουν κοπεί από την κατωφλίωση αντί να είναι αυτά που περνάνε από το κατώφλι. Για να πάρουμε τα επιθυμητά αποτελέσματα αρκεί να αντιστρέψουμε το αποτέλεσμα στις περιοχές κειμένου ή να θέσουμε την τιμή του κατωφλίου ως άνω φράγμα αντί για κάτω που είναι στις άλλες περιπτώσεις. Ο λόγος που συμβαίνει αυτό είναι ότι το κατώφλι γ_i υπολογίζεται με βάση την παράγωγο της εικόνας όπως αυτή προσεγγίζεται από τα φίλτρα g_1 και g_2 , οπότε για την τιμή του παίζει ρόλο η διαφορά μεταξύ των γραμμάτων και του background. Το κατώφλι όμως αυτό, εφαρμόζεται στη συνέχεια στις περιοχές της αρχικής εικόνας όπως φαίνεται από την εξ. (4.14). Επομένως από το κατώφλι περνάει η περιοχή της εικόνας που έχει τιμές φωτεινότητας μεγαλύτερες από το κατώφλι.

Τέλος, στις εικόνες του σχήματος 4.5, βλέπουμε ένα παράδειγμα που δείχνει τα όρια της συγκεκριμένης μεθόδου. Στην αρχική εικόνα υπάρχουν γράμματα με τιμές φωτεινότητας μεγαλύτερες, μικρότερες αλλά και παρόμοιες με το αντίστοιχο background. Στο τελικό αποτέλεσμα τα γράμματα που έχουν τιμές μεγαλύτερες από το αντίστοιχο background έχουν εντοπιστεί σωστά. Για τα γράμματα με φωτεινότητα μικρότερη από το αντίστοιχο background, αυτά διαγράφονται στην τελική εικόνα τους μέσω του background που περνάει από το κατώφλι όπως αναμενόταν. Τα γράμματα όμως με τιμές παρόμοιες με το αντίστοιχο background (U, R, O, G), δεν εντοπίζονται καθόλου. Μάλιστα από τη διαδικασία θολώσης τα γράμματα αυτά δεν είναι πλέον ευδιάκριτα και δεν υπάρχουν στην εικόνα που λαμβάνεται με την εφαρμογή του closing, οπότε είναι και αναμενόμενο να μην εντοπιστούν στην τελική εικόνα. Αν επιλέξουμε μην θολώσουμε την εικόνα τότε δεν καταπιέζονται οι δευτερεύουσες ακμές της



$B=\text{ones}(4 \times 10)$



$B=\text{ones}(6 \times 18)$



$B=\text{ones}(10 \times 20)$



$B=\text{ones}(15 \times 5)$

Σχήμα 4.3: Η μέθοδος δίνει διαφορετικά αποτελέσματα για διαφορετικά μορφολογικά στοιχεία με τα οποία γίνεται το closing. Σε κάθε γραμμή φαίνονται το αποτέλεσμα του closing στην εικόνα της κατωφλιωμένης παραγωγής και το τελικό αποτέλεσμα, ενώ αναγράφονται και οι διαστάσεις του στοιχείου B



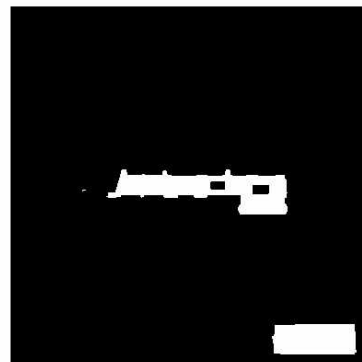
Αρχική Εικόνα



Μορφολογική Παράγωγος



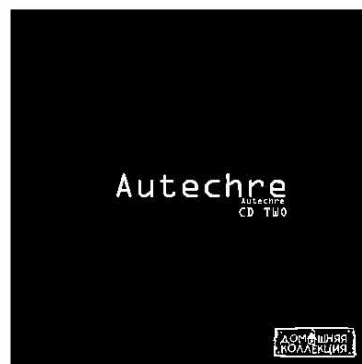
Κατωφλίωση Παραγώγου



Περιοχές μετά το closing



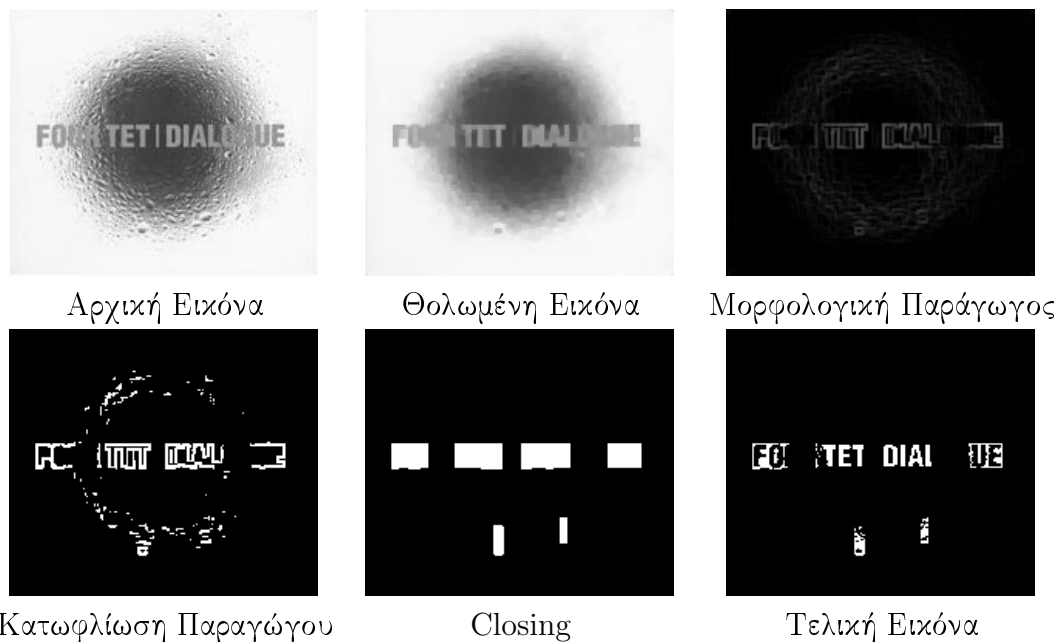
Τελική Εικόνα



Αντιστροφή περιοχής - Εξαγωγή Κειμένου

Σχήμα 4.4: Παράδειγμα σε εικόνα με σκούρα γράμματα

εικόνας και τα αποτελέσματα επηρεάζονται άμεσα όπως φαίνεται από την εικόνα 4.6.



Σχήμα 4.5: Παράδειγμα σε εικόνα με έντονα μεταβλητό background

4.3 Αυτόματη Απομάκρυνση Κειμένου από Εικόνες

Στην ενότητα 4.2 είδαμε ένα μοντέλο για εύρεση κειμένου σε εικόνες. Το μοντέλο αυτό, καθώς και άλλα παρόμοια, δίνουν ικανοποιητικά αποτελέσματα στις περισσότερες περιπτώσεις. Οι θέσεις των γραμμάτων εντοπίζονται σωστά και η μορφή του κάθε γράμματος εμφανίζεται στην έξοδο του συστήματος με λεπτομέρεια. Έτσι αν η δυαδική εικόνα της εξόδου δοθεί σε ένα σύστημα OCR, τότε το σύστημα αυτό αναγνωρίζει με επιτυχία το κείμενο και το μετατρέπει σε κώδικα ASCII για την περαιτέρω επεξεργασία.

Στην παρούσα εργασία όμως, ο σκοπός δεν είναι η εξαγωγή του κειμένου έτσι ώστε αυτό να μπορεί να αναγνωριστεί από ένα σύστημα OCR, αλλά ο ακριβής εντοπισμός του κειμένου έτσι ώστε αυτό να απομακρυνθεί από την εικόνα με τεχνικές Inpainting. Στην ουσία, η έξοδος του κειμένου αποτελεί και την Inpainting περιοχή - μάσκα. Έτσι ο προσδιορισμός του κειμένου πρέπει να είναι απόλυτα ακριβής και όχι απλώς να φαίνεται η δομή του κάθε γράμματος



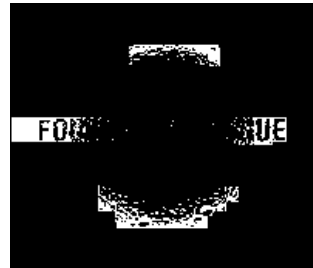
Μορφολογική Παράγωγος



Κατωφλίωση Παράγωγου



Closing



Τελική Εικόνα

Σχήμα 4.6: Αποτελέσματα χωρίς τη διαδικασία θόλωσης (blurring)

ώστε αυτό να μπορεί στη συνέχεια να αναγνωριστεί. Οι εικόνες του σχήματος 4.7, δείχνουν παραστατικά αυτή τη μικρή αλλά ουσιώδη διαφορά.



Αρχική Εικόνα



Εξαγωγή Κειμένου

Σχήμα 4.7: Η μορφή των γραμμάτων φαίνεται ξεκάθαρα, αλλά δεν έχει εντοπισθεί τελείως η θέση τους

Στις εικόνες του σχήματος 4.7, βλέπουμε ότι το κείμενο έχει εξαχθεί έτσι ώστε το κάθε γράμμα να φαίνεται ξεκάθαρα ποιο γράμμα είναι, πράγμα που είναι και το ζητούμενο στην αναγνώριση χαρακτήρων. Όμως, όπως φαίνεται εύκολα, το κείμενο δεν έχει εντοπιστεί ολόκληρο, (τα γράμματα στην έξοδο είναι πιο

λεπτά) δηλαδή υπάρχουν pixel της αρχικής εικόνας που αντιστοιχούν σε κείμενο, και τα οποία δεν εμφανίζονται στην έξοδο του συστήματος. Γίνεται φανερό ότι αν θέλουμε να απομακρύνουμε το κείμενο με κάποια Inpainting μέθοδο, δεν μπορούμε να θεωρήσουμε την έξοδο του συστήματος ως την Inpainting περιοχή, γιατί τότε δεν θα απομακρύνουμε όλα τα pixel κειμένου αλλά και γιατί τα pixel κειμένου που δεν υπάρχουν στην Inpainting θα επηρεάσουν αρνητικά το αποτέλεσμα του Inpainting καθώς θα είναι τα συνοριακά pixel της Inpainting περιοχής. Έτσι γίνεται φανερό ότι η αυτόματη απομάκρυνση γραμμμάτων από μία εικόνα χρειάζεται έναν πολύ ακριβή εντοπισμό τους.

Θεωρούμε ότι ο ακριβέστερος προσδιορισμός των γραμμμάτων πρέπει να γίνει ξεχωριστά (αλλά πάλι αυτόματα), από τη διαδικασία εντοπισμού τους, δηλαδή ότι οι μέθοδοι που προαναφέραμε δεν επιδέχονται κάποιας σημαντικής βελτιστοποίησης. Για παράδειγμα, μια αλλαγή του κατωφλίου ώστε να περάσουν περισσότερα pixel γραμμμάτων, θα είχε ως αποτέλεσμα να περάσουν και περιοχές οι οποίες δεν ανήκουν σε γράμματα και τις οποίες είναι λάθος να τις θεωρήσουμε σαν μέρος της Inpainting περιοχής.

Επομένως το πρόβλημα τώρα μπορεί να διατυπωθεί ως εξής: Δεδομένης της αρχικής εικόνας και μιας πρώτης (αλλά καλής) εκτίμησης των γραμμμάτων να εντοπιστούν τα γράμματα με απόλυτα ακριβή τρόπο, ώστε αυτός ο εντοπισμός να ορίζει αυτόματα την Inpainting περιοχή.

Μία πρώτη σκέψη θα ήταν, στην αρχική εκτίμηση των γραμμμάτων να εφαρμόσουμε έναν τελεστή μορφολογικής διαστολής, (dilation), έτσι ώστε στη νέα εικόνα που θα προκύψει να περιέχονται περισσότερα pixel κειμένου. Η μέθοδος αυτή μπορεί να εφαρμοστεί αλλά δεν είναι γενική καθώς προκύπτουν δύο σημαντικά προβλήματα. Πρώτον, ποιο είναι το κατάλληλο μέγεθος της διαστολής που πρέπει να εφαρμόσουμε, και δεύτερο μια μεγάλη διαστολή μπορεί να αυξήσει πολύ το μέγεθος της Inpainting περιοχής, η οποία να συμπεριλάβει και περιοχές που είναι απαραίτητες για την καλή ανακατασκευή. Για παράδειγμα οι περιοχές ανάμεσα στα γράμματα είναι πολύ χρήσιμες ειδικά για την περίπτωση του Inpainting περιεχομένου, όπως αναφέραμε και στην αρχή του κεφαλαίου, ενώ με μία μεγάλη διαστολή αυτές οι περιοχές μπορεί να συμπεριληφθούν στην Inpainting περιοχή.

Αυτό που ζητείται είναι η διαστολή της περιοχής αλλά έτσι ώστε να καλύψει μόνο της περιοχές όπου υπάρχει κείμενο. Αυτό το ζητούμενο μας οδηγεί σε 2 μορφολογικούς τελεστές της άλγεβρας πλεγμάτων (lattice algebra): Τη διαστολή υπό συνθήκη, (conditional dilation), και το reconstruction opening.

Η διαστολή υπό συνθήκη ορίζεται για σύνολα ως εξής:

$$\delta_B(M|X) \equiv (M \oplus B) \cap X \quad (4.18)$$

όπου B είναι το μορφολογικό στοιχείο που χρησιμοποιούμε, M είναι το

σύνολο που θέλουμε να διασταλεί και X είναι ένα άλλο σύνολο. Ο τελεστής της διαστολής υπό συνθήκη έχει νόημα μόνο όταν $M \subset X$, και η φυσική του ερμηνεία είναι ότι το σύνολο M παίζει το ρόλο του δείκτη (marker), που πρέπει να διασταλεί αλλά να παραμένει μέσα στα όρια του συνόλου X .

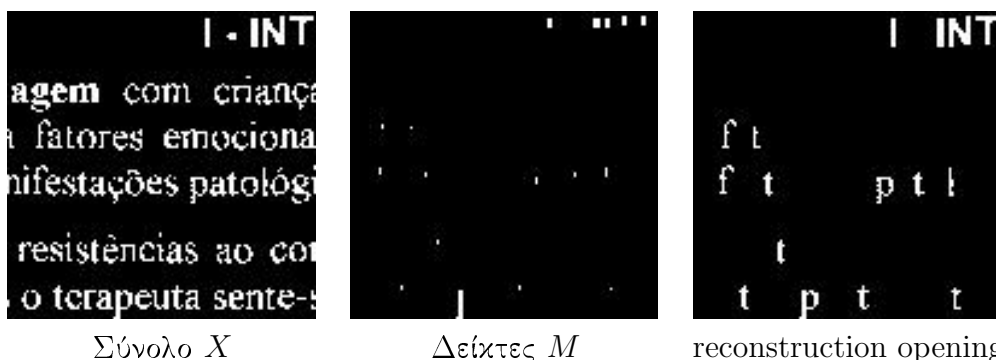
Αν τώρα έχουμε ένα σύνολο $X = \bigcup_i X_i$ ξένων συνεκτικών συνιστωσών και ένας δείκτης M που ανήκει σε κάποιο X_i , τότε ως reconstruction opening ορίζεται ο τελεστής

$$\rho^-(M|X) \equiv \eta \text{ συνεκτική συνιστώσα του } X \text{ που περιέχει τον } M \quad (4.19)$$

Εύκολα μπορούμε να δείξουμε ότι το reconstruction opening αποτελεί τη τελική κατάσταση μιας σειράς επαναλαμβανόμενων διαστολών υπό συνθήκη.

$$\lim_{t \rightarrow \infty} \delta_B(\dots(\delta_B(\delta_B(M|X)|X)\dots)|X) = \rho^-(M|X) \quad (4.20)$$

Στην εικόνα 4.8 βλέπουμε ένα παράδειγμα του reconstruction opening.



Σχήμα 4.8: Παράδειγμα reconstruction opening $\rho^-(M|X)$

Στην περίπτωσή μας, δεν μπορούμε να εφαρμόσουμε την τεχνική του reconstruction opening καθώς δεν διαθέτουμε την του περιγράμματος X . Ο αλγόριθμος που εφαρμόζουμε για την καλή διαστολή του συνόλου των pixel κειμένου στηρίζεται στην υπόθεση, ότι τα pixel κειμένου έχουν παρόμοιες τιμές φωτεινότητας μεταξύ τους, και αποτελείται από τα παρακάτω βήματα:

Έστω M η δυαδική εικόνα που αποτελεί την πρώτη πρόβλεψη για τις θέσεις των χαρακτήρων και I η αρχική εικόνα.

Αρχικά $M^1 = M$

1. Εφαρμόζουμε διαστολή με έναν μοναδιαίο δίσκο στην εικόνα M^n . $D = M^n \oplus B$.

$$2. \text{ Υπολογίζουμε την μέση τιμή } s = \frac{\sum(I \cdot M^n)}{\sum M^n}$$

3. Υπολογίζουμε τη νέα πρόβλεψη

$$M^{n+1} = \{(i, j) \in D : |I(i, j) - s| < \gamma\},$$

όπου γ ένα κατώφλι που εκφράζει μια εκτίμηση για τη διαφορά στη φωτεινότητα των pixel χαρακτήρων από το background π.χ. 30.

4. Επανάληψη των βημάτων μέχρι τη σύγκλιση όπου $M^{n+1} = M^n$

Στο τέλος του αλγορίθμου η περιοχή κειμένου θα έχει διασταλεί έτσι ώστε να περιλαμβάνει όλα τα σημεία με τιμές φωτεινότητας παρόμοιες με αυτές της αρχικής πρόβλεψης οπότε θα έχει συμπεριλάβει όλα τα σημεία που αντιστοιχούν σε κείμενο.

Αν το κείμενο αποτελείται από περιοχές με παρόμοια φωτεινότητα στο εσωτερικό τους, τότε η τελική πρόβλεψη μπορεί να γίνει αρχίζοντας και από μια πρόβλεψη που περιέχει λίγα μόνο σημεία, και δεν δείχνει ξεκάθαρα τι είναι το κάθε γράμμα. Αυτό μας επιτρέπει να αλλάξουμε το κατώφλι στις μεθόδους εντοπισμού που προαναφέραμε έτσι ώστε να αποκλείσουμε άλλες μικρές περιοχές που δεν είναι χαρακτήρες αλλά βγαίνουν στην έξοδο του συστήματος.

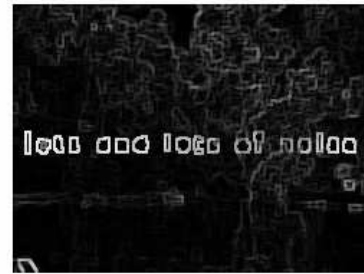
4.3.1 Αποτελέσματα

Ακολουθούν παραδείγματα της συγκεκριμένης μεθόδου, όπου περιλαμβάνει αυτόματο εντοπισμό των γραμματικών χαρακτήρων και απομάκρυνση τους με τεχνικές Inpainting.

Στο σχήμα 4.9 βρίσκουμε αρχικά τις πιθανές θέσεις των γραμμάτων με τη μορφολογική μέθοδο της ενότητας 4.2.2. Η εικόνα που παίρνουμε περιέχει τις θέσεις των γραμμάτων, αλλά ο εντοπισμός τους δεν είναι πλήρης ενώ παράλληλα περιέχονται και κάποια pixel που δεν αντιστοιχούν σε κείμενο. Επομένως στη συνέχεια εφαρμόζεται ο προτεινόμενος αλγόριθμος για τον πλήρη εντοπισμό της θέσεως των γραμμάτων. Για την εξάλειψη των μεμονωμένων pixel που δεν αντιστοιχούν σε κείμενο χρησιμοποιούμε την τεχνική του *area opening*. Με την τεχνική αυτή μπορούμε σε μία δυαδική εικόνα να εξαλείψουμε τις συνεκτικές συνιστώσες που έχουν εμβαδόν κάτω από ένα συγκεκριμένο κατώφλι. Η υλοποίησή της τεχνικής είναι πολύ απλή αφού περιλαμβάνει αρχικά την εύρεση των συνεκτικών συνιστωσών της δυαδικής εικόνας και στη συνέχεια κάθε συνιστώσα αποκόπτεται αν το εμβαδόν της είναι κάτω από ένα κατώφλι (π.χ. 10). Στην τελική εικόνα που προκύπτει μπορούμε να εφαρμόσουμε έναν τελεστή μορφολογικής διαστολής (*dilation*), με έναν μικρό δίσκο (ακτίνας 1), έτσι



Αρχική Εικόνα (greyscale)



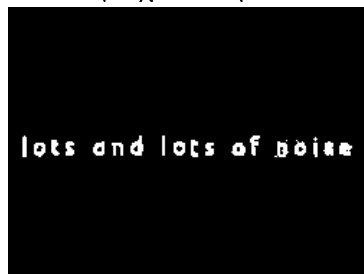
Μορφολογική Παράγωγος



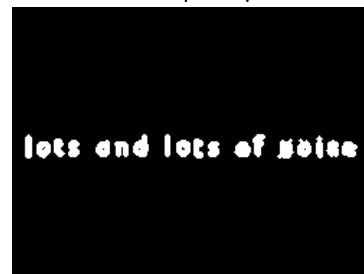
Περιοχές Κειμένου



Εξαγωγή Κειμένου



Ανακατασκευή Μάσκας



Τελική Μάσκα (Inpainting περιοχή)



Αρχική Εικόνα



Απομάκρυνση Γραμμάτων (Inpainting υφής)

Σχήμα 4.9: Παράδειγμα αυτόματου εντοπισμού και απομάκρυνση γραμμάτων με Inpainting υφής

ώστε να είμαστε σίγουροι ότι το κείμενο έχει εντοπισθεί πλήρως. Στη συνέχεια εφαρμόζουμε Inpainting υφής με τον αλγόριθμο των Efros-Leung στα 3 κανάλια της εικόνας, το οποίο δίνει πολύ καλά αποτελέσματα, όπως φαίνεται και στην τελική εικόνα.

Πρέπει να σημειώσουμε ότι η επιλογή του Inpainting υφής έγινε, αφού προηγουμένως είχαμε δει ότι το κείμενο βρίσκεται σε περιοχή με έντονη υφή. Ενδιαφέρον παρουσιάζει και η αυτοματοποίηση αυτού του σταδίου δηλαδή να επιλέγεται αυτόματα ποια μέθοδος Inpainting μπορεί να εφαρμοστεί. Για να γίνει κάτι τέτοιο χρειάζεται ένας τρόπος (αυτόματου) προσδιορισμού για το αν η γειτονιά της Inpainting περιοχής είναι περιοχή υφής ή περιεχομένου. Για το πρόβλημα αυτό υπάρχει σχετική βιβλιογραφία π.χ. [29], [53]. Μία λύση πάντως θα ήταν η εφαρμογή του ταυτόχρονου Inpainting υφής και περιεχομένου 3.4 που καλύπτει όλες τις περιπτώσεις αλλά εισάγει υπολογιστική πολυπλοκότητα χωρίς αυτό να είναι πάντα αναγκαίο.



Αρχική Εικόνα Εντοπισμός κειμένου Τελική dilated Μάσκα Inpainted Εικόνα

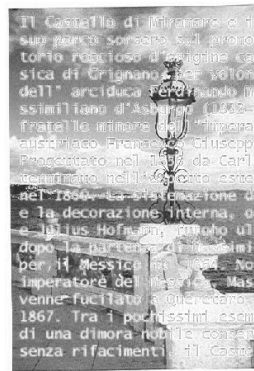
Σχήμα 4.10: Παράδειγμα αυτόματου εντοπισμού και απομάκρυνση γραμμάτων με ταυτόχρονο Inpainting περιεχομένου και υφής

Στο σχήμα 4.10 παρουσιάζονται τα βασικά στάδια ενός παράδειγμα με αυτόματο εντοπισμό του κειμένου και απομάκρυνσής του, με ταυτόχρονο Inpainting υφής και περιεχομένου. Αναλυτικά όλα τα στάδια έχουν παρουσιασθεί πριν (σχήματα 4.2, 3.20).

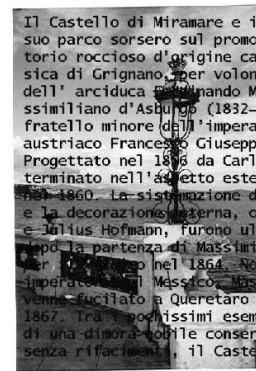
Τέλος στο σχήμα 4.11 βλέπουμε ένα ακόμα παράδειγμα. Για τον προσδιορισμό του κειμένου αν αυτός γινόταν με βάση την greyscale εικόνα τα αποτελέσματα δεν είναι καθόλου ικανοποιητικά λόγω των παραπλήσιων τιμών φωτεινότητας που έχει το κείμενο με το background. Το γεγονός ότι το κείμενο έχουν κόκκινο χρώμα μας οδηγεί στο συμπέρασμα ότι στην R-συνιστώσα το κείμενο θα έχει μεγάλες τιμές φωτεινότητας ενώ στην G-συνιστώσα πολύ μικρές. Έτσι βρίσκουμε το κείμενο και για τις δύο εικόνες και οι δύο μάσκες που προκύπτουν αθροίζονται λογικά ώστε να προκύψει η τελική Inpainting περιοχή. Το Inpainting στη συνέχεια γίνεται με το BCBS μοντέλο (ενότητα 2.4).



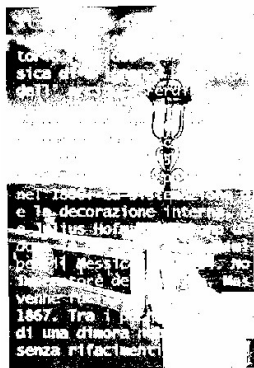
Αρχική Εικόνα



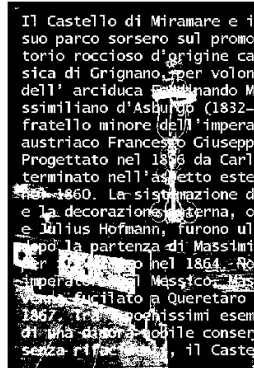
R-συνιστώσα



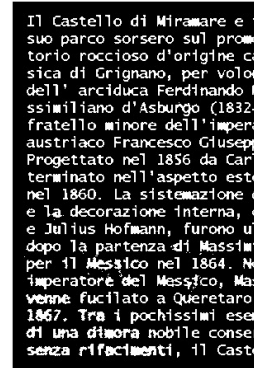
G-συνιστώσα



Εύρεση Κειμένου
R-συνιστώσα



Εύρεση Κειμένου
G-συνιστώσα



Τελική Μάσκα
(Λογικό ΚΑΙ)



Τελική Εικόνα

Σχήμα 4.11: Παράδειγμα αυτόματου εντοπισμού και απομάκρυνση γραμμάτων με Inpainting περιεχομένου

Κεφάλαιο 5

Συμπεράσματα

5.1 Συμβολή της Διπλωματικής Εργασίας

Όσον αφορά την πρωτοτυπία η εργασία αυτή συνεισφέρει κυρίως στα εξής σημεία:

- Συνδιασμό των τεχνικών Inpainting και των τεχνικών αυτόματου εντοπισμού κειμένου σε εικόνες και πρόταση ενός νέου αλγορίθμου για αυτόματο εντοπισμό και απομάκρυνση κειμένου από εικόνες. Από όσο ξέρουμε μία τέτοια εφαρμογή δεν υπάρχει μέχρι στιγμής στη βιβλιογραφία, ενώ στην εφαρμογή αυτή προτείνεται και ένας τρόπος ώστε ο εντοπισμός της θέσης του κειμένου σε μία εικόνα να είναι πλήρης και ακριβής.
- Συσχέτιση όλων των Inpainting μοντέλων και περιεχομένου και υψής με το θεμελιώδες πρόβλημα της επεξεργασίας εικόνας. Όλα τα μοντέλα Inpainting που υπάρχουν μέχρι στιγμής σχετίζονται με το θεμελιώδες πρόβλημα της επεξεργασίας εικόνας αφού το καθένα μοντελοποιεί την εικόνα με κάποιον από τους γενικούς τρόπους, με τους οποίους μοντελοποιούνται γενικά οι εικόνες στην επεξεργασία εικόνας (προσομοίωση φυσικών διαδικασιών, τυχαία πεδία, χώροι συναρτήσεων).

Εκτός όμως από τα στοιχεία πρωτοτυπίας της η διπλωματική εργασία ασχολήθηκε και με τα εξής αντικείμενα-κατευθύνσεις:

- Παρουσίαση όλων των PDE μοντέλων, και υλοποίηση των σημαντικότερων από αυτά, που χρησιμοποιούνται για Inpainting περιεχομένου, καθώς και ανάλυση αριθμητικών σχημάτων που χρησιμοποιούνται για την επίλυσή τους.

- Παρουσίαση και υλοποίηση των σημαντικότερων αλγορίθμων σύνθεσης υφής και το πως μπορούν αυτοί να χρησιμοποιηθούν στο Inpainting πρόβλημα με ιδιαίτερα εντυπωσιακά αποτελέσματα.
- Ανάλυση του προβλήματος διαχωρισμού περιεχομένου και υφής και χρησιμοποίησή του για τη μέθοδο του ταυτόχρονου Inpainting περιεχομένου και υφής.
- Παρουσίαση του προβλήματος αυτόματου εντοπισμού θέσεως κειμένου σε εικόνα και υλοποίηση μίας απλής αλλά αποτελεσματικής μεθόδου για την επίλυσή του.

Γενικά θεωρούμε ότι τα 3 πρώτα κεφάλαια της εργασίας, αποτελούν ένα καλό και πλήρες εγχειρίδιο για όποιον θέλει να ασχοληθεί με το Inpainting πρόβλημα, γιατί συγκεντρώνουν μαζί όλα τα μοντέλα Inpainting που υπάρχουν μέχρι σήμερα. Ιδιαίτερη σημασία έχει το γεγονός ότι τα σημαντικότερα Inpainting μοντέλα που αναφέρθηκαν, υλοποιήθηκαν σε υπολογιστή κατά τη διάρκεια εκπόνησης της εργασίας. Το γεγονός αυτό μας βοήθησε να αποκτήσουμε εμπειρία πάνω στο πρόβλημα και να μπορέσουμε να συγκρίνουμε, ως προς την απόδοση αλλά και τη μεθοδολογία, τα διάφορα Inpainting μοντέλα αλλά και να τα συνδιάσουμε μεταξύ τους, για καλύτερα αποτελέσματα. Παράλληλα πρέπει να σημειώσουμε ότι μία τέτοια πλήρης καταγραφή των μεθόδων για το Inpainting πρόβλημα, δεν υπάρχει μέχρι στιγμής στη βιβλιογραφία.

5.2 Κατευθύνσεις για Μελλοντική Έρευνα

Το Inpainting πρόβλημα, όπως έχει αναφερθεί, είναι ένα πολύ πρόσφατο πρόβλημα της όρασης υπολογιστών. Παρόλο που έχει γίνει σημαντική πρόοδος, πιστεύουμε ότι υπάρχουν πολλά περιθώρια για περαιτέρω έρευνα καθώς πολλά ερωτήματα έχουν μείνει αναπάντητα. Ορισμένες κατευθύνσεις που μπορεί να στραφεί η μελλοντική έρευνα είναι οι παρακάτω:

- *Εύρεση καλύτερων PDE μοντέλων για Inpainting περιεχομένου*

Τα PDE μοντέλα 3ου βαθμού που εξετάσαμε (BCBS, CDD, Euler's Elastica), δεν είναι ευσταθή ενώ η σύγκρισή τους είναι αργή και όχι απαραίτητα εξασφαλισμένη. Χρειάζεται να βρεθούν νέα PDE σχήματα που να είναι ευσταθή για μεγάλα χρονικά βήματα (ή ακόμα καλύτερα ανεξαρτήτως χρονικού βήματος), τα οποία να συγκλίνουν γρηγορότερα προς την επιθυμητή λύση.

- *Εκμετάλευση της πληροφορίας όλης της εικόνας και όχι μόνο της γειτονιάς της Inpainting περιοχής στις περιπτώσεις Inpainting περιεχομένου*
Όπως αναφέρθηκε, οι συντηρητές έργων τέχνης αποφασίζουν για το πως θα ανακτήσουν τη ζητούμενη περιοχή με βάση όλη την εικόνα και όχι μόνο τη γειτονιά της περιοχής αυτής. Κάτι τέτοιο γίνεται μόνο με τις μεθόδους Inpainting υφής και όχι με τις μεθόδους Inpainting περιεχομένου.
- *Σε περιπτώσεις Inpainting υφής, αυτόματος προσδιορισμός της δομικής μονάδας υφής texel*
Όπως αναφέρθηκε στο κεφάλαιο 3, η παράμετρος που πρέπει να επιλέξουμε για τη σύνθεση υφής είναι το μέγεθος του παραθύρου, και το οποίο πρέπει να έχει διάσταση τουλάχιστον ίση από τη μεγαλύτερη διάσταση της δομικής μονάδας της υφής texel, για να έχουμε καλά αποτελέσματα. Μία αυτόματη επιλογή του παραθύρου έχει πολλά οφέλη, καθώς το μόνο που πρέπει να προσδιοριστεί στη συνέχεια είναι η Inpainting περιοχή. Η σύνθεση υφής με ανάλυση σε πολλαπλές κλίμακες (ενότητα 3.2.2, [64]), δίνει μία λύση, αλλά ακόμα καλύτερο θα ήταν ο αυτόματος προσδιορισμός των texel.
- *Εφαρμογή και άλλων Inpainting μοντέλων για ψηφιακό zoom-in*
Το TV Inpainting μοντέλο μπορεί να χρησιμοποιηθεί όπως είδαμε για το ψηφιακό zoom-in σε μία εικόνα με παραπλήσια αποτελέσματα της μεθόδου bicubic interpolation. Το γεγονός ότι το ψηφιακό zoom-in με Inpainting μεθόδους, γίνεται με μία εντελώς διαφορετική θεώρηση της εικόνας και του καλού τελικού αποτελέσματος, από ότι με τις μεθόδους παρεμβολής, οδηγεί στην υπόθεση ότι κάποια άλλη μέθοδος Inpainting περιεχομένου πιθανώς να δίνει ακόμα καλύτερα αποτελέσματα.
- *Συνδιασμός Inpainting περιεχομένου και υφής για αποτελεσματικότερη κωδικοποίηση και συμπίεση εικόνας*
Η εφαρμογές της συμπίεσης εικόνας περιεχομένου, της σύνθεσης υφής αλλά και του ταυτόχρονου Inpainting περιεχομένου και υφής μπορούν να συνδιαστούν για υψηλούς και με μικρές απώλειες βαθμούς συμπίεσης εικόνας. Αυτό που πάλι έχει ιδιαίτερο ενδιαφέρον είναι η διαφορετική αντιμετώπιση του προβλήματος από τα ήδη υπάρχοντα πρότυπα συμπίεσης (π.χ. JPEG).
- *Αυτόματος προσδιορισμός της Inpainting περιοχής - μάσκας για περισσότερες περιπτώσεις*
Όπως, έχει αναφερθεί ο προσδιορισμός της Inpainting περιοχής γίνεται από τον χρήστη αφού αυτή είναι κάτι το υποκειμενικό. Ορισμένες όμως

κατηγορίες αντικειμένων μπορούν να διορθωθούν. Πέρα από το κείμενο σε μία εικόνα, άλλες κατηγορίες αντικειμένων που μπορούν να τυποποιηθούν είναι οι ρωγμές σε έναν πίνακα ζωγραφικής ή μία φωτογραφία, τα χαμένα πακέτα μιας εικόνας που προκύπτουν από τη μετάδοση κ.λ.π.

Οι κατευθύνσεις αυτές μπορούν να αποτελέσουν αντικείμενο μελλοντικής έρευνας.

Βιβλιογραφία

- [1] P. Perez A. Criminisi και K. Toyama. Object removal by exemplar based inpainting. Στο *Proc. IEEE Computer Society Conf. on Computer Vision and Pattern Recognition*, τόμος 22, σελίδες 721–728, Ιούνιος 2003.
- [2] Y. Weiss A. Levin, A. Zomet. Learning how to inpaint from global image statistics. *Intl. Conf. on Computer Vision, Nice, France*, Οκτώβριος 2003.
- [3] Jr. A. Yezzi A. Tsai και A. S. Welski. Curve evolution implementation of the mumford - shah functional for images segmentation, denoising, interpolation and magnification. *IEEE Trans. Image Proc.*, 10(8):1169–1186, 2001.
- [4] W.J. Fitzgerald P.J.W. Rayner A.C. Kokaram, R.D. Morris. Detection of missing data in image sequences. *IEEE Transactions on Image Processing*, 11(4):1496–1508, 1995.
- [5] L. Ambrosio και V. M. Tortorelli. Approximation of functionals depending on jumps by elliptic functionals via γ - convergence. *Comm. Pure Applied Math*, 43:999–1036, 1990.
- [6] L. Ambrosio και V. M. Tortorelli. On the approximation of free discontinuity problems. *Boll. Un. Mat. Ital.*, 6-B:105–123, 1992.
- [7] J. Bergen. *Spatial Vision, Chapter: Theories of Visual Texture Perception*, σελίδες 114–133. CRC Press, 1991.
- [8] D. Bloomberg. Image analysis using threshold reduction. Στο *SPIE Conf. on Image Algebra and Morphological Image Processing*, τόμος 1568, σελίδες 38–52, San Diego, CA, Ιούλιος 1991.
- [9] D. Bloomberg. Multiresolution morphological image analysis of document images. Στο *SPIE Conf. on Visual Communications and Image Processing*, τόμος 1818, σελίδες 648–662, Boston, MA, Νοέμβριος 1992.

- [10] D. Bloomberg και P. Maragos. Generalized hit-miss operators with applications to document image analysis. Στο *SPIE Conf. on Image Algebra and Morphological Image Processing*, San Diego, CA, Ιούλιος 1990.
- [11] J.S. De Bonet. Multiresolution sampling procedure for analysis and synthesis of texture images. Στο *Proc. ACM Conf. on Comp. Graphics (SIGGRAPH)*, σελίδες 361–368, 1997.
- [12] V. Caselles. G. Sapiro C. Ballester, M. Bertalmio και J. Verdera. Filling-in by joint interpolation of vector fields and gray levels. *IEEE Transactions On Image Processing*, 10(8):1200–1211, Αύγουστος 2001.
- [13] V. Caselles. G. Sapiro C. Ballester, M. Bertalmio και J. Verdera. A variational model for filling-in gray level and color images. *In Proc. 8th IEEE Intl. Conf. on Computer Vision, Vancouver BC Canada*, σελίδες 10–16, 2001.
- [14] T. Chan και J. Shen. Mathematical models for local non-textured inpaintings. *UCLA CAM Reports*, Μάρτιος 2000.
- [15] T. Chan και J. Shen. Non-texture inpainting by cutvature-driven diffusions (cdd). *UCLA CAM Report 00-35*, Σεπτέμβριος 2000.
- [16] A. Efros και W.T. Freeman. Image quilting for texture synthesis and transfer. Στο *Proc. ACM Conf. Comp. Graphics (SIGGRAPH)*, σελίδες 341–346, Eugene Fiume, Αύγουστος 2001.
- [17] A. Efros και T. Leung. Texture synthesis by non-parametric sampling. Στο *IEEE Intl. Conf. on Computer Vision*, Corfu, Greece, Σεπτέμβριος 1999.
- [18] G. Emile-Male. *The Restorer's Handbook of Easel Painting*. New York, 1976.
- [19] S. Esedoglu και J. Shen. Digital inpainting based on the mumford-shah-euler image model. *European Journal of Applied Math.*, 13:353–370, 2002.
- [20] D. Garber. *Computational Models for Texture Analysis and Texture Synthesis*. Διδακτορική Διατριβή, 1981.
- [21] S. Geman και D. Geman. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, ΠAMI-6(6):721–741, Νοέμβριος 1984.

- [22] A. Gersho και R.M. Gray. *Vector Quantization and Signal Compression*. Kluwer Academic Publishers, 1992.
- [23] R. C. Gonzalez και R. E. Woods. *Digital Image Processing*. Addison-Wesley, New York, 1992.
- [24] Y. Hasan και L. Karam. Morphological text extraction from images. *IEEE Transactions on Image Processing*, 9(11):1978–1983, Νοέμβριος 2000.
- [25] D. Heeger και J. Bergen. Pyramid-based texture analysis synthesis. Στο *Proc. ACM Conf. on Comp. Graphics (SIGGRAPH)*, τόμος 29, σελίδες 229–238, 1995.
- [26] D.H. Hubel και T.N. Wiesel. Binocular interaction in striate cortex of kittens reared with artificial squint. *Journal of Neurophysiology*, 28:1041–1059, 1965.
- [27] M. Bertalmio J. Verdera, V. Caselles και G. Sapiro. Inpainting surface holes. Στο *Proceeding IEEE Intl. Conf. on Image Processing*, Barcelona, Spain, Σεπτέμβριος 2003.
- [28] B. Julesz και J.R. Bergen. Textons, the fundamental elements in preattentive vision and perception of textures. 1983.
- [29] A. K. Jain K. Karu και R. M. Bolle. Is there any texture in the image? *Pattern Recognition*, 29(9):1437–1446, 1996.
- [30] S. H. Kang. *Mathematical Approaches to Color Denoising and Image Inpainting Problems*. Διδακτορική Διατριβή, UCLA, 2002.
- [31] G. Kanizsa. *Organization in Vision*. Praeger, New York, 1979.
- [32] G. Karagiannis και G. Steinhauer. *Machine Learning and Pattern Recognition*. N.T.U.A., Athens, Νοέμβριος 2001.
- [33] S.H. Park K.I. Kim, K. Jung και H.J. Kim. Support vector machines for texture classification. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 24(11):1542–1550, 2002.
- [34] N. Fusco L. Ambrosio και D. Pallara. *Functions of Bounded Variation and Free Discontinuity Problems*. Oxford Univ. Press, Oxford, U.K., 2000.

- [35] Y. Xu B. Guo L. Liang, C. Liu και H. Y. Shum. Real-time texture synthesis by patch-based sampling. Τεχνική Αναφορά 2001-40, Microsoft Research, Μάρτιος 2001.
- [36] S. Osher L. Rudin και E. Fatemi. Nonlinear total variation based noise removal algorithms. *Phys. D*, 60:259–268, 1992.
- [37] A. L. Bertozzi M. Bertalmio και G. Sapiro. Navier-stokes, fluid dynamics and image and video inpainting. *Proc. of IEEE Computer Society Conf. on Computer Vision and Pattern Recognition*, σελίδες 355–362, 2001.
- [38] C. Ballester G. Sapiro M. Bertalmio, V. Casseles. Image inpainting. ACM Conf. Computer Graphics (SIGGRAPH), σελίδες 417–424, Ιούλιος 2000.
- [39] G. Sapiro M. Bertalmio, L. Vese και S. Osher. Simultaneous structure and texture image inpainting. *IEEE Transactions on Image Processing*, 12(8):882–889, Αύγουστος 2003.
- [40] D. Mumford M. Nitzberg και T. Shiota. *Filtering, Segmentation and Depth*, τόμος 662. Springer-Verlag, Berlin, 1993.
- [41] P. Maragos. *Computer Vision*. N.T.U.A., 2002.
- [42] A. Marquina και S. Osher. *Lecture Notes in Computer Science, A new time dependent model based on level set motion for nonlinear deblurring and noise removal*, τόμος 1682, σελίδες 429–434. 1999.
- [43] D. Marr και E. Hildreth. Theory of edge detection. *Proc. Royal Soc. London*, σελίδες 187–217, 1980.
- [44] S. Masnou και J. M. Morel. Level-lines based disocclusion. Στο *5th IEEE Int'l Conf. on Image Processing*, αριθμός 3, σελίδες 259–263, 1998.
- [45] Y. Meyer. Oscillating pattern in image processing and nonlinear evolution equations. *AMS University Lecture Series*, 22, 2002.
- [46] D. Mumford. *Algebraic Geometry and its Applications*, σελίδες 491–506. Springer - Verlag, New York, 1994.
- [47] D. Mumford και J. Shah. Optimal approximations by piecewise smooth functions and associated variational problems. *Comm. Pure Applied Math. Volume: XLII*, σελίδες 577–685, 1989.

-
- [48] O. Okun και M. Pietikainen. A survey of texture based methods for document layout analysis. *Machine Perception and Artificial Intelligence*, 40:165–177, 2000.
- [49] O. Okun και M. Pietikainen. Edge-based method for text detection from complex document images. 2001.
- [50] P. Perona και J. Malik. Scale-space and edge detection using anisotropic diffusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(7):629–639, Ιούλιος 1990.
- [51] J. Portilla και E.P. Simoncelli. Texture representation and synthesis using corellation of complex wavelet coefficient magnitudes. *TR 54, CSIC, Madrid*, Απρίλιος 1999.
- [52] L. Rudin και S. Osher. Total variation based image restoration with free local constraints. Στο *1st IEEE Intl. Conf. Image Processing*, τόμος 1, σελίδες 31–35, 1994.
- [53] G. Sapiro S. D. Rane και M. Bertalmio. Structure and texture filling-in of missing image blocks in wireless transmission and compression applications. *IEEE Transactions on Image Processing*, 12(3):296–303, 2003.
- [54] J. Serra. *Image Analysis and Mathematical Morphology*. Acad. Press, New York, 1982.
- [55] T. Chan S.H. Kang και S. Soatto. Landmark based inpainting from multiple views. *UCLA CAM Report 02-11*, Μάρτιος 2002.
- [56] C.E. Shanon. A mathematical theory of communication. *Bell Sys. Tech. Journal*, 27, 1948.
- [57] J. Shen. Inpainting and the fundamental problem of image processing. *SIAM News*, 36(5), Ιούνιος 2003.
- [58] E.P. Simoncelli και J. Portilla. Texture characterization via joint statistics of wavelet coefficient magnitudes. Στο *Proc. 5th Intl. Conf. on Image Processing*, 1998.
- [59] L. Vese T. Chan, J. Shen. Variational pde models in image processing. *Notices of the AMS*, 50(1):14–26, Ιανουάριος 2003.
- [60] S. Kang T. Chan και J. Shen. Euler’s elastica and curvature based inpaintings. *UCLA CAM Reports 01-12*, 2001.

-
- [61] S. Osher T. Chan και J. Shen. The digital tv filter and nonlinear denoising. *IEEE Transactions on Image Processing*, 10(2):231–241, Φεβρουάριος 2001.
- [62] G. Sapiro V. Caselles, J. M. Morel και A. Tannenbaum. Introduction to the special issue on partial differential equations and geometry-driven diffusion in image processing and analysis. *IEEE Transactions on Image Processing*, 7(3):269–273, Μάρτιος 1998.
- [63] L. Vese και S. Osher. Modeling textures with total variation minimization and oscillating pattern in image processing. *UCLA CAM Report 02-19*, Μάιος 2002.
- [64] L.Y. Wei και M. Levoy. Fast texture synthesis using tree-structured vector quantization. Στο *Proc. of SIGGRAPH*, σελίδες 479–488, Ιούλιος 2000.