



Πανεπιστήμιο Πειραιώς – Τμήμα Πληροφορικής
Πρόγραμμα Μεταπτυχιακών Σπουδών
«Προηγμένα Συστήματα Πληροφορικής»

Μεταπτυχιακή Διατριβή

Τίτλος Διατριβής	Αναγνώριση και εξαγωγή ενσωματωμένου στατικού κειμένου από ψηφιακό σήμα εικόνας, με εφαρμογή ανομοιογενούς - πολυδιακριτικού μετασχηματισμού κυματίων, στις διαστάσεις του χώρου και του χρόνου.
Όνοματεπώνυμο Φοιτητή	Χαιρεκάκης Θεμιστοκλής του Γεωργίου
Αριθμός Μητρώου	ΠΣΠ/07023
Κατεύθυνση	Αναγνώριση Προτύπων & Μηχανική Μάθηση
Επιβλέπων	Γεώργιος Τσιχριντζής, Καθηγητής

Πανεπιστήμιο Πειραιώς-Τμήμα Πληροφορικής
Πρόγραμμα Μεταπτυχιακών Σπουδών στα
Προηγμένα Συστήματα Πληροφορικής

Ημερομηνία Παράδοσης **Ιούλιος 2010**

РАМЕТЪМО РЕПАА

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΡΑΙΑ

Τριμελής Εξεταστική Επιτροπή

(υπογραφή)

Όνομα Επώνυμο
Βαθμίδα

(υπογραφή)

Όνομα Επώνυμο
Βαθμίδα

(υπογραφή)

Όνομα Επώνυμο
Βαθμίδα

РАМЕТЪМО РЕПАА

РАМЕТЪМО РЕПАА

ΜΕΤΑΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

Τίτλος εργασίας: Αναγνώριση και εξαγωγή ενσωματωμένου στατικού κειμένου από ψηφιακό σήμα εικόνας, με εφαρμογή ανομοιογενούς - πολυδιακριτικού μετασχηματισμού κυματίων, στις διαστάσεις του χώρου και του χρόνου.

Πίνακας Περιεχομένων

Περίληψη:	3
Abstract:	3
1. Εισαγωγή.....	4
1.1. Ανίχνευση κειμένου (Text Detection):.....	5
1.2. Χωρικός εντοπισμός κειμένου (Text Localization):.....	6
1.2.1. Μέθοδος Περιοχής	6
1.2.2. Μέθοδος Υφής.....	7
2. Εισαγωγή στην κυματιδική ανάλυση:	9
2.1. Ιστορική διαδρομή	9
2.2. Εφαρμογές	10
2.3. Ανάλυση κατά Fourier	11
2.4. Short-Time Fourier Analysis.....	11
3. Κυματιδική ανάλυση	12
3.1. Κυματιδική ανάλυση - πλεονεκτήματα	14
3.2. Κυματιδική ανάλυση - ορισμός.....	15
3.3. Συνεχής μετασχηματισμός κυματιδίων – Continuous wavelet transform.....	15
3.4. Διακλιμάκωση.....	16
3.5. Μετάθεση.....	17
3.6. Διακριτός μετασχηματισμός κυματιδίων – Discrete wavelet transform.....	18
3.6.1. Εφαρμογές της Διακριτής Ανάλυσης κυματιδίων	18
3.7. Εφαρμογή ψηφιακών φίλτρων σε ένα στάδιο: Προσέγγιση και Λεπτομέρειες.....	19
3.8. Αποσύνθεση πολλαπλών επιπέδων	20
3.9. Κυματιδική αναδόμηση.....	21
3.10. Φίλτρα αναδόμησης	21
3.11. Αναδόμηση Προσεγγίσεων & Λεπτομερειών	22
3.12. Συσχέτιση φίλτρων διακριτής ανάλυσης – σχήματος κυματιδίων	23
3.13. Ολοκληρωμένη διαδικασία Αποδόμησης – Αναδόμησης πολλαπλών επιπέδων.....	23
3.14. Παράδειγμα διακριτής κυματιδικής ανάλυσης μονοδιάστατου σήματος	23
3.15. Διακριτή κυματιδική ανάλυση δισδιάστατου σήματος	27
3.16. Διακριτή κυματιδική ανάλυση 3D σήματος.....	32

4. Detecting Video Texts Using Spatial-Temporal Wavelet Transform:.....	38
4.1. Ανίχνευση και ανάδειξη των ακμών.....	38
4.2. Εξαγωγή χαρακτηριστικών.....	42
4.3. Ταξινόμηση:.....	48
4.4. Εξαγωγή κειμένου:.....	49
4.5.Τυποποίηση Συνδεδεμένων Στοιχείων.....	50
5. 3D Wavelet Toolbox for Textual Information Extraction in Video Sequences.....	60
5.1. Προγραμματιστικό περιβάλλον:.....	60
5.2. Διεπαφή χρήστη - Video W3D Toolbox:.....	61
5.2.1. Επιμέρους Συστατικά Στοιχεία Διαπαφής χρήστη - Video W3D Toolbox:.....	62
5.3. Παραδείγματα εφαρμογής σε πραγματικές συνθήκες:.....	66
5.3.1. Παράδειγμα 1.....	66
5.3.2. Παράδειγμα 2.....	74
6. Συμπεράσματα-Περίληψη:.....	76
7. Βιβλιογραφία:.....	78

Περίληψη:

Η αναγνώριση και εξαγωγή ενσωματωμένου κειμένου από ψηφιακό σήμα video, αποτελεί σημαντικό πεδίο έρευνας στον τομέα της πληροφορικής, διότι το κείμενο εμπεριέχει σημασιολογική πληροφορία η οποία δύναται να χαρακτηρίσει και να περιγράψει το υποκείμενο οπτικοακουστικό υλικό.

Η επιστημονική βιβλιογραφία, προτείνει διάφορες μεθόδους εντοπισμού, ανάδειξης και οπτικής αναγνώρισης πληροφορίας κειμένου σε video. Κεντρικός στόχος των μεθόδων αυτών είναι η απαλοιφή της συνήθως πολύπλοκης πληροφορίας που ορίζει το φόντο και η τελική ανάδειξη της πληροφορίας κειμένου.

Αντικείμενο της παρούσης εργασίας αποτελεί η μελέτη και η επέκταση των προαναφερθέντων μεθόδων, με σκοπό την ανάλυση, αναγνώριση και εξαγωγή της πληροφορίας στατικού και μηχανικά ενσωματωμένου κειμένου (*video embedded textual information*). Παρότι στην εργασία παρουσιάζονται τα πλεονεκτήματα αλλά και οι ανεπάρκειες υφιστάμενων προσεγγίσεων όπως η **περιοχική προσέγγιση** (*Region Based Approach*) και η **προσέγγιση υφής** (*Texture Based Approach*), γίνεται σαφές ότι η χρήση **Ανομοιογενούς – Πολλαπλών Αναλύσεων Μετασχηματισμού Κυματίων στο Χώρο και το Χρόνο** (*3-Dimensional Multiresolution Inhomogeneous Wavelet Transform*), μπορεί να αναδείξει με μεγαλύτερη ακρίβεια τα επιθυμητά χαρακτηριστικά.

Λαμβάνοντας υπόψη το συγκεκριμένο συμπέρασμα, προτείνεται μια καινοτόμος και αποτελεσματική διαδικασία επεξεργασίας, η οποία υλοποιείται μέσω της δημιουργίας ολοκληρωμένου προϊόντος λογισμικού (**3D Wavelet Toolbox for Textual Information Extraction in Video Sequences - W3D Toolbox**), που επιτρέπει την προβολή, παραμετροποίηση και συνδυασμένη χρήση των προκρινόμενων μεθόδων σε πραγματικά δεδομένα.

Abstract:

Detection and extraction of machine embedded text over digital video signal, is a significant research field in information technology, because textual information contains semantic information which may describe and classify the subjacent audiovisual content.

Scientific bibliography on this field suggests various methods of detection, emergence and optical recognition of video encoded textual information. The main target of such methods is the elimination of the usually complicated information which describes background and the extraction of textual information.

The objective of this project is the study and extension of the aforementioned methodologies in order to analyze, detect and extract video embedded, spatially and temporally static, textual information. Although this project presents advantages and deficiencies of such approaches as region based and texture based approach, it finally becomes clear that the use of an innovative 3-Dimensional Multiresolution Inhomogeneous Wavelet Transform technique in Spatial and Temporal domains may extract more efficiently the desired characteristics.

Following the above presumption, we suggest an innovative and efficient video content processing algorithm and additionally an integrated software product (**3D Wavelet Toolbox for Textual Information Extraction in Video Sequences - W3D Toolbox**), which permits the visualization, configuration and compound use of designated methodologies on real world data.

Εισαγωγή

Η πληροφορία κειμένου σε αρχεία πολυμέσων (εικόνα, video), αποτελεί ένα διακριτό είδος οπτικού αντικειμένου. Ο συνδυασμός των απεικονιζόμενων χαρακτήρων για την παραγωγή λέξεων και προτάσεων, δημιουργεί ένα επιπλέον επίπεδο πληροφορίας, που κωδικοποιεί και περιγράφει αφηρημένα και πλησιέστερα στην ανθρώπινη αντίληψη, την υποκείμενη πληροφορία.

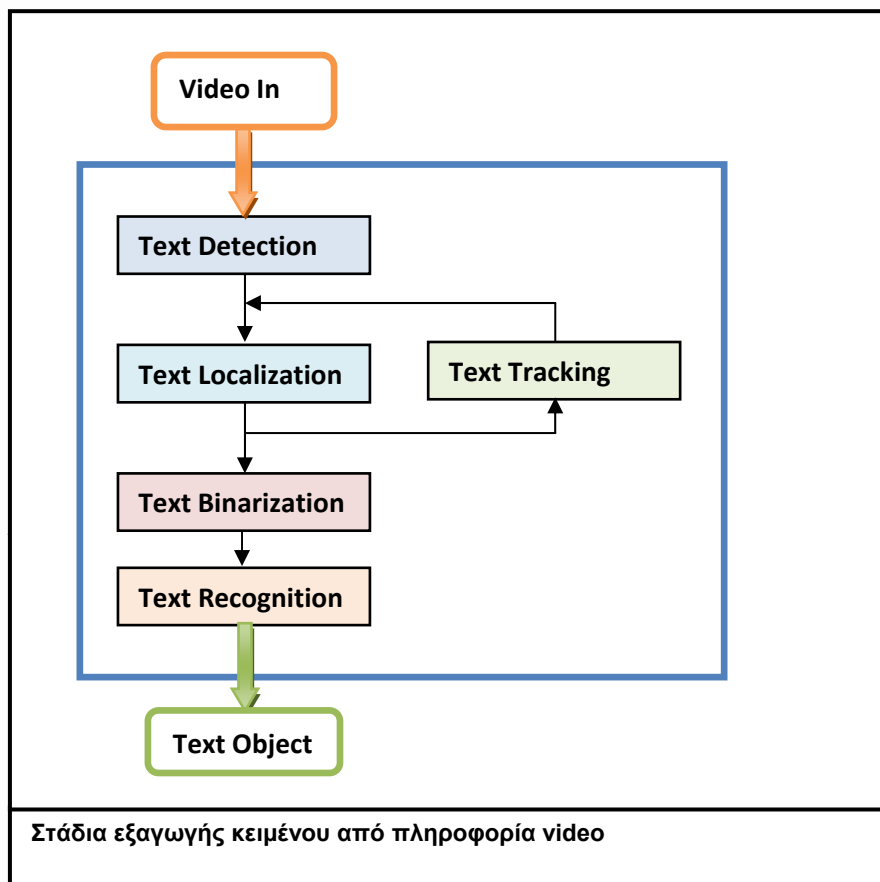
Το κείμενο στα αρχεία πολυμέσων διαχωρίζεται σε δύο βασικές κατηγορίες (Jung et al., 2004):

1. Στο **τεχνητά ενσωματωμένο κείμενο** (*Caption text*), για παράδειγμα:
 - a. Οι επικεφαλίδες κειμένου στις τηλεοπτικές και ειδησεογραφικές εκπομπές, όπου συνήθως καταγράφεται η ημερομηνία και η σύντομη λεκτική περιγραφή της οπτικής πληροφορίας.
 - b. Οι ενσωματωμένοι υπότιτλοι σε κινηματογραφικά και τηλεοπτικά προϊόντα, όπου συνήθως καταγράφεται το πλήρες κείμενο του σεναρίου.
2. Στο **κείμενο που ενυπάρχει φυσικά στο πεδίο λήψης της κάμερας** κατά τη διάρκεια της καταγραφής (*Scene Text*), για παράδειγμα η οδική σήμανση, οι διαφημιστικές πινακίδες, οι πινακίδες κυκλοφορίας οχημάτων κ.λπ.)

Τα τελευταία χρόνια, λόγω της ταχέως αυξανόμενης παραγωγής ψηφιακών αρχείων πολυμέσων, πρόεκυψε η ανάγκη δημιουργίας βάσεων μεταδεδομένων, οι οποίες θα παράσχουν δυνατότητες αναζήτησης και ανεύρεσής του οπτικοακουστικού υλικού βάσει του θεματικού του περιεχομένου (*content-based video retrieval*), (Aslandogan and Yu, 1999). Όμως παρά το γεγονός ότι το εύρος των εφαρμογών που θα μπορούσαν να προκύψουν από την εφαρμογή μεθόδων εντοπισμού και αναγνώρισης κειμένου σε αρχεία video είναι σημαντικό, το συγκεκριμένο πεδίο θεωρείται ιδιαίτερα απαιτητικό και ο αριθμός των σχετικών εφαρμογών είναι ακόμη αρκετά περιορισμένος.

Γι' αυτό το λόγο, ο εντοπισμός και η αναγνώριση κειμένου σε ψηφιακά αρχεία video συνιστά ενδιαφέρον πεδίο έρευνας διότι παρουσιάζει μεγαλύτερες δυσκολίες από την αναγνώριση κειμένου σε ψηφιακά αρχεία εικόνας. Αυτό συμβαίνει επειδή στα αρχεία video, τα καρέ (frames), έχουν συνήθως πολύ χαμηλότερη ανάλυση αλλά και αντίθεση σε σχέση με το φόντο, επίσης παρατηρείται συχνά το φαινόμενο της διάχυσης χρώματος (*color bleeding*), ένα φαινόμενο κατά το οποίο αντικείμενα χρωματίζονται μέσω ανάκλασης από γειτονικές επιφάνειες, επιπρόσθετα το φόντο είναι κατά κανόνα ιδιαίτερα περίπλοκο, ενώ το χρώμα, η θέση η διεύθυνση και η διάταξη του κειμένου είναι μεταβλητή και άγνωστη. Παρά τις προαναφερθείσες δυσχέρειες, υπάρχει το πολύ σημαντικό πλεονέκτημα του χρονικού πλεονασμού (*temporal redundancy*), αφού η πληροφορία κειμένου παραμένει για αρκετό διάστημα, ώστε να γίνει αντιληπτή από τον άνθρωπο. Η χρονική διάσταση της πληροφορίας κειμένου σε ψηφιακό σήμα video, άρχισε να αξιοποιείται επαρκώς μόλις πρόσφατα, γεγονός που συνεισέφερε σημαντικά στην πρόοδο που παρατηρείται στο συγκεκριμένο πεδίο. Η διαδικασία εξαγωγής κειμένου από αρχεία βίντεο, μπορεί να αναλυθεί σε πέντε διακριτά στάδια:

1. Ανίχνευση κειμένου (Text Detection)
2. Χωρικός εντοπισμός κειμένου (Text Localization)
3. Ιχνηλάτηση κειμένου (Text Tracking)
4. Δυαδική κωδικοποίηση κειμένου (Text Binarization)
5. Οπτική αναγνώριση κειμένου (Optical Character Recognition - OCR)



Η ύπαρξη ή μη κειμένου στα επιμέρους καρέ ενός αρχείου βίντεο, δεν είναι γνωστή εκ των προτέρων, σε αντίθεση με τα μεμονωμένα αρχεία εικόνας, όπου η ύπαρξη κειμένου γίνεται άμεσα αντιληπτή. Κρίνεται λοιπόν απαραίτητο, ένα στάδιο επεξεργασίας, κατά το οποίο θα εντοπίζονται τα καρέ που είναι υποψήφια για περαιτέρω ανάλυση. Έχουν αναπτυχθεί διάφορες τεχνικές για την ανίχνευση καρέ με πληροφορία κειμένου, ακολουθούν ενδεικτικά σύντομες αναφορές:

1. Οι **Smith και Kanade, (1995)**, αντιμετωπίζουν τις περιοχές κειμένου ως πυκνές συστάδες ακμών (λόγω της κατά κανόνα μεγάλης αντίθεσής τους με το υπόβαθρο), οι οποίες με τη σειρά τους εμπεριέχονται σε περιοχές ορθογώνιου παραλληλόγραμμου σχήματος. Οι συγγραφείς προτείνουν την ανάδειξη των κατακόρυφων ακμών, την ομαλοποίηση τους ώστε να συνδεθούν τυχόν κατακερματισμένες περιοχές και την δημιουργία χαρακτηριστικών (bounding boxes) μέσω ανίχνευσης συστάδων (cluster detection) στο χάρτη ακμών. Τα χαρακτηριστικά τα οποία τηρούν συγκεκριμένες προϋποθέσεις αναλογίας, μεγέθους και πυκνότητας πληροφορίας, χαρακτηρίζονται ως περιοχές κειμένου.
2. Ο **Lim et Al., (2000)**, προτείνουν μια απλή και γρήγορη μέθοδο, η οποία βασίζεται στην υψηλότερη τιμή φωτεινής έντασης (*intensity*) που συνήθως χαρακτηρίζει τις περιοχές κειμένου. Σύμφωνα με την προσέγγισή τους, τα καρέ τα οποία περιέχουν ικανό αριθμό εικονοστοιχείων (*pixels*) με τιμές φωτεινότητας μεγαλύτερες από μια τιμή κατωφλίου, ενώ παράλληλα διαφοροποιούνται χρωματικά επαρκώς σε σχέση με τα γειτονικά τους, χαρακτηρίζονται ως καρέ τα οποία περιέχουν πληροφορία κειμένου.

Χωρικός εντοπισμός κειμένου (Text Localization):

Στο συγκεκριμένο βήμα, πραγματοποιείται ομαδοποίηση των επιμέρους περιοχών κειμένου, τα οποία ακολούθως περιβάλλονται από ένα σύνολο χαρακτηριστικών (bounded boxes), ώστε να οριστούν επακριβώς τα όριά τους. Οι διαφορές των δύο πρώτων σταδίων είναι δυσδιάκριτες και μπορούμε να θεωρήσουμε ότι μια εφαρμογή χωρικού εντοπισμού κειμένου δύναται να πραγματοποιεί και την ανίχνευσή του κειμένου αρκεί αυτή να μπορεί να γίνει σε πραγματικό χρόνο. Ο χωρικός εντοπισμός του κειμένου υλοποιείται μέσω δύο βασικών μεθόδων:

Μέθοδος Περιοχής

Η μέθοδος περιοχής (Region Based Approach), αναγνωρίζει τις περιοχές κειμένου μέσω χαρακτηριστικών τους, που τις διαφοροποιούν από το φόντο, όπως το χρώμα και η φωτεινότητα. Η συγκεκριμένη μέθοδος θεωρείται ότι λειτουργεί «από κάτω προς τα επάνω» (Bottom-up), διότι αρχικά αναδεικνύει μικρές κατακερματισμένες περιοχές, τις οποίες εν συνεχεία συσσωματώνει, ώστε να ορίσει το τελικό σύνολο των επιθυμητών χαρακτηριστικών. Η μέθοδος περιοχής, με τη σειρά της, υλοποιείται μέσω δύο βασικών προσεγγίσεων, την προσέγγιση συνδεδεμένων δομικών στοιχείων (Connected-Components based methods) και την προσέγγιση εντοπισμού μέσω ακμών (edge-based methods).

Στην προσέγγιση συνδεδεμένων δομικών στοιχείων, οι μικρότερες περιοχές ομαδοποιούνται σε μεγαλύτερες μέσω γεωμετρικής ανάλυσης κατά την οποία αξιολογείται η χωρική διάταξή τους ώστε να απαλειφθούν οι περιοχές φόντου.

1. Οι **Subramanian et al., (2007)**, προτείνουν μια μέθοδο εντοπισμού κειμένου, η οποία στοχεύει στην ανίχνευση των αποτυπώσεων των χαρακτήρων (Character-Stroke Detection). Οι συγγραφείς αξιοποιούν το σταθερό πλάτος αποτύπωσης και την υψηλή αντίθεση των χαρακτήρων κειμένου σε σχέση με το φόντο. Προτείνουν μια μέθοδο τύπου (bottom-up), με τη χρήση της οποίας επιχειρούν να άρουν παραδοσιακές δυσκολίες, όπως η εύρεση ικανοποιητικής τιμής κατωφλίου και η ανίχνευση προτύπων κειμένου. Για να πετύχουν το στόχο τους εφαρμόζουν ένα ευρηματικό αλγόριθμο ο οποίος αναλύει τις τιμές φωτεινότητας των εικονοστοιχείων κατά μήκος οριζόντιων γραμμών, ώστε να ανιχνεύσει αποτυπώσεις κειμένου και εν συνεχεία να τις χωροθετήσει.
2. Οι **Mancas και Gosselin, (2006)**, προτείνουν μια μεθοδολογία, αποτελεσματική στην ύπαρξη ακανόνιστου φωτισμού, θάμβωσης και πολύπλοκου φόντου, η οποία συνδυάζει χωρική και χρωματική πληροφορία για την αναγνώριση φυσικά ενυπάρχοντος κειμένου σε πληροφορία βίντεο. Οι συγγραφείς προτείνουν έναν αλγόριθμο τμηματοποίησης χρώματος (color segmentation), βάσει της χωρικής πληροφορίας (spatial information). Χρησιμοποιούνται τεχνικές συσταδοποίησης βασιζόμενες στην Ευκλείδεια απόσταση (Euclidean distance) και την ομοιότητα βάσει συνημίτονου (Cosine similarity based clustering), οι οποίες εφαρμόζονται συμπληρωματικά στα τρία βασικά χρώματα RGB, ώστε να διαχωριστεί η πληροφορία εισόδου σε τρεις συστάδες: προσκήνιο κειμένου, παρασκήνιο κειμένου, θόρυβος. Για την τελική εξαγωγή πληροφορίας κειμένου, γίνεται χρήση φίλτρων τύπου Log-Gabor, τα οποία έχουν το πλεονέκτημα να συνδυάζουν τη χρωματική και τη χωρική πληροφορία.

Στην προσέγγιση εντοπισμού μέσω ακμών, αξιοποιείται η έντονη αντίθεση του κειμένου από το φόντο. Για την ανάδειξη των ακμών, χρησιμοποιείται συνήθως ένα φίλτρο ακμών (π.χ. Sobel), εν συνεχεία οι ακμές συγχωνεύονται είτε μέσω εξομάλυνσης (smoothing), είτε μέσω μορφολογικών διεργασιών (morphological operations: dilation, erosion, opening, closing), ενώ στο τέλος χρησιμοποιούνται ευρετικές μέθοδοι για τον αποκλεισμό των περιοχών φόντου. Ακολουθούν αναφορές και περιγραφή δημοσιεύσεων οι οποίες βασίζονται στην προσέγγιση εντοπισμού μέσω ακμών:

1. Οι **Dubey et al., (2006)**, προτείνουν μια πολυχρηστική εφαρμογή εντοπισμού κειμένου που βασίζεται στην ανάδειξη ακμών. Οι συγγραφείς προτείνουν τη χρήση πυρήνα Sobel για την ανάδειξη ακμών **μονάχα κατά την κατακόρυφη διεύθυνση**, θεωρώντας ότι ανταποκρίνεται καλύτερα στα ιδιαίτερα χαρακτηριστικά του κειμένου. Κατόπιν περιγράφεται μια καινοτόμος μεθοδολογία **μορφολογικής ανάλυσης** για τη

διασύνδεση των υποψηφίων περιοχών κειμένου, η οποία βασίζεται στη φωτεινή ένταση των εικονοστοιχείων, ενώ παράλληλα λαμβάνει υπόψη της γεωμετρικά χαρακτηριστικά, όπως η απόσταση μεταξύ των χαρακτήρων και το ύψος τους. Στο τέλος αναπτύσσεται μια τεχνική **κατάτμησης** (segmentation) και επαλήθευσης των υποψηφίων περιοχών, η οποία αξιοποιεί χαρακτηριστικά όπως η τιμή φωτεινότητας, η αλληλεπικάλυψη και η ταχύτητα μεταβολής της φωτεινής έντασης.



2. Οι **Lyu et al., (2005)**, προτείνουν μια διεξοδική μέθοδο ανίχνευσης, εντοπισμού και εξαγωγής πολύγλωσσου κειμένου από πληροφορία βίντεο. Δίδεται έμφαση στην ανάδειξη των ιδιαίτερων γνωρισμάτων των χαρακτήρων κειμένου σε σχέση με τον τύπο γραφής και στη μελέτη των διαφορετικών προσεγγίσεων που απαιτούνται για την επεξεργασία του. Οι συγγραφείς διαχωρίζουν τα χαρακτηριστικά του κειμένου, σε αυτά που είναι ανεξάρτητα της γλώσσας και εκείνα που εξαρτώνται από αυτή, αναλύοντας παράλληλα τις επιπτώσεις τους στην επεξεργασία του κειμένου, μέσω συγκριτικής μελέτης των λατινικών χαρακτήρων και των κινέζικων ιδεογραμμάτων.

Οι συγγραφείς προτείνουν μια μεθοδολογία που βασίζεται στη διαδοχική - πολυδιακριτική ανάλυση χάρτη ακμών (edge map) του σήματος εισόδου και την ακολουθιακή ανατροφοδότηση του αποτελέσματος στα επόμενα βήματα (sequential multiresolution analysis). Το σκεπτικό της πρότασης βασίζεται στην θεώρηση ότι οι περιοχές κειμένου που ανιχνεύονται σε κάθε βήμα θεωρούνται δεδομένες και δεν θα πρέπει να απασχολούν τον αλγόριθμο στα επόμενα βήματα.

Μέθοδος Υφής

Η **μέθοδος υφής (Texture Based Approach)**, βασίζεται στη διαπίστωση ότι το κείμενο στις εικόνες έχει διαφορετική «υπογραφή υφής» (textural signature) από ότι το φόντο. Η συγκεκριμένη μέθοδος θεωρείται ότι λειτουργεί «από επάνω προς τα κάτω» (**Top-down**), διότι αρχικά εξάγει χαρακτηριστικά υφής και κατόπιν οριοθετεί τις περιοχές κειμένου. Για την ανάδειξη των ειδικών χαρακτηριστικών της υφής του κειμένου, αξιοποιούνται μέθοδοι όπως η ταχεία ανάλυση κατά Fourier (FFT), η ανάλυση κυματίων (Wavelet Transform), τα φίλτρα Gabor, η χωρική απόκλιση (Spatial Variance), οι Ταξινομητές Πυρήνα (SVM) κ.α.

1. Οι **Gilavata et al., (2006)**, αντιμετωπίζουν το πρόβλημα, εφαρμόζοντας **σύνολα αλγόριθμων ασαφούς συσταδοποίησης** (fuzzy clustering ensembles - FCE), με σκοπό τη βελτίωση της ποιότητας των ανεξάρτητων συστάδων και την επαυξητική σύλληψη χρονικής πληροφορίας στατικού κειμένου σε βίντεο. Για ένα σύνολο διαδοχικών καρέ, εξάγονται διανύσματα χαρακτηριστικών (feature vectors) μέσω εφαρμογής μετασχηματισμού κυματίων στα επιμέρους καρέ, τα οποία με τη σειρά τους ομαδοποιούνται μέσω εφαρμογής αλγορίθμων ασαφούς συσταδοποίησης κεντροειδούς

μέσου (fuzzy c-means clustering - FCM). Κατόπιν, τα ανεξάρτητα αποτελέσματα της ομαδοποίησης, συγχωνεύονται σε τρεις κλάσεις με χρήση του αλγόριθμου FCE και οι χωρικές περιοχές που ανήκουν στην κλάση «κειμένου», αναδεικνύονται.

2. Οι **Aradhye και Myers, (2007)**, παρουσιάζουν μια μέθοδο μετα-επεξεργασίας (post-processing), η οποία βασίζεται αφενός σε έναν αλγόριθμο ανίχνευσης μεταβολών στην εμφάνιση περιοχών κειμένου, ο οποίος υλοποιείται μέσω πολλαπλών αναλύσεων κυματίων στον άξονα του χρόνου (Multiscale detection of videotext events) και αφετέρου σε έναν ακολουθιακό αλγόριθμο χωρικού κατακερματισμού στον κάθετο και οριζόντιο άξονα με χρήση συσταδοποίησης τύπου k-means.
3. Οι **Kim et al. (2007)**, συνδυάζουν τη χρήση μηχανών διανυσματικής υποστήριξης (SVM), για την επιλογή περιοχών κειμένου, με τη χρήση του αλγόριθμου συνεχούς προσαρμοζόμενης μετατόπισης μέσου (Continuously Adaptive Mean Shift Algorithm - CAMSHIFT) για την ανάλυση των επιλεγμένων περιοχών.
4. Οι **Jain και Zhong (1996)**, προτείνουν τη χρήση νευρωνικού δικτύου, για την εκπαίδευση και το διαχωρισμό της εικόνας εισόδου σε υποπεριοχές φόντου (background), μισού τόνου (half tone) καθώς και κειμένου μαζί με περιοχές γραμμικών γραφημάτων. Οι περιοχές κειμένου και γραμμικών γραφημάτων, διαχωρίζονται περαιτέρω στις συνιστώσες τους μέσω ανάλυσης συνδεσιμότητας. Οι συγγραφείς προτείνουν επίσης τη χρήση μεθοδολογίας για το διαχωρισμό πολύγλωσσου κειμένου στις συνιστώσες γλώσσες του.

Στην παρούσα εργασία θα επικεντρωθούμε στο στάδιο του χωρικού εντοπισμού, στατικού στο χρόνο, μηχανικά ενσωματωμένου κειμένου, καθώς και στα στάδια ενίσχυσης της ποιότητας, δαδοποίησης και οπτικής αναγνώρισής του.

Αρχικά παρουσιάζονται συνοπτικά οι υφιστάμενες τεχνικές αναγνώρισης φυσικού και μηχανικά ενσωματωμένου κειμένου σε ψηφιακό σήμα video. Στη συνέχεια παρατίθενται οι προκρινόμενες μεθοδολογίες (Wang και Chen, 2006; Burns et al., 1996). Η διαδικασία επεξεργασίας που προτείνεται στην παρούσα μελέτη, βασίζεται στην ανάλυση του πηγαίου σήματος, μέσω της εφαρμογής μετασχηματισμού κυματίων (*wavelet transform*). Σε πρώτη φάση αναλύεται συνοπτικά το θεωρητικό υπόβαθρο του διακριτού μετασχηματισμού κυματίου (*discrete wavelet transform*) μονοδιάστατου και δισδιάστατου (*ψηφιακή εικόνα*) σήματος. Στη συνέχεια πραγματοποιείται εκτενής αναφορά στην ανάλυση (spatio-temporal analysis) σήματος τριών διαστάσεων (video). Ιδιαίτερη έμφαση δίδεται στις καινοτόμες τεχνικές οι οποίες αναπτύχθηκαν προκειμένου να υλοποιηθεί πολυεπίπεδη και ανομοιογενής ανάλυση του σήματος στο πεδίο του χώρου και του χρόνου (Burns et al., 1996). Η συγκεκριμένη προσέγγιση κρίνεται ως ιδιαίτερως ενδιαφέρουσα, διότι σε αντίθεση με τη συμβατική ομοιογενή ανάλυση, αναγνωρίζει τη διακριτή φύση του σήματος στο πεδίο του χώρου και του χρόνου, επιτρέποντας την ασύμμετρη επεξεργασία του και αναδεικνύοντας με μεγαλύτερη ακρίβεια τα επιθυμητά χαρακτηριστικά του στον παραγόμενο χάρτη ακμών (saliency map).

Ακολούθως περιγράφονται λεπτομερώς οι αλγόριθμοι που επιλέχθηκαν για την ανάλυση του χάρτη ακμών. Θα παρουσιαστούν τεχνικές για τη δημιουργία διανύσματος χαρακτηριστικών με χρήση πίνακα συνάφειας (co-occurrence matrix GLCM), (Haralick et al, 1973; Fuan Tsai et al, 2007), στις τρεις διαστάσεις, θα υλοποιηθεί ταξινομητής (classifier) με χρήση Ταξινομητών Πυρήνα (SVM), (Gunn S.R., 1992; Theodoridis και Koutroumbas, 2009) για την αναγνώριση περιοχών κειμένου και θα εφαρμοστούν αλγόριθμοι μορφολογικής ανάλυσης για την τελική ανάδειξή του.

Στο τελευταίο Κεφάλαιο παρουσιάζεται το ολοκληρωμένο λογισμικό ανάλυσης κειμένου που αναπτύχθηκε στα πλαίσια της εργασίας και φέρει το όνομα 3D Wavelet Toolbox for Textual Information Extraction in Video Sequences – **Video W3D Toolbox**. Στο κεφάλαιο αυτό παρέχεται και σύντομο εγχειρίδιο χρήσης του λογισμικού. Το υπολογιστικό αυτό εργαλείο επιτρέπει την αξιολόγηση του συνόλου των προτεινόμενων μεθόδων με χρήση πραγματικών αρχείων video, γι' αυτό το λόγο θα παρατεθούν παραδείγματα και θα εξαχθούν συμπεράσματα για την αποτελεσματικότητα αλλά και τις αδυναμίες των προτεινόμενων μεθοδολογιών. Τέλος θα γίνουν προτάσεις για τη μελλοντικές επεκτάσεις και βελτιώσεις.

Εισαγωγή στην κυματιδιακή ανάλυση:

Ένα **κυματίδιο (wavelet)**, μπορεί να περιγραφεί ως μια ταλάντωση με πλάτος το οποίο ξεκινάει από το μηδέν εν συνεχεία αυξάνεται και κατόπιν ελαττώνεται έως ότου μηδενιστεί ξανά. Μπορεί να ειπωθεί ως μια «συνοπτική ταλάντωση», όπως μια καταγραφή σε σειсмоγράφο ή καρδιογράφο. Γενικά τα κυματίδια είναι ηθελημένα δομημένα, με τρόπο που να τους προσδίδει ιδιότητες τέτοιες, ώστε να είναι χρήσιμα στην επεξεργασία σήματος. Τα κυματίδια μπορούν να συνδυαστούν μέσω μια τεχνικής «μετατόπισης, πολλαπλασιασμού και αθροίσματος» η οποία ονομάζεται συνέλιξη (convolution), με τμήματα απροσδιόριστου σήματος ώστε να εξάγουν επιθυμητή πληροφορία.

Για παράδειγμα ένα κυματίδιο θα μπορούσε να οριστεί έχοντας συχνότητα ντο ματζόρε και διάρκεια ενός $32^{ου}$. Εάν χρησιμοποιούσαμε το συγκεκριμένο κυματίδιο σε περιοδική συνέλιξη με το σήμα ενός μουσικού κομματιού, τα αποτελέσματα των συνελίξεων θα ήταν χρήσιμα για τον προσδιορισμό των σημείων που η συγκεκριμένη νότα περιγράφει το μουσικό κομμάτι. Μαθηματικά, ένα κυματίδιο θα ανταποκριθεί εάν το άγνωστο προς ανάλυση σήμα περιέχει πληροφορία παρόμοιας συχνότητας, όπως ένα πιρούνι αντηχεί με ηχητικά κύματα που αντιστοιχούν στην ιδιοσυχνότητά του. Η έννοια της αντήχησης-ανταπόκρισης βρίσκεται στον πυρήνα πολλών πρακτικών εφαρμογών στη θεωρία κυματιδίων (Wavelet Theory).

Επειδή τα κυματίδια είναι ένα μαθηματικό εργαλείο μπορούν να χρησιμοποιηθούν για την εξαγωγή πληροφορίας από διαφόρων ειδών δεδομένα, συμπεριλαμβανομένων των σημάτων εικόνας και βίντεο. Για την πλήρη ανάλυση δεδομένων απαιτούνται συνήθως σύνολα κυματιδίων. Ένα σύνολο συμπληρωματικών κυματιδίων θα αποδομήσει τα δεδομένα χωρίς κενά ή αλληλοεπικάλυψη με τρόπο ώστε η αποδόμηση να είναι μαθηματικά αναστρέψιμη. Έτσι σύνολα συμπληρωματικών κυματιδίων είναι χρήσιμα σε αλγόριθμους συμπίεσης/αποσυμπίεσης, όπου είναι επιθυμητή η ανάκτηση της αρχικής πληροφορίας με τις ελάχιστον απώλειες.

Περισσότερο τεχνικά, ένα κυματίδιο είναι μια μαθηματική συνάρτηση η οποία χρησιμοποιείται για να διαιρέσει μια δεδομένη συνάρτηση ή συνεχές ως προς το χρόνο σήμα σε συνιστώσες διαφορετικής κλίμακας. Συνήθως κάποιος μπορεί να προσδώσει ένα εύρος συχνοτήτων σε κάθε τέτοια συνιστώσα. Κάθε συνιστώσα διαφορετικής κλίμακας μπορεί να μελετηθεί με ανάλυση η οποία θα ταιριάζει στην κλίμακά της. Ένας κυματιδιακός μετασχηματισμός (wavelet transform), αποτελεί την αναπαράσταση μιας συνάρτησης με χρήση κυματιδίων. Τα κυματίδια αυτά (daughter wavelets) είναι αντίγραφα υπό κλίμακα και μετατροπή, μιας πεπερασμένης και γοργά αποσβενόμενης κυματομορφής, γνωστής και ως (mother wavelet). Ο κυματιδιακός μετασχηματισμός πλεονεκτεί έναντι του παραδοσιακού μετασχηματισμού κατά Fourier, στην αναπαράσταση συναρτήσεων με ασυνέχειες και απότομες κορυφές, καθώς και στην ακριβή αποσύνθεση και ανασύνθεση πεπερασμένων, μη περιοδικών και μη στατικών σημάτων.

Ο μετασχηματισμός κυματιδίων, μπορεί να διαχωριστεί στον διακριτό μετασχηματισμό κυματιδίων (discrete wavelet transforms DWT) και στον συνεχή μετασχηματισμό κυματιδίων (continuous wavelets transforms). Οι δύο μετασχηματισμοί είναι συνεχείς ως προς τον χρόνο (αναλογικοί) και μπορούν να χρησιμοποιηθούν για την αναπαράσταση συνεχών ως προς τον χρόνο σημάτων. Η διαφορά μεταξύ τους είναι ότι ο συνεχής μετασχηματισμός εφαρμόζει κατά την ανάλυση τις τιμές που προκύπτουν υπό όλες τις δυνατές κλίμακες και μετατροπές του mother wavelet, ενώ ο διακριτός μετασχηματισμός χρησιμοποιεί ένα υποσύνολο των κλιμάκων και μετατροπών.

Ιστορική διαδρομή

Η θεωρία της κυματιδιακής ανάλυσης είναι μια αρκετά νέα μεθοδολογία, το μαθηματικό υπόβαθρό της όμως, είναι κατά πολύ παλαιότερο και βασίζεται στο έργο του Joseph Fourier κατά το 19^ο αιώνα, έργο το οποίο ανέδειξε θεωρίες ανάλυσης σήματος μέσω της συχνότητας και αποδείχθηκε ιδιαίτερα σημαντικό καθώς επηρέασε σημαντικά τη μετέπειτα έρευνα. Σχετικά πρόσφατα η έρευνα μετατέθηκε από την ανάλυση κατά συχνότητα στην ανάλυση κατά κλίμακα, διότι κατέστη σαφές ότι μια προσέγγιση εντοπισμού διακυμάνσεων σε διαφορετικές κλίμακες

είναι περισσότερο αποτελεσματική. Η πρώτη αναφορά σε έννοιες σχετικές με τα «κυματίδια» έγινε σε μια εργασία του Alfred Haar (1910). Αξιωματικότερες συνεισφορές στη θεωρία κυματιδίων αποτελούν, η ανακάλυψη του συνεχούς μετασχηματισμού κυματιδίων από τον Zweig (Zweig et al., 1975), (αρχικά αποκλήθηκε μετασχηματισμός cochlea- cochlear transform και ανακαλύφθηκε κατά τη μελέτη της αντίδρασης του βιολογικού συστήματος ακοής στον ήχο), η διαμόρφωση αυτού που αποκαλείται συνεχής μετασχηματισμός κυματιδίων από τους Pierre Gouillaud, Grossman και Morlet, (Morlet et al., 1982; Gouillaud et al., 1984; Grossman and Morlet, 1984). Οι συγγραφείς αρχικά χρησιμοποίησαν τον ορισμό *ondelette*, ο οποίος σημαίνει «μικρό κύμα» στα γαλλικά, ο ορισμός αυτός σύντομα καθιερώθηκε στα αγγλικά με τη σημερινή του ονομασία wavelets. Η εργασία του Jan-Olov Strömberg στον διακριτό μετασχηματισμό (Stromberg, 1983), η μελέτη περί ορθογωνικών κυματιδίων με συμπαγή υποστήριξη του Daubechies (1988), το πλαίσιο πολυδιακριτικής ανάλυσης του Mallat (1989), η ερμηνεία του συνεχούς μετασχηματισμού στα πεδία του χρόνου και της συχνότητας της Nathalie Delprat (Delprat et al., 1992), ο μετασχηματισμός αρμονικών κυματιδίων (Harmonic wavelet transform), (Newland, 1993) κ.α.

Εφαρμογές

Πολλές εφαρμογές χρησιμοποιούν την κυματιδική ανάλυση. Οι συνήθεις στόχοι αφορούν την απλοποίηση και τον καθαρισμό σημάτων και εικόνων, διαδικασίες οι οποίες αφορούν την αφαίρεση θορύβου και τη συμπίεση. Υπάρχουν πολλές εφαρμογές στην ωκεανογραφία, στη συμπίεση δακτυλικών αποτυπωμάτων, στη φαρμακευτική, στα ηλεκτροκαρδιογραφήματα, τα εγκεφαλογραφήματα κ.α.

Πολλά πεδία χρησιμοποιούν τεχνικές κυματιδικής ανάλυσης για την μελέτη κανονικοτήτων:

- Η βιολογία για την αναγνώριση της μεμβράνης των κυττάρων, ώστε να διαχωρίσει τα φυσιολογικά από τα παθολογικά άτομα.
- Η μεταλλουργία για την ταξινόμηση τραχέων περιοχών.
- Η Οικονομική Επιστήμη, για την ανίχνευση ιδιοτήτων στις διακυμάνσεις των τιμών.
- Η περιγραφή της κυκλοφορίας δεδομένων στο διαδίκτυο, για τον σχεδιασμό διαδικτυακών υπηρεσιών.

Εκδοχές της κυματιδικής ανάλυσης στις τρεις διαστάσεις, έχουν αξιοποιηθεί πιο πρόσφατα, ακολουθούν μερικές από αυτές:

- αλγόριθμοι συμπίεσης βίντεο,
- εφαρμογές τηλεδιάσκεψης
- βιντεοτηλέφωνα
- τηλεόραση υψηλής ευκρίνειας κ.α.

Η κυματιδική ανάλυση μπορεί να ανιχνεύσει την ύπαρξη πολύ σύντομων σημάτων και να τα εντοπίσει. Γι' αυτό, όσον αφορά στην ανάλυση κατά τη διάσταση του χρόνου κύριοι στόχοι είναι ακόλουθοι:

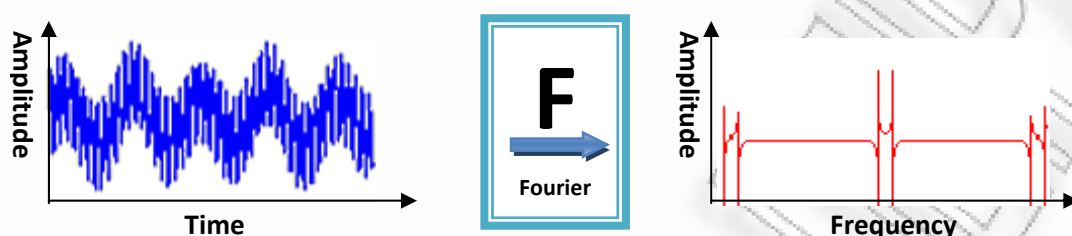
- Αναγνώριση ριζων η ακμών.
- Μελέτη πολύ σύντομων φαινομένων μεταβατικών διαδικασιών.

Ακολουθούν μερικές πρακτικές εφαρμογές:

- Βιομηχανική επίβλεψη οδοντοτροχών.
- Έλεγχος σύντομων παθολογικών γεγονότων, όπως κρίσεις επιληπτικές και μη.
- Απεικόνιση SAR.
- Αυτόματη αναγνώριση στόχου.

Ανάλυση κατά Fourier

Το περισσότερο γνωστό εργαλείο για την ανάλυση σημάτων είναι η ανάλυση κατά Fourier, η οποία διασπά το σήμα σε συστατικά ημιτονοειδή σήματα διαφόρων συχνοτήτων. Η ανάλυση κατά Fourier είναι μια μαθηματική μέθοδος μετατροπής της θέασης ενός σήματος από θέαση βάσει χρόνου σε θέαση βάσει συχνότητας.

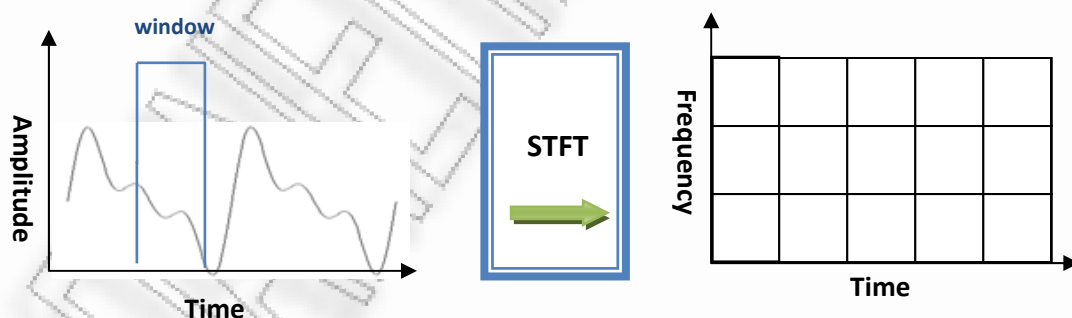


Για πολλά σήματα η ανάλυση κατά Fourier είναι εξαιρετικά χρήσιμη διότι η συχνοτική πληροφορία είναι συνήθως μεγάλης σημασίας.

Η ανάλυση κατά Fourier έχει ένα σημαντικό μειονέκτημα. Με τη μετατροπή δεδομένου σήματος στο πεδίο των συχνοτήτων, η χρονική πληροφορία χάνεται. Εάν προσπαθήσουμε να αναλύσουμε ένα μετασχηματισμένο κατά Fourier σήμα, είναι αδύνατο να προσδιορίσουμε χρονικά οποιοδήποτε γεγονός σε αυτό. Εάν το εξεταζόμενο σήμα είναι στατικό ως προς το χρόνο (stationary signal), το γεγονός αυτό δεν είναι τόσο σημαντικό. Παρόλα αυτά τα περισσότερο ενδιαφέροντα σήματα περιέχουν αριθμό μη στατικών ή μεταβατικών χαρακτηριστικών, όπως, μετατόπιση, τάσεις, απότομες μεταβολές κ.α. Τα χαρακτηριστικά αυτά αποτελούν το περισσότερο σημαντικό μέρος τέτοιων σημάτων και η ανάλυση κατά Fourier δεν είναι το ενδεδειγμένο εργαλείο για την ανάδειξή τους.

Short-Time Fourier Analysis

Σε μια προσπάθεια να διορθωθεί αυτή η ανεπάρκεια, ο Dennis Gabor (1946), προσαρμόσε τον αλγόριθμο μετατροπής κατά Fourier, ώστε να αναλύει μονάχα μικρές περιοχές του σήματος κάθε φορά, εφαρμόζοντας την τεχνική του κινούμενου παραθύρου (signal windowing). Η προσαρμογή που πρότεινε ο Gabor, ονομάστηκε Ταχύς Μετασχηματισμός κατά Fourier (Short-Time Fourier Transform - STFT), (Cohen, 1995) και αποτυπώνει δεδομένο σήμα ως συνάρτηση δύο διαστάσεων στο χρόνο και τη συχνότητα:



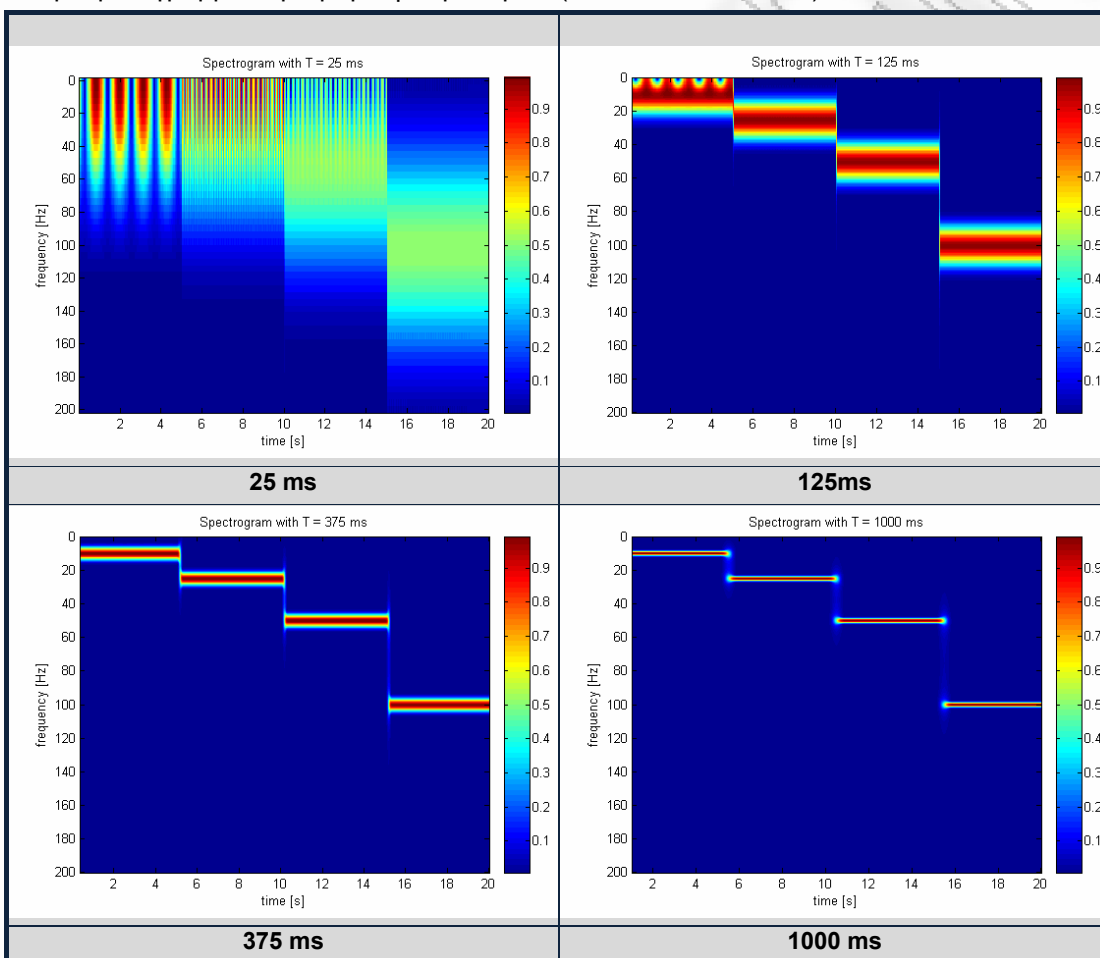
Η τεχνική STFT παρουσιάζει ένα είδος συγκερασμού μεταξύ της θέσης ενός σήματος βάσει του χρόνου και βάσει συχνότητας. Παρέχει πληροφορία για τις συχνότητες και το χρόνο που εμφανίζεται κάποιο γεγονός. Παρόλα αυτά η συγκεκριμένη πληροφορία παρέχεται με περιορισμένη ακρίβεια, η οποία εξαρτάται από το επιλεγόμενο μέγεθος παραθύρου. Το μειονέκτημα εντοπίζεται στο ότι το μέγεθος του χρονοπαραθύρου που επιλέγεται είναι το ίδιο για όλες τις συχνότητες. Πολλά σήματα απαιτούν μια ποιο ευέλικτη προσέγγιση, όπου το μέγεθος

του παραθύρου θα είναι μεταβλητό, ώστε να μπορεί να προσδιοριστεί με μεγαλύτερη ακρίβεια ο χρόνος ή η συχνότητα.

Εάν για παράδειγμα δημιουργήσουμε ένα σήμα $x(t)$, το οποίο αποτελείται από τέσσερις κυματομορφές σε σειρά, οι οποίες έχουν συχνότητες (10,25,50,100 Hz):

$$x(t) = \begin{cases} \cos(2\pi 10t); & 0 \leq t < 5s \\ \cos(2\pi 25t); & 5 \leq t < 10s \\ \cos(2\pi 50t); & 10 \leq t < 15s \\ \cos(2\pi 100t); & 15 \leq t < 20s \end{cases}$$

Τα φασματογράμματα για μεγέθη παραθύρων (25,125,375,1000 ms) είναι τα ακόλουθα:



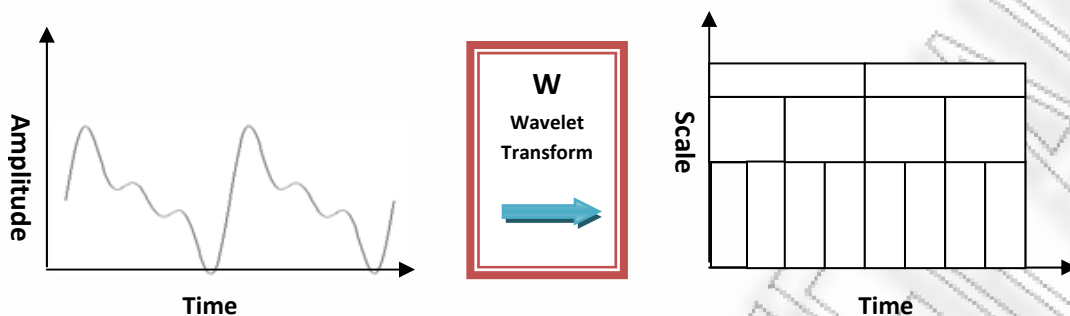
Παρατηρούμε ότι για μέγεθος παραθύρου 25 ms, μπορούμε να προσδιορίσουμε επακριβώς τα χρονικά σημεία στα οποία μεταβάλλεται το σήμα, αντιθέτως ο προσδιορισμός των συχνοτήτων δεν μπορεί να γίνει με ακρίβεια. Για μέγεθος παραθύρου 1000ms παρατηρούμε το αντίστροφο, καθώς οι συχνότητες διακρίνονται επακριβώς, ενώ η χρονική ανάλυση του σήματος είναι δυσχερής.

Κυματιδιακή ανάλυση

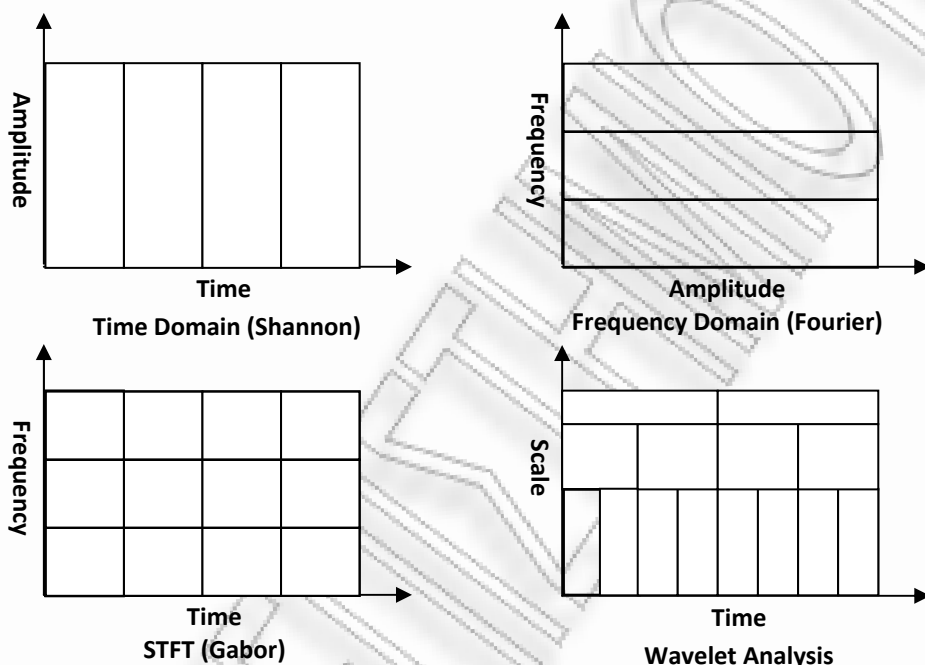
Η κυματιδιακή ανάλυση καλύπτει τα προαναφερθέντα μειονεκτήματα, αποτελώντας το επόμενο λογικό βήμα στην επεξεργασία σήματος: μια παραθυρική τεχνική η οποία χρησιμοποιεί μεταβλητού μεγέθους περιοχές. Η κυματιδιακή ανάλυση επιτρέπει τη χρήση μεγαλύτερων

χρονικών διαστημάτων όπου είναι επιθυμητή η ακριβής ανάλυση πληροφορίας περιοχών χαμηλών συχνοτήτων και βραχύτερων χρονικών περιοχών για τις υψηλές συχνοότητες.

ΓΑΛΕΡΙΟ ΤΗΛΩΟ ΓΕΡΑΝ



ακολουθως παρουσιάζεται διάγραμμα της κυματιδικής ανάλυσης σε αντιπαραβολή με τις θεάσεις του σήματος βάσει χρόνου (Shannon, 1949), συχνότητας (Fourier) και ταχέως μετασχηματισμού κατά Fourier (STFT-Gabor):

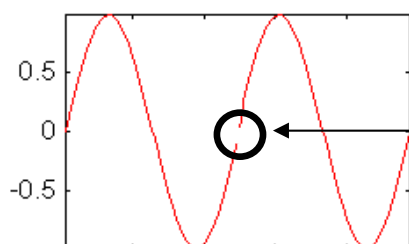


Η κυματιδική ανάλυση δεν αξιοποιεί τις διαστάσεις χρόνου-συχνότητας, αλλά τις διαστάσεις χρόνου-κλίμακας (time-scale).

Κυματιδική ανάλυση - πλεονεκτήματα

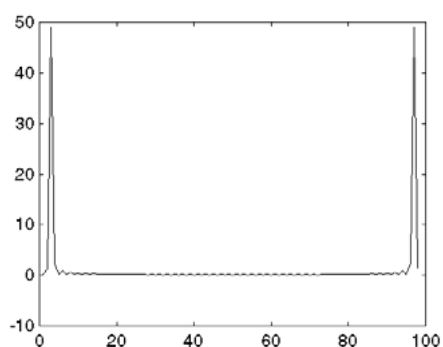
Το σημαντικότερο πλεονέκτημα της κυματιδικής ανάλυσης είναι η δυνατότητά της να αναλύει τοπικά περιοχές μεγαλύτερων σημάτων.

Εάν θεωρήσουμε ένα ημιτονοειδές σήμα το οποίο διαθέτει μια πολύ μικρή ασυνέχεια, κάτι το οποίο θα ήταν δυνατόν να συμβεί εύκολα στην πράξη από μια διακύμανση ισχύος ή από παρουσία θορύβου σε κάποιο διακόπτη.

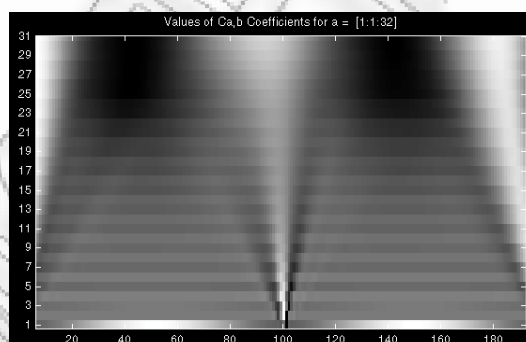


Ημιτονοειδές σήμα με μικρή ασυνέχεια

Εάν εφαρμόσουμε μετασχηματισμό κατά Fourier και παρουσιάσουμε το φασματογράφημα των συντελεστών της, παρατηρούμε μονάχα δυο κορυφές οι οποίες αναπαριστούν το σήμα στους άξονες πλάτους συχνότητας. Αντίθετα, εάν εφαρμόσουμε κυματιδιακή ανάλυση και απεικονίσουμε το διάγραμμα των συντελεστών της, διαπιστώνουμε ότι εμφανίζεται ξεκάθαρα η θέση της ασυνέχειας στον χρόνο.



Συντελεστές ανάλυσης Fourier



Συντελεστές κυματιδιακής ανάλυσης

Η ανάλυση κυματιδίων είναι δυνατό να αναδείξει χαρακτηριστικά σημάτων, τα οποία άλλες τεχνικές αποτυγχάνουν να αναδείξουν, όπως τάσεις, καταρρεύσεις, ασυνέχειες και ταυτότητες. Λόγω της διαφορετικής θέασης των δεδομένων, επιτρέπει την συμπίεση και την αφαίρεση θορύβου από σήματα με τη μικρότερη δυνατή υποβάθμιση.

Κυματιδιακή ανάλυση - ορισμός

Ορίζεται ως κυματίδιο (wavelet) μια κυματομορφή δραστικά περιορισμένης διάρκειας η οποία έχει μηδενική μέση τιμή.

Εάν συγκρίνουμε ένα κυματίδιο με ένα ημιτονοειδές κύμα, το οποίο αποτελεί τη βάση της ανάλυσης κατά Fourier, παρατηρούμε ότι το ημιτονοειδές κύμα έχει άπειρη διάρκεια αλλά και ομαλή-προβλέψιμη μορφή σε αντίθεση με ένα κυματίδιο το οποίο είναι ακανόνιστο και ασύμμετρο. Η ανάλυση κατά Fourier απαιτεί την αποσύνθεση ενός σήματος σε ημιτονοειδείς συνιστώσες μεταβλητής συχνότητας, ενώ η κυματιδιακή ανάλυση απαιτεί την αποσύνθεση του αρχικού σήματος σε σειρά συνιστώσων που ορίζονται από αντίγραφα του αρχικού (mother) κυματιδίου υπό κλίμακα και μετάθεση.

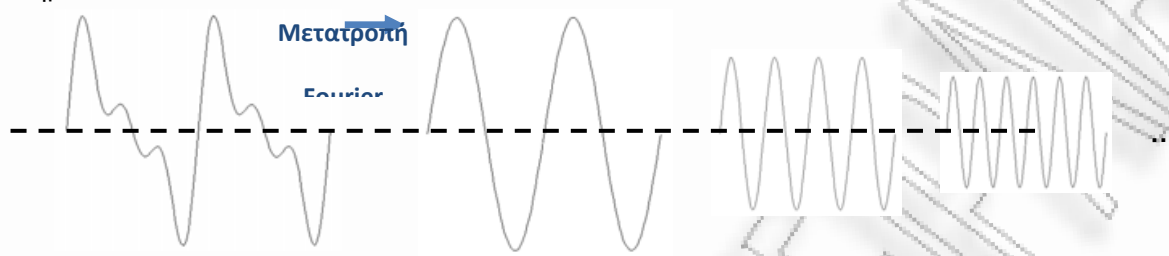
Συνεχής μετασχηματισμός κυματιδίων – Continuous wavelet transform

Η διαδικασία της ανάλυσης κατά Fourier αναπαρίσταται μαθηματικά μέσω της ακόλουθης συνάρτησης:

$$F(\omega) = \int_{-\infty}^{\infty} f(t) e^{-i\omega t} dt$$

Η ανάλυση ορίζεται ως το άθροισμα καθ' όλη τη χρονική διάρκεια, σήματος $f(t)$, πολλαπλασιαζόμενου με ένα μιγαδικό εκθετικό, ο οποίος αποδεικνύεται ότι μπορεί να διασπαστεί σε αριθμό πραγματικών και φανταστικών ημιτονοειδών συνιστώσων.

Ως αποτέλεσμα του μετασχηματισμού, προκύπτουν οι συντελεστές κατά Fourier (Fourier Coefficients - $F(\omega)$), οι οποίοι όταν με τη σειρά τους πολλαπλασιαστούν με ημιτονοειδές σήμα συχνότητας ω , αποδίδουν ξανά τις ημιτονοειδείς συνιστώσες που απαρτίζουν το αυθεντικό σήμα.



Αυθεντικό Σήμα

Ημιτονοειδείς συνιστώσες διαφόρων συχνοτήτων

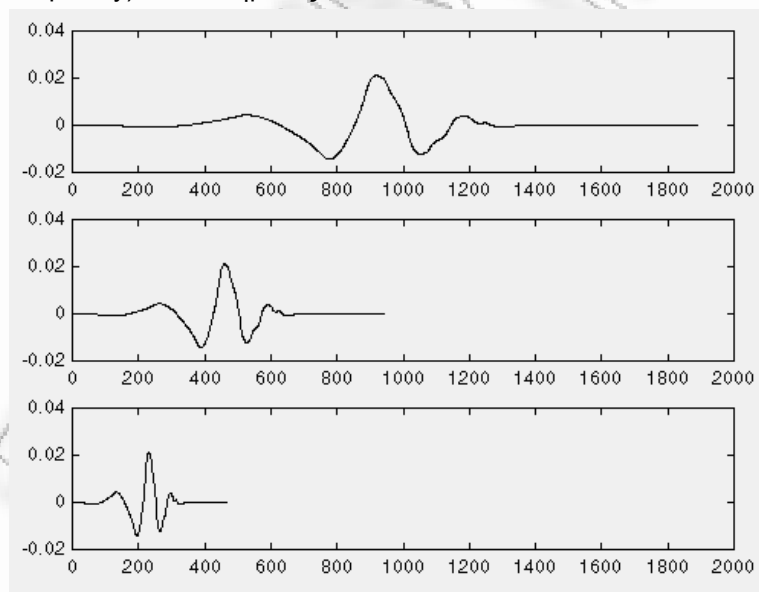
Αντίστοιχα ο συνεχής μετασχηματισμός κυματιδίων (CWT), ορίζεται ως το άθροισμα καθ' όλη τη διάρκεια του σήματος, πολλαπλασιαζόμενου με αντίγραφα της συνάρτησης κυματιδίου υπό κλίμακα και μετάθεση ψ :

$$C(\text{scale, position}) = \int_{-\infty}^{\infty} f(t)\psi(\text{scale, position, } t)dt$$

Το αποτέλεσμα της ανάλυσης CWT, είναι ένα πλήθος κυματιδιακών συντελεστών (wavelet coefficients), οι οποίοι ορίζουν μια συνάρτηση θέσης-κλίμακας. Πολλαπλασιάζοντας κάθε συντελεστή με την κατάλληλη έκδοση του μητρικού κυματιδίου υπό κλίμακα και μετατόπιση προκύπτουν τα συστατικά κυματίδια του αρχικού σήματος.

Διακλιμάκωση

Με την έννοια «κλιμάκωση κυματιδίου» εννοούμε την έκταση ή συμπίεση δεδομένου κυματιδίου στο πλαίσιο της κυματιδιακής ανάλυσης. Εισάγουμε τον όρο του «συντελεστή διακλιμάκωσης – scale factor» τον οποίο συμβολίζουμε ως a . Όσο μικρότερος είναι ο συντελεστής διακλιμάκωσης τόσο περισσότερο «συμπίεσμένο» το δεδομένο κυματίδιο. Επίσης ο συντελεστής διακλιμάκωσης a , σχετίζεται αντιστρόφως με την κυκλική συχνότητα (radial frequency) ω του σήματος.



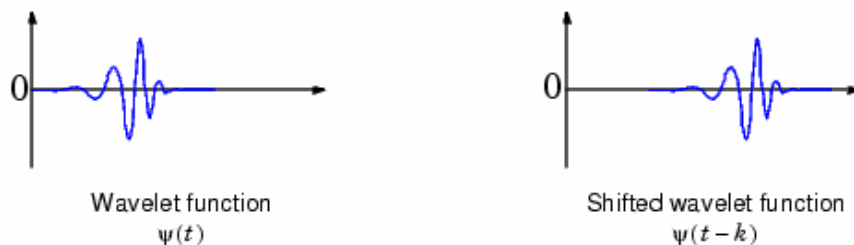
$$f(t) = \psi(t) \quad ; \quad a = 1$$

$$f(t) = \psi(2t) \quad ; \quad a = \frac{1}{2}$$

$$f(t) = \psi(4t) \quad ; \quad a = \frac{1}{4}$$

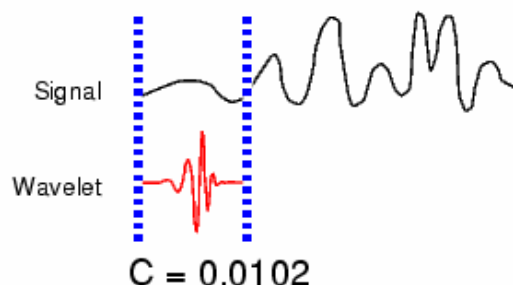
Μετάθεση

Με την έννοια «μετάθεση κυματιδίου» εννοούμε την καθυστέρηση ή επίσπευση της απαρχής δεδομένου κυματιδίου, μαθηματικά η καθυστέρηση μιας συνάρτησης $f(t)$ κατά k , αναπαρίσταται ως $f(t - k)$.

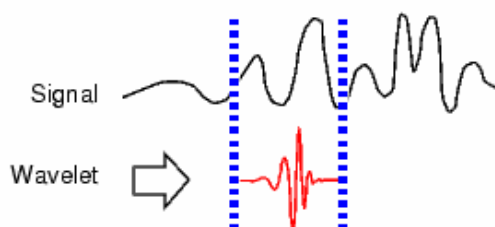


Πρακτικά η διαδικασία της κυματιδιακής ανάλυσης

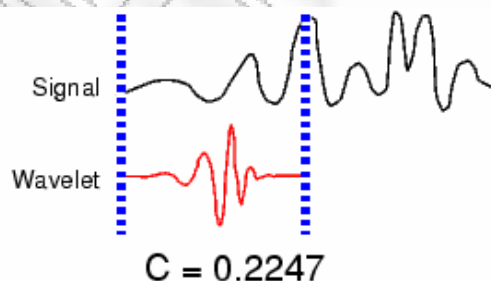
- τοποθετεί το πρωτογενές κυματίδιο στην αρχή του σήματος
- υπολογίζει μια τιμή συσχέτισης C , η οποία αναπαριστά το βαθμό ομοιότητας του κυματιδίου με τη συγκρινόμενη περιοχή του σήματος



- το κυματίδιο μετατοπίζεται σταδιακά καθ' όλη τη διάρκεια του σήματος και καταγράφονται οι τιμές συσχέτισης



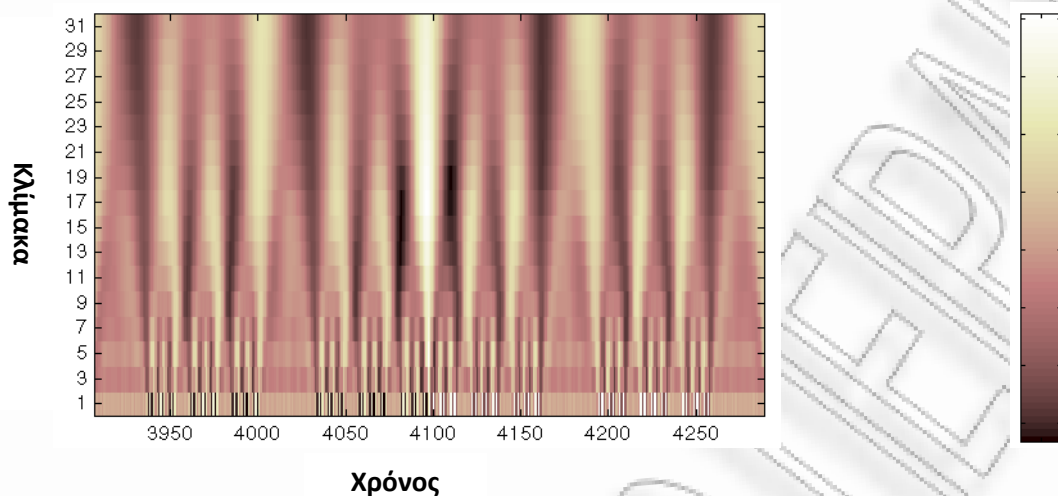
- το κυματίδιο διαστέλλεται σε κάθε βήμα και επαναλαμβάνονται οι προηγούμενες διαδικασίες.



Για την οπτικοποίηση των συντελεστών που παρήχθησαν σε διαφορετικές κλίμακες και θέσεις προτείνεται το διάγραμμα διακλιμάκωσης (scaleogram) όπου:

- στον άξονα x αναπαρίσταται θέση στο σήμα (χρόνος).
- Στον άξονα y αναπαρίσταται η κλίμακα (scale).

- Με χρωματική κλίμακα στις αντίστοιχες θέσεις x-y, αναπαρίστανται το μέτρο των συντελεστών κυματιδίων C.



Ο συντελεστής διακλιμάκωσης λαμβάνει τιμές από 1 έως 31.

- Χαμηλός συντελεστής a = συμπιεσμένο κυματίδιο = ταχέως μεταβαλλόμενες λεπτομέρειες = υψηλή συχνότητα ω
- Υψηλός συντελεστής a = διεσταλμένο κυματίδιο = αργά μεταβαλλόμενα αδρά χαρακτηριστικά = χαμηλή συχνότητα ω

Ο συνεχής μετασχηματισμός κυματιδίων διαφέρει από τον διακριτό στο ότι μπορεί να χρησιμοποιηθεί οποιαδήποτε τιμή διακλιμάκωσης, το εύρος τους καθορίζεται από τον χρήστη ανάλογα με τις ανάγκες και τη διαθέσιμη επεξεργαστική ισχύ. Επίσης η μετάθεση του κυματιδίου ανάλυσης είναι ομαλή και πραγματοποιείται σε όλη την έκταση του σήματος.

Διακριτός μετασχηματισμός κυματιδίων – Discrete wavelet transform

Η διαδικασία υπολογισμού συντελεστών για κάθε δυνατή κλίμακα απαιτεί ιδιαίτερα μεγάλη επεξεργαστική ισχύ και παράγει πλεόνασμα δεδομένων. Η λογική πίσω από τον διακριτό μετασχηματισμό, είναι η επιλογή μονάχα ενός υποσυνόλου κλιμάκων και θέσεων για την εκτέλεση υπολογισμών. Η διακριτή ανάλυση κυματιδίων υλοποιείται μέσω φίλτρων, ενώ το βήμα επιλογής διαδοχικών κλιμάκων και θέσεων βασίζεται σε τετραγωνικές δυνάμεις (dyadic scales-positions), παραδοχή η οποία εξασφαλίζει ακριβή και ικανοποιητικά αποτελέσματα. Η συγκεκριμένη προσέγγιση διατυπώθηκε αρχικά από τον Mallat, (1989).

Εφαρμογές της Διακριτής Ανάλυσης κυματιδίων

Γενικά, η προσέγγιση του διακριτού μετασχηματισμού (DWT), χρησιμοποιείται για συμπύεση ψηφιακών δεδομένων, εάν το σήμα έχει υποστεί δειγματοληψία, ενώ ο συνεχής μετασχηματισμός (CWT), χρησιμοποιείται ως επί το πλείστον για ανάλυση σημάτων. Γι' αυτούς τους λόγους η προσέγγιση DWT χρησιμοποιείται κυρίως στη μηχανική και την επιστήμη των υπολογιστών, ενώ η CWT στην επιστημονική έρευνα.

Η κυματιδιακή ανάλυση έχει προσαρμοστεί στις ανάγκες μεγάλου αριθμού εφαρμογών, συχνά αντικαθιστώντας τη συμβατική μετατροπή κατά Fourier. Η τελευταία διαπίστωση αφορά πολλές περιοχές της Φυσικής επιστήμης, όπως η μοριακή δυναμική (molecular dynamics), οι εξ υπαρχής (ab initio) υπολογισμοί, η αστροφυσική, ο χωρικός εντοπισμός μήτρας πυκνότητας (density-matrix), σεισμολογία-γεωφυσική, οπτική, τύρβωση (turbulence) και κβαντική μηχανική (quantum mechanics). Αφορά επίσης την επεξεργασία εικόνας, πίεση αίματος, ηλεκτροκαρδιογραφήματα, ανάλυση γωνιώματος, ανάλυση πρωτεϊνών, κλιματολογία, επεξεργασία σήματος εν γένει, αναγνώριση φωνής, γραφικά υπολογιστών και η πολυκλασματική ανάλυση (multifractal analysis). Στα πεδία της τεχνητής όρασης και της

επεξεργασίας εικόνας, οι έννοιες της αναπαράστασης υπό κλίμακα στο χώρο (space-scale representation) και της εξαγωγής τελεστών παραγώγων κατά Gauss (Gaussian Derivative Operators), εκτιμώνται ως τυπικές πολυδιακριτικές αναπαραστάσεις.

Μια συνήθης χρήση της προσέγγισης κυματιδίων είναι η συμπίεση δεδομένων. Η κυματιδιακή ανάλυση μπορεί να χρησιμοποιηθεί ώστε να μετασχηματίσει δεδομένα, τα οποία στην συνέχεια κωδικοποιούνται, με αποτέλεσμα την αποτελεσματική συμπίεση της αρχικής πληροφορίας. Για παράδειγμα, ο αλγόριθμος JPEG 2000, είναι ένα πρότυπο συμπίεσης εικόνας που χρησιμοποιεί διορθογωνικά κυματίδια. Αυτό σημαίνει ότι παρότι το πλαίσιο είναι υπερπλήρες, είναι ένα *tight frame*, επομένως οι ίδιες συναρτήσεις πλαισίου μπορούν να χρησιμοποιηθούν για την ανάλυση και την σύνθεση του αρχικού σήματος.

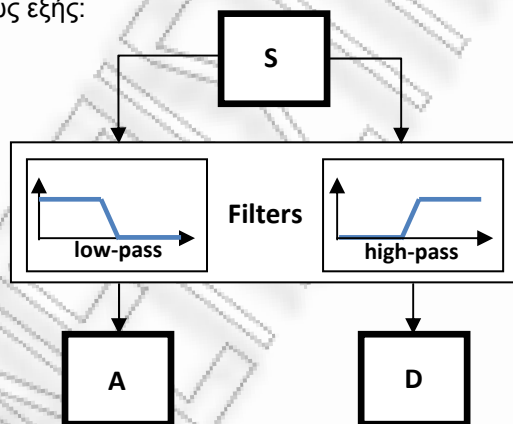
Μια συναφής χρήση είναι αυτή της εξομάλυνσης/αφαίρεσης θορύβου από δεδομένα μέσω εφαρμογής τιμών κατωφλίου στους συντελεστές κυματιδίων, διαδικασία η οποία αποκαλείται συρρίκνωση κυματιδίων (wavelet shrinkage). Εφαρμόζοντας προσαρμοστικά τιμές κατωφλίου στους συντελεστές κυματιδίων, οι οποίες αντιστοιχούν σε ανεπιθύμητες συνιστώσες συχνότητας, μπορούμε να επιτύχουμε εξομάλυνση, ή μείωση του θορύβου της αρχικής πληροφορίας.

Ο μετασχηματισμός κυματιδίων, έχει αρχίσει να χρησιμοποιείται σε εφαρμογές επικοινωνιών όπως η διάταξη διαμόρφωσης (Wavelet OFDM) η οποία χρησιμοποιείται στην τεχνολογία HD-PLC (τεχνολογία που αναπτύχθηκε από την PANASONIC και αποσκοπεί στη διαμόρφωση σήματος Ethernet στο οικιακό δίκτυο διανομής ηλεκτρισμού - powerline). Η συγκεκριμένη τεχνολογία πλεονεκτεί της προσέγγισης ταχέως μετασχηματισμού κατά Fourier (FFT), διότι επιτυγχάνει καλύτερο διαχωρισμό του σήματος και απαιτεί μικρότερο επίφορτο (overhead).

Εφαρμογή ψηφιακών φίλτρων σε ένα στάδιο: Προσέγγιση και Λεπτομέρειες

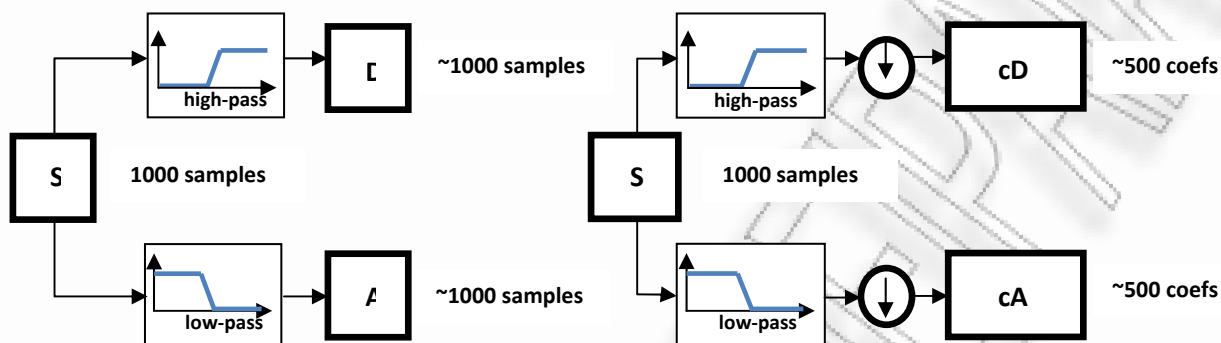
Στην θεωρία της κυματιδιακής ανάλυσης αναφερόμαστε συχνά στις έννοιες προσέγγιση (approximations) και λεπτομέρειες (details). Οι προσεγγίσεις ορίζονται ως οι υψηλής κλίμακας, χαμηλής συχνότητας συνιστώσες του σήματος. Οι λεπτομέρειες ορίζονται ως οι χαμηλής κλίμακας, υψηλής συχνότητας συνιστώσες.

Η βασική διαδικασία διέλευσης μέσα από ψηφιακά φίλτρα για τις ανάγκες κυματιδιακής ανάλυσης έχει ως εξής:

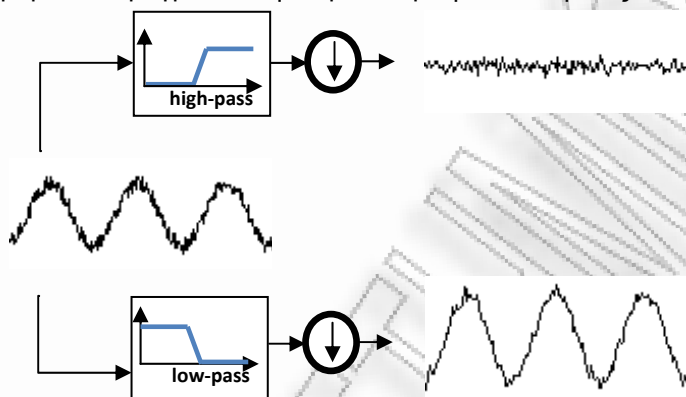


Το αρχικό σήμα S διέρχεται μέσω των δύο συμπληρωματικών φίλτρων (low-pass, high-pass) και διαχωρίζεται σε δύο σήματα (approximation-details).

Παρατηρούμε ότι η πληροφορία που λαμβάνουμε έχει διπλάσιο μέγεθος από την αρχική, μπορούμε να άρουμε το συγκεκριμένο πλεονασμό, διατηρώντας την πλήρη αναπαράσταση της αρχικής πληροφορίας, εάν καταγράψουμε στις εξόδους μονάχα τις άρτιες θέσεις πληροφορίας (downsampling). Η ολοκληρωμένη διαδικασία (δεξιά), παράγει συντελεστές διακριτού μετασχηματισμού κυματιδίων.



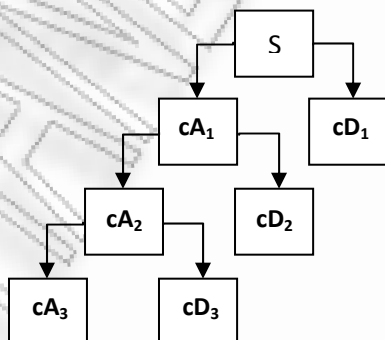
Εάν για παράδειγμα δημιουργήσουμε ημιτονοειδές σήμα, προσθέσουμε υψηλής συχνότητας θόρυβο και πραγματοποιήσουμε διακριτή ανάλυση ενός σταδίου, παρατηρούμε:



Παρατηρούμε ότι οι συντελεστές λεπτομερειών (detail coefficients) cD , αποτελούνται κυρίως από υψηλής συχνότητας θόρυβο, ενώ οι συντελεστές προσέγγισης (approximation coefficients) cA , περιέχουν πολύ λιγότερο θόρυβο από ότι το αρχικό σήμα. Το πλήθος των συντελεστών είναι μεγαλύτερο από το μισό του αρχικού δείγματος, λόγω του ότι η διαδικασία φιλτραρίσματος υλοποιείται μέσω της συνέλιξης του αρχικού σήματος με κάθε φίλτρο.

Αποσύνθεση πολλαπλών επιπέδων

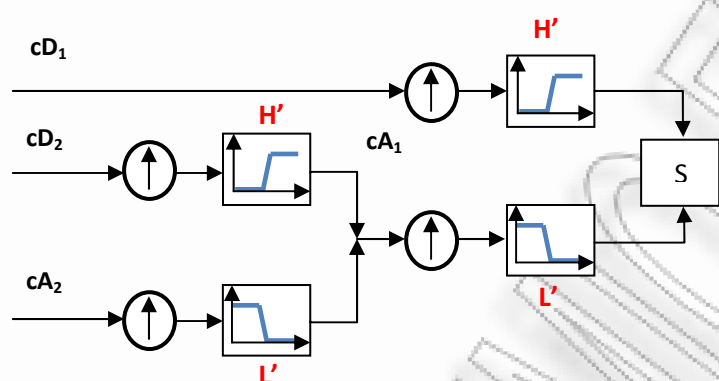
Η διαδικασία αποσύνθεσης μπορεί να είναι επαναληπτική, έτσι ώστε να αποσυντίθενται διαδοχικές προσεγγίσεις, ώστε να προκύπτουν πολλές συνιστώσες χαμηλότερης ανάλυσης. Η διαδικασία αυτή αποκαλείται δένδρο κυματιδιακής αποσύνθεσης (wavelet decomposition tree).



Λόγω του ότι η διαδικασία έχει επαναληπτική φύση, θεωρητικά μπορεί να συνεχιστεί επ' άπειρον, πρακτικά όμως μπορεί να συνεχιστεί έως ότου παραχθεί σήμα λεπτομερειών με ένα μόνο δείγμα πληροφορίας ή μέχρι να ικανοποιηθούν κατάλληλα κριτήρια.

Κυματιδιακή αναδόμηση

Ο διακριτός μετασχηματισμός κυματιδίων εμπεριέχει δύο βασικές διαδικασίες, τη διαδικασία της αποσύνθεσης ή ανάλυσης (decomposition-analysis) για στην οποία αναφερθήκαμε προηγουμένως καθώς και τη διαδικασία της αναδόμησης ή σύνθεσης (reconstruction-synthesis) αφορά την επαναδημιουργία του αρχικού σήματος από τους συντελεστές κυματιδίων, με τις μικρότερες δυνατές απώλειες και τυπικά αποκαλείται *αντίστροφος διακριτός μετασχηματισμός κυματιδίων* (inverse discrete wavelet transform-IDWT).

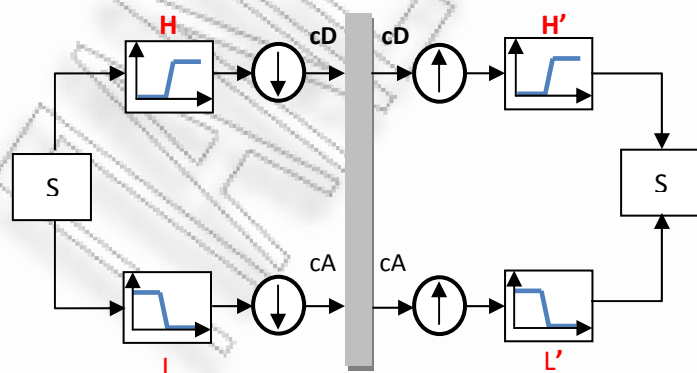


Ενώ η διαδικασία της ανάλυσης περιλαμβάνει τα στάδια της διέλευσης από φίλτρο και της μείωσης των δειγμάτων (υποδειγματοληψία), η διαδικασία της σύνθεσης περιλαμβάνει τα στάδια της αύξησης των δειγμάτων (υπερδειγματοληψία) και εν συνεχεία της διέλευσης από κατάλληλο φίλτρο. Κατά τη διαδικασία της αύξησης των δειγμάτων παρεμβάλλονται δείγματα μηδενικής πληροφορίας ανάμεσα στα πραγματικά δείγματα.

Φίλτρα αναδόμησης

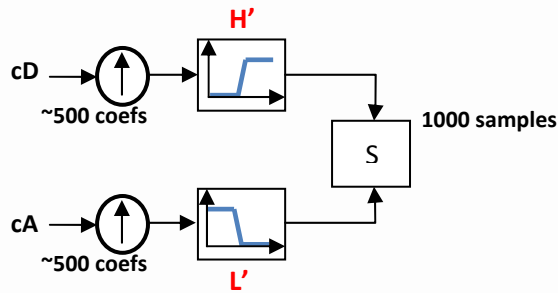
Η διαδικασία της διέλευσης μέσω φίλτρων κατά την αναδόμηση ενός σήματος απαιτεί φίλτρα ειδικού σχεδιασμού. Τα συγκεκριμένα φίλτρα θα πρέπει να είναι σχετικά μα όχι όμοια με εκείνα της αποδόμησης, ώστε να αναιρούν το φαινόμενο τη οδόντωσης (aliasing), το οποίο υπεισέρχεται κατά τη διαδικασία της αποδόμησης.

Τα τέσσερα φίλτρα που περιλαμβάνονται στο σύστημα που ορίζεται από τα χαμηλοπερατά και υψηροπερατά φίλτρα αποδόμησης και αναδόμησης (L-H, L'-H' αντίστοιχα), ονομάζονται quadrature mirror filters.

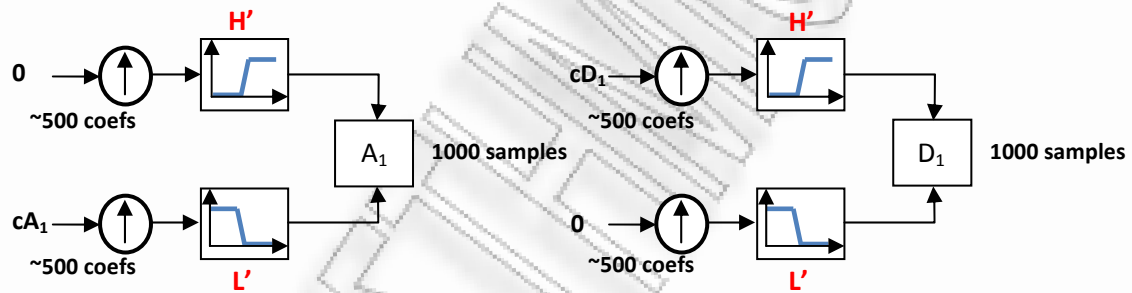


Αναδόμηση Προσεγγίσεων & Λεπτομερειών

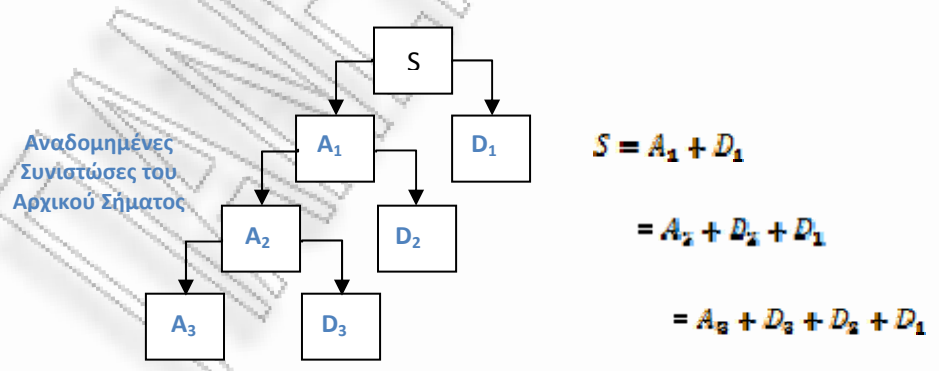
Για την αναδόμηση του αρχικού σήματος απαιτούνται οι συντελεστές προσεγγίσεων και λεπτομερειών:



Είναι δυνατόν να ανακατασκευάσουμε τις ίδιες τις προσεγγίσεις αλλά και τις λεπτομέρειες του αρχικού σήματος από τα αντίστοιχα διανύσματα συντελεστών, εισάγοντας διανύσματα μηδενικών τιμών στις κατάλληλες εισόδους. Η διαδικασία αναδόμησης **προσεγγίσεων** και **λεπτομερειών** παρουσιάζεται αντιστοίχως στα ακόλουθα σχήματα:



Τα αναδομημένα σήματα προσεγγίσεων (A_1) και λεπτομερειών (D_1), είναι πραγματικές συνημιτόνες του αρχικού σήματος S , αφού μπορούμε να τα αθροίσουμε ώστε να παράξουμε το αρχικό $A_1 + D_1 = S$. Παρατηρούμε επίσης ότι οι αρχικοί συντελεστές είναι απαραίτητο να αναδομηθούν πρώτα και να συνδυαστούν στην συνέχεια. Η ίδια τεχνική μπορεί να εφαρμοστεί και στην πολυεπίπεδη ανάλυση, επιτρέποντας την σύνθεση του αρχικού σήματος με διαφορετικούς τρόπους:



Συσχέτιση φίλτρων διακριτής ανάλυσης – σχήματος κυματιδίων

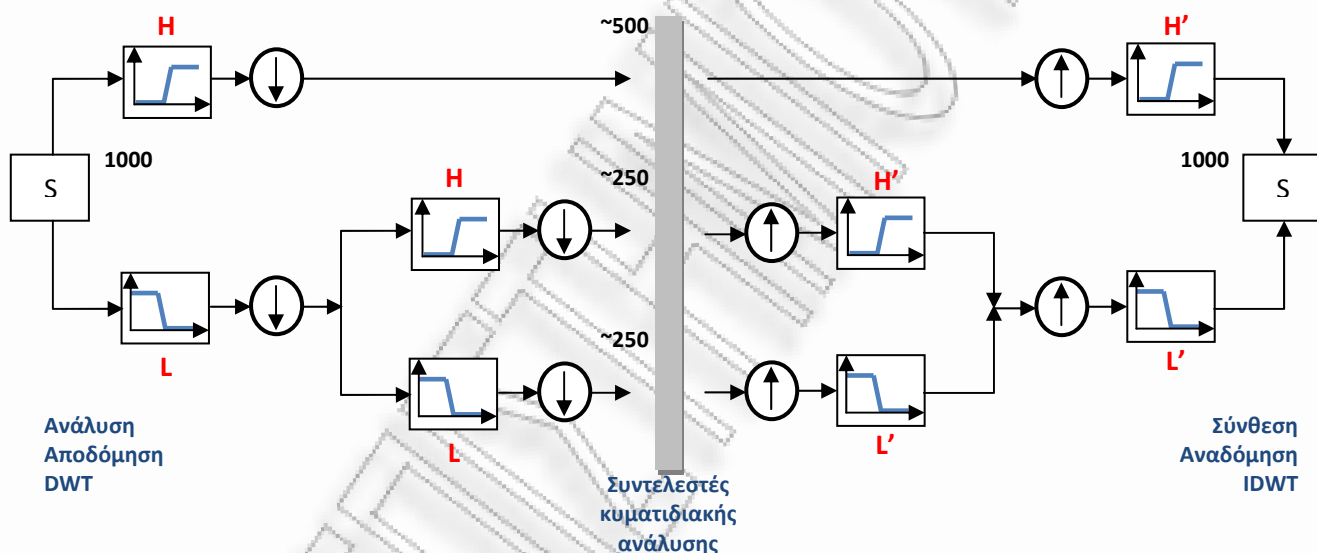
Η επιλογή των φίλτρων διακριτής ανάλυσης καθορίζει το σχήμα του κυματιδίου που θα χρησιμοποιηθεί στην ανάλυση. Συνήθως σχεδιάζεται πρώτα το σύστημα φίλτρων (quadrature mirror filters) και στη συνέχεια με αλληπάλληλες αυξήσεις δειγμάτων και συνελίξεις μεταξύ των φίλτρων προκύπτει η μορφή του σχετικού κυματιδίου.

Η συνάρτηση κυματιδίων $\phi(\omega)$, καθορίζεται από το υψηροπαστό φίλτρο, το οποίο παράγει τις λεπτομέρειες κατά της κυματιδιακής ανάλυσης του αρχικού σήματος.

Η συνάρτηση διακλιμάκωσης $\psi(\omega)$, είναι μια συμπληρωματική συνάρτηση η οποία συσχετίζεται με κάποια, μα όχι όλα τα κυματίδια και καθορίζεται από το χαμηλοπαστό φίλτρο, επομένως συσχετίζεται με τις λεπτομέρειες της κυματιδιακής ανάλυσης.

Ολοκληρωμένη διαδικασία Αποδόμησης – Αναδόμησης πολλαπλών επιπέδων

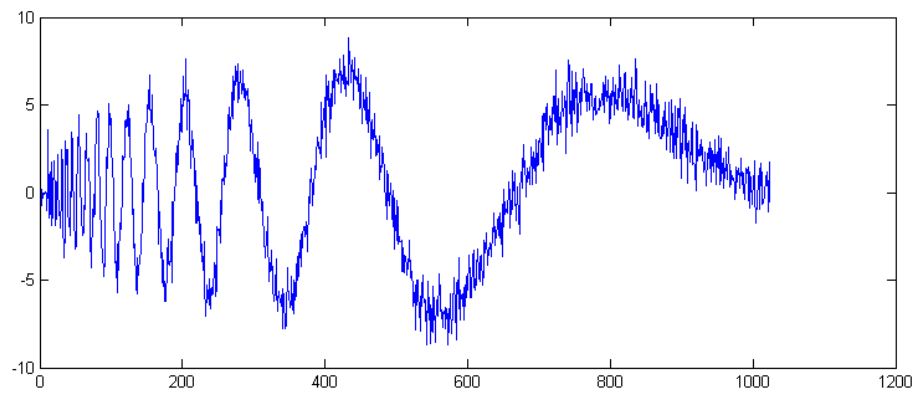
Η ολοκληρωμένη διαδικασία παρουσιάζεται στο ακόλουθο διάγραμμα



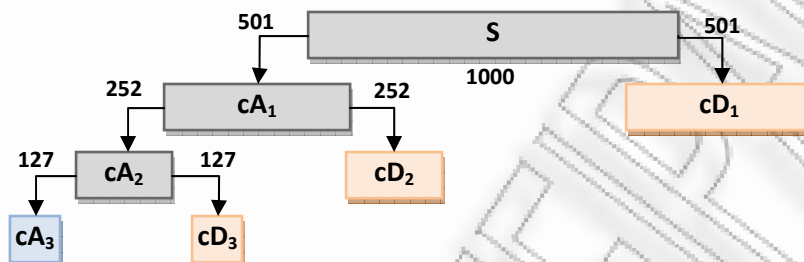
Σκοπός της διαδικασίας αυτής δεν είναι βέβαια η ανάλυση του αρχικού σήματος και η εν συνεχεία σύνθεσή του ώστε να παραχθεί εκ νέου η αρχική πληροφορία, αλλά η επεξεργασία των συντελεστών της ανάλυσης ώστε να προκύψει ωφέλιμο αποτέλεσμα όπως η ελάττωση του θορύβου, η συμπίεση της αρχικής πληροφορίας κ.α.

Παράδειγμα διακριτής κυματιδιακής ανάλυσης μονοδιάστατου σήματος

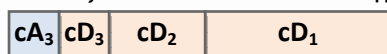
Δημιουργούμε δοκιμαστικό σήμα Doppler, το οποίο εμφανίζεται στο ακόλουθο διάγραμμα:



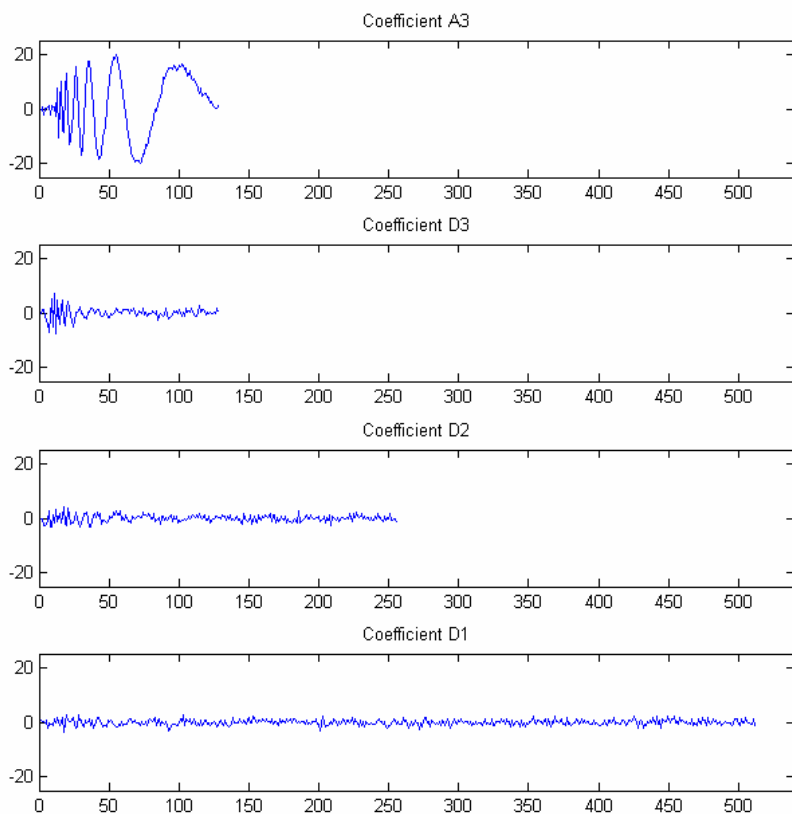
Εφαρμόζουμε διακριτή πολυεπίπεδη κυματιδική ανάλυση τριών επιπέδων, χρησιμοποιώντας κυματίδιο τύπου db1.



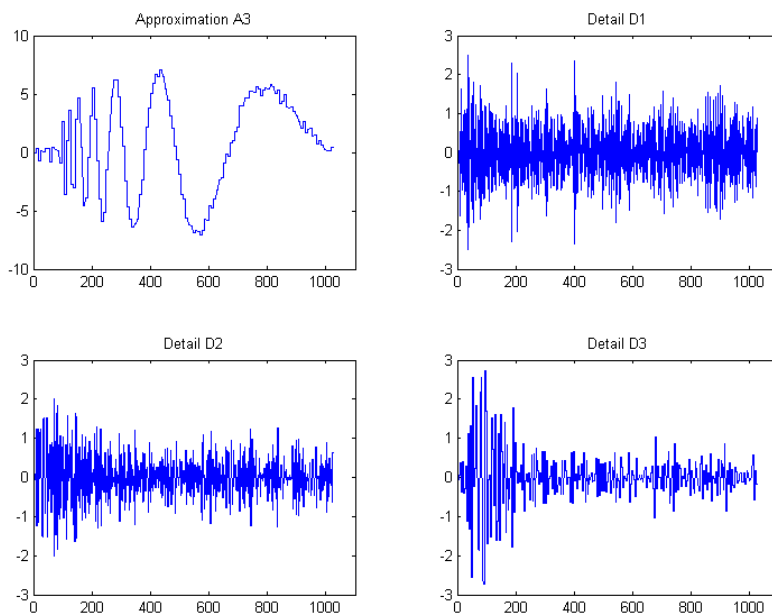
Οι παραγόμενοι συντελεστές κωδικοποιούνται σε ένα διάνυσμα, με σχετικά μήκη και θέσεις πληροφορίας, που απεικονίζονται στο ακόλουθο διάγραμμα:



Οι συντελεστές προσέγγισης, καθώς και οι συντελεστές λεπτομερειών για κάθε επίπεδο ανάλυσης παρουσιάζονται στο ακόλουθο σχήμα. Παρατηρούμε ότι σταδιακά σε κάθε επίπεδο, τμήμα του θορύβου κωδικοποιείται στους συντελεστές λεπτομερειών, και αφαιρείται από το αρχικό σήμα, ώστε να προκύψει η πληροφορία προσέγγισης, η οποία με τη σειρά της κωδικοποιείται στους συντελεστές προσέγγισης κατά το τελικό επίπεδο της ανάλυσης. Παρατηρούμε επίσης τον υποδιπλασιασμό του μήκους της πληροφορίας συντελεστών σε κάθε επίπεδο ανάλυσης (ο υποδιπλασιασμός δεν είναι απόλυτος μιας και το μήκος του ανύσματος συντελεστών επαυξάνεται σε συσχέτιση με το μήκος του φίλτρου ανάλυσης κατά τη συνέλιξή του με το σήμα εισόδου στο τρέχον επίπεδο).



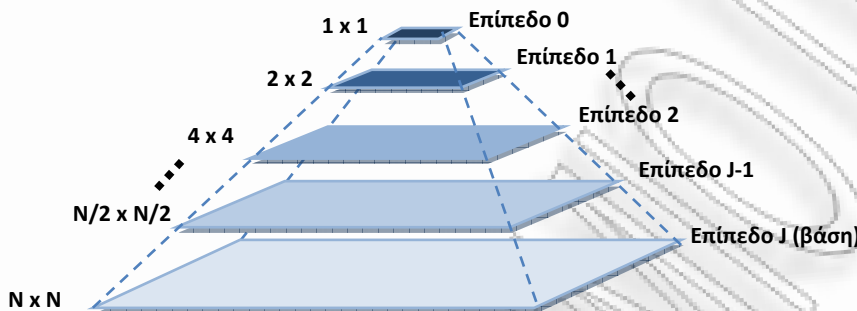
Από την επιμέρους σύνθεση των συνιστωσών συντελεστών ($cA_3 \rightarrow A_3$, $cD_3 \rightarrow D_3$, $cD_2 \rightarrow D_2$, $cD_1 \rightarrow D_1$), προκύπτουν τα ακόλουθα σχήματα:



Διακριτή κυματιδιακή ανάλυση δισδιάστατου σήματος

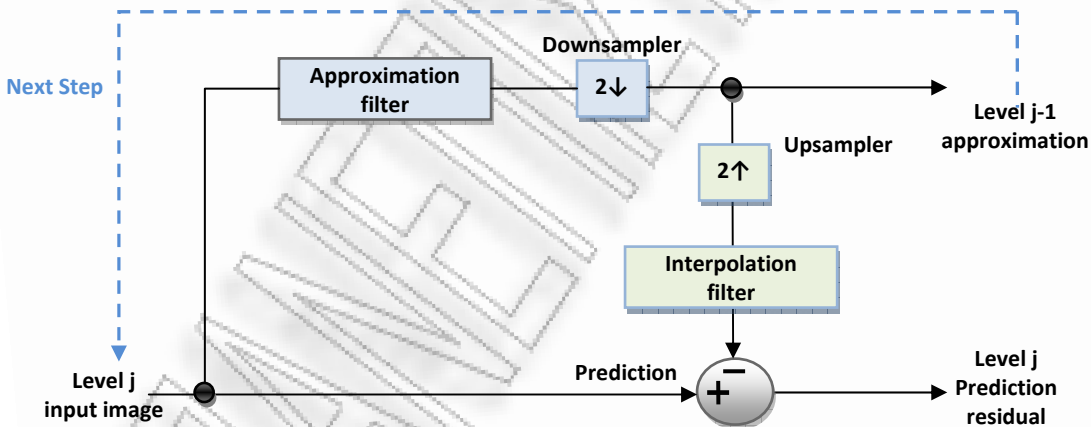
Η ανάλυση δισδιάστατου σήματος με χρήση κυματιδίων, θεωρείται από τις πλέον ολοκληρωμένες προσεγγίσεις για τη συμπίεση, μετάδοση, ανάλυση και επεξεργασία της πληροφορίας ψηφιακών εικόνων. Η χρήση κυματιδίων στην επεξεργασία εικόνας ενσωματώνει την πολυδιακριτική θεωρία (multiresolution theory) του Mallat (1989), η οποία με τη σειρά της ενοποιεί θεωρητικά πεδία, όπως η επεξεργασία σήματος, η επεξεργασία με χρήση ορθογωνικών φίλτρων (quadrature mirror filtering) και η ανάλυση εικόνας μέσω πυραμιδικής αναπαράστασης (pyramidal image processing).

Η ανάλυση εικόνας μέσω **πυραμιδικής αναπαράστασης** (Burt και Adelson 1983; Adelson et al., 1987), βασίζεται στη θεώρηση ότι συγκεκριμένες υφές ή αντικείμενα αναδεικνύονται καλύτερα σε διαφορετικές αναλύσεις. Υφές ή αντικείμενα μικρού μεγέθους ή χαμηλής αντίθεσης, αναδεικνύονται καλύτερα ως λεπτομέρειες, σε υψηλής ανάλυσης εκδοχές εικόνων, αντίθετα υφές ή αντικείμενα μεγάλου μεγέθους ή υψηλής αντίθεσης, αναδεικνύονται καλύτερα ως αδρά χαρακτηριστικά, χαμηλότερης ανάλυσης εκδοχών ψηφιακών εικόνων.



Όπως φαίνεται και στο παραπάνω σχήμα, στη βάση της πυραμίδας (επίπεδο J) κωδικοποιείται η αρχική πληροφορία υψηλής ανάλυσης, σταδιακά και σε κάθε επίπεδο έως την κορυφή (επίπεδο 0), το μέγεθος κάθε διάστασης υποδιπλασιάζεται ενώ η ανάλυση ελαττώνεται.

Η διαδικασία κωδικοποίησης της πληροφορίας στα διάφορα επίπεδα πραγματοποιείται μέσω μιας επαναληπτικής διαδικασίας, η οποία παρουσιάζεται στο ακόλουθο διάγραμμα:



Σύμφωνα με αυτή τη διαδικασία, σε κάθε στάδιο επεξεργασίας, υπολογίζεται μια μειωμένη ανάλυσης προσέγγιση της εικόνας εισόδου j , η οποία διέρχεται από κατάλληλο φίλτρο (approximation filter), ώστε να εξαλειφθεί το φαινόμενο της οδόντωσης (aliasing), ενώ στη συνέχεια ακολουθεί υποδειγματοληψία και εξαγωγή του αποτελέσματος ως *προσέγγιση επιπέδου $j-1$* , το οποίο ανατροφοδοτεί την επαναληπτική διαδικασία. Το ίδιο αποτέλεσμα χρησιμοποιείται παράλληλα ως προβλεπόμενη εικόνα (prediction image), αφού προηγηθεί υπερδειγματοληψία και διέλευση μέσω φίλτρου παρεμβολής (interpolation filter), για την εξάλειψη της εμφάνισης τοπικού ορθογωνισμού (blocking effect). Η διαφορά μεταξύ προβλεπόμενης εικόνας και εικόνας εισόδου κωδικοποιείται ως *προβλεπόμενη υπολειμματική*

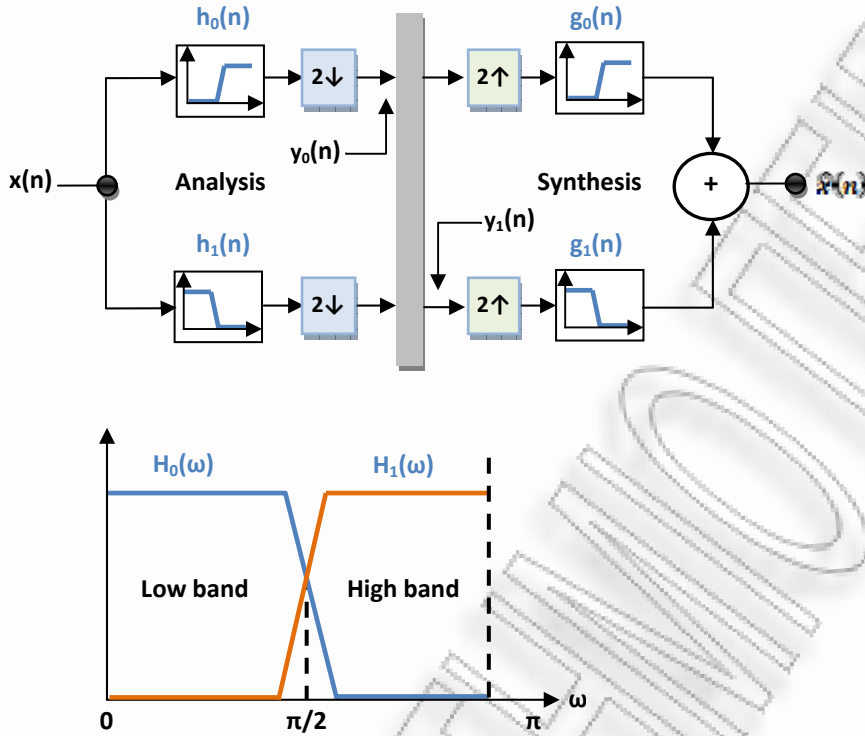
εικόνα (prediction residual) για το επίπεδο j . Από τη σύνθεση του συνόλου των υπολειμματικών εικόνων για όλα τα επίπεδα, προκύπτει η αρχική εικόνα εισόδου.

Στο ακόλουθο σχήμα παρατηρούμε την αρχική εικόνα εισόδου, καθώς και τις υπολειμματικές εικόνες που προέκυψαν, μετά από ανάλυση πυραμιδικής αναπαράστασης 3 επιπέδων:

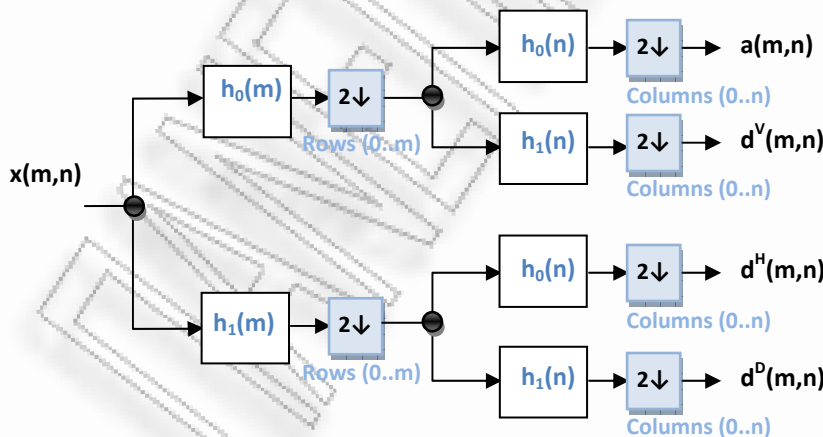


Η ανάλυση εικόνας μέσω **κωδικοποίησης υποζωνών** (subband coding), (Woods and O'Neil, 1986; Civanlar and Pun, 1992) είναι μια ακόμα μέθοδος πολυδιακριτικής ανάλυσης, η οποία με τη σειρά της άπτεται της κυματιδικής ανάλυσης εικόνας, ενώ προτείνεται και 3D εκδοχή της για την επεξεργασία video (Taubman and Zakhor, 1994). Κατά τη συγκεκριμένη τεχνική, μια εικόνα αναλύεται μέσω ζωνοπερατών φίλτρων σε συνιστώσες πληροφορίας περιορισμένης ζώνης (subbands). Στις επιμέρους συνιστώσες, εφαρμόζεται υποδειγματοληψία, χωρίς απώλεια πληροφορίας, αφού αυτές καταλαμβάνουν μικρότερο εύρος ζώνης από την αρχική πληροφορία. Η επανασύνθεση του αρχικού σήματος επιτυγχάνεται με εφαρμογή υπερδειγματοληψίας, χρήση κατάλληλων φίλτρων και άθροιση των επιμέρους συνιστωσών υποζωνών.

Στα ακόλουθα διαγράμματα αποτυπώνεται σύστημα κωδικοποίησης και αποκωδικοποίησης δύο υποζωνών μονοδιάστατου σήματος, καθώς και οι ιδιότητες διαχωρισμού φάσματος:



Το σύστημα κωδικοποίησης μονοδιάστατου σήματος μπορεί να επεκταθεί στις δύο διαστάσεις για την επεξεργασία σήματος εικόνας. Η κωδικοποίηση πραγματοποιείται σε δύο στάδια, ένα για κάθε διάσταση (κατακόρυφη και οριζόντια), με χρήση διαχωρισμών φίλτρων. Σε κάθε στάδιο μετά τη διέλευση από χαμηλοπερατό και υψηλοπερατό φίλτρο, εφαρμόζεται υποδειγματοληψία. Με το πέρας της διαδικασίας, προκύπτουν τέσσερις υποζώνες οι οποίες κωδικοποιούν, την προσέγγιση $a(m,n)$, τις κατακόρυφες λεπτομέρειες $d^V(m,n)$, τις οριζόντιες λεπτομέρειες $d^H(m,n)$ και τις διαγώνιες λεπτομέρειες $d^D(m,n)$ σήματος εικόνας διαστάσεων $m \times n$, όπως φαίνεται στο ακόλουθο διάγραμμα:



Οι θεωρίες οι οποίες αναπτύχθηκαν προηγουμένως ενοποιούνται και επεκτείνονται στην **κυματιδιακή ανάλυση δισδιάστατου σήματος**. Κατά την εφαρμογή της χρησιμοποιούνται ψηφιακά φίλτρα και υποδειγματοληψία σε κάθε διάσταση, αρχικά πραγματοποιείται μονοδιάστατη κυματιδιακή ανάλυση των γραμμών του σήματος $f(x,y)$, η οποία ακολουθείται από

ανάλυση των εναπομεινάντων σηλών. Ως αποτέλεσμα προκύπτουν τέσσερα διανύσματα συντελεστών, ένα προσέγγισης και τρία λεπτομερειών (οριζόντιες, κατακόρυφες και διαγώνιες). Εάν η συστοιχία φίλτρων (filter bank), η οποία αξιοποιείται για ανάλυση ενός επιπέδου, ανατροφοδοτηθεί με το παραγόμενο διάνυσμα προσέγγισης επαναληπτικά, προκύπτει σύστημα πολυεπίπεδης ανάλυσης.

Για την υλοποίηση της ανάλυσης δύο διαστάσεων απαιτείται μια δισδιάστατη συνάρτηση διακλιμάκωσης $\varphi(x,y)$ και τρεις δισδιάστατες συναρτήσεις κυματιδίων $\psi^H(x,y)$, $\psi^V(x,y)$, $\psi^D(x,y)$, οι οποίες προκύπτουν μέσω των αντίστοιχων μονοδιάστατων ως εξής:

$$\varphi(x,y) = \varphi(x)\varphi(y)$$

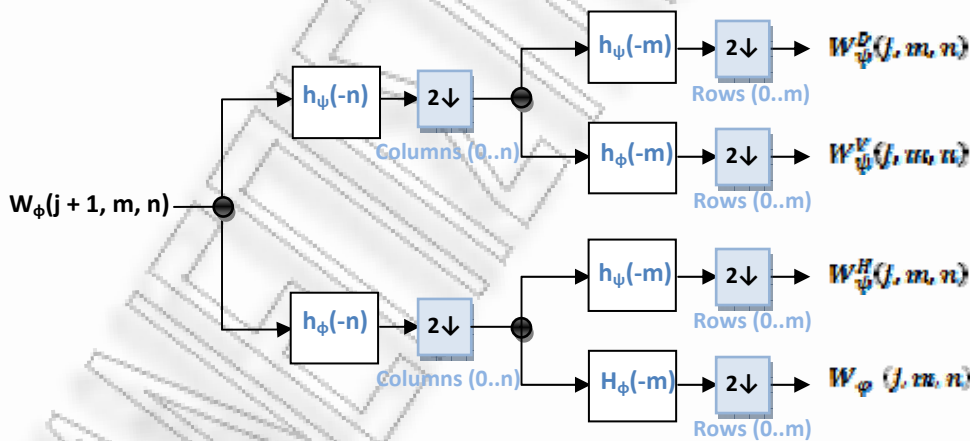
$$\psi^H(x,y) = \psi(x)\psi(y)$$

$$\psi^V(x,y) = \varphi(x)\psi(y)$$

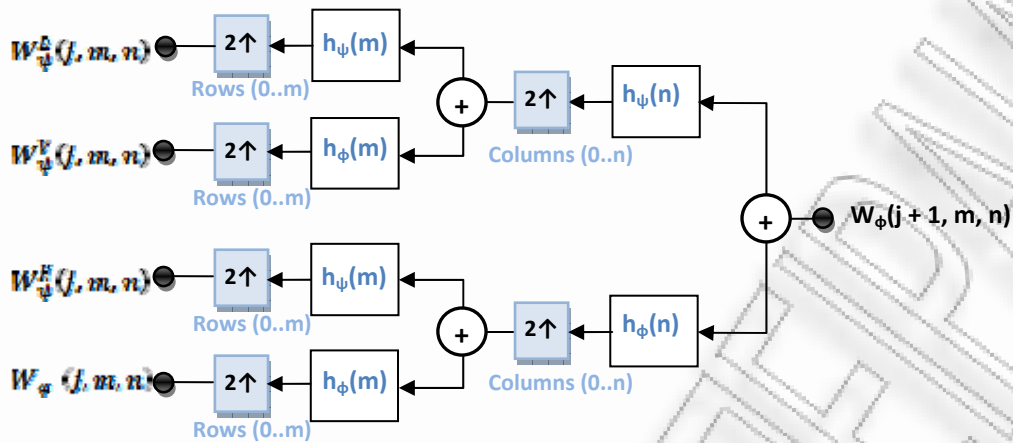
$$\psi^D(x,y) = \psi(x)\varphi(y)$$

Οι παραγόμενες συναρτήσεις κυματιδίων έχουν δύο σημαντικές ιδιότητες, είναι κατευθυντικές και μπορούν να υπολογίσουν τις διακυμάνσεις της φωτεινότητας στις διευθύνσεις που τους αντιστοιχούν (οριζόντια, κατακόρυφη, διαγώνια).

Στο ακόλουθο διάγραμμα παρουσιάζεται συνοπτικά η διαδικασία κυματιδιακής ανάλυσης δύο διαστάσεων. Παρατηρούμε ότι η εικόνα $f(x,y)$, εισέρχεται στο σύστημα ως $W_\varphi(j,m,n)$, όπου J το επίπεδο ανάλυσης και m,n οι διαστάσεις της. Στη συνέχεια εφαρμόζεται συνέλιξη κατά γραμμές, με υπηπερατό και χαμηλοπερατό φίλτρο, $h_\psi(-n)$ και $h_\phi(-n)$ αντίστοιχα, και εν τέλει υποδειγματοληψία κατά την κατακόρυφη διάσταση (στήλες). Ως αποτέλεσμα προκύπτουν δύο εικόνες με υποδιπλάσια οριζόντια ανάλυση, στις οποίες εφαρμόζεται συνέλιξη κατά στήλες και υποδειγματοληψία κατά την οριζόντια διάσταση (γραμμές). Οι εικόνες συντελεστών που προκύπτουν έχουν υποτετραπλάσιο μέγεθος από την εικόνα εισόδου, συμβολίζονται ως W_ψ^D , W_ψ^H , W_ψ^V , W_ϕ και αναπαριστούν αντίστοιχα την προσέγγιση και τις λεπτομέρειες, της εικόνας εισόδου, κατά την οριζόντια, κατακόρυφη και διαγώνια διεύθυνση.



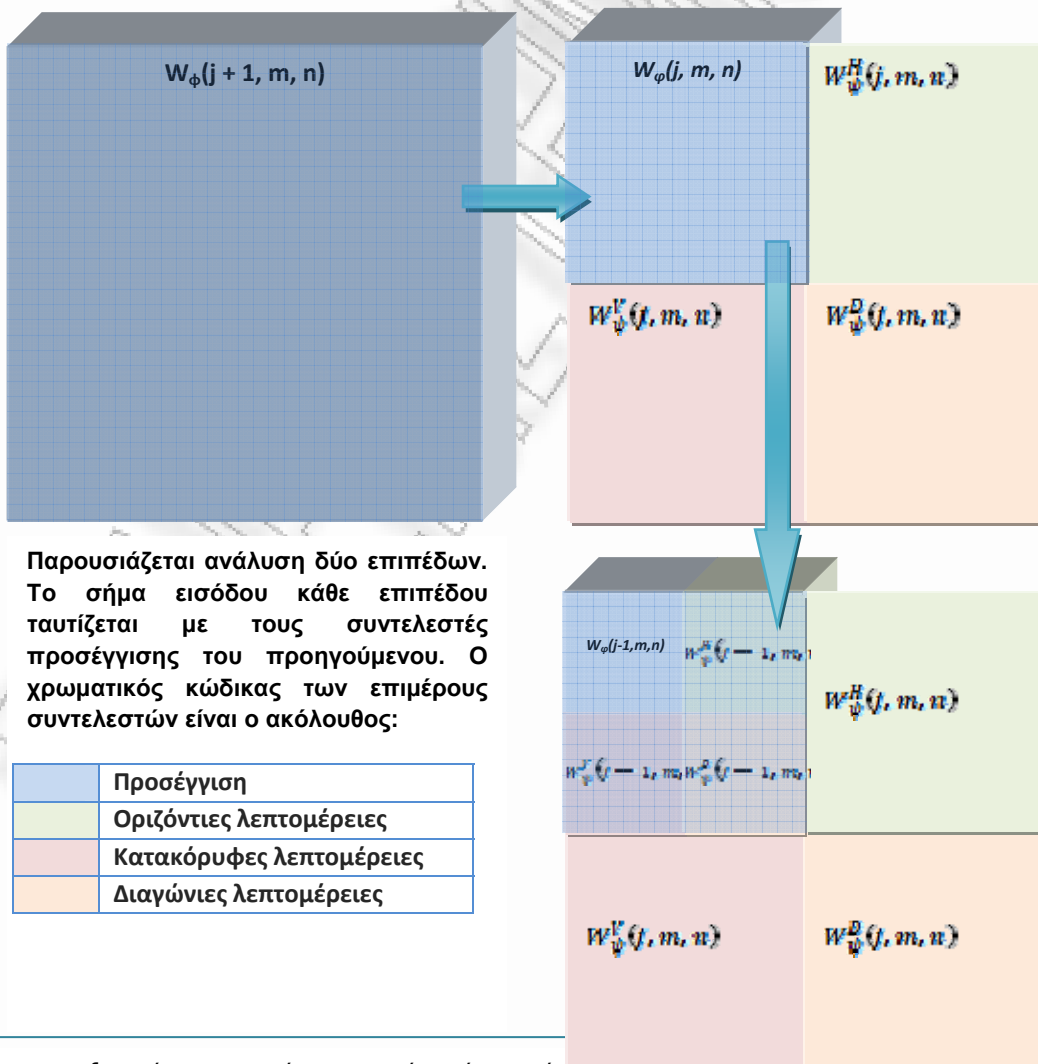
Η συμπληρωματική διαδικασία κυματιδιακής σύνθεσης παρουσιάζεται στο ακόλουθο διάγραμμα, στους συντελεστές εισόδου ανά ζεύγη, $(W_\psi^D(j, m, n), W_\psi^V(j, m, n)) - (W_\psi^H(j, m, n), W_\phi(j, m, n))$, εφαρμόζεται υπερδειγματοληψία κατά την οριζόντια διάσταση, εν συνεχεία συνέλιξη κατά στήλες με υπηπερατό και χαμηλοπερατό φίλτρο, $h_\phi(n) - h_\psi(n)$ και εν τέλει άθροιση των αποτελεσμάτων. Στο ζεύγος των ενδιάμεσων αποτελεσμάτων εφαρμόζεται αντίστοιχη διαδικασία, υπορδειγματοληψία κατά την κατακόρυφη διάσταση, συνέλιξη κατά γραμμές και άθροιση των συνιστωσών, με τελικό αποτέλεσμα την σύνθεση του αρχικού σήματος της εικόνας εισόδου $W_\varphi(j, m, n)$.



Η συνολική παραγόμενη πληροφορία συντελεστών, μετά το πέρας της διαδικασίας πολυεπίπεδης κυματιδικής ανάλυσης, καταλαμβάνει προσεγγιστικά, ίδιο μέγεθος με την αρχική πληροφορία. Πρακτικά όμως η παραγόμενη πληροφορία είναι ελαφρώς μεγαλύτερη της αρχικής λόγω του ότι το μήκος κάθε διάστασης δεν υποδιπλασιάζεται ακριβώς σε κάθε στάδιο, αλλά

διέπεται από την ακόλουθη σχέση:
$$S_j = \left\lfloor \frac{S_{j+1} + S_f - 1}{2} \right\rfloor$$
, όπου: S_j = το μήκος της πληροφορίας κατά διάσταση, S_f = το μήκος του φίλτρου ανάλυσης, j = το επίπεδο ανάλυσης.

Στο ακόλουθο διάγραμμα απεικονίζεται η παραγόμενη πληροφορία συντελεστών, αναλογικά με το μέγεθός της, για κάθε επίπεδο ανάλυσης:



Παρουσιάζεται ανάλυση δύο επιπέδων. Το σήμα εισόδου κάθε επιπέδου ταυτίζεται με τους συντελεστές προσέγγισης του προηγούμενου. Ο χρωματικός κώδικας των επιμέρους συντελεστών είναι ο ακόλουθος:

	Προσέγγιση
	Οριζόντιες λεπτομέρειες
	Κατακόρυφες λεπτομέρειες
	Διαγώνιες λεπτομέρειες

Διακριτή κυματιδιακή ανάλυση 3D σήματος

Η κυματιδιακή ανάλυση στις τρεις διαστάσεις (Burns, 1993), αποτελεί το κεντρικό σημείο μελέτης της παρούσας εργασίας, καθώς και τον πυρήνα της εφαρμογής **3D TIEVID**, η οποία αναπτύχθηκε παράλληλα με αυτή.

Οι Burns et al., 1996, μέσω της εργασίας τους «A Wavelet Multiresolution Analysis for Spatio-Temporal Signals» («Πολυδιακριτική Κυματιδιακή Ανάλυση για Χώρο-Χρονικά Σήματα»), αναλύουν το πρόβλημα και προτείνουν σχετικές μεθοδολογίες. Η εργασία τους αξιοποιήθηκε ως το θεωρητικό υπόβαθρο, για την υλοποίηση του προτεινόμενου αλγόριθμου, διότι παράλληλα με τη συμβατική προσέγγιση, προτείνεται μια μεθοδολογία **μη συμβατικής και ανομοιογενούς προσέγγισης** των χώρο-χρονικών συνιστωσών του σήματος. Η εφαρμογή της συγκεκριμένης μεθοδολογίας επεκτείνεται και ενισχύει τη γενικότερη πρόταση των (Wang και Chen, 2006) για την εξαγωγή πληροφορίας κειμένου από βίντεο μέσω κυματιδιακού μετασχηματισμού.

Η συμβατική προσέγγιση 3D πολυδιακριτικής ανάλυσης αντιμετωπίζει την πληροφορία στα πεδία τη χώρου και του χρόνου, θεωρώντας ότι αυτή κατέχει ομοιογενή χαρακτηριστικά. Η θεώρηση αυτή δυσχεραίνει το συνδυασμό των χαρακτηριστικών της συχνότητας του φίλτρου ανάλυσης με εκείνη του σήματος. Επίσης, η συμβατική ανάλυση καταλήγει στη δημιουργία δομής ανάλυσης οκτώ υποζωνών για κάθε επίπεδο, περιορίζοντας σε ταύτιση τα επίπεδα ανάλυσης στο χώρο και το χρόνο. Οι συγγραφείς προτείνουν έναν αλγόριθμο πολυδιακριτικής ανάλυσης, ο οποίος αξιοποιεί ανομοιογενή φίλτρα στο χώρο και το χρόνο, και παράλληλα μια δομή ορθογωνικής κωδικοποίησης υποζωνών η οποία διαχωρίζει τις διαδικασίες χωρικής και χρονικής ανάλυσης.

Κατά τη συμβατική θεώρηση η συνάρτηση διακλιμάκωσης αλλά και οι συναρτήσεις κυματιδίων, προκύπτουν από τα προϊόντα των αντίστοιχων μονοδιάστατων συναρτήσεων:

$$\begin{aligned}\varphi(x, y, t) &= \varphi(x)\varphi(y)\varphi(t) \\ \varphi^1(x, y, t) &= \varphi(x)\varphi(y)\varphi^1(t) \\ \varphi^2(x, y, t) &= \varphi(x)\varphi(y)\varphi^2(t) \\ \varphi^3(x, y, t) &= \varphi(x)\varphi(y)\varphi^3(t) \\ \varphi^4(x, y, t) &= \varphi(x)\varphi(y)\varphi^4(t) \\ \varphi^5(x, y, t) &= \varphi(x)\varphi(y)\varphi^5(t) \\ \varphi^6(x, y, t) &= \varphi(x)\varphi(y)\varphi^6(t) \\ \varphi^7(x, y, t) &= \varphi(x)\varphi(y)\varphi^7(t)\end{aligned}$$

Παρατηρούμε ότι για την υλοποίηση ανάλυσης τριών διαστάσεων, προκύπτουν μια συνάρτηση διακλιμάκωσης και επτά συναρτήσεις κυματιδίων. Όπως προαναφέρθηκε όμως, από τη χρήση κοινών συναρτήσεων (και κατά συνέπεια κοινών φίλτρων ($h(n), g(n)$)), για την ανάλυση της πληροφορίας στα πεδία του χώρου και του χρόνου, προκύπτουν περιορισμοί στην επάρκεια της μεθοδολογίας.

Γι' αυτό το λόγο οι συγγραφείς αποδεικνύουν την ορθότητα της χρήσης διαφορετικών συναρτήσεων και φίλτρων σε κάθε πεδίο και προτείνουν την ανομοιομορφή κυματιδιακή ανάλυση τριών διαστάσεων, όπου η συνάρτηση διακλιμάκωσης και οι συναρτήσεις κυματιδίων, προκύπτουν πλέον από τα αντίστοιχα ζεύγη μονοδιάστατων ($\varphi - \tilde{\varphi}$, $\psi - \tilde{\psi}$) ως εξής:

$$\begin{aligned}\varphi(x, y, t) &= \varphi(x)\varphi(y)\tilde{\varphi}(t) \\ \psi^1(x, y, t) &= \varphi(x)\varphi(y)\tilde{\psi}^1(t) \\ \varphi^2(x, y, t) &= \varphi(x)\varphi(y)\tilde{\varphi}^2(t) \\ \varphi^3(x, y, t) &= \varphi(x)\varphi(y)\tilde{\varphi}^3(t) \\ \varphi^4(x, y, t) &= \varphi(x)\varphi(y)\tilde{\varphi}^4(t) \\ \varphi^5(x, y, t) &= \varphi(x)\varphi(y)\tilde{\varphi}^5(t) \\ \varphi^6(x, y, t) &= \varphi(x)\varphi(y)\tilde{\varphi}^6(t) \\ \psi^7(x, y, t) &= \varphi(x)\varphi(y)\tilde{\psi}^7(t)\end{aligned}$$

Τα δύο ζεύγη ορθογωνικών κατοπτρικών φίλτρων (quadrature mirror filters - QMF), που χρησιμοποιούνται στο μετασχηματισμό $(h(n), g(n))$ και $(\bar{h}(n), \bar{g}(n))$, προκύπτουν από τις ακόλουθες εξισώσεις:

$$h(n) = \int_{-\infty}^{\infty} \varphi\left(\frac{x}{2}\right) \varphi(x-n) dx$$

$$\bar{h}(n) = \int_{-\infty}^{\infty} \varphi\left(\frac{t}{2}\right) \varphi(t-n) dt$$

και

$$g(n) = (-1)^{1-n} h(1-n)$$

$$\bar{g}(n) = (-1)^{1-n} \bar{h}(1-n)$$

όπου

$$\bar{h}(n) = h(-n)$$

$$\bar{\bar{h}}(n) = \bar{h}(-n)$$

Οι συντελεστές προσέγγισης και λεπτομερειών για το επίπεδο ανάλυσης j , προκύπτουν μετά από συνέλιξη των συναρτήσεων διακλιμάκωσης και κυματιδίων, με τους συντελεστές προσέγγισης του επιπέδου $j-1$, και κατόπιν υποδειγματοληψίας του αποτελέσματος, κατά συντελεστή 2 ($2l, 2m, 2n$):

Όπου $\alpha_{j,l,m,n}^i$, ο συντελεστής προσέγγισης στη θέση (l,m,n) σε ένα κυβικό πλέγμα δειγματοληψίας, ενώ όπου οι συντελεστές λεπτομερειών.

$$\alpha_{j,l,m,n}^1 = [\alpha_{j+1,p,q,r} * \bar{h}(p) * \bar{h}(q) * \bar{h}(r)](2l, 2m, 2n)$$

$$\alpha_{j,l,m,n}^2 = [\alpha_{j+1,p,q,r} * \bar{h}(p) * \bar{h}(q) * \bar{g}(r)](2l, 2m, 2n)$$

$$\alpha_{j,l,m,n}^3 = [\alpha_{j+1,p,q,r} * \bar{h}(p) * \bar{g}(q) * \bar{h}(r)](2l, 2m, 2n)$$

$$\alpha_{j,l,m,n}^4 = [\alpha_{j+1,p,q,r} * \bar{h}(p) * \bar{g}(q) * \bar{g}(r)](2l, 2m, 2n)$$

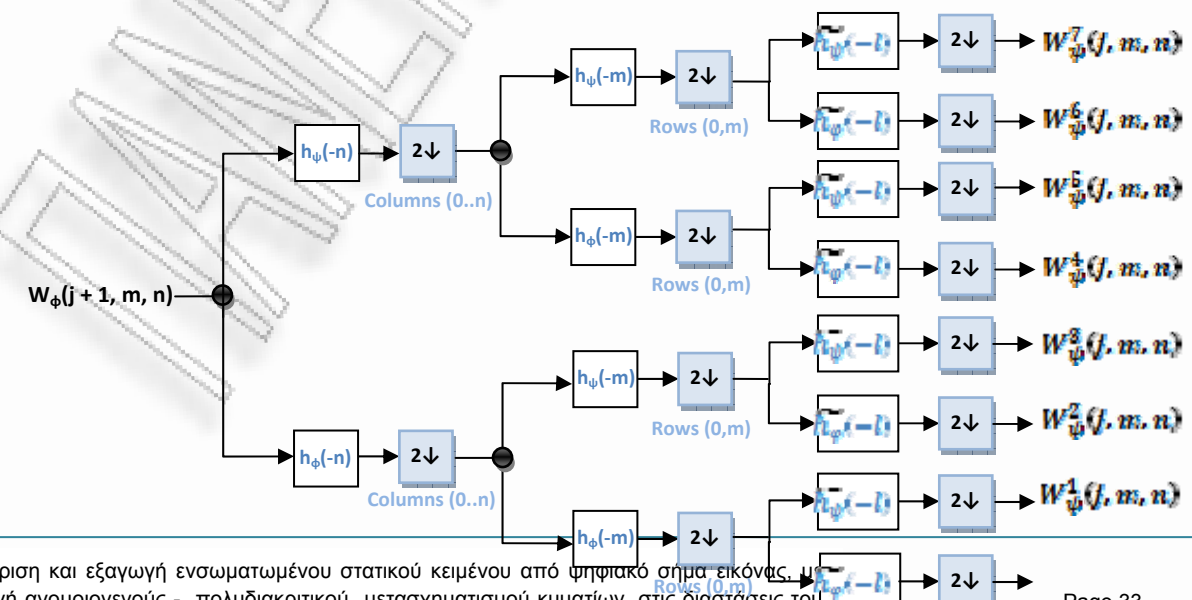
$$\alpha_{j,l,m,n}^5 = [\alpha_{j+1,p,q,r} * \bar{g}(p) * \bar{h}(q) * \bar{h}(r)](2l, 2m, 2n)$$

$$\alpha_{j,l,m,n}^6 = [\alpha_{j+1,p,q,r} * \bar{g}(p) * \bar{h}(q) * \bar{g}(r)](2l, 2m, 2n)$$

$$\alpha_{j,l,m,n}^7 = [\alpha_{j+1,p,q,r} * \bar{g}(p) * \bar{g}(q) * \bar{h}(r)](2l, 2m, 2n)$$

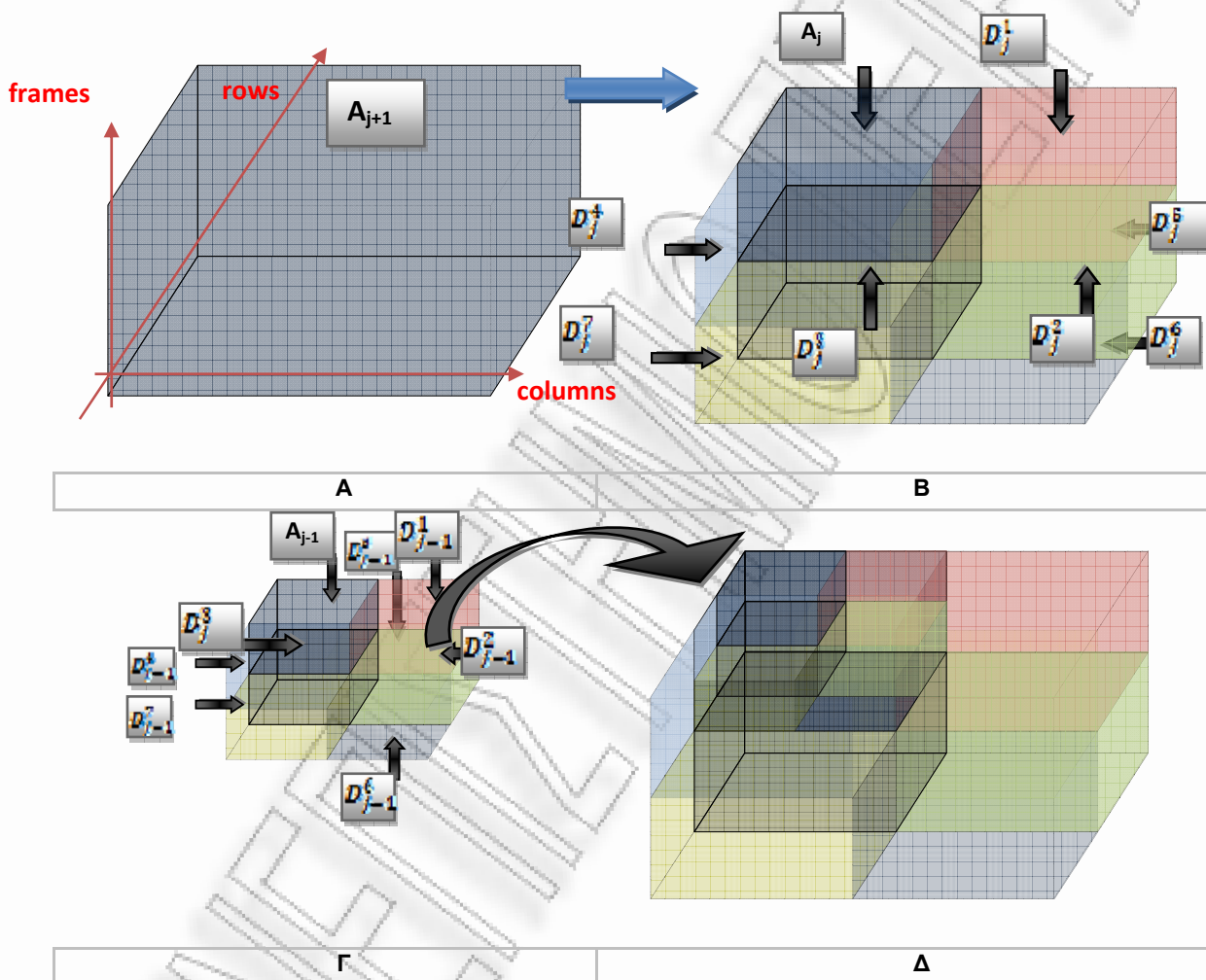
$$\alpha_{j,l,m,n}^8 = [\alpha_{j+1,p,q,r} * \bar{g}(p) * \bar{g}(q) * \bar{g}(r)](2l, 2m, 2n)$$

Από την προηγούμενη διαδικασία προκύπτουν οκτώ συνδυασμοί, οι οποίοι αποτελούν τις απολήξεις της χαρακτηριστικής δομής 3D συμβατικής ανάλυσης οκτώ κλάδων (oct-tree decomposition). Η ανωτέρω ανάλυση αναφέρεται ως συμβατική επειδή αποτελεί απλή επέκταση της ανάλυσης δύο διαστάσεων, αφού κάθε κλάδος της αναλύεται περεταίρω σε δύο ακόμα υποζώνες στο πεδίο του χρόνου, όπως φαίνεται και στο ακόλουθο σχήμα:



$$W_{\phi}^{*j, m, n}$$

Το εύρος συχνοτήτων, που καταλαμβάνει στο χώρο και το χρόνο η προσέγγιση του σήματος εισόδου, στο επίπεδο ανάλυσης j , συμβολίζεται ως W_{ϕ}^{*j} . Η ανομοιογενής κυματιδιακή ανάλυση τριών διαστάσεων, αναλύει το σήμα σε σύνολο ανεξάρτητων χώρο-χρονικών καναλιών συχνότητας. Γι' αυτό το λόγο, τα συχνοτικά χαρακτηριστικά των φίλτρων ανάλυσης, μπορούν να διαφοροποιηθούν ανεξάρτητα, μέσω της επιλογής κατάλληλου τύπου φίλτρου, ενώ παράλληλα διατηρείται η ορθογωνικότητα των συναρτήσεων των κυματιδίων βάσης.

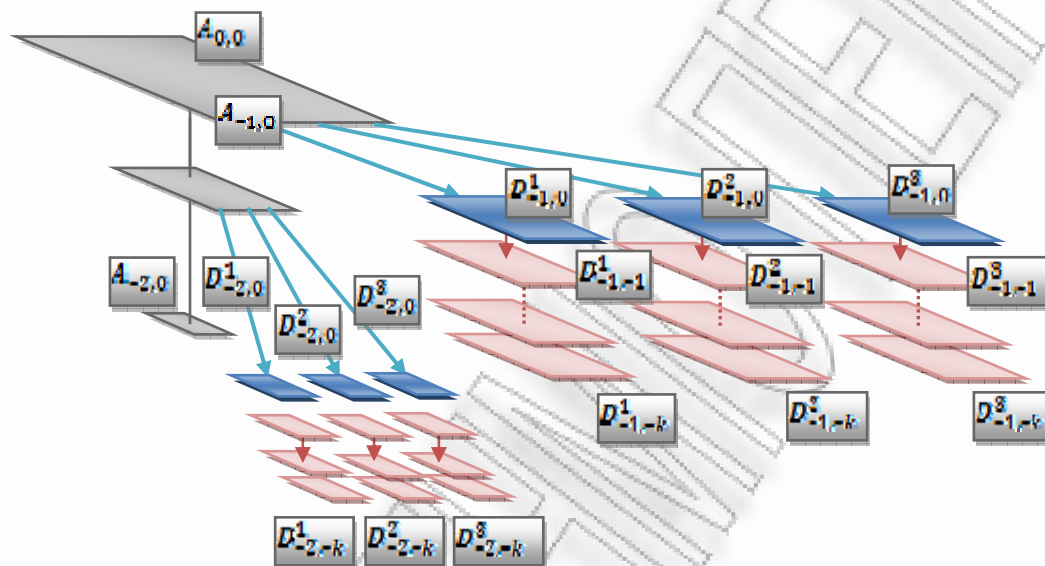


Στο προηγούμενο διάγραμμα, απεικονίζεται η διαδικασία 3Δ συμβατικής κυματιδιακής ανάλυσης, σε δύο επίπεδα. Στο σήμα εισόδου A_{j+1} (Α), εφαρμόζεται κυματιδιακή ανάλυση, οπότε προκύπτουν οκτώ τριών διαστάσεων συντελεστές στο επίπεδο j (Β). Στους συντελεστές προσέγγισης A_j , εφαρμόζεται εκ νέου κυματιδιακή ανάλυση (Γ), οπότε ο χώρος που καταλαμβάνουν (W_{ϕ}^{*j}), υποκαθίσταται από τους οκτώ παραγόμενους συντελεστές του επιπέδου $j-1$ (Δ).

Η δομή 3Δ συμβατικής ανάλυσης οκτώ κλάδων, οδηγεί σε σύμπτωση του πλήθους των επιμέρους αναλύσεων, στο χώρο και το χρόνο. Τη συμπεριφορά αυτή μπορούμε να τη διαπιστώσουμε στα προηγούμενα σχήματα, όπου το μήκος κάθε διάστασης υποδιπλασιάζεται ομοίμορφα σε κάθε διάσταση, ενώ ο όγκος που καταλαμβάνουν οι συντελεστές υποοκταπλασιάζεται (2^3) σε κάθε επίπεδο ανάλυσης. Οι συγγραφείς θεωρούν, ότι με τη συγκεκριμένη μεθοδολογία, δεν είναι δυνατόν να κατασκευαστούν φίλτρα, τα οποία θα διαχωρίσουν κινούμενα αντικείμενα με ανόμοιες χαρακτηριστικές συχνότητας στο χώρο και το

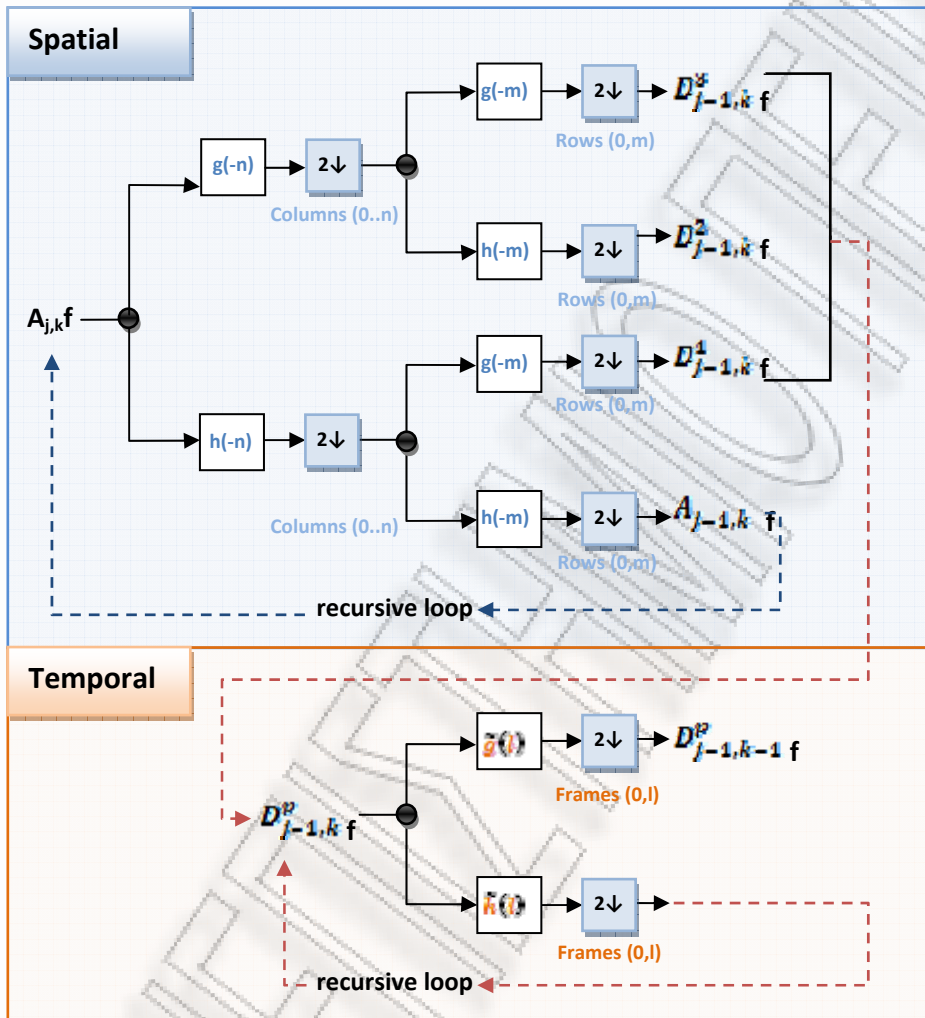
χρόνο, όπως μικρά και αργά κινούμενα αντικείμενα (υψηλή χωρική -χαμηλή χρονική συχνότητα) ή αδρά και γοργά κινούμενα αντικείμενα (χαμηλή χωρική – υψηλή χρονική συχνότητα).

Για τους παραπάνω λόγους, προτείνουν μια μη συμβατική μεθοδολογία ανομοιογενούς κυματιδιακού μετασχηματισμού, η οποία διαχωρίζει τους επιμέρους μηχανισμούς ανάλυσης στα πεδία του χώρου και του χρόνου. Η προτεινόμενη μεθοδολογία επιτρέπει τη χρονική ανάλυση σήματος σε επιθυμητό αριθμό επιπέδων για δεδομένη χωρική ανάλυση. Οι συγγραφείς αποδεικνύουν ότι διατηρείται η ορθογωνικότητα των συναρτήσεων των κυματιδίων βάσης, καθώς και η δυνατότητα σύνθεσης του αρχικού σήματος από τους παραγόμενους συντελεστές.



Στο προηγούμενο σχήμα, αναπαρίσταται η προτεινόμενη διαδικασία ανάλυσης. Στο σήμα εισόδου $A_{0,0}$, εφαρμόζεται χωρική κυματιδιακή ανάλυση, οπότε προκύπτουν τέσσερις ακολουθίες συντελεστών για το χωρικό επίπεδο $j=-1$, η ακολουθία προσέγγισης $A_{-1,0}$, καθώς και οι τρεις ακολουθίες λεπτομερειών $D_{-1,0}^1, D_{-1,0}^2, D_{-1,0}^3$. Η διαδικασία πραγματοποιείται μέσω συνέλιξης σε κάθε διάσταση (γραμμές, στήλες), των επιμέρους καρέ του σήματος $A_{0,0}$, με τις αντεστραμμένες εκδοχές των χωρικών φίλτρων h και g , ενώ στη συνέχεια πραγματοποιείται υποδειγματοληψία, με αποτέλεσμα τον υποδιπλασιασμό του μήκους των χωρικών διαστάσεων. Η παραγόμενη ακολουθία προσέγγισης $A_{-1,0}$, ανατροφοδοτεί τη διαδικασία χωρικής ανάλυσης για τον επιθυμητό αριθμό χωρικών επιπέδων $A_{-j,0} \quad j = 1,2,3 \dots$, ώστε να παραχθεί η πληροφορία υποβάθρου, η οποία ουσιαστικά κωδικοποιεί και καταγράφει τις λεπτομέρειες του σήματος εισόδου, για διαδοχικά χαμηλότερες χωρικές αναλύσεις και για σταθερή χρονική ανάλυση στο επίπεδο χρονικής ανάλυσης $k=0$.

Σε δεύτερη φάση, για κάθε επίπεδο χωρικής ανάλυσης, εφαρμόζεται διαδοχικά στους παραχθέντες συντελεστές λεπτομερειών ($D_{-j,0}^p \quad p=1,2,3$), συνέλιξη στη διάσταση του χρόνου (διαδοχικά καρέ) με τις αντεστραμμένες εκδοχές των χρονικών φίλτρων \tilde{h} και \tilde{g} και κατόπιν υποδειγματοληψία, με αποτέλεσμα τον υποδιπλασιασμό του μήκους της χρονικής διάστασης. Οι παραγόμενες χρονικές ακολουθίες προσέγγισης, ανατροφοδοτούν τη διαδικασία για το πλήθος των οριζόμενων επιπέδων χρονικής ανάλυσης, ενώ παράλληλα καταγράφονται οι χρονικές ακολουθίες λεπτομερειών, $D_{-j,-k}^p \quad p=1,2,3$

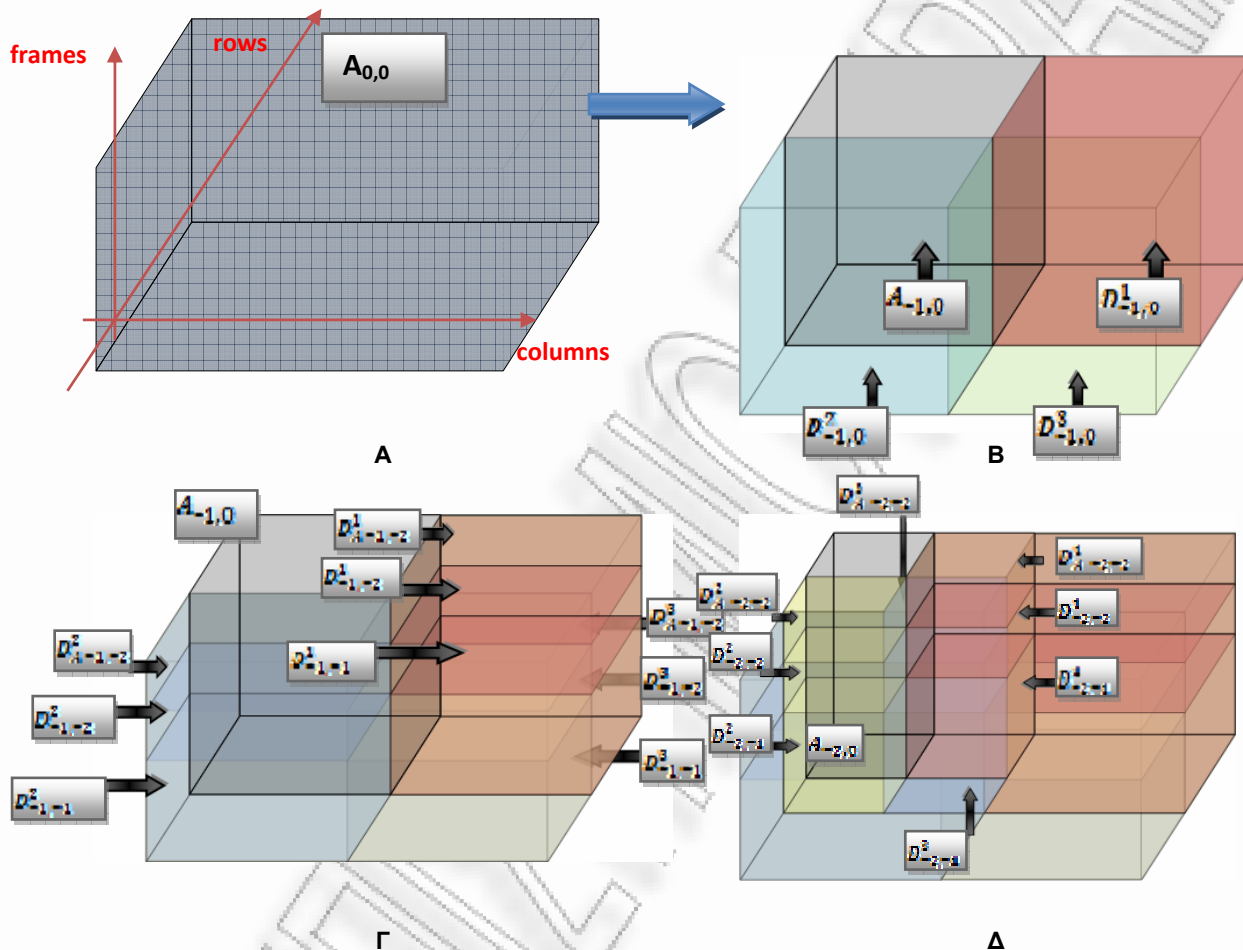


Στο ακόλουθο διάγραμμα, απεικονίζεται η διαδικασία 3D μη συμβατικής κυματιδικής ανάλυσης, σε δύο χωρικά και δύο χρονικά επίπεδα. Στο σήμα εισόδου $A_{j,k} f$ (\mathbf{A}), εφαρμόζεται χωρική ανάλυση, οπότε προκύπτουν τέσσερις τριών διαστάσεων συντελεστές στα επίπεδα ανάλυσης $(-1,0)$. Στους συντελεστές αυτούς έχουν κωδικοποιηθεί, για όλα τα καρέ του σήματος εισόδου, οι ακολουθίες των χωρικών συντελεστών προσέγγισης $A_{-1,0}$ καθώς και των οριζόντιων, κατακόρυφων και διαγώνιων λεπτομερειών ($D_{-1,0}^1, D_{-1,0}^2, D_{-1,0}^3$), αντίστοιχα (\mathbf{B}).

Στους συντελεστές λεπτομερειών, εφαρμόζεται διαδοχικά χρονική κυματιδική ανάλυση, οπότε προκύπτουν με τη σειρά τους ζεύγη ακολουθιών συντελεστών: ($D_{-1,-1}^p f$ & $D_{-1,-1}^p f$, $p=1,2,3$), τα οποία αναπαριστούν αντίστοιχα, την αργά και γοργά ($\mathbf{A\&D}$) μεταβαλλόμενη πληροφορία λεπτομερειών, κατά την οριζόντια, κατακόρυφη και διαγώνια διεύθυνση ($p=1,2,3$). Η αργά μεταβαλλόμενη πληροφορία λεπτομερειών $D_{-1,-1}^p$, ανατροφοδοτεί τη διαδικασία

χρονικής ανάλυσης, οπότε προκύπτουν οι συντελεστές (D_A^p $-1,-2 f$ & $D_{-1,-2}^p$, $p=1,2,3$), οι οποίοι εμφανίζονται στο διάγραμμα (Γ).

Στην ακολουθία προσέγγισης χωρικής ανάλυσης $A_{-1,0}$, πρώτου επιπέδου, εφαρμόζεται εκ νέου χωρική και χρονική ανάλυση κατά τα προαναφερθέντα, οπότε προκύπτουν οι συντελεστές πρώτου και δεύτερου επιπέδου χρονικής ανάλυσης, για επίπεδο χωρικής ανάλυσης 2 (Δ).



Detecting Video Texts Using Spatial-Temporal Wavelet Transform:

Η μεθοδολογίες που προτείνονται στην εργασία των Wang και Chen (2006), χρησιμοποιήθηκαν ως το γενικό θεωρητικό υπόβαθρο της παρούσας εργασίας. Οι συγγραφείς διαπιστώνουν ότι οι περισσότερες εργασίες στο πεδίο της εξαγωγής πληροφορίας κειμένου από βίντεο, χρησιμοποιούν χαμηλού επιπέδου χαρακτηριστικά του κειμένου, όπως το χρώμα, οι ακμές ή υφή κ.α. Στη βιβλιογραφική τους έρευνα παρουσιάζουν εργασίες, οι οποίες αξιοποιούν την ανάδειξη ακμών για τον εντοπισμό περιοχών κειμένου, χρησιμοποιώντας είτε συμβατικές μεθόδους (πυρήνες ανάδειξης ακμών sobel, canny κ.α.) [2-4], είτε μη συμβατικές μεθόδους όπως η χρήση χωρικού μετασχηματισμού κυματιδίων [5] και η αξιοποίηση χαρακτηριστικών του στατικού κειμένου στο χρόνο [6].

Η συγκεκριμένη εργασία εξετάζει τα ιδιαίτερα χαρακτηριστικά του κειμένου σε βίντεο διαπιστώνοντας ότι οι περιοχές κειμένου:

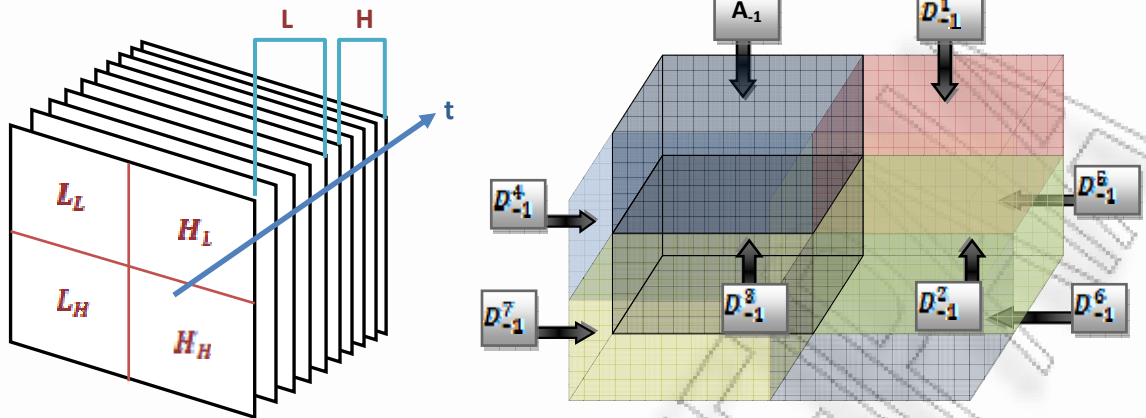
- Είναι αυστηρά ορισμένες, καθώς οι χαρακτήρες που τις απαρτίζουν συγκεντρώνονται σε πυκνά ενοποιημένα σύνολα, ενώ παράλληλα περιέχουν σε μεγάλο ποσοστό πληροφορία ακμών, η οποία αντανακλάται σε πληροφορία υψηλής συχνότητας στο χώρο.
- Διατηρούνται για ικανό αριθμό καρέ ώστε να είναι δυνατή η ανάγνωσή τους. Το στατικό κείμενο διατηρείται στο χώρο δίχως μετατόπιση, αλλοίωση ή αλλαγή μεγέθους και υποδηλώνει χαμηλής συχνότητας πληροφορία στο πεδίο του χρόνου. Αντίθετα το δυναμικό κείμενο μεταβάλλει τα χαρακτηριστικά του κατά την αλληλουχία των καρέ, ενώ υποδηλώνει υψηλής συχνότητας πληροφορία στο πεδίο του χρόνου.
- Συνήθως είναι ευθυγραμμισμένες κατά την οριζόντια ή την κατακόρυφη διεύθυνση.

Η μεθοδολογία που προτείνεται, μπορεί να κατηγοριοποιηθεί στις προσεγγίσεις μέσω υφής και περιλαμβάνει τα ακόλουθα στάδια επεξεργασίας:

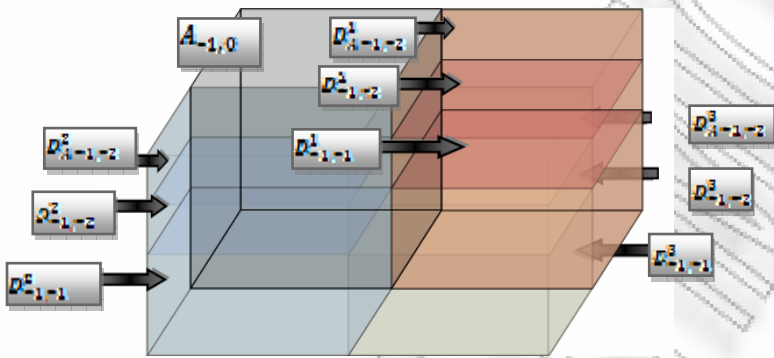
- Ανίχνευση και ανάδειξη των ακμών της πληροφορίας κειμένου.
- Επιλογή και εξαγωγή των κατάλληλων χαρακτηριστικών κειμένου.
- Ταξινόμηση των χαρακτηριστικών σε χαρακτηριστικά κειμένου ή μη.

Ανίχνευση και ανάδειξη των ακμών

Οι συγγραφείς θεωρούν ότι η συνδυασμένη χρήση κυματιδιακής ανάλυσης δύο διαστάσεων στο πεδίο του χώρου, και μιας διάστασης στο πεδίο του χρόνου, επιτρέπει αφενός την εξαγωγή υψηλής συχνότητας χωρικής πληροφορίας (ακμές κειμένου), και αφετέρου τον εντοπισμό υψηλής συχνότητας (δυναμικού) ή χαμηλής συχνότητας (στατικού) κειμένου στο πεδίο του χρόνου. Η εργασία των Burns et al., 1996, παρέχει το θεωρητικό υπόβαθρο για την υλοποίηση διαφόρων εκδόσεων της προτεινόμενης «συγχώνευσης» με στόχο την υλοποίηση 3D κυματιδιακής ανάλυσης, ενώ εκτενής αναφορά σε αυτή πραγματοποιείται στην ενότητα 3.16 *Διακριτή κυματιδιακή ανάλυση 3D σήματος*. Οι Wang και Chen προτείνουν τη χρήση του **συμβατικού αλγόριθμου ανομοιογενούς κυματιδιακής ανάλυσης τριών διαστάσεων**:



Στην παρούσα εργασία , θα προτιμήσουμε την υλοποίηση και χρήση του πολυπλοκότερου **μη συμβατικού αλγόριθμου ανομοιογενούς κυματιδιακής ανάλυσης τριών διαστάσεων**, ο οποίος επιτρέπει την αποσύνδεση του μηχανισμού χωρικού και χρονικού μετασχηματισμού, με αποτέλεσμα την δραστηκή βελτίωση της απόδοσης της κυματιδιακής ανάλυσης, όπως αναλύεται εκτενώς στην ενότητα 3.16. (Διακριτή κυματιδιακή ανάλυση 3D σήματος).



Στο ακόλουθο διάγραμμα παρουσιάζονται συγκριτικά η ομοιογενής συμβατική κυματιδιακή ανάλυση τριών διαστάσεων, δύο επιπέδων (αριστερά) και η μη συμβατική δύο χωρικών – δύο χρονικών επιπέδων (δεξιά). Η ομοιογενής προσέγγιση παράγει οκτώ ακολουθίες συντελεστών σε κάθε επίπεδο. Οι συγγραφείς προτείνουν το συνδυασμό των συντελεστών D_j^2, D_j^4, D_j^6 για την εξαγωγή χάρτη ακμών στατικού κειμένου είτε τον συνδυασμό των συντελεστών D_j^3, D_j^5, D_j^7 , για την εξαγωγή χάρτη ακμών δυναμικού κειμένου. Στους προτεινόμενους συνδυασμούς, κωδικοποιούνται αφενός οι αργά μεταβαλλόμενες λεπτομέρειες (ακμές), κατά την οριζόντια κατακόρυφη και διαγώνια διεύθυνση (στατικό κείμενο) και αφετέρου, οι αντίστοιχες γοργά μεταβαλλόμενες εκδόσεις (δυναμικό κείμενο).

$$F_z(x, y) = |a^2(x, y)| + |a^4(x, y)| + |a^6(x, y)|,$$

$$F_H(x, y) = |a^3(x, y)| + |a^5(x, y)| + |a^7(x, y)|$$

Στην παρούσα εργασία θα μελετήσουμε αποκλειστικά την εξαγωγή στατικού κειμένου, χρησιμοποιώντας τη μεθοδολογία της μη συμβατικής ανάλυσης. Αξιοποιούμε τις κατάλληλες ακολουθίες συντελεστών (αργά μεταβαλλόμενες χωρικές λεπτομέρειες στις τρεις διευθύνσεις), συνδυάζοντας τους συντελεστές $D_{j,k,p}^p = 1, 2, 3$, εν προκειμένω, για δύο επίπεδα χωρικής και δύο επίπεδα χρονικής ανάλυσης, τους συντελεστές:

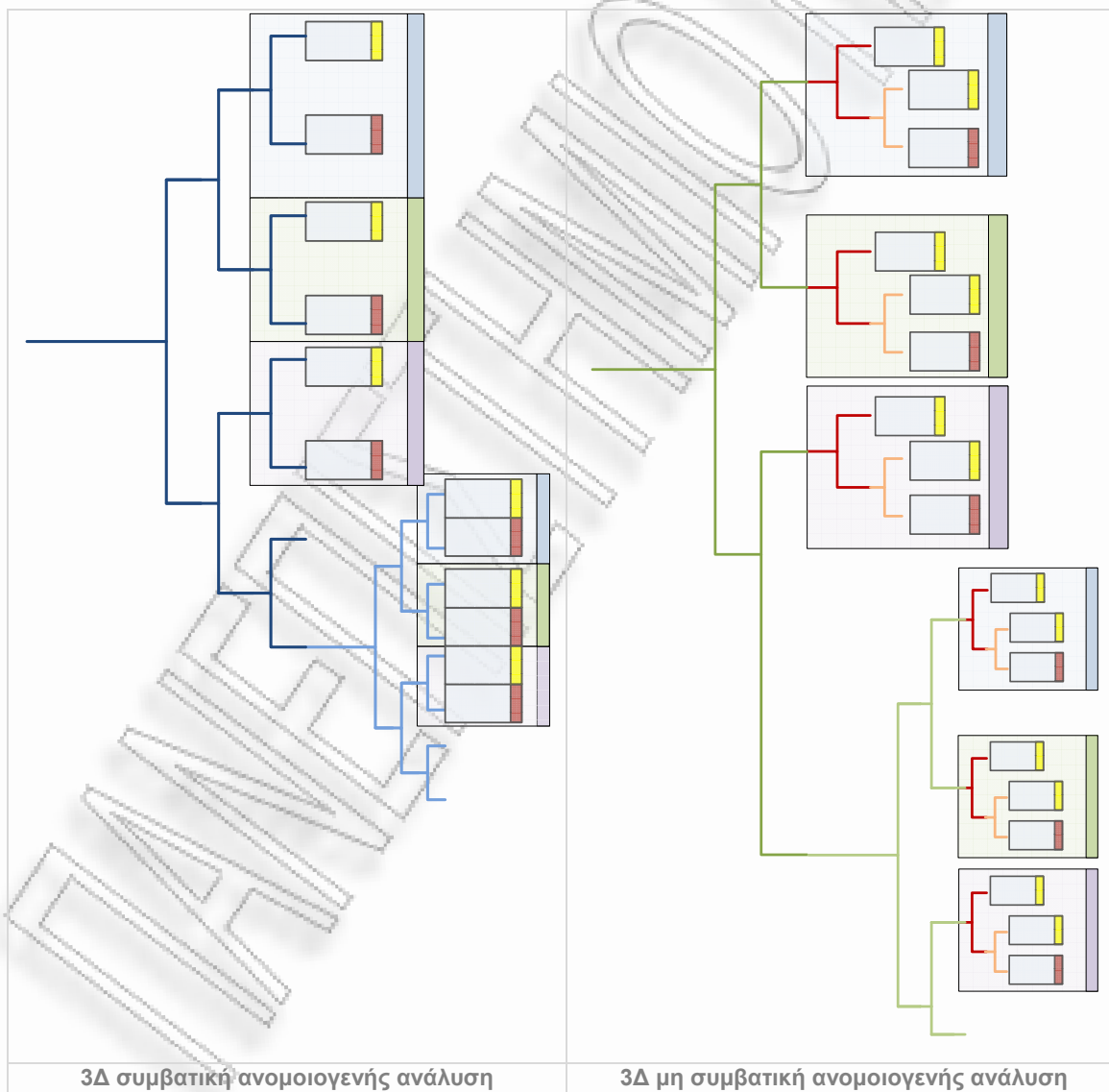
κωδικοποίηση	Επίπεδο χωρικής ανάλυσης	Χαρακτηριστικά χωρικής ανάλυσης	Επίπεδο χρονικής ανάλυσης	Χαρακτηριστικά χρονικής ανάλυσης
$D_A^1 -2,2$	2	Οριζόντιες λεπτομέρειες	2	Στατική περιγραφή
$D_A^2 -2,2$	2	Κατακόρυφες λεπτομέρειες	2	Στατική περιγραφή
$D_A^3 -2,2$	2	Διαγώνιες λεπτομέρειες	2	Στατική περιγραφή
$D_A^1 -1,2$	1	Οριζόντιες λεπτομέρειες	2	Στατική περιγραφή
$D_A^1 -1,2$	1	Κατακόρυφες λεπτομέρειες	2	Στατική περιγραφή
$D_A^1 -1,2$	1	Διαγώνιες λεπτομέρειες	2	Στατική περιγραφή

Όλες οι υπόλοιπες ακολουθίες συντελεστών, οι οποίες κωδικοποιούν μη ενδιαφέρουσες συνιστώσες του αρχικού σήματος, μηδενίζονται και κατόπιν εφαρμόζεται αντίστροφος μετασχηματισμός κυματιδίων, οπότε προκύπτει 3D αναδομημένο σήμα, ιδίων διαστάσεων με το σήμα εισόδου, από το οποίο αξιοποιούμε την απόλυτη τιμή της καταγεγραμμένης πληροφορίας, ώστε να αποκομίσουμε ένα χάρτη ακμών (saliency map) F_s , για κάθε αντίστοιχο καρέ του αρχικού σήματος. Σε κάθε τέτοιο χάρτη καταγράφονται ενισχυμένες ως προς την τιμή τους (φωτεινή ένταση) οι αντίστοιχες θέσεις του αρχικού σήματος, οι οποίες συνδυάζουν όλες τις προδιαγραφές του στατικού κειμένου (αργά κινούμενες λεπτομέρειες, με απόκριση υψηλής φωτεινής έντασης-ακμών κατά την ανάλυση και στις τρεις χωρικές διευθύνσεις), ενώ οι θέσεις οι οποίες δεν τηρούν όλες τις προϋποθέσεις καταγράφονται ως θέσεις φόντου μέσω της υποβάθμισης της τιμής της φωτεινής τους έντασης.

Χωρικές ιδιότητες συντελεστών	
	Διαγώνιες ακμές
	Κατακόρυφες ακμές
	Οριζόντιες ακμές

Χρονικές ιδιότητες συντελεστών	
	Γοργά κινούμενοι – δυναμικοί
	Αργά κινούμενοι – στατικοί

	1 ^ο επίπεδο χώρο-χρονικής ανάλυσης		1 ^ο επίπεδο χωρικής ανάλυσης
	2 ^ο επίπεδο χώρο-χρονικής ανάλυσης		2 ^ο επίπεδο χωρικής ανάλυσης
			1 ^ο επίπεδο χρονικής ανάλυσης
			2 ^ο επίπεδο χρονικής ανάλυσης



Εξαγωγή χαρακτηριστικών

Στο επόμενο βήμα επεξεργασίας, θα χρειαστεί να εξαχθούν χαρακτηριστικές τιμές με βάση την 3D πληροφορία του χάρτη ακμών που παρήχθη στο προηγούμενο στάδιο. Οι χαρακτηριστικές τιμές που προκύπτουν, χρησιμοποιούνται ώστε να περιγράψουν τις ιδιότητες διακριτών τάξεων (classes), οι οποίες εξάγονται στο επόμενο στάδιο, αυτό της ταξινόμησης (classification). Διανύσματα χαρακτηριστικών με παρόμοιες τιμές ανήκουν στην ίδια κλάση, ιδιότητες όπως η ένταση, η αντίθεση και η κατευθυντικότητα, αποτελούν τα κριτήρια για το διαχωρισμό των περιοχών κειμένου από τις περιοχές υποβάθρου.

Τα διανύσματα χαρακτηριστικών εξάγονται μέσω επεξεργασίας της **μήτρας συνάφειας** (co-occurrence matrix), η οποία προκύπτει με τη σειρά της από την ανάλυση του χάρτη ακμών, και περιγράφουν την υφή των αναλυόμενων περιοχών.

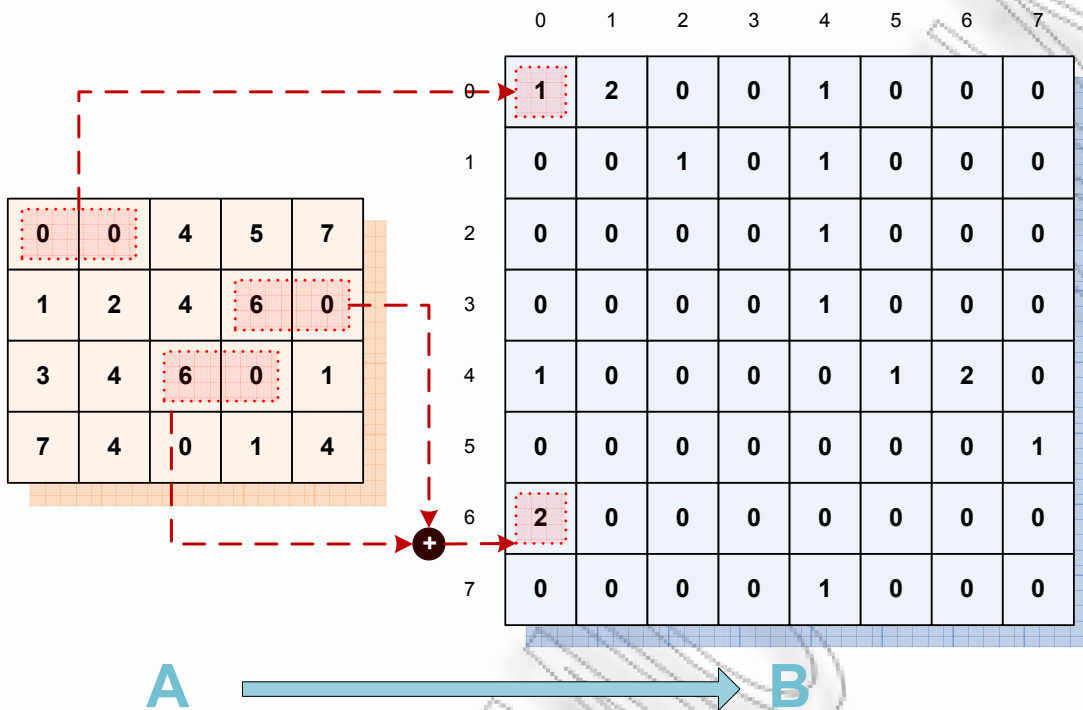
Η μήτρα συνάφειας M_d ή μήτρα διασποράς στην κλίμακα του γκρι (Gray Level Co-occurrence Matrix - GLCM), είναι το αποτέλεσμα μιας μεθοδολογίας ανάλυσης (Haralick et al, 1973) για τον εντοπισμό και την κωδικοποίηση συνδυασμού θέσεων πληροφορίας εισόδου, οι οποίες συσχετίζονται μέσω ορισμένης χωρικής συσχέτισης (διεύθυνση - απόσταση), ώστε οι τιμές τους να ταυτίζονται με την τιμή των δεικτών θέσης της μήτρας συνάφειας. Από τη στατιστική ανάλυση κάθε μήτρας GLCM, η οποία προκύπτει από διαδοχικά ληφθέντα χωρικά δείγματα του χάρτη ακμών, προκύπτουν διανύσματα χαρακτηριστικών τα οποία με τη σειρά τους χαρακτηρίζουν τις αντίστοιχες περιοχές δειγματοληψίας.

Ο υπολογισμός της μήτρας συνάφειας (GLCM) πληροφορίας εικόνας δύο διαστάσεων $I_{n \times m}$ για διάνυσμα χωρικής απόκλισης $d = (\Delta r, \Delta c)$, όπου Δr =row distance, Δc =column distance, ορίζεται μαθηματικά ως εξής:

$$M_{\Delta r, \Delta c}(i, j) = \sum_{p=1}^n \sum_{q=1}^m \begin{cases} 1, & \text{if } I(p, q) = i \text{ and } I(p + \Delta r, q + \Delta c) = j \\ 0, & \text{otherwise} \end{cases}$$

Παράδειγμα υπολογισμού 2D GLCM:

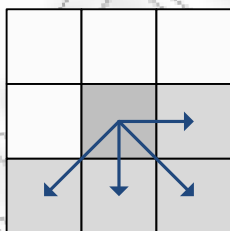
Για πληροφορία εισόδου $I_{4 \times 5}$ (A) και $(\Delta r, \Delta c) = (0, 1)$, προκύπτει η μήτρα συνάφειας GLCM M_d (B).



Κάθε στοιχείο της μήτρας (i,j) GLCM ορίζει τον αριθμό των περιπτώσεων κατά τις οποίες ένα στοιχείο του πίνακα εισόδου με τιμή i εμφανίζεται οριζόντια συσχετιζόμενο και σε απόσταση μιας θέσης από ένα στοιχείο με τιμή j .

Παρατηρούμε ότι το μέγεθος της μήτρας συνάφειας S_{GLCM} εξαρτάται αποκλειστικά από το μήκος της πληροφορίας (n bits) που απαιτείται για να κωδικοποιηθούν οι θέσεις εισόδου $I_{n \times m}$: $S_{GLCM} = 2^n * 2^m$. Για εικόνα κωδικοποιημένη στη κλίμακα του γκρι (8bit) απαιτείται μήτρα μεγέθους $S_{GLCM} = 2^8 * 2^8$, οι συγγραφείς (Wang και Chen) προτείνουν για λόγους υπολογιστικής απόδοσης την ελάττωση της πληροφορίας εισόδου στα 3 bit (8 στάθμες του γκρι), πρόταση που θα ακολουθήσουμε και στην παρούσα εργασία.

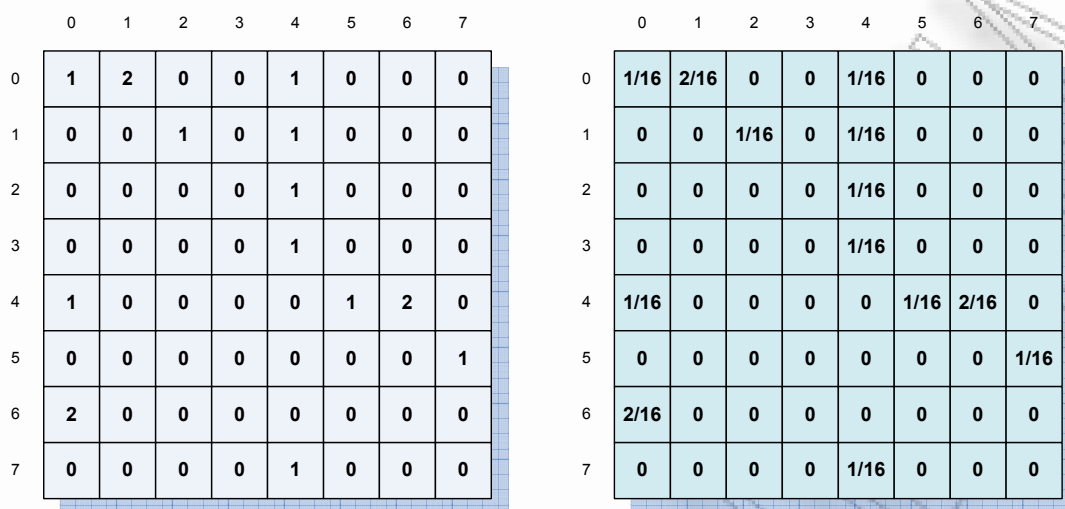
Το διάνυσμα χωρικής απόκλισης (distance vector) d , για τις γειτονικές θέσεις ενός εικονοστοιχείου ενδιαφέροντος κωδικοποιείται ως εξής:



Από τη μήτρα συνάφειας προκύπτει η μήτρα συνδεδεμένων πιθανοτήτων (joint probability matrix) P_d , διαιρώντας κάθε στοιχείο του M_d με το συνολικό αριθμό συναφών ζευγών N που καταγράφονται σε αυτόν:

$$N = \sum_{i=0 \sim 7} \sum_{j=0 \sim 7} M_d(i,j) \quad \text{ενώ} \quad P_d = \frac{M_d(i,j)}{N}$$

Στη μήτρα συνδεδεμένων πιθανοτήτων (Γ) καταγράφεται σε κάθε θέση η πιθανότητα ύπαρξης, του αντίστοιχου με τη μήτρα συνάφειας ζεύγους τιμών.



Όπου
$$N = \sum_{i=0}^7 \sum_{j=0}^7 M_2(i,j) = 10$$

Για τις ανάγκες του εντοπισμού υψής κειμένου, οι συγγραφείς προτείνουν την εξαγωγή διανυσμάτων χαρακτηριστικών από τρεις μήτρες συνδεδεμένων πιθανοτήτων για κάθε δείγμα του χάρτη ακμών, στις διευθύνσεις:

Διάνυσμα απόστασης	Μοίρες	Γειτονικά στοιχεία κατά την κάτωθι διεύθυνση
$d(0,1)$	0°	Οριζόντια
$d(1,0)$	90°	Κατακόρυφη
$d(1,1)$	45°	Διαγώνια

Για την τελική εξαγωγή των χαρακτηριστικών πραγματοποιείται στατιστική ανάλυση των τιμών κάθε μήτρας συνδεδεμένων πιθανοτήτων, υπολογίζοντας τα εξής μεγέθη:

1.	Μέγιστη Πιθανότητα	Maximum Probability	$M = \max_{i,j} F_d(i,j)$
2.	Ενέργεια	Energy	$E = \sum_{i=0}^7 \sum_{j=0}^7 F_d^2(i,j)$
3.	Εντροπία	Entropy	$e = \sum_i \sum_j (P_d(i,j) \log P_d(i,j))$

Το μέτρο της μέγιστης πιθανότητας $M = \max_{i,j} P_a(i,j)$, καταγράφει την τιμή της μέγιστης συχνότητας εμφάνισης ζεύγους τιμών σε δεδομένο παράθυρο εισόδου. Υψηλές τιμές της υπονοούν την κυριαρχία ενός ζεύγους τιμών έναντι των υπολοίπων.

$$E = \sum_{i=0}^7 \sum_{j=0}^7 P_a^2(i,j)$$

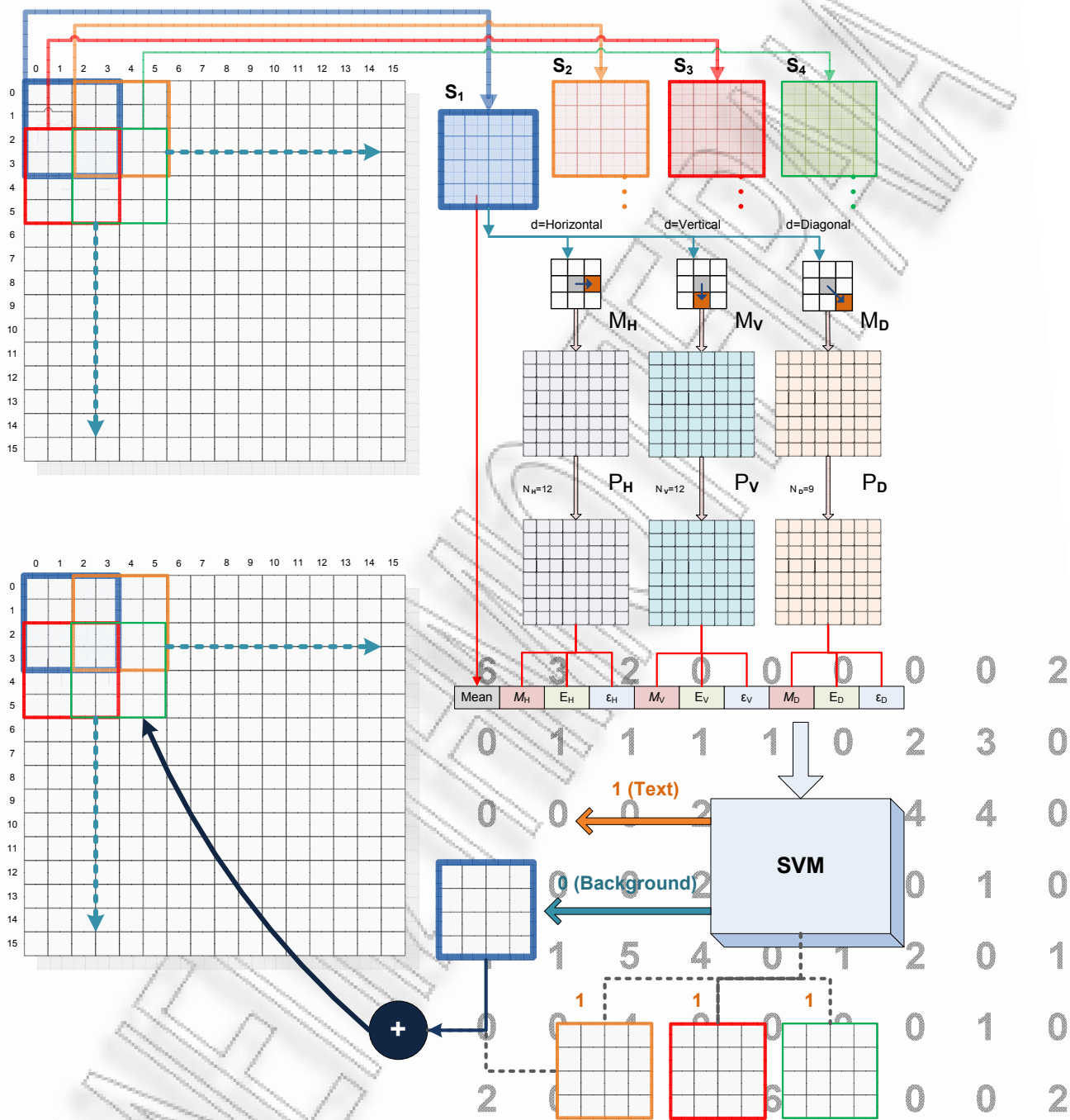
Το μέτρο της ενέργειας, το οποίο καλείται επίσης Angular Second Moment (ASM) είναι ένα μέγεθος συσχετιζόμενο με την κανονικότητα, και κατά τον υπολογισμό του, η τιμή κάθε καταγραφής $P_a(i,j)$, χρησιμοποιείται ως βάρος του εαυτού της.

$$e = \sum_{i=0}^7 \sum_{j=0}^7 (P_a(i,j) \log P_a(i,j))$$

Το μέτρο της εντροπίας, είναι ένα μέγεθος συσχετιζόμενο με το βαθμό της χωρικής αταξίας των τιμών της μήτρας συνάφειας.

Παράλληλα με τον υπολογισμό και την καταγραφή των χαρακτηριστικών από κάθε μήτρα συνδεδεμένων πιθανοτήτων (3x3), στο διάνυσμα χαρακτηριστικών καταχωρείται επιπρόσθετα και η μέση τιμή φωτεινότητας των τιμών του παραθύρου εισόδου, θεωρώντας ότι αυτή θα είναι υψηλότερη για τα δείγματα περιοχών κειμένου.

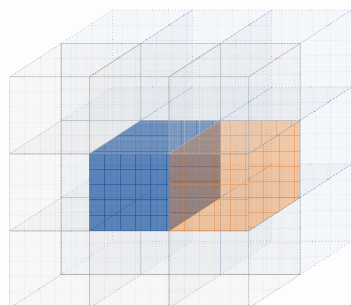
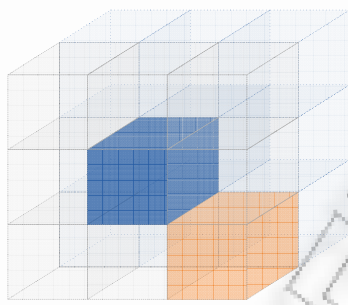
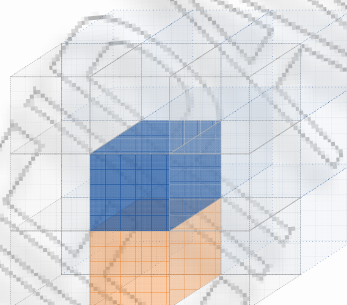
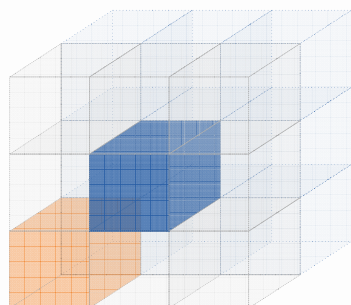
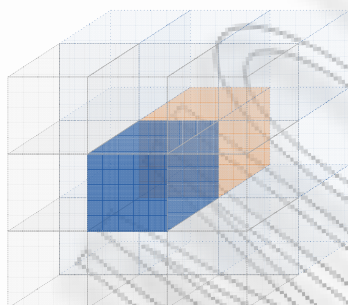
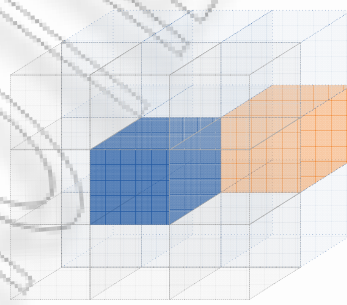
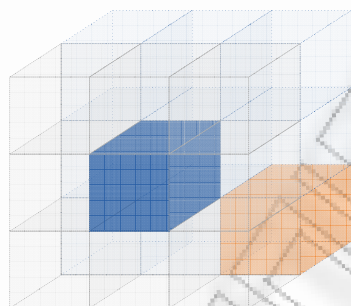
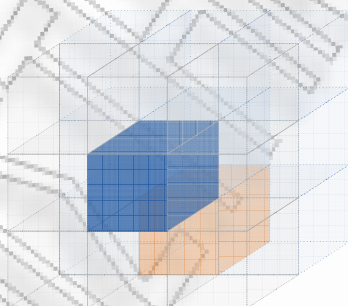
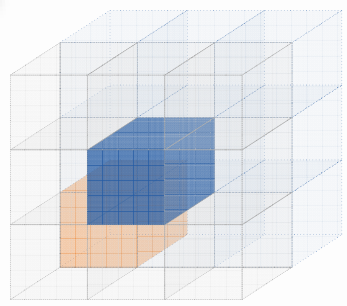
Όλα τα βήματα της μεθοδολογίας 2Δ ανάλυσης και εξαγωγής χαρακτηριστικών μέσω χρήσης μητρών συνάφειας, καθώς και η διαδικασία ταξινόμησης σε τάξεις μέσω χρήσης SVM, παρουσιάζεται στο ακόλουθο αριθμητικό παράδειγμα υπολογισμού:



Στην παρούσα εργασία προτείνουμε μια διαφορετική προσέγγιση εξαγωγής χαρακτηριστικών από εκείνη των Wang και Chen, προτιμώντας να αξιοποιήσουμε μήτρα GLCM ανάλυσης τρισδιάστατου σήματος (Fuaη Tsai et al, 2007). Με αυτό τον τρόπο αξιοποιούμε τη συσχέτιση των τιμών φωτεινότητας στη διάσταση του χρόνου, έχοντας κάνει ήδη την παραδοχή ότι θα αναλύσουμε χρονοσειρές εικόνων στις οποίες έχει εντυπωθεί στατικό κείμενο.

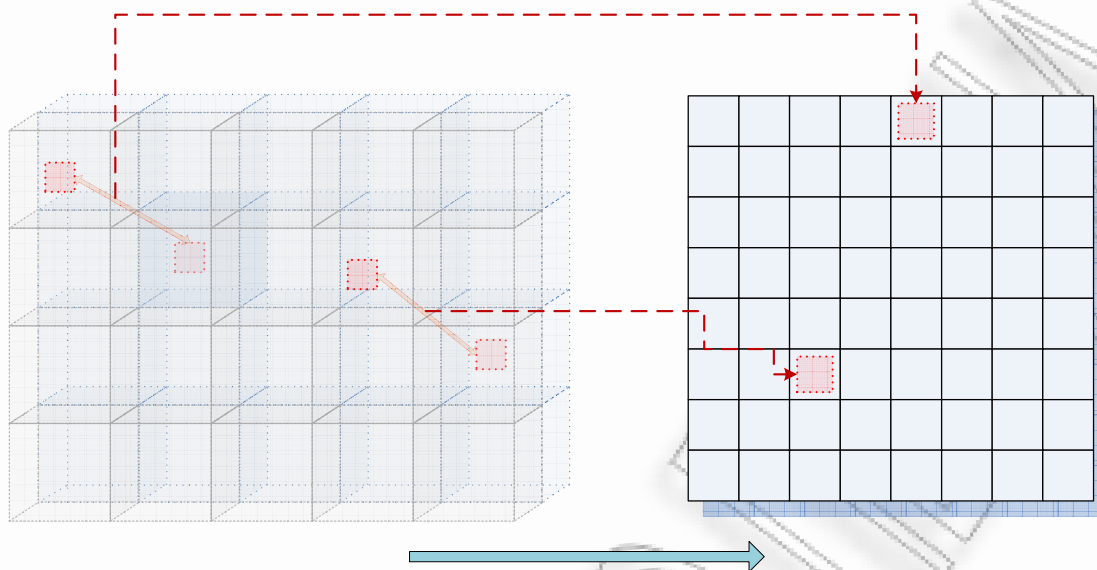
Με τον όρο «**3D μήτρα συνάφειας**» (**3D GLCM**), αναφερόμαστε σε μήτρα δύο διαστάσεων η οποία κωδικοποιεί τη συσχέτιση τιμών συνδυασμού θέσεων, οι οποίες έχουν συγκεκριμένη χωρική συσχέτιση στον τρισδιάστατο χώρο.

Το διάνυσμα 3D χωρικής απόκλισης (3D distance vector) $\mathbf{d} = (dx, dy, dz)$, για τις γειτονικές θέσεις ενός εικονοστοιχείου ενδιαφέροντος 3D πληροφορίας κωδικοποιείται ως εξής:

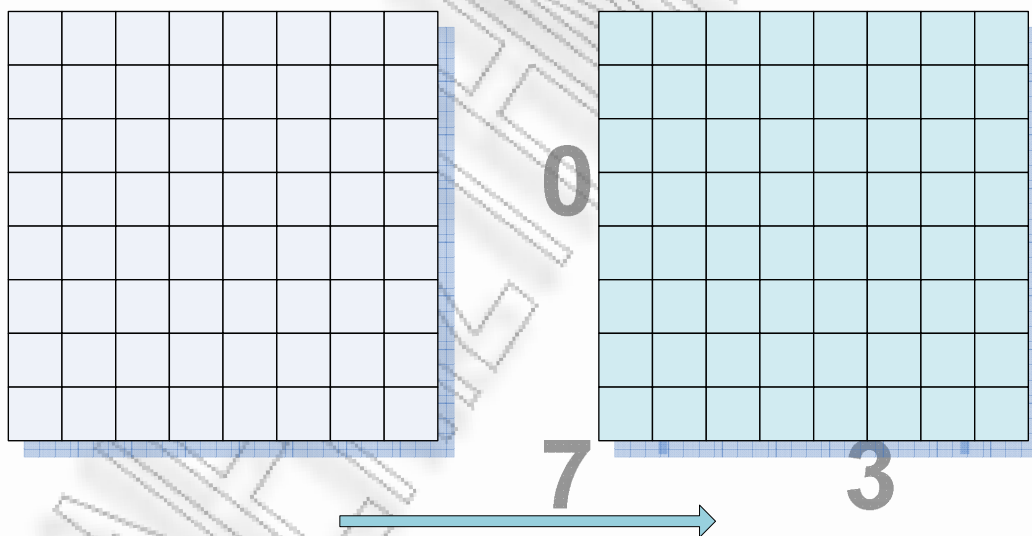

 $\mathbf{d}=(0,1,0)$

 $\mathbf{d}=(1,1,0)$

 $\mathbf{d}=(1,0,0)$

 $\mathbf{d}=(1,-1,0)$

 $\mathbf{d}=(0,0,1)$

 $\mathbf{d}=(0,1,1)$

 $\mathbf{d}=(1,1,1)$

 $\mathbf{d}=(1,0,1)$

 $\mathbf{d}=(1,-1,1)$

Παράδειγμα υπολογισμού 3D GLCM:

Για πληροφορία εισόδου $4 \times 4 \times 2$ (A) και $\mathbf{d} = (\Delta x, \Delta y, \Delta z) = (0, 1, 1)$, προκύπτει η μήτρα συνάφειας GLCM \mathbf{M}_d (B).



Ενώ η μήτρα συνδεδεμένων πιθανοτήτων (Γ) προκύπτει ως εξής:



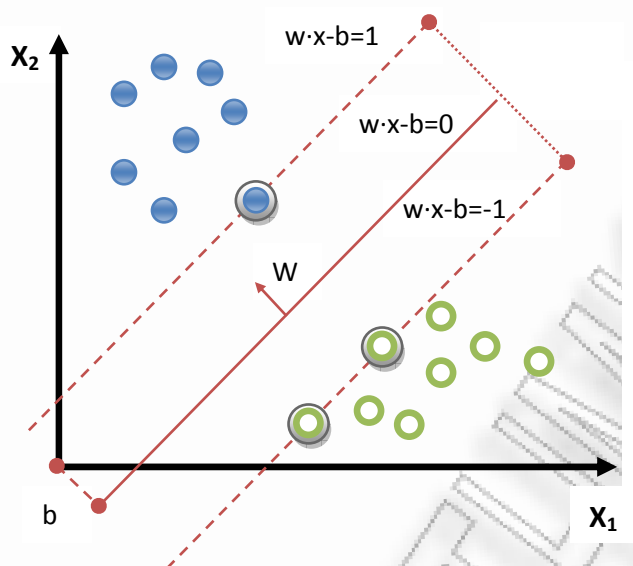
Στην εφαρμογή μας και με το πέρας της επεξεργασίας μέσω 3D μήτρας GLCM, προκύπτει για κάθε δείγμα εισόδου, διάνυσμα χαρακτηριστικών $V_{28 \times 1}$. Κάθε τέτοιο διάνυσμα θα πρέπει να ταξινομηθεί σε μια από τις δύο επιθυμητές κλάσεις (υπόβαθρο, κείμενο).

Ταξινόμηση:

Η ταξινόμηση των διανυσμάτων χαρακτηριστικών πραγματοποιείται με χρήση Ταξινομητών Πυρήνα (Support Vector Machines - SVM), (Gunn S.R., 1992). Οι ταξινομητές πυρήνα αποτελούνται από ένα σύνολο συσχετιζόμενων μεθόδων μηχανικής μάθησης με επίβλεψη, ενώ

αξιοποιούνται για την ταξινόμηση (classification) ή την παλινδρόμηση (regression) δεδομένων εισόδου.

Για την χρήση Ταξινομητή Πυρήνα, καταρχάς απαιτείται η εκπαίδευσή του με δεδομένα εισόδου για τα οποία γνωστοποιείται η κατηγορία στην οποία ανήκουν (training set data). Στη συνέχεια οι αλγόριθμοι υποβάθρου διαμορφώνουν ένα μοντέλο το οποίο αντανάκλα την σημειακή απεικόνιση των δεδομένων δοκιμής στο επίπεδο, με τρόπο ώστε αυτά να διαχωρίζονται χωρικά κατά κατηγορία με όσο το δυνατόν μεγαλύτερο διάκενο (maximum-margin hyperplane) . Όταν αυτό το μοντέλο διαμορφωθεί, ο ταξινομητής (linear classifier) είναι έτοιμος να υποδεχθεί άγνωστης τάξεως διανύσματα, να τα απεικονίσει στον ίδιο χώρο ως σημεία, και να προβλέψει την κατηγορία στην οποία ανήκουν.



Στο διάγραμμα (αριστερά), παρατηρούμε τις σημειακές αναπαραστάσεις διανυσμάτων χαρακτηριστικών, τα οποία ανήκουν σε δύο κλάσεις και προβάλλονται σε επίπεδο, μέσω εκπαιδευμένου SVM, με τέτοιο τρόπο ώστε να διαχωρίζονται κατά το μέγιστο. Τα διανύσματα τα οποία απεικονίζονται εκατέρωθεν, στα όρια του διακένου αποκαλούνται διανύσματα υποστήριξης (support vectors).

Εξαγωγή κειμένου:

Κατά το τελικό στάδιο της επεξεργασίας, εφαρμόζεται ένα σύνολο μεθοδολογιών, βασιζόμενο στην εργασία των Wang και Chen, το οποίο όμως εμπλουτίζουμε και επεκτείνουμε ώστε να επιτύχουμε την εξαγωγή του τελικού προϊόντος, το οποίο δεν είναι άλλο από μια δυαδική εικόνα, όπου απεικονίζεται διακριτά το μηχανικά εντύπωμένο κείμενο (τιμή 1) και το υπόβαθρο (τιμή 0), με τρόπο τέτοιο, ώστε να είναι δυνατή η ορθή αναγνώρισή του από αλγόριθμο οπτικής αναγνώρισης χαρακτήρων (OCR).

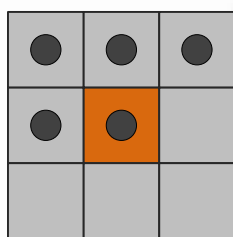
Κατά το τελικό στάδιο επεξεργασίας χρησιμοποιούμε ως πληροφορία εισόδου, το αρχικό σήμα βίντεο στην κλίμακα του γκρι, τον χάρτη ακμών ο οποίος προήλθε από ανάλυση επεξεργασία και σύνθεση μέσω της θεωρίας κυματιδιακής ανάλυσης σήματος 3D, καθώς και τη δισδιάστατη εικόνα η οποία αντανάκλα την ταξινόμηση της υψής, η οποία προέκυψε από την ανάλυση υψής μέσω 3D GLCM και την εν συνεχεία ταξινόμησή της από τον εκπαιδευμένο ταξινομητή SVM. Για να αξιοποιήσουμε την προαναφερθείσα πληροφορία εισόδου ώστε να επιτύχουμε το επιθυμητό αποτέλεσμα, συνδυάζουμε μεθοδολογίες μορφολογικής ανάλυσης και επεξεργασίας από το πεδίο της μηχανικής όρασης (machine vision) όπως: Τυποποίηση Συνδεδεμένων Στοιχείων (Connected Components Labeling), καθώς και Μορφολογική Διάβρωση (Erosion), Διαστολή (Dilation), Άνοιγμα (Opening), Κλείσιμο (Closing), αλλά και μεθοδολογίες όπως η ανάλυση ιστογράμματος κ.α., οι οποίες συνδυάστηκαν εμπειρικά με θετικά αποτελέσματα. Ακολουθεί συνοπτική παρουσίαση των προαναφερθέντων μεθοδολογιών:

Τυποποίηση Συνδεδεμένων Στοιχείων

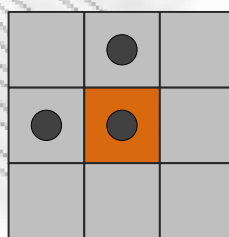
Η τυποποίηση συνδεδεμένων στοιχείων ή ανάλυση συνδεδεμένων στοιχείων (Connected Component Labeling – Connected Components Analysis), είναι μια αλγοριθμική εφαρμογή της θεωρίας γραφημάτων (graph theory) η οποία βρίσκει εφαρμογή στο πεδίο της μηχανικής όρασης και αποσκοπεί στην ανίχνευση και σηματοδότηση συνδεδεμένων περιοχών δυσδιάστατων δυαδικών εικόνων (Hopcroft and Tarjan, 1973; Samet, 1981), η χρήση της όμως μπορεί όμως να επεκταθεί στην ανάλυση έγχρωμων εικόνων αλλά και δεδομένων εισόδου υψηλότερων διαστάσεων (Lumina, 1983; Samet and Tamminen, 1988).

Ο αλγόριθμος για την τυποποίηση συνδεδεμένων στοιχείων απαιτεί ένα δομικό στοιχείο (structuring element), το οποίο ορίζει τη σχέση χωρικής γειτονίας, βάσει της οποίας θα εξεταστεί η τιμή του εξεταζόμενου εικονοστοιχείου (σάρωση ανά στήλες & γραμμές) με τα οριζόμενα ως γειτονικά του, ώστε να θεωρηθεί ότι υφίσταται σύνδεση. Εάν υπάρχουν γειτονικά στοιχεία τα οποία έχουν σημανθεί, το εξεταζόμενο αποκτά την μικρότερη τιμή σήμανσης που ανιχνεύτηκε στην γειτονία του, διαφορετικά του αποδίδεται η αμέσως επόμενη από εκείνες που έχουν ήδη εκχωρηθεί (**B**).

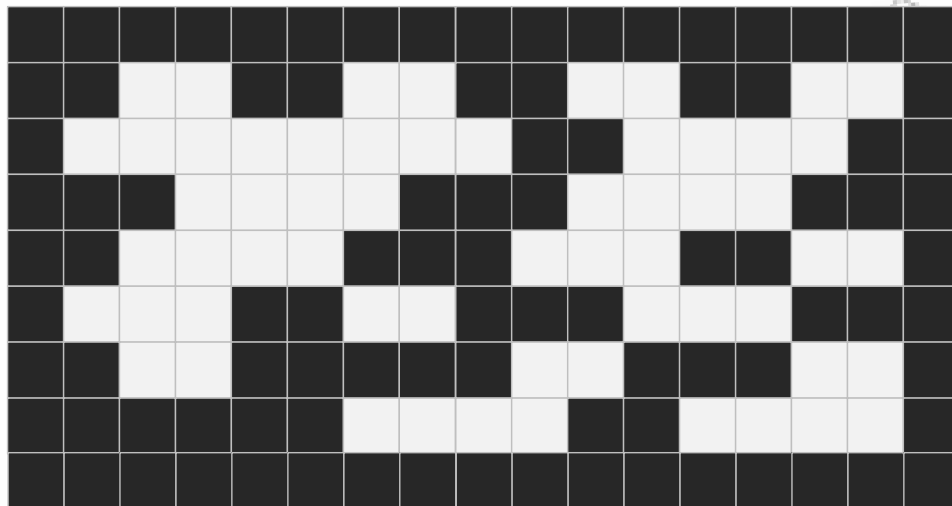
Σε δεύτερη φάση, ο αλγόριθμος σαρώνει μόνο τα δεδομένα τα οποία έχουν ήδη σηματοδοτηθεί και τους αποδίδει τη χαμηλότερη τιμή σήμανσης (**Γ**). Στα ακόλουθα διαγράμματα παρατηρούμε τη διαδικασία CCL διπλής σαρώσεως, για δυαδική εικόνα εισόδου $I_{9 \times 17}$ (**A**), με δομικό στοιχείο 8 διασυνδέσεων (8-connectivity):



Διασύνδεση 8 διασυνδέσεων
(8-connectivity)



Διασύνδεση 4 διασυνδέσεων
(4-connectivity)



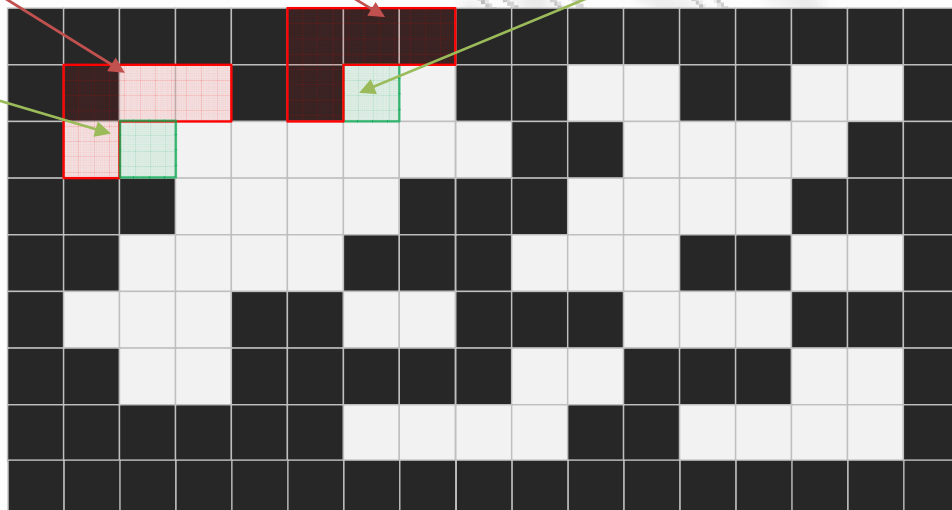
Εικόνα εισόδου (A)

8-connectivity
 $\min(v(i,j) > 0) = 1$

8-connectivity
 $\min(v(i,j) > 0) = \text{NaN}$

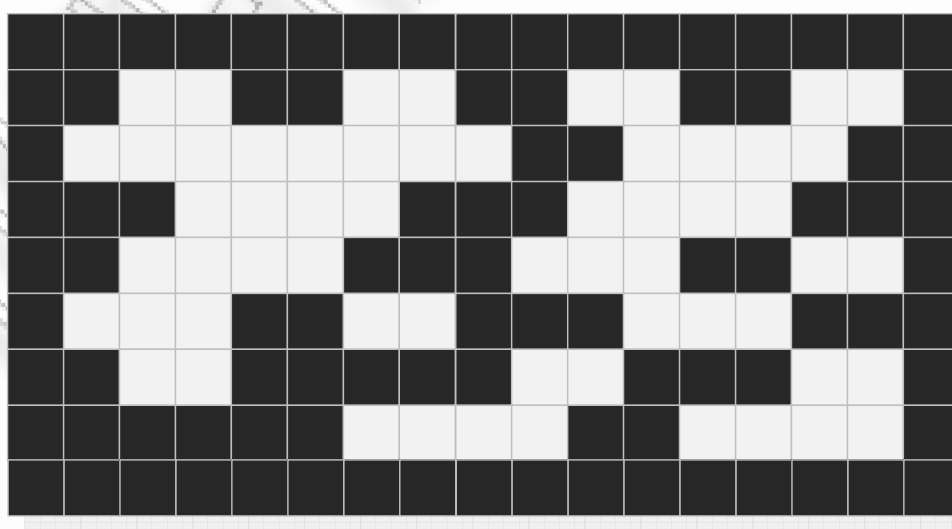
Εξεταζόμενο
 εικονοστοιχείο

Εξεταζόμενο
 εικονοστοιχείο



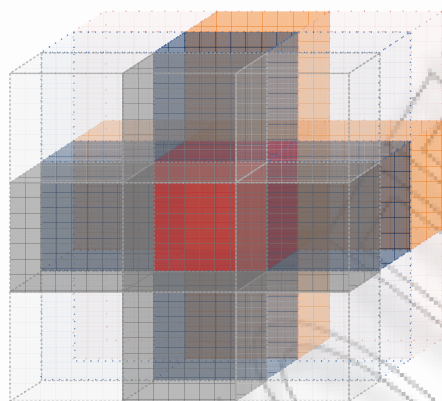
1^η σάρωση (B)

1	1
1	1
	1
1	1
1	1
1	1



2^η σάρωση (Γ)

Η εφαρμογή του αλγορίθμου μπορεί εύκολα να επεκταθεί στην επεξεργασία πολυδιάστατων δεδομένων αξιοποιώντας δομικά στοιχεία ανάλογων διαστάσεων. Στην παρούσα εργασία υλοποιούμε τυποποίηση συνδεδεμένων στοιχείων τριών διαστάσεων, χρησιμοποιώντας για τον έλεγχο των συνδεδεμένων εικονοστοιχείων το δομικό στοιχείο που απεικονίζεται στο ακόλουθο διάγραμμα:

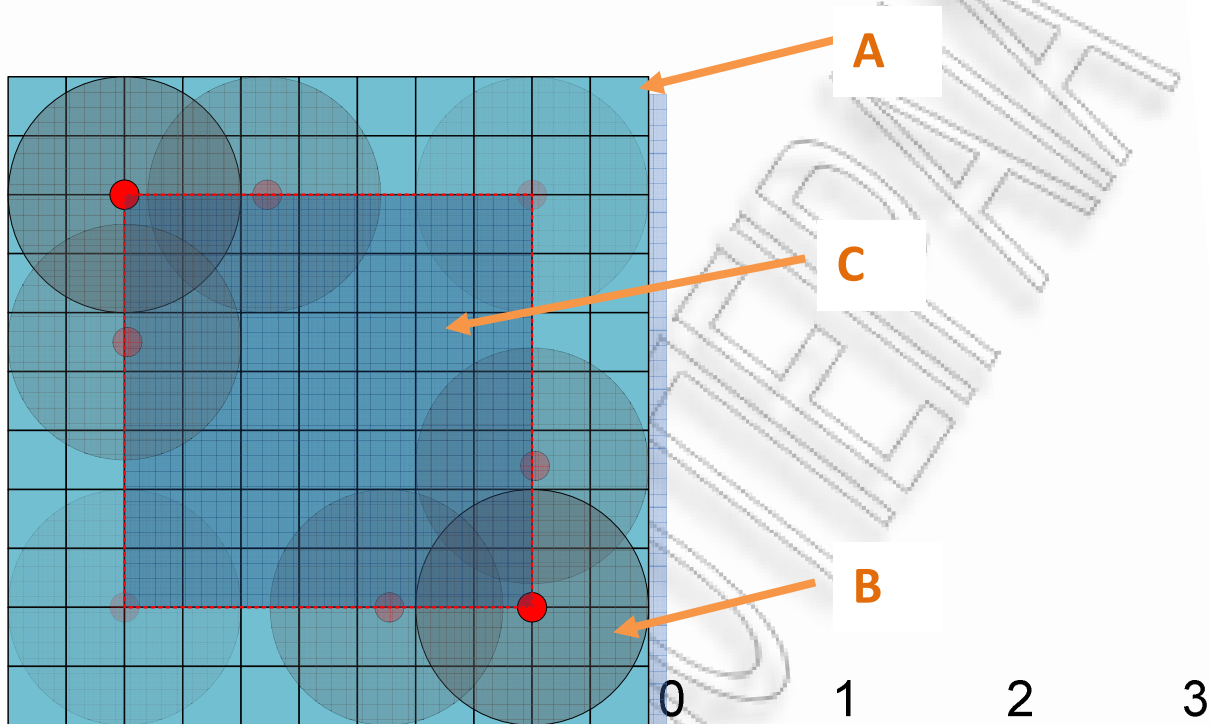


Cross shaped 3D structure element

Μορφολογικές Διεργασίες

Η Μαθηματική Μορφολογία (Mathematical Morphology-MM), ορίζεται ως μια θεωρία και σύνολο τεχνικών, για την ανάλυση και επεξεργασία γεωμετρικών δομών. Η συγκεκριμένη θεωρία εφαρμόζεται στην επεξεργασία ψηφιακών εικόνων. Όπως και στην μεθοδολογία τυποποίησης συνδεδεμένων στοιχείων, έτσι και στις μορφολογικές διεργασίες απαιτείται η χρήση ενός προκαθορισμένου σχήματος (δομικού στοιχείου) για την ανάλυση των δεδομένων εισόδου. Το οριζόμενο δομικό στοιχείο σαρώνει την εικόνα εισόδου και ο τρόπος με το οποίο αλληλεπιδρά χωρικά με τις γεωμετρικές δομές της, ορίζει το αποτέλεσμα. Οι βασικοί τελεστές της Μαθηματικής Μορφολογίας για εικόνα **A** και δομικό στοιχείο **B** είναι οι ακόλουθοι:

Μορφολογική Διάβρωση (Erosion)

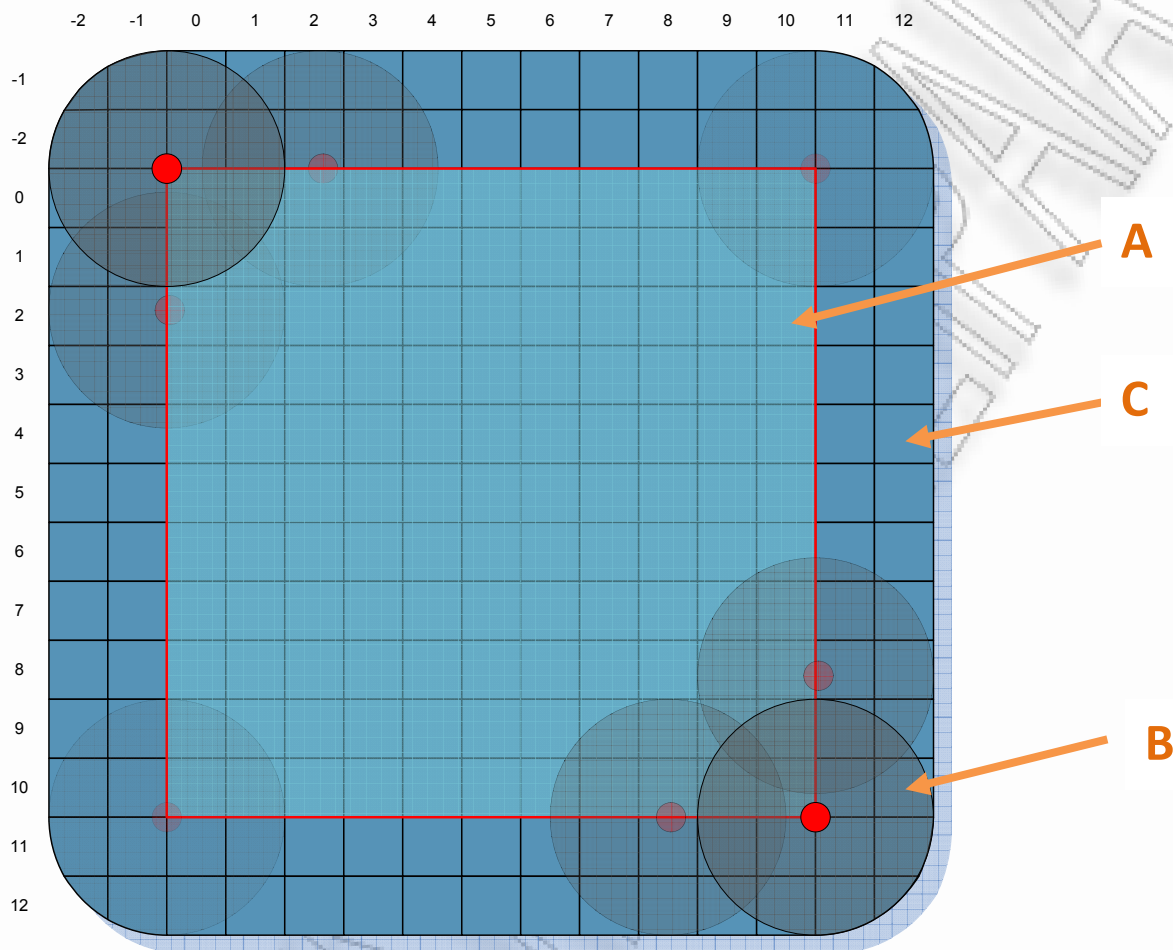


Όπως παρατηρούμε στο παραπάνω σχήμα η μορφολογική διάβρωση σχήματος A από δομικό στοιχείο B, ορίζεται από το γεωμετρικό τόπο των σημείων που διατρέχει το κέντρο του δομικού στοιχείου, όταν αυτό κινείται εντός του σχήματος A:

$$A \ominus B = \bigcap_{B \subseteq A} A_x$$

Το αποτέλεσμα της μορφολογικής διάβρωσης είναι η εκλέπτυνση των γραμμικών τμημάτων και η διεύρυνση των διάκενων γεωμετρικών σχημάτων (όπως οι αλφαριθμητικοί χαρακτήρες).

Μορφολογική Διαστολή (Dilation)



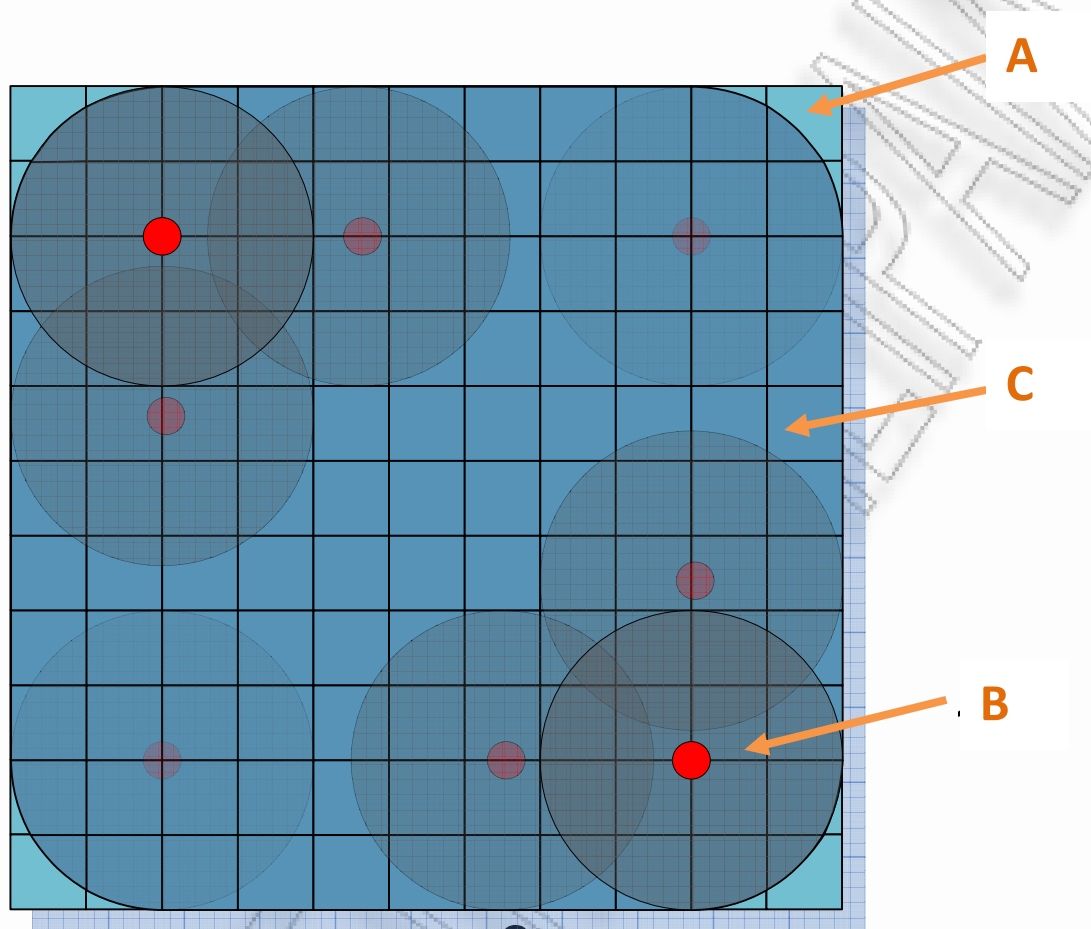
Όπως παρατηρούμε στο παραπάνω σχήμα η μορφολογική διαστολή σχήματος A από δομικό στοιχείο B, ορίζεται από το γεωμετρικό τόπο των σημείων που καταλαμβάνει το B όταν το κέντρο του κινείται εντός του σχήματος A:

$$A \oplus B = A_2$$

Το αποτέλεσμα της μορφολογικής διαστολής είναι το αντίθετο της διάβρωσης, δηλαδή πολύ λεπτές γραμμές γίνονται πιο αδρές.

Θα χρησιμοποιήσουμε δύο ακόμα παραγόμενους τελεστές μορφολογικής ανάλυσης, οι οποίοι περιγράφονται συνοπτικά ακολούθως:

Μορφολογικό Άνοιγμα (Opening)



2

Το Μορφολογικό Άνοιγμα σχήματος A από δομικό στοιχείο B ορίζεται ως το αποτέλεσμα της εφαρμογής Μορφολογικής Διάβρωσης στο σχήμα A με χρήση δομικού στοιχείου B, ακολουθούμενης από την εφαρμογή Μορφολογικής Διαστολής στο αποτέλεσμά της με εκ νέου χρήση του ίδιου δομικού στοιχείου:

$$A \circ B = (A \ominus B) \oplus B = \bigcup_{F \in B} A_F$$

1

Το αποτέλεσμα του μορφολογικού ανοίγματος, είναι η εξάλειψη των μικρής διαστάσεως λεπτομερειών (πιθανός θόρυβος) του αναλυόμενου σχήματος, ακολουθούμενη από ανάδειξη της βασικής δομής του αλλά και ταυτόχρονη παραμόρφωσή των λεπτομερειών (απαλοιφή ακμών).

2

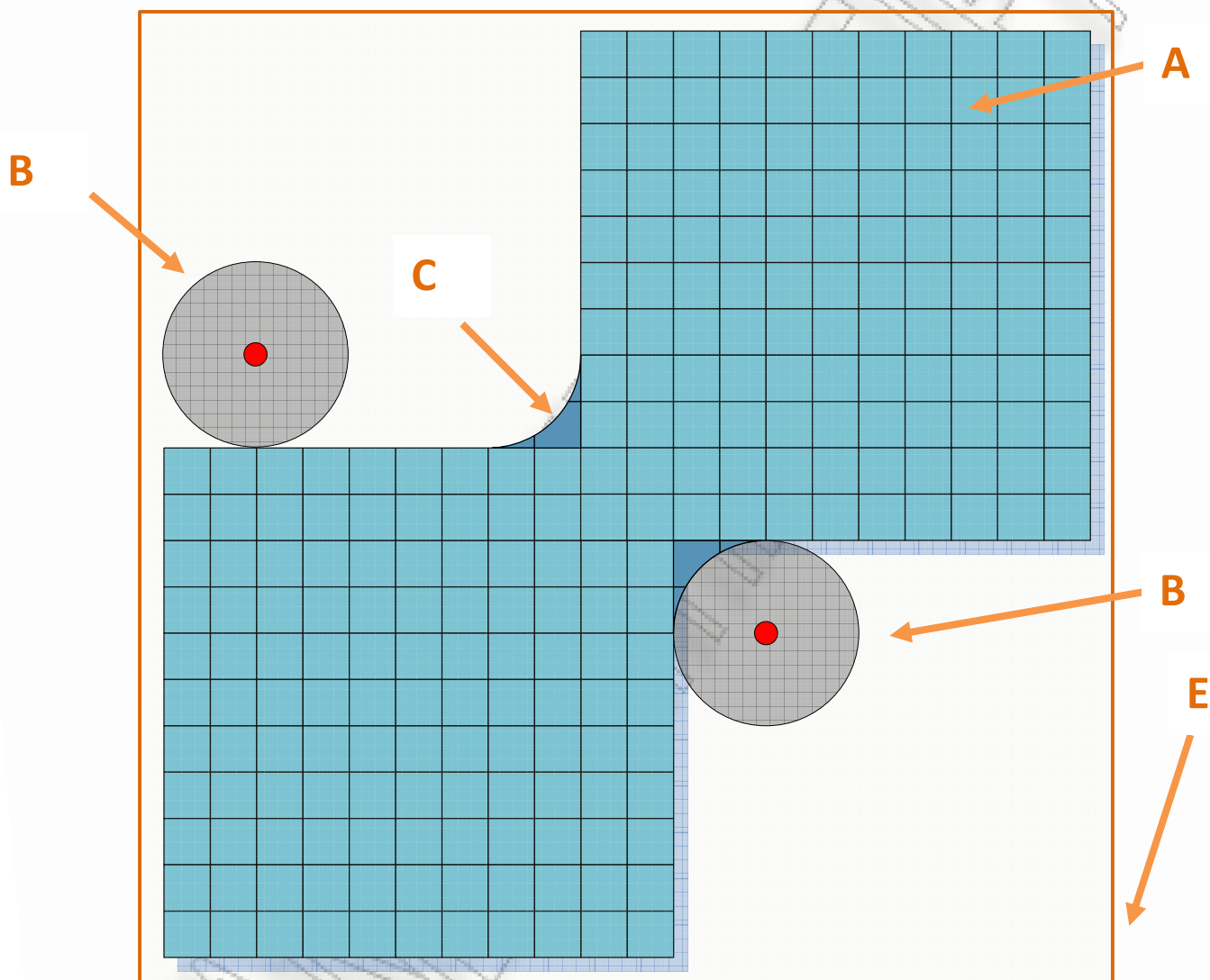
Μορφολογικό Κλείσιμο (Closing)

Το Μορφολογικό Κλείσιμο σχήματος A από δομικό στοιχείο B ορίζεται ως το αποτέλεσμα της εφαρμογής Μορφολογικής Διαστολής στο σχήμα A με χρήση δομικού στοιχείου B, ακολουθούμενης από την εφαρμογή Μορφολογικής Διάβρωσης στο αποτέλεσμά της με εκ νέου χρήση του ίδιου δομικού στοιχείου:

3

$$A \circ B = (A \oplus B) \ominus B = (A^c \circ B^c)^c$$

Η προηγούμενη μαθηματική σχέση υποδηλώνει ότι το αποτέλεσμα του Μορφολογικού Κλεισίματος σχήματος A , ισούται με το συμπλήρωμα των γεωμετρικών τόπων που καταλαμβάνει το B , όταν αυτό κινείται εκτός του σχήματος A , αλλά εντός του Ευκλείδειου Χώρου E που περιέχει το A , το οποίο εμφανίζεται στο ακόλουθο σχήμα ως η συνένωση των πράσινων (A) και των μπλε περιοχών.



Διάγραμμα διαδικασιών εξαγωγής κειμένου:

Στο ακόλουθο διάγραμμα, παρουσιάζονται οι μεθοδολογίες που χρησιμοποιούμε στην παρούσα εργασία, κατά την τελική διαδικασία της εξαγωγής κειμένου:

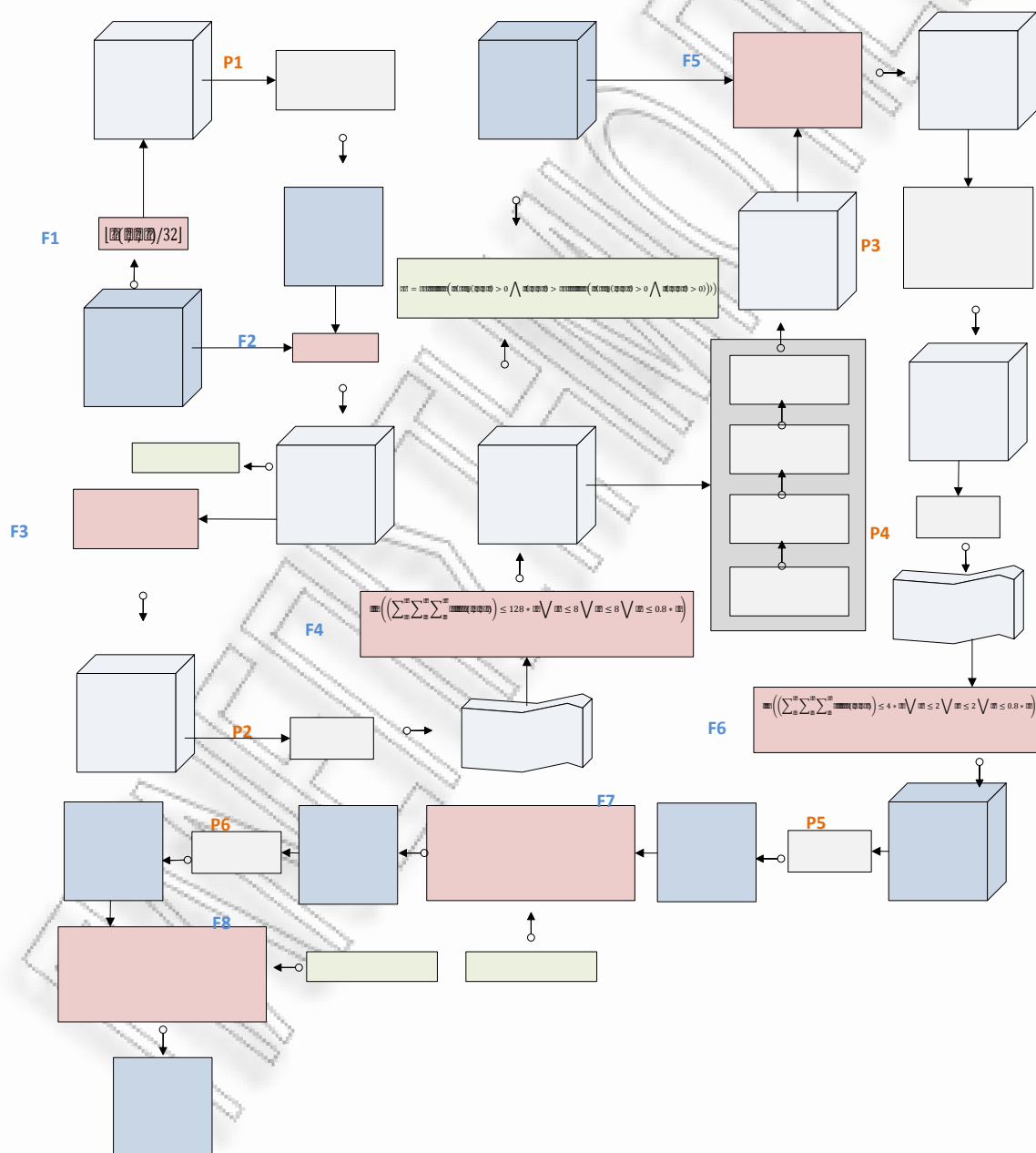
- Στον χάρτη ακμών $E(X, Y, Z)$, εφαρμόζεται μείωση της στάθμης πληροφορίας από 8bit σε 3bit ($F1$), εφαρμογή 3D GLCM ανάλυσης, ακολουθούμενη από ταξινόμηση με χρήση SVM ($P1$), και εξαγωγή των κλάσεων (0,1) μέσω απεικόνισής τους στον δυαδικό δισδιάστατο πίνακα $G(X, Y)$.

- Αξιοποιώντας το αποτέλεσμα της ταξινόμησης $G(X, Y)$, απαλείφεται η πληροφορία υποβάθρου (**F2**) του χάρτη ακμών.
- Από το αποτέλεσμα **Ef0(X, Y, Z)**, απαλείφεται η πληροφορία, η οποία έχει τιμή μικρότερη του 45% της τιμής, του στατιστικού του μέσου m (**F3**).
- Στο αποτέλεσμα **Ef1(X, Y, Z)**, εφαρμόζεται 3D τυποποίηση συνδεδεμένων στοιχείων, απαλείφοντας τα στοιχεία εκείνα τα οποία δεν τηρούν τις κατάλληλες προϋποθέσεις ($\text{Pixels} \leq 128 * Z$, $X' \leq 8$, $Y' \leq 8$, $Z' \leq 80\% * Z$).
- Σε αυτό το στάδιο, στον επεξεργασμένο χάρτη ακμών **Ef2(X, Y, Z)**, κωδικοποιείται η πληροφορία ακμών, της οποίας τα εικονοστοιχεία, αντιστοιχούν χωρικά, στις περιοχές της αρχικής πληροφορίας $I(X, Y, Z)$, όπου ενυπάρχει πληροφορία κειμένου με αυξημένη βεβαιότητα. Αξιοποιώντας την παραδοχή που κάναμε αρχικά, όταν θεωρήσαμε ότι η διασπορά της τιμής φωτεινότητας στις περιοχές κειμένου πρέπει να είναι ιδιαίτερα μικρή, υπολογίζουμε μια τιμή κατωφλίου m' , την οποία χρησιμοποιούμε στη διαδικασία **P3**:
 - Υπολογίζουμε το στατιστικό μέσο μ' των τιμών των εικονοστοιχείων της αρχικής πληροφορίας I , για τα οποία, στις αντίστοιχες θέσεις του επεξεργασμένου χάρτη ακμών **Ef2**, καταγράφεται πληροφορία.
 - Υπολογίζουμε την τιμή κατωφλίου m' , ως τον στατιστικό μέσο των τιμών των εικονοστοιχείων της αρχικής πληροφορίας I , οι τιμές των οποίων είναι μεγαλύτερες της μ' , και για τα οποία, στις αντίστοιχες θέσεις του επεξεργασμένου χάρτη ακμών **Ef2**, καταγράφεται πληροφορία.

$$m' = \text{median} \left(I(Ef_2(i, j, k)) > 0 \wedge I(i, j, k) > \text{median} \left(I(Ef_2(i, j, k)) > 0 \wedge I(i, j, k) > 0 \right) \right)$$

- Στο αποτέλεσμα **Ef2(X, Y, Z)**, εφαρμόζεται διαδοχικά μορφολογική διαστολή, διάβρωση, άνοιγμα και κλείσιμο, χρησιμοποιώντας σε κάθε βήμα διαφορετικής διάστασης δομικά στοιχεία, τα οποία επιλέξαμε εμπειρικά, ώστε να προκύψει το βέλτιστο προϊόν **Ef3(X, Y, Z)**. Στόχος μας είναι η ανάδειξη και η εντοπίηση των εικονοστοιχείων του 3D χάρτη ακμών, τα οποία περιγράφουν το σχήμα των χαρακτήρων κειμένου, περιορίζοντας στο ελάχιστο τη μορφολογική παραμόρφωσή τους.
- Αξιοποιώντας το αποτέλεσμα των μορφολογικών διεργασιών στον χάρτη ακμών **Ef3**, απαλείφουμε την πληροφορία υποβάθρου στο βίντεο I (**F5**), παράγοντας το αποτέλεσμα **If0**.
- Ακολουθεί η ανάλυση και επεξεργασία του ιστογράμματος της **θεωρούμενης περιοχής κειμένου (If0)**, μέσω της διαδικασίας **P3**. Θεωρούμε ότι στην περιοχή κειμένου, τα εικονοστοιχεία κειμένου αποτελούν πλειοψηφία, ενώ η τιμή φωτεινότητας που κωδικοποιείται σε αυτά, έχει υψηλή τιμή, ώστε να διευκολύνει την ανάγνωση των περιγραφόμενων χαρακτήρων. Βάσει των προηγούμενων θεωρήσεων εντοπίζουμε στο ιστόγραμμα την περιοχή όπου κωδικοποιείται η υψηλότερης συχνότητας πληροφορία η οποία αντιστοιχεί σε τιμές μεγαλύτερες από m' και απαλείφουμε την υπόλοιπη, θεωρώντας την πληροφορία υποβάθρου.
- Σε αυτό το στάδιο θεωρούμε ότι έχουν διαχωριστεί οι χαρακτήρες και εφαρμόζουμε εκ νέου 3D τυποποίηση συνδεδεμένων στοιχείων **P4**, με περισσότερο αυστηρά κριτήρια απαλοφής του εναπομείναντος θορύβου ($\text{Pixels} \leq 2 * Z$, $X' \leq 2$, $Y' \leq 2$, $Z' > 80\% * Z$).
- Επεξεργαζόμαστε την 3D πληροφορία που προκύπτει $O(X, Y, Z)$, ώστε να καταλήξουμε σε μια 2D δυαδική εικόνα $O(X, Y)$, η οποία θα αποτελέσει το τελικό προϊόν με το οποίο θα τροφοδοτήσουμε το στάδιο της οπτικής αναγνώρισης χαρακτήρων (OCR). Το στάδιο αυτό είναι παραμετροποιήσιμο, μέσω της εφαρμογής που αναπτύξαμε και θα παρουσιάσουμε ακολούθως, ενώ περιλαμβάνει τα ακόλουθα στάδια:

- Επιλογή της διαδικασίας μετατροπής από 3Δ → 2Δ (P5), μεταξύ του υπολογισμού της μέσης τιμής (average) ή του στατιστικού μέσου (median) και εφαρμογή της στη διάσταση του χρόνου (άξονας Z).
- Επιλογή μιας τιμής κατωφλίου **signal_stability** ($0 \leq \text{signal_stability} \leq 1$), για τη μετατροπή της εικόνας σε δυαδική, βάσει της χρονικής σταθερότητας (temporal stability), της κωδικοποιημένης πληροφορίας (F7).
- Εφαρμογή φίλτρων εξομάλυνσης και ενίσχυσης αντίθεσης (P6).
- Επιλεκτική εφαρμογή της παραμέτρου **thin_perc** ($0 \leq \text{thin_perc} \leq 1$), βάσει της επιθυμητής εκλέπτυνσης του σχήματος των χαρακτήρων του κειμένου (F8).
- Εξαγωγή της δυαδικής εικόνας O(X,Y).



ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΡΡΑΙΑ

3D Wavelet Toolbox for Textual Information Extraction in Video Sequences

Για τις ανάγκες της παρούσας εργασίας και προκειμένου να υλοποιηθούν προγραμματιστικά οι προτεινόμενες μεθοδολογίες, δημιουργήσαμε ολοκληρωμένο προϊόν λογισμικού (**3D Wavelet Toolbox for Textual Information Extraction in Video Sequences – W3D Toolbox**), το οποίο παρουσιάζεται ακολούθως.

Προγραμματιστικό περιβάλλον:

Για την υλοποίηση της εφαρμογής χρησιμοποιήθηκαν ποικίλες προγραμματιστικές προσεγγίσεις, με κεντρική επιλογή τη χρήση της γλώσσας **Python**.

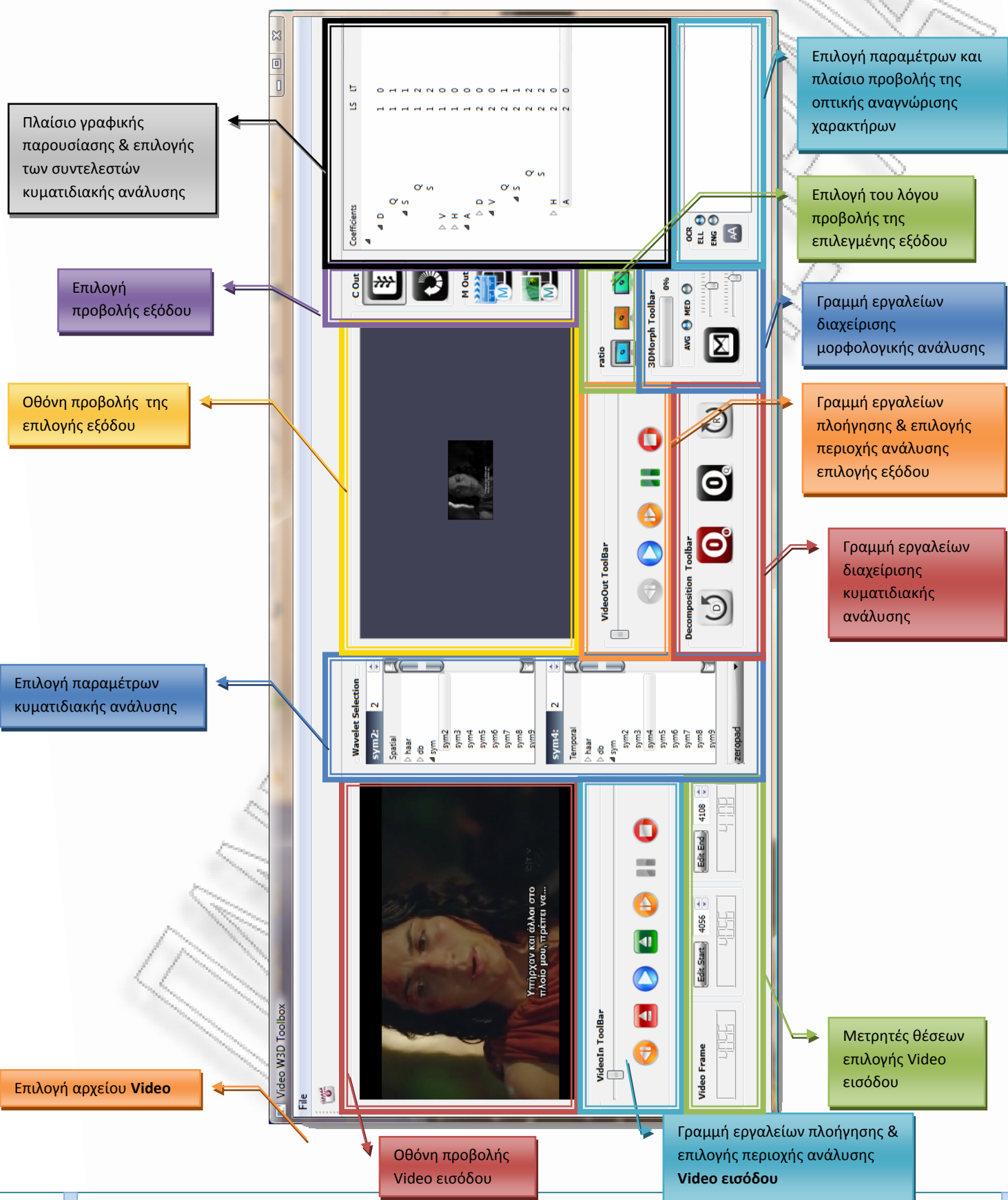
Η γλώσσα προγραμματισμού **Python**, είναι μια ισχυρή, αντικειμενοστρεφής γλώσσα, με ιδιαίτερα αποτελεσματική σύνταξη, η οποία οδηγεί στη δημιουργία ευανάγνωστου κώδικα. Σημαντικό πλεονέκτημά της αποτελεί η ύπαρξη πλήθους βιβλιοθηκών, οι οποίες επιτελούν αποτελεσματικά, χρήσιμες διεργασίες, όπως η βιβλιοθήκη **numpy** (υλοποίηση N-Διάστατων πινάκων, γραμμική άλγεβρα, δυνατότητα επέκτασης μέσω ενσωμάτωσης κώδικα χαμηλότερου επιπέδου σε C/C++ και Fortran) ή η βιβλιοθήκη **scipy**, η οποία αξιοποιεί τους N-διάστατους πίνακες της βιβλιοθήκης numpy για την υλοποίηση αλγορίθμων και μαθηματικών εργαλείων, όπως η **Τυποποίηση Συνδεδεμένων Στοιχείων**,

οι **Μορφολογικές Διεργασίες**, η εκπαίδευση και χρήση **Ταξινομητών Πυρήνα** κ.α.

Η επεξεργασία σήματος βίντεο, είναι ιδιαίτερα απαιτητική σε υπολογιστική ισχύ αλλά και δαπανηρή όσον αφορά στην χρήση μνήμης RAM. Στην παρούσα εργασία αποφασίσαμε να αξιοποιήσουμε και να υλοποιήσουμε προγραμματιστικά καινοτόμες τεχνικές επεξεργασίας, όπως η **3D μη συμβατική ανομοιογενής ανάλυση κυματιδίων**, η **3D ανάλυση μέσω πίνακα συνάφειας και συνδεδεμένων πιθανοτήτων** καθώς και ή **πρωτότυπη διαδικασία εξαγωγής κειμένου**. Τμήματα των προαναφερθέντων τεχνικών χρειάστηκε να υλοποιηθούν εξ αρχής, λαμβάνοντας υπ' όψιν τις αυξημένες υπολογιστικές απαιτήσεις που θα έπρεπε να ικανοποιηθούν. Γι' αυτό το λόγο αξιοποιήσαμε τις σημαντικές δυνατότητες επέκτασης που προσφέρει η Python ώστε να ενσωματώσουμε τμήματα κώδικα γραμμένα σε γλώσσα **Cython** και **C**. Η γλώσσα προγραμματισμού Cython, είναι μια γλώσσα-επέκταση της Python, η οποία διευκολύνει τη συγγραφή βιβλιοθηκών της, παρέχοντας συντακτικές διατυπώσεις, οι οποίες επιτρέπουν την κλήση ρουτινών αλλά και τον ορισμό τύπων της γλώσσας C, παρέχοντας παράλληλα κατάλληλο μεταγλωττιστή, με σκοπό την παραγωγή χαμηλότερου επιπέδου αποδοτικού κώδικα.

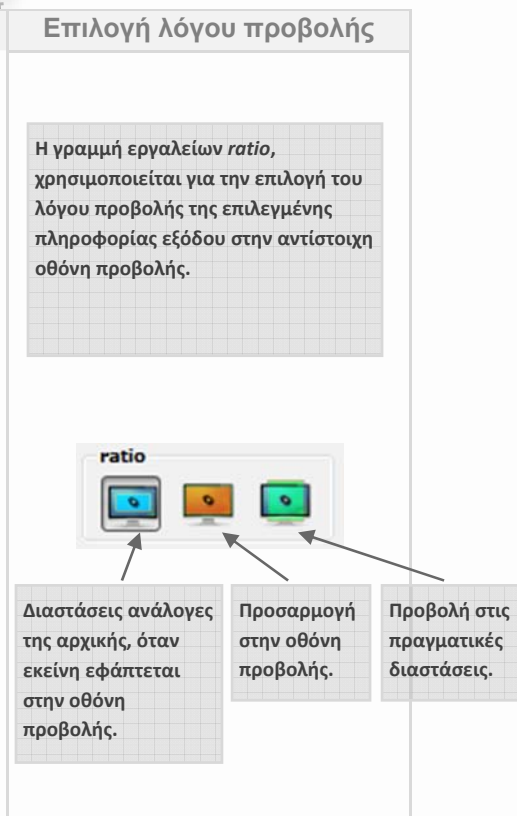
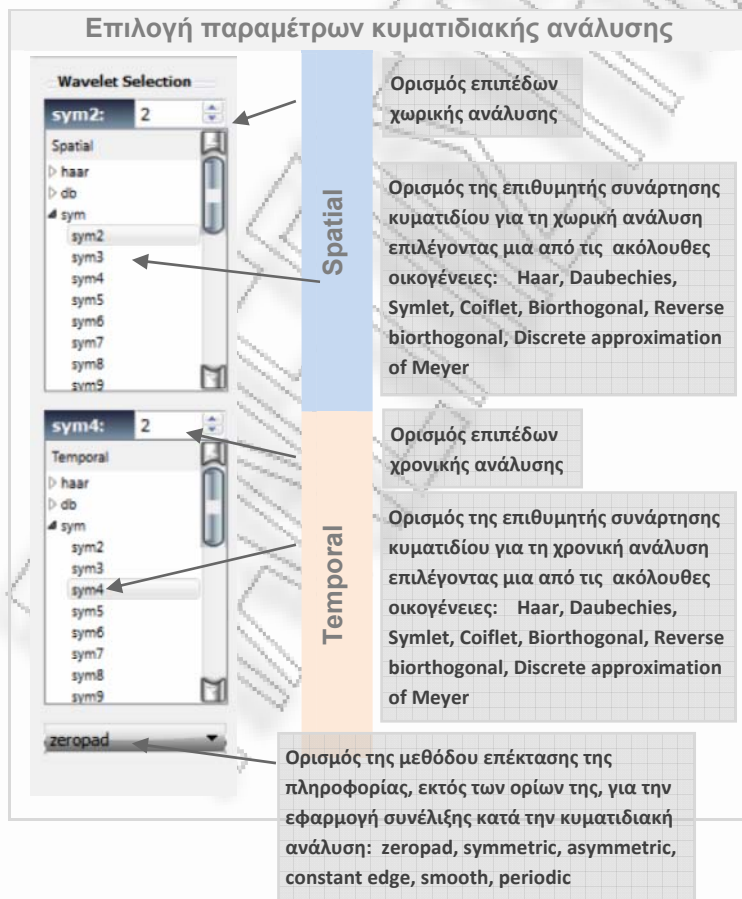
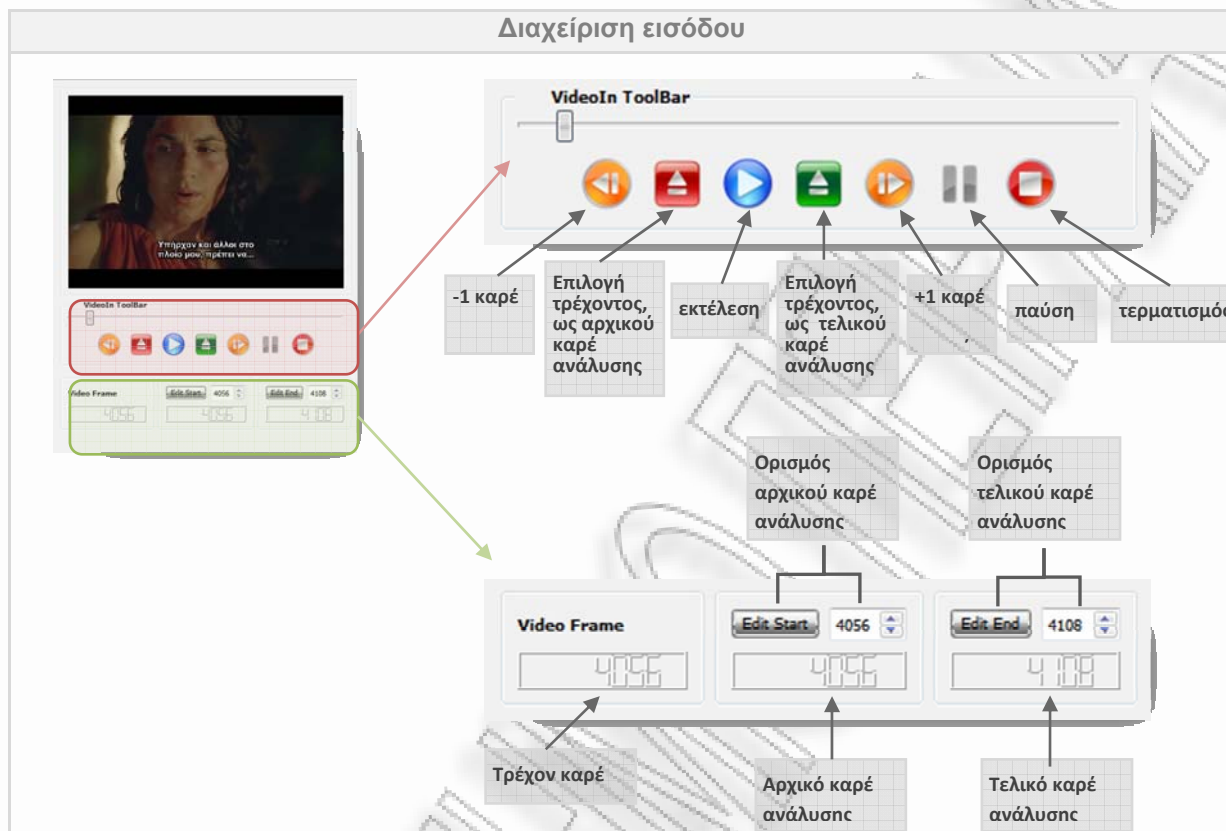
Η διεπαφή (interface) προγράμματος - χρήστη, σχεδιάστηκε μέσω του πλαισίου δημιουργίας διεπαφών χρήστη **Qt (Qt UI framework)**. Το συγκεκριμένο πακέτο διαχωρίζει το δημιουργικό τμήμα της διαδικασίας σχεδίασης από το προγραμματιστικό, εξάγοντας το γραφικό περιβάλλον με μορφή κώδικα (Python), ο οποίος στη συνέχεια εμπλουτίζεται ώστε να συνδεθούν τα γραφικά με τις επιθυμητές δραστηριότητες. Η σύνδεση αυτή καθίσταται δυνατή με τη χρήση της βιβλιοθήκης (Python) **PyQt**.

Διεπαφή χρήστη - Video W3D Toolbox:



Αναγνώριση και εξαγωγή ενσωματωμένου στατικού κειμένου από ψηφιακό σήμα εικόνας, με εφαρμογή ανομοιογενούς - πολυδιακριτικού μετασχηματισμού κυματιών, στις διαστάσεις του χώρου και του χρόνου.

Επιμέρους Συστατικά Στοιχεία Διαπαφής χρήστη - Video W3D Toolbox:



Γραμμή εργαλείων κυματιδιακής ανάλυσης



Εκτέλεση 3D κυματιδιακής ανάλυσης στην επιλεγμένη περιοχή του video, εμφάνιση του δενδρογράμματος επιλογής συντελεστών, και προβολή του επιλεγμένου συντελεστή ως video στην οθόνη εξόδου.

Μηδενισμός των τιμών του επιλεγμένου συντελεστή

Μηδενισμός των τιμών του συντελεστή προσέγγισης καθώς και των συντελεστών γοργά κινούμενων λεπτομερειών για όλα τα επίπεδα ανάλυσης και όλες τις διευθύνσεις (horizontal, vertical, diagonal)

Εκτέλεση 3D κυματιδιακής σύνθεσης βάσει των επεξεργασμένων συντελεστών και προβολή του αποτελέσματος (video ιδίων διαστάσεων με την επιλεγμένη περιοχή του video εισόδου) στην οθόνη εξόδου.

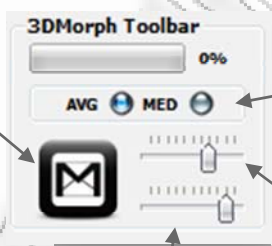
Γραμμή εργαλείων διαχείρισης μορφολογικής ανάλυσης

Πατώντας το κουμπί M (Morphological Operations on Decomposed Data), ενεργοποιούμε μια σειρά διαδικασιών, οι οποίες εμφανίζονται αναλυτικά στο διάγραμμα **διαδικασιών εξαγωγής κειμένου**. Οι διαδικασίες αυτές είναι οι ακόλουθες:

- Εξαγωγή χαρακτηριστικών μέσω 3D ανάλυσης GLCM
- Ταξινόμηση με χρήση SVM
- Εξαγωγή κειμένου με τυποποίηση συνδεδεμένων στοιχείων, μορφολογική και στατιστική ανάλυση.

Το αποτέλεσμα των προηγούμενων διαδικασιών είναι:

- Ένα video ιδίων διαστάσεων με την επιλεγμένη περιοχή του video εισόδου, όπου σε κάθε καρέ αποτυπώνεται η αντίστοιχη πληροφορία κειμένου.
- Μια εικόνα 2-διαστάσεων η οποία κωδικοποιεί το τελικό προϊόν της ανάλυσης και προκύπτει από τη συγχώνευση των επιμέρους καρέ, με τη μέθοδο που περιγράφεται στο διάγραμμα **διαδικασιών εξαγωγής κειμένου**.



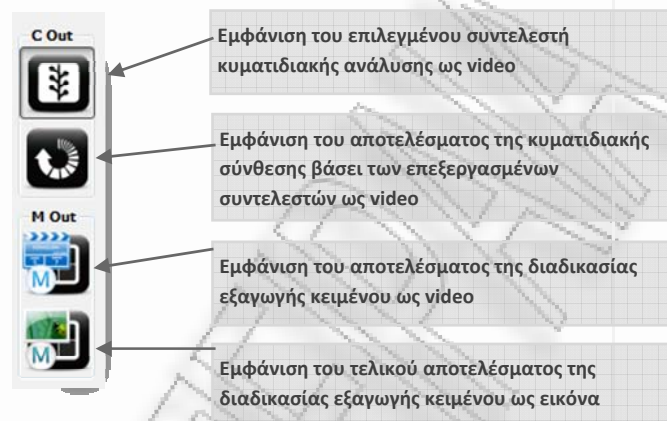
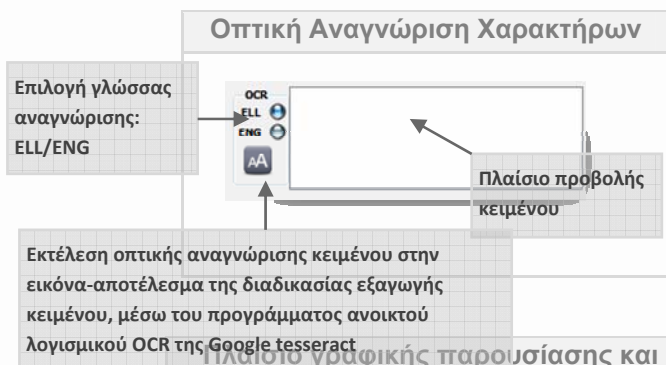
Επιλογή μεθόδου συγχώνευσης μεταξύ μέσης τιμής και στατιστικού μέσου και εφαρμογή της στη διάσταση του χρόνου

Επιλογή μιας τιμής κατωφλίου $signal_stability$ ($0 \leq signal_stability \leq 1$), για τη μετατροπή της εικόνας σε δυαδική, βάσει της χρονικής σταθερότητας (temporal stability), της κωδικοποιημένης πληροφορίας.

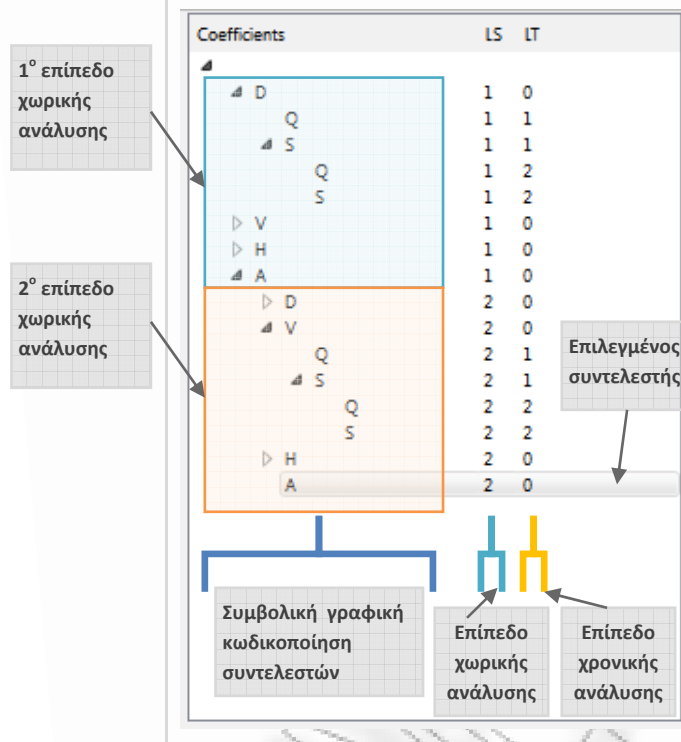
Επιλεκτική εφαρμογή της παραμέτρου $thin_perc$ ($0 \leq thin_perc \leq 1$), βάσει της επιθυμητής εκλέπτυνσης του σχήματος των χαρακτήρων του κειμένου.

Το τελικό στάδιο παραγωγής της εικόνας 2-διαστάσεων είναι παραμετροποιήσιμο από τον χρήστη.

Επιλογή προβολής εξόδου



Πλαίσιο γραφικής παρουσίασης και επιλογής συντελεστών κυματιδιακής ανάλυσης



Στο συγκεκριμένο παράδειγμα παρουσιάζεται το δενδροδιάγραμμα των συντελεστών κυματιδιακού μετασχηματισμού για 2 επίπεδα χωρικής & 2 επίπεδα χρονικής ανάλυσης

Οι διευθύνσεις χωρικής ανάλυσης κωδικοποιούνται ως εξής:

- H: Οριζόντιες λεπτομέρειες
- V: Κατακόρυφες λεπτομέρειες
- D: Διαγώνιες λεπτομέρειες

Ενώ η προσέγγιση του αρχικού σήματος κωδικοποιείται με τον χαρακτήρα A (Approximation)

Οι διευθύνσεις χρονικής ανάλυσης κωδικοποιούνται ως εξής:

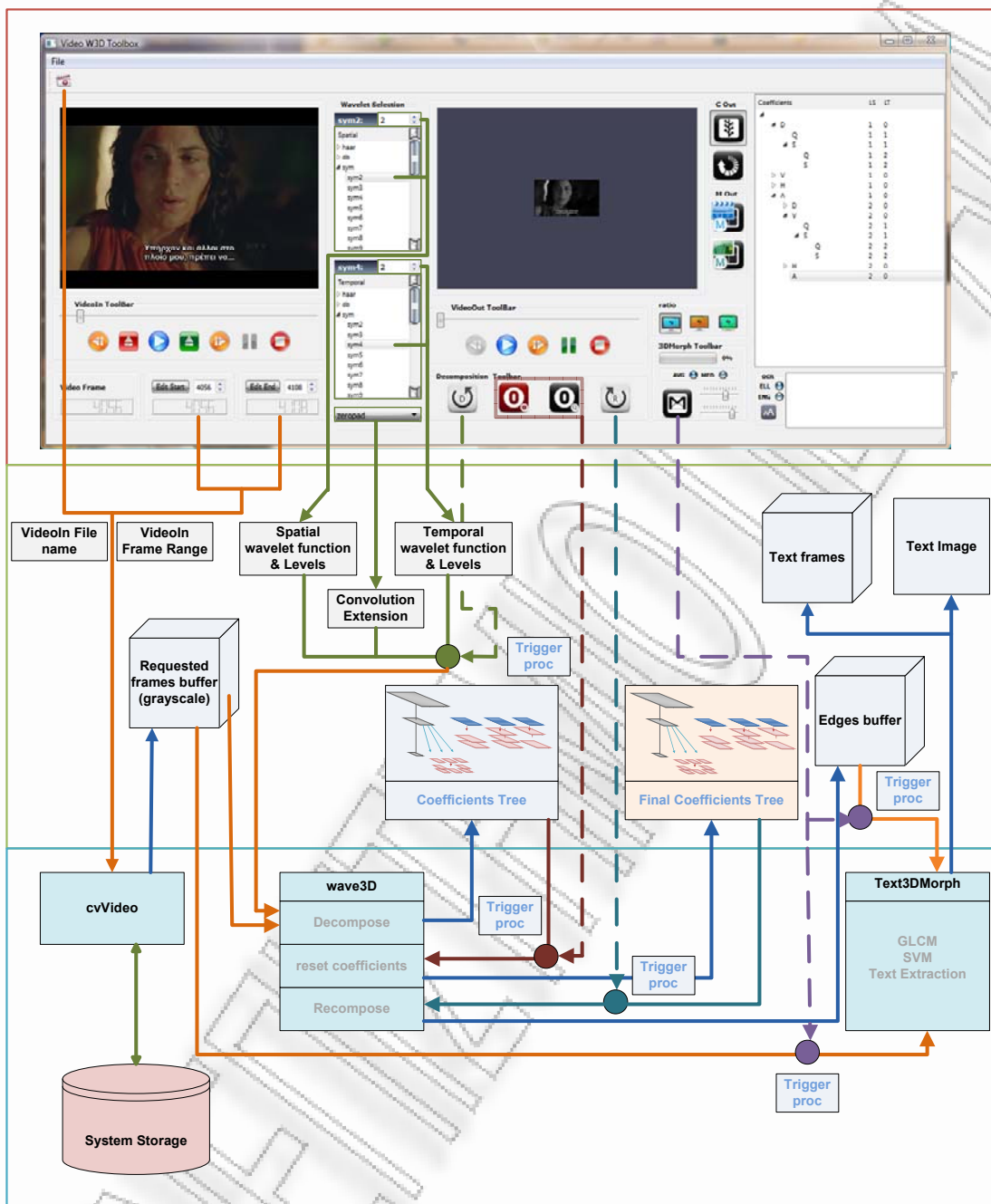
- Q: Γοργά κινούμενες λεπτομέρειες
- S: Αργά κινούμενες λεπτομέρειες

Για τις ανάγκες της εφαρμογής έχουν δημιουργηθεί τρεις βιβλιοθήκες (**cvVideo**, **wave3D**, **Text3DMorph**), οι οποίες ενεργοποιούνται μέσω του γραφικού περιβάλλοντος, δίνοντας στο χρήστη τη δυνατότητα πειραματισμού στα επιμέρους στάδια της διαδικασίας. Ο συνολικός σχεδιασμός προσδίδει στην εφαρμογή το χαρακτηρισμό της «*εργαλειοθήκης*» (toolbox), λόγω της ευχέρειας διαχείρισης επιμέρους «*εργαλείων*» διαδραστικά και με συγκεκριμένο τελικό στόχο (την εξαγωγή και αναγνώριση του εντυπωμένου κειμένου). Η διαδικασία χειρισμού του προγράμματος περιλαμβάνει τα εξής βήματα:

1. Επιλογή αρχείου video, πλοήγηση και ορισμός της περιοχής αναγνώρισης.
2. Επιλογή των επιθυμητών συναρτήσεων, των επιθυμητών επιπέδων κυματιδιακής ανάλυσης στο χώρο και το χρόνο και της μεθόδου επέκτασης του σήματος (για τις ανάγκες εφαρμογής της συνέλιξης) από την περιοχή *wavelet selection*.

3. Επιλογή *start decomposition* από τη γραμμή εργαλείων *Decomposition Toolbox*. Η συγκεκριμένη επιλογή καλεί τη βιβλιοθήκη *cnVideo*, η οποία απομονώνει την πληροφορία της περιοχής αναγνώρισης και τη μετατρέπει από έγχρωμη, στην κλίμακα του γκρι (*grayscale buffer*). Στη συνέχεια καλείται η βιβλιοθήκη *wave3D*, η οποία εφαρμόζει κυματιδιακή ανάλυση με τις επιθυμητές παραμέτρους, επιστρέφοντας τις συνιστώσες της (*Coefficient Tree*).
4. Στο πλαίσιο γραφικής παρουσίασης & επιλογής των συντελεστών κυματιδιακής ανάλυσης δημιουργείται το δένδροδιάγραμμα των συντελεστών, από όπου μπορούμε να επιλέξουμε την οπτική προβολή του επιθυμητού συντελεστή.
5. Από τη γραμμή εργαλείων *Decomposition Toolbox*, είναι δυνατόν να μηδενίζουμε τον επιλεγμένο συντελεστή, μέσω του κομβίου *Zero Selected Coefficient*. Επειδή οι ανάγκες αναγνώρισης στατικού κειμένου απαιτούν το μηδενισμό των τιμών του συντελεστή προσέγγισης καθώς και των συντελεστών των γοργά κινούμενων λεπτομερειών για όλα τα επίπεδα ανάλυσης και όλες τις διευθύνσεις, έχει προβλεφθεί η ύπαρξη κομβίου (*Zero Quick Coefficient*), για τον μηδενισμό όλων των σχετικών συντελεστών.
6. Αφού έχουμε καταλήξει στην τελική μορφή των συντελεστών κυματιδιακής ανάλυσης (*Final Coefficients Tree*), μπορούμε να επιλέξουμε τη σύνθεσή τους μέσω της γραμμής εργαλείων *Decomposition Toolbox*, πατώντας το κομβίο *Recompose*. Η βιβλιοθήκη *wave3D* εφαρμόζει την κυματιδιακή σύνθεση και επιστρέφει 3D σήμα *video* ιδίων διαστάσεων με το πρωτότυπο, όπου ιδανικά θα πρέπει να αποτυπώνονται μονάχα οι ακμές των χαρακτήρων κειμένου (*Edges buffer*). Η συγκεκριμένη πληροφορία παρουσιάζεται στην οθόνη εξόδου.
7. Όλα τα στάδια της εξαγωγής του κειμένου (3D ανάλυση GLCM, κατηγοριοποίηση με χρήση SVM, μορφολογικές διεργασίες και στατιστική ανάλυση), από την αρχική πληροφορία μέσω του χάρτη ακμών, πραγματοποιούνται από τη βιβλιοθήκη *Text3DMorph*, η οποία ενεργοποιείται μέσω του κομβίου *Morphological Operations on Decomposed Data*. Το αποτέλεσμα των προηγούμενων διαδικασιών είναι η 3D πληροφορία *Text frames* και η 2D εικόνα *Text image*. Οι δύο πληροφορίες οπτικοποιούνται κατ' επιλογήν του χρήστη στην οθόνη εξόδου.
8. Η τελική εικόνα κειμένου (*Text Image*), μπορεί να αναπαραχθεί κατ' επανάληψη, με παραμετροποίηση κρίσιμων παραμέτρων (μέθοδος συγχώνευσης, παράμετρος χρονικής ευστάθειας, παράμετρος εκλέπτυνσης), μέσω της γραμμής εργαλείων *3D Morph Toolbar*.
9. Η οπτική αναγνώριση χαρακτήρων στην εικόνα *Text Image*, εφαρμόζεται μέσω της γραμμής εργαλείων *Οπτικής Αναγνώρισης Χαρακτήρων-OCR*, η οποία ενεργοποιεί την εξωτερική εφαρμογή *tesseract*, μια εφαρμογή ανοικτού λογισμικού υπό την αιγίδα της Google. Το αποτέλεσμα της αναγνώρισης επιστρέφεται και εμφανίζεται ως κείμενο στη σχετική περιοχή της διεπαφής χρήστη, της εφαρμογής *Video3D Toolbox*.

Όλες οι προηγούμενες διαδικασίες αναπαρίστανται γραφικά στο ακόλουθο διάγραμμα:



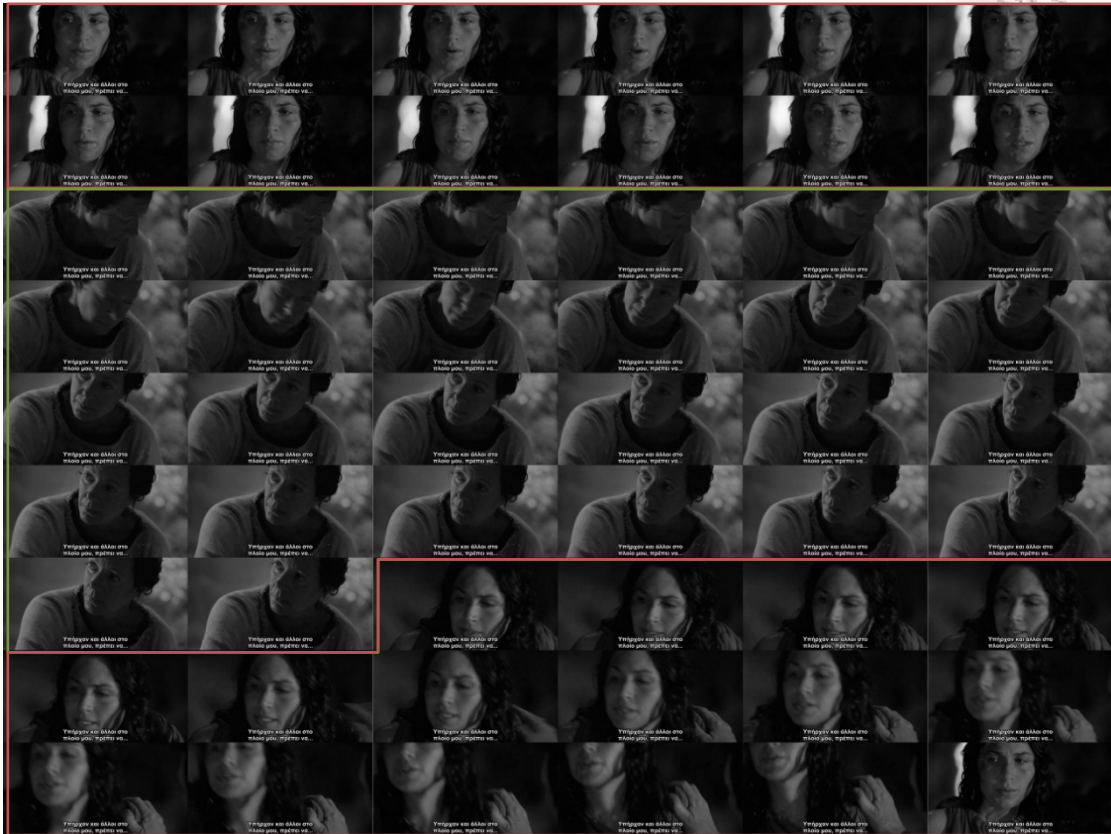
Παραδείγματα εφαρμογής σε πραγματικές συνθήκες:

Ακολουθούν παραδείγματα ανάλυσης σήματος video με εντυπωμένο στατικό κείμενο, όπου παρουσιάζονται τα δεδομένα εισόδου, η πληροφορία εξόδου των επιμέρους αναλύσεων, καθώς και το τελικό προϊόν, δηλαδή η εικόνα κειμένου και το αποτέλεσμα της οπτικής αναγνώρισης χαρακτήρων.

Παράδειγμα 1

Στο πρώτο παράδειγμα χρησιμοποιούμε βίντεο, υψηλής ευκρίνειας, κωδικοποιημένο σε συμπίεση xvid με υψηλή ταχύτητα δυαδικών ψηφίων (bitrate). Το υπόβαθρο είναι στατικό, στη διάρκεια όμως των 53 επιλεγμένων καρέ εμφανίζονται 3 διαφορετικές σκηνές.

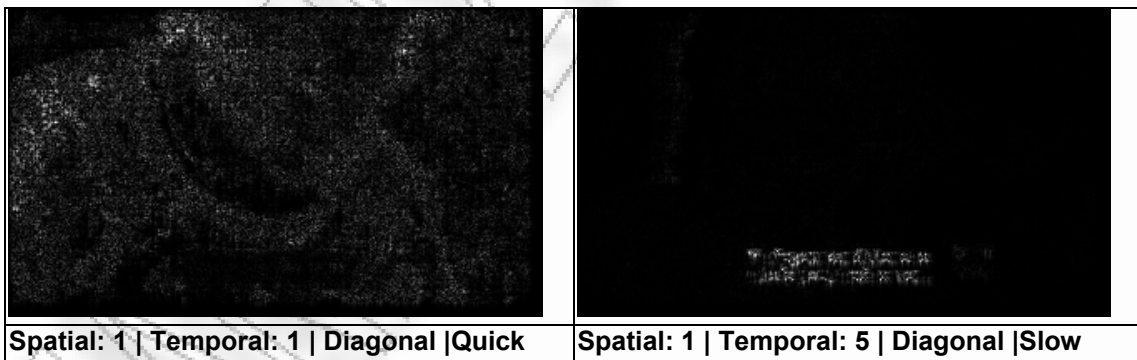
Βίντεο Εισόδου (53 καρέ)



Στο παραπάνω σχήμα διακρίνονται οι τρεις διαφορετικές σκηνές στις οποίες όμως εντυπώνεται ο ίδιος υπότιτλος.

Ανάλυση Κυματιδίων & Συντελεστές

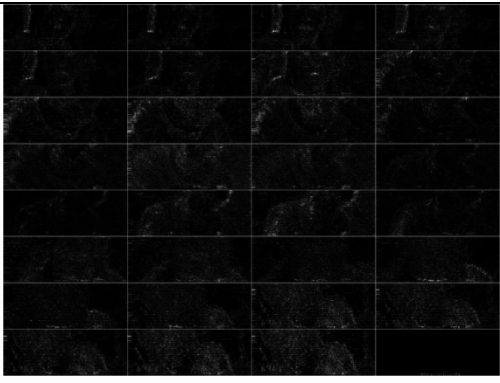

Εφαρμόζουμε 3 επιπέδων χωρική και 5 επιπέδων χρονική κυματιδιακή ανάλυση, χρησιμοποιώντας κυματιδιακή συνάρτηση Daubechies 10 & Daubechies 6 αντίστοιχα.



Spatial: 1 Temporal: 1 Vertical Quick	Spatial: 1 Temporal: 5 Vertical Slow
Spatial: 1 Temporal: 1 Horizontal Quick	Spatial: 1 Temporal: 5 Horizontal Slow

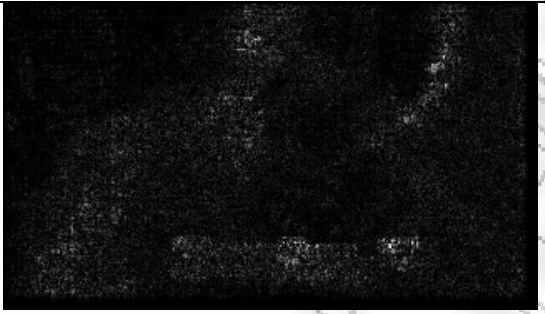
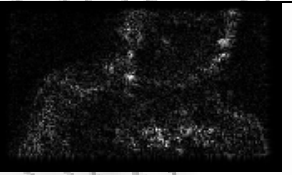




Στις προηγούμενες εικόνες παρατηρούμε σε αντιπαράθεση μεμονωμένα καρέ των συντελεστών ανάλυσης, για γοργά και αργά κινούμενες λεπτομέρειες ταξινομημένα ανά κατεύθυνση (Horizontal, Vertical, Diagonal), για κοινή χωρική ανάλυση 1, αλλά διαφορετική χρονική ανάλυση. Διαπιστώνουμε την αποτελεσματικότητα του αλγορίθμου όσον αφορά τον διαχωρισμό κινούμενων και στάσιμων λεπτομερειών (ανάλυση στο πεδίο του χρόνου), αλλά και τον εντοπισμό και διαχωρισμό του κειμένου από το υπόβαθρο (χωρική ανάλυση).

Spatial: 1 Temporal: 1 Diagonal Quick	Spatial: 1 Temporal: 5 Diagonal Slow
Spatial: 1 Temporal: 1 Diagonal Quick	Spatial: 1 Temporal: 5 Diagonal Slow

Spatial: 1 Temporal: 1 Vertical Quick	Spatial: 1 Temporal: 5 Vertical Slow
	
Spatial: 1 Temporal: 1 Horizontal Quick	Spatial: 1 Temporal: 5 Horizontal Slow

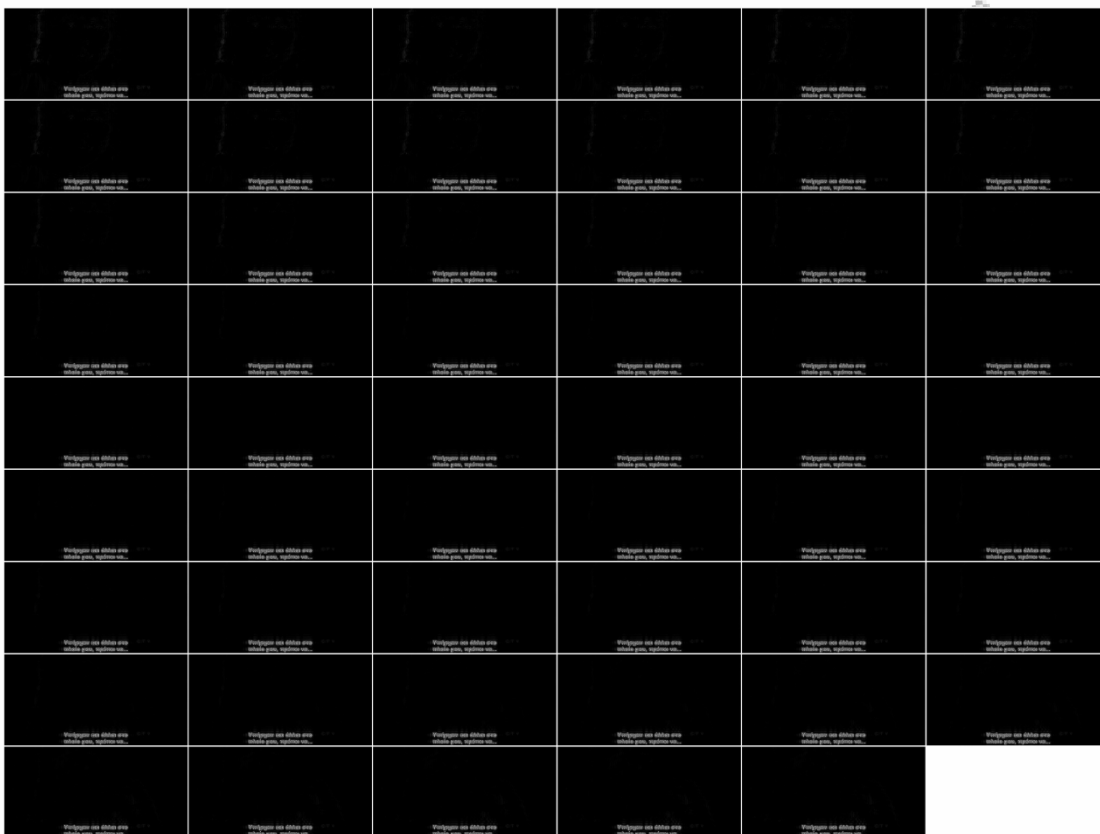
Η διαπίστωση αυτή ενισχύεται από την εξέταση του συνόλου των καρέ για τις αντίστοιχες περιπτώσεις.

Εάν συγκρίνουμε ομάδες καρέ συντελεστών (καρέ N° 12) ίδιας διεύθυνσης και χρονικής ανάλυσης, αλλά διάφορης χωρικής ανάλυσης, διαπιστώνουμε ότι όσο αυξάνεται η χωρική ανάλυση, τόσο πιο αδρά κωδικοποιούνται οι λεπτομέρειες.

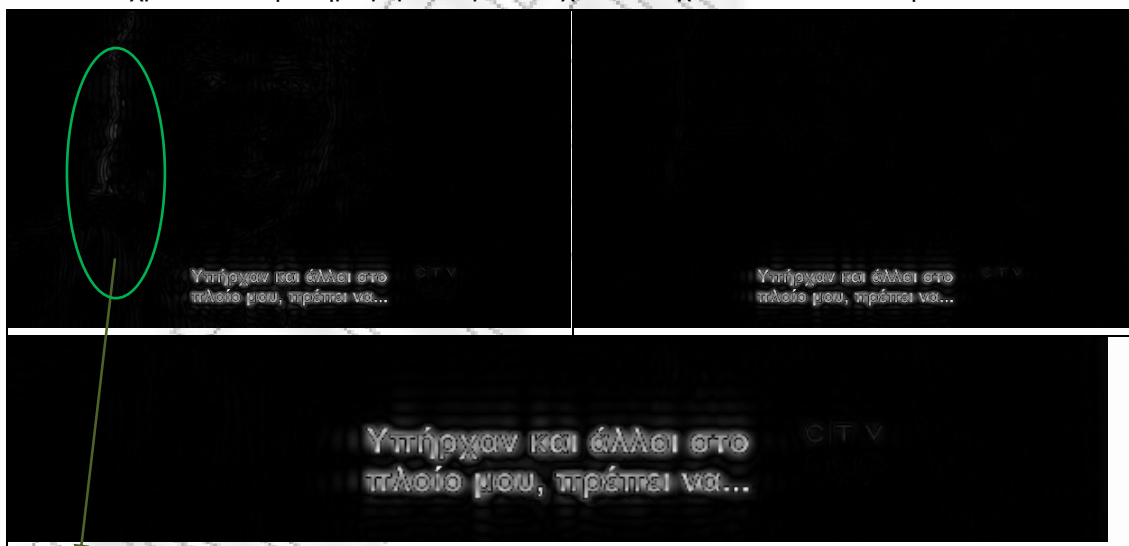
		
Spatial: 1 Temporal: 2 Diagonal Quick	Spatial: 2 Temporal: 2 Diagonal Quick	Spatial: 3 Temporal: 2 Diagonal Quick
		
Spatial: 1 Temporal: 5 Horizontal Slow	Spatial: 2 Temporal: 5 Horizontal Slow	Spatial: 3 Temporal: 5 Horizontal Slow

Σύνθεση Κυματιδίων

Εάν μηδενίσουμε όλους τους γοργά κινούμενους συντελεστές καθώς και τον συντελεστή προσέγγισης και εφαρμόσουμε κυματιδιακή σύνθεση, προκύπτει ο χάρτης ακμών που εμφανίζεται ακολούθως.

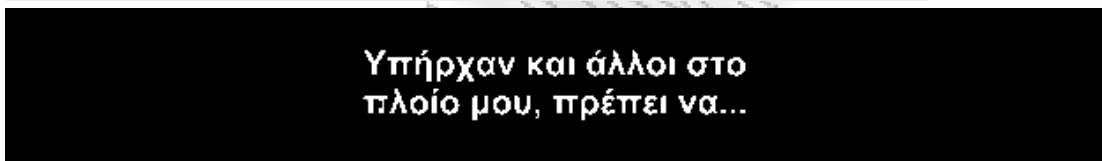
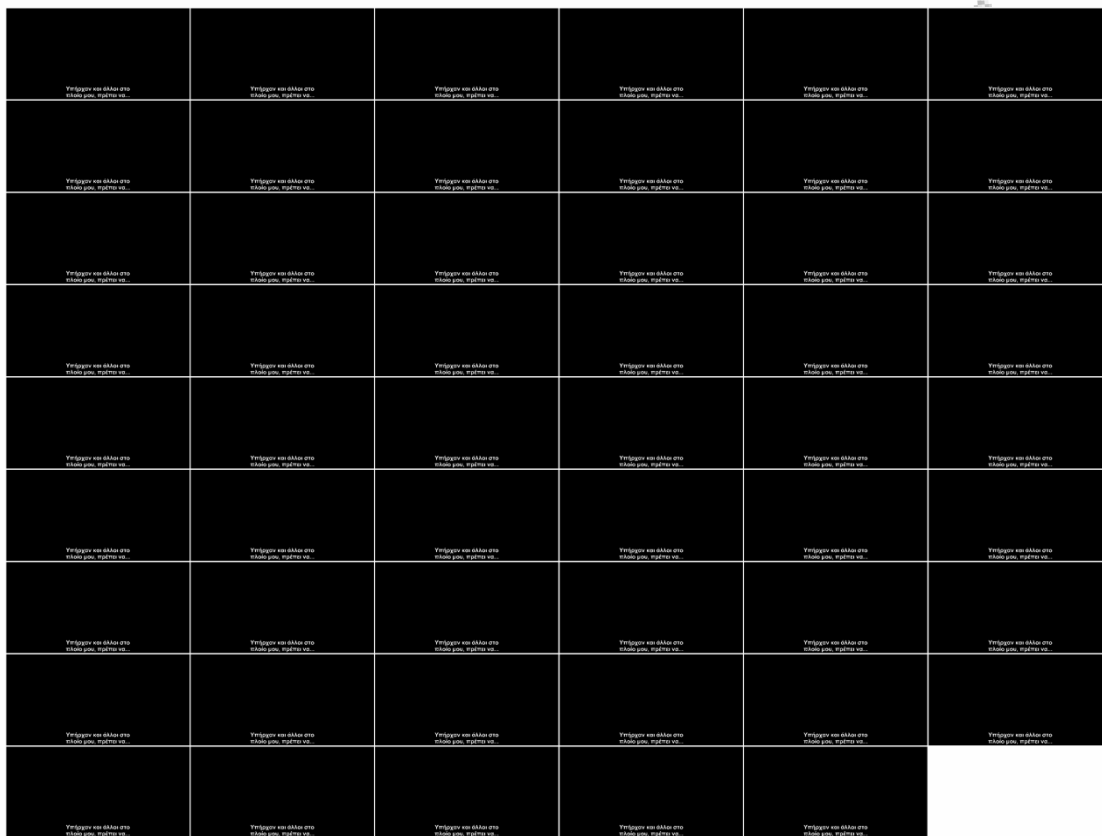


Παρατηρούμε ότι στη συγκεκριμένη περίπτωση η πληροφορία είναι ιδιαίτερα στατική στο πεδίο του χρόνου και η πληροφορία κειμένου έχει αναδειχθεί πολύ ικανοποιητικά:

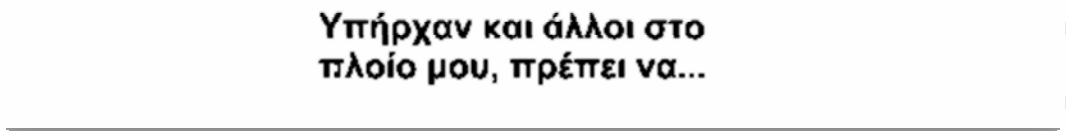


Παρατηρούμε μονάχα λίγο θόρυβο στα αρχικά καρέ ο οποίος θα εξαφανιστεί στα επόμενα στάδια.

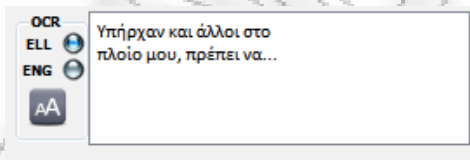
Εκτελούμε τα υπόλοιπα στάδια της εξαγωγής του κειμένου (3Δ ανάλυση GLCM, κατηγοριοποίηση με χρήση SVM, μορφολογικές διεργασίες και στατιστική ανάλυση) και παρουσιάζουμε το τελικό βίντεο με εντυπωμένη την εξαχθείσα πληροφορία κειμένου:



Καθώς και την τελική εικόνα κειμένου:

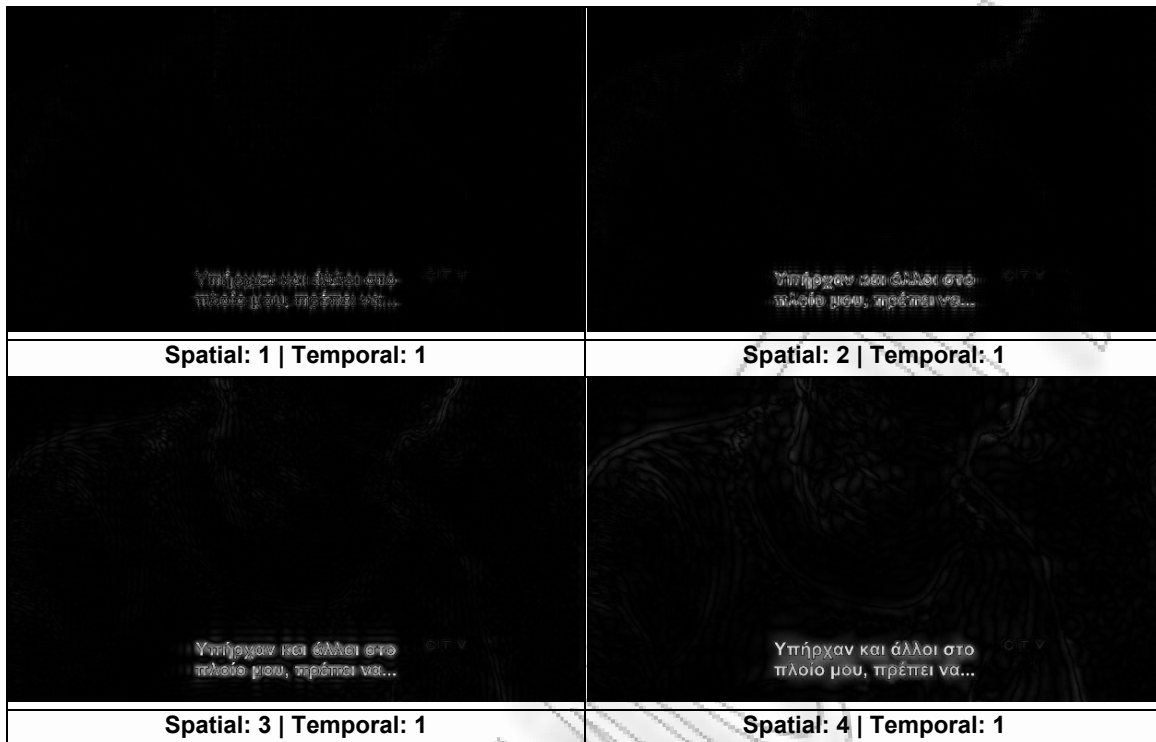


Η οποία αναγνωρίζεται επακριβώς από το πρόγραμμα OCR – tesseract:



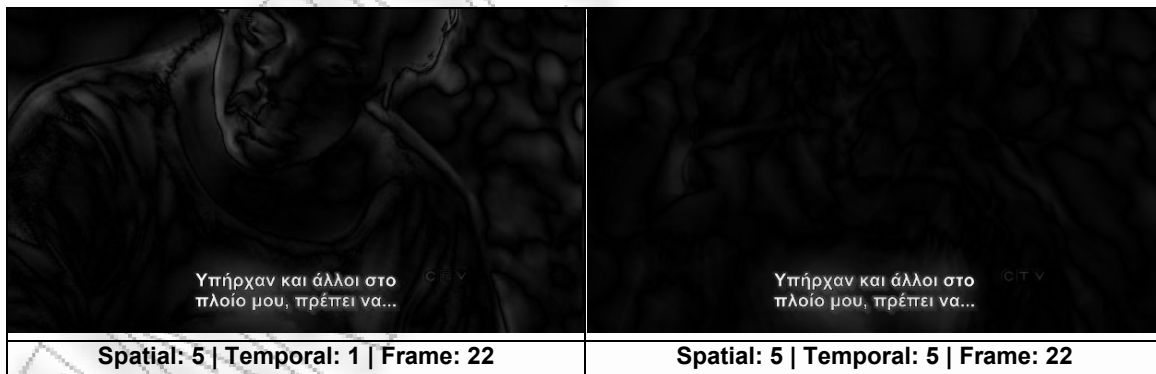
Συγκριτική μελέτη παραμέτρων Κυματιδιακής Ανάλυσης:

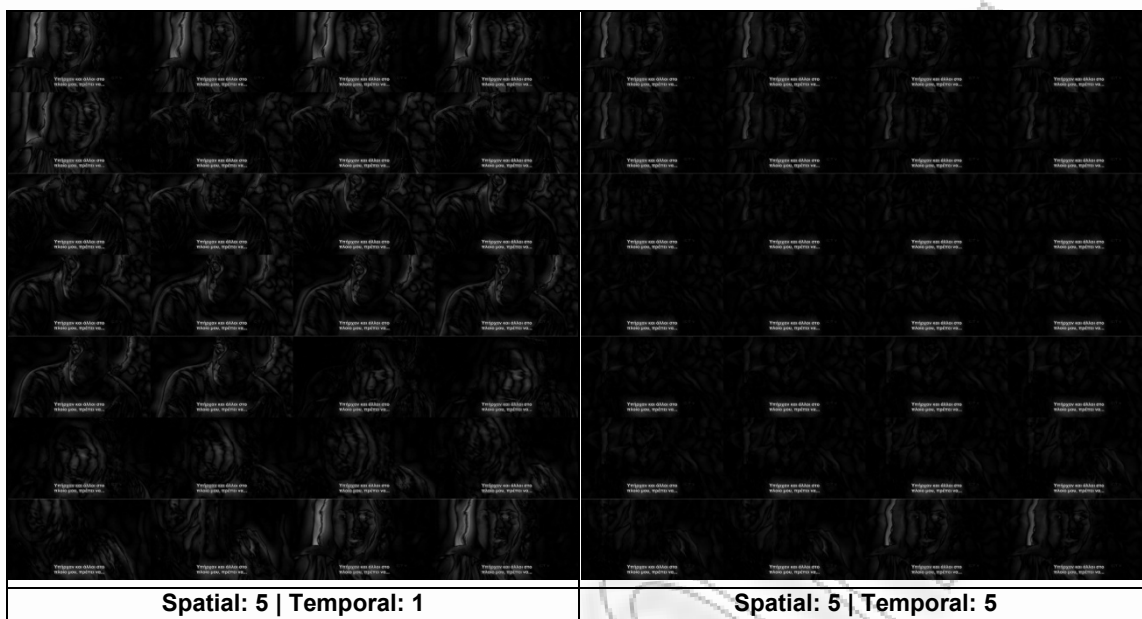
Για την ίδια πληροφορία εισόδου, εφαρμόζουμε κυματιδιακή ανάλυση και σύνθεση, χρησιμοποιώντας κυματιδιακή συνάρτηση Daubechies 10 & Daubechies 6 για τη διάσταση του χώρου και του χρόνου αντίστοιχα:



Σταδιακά αυξάνουμε το επίπεδο χωρικής ανάλυσης και διαπιστώνουμε ότι, για χαμηλές τιμές του, η περιοχή κειμένου περιγράφεται ασθενώς, ενώ για υψηλές τιμές οι ακμές τείνουν να συγχωνευτούν περιγράφοντας το κείμενο λεπτομερώς με υψηλές τιμές φωτεινότητας. Επιπρόσθετα, λόγω της ανάδειξης χωρικών θεάσεων του σήματος υπό χαμηλή ανάλυση, εμφανίζεται αυξανόμενης φωτεινής έντασης άλως γύρω από τις αναδεικνυόμενες περιοχές, ενώ παράλληλα παρατηρείται ανάδειξη αδρών φωτεινών μορφών στο υπόβαθρο.

Στις αναλυόμενες σκηνές ο αλγόριθμος εμφανίζει ιδιαίτερα καλή συμπεριφορά και ο 3D χάρτης ακμών παραμένει στατικός στο χρόνο χωρίς την παρουσία θορύβου. Γι' αυτό το λόγο ενισχύουμε το θόρυβο υποβάθρου εφαρμόζοντας υψηλή τιμή χωρικής ανάλυσης (επίπεδο 5), ώστε να πειραματιστούμε με την αποτελεσματικότητα της χρονικής ανάλυσης για διάφορα επίπεδα εφαρμογής της.





Όπως παρατηρούμε στις προηγούμενες εικόνες οι υψηλές τιμές επιπέδου χρονικής ανάλυσης υποβαθμίζουν το θόρυβο της πληροφορίας υποβάθρου, χωρίς να υποβαθμίζουν την πληροφορία κειμένου.

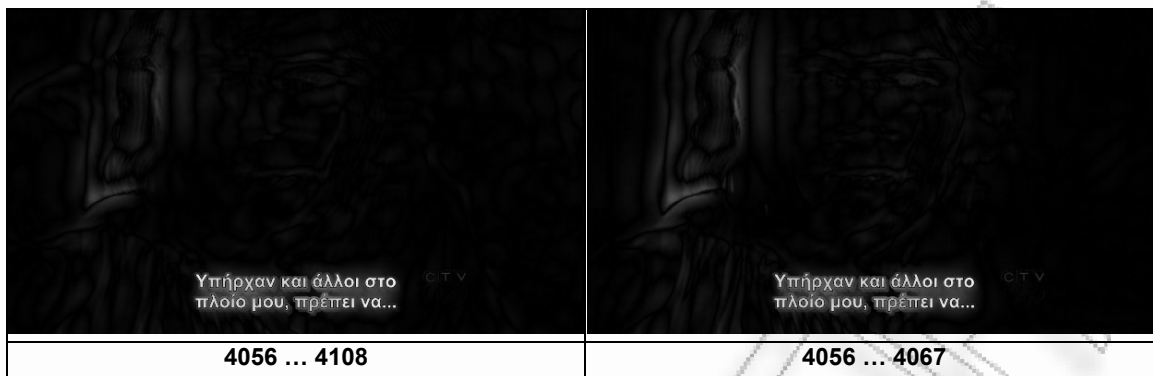
Ακολούθως θα εφαρμόσουμε κυματιδικό μετασχηματισμό εφαρμόζοντας διαφόρων ειδών συναρτήσεις, διατηρώντας σταθερή την επιλογή των επιπέδων ανάλυσης (Spatial: 3, Temporal: 2).



Παρατηρούμε ότι η επιλογή συνάρτησης χωρικής κυματιδικής ανάλυσης παίζει καταλυτικό ρόλο στην ποιότητα περιγραφής της υποκείμενης πληροφορίας. Το ίδιο παρατηρούμε δοκιμάζοντας διαφορετικού τύπου συναρτήσεις για τη διαδικασία της χρονικής ανάλυσης.

Συγκριτική μελέτη αποτελεσμάτων ανάλογα με τη χρονική διάρκεια του σήματος:

Εφαρμόζουμε κυματιδική ανάλυση (χωρική: 5 επίπεδα - db8, χρονική: 3 επίπεδα – db4), κατά πρώτον στην εξεταζόμενη περιοχή του βίνεο του παραδείγματος 1 (καρέ 4056..4108) και κατά δεύτερον στην περιοχή 4056..4067, δηλαδή μονάχα στην πρώτη σκηνή:



Η παρουσία θορύβου στο υπόβαθρο είναι περισσότερο έντονη, όπως αναμένουμε, στη ανάλυση της μικρότερης περιοχής (1^η σκηνή – δεξιά εικόνα), διότι η ύπαρξη δυναμικών εναλλαγών στο υπόβαθρο, διευκολύνει το διαχωρισμό και την υποβάθμισή του σε σχέση με το στατικό κείμενο, κατά τη διαδικασία της χρονικής ανάλυσης.

Παράδειγμα 2

Στο δεύτερο παράδειγμα χρησιμοποιούμε βίντεο, χαμηλής ευκρίνειας προερχόμενο από πηγή VHS. Στο υπόβαθρο εμφανίζεται δυναμικά μεταβαλλόμενη στο χρόνο, ισχυρής φωτεινής έντασης πηγή, η οποία υπερκαλύπτει τμήματα της πληροφορίας κειμένου.



Εκτός αυτού μπορούμε να παρατηρήσουμε την σποραδική εμφάνιση θορύβου, η οποία είναι αρκετά έντονη και επηρεάζει την περιοχή κειμένου σε κάποια καρέ.



Με τη επιλογή κατάλληλου συνδυασμού φίλτρων: *symlets 6, 3 επίπεδα χωρική και symlets10, 5 επίπεδα χρονική ανάλυση*, το αποτέλεσμα της κυματιδικής ανάλυσης –επεξεργασίας και σύνθεσης, αναδεικνύει την πληροφορία κειμένου:

Γιατί μελετάς τους Ινδιάνους;	Γιατί μελετάς τους Ινδιάνους;	Γιατί μελετάς τους Ινδιάνους;	Γιατί μελετάς τους Ινδιάνους;	Γιατί μελετάς τους Ινδιάνους;	Γιατί μελετάς τους Ινδιάνους;	Γιατί μελετάς τους Ινδιάνους;	Γιατί μελετάς τους Ινδιάνους;
Γιατί μελετάς τους Ινδιάνους;	Γιατί μελετάς τους Ινδιάνους;	Γιατί μελετάς τους Ινδιάνους;	Γιατί μελετάς τους Ινδιάνους;	Γιατί μελετάς τους Ινδιάνους;	Γιατί μελετάς τους Ινδιάνους;	Γιατί μελετάς τους Ινδιάνους;	Γιατί μελετάς τους Ινδιάνους;
Γιατί μελετάς τους Ινδιάνους;	Γιατί μελετάς τους Ινδιάνους;	Γιατί μελετάς τους Ινδιάνους;	Γιατί μελετάς τους Ινδιάνους;	Γιατί μελετάς τους Ινδιάνους;	Γιατί μελετάς τους Ινδιάνους;	Γιατί μελετάς τους Ινδιάνους;	Γιατί μελετάς τους Ινδιάνους;
Γιατί μελετάς τους Ινδιάνους;	Γιατί μελετάς τους Ινδιάνους;	Γιατί μελετάς τους Ινδιάνους;	Γιατί μελετάς τους Ινδιάνους;	Γιατί μελετάς τους Ινδιάνους;	Γιατί μελετάς τους Ινδιάνους;	Γιατί μελετάς τους Ινδιάνους;	

Ενώ το στάδιο της εξαγωγής κειμένου εξαγει αποκλειστικά την επιθυμητή πληροφορία

Γιατί μελετάς τους Ινδιάνους;	Γιατί μελετάς τους Ινδιάνους;	Γιατί μελετάς τους Ινδιάνους;	Γιατί μελετάς τους Ινδιάνους;	Γιατί μελετάς τους Ινδιάνους;	Γιατί μελετάς τους Ινδιάνους;	Γιατί μελετάς τους Ινδιάνους;	Γιατί μελετάς τους Ινδιάνους;
Γιατί μελετάς τους Ινδιάνους;	Γιατί μελετάς τους Ινδιάνους;	Γιατί μελετάς τους Ινδιάνους;	Γιατί μελετάς τους Ινδιάνους;	Γιατί μελετάς τους Ινδιάνους;	Γιατί μελετάς τους Ινδιάνους;	Γιατί μελετάς τους Ινδιάνους;	Γιατί μελετάς τους Ινδιάνους;
Γιατί μελετάς τους Ινδιάνους;	Γιατί μελετάς τους Ινδιάνους;	Γιατί μελετάς τους Ινδιάνους;	Γιατί μελετάς τους Ινδιάνους;	Γιατί μελετάς τους Ινδιάνους;	Γιατί μελετάς τους Ινδιάνους;	Γιατί μελετάς τους Ινδιάνους;	Γιατί μελετάς τους Ινδιάνους;
Γιατί μελετάς τους Ινδιάνους;	Γιατί μελετάς τους Ινδιάνους;	Γιατί μελετάς τους Ινδιάνους;	Γιατί μελετάς τους Ινδιάνους;	Γιατί μελετάς τους Ινδιάνους;	Γιατί μελετάς τους Ινδιάνους;	Γιατί μελετάς τους Ινδιάνους;	Γιατί μελετάς τους Ινδιάνους;

Γιατί μελετάς τους Ινδιάνους;

Συμπεράσματα-Περίληψη:

Βάσει της θεωρητικής ανάλυσης που αναπτύχθηκε στα κεφάλαια της παρούσας εργασίας αλλά και κατά την εφαρμογή των προκρινόμενων μεθοδολογιών μέσω του εργαλείου **Video W3D Toolbox**, αποδείξαμε ότι η εφαρμογή **3Δ μη συμβατικής ανομοιογενούς ανάλυσης κυματιδίων** στα πεδία του χώρου και του χρόνου, μπορεί να αναδείξει την πληροφορία εντυπωμένου στατικού κειμένου σε video με πολύ καλά αποτελέσματα.

Παράλληλα, η εφαρμογή της προτεινόμενης μεθοδολογίας εξαγωγής κειμένου με χρήση και εφαρμοσμένη επέκταση υφιστάμενων τεχνικών στις 3 διαστάσεις (μήτρα GLCM, κατηγοριοποίηση με χρήση SVM, μορφολογικές διεργασίες), καθώς και πρωτότυπης στατιστικής ανάλυσης στο παραγόμενο αποτέλεσμα, αξιοποιεί τα αποτελέσματα της κυματιδιακής ανάλυσης, ώστε να παράξει προϊόν κατάλληλο για επεξεργασία με χρήση εφαρμογής οπτικής αναγνώρισης χαρακτήρων.

Το βασικό πλεονέκτημα της προτεινόμενης μεθοδολογίας κυματιδιακής ανάλυσης είναι η αξιοποίηση της χρονικής διάστασης του σήματος και των ειδικών χαρακτηριστικών των περιοχών κειμένου ως προς το χρόνο, όπως η στατικότητα σε σχέση με τη πληροφορία υποβάθρου και ο σχετικά μεγάλος χρόνος παραμονής τους για λόγους αναγνωσιμότητας.

Για δείγματα video με μικρή χρονική διάρκεια, σε συνδυασμό με την ύπαρξη ιδιαίτερα στατικού υποβάθρου, υψηλής φωτεινότητας ή κειμένου χαμηλής φωτεινής αντίθεσης τα αποτελέσματα είναι λιγότερο ικανοποιητικά, καθώς αναδεικνύονται περιοχές οι οποίες δεν καταγράφουν κείμενο, με αποτέλεσμα η οπτική αναγνώριση χαρακτήρων να καθίσταται σφαλματική ή αδύνατη. Γι' αυτούς τους λόγους προτείνεται βελτίωση του σταδίου εξαγωγής κειμένου και ιδιαίτερα του σταδίου στατιστικής ανάλυσης, ώστε να ληφθούν υπόψη σε μεγαλύτερο βαθμό, τα ιδιαίτερα γεωμετρικά και χωρικά χαρακτηριστικά του κειμένου, όπως η υψηλή χωρική πυκνότητα στο σύνολο αλλά και στους επιμέρους χαρακτήρες, το επιτρεπτό εύρος του λόγου συσχέτισης των δύο διαστάσεων (ύψος – πλάτος χαρακτήρων), καθώς και η αξιοποίηση της χρωματικής πληροφορίας κατά την κυματιδιακή αλλά και τη στατιστική ανάλυση.

Το ολοκληρωμένο εργαλείο εφαρμογής – επίδειξης των προκρινόμενων μεθοδολογιών που δημιουργήσαμε (**Video W3D Toolbox**) για τις ανάγκες της παρούσας εργασίας, απαιτεί την εισαγωγή από το χρήστη, του χρονικού εύρους των περιοχών του video, όπου καταγράφεται συγκεκριμένη πληροφορία κειμένου προς ανάλυση. Προτείνεται η εισαγωγή ενός πρώτου σταδίου ταχείας ανάλυσης στο σύνολο του video ενδιαφέροντος, ώστε να εξαχθούν οι θέσεις των καρέ εμφάνισης και λήξης κάθε υπότιτλου, με στόχο να αυτοματοποιηθεί η διαδικασία εξαγωγής του συνόλου της εντυπωμένης πληροφορίας στατικού, μηχανικά εντυπωμένου κειμένου (υπότιτλοι), ως ascii πληροφορία η οποία θα καταγράψει το κείμενο αλλά και τις θέσεις ή το χρόνο όπου αυτό εμφανίζεται στο video εισόδου.

Στην παρούσα εργασία προσπαθήσαμε να αναγνωρίσουμε μονάχα στατικό, μηχανικά εντυπωμένο κείμενο, η προτεινόμενη από τη βιβλιογραφία κυματιδιακή ανάλυση όμως, δίνει τη δυνατότητα αναγνώρισης ακόμα και δυναμικού ως προς το χώρο και το χρόνο μηχανικά εντυπωμένου αλλά και φυσικού κειμένου σκηνής. Η δυνατότητα αυτή είναι πολύ σημαντική για την πληρέστερη περιγραφή του υποκείμενου οπτικοακουστικού υλικού μέσω καταχώρησης της παραγόμενης πληροφορίας μεταδεδομένων σε κατάλληλες ψηφιακές βάσεις δεδομένων.

Επιπρόσθετα προτείνεται η δημιουργία ή αξιοποίηση μιας επαρκέστερης εφαρμογής οπτικής αναγνώρισης χαρακτήρων καθότι η εφαρμογή ανοικτού λογισμικού tesseract, η οποία αξιοποιήθηκε στην παρούσα εργασία κρίθηκε μη επαρκής ιδιαίτερα κατά την αναγνώριση χαρακτήρων του ελληνικού αλφαβήτου.

Βιβλιογραφία:

- Adelson E. H., E. Simoncelli, and R. Hingorani, **Orthogonal pyramid transforms for image coding**, in *Proc. SPIE* (Cambridge, MA), pp. 50-58, (Oct.1987)
- Aradhye H.B. and G.K. Myers, **Exploiting Videotext “Events” for Improved Videotext Detection**, *Proceedings of Ninth International Conference on Document Analysis and Recognition*, IEEE pp. 894-898, (2007)
- Aslandogan Y.A. and C.T. Yu, **Techniques and Systems for Image and Video Retrieval**, *IEEE Transactions on Knowledge and Data Engineering*, Vol. 11(1), 56–63 (1999)
- Burns, Thomas J., **A Non homogeneous, Spatio-Temporal Wavelet Multiresolution Analysis and Its Application to the Analysis of Motion**. *PhD dissertation*, Air Force Institute of Technology, (1993)
- Burns T. J., S. K. Rogers, M. E. Oxley and D.W Ruck, **A wavelet multiresolution analysis for spatio-temporal signals**, *IEEE Transactions on Aerospace and Electronic Systems*, vol. 32, issue 2, pp. 628-649, (1996)
- Burt P. J. and E. H. Adelson. **The Laplacian pyramid as a compact image code**, *IEEE Trans. Commun.* vol. COM-31, pp. 532-540, (Apr. 1983)
- Chen D., J. M. Odobez and H. Bourlard, **Text detection and recognition in images and video frames**, *Pattern Recognition*, Vol. 37, No. 3, pp. 595-608, (March 2004).
- Civanlar M. R. and A. Pun, **Scalable video coding in frequency domain**, in *Proc. SPIE Symp. visual Commun., Image Processing* (Boston, MA), vol. 1818, pt. 3, pp. 1124-1134. (Nov. 1992)
- Cohen L., **Time-frequency analysis**, *Prentice Hall*, New Jersey, (1995)
- Daubechies, I., **Orthonormal bases of compactly supported wavelets**, *Comm. Pure & Appl. Math.*, 41, 909-996 (1988)
- Delprat N., B. Escudié, P. Guillemain, R. Kronland-Martinet, P. Tchamitchian, and B. Torrési, **Asymptotic wavelet and Gabor analysis: extraction of instantaneous frequencies**. *IEEE Trans. Inf. Th.*, 38:644-664, (1992)
- Dubey P., **Edge Based Text Detection for Multi-purpose Application**, *Proceedings of International Conference Signal Processing*, IEEE, Vol. 4, (2006)
- Fuan Tsai , Chun-Kai Chang , Jian-Yeo Rau , Tang-Huang Lin , Gin-Ron Liu, **3D computation of gray level co-occurrence in hyperspectral image cubes**, *Proceedings of the 6th international conference on Energy minimization methods in computer vision and pattern recognition*, August 27-29, (2007)
- Gabor D., **Theory of communications**, *J. Inst. Elec. Eng.*, 93,429-457,(1946)
- Gllavata J., E. Qeli and B. Freisleben, **Detecting Text in Videos Using Fuzzy Clustering Ensembles**, *Proceedings of the Eighth IEEE International Symposium on Multimedia*, pp. 283-290, (2006)
- Gopuillaud P, A. Grosssmann , J. Morlet, **Cycle-Octave and Related Transforms in Seismic Signal Analysis**, *Geoexploration*, 23, 85-102 (1984).
- Grossmann A. and J. Morlet, **Decomposition of Hardy Functions to Square Integrable Wavelets of Constant Shape**, *SIAM J.Math.Anal.*, 15, 723-736 (1984).
- Gunn S.R., **Support Vector Machines for Classification and Regression**, *Technical Report*, Dept. of Electronics and Computer Science, University of Southampton (1998)
- Haar A. **Zur Theorie der orthogonalen Funktionensysteme**, *Mathematische Annalen*, 69, pp 331–371, (1910).
- Haralick Robert M., K. Shanmugam, Itshak Dinstein, **Textural Features for Image Classification**. *IEEE Transactions on Systems, Man, and Cybernetics* SMC-3 (6): 610–621 (1973)

- Hopcroft, J., R. Tarjan, **Efficient algorithms for graph manipulation**, *Communications of the ACM*, 16: 372–378, (1973)
- Hua X. S., P. Yin and H. J. Zhang, **Efficient video text recognition using multiple frame integration**, *IEEE International Conference on Image Processing*, (September 2002).
- Jain A.K. and Y. Zhong, **Page segmentation using texture analysis**, *Pattern Recognition*, Vol. 29 (5), pp. 743–770, (1996)
- Jung K., K.I. Kim and A.K. Jain, **Text information extraction in images and video: a survey**, *Pattern Recognition*, Vol. 37 (5), pp. 977-997, (2004)
- Kim K.I., K. Jung and J.H. Kim, **Texture-based approach for text detection in image using support vector machine and continuously adaptive mean shift algorithm**, *IEEE Transaction Pattern Analysis and Machine Intelligence*, Vol. 25, No. 12, pp. 1631-1638, (2003)
- Li H., D. Doermann and O. Kia, **Automatic text detection and tracking in digital video**, *IEEE Transactions on Image Processing*, Vol. 9, No. 1, pp. 147-156, (January 2000).
- Lienhart R. and A. Wernicke, **Localizing and segmenting text in images and videos**, *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 12, No. 4, pp. 256-268, (April 2002).
- Lim Y.K., S.H. Choi and S.W. Lee, **Text extraction in MPEG compressed video for content-based indexing**, *Proceedings of International Conference on Pattern Recognition*, 2000, pp. 409–412.
- Lumina R., **A new three-dimensional connected components algorithm**, *Comput. Vision, Graphics, Image Processing*, vol. 23, pp. 207-217, (Aug. 1983)
- Lyu M.R., J. Song and M. Cai, **A comprehensive method for multilingual video text detection, localization, and extraction**, *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 15 (2), pp. 243-255, (2005)
- Mallat, S., **Multiresolution approximation and wavelet orthonormal bases of $L^2(\mathbb{R})$** , *Trans. Amer. Math. Soc.*, 315, 69-88 (1989)
- Mallat S.G., **A theory for multiresolution signal decomposition: the wavelet representation**. *IEEE Transactions on PAMI* 11, pp. 674–693 (1989)
- Mancas-Thilou C. and B. Gosselin, **Spatial and Color Spaces Combination for Natural Scene Text Extraction**, *Proceedings of IEEE International Conference on Image Processing*, pp. 985-988, (2006)
- Morlet J, G. Arens, I. Fougéau, and D. Giard, **Wave propagation and sampling theory**, *Geophysics*, 47, 203-206 (1982)
- Newland David E., **Harmonic wavelet analysis**, *Proceedings of the Royal Society of London, Series A (Mathematical and Physical Sciences)*, vol. 443, no. 1917, p. 203–225, (1993)
- Samet H., **Connected component labeling using quadrees**, *J. ACM*, vol. 28, pp. 487-501, (July 1981)
- Samet H., M. Tamminen, **Efficient Component Labeling of Images of Arbitrary Dimension Represented by Linear Bintreees**, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v.10 n.4, p.579-586, (July 1988)
- Shannon C. E., **The Mathematical Theory of Communication**, *Urbana, IL:University of Illinois Press*, (1949, reprinted 1998)
- Smith M. and T. Kanade, **Video Skimming for Quick Browsing Based on Audio and Image Characterization**, *Tech. Report CMU-CS-95-186*, Carnegie Mellon Univ., Pittsburgh, (1995).
- Stromberg J.-O., **A modified Franklin system and high order spline systems on \mathbb{R}^n as unconditional bases for Hardy spaces**, *Conference on Harmonic Analysis in honor of A. Zygmund*, pages 475–494, Wadsworth International Group, Belmont, California (1983)
- Subramanian K., P. Natarajan, M. Decerbo and D. Castañón, **Character-stroke detection for text-localization and extraction**, *Proceedings of the International Conference on Document Analysis and Recognition, ICDAR 1*, art. no. 4378671, pp. 33-37, (2007)

- Taubman D., and A. Zakhor, **Multi-rate 3-D subband coding of video**, *IEEE Transactions on Image Processing* 3, 5, 572-588, (Sept. 1994)
- Theodoridis Sergios and Konstantinos Koutroumbas, **"Pattern Recognition"**, 4th Edition, *Academic Press*, (2009).
- Wang Y.K. and J.M. Chen, **Detecting Video Texts Using Spatial-Temporal Wavelet Transform**, Proceedings - International Conference on Pattern Recognition 4, art. no. 1699950, pp. 754-757, (2006)
- Wong E. K. and M. Chen, **A new robust algorithm for video text Extraction**, *Pattern Recognition*, Vol. 36, No. 6, pp. 1397-1406, June 2003.
- Woods J. and S. O'Neil, **Subband coding of images**, *IEEE Trans. Acoust.. Speech, Signal Processing*, vol. 34, pp. 1278-1288, (Oct. 1986)
- Zweig G., R. Lipes, and J. Pierce, **The Cochlear Compromise**, *J. Acoust. Soc. Amer.* 59, 975 (1976).