

ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ  
ΤΜΗΜΑ ΗΛΕΚΤΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ  
ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ  
ΤΟΜΕΑΣ ΠΛΗΡΟΦΟΡΙΚΗΣ

ΔΙΔΑΚΤΟΡΙΚΗ ΔΙΑΤΡΙΒΗ

ΑΝΑΛΥΣΗ ΚΑΙ ΚΩΔΙΚΟΠΟΙΗΣΗ  
ΚΙΝΟΥΜΕΝΩΝ ΚΑΙ ΑΚΙΝΗΤΩΝ ΕΙΚΟΝΩΝ  
ΜΕ ΧΡΗΣΗ ΝΕΥΡΩΝΙΚΩΝ ΔΙΚΤΥΩΝ



Δημήτριος Κ. Καλογεράς

Διπλ. Ηλεκτρολόγος Μηχανικός ΕΜΠ,

Αθήνα, Ιανουάριος 1996.

## ΔΙΔΑΚΤΟΡΙΚΗ ΔΙΑΤΡΙΒΗ

Δημήτριος Κ. Καλογεράς

Διπλ. Ηλεκτρολόγου Μηχ/κού ΕΜΠ

## ΑΝΑΛΥΣΗ ΚΑΙ ΚΩΔΙΚΟΠΟΙΗΣΗ ΚΙΝΟΥΜΕΝΩΝ ΚΑΙ ΑΚΙΝΗΤΩΝ ΕΙΚΟΝΩΝ ΜΕ ΧΡΗΣΗ ΝΕΥΡΩΝΙΚΩΝ ΔΙΚΤΥΩΝ

Επιβλέπων Καθηγητής: Σ. Κόλλιας

Τριμελής συμβουλευτική επιτροπή: Χ. Χαλκιάς, Γ. Καραγιάννης, Σ. Κόλλιας

Επταμελής εξεταστική επιτροπή:

Χ. Χαλκιάς ,  
Καθηγητής ΕΜΠ

Γ. Καραγιάννης,  
Καθηγητής ΕΜΠ

Σ. Κόλλιας,  
Αν. Καθηγητής ΕΜΠ

Γ. Τζιρίτας,  
Καθηγητής ΙΤΕ

Β. Μάγκλαρης,  
Καθηγητής ΔΠΘ

Α. Σταφυλοπάτης,  
Αν. Καθηγητής ΕΜΠ

Γ. Στασινόπουλος  
Καθηγητής ΕΜΠ

Η εργασία αυτή αφιερώνεται  
στους γονείς μου *Κωνσταντίνο και Μαρία*

# ΠΡΟΛΟΓΟΣ

## ΠΡΟΛΟΓΟΣ

Η διατριβή αυτή εκπονήθηκε στο Εργαστήριο Ψηφιακής Επεξεργασίας Εικόνας του Τομέα Πληροφορικής του Τμήματος Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών του ΕΜΠ. Μέρος της εργασίας εκπονήθηκε σε συνεργασία με το Καθ. Σ. Βαρουφάκη στο εργαστήριο Νευρωνικών Δικτύων του Ε.ΚΕ.ΦΕ “Δημόκριτος”, του οποίου ήμουν υπότροφος κατά το διάστημα Μάρτιος 1991 - Απρίλιος 1995. Ολοκληρώνοντας την διατριβή μου θα ήθελα να ευχαριστήσω τον Αν. Καθηγητή Στέφανο Κόλλια για την θερμή συμπαράσταση και την επιστημονική καθοδήγησή του. Η κατανόησή του σε όλους τους τομείς ήταν διαρκής κατά την εκπόνηση της διατριβής. Το περιβάλλον που δημιούργησε στο εργαστήριο ΨΕΕ με την συνεχή υποστήριξη σε υλικό και λογισμικό συνέβαλε στην απρόσκοπτη εργασία μέσα στο εργαστήριο.

Θερμά επίσης ευχαριστώ και τους καθηγητές μου στο ΕΜΠ και μέλη της συμβουλευτικής επιτροπής κ.κ. Χ. Χαλκιά και Γ. Καραγιάννη, καθώς και τους υπόλοιπους καθηγητές της επταμελούς εξεταστικής επιτροπής κ.κ. Γ. Στασινόπουλο, Β. Μάγκλαρη, Α. Σταφυλοπάτη από το ΕΜΠ, και κ. Γ. Τζιρίτα από το ΙΤΕ, για την τιμή που μου έκαναν συμμετέχοντας σ’ αυτήν. Ευχαριστώ τον κ. Γ. Καραγιάννη επειδή διέθεσε χώρο και υλικά από το εργαστήριο για την ολοκλήρωση απαραίτητων πειραμάτων. Ευχαριστώ τον κ. Α. Σταφυλοπάτη για τα βιβλία και περιοδικά της βιβλιοθήκης του ήταν πάντα στη διάθεσή μου όλα αυτά τα χρόνια. Οι συζητήσεις που είχαμε μαζί σε θέματα νευρωνικών δικτύων ήταν ιδιαίτερα αποκαλυπτικές και βοήθησαν την έρευνα στο συγκεκριμένο πεδίο.

Ιδιαίτερα ευχαριστώ τον κ. Χ. Χαλκιά για την δυνατότητα συμμετοχής μου στην ελληνική ομάδα στην Συνάντηση των 7 πύο Αναπτυγμένων Χωρών στο Κόσμο (G7) στην έκθεση για τις λεωφόρους της πληροφορικής που έγινε στις Βρυξέλλες τον Φεβρουάριο του

1995. Στην έκθεση παρουσιάστηκε από την πλευρά του εργαστηρίου ΨΕΕ ένα σύστημα μετάδοσης εικόνας σε πραγματικό χρόνο μέσω δικτύου υπολογιστών χρησιμοποιώντας τεχνολογία MPEG-1. Υπήρχε δυνατότητα για χειρισμό του συστήματος και επιλογής καναλιών μέσω WWW. Από ερευνητική άποψη το εγχείρημα δεν παρουσίαζε πρωτοτυπία, αλλά από αναπτυξιακή ήταν διεθνής πρωτοτυπία. Πολλά προβλήματα παρουσιάστηκαν, μερικά επιλύθηκαν ενώ άλλα κατέδειξαν τον δύσκολο δρόμο σε εφαρμογές Video On Demand, και θέματα συγχρονισμού δέκτη με πομπό σε ασύγχρονα δίκτυα δεδομένων.

Η εργασία αυτή θα ήταν πολύ δυσκολότερη για μένα χωρίς τη διαρκή συμπαράσταση των συναδέλφων και φίλων μου της ομάδας Επεξεργασίας Εικόνων, Τάσου Ντελόπουλου, Αντρέα Τυράκη, Νίκου Παναγιωτίδη, Σπύρου Μιχαηλίδη, Βασίλη Αλεξόπουλου και Μαρίας Ραγκούση από το εργαστήριο ΨΕΣ η οποία ήταν πάντα πρόθυμη για επίλυση αποριών σε βασικά θέματα επεξεργασίας σημάτων. Για το ωραίο κλίμα συνεργασίας κι αλληλοϋποστήριξης που με την καθημερινή τους παρουσία έχτισαν, για τις ενδιαφέρουσες επιστημονικές συζητήσεις που είχαμε, αλλά και για το χιούμορ, το κέφι, και την όρεξή τους για δουλειά - που ως γνωστόν είναι ... κολλητικά - τους ευχαριστώ θερμά.

Ταυτόχρονα με την εργασία στο εργαστήριο ΨΕΕ εργάστηκα στο Ε.Μ.Π σε διάφορα ερευνητικά προγράμματα. Η πλειοψηφία τους σχετιζόταν με γενική επεξεργασία εικόνας. Στο πρόγραμμα "Υδροσκόπιο" ανέλαβα την υλοποίηση ενός δικτύου δεδομένων ευρείας περιοχής για την εξυπηρέτηση της πανελληνίας υδρολογικής και μετεωρολογικής βάσης δεδομένων. Το δίκτυο κάλυπτε τρία Α.Ε.Ι και εννέα δημόσιους φορείς και αποτελούσε το μεγαλύτερο ερευνητικό δίκτυο δεδομένων στον Ελληνικό χώρο την δεδομένη χρονική στιγμή. Στην συνέχεια εργάστηκα στο Ε.Μ.Π στο έργο της υλοποίησης του νέου δικτύου δεδομένων σε συνεργασία με τον Καθ. Β. Μάγκλαρη. Θα ήθελα από αυτό το σημείο να εκφράσω τις ευχαριστίες μου στην ομάδα του Κέντρου Ελέγχου Δικτύου η οποία ανέχθηκε την εκτέλεση της διατριβής και με αναπλήρωσε αρκετές φορές στο έργο της διαχείρισης του δικτύου. Τους ευχαριστώ θερμά.

Θέλω επίσης να ευχαριστήσω τους φίλους του Εργαστηρίου Λογισμικού Αχιλλέα Βολιώτη, Αλέξη Ζάβρα, Γιώργο Ευθυβουλίδη, Παναγιώτη Χριστιά, Παναγιώτη Σέμπο, αλλά

και τον υπεύθυνο καθηγητή κ. Ε. Σκορδαλάκη, γιατί η ύπαρξη της σύνδεσης με το Interpreter, και η παροχή των υπηρεσιών που αυτό συνεπάγεται, καθοριστική και για την εκπόνηση της παρούσας εργασίας αλλά και για την όλη ζωή του εργαστηρίου και του Τομέα, είναι δικό τους έργο και στηρίζεται στην ανιδιοτελή προσφορά τους. Ευχαριστώ την Ναυσικά Τζάνου και τον Αναστάσιο Ντελόπουλο για το δύσκολο και βαρετό έργο της ανάγνωσης και διόρθωσης του κειμένου. Πολλές φορές ήταν αδύνατη η διόρθωση των λαθών μετά από την συνεχή ανάγνωση των κειμένων.

Δημήτριος Κ. Καλογεράς

Ιανουάριος 1996

# Πίνακας Περιεχομένων

<b>ΠΡΟΛΟΓΟΣ</b>	<b>iv</b>
<b>I Εισαγωγή - Νευρωνικά δίκτυα στην Ανάλυση Εικόνων</b>	<b>1</b>
<b>1 Εισαγωγή</b>	<b>2</b>
1.1 Γενικά στοιχεία για την Ψηφιακή Επεξεργασία Εικόνας . . . . .	3
1.2 Οργάνωση και συνεισφορά της διατριβής . . . . .	6
<b>2 Μοντέλα νευρωνικών δικτύων στην ανάλυση εικόνων</b>	<b>11</b>
2.1 Εισαγωγικά στοιχεία για την επεξεργασία με νευρωνικά δίκτυα . . . . .	11
2.2 Κατηγοριοποίηση νευρωνικών δικτύων . . . . .	13
2.3 Πολυεπίπεδα perceptrons . . . . .	14
2.4 Διανυσματικός Κβαντιστής με εκμάθηση . . . . .	17
<b>3 Τεχνική Fractals στην ανάλυση και αναγνώριση εικόνων</b>	<b>23</b>
3.1 Εισαγωγή . . . . .	23
3.2 Εύρεση της κλασματικής διάστασης με την μέθοδο Μέτρησης με “Κύβους”	27
3.3 Εκτίμηση κλασματικής Διάστασης D διδιάστατων αντικειμένων . . . . .	28
3.3.1 Μέθοδος της συνεχούς επιστρωμάτωσης . . . . .	29
3.4 Κλασματικός Πίνακας . . . . .	30
3.5 Πολυκαναλικός Κλασματικός Πίνακας . . . . .	33
3.6 Μια πολυμοντελική προσέγγιση για την ταξινόμηση εικόνων με νευρωνικά δίκτυα . . . . .	37
3.7 Συμπεράσματα . . . . .	40

4	Αποτελεσματική κατηγοριοποίηση εικόνων με χρήση νευρωνικών δικτύων και πολυδιακριτής ανάλυσης	44
4.1	Εισαγωγή	44
4.2	2-Δ Πολυδιακριτική Ανάλυση	46
4.3	Πιθανοτικά Δίκτυα	49
4.4	Νευρωνικά δίκτυα για την πολυδιακριτή κατηγοριοποίηση εικόνων	51
4.5	Αποτελέσματα Προσομιώσεων - Συγκριτικές μελέτες	55
4.6	Συμπεράσματα	58
<b>II</b>	<b>Κωδικοποίηση και Συμπύεση Κινούμενων Εικόνων</b>	<b>62</b>
5	Πρότυπα Κωδικοποίησης Κινούμενων Εικόνων	63
5.1	Εισαγωγή	63
5.2	H.261	67
5.3	MPEG-1	69
5.4	MPEG-2	72
6	Συμπύεση εικόνων σε χαμηλούς ρυθμούς μετάδοσης	82
6.1	Εισαγωγή	82
6.2	Κωδικοποίηση Με Μοντέλα (KMM)	85
6.2.1	Κωδικοποίηση με αντικείμενα	86
6.2.2	Κωδικοποίηση με γνωστά αντικείμενα	90
6.3	Υβριδικός Κωδικοποιητής με Περιοχές Ενδιαφέροντος (ΔΜΣ-ΑΚ-ΠΕ)	92
6.4	Προσαρμοζόμενη Επιλογή Περιοχών Ενδιαφέροντος με Χρήση Νευρωνικών Δικτύων	96
6.5	Θέματα Υλοποίησης - Αριθμητικά Αποτελέσματα	98
6.6	Συμπεράσματα - Θέματα για περαιτέρω έρευνα	102
7	Συμπεράσματα της διατριβής	112



# Μέρος I

## Εισαγωγή - Νευρωνικά δίκτυα στην Ανάλυση Εικόνων

# Εισαγωγή

Αντικείμενο της παρούσας εργασίας είναι η ανάπτυξη εργαλείων ειδικής χρήσης για την ψηφιακή επεξεργασία εικόνας. Η ψηφιακή επεξεργασία εικόνας αποτελεί ένα ευρύτατο ερευνητικό πεδίο, όπου ποικιλία μεθόδων έχει δοκιμαστεί σε εξίσου μεγάλη ποικιλία προβλημάτων. Παρ' όλη την ερευνητική δραστηριότητα στην περιοχή αυτή, ένα πλήθος προβλημάτων παραμένουν ανοιχτά, καθώς οι μέχρι σήμερα χρησιμοποιούμενες τεχνικές δείχνουν να έχουν εξαντλήσει τις δυνατότητές τους. Είναι συνεπώς λογικό να αναζητηθεί διέξοδος μέσω λύσεων που διαφοροποιούνται ουσιαστικά από τις κλασικά χρησιμοποιούμενες. Μία τέτοια προσέγγιση επιχειρείται στην παρούσα εργασία. Διερευνάται το εφαρμόσιμο ορισμένων ειδικών μεθόδων σε συγκεκριμένα προβλήματα της ψηφιακής επεξεργασίας εικόνας. Συγκεκριμένα, αναπτύσσονται σύγχρονες τεχνικές που στηρίζονται σε μη γραμμική ανάλυση με fractals και νευρωνικά δίκτυα πολυδιακριτής ανάλυσης για την επίλυση προβλημάτων ανάλυσης και κατηγοριοποίησης εικόνας. Παρουσιάζονται νέες τεχνικές κωδικοποίησης κινούμενων εικόνων σε χαμηλούς ρυθμούς μετάδοσης που βασίζονται σε ένα αμάγαλμα παραδοσιακών τεχνικών κωδικοποίησης και έξυπνων νευρωνικών δικτύων. *A priori* γνώση εμφυτεύεται στον κωδικοποιητή για την κωδικοποίηση συγκεκριμένου τύπου κινούμενων εικόνων (video). Τα πειραματικά αποτελέσματα ελέγχου της επίδοσης των προτεινόμενων μεθόδων πάνω σε πραγματικά δεδομένα εικόνων επιβεβαιώνουν τις δυνατότητες που οι μέθοδοι αυτές προσφέρουν στον ερευνητή για την επιτυχή αντιμετώπιση ορισμένων προβλημάτων της ψηφιακής επεξεργασίας εικόνας.

## 1.1 Γενικά στοιχεία για την Ψηφιακή Επεξεργασία Εικόνων

Η ψηφιακή επεξεργασία σημάτων έχοντας τις ρίζες της στα μαθηματικά του 17<sup>ου</sup> και 18<sup>ου</sup> αιώνα και σημερινό στήριγμά της την πληθώρα των ψηφιακών συστημάτων, αποτελεί ένα γοργά εξελισσόμενο ερευνητικό πεδίο με εφαρμογές που αλλάζουν την εικόνα, (ίσως και την ηθική), της σύγχρονης κοινωνίας.

Ένας από τους βασικότερους κλάδους της επεξεργασίας σημάτων είναι η ψηφιακή επεξεργασία εικόνων. Θα μπορούσε να ειπωθεί ότι οι ερευνητές του κλάδου αυτού έχουν σαν απώτερο στόχο (ή ελπίδα) την υλοποίηση της απόλυτης μηχανής με δυνατότητες εφάμιλλες των οπτικών λειτουργιών των ζώντων οργανισμών, επίτευγμα για το οποίο απαιτούνται νέα θεωρητικά και τεχνολογικά άλματα. Πρόσφατες πάντως τεχνολογικές εξελίξεις σε πολλά πεδία της ψηφιακής επεξεργασίας εικόνων έχουν επιτρέψει την επίλυση προβλημάτων που μέχρι πρόσφατα περιόριζαν την ουσιαστική αξιοποίηση της πληροφορίας σε μορφή εικόνων. Μερικά από αυτά τα πεδία είναι τα εξής:

### α) Αναπαράσταση και μοντελοποίηση εικόνων.

Η αναπαράσταση εικόνων ασχολείται με τον χαρακτηρισμό της ποσότητας που κάθε σημείο της εικόνας (picture element ή pixel) παριστάνει. Οι εικόνες μπορεί να παριστούν φωτεινότητες αντικειμένων σε σκηνές (π.χ. φωτογραφίες λαμβανόμενες από κλασικές φωτογραφικές μηχανές), τα χαρακτηριστικά απορρόφησης του ανθρώπινου ιστού (ακτινογραφίες, X-ray imaging), σήματα ραντάρ (radar imaging), την θερμοκρασιακή εικόνα μίας περιοχής (infrared imaging), ή το βαρυτικό πεδίο μίας περιοχής (geophysical imaging). Γενικά, κάθε διδιάστατη συνάρτηση που περιέχει πληροφορία μπορεί να θεωρηθεί ως εικόνα. Τα μοντέλα εικόνων παρέχουν μία λογική και ποσοτική περιγραφή των ιδιοτήτων της συνάρτησης αυτής.

### β) Βελτίωση της ποιότητας εικόνων.

Η βελτίωση της ποιότητας εικόνων (Image Enhancement) αποσκοπεί στο να δώσει στην εικόνα μία μορφή περισσότερο κατάλληλη για περαιτέρω επεξεργασία ή ανάλυση. Οι

μέθοδοι για την βελτίωση ποιότητας χωρίζονται σε δύο βασικές κατηγορίες, επεξεργαζόμενες την εικόνα στο πεδίο του χώρου ή της συχνότητας. Οι μέθοδοι αυτές επεκτείνονται και ως προς το χρόνο, στην περίπτωση που επεξεργαζόμαστε ακολουθίες εικόνων (σήμα video). Στο πεδίο του χώρου οι τεχνικές αυτές εκμεταλλεύονται τις γεωμετρικές ιδιότητες της εικόνας, στο πεδίο της συχνότητας επεξεργάζονται το φάσμα της εικόνας, ενώ στο πεδίο του χρόνου εκμεταλλεύονται το συσχετισμό της πληροφορίας από εικόνα σε εικόνα.

#### γ) Αποκατάσταση εικόνων.

Η αποκατάσταση εικόνων (Image Restoration) βασίζεται στις θεωρίες εκτίμησης (estimation theory) και σχεδίασης ψηφιακών φίλτρων. Σκοπός της είναι η αφαίρεση παραμορφώσεων, όπως του τυχαίου θορύβου, του θολώματος της εικόνας, της έλλειψης καλής αντίθεσης (contrast), διαφόρων παρεμβολών και της μη ομοιογένειας των πεδίων της εικόνας. Οι τεχνικές αποκατάστασης εικόνων συνήθως βασίζονται σε μαθηματικά στοχαστικά μοντέλα παράστασης των εικόνων. Χρησιμοποιούν γραμμικές μεθόδους, όπως φίλτρα αντιστροφής, φίλτρα Wiener και Kalman, αλλά και μη γραμμικές όπως ομομορφικά φίλτρα, μεθόδους μέγιστης εντροπίας και εκτίμησης με το κριτήριο του Bayes, μορφολογικές μεθόδους.

#### δ) Ανάλυση εικόνων.

Η ανάλυση εικόνων (Image Analysis), αποτελεί τη βασική λειτουργία της όρασης υπολογιστών και αναφέρεται στην εξαγωγή της χρήσιμης πληροφορίας από μία εικόνα ή ένα σύνολο εικόνων, όπως και στην κατανόηση ή κατάταξη της πληροφορίας αυτής. Χρησιμοποιεί τεχνικές επεξεργασίας εικόνων αλλά και τεχνικές αναγνώρισης προτύπων και τεχνητής νοημοσύνης. Μερικοί από τους σκοπούς της ανάλυσης εικόνων είναι:

- i) Η επίτευξη μίας πιο αποτελεσματικής παράστασης της εικόνας. Για λόγους π.χ. συμπίεσης η εικόνα μπορεί να σταλεί υπό μορφή σχίτσου (περιγραμμάτων των περιοχών της) ή μέσω ιεραρχικών διαγραμμάτων.
- ii) Η έμφαση συγκεκριμένων χαρακτηριστικών της εικόνας, που είναι χρήσιμα για ανίχνευση, εκτίμηση, κατάταξη και αναγνώριση διδιάστατων ή τριδιάστατων αντικειμένων από διδιάστατες ή τριδιάστατες εικόνες. Η εκτίμηση των παραμέτρων ενός αντικειμένου, σαν το μέγεθος, την ταχύτητα ή την επιφάνειά του μπορεί να προηγείται ή να έπεται της

ανίχνευσης του αντικειμένου.

iii) Η χρησιμοποίηση μετασχηματισμών για καλύτερη απεικόνιση και στη συνέχεια εκτίμηση του αντικειμένου.

ε) Κωδικοποίηση εικόνων.

Η κωδικοποίηση εικόνων (Image Coding) είναι απαραίτητη για την συμπίεση του τεράστιου όγκου πληροφορίας των εικόνων και ιδιαίτερα των ακολουθιών αυτών, για μετάδοση ή αποθήκευση. Η συμπίεση επιτυγχάνεται με την εκμετάλλευση του πλεονασμού που υπάρχει στις εικόνες. Η αποτελεσματική κωδικοποίηση εικόνων αποκτά ιδιαίτερο και επιτακτικό ενδιαφέρον στις μέρες μας που γεννιέται η ψηφιακή τηλεόραση, η τηλεόραση υψηλής ευκρίνεια (HDTV), και σε multimedia εφαρμογές που στηρίζονται ιδιαίτερα στην οπτική πληροφορία. Επιπλέον η απαίτηση για μετάδοση εικόνων διαμέσου κοινών τηλεφωνικών γραμμών έχει δημιουργήσει ένα νέο πεδίο έρευνας: κωδικοποίηση εικόνων σε πολύ χαμηλούς ρυθμούς. Τό νέο πεδίο έχει φέρει στην επιφάνεια νέες τεχνικές οι οποίες βασίζονται σε τεχνικές ανάλυσης και βελτίωσης εικόνων. Υπό αυτή την έννοια παρατηρείται μία σύγκλιση των διάφορων περιοχών της επεξεργασίας εικόνων για την λύση δύσκολων προβλημάτων.

Οι εφαρμογές της ψηφιακής επεξεργασίας εικόνων καλύπτουν ένα ευρύ φάσμα της σημερινής τεχνολογίας όπως:

- Διαστημικές εφαρμογές: μετάδοση, βελτίωση ποιότητας και αποκατάσταση, κατανόηση εικόνων που λαμβάνονται από διαστημόπλοια.
- Ιατρικές και βιοϊατρικές εφαρμογές : ανάκτηση, επεξεργασία, βελτίωση ποιότητας, μετάδοση, αποθήκευση ιατρικών εικόνων (π.χ. ακτινογραφιών) μαζί με τα σήματα (π.χ. καρδιογραφήματα) και κείμενα (π.χ. ιστορικό ασθενών) σε βάσεις δεδομένων (ή και βιντεοδίσκους), τομογραφία, αυτόματη διάγνωση ασθενειών από ιατρικές εικόνες (τηλεϊατρική).
- Αυτοματοποίηση βιομηχανικών εφαρμογών, ρομποτική όραση, όραση υπολογιστών. ανάπτυξη υπολογιστών “πέμπτης γενιάς”.
- Τηλεόραση υψηλής ευκρίνειας, τηλεσυνέντευξη, multimedia εφαρμογές.

- Δορυφορική φωτογράφιση γήινης επιφάνειας: αποθήκευση τεράστιου όγκου συμπιεσμένων ψηφιακών εικόνων, αυτόματη χαρτογράφηση, αναγνώριση του είδους καλλιέργειας γεωργικών εκτάσεων, σεισμολογικές έρευνες.

## 1.2 Οργάνωση και συνεισφορά της διατριβής

Το αντικείμενο της παρούσας διατριβής είναι η ανάπτυξη νέων αποτελεσματικών τεχνικών στην περιοχή της ανάλυσης ακίνητων εικόνων και κωδικοποίησης κινούμενων εικόνων. Ήδη αναφέραμε ότι σκοπός της ανάλυσης εικόνων είναι κυρίως η κατάλληλη εξαγωγή χαρακτηριστικών από την εικόνα ώστε να είναι δυνατή η κατάτμησή της σε ομογενείς περιοχές, η ταξινόμηση των περιοχών αυτών, η ανίχνευση και αναγνώριση διαφόρων αντικειμένων μέσα στην εικόνα.

Ένα σημαντικό στοιχείο των μεθόδων ανάλυσης εικόνων είναι το μοντέλο αναπαράστασης που χρησιμοποιούν. Ένα μοντέλο το οποίο θα μπορούσε να αναπαραστήσει με “καλό τρόπο” τόσο τεχνητά αντικείμενα όσο και φυσικές εικόνες είναι πολύ πιθανό να είναι κατάλληλο για τις διαδικασίες ανάλυσης εικόνας. Επιπλέον, είναι επιθυμητό οι διαδικασίες ανάλυσης εικόνων να λαμβάνουν υπόψη και την χρωματική πληροφορία η οποία είναι δυνατόν να είναι συζευγμένη με την ήδη υπάρχουσα πληροφορία αναπαράστασης της εικόνας. Η εκμετάλλευση της πληροφορίας χρώματος μπορεί να βοηθήσει αρκετά στην περίπτωση που η πληροφορία της υψής και της φωτεινότητας δεν είναι αρκετές. Ένα από τα βασικά προβλήματα της ανάλυσης εικόνων είναι οι συνήθως μεγάλες διαστάσεις τους, με αποτέλεσμα την χρονοβόρα επεξεργασία, ιδιαίτερα στην περίπτωση επεξεργασίας πολυκαναλικών εικόνων. Άλλο σημαντικό πρόβλημα είναι η παρουσία θορύβου στις εικόνες, η οποία δυσχεραίνει την ανίχνευση και ταξινόμηση των διαφόρων αντικειμένων που αυτές περιέχουν. Επίσης θα πρέπει να αναφέρουμε το πρόβλημα των πιθανών παραμορφώσεων αυτών των αντικειμένων παράλληλα με δυνατές αλλαγές στις σχετικές θέσεις που αυτά εμφανίζονται μέσα στην εικόνα, γεγονότα που μεγιστοποιούν την δυσκολία αναγνώρισης. Τα μοντέλα αναπαράστασης που βασίζονται στα fractals μπορούν να αναπαραστήσουν με μεγαλύτερη πιστότητα αντικείμενα και φυσικές σχημές απ’ ό,τι με παραδοσιακές τεχνικές.

Η αναπαράσταση με fractals προσδίδει αναισθησία σε αλλαγές της κλίμακας μέτρησης γιατί τα αντικείμενα που χαρακτηρίζονται ως fractals περιγράφονται από μια διδιαστατή ποσότητα η οποία είναι ανεξάρτητη από την κλίμακα ανάλυσης. Ένα τεχνητό νευρωνικό δίκτυο τύπου LVQ χρησιμοποιήθηκε για το συγγενές πρόβλημα κατηγοριοποίησης εικόνων. Το πρόβλημα της ανάλυσης / κατηγοριοποίησης εικόνων είναι δυνατόν να προσεγγιστεί και με ανάλυση subband. Η ανάλυση subband μπορεί να επιβάλει περιορισμούς στην κατασκευή των κατηγοριοποιητών με αποτέλεσμα να βελτιωθεί η λειτουργία τους. Το δεύτερο πρόβλημα που αντιμετωπίζεται είναι η κωδικοποίηση κινουμένων εικόνων σε πολύ χαμηλούς ρυθμούς. Οι υπάρχοντες κωδικοποιητές παρουσιάζουν προβλήματα ποιότητας σε πολύ χαμηλούς ρυθμούς. Κωδικοποιητές οι οποίοι έχουν γνώση των αντικειμένων που περιέχονται στις υπό κωδικοποίηση εικόνες μπορούν να αποτελέσουν εναλλακτική λύση. Η γνώση ενταμιεύεται στον κωδικοποιητή με την μορφή νευρωνικών δικτύων τύπου LVQ. Ως συνέπεια των προηγούμενων, η παρούσα εργασία συντίθεται από τα εξής δύο μέρη:

### Μέρος I

Στο χώρο της ανάλυσης εικόνων, η κατάταξη (classification) περιοχών/αντικειμένων αντιμετωπίζει το πρόβλημα προσδιορισμού του τύπου της περιοχής ή της ταυτότητας του αντικειμένου, συνήθως σε σχέση με κάποιους γνωστούς τύπους αντικειμένων ή περιοχών. Αυτό το πρόβλημα επιλύεται με εφαρμογή της θεωρίας αναγνώρισης προτύπων. Οι τεχνικές κατάταξης (ταξινόμησης, κατηγοριοποίησης) είναι είτε με δάσκαλο (supervised), όπου σε ένα στάδιο εκμάθησης (training) ο ταξινομητής εκπαιδεύεται να αναγνωρίζει τις κλάσεις των αντικειμένων, είτε χωρίς δάσκαλο (clustering). Μία ειδική κατηγορία ταξινομητών είναι τα τεχνητά νευρωνικά δίκτυα (artificial neural networks). Τα νευρωνικά δίκτυα είναι συστήματα με μεγάλη παραλληλία, που προσπαθούν να προσομοιάσουν την λειτουργία των βιολογικών νευρώνων. Λόγω της παραλληλίας τους αλλά και άλλων ιδιοτήτων που έχουν, τα νευρωνικά δίκτυα θα χρησιμοποιηθούν εκτεταμένα στην εργασία αυτή, με διάφορες αρχιτεκτονικές σαν μέρη των σχημάτων των νέων τεχνικών που θα εισάγουμε. Έτσι το πρώτο Μέρος, εκτός από αυτό το εισαγωγικό για την διατριβή κεφαλαίο (**Κεφάλαιο 1**), αφιερώνεται στην παρουσίαση των νευρωνικών δικτύων, με έμφαση στη συνεισφορά τους στην ανάλυση εικόνων.

Στο **Κεφάλαιο 2**, παρουσιάζουμε τα νευρωνικά δίκτυα, με έμφαση στα πολυεπίπεδα perceptrons και τους διανυσματικούς κβαντιστές με εκμάθηση (LVQ). Αναλύονται οι δύο σημαντικές φάσεις λειτουργίας των νευρωνικών δικτύων με επίβλεψη: η φάση εκπαίδευσης και η φάση ελέγχου. Τα νευρωνικά δίκτυα συνήθως δουλεύουν σαν ταξινομητές σε προβλήματα αναγνώρισης προτύπων με είσοδο κατάλληλα χαρακτηριστικά που εξάγονται από τα πρότυπα αυτά. Το πρόβλημα της ανάλυσης / κατηγοριοποίησης εικόνων είναι πρόβλημα αναγνώρισης προτύπων. Στο πρόβλημα της ανάλυσης εικόνων στο **Κεφάλαιο 3** παρουσιάζονται τα fractals και μία από τις χαρακτηριστικές τους ποσότητες: η κλασματική διάσταση. Τα fractals είναι ένα μοντέλο αναπαράστασης αντικειμένων και φυσικών σκηνών. Παρουσιάζονται διάφορες μέθοδοι για την εκτίμηση της κλασματικής διάστασης όπως η μέθοδος μέτρησης με "Κύβους". Επέκταση της κλασματικής διάστασης σε προβλήματα δύο διαστάσεων και εισαγωγή της μεθόδου της συνεχούς επιστρωμάτωσης δίνεται στην συνέχεια. Ο κλασματικός πίνακας είναι μία εναλλακτική προσέγγιση στην χρησιμοποίηση της κλασματικής διάστασης σε δύο διαστάσεις. Ο κλασματικός πίνακας στην πολυκαναλική περίπτωση εξελίσσεται ως πολυκαναλικός κλασματικός πίνακας. Εξετάζονται δύο περιπτώσεις χρωματικής αναπαράστασης των εικόνων και διερευνώνται οι δυνατότητες του πολυκαναλικού κλασματικού πίνακα στην κατάτμηση συνθετικών εικόνων. Στην συνέχεια παρουσιάζεται περιληπτικά μία πολυμοντελική μέθοδος που βασίζεται σε συσχετίσεις τρίτης τάξης και στον κλασματικό πίνακα για την κατηγοριοποίηση πραγματικών εικόνων. Εξάγονται συμπεράσματα για τις επιδόσεις της κάθε μεθόδου ατομικά και σε συνεργασία.

Στο **Κεφάλαιο 4** αντιμετωπίζεται το πρόβλημα της ανάλυσης εικόνων μεγάλων διαστάσεων. Η εισαγωγή της πολυδιακριτής ανάλυσης σημάτων (subband analysis) δίνει τη δυνατότητα για παράλληλα σχήματα επεξεργασίας των εικόνων όχι μόνο στις αρχικές τους διαστάσεις αλλά και σε πολλαπλά διακριτικά επίπεδα. Στην συνέχεια δίνεται ο ορισμός του πιθανοτικού δικτύου. Ακολούθως περιγράφεται ένα αποτελεσματικό σχήμα κατηγοριοποίησης με απλά feedforward νευρωνικά δίκτυα τα οποία εκπαιδεύονται σε εικόνες μικρότερων διακριτικών επιπέδων, ξεκινώντας από ένα μικρότερο διακριτικό επίπεδο όπου το πρόβλημα περιγράφεται ικανοποιητικά (οι μικρότερης διάστασης εικόνες διατηρούν



επαρκές ποσοστό πληροφορίας). Ακολουθεί ο έλεγχος κατηγοροποίησης στο συγκεκριμένο διακριτικό επίπεδο και εάν τα αποτελέσματα δεν είναι ικανοποιητικά, αναφέρεται τρόπος αναδρομικής κατασκευής μεγαλύτερων δικτύων που θα χειρίζονται εικόνες των μεγαλύτερων διακριτικών επιπέδων. Η συγκεκριμένη αρχιτεκτονική στην περίπτωση των πιθανοτικών δικτύων καταλήγει στο ιεραρχικό πιθανοτικό δίκτυο. Στο τέλος παρουσιάζεται περιληπτικά αντίστοιχη εργασία με πολυεπίπεδα perceptron και δίνονται συγκριτικά αποτελέσματα.

## Μέρος II

Στο δεύτερο μέρος στο **Κεφάλαιο 5** παρουσιάζονται οι ευρέως διαδεδομένοι κωδικοποιητές κινούμενων εικόνων, H.261, MPEG-1, και MPEG-2, που βασίζονται στο Διακριτό Μετασχηματισμό Συνημίτονου (ΔΜΣ) και στην αντιστάθμιση κίνησης. Στο **Κεφάλαιο 6** αντιμετωπίζεται το πρόβλημα της συμπίεσης / κωδικοποίησης κινούμενων εικόνων σε πολύ χαμηλούς ρυθμούς για την μετάδοση μέσα από κοινές τηλεφωνικές γραμμές. Το πρόβλημα αυτό είναι δύσκολο επειδή απαιτεί μεγάλους λόγους συμπίεσης. Εξαιτίας της δυσκολίας του προβλήματος στην γενική μορφή του, επιλέγεται να λυθεί το πρόβλημα με περιορισμένο τύπο κινούμενων εικόνων αυτές των εικόνων τηλεσυνδιάσκεψης. Μια προσέγγιση για την λύση του προβλήματος βασίζεται στις τεχνικές Κωδικοποίησης Με Μοντέλο (ΚΜΜ) οι οποίες παρουσιάζονται εισαγωγικά σε αυτό το κεφάλαιο. Κωδικοποίηση με συμπαγή ή παραμορφώσιμα αντικείμενα με διδιάστατη ή τριδιάστατη κίνηση είναι μερικές από τις περιπτώσεις που εξετάζονται. Η κωδικοποίηση με γνωστά αντικείμενα είναι η περίπτωση όπου έχουμε συγκεκριμένου τύπου αντικείμενα (π.χ πρόσωπα) για τα οποία κωδικοποιούνται οι παράμετροι του μοντέλου που τα περιγράφει. Στη συνέχεια εξετάζεται μία διαφορετική προσέγγιση όπου αναλύονται τα προβλήματα μειωμένης ποιότητας των διεθνών προτύπων συμπίεσης εικόνων σε χαμηλούς ρυθμούς μετάδοσης σε συνδυασμό με τις ιδιότητες των εικόνων τηλεσυνδιάσκεψης. Εισάγεται η έννοια των Περιοχών Ενδιαφέροντος ως μέσο για την καταπολέμηση πολλών από τα προβλήματα ποιότητας. Οι Περιοχές Ενδιαφέροντος προκαθορίζουν την ποιότητα σε περιοχές οι οποίες είναι σημαντικές για υποκειμενικά καλή οπτική αντίληψη της εικόνας. Οι κωδικοποιητές με περιοχές ενδιαφέροντος παρουσιάζουν ομοιότητες με τους ΚΜΜ. Ο μηχανισμός για την επιλογή των περιοχών ενδιαφέροντος

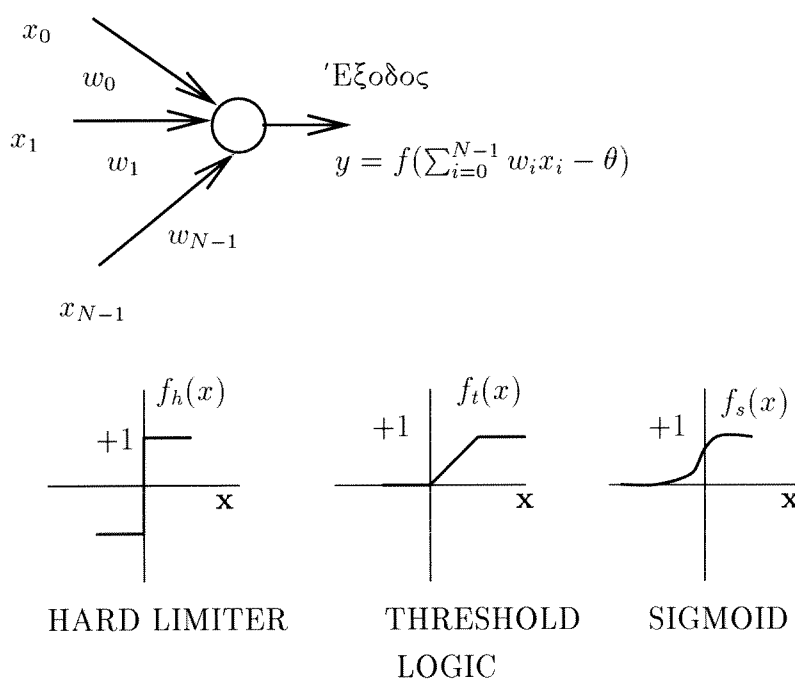
γίνεται με την βοήθεια νευρωνικών δικτύων. Το νευρωνικό δίκτυο εκπαιδεύεται και στην συνέχεια μπορεί να επιλέγει τις περιοχές ενδιαφέροντος σε πραγματικό χρόνο. Οι επιδόσεις του κωδικοποιητή παρουσιάζονται στην συνέχεια και γίνονται συγκρίσεις αφενός με τους συμβατικούς κωδικοποιητές video και αφετέρου με άλλους κωδικοποιητές που προτείνονται από την επιστημονική κοινότητα. Για τις συγκρίσεις χρησιμοποιούνται τυπικές ακολουθίες τηλεσυνδιάσκεψης.

# Μοντέλα νευρωνικών δικτύων στην ανάλυση εικόνων

## 2.1 Εισαγωγικά στοιχεία για την επεξεργασία με νευρωνικά δίκτυα

Τα τεχνητά νευρωνικά δίκτυα αποτελούν ένα πεδίο εντατικής έρευνας για αρκετά χρόνια, με στόχο την υλοποίηση μηχανών με δυνατότητες αντίστοιχες των ανθρώπινων, κυρίως στις περιοχές της αναγνώρισης φωνής και εικόνας. Τα τεχνητά νευρωνικά δίκτυα, ή απλούστερα νευρωνικά δίκτυα, έχουν παρουσιαστεί στη διεθνή βιβλιογραφία με πολλά ονόματα όπως connectionist models, parallel distributed processing models, neuromorphic systems, κάτι που συνδέεται με το γεγονός ότι το πεδίο αυτό οικοδομείται βάση μίας συνισταμένης προσπάθειας πολλών διαφορετικών επιστημονικών χώρων, όπως της ψυχολογίας, βιολογίας, νευροφυσικής, αλλά και των μαθηματικών, φυσικής, αυτομάτου ελέγχου και πληροφορικής.

Στην ουσία, όλα τα νευρωνικά μοντέλα προσπαθούν να επιτύχουν μεγάλη υπολογιστική απόδοση με την χρήση πυκνής διασύνδεσης στοιχειωδών επεξεργαστικών μονάδων, σε μία προσπάθεια προσομοίωσης των βιολογικών νευρικών συστημάτων [2, 5]. Στο σχήμα 2.1 δείχνεται η λειτουργία μίας απλής βασικής επεξεργαστικής μονάδας (κόμβος, νευρώνας) η οποία χρησιμοποιείται συνήθως στα διάφορα νευρωνικά μοντέλα. Παρατηρούμε ότι αθροίζει βεβαρημένες εισόδους, και περνά το άθροισμα από μία μη γραμμική συνάρτηση (τρεις



Σχήμα 2.1: Βασική επεξεργαστική μονάδα (νευρώνας).

περιπτώσεις τέτοιων μη γραμμικών εξισώσεων φαίνονται στο ίδιο σχήμα). Ο νευρώνας χαρακτηρίζεται από το εσωτερικό του offset  $\theta$  και το είδος της μη γραμμικότητας. Πιο πολύπλοκοι νευρώνες μπορεί να χρησιμοποιούν χρονική ολοκλήρωση των τιμών της εισόδου τους και/ή πιο πολύπλοκες μαθηματικές συναρτήσεις από την απλή άθροιση των βεβαρημένων εισόδων.

Τα πλεονεκτήματα των νευρωνικών δικτύων βασίζονται κυρίως στον *παράλληλισμό* της λειτουργίας των βασικών επεξεργαστικών μονάδων, και στην *ανθεκτικότητά* τους σε πιθανή καταστροφή μερικών νευρώνων, η οποία οφείλεται στη πυκνή διασύνδεσή τους. Κάθε νευρώνας λειτουργεί ανεξάρτητα από τους άλλους. Η λειτουργία του εκτελείται με αναλογικό τρόπο, και διέπεται από συνήθεις διαφορικές εξισώσεις. Διακριτή παράσταση χρησιμοποιείται όταν η προσομοίωση του συστήματος γίνεται στον υπολογιστή. Η άθροιση των βεβαρημένων εισόδων αποτελεί ένα γραμμικό φιλτράρισμα, όπου οι συντελεστές του φίλτρου είναι τα βάρη των συνδέσεων του νευρώνα με τους γείτονές του. Τα βάρη αυτά προσδιορίζουν και την λειτουργία του κυκλώματος. Θετικές τιμές των βαρών

αντιστοιχούν σε διεγερτικές (excitatory) συνδέσεις, ενώ αρνητικές τιμές σε απαγορευτικές (inhibitory) συνδέσεις. Όπως θα δούμε στη συνέχεια, ο τρόπος της διασύνδεσης των νευρώνων καθορίζει την αρχιτεκτονική του δικτύου, ενώ τα βάρη των συνδέσεων του δικτύου περικλείουν αυτό που ονομάζεται γνώση του δικτύου. Οι μεταβολές των βαρών γίνονται με βάση κάποιο συγκεκριμένο κανόνα που ονομάζεται αλγόριθμος μάθησης.

## 2.2 Κατηγοριοποίηση νευρωνικών δικτύων

Ας δούμε αναλυτικότερα τις συνιστώσες που περιγράφουν τη δομή και τη λειτουργία ενός νευρωνικού μοντέλου. Αυτές είναι οι εξής:

- α) Σύνολο επεξεργαστικών μονάδων (νευρώνες, κόμβοι).
- β) Τρόπος διασύνδεσης των νευρώνων.
- γ) Κανόνας μετάδοσης της πληροφορίας μέσα στο δίκτυο (propagation rule).
- δ) Κανόνας μάθησης (learning rule).
- ε) Περιβάλλον στο οποίο λειτουργεί το δίκτυο (μορφή εισόδων/εξόδων).
- στ) Λειτουργία του νευρώνα.

Ο καθορισμός των παραπάνω συνιστωσών οδηγεί σε ειδικές κατηγορίες νευρωνικών δικτύων. Ειδικότερα έχουμε:

**Στατικά - Δυναμικά συστήματα:** Τα στατικά δίκτυα (static) χαρακτηρίζονται από εξισώσεις νευρώνων “χωρίς μνήμη” (δηλ. εξαρτώνται μόνο από την τρέχουσα είσοδο), ενώ τα δυναμικά (dynamic) χρησιμοποιούν και περασμένες τιμές εισόδων.

**Δίκτυα εκπαιδευόμενα με/χωρίς χρήση εξωτερικού “δασκάλου” :** Οι τεχνικές μάθησης των δικτύων χωρίζονται σε τεχνικές με δάσκαλο (supervised), όπου παρέχεται επιπρόσθετη πληροφορία για την επιθυμητή απόκριση του δικτύου, και σε τεχνικές χωρίς δάσκαλο (unsupervised).

**Σταθερά - Προσαρμοζόμενα:** Στα σταθερά δίκτυα (fixed) τα βάρη των συνδέσεων επιλέγονται με κάποιο τρόπο και παραμένουν σταθερά καθ’ όλη την διάρκεια της λειτουργίας τους. Τα προσαρμοζόμενα δίκτυα (adaptive) μεταβάλλουν τα βάρη τους

σύμφωνα με νεότερες πληροφορίες που λαμβάνουν.

**Χωρίς ανάδραση - Με ανάδραση:** Στα δίκτυα με ανάδραση (recurrent) ο κάθε νευρώνας μπορεί να συνδέεται με οποιονδήποτε άλλο στο δίκτυο και να ανατροφοδοτεί τον εαυτό του. Τα δίκτυα χωρίς ανάδραση (feedforward) χαρακτηρίζονται από στρώματα νευρώνων, με την πληροφορία να διαδίδεται κατά συγκεκριμένη φορά μέσα στο δίκτυο.

Οι πιο συνηθισμένες εφαρμογές των νευρωνικών δικτύων είναι σε αναγνώριση και κατάταξη προτύπων (φωνής-εικόνας), στον αυτόματο έλεγχο και σε προβλήματα βελτιστοποίησης. Επίσης έχουν ευρέως χρησιμοποιηθεί σαν συσχειστική μνήμη (associative memory), η οποία, π.χ. αναγνωρίζει παραμορφωμένες εικόνες αλφαριθμητικών χαρακτήρων έχοντας με κάποιο τρόπο μάθει να αναγνωρίζει τις κανονικές τους παραστάσεις.

Δύο ευρέως διαδεδομένα μοντέλα είναι του Hopfield και το πολυεπίπεδο perceptron. Το δίκτυο του Hopfield αποτελείται από ένα πλήρως συνδεδεμένο στρώμα νευρώνων (δίκτυο με ανάδραση) που λειτουργούν σε συνεχή ή διακριτό χρόνο. Το δίκτυο λειτουργεί επαναληπτικά δηλ. ξεκινά από μία αρχική κατάσταση και ο κάθε νευρώνας ανανεώνει την έξοδό του σύμφωνα με τις μεταβολές των εξόδων των νευρώνων που συνδέονται με αυτόν. Τα δίκτυα Hopfield περιγράφονται από μία συνάρτηση ενέργειας. Με οποιαδήποτε αρχική κατάσταση και κατάλληλους περιορισμούς στα βάρη (συμμετρία) το δίκτυο αυτό μπορεί να συγκλίνει σε κάποιο τοπικό ελάχιστο της συνάρτησης ενέργειας, ιδιότητα σημαντική για πολλές εφαρμογές είτε αναγνώρισης προτύπων, είτε βελτιστοποίησης συναρτήσεων (π.χ. πρόβλημα περιοδεύοντος πωλητή). Το άλλο σημαντικό μοντέλο, το οποίο θα χρησιμοποιηθεί ευρέως στην εργασία αυτή, είναι το πολυεπίπεδο perceptron, το οποίο θα προσεγγίσουμε αναλυτικότερα στην επόμενη ενότητα.


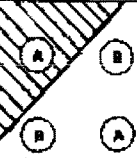



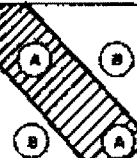

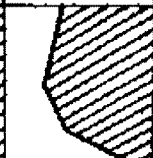



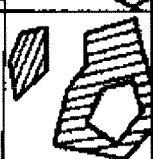
## 2.3 Πολυεπίπεδα perceptrons

Η κατηγορία δικτύων που θα εξετάσουμε στην ενότητα αυτή περιλαμβάνει δίκτυα χωρίς ανάδραση (feedforward networks), με στρώματα νευρώνων (multilayered networks). Η ιστορία των δικτύων αυτών ξεκίνησε την δεκαετία του 50, με την εισαγωγή από τον

Rosenblatt του perceptron [7]. Το perceptron λειτουργεί ακριβώς όπως ο νευρώνας του σχήματος 2.1, με το αρχικό μοντέλο του Rosenblatt να χρησιμοποιεί την hard-limiting μη γραμμικότητα. Υπάρχουν δύο τρόποι θεώρησης της λειτουργίας ενός perceptron. Ο ένας είναι σαν μία συνάρτηση διαχωρισμού προτύπων δύο διαφορετικών κλάσεων σε προβλήματα αναγνώρισης προτύπων, και ο άλλος σαν μία διαδική λογική μονάδα, ικανή να υλοποιήσει ένα πλήθος λογικών συναρτήσεων όπως τις AND, OR και COMPLEMENT. Σαν ταξινομητής, το perceptron δύναται να διαχωρίσει τον χώρο των προτύπων σε δύο περιοχές με ένα γραμμικό όριο απόφασης (π.χ. ευθεία για ένα διδιάστατο χώρο προτύπων, υπερεπίπεδο στον  $n$ -διάστατο χώρο). Αρα είναι κατάλληλο μόνο για επίλυση προβλημάτων που είναι γραμμικά διαχωρίσιμα. Επιπλέον, σαν λογική μονάδα δεν μπορεί να υλοποιήσει τη συνάρτηση EXCLUSIVE OR. Στο πάνω μέρος του σχήματος 2.2, δείχνουμε σχηματικά τις δυνατότητες διαχωρισμού περιοχών για ένα μοναδικό perceptron. Οι περιοχές ονομάζονται A και B, και γραμμοσκιασμένη είναι πάντα η περιοχή A. Όπως φαίνεται και στο σχήμα, οι δυνατότητες του perceptron είναι περιορισμένες σε γραμμικά προβλήματα. Η λύση πιο πολύπλοκων προβλημάτων θα απαιτούσε την διασύνδεση πολλών στοιχειωδών perceptron. Τα δεδομένα αυτά, όπως και η μη ύπαρξη κατάλληλου αλγόριθμου μάθησης εκείνη την εποχή, παρατηρήθηκαν από τους Minsky και Papert [6], με αποτέλεσμα την ανακοπή της ερευνητικής προσπάθειας στην κατεύθυνση αυτή.

Μόλις την δεκαετία του '80, αναπτύχθηκε αλγόριθμος ικανός να μεταβάλλει τα βάρη σε ένα δίκτυο με πολλά επίπεδα από perceptrons [8]. Η hard-limiting μη γραμμικότητα αντικαταστάθηκε με την σιγμοειδή (βλέπε σχήμα 2.1), η οποία είναι παραγωγίσιμη. Στο σχήμα 2.3 παρουσιάζουμε ένα πολυεπίπεδο perceptron δίκτυο.

Το εικονιζόμενο δίκτυο ονομάζεται δίκτυο τριών επιπέδων. Τα σήματα ή πρότυπα εισόδου παρουσιάζονται στο πρώτο επίπεδο του δικτύου (επίπεδο εισόδου), το οποίο δεν προσμετράται στον συνολικό αριθμό των επιπέδων. Τα σήματα διαδίδονται κατά μία μόνο φορά μέσω των ενδιάμεσων επιπέδων (εδώ δύο κρυφά επίπεδα), ενώ οι έξοδοί του δημιουργούνται στο τελευταίο στρώμα στην κορυφή (επίπεδο εξόδου). Τα ενδιάμεσα ή κρυφά επίπεδα (hidden layers) παίζουν σημαντικό ρόλο στη λειτουργία του δικτύου, διότι επιτρέπουν σε αυτό να σχηματίζει εσωτερικές παραστάσεις και αντιστοιχίες των εισόδων

STRUCTURE	TYPES OF DECISION REGIONS	EXCLUSIVE OR PROBLEM	CLASSES WITH MESHED REGIONS	MOST GENERAL REGION SHAPES
	HALF PLANE BOUNDED BY HYPERPLANE			
	CONVEX OPEN OR CLOSED REGIONS			
	ARBITRARY (Complexity Limited By Number of Nodes)			

Σχήμα 2.2: Δυνατότητες ταξινόμησης των πολυεπίπεδων νευρωνικών δικτύων.

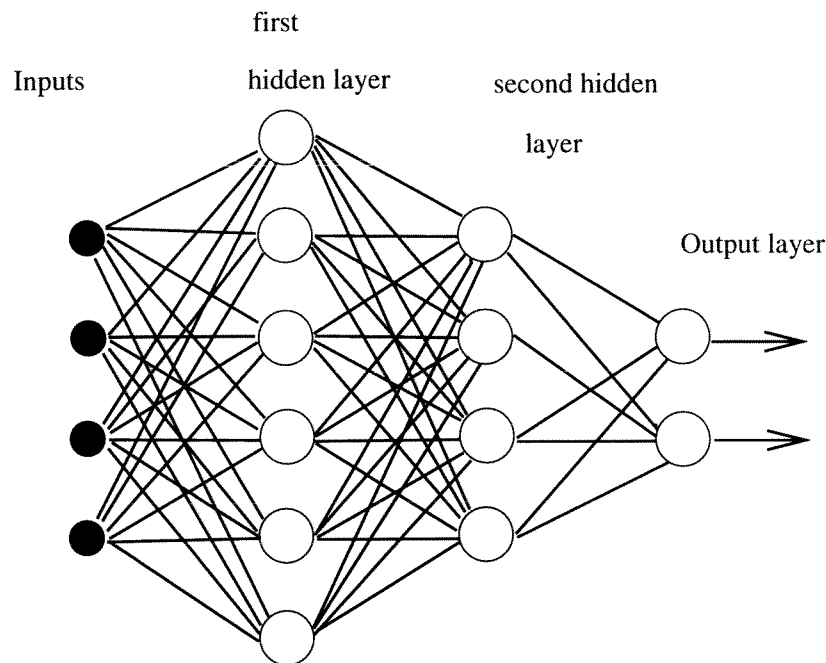
και εξόδων. Οι νευρώνες των στρωμάτων αυτών δρουν σαν ανιχνευτές χαρακτηριστικών (feature detectors), οι οποίοι συλλαμβάνουν τα σημαντικά χαρακτηριστικά της πληροφορίας της εισόδου. Ο αλγόριθμος μάθησης μεταβάλλει τα βάρη ώστε να μειωθεί το μέσο τετραγωνικό σφάλμα μεταξύ των πραγματικών και των επιθυμητών εξόδων του δικτύου για όλα τα πρότυπα εισόδου,

$$E = \frac{1}{2} \sum_{r=1}^K \sum_{m=1}^M [d(m) - y(m)]_r^2 \quad (2.1)$$

όπου το πλήθος των προτύπων, το πλήθος των εξόδων,  $y(m)$  οι πραγματικές έξοδοι και  $d(m)$  οι επιθυμητές έξοδοι. Οι επιθυμητές έξοδοι πρέπει να δίδονται, άρα ο αλγόριθμος χρειάζεται εξωτερική πληροφορία (supervised). Ο κλασικός αλγόριθμος μάθησης αυτών των δικτύων ονομάζεται backpropagation [2, 8], ο οποίος ελαχιστοποιεί την παραπάνω ποσότητα και είναι τύπου βαθύτατης καθόδου της παραγώγου του σφάλματος (steepest/gradient descent). Ο αλγόριθμος ξεκινά με μικρές τυχαίες τιμές για τα βάρη, υπολογίζει το σφάλμα για όλα τα πρότυπα εφόσον είναι δεδομένες οι επιθυμητές τιμές, και αρχικά μεταβάλλει τα βάρη των νευρώνων εξόδου ως εξής:

$$\Delta w_{ij} = -\mu \frac{\partial E}{\partial w_{ij}} \quad (2.2)$$





Σχήμα 2.3: Πολυεπίπεδο perceptron.

Για να μεταβληθούν κατόπιν τα βάρη των πιο κάτω κρυφών επιπέδων χρησιμοποιείται ο κανόνας αλυσίδας για τον υπολογισμό της μερικής παραγώγου του "σφάλματος" στις εξόδους των αντίστοιχων νευρώνων των κρυφών επιπέδων. Συνεπώς υπάρχει μία προς τα πίσω διάδοση του σφάλματος εξόδου, από όπου και το όνομα backpropagation.

## 2.4 Διανυσματικός Κβαντιστής με εκμάθηση

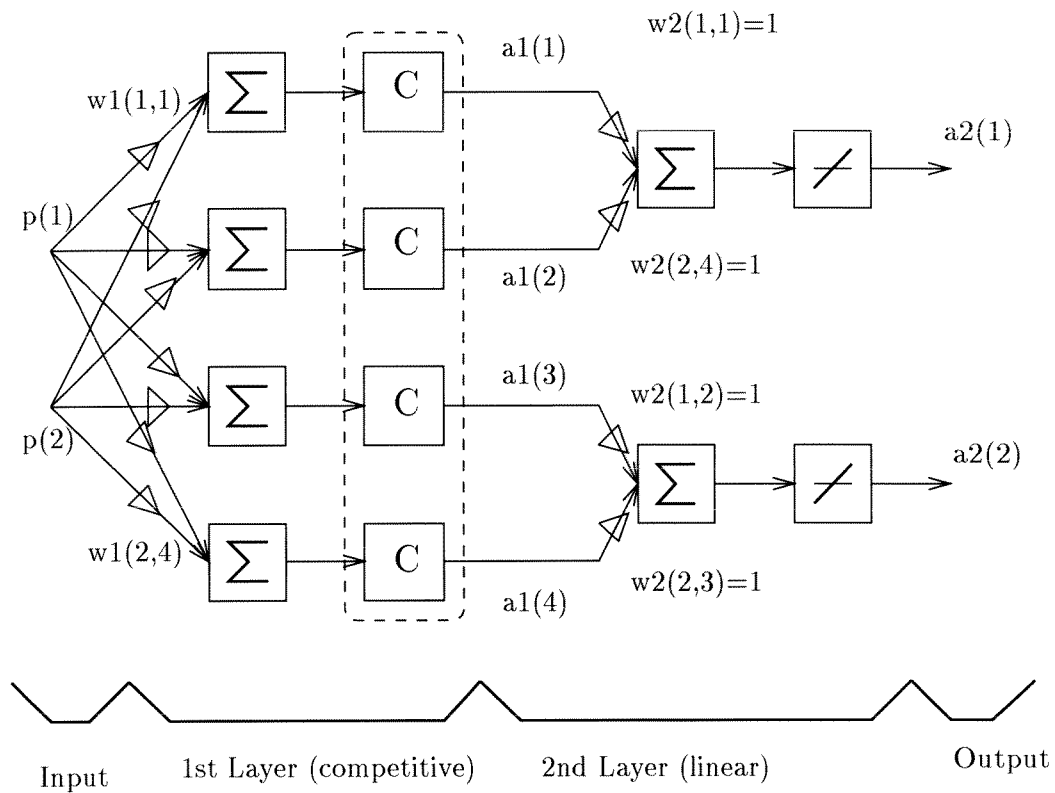
Ένας νεότερος τύπος νευρωνικού δικτύου βασίζεται στον αλγόριθμο Διανυσματικού Κβαντισμού με Εκμάθηση (Learning Vector Quantizer, LVQ) που προτάθηκε από τον Kohonen [3]. Πρόκειται για ένα τεχνητό νευρωνικό δίκτυο δύο επιπέδων, όπως φαίνεται στο σχήμα 2.4. Ως παράδειγμα, στο σχήμα 2.4 φαίνεται ένας κατηγοριοποιητής LVQ που δέχεται διανύσματα εισόδου  $\mathbf{p}$  δύο συνιστωσών και τα κατηγοριοποιεί σε δύο κατηγορίες. Το διάνυσμα εξόδου  $\mathbf{a}_2$  έχει, συνεπώς, δύο συνιστώσες. Όταν κατηγοριοποιείται δεδομένο διάνυσμα εισόδου μία μόνο από τις συνιστώσες αυτές γίνεται ίση με 1 ενώ η άλλη μένει στο 0, δηλώνοντας έτσι την κατηγορία της εισόδου. Οι εισοδοί πολλαπλασιάζονται με

τα βάρη  $w_1$  και τροφοδοτούν το πρώτο επίπεδο, που είναι και το ανταγωνιστικό (*competitive*) επίπεδο του δικτύου. Στο εικονιζόμενο παράδειγμα αυτό περιλαμβάνει τέσσερις νευρώνες. Το διάνυσμα εξόδου του πρώτου αυτού επιπέδου,  $a_1$ , τροφοδοτεί το επόμενο επίπεδο που είναι το γραμμικό (*linear*) επίπεδο. Στο πρώτο επίπεδο, τα διανύσματα εισόδου χωρίζονται σε υποκατηγορίες ενώ η λειτουργία του δεύτερου επιπέδου είναι να ομαδοποιήσει τις υποκατηγορίες σε τελικές κατηγορίες. Γι' αυτό τα βάρη  $w_2$  παίρνουν τις τιμές 0 ή 1 υλοποιώντας την ομαδοποίηση αυτή. Στο εικονιζόμενο παράδειγμα, οι δύο πρώτοι νευρώνες του πρώτου επιπέδου (και οι υποκατηγορίες που αυτοί ορίζουν) αποτελούν την πρώτη τελική κατηγορία, γι' αυτό και τα αντίστοιχα βάρη  $w_2(1, 1)$ ,  $w_2(1, 2)$  είναι ίσα με 1. Αντίστοιχα, οι δύο επόμενοι νευρώνες του πρώτου επιπέδου ανήκουν στην δεύτερη τελική κατηγορία, οπότε και τα  $w_2(2, 3)$ ,  $w_2(2, 4)$  είναι ίσα με 1. Τα υπόλοιπα βάρη,  $w_1(1, 3)$ ,  $w_1(1, 4)$  και  $w_2(2, 1)$ ,  $w_2(2, 2)$  είναι ίσα με μηδέν.

Το δίκτυο αυτό πρέπει κατ' αρχήν να εκπαιδευτεί από ένα σύνολο διανυσμάτων εκμάθησης και στη συνέχεια να λειτουργήσει ως κατηγοριοποιητής, παράγοντας μία απόφαση  $a_2$  για κάθε άγνωστο διάνυσμα χαρακτηριστικών με το οποίο τροφοδοτούμε την είσοδό του.

Η κύρια ιδιότητα που οδήγησε στην επιλογή του συγκεκριμένου κατηγοριοποιητή LVQ είναι το γεγονός ότι σχεδιάστηκε με τη φιλοσοφία περισσότερο του επιτυχημένου κατηγοριοποιητή, και λιγότερο του επιτυχημένου διανυσματικού κβαντιστή. Πιο συγκεκριμένα, τα “κέντρα” ή αλλιώς τα “αντιπροσωπευτικά διανύσματα” (*reference vectors*) της κάθε κατηγορίας δεν τοποθετούνται με κριτήριο να αντιπροσωπεύουν με τον καλλίτερο τρόπο την πλειοψηφία των μελών της κατηγορίας τους, όπως είναι το επιθυμητό για έναν καλό διανυσματικό κβαντιστή ώστε να ελαχιστοποιεί το σφάλμα κβαντισμού (*distortion*). Αντίθετα, τοποθετούνται ώστε να δημιουργούν διαχωριστικές επιφάνειες όσο το δυνατόν πλησιέστερες στο άριστο υπό την Bayesian άποψη, οι οποίες δηλαδή ελαχιστοποιούν το σφάλμα κατηγοριοποίησης.

Το επιθυμητό αυτό χαρακτηριστικό επιτυγχάνεται κατά την εκπαίδευση του δικτύου δηλαδή από πλευράς κβαντιστού, με το σχεδιασμό του codebook. Κατά τη φάση αυτή τοποθετούνται μέσα στον (πολυδιάστατο εδώ) χώρο προτύπων τα πρότυπα εκείνα που είναι



Σχήμα 2.4: Σχηματικό διάγραμμα ενός κατηγοριοποιητή LVQ 2 κατηγοριών

οι αντιπρόσωποι των κατηγοριών. Με βάση τις αποστάσεις από τους αντιπροσώπους θα γίνεται στο εξής η κατηγοριοποίηση. Η εκπαίδευση περιλαμβάνει δύο στάδια. Στο πρώτο στάδιο, που λειτουργεί ως αρχικοποίηση του δεύτερου, ένας αλγόριθμος *ομαδοποίησης* (*clustering*) χωρίς επίβλεψη (*unsupervised*) χρησιμοποιείται επί των διανυσμάτων του συνόλου εκπαίδευσης για να τοποθετηθούν στο χώρο προτύπων τα κέντρα των κατηγοριών. Στο σημείο αυτό μπορούν να χρησιμοποιηθούν αλγόριθμοι όπως ο K-Means ή ο LBG. Έστω  $\mathbf{r}_n^{(0)}$ ,  $n = 1, 2, \dots, N$  οι θέσεις των  $N$  προτύπων μετά το στάδιο αυτό. Το επόμενο στάδιο της φάσης εκπαίδευσης επαναληπτικά διορθώνει τις αρχικές θέσεις, έτσι ώστε το σφάλμα κατηγοριοποίησης να φθίνει, αποτέλεσμα που σημαίνει ότι οι προκύπτουσες διαχωριστικές επιφάνειες πλησιάζουν τις, υπό Bayesian έννοια άριστες.

Για το σκοπό αυτό τα διανύσματα χαρακτηριστικών του συνόλου εκπαίδευσης, έστω  $\mathbf{t}_i$ ,  $i = 1, 2, \dots, T$ , χρησιμοποιούνται όλα κατ' επανάληψιν. Έστω  $j = 0, 1, \dots, J$  ο

μετρητής αυτών των επαναλήψεων καθεμία από τις οποίες περιλαμβάνει  $T$  ακριβώς βήματα. Κατά την  $j$ -οστή, έστω, επανάληψη, όλα τα  $\mathbf{t}_i$ ,  $i = 1, 2, \dots, T$  κατηγοριοποιούνται συγκρινόμενα με τις τρέχουσες θέσεις των αντιπροσώπων των κατηγοριών, έστω  $\mathbf{r}_n^{(i-1,j)}$ ,  $n = 1, 2, \dots, N$ , χρησιμοποιώντας τον κατηγοριοποιητή 1 - Nearest Neighbor (1-NN).

Λανθασμένη κατηγοριοποίηση ενός  $\mathbf{t}_i$  προκαλεί διόρθωση (ανανέωση) των θέσεων των δύο αντιπροσώπων που εμπλέκονται, δηλαδή του πλησιέστερου (λανθασμένου) γείτονα, έστω  $\mathbf{r}_w^{(i-1,j)}$ , και του σωστού (αλλά όχι πλησιέστερου)  $\mathbf{r}_c^{(i-1,j)}$ . Η διόρθωση γίνεται ανάλογα με την απόσταση, σύμφωνα με τη σχέση

$$\begin{aligned} \mathbf{r}_c^{(i,j)} &= \mathbf{r}_c^{(i-1,j)} + \alpha(i) \{ \mathbf{t}_i - \mathbf{r}_c^{(i-1,j)} \}, \\ \mathbf{r}_w^{(i,j)} &= \mathbf{r}_w^{(i-1,j)} - \alpha(i) \{ \mathbf{t}_i - \mathbf{r}_w^{(i-1,j)} \}. \end{aligned} \quad (2.3)$$

Είναι φανερό ότι κάθε τέτοια επανάληψη διαρκεί  $T$  χρονικές στιγμές, αν ως χρονική στιγμή θεωρηθεί η εξέταση ενός διανύσματος  $\mathbf{t}_i$ . Η σύγκλιση των θέσεων των κέντρων με τον αριθμό επαναλήψεων επιτυγχάνεται, αν το βήμα μετακίνησης  $\alpha(i)$  είναι γνησίως φθίνουσα συνάρτηση του μετρητή διανύσματος ή “χρόνου”,  $i$ , [1].

Στην αρχή κάθε επανάληψης οι θέσεις των εκπροσώπων αρχικοποιούνται είτε από την ομαδοποίηση που έχει προηγηθεί, προκειμένου για την πρώτη επανάληψη  $j = 1$ , είτε από τις θέσεις στις οποίες κατέληξε η προηγούμενη ( $j - 1$ ) επανάληψη, δηλαδή:

$$\begin{aligned} \mathbf{r}_n^{(0,j)} &= \mathbf{r}_n^{(T,j-1)}, \quad \forall n, \quad j = 2, \dots, J, \\ \mathbf{r}_n^{(0,1)} &= \mathbf{r}_n^{(0)}, \quad \forall n. \end{aligned} \quad (2.4)$$

Επίσης, μετά από κάθε κατηγοριοποίηση ενός  $\mathbf{t}_i$ , δύο συνθήκες πρέπει να πληρούνται επιπλέον της λανθασμένης απόφασης, για να ξεκινήσει η διόρθωση θέσης που περιγράφει η σχέση (2.3):

- (i) το υπό εξέταση διάνυσμα  $\mathbf{t}_i$  πρέπει να βρίσκεται εντός ενός πλαισίου πεπερασμένου εύρους, το οποίο κείται περί το νοητό υπερεπίπεδο που ισαπέχει από τους δύο εμπλεκόμενους αντιπροσώπους,  $\mathbf{r}_c^{(i-1,j)}$  και  $\mathbf{r}_w^{(i-1,j)}$ , και

- (ii) δεν πρέπει να υπάρχει άλλος λανθασμένος αντιπρόσωπος, έστω  $\mathbf{r}_w^{(i-1,j)}$ , που να κείται πλησιέστερα στο  $\mathbf{t}_i$  από ό,τι ο σωστός αντιπρόσωπος  $\mathbf{r}_c^{(i-1,j)}$ , δηλαδή ο σωστός αντιπρόσωπος πρέπει να είναι ο δεύτερος πλησιέστερος.

Οι κανόνες που περιγράφηκαν, μαζί με τις δύο αυτές επιπλέον συνθήκες, μας δίνουν εκείνη την εκδοχή του LVQ που ονομάζεται LVQ2, [4].

Με την ολοκλήρωση της φάσης εκμάθησης, έχουμε διαθέσιμο το τελευταίο σύνολο αντιπροσώπων,  $\mathbf{r}_n^{(T,J)}$ ,  $n = 1, 2, \dots, N$ . Αυτά τα σταθερά πλέον σημεία του χώρου προτύπων χρησιμοποιούνται κατά τη φάση λειτουργίας (*operation mode*) του δικτύου, για την κατηγοριοποίηση νεοεισερχόμενων διανυσμάτων. Συγκεκριμένα, κάθε νεοεισερχόμενο διάνυσμα κατηγοριοποιείται με τον κανόνα 1-NN, και ανατίθεται στην κατηγορία, έστω  $n$ , του πλησιέστερου αντιπροσώπου  $\mathbf{r}_n^{(T,J)}$ . Η έξοδος, συνεπώς, του κατηγοριοποιητή είναι η ακολουθία αποφάσεων,  $d_i \in \{1, 2, \dots, N\}$ , για κάθε μέλος της ακολουθίας ελέγχου (*test set*) από διανύσματα χαρακτηριστικών.

## Αναφορές

- [1] J.S.Baras, A.LaVigna, “Convergence of Kohonen’s LVQ,” *Proc. IJCNN’90*, vol. 3, pp. 17-20, 1990.
- [2] D. Hush and B. Horne, *Progress in supervised neural networks*. IEEE Signal Processing Magazine, vol. 10.1, pp. 8-39, January 1993.
- [3] T.Kohonen, “Self-Organization and Associative Memory,” 2nd Ed.. *Springer-Verlag*, New York, 1988.
- [4] T.Kohonen, “Improved Versions of LVQ,” *Proc. IJCNN’90*, vol. 1, pp. 545-550, June 1990.
- [5] R. P. Lippmann, *An introduction to computing with neural nets*. IEEE ASSP Magazine, vol. 4.2, pp. 4-22, April 1987.
- [6] M. Minsky and S. Papert, *Perceptrons*. MIT Press, Cambridge, MA. 1969.
- [7] F.Rosenblatt, “The perceptron: A probabilistic model for information storage and organization in the brain,” *Psychological Review*, vol. 65, pp. 386-408, 1958.
- [8] D.Rumelhart, G.Hinton, R.Williams. “Learning internal representations by error propagation,” In D.Rumelhart and J.McClelland, eds, *Parallel distributed processing: explorations in the micro structure of cognition*, pp. 318-362. MIT Press, 1986.

# Τεχνική Fractals στην ανάλυση και αναγνώριση εικόνων

### 3.1 Εισαγωγή

Ο κόσμος που μας περιβάλλει αποτελείται, εκτός από τα ανθρώπινα κατασκευάσματα, από περίπλοκες ακανόνιστες και τραχιές επιφάνειες. Εάν θέλουμε να κατασκευάσουμε μηχανές που να λειτουργούν μέσα στο φυσικό περιβάλλον χρειαζόμαστε ένα πλαίσιο - μοντέλο περιγραφής που θα είναι ικανό να περιγράψει τις μορφές που παρουσιάζονται στο φυσικό περιβάλλον. Είναι δυνατόν να περιγραφεί από ένα μοντέλο το σχήμα μίας τσαλακωμένης εφημερίδας, το κλαδί ενός δέντρου ή η κορυφογραμμή ενός βουνού;

Οι υπάρχουσες μέθοδοι αναπαράστασης βασίζονται περισσότερο στην ιδέα του Πλάτωνα για ιδανικά σχήματα, π.χ. σφαίρες, κυλίνδρους και κύβους για την περιγραφή τρισδιάστατων αντικειμένων. Τέτοια πρότυπα δουλεύουν καλά με τεχνητά αντικείμενα ή σε τεχνητά περιβάλλοντα. Εάν όμως προσπαθήσουμε να περιγράψουμε την μορφή ενός φυσικού αντικειμένου τότε το έργο γίνεται απελπιστικά πολύπλοκο. Ευριστικές μέθοδοι, όπως η παρεμβολή επιφάνειας (surface interpolation) [7] και μορφοποίηση από σκίαση (shape from shading) [6], εκμεταλλεύονται την έννοια της "ομαλότητας" για να συσχετίσουν γειτονικά σημεία μίας επιφάνειας. Οι περιγραφείς εικόνων υφής (texture descriptors) εφαρμόζονται μόνο σε εικόνες επιφάνειας και αγνοούν τις περιπτώσεις τρισδιάστατων εικόνων. Οι συναρτήσεις fractal είναι υποψήφιες ως μοντέλο περιγραφής τρισδιάστατων

εικόνων και επιφανειών γιατί πολλές φυσικές διαδικασίες δημιουργούν γραμμές, επιφάνειες ή και τρισδιάστατες παραστάσεις που είναι fractal. Τα fractals έχουν χρησιμοποιηθεί για να περιγράψουν πλήθος φυσικών αντικειμένων, σύννεφα, δένδρα, ροή υγρών κ.ο.κ. Κανένα από αυτά δεν είναι ιδανικά fractals, παρουσιάζουν όμως τις ιδιότητες αυτών σε μία κλίμακα αναλύσεων [12]. Ο Pentland [15] έδειξε ότι συναρτήσεις fractals αποτελούν ένα εργαλείο περιγραφής τρισδιάστατων φυσικών σχημάτων, π.χ λιβάδια, δάση, το οποίο παρουσιάζει πλεονεκτήματα σε σχέση με την μοντελοποίηση με πρότυπα αντικείμενα, π.χ. σφαίρες, κυλίνδρους, κύβους κ.ο.κ.

Η λέξη fractal, πρωτοαναφέρθηκε στην βιβλιογραφία από τον Mandelbrot, και προέρχεται από την λέξη *fractus* που σημαίνει σπασμένο. Με τον συγκεκριμένο όρο ο Mandelbrot ήθελε να αναφερθεί σε αντικείμενα τα οποία παρουσίαζαν τοπολογικές ανωμαλίες. Ο Mandelbrot [12] ορίζει ως fractals τα σύνολα των οποίων η διάσταση Hausdorff είναι μεγαλύτερη από την αντίστοιχη Ευκλείδεια; δηλαδή σύνολα των οποίων οι μετρικές ιδιότητες απαιτούν τον ορισμό μίας “διάστασης” η οποία είναι μεγαλύτερη από την γνώριμη ευκλείδεια διάσταση που χρησιμοποιούμε για να περιγράψουμε την τοπολογική διάσταση. Στα fractals η νέα διάσταση παίρνει κλασματικές τιμές από όπου και το όνομα τους [12, 13]. Κατά τον Barnsley [1] τα fractals είναι γεωμετρικά σχήματα των οποίων τα χαρακτηριστικά εμφανίζονται αναλλοίωτα σε μια κλίμακα αναλύσεων και ως εκ τούτου μπορούν να περιγραφούν από γραμμικούς μετασχηματισμούς. Ο Falconer [3] ορίζει τα fractals ως σύνολα τα οποία έχουν τις παρακάτω ιδιότητες:

1. Έχουν καθορισμένη δομή σε οποιαδήποτε ανάλυση παρατήρησης.
2. Τα σύνολα αυτά είναι τόσο ακανόνιστα ώστε να μην μπορούν να χαρακτηρισθούν με γεωμετρικές σχέσεις τόσο τοπικά όσο και συνολικά.
3. Τα σύνολα παρουσιάζουν ιδιότητες ομοιότητας ως προς τον εαυτό τους (self-similarity).
4. Η διάσταση fractal είναι συνήθως μεγαλύτερη από την αντίστοιχη ευκλείδεια.
5. Είναι δυνατός ένας αναδρομικός ορισμός για την κατασκευή των συγκεκριμένων συνόλων.



Παραδείγματα συνόλων fractal είναι η γραμμή Van Koch και το τρίγωνο του Sierpinski. Η γραμμή του Van Koch κατασκευάζεται εάν σε κάθε ευθύγραμμο τμήμα αντικαθίσταται το δεύτερο τρίτο με τις δύο πλευρές ενός ισόπλευρου τριγώνου. Το τρίγωνο του Sierpinski κατασκευάζεται εάν κάθε ισόπλευρο τρίγωνο αντικαθίσταται από τρία του μισού ύψους.

Η κλασματική διάσταση σχετίζεται άμεσα με την μορφολογία των γραμμών ή των επιφανειών που περιγράφονται από συναρτήσεις fractal. Όσο πιο τραχιές οι επιφάνειες τόσο μεγαλύτερη η κλασματική διάσταση τους. Τέτοιες επιφάνειες είναι γενικεύσεις τυχαίας κίνησης Brown, δηλαδή “τυχαίου περιπάτου”, μιας διαδικασίας που παρουσιάζει δύο σημαντικές ιδιότητες: α) κάθε τμήμα επιφάνειας είναι “στατιστικά όμοιο” με όλα τα άλλα και β) τμήματα σε διαφορετικές αναλύσεις είναι στατιστικά δυσδιάκριτα μεταξύ τους. Με άλλα λόγια οι ιδιότητες μιας επιφάνειας, π.χ εμβαδόν, παραμένουν οι ίδιες κάτω από μία κλίμακα αναλύσεων. Μια τυχαία κίνηση  $I(x)$  είναι κλασματική κίνηση Brown εάν για όλα τα  $x$  και  $\Delta x$  ισχύει:

$$Pr \left( \frac{I(x + \delta x) - I(x)}{\|\delta x\|^{(T-D-1)}} < y \right) = F(y) \quad (3.1)$$

όπου  $F(y)$  είναι μία αθροιστική κατανομή πιθανότητας,  $T$  είναι η τοπολογική διάσταση της κίνησης  $I$  και  $D$  η κλασματική διάσταση της γραμμής.

Η κλασματική διάσταση  $D$  μπορεί να υπολογισθεί από τα στατιστικά δεύτερης τάξης της  $I(x)$  με χρήση της σχέσης (3.1) ή από φάσμα Fourier  $P(f)$  της  $I(x)$ . Η σχέση (3.1) μπορεί να γραφεί ως [15]:

$$E(|\Delta I_{\delta x}|) \|\delta x\|^{(T-D-1)} \approx E(|\Delta I_{\delta x=1}|) \quad (3.2)$$

Η ποσότητα  $E(|\Delta I_{\delta x}|)$  είναι η αναμενόμενη τιμή της αλλαγής της φωτεινότητας όταν μεταβάλλεται η απόσταση κατά  $\Delta x$ . Κατά συνέπεια υπολογίζουμε για διάφορα  $\Delta x$  την ποσότητα  $E(|\Delta I_{\delta x}|)$  και λύνουμε με ελάχιστα τετράγωνα την λογαριθμική μορφή της σχέσης (3.2). Εναλλακτικά μπορούμε να χρησιμοποιήσουμε μια έκφραση για το φάσμα Fourier της  $I(x)$ , δηλ.  $P(f) = f^{2*(D-T)-3}$  [12].

Στην περίπτωση διδιάστατων αντικειμένων μία ποσότητα που σχετίζεται άμεσα με

την κλασματική διάσταση  $D$  είναι η επιφάνεια των αντικειμένων, ενώ στη μονοδιάστατη περίπτωση το μήκος μιας γραμμής σχετίζεται με την κλασματική διάσταση  $D$ . Διάφορες μέθοδοι για την μέτρηση του μήκους μιας γραμμής έπονται από την σχέση (3.2) και έχουν προταθεί από τον Mandelbrot στο πρόβλημα μέτρησης του μήκους των ακτογραμμών. Το συγκεκριμένο πρόβλημα σχετίζεται άμεσα με τα fractals αφού οι ακτογραμμές μιας περιοχής μπορούν να θεωρηθούν σχήματα fractals εάν παρατηρηθούν από διάφορα ύψη. Οι διάφορες μέθοδοι δίνονται παρακάτω:

1. Η συγκεκριμένη μέθοδος στηρίζεται στην προσέγγιση μίας γραμμής με μία πολυγωνική, της οποίας κάθε τμήμα έχει σταθερό μήκος  $\epsilon$ . Δεδομένου ενός πρότυπου ευθύγραμμου τμήματος μήκους  $\epsilon$  διανύουμε την υπό μέτρηση γραμμή. Ο αριθμός των επαναλήψεων πολλαπλασιασμένος με  $\epsilon$  αποτελεί μια πρώτη προσέγγιση του μήκους  $L(\epsilon)$ . Σημειώνεται ότι, όσο μειώνεται το μήκος του κανόνα τόσο μεγαλώνει η ποσότητα  $L(\epsilon)$ .
2. Ας θεωρηθούν όλα τα σημεία τα οποία απέχουν όχι περισσότερο από  $\epsilon$  από την προς μέτρηση γραμμή. Τα σύνολα των σημείων σχηματίζουν μία ταινία πάχους  $2\epsilon$ . Εάν μετρηθεί η επιφάνεια της ταινίας και διαιρεθεί με  $\epsilon$  προκύπτει το μήκος της γραμμής  $L(\epsilon)$ . Σημειώνεται πάλι ότι η ποσότητα  $L(\epsilon)$  είναι αντιστρόφως ανάλογη του  $\epsilon$ .
3. Καλύπτεται η ακτογραμμή με τον ελάχιστο αριθμό δίσκων ακτίνας  $\epsilon$ , οι οποίοι δεν είναι κατ' ανάγκη τοποθετημένοι πάνω στην γραμμή. Τότε  $L(\epsilon)$  είναι η συνολική επιφάνεια των δίσκων διαιρεμένη με  $\epsilon$ .

Η τρίτη μέθοδος αποτελεί την βασική ιδέα για την μέθοδο Μέτρησης με "Κύβους" (Box - Counting) όπως αναφέρεται στο [1], η οποία είναι γενέτειρα πολλών μεθόδων που εμφανίστηκαν στην συνέχεια. Χρησιμοποιώντας τις παραπάνω μεθόδους το μήκος μίας ακτογραμμής δίνεται από την σχέση [12]:

$$L(\epsilon) \approx F\epsilon^{1-D} \rightarrow L(\epsilon) \approx \epsilon^{-D}L(1) \quad (3.3)$$

Όπου  $F$  και  $D$  και σταθερές.

Στην συνέχεια δίνεται η δομή του παρόντος κεφαλαίου. Στην ενότητα 3.2 παρουσιάζεται η μέθοδος Μέτρησης με “Κύβους” για την εκτίμηση της κλασματικής διάστασης ενώ στην ενότητα 3.3 παρουσιάζονται μέθοδοι υπολογισμού της κλασματικής διάστασης σε διδιάστατα αντικείμενα και επεκτάσεις της κλασματικής διάστασης για την ανάλυση και κατάτμηση εικόνων. Η μέθοδος της επιστρωμάτωσης [14] στην ενότητα 3.3.1 μαζί με την μέθοδο του κλασματικού πίνακα [10] ως επέκταση της κλασματικής διάστασης αναλύεται στην ενότητα 3.4 ως βάση για τον ορισμό του κλασματικού πίνακα στην πολυκαναλική περίπτωση. Στην ενότητα 3.5 παρουσιάζεται το πολυκαναλικό μοντέλο και δίδονται αποτελέσματα ανάλυσης πραγματικών εικόνων για διάφορα μεγέθη block ανάλυσης της αρχικής εικόνας. Τέλος παρουσιάζεται μία περίπτωση κατάτμησης συνθετικής εικόνας και γίνεται σύγκριση με μια πολυμοντελική προσέγγιση στην κατηγοριοποίηση εικόνων με σύνολο πραγματικών εικόνων [16].

## 3.2 Εύρεση της κλασματικής διάστασης με την μέθοδο Μέτρησης με “Κύβους”

Έστω ο αριθμός  $(, \epsilon)$  είναι ο ελάχιστος αριθμός δίσκων σφαιρών ακτίνας  $\epsilon$  για την κάλυψη ενός συνόλου  $A$ , όπου το σύνολο  $A$  μπορεί να είναι μια γραμμή ή μια επιφάνεια. Τότε από την σχέση (3.3) έχουμε:

$$D = \frac{\text{Ln}(N(A, \epsilon)) - \text{Ln}F}{\text{Ln}(1/\epsilon)} \quad (3.4)$$

Ο όρος  $\text{Ln}F/\text{Ln}(1/\epsilon)$  τείνει στο μηδέν όπως  $\epsilon \rightarrow 0$ . Αρα:

$$D = \lim_{\epsilon \rightarrow 0} \left\{ \frac{\text{Ln}(N(A, \epsilon))}{\text{Ln}(1/\epsilon)} \right\} \quad (3.5)$$

Η ποσότητα  $D$  είναι η κλασματική διάσταση του συνόλου  $A$ . Στην περίπτωση συνόλων του χώρου  $R^m$  με χρήση Ευκλείδειας μετρικής, όπου αντί για σφαίρες χρησιμοποιούμε “κύβους” πλευράς  $(1/2^n)$  οι οποίοι τέμνουν σε τουλάχιστον ένα σημείο το  $A$ , η ποσότητα  $D$  δίνεται από την παρακάτω σχέση:

$$D = \lim_{n \rightarrow 0} \left\{ \frac{\text{Ln}(N(A))}{\text{Ln}(2^n)} \right\} \quad (3.6)$$

Εάν η μέθοδος Μέτρησης με "Κύβους" συνδυαστεί με μετρήσεις σχετικής πυκνότητας πληθυσμού των δίσκων τότε μπορούν να ορισθούν διαστάσεις ανώτερης τάξης [5, 4]. Οι διαστάσεις ανώτερης τάξης για ιδανικά σύνολα fractal ταυτίζονται με την κλασματική διάσταση. Οι διαστάσεις ανώτερης τάξης μπορούν να ξεπεράσουν προβλήματα που παρουσιάζονται με την κατηγοριοποίηση αντικειμένων με βάση την κλασματική διάσταση στην περίπτωση που δύο διαφορετικά σύνολα έχουν την ίδια κλασματική διάσταση. Η πληροφοριακή διάσταση (information dimension) και η συσχετιστική διάσταση (correlation dimension) μπορούν να βοηθήσουν στον αποσυσχετισμό των συνόλων. Τροποποιήσεις της μεθόδου "Μέτρησης με Κύβους" έχουν προταθεί για τον υπολογισμό των διαστάσεων ανώτερης τάξης.

### 3.3 Εκτίμηση κλασματικής Διάστασης D διδιάστατων αντικειμένων

Μία χαρακτηριστική ιδιότητα διδιάστατων αντικειμένων που μπορεί να συσχετιστεί με την κλασματική διάσταση είναι η επιφάνεια τους. Αντίστοιχα με την σχέση (3.3) ισχύει [12] η σχέση  $A(\epsilon) \approx F\epsilon^{2-D}$ . Η τρίτη μέθοδος για την μέτρηση του μήκους μίας ακτογραμμής, που αναφέρθηκε προηγουμένως, μπορεί να επεκταθεί σε δύο διαστάσεις. Εάν θεωρήσουμε το σύνολο των σημείων τα οποία σχηματίζουν ένα στρώμα πάχους  $2\epsilon$  γύρω από την υπό εξέταση επιφάνεια  $A$ , διαιρώντας τον όγκο  $V(A, \epsilon)$ , που καταλαμβάνει το σύνολο των σημείων με  $2\epsilon$ , έχουμε μία εκτίμηση της κλασματικής διάστασης του υπό εξέταση αντικειμένου. Η περιγραφή διδιάστατων ψηφιακών εικόνων με την χρήση μίας μόνο βαθμωτής ποσότητας όπως η κλασματική διάσταση δεν έχει αποδειχτεί αποδοτική για προβλήματα ανάλυσης και κατάτμησης εικόνων. Η εργασία [11] επεκτείνει την κλασματική κίνηση Brown και τον κλασματικό θόρυβο Gauss στο διακριτό πεδίο. Υπολογίζεται η συσχέτιση του θορύβου στο διακριτό πεδίο. Ο υπολογισμός του φάσματος ισχύος σε κλειστή μορφή είναι δυνατός μόνο σε ειδικές περιπτώσεις. Υπολογίζεται ένας εκτιμητής μέγιστης

πιθανοφάνειας για την κλασματική διάσταση  $D$  μαζί με το αντίστοιχο φράγμα Cramer - Rao. Στην περίπτωση διδιάστατων εικόνων η εικόνα χωρίζεται σε μικρά κομμάτια όπου υπολογίζεται οριζόντια, κατακόρυφα και διαγώνια η κλασματική διάσταση  $D$ . Οι μετρήσεις χρησιμοποιούνται στην διαδικασία της ανάλυσης εικόνας. Η μέθοδος της συνεχούς επιστρωμάτωσης [14] για τον υπολογισμό της επιφάνειας αντικειμένων και ο ορισμός της κλασματικής υπογραφής, ενός διανύσματος εκτιμήσεων της κλασματικής διάστασης σε μία κλίμακα αναλύσεων, έχει δώσει ικανοποιητικά αποτελέσματα στο πρόβλημα της ανάλυσης εικόνας.

### 3.3.1 Μέθοδος της συνεχούς επιστρωμάτωσης

Ορίζουμε ως επιστρωμάτωση μίας επιφάνειας, δύο επιφάνειες, την άνω  $u_\epsilon$  και την κάτω  $b_\epsilon$  οι οποίες περιβάλλουν την υπό εξέταση επιφάνεια σε απόσταση  $\epsilon$ . Εάν η αρχική επιφάνεια είναι μία εικόνα συνεχούς γκριζου τόνου (gray level image)  $g(i,j)$  και  $b_0 = u_0 = g(i,j)$  τότε οι πάνω και οι κάτω επιφάνειες για τα διάφορα  $\epsilon$  δίνονται ως εξής:

$$u_\epsilon(i,j) = \max \left\{ u_{\epsilon+1}(i,j) + 1, \max_{|(m,n)-(i,j)| \leq 1} u_{\epsilon-1}(m,n) \right\} \quad (3.7)$$

και

$$b_\epsilon(i,j) = \max \left\{ b_{\epsilon+1}(i,j) + 1, \max_{|(m,n)-(i,j)| \leq 1} b_{\epsilon-1}(m,n) \right\} \quad (3.8)$$

Τα σημεία  $(m,n)$  επιλέγησαν να είναι τα τέσσερα γειτονικά σημεία γύρω από τα  $(i,j)$  σε απόσταση μικρότερη από ένα. Παρόμοιες εκφράσεις υπάρχουν σε περίπτωση που θεωρήσουμε γειτονιά των 8 σημείων. Ένα σημείο  $(x,y,f)$  θα ανήκει στην επιστρωμάτωση για κάποιο  $\epsilon$  εάν  $b_\epsilon(x,y) \leq f \leq u_\epsilon(x,y)$ . Ο ορισμός της επιστρωμάτωσης μίας επιφάνειας για δεδομένο  $\epsilon$  περιέχει όλα τα σημεία της επιστρωμάτωσης για  $\epsilon-1$  και τα σημεία σε απόσταση 1 από την επιστρωμάτωση για  $\epsilon-1$ . Για παράδειγμα η σχέση (3.7) εγγυάται ότι η επιφάνεια  $u_\epsilon$  είναι κατά 1 μεγαλύτερη από την  $u_{\epsilon-1}$  και σε απόσταση μικρότερη από 1 κατά την οριζόντια ή την κατακόρυφη κατεύθυνση σε κάθε σημείο  $(i,j)$ . Ο όγκος της επιστρωμάτωσης  $\epsilon$  υπολογίζεται από τις επιφάνειες  $u_\epsilon$  και  $b_\epsilon$ :

$$v_\epsilon = \sum_{i,j} (u_\epsilon(i,j) - b_\epsilon(i,j)) \quad (3.9)$$

Το εμβαδόν της επιφάνειας για δεδομένο  $\epsilon$  δίνεται από τον όγκο που προστέθηκε στην επιστρωμάτωση  $\epsilon-1$  από τις δύο επιφάνειες διά 2:

$$A(\epsilon) = \frac{v_\epsilon - v_{\epsilon-1}}{2} \quad (3.10)$$

Η μέθοδος της επιστρωμάτωσης διαφέρει με την μέθοδο 3 στην ενότητα 1, η οποία προτείνει ότι  $A(\epsilon) = v/2\epsilon$ . Αυτό γίνεται ώστε ο όγκος  $v$  να μην εξαρτάται από τις λεπτομέρειες των προηγούμενων επιστρωματώσεων. Αφαιρώντας τον όγκο  $v_{\epsilon-1}$  απομονώνουμε τις λεπτομέρειες οι οποίες αλλάζουν από την κλίμακα  $\epsilon-1$  στην κλίμακα  $\epsilon$ . Στην περίπτωση ενός ιδανικού fractal αντικειμένου και οι δυο τρόποι υπολογισμού δίνουν τα ίδια αποτελέσματα γιατί οι αλλαγές είναι ανεξάρτητες της κλίμακας: στην περίπτωση όμως μη ιδανικών fractal αντικειμένων η απομόνωση των λεπτομερειών της μικρότερης κλίμακας είναι αναγκαία.

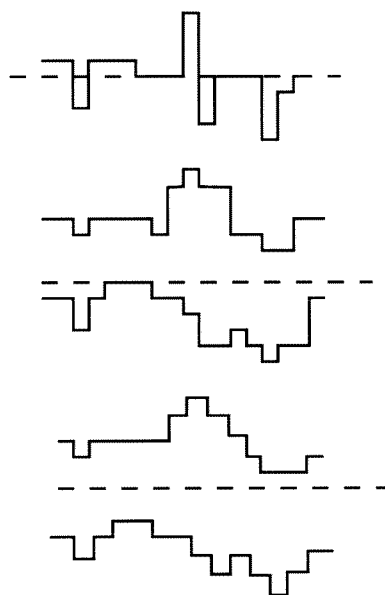
Η μέθοδος της επιστρωμάτωσης μπορεί να εφαρμοστεί και σε μία διάσταση στην μέτρηση του μήκους μίας γραμμής. Μια γραφική επισκόπηση της μεθόδου φαίνεται στο σχήμα 3.1. Οι επιστρωματώσεις για διάφορα  $\epsilon$  είναι:  $A(\epsilon=1)=47$ ,  $A(\epsilon=2)=78$ , και τα αντίστοιχα μήκη για τις ίδιες κλίμακες είναι:  $L(1)=(47-0)/2=23.5$  και  $L(2)=(78-47)/2=15.5$

### 3.4 Κλασματικός Πίνακας

Ενας τρόπος για να οριστεί μία διαδικασία περιγραφής αντικειμένων με βάση τα fractals είναι να συσχετισθεί μία μετρούμενη ποσότητα με την κλίμακα μέτρησής της. Δηλαδή να γραφτεί μια σχέση της μορφής:

$$F(\epsilon) = \epsilon^{-D} F(\epsilon = 1) \quad (3.11)$$

Η ποσότητα  $F(\epsilon)$  μπορεί να είναι το μήκος μίας γραμμής ή το εμβαδόν μίας επιφάνειας. Είναι δυνατόν όμως να θέλουμε να περιγράψουμε διανυσματικές ποσότητες



Σχήμα 3.1: Μια μονοδιάστατη περίπτωση της μεθόδου επικάλυψης. Η συνάρτηση  $g$  και οι επικαλύψεις για  $\epsilon=1,2$ . Οι όγκοι (επιφάνειες)  $v(1)=47$  και  $v(2)=78$ . Τα αντίστοιχα μήκη είναι  $L(1) = (47-0)/2 = 23.5$  και  $L(2) = (78-47)/2=13.5$

$\mathbf{F}(\epsilon) = [F_1(\epsilon), F_2(\epsilon), \dots, F_n(\epsilon)]$  σε σχέση με την κλίμακα μέτρησης. Σε αυτήν την περίπτωση έχουμε:

$$[\mathbf{F}(\epsilon)] = \epsilon^{-[D]} [\mathbf{F}(1)] \quad (3.12)$$

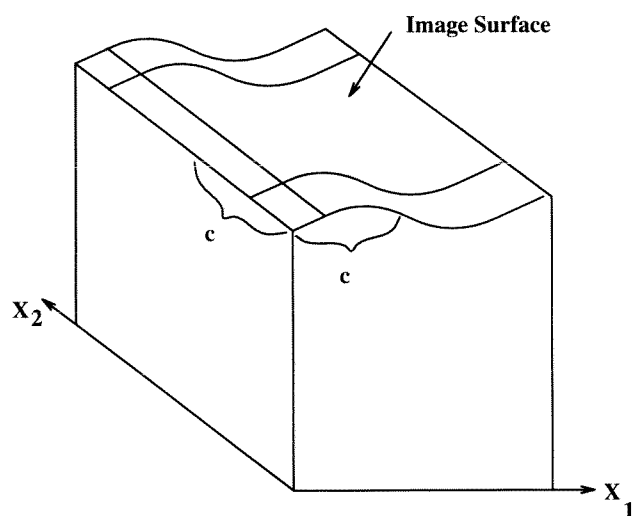
όπου ο πίνακας  $D$  λέγεται κλασματικός πίνακας fractal. Εάν  $n = 2$  τότε  $\mathbf{F}(\epsilon) = [F_1(\epsilon), F_2(\epsilon)]$  όπου  $F_1(\epsilon), F_2(\epsilon)$  μπορεί να θεωρηθεί το διάνυσμα μέτρησης του μήκους της προβολής μιας εικόνας στην οριζόντια και κατακόρυφη διεύθυνση για διάφορες κλίμακες μέτρησης  $\epsilon$ . Το σχήμα 3.2 δείχνει την περίπτωση για  $n=2$ .

Ο κλασματικός πίνακας είναι ανεξάρτητος της κλίμακας  $\epsilon$  αλλά εξαρτάται από το σύστημα συντεταγμένων. Εάν θέσουμε:

$[D] = d_{i,j}$  για  $i,j = 1,2$  και  $s = \log(\epsilon)$  και  $[Y(s)] = [F(\exp(s))]$  τότε έχουμε:

$$[Y(s)] = \exp(-s[D])[Y(0)] \quad (3.13)$$

Είναι πλέον εύκολο να δούμε ότι το αριστερό μέρος της σχέσης (3.13) είναι η λύση της ακόλουθης διαφορικής εξίσωσης:



Σχήμα 3.2: Γενική Ιδέα του κλασματικού πίνακα: Διάσχιση κατά μήκος των αξόνων  $X_1$  και  $X_2$

$$\begin{pmatrix} \dot{Y}_1(s) \\ \dot{Y}_2(s) \end{pmatrix} = \begin{pmatrix} d_{11} & d_{12} \\ d_{21} & d_{22} \end{pmatrix} * \begin{pmatrix} Y_1(s) \\ Y_2(s) \end{pmatrix} \quad (3.14)$$

όπου

$$Y_i(s) = F_i(s) \text{ for } i,j = 1,2$$

Εάν χρησιμοποιήσουμε δύο γραμμικά μοντέλα παρεμβολής θα έχουμε:

$$\frac{\Delta Y_1(s)}{\Delta s} = -(d_{11}Y_1(s) + d_{12}Y_2(s)) + e_1(s) \quad (3.15)$$

$$\frac{\Delta Y_2(s)}{\Delta s} = -(d_{21}Y_1(s) + d_{22}Y_2(s)) + e_2(s) \quad (3.16)$$

όπου  $e_1$  and  $e_2$  είναι οι όροι διατάραξης οι οποίοι θεωρούνται ανεξάρτητοι, έτσι ώστε:

$$E(e_i(s)e_j(t)) = 0 \text{ όταν } s \neq t \text{ και } E(e_i(s)) = 0 \text{ για } i,j = 1,2$$

Κατόπιν είναι δυνατόν να λύσουμε το σύστημα των σχέσεων (3.15),(3.16) χρησιμοποιώντας κάποια ευέλικτη μέθοδο ελαχίστων τετραγώνων.



Για την αξιολόγηση του μοντέλου ορίζουμε το κανονικοποιημένο σφάλμα μοντέλου (ΚΣΜ) να είναι το άθροισμα των διατάραξης του μοντέλου  $e_i(s)$  προς το άθροισμα των τετραγώνων του δεξιού μέλους των σχέσεων (3.15) και (3.16).

$$KSM = \frac{\sum_s (e_1^2(s) + e_2^2(s))}{\sum_s \left\{ \left( \frac{\Delta Y_1(s)}{\Delta s} \right)^2 + \left( \frac{\Delta Y_2(s)}{\Delta s} \right)^2 \right\}} \quad (3.17)$$

Αποτελέσματα της μεθόδου στην ανάλυση και κατάτμηση εικόνων υπάρχουν στην αναφορά [10].

### 3.5 Πολυκαναλικός Κλασματικός Πίνακας

Ολες οι αναλύσεις μέχρι στιγμής έχουν συγκεντρωθεί στην ανάλυση εικόνων αποχρώσεων του γκριζου (greyscale images) ή δίτιμων εικόνων. Από τα μοντέλα που παρουσιάστηκαν μέχρι στιγμής το μοντέλο του κλασματικού πίνακα επιλέχθηκε για επέκταση στην περίπτωση των έγχρωμων εικόνων. Ο πολυκαναλικός κλασματικός πίνακας παράγεται εάν γράψουμε την σχέση (3.12) για κάθε χρωματική συνιστώσα, δημιουργώντας έτσι ένα πίνακα από τρεις (2 x 2) πίνακες. Αυτό είναι ανάλογο με την διαδικασία συλλογής μετρήσεων  $F_i(\epsilon)$  για  $i = 1, 2, 3$  για διάφορες κλίμακες μέτρησης (αναλύσεων) και για όλες τις χρωματικές συνιστώσες. Οπότε έχουμε:

$$\begin{bmatrix} F_{11}(\epsilon) \\ F_{12}(\epsilon) \\ F_{21}(\epsilon) \\ F_{22}(\epsilon) \\ F_{31}(\epsilon) \\ F_{32}(\epsilon) \end{bmatrix} = \epsilon \begin{bmatrix} d_{11} & d_{12} & 0 & 0 & 0 & 0 \\ d_{21} & d_{22} & 0 & 0 & 0 & 0 \\ 0 & 0 & d_{33} & d_{34} & 0 & 0 \\ 0 & 0 & d_{43} & d_{44} & 0 & 0 \\ 0 & 0 & 0 & 0 & d_{55} & d_{65} \\ 0 & 0 & 0 & 0 & d_{65} & d_{66} \end{bmatrix} * \begin{bmatrix} F_{11}(1) \\ F_{12}(1) \\ F_{21}(1) \\ F_{22}(1) \\ F_{31}(1) \\ F_{32}(1) \end{bmatrix} \quad (3.18)$$

Γιατί συνεχίζεται ;

Στην ανάλυση έγχρωμων εικόνων χρησιμοποιούνται ποικίλες αναπαραστάσεις των έγχρωμων εικόνων, όπως η RGB (Κόκκινο, Πράσινο, Μπλέ) και η  $YCrCb$  (Φωτεινότητα, Χρωματικότητα). Η τελευταία αναπαράσταση προκύπτει από γραμμικό μετασχηματισμό της πρώτης [14]. Είναι γνωστό όμως [1] ότι η κλασματική διάσταση ενός γραμμικού μετασχηματισμού δύο σημάτων (συνόλων) δεν είναι ίση με τον γραμμικό μετασχηματισμό των κλασματικών διαστάσεων των σημάτων (συνόλων). Οπότε αναμένεται ότι η περιγραφή έγχρωμων εικόνων στο σύστημα (RGB) να δίνει καλύτερα αποτελέσματα στην ανάλυση με fractals. Ο ισχυρισμός αυτός επιβεβαιώνεται και από τα αριθμητικά αποτελέσματα που θα παρουσιαστούν στη συνέχεια.

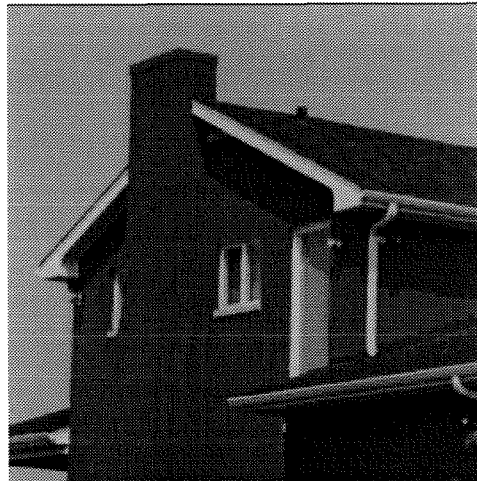
Οι χρωματικές συνιστώσες είναι συνήθως συσχετισμένες μεταξύ τους, μια ιδιότητα την οποία μπορούμε να εκμεταλλευτούμε εάν επιτρέψουμε στη σχέση (3.18) τα μηδενικά στοιχεία να πάρουν μη μηδενικές τιμές. Οπότε η σχέση (3.18) μετατρέπεται στην ακόλουθη:

$$[\vec{F}_{ij}(\epsilon)] = \epsilon^{-[D]} * [\vec{F}_{ij}(1)] \text{ για } i=1,2,3 \text{ } j=1,2 \quad (3.19)$$

Ένα στοιχείο που χρήζει προσοχής, γενικότερα των μεθόδων με βάση τα fractals, είναι το εύρος των περιοχών πάνω στις οποίες εφαρμόζονται οι μέθοδοι. Στην παρούσα ανάλυση μελετήθηκαν δύο ειδών περιοχές  $8 \times 8$  και  $16 \times 16$ . Εάν η μέθοδος πρόκειται να χρησιμοποιηθεί για ανάλυση και κατάτμηση εικόνων, τότε το μέγεθος του block εικόνας είναι σημαντικό στοιχείο της ανάλυσης. Η επιλογή μεγάλου μεγέθους block εικόνας οδηγεί σε μεγάλες γραμμές οι οποίες εκτιμάται ότι περιέχουν απότομες αλλαγές της φωτεινότητας οι οποίες είναι σχετικά δύσκολο να μοντελοποιηθούν. Αντίθετα block μικρού μεγέθους εμφανίζει γραμμές “χαμηλού μετώπου” οι οποίες οδηγούν σε υποορισμένα συστήματα εξισώσεων της μορφής (3.18). Μια διαδικασία επιλογής μεταβαλλόμενου μεγέθους του block μπορεί να κατασκευαστεί. Αρχικά επιλέγεται μέγεθος block ( $16 \times 16$ ) και στη συνέχεια ελέγχεται εάν η ανάλυση σε τέσσερα block είναι εφικτή και οδηγεί σε μικρότερο σφάλμα μοντελοποίησης στην σχέση (3.17).

Τα μοντέλα που αναπτύχθηκαν παραπάνω μπορούν να χρησιμοποιηθούν για να τροφοδοτήσουν τις εισόδους είτε κλασικών κατηγοροποιητών [2] όπως κατηγοριοποιητής ελάχιστης απόστασης, Bayes, K πλησιέστερων γειτόνων, είτε νευρωνικών δικτύων τύπου Back Propagation και Διανυσματικού Κβαντιστή με Εκμάθηση (LVQ) . Επιπλέον η διαδικασία της κατάτμησης εικόνας μπορεί να θεωρηθεί ως το επόμενο βήμα της διαδικασίας κατηγοριοποίησης.

Μιά έγχρωμη εικόνα ενός σπιτιού, η οποία φαίνεται στο σχήμα 3.3 χρησιμοποιήθηκε για να διαλέξουμε τριών ειδών έγχρωμες εικόνες υψής οι οποίες φαίνονται στα σχήματα 3.4,3.5 και 3.6.



Σχήμα 3.3: Αρχική εικόνα



Σχήμα 3.4: Κόκκινο Πράσινο Μπλέ (ως αποχρώσεις του γκριζου) του Κάτω Δεξιού Τοίχου

Αυτές οι εικόνες χρησιμοποιήθηκαν για τον υπολογισμό των τιμών  $F(\epsilon)$  στις σχέσεις (3.3), (3.18) και (3.19) για διάφορα μεγέθη εικόνας. Χρησιμοποιώντας τεχνική Ελαχίστων Τετραγώνων λύθηκαν τα παραπάνω συστήματα εξισώσεων. Οι πίνακες 3.1 και 3.2 δείχνουν την μέση τιμή και τυπική απόκλιση του κανονικοποιημένου σφάλματος μοντελοποίησης (ΚΣΜ), όπως ορίζεται στην σχέση (3.17) για μεγέθη block 8 x 8 και 16 x 16 για



Σχήμα 3.5: Κόκκινο Πράσινο Μπλέ (ως αποχρώσεις του γκριζου) Κόκκινο Πράσινο Μπλέ (ως αποχρώσεις του γκριζου) της οροφής



Σχήμα 3.6: Κόκκινο Πράσινο Μπλέ (ως αποχρώσεις του γκριζου) του Μπροστινού Τοίχου

τις ακόλουθες περιπτώσεις: (α) χρησιμοποίηση μόνο της φωτεινότητας της εικόνας της σχέσης (3.3), (β) χρησιμοποίηση της αναπαράστασης (RGB) στην σχέση (3.19) και (γ) χρησιμοποίηση της αναπαράστασης (YCrCb) στην σχέση (3.19). Εύκολα επιβεβαιώνεται ότι η αναπαράσταση (RGB) στην σχέση (3.19) οδηγεί σε μικρότερο σφάλμα μοντελοποίησης. Αναφορικά με το μέγεθος του block ανάλυσης το σχήμα 3.7 δείχνει το σφάλμα μοντελοποίησης για μεγέθη block 8 x 8 και 16 x 16. Τα αποτελέσματα δείχνουν ότι το σφάλμα μοντελοποίησης είναι γενικά μικρότερο στην περίπτωση block μεγέθους 8 x 8 από την αντίστοιχη περίπτωση των block μεγέθους 16 x 16.

	Gray	YCrCb	RGB(12)	RGB(36)
8 x 8	0.1834	0.7864	0.1706	$1.068 \cdot 10^{-13}$
16 x 16	0.2261	2.5349	0.2017	0.071

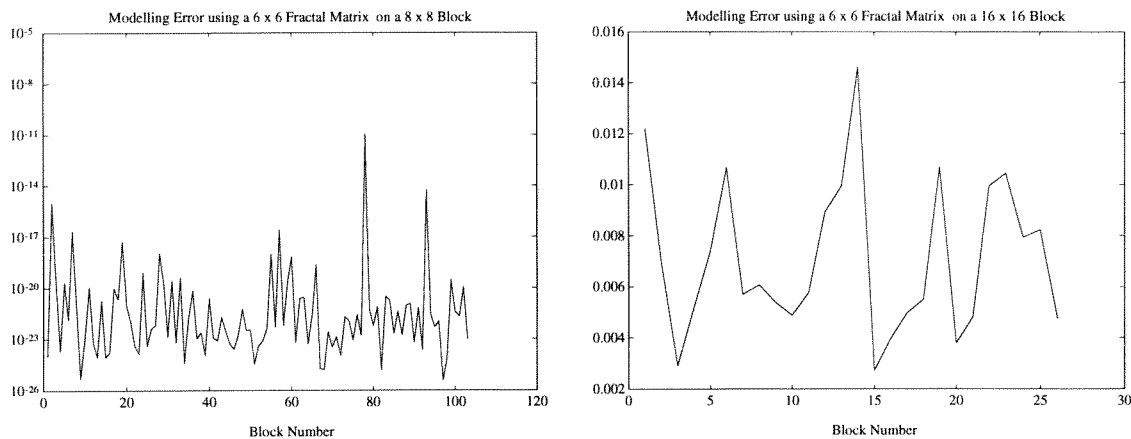
Πίνακας 3.1: Μέση τιμή του κανονικοποιημένου σφάλματος μοντελοποίησης

Πραγματοποιήθηκε διαδικασία κατηγοροποίησης χρησιμοποιώντας νευρωνικά δίκτυα όπως πολυεπίπεδο perceptron και διανυσματικού κβαντιστή με εκμάθηση στις εικόνες υψής των σχημάτων 3.4 και 3.6. Τα μισά από τα διαθέσιμα blocks χρησιμοποιήθηκαν για εκπαίδευση ενώ τα υπόλοιπα χρησιμοποιήθηκαν για έλεγχο. Το ποσοστό επιτυχίας ήταν γύρω στο 80 % και για τους δύο τύπους νευρωνικών δικτύων όταν χρησιμοποιήθηκε μόνο η φωτεινότητα στην σχέση (3.18), ενώ το ποσοστό επιτυχίας ήταν πάνω από 90 % όταν

	Gray	YCRCB	RGB(12)	RGB(36)
8 x 8	0.1181	2.8237	0.0725	$1.083 \cdot 10^{-12}$
16 x 16	0.1202	5.3686	0.0813	0.031

Γιατί blocks;

Πίνακας 3.2: Τυπική απόκλιση του κανονικοποιημένου σφάλματος μοντελοποίησης

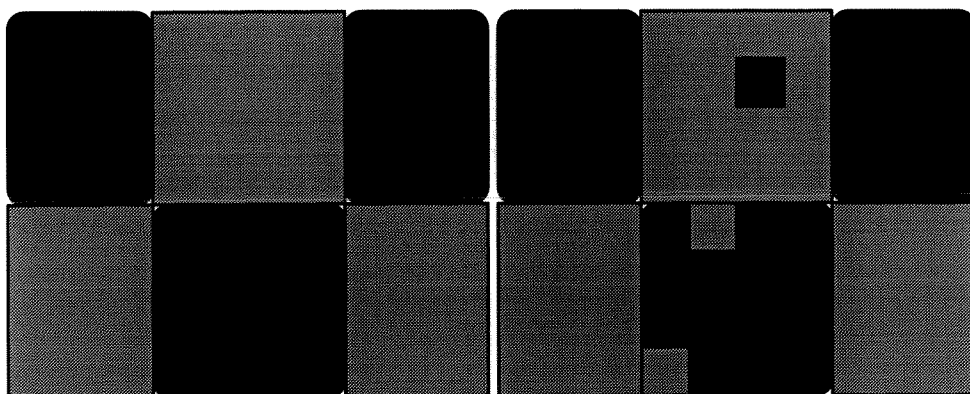


Σχήμα 3.7: Comparative plot for 8 x 8 and 16 x 16 blocks

χρησιμοποιήθηκε στην σχέση (3.19) η αναπαράσταση (RGB) (η οποία έδινε και μεγαλύτερα διανύσματα χαρακτηριστικών). Τέλος, χρησιμοποιήσαμε τον κατηγοριοποιητή (που προηγουμένως εκπαιδεύσαμε με βάση την (RGB) αναπαράσταση) σε μια διαδικασία κατάτμησης εικόνας όπως αυτής που φαίνεται στο σχήμα 3.8. Το σχήμα 3.8 κατασκευάστηκε από εικόνες υψής των σχημάτων 3.4 και 3.6. Επειδή στην φυσική οπτική μορφή η διάκριση των βασικών κομματιών είναι δύσκολη εξαιτίας της συγγενούς υψής και χρωματικότητας, οι δύο περιοχές εμφανίζονται στην μορφή (γκρίζου - μαύρου).

### 3.6 Μια πολυμοντελική προσέγγιση για την ταξινόμηση εικόνων με νευρωνικά δίκτυα

Επειδή η ανάλυση εικόνων με fractals εξετάζει τα αντικείμενα σε πολλαπλές αναλύσεις, από τις οποίες εξάγεται η κλασματική διάσταση ή ο κλασματικός πίνακας, παρουσιάζει το χαρακτηριστικό να είναι αναλλοίωτη ως προς την αλλαγή κλίμακας. Η ιδιότητα αυτή είναι



Σχήμα 3.8: Διαδικασία Κατηγοροποίησης: Αρχική και τελική εικόνα

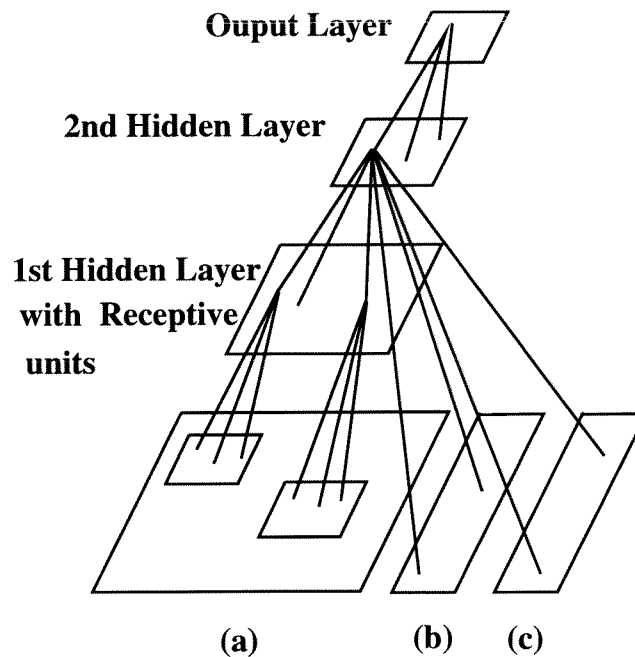
επιθυμητή σε ένα σύστημα κατηγοριοποίησης /αναγνώρισης γιατί αυξάνει την ανθεκτικότητα του συστήματος. Επιπλέον είναι επιθυμητό το σύστημα να έχει και τις ακόλουθες ιδιότητες: (α) αναλλοιωσιμότητα ως προς την μετατόπιση και (β) αναλλοιωσιμότητα ως προς την στροφή. Η εργασία [16] παρουσιάζει μια αναλλοίωτη διδιάστατη αναπαράσταση αντικειμένων βασισμένη σε συσχετίσεις τρίτης τάξης και μια μέθοδο με ροπές υψηλής τάξης. Προτείνονται νευρωνικά δίκτυα υψηλής τάξης τα οποία παρουσιάζουν επιπλέον ανθεκτικότητα στον προσθετικό γκαουσιανό θόρυβο.

Στην παρούσα ενότητα παρουσιάζεται μία συνολική προσέγγιση για το πρόβλημα της ταξινόμησης εικόνων όπου χρησιμοποιούμε τρεις διαφορετικούς τύπους αναλλοίωτων χαρακτηριστικών για εισαγωγή σε ένα νευρωνικό δίκτυο το οποίο θα υλοποιεί την ταξινόμηση. Οι τρεις τύποι χαρακτηριστικών που θα χρησιμοποιηθούν είναι οι εξής:

- Οι αναλλοίωτες συναρτήσεις ροπών [17].
- Οι αναλλοίωτες παραστάσεις των σχέσεων [18].
- Οι συντελεστές που προκύπτουν από fractal μοντελοποίηση των εικόνων.

Στο σχήμα 3.9 παρουσιάζουμε μία δομή νευρωνικού δικτύου το οποίο μπορεί να χειρίζεται και τους τρεις τύπους χαρακτηριστικών. Η είσοδος-(α) του δικτύου δέχεται την αναλλοίωτη παράσταση διαστάσεων  $L_1 \times L_2$  [18]. Το πρώτο κρυφό επίπεδο έχει νευρώνες που βλέπουν περιορισμένες περιοχές (παράθυρα) της εισόδου-(α) (receptive field type neurons), της 2-Δ παράστασης που βασίζεται στα τριπλά γινόμενα. Το δεύτερο κρυφό επίπεδο συνδυάζει αυτή την πληροφορία με τα χαρακτηριστικά των fractals και των ροπών,

δεδομένου ότι οι νευρώνες σε αυτό το επίπεδο συνδέονται με τις εισόδους-( $\beta$ ) και -( $\gamma$ ) που περιέχουν τα χαρακτηριστικά αυτά (βλέπε σχήμα 3.9).



Σχήμα 3.9: Νευρωνικό δίκτυο για ταξινόμηση εικόνων βασιζόμενο στα χαρακτηριστικά: (α) Συσχετίσεις τρίτης τάξης (β) Συντελεστές fractal (c) Συντελεστές ροπών.

Θα πρέπει να σημειωθεί ότι δεν είναι αναγκαίες πάντα και οι τρεις αναπαραστάσεις για να λυθεί ένα συγκεκριμένο πρόβλημα, αλλά μπορεί να χρησιμοποιηθεί οποιοσδήποτε συνδυασμός εισόδων (χαρακτηριστικών).

Στο πείραμα αυτό θα χρησιμοποιήσουμε αντί συνθετικού παραδείγματος, μία πραγματική εφαρμογή ώστε να εξετάσουμε την απόδοση και την δυνατότητα γενίκευσης των νευρωνικών δικτύων. Ο έλεγχος των κολλήσεων σε τυπωμένα κυκλώματα είναι το προς επίλυση πρόβλημα. Να τονιστεί ότι στην εφαρμογή αυτή οι κλασικές τεχνικές αναγνώρισης δεν είχαν επιδείξει επιτυχή συμπεριφορά. Gray-scale εικόνες 2- $\Delta$  που παριστάνουν το ύψος της κόλλησης σα συνάρτηση της θέσης μέσα στην περιοχή της κόλλησης, συλλέχτηκαν με τη βοήθεια ενός οπτικού laser scanner και χρησιμοποιήθηκαν ως σήματα που έπρεπε να κατηγοριοποιηθούν σε δύο ομάδες, των καλών και των ελαττωματικών κολλήσεων (οι δεύτερες περιέχουν ανεπαρκή ποσότητα μολύβδου). Ένα δείγμα από κάθε

κατηγορία φαίνεται στο σχήμα 3.10.



Σχήμα 3.10: Παραδείγματα καλής (α) και ελαττωματικής (β) κόλλησης.

100 εικόνες διαχωρίστηκαν σε δύο σύνολα, μάθησης και ελέγχου. Οι αναλλοίωτες παραστάσεις που βασίζονται σε συσχετίσεις τρίτης τάξης έδειξαν ικανότητα κατηγοριοποίησης των εικόνων κολλήσεων με ποσοστά από 90% έως 96%.

Η χρήση του κλασματικού πίνακα με χρήση μόνο της φωτεινότητας, έδωσε ποσοστά επιτυχίας από 80% έως 84%, ενώ τα χαρακτηριστικά των ροπών έδωσαν ποσοστά μεταξύ του 70% και 75%. Ακολούθως εξετάστηκαν δίκτυα με διάφορους συνδυασμούς χαρακτηριστικών, με πιο αποτελεσματικό τον συνδυασμό των παραστάσεων με βάση τις τρίτης τάξης συσχετίσεις και τους συντελεστές fractal. Στον πίνακα 3.3 συνοψίζονται τα παραπάνω αποτελέσματα ταξινόμησης.

### 3.7 Συμπεράσματα

Στο κεφάλαιο αυτό παρουσιάστηκε η έννοια των fractals και της κλασματικής διάστασης. Παρουσιάστηκαν μέθοδοι μέτρησης της κλασματικής διάστασης των fractals για αντικείμενα και σύνολα σε μία και δύο διαστάσεις. Έγινε αναφορά σε επεκτάσεις της



Μέθοδος	Ποσοστό επιτυχίας ταξινόμησης
Συσχετίσεις τρίτης τάξης	90% - 96%
Συντελ. fractal	80% - 84%
Συντελ. ροπών	70% - 75%
Συσχετίσεις τρίτης τάξης & fractals	94% - 96%

Πίνακας 3.3: Ποσοστά επιτυχίας της ταξινόμησης του πολυμοντελικού σχήματος.

κλασματικής διάστασης για διαδικασίες κατηγοροποίησης εικόνων υψής. Ορίστηκε ο πολυκαναλικός κλασματικός πίνακας σε έγχρωμες εικόνες υψής. Εξετάστηκαν περιπτώσεις συστήματων χρωματικών συντεταγμένων (RGB) και (YCrCb) για διάφορα μεγέθη block. Ο πολυκαναλικός κλασματικός πίνακας κατέληξε σε μικρότερο σφάλμα μοντελοποίησης από τον αντίστοιχο κλασματικό πίνακα. Παρουσιάστηκαν περιπτώσεις αναγνώρισης / κατηγοριοποίησης συνθετικών και πραγματικών εικόνων.

Ποια η συνθετικά που fractals;  
 Παιδιά παρατήσαν 3d εικόνες  
 Fractal

Διαφ.

Καταφέρω να / κατηγοριοποιήσω

## Αναφορές

- [1] M.F. Barnsley, "Fractals Everywhere", *Academic Press*, Boston
- [2] R. Duda and P. Hart, "Pattern Classification and Scene Analysis", *John Willey & Sons*, 1973.
- [3] k. Falconer, "Fractal Geometry: Mathematical Foundation and Foundation", *John Willey Sons*, New York
- [4] P. Grassberg and I. Procaccia, "Measuring The Strangeness of Strange Attractors", *Physica 9*, pp. 189 - 208, 1983.
- [5] H.G.E Hentscel and I. Procaccia, "The Infinite Number of Generalised Dimensions of Fractals and Strange Attractors", *Physica 8D*, pp. 435 - 444. 1988.
- [6] B.K.P Horn and I.K. Ikeuchi, "Numerical shape from shading and occluding boundaries", *Artificial Inteligence*, Special Issue on Computer Vision, vol. 15 pp. 141 - 184, 1981
- [7] W. E. L. Grimson, "Computing shape using theory of human stereo vision", *Ph.D thesis*, Dep. Mathematics, Mass. Inst. Techn., MA 1980
- [8] A. K. Jain, " Fundamentals of Digital Image Processing ". *Prentice Hall*. 1989.
- [9] D. Kalogeras and S. Kollias "Image Classification and Segmentation Using Multichannel Fractal Modelling", *SPIE Visual Comm. and Image Proc.*, Nov. 1992. Boston. U.S.A
- [10] H. Kaneko. " A Generalised Fractal Dimension And Its Application To Texture Analysis". *ICASSP-89*. Glasgow. U.K., May 1989.

- [11] T. Lundal, W. Ohley, S. Kay and R. Siffert, "Fractional Brownian Motion: A Maximum Likelihood Estimator and Its Application to Image Texture", *IEEE Trans. on Medical Imaging*, vol. MI-5, no. 3, 1986.
- [12] B. B. Mandelbrot, *The Fractal Geometry of Nature*, San Fransisco: Freeman, 1982.
- [13] B.B. Mandelbrot and J. van Ness, "Fractional Brownian Motion, Fractional Noise and Application", *SIAM Review*, vol. 10, no. 4, pp. 422 - 437.
- [14] S. Peleg, J. Naor, R. Hartley and D. Avnir, "Multiple Resolution Texture Analysis and Classification", *IEEE Trans. Pattern. Anal. Mach. Intell.*, PAMI-6, pp.518-523. July 1984.
- [15] A. P. Pentland, "Fractal-Based Description of Natural Scenes", *IEEE Trans. Pattern Anal. Mach. Intell.*, PAMI-6, pp. 661-674, Nov. 1984.
- [16] A. Tirakis, D. Kalogeras and S. Kollias "A multimodel Approach to Image Classification Using Neural Networks", *Journal of Neural Network World*, vol. 2, pp. 269 - 287, 1992
- [17] Α. Τυράκης, "Ανναλοιώτη και Πολυδιακριτική Παράσταση Διδιάστατων σημάτων - Εφαρμογές στην Ανάλυση Εικόνων με Νευρωνικά Δίκτυα", *Διδακτορική Διατριβή*. Κεφ. 2, Οκτ. 1993
- [18] Α. Τυράκης "Ανναλοιώτη και Πολυδιακριτική Παράσταση Διδιάστατων σημάτων - Εφαρμογές στην Ανάλυση Εικόνων με Νευρωνικά Δίκτυα", *Διδακτορική Διατριβή*. Κεφ. 4, Οκτ. 1993

# Αποτελεσματική κατηγοριοποίηση εικόνων με χρήση νευρωνικών δικτύων και πολυδιακριτής ανάλυσης

## 4.1 Εισαγωγή

Υπάρχουσες βιομηχανικές εφαρμογές σε αναγνώριση / κατηγοριοποίηση προτύπων χρησιμοποιούν συνήθως μοντέλα τα οποία βασίζονται σε αποθήκευση γεωμετρικών χαρακτηριστικών για αναγνώριση / κατηγοριοποίηση με πρότυπα (patern matching). Ένα πρόβλημα που συνήθως ανακύπτει σε αυτές τις εφαρμογές σχετίζεται με τον όγκο των δεδομένων στην περίπτωση εικόνων με αποχρώσεις του γκριζου ή, πολύ περισσότερο, έγχρωμων και πολυκαναλικών εικόνων. Πολλές τεχνικές έχουν εφαρμοσθεί για την μείωση της πολυπλοκότητας του προβλήματος (*curse of dimensionality*), όπως ιδιαίζουσες ιδιοτιμές, ανάλυση με βάση τις πρωτεύουσες τιμές κ.α. Τα τεχνητά νευρωνικά δίκτυα, π.χ τα πολυεπίπεδα perceptrons, έχουν αποδειχτεί ικανά για να εξάγουν κατάλληλα βαθμωτά ή διδιάστατα χαρακτηριστικά [20, 27] ενώ ταυτόχρονα, εξαιτίας της μαζικότητας και παραλληλίας τους, μπορούν να είναι αποδοτικά σε προβλήματα μεγάλων διαστάσεων όπως αυτά που αναφέραμε προηγουμένως. Αλγόριθμοι εκμάθησης με επίβλεψη και με ανάδραση σφάλματος (error-feedback) όπως ο αλγόριθμος backpropagation, χρησιμοποιούνται εν γένει για την εκμάθηση των πολυεπίπεδων feedforward δικτύων. Τόσο τα πολυεπίπεδα perceptron όσο

και τα πιθανοτικά νευρωνικά δίκτυα είναι ικανά να προσεγγίσουν οσοδήποτε καλά κατηγοριοποιητές τύπου Bayes ή K πλησιέστερων γειτόνων [14, 21]. Σε αυτή την περίπτωση τα δομημένα νευρωνικά δίκτυα κατηγοριοποιητές είναι ιδιαίτερα αποτελεσματικά όταν εφαρμόζονται απευθείας στις τιμές των pixels των εικόνων-προτύπων και όχι σ' ένα σύνολο χαρακτηριστικών που έχουν εξαχθεί από αυτές [1, 15].

Εντούτοις, οι πραγματικές εικόνες είναι γενικά μεγάλων διαστάσεων. Συνεπώς, ο αναγκαίος αριθμός βαρών διασύνδεσης, ειδικά μεταξύ των μονάδων εισόδου και αυτών του πρώτου κρυφού επιπέδου, μπορεί να είναι τεράστιος, καταλήγοντας σε απογορευτικούς χρόνους μάθησης, όπως και σε μη ικανοποιητική γενίκευση [5, 13].

Η παράσταση σημάτων σε πολλαπλά διακριτικά επίπεδα έχει πρόσφατα χρησιμοποιηθεί σε πολλές εφαρμογές όπως κωδικοποίηση subband, όραση υπολογιστών για εκτίμηση κίνησης και προβλήματα στερεοσκοπικής όρασης, εύρεσης ακμών, εκτίμηση του pitch κτλ. Το βασικό πρόβλημα στο πεδίο αυτό είναι η αποτελεσματικότερη δυνατή αναπαράσταση των σημάτων στα διάφορα διακριτικά επίπεδα, ή (με άλλα λόγια), η κατάλληλη για ένα συγκεκριμένο πρόβλημα επιλογή συνάρτησης wavelet, ή ισοδύναμα μίας συστοιχίας φίλτρων, η διαδοχική εφαρμογή της οποίας δίνει την αναπαράσταση σε διαδοχικά διακριτά επίπεδα του σήματος εισόδου σε διαδοχικά διακριτά επίπεδα.

Οι πολυδιακριτές παραστάσεις έχουν αποδειχθεί ιδιαίτερα αποτελεσματικές για την ανάλυση του πληροφοριακού περιεχομένου των εικόνων [3, 17]. Ειδικότερα η πολυδιακριτή ανάλυση οδηγεί σε μία αναλλοίωτη, ως προς το μέγεθος, κατανόηση της εικόνας δεδομένου ότι οι χαμηλών διακριτικότητας αναλύσεις περιέχουν τα χονδροειδή χαρακτηριστικά των εικόνων. Στη συνέχεια προτείνουμε μία βέλτιστη πολυδιακριτή ανάλυση των εικόνων εισόδου στο νευρωνικό δίκτυο, που επιτρέπει μία σημαντική ελάττωση του μεγέθους του δικτύου η οποία είναι απαραίτητη για την επίλυση του προβλήματος κατηγοριοποίησης. Προφανώς υπάρχει κάποια απώλεια πληροφορίας καθώς λαμβάνουμε εικόνες χαμηλότερων διακριτικών επιπέδων. Εντούτοις, όταν αυξάνουμε ακολούθως το επίπεδο διακριτικότητας, περισσότερη πληροφορία προστίθεται στις εικόνες με συνέπεια την διαδοχικά ικανοποιητικότερη κατηγοριοποίηση. Τέτοιες coarse-to-fine στατηγικές έχουν μελετηθεί στη χαμηλού επιπέδου επεξεργασία εικόνων, όπως template matching, και είναι ιδιαίτερα υποσχόμενες

για αλγορίθμους αναγνώρισης προτύπων.

Ο συνδυασμός της πολυδιακριτής ανάλυσης με νευρωνικά δίκτυα μπορεί να οδηγήσει σε αποτελεσματικές αρχιτεκτονικές στην κατηγοριοποίηση εικόνων - προτύπων. Ειδικότερα σε νευρωνικά δίκτυα τύπου πολυεπιπεδικού perceptron καταλήγει σε μία ουσιαστική μείωση του αριθμού των βαρών διασύνδεσης και φυσικά του χρόνου μάθησης των δικτύων. Στην περίπτωση πιθανοτικών νευρωνικών δικτύων (probabilistic neural network) η μείωση του αριθμού των βαρών διασύνδεσης οδηγεί σε μείωση απαιτήσεων μνήμης.

Η δομή του παρόντος κεφαλαίου είναι η ακόλουθη: στην ενότητα 4.2 γίνεται μια συνοπτική παρουσίαση της πολυδιακριτής ανάλυσης · στην ενότητα 4.3 παρουσιάζεται το πιθανοτικό νευρωνικό δίκτυο. Στην ενότητα 4.4 παρουσιάζεται μία δυνατή μίξη των νευρωνικών δικτύων και της πολυδιακριτικής ανάλυσης · εισάγεται το ιεραρχικό πιθανοτικό δίκτυο και δίνονται αποτελέσματα στην κατηγοριοποίηση πραγματικών εικόνων. Στην ενότητα 4.5 παρουσιάζονται αποτελέσματα προσομοιώσεων χρησιμοποιώντας πραγματικές εικόνες και γίνονται συγκρίσεις με πολυδιακριτή παραλλαγή του πολυεπιπεδικού perceptron.

## 4.2 2-Δ Πολυδιακριτική Ανάλυση

Η πολυρυθμική επεξεργασία σημάτων έχει πρόσφατα χρησιμοποιηθεί με επιτυχία σε διάφορες εφαρμογές, όπως επεξεργασία και κωδικοποίηση φωνής και εικόνας, φασματική ανάλυση, ταυτοποίηση χρονικά μεταβαλλόμενων συστημάτων, συστήματα κεραιών και ραντάρ. Παράλληλα, αυξανόμενο είναι το ενδιαφέρον για πολυδιακριτές παραστάσεις σημάτων. Οι παραστάσεις σε διαδοχικά διακριτικά επίπεδα έχουν αποδειχθεί αποτελεσματικές στην ανάλυση της πληροφορίας σημάτων και ειδικότερα εικόνων, δημιουργώντας μία νέα οπτική γωνία αντιμετώπισης του προβλήματος της κωδικοποίησης της πληροφορίας για μετάδοση ή αποθήκευση. Η αναπαράσταση σημάτων σε πολλά διακριτικά επίπεδα γνώρισε νέα ώθηση με την χρήση του διακριτού μετασχηματισμού wavelet ( Discrete Wavelet Transform), υλοποιημένου με συστοιχίες φίλτρων (filter banks) χρησιμοποιώντας τα γνωστά quadrature mirror filters (QMFs) [17, 22]. Στην περιοχή της επεξεργασίας εικόνων, οι

παραπάνω ιδέες συγκλίνουν στην επεξεργασία κατά ζώνες (subband processing) [26, 28].

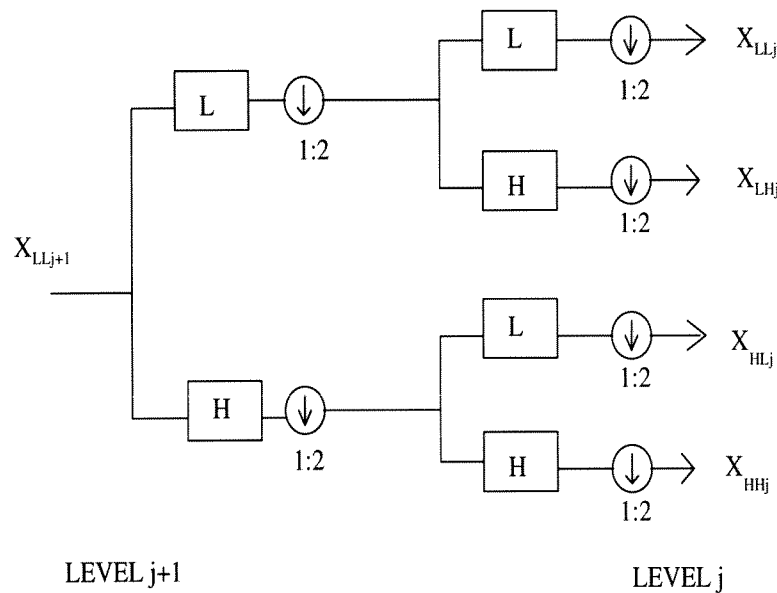
Η επεξεργασία subband αποτελεί ένα πεδίο εκτεταμένης έρευνας. Η κύρια ιδέα είναι ο διαδοχικός διαχωρισμός του αρχικού σήματος σε σήματα με συμπληρωματικό φασματικό περιεχόμενο και η επεξεργασία κάθε σήματος subband ανεξάρτητα, με ακρίβεια ανάλογη των στατιστικών χαρακτηριστικών του συγκεκριμένου φασματικού περιεχομένου του. Ο τρόπος επεξεργασίας των σημάτων subband εξαρτάται από την συγκεκριμένη εφαρμογή. Σε εφαρμογές κωδικοποίησης για παράδειγμα, κάθε σήμα subband αναλύεται με τέτοιο τρόπο ώστε να μεταδοθεί τελικά η λιγότερη δυνατή πληροφορία που αρκεί για ικανοποιητική ανακατασκευή στον δέκτη. Στην περίπτωση αυτή, τα σήματα subband διαταράσσονται εξαιτίας προσθετικού θορύβου και θορύβου λόγω κβαντισμού. Επιπλέον, είναι πιθανό ότι κάποια από τα σήματα subband χάνονται κατά την μετάδοση ή δεν μεταδίδονται για λόγους συμπίεσης της μεταδιδόμενης πληροφορίας ή λόγω του περιορισμένου εύρους του καναλιού μετάδοσης, με αποτέλεσμα στον δέκτη να χρησιμοποιείται ένας περιορισμένος αριθμός σημάτων subband για την ανακατασκευή.

Ο διαχωρισμός του αρχικού σήματος (εικόνας) και η μετάβαση από ένα επίπεδο ανάλυσης σε ένα επίπεδο μικρότερης ανάλυσης υλοποιείται με την υποδειγματοληψία ανά δύο σε κάθε κατεύθυνση με την χρήση κατάλληλης συστοιχίας φίλτρων ανάλυσης QMF. Τα συγκεκριμένα φίλτρα είναι τύπου FIR με σταθερούς συντελεστές · τυπικά φίλτρα με 8 ή 16 συντελεστές δίνονται στο [4]. Με αυτό τον τρόπο παράγονται τέσσερις εικόνες διαστάσεων  $(/12 \times /2)$ . Μία από αυτές αντιστοιχεί στην βαθυπερατή έκδοση της αρχικής εικόνας. Μπορούμε να επαναλάβουμε την ίδια διαδικασία μόνο στο βαθυπερατό αντίγραφο της εικόνας μέχρι να φτάσουμε στην επιθυμητή διακριτική ανάλυση. Αυτή η διαδικασία φαίνεται στο σχήμα 4.1. Εστω  $x_0$  μία  $N \times N$  2-Δ εικόνα. Χρησιμοποιώντας κατάλληλα FIR φίλτρα τέλει ανακατασκευής  $h_L(n)$  και  $h_H(n)$ , όπου  $h_L(n)$  ένα κατωδιαβατό και  $h_H(n)$  ένα υψιπερατό φίλτρο, μπορούμε να αναλύσουμε την αρχική εικόνα σε τέσσερις μικρότερης διακριτότητας  $\frac{N}{2} \times \frac{N}{2}$  υποεικόνες [17]. Εφαρμόζοντας για παράδειγμα, το κατωδιαβατό φίλτρο  $h_L(n)$  κατά την οριζόντια και ακολούθως την κατακόρυφη κατεύθυνση της αρχικής εικόνας (έχοντας θεωρήσει την διαχωρίσιμη περίπτωση αρχικά για λόγους

απλότητας), λαμβάνουμε την προσεγγιστική εικόνα (approximation image) στο συγκεκριμένο μικρότερης διακριτότητας επίπεδο  $j = -1$ , την οποία συμβολίζουμε ως εξής

$$x_{-1}^{LL}(m, n) = \sum_{k=1}^N \sum_{l=1}^N h_L(2m - k)h_L(2n - l)x_0(k, l) \quad (4.1)$$

Εφαρμόζοντας όλους τους δυνατούς συνδυασμούς των FIR φίλτρων  $h_L(n)$  και  $h_H(n)$ , λαμβάνουμε τις τρεις μικρότερης διακριτότητας εικόνες λεπτομέρειας (detail images), που τις συμβολίζουμε ως  $\mathbf{x}_{-1}^{LH}$ ,  $\mathbf{x}_{-1}^{HL}$  και  $\mathbf{x}_{-1}^{HH}$ . Η παραπάνω διαδικασία παρουσιάζεται στο σχήμα 4.1.



Σχήμα 4.1: Συστοιχία φίλτρων ανάλυσης των εικόνων εισόδου στο αμέσως μικρότερο διακριτικό επίπεδο.

Επιπλέον, αν η παραπάνω διαδικασία ανάλυσης εφαρμοστεί διαδοχικά στις προσεγγιστικές εικόνες του κάθε επιπέδου, λαμβάνουμε την πολυδιακριτή προσέγγιση (multiresolution approximation) της αρχικής εικόνας, έχοντας στη διάθεσή μας υποεικόνες διαδοχικά μειωμένων διαστάσεων.

Κάθε διαδικασία υποβιβασμού της διακριτικής ανάλυσης εισάγει απώλεια πληροφορίας. Είναι επιθυμητό να γίνει ελαχιστοποίηση της απώλειας. Ένα προσαρμοζόμενο σχήμα το οποίο σε κάθε βήμα της παραπάνω διαδικασίας διαλέγει την εικόνα με το μεγαλύτερο πόσο πληροφορίας αντί του πάντα επιλεγόμενου βαθυπερατού παρουσιάζεται στο [24]. Τα



φίλτρα που σχεδιάζονται με αυτήν την μέθοδο δημιουργούν εικόνες αποσυσχετισμένες μεταξύ τους κατάλληλες για κατηγοριοποίηση εικόνων σε χαμηλότερες αναλύσεις.

### 4.3 Πιθανοτικά Δίκτυα

Πολλά τεχνητά νευρωνικά δίκτυα στην έξοδό τους δίνουν τιμές οι οποίες μπορεί να θεωρηθεί ότι προσεγγίζουν Bayes *a posteriori* τιμές πιθανοτήτων. Όταν οι εκτιμήσεις είναι ακριβείς και αθροίζονται στη μονάδα είναι δυνατόν να χρησιμοποιηθούν ως πιθανότητες. Οι έξοδοι πολυεπίπεδων perceptron μπορεί να θεωρηθεί ότι εκτιμούν πιθανότητες Bayes όταν οι έξοδοι των νευρωνικού δικτύου αντιστοιχούν σε  $M$  διαφορετικά σύνολα, για τα οποία μία έξοδος του νευρωνικού δικτύου είναι 1 και οι άλλες 0, ενώ η συνάρτηση κόστους που ελαχιστοποιείται από το νευρωνικό δίκτυο είναι μέσου τετραγωνικού σφάλματος ή συνάρτηση εντροπίας [14]. Αλγόριθμοι εκμάθησης με δάσκαλο βασιζόμενοι στο κριτήριο Kullback Leibler (KL) [16, 14], με γενικευμένες σιγμοειδείς συναρτήσεις έχουν αποδειχθεί ικανοί να παράγουν Bayesian *a posteriori* πιθανότητες ή εκτιμήσεις πιθανοφάνειας οι οποίες μπορούν να χρησιμοποιηθούν τελικά σε εφαρμογές κατηγοριοποίησης. Η ακρίβεια της εκτίμησης των πιθανοτήτων εξαρτάται από: α) την πολυπλοκότητα του νευρωνικού δικτύου, β) το σύνολο εκμάθησης και γ) το βαθμό στον οποίο τα δεδομένα εκμάθησης αντιπροσωπεύουν κατανομές πιθανοφάνειας ή *a priori* πιθανότητες του κάθε στοιχείου να ανήκει σε ένα από τα υποψήφια σύνολα. Αντίθετα από την περίπτωση πολυεπίπεδων perceptron που ταξινομούν τα διανύσματα εισόδου μαθαίνοντας πολυδιάστατες επιφάνειες αποφάσεων τα πιθανοτικά νευρωνικά δίκτυα (ΠΝΔ) [21] ταξινομούν τα διανύσματα εισόδου εκτιμώντας μία κατανομή πυκνότητας πιθανότητας (κ.π.π) χωρίς δεδομένες τιμές παραμέτρων. Η δομή του δικτύου είναι παρόμοια με αυτήν του πολυεπίπεδου perceptron με την διαφορά ότι η σιγμοειδής συνάρτηση ενεργοποίησης έχει αντικατασταθεί από μία εκθετική. Σημαντικά πλεονέκτημα του ΠΝΔ είναι ότι η διαδικασία εκμάθησης διαρκεί μόνον ένα κύκλο και ότι οι επιφάνειες απόφασης προσεγγίζουν την καλύτερη δυνατή επιφάνεια απόφασης κατά Bayes καθώς ο αριθμός των δεδομένων εκμάθησης μεγαλώνει. Τα ΠΝΔ βασίζονται στην ιδέα ότι στο όριο κάθε συνεχής και ομαλή συνάρτηση πυκνότητας

πιθανότητας π.χ  $f_A$  ενός συνόλου διανυσμάτων  $\mathbf{X}$  μπορεί να προσεγγιστεί από το άθροισμα γκαουσιανών διανυσματικών κατανομών με κέντρο το κάθε διάνυσμα εκπαίδευσης [19]. Σε αυτήν την περίπτωση η  $f_A(\mathbf{X})$  θα είναι ίση με:

$$\frac{1}{(2\pi)^{p/2}\sigma^p} \frac{1}{m} \sum_{i=1}^m \exp[-(\mathbf{X} - \mathbf{X}_{\alpha i})^t(\mathbf{X} - \mathbf{X}_{\alpha i})/2\sigma^2] \quad (4.2)$$

όπου

$i$  = αύξων αριθμός των διανυσμάτων εκμάθησης,

$m$  = συνολικός αριθμός των διανυσμάτων εκμάθησης,

$\mathbf{X}_{\alpha i}$  =  $i$ -οστό διάνυσμα εκμάθησης από την κατηγορία  $A$ ,

$p$  = διάσταση του διανύσματος εισόδου  $\mathbf{X}$ ,

$\sigma$  = “παράγοντας εξομάλυνσης” (τυπική απόκλιση).

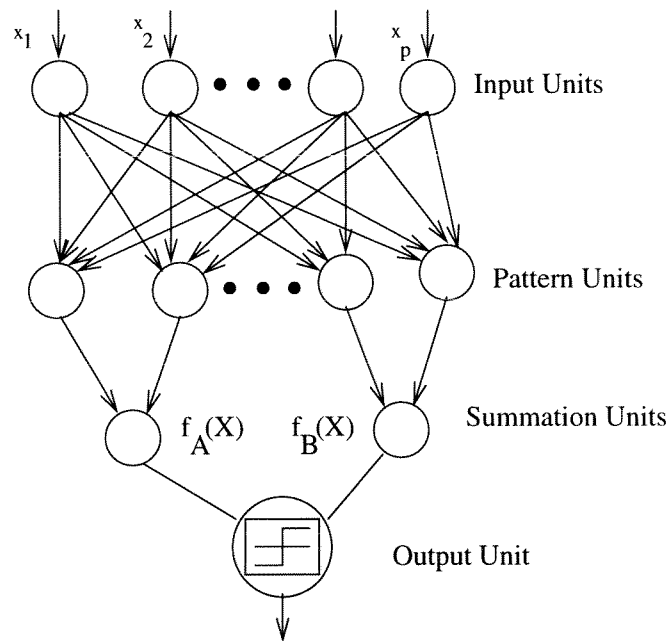
Το σχήμα 4.2 δείχνει ένα νευρωνικό δίκτυο το οποίο μπορεί να χρησιμοποιηθεί για την εκτίμηση μίας κ.π.π ενός συνόλου διανυσμάτων  $\mathbf{X}$  και την μετέπειτα κατηγοριοποίηση αυτού σε δύο κατηγορίες  $A$  και  $B$ . Οι μονάδες εισόδου είναι μονάδες τροφοδοσίας οι οποίες τροφοδοτούν με  $p$  στοιχεία όλες τις μονάδες χαρακτηριστικών (pattern units). Ο αριθμός των μονάδων χαρακτηριστικών είναι ίσος με τον αριθμό των διανυσμάτων εκπαίδευσης  $m$ . Η  $i$  μονάδα χαρακτηριστικού υπολογίζει το εσωτερικό γινόμενο του διανύσματος εισόδου και του διανύσματος των βαρών διασύνδεσης  $\mathbf{W}_i$ . Επιπλέον τροφοδοτεί το αποτέλεσμα στην εκθετική συνάρτηση ενεργοποίησης όπως φαίνεται στην συνέχεια:

$$\exp[(\mathbf{W}_i \mathbf{X}^t - 1)/\sigma^2] \quad (4.3)$$

Υποθέτοντας ότι τόσο το διάνυσμα  $\mathbf{X}$  όσο και το  $\mathbf{W}_i$  είναι κανονικοποιημένα στη μονάδα, η έξοδος της  $i$  μονάδας χαρακτηριστικού είναι αντίστοιχη με εκτίμηση πιθανότητας:

$$p_i(\mathbf{X}) = \exp[-(\mathbf{X} - \mathbf{W}_i)^t(\mathbf{X} - \mathbf{W}_i)/2\sigma^2] \quad (4.4)$$

Οι μονάδες άθροισης υπολογίζουν μέσω της σχέσης (4.4) τις κ.π.π κάθε συνόλου με τρόπο αντίστοιχο της σχέσης (4.2). Οι μονάδες εξόδου υπολογίζουν τη μεγαλύτερη από τις κ.π.π του κάθε συνόλου. Το δίκτυο εκπαιδεύεται θέτοντας το διάνυσμα  $\mathbf{W}_i$  της  $i$ -οστής



Σχήμα 4.2: Πιθανοτικό Νευρωνικό Δίκτυο για κατηγοριοποίηση αντικειμένων σε δύο κατηγορίες

μονάδας χαρακτηριστικών ίσο με το αντίστοιχο διάνυσμα  $\mathbf{X}$  του συνόλου εκμάθησης και συνδέοντας την μονάδα χαρακτηριστικού στην κατάλληλη μονάδα άθροισης. Η ποσότητα εξομάλυνσης  $\sigma$  ελέγχει τη συνάρτηση ενεργοποίησης: μεγάλες (μικρές) τιμές του  $\sigma$  μειώνουν (αυξάνουν) την ευαισθησία της εκθετικής συνάρτησης.

Στην συνέχεια θα προσπαθήσουμε να εισάγουμε μια αρχιτεκτονική που εκμεταλλεύεται την πολυδιακριτή ανάλυση που περιγράφηκε στην ενότητα 4.2.

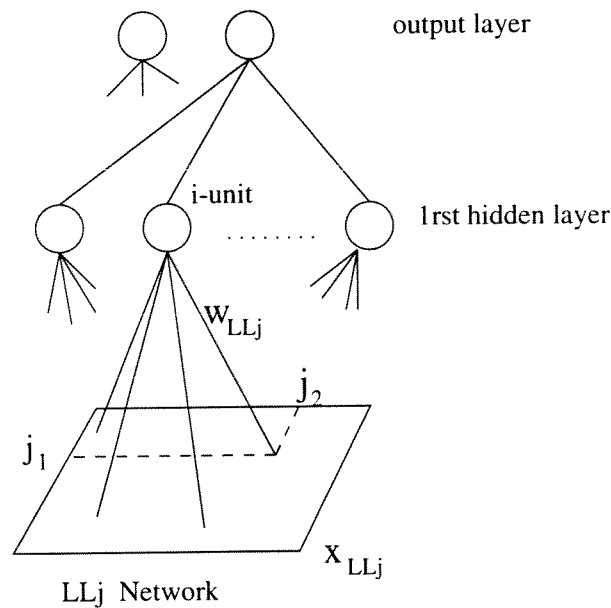
#### 4.4 Νευρωνικά δίκτυα για την πολυδιακριτή κατηγοριοποίηση εικόνων

Τα πιθανοτικά δίκτυα σε αντιδιαστολή με τα πολυεπίπεδα perceptron παρουσιάζουν μικρούς χρόνους εκπαίδευσης λόγω του ενός κύκλου εκμάθησης. Το βασικό μειονέκτημα τους είναι οι ανάγκες σε μνήμη, επειδή χρειάζεται να αποθηκεύσουν ολόκληρο το σύνολο

εκμάθησης στις μονάδες χαρακτηριστικών. Εάν το σύνολο εκμάθησης είναι εικόνες αποχρώσεων του γκριζου ή έγχρωμες εικόνες τότε οι ανάγκες σε μνήμη αυξάνονται σε πολύ μεγάλο βαθμό η δε ικανότητα γενίκευσης μειώνεται σε σημαντικό βαθμό. Ως εκ τούτου είναι επιθυμητό να περιορίσουμε τις απαιτήσεις σε μνήμη με όσο το δυνατόν μικρότερη επιβάρυνση των επιδόσεων του πιθανοτικού νευρωνικού δικτύου. Η πολυδιακριτή ανάλυση μπορεί να βοηθήσει σε αυτήν την κατεύθυνση. Εάν αντί της αρχικής εικόνας διαλέξουμε μία εικόνας μικρότερης διακριτότητας η οποία έχει δημιουργηθεί με τον “βέλτιστο τρόπο” [24] ώστε να εισάγει την ελάχιστη απώλεια πληροφορίας είναι δυνατόν να έχουμε ικανοποιητική απόδοση του πιθανοτικού δικτύου. Στην συνέχεια θα δώσουμε μία μεθοδολογία για την εκπαίδευση ενός feedforward νευρωνικού δικτύου και θα εισάγουμε το ιεραρχικό πιθανοτικό δίκτυο.

Ας θεωρήσουμε ότι ένα feedforward νευρωνικό δίκτυο χρησιμοποιείται για την κατηγοριοποίηση της αρχικής εικόνας  $x_0$ . Στην περίπτωση αυτή η εκμάθηση του δικτύου μπορεί να υλοποιηθεί σε εικόνες διαδοχικών, μικρότερης διακριτότητας, επιπέδων αρχίζοντας την διαδικασία από ένα συγκεκριμένο, μικρότερης διακριτότητας, επίπεδο όπου το πρόβλημα περιγράφεται ικανοποιητικά (με άλλα λόγια οι υποεικόνες του συγκεκριμένου επιπέδου περιέχουν αρκετή πληροφορία για την κατηγοριοποίηση). Η διαδικασία εκμάθησης που προτείνουμε χρησιμοποιεί αρχικά ένα μάλλον απλό πλήρως διασυνδεδεμένο feedforward νευρωνικό δίκτυο, με ένα κρυφό επίπεδο, για να κατηγοριοποιήσει μία αρκετά μικρής διακριτότητας παράσταση της αρχικής εικόνας και ακολούθως αναδρομικά κατασκευάζει την αρχιτεκτονική νέων δικτύων ικανών να χειριστούν τις εικόνες υψηλότερων διακριτικών επιπέδων.

Αναλυτικότερα, η προτεινόμενη διαδικασία κατηγοριοποίησης ξεκινά με την εκμάθηση ενός δικτύου στο, έστω, διακριτικό επίπεδο  $j$  με  $j \leq -1$ , (δίκτυο  $LL_j$ ) για την κατηγοριοποίηση των προσεγγιστικών εικόνων  $x_j^{LL}$  του συγκεκριμένου επιπέδου. Το δίκτυο αυτό παρουσιάζεται στο σχήμα 4.3. Μετά την εκμάθηση, η αποτελεσματικότητα του δικτύου εξετάζεται με την χρησιμοποίηση για κατηγοριοποίηση από αυτό ενός συνόλου προσεγγιστικών εικόνων του συγκεκριμένου επιπέδου  $j$ . Αν η απόδοση του δικτύου δεν είναι σε ικανοποιητικό επίπεδο η διαδικασία εκμάθησης επαναλαμβάνεται στο αμέσως επόμενο



Σχήμα 4.3: Το νευρωνικό δίκτυο του επιπέδου  $j$ , για ταξινόμηση των προσεγγιστικών εικόνων του επιπέδου αυτού.

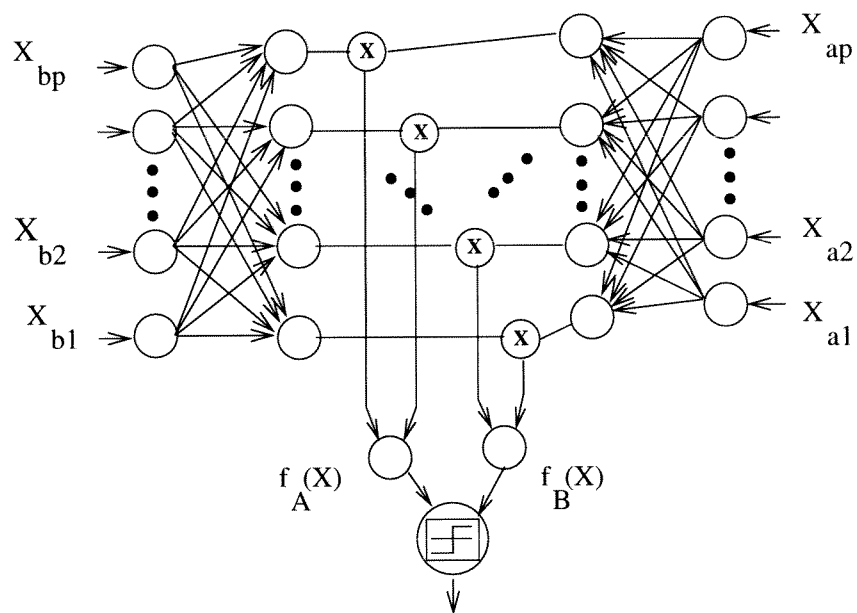
μεγαλύτερο διακριτικό επίπεδο, έστω  $j + 1$ . Δεδομένου ότι η πληροφορία της προσεγγιστικής εικόνας του επιπέδου  $j + 1$  ισοδυναμεί με την πληροφορία που έχουν η προσεγγιστική και οι τρεις εικόνες λεπτομέρειας του επιπέδου  $j$ , μπορούμε να εκπαιδεύσουμε άλλα τρία δίκτυα ( $LH_j, HL_j, HH_j$ ), παράλληλα με το προηγούμενο δίκτυο του σχήματος 4.3, στα οποία ανατίθεται η κατηγοριοποίηση των εικόνων λεπτομέρειας του επιπέδου  $j$ . Θα ήταν ακολούθως επιθυμητό, το νευρωνικό δίκτυο του επιπέδου  $j + 1$  (δίκτυο  $LL_{j+1}$ ) να εμπεριέχει α-priori στη δομή του την περισσότερη δυνατή πληροφορία για το πρόβλημα κατηγοριοποίησης που ήδη έχουν συγκεντρώσει τα προηγούμενα δίκτυα του επιπέδου  $j$  (δίκτυα  $LL_j, LH_j, HL_j, HH_j$ ) κατά την εκμάθησή τους.

Στην περίπτωση του πιθανοτικού δικτύου μπορούμε να υποθέσουμε ότι η αρχική εικόνα  $\mathbf{X}$  σε κάποια διακριτική ανάλυση  $j+1$  χωρίζεται σε δύο μέρη  $\mathbf{X} = [\mathbf{X}_a \mathbf{X}_b]$ . το πρώτο είναι η προσεγγιστική εικόνα  $\mathbf{x}_j^{L}$  ενώ το δεύτερο είναι οι υπόλοιπες τρεις εικόνες λεπτομέρειας  $\mathbf{x}_j^{HL}, \mathbf{x}_j^{HL}, \mathbf{x}_j^{HH}$  μικρότερης ανάλυσης. Αντίστοιχα και τα βάρη διασύνδεσης μπορούν να χωρισθούν με όμοιο τρόπο  $\mathbf{W}_i = [\mathbf{W}_{ai} \mathbf{W}_{bi}]$ . Σε αυτήν την περίπτωση η έξοδος της  $i$ -οστής χαρακτηριστικής μονάδας μπορεί να τροποποιηθεί ως εξής:

$$\begin{aligned}
 p_i(\mathbf{X}) &= \exp[(\mathbf{W}_i \mathbf{X}^t - 1)/\sigma^2] \\
 &= \exp[(\mathbf{W}_{ai} \ \mathbf{W}_{bi})[\mathbf{X}_{ai} \ \mathbf{X}_{bi}]^t - 1)/\sigma^2] \\
 &= k \cdot \exp[(\mathbf{W}_{ai} \mathbf{X}_{ai} - 1) + (\mathbf{W}_{bi} \mathbf{X}_{bi} - 1)]/\sigma^2 \\
 &= k \cdot p_{ai}(\mathbf{X}) \cdot p_{bi}(\mathbf{X})
 \end{aligned}
 \tag{4.5}$$

όπου  $k$  συντελεστής ίσος με  $\exp(1/\sigma^2)$ .

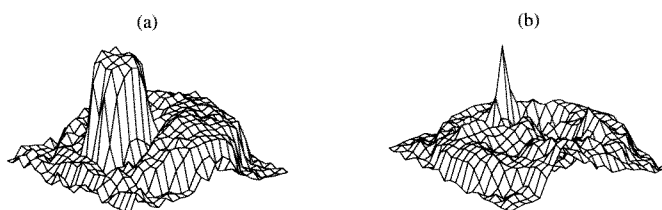
Αρα εάν θέλουμε να συνδυάσουμε δύο δίκτυα τα οποία εκπαιδεύτηκαν να κατηγοριοποιούν προσεγγιστικές και εικόνες λεπτομέρειες σε επίπεδο  $j$  αρκεί να πολλαπλασιάσουμε τις εξόδους των αντιστοιχών χαρακτηριστικών μονάδων και στην συνέχεια να τροφοδοτήσουμε κατάλληλα τις μονάδες άθροισης. Αυτό φαίνεται στο σχήμα 4.4.



Σχήμα 4.4: Πιθανοτικό νευρωνικό δίκτυο

## 4.5 Αποτελέσματα Προσομιώσεων - Συγκριτικές μελέτες

Μία εφαρμογή που αντιστοιχεί σε ένα πραγματικό πρόβλημα χρησιμοποιείται στην ενότητα αυτή για να εξετασθεί η αποτελεσματικότητα της προτεινόμενης διαδικασίας πολυδιακριτής κατηγοριοποίησης, εφαρμογή κατά την οποία ελέγχονται οι μεταλλικές συνδέσεις στην κατασκευή τυπωμένων πλακετών κυκλωμάτων. Διδιάστατες gray scale εικόνες ( $23 \times 23$  pixels), που παριστάνουν το πλάτος ή την ένταση σαν συναρτήσεις της θέσης στις μεταλλικές συνδέσεις, ελήφθησαν από έναν οπτικό laser scanner και χρησιμοποιήθηκαν ως σήματα για κατηγοριοποίηση σε δύο κατηγορίες: α) “καλές” και β) “ανεπαρκείς” συνδέσεις, όπου η δεύτερη περιέχει ανεπαρκή ποσότητα μετάλλου. Μία εικόνα-πρότυπο καλής και μία ανεπαρκούς σύνδεσης παρουσιάζονται στο σχήμα 4.5:



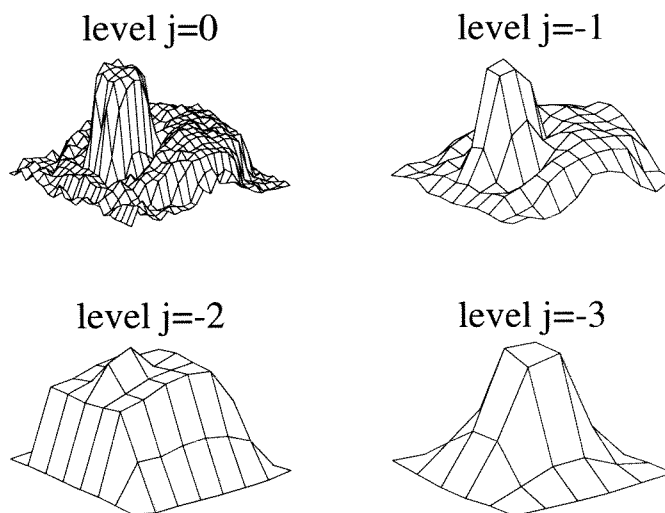
Σχήμα 4.5: Εικόνες καλής (a) και ελαττωματικής (b) κόλλησης -  $23 \times 23$  pixels.

Μία βάση δεδομένων 50 καλών και 50 ανεπαρκών εικόνων-προτύπων χρησιμοποιήθηκε στα επόμενα πειράματα. 25 καλά και 25 ανεπαρκή πρότυπα διαλέχτηκαν για να σχηματίσουν το σύνολο εκμάθησης ενώ τα υπόλοιπα σχημάτισαν το σύνολο ελέγχου.

Αρχικά χρησιμοποιήθηκε ένα πιθανοτικό νευρωνικό δίκτυο αποτελούμενο από  $(23)^2 = 529$

μονάδες εισόδου και 100 μονάδες χαρακτηριστικών. Το δίκτυο εκπαιδεύτηκε χρησιμοποιώντας το σύνολο εκμάθησης των αρχικών  $23 \times 23$  εικόνων. Ο μεγάλος αριθμός των βαρών του δικτύου σε σχέση με τον μικρό αριθμό των προτύπων εκμάθησης, οδήγησε σε ανεπιτυχή εκμάθηση. Το δίκτυο επέδειξε ποσοστό επιτυχούς ταξινόμησης ίσο με 79 %.

Στη συνέχεια, χρησιμοποιήσαμε τα FIR 8-tap QMF φίλτρα  $h_L(n)=[0.0094,-0.0707,0.0694,.49,0.49,0.0694,-0.0707,0.0094]$ , και  $h_H(n) = (-1)^{1-n}h_L(1-n)$  με αντίστοιχη υποδειγματοληψία κατά 2 λαμβάνοντας τις υποεικόνες σε κάθε μικρότερο διακριτικό επίπεδο (βλέπε σχήμα 4.1). Η καλή κόλληση του σχήματος 4.5 και οι προσεγγιστικές εικόνες σε τρία διαδοχικά επίπεδα μικρότερης διακριτότητας παρουσιάζονται στο σχήμα 4.6.

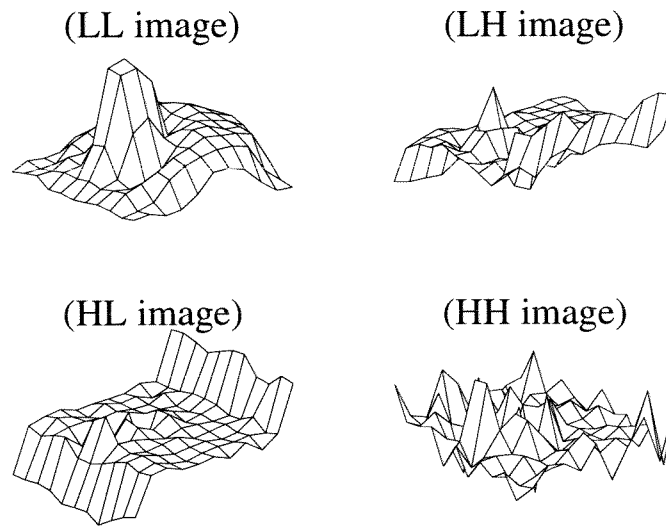


Σχήμα 4.6: Αρχική και προσεγγιστικές εικόνες σε τρία διαδοχικά μικρότερα διακριτικά επίπεδα.

Παρατηρούμε ότι οι λεπτομέρειες της ήδη μικρών διαστάσεων αρχικής εικόνας διακρίνονται ικανοποιητικά μόνο στο επίπεδο  $j = -1$ . Για το λόγο αυτό προσπαθήσαμε να εφαρμόσουμε την κατηγοριοποίηση αρχικά στο επίπεδο αυτό. Στο σχήμα 4.7 δείχνουμε την προσεγγιστική και τις τρεις εικόνες λεπτομερειών ( $11 \times 11$  pixels) στο επίπεδο  $j = -1$ , που αντιστοιχούν πάλι στην κόλληση του σχήματος 4.5.

Στη συνέχεια χρησιμοποιήσαμε δύο πιθανοτικά δίκτυα της μορφής του σχήματος 4.3, για να κατηγοριοποιήσουν ανεξάρτητα τις προσεγγιστικές ( $LL$ ) εικόνες και τις εικόνες





Σχήμα 4.7: Προσεγγιστική (LL) και εικόνες λεπτομέρειας (LH,HL,HH) για την εικόνα του Σχήματος 6.4(α) (11x11 pixels).

λεπτομέρειας  $LH_{-1}, HL_{-1}, HH_{-1}$  αυτού του διακριτικού επιπέδου. Τα δίκτυα έχουν 121 και  $3 \times 121$  μονάδες εισόδου αντίστοιχα, δεδομένου ότι οι διαστάσεις των εικόνων στο συγκεκριμένο επίπεδο  $j = -1$  είναι  $11 \times 11$  pixels. Το δίκτυο  $LL_{-1}$  κατηγοριοποίησε το αντίστοιχο σύνολο ελέγχου με επιτυχία 86.6%. Η αντίστοιχη επιτυχία του δικτύου  $LH_{-1}, HL_{-1}, HH_{-1}$  ήταν 66.6%.

Η διαφορά αυτή εξηγείται από το γεγονός ότι οι προσεγγιστικές εικόνες περιέχουν μεγαλύτερη πληροφορία από αυτή των εικόνων λεπτομέρειας στο επίπεδο αυτό. Ακολούθως, σχηματίσαμε το δίκτυο του επόμενου διακριτικού επιπέδου  $j = 0$ , συνδυάζοντας τα δύο προηγούμενα δίκτυα με βάση την σχέση (4.6). Το δίκτυο δεν χρειάστηκε περαιτέρω εκπαίδευση και τελικά κατηγοριοποίησε τις εικόνες-πρότυπα του συνόλου ελέγχου με επιτυχία 94 %. Διαφορετικοί συντελεστές χρησιμοποιήθηκαν για τα πιθανοτικά δίκτυα,  $\sigma_1 = 0.0001$  και  $\sigma_2 = 0.01$  για το πρώτο και δεύτερο αντίστοιχα. Η μικρότερη τιμή του δεύτερου συντελεστή οφείλεται το γεγονός ότι κατά τον έλεγχο των δικτύων σε ανεξάρτητη λειτουργία το πρώτο πέτυχε μεγαλύτερο ποσοστό επιτυχίας σε σχέση με το δεύτερο.

Στη συνέχεια παραθέτουμε συγκριτικά αποτελέσματα από την εργασία [24] στην οποία

παρουσιάζεται μια αρχιτεκτονική που βασίζεται στο πολυεπιπεδικό perceptron για να μειώσει τον αριθμό των βαρών κατά την διάρκεια της εκμάθησης. Στην εν λόγω εργασία παρουσιάστηκαν ειδικές σχέσεις για την “μεταφορά της γνώσης” από τα μικρότερα στα μεγαλύτερα διακριτικά επίπεδα οι οποίες οδηγούν στην μείωση του χρόνου εκμάθησης και στην αύξηση των επιδόσεων στα σύνολα ελέγχου.

Διακριτικό επίπεδο	Εικόνα	Πιθανοτικό δίκτυο, %	Multires. Perc. %
$j = 0$	$X[i][j]$	79	-
$j = -1$	$LL_{-1}$	86	80
	$HH_{-1}$	66.7	67
$j = 0$	$LL_{-11} + HH_{-1}$	94	91

Πίνακας 4.1: Συγκριτικά πειραματικά αποτελέσματα του πολυδιακριτού σχήματος κατηγοριοποίησης.

Επιπλέον η χρήση βέλτιστων φίλτρων ανακατασκευής στην εργασία [24] δίνει αποτελέσματα καλύτερα κατά 1 - 2 %.

## 4.6 Συμπεράσματα

Σε αυτό το κεφάλαιο παρουσιάστηκαν εισαγωγικά στοιχεία της πολυδιακριτής ανάλυσης και του πιθανοτικού δικτύου. Προτάθηκε τροποποίηση του πιθανοτικού δικτύου για λειτουργία σε πολυδιακριτό περιβάλλον. Δείχθηκαν τα πλεονεκτήματα της λύσης και έγιναν συγκρίσεις με παρόμοια εργασία. Η σύγκριση με τα πιθανοτικά δίκτυα αναδύει τα σημαντικά στοιχεία τους: απλότητα υλοποίησης, εύκολη προσαρμογή σε πολυδιακριτό περιβάλλον και αυξημένη ταχύτητα εκμάθησης.

Βασία η αμβροσία ΗΗ

## Αναφορές

- [1] ANNIE, “ Applications of Neural Networks to Industrial Pattern Recognition problems”, *Esprit Project 2092 Handbook*, 1991.
- [2] M. Morrison and Y. Attikiouzel, “A Probabilistic Neural Network Based Image Segmentation Network for Magnetic Resonance Images”, *IJCNN 92*, Portland, 1992.
- [3] I. Daubechies, “Orthonormal bases of compactly supported wavelets”, *Comm. Pure Appl. Math.*, vol.41, pp.909-996, 1988.
- [4] I. Daubechies et al, “Image coding using the wavelet transform”, *IEEE Trans. on Image Processing*, vol. 1, pp. 205-220, 1992.
- [5] A. Delopoulos, A. Tirakis and S. Kollias, “Invariant Image Classification using triple-correlation Neural Networks”, *IEEE Trans. on Neural Networks*, May 1994.
- [6] R. Fisher, “From surfaces to objects, computer vision and three dimensional analysis”, *Wiley*, 1989.
- [7] J. Hwang, E.Y. Chen and A.F. Lippman, “Probabilistic Image Modelling via neural Networks” *ICASSP 93* Minneapolis, Minnesota.
- [8] C. Hand et al, “A neural-network feature detector using a multiresolution pyramid”, in R. Linngard, ed., *Neural networks for vision. speech and natural language*. Chappman and Hall, 1992.
- [9] D. R. Hush and B.G. Horne, “Progres in Supervised Neural Networks”. *IEEE Signal Processing Magazine*, pp. 8-39, Jan 1993.

- [10] I. Ng, J. Kittler and J. Illingworth, "Supervised segmentation using a multiresolution data representation", *Signal Processing*, vol. 31, pp. 133-164, 1993.
- [11] S. Kollias and D. Kalogeras, "A Multiresolution Probabilistic Neural Network for Image Segmentation", *ICASSP-94*, Adelaid, 1994.
- [12] S. Kollias and D. Anastassiou, "An adaptive least squares algorithm for the efficient training of multilayered networks", *IEEE Trans. on CAS*, vol. 36, pp. 1092-1101, 1989.
- [13] S. Kollias, A. Stafylopatis and A. Tirakis, "Performance of Higher Order Neural Networks in Invariant Recognition", *Neural Networks: Advances and Applications*, pp. 79-108, North Holland, 1991.
- [14] R. P. Lippmann, "Neural Network Classifiers Estimate Bayesian a Posteriori Probabilities", *Neural Computation*, vol. 3 pp. 461 - 483, 1992.
- [15] Y. LeCun, "Generalization and Network Design Strategies", *Connectionism in Perspective*, North Holland, Switzerland, pp. 143 - 155, 1989.
- [16] J. Makhoul, "Pattern Recognition Properties of Neural Networks", *IEEE-SP Workshop on Neural Networks for Signal Processing*, pp. 173-186, Princeton NJ, 1991.
- [17] S. Mallat, "A theory for multiresolution signal decomposition: the wavelet representation", *IEEE Trans. on PAMI*, vol.11, pp. 674-693, 1989.
- [18] J. Moody and C. Darken, "Fast learning in networks of locally-tuned processing elements", *Neural Computation*, vol. 1, pp. 281-293, 1989.
- [19] E. Parzen "On estimation of a probability density faction and mode", *Ann. math Stat.*, vol. 33, pp. 1063-1076, 1965.
- [20] Y. Pao, "Adaptive pattern recognition and neural networks", *Addison Wesley Co.*, 1989.

- [21] D. F. Specht, "Probabilistic Neural Networks and the polynomial Adaline as Complementary Techniques for Classification", *IEEE Trans. On Neural Networks*, vol. 1, no. 1, Mar. 1990.
- [22] P.P. Vaidyanathan, "Multirate Digital Filters, Filter Banks, Polyphase Networks, and Applications: A Tutorial", *Proceedings of IEEE*, Vol.78, No.1, January 1990.
- [23] S. Tanimoto and T. Pavlidis, "A hierarchical data structure for picture processing", *CGIP 4*, pp. 104-109, 1975.
- [24] Α. Τυράκης "Αναλοιώτη και Πολυδιακριτική Παράσταση Διδιάστατων σημάτων - Εφαρμογές στην Ανάλυση Εικόνων με Νευρωνικά Δίκτυα", *Διδακτορική Διατριβή*, Κεφ. 5, Οκτ. 1993
- [25] Α. Tirakis Α. Delopoulos and S. Kollias, "2-D Filter Bank Design For Optimal Subband Reconstruction", *IEEE Trans on Image Processing*, to be published.
- [26] M. Vetterli, "Multi-dimensional Sub-band Coding: Some Theory and Algorithms", *Signal Processing*, Vol. 6, pp. 97-112, April 1984.
- [27] B. Widrow and R. Winter, "Neural Nets for adaptive filtering and adaptive pattern recognition", *IEEE Computer*, vol. 21, no. 3, pp. 25-39, 1988.
- [28] P. Woods et al. "The use of geometric and gray level models for industrial inspection", *Pattern Recognition Letters*, pp. 1-17, 1987.

## Μέρος II

# Κωδικοποίηση και Συμπύεση Κινούμενων Εικόνων

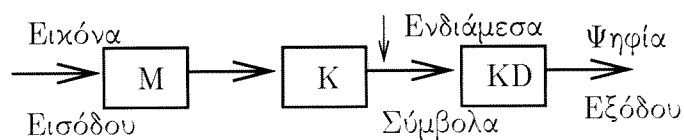
# Πρότυπα Κωδικοποίησης Κινούμενων Εικόνων

## 5.1 Εισαγωγή

Η ανάγκη για την συμπίεση εικόνων είναι προφανής σε όλες τις εφαρμογές όπου απαιτείται μετάδοση και αποθήκευση ψηφιακών εικόνων. Για παράδειγμα μία σελίδα χαρτιού A4 η οποία έχει σαρωθεί με ανάλυση 300 pixels/inch και 1 bit/pixel δημιουργεί δεδομένα όγκου 8.4 Mbits για τα οποία ο χρόνος μετάδοσης είναι 15 min μέσα από μία γραμμή 9600 baud. Παρόμοια, εάν ένα έγχρωμο φιλμ 35mm σαρωθεί με ανάλυση 12 micron, θα δημιουργηθεί μία ψηφιακή εικόνα διαστάσεων 3656 × 2664. Εάν στην συνέχεια χρησιμοποιηθεί δειγματοληψία 8 bit/pixel για κάθε χρωματική συνιστώσα δημιουργούνται δεδομένα όγκου 233 Mbits. Λαμβάνοντας υπόψη την αποθηκευτική ικανότητα ενός CD που είναι περίπου 5 Gbits, είναι δυνατή η αποθήκευση 600 μαυρόασπρων σελίδων ή 21 έγχρωμων εικόνων ψηφιοποιημένων από φιλμ των 35mm. Ειδικά για την περίπτωση ακίνητων εικόνων διάφορα διεθνή πρότυπα έχουν θεσπιστεί από την ITU (πρώην CCITT) όπως Group 3 και Group 4 [3, 4] και ISO/IEC/CCITT JPEG [6, 9]. Στην περίπτωση κινούμενης εικόνας video η ψηφιακή μετάδοση έγχρωμου video NTSC, διαστάσεων 480 γραμμών και 720 pixels με 8 bit/χρώμα συχνότητας, 30 πλαισίων/sec απαιτεί χωρητικότητα γραμμής 248 Mbit/sec. Παρόμοια, video ποιότητας HDTV διαστάσεων 3112 γραμμών και 4096 pixels με 8 bit/χρώμα συχνότητας, 30 πλαισίων/sec απαιτεί χωρητικότητα γραμμής 1.5

Gbit/sec. Ένα CD με τη χωρητικότητα που αναφέραμε προηγουμένως μπορεί να αποθηκεύσει 20 sec video NTSC και 3 sec video HDTV. Δεδομένου ότι η ταχύτητα ανάγνωσης δεδομένων από ένα CD είναι 1.5 Mbps (παρόλο που έχουν εμφανιστεί συσκευές με τετραπλάσιες δυνατότητες) είναι φανερό ότι απαιτείται συμπίεση της τάξης 200:1 για την περίπτωση video NTSC. Με αυτό το βαθμό συμπίεσης 3400 sec (περίπου 1 ώρα) video μπορούν να αποθηκευτούν σε ένα CD. Η διεθνής επιτροπή προτύπων αναγνωρίζοντας την ανάγκη για αλγόριθμους συμπίεσης video έφτιαξε ορισμένα πρότυπα: το H.261 [11] για μετάδοση εικόνας σε πολλαπλάσια των 64 kbit/sec, το MPEG-1 [22] για κωδικοποίηση στα 1.5 Mbits/sec και το MPEG-2 [23] για κωδικοποίηση στα 4 - 10 Mbit/sec.

Η συμπίεση ψηφιακών εικόνων είναι δυνατή επειδή: α) υπάρχει χωρική πλεονάζουσα πληροφορία λόγω της χωρικής συσχέτισης γειτονικών pixels β) υπάρχει συχνοτική πλεονάζουσα πληροφορία εξαιτίας της συσχέτισης μεταξύ των χρωματικών συνιστωσών και γ) υπάρχει ψυχοοπτική πλεονάζουσα πληροφορία εξαιτίας των φυσικών χαρακτηριστικών του ανθρώπινου οπτικού συστήματος. Τέλος (δ) στην περίπτωση κινούμενης εικόνας (video) επιπλέον συμπίεση μπορεί να επιτευχθεί εξαιτίας της χρονικής συσχέτισης γειτονικών εικόνων. Όσο μεγαλύτερη είναι η πλεονάζουσα πληροφορία τόσο μεγαλύτερη είναι η συμπίεση που μπορεί να υποστεί η αρχική πηγή της πληροφορίας. Η αφαίρεση της πλεονάζουσας πληροφορίας (στατιστικής υψής στις περιπτώσεις α) β) και δ) αναφέρεται ως κωδικοποίηση πηγής στην θεωρία πληροφορίας. Ένα τυπικό σύστημα κωδικοποιητή πηγής είναι σαν αυτό του σχήματος 5.1.



Σχήμα 5.1: Τυπικό σύστημα κωδικοποίησης πηγής

Η μονάδα M μετασχηματίζει τα δεδομένα εισόδου σε μία μορφή η οποία είναι ευνοϊκότερη για συμπίεση. Τυπικοί μετασχηματισμοί είναι: α) γραμμικοί μετασχηματισμοί πρόβλεψης (οι οποίοι μετασχηματίζουν τις αρχικές τιμές των pixels σε σφάλματα πρόβλεψης



καλύτερη δυνατή ποιότητα για δεδομένο λόγο συμπίεσης εκμεταλλευόμενα και την ψυχοοπτική πλεονάζουσα πληροφορία. Στη συμπίεση ακίνητων εικόνων έχουν εφαρμοστεί και οι δύο μέθοδοι συμπίεσης. Στην κωδικοποίηση κινούμενων εικόνων οι απαιτούμενοι βαθμοί συμπίεσης είναι μεγαλύτεροι από 200:1 και επιτυγχάνονται, μέχρι στιγμής, μόνο με αλγόριθμους που ανήκουν στη δεύτερη κατηγορία.

Η διαδικασία προτυποποίησης αλγόριθμων συμπίεσης ψηφιακού video για αποθήκευση και μετάδοση video χρειάστηκε πάνω από 10 χρόνια. Ψηφιακή τηλεόραση, τηλεσυνδιάσκεψη, video-τηλέφωνα, εφαρμογές πολυμέσων είναι περιπτώσεις εφαρμογών των προτύπων συμπίεσης ψηφιακού video. Σε αυτό το κεφάλαιο θα περιγράψουμε τα πρότυπα που έχουν υιοθετηθεί από την διεθνή επιτροπή προτύπων (ISO πρώην CCITT) για την μετάδοση video: το H.261 και το MPEG-1.

Η δομή του κεφαλαίου είναι η ακόλουθη: Στην Ενότητα 5.2 παρουσιάζεται το πρότυπο H.261 για τη μετάδοση video σε ταχύτητες  $p \times 64$  kbps. Το πρότυπο αυτό αν και δε γνώρισε μεγάλη επιτυχία εξαιτίας της μεγάλης οικονομικής δυσκολίας για εγκατάσταση σταθμών τηλεσυνδιάσκεψης, αποτέλεσε την βάση για τα επικείμενα πρότυπα. Σήμερα δεδομένου της υπολογιστικής ισχύος των προσωπικών υπολογιστών είναι δυνατή η υλοποίηση του προτύπου σε λογισμικό με την βοήθεια υλικού για την σύλληψη κινούμενων εικόνων στις απαιτούμενες διαστάσεις. Το πρότυπο H.261 Το πρότυπο MPEG-1 ως συνέχεια του H.261 παρουσιάζεται στην Ενότητα 5.3. Το πρότυπο MPEG-1 γνώρισε γρήγορα παγκόσμιο ενδιαφέρον. Πολλές εταιρίες σχετιζόμενες με την βιομηχανία του video υλοποίησαν πολύ γρήγορα το πρότυπο σε ειδικές κάρτες για κωδικοποίηση και αποκωδικοποίηση σε πραγματικό χρόνο. Ταυτόχρονα η ύπαρξη λογισμικού για συνεχή πειραματισμό σε ανοικτά θέματα υλοποίησης επιτάχυνε την αποδοχή του προτύπου από την επιστημονική κοινότητα. Σήμερα οι κάρτες οθόνης προσωπικών υπολογιστών είναι εφοδιασμένες με ψηφίδες (VLSI chip) για αποκωδικοποίηση MPEG-1 σε πραγματικό χρόνο διαστάσεων τηλεοπτικού σήματος έτσι ώστε η απόσταση τηλεοπτικών συσκευών και υπολογιστών να είναι πλέον ελάχιστη. Η επιτυχία του προτύπου MPEG-1 επηρέασε σημαντικά στην επέκτασή του προτύπου MPEG-2 το οποίο παρουσιάζεται στην Ενότητα 5.4. Η διεθνής επιτροπή ήθελε ένα πρότυπο του οποίου η ποιότητα να είναι όχι απλά συγκρίσιμη με την ποιότητα

του video στα τηλεοπτικά studio αλλά εφάμιλλη της ποιότητας της κινηματογραφικής εικόνας ώστε να είναι δελεαστική η αλλαγή μεγάλου εγκαταστημένου υλικού. Επιπλέον η δυνατότητα του προτύπου για αποκωδικοποίηση τηλεοπτικού σήματος υψηλής ευκρίνειας (HDTV) επέδρασε σημαντικά στην βιομηχανία κατασκευής τηλεοπτικών δεκτών. Η παρεχόμενη συμβατότητα προς τα πίσω (backward compatibility) με το MPEG-1 σε ορισμένα επίπεδα λειτουργίας (ανάλυση PAL / NTSC) δημιούργησε τεράστιο ενδιαφέρον στους κατασκευαστές αφού είναι δυνατόν τεχνολογία και υλικά να μεταφερθούν από το προηγούμενο πρότυπο.

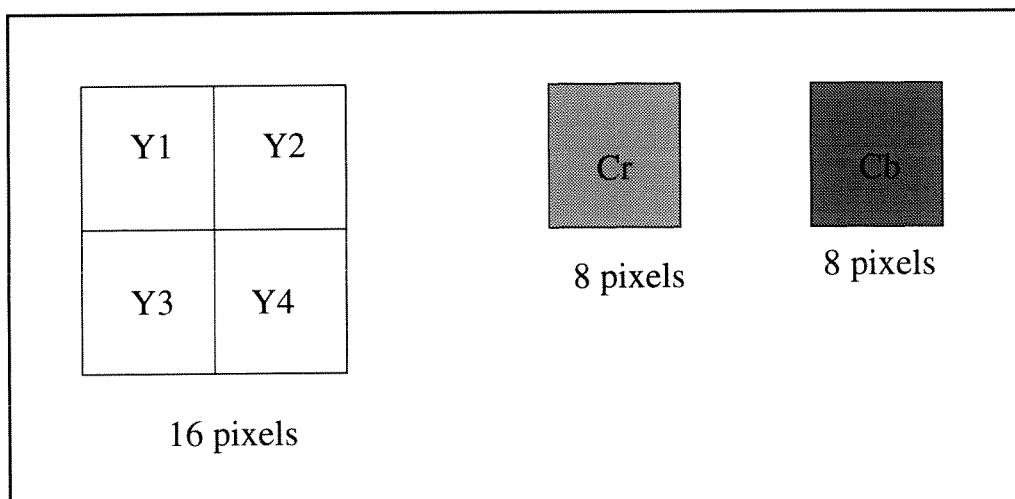
## 5.2 H.261

Το πρότυπο H.261 αναπτύχθηκε για να χρησιμοποιηθεί σε εφαρμογές τηλεσυνδιάσκεψης πάνω από δίκτυα ISDN σε ρυθμούς  $p \times 64$  kbps,  $p = 1, \dots, 30$ . Το πρότυπο εκτός του ότι αποτέλεσε τη βάση για τα πρότυπα MPEG-1/2 προσφέρει δύο σημαντικά χαρακτηριστικά: α) Προδιαγράφει την μέγιστη καθυστέρηση κωδικοποίησης στα 150 msec και β) η υλοποίηση του αλγορίθμου είναι εφικτή με VLSI chips. Στην προσπάθεια για ενιαίο τρόπο αντιμετώπισης των τηλεοπτικών σημάτων NTSC και PAL το πρότυπο προδιαγράφει μια τυπική εικόνα εισόδου (TEE) CIF ( Common Interchange Format). Προαιρετικά για εφαρμογές χαμηλού ρυθμού μετάδοσης ορίζεται το QCIF το οποίο έχει τις μισές διαστάσεις του CIF. Ο πίνακας 5.1 δείχνει τα χαρακτηριστικά των εικόνων CIF και QCIF, οι αριθμοί εντός των παρενθέσεων δείχνουν τις στρογγυλεμένες τιμές σε πολλαπλάσια του 8. Χαρακτηριστικά αναφέρουμε ότι για ρυθμούς 30 πλαίσια/sec ο ρυθμός μετάδοσης για ασυμπίεστα δεδομένα εισόδου είναι 37.3 Mbps και 9.35 Mbps για CIF και QCIF αντίστοιχα. Ακόμα και σε ρυθμούς των 10 πλαισίων/sec με εικόνες εισόδου QCIF απαιτείται λόγος συμπίεσης 48:1 για μετάδοση σε γραμμή 64kbps.

Το πρότυπο ορίζει ότι τα δεδομένα έχουν μια ιεραρχημένη δομή. Κάθε εικόνα αποτελείται από 12 GOB (group-of-blocks). Κάθε GOB αποτελείται από 33 MacroBlock (MB) το οποίο αποτελείται από υποεικόνες B (blocks)) διαστάσεων  $8 \times 8$ . Το macroblock αποτελεί το μικρότερο κομμάτι εικόνας στο οποίο μπορεί να διαφοροποιηθεί ο τρόπος συμπίεσης.

	CIF	QCIF
Αριθμός pixel / γραμμή		
Lum (Y)	360 (352)	180 (176)
Chroma (U,V)	180 (176)	90 (88)
Αριθμός γραμμών / εικόνα		
Lum (Y)	288	144
Chroma (U,V)	144	72
Interlacing	1:1	1:1
Πλαίσια Εικόνων / sec	30,15,10, 7.5	30,15,10, 7.5
Λόγος πλάτους/ύψος	4:3	4:3

Πίνακας 5.1: Τυπικές εικόνες εισόδου στο H.261



Σχήμα 5.2: Ένα MB του MPEG-1

Κάθε MB αποτελείται από τέσσερα Block φωτεινότητας Y και τα αντίστοιχα Block χρωματικότητας U και V, τα οποία έχουν προκύψει από υποδειγματοληψία τεσσάρων Block χρωματικότητας διαστάσεων ίδιων με αυτά της φωτεινότητας. Ένα MB φαίνεται στο σχήμα 5.2.

Κατά την διαδικασία προτυποποίησης έγιναν συγκριτικές μελέτες μεταξύ διάφορων αλγορίθμων συμπίεσης [1]. Το πρότυπο ορίζει ένα υβριδικό αλγόριθμο κωδικοποίησης τον Διακριτό Μετασχηματισμό Συνημίτονου με Αντιστάθμιση κίνησης (ΔΜΣ-ΑΚ). Ένα διάγραμμα του κωδικοποιητή και του αποκωδικοποιητή φαίνεται στο σχήμα 5.2.

Προηγούμενα κωδικοποιημένα πλαίσια εικόνων φυλάγονται στην μνήμη (Frame Store)

τόσο στον κωδικοποιητή όσο και στον αποκωδικοποιητή. Η μονάδα εκτίμησης κίνησης ανιχνεύει και υπολογίζει την κίνηση στην εικόνα. Υπολογίζεται ένα διάνυσμα κίνησης από το προηγούμενο πλαίσιο εικόνας με ακρίβεια 1 pixel για κάθε MB. Το πρότυπο δεν προδιαγράφει την μέθοδο, αλλά μπορεί να χρησιμοποιηθεί κάποια από τις γνωστές μεθόδους [2]. Η χρονική πρόβλεψη της τιμής κάθε pixel σε κάθε MB αφαιρείται από την τρέχουσα τιμή και προκύπτει το χρονικά αντισταθμισμένο σφάλμα πρόβλεψης. Ένας  $8 \times 8$  ΔΜΣ εφαρμόζεται σε κάθε Β σφαλμάτων εκτίμησης. Στη συνέχεια ακολουθεί ο κβαντιστής (Q) και κατόπιν ο κωδικοποιητής εντροπίας (VLC). Κωδικοποιούνται επιπρόσθετα στην περίπτωση μη μηδενικής κίνησης και τα διανύσματα κίνησης. Ένας απομονωτής (video buffer) βρίσκεται στην έξοδο των δεδομένων για να εγγυάται σταθερή ροή δεδομένων. Οι πίνακες κβαντισμού μπορούν να μεταβάλλονται σε κάθε MB ώστε να αποφεύγεται η υπερχείλιση ή το άδειασμα του απομονωτή. Σε αυτήν την περίπτωση κωδικοποιείται και το “ύψος του σκαλοπατιού” του κβαντιστή. Επιπλέον υπάρχει και ο τρόπος λειτουργίας όπου δεν κωδικοποιείται το σφάλμα πρόβλεψης (Interframe Coding) αλλά η αρχική εικόνα (Intraframe Coding). Η εναλλαγή του τρόπου λειτουργίας δεν είναι προκαθορισμένη και μπορεί να συμβαίνει για λόγους εξοικονόμησης εύρους ζώνης (Interframe Coding) ή για λόγους βελτίωσης της ποιότητας της παραγόμενης ακολουθίας εικόνων. Ο αποκωδικοποιητής χρησιμοποιεί την αντίστροφη διαδικασία.

### 5.3 MPEG-1

Το πρότυπο MPEG-1 φτιάχτηκε ως ένα γενικό πρότυπο για κωδικοποίηση και αποθήκευση σε ψηφιακά μέσα αποθήκευσης κινούμενων εικόνων και ήχου σε ρυθμούς γύρω στα 1.5 Mb/s. Οι εφαρμογές που μπορούν να χρησιμοποιήσουν το πρότυπο κυμαίνονται από διαλογικές εφαρμογές πάνω σε CD-ROM (CD-i) έως μετάδοση τηλεοπτικού σήματος σε τηλεπικοινωνιακά δίκτυα. Με αυτό τον τρόπο τα χαρακτηριστικά εισόδου μπορούν να μεταβάλλονται ανάλογα με τις απαιτήσεις της εφαρμογής. Επειδή το πρότυπο στόχευε κυρίως σε εφαρμογές πολυμέσων υπήρχε η απαίτηση να παρέχεται τυχαία προσπέλαση του video, γρήγορη κίνηση μπρός/πίσω, ανάποδη εκτέλεση του video, και δυνατότητα για

απομονωμένη επεξεργασία των πλαισίων εικόνας.

Οι απαιτήσεις για αποκωδικοποίηση τηλεοπτικού σήματος που έχει επιβάλει το πρότυπο ονομάζονται Σύνολο Ελάχιστων Προδιαγραφών (Constrained Parameters Set) το οποίο είναι:

- Μέγιστος αριθμός γραμμών/εικόνα: 576
- Μέγιστος αριθμός pixel/γραμμή: 720
- Μέγιστος αριθμός πλαισίων/sec: 30
- Μέγιστος αριθμός MB (macroblock)/εικόνα: 396
- Μέγιστος αριθμός MB/sec: 9900
- Μέγιστος ρυθμός: 1.86 Mbps
- Μέγιστο μέγεθος του απομονωτή του αποκωδικοποιητή: 376.832 bits.

Το σήμα εισόδου θεωρείται ότι είναι προοδευτικό και δεν αποτελείται από πεδία εικόνας (interlaced fields). Το πρότυπο, αντίστοιχα με το H.261, μετατρέπει τις εικόνες εισόδου σε εικόνες τύπου SIF (Standard Input Format) διαστάσεων  $352 \times 240$  στον κωδικοποιητή ώστε ο ρυθμός των 1.5 Mbps να είναι εφικτός. Για να μετατραπεί το τηλεοπτικό σήμα σε αυτές τις διαστάσεις τα περιττά πεδία εικόνας φιλτράρονται ενώ τα άρτια πετιούνται. Το χρωματικό σύστημα συντεταγμένων είναι το YCrCb με 8 bit/pixel και υποδειγματοληψία ανά 2 στις χρωματικές συνιστώσες στην οριζόντια και στην κατακόρυφη διεύθυνση. Το πρότυπο δεν προδιαγράφει με Block διάγραμμα τον αποκωδικοποιητή αλλά ορίζει ένα συντακτικό από το οποίο γίνεται φανερό ότι ο κωδικοποιητής είναι τύπου ΔΜΣ-ΑΚ όμοιος με αυτόν του H.261.

Το πρότυπο παρόμοια με το H.261 ορίζει μια ιεραρχημένη δομή 6 επιπέδων στα δεδομένα: Ακολουθία εικόνων, Group εικόνων, Εικόνες, Εικονολορίδα (Slice), Macroblock (MB), Block (B).

- Ακολουθία εικόνων. Μία ακολουθία εικόνων η οποία αποτελείται από Group Εικό-  
νων.

- Group Εικόνων. Μία εκ των προτέρων καθορισμένη αλληλουχία εικόνων όπως φαίνεται στο σχήμα 5.4
- Εικόνες. Το πρότυπο ορίζει I-εικόνες, P-εικόνες, B-εικόνες και D-εικόνες. Ο κωδικοποιητής κωδικοποιεί αυτούσιες I-εικόνες (Interframe Coding) σε τακτά χρονικά διαστήματα ώστε να ικανοποιούνται οι απαιτήσεις για τυχαία προσπέλαση video. Στον τρόπο λειτουργίας με αυτούσιες εικόνες η μέθοδος συμπίεσης είναι ΔΜΣ + κβαντιστής + κωδικοποιητής εντροπίας, αντίστοιχη του σχήματος 5.1. Επιπρόσθετα κωδικοποιούνται P-εικόνες (Predicted frames) όπου κωδικοποιείται το σφάλμα πρόβλεψης με χρονική αντιστάθμιση από πλαίσια εικόνων από το παρελθόν. Στις B-εικόνες κωδικοποιείται επίσης το σφάλμα πρόβλεψης, μόνο που η πρόβλεψη μπορεί να προέρχεται από μελλοντικά ή/και παρελθόντα πλαίσια εικόνων.
- Εικονολωρίδες. Οι εικονολωρίδες είναι ένα σύνολο από MB. Χρησιμοποιούν στην ανάνηψη από σφάλματα στή συμβολοσειρά του MPEG-1.
- MB. Τα MB ορίζονται αντίστοιχα με το H.261 ενώ επιπλέον ορίζονται κατηγορίες MB ανάλογα:
  - α) με τον τρόπο κωδικοποίησης της εικόνας στην οποία ανήκουν (Intra-MB ή Inter-MB); Intra-MB μπορούν να χρησιμοποιηθούν σε P-εικόνες ή B-εικόνες εάν ο αλγόριθμος αντιστάθμισης κίνησης αποτύχει.
  - β) με την κατηγορία και την τιμή του διανύσματος κίνησης που χρησιμοποιείται για την αντιστάθμιση κίνησης (Inter-Forward, Inter-Backward, Inter-(Forward+Backward) · Μηδενικά διανύσματα κίνησης δεν κωδικοποιούνται.
  - γ) τη τιμή του σφάλματος πρόβλεψης σε περίπτωση (Interframe Coding): σε MB με σχεδόν μηδενικό σφάλμα πρόβλεψης μεταδίδονται μόνο τα διανύσματα κίνησης.
  - δ) την ύπαρξη ελεγχτή προσαρμοζόμενου κβαντιστή · MB στα οποία αλλάζουν τα χαρακτηριστικά του κβαντιστή μεταδίδονται πάντα.
- Τα B είναι πίνακες συντελεστών ΔΜΣ διαστάσεων  $8 \times 8$ .

Το πρότυπο χρησιμοποιεί ομοιόμορφο κβαντιστή του οποίου το βήμα μεταβάλλεται ανάλογα με την συχνότητα του κάθε συντελεστή ΔΜΣ. Τα διανύσματα κίνησης (ΔΚ) υπολογίζονται γίνεται με ακρίβεια 1/2 pixel.

Συνοψίζοντας δίνουμε στον πίνακα 5.2 τις βασικές διαφορές μεταξύ των προτύπων H.261 και του MPEG-1.

H.261	MPEG-1
Σειριακή προσπέλαση	Τυχαία προσπέλαση
Ενας βασικός ρυθμός μετάδοσης hline Εικόνες CIF και QCIF	Μεταβλητός ρυθμός μετάδοσης Μεταβλητές διαστάσεις εικόνων
I-εικόνες και P-εικόνες	I, P, B εικόνες
Εκτίμηση με αναφορά 1 πλαίσιο εικόνας	Εκτίμηση με αναφορά σε 1 ή περισσότερα πλαίσια εικόνας
Ακρίβεια 1 pixel στα ΔΚ	Ακρίβεια 1/2 pixel στα ΔΚ
Φίλτρο Βρόγχου	-
Μεταβαλλόμενο κατώφλι + ομοιόμορφος κβαντιστής	Πίνακας κβαντισμού
-	Δομή εικονολωρίδας

Πίνακας 5.2: Διαφορές μεταξύ του H.261 και του MPEG-1

## 5.4 MPEG-2

Αν και το πρότυπο MPEG-1 μπορεί να χειριστεί video παρόμοιο με αυτό που χρησιμοποιείται στην εκπομπή τηλεοπτικών προγραμμάτων, η ποιότητα του αποκωδικοποιημένου video δεν είναι ικανοποιητική. Υποκειμενικά πειράματα έδειξαν ότι η συμπίεση εικόνων διαστάσεων ανάλογων του CCIR - 601 format είναι δυνατή σε ρυθμούς 4 - 6 Mbps. Το πρότυπο MPEG-2 προτάθηκε ως επέκταση του MPEG-1 για να ικανοποιήσει τις ανάγκες για συμπίεση σε διάφορους ρυθμούς μετάδοσης (από 2 - 20 Mbits) και σε ποικίλες αναλύσεις. Τα κύρια χαρακτηριστικά του είναι: α) επεξεργασία πεδίων και πλαισίων εικόνας, εναλλακτική δειγματοληψία των χρωματικών συνιστωσών, β) παραγωγή κλιμακωτής συμβολοσειράς εξόδου (scalable bitstream), γ) υλοποίηση προσαρμοζόμενου κβαντιστή μεγαλύτερης ακρίβειας από αυτόν του MPEG-1, δ) εναλλακτική μέθοδος σάρωσης του πίνακα του ΔΜΣ και ε) διαφορετικούς πίνακες εντροπίας για τους συντελεστές των MB

τύπου I.

Επειδή η υλοποίηση του πλήρους συντακτικού του MPEG-2 είναι προς στιγμή αδύνατη εξαιτίας της μεγάλης πολυπλοκότητάς του, έχουν οριστεί πέντε υποσύνολα λειτουργίας, (ΥΛ) (profiles), για να γίνει εφικτή η βιομηχανική υλοποίηση του σε μορφή ολοκληρωμένων κυκλωμάτων. Τα ΥΛ είναι: το Απλό, το Κύριο, το κλιμακωτό SNR (SNR scalable), το Χωροκλιμακωτό (spatial scalable) και το ανώτατο ΥΛ. Κατά τον ορισμό των ΥΛ ορίζονται και επίπεδα λειτουργίας (level) τα οποία επηρεάζουν τις τιμές μεταβλητών [23] του τρόπου λειτουργίας των ΥΛ.

Τα επιπρόσθετα χαρακτηριστικά και τα ΥΛ του MPEG-2 μπορούν να χωριστούν σε δύο κατηγορίες: α) Μη κλιμακούμενοι τρόποι λειτουργίας και β) κλιμακούμενοι τρόποι λειτουργίας.

### Μη κλιμακούμενοι τρόποι λειτουργίας

Στην παρούσα κατηγορία εντάσσονται όλες οι επεκτάσεις που έγιναν στο MPEG-2 ώστε αφενός να παρέχεται προς τα πίσω συμβατότητα με το MPEG-1 και αφετέρου να υπάρχει δυνατότητα για αντιμετώπιση των ειδικών χαρακτηριστικών του διαπλεκόμενου video.

*Πλαίσια και Πεδία Εικόνων:* Το πρότυπο MPEG-2 εισάγει την έννοια των πλαισίων και των πεδίων εικόνων καθώς και την πρόβλεψη πεδίων και πλαισίων εικόνων για να αντιμετωπισθούν οι ανάγκες για διαπλεκόμενο (interlaced) και προοδευτικό (progressive) video. Το διαπλεκόμενο video αποτελείται από πεδία εικόνας που περιέχουν τις άρτιες και περιττές γραμμές εναλλάξ. Τα πεδία μπορούν να κωδικοποιηθούν είτε ανεξάρτητα είτε μαζί ως πλαίσιο εικόνας. Επιπλέον κατά την διάρκεια του ΔΜΣ είναι δυνατόν να επιλεγεί για τα πλαίσια εικόνας σε επίπεδο MB μετασχηματισμός χωριστά του κάθε πεδίου (αρκεί να μην υπάρχει επικάλυψη) ή συνολικά ως πλαίσιο εικόνας. Η επιλογή είναι δυνατόν να βασίζεται στο κινητικό περιεχόμενο του MB κάθε εικόνας. Ακολουθίες εικόνων με μικρή κίνηση μπορούν να κωδικοποιηθούν σε πλαίσια MB ενώ ακολουθίες εικόνων με μεγάλη κίνηση μπορούν να κωδικοποιηθούν χρησιμοποιώντας πεδία εικόνας. Η δομή ενός MB φωτεινότητας και οι διάφοροι τρόποι λειτουργίας κατά τον ΔΜΣ φαίνονται στα σχήματα 5.5 και 5.6 αντίστοιχα.



*Πρόβλεψη Κίνησης με Πεδία ή Πλαίσια Εικόνων:* Υπάρχουν δύο βασικοί τρόποι πρόβλεψης της κίνησης: απλή πρόβλεψη με πλαίσια και απλή πρόβλεψη με πεδία. Στην πρόβλεψη κίνησης με πεδία χρησιμοποιούνται μόνο προηγούμενα πεδία εικόνων. Παρόμοια ισχύουν και για την περίπτωση της απλής πρόβλεψης με πλαίσια εικόνων. Επιπλέον είναι δυνατό στην πρόβλεψη κίνησης στα πλαίσια εικόνων να χρησιμοποιηθούν και πεδία εικόνων ανά MB. Η επιλογή των διάφορων τρόπων λειτουργίας βασίζεται και πάλι στο κινητικό περιεχόμενο της εικόνας.

Άλλοι δύο τρόποι πρόβλεψης / εκτίμησης κίνησης που παρέχονται στο MPEG-2 είναι: ο τρόπος λειτουργίας 16 x 8 και ο τρόπος λειτουργίας dual-prime. Στην πρώτη περίπτωση που χρησιμοποιείται μόνο στην πρόβλεψη κίνησης σε πεδία εικόνων για κάθε MB 16 x 8 χρησιμοποιούνται δύο διανύσματα κίνησης: ένα για την πάνω μισή περιοχή που ανήκει στο περιττό πεδίο και ένα για την κάτω μισή περιοχή που ανήκει στο άρτιο πεδίο. Ο τρόπος λειτουργίας dual-prime προβλέπει εκτός από το διάνυσμα κίνησης (να κωδικοποιείται) και ένα διαφορικό διάνυσμα. Για κάθε MB δημιουργούνται δύο προβλέψεις από δύο πεδία εικόνων, οι οποίες στην συνέχεια αθροίζονται για να δημιουργηθεί η τελική πρόβλεψη. Ο τρόπος λειτουργίας dual-prime χρησιμοποιείται μόνο για εικόνες τύπου P.

*Χρωματική Πληροφορία:* Επειδή το MPEG-2 προορίζεται για εφαρμογές με μεγάλες απαιτήσεις ποιότητας ορίστηκαν επιπλέον τρόποι υποδειγματοληψίας των χρωματικών συνιστωσών εκτός της βασικής 4:2:0 που προδιαγράφει το πρότυπο MPEG-1. Οι επιπλέον δυνατότητες είναι υποδειγματοληψία μόνο στην οριζόντια κατεύθυνση 4:2:2 και μηδενική υποδειγματοληψία 4:4:4. Στο σχήμα 5.7 παρουσιάζονται τις διαφορετικές υποδειγματοληψίες για ένα MB το οποίο αποτελείται από τέσσερα B φωτεινότητας και τα αντίστοιχα χωρικά εξαρτώμενα B χρωματικότητας. Εν κατακλείδι ένα MB στο MPEG-2 μπορεί να περιέχει 6 (4 Φωτεινότητας, 1 Cr και 1 Cb) 8 (4 Φωτεινότητας, 2 Cr και 2 Cb) ή 12 (4 Φωτεινότητας, 4 Cr και 4 Cb) 8 x 8 block.

Το Κύριο Υποσύνολο λειτουργίας (ΥΛ) είναι μια κατευθείαν επέκταση του προτύπου MPEG-1 για το χειρισμό διαπλεκόμενου (interlaced) video διατηρώντας τα βασικά χαρακτηριστικά του MPEG-1. Η δομή του κωδικοποιητή είναι όμοια με αυτή του MPEG-1. Αντίθετα στο απλό ΥΛ ο κωδικοποιητής δεν χειρίζεται εικόνες τύπου B. Αυτό σημαίνει

οτι οι αποκωδικοποιητές που εργάζονται στο απλό ΥΛ δεν χρειάζονται τον απαραίτητο αποθηκευτικό χώρο για αποκωδικοποίηση των πλαισίων τύπου B.

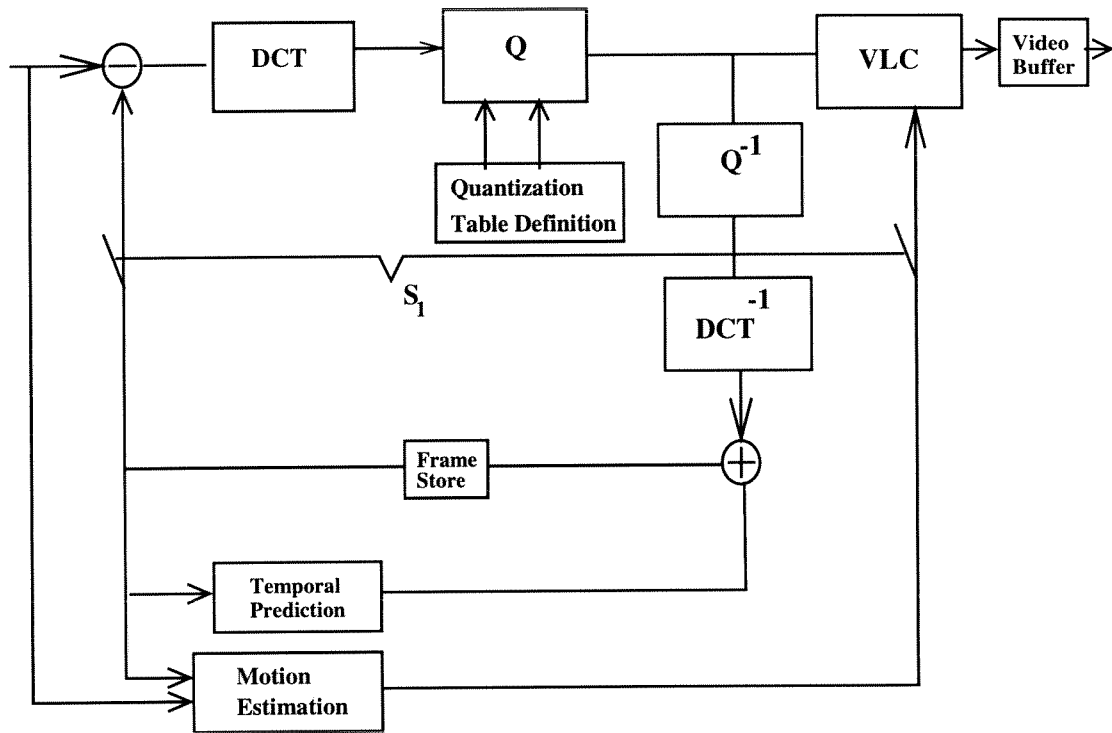
### Κλιμακούμενοι τρόποι λειτουργίας

Με την έννοια των κλιμακούμενων τρόπων λειτουργίας αναφερόμαστε στη δυνατότητα για αποκωδικοποίηση σε ορισμένη ανάλυση της συμπιεσμένης συμβολοσειράς εισόδου. Αυτό είναι χρήσιμο γιατί θα υπάρχουν αποκωδικοποιητές με διαφορετικές υπολογιστικές δυνατότητες που θα μπορούν να αποκωδικοποιούν και να ανακατασκευάζουν video σε ποικίλες χωροχρονικές αναλύσεις. Το ελάχιστο υποσύνολο μιας συμπιεσμένης συμβολοσειράς εισόδου που μπορεί δυνατοτήτες να αποκωδικοποιηθεί ονομάζεται βασικό επίπεδο (basic layer) video. Τα επίπεδα τα οποία βελτιώνουν την χωροχρονική ανάλυση του βασικού επιπέδου ονομάζονται ενισχυτικά επίπεδα. Η γενική δομή ενός κλιμακωτού κωδικοποιητή video φαίνεται στο σχήμα 5.8 (Παρουσιάζεται η λύση για δύο αναλύσεις). Το αρχικό σήμα εισόδου υποδειγματοληπτείται χρονικά ή χωρικά και κωδικοποιείται στο βασικό επίπεδο ανάλυσης με μειωμένο ρυθμό μετάδοσης. Μία υπερδειγματοληπτημένη ανακατασκευασμένη έκδοση του βασικού επιπέδου αποτελεί την εκτίμηση για το σήμα εισόδου σε υψηλή ανάλυση. Το σφάλμα πρόβλεψης κωδικοποιείται και παράγεται συμβολοσειρά του ενισχυτικού επιπέδου. Εάν ένας αποκωδικοποιητής δεν είναι ικανός ή δεν θέλει να αποκωδικοποιήσει video σε μέγιστη ανάλυση, μπορεί να αποκωδικοποιήσει το βασικό επίπεδο ανάλυσης. Οι κλιμακούμενοι τρόποι λειτουργίας στο MPEG-2 παίρνουν την ακόλουθη μορφή:

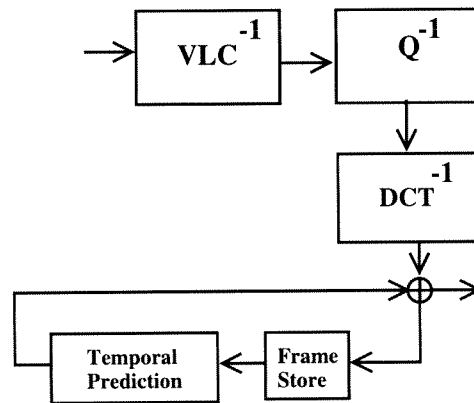
- Η χωροκλιμακούμενη λειτουργία η οποία αντιστοιχεί στο ομόνυμο ΥΛ παρέχει τη δυνατότητα για αποκωδικοποίηση video σε διάφορες αναλύσεις χωρίς αποκωδικοποίηση και υποδειγματοληψία. Το βασικό επίπεδο αποτελεί ένα αντίγραφο του αρχικού σήματος σε μικρότερη χωρική ανάλυση. Τα ενισχυτικά επίπεδα παρέχουν διαδοχικά πληροφορία υψηλότερου συχνοτικού περιεχομένου. Το βασικό επίπεδο δημιουργείται με χωρική υποδειγματοληψία του αρχικού σήματος ενώ το ενισχυτικό επίπεδο είναι η διαφορά του αρχικού σήματος και του υπερδειγματοληπτημένου ανακατασκευασμένου από το βασικό επίπεδο.

- Η κλιμακούμενη SNR λειτουργία η οποία αντιστοιχεί στο ομόνυμο ΥΛ προσφέρει την δυνατότητα για αποκωδικοποίηση με διαφορετικά χαρακτηριστικά του κβαντιστή. Το βασικό επίπεδο δημιουργείται χρησιμοποιώντας κβαντιστή με μεγάλο βήμα χονδροειδούς ανάλυσης και έχει ανάλυση ίδια με αυτή του ενισχυτικού επιπέδου. Στο ενισχυτικό επίπεδο κωδικοποιείται απλά η διαφορά μεταξύ του αρχικού σήματος και του βασικού επιπέδου.
- Η χρονοκλιμακούμενη λειτουργία η οποία δεν έχει αντιστοιχηθεί ακόμα σε κάποιο ΥΛ αναφέρεται στη δυνατότητα για αποκωδικοποίηση με διαφορετικό ρυθμό πλαισίων / sec χωρίς να είναι αναγκαία η αποκωδικοποίηση κάθε πλαισίου εικόνας.

Είναι δυνατό οι παραπάνω τρόποι λειτουργίας να συνδυασθούν με αποτέλεσμα έναν υβριδικό τρόπο λειτουργίας. Ένα σημαντικό πλεονέκτημα των κλιμακούμενων τρόπων λειτουργίας είναι ότι μπορούν να προσφέρουν αυξημένη αντοχή σε περιβάλλον θορύβου αφού το βασικό επίπεδο μπορεί να μεταδοθεί με καλύτερη δυνατότητα διόρθωσης σφαλμάτων σε σχέση με το ενισχυμένο επίπεδο.

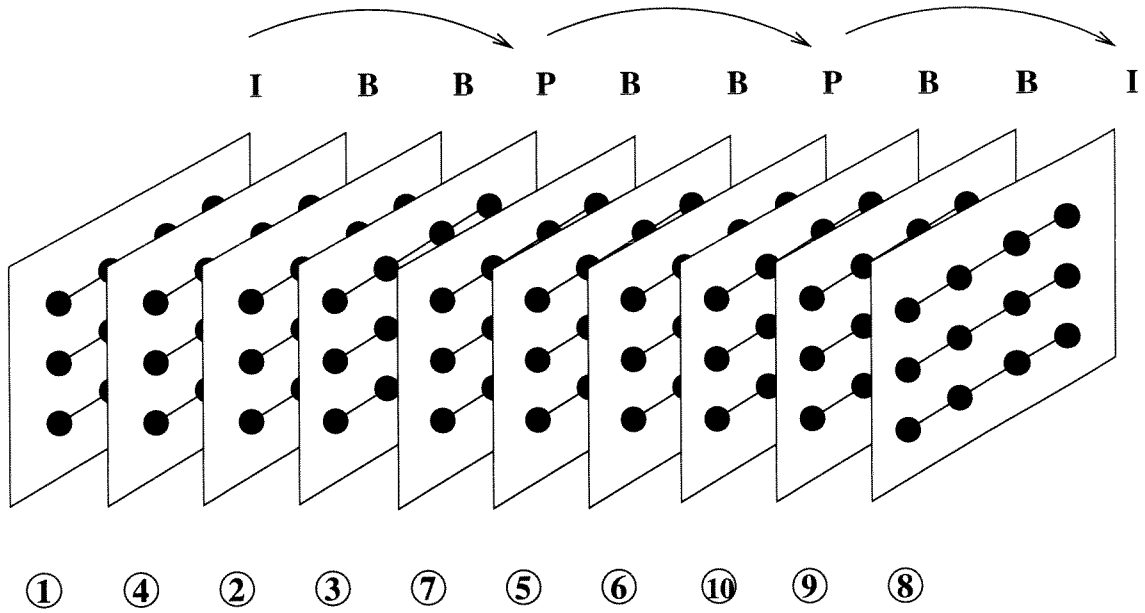


**ENCODER**

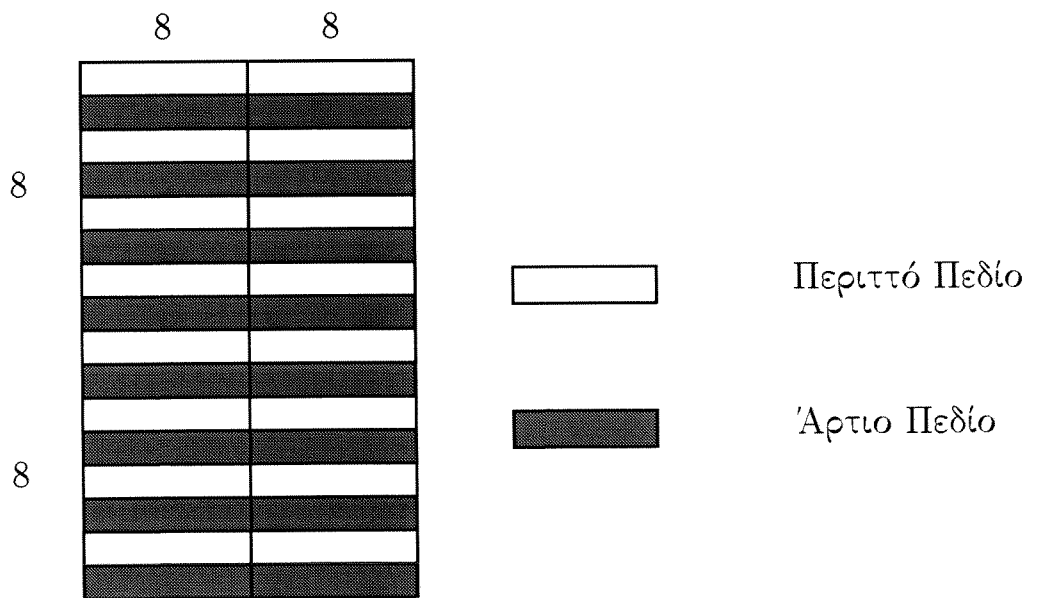


**DECODER**

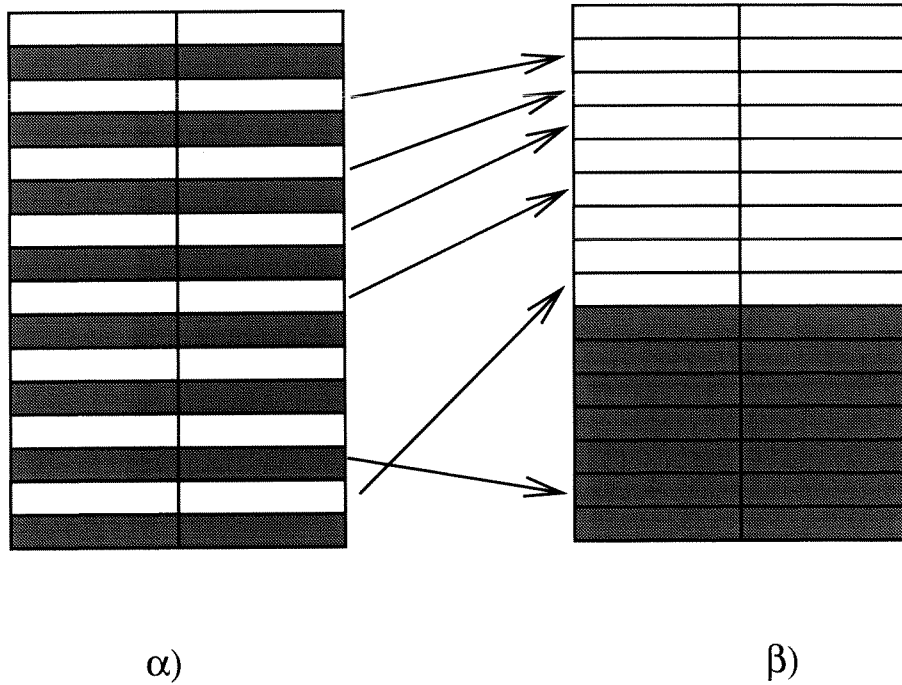
Σχήμα 5.3: Block διαγράμμα κωδικοποιητή /αποκωδικοποιητή



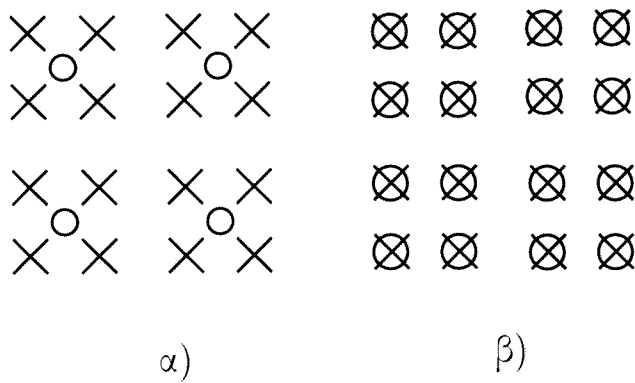
Σχήμα 5.4: I-εικόνες, P-εικόνες, και B-εικόνες χρησιμοποιούνται στην κωδικοποίηση του MPEG-1



Σχήμα 5.5: Ένα MB φωτεινότητας σε ένα πλαίσιο εικόνας

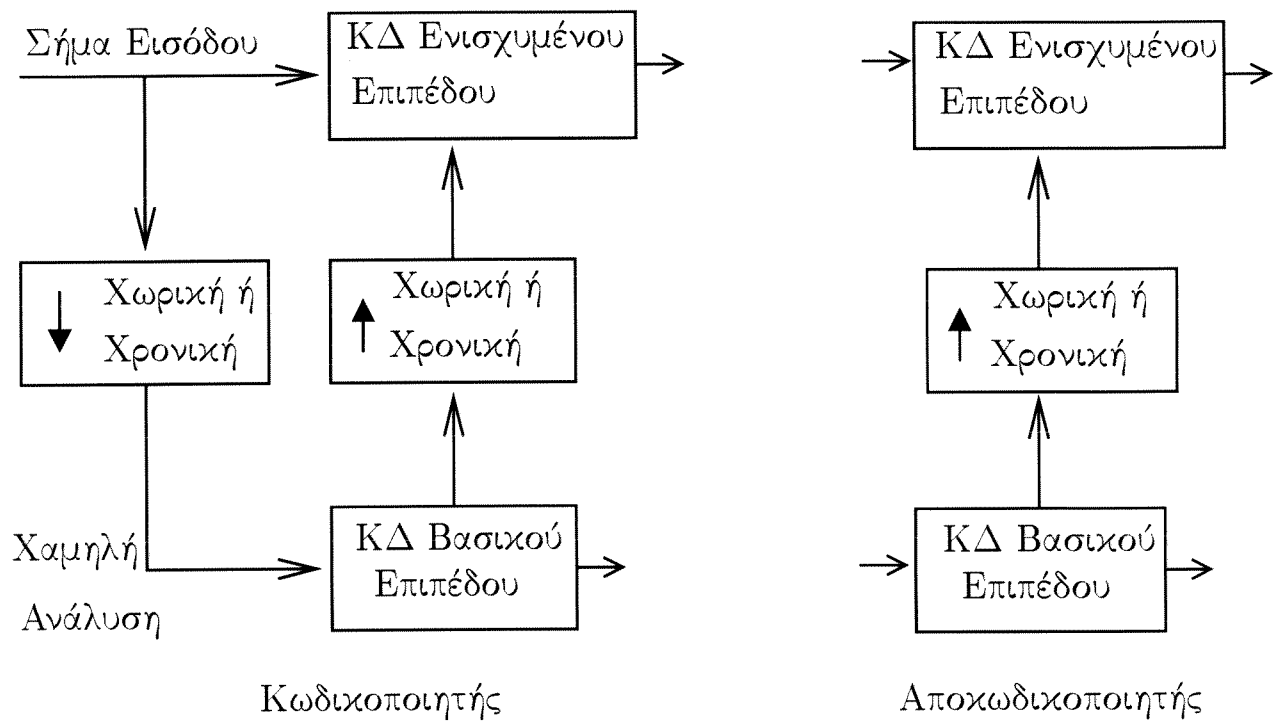


Σχήμα 5.6: ΔΜΣ σε πεδία (α) και πλαίσια εικόνων



- × pixel Φωτεινότητας
- pixel Χρωματικότητας

Σχήμα 5.7: α) format 4:2:0 β) format 4:2:2



Σχήμα 5.8: Κλιμακωτός κωδικοποιητής Video

## Αναφορές

- [1] CCITT SG XV, “Description of reference model 8 (RM8)”, Doc. 525, June 1989.
- [2] F. Dufaux and F. Mosheni, “Motion Estimation Techniques for Digital TV”, *Proc. of the IEEE*, vol. 83, no. 2, June 1995.
- [3] “Recommendation T.4 - Standardization of Group 3 facsimile apparatus for document transmission”, ITU, Geneva.
- [4] “Recommendation T.6 - Facsimile coding schemes and coding control functions for Group 4 facsimile apparatus”, ITU, Geneva.
- [5] CCITT Recommendation H.261, “Video Codec For Audiovisual at  $p \times 64$  kb/s,” Geneva 1990.
- [6] ISO 10918-1, “Information Technology - Digital Compression and coding of continuous-tone still images, Requirements and guidelines”, Geneva, 1994
- [7] ISO/CD 11172-2, “ Coding Of Moving Pictures And Associated Audio For Digital Storage Media At Up To About 1.5 Mbps.”, March 1991.
- [8] ISO/IEC 13818-2, “Generic Coding of Moving Pictures And Associated Audio. Recommendation H.262, Committee Draft” , May 1994.
- [9] W. P. Pennebaker and J. L. Mitchell, “JPEG Still Image Data Compression Standard”, New York, *Van Nostrand Reinhold*, 1993



# Συμπίεση εικόνων σε χαμηλούς ρυθμούς μετάδοσης

## 6.1 Εισαγωγή

Οι τεχνικές που παρουσιάστηκαν στο προηγούμενο κεφάλαιο παγιωμένες μέσω των διεθνών προτύπων επεξεργάζονται είτε μεμονωμένα pixel είτε block από pixels με βάση γνωστά στατιστικά μοντέλα. Οι συγκεκριμένες τεχνικές αποδίδουν ικανοποιητικά αποτελέσματα σε ρυθμούς μετάδοσης πάνω από 1.2 Mbit/sec. Εντούτοις, τα ήδη γνωστά πρότυπα παρουσιάζουν κακή ποιότητα εικόνας σε πολύ χαμηλούς ρυθμούς μετάδοσης π.χ 15 kbps. Τέτοιοι ρυθμοί μετάδοσης είναι απαραίτητες για μετάδοση κινουμένων εικόνων μέσω κοινών τηλεπικοινωνιακών γραμμών με χρήση απλών modem. Οι προσπάθειες για δημιουργία μεθόδων που πετυχαίνουν μεγάλους λόγους συμπίεσης συχνά εκμεταλλεύονται τις ιδιότητες του ανθρώπινου οπτικού συστήματος και συνήθως αναφέρονται ως “τεχνικές κωδικοποίησης δεύτερης γενιάς” [18]. Η διεθνής επιτροπή ITU (πρώην CCITT) εκτιμώντας το έντονο βιομηχανικό ενδιαφέρον προχώρησε σε ανοικτό διαγωνισμό για ένα νέο πρότυπο συμπίεσης κινούμενων εικόνων σε χαμηλούς ρυθμούς μετάδοσης το οποίο θα ονομάζεται MPEG-4. Το νέο πρότυπο εν αντιθέσει με τα ήδη υπάρχοντα δεν είναι υποχρεωμένο να παρέχει προς τα πίσω συμβατότητα. Οι υποψήφιες μέθοδοι προέρχονται από τεχνικές οι οποίες ανήκουν στην κατηγορία αλγόριθμων συμπίεσης δεύτερης γενιάς

και βασίζονται σε πιο ρεαλιστικά μοντέλα εκτίμησης κίνησης. Οι τεχνικές αυτές προσανατολίζονται σε τεχνικές περιγραφής περιγράμματος / υψής και προσπαθούν να περιγράψουν μια εικόνα με αρχέτυπα περιγραμμάτων και υψής. Αρχικά γίνεται κατάτμηση (segmentation) σε περιοχές σχετικά ομοιόμορφης υψής οι οποίες περικλείονται από περιγράμματα (contours). Στην συνέχεια κωδικοποιούνται ξεχωριστά τα περιγράμματα και οι περιοχές με αποτέλεσμα την επίτευξη μεγάλων λόγων συμπίεσης. Στην κατηγορία αυτή ανήκουν και οι αλγόριθμοι κωδικοποίησης με μοντέλο KMM (Model Based Coding) ή αλγόριθμοι ανάλυσης / σύνθεσης οι οποίοι χαρακτηρίζονται από τα παρακάτω βήματα:

- Ανάλυση εικόνας: Σε κάθε πλαίσιο εικόνας προς κωδικοποίηση γίνεται κατάτμηση (segmentation) σε αμιγή κινούμενα αντικείμενα με την βοήθεια της ίδιας διαδικασίας από προηγούμενα πλαίσια. Κάθε κινούμενο αντικείμενο στο τρέχον πλαίσιο εικόνας χαρακτηρίζεται από ένα σύνολο χαρακτηριστικών π.χ σχήμα (περίγραμμα), επιφάνεια και διάνυσμα κίνησης. Αλγόριθμοι αντιστάθμισης κίνησης μπορούν να εφαρμοσθούν ώστε να γίνει ανάλυση σε χρονικά αντισταθμισμένο πλαίσιο εικόνας.
- Σύνθεση εικόνας: Κάθε πλαίσιο εικόνας συντίθεται από την εκτίμηση των παραμέτρων ανάλυσης, δηλ. του περιγράμματος και της κίνησης. Υπολογίζεται η διαφορά της συνθετικής εικόνας από την πραγματική (σφάλμα πρόβλεψης) κι ακόμη υπολογίζεται και ελέγχεται η εγκυρότητα του μοντέλου ανάλυσης. Σε περίπτωση αποτυχίας του μοντέλου ανάλυσης μεταδίδεται το σφάλμα πρόβλεψης.
- Κωδικοποίηση: Οι παράμετροι του μοντέλου ανάλυσης καθώς και το σφάλμα πρόβλεψης κωδικοποιούνται με κωδικοποιητές εντροπίας.

Αν και οι κωδικοποιητές ΔΜΣ-ΑΚ εμφανίζονται να μην είναι τόσο ανταγωνιστικοί, είναι δυνατόν με κατάλληλες τροποποιήσεις, οι οποίες προέρχονται από τις τεχνικές KMM, να αποτελέσουν ένα σοβαρό διεκδικητή για το υποκατασκευή πρότυπο MPEG-4 [24]. Σε αυτό το κεφάλαιο προτείνεται μία μέθοδος συμπίεσης για ακολουθίες κινούμενων εικόνων (video) σε χαμηλούς ρυθμούς μετάδοσης  $\leq 64\text{ kbit/sec}$  η οποία βασίζεται σε αλγόριθμους συμπίεσης H.261 [11] και MPEG-1 [22]. Ο προτεινόμενος αλγόριθμος συμπίεσης εκμεταλλεύεται προκαθορισμένους περιορισμούς ποιότητας στο σήμα εισόδου οι οποίοι

μπορούν να χρησιμοποιηθούν για την κατασκευή του κβαντιστή του σχήματος 5.2. Παρόλο που ο κβαντιστής μπορεί να κατασκευαστεί με βάση το περιεχόμενο της εικόνας [29] είναι δυνατόν με την χρήση νευρωνικών δικτύων να επιτύχουμε οπτικά προσαρμοζόμενους κβαντιστές. Η χρήση διαφορετικών κβαντιστών σε μία εικόνα δεν έχει εφαρμοσθεί ευρέως παρά μόνο σε αλγόριθμους μεταβλητού ρυθμού μετάδοσης, οι οποίοι μεταβάλλουν τα χαρακτηριστικά του κβαντιστή τοπικά στην εικόνα προκειμένου να πετύχουν περιορισμούς στον συνολικό ρυθμό μετάδοσης σε αλγόριθμους μετάδοσης κινούμενης εικόνας (video). Ο αλγόριθμος που προτείνεται θεωρεί ότι σε κάθε πλαίσιο εικόνας τα διάφορα τμήματα της έχουν διαφορετική σημασία, γεγονός που δεν συμβαίνει στην κοινή περίπτωση εκπομπής τηλεοπτικού σήματος.

Ο ορισμός Περιοχών Ενδιαφέροντος (ΠΕ) σε οπτικά προσαρμοζόμενους κβαντιστές μπορεί να είναι πολύ αποδοτικός σε περιπτώσεις υπηρεσιών με χαμηλούς ρυθμούς μεταφοράς, όπως μετάδοση ειδικών ιατρικών εικόνων (τηλεϊατρική), τηλεσυνδιάσκεψη, τηλεπισκόπηση, ανεύρεση στοιχείων βάσεων δεδομένων εικόνων με βάση το περιεχόμενο των εικόνων. Σε αυτές τις εφαρμογές μπορεί να θεωρηθεί ότι μια συνολική σταθερή ποιότητα ανακατασκευής δεν είναι επιθυμητή.

Στην περίπτωση εικόνων τηλεσυνδιάσκεψης η ποιότητα των τμημάτων κεφαλιού και ώμων είναι πιο κρίσιμη από την ποιότητα του φόντου η οποία μπορεί να υποστεί μεγαλύτερες αλλοιώσεις. Παρόμοια στην επιτήρηση χώρων και αντικειμένων με video μερικά τμήματα της εικόνας θεωρούνται υψηλότερης προτεραιότητας σε σχέση με την υπόλοιπη εικόνα. Στην περίπτωση τηλεϊατρικής, έγχρωμες εικόνες δερματολογικών παθήσεων που αποστέλλονται από ένα κέντρο υγείας σε ένα κεντρικό νοσοκομείο για εξέταση, (Οι συγκεκριμένες εικόνες) εμπεριέχουν περιοχές φόντου οι οποίες ελάχιστα συνεισφέρουν στην χρήσιμη πληροφορία και από περιοχές που βρίσκονται στο προσκήνιο και είναι υψηλού ενδιαφέροντος. Εάν εφαρμοσθεί σχετικά “χωδροειδής” (λεπτομερής) κβαντισμός στην περιοχές χαμηλού (υψηλού) ενδιαφέροντος είναι δυνατόν να γίνει μια σημαντική αύξηση του λόγου συμπίεσης.

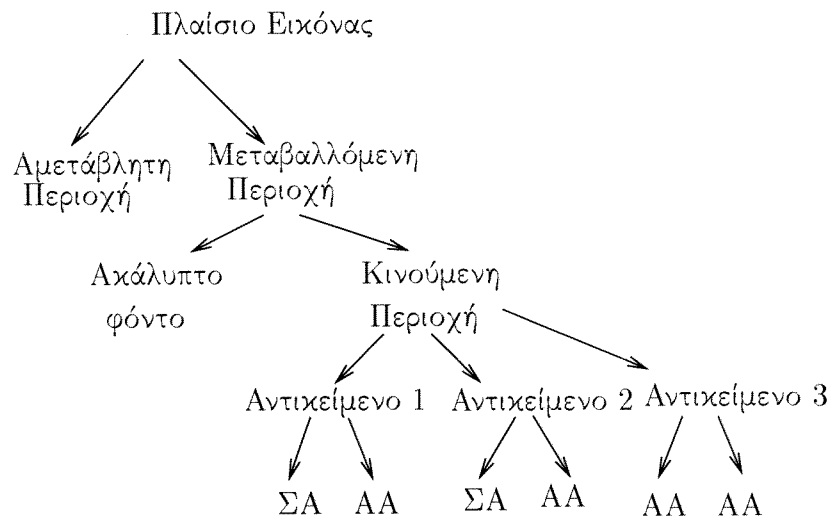
Η παραπάνω παρατήρηση δείχνει ότι μία μεταβλητή ποιότητα ανακατασκευής είναι δυνατή εάν ορισθούν εκ των προτέρων επίπεδα ανακατασκευής με βάση τα περιεχόμενα της

εικόνας. Αυτή η διαδικασία “εστίασης ενδιαφέροντος” απαιτεί δύο βασικές αλγοριθμικές μονάδες. Η πρώτη είναι η μονάδα κατάτμησης που δίνει πληροφορίες για την επιλογή και τοποθέτηση των ΠΕ με βάση *a priori* κριτήρια [26]. Στην περίπτωση εικόνων τηλεσυνδιάσκεψης μπορεί να είναι οι συντεταγμένες του εικονιζόμενου προσώπου και των ώμων του προσώπου της εικόνας. Η δεύτερη αλγοριθμική μονάδα είναι ένας ήδη υπάρχον προσαρμοζόμενος αλγόριθμος με απώλειες όπου η πληροφορία για τις περιοχές ενδιαφέροντος δημιουργεί μια κατάλληλη χωρικά ανομοιογενή ποιότητα ανακατασκευής ώστε να επιτευχθεί ένας συγκεκριμένος ρυθμός μετάδοσης ή μία δεδομένη ποιότητα ανακατασκευής.

Η δομή του κεφαλαίου είναι η ακόλουθη: Στην Ενότητα 6.2 παρουσιάζονται περιληπτικά οι βασικές μέθοδοι της ΚΜΜ: α) κωδικοποίηση με αντικείμενα β) κωδικοποίηση με βάση γνωστά αντικείμενα και γ) κωδικοποίηση με σημασιολογικά αντικείμενα. Ένας υβριδικός κωδικοποιητής ΔΜΣ-ΑΚ με χαρακτηριστικά ΚΜΜ, ο ΔΜΣ-ΑΚ με Περιοχές Ενδιαφέροντος, παρουσιάζεται στην Ενότητα 6.3. Ο κωδικοποιητής ΔΜΣ-ΑΚ με Περιοχές Ενδιαφέροντος εμφανίζεται αποτελεσματικός στην καταπολέμηση των βασικών μειονεκτημάτων των υπαρχόντων προτύπων σε χαμηλούς ρυθμούς μετάδοσης σε περιβάλλον εικόνων τηλεσυνδιάσκεψης. Ένας προσαρμοζόμενος αλγόριθμος εύρεσης των περιοχών ενδιαφέροντος παρουσιάζεται στην Ενότητα 6.4 με την βοήθεια νευρωνικών δικτύων. Στην Ενότητα 6.5 παρουσιάζονται: α) πειραματικά αποτελέσματα χρησιμοποιώντας πραγματικές σειρές εικόνων τηλεσυνδιάσκεψης και β) παράμετροι υλοποίησης του νευρωνικού δικτύου. Τέλος δίνονται κατευθύνσεις για μελλοντική έρευνα.

## 6.2 Κωδικοποίηση Με Μοντέλα (ΚΜΜ)

Μια βασική αδυναμία των ΔΜΣ-ΑΚ κωδικοποιητών έγκειται στην υπόθεση για 2-Δ μεταφορική κίνηση των block της εικόνας. Αυτή η υπόθεση αποδεικνύεται ελλιπής στην περίπτωση ακριβούς περιγραφής της κίνησης μιας εικόνας όταν υπάρχει στροφή και μεγέθυνση. Επιπλέον στην περίπτωση που υπάρχουν κινούμενα αντικείμενα μέσα στην εικόνα των οποίων τα όρια δεν συμπίπτουν με τα Block της εικόνας παρουσιάζεται το φαινόμενο της λανθασμένης εκτίμησης της κίνησης [28]. Οι μέθοδοι ΚΜΜ χρησιμοποιούν χωρικούς,

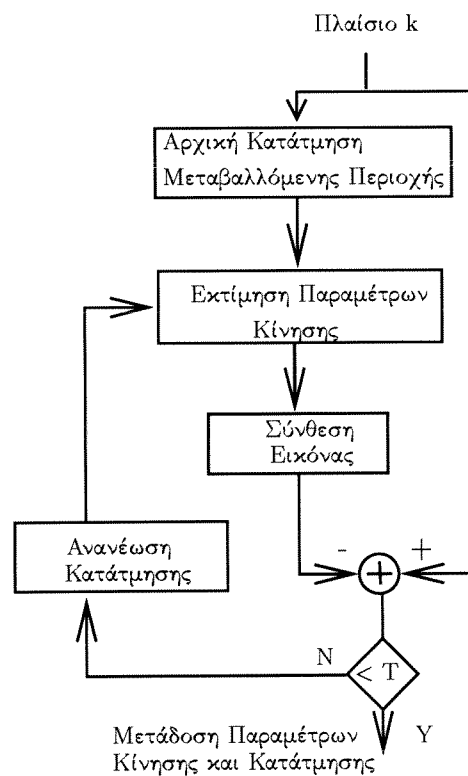


Σχήμα 6.1: Κατάτμηση πλαισίου Εικόνας σε αντικείμενα

γραμμικούς και συγγενείς μετασχηματισμούς καθώς και κατάτμηση της εικόνας σε διάφορα αντικείμενα για την λεπτομερειακή και πραγματική καταγραφή της κίνησης μέσα σε μία εικόνα. Οι μέθοδοι KMM χωρίζονται σε δύο βασικές κατηγορίες: α) κωδικοποίηση με αντικείμενα και β) κωδικοποίηση με γνωστά αντικείμενα.

### 6.2.1 Κωδικοποίηση με αντικείμενα

Η μέθοδος κωδικοποίησης με αντικείμενα βασίζεται σε θεωρία αναπαράστασης του 3-Δ χώρου με άγνωστα αντικείμενα τα οποία περιγράφονται με διαφορετικά μοντέλα. Από αλγοριθμικής πλευράς κάθε πλαίσιο εικόνας χωρίζεται στην “αμετάβλητη περιοχή”, στο “φόντο χωρίς επικάλυψη” και σε ένα αριθμό από άγνωστα αντικείμενα τα οποία είναι είτε “συμβατά” (ΣΑ) είτε “ασύμβατα” (ΣΑ) με κάποιο μοντέλο περιγραφής αντικειμένων (σχήμα 6.1). Τα άγνωστα αντικείμενα μπορούν να περιγραφούν [21] ως: α) 2-Δ συμπαγή ή παραμορφώσιμα αντικείμενα με 2-Δ κίνηση β) 2-Δ συμπαγή αντικείμενα με 3-Δ κίνηση [12, 6] ή γ) 3-Δ συμπαγή ή παραμορφώσιμα αντικείμενα με 3-Δ κίνηση [21, 27, 15]. Παρουσιάζονται στην συνέχεια οι σχετικές μέθοδοι.



Σχήμα 6.2: Διάγραμμα ροής

### 2- $\Delta$ /3- $\Delta$ Συμπαγή αντικείμενα με 3- $\Delta$ Κίνηση

Δεδομένων των παρουσών υποθέσεων κίνησης και αντικειμένων (2- $\Delta$ /3- $\Delta$  συμπαγή αντικείμενα με 3- $\Delta$  κίνηση) είναι δυνατή η ταυτόχρονη κατάτμηση των αντικειμένων της εικόνας και η εκτίμηση της κίνησης τους στο πλαίσιο μίας εικόνας [25]. Για κάθε ανεξάρτητα κινούμενο αντικείμενο οι πληροφορίες της κίνησης και του περιγράμματός του (segmentation mask) αποτελούν μια περιγραφή του αντικειμένου και μπορούν να χρησιμοποιηθούν ως μια αντισταθμισμένη εκτίμηση με κίνηση του επόμενου πλαισίου εικόνας. Για την περίπτωση της περιοχής του φόντου μεταδίδεται και το σφάλμα σύνθεσης το οποίο βελτιώνει την ποιότητα στις περιοχές που δεν γίνεται περιγραφή με μοντέλα αντικειμένων.

Στη γενικότερη μορφή του ένας κωδικοποιητής με αντικείμενα περιγράφεται από το σχήμα 6.2 και υλοποιεί τα παρακάτω βήματα

1. *Ανάλυση:* Αρχικά κάθε εικόνα χωρίζεται σε περιοχές μεταβαλλόμενες και αμετάβλητες στις οποίες γίνεται επεξεργασία για τις κινούμενες περιοχές οι οποίες μπορεί

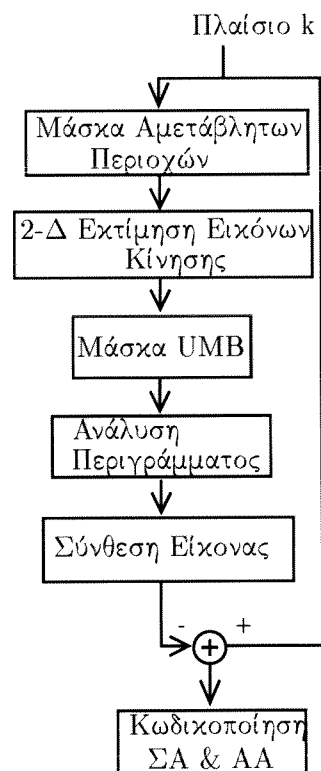
να περιέχουν ένα ή περισσότερα αντικείμενα. Στην συνέχεια πραγματοποιείται εκτίμηση κίνησης και κατάτμηση εικόνας (segmentation) μεταξύ του  $k$  πλαισίου εικόνας και του ανακατασκευασμένου  $k-1$ , για την εύρεση των καλύτερων παραμέτρων κίνησης και των περιγραμμάτων των αντικειμένων (περιοχών) της κατατμημένης εικόνας. Για κάθε ανεξάρτητο κινούμενο αντικείμενο οι πληροφορίες της κίνησης και του περιγράμματός του (segmentation mask) αποτελούν μια περιγραφή του αντικειμένου και μπορούν να χρησιμοποιηθούν ως μια αντισταθμισμένη εκτίμηση με κίνηση του επόμενου πλαισίου εικόνας. Η σύνθεση του  $k$ -οστού πλαισίου εικόνας είναι δυνατή χρησιμοποιώντας τις παραπάνω παράμετρους. Τα αντικείμενα των οποίων το σφάλμα σύνθεσης είναι μεγαλύτερο από ένα κατώφλι  $T$ , κατηγοριοποιούνται ως ασύμβατα αντικείμενα τα οποία στη συνέχεια διαχωρίζονται σε περισσότερα αντικείμενα. Η διαδικασία συνεχίζεται μέχρι να ικανοποιηθεί κάποιο κριτήριο ποιότητας.

2. *Σύνθεση*: Με βάση τις παραμέτρους κίνησης και τα στοιχεία κατάτμησης εικόνας γίνεται ανασύνθεση του παρόντος πλαισίου εικόνας από το προηγούμενο ανακατασκευασμένο πλαίσιο εικόνας. Υπολογίζεται το σφάλμα σύνθεσης και αποστέλεται στον αποκωδικοποιητή.
3. *Κωδικοποίηση*: Κωδικοποιούνται οι παράμετροι κίνησης, κατάτμησης της εικόνας και του σφάλματος σύνθεσης με κωδικοποιητή εντροπίας

Το μειονέκτημα του παρόντος αλγορίθμου είναι η μεγάλη υπολογιστική πολυπλοκότητα στην ταυτόχρονη εκτίμηση της κίνησης και του περιγράμματος των αντικειμένων.

## 2-Δ Παραμορφώσιμα αντικείμενα με 2-Δ Κίνηση

Μείωση της υπολογιστικής πολυπλοκότητας μπορεί να επιτευχθεί εάν υποθέσουμε μεταφορική κίνηση και παραμορφώσιμα 2-Δ αντικείμενα. Αυτή η υπόθεση έχει επίσης υιοθετηθεί και από το μοντέλο COST211ter για να καταστεί δυνατή η υλοποίηση σε κυκλωματική μορφή. Η βασική ιδέα του αλγορίθμου είναι η κατάτμηση της εικόνας σε τριών ειδών περιοχές: α) αμετάβλητες περιοχές β) κινούμενες περιοχές, των οποίων κάθε pixel βρίσκεται στο προηγούμενο πλαίσιο της εκάστοτε εικόνας και γ) περιοχές φόντου χωρίς επικάλυψη



Σχήμα 6.3: Διάγραμμα ροής

και χωρίς επαναλαμβανόμενα pixel. Πραγματοποιείται εκτίμηση κίνησης χρησιμοποιώντας μια ιεραρχική μέθοδο block matching [8]. Παράγεται με αυτόν τον τρόπο ένα αραιό πεδίο εκτίμησης κίνησης (sparse motion field) για τις περιοχές των block της εικόνας. Ένα πυκνό πεδίο εκτίμησης κίνησης (dense motion field) παράγεται με γραμμική παρεμβολή. Τα “Συμβατά Αντικείμενα” κωδικοποιούνται με τα αντίστοιχα 2-Δ διανύσματα κίνησης και την πληροφορία του περιγράμματος τους. Το σχήμα και το σφάλμα πρόβλεψης μεταδίδεται για τα μη συμβατά αντικείμενα. Ο συνολικός αλγόριθμος είναι γνωστός στην βιβλιογραφία ως object-based analysis-synthesis coder, υλοποιεί τα παρακάτω βήματα (σχήμα 6.3):

1. *Μάσκα Αμετάβλητων Περιοχών*: είναι μία δυαδική μάσκα (μορφολογική ή μάσκα ενδιάμεσης τιμής) για το καθορισμό των μεταβαλλόμενων ή αμετάβλητων περιοχών σε ένα πλαίσιο εικόνας σε σχέση με το προηγούμενο.

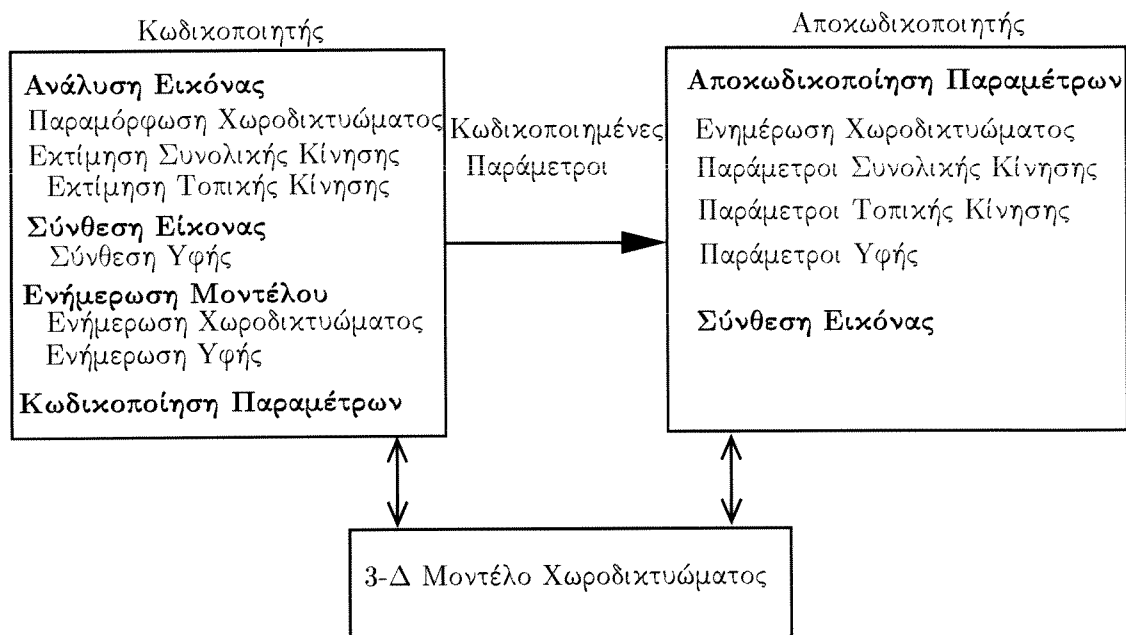


2. *2-Δ Εκτίμηση Κίνησης*: των pixel στις μεταβαλλόμενες περιοχές χρησιμοποιώντας την ιεραρχική μέθοδο block matching.
3. *Μάσκα UMB*: υπολογίζει τις περιοχές με κίνηση, τις περιοχές χωρίς επικάλυψη και τις αμετάβλητες περιοχές του κάθε πλαισίου εικόνας.
4. *Ανάλυση του περιγράμματος* κάθε περιοχής χρησιμοποιώντας συναρτήσεις splines.
5. *Σύνθεση των κινούμενων περιοχών* χρησιμοποιώντας τα στοιχεία κίνησης και περιγράμματος από τα προηγούμενα βήματα.
6. *Κωδικοποίηση των Συμβατών και Ασύμβατων Περιοχών* (των διανυσμάτων κίνησης και περιγραμμάτων τους) χρησιμοποιώντας κωδικοποιητές εντροπίας.

### 6.2.2 Κωδικοποίηση με γνωστά αντικείμενα

Η εκτίμηση της κίνησης και της δομής ενός άγνωστου 3-Δ αντικειμένου από δύο πλαίσια εικόνας είναι ένα δύσκολο πρόβλημα το οποίο αφενός έχει μεγάλη υπολογιστική πολυπλοκότητα και αφετέρου είναι ευαίσθητο στη παρουσία θορύβου. Άλλες μέθοδοι που βασίζονται σε απλούστερα 2-Δ μοντέλα χρησιμοποιούν το μεταφορικό μοντέλο κίνησης το οποίο δεν ισχύει πάντοτε · παρουσιάζουν όμως το πλεονέκτημα να είναι μικρής υπολογιστικής πολυπλοκότητας και συνεπώς πιο ελκυστικές για πραγματικές εφαρμογές. Συμβιβασμό μεταξύ των παραπάνω μεθόδων αποτελούν οι μέθοδοι κωδικοποίησης με γνωστά αντικείμενα στην περίπτωση που υπάρχει *a priori* γνώση για το περιεχόμενο της εικόνας που θα κωδικοποιηθεί. Η γνώση μπορεί να υπάρχει με την μορφή χωροδικτύωματος (wireframe) των αντικειμένων που θα κωδικοποιηθούν.

Γενικά στην κωδικοποίηση με γνωστά αντικείμενα θεωρείται δεδομένο (και γνωστό) ένα γενικό χωροδικτύωμα των προς κωδικοποίηση αντικειμένων τόσο στον κωδικοποιητή (πομπό) όσο και στο αποκωδικοποιητή (παραλήπτης). Ο κωδικοποιητής επιλέγει αρχικά ένα χωροδικτύωμα το οποίο στη συνέχεια το τροποποιεί (παραμορφώνει) για να το προσαρμόσει στο μέγεθος του ζητούμενου αντικειμένου. Η συνολική κίνηση του αντικειμένου μπορεί να προσεγγισθεί με μετατοπίσεις των ακμών του χωροδικτύωματος ενώ η τοπική



Σχήμα 6.4: Block διάγραμμα του Κωδικοποιητή με γνωστά αντικείμενα

κίνηση με τοπικές μετατοπίσεις των ακμών έχοντας ως περιορισμό κάποιες γεωμετρικές ιδιότητες του χωροδικτυώματος. Εναλλακτικά στην περίπτωση εικόνων με ανθρώπινα πρόσωπα έχουν αναπτυχθεί τεχνικές με σημασιολογικά χαρακτηριστικά [5]. Μία από τις κύριες εφαρμογές της κωδικοποίησης με γνωστά αντικείμενα είναι οι συσκευές τηλεσυνδιάσκεψης γι' αυτό και τα περισσότερα χωροδικτυώματα που έχουν σχεδιαστεί είναι για πρόσωπα.

Ένα block διάγραμμα ενός κωδικοποιητή / αποκωδικοποιητή με γνωστά αντικείμενα δίνεται στο σχήμα 6.4. Ο κωδικοποιητής αποτελείται από τέσσερις βασικές μονάδες : α) τη Μονάδα Ανάλυσης Εικόνας η οποία περιλαμβάνει την παραμόρφωση του χωροδικτυώματος και την εκτίμηση της συνολικής και τοπικής κίνησης της εικόνας β) τη Μονάδα Σύνθεσης Εικόνας η οποία περιλαμβάνει την σύνθεση υφής για το χωροδικτύωμα γ) τη Μονάδα Ενημέρωση του χωροδικτυώματος και της υφής και δ) την Μονάδα Κωδικοποίησης των παραμέτρων.

Από τις παραπάνω μονάδες ενδιαφέρον παρουσιάζει η μονάδα ανάλυσης εικόνας και ιδιαίτερα το υποσύστημα εκτίμησης της τοπικής κίνησης του προσώπου η οποία οφείλεται στην κίνηση των μυών του. Με την βοήθεια των χαρακτηριστικών μονάδων (XM)(action

units) προσώπου [1, 2, 5] οι οποίες ορίζονται στο Σύστημα Κωδικοποίησης Χαρακτηριστικών Προσώπου (FACS)[9] είναι δυνατό να εκτιμηθεί η τοπική κίνηση του προσώπου [10]. Το σύστημα FACS χρησιμοποιεί 44 ΧΜ για να περιγράψει όλες τις ανθρώπινες εκφράσεις. Εντούτοις μικρός μόνο αριθμός τέσσερις χρειάζεται για την κωδικοποίηση της τοπικής κίνησης του προσώπου.

### 6.3 Υβριδικός Κωδικοποιητής με Περιοχές Ενδιαφέροντος (ΔΜΣ-ΑΚ-ΠΕ)

Μία σημαντική εφαρμογή της συμπίεσης κινούμενων εικόνων σε χαμηλούς ρυθμούς μετάδοσης (15 kbps) είναι η τηλεσυνδιάσκεψη. Το αυξημένο ενδιαφέρον οφείλεται στην εκτιμώμενη ζήτηση που θα έχουν οπτικοακουστικές εφαρμογές διαμέσω κοινών τηλεπικοινωνιακών γραμμών και δικτύων κινητής τηλεφωνίας. Οι κωδικοποιητές με μοντέλο που παρουσιάστηκαν στις προηγούμενες ενότητες αν και δεν περιορίζονται στην συγκεκριμένη εφαρμογή έχουν αναπτυχθεί με σκοπό την βελτίωση της ποιότητας στην αναπαραστάση εικόνων τηλεσυνδιάσκεψης. Αν και οι μέθοδοι συμπίεσης που βασίζονται στον ΔΜΣ-ΑΚ δεν παρέχουν ικανοποιητική ποιότητα σε χαμηλούς ρυθμούς μετάδοσης είναι δυνατόν με κατάλληλες τροποποιήσεις που προέρχονται από τους ΚΜΜ να αποτελέσουν ένα διεκδικητή του επερχόμενου προτύπου MPEG-4. Η διαδικασία για την βελτίωση των συμβατικών κωδικοποιητών ΔΜΣ-ΑΚ δρομολογείται αφενός από την ανάλυση των ελαττωμάτων τους και αφετέρου από τα ειδικά χαρακτηριστικά που παρουσιάζουν οι εικόνες τηλεσυνδιάσκεψης.

Τα μειονέκτηματά των υπάρχοντων προτύπων συμπίεσης video που βασίζονται στον ΔΜΣ-ΑΚ αναφέρονται στην συνέχεια [4, 19]:

- Οι τεχνικές για τον καταμερισμό των bit στους κωδικοποιητές ΔΜΣ-ΑΚ είναι αμετάβλητες σε κάθε υπό εξέταση εικόνα. Η επιθυμητή υψηλή συμπίεση μπορεί να επιτευχθεί με “τραχύ” (coarse) κβαντισμό ώστε να αποφευχθεί υπερχειλίση του απομονωτή εξόδου. Αυτή η τεχνική έχει ως συνέπεια την αλλοίωση της ποιότητας συνολικά σε όλη την εικόνα.

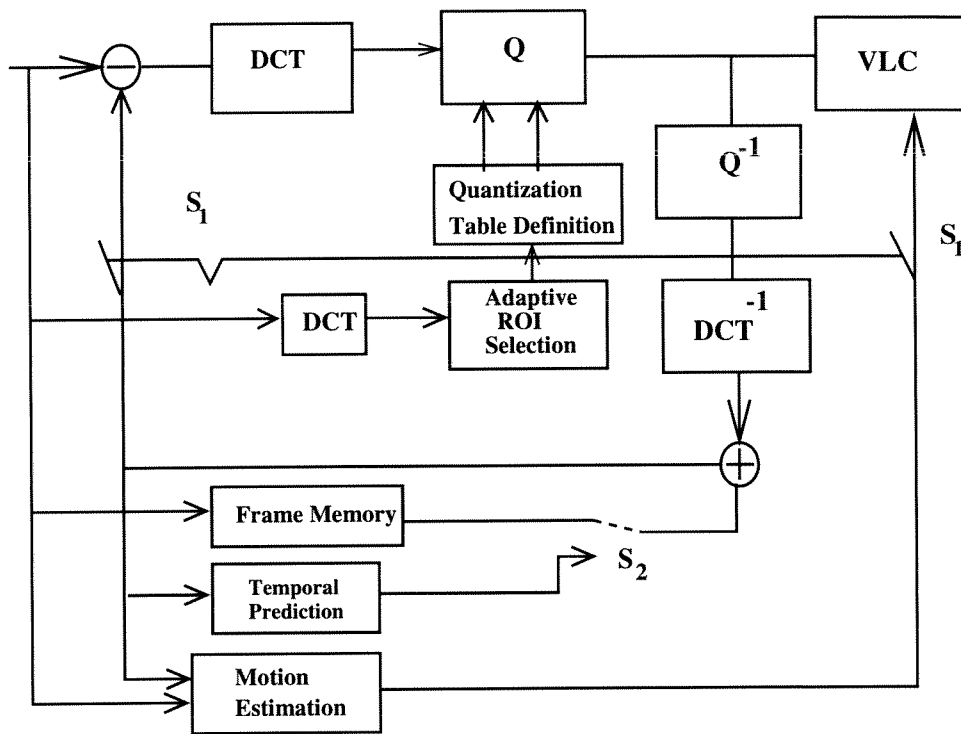
- Χαμηλοί ρυθμοί μετάδοσης κατά την διάρκεια interframe coding θέτουν αριθμητικούς περιορισμούς στο σφάλμα πρόβλεψης με αποτέλεσμα η ανανέωση των block της εικόνας να είναι σπάνια. Επιπλέον εξαιτίας της αναδρομικής υφής του ΔΜΣ-ΑΚ τα σφάλματα ανακατασκευής συσσωρεύονται γεγονός που οδηγεί στη σημαντική αλλοίωση των εικόνων.
- Η πρόβλεψη κίνησης που προέρχεται από εικόνες ανακατασκευασμένες με αντιστάθμιση κίνησης δημιουργεί blocky σχήματα. Ταυτόχρονα, ο κβαντισμός του σφάλματος πρόβλεψης δεν επιτρέπει τη μετάδοση block τμημάτων ανανέωσης με αποτέλεσμα την επιδείνωση της ποιότητας.

Η σχεδίαση ενός κωδικοποιητή για χρήση τηλεσυνδιάσκεψης ο οποίος θα εκμεταλλευόταν τις ιδιότητες των προς μετάδοση εικόνων θα μπορούσε να εξαλείψει πολλά από τα προαναφερθέντα μειονεκτήματα. Μερικές από αυτές τις ιδιότητες αναφέρονται στην συνέχεια.

- *Σταθερό περιεχόμενο εικόνας.* Οι τυπικές ακολουθίες εικόνων σε περιβάλλον τηλεσυνδιάσκεψης αποτελούνται από το πρόσωπο και τους ώμους του ομιλητή. Η πληροφορία για *a priori* ύπαρξη αντικειμένων μπορεί να συμπεριληφθεί στην κατασκευή του κωδικοποιητή.
- *Περιορισμένη κίνηση.* Η κίνηση μεταξύ των πλαισίων των εικόνων δημιουργείται από την κίνηση του ομιλητή ενώ η κάμερα παραμένει σχετικά ακίνητη. Η κίνηση του ομιλητή αποτελείται από την κίνηση του κεφαλιού και των ώμων των χαρακτηριστικών του προσώπου.
- *Ειδικές ανάγκες για την οπτική πληροφορία.* Η διαπροσωπική επικοινωνία με video δεν απαιτεί συνήθως υψηλή ανάλυση. Είναι βασικό να προσφέρει την αίσθηση της επικοινωνίας μέσω της εικόνας. Για αυτό χρησιμοποιούνται συνήθως μικρότερες διαστάσεις όπως QCIF ενώ ο ρυθμός πλαισίων/sec είναι μειωμένος στα 10 ή στα 6 Hz.

Προκειμένου να καταστεί δυνατή η εκμετάλλευση των χαρακτηριστικών εισόδων τηλεσυνδιάσκεψης και να ξεπεραστούν τα προβλήματα των κωδικοποιητών ΔΜΣ-ΑΚ, εισάγεται ο κωδικοποιητής ΔΜΣ-ΑΚ με Περιοχές Ενδιαφέροντος. Η κωδικοποίηση με Περιοχές Ενδιαφέροντος παρέχει τα μέσα για τον καθορισμό περιοχών υψηλού / χαμηλού ενδιαφέροντος μεταβάλλοντας τα χαρακτηριστικά του κβαντιστή. Ο λόγος των επιφανειών υψηλού προς χαμηλού ενδιαφέροντος, ο οποίος στις περιπτώσεις τηλεσυνδιάσκεψης είναι ο λόγος block εικόνας κεφαλιού / ώμων προς block που δεν ανήκουν σε αυτά, μεταβάλλει το ποσοστό συμπίεσης. Είναι πλέον δυνατόν να επιτευχθεί μεγάλη μείωση των κωδικοποιημένων Block διατηρώντας παράλληλα επαρκή ποιότητα για εφαρμογές τηλεσυνδιάσκεψης. Με αυτήν την τεχνική είναι πλέον δυνατό να ξεπεραστεί το πρώτο μειονέκτημα του ΔΜΣ-ΑΚ. Η εισαγωγή μιας μνήμης πλαισίου εικόνας (Frame Buffer) μαζί με ένα διακόπτη  $S_2$ , υλοποιεί ένα ανοικτό βρόγχο πρόβλεψης [20] παρέχοντας τη δυνατότητα για ανίχνευση / εκτίμηση κίνησης με βάση τα αυθεντικά πλαίσια εικόνας και όχι με τα ανακατασκευασμένα με χρονική αντιστάθμιση. Συνεπώς, είναι δυνατόν να περιορισθεί η συσσώρευση των λαθών, η οποία αποτελεί το δεύτερο μειονέκτημα του σχήματος ΔΜΣ-ΑΚ. Επιπλέον επειδή οι ακμές ορίζονται να ανήκουν στην κατηγορία περιοχών υψηλού ενδιαφέροντος, αναμένεται να περιορισθεί το φαινόμενο του “φωτοστέφανου”, (ring effect) το οποίο εμφανίζεται στην κωδικοποίηση των ακμών σε χαμηλούς ρυθμούς.

Ένα block διάγραμμα του κωδικοποιητή ΔΜΣ-ΑΚ-ΠΕ εμφανίζεται στο σχήμα 6.5. Έχει προστεθεί μία μονάδα επιλογής των περιοχών ενδιαφέροντος (ΠΕ) η οποία λειτουργεί στο πεδίο του ΔΜΣ για ευκολότερη εισαγωγή στο σχήμα του κωδικοποιητή χωρίς να είναι απαραίτητη φάση προεπεξεργασίας της εικόνας. Η μονάδα ΔΜΣ (DCT) τροφοδοτεί την μονάδα ΠΕ με τα αυθεντικά πλαίσια εικόνας κατά την διάρκεια (intraframe coding), όπου διατίθενται μόνο τα σφάλματα πρόβλεψης. Ο διακόπτης  $S_2$  υλοποιεί τον προαναφερθέντα ανοικτό βρόγχο πρόβλεψης. Συνεπώς η μονάδα επιλογής ΠΕ τροφοδοτείται πάντα με τα πραγματικά pixels της εικόνας στη φάση interframe και intraframe κωδικοποίησης, η οποία στη συνέχεια ρυθμίζει τη λειτουργία του κβαντιστή. Η μονάδα επιλογής ΠΕ κατηγοριοποιεί τα blocks (B) μιας εικόνας είτε στην κλάση που απαιτείται υψηλή ποιότητα ανακατασκευής είτε στην κατηγορία χαμηλού ενδιαφέροντος. Η χρήση δύο κατηγοριών



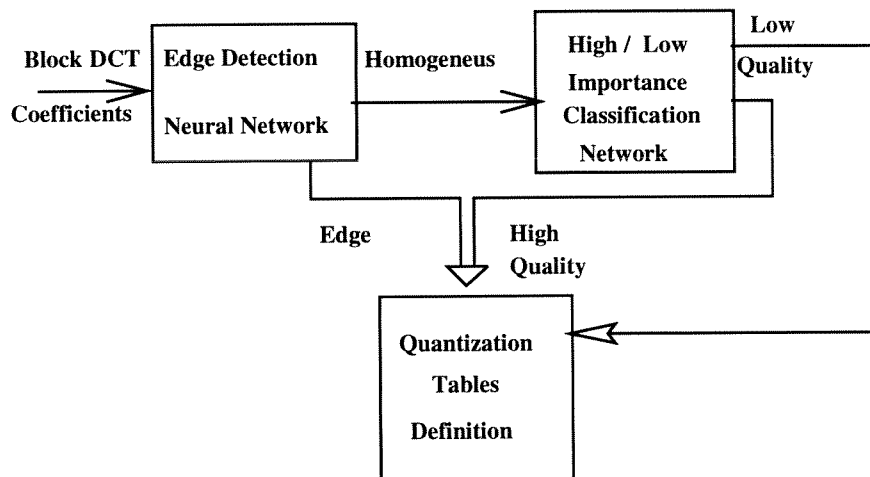
Σχήμα 6.5: Προσαρμοζόμενος Κωδικοποιητής ΔΜΣ-AK με Π.Ε

δεν περιορίζει τον αλγόριθμο. Στην ενότητα 6.6 εξετάζεται η περίπτωση για περισσότερες κατηγορίες. Πρέπει να τονισθεί ότι εφόσον η επιλογή των ΠΕ γίνεται σε κάθε πλαίσιο εικόνας στη φάση interframe coding ο αλγόριθμος ΔΜΣ-AK-ME μπορεί να βελτιώσει το σηματοθορυβικό λόγο των περιοχών υψηλού ενδιαφέροντος ακόμα και εάν ορισμένα Block ανήκαν σε περιοχές χαμηλού ενδιαφέροντος σε προηγούμενα πλαίσια εικόνας.

Η εισαγωγή περιοχών ενδιαφέροντος (ΠΕ) σε ένα κωδικοποιητή ΔΜΣ-AK μπορεί να συσχετιστεί με τον κωδικοποιητή με αντικείμενα όπου η διαδικασία της κατάτμησης αντιστοιχίζεται με την επιλογή ΠΕ. Η κωδικοποίηση των συμβατών αντικειμένων γίνεται με την επιλογή κατηγορίας υψηλού ενδιαφέροντος ενώ το αντίθετο συμβαίνει με τα ασύμβατα αντικείμενα. Η υλοποίηση της κατάτμησης των ΠΕ από την συνολική εικόνα γίνεται με την χρήση νευρωνικών δικτύων η οποία αναλύεται στην συνέχεια ενώ τα αποτελέσματα της κατάτμησης φαίνονται στην ενότητα 6.5. Τα υπόλοιπα τμήματα του διαγράμματος ροής 6.2 έχουν τα αντίστοιχα στο σχήμα 6.5.

## 6.4 Προσαρμοζόμενη Επιλογή Περιοχών Ενδιαφέροντος με Χρήση Νευρωνικών Δικτύων

Σε αυτή την ενότητα προτείνεται μία αρχιτεκτονική με νευρωνικά δίκτυα για την διαδικασία επιλογής ΠΕ σε εφαρμογές κωδικοποίησης σε χαμηλούς ρυθμούς  $\leq 64$ . Το προτεινόμενο σύστημα αποτελείται από μία αρχιτεκτονική δύο επιπέδων όπως φαίνεται στο σχήμα 6.6. Η συγκεκριμένη αρχιτεκτονική επιτελεί αφενός μια διαδικασία εύρεσης ακμών σε πρώτο επίπεδο και αφετέρου διαχωρισμού σε περιοχές υψηλού / χαμηλού ενδιαφέροντος στο δεύτερο επίπεδο. Η χρήση αρχιτεκτονικής δύο επιπέδων υιοθετήθηκε επειδή είχε καλύτερα αποτελέσματα από την αρχιτεκτονική ενός επιπέδου. Επιπλέον κάθε μια από τις αρχιτεκτονικές τροφοδοτείται με διαφορετικά σύνολα συντελεστών ΔΜΣ στην είσοδο των νευρωνικών δικτύων.



Σχήμα 6.6: Αρχιτεκτονική δύο επιπέδων με νευρωνικά δίκτυα

Το πρώτο επίπεδο αποτελείται από μία αρχιτεκτονική νευρωνικού δικτύου χωρίς ανάδραση το οποίο επιτελεί μια διαδικασία εύρεσης ακμών στο πεδίο της συχνότητας κατηγοριοποιώντας κάθε block εικόνας στις κατηγορίες “ακμή” και “μή ακμή”. Στην κατηγορία “μή ακμή” ανήκουν ομοιογενείς περιοχές οι οποίες δεν περιέχουν πληροφορία ακμών, ενώ στην κατηγορία “ακμή” ανήκουν περιοχές οι οποίες έχουν κάποια πληροφορία υψηλού συχνοτικού περιεχόμενου. Η διαδικασία ανίχνευσης ακμών μπορεί να πραγματοποιηθεί σε

ένα στάδιο προεπεξεργασίας με χρήση συμβατικών χωρικών τελεστών [14]. Υιοθετείται μια αποδοτικότερη διαδικασία η οποία είναι συμβατή με το σχήμα 6.6 και στην οποία τα blocks που περιέχουν ακμές επιλέγονται κατευθείαν από τους συντελεστές ΔΜΣ κατά την διάρκεια ενός σχήματος σαν το MPEG. Ειδικότερα το νευρωνικό δίκτυο χρησιμοποιείται για να εξάγει το υψηλό συχνοτικό περιεχόμενο και των τριών συντελεστών των έγχρωμων block εικόνας, δηλ. της φωτεινότητας ( $Y$ ) και των χρωματικών συντελεστών ( $C_b, C_r$ ). Ένα μεγάλο μέρος ή ίσως όλοι οι συντελεστές του ΔΜΣ πρέπει να χρησιμοποιηθούν για να γίνει δυνατή η ανίχνευση των ακμών προς όλες τις κατευθύνσεις. Συνεπώς ο αριθμός των εισόδων μπορεί να είναι μέχρι 192 ( $3 \times 64$ ) ενώ ο αριθμός των εξόδων είναι ίσος με τον αριθμό των κατηγοριών εξόδων.

Μια διαδικασία εκμάθησης με επίβλεψη (supervised learning) χρησιμοποιείται για την διαδικασία της ανίχνευσης ακμών. Ένα προκαθορισμένο σύνολο εκμάθησης από χαρακτηριστικές εικόνες επιλέγεται και εφαρμόζεται κάποιος συμβατικός τελεστής εύρεσης ακμών, π.χ laplace, μορφολογικός κ.λ.π., και στην συνέχεια τα αποτελέσματα αξιολογούνται εκ των υστέρων από ειδικούς για την βελτίωση της ποιότητας των ακμών. Τα blocks των ΔΜΣ της εικόνας μαζί με τον χαρακτηρισμό του κάθε block χρησιμοποιούνται για την εκμάθηση του δικτύου. Μετά τη διαδικασία εκμάθησης το δίκτυο είναι έτοιμο να κατηγοριοποιήσει όλα τα Blocks εικόνας, τα οποία είναι παρόμοια με αυτά τα οποία χρησιμοποιήθηκαν στην διαδικασία εκμάθησης, σε κατηγορίες “ακμής” και “μή ακμής”. Όσα ανήκουν στην πρώτη κατηγορία κατατάσσονται στις περιοχές υψηλού ενδιαφέροντος.

Εντούτοις εάν το block βρεθεί να ανήκει σε ομοιόμορφη περιοχή (“μή ακμή”) τροφοδοτείται ως είσοδος στην δεύτερη προτεινόμενη αρχιτεκτονική νευρωνικού δικτύου, το οποίο οριστικά το κατηγοριοποιεί σε περιοχή χαμηλού ή υψηλού ενδιαφέροντος. Αυτό το βήμα είναι απαραίτητο για να ληφθούν υπόψη και χαρακτηριστικά εκτός της συχνοτικής πληροφορίας π.χ. χρώμα, μέση τιμή φωτεινότητας, τα οποία μπορεί να είναι σημαντικά στις υπό εξέταση εφαρμογές.

Το δεύτερο επίπεδο αποτελείται από νευρωνικό δίκτυο παρόμοιο με αυτό του πρώτου επιπέδου με την έννοια ότι και αυτό χρησιμοποιεί τους συντελεστές ΔΜΣ στη μονάδα

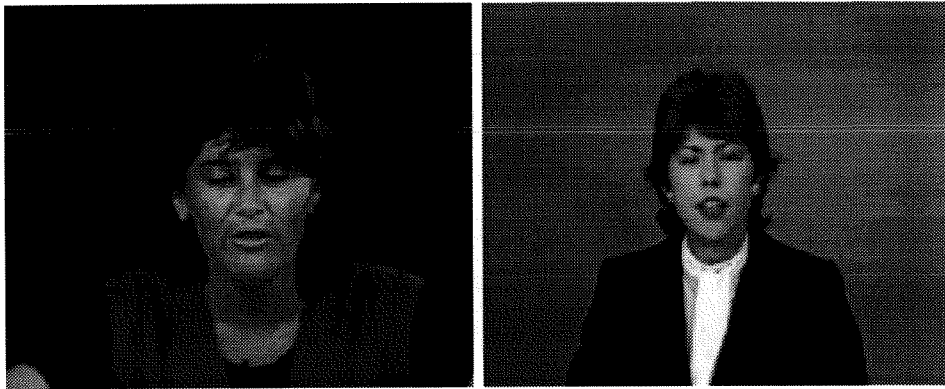


εισόδου. Ο αριθμός των χαρακτηριστικών εισόδου, σε αυτήν την περίπτωση όμως, είναι γενικά μικρότερος από τον αντίστοιχο της πρώτης αρχιτεκτονικής η οποία έπρεπε να ανιχνεύει ακμές προς διάφορες κατευθύνσεις. Ειδικότερα τα δεδομένα εισόδου που τροφοδοτούν το δίκτυο είναι η DC (συνεχής τιμή) και ένας μικρός αριθμός AC συντελεστών διατεταγμένων σε σειρά zig-zag για κάθε συνιστώσα της εικόνας. Διαδικασία εκμάθησης με επίβλεψη χρησιμοποιήθηκε για την εκμάθηση του νευρωνικού δικτύου χρησιμοποιώντας εικόνες από την υπό εξέταση εφαρμογή. Κάθε φορά που ένα block κατηγοριοποιείται σε περιοχή υψηλού ενδιαφέροντος επιλέγεται χβαντισμός με μικρό βήμα ενώ στην αντίθετη περίπτωση (περιοχή χαμηλού ενδιαφέροντος) επιλέγεται χβαντισμός με μεγάλο βήμα. Ο αλγόριθμος του διανυσματικού χβαντιστή με εκμάθηση, οποίος παρουσιάστηκε στην ενότητα 2.4 [16] χρησιμοποιήθηκε και για τις δύο αρχιτεκτονικές.

## 6.5 Θέματα Υλοποίησης - Αριθμητικά Αποτελέσματα

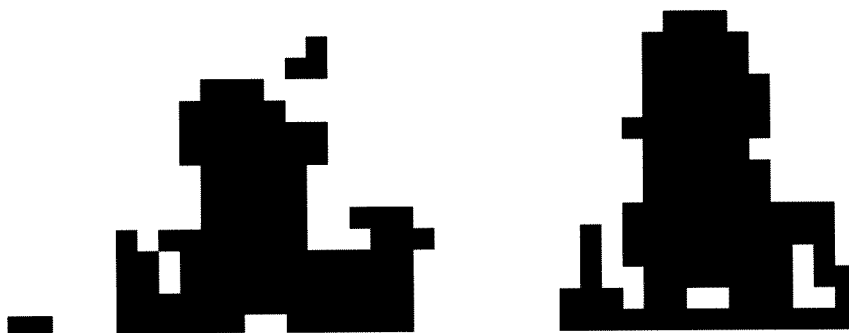
Η απόδοση του προτεινόμενου προσαρμοζόμενου ΔΜΣ-ΑΚ-ΠΕ αλγορίθμου μελετήθηκε χρησιμοποιώντας τις ακολουθίες Miss America και Claire σε ρυθμούς των 10 πλαισίων/sec. Τα πρώτα πέντε πλαίσια της ακολουθίας Miss America (π.χ το πλαίσιο # 7 φαίνεται στο σχήμα 6.7(α)) και τα πρώτα δύο της ακολουθίας Claire (π.χ το πλαίσιο # 7 φαίνεται στο σχήμα 6.7(β)) χρησιμοποιήθηκαν στο σύνολο εκμάθησης της αρχιτεκτονικής των νευρωνικών δικτύων που περιγράφηκε στην προηγούμενη ενότητα. Τα υπόλοιπα 145 και 148 πλαίσια εικόνων και των δύο ακολουθιών χρησιμοποιήθηκαν σαν σύνολο ελέγχου.

Μορφολογικοί τελεστές εφαρμόστηκαν στο σύνολο εκμάθησης ώστε να διοχετεύσουν το νευρωνικό δίκτυο του πρώτου επιπέδου με τις προεπιλεγμένες κατηγορίες “ακμής” / “μή ακμής”. Η επιλογή ομοιογενών block σε κατηγορίες υψηλού / χαμηλού ενδιαφέροντος έγινε με ανθρώπινη αρωγή για την εκμάθηση του νευρωνικού δικτύου του δεύτερου επιπέδου. Τα σχήματα 6.8(α) και 6.8(β) δείχνουν τις ακμές, που επιλέχθηκαν από την πρώτη αρχιτεκτονική, να ανήκουν σε ΠΕ στο πλαίσιο # 7 και των δύο ακολουθιών. Η dc και οι πρώτοι 31 ac συντελεστές σαρωμένοι κατά zig-zag από τις συνιστώσες της εικόνας χρησιμοποιήθηκαν ως διάνυσμα εισόδου. Φαίνεται ότι όλες οι σημαντικές ακμές



Σχήμα 6.7: (α) Frame #7      Σχήμα 6.7 (β) Frame #7

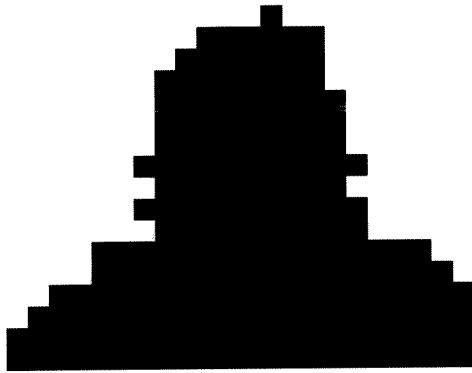
ανιχνεύτηκαν με μεγάλη ακρίβεια από το νευρωνικό δίκτυο. Τα σχήματα 6.9(α) και 6.9(β) δείχνουν την τελική κατηγοριοποίηση (δηλ. μετά την εφαρμογή και του δεύτερου νευρωνικού δικτύου) των εικόνων 6.7(α) και 6.7(β) σε περιοχές υψηλού (μαύρη περιοχή) /χαμηλού (άσπρη περιοχή) ενδιαφέροντος. Η dc τιμή και 8 ac συντελεστές σαρωμένοι κατά zig-zag χρησιμοποιήθηκαν για διάνυσμα εισόδου. Η ποιότητα της τελικής κατηγοριοποίησης είναι πολύ καλή συγκρίνοντας τους χάρτες κατηγοριοποίησης με τις αρχικές εικόνες.



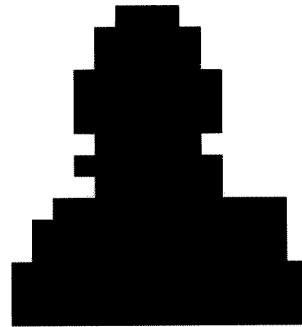
Σχήμα 6.8: (α)

Σχήμα 6.8 (β)

Η εξαχθείσα πληροφορία από την επιλογή των ΠΕ ενσωματώθηκε αργότερα σε ένα κωδικοποιητή ΔΜΣ-ΑΚ όπως του σχήματος 6.5 ο οποίος υλοποιήθηκε χρησιμοποιώντας συντακτικό MPEG-1 [22]. Οι χάρτες κατηγοριοποίησης κωδικοποιήθηκαν με 1-bit για κάθε block της εικόνας και αποθηκεύθηκαν στην επικεφαλίδα κάθε εικόνας. Σε κατάσταση



Σχήμα 6.9: (α)



Σχήμα 6.9 (β)

λειτουργίας μεταβλητού ρυθμού (variable bitrate) ο αποκωδικοποιητής ήταν δυνατόν να αποκωδικοποιεί τον πίνακα κατηγοριοποίησης και κατάλληλα να ανακατασκευάζει το κάθε block της εικόνας. Σε κατάσταση λειτουργίας σταθερού ρυθμού ένας ελεγκτής κβαντισμού και ρυθμού μετάδοσης ανάλογος του MPEG-2 [23] ενεργοποιήθηκε με διαφορετικούς απομονωτές για τις περιπτώσεις περιοχών υψηλού / χαμηλού ενδιαφέροντος.

Ο αλγόριθμος ΔΜΣ-ΑΚ-ΠΕ εφαρμόστηκε στην συνέχεια και στις δύο ακολουθίες σε διάφορους ρυθμούς από 15 έως 60 kbit/sec. Ο πίνακας 6.1 δείχνει το μέσο Σηματοθορυβικό Λόγο από Ακρη Σε Ακρη (Peak Signal to Noise Ratio (PSNR)) (ΣΛΑΣΑ) από την λειτουργία του κωδικοποιητή ΔΜΣ-ΑΚ-ΠΕ στην ακολουθία Miss America. Επιπλέον ο πίνακας 6.1 δείχνει το μέσο αριθμό bits ως ποσοστό των συνολικών bits που δαπανήθηκαν στον τρόπο λειτουργίας με αυθεντικές εικόνες (Intraframe) και εικόνες σφαλμάτων (Interframe). Εύκολα βλέπουμε ότι ο αλγόριθμος ΔΜΣ-ΑΚ-ΠΕ έχει βελτιώσει τον ΣΛΑΣΑ πάνω από 4 db για τον τρόπο λειτουργίας intraframe ενώ για τον τρόπο λειτουργίας interframe η βελτίωση μεταβάλλεται από 2 - 4 db ανάλογα με τον ρυθμό μετάδοσης. Ο πίνακας 6.2 δείχνει τα αντίστοιχα αποτελέσματα για την ακολουθία Claire τα οποία δείχνουν παρόμοια συμπεράσματα. Επιπλέον τα αποτελέσματα της βελτίωσης για ΣΛΑΣΑ φαίνονται περαιτέρω στο σχήμα 6.10. Το σχήμα 6.11 δείχνει την μεταβολή του ΣΛΑΣΑ σε σχέση με τον αριθμό του πλαισίου για την ακολουθία Miss America ενώ το σχήμα 6.12 δείχνει το ίδιο για την ακολουθία Claire. Είναι ενδιαφέρον να παρατηρήσουμε ότι το

Bitrate kbps	ΔΜΣ-ΑΚ-ΠΕ						ΔΜΣ-ΑΚ	
	Intra			Inter			Intra	Inter
	$\frac{\# \text{ΠΕ bits}}{\# \text{Συν. Bits}}$	%	Μεσ. PSNR	$\frac{\# \text{ΠΕ bits}}{\# \text{Συν. Bits}}$	%	Μεσ. PSNR	Μεσ. PSNR	Μεσ. PSNR
16	58.1		38.3	30.2		35.5	34.2	33.7
20	59.8		39.3	38.6		35.7	34.4	33.9
25	59.8		39.3	44.1		36.9	34.7	32.7
30	61.2		40.5	55.0		36.7	35.2	33.4
35	59.7		39.3	42.9		37.5	35.2	34.1
40	61.1		40.5	54.6		38.3	35.7	34.5
45	60.5		39.9	42.2		39.2	35.7	36.0
50	61.2		40.5	48.6		39.1	36.1	36.3
55	60.5		39.9	44.1		39.6	35.9	35.3
60	61.2		40.5	41.4		39.3	36.3	36.6

Πίνακας 6.1: Μέση Τιμή του PSNR και του ποσοστού των bits που χρησιμοποιούνται στον ΔΜΣ-ΑΚ-ΠΕ σε σχέση με τον ΔΜΣ-ΑΚ (Ακολουθία Miss America )

60 % των block σε λειτουργία intraframe ανήκουν στην κατηγορία υψηλού ενδιαφέροντος ενώ στον τρόπο λειτουργίας interframe το ποσοστό κυμαίνεται από 30% έως 50 %. Το σχήμα 6.13 δείχνει πλαίσια εικόνας ανακατασκευασμένα σε κωδικοποίηση intraframe και interframe σε ρυθμούς 20 και 60 kbit/sec. Οι αντίστοιχες ανακατασκευασμένες εικόνες για τον συμβατικό κωδικοποιητή /αποκωδικοποιητή ΔΜΣ-ΑΚ στα 20 kbit/sec φαίνονται στο σχήμα 6.14. Είναι φανερό ότι οι εικόνες από τον αποκωδικοποιητή ΔΜΣ-ΑΚ-ΠΕ είναι πολύ καλύτερες από αυτές του ΔΜΣ-ΑΚ. Το σχήμα 6.15 δείχνει ανακατασκευασμένες εικόνες της ακολουθίας Claire σε ρυθμό 16 kbit/sec (Πλαίσια 1, 99 και μεγεθυμένο 99 στην περιοχή του προσώπου). Πρέπει να σημειωθεί ότι τα προαναφερθέντα αποτελέσματα δείχνουν ότι ο προτεινόμενος αλγόριθμος ΔΜΣ-ΑΚ-ΠΕ επιτυγχάνει τιμές ΣΛΑΣΑ και ανακατασκευασμένες εικόνες παρόμοιες με αυτές των λεγόμενων τεχνικών κωδικοποίησης δεύτερης γενιάς. Αυτό μπορεί να επιβεβαιωθεί συγκρίνοντας τα παραπάνω αποτελέσματα της ακολουθίας Claire στα 16 kbit/sec με αυτά που παρουσιάζονται [7], στα Σχήματα 9 - 14, που αναφέρονται στο ίδιο πείραμα.

Bitrate kbps	ΔΜΣ-ΑΚ-ΠΕ						ΔΜΣ-ΑΚ	
	Intra			Inter			Intra	Inter
	$\frac{\# \text{ΠΕ bits}}{\# \text{ΠΕ Bits}}$	%	Μεσ. PSNR	$\frac{\# \text{bits}}{\# \text{Συν. Bits}}$	%	Μεσ. PSNR	Μεσ. PSNR	Μεσ. PSNR
16	54.1		36.6	38.7		34.0	32.9	32.1
20	55.4		37.7	47.9		35.5	33.2	32.4
25	55.6		37.2	51.4		35.6	33.5	32.1
30	56.9		37.4	60.1		36.7	34	32.9
35	55.5		37.7	51.3		37	33.8	32.7
40	56.9		39	60		38.1	34.3	33.4
45	56.2		38.3	46.4		37.8	34.3	33.6
50	56.9		39	51.7		38.4	34.6	34.5
55	56.3		38.3	48.9		37.6	34.4	34.0
60	57		39	39.7		38.7	34.8	34.7

Πίνακας 6.2: Μέση Τιμή του PSNR και του ποσοστού των bits που χρησιμοποιούνται στον ΔΜΣ-ΑΚ-ΠΕ σε σχέση με τον ΔΜΣ-ΑΚ ( Ακολουθία Claire )

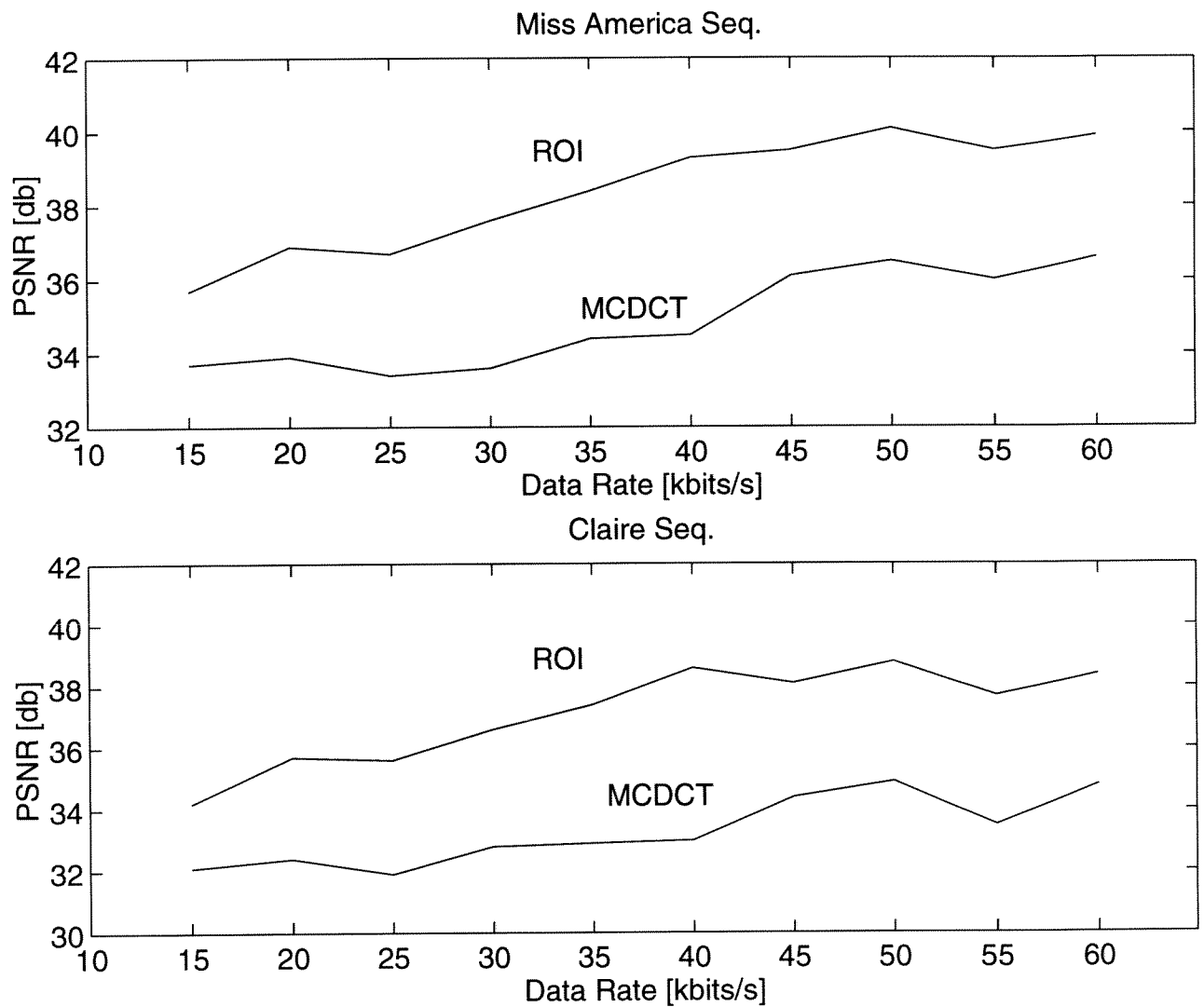
## 6.6 Συμπεράσματα - Θέματα για περαιτέρω έρευνα

Μια προσαρμοζόμενη τεχνική η οποία καθορίζει περιοχές ενδιαφέροντος σε μία ακολουθία εικόνων για κωδικοποίηση σε χαμηλούς ρυθμούς παρουσιάστηκε σε αυτό το κεφάλαιο. Μια αρχιτεκτονική με νευρωνικά δίκτυα σχεδιάστηκε για επιλογή των ΠΕ με βάση τους συντελεστές ΔΜΣ του block κάθε εικόνας. Ο αλγόριθμος υλοποιήθηκε χρησιμοποιώντας συντακτικό του MPEG-1 με σηματοθρομβικούς λόγους και ποιότητα ανακατασκευής συγκρίσιμη με ( και μερικές φορές καλύτερες από ) αυτή των υπάρχοντων τεχνικών κωδικοποίησης σε χαμηλούς ρυθμούς. Η προτεινόμενη αρχιτεκτονική των δύο επιπέδων φάνηκε να είναι ικανή να κατηγοριοποιήσει block εικόνας σε περιοχές χαμηλού / υψηλού επιπέδου και επιτρέποντας στην συνέχεια κβαντισμό με διαφορετικό πίνακα κβαντισμού για κάθε κατηγορία. Η κατηγοριοποίηση επιτυγχάνεται με χρήση νευρωνικών δικτύων τα οποία εκπαιδεύονται με τεχνικές με επίβλεψη χρησιμοποιώντας μικρό αριθμό από πλαίσια εικόνας.

Σε εφαρμογές τηλεσυνδιάσκεψης που εφαρμόστηκε ο αλγόριθμος ΔΜΣ-ΑΚ-ΠΕ διαχωρίστηκαν οι περιοχές του κεφαλιού-ώμων από το φόντο. Έρευνα στην ανάλυση εικόνων προσώπων εστιάζει στη μοντελοποίηση των προσώπων με χωροδικτυώματα και μονάδες

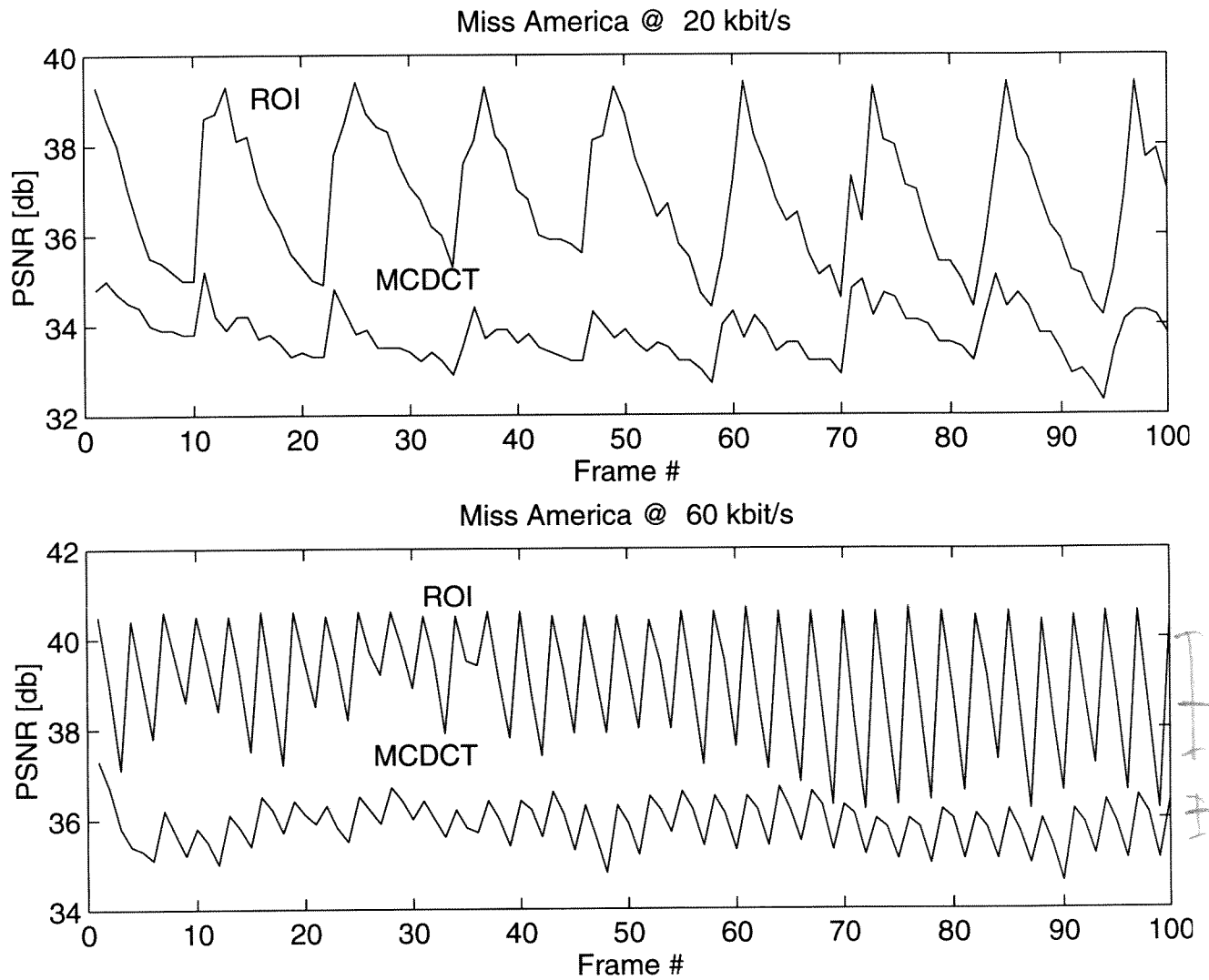
χαρακτηριστικών (action units) [9] για τον υπολογισμό της τοπικής κίνησης του προσώπου [25]. Ιδιαίτερου ενδιαφέροντος είναι η κίνηση των ματιών και του προσώπου. Είναι γνωστό [3] ότι κοιτάζοντας την εικόνα ενός προσώπου οι παρατηρητές εστιάζουν κυρίως στα μάτια (58 % του χρόνου ) και στη συνέχεια στο στόμα (13 % του χρόνου ). Οι υπόλοιπες περιοχές αθροίζονται σε 1 % του συνολικού χρόνου παρατήρησης.

Συνεκτιμώντας αυτά τα γεγονότα είναι δυνατό να επεκταθεί η τεχνική ΔΜΣ-ΑΚ-ΠΕ να δημιουργεί και να κωδικοποιεί περισσότερες από δύο κατηγορίες περιοχών. Συνεχίζοντας την διαδικασία επιλογής ΠΕ, μια περαιτέρω διαδικασία επιλογής περιοχών ύψιστης σημασίας (περιοχή ματιών), περιοχών υψηλότερης σημασίας (περιοχή στόματος) είναι δυνατή. Το δίκτυο που περιγράφηκε στην ενότητα 6.4 μπορεί να επεκταθεί σε μία ιεραρχική δομή χρησιμοποιώντας δύο επιπλέον επίπεδα δικτύων μετά από τα προηγούμενα δύο. Το πρώτο θα εξαγάγει την περιοχή των ματιών ενώ το δεύτερο θα εξάγει την περιοχή του στόματος από τις προηγούμενες επιλεγμένες περιοχές. Η εκπαίδευση των νευρωνικών δικτύων που βασίζεται τόσο σε συντελεστές ΔΜΣ καθώς και σε μονάδες χαρακτηριστικών (action units) είναι ένα πεδίο της παρούσας και της μελλοντικής έρευνας.



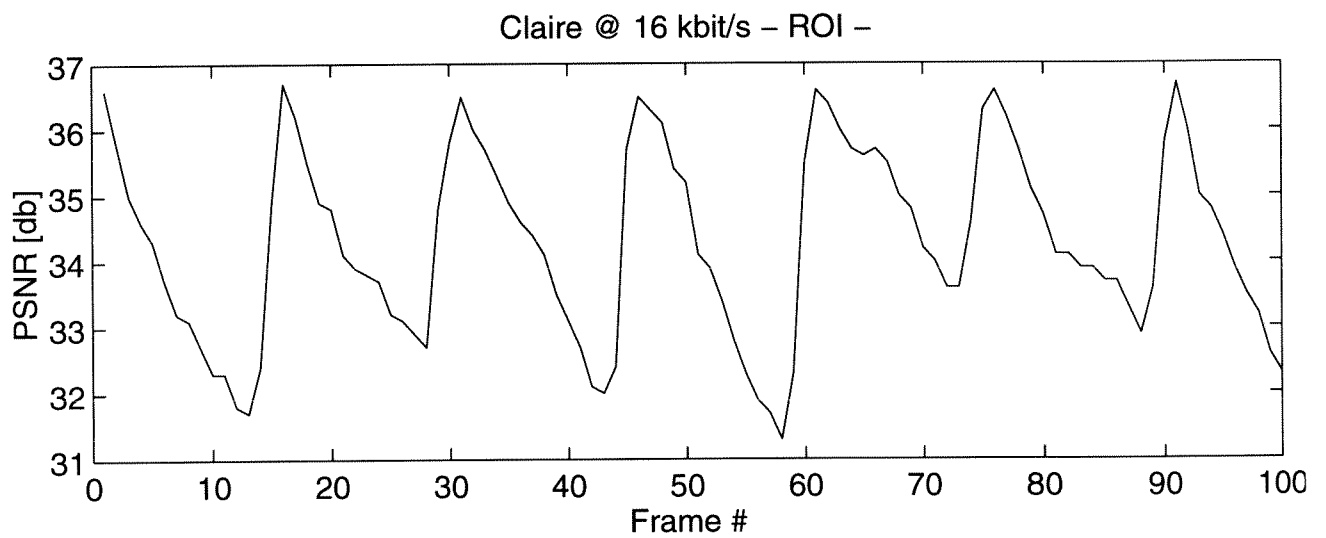
Σχήμα 6.10: Διάγραμμα PSNR ως προς bit rate για τις ακολουθίες Miss America και Claire χρησιμοποιώντας τον κωδικοποιητή ΔΜΣ-ΑΚ-ΠΕ (ROI MC-DCT) και τον κλασικό ΔΜΣ-ΑΚ (MC-DCT) κωδικοποιητή

CIF η QCIF

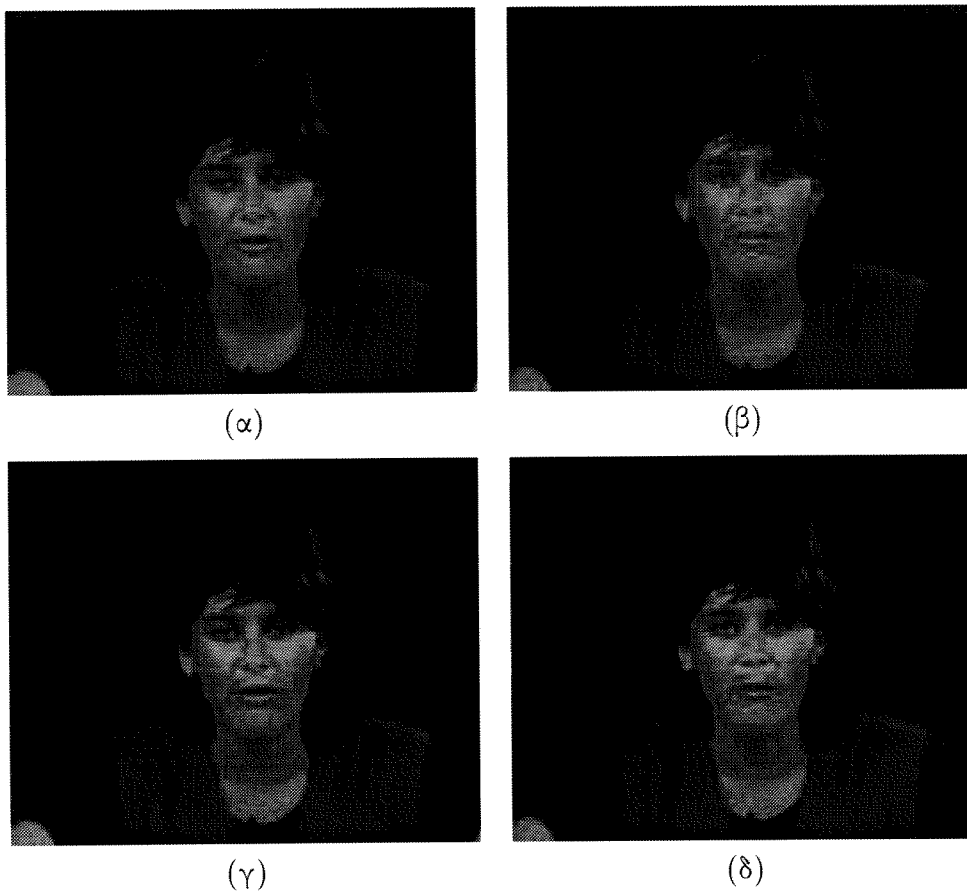


Σχήμα 6.11: Διάγραμμα PSNR ως προς Αρ. Πλαισίου (Frame Number) για την ακολουθία Miss America στα 20 και 60 kbit/sec χρησιμοποιώντας τον κωδικοποιητή ΔΜΣ-ΑΚ-ΠΕ (ROI MC-DCT) και τον κλασικό ΔΜΣ-ΑΚ (MC-DCT) κωδικοποιητή

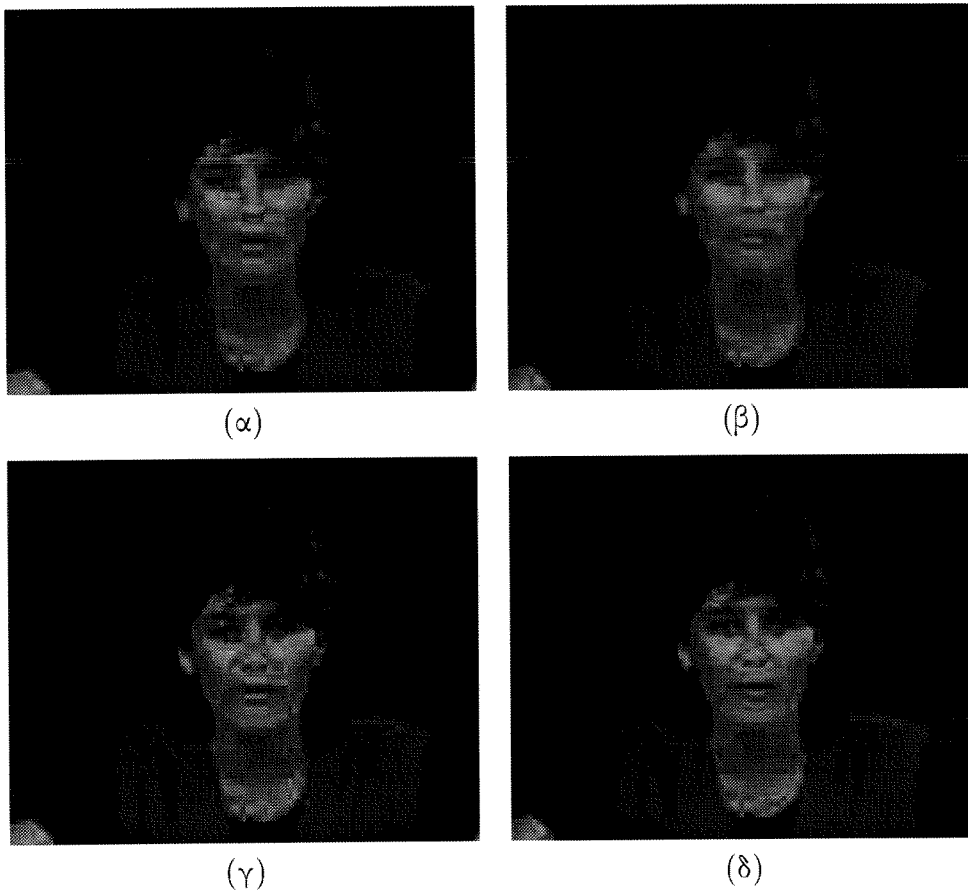




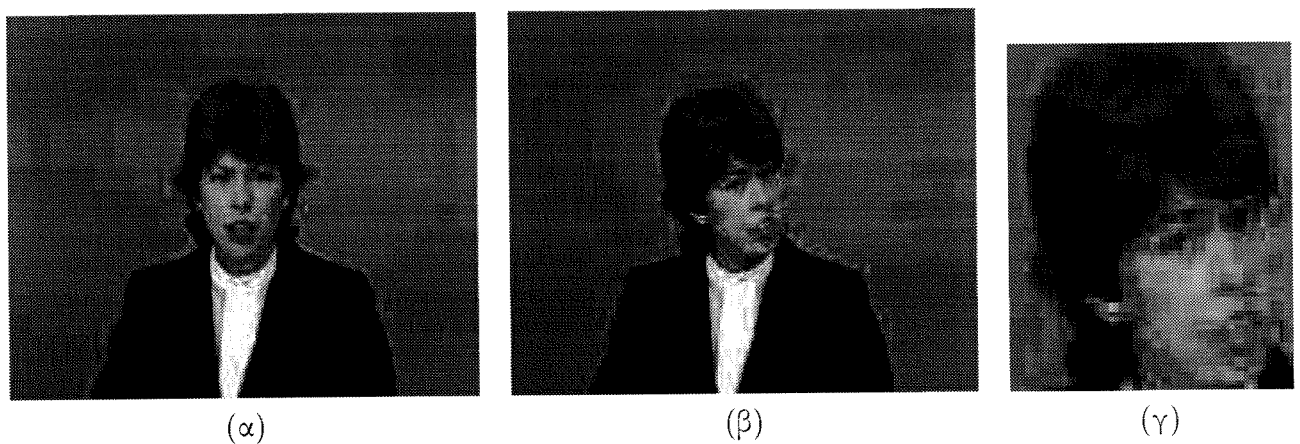
Σχήμα 6.12: Διάγραμμα PSNR ως προς Αρ. Πλαισίου (Frame Number) για την ακολουθία Claire στα 16 kbit/sec χρησιμοποιώντας τον κωδικοποιητή ΔΜΣ-ΑΚ-ΠΕ (ROI MC-DCT) και τον κλασικό ΔΜΣ-ΑΚ (MC-DCT) κωδικοποιητή



Σχήμα 6.13: Αποκωδικοποιημένα Πλαίσια εικόνας της ακολουθίας Miss America χρησιμοποιώντας τον κωδικοποιητή ΔΜΣ-ΑΚ-ΠΕ (ROI-MC-DCT) στα 20 kbps (α) intraframe, (β) interframe και στα 60 kbis/sec (γ) intraframe, (δ) interframe.



Σχήμα 6.14: Αποκωδικοποιημένα Πλαίσια εικόνας της ακολουθίας Miss America χρησιμοποιώντας τον κωδικοποιητή ΔΜΣ-ΑΚ (MC-DCT) στα 20 kbps (α) intraframe, (β) interframe και στα 60 kbis/sec (γ) intraframe, (δ) interframe.



Σχήμα 6.15: Αποκωδικοποιημένα Πλαίσια εικόνας της ακολουθίας Claire στα 16 kbit/sec χρησιμοποιώντας ΔΜΣ-ΑΚ-ΠΕ (ROI-MC-DCT) ; (α) Αρ. Πλαισίου 1 (β) Αρ. Πλαισίου 99 (β) μεγέθυνση στην περιοχή του προσώπου στον Αρ. Πλαισίου 99

## Αναφορές

- [1] K. Aizawa, C. S. Choi, H. Harashima and T. S. Huang, “Human facial motion analysis and syntesis with application to model based coding”, in *Motion Analysis and Image Sequence Processing*, M. I. Sezan and R.L. Lagendijk, eds, Norwell, MA: Kluwer 1993.
- [2] K. Aizawa and T. Huang, “Model-Based Image Coding: Advanced Video Coding Techniques for Very Low Bitrate Applications”, *Proc. of the IEEE*, vol. 83, no. 2, pp. 259-271, Feb. 1995.
- [3] M. Argyle and M.Cook, “Gaze and Mutual Gaze”, *Cambridge Univ. Press*, 1976
- [4] R. Baker, “Waveform Based Very Low Rate Video Coding”, Keynote address presented at the *Intl. Note Workshop Very Low Bitrate Video Compression* (Urbana IL), May 1, 1993.
- [5] C. S. Choi, K. Aizawa, H. Harashima and T. Takebe, “Analysis and synthesis of facial imagr sequences in model-based image coding”, *IEE Trans, Circ. and Syst: Video Tech.*, vol. 4, pp. 257-275, Sept. 1994.
- [6] N. Diehl, “Model-based Image sequence coding, in *Motion Analysis and Image Sequence Processing*, M. I. Sezan and R.L. Lagendijk, eds, Norwell, MA: Kluwer 1993.
- [7] T. Ebrahimi, E. Reusens and W. Li, “New Trends in Very Low Bitrate Coding”, *Proc. of the IEEE*, vol. 83, no. 6, June 1995.

- [8] P. Gerken, "Object oriented analysis-synthesis coding based on moving two-dimensional objects", *IEEE Trans. Circ. and Syst: Video Tech.*, vol. 4, pp. 228-237, Sept. 1994.
- [9] P. Ekman and W. V. Friesen, "Facial Image Coding System", *Consulting Psychologist Press*, 1977.
- [10] R. Forchheimer and T. Kronander, "Image coding from waveforms to animation" *IEEE Trans. Acoust. Speech Sign. Proc.*, vol. 37, no. 12, pp. 2008-2023, Dec. 1993.
- [11] CCITT Recommendation H.261, "Video Codec For Audiovisual at  $p \times 64$  kb/s," Geneva 1990.
- [12] M. Hotter, "Object oriented analysis-synthesis coding based on moving two-dimensional objects", *Signal Proc: Image Comm.*, vol. 2, no. 4, pp. 409-428, Dec 1990.
- [13] D. R. Hush and B. G. Horne, "What's new Since Lippmann", *IEEE Acoustics, Speech and Signal Processing Magazine*, vol. 10, no. 1, January 1993.
- [14] A.K.Jain, "Fundamentals of Digital Image Processing", Eng. Cliffs, N. Jersey, *Prentice Hall*, 1989.
- [15] R. Koch, "Dynamic 3-D scene analysis through synthesis feedback control", *IEEE Trans. Patt. Analysis and Masc. Intel.*, vol. 15, pp. 556-568, June 1993.
- [16] T. Kohonen, "Improved Versions of Learning Vector Quantization", *Proc. of the Inter. Joint Conf. on Neural Networks*, pp. I 545 - 550, June 1990.
- [17] D. Kalogeras and S. Kollias, "Adaptive Windowed Jpeg Coding Of Still Images Using Neural Networks", *1995 IEEE Workshop on Nonlinear Signal and Image Processing*, vol. 1, pp. 340-343, Halkidiki 1995.
- [18] M. Kunt, A. Ikonopoulos and M. Kocher, "Second-generation image coding techniques", *Proc. IEEE*, vol. 73, pp. 549-574, 1985.

- [19] H. Li, A. Lundmark and R. Forchheimer, "Image Sequence Coding at Very Low Bitrates: A Review", *IEEE Trans. on Image Proc.*, vol. 3, no. 5, pp. 589 - 609, Sept. 1994.
- [20] H. Li, A. Lundmark and R. Forchheimer, "Triangle-based Motion Compensation And Its Application To Very Low Bit Rate Coding", submitted to *IEEE Trans. On Image Proc.*
- [21] H. G. Musmann, M. Hotter and J. Ostermann, "Object-oriented analysis synthesis coding of moving images", *Signal Proc: Image Comm.*, vol. 1, pp. 339-356, June 1994 .
- [22] ISO/CD 11172-2, "Coding Of Moving Pictures And Associated Audio For Digital Storage Media At Up To About 1.5 Mbps.", March 1991.
- [23] ISO/IEC 13818-2, "Generic Coding of Moving Pictures And Associated Audio. Recommendation H.262, Committee Draft" , May 1994.
- [24] Int. Org. for Standard., ISO/IEC JTC1/SC29/WG11, "First draft of MPEG-4 requirements", March 1994.
- [25] A. M. Tekalp, "Digital Video Processing", Eng. Cliffs, N. Jersey, *Prentice Hall*, 1995.
- [26] E. Nguyen and C. Labit. "Adaptive Region - Based Quantization in Subband Coding Using A Priori Levels Of Interest", *Proc. of the Picture Coding Symposium PCS'94*, September 1994.
- [27] J. Ostermann, "An Analysis-synthesis coder based on moving flexible 3-D objects", *Proc. Pict. Cod. Symp.*, Lausanne, Switzerland, Mar. 1993.
- [28] D. Pearson, "Developments In Model-Based Video Coding", *Proc. of the IEEE*, vol. 83, no. 6, pp. 892-906, June 1995.
- [29] A.B. Watson, "Visually optimal DCT quantization matrices for individual images" in J.A. Storer and M. Cohn, Eds., *Proc. Data Compression Conf.*, IEEE Computer Society Press, 1993. pp. 178-187.

# Συμπεράσματα της διατριβής

Στην εργασία αυτή αναπτύξαμε νέες μεθόδους στην περιοχή της ανάλυσης εικόνων, και παρουσιάσαμε κατάλληλες αρχιτεκτονικές νευρωνικών δικτύων για αναγνώριση και ταξινόμηση διδιάστατων σημάτων.

Στο **Μέρος I** παρουσιάστηκαν τα νευρωνικά δίκτυα, με έμφαση στο πολυεπίπεδο perceptron και τον διανυσματικό κβαντιστή με επίβλεψη. Το πρόβλημα της ανάλυσης / κατηγοριοποίησης εικόνας είναι ένα πρόβλημα αναγνώρισης προτύπων και η λύση του επηρεάζεται σημαντικά από τα χαρακτηριστικά που επιλέγονται ως πρότυπα. Η ύπαρξη ενός μοντέλου περιγραφής αντικειμένων και φυσικών σκηνών μελετήθηκε στο συγκεκριμένο πρόβλημα. Τα fractals και κατ' επέκταση ο πολυκαναλικός κλασματικός πίνακας είναι ένα υποψήφιο μοντέλο για μοντελοποίηση ακόμη και έγχρωμων αντικειμένων. Δείχθηκε ότι ο πολυκαναλικός κλασματικός πίνακας οδηγεί σε μείωση του σφάλματος μοντελοποίησης. Επιβεβαιώθηκε ο ισχυρισμός ότι η αναπαράσταση στο χρωματικό σύστημα RGB υπερτερεί έναντι του YCrCb. Έγιναν πειράματα κατάτμησης τεχνητής συνθετικής εικόνας και παρουσιάστηκαν συγκριτικά αποτελέσματα με ένα πολυμοντελικό σύστημα που βασίζεται τόσο σε fractals όσο και σε συσχετίσεις τρίτης τάξης. Συνεχίζοντας αντιμετωπίζουμε το πρόβλημα της ανάλυσης εικόνων μεγάλων διαστάσεων. Η εισαγωγή της πολυδιακριτής ανάλυσης σημάτων (subband analysis) δίνει την δυνατότητα για παράλληλα σχήματα επεξεργασίας των εικόνων όχι μόνο στις αρχικές τους διαστάσεις αλλά σε πολλαπλά διακριτικά επίπεδα. Είναι δυνατόν η πολυδιακριτή ανάλυση να ληφθεί υπόψη στην κατασκευή νευρωνικών αρχιτεκτονικών. Οι νέες αρχιτεκτονικές μπορούν να ξεκινούν την ανάλυση της εικόνας σε μικρότερο διακριτικό επίπεδο και σε περίπτωση μη ικανοποιητικής

λειτουργίας να συνεχίζεται η ανάλυση σε μεγαλύτερα διακριτικά επίπεδα εκμεταλλευόμενα την υπάρχουσα γνώση. Το πιθανοτικό δίκτυο παρουσιάστηκε στη συνέχεια και προτάθηκε κατόπιν το ιεραρχικό πιθανοτικό δίκτυο, μια αρχιτεκτονική που υλοποιεί την πολυδιακριτική ανάλυση. Το ιεραρχικό πιθανοτικό δίκτυο έχει μικρότερες απαιτήσεις μνήμης και παρουσιάζει πολύ καλές επιδόσεις στην διάρκεια των πειραμάτων. Στη συνέχεια γίνεται σύγκριση με αρχιτεκτονική πολυεπίδου perceptron. Το ιεραρχικό πιθανοτικό δίκτυο είναι ταχύτερο από το αντίστοιχο πολυδρακρικό perceptron, οι δε επιδόσεις κατά τι υποδεέστερες.

Τέλος στο **Μέρος II** παρουσιάστηκαν αρχικά οι βασικές αρχές ενός συστήματος κωδικοποίησης / συμπίεσης εικόνων. Έγινε εκτενής αναφορά στα χαρακτηριστικά των διεθνών προτύπων H.261, MPEG-1 και MPEG-2 των οποίων κοινό χαρακτηριστικό είναι η δομή του ΔΜΣ-ΑΚ. Στη συνέχεια εξετάστηκε το σημαντικό πρόβλημα της κωδικοποίησης / συμπίεσης εικόνων σε πολύ χαμηλούς ρυθμούς  $\leq 64$  kbit/sec ώστε να είναι εφικτή η μετάδοση από κοινές τηλεφωνικές γραμμές. Η τηλεσυνδιάσκεψη είναι μια εφαρμογή που προσελκύει το μεγαλύτερο ενδιαφέρον. Οι τεχνικές ΚΜΜ αποτελούν μία προσέγγιση για την αντιμετώπιση του προβλήματος και έχουν δείξει ικανοποιητικά αποτελέσματα σε περιβάλλον τηλεσυνδιάσκεψης. Οι τεχνικές ΚΜΜ δεν παρουσιάζουν τις αδυναμίες των διεθνών προτύπων στο συγκεκριμένο πρόβλημα οι οποίες αναφέρονται στην συνέχεια. Μία δεύτερη προσέγγιση βασίζεται στους ΔΜΣ-ΑΚ με σκοπό αφενός την άμβλυση των προαναφερθέντων μειονεκτημάτων και αφετέρου την εκμετάλλευση των χαρακτηριστικών των εικόνων τηλεσυνδιάσκεψης. Η έννοια των περιοχών ενδιαφέροντος στην περίπτωση των εικόνων τηλεσυνδιάσκεψης παρέχει την δυνατότητα καθορισμού ποιότητας σε περιοχές οι οποίες είναι σημαντικές για υποκειμενικά καλή οπτική αντίληψη της εικόνας. Η έννοια των περιοχών ενδιαφέροντος μπορεί να συσχετισθεί με τις τεχνικές ΚΜΜ. Οι ομοιότητες είναι εμφανείς από την σύγκριση των διαγραμμάτων των κωδικοποιητών με περιοχές και των κωδικοποιητών με μοντέλο. Ο μηχανισμός για την επιλογή των περιοχών ενδιαφέροντος γίνεται με την βοήθεια νευρωνικών δικτύων. Επιλέχθηκε αρχιτεκτονική δύο επιπέδων η οποία σε πρώτη φάση επιλέγει τις ακμές που ορίζουν τα αντικείμενα μέσα στην εικόνα ενώ σε δεύτερη φάση γίνεται επιλογή ειδικών ομοιόμορφων περιοχών οι οποίες



έχουν προδιαγραφεί κατά την φάση της εκπαίδευσης. Η φάση της επιλογής δημιουργεί εικόνες απεικονιστικής κατάτμησης των ενδιαφερόντων περιοχών. Με βάση αυτούς τους πίνακες ρυθμίζεται η λειτουργία του κβαντιστή στο σχήμα κωδικοποίησης. Οι περιοχές υψηλού (χαμηλού) ενδιαφέροντος κωδικοποιούνται με καλή (κακή) ποιότητα. Στην συνέχεια έγιναν πειράματα με πραγματικές ακολουθίες εικόνων. Ο κωδικοποιητής με περιοχές ενδιαφέροντος προσφέρει υψηλότερη ποιότητα σε σχέση με τους κλασικούς κωδικοποιητές τόσο σε οπτική απεικόνιση όσο και με την μορφή PSNR. Επιπλέον από συγκρίσεις που γίνονται με άλλους κωδικοποιητές δεύτερης γενιάς πολύ χαμηλών ρυθμών φαίνεται ότι ο προτεινόμενος αλγόριθμος είναι εξαιρετικά συναγωνιστικός και μερικές φορές καλύτερος.