

Πολυτεχνείο Κρήτης
Τμήμα Ηλεκτρονικών Μηχανικών και
Μηχανικών Ηλεκτρονικών Υπολογιστών



Διπλωματική εργασία

«Μετατροπή της χωροχρονικής ανάλυσης MPEG με χρήση
διανυσμάτων κίνησης»

Χανιά, Μάρτιος 2005

Ζουρίδης Γιώργος

Επιβλέπων Καθηγητής: Ζερβάκης Μιχάλης (καθηγητής)

Εξεταστική Επιτροπή: Πετράκης Ευριπίδης (καθηγητής)
Χριστοδουλάκης Σταύρος (καθηγητής)

Αρχικά θα ήθελα να ευχαριστήσω τον επιβλέποντα καθηγητή της εργασίας μου κ. Ζερβάκη Μιχάλη, για την συνεργασία και την ιδιαίτερη κατανόηση του όλη τη διάρκεια της εργασίας αυτής. Επίσης θα ήθελα να ευχαριστήσω όλους εκείνους που λίγο ή πολύ με υποστήριξαν σε όλη μου την πορεία στο Πολυτεχνείο, και που χωρίς αυτούς τίποτα δεν θα ήταν εύκολο. Επίσης θα ήθελα να ευχαριστήσω τους γονείς μου για τα πράγματα αυτά που μου προσέφεραν και είναι αδύνατο να περιγράψω, καθώς και για την πίστη τους σε εμένα.

Περιεχόμενα

Περιεχόμενα.....	3
Περίληψη.....	5
Εισαγωγή.....	6
Συστήματα Mpeg.....	9
Γενικά.....	9
Χρονική κωδικοποίηση(Inter-frame coding).....	9
Χωρική κωδικοποίηση(Intra-frame coding).....	10
Διαβάθμιση(Scalability).....	13
Υπολογισμός Διανυσμάτων κίνησης κατά Horn-Schunk.....	15
Χωρική μετατροπή της ανάλυσης.....	20
Υποδειγματοληψία(Down-Sampling).....	21
Μέθοδος υποδειγματοληψίας με τελεστή averaging.....	21
Υπερδειγματοληψία (Up-Sampling).....	24
Παρεμβολή(Interpolation) στον χώρο.....	25
Γενικά.....	25
Μέθοδος χωρικής παρεμβολής (Interpolation) κατά J. Mateos A. Katsaggelos και R. Molina.....	27
Γενικά.....	27
Εφαρμογή διανυσμάτων κίνησης.....	28
Ανάπτυξη αλγορίθμου.....	30
Μέθοδος χωροχρονικής παρεμβολής κατά A.Smolc, K.Rummmler, J.Ohm, R.Schafer και S.Bauer.....	34
Παραγωγή διανυσμάτων κίνησης για video υψηλής ανάλυσης με Median Interpolation.....	35
Μέθοδος χωρικής παρεμβολής (De-interlacing).....	36

Παρεμβολή(Interpolation) στον χρόνο.....	39
Μέθοδος ενδιάμεσου με median φίλτρο.....	40
Πειραματικά Αποτελέσματα.....	43
Συμπεράσματα-Μελλοντικές Επεκτάσεις.....	60
Βιβλιογραφία.....	62

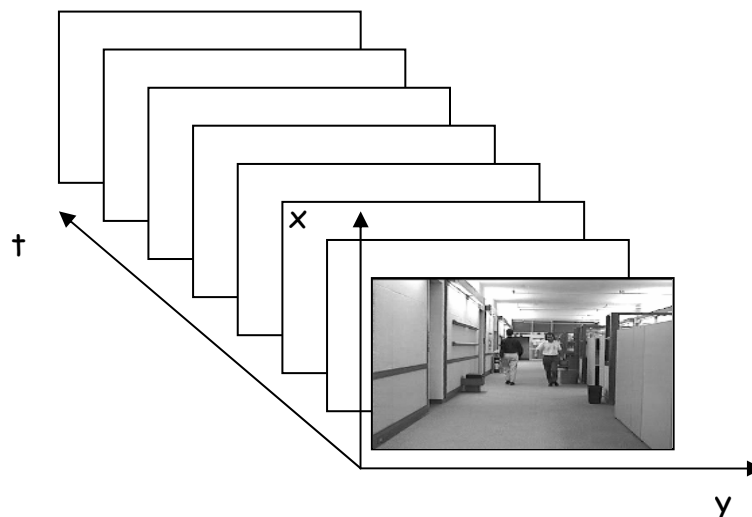
Περίληψη

Πολλές εφαρμογές που σχετίζονται με αναπαράσταση εικόνων και βίντεο, απαιτούν μεγάλη πυκνότητα πληροφορίας τόσο στον χώρο όσο και στον χρόνο. Επιπλέον η συνεχής ανάπτυξη συστημάτων για μεταφορά δεδομένων που σχετίζονται με πολυμέσα μέσω του διαδικτύου και όχι μόνο, απαιτεί μετάδοση όσο το δυνατόν πιο συμπιεσμένης πληροφορίας αλλά και προσαρμοστικότητα στα κανάλια επικοινωνίας, η οποία επιτυγχάνεται με την αλλαγή της μορφής των συστημάτων (transcoding). Οι επιδιώξεις αυτές συμβαδίζουν και με τους στόχους συστημάτων υψηλής συμπίεσης βίντεο όπως είναι τα πρότυπα MPEG. Για την εκπλήρωση των παραπάνω στόχων, η μετατροπή της χωρικής και χρονικής ανάλυσης του συμπιεσμένου βίντεο με διατήρηση της ποιότητας σε υψηλά επίπεδα, είναι μια αναγκαία διαδικασία.

Στην εργασία αυτή υλοποιήθηκαν μέθοδοι που επιτυγχάνουν την μετατροπή της χωροχρονικής ανάλυσης χρησιμοποιώντας την πληροφορία κίνησης που εκφράζεται από τα διανύσματα κίνησης. Για την εξαγωγή των διανυσμάτων κίνησης, χρησιμοποιήθηκε ένα επαναληπτικό σχήμα που προκύπτει από την ανάλυση της οπτικής ροής κατά Horn-Schunk. Στην διαδικασία μετατροπής της ανάλυσης, αρχικά αναπτύχθηκε η διαδικασία της υποδειγματοληψίας (down sampling) η οποία μειώνει την χωρική ανάλυση του βίντεο. Έπειτα υλοποιήθηκε ο επαναληπτικός αλγόριθμος κατά J. Mateos A. Katsaggelos και R. Molina, ο οποίος υπολογίζει καρέ υψηλής ανάλυσης από βίντεο με μικρότερη πυκνότητα χωρικής πληροφορίας. Επίσης αναπτύχθηκαν μέθοδοι χωροχρονικής ενίσχυσης της ανάλυσης βίντεο, σύμφωνα με τους A.Smolc, K.Rummler, J.Ohm, R.Schafer και S.Bauer, με χαμηλότερη υπολογιστική πολυπλοκότητα, οι οποίοι συγκρίθηκαν με τον επαναληπτικό αλγόριθμο. Τέλος επιχειρήθηκε μία προσπάθεια συνδυασμού των δύο μεθόδων, που στοχεύει στην μετάδοση λιγότερης πληροφορίας του βίντεο πριν την ενίσχυση της ανάλυσης του.

Εισαγωγή

Σκοπός της διπλωματικής αυτής εργασίας είναι η αλλαγή της χωρο-χρονικής ανάλυσης βίντεο η οποία εξυπηρετεί πολλές εφαρμογές, κυρίως τηλεπικοινωνιακές, καθώς και η μελέτη της συμπεριφοράς των μεθόδων ενίσχυσης της χωρικής και χρονικής ανάλυσης ενός βίντεο. Ως βίντεο ορίζεται μια ροή από εικόνες (x, y) στο χρόνο (t) . Οι εικόνες λέγονται καρέ του βίντεο και είναι η δισδιάστατη διακριτή αναπαράσταση του φυσικού χώρου (Σχήμα 1). Για την συρρίκνωση της πληροφορίας του βίντεο χρησιμοποιείται η διαδικασία της υποδειγματοληψίας (down sampling). Η υποδειγματοληψία είναι μια διαδικασία ελάττωσης της πληροφορίας του βίντεο τόσο σε χρονικό όσο και σε χωρικό επίπεδο. Στο πρώτο απλά γίνεται δειγματοληψία των καρέ του βίντεο και στο δεύτερο κάτι αντίστοιχο στα pixels για κάθε καρέ. Επομένως για την επαναφορά των εικόνων στις κανονικές τους διαστάσεις χρειάζεται να παρεμβάλουμε την πληροφορία που χάθηκε, καλύπτοντας τα κενά που δημιουργεί η υποδειγματοληψία. Αυτή η διαδικασία ονομάζεται παρεμβολή, μέθοδοι της οποίας θα αναπτυχθούν και θα μελετηθούν σε αυτήν την εργασία.



Σχήμα 1. Αναπαράσταση του βίντεο ως αλληλουχία εικόνων στις διαστάσεις του χωροχρόνου.

Για ένα φυσικό βίντεο η αύξηση της χωρικής ανάλυσης σε επίπεδο προγραμματισμού συναντά πολλές δυσκολίες αφού χρειάζεται να παρεμβληθεί μεγάλη ποσότητα πληροφορίας, ενώ και η ενίσχυση της χρονικής ανάλυσης δεν έχει ανοιχτό πεδίο ανάπτυξης. Σε εφαρμογές όμως συμπίεσμένου βίντεο όπως τα πρότυπα MPEG και όλες οι προεκτάσεις τους, η πληροφορία κίνησης μπορεί να βοηθήσει σημαντικά στο πρόβλημα.

Τα πρότυπα MPEG σκοπεύουν στην όσο το δυνατόν καλύτερη συμπίεση του βίντεο και περιγράφουν τον τρόπο με τον οποίο θα είναι αποθηκευμένη η πληροφορία του βίντεο στο μέσο αποθήκευσης μιας ψηφιακής συσκευής καθώς και τον τρόπο με τον οποίο γίνεται η μετάδοσή του. Η συμπίεση του video είναι απαραίτητη προκειμένου να υπάρχει επάρκεια χώρου στα υπάρχοντα αποθηκευτικά συστήματα και να είναι δυνατή η μετάδοση δημιουργώντας όσο το δυνατό λιγότερη κίνηση στο κανάλι επικοινωνίας. Οι μέθοδοι μείωσης και κυρίως ενίσχυσης της ανάλυσης σε χωρικό και χρονικό επίπεδο που θα αναλυθούν, εκτός από την προσαρμοστικότητα που προσφέρουν σε τηλεπικοινωνιακές εφαρμογές βίντεο, σκοπεύουν και στην επιπλέον συμπίεση της πληροφορίας προσφέροντας ακόμα μεγαλύτερη αποσυμφόρηση κατά την μετάδοση του βίντεο. Ήδη το πρότυπο MPEG 4 χρησιμοποιεί παρόμοιες μεθόδους ως μέρος μιας διαδικασίας που ονομάζεται διαβάθμιση (Scalability) για την εξυπηρέτηση των σκοπών που αναφέρθηκαν. Γενικός στόχος της εξέλιξης νέων τρόπων και μεθόδων που υπακούουν στα συστήματα MPEG είναι η προσαρμοστικότητα και η οικονομία στην μετάδοση της πληροφορίας. Έτσι οι προσεγγίσεις της εργασίας θα μπορούσαν να χρησιμοποιηθούν από τα πρότυπα MPEG. Ως κανάλι επικοινωνίας θεωρείται οποιοδήποτε δίκτυο υπολογιστών ή συσκευών δια μέσου του οποίου θα γίνει η μετάδοση του βίντεο.

Τόσο ο επαναληπτικός αλγόριθμος κατά J. Mateos A. Katsaggelos και R. Molina, ο οποίος υπολογίζει καρέ υψηλής ανάλυσης από εικόνες χαμηλής ανάλυσης, όσο και η μέθοδος ενίσχυσης της χωροχρονικής ανάλυσης σύμφωνα με τους A.Smolic, K.Rummier, J.Ohm, R.Schafer και S.Bauer, χρησιμοποιούν την πληροφορία κίνησης ανάμεσα στα καρέ που εκφράζεται από τα διανύσματα κίνησης. Έτσι υπήρχε η ανάγκη υλοποίησης ενός αλγορίθμου υπολογισμού των διανυσμάτων κίνησης αυτών. Ανάμεσα σε αλγορίθμους βασισμένους σε πλαίσια (block matching) και σε αλγορίθμους που υπολογίζουν

την οπτική ροή, δηλαδή το σύνολο των διανυσμάτων για όλα τα pixels του καρέ, χρησιμοποιήθηκε η δεύτερη κατηγορία με την ανάπτυξη και την υλοποίηση του αλγόριθμου κατά Horn και Shunk που βασίζεται στην κανονικοποίηση της εξίσωσης οπτικής ροής. Η εξίσωση καταλήγει σε ένα επαναληπτικό σχήμα το οποίο προσφέρει την οριζόντια και κατακόρυφη μετατόπιση των pixels ανάμεσα σε δύο καρέ του βίντεο. Οι μετατοπίσεις αυτές χρησιμοποιήθηκαν σε μορφή πίνακα ίσων διαστάσεων με την εικόνα, από τους αλγόριθμους που αποτελούν τη βάση της εργασίας.

Για τις ανάγκες της υλοποίησης του επαναληπτικού αλγόριθμου αύξησης της χωρικής ανάλυσης, αναπτύχθηκε η διαδικασία της υποδειγματοληψίας με τελεστή averaging η οποία μειώνει την χωρική ανάλυση ενός καρέ, καθώς και η υπερδειγματοληψία (up sampling) διαδικασία η οποία προετοιμάζει το έδαφος για την αύξηση της χωρικής ανάλυσης μέσω της παρεμβολής. Επιπλέον, για τους σκοπούς της μεθόδου ενίσχυσης της χωροχρονικής ανάλυσης κατά A.Smolic, K.Rummier, J.Ohm, R.Schafer και S.Bauer υλοποιήθηκε και μία μέθοδος η οποία αυξάνει την πυκνότητα της πληροφορίας των διανυσμάτων κίνησης σε χωρικό επίπεδο. Αυτή η χωρική παρεμβολή των πεδίων των διανυσμάτων κίνησης θα εφαρμοστεί στη συνέχεια και στον επαναληπτικό αλγόριθμο, ως συνδυασμός των δύο μεθόδων.

Συστήματα MPEG

Γενικά

Προκειμένου να κατανοήσουμε τη φύση του προβλήματος ας δούμε συνοπτικά τα σημεία των αλγόριθμων συμπίεσης MPEG από την αρχική τους μορφή μέχρι και την πιο σύγχρονη δηλαδή του MPEG7.

Αρχικά πρέπει να γίνει κατανοητό ότι οι επεκτάσεις του MPEG προτύπου βασίζονται στο προηγούμενο πρότυπο. Δηλαδή το MPEG7 βασίζεται στο MPEG4, αυτό στο MPEG2 και ούτω καθεξής. Έτσι για να μπορέσει κανείς να κατανοήσει το MPEG7 θα πρέπει να γνωρίζει καλά πρώτα όλα τα προηγούμενα πρότυπα συμπίεσης βίντεο και εικόνας στα οποία βασίζεται. Αυτό ξεφεύγει από τα πλαίσια αυτής της εργασίας, για αυτό το λόγο θα περιγραφούν τα σημεία τα οποία μελετήσαμε και μας ενδιαφέρουν.

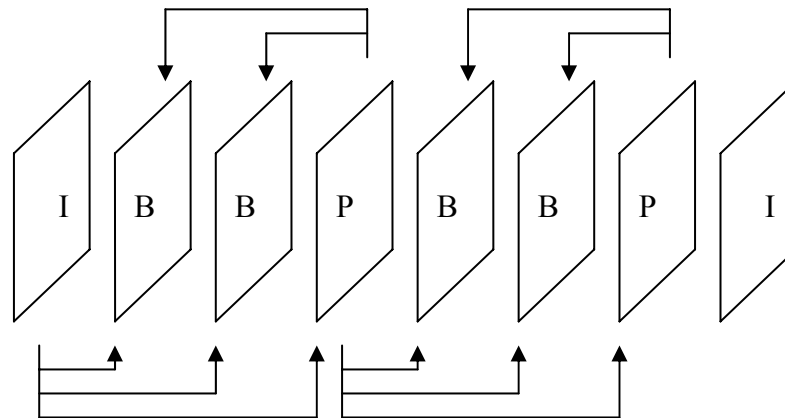
Χρονική κωδικοποίηση(Inter-frame coding)

Το πρότυπο Mpeg1 καθώς και τα Mpeg2, Mpeg4 και 7 χρησιμοποιούν τριών ειδών καρέ. Τα I-καρέ, τα B-καρέ και τα P-καρέ. Το πρώτο είναι το καρέ αναφοράς και αποτελεί το πρώτο καρέ μιας σκηνής σε ένα video, το οποίο δεν εξαρτάται από τα υπόλοιπα καρέ του video. Το Mpeg1 προσφέρει χαμηλή συμπίεση και χρησιμοποιείται προκειμένου να μπορεί να γίνει γρήγορη τυχαία πρόσβαση. Για την συμπίεσή του χρησιμοποιείται ο προκαλών απώλεια ενέργειας JPEG αλγόριθμος συμπίεσης (lossy Jpeg). Μέχρι το επόμενο καρέ αναφοράς, τα υπόλοιπα καρέ είναι τύπου B ή P. Τα P-καρέ χρησιμοποιούν πληροφορία από το προηγούμενο I-καρέ ή και τα προηγούμενα P-καρέ και προσφέρουν μέτρια συμπίεση. Τα B-καρέ είναι αυτά που προσφέρουν τη μεγαλύτερη συμπίεση και χρησιμοποιούν πληροφορία από προηγούμενα ή και μελλοντικά I-καρέ και B-καρέ (Σχήμα 2).

Είναι κατανοητό λοιπόν ότι προκειμένου να επιτευχθεί συμπίεση, χρησιμοποιούμε χρονική πληροφορία, δηλαδή την εκ των προτέρων γνώση ότι τα περισσότερα καρέ σε ένα βίντεο δεν είναι ανεξάρτητα μεταξύ τους (I-καρέ) αλλά είναι σχεδόν ίδια με τα προηγούμενα και επόμενα, επομένως μπορούμε να

εξοικονομήσουμε χώρο αν αξιοποιήσουμε την σχέση μεταξύ τους. Τα I-καρέ εμφανίζονται όταν αλλάζει σκηνή στο βίντεο ή όταν οι απαιτήσεις επιβάλλουν I-καρέ για γρήγορη αναζήτηση. Τα υπόλοιπα καρέ του βίντεο είναι τύπου B και P.

Η διαδικασία προσδιορισμού των διανυσμάτων κίνησης ξεκινάει από το πρότυπο Mpeg1. Τα διανύσματα κίνησης περιγράφουν την κίνηση κάθε σημείου (pixel) ή περιοχής (block) ανάμεσα στα καρέ του βίντεο. Τα διανύσματα εκφράζουν την προσπάθεια εισαγωγής της χρονικής πληροφορίας στον αλγόριθμο συμπίεσης Mpeg. Το σύνολο των διανυσμάτων αυτών μεταξύ δυο καρέ του βίντεο καλείται οπτική ροή (optical flow).



Σχήμα 2. Πιθανή αλληλουχία καρέ μεταξύ δυο I καρέ

Χωρική κωδικοποίηση(Intra-frame coding)

Εκτός από χρονική υπόσταση της κωδικοποίησης στα συστήματα MPEG, υπάρχει και η χωρική. Η κωδικοποίηση αυτή χρησιμοποιείται από τα καρέ τύπου I και συμπιέζει την πληροφορία μέσα στο ίδιο το καρέ. Τρεις από τις σημαντικότερες διεργασίες που αποσκοπούν στη συμπίεση αυτή, είναι ο συνημιτονικός μετασχηματισμός DCT (District Cosine Transform), η κβαντοποίηση και η κωδικοποίηση εντροπίας (Entropy coding). Πρόκειται για μεθόδους κωδικοποίησης οι οποίοι έχουν υιοθετηθεί από τα πρότυπα JPEG.

Ο μετασχηματισμός DCT όταν εφαρμόζεται στα δεδομένα της εικόνας που μέχρι τότε ήταν απλές τιμές εντάσεων, μεταφέρει την εικόνα από το χωρικό επίπεδο στο επίπεδο της συχνότητας. Η εικόνα χωρίζεται σε περιοχές

(blocks) των 64 pixels (8x8), και ο μετασχηματισμός DCT εφαρμόζεται σε κάθε έναν από αυτούς. Η εξίσωση που εκφράζει τον μετασχηματισμό φαίνεται παρακάτω.

$$F(u, v) = \frac{C(u)}{2} \frac{C(v)}{2} \sum_{x=0}^7 \sum_{y=0}^7 f(x, y) \cos \frac{(2x+1)u\pi}{16} \cos \frac{(2y+1)v\pi}{16}$$

Ο συντελεστής $F(0,0)$, αποτελεί την μέση τιμή της μετασχηματισμένης μορφής της εικόνας και ονομάζεται συντελεστής DC ενώ οι υπόλοιποι 63 ονομάζονται συντελεστές AC. Σε αυτό το σημείο κωδικοποίησης καμία πληροφορία της αρχικής εικόνας δεν έχει χαθεί, οπότε εφαρμόζοντας τον αντίστροφο μετασχηματισμό παράγεται αυτούσια η αρχική εικόνα. Το πλεονέκτημα του μετασχηματισμού DCT είναι το γεγονός ότι οι περισσότεροι συντελεστές του έχουν μηδενικές τιμές και μεγάλο μέρος των υπολοίπων έχουν πολύ μικρές τιμές, ενώ πετυχαίνει και την συγκέντρωση ενέργειας σε λίγους συντελεστές. Αυτήν την ιδιότητα εκμεταλλεύεται αργότερα η κωδικοποίηση εντροπίας.

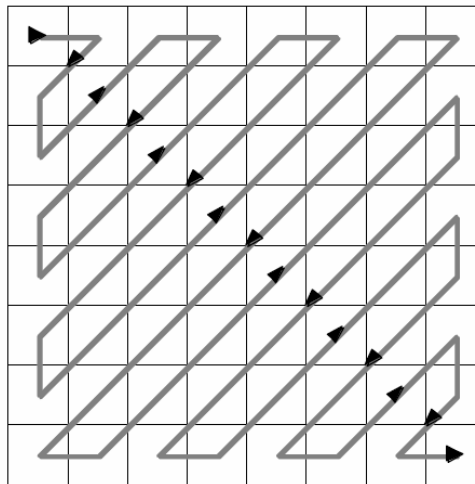
Μετά τον DCT μετασχηματισμό ακολουθεί η διαδικασία της κβαντοποίησης. Η κβαντοποίηση προσφέρει μεγάλη συμπίεση, με κόστος την απώλεια ενέργειας. Οι συντελεστές εντροπίας $Q(u, v)$ υπολογίζονται σύμφωνα με τους συντελεστές του μετασχηματισμού DCT $F(u, v)$ επί τον συντελεστή ελέγχου l , το βήμα κβαντοποίησης $QUANT$ και τον πίνακα κβαντοποίησης $W(u, v)$ με τον εξής τρόπο:

$$Q(u, v) = \frac{l \times F(u, v)}{QUANT \times W(u, v)}$$

Το βήμα κβαντοποίησης είναι αυτό που καθορίζει το ποσοστό της συμπίεσης, σε σχέση με την επιθυμητή ποιότητα των καρέ του βίντεο. Σε αυτό συμβάλει και ο πίνακας κβαντοποίησης, οπότε χρησιμοποιούνται διαφορετικοί πίνακες ανάλογα με τις απαιτήσεις της εφαρμογής. Μετά την κβαντοποίηση, μεγαλύτερος αριθμός συντελεστών αποκτά μηδενικές τιμές ενώ μικραίνει τις τιμές των υπολοίπων. Η κβαντοποίηση όταν εφαρμόζεται με μεγάλο ποσοστό συμπίεσης έχει επιπτώσεις ως προς την ποιότητα της εικόνας στην επαναφορά

αυτής. Πρόκειται για μία μη αναστρέψιμη διαδικασία. Έτσι εισάγεται το δίλημμα μεταξύ της ποιότητας και της οικονομίας.

Μετά την διαδικασία της κβαντοποίησης και πριν την κωδικοποίηση εντροπίας, εφαρμόζεται η κωδικοποίηση zig zag, όπως φαίνεται στο σχήμα 3, η οποία παρατάσσει τους συντελεστές με τέτοιο τρόπο ώστε να μην έχουν μεγάλες συχνοτικές διαφορές, με αποτέλεσμα πολλές γειτονικές τιμές να είναι μηδέν. Η κωδικοποίηση εντροπίας λαμβάνει τον συρμό από αριθμούς της σάρωσης zig zag και τον μετατρέπει σε δυαδική πληροφορία, συμπυκνώνοντας κατάλληλα την επαναλαμβανόμενη πληροφορία.

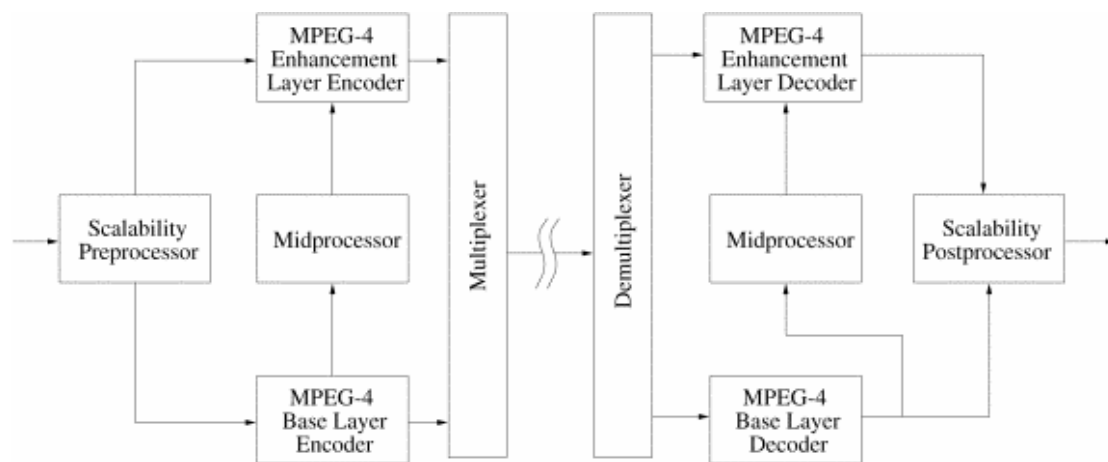


Σχήμα 3. Η σάρωση Zig Zag

Την επαναφορά των καρέ στην αρχική τους μορφή αναλαμβάνει η διαδικασία της αποκωδικοποίησης (decoding). Οι εικόνες αποκωδικοποιούνται με την εφαρμογή των αντίστροφων εργασιών που είχαμε κατά την κωδικοποίηση. Αυτές εφαρμόζονται με την αντίστροφη χρονικά σειρά από αυτήν την κωδικοποίησης.

Διαβάθμιση(Scalability)

Η διαβάθμιση (scalability) στον χώρο και στον χρόνο είναι μία από τις ιδιότητες των σύγχρονων συστημάτων MPEG. Έχει σαν στόχο την προσαρμοστικότητα του βίντεο για την χρήση του σε διαφορετικές συνθήκες μετάδοσης. Έτσι η πληροφορία του βίντεο στα πρότυπα MPEG 4 και MPEG 7 χωρίζεται σε δύο επίπεδα: στο βασικό επίπεδο(base layer) και στο ενισχυμένο επίπεδο(enhancement layer). Μια αναπαράσταση της διαδικασίας του προτύπου MPEG-4 φαίνεται στο σχήμα 3.



Σχήμα 4. Σχηματική αναπαράσταση της ιδιότητας της διαβάθμισης (Scalability) του προτύπου MPEG-4.

Τα πρότυπα MPEG 4 και 7 χρησιμοποιούν ένα γενικό πλαίσιο εργασίας για την διαβάθμιση στον χώρο και στον χρόνο. Όπως φαίνεται στο σχήμα 3 χρησιμοποιείται ένας προεπεξεργαστής(Scalability Preprocessor) για να υλοποιήσει την επιθυμητή κατάσταση. Για παράδειγμα στο χωρικό μέρος της διαδικασίας ο προεπεξεργαστής υποδειγματολυπτεί την είσοδο για να δημιουργήσει το βασικό επίπεδο το οποίο μετέπειτα θα κωδικοποιηθεί. Ο μεσοεπεξεργαστής(Midprocessor) λαμβάνει την κωδικοποιημένη έξοδο του βασικού επιπέδου και αυξάνει την ανάλυσή του για την μετατροπή στο ενισχυμένο επίπεδο. Στην περίπτωση της χρονικής διαβάθμισης τα καρέ χωρίζονται σε δύο συρμούς οι οποίοι οδηγούν στα δύο επίπεδα και ο μεσοεπεξεργαστής παρακάμπτεται. Ο ένας συρμός αποτελεί την είσοδο στον κωδικοποιητή του

βασικού επιπέδου, ενώ ο άλλος επεξεργάζεται από τον κωδικοποιητή του ενισχυμένου επιπέδου.

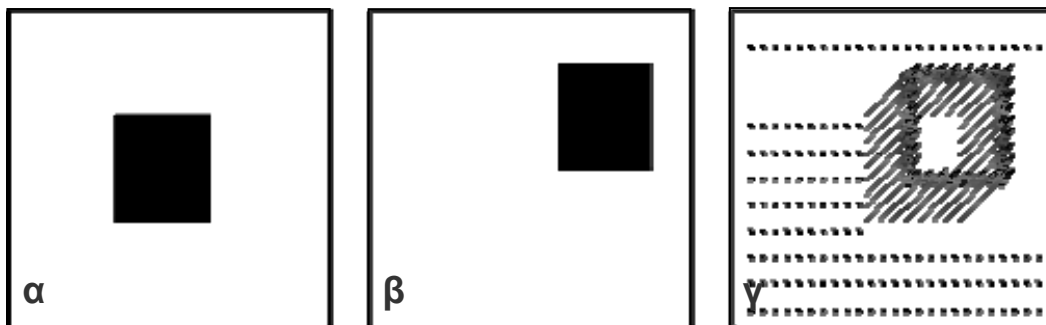
Στη χωρική διαβάθμιση, η χαμηλής περιεκτικότητας πληροφορία του βασικού επιπέδου μετατρέπεται σε πυκνότερη στο ενισχυμένο επίπεδο. Κάθε καρέ αυξάνει την ανάλυσή του αρχικά μέσω της υπερδειγματοληψίας (up sampling) και στην συνέχεια μέσω της διαδικασίας της παρεμβολής (interpolation) μετατρέπεται σε ένα υψηλής ανάλυσης καρέ, και με την βοήθεια των διανυσμάτων κίνησης. Στο πρότυπο MPEG-4 και στα επόμενα αυτού, όπου κάθε αντικείμενο αντιμετωπίζεται ξεχωριστά, είναι δυνατή η ενίσχυση επιλεγμένων αντικειμένων αφού πρώτα αυτά έχουν πάρει μορφή παραθύρου. Οι μέθοδοι που χρησιμοποιούνται από τα συστήματα MPEG 4 και 7 για την χωρική ενίσχυση των καρέ έχουν την ίδια βάση και παρόμοια μεθοδολογία με τις μεθόδους που θα αναπτυχθούν σε αυτήν την εργασία.

Στη χρονική διάσταση πραγματοποιείται επιπλέον ενίσχυση της πυκνότητας της πληροφορίας. Με την χρήση της χωρικής πληροφορίας των υπαρχόντων καρέ, αλλά και με την βοήθεια των διανυσμάτων κίνησης είναι δυνατή η εξαγωγή νέων καρέ τα οποία παρεμβάλλονται χρονικά ανάμεσα στα υπόλοιπα.

Υπολογισμός Διανυσμάτων κίνησης κατά Horn Shunk

Γενικά

Πριν την παρουσίαση της μεθόδου υπολογισμού των διανυσμάτων κίνησης είναι αναγκαίος ο ορισμός της έννοιας της οπτικής ροής. Ως οπτική ροή (optical flow) ορίζουμε το σύνολο των διανυσμάτων της κίνησης (u, v) μεταξύ δυο διαδοχικών εικόνων (καρέ) ενός video όπως φαίνεται και στο Σχήμα 5.



Σχήμα 5.

Η οπτική ροή (γ) μεταξύ των καρέ (α) και (β)

Αν μοντελοποιήσουμε το βίντεο ως συνάρτηση $I(\mathbf{x}, t)$ με $\mathbf{x} = (x, y)$, οπτική ροή είναι το σύνολο των διανυσμάτων $\mathbf{v}(\mathbf{x}, t) = (u(\mathbf{x}, t), v(\mathbf{x}, t))$ τέτοιο ώστε να ισχύει ότι:

$$I(\mathbf{x}, t)\mathbf{v} + I_t(\mathbf{x}, t) = 0$$

Υπολογισμός αλγορίθμου

Σύμφωνα με τους Horn και Schunck, η οπτική ροή ορίζεται σε σχέση με τη χωρική και τη χρονική αλλαγή της έντασης της φωτεινότητας ανάμεσα στα καρέ του βίντεο. Οπτική ροή είναι η διανομή της φαινομενικής ταχύτητας της κίνησης του χρωματικού μοτίβου σε μια εικόνα.

Η ανοιχτότητα του προβλήματος είναι αναπόφευκτη εφόσον δεν είναι ένα πρόβλημα πλήρως ορισμένο. Για αυτό το λόγο γίνονται κάποιες βολικές υποθέσεις προκειμένου να γίνει μια εκτίμηση και να μπορέσουν να υπολογι-

στούν τα διανύσματα κίνησης. Έτσι η ταχύτητα ανάμεσα σε δυο καρέ του βίντεο, μπορεί να υπολογιστεί από το ολοκλήρωμα:

$$F_t = \int_D (\nabla I \cdot \mathbf{v} + I_t)^2 + \lambda^2 (\|\nabla u\|_2^2 + \|\nabla v\|_2^2) d\mathbf{x}$$

που ορίζεται στα όρια της εικόνας του καρέ. Ιδανικά το ολοκλήρωμα αυτό είναι μηδέν. Με πραγματικά δεδομένα όμως, ο πρώτος όρος $(\nabla I \cdot \mathbf{v} + I_t)^2$ είναι διάφορος του μηδενός λόγω θορύβου, παραμορφώσεων και άλλων φαινομένων που έχουν να κάνουν με τη φύση του βίντεο και ο δεύτερος όρος $\|\nabla u\|_2^2 + \|\nabla v\|_2^2$ είναι επίσης διάφορος του μηδενός, για εκτίμηση που δε συμφωνεί με τα γειτονικά διανύσματα κίνησης αφού θεωρούμε ότι ιδεατά διανύσματα κίνησης είναι ομαλά (smoothness constraint). Έτσι το ολοκλήρωμα αυτό είναι μηδέν για επιλογή διανυσμάτων κίνησης τέτοιων ώστε να εκφράζουν τόσο την κίνηση με κριτήριο την ένταση (πρώτος όρος) όσο και την κίνηση με κριτήριο την ομοιομορφία των διανυσμάτων στα πλαίσια της περιοχής γύρω του (δεύτερος όρος). Ο παράγοντας λ υπάρχει για να μπορεί να ρυθμίζει την ισχύ των δυο δεσμεύσεων που αποτελούν τη βάση του υπολογισμού των διανυσμάτων κίνησης.

Αναπτύσσοντας το ολοκλήρωμα λοιπόν έχουμε:

$$F_t = \int_D \lambda (I_x u + I_y v + I_t)^2 + (u_x^2 + u_y^2 + v_x^2 + v_y^2) d\mathbf{x}$$

Οπότε σε διακριτή μορφή το ολοκλήρωμα θα πάρει τη μορφή:

$$F_t = \sum_{i,j} \lambda (I_x u + I_y v + I_t)^2 + \frac{1}{4} ((u_{i+1,j} - u_{ij})^2 + (u_{i,j+1} - u_{ij})^2 + (v_{i+1,j} - v_{ij})^2 + (v_{i,j+1} - v_{ij})^2)$$

Έτσι προκειμένου να ελαχιστοποιήσουμε το ολοκλήρωμα αρκεί:

$$\nabla F_t = 0 \quad \text{ή αλλιώς} \quad \begin{cases} \frac{\partial F_t}{\partial u} = 0 \\ \frac{\partial F_t}{\partial v} = 0 \end{cases}$$

Η μερική παράγωγος $\frac{\partial F_t}{\partial u}$ υπολογίζεται παρακάτω:

$$\frac{\partial F_t}{\partial u} = 2\lambda(I_x u_{ij} + I_y v_{ij} + I_t)I_x - \frac{1}{2}[(u_{i+1,j} - u_{ij}) + (u_{i,j+1} - u_{ij})] + \frac{1}{2}[(u_{ij} - u_{i-1,j}) + (u_{ij} - u_{i,j-1})] \Leftrightarrow$$

$$\frac{\partial F_t}{\partial u} = 2\lambda(I_x u_{ij} + I_y v_{ij} + I_t)I_x + 2u_{ij} - 2\left(\frac{1}{4}(u_{i+1,j} + u_{i,j+1} + u_{i-1,j} + u_{i,j-1})\right) \Leftrightarrow$$

$$\frac{\partial F_t}{\partial u} = 2\lambda(I_x u_{ij} + I_y v_{ij} + I_t)I_x + 2u_{ij} - 2\bar{u}_{ij}$$

Ομοίως υπολογίζουμε και την μερική παράγωγο $\frac{\partial F_t}{\partial v}$ οπότε έχουμε:

$$\frac{\partial F_t}{\partial v} = 2\lambda(I_x u_{ij} + I_y v_{ij} + I_t)I_y + 2v_{ij} - 2\bar{v}_{ij}$$

Οπότε το σύστημα που έχουμε να λύσουμε είναι:

$$\begin{cases} 2\lambda(I_x u_{ij} + I_y v_{ij} + I_t)I_x + 2u_{ij} - 2\bar{u}_{ij} = 0 \\ 2\lambda(I_x u_{ij} + I_y v_{ij} + I_t)I_y + 2v_{ij} - 2\bar{v}_{ij} = 0 \end{cases} \Leftrightarrow \begin{cases} (1 + \lambda I_x^2)u_{ij} + \lambda I_x I_y v_{ij} = \bar{u} - \lambda I_x I_t \\ \lambda I_x I_y u_{ij} + (1 + \lambda I_y^2)v_{ij} = \bar{v} - \lambda I_y I_t \end{cases}$$

Σε μορφή πινάκων μπορεί να γραφτεί:

$$\begin{bmatrix} 1 + \lambda I_x^2 & \lambda I_x I_y \\ \lambda I_x I_y & 1 + \lambda I_y^2 \end{bmatrix} \begin{bmatrix} u_{ij} \\ v_{ij} \end{bmatrix} = \begin{bmatrix} \bar{u} - \lambda I_x I_t \\ \bar{v} - \lambda I_y I_t \end{bmatrix}$$

Για την επίλυση του παραπάνω συστήματος εφαρμόσαμε την επαναληπτική μέθοδο Jacobi (Jacobi iterations), όπου η λύση του συστήματος δίνεται από τον τύπο:

$$x^{(m+1)} = D^{-1}(L + U)x^{(m)} + D^{-1}b \mid m = 0, 1, 2, \dots$$

Ο πίνακας των συντελεστών των u, v ονομάζεται A και ο πίνακας των σταθερών όρων b . Από αυτούς κατασκευάζουμε τους πίνακες D, L και U ώστε να ισχύει $A = D - L - U$ όπου D να είναι διαγώνιος, L αυστηρά κάτω τριγωνικός και U αυστηρά πάνω τριγωνικός πίνακας. Έτσι οι πίνακες του σχήματος είναι οι παρακάτω:

$$D = \begin{bmatrix} 1 + \lambda I_x^2 & 0 \\ 0 & 1 + \lambda I_y^2 \end{bmatrix} \quad U = \begin{bmatrix} 0 & -\lambda I_x I_y \\ 0 & 0 \end{bmatrix} \quad L = \begin{bmatrix} 0 & 0 \\ -\lambda I_x I_y & 0 \end{bmatrix}$$

$$b = \begin{bmatrix} \bar{u} - \lambda I_x I_t \\ \bar{v} - \lambda I_y I_t \end{bmatrix} \text{ και } D^{-1} = \begin{bmatrix} \frac{1}{1 + \lambda I_x^2} & 0 \\ 0 & \frac{1}{1 + \lambda I_y^2} \end{bmatrix}$$

Έτσι έχουμε:

$$x^{(m+1)} = D^{-1}(L + U)x^{(m)} + D^{-1}b$$

$$\begin{bmatrix} u^{(m+1)} \\ v^{(m+1)} \end{bmatrix} = \begin{bmatrix} 1/(1 + \lambda I_x^2) & 0 \\ 0 & 1/(1 + \lambda I_y^2) \end{bmatrix} \begin{bmatrix} 0 & -\lambda I_x I_y \\ -\lambda I_x I_y & 0 \end{bmatrix} \begin{bmatrix} u^{(m)} \\ v^{(m)} \end{bmatrix} + \begin{bmatrix} 1/(1 + \lambda I_x^2) & 0 \\ 0 & 1/(1 + \lambda I_y^2) \end{bmatrix} \begin{bmatrix} \bar{u} - \lambda I_x I_t \\ \bar{v} - \lambda I_y I_t \end{bmatrix}$$

$$\begin{bmatrix} u^{(m+1)} \\ v^{(m+1)} \end{bmatrix} = \begin{bmatrix} 0 & \frac{-\lambda I_x I_y}{1 + \lambda I_x^2} \\ \frac{-\lambda I_x I_y}{1 + \lambda I_y^2} & 0 \end{bmatrix} \begin{bmatrix} u^{(m)} \\ v^{(m)} \end{bmatrix} + \begin{bmatrix} \frac{\bar{u} - \lambda I_x I_t}{1 + \lambda I_x^2} \\ \frac{\bar{v} - \lambda I_y I_t}{1 + \lambda I_y^2} \end{bmatrix}$$

$$\begin{bmatrix} u^{(m+1)} \\ v^{(m+1)} \end{bmatrix} = \begin{bmatrix} \frac{-\lambda I_x I_y}{1 + \lambda I_x^2} v^{(m)} \\ \frac{-\lambda I_x I_y}{1 + \lambda I_y^2} u^{(m)} \end{bmatrix} + \begin{bmatrix} \frac{\bar{u} - \lambda I_x I_t}{1 + \lambda I_x^2} \\ \frac{\bar{v} - \lambda I_y I_t}{1 + \lambda I_y^2} \end{bmatrix}$$

$$\begin{bmatrix} u^{(m+1)} \\ v^{(m+1)} \end{bmatrix} = \begin{bmatrix} \frac{\bar{u} - \lambda I_x I_t - \lambda I_x I_y v^{(m)}}{1 + \lambda I_x^2} \\ \frac{\bar{v} - \lambda I_y I_t - \lambda I_x I_y u^{(m)}}{1 + \lambda I_y^2} \end{bmatrix} \dots \dots \dots (1)$$

Για $x^{(0)}$ επιλέγονται αυθαίρετες τιμές, συνήθως μηδέν. Βέβαια όταν πρόκειται για παραπάνω των δυο καρέ, στην ακολουθία του βίντεο που ζητάμε να βρούμε τα διανύσματα κίνησης, τότε μια καλή αρχικοποίηση είναι η λύση του προηγούμενου ζευγαριού. Αυτό είναι προφανές δεδομένου ότι περιμένουμε μικρές αλλαγές στην κίνηση μεταξύ δυο διαδοχικών καρέ.

Για τον υπολογισμό του παραπάνω σχήματος έγιναν οι ακόλουθες παραδοχές:

- Ότι η κίνηση ανάμεσα στα καρέ του βίντεο είναι αρκετά μικρή με αποτέλεσμα να μπορούμε να θεωρήσουμε ότι:

$$\frac{dI}{dt} \approx 0$$

- Ότι υπάρχει η παράγωγος της εικόνας σε κάθε σημείο ως προς x και ως προς y . Δηλαδή:

$$\frac{\partial x}{\partial t} \neq 0 \text{ και } \frac{\partial y}{\partial t} \neq 0$$

Τα αποτελέσματα του επαναληπτικού σχήματος που φαίνονται στην σχέση (1) θα χρησιμοποιηθούν ως διανύσματα κίνησης στις μεθόδους μετατροπής της χωροχρονικής ανάλυσης που θα ακολουθήσουν.

Χωρική μετατροπή της ανάλυσης

Η ανάλυση ενός βίντεο είναι όρος που σχετίζεται με την πυκνότητα της πληροφορίας που περιέχεται σε αυτό. Με τον όρο χωρική μετάλλαξη εννοούμε την αλλαγή της ποσότητας της πληροφορίας σε κάθε καρέ του βίντεο. Για την αύξηση του αριθμού των pixel σε κάθε καρέ εφαρμόζεται αρχικά η διαδικασία της υπερδειγματοληψίας (up sampling) και έπειτα ακολουθεί η παρεμβολή (interpolation), ενώ για την μείωση του μεγέθους των καρέ εφαρμόζεται η υποδειγματοληψία (down-sampling). Οι δύο εφαρμογές που έχουν να κάνουν με την δειγματοληψία, είναι άμεσα σχετιζόμενες αφού όπως θα αναλυθεί παρακάτω συνδέονται με την αντιστροφή ενός πίνακα, και επιπλέον πρόκειται για αρκετά απλές διαδικασίες. Η παρεμβολή ακολουθεί την υπερδειγματοληψία και συμπληρώνει τη διαδικασία αύξησης της χωρικής ανάλυσης. Με αυτή τη σειρά θα παρουσιαστούν αυτοί οι μέθοδοι στη συνέχεια της εργασίας.



Σχήμα 6. (α) Καρέ μεγέθους 352x240 pixels. (β) Το ίδιο καρέ σε μέγεθος 176x120 pixels.

Υποδειγματοληψία(down-sampling)

Η διαδικασία της υποδειγματοληψίας χρησιμοποιείται ιδιαίτερα σε σύγχρονες τηλεπικοινωνιακές εφαρμογές όπως η τηλεδιάσκεψη και η εκμάθηση από απόσταση, ως μέρος μιας άλλης διαδικασίας που ονομάζεται διακωδικοποίηση(transcoding). Η διακωδικοποίηση συμβάλει στην μετατροπή ενός συρμού από bit κωδικοποιημένου video σε ένα άλλο με διαφοροποιημένο ρυθμό μετάδοσης bit (bit rate) και ρυθμό μετάδοσης καρέ (frame rate), ενώ διαφορετική μπορεί να είναι και η μέθοδος κωδικοποίησης(π.χ. μετατροπή από H.26x σε MPEGx). Επιπλέον χαρακτηριστικά του βίντεο που επιδέχονται αλλαγής είναι και η χωρική και χρονική ανάλυση με μεθόδους παρόμοιες με αυτές που θα αναπτυχθούν στην συνέχεια της εργασίας. Σκοπός της διακωδικοποίησης είναι η προσαρμογή του περιεχομένου του video σε ποικιλόμορφες συνθήκες καναλιών και η δυνατότητα υποστήριξης από τον εξοπλισμό κάθε χρήστη (UMA universal multimedia access). Η υπολογιστική πολυπλοκότητα και η ποιότητα της εικόνας είναι δύο σημαντικές υποθέσεις σε αυτήν την διαδικασία και χρειάζεται να προσδιορίζονται με τέτοιο τρόπο ώστε να συμφωνούν με τις ποικίλες απαιτήσεις σε πρακτικές εφαρμογές. Για την επίτευξη αυτού του στόχου η διαδικασία της μείωσης της χωρικής ανάλυσης είναι σημαντικός παράγοντας.

Στο πεδίο της υλοποίησης η υποδειγματοληψία στο χώρο περιλαμβάνει απλές μεθόδους. Κάποιες από αυτές λειτουργούν σε επίπεδα blocks ή macroblocks και άλλες σε επίπεδο pixel. Η μέθοδος που υλοποιήθηκε σε αυτήν την εργασία είναι απλή και αποτελεσματική. Πρόκειται για την υποδειγματοληψία με τελεστή averaging και θα αναλυθεί παρακάτω.

Μέθοδος υποδειγματοληψίας με τελεστή averaging

Η μέθοδος που θα εξετάσουμε είναι από τις πιο συνήθεις που χρησιμοποιούνται για την επίτευξη της υποδειγματοληψίας στον χώρο σε ένα video. Αποτελείται από δύο σκέλη. Στο πρώτο σκέλος το αρχικό καρέ προς επεξερ-

γασία φιλτράρεται με ένα φίλτρο averaging. Στο δεύτερο σκέλος στο φιλτραρισμένο πλέον υψηλής ανάλυσης καρέ εφαρμόζεται η διαδικασία της υποδειγματοληψίας. Να σημειωθεί ότι οι διαστάσεις του καρέ σε pixels μειώθηκαν στην μισή τιμή τόσο για το ύψος όσο και για το πλάτος.

Το φιλτράρισμα του υψηλής ανάλυσης καρέ συντελέστηκε με εφαρμογή της μάσκας averaging H της οποίας η μορφή φαίνεται παρακάτω. Ξεκινώντας από την 2×2 μάσκα Στόχος του φίλτρου είναι η διαμόρφωση της έντασης κάθε pixel έτσι ώστε να έχει τιμή ίση με τον μέσο όρο των εντάσεων των τριών γειτονικών του pixel αλλά και του ιδίου. Πλέον κάθε pixel του καρέ θα αντιπροσωπεύει τον μέσο όρο των τιμών των τεσσάρων pixels του block στο οποίο ανήκει.

$$h = \begin{bmatrix} 1/4 & 1/4 \\ 1/4 & 1/4 \end{bmatrix} \rightarrow h_e = \left\{ \begin{array}{cccc} \overbrace{1/4 & 0 & \dots & 1/4}^M \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \underbrace{1/4 & \dots & \dots & 1/4}_N \end{array} \right\}_N$$

$$H_j = \left\{ \begin{array}{cccc} \overbrace{h_{j0} & h_{jM-1} & \dots & h_{j1}}^M \\ h_{j1} & h_{j0} & \dots & h_{j2} \\ \vdots & \vdots & \ddots & \vdots \\ h_{jN-1} & \dots & \dots & h_{j0} \end{array} \right\}_N \rightarrow H_{bc} = \left\{ \begin{array}{cccc} \overbrace{H_0 & H_{N-1} & \dots & H_1}^{M \times N} \\ H_1 & H_0 & \dots & H_2 \\ \vdots & \vdots & \ddots & \vdots \\ H_{M-1} & \dots & \dots & H_0 \end{array} \right\}_{M \times N}$$

$$H_{bc} = \left\{ \begin{array}{ccccccccc} \overbrace{1/4 & 1/4 & 0 & \dots & 1/4 & 1/4 & 0 & \dots & 0}^{M \times N} \\ 0 & 1/4 & 1/4 & 0 & \dots & 1/4 & 1/4 & \dots & 0 \\ \vdots & 0 & \ddots & \ddots & 0 & 0 & \ddots & \ddots & \vdots \\ 0 & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & 0 \\ 0 & \dots & 0 & \dots & 0 & 0 & \dots & 0 & 1/4 \end{array} \right\}_{M \times N}$$

Σχήμα 7. Η διαδικασία μετατροπής της μάσκας σε μορφή block circulant

Μετά την εφαρμογή του φίλτρου ακολουθεί η διαδικασία της υποδειγματοληψίας. Πρόκειται για μία δειγματοληψία που γίνεται στο διαμορφωμένο υψηλής ανάλυσης καρέ. Σε επίπεδο υλοποίησης η ενέργεια αυτή εφαρμόστηκε με τη βοήθεια ενός πίνακα A που ονομάζεται τελεστής υποδειγματοληψίας.

Έστω ότι \underline{F}_k είναι το k -στο υψηλής ανάλυσης καρέ με row order κωδικοποίηση και \underline{X}_k το αντίστοιχο καρέ χαμηλής ανάλυσης σε μορφή διανύσματος επίσης. Αν το αρχικό μέγεθος του υψηλής ανάλυσης καρέ είναι $M \times N$ pixels τότε κάθε καρέ χαμηλής ανάλυσης επιδιώκουμε να έχει $M/2 \times P/2$ pixels. Τότε ο πίνακας A μεγέθους $(M/2 \times N/2) \times (M \times N)$ είναι ο πίνακας υποδειγματοληψίας (sub-sampling operator) με την μορφή που φαίνεται στο σχήμα 7.

$$A = \left[\begin{array}{cccccc} \overbrace{1 & 0 & 0 & 0 & \dots & 0}^{M \times N} \\ 0 & 0 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 & 0 \\ 0 & 0 & \dots & 0 & 0 & 1 \end{array} \right] \left. \vphantom{\begin{array}{c} 1 \\ 0 \\ \vdots \\ 0 \\ 0 \end{array}} \right\} M/2 \times N/2$$

Σχήμα 8. Ο πίνακας υποδειγματοληψίας

Η κωδικοποίηση των εντάσεων του καρέ κατά row order είναι στην ουσία μια μετατροπή της πληροφορίας από μορφή πίνακα σε μορφή διανύσματος. Οι τιμές του διανύσματος αυτού είναι οι εντάσεις κάθε γραμμής της εικόνας η μία κάτω από την άλλη όπως φαίνεται στο σχήμα 8.

$$\begin{bmatrix} a_{11} & \dots & a_{1N} \\ \vdots & \ddots & \vdots \\ a_{M1} & \dots & a_{MN} \end{bmatrix} \rightarrow \begin{bmatrix} a_{11} \\ a_{12} \\ \vdots \\ a_{1N} \\ a_{21} \\ \vdots \\ a_{2N} \\ \vdots \\ \vdots \\ a_{MN} \end{bmatrix}$$

Σχήμα 9. Μετατροπή του καρέ από μορφή πίνακα σε μορφή διανύσματος(row order)

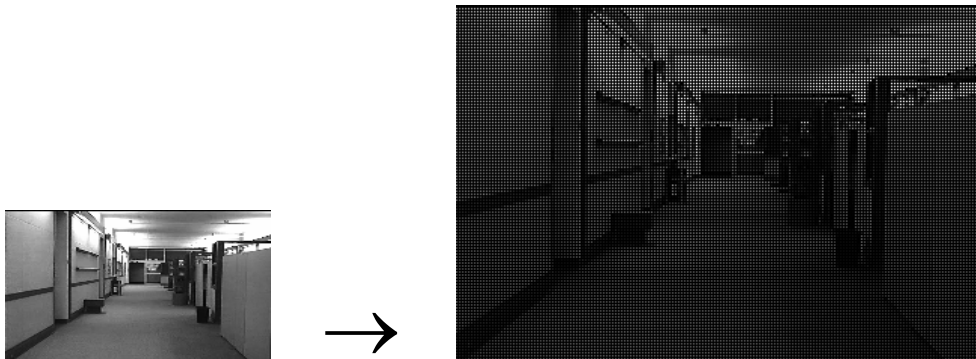
Έτσι η τελική μορφή της εξίσωσης η οποία αντιπροσωπεύει τη μέθοδο υποδειγματοληψίας που εξετάσαμε είναι της μορφής:

$$\underline{X}_k = A H \underline{F}_k$$

Σε επίπεδο υλοποίησης, η εφαρμογή του φίλτρου έγινε με τη συνέλιξη αυτού σε μορφή μάσκας πάνω στην εικόνα η οποία έχει μορφή πίνακα και όχι διανύσματος. Η συνέλιξη εφαρμόστηκε με την ολίσθηση της μάσκας με μορφή παραθύρου, υπολογίζοντας τις νέες τιμές έντασης για κάθε pixel.

Υπερδειγματοληψία (Up-sampling)

Η υπερδειγματοληψία (up-sampling) αποτελεί το πρώτο στάδιο στην διαδικασία αύξησης της χωρικής ανάλυσης ενός βίντεο και διαμορφώνει τα καρέ με τέτοιο τρόπο, ώστε να είναι δυνατή η μετέπειτα εφαρμογή της παρεμβολής(interpolation). Στην υπερδειγματοληψία, κάθε χαμηλής ανάλυσης καρέ του βίντεο αυξάνει τις διαστάσεις του με τέτοιο τρόπο, ώστε η ήδη υπάρχουσα πληροφορία να περιβάλλεται από pixels με κενή πληροφορία (τιμές εντάσεως στο μηδέν) όπως φαίνεται στο σχήμα 9. Αξίζει να σημειωθεί ότι κατά τον διπλασιασμό των διαστάσεων ενός καρέ, η νέα εικόνα που λαμβάνεται αποτελείται κατά 3/4 από άγνωστη πληροφορία και το υπόλοιπο 1/4 προέρχεται από το καρέ εισόδου.



Σχήμα 10. Το αποτέλεσμα της υπερδειγματοληψίας(up-sampling) ενός καρέ. Οι διαστάσεις του διπλασιάζονται όμως χρειάζεται επιπλέον επεξεργασία για την απόδοση του καρέ σε ποιο υψηλή ανάλυση.

Αλγεβρικά, η υπερδειγματοληψία εφαρμόζεται μέσω του τελεστή υπερδειγματοληψίας (up-sampling operator). Αν A είναι ο $(M \times N) \times (M/2 \times N/2)$ πίνακας υποδειγματοληψίας που αναπτύχθηκε στην προηγούμενη ενότητα τότε ο ανάστροφός του πίνακας A' αποτελεί τον πίνακα υπερδειγματοληψίας. Η εφαρμογή του πίνακα αυτού στην χαμηλής ανάλυσης εικόνα \underline{X}_k θα μας δώσει το αποτέλεσμα που φαίνεται στο σχήμα 10.

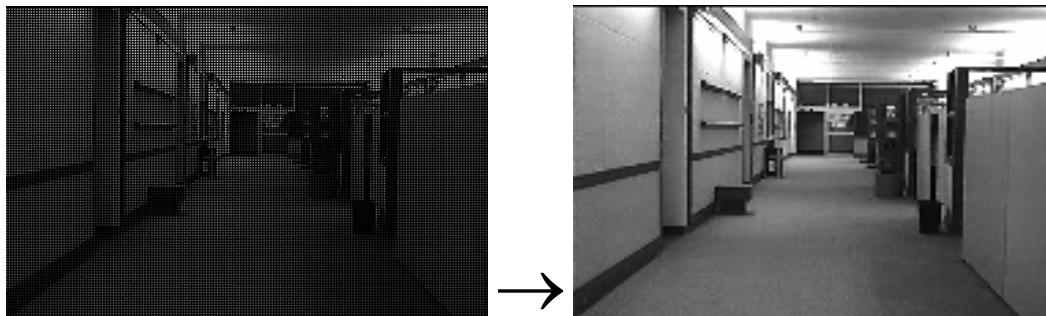
Παρεμβολή (Interpolation) στον χώρο

Η παρεμβολή έχει σκοπό την κάλυψη της άγνωστης πληροφορίας που δημιουργείται κατά την υπερδειγματοληψία ενός καρέ. Στόχος της διαδικασίας αυτής είναι η όσο το δυνατόν καλύτερη απόδοση του αρχικού υψηλής ανάλυσης καρέ όταν πρόκειται για εφαρμογές όπου χρειάζεται επαναφορά του βίντεο στις αρχικές του διαστάσεις. Σε εφαρμογές στις οποίες δεν υπάρχει αρχική πληροφορία στόχος είναι μία πιο λεπτομερής απεικόνιση σε μεγαλύτερη ανάλυση.

Το θέμα της παρεμβολής απασχόλησε αρχικά την επεξεργασία εικόνας με αποτέλεσμα την ανάπτυξη πληθώρας μεθόδων με σκοπό την βέλτιστη ανάπλαση μιας εικόνας σε μεγαλύτερη ανάλυση. Οι μέθοδοι αυτοί χρησιμοποιούν κυρίως την χωρική πληροφορία της ίδιας της εικόνας, για να προσδιορίσουν τις τιμές των εντάσεων που χρειάζεται να συμπληρωθούν. Όταν πρόκειται όμως για ψηφιακό βίντεο, η εφαρμογή των μεθόδων αυτών σε κάθε καρέ δεν επαρκεί για να φτάσουμε στο βέλτιστο αποτέλεσμα. Η λύση βρίσκεται στην πολύ σημαντική ιδιότητα των συστημάτων MPEG, τα διανύσματα κίνησης. Έτσι η υπάρχουσα ανάπτυξη των χωρικών μεθόδων σε συνδυασμό με την πληροφορία που παρέχουν τα διανύσματα κίνησης, προσφέρουν τόσο σε χωρικό όσο και σε χρονικό επίπεδο μια πιο ικανοποιητική μετάβαση του βίντεο σε μεγαλύτερη ανάλυση.

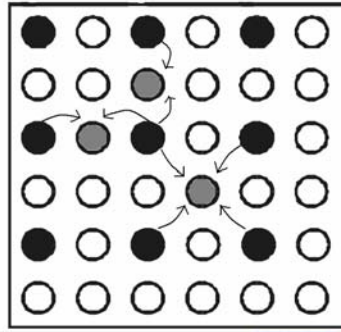
Ανάμεσα σε μεθόδους χωρικής παρεμβολής που χρησιμοποιούν συνέλιξη (B-splines, Cubic convolution) αλλά και άλλες χωρικές μεθόδους ενίσχυ-

σης ανάλυσης εικόνας, προτιμήθηκε πρώτα ένας επαναληπτικός αλγόριθμος μεγάλης πολυπλοκότητας ο οποίος χρησιμοποιεί πληροφορία κίνησης του βίντεο, καθώς και μια μέθοδος πιο ευέλικτη και γρήγορη, που επίσης εκμεταλλεύεται τα διανύσματα κίνησης. Αυτές οι δύο μέθοδοι θα αναλυθούν στην συνέχεια και θα συγκριθούν με μία μέθοδο που χρησιμοποιεί μόνο χωρική πληροφορία, την γραμμική παρεμβολή.



Σχήμα 11. Παράδειγμα εφαρμογής της γραμμικής παρεμβολής(bilinear interpolation) σε καρέ στο οποίο έχει εφαρμοστεί υπερδειγματοληψία(upsampling).

Μία από τις πιο απλές μεθόδους χωρικής παρεμβολής είναι η γραμμική (bilinear interpolation). Στη γραμμική παρεμβολή κάθε pixel με μηδενική πληροφορία έντασης, ανάλογα με την περιοχή που βρίσκεται παίρνει την τιμή της έντασης που προκύπτει από τον μέσο όρο των εντάσεων των γειτονικών του pixels όπως φαίνεται στο σχήμα 10. Το pixel προς κάλυψη μπορεί να βρίσκεται ανάμεσα σε δύο pixels που προέρχονται από το χαμηλής ανάλυσης καρέ οριζόντια ή κάθετα, είτε μπορεί να βρίσκεται ανάμεσα σε τέσσερα τέτοια pixels. Σε κάθε μια από τις τρεις περιπτώσεις η τιμή της έντασης του pixel συμπληρώνεται με τον μέσο όρο είτε των δύο είτε των τεσσάρων pixels τα οποία αποτελούν τη γνήσια πληροφορία.



Σχήμα 12. Απεικόνιση της λειτουργίας της γραμμικής παρεμβολής (bilinear interpolation).

Τα αποτελέσματα της γραμμικής παρεμβολής είναι ικανοποιητικά αλλά δεν αρκούν για τις ανάγκες των εφαρμογών του ψηφιακού βίντεο. Η μέθοδος αυτή είναι χρήσιμη αν συνδυαστεί με αλγορίθμους που χρησιμοποιούν διανύσματα κίνησης. Παρακάτω θα παρουσιαστούν τέτοιες μέθοδοι αναλυτικά.

Μέθοδος χωρικής παρεμβολής κατά J. Mateos A. Katsaggelos και R. Molina

Γενικά

Η μέθοδος σύμφωνα με τους J. Mateos A. Katsaggelos και R. Molina προσφέρει τη δυνατότητα χωρικής παρεμβολής σε ένα συμπιεσμένο βίντεο με χρήση διανυσμάτων κίνησης. Η μέθοδος αυτή βασίζεται στην ανάπτυξη μιας μαθηματικής εξίσωσης που καταλήγει σε έναν επαναληπτικό αλγόριθμο, ο οποίος πετυχαίνει τη βελτίωση της ανάλυσης του βίντεο.

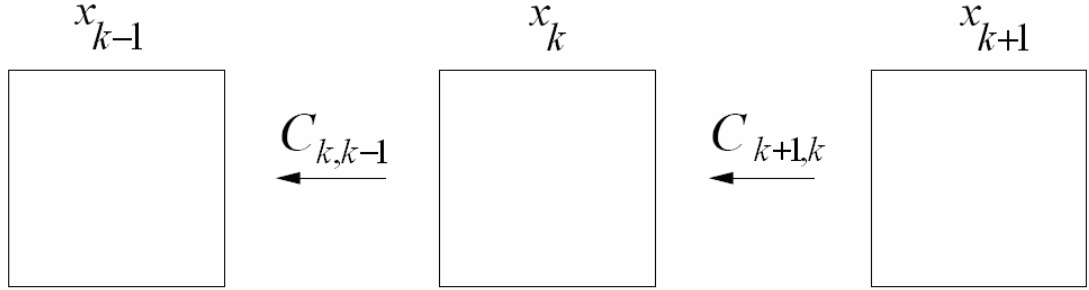
Για την προσέγγιση του προβλήματος και την ανάπτυξη του αλγορίθμου είναι αναγκαία η εφαρμογή μιας σειράς βημάτων. Αρχικά χρειάζεται να αναπτυχθεί ένα μοντέλο που να περιγράφει τη σχέση μεταξύ των καρέ χαμηλής και υψηλής ανάλυσης. Έπειτα πρέπει να προσδιοριστεί η σχέση η οποία εκφράζει την εφαρμογή των διανυσμάτων κίνησης σε ένα καρέ, που θα έχει ως αποτέλεσμα την απόδοση του επόμενου ή του προηγούμενου καρέ του βίντεο. Τέλος πρέπει να εξαχθεί η μαθηματική εξίσωση η οποία περιέχει τις

πράξεις των δύο πρώτων βημάτων και δίδει το επιθυμητό αποτέλεσμα. Όμως στην προσπάθεια αυτή της εκτίμησης του υψηλής ανάλυσης βίντεο, πολύ σημαντικό παράγοντα για την εξαγωγή όσο το δυνατόν καλύτερων αποτελεσμάτων αποτελεί η ακρίβεια των διανυσμάτων κίνησης. Η μέθοδος για την εκτίμηση της ροής του βίντεο θα πρέπει να αντιμετωπίζει ικανοποιητικά τα προβλήματα που προκύπτουν από την εφαρμογή των διανυσμάτων κίνησης όπως η παρουσία κενών pixels και η αστοχία.

Στο επίπεδο υλοποίησης για την παραγωγή των διανυσμάτων κίνησης χρησιμοποιήθηκε ο επαναληπτικός αλγόριθμος που προκύπτει από την ανάπτυξη της εξίσωσης οπτικής ροής σύμφωνα με Horn και Schunck όπως αναλύθηκε σε προηγούμενη ενότητα, ενώ για την περιγραφή της σχέσης μεταξύ των χαμηλής και υψηλής ανάλυσης καρέ χρησιμοποιήθηκε ένα γραμμικό μοντέλο. Ο αλγόριθμος που προκύπτει αξιοποιεί μια σειρά από καρέ του χαμηλής ανάλυσης βίντεο για τον υπολογισμό ενός καρέ υψηλής ανάλυσης. Τα καρέ αυτά μπορεί να βρίσκονται χρονικά πίσω ή και μπροστά από το καρέ υψηλής ανάλυσης που υπολογίζεται κάθε φορά. Επιπλέον το αρχικό υψηλής ανάλυσης καρέ του επαναληπτικού αυτού αλγορίθμου υπολογίζεται με τη μέθοδο της γραμμικής παρεμβολής.

Εφαρμογή διανυσμάτων κίνησης

Πριν την παρουσίαση και ανάλυση του αλγορίθμου χρειάζεται να αναπτυχθεί ένα μοντέλο το οποίο θα προσδιορίζει τη σχέση των διαδοχικών καρέ μέσω της εφαρμογής των διανυσμάτων κίνησης. Ο τελεστής ο οποίος προσδιορίζει τη σχέση μεταξύ των διαδοχικών καρέ του βίντεο είναι ο πίνακας C . Πρόκειται για τον πίνακα που προκύπτει από τα διανύσματα κίνησης και εκφράζει την εφαρμογή αυτών στα αντίστοιχα καρέ σε μια διαδικασία που ονομάζεται εφαρμογή κίνησης (motion compensation). Έτσι αν x_k και x_{k+1} είναι δύο διαδοχικά καρέ τότε η εφαρμογή του πίνακα $C_{k+1,k}$ στο καρέ x_{k+1} έχει σαν αποτέλεσμα το καρέ x_k όπως φαίνεται και στο σχήμα 13.



Σχήμα 13. Η σχέση μεταξύ των διαδοχικών καρέ μέσω της εφαρμογής κίνησης (motion compensation).

Αν $u_{k+1,k}$ και $v_{k+1,k}$ είναι τα διανύσματα κίνησης μεταξύ των καρέ x_{k+1} και x_k σε μορφή πίνακα μεγέθους $M \times N$ που εκφράζουν την οριζόντια και κάθετη αντίστοιχα μετακίνηση των pixels για κάθε καρέ τότε ο πίνακας $C_{k+1,k}$ μεγέθους $(M \times N) \times (M \times N)$ έχει τη μορφή που φαίνεται στο σχήμα 13 προκύπτει από τα διανύσματα κίνησης.

$$u_{k+1,k} = \begin{bmatrix} u_{11} & \cdots & u_{1N} \\ \vdots & \ddots & \vdots \\ u_{M1} & \cdots & u_{MN} \end{bmatrix} \quad v_{k+1,k} = \begin{bmatrix} v_{11} & \cdots & v_{1N} \\ \vdots & \ddots & \vdots \\ v_{M1} & \cdots & v_{MN} \end{bmatrix}$$

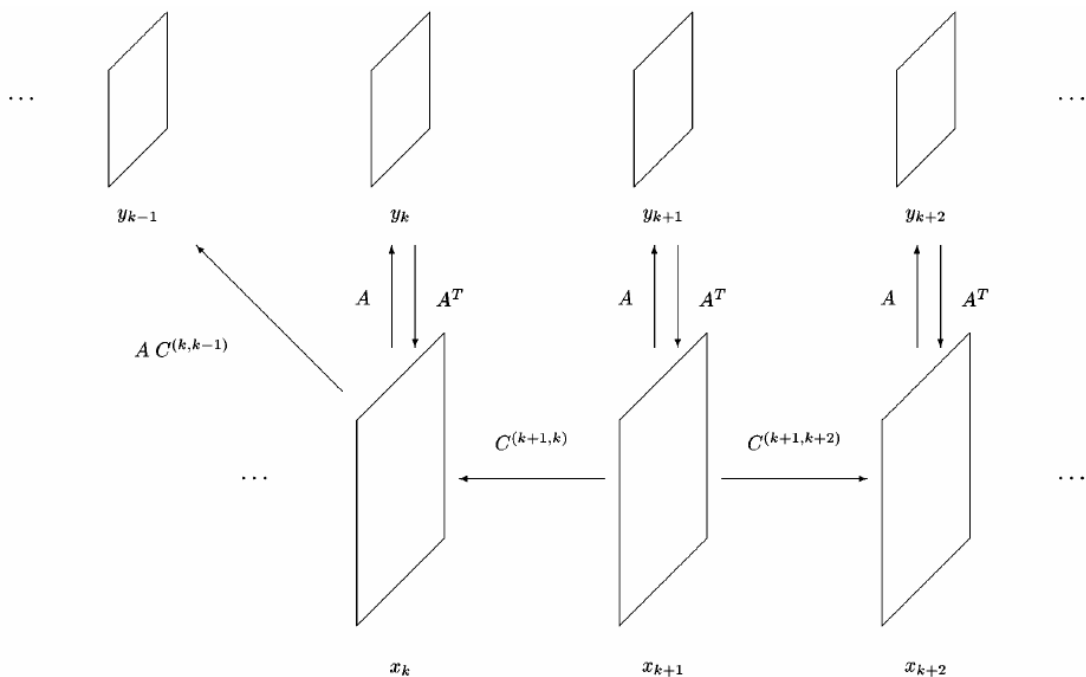
$$C_{k+1,k} = \left[\begin{array}{cccccc} \overbrace{1 & 0 & 0 & 0 & \cdots & 0}^{M \times N} \\ 0 & 1 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 1 & 0 \\ 0 & 0 & \cdots & 0 & 0 & 1 \end{array} \right] \Bigg\}^{M \times N}$$

Σχήμα 14. Ο πίνακας εφαρμογής κίνησης $C_{k+1,k}$ που διαμορφώνεται μέσω των διανυσμάτων κίνησης $u_{k+1,k}$ και $v_{k+1,k}$.

Ο πίνακας $C_{k+1,k}$ εφαρμόζεται στο καρέ x_{k+1} το οποίο βρίσκεται σε μορφή διανύσματος. Κάθε γραμμή του περιλαμβάνει μια μονάδα η οποία δείχνει σε κάποιο ρixel του καρέ. Η θέση της μονάδας αυτής πάνω στην γραμμή του πίνακα είναι αυτή που καθορίζει την καινούργια θέση του ρixel αναφοράς, και προκύπτει από τις τιμές μετατόπισης των διανυσμάτων κίνησης.

Ανάπτυξη αλγορίθμου

Το δεύτερο βήμα το οποίο είναι αναγκαίο να πραγματοποιηθεί για να είναι δυνατή η υλοποίηση του αλγορίθμου, είναι η ανάπτυξη ενός μοντέλου που θα περιγράφει τη σχέση μεταξύ των καρέ υψηλής και χαμηλής ανάλυσης.



Σχήμα 15. Σχέση μεταξύ των καρέ υψηλής και χαμηλής ανάλυσης.

Σύμφωνα με το σχήμα 15 η σχέση μεταξύ των καρέ χαμηλής και υψηλής ανάλυσης εκφράζεται μέσω του πίνακα A και του ανάστροφου αυτού A^T

οι οποίοι είναι οι πίνακες υποδειγματοληψίας (down-sampling) και υπερδειγματοληψίας (up-sampling) αντίστοιχα όπως έχουν αναλυθεί σε προηγούμενη ενότητα.

Αν x_k είναι το k -στο υψηλής ανάλυσης καρέ μεγέθους $2M \times 2N$ pixels και y_k το αντίστοιχο καρέ χαμηλής ανάλυσης μεγέθους $M \times N$ pixels σε μια ακολουθία ενός βίντεο, τότε η σχέση μεταξύ των καρέ με διαφορετική ανάλυση θα εκφράζεται από τις σχέσεις:

$$y_k = Ax_k \quad \text{και} \quad x_k = A^T y_k$$

ενώ ο πίνακας εφαρμογής των διανυσμάτων κίνησης $C^{(k+1,k)}$ περιέχει πληροφορία για την κίνηση από το καρέ υψηλής ανάλυσης x_{k+1} στο γειτονικό του καρέ x_k , οπότε θα ισχύει η σχέση:

$$x_k = C^{(k+1,k)} x_{k+1}$$

Όπως φαίνεται και στο σχήμα είναι δυνατή και η μετάβαση από ένα καρέ υψηλής ανάλυσης σε καρέ χαμηλής ανάλυσης που βρίσκεται χρονικά σε διαφορετική θέση με μια εξίσωση της μορφής:

$$y_{k-1} = AC^{(k,k-1)} x_k$$

Βασισμένα στα παραπάνω μοντέλα και με σκοπό τη μείωση της επιρροής της διαδικασίας της συμπίεσης κατά την ενίσχυση της χωρικής ανάλυσης, ο επαναληπτικός αλγόριθμος θα έχει τη μορφή:

$$x_k^{l+1} = x_k^l - \beta \nabla S_{(x)}$$

όπου ο ακέραιος l εκφράζει τις επαναλήψεις του αλγορίθμου και η ποσότητα $S_{(x)}$ αποτελείται από τους όρους που φαίνονται παρακάτω.

$$S_{(x)} = \arg \min_x \left\{ \sum_{i=-M/2}^{M/2} \|AC_{k,k-i} x_k - y_{k-i}\|^2 + \lambda_1 \|Q_1 x_k\|^2 + \lambda_2 \|Q_2 x_k\|^2 + \lambda_3 \|x_k - x_k^{mc}\|^2 \right\} \quad (1)$$

Ο πρώτος όρος $\sum_{i=-M/2}^{M/2} \|AC_{k,k-i}x_k - y_{k-i}\|^2$ αντιμετωπίζει τον θόρυβο που

εισέρχεται κατά την διαδικασία της συμπίεσης, και των λαθών που προκύπτουν κατά την μετατροπή της ανάλυσης. Η ευελιξία του μοντέλου που έχει αναπτυχθεί προσφέρει τη δυνατότητα της χρησιμοποίησης πολλαπλών γειτονικών καρτέ. Η αξιοποίηση αυτή ενός αριθμού από παρελθοντικά αλλά και μελλοντικά καρτέ προσφέρει μια καλύτερη προσέγγιση στη λύση του προβλήματος.

Οι όροι $\lambda_1 \|Q_1 x_k\|^2$ και $\lambda_2 \|Q_2 x_k\|^2$ χρησιμοποιούνται για να καλύψουν δύο ανωμαλίες που μπορούν να προκληθούν από τις συνήθεις τεχνικές συμπίεσης, κατά την μετάβαση του καρτέ σε υψηλή χωρική ανάλυση. Η πρώτη ανωμαλία αφορά στην εφαρμογή της κβαντοποίησης σε περιοχές με υψηλή συχνότητα με αποτέλεσμα την οπτική αλλοίωση της εικόνας, ενώ η δεύτερη προκαλείται από την ανεξάρτητη επεξεργασία των περιοχών (blocks) της εικόνας η οποία εισάγει λάθη στα όρια των περιοχών αυτών. Οι τελεστές Q_1 και Q_2 αποτελούν υπερπαρατά φίλτρα, ενώ οι αριθμοί λ_1 και λ_2 ελέγχουν την βαρύτητά τους. Κατά την υλοποίηση του αλγορίθμου ο όρος $\lambda_2 \|Q_2 x_k\|^2$ παραλήφθηκε αφού η προσέγγιση που ακολουθήθηκε ήταν βασισμένη σε pixels και όχι σε περιοχές από pixels (blocks).

Τέλος ο όρος $\lambda_3 \|x_k - x_k^{mc}\|^2$ εισάγει στην εξίσωση την πολύτιμη πληροφορία των διανυσμάτων κίνησης. Το καρτέ x_k^{mc} είναι το καρτέ που προκύπτει από την εφαρμογή των διανυσμάτων κίνησης στο καρτέ υψηλής ανάλυσης x_{k-1} του οποίου ο υπολογισμός προηγήθηκε. Το αποτέλεσμα του όρου αυτού συμβάλει καθοριστικά στον σκοπό του αλγορίθμου που δεν είναι άλλος από την καλύτερη απόδοση του υψηλής ανάλυσης καρτέ.

Η σχέση (1) μπορεί να αναπτυχθεί ως εξής:

$$S_{(x)} = \arg \min_x \left\{ \sum_{i=-M/2}^{M/2} (AC_{k,k-i}x_k - y_{k-i})^T (AC_{k,k-i}x_k - y_{k-i}) + \lambda_1 (Q_1 x_k)^T (Q_1 x_k) \right.$$

$$+ \lambda_2 (Q_2 x_k)^T (Q_2 x_k) + \lambda_3 (x_k - x_k^{mc})^T (x_k - x_k^{mc}) \}$$

Διαφορίζοντας την εξίσωση θα έχουμε:

$$\nabla S_{(x)} = 2C_{k,k-i}^T A^T A C_{k,k-i} x_k - 2C_{k,k-i}^T A^T y_{k-i} + 2\lambda_1 Q_1^T Q_1 x_k + 2\lambda_2 Q_2^T Q_2 x_k + 2\lambda_3 (x_k - x_k^{mc})$$

Οπότε τελικά προκύπτει ο επαναληπτικός και συγκλινόμενος αλγόριθμος που έχει την μορφή:

$$x_k^{l+1} = x_k^l - \beta \left[\sum_{i=-M_2}^{M_1} C_{k,k-i}^T A^T (A C_{k,k-i} x_k^l - y_{k-i}) + \lambda_1 Q_1^T Q_1 x_k^l + \lambda_2 Q_2^T Q_2 x_k^l + \lambda_3 (x_k - x_k^{mc}) \right]$$

Οι τιμές x_k^l και x_k^{l+1} αποτελούν τα ενισχυμένης ανάλυσης καρέ στις επαναλήψεις l και $l+1$ αντίστοιχα. Ο συντελεστής β είναι μια παράμετρος που ελέγχει την σύγκλιση και τον ρυθμό σύγκλισης του αλγορίθμου. Ο αλγόριθμος συγκλίνει με βάση τον όρο $\|x_k^{l+1} - x_k^l\|^2$. Με το πέρασμα των επαναλήψεων, η διαφορά αυτή μεταξύ των γειτονικών αποτελεσμάτων του εξαγόμενου υψηλής ανάλυσης καρέ μικραίνει, ώσπου φθάνει σε μια τιμή σύγκλισης.

Το άθροισμα που περιέχεται στον αλγόριθμο χρησιμοποιεί αριθμό ίσο με M_2 προηγούμενα χρονικά καρέ και M_1 επόμενα καρέ, από το καρέ προς επεξεργασία. Έτσι ο αλγόριθμος αντιμετωπίζει πιο αποτελεσματικά πιθανά λάθη που εισάγονται κατά την συμπίεση και αποσυμπίεση του βίντεο.

Το υπεραπλοποιημένο φίλτρο που χρησιμοποιήθηκε είναι το Laplacian το οποίο σε μορφή πίνακα έχει το περιεχόμενο που φαίνεται στο σχήμα 16.

$$q_1 = \begin{bmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{bmatrix} \quad Q_1 = \left[\begin{array}{cccccc} \overbrace{-4 & 1 & \dots & 1 & \dots & 0}^{2M \times 2N} \\ 1 & -4 & 1 & \dots & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & 1 & 0 \\ 0 & \dots & 1 & \dots & 1 & -4 \end{array} \right] \left. \vphantom{\begin{array}{c} \\ \\ \\ \\ \end{array}} \right\} 2M \times 2N$$

Σχήμα 16. Το υψιπερατό φίλτρο Laplacian σε μορφή μάσκας και σε μορφή πίνακα.

Αξίζει να σημειωθεί ότι ο πίνακας εφαρμογής των διανυσμάτων κίνησης $C_{k,k-i}^T$ ο οποίος είναι ο ανάστροφος του πίνακα που περιέχει πληροφορία για την κίνηση από το καρέ x_k στο καρέ x_{k-i} , ισοδυναμεί με τον πίνακα που περιέχει την πληροφορία κίνησης από το καρέ x_{k-i} στο καρέ x_k δηλαδή με τον πίνακα $C_{k-i,k}$.

Σε επίπεδο υλοποίησης το υψιπερατό φίλτρο εφαρμόστηκε ως μάσκα με την διαδικασία της συνέλιξης, ενώ η εφαρμογή των διανυσμάτων κίνησης (motion compensation) δεν έγιναν μέσω του μοντέλου του πίνακα, αλλά με ολίσθηση των εντάσεων για κάθε pixel σύμφωνα με τις μετατοπίσεις που προέκυψαν από τον αλγόριθμο κατά Horn και Sunk.

Μέθοδος χωροχρονικής παρεμβολής κατά A.Smolic, K.Rummler, J.Ohm, R.Schafer και S.Bauer

Η προσπάθεια για την ανάπτυξη αξιόπιστων συστημάτων μεταφοράς δεδομένων πολυμέσων, οδήγησε στην ανάπτυξη αλγορίθμων μετατροπής της ανάλυσης με μεγάλο βαθμό πολυπλοκότητας. Αλγόριθμοι όπως αυτός που αναλύθηκε στο προηγούμενο κεφάλαιο, προσφέρουν μεν αποτελέσματα υψηλής ποιότητας και αξιοπιστίας, αντιμετωπίζοντας την επιρροή του θορύβου ή λαθών που προκαλούνται κατά την μετατροπή του βίντεο από υψηλή σε χαμηλή ανάλυση και αντιστρόφως, εντούτοις έχουν το σημαντικό μειονέκτημα του μεγάλου υπολογιστικού κόστους.

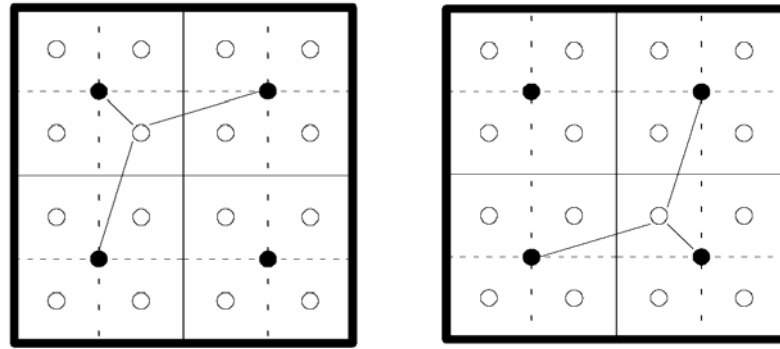
Σε αυτήν την ενότητα θα αναπτυχθούν μέθοδοι χαμηλού υπολογιστικού κόστους που υπολογίζουν βίντεο ενισχυμένης χωροχρονικής ανάλυσης. Η μέθοδος χωρικής παρεμβολής που θα αναλυθεί ενισχύει την ανάλυση, αξιοποιώντας την χρονική μετατόπιση των pixels ή οποία εκφράζεται μέσω των διανυσμάτων κίνησης, και για να αντιμετωπίσει πιθανές αστοχίες αυτών, χρη-

σιμοποιεί ένα μη γραμμικό φίλτρο που εξισορροπεί την έξοδο. Στην μέθοδο χρονικής παρεμβολής, στην υπάρχουσα ακολουθία από καρέ προστίθενται και νέα, αυξάνοντας έτσι την χρονική ανάλυση του βίντεο. Και σε αυτή τη μέθοδο καθοριστικό ρόλο για την παραγωγή των ενδιάμεσων καρέ αποτελεί η πληροφορία κίνησης.

Οι μέθοδοι κατά A.Smolic, K.Rummeler, J.Ohm, R.Schafer και S.Bauer χρησιμοποιούν διανύσματα κίνησης τα οποία αντιστοιχούν στο υψηλής ανάλυσης βίντεο. Έτσι υπάρχει η ανάγκη της ενίσχυσης της πυκνότητας της πληροφορίας κίνησης. Για την επίτευξη αυτού του στόχου αναπτύχθηκε ένα σχήμα παρεμβολής(motion interpolation) το οποίο αυξάνει την πληροφορία των διανυσμάτων κίνησης, με τη χρήση ενός φίλτρου μεσαίου(median) το οποίο προλαμβάνει πιθανά λάθη και αστοχίες της κίνησης.

Παραγωγή διανυσμάτων κίνησης για βίντεο υψηλής ανάλυσης με Median Interpolation

Η ανάγκη ελάττωσης της πληροφορίας κατά την μετάδοση ενός βίντεο, με σκοπό να έχει όσο το δυνατόν πιο συμπιεσμένη μορφή ώστε να εξυπηρετεί τις ανάγκες των καναλιών μετάδοσης, μπορεί να οδηγήσει σε δειγματοληψία ακόμα και στο πεδίο των διανυσμάτων κίνησης. Συγκρινόμενα με την χωρική πληροφορία ενός καρέ, τα διανύσματα κίνησης που εκφράζουν την μετατόπιση των pixels αυτού δεν έχουν ασφαλώς τον ίδιο όγκο πληροφορίας. Όμως σε συστήματα όπως το Mpeg η παρουσία τους σε αναλογία με την χωρική πληροφορία ενός βίντεο είναι αισθητή. Έτσι η ανάπτυξη μιας μεθόδου παρεμβολής(interpolation) στα διανύσματα κίνησης, της οποίας το αποτέλεσμα θα προσέγγιζε το αρχικό μέγεθος των μετατοπίσεων.



Σχήμα 17. Απεικόνιση της λειτουργίας του median φίλτρου για την παραγωγή διανυσμάτων υψηλής ανάλυσης.

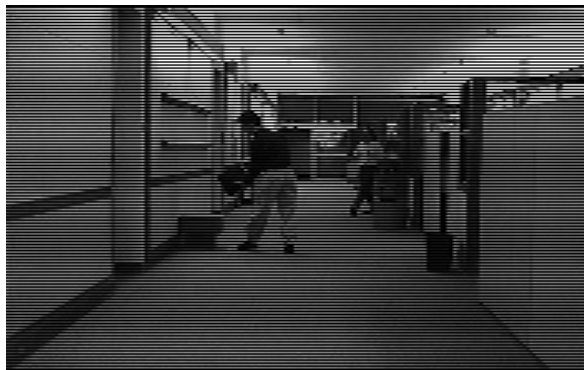
Όπως φαίνεται και στο σχήμα 17 κάθε νέα μετατόπιση που αντιστοιχεί σε pixel του υψηλής ανάλυσης καρέ, υπολογίζεται μέσω των τριών κοντινότερων τιμών των διανυσμάτων μικρής πυκνότητας. Στην εικόνα οι άδειες κουκκίδες αντιστοιχούν στις νέες τιμές των διανυσμάτων ενώ οι μαύρες κουκκίδες στις τιμές που μεταδόθηκαν. Οι νέες τιμές υπολογίζονται μέσω της εφαρμογής της μη γραμμικής διαδικασίας του “μεσαίου”(median). Το φίλτρο λαμβάνει σαν είσοδο τις τρεις κοντινότερες τιμές των διανυσμάτων χαμηλής ανάλυσης, και διαλέγει από αυτές την τιμή που αντιστοιχεί στον “μεσαίο”. Η τιμή αυτή προκύπτει κατά την ταξινόμηση των τιμών αυτών, όπου η μεσαία τιμή στην ταξινομημένη σειρά αποτελεί την έξοδο του φίλτρου. Ο λόγος για τον οποίο χρησιμοποιήθηκε αυτή η μη γραμμική διαδικασία αντί μιας γραμμικής στηρίζεται στο γεγονός ότι με την πρώτη αντιμετωπίζονται καλύτερα οι σημειακές λανθασμένες μετατοπίσεις.

Μέθοδος χωρικής παρεμβολής (De-interlacing)

Η μέθοδος που θα αναπτυχθεί ολοκληρώνει την διαδικασία αύξησης της χωρικής ανάλυσης ενός βίντεο, με την κάλυψη των άγνωστων τιμών έντασης που προκύπτουν από την αύξηση των διαστάσεων των καρέ του βίντεο. Η μέθοδος συνδυάζει τη γραμμική μέθοδο παρεμβολής, με την χρήση της πληροφορίας που δίνει η μετατόπιση των pixels ανάμεσα σε δύο καρέ για την εξαγωγή καλύτερων αποτελεσμάτων. Παίρνει σαν είσοδο δύο συνεχόμενα

καρέ ενός βίντεο τα οποία περιέχουν γραμμές με μηδενικές τιμές έντασης (interlace input). Οι τιμές αυτές έντασης του καρέ εξόδου υπολογίζονται μέσω της μη γραμμικής διαδικασίας του μεσαίου (median).

Τα καρέ εισόδου όπως αναφέρθηκε έχουν μια συγκεκριμένη μορφή (interlace). Όπως φαίνεται και στο σχήμα 18 οι γραμμές του καρέ αντιστοιχούν ανα μία σε γραμμές των οποίων τα pixels έχουν μηδενικές τιμές έντασης. Οι υπάρχουσες τιμές των υπολοίπων γραμμών του καρέ προέρχονται από το καρέ χαμηλής ανάλυσης και αποτελούν τη γνήσια πληροφορία.

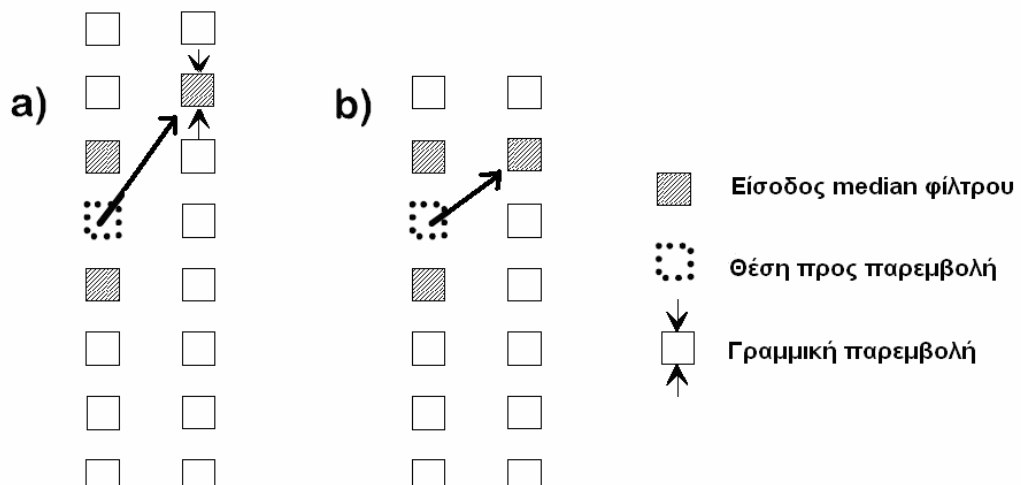


Σχήμα 18. Η μορφή του καρέ εισόδου (interlace input).

Η μέθοδος χρησιμοποιεί δύο διαδοχικά καρέ του βίντεο τα οποία βρίσκονται στη μορφή που περιγράφηκε. Το ενισχυμένης ανάλυσης καρέ που θα ληφθεί σαν έξοδος, προκύπτει από το καρέ εισόδου της χρονικής στιγμής t όπως φαίνεται και στο σχήμα 20. Η πληροφορία του καρέ εισόδου $t+1$ που ακολουθεί αξιοποιείται σε συνδυασμό με τα διανύσματα κίνησης μεταξύ των δύο καρέ, τα οποία εκφράζονται ως μετατόπιση για κάθε pixel με τη μορφή του βέλους.

Για να μετατραπεί το καρέ εισόδου t σε καρέ υψηλής ανάλυσης, χρειάζεται να συμπληρωθούν τα pixels με μηδενική ένταση στις γραμμές που αυτά υπάρχουν. Έτσι σύμφωνα με το σχήμα 20 κάθε τέτοιο pixel παίρνει την τιμή που προκύπτει από την έξοδο του φίλτρου του μεσαίου (median). Το φίλτρο αυτό λαμβάνει τις τιμές έντασης τριών pixels, τα δύο αντιστοιχούν στα pixels που βρίσκονται άνω και κάτω από το pixel προς παρεμβολή και αποτελούν

μέρος της γνήσιας πληροφορίας και το τρίτο λαμβάνεται από το επόμενο καρέ $t+1$ ως εξής. Με την εφαρμογή των διανυσμάτων κίνησης στο pixel προς επεξεργασία για την μετάβαση στο καρέ $t+1$, δηλαδή την οριζόντια και την κάθετη μετακίνηση του pixel που εκφράζονται από τις τιμές του πινάκων u και v όπως αναλύθηκαν σε προηγούμενη ενότητα. Τότε η τιμή της έντασης του pixel του καρέ $t+1$ στο οποίο δείχνει η μετατόπιση αυτή, αναλόγως με τη θέση που βρίσκεται αποτελεί την τρίτη τιμή εισόδου του φίλτρου.



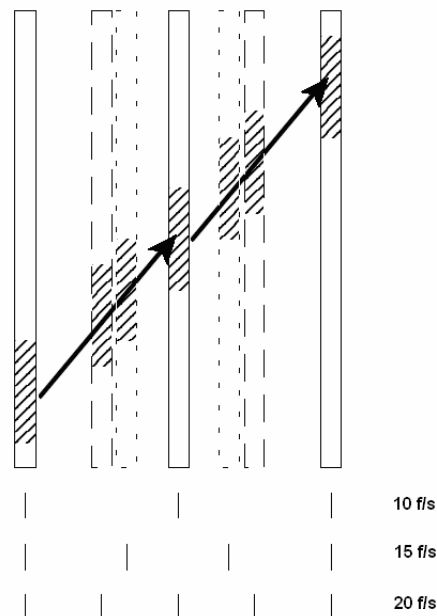
Σχήμα 19. Οι είσοδοι του median φίλτρου α) Για άρτια μετατόπιση. β) Για περιττή μετατόπιση.

Όταν η κατακόρυφη μετατόπιση αντιστοιχεί σε ζυγό αριθμό, αυτό σημαίνει ότι η τιμή του καρέ $t+1$ υπάρχει οπότε αυτή η τιμή της έντασης εισάγεται στο φίλτρο. Όταν όμως η κατακόρυφη μετατόπιση συμβαίνει κατά μονό αριθμό, τότε η τιμή εισόδου του φίλτρου λαμβάνεται από τον γραμμικό συνδυασμό των εντάσεων των δύο pixels που περιβάλλουν τη θέση με μηδενική πληροφορία.

Χρονική μετατροπή ανάλυσης

Στα πλαίσια της διακωδικοποίησης (transcoding), η ανάγκη για προσαρμογή των συστημάτων αναπαραγωγής βίντεο σε τηλεπικοινωνιακό επίπεδο, δεν εξαντλείται μόνο στην μετατροπή της ανάλυσης στον χώρο. Η μετατροπή της χρονικής ανάλυσης παίζει εξίσου σημαντικό ρόλο στην διαδικασία αύξησης ή μείωσης της πυκνότητας της πληροφορίας ενός βίντεο. Η έννοια της χρονικής πυκνότητας της πληροφορίας για ένα βίντεο αντιστοιχεί με τον αριθμό των καρέ που προβάλλονται σε αυτό, σε σχέση με την αρχική πυκνότητα των καρέ του γνήσιου βίντεο. Ανάλογος με τον αριθμό των καρέ που απαρτίζουν το βίντεο είναι και ο ρυθμός μετάδοσης των καρέ (frame rate), ο οποίος μετριέται συνήθως σε καρέ ανά δευτερόλεπτο.

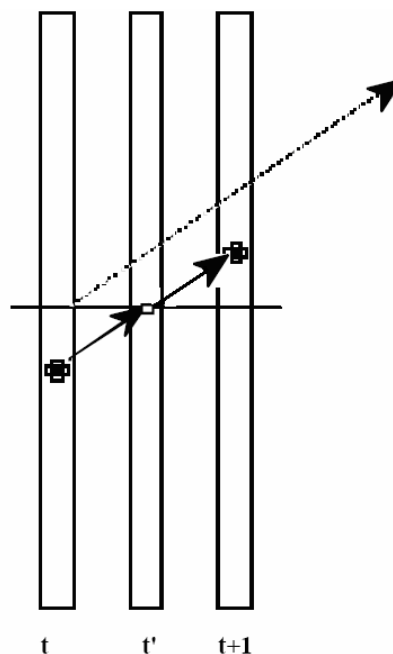
Στην συνέχεια θα αναπτυχθεί μια χαμηλής πολυπλοκότητας μέθοδος, η οποία παρεμβάλλει επιπρόσθετα καρέ ανάμεσα στα υπάρχοντα, αυξάνοντας έτσι την χρονική ανάλυση του βίντεο. Καθοριστική σημασία σε αυτήν την μέθοδο έχει η πληροφορία κίνησης ανάμεσα στα καρέ που εκφράζεται με τα διανύσματα κίνησης.



Σχήμα 20. Χρονικές θέσεις των καρέ για διάφορους ρυθμούς μετάδοσης.

Μέθοδος παρεμβολής ενδιάμεσου(Intermediate image Interpolation)

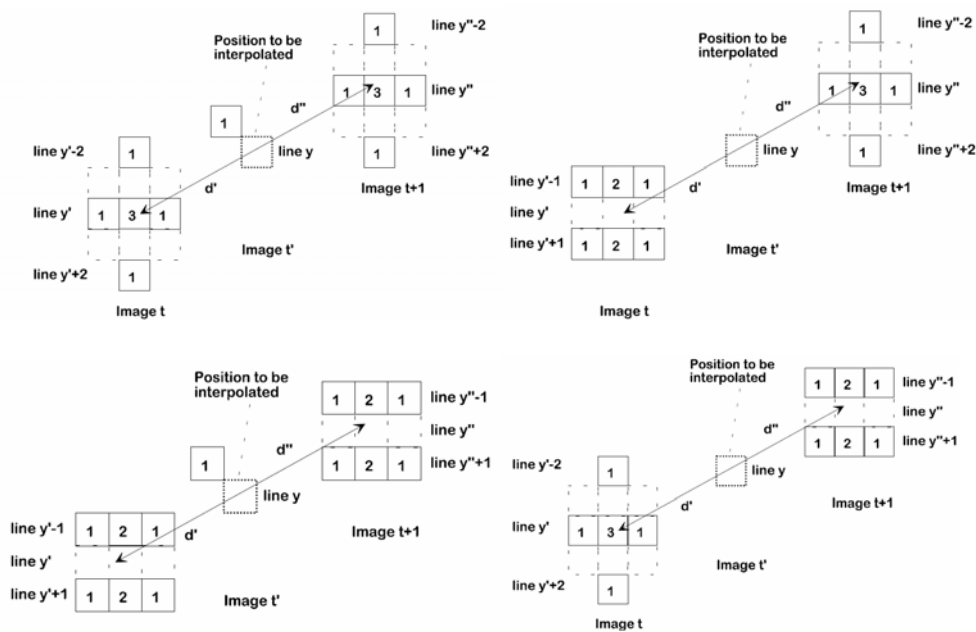
Σκοπός της μεθόδου αυτής είναι η δημιουργία νέων καρέ ανάμεσα σε αυτά που υπάρχουν, με αποτέλεσμα την ενίσχυση της χρονικής ανάλυσης ενός συμπιεσμένου βίντεο. Η μέθοδος βασίζεται στην τμηματοποίηση (scaling) των διανυσμάτων κίνησης ανάμεσα σε δύο καρέ, έτσι ώστε να παραχθούν δύο νέα πεδία διανυσμάτων κίνησης όπως φαίνεται και στο σχήμα 21. Μετά από αυτό το βήμα ακολουθεί η δημιουργία του καρέ t' το οποίο παρεμβάλλεται χρονικά ανάμεσα στα καρέ t και $t+1$ τα οποία είτε αποτελούνται εξ ολοκλήρου από γνήσια πληροφορία είτε βρίσκονται σε μορφή interlaced που περιγράφηκε παραπάνω.



Σχήμα 21. Απεικόνιση της τμηματοποίησης των διανυσμάτων κίνησης (motion scaling) για την παραγωγή του καρέ παρεμβολής t' .

Το πρώτο βήμα για την ανάπτυξη της μεθόδου χρονικής παρεμβολής είναι ο υπολογισμός των δύο νέων πεδίων διανυσμάτων κίνησης που προκύπτουν από το διάνυσμα κίνησης ανάμεσα στα καρέ t και $t+1$. Αν d είναι ο πίνακας διανυσμάτων κίνησης μεταξύ των καρέ t και $t+1$, τότε ο πίνακας διανυ-

σμάτων που εκφράζει την μετατόπιση από το καρέ t στο καρέ t' υπολογίζεται από τη σχέση $d' = \text{nint}[d \cdot m]$, όπου m ισούται με $m = 2^{-n}$, ενώ ο τελεστής nint ισοδυναμεί με την εύρεση του κοντινότερου ακέραιου (nearest integer). Τότε ο πίνακας των διανυσμάτων κίνησης μεταξύ των καρέ t' και t θα υπολογίζεται από τη σχέση $d'' = d - d'$. Ο αριθμός n είναι ένας ακέραιος ο οποίος εκφράζει τον βαθμό τμηματοποίησης του αρχικού διανύσματος. Αν $n=1$ τότε η μετατόπιση τόσο η οριζόντια όσο και η κάθετη χωρίζεται σε δύο ίσες συνιστώσες, περίπτωση που αντιστοιχεί με την αλλαγή του ρυθμού μετάδοσης καρέ σε 20 f/s στο σχήμα 20. Για την παραγωγή όμως ενδιάμεσου καρέ όπως φαίνεται στην περίπτωση των 15 f/s χρειάζεται διαίρεση της μετατόπισης κατά $1/4$, οπότε ο αριθμός n θα ισούται με 2.



Σχήμα 22. Φίλτρα για τις διάφορες περιπτώσεις *interlaced* εισόδου.

Το νέο καρέ σχηματίζεται με την εφαρμογή των διανυσμάτων κίνησης d' που υπολογίστηκαν, από το καρέ t στο καρέ t' (motion compensation). Όμως τα καρέ που περιβάλλουν αυτό που πρόκειται να υπολογιστεί μπορεί να έχουν ολοκληρωμένη πληροφορία ή να είναι ημιτελή (*interlace*). Στην δεύτερη περίπτωση η εμφάνιση πολλών *pixel* στο νέο καρέ με μηδενική ένταση είναι αναπόφευκτη. Έτσι ανάλογα με το αν τα νέα διανύσματα κίνησης d' και

d'' δείχνουν σε διαθέσιμα ή μη pixels, διακρίνονται τέσσερις περιπτώσεις για τον υπολογισμό κάθε τέτοιου pixel του νέου καρέ. Στην πρώτη περίπτωση και τα δύο pixels στα οποία δείχνουν τα διανύσματα κίνησης είναι διαθέσιμα, στις επόμενες δύο μόνο το ένα από τα δύο pixels είναι διαθέσιμο, ενώ στην τελευταία περίπτωση και κανένα από τα δύο δεν είναι διαθέσιμο. Για τον υπολογισμό των pixel με μηδενική ένταση χρησιμοποιήθηκαν δύο εφαρμογές median φίλτρων. Όπως διακρίνεται και στο σχήμα 22 ανάλογα με το αν το pixel του καρέ είναι διαθέσιμο ή όχι, γίνεται χρήση του αντίστοιχου φίλτρου. Το φίλτρο μεσαίου παίρνει σαν εισόδους τις τιμές και από τα δύο καρέ αναλόγως με την κάθε περίπτωση, καθώς και μία τιμή που έχει υπολογιστεί ήδη και προέρχεται από το ενδιάμεσο καρέ t' η οποία βρίσκεται στην πάνω αριστερή γωνία από το pixel προς υπολογισμό. Η τιμή αυτή χρησιμοποιείται μόνο στην περίπτωση κατά την οποία ο αριθμός των εισόδων του φίλτρου είναι περιττός.

Χρειάζεται να σημειωθεί ότι κατά την εφαρμογή των διανυσμάτων κίνησης για την παραγωγή του νέου καρέ, εμφανίζονται και πάλι pixels μηδενικής έντασης, τα οποία αντιμετωπίζονται με τον ίδιο τρόπο που περιγράφηκε παραπάνω. Επιπλέον το διάνυσμα d' επεξεργάστηκε έτσι ώστε να αντιστοιχεί με την κίνηση από το καρέ t' στο καρέ t , λόγω απαίτησης της μεθόδου.

Πειραματικά Αποτελέσματα

Η ανάλυση των πειραματικών αποτελεσμάτων που προέκυψαν ύστερα από την υλοποίηση των μεθόδων που παρουσιάστηκαν, αποτελεί το βασικότερο κριτήριο για την ερμηνεία αλλά και την σύγκριση των αλγορίθμων ενίσχυσης της χωροχρονικής ανάλυσης. Κατά την ανάλυση των αποτελεσμάτων δόθηκε μεγάλη βάση στην ποιοτική τους διάσταση αλλά και στην υπολογιστική πολυπλοκότητα, που είναι ανάλογη με το κόστος της χρονικής καθυστέρησης, των αλγορίθμων από τους οποίους εξήχθησαν.

Για την υλοποίησή όλων των μεθόδων που παρουσιάστηκαν χρησιμοποιήθηκε αποκλειστικά η γλώσσα προγραμματισμού C για λόγους ευελιξίας και ταχύτητας. Για λόγους ευκολίας και άμεσης αντιπαράθεσης αποτελεσμάτων, τόσο η είσοδος όσο και η έξοδος του προγράμματος είναι τα καρέ του βίντεο υπό την μορφή ασπρόμαυρων ασυμπίεστων εικόνων.

Ως μέτρο σύγκρισης των αποτελεσμάτων των αλγορίθμων, χρησιμοποιήθηκε το ενεργειακό φάσμα (power spectrum) της διαφοράς των καρέ υψηλής ανάλυσης του γνήσιου βίντεο g_k , με τα καρέ που εξήχθησαν από κάθε έναν από τους αλγορίθμους, a_k για τον επαναληπτικό αλγόριθμο και b_k για τον δεύτερο αλγόριθμο καθώς επίσης και με τα καρέ που υπολογίζονται με τη γραμμική μέθοδο l_k . Το ενεργειακό φάσμα υπολογίζεται ως εξής:

$$\|g_k - a_k\|^2 = \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} (g_{ij} - a_{ij})^2. \text{ Έτσι συγκρίνονται οι τιμές } \|g_k - a_k\|^2, \|g_k - b_k\|^2 \text{ και}$$

$\|g_k - l_k\|^2$ ώστε να προσδιοριστεί ποια σειρά αποτελεσμάτων προσεγγίζει καλύτερα την γνήσια πληροφορία του βίντεο. Πρώτα θα συγκριθούν τα αποτελέσματα κάθε αλγορίθμου ξεχωριστά, με τα καρέ που προκύπτουν από την γραμμική παρεμβολή και έπειτα θα συγκριθούν μεταξύ τους. Σημειώνεται ότι σαν είσοδο στους αλγορίθμους ενίσχυσης της χωρικής ανάλυσης χρησιμοποιήθηκαν καρέ προερχόμενα από δύο βίντεο τα οποία περιείχαν σκηνές από ακίνητη και κινητή κάμερα, για την καλύτερη αξιοπιστία των αποτελεσμάτων.

Σύγκριση με γραμμική μέθοδο

Τα καρέ υψηλής ανάλυσης που εξήχθησαν από τον επαναληπτικό αλγόριθμο κατά J. Mateos A. Katsaggelos και R. Molina διακρίθηκαν σε δύο περιπτώσεις που αφορούν την είσοδο και συγκεκριμένα τον όρο $x_k - x_k^{mc}$ που εισάγει πληροφορία από προηγούμενο υψηλής ανάλυσης καρέ. Όταν το εφαρμοσμένο από τα διανύσματα κίνησης καρέ x_k^{mc} προκύψει από προηγούμενο καρέ γνήσιας πληροφορίας τότε η βελτίωση του καρέ εξόδου της μεθόδου σε σχέση με το γραμμικά υπολογισμένο καρέ φθάνει και το 90%. Η έννοια της ποσοστιαίας βελτίωσης προκύπτει από την σύγκριση των ενεργειακών φασμάτων ως εξής:

$$\frac{\|g_k - l_k\|^2 - \|g_k - a_k\|^2}{\|g_k - a_k\|^2} \cdot 100\%.$$

Όταν η γνήσια πληροφορία

προηγούμενων καρέ δεν χρησιμοποιηθεί, τότε και πάλι υπάρχει βελτίωση των αποτελεσμάτων σε αρκετά μικρότερο ποσοστό όπως αναμενόταν της τάξης του 1%. Στο πεδίο της χρονικής καθυστέρησης που είναι ανάλογη με την υπολογιστική πολυπλοκότητα, ο επαναληπτικός αλγόριθμος εισάγει σαφώς μεγαλύτερη καθυστέρηση. Οι χρόνοι υλοποίησης που μετρήθηκαν έδειξαν ότι η καθυστέρηση του αλγορίθμου είναι πολλαπλάσια από αυτήν της γραμμικής μεθόδου.

Στην συνέχεια πραγματοποιήθηκε σύγκριση των αποτελεσμάτων του αλγορίθμου κατά A.Smolc, K.Rummler, J.Ohm, R.Schafer και S.Bauer για την αύξηση της χωρικής ανάλυσης, με αυτά της γραμμικής παρεμβολής. Ακολουθώντας την ίδια διαδικασία σύγκρισης βρέθηκε ότι τα καρέ υψηλής ανάλυσης που υπολογίζονται με την αναφερόμενη μέθοδο, προσεγγίζουν περισσότερο την γνήσια πληροφορία σε σχέση με τα καρέ που προκύπτουν από την γραμμική μέθοδο. Με βάση το ενεργειακό φάσμα πάλι, υπολογίστηκε ότι κατά μέσο όρο η βελτίωση είναι της τάξης του 65%. Όσον αφορά την καθυστέρηση, δεν προέκυψαν σημαντικές διαφορές στον χρόνο υλοποίησης αφού η προτεινόμενη μέθοδος όπως και η γραμμική μέθοδος έχουν χαμηλή υπολογιστική πολυπλοκότητα.



Σχήμα 23. Παραγωγή του καρέ 091 (Hall), Γνήσιο καρέ , Γραμμική παρεμβολή , Επαναληπτικός αλγόριθμος.



Σχήμα 24. Παραγωγή του καρέ 121 (Hall), Γνήσιο καρέ , Γραμμική παρεμβολή , Επαναληπτικός αλγόριθμος.



Σχήμα 25. Παραγωγή του καρέ 013 (garden) : Γνήσιο καρέ , Γραμμική παρεμβολή , Επαναληπτικός αλγόριθμος



Σχήμα 26. Παραγωγή του καρέ 048 (coast) : Γνήσιο καρέ , Γραμμική παρεμβολή , Επαναληπτικός αλγόριθμος



Σχήμα 27. Παραγωγή του καρέ 013 (garden) , Γνήσιο καρέ , Γραμμική παρεμβολή , Μέθοδος De-interlace



Σχήμα 28. Παραγωγή του καρέ 090 (Hall) , Γνήσιο καρέ , Γραμμική παρεμβολή , Μέθοδος De-interlace



Σχήμα 29. Παραγωγή του καρέ 120 (Hall) , Γνήσιο καρέ , Γραμμική παρεμβολή , Μέθοδος De-interlace



Σχήμα 30. Παραγωγή του καρέ 48 (Coast) , Γνήσιο καρέ , Γραμμική παρεμβολή , Μέθοδος De-interlace

Σύγκριση μεταξύ τους

Αρχικά θα πρέπει να συγκριθούν οι δύο μέθοδοι σύμφωνα με το εύρος της λειτουργικότητάς τους. Η μέθοδος κατά J. Mateos A. Katsaggelos και R. Molina υπολογίζει βίντεο υψηλής χωρικής ανάλυσης με την χρήση των διανυσμάτων κίνησης του γνήσιου υψηλής ανάλυσης βίντεο. Αντίθετα η μέθοδος κατά A.Smolc, K.Rummler, J.Ohm, R.Schafer και S.Bauer περιλαμβάνει όπως είδαμε μία εφαρμογή στην οποία ενισχύεται η πυκνότητα πληροφορίας των χαμηλής ανάλυσης διανυσμάτων κίνησης ώστε να αντιστοιχούν σε βίντεο υψηλής ανάλυσης. Επιπλέον η ίδια μέθοδος περιλαμβάνει και ενίσχυση της χρονικής ανάλυσης, με την παραγωγή νέων καρέ τα οποία παρεμβάλλονται ανάμεσα στα υπάρχοντα, ενώ η μέθοδος κατά J. Mateos A. Katsaggelos και R. Molina δεν περιλαμβάνει ενίσχυση της ανάλυσης στον χρόνο.

Συγκρίνοντας τις δύο μεθόδους στην λειτουργία της αύξησης της χωρικής ανάλυσης ως προς την ποιότητα των αποτελεσμάτων τους, ο επαναληπτικός αλγόριθμος προτιμάται έναντι του δεύτερου αλγορίθμου σε περίπτωση που υπάρχει διαθέσιμο μέρος της πληροφορίας του γνήσιου βίντεο. Όταν δεν υπάρχει καμία τέτοια πληροφορία τότε ο προτεινόμενος αλγόριθμος προσφέρει καλύτερη προσέγγιση στο υψηλής ανάλυσης βίντεο.

Στον πίνακα που ακολουθεί φαίνονται ενδεικτικές τιμές των φασμάτων ισχύος των διαφορών των εξόδων κάθε αλγορίθμου με τα γνήσια καρέ για τα τρία βίντεο.

	Επαναληπτικός Αλγόριθμος	Μέθοδος De-interlace	Γραμμική μέθοδος	Συνδυασμένος αλγό- ριθμος
Hall	2.247.324	11.003.157	23.840.283	4.512.371
Garden	70.880.243	82.936.312	151.395.889	81.614.104
Coast	15.312.392	28.016.802	35.001.026	17.632.098



Σχήμα 31. Παραγωγή του καρέ 120 (Hall), Γνήσιο καρέ , Μέθοδος De-interlace , Επαναληπτικός αλγόριθμος



Σχήμα 32. Παραγωγή του καρέ 013 (garden) , Γνήσιο καρέ , Μέθοδος De-interlace, Επαναληπτικός αλγόριθμος

Αξίζει να αναφερθεί ότι η μέθοδος κατά J. Mateos A. Katsaggelos και R. Molina συνδυάστηκε με ένα μέρος της δεύτερης μεθόδου. Αντί για τα γνήσια υψηλής ανάλυσης διανύσματα κίνησης, χρησιμοποιήθηκαν αυτά που υπολογίστηκαν από την εφαρμογή της μεθόδου ενίσχυσης της πληροφορίας κίνησης, στα αρχικά χαμηλής ανάλυσης διανύσματα. Τα αποτελέσματα που προέκυψαν ήταν ικανοποιητικά αφού δεν διέφεραν σημαντικά από εκείνα που χρησιμοποιούσαν τα γνήσια υψηλής ανάλυσης διανύσματα. Παραδείγματα εξόδου φαίνονται παρακάτω.



Σχήμα 33. Παραγωγή του καρτέ 119 (Hall) με ενισχυμένα διανύσματα κίνησης, Γνήσιο καρτέ, Γραμμική παρεμβολή, Επαναληπτικός αλγόριθμος

Παρακάτω φαίνονται παραδείγματα εξόδου της μεθόδου χρονικής παρεμβολής. Ανάμεσα σε δύο συνεχόμενα καρέ η μέθοδος παράγει ένα ενδιάμεσο καρέ που προκύπτει από την εφαρμογή διανυσμάτων κίνησης τα οποία έχουν χωριστεί, διαιρώντας έτσι την κίνηση σε ίσα μέρη, ή κατά $1/4$ και $3/4$ αντίστοιχα.



Σχήμα 34. Παραγωγή ενδιάμεσου καρέ (Hall) με motion scaling κατά $1/4$, Γνήσιο καρέ 98, Ενδιάμεσο καρέ εξόδου, Γνήσιο καρέ 99



Σχήμα 35. Παραγωγή ενδιάμεσου καρέ (Hall) με motion scaling κατά 1/2, Γνήσιο καρέ 98 , Ενδιάμεσο καρέ εξόδου , Γνήσιο καρέ 99

Συμπεράσματα - Μελλοντικές επεκτάσεις

Οι μέθοδοι που εξετάσαμε παράγουν βίντεο υψηλής χωρικής και χρονικής ανάλυσης, αυξάνοντας την πυκνότητα πληροφορίας ενός χαμηλής ανάλυσης βίντεο. Οι μέθοδοι αυτοί οι οποίοι αξιοποιούν την χρονική πληροφορία των διανυσμάτων κίνησης, έδειξαν ότι μπορούν να προσεγγίσουν καλύτερα τα επιθυμητά αποτελέσματα του υψηλής ανάλυσης βίντεο, σε σχέση με μεθόδους που χρησιμοποιούν μόνο χωρική πληροφορία, όπως η γραμμική.

Οι αλγόριθμοι στάθηκαν ικανοί στο να αντιμετωπίσουν λάθη που εισάγονται από πιθανή αστοχία των διανυσμάτων κίνησης. Επιπλέον η μέθοδος εκτίμησης της οπτικής ροής που χρησιμοποιήθηκε για την παραγωγή των διανυσμάτων, παρείχε αρκετά ικανοποιητικά αποτελέσματα.

Στην μελέτη αυτή χρησιμοποιήθηκαν ασπρόμαυρες σκηνές (grayscale). Ως μελλοντική επέκταση στις μεθόδους που αναπτύχθηκαν θα μπορούσε να είναι η χρήση του χρώματος στο βίντεο. Το χρώμα στην εικόνα περιέχει μεγάλη πληροφορία που είναι πάρα πολύ χρήσιμη και σημαντική. Η μελέτη και ο συνδυασμός της συμπεριφοράς τους σε κάθε χρωματική συνιστώσα ξεχωριστά θα έδινε καλύτερες προσεγγίσεις στα αποτελέσματα. Επίσης η χρήση του χρώματος της εικόνας θα βοηθούσε και στην διαδικασία εκτίμησης των διανυσμάτων κίνησης, πράγμα που σημαίνει ότι η ακρίβειά τους θα ήταν μεγαλύτερη βοηθώντας έμμεσα τους αλγορίθμους ενίσχυσης της χωροχρονικής ανάλυσης.

Επιπλέον σημαντική βελτίωση στην απόδοση των αλγορίθμων θα αποτελούσε και μία πιο ακριβής διαδικασία για τον υπολογισμό των διανυσμάτων κίνησης, οι οποίοι παίζουν βασικό ρόλο στην εξαγωγή καλύτερων αποτελεσμάτων των αλγορίθμων. Θα μπορούσε να μελετηθεί η συμπεριφορά τους με την χρήση μιας διαφορετικής προσέγγισης για την εκτίμησης της κίνησης όπως των block matching αλγορίθμων.

Σημαντικές ελλείψεις υπάρχουν σε θέματα που αφορούν στην επιρροή του θορύβου που μπορεί να εισαχθεί κατά την μετάδοση και κωδικοποίηση του βίντεο. Οι εφαρμογές αυτές συμβαίνουν σε πραγματικό χρόνο, και η προ-

σομοίωσή τους δεν ήταν δυνατή, έτσι δεν υπήρχε η δυνατότητα μελέτης σχετικά με την αντιμετώπισή του θορύβου.

Επιπλέον σημαντική αδυναμία του επαναληπτικού αλγορίθμου ενίσχυσης της χωρικής ανάλυσης, αποτελεί το μεγάλο υπολογιστικό κόστος. Την αδυναμία αυτή αντιμετωπίζει ο προτεινόμενος αντίστοιχος αλγόριθμος, ο οποίος μειώνει την χρονική καθυστέρηση σε μεγάλο βαθμό εξυπηρετώντας έτσι τις εφαρμογές που απαιτούν ταχύτητα σε πραγματικό χρόνο.

ΒΙΒΛΙΟΓΡΑΦΙΑ

- [1]. F.Pereira,T.Ebrahimi, The MPEG-4 Book
- [2]. A.Smolcic,K.Rummeler,J.Ohm,R.Schafer,S.Bauer ,MPEG-4 Video Transmission over DAB\DMB :Joined Optimization of Encoding and Format Conversion Berlin, 1998.
- [3]. J.Mateos,A.Katsaggelos,R.Molina , High Resolution Image Reconstruction from Compressed Video Sequences.
- [4]. A.Segall,R.Molina,A.Katsaggelos,J.Mateos, Bayesian High Resolution Reconstruction of low Resolution Compressed Video, 2001.
- [5]. B.Tom,A.Katsaggelos, Resolution Enhancement of Monochrome and Color Video Using Motion Compensation, IEEE vol. 10 No. 2 2001.
- [6]. A.Segall,A.Katsaggelos, Enhancement of Compressed Video Using Visual Quality Measurements
- [7]. B. K. Horn and B. G. Schunck, Determining Optical Flow, Artificial Intelligence, Vol. 17, 1981, pp. 185- 204.
- [8]. L.Hendriks,L.J van Vliet, Improving resolution to reduce aliasing in an undersampled image sequence
- [9]. E.Sil Lee and M.Gi Kang ,Regularized Adaptive High-Resolution Image Reconstruction Considering Inaccurate Subpixel Registration IEEE 2003