



ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ  
ΤΜΗΜΑ  
ΜΗΧΑΝΙΚΩΝ Η/Υ ΤΗΛΕΠΙΚΟΙΝΩΝΙΩΝ ΚΑΙ ΔΙΚΤΥΩΝ

## ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

**« Εκτίμηση κίνησης βασισμένη στην ελαχιστοποίηση του μέσου τετραγωνικού σφάλματος στα πεδία του χώρου και της συχνότητας »**

**Κουτσάγια Δέσποινα**  
email : dekoutsa@inf.uth.gr

*Επιβλέποντες καθηγητές : κος. Χούστης Ηλίας  
κα. Μπριασούλη Αλεξία*

**Βόλος, Ιούλιος 2008**



**ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ  
ΒΙΒΛΙΟΘΗΚΗ & ΚΕΝΤΡΟ ΠΛΗΡΟΦΟΡΗΣΗΣ  
ΕΙΔΙΚΗ ΣΥΛΛΟΓΗ «ΓΚΡΙΖΑ ΒΙΒΛΙΟΓΡΑΦΙΑ»**

Αριθ. Εισ.: 6421/1  
Ημερ. Εισ.: 11-07-2008  
Δωρεά: Συγγραφέα  
Ταξιθετικός Κωδικός: ΠΤ - ΜΗΥΤΔ  
2008  
ΚΟΥ

## Ευχαριστίες

Ολοκληρώνοντας την παρούσα διπλωματική εργασία, δεν θα μπορούσα να μην αναγνωρίσω την πολύτιμη βοήθεια που μου προσφέρθηκα και την άψογη συνεργασία που διατηρήθηκε καθ' όλη την διάρκεια εκπόνησης της, και να ευχαριστήσω την κα. Μπριασούλη Αλεξία καθώς και τον κο. Χούστη Ηλία. Επίσης θα ήθελα να ευχαριστήσω την οικογένεια μου και τους φίλους μου για την βοήθεια και την πολύτιμη υποστήριξη τους.

# Περιεχόμενα

<b>1</b>	<b>ΕΙΣΑΓΩΓΗ</b> .....	<b>5</b>
1.1.	Σκοπός της πτυχιακής .....	6
<b>2</b>	<b>Η ΨΗΦΙΑΚΗ ΕΙΚΟΝΑ</b> .....	<b>7</b>
2.1.	Η εικόνα ως δισδιάστατο σήμα .....	7
2.2.	Η επεξεργασία της ψηφιακής εικόνας .....	7
2.2.1.	Επεξεργασία εικόνας στο πεδίο του χώρου .....	9
2.2.2.	Επεξεργασία εικόνας στο πεδίο της συχνότητας .....	9
2.3.	Το φάσμα της εικόνας .....	10
<b>3</b>	<b>ΤΟ ΒΙΝΤΕΟ</b> .....	<b>15</b>
3.1.	Το βίντεο ως σήμα .....	15
3.2.	Χρονική κατάτμηση του βίντεο σε πλάνα .....	16
3.2.1.	Κωδικοποίηση του βίντεο (video coding) .....	17
3.3.	Η επεξεργασία του βίντεο .....	19
<b>4</b>	<b>ΕΚΤΙΜΗΣΗ ΚΙΝΗΣΗΣ</b> .....	<b>21</b>
4.1.	Εκτίμηση κίνησης στο πεδίο του χώρου (motion estimation in spatial domain) .....	22
4.2.	Εκτίμηση κίνησης στο πεδίο της συχνότητας (motion estimation in frequency domain). .....	22
4.3.	Μέθοδοι εκτίμησης της κίνησης .....	23
4.3.1.	Σύγκριση μπλοκ – Block Matching .....	23
4.3.2	Οπτική ροή ( Optical Flow ) .....	26

4.3.3	Τεχνικές στο πεδίο της συχνότητας και Συσχέτιση Φάσης (Frequency-Domain Techniques and phase correlation) .....	32
<b>5</b>	<b>ΕΚΤΙΜΗΣΗ ΚΙΝΗΣΗΣ ΣΤΟ ΠΕΔΙΟ FOURIER ΚΑΙ ΕΚΤΙΜΗΣΗ ΚΙΝΗΣΗΣ ΑΠΟ ΤΗΝ ΦΑΣΗ ΤΟΥ FOURIER .....</b>	<b>36</b>
5.1.	Ελαχιστοποίηση σφάλματος – Error Minimization .....	36
5.2.	Εκτίμηση της κίνησης στο πεδίο του μετασχηματισμού Fourier ( Fourier Transform domain) .....	39
5.3.	Εκτίμηση της κίνησης από την φάση του Fourier .....	44
<b>6</b>	<b>ΠΕΙΡΑΜΑΤΑ .....</b>	<b>47</b>
6.1.	Εκτίμηση κίνησης με ελαχιστοποίηση σφάλματος σε συνθετική και πραγματική ακολουθία .....	47
6.2.	Εκτίμηση κίνησης από την φάση των συνιστωσών Fourier .....	53
<b>7</b>	<b>ΣΥΜΠΕΡΑΣΜΑΤΑ .....</b>	<b>58</b>
<b>8</b>	<b>ΒΙΒΛΙΟΓΡΑΦΙΑ .....</b>	<b>59</b>

# Κεφάλαιο 1

## Εισαγωγή

Η εποχή μας χαρακτηρίζεται από έναν καταγισμό πληροφορίας, που δεν έχει προηγούμενο στην ανθρώπινη ιστορία. Τεράστιος όγκος πληροφοριών παράγεται, αποθηκεύεται και διακινείται εξαιτίας της αλματώδους προόδου που έχει επιτευχθεί στους τομείς της πληροφορικής και των τηλεπικοινωνιών που εκφράστηκε κυρίως με την επανάσταση που προκάλεσε το διαδίκτυο με τον Παγκόσμιο Ιστό. Μέσα στο περιβάλλον αυτό της ανεπανάληπτης διαθεσιμότητας της πληροφορίας, το ζήτημα της πρόσβασης στην πληροφορία αποτελεί προτεραιότητα και αντικείμενο μεγάλης και εντεινόμενης ερευνητικής δραστηριότητας. Αντικειμενικός στόχος αποτελεί η αποτελεσματικότερη ανάκτηση της πληροφορίας από τον ενδιαφερόμενο σε όποια μορφή κι αν βρίσκεται.

Έναν πολύ σημαντικό φορέα πληροφορίας και μέσο επικοινωνίας για τον άνθρωπο αποτελούν οι εικόνες. Η ανάπτυξη των εικόνων, αναφορικά με τον αριθμό, τη διαθεσιμότητα και τη σημασία τους για τον άνθρωπο υπήρξε αλματώδης. Ενδεικτικό της τεράστιας αύξησης των εικόνων είναι το γεγονός ότι κάθε ιστοσελίδα στον Παγκόσμιο Ιστό, που διαρκώς γιγαντώνεται, έχει υπολογιστεί ότι περιέχει ένα μεγάλο αριθμό εικόνων. Η εκρηκτική ανάπτυξη των συλλογών με εικόνες, που υπήρξε αποτέλεσμα της προόδου της τεχνολογίας στον τομέα της δημιουργίας και αποθήκευσης της ψηφιακής εικόνας, έκανε επιτακτική την ανάγκη εύρεσης και εξέλιξης νέων τεχνικών για την διαχείριση και αποτελεσματική ανάκτηση των εικόνων.

Η εντυπωσιακή ανάπτυξη της μικροηλεκτρονικής και των υπολογιστών είχε καθοριστική επίδραση στην ψηφιακή επεξεργασία σημάτων και εικόνων. Οι τεχνικές ψηφιακής επεξεργασίας σημάτων χρησιμοποιούνται σήμερα σε πολλές περιοχές της επιστήμης και της τεχνολογίας, όπως για παράδειγμα τις επικοινωνίες, την αεροναυτική, την σεισμολογία, την βιοϊατρική τεχνολογία.

Ως σήμα ορίζουμε το σύνολο των τιμών που λαμβάνει μια φυσική ποσότητα. Μαθηματικά αυτό εκφράζεται ως συνάρτηση ή ακολουθία μιας ή περισσότερων ανεξάρτητων μεταβλητών. Τα σήματα περιέχουν πληροφορία σχετικά με την συμπεριφορά ή τη φύση ενός φαινομένου.

Τα σήματα χωρίζονται σε κατηγορίες ανάλογα με τον τύπο και το πλήθος των μεταβλητών. Έτσι, ανάλογα με το αν η ανεξάρτητη μεταβλητή είναι συνεχής ή διακριτή, έχουμε σήματα συνεχούς ή διακριτού χρόνου. Ανάλογα με το πλήθος των ανεξάρτητων μεταβλητών έχουμε σήματα μιας μεταβλητής ή διάστασης, δυο μεταβλητών ή διαστάσεων και πολλών μεταβλητών ή διαστάσεων. Παράδειγμα μονοδιάστατου σήματος είναι η ομιλία, όπου η ανεξάρτητη μεταβλητή είναι ο χρόνος και η εξαρτημένη μεταβλητή η ακουστική πίεση. Μια εικόνα (φωτογραφία) αποτελεί χαρακτηριστικό παράδειγμα σήματος δυο διαστάσεων. Εδώ η εξαρτημένη μεταβλητή είναι η φωτεινότητα της εικόνας και οι δυο ανεξάρτητες μεταβλητές είναι οι δυο χωρικές συντεταγμένες. Τέλος, παράδειγμα σήματος τριών διαστάσεων είναι η ακολουθία εικόνων (video), όπου οι δυο ανεξάρτητες μεταβλητές είναι χωρικές και η τρίτη είναι ο χρόνος. Η εξαρτημένη μεταβλητή είναι και εδώ η φωτεινότητα της κάθε εικόνας.

Καθένα από τα παραπάνω σήματα είναι απαραίτητο να το επεξεργαστούμε με στόχο να βελτιώσουμε την ποιότητά του, να εξάγουμε την χρήσιμη πληροφορία και να εξαλείψουμε κατά το δυνατόν τον θόρυβο, να το αποθηκεύουμε ή να το μεταδώσουμε.

Σε κάθε περίπτωση είναι προφανές ότι οι εικόνες έχουν σημαντική θέση στη ζωή του ανθρώπου γιατί με αυτές κινείται, σκέφτεται, ενεργεί επικοινωνεί και εξάγει συμπεράσματα. Στον κλάδο της πληροφορικής ο όγκος οπτικών πληροφοριών είναι τεράστιος και η εξαγωγή συμπερασμάτων σε εφαρμογές απαιτεί ταχύτητα και αξιοπιστία. Τα τελευταία χρόνια παρατηρείται ανάπτυξη αποδοτικών τεχνικών επεξεργασίας και ανάλυσης καθώς και μια σημαντική πρόοδο στις τεχνολογίες των συστημάτων καταγραφής των εικόνων.

### 1.1. Σκοπός της πτυχιακής

Στην παρούσα διπλωματική θα ασχοληθούμε με την ανίχνευση της κίνησης σε βίντεο. Παρουσιάζουμε μία νέα προσέγγιση στην εκτίμηση της κίνησης ελαχιστοποιώντας το μέσο τετραγωνικό σφάλμα (Mean Squared Error) στα πεδία του χώρου (spatial domain) και της συχνότητας (frequency domain) και δείχνουμε ότι η ολική χωρική φύση του Fourier Transform οδηγεί σε ένα σφάλμα εκτίμησης της κίνησης που είναι αρκετά μικρότερο από αυτό που λαμβάνεται μέσω της χωρικής εκτίμησης κινήσεων (spatial estimation). Από την άλλη πλευρά όμως η χωρική ανάλυση είναι χρήσιμη για την ακριβή κατάτμηση. Περιγράφουμε μία νέα υβριδική προσέγγιση που συνδυάζει τις ανωτέρω δυο εκτιμήσεις της κίνησης και της κατάτμησης. Εξετάζουμε την σταθερότητα της ελαχιστοποίησης των όρων σφαλμάτων και στα δυο πεδία θεωρητικά και πειραματικά. Πειράματα με πραγματικές και συνθετικές ακολουθίες καταδεικνύουν τις ικανότητες του προτεινόμενου αλγορίθμου.

Παραδοσιακά, η ανάλυση της κίνησης βασίζεται στην χωρική αναπαράσταση του βίντεο (video) [1], [2]. Σε αυτήν την προσέγγιση συνδυάζουμε πληροφορίες και από το πεδίο του χώρου και από το πεδίο της συχνότητας. Χρησιμοποιώντας τα δεδομένα της συχνότητας αποφεύγουμε τις ανακρίβειες και την ευαισθησία του θορύβου που παρουσιάζουν οι χωρικές μέθοδοι [1]. Με την χωρική πληροφορία επιτυγχάνουμε αξιόπιστη κατάτμηση της εικόνας.

Η υπάρχουσα έρευνα για την ανάλυση της κίνησης μέσω του συνδυασμού των δεδομένων των πεδίων του χώρου και της συχνότητας, βασίζεται στον συσχετισμό φάσης (phase correlation) και στις αντίστροφες τεχνικές απεικόνισης (inverse imaging techniques) [3]. Ακολουθούμε μία διαφορετική προσέγγιση ελαχιστοποιώντας τα σφάλματα στα πεδία της συχνότητας και του χώρου για εκτίμηση της κίνησης και κατάτμηση αντίστοιχα.

Στις ακόλουθες ενότητες θα ακολουθήσει μία αναφορά στην ψηφιακή εικόνα και στην επεξεργασία της, μία ανάλυση του βίντεο και της κίνησης που υπάρχει σε αυτό. Επίσης αναλύεται η εκτίμηση της κίνησης στα πεδία του χώρου και της συχνότητας καθώς και διάφοροι αλγόριθμοι υπολογισμού. Και τέλος, ακολουθεί η εκτίμηση της κίνησης στο πεδίο του Fourier και η εκτίμηση της κίνησης από την φάση του Fourier.

## Κεφάλαιο 2 Η ψηφιακή εικόνα

Η εικόνα αποτελεί στην εποχή μας μία από τις σημαντικότερες πηγές πληροφορίας. Τη συναντούμε ως εικόνα ακίνητη (φωτογραφία) ή κινούμενη (τηλεόραση), ασπρόμαυρη ή έγχρωμη. Η ψηφιακή εικόνα αποτελεί ό,τι πιο σύγχρονο, τόσο στο χώρο της ενημέρωσης (Internet) και της εκπαίδευσης (multimedia) όσο και στο χώρο του θεάματος και της ψυχαγωγίας (ψηφιακή τηλεόραση, DVD κλπ.).

Η ψηφιακή εικόνα είναι ένα από τα δισδιάστατα ψηφιακά σήματα που παρουσιάζουν σημαντικό ενδιαφέρον στη σημερινή κοινωνία της πληροφορικής. Τα στοιχειώδη συστατικά της ψηφιακής εικόνας είναι το εικονοστοιχείο (pixel), το μέγεθος της εικόνας, οι πράξεις που μπορούμε να κάνουμε πάνω σε αυτή καθώς και το πλήθος των υπολογισμών που απαιτούνται.

### 2.1 Η εικόνα ως δισδιάστατο σήμα

Η ψηφιακή εικόνα αναπαρίσταται ως μία δισδιάστατη συνάρτηση φωτεινότητας  $f(x, y)$  όπου  $x, y$  είναι οι συντεταγμένες στο επίπεδο. Η ψηφιακή εικόνα μπορεί να θεωρηθεί σαν ένας πίνακας στοιχείων (picture elements- pixels) . Η συνάρτηση  $f$  μπορεί να θεωρηθεί σαν ένα διάνυσμα με τρεις συνιστώσες όπου η καθεμία αντιστοιχεί σε ένα βασικό χρώμα. Ο συνδυασμός των τριών συνιστωσών δίνει το ακριβές χρώμα ενός σημείου. Για την επιλογή των τριών βασικών χρωμάτων υπάρχουν πολλά χρώματα με πιο διαδεδομένο το RGB ( Red Green Blue ).

Τα σήματά θεωρούνται ως συναρτήσεις μιας μόνο ανεξάρτητης μεταβλητής, του χρόνου. Η ανάγκη για επεξεργασία της πληροφορίας της εικόνας οδηγεί στη μαθηματική περιγραφή του περιεχομένου της. Έτσι, την ασπρόμαυρη εικόνα μπορούμε να τη δούμε ως μία δισδιάστατη συνάρτηση των συντεταγμένων του χώρου  $x$  και  $y$  :  $f \rightarrow f(x, y)$   
Αυτό σημαίνει ότι η  $f$  περιγράφει μία επιφάνεια, η οποία έχει μεγάλη τιμή όπου η εικόνα είναι πιο λευκή ενώ για μικρές τιμές της  $f$  η εικόνα πλησιάζει το μαύρο. Στην ουσία η  $f$  περιγράφει την αμαύρωση της εικόνας σε κάθε θέση  $(x, y)$  .

### 2.2 Η επεξεργασία της ψηφιακής εικόνας

Ένα σύστημα επεξεργασίας εικόνων καταγράφει εικόνες, μία προς μία ή και ακολουθία, επεξεργάζεται τις εικόνες με ψηφιακό τρόπο και εξάγει νέες εικόνες. Η επεξεργασία της ψηφιακής εικόνας περιλαμβάνει στη γενικότερη περίπτωση πολλά και διαφορετικά στάδια ανάλογα με τη μορφή και την αρχική κατάσταση της ψηφιακής εικόνας και ανάλογα με το επιθυμητό αποτέλεσμα.



Η ψηφιακή επεξεργασία εικόνας αποσκοπεί στην βελτίωση της εμφάνισής των εικόνων, στην καλύτερη παρατήρησή τους από τον άνθρωπο, στην προετοιμασία τους για αυτόματη αναγνώριση καθώς και στην μέτρηση των χαρακτηριστικών δομών που υπάρχουν σε αυτές.

Η πραγματική εικόνα καταγράφεται από το σύστημα απόκτησης της ψηφιακής εικόνας (image acquisition system). Θα ακολουθήσει το στάδιο της εξαγωγής του θορύβου (filtering) και της διόρθωσης (αποκατάστασης - restoration) σφαλμάτων που προκλήθηκαν είτε από το μέσο μεταξύ

εικόνας και συστήματος απόκτησης είτε και από το ίδιο το σύστημα απόκτησης. Τα σφάλματα αυτά θα πρέπει να διορθωθούν πριν προχωρήσουμε σε επόμενο στάδιο στην επεξεργασία της εικόνας. Σφάλματα που παρουσιάζονται στην εικόνα κατά τη διάρκεια της απόκτησής της είναι ο θερμικός θόρυβος ή/και ο θόρυβος αιχμών (από το σύστημα απόκτησης) και η θόλωση της εικόνας. Η θόλωση προκαλείται συνήθως από την οπτική ανομοιογένεια του μέσου (αέρας) ή τη σχετική κίνηση της εικόνας ως προς το σύστημα απόκτησης.

Οι αιτίες δημιουργίας σφαλμάτων κατά την απόκτηση της εικόνας έχουν καταγραφεί στη βιβλιογραφία εκτενώς. Καλύτερος τρόπος για τη διόρθωση των σφαλμάτων αυτών είναι να περιγραφεί μαθηματικά η αιτία που τα προκαλεί και να υλοποιηθεί με το σύστημα αποκατάστασης η αντίστροφη διαδικασία. Υπάρχουν φυσικά περιπτώσεις όπου η διαδικασία απόκτησης είναι αρκετά καλή με αποτέλεσμα η εικόνα να μη χρειάζεται αποκατάσταση. Σε άλλες περιπτώσεις το στάδιο της αποκατάστασης είναι το τελευταίο πριν την επίδειξη της εικόνας (τηλεόραση). Η βελτίωση της εικόνας αφορά κυρίως στον τονισμό συγκεκριμένων στοιχείων της εικόνας με σκοπό τη διευκόλυνση των διαδικασιών που θα ακολουθήσουν. Τα επόμενα στάδια εντάσσονται σε ότι καλούμε ανάλυση της εικόνας. Σημαντικά θέματα αποτελούν η εξαγωγή των κατάλληλων χαρακτηριστικών για αναγνώριση και ταξινόμηση αντικειμένων που υπάρχουν μέσα στην εικόνα, η συμπίεση της εικόνας για γρήγορη μετάδοσή της ή οικονομική αποθήκευση, ο μετασχηματισμός της σε άλλους χώρους, όπως αυτός των συχνοτήτων, για την αποκάλυψη πληροφορίας που στον αρχικό χώρο δεν είναι εμφανής, οι γεωμετρικοί μετασχηματισμοί που είναι απαραίτητοι για τη σύγκριση εικόνων, τη σύνθεση της πληροφορίας σκηνών που έχουν ληφθεί από διαφορετική γωνία καθώς και τη σύνθεση εγκάρσιων τομών αντικειμένων από προβολικές εικόνες. Το τελευταίο στάδιο είναι η τεχνητή όραση. Εδώ περιλαμβάνονται όλα τα αυτοματοποιημένα συστήματα επεξεργασίας εικόνας, αναγνώρισης αντικειμένων και λήψης αποφάσεων για συγκεκριμένες ενέργειες (robot).

Οι διαδικασίες ψηφιακής επεξεργασίας εφαρμόζονται στο πεδίο του χώρου (spatial domain) και στο πεδίο της συχνότητας (frequency domain). Η μετάβαση από το πεδίο του χώρου στο πεδίο της συχνότητας υλοποιείται με τη βοήθεια του δισδιάστατου μετασχηματισμού Fourier. Πραγματοποιούνται μετρήσεις σε χαρακτηριστικά της εικόνας όπως η φωτεινότητα, η θέση, το μέγεθος και το σχήμα.

### 2.2.1 Επεξεργασία εικόνας στο πεδίο του χώρου

Ο χώρος είναι το πεδίο όπου η εικόνα είναι αντιληπτή από τον άνθρωπο. Οι αλλαγές στην θέση των εικονοστοιχείων (pixels) συνεπάγονται αλλαγές στο περιεχόμενο της εικόνας και οι αποστάσεις στο πεδίο του χώρου αντιπροσωπεύουν πραγματικές αποστάσεις. Στο πεδίο του χώρου πραγματοποιείται απευθείας επεξεργασία της εικόνας, δηλαδή του πίνακα με τις τιμές φωτεινότητας των pixels που την αποτελούν.

### 2.2.2 Επεξεργασία εικόνας στο πεδίο της συχνότητας

Η θεώρηση των σημάτων και συστημάτων διακριτού χρόνου από το πεδίο της συχνότητας, μας δίνει μια εντελώς διαφορετική δυναμική κατανόησης και επεξεργασίας τους. Στο πεδίο της συχνότητας οι μεταβολές στην θέση της εικόνας έχουν ως αποτέλεσμα μεταβολές στην συχνότητα. Αν επεξεργαστούμε την εικόνα σε αυτό το πεδίο παίρνουμε τον ρυθμό με τον οποίο μεταβάλλονται στο πεδίο της συχνότητας οι τιμές της φωτεινότητας της εικόνας. Για τον σκοπό αυτό, χρειάζεται να μετασχηματίσουμε την εικόνα ώστε να πάρουμε την αναπαράσταση της στο πεδίο της συχνότητας, να την επεξεργαστούμε και να υπολογίσουμε τον αντίστροφο μετασχηματισμό ώστε αν μεταφέρουμε την εικόνα ξανά στο πεδίο του χώρου.

Για τα (μονοδιάστατα) σήματα διακριτού χρόνου η μετάβαση στο πεδίο της συχνότητας επιτυγχάνεται με την βοήθεια του μετασχηματισμού Fourier διακριτού χρόνου, ενώ για εικόνες εφαρμόζεται ο δισδιάστατος μετασχηματισμός Fourier (2D Fourier Transform). Για ψηφιακή επεξεργασία οποιουδήποτε σήματος (μονοδιάστατο, εικόνα, βίντεο), χρησιμοποιείται ο διακριτός μετασχηματισμός Fourier (Discrete Fourier Transform - DFT).

#### *Fourier στην εικόνα*

Ο μετασχηματισμός Fourier δίνεται από την σχέση :

$$F(\omega) = \int_{-\infty}^{\infty} f(x)e^{-j\omega x} dx$$

Ο αντίστροφος μετασχηματισμός Fourier δίνεται από την σχέση :

$$f(x) = \int_{-\infty}^{\infty} F(\omega) e^{j\omega x} d\omega$$

Το φάσμα της συχνότητας ορίζεται δίνεται από την σχέση :

$$F(\omega) = \sqrt{R^2(\omega) + I^2(\omega)}$$

Η φάση ορίζεται ως :

$$\varphi(\omega) = \tan^{-1} \frac{I(\omega)}{R(\omega)}$$

Στην επεξεργασία εικόνας, ο δισδιάστατος μετασχηματισμός Fourier ορίζεται ως εξής :

$$F(\bar{\omega}) = \int_{-\infty}^{\infty} f(\bar{x}) e^{-j\bar{\omega}\bar{x}} d\bar{x} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x_1, x_2) e^{-j(\omega_1 x_1 + \omega_2 x_2)} dx_1 dx_2$$

$$f(\bar{x}) = \int_{-\infty}^{\infty} F(\bar{\omega}) e^{j\bar{\omega}\bar{x}} d\bar{\omega} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} F(\omega_1, \omega_2) e^{j(\omega_1 x_1 + \omega_2 x_2)} d\omega_1 d\omega_2$$

### 2.3 Το φάσμα της εικόνας

Ο μετασχηματισμός ενός σήματος από το πεδίο του χρόνου στο πεδίο των συχνοτήτων προσφέρει σημαντικές δυνατότητες για τη μελέτη του σήματος αφού μπορούμε να το εξετάσουμε από μία άλλη σκοπιά. Τον διακριτό μετασχηματισμό Fourier (DFT) τον υλοποιούμε γρήγορα με τον αλγόριθμο που καλείται FFT. Στα δισδιάστατα σήματα, όπως είναι οι εικόνες, υπάρχει ακριβώς η ίδια δυνατότητα μετασχηματισμού. Φυσικά, αφού η εικόνα αποτελεί σήμα που περιγράφει τον τρόπο μεταβολής της φωτεινότητας (ή του χρώματος) στο χώρο και όχι στο χρόνο, ο διακριτός μετασχηματισμός Fourier θα μας μεταφέρει στην περιοχή των χωρικών συχνοτήτων (διαστάσεις αντιστρόφου χώρου). Κατά τα άλλα ο δισδιάστατος διακριτός μετασχηματισμός Fourier έχει όλες τις ιδιότητες του μονοδιάστατου μετασχηματισμού. Ο δισδιάστατος μετασχηματισμός Fourier  $p(k_1, k_2)$ , μιας εικόνας  $q(n_1, n_2)$  μεγέθους  $N \times N$  εικονοστοιχείων, ορίζεται ως:

$$p(k_1, k_2) = \sum_{n_1=0}^{N-1} \sum_{n_2=0}^{N-1} q(n_1, n_2) W_N^{k_1 n_1} W_N^{k_2 n_2} \quad 0 \leq k_1, k_2 \leq N - 1$$

Ενώ ο αντίστροφος μετασχηματισμός είναι ο εξής :

$$q(k_1, k_2) = \frac{1}{N^2} \sum_{k_1=0}^{N-1} \sum_{k_2=0}^{N-1} p(k_1, k_2) W_N^{-k_1 n_1} W_N^{-k_2 n_2} \quad 0 \leq n_1, n_2 \leq N - 1$$

Από την πρώτη σχέση γίνεται σαφές ότι για μία εικόνα  $q$  με  $N \times N = N^2$  εικονοστοιχεία, δημιουργείται ένα φάσμα  $p$  με  $N \times N = N^2$  όρους, που στη γενική τους περίπτωση είναι μιγαδικοί αριθμοί. Κάθε ένας από τους όρους  $p(k_1, k_2)$  απαιτεί για τον υπολογισμό του  $N^2$  πολλαπλασιασμούς και σχεδόν άλλες τόσες προσθέσεις. Επομένως, για τον υπολογισμό όλων των όρων του φάσματος απαιτείται ένα πλήθος πράξεων της τάξης  $N^4$ . Επειδή ο δισδιάστατος DFT είναι διαχωρίσιμη διαδικασία, το πλήθος των πράξεων είναι της τάξης του  $N^3$ .

$$p(k_1, k_2) = \sum_{n_1=0}^{N-1} W_N^{k_1 n_1} \sum_{n_2=0}^{N-1} q(n_1, n_2) W_N^{k_2 n_2} \quad 0 \leq k_1, k_2 \leq N - 1$$

που σημαίνει ότι μπορούμε να υπολογίσουμε πρώτα το εσωτερικό άθροισμα της σχέσης. Η πράξη αυτή ισοδυναμεί με το να υπολογίσουμε τον DFT κάθε μιας γραμμής της εικόνας και στο αποτέλεσμα που προκύπτει να υπολογίσουμε τον DFT της κάθε στήλης. Η χρησιμοποίηση του FFT θα μειώσει τις απαιτούμενες πράξεις για τον υπολογισμό του φάσματος σε πλήθος της τάξης του  $2N^2 \log 2N$ . Δεδομένου ότι οι όροι  $p(k_1, k_2)$  του φάσματος είναι μιγαδικοί αριθμοί, για να μελετήσουμε το φάσμα μιας εικόνας, συνήθως μελετούμε το μέτρο των χωρικών συνιστωσών και τη φάση τους ξεχωριστά.

Στο παρακάτω σχήμα φαίνεται το μέτρο και η φάση του φάσματος φυσικών εικόνων. Συγκρίνοντας το μέτρο των δύο εικόνων μπορούμε να πούμε ότι παρουσιάζει τις περισσότερες αρμονικές στις χαμηλές χωρικές συχνότητες (κέντρο των εικόνων του μέτρου). Αντίθετα οι εικόνες της φάσης παρουσιάζουν έντονη τυχαιότητα και δεν μπορεί να συγκριθούν.



Εικόνα

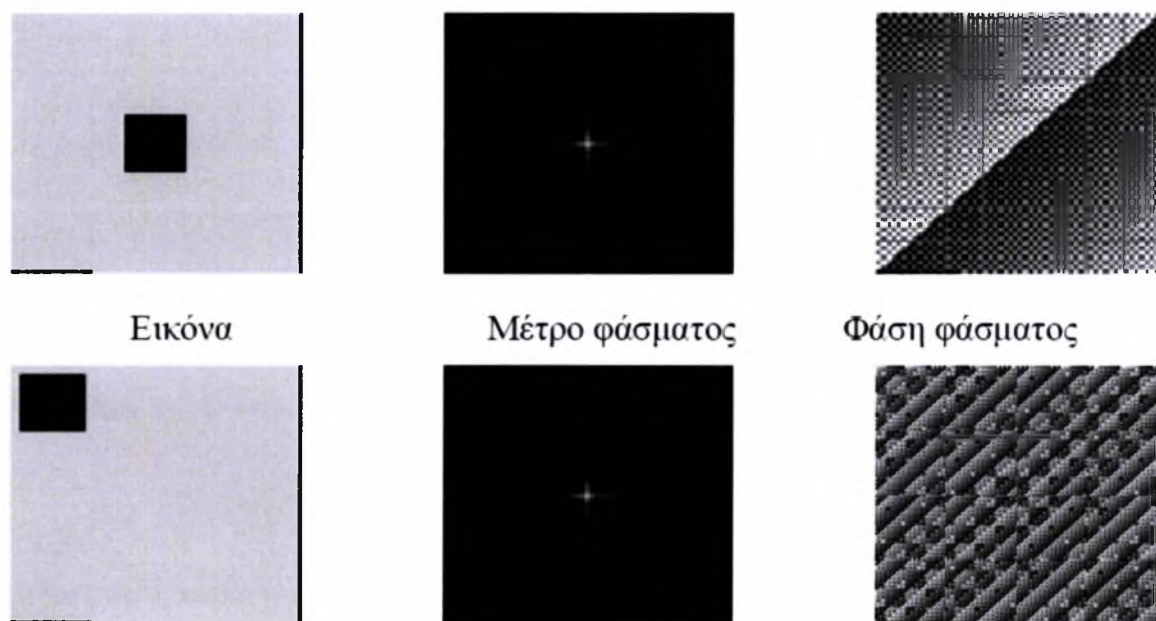
μέτρο φάσματος

φάση φάσματος

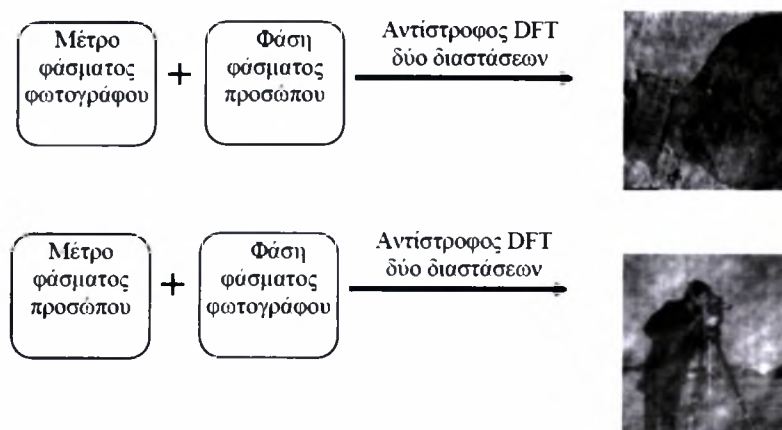
Παραπάνω παρατίθενται το μέτρο και η φάση του φάσματος δυο φυσικών εικόνων.

Τέλος, πρέπει να επισημάνουμε τη σημασία που έχει η φάση για την εικόνα. Στο Σχήμα 1 παρακάτω παρουσιάζονται το μέτρο και η φάση του φάσματος για το ίδιο σχήμα (ένα τετράγωνο) που είναι τοποθετημένο σε δύο διαφορετικές θέσεις της εικόνας. Το μέτρο του φάσματος και για τις δύο εικόνες είναι το ίδιο, ενώ η φάση είναι διαφορετική. Το γεγονός αυτό αποτελεί ένδειξη ότι η φάση περιέχει σημαντική πληροφορία για τη θέση ενός σχήματος μέσα στην εικόνα. Επίσης πρέπει να σημειωθεί ότι η εικόνα κρατά τα γενικά της χαρακτηριστικά αν το μέτρο του φάσματος μεταβληθεί αλλά η φάση διατηρηθεί αναλλοίωτη. Έτσι η ανασύνθεση εικόνας, χρησιμοποιώντας την πληροφορία της φάσης του φάσματός της και μέτρο από φάσμα οποιασδήποτε φυσικής σκηνής, δίνει ικανοποιητική αναπαράσταση της εικόνας όπως φαίνεται στο Σχήμα 2. Συμπερασματικά θα πρέπει να τονίσουμε ότι στην ψηφιακή επεξεργασία εικόνας

με διάφορα γραμμικά φίλτρα, αυτά θα πρέπει να μην επηρεάζουν τη φασική σχέση των χωρικών αρμονικών της εικόνας. Τέτοια φίλτρα είναι τα δισδιάστατα FIR . Τα FIR φίλτρα ως γνωστό παρουσιάζουν γραμμική απόκριση φάσης με αποτέλεσμα να διατηρούν τη φασική σχέση των αρμονικών του σήματος και επομένως τη μορφή του.



**Σχήμα 1.** Η πληροφορία για τη θέση του τετραγώνου βρίσκεται στη φάση του φάσματος της εικόνας αφού αυτή διαφοροποιείται και όχι στο μέτρο που παραμένει το ίδιο.



**Σχήμα 2.** Η φάση του φάσματος μιας εικόνας είναι καθοριστική στην ανασύνθεση της εικόνας. Το μέτρο του φάσματος παίζει δευτερεύοντα ρόλο.

Η μετατροπή των εικόνων στο χώρο φάσματος, είναι χρήσιμη για πληθώρα εφαρμογών ανάλυσης εικόνας (όπως η απαλοιφή θορύβου και το χωρικό φιλτράρισμα). Για τους μετασχηματισμούς Fourier, η μηδενική συχνότητα είναι το κέντρο της μετασχηματισμένης εικόνας και οι σταδιακά αυξανόμενες συχνότητες απεικονίζονται ακτινικά προς τα έξω. Μια εφαρμογή του μετασχηματισμού συχνότητας είναι η εφαρμογή φίλτρων στον χώρο της συχνότητας για ενίσχυση εικόνας, ανακατασκευή εικόνας και άλλες διαδικασίες. Η βασική αρχή των διαδικασιών φιλτραρίσματος είναι να ελαττωθούν ή να εξαλειφθούν κάποιες συχνότητες ενώ ταυτόχρονα να διατηρηθούν κάποιες άλλες. Όταν εφαρμοστεί ο αντίστροφος μετασχηματισμός στην φιλτραρισμένη εικόνα, το αποτέλεσμα είναι η αρχική εικόνα χωρίς ένα μέρος των συχνοτήτων.

Εφόσον μία εικόνα έχει μετασχηματιστεί στο χώρο συχνοτήτων, μέσω γρήγορου μετασχηματισμού Fourier, ορισμένα πρότυπα θορύβου στην αρχική εικόνα πλέον εμφανίζονται ως ισχυρές ενεργειακές συγκεντρώσεις στην εικόνα συχνότητας. Με το μαρκάρισμα αυτών των συνιστωσών, επιτυγχάνεται το φιλτράρισμα των προτύπων του θορύβου εφόσον η εικόνα μετασχηματιστεί πάλι στον φασματικό χώρο με τον ανάστροφο μετασχηματισμό Fourier. Η θεωρία Fourier δηλώνει ότι οποιοδήποτε σήμα στις οπτικές εικόνες μπορεί να εκφραστεί ως μία ημιτονοειδής σειρά. Παραδείγματος χάριν η εικόνα με ημιτονοειδές πρότυπο που παρουσιάζεται κατωτέρω μπορεί να αναλυθεί σε έναν ενιαίο όρο Fourier που κωδικοποιεί τη χωρική συχνότητα, το πλάτος (θετικό ή αρνητικό) και τη φάση. Το μέγεθος του πλάτους (amplitude) αντιστοιχεί στην αντίθεση (contrast) ή την διαφορά μεταξύ των σκοτεινότερων και φωτεινότερων τόνων της εικόνας. Ένα αρνητικό μέγεθος αντιπροσωπεύει μια αντίθεση-αντιστροφή. Η φάση αντιπροσωπεύει πως οι ημιτονοειδείς συνιστώσες του σήματος μετατοπίζονται σε σχέση με την αρχική τους θέση. Δηλαδή πόσο η ημιτονοειδής συνιστώσα μετατοπίζεται αριστερά ή δεξιά. Ένας μετασχηματισμός Fourier κωδικοποιεί όχι μόνο μία ημιτονοειδή συνιστώσα αλλά μια ολόκληρη σειρά μέσω ενός φάσματος χωρικών συχνοτήτων από το μηδέν μέχρι την συχνότητα Nyquist.

Κάθε pixel της εικόνας Fourier είναι μια τιμή χωρικής συχνότητας της οποίας το πλάτος κωδικοποιείται από την αντίθεση δυο γειτονικών pixels. Όσο φωτεινότερες είναι οι κηλίδες στην εικόνα Fourier τόσο υψηλότερη η αντίθεση στην εικόνα φωτεινότητας.

Ο μετασχηματισμός Fourier έχει δυο σημαντικές λειτουργίες στην ψηφιακή επεξεργασία εικόνας. Η πρώτη είναι ότι ο δισδιάστατος μετασχηματισμός Fourier είναι η γέφυρα μεταξύ της χωρικής περιοχής και της περιοχής συχνότητας. Η δεύτερη είναι ότι ο δισδιάστατος μετασχηματισμός Fourier συνδέεται πολύ με το θεώρημα συνελίξεων και δειγματοληψίας. Το πρώτο είναι η βάση της επεξεργασίας εικόνας και το δεύτερο είναι η γέφυρα μεταξύ του συνεχούς (αληθινού) κόσμου και του διακριτού κόσμου.

Η χωρική περιοχή και η περιοχή συχνότητας είναι ακριβώς όπως δυο γλώσσες εφαρμοσμένης μηχανικής, οι οποίες μπορούν και οι δυο να περιγράψουν το ίδιο φαινόμενο. Επομένως, ο μετασχηματισμός κατά Fourier είναι πολύ χρήσιμος όχι μόνο περιεκτικά να καταλάβει (την ψηφιακή) ανάλυση εικόνας αλλά και (στην ψηφιακή) επεξεργασία εικόνας.

Εκτός από το δισδιάστατο μετασχηματισμό Fourier, στην επεξεργασία της ψηφιακής εικόνας χρησιμοποιούνται και πλήθος άλλων μετασχηματισμών με ενδιαφέρουσες ιδιότητες. Ο σημαντικότερος εξ' αυτών είναι ο δισδιάστατος μετασχηματισμός συνημίτονου (Discrete Cosine Transform - DCT). Οι πολύ καλές ιδιότητες του DCT τον καθιστούν πολύ χρήσιμο σε

διαδικασίες συμπίεσης εικόνας. Αυτό οφείλεται στο γεγονός ότι παρουσιάζει συγκέντρωση της ενέργειας του σήματος σε λίγους μόνο συντελεστές στον μετασχηματισμένο χώρο. Ο DCT μπορεί να υλοποιηθεί ταχύτατα μέσω του FFT, ή και απευθείας με τεχνικές αποδεκατισμού (decimation) ανάλογες εκείνων του FFT .

## Κεφάλαιο 3 Το βίντεο

### 3.1. Το βίντεο ως σήμα

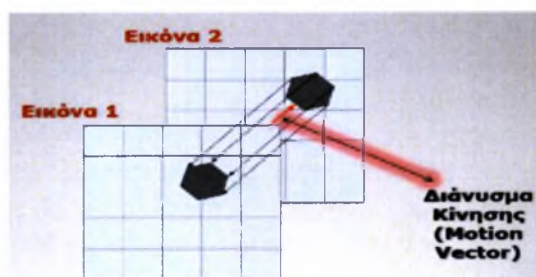
Τα βίντεο όπως γνωρίζουμε είναι μια σειρά από εικόνες που διαδέχονται η μια την άλλη. Η διαδοχή αυτή γίνεται αρκετά γρήγορα και με σταθερό ρυθμό δίνοντας έτσι στο θεατή τους την ψευδαίσθηση της κινούμενης εικόνας. Το σύνολο των εικόνων από τις οποίες αποτελείται ένα βίντεο ονομάζονται frames ή καρέ. Με μια αντίστοιχη θεώρηση με αυτή που έγινε για τις εικόνες μπορούν να οριστούν και τα βίντεο ως σήματα τριών διαστάσεων. Απλά προστίθεται στη συνάρτηση μιας εικόνας  $f(x, y)$  μια ακόμα παράμετρος, ο χρόνος  $t$ . Ας θεωρήσουμε λοιπόν ότι έχουμε το βίντεο  $F(x, y, t)$ . Η τιμή της  $F$  είναι η φωτεινότητα του εικονοστοιχείου με χωρικές συντεταγμένες  $(x, y)$  στο καρέ  $t$  του βίντεο. Τα  $x$  και  $y$  παίρνουν τιμές από 0 έως το πλάτος και ύψος αντίστοιχα του κάθε καρέ του βίντεο, ενώ ο χρόνος  $t$  παίρνει διακριτές τιμές από 0, που αποτελεί το πρώτο καρέ, έως  $T$  που αποτελεί το τελικό. Αν η παράμετρος  $t$  είναι σταθερή και ίση με  $t_0$ , αναφερόμαστε σε ένα συγκεκριμένο καρέ. Έτσι το σήμα  $F(x, y, t_0)$  μπορούμε να θεωρήσουμε ότι αποτελεί ένα δισδιάστατο σήμα  $G_0(x, y)$ . Αυτό αναπαριστά το συγκεκριμένο καρέ με αριθμό  $t_0$ , δηλαδή μία στατική εικόνα. Έτσι μπορούμε να εφαρμόσουμε πάνω στο  $G_0$  κάθε γνωστό αλγόριθμο από την επεξεργασία εικόνας. Αυτή είναι γενικά από μαθηματική πλευρά η διαδικασία με την οποία προχωράμε στην ψηφιακή επεξεργασία του βίντεο βασιζόμενοι σε τεχνικές της ψηφιακής επεξεργασίας εικόνας [23].

Όλα τα σήματα που συναντώνται στη φύση είναι αναλογικά, δηλαδή συνεχή. Προκειμένου όμως να επεξεργαστούμε αναλογικά σήματα με ψηφιακά μέσα, όπως ένας υπολογιστής, θα πρέπει προηγουμένως να τα ψηφιοποιήσουμε. Το ίδιο ισχύει για εικόνες και βίντεο. Αν επεξεργαζόμαστε εικόνες αυτό γίνεται σε δυο βήματα. Το πρώτο είναι η δειγματοληψία και το δεύτερο η κβάντιση. Αν έχουμε βίντεο απαιτείται να καθοριστεί μια ακόμα παράμετρος, ο αριθμός των καρέ ανά δευτερόλεπτο.

Κατά την ψηφιοποίηση ενός βίντεο πέρα από την ανάλυση και το βάθος χρώματος που έχουμε, υπάρχει ακόμα μια πολύ σημαντική παράμετρος που πρέπει να υπολογίζεται. Η τρίτη παράμετρος είναι ο αριθμός των καρέ ανά δευτερόλεπτο (FPS - Frames Per Sec) που χρησιμοποιούνται για την ψηφιοποίηση του βίντεο. Η τιμή αυτής της παραμέτρου συνήθως κυμαίνεται ανάμεσα στο 25 και 30 FPS. Τιμές μικρότερες από τα 25 FPS ρίχνουν πολύ την ποιότητα του βίντεο αφού οι εναλλαγές των εικόνων γίνονται αντιληπτές από το ανθρώπινο μάτι. Αυτό έχει ως αποτέλεσμα μια ομαλή κίνηση να φαίνεται στο βίντεο σπασμωδική. Αντίστοιχα τιμές μεγαλύτερες από τα 30 FPS δεν γίνονται αντιληπτές από το ανθρώπινο μάτι. Δηλαδή όχι μόνο δεν ανεβάζουν την ποιότητα του βίντεο αλλά αντίθετα μεταφέρουν και μεγάλη πλεονάζουσα πληροφορία που δεν γίνεται αντιληπτή. Για παράδειγμα στο σύστημα τηλεόρασης NTSC έχουμε 30 FPS ενώ στο PAL έχουμε 25 FPS. Βέβαια σε ειδικές περιπτώσεις εφαρμογών μπορούν να απαιτούνται και τιμές μεγαλύτερες από τα 30 FPS. Γενικά η επιλογή της κατάλληλης τιμής του αριθμού των καρέ ανά δευτερόλεπτο εξαρτάται αποκλειστικά από την εφαρμογή.



Σε ένα βίντεο είναι πιθανή η ύπαρξη κίνησης. Αυτό που μας ενδιαφέρει είναι να υπολογίσουμε την κίνηση (motion estimation) σε βίντεο. Ο υπολογισμός της κίνησης εξετάζει την κίνηση των αντικειμένων σε μία ακολουθία εικόνων ώστε να ληφθεί το διάνυσμα που αντιπροσωπεύει την κίνηση.



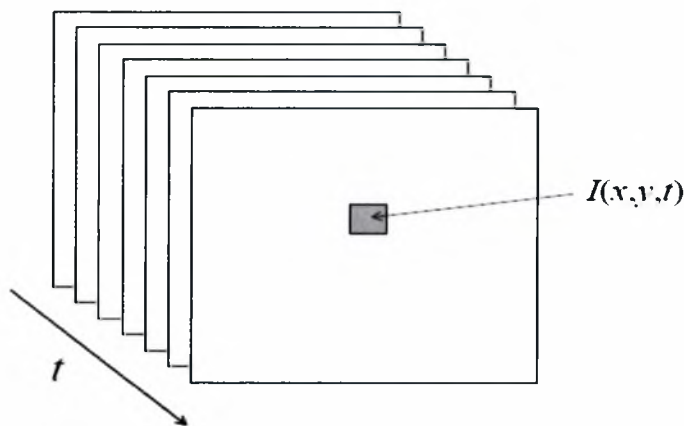
### 3.2 Χρονική κατάτμηση του βίντεο σε πλάνα

Το πλάνο ορίζεται ως η δομική μονάδα του βίντεο που αποτελείται από διαδοχικές κινούμενες εικόνες που παρουσιάζουν χρονική συνέχεια. Το πλάνο είναι η καταγεγραμμένη ακολουθία από συνεχόμενα καρέ που προκύπτουν από μία συγκεκριμένη κάμερα με εγγραφή χωρίς διακοπή. Στην αγγλική βιβλιογραφία χρησιμοποιείται περισσότερο ο όρος “shot” και μερικές φορές ο όρος “scene”. Πολλά συνεχόμενα πλάνα που συσχετίζονται μεταξύ τους, εστιάζοντας σε κάποιο συγκεκριμένο αντικείμενο ή γενικότερα σε κάποιο κοινό θέμα ενδιαφέροντος αποτελούν μία σκηνή (story), δηλαδή μία μικρή αυτοτελή ιστορία. Οι σκηνές αυτές μπορούν να ταξινομηθούν σε διάφορα επίπεδα σημαντικότητας και περιεχομένου και έτσι να υλοποιηθεί ένα πολυεπίπεδο μοντέλο για μια ιεραρχική δομή του βίντεο. Μέσα σε ένα βίντεο οι διάφορες γειτονικά πλάνα, που μπορεί να αφορούν την ίδια ή και διαφορετική σκηνή, συνδέονται μεταξύ τους με κάποιους συγκεκριμένους τρόπους μετάβασης, που ονομάζονται μεταβάσεις πλάνου (shot transitions) [22].

Επομένως η χρονική κατάτμηση του βίντεο σε πλάνα αφορά στην ανίχνευση των μεταβάσεων μεταξύ των πλάνων, την ταξινόμηση τους σε μια από τις προηγούμενες κατηγορίες και τον εντοπισμό των ορίων των πλάνων (shot boundaries). Ο πιο διαδεδομένος τρόπος αποθήκευσης και κωδικοποίησης κινούμενων εικόνων είναι η κωδικοποίηση MPEG, πρότυπο το οποίο βασίζεται στην κίνηση των blocks από κάθε πλαίσιο στο επόμενο. Αν εκμεταλλευτούμε την πληροφορία που υπάρχει αποθηκευμένη στο MPEG, όπως είναι τα διανύσματα κίνησης, ο τύπος και οι ιδιότητες των Macroblocks, ή το σφάλμα κωδικοποίησης, η ανίχνευση των πλάνων μπορεί να γίνει απευθείας στο κωδικοποιημένο πεδίο του MPEG προτύπου, εξοικονομώντας χρόνο αλλά επιτυγχάνοντας και καλύτερα αποτελέσματα. Δεδομένου ότι πρόκειται για εφαρμογές πάνω σε μεγάλες βάσεις δεδομένων από βίντεο μεγάλης διάρκειας και δεδομένου ότι η ανάλυση των πλάνων αποτελεί μόνο το πρώτο βήμα στην ανάλυση και την περιγραφή του

περιεχομένου του βίντεο, η ταχύτητα των υπολογισμών αποτελεί βασικό παράγοντα της ποιότητας της μεθόδου.

Το παρακάτω σχήμα δείχνει μία ακολουθία εικόνων στον χρόνο.



### 3.2.1. Κωδικοποίηση του βίντεο (video coding)

Το βίντεο είναι ένα σήμα του οποίου η οργάνωση μπορεί να θεωρηθεί ότι εκτείνεται προς τρεις διαστάσεις, **οριζόντια** όπου το σήμα μεταφέρει την πληροφορία φωτεινότητας και χρώματος σαρώνει μία οριζόντια γραμμή στην οθόνη, **κάθετα** όπου το σήμα οργανώνει σε οριζόντιες γραμμές σάρωσης τη μία κάτω από την άλλη και **χρονικά** όπου το σήμα οργανώνει σε διαδοχικά πλαίσια (frames) που δημιουργούν την αίσθηση της κίνησης.

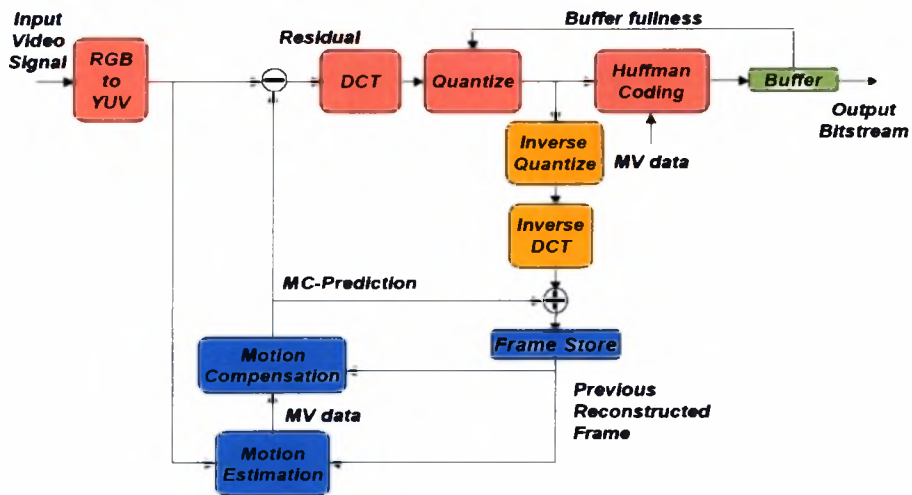
Στην κωδικοποίηση ενός βίντεο για συμπίεση, η βασική ιδέα είναι να εξερευνήσουμε τα πλεονάζουσα δεδομένα, δηλαδή την περιττή πληροφορία. Υπάρχουν δυο τύποι πλεονασμού σε μία κινούμενη εικόνα :

- *Χωρικός Πλεονασμός (Spatial Redundancy)*, όπου εκμεταλλευόμαστε την ομοιότητα μεταξύ των πιο γειτονικών εικονοστοιχείων.
- *Χρονικός Πλεονασμός (Temporal Redundancy)*, όπου εκμεταλλευόμαστε την ομοιότητα ανάμεσα στα επιτυχημένα πλαίσια.

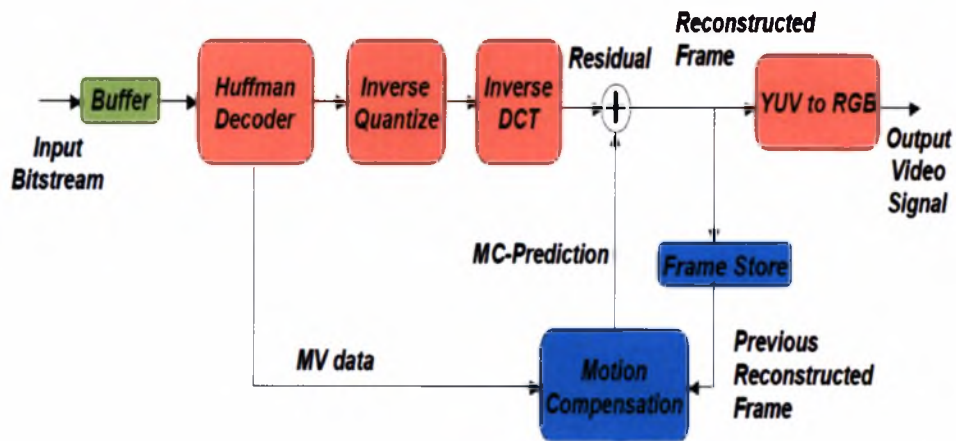
Ο λόγος που προκαλείται ο χρονικός πλεονασμός είναι η κίνηση που έχουν τα εικονοστοιχεία από πλαίσιο σε πλαίσιο σε μία κινούμενη εικόνα. Η κίνηση μπορεί να είναι είτε εστίαση (zoom), είτε περιστροφή (rotation), είτε μετάθεση (translation). Δυο στάδια επεξεργασία ακολουθούνται, επεξεργασία για μείωση του χρονικού πλεονασμού και επεξεργασία για μείωση του χωρικού πλεονασμού.

Παρακάτω παρατίθενται η κωδικοποίηση και η αποκωδικοποίηση ενός βίντεο.

## Video encoder



## Video decoder



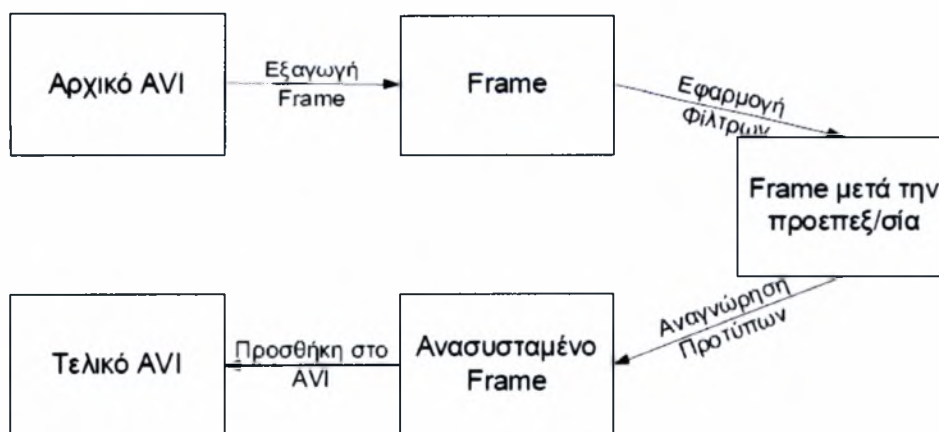
### 3.3 Η επεξεργασία του βίντεο

Τα τελευταία χρόνια ο διαρκώς αυξανόμενος όγκος διαθέσιμου περιεχομένου βίντεο έχει οδηγήσει στην ανάγκη για την αυτόματη ανάλυση, σύνοψη και εξαγωγή πληροφορίας από το βίντεο [26], [27]. Κάθε ακολουθία κινούμενων εικόνων αποτελείται από ένα σύνολο πλάνων, που περιέχουν καρέ με χρονική συνέχεια, ενώ τα γειτονικά πλάνα συνδέονται στα όριά τους με κάποιο είδος μετάβασης. Η ανίχνευση των ορίων αυτών και η χρονική κατάτμηση της ακολουθίας αποτελούν συνήθως το πρώτο βήμα για κάθε είδους ανάλυση του βίντεο, ενώ το επόμενο και πιο σημαντικό βήμα αφορά στη σύνοψη ή δημιουργία περίληψης από το βίντεο. Η τελευταία επιτυγχάνεται συνήθως με την επιλογή ενός αριθμού χαρακτηριστικών καρέ από κάθε πλάνο, ούτως ώστε να αναπαρασταθεί το περιεχόμενό του με σύντομο και ταυτόχρονα περιεκτικό τρόπο.

Η επεξεργασία ενός βίντεο έχει να κάνει με την εφαρμογή αλγορίθμων αναγνώρισης προτύπων προκειμένου να αναγνωριστούν οι διάφορες μορφές και τα αντικείμενα στα βίντεο που επεξεργαζόμαστε. Για να επιτευχθεί αυτό πρέπει να ακολουθηθεί μια σειρά από ενέργειες.

Το πρώτο βήμα που πρέπει να γίνει είναι ο χωρισμός του βίντεο στα καρέ από τα οποία αποτελείται. Στη συνέχεια γίνεται η επεξεργασία κάθε ενός από αυτά ξεχωριστά σαν να ήταν μια ανεξάρτητη εικόνα. Πριν από την αναγνώριση προηγείται το στάδιο της προ-επεξεργασίας των εικόνων. Σε αυτό εφαρμόζονται μια σειρά από φίλτρα και μετασχηματισμοί στα καρέ του βίντεο προκειμένου να αποκτήσουν την επιθυμητή μορφή και χρήσιμες ιδιότητες. Για παράδειγμα απομακρύνεται ο θόρυβος, αναδεικνύονται οι ακμές, τα περιγράμματα, οι υφές και γενικότερα τα χαρακτηριστικά που μας ενδιαφέρουν. Στη συνέχεια ακολουθεί το στάδιο της ανάλυσης της εικόνας. Συγκεκριμένα εφαρμόζονται αλγόριθμοι αναγνώρισης προτύπων που βασίζονται στην αναγνώριση υφής. Τα αποτελέσματα που παίρνουμε από το στάδιο της αναγνώρισης προτύπων είναι μια σειρά από εικόνες. Έτσι το τελικό στάδιο της επεξεργασίας μας είναι η δημιουργία ενός νέου βίντεο τα καρέ του οποίου αποτελούν οι εικόνες που έχουν δημιουργηθεί.

Στο παρακάτω σχήμα φαίνεται σχηματικά η παραπάνω διαδικασία η οποία ακολουθείται για κάθε καρέ του βίντεο.



Οι τεχνικές της ψηφιακής επεξεργασίας εικόνας χωρίζονται, ανάλογα με το πεδίο στο οποίο εφαρμόζονται, σε δυο βασικές κατηγορίες. Έτσι υπάρχουν αυτές που εφαρμόζονται στο χωρικό πεδίο (spatial domain), δηλαδή απευθείας στα εικονοστοιχεία της εικόνας, και αυτές που εφαρμόζονται στο πεδίο των συχνοτήτων (frequency domain). Το πεδίο των συχνοτήτων μας δίνει τις «εναλλαγές» στις τιμές της φωτεινότητας στα εικονοστοιχεία μιας εικόνας αντί για τις ίδιες τις τιμές τους. Για τις τεχνικές στο πεδίο της συχνότητας απαιτείται μετατροπή της εικόνας από το πεδίο του χώρου στο πεδίο των συχνοτήτων. Αυτό γίνεται με την χρήση του δισδιάστατου διακριτού μετασχηματισμού Fourier (2D Discrete Fourier Transform). Μετά την εφαρμογή των φίλτρων στο πεδίο της συχνότητας απαιτείται ο αντίστροφος διακριτός μετασχηματισμός Fourier για την απεικόνιση του αποτελέσματος.

## Κεφάλαιο 4      Εκτίμηση κίνησης

Γενικά ο όρος κίνηση χρησιμοποιείται για να περιγράψει το συνεκτικό συστατικό των κινήσεων των διαφορετικών συστατικών μερών ενός αντικειμένου, από ένα παραμετροποιημένο μοντέλο κίνησης. Η διαδικασία υπολογισμού αυτών των παραμέτρων είναι γνωστή ως εκτίμηση κίνησης (motion estimation). Συνήθως, η κίνηση αναφέρεται στην προφανή κίνηση του φόντου (background), συμπεριλαμβανομένου αυτής που προκαλείται από την κάμερα (camera) [24], [25]. Η εκτίμηση της κίνησης χρησιμοποιείται σε πολλές εφαρμογές όπως στην συμπίεση βίντεο (video compression), στην κατάτμηση (segmentation), στην εγγραφή εικόνας (image registration), στην σταθεροποίηση κάμερας (camera stabilization) και σε πολλές άλλες. Η πιθανότητα ύπαρξης μεγάλων κινήσεων, διαφορετικά κινούμενων αντικειμένων στο χώρο προβολής, και της εμφάνισης και εξαφάνισης περιοχών εικόνας κάνουν το πρόβλημα της εκτίμησης πολύ δύσκολο ειδικά κάτω από τους περιορισμούς των περιορισμένων υπολογιστικών χώρων.

Πριν αναλυθούν με λεπτομέρειες οι τεχνικές εκτίμησης κίνησης, θα αναλύσουμε θεωρητικά την ιδέα της κίνησης στο πλαίσιο της επεξεργασίας της ακολουθίας εικόνων [17]. Οι διατυπώσεις από την άποψη είτε της στιγμιαίας ταχύτητας είτε της μετατόπισης είναι δυνατές. Λόγω της διακριτής φύσης της ακολουθίας των εικόνων κατά την διάρκεια του χρόνου, η στιγμιαία ταχύτητα  $u$  ενός εικονοστοιχείου και η μετατόπιση του  $d$ , σχετίζονται με έναν σταθερό παράγοντα  $\Delta t$  που ανταποκρίνεται στο χρονικό διάστημα δειγματοληψίας. Σε αυτήν την περίπτωση, αυτές οι δυο ποσότητες είναι εναλλάξιμες, οπότε και οι δυο διατυπώσεις είναι ισότιμες. Όμως η τελευταία διατύπωση δεν είναι πλέον διαθέσιμη όταν η κίνηση μιας ομάδας από εικονοστοιχεία μοντελοποιείται από ένα σύνολο παραμέτρων κίνησης. Σε αυτήν την περίπτωση, οι διατυπώσεις της στιγμιαίας ταχύτητας και της μετατόπισης ίσως να οδηγούν σε διαφορετικά μοντέλα (distinct models). Διατύπωση από την άποψη της μετακίνησης υιοθετείται, και ο όρος διάνυσμα κίνησης πρέπει να γίνεται ανάλογα κατανοητός.

Σε μία ψηφιακή ακολουθία εικόνων, η 4D χωρο-χρονική συνέχεια προβάλλεται σε ένα 3D διακριτό πλέγμα δειγμάτων. Μία διάκριση γίνεται ανάμεσα σε δυο οντότητες, στο πεδίο της 2D κίνησης (2D motion field) και της οπτικής ροής (optical flow field) τα οποία αναλύονται παρακάτω. Στην ιδανική περίπτωση το πεδίο της ροής αντιστοιχεί στο δισδιάστατο πεδίο της κίνησης. Όμως στην πράξη αυτό δεν εγγυάται.

Στην κωδικοποίηση του βίντεο, οι τεχνικές εκτίμησης της κίνησης υπολογίζουν την τροχιά (trajectory) των εικονοστοιχείων στις διαδοχικές εικόνες προκειμένου να εκφραστεί η ένταση (intensity) της τρέχουσας εικόνας από προηγούμενη πληροφορία. Επομένως υπολογίζεται η οπτική ροή.

#### **4.1. Εκτίμηση κίνησης στο πεδίο του χώρου ( motion estimation in spatial domain )**

Οι περισσότερες κοινές μέθοδοι στο πεδίο του χώρου (spatial domain) περιλαμβάνουν εκείνες που βασίζονται στην ελαχιστοποίηση του SSD (sum of squared difference) ή της μέτρησης σφάλματος SAD (sum of absolute difference). Η ελαχιστοποίηση του SSD ολοκληρώνεται με μεθόδους καθόδου κλίσης (gradient descent methods) όπως η Marquardt-Levenburg. Δεδομένου ότι αυτή η τεχνική είναι αναδρομική, υπολογιστικά είναι πολύ αργή. Η ελαχιστοποίηση της μέτρησης σφάλματος SAD ολοκληρώνεται με άμεση αναζήτηση (direct search) του χώρου παραμέτρων. Όμως η πολυπλοκότητα της αναζήτησης αυξάνεται εκθετικά με τον αριθμό των παραμέτρων. Αυτές οι τεχνικές υποφέρουν από το μειονέκτημα ότι μπορεί να κολλήσουν σε τοπικά ελάχιστα, αν και είναι λιγότερο πιθανό πολλαπλά αναλυόμενα πλαίσια (multi resolution framework).

Οι μέθοδοι που βασίζονται στα κοινά γνωρίσματα (feature based methods) στην εξαγωγή (extraction) και στην καταδίωξη (tracking) των χαρακτηριστικών γνωρισμάτων. Όμως η εξαγωγή των αξιόπιστων χαρακτηριστικών γνωρισμάτων στην παρουσία των κλειστών συνενώσεων (occlusion junctions), και στον χειρισμό της εμφάνισης και εξαφάνισης γνωρισμάτων είναι πολύ δύσκολη. Μια πολύ σχετική κατηγορία μεθόδων χρησιμοποιεί διανύσματα φραγμού κίνησης (block motion vectors) αντί των συντεταγμένων των χαρακτηριστικών σημείων, που είναι πολύ κατάλληλα για MPEG-2 και MPEG-4 κωδικοποίηση αφού τα διανύσματα κίνησης υπολογίζονται ήδη. Η τάξη και η ακρίβεια αυτών των μεθόδων περιορίζεται από την τάξη και την ακρίβεια των διανυσμάτων φραγμού κίνησης. Αν αυτά τα διανύσματα δεν είναι διαθέσιμα, το υπολογιστικό κόστος της εύρεσης τους για μια λογική τάξη διανυσμάτων κίνησης και για ακρίβεια εικονοστοιχείου μπορεί να είναι απαγορευτικό.

Ακολουθεί μια αναλυτική περιγραφή δυο μεθόδων για την εκτίμηση της κίνησης στο πεδίο του χώρου, της μεθόδου της Σύγκρισης Μπλοκ (Block Matching method) και της μεθόδου της Οπτικής Ροής (Optical Flow method).

#### **4.2. Εκτίμηση κίνησης στο πεδίο της συχνότητας ( motion estimation in frequency domain )**

Η συσχέτιση φάσης είναι πολύ δημοφιλής και αποδοτική μέθοδος για την εκτίμηση της κίνησης. Η λήψη δείγματος του φάσματος του μεγέθους του Fourier σε λογαριθμικό πλέγμα έχει εισαχθεί για να υπολογιστεί η περιστροφή (rotation) χρησιμοποιώντας συσχέτιση φάσης (phase correlation) και έχει χρησιμοποιηθεί στην εγγραφή εικόνας (image registration) και στην εκτίμηση της σφαιρικής κίνησης (global motion estimation).

Η εκτίμηση των σχετικών παραμέτρων στο πεδίο συχνότητας βασίζεται στο Σχετικό Θεώρημα του Μετασχηματισμού Φουριέ (Affine Theorem of Fourier Transform). Η ιδιότητα της μετατόπισης Fourier χρησιμοποιείται για να επιτύχει σταθερότητα μετάθεσης (translation invariance) λαμβάνοντας το μέγεθος του φάσματος Fourier των εικόνων.

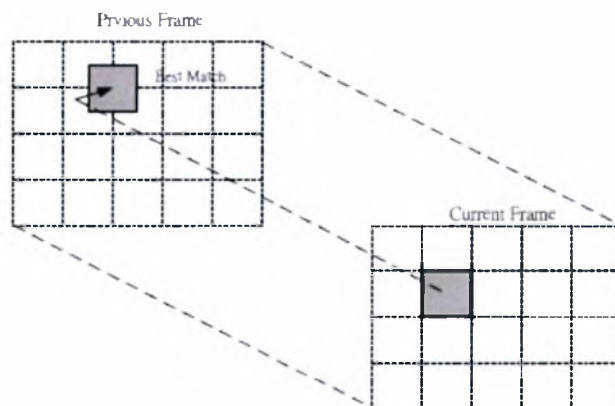
### 4.3. Μέθοδοι εκτίμησης της κίνησης

Τα τελευταία χρόνια έχουν αναπτυχθεί πολλές μέθοδοι με τις οποίες μπορούμε να διεξάγουμε την εκτίμηση της κίνησης (motion estimation) [6]. Οι μέθοδοι χωρίζονται σε τρεις γενικές κατηγορίες :

- Σύγκριση μπλοκ ( Block Matching )
- Οπτική Ροή ( Optical flow )
- Τεχνικές στο πεδίο της συχνότητας και Συσχέτιση Φάσης ( Frequency-Domain Techniques and phase correlation)

Ακολουθεί μια αναλυτική περιγραφή των κατηγοριών εκτίμησης της κίνησης.

#### 4.3.1. Σύγκριση μπλοκ – Block Matching



Όπως αναφέραμε προηγουμένως, οι αλγόριθμοι σύγκρισης μπλοκ βασίζονται στο ταίριασμα των μπλοκ (blocks) ανάμεσα σε δυο εικόνες και ο σκοπός τους είναι η ελαχιστοποίηση ενός μέτρου ομοιότητας [14], [15]. Πιο συγκεκριμένα, υπολογίζονται τα διανύσματα κίνησης ελαχιστοποιώντας μερικές συναρτήσεις κόστους. Αυτοί οι αλγόριθμοι βρίσκουν την καλύτερη αντιστοιχία μεταξύ του τρέχοντος μπλοκ (block) και των υποψηφίων στο προηγούμενο πλαίσιο (frame). Τυπικά κάθε μακρομπλοκ ( $16 \times 16$  pixels) στο νέο πλαίσιο (frame) συγκρίνεται με τις μετατοπισμένες περιοχές ίδιου μεγέθους του προηγούμενου



αποκωδικοποιημένου πλαισίου, και η μετατόπιση που οδηγεί στο ελάχιστο σφάλμα επιλέγεται ως το καλύτερο διάνυσμα κίνησης για αυτό το μακρομπλοκ.

Η εικόνα χωρίζεται σε μπλοκ και τα ίδια διανύσματα μετατόπισης ορίζονται σε όλα τα εικονοστοιχεία (pixels) μέσα σε ένα μπλοκ. Το μοντέλο της κίνησης συχνά θεωρεί ότι μία εικόνα αποτελείται από άκαμπτα αντικείμενα, όπου κάθε στοιχείο τους έχει την ίδια κίνηση. Αν και αυτό το μοντέλο είναι καθαρά περιοριστικό, δικαιολογείται από το γεγονός ότι η σύνθετη κίνηση μπορεί να αναλυθεί σε ένα άθροισμα μετατοπισμένων συνιστωσών. Ο σκοπός της εκτίμησης της κίνησης είναι να βρεθεί το διάνυσμα μετατόπισης. Για να προσδιορίσουμε την μετατόπιση ενός συγκεκριμένου εικονοστοιχείου (pixel)  $p$  στο πλαίσιο την χρονική στιγμή  $t$ , αναφερόμαστε σε ένα μπλοκ από εικονοστοιχεία κεντραρισμένα στο  $p$ . Το πλαίσιο την χρονική στιγμή  $t + dt$  ελέγχεται για το καλύτερο μπλοκ αντιστοίχισης ίδιου μεγέθους. Στην διαδικασία αντιστοίχισης υποτίθεται ότι τα εικονοστοιχεία μετακινούνται ομαλά. Το ταίριασμα γίνεται είτε μεγιστοποιώντας την συνάρτηση διαγώνιου συσχετισμού (cross correlation function) είτε ελαχιστοποιώντας ένα κριτήριο σφάλματος. Το πιο κοινά χρησιμοποιούμενο κριτήριο σφάλματος είναι το μέσο τετραγωνικό σφάλμα (mean squared error) και η ελάχιστη απόλυτη διαφορά (minimum absolute difference).

- **Το μέσο τετραγωνικό σφάλμα (mean squared error)**

$$MSE(d_1, d_2) = \frac{1}{n^2} \sum_{(i,j) \in B} [I_1(i, j) - I_2(i + d_1, j + d_2)]^2$$

όπου  $n$  είναι η διάσταση του θεωρούμενου τετραγωνικού μπλοκ  $B$ . Το διάνυσμα μετατόπισης  $d = [d_1 \ d_2]^T$  υπολογίζεται με την ελαχιστοποίηση του κριτηρίου.

- **Η μέση απόλυτη διαφορά (mean absolute difference)**

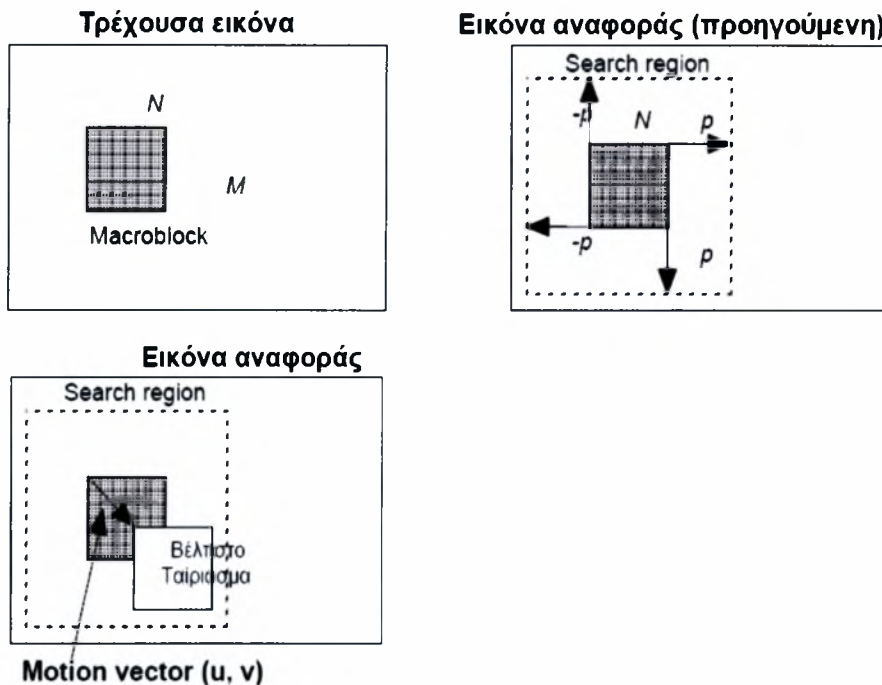
$$MAD(d_1, d_2) = \frac{1}{n^2} \sum_{(i,j) \in B} |I_1(i, j) - I_2(i + d_1, j + d_2)|$$

Το βασικό μοντέλο της εκτίμησης κίνησης είναι η πρόβλεψη του επόμενου καρέ με βάση το προηγούμενο.

$$I(x, y, t) = I(x, y, t - 1) + e(x, y, t)$$

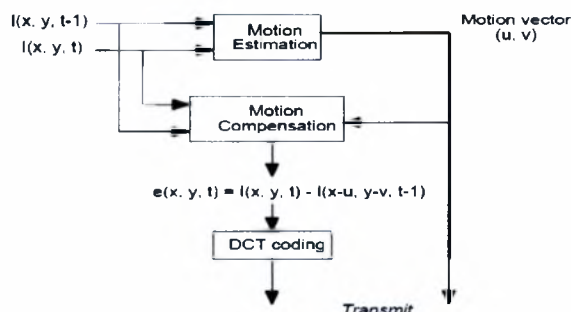
Δεν εντοπίζει σύνθετες κινήσεις και η πρόβλεψη της κίνησης γίνεται με την βοήθεια του διανύσματος κίνησης  $(u, v)$ . Δηλαδή έχουμε την εξίσωση :

$$I(x, y, t) = I(x + u, y + v, t - 1) + e(x, y, t)$$

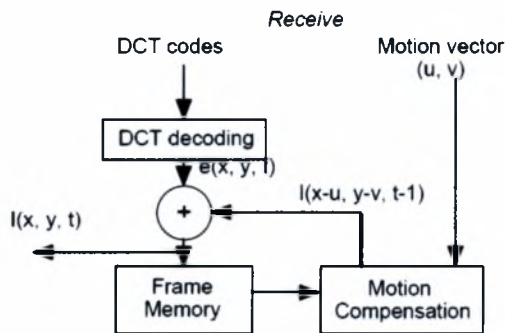


Ένα από τα ζητήματα στην εκτίμηση κίνησης αποτελεί το μέγεθος της περιοχής αναζήτησης. Η τεχνική της σύγκρισης μπλοκ (block matching) μπορεί να είναι υπολογιστικά πολύ απαιτητική αν αναλύονται όλες οι μετατοπίσεις σε κάθε μακρομπλοκ. Για παράδειγμα, προκειμένου να αναλύσουμε τις μετατοπίσεις σε  $\pm 15$  εικονοστοιχεία (pixels) στις οριζόντιες και κάθετες κατευθύνσεις απαιτούνται  $31 \times 31 = 961$  μετατοπίσεις, κάθε μία από τις οποίες περιλαμβάνει  $16 \times 16 = 256$  διαφορετικούς υπολογισμούς ένα μακρομπλοκ. Αυτό είναι γνωστό ως εξαντλητική αναζήτηση σύγκρισης μπλοκ (exhaustive search block matching).

Παρακάτω παρουσιάζεται η κωδικοποίηση και η αποκωδικοποίηση της εκτίμησης κίνησης.



**Σχήμα 1. Κωδικοποίηση**



Σχήμα 2. Αποκωδικοποίηση

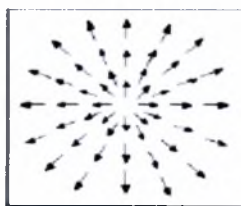
### 4.3.2 Οπτική ροή ( Optical Flow )

Η οπτική ροή χρησιμοποιείται για να υπολογίσει την κίνηση των εικονοστοιχείων (pixels) μιας ακολουθίας εικόνων. Παρέχει μία πυκνή (σημείο προς σημείο) αντιστοίχιση.

#### Δύο σημαντικοί ορισμοί

- **Motion field** : η δισδιάστατη προβολή μιας τρισδιάστατης κίνησης στο επίπεδο της εικόνας

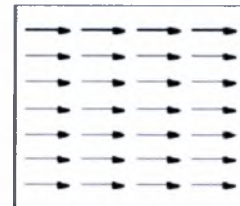
#### *Παραδείγματα Πεδίων Κίνησης – Motion fields*



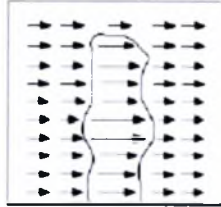
Κίνηση προς τα εμπρός



περιστροφή

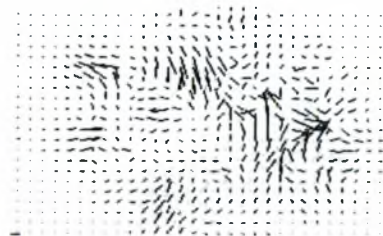
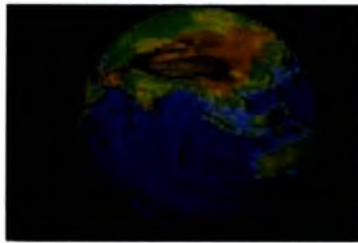


οριζόντια μετάθεση



τα γειτονικά αντικείμενα κινούνται πιο γρήγορα

- **Optical Flow Field** : η προφανής κίνηση των φωτεινών σχηματισμών σε μία ακολουθία εικόνων. Είναι η προβολή του πεδίου κίνησης στην δισδιάστατη εικόνα με την οποία μπορούμε να υπολογίσουμε το δισδιάστατο πεδίο κίνησης.



### Ορισμός οπτικής ροής

Η προβολή της κίνησης ενός σημείου πάνω στο επίπεδο προβολής είναι ένα καθαρά θεωρητικό μέγεθος μιας και δεν μπορεί να μετρηθεί. Στη θέση της χρησιμοποιείται συνήθως αυτό που ονομάζουμε οπτική ροή . Η οπτική ροή είναι η φαινομενική ταχύτητα της κίνησης των φωτεινών σχηματισμών πάνω στην εικόνα, αυτό δηλαδή που βλέπουμε. Η οπτική ροή δεν συμπίπτει πάντα με την πραγματική δισδιάστατη ροή. Για παράδειγμα όταν η απεικόνιση ενός αντικειμένου δεν αλλάζει καθόλου καθώς περιστρέφεται γύρω από τον εαυτό του, το αποτέλεσμα είναι το πεδίο της οπτικής ροής να είναι παντού μηδενικό.

Η οπτική ροή είναι το πεδίο μετατόπισης για κάθε ένα pixel σε μία ακολουθία εικόνων. Σε κάθε pixel αντιστοιχεί το διάνυσμα της ταχύτητας ( $\frac{dx}{dt}, \frac{dy}{dt}$ ) το οποίο δείχνει πόσο γρήγορα ένα pixel κινείται διαμέσου της εικόνας καθώς και την κατεύθυνση της κίνησης. Έχουμε καλύτερη εκτίμηση από την επανόρθωση της κίνησης γιατί χρησιμοποιούμε ένα διάνυσμα για κάθε pixel αντί ένα διάνυσμα για κάθε block από pixels.

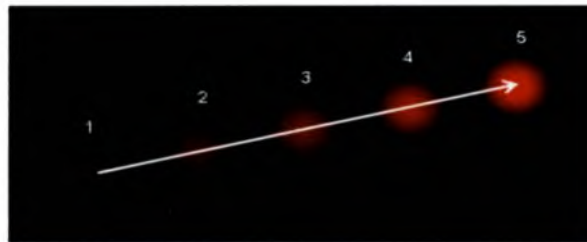
Η οπτική ροή είναι μια προσέγγιση της προφανούς κίνησης των αντικειμένων μέσα σε μία εικόνα [19]. Οι αλγόριθμοι που χρησιμοποιούνται για τον υπολογισμό της οπτικής ροής τείνουν να βρίσκουν συσχετισμούς ανάμεσα σε γειτονικά-κοντινά πλαίσια (frames) σε ένα βίντεο, παράγοντας ένα διανυσματικό πεδίο που δείχνει κατά πόσο μετακινήθηκε κάθε pixel ή περιοχή της αρχικής εικόνας στην δεύτερη εικόνα. Χαρακτηριστικά η κίνηση αντιπροσωπεύεται ως διανύσματα που προέρχονται ή καταλήγουν σε pixels σε μία ψηφιακή ακολουθία εικόνας.

Ο όρος οπτική ροή αναφέρεται στην αποσπασματική κίνηση του pixel που καθορίζεται διαμέσου του επιπέδου της εικόνας, αλλά ο όρος συχνά (ανακριβώς/λανθασμένα) επίσης χρησιμοποιείται για να δείξει την διαδικασία του υπολογισμού της οπτικής ροής από τα στοιχεία της εικόνας, ή ακόμα και μια ιδιαίτερη κατηγορία αλγορίθμων που χρησιμοποιούνται για αυτόν τον σκοπό [16].

Ο υπολογισμός της οπτικής ροής είναι χρήσιμος στην αναγνώριση σχεδίων (pattern recognition), computer vision, και άλλες εφαρμογές επεξεργασίας εικόνας. Συσχετίζεται πολύ με την εκτίμηση κίνησης και την επανόρθωση κίνησης. Συχνά ο όρος οπτική ροή χρησιμοποιείται για να περιγράψει ένα πυκνό πεδίο κίνησης με διανύσματα σε κάθε pixel, σε αντίθεση με την εκτίμηση κίνησης ή την επανόρθωση, που χρησιμοποιούν διανύσματα για blocks από pixels, όπως στις μεθόδους συμπίεσης βίντεο όπως τα MPEG.

Ένα βασικό πρόβλημα στην επεξεργασία ακολουθιών με εικόνες είναι ο υπολογισμός της οπτικής ροής [7]. Η οπτική ροή συχνά είναι μία χρήσιμη αναπαράσταση της κίνησης της εικόνας.

### Υπολογισμός οπτικής ροής



**Σχήμα 5.** Το διάνυσμα οπτικής ροής ενός κινούμενου αντικειμένου σε μία ακολουθία βίντεο

Οι ακολουθίες χρονικά προσδιοριζόμενων εικόνων επιτρέπουν τον υπολογισμό της προβολής της διδιάστατης κίνησης της εικόνας είτε ως στιγμιαίες ταχύτητες της εικόνας είτε ως διακριτές μετατοπίσεις της εικόνας. Αυτό αναφέρεται ως πεδίο της οπτικής ροής. Ο υπολογισμός της οπτικής ροής βελτιστοποιείται όταν αναφέρεται σε ένα εικονικό σύστημα [15], [18].

Για να υπολογίσουμε την οπτική ροή χρησιμοποιούμε δυο παραδοχές.

#### 1<sup>η</sup> παραδοχή :

Η γενική παραδοχή σχεδόν όλων των αλγορίθμων οπτικής ροής που χρησιμοποιείται για τον υπολογισμό της είναι ότι η φωτεινότητα των κινούμενων αντικειμένων, ή η ένταση της εικόνας  $I$ , δεν αλλάζει αλλά παραμένει σταθερή στον χρόνο.

## 2<sup>η</sup> παραδοχή :

Η ένταση της εικόνας είναι συνεχής και διαφοροποιήσιμη.

Έστω ότι  $(x, y)$  είναι ένα σημείο σε μια εικόνα που αντιστοιχεί στην χρονική στιγμή  $t$  και έστω ότι  $I(x, y, t)$  είναι η έντασή του. Έστω ότι το σημείο  $(x, y)$  στην αμέσως επόμενη εικόνα, τη χρονική στιγμή  $t + dt$ , έχει μετακινηθεί κατά  $(dx, dy)$ . Η θέση του θα είναι  $(x + dx, y + dy)$  και η έντασή του  $I(x + dx, y + dy, t + dt)$ . Αν  $(u, v)$  είναι η οπτική ροή στο σημείο  $(x, y)$  τη χρονική στιγμή  $t$  τότε θα ισχύει για τη μεταβολή του σημείου  $(x, y)$  τη χρονική στιγμή  $t$  :  $(dx, dy) = (udt, vdt)$   
Δηλαδή

$$u = \frac{dx}{dt} \quad v = \frac{dy}{dt}$$

Θεωρώντας ότι η ένταση στο σημείο  $(x, y)$  τη χρονική στιγμή  $t$  θα είναι ίδια με την ένταση στο σημείο  $(x + dx, y + dy)$  τη χρονική στιγμή  $t + dt$ , καταλήγουμε στην σχέση:

$$I(x, y, t) = I(x + dx, y + dy, t + dt)$$



$$I(x, y, t) = I(x + dx, y + dy, t + dt)$$

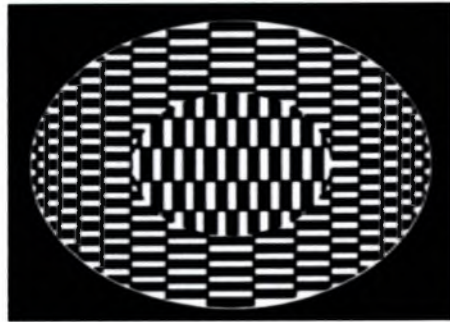
Με την 1<sup>η</sup>ς τάξης επέκταση κατά Taylor παίρνουμε :

$$I(x, y, t) = I(x, y, t) + \frac{dI}{dx} dx + \frac{dI}{dy} dy + \frac{dI}{dt} dt$$

Δηλαδή  $I_x dx + I_y dy + I_t dt = 0$

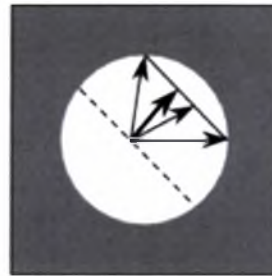
Επομένως καταλήγουμε στην ακόλουθη εξίσωση :

$$I_x u + I_y v + I_t = 0$$





(a)



(b)

- (a) Χαρακτηριστικό γνώρισμα γραμμών που παρατηρείται μέσω ενός μικρού ανοίγματος στο χρόνο  $t$
- (b) Στο χρόνο  $t + dt$  το χαρακτηριστικό γνώρισμα έχει μετακινηθεί προς μια νέα θέση. Δεν είναι δυνατό να καθοριστεί ακριβώς η θέση που έχει μετακινηθεί κάθε σημείο. Από τις τοπικές μετρήσεις εικόνας μόνο η συνιστώσα της ροής που είναι κάθετη στην χαρακτηριστική γραμμή μπορεί να υπολογιστεί.

### Αλγόριθμοι υπολογισμού οπτικής ροής

Χωρίς αμφιβολία, ένα θεμελιώδες πρόβλημα στην επεξεργασία των ακολουθιών εικόνας είναι η μέτρηση της οπτικής ροής ή της ταχύτητας της εικόνας. Ο στόχος είναι να υπολογιστεί μια προσέγγιση στο δισδιάστατο πεδίο της κίνησης, μια προβολή των τρισδιάστατων ταχυτήτων των σημείων επιφάνειας πάνω στην επιφάνεια απεικόνισης, από τα χωροχρονικά σχέδια της εικόνας.

Η οπτική ροή είναι η κατανομή των φαινομενικών ταχυτήτων των εντάσεων μιας εικόνας. Ο υπολογισμός της οπτικής ροής είναι σημαντικός για την χαμηλού επιπέδου ανάλυση της κίνησης και χρησιμοποιείται για πολλές εφαρμογές στην ψηφιακή επεξεργασία εικόνας.

Για περισσότερο από δυο δεκαετίες έρευνας έχουν αναπτυχθεί πάρα πολλοί αλγόριθμοι υπολογισμού της οπτικής ροής. Οι περισσότερες τεχνικές δουλεύουν μόνο με δυο επακόλουθα πλαίσια μιας ακολουθίας εικόνων. Χαρακτηριστικές είναι οι διαφορικές μέθοδοι (differential methods), οι μέθοδοι ταιριάσματος (matching methods), οι μέθοδοι βασισμένοι στην ενέργεια (energy-based methods) και οι μέθοδοι βασισμένοι στην φάση (phase-based methods).

Παρά τις διαφορές τους, οι τεχνικές ακολουθούν τρία βασικά στάδια επεξεργασίας. Το πρώτο είναι το προφιλτράρισμα όπου χρησιμοποιούνται χαμηλής διέλευσης φίλτρα προκειμένου να εξαχθεί η δομή ενδιαφέροντος των σημάτων έτσι ώστε να ενισχυθεί. Το δεύτερο στάδιο αφορά την εξαγωγή βασικών μετρήσεων, όπως τα χωροχρονικά παράγωγα ή οι τοπικές επιφάνειες συσχετισμού, για να μετρηθούν τα κανονικά χαρακτηριστικά της ταχύτητας. Στο τελευταίο στάδιο γίνεται η ολοκλήρωση αυτών των μετρήσεων για να παραχθεί ένα δισδιάστατο πεδίο ροής, το οποίο συχνά περιλαμβάνει υποθέσεις για την ομαλότητα του ελλοχεύοντος πεδίου ροής.

Οι πιο γνωστοί μέθοδοι υπολογισμού οπτικής ροής είναι η μέθοδος Lucas Kanade και η μέθοδος Horn Schunck.



### 4.3.3. Τεχνικές στο πεδίο της συχνότητας και Συσχέτιση Φάσης (Frequency-Domain Techniques and phase correlation)

Για την εκτίμηση κίνησης έχει αναπτυχθεί και άλλη κατηγορία τεχνικών, που επεξεργάζονται τα σήματα στο πεδίο της συχνότητας. Συνήθως βασίζονται στην αρχή του κυκλικού συσχετισμού και προσφέρουν καλά τεκμηριωμένα πλεονεκτήματα από την άποψη της υπολογιστικής αποδοτικότητας λόγω της απασχόλησης των γρήγορων αλγορίθμων. Η πιο γνωστή μέθοδος αυτής της κατηγορίας είναι η συσχέτιση φάσης (phase correlation) που έχει γίνει μια από τις μεθόδους εκτίμησης κίνησης που επιλέγεται για ένα ευρύ φάσμα μεταδιδόμενων εφαρμογών [9], [13]. Εκτός από την υπολογιστική αποδοτικότητα, η συσχέτιση φάσης προσφέρει σημαντικά πλεονεκτήματα από την άποψη της ισχυρής απόκρισης στις άκρες και των εξεχόντων χαρακτηριστικών της εικόνας, της ανοσίας του στις αλλαγές της φωτεινότητας και των κινούμενων σκιών, και της ικανότητας του να μετράει τις μεγάλες μετατοπίσεις.

Οι τεχνικές αυτές βασίζονται στην σχέση μεταξύ των μετασχηματισμένων συντελεστών των μετατοπισμένων εικόνων και δεν χρησιμοποιούνται ευρέως για την κωδικοποίηση των ακολουθιών εικόνας. Η εκτίμηση της κίνησης γίνεται παίρνοντας τον μετασχηματισμό του μπλοκ πρώτα στο πεδίο της συχνότητας.

Οι τεχνικές στο πεδίο της συχνότητας υπολογίζουν την κίνηση απευθείας από τον χάρτη της συσχέτισης της φάσης ο οποίος μας δίνει έναν πιο σταθερό υπολογισμό του διανύσματος της κίνησης και ένα πεδίο κίνησης με πολύ λιγότερη εντροπία. Αυτές οι τεχνικές είναι υπολογιστικά πού αποδοτικές και αποδίδει καλύτερα σε σχέση με την μέθοδο Σύγκρισης μπλοκ (block matching), που έχει την ίδια πολυπλοκότητα στις περισσότερες περιπτώσεις ειδικά στην μετάφραση και στην μεγάλης κλίμακας κίνηση.

Σε αντίθεση με την μέθοδο Σύγκρισης μπλοκ (block matching), που ψάχνει τα μπλοκ από τις αντιστοιχίες της φωτεινότητας, οι τεχνικές στο πεδίο της συχνότητας και η συσχέτιση φάσης υπολογίζουν την μετακίνηση μεταξύ των δυο πεδίων απευθείας από τις φάσεις τους. Στην επεξεργασία εικόνας η συσχέτιση φάσης (cross correlation) είναι μια γρήγορη προσέγγιση στο πεδίο της συχνότητας για την εκτίμηση της σχετικής μεταθετικής μετακίνησης (translating movement) ανάμεσα σε δυο εικόνες.

#### Μέθοδος συσχέτισης φάσης - Phase-correlation method

Αρχικά παρουσιάζουμε την σχέση μεταξύ των δισδιάστατων μετασχηματισμών Fourier (2D Fourier Transform) δύο διαδοχικών πλαισίων που παρουσιάζουν μια μεταθετική κίνηση (translational motion).

Υποθέτουμε μια μεταθετική μετατόπιση μεταξύ των δυο πλαισίων  $k$  και  $k + 1$  κεντραρισμένη στο εικονοστοιχείο (pixel)  $\bar{n} = [n_1, n_2]$ .

$$s(n_1, n_2, k) = s(n_1 + d_1, n_2 + d_2, k + 1)$$

Ο 2D μετασχηματισμός Fourier είναι :  $S_k(f_1, f_2) = S_{k+1}(f_1, f_2)e^{j2\pi(d_1f_1+d_2f_2)}$

Επομένως η μετατόπιση στο πεδίο του χώρου απεικονίζεται ως η αλλαγή της φάσης στο πεδίο του φάσματος. Ο διαγώνιος συσχετισμός (cross correlation) ανάμεσα στα δυο πλαίσια είναι :

$$c_{k,k+1}(n_1, n_2) = s(n_1, n_2, k + 1)^{**} s(-n_1, -n_2, k)$$

Ο μετασχηματισμός Fourier του δίνεται από την εξίσωση :

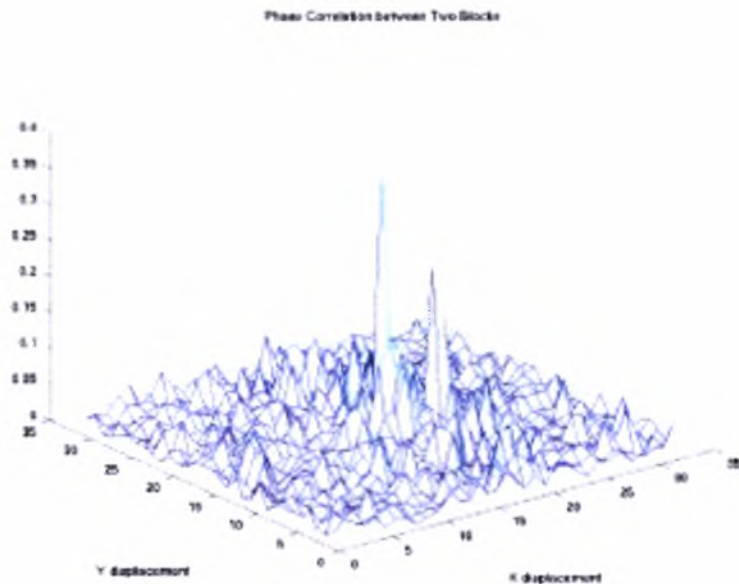
$$C_{k,k+1}(f_1 + f_2) = S_{k+1}(f_1 + f_2)S_k^*(f_1, f_2)$$

Προκειμένου να αφαιρέσουμε την επιρροή της αλλαγής της φωτεινότητας κατά την διάρκεια της ανάλυσης της φάσης, ομαλοποιούμε το φάσμα της διαγώνιας δύναμης (cross-power spectrum) από το μέγεθος του, και λαμβάνουμε την φάση του :

$$\Phi[C_{k,k+1}(f_1, f_2)] = \frac{S_{k+1}(f_1, f_2)S_k^*(f_1, f_2)}{|S_{k+1}(f_1, f_2)S_k^*(f_1, f_2)|}$$

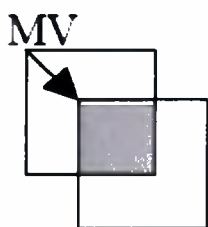
Συνδυάζοντας τις παραπάνω εξισώσεις, παίρνουμε :  $\Phi[C_{k,k+1}(f_1, f_2)] = e^{-j2\pi(d_1f_1+d_2f_2)}$  όπου ο δισδιάστατος αντίστροφος μετασχηματισμός δίνεται από την εξίσωση :  $c_{k,k+1}(n_1, n_2) = \delta(n_1 - d_1, n_2 - d_2)$

Κατά συνέπεια, βρίσκοντας την θέση του παλμού της παραπάνω εξίσωσης, είμαστε σε θέση να ανακαλύψουμε την μετατόπιση, η οποία είναι το διάνυσμα της κίνησης. Στην πράξη η κίνηση δεν είναι καθαρά μεταφραστική (translational), για αυτό παίρνουμε τον συσχετισμό της φάσης παρόμοιο με αυτόν που απεικονίζεται στο παρακάτω σχήμα. Σε αυτήν την περίπτωση βρίσκουμε τον παλμό, βρίσκοντας το υψηλότερο μέγιστο ή μερικά υποψήφια.



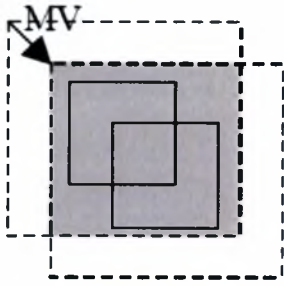
**Σχήμα 3.** *Συσχετισμός φάσης μεταξύ δυο μπλοκ*

Στην πράξη, το τρέχον πλαίσιο διαιρείται σε μπλοκ  $16 \times 16$  και ο υπολογισμός της συσχέτισης της φάσης εκτελείται για κάθε μπλοκ. Προκειμένου να υπολογιστεί σωστά ο διαγώνιος συσχετισμός (cross correlation) των αντίστοιχων μπλοκ στα αντίστοιχα πλαίσια, επεκτείνουμε τα μπλοκ στο μέγεθος  $32 \times 32$ , κεντραρισμένα γύρω από τα καθορισμένα μπλοκ των  $16 \times 16$  για να υπολογίσουμε τον συσχετισμό φάσης. Αν το κάνουμε αυτό μόνο για τα μπλοκ των  $16 \times 16$ , οι συσχετισμοί τους μπορεί να είναι πολύ χαμηλοί για συγκεκριμένη κίνηση λόγω της μικρής επικαλύπτουσας περιοχής όπως φαίνεται στο παρακάτω σχήμα.



**Σχήμα 4.** *Περιοχή συσχέτισης χρησιμοποιώντας το μπλοκ των  $16 \times 16$*

Μόλις επεκταθεί το μέγεθος των μπλοκ σε  $32 \times 32$ , η περιοχή επικάλυψης αυξάνεται για καλύτερη εκτίμηση του συσχετισμού όπως φαίνεται στο παρακάτω σχήμα.



**Σχήμα 5. Περιοχή συσχέτισης χρησιμοποιώντας το μπλοκ των  $32 \times 32$**

Προφανώς θα υπάρξει επικάλυψη μεταξύ των εκτεταμένων μπλοκ και ένα εικονοστοιχείο συνήθως υπάρχει σε πολλαπλά μπλοκ αντικειμένων.

## Κεφάλαιο 5 Εκτίμηση κίνησης σε πεδίο Fourier και εκτίμηση κίνησης από την φάση του Fourier

Όπως αναφέραμε και στον πρόλογο, σκοπός αυτής της πτυχιακής είναι η εκτίμηση της κίνησης σε βίντεο με χρήση του μετασχηματισμού Fourier και συγκεκριμένα της φάσης του. Μέχρι στιγμής οι υπάρχουσες τεχνικές για την ανάλυση της κίνησης βασίζονται στην συσχέτιση φάσης (phase correlation) και στις αντίστροφες τεχνικές απεικόνισης (inverse image techniques). Σε αντίθεση με τις υπάρχουσες μεθόδους και τις έρευνες, η μέθοδος μας ακολουθεί μια διαφορετική προσέγγιση ελαχιστοποιώντας τα σφάλματα στο πεδίο του Fourier (Fourier domain) και του χώρου (spatial domain) αντίστοιχα. Κι αυτό γιατί μπορεί να δειχθεί ότι το σφάλμα στο πεδίο του Fourier (Fourier error) είναι πιο σταθερό στον πρόσθετο θόρυβο από το χωρικό σφάλμα (spatial error).

Αρχικά, σε αυτήν την ενότητα θα παρουσιάσουμε την ελαχιστοποίηση του σφάλματος (error minimization) στα πεδία του χώρου και της συχνότητας και θα δείξουμε ότι με την προσθήκη θορύβου το σφάλμα στο πεδίο της συχνότητας είναι πιο σταθερό. Στη συνέχεια, θα περιγράψουμε αναλυτικά την εκτίμηση της κίνησης σε βίντεο στο πεδίο του Fourier καθώς και την εκτίμηση της χρησιμοποιώντας την φάση των συνιστωσών Fourier.

Όπως γνωρίζουμε, η κίνηση στο βίντεο δεν είναι μια αυθαίρετη αλληλουχία των εικόνων, αλλά μια ακολουθία εικόνων που μεταφέρουν μια συνεπή ερμηνεία της φυσικής σκηνής. Η καλύτερη μέθοδος υπολογισμού της κίνησης σε βίντεο μπορεί να είναι η εξαντλητική (full-search) μέσω συσχέτισης, όμως είναι υπολογιστικά πολύ απαιτητική και χρονοβόρα. Για αυτό και έχουμε καταφύγει στην έρευνα και εφαρμογή πιο γρήγορων και λιγότερο χρονοβόρων τεχνικών.

### 5.1. Ελαχιστοποίηση σφάλματος – Error Minimization

Αρχικά παρουσιάζουμε την ελαχιστοποίηση του χωρικού τετραγωνικού σφάλματος (spatial squared error) με  $M$  αντικείμενα  $s_i(\vec{r})$  να κινούνται σε μαύρο φόντο. Αν το πρώτο πλαίσιο του βίντεο (video frame)  $N1 \times N2$ , έχει φωτεινότητα (luminance)  $f(\vec{r}, 1)$  στο εικονοστοιχείο  $\vec{r} = (x, y)$  και υπάρχουν  $M$  αντικείμενα  $s_i(\vec{r})$ , τότε έχουμε  $f(\vec{r}, 1) = \sum_{i=1}^M s_i(\vec{r})$

Αν κάθε αντικείμενο αντικαθίσταται από  $\vec{r}_i(t)$  από το πλαίσιο 1 έως το πλαίσιο  $t$ , το πλαίσιο  $t$  έχει φωτεινότητα  $\sum_{i=1}^M s_i(\vec{r} - \vec{r}_i(t))$ . Το τετραγωνικό σφάλμα στο πεδίο του χώρου ανάμεσα στο πλαίσιο  $t$  και στο πλαίσιο 1, με τα αντικείμενα να μετατοπίζονται κατά  $\vec{r}_k(t)$ , είναι ίσο με :

$$J_{spatial}(\vec{r}) = \left\| \sum_{i=1}^M [s_i(\vec{r} - \vec{r}_i(t)) - s_i(\vec{r} - \vec{r}_k(t))] \right\|^2$$

Οι χωρικές μέθοδοι ελαχιστοποιούν  $J_{spatial}(\bar{\mathbf{r}})$  σε όλα τα εικονοστοιχεία  $\bar{\mathbf{r}}$  προκειμένου να υπολογιστούν οι σωστές μετατοπίσεις. Το σφάλμα ελαχιστοποιείται στο  $\bar{\mathbf{r}}_k(t) = \bar{\mathbf{r}}_i(t)$  με  $1 \leq i \leq M$ . Κατά συνέπεια, η κίνηση υπολογίζεται σε κάθε εικονοστοιχείο και η κατάτμηση επιτυγχάνεται ταυτόχρονα.

Ο μετασχηματισμός Fourier (Fourier Transform-FT) του πλαισίου 1 είναι  $F(\bar{\omega}, \mathbf{1}) = \sum_{i=1}^M S_i(\bar{\omega})$  και του πλαισίου  $t$

$$F(\bar{\omega}, t) = \sum_{i=1}^M S_i(\bar{\omega}) e^{-j\bar{\omega}^T \bar{\mathbf{r}}_i(t)}$$

και το αντίστοιχο τετραγωνικό σφάλμα δίνεται από την σχέση :

$$J_{freq}(\bar{\omega}) = \left\| \sum_{i=1}^M S_i(\bar{\omega}) \left[ e^{-j\bar{\omega}^T \bar{\mathbf{r}}_i(t)} - e^{-j\bar{\omega}^T \bar{\mathbf{r}}_k(t)} \right] \right\|^2$$

Ομοίως και με το τετραγωνικό σφάλμα στο πεδίο του χώρου, το σφάλμα στο πεδίο του Fourier ελαχιστοποιείται όταν  $\bar{\mathbf{r}}_k(t) = \bar{\mathbf{r}}_i(t)$ . Σύμφωνα με το θεώρημα του Parseval, η συνολική ενέργεια ενός σήματος, στην προκειμένη περίπτωση το σφάλμα, υπολογισμένο στο πεδίο του χώρου είναι ίσο με την ενέργεια στο πεδίο του Fourier [4]. Αυτό αποδεικνύεται εύκολα μαθηματικά, αλλά είναι επίσης και μια δήλωση της φυσικής αρχής της συντήρησης της ενέργειας. Κατά συνέπεια, τα σφάλματα στο πεδίο του χώρου και της συχνότητας είναι ίσα, και οι ελαχιστοποιήσεις τους δίνουν τα ίδια αποτελέσματα. Όμως όπως δείχνουμε παρακάτω, αυτά τα σφάλματα συμπεριφέροντε πολύ διαφορετικά στην παρουσία πρόσθετου θορύβου στην εικόνα και πλέον δίνουν διαφορετικά αποτελέσματα.

Παρακάτω παρουσιάζουμε την επίδραση του θορύβου στην ελαχιστοποίηση του σφάλματος.

Τα πλαίσια με θόρυβο την χρονική στιγμή  $t$  είναι  $f_n(\bar{\mathbf{r}}, t) = \sum_{i=1}^M s_i(\bar{\mathbf{r}} - \bar{\mathbf{r}}_i(t), 1) + n(\bar{\mathbf{r}})$  με πρόσθετο θόρυβο  $n(\bar{\mathbf{r}})$ .

Ο θόρυβος ακολουθεί μια μηδέν-μέση Γκαουσιανή πυκνότητα πιθανότητας (zero-mean Gaussian probability density)  $n(\bar{\mathbf{r}}) \sim N(\mathbf{0}, \sigma_n^2)$ . Αυτό είναι το πιο κοινό μοντέλο για την μέτρηση του θορύβου στα βίντεο [5]. Έτσι, το χωρικό σφάλμα γίνεται :

$$\begin{aligned} J_{spatial,n}(\bar{\mathbf{r}}) &= \left\| f(\bar{\mathbf{r}}, t) - \hat{f}(\bar{\mathbf{r}}, t) - n(\bar{\mathbf{r}}) \right\|^2 \\ &= \left\| \sum_{i=1}^M [s_i(\bar{\mathbf{r}} - \bar{\mathbf{r}}_i(t)) - s_i(\bar{\mathbf{r}} - \bar{\mathbf{r}}_k(t))] \right\|^2 + \|n(\bar{\mathbf{r}})\|^2 \\ &\quad - 2n(\bar{\mathbf{r}}) \left( \sum_{i=1}^M s_i(\bar{\mathbf{r}} - \bar{\mathbf{r}}_i(t)) - s_i(\bar{\mathbf{r}} - \bar{\mathbf{r}}_k(t)) \right) \end{aligned}$$

Αφού ο πρόσθετος θόρυβος είναι μια τυχαία διαδικασία, οι τιμές του αλλάζουν για κάθε «πραγματοποίηση» του βίντεο. Κατά συνέπεια, εξετάζουμε το μέσο του τετραγωνικού σφάλματος, όσον αφορά τον θόρυβο  $n(\vec{r})$  :

$$E_n [J_{spatial,n}(\vec{r})] = E_n \left[ \left\| \sum_{i=1}^M [s_i(\vec{r} - \vec{r}_i(t)) - s_i(\vec{r} - \vec{r}_k(t))] \right\|^2 + \|n(\vec{r})\|^2 - 2n(\vec{r}) \left( \sum_{i=1}^M s_i(\vec{r} - \vec{r}_i(t)) - s_i(\vec{r} - \vec{r}_k(t)) \right) \right] = J_{spatial}(\vec{r}) + \sigma_n^2$$

Για μηδενικό μέσο θόρυβο, δηλαδή  $E_n [n(\vec{r})] = 0$ , με διαφορά (variance)  $\sigma_n^2$  σε κάθε εικονοστοιχείο  $\vec{r}$ . Δεδομένου ότι το λάθος «μολύνεται» από αυτόν τον πρόσθετο θόρυβο σε κάθε εικονοστοιχείο, η επιρροή του θορύβου είναι σημαντική και αναμένεται να μειώσει (degrade) την διαδικασία εκτίμησης κίνησης καθώς η διαφορά (variance)  $\sigma_n^2$  αυξάνεται. Στο πεδίο της συχνότητας, η εξίσωση σφάλματος γίνεται :

$$J_{freq,n}(\vec{\omega}) = \|F(\vec{\omega}, t) - \hat{F}_n(\vec{\omega}, t)\|^2 = \left\| \sum_{i=1}^M S_i(\vec{\omega}) \left[ e^{-j\vec{\omega}^T \vec{r}_i(t)} - e^{-j\vec{\omega}^T \vec{r}_k(t)} \right] - N(\vec{\omega}) \right\|^2$$

$$= J_{freq}(\vec{\omega}) + \|N(\vec{\omega})\|^2 - 2N(\vec{\omega}) \left( \sum_{i=1}^M S_i(\vec{\omega}) \left[ e^{-j\vec{\omega}^T \vec{r}_i(t)} - e^{-j\vec{\omega}^T \vec{r}_k(t)} \right] \right)$$

Όπου ο θόρυβος  $FT N(\vec{\omega}) = FT [n(\vec{r})]$  ακολουθεί μια μηδενική μέση Γκαουσιανή κατανομή με  $E_N [N(\vec{\omega})] = 0$  και  $E_N [N^2(\vec{\omega})] = \sigma_n^2 \delta(\vec{\omega})$ . Έτσι, το μέσο τετραγωνικό σφάλμα στο πεδίο της συχνότητας δίνεται από την εξίσωση :  $E_n [J_{freq,n}(\vec{\omega})] = J_{freq}(\vec{\omega}) + \sigma_n^2 \delta(\vec{\omega})$ . Αυτό δείχνει ότι η επιρροή του θορύβου μόνο για  $\vec{\omega} = 0$ , αφού είναι κάθετο στο  $\delta(\vec{\omega})$ . Οπότε αναμένεται να επηρεαστεί το μέσο τετραγωνικό σφάλμα σε πολύ μικρότερο βαθμό από ότι στο πεδίο το χώρου. Αυτό φαίνεται και στα πειράματά μας, όπου ο θόρυβος επηρεάζει σημαντικά το χωρικό σφάλμα, ενώ το σφάλμα στο πεδίο της συχνότητας παραμένει πιο σταθερό για αυξανόμενη τιμή θορύβου.

Κατά συνέπεια, η εκτίμηση της κίνησης που αποκτάται από το τετραγωνικό σφάλμα στο πεδίο του Fourier είναι περισσότερο σταθερές από το χωρικό τετραγωνικό σφάλμα, καθιστώντας έτσι την πληροφορία στο πεδίο του Fourier πιο χρήσιμη για την εκτίμηση της κίνησης σε βίντεο (video motion estimation).

Οι χωρικές μέθοδοι υποφέρουν από το πρόβλημα της τοπικότητας "localization problem", αφού η ολική του φύση δεν μπορεί να προσδιορίσει σε ποια εικονοστοιχεία αντιστοιχούν οι μετατοπίσεις που υπολογίζουμε. Λόγω αυτού του προβλήματος η χωρική πληροφορία είναι απαραίτητο να χρησιμοποιηθεί για κατάτμηση της κίνησης. Μόλις οι μετατοπίσεις αντικειμένων  $\vec{r}_i$  υπολογιστούν, όλα τα εικονοστοιχεία του πλαισίου μετατοπίζονται κατά αυτήν την ποσότητα, έχουμε δηλαδή

$$d_i(\bar{r}) = s_i(\bar{r} - \bar{r}_i(t)) - s_i(\bar{r} - \bar{r}_k(t)) - n(\bar{r}) \quad (1 \leq i, k \leq M):$$

$$d_i(\bar{r}) = n(\bar{r}), \quad i = k$$

$$d_i(\bar{r}) = m_i(\bar{r}) + n(\bar{r}), \quad i \neq k$$

Όπου  $m_i(\bar{r}) = s_i(\bar{r} - \bar{r}_i(t)) - s_i(\bar{r} - \bar{r}_k(t))$  για  $i \neq k$ . Όταν ένα εικονοστοιχείο μετατοπίζεται σωστά, τότε  $d_i(\bar{r}) = n(\bar{r})$ , όπου είναι Γκαουσιανός θόρυβος  $N(0, \sigma_n^2)$ . Αν μετατοπίζεται λανθασμένα, ακολουθεί μια μη Γκαουσιανή κατανομή, αφού  $m_i(\bar{r})$  είναι μια άγνωστη εξαρτημένη από το πλαίσιο ποσότητα. Έτσι προκειμένου να προσδιορίσουμε την περιοχή του αντικειμένου  $i$  πρέπει μόνο να εξετάσουμε την Γκαουσιανότητα του  $d_i(\bar{r})$ . Αυτό γίνεται εύκολα υπολογίζοντας το kurtosis  $kurt(y) = E\{y^4\} - 3(E\{y^2\})^2$  που είναι μηδέν όταν η τυχαία μεταβλητή  $y$  είναι Γκαουσιανή.

## 5.2. Εκτίμηση της κίνησης στο πεδίο του μετασχηματισμού Fourier ( Fourier Transform domain)

Η σφαιρική κίνηση (global motion), η κίνηση με σταθερή ταχύτητα (constant-velocity) και η μεταθετική κίνηση (translational motion) σε μια ακολουθία εικόνων όπου όλα τα εικονοστοιχεία κινούνται με την ίδια ταχύτητα καταλήγει σε ένα επίπεδο φάσμα Fourier, δηλαδή συχνότητες με μη μηδενικό πλάτος μετασχηματισμού Fourier προσδιορίζονται σε ένα επίπεδο κάθετο στην διεύθυνση της κίνησης [20], [21].

Αρχικά περιγράφουμε την ομοιόμορφη μεταθετική κίνηση μιας εικόνας στο πεδίο του Fourier. Έστω  $u_0(x_1, x_2)$  η συνεχής ένταση της εικόνας σαν συνάρτηση των συνεχών χωρικών συντεταγμένων  $(x_1, x_2)$ . Ομοίως  $u(x_1, x_2, x_3)$  η συνεχής συνάρτηση έντασης μιας εικόνας όπου η  $x_3$  προσδιορίζει την συνιστώσα του χρόνου. Μια ομοιόμορφη μετάθεση (translation) των εντάσεων  $u_0$  σε όλες τις χωρικές συνιστώσες με σταθερή ταχύτητα  $[u_1, u_2]^T$  καταλήγει σε μια συνάρτηση εικόνας εξαρτημένη από τον χρόνο :  $u(x_1, x_2, x_3) = u_0(x_1 - u_1 x_3, x_2 - u_2 x_3)$ . Υποθέτοντας ότι η εικόνα  $u_0$  έχει άπειρες διαστάσεις, είναι απλό να δείξουμε την ακόλουθη σχέση :

$$u_0(x_1 - u_1 x_3, x_2 - u_2 x_3) = [u_0(x_1, x_2)\delta(x_3)] * \delta(x_1 - u_1 x_3, x_2 - u_2 x_3)$$

Όπου το "\*" προσδιορίζει την τρισδιάστατη συνέλιξη ενώ το  $\delta(\cdot)$  την συνάρτηση δέλτα (Dirac impulse) . Εφαρμόζοντας τον μετασχηματισμό Fourier σε κάθε όρο της παραπάνω συνέλιξης παίρνουμε :

$$F\{u(x_1, x_2, x_3)\} = U_0(f_1, f_2)\delta(f_1 u_1 + f_2 u_2 + f_3)$$

Αφού

$$F\{u_0(x_1, x_2)\delta(x_3)\} = F\{u_0(x_1, x_2)\} = U_0(f_1, f_2) \quad \forall f_3$$



$$F\{\delta(x_1 - u_1x_3, x_2 - u_2x_3)\} = \delta(f_1u_1 + f_2u_2 + f_3)$$

Εμφανώς, το φάσμα Fourier της μεταθετικής εικόνας περιορίζεται στο επίπεδο  $f_1u_1 + f_2u_2 + f_3 = 0$  στο χωρο-χρονικό πεδίο συχνότητας  $(f_1, f_2, f_3)$  από όπου μπορούμε να ανακαλύψουμε την διεύθυνση (direction) και το μέγεθος (amplitude) της κίνησης. Όμως αν η κίνηση δεν είναι μεταθετική ή η ταχύτητα ποικίλει στον χρόνο, δεν προσδιορίζεται το επίπεδο στο πεδίο της συχνότητας.

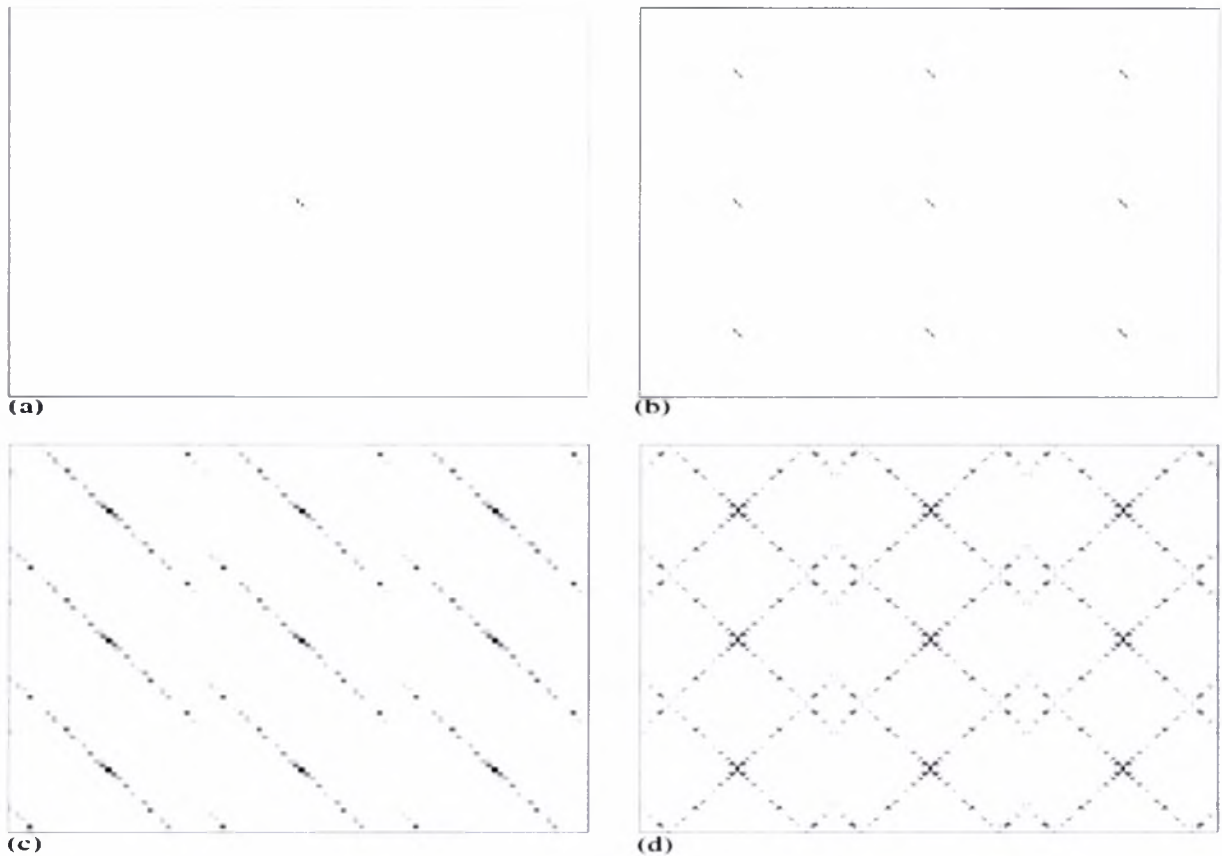
Στην πράξη, η ακολουθία εικόνων προσδιορίζεται ως  $u(n_1, n_2, n_3)$  όπου προέρχεται από ομοιόμορφη μετάθεση της εικόνας  $u_0(x_1, x_2)$  με διδιάστατη μετατόπιση  $(d_1, d_2)$  μεταξύ κάθε δυο διαδοχικών πλαισίων. Επομένως έχουμε την σχέση  $u(n_1, n_2, n_3) = u_0(n_1 - d_1n_3, n_2 - d_2n_3)$ . Αντικαθιστώντας την συνέλιξη με το διακριτό στον χρόνο ισότιμο του, δηλαδή το CSFT (convolution sum Fourier Transform) με το DSFT (discrete-sspace Fourier Transform) καταλήγουμε στην σχέση :

$$F_{DSFT}\{u[n_1, n_2, n_3]\} = U_0(f_1, f_2)\delta(f_1d_1 + f_2d_2 + f_3)$$

Όπου  $F_{DSFT}$  προσδιορίζει το DSFT και  $U_0(f_1, f_2) = F_{DSFT}\{u_0(n_1, n_2)\}$ . Λόγω της περιοδικής φύσης του DSFT αυτό το φασματικό επίπεδο είναι επίσης περιοδικό. Παρόμοια σχέση ισχύει για το DFT και την κυκλική συνέλιξη εφαρμοσμένη στην εικόνα  $u_0$ , εκτός από το ότι το Kronecker Δέλτα αντικαθιστά το Dirac impulse ενώ οι ακέραιοι δείκτες  $k_1, k_2, k_3$  αντικαθιστούν τις συχνότητες  $f_1, f_2, f_3$  :  $F_{DFT}\{u[n_1, n_2, n_3]\} = U_0[k_1, k_2]\delta[k_1d_1 + k_2d_2 + k_3]$ .

Το  $\delta[k_1d_1 + k_2d_2 + k_3]$  στην παραπάνω σχέση είναι μια έκδοση δειγματοληψίας (sampled version) του  $\delta[f_1d_1 + f_2d_2 + f_3]$ . Με άλλα λόγια περιγράφει δείγματα του επιπέδου  $f_1d_1 + f_2d_2 + f_3 = 0$  στις διακριτές συχνότητες  $(k_1, k_2, k_3)$ . Γενικά για διακριτές τιμές των  $d_1$  και  $d_2$  η ένταση αυτών των δειγμάτων θα είναι υψηλότερη, ενώ για μη διακριτές τιμές θα είναι χαμηλότερη.

Στο παρακάτω σχήμα παρουσιάζεται η συμπεριφορά του φάσματος του Fourier Transform, του DSFT (discrete-sspace Fourier Transform) και του DFT (discrete Fourier Transform) στην διδιάστατη περίπτωση.



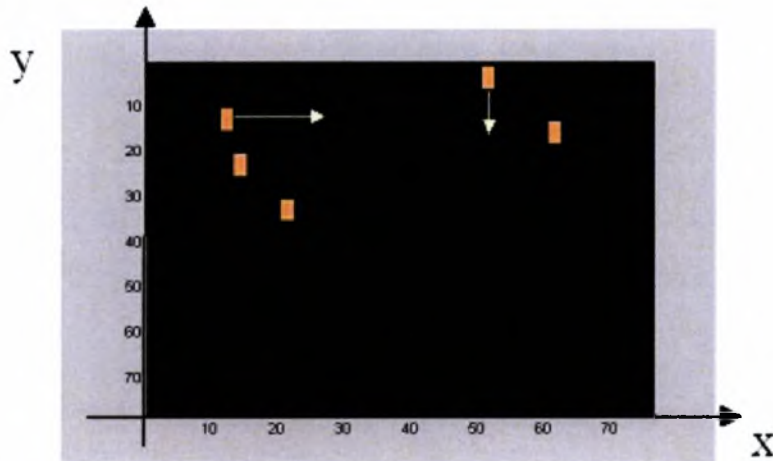
Ακολουθεί ένα παράδειγμα που δείχνει τον μετασχηματισμό Fourier ενός μετακινούμενου σήματος ή μιας εικόνας.

Αν ένα σήμα ή μια εικόνα μετακινείται στο πεδίο του χώρου με ορισμένη ταχύτητα  $u_0$ , το φάσμα του ίδιου στατικού σήματος στο πεδίο του Fourier μετατοπίζεται κατά μήκος του χρονικού διανύσματος κατά μία τιμή ίση με την ταχύτητα  $V_0$ . Για παράδειγμα, για ένα ορθογώνιο παραλληλόγραμμο έχουμε το φάσμα στατικού ορθογωνίου (static rectangle spectrum) και το φάσμα του μετακινούμενου ορθογωνίου (moving rectangle spectrum).

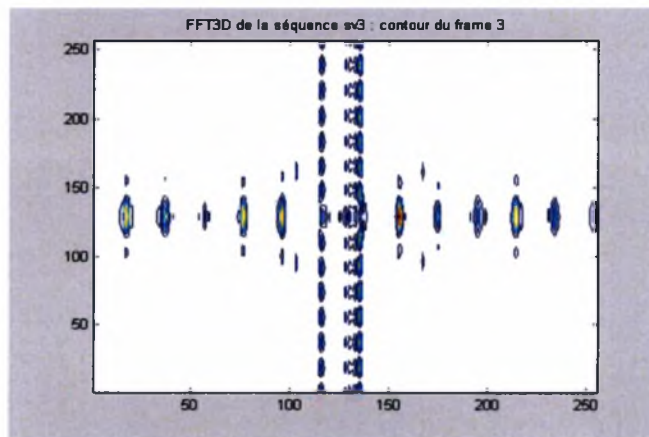
Παρακάτω αναλύουμε τον μετασχηματισμό Fourier ενός ορθογωνίου σε μετάθεση (translation). Αναλύουμε μέσω του μετασχηματισμού Fourier μία ακολουθία πέντε ορθογωνίων

σε μετάθεση. Τα τρία είναι σε οριζόντια μετάθεση με ταχύτητες  $V_h = 1,3$  και  $10 \text{ pix/fr}$  και τα δυο είναι σε κάθετη μετάθεση με ταχύτητες  $V_v = 1$  και  $3 \text{ pix/fr}$ .

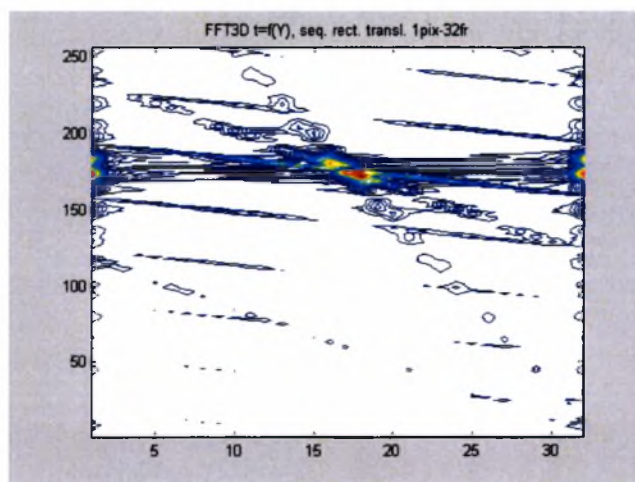
Στο ακόλουθο σχήμα παρουσιάζονται 16 ακολουθίες πλαισίων σε οριζόντια ( $V_h = 1,3$  και  $10 \text{ pix/fr}$ ) και κάθετη ( $V_v = 1$  και  $3 \text{ pix/fr}$ ) μετάθεση.



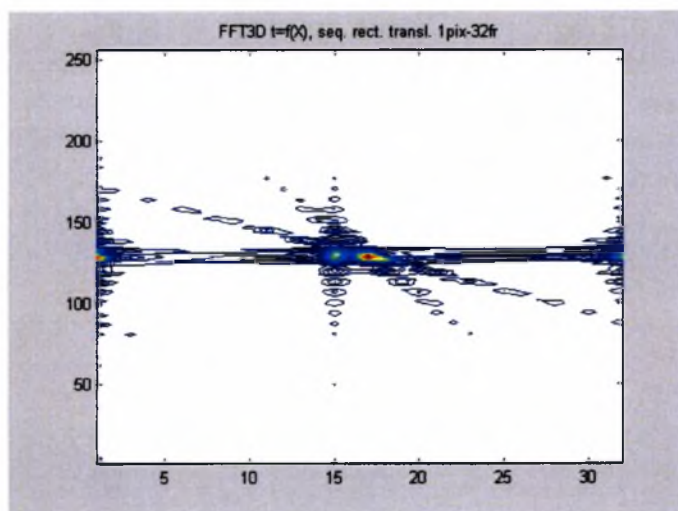
Στο ακόλουθο σχήμα παρουσιάζεται το φάσμα του διανύσματος  $(k_x, k_y)$  ενός μετακινούμενου ορθογωνίου (frame 3).



Παρακάτω ακολουθεί το φάσμα του διανύσματος της συχνότητας  $(k_y, k_t)$  τριων οριζόντια μετακινούμενων ορθογωνίων με σταθερή ταχύτητα.



Παρακάτω ακολουθεί το φάσμα του διανύσματος της συχνότητας  $(k_x, k_t)$  δυο κάθετα μετακινούμενων ορθογωνίων με σταθερή ταχύτητα.



### 5.3. Εκτίμηση της κίνησης από την φάση του Fourier

Η τεχνική την οποία παρουσιάζουμε εκμεταλλεύεται την περιοδικότητα στις χωροχρονικές-συχνοτικές αναπαραστάσεις (spatio-temporal-frequency representations) της εικόνας που προκύπτουν από ορισμένους τύπους κίνησης της εικόνας. Αποδεικνύεται ότι ο χωρο-χρονικός μετασχηματισμός Fourier μιας ακολουθίας εικόνων στην οποία το περιεχόμενο της εικόνας κινείται με σταθερή ταχύτητα, οδηγεί σε μια αναπαράσταση χωροχρόνου-συχνότητας που είναι ίση με το χωρικό μετασχηματισμό Fourier της εικόνας πολλαπλασιασμένο με μια Δέλτα συνάρτηση στο πεδίο του χρόνου-συχνότητας [12]. Αυτή η συνάρτηση Δέλτα εξαρτάται από την ταχύτητα της εικόνας που μπορεί να υπολογιστεί αν κάποιος ξέρει την θέση της Δέλτα συνάρτησης και οποιαδήποτε χωρική συχνότητα. Επειδή αυτή η προσέγγιση είναι βασισμένη στην κίνηση της εικόνας, παρά στην κίνηση των αντικειμένων υποθέτουμε ομοιόμορφο φόντο όταν υπολογίζεται η κίνηση των αντικειμένων.

Μια εναλλακτική προσέγγιση της χωροχρονικής συχνότητας, χρησιμοποιεί τον χωρικό μετασχηματισμό Fourier μαζί με τον μετασχηματισμό Hough, και όχι τον χωρο-χρονικό μετασχηματισμό Fourier. Η ταχύτητα σε κάθε σημείο στην εικόνα υπολογίζεται μεταχειρίζοντας μια τοπική περιοχή ως μια ευδιάκριτη υπο-εικόνα (sub-image) που μετατοπίζεται με κάποια ταχύτητα, και προσδιορίζοντας τις συνιστώσες Fourier που εκθέτουν τις αλλαγές των διαστάσεων και της φάσης. Η ίδια ανίχνευση της ταχύτητας ολοκληρώνεται χρησιμοποιώντας ένα κατάλληλο μετασχηματισμό Hough, όπου ενσωματώνει την σχέση μεταξύ της ταχύτητας και της αλλαγής της φάσης [8], [11]. Η ταχύτητα μετριέται με την εντόπιση των τοπικών μέγιστων στο διάστημα Hough. Επειδή ο υπολογισμός της ταχύτητας βασίζεται στην πληροφορία της φάσης Fourier, μπορούμε να υπολογίσουμε την ταχύτητα οποιασδήποτε συνάρτησης σήματος εκτός από την ομοιόμορφη περιοχή. Γενικά, τα διανύσματα ταχύτητας υπολογίζονται με ακρίβεια υπο-εικονοστοιχείου (sub-pixel).

#### Επισκόπηση της προσέγγισης

Ο διακριτός μετασχηματισμός Fourier (discrete Fourier Transform)  $F f(x, y)$  μιας δισδιάστατης συνάρτησης  $f(x, y)$  δίνεται από την εξίσωση :  $F f(x, y) = F(k_x, k_y) = \sum_x \sum_y f(x, y) e^{ik_x x + k_y y}$

Και ο αντίστροφος διακριτός μετασχηματισμός Fourier (inverse discrete Fourier Transform), δίνεται από την σχέση :

$$f(x, y) = F^{-1} (F(k_x, k_y)) = \frac{1}{(2\pi)^2} \sum_{k_x} \sum_{k_y} |F(k_x, k_y)| e^{i\varphi(k_x, k_y)} e^{-ik_x x + k_y y}$$

Όπου  $|F(k_x, k_y)|$  είναι το πραγματικό το πραγματικό εκτιμημένο φάσμα εύρους (real valued amplitude spectrum) και  $\varphi(k_x, k_y)$  είναι το πραγματικό εκτιμημένο φάσμα συχνότητας.

Μια συνάρτηση  $f(x, y)$  που μετατοπίζεται με σταθερή ταχύτητα  $(v_x, v_y)$  γίνεται  $f(x - v_x \delta t, y - v_y \delta t)$ . Σύμφωνα με την ιδιότητα της μετατόπισης, ο μετασχηματισμός Fourier της μετατοπισμένης συνάρτησης δίνεται από την σχέση :

$$F(f(x - v_x \delta t, y - v_y \delta t)) = |F(k_x, k_y)| e^{i\varphi(k_x, k_y)} e^{-i(k_x v_x \delta t + k_y v_y \delta t)}$$

Κατά συνέπεια, μια χωρική μετατόπιση  $(v_x \delta t, v_y \delta t)$  ενός σήματος στο πεδίο του χώρου, δηλαδή η συνάρτηση  $f(x, y)$  μετατοπισμένη σε  $f(x - v_x \delta t, y - v_y \delta t)$ , παράγει μόνο μια αλλαγή στην φάση των συνιστωσών Fourier στο πεδίο της συχνότητας. Αυτή η αλλαγή της φάσης είναι  $e^{-i(k_x v_x \delta t + k_y v_y \delta t)}$ . Προκειμένου να υπολογίσουμε την ταχύτητα ενός σήματος που μετατοπίζεται με σταθερή ταχύτητα στην εικόνα, απλά πρέπει αν προσδιορίσουμε σύνολο των συνιστωσών της συχνότητας  $k_x$  και  $k_y$ , που έχουν τροποποιηθεί με την ίδια μετατόπιση φάσης, δηλαδή  $e^{-i(k_x v_x \delta t + k_y v_y \delta t)}$ .

Το φάσμα της φάσης για την μετατόπιση την χρονική στιγμή  $t + \delta t$  είναι ίσο με το φάσμα της φάσης την χρονική στιγμή  $t$  πολλαπλασιασμένο με την αλλαγή της φάσης όπως φαίνεται παρακάτω :

$$e^{i\varphi_{t+\delta t}(k_x, k_y)} = e^{-i(k_x v_x \delta t + k_y v_y \delta t)} e^{i\varphi_t(k_x, k_y)} = e^{i(\varphi_t(k_x, k_y) - (k_x v_x \delta t + k_y v_y \delta t))}$$

Η φάση την χρονική στιγμή  $t + \delta t$  είναι ίση με την αρχική φάση την χρονική στιγμή  $t$  μείον την ποσότητα  $(k_x v_x \delta t + k_y v_y \delta t)$ . Δηλαδή,  $\varphi_{t+\delta t}(k_x, k_y) = \varphi_t(k_x, k_y) - (k_x v_x \delta t + k_y v_y \delta t)$ .

Από αυτήν την σχέση προκύπτει ο ορισμός της ταχύτητας  $v_x$  και  $v_y$  όπως φαίνεται παρακάτω :

$$v_y = \frac{1}{k_y \delta t} (\varphi_t(k_x, k_y) - \varphi_{t+\delta t}(k_x, k_y) - k_x v_x \delta t)$$

Αυτή η εξίσωση εκφυλίζεται αν  $k_y = 0$ . Σε αυτήν την περίπτωση αντικαθιστούμε  $k_y = 0$  και χρησιμοποιούμε την ακόλουθη έκφραση :  $v_x = \frac{1}{k_x \delta t} (\varphi_t(k_x, k_y) - \varphi_{t+\delta t}(k_x, k_y))$ .

Αν έχουμε δυο εικόνες τις χρονικές στιγμές  $t = t_0$  και  $t = t_0 + \delta t$ , μπορούμε να υπολογίσουμε τα  $\varphi_{t_0}$  και  $\varphi_{t_0+\delta t}$ . Μεταχειρίζοντας την παραπάνω εξίσωση σαν ένα μετασχηματισμό Hough, με ένα διδιάστατο χώρο του μετασχηματισμού Hough ορισμένο στα  $v_x, v_y$  μπορούμε να υπολογίσουμε το  $v_y$  για όλες τις πιθανές τιμές των  $v_x$  και για όλες τις γνωστές τιμές των  $k_x, k_y, \varphi_t(k_x, k_y), \varphi_{t+\delta t}(k_x, k_y)$ . Το τοπικό μέγιστο των  $v_x, v_y$  στο χώρο του μετασχηματισμού Hough δηλώνει τις συνιστώσες Fourier που περιλαμβάνουν σήματα στο πεδίο του χώρου τα οποία μετακινούνται με σταθερή ταχύτητα  $v_x, v_y$ . Το  $\varphi(k_x, k_y)$  αντιπροσωπεύει την απόλυτη φάση της συχνότητας  $(k_x, k_y)$ . Όμως στο πεδίο του Fourier, η φάση οριοθετείται σε  $\pm 2\pi$  και οι τιμές της φάσης θα κόβονται ('wrap') καθώς περνούν αυτό το κατώφλι. Στην πραγματικότητα, οι αλλαγές της φάσης αναπαρίστανται σαν ακέραια πολλαπλάσια του  $2\pi$  που προστίθενται στις wrapped τιμές (με phase unwrapping).

Η πολυπλοκότητα για τον υπολογισμό της ταχύτητας σε μια περιοχή είναι  $O(nm^2)$  όπου  $m$  η διάσταση του πεδίου Fourier και  $n$  η διάσταση του χώρου μετασχηματισμού του Hough, και αυτό γιατί η εξίσωση

$$v_y = \frac{1}{k_y \delta t} (\varphi_t(k_x, k_y) - \varphi_{t+\delta t}(k_x, k_y) - k_x v_x \delta t)$$
 πρέπει να υπολογιστεί για όλες τις χωρικές συχνότητες  $x$  και  $y$ , και για όλες τις πιθανές τιμές του  $v_x$ . Επομένως, η πολυπλοκότητα του υπολογισμού της συνολικού πεδίου οπτικής ροής είναι  $O(nm^4)$ , αφού η διάσταση του πεδίου του Fourier  $m$  είναι ίδια με αυτή της αρχικής εικόνας.

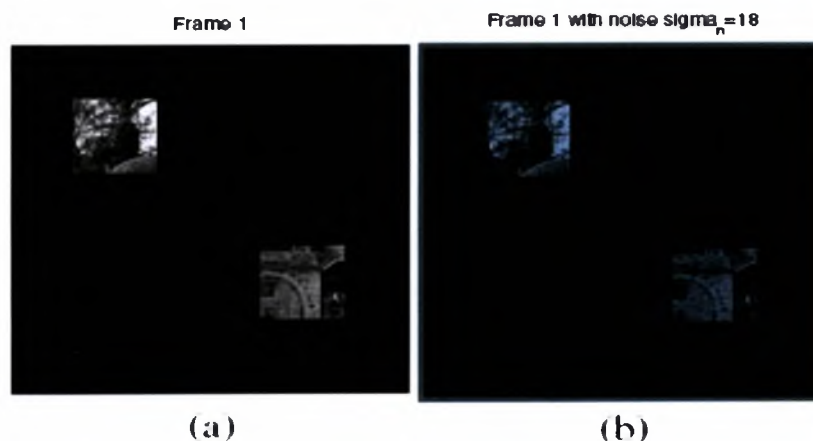
## Κεφάλαιο 6 Πειράματα

### 6.1. Εκτίμηση κίνησης με ελαχιστοποίηση σφάλματος σε συνθετική και πραγματική ακολουθία

Αρχικά, θα παρουσιάσουμε τα πειράματα που αφορούν την ελαχιστοποίηση του σφάλματος τόσο σε μία συνθετική ακολουθία (synthetic sequence) όσο και σε μία πραγματική ακολουθία (real-sequence).

#### Συνθετική ακολουθία - Synthetic sequence

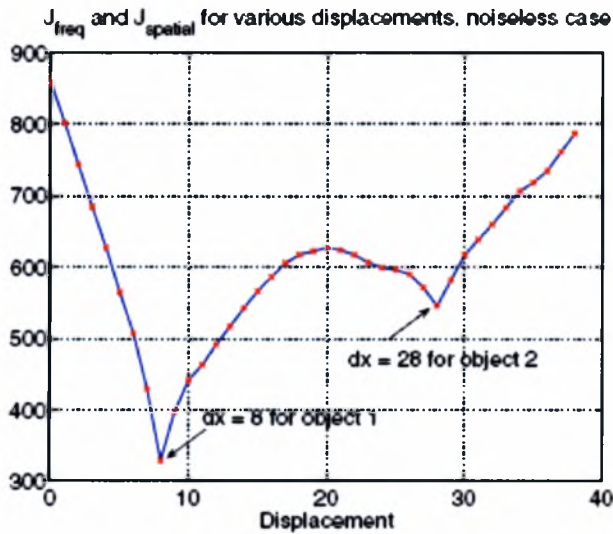
Στο παρακάτω σχήμα, παρουσιάζεται μια συνθετική ακολουθία δυο κινούμενων αντικειμένων με διαφορετικές οριζόντιες μετατοπίσεις. Αποκτώντας το τοπικό ελάχιστο του  $J_{spatial}$  και του  $J_{freq}$  παίρνουμε τους υπολογισμούς των κινήσεών τους, καθώς αυτοί πρέπει να ελαχιστοποιούνται στο  $\bar{r} = \bar{r}_1$  και στο  $\bar{r} = \bar{r}_2$ . Στο σχήμα 1a έχουμε απουσία θορύβου, ενώ στο 1b όχι.



Σχήμα 1. Συνθετική ακολουθία πλαίσιο 1 : a) χωρίς θόρυβο (noiseless) b) θόρυβο με  $\sigma_n^2 = 18$

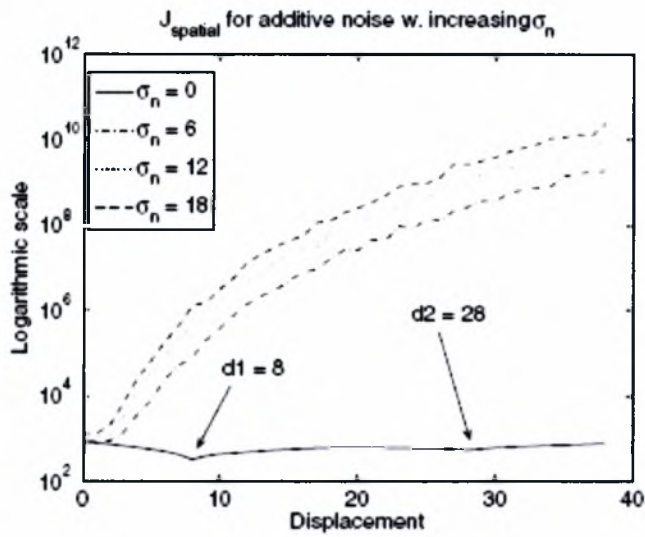


Το Σχήμα 2 δείχνει ότι τα σφάλματα ελαχιστοποιούνται στις δυο περιοχές  $\bar{r} = \bar{r}_1$  και  $\bar{r} = \bar{r}_2$  που αναφέραμε προηγουμένως, και ότι οι καμπύλες των δυο σφαλμάτων συμπίπτουν με την απουσία θορύβου, όπως αναμένεται σύμφωνα με το θεώρημα του Parseval.



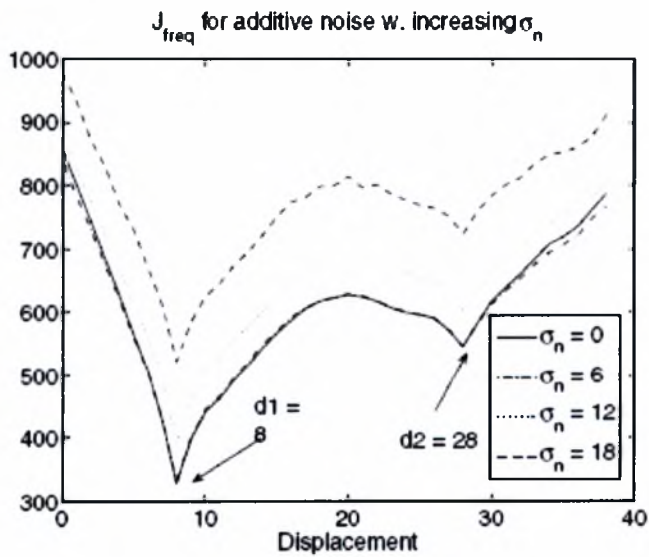
**Σχήμα 2. Συνθετική ακολουθία : Τετραγωνισμένο σφάλμα στο χώρο και στη συχνότητα (Squared FT and spatial errors). Περίπτωση με μηδενικό θόρυβο.**

Στο Σχήμα 3 φαίνεται η καταστροφική επιρροή του θορύβου στην διαδικασία ελαχιστοποίησης του  $J_{spatial}$ , καθώς ο υπολογισμός της κίνησης γίνεται εσφαλμένα ακόμη και για μικρές τιμές πρόσθετου θορύβου. Το σφάλμα γίνεται τόσο υψηλό που πρέπει να το σχεδιάσουμε σε λογαριθμική κλίμακα.



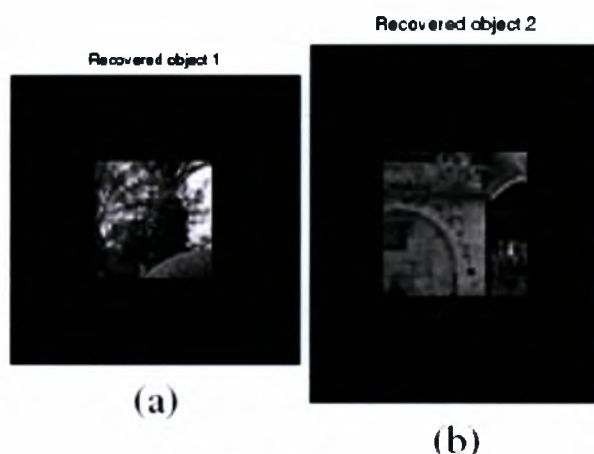
Σχήμα 3. Συνθετική ακολουθία  $J_{spatial}$  για  $0 \leq \sigma_n^2 \leq 19$

Από την άλλη μεριά όμως, όπως φαίνεται στο Σχήμα 4 το  $J_{freq}$  παρουσιάζει αξιοσημείωτη ευρωστία (robustness) στον ίδιο πρόσθετο θόρυβο.



Σχήμα 4. Συνθετική ακολουθία  $J_{freq}$  για  $0 \leq \sigma_n^2 \leq 19$

Όπως αναφέραμε και στις προηγούμενες ενότητες χρησιμοποιούμε το  $J_{freq}$  για τον υπολογισμό της κίνησης. Η χωρική τοπικότητα (localization) των υπολογισμένων κινήσεων παρουσιάζεται χρησιμοποιώντας την χωρική πληροφορία (spatial data). Οι υπολογισμένες μετατοπίσεις χρησιμοποιούνται για να παραμορφώσουν (wrap) ολόκληρο το πλαίσιο την χρονική στιγμή  $t$ , και το παραμορφωμένο πλαίσιο συγκρίνεται με το αρχικό, δίνοντας ακριβή κατάτμηση αντικειμένων (accurate object segmentation) όπως φαίνεται στο ακόλουθο Σχήμα 5.

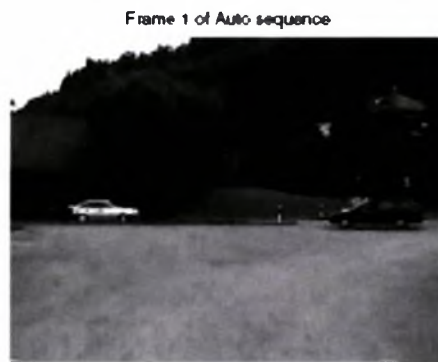


**Σχήμα 5. Συνθετική ακολουθία : Ανακτημένα αντικείμενα με χωρική συγχώνευση (recovered object with spatial fusion) a) αντικείμενο 1 b) αντικείμενο 2**

Τα αποτελέσματα της κατάτμησης μπορεί να χρησιμοποιηθούν για να μειώσουν τις τιμές των τετραγωνισμένων σφαλμάτων. Σε αυτό το πείραμα, οι περιοχές των δυο κινούμενων αντικειμένων χωρίζονται καθ' όλη τη διάρκεια της ακολουθίας, και τα ανταποκρινόμενα σφάλματα στο χώρο και στην συχνότητα (FT and spatial errors) υπολογίζονται ξανά. Στην περίπτωση που δεν υπάρχει θόρυβος, τα χωρικά σφάλματα είναι ίσα με  $2\sigma_n^2$  σε κάθε εικονοστοιχείο και τα FT σφάλματα είναι ανάλογα στο  $\sigma_n^2 \delta(\omega)$ .

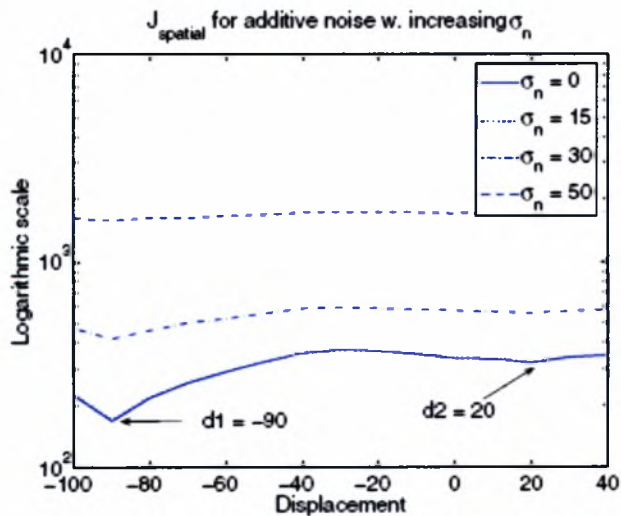
## Πραγματική ακολουθία – Real Sequence : Ακολουθία Αυτοκινήτου - Auto Sequence

Παρουσιάζεται ένα πείραμα με μία πραγματική ακολουθία από δυο αυτοκίνητα που κινούνται αντίθετα μεταξύ τους, όπως φαίνεται στο Σχήμα 6.



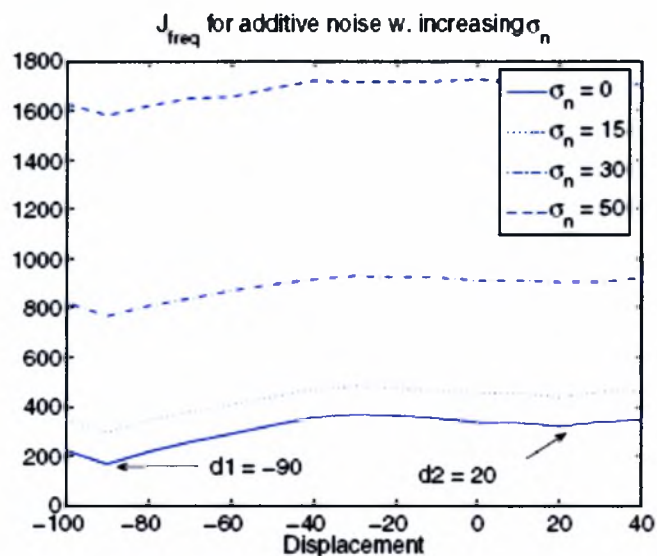
Σχήμα 6. Πλαίσιο 1 ακολουθίας αυτοκινήτου – auto sequence

Στην περίπτωση που δεν υπάρχει θόρυβος,  $J_{spatial} = J_{freq}$ , και ελαχιστοποιούνται για τις σωστές μετατοπίσεις,  $d_1 = 90$  και  $d_2 = 20$ . Στο Σχήμα 7, βλέπουμε το χωρικό σφάλμα σε λογαριθμική κλίμακα.



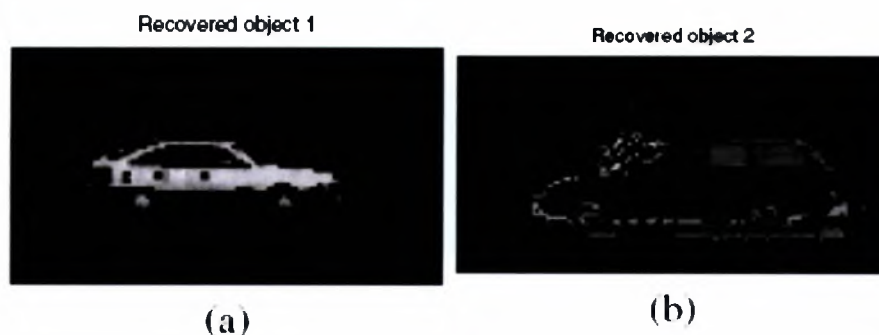
Σχήμα 7. Ακολουθία αυτοκινήτου :  $J_{spatial}$  για  $0 \leq \sigma_n^2 \leq 29$

Το χωρικό σφάλμα (spatial error) είναι λιγότερο σταθερό την παρουσία του θορύβου από το σφάλμα FT (FT error). Αυτό φαίνεται σχηματικά παρακάτω.



Σχήμα 8. Ακολουθία αυτοκινήτων :  $J_{freq}$  για  $0 \leq \sigma_n^2 \leq 29$

Το ακόλουθο Σχήμα 9, δείχνει ότι χρησιμοποιώντας την χωρική πληροφορία (spatial data) επιτυγχάνουμε ακριβή εξαγωγή αντικειμένων (correct object extraction).

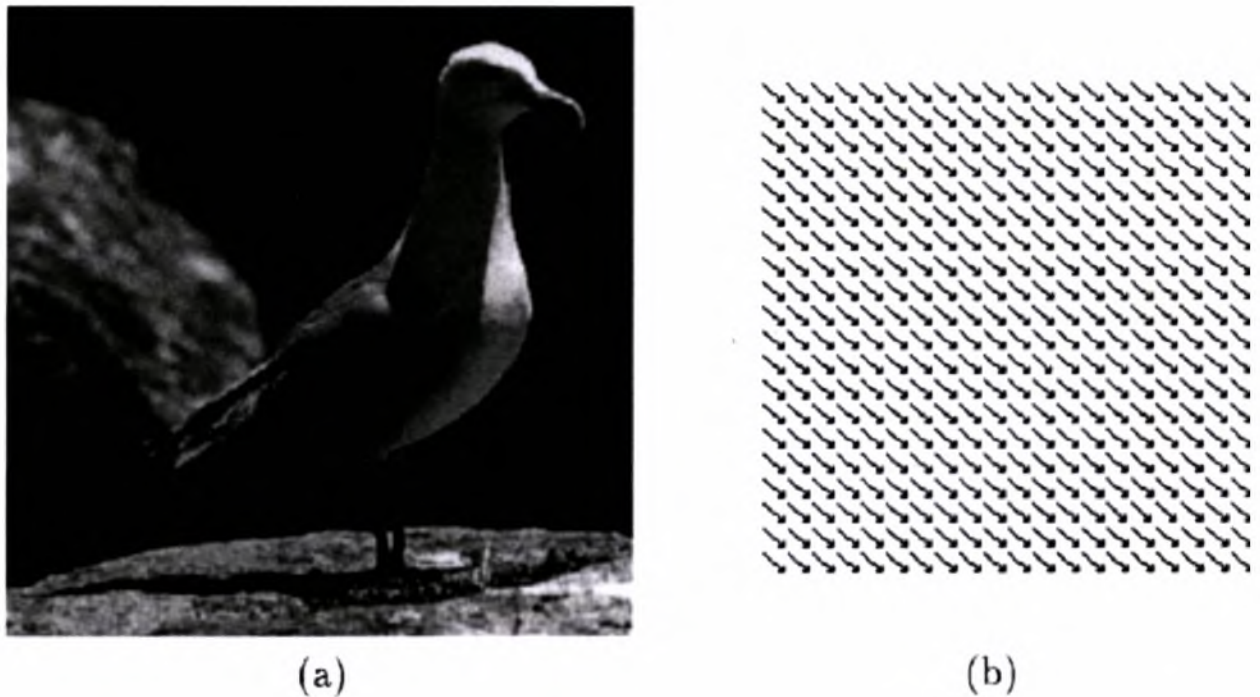


Σχήμα 9. Κατάτμηση (Segmentation) a) αυτοκίνητο 1 b) αυτοκίνητο 2

## 6.2. Εκτίμηση κίνησης από την φάση των συνιστωσών Fourier

Παρακάτω παρουσιάζονται τα αποτελέσματα που προκύπτουν από την εκτίμηση της κίνησης μέσω της φάσης των συνιστωσών Fourier. Πρέπει να σημειώσουμε ότι όλα τα αποτελέσματα υπολογίζονται με μέγεθος παραθύρου 64x64 εικονοστοιχεία.

Το Σχήμα 1 δείχνει το αποτέλεσμα της αποδοχής της τεχνικής σε δυο εικόνες όπου η δεύτερη είναι η πρώτη μετατοπισμένη στην  $x$  και  $y$  διεύθυνση κατά δυο εικονοστοιχεία. Σκοπός αυτού του πειράματος είναι αποδείξει την ακρίβεια (accuracy) και την επαναληπτικότητα (repeatability) της τεχνικής σε πραγματικά δεδομένα με ένα γνωστό πεδίο ροής. Στο Σχήμα 1a παρουσιάζεται η εικόνα που χρησιμοποιούμε για το πείραμα μας, και στο 1b το υπολογισμένο πεδίο ροής χρησιμοποιώντας την πληροφορία της φάσης.



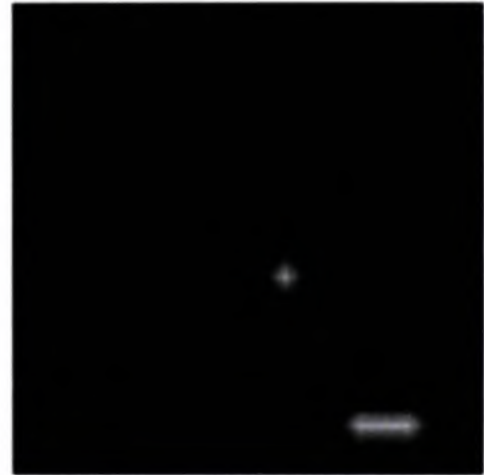
**Σχήμα 1. a) Ακολουθία πειράματος : η οπτική ροή υπολογίζεται χρησιμοποιώντας αυτήν την εικόνα που είναι μετατοπισμένη κατά δυο εικονοστοιχεία στις διευθύνσεις  $x$  και  $y$  b) το πεδίο οπτικής ροής υπολογίζεται χρησιμοποιώντας την πληροφορία της φάσης**

Τα Σχήματα 2a και 2b δείχνουν την απόλυτη τιμή σφάλματος του μεγέθους (absolute value of error in magnitude) και της διεύθυνσης (absolute value of error in direction) του υπολογισμένου πεδίου ροής. Το σφάλμα είναι ανάλογο με το σκότος. Το μέγιστο σφάλμα μεγέθους (maximum

magnitude error) είναι 0.141 pixels και το μέγιστο σφάλμα διεύθυνσης (maximum direction error) είναι 0.025 radians.



(a)



(b)

**Σχήμα 2.** *a) απόλυτη τιμή σφάλματος των διαστάσεων της υπολογισμένης ροής b) απόλυτη τιμή σφάλματος της διεύθυνσης της υπολογισμένης ροής*

Ακολουθεί η απεικόνιση των αποτελεσμάτων εφαρμογής της τεχνικής σε εικόνες μιας πραγματικής σκηνής που αποκτήθηκαν με χρήση κάμερας. Η κίνηση της κάμερας είναι μια απλή τρισδιάστατη μετάθεση (translation). Στο Σχήμα 3a και 3b παρουσιάζονται τα πλαίσια 40 και 41 της πραγματικής κίνησης της ακολουθίας (ground-truth motion sequence).



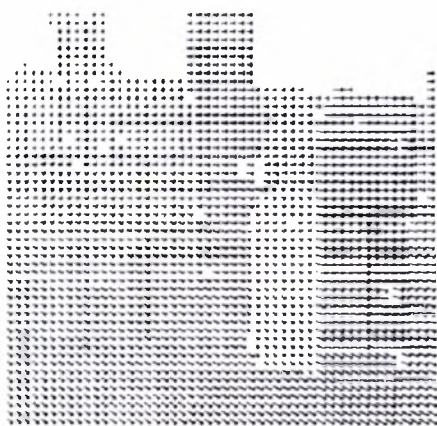
(a)



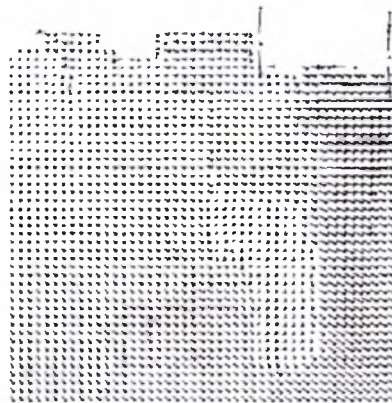
(b)

**Σχήμα 3. a) και b) τα πλαίσια 40 και 41 της ακολουθίας**

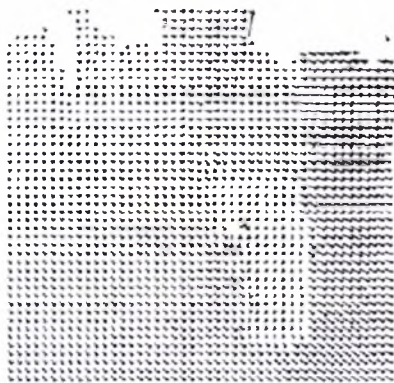
Στα σχήματα 3(d)-(f) δείχνουν το πεδίο οπτικής ροής χρησιμοποιώντας την πληροφορία της φάσης.



(c)



(d)



(e)

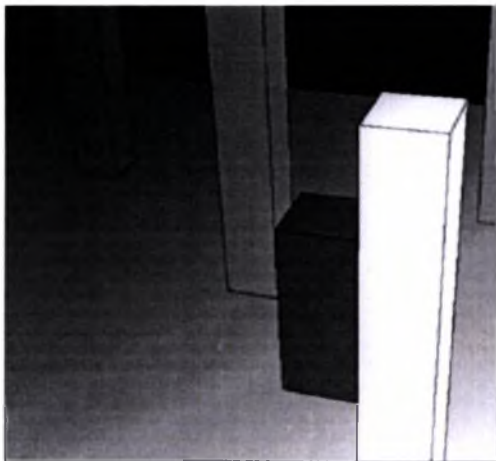


(f)

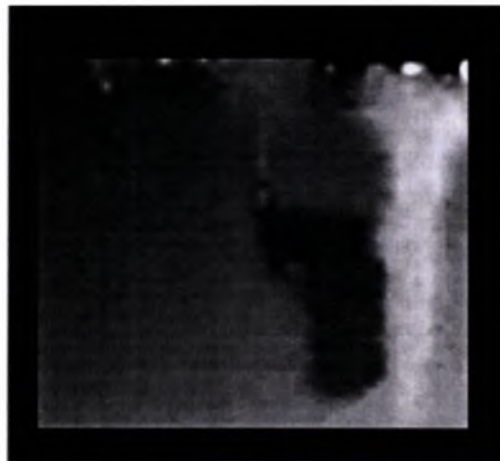
**Σχήμα 3 (d)-(f) . Υπολογισμός πεδίου οπτικής ροής με χρήση της φάσης.**



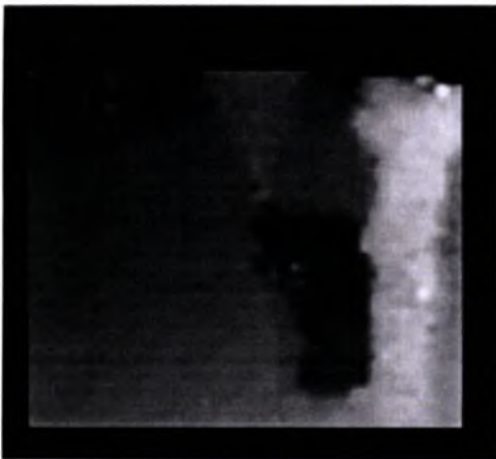
Το Σχήμα 4 (a) δείχνει την πραγματική διάσταση (magnitude) της οπτικής ροής που εξάγεται από τα παρακάτω ground-truth δεδομένα, ενώ στα σχήματα 4 (b)-(d) παρουσιάζεται το μέγεθος του πεδίου της οπτικής ροής υπολογισμένο με την χρήση της φάσης.



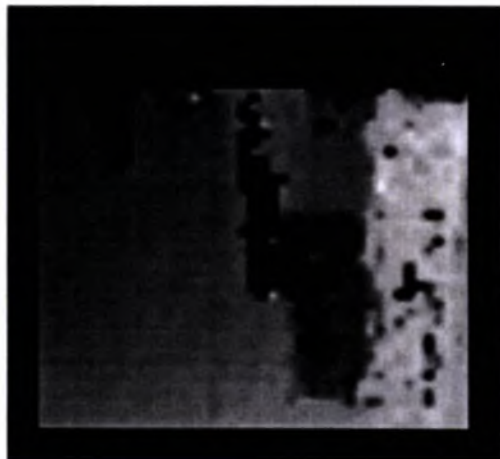
(a)



(b)



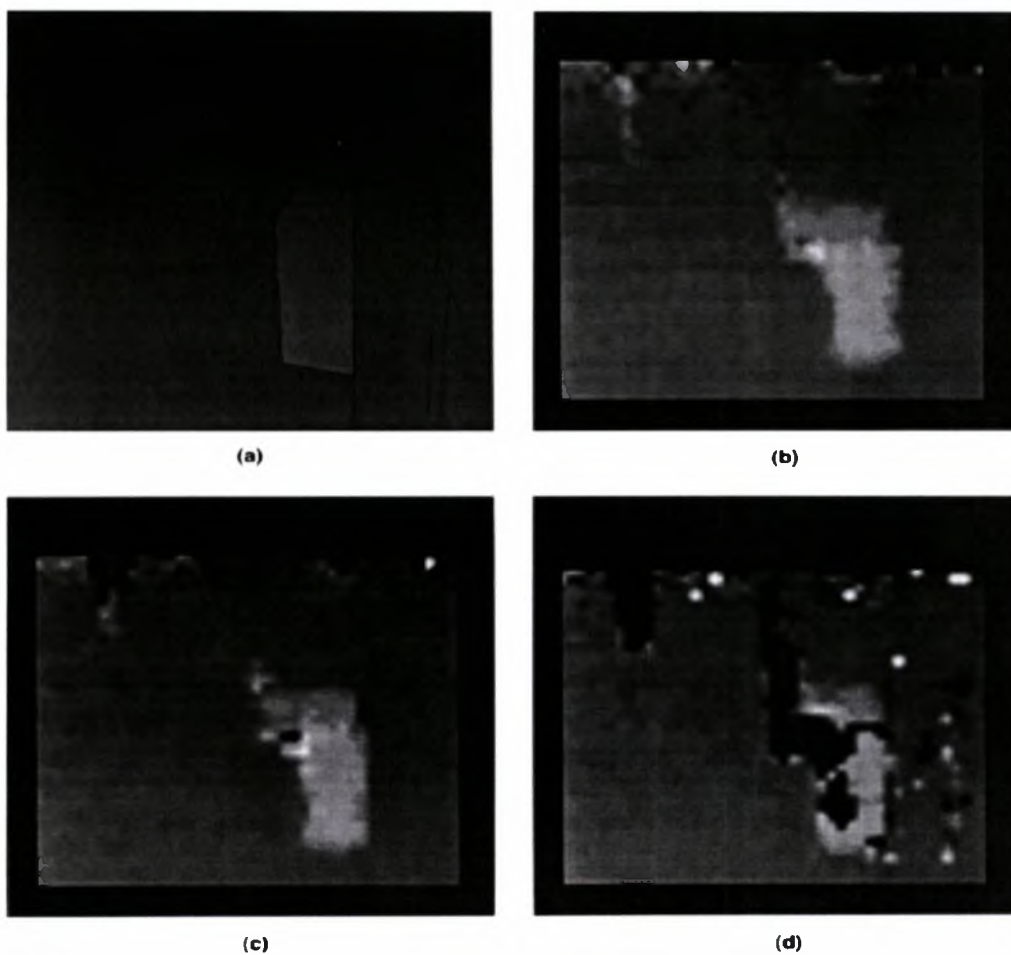
(c)



(d)

*Σχήμα 4.*

Στο ακόλουθο σχήμα 5(a) παρουσιάζεται η πραγματική διεύθυνση (real direction) του πεδίου οπτικής ροής που εξάγεται από τα ground-truth δεδομένα, ενώ στα σχήματα 5(b)-(d) παρουσιάζεται η διεύθυνση του υπολογισμένου πεδίου οπτικής ροής με χρήση της φάσης.



**Σχήμα 5.**

## Κεφάλαιο 7      Συμπεράσματα

Στην παρούσα πτυχιακή, παρουσιάσαμε μια υβριδική προσέγγιση ελαχιστοποίησης σφάλματος (error minimization) για την εκτίμηση της κίνησης και την κατάτμηση. Δείξαμε ότι το σφάλμα στο πεδίο της συχνότητας (FT error) είναι πολύ σταθερό στην παρουσία θορύβου και πρέπει να χρησιμοποιείται σε εφαρμογές που απαιτούν τον υπολογισμό της κίνησης, ενώ το σφάλμα στο πεδίο του χώρου (spatial error) μπορεί να χρησιμοποιηθεί για να παρουσιάσει την κατάτμηση αντικειμένων (object segmentation). Αναφερθήκαμε στο πρόβλημα της εκτίμησης της κίνησης για μία ακολουθία εικόνων σε βίντεο, και αναλύσαμε τις κλασσικές τεχνικές για τον υπολογισμό της. Υπολογίσαμε την κίνηση σε βίντεο χρησιμοποιώντας τον μετασχηματισμό Fourier και συγκεκριμένα της φάσης του παραθέτοντας τα αντίστοιχα πειράματα. Τα πειράματα για την εκτίμηση της κίνησης και την κατάτμηση των αντικειμένων, σε περίπτωση απουσίας και παρουσίας θορύβου, καταδεικνύουν την αποδοτικότητα και την σταθερότητα της προτεινόμενης μεθόδου μας, ακόμα και όταν η ακολουθία είναι κάτω από την επιρροή του θορύβου.

## Κεφάλαιο 8 Βιβλιογραφία

- [1] J. Barron, D. Fleet, and S. Beauchemin. Performance of optical flow techniques. *International Journal of Computer Vision*, 12(1):43–77, Dec. 1994
- [2] J. Y. A. Wang and E. H. Adelson. Representing moving images with layers. *IEEE Transactions on Image Processing*, 3:623–638, Sept 1994
- [3] A. Briassouli and N. Ahuja. Fusion of frequency and spatial domain information for motion analysis. In *ICPR 2004, Proceedings of the 17th International Conference on Pattern Recognition*, volume 2, pages 175–178, Aug. 2004.
- [4] A. Oppenheim and R. Schaffer. *Digital Signal Processing*. Prentice-Hall, Inc., Englewood Cliffs, NJ, 1975.
- [5] R. C. Gonzalez and R. E. Woods. *Digital Image Processing*. Addison-Wesley Longman Publishing Co., Boston, MA, 2001.
- [6] Christoph Stiller and Janusz Konrad. Estimating motion in image sequences. *A tutorial on modeling and computation of 2D motion. IEEE Signal Processing Magazine* 1053-5888/98/\$10.00, 1998 IEEE
- [7] J.L. Barron and D.J. Fleet and S.S. Beauchemin. Performance of Optical Flow Techniques. *International Journal of Computer Vision*, 12:1, 43-77 (1994)
- [8] M.P. Cagigal, L. Vega, P. Prieto, Object movement characterization from low-light-level images, *Optical Engineering* 33 (8) (1994) 2810–2812
- [9] V. Argyriou and T. Vlachos Centre for Vision, Speech and Signal Processing University of Surrey Guildford GU2 7XH, United Kingdom. A study of sub-pixel motion estimation using phase correlation.
- [10] M. Otte and H.-H. Nagel, Optical flow estimation: advances and comparisons, in: J.O. Eklundh (ed.), *Lecture Notes in Computer Science, Computer Vision—ECCV '94*, Springer, Berlin, 1994, pp. 51–60.
- [11] D.J. Fleet, A.D. Jepson, Computation of component image velocity from local phase information, *International Journal of Computer Vision* 5 (1990) 77–104.

- [12] D.J. Fleet, A.D. Jepson, Stability in phase information, *IEEE Trans. Pattern Analysis and Machine Intelligence* 15 (12) (1993) 1253–1268.
- [13] J. Watkinson, *The Engineer's Guide to Motion Compensation*, Snell & Wilcox, 1994
- [14] A. Lundmark, H. Li, and R. Forchheimer, "Motion vector certainty reduces bit rate in backward motion estimation video coding," in *SPIE VCIP*, Jun. 2000.
- [15] H. H. Nagel, *Displacement vectors derived from second-order intensity variations in image sequences*, *CGIP* (1983), no. 21, 85–117.
- [16] P. Anandan and Michael J. Black, *The robust estimation of multiple motions: Parametric and piecewise smooth flow fields*, *Computer Vision and Image Understanding* 63 (1996), no. 1, 75–104.
- [17] Batali, J. and Ullman, S., Motion detection and analysis, *Proc. of the ARPA Image Understanding Workshop*, 7-8 November 1979 (Science Applications Inc., Arlington, VA 1979) pp.69-75.
- [18] E.H. Adelson and J.R. Bergen, "Spatiotemporal energy models for the perception of motion," *J. Optical Soc. America*, vol. 2, no. 2, pp. 284-299, 1985.
- [19] A Del Bimbo, P. Nesi and J.L. C. *Sanz*, "Analysis of optical flow constraints," *IEEE Trans. on Image Processing*, vol. 4, no. 4, pp. 460- 469, 1995.
- [20] Foroosh H, Balci M: Sub-Pixel Registration and Estimation of Local Shifts Directly in the Fourier Domain. *Proc IEEE Conf Image Processing ICIP-2004*, 1955-1958, 2004.
- [21] Reddy BS, Chatterji BN: An FFT-Based Technique for Translation, Rotation, and Scale-Invariant Image Registration. *IEEE Trans Image Processing*, 5(8): 1266-1271, 1996.
- [22] ISO/IEC Intl. Std. 15444, Information technology – JPEG 2000 image coding system, particularly Part 3: Motion JPEG 2000 (Sept. 2002, with subsequent amendments).
- [23] M. Irani, P. Anandan, J. Bergen, R. Kumar, and S. Hsu. Efficient representations of video sequences and their applications. *Signal Processing: Image Communication, special issue on Image and Video Semantics: Processing, Analysis, and Application*, 8(4), May 1996.
- [24] Chellapa, R., Quian, G, and Srinivasan, S. "Structure from motion: sparse versus dense correspondence methods". *IEEE Proceedings 1999 International Conference on Image Processing*, Vol 2, pp:492-499, 1999.

- [25] Phil Tudor, Progressive transmission of interlaced pictures. BBC Research & Development Technical Memorandum No. T-1231, August 1993.
- [26] Thomas, G.A, Application of motion compensation to video processing in the TV studio, IEE Colloquium on video processing in TV studio, 1991 pg 5/1-5/4.
- [27] A.J Patti, M.I Sezan and A.M Tekalp, Robust methods for High-quality stills from interlaced video in the presence of dominant motion, IEEE Tr on Circuits and Systems for video technology, Vol7 , No 2 , April 1997 pg 328-342.



ΠΑΝΕΠΙΣΤΗΜΙΟ  
ΘΕΣΣΑΛΙΑΣ



004000091573