



ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ

Τμήμα Μηχανικών Ηλεκτρονικών Υπολογιστών Τηλεπικοινωνιών και Δικτύων

Διπλωματική εργασία

# Εκτίμηση και Κατάτμηση Κίνησης για Βίντεο με Διαφάνεια

Ξαυδοπούλου Παρθένα

Επιβλέποντες Καθηγητές :

Χούστης Ηλίας

Μπριασούλη Αλεξία

Βόλος 2008



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

Αριθ. Εισ.: 6569/1  
Ημερ. Εισ.: 03-10-2008  
Δωρεά: Συγγραφέα  
Ταξιθετικός Κωδικός: ΠΤ - ΜΗΥΤΔ  
2008  
ΞΑΝ

## Ευχαριστίες

Θέλω να ευχαριστήσω τους επιβλέποντες καθηγητές μου, τον κύριο Ηλία Χούστη και την κυρία Αλεξία Μπριασούλη, για την πολύτιμη βοήθεια τους για την ολοκλήρωση της πτυχιακής εργασίας.

Επίσης, θα ήθελα να ευχαριστήσω την οικογένεια μου και τους φίλους μου που είναι πλάι μου και με στηρίζουν. Τέλος, θέλω να ευχαριστήσω τον αδερφό μου Στάθη που με βοήθησε στην υλοποίηση της παρουσιάσεως της εργασίας.



# Περιεχόμενα

Σελ.

|   |    |
|---|----|
| <b>1. Εισαγωγή</b> .....  | 5  |
| <b>2. Εκτίμηση κίνησης</b> .....  | 8  |
| 2.1. Μοντέλα κίνησης.....   | 9  |
| 2.2. Κριτήρια εκτίμησης.....  | 10 |
| 2.3. Μέθοδοι εκτίμησης κίνησης στο χώρο.....  | 11 |
| 2.3.1. Ταίριασμα μπλοκ (block matching).....  | 11 |
| 2.3.2. Οπτική ροή (optical flow).....   | 13 |
| 2.3.3. Μέθοδοι υπολογισμού οπτικής ροής.....  | 17 |
| 2.3.3.1. Διαφορικές τεχνικές.....   | 17 |
| 2.3.3.2. Ταίριασμα βασισμένο στην περιοχή (region based matching).....  | 19 |
| 2.3.3.3. Μέθοδοι βασισμένες στην ενέργεια.....  | 20 |
| 2.4. Μέθοδοι εκτίμησης κίνησης στη συχνότητα.....   | 20 |
| 2.4.1. Συσχέτιση φάσης (phase correlation).....   | 21 |
| <b>3. Διαφάνεια</b> .....   | 23 |
| 3.1. Τυπικά χαρακτηριστικά της διαφάνειας.....  | 25 |
| 3.2. Διαφανής κίνηση.....   | 28 |
| <b>4. Μέθοδοι εκτίμησης διαφανούς κίνησης</b> .....   | 32 |
| 4.1. Διαχωρισμός διαφανών επιπέδων με τη δια μέσου ανταλλαγής πληροφορίας ανάμεσα στα επίπεδα (separating transparent layers through layer information exchange)..... | 32 |
| 4.2. Υπολογισμός «φρασσόμενων» και διαφανών κινήσεων (computing occluding and transparent motions).....   | 34 |
| <b>5. Εκτίμηση και κατάτμηση διαφανούς κίνησης</b> .....  | 37 |
| 5.1. Εκτίμηση μετατόπισης σε χώρο Fourier.....  | 37 |
| 5.2. Μια ελαχίστων τετραγώνων κατάτμηση για μετατόπιση στο χώρο συχνοτήτων.....   | 40 |
| 5.2.1. Συστηματοποίηση για την ελαχίστων τετραγώνων κατάτμηση κίνησης..   | 43 |
| 5.3. Ολοκλήρωση με χωρικούς υπολογισμούς.....   | 44 |
| 5.3.1. Συσχετισμός της λύσης ελαχίστων τετραγώνων και της αρχικής.....  | 44 |
| 5.3.2. Περιοχές δραστηριότητας (activity areas).....  | 45 |
| 5.4. Περιστρεφόμενες και μετατοπιζόμενες κινήσεις.....  | 47 |
| 5.4.1. Ένα αντικείμενο.....   | 48 |
| 5.4.2. Πολλαπλά αντικείμενα.....  | 49 |

|  |    |
|--|----|
| <b>6. Πειράματα</b> .....                        | 52 |
| 6.1. Συνθετική ακολουθία 1 .....                 | 52 |
| 6.2. Συνθετική ακολουθία 2 .....                 | 53 |
| 6.3. Συνθετική ακολουθία 3 .....                 | 54 |
| 6.4. Χώρος στάθμευσης αυτοκινήτων .....          | 55 |
| 6.5. Περίπατος .....                             | 56 |
| 6.6. Περίπατος σε ελαφρώς κεκλιμένο επίπεδο..... | 57 |
| 6.7. Είσοδος κτιρίου .....                       | 58 |
| 6.8. Ταξί.....                                   | 59 |
| <br>   |    |
| <b>7. Συμπεράσματα</b> .....                     | 61 |
| <br>   |    |
| <b>Παράρτημα</b> .....                           | 62 |
| <br>   |    |
| <b>Βιβλιογραφία</b> .....                        | 63 |

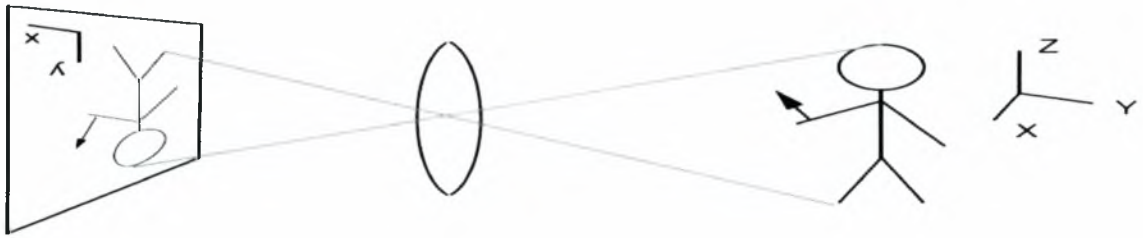
# 1. Εισαγωγή

Η επεξεργασία εικόνων και βίντεο είναι ένα πεδίο που συνεχίζει να μεγαλώνει, με νέες εφαρμογές να αναπτύσσονται συνεχώς με αλματώδη βήματα. Είναι μια ενδιαφέρουσα περιοχή με ποικίλες εφαρμογές, όπως η ψυχαγωγία, η ασφάλεια (surveillance), οι επικοινωνίες. Το internet, με την ευκολία χρησιμοποίησης του μέσω των παγκόσμιων φυλλομετρητών, συνδυασμένο με τις προόδους στα χαρακτηριστικά των υπολογιστών και του εύρους ζώνης του δικτύου έχει φέρει τον κόσμο στα γραφεία μας και μέσα στα ίδια τα σπίτια μας. Μία από τις πιο ενδιαφέρουσες πτυχές αυτής της πληροφοριακής επανάστασης είναι η ικανότητα που σου δίνουν να στέλνεις και να λαμβάνεις σύνθετα δεδομένα τα οποία ξεπερνούν το συνηθισμένο γραπτό κείμενο. Η οπτική πληροφορία, μεταδιδόμενη στη μορφή των ψηφιακών εικόνων, έχει γίνει η κυριότερη μέθοδος επικοινωνίας στον 21<sup>ο</sup> αιώνα.

Η αναπαράσταση εικόνων από των υπολογιστή (computer imaging) μπορεί να οριστεί ως η απόκτηση και επεξεργασία οπτικής πληροφορίας από τον υπολογιστή. Η σημασία του computer imaging προέρχεται από το γεγονός ότι στην καθημερινή μας ζωή χρησιμοποιούμε την όραση ως μία από τις κυριότερες πηγές απόκτησης πληροφορίας για τον έξω κόσμο. Παρατηρώντας τις αλλαγές μιας σκηνής με το χρόνο, μπορούμε να ανακαλύψουμε την τρισδιάστατη δομή της σκηνής, να κάνουμε προβλέψεις για πιθανές συγκρούσεις, να συμπεράνουμε τις υλικές ιδιότητες των αντικειμένων, όπως η ακαμψία και η διαφάνεια. Αρκετή από αυτήν την πληροφορία αποκαλύπτεται από την κίνηση διαφόρων μερών της σκηνής. Μια εικόνα αξίζει όσο χίλιες λέξεις. Η αναπαράσταση μιας εικόνας από τον υπολογιστή απαιτεί το ισοδύναμο πολλών χιλιάδων λέξεων από δεδομένα και χωρίς ένα αντίστοιχο μέγεθος πληροφοριών τα μέσο επικοινωνίας θα ήταν απαγορευτικά ανεπαρκές. Η τεράστια ποσότητα δεδομένων που απαιτείται για τις εικόνες είναι ο πρωταρχικός λόγος για την ανάπτυξη πολλών υποπεριοχών μέσα στο πεδίο του computer imaging, όπως η κατάτμηση εικόνας και η συμπίεση εικόνας.

Οι εικόνες συνιστούν την οπτική προβολή του τρισδιάστατου κόσμου πάνω σε δισδιάστατη επιφάνεια αντίληψης του φωτός. Αυτή η επιφάνεια θα μπορούσε να είναι, για παράδειγμα, ένα κομμάτι από φιλμ, ή μια σειρά από αισθητήρες φωτός σε μια κάμερα τηλεόρασης. Η φωτεινότητα της εικόνας σε κάθε σημείο δηλώνει πόσο φως πέφτει πάνω στη επιφάνεια σε αυτή τη θέση του χώρου σε μια συγκεκριμένη στιγμή. Όταν ένα αντικείμενο κινείται κατά μήκος αυτής της επιφάνειας προβολής, η δισδιάστατη προβολή του αντικειμένου κινείται μέσα στην εικόνα. Η κίνηση της προβολής κάθε σημείου στον κόσμο αναφέρεται ως ταχύτητα της εικόνας ή πεδίο κίνησης (motion field). Αυτό απεικονίζεται στην Εικόνα 1:





**Εικόνα 1:** Η κίνηση της προβολής κάθε σημείου στον κόσμο αναφέρεται ως ταχύτητα της εικόνας ή πεδίο κίνησης.

Η αύξηση της μνήμης των υπολογιστών και της ταχύτητας των επεξεργαστών έχει προκαλέσει την επέκταση των αυτοματοποιημένων συστημάτων για τις ανάγκες της καθημερινής μας ζωής. Τα συστήματα που επεξεργάζονται εικόνες και βίντεο για να εξάγουν πληροφορίες για τον κόσμο είναι ένα παράδειγμα τέτοιων συστημάτων υπολογισμών, των οποίων η χρήση γίνεται πιο εφικτή με τις προόδους στη δύναμη των υπολογιστών. Ο τομέας αυτών των συστημάτων στοχεύει στην επεξεργασία οπτικής πληροφορίας, η οποία γενικά χαρακτηρίζεται από μεγάλο μέγεθος και σύνθετη δομή. Άλλος λόγος που εμποδίζει την ανάπτυξη αυτών των συστημάτων είναι η έλλειψη πλήρους κατανόησης του τρόπου που τα ανθρώπινα όντα επεξεργάζονται την οπτική πληροφορία. Εντούτοις, τα συστήματα που επεξεργάζονται εικόνες και βίντεο για να εξάγουν πληροφορίες για τον κόσμο έχουν ενσωματωθεί σε πολλές πραγματικές εφαρμογές, όπως συστήματα παρακολούθησης, πλοήγηση ρομπότ, στην ιατρική και συστήματα επαλήθευσης ταυτότητας.

Ο εντοπισμός αντικείμενων είναι ένα βασικό θέμα στα συστήματα που επεξεργάζονται εικόνες και βίντεο, το οποίο στοχεύει στην ανίχνευση της θέσης ενός κινούμενου αντικείμενου από μια ακολουθία βίντεο. Η ανίχνευση κίνησης αναπτύχθηκε αρχικά για την κωδικοποίηση βίντεο. Τα διανύσματα κίνησης μπορούν να χρησιμοποιηθούν για να προβλέψουν αλλαγές στη σκηνή μεταξύ δύο ή περισσότερων πλαισίων βίντεο. Το μέγεθος των δεδομένων του βίντεο μειώνεται με την κωδικοποίηση μόνο του τρέχοντος πλαισίου και των διανυσμάτων κίνησης, από τα οποία πολλά μελλοντικά πλαίσια μπορούν να ανακτηθούν. Η ανίχνευση της κίνησης για αντικείμενα που παρακολουθούνται έχει διαφορετικές απαιτήσεις. Στην παρακολούθηση ενός αντικείμενου, υπάρχει μια ανάγκη για ερμηνεία της πληροφορίας που δίνεται από τα διανύσματα κίνησης. Πρέπει επίσης να ανιχνευθούν μόνο τα κινούμενα αντικείμενα και να φιλτραριστεί ο θόρυβος, οι κινήσεις της κάμερας, και άλλες μικρές κινήσεις (φύλλα που πέφτουν, σταγόνες της βροχής). Η ομαδοποίηση ή κάποιου άλλου είδους συσχέτισης των διανυσμάτων κίνησης σε μια εποικοδομητική αναπαράσταση σκηνής είναι ο κύριος στόχος της ανίχνευσης κίνησης για ένα αντικείμενο που παρακολουθείται.

Τυπικά προβλήματα σε αλγορίθμους ανάλυσης στιγμιότυπων σκηνής είναι: (1) Θόρυβος με χαρακτηριστικά γνωρίσματα παρόμοια με αυτά των αντικείμενων που παρακολουθούνται (παραδείγματος χάρη το χιόνι που πέφτει). (2) Αντικείμενα που μοιάζουν είτε ως φόντο ή ως θόρυβος. Αυτό μπορεί να συμβεί όταν ένα αντικείμενο μπαίνει σε μια σκιασμένη περιοχή ή σε ένα σύννεφο ομίχλης. (3) Δύο ή



περισσότερα αντικείμενα, τα οποία εμφανίζονται ως ενιαίο αντικείμενο και αντίστροφα.

Στην παρούσα διπλωματική εργασία θα ασχοληθούμε με την εκτίμηση και κατάτμηση κίνησης για βίντεο με διαφάνεια. Παρουσιάζουμε μια πρωτότυπη μέθοδο στο χώρο του Fourier για την εκτίμηση της κίνησης, που βασίζεται σε τεχνικές συσχέτισης φάσης (phase correlation), οι οποίες έχουν αποδειχτεί ότι δίνουν εύρωστα αποτελέσματα για την εγγραφή της εικόνας. Αλλά ενώ η συσχέτιση φάσης για την εγγραφή της εικόνας εξετάζει μόνο ένα κινούμενο αντικείμενο, η μέθοδός μας έχει σχεδιαστεί για την εκτίμηση πολλών μετατοπιζόμενων και περιστρεφόμενων κινήσεων με διαφάνεια. Επίσης, σε αντίθεση με τις υπάρχουσες μεθόδους του χώρου Fourier για πολλαπλή εκτίμηση κίνησης, που περιορίζονται στην περίπτωση των μετατοπίσεων, προτείνουμε ένα σύστημα που υπολογίζει και μετατοπιζόμενες και περιστρεφόμενες-μετατοπιζόμενες κινήσεις. Η διατύπωση του προβλήματος μας επίσης επιτρέπει την εκτίμηση του αριθμού των μετακινούμενων αντικειμένων. Για την περίπτωση των πολλαπλών μετατοπίσεων, παρέχουμε μία πρωτότυπη, κομψή μέθοδο για την ταυτόχρονη κατάτμηση του φόντο και των μετακινούμενων αντικειμένων, που επιτυγχάνεται λύνοντας ένα γραμμικό σύστημα στο πνεύμα των ελαχίστων τετραγώνων.

## 2. Εκτίμηση κίνησης

Η κίνηση είναι μια κατάσταση, η οποία δείχνει την αλλαγή της θέσης. Η κίνηση είναι μια κυρίαρχη πηγή χρονικών διακυμάνσεων σε ακολουθίες εικόνων. Για να μοντελοποιηθεί και να υπολογιστεί η κίνηση, απαιτείται μια κατανόηση για το πώς οι εικόνες (και συνεπώς η κίνηση στην εικόνα) σχηματίζονται. Συνεπώς, οι παράμετροι της κάμερας, όπως η τρισδιάστατη κίνηση (περιστροφή, μετατόπιση) ή εστιακό μήκος, παίζουν σημαντικό ρόλο στην μοντελοποίηση της κίνησης στην εικόνα. Αν αυτές οι παράμετροι είναι γνωστές, μόνο η κίνηση του αντικειμένου χρειάζεται να βρεθεί. Παρόλα αυτά, αυτό το σενάριο είναι μάλλον απίθανο, και η κίνηση του αντικειμένου και η κίνηση της κάμερας πρέπει να υπολογιστούν. Στις πραγματικές σκηνές βίντεο, η κίνηση μπορεί να είναι ένας σύνθετος συνδυασμός μετατόπισης και περιστροφής. Τέτοια κίνηση είναι δύσκολο να υπολογιστεί και μπορεί να απαιτήσει πολύ επεξεργασία.

Η εκτίμηση κίνησης είναι η διαδικασία εντοπισμού των διανυσμάτων κίνησης που περιγράφουν το μετασχηματισμό από μια δισδιάστατη εικόνα σε μια άλλη, συνήθως από συνεχόμενα πλαίσια (frames) σε μια ακολουθία βίντεο. Είναι ένα άσχημα τεθειμένο πρόβλημα δεδομένου ότι η κίνηση είναι σε τρεις διαστάσεις ενώ οι εικόνες είναι μια προβολή του τρισδιάστατου χώρου σε δισδιάστατο επίπεδο. Τα διανύσματα κίνησης μπορεί να σχετίζονται με ολόκληρη την εικόνα (γενική εκτίμηση κίνησης-global motion estimation) ή με συγκεκριμένα μέρη, όπως ορθογώνια μπλοκ, αυθαίρετα σχηματισμένες περιοχές, ή ακόμα και ανά εικονοκύτταρο (pixel). Τα διανύσματα κίνησης μπορούν να αναπαρασταθούν από ένα μετατοπιζόμενο μοντέλο ή από πολλά άλλα μοντέλα που μπορούν να προσεγγίσουν την κίνηση μιας πραγματικής βιντεοκάμερας, όπως η περιστροφή και η μετατόπιση και στις τρεις διαστάσεις.

Για τον υπολογισμό των τροχιών της κίνησης, χρειάζεται τρία βασικά στοιχεία να διευκρινιστούν. Πρώτα, τα ελλοχεύοντα μοντέλα πρέπει να επιλεγούν, παραδείγματος χάριν το μοντέλο κίνησης (περιγραφή, περιοχή υποστήριξης), το μοντέλο σχετιζόμενης κίνησης και δεδομένα εικόνας (μοντέλο παρατήρησης), το μοντέλο ορίου κίνησης, μοντέλο φραξίματος. Δεύτερον, ένα κριτήριο εκτίμησης πρέπει να προσδιοριστεί. Ένα τέτοιο κριτήριο μπορεί να έχει διαφορετικές μορφές, όπως ένα απλό μέσο τετραγωνικό σφάλμα σε ένα μπλοκ, ένα εύρωστο κριτήριο (με κορεσμό για μεγάλα σφάλματα). Τρίτον, μια στρατηγική αναζήτησης πρέπει να εφαρμοστεί για να διευκρινιστούν οι παράμετροι κίνησης που βελτιστοποιούν το επιλεγόμενο κριτήριο.

Οι περισσότεροι από τους αλγορίθμους εκτίμησης κίνησης κάνουν τις ακόλουθες υποθέσεις:

1. Τα αντικείμενα κινούνται κάνοντας μετατοπίσεις σε ένα επίπεδο που είναι παράλληλο με το επίπεδο της κάμερας, δηλαδή τα αποτελέσματα του ζουμ της κάμερας, και οι περιστροφές των αντικειμένων δεν λαμβάνονται υπόψη.
2. Ο φωτισμός είναι χωρικά και χρονικά ομοιόμορφος.

## 2.1 Μοντέλα κίνησης

Το απλούστερο μοντέλο κίνησης είναι να υποθέσουμε ότι η ταχύτητα είναι σταθερή σε όλη την περιοχή

$$\begin{aligned}u_x(x, y) &= a \\u_y(x, y) &= b\end{aligned}$$

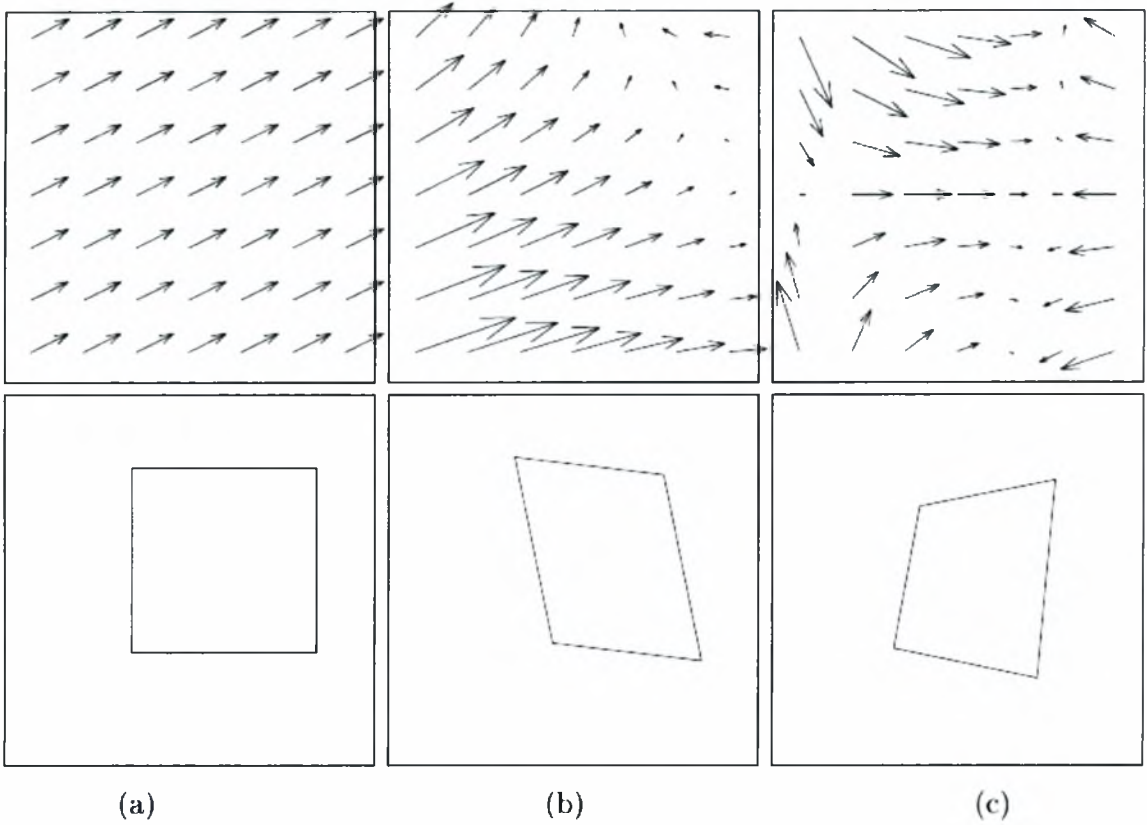
Όπου  $x$  και  $y$  είναι οι χωρικές συντεταγμένες,  $u_x$  και  $u_y$  είναι οι συνιστώσες  $x$  και  $y$  της ταχύτητας,  $a$  και  $b$  είναι οι παράμετροι του μοντέλου. Γεωμετρικά, αυτό το μοντέλο κίνησης αντιστοιχεί σε αντικείμενα που υποβάλλονται σε μια καθαρή μετατόπιση σε ορθογραφική προβολή. Μια ισχυρότερη εναλλακτική λύση είναι το σχετικό (affine) μοντέλο κίνησης

$$\begin{aligned}u_x(x, y) &= ax + by + c \\u_y(x, y) &= dx + ey + f\end{aligned}$$

Το οποίο εφαρμόζεται σε επίπεδες μικρές περιοχές που υποβάλλονται σε «άκαμπτες» κινήσεις σωμάτων, δηλαδή μετατόπιση και περιστροφή, κάτω από ορθογραφική προβολή. Για μια εναλλακτική προβολή (perspective projection) χρειάζονται οχτώ παράμετροι κίνησης

$$\begin{aligned}u_x(x, y) &= a_1 + a_2x + a_3y + a_7x^2 + a_8xy \\u_y(x, y) &= a_4 + a_5x + a_6y + a_7xy + a_8y^2\end{aligned}$$

Η χρησιμότητα αυτών των μοντέλων εξαρτάται φυσικά από την εκάστοτε εφαρμογή, αλλά είναι χρήσιμο να παρατηρηθεί ότι πολλές επιφάνειες μπορούν να προσεγγιστούν ως επίπεδες και αν η απόσταση στη σκηνή είναι μεγαλύτερη από ότι η διακύμανση στην απόσταση μέσα στην σκηνή, τότε η εναλλακτική προβολή (perspective projection) μπορεί να προσεγγιστεί από την ορθογραφική προβολή.



**Εικόνα 2:** Παραδείγματα παραμετρικών διανυσματικών πεδίων κίνησης και αντίστοιχες αντιστάθμισης κίνησης προβλέψεις για ένα κεντροθετημένο τετράγωνο: (a) μετατοπιζόμενο, (b) affine, (c) projective linear.

### Τοπική (global) και τοπική (local) κίνηση

Η γενική κίνηση (global) είναι η κίνηση σε μια μοναδική κατεύθυνση που προκύπτει από ένα ερέθισμα του οποίου τα στοιχεία κινούνται σε μια ποικιλία διαφορετικών κατευθύνσεων.

Η τοπική κίνηση (local) είναι η κίνηση ενός τυχαίου τομέα σημείων που αποτελείται από πολλές κινήσεις σε πολλές κατευθύνσεις και πολλές εκτάσεις.

## 2.2 Κριτήρια εκτίμησης

Ποικίλες αναπαραστάσεις κίνησης μπορούν να χρησιμοποιηθούν για να διατυπώσουν ένα κριτήριο αξιολόγησης. Δεν υπάρχει ωστόσο ένα μοναδικό κριτήριο για την εκτίμηση της κίνησης. Η δυσκολία στη δημιουργία ενός κριτηρίου προκαλείται κυρίως από το γεγονός ότι η κίνηση στις εικόνες δεν είναι άμεσα παρατηρήσιμη και ότι μπορεί να υπάρχουν παραπάνω από μία κινήσεις στην εικόνα. Μερικά από τα κριτήρια που μπορούν να χρησιμοποιηθούν είναι τα εξής:

## Μέσο τετραγωνικό σφάλμα (MSE)

Όταν ο εκτιμητής  $\hat{\theta}$  είναι άγνωστος, το μέσο τετραγωνικό σφάλμα είναι μια τυχαία μεταβλητή της οποίας η τιμή πρέπει να εκτιμηθεί. Αυτό γίνεται από το μέσο των δειγμάτων:

$$\text{MSE}(\hat{\theta}) = \frac{1}{n} \sum_{j=1}^n (\theta_j - \theta)^2 \quad (1)$$

όπου  $\theta_j$  είναι υλοποιήσεις του εκτιμητή  $\hat{\theta}$  μεγέθους  $n$ .

## Άθροισμα απόλυτων διαφορών (SAD)

Το άθροισμα απόλυτων διαφορών είναι μια ευρέως χρησιμοποιούμενη, πολύ απλή μετρική για βίντεο που χρησιμοποιείται στο ταίριασμα μπλοκ (block-matching) στην εκτίμηση κίνησης σε βίντεο. Λειτουργεί με τη λήψη της απόλυτης τιμής της διαφοράς μεταξύ κάθε εικονοκυττάρου στο αρχικό μπλοκ και του αντίστοιχου εικονοκυττάρου στο μπλοκ που χρησιμοποιείται για σύγκριση. Αυτές οι διαφορές αθροίζονται και δημιουργούν μια απλή μετρική ομοιότητας μπλοκ, την νόρμα  $\|L\|_1 = \sum_{i=1}^n L_i$ , της εικόνα διαφοράς.

## 2.3 Μέθοδοι εκτίμησης κίνησης στο χώρο

Οι περισσότερες κοινές μέθοδοι του χωρικού τομέα περιλαμβάνουν αυτές που βασίζονται στην ελαχιστοποίηση λάθους της μετρικής του αθροίσματος απόλυτων διαφορών (sum of absolute difference, SAD) ή του αθροίσματος της τετραγωνικής διαφοράς (sum of squared difference, SSD). Η ελαχιστοποίηση του αθροίσματος της τετραγωνικής διαφοράς, SSD, επιτυγχάνεται με μεθόδους καθόδου κλίσης (gradient descent methods). Δεδομένου ότι αυτή η μέθοδος είναι επαναληπτική και κάθε επανάληψη περιλαμβάνει την στρέβλωση της εικόνας και τον υπολογισμό των παραγώγων, είναι υπολογιστικά πολύ απαιτητική και αργή. Η ελαχιστοποίηση της μετρικής λάθους SAD επιτυγχάνεται με άμεση αναζήτηση των παραμέτρων του χώρου. Αλλά η πολυπλοκότητα της αναζήτησης αυξάνει εκθετικά με τον αριθμό των παραμέτρων. Αυτές οι τεχνικές έχουν το μειονέκτημα ότι μπορεί να κολλήσουν σε τοπικά ελάχιστα, αν και είναι λιγότερο πιθανό σε πολλαπλά πλαίσια ευκρίνειας (multi resolution framework).

Για την εκτίμηση κίνησης στο χώρο χρησιμοποιούνται μέθοδοι που βασίζονται στο ταίριασμα μπλοκ (block matching) και μέθοδοι οπτικής ροής.

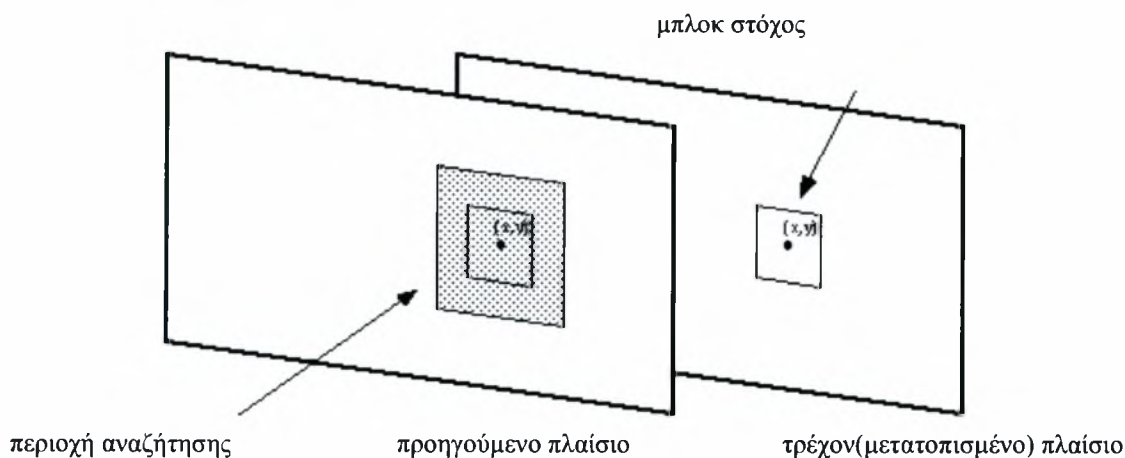
### 2.3.1 Ταίριασμα μπλοκ (block matching)

Ο αλγόριθμος ταίριασματος μπλοκ είναι μια τυποποιημένη τεχνική για την κωδικοποίηση της κίνησης σε μια ακολουθία από βίντεο. Στοχεύει στην ανίχνευση



της κίνησης μεταξύ δύο εικόνων, που θα περιγράφονται σε μια μορφή μπλοκ. Τα μπλοκ καθορίζονται συνήθως με την διαίρεση του πλαισίου της εικόνας σε μη επικαλυπτόμενα τετραγωνικά μέρη. Κάθε μπλοκ από το τρέχον πλαίσιο συγκρίνεται με ένα μπλοκ στο πλαίσιο προορισμού μετατοπίζοντας το τρέχον μπλοκ πέρα από μια προκαθορισμένη γειτονιά εικονοκυττάρων (pixels) στο πλαίσιο προορισμού. Σε κάθε μετατόπιση, το άθροισμα των αποστάσεων μεταξύ των γκρίζων τιμών των δύο μπλοκ υπολογίζεται. Η μετατόπιση η οποία δίνει τη μικρότερη συνολική απόσταση θεωρείται η καλύτερη σύγκριση (match). Ιδανικά, δύο συγκρινόμενα μπλοκ έχουν τα αντίστοιχα εικονοκύτταρα τους ακριβώς ίσα. Αυτό ισχύει σπάνια γιατί τα κινούμενα αντικείμενα αλλάζουν το σχήμα τους, όπως το βλέπει ένας εξωτερικός παρατηρητής, το φως που αντικατοπτρίζεται από την επιφάνεια του αντικειμένου επίσης αλλάζει, και ακόμα στον αληθινό κόσμο πάντα υπάρχει θόρυβος. Επιπλέον, σε σκηνές που περιέχουν κίνηση υπάρχουν φραξίματα (occlusion) μεταξύ των αντικειμένων, εξαφανίσεις αντικειμένων, εμφανίσεις νέων. Παρά τα προβλήματα της εικονοκύτταρο με εικονοκύτταρο ομοιότητας, είναι γρήγορο να υπολογιστεί και χρησιμοποιείται ευρέως για την αναζήτηση περιοχών σύγκρισης.

Κατά τη διάρκεια της σύγκρισης των μπλοκ, κάθε μπλοκ στόχος του τρέχοντος πλαισίου (δηλαδή του μετατοπιζόμενου πλαισίου) συγκρίνεται με ένα προηγούμενο πλαίσιο έτσι ώστε να βρεθεί ένα μπλοκ που είναι παρόμοιο. Η σύγκριση των μπλοκ πραγματοποιείται μόνο στο φωτεινό τμήμα των πλαισίων. Η αναζήτηση μπορεί να πραγματοποιηθεί σε όλο το προηγούμενο πλαίσιο, αλλά συνήθως περιορίζεται σε μια μικρότερη περιοχή αναζήτησης που κεντροθετείται γύρω από τη θέση του μπλοκ στόχου στο τρέχον πλαίσιο (όπως φαίνεται στην Εικόνα 3).



**Εικόνα 3:** Κατά τη διάρκεια της σύγκρισης των μπλοκ, κάθε μπλοκ στόχος του τρέχοντος πλαισίου συγκρίνεται με ένα προηγούμενο πλαίσιο έτσι ώστε να βρεθεί ένα μπλοκ που είναι παρόμοιο. Η αναζήτηση συνήθως περιορίζεται σε μια μικρότερη περιοχή αναζήτησης που κεντροθετείται γύρω από τη θέση του μπλοκ στόχου στο τρέχον πλαίσιο.



Αυτή η πρακτική θέτει ένα άνω όριο, γνωστό ως μέγιστη μετατόπιση, στο πόσο μακριά μπορούν να κινηθούν τα αντικείμενα μεταξύ των πλαισίων. Η μέγιστη μετατόπιση διευκρινίζεται ως ο μέγιστος αριθμός εικονοκυττάρων στην οριζόντια και κάθετη κατεύθυνση όπου ένα υποψήφιο μπλοκ μπορεί να είναι από τη θέση του μπλοκ στόχου στο αρχικό πλαίσιο.

Η περιοχή αναζήτησης δεν χρειάζεται να είναι τετραγωνική. Αν το μέγεθος του μπλοκ είναι  $b$  και η μέγιστη μετατόπιση στην οριζόντια και κάθετη κατεύθυνση είναι  $dx$  και  $dy$  αντίστοιχα, τότε η περιοχή αναζήτησης θα είναι μεγέθους  $(2dx+b)(2dy+b)$ . Όσο πιο μεγάλη είναι η επιτρεπόμενη μετατόπιση, τόσο μεγαλύτερη η πιθανότητα εύρεσης μιας καλής ομοιότητας.

Μερικά από τα συχνότερα χρησιμοποιούμενα κριτήρια για τη σύγκριση μπλοκ είναι η μέση απόλυτη διαφορά (MAD) και το μέσο τετραγωνικό σφάλμα (MSE).

Η μέση απόλυτη διαφορά (MAD) είναι το πιο δημοφιλές κριτήριο. Ισοδύναμα εικονοκύτταρα από κάθε μπλοκ συγκρίνονται και οι διαφορές τους συγκρίνονται, όπως περιγράφεται στην ακόλουθη εξίσωση:

$$\frac{1}{mn} \sum_{p=1}^m \sum_{q=1}^n |A[p, q] - B[p, q]| \quad (2)$$

Για δύο μπλοκ  $A, B$  μεγέθους  $n \times m$ .  $A[p, q]$  είναι η τιμή του εικονοκυττάρου στην  $p$  σειρά και στην  $q$  στήλη του μπλοκ  $A$ .

Όσο χαμηλότερη είναι η τιμή της μετρικής MAD τόσο καλύτερη είναι η σύγκριση και το υποψήφιο μπλοκ με την μικρότερη τιμή MAD πρέπει να επιλεγεί.

Το μέσο τετραγωνικό σφάλμα (MSE) είναι παρόμοιο με τη συνάρτηση μέσης απόλυτης διαφοράς, εκτός από το ότι η διάφορα μεταξύ των εικονοκυττάρων τετραγωνίζεται πριν αθροιστεί. Δηλαδή,

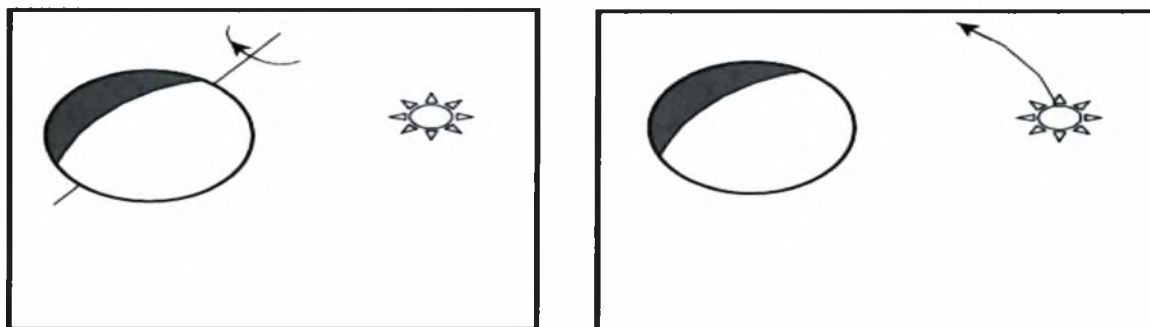
$$\frac{1}{mn} \sum_{p=1}^m \sum_{q=1}^n (A[p, q] - B[p, q])^2 \quad (3)$$

Όσο πιο μικρή είναι τιμή που επιστρέφει η παραπάνω συνάρτηση τόσο καλύτερο είναι το ταίριασμα.

### 2.3.2 Οπτική ροή (optical flow)

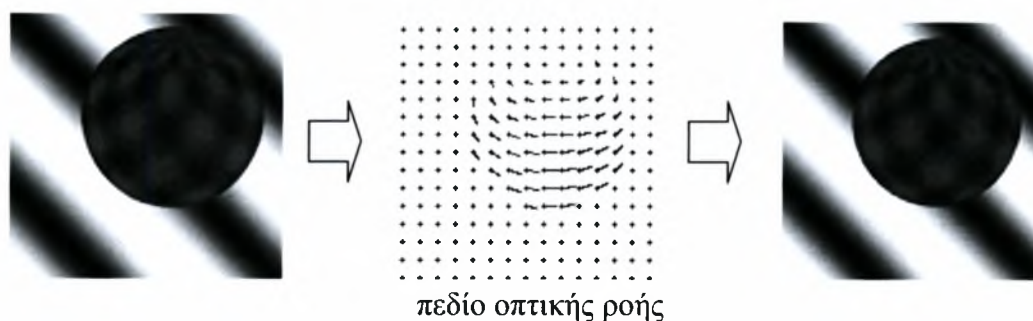
Η οπτική ροή είναι η κατανομή των εμφανών ταχυτήτων κίνησης των προτύπων φωτεινότητας σε μια εικόνα. Η οπτική ροή μπορεί να προκύψει από τη σχετική κίνηση των αντικειμένων και του παρατηρητή. Συνεπώς, η οπτική ροή μπορεί να δώσει σημαντική πληροφορία για τη χωρική διάταξη των παρατηρούμενων αντικειμένων και το ρυθμό αλλαγής αυτής της διάταξης. Οι ασυνέχειες στην οπτική ροή μπορούν να βοηθήσουν στην κατάρτιση των εικόνων σε περιοχές που αντιστοιχούν στα διαφορετικά αντικείμενα.

Οι δυσκολίες της αντιστοίχισης σημείων σε κινούμενα αντικείμενα, που ακολουθείται από τον υπολογισμό της θέσης και των ταχυτήτων των αντικειμένων με επαρκή ακρίβεια, έχει οδηγήσει σε μια τοπική προσέγγιση του προβλήματος όπου η αλλαγή της πυκνότητας σε ένα εικονοκύτταρο λαμβάνεται υπόψη. Οι παρακάτω εικόνες (Εικόνα 4(α)-(β)) δείχνουν ότι μια περιστρεφόμενη σφαίρα με προκαθορισμένες θέσεις του φωτός και της κάμερας δε θα δείξει καμία αλλαγή στις εντάσεις του εικονοκυττάρου, ενώ μια σφαίρα που παραμένει σταθερή σε σχέση με μια κινούμενη πηγή φωτός θα παράγει αλλαγές στην ένταση. Υποθέτουμε ότι ο φωτισμός είναι σταθερός, και η κίνηση των ακίνητων αντικειμένων λαμβάνεται υπόψη σε σχέση με μια προκαθορισμένη θέση της κάμερας.



**Εικόνα 4:** Μια περιστρεφόμενη σφαίρα με προκαθορισμένες θέσεις του φωτός και της κάμερας δε θα δείξει καμία αλλαγή στις εντάσεις του εικονοκυττάρου, ενώ μια σφαίρα που παραμένει σταθερή σε σχέση με μια κινούμενη πηγή φωτός θα παράγει αλλαγές στην ένταση. (α) Κίνηση αλλά όχι οπτική ροή. (β) Οπτική ροή αλλά όχι κίνηση.

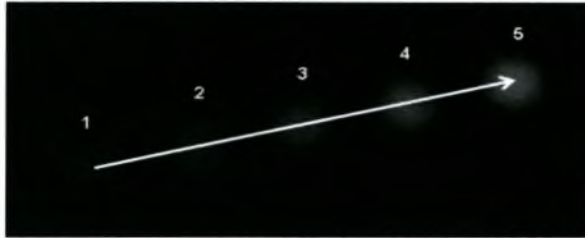
Σε κάθε εικονοκύτταρο αντιστοιχεί το διάνυσμα της ταχύτητας ( $dx/dt, dy/dt$ ) το οποίο δείχνει πόσο γρήγορα ένα εικονοκύτταρο κινείται διαμέσου της εικόνας καθώς και την κατεύθυνση της κίνησης. Η οπτική ροή είναι ένα πεδίο ταχύτητας το οποίο παραμορφώνει μια εικόνα σε μια άλλη (παρόμοια) εικόνα. Αυτό φαίνεται στην παρακάτω εικόνα (Εικόνα 5). Η σφαίρα περιστρέφεται από τα αριστερά στα δεξιά, παράγοντας το πεδίο της οπτικής ροής που φαίνεται στο κέντρο.



**Εικόνα 5:** Η οπτική ροή είναι ένα πεδίο ταχύτητας το οποίο παραμορφώνει μια εικόνα σε μια άλλη (παρόμοια) εικόνα. Η σφαίρα περιστρέφεται από τα αριστερά στα δεξιά, παράγοντας το πεδίο της οπτικής ροής που φαίνεται στο κέντρο.

## Υπολογισμός οπτικής ροής

Για τον υπολογισμό της οπτικής ροής υποθέτουμε ότι ισχύει ότι η φωτεινότητα των κινούμενων αντικειμένων, ή η ένταση της εικόνας  $I$ , δεν αλλάζει αλλά παραμένει σταθερή στον χρόνο, και ότι η ένταση της εικόνας είναι συνεχής και διαφοροποιήσιμη.



**Εικόνα 6:** Το διάνυσμα οπτικής ροής ενός κινούμενου αντικειμένου σε μια ακολουθία βίντεο.

Υποθέτουμε ότι η ένταση της εικόνας δίνεται από την  $I(x,y,t)$ , όπου η ένταση είναι συνάρτηση του χρόνου  $t$  καθώς επίσης του σημείου  $(x, y)$  που αντιστοιχεί στη χρονική στιγμή  $t$ .

Το σημείο  $(x, y)$  που βρίσκεται στην εικόνα τη χρονική στιγμή  $t$ , μετά από ένα διάστημα  $dt$  μετατοπίζεται κατά μια απόσταση  $(dx, dy)$  στην εικόνα. Η ένταση του δίνεται από τη σχέση:

$$I(x + dx, y + dy, t + dt) = I(x, y, t) + \frac{\partial I}{\partial x} dx + \frac{\partial I}{\partial y} dy + \frac{\partial I}{\partial t} dt \quad (4)$$

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:

$$I(x + dx, y + dy, t + dt) = I(x, y, t) \quad (5)$$

Από τις εξισώσεις (4) και (5) καταλήγουμε στη σχέση:

$$\frac{\partial I}{\partial x} dx + \frac{\partial I}{\partial y} dy + \frac{\partial I}{\partial t} dt = 0 \quad (6)$$

Διαιρώντας την (6) με  $dt$  και γνωρίζοντας ότι  $u = \frac{dx}{dt}$  και  $v = \frac{dy}{dt}$ , όπου  $u$  και  $v$  είναι οι ταχύτητες του αντικειμένου κατά μήκος της  $x$  και  $y$  κατεύθυνσης αντίστοιχα καταλήγουμε στην σχέση:

$$I_x u + I_y v + I_t = 0 \quad (7)$$

όπου  $I_x = \frac{\partial I}{\partial x}$ ,  $I_y = \frac{\partial I}{\partial y}$  και  $I_t = \frac{\partial I}{\partial t}$ . Η σχέση (7) είναι η σχέση περιορισμού της οπτικής ροής (optical flow constraint equation) και χρησιμοποιείται για τον υπολογισμό της οπτικής ροής. Η προσέγγιση των  $u$  και  $v$  δεν είναι εύκολη αφού έχουμε μία εξίσωση με δυο αγνώστους. Αυτό το πρόβλημα είναι γνωστό ως πρόβλημα ανοιγμάτων (aperture problem).

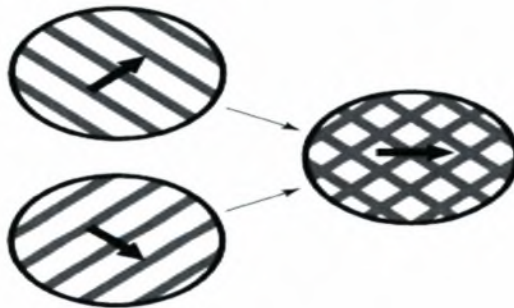
### Πρόβλημα ανοιγμάτων (Aperture problem)

Στην εξίσωση της οπτικής ροής υπάρχουν δυο άγνωστοι ( $u$ ,  $v$ ) αλλά μία μόνο εξίσωση.

$$I_x u + I_y v + I_t = 0 \quad (8)$$

Αυτό έχει ως αποτέλεσμα να μην μπορούμε να προσδιορίσουμε μοναδικά το πεδίο κίνησης. Μπορούμε μόνο να ανακτήσουμε στην διεύθυνση της χωρικής κλίσης της εικόνας. Η κίνηση που είναι κάθετη στην διεύθυνση της χωρικής κλίσης της εικόνας δεν μπορεί να υπολογιστεί σε μία τοπική περιοχή.

Το πρόβλημα ανοιγμάτων μελετάται συχνά με την επίθεση δύο πλεγμάτων που κινούνται σε διαφορετικές κατευθύνσεις. Όπως φαίνεται στην Εικόνα 7, το ένα πλέγμα κινείται πάνω και δεξιά και το άλλο κάτω και δεξιά. Το προκύπτον σχέδιο πλέγματος μπορεί να εμφανιστεί να κινείται σε μια κατεύθυνση διαφορετική από και τα δύο επιμέρους συστατικά του μέρη. Σε αυτήν την περίπτωση η κίνηση είναι άμεσα προς τα δεξιά.



**Εικόνα 7:** Aperture problem. Το ένα πλέγμα κινείται πάνω και δεξιά και το άλλο κάτω και δεξιά. Το προκύπτον σχέδιο πλέγματος μπορεί να εμφανιστεί να κινείται σε μια κατεύθυνση διαφορετική από και τα δύο επιμέρους συστατικά του μέρη.

Για αυτό χρειαζόμαστε και άλλες εξισώσεις για την προσέγγιση του διανύσματος ταχύτητας.

### 2.3.3 Μέθοδοι υπολογισμού οπτικής ροής

Ακολουθεί μια σύντομη περιγραφή των διαφόρων μεθόδων που χρησιμοποιούνται για τον υπολογισμό της οπτικής ροής, καθώς επίσης και αρκετές από τις λεπτομέρειες εφαρμογής.

#### 2.3.3.1 Διαφορικές τεχνικές

Οι διαφορικές τεχνικές υπολογίζουν την ταχύτητα από χωροχρονικές παραγώγους της έντασης της εικόνα ή φιλτραρισμένες εκδοχές της εικόνας (χρησιμοποιώντας χαμηλοπερατά ή βαθυπερατά φίλτρα). Αρχικά στις πρώτες προσπάθειες ανάπτυξης των τεχνικών, χρησιμοποιούνταν πρώτης τάξης παράγωγοι και βασίζονταν στην μετατόπιση της εικόνας, δηλαδή

$$I(\mathbf{x}, t) = I(\mathbf{x} - \mathbf{v}t, 0) \quad (9)$$

Όπου  $\mathbf{v} = (u, v)^T$ . Γενικά από μια υπόθεση ότι η ένταση είναι προστατευμένη,  $\frac{dI(\mathbf{x}, t)}{dt} = 0$ , η εξίσωση περιορισμού κλίσης εύκολα παράγεται από:

$$\nabla I(\mathbf{x}, t) \cdot \mathbf{v} + I_t(\mathbf{x}, t) = 0 \quad (10)$$

όπου το  $I_t(\mathbf{x}, t)$  δηλώνει τη μερική χρονική παράγωγο του  $I(\mathbf{x}, t)$ ,  $\nabla I(\mathbf{x}, t) = (I_x(\mathbf{x}, t), I_y(\mathbf{x}, t))^T$  και το  $\nabla I \cdot \mathbf{v}$  δηλώνει το εσωτερικό γινόμενο.

Υπάρχουν δύο άγνωστες συνιστώσες του  $\mathbf{v}$  στην (9), περιορισμένες από μόνο μια γραμμική εξίσωση. Επιπλέον περιορισμοί είναι απαραίτητοι για την επίλυση και των δύο συνιστωσών του  $\mathbf{v}$ .

Οι δεύτερης τάξης διαφορικές μέθοδοι χρησιμοποιούν δεύτερης τάξης παραγώγους (τον πίνακα Hessian του  $I$ ) για να περιορίσουν την διαδιάστατη ταχύτητα:

$$\begin{bmatrix} I_{xx}(\mathbf{x}, t) & I_{yx}(\mathbf{x}, t) \\ I_{xy}(\mathbf{x}, t) & I_{yy}(\mathbf{x}, t) \end{bmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} + \begin{pmatrix} I_{tx}(\mathbf{x}, t) \\ I_{ty}(\mathbf{x}, t) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad (11)$$

Η εξίσωση (11) μπορεί να παραχθεί από την (9), ή από την διατήρηση του  $\nabla I(\mathbf{x}, t), \frac{d\nabla I(\mathbf{x}, t)}{dt} = 0$ . Η διατήρηση του  $\nabla I(\mathbf{x}, t)$  υπονοεί ότι η πρώτη τάξης παραμόρφωση της έντασης (περιστροφή ή διαστολή) δε θα έπρεπε να είναι παρούσα. Αυτός επομένως είναι ένας ισχυρότερος περιορισμός από αυτόν της εξίσωσης (10) στα επιτρεπτά πεδία κίνησης. Για να εκτιμήσεις την ταχύτητα της εικόνας, υποθέτοντας  $\frac{d\nabla I(\mathbf{x}, t)}{dt} = 0$ , οι περιορισμοί στη (11) μπορούν να χρησιμοποιηθούν μόνοι τους ή μαζί με την (10) για να παράγουν ένα υπέρ προσδιορισμένο σύστημα γραμμικών εξισώσεων. Παρόλα αυτά, αν το πρόβλημα ανοιγμάτων (aperture problem) επικρατεί σε μια τοπική γειτονιά (δηλαδή αν η ένταση είναι μιας



διάστασης), τότε εξαιτίας της ευαισθησίας της αριθμητικής παραγωγίσης, οι δεύτερης τάξης παράγωγοι δεν μπορούν να υπολογιστούν με αρκετή ακρίβεια για να προσδιορίσουν τις εφαπτόμενες συνιστώσες του  $\mathbf{v}$ . Συνεπώς οι εκτιμήσεις ταχύτητας από δεύτερης τάξης μεθόδους είναι περισσότερο ανεπαρκείς και λιγότερο ακριβείς από τις εκτιμήσεις των μεθόδων πρώτης τάξης.

Φυσικά μια απαίτηση των διαφορικών τεχνικών είναι ότι το  $I(\mathbf{x}, t)$  πρέπει να είναι διαφοροποιήσιμο. Αυτό υπονοεί ότι η χρονική εξομάλυνση στους σένσορες χρειάζεται για να αποφευχθεί η διαστρέβλωση (aliasing) της εικόνας και ότι η αριθμητική παραγωγή πρέπει να γίνει προσεκτικά.

Παρακάτω παρατίθενται δύο διαφορικές μέθοδοι. Περιλαμβάνουν πρώτης και δεύτερης τάξης περιορισμούς, όπως επίσης και τοπικές και καθολικές μεθόδους που συνδυάζουν τους τοπικούς περιορισμούς.

### Horn και Schunck

Οι Horn και Schunck συνδύασαν τον περιορισμό κλίσης της (10) με ένα καθολικό όρο ομαλότητας για να περιορίσουν το εκτιμώμενο πεδίο ταχύτητας  $\mathbf{v}(\mathbf{x}, t) = (u(\mathbf{x}, t), v(\mathbf{x}, t))$  ελαχιστοποιώντας το:

$$\int_D (\nabla I \cdot \mathbf{v} + I_t)^2 + \lambda^2 (\|\nabla u\|_2^2 + \|\nabla v\|_2^2) dx \quad (12)$$

Ορισμένο σε έναν τομέα  $D$ , όπου το μέγεθος του  $\lambda$  αντικατοπτρίζει την επιρροή του όρου ομαλότητας. Επαναληπτικές εξισώσεις χρησιμοποιούνται για να ελαχιστοποιήσουν την (12) και πετυχαίνουν ταχύτητα εικόνας:

$$u^{k+1} = \bar{u}^k - \frac{I_x [I_x \bar{u}^k + I_y \bar{v}^k + I_t]}{a^2 + I_x^2 + I_y^2} \quad (13)$$

$$v^{k+1} = \bar{v}^k - \frac{I_y [I_x \bar{u}^k + I_y \bar{v}^k + I_t]}{a^2 + I_x^2 + I_y^2} \quad (14)$$

όπου το  $k$  δηλώνει τον αριθμό επανάληψης  $u^0, v^0$  δηλώνουν αρχικές εκτιμήσεις ταχύτητας οι οποίες θέτονται ίσες με μηδέν, και τα  $\bar{u}^k, \bar{v}^k$  δηλώνουν γειτονικούς μέσους όρους των  $u^k, v^k$ .

### Lucas και Kanade

Εφαρμόζεται ένα σταθμισμένο ελαχίστων τετραγώνων ταίριασμα των τοπικών πρώτης τάξης περιορισμών της σχέσης (10) σε ένα σταθερό μοντέλο για το  $\mathbf{v}$ , σε κάθε μικρή χωρική γειτονιά  $\Omega$  ελαχιστοποιώντας το:



$$\sum_{\mathbf{x} \in \Omega} W^2(\mathbf{x}) [\nabla I \cdot \mathbf{v} + I_t]^2 \quad (15)$$

όπου το  $W(\mathbf{x})$  δηλώνει μια συνάρτηση παραθύρου που δίνει περισσότερη επιρροή σε περιορισμούς που βρίσκονται στο κέντρο της γειτονιάς παρά σε αυτούς που είναι στη περιφέρεια. Η λύση της (15) δίνεται από:

$$A^T W^2 A \mathbf{v} = A^T W^2 \mathbf{b} \quad (16)$$

όπου για  $n$  σημεία  $\mathbf{x}_i \in \Omega$ , σε μια μεμονωμένη στιγμή  $t$ ,

$$\begin{aligned} A &= [\nabla I(\mathbf{x}_1), \dots, \nabla I(\mathbf{x}_n)]^T \\ W &= \text{diag}[W(\mathbf{x}_1), \dots, W(\mathbf{x}_n)] \\ \mathbf{b} &= -[I_t(\mathbf{x}_1), \dots, I_t(\mathbf{x}_n)]^T \end{aligned}$$

Η λύση της (16) είναι  $\mathbf{v} = (A^T W^2 A)^{-1} A^T W^2 \mathbf{b}$ , η οποία λύνεται όταν το  $A^T W^2 A$  είναι nonsingular, αφού είναι ένας πίνακας  $2 \times 2$ :

$$A^T W^2 A = \begin{bmatrix} \sum W^2(\mathbf{x}) I_x^2(\mathbf{x}) & \sum W^2(\mathbf{x}) I_x(\mathbf{x}) I_y(\mathbf{x}) \\ \sum W^2(\mathbf{x}) I_y(\mathbf{x}) I_x(\mathbf{x}) & \sum W^2(\mathbf{x}) I_y^2(\mathbf{x}) \end{bmatrix} \quad (17)$$

### 2.3.3.2 Ταίριασμα βασισμένο την περιοχή (region based matching)

Η ακριβής αριθμητική παραγωγή μπορεί να είναι μη πρακτική λόγω του θορύβου, γιατί μόνο ένας μικρός αριθμός πλαισίων υπάρχει ή λόγω της διαστρέβλωσης κατά τη διαδικασία απόκτησης της εικόνας. Σε αυτές τις περιπτώσεις οι διαφορικές προσεγγίσεις μπορεί να είναι ακατάλληλες και είναι φυσική η στροφή σε μεθόδους ταιριάσματος βασισμένο στην περιοχή. Τέτοιες προσεγγίσεις ορίζουν την ταχύτητα  $\mathbf{v}$  ως μια μετατόπιση  $\mathbf{d} = (d_x, d_y)$ , η οποία παράγει το καλύτερο ταίριασμα μεταξύ περιοχών εικόνων σε διαφορετικές στιγμές. Η αναζήτηση του καλύτερου ταιριάσματος ισοδυναμεί με την μεγιστοποίηση ενός μέτρου ομοιότητας (μεγαλύτερο του  $\mathbf{d}$ ), όπως η ομαλοποιημένη συσχέτιση ή με την ελαχιστοποίηση ενός μέτρου απόστασης, όπως το άθροισμα των τετραγωνισμένων αποστάσεων (SSD)

$$\begin{aligned} \text{SSD}_{1,2}(\mathbf{x}; \mathbf{d}) &= \sum_{j=-n}^n \sum_{i=-n}^n W(i,j) [I_1(\mathbf{x} + (i,j)) - I_2(\mathbf{x} + \mathbf{d} + (i,j))]^2 \\ &= W(\mathbf{x}) * [I_1(\mathbf{x}) - I_2(\mathbf{x} + \mathbf{d})]^2 \end{aligned} \quad (18)$$

όπου το  $W$  δηλώνει ένα διακριτή δισδιάστατη συνάρτηση παραθύρου και η  $\mathbf{d} = (d_x, d_y)$  παίρνει ακέραιες τιμές.

### 2.3.3.3 Μέθοδοι βασισμένες στην ενέργεια

Αυτές οι τεχνικές βασίζονται στη συνολική παραγωγή ενέργειας από φίλτρα συντονισμένης ταχύτητας. Αναφέρονται επίσης και ως μέθοδοι βασισμένες στη συχνότητα λόγω του ότι τα φίλτρα συντονισμένης ταχύτητας σχεδιάζονται στον τομέα Fourier. Ο μετασχηματισμός Fourier ενός μετατοπισμένου δισδιάστατου προτύπου είναι:

$$\hat{I}(\mathbf{k}, \omega) = \hat{I}_0(\mathbf{k})\delta(\omega + \mathbf{v}^T \mathbf{k}) \quad (19)$$

όπου  $\hat{I}_0(\mathbf{k})$  είναι ο μετασχηματισμός Fourier του  $I(\mathbf{x}, 0)$ , η  $\delta(k)$  είναι συνάρτησης δέλτα, το  $\omega$  δηλώνει τη χρονική συχνότητα και το  $\mathbf{k} = (k_x, k_y)$  είναι η χωρική συχνότητα. Αυτό δείχνει ότι η μη μηδενική ισχύς που σχετίζεται με ένα μετατοπιζόμενο δισδιάστατο πρότυπο (pattern) βρίσκεται σε ένα επίπεδο δια μέσου του αρχικού στο χώρο της συχνότητας.

## 2.4 Μέθοδοι εκτίμησης κίνησης στη συχνότητα

Οι μέθοδοι αυτές εκμεταλλεύονται αρκετές ιδιότητες του μετασχηματισμού Fourier. Ψάχνουν για το βέλτιστο «ταίριασμα» σύμφωνα με την πληροφορία από τον τομέα της συχνότητας.

Ο μετασχηματισμός Fourier μιας εικόνας  $f(x,y)$  είναι μια σύνθετη συνάρτηση. Κάθε τιμή της συνάρτησης έχει ένα πραγματικό μέρος  $R(\omega_x, \omega_y)$  και ένα φανταστικό μέρος  $I(\omega_x, \omega_y)$  σε κάθε συχνότητα  $(\omega_x, \omega_y)$  του φάσματος συχνοτήτων

$$F(\omega_x, \omega_y) = R(\omega_x, \omega_y) + i I(\omega_x, \omega_y).$$

Αυτό μπορεί να εκφραστεί εναλλακτικά σε εκθετική μορφή ως:

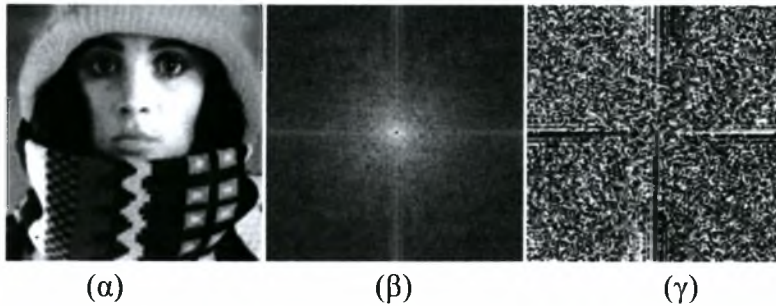
$$F(\omega_x, \omega_y) = |F(\omega_x, \omega_y)|e^{i\phi(\omega_x, \omega_y)} \quad (20)$$

όπου το  $|F(\omega_x, \omega_y)|$  είναι το μέγεθος του μετασχηματισμού Fourier και  $\phi(\omega_x, \omega_y)$  είναι η γωνία φάσης. Το φάσμα της συχνότητας ορίζεται δίνεται από την σχέση:

$$|F(\omega_x, \omega_y)| = \sqrt{R(\omega_x, \omega_y)^2 + I(\omega_x, \omega_y)^2} \quad (21)$$

Η γωνία φάσης περιγράφει το ποσό της μετατόπισης της φάσης σε κάθε συχνότητα και περιγράφεται από την:

$$\phi(\omega_x, \omega_y) = \tan^{-1}[I(\omega_x, \omega_y)/R(\omega_x, \omega_y)] \quad (22)$$



**Εικόνα 8:** Ο μετασχηματισμός Fourier μιας εικόνας μπορεί να είναι σύνθετος. (α) Αρχική εικόνα. (β) Το μέγεθος σε μια κλιμακούμενη μορφή ως  $\log|F(\omega_x, \omega_y)|$ . (γ) Η φάση  $\phi(\omega_x, \omega_y)$ .

Ο συσχετισμός φάσης (phase correlation) είναι μια αρκετά δημοφιλής και αποδοτική τεχνική για την εκτίμηση κίνησης στο χώρο των συχνοτήτων.

### 2.4.1 Συσχέτιση φάσης

Η συσχέτιση φάσης λειτουργεί σε ένα ζευγάρι εικόνων  $f_t$  και  $f_{t+1}$  ίδιων διαστάσεων που ανήκουν σε συνεχόμενα πλαίσια μιας κινούμενης ακολουθίας που δειγματοληπτήθηκε τις στιγμές  $t, t+1$ . Η εκτίμηση της κίνησης στηρίζεται στην ανίχνευση του μεγίστου της συνάρτησης διαγώνιας συσχέτισης (cross-correlation) μεταξύ των εικόνων  $f_t$  και  $f_{t+1}$ . Δεδομένου ότι όλες οι συναρτήσεις που περιλαμβάνονται είναι διακεκριμένες, η διαγώνια συσχέτιση είναι κυκλική, και μπορεί να πραγματοποιηθεί ως πολλαπλασιασμός στο πεδίο συχνότητας χρησιμοποιώντας γρήγορες εφαρμογές. Η πραγματική εκτιμώμενη επιφάνεια συσχετισμού ορίζεται ως:

$$c_{t,t+1}(k,l) = F^{-1} \left( \frac{F_t^* F_{t+1}}{|F_t^* F_{t+1}|} \right) \quad (23)$$

όπου  $F_t$  και  $F_{t+1}$  είναι αντίστοιχα οι δισδιάστατοι διακριτοί μετασχηματισμοί Fourier των  $f_t$  και  $f_{t+1}$ . Το  $F^{-1}$  δηλώνει τον αντίστροφο μετασχηματισμό Fourier και το  $*$  δηλώνει το συζυγή. Οι συντεταγμένες  $(k_m, l_m)$  του μεγίστου της πραγματικής εκτιμώμενης τιμής της διάταξης  $c_{t,t+1}$  μπορούν να χρησιμοποιηθούν σαν εκτίμηση των οριζόντιων και κάθετων συνιστωσών της κίνησης μεταξύ των  $f_t$  και  $f_{t+1}$  ως ακολούθως:

$$(k_m, l_m) = \arg \max \{c_{t,t+1}(k,l)\} \quad (24)$$

Η ακρίβεια των μετρήσεων της κίνησης στα υπό εικονοκύτταρα (subpixels) επιτυγχάνεται με την προσαρμογή των διαχωρίσιμων μεταβλητών που εκτελείται στη γειτονιά του μεγίστου, χρησιμοποιώντας μιας διάστασης τετραγωνικές συναρτήσεις.

Χρησιμοποιώντας την εξίσωση (24), πρότυπες συναρτήσεις εφαρμόζονται στις τριπλέτες:

$$\{c_{t,t+1}(k_m - 1, l_m), c_{t,t+1}(k_m, l_m), c_{t,t+1}(k_m + 1, l_m)\}$$

και

$$\{c_{t,t+1}(k_m, l_m - 1), c_{t,t+1}(k_m, l_m), c_{t,t+1}(k_m, l_m + 1)\}$$

Η θέση του μεγίστου της εφαρμοζόμενης συνάρτησης παρέχει την απαιτούμενη εκτίμηση κίνησης του υπό εικονοκυττάρου (dx,dy). Για παράδειγμα, εφαρμόζοντας μια παραβολική συνάρτηση οριζόντια στην αριστερή μεριά της παραπάνω εξίσωσης, παράγεται μια κλειστού τύπου λύση για την οριζόντια συνιστώσα της εκτίμησης κίνησης dx:

$$dx = \frac{c_{t,t+1}(k_m + 1, l_m) - c_{t,t+1}(k_m - 1, l_m)}{2(2c_{t,t+1}(k_m, l_m) - c_{t,t+1}(k_m + 1, l_m) - c_{t,t+1}(k_m - 1, l_m))} \quad (25)$$

Το κλασματικό μέρος dy της κάθετης συνιστώσας μπορεί να επιτευχθεί με παρόμοιο τρόπο χρησιμοποιώντας το δεξί μέρος της παραπάνω εξίσωσης.



**Εικόνα 9:** Συσχετισμός φάσης μεταξύ δύο μπλοκ.

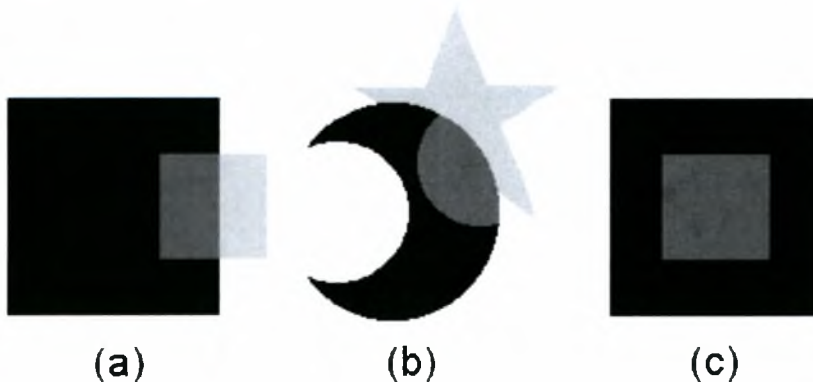
### 3. Διαφάνεια

Η διαφάνεια είναι η ιδιότητα των αντικειμένων να αφήνουν το φως να περνάει δια μέσου αυτών. Τα διαφανή υλικά είναι καθαρά, μπορείς να δεις μέσα από αυτά. Παραδείγματα διαφανών υλικών είναι ο αέρας, υγρά όπως το νερό και μερικά πλαστικά. Όπου ο βαθμός της διαφάνειας ποικίλλει σύμφωνα με το μήκος κύματος του φωτός, η εικόνα που φαίνεται μέσω του υλικού είναι παραλλαγμένη χρωματικά.



**Εικόνα 10:** Μια πεταλούδα με διαφανή φτερά. Το σώμα της φαίνεται καθαρά μέσα από τα φτερά.

Η αντίληψη για τη διαφάνεια μόνο αόριστα περιορίζεται από τους νόμους της οπτικής. Στην πραγματικότητα, η σχηματική ενότητα είναι ίσως εξίσου σημαντική ως ένας παράγοντας για την αντίληψη της διαφάνειας όπως η σχέση μεταξύ της φωτεινότητας των διαφορετικών περιοχών εικόνας. Εάν υπάρχει μια απότομη αλλαγή της μορφής στα όρια μεταξύ των εικόνων, η αντίληψη για τη διαφάνεια μπορεί να μην υφίσταται, ακόμα κι αν η διαφάνεια είναι πραγματικά παρούσα (εικόνα 11b). Υπάρχουν άλλες περιπτώσεις όπου η διαφάνεια υπάρχει πραγματικά αλλά δεν γίνεται αντιληπτή. Παραδείγματος χάριν, ένα τετραγωνικό φίλτρο που τοποθετείται σε ένα ολοφάνερο φόντο θα θεωρηθεί κανονικά ως χρωματισμένο μπάλωμα, πιθανώς επειδή αυτό είναι η απλούστερη εξήγηση (εικόνα 11c).



**Εικόνα 11:** Η κανονική αντίληψη για τη διαφάνεια φαίνεται στο a. Οι περισσότεροι θα αναφέρουν ότι βλέπουν ένα μικρό διαφανές τετράγωνο πάνω από ένα μεγάλο σκοτεινό τετράγωνο. Παρόλα αυτά, η αντίληψη της διαφάνειας μπορεί να «διαταραχτεί», όπως φαίνεται στο b, από μια ξαφνική αλλαγή του σχήματος. Η



διαφάνεια δεν θα γίνει αντιληπτή ούτε στο c. Ακόμα κι αν το μικρότερο τετράγωνο είναι στην πραγματικότητα διαφανές, η σκηνή μπορεί να εξηγηθεί απλούστερα χωρίς διαφάνεια.

Είναι δυνατό να προκληθεί μια αντίληψη για τη διαφάνεια στις στατικές εικόνες και μερικοί γενικοί περιορισμοί στις σχέσεις φωτεινότητας έχουν περιγραφεί. Η αντίληψη για τη διαφάνεια μπορεί συχνά να ενισχυθεί όταν οι διαφορές στην κίνηση βοηθούν να διαχωριστούν οι δύο επιφάνειες. Δεν είναι αυστηρά απαραίτητο να υπάρξουν δύο ευδιάκριτες επιφάνειες για να παραχθεί διαφάνεια στην κίνηση. Για παράδειγμα, μια σκιά μπορεί να κινείται δια μέσου ενός αντικειμένου. Η κίνηση της σκιάς, αν και μπορεί να παρατηρηθεί, δεν αποδίδεται στο αντικείμενο που βρίσκεται από κάτω, ούτε αποδίδεται στην παρουσία μιας δεύτερης επιφάνειας. Αυτό το παράδειγμα μπορεί να περιγραφεί ως προϊόν μιας εικόνας ανάκλασης (δηλαδή, της χαρακτηριστικής ιδιότητας, υψής του αντικειμένου ή του χρώματός του) και μιας εικόνας φωτισμού (δηλαδή, το χωρικό πρότυπο του φωτός και της σκιάς). Αυτή η απλή μαθηματική σχέση περιγράφει επίσης τη φυσική των μερικώς διαφανών αντικειμένων όταν δεν υπάρχει καμία διάχυτη αντανάκλαση. Αυτή η κατάσταση θα αναφέρεται ως πολλαπλασιαστική διαφάνεια (Εικόνα 13). Ένα άλλο παράδειγμα είναι αυτό της κατοπτρικής αντανάκλασης από τις στιλπνές επιφάνειες όπως τα φύλλα, η γούνα, το δέρμα, η τρίχα, ή το σελοφάν. Η κατοπτρική αντανάκλαση από μια επίπεδη επιφάνεια όπως μια διεπαφή αέρα-νερού, παράγει ένα αποτέλεσμα στο οποίο μια ανακλώμενη εικόνα συνδυάζεται προσθετικά με τις εικόνες των συγχωνευμένων αντικειμένων (Εικόνα 12). Η κίνηση της ανακλώμενης εικόνας είναι γενικά διαφορετική από αυτή του ανακλώμενου αντικειμένου. Αυτή η κατάσταση αναφέρεται ως προσθετική διαφάνεια. Γενικά ο όρος διαφανής κίνηση αναφέρεται σε μια κατάσταση όπου πολλαπλές ευδιάκριτες κινήσεις φαίνονται σε μια ενιαία θέση.

Η διαφάνεια είναι πλήρως αχρωματική στη φύση. Οι συνδυασμοί χρώματος που είναι απίθανο να προκύψουν στις πραγματικές σκηνές δίνουν ακόμα την αντίληψη για τη διαφάνεια.



**Εικόνα 12:** Η κατοπτρική αντανάκλαση από μια επίπεδη επιφάνεια όπως μια διεπαφή αέρα-νερού, παράγει ένα αποτέλεσμα στο οποίο μια ανακλώμενη εικόνα συνδυάζεται προσθετικά με τις εικόνες των συγχωνευμένων αντικειμένων. Ένα παράδειγμα διαφάνειας λόγω προσθετικής επίθεσης (επίθεση σκηνής ανακλώμενου βουνού και νερού).

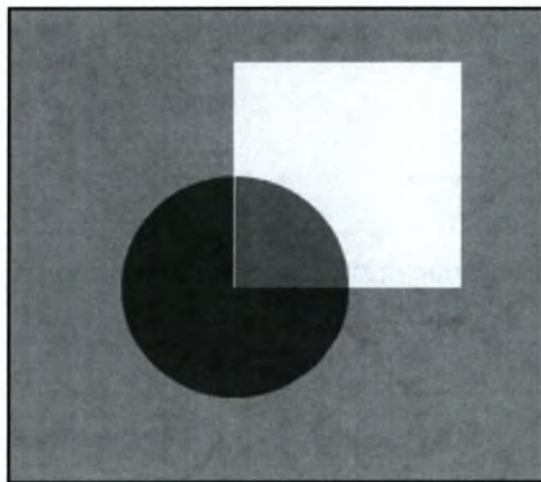




**Εικόνα 13:** Μια σκιά μπορεί να κινείται δια μέσου ενός αντικείμενου, η κίνηση της σκιάς, αν και μπορεί να παρατηρηθεί, δεν αποδίδεται στο αντικείμενο που βρίσκεται από κάτω, ούτε αποδίδεται στην παρουσία μιας δεύτερης επιφάνειας. Ένα παράδειγμα διαφάνειας λόγω πολλαπλασιαστικής επίθεσης (επίθεση κατασκευασμένης σκιάς από δέντρο και της αντανάκλασης του εδάφους και του ζώου).

### 3.1 Τυπικά χαρακτηριστικά της διαφάνειας

Η Εικόνα 14 δείχνει ένα παράδειγμα οπτικής διαφάνειας. Η Εικόνα 14 μπορεί να προκύψει από διαφορετικές αιτίες. Για παράδειγμα, ένα τετράγωνο από τσιγαρόχαρτο μπορεί να είναι μπροστά από ένα σκούρο γκρι κύκλο, ή μια κυκλική σκιά μπορεί να πέσει πάνω σε ένα αεροπλάνο που έχει ένα ανοιχτόχρωμο γκρι τετράγωνο, ή ένα σκούρο κυκλικό φίλτρο μπορεί να βρίσκεται πάνω από ένα ανοιχτόχρωμο γκρι τετράγωνο. Κάποιος μπορεί να θεωρήσει την Εικόνα 14 ως συνδυασμό δύο πρωταρχικών (primitive) εικόνων.

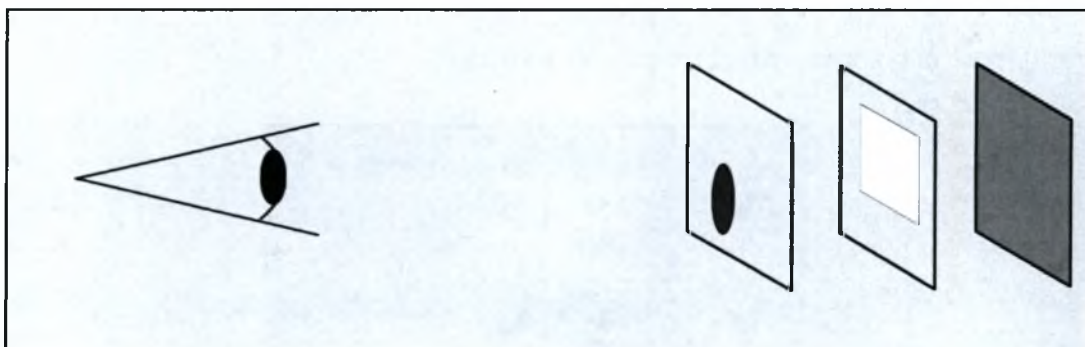


**Εικόνα 14:** Ένα παράδειγμα οπτικής διαφάνειας, η οποία μπορεί να προκύψει από διαφορετικές αιτίες. Ένα τετράγωνο από τσιγαρόχαρτο μπορεί να είναι μπροστά από ένα σκούρο γκρι κύκλο, ή μια κυκλική σκιά μπορεί να πέσει πάνω σε ένα αεροπλάνο που έχει ένα ανοιχτόχρωμο γκρι τετράγωνο, ή ένα σκούρο κυκλικό φίλτρο μπορεί να βρίσκεται πάνω από ένα ανοιχτόχρωμο γκρι τετράγωνο.

Πολλά φυσικά φαινόμενα μπορούν να έχουν ως αποτέλεσμα τη διαφάνεια. Για παράδειγμα, τα σκούρα φίλτρα, κατοπτρικές ανακλάσεις, ριπές καπνού, τούλια από κουρτίνες και σκιές, όλα συνδυασμένα με διάφορα σχήματα πίσω τους σε ένα διαφανή τρόπο.

Όταν μια εικόνα έχει σχηματιστεί από τον συνδυασμό δύο αρχικών εικόνων, τότε είναι συνήθως πιο συντετό η εικόνα να περιγραφτεί αναφορικά με αυτόν τον συνδυασμό, κατά συνέπεια υπάρχουν περισσότερα πλεονεκτήματα για ένα οπτικό σύστημα να αναλύσει την εικόνα στις αρχικές εικόνες από τις οποίες προήλθε σύμφωνα με κάποιο συνδυαστικό κανόνα. Αυτή η οικονομία δεν εξαρτάται από την ανάθεση μιας μοναδικής φυσικής ερμηνείας στις αρχικές εικόνες, η Εικόνα 14 μπορεί να αναλυθεί σε ένα κύκλο και ένα τετράγωνο.

Αρχικά η οπτική διαφάνεια θα μπορούσε να αναλυθεί σε ένα «προ-φυσικής» επίπεδο, το οποίο δεν περιλαμβάνει τη φυσική ιδιομορφία μιας πλήρους εσωτερικής ανάλυσης εικόνας. Η αναπαράσταση σε αυτό το επίπεδο αποτελείται από ένα σύνολο επιπέδων της αρχικής εικόνας που διατάσσονται σε βάθος. Κάθε επίπεδο περιέχει «γεμάτες» περιοχές που τροποποιούν την εμφάνιση των επιπέδων που βρίσκονται από κάτω τους, και «άδειες» περιοχές που είναι τελείως καθαρές. Οι «γεμάτες» περιοχές διαφορετικών επιπέδων συνδυάζονται μεταξύ τους σύμφωνα με απλούς κανόνες όπως ο πολλαπλασιασμός και η πρόσθεση. Η Εικόνα 15 δείχνει ένα παράδειγμα των επιπέδων από τα οποία προήλθε η Εικόνα 14.

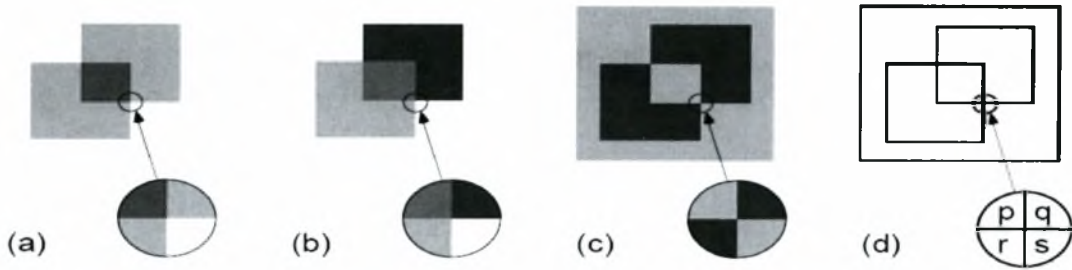


**Εικόνα 15:** Ένα παράδειγμα των επιπέδων από τα οποία προήλθε η Εικόνα 14.

### **Η σημασία των X συνδέσεων (junction)**

Όταν εικόνες σε ευδιάκριτα επίπεδα επικαλύπτονται, συνήθως δίνουν αφορμή για X συνδέσεις στην εικόνα, που συμβάλλουν σημαντικά στην αντίληψη για τη διαφάνεια από τους ανθρώπους. Αυτές οι X συνδέσεις μπορούν να δώσουν αρκετές πληροφορίες για τη φύση της αλληλεπίδρασης στη διαφάνεια, και τη διάταξη του βάθους στα διάφορα επίπεδα. Για παράδειγμα, οι εικόνες 16(a-c) περιέχουν 3 εικόνες, τις οποίες το ανθρώπινο οπτικό σύστημα ερμηνεύει με 3 διαφορετικούς τρόπους. Η διαφάνεια φαίνεται στην Εικόνα 16(a), η οποία εξηγείται σα να περιέχει δύο σκούρα φίλτρα, όμως η διάταξη του βάθους είναι διφορούμενη: κάθε τετράγωνο φαίνεται σα να βρίσκεται μπροστά από το άλλο. Η διαφάνεια φαίνεται επίσης στην Εικόνα 16(b), αλλά σε αυτή την περίπτωση η διάταξη του βάθους δεν είναι διφορούμενη: το

τετράγωνο που βρίσκεται κάτω αριστερά είναι μπροστά από το άλλο. Η διαφάνεια δεν φαίνεται στην 16(c), η οποία φαίνεται σαν ένα ζωγραφισμένο σχέδιο ενός μόνο στρώματος.



**Εικόνα 16:** (a) Η διαφάνεια εξηγείται σα να περιέχει δύο σκούρα φίλτρα, όμως η διάταξη του βάθους είναι διαφορούμενη: κάθε τετράγωνο φαίνεται σα να βρίσκεται μπροστά από το άλλο. (b) Η διαφάνεια φαίνεται, αλλά η διάταξη του βάθους δεν είναι διαφορούμενη: το τετράγωνο που βρίσκεται κάτω αριστερά είναι μπροστά από το άλλο. (c) Η διαφάνεια δεν φαίνεται. Φαίνεται σαν ένα ζωγραφισμένο σχέδιο ενός μόνο στρώματος. (d)  $p$ ,  $q$ ,  $r$ ,  $s$  είναι οι φωτεινότητες στις τέσσερις περιοχές που περιβάλλουν την  $X$  σύνδεση.

Τα ποιοτικά χαρακτηριστικά της διαφάνειας μπορούν να συσχετιστούν με τα ποιοτικά χαρακτηριστικά των  $X$  συνδέσεων. Έστω ότι τα  $p$ ,  $q$ ,  $r$ ,  $s$  είναι οι φωτεινότητες στις τέσσερις περιοχές που περιβάλλουν την  $X$  σύνδεση, όπως φαίνεται στην Εικόνα 16(d). Στην Εικόνα 16(a) ισχύει  $p < q$  και  $r < s$ , που σημαίνει ότι η κάθετη ακμή διατηρεί το ίδιο σήμα και στα δύο μισά της σύνδεσης  $X$ . Παρόμοια,  $p < r$  και  $q < s$ , που σημαίνει ότι η οριζόντια ακμή, επίσης διατηρεί το ίδιο σήμα και στα δύο μισά της σύνδεσης  $X$ . Αυτή η σύνδεση θα λέγεται «μη-ανάστροφη» γιατί και οι δύο ακμές διατηρούν το σήμα τους. Στην Εικόνα 16(b) η κάθετη ακμή αλλάζει το σήμα δια μέσου του  $X$  συνδέσμου, ενώ η οριζόντια ακμή διατηρεί το σήμα της. Αυτή η σύνδεση θα λέγεται «προς μία κατεύθυνση ανάστροφη». Τελικώς, στην Εικόνα 16(c) και η οριζόντια και η κάθετη ακμή αλλάζει το σήμα δια μέσου της σύνδεσης  $X$ . Αυτή θα λέγεται «προς δύο κατευθύνσεις ανάστροφη».

Το ανθρώπινο οπτικό σύστημα φαίνεται να χρησιμοποιεί ευριστικές μεθόδους που σχετίζονται με αυτές τις διαφορετικές κατηγορίες της σύνδεσης  $X$ . Οι «μη-ανάστrofες» συνδέσεις υποστηρίζουν την αντίληψη για διαφάνεια, αφήνοντας όμως τη διάταξη βάθους των επιπέδων διαφορούμενη. Οι «προς μια κατεύθυνση ανάστrofες» συνδέσεις επίσης υποστηρίζουν τη διαφάνεια, και επιπλέον επιβάλλουν μία μοναδική διάταξη βάθους. Οι «προς δύο κατευθύνσεις ανάστrofες» συνδέσεις δεν υποστηρίζουν διαφάνεια.

### Υπολογιστική ανάλυση

Η διαφάνεια θα εξεταστεί και από υπολογιστική άποψη, για την κατανόηση των ευριστικών μεθόδων που περιγράφηκαν παραπάνω. Αρχικά χρησιμοποιείται ένα

πλαίσιο για τον χαρακτηρισμό του συνδυασμού των διαφανών επιπέδων, και έστω τα επίπεδα είναι τα  $I_1, I_2, \dots, I_n$ . Κάθε επίπεδο μπορεί να μειώσει τη φωτεινότητα από το επίπεδο που βρίσκεται από κάτω του με ένα παράγοντα  $a$ ,  $0 < a \leq 1$ , και μπορεί να συμβάλει στην εκπομπή μιας ποσότητας  $e$ ,  $e \geq 0$ . Η μείωση και η εκπομπή είναι συναρτήσεις της θέσεως  $a(x,y)$ ,  $e(x,y)$ . Μια «άδεια» περιοχή έχει  $a=1$  και  $e=0$ .

Αν το επίπεδο  $n-1$  έχει ένα πρότυπο φωτεινότητας  $I_{n-1}(x,y)$ , τότε το πρότυπο φωτεινότητας στο επίπεδο  $n$  είναι:

$$I_n(x,y) = a_n(x,y) \cdot I_{n-1}(x,y) + e_n(x,y) \quad (26)$$

Ένα ευρύ φάσμα τιμών για τις  $(p, q, r, s)$  έχουν ως αποτέλεσμα επιτρεπτά παραδείγματα διαφάνειας. Παρόλα αυτά υπάρχουν και κάποιοι περιορισμοί. Τα επιτρεπτά όρια διακύμανσης για τα  $a$  και  $e$  υποδηλώνουν ότι μια «γεμάτη» περιοχή πρέπει να μειώσει ή να αφήσει αμετάβλητο το εύρος της διακύμανσης της φωτεινότητας σε ένα πιο χαμηλό επίπεδο. Έστω ότι η σύνδεση  $X$  προκύπτει από την επικάλυψη «γεμάτων» περιοχών σε δύο επίπεδα, και μένει να καθοριστεί αν η ακμή του μπροστινού επιπέδου είναι κάθετη ή οριζόντια, και ποιο μισό της ακμής είναι γεμάτο.

Οι τέσσερις πιθανές τοπικές υποθέσεις για την «γεμάτη» μπροστινή περιοχή είναι: (i) βρίσκεται πάνω από την οριζόντια γραμμή, (ii) βρίσκεται κάτω από την οριζόντια γραμμή, (iii) βρίσκεται στα αριστερά της κάθετης γραμμής, (iv) βρίσκεται στα δεξιά της κάθετης γραμμής. Οι παραπάνω όροι μεταφράζονται για τον παράγοντα μείωσης στις παρακάτω συνθήκες ανισότητας,

1. Η υπόθεση (i) είναι φυσικά αληθοφανής αν  $0 < (p - q)/(r - s) \leq 1$ ,
2. Η υπόθεση (ii) είναι αληθοφανής αν  $0 < (r - s)/(p - q) \leq 1$ ,
3. Η υπόθεση (iii) είναι αληθοφανής αν  $0 < (p - r)/(q - s) \leq 1$ , και
4. Η υπόθεση (iv) είναι αληθοφανής αν  $0 < (q - s)/(p - r) \leq 1$ .

Το γεγονός ότι αυτά τα κλάσματα είναι μη αρνητικά οδηγεί στις ανάστροφης-ακμής (edge-reversal) ευριστικές μεθόδους που σημειώθηκαν παραπάνω. Κατά συνέπεια, μια ακμή που είναι διαφανώς φρασσόμενη δεν μπορεί να αντιστρέψει το σήμα, ενώ μια ακμή που είναι μπροστά μπορεί ή δε μπορεί να αντιστρέψει το σήμα. Συνεπάγεται ότι οι «προς τις δύο κατευθύνσεις» συνδέσεις έχουν δύο συνεπείς ερμηνείες, ενώ οι «προς μία κατεύθυνση» συνδέσεις έχουν μόνο μία. Μία «προς τις δύο κατευθύνσεις» σύνδεση θα απαιτούσε ότι και οι καθετές και οι οριζόντιες ακμές να είναι μπροστά από τις άλλες, το οποίο είναι αδύνατο. Συνεπώς, δεν υπάρχει διαφάνεια.

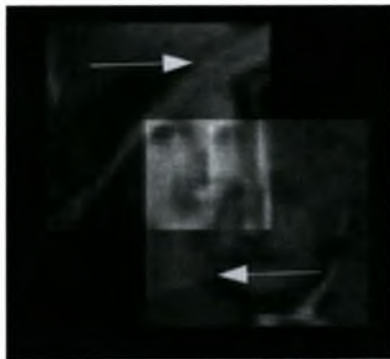
### 3.2 Διαφανής κίνηση (transparent motion)

Μια ιδιαίτερα σημαντική περίπτωση όπου η κατανομή των κατευθύνσεων των κινήσεων δεν παράγει αυτόματα ένα ενιαίο γενικό αποτέλεσμα συμβαίνει στην περίπτωση της διαφανής κίνησης. Όταν υπάρχει ένα αρκετά μεγάλο χάσμα στη κατανομή των τοπικών κατευθύνσεων, η κατανομή μπορεί να αναλυθεί σε δύο



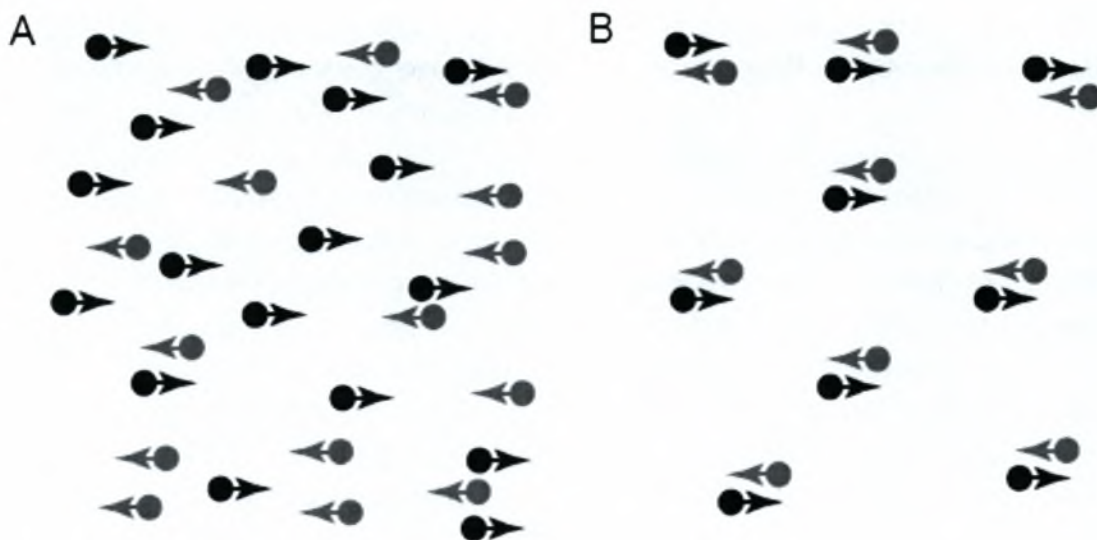
ξεχωριστές γενικές κινήσεις που φαίνονται ως χωρικά συμπεριεκτικές (co-extensive) και ως επιθετούμενες (superimposed) κατά τρόπο διαφανή. Το φαινόμενο της διαφάνειας είναι μια σημαντική περίπτωση δοκιμής για τη φύση του υπολογισμού και της αντιπροσώπευσης της κίνησης.

Η διαφανής κίνηση ορίζεται ως η αντίληψη για περισσότερα από έναν πεδία ταχύτητας στο ίδιο μέρος του οπτικού διαστήματος. Διαφανής κίνηση υπάρχει όποτε δύο διαφορετικές κινήσεις συμβαίνουν στην ίδια τοπική περιοχή της εικόνας. Είναι αρκετά συνηθισμένη σε φυσικές εικόνες, αφού μπορείς να τη βρεις υπό ποικίλους τρόπους μετακίνησης. Ένα προφανές παράδειγμα διαφανής κίνησης είναι η εικόνα που κάποιος λαμβάνει καθώς κοιτά δια μέσου του παραθύρου ενός κινούμενου οχήματος στη βροχή με το νερό να αυλακώνει τα τζάμια. Ένα λιγότερο εμφανές, αλλά περισσότερο συνηθισμένο, παράδειγμα διαφάνειας συμβαίνει όταν μια σκιά κινείται κατά μήκος ενός κατασκευασμένου φόντου. Αν μόνο ένα μοναδικό διάνυσμα κίνησης επιτρέπεται σε κάθε τοπική περιοχή της εικόνας, τότε μπορεί να παρατηρηθεί είτε το όριο της σκιάς που σέρνει τη σύσταση μαζί της, ή το όριο της σκιάς να παραμένει σταθερό μαζί με τη σύσταση. Ένα συνθετικό παράδειγμα παρουσίασης διαφανούς κίνησης μπορεί να κατασκευαστεί με την μετάθεση και πρόσθεση πολλών εικόνων μαζί, όπως φαίνεται στην Εικόνα 17.



**Εικόνα 17:** Πλαίσιο από μια ακολουθία εικόνας με δύο διαφανώς συνδυασμένες κινούμενες εικόνες. Κάθε εικόνα κινείται σε διαφορετική κατεύθυνση, όπως υποδεικνύεται από τα βέλη.

Οι διαφανείς κινήσεις είναι οι κινήσεις που προκύπτουν από δύο επιφάνειες ή αντικείμενα που επικαλύπτονται φυσικά έτσι ώστε κάποιος να μη μπορεί να προκαλέσει κατάτμηση των επιφανειών μόνο από βασισμένα στη θέση παραδείγματα (όπως φαίνεται στην Εικόνα 18). Αντίθετα από τα περισσότερα εργαστηριακά γεγονότα (stimuli), τα πραγματικά αντικείμενα κινούνται κανονικά κατά μήκος στοιβαγμένων περιβαλλόντων που, δυνάμει των δικών τους μετακινήσεων, οι κινήσεις της κεφαλής μας, και ο θόρυβός, θα μπορούσαν επίσης να συμβάλλουν με πιθανά σήματα κινήσεων στην ανάλυση της σκηνής. Ως εκ τούτου, η ανάγκη για αφαίρεση, ή τουλάχιστον κατάτμηση, των πρόσθετων σημάτων είναι σημαντική. Η διαφάνεια στην κίνηση επομένως παρέχει μια πολύτιμη περίπτωση δοκιμής για τα πιθανά πρότυπα της ενσωμάτωσης της κίνησης.



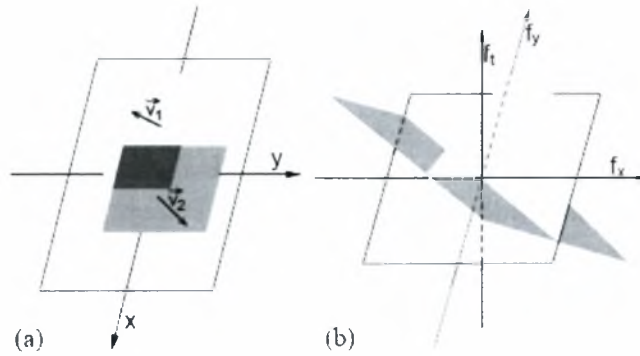
**Εικόνα 18:** Ένα γεγονός διαφανούς κίνησης. (A) Δύο σύνολα τυχαία τοποθετημένων σημείων που κινούνται σε διαφορετικές κατευθύνσεις και με διαφορετικές ταχύτητες φαίνονται να κινούνται ομαλά το ένα μέσω του άλλου. Σε τέτοια πειράματα τα σημεία δεν είναι συνήθως διαφορετικά χρωματισμένα, κατά συνέπεια, ο μόνος τρόπος για την κατάτμηση των δύο επιφανειών είναι μόνο μέσω των παραδειγμάτων κίνησης. (B) Σε αυτήν την περίπτωση τα σημεία είναι τοπικά ταιριασμένα, η διαφάνεια εξουδετερώνεται και μόνο ένα τρεμούλιασμα γίνεται αντιληπτό.

Η διαφανής κίνηση είναι ένα προκλητικό πρόβλημα από υπολογιστική άποψη, επειδή οι περισσότεροι αλγόριθμοι ανάλυσης κινούμενων εικόνων έχουν δυσκολίες με αυτή. Κάθε ένας περιλαμβάνει τον υπολογισμό των τοπικών συνιστωσών κίνησης και την ένωσή τους, για να εξομαλύνει ή να βρει το μέσο όρο του τοπικού πεδίου κίνησης. Αυτός ο υπολογισμός πραγματοποιείται για να βελτιώσει τα ποσοστά σήματος-θορύβου, για να παραποιήσουν την κίνηση σε περιοχές της εικόνας όπου υπάρχουν αραιά δεδομένα και για να λύσουν το πρόβλημα ανοιγμάτων (aperture problem). Ένα ατυχές αποτέλεσμα αυτού του υπολογισμού είναι ότι κάθε σημείο στην εικόνα μπορεί να έχει μόνο ένα μοναδικό διάνυσμα κίνησης ορισμένο σε αυτό και ως αποτέλεσμα αυτοί οι αλγόριθμοι δεν «αναγνωρίζουν» τη διαφανή κίνηση. Όταν στην κίνηση υπάρχει διαφάνεια, οι γκρι τιμές των διαφορετικών αντικειμένων επιθέτονται (τοποθετούνται το ένα πάνω στο άλλο) και η σταθερότητα φωτεινότητας των σημείων κατά μήκος των τροχιών της εικόνας τους, χρησιμοποιημένη για την εκτίμηση κίνησης, δεν ισχύει πλέον. Επιπλέον, δύο διαφορετικά διανύσματα κίνησης μπορεί να υπάρχουν στην ίδια χωρική θέση. Επομένως, μέθοδοι εκτίμησης κίνησης που αντιμετωπίζουν ρητά το ζήτημα της διαφάνειας πρέπει να αναπτυχθούν.

Στον χώρο των συχνотήτων, δύο διαφανή κινούμενα αντικείμενα συνοψίζονται σε δύο διατάξεις, αποκαλούμενες διατάξεις κίνησης (motion plans), των οποίων οι παράμετροι είναι οι ταχύτητες των αντικειμένων. Με τις μεθόδους συχνότητας (που βασίζονται στο εύρος ή φάση του μετασχηματισμού Fourier) είναι δυνατό να αντιμετωπιστούν τα φαινόμενα διαφάνειας.



Η εκτίμηση κίνησης βασισμένη στο μετασχηματισμό Fourier μιας ακολουθίας προσφέρει διάφορα πλεονεκτήματα: (1) Είναι εύρωστη σε γενικές αλλαγές φωτεινότητας. (2) Οι ανακρίβειες στην εκτίμηση κίνησης κοντά στα όρια του αντικειμένου αποφεύγονται, δεδομένου ότι οι εκτιμήσεις δεν είναι βασισμένες στις χωρικά τοπικές διαφορές φωτεινότητας ανάμεσα σε δύο συνεχόμενα πλαίσια (inter frame), π.χ. περιλαμβάνουν την οπτική ροή, και μάλλον, προέρχονται από τη γενική κατανομή έντασης. (3) Υπολογιστικά, πολλοί αλγόριθμοι είναι διαθέσιμοι για τον αποδοτικό υπολογισμό του μετασχηματισμού Fourier.



**Εικόνα 19:** (a) Παράδειγμα μιας ακολουθίας εικόνας που έχει δύο διαφανή κινούμενα αντικείμενα. (b) Συντελεστής του ιδανικού μετασχηματισμού Fourier

## 4. Μέθοδοι εκτίμησης διαφανούς κίνησης

Παρακάτω παρατίθενται δύο μέθοδοι που έχουν δημοσιευτεί για τον διαχωρισμό επιπέδων που περιέχουν διαφανή κίνηση καθώς και την εκτίμηση κίνησης για εικόνες με διαφάνεια.

### 4.1 Διαχωρισμός διαφανών επιπέδων με τη δια μέσου ανταλλαγή πληροφορίας ανάμεσα στα επίπεδα (separating transparent layers through layer information exchange)

Το πρόβλημα που εξετάζει αυτή η μέθοδος είναι ο διαχωρισμός δύο αυθαίρετα επιθετούμενων (superimposed) επιπέδων (είτε σε εικόνες, είτε σε βίντεο), χωρίς καμιά προηγούμενη γνώση για τη διαδικασία μίξης. Έχοντας δύο αρχικούς συνδυασμούς των προϊόντων μίξης  $I_1$  και  $I_2$ , που παράχθηκαν με ένα άγνωστο τρόπο, επιδιώκεται ένας διαχωρισμός των επιπέδων σε απεικονίσεις των  $L_1$  και  $L_2$  οι οποίες ελαχιστοποιούν τους δομικούς συσχετισμούς κατά μήκος των δύο επιπέδων σε κάθε σημείο της εικόνας. Τέτοιος διαχωρισμός επιτυγχάνεται με τη μεταφορά της τοπικής δομής από τη μια εικόνα στην άλλη εκεί που σχετίζεται πολύ με την υποκείμενη τοπική δομή της άλλης εικόνας, και αντίστροφα. Αυτή η αμφίδρομη λειτουργία μεταφοράς, που την ονομάζουν «ανταλλαγή πληροφορίας μεταξύ των επιπέδων» (layer information exchange), εκτελείται ελαττώνοντας τα μεγέθη των παραθύρων, από γενικά παράθυρα εικόνας (δηλαδή ολόκληρη η εικόνα) σε τοπικά παράθυρα εικόνας, ανιχνεύοντας κατά συνέπεια τις συσχετισμένες δομές στις ποικίλες κλίμακες κατά μήκος των θέσεων των εικονοκυττάρων.

#### Ομοιόμορφες συναρτήσεις μίξης

Υποθέτοντας ομοιόμορφες συναρτήσεις μίξης, έχουμε:

$$I_1(i) = a_1 \cdot L_1(i) + a_2 \cdot L_2(i) \text{ και } I_2(i) = b_1 \cdot L_1(i) + b_2 \cdot L_2(i)$$

Υπάρχει μια σταθερή ποσότητα  $\sigma$  τέτοια ώστε  $\hat{L}_1(i) = I_1(i) - \sigma I_2(i)$  θα περιέχει μόνο τη γεωμετρική δομή του  $L_1(i)$ , χωρίς κανένα ίχνος του  $L_2(i)$ .

Ο παράγοντας  $\sigma$  μπορεί να ανακτηθεί ελαχιστοποιώντας την παρακάτω αντικειμενική συνάρτηση:

$$\sigma = \operatorname{argmin}(\operatorname{GNGC}(I_2, I_1 - \sigma I_2)) \quad (28)$$

διαφοροποιώντας την παραπάνω αντικειμενική συνάρτηση ως προς  $\sigma$  και εξισώνοντας την με μηδέν (δηλαδή,  $\frac{\partial}{\partial \sigma} \operatorname{GNGC}(I_2, I_1 - \sigma I_2) = 0$ ) παράγει μια αναλυτική έκφραση για το  $\sigma$ :

$$\sigma = \frac{\sum_{i=0}^N C_i(I_1, I_2) \cdot V_i(I_2)}{\sum_{i=1}^N V_i^2(I_2)} \quad (29)$$

όπου  $C_i$  και  $V_i$  είναι οι τοπικές συνδιακυμάνσεις και διακυμάνσεις.

Στη συνέχεια υπολογίζεται και το δεύτερο επίπεδο με τον ίδιο τρόπο. Το δεύτερο επίπεδο:

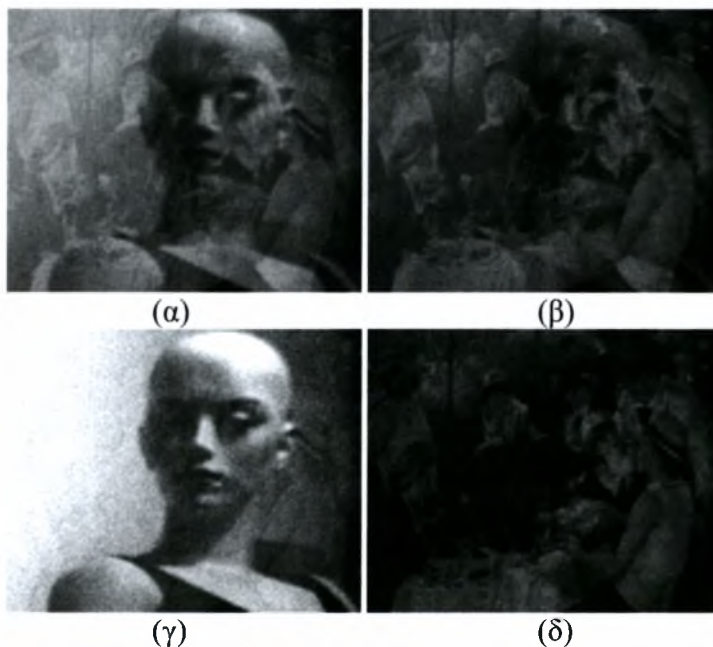
$$\hat{L}_2 = I_2 - \eta \hat{L}_1$$

Ανακτάται αναζητώντας το  $\eta$  που ελαχιστοποιεί το  $GNGC(\hat{L}_1, I_2 - \eta \hat{L}_1)$ .

### Γενίκευση για χωρικά μεταβαλλόμενες συναρτήσεις μίξης

Η γενική διαδικασία για ανταλλαγή επιπέδων που περιγράφηκε στην προηγούμενη ενότητα μπορεί να εφαρμοστεί σε μια μικρή περιοχή ανάλυσης για να υπολογίσει το  $\sigma(i)$  και το  $\eta(i)$  στο αντίστοιχο εικονοκύτταρο  $i$ . Αυτοί οι παράγοντες μεταφοράς υπολογίζονται επαναλαμβανόμενα για κάθε εικονοκύτταρο  $i = 1 \dots N$ , χρησιμοποιώντας ένα παράθυρο κεντροθετημένο γύρω από κάθε εικονοκύτταρο της εικόνας. Αυτό οδηγεί σε μια χωρικά μεταβαλλόμενη ανταλλαγή πληροφορίας μεταξύ των επιπέδων:

$$\hat{L}_1 = I_1(i) - \sigma(i)I_2(i) \text{ και } \hat{L}_2 = I_2(i) - \eta(i)\hat{L}_1(i).$$



**Εικόνα 20:** Ανάκτηση των διαφανών στρώματων από πολωμένες εικόνες. (α)-(β) Δύο αληθινές εικόνες που λαμβάνονται με διαφορετικές πολώσεις, που παρουσιάζουν την αντανάκλαση της Sheila σε έναν πίνακα του Renoir. (γ)-(δ) Τα ανακτημένα διαφανή επίπεδα χρησιμοποιώντας τη μέθοδο διαχωρισμού επιπέδων.

Η μέθοδος που παρουσιάστηκε δεν επεξεργάζεται ολόκληρο το πλαίσιο ταυτόχρονα, αλλά επεξεργάζεται μια μικρή περιοχή του πλαισίου κάθε φορά, επαναλαμβανόμενα για κάθε εικονοκύτταρο του πλαισίου. Αυτή η διαδικασία όμως είναι επιρρεπής σε τοπικές ανακρίβειες, όπως αλλαγές στη φωτεινότητα, ασυνέχειες στην κίνηση, «φράξιμο» ενός αντικειμένου από κάποιο άλλο. Συνεπώς οι εκτιμήσεις για τα όρια των επιπέδων και τον διαχωρισμό αυτών που προσφέρει αυτή η μέθοδος είναι μη αξιόπιστα. Αυτό φαίνεται και στην εικόνα 20(γ) και 20(δ), όπου ο διαχωρισμός δεν ήταν απόλυτα επιτυχής, γιατί σε κάθε επίπεδο διακρίνεται μέρος του

άλλου επιπέδου. Επίσης, το υπολογιστικό κόστος αυτής της μεθόδου είναι πολύ υψηλό, γιατί απαιτεί επεξεργασία κάθε εικονοκουττάρου ξεχωριστά για κάθε πλαίσιο.

## 4.2 Υπολογισμός «φρασσόμενων» και διαφανών κινήσεων (computing occluding and transparent motions)

Αυτή η μέθοδος χρησιμοποιείται για τον εντοπισμό και την παρακολούθηση «φρασσόμενων» και διαφανών κινούμενων αντικειμένων, που χρησιμοποιούν χρονική ολοκλήρωση (temporal integration) χωρίς να υποθέτουν σταθερότητα κίνησης. Κάθε νέο πλαίσιο στην ακολουθία συγκρίνεται με μια δυναμική εσωτερική περιγραφή της εικόνας του αντικειμένου που παρακολουθείται. Η εσωτερική περιγραφή της εικόνας κατασκευάζεται με την χρονική ολοκλήρωση των πλαισίων μετά την εγγραφή που είναι βασισμένη στον υπολογισμό της κίνησης. Η χρονική ολοκλήρωση διατηρεί την ευκρίνεια του αντικειμένου που παρακολουθείται θολώνοντας τα αντικείμενα που έχουν άλλες κινήσεις. Η σύγκριση νέων πλαισίων στην εσωτερική περιγραφή της εικόνας αναγκάζει τον αλγόριθμο ανάλυσης κίνησης να συνεχίσει να παρακολουθεί το ίδιο αντικείμενο στα επόμενα πλαίσια, και να βελτιώσει την κατάτμηση.

### Ανίχνευση των πολλαπλών κινούμενων αντικειμένων σε ζευγάρια εικόνας

Για την ανίχνευση διαφορετικώς κινούμενων αντικειμένων σε ένα ζευγάρι εικόνας μια μοναδική κίνηση υπολογίζεται αρχικά και το αντικείμενο που αντιστοιχεί σε αυτήν την κίνηση προσδιορίζεται. Αυτή η κίνηση αναφέρεται εδώ ως κυρίαρχη κίνηση, και το αντίστοιχο αντικείμενο το κυρίαρχο αντικείμενο. Μόλις ανιχνευθεί ένα κυρίαρχο αντικείμενο, αποκλείεται από την περιοχή της ανάλυσης και η διαδικασία επαναλαμβάνεται στην υπόλοιπη περιοχή για να βρεθούν άλλα αντικείμενα και οι κινήσεις τους.

### Παρακολούθηση του κυρίαρχου αντικειμένου

Μια περίληψη του αλγορίθμου για τον εντοπισμό και παρακολούθηση του κυρίαρχου αντικειμένου σε μια ακολουθία εικόνας είναι ο παρακάτω:

Για κάθε πλαίσιο στην ακολουθία (ξεκινώντας από  $t = 0$ ) κάνε:

1. Υπολόγισε τις παραμέτρους της κυρίαρχης κίνησης μεταξύ της εσωτερικής περιγραφής της εικόνας του αντικειμένου που παρακολουθείται  $Au(t)$  και του καινούριου πλαισίου  $I(t+1)$ , στην περιοχή  $M(t)$  του αντικειμένου που παρακολουθείται.
2. Στρέβλωσε την τρέχουσα εσωτερική περιγραφή της εικόνας  $Au(t)$  και την τρέχουσα μάσκα κατάτμησης  $M(t)$  προς το νέο πλαίσιο  $I(t+1)$  σύμφωνα με τις υπολογισμένες παραμέτρους κίνησης.
3. Προσδιόρισε τις στάσιμες περιοχές στις καταχωρημένες εικόνες, χρησιμοποιώντας την καταχωρημένη μάσκα  $M(t)$  ως μια αρχική εικασία. Αυτή θα είναι η κατατετμημένη περιοχή  $M(t+1)$  του αντικειμένου που παρακολουθείται στο πλαίσιο  $I(t+1)$ .
4. Υπολόγισε την ενημερωμένη εσωτερική περιγραφή της εικόνας  $Au(t+1)$  χρησιμοποιώντας την εξίσωση:

$$Au(0) \stackrel{\text{def}}{=} I(0)$$

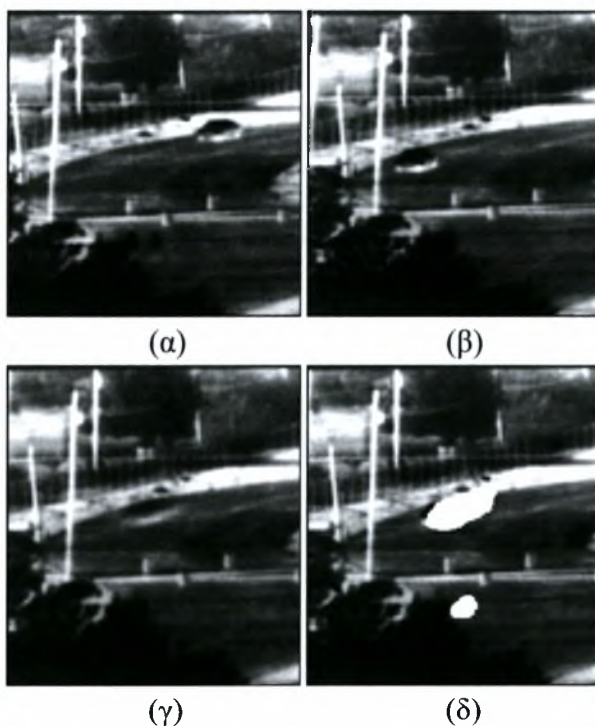
$$Au(t + 1) \stackrel{\text{def}}{=} (1 - w) \cdot I(t + 1) + w \cdot \text{register}(Au(t), I(t + 1))$$

και συνέχισε τη διαδικασία στο επόμενο πλαίσιο.

### Παρακολούθηση άλλων φρασσόμενων ή διαφανών αντικειμένων

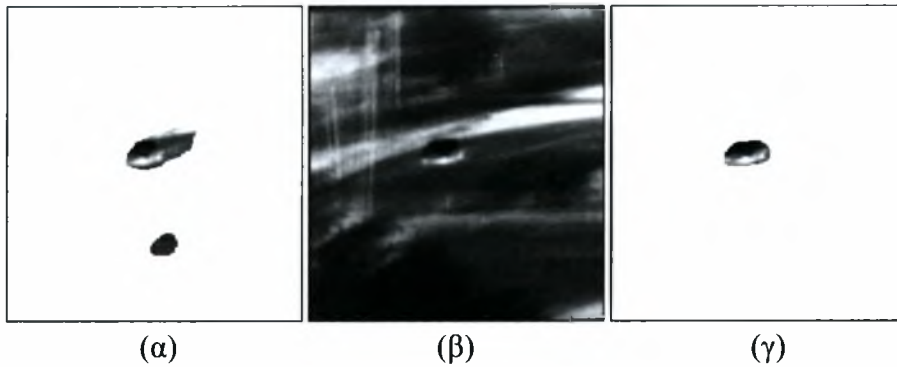
Μετά την ανίχνευση και την παρακολούθηση του πρώτου αντικειμένου, η προσοχή στρέφεται σε άλλα αντικείμενα. Αυτό γίνεται με την εφαρμογή του αλγορίθμου παρακολούθησης για ακόμα μια φορά, αυτή τη φορά στην υπόλοιπη εικόνα, μετά τον αποκλεισμό του πρώτου αντικειμένου που εντοπίστηκε από την περιοχή της ανάλυσης.

Το σχέδιο επαναλαμβάνεται κατ' επανάληψη, μέχρι να μην μπορεί να εντοπιστεί κανένα άλλο αντικείμενο.



**Εικόνα 21:** Εντοπισμός και παρακολούθηση του κυρίαρχου αντικειμένου σε μια υπέρυθρη ακολουθία εικόνας χρησιμοποιώντας χρονική ολοκλήρωση. (α)-(β) Το πρώτο και το τελευταίο πλαίσιο της ακολουθίας. Και το φόντο και το αυτοκίνητο κινούνται, και ένας άνθρωπος κινείται στο κάτω μέρος του δρόμου (εμφανίζεται σαν ένα σκούρο σημάδι στο χαμηλότερο μέρος της εικόνας α). (γ) Η εσωτερική απεικόνιση του παρακολουθούμενου αντικειμένου (το φόντο). Το φόντο παραμένει έντονο με λιγότερο θόρυβο, ενώ το αυτοκίνητο και ο πεζός θολώνουν. (δ) Το κατατετμημένο ακολουθούμενο αντικείμενο (το φόντο). Οι άσπρες περιοχές είναι αυτές που αποκλείστηκαν από την περιοχή που παρατηρείται.





**Εικόνα 22:** Εντοπισμός και παρακολούθηση του δεύτερου αντικειμένου χρησιμοποιώντας χρονική ολοκλήρωση. (α) Η αρχική περιοχή ανάλυσης μετά την εξαγωγή του πρώτου κυρίαρχου αντικειμένου. (β) Η εσωτερική αναπαράσταση του δεύτερου παρακολουθούμενου αντικειμένου (το αυτοκίνητο). Το αυτοκίνητο παραμένει έντονο ενώ το φόντο και ο πεζός θολώνουν. (γ) Κατάτμηση του παρακολουθούμενου αυτοκινήτου.

Η μέθοδος που αναπτύχθηκε παραπάνω για να επιλύσει το πρόβλημα προσπαθεί να βρει πρώτα ένα κυρίαρχο κινούμενο αντικείμενο σε ένα πλαίσιο, βγάζει αυτό το αντικείμενο από την εικόνα και συνεχίζει το ψάξιμο για να βρει το κυρίαρχο αντικείμενο (αν υπάρχει άλλο) στην εικόνα που μένει, επαναλαμβάνοντας τον αλγόριθμο. Συνεπώς, το υπολογιστικό κόστος αυτής της μεθόδου είναι αρκετά μεγάλο, καθώς απαιτεί το τρέξιμο του αλγορίθμου πολλές φορές για να βρει όλα τα κινούμενα αντικείμενα. Στη μέθοδο που θα αναλύσουμε παρακάτω όμως τα κινούμενα αντικείμενα εντοπίζονται και αφαιρούνται από την εικόνα όλα μαζί χωρίς να χρειάζεται η διαδικασία να επαναληφθεί παραπάνω από μία φορές.

Οι μέθοδοι που αναπτύχθηκαν παραπάνω προσπαθούν να επιλύσουν το πρόβλημα της εκτίμησης κίνησης εικόνων με διαφάνεια, αλλά δε το καταφέρνουν πλήρως.

Οι μέθοδοι του χώρου Fourier ή οι χωροχρονικές-του χώρου συχνότητας (spatiotemporal-frequency domain) μέθοδοι έχουν αναπτυχθεί για να συμπληρώσουν, να ξεπεράσουν τα προβλήματα των χωρικών μεθόδων. Αυτές οι μέθοδοι βασίζονται στη μετατόπιση της φάσης που εισάγεται στην αναπαράσταση του τομέα συχνότητας από τις χωρικές μετατοπίσεις. Επεξεργάζονται όλα τα πλαίσια ταυτόχρονα, οπότε είναι εύρωστες σε τοπικές ανακρίβειες, όπως τοπικό φράξιμο ενός αντικειμένου από κάποιο άλλο, αλλαγές στη φωτεινότητα. Η γενική, ταυτόχρονη επεξεργασία όλων των πλαισίων τις κάνει λιγότερο ευαίσθητες σε ομαλά χαρακτηριστικής ιδιότητας, υφής περιοχές και γενικά παράγουν καλύτερα αποτελέσματα. Επίσης, οι πολυάριθμες μέθοδοι που είναι διαθέσιμες για τον γρήγορο υπολογισμό των μετασχηματισμών συχνότητας, τις καταστεί υπολογιστικά αποδοτικές.

## 5. Εκτίμηση και κατάτμηση διαφανούς κίνησης

Στις προηγούμενες ενότητες παρουσιάσαμε μεθόδους για τον υπολογισμό της κίνησης, όπως επίσης και έρευνες (papers) που έχουν δημοσιευτεί για τον υπολογισμό και την κατάτμηση κίνησης για βίντεο με διαφάνεια. Σε αντίθεση με τις υπάρχουσες όμως μεθόδους, η μέθοδος που παρουσιάζουμε παρακάτω επιτρέπει την ταυτόχρονη εξαγωγή του φόντο και όλων των αντικειμένων που υφίστανται περιστροφές και μετατοπίσεις με τη βοήθεια μιας τεχνικής ελαχίστων τετραγώνων που πραγματοποιείται εξ ολοκλήρου στο χώρο Fourier.

### 5.1 Εκτίμηση μετατόπισης σε χώρο Fourier

Εδώ θα παρουσιάσουμε τη διατύπωση του προβλήματος για ένα βίντεο που περιέχει πολλαπλά αντικείμενα που μετατοπίζονται σε σχέση με ένα σταθερό φόντο, με τη μετατόπιση να μεταβάλλεται με το χρόνο. Κάθε πλαίσιο περιέχει  $M$  αντικείμενα, με φωτεινότητα  $s_i(\vec{r})$ ,  $1 \leq i \leq M$  στο εικονοκτύταρο  $\vec{r} = (x, y)$ , και μετατόπιση  $\vec{d}_i(k)$  από το πλαίσιο  $l$  στο πλαίσιο  $k$ . Ο μετασχηματισμός Fourier (FT) κάθε αντικειμένου  $i$  είναι  $S_i(\vec{\omega}) = |S_i(\vec{\omega})|e^{j\theta_i(\vec{\omega})}$ , όπου  $\vec{\omega} = [2\pi m/N_1, 2\pi n/N_2]^T$ ,  $m, n \in \mathbb{Z}$  είναι η δισδιάστατη χωρική συχνότητα,  $N_1 \times N_2$  το μέγεθος της εικόνας,  $|S_i(\vec{\omega})|$  το μέγεθος του μετασχηματισμού Fourier (FT), και  $\theta_i(\vec{\omega})$  η φάση του μετασχηματισμού. Κάθε πλαίσιο του βίντεο αναπαριστάται στο χώρο ως το άθροισμα του στατικού φόντο, που δηλώνεται με  $s_b(\vec{r})$ , και τα  $M$  αντικείμενα,  $s_i(\vec{r})$ ,  $1 \leq i \leq M$ , έτσι το πλαίσιο  $l$  είναι:

$$\alpha(\vec{r}, l) = s_b(\vec{r}) + \sum_{i=1}^M s_i(\vec{r}) + e_{\text{mod}}(\vec{r}, l) + u_{\text{noise}}(\vec{r}, l) \quad (30)$$

όπου  $\alpha(\vec{r}, l)$  δηλώνει τη τιμή της φωτεινότητας του πρώτου πλαισίου, στη θέση του εικονοκυττάρου  $\vec{r}$ . Στην πραγματικότητα, ένα προσθετικό μοντέλο για τα πλαίσια ενός βίντεο, δεν είναι εξ ολοκλήρου ακριβές: οι τιμές των εικονοκυττάρων του φόντο δεν προστίθενται στα εικονοκύτταρα του αντικειμένου στην περιοχή που βρίσκεται το αντικείμενο, αλλά καλύπτουν τις τιμές τους. Συνεπώς το προσθετικό μοντέλο της εξίσωσης (30) περιλαμβάνει ένα λάθος διαμόρφωσης (modeling error)  $e_{\text{mod}}$ , το οποίο αποτελείται από εικοκύτταρα του φόντο που καλύπτονται από τα αντικείμενα σε ένα πλαίσιο, και είναι ακάλυπτα σε άλλα πλαίσια. Στην περίπτωσή μας όμως εξετάζουμε αντικείμενα που έχουν διαφάνεια, δηλαδή δεν καλύπτουν το φόντο. Οπότε το λάθος διαμόρφωσης θα θεωρηθεί μηδενικό και θα καταλήξουμε στην εξίσωση (31).

Ο όρος  $u_{\text{noise}}$  αναπαριστά τη μέτρηση του θορύβου, που εισήχθει κατά τη διάρκεια της απόκτησης της εικόνας ή του βίντεο. Ο θόρυβος, που είναι μια τυχαία ποσότητα, έχει μοντελοποιηθεί στατιστικά στη βιβλιογραφία, και ποικίλες μέθοδοι έχουν αναπτυχθεί για την αφαίρεσή του. Εμείς θα θεωρήσουμε ότι τα δεδομένα που υποβάλλονται σε επεξεργασία είτε περιέχουν αμελητέα ποσότητα θορύβου, είτε ότι ο

θόρυβος έχει αφαιρεθεί, όπως στην περίπτωση σχετικής εργασίας αφαίρεσής του. Έτσι, ο θόρυβος  $u_{\text{noise}}$  δεν εμφανίζεται στις μεθόδους για την εκτίμηση κίνησης. Οι υπόλοιποι όροι στην εξίσωση (30) δεν είναι τυχαίοι, έτσι σαν επακόλουθο θα αναφερόμαστε σε ντετερμινιστικές ποσότητες για την εκτίμηση κίνησης. Το πλαίσιο  $k$ ,  $1 \leq k \leq N$  τότε είναι:

$$\alpha(\bar{r}, k) = s_b(\bar{r}) + \sum_{i=1}^M s_i(\bar{r} - \bar{d}_i(k)) \quad (31)$$

Θα πρέπει να σημειωθεί ότι ο αριθμός των αντικειμένων  $M$  είναι αρχικά άγνωστος, έτσι η μόνη πληροφορία του συστήματος που χρησιμοποιείται είναι τα πλαίσια του βίντεο. Για την ειδική περίπτωση των χρονικά συνεχών μετατοπίσεων ανάμεσα σε δύο συνεχόμενα πλαίσια (inter-frame), ο αριθμός των αντικειμένων μπορεί να εκτιμηθεί εκ των προτέρων. Για μετατοπίσεις που ποικίλλουν με το χρόνο, όπως στην εξίσωση (31), ο αριθμός των αντικειμένων εκτιμάται σε ένα μεταγενέστερο στάδιο, αλλά μετρώντας τον αριθμό των μετατοπίσεων που έχουν εξαχθεί. Η μέθοδος που προτείνουμε για την εκτίμηση της μετατόπισης βασίζεται στην ιδιότητα της μετατόπισης φάσης του μετασχηματισμού Fourier, όμοια με τις τεχνικές συσχετισμού φάσης, που έχουν χρησιμοποιηθεί για την εγγραφή της εικόνας. Στην περίπτωσή μας το πρόβλημα είναι περισσότερο περίπλοκο, λόγω της παρουσίας πολλαπλών κινούμενων αντικειμένων. Ο δισδιάστατος χωρικός μετασχηματισμός Fourier της εξίσωσης (31) είναι:

$$A(\bar{\omega}, k) = S_b(\bar{\omega}) + \sum_{i=1}^M S_i(\bar{\omega}) e^{-j\bar{\omega}^T \bar{d}_i(k)} \quad (32)$$

Έπειτα εξετάζουμε το λόγο των μετασχηματισμών Fourier των πλαισίων 1 και  $k$ :

$$\Phi(\bar{\omega}, k) = \frac{A(\bar{\omega}, k)}{A(\bar{\omega}, 1)} = \Gamma_b(\bar{\omega}) + \sum_{i=1}^M \Gamma_i(\bar{\omega}) e^{-j\bar{\omega}^T \bar{d}_i(k)} \quad (33)$$

όπου

$$\Gamma_b(\bar{\omega}) = \frac{S_b(\bar{\omega})}{A(\bar{\omega}, 1)}, \quad \Gamma_i(\bar{\omega}) = \frac{S_i(\bar{\omega})}{A(\bar{\omega}, 1)} \quad (34)$$

Στην εξίσωση (33), τα εκτοπίσματα εμφανίζονται σε ένα άθροισμα σταθμισμένων εκθετικών, έτσι ο ανάστροφος μετασχηματισμός Fourier του  $\Phi(\bar{\omega}, k)$  είναι:

$$\varphi_k(\bar{r}) = \gamma_b(\bar{r}) + \sum_{i=1}^M \gamma_i(\bar{r}) \delta(\bar{r} - \bar{d}_i(k)) \quad (35)$$

Δεδομένου ότι το  $\varphi_k(\bar{r})$  είναι ένα άθροισμα σταθμισμένων συναρτήσεων δέλτα, έχει  $M$  κορυφές στο  $\bar{r} = \bar{d}_i(k)$ , από το οποίο μπορούμε να εκτιμήσουμε τις  $M$  μετατοπίσεις. Πρακτικά, οι ψηφιακές εικόνες (πλαίσια) επεξεργάζονται, έτσι ο μετασχηματισμός Fourier είναι στην πραγματικότητα ένας διακριτός μετασχηματισμός Fourier(DDF), όποιος εφαρμόζεται μέσω του γρήγορου μετασχηματισμού Fourier(FFT). Επίσης, πρακτικά οι συναρτήσεις δέλτα είναι συναρτήσεις ώθησης, και για την εικόνα θεωρείται ότι περιοδικά κατασκευάζεται μια πανομοιότυπή της. Παρόλο αυτά, οι κορυφές γύρω από το  $\bar{d}_i(k)$  γίνονται ξεκάθαρες και παρέχουν αξιόπιστο εντοπισμό των πραγματικών εκτοπίσεων. Για περαιτέρω ενίσχυση της ευκρίνειας των κορυφών, εξετάζουμε το τετραγωνισμένο μέγεθος της εξίσωσης (35), που δίνεται από:

$$|\varphi_k(\bar{r})|^2 = \sum_{i=1}^M |\gamma_i(\bar{r})|^2 \delta^2(\bar{r} - \bar{d}_i(k)) + 2\Re \left[ \gamma_b(\bar{r}) + \sum_{i=1}^M \gamma_i(\bar{r}) \delta(\bar{r} - \bar{d}_i(k)) \right] + |\gamma_b(\bar{r})|^2 \quad (36)$$

Από την εξίσωση (36) βλέπουμε ότι οι κορυφές του  $|\varphi_k(\bar{r})|^2$  τώρα προέρχονται από όρους που περιλαμβάνουν συναρτήσεις δέλτα γύρω από το  $\bar{d}_i(k)$ , και επίσης από όρους τετραγωνισμένων συναρτήσεων δέλτα ( $\sum_{i=1}^M |\gamma_i(\bar{r})|^2 \delta^2(\bar{r} - \bar{d}_i(k))$ ). Οι όροι που υπολείπονται  $|\gamma_b(\bar{r})|^2$  στην εξίσωση (36) δεν υποβιβάζουν σημαντικά την ακρίβεια των κορυφών στο  $\varphi_k(\bar{r})$ , καθώς επηρεάζουν όλες τις συντεταγμένες  $\bar{r}$ , αλλά δεν έχουν καμιά ωστική φύση γύρω από κανένα συγκεκριμένο  $\bar{r}$ . Αυτό το γεγονός, συνδυασμένο με την παρουσία συναρτήσεων δέλτα γύρω από κάθε  $\bar{d}_i(k)$  μετά από τη μη γραμμική επεξεργασία, μας οδηγεί στο να περιμένουμε μια ενίσχυση των κορυφών γύρω από τα πραγματικά εκτοπίσματα. Επιπρόσθετα, η ευρωστία των βασισμένων στη φάση μεθόδων του μετασχηματισμού Fourier για την εγγραφή και την εκτίμηση του εκτοπίσματος επαληθεύεται από τα αποτελέσματα των πειραμάτων μας, όπου τα εκτοπίσματα εξάγονται με ακρίβεια στα πειράματα, συμπεριλαμβανομένου αληθινών ακολουθιών βίντεο.

Όταν η προτεινόμενη μέθοδος εξετάζει το μετασχηματισμό Fourier των πλαισίων  $l$  και  $k$ , βρίσκει το συνολικό εκτόπισμα  $\bar{d}_i(k)$  ανάμεσα στα δύο αυτά πλαίσια (για κάθε αντικείμενο  $i$ ), και εκτιμά τα αντίστοιχα εκτοπίσματα ανάμεσα σε συνεχόμενα πλαίσια ως  $\bar{d}_i(k)/(k-1)$ . Αν το πραγματικό εκτόπισμα μεταξύ των πλαισίων  $(1,2),(2,3),\dots,(k-1,k)$  δεν είναι σταθερό, δηλαδή δεν είναι ίσο με το  $\bar{d}_i(k)/(k-1)$ , επιπλέον επεξεργασία είναι απαραίτητη για να εξαχθούν οι σωστές, μεταβαλλόμενες με το χρόνο τιμές. Στην πράξη, μπορούμε να εκτιμήσουμε μεταβαλλόμενες με το χρόνο μετατοπίσεις είτε (α) με την εξαγωγή των μετατοπίσεων μεταξύ των πλαισίων που είναι κοντά το ένα με το άλλο, έτσι οι μετατοπίσεις μεταξύ συνεχόμενων πλαισίων είναι προσεγγιστικά συνεχείς, ή (β) με τον εντοπισμό των μετατοπίσεων μεταξύ δύο πλαισίων, δηλαδή του πλαισίου  $l$  και του  $k$ , και μετά να γίνεται η εκτίμηση των μετατοπίσεων μεταξύ των πλαισίων  $(l, \lfloor k/2 \rfloor)$  και  $(\lfloor k/2 \rfloor +$



1, k), μέχρι οι εκτιμήσεις των μετατοπίσεων σε αυτά τα μικρότερα διαστήματα να γίνουν ίσα μεταξύ τους<sup>1</sup>.

Ο προτεινόμενος αλγόριθμος εκτίμησης της μετατόπισης είναι υπολογιστικά αποδοτικός εξαιτίας των πολλών αλγορίθμων που είναι διαθέσιμοι για την γρήγορη εκτίμηση του μετασχηματισμού Fourier, δηλαδή ο γρήγορος μετασχηματισμός Fourier ή οι παραλλαγές του. Για την εκτίμησης της μετατόπισης μεταξύ N πλαισίων μεγέθους  $L = N_1 \times N_2$ , χρειαζόμαστε να υπολογίσουμε το γρήγορο μετασχηματισμός Fourier όλων των πλαισίων. Αυτό έχει υπολογιστική πολυπλοκότητα τάξεως  $O(N \cdot L \log_2 L)$ . Για την εξίσωση (36), υπολογίζουμε το τετραγωνισμένο μέγεθος των λόγων των γρήγορων μετασχηματισμών Fourier, το οποίο προσθέτει L πολλαπλασιασμούς και προσθέσεις, και ο αντίστροφος μετασχηματισμός επίσης προσθέτει  $O(L \log_2 L)$  υπολογισμούς, επομένως η πολυπλοκότητα είναι  $O(L^2 \log_2 L)$ . Η εκτίμηση της κίνησης απαιτεί τον εντοπισμό των κορυφών του  $\varphi_k(\bar{r})$ , το οποίο συνεπάγεται την αναζήτηση σε όλες τις L τιμές αυτής της επιφάνειας (που έχει ήδη εκτιμηθεί) και μπορεί να γίνει υπολογιστικά δαπανηρή για μεγάλα πλαίσια.

## 5.2 Μια ελαχίστων τετραγώνων κατάτμηση για μετατόπιση στο χώρο της συχνότητας

Σε αυτή την ενότητα, παρουσιάζουμε μια πρωτότυπη μέθοδο που επιτυγχάνει την ταυτόχρονη εξαγωγή του φόντο και των πολλαπλών, ανεξάρτητα μετατοπιζόμενων αντικειμένων σε ένα πλαίσιο, χρησιμοποιώντας δεδομένα συχνότητας. Αναπαριστούμε την ακολουθία όπως στην προηγούμενη ενότητα, όπου ο μετασχηματισμός Fourier κάθε πλαισίου δίνεται από την εξίσωση (32), για κάθε συχνότητα  $\bar{\omega}$  (υπάρχουν συνολικά  $N_1 \cdot N_2$  συχνότητες) στο πλαίσιο k, ( $1 \leq k \leq N$ ). Στοιβάζοντας τους μετασχηματισμούς Fourier όλων των N πλαισίων βίντεο, για οποιαδήποτε συχνότητα  $\bar{\omega}$ , επιτυγχάνουμε το γραμμικό σύστημα:

$$\mathbf{A} = \mathbf{G}\mathbf{S} \quad (37)$$

όπου  $\mathbf{A} = [A(\bar{\omega}, 1), A(\bar{\omega}, 2), \dots, A(\bar{\omega}, N)]^T$  είναι οι  $N \times 1$  μετασχηματισμοί Fourier των πλαισίων, το  $(M + 1) \times 1$  διάνυσμα  $\mathbf{S} = [S_b(\bar{\omega}), S_1(\bar{\omega}), \dots, S_M(\bar{\omega})]^T$  περιέχει τους μετασχηματισμούς Fourier από το φόντο και του αντικειμένου, στη συχνότητα  $\bar{\omega}$ , και  $\mathbf{G}$  είναι ένας  $N \times (M + 1)$  πίνακας με την πληροφορία της κίνησης, που δίνεται από:

$$\mathbf{G} = \begin{pmatrix} 1 & 1 & \dots & 1 \\ 1 & e^{-j\bar{\omega}^T \bar{d}_1(2)} & \dots & e^{-j\bar{\omega}^T \bar{d}_M(2)} \\ \vdots & \ddots & \vdots & \vdots \\ 1 & e^{-j\bar{\omega}^T \bar{d}_1(N)} & \dots & e^{-j\bar{\omega}^T \bar{d}_M(N)} \end{pmatrix} \quad (38)$$

<sup>1</sup> Η παράσταση  $\lfloor k/2 \rfloor$  αναπαριστά το κοντινότερο μικρότερο ακέραιο αριθμό της  $k/2$



για κάθε συχνότητα  $\bar{\omega}$ . Η πρώτη γραμμή του πίνακα  $\mathbf{G}$  περιέχει άσους επειδή υπάρχει μηδενικό εκτόπισμα από το πλαίσιο 1 στο πλαίσιο 1. Το γραμμικό σύστημα της εξίσωσης (37) είναι σχεδόν πάντα υπερπροσδιορισμένο, επειδή ο αριθμός των πλαισίων  $N$  που είναι διαθέσιμα είναι συνήθως μεγαλύτερος από τον αριθμό των κινούμενων αντικειμένων  $M$ . Μόλις οι μετατοπίσεις  $\bar{d}_i(k)$  του αντικειμένου εκτιμηθούν, οι μόνοι άγνωστοι στην εξίσωση (37) είναι οι  $\mathbf{S}$ , στη συχνότητα  $\bar{\omega}$ .

Μόλις οι μετατοπίσεις του αντικειμένου υπολογιστούν, εξάγουμε το φόντο και τα κινούμενα αντικείμενα λύνοντας το γραμμικό σύστημα της εξίσωσης (37) για κάθε μία από τις  $N_1 \times N_2$  συχνότητες  $\bar{\omega}$  ξεχωριστά. Υπολογίζουμε τον  $N \times (M + 1)$  πίνακα  $\mathbf{G}$  της εξίσωσης (38) για κάθε  $\bar{\omega}$  και για τις εξαγόμενες μετατοπίσεις  $\bar{d}_i(k)$ . Το αντικαθιστούμε στην εξίσωση (37), μαζί με το  $N \times 1$  διάνυσμα  $\mathbf{A}$ , περιέχοντας τις  $N$  τιμές μετασχηματισμού Fourier κάθε πλαισίου σε αυτή την  $\bar{\omega}$ , και λύνουμε το γραμμικό σύστημα που προκύπτει όπως θα λύναμε ένα σύστημα ελαχίστων τετραγώνων. Αυτό οδηγεί στο διάνυσμα  $\mathbf{S}$ , το οποίο περιέχει τους υπολογισμούς των μετασχηματισμών Fourier για το φόντο και τα κινούμενα αντικείμενα, για αυτήν τη συχνότητα  $\bar{\omega}$ . Επαναλαμβάνοντας αυτήν την διαδικασία για όλες τις  $N_1 \times N_2$  συχνότητες, παίρνουμε το δισδιάστατο μετασχηματισμό Fourier του φόντο,  $S_b$ , και των  $M$  αντικειμένων  $S_i$ . Η χωρική αναπαράσταση του φόντο και των αντικειμένων μπορεί αμέσως να ληφθεί από τον ανάστροφο μετασχηματισμό Fourier του  $S_b$  και κάθε  $S_i$ . Πρέπει να σημειωθεί ότι αυτή η διαδικασία δεν είναι υπολογιστικά δαπανηρή, αφού το σύστημα ελαχίστων τετραγώνων λύνεται χρησιμοποιώντας την Μοναδικής Τιμής Αποσύνθεση (Singular Value Decomposition), για την οποία υπάρχουν αρκετές υπολογιστικά ικανές μέθοδοι.

1) *Υπολογίζοντας τον αριθμό των κινούμενων αντικειμένων:* Η ακολουθία βίντεο αναπαριστάται στο χώρο ως ένα άθροισμα του φόντο, των  $M$  κινούμενων αντικειμένων. Η μέθοδός μας δεν απαιτεί προηγούμενη γνώση του αριθμού των κινούμενων αντικειμένων, αφού ο αριθμός τους μπορεί να εξαχθεί ταυτόχρονα με τις μετατοπίσεις, απλά μετρώντας τον αριθμό των κορυφών στην εξίσωση (36). Δεδομένου ότι υπάρχουν πολλά αντικείμενα στη σκηνή, και το φόντο είναι παρόν, οι κορυφές στο  $\bar{r} = \bar{d}_i(k)$  περιβάλλονται από πλευρικούς λοβούς. Στην πράξη αυτό δεν εισάγει ανακρίβειες, γιατί οι προσκληθείσες από την κίνηση κορυφές αντιστοιχούν σε συναρτήσεις δέλτα, οι οποίες είναι σημαντικά πιο έντονες από ότι ο θόρυβος γύρω από κάθε θέση  $\bar{r} = \bar{d}_i(k)$ .

Για χρονικά σταθερές μετατοπίσεις ανάμεσα σε συνεχόμενα πλαίσια, ο αριθμός των μετακινούμενων αντικειμένων μπορεί να εξαχθεί από πριν, πριν υπολογιστούν οι μετατοπίσεις τους. Σε αυτήν την περίπτωση, το εκτόπισμα από το πλαίσιο 1 στο πλαίσιο  $k$  είναι  $\bar{d}_i(k) = (k - 1)\bar{d}_i$ , όπου  $\bar{d}_i$  είναι το σταθερό εκτόπισμα ανάμεσα σε συνεχόμενα για το αντικείμενο  $i$ . Τότε ο πίνακας  $\mathbf{G}$  της εξίσωσης (38) γίνεται πίνακας Vandermonde για κάθε συχνότητα  $\bar{\omega}$ , δηλαδή τα στοιχεία σε κάθε μία από τις γραμμές του σχηματίζουν μια γεωμετρική πρόοδο, όπως ακολουθεί:

$$\mathbf{G} = \begin{pmatrix} 1 & 1 & \dots & 1 \\ 1 & e^{-j\bar{\omega}^T \bar{d}_1} & \dots & e^{-j\bar{\omega}^T \bar{d}_M} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & e^{-j(N-1)\bar{\omega}^T \bar{d}_1} & \dots & e^{-j(N-1)\bar{\omega}^T \bar{d}_M} \end{pmatrix} \quad (39)$$

Ο πίνακας αυτοσυσχέτισης για τους μετασχηματισμούς Fourier για κάθε πλαίσιο (σε μια σταθερή συχνότητα  $\bar{\omega}$ )  $\mathbf{A} = \mathbf{G}\mathbf{S}$  μπορεί να εκφραστεί από το  $\mathbf{R}_A = \mathbf{G}\mathbf{R}_S\mathbf{G}^H$ , όπου  $\mathbf{R}_S$  είναι ο πίνακας συσχέτισης του  $\mathbf{S}$ , δηλαδή οι μετασχηματισμοί Fourier του φόντο και των αντικειμένων. Η τάξη του  $\mathbf{R}_A$  είναι ίση με την τάξη του  $\mathbf{G}$ . Για σταθερές μετατοπίσεις ανάμεσα σε δύο συνεχόμενα πλαίσια, ο  $\mathbf{G}$  έχει  $M+1$  ανεξάρτητες στήλες, έτσι η τάξη του δίνει τον αριθμό των ανεξάρτητα κινούμενων αντικειμένων. Οι μοναδικές (singular) τιμές του χωρίς θόρυβο δεδομένων πίνακα συσχέτισης  $\mathbf{R}_A$  είναι  $M+1$ :  $\{\sigma_0^2, \sigma_1^2, \dots, \sigma_M^2\}$ . Όταν τυχαίος θόρυβος είναι παρόν, στη βιβλιογραφία υποτίθεται ότι είναι Προσθετικός, Λευκός και Γκαουσιανός, για απλότητα και χωρίς απώλεια γενικότητας. Οι περιπτώσεις μη-λευκού Προσθετικού Γκαουσιανού θορύβου και ακόμη γενικότεροι, έχουν μελετηθεί και έχουν αντιμετωπιστεί. Αν θεωρούσαμε ότι μπορεί να υπάρξει «φράξιμο» του φόντο από τα αντικείμενα, θα υπήρχε μη τυχαίος θόρυβος και θα προερχόταν από το λάθος διαμόρφωσης. Αλλά στην περίπτωσή μας τα αντικείμενα είναι διαφανή οπότε δεν καλύπτουν το φόντο και έτσι δεν εισάγεται κάποια ποσότητα λάθους στις τιμές  $\{\sigma_0^2, \sigma_1^2, \dots, \sigma_M^2\}$ . Η Μοναδικής Τιμής Αποσύνθεση (Singular Value Decomposition) του  $\mathbf{R}_A$  είναι:

$$\mathbf{R}_A = \mathbf{U}_A \begin{pmatrix} \sigma_0^2 & & & 0 \\ & \sigma_1^2 & & \\ & & \ddots & \\ 0 & & & \sigma_M^2 \end{pmatrix} \mathbf{V}_A^H \quad (40)$$

όπου  $\mathbf{U}_A$  και  $\mathbf{V}_A$  είναι οι πίνακες ιδιοδιανύσματος που αντιστοιχούν στο  $\mathbf{A}$ . η προηγούμενη ανάλυση εστιάζει στη δομή του πίνακα αυτοσυσχέτισης για μια σταθερή συχνότητα  $\bar{\omega}$ . Αυτό γίνεται επειδή η τάξη του  $\mathbf{R}_A$  προσδιορίζεται μόνο από τα εκτοπίσματα του αντικειμένου, δηλαδή παραμένει ίδια για κάθε  $\bar{\omega}$  (εκτός των τετριμμένων περιπτώσεων της  $\bar{\omega} = 0$  και  $\bar{\omega}^T \bar{d}_i(k) = 2\pi k$ ,  $k \in \mathcal{Z}$ ). Στην πράξη, υπολογίζουμε τον αριθμό των κινούμενων αντικειμένων από το  $\mathbf{R}_A$  για όλες τις συχνότητες  $\bar{\omega}$ , και κρατάμε τον υπολογισμό του  $M$  που επιστρέφεται από την πλειοψηφία των  $\mathbf{R}_A$ . Για να βρούμε τον αριθμό των κινούμενων αντικειμένων (για κάθε  $\mathbf{R}_A$ , δηλαδή για κάθε  $\bar{\omega}$ ) υπολογίζουμε τον αριθμό των μοναδικών τιμών στο 75<sup>ο</sup> εκατοστημόριο (δηλαδή, τις υψηλότερες 25% των μοναδικών τιμών). Στα πειράματά μας, αυτό αποδεικνύεται ότι είναι ένας αξιόπιστος υπολογισμός του αριθμού των κινούμενων αντικειμένων. Προηγούμενη γνώση του αριθμού των κινούμενων αντικειμένων βοηθάει στον έλεγχο του αριθμού των κορυφών που βρίσκονται από την εξίσωση (36), μπορεί επίσης να είναι μια παράμετρος ενδιαφέροντος για το αν, παραδείγματος χάριν, ενδιαφερόμασταν μόνο για το πόσα κινούμενα αντικείμενα βρίσκονται στη σκηνή και όχι για τον πλήρη χαρακτηρισμό των κινήσεών τους.

### 5.2.1 Συστηματοποίηση για την ελαχίστων τετραγώνων κατάτμηση της κίνησης

Η λύση ελαχίστων τετραγώνων της εξίσωσης (37) είναι η λύση που ελαχιστοποιεί το μέσο τετραγωνικό σφάλμα (Mean Squared Error)  $\|\mathbf{GS} - \mathbf{A}\|^2$  μεταξύ των χωρίς θόρυβο δεδομένων  $\mathbf{GS}$ , και των αληθινών δεδομένων  $\mathbf{A}$ . Η λύση ελαχίστων τετραγώνων δίνεται από την ανάλυση Μοναδικής Τιμής Αποσύνθεση (SVD) του  $\mathbf{G}$ , δηλαδή του  $\mathbf{G} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^H$ :

$$\hat{\mathbf{S}} = (\mathbf{G}^H\mathbf{G})^{-1}\mathbf{G}^H\mathbf{A} = \mathbf{V}\mathbf{\Sigma}^{-1}\mathbf{U}^H\mathbf{A} \quad (41)$$

όπου ο  $\mathbf{\Sigma}^{-1}$  είναι ένας διαγώνιος  $(M+1) \times (M+1)$  πίνακας με τιμές  $1/\sigma_i$  για  $\sigma_i \neq 0$ , και 0 για  $\sigma_i = 0$ . Αυτό είναι ένα ανάστροφο πρόβλημα, που είναι έμφυτα λάθος τοποθετημένο, δηλαδή οι λύσεις του είναι πολύ ευαίσθητες στο θόρυβο και πολύ εύκολα μπορούν να γίνουν ασταθείς.

Ο  $\mathbf{\Sigma}^{-1}$  είναι διαγώνιος με τιμές  $1/\sigma_i$ . Από την εξίσωση (41) βλέπουμε ότι οι μικρότερες μεμονωμένες τιμές  $\sigma_i$ , που αντιστοιχούν στα υψηλής συχνότητας στοιχεία του βίντεο, ενισχύουν πολύ τον όρο  $\mathbf{A}$ . Αυτό εμφανίζεται στη λύση για το διάνυσμα του αντικειμένου  $\hat{\mathbf{S}}$  υπό τη μορφή μεγάλων ταλαντώσεων, που την καθιστούν άχρηστη.

Μια πολύ γνωστή τεχνική που αντιμετωπίζει αυτήν την αστάθεια είναι ο αλγόριθμος συστηματοποίησης Tikhonov. Ουσιαστικά, η συστηματοποίηση Tikhonov επιβάλλει έναν περιορισμό στο μέγεθος της λύσης  $\mathbf{S}$ , για να εξαλείψει λύσεις με μεγέθη που τείνουν στο άπειρο. Συνεπώς, αντί να ελαχιστοποιούμε το  $\|\mathbf{GS} - \mathbf{A}\|^2$ , ελαχιστοποιούμε το  $\|\mathbf{GS} - \mathbf{A}\|^2 + \lambda\|\mathbf{S}\|^2$ , όπου  $\lambda$  είναι μια θετική σταθερά που ελέγχει το μέγεθος του διανύσματος της λύσης<sup>2</sup>. Η λύση ελαχίστων τετραγώνων γίνεται τότε:

$$\mathbf{S} = (\mathbf{G}^H\mathbf{G} + \lambda\mathbf{I})^{-1}\mathbf{G}^H\mathbf{A} = \sum_{i=1}^{M+1} \frac{\sigma_i}{\sigma_i^2 + \lambda} \bar{\mathbf{v}}_i \bar{\mathbf{u}}_i^H \mathbf{A} \quad (42)$$

έτσι η επίδραση του  $\sigma_i \cong 0$  αμβλύνεται από την παράμετρο συστηματοποίησης  $\lambda$ . σημειώστε ότι για  $\lambda = 0$ , η εξίσωση (42) ανάγεται στην λύση για την ιδανική (χωρίς θόρυβο) περίπτωση:

$$\mathbf{S} = (\mathbf{G}^H\mathbf{G})^{-1}\mathbf{G}^H\mathbf{A} = \sum_{i=1}^{M+1} \frac{1}{\sigma_i} \bar{\mathbf{v}}_i \bar{\mathbf{u}}_i^H \mathbf{A} \quad (43)$$

Από την εξίσωση (42) είναι προφανές ότι η συστηματοποιημένη λύση είναι περισσότερο σταθερή: όταν μια μεμονωμένη τιμή  $\sigma_i \rightarrow 0$ , η λύση δεν τείνει στο

<sup>2</sup> Χρησιμοποιούμε την  $L_2$  νόρμα  $\|\mathbf{x}\| = \sqrt{x_1^2 + \dots + x_N^2}$  για το  $N \times 1$  διάνυσμα  $\mathbf{x}$

άπειρο, λόγω της παραμέτρου συστηματοποίησης  $\lambda$ . Παρόλα αυτά, μεγάλες τιμές της  $\lambda$  επίσης μειώνουν την ακρίβεια των λύσεων ελαχίστων τετραγώνων, γιατί τότε, οι συστηματοποιημένες μεμονωμένες τιμές  $\sigma_i/\sigma_i^2 + \lambda$  θα παρεκκλίνουν περισσότερο από τις πραγματικές τιμές  $1/\sigma_i$ . Οι εκτιμήσεις ελαχίστων τετραγώνων γίνονται μικρότερες λόγω του  $\lambda$ , έτσι στο πρόβλημά μας, δηλαδή αυτό της εξαγωγής του αντικείμενου για την κατάτμηση της κίνησης, οι προκύπτουσες εκτιμήσεις για το αντικείμενο και το φόντο είναι πιο σκούρες από ότι στην πραγματικότητα. Κατά συνέπεια, υπάρχει ένας "συγκερασμός" στην επιλογή αυτής της παραμέτρου συστηματοποίησης. Επιπλέον, η ιδανική της τιμή δεν μπορεί να διευκρινιστεί από πριν, γιατί απαιτεί την γνώση της πραγματικής λύσης.

### 5.3 Ενσωμάτωση με χωρικούς υπολογισμούς

Οι εκτιμήσεις των αντικειμένων που παρουσιάστηκαν παραπάνω διαφέρουν από τις πραγματικές τους τιμές εξαιτίας της επίδρασης του φόντο και του προσεγγιστικού λάθους που εμφανίστηκε από τη συστηματοποίηση. Για να μειωθούν αυτά τα λάθη, επεξεργαζόμαστε τα αποτελέσματα από το χώρο συχνοτήτων με επιπλέον χωρική πληροφορία. Θα πρέπει να σημειωθεί ότι οι μέθοδοι που ακολουθούν βασίζονται σε στατιστικές ιδιότητες του χαμηλού τυχαίου θορύβου στα πλαίσια του βίντεο. Αυτός ο θόρυβος δε λήφθηκε υπόψη κατά τη διάρκεια του σταδίου εκτίμησης της κίνησης, καθώς είναι πολύ μικρός. Παρόλα αυτά, οι στατιστικές του ιδιότητες είναι χρήσιμες για την υλοποίηση του αλγορίθμου κατάτμησης κίνησης με έναν ηθικό, μη ειδικό τρόπο, που μπορεί να εφαρμοστεί σε οποιαδήποτε ακολουθία βίντεο.

#### 5.3.1 Συσχετισμός της λύσης ελαχίστων τετραγώνων και της αρχικής

Προτείνουμε να επεξεργαστούμε τις λύσεις του τομέα συχνοτήτων με χωρικά δεδομένα συσχετίζοντας κάθε λύση ελαχίστων τετραγώνων  $s_i(x, y)$  ( $1 \leq i \leq M$ ), στον χωρικό τομέα, με το αρχικό πλαίσιο  $\alpha(x, y, 1)$ . Η ομαλοποιημένη διαγώνια συσχέτιση (cross-correlation)  $c_b(x, y)$  του  $s_i(x, y)$  με το  $\alpha(x, y, 1)$ , υπολογισμένη σε μια τετράγωνη γειτονιά  $\mathcal{N}_b(x, y)$  γύρω από το εικονοκύτταρο  $(x, y)$ , πετυχαίνει υψηλές τιμές σε εικονοκύτταρα που ανήκουν στο αντικείμενο  $i$ , και χαμηλές τιμές σε οπουδήποτε αλλού. Όπως αναφέρθηκε στην ενότητα 5.1 (κάτω από την εξίσωση 30) η παρουσία τυχαίου θορύβου μέτρησης έχει αγνοηθεί σε άλλα στάδια της μεθόδου, γιατί είναι πολύ μικρός. Ωστόσο, αυτή η πηγή τυχειότητας μπορεί να χρησιμοποιηθεί για την κατανόηση της κατανομής των συντελεστών συσχέτισης  $c_b(x, y)$ . Συγκεκριμένα, οι χαμηλές τιμές του τυχαίου θορύβου μέτρησης χρησιμοποιούνται για να ερμηνεύσουν κάθε συντελεστή  $c_b(x, y)$  ως ένα άθροισμα τυχαίων μεταβλητών, των οποίων οι μεμονωμένες «διαφορές» (individual variances) είναι μικρές (λόγω των χαμηλών επιπέδων θορύβου). Δεδομένου ότι κάθε μεμονωμένη «διαφορά» είναι



μικρή συγκρινόμενη με το άθροισμα των «διαφορών», οι συνθήκες Lindeberg<sup>3</sup> ικανοποιούνται. Έπειτα, το Κεντρικό Θεώρημα Ορίου (Central Limit Theorem) τηρείται, και η κατανομή αυτών των ομαλοποιημένων αθροισμάτων συγκλίνουν στη Κανονική κατανομή (Normal distribution), εφόσον έχουμε ένα μεγάλο αριθμό τιμών συσχετισμού (μια συνθήκη που ικανοποιείται στην πράξη, αφού οι τιμές συσχετισμού είναι της ίδιας τάξης με τον αριθμό των εικονοκυττάρων του πλαισίου). Συνεπώς, οι συντελεστές  $c_b(x, y)$  υπολογισμένοι σε όλα τα  $(x, y)$ , οδηγούν σε ένα «χάρτη συσχέτισης»  $C(x, y)$ , του οποίου οι τιμές θεωρούνται ότι προσεγγίζουν μια κανονική κατανομή. Υποθέτουμε ότι οι συντελεστές είναι εργοδικοί (στο μέσο όρο και στη «διαφορά»), έτσι ο μέσος όρος της κατανομής προσεγγίζεται από δείγμα μέσου όρου  $\mu_b = \frac{1}{N} \sum_{x,y} c_b(x, y)$ , και η «διαφορά» από το δείγμα διαφοράς  $\sigma_b^2 = \frac{1}{N} \sum_{x,y} (c_b(x, y) - \mu_b)^2$ .

Για κανονικά καταναμημένους συντελεστές συσχετισμού, οι οποίοι έχουν τυποποιηθεί,  $c_s(x, y) = \frac{c_b(x,y) - \mu_b}{\sigma_b}$ , θεωρούμε ότι τα εικονοκύτταρα του πλαισίου με υψηλή τιμή συσχέτισης ανήκουν στο αντικείμενο που εξάχθηκε από την ελαχίστων τετραγώνων λύση:

$$\text{Prob}((x, y) \in \text{object}) = P(c_s(x, y) > \eta) = Q(\eta) \quad (44)$$

όπου  $Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-t^2/2} dt$  δίνει την πιθανότητα ουρών της κανονικής κατανομής. Το κατώφλι για τις τιμές συσχέτισης που είναι υψηλότερες από  $a\%$  δίνεται από  $\eta = Q^{-1}(a)$ . Υποθέτουμε ότι οι συντελεστές του 90<sup>ου</sup> εκατοστημορίου ( $a = 0.9$ ) ανήκουν στο κινούμενο αντικείμενο, ή ισοδύναμα, ότι  $P(c_s(x, y) > \eta) = a$  για  $a > 0.1$ . Συνεπώς, έχουμε μια γενική, του χωρικού τομέα μέθοδο, που συνδυάζει τα αποτελέσματα της κατάτμησης κίνησης στον τομέα συχνοτήτων με την πληροφορία φωτεινότητας, ούτως ώστε να καθοριστούν οι εξαγμένες περιοχές του αντικειμένου.

### 5.3.2 Περιοχές δραστηριότητας

Η διαδικασία συσχέτισης που περιγράφηκε παραπάνω συνδυάζει τις βασισμένες στη συχνότητα ελαχίστων τετραγώνων λύσεις για τα κινούμενα αντικείμενα με χωρικά δεδομένα, έτσι ώστε να βελτιωθούν οι εκτιμήσεις του αντικειμένου. Ωστόσο, η διαδικασία συσχέτισης μπορεί να μην είναι αρκετά ακριβής, γιατί μπορεί να μπερδεύει κάποια εικονοκύτταρα του φόντο για εικονοκύτταρα του αντικειμένου. Για αυτόν τον λόγο, αναπτύσσουμε άλλη χωρική μέθοδο που συμπληρώνει τα προηγούμενα παραγόμενα αποτελέσματα κατάτμησης. Το πλαίσιο  $k$  στρεβλώνεται με την εφαρμογή του αντιθέτου κάθε εκτίμησης κίνησης  $\bar{d}_i(k)$  στις θέσεις  $(x, y)$  του εικονοκυττάρου, και με τον υπολογισμό των τιμών φωτεινότητας

<sup>3</sup> Οι συνθήκες Lindeberg είναι οι συνθήκες που αφορούν τις «διαφορές» (variance) του αθροίσματος των αμοιβαίων ανεξάρτητων τυχαίων μεταβλητών που χρειάζεται να ικανοποιούνται έτσι ώστε το θεώρημα Κεντρικού ορίου (Central limit theorem) να μπορεί να χρησιμοποιηθεί. Ορίζουν ότι οι ξεχωριστές «διαφορές»  $\sigma_i^2$  είναι μικρές συγκρινόμενες με το  $\sum_{i=1}^n \sigma_i^2$



στις στρεβλωμένες θέσεις του εικονοκυττάρου μέσω κυβικής παρεμβολής. Όπως στην προηγούμενη ενότητα, σε αυτό το στάδιο λαμβάνουμε υπόψη το μικρό τυχαίο θόρυβο μέτρησης που είναι παρόν στο βίντεο. Έπειτα, μετά τη στρέβλωση, το πλαίσιο  $k$  γίνεται  $a_w(\bar{r}, k)$  και οι περιοχές στο  $a_w(\bar{r}, k)$  και  $a(\bar{r}, 1)$  που αντιστοιχούν στο αντικείμενο  $i$  θα διαφέρουν μόνο από το θόρυβο μέτρησης, ενώ τα μέρη του πλαισίου που έχουν λανθασμένα στρεβλωθεί θα διαφέρουν από μια υψηλότερη τιμή, που εξαρτάται από την ακολουθία. Ένα εικονοκύτταρο που έχει σωστά μετατοπιστεί στο στρεβλωμένο πλαίσιο  $k$  θα έχει φωτεινότητα:

$$a_w(\bar{r}, k) = a(\bar{r}, 1) + u_{\text{noise}}(\bar{r}, k) \quad (45)$$

όπου  $u_{\text{noise}}(\bar{r}, k)$  αναπαριστά τον τυχαίο θόρυβο μέτρησης, ενώ ένα λανθασμένα μετατοπισμένο εικονοκύτταρο είναι:

$$a_w(\bar{r}, k) = a(\bar{r}, 1) + w(\bar{r}, k) + u_{\text{noise}}(\bar{r}, k) \quad (46)$$

όπου  $w(\bar{r}, k)$  είναι μια άγνωστη τιμή, που εισάγεται από τη λανθασμένη στρέβλωση. Το πρόβλημα για το αν ένα εικονοκύτταρο ανήκει στο αντικείμενο  $i$  ή όχι μπορεί τώρα να διατυπωθεί ως μια δυαδική δοκιμή υπόθεσης:

$$\begin{aligned} H_0: d(\bar{r}, k) &= u_{\text{noise}}(\bar{r}, k) \\ H_1: d(\bar{r}, k) &= w(\bar{r}, k) + u_{\text{noise}}(\bar{r}, k) \end{aligned} \quad (47)$$

όπου  $d(\bar{r}, k) = a_w(\bar{r}, k) - a(\bar{r}, 1)$ . Κάτω από το  $H_0$ ,  $d(\bar{r}, k)$  ακολουθεί η κατανομή του θορύβου μέτρησης, που έχει μοντελοποιηθεί ως Γκαουσιανός, αλλά κάτω από το  $H_1$ , η κατανομή του αλλάζει σημαντικά, γιατί μία άγνωστη, εξαρτώμενη από το εικονοκύτταρο ποσότητα  $w(\bar{r}, k)$  προστίθεται στο  $d(\bar{r}, k)$ . Συνεπώς για να διευκρινίσουμε αν το  $d(\bar{r}, k)$  ανήκει στο  $H_0$  ή στο  $H_1$ , αρκεί να εξεταστεί η μη-γκαουσιανότητα (non-gaussianity) των δεδομένων. Το κλασικό μέτρο μη-γκαουσιανότητας μιας τυχαίας μεταβλητής  $y$  είναι η κύρτωση, που ορίζεται ως:

$$\text{kurt}(y) = E\{y^4\} - 3(E\{y^2\})^2 \quad (48)$$

Η τέταρτη στιγμή μιας γκαουσιανής τυχαίας μεταβλητής είναι  $E\{y^4\} = 3(E\{y^2\})^2$ , έτσι η κύρτωσή του είναι ίση με μηδέν. Μη μηδενικές τιμές της κύρτωσης  $\text{kurt}(d(\bar{r}, k))$  δείχνουν ότι το εικονοκύτταρο έχει λανθασμένα μετατοπιστεί, και μηδενικές (ή κοντα στο μηδέν τιμές) δείχνουν ότι το εικονοκύτταρο είχε σωστά μετατοπιστεί, και ανήκει στο αντικείμενο  $i$ .

Για να υπολογίσουμε την κύρτωση σε κάθε εικονοκύτταρο  $\bar{r}$ , χρειάζεται να υπολογίσουμε τους μέσους όρους του  $d(\bar{r}, k)$  σε όλα τα πλαίσια του βίντεο  $1 \leq k \leq N$ . Υποθέτουμε ότι οι διαφορές  $d(\bar{r}, k)$  σχηματίζουν μια ergodic διαδικασία, έτσι τα σύνολα των μέσων όρων τους μπορούν να προσεγγιστούν από τους χρονικούς μέσους

όρους τους. Συνεπώς, στην πράξη, υπολογίζουμε την κύρτωση κάθε εικονοκυττάρου  $\bar{r}$  από:

$$\text{kurt}(d(\bar{r})) = \frac{1}{N} \sum_{k=1}^N d(\bar{r}, k)^4 - 3 \left[ \frac{1}{N} \sum_{k=1}^N d^2(\bar{r}, k) \right]^2 \quad (49)$$

Αυτό οδηγεί σε έναν αξιόπιστο και ακριβή εντοπισμό των περιοχών δραστηριότητας.

## 5.4 Περιστρεφόμενες και μετατοπιζόμενες κινήσεις

Οι αναπαραστάσεις των εικόνων στον τομέα συχνοτήτων χρησιμοποιούνται για την ανάκτηση πιο περίπλοκων μετασχηματισμών, όπως περιστροφές συνδυασμένες με μετατοπίσεις. Τέτοιες μέθοδοι έχουν αναπτυχθεί κυρίως για την εγγραφή εικόνας, έτσι έχουν σχεδιαστεί για την ανάκτηση ενός μόνο μετασχηματισμού. Θα δείξουμε πως αυτές μπορούν ακόμα να χρησιμοποιηθούν για την ανάκτηση πολλών περιστρεφόμενων και μετατοπιζόμενων κινήσεων σε ακολουθίες από βίντεο. Όπως στην ενότητα 5.1 θα σχεδιάσουμε τον αλγόριθμο εκτίμησης κίνησης υποθέτοντας ότι ο θόρυβος μέτρησης είναι αμελητέος. Για να γίνουν οι υπολογισμοί της περιστροφής και της μετατόπισης περισσότερο ακριβείς, αφαιρούμε το φόντο από τα πλαίσια του βίντεο.

Υπάρχουν πολλές μέθοδοι για την αφαίρεση του φόντο, οι οποίοι μπορούν να χρησιμοποιηθούν για να λάβουμε ένα πιο ακριβές μοντέλο για κάθε πλαίσιο. Μέθοδοι που διαμορφώνουν το φόντο ως προϊόν Γκαουσιανών μίξεων, δίνουν πολύ καλά αποτελέσματα, αλλά απαιτούν ένα στάδιο κατάρτισης και έχουν υψηλό υπολογιστικό κόστος. Απλούστερες μέθοδοι, όπως το μεσαίο φιλτράρισμα (median filtering), παρέχει αποδεκτή ακρίβεια, με μικρό υπολογιστικό κόστος και περιορισμένες απαιτήσεις μνήμης. Συγκεκριμένα, τα εικονοκύτταρα του φόντο υπολογίζονται από το μέσο των τιμών φωτεινότητας κάθε εικονοκυττάρου από τα  $N$  εικονοκύτταρα. Όταν ένα κινούμενο αντικείμενο καλύπτει ένα εικονοκύτταρο  $\bar{r}$  του φόντο σε κάποια πλαίσια, η φωτεινότητα του  $\bar{r}$  είναι η υψηλότερη ή η χαμηλότερη (outlier), συγκρινόμενη με την τιμή του στα πλαίσια που δεν κρύβεται από το κινούμενο αντικείμενο (δηλαδή η πραγματική τιμή του φόντο). Ο μέσος των τιμών φωτεινότητας ενός εικονοκυττάρου σε  $N$  πλαίσια απορρίπτει τα σημεία που βρίσκονται μακριά από κεντροειδή (outliers) και κρατά μόνο τις τιμές του φόντο. Αυτό απαιτεί ότι κάθε εικονοκύτταρο του φόντο αποκαλύπτεται κατά τη διάρκεια της ακολουθίας βίντεο. Σε ακολουθίες χαμηλής ποιότητας, όπου το βίντεο παρουσιάζει τρεμουλιάσματα και η φωτεινότητά του αλλάζει σημαντικά, η «τοπική» πληροφορία του πλαισίου μπορεί επίσης να χρησιμοποιηθεί, για τη επίτευξη πιο ακριβούς μοντέλου του φόντου και για την αποτελεσματική αφαίρεσή του.

### 5.4.1 Ένα αντικείμενο

Θεωρείστε ένα πλαίσιο βίντεο με ένα αντικείμενο, το οποίο υποβάλλεται σε περιστροφή  $\theta$  γύρω από το κέντρο του πλαισίου, και μια επακόλουθη μετατόπιση  $\vec{d} = (d_x, d_y)$  μεταξύ του πρώτου και δεύτερου πλαισίου, με το φόντο να είναι απών. Η φωτεινότητά του  $a(x, y, 1)$  στο πρώτο πλαίσιο γίνεται  $a(x, y, 2) = a(x \cos \theta + y \sin \theta - d_x, -x \sin \theta + y \cos \theta - d_y)$  μετά την περιστροφή και τη μετατόπισή του. Αυτή η μετατόπιση του πλαισίου θα μπορούσε να είχε προκληθεί από μια μετατόπιση του  $(x_0, y_0)$ , ακολουθούμενη από την περιστροφή του αντικειμένου γύρω από το κέντρο του, όπως ακολουθεί:

$$\begin{aligned} a(x, y, 2) &= a((x - x_0) \cos \theta + (y - y_0) \sin \theta, -(x - x_0) \sin \theta + (y - y_0) \cos \theta) \\ &= a(x \cos \theta + y \sin \theta - (x_0 \cos \theta + y_0 \sin \theta), -x \sin \theta + y \cos \theta \\ &\quad - (-x_0 \sin \theta + y_0 \cos \theta)) \end{aligned} \quad (50)$$

Αυτό είναι ισοδύναμο με την πρώτη διατύπωση, με  $[d_x, d_y] = [x \cos \theta + y \sin \theta, -x_0 \sin \theta + y_0 \cos \theta]$ , έτσι στο επόμενο θα θεωρήσουμε, χωρίς απώλεια της γενικότητας, ότι το αντικείμενο πρώτα περιστρέφεται (γύρω από το κέντρο του πλαισίου) και μετά μετατοπίζεται. Ο μετασχηματισμός Fourier του πλαισίου 1 είναι  $A(\omega_x, \omega_y, 1)$  και για το πλαίσιο 2:

$$A(\omega_x, \omega_y, 2) = A(\omega_x \cos \theta + \omega_y \sin \theta, -\omega_x \sin \theta + \omega_y \cos \theta) e^{-j(\omega_x d_x + \omega_y d_y)} \quad (51)$$

Από την εξίσωση (51), είναι εμφανές ότι η μετατόπιση εμφανίζεται μόνο στη φάση του μετασχηματισμού Fourier. Συνεπώς, με τη λήψη του μεγέθους του μετασχηματισμού Fourier, η επίδραση της μετατοπιζόμενης κίνησης εξαλείφεται, και μόνο η πληροφορία για την περιστροφή παραμένει. Αν το μέγεθος του μετασχηματισμού Fourier του πρώτου πλαισίου στον πολικό τομέα εκφράζεται ως  $|A(\rho, \psi, 1)|$ , τότε το μέγεθος του δεύτερου πλαισίου είναι:

$$|A(\rho, \psi, 2)| = |A(\rho, \psi - \theta, 1)| \quad (52)$$

Για πολικές συντεταγμένες  $\omega_x = \rho \cos \psi$ ,  $\omega_y = \rho \sin \psi$ . Κατά συνέπεια, στο μέγεθος του μετασχηματισμού Fourier στις πολικές συντεταγμένες, η περιστροφή  $\theta$  γίνεται μια απλή μετατόπιση της  $\psi$  συντεταγμένης. Συνεπώς, το  $\theta$  μπορεί να ανακτηθεί από την εξίσωση (52) υπολογίζοντας τη μετατόπιση του  $\psi$ , στο  $|A(\rho, \psi, 1)|$ , στο  $\psi - \theta$  στο  $|A(\rho, \psi, 2)|$ . Η μετατόπιση υπολογίζεται χρησιμοποιώντας τη μέθοδο που περιγράφηκε στην ενότητα 5.1. Σημειώστε ότι αυτό απαιτεί το μετασχηματισμό του πλαισίου (1) από το χωρικό τομέα στον τομέα συχνότητας, (2) την εξαγωγή του μεγέθους του μετασχηματισμού Fourier, (3) τη διατύπωση του μεγέθους του μετασχηματισμού Fourier σε πολικές συντεταγμένες και (4) το μετασχηματισμό

Fourier του μεγέθους του πολικού μετασχηματισμού Fourier, για την εξαγωγή της γωνίας περιστροφής.

### 5.4.2 Πολλαπλά αντικείμενα

Θεωρούμε ένα βίντεο που περιέχει πολλαπλά αντικείμενα, που υποβάλλονται σε ένα συνδυασμό από περιστροφικές και μετατοπιζόμενες κινήσεις. Όπως περιγράφηκε παραπάνω, θεωρούμε, χωρίς απώλεια της γενικότητας, τα αντικείμενα πρώτα υποβάλλονται σε περιστροφές γύρω από το κέντρο των πλαισίων της εικόνας, και μετά σε μετατοπίσεις. Επίσης, το φόντο αφαιρείται για την επίτευξη αξιόπιστων υπολογισμών περιστροφής και μετατόπισης. Έπειτα, το προσθετικό μοντέλο  $a_{RT}(x, y, 1) = \sum_{i=1}^M s_i(x, y)$  για το πλαίσιο 1 είναι ακριβές και για τα πλαίσια  $k$ ,  $2 \leq k \leq N$  το προσθετικό μοντέλο είναι:

$$a_{RT}(x, y, k) = \sum_{i=1}^M s_i(x \cos(\theta_i(k)) + y \sin(\theta_i(k)) - d_i^x(k), -x \sin(\theta_i(k)) + y \cos(\theta_i(k)) - d_i^y(k)) \quad (53)$$

όπου το  $\theta_i(k)$  δίνει την περιστροφή κάθε αντικειμένου από το πλαίσιο 1 στο πλαίσιο  $k$  και  $\bar{d}_i(k) = [d_i^x(k), d_i^y(k)]$  δίνει τη μετατόπισή του. Ο μετασχηματισμός Fourier της παραπάνω εξίσωσης δίνεται από την:

$$A_{RT}(\omega_x, \omega_y, k) = \sum_{i=1}^M S_i(\omega_x \cos(\theta_i(k)) + \omega_y \sin(\theta_i(k)), -\omega_x \sin(\theta_i(k)) + \omega_y \cos(\theta_i(k))) e^{-j(\omega_x d_i^x(k) + \omega_y d_i^y(k))} \quad (54)$$

Σε πολικές συντεταγμένες  $\omega_x = \rho \cos \psi$ ,  $\omega_y = \rho \sin \psi$  έχουμε:

$$A_{RT}(\rho, \psi, k) = \sum_{i=1}^M S_i(\rho, \psi - \theta_i(k)) e^{-jQ(\rho, \psi - \theta_i(k))} \quad (55)$$

όπου  $Q(\rho, \psi - \theta_i(k)) = a \cos(\psi - \theta_i(k)) + \beta \sin(\psi - \theta_i(k))$ , για  $a = \rho[d_i^x(k) \cos \theta_i(k) + d_i^y(k) \sin \theta_i(k)]$  και  $\beta = \rho[d_i^y(k) \cos \theta_i(k) - d_i^x(k) \sin \theta_i(k)]$  (Παράρτημα 1). Θέτοντας  $P_i(\rho, \psi) = S_i(\rho, \psi) e^{-jQ(\rho, \psi)}$ , πετυχαίνουμε:

$$A_{RT}(\rho, \psi, k) = \sum_{i=1}^M P_i(\rho, \psi - \theta_i(k)) \quad (56)$$

Μετά από αυτούς τους μετασχηματισμούς, οι πολλαπλές περιστροφές στο πλαίσιο  $k$  αναπαριστώνται από  $M$  «μετατοπίσεις»  $\theta_i(k)$  της πολικής συντεταγμένης  $\psi$ , η οποία μπορεί εύκολα να υπολογιστεί από τη φάση του μετασχηματισμού Fourier του όπως στην ενότητα 5.1.

Μόλις οι γωνίες περιστροφής  $\theta_i(k)$  εξαχθούν, χρειάζεται επίσης να υπολογίσουμε τις αντίστοιχες μετατοπίσεις  $\bar{d}_i(k)$ . Για κάθε αντικείμενο  $i$ , ( $1 \leq i \leq M$ ), το πλαίσιο  $k$  «από-περιστρέφεται» από τη γωνία  $\theta_i(k)$  υπό εξέταση. Ως αποτέλεσμα, στο «από-περιστρεφόμενο» πλαίσιο, το αντικείμενο  $i$  έχει υποβληθεί σε καθαρή μετατόπιση, ενώ τα άλλα αντικείμενα έχουν υποβληθεί σε μια καινούρια περιστροφή  $\theta_j(k) - \theta_i(k)$  (για  $i \neq j$ ), εκτός από τη μετατόπιση. Μετά την «από-περιστροφή» από τη  $\theta_i(k)$  (που είναι ισοδύναμο με το να περιστραφεί με γωνία  $-\theta_i(k)$ ) το πλαίσιο  $k$  γίνεται:

$$a'_{RT}(x, y, k) = s_i \left( x - T_i^x(k), y - T_i^y(k) \right) + \sum_{j=1, j \neq i}^M s_j \left( x' - t_j^x(k), y' - t_j^y(k) \right) \quad (57)$$

όπου μετά από μερική άλγεβρα, παίρνουμε:

$$\begin{aligned} T_i^x(k) &= d_i^x(k) \cos(\theta_i(k)) - d_i^y(k) \sin(\theta_i(k)) \\ T_i^y(k) &= d_i^x(k) \sin(\theta_i(k)) + d_i^y(k) \cos(\theta_i(k)) \\ x' &= x \cos(\theta_j(k) - \theta_i(k)) + y \sin(\theta_j(k) - \theta_i(k)) \\ y' &= -x \sin(\theta_j(k) - \theta_i(k)) + y \cos(\theta_j(k) - \theta_i(k)) \\ t_j^x(k) &= d_j^x(k) \cos \theta_i(k) - d_j^y(k) \sin \theta_i(k) \\ t_j^y(k) &= d_j^x(k) \sin \theta_i(k) + d_j^y(k) \cos \theta_i(k) \end{aligned}$$

Στην εξίσωση (57) βλέπουμε ότι οι μετατοπίσεις όλων των αντικειμένων επηρεάζονται από την «από-περιστροφή». Ωστόσο, μόνο το αντικείμενο  $i$  υποβάλλεται σε καθαρή μετατόπιση  $[T_i^x(k), T_i^y(k)]$ . Τα άλλα αντικείμενα υποβάλλονται σε μετατόπιση μετά την περιστροφή με γωνία  $\theta_j(k) - \theta_i(k)$ , δηλαδή οι μετατοπίσεις τους συμβαίνουν στις περιστρεφόμενες συντεταγμένες  $(x', y')$  και η κίνησή τους είναι ακόμα περιστρεφόμενη-μετατοπιζόμενη. Συνεπώς, η μέθοδος της ενότητας 5.1 εφαρμόζεται μόνο σε καθαρή μετατόπιση του αντικειμένου  $i$  και θα δώσει εμφανείς κορυφές γύρω από το  $[T_i^x(k), T_i^y(k)]$ . Η αρχική μετατόπιση  $\bar{d}_i(k)$  μπορεί να υπολογιστεί εύκολα από το  $[T_i^x(k), T_i^y(k)]$ , δεδομένου ότι η  $\theta_i(k)$  είναι γνωστή. Οι περιστροφικές-μετατοπιζόμενες κινήσεις των άλλων αντικειμένων μπορεί να εισάγουν κάποιο θόρυβο σε αυτόν τον υπολογισμό της μετατόπισης, αλλά δεν αναμένεται να οδηγήσουν σε σημαντικές κορυφές, αφού δεν αντιστοιχούν σε καθαρές μετατοπίσεις.

Τελικά, για τη σωστή αντιστοιχία των υπολογισμένων μετατοπίσεων σε γωνίες περιστροφής  $\theta_i(k)$ , στρεβλώνουμε και παρεμβάλλουμε (με κυβική παρεμβολή) το πλαίσιο  $k$  με όλους τους πιθανούς συνδυασμούς των εξαγόμενων περιστρεφόμενων και μετατοπιζόμενων κινήσεων. Το συγκρίνουμε έπειτα με το



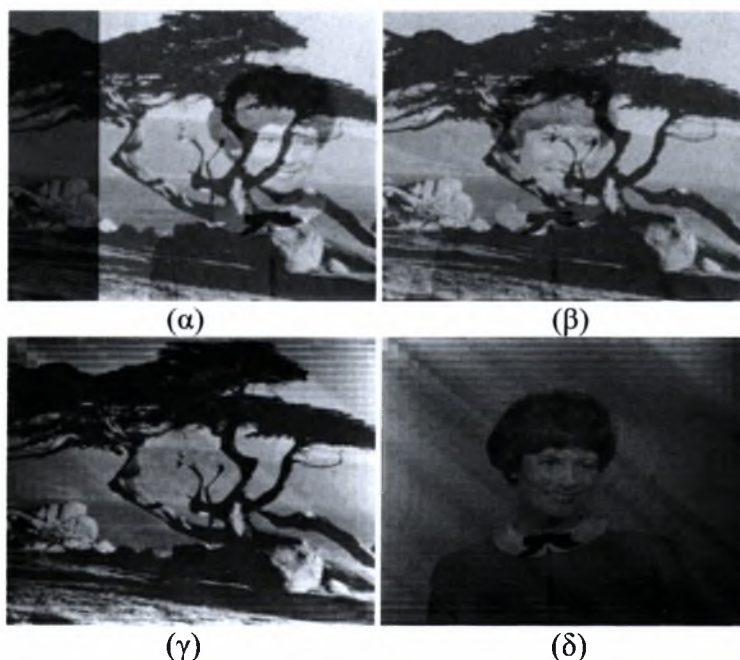
πλαίσιο 1 με τη διαδικασία που περιγράφηκε στην 5.3.2. Οι σωστά στρεβλωμένες περιοχές αντιστοιχούν στο αντικείμενο  $i$  και στις κινήσεις  $\theta_i(\mathbf{k})$  και  $\bar{d}_i(\mathbf{k})$ . Σημειώστε ότι, με αυτόν τον τρόπο, εξάγουμε συγχρόνως τα κινούμενα αντικείμενα, δηλαδή εκτελούμε κατάτμηση κίνησης στο χωρικό τομέα.

## 6. Πειράματα

Εκτελούμε πειράματα με συνθετικές και αληθινές ακολουθίες, για να αποδείξουμε τις ικανότητες εκτίμησης μετατόπισης και περιστροφής της μεθόδου μας. Επίσης δείχνουμε τη βασισμένη στο Fourier κατάτμηση κίνησης, και τα αποτελέσματα της κατάτμησης μετά την συγχώνευση της χωρικής και συχνοτικής πληροφορίας. Στις αληθινές ακολουθίες, η θεμελιωμένη αλήθεια για τις μετατοπίσεις και τις περιστροφές παράγεται μέσω χειροκίνητου ταιριάσματος των χαρακτηριστικών.

### 6.1 Συνθετική ακολουθία 1

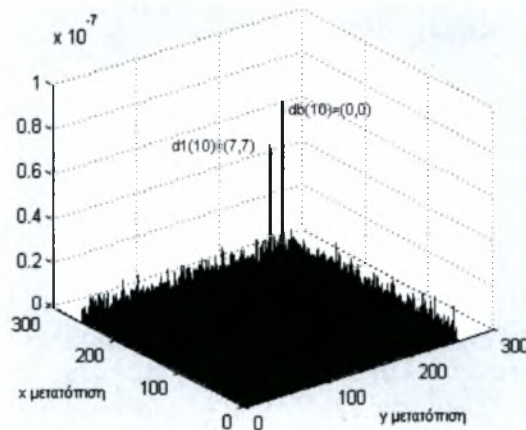
Αρχικά διεξάγουμε πειράματα με μια συνθετική ακολουθία, όπου ένα κορίτσι μετατοπίζεται με σταθερή ταχύτητα 7.2 προς τα δεξιά, σε 10 πλαίσια. Το 75<sup>ο</sup> εκατοστημόριο των μοναδικών τιμών του πίνακα συσχέτισης περιέχει δύο τιμές, 2.42 και 0.34 (από το φόντο και το κορίτσι), ενώ οι υπόλοιπες είναι κοντά στο μηδέν. Συνεπώς περιμένουμε να έχουμε ένα κινούμενο αντικείμενο. Η μέθοδος εκτίμησης κίνησης μετασχηματισμού Fourier της ενότητας 5.1 υπολογίζει ότι η μετατόπιση είναι οριζόντια, προς τα δεξιά, και ίση με 7.1, η οποία είναι πολύ κοντά στην αληθινή τιμή. Οι εικόνες 23(γ),(δ) δείχνουν την ελαχίστων τετραγώνων κατάτμηση στον χώρο Fourier, όπου το φόντο και το κινούμενο αντικείμενο διαχωρίστηκαν επιτυχώς. Το ανακτημένο φόντο και το αντικείμενο είναι σκουρότερα από τα αρχικά λόγω της συστηματοποίησης (regularization).



**Εικόνα 23.** Συνθετική ακολουθία: (α) Πλαίσιο 1. (β) Πλαίσιο 10. Λύσεις ελαχίστων τετραγώνων. (γ) Φόντο. (δ) Αντικείμενο.

## 6.2 Συνθετική ακολουθία 2

Σε αυτό το πείραμα δημιουργήσαμε μια συνθετική ακολουθία, αποτελούμενη από μια κατασκευασμένη εικόνα που αναπαριστά έναν καμεραμάν να μετατοπίζεται στις  $x$  και  $y$  κατευθύνσεις ενάντια σε ένα κατασκευασμένο φόντο (που αναπαριστά το δρόμο μίας πόλης) πάνω σε 10 πλαίσια. Δεδομένου ότι η μετατόπισή της είναι σταθερή, εφαρμόζουμε τη μέθοδο της ενότητας 5.2(1), και βρίσκουμε δύο μοναδικές τιμές που είναι μεγαλύτερες τις μονάδας, ενώ όλες οι υπόλοιπες είναι κοντά στο μηδέν. Έτσι αναμένουμε να εξάγουμε 1 αντικείμενο. Αυτό επαληθεύεται στη διαδικασία υπολογισμού της κίνησης, όπου υπάρχει μία κορυφή που αντιστοιχεί σε μη μηδενική μετατόπιση. Η μετατόπιση μέσα στο πλαίσιο είναι (7,7) Η μέθοδος μετασχηματισμού Fourier υπολογίζει την κίνησή του (7.1,7.1) αντίστοιχα. Έπειτα εφαρμόζουμε την ελαχίστων τετραγώνων κατάτμηση αντικειμένου. Η Εικόνα 25(γ) δείχνει ότι το φόντο έχει επιτυχώς διαχωριστεί από το αντικείμενο, το οποίο έχει επίσης αποσπαστεί επιτυχώς Εικόνα 25(δ). Σημειώστε ότι όλες οι λεπτομέρειες στη σύσταση του αντικειμένου διατηρούνται, και το σχήμα του εξάγεται, χωρίς την επίδραση του ορίου. Στην Εικόνα 24 φαίνεται η τρισδιάστατη επιφάνεια  $\varphi_{10}(\vec{r})$  από την συγκεκριμένη ακολουθία, όπου ο καμεραμάν μετατοπίζεται από το πλαίσιο 1 στο πλαίσιο 10. Παρόλο που υπάρχει το φόντο, το οποίο δημιουργεί μια κυρίαρχη κορυφή ( $\vec{d}_b = (0,0)$ ) η κορυφή που αντιστοιχεί στο αντικείμενο που κινείται μπορεί να εξαχθεί καθαρά. Είναι ίση με  $\vec{d}_1(10) = (63,63)$ , αντιστοιχώντας σε οριζόντια μέσα στο κάθε πλαίσιο μετατόπιση  $d_1=7$  και σε κάθετη μετατόπιση μέσα σε κάθε πλαίσιο ίση με  $d_2=7$ . Πρέπει να σημειωθεί ότι το κέντρο του συστήματος συντεταγμένων θεωρείται ότι είναι το κέντρο της επιφάνειας  $\varphi_k(\vec{r})$ , και οι μετατοπίσεις υπολογίζονται σύμφωνα με αυτό. Η φωτεινότητα των ανακτημένων αντικειμένων είναι ελαφρώς σκοτεινότερη από τα αρχικά αντικείμενα, λόγω της διαδικασίας συστηματοποίησης (regularization process, ενότητα 5.2.1 ).



**Εικόνα 24.** Επιφάνεια  $\varphi_{10}(\vec{r})$ , πλαίσια 1 και 10. Δύο κορυφές είναι κυρίαρχες, μία για το σταθερό φόντο (η κορυφή  $db$ ) και η άλλη ( $d1$ ) για τον καμεραμάν.



(α)



(β)



(γ)



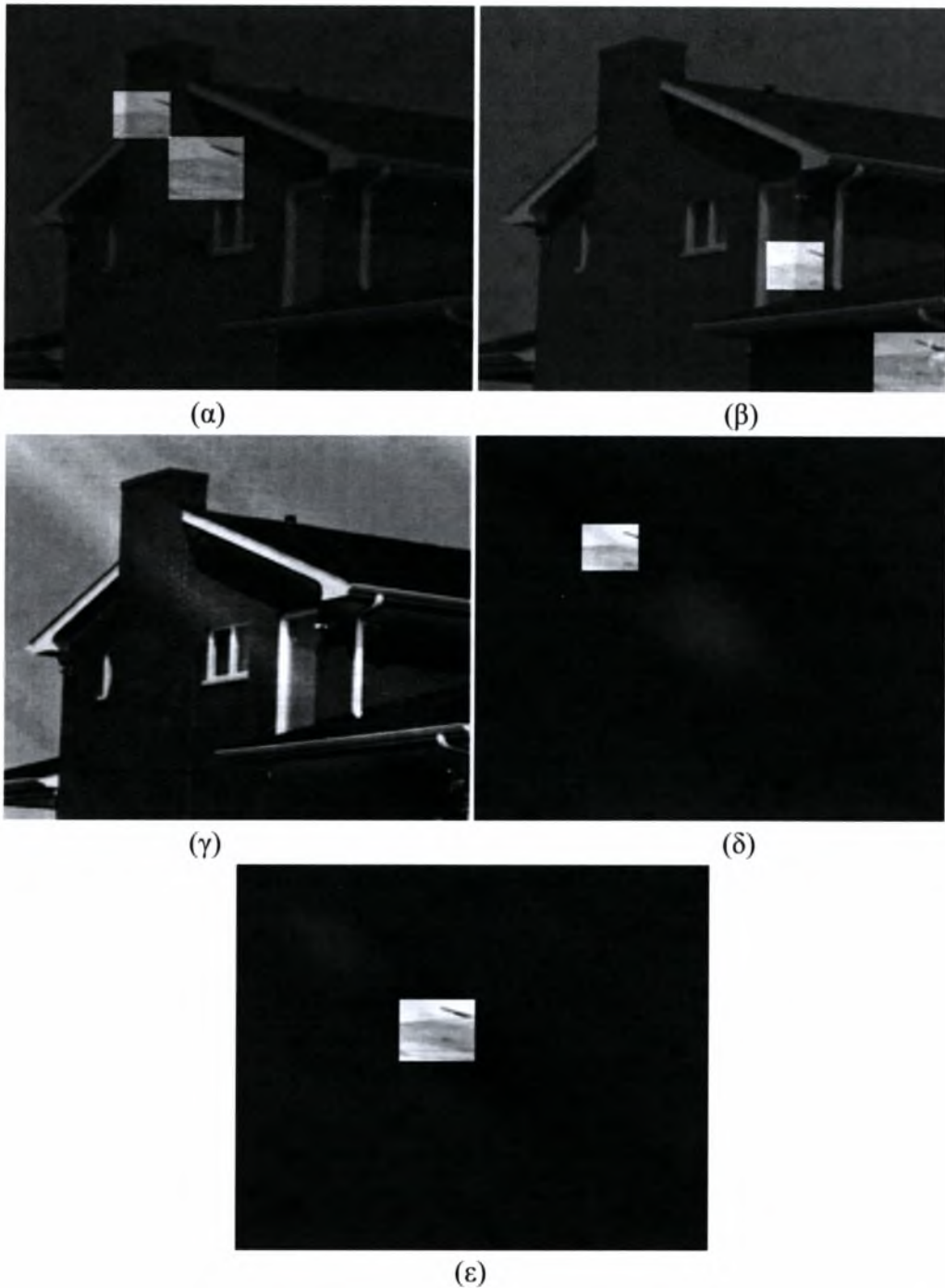
(δ)

**Εικόνα 25.** Συνθετική ακολουθία: (α) Πλαίσιο 1. (β) Πλαίσιο 10. Λύσεις ελαχίστων τετραγώνων: (γ) Φόντο. (δ) Αντικείμενο

### 6.3 Συνθετική ακολουθία 3

Σε αυτό το πείραμα χρησιμοποιούνται δύο κατασκευασμένα τετράγωνα (μιας εικόνας που απεικονίζει ένα αεροπλάνο) που μετατοπίζονται στις  $x$  και  $y$  κατευθύνσεις ενάντια σε ένα επίσης κατασκευασμένο ακίνητο φόντο (που αναπαριστά ένα σπίτι) πάνω σε 15 πλαίσια. Κατά τη διαδικασία εκτίμησης της κίνησης υπάρχουν δύο κορυφές που αντιστοιχούν σε μη μηδενικές μετατοπίσεις, οπότε συμπεραίνουμε ότι θα εξάγουμε δύο αντικείμενα. Η μετατόπιση ανάμεσα στα πλαίσια για το μεγάλο τετράγωνο είναι (9,9) και για το μικρό (7,7). Η μέθοδος μετασχηματισμού Fourier υπολογίζει τις κινήσεις τους (9.1,9.1) και (6.7,6.7) αντίστοιχα. Μετά εφαρμόζουμε την ελαχίστων τετραγώνων κατάτμηση αντικειμένου. Στην Εικόνα 26(γ) φαίνεται η επιτυχής κατάτμηση του φόντο από τα αντικείμενα, και στις εικόνες 26(δ),(ε) φαίνονται τα ανακτημένα αντικείμενα. Το σχήμα και η «υφή» των αντικειμένων σώζονται και ανακτώνται ακριβώς, αποδεικνύοντας ότι η ελαχίστων τετραγώνων, βασισμένη στο Fourier κατάτμηση μπορεί να προκαλέσει την ανάκτηση των αντικειμένων άσχετα από το σχήμα τους, το μέγεθός τους και των χαρακτηριστικών ιδιοτήτων τους. Τα αποτελέσματα ελαχίστων τετραγώνων συσχετίζονται με το πρώτο πλαίσιο (Εικόνα 26(α)), και οι τιμές συσχέτισης συγκρίνονται με ένα κατώτατο όριο 0.842 (ενότητα 4.1), προκειμένου να εντοπιστούν τα κινούμενα αντικείμενα.





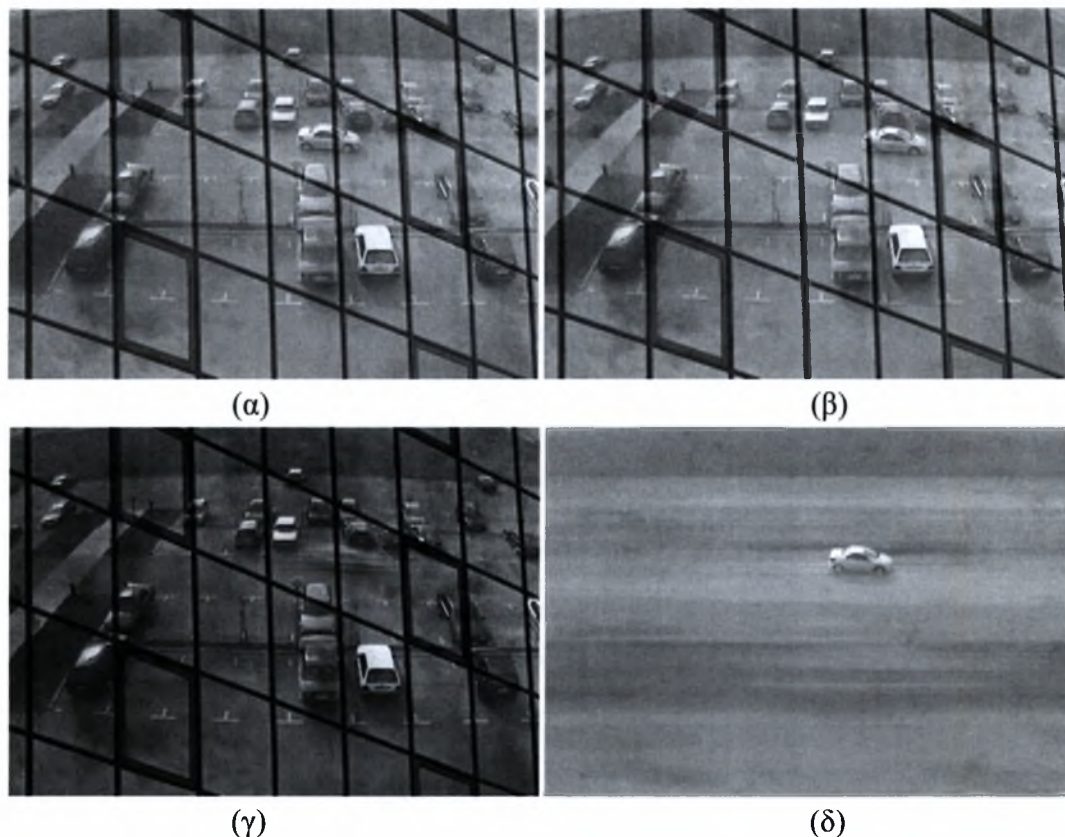
**Εικόνα 26.** Συνθετική ακολουθία: (α) Πλαίσιο 1. (β) Πλαίσιο 15. Λύσεις ελαχίστων τετραγώνων. (γ) Φόντο. (δ) Αντικείμενο 1. (ε) Αντικείμενο 2.

#### 6.4 Χώρος στάθμευσης αυτοκινήτων

Μια αληθινή ακολουθία από ένα χώρο στάθμευσης αυτοκινήτων, με 50 πλαίσια στα οποία ένα αυτοκίνητο μετατοπίζεται δεξιά, εξετάζεται (εικόνες 27(α),(β)). Η ανάλυση SVD του πίνακα συσχέτισης του μετασχηματισμού Fourier δείχνει ότι υπάρχει μόνο ένα μετατοπιζόμενο αντικείμενο, γιατί οι δύο πρώτες μοναδικές τιμές του είναι ίσες με 6.8 και 7.22, ενώ οι υπόλοιπες είναι κοντά στο 0. Η



μέθοδος μετασχηματισμού Fourier οδηγεί σε έναν υπολογισμό της μετατόπισης ανάμεσα στα πλαίσια ίσο με 1.9 εικονοκύτταρα (pixels), ο οποίος είναι πολύ κοντά στην πραγματική μετατόπιση που είναι ίση με 1.95 εικονοκύτταρα. Η ελαχίστων τετραγώνων κατάτμηση στις εικόνες 27(γ),(δ) είναι αρκετά ακριβής. Το σχήμα του αυτοκινήτου και οι λεπτομέρειές του όπως τα λάστιχά του, τα παράθυρα και το παρμπρίζ του έχουν εξαχθεί. Στην Εικόνα 27(γ) το εξαγόμενο φόντο φαίνεται θολό σε μερικές περιοχές όπου βρισκόταν το κινούμενο αυτοκίνητο, επειδή είχε φραχτεί (occluded) από αυτό. Στην πραγματικότητα, το αυτοκίνητο έχει πολύ μικρή μετατόπιση ανάμεσα στα πλαίσια, έτσι το φόντο φαίνεται θολό σε εκείνη την περιοχή όπου βρισκόταν το αυτοκίνητο στην ελαχίστων τετραγώνων λύση.

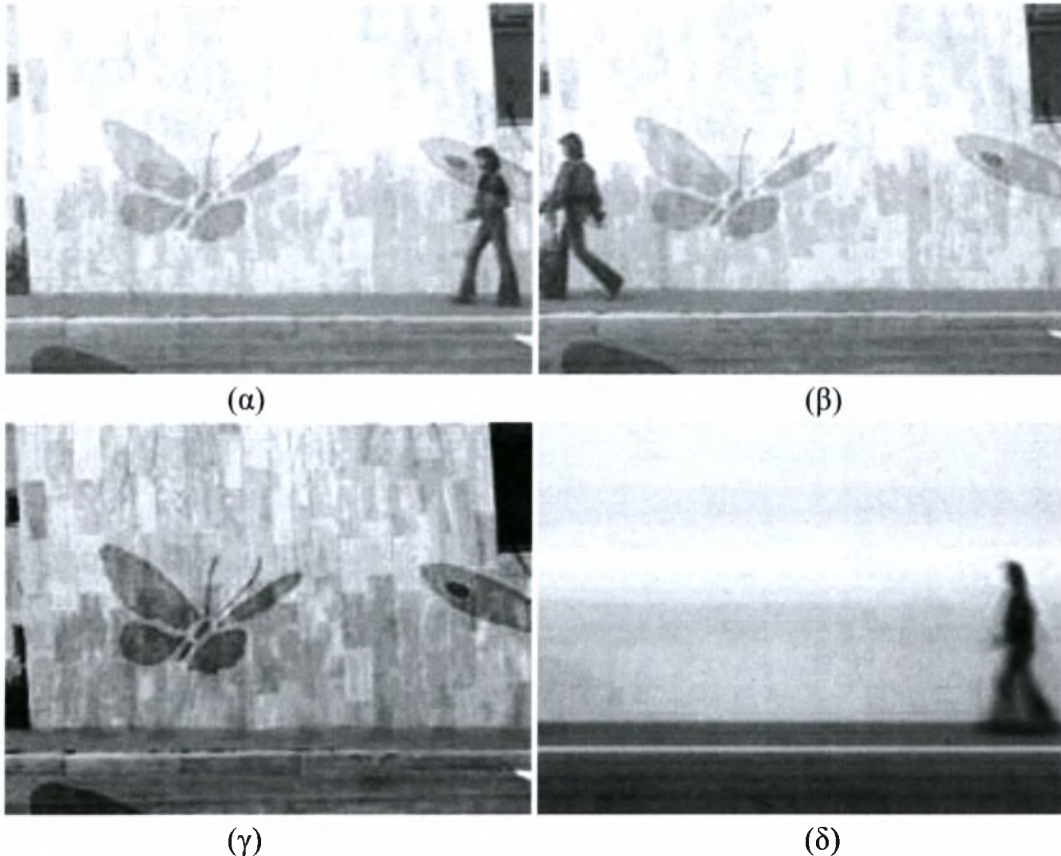


**Εικόνα 27.** Ακολουθία χώρου στάθμευσης αυτοκινήτων. (α) Πλαίσιο 30. (β) Πλαίσιο 80. Λύσεις ελαχίστων τετραγώνων: (γ) Φόντο. (δ) Αυτοκίνητο.

## 6.5 Περίπατος

Σε αυτό το πείραμα εξετάζουμε μια αληθινή ακολουθία, όπου ένα κορίτσι κάνει τον περίπατό του έχοντας ως σταθερό φόντο τον τοίχο ενός κτιρίου πάνω στον οποίο φαίνονται οι αντανάκλασεις από πεταλούδες. Το κορίτσι ξεκινάει τον περίπατό του στο πλαίσιο 1 (Εικόνα 28(α)) και σταματάει στο πλαίσιο 80 (Εικόνα 28(β)). Η μετατόπιση του κοριτσιού υπολογίζεται ανάμεσα σε διαδοχικά πλαίσια και συγκρίνεται με την πραγματική τιμή, η οποία επιτυγχάνεται με χειρωνακτική παρακολούθηση των σημείων στα οποία κινείται το αντικείμενο. Η μετατόπιση υπολογίζεται ότι είναι οριζόντια και ίση με (0,1.8) που είναι κοντά στην πραγματική

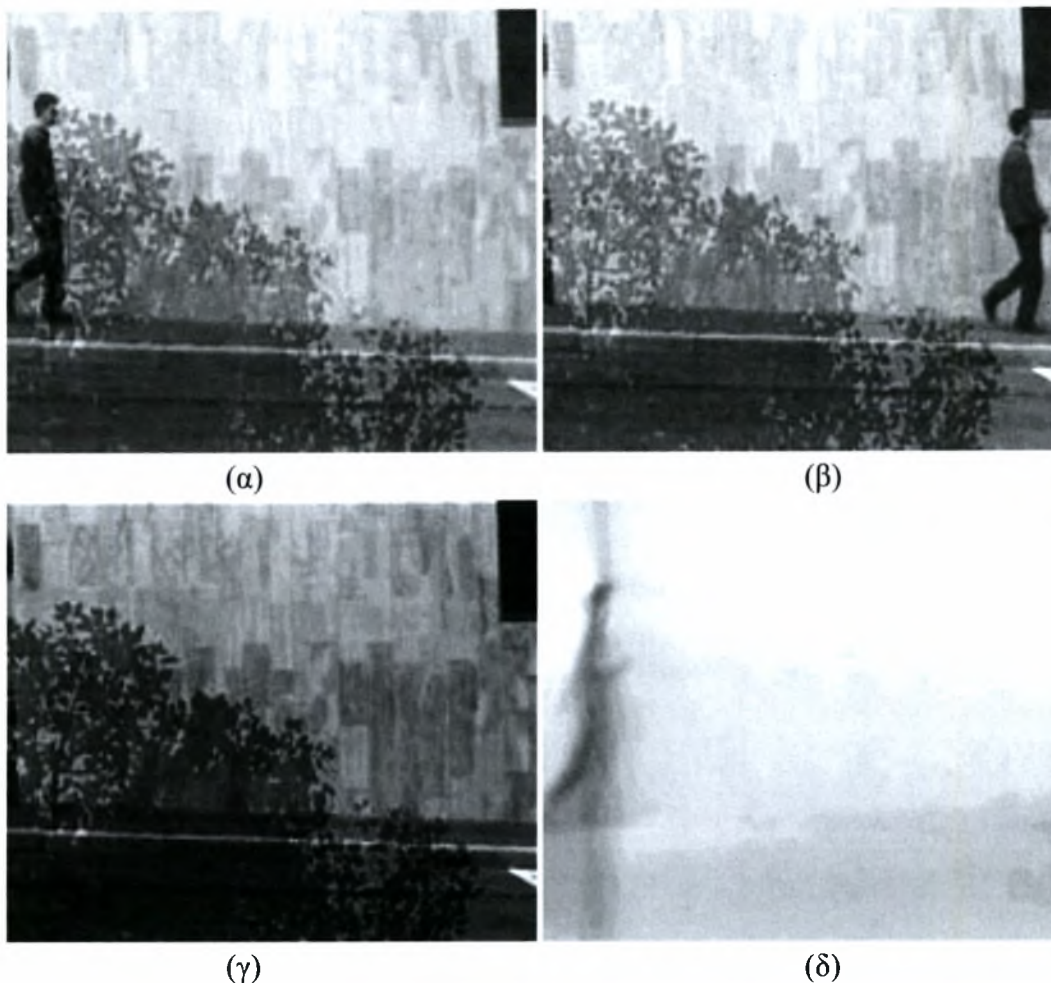
τιμή (0,1.875). η ελαχίστων τετραγώνων κατάτμηση είναι αρκετά ακριβής. Το φόντο έχει εξαχθεί επιτυχώς από το αντικείμενο που κινείται, όπως φαίνεται στην Εικόνα 28(γ). Το σχήμα του κοριτσιού διακρίνεται πολύ καλά, αλλά λεπτομέρειες που αφορούν στην υφή του στη σύσταση του δεν είναι εμφανείς. Αυτό συμβαίνει γιατί το κορίτσι δε διατηρούσε το σχήμα του καθ' όλη τη διάρκεια της ακολουθίας, αλλά κουνούσε τα χέρια και τα πόδια της συνεχώς για να περπατήσει.



**Εικόνα 28.** Περίπατος. (α) Πλαίσιο 1. (β) Πλαίσιο 80. (γ) Ανακτημένο φόντο με τις διαφανείς πεταλούδες. (δ) Κορίτσι.

## 6.6 Περίπατος σε ελαφρώς κεκλιμένο επίπεδο

Σε αυτήν την ακολουθία ένας άντρας περπατάει σε έναν κεκλιμένο δρόμο, στον οποίο πέφτει ο ίσκιος από ένα δέντρο. Η μετατόπιση υπολογίζεται ότι είναι 2.1 στην οριζόντια κατεύθυνση και 0.1 στην κάθετη κατεύθυνση. Μετά την κατάτμηση των πλαισίων της ακολουθίας το φόντο έχει εξαχθεί επιτυχώς (Εικόνα 29(γ)). Το σχήμα του άντρα έχει εξαχθεί αρκετά καλά, αλλά είναι θολό και δε διακρίνονται λεπτομέρειές του γιατί ο άντρας κουνάει καθ' όλη τη διάρκεια του περιπάτου του τα χέρια και τα πόδια του, δεν έχει δηλαδή ένα σταθερό σχήμα. Επίσης λόγω της συστηματοποίησης τα αποτελέσματα από την ελαχίστων τετραγώνων κατάτμηση είναι σκουρότερα από τα αρχικά.

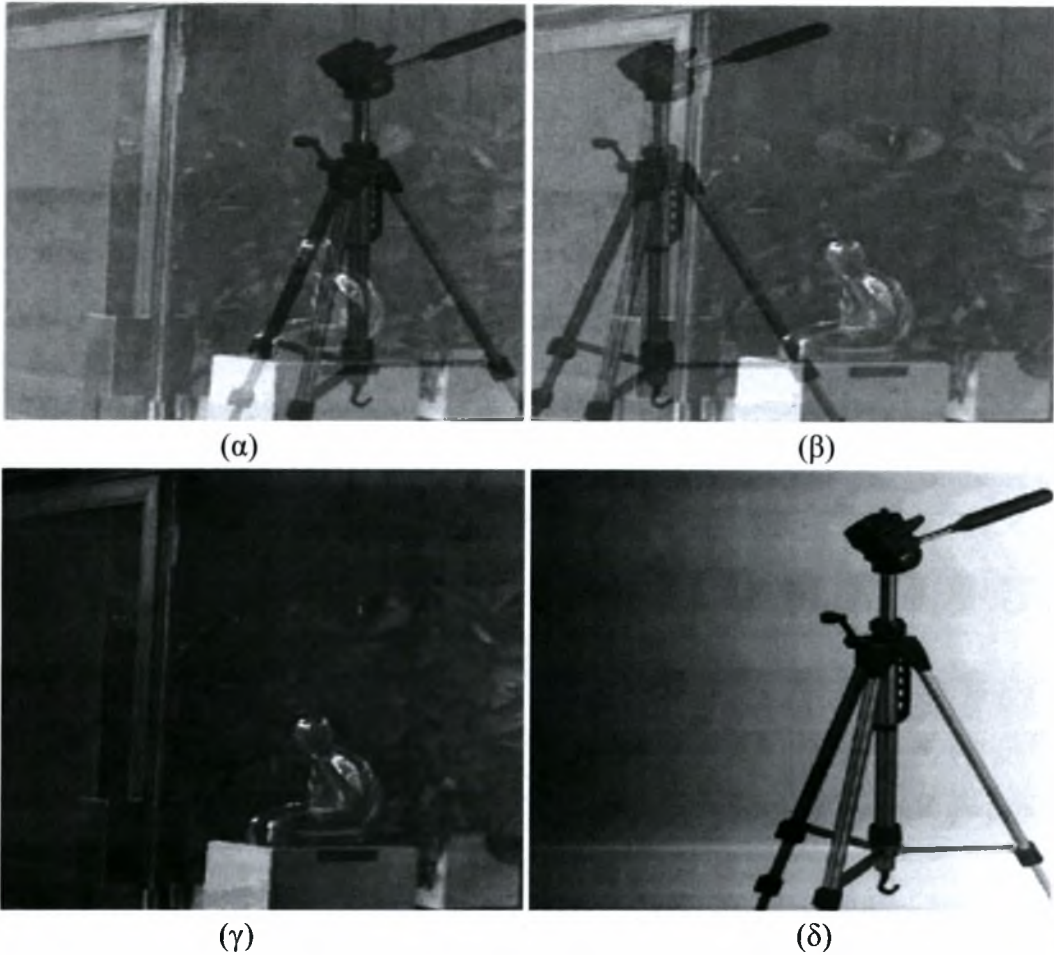


**Εικόνα 29.** Περίπατος σε ελαφρώς κεκλιμένο επίπεδο. (α) Πλαίσιο 1. (β) Πλαίσιο 72. (γ) Ανακτημένο φόντο. (δ) Άντρας.

## 6.7 Είσοδος κτιρίου

Σε αυτό το πείραμα, μια κάμερα τοποθετήθηκε μπροστά από μια περιστρεφόμενη πόρτα ενός κτιρίου. Το πλάνο της κάμερας περιλαμβάνει το σκηνικό που αντανακλάται από την περιστρεφόμενη πόρτα (τη γενική εικόνα πίσω από την κάμερα, συμπεριλαμβανομένης και της ίδιας της κάμερας) και το σκηνικό που γίνεται αντιληπτό δια μέσου της περιστρεφόμενης πόρτας (το χολ εισόδου του κτιρίου). Το σκηνικό που παρατηρούμε μέσα στο κτίριο (ένα άγαλμα και μερικά φυτά) είναι στάσιμο, ενώ το αντανακλώμενο εξωτερικό σκηνικό, δηλαδή η αντανάκλαση της κάμερας πάνω στην πόρτα, λόγω της μεταβολής του φωτισμού και λόγω του ότι η πόρτα ανοιγοκλείνει δεν είναι στάσιμο. Στις εικόνες 30(α)-(β) φαίνονται αντίστοιχα το πρώτο και το τελευταίο πλαίσιο της ακολουθίας. Εφαρμόζοντας την ελαχίστων τετραγώνων κατάτμηση αντικειμένου παίρνουμε το επιτυχώς εξαγόμενο φόντο και το αντικείμενο, όπως φαίνονται στις εικόνες 30(γ)-(δ).





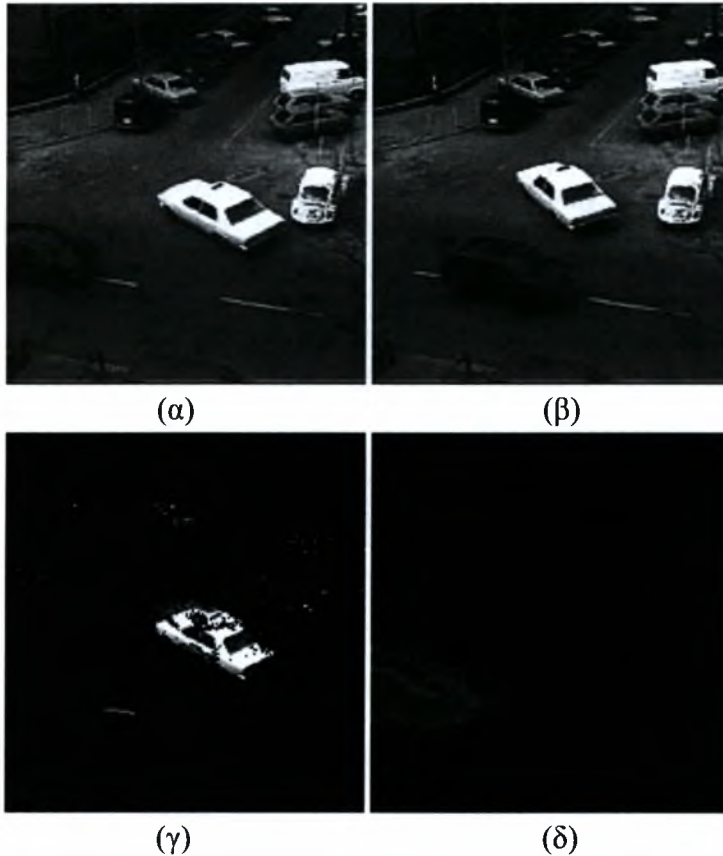
**Εικόνα 30.** Είσοδος κτιρίου. (α)-(β) Το πρώτο και το τελευταίο πλαίσιο της ακολουθίας. (γ) Το ανακτημένο φόντο μετά την εξαγωγή του τριπόδου. (δ) Το ανακτημένο τρίποδο.

## 6.8 Ταξί

Εξετάζουμε μια ακολουθία, όπου ένα ταξί υποβάλλεται σε μια περιστροφή και ένα αυτοκίνητο μετατοπίζεται σε είκοσι πλαίσια. Η λογαριθμική-πολική μέθοδος Fourier της ενότητας 5.4 δίνει την γωνία περιστροφής ίση με  $9.8^\circ$  μεταξύ των πλαισίων 20 και 1 για το ταξί, που είναι κοντά στην με το χέρι παραγμένη γωνία των  $9.5^\circ$ . Τότε από- περιστρέφουμε το πλαίσιο 20 (μετά την εξαγωγή του φόντο), και υπολογίζουμε τη μετατόπιση του ταξί να είναι ίση με (2,3). Εφαρμόζοντας τη μέθοδο εκτίμησης κίνησης Fourier της ενότητας 5.1, βγάζουμε το εκτόπισμα του αυτοκινήτου στα δεξιά (από το πλαίσιο 1 στο 20) να είναι ίσο με (53,12), και το αληθινό είναι (54,10). Συνεπώς, βλέπουμε ότι η προτεινόμενη μέθοδος πράγματι οδηγεί σε ακριβείς υπολογισμούς μετατόπισης και περιστροφής, για μια αληθινή ακολουθία, που μπορεί να προκύψει σε πολλές πρακτικές εφαρμογές.

Δεδομένου ότι υπάρχει περιστρεφόμενη κίνηση στην ακολουθία, χρησιμοποιούμε μόνο χωρικά δεδομένα για την εξαγωγή του αντικειμένου. Στρεβλώνουμε το πλαίσιο 20 με κάθε μία από τις κινήσεις που είναι υπό εξέταση, και μετά συγκρίνουμε το στρεβλωμένο πλαίσιο με το πρώτο, όπως περιγράφηκε στην

ενότητα 5.3.2. αυτό δίνει τα τελικά αποτελέσματα κατάτμησης των εικόνων 31(γ)-(δ). μερικό από το φόντο γύρω από τα αυτοκίνητα έχει ανακτηθεί επίσης, γιατί το χρώμα από αυτά τα αυτοκίνητα είναι αρκετά παρόμοιο με αυτό του δρόμου, κάνοντας δύσκολη τη διάκριση του ενός από το άλλο. Παρόλα αυτά, οι περιστροφές και οι μετατοπίσεις του αντικειμένου έχουν υπολογιστεί σωστά, και κατά συνέπεια, τα αντικείμενα έχουν εντοπιστεί σωστά.



**Εικόνα 31.** Ακολουθία ταξί. (α) Πλαίσιο 1. (β) Πλαίσιο 20. (γ) Κατατετμημένο ταξί. (δ) Κατατετμημένο αυτοκίνητο στα δεξιά. Μόνο χωρικά δεδομένα χρησιμοποιήθηκαν για την κατάτμηση γιατί η ακολουθία περιέχει και μετατόπιση και περιστροφή.



## 7. Συμπεράσματα

Μια νέα υβριδική μέθοδος για την ανάλυση κίνησης παρουσιάστηκε. Ο υπολογισμός της κίνησης επιτυγχάνεται στο πεδίο της συχνότητας, και η κατάτμηση της ακολουθίας βασίζεται και σε χωρικά δεδομένα και σε δεδομένα στο χώρο της συχνότητας. Η προτεινόμενη μέθοδος αποφεύγει τα προβλήματα των χωρικών μεθόδων, όπως η ευαισθησία σε γενικές αλλαγές φωτισμού, ανακρίβειες στα όρια του αντικειμένου και ασυνέχειες στην κίνηση. Η χρήση του μετασχηματισμού Fourier μπορεί να εφαρμοστεί μέσω του γρήγορου μετασχηματισμού Fourier (Fast Fourier Transform), έτσι ώστε να είναι υπολογιστικά αποδοτικός. Για την περίπτωση των αμιγώς μετατοπιζόμενων κινήσεων, μια νέα διατύπωση παρουσιάζεται για την εξαγωγή των αντικειμένων, μέσω της λύσης ελαχίστων τετραγώνων ενός γραμμικού συστήματος στον τομές Fourier. Ένα σημαντικό πλεονέκτημα αυτής της προσέγγισης είναι ότι τα αποτελέσματα είναι ανεξάρτητα από τα σχήματα, τις ασυνέχειες στην κίνηση των αντικειμένων, και είναι ανθεκτική σε τοπικές ανακρίβειες, όπως το κλείσιμο ενός αντικειμένου από ένα άλλο (occlusion), λόγω του πλεονασμού στο γραμμικό σύστημα. Λόγω του λάθους διαμόρφωσης και της διαδικασίας συστηματοποίησης που χρησιμοποιείται, τα αποτελέσματα της ελαχίστων τετραγώνων κατάτμησης περιέχουν κάποιο θόρυβο, για αυτό βελτιώνονται περαιτέρω χρησιμοποιώντας χωρική πληροφορία κατά τρόπο μη ειδικό. Η μέθοδος μετασχηματισμού Fourier εκτίμησης κίνησης έχει γενικευτεί σε περισσότερο περίπλοκες κινήσεις, που περιλαμβάνουν συνδυασμούς από περιστροφές και μετατοπίσεις. Τα πειράματα με συνθετικές και αληθινές ακολουθίες δείχνουν ότι η προτεινόμενη προσέγγιση μπορεί πράγματι να υπολογίσει και τις μετατοπίσεις και τις περιστροφές αξιόπιστα. Σε αυτή την περίπτωση αμιγώς χωρικές μέθοδοι χρησιμοποιούνται για την εξαγωγή του αντικειμένου, οι οποίες οδηγούν σε ακριβή εξαγωγή του αντικειμένου, όπως φαίνεται στα αποτελέσματα των πειραμάτων. Μελλοντικές κατευθύνσεις της έρευνας περιλαμβάνουν την κοινή χρήση χωρικών και συχνοτικών δεδομένων για την ανάλυση ακόμη περισσότερο περίπλοκων κινήσεων, που περιλαμβάνουν τυχαίες παραλλαγές, καθώς επίσης και την εξέταση και την ανάλυση της κίνησης εύκαμπτων αντικειμένων.

## ΠΑΡΑΡΤΗΜΑ

### ΠΑΡΑΓΩΓΟΣ ΤΗΣ ΣΥΝΑΡΤΗΣΗΣ (derivation) $Q(\rho, \theta)$

Για την περίπτωση των πολλαπλών περιστροφών και μετατοπίσεων, δε μπορούμε να εξάγουμε αμέσως τις γωνίες περιστροφής από τη φάση του μετασχηματισμού Fourier σε πολικές συντεταγμένες (όπως στην ενότητα 5.4.1). Αυτό γίνεται επειδή έχουμε ένα άθροισμα των μετασχηματισμών Fourier του κινούμενου αντικειμένου, έτσι η μετατόπιση που προκάλεσε αλλαγές στη φάση  $e^{-j(\omega_x d_i^x(k) + \omega_y d_i^y(k))}$  δεν εξαφανίζεται όταν παίρνουμε το μέγεθος του μετασχηματισμού Fourier του πλαισίου. Ο μετασχηματισμός Fourier του πλαισίου  $k$ , μετά την υποβολή κάθε αντικειμένου  $i$  σε περιστροφή μια  $\theta_i(k)$  και μια μετατόπιση  $\bar{d}_i(k)$ , δίνεται από:

$$\begin{aligned} A_{RT}(\omega_x, \omega_y, k) &= \sum_{i=1}^M S_i(\omega_x \cos(\theta_i(k)) + \omega_y \sin(\theta_i(k)), -\omega_x \sin(\theta_i(k)) \\ &\quad + \omega_y \cos(\theta_i(k))) e^{-j(\omega_x d_i^x(k) + \omega_y d_i^y(k))} \end{aligned} \quad (58)$$

Οι εκθετικοί όροι χρειάζεται να εκφραστούν συναρτήσει του  $\psi - \theta_i(k)$ , έτσι υποθέτουμε ότι:

$$\begin{aligned} \omega_x d_i^x(k) + \omega_y d_i^y(k) &= \rho [\cos \psi d_i^x(k) + \sin \psi d_i^y(k)] \\ &= \alpha \cos(\psi - \theta_i(k)) + \beta \sin(\psi - \theta_i(k)) \end{aligned} \quad (59)$$

Όπου  $\omega_x = \rho \cos \psi$  και  $\omega_y = \rho \sin \psi$ . Έπειτα έχουμε:

$$\begin{aligned} \alpha \cos(\psi - \theta_i(k)) + \beta \sin(\psi - \theta_i(k)) &= \alpha \cos \psi \cos \theta_i(k) + \alpha \sin \psi \sin \theta_i(k) + \beta \sin \psi \cos \theta_i(k) \\ &\quad - \beta \sin \theta_i(k) \cos \psi \\ &= [\alpha \cos \theta_i(k) - \beta \sin \theta_i(k)] \cos \psi \\ &\quad + [\alpha \sin \theta_i(k) + \beta \sin \psi \cos \theta_i(k)] \sin \psi \end{aligned} \quad (60)$$

Από τις (59),(60) έχουμε  $\rho d_i^x(k) = \alpha \cos \theta_i(k) - \beta \sin \theta_i(k)$ ,  $\rho d_i^y(k) = \alpha \sin \theta_i(k) + \beta \cos \theta_i(k)$  από τις οποίες πετυχαίνουμε:

$$\begin{aligned} \alpha &= \rho [d_i^x(k) \cos \theta_i(k) + d_i^y(k) \sin \theta_i(k)] \\ \beta &= \rho [d_i^y(k) \cos \theta_i(k) - d_i^x(k) \sin \theta_i(k)] \end{aligned} \quad (61)$$

## Βιβλιογραφία

1. B. Horn and B. Schunck, “*Determining optical flow*,” *Artificial Intelligence*, vol. 17, pp. 185–203, 1981.
2. J. L. Barron, D. J. Fleet, S. S. Beauchemin, and T. A. Burkitt, “*Performance of optical flow techniques*”, *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, June 1992, pp. 236–242.
3. B. D. Lucas and T. Kanade, “*An iterative image registration technique with an application to stereo vision*”, *Proceedings of Imaging understanding workshop*, 1981, pp. 121-130.
4. W. Chen, G. B. Giannakis, and N. Nandhakumar, “*A harmonic retrieval framework for discontinuous motion estimation*”, *IEEE Transactions on Image Processing*, vol. 7, no. 9, pp. 1242–1257, Sept 1998.
5. E. E. Osborne, “*On least squares solutions of linear equations*” *Journal of the ACM*, vol. 8, pp. 628–636, 1961.
6. C. Stiller, J. Conrad and R. Bosch, “*Estimating motion in image sequences: A tutorial on modeling and computation of 2D motion*”, *IEEE Signal Processing Magazine*, 1999, pp 70-91.
7. Robert J. Snowden and Frans A. J. Verstraten, “*Motion Transparency: making models of motion perception transparent*”, *Trends in Cognitive Sciences* October 1999, pp 369-377.
8. M. Shizawa and K. Mase, “*Multiple optical flow-fundamental constraint equations and a unified computational theory for detecting motion transparency and motion boundaries*”, *Systems and computers in Japan*, 1994, pp 52-74.
9. E. P. Simoncelli, “*Distributed representation and analysis of visual motion*”, *Massachusetts Institute of Technology, Cambridge, MA*, 1993.
10. G. Farneback, “*Spatial Domain Methods for Orientation and Velocity Estimation*”, Lic. Thesis LiU-TekLic -1999:13, Dept. EE, Linköping University, SE-581 83 Linköping, Sweden, March 1999. Thesis No. 755, ISBN 91-7219-441-3.
11. M. Irani, B. Rousso and S. Peleg, “*Computing occluding and transparent motions*”, *International Journal of Computer Vision*, February 1994, pp 5-16.
12. E. H. Adelson and P. Anandan, “*Ordinal characteristics of transparency*”, *AAAI-90 Workshop on Qualitative Vision*, Boston, MA, July 1990, pp. 77—81.

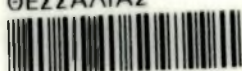
13. T Darrell and E P Simoncelli, “*Separation of Transparent Motion into Layers using Velocity-Tuned Mechanisms*”, Third European Conf on Computer Vision, Stockholm, 1994.
14. B. Sarel, M. Irani, “*Separating transparent layers through layer information exchange*”, In: ECCV, 2004, vol. 4, pp. 328-341.
15. A. Briassouli, N. Ahuja, “*Integration of Frequency and Space for Multiple Motion Estimation and Shape-Independent Object Segmentation*”, Circuits and Systems for Video Technology, IEEE Transactions on Volume 18, Issue 5, May 2008, pp 657 – 669.
16. R. J. Snowden, S. Treue, R. E. Erickson and R. A. Andersen, “*The response of area MT and VI neurons to transparent motion*”, Journal of Neuroscience , 1991, pp 2768-2785.

### Ιστοχώροι

17. <http://www.questia.com/read/99402314?title=Advances%20in%20Computer%20Vision%20-%20Vol.%201>
18. <http://decsai.ugr.es/mia/complementario/tl/book3chaps.html>
19. [http://www.ece.cmu.edu/~ee899/project/deepak\\_mid.htm](http://www.ece.cmu.edu/~ee899/project/deepak_mid.htm)
20. [http://en.wikipedia.org/wiki/Optical\\_flow](http://en.wikipedia.org/wiki/Optical_flow)
21. [http://en.wikipedia.org/wiki/Phase\\_correlation](http://en.wikipedia.org/wiki/Phase_correlation)



ΠΑΝΕΠΙΣΤΗΜΙΟ  
ΘΕΣΣΑΛΙΑΣ



004000091627

