

## Κεφάλαιο 5

# ΤΕΧΝΙΚΕΣ ΕΚΤΙΜΗΣΗΣ ΔΙΔΙΑΣΤΑΤΗΣ ΚΙΝΗΣΗΣ

Η εκτίμηση της κίνησης στο επίπεδο της εικόνας (διδιάστατη κίνηση) ή η εκτίμηση κίνησης αντικειμένου (τρισδιάστατη κίνηση), είναι ένα από τα θεμελιώδη προβλήματα στην ψηφιακή επεξεργασία βίντεο και έχει αποτελέσει το αντικείμενο σημαντικής ερευνητικής προσπάθειας [Hua81, Hua83, Agg88, Sin91, Fle92, Sez93].

### 5.1 Το Πρόβλημα της Εκτίμησης Κίνησης

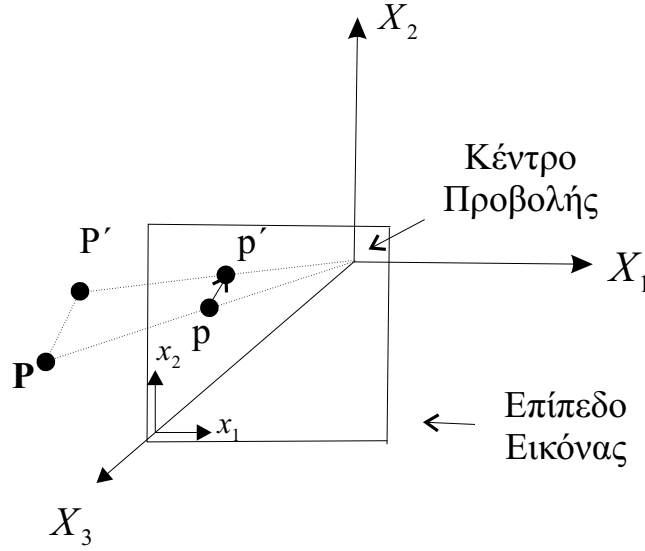
Αυτό το υποκεφάλαιο παρέχει: α) μια γενική εισαγωγή στο πρόβλημα της εκτίμησης της διδιάστατης κίνησης και β) συγκεκριμένους αλγόριθμους εκτίμησης κίνησης βασισμένους στην εξίσωση οπτικής ροής. Άλλες ποικίλες μη παραμετρικές προσεγγίσεις για την εκτίμηση της 2-D κίνησης θα καλυφθούν παρακάτω σ' αυτό το κεφάλαιο. Στην παράγραφο 5.1.1, θα δοθεί έμφαση στη διάκριση μεταξύ της διδιάστατης και της φαινομενικής κίνησης (οπτική ροή ή αντιστοίχιση). Στην παράγραφο 5.1.2, το πρόβλημα της εκτίμησης της διδιάστατης κίνησης διατυπώνεται ως ένα 'κακώς διατυπωμένο' πρόβλημα. Μια σύντομη προεπισκόπηση των μοντέλων του πεδίου κίνησης και των κατηγοριών μεθόδων εκτίμησης κίνησης παρουσιάζονται στην παράγραφο 5.1.3. Στην παράγραφο 5.2, συζητιούνται μέθοδοι εκτίμησης της οπτικής ροής βασισμένες στην εξίσωση οπτικής ροής (γνωστές και σαν διαφορικές μέθοδοι). Η εκτίμηση διδιάστατης κίνησης είναι συχνά το πρώτο βήμα προς την ανάλυση της τρισδιάστατης κίνησης. Τέλος αποτελεί μια αναπόσπαστη συνιστώσα του φιλτραρίσματος με αντιστάθμιση κίνησης και της συμπίεσης.

#### 5.1.1 Σχέση διδιάστατης κίνησης και φαινόμενης κίνησης

Επειδή οι εικόνες που μεταβάλλονται με το χρόνο είναι διδιάστατες προβολές τρισδιάστατων σκηνών, όπως περιγράφηκε σε προηγούμενα κεφάλαια, η διδιάστατη κίνηση αναφέρεται στην προβολή της τρισδιάστατης κίνησης στο επίπεδο της εικόνας, η οποία δημιουργεί ένα διδιάστατο διανυσματικό πεδίο κίνησης. Επιθυμούμε να κάνουμε εκτίμηση του πεδίου της διδιάστατης κίνησης (στιγμιαία ταχύτητα ή μετατόπιση) από εικόνες μεταβαλλόμενες με το χρόνο και δειγματοληπτούμενες σε ένα χωροχρονικό πλέγμα  $\Lambda^3$ . Τα πεδία διδιάστατης ταχύτητας ή μετατόπισης μπορεί να μην είναι πάντοτε δυνατό να παρατηρηθούν για διάφορους λόγους, οι οποίοι παρατίθενται παρακάτω. Αντιθέτως, αυτό το οποίο παρατηρούμε είναι το πεδίο της επονομαζόμενης 'φαινόμενης' κίνησης (οπτική ροή ή αντιστοίχιση). Σκοπός αυτής της παραγράφου είναι να καταστήσει σαφή τη διάκριση μεταξύ των πεδίων διδιάστατης ταχύτητας και της οπτικής ροής, και των πεδίων διδιάστατης μετατόπισης και αντιστοίχισης, αντίστοιχα.

## Διδιάστατη Κίνηση

Η διδιάστατη κίνηση, η οποία καλείται και ‘προβαλλόμενη κίνηση’, αναφέρεται στην προοπτική ή την ορθογραφική προβολή της τρισδιάστατης κίνησης στο επίπεδο της εικόνας. Η τρισδιάστατη κίνηση μπορεί να χαρακτηριστεί σε σχέση με είτε την 3-D στιγμιαία ταχύτητα (από εδώ και στο εξής απλώς ταχύτητα) ή την 3-D μετατόπιση των σημείων του αντικειμένου. Εκφράσεις των προβολών των διανυσμάτων της τρισδιάστατης μετατόπισης και ταχύτητας με την υπόθεση της στερεάς κίνησης, αναπτύσσονται σε παρακάτω κεφάλαια.



Σχήμα 5.1: Το διδιάστατο διάνυσμα μετατόπισης

Η βασική ιδέα του διδιάστατου διανύσματος μετατόπισης φαίνεται στην εικόνα 5.1. Ας θεωρήσουμε ότι το αντικείμενο που βρίσκεται στο σημείο  $\mathbf{P}$  σε χρόνο  $t$ , μετακινείται στο σημείο  $\mathbf{P}'$  στο χρόνο  $t'$ . Η προοπτική προβολή των σημείων  $\mathbf{P}$  και  $\mathbf{P}'$  στο επίπεδο της εικόνας δίνει τα αντίστοιχα σημεία εικόνας  $\mathbf{p}$  και  $\mathbf{p}'$ . Το σχήμα 5.2, απεικονίζει μια διδιάστατη όψη της κίνησης του σημείου της εικόνας από το σημείο  $\mathbf{p}_t$  σε χρόνο  $t$  στο σημείο  $\mathbf{p}_{t'}$  στο χρόνο  $t'$ , ως την προοπτική προβολή της τρισδιάστατης κίνησης των αντίστοιχων σημείων του αντικειμένου. Σημειώστε ότι εξαιτίας της διαδικασίας της προβολής, όλα τα τρισδιάστατα διανύσματα μετατόπισης των οποίων οι κορυφές βρίσκονται στη διακεκομμένη γραμμή θα δώσουν το ίδιο διδιάστατο διάνυσμα μετατόπισης.

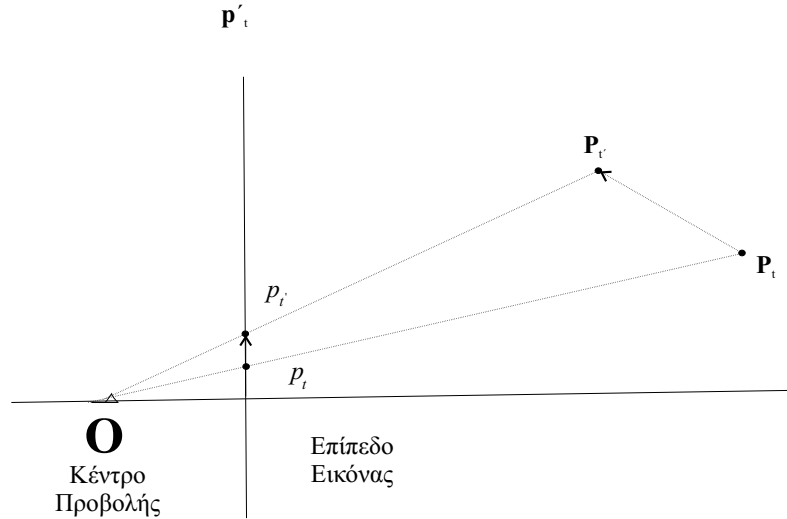
Η προβαλλόμενη μετατόπιση μεταξύ των χρόνων  $t$  και  $t' = t + \ell \Delta t$ , όπου  $\ell$  είναι ένας αθέρατος και  $\Delta t$  είναι το χρονικό διάστημα δειγματοληψίας, μπορεί να οριστεί για όλα τα σημεία  $(\mathbf{x}, t) \in \mathbf{R}^3$ , δίνοντας μια συνάρτηση πραγματικών τιμών του διδιάστατου διανύσματος μετατόπισης  $\mathbf{d}_c(\mathbf{x}, t; \ell \Delta t)$  των συνεχόμενων χωρο-χρονικών μεταβλητών. Το διδιάστατο διάνυσμα μετατόπισης αναφέρεται σε μια δειγματοληπτούμενη αναπαράσταση της συνάρτησης αυτής, που δίνεται από τη σχέση:

$$\mathbf{d}_p(\mathbf{x}, t; \ell \Delta t) = \mathbf{d}_c(\mathbf{x}, t; \ell \Delta t), \quad (\mathbf{x}, t) \in \Lambda^3 \quad (5.1)$$

ή, ισοδύναμα,

$$\mathbf{d}(\mathbf{n}, k; \ell) = \mathbf{d}_p(\mathbf{x}, t; \ell \Delta t)|_{[x_1 \ x_2 \ t]^T = \mathbf{V}[n_1 \ n_2 \ k]^T}, \quad (\mathbf{n}, k) \in Z^3 \quad (5.2)$$

όπου  $\mathbf{V}$  είναι ο πίνακας δειγματοληψίας του πλέγματος  $\Lambda^3$ . Έτσι, το πεδίο της διδιάστατης μετατόπισης είναι μια συλλογή διδιάστατων διανυσμάτων μετατόπισης,  $\mathbf{d}(\mathbf{x}, t; \ell \Delta t)$ , όπου  $(\mathbf{x}, t) \in \Lambda^3$ .



Σχήμα 5.2: Η προβαλλόμενη κίνηση της εικόνας από το σημείο  $p$  στο  $p'$

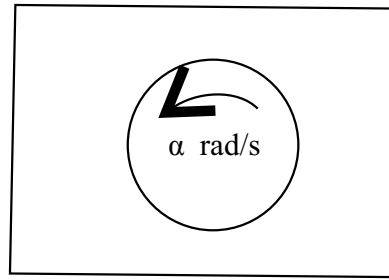
Η συνάρτηση της προβαλλόμενης ταχύτητας  $\mathbf{v}_c(\mathbf{x}, t)$  στο χρόνο  $t$ , και το πεδίο του διανύσματος της διδιάστατης ταχύτητας  $\mathbf{v}_p(\mathbf{x}, t) = v(\mathbf{n}, k)$ , για  $[x_1 \ x_2 \ t]^T = \mathbf{V}[n_1, n_2, k]^T \in \Lambda^3$  και  $(\mathbf{n}, k) \in \mathbf{Z}^3$ , μπορούν παρομοίως να οριστούν με βάση την τρισδιάστατη στιγμιαία ταχύτητα  $(\dot{X}_1, \dot{X}_2, \dot{X}_3)$ , όπου η στιγμή  $(\cdot)$  δηλώνει μια χρονική παράγωγο.

### Αντιστοίχιση και Οπτική Ροή

Η μετατόπιση των συντεταγμένων  $\mathbf{x}$  του επιπέδου της εικόνας από το χρόνο  $t$  στο  $t'$  που βασίζεται στις μεταβολές του  $s_c(\mathbf{x}, t)$ , ονομάζεται διάνυσμα αντιστοίχισης. Ένα διάνυσμα οπτικής ροής ορίζεται ως ο χρονικός ρυθμός αλλαγής των συντεταγμένων του επιπέδου της εικόνας,  $(v_1, v_2) = (dx_1/dt, dx_2/dt)$ , σε ένα συγκεκριμένο σημείο  $(\mathbf{x}, t) \in \mathbf{R}^3$ , όπως καθορίζεται από τις χωρο-χρονικές μεταβολές του προτύπου της έντασης  $s_c(\mathbf{x}, t)$ , το οποίο αντιστοιχεί στο διάνυσμα της στιγμιαίας ταχύτητας των εικονοστοιχείων. (Θεωρητικά, τα διανύσματα οπτικής ροής και αντιστοίχισης είναι ταυτόσημα όταν το όριο  $\Delta t = t' - t$  τείνει στο 0). Στην πραγματικότητα, ορίζουμε το πεδίο της αντιστοίχισης (οπτικής ροής) ως ένα διάνυσμα του πεδίου των μετατοπίσεων (ταχυτήτων) των εικονοστοιχείων με βάση τις παρατηρούμενες μεταβολές του διδιάστατου προτύπου της έντασης της εικόνας στο χωρο-χρονικό πλέγμα  $\Lambda^3$ . Τα πεδία της αντιστοίχισης και της οπτικής ροής είναι επίσης γνωστά και ως τα πεδία της 'φαινόμενης διδιάστατης μετατόπισης' και της 'φαινόμενης διδιάστατης ταχύτητας', αντίστοιχα.

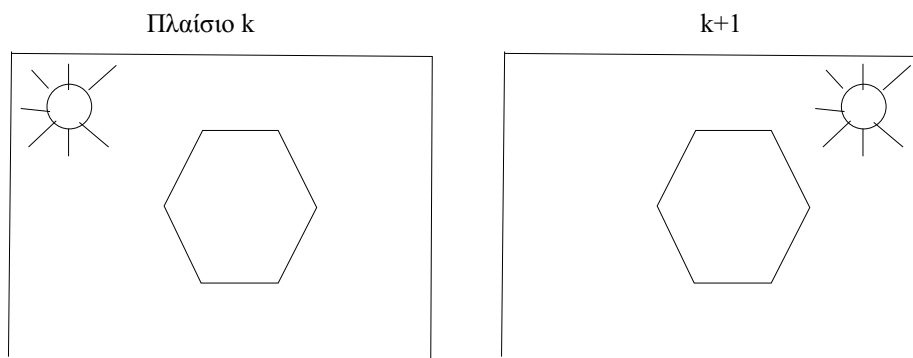
Το πεδίο της αντιστοίχισης (οπτικής ροής) είναι, γενικώς, διαφορετικό από το πεδίο της διδιάστατης μετατόπισης (διδιάστατης ταχύτητας) λόγω [Ver89]:

- Έλλειψης επαρκούς γνώσης της χωρικής διεύθυνσης της εικόνας: Πρέπει να υπάρχει επαρκής μεταβολή της στάθμης του γκρίζου (χρώμα) στην περιοχή κίνησης για να είναι παρατηρήσιμη η πραγματική κίνηση. Ένα παράδειγμα μη παρατηρήσιμης κίνησης δείχνεται στη εικόνα 5.3, όπου ένας κύκλος με ομοιόμορφη κατανομή φωτεινής έντασης περιστρέφεται γύρω από το κέντρο του. Αυτή η κίνηση δεν παράγει καμιά οπτική ροή, και έτσι είναι μη παρατηρήσιμη.
- Αλλαγών στον εξωτερικό φωτισμό: Μια παρατηρήσιμη οπτική ροή μπορεί να μην αντιστοιχεί πάντοτε σε πραγματική κίνηση. Για παράδειγμα, αν ο εξωτερικός φωτισμός αλλάζει από πλαίσιο σε πλαίσιο, όπως φαίνεται στην εικόνα 5.4, τότε θα παρατηρηθεί μια οπτική ροή



Σχήμα 5.3: Μη παρατηρήσιμη οπτική κίνηση

ακόμη και αν δεν υπάρχει κίνηση. Επομένως, αλλαγές στον εξωτερικό φωτισμό εξασθενούν τη δυνατότητα της σωστής εκτίμησης του πραγματικού πεδίου της διδιάστατης κίνησης.



Σχήμα 5.4: Οπτική ροή που δημιουργείται από αλλαγές στον εξωτερικό φωτισμό

Σε μερικές περιπτώσεις, η φωτοσκίαση μπορεί να διαφέρει από πλαίσιο σε πλαίσιο ακόμη και αν δεν υπάρχει αλλαγή στον εξωτερικό φωτισμό. Για παράδειγμα, αν ένα αντικείμενο περιστρέφεται, το κάθετο στην επιφάνεια διάνυσμά του αλλάζει, κάτι το οποίο έχει ως αποτέλεσμα αλλαγή στη φωτοσκίαση. Αυτή η αλλαγή στη φωτοσκίαση μπορεί να κάνει την ένταση των εικονοστοιχείων κατά μήκος της τροχιάς της κίνησης να μεταβάλλεται, το οποίο πρέπει να λαμβάνεται υπόψη κατά την εκτίμηση της διδιάστατης κίνησης.

Συμπερασματικά, τα πεδία της διδιάστατης μετατόπισης και ταχύτητας είναι προβολές των αντιστοιχών τρισδιάστατων πεδίων στο επίπεδο της εικόνας, ενώ τα πεδία της αντιστοίχισης και της οπτικής ροής είναι οι συναρτήσεις ταχύτητας και μετατόπισης όπως παρατηρούνται από το πρότυπο της χρονικά μεταβαλλόμενης έντασης της εικόνας. Όμως, εφόσον μπορούμε να παρατηρήσουμε μόνο τα πεδία της οπτικής ροής και της αντιστοίχισης, θα θεωρήσουμε ότι είναι τα ίδια με το πεδίο της διδιάστατης κίνησης στο υπόλοιπο αυτού του βιβλίου, εκτός αν αναφέρεται διαφορετικά.

### 5.1.2 Εκτίμηση Διδιάστατης κίνησης

Σε αυτήν την παράγραφο, αρχικά, θα εκθέσουμε τα προβλήματα της εκτίμησης της αντιστοίχισης και της οπτικής ροής ενός εικονοστοιχείου. Στη συνέχεια, θα συζητήσουμε την ‘κακώς διατυπωμένη’ φύση αυτών των προβλημάτων, και θα αναφέρουμε μερικά μοντέλα πεδίων εκ των προτέρων διδιάστατης κίνησης.

Το πρόβλημα της εκτίμησης της διδιάστατης κίνησης μπορεί να τεθεί είτε ως:

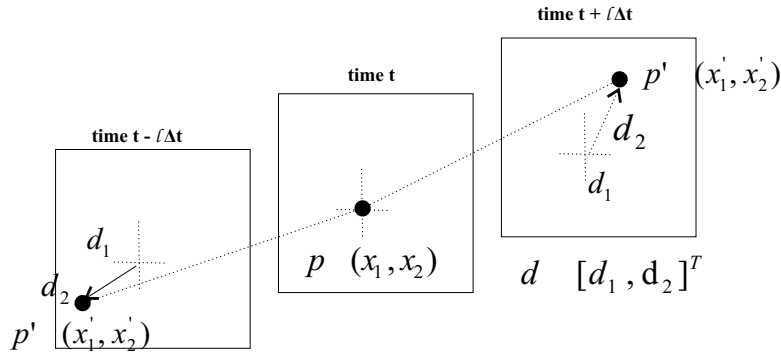
α) εκτίμηση των διανυσμάτων αντιστοίχισης του επιπέδου της εικόνας

$\mathbf{d}(\mathbf{x}, t; \ell\Delta t) = [d_1(\mathbf{x}, t; \ell\Delta t) \ d_2(\mathbf{x}, t; \ell\Delta t)]^T$  μεταξύ των χρόνων  $t$  και  $t + \ell\Delta t$ , για όλα τα  $(\mathbf{x}, t) \in \Lambda^3$  και όπου  $\ell$  είναι ένας ακέραιος, ή

β) εκτίμηση των διανυσμάτων οπτικής ροής  $\mathbf{v}(\mathbf{x}, t) = [v_1(\mathbf{x}, t) \ v_2(\mathbf{x}, t)]^T$  για όλα τα  $(\mathbf{x}, t) \in \Lambda^3$ .

Παρατηρήστε ότι ο δείκτης ' $p$ ' δεν χρησιμοποιείται για λόγους σημειογραφικής απλότητας. Τα διανύσματα της αντιστοίχισης και της οπτικής ροής συνήθως διαφέρουν από εικονοστοιχείο σε εικονοστοιχείο (χωρο-μεταβαλλόμενη κίνηση), όπως για παράδειγμα εξαιτίας περιστροφής των αντικειμένων στη σκηνή, και ως συνάρτηση του χρόνου, εξαιτίας επιτάχυνσης των αντικειμένων.

*Το πρόβλημα της αντιστοίχισης:* Το πρόβλημα της αντιστοίχισης μπορεί να θεωρηθεί ως ένα πρόβλημα εκτίμησης κίνησης προς τα μπρος ή προς τα πίσω, ανάλογα με το αν το διάνυσμα της κίνησης ορίζεται από το χρόνο  $t$  προς το χρόνο  $t + \ell\Delta t$  ή από το χρόνο  $t$  προς τα πίσω προς το χρόνο  $t - \ell\Delta t$ , όπως δείχνει η εικόνα 5.5.



Σχήμα 5.5: Το πρόβλημα της αντιστοίχισης προς τα εμπρός και προς τα πίσω

*Εκτίμηση προς τα μπρος:* Δοθέντων των χωρο-χρονικών δειγμάτων  $s_p(\mathbf{x}, t)$  στους χρόνους  $t$  και  $t + \ell\Delta t$ , τα οποία συνδέονται με τη σχέση

$$s_p(x_1, x_2, t) = s_c(x_1 + d_1(\mathbf{x}, t; \ell\Delta t), x_2 + d_2(\mathbf{x}, t; \ell\Delta t), t + \ell\Delta t) \quad (5.3)$$

ή, ισοδύναμα,

$$s_k(x_1, x_2) = s_{k+\ell}(x_1 + d_1(\mathbf{x}), x_2 + d_2(\mathbf{x})), \quad \text{έτσι ώστε } t = k\Delta t$$

βρείτε το διάνυσμα αντιστοίχισης πραγματικής τιμής  $\mathbf{d}(\mathbf{x}) = [d_1(\mathbf{x}) \ d_2(\mathbf{x})]^T$ , όπου τα χρονικά ορίσματα  $\mathbf{d}(\mathbf{x})$  δε λαμβάνονται υπόψη.

*Εκτίμηση προς τα πίσω:* Αν ορίσουμε τα διανύσματα αντιστοίχισης από το χρόνο  $t$  προς το χρόνο  $t - \ell\Delta t$ , τότε το μοντέλο της διδιάστατης κίνησης γίνεται

$$s_k(x_1, x_2) = s_{k-\ell}(x_1 + d_1(\mathbf{x}), x_2 + d_2(\mathbf{x})), \quad \text{έτσι ώστε } t = k\Delta t$$

Εναλλακτικά, το διάνυσμα της κίνησης μπορεί να οριστεί στο χρονικό διάστημα  $(t - \ell\Delta t, t)$ . Τότε θα έχουμε :

$$s_k(x_1, x_2) = s_{k-\ell}(x_1 - d_1(\mathbf{x}), x_2 - d_2(\mathbf{x})), \quad \text{έτσι ώστε } t = k\Delta t$$

Αν και θα εξεταστούν και οι δυο τρόποι εκτίμησης της κίνησης, η προς τα πίσω μέθοδος εκτίμησης της κίνησης είναι πιο εύχρηστη για αντιστάθμιση εμπρόσθιας κίνησης, η οποία εμπλέκεται συνήθως στη συμπίεση βίντεο με πρόβλεψη. Παρατηρήστε ότι επειδή  $\mathbf{x} \pm \mathbf{d}(\mathbf{x})$  δεν ανταποκρίνεται γενικώς σε πλέγμα, τα δεξιά μέλη των εκφράσεων δίνονται σε όρους συνεχούς βίντεο  $s_c(x_1, x_2, t)$ ,

το οποίο δεν είναι διαθέσιμο. Έτσι, οι περισσότερες μέθοδοι εκτίμησης αντιστοίχισης περιλαμβάνουν και κάποιο μέθοδο παρεμβολής. Το πρόβλημα της αντιστοίχισης συναντάται επίσης και στην εκτίμηση της στερεοσκοπικής διαφοράς, όπου έχουμε ένα αριστερό-δεξιό ζευγάρι αντί για ένα χρονικό ζευγάρι εικόνων.

**Ταύτιση Εικόνας:** Το πρόβλημα της ταύτισης είναι μια ειδική περίπτωση του προβλήματος της αντιστοίχισης, όπου τα δυο πλαίσια μετατοπίζονται συνολικά το ένα σε σχέση με το άλλο, για παράδειγμα, πολλαπλά στιγμιότυπα μιας στατικής σκηνής με μια μετατοπιζόμενη κάμερα.

**Εκτίμηση Οπτικής Ροής:** Δοθέντων των δειγμάτων  $s_p(x_1, x_2, t)$  σε ένα τρισδιάστατο πλέγμα  $\Lambda^3$ , βρείτε τη διδιάστατη ταχύτητα  $\mathbf{v}(\mathbf{x}, t)$  για όλα τα  $(\mathbf{x}, t) \in \Lambda^3$ . Βεβαίως, η εκτίμηση των διανυσμάτων της οπτικής ροής και της αντιστοίχισης από δυο πλαίσια είναι ίδια, με  $\mathbf{d}(\mathbf{x}, t; \ell\Delta t) = \mathbf{v}(\mathbf{x}, t)\ell\Delta t$ , υποθέτοντας ότι η ταχύτητα παραμένει σταθερή κατά τη διάρκεια κάθε χρονικού διαστήματος  $\ell\Delta t$ . Σημειώστε ότι στην περίπτωση επιτάχυνσης χρειάζεται να θεωρήσουμε περισσότερα από δυο πλαίσια τη φορά για την εκτίμηση της οπτικής ροής.

Το πρόβλημα της εκτίμησης της διδιάστατης κίνησης, θεωρώντας το είτε ως πρόβλημα αντιστοίχισης είτε ως πρόβλημα εκτίμησης οπτικής ροής, βασισμένο μόνο σε δυο πλαίσια, είναι ένα ‘κακώς διατυπωμένο’ πρόβλημα, όταν απουσιάζουν και άλλες επιπρόσθετες υποθέσεις για τη φύση της κίνησης. Ένα πρόβλημα καλείται κακώς διατυπωμένο αν δεν υπάρχει μια μοναδική λύση, και/ή η λύση ή οι λύσεις δεν εξαρτώνται συνεχώς από τα δεδομένα [Ber88]. Η εκτίμηση της διδιάστατης κίνησης υποφέρει από όλα τα προβλήματα ύπαρξης, μοναδικότητας και συνέχειας:

- **Ύπαρξη λύσης:** Καμία αντιστοίχιση δεν μπορεί να υπάρξει για καλυμμένα/μη καλυμμένα σημεία φόντου. Αυτό είναι γνωστό ως το πρόβλημα της κάλυψης (occlusion).
- **Μοναδικότητα λύσης:** Αν οι συνιστώσες της μετατόπισης (ή της ταχύτητας) σε κάθε εικονοστοιχείο λογίζονται ως ανεξάρτητες μεταβλητές, τότε ο αριθμός των αγνώστων είναι διπλάσιος του αριθμού των παρατηρήσεων. Αυτό οδηγεί και στο λεγόμενο πρόβλημα του διαφράγματος (aperture).
- **Συνέχεια της λύσης:** Η εκτίμηση της κίνησης είναι πάρα πολύ ευαίσθητη στην παρουσία του θορύβου παρατήρησης στις εικόνες τις προερχόμενες από βίντεο. Ένα μικρό ποσό θορύβου μπορεί να προκαλέσει σημαντική απόκλιση στην εκτιμώμενη κίνηση.

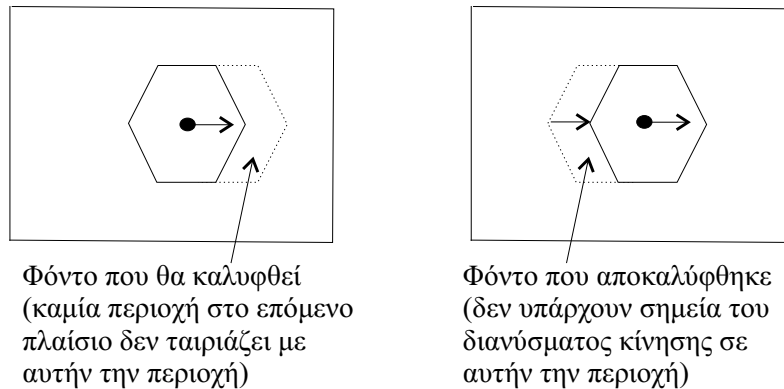
Τα προβλήματα της κάλυψης και του διαφράγματος περιγράφονται λεπτομερώς στις επόμενες παραγράφους.

## Το πρόβλημα της κάλυψης

Το πρόβλημα αυτό αναφέρεται στην κάλυψη/αποκάλυψη του φόντου εξαιτίας της τρισδιάστατης περιστροφής και μετατόπισης ενός αντικειμένου, το οποίο καταλαμβάνει ένα μόνο μέρος του οπτικού πεδίου. Η γενική ιδέα του καλυμμένου και του μη καλυμμένου φόντου φαίνεται στην εικόνα 5.6, όπου το αντικείμενο που σχηματίζεται από τις συμπαγείς γραμμές μετατοπίζεται κατά τη  $x_1$  διεύθυνση στο χρονικό διάστημά από  $t$  σε  $t'$ .

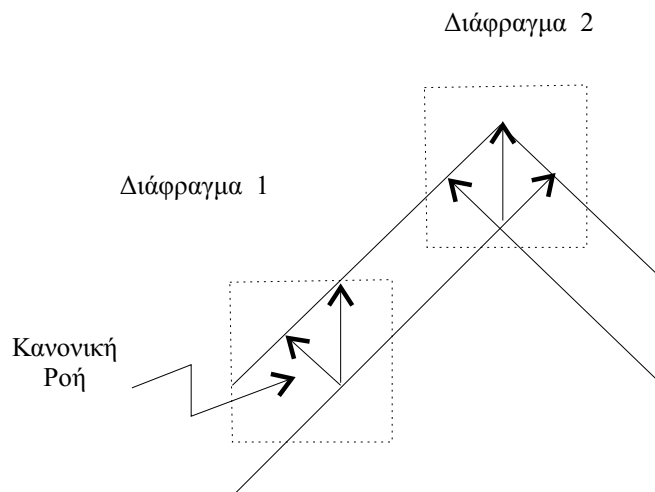
## Το πρόβλημα του διαφράγματος

Το πρόβλημα του διαφράγματος είναι παρελκόμενο του γεγονότος ότι η λύση στο πρόβλημα της εκτίμησης της διδιάστατης κίνησης δεν είναι μοναδική. Αν τα διανύσματα της κίνησης σε κάθε εικονοστοιχείο θεωρηθούν ως ανεξάρτητες μεταβλητές, τότε ο αριθμός των αγνώστων είναι διπλάσιος του αριθμού των εξισώσεων, όπως αυτές δίνονται από την εξίσωση (5.3). Ο αριθμός των εξισώσεων είναι ίσος με τον αριθμό των εικονοστοιχείων της εικόνας, αλλά για κάθε εικονοστοιχείο το διάνυσμα της κίνησης έχει δυο συνιστώσες.



Σχήμα 5.6: Το πρόβλημα της κάλυψης

Η θεωρητική ανάλυση, του υποκεφαλαίου 5.2 καταδεικνύει ότι μπορούμε να ερμηνεύσουμε, σε κάθε εικονοστοιχείο, μόνο την κίνηση που είναι ορθογώνια στη χωρική κλίση της εικόνας και η οποία καλείται κανονική ροή. Το πρόβλημα του διαφράγματος φαίνεται στην εικόνα 5.7. Υποθέστε ότι έχουμε μια γωνία ενός αντικειμένου που κινείται στη  $x_2$  διεύθυνση (προς τα πάνω). Αν εκτιμήσουμε την κίνηση με βάση ένα τοπικό παράθυρο, που υποδεικνύεται από το Διάφραγμα 1, τότε δεν είναι δυνατό να καθορίσουμε αν η εικόνα κινείται προς τα πάνω ή κάθετα προς την ακμή. Η κίνηση που είναι κάθετη προς την ακμή καλείται κανονική ροή.



Σχήμα 5.7: Το πρόβλημα του διαφράγματος

Παρ' όλα αυτά, αν παρατηρήσουμε το Διάφραγμα 2, είναι πιθανό να εκτιμήσουμε τη σωστή κίνηση, αφού η εικόνα έχει κλίση προς δυο κατακόρυφες διευθύνσεις μέσα σ' αυτό το διάφραγμα. Έτσι, είναι δυνατό να ξεπεράσουμε το πρόβλημα του διαφράγματος, εκτιμώντας την κίνηση με βάση ένα τμήμα από εικονοστοιχεία το οποίο περιέχει αρκετή διαβάθμιση του γκρι. Πρέπει να τονιστεί ότι γίνεται η σιωπηρή υπόθεση ότι όλα τα εικονοστοιχεία μεταφέρονται από το ίδιο διάνυσμα κίνησης. Μια λιγότερο περιοριστική προσέγγιση θα ήταν να αναπαρασταθεί η διαχύμανση των διανυσμάτων κίνησης από εικονοστοιχείο σε εικονοστοιχείο από κάποια παραμετρικά ή μη παραμετρικά μοντέλα πεδίων διδιάστατης κίνησης.

### 5.1.3 Μοντέλα Πεδίου Διδιάστατης Κίνησης

Εξαιτίας της κακώς διατυπωμένης φύσης του προβλήματος, οι αλγόριθμοι εκτίμησης της κίνησης χρειάζονται επιπρόσθετες υποθέσεις (μοντέλα) γύρω από τη δομή του πεδίου της διδιάστατης κίνησης. Για το λόγο αυτό, παρέχεται μια σύντομη επισκόπηση των μοντέλων αυτών.

#### Παραμετρικά Μοντέλα

Τα παραμετρικά μοντέλα αποσκοπούν στο να περιγράψουν την ορθογραφική ή προοπτική προβολή της τρισδιάστατης κίνησης (μετατόπιση ή ταχύτητα) μιας επιφάνειας στο επίπεδο της εικόνας. Γενικά, τα μοντέλα του πεδίου της παραμετρικής διδιάστατης κίνησης εξαρτώνται από την αναπαράσταση της τρισδιάστατης επιφάνειας. Για παράδειγμα, ένα πεδίο διδιάστατης κίνησης που προκύπτει από την τρισδιάστατη στερεά κίνηση μιας επίπεδης επιφάνειας υπό ορθογραφική προβολή μπορεί να περιγραφεί από ένα αφινικό μοντέλο 6 παραμέτρων, ενώ υπό προοπτική προβολή μπορεί να περιγραφεί από ένα μη γραμμικό μοντέλο 8 παραμέτρων [Ana93]. Υπάρχουν, επίσης, και πιο πολύπλοκα μοντέλα για τετραγωνικές επιφάνειες [Agg88].

Μια υποκατηγορία των παραμετρικών μοντέλων είναι και τα λεγόμενα σχεδόν-παραμετρικά μοντέλα, τα οποία θεωρούν το βάθος κάθε τρισδιάστατου σημείου ως έναν ανεξάρτητο άγνωστο. Τότε οι 6 παράμετροι της τρισδιάστατης κίνησης περιορίζουν το διάνυσμα της τοπικής ροής της εικόνας να εκτείνεται κατά μήκος μιας συγκεκριμένης γραμμής, ενώ γνώση της τιμής του τοπικού βάθους απαιτείται για να υπολογιστεί η ακριβής τιμή του διανύσματος κίνησης [Ana93]. Αυτά τα μοντέλα μπορούν να εξυπηρετήσουν ως περιορισμοί για να ρυθμίσουν την εκτίμηση των διανυσμάτων της διδιάστατης κίνησης και να οδηγήσουν έτσι σε ταυτόχρονους σχηματισμούς της εκτίμησης της διδιάστατης και της τρισδιάστατης κίνησης.

#### Μη Παραμετρικά Μοντέλα

Το κυριότερο μειονέκτημα των παραμετρικών μοντέλων είναι ότι είναι εφαρμόσιμα μόνο στην περίπτωση της τρισδιάστατης στερεάς κίνησης. Από την άλλη, οι μη παραμετρικοί περιορισμοί ομοιομορφίας (εξομάλυνσης) μπορούν να εφαρμοστούν στο πεδίο της διδιάστατης κίνησης χωρίς να εμπλέκουν τα μοντέλα της τρισδιάστατης στερεάς κίνησης. Οι μη παραμετρικοί περιορισμοί μπορούν να ταξινομηθούν ως ντετερμινιστικοί σε αντίθεση με τα στοχαστικά μοντέλα ομαλότητας. Ακολουθεί μια σύντομη προεπισκόπηση των μη παραμετρικών προσεγγίσεων.

- *Μοντέλο της κίνησης κατά τμήματα:* Γίνεται η υπόθεση ότι η εικόνα συντίθεται από μετακινούμενα τμήματα. Θα αναφέρουμε δυο προσεγγίσεις για την εύρεση της μετατόπισης των τμημάτων από πλαίσιο σε πλαίσιο: τη μέθοδο συσχέτισης φάσης και τη μέθοδο του ταιριάσματος τμημάτων. Στην προσέγγιση συσχέτισης φάσης, ο γραμμικός όρος της διαφοράς της Fourier φάσης μεταξύ δυο συνεχόμενων πλαισίων καθορίζει την εκτίμηση κίνησης. Η μέθοδος ταιριάσματος τμημάτων φάχνει για τη θέση που ταιριάζει καλύτερα, ενός τμήματος συγκεκριμένου μεγέθους στο επόμενο (και/ή στο προηγούμενο) πλαίσιο(α), με βάση ένα κριτήριο απόστασης. Η βασική μορφή και των δύο μεθόδων εφαρμόζεται μόνο στην περίπτωση της κίνησης γραμμικής μετατόπισης. Παρ' όλα αυτά, μια γενικευμένη μέθοδος ταιριάσματος τμημάτων μπορεί να ενσωματώσει και άλλους χωρικούς μετασχηματισμούς. Η εκτίμηση της κίνησης που βασίζεται στα τμήματα καλύπτεται στην παράγραφο 5.2.
- *Μέθοδοι βασισμένες στην εξίσωση οπτικής ροής (EOP):* Οι μέθοδοι που βασίζονται στην EOP επιχειρούν να δώσουν μια εκτίμηση του πεδίου της οπτικής ροής σε σχέση με την κλίση της χωρο-χρονικής έντασης της εικόνας. Με μονόχρωμες εικόνες, η EOP πρέπει να χρησιμοποιηθεί σε συνδυασμό με τον κατάλληλο χωρο-χρονικό περιορισμό εξομάλυνσης, ο οποίος



απαιτεί ότι το διάνυσμα της μετατόπισης μεταβάλλεται πολύ αργά στο χώρο μιας γειτονιάς. Στην περίπτωση των έγχρωμων εικόνων, η ΕΟΡ μπορεί να χρησιμοποιηθεί για κάθε ζώνη χρώματος ξεχωριστά, κάτι που μπορεί πιθανόν να περιορίσει το διάνυσμα της μετατόπισης σε τρεις διαφορετικές κατευθύνσεις [Oht 90]. Παρ' όλα αυτά, στις περισσότερες περιπτώσεις, ένας κατάλληλος περιορισμός εξομάλυνσης εξακολουθεί να είναι απαραίτητος προκειμένου να πετύχουμε ικανοποιητικά αποτελέσματα. Καθολικοί περιορισμοί εξομάλυνσης προκαλούν ανακριβή εκτίμηση κίνησης στα όρια των αντικειμένων λόγω κάλυψης. Περισσότερο προχωρημένοι κατευθυντικοί περιορισμοί εξομάλυνσης επιτρέπουν την ύπαρξη ξαφνικών ασυνεχειών στο πεδίο της κίνησης.

- *Αναδρομικές μέθοδοι με βάση τα εικονοστοιχεία:* Οι μέθοδοι αυτές είναι εκτιμητές μετατόπισης τύπου πρόβλεψης - διόρθωσης. Η πρόβλεψη λαμβάνεται ως η τιμή της εκτίμησης κίνησης στη θέση του προηγούμενου εικονοστοιχείου ή ως ο γραμμικός συνδυασμός των εκτιμήσεων κίνησης στην περιοχή της γειτονιάς του τρέχοντος εικονοστοιχείου. Η ενημέρωση βασίζεται στην ελαχιστοποίηση της κλίσης της διαφοράς του μετατοπισμένου πλαισίου (ΔΜΠ) σε αυτό το εικονοστοιχείο. Το βήμα της πρόβλεψης μπορεί γενικά να θεωρηθεί ως ένας υπονοούμενος περιορισμός εξομάλυνσης. Επέκταση αυτής της προσέγγισης στην εκτίμηση που βασίζεται σε τμήματα οδηγεί στις λεγόμενες στρατηγικές εκτίμησης τύπου-Wiener. Οι αναδρομικές μέθοδοι εκτίμησης κίνησης με βάση τα εικονοστοιχεία παρουσιάζονται στη παράγραφο 5.3.
- *Bayesian μέθοδοι:* Οι μέθοδοι αυτές εκμεταλλεύονται τους πιθανοτικούς περιορισμούς εξομάλυνσης, συνήθως στη μορφή ενός τυχαίου πεδίου Gibbs, προκειμένου να κάνουν εκτίμηση του πεδίου της μετατόπισης. Το κυριότερο μειονέκτημά τους είναι ο μεγάλος αριθμός υπολογισμών που απαιτείται.

#### 5.1.4 Μετρικές ποιότητας εκτιμητών κίνησης

Υπολογίζουμε την ποιότητα των εκτιμητών κίνησης με βάση τον λόγο σήματος-προς-θόρυβο (PSNR) της τελικής διαφοράς μετατοπισμένου πλαισίου - (ΔΜΠ) μεταξύ του αρχικού και τελικού πλαισίου, που ορίζεται από :

$$PSNR = 10 \log_{10} \frac{255 \times 255}{\sum [s_k(n_1, n_2) - s_{k+l}(n_1 + d_1(n_1, n_2), n_2 + d_2(n_1, n_2))]^2} \quad (5.4)$$

όπου  $d_1, d_2$  είναι οι συνιστώσες των εκτιμητών κίνησης σε κάθε εικονοστοιχείο. Επίσης, υπολογίζουμε την εντροπία του εκτιμώμενου πεδίου της διδιάστατης κίνησης, που δίνεται από την:

$$H = - \sum_{d_1} P(d_1) \log_2 P(d_1) - \sum_{d_2} P(d_2) \log_2 P(d_2) \quad (5.5)$$

όπου τα  $P(d_1), P(d_2)$  δηλώνουν τη σχετική συχνότητα εμφάνισης της οριζόντιας και κάθετης συνιστώσας του διανύσματος κίνησης  $\mathbf{d}$ . Η εντροπία, εκτός από το να αποτελεί ένα μέτρο της εξομάλυνσης του πεδίου της κίνησης, είναι ιδιαίτερου ενδιαφέροντος στη συμπίεση βίντεο με αντιστάθμιση κίνησης, όπου είναι επιθυμητή η ελαχιστοποίηση τόσο της εντροπίας του πεδίου της κίνησης (για φθηνότερη εκπομπή), όσο και της ενέργειας της ΔΜΠ.

## 5.2 Εκτίμηση Κίνησης Κατά Τμήματα

Η κατά τμήματα εκτίμηση και αντιστάθμιση κίνησης είναι πολύ δημοφιλής και έχει υιοθετηθεί στα διεθνή πρότυπα για τη συμπίεση ψηφιακού βίντεο, όπως τα H.261 και MPEG 1/2. Παρόλο που αυτά τα πρότυπα δεν ορίζουν πλήρως κάποια συγκεκριμένη μέθοδο εκτίμησης, η κατά τμήματα εκτίμηση

κίνησης έχει γίνει μία φυσική επιλογή. Η κατά τμήματα εκτίμηση κίνησης χρησιμοποιείται, επίσης, ευρέως σε αρκετές άλλες εφαρμογές ψηφιακού βίντεο όπως το φιλτράρισμα που βασίζεται στην αντιστάθμιση κίνησης για μετατροπές προτύπων.

Αρχίζουμε με μία σύντομη εισαγωγή των κατά τμήματα μοντέλων εκτίμησης κίνησης στη παράγραφο 5.2.1. Σ' αυτό το κεφάλαιο παρουσιάζουμε δύο στρατηγικές εκτίμησης κίνησης. Η πρώτη προσέγγιση που θα συζητηθεί στην παράγραφο 5.2.2 είναι μία τεχνική στο πεδίο των χωρικών συχνοτήτων και ονομάζεται μέθοδος συσχέτισης φάσης. Η δεύτερη, που θα παρουσιασθεί στην παράγραφο 5.2.3 είναι μία προσέγγιση αναζήτησης στο χωρικό πεδίο. Και οι δύο μέθοδοι μπορούν να υλοποιηθούν ιεραρχικά χρησιμοποιώντας μια πολυδιακριτική περιγραφή του βίντεο.

### 5.2.1 Μοντέλα κίνησης των τμημάτων

Το μοντέλο κίνησης κατά τμήματα υποθέτει ότι η εικόνα αποτελείται από κινούμενα τμήματα που συνήθως αντιστοιχίζονται σε κινούμενα αντικείμενα. Λαμβάνουμε υπόψη δύο τύπους τμηματικής κίνησης: α) απλή  $2 - D$  μετατόπιση και β) διάφορες  $2 - D$  παραμορφώσεις των τμημάτων.

#### Μεταφορική Κίνηση Κατά Τμήματα

Η απλούστερη μορφή αυτού του μοντέλου είναι αυτή των τμημάτων που μετακινούνται, περιορίζοντας τη κίνηση του κάθε τμήματος σε μία απλή μετατόπιση. Τότε ένα  $N \times N$  τμήμα  $B$  στο πλαίσιο  $\mathbf{k}$  κεντραρισμένο γύρω από το εικονοστοιχείο  $\mathbf{v} = (n_1, n_2)$  μοντελοποιείται ως μία γενική μετατοπισμένη έκδοση ενός ίδιου μεγέθους τμήματος στο πλαίσιο  $k+l$ , για κάποιο ακέραιο  $l$ :

$$s(n_1, n_2, k) = s_c(x_1 + d_1, x_2 + d_2, t + l\Delta t) \left| \begin{bmatrix} \mathbf{x} \\ t \end{bmatrix} = \mathbf{v} \begin{bmatrix} \mathbf{n} \\ k \end{bmatrix} \right| \quad d_1, d_2 \in \mathbf{R} \quad (5.6)$$

για όλα τα  $(n_1, n_2) \in B$ , όπου τα  $d_1$  και  $d_2$  είναι οι συνιστώσες των διανυσμάτων μετατόπισης για το τμήμα  $B$ . Το δεξί μέρος της εξίσωσης (5.6) δίνεται σε σχέση με τις συνεχείς χρονικά μεταβαλλόμενες εικόνες  $s_c(x_1, x_2, t)$ , επειδή τα  $d_1$  και  $d_2$  παίρνουν πραγματικές τιμές. Υποθέτοντας ότι τα  $d_1$  και  $d_2$  κβαντίζονται στο πλησιέστερο ακέραιο, το μοντέλο (5.6) μπορεί να απλοποιηθεί σε

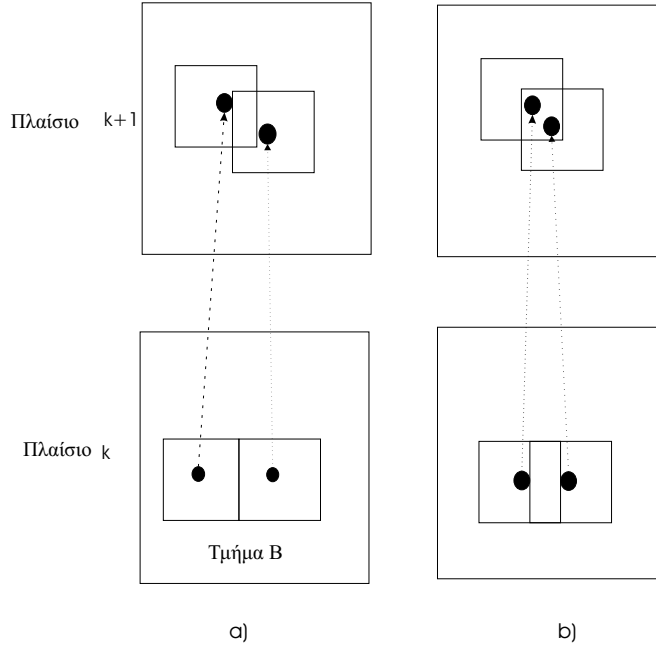
$$s(n_1, n_2, k) = s(n_1 + d_1, n_2 + d_2, k + l) \quad d_1, d_2 \in \mathbf{Z} \quad (5.7)$$

Παρατηρήστε ότι είναι πιθανό να πάρουμε ακρίβεια εικονοστοιχείων  $\frac{1}{2L}$  στις εκτιμήσεις της κίνησης, εάν στο πλαίσιο  $k+l$  στην (5.7) γίνει παρεμβολή κατά ένα παράγοντα  $L$ .

Στο μοντέλο (5.6), τα τμήματα  $B$  μπορεί να είναι επικαλυπτόμενα ή μη, όπως φαίνεται στην εικόνα 5.8 (α) και (β) αντιστοίχως. Στη περίπτωση της μη επικάλυψης, σε ολόκληρο το τμήμα αντιστοιχεί ένα μόνο διάνυσμα κίνησης. Έτσι, η εκτίμηση κίνησης μπορεί να επιτευχθεί αντιγράφοντας την πληροφορία διαβάθμισης του γκρίζου ή χρωματική πληροφορία από τα αντίστοιχα τμήματα, στο πλαίσιο  $k+l$ , εικονοστοιχείο με εικονοστοιχείο. Στην περίπτωση των επικαλυπτόμενων τμημάτων, μπορούμε είτε να υπολογίσουμε το μέσο όρο των διανυσμάτων κίνησης μέσα στις επικαλυπτόμενες περιοχές ή να επιλέξουμε ένα από τα εκτιμώμενα διανύσματα κίνησης. Η αντιστάθμιση κίνησης στην περίπτωση των επικαλυπτόμενων τμημάτων έχει συζητηθεί στο [Sul93], όπου είχε προταθεί μία προσέγγιση χρήσης πολλαπλών υποθέσεων.

Η δημοτικότητα της αντιστάθμισης και εκτίμησης κίνησης που βασίζεται στο μοντέλο των μετατοπισμένων τμημάτων προέρχεται από :

- τις χαμηλές απαιτήσεις για την αναπαράσταση του πεδίου κίνησης, καθώς ένα διάνυσμα κίνησης απαιτείται για κάθε τμήμα, και
- ύπαρξη φθηνών υλοποιήσεων VLSI.



Σχήμα 5.8: Μοντέλα κίνησης των τμημάτων: α) μη επικαλυπτόμενα β) επικαλυπτόμενα τμήματα

Ωστόσο, η αντιστάθμιση κίνησης χρησιμοποιώντας μετατοπισμένα πλαίσια: α) αποτυγχάνει στο ζουμ, στην περιστροφική κίνηση και κάτω από τοπικές παραμορφώσεις και β) έχει ως αποτέλεσμα σοβαρές παραμορφώσεις τμηματοποίησης της εικόνας (blocking-artifacts), ειδικά για εφαρμογές με χαμηλό ρυθμό bits, καθώς τα όρια των αντικειμένων γενικά δε συμφωνούν με τα όρια των αντικειμένων, και μπορεί να αποδοθούν στα γειτονικά τμήματα αρκετά διαφορετικά διανύσματα κίνησης.

### Γενικευμένη/Παραμορφώσιμη Κίνηση Κατά Τμήματα

Για να γενικεύσουμε το μοντέλο της μετατοπισμένης τμηματικής κίνησης (5.6), βλέπουμε ότι μπορεί να χαρακτηριστεί από έναν απλό (χωρικό) μετασχηματισμό των συντεταγμένων των εικονοστοιχείων από πλαίσιο-σε-πλαίσιο της μορφής

$$\begin{aligned} x'_1 &= x_1 + d_1 \\ x'_2 &= x_2 + d_2 \end{aligned} \quad (5.8)$$

όπου  $(x_1, x_2), (x'_1, x'_2)$  είναι οι συντεταγμένες ενός σημείου στα πλαίσια  $k, k+l$  αντίστοιχα. Ο χωρικός μετασχηματισμός (5.8) μπορεί να γενικευθεί για να περιλάβει αφφινικούς μετασχηματισμούς συντεταγμένων που δίνονται από τις σχέσεις :

$$\begin{aligned} x'_1 &= a_1x_1 + a_2x_2 + d_1 \\ x'_2 &= a_3x_1 + a_4x_2 + d_1 \end{aligned} \quad (5.9)$$

Ο αφφινικός μετασχηματισμός (5.9) μπορεί να χειρισθεί περιστροφή των τμημάτων, καθώς και 2-D παραμόρφωση των τετραγώνων (ορθογωνίων) σε παραλληλόγραμμα όπως δείχνει το σχήμα 5.9. Άλλοι χωρικοί μετασχηματισμοί περιλαμβάνουν τον προοπτικό και τον διγραμμικό μετασχηματισμό συντεταγμένων. Ο προοπτικός μετασχηματισμός δίνεται από τις σχέσεις:

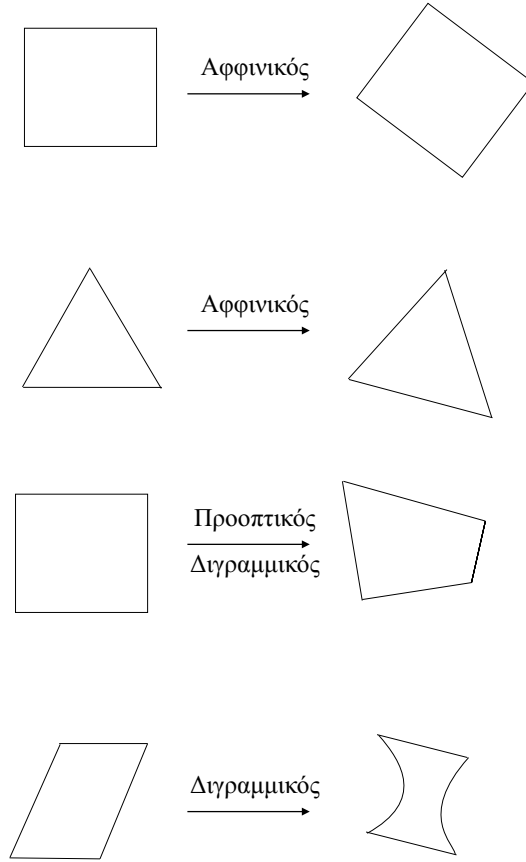
$$x'_1 = \frac{a_1x_1 + a_2x_2 + a_3}{a_7x_1 + a_8x_2 + 1}$$

$$x_2' = \frac{a_4x_1 + a_5x_2 + a_6}{a_7x_1 + a_8x_2 + 1} \quad (5.10)$$

Ο διγραμμικός μετασχηματισμός μπορεί να εκφρασθεί ως :

$$\begin{aligned} x_1' &= a_1x_1 + a_2x_2 + a_3x_1x_2 + a_4 \\ x_2' &= a_5x_1 + a_6x_2 + a_7x_1x_2 + a_8 \end{aligned} \quad (5.11)$$

Οι αφφινικοί και οι προοπτικοί μετασχηματισμοί συντεταγμένων αντιστοιχούν στις ορθογραφικές και προοπτικές προβολές μίας 3-D στερεής κίνησης μίας επίπεδης επιφάνειας αντίστοιχα. Ωστόσο, ο διγραμμικός μετασχηματισμός δε σχετίζεται με κάποια φυσική 3-D κίνηση. Οι αλγεβρικές και γεωμετρικές ιδιότητες αυτών των μετασχηματισμών έχουν αναπτυχθεί στο [Wo 90].



Σχήμα 5.9: Παραδείγματα χωρικών μετατοπίσεων

Σημειώνουμε ότι μερικά μοντέλα κίνησης κατά τμήματα, συμπεριλαμβανομένων των (5.8) και (5.9) ανήκουν στην κατηγορία των παραμετρικών μοντέλων κίνησης (που συζητήθηκε στην παράγραφο 5.1) και μπορεί να θεωρηθούν ως τοπικοί περιορισμοί κανονικοποίησης σε αυθαίρετα πεδία μετατόπισης για να αντιμετωπίσουμε το πρόβλημα του διαφράγματος, σε συνδυασμό με τις μεθόδους που βασίζονται στην εξίσωση οπτικής ροής ή στις αναδρομικές μεθόδους με βάση τα εικονοστοιχεία.

### 5.2.2 Μέθοδος Συσχέτισης Φάσης

Παίρνοντας τον 2-D μετασχηματισμό Fourier και των δύο πλευρών του μοντέλου της διακριτής κίνησης (5.7) με  $l = 1$  πάνω σε ένα τμήμα  $B$ , έχουμε:

$$S_k(f_1, f_2) = S_{k+1}(f_1, f_2) \exp(j2\pi(d_1f_1 + d_2f_2)) \quad (5.12)$$

όπου με  $S_k(f_1, f_2)$  δηλώνεται ο  $2 - D$  μετασχηματισμός Fourier του πλαισίου  $k$  αναφορικά με τις χωρικές συντεταγμένες  $x_1$  και  $x_2$ . Έπεται ότι στην περίπτωση μεταφορικής κίνησης, η διαφορά των Fourier φάσεων των αντίστοιχων τμημάτων,

$$\arg\{S(f_1, f_2, k)\} - \arg\{S(f_1, f_2, k + 1)\} = 2\pi(d_1 f_1 + d_2 f_2)$$

ορίζει ένα χώρο στις μεταβλητές  $(f_1, f_2)$ . Τότε, το διαπλαισιακό διάνυσμα κίνησης μπορεί να εκτιμηθεί από τον προσανατολισμό του επιπέδου (5.13). Αυτή η φαινομενικά άμεση προσέγγιση έχει δύο σοβαρά προβλήματα: α) η εκτίμηση του προσανατολισμού του επιπέδου γενικά απαιτεί  $2 - D$  εκτύλιξη φάσης (phase-unwrapping), η οποία δεν είναι κάτι τετριμμένο, και β) δεν είναι συνήθως εύκολο να αναγνωρίσουμε τα διανύσματα κίνησης για περισσότερα από ένα κινούμενα αντικείμενα μέσα σε ένα τμήμα. Η μέθοδος της ετεροσυσχέτισης λύνει και τα δύο προηγούμενα προβλήματα [Tho 87]. Άλλες μέθοδοι εκτίμησης κίνησης στο πεδίο των συχνοτήτων περιλαμβάνουν αυτές που βασίζονται σε ανάλυση του 3-D πεδίου χωροχρονικών συχνοτήτων με χρήση κατανομών Wigner ή ένα σετ από Gabor φίλτρα [Hee 87].

Η μέθοδος συσχέτισης φάσης εκτιμά τη σχετική μετατόπιση μεταξύ δύο τμημάτων της εικόνας μέσω μιας κανονικοποιημένης συνάρτησης ετεροσυσχέτισης που υπολογίζεται στο διδιάστατο χώρο Fourier. Βασίζεται ακόμα στην αρχή ότι μια σχετική μετατόπιση στο χωρικό πεδίο έχει ως αποτέλεσμα την ύπαρξη ενός γραμμικού όρου φάσης στο πεδίο Fourier. Ακολούθως, πρώτα δείχνουμε πως προκύπτει η συνάρτηση συσχέτισης φάσης και στη συνέχεια αναφερόμαστε σε ζητήματα που σχετίζονται με την υλοποίησή της. Επέκταση της μεθόδου αυτής μπορεί να χρησιμοποιηθεί και για περιστροφική κίνηση.

### Συνάρτηση Συσχέτισης Φάσης

Η συνάρτηση ετεροσυσχέτισης μεταξύ δύο πλαισίων  $k$  και  $k + l$  ορίζεται ως

$$c_{k,k+1}(n_1, n_2) = s(n_1, n_2, k + 1) ** s(-n_1, -n_2, k) \quad (5.13)$$

όπου με  $**$  δηλώνεται ο  $2 - D$  τελεστής συνέλιξης. Παίρνοντας το μετασχηματισμό Fourier και των δύο πλευρών, παίρνουμε τη μιγαδική έκφραση του φάσματος ετεροσυσχέτισης :

$$C_{k,k+1}(f_1, f_2) = S_{k+1}(f_1, f_2) S_k^*(f_1, f_2) \quad (5.14)$$

Κανονικοποιώντας το  $C_{k,k+1}(f_1, f_2)$  με το μέτρο του, παίρνουμε τη φάση του φάσματος ετεροσυσχέτισης

$$\tilde{C}_{k,k+1}(f_1, f_2) = \frac{S_{k+1}(f_1, f_2) S_k^*(f_1, f_2)}{|S_{k+1}(f_1, f_2) S_k^*(f_1, f_2)|} \quad (5.15)$$

Υποθέτοντας κίνηση γραμμικής μετατόπισης αντικαθιστούμε την (5.11) στην (5.7) και παίρνουμε :

$$\tilde{C}_{k,k+1}(f_1, f_2) = \exp(-j2\pi(f_1 d_1 + f_2 d_2)) \quad (5.16)$$

Παίρνοντας τον αντίστροφο 2-D μετασχηματισμό Fourier της έκφρασης, παίρνουμε τη συνάρτηση συσχέτισης φάσης

$$\tilde{c}_{k,k+1}(n_1, n_2) = \delta(n_1 - d_1, n_2 - d_2) \quad (5.17)$$

Παρατηρούμε ότι η συνάρτηση συσχέτισης φάσης αποτελείται από έναν παλμό, του οποίου η θέση παράγει το διάνυσμα μετατόπισης.

## Θέματα Υλοποίησης

Η υλοποίηση της μεθόδου φασικής συσχέτισης στον υπολογιστή απαιτεί την αντικατάσταση του  $2 - D$  μετασχηματισμού Fourier από τον αντίστοιχο διακριτό μετασχηματισμό ( $2 - D$  DFT), με αποτέλεσμα τον ακόλουθο αλγόριθμο:

1. Υπολογίστε το  $2 - D$  DFT των αντίστοιχων τμημάτων από το πλαίσιο  $k$  και το πλαίσιο  $k + 1$
2. Υπολογίστε την φάση του φάσματος διασταυρούμενης ισχύος όπως στην 5.16
3. Υπολογίστε τω αντίστροφο  $2 - D$  DFT της  $\tilde{C}_{k,k+1}(f_1, f_2)$ , για να προκύψει η συνάρτηση συσχέτισης φάσης  $\tilde{c}_{k,k+1}(n_1, n_2)$
4. Ανιχνεύστε τις θέσεις της(ων) κορυφής(ών) στη συνάρτηση συσχέτισης φάσης

Ιδανικά, θα περιμέναμε να παρατηρήσουμε έναν μόνο παλμό στη συνάρτηση φασικής συσχέτισης, ο οποίος θα υποδείκνυε τη σχετική μετατόπιση μεταξύ των δύο τμημάτων. Πρακτικά, ένας αριθμός από παράγοντες συνεισφέρουν στον εκφυλισμό της συνάρτησης συσχέτισης φάσης με αποτέλεσμα να περιέχει μία ή περισσότερες κορυφές. Αυτοί είναι: η χρήση του  $2 - D$  DFT αντί ενός  $2 - D$  μετασχηματισμού Fourier, η παρουσία περισσότερων του ενός κινούμενων αντικειμένων μέσα στο τμήμα και η παρουσία θορύβου παρατήρησης.

Η χρήση του  $2 - D$  DFT αντί του  $2 - D$  μετασχηματισμού Fourier έχει έναν αριθμό από συνέπειες:

- **Φαινόμενα ορίων:** Για να πάρουμε έναν ιδανικό παλμό, η μετατόπιση πρέπει να είναι κυκλική. Καθώς τα πράγματα που εξαφανίζονται από τη μία πλευρά του παραθύρου γενικά δεν εμφανίζονται ξανά στην άλλη άκρη, οι παλμοί υποβαθμίζονται σε τοπικές κορυφές. Επίσης, είναι γνωστό ότι ο  $2 - D$  DFT υποθέτει περιοδικότητα και στις δύο κατευθύνσεις. Ασυνέχειες από τα αριστερά προς τα δεξιά όρια, και από πάνω προς τα κάτω, μπορεί να εισάγουν ψεύτικες κορυφές.
- **Διαρροή φάσματος** εξαιτίας μη ακεραίων διανυσμάτων κίνησης: Για να παρατηρήσουμε έναν τέλειο παλμό, οι συνιστώσες του διανύσματος μετατόπισης πρέπει να αντιστοιχούν σε ακέραια πολλαπλάσια της θεμελιώδους συχνότητας. Διαφορετικά, ο παλμός θα εκφυλισθεί σε μία κορυφή λόγω του γνωστού φαινομένου της διαρροής φάσματος.
- **Περιοχή υποστήριξης των εκτιμητών μετατόπισης:** Καθώς ο  $2 - D$  DFT είναι περιοδικός με μέγεθος τμήματος  $(N_1, N_2)$ , οι εκτιμητές μετατόπισης πρέπει να εκτυλιχθούν ως εξής:

$$\hat{d}_i = \begin{cases} d_i & \text{εάν } |d_i| \leq N_i/2, N_i \text{ άρτιος, ή } |d_i| \leq (N_i - 1)/2, N_i \text{ περιττός} \\ d_i - N_i & \text{αλλιώς} \end{cases} \quad (5.18)$$

για να χειρίζονται και τις αρνητικές μετατοπίσεις. Έτσι, η εμβέλεια των εκτιμητών είναι  $[-N_i/2 + 1, N_i/2]$  για  $N_i$  άρτιους. Για παράδειγμα, για την εκτίμηση μετατοπίσεων μέσα σε ένα εύρος  $[-31, 32]$ , το μέγεθος του πλαισίου πρέπει να είναι τουλάχιστον  $64 \times 64$ .

Το μέγεθος του τμήματος είναι μία από τις πιο σημαντικές παραμέτρους σε ένα αλγόριθμο κατά τμήματα εκτίμησης κίνησης. Η επιλογή του μεγέθους του τμήματος συνήθως περιλαμβάνει μία συναλλαγή δύο συγκρουόμενων απαιτήσεων. Το παράθυρο πρέπει να είναι αρκετά μεγάλο για να είναι ικανό να υπολογίζει μεγάλα διανύσματα μετατόπισης. Από την άλλη πλευρά, πρέπει να είναι αρκετά μικρό ώστε το διάνυσμα μετατόπισης να παραμένει σταθερό μέσα στο παράθυρο. Αυτές οι δύο αντιφατικές συνθήκες μπορούν να διευθετηθούν με τις ιεραρχικές μεθόδους [Etk 93].

Η μέθοδος συσχέτισης φάσης έχει κάποιες επιθυμητές ιδιότητες:

- Αλλαγές στην ένταση από πλαίσιο σε πλαίσιο: Η μέθοδος της συχέτισης φάσης είναι σχετικά μη ευαίσθητη σε αλλαγές στο φωτισμό, επειδή οι μετατοπίσεις στη μέση τιμή ή ο πολλαπλασιασμός με μία σταθερά δεν επηρεάζουν τη φάση του μετασχηματισμού Fourier. Καθώς η συνάρτηση συσχέτισης φάσης είναι κανονικοποιημένη με το μέτρο του μετασχηματισμού Fourier, η μέθοδος δεν είναι ευαίσθητη στις μειώσεις του μέτρου του μετασχηματισμού.
- Πολλαπλά κινούμενα αντικείμενα: Είναι ενδιαφέρον να γνωρίσουμε τι συμβαίνει όταν πολλαπλά κινούμενα αντικείμενα με διαφορετικές ταχύτητες βρίσκονται στο ίδιο παράθυρο. Πειράματα έχουν δείξει ότι παρατηρούνται πολλές κορυφές, με κάθε μία από αυτές να υποδεικνύει την κίνηση ενός συγκεκριμένου αντικειμένου [Tho 87]. Η ανίχνευση των σημαντικών κορυφών παράγει έπειτα μία λίστα από υποψήφια διανύσματα μετατόπισης για κάθε εικονοστοιχείο μέσα στο τμήμα. Μία επιπλέον αναζήτηση απαιτείται για να βρεθεί ποιο διάνυσμα μετατόπισης ανήκει σε ποιο εικονοστοιχείο μέσα στο τμήμα. Αυτό μπορεί να επαληθευθεί δοκιμάζοντας το μέτρο της διαφοράς του μετατοπισμένου πλαισίου με καθένα από τα υποψήφια διανύσματα.

### 5.2.3 Μέθοδοι ταιριάσματος τμημάτων

Με τις μεθόδους των τμημάτων, υποθέτουμε ότι η μετατόπιση των αντικειμένων είναι σταθερή μέσα σε ένα μικρό  $2 - D$  τμήμα  $B$  της εικόνας. Αυτή η υπόθεση παρουσιάζει δυσκολίες στις σκηνές με πολλαπλά κινούμενα αντικείμενα ή σε σκηνές όπου διαφορετικά κομμάτια του ίδιου αντικειμένου κινούνται σε διαφορετικές κατευθύνσεις. Εάν το μέγεθος του τμήματος μειωθεί, τότε αυτή η υπόθεση γίνεται πιο έγκυρη. Παρά ταύτα, το τμήμα της υπολογισιμότητας και της μετάδοσης της πληροφορίας της μετατόπισης αυξάνεται.

Η μετατόπιση  $\bar{d}$  μπορεί να εκτιμηθεί με συσχέτιση ή με τεχνικές ταιριάσματος. Έτσι, η μετατόπιση  $\bar{d} = (i, j)$  μπορεί να επιλεγεί ώστε να ελαχιστοποιεί κάποια μετρική του λάθους πρόβλεψης:

$$PE(\bar{d}) = \sum (N(s(n_1, n_2, t) - s(n_1 - i, n_2 - j, t - \tau))) \quad (5.19)$$

όπου  $N()$  είναι μία μετρική (νόρμα), όπως το μέτρο  $L_1$  ή η τετραγωνική νόρμα  $L_2$ . Έχοντας βρει το καλύτερο διάνυσμα, όλα τα εικονοστοιχεία  $z \in B$  προβλέπονται, κωδικοποιούνται και μεταδίδονται χρησιμοποιώντας αυτή την τιμή του  $\bar{d}$ .

Ως παράδειγμα μπορούμε να θεωρήσουμε ένα τμήμα  $M \times N$  εικονοστοιχείων γύρω από ένα εικονοστοιχείο  $x_0$  στο παρόν πλαίσιο τη στιγμή  $t$ . Υποθέτοντας μέγιστες οριζόντιες και κάθετες μετατοπίσεις των  $d_{max}$  εικονοστοιχείων, η περιοχή ψαξίματος του ελαχίστου της  $PE(\bar{d})$  δίνεται από τη σχέση :

$$x_0 \pm \left( \frac{\frac{1}{2}(M-1) + d_{max}}{\frac{1}{2}(N-1) + d_{max}} \right)$$

Ο αριθμός των εικονοστοιχείων σ' αυτή τη περιοχή δίνεται από τη σχέση :

$$(M + 2d_{max})(N + 2d_{max}) \quad (5.20)$$

Έτσι εάν το μέγεθος του τμήματος είναι  $9 \times 9$  και χρησιμοποιείται η μέγιστη μετατόπιση του  $d_{max} = 10$ , η περιοχή αναζήτησης στο προηγούμενο πλαίσιο θα περιέχει  $29 \times 29$  εικονοστοιχεία. Μία αφελής μέθοδος αναζήτησης θα υπολόγιζε την εξίσωση (5.19) για κάθε μετατόπιση των εικονοστοιχείων στην οριζόντια και στην κάθετη κατεύθυνση και θα διάλεγε την ελάχιστη. Αυτό θα απαιτούσε  $(2d_{max} + 1)^2$  υπολογισμούς της εξίσωσης (5.19). Επιπλέον, αν δεν περιλαμβάνονται μετατοπίσεις κλασματικών αποστάσεων των εικονοστοιχείων, η ακρίβεια με την οποία παίρνουμε το  $D$  περιορίζεται στο ένα εικονοστοιχείο. Αρκετές απλοποιήσεις έχουν προταθεί στη βιβλιογραφία. Ένας τύπος αυτών των απλοποιήσεων είναι η χρήση μιας υπολογιστικά απλής νόρμας στην (5.19):

$$PE(x_0, i, j) = \frac{1}{MN} \sum_{|m| \leq \frac{M}{2}, |n| \leq \frac{N}{2}} \sum |b(x_{mn}, t) - b(x_{m+i, n+j}, t - \tau)|^2 \quad (5.21)$$

$$PE(x_0, i, j) = \frac{1}{MN} \sum_{|m| \leq \frac{M}{2}, |n| \leq \frac{N}{2}} \sum |b(x_{mn}, t) - b(x_{m+i, n+j}, t - \tau)| \quad (5.22)$$

όπου

$$\begin{aligned} -d_{max} &\leq i, j \leq +d_{max} \\ x_{mn} &= x_0 + [m, n]'. \end{aligned} \quad (5.23)$$

Ο δεύτερος ορισμός του  $PE(\bar{d})$  στην εξίσωση (5.22) έχει το πλεονέκτημα ότι δεν απαιτεί καθόλου πολλαπλασιασμούς. Επίσης, έχει παρατηρηθεί πειραματικά ότι ο ακριβής ορισμός του  $PE$  δεν έχει σημαντικές συνέπειες στη ποσότητα της αναζήτησης ή στην ακρίβεια του υπολογισμού του  $D$  και, έτσι λόγω απλότητας, το κριτήριο (5.22) γενικά προτιμάται.

Πολλές μέθοδοι έχουν μελετηθεί για την απλοποίηση της διαδικασίας της αναζήτησης. Εμείς περιγράφουμε τρεις από αυτές: (α) Διδιάστατη λογαριθμική αναζήτηση, (β) αναζήτηση τριών βημάτων, και (γ) τροποποιημένη κατεύθυνση αναζήτησης. Σε όλα αυτά, ο στόχος είναι να απαιτούνται όσο το δυνατόν λιγότερες μετατοπίσεις και έτσι να υπολογίσουμε την  $PE$  όσο το δυνατόν λιγότερες φορές (κάθε μετατόπιση απαιτεί έναν υπολογισμό της  $PE$ ). Κατά τη μείωση του αριθμού των μετατοπίσεων η υπόθεση που συνήθως γίνεται είναι ότι η  $PE(x_0, i, j)$  αυξάνεται μονότονα καθώς η μετατόπιση  $(i, j)$  απομακρύνεται από τη θέση του ελαχίστου.

### Μέθοδος της Διδιάστατης Λογαριθμικής Αναζήτησης

Στη διδιάστατη λογαριθμική αναζήτηση, ο αλγόριθμος ακολουθεί την κατεύθυνση της ελάχιστης παραμόρφωσης. Σε κάθε βήμα, πέντε μετατοπίσεις ελέγχονται όπως φαίνεται στο σχήμα 5.10. Η απόσταση μεταξύ των σημείων όπου γίνεται η αναζήτηση μειώνεται εάν το ελάχιστο βρίσκεται είτε στο κέντρο της περιοχής αναζήτησης, ή στα όρια αυτής της περιοχής. Για το παράδειγμα που φαίνεται στο σχήμα 5.10 πέντε βήματα απαιτούνται για να πάρουμε το διάνυσμα μετατόπισης στο σημείο  $(i, j) = (2, 6)$ .

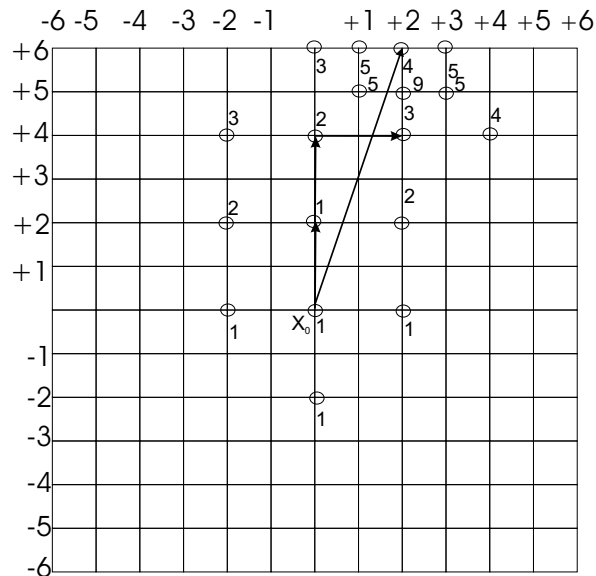
### Μέθοδος της Αναζήτησης Τριών Βημάτων

Στη μέθοδο της αναζήτησης τριών βημάτων, οχτώ τυχαία τοποθετημένα εικονοστοιχεία γύρω από το εικονοστοιχείο στο  $x_0$  ελέγχονται στο πρώτο βήμα. Στο δεύτερο βήμα, πάλι χρησιμοποιούνται οχτώ εικονοστοιχεία γύρω από το εικονοστοιχείο της ελάχιστης  $PE$  η οποία βρίσκεται στο πρώτο βήμα, αλλά σε αυτό το βήμα η διάταξη αυτών είναι καλύτερη από το πρώτο βήμα. Το σχήμα 5.11 δείχνει τα σημεία  $x_0 + (3, 3)'$  και  $x_0 + (3, 5)'$  ως πρώτες και δεύτερες προσεγγίσεις αντίστοιχα. Αυτή η διαδικασία συνεχίζεται για ακόμη ένα βήμα, με αποτέλεσμα το τελικό διάνυσμα μετατόπισης στο εικονοστοιχείο  $x_0 + (2, 6)'$ .

### Προσαρμοσμένη Κατεύθυνση Αναζήτησης

Η προσαρμοσμένη κατεύθυνση αναζήτησης, στην απλή μορφή, αναζητά την κατεύθυνση της ελάχιστης  $PE(x_0, i, j)$ . Είναι μία μέθοδος δύο βημάτων. Στο πρώτο βήμα, μία αναζήτηση γίνεται κατά μήκος της οριζόντιας κατεύθυνσης. Πρώτα, υπολογίζεται το ελάχιστο μεταξύ των  $PE(x_0, -1, 0)$ ,  $PE(x_0, 0, 0)$  και  $PE(x_0, 1, 0)$ . Εάν το  $PE(x_0, 1, 0)$  είναι το ελάχιστο, τότε το  $PE(x_0, 2, 0)$  υπολογίζεται επίσης, και συγκρίνεται. Ακολουθώντας αυτόν τον τρόπο, ο αλγόριθμος προσδιορίζει το ελάχιστο στην οριζόντια κατεύθυνση, όταν βρίσκεται μεταξύ των δύο υψηλότερων τιμών για τα





Σχήμα 5.10: Αναπαράσταση της διδιάστατης λογαριθμικής αναζήτησης. Οι μετακινήσεις στη περιοχή αναζήτησης του προηγούμενου πλαισίου φαίνονται σε σχέση με το εικονοστοιχείο  $x_0$  στο παρόν πλαίσιο. Εδώ τα διανύσματα μετατόπισης  $(0, 2)'$ ,  $(0, 4)'$ ,  $(2, 4)'$ ,  $(2, 5)'$ ,  $(2, 6)'$  βρίσκονται στα βήματα 1, 2, 3, 4 και 5. Το  $d_{max}$  είναι 6 εικονοστοιχεία.

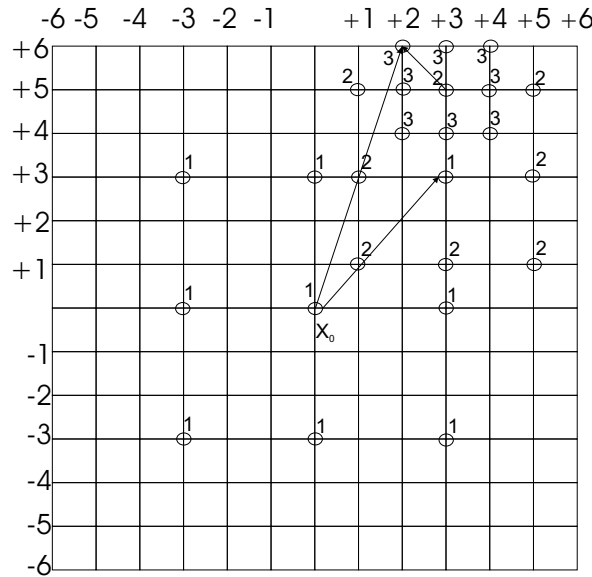
γειτονικά εικονοστοιχεία. Στο δεύτερο βήμα, χρησιμοποιείται η ίδια τεχνική για να υπολογιστεί το ελάχιστο στην κατακόρυφη κατεύθυνση. Το σχήμα 5.12 δείχνει ένα παράδειγμα της μεθόδου προσαρμοσμένης κατεύθυνσης αναζήτησης στην οποία το αποτέλεσμα των δύο βημάτων είναι τα εικονοστοιχεία  $x_0 + (2, 0)'$  και  $x_0 + (2, 6)'$  αντίστοιχα.

### Θέματα Υλοποίησης

Είναι φανερό ότι και οι τρεις μέθοδοι αναζήτησης βασίζονται σε κάποιους ευριστικούς αλγορίθμους και αντιμετωπίζουν δυσκολίες σε πολλές περιπτώσεις. Όπως προαναφέρθηκε, η υπολογιστική πολυπλοκότητα της μεθόδου αναζήτησης σχετίζεται με τον αριθμό των μετατοπίσεων για τις οποίες η  $PE$  πρέπει να υπολογισθεί. Ο πίνακας 5.1 δείχνει μία σύγκριση του αριθμού των μετατοπίσεων αναζήτησης και του αριθμού των διαδοχικών βημάτων που απαιτούνται για κάθε μία από της παραπάνω μεθόδους αναζήτησης. Πρέπει να σημειωθεί ότι η ακρίβεια της εκτίμησης μετατόπισης της μεθόδου των τμημάτων γενικά περιορίζεται στα 0.5 εικονοστοιχεία. Θα μπορούσε να αυξηθεί με το περιλάβει παρεμβολόμενες τιμές των εικονοστοιχείων και έτσι να φαίνεται καλύτερα το πλέγμα στα σχήματα 5.10-5.12. Ωστόσο, η υπολογιστική πολυπλοκότητα είναι συνήθως απαγορευτική.

## 5.3 Μέθοδοι που χρησιμοποιούν την Εξίσωση Οπτικής Ροής

Σε αυτό το υποκεφάλαιο θα μελετήσουμε αρχικά την εξίσωση οπτικής ροής (EOP). Στη συνέχεια, θα συζητηθούν οι μέθοδοι εκτίμησης οπτικής ροής που χρησιμοποιούν την EOP.



Σχήμα 5.11: Αναπαράσταση της διαδικασίας των τριών βημάτων. Εδώ,  $(3, 3)'$ ,  $(3, 5)'$ ,  $(2, 6)'$  είναι τα διανύσματα μετατόπισης που βρέθηκαν στο 1, 2 και 3. Το  $d_{max}$  είναι 6 εικονοστοιχεία.

Πίνακας 5.1: Υπολογιστική πολυπλοκότητα αλγορίθμων ταιριάσματος τμημάτων

Διαδικασία αναζήτησης	Απαιτούμενος αριθμός σημείων αναζήτησης		Απαιτούμενος αριθμός σειριακών βημάτων	
	$\alpha$	$\beta$	$\alpha$	$\beta$
Λογαριθμική αναζήτηση	18	21	5	7
Αναζήτηση τριών βημάτων	25	25	3	3
Προσαρμοσμένη κατεύθυνση αναζήτησης	12	15	9	12

- α) Για ένα χωρικό διάνυσμα μετατόπισης  $(2, -6)'$   
 β) Για την χειρότερη περίπτωση

### 5.3.1 Η Εξίσωση Οπτικής Ροής

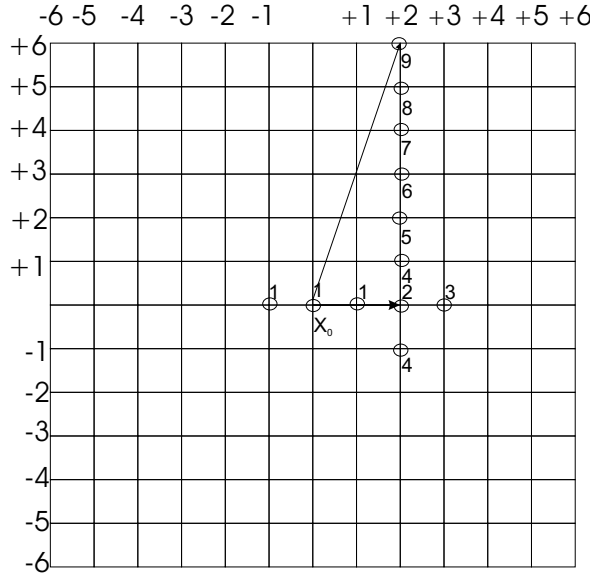
Ας υποθέσουμε ότι  $s_c(x_1, x_2, t)$  είναι η συνεχής χωρο-χρονική κατανομή έντασης. Αν η ένταση παραμένει σταθερή κατά μήκος της τροχιάς της κίνησης, θα έχουμε :

$$\frac{ds_c(x_1, x_2, t)}{dt} = 0 \quad (5.24)$$

όπου τα  $x_1, x_2$  μεταβάλλονται με το  $t$  σύμφωνα με την τροχιά της κίνησης. Η διαφορίση στην εξίσωση (5.24) είναι ολική και δηλώνει το ρυθμό με τον οποίο μεταβάλλεται η ένταση κατά μήκος της τροχιάς της κίνησης. Χρησιμοποιώντας τον κανόνα της αλυσίδας, μπορούμε να γράψουμε:

$$\frac{\partial s_c(\mathbf{x}; t)}{\partial x_1} v_1(\mathbf{x}, t) + \frac{\partial s_c(\mathbf{x}; t)}{\partial x_2} v_2(\mathbf{x}, t) + \frac{\partial s_c(\mathbf{x}; t)}{\partial t} = 0 \quad (5.25)$$

όπου  $v_1(\mathbf{x}, t) = dx_1/dt$ ,  $v_2(\mathbf{x}, t) = dx_2/dt$  δηλώνουν τις συνιστώσες του διανύσματος ταχύτητας των συνεταγμένων σε σχέση με τις συνεχείς χωρικές συντεταγμένες. Η έκφραση (5.25) είναι γνωστή



Σχήμα 5.12: Αναπαράσταση ενός απλοποιημένου μοντέλου προσαρμοσμένης αναζήτησης. Εδώ το διάνυσμα μετατόπισης είναι το  $(2, 6)'$  και βρίσκεται στο βήμα 9. Το  $d_{max}$  είναι 6 εικονοστοιχεία.

ως η εξίσωση της οπτικής ροής ή σαν περιορισμός της οπτικής ροής.

Διαφορετικά η εξίσωση (5.25) μπορεί να γραφεί ως εξής:

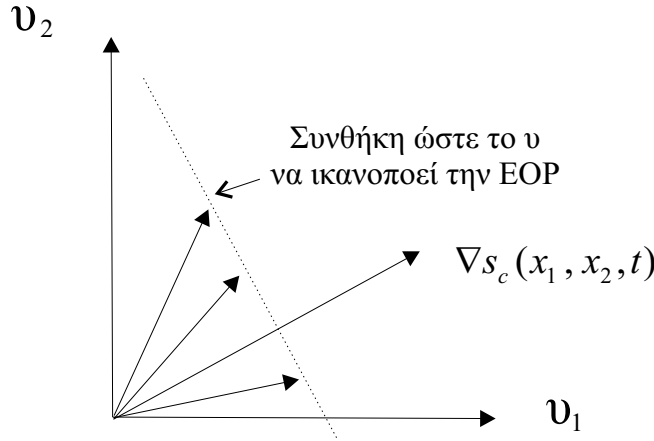
$$\langle \nabla s_c(\mathbf{x}; t), \mathbf{v}(\mathbf{x}, t) \rangle + \frac{\partial s_c(\mathbf{x}; t)}{\partial t} = 0 \quad (5.26)$$

όπου  $\nabla s_c(\mathbf{x}; t) = [\frac{\partial s_c(\mathbf{x}; t)}{\partial x_1} \frac{\partial s_c(\mathbf{x}; t)}{\partial x_2}]^T$  και η έκφραση  $\langle \cdot, \cdot \rangle$  δηλώνει το εσωτερικό γινόμενο διανυσμάτων.

Στην πράξη, η εξίσωση EOP (5.25) δεν είναι αρκετή για το μοναδικό προσδιορισμό του πεδίου της διδιάστατης ταχύτητας (ροής), ακριβώς όπως συμβαίνει και με την εξίσωση (5.3). Η EOP παράγει μια βαθμωτή εξίσωση με δυο αγνώστους,  $v_1(\mathbf{x}, t)$  και  $v_2(\mathbf{x}, t)$  για κάθε  $(\mathbf{x}, t)$ . Πιο λεπτομερής μελέτη της (5.26) αποκαλύπτει ότι μπορούμε να εκτιμήσουμε μόνο τη συνιστώσα του διανύσματος της ροής το οποίο βρίσκεται επί της διεύθυνσης της χωρικής κλίσης της εικόνας  $\frac{\nabla s_c(\mathbf{X}; t)}{\|\nabla s_c(\mathbf{X}; t)\|}$ , η οποία ονομάζεται κανονική ροή  $v_\perp(\mathbf{x}, t)$ , και αυτό γιατί η συνιστώσα η οποία είναι ορθογώνια στην κλίση αυτή εξαφανίζεται με την εφαρμογή του εσωτερικού γινόμενου. Αυτό φαίνεται στην εικόνα 5.13, όπου όλα τα διανύσματα των οποίων οι κορυφές βρίσκονται επί της διακεκομμένης γραμμής ικανοποιούν την εξίσωση (5.26). Η κανονική ροή μπορεί να υπολογιστεί από την (5.26) ως εξής:

$$v_\perp(\mathbf{x}, t) = \frac{-\frac{\partial s_c(\mathbf{X}; t)}{\partial t}}{\|\nabla s_c(\mathbf{x}; t)\|} \quad (5.27)$$

Έτσι, η EOP (5.25) θέτει έναν περιορισμό στη συνιστώσα του διανύσματος της ροής το οποίο είναι στη διεύθυνση της χωρικής κλίσης της έντασης της εικόνας σε κάθε σημείο (εικονοστοιχείο), και ο οποίος είναι σύμφωνος με το πρόβλημα του διαφράγματος. Παρατηρήστε ότι η EOP προσέγγιση απαιτεί ότι πρώτα, η χωρο-χρονική ένταση της εικόνας είναι διαφορίσιμη, και δεύτερον, ότι υπάρχουν οι μερικές παράγωγοι της έντασης. Παρακάτω, παρουσιάζουμε αρκετές προσεγγίσεις εκτίμησης της οπτικής ροής από εκτιμήσεις της κανονικής ροής.



Σχήμα 5.13: Υπολογισμός της οπτικής ροής

### 5.3.2 Μέθοδοι Διαφορικών Δευτέρας Τάξης

Στην αναζήτηση ενός άλλου περιορισμού ο οποίος θα περικλείει και τις δυο συνιστώσες του διάνυσματος της ροής σε κάθε εικονοστοιχείο, αρκετοί ερευνητές [Nag 87, Ura 88] πρότειναν μεταξύ άλλων τη διατήρηση της χωρικής κλίσης της εικόνας  $\|\nabla s_c(\mathbf{x}; t)\|$ , η οποία δίνεται από την εξίσωση

$$\frac{d\nabla s_c(\mathbf{x}; t)}{dt} = 0 \quad (5.28)$$

Μια εκτίμηση του πεδίου της ροής, που παίρνεται από την (5.28), δίνεται από την εξίσωση (5.29)

$$\begin{bmatrix} \hat{v}_1(\mathbf{x}; t) \\ \hat{v}_2(\mathbf{x}; t) \end{bmatrix} = \begin{bmatrix} \frac{\partial^2 s_c(\mathbf{X}, t)}{\partial x_1^2} & \frac{\partial^2 s_c(\mathbf{X}, t)}{\partial x_2 \partial x_1} \\ \frac{\partial^2 s_c(\mathbf{X}, t)}{\partial x_1 \partial x_2} & \frac{\partial^2 s_c(\mathbf{X}, t)}{\partial x_2^2} \end{bmatrix}^{-1} \begin{bmatrix} -\frac{\partial^2 s_c(\mathbf{X}, t)}{\partial t \partial x_1} \\ -\frac{\partial^2 s_c(\mathbf{X}, t)}{\partial t \partial x_2} \end{bmatrix} \quad (5.29)$$

Παρ' όλα αυτά, ο περιορισμός (5.28) δεν μπορεί να εφαρμοστεί σε ορισμένες συνήθεις κινήσεις όπως είναι η περιστροφή και το ζουμ. Επιπλέον, οι μερικές παράγωγοι δευτέρας τάξεως δεν μπορούν πάντοτε να υπολογιστούν με απόλυτη ακρίβεια. Αυτό το πρόβλημα αναλύεται στην παράγραφο 5.1.3. Ως αποτέλεσμα, η εξίσωση (5.29) δεν παράγει πάντοτε αξιόπιστες εκτιμήσεις ροής.

### 5.3.3 Το Μοντέλο της Κίνησης Κατά Τμήματα

Μια άλλη προσέγγιση για να υπερκεράσουμε το πρόβλημα του διαφράγματος είναι να κάνουμε την υπόθεση ότι το διάνυσμα της κίνησης είναι σταθερό σε ένα συγκεκριμένο τμήμα της εικόνας, το οποίο τμήμα συνήθως αντιστοιχεί σε ένα αντικείμενο και δηλώνεται με το γράμμα  $\mathcal{B}$  [Luc 81]. Σε αυτήν την περίπτωση έχουμε :

$$\mathbf{v}(\mathbf{x}, t) = \mathbf{v}(t) = [v_1(t) \ v_2(t)]^T, \text{ for } \mathbf{x} \in \mathcal{B} \quad (5.30)$$

Αν και ένα τέτοιο μοντέλο δεν μπορεί να χειριστεί την περιστροφική κίνηση, είναι δυνατό να εκτιμήσουμε μοναδικά ένα διάνυσμα μετατόπισης με την προϋπόθεση ότι το τμήμα των εικονοστοιχείων περιέχει επαρκή διαβάθμιση των τόνων του γκρι.

Ας ορίσουμε, τώρα, το σφάλμα που παράγει η εξίσωση οπτικής ροής για το τμήμα της εικόνας  $\mathcal{B}$  :

$$E = \sum_{\mathbf{x} \in \mathcal{B}} \left( \frac{\partial s_c(\mathbf{x}, t)}{\partial x_1} v_1(t) + \frac{\partial s_c(\mathbf{x}, t)}{\partial x_2} v_2(t) + \frac{\partial s_c(\mathbf{x}, t)}{\partial t} \right)^2 \quad (5.31)$$

Το λάθος  $E$  πρέπει να ελαχιστοποιείται για κάποιο  $\hat{v}_1(t)$  και  $\hat{v}_2(t)$ . Υπολογίζοντας τις μερικές παραγώγους του σφάλματος  $E$  σε σχέση με τα  $v_1(t)$  και  $v_2(t)$ , αντίστοιχα, και θέτοντάς αυτές ίσες με μηδέν, θα έχουμε :

$$\begin{aligned} \sum_{\mathbf{x} \in \mathcal{B}} \left( \frac{\partial s_c(\mathbf{x}, t)}{\partial x_1} \hat{v}_1(t) + \frac{\partial s_c(\mathbf{x}, t)}{\partial x_2} \hat{v}_2(t) + \frac{\partial s_c(\mathbf{x}, t)}{\partial t} \right) \frac{\partial s_c(\mathbf{x}, t)}{\partial x_1} &= 0 \\ \sum_{\mathbf{x} \in \mathcal{B}} \left( \frac{\partial s_c(\mathbf{x}, t)}{\partial x_1} \hat{v}_1(t) + \frac{\partial s_c(\mathbf{x}, t)}{\partial x_2} \hat{v}_2(t) + \frac{\partial s_c(\mathbf{x}, t)}{\partial t} \right) \frac{\partial s_c(\mathbf{x}, t)}{\partial x_2} &= 0 \end{aligned}$$

όπου το  $\hat{\cdot}$  δηλώνει την εκτίμηση της αντίστοιχης ποσότητας. Επιλύοντας αυτές τις εξισώσεις ταυτόχρονα, προκύπτει ότι

$$\begin{bmatrix} \hat{v}_1(t) \\ \hat{v}_2(t) \end{bmatrix} = \begin{bmatrix} \sum_{\mathbf{x} \in \mathcal{B}} \frac{\partial s_c(\mathbf{x}, t)}{\partial x_1} \frac{\partial s_c(\mathbf{x}, t)}{\partial x_1} & \sum_{\mathbf{x} \in \mathcal{B}} \frac{\partial s_c(\mathbf{x}, t)}{\partial x_1} \frac{\partial s_c(\mathbf{x}, t)}{\partial x_2} \\ \sum_{\mathbf{x} \in \mathcal{B}} \frac{\partial s_c(\mathbf{x}, t)}{\partial x_2} \frac{\partial s_c(\mathbf{x}, t)}{\partial x_1} & \sum_{\mathbf{x} \in \mathcal{B}} \frac{\partial s_c(\mathbf{x}, t)}{\partial x_2} \frac{\partial s_c(\mathbf{x}, t)}{\partial x_2} \end{bmatrix}^{-1} \begin{bmatrix} - \sum_{\mathbf{x} \in \mathcal{B}} \frac{\partial s_c(\mathbf{x}, t)}{\partial x_1} \frac{\partial s_c(\mathbf{x}, t)}{\partial t} \\ - \sum_{\mathbf{x} \in \mathcal{B}} \frac{\partial s_c(\mathbf{x}, t)}{\partial x_2} \frac{\partial s_c(\mathbf{x}, t)}{\partial t} \end{bmatrix} \quad (5.32)$$

Είναι δυνατό να αυξήσουμε την επίδραση των περιορισμών προς το κέντρο του τμήματος  $\mathcal{B}$  αντικαθιστώντας όλα τα αθροίσματα με σταθμισμένα αθροίσματα. Μια κατάλληλη συνάρτηση σταθμίσεως μπορεί να είναι της μορφής ενός διδιάστατου τριγωνικού παραθύρου. Η ακρίβεια των εκτιμήσεων της ροής εξαρτάται από την ακρίβεια των εκτιμώμενων χωρικών και χρονικών μερικών παραγώγων (δες παράγραφο ;;). Στη συνέχεια θα συζητήσουμε τη μέθοδο των Horn και Schunck [Hor 81], η οποία θέτει ένα λιγότερα αυστηρό καθολικό περιορισμό εξομάλυνσης στο πεδίο της ταχύτητας.

### 5.3.4 Η Μέθοδος Horn-Schunck

Οι Horn και Schunck αναζητούν ένα πεδίο κίνησης που ικανοποιεί την ΕΟΡ με τη μικρότερη μεταβολή των διανυσμάτων ροής από εικονοστοιχείο σε εικονοστοιχείο. Ας υποθέσουμε ότι η εξίσωση

$$\mathcal{E}_{of}(\mathbf{v}(\mathbf{x}, t)) = \langle \nabla s_c(\mathbf{x}, t), \mathbf{v}(\mathbf{x}, t) \rangle + \frac{\partial s_c(\mathbf{x}, t)}{\partial t} \quad (5.33)$$

είναι το σφάλμα της εξίσωσης οπτικής ροής. Η ΕΟΡ ικανοποιείται όταν το σφάλμα  $\mathcal{E}_{of}(\mathbf{v}(\mathbf{x}, t))$  είναι ίσο με μηδέν. Με την παρουσία κάλυψης και θορύβου προσδοκούμε να ελαχιστοποιήσουμε το τετράγωνο του σφάλματος  $\mathcal{E}_{of}(\mathbf{v}(\mathbf{x}, t))$  ώστε να επιβληθεί ο περιορισμός της οπτικής ροής.

Η μεταβολή των διανυσμάτων ταχύτητας από εικονοστοιχείο σε εικονοστοιχείο μπορεί να ποσοτικοποιηθεί από το άθροισμα των τετραγώνων του μεγέθους των χωρικών κλίσεων των συνιστωσών του διανύσματος της ταχύτητας, και δίνεται από την εξίσωση:

$$\mathcal{E}_s(\mathbf{v}(\mathbf{x}, t)) = \|\nabla v_1(\mathbf{x}, t)\|^2 + \|\nabla v_2(\mathbf{x}, t)\|^2 = \left( \frac{\partial v_1}{\partial x_1} \right)^2 + \left( \frac{\partial v_1}{\partial x_2} \right)^2 + \left( \frac{\partial v_2}{\partial x_1} \right)^2 + \left( \frac{\partial v_2}{\partial x_2} \right)^2 \quad (5.34)$$

όπου υποθέτουμε ότι οι τρισδιάστατες χρονικές συντεταγμένες είναι συνεχείς μεταβλητές. Μπορεί εύκολα να διαπιστωθεί ότι όσο πιο ομαλό είναι το πεδίο της ταχύτητας, τόσο μικρότερο είναι το  $\mathcal{E}_s(\mathbf{v}(\mathbf{x}, t))$ .

Κατόπιν, η μέθοδος Horn και Schunck ελαχιστοποιεί ένα σταθμισμένο άθροισμα του σφάλματος της ΕΟΡ  $\mathcal{E}_{of}$  και του μέτρου της μεταβολής της ταχύτητας από εικονοστοιχείο σε εικονοστοιχείο  $\mathcal{E}_s$

$$\min_{\mathbf{v}(\mathbf{x}, t)} \int_{\mathcal{A}} (\mathcal{E}_{of}^2(\mathbf{v}) + \alpha^2 \mathcal{E}_s^2(\mathbf{v})) d\mathbf{x} \quad (5.35)$$

για να εκτιμήσει το διάνυσμα της ταχύτητας σε κάθε σημείο  $\mathbf{x}$ , όπου το  $\mathcal{A}$  δηλώνει περιοχή υποστήριξης συνεχούς εικόνας. Η παράμετρος  $\alpha^2$ , η οποία συνήθως επιλέγεται ευριστικά, ελέγχει την ισχύ του περιορισμού της εξομάλυνσης. Μεγαλύτερες τιμές του  $\alpha^2$  αυξάνουν την επίδραση που έχει ο περιορισμός.

Ελαχιστοποίηση της (5.35), χρησιμοποιώντας το λογισμό των μεταβολών, απαιτεί την ταυτόχρονη επίλυση δυο εξισώσεων

$$\begin{aligned} \left(\frac{\partial s_c}{\partial x_1}\right)^2 \hat{v}_1(\mathbf{x}, t) + \frac{\partial s_c}{\partial x_1} \frac{\partial s_c}{\partial x_2} \hat{v}_2(\mathbf{x}, t) &= \alpha^2 \nabla^2 \hat{v}_1(\mathbf{x}, t) - \frac{\partial s_c}{\partial x_1} \frac{\partial s_c}{\partial t} \\ \frac{\partial s_c}{\partial x_1} \frac{\partial s_c}{\partial x_2} \hat{v}_1(\mathbf{x}, t) + \left(\frac{\partial s_c}{\partial x_2}\right)^2 \hat{v}_2(\mathbf{x}, t) &= \alpha^2 \nabla^2 \hat{v}_2(\mathbf{x}, t) - \frac{\partial s_c}{\partial x_2} \frac{\partial s_c}{\partial t} \end{aligned} \quad (5.36)$$

όπου το  $\nabla^2$  δηλώνει τη Λαπλασιανή και το  $\hat{\cdot}$  δηλώνει την εκτιμώμενη αντίστοιχη ποσότητα. Στην υλοποίηση των Horn και Schunck [Hor 81], οι Λαπλασιανές συνιστώσες της ταχύτητας έχουν προσεγγισθεί από υψηλοπερατά FIR φίλτρα οδηγούν σε επαναληπτική μέθοδο Gauss-Seidel :

$$\begin{aligned} \hat{v}_1^{(n+1)}(\mathbf{x}, t) &= \bar{v}_1^{(n)}(\mathbf{x}, t) - \frac{\frac{\partial s_c}{\partial x_1} \bar{v}_1^{(n)}(\mathbf{x}, t) + \frac{\partial s_c}{\partial x_2} \bar{v}_2^{(n)}(\mathbf{x}, t) + \frac{\partial s_c}{\partial t}}{\alpha^2 + \left(\frac{\partial s_c}{\partial x_1}\right)^2 + \left(\frac{\partial s_c}{\partial x_2}\right)^2} \\ \hat{v}_2^{(n+1)}(\mathbf{x}, t) &= \bar{v}_2^{(n)}(\mathbf{x}, t) - \frac{\frac{\partial s_c}{\partial x_1} \bar{v}_1^{(n)}(\mathbf{x}, t) + \frac{\partial s_c}{\partial x_2} \bar{v}_2^{(n)}(\mathbf{x}, t) + \frac{\partial s_c}{\partial t}}{\alpha^2 + \left(\frac{\partial s_c}{\partial x_1}\right)^2 + \left(\frac{\partial s_c}{\partial x_2}\right)^2} \end{aligned} \quad (5.37)$$

όπου  $n$  είναι ο απαριθμητής των επαναλήψεων, τα  $\bar{v}_i$  δηλώνουν σταθμισμένους τοπικούς μέσους όρους (εξαιρώντας το παρόν εικονοστοιχείο) και όλες οι μερικές παράγωγοι υπολογίζονται στο σημείο  $(\mathbf{x}, t)$ . Οι αρχικές εκτιμήσεις των ταχυτήτων  $v_1^{(0)}(\mathbf{x}, t)$  και  $v_2^{(0)}(\mathbf{x}, t)$  λαμβάνονται συνήθως μηδέν. Στον παραπάνω τύπο γίνεται η υπόθεση συνεχούς χωρο-χρονικής κατανομής της έντασης. Στην υλοποίηση σε H/Y, όλες οι χωρικές και χρονικές κλίσεις της εικόνας πρέπει να εκτιμηθούν αριθμητικώς από τα παρατηρούμενα δείγματα της εικόνας, ένα ζήτημα που θα συζητηθεί στη συνέχεια.

### 5.3.5 Εκτίμηση των Κλίσεων

Θα συζητήσουμε δυο μεθόδους εκτίμησης των χωροχρονικών κλίσεων της εικονοσειράς που απαιτούνται για την εφαρμογή των μεθόδων εύρεσης κίνησης που στηρίζονται στην EOP. Η πρώτη κάνει χρήση πεπερασμένων διαφορών, ενώ η δεύτερη βασίζεται σε πολυωνυμική προσαρμογή.

#### Εκτίμηση κλίσης με τη χρήση πεπερασμένων διαφορών

Μια προσέγγιση στην εκτίμηση των μερικών παραγώγων από μια διακριτή εικόνα  $s(n_1, n_2, k)$  είναι να τις προσεγγίσουμε από τις αντίστοιχες προς τα μπρος ή προς τα πίσω πεπερασμένες διαφορές. Για να πετύχουμε περισσότερο αξιόπιστες εκτιμήσεις των μερικών παραγώγων, μπορούμε να υπολογίσουμε τις μέσες προς τα μπρος και προς τα πίσω πεπερασμένες διαφορές, κάτι που ονομάζεται η μέση διαφορά. Επιπλέον, μπορούμε να υπολογίσουμε μια τοπική μέση τιμή των μέσων διαφορών για την εξάλειψη των επιδράσεων του θορύβου παρατήρησης. Οι Horn και Schunck [Hor 81] πρότειναν την εύρεση μέσου όρου για τέσσερις μέσες πεπερασμένες διαφορές

$$\begin{aligned} \frac{\partial s_c(x_1, x_2, t)}{\partial x_1} &\approx \frac{1}{4} \{s(n_1 + 1, n_2, k) - s(n_1, n_2, k) + s(n_1 + 1, n_2 + 1, k) \\ &\quad - s(n_1, n_2 + 1, k) + s(n_1 + 1, n_2, k + 1) - s(n_1, n_2, k + 1) \\ &\quad + s(n_1 + 1, n_2 + 1, k + 1) - s(n_1, n_2 + 1, k + 1)\} \\ \frac{\partial s_c(x_1, x_2, t)}{\partial x_2} &\approx \frac{1}{4} \{s(n_1, n_2 + 1, k) - s(n_1, n_2, k) + s(n_1 + 1, n_2 + 1, k) \\ &\quad - s(n_1, n_2 + 1, k + 1) + s(n_1, n_2, k + 1) - s(n_1 + 1, n_2, k + 1) \\ &\quad + s(n_1 + 1, n_2 + 1, k + 1) - s(n_1 + 1, n_2 + 1, k + 1)\} \end{aligned}$$

$$\begin{aligned}
 & -s(n_1 + 1, n_2, k) + s(n_1, n_2 + 1, k + 1) - s(n_1, n_2, k + 1) \\
 & + s(n_1 + 1, n_2 + 1, k + 1) - s(n_1 + 1, n_2, k + 1) \} \\
 \frac{\partial s_c(x_1, x_2, t)}{\partial t} \approx & \frac{1}{4} \{ s(n_1, n_2, k + 1) - s(n_1, n_2, k) + s(n_1 + 1, n_2, k + 1) \\
 & - s(n_1 + 1, n_2, k) + s(n_1, n_2 + 1, k + 1) - s(n_1, n_2 + 1, k) \\
 & + s(n_1 + 1, n_2 + 1, k + 1) - s(n_1 + 1, n_2 + 1, k) \} \quad (5.38)
 \end{aligned}$$

Πολλές άλλες στρατηγικές εύρεσης μέσων όρων υπάρχουν για την εκτίμηση μερικών παραγώγων χρησιμοποιώντας πεπερασμένες διαφορές [Lim 90]. Χωρική και χρονική προ-εξομάλυνση του βίντεο, με φιλτράρισμα χρησιμοποιώντας Γκαουσιανούς πυρήνες συνήθως βοηθά στην εκτίμηση κλίσης.

### Εκτίμηση κλίσης με τοπικές πολυωνυμικές προσαρμογές

Μια εναλλακτική μέθοδος είναι να προσεγγίσουμε το  $s_c(x_1, x_2, t)$  τοπικά με χρήση ενός γραμμικού συνδυασμού πολυωνύμων χαμηλής τάξης. Έτσι :

$$s_c(x_1, x_2, t) \approx \sum_{i=0}^{N-1} \alpha_i \phi_i(x_1, x_2, t) \quad (5.39)$$

όπου  $\phi_i(x_1, x_2, t)$  είναι τα πολυώνυμα βάσης,  $N$  είναι ο αριθμός των συναρτήσεων βάσης που χρησιμοποιούνται στην πολυωνυμική προσέγγιση και  $\alpha_i$  είναι οι συντελεστές του γραμμικού συνδυασμού. Εδώ, θα θέσουμε το  $N$  ίσο με 9, με την ακόλουθη επιλογή των συναρτήσεων βάσης:

$$\phi_i(x_1, x_2, t) = 1, x_1, x_2, t, x_1^2, x_2^2, x_1 x_2, x_1 t, x_2 t \quad (5.40)$$

οι οποίες προτάθηκαν από τον Lim [Lim 90]. Έτσι, η εξίσωση (5.39) γράφεται

$$s_c(x_1, x_2, t) \approx \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \alpha_3 t + \alpha_4 x_1^2 + \alpha_5 x_2^2 + \alpha_6 x_1 x_2 + \alpha_7 x_1 t + \alpha_8 x_2 t \quad (5.41)$$

Οι συντελεστές  $\alpha_i$ ,  $i = 0, \dots, 8$ , εκτιμώνται με τη μέθοδο των ελαχίστων τετραγώνων, η οποία ελαχιστοποιεί το σφάλμα συνάρτησης:

$$e^2 = \sum_{n_1} \sum_{n_2} \sum_k \left( s(n_1, n_2, k) - \sum_{i=0}^{N-1} \alpha_i \phi_i(x_1, x_2, t) \Big|_{[x_1 \ x_2 \ t]^T = \mathbf{V} [n_1 \ n_2 \ k]^T} \right)^2 \quad (5.42)$$

Το άθροισμα πάνω από το  $(n_1, n_2, k)$  μεταφέρεται στην τοπική γειτονιά του εικονοστοιχείου για το οποίο εφαρμόζεται η πολυωνυμική προέγγιση. Μια τυπική περίπτωση περιλαμβάνει 50 εικονοστοιχεία,  $5 \times 5$  χωρικά παράθυρα σε δυο συνεχόμενα πλαίσια.

Από τη στιγμή που οι συντελεστές  $\alpha_i$  έχουν εκτιμηθεί, οι συνιστώσες της κλίσης είναι δυνατό να βρεθούν με απλή διαφορίση,

$$\frac{\partial s_c(x_1, x_2, t)}{\partial x_1} \approx \alpha_1 + 2\alpha_4 x_1 + \alpha_6 x_2 + \alpha_7 t \Big|_{x_1=x_2=t=0} = a_1 \quad (5.43)$$

$$\frac{\partial s_c(x_1, x_2, t)}{\partial x_2} \approx \alpha_2 + 2\alpha_5 x_2 + \alpha_6 x_1 + \alpha_8 t \Big|_{x_1=x_2=t=0} = a_2 \quad (5.44)$$

$$\frac{\partial s_c(x_1, x_2, t)}{\partial t} \approx \alpha_3 + \alpha_7 x_1 + \alpha_8 x_2 \Big|_{x_1=x_2=t=0} = a_3 \quad (5.45)$$

Ομοίως, οι δευτέρας τάξης μερικές παράγωγοι μπορούν εύκολα να εκτιμηθούν χρησιμοποιώντας τους συντελεστές  $\alpha_4$  ως  $\alpha_8$ .

### 5.3.6 Προσαρμοστικές μέθοδοι

Η μέθοδος των Horn-Schunck εφαρμόζει τους περιορισμούς της οπτικής ροής και της εξομάλυνσης καθολικά σε όλη την εικόνα, ή σε όλο το παράθυρο εκτίμησης κίνησης. Αυτό έχει δυο ανεπιθύμητα αποτελέσματα:

1. Ο περιορισμός της εξομάλυνσης δεν εφαρμόζεται σε διεύθυνση κάθετη σ' ένα όριο αντικειμένου. Έτσι, ένας καθολικός περιορισμός εξομάλυνσης θολώνει τις 'άκρες της κίνησης'. Για παράδειγμα, αν ένα αντικείμενο κινείται πάνω σε ένα ακίνητο φόντο, υπάρχει μια απότομη αλλαγή στο πεδίο της κίνησης στα σύνορα του αντικειμένου. Οι άκρες της κίνησης μπορούν να διατηρηθούν εφαρμόζοντας τον περιορισμό της εξομάλυνσης μόνο στις διευθύνσεις κατά τις οποίες οι εντάσεις των εικονοστοιχείων δε μεταβάλλονται σημαντικά. Αυτή είναι η βασική ιδέα του λεγόμενου κατευθυνόμενου ή προσανατολισμένου περιορισμού εξομάλυνσης.
2. Η μη προσαρμοστική μέθοδος, επίσης, επιβάλλει την συνθήκη της οπτικής ροής στις περιοχές κάλυψης, όπου δε θα έπρεπε να ισχύει. Αυτό μπορεί να επιτευχθεί μεταβάλλοντας το  $\alpha$  προσαρμοστικά προκειμένου να ελέγχεται η σχετική ισχύς των περιορισμών της οπτικής ροής και της εξομάλυνσης. Για παράδειγμα, στις περιοχές κάλυψης, όπως είναι οι περιοχές που περιβάλλονται από διακεκομμένες γραμμές στην εικόνα 5.6, ο περιορισμός της οπτικής ροής μπορεί να μην εφαρμοστεί καθόλου, ενώ από την άλλη η συνθήκη της εξομάλυνσης να είναι σε πλήρη ισχύ.

Αρκετοί ερευνητές πρότειναν την εφαρμογή του περιορισμού της εξομάλυνσης κατά μήκος των συνόρων όχι όμως και κάθετα στα σύνορα της περιοχής κάλυψης. Ο Hildreth [Hil 84] ελαχιστοποίησε το κριτήριο της συνάρτησης των Horn-Schunck που δίνεται από την εξίσωση (5.35) κατά μήκος της περιμέτρου του αντικειμένου. Οι Nagel και Enkelman [Nag 86, Enk 88] εισήγαγαν την ιδέα της κατευθυντικής εξομάλυνσης, η οποία περιορίζει την εξομάλυνση κατά την κατεύθυνση της χωρικής κλίσης της εικόνας. Ο Fogel [Fog 91] χρησιμοποίησε κατευθυντικούς περιορισμούς εξομάλυνσης με προσαρμοστική στάθμιση σε ιεραρχική δομή εικόνας. Σημειώστε ότι οι μέθοδοι προσαρμοστικών σταθμίσεων απαιτούν συγκεκριμένες στρατηγικές για την ανίχνευση των κινούμενων ορίων του αντικειμένου. Ο Snyder [Sny 91] πρότεινε μια γενική διατύπωση του περιορισμού εξομάλυνσης, ο οποίος περιέχει μερικές από τις παραπάνω ως ειδικές περιπτώσεις.

#### Περιορισμός Κατευθυντικής Εξομάλυνσης

Ο περιορισμός της κατευθυνόμενης εξομάλυνσης δίνεται από την εξίσωση

$$\mathcal{E}_{ds}^2(\mathbf{v}(\mathbf{x}, t)) = (\nabla v_1)^T \mathbf{W} (\nabla v_1) + (\nabla v_2)^T \mathbf{W} (\nabla v_2) \quad (5.46)$$

όπου  $\mathbf{W}$  είναι ένας πίνακας βαρών που 'τιμωρεί' μεταβολές στο πεδίο της κίνησης που εξαρτώνται από τις χωρικές αλλαγές του επιπέδου του γκρι στο βίντεο. Διάφορες παραλλαγές του πίνακα βαρών  $\mathbf{W}$  υπάρχουν [Nag 86, Nag 87, Enk 88]. Για παράδειγμα, μπορεί να επιλεγεί:

$$\mathbf{W} = \frac{\mathbf{F} + \delta \mathbf{I}}{\text{trace}(\mathbf{F} + \delta \mathbf{I})} \quad (5.47)$$

όπου  $\mathbf{I}$  είναι ο μοναδιαίος πίνακας και αναπαριστά έναν καθολικό όρο εξομάλυνσης για να βεβαιώνει ότι υπάρχει ένας μη μηδενικός πίνακας βαρών σε χωρικά ομοιόμορφες περιοχές,  $\delta$  είναι ένας βαθμωτός συντελεστής, και



$$\mathbf{F} = \begin{bmatrix} (\frac{\partial s_c}{\partial x_1})^2 + b^2((\frac{\partial^2 s_c}{\partial x_1^2})^2 + (\frac{\partial^2 s_c}{\partial x_1 \partial x_2})^2) & \frac{\partial s_c}{\partial x_1} \frac{\partial s_c}{\partial x_1} + b^2 \frac{\partial^2 s_c}{\partial x_1 \partial x_2} (\frac{\partial^2 s_c}{\partial x_1^2} + \frac{\partial^2 s_c}{\partial x_1^2}) \\ \frac{\partial s_c}{\partial x_1} \frac{\partial s_c}{\partial x_1} + b^2 \frac{\partial^2 s_c}{\partial x_1 \partial x_2} (\frac{\partial^2 s_c}{\partial x_1^2} + \frac{\partial^2 s_c}{\partial x_1^2}) & (\frac{\partial s_c}{\partial x_2})^2 + b^2((\frac{\partial^2 s_c}{\partial x_1 \partial x_2})^2 + (\frac{\partial^2 s_c}{\partial x_2^2})^2) \end{bmatrix}$$

με το  $b^2$  σταθερά.

Τότε, η μέθοδος της κατευθυνόμενης εξομάλυνσης ελαχιστοποιεί τη συνάρτηση κριτηρίου

$$\min_{\mathbf{v}(\mathbf{x},t)} \int_{\mathcal{A}} (\mathcal{E}_{of}^2(\mathbf{v}) + \alpha^2 \mathcal{E}_{ds}^2(\mathbf{v})) d\mathbf{x} \quad (5.48)$$

για να βρει μια εκτίμηση του πεδίου της κίνησης, όπου το  $\mathcal{A}$  είναι η περιοχή υποστήριξης της εικόνας και το  $\alpha^2$  την παράμετρο εξομάλυνσης. Παρατηρήστε ότι η μέθοδος των Horn-Schunck (5.35) είναι μια ειδική περίπτωση του τύπου αυτού με  $\delta = 1$  και  $\mathbf{F} = 0$ . Η επαναληπτική μέθοδος των Gauss-Seidel μπορεί να χρησιμοποιηθεί για την ελαχιστοποίηση της (5.48) όπως περιγράφηκε στο [Enk 88], όπου ο όρος ανανέωσης σε κάθε επανάληψη υπολογίζεται με χρήση ενός γραμμικού αλγόριθμου. Η απόδοση της μεθόδου της κατευθυνόμενης εξομάλυνσης εξαρτάται από το πόσο ακριβής είναι η εκτίμηση των απαιτούμενων μερικών παραγώγων της έντασης της εικόνας.

### Ιεραρχική Προσέγγιση

Ο Fogel [Fog 91] χρησιμοποίησε τις ιδέες της κατευθυνόμενης εξομάλυνσης και της προσαρμοστικής στάθμισης σε έναν κομψό ιεραρχικό σχηματισμό. Μια πολλαπλής ανάλυσης αναπαράσταση  $s_c^a(x_1, x_2, t)$  του βίντεο ορίστηκε ως

$$s_c^a(x_1, x_2, t) = \int_{\mathcal{A}} s_c(x_1, x_2, t) h(\frac{\mu - x_1}{\alpha}, \frac{\eta - x_2}{\alpha}) d\mu d\eta \quad (5.49)$$

όπου  $\alpha$  είναι η παράμετρος ευκρίνειας,  $\mathcal{A}$  δηλώνει την υποστήριξη εικόνας, και

$$h(x_1, x_2) = \begin{cases} A \exp\{\frac{-(C^2 x_1^2 + D^2 x_2^2)}{B^2 - (C^2 x_1^2 + D^2 x_2^2)}\} & C^2 x_1^2 + D^2 x_2^2 < B^2 \\ 0 & \text{διαφορετικά} \end{cases} \quad (5.50)$$

δηλώνει ένα χαμηλοπερατό φίλτρο με παραμέτρους  $A$ ,  $B$ ,  $C$  και  $D$ . Εύκολα μπορεί να φανεί ότι η χωρική ανάλυση του  $s_c^a(x_1, x_2, t)$  ελαττώνεται καθώς το  $\alpha$  αυξάνεται. Παρατηρήστε ακόμη ότι οι χωρικές μερικές παράγωγοι του  $s_c^a(x_1, x_2, t)$  μπορούν να υπολογιστούν με συσχέτιση του  $s_c(x_1, x_2, t)$  με τις μερικές παραγώγους του  $h(x_1, x_2)$  οι οποίες μπορούν να υπολογιστούν αναλυτικά. Ένα μη γραμμικό πρόβλημα βελτιστοποίησης επιλύθηκε σε κάθε επίπεδο ευκρίνειας χρησιμοποιώντας τη μέθοδο του Quasi-Newton. Η εκτίμηση που επετεύχθη σε ένα επίπεδο ευκρίνειας χρησιμοποιήθηκε ως η αρχική εκτίμηση στο επόμενο υψηλότερο επίπεδο ευκρίνειας.

## 5.4 Αναδρομικές Μέθοδοι εκτίμησης διδιάστατης κίνησης

Όλες οι μέθοδοι εκτίμησης κίνησης, σε οποιαδήποτε μορφή, εμπλέκουν τον περιορισμό της οπτικής ροής συνοδεύόμενο από κάποιο περιορισμό εξομάλυνσης. Οι αναδρομικές μέθοδοι με βάση τα εικονοστοιχεία είναι εκτιμητές πρόβλεψης και διόρθωσης, της μορφής:

$$\hat{\mathbf{d}}_a(\mathbf{x}, t; \Delta t) = \hat{\mathbf{d}}_b(\mathbf{x}, t; \Delta t) + \mathbf{u}(\mathbf{x}, t; \Delta t) \quad (5.51)$$

όπου  $\hat{\mathbf{d}}_a(\mathbf{x}, t; \Delta t)$  δηλώνει το εκτιμώμενο διάνυσμα κίνησης στη θέση  $x$ , τη στιγμή  $t$ ,  $\hat{\mathbf{d}}_b(\mathbf{x}, t; \Delta t)$  δηλώνει τον προβλεπόμενο εκτιμητή κίνησης, και  $\mathbf{u}(\mathbf{x}, t; \Delta t)$  είναι ο όρος ενημέρωσης. Οι δείκτες 'a' και 'b' δηλώνουν τους όρους πριν και μετά την ενημέρωση στη θέση του εικονοστοιχείου  $(\mathbf{x}, t)$ . Το βήμα πρόβλεψης σε κάθε εικονοστοιχείο, υπονοεί ένα τοπικό περιορισμό εξομάλυνσης στους εκτιμητές, και το βήμα ενημέρωσης επιβάλλει τον περιορισμό οπτικής ροής.

Ο εκτιμητής (5.51) συνήθως εφαρμόζεται με ένα αναδρομικό τρόπο, εκτελώντας μία ή περισσότερες επαναλήψεις στο  $(x, t)$  και προχωρώντας στο επόμενο εικονοστοιχείο στη κατεύθυνση της αναζήτησης. Πρώιμες αναδρομικές μέθοδοι με βάση τα εικονοστοιχεία εστιάζονταν στην ευκολία της υλοποίησης τους και στη λειτουργία τους σε πραγματικό χρόνο, και γι' αυτό χρησιμοποιούσαν απλές εξισώσεις πρόβλεψης και ενημέρωσης. Γενικά, η καλύτερη διαθέσιμη εκτίμηση στο προηγούμενο εικονοστοιχείο πάρθηκε ως η προβλεπόμενη εκτίμηση για το επόμενο εικονοστοιχείο, ακολουθούμενη από μία ενημέρωση με βάση την κλίση για να ελαχιστοποιήσει το τετράγωνο της διαφοράς του μετατοπισμένου πλαισίου σ' αυτό το εικονοστοιχείο. Αργότερα, προτάθηκαν, πιο εκλεπτυσμένες μέθοδοι πρόβλεψης και ενημέρωσης που απαιτούν περισσότερους υπολογισμούς.

Αρχίζουμε αυτό το υποκεφάλαιο με μία λεπτομερή συζήτηση της σχέσης μεταξύ της ελαχιστοποίησης της διαφοράς του μετατοπισμένου πλαισίου και του περιορισμού της οπτικής ροής στην παράγραφο 5.4.1. Δίνουμε έμφαση στο γεγονός ότι στο βήμα ενημέρωσης, το οποίο ελαχιστοποιεί τη διαφορά του μετατοπισμένου πλαισίου στη συγκεκριμένη θέση του εικονοστοιχείου, πράγματι επιβάλλει τον περιορισμό της οπτικής ροής σ' αυτό το εικονοστοιχείο. Η παράγραφος 5.4.2 είναι μία περίληψη των μεθόδων ελαχιστοποίησης με βάση τη κλίση, οι οποίες είναι αναπόσπαστο κομμάτι των βασικών αναδρομικών μεθόδων με βάση τα εικονοστοιχεία. Μία επέκταση των βασικών αναδρομικών μεθόδων με βάση τα εικονοστοιχεία, η οποία καλείται εκτίμηση Wiener, καλύπτεται στη παράγραφο 5.4.3.

#### 5.4.1 Διαφορά Μετατοπισμένου Πλαισίου

Η θεμελιώδης αρχή σε όλες σχεδόν τις μεθόδους εκτίμησης κίνησης, γνωστή και ως περιορισμός οπτικής ροής, είναι ότι η ένταση της εικόνας παραμένει αναλλοίωτη από πλαίσιο σε πλαίσιο κατά μήκος της πραγματικής κίνησης (ή υπάρχουν αλλαγές με κάποιο γνωστό ή προβλέψιμο τρόπο). Ο περιορισμός της οπτικής ροής μπορεί να εμπλακεί με τη μορφή της εξίσωσης οπτικής ροής ή μπορεί να εξαχθεί με την ελαχιστοποίηση της διαφοράς μετατοπισμένου πλαισίου (ΔΜΠ - DFD), όπως στις μεθόδους προσαρμογής τμημάτων και στις αναδρομικές μεθόδους με βάση τα εικονοστοιχεία.

Έστω η ΔΜΠ μεταξύ των χρονικών στιγμών  $t$  και  $t + \Delta t$  :

$$\Delta s(\mathbf{x}, \mathbf{d}) = s_c(\mathbf{x} + \mathbf{d}(\mathbf{x}, t; \Delta t), t + \Delta t) - s_c(\mathbf{x}, t) \quad (5.52)$$

όπου  $s_c(x_1, x_2, t)$  δείχνει την χρονικά μεταβαλλόμενη κατανομή της εικόνας, και

$$\mathbf{d}(\mathbf{x}, t; \Delta t) \doteq \mathbf{d}(\mathbf{x}) = [d_1(\mathbf{x}) \quad d_2(\mathbf{x})]^T \quad (5.53)$$

δείχνει το πεδίο του διανύσματος μετατόπισης μεταξύ των χρονικών στιγμών  $t$  και  $t + \Delta t$ . Παρατηρούμε πως α) εάν οι συνιστώσες του  $\mathbf{d}(\mathbf{x})$  υποθέτουν μη ακέραιες τιμές, απαιτείται παρεμβολή για να υπολογίσουμε τη ΔΜΠ σε κάθε θέση εικονοστοιχείο και β) εάν το  $\mathbf{d}(\mathbf{x})$  ήταν ίσο με το πραγματικό διάνυσμα μετατόπισης στο σημείο  $\mathbf{x}$  και δεν υπήρχαν λάθη παρεμβολής, η ΔΜΠ τείνει στη μηδενική τιμή στη περιοχή κάτω από το περιορισμό της οπτικής ροής.

Παρακάτω, αναλύουμε το  $s_c(\mathbf{x} + \mathbf{d}(\mathbf{x}), t + \Delta t)$  σε μία σειρά Taylor γύρω από το  $(\mathbf{x}; t)$ . Για μικρά  $\mathbf{d}(\mathbf{x})$  και  $\Delta t$ , όπως:

$$\begin{aligned} s_c(x_1 + d_1(\mathbf{x}), x_2 + d_2(\mathbf{x}); t + \Delta t) &= s_c(\mathbf{x}; t) + d_1(\mathbf{x}) \frac{\partial s_c(\mathbf{x}; t)}{\partial x_1} \\ &+ d_2(\mathbf{x}) \frac{\partial s_c(\mathbf{x}; t)}{\partial x_2} + \Delta t \frac{\partial s_c(\mathbf{x}; t)}{\partial t} + \varepsilon \end{aligned} \quad (5.54)$$

όπου  $\varepsilon$  δηλώνει όρους ανώτερης τάξης. Αντικαθιστώντας την (5.52) στην (5.51), και παραμελώντας τους όρους ανώτερης τάξης έχουμε:

$$\Delta s(\mathbf{x}, \mathbf{d}) = d_1(\mathbf{x}) \frac{\partial s_c(\mathbf{x}; t)}{\partial x_1} + d_2(\mathbf{x}) \frac{\partial s_c(\mathbf{x}; t)}{\partial x_2} + \Delta t \frac{\partial s_c(\mathbf{x}; t)}{\partial t} \quad (5.55)$$

Μελετούμε τη σχέση μεταξύ ΔΜΠ και ΕΟΡ σε δύο περιπτώσεις:

1. Το όριο του  $\Delta t$  τείνει στο 0: Θέτοντας  $\Delta s(\mathbf{x}, d) = 0$ , διαιρώντας και τις δύο πλευρές της 5.55 με το  $\Delta t$ , και παίρνοντας το όριο καθώς το  $\Delta t$  τείνει στο 0, παίρνουμε την ΕΟΡ :

$$u_1(\mathbf{x}, t) \frac{\partial s_c(\mathbf{x}; t)}{\partial x_1} + u_2(\mathbf{x}, t) \frac{\partial s_c(\mathbf{x}; t)}{\partial x_2} + \frac{\partial s_c(\mathbf{x}; t)}{\partial t} = 0 \quad (5.56)$$

όπου  $\mathbf{v}(\mathbf{x}, t) = [u_1(\mathbf{x}, t) u_2(\mathbf{x}, t)]^T$  υποδηλώνει το διάνυσμα της ταχύτητας τη στιγμή  $t$ . Έτσι, η εκτίμηση της ταχύτητας χρησιμοποιώντας την ΕΟΡ και η εκτίμηση της μετατόπισης θέτοντας την ίση με μηδέν είναι ισοδύναμες, καθώς το  $\Delta t$  τείνει στο μηδέν.

2. Για πεπερασμένο  $\Delta t$ : Μία εκτίμηση του διανύσματος μετατόπισης  $\hat{d}(\mathbf{x})$  μεταξύ οποιωνδήποτε δύο πλαισίων τα οποία διαφέρουν κατά  $\Delta t$  μπορεί να εξαχθεί από την 5.55 με αρκετούς τρόπους:

- (α') Αναζήτηση για το  $\hat{d}(\mathbf{x})$ , η οποία θα θέσει το αριστερό μέλος της (5.55), που δίνεται από την (5.52), ίσο με μηδέν πάνω σε ένα τμήμα από εικονοστοιχεία (μέθοδοι ταιριάσματος τμημάτων).
- (β') Ο υπολογισμός του  $\hat{d}(\mathbf{x})$ , ο οποίος θα θέσει το αριστερό μέλος της (5.55) ίσο με μηδέν, από εικονοστοιχείο σε εικονοστοιχείο, χρησιμοποιώντας μία μέθοδο βελτιστοποίησης βασισμένη στη κλίση (αναδρομική μέθοδος με βάση τα εικονοστοιχεία).
- (γ') Θέτοντας  $\Delta t = 1$  και  $\Delta s(\mathbf{x}, \hat{\mathbf{d}}) = 0$ , λύνοντας για  $\hat{\mathbf{d}}(\mathbf{x})$  χρησιμοποιώντας ένα σύνολο γραμμικών εξισώσεων που προκύπτουν από το δεξιό μέλος της (5.55,) χρησιμοποιώντας ένα τμήμα από εικονοστοιχεία.

Και οι τρεις προσεγγίσεις μπορεί ναδειχθεί ότι είναι ταυτόσημες εάν α) οι τοπικές διακυμάνσεις της χωροχρονικής έντασης της εικόνας είναι γραμμικές, και β) η ταχύτητα είναι σταθερή μέσα στο χρονικό διάστημα  $\Delta t$ , δηλαδή :

$$\hat{d}_1(\mathbf{x}) = \hat{u}_1(\mathbf{x}, t) \Delta t \quad (5.57)$$

και

$$\hat{d}_2(\mathbf{x}) = \hat{u}_2(\mathbf{x}, t) \Delta t. \quad (5.58)$$

Στη πράξη, η ΔΜΠ,  $\Delta s(\mathbf{x}, \mathbf{d})$ , σπάνια μηδενίζεται για κάποιες τιμές του  $\hat{\mathbf{d}}(\mathbf{x})$ , επειδή: α) υπάρχει θόρυβος παρατήρησης, β) υπάρχει επικάλυψη (πρόβλημα καλυμμένου και ακάλυπτου φόντου), γ) δημιουργούνται λάθη από το βήμα της παρεμβολής στην περίπτωση μη ακέραιων διανυσμάτων μετατόπισης, και δ) ο φωτισμός της εικόνας μπορεί να διαφέρει από πλαίσιο σε πλαίσιο. Συνεπώς, στοχεύουμε να ελαχιστοποιήσουμε την απόλυτη τιμή ή το τετράγωνο του  $\Delta s$  (5.52) ή το αριστερό μέλος της ΕΟΡ (5.56) για την εκτίμηση του διδιάστατου πεδίου κίνησης. Οι αναδρομικές μέθοδοι με βάση τα εικονοστοιχεία που παρουσιάζονται εδώ χρησιμοποιούν τεχνικές βελτιστοποίησης που βασίζονται στη κλίση για να ελαχιστοποιήσουν το τετράγωνο του  $\Delta s$  (αντίθετα με τη μέθοδο αναζήτησης που χρησιμοποιείται στην προσαρμογή τμημάτων) με ένα υπονοούμενο περιορισμό εξομάλυνσης στο βήμα της πρόβλεψης.

#### 5.4.2 Αλγόριθμοι εκτίμησης διδιάστατης κίνησης με βάση την βέλτιστη κάθοδο

Πριν από την εκτίμηση κίνησης με χρήση αναδρομικών μεθόδων βασισμένων στα εικονοστοιχεία, προηγείται συνήθως ένα στάδιο αναζήτησης αλλαγών φωτεινότητας, όπου η διαφορά πλαισίου

σε κάθε εικονοστοιχείο συγκρίνεται με κάποιο κατώφλι. Η εκτίμηση εκτελείται μόνο πάνω στα εικονοστοιχεία που ανήκουν στην μεταβληθείσα περιοχή. Οι αναδρομικοί αλγόριθμοι με βάση τα εικονοστοιχεία που χρησιμοποιούν την τεχνική της βέλτιστης καθόδου εκτιμούν το βήμα ενημέρωσης  $\mathbf{u}(\mathbf{x}, t; \Delta t)$  στην (5.51) σε κάθε εικονοστοιχείο στη μεταβληθείσα περιοχή, ελαχιστοποιώντας μία θετικά ορισμένη συνάρτηση  $E$  της διαφοράς πλαισίου  $\Delta s(\mathbf{x}, \mathbf{d})$ , όσον αφορά το  $\mathbf{d}$ . Η συνάρτηση  $E$  πρέπει να είναι θετικά ορισμένη, όπως η τετραγωνική συνάρτηση, έτσι ώστε το ελάχιστο να προκύπτει όταν η ΔΜΠ είναι μηδενική. Η ΔΜΠ συγκλίνει στο μηδέν τοπικά, όταν  $\hat{\mathbf{d}}_a(\mathbf{x}, t; \Delta t)$  συγκλίνει στην πραγματική μετατόπιση. Έτσι, το βήμα ενημέρωσης αντιστοιχεί στην επιβολή του περιορισμού της οπτικής ροής τοπικά. Παρακάτω, παρουσιάζουμε κάποιες παραλλαγές της βασικής μεθόδου εκτίμησης με αναδρομικές μεθόδους που βασίζονται στα εικονοστοιχεία με χρήση της απότομης καθόδου.

### Αλγόριθμος Netravali-Robbins

Ο αλγόριθμος Newton-Raphson βρίσκει μία εκτίμηση του διανύσματος μετατόπισης, το οποίο ελαχιστοποιεί το τετράγωνο της ΔΜΠ σε κάθε εικονοστοιχείο, χρησιμοποιώντας μία μέθοδο καθόδου της κλίσης. Η συνάρτηση κριτηρίου που θα ελαχιστοποιηθεί δίνεται από την:

$$E(\mathbf{x}; d) = [\Delta s(\mathbf{x}, \mathbf{d})]^2 \quad (5.59)$$

Από τη παράγραφο του παραρτήματος η ελαχιστοποίηση του  $E(\mathbf{x}; d)$  σε σχέση με το  $\mathbf{d}$ , στο εικονοστοιχείο  $\mathbf{x}$ , από τη μέθοδο της απότομης καθόδου παράγει την αναδρομική σχέση

$$\hat{\mathbf{d}}^{(i+1)}(\mathbf{x}) = \hat{\mathbf{d}}^{(i)}(\mathbf{x}) - \left(\frac{1}{2}\right) \epsilon \nabla_{\mathbf{d}} [\Delta s(\mathbf{x}, \mathbf{d})|_{\mathbf{d}=\hat{\mathbf{d}}^{(i)}}]^2 = \hat{\mathbf{d}}^{(i)}(\mathbf{x}) - \epsilon \Delta s(\mathbf{x}, \hat{\mathbf{d}}^{(i)}) \nabla_{\mathbf{d}} \Delta s(\mathbf{x}, \mathbf{d})|_{\mathbf{d}=\hat{\mathbf{d}}^{(i)}} \quad (5.60)$$

όπου  $\nabla \mathbf{d}$  είναι η κλίση σε σχέση με το  $\mathbf{d}$ , και το  $\epsilon$  είναι το μέγεθος βήματος. Θυμηθείτε ότι η αρνητική κλίση δείχνει προς την κατεύθυνση της βέλτιστης καθόδου.

Τώρα θα συζητήσουμε τον υπολογισμό της  $\nabla_{\mathbf{d}} \Delta s(\mathbf{x}, \mathbf{d})$ . Από την (5.52), μπορούμε να γράψουμε

$$\Delta s(\mathbf{x}, \mathbf{d}) - \Delta s(\mathbf{x}, \hat{\mathbf{d}}^{(i)}) = s_c(\mathbf{x} + \mathbf{d}, t + \Delta t) - s_c(\mathbf{x} + \hat{\mathbf{d}}^{(i)}, t + \Delta t) \quad (5.61)$$

Τώρα, αναλύοντας την ένταση  $s_c(\mathbf{x} + \mathbf{d}, t + \Delta t)$  σε ένα αυθαίρετο σημείο  $\mathbf{x} + \mathbf{d}$  σε μία σειρά Taylor γύρω από το  $\mathbf{x} + \hat{\mathbf{d}}^{(i)}$ , έχουμε

$$s_c(\mathbf{x} + \mathbf{d}, t + \Delta t) = s_c(\mathbf{x} + \hat{\mathbf{d}}^{(i)}, t + \Delta t) + (\mathbf{d} - \hat{\mathbf{d}}^{(i)})^T \nabla_{\mathbf{x}} s_c(\mathbf{x} - \mathbf{d}; t - \Delta t)|_{\mathbf{d}=\hat{\mathbf{d}}^{(i)}} + o(\mathbf{x}, \hat{\mathbf{d}}^{(i)}) \quad (5.62)$$

όπου  $o(\mathbf{x}, \hat{\mathbf{d}}^{(i)})$  οι όροι υψηλότερης τάξης της σειράς. Αντικαθιστώντας την ανάπτυξη της σειράς Taylor στην (5.60), έχουμε την γραμμική έκφραση της ΔΜΠ

$$\Delta s(\mathbf{x}, \mathbf{d}) = \Delta s(\mathbf{x}, \hat{\mathbf{d}}^{(i)}) + \nabla_{\mathbf{x}}^T s_c(\mathbf{x} - \hat{\mathbf{d}}^{(i)}; t - \Delta t)(\mathbf{d} - \hat{\mathbf{d}}^{(i)}) + o(\mathbf{x}, \hat{\mathbf{d}}^{(i)}) \quad (5.63)$$

όπου  $\nabla_{\mathbf{x}} s_c(\mathbf{x} - \hat{\mathbf{d}}^{(i)}; t - \Delta t) \doteq \nabla_{\mathbf{x}} s_c(\mathbf{x} - \mathbf{d}; t - \Delta t)|_{\mathbf{d}=\hat{\mathbf{d}}^{(i)}}$ . Χρησιμοποιώντας την (5.62) και αγνοώντας τους όρους υψηλής τάξης, μπορούμε να εκφράσουμε την κλίση της ΔΜΠ σε σχέση με το  $\mathbf{d}$  με όρους της χωρικής κλίσης της έντασης της εικόνας ως εξής:

$$\nabla_{\mathbf{d}} \Delta s(\mathbf{x}, \mathbf{d})|_{\mathbf{d}=\hat{\mathbf{d}}^{(i)}} = \nabla_{\mathbf{x}}^T s_c(\mathbf{x} - \hat{\mathbf{d}}^{(i)}; t - \Delta t) \quad (5.64)$$

και ο αναδρομικός εκτιμητής βασισμένος στα εικονοστοιχεία γίνεται:

$$\hat{\mathbf{d}}^{(i+1)}(\mathbf{x}) = \hat{\mathbf{d}}^{(i)}(\mathbf{x}) - \epsilon \Delta s(\mathbf{x}, \hat{\mathbf{d}}^{(i)}) \nabla_{\mathbf{x}} s_c(\mathbf{x} - \hat{\mathbf{d}}^{(i)}; t - \Delta t) \quad (5.65)$$

Στη (5.65), ο πρώτος και ο δεύτερος όρος είναι ο όρος πρόβλεψης και ενημέρωσης αντίστοιχα. Σημειώστε ότι ο υπολογισμός της διαφοράς πλαισίου  $\Delta s(\mathbf{x}, \hat{\mathbf{d}}^{(i)})$  και το χωρικό διάνυσμα της κλίσης μπορεί να απαιτεί παρεμβολή της τιμής της έντασης για μη ακέραιες εκτιμήσεις της μετατόπισης.

Το πρόβλημα του διαφράγματος είναι επίσης εμφανές στους αναδρομικούς αλγορίθμους με βάση τα εικονοστοιχεία. Ο όρος ενημέρωσης είναι ένα διάνυσμα κατά μήκος της χωρικής κλίσης της έντασης της εικόνας. Προφανώς, καμία διόρθωση δεν εκτελείται στη κάθετη διεύθυνση του διανύσματος κλίσης.

Σε μία προσπάθεια να απλοποιήσουν περαιτέρω τη δομή του εκτιμητή, οι Netravali και Robbins πρότειναν τον τροποποιημένο τύπο εκτίμησης,

$$\hat{\mathbf{d}}^{(i+1)}(\mathbf{x}) = \hat{\mathbf{d}}^{(i)}(\mathbf{x}) - \epsilon \operatorname{sgn}(\Delta s(\mathbf{x}, \hat{\mathbf{d}}^{(i)})) \operatorname{sgn}(\nabla_{\mathbf{x}} s_c(\mathbf{x} - \hat{\mathbf{d}}^{(i)}; t - \Delta t)), \quad (5.66)$$

όπου ο όρος ενημέρωσης παίρνει μία από τις τρεις τιμές  $\pm\epsilon$  και μηδέν. Σ' αυτή την περίπτωση, οι εκτιμητές κίνησης ενημερώνονται μόνο κατά τις διευθύνσεις 0, 45, 90, 135, ... μοιρών.

Η σύγκλιση και ο ρυθμός σύγκλισης του αλγορίθμου Netravali-Robbins εξαρτάται από την επιλογή της παραμέτρου  $\epsilon$  του μεγέθους του βήματος. Για παράδειγμα, εάν  $\epsilon = \frac{1}{16}$ , τότε τουλάχιστον 32 επαναλήψεις απαιτούνται για τον υπολογισμό μίας μετατόπισης 2 εικονοστοιχείων. Από την άλλη πλευρά, μία επιλογή ενός μεγάλου μεγέθους βήματος μπορεί να οδηγήσει σε ταλαντώσεις της εκτίμησης κίνησης. Αρκετές στρατηγικές μπορούν να οδηγήσουν στη βελτίωση της σύγκλισης του αλγορίθμου.

### Αλγόριθμος Walker-Rao

Οι Walker και ο Rao πρότειναν ένα προσαρμοστικό μέγεθος βήματος υποκινούμενοι από τις παρακάτω παρατηρήσεις: α) Στη γειτονιά μίας ακμής όπου το  $|\nabla s_c(x_1, x_2, t)|$  είναι μεγάλο, το  $\epsilon$  πρέπει να είναι μικρό εάν η ΔΜΠ είναι μικρή, έτσι ώστε να μη χρειασθεί να κάνουμε μία αδικαιολόγητα μεγάλη ενημέρωση. Επιπλέον, στη γειτονιά μίας ακμής, η ακρίβεια της εκτίμησης κίνησης έχει ζωτική σημασία, το οποίο επίσης απαιτεί μικρό μέγεθος βήματος. β) Στις ομοιόμορφες περιοχές της εικόνας όπου  $|\nabla s_c(x_1, x_2, t)|$  είναι μικρό, χρειαζόμαστε μεγάλο μέγεθος βήματος όταν η ΔΜΠ είναι μεγάλη. Και οι δύο αυτές απαιτήσεις μπορούν να καλυφθούν από ένα μέγεθος βήματος της μορφής

$$\epsilon = \frac{1}{2 \|\nabla_{\mathbf{x}} s_c(\mathbf{x} - \hat{\mathbf{d}}^i; t - \Delta t)\|^2} \quad (5.67)$$

Επιπρόσθετα, οι Walker και Rao εισήγαγαν τους ακόλουθους ευριστικούς κανόνες:

1. Εάν η ΔΜΠ είναι μικρότερη από κάποιο κατώφλι, ο όρος ενημέρωσης τίθεται ίσος με μηδέν
2. Εάν η ΔΜΠ υπερβαίνει κάποιο κατώφλι, αλλά το μέτρο της χωρικής κλίσης της εικόνας είναι μηδέν, τότε ο όρος ενημέρωσης πάλι τίθεται ίσος με μηδέν.
3. Εάν η απόλυτη τιμή του όρου ενημέρωσης (για κάθε συνιστώσα) είναι μικρότερη από το  $\frac{1}{16}$ , τότε τίθεται ίσος με  $\pm \frac{1}{16}$ .
4. Εάν η απόλυτη τιμή του όρου ενημέρωσης (για κάθε συνιστώσα) είναι πάνω από 2, τότε τίθεται ίση με  $\pm 2$

Οι Caffario και Rocca επίσης, ανέπτυξαν μία παρόμοια έκφραση του μεγέθους βήματος [Caf83]

$$\epsilon = \frac{1}{\|\nabla_{\mathbf{x}} s_c(\mathbf{x} - \hat{\mathbf{d}}^i; t - \Delta t)\|^2 + \eta^2} \quad (5.68)$$

η οποία περιλαμβάνει έναν όρο μεροληψίας (bias)  $\eta^2$  για την αποφυγή της διαίρεσης με το μηδέν σε περιοχές σταθερής έντασης όπου η χωρική κλίση είναι ίση με το μηδέν. Μία τυπική τιμή για το  $\eta$  είναι  $\eta^2 = 100$ .

Πειραματικά αποτελέσματα δείχνουν ότι η χρήση ενός προσαρμοστικού μέγεθους βήματος βελτιώνει περισσότερο τη σύγκλιση του αλγορίθμου Netravali-Robbins. Έχει βρεθεί ότι πέντε επαναλήψεις ήταν αρκετές για να πετύχουμε ικανοποιητικά αποτελέσματα στις περισσότερες περιπτώσεις.

### Επέκταση στο Μοντέλο Κίνησης Κατά Τμήματα

Είναι πιθανό να επιβάλλουμε έναν ισχυρότερο περιορισμό ομαλότητας στις εκτιμήσεις σε κάθε εικονοστοιχείο  $\mathbf{x}$ , υποθέτοντας ότι η μετατόπιση παραμένει σταθερή τοπικά πάνω από ολισθαίνουσα περιοχή υποστήριξης  $B$  γύρω από κάθε εικονοστοιχείο. Μπορούμε, τότε, να ελαχιστοποιήσουμε τη ΔΜΠ πάνω από την περιοχή υποστήριξης  $B$ , ως εξής:

$$E(\mathbf{x}, \mathbf{d}) = \sum_{\mathbf{x}_B \in B} [\Delta s(\mathbf{x}_B, \mathbf{d})]^2 \quad (5.69)$$

αντί σε κάθε εικονοστοιχείο. Παρατηρήστε ότι η περιοχή υποστήριξης  $B$  πρέπει να είναι ‘αιτιατή’ (στη λογική της αναδρομικής υπολογισιμότητας) για να διατηρήσουμε την αναδρομική φύση του αλγορίθμου. Μία τυπική αιτιατή περιοχή υποστήριξης με  $N = 7$  εικονοστοιχεία φαίνεται στο σχήμα 5.14.

1	2	3	4
5	6	X	

Σχήμα 5.14: Μία αιτιατή περιοχή υποστήριξης  $B$  για  $N = 7$ .

Ακολουθώντας βήματα παρόμοια με αυτά του αλγορίθμου που εκτελείται σε εικονοστοιχεία (αντί σε τμήματα), η ελαχιστοποίηση με τη μέθοδο της απότομης καθόδου αυτής της συνάρτησης κριτηρίου παράγει την επανάληψη:

$$\hat{\mathbf{d}}^{(i+1)}(\mathbf{x}) = \hat{\mathbf{d}}^{(i)}(\mathbf{x}) - \epsilon \sum_{\mathbf{x}_B \in B} \Delta s(\mathbf{x}_B, \hat{\mathbf{d}}^{(i)}) \nabla_{\mathbf{x}} s_c(\mathbf{x}_B - \hat{\mathbf{d}}^{(i)}; t - \Delta t) \quad (5.70)$$

Παρατηρήστε ότι αυτή η σχέση είναι ισοδύναμη με αυτήν του ταιριάσματος τμημάτων, εκτός από το σχήμα της περιοχής υποστήριξης  $B$ . Εδώ, η λύση αναζητείται για τη χρήση ελαχιστοποίησης απότομης καθόδου αντί για μια στρατηγική αναζήτησης.

### 5.4.3 Αλγόριθμοι με βάση την εκτίμηση Wiener

Οι μέθοδοι που βασίζονται στην εκτίμηση Wiener είναι επεκτάσεις του αλγορίθμου των Netravali-Robbins στη περίπτωση της κίνησης κατά τμήματα, όπου οι όροι ανώτερης τάξης στη γραμμική μορφή της έκφρασης της ΔΜΠ (5.62) δεν αγνοούνται. Σε αντίθεση, ένα γραμμικό σφάλμα ελαχίστων τετραγώνων (LLSE) ή ένα γραμμικό ελάχιστο μέσο τετραγωνικό σφάλμα (LMMSE) εκτιμά τον όρο ενημέρωσης

$$\mathbf{u}^{(i)}(\mathbf{x}) = \hat{\mathbf{d}}(\mathbf{x}) - \hat{\mathbf{d}}^{(i)}(\mathbf{x}) \quad (5.71)$$

όπου  $\mathbf{d}(\mathbf{x})$  δείχνει το πραγματικό διάνυσμα μετατόπισης, παράγεται με βάση μία γειτονιά  $B$  γύρω από ένα εικονοστοιχείο  $\mathbf{x}$ . Παρακάτω, θα δώσουμε την παραγωγή του εκτιμητή Wiener για τη περίπτωση  $N$  παρατηρήσεων με ένα κοινό διάνυσμα μετατόπισης  $\mathbf{d}(\mathbf{x})$ .

Γράφοντας την γραμμική ΔΜΠ με  $\Delta s(\mathbf{x}_B, \mathbf{d}) = 0$ , σε όλα τα  $N$  εικονοστοιχεία  $\mathbf{x}_B$  μέσα στη περιοχή υποστήριξης  $B$ , έχουμε  $N$  εξισώσεις με δύο αγνώστους (τις συνιστώσες του  $\mathbf{u}^{(i)}$ ) που δίνονται από

$$\begin{aligned} -\Delta s(\mathbf{x}_B(1), \hat{\mathbf{d}}^{(i)}(\mathbf{x})) &= \nabla^T s_c(\mathbf{x}_B(1) - \hat{\mathbf{d}}^{(i)}(\mathbf{x}), t - \Delta t) \mathbf{u}^{(i)} + o(\mathbf{x}_b(1), \hat{\mathbf{d}}^{(i)}(\mathbf{x})) \\ -\Delta s(\mathbf{x}_B(2), \hat{\mathbf{d}}^{(i)}(\mathbf{x})) &= \nabla^T s_c(\mathbf{x}_B(2) - \hat{\mathbf{d}}^{(i)}(\mathbf{x}), t - \Delta t) \mathbf{u}^{(i)} + o(\mathbf{x}_b(2), \hat{\mathbf{d}}^{(i)}(\mathbf{x})) \\ &\vdots = \vdots \\ -\Delta s(\mathbf{x}_B(N), \hat{\mathbf{d}}^{(i)}(\mathbf{x})) &= \nabla^T s_c(\mathbf{x}_B(N) - \hat{\mathbf{d}}^{(i)}(\mathbf{x}), t - \Delta t) \mathbf{u}^{(i)} + o(\mathbf{x}_b(N), \hat{\mathbf{d}}^{(i)}(\mathbf{x})) \end{aligned} \quad (5.72)$$

όπου  $\mathbf{x}_B(1), \dots, \mathbf{x}_B(N)$  δείχνουν μία διάταξη των εικονοστοιχείων μέσα στη περιοχή υποστήριξης  $B$  (που φαίνεται στο σχήμα 5.14). Αυτές οι εξισώσεις μπορούν να εκφραστούν σε μορφή διανυσμάτων - πινάκων ως:

$$\mathbf{z} = \Phi \mathbf{u}(\mathbf{x}) + \mathbf{n} \quad (5.73)$$

όπου

$$\mathbf{z} = \begin{pmatrix} -\Delta s(\mathbf{x}_B(1), \hat{\mathbf{d}}^{(i)}(\mathbf{x})) \\ -\Delta s(\mathbf{x}_B(2), \hat{\mathbf{d}}^{(i)}(\mathbf{x})) \\ \vdots \\ -\Delta s(\mathbf{x}_B(N), \hat{\mathbf{d}}^{(i)}(\mathbf{x})) \end{pmatrix}$$

$$\Phi = \begin{pmatrix} \frac{\partial s_c(\mathbf{x}_B(1) - \hat{\mathbf{d}}^{(i)}, t - \Delta t)}{\partial x_1} & \frac{\partial s_c(\mathbf{x}_B(1) - \hat{\mathbf{d}}^{(i)}, t - \Delta t)}{\partial x_2} \\ \frac{\partial s_c(\mathbf{x}_B(2) - \hat{\mathbf{d}}^{(i)}, t - \Delta t)}{\partial x_1} & \frac{\partial s_c(\mathbf{x}_B(2) - \hat{\mathbf{d}}^{(i)}, t - \Delta t)}{\partial x_2} \\ \vdots & \vdots \\ \frac{\partial s_c(\mathbf{x}_B(N) - \hat{\mathbf{d}}^{(i)}, t - \Delta t)}{\partial x_1} & \frac{\partial s_c(\mathbf{x}_B(N) - \hat{\mathbf{d}}^{(i)}, t - \Delta t)}{\partial x_2} \end{pmatrix}$$

και

$$\mathbf{n} = \begin{pmatrix} o(\mathbf{x}_B(1), \hat{\mathbf{d}}^{(i)}) \\ o(\mathbf{x}_B(2), \hat{\mathbf{d}}^{(i)}) \\ \vdots \\ o(\mathbf{x}_B(N), \hat{\mathbf{d}}^{(i)}) \end{pmatrix}$$

Υποθέτοντας ότι ο όρος ενημέρωσης  $\mathbf{u}(\mathbf{x})$  και θόρυβος  $\mathbf{n}$  είναι ασυσχέτιστα τυχαία διανύσματα, και χρησιμοποιώντας της αρχή της ορθογωνιότητας, η LMMSE εκτίμηση του όρου ενημέρωσης δίνεται από [Bie87]

$$\hat{\mathbf{u}}(\mathbf{x}) = [\Phi^T \mathbf{R}_n^{-1} \Phi + \mathbf{R}_u^{-1}]^{-1} \Phi^T \mathbf{R}_n^{-1} \mathbf{z} \quad (5.74)$$

Σημειώστε ότι η λύση απαιτεί γνώση των πινάκων συμεταβλητότητας, του όρου ενημέρωσης  $\mathbf{R}_u$  και του θορύβου  $\mathbf{R}_n$ . Όταν δεν έχουμε ακριβή γνώση αυτών των ποσοτήτων, θα κάνουμε την υπόθεση απλοποίησης ότι και τα δύο διανύσματα έχουν μέσες τιμές μηδέν, και οι συνιστώσες τους είναι ασυσχέτιστες μεταξύ τους. Δηλαδή,  $\mathbf{R}_u = \sigma_u^2 \mathbf{I}$  και  $\mathbf{R}_n = \sigma_n^2 \mathbf{I}$ , όπου  $\mathbf{I}$  ένας μοναδιαίος πίνακας  $2 \times 2$  και  $\sigma_u^2$  και  $\sigma_n^2$  είναι οι μεταβλητότητες των συνιστωσών των δύο διανυσμάτων, αντίστοιχα. Τότε η LMMSE εκτίμηση (5.76) απλοποιείται σε

$$\hat{\mathbf{u}}(\mathbf{x}) = [\Phi^T \Phi + \mu \mathbf{I}]^{-1} \Phi^T \mathbf{z} \quad (5.75)$$

όπου  $\mu = \frac{\sigma_n^2}{\sigma_u^2}$  καλείται παράμετρος απόσβεσης. Η εξίσωση (5.75) δίνει την εκτίμηση ελαχίστων τετραγώνων του όρου ενημέρωσης. Οι υποθέσεις που χρησιμοποιούνται για να οδηγηθούμε σε ένα απλοποιημένο εκτιμητή δεν είναι γενικά αληθείς. Για παράδειγμα, ο θόρυβος  $n$  δεν είναι ασυσχέτιστος με τον όρο ενημέρωσης  $u$ , και τα λάθη ενημέρωσης και γραμμοποίησης σε κάθε εικονοστοιχείο δεν είναι ασυσχέτιστα το ένα με το άλλο. Ωστόσο, πειραματικά αποτελέσματα δείχνουν μεγαλύτερη απόδοση σε σύγκριση με της αναδρομικές μεθόδους με βάση τα εικονοστοιχεία.

Έχοντας αποκτήσει μία εκτίμηση του όρου ενημέρωσης, η Wiener αναδρομική μέθοδος μπορεί να γραφεί ως

$$\hat{\mathbf{d}}^{(i+1)}(\mathbf{x}) = \hat{\mathbf{d}}^{(i)}(\mathbf{x}) + [\Phi^T \Phi + \mu \mathbf{I}]^{-1} \Phi^T \mathbf{z} \quad (5.76)$$

Έχειδειχθεί ότι οι εκτιμητές Wiener συνδέονται με τους αλγόριθμους Walker-Rao και Caffario-Rocca [Bie87]. Αυτό γίνεται εύκολα αντιληπτό γράφοντας την (5.77) στην ειδική περίπτωση που  $N = 1$  σαν:

$$\hat{\mathbf{d}}^{(i+1)}(\mathbf{x}) = \hat{\mathbf{d}}^{(i)}(\mathbf{x}) - \frac{\Delta s(\mathbf{x}, \hat{\mathbf{d}}^{(i)}) \nabla^T s_c(\mathbf{x} - \hat{\mathbf{d}}^{(i)}; t - \Delta t)}{|\nabla^T s_c(\mathbf{x} - \hat{\mathbf{d}}^{(i)}; t - \Delta t)|^2 + \mu} \quad (5.77)$$

Ο αποκαλούμενος και ως απλοποιημένος Caffario-Rocca αλγόριθμος προκύπτει από την εξίσωση όταν  $\mu = 100$ . Για  $\mu = 0$  και πολλαπλασιάζοντας τον όρο ενημέρωσης με  $\frac{1}{2}$  παίρνουμε τον αλγόριθμο Walker-Rao. Οι ιδιότητες της σύγκλισης των εκτιμητών Wiener έχουν αναλυθεί στο [Bor91].

Η εκτίμηση Wiener που παρουσιάζεται σε αυτή την παράγραφο εμπλέκει την βέλτιστη εκτίμηση του διανύσματος κίνησης από την προηγούμενη επανάληψη ως την εκτίμηση για την επόμενη επανάληψη. Ένας βελτιωμένος αλγόριθμος με χωροχρονικούς διανυσματικούς εκτιμητές με αντιστάθμιση κίνησης προτάθηκε στο [Dri91]. Οι αναδρομικοί αλγόριθμοι βασισμένοι στα εικονοστοιχεία έχουν επίσης επεκταθεί ώστε να περιλαμβάνουν και την περιστροφική κίνηση [Bie88]. Ως τελική παρατήρηση σημειώνουμε ότι όλοι οι αναδρομικοί αλγόριθμοι με βάση τα εικονοστοιχεία μπορούν να εφαρμοσθούν ιεραρχικά, με χρήση πολλαπλών αναλύσεων (multiresolutional) αναπαράστασης των εικόνων για να πάρουμε βελτιωμένα αποτελέσματα. Οι αναδρομικοί αλγόριθμοι έχουν εξελιχθεί σε Bayesian μεθόδους εκτίμησης κίνησης, εμπλέκοντας στοχαστικά μοντέλα πεδίων κίνησης.

## 5.5 ΠΑΡΑΡΤΗΜΑ

### Μέθοδος βελτιστοποίησης συναρτήσεων πολλών μεταβλητών

Ο πιο εύκολος τρόπος για την ελαχιστοποίηση μίας συνάρτησης  $f(u_1, \dots, u_n)$   $n$  μεταβλητών  $u_1, \dots, u_n$  είναι να υπολογίσουμε τις μερικές παραγώγους αντίστοιχα για κάθε μεταβλητή και να λύσουμε το σύστημα εξισώσεων:

$$\begin{aligned} \frac{\partial f(u)}{\partial u_1} &= 0 \\ &\vdots \\ \frac{\partial f(u)}{\partial u_n} &= 0 \end{aligned} \quad (5.78)$$

ταυτόχρονα για  $u_1, \dots, u_n$ . Αυτό το σύνολο των εξισώσεων μπορεί να εκφραστεί ως μία διανυσματική εξίσωση :

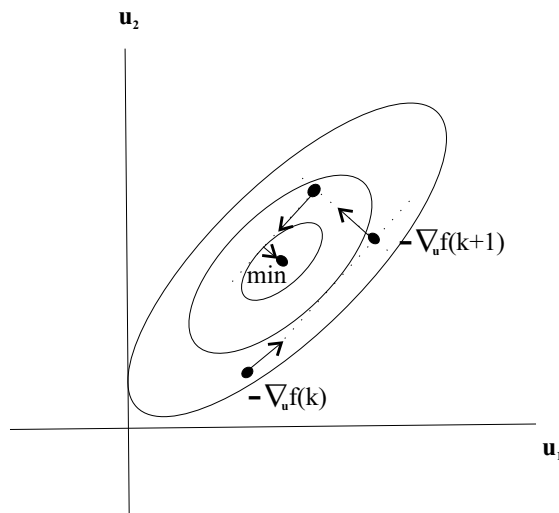
$$\nabla_{\mathbf{u}} f(\mathbf{u}) = \mathbf{0} \quad (5.79)$$



## 5.5. ΠΑΡΑΡΤΗΜΑ

### ΜΕΘΟΔΟΣ ΒΕΛΤΙΣΤΟΠΟΙΗΣΗΣ ΣΥΝΑΡΤΗΣΕΩΝ ΠΟΛΛΩΝ ΜΕΤΑΒΛΗΤΩΝ

όπου  $\nabla_{\mathbf{u}}$  είναι ο τελεστής κλίσης (grad) που αντιστοιχεί στο διάνυσμα  $\mathbf{u}$ . Επειδή είναι δύσκολο να ορίσουμε μία κλειστή έκφραση για τη συνάρτηση κριτηρίου  $f(u_1, \dots, u_n)$  για την εκτίμηση κίνησης, και/ή να λύσουμε το σύνολο των εξισώσεων (5.57) σε μία κλειστή μορφή, καταφεύγουμε στις επαναληπτικές (αριθμητικές) μεθόδους. Για παράδειγμα, η ΔΜΠ είναι μία συνάρτηση από εντάσεις εικονοστοιχείων, οι οποίες δεν είναι δυνατό να εκφρασθούν σε κλειστή μορφή.



Σχήμα 5.15: Αναπαράσταση της μεθόδου της απότομης κλίσης.

#### 5.5.1 Μέθοδος της Βέλτιστης Καθόδου

Η μέθοδος της βέλτιστης καθόδου είναι πιθανότατα η απλούστερη μέθοδος αριθμητικής βελτιστοποίησης. Η μέθοδος αυτή ενημερώνει τη παρούσα εκτίμηση της θέσης του ελάχιστου στη κατεύθυνση της αρνητικής κλίσης, η οποία καλείται κατεύθυνση της βέλτιστης καθόδου. Θυμηθείτε ότι το διάνυσμα της κλίσης  $\nabla_{\mathbf{u}} f(\mathbf{u})$  δείχνει στην κατεύθυνση της μέγιστης ανόδου. Δηλαδή, στη μία διάσταση (συνάρτηση μίας μεταβλητής), το πρόσημο θα είναι θετικό σε μία 'ανηφορική' πλαγιά. Έτσι, η κατεύθυνση της βέλτιστης καθόδου είναι ακριβώς η αντίθετη, όπως φαίνεται στο σχήμα 5.15.

Με σκοπό να πάμε πιο κοντά στο ελάχιστο, ενημερώνουμε την παρούσα εκτίμηση ως εξής:

$$\mathbf{u}^{(k+1)} = \mathbf{u}^{(k)} - a \nabla_{\mathbf{u}} f(\mathbf{u}^{(k)}) \quad (5.80)$$

όπου  $a$  είναι κάποιος θετικός αριθμός, γνωστός και ως μέγεθος βήματος. Το μέγεθος βήματος είναι σημαντικό για τη σύγκλιση των επαναλήψεων, επειδή εάν το  $a$  είναι πολύ μικρό, μετακινούμαστε κατά μία μικρή ποσότητα κάθε φορά, και θα χρειαστούν πολλές επαναλήψεις μέχρι τη σύγκλιση. Από την άλλη πλευρά, αν είναι πολύ μεγάλο, ο αλγόριθμος μπορεί να γίνει ασταθής και να ταλαντεύεται γύρω από το ελάχιστο. Στη μέθοδο της απότομης καθόδου, το μέγεθος βήματος συνήθως βρίσκεται ευριστικά.

#### 5.5.2 Μέθοδος Newton-Raphson

Η βέλτιστη τιμή του βήματος  $a$  μπορεί να εκτιμηθεί χρησιμοποιώντας τη γνωστή από την εύρεση ριζών μέθοδο Newton-Raphson. Εδώ, για λόγους απλότητας θα δείξουμε τη παραγωγή μίας συνάρτησης μίας μεταβλητής. Στη μία διάσταση, θα θέλαμε να βρούμε μια ρίζα της  $f'(u)$ . Σ' αυτό το σημείο, αναλύουμε την  $f'(u)$  σε μία σειρά (Taylor) γύρω από το σημείο  $u^{(k)}$  για να πάρουμε

$$f'(u^{(k+1)}) = f'(u^{(k)}) + (u^{(k+1)} - u^{(k)}) f''(u^{(k)}) \quad (5.81)$$

Καθώς επιθυμούμε η  $u^{(k+1)}$  να είναι ρίζα της  $f'(n)$ , θέτουμε  $f'(u^{(k)}) + (u^{(k+1)} - u^{(k)})f''(u^{(k)}) = 0$ . Λύνοντας ως προς  $u^{(k+1)}$ , έχουμε

$$u^{(k+1)} = u^{(k)} - \frac{f'(u^{(k)})}{f''(u^{(k)})} \quad (5.82)$$

Αυτό το αποτέλεσμα μπορεί να γενικευθεί για την περίπτωση μίας συνάρτησης πολλών μεταβλητών ως εξής:

$$\mathbf{u}^{(k+1)} = \mathbf{u}^{(k)} - \mathbf{H}^{-1} \nabla_{\mathbf{u}} f(\mathbf{u}^{(k)}) \quad (5.83)$$

όπου  $\mathbf{H}$  είναι ο Ιακωβιανός πίνακας με στοιχεία  $H_{ij} = [\frac{\partial^2 f(\mathbf{u})}{\partial u_i \partial u_j}]$ . Η μέθοδος Newton-Raphson βρίσκει μία αναλυτική έκφραση για τη παράμετρο μεγέθους βήματος σε σχέση με τις μερικές παραγώγους δεύτερης τάξης της συνάρτησης προς ελαχιστοποίηση. Όταν δεν είναι διαθέσιμη μία κλειστή μορφή συνάρτησης, ο Ιακωβιανός πίνακας μπορεί να εκτιμηθεί χρησιμοποιώντας αριθμητικές μεθόδους.

### 5.5.3 Σύγκριση Τοπικών και Ολικών Ελαχίστων

Η προσέγγιση με την κλίση της καθόδου έχει ένα σοβαρό μειονέκτημα : η λύση εξαρτάται από το αρχικό σημείο. Εάν αρχίσουμε σε μία 'κοιλιάδα', θα κολλήσει στο πάτο της, ακόμη και αν είναι ένα τοπικό ελάχιστο. Επειδή το διάνυσμα της κλίσης είναι μηδενικό ή σχεδόν μηδενικό, πάνω ή γύρω από το τοπικό ελάχιστο, η ενημέρωση γίνεται πολύ μικρή ώστε η μέθοδος να μπορέσει να ξεφύγει από το τοπικό ελάχιστο. Μία λύση σ' αυτό το πρόβλημα είναι να αρχικοποιούμε τον αλγόριθμο σε αρκετά διαφορετικά σημεία, και έπειτα να παίρνουμε τη λύση που δίνει τη μικρότερη τιμή της συνάρτησης κριτηρίου.

Περισσότερο εκλεπτυσμένες μέθοδοι, όπως της προσομοιωμένης ανόπτυσης, υπάρχουν στην βιβλιογραφία για να φτάσουμε στο ολικό ελάχιστο ανεξαρτήτως του αρχικού σημείου. Ωστόσο, αυτές οι μέθοδοι συνήθως απαιτούν περισσότερο χρόνο επεξεργασίας.

## 5.6 Ασκήσεις

1. Για μια έγχρωμη εικόνα, η εξίσωση οπτικής ροής (5.25) μπορεί να γραφεί για καθένα από τα κανάλια  $R$ ,  $G$  και  $B$  ξεχωριστά. Διατυπώστε τις συνθήκες στις  $(R, G, B)$  εντάσεις, έτσι ώστε να έχουμε τουλάχιστον δύο γραμμικά ανεξάρτητες εξισώσεις σε κάθε εικονοστοιχείο. Πόσο ισχύουν οι συνθήκες αυτές για τις γενικές έγχρωμες εικόνες;
2. Διατυπώστε τις συνθήκες για την χωροχρονική ένταση της εικόνας και την ταχύτητα, για τις οποίες η εξίσωση οπτικής ροής μπορεί να χρησιμοποιηθεί για εκτίμηση μετατόπισης. Γιατί χρειαζόμαστε την υπόθεση της μικρής κίνησης;
3. Ποιες είναι οι συνθήκες για την ύπαρξη της κανονικής ροής (5.27); Μπορούμε πάντα να ανακτήσουμε την οπτική ροή από την κανονική ροή; Συζητήστε τη σχέση μεταξύ των χωρικών κλίσεων της εικόνας και του προβλήματος του διαφράγματος.
4. Προτείνετε μεθόδους για την ανίχνευση occlusion.
5. Δείξτε πώς προκύπτει η (5.29) από την (5.28).
6. Δείξτε ότι ο περιορισμός (5.28) δεν ισχύει όταν υπάρχει περιστροφή ή ζουμ.
7. Βρείτε τις εκτιμήσεις ελαχίστων τετραγώνων των  $\alpha_1$ ,  $\alpha_2$  και  $\alpha_3$  στις (5.43)-(5.45).

8. Πώς μπορούμε να αντιμετωπίσουμε το φαινόμενο ορίων στη μέθοδο συσχέτισης φάσης; Μπορούμε να χρησιμοποιήσουμε τον διακριτό μετασχηματισμό συνημιτόνου, αντί για τον DFT;
9. Προτείνετε ένα μοντέλο για να ποσοτικοποιηθεί η διαρροή φάσματος εξαιτίας της κίνησης σε επίπεδο μικρότερο του εικονοστοιχείου στη μέθοδο συσχέτισης φάσης.
10. Διατυπώστε τις ιδιότητες συμμετρίας του DFT για  $N$  άρτιο και περιττό αντίστοιχα. Επιβεβαιώστε την εξίσωση (5.18).
11. Δείξτε πώς προκύπτει η σχέση (5.83). Πώς θα υπολογίζατε το βέλτιστο μέγεθος βήματος για τον αλγόριθμο Netravali-Robbins;
12. Βρείτε μια αναλυτική έκφραση για την (5.70). Σχολιάστε τη σχέση μεταξύ της προσαρμογής τμημάτων και του αλγορίθμου που δίνεται από την (5.70).
13. Σχολιάστε την ισχύ των υποθέσεων που έγιναν για προκύψει ο εκτιμητής τύπου Wiener (5.76).
14. Αποδείξτε πώς μπορεί να προκύψει η (5.77) από την (5.76).



# Βιβλιογραφία

- [1] [Agg 88] J.K. Aggarwal and N. Nandhakumar, "On the computation of motion from sequences of images", Proc. IEEE, vol.76, pp.917-935, Aug. 1988.
- [2] [Ana 93] P. Anandan, J.R. Bergen, K.J. Hanna, and R. Hingorani, "Hierarchical model-based motion estimation," in Motion Analysis and Image Sequence Processing, M.I. Sezan and R.L. Lagendijk, eds., Norwell, MA: Kluwer, 1993.
- [3] [Bar 94] J.L. Barron, D.J. Fleet, and S.S. Beauchemin, "Systems and experiment :Performance of optical flow techniques", Int. J. Comp. Vision, vol.12:1, pp. 43-77, 1994.
- [4] [Ber 88] M. Bertero, T.A. Poggio and V. Torre, "Ill-posed problems in early vision," Proc. IEEE, vol. 76, pp. 869-889, August 1988.
- [5] [Enk 88] W. Enkelmann, "Investigations of multigrid algorithms for the estimation of optical flow fields in image sequences," Comp. Vis. Graph Image Proc., vol. 43, pp. 150-177, 1988.
- [6] [Fle 92] D.J. Fleet, Measurement of Image Velocity, Norwell, MA:Kluwer, 1992.
- [7] [Fog 91] S.V. Fogel, "Estimation of velocity vector fields from time-varying image sequences", CVGIP: Image Understanding, vol. 53, pp. 253-287, 1991.
- [8] [Hil 84] E.C. Hilreth, "Computations underlying the measurement of visual motion", Artif. Intel., vol. 23, pp. 309-354, 1984.
- [9] [Hor 81] B.K.P. Horn and B.G. Schunck, "Determinin optical flow", Artif. Intel., vol. 17, pp. 185-203, 1981.
- [10] [Hua 81] T.S. Huang, ed., Image Sequence Analysis, Springer Verlag, 1981.
- [11] [Hua 83] T.S. Huang, ed., Image Sequence Processing and Dynamic Scene Analysis, Berlin, Germany: Springer Verlag, 1983.
- [12] [Lim 90] J.S. Lim, ed., Two-Dimensional Signal and Image Processing, Englewood Cliffs, NJ: Prentice Hall, 1990.
- [13] [Luc 81] B.D. Lucas and T. Kanade, "An iterative image resgistration technique with an application to stereo vision", Proc. DARPA Image Understanding Workshop, pp. 121-130, 1981.
- [14] [Nag 86] H.H. Nagel and W. Enkelmann, "An investigation of smoothness constraints for the estimation of displacement vector fields from image sequences", IEEE Trans. Patt. Anal. Mach. Intel., vol. 8, pp. 565-593, 1986.

- [15] [Nag 87] H.H. Nagel, "On the estimation of optical flow: Relations between different approaches and some new results", *Artificial Intelligence*, vol. 33, pp. 299-324, 1987.
- [16] [Oht 90] N. Ohta, "Optical flow detection by color images", *NEC Res. and Dev.*, no. 97, pp. 78-84, Apr. 1990.
- [17] [Sez 93] M.I. Sezan and R.L. Lagendijk, eds., *Motion Analysis and Image Sequence Processing*, Norwell, MA: Kluwer, 1993.
- [18] [Sin 91] A. Singh, *Optic flow computation*, Los Alamitos, CA: IEEE Computer Soc. Press, 1991.
- [19] [Sny 91] M.A. Snyder, "On the mathematical foundations of smoothness constraints for the determination of optical flow and for surface reconstruction", *IEEE Trans. Patt. Anal. Mach. Intel.*, vol. 13, pp. 1105-1114, 1991.
- [20] [Ura 88] S. Uras, F. Girosi, A. Verri, and V. Torre, "A computational approach to motion perception", *Biol. Cybern.*, vol. 60, pp. 79-97, 1988.
- [21] [Ver 89] A. Verri and T. Poggio, "Motion field and optical flow: Qualitative properties", *IEEE Trans. Patt. Anal. Mach. Intel.*, vol. PAMI-11, pp. 490-498, May 1989.
- [22] [Sul 93] G. Sullivan, "Multi-hypothesis motion compensation for low bit-rate video coding", *Proc. IEEE Int. Conf. ASSP*, Minneapolis, MN, vol. 5, pp. 437-440, May 1993.
- [23] [Wol 90] G. Wolberg, *Digital Image Warping*, Los Alamitos, CA: IEEE Computer Soc. Press, 1990.
- [24] [Tho 87] G.A. Thomas and B.A. Hons, "Television motion measurement for DATV and other applications", *Tech. Rep. BBC-RD-1987-11*, 1987.
- [25] [Jac 87] L. Jacobson and H. Wechsler, "Derivation of optical flow using a spatiotemporal frequency approach", *Comp. Vision Graph. Image Proc.*, vol. 38, pp. 57-61, 1987.
- [26] [Hee 87] D.J. Heeger, "Model for the extraction of optical flow", *J. Opt. Soc. Am. A*, vol. 4, no. 8, pp. 1455-1471, Aug. 1987.
- [27] [Cas 87] E. De Castro and C. Morandi, "Registration of translated and rotated images using finite Fourier transforms", *IEEE Trans. Patt. Anal. Mach. Intel.*, vol. 9, no. 5, pp. 700-703, Sep 1987.
- [28] [Bie 87] J. Biemond, L. Looijenga, D.E. Boekee and R.H.J.M. Plompen, "A per-recursive Wiener-based displacement estimation algorithm", *Sign. Proc.*, vol. 13, pp. 399-412, Dec. 1987.
- [29] [Bie 88] J. Biemond, J.N. Driessen, A.M. Geurtz, and D.E. Boekee, "A per-recursive Wiener based algorithm for the simultaneous estimation of rotation and translation", *Proc. SPIE Conf. Visual Commun. Image Proc.*, pp. 917-924, Cambridge MA, 1988.
- [30] [Bie 87] L. Biemond, L. Looijenga, D.E Boekee and R.H.J.M. Plompen, "A pel-recursive Wiener-based displacement estimation alogithm", *Sign. Proc.*, vol. 13, pp. 399-412, Dec. 1987.

- [31] [Bor 91] L. Boroczky, Pel-Recursive Motion Estimation for Image Coding, Ph.D. Thesis, Delft University of Technology, 1991.
- [32] [Caf 83] C. Cafforio and F. Rocca, "The differential method for image motion estimation", in Image Sequence Processing and Dynamic Scene Analysis, T.S. Huang, ed., pp. 104-124, Berlin, Germany:Springer-Verlag, 1983.
- [33] [Dri 91] J.N. Driessen, L. Boroczky, and J. Biemond, "Pel-recursive motion field estimation from image sequences", J. Vis. comm. Image Rep., vol. 2, no. 3, pp. 259-280, 1991.
- [34] [Fle 87] R. Fletcher, Practical Methods of Optimization, Vol. 1, 2nd ed., Chichester, UK: John Wiley and Sons, 1987.
- [35] [Rob 83] J.D. Robbins and a.N. Netravali, "Recursive motion compensation: A review", in Image Sequence Processing and Dynamic Scene Analysis, T.S. Huang, ed., pp. 76-103, Berlin, GermanyQ Springer-Verlag, 1983.
- [36] [Wal 84] D.R. Walker and K.R. Rao, "Improved pel-recursive motion compensation", IEEE Trans. commun., vol. COM-32, pp. 1128-1134, Oct. 1984.