

## Εθνικό Μετσόβιο Πολυτέχνειο

Σχολή Ηλεκτρολογών Μηχανικών και Μηχανικών Υπολογίστων

Τομέας Σηματών, Ελέγχου και Ρομποτικής

Εικονική Ανακατασκευή και Ταυτοποίηση Αντικειμένων στο Χώρο με Τεχνικές Υπολογιστικής Όρασης και Τριδιάστατα Γραφικά

Διδακτορική Διατριβή

του

### ΑΛΗΦΡΑΓΚΗ Ε. ΜΑΤΘΑΙΟΥ

Διπλωματούχου Ηλεκτρολόγου Μηχανικού & Μηχανικού Υπολογιστών

Αθήνα, Δεκέμβριος 2010



Σχολή Ηλεκτρολογών Μηχανικών και Μηχανικών Υπολογιστών ΤΟΜΕΑΣ ΣΗΜΑΤΩΝ, ΕΛΕΓΧΟΥ ΚΑΙ ΡΟΜΠΟΤΙΚΗΣ

#### Εικονική Ανακατασκευή και Ταυτοποίηση Αντικειμένων στο Χώρο με Τεχνικές Υπολογιστικής Όρασης και Τριδιάστατα Γραφικά.

Διδακτορική Διατριβή του ΑΛΗΦΡΑΓΚΗ Ε. ΜΑΤΘΑΙΟΥ

Διπλωματούχου Ηλεκτρολόγου Μηχανικού & Μηχανικού Υπολογιστών

Συμβουλευτική Επιτροπή:

Ομότ. Καθ. Σ. Τζαφέστας Καθ. Π. Μαραγκός Επίκ. Καθ. Κ. Τζαφέστας

#### Επταμελής Εξεταστική Επιτροπή:

..... ..... .... Σ. Τζαφέστας Π.Μαραγκός Κ. Τζαφέστας Ομότ. Καθηγητής ΕΜΠ Καθηγητής ΕΜΠ Επίκ. Καθηγητής ΕΜΠ

. . . . . . . . . Γ. Παπαβασιλόπουλος Καθηγητής ΕΜΠ

..... Σ. Κόλλιας

..... Κ.Παλυβού Καθηγητής ΕΜΠ Καθηγήτρια ΑΠΘ

. . . . . . . . . Γ. Ποταμιάνος Ερευνητής Α', ΕΚΕΦΕ «Δημόκριτος»

Αθήνα, Δεκέμβριος 2010



Η παρούσα διδακτορική διατριδή πραγματοποιήθηκε στα πλαίσια του ερευνητικού προγράμματος ΠΕΝΕΔ-2003, της Γενικής Γραμματείας Έρευνας και Τεχνολογίας. Το πρόγραμμα συγχρηματοδοτήθηκε κατά 80% από την Ευρωπαϊκή Ένωση και 20% από το Ελληνικό Δημόσιο.

This Ph.D. thesis was supported by research grant PENED2003 of the Greek Ministry of Development-GSRT. It is co-financed by E.U.-European Social Fund (80%) and National Resources (20%).

#### ΑΛΗΦΡΑΓΚΗΣ Ε. ΜΑΤΘΑΙΟΣ

Διπλωματούχος Ηλεκτρολόγος Μηχανικός & Μηχανικός Υπολογιστών, ΕΜΠ. Copyright ©Αληφραγκής Ματθαίος, 2010 Με επιφύλαξη παντός δικαιώματος. All rights reserved

Απαγορεύεται η αντιγραφή, αποθήκευση και διανομή της παρούσας εργασίας, εξ ολοκλήρου ή τμήματος αυτής, για εμπορικό σκοπό. Επιτρέπεται η ανατύπωση, αποθήκευση και διανομή για σκοπό μη κερδοσκοπικό, εκπαιδευτικής ή ερευνητικής φύσης υπό την προϋπόθεση να αναφέρεται η πηγή προέλευσης και να διατηρείται το παρόν μήνυμα. Ερωτήματα που αφορούν στη χρήση της εργασίας για κερδοσκοπικό σκοπό πρέπει να απευθύνονται προς το συγγραφέα.

Οι απόψεις και τα συμπεράσματα που περιέχονται σε αυτό το έγγραφο εκφράζουν τον συγγραφέα και δεν πρέπει να ερμηνευτεί ότι αντιπροσωπεύουν τις επίσημες θέσεις του Εθνικού Μετσόβιου Πολυτεχνείου.

#### Ευχαριστίες

Θα ήθελα να ευχαριστήσω θερμά τον επιβλέποντα καθηγητή της παρούσας διατριβής κ. Κωνσταντίνο Τζαφέστα, Επίκουρο Καθηγητή ΕΜΠ για τη διαρκή καθοδήγηση και βοήθεια του σε κάθε φάση της δημιουργίας της. Επίσης τα υπόλοιπα μέλη της συμβουλευτικής επιτροπής κ. Σπυρίδων Τζαφέστα, Ομότιμο Καθηγητή ΕΜΠ και κ. Πέτρο Μαραγκό, Καθηγητή ΕΜΠ για τις υποδείξεις και τις παρατηρήσεις τους στη διάρκεια εκπόνησης της.

Τέλος θέλω να εκφράσω την ευγνωμοσύνη μου στην οικογένεια μου Ευάγγελο, Γεωργία και Σοφία Αληφραγκή καθώς και στη σύντροφο μου Τάσσιου Ιωάννα για τη συνεχή υποστήριξη και συμπαράσταση τους όλα αυτά τα χρόνια.

## Περιεχόμενα

1	Εισ	αγωγή	1
	1.1	Στόχοι της Διατριβής	1
	1.2	Τριδιάστατα Μοντέλα από Εικόνες	3
		1.2.1 Αντιστοίχιση Χαρακτηριστικών	4
		1.2.2 Δομή από Κίνηση	5
		1.2.3 Πυκνή Στέρεο Αντιστοίχιση	6
	1.3	Ερευνητική Συνεισφορά της Διατριβής	7
	1.4	Δομή της Διατριβής	8
I á	Τε <u>;</u> στατ	χνικές Αντιστοίχισης Χαρακτηριστικών και Τριδι τη Ανακατασκευή σε Ζεύγη Εικόνων	- 11
2	Βασ	σικά Μοντέλα και Τεχνικές Βαθμονόμησης Κάμερας	13
	2.1	Το Βασικό Μοντέλο Οπής	13
	2.2	Περιστροφή και Μεταφορά Κάμερας	16
	2.3	Κάμερα με αισθητήρα τύπου CCD	16
	2.4	Βαθμονόμηση Κάμερας	17
	2.5	Βασικές Εξισώσεις	18
		2.5.1 Υπολογισμός Ομοιογραφήματος (homography) μεταξύ Ε-	
		πίπεδου Μοντέλου και της Εικόνας του	19
		2.5.2 Περιορισμοί στις Εσωτερικές Παραμέτρους της Κάμερας	20
	2.6	Επίλυση του προβλήματος Βαθμονόμησης της Κάμερας	21
	2.7	Ακτινική Παραμόρφωση	23
	2.8	Εξοπλισμός	24
	2.9	Αλγόριθμος Βαθμονόμησης	25
	2.10	) Βαθμονόμηση: Μετρήσεις και Αποτελέσματα	26
		2.10.1Εσωτερικές Παράμετροι μετά την Ελαχιστοποίηση του	
		Γεωμετρικού Λάθους	28
		2.10.2Παράμετροι Ακτινικής Παραμόρφωσης	28
		2.10.3Εξωτερικές Παράμετροι	28

	2.11	Ι Συμπεράσματα Κεφαλαίου	28
3	Τρι σικι	διάστατη Ανακατασκευή - Βιβλιογραφική Επισκόπηση - Κλα- ές και Σύγχρονες Προσεγγίσεις	31
	3.1	Επεξεργασία Εικόνων - Ανάκτηση Χαρακτηριστικών - Αρχή της	
		Συσχέτισης	32
		3.1.1 Εντοπισμός Χαρακτηριστικών Ακμών - Επισκόπηση Κλασ-	
		σικών Μεθόδων	32
		3.1.2 Αντιστοίχιση Κύριων Χαρακτηριστικών - Αρχή της Συ-	
		σχέτισης	33
	3.2	Αναλλοίωτα Χαρακτηριστικά Εικόνας	35
		3.2.1 Εντοπισμός Ακρότατων σε Χώρο - Κλίμακας	37
		3.2.2 Εντοπισμός Τοπικών Ακρότατων	39
		3.2.3 Αποκρίσεις Ακμών	43
		3.2.4 Συνάρτηση Τοπικής Περιγραφής Εικόνας	44
		3.2.5 Δομές Περιγραφής Χαρακτηριστικών Σημείων	45
	3.3	Αυτόματη Βαθμονόμηση Κάμερας, Επανόρθωση Εικόνων	48
		3.3.1 Υπολογισμός Θεμελιώδους Πίνακα	49
		3.3.2 Υπολογισμός της Μήτρας Προβολής Κάμερας από τον	
		Ουσιώδη Πίνακα	50
	3.4	Συμπεράσματα - Κεφαλαίου	52
л	Σте	οροσκοπική Αντιστοίνιση και Τοιδιάστατη Ανακατασκουή με	
Ŧ	Xor	οση Μεθόδων Ομοιότητας Φάσης	53
	4.1	Επεξεργασία Εικόνων με Μοντέλα Τοπικής Ενέργειας	53
	42	Εξαγωνή Χαρακτηριστικών με Χρήση Τεχνικών Ομοιότητας Φάσης	56
	1.2	4.2.1 Ανάκτηση Οιμοιότητας Φάσης στις διάφορες κατευθύνσεις	59
		4.2.2 Αντιστοίνιση εικόνων από Τοπική Πληροφορία Συγγότητας	59
	43	Αναλιπικό Σήμα	61
	4.4	Μετασχηματισμός Riesz Μονογενές Σήμα και Στερερακοπική	01
	1.1	Αντιστοίχιση	62
	4.5	Πειοαιματική Αξιολόνηση - Εντοπισμός Χαρακτηριστικών Λομών	
	110	- Ταίριασμα Εικόνων - Αυτοβαθμονόμηση	66
	4.6	Τοιδιάστατη Ψηφιακή Ανακατασκευή	71
	4.7	Επανόρθωση Εικόνων - Αυτόματη Βαθμονόμηση	72
	4.8	Ανακατασκευή Υπάρχουσας Δομής με χρήση Ουσιώδη Πίνακα	_
		(Essential Matrix)	74
	4.9	Συμπεράσματα Κεφαλαίου	76

II	Στοχαστικές	Μέθοδοι	Αντιστοίχισης	και	Ανάκτηση
Πλ	ηροφορίας Δα	ομής σε Α	κολουθίες Εικό	νων	77

5	33 <b>0</b>	ορητικό Υπόβαθρο Στοχαστικών Μεθόδων Ανάκτησης Πληρο	-
	φορ	ρίας Δομής από Εικόνες	79
	5.1	Τυχαία Μαρκοβιανά Μοντέλα	79
	5.2	Ετικετοποίηση στην Ανάλυση Εικόνας	80
	5.3	Μοντελοποίηση Πληροφορίας Πλαισίου με Τυχαία Μαρκοβιανά	
		Μοντέλα	82
	5.4	Πλαίσιο Λήψης Απόφασης MAP-MRF - Τεχνικές Βελτιστοποίησης	; 83
	5.5	Συναρτήσεις Ενέργειας - Βέλτιστες Μέθοδοι Ελαχιστοποίησης.	89
		5.5.1 Κατάτμηση και Χώροι Κίνησης	94
		5.5.2 Βελτιστοποίηση με Αποκοπή Γράφων	95
		5.5.3 Αναζήτηση Βέλτιστης Κίνησης $\alpha - \beta$ Αντιμετάθεσης	97
		5.5.4 Βέλτιστη Κίνηση Διαστολής	100
	5.6	Συμπεράσματα Κεφαλαίου	103
6	Βέλ	τιστη Αντιστοίχιση Εικόνων με Χρήση Πολυδιάστατου Δια	-
	νύσ	ματος Δομής	105
	6.1	Εισαγωγή	105
	6.2	Γενικεύσεις Αναλυτικού Σήματος	105
	6.3	Γεωμετρική Πληροφορία Μονογενούς Σήματος: Θέση Προσα-	
		νατολισμός και Φάση	107
	6.4	Μέθοδοι Βελτιστοποίησης σε Στέρεο Αντιστοίχιση	110
		6.4.1 Ενεργειακό Μοντέλο	110
		6.4.2 Αποκοπή Γράφων	114
	6.5	Στέρεο Αντιστοίχιση - Πειραματική Αξιολόγηση	115
		6.5.1 Εισαγωγή	115
		6.5.2 Πειραματικά Δεδομένα - Αποτελέσματα - Αξιολόγηση	116
		6.5.3 Δείκτης Δομικής Ομοιότητας	127
		6.5.4 Πειραματική Αξιολογήση βασισμένη στο Δεικτή Δομικής	100
	6.6		130
	0.0		132
7	Πολ	υοπτικό Ταίριασμα Εικόνων - Ανάκτηση Δομής από Κίνηση	135
	7.1	Εισαγωγή	135
	7.2	Συνάρτηση Ενέργειας Πολυοπτικής Αντιστοίχισης	136
	7.3	Πειραματική Αξιολόγηση Πολυοπτικής Αποκατάστασης Ανομοι-	
		ότητας Βάθους	139
	7.4	Συμπεράσματα Κεφαλαίου	141

8	Σύν	οψη Διατριβής																					145
	8.1	Συμπεράσματα					•			•				•	•			•	•		•		145
	8.2	Μελλοντικές Κατευθύνσεις	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	148
Bι	6λιο <sup>,</sup>	γραφία																					150

### Βι6λιογραφία

# Κατάλογος Σχημάτων

2.1	Γεωμετρία κάμερας οπής	14
2.2	Σύστημα συντεταγμένων εικόνας-καμερας	15
2.3	Κώνος εσωτερικών παραμέτρων κάμερας	21
2.4	Επίδραση της Ακτινικής Παραμόρφωσης και Διόρθωση	<b>24</b>
2.5	Φωτογραφίες αντικειμένου βαθμονόμησης	27
2.6	Διόρθωση Παραμόρφωσης	29
2.7	Απεικόνιση εξωτερικών παραμέτρων	30
3.1	Συσχέτιση σημείων σε δύο εικόνες	34
3.2	Πυραμιδική αναπαράσταση εικόνας	39
3.3	Εντοπισμός τοπικών ακρότατων	40
3.4	Επιλογή χαρακτηριστικών σημείων	43
3.5	Δομή περιγραφής χαρακτηριστικών σημείων (SIFT)	46
3.6	Λύσεις βαθμονομημένης ανακατασκευής	51
4.1	Πειραματικές φυσικές εικόνες	67
4.2	Αποτελέσματα εντοπισμού ακμών από δύο φυσικές εικόνες	69
4.3	Τριδιάστατη ψηφιακή ανακατασκευή από ζευγάρι πειραματι-	
	κών στέρεο εικόνων	73
4.4	Επανόρθωση εικόνων με αυτόματη βαθμονόμηση	75
5.1	Βασικά είδη συναρτήσεων κόστους	86
5.2	Τυπική και ευρεία κίνηση σε σύνολο ετικετών	95
5.3	Ψευδοκώδικας κίνησης τύπου αντιμετάθεσης/διαστολής	96
5.4	Παράδειγμα γράφου $\mathcal{G}_{lphaeta}$ για 1Δ εικόνα $\ldots$	98
5.5	Αναπαράσταση ιδιοτήτων αποκοπής (cut) ${\mathcal C}$ σε ένα γράφο ${\mathcal G}_{lpha}$ .	100
5.6	Παράδειγμα γράφου $\mathcal{G}_{lpha}$ για 1Δ εικόνα	101
5.7	Αναπαράσταση ιδιοτήτων ελάχιστης αποκοπής (cut) C σε ένα	
	γράφο $\mathcal{G}_{\alpha\beta}$	104
6.1	Σχηματική αναπαράσταση του βασικού διανύσματος χαρακτη-	
	ριστικών σε χωρίο εικόνας γύρω από μία ακμή	108

6.2 6.3	Κατεύθυνση, φάσης και προσανατολισμός σε στέρεο εικόνες Συλλογή πειραματικών εικόνων για επαλήθευση αλγορίθμων στέρεο ταιριάσματος μαζί με τους ιδανικούς χάρτες ανομοιότη-	109
6.4	τας βάθους	117
65	γιση)	118
0.5	σέγγιση)	119
6.6	Πειραματικές φυσικές εικόνες	121
6.7	Συγκριτική εφαρμογή νέας συνάρτησης ενέργειας σε πειραμα- τικές εικόνες με αρχιτεκτονικά θέματα αρχαιολογικού ενδια-	
6.8	φέροντος	123
0.0	ότητας με την τιμή αποκοπής (truncation)	125
6.9	Δυο χαρακτηριστικά παραδείγματα πειραματικών στέρεο εικόνων	
0.10	(Aloe, Rocks)	126
6.10 6.11	Υπολογιστικό σύστημα μέτρησης δείκτη δομικής ομοιότητας Διαγράμματα μεταβολής δείκτη δομικής ποιότητας (SSIM) σε	127
	συνάρτηση με τις τιμές αποκοπής του αθροιζόμενου κόστους σύγκρισης	131
6.12	ΖΧάρτες ανομοιότητας βάθους πειραματικών ομάδων στέρεο ει- κόνων	133
7.1	Γραφική αναπαράσταση του περιορισμού ορατότητας	140
7.2 7.3	Συγκριτικά γραφήματα μεταβολής του δείκτη δομικής ποιότητας Χάρτες ανομοιότητας βάθους πειραματικών στέρεο εικόνων με	142
	μέγιστο δείκτη δομικής ποιότητας SSIM	143

\_\_\_\_\_

# Κατάλογος Πινάκων

5.1	Πίνακας ανάθεσης βαρών στις πλευρές ενός γράφου 99
5.2	Πίνακας ανάθεσης βαρών στις πλευρές ενός γράφου 102
6.1	Συγκριτικός πίνακας RMS σφάλματος
7.1	Συγκριτικός πίνακας RMS σφάλματος χαρτών ανομοιότητας βάθους
	με πολυοπτική αντιστοίχιση

## Περίληψη Διατριβής

Η παρούσα διατριδή εκπονήθηκε στο πλαίσιο ενός ερευνητικού έργου ΠΕ-ΝΕΔ, και αφορά στην ανάπτυξη καινοτόμων μεθόδων πληροφορικής από τα πεδία της ψηφιακής επεξεργασίας σήματος, ανάλυσης εικόνας, τεχνητής νοημοσύνης με όραση υπολογιστών και εικονικής πραγματικότητας, για ψηφιακή αποκατάσταση και ανακατασκευή κτιριακών αρχαιολογικών ευρημάτων από τον προϊστορικό οικισμό της Θήρας. Η διατριδή πραγματοποιήθηκε σε συνεργασία με παράλληλες προσπάθειες από τα πεδία αρχαιολογίας και αρχιτεκτονικής για αναλογική ανακατασκευή με σχεδιαστικές μεθόδους.

Η έρευνα προβλημάτων της υπολογιστικής όρασης αποτελεί βασικό προσανατολισμό της παρούσας διατριβής. Το αξιόπιστο ταίριασμα χαρακτηριστικών από διαδοχικές εικόνες, η ακριβής μοντελοποίηση και ο υπολογισμός των οπτικών χαρακτηριστικών της εικόνας, καθώς και η τρισδιάστατη ψηφιακή αποτύπωση της σκηνής, αξιοποιώντας την ανακτώμενη μετρική πληροφορία, αποτελούν τομείς μελέτης και έρευνας της εν λόγω διατριβής.

Μία σημαντική κατηγορία μεθόδων στερεοσκοπικής αντιστοίχισης αποτελούν οι μέθοδοι συσχέτισης φωτεινότητας εικονοστοιχείων. Στην παρούσα διατριβή προτείνεται ένα νέο μέτρο συσχέτισης χαρακτηριστικών που βασίζεται στην τοπική αναπαράσταση της πληροφορίας της εικόνας μέσω του μονογενούς σήματος. Αξιοποιήθηκαν η ενεργειακή, η δομική και η γεωμετρική πληροφορία του μονογενούς σήματος στην κατασκευή του μέτρου συσχέτισης από εικόνα σε εικόνα καθώς επίσης και φίλτρα ομοιότητας φάσης στον εντοπισμό χαρακτηριστικών γεωμετρικών δομών στις εικόνες.

Η συγκριτική πειραματική μελέτη της τριδιάστατης αποτύπωσης που προκύπτει από τη χρήση του προτεινόμενου μέτρου σε σχέση τόσο με κλασικές μεθόδους συσχέτισης φωτεινότητας, όσο και με πιο σύγχρονες μεθόδους αντιστοίχισης αναλλοίωτων χαρακτηριστικών σε κλίμακα και περιστροφή της εικόνας, διενεργήθηκε σε φυσικές εικόνες κάτω από τυχαίες συνθήκες φωτισμού με πολύπλοκη απεικονιζόμενη δομή αρχαιολογικού ενδιαφέροντος. Χρησιμοποιήθηκαν στέρεο εικόνες από το Ερεχθείο στην Ακρόπολη και από το Ηρώδειο. Η εφαρμογή του προτεινόμενου μέτρου στερεοσκοπικής συσχέτισης οδήγησε σε περισσότερο αξιόπιστα αποτελέσματα τριδιάστατης ψηφιακής αποτύπωσης και υπολογισμού των οπτικών παραμέτρων της εικόνας.

Τέλος, στην προσπάθεια ανάκτησης πληροφορίας βάθους από όλο το επίπεδο της εικόνας, αξιοποιήθηκαν Μπεϋζιανά μοντέλα βέλτιστης απόφασης. Σε αυτό το πλαίσιο, το πρόβλημα στερεοσκοπικής αντιστοίχισης εκφράζεται ως ένα πρόβλημα ελαχιστοποίησης συναρτήσεων κόστους. Στην παρούσα διατριβή προτείνεται μία τροποποιημένη μορφή συνάρτησης ενέργειας στερεοσκοπικής συνέπειας ή κόστους αντιστοίχισης που συνδυάζει σε ένα ενιαίο μέτρο ομοιότητας την τοπική γεωμετρική, δομική και ενεργειακή πληροφορία της εικόνας, όπως προκύπτουν από την αναπαράσταση του μονογενούς σήματος. Η συγκριτική αξιολόγηση των αποτελεσμάτων επιβεβαιώνει το γεγονός ότι η χρήση της προτεινόμενης συνάρτησης ενέργειας κόστους οδηγεί στον υπολογισμό τιμών βάθους με μικρότερα επίπεδα σφάλματος. Παράλληλα, στη διαδικασία αξιολόγησης της επίδοσης της προτεινόμενης συνάρτησης κόστους, προτείνεται η χρήση του δείκτη δομικής ομοιότητας (Structural SIMilarity index, SSIM index) που βασίζεται στην ανθρώπινη αντιληπτική ικανότητα. Στην παρούσα διατριβή προτείνεται η χρήση του δείκτη αυτού ως μέτρου ομοιότητας των υπολογιζόμενων τιμών βάθους με τις αντίστοιχες ιδανικές, γεγονός που παρέχει ουσιαστικά ένα μέτρο εκτίμησης της ποιότητας στερεοσκοπικής αποτύπωσης.

## Κεφάλαιο 1

## Εισαγωγή

## 1.1 Στόχοι της Διατριβής

Η τριδιάστατη αποκατάσταση, αναπαράσταση και μοντελοποίηση απεικονιζόμενων σκηνών αποτελεί εδώ και δεκαετίες επίκεντρο πληθώρας ερευνητικών εργασιών στο πεδίο της υπολογιστικής όρασης. Παραδοσιακές τεχνολογικές εφαρμογές ρομποτικής, τεχνητής νοημοσύνης αποτελούν βασικά παραδείγματα αξιοποίησης τριδιάστατων γραφικών με υπολογιστές. Η ακρίβεια στη τριδιάστατη γραφική μοντελοποίηση και ανακατασκευή αποτελεί ένα από τα κύρια ζητούμενα. Πολλές φορές κλασικές μέθοδοι χρησιμοποιούνται με ικανοποιητικά αποτελέσματα σε αυστηρά όμως ελεγχόμενο και δομημένο περιβάλλον.

Στις μέρες μας, καθώς οι επιστημονικές κοινότητες της υπολογιστικής όρασης και των γραφικών με υπολογιστές αναπτύσσονται, το ενδιαφέρον για εφαρμογές ψηφιακής τριδιάστατης ανακατασκευής και ταυτοποίησης μεγαλώνει. Στη διαδικασία αυτή, βέβαια, συναντώνται πολλά εμπόδια που σχετίζονται με την εγγενή δυσκολία να ανακατασκευαστεί τριδιάστατα ένας σύνθετος κόσμος. Η απαίτηση για ρεαλιστική ανακατασκευή και απεικόνιση του τριδιάστατου κόσμου είναι πολλές φορές υπολογιστικά απαγορευτική, και τα προκύπτοντα αποτελέσματα είναι συχνά μη αποδεκτά, με αποσπασματική απεικόνιση του πραγματικού κόσμου.

Πρόσφορο έδαφος βρίσκει η απαίτηση για συστήματα που να μπορούν να απεικονίζουν τριδιάστατα συγκεκριμένα θέματα αξιοποιώντας ένα σύνολο διδιάστατων φυσικών εικόνων. Το σημαντικό στοιχείο και ζητούμενο μιας τέτοιας απαίτησης είναι η εκλαμβανόμενη ποιότητα και ακρίβεια των ανακατασκευασμένων τριδιάστατων μοντέλων. Η ικανοποίηση μίας τέτοιας απαίτησης οδηγεί σε λεπτομερείς τριδιάστατες απεικονίσεις ενός συγκεκριμένου θέματος, το οποίο έχει φωτογραφηθεί πολλαπλές φορές και από διάφορες οπτικές γωνίες. Η τεχνική αυτή είναι αρκετά απλή, αφού δε χρειάζεται κάτι πιο ακριβό ή δύσχρηστο από μία φωτογραφική μηχανή και ένα σύστημα στήριξης (τρίποδο).

Η παρούσα διατριβή εντάσσεται στο ευρύτερο πλαίσιο των στόχων του έργου ΠΕΝΕΔ βασικό μέρος των οποίων συνίσταται στην αρχιτεκτονική αποκατάσταση και ανακατασκευή κτιριακών δομών στον προϊστορικό οικισμό στο Ακρωτήρι Σαντορίνης.

Στην παρούσα διατριδή επιχειρείται αρχικά η ανασκόπηση μεθόδων αυτόματης τριδιάστατης απεικόνισης και αποκατάστασης σύνθετων φωτογραφημένων θεμάτων. Βασική ομάδα φωτογραφημένων θεμάτων που θα χρησιμοποιηθεί για την αποτίμηση και τη συγκριτική αξιολόγηση των μεθόδων είναι αρχιτεκτονικές δομές αρχαιολογικού ενδιαφέροντος. Τα θέματα αυτά συχνά παρουσιάζουν μία αυξημένη πολυπλοκότητα στη δομή με αρκετά ελλιπή μέρη, λόγω φυσικής φθοράς του πραγματικού μνημείου. Συνεπώς, η φωτογράφηση τους οδηγεί σε εικόνες αρκετά δύσκολες στην επεξεργασία με πολλά στοιχεία υφής, χωρίς διακριτές γεωμετρικές δομές (ακμές, γωνίες), και δυσδιάκριτα όρια των αντικειμένων εντός της εικόνας. Η εξαγωγή της μέγιστης αξιοποιήσιμης πληροφορίας από απλές φωτογραφίες σε σχέση τόσο με τα δομικά τους στοιχεία και την ταυτοποίησή τους στο χώρο, όσο και με τα μετρικά χαρακτηριστικά τους, συμβάλλει σημαντικά σε παράλληλες αρχαιολογικές και αρχιτεκτονικές μελέτες.

Οι χώροι αυτοί πολλές φορές είναι δύσκολα προσβάσιμοι από τον άνθρωπο ή η χρήση εξειδικευμένου εξοπλισμού όπως τριδιάστατοι σαρωτές παράγει πολλές φορές φτωχά αποτελέσματα λόγω του μεγάλου βαθμού σκέδασης σε συγκεκριμένα υλικά. Η φωτογράφησή τους αποτελεί το μόνο τρόπο για την ψηφιακή αποτύπωσή τους και την επακόλουθη διαδικασία εξαγωγής οπτικών, δομικών χαρακτηριστικών τους, αξιοποιώντας μεθόδους υπολογιστικής όρασης.

Η αναπαράσταση της τριδιάστατης δομής μίας απεικονιζόμενης σκηνής γίνεται αφενός με την αποκατάσταση και ταυτοποίηση βασικών οπτικών χαρακτηριστικών της εικόνας, όπως ακμών, και αφετέρου με τον υπολογισμό του σχετικού βάθους για όλα τα εικονοστοιχεία, κατασκευάζοντας πυκνούς χάρτες ανομοιότητας βάθους (disparity maps). Στην πρώτη περίπτωση ανακτάται επιπλέον και η μετρική πληροφορία των γεωμετρικών χαρακτηριστικών της σκηνής με σημείο αναφοράς τη θέση της κάμερας, είτε με χρήση του πίνακα βαθμονόμησης της κάμερας (εξωτερικές και εσωτερικές παραμέτρους), είτε με χρήση μεθόδων αυτόματης βαθμονόμησης. Βασική παραδοχή στα πειράματά μας αποτελεί η διατήρηση σταθερών παραμέτρων της κάμερας. Το γεγονός αυτό σημαίνει ότι δεν αξιοποιούνται δυνατότητες μεταβλητής εστίασης στην κάμερα (zoom).

Στη δεύτερη περίπτωση υλοποιείται η πυκνή αποκατάσταση βάθους μέσω

χαρτών ανομοιότητας βάθους. Στέρεο ανομοιότητα βάθους (disparity), καλείται η απόλυτη διαφορά της θέσης ενός εικονοστοιχείου σε ένα στέρεο ζευγάρι εικόνων. Το μέτρο της απόστασης αυτής είναι αντιστρόφως ανάλογο του βάθους στο εν λόγω σημείο της εικόνας [49].

Ένας επιπλέον στόχος της παρούσης διατριβής είναι να επιβεβαιώσει, καθ'όλη τη διαδικασία της τριδιάστατης αποκατάστασης και αποτύπωσης, την αξιοπιστία τεχνικών, που βασίζονται σε στατιστικά μοντέλα αποφάσεων [29]. Η εφαρμογή τέτοιων μεθόδων, ειδικά σε προβλήματα αντιστοίχισης εικονοστοιχείων ή χαρακτηριστικών, οδηγεί σε εμφανή βελτίωση των αποτελεσμάτων.

Τέλος, στόχο αποτελεί η συγκριτική μελέτη των αποτελεσμάτων που προκύπτουν αφενός από προτεινόμενες προσεγγίσεις επεξεργασίας και ψηφιακής τριδιάστατης ανακατασκευής και αφετέρου από τις αντίστοιχες κλασικές μεθόδους που βασίζονται κυρίως στην επισφαλή, όπως αποδεικνύεται, πληροφορία φωτεινότητας.

### 1.2 Τριδιάστατα Μοντέλα από Εικόνες

Παρακάτω παρουσιάζεται μία σύντομη επισκόπηση θεμάτων σχετικών με αυτά που εξετάζονται στην παρούσα διατριβή.

Είναι γεγονός ότι ο υπολογισμός τριδιάστατων μοντέλων από σειρά εικόνων, χωρίς καμία επιπρόσθετη πληροφορία σχετική με τα γεωμετρικά χαρακτηριστικά κάμερας και της απεικονιζόμενης σκηνής, αποτελεί μία εξαιρετικά δύσκολη εργασία. Γεγονός είναι ότι σε μία ρεαλιστική πειραματική διάταξη, τόσο η γεωμετρία της σκηνής, όσο και η γεωμετρία της κάμερας παραμένουν άγνωστα. Μόνο γενικές υποθέσεις μπορούν να διατυπωθούν, όπως για παράδειγμα, τμηματικά συνεχείς επιφάνειες με διαχεόμενη φωτεινή ανάκλαση για τη φωτογραφημένη σκηνή ή μοντέλο οπής για την κάμερα. Βασική μεθοδολογία επίλυσης του προβλήματος συνιστά ο διαχωρισμός του αρχικού προβλήματος σε υποπροβλήματα που επιλύονται σε ξεχωριστές ενότητες [62].

- Εντοπισμός ακμών ή άλλων γεωμετρικών διακριτών δομών.
- Ορθοποίηση εικόνων (rectification) με στόχο την απλοποίηση του προβλήματος αντιστοίχισης. Τα αντίστοιχα σημεία των ορθοποιημένων εικόνων κινούνται σε παράλληλες ευθείες (επιπολικές γραμμές) κατά μήκος της εικόνας.
- Αντιστοίχιση εντοπισμένων χαρακτηριστικών.
- Ανάκτηση βάθους των χαρακτηριστικών, με τριγωνοποίηση ή χάρτες ανομοιότητας βάθους.

 Ανάκτηση εσωτερικών παραμέτρων (οπτικά χαρακτηριστικά) και εξωτερικών παραμέτρων (τριδιάστατη θέση) κάμερας.

Συχνά απαιτείται ανατροφοδότηση αποτελεσμάτων από κάθε υποπρόβλημα προκειμένου να εξαχθούν απαραίτητα συμπεράσματα σχετικά με βασικά γεωμετρικά και μετρικά χαρακτηριστικά των εικόνων ή της κάμερας. Με αυτόν τον τρόπο ανακτάται σταδιακά μεγαλύτερος όγκος πληροφορίας. Επομένως, στην τελική φάση, η σύζευξη της τμηματικής πληροφορίας οδηγεί σε βελτιωμένα αποτελέσματα.

#### 1.2.1 Αντιστοίχιση Χαρακτηριστικών

Το αρχικό και κομβικό πρόβλημα που καλείται να αντιμετωπίσει κάθε ερευνητική εργασία σε αυτόν τον τομέα είναι το πρόβλημα της αντιστοίχισης (matching) από διαφορετικές εικόνες. Δοσμένου ενός αρχικού χαρακτηριστικού σε μία εικόνα (ακμή, καμπύλη, γωνία κτλ), το πρόβλημα αυτό συνίσταται στην αναζήτηση σε άλλη εικόνα της αντίστοιχης θέσης του χαρακτηριστικού αυτού. Τα αντίστοιχα χαρακτηριστικά αποτελούν προβολές ενός κοινού τριδιάστατου χαρακτηριστικού, το οποίο στη συνέχεια μπορεί να αναπαρασταθεί γραφικά. Εν γένει, το πρόβλημα αυτό είναι ασθενώς ορισμένο και στερείται μοναδικής λύσης. Είναι αναγκαίο λοιπόν να γίνουν κάποιες υποθέσεις, ώστε να καταστεί δυνατή η αυτόματη επίλυσή του. Μία από τις απαραίτητες υποθέσεις είναι ότι οι εικόνες δεν παρουσιάζουν μεγάλες διαφορές αναφορικά με την μέση φωτεινότητα ή τη διάταξη της σκηνής. Σε αυτήν την περίπτωση, τόσο οι συντεταγμένες των χαρακτηριστικών, όσο και η κατανομή της φωτεινότητας σε γειτονική περιοχή του εντοπισμένου χαρακτηριστικού παρουσιάζουν σημαντικές ομοιότητες για το σύνολο των υπό εξέταση εικόνων. Η υιοθέτηση της υπόθεσης αυτής απαιτεί την ομαλή κίνηση της κάμερας μπροστά από το θέμα αποφεύγοντας πιθανές αλληλοκαλύψεις (occlusions). Η χρήση των παραπάνω υποθέσεων στην εξέλιξη της ανάλυσης συντελεί στη δραστική μείωση του εύρους αναζήτησης αντίστοιχων εικονοστοιχείων και στην αντιστοίχιση με διασταυρωμένη συσχέτιση φωτεινότητας που οδηγεί σε περισσότερο αξιόπιστες αντιστοιχίσεις, με μικρότερο υπολογιστικό κόστος.

Είναι σαφές ότι όλα τα πιθανά χαρακτηριστικά της εικόνας δεν μπορεί να θεωρηθούν κατάλληλα για αντιστοίχιση. Μεμονωμένα σημεία χρησιμοποιούνται ως πληροφορία άμεσα και εύκολα αξιοποιήσιμη, ενώ συχνά επίσης συναντάται με μεγάλη επιτυχία και η χρήση γεωμετρικών δομών όπως τμήματα ευθείας [12], [60] ή ολόκληρων περιοχών για εύρεση αντιστοιχίσεων από εικόνα σε εικόνα. Η χρήση υποψήφιων προς αντιστοίχιση χαρακτηριστικών εικονοστοιχείων από ομογενείς ως προς τη φωτεινότητα περιοχές της εικόνας χωρίς να υπάρχει τρόπος διαφοροποίησής τους, οδηγεί σε μη αποδεκτά αποτελέσματα όσον αφορά τον υπολογισμό του βάθους των σημείων αυτών. Ανακύπτει λοιπόν η ανάγκη χρήσης ενός φίλτρου εντοπισμού εικονοστοιχείων κατάλληλων προς αντιστοίχιση. Η αναζήτηση αυτή επιτυγχάνεται ικανοποιώντας δύο βασικά κριτήρια όσον αφορά τη διαδικασία εξαγωγής συγκεκριμένων σημείων στις εικόνες, η οποία πρέπει να είναι:

- 1. Ανεξάρτητη από τη θέση της κάμερας και
- 2. Ανεξάρτητη από τις πιθανές διακυμάνσεις φωτεινότητας στην γειτονιά τους. Πληθώρα φίλτρων εντοπισμού χαρακτηριστικών σημείων ή περιοχών σε μια εικόνα συναντάται στη βιβλιογραφία (για παράδειγμα στα [59] και [11]). Οι Schmid et al [13] κατέληξαν ότι το φίλτρο εντοπισμού γωνιών του Harris καταλήγει σε βέλτιστα αποτελέσματα αναφορικά με τα παραπάνω κριτήρια.

Οι εικόνες αρχαιολογικού ενδιαφέροντος συχνά εμφανίζουν μεγάλο βαθμό πολυπλοκότητας στη δομή και αρκετά μεγάλη διακύμανση στη φωτεινότητα, ειδικά εάν οι φωτογραφίες απεικονίζουν σκηνές εξωτερικού χώρου χωρίς καμία δυνατότητα ελέγχου του φωτισμού. Κάτω από αυτές τις συνθήκες μεγάλη επιτυχία τόσο στον ακριβή εντοπισμό κατάλληλων χαρακτηριστικών, όσο και στην αντιστοίχισή τους επέδειξαν φίλτρα συσχέτισης φάσης [55], [54], όπως αυτά παρουσιάζονται στην παρούσα διατριβή.

Η διαδικασία της αντιστοίχισης είναι στενά συνδεδεμένη επίσης με το θεμελιώδες πρόβλημα υπολογισμού δομής από την κίνηση της κάμερας. Ζευγάρια αντίστοιχων χαρακτηριστικών οδηγούν στον υπολογισμό της τριδιάστατης γεωμετρίας τόσο της κάμερας όσο και της απεικονιζόμενης σκηνής.

#### 1.2.2 Δομή από Κίνηση

Ερευνητές έχουν παρουσιάσει τα τελευταία χρόνια πληθώρα εργασιών σχετικών με την αυτόματη εξαγωγή της τριδιάστατης δομής από αλληλουχία εικόνων. Το πρόβλημα αυτό καλείται υπολογισμός δομής από κίνηση και μπορεί να ορισθεί ως εξής: Δοσμένης μίας αλληλουχίας εικόνων μίας σκηνής, όπως αυτές παράγονται από την άγνωστη κίνηση στο χώρο μίας κάμερας, ζητείται να ανακατασκευαστεί ψηφιακά η τριδιάστατη γεωμετρία της σκηνής. Η επίτευξη αυτού του στόχου προϋποθέτει την ταυτόχρονη ανάκτηση της γεωμετρίας της κάμερας. Στο αρχικό στάδιο της παρούσας διατριβής αξιοποιήθηκε ζευγάρι στέρεο εικόνων όπως προέκυψε από μία ευθύγραμμη κίνηση της κάμερας είτε με γνωστές, είτε με άγνωστες παραμέτρους. Διαπιστώθηκε αμέσως η στενή σχέση της ποιότητας των αποτελεσμάτων με το είδος του ακολουθούμενου αλγορίθμου στην αντιστοίχιση χαρακτηριστικών. Στην παρούσα διατριβή προτείνεται μια νέα μέθοδος αντιστοίχισης η οποία βασίζεται σε επεξεργασία και μετασχηματισμό των ψηφιακών δεδομένων στο χώρο της φάσης έτσι ώστε να ελαχιστοποιηθεί η επιρροή εξωτερικών τυχαίων συνθηκών φωτισμού [46]. Το σύνολο των αντίστοιχων χαρακτηριστικών της εικόνας αποτελούν τα απαραίτητα δεδομένα για την αυτο-βαθμονόμηση της κάμερας και την επακόλουθη ανάκτηση της μετρικής πληροφορίας της σκηνής. Χαρακτηριστικές εργασίες προς αυτή την κατεύθυνση είναι των Spetsakis και Aloimonos [45], Szeliski και Kang [14].

#### 1.2.3 Πυκνή Στέρεο Αντιστοίχιση

Η επίλυση του προβλήματος ανάκτησης δομής από κίνηση οδηγεί σε εξαγωγή ενός πλήθους τριδιάστατων χαρακτηριστικών, που μπορεί να οδηγήσει παρά την πιθανή μετρική ακρίβεια των χαρακτηριστικών, σε μία αποσπασματική εικόνα της πραγματικότητας και σε ατελή τριδιάστατα μοντέλα.

Η χαμηλή εκλαμβανόμενη ποιότητα της τριδιάστατης αποτύπωσης, μπορεί να διορθωθεί με την προσθήκη στα δεδομένα επιπρόσθετων εικόνων από την υπό εξέταση σκηνή. Σε συνδυασμό λοιπόν με τη γνωστή πλέον γεωμετρία της κάμερας, είναι εφικτή η ανάκτηση ζευγαριών αντιστοίχισης σε όλη την επιφάνεια της εικόνας. Η προσέγγιση αυτή οδηγεί στον υπολογισμό τριδιάστατου μοντέλου επιφάνειας [43].

Οι επιστημονικές προσεγγίσεις στο πρόβλημα ποικίλουν. Μία αρχική τους κατηγοριοποίηση είναι σε τεχνικές βασισμένες σε χαρακτηριστικά στοιχεία ή σε τεχνικές συσχέτισης (correlation based techniques). Προσεγγίσεις βασισμένες σε χαρακτηριστικά εικόνων έχουν προταθεί από τους Marr και Poggio [20], Polard, Mayhew και Frisby [73] (μέθοδοι αριθμητικού *relaxation*), Ohta και Kanade [82] (μέθοδοι δυναμικού προγραμματισμού). Επιτυχημένες τεχνικές συσχέτισης έχουν προταθεί από τους Okutomi και Kanade [41] και Cox et al. [32]. Η τελευταία έχει δεχθεί σημαντικές βελτιώσεις από τους Koch [65] και Falkenhagen [1].

Μία επιπρόσθετη κατηγορία αλγορίθμων αντιστοίχισης συνίσταται στην εφαρμογή στατιστικών μοντέλων απόφασης. Κάθε υποψήφιο προς αντιστοίχιση σημείο χαρακτηρίζεται από μία πιθανότητα επιτυχούς ταιριάσματος. Η πιθανότητα αυτή αναλαμβάνει το ρόλο μίας τιμής κόστους σε ένα ευρύτερο γράφο υποψήφιων αντιστοιχιών. Το πρόβλημα λοιπόν της αντιστοίχισης μετατρέπεται σε ένα πρόβλημα εύρεσης ελάχιστου γράφου από πλευράς κόστους, [81]. Ένας τέτοιος γράφος περιέχει συστοιχίες αντίστοιχων σημείων.

Μία αρχική προσέγγιση στην τριδιάστατη αποτύπωση αποτελεί η χρήση χαρτών ανομοιότητας βάθους, οι οποίοι δίνουν μια διδιάστατη απεικόνιση του βάθους μέσω της απεικόνισης των τιμών βάθους σε κλίμακα του γκρι. Οι χάρτες ανομοιότητας βάθους αποτελούν επίσης ένα ιδιαίτερα εύχρηστο εργαλείο τόσο για την ανάκτηση της δομής της σκηνής όσο και την αξιολόγηση των ακολουθούμενων αλγορίθμων αντιστοίχισης.

## 1.3 Ερευνητική Συνεισφορά της Διατριβής

Πριν την αναλυτική διερεύνηση όλων των παραπάνω θεμάτων είναι αναγκαία η παρουσίαση των βασικών καινοτομιών της παρούσας διατριβής. Οι βασικές επιστημονικές καινοτομίες που συνιστούν την ερευνητική συμβολή της παρούσας διατριβής είναι οι ακόλουθες.

- Εισαγωγή φίλτρωυ σύγκρισης φάσης με κύριο στόχο τον εντοπισμό αξιόπιστων χαρακτηριστικών σε σύνθετες εικόνες εξωτερικού χώρου με τυχαίες συνθήκες σκίασης και φωτισμού.
- Ανάπτυξη μίας δομημένης διαδικασίας τριδιάστατης ψηφιακής αποτύπωσης (ανακατασκευής) από στέρεο εικόνες, βασισμένη σε αποκλειστική επεξεργασία των εικόνων στο χώρο της φάσης με την εισαγωγή ενός τροποποιημένου μέτρου συσχέτισης χαρακτηριστικών βασισμένου σε μονογενή φίλτρα. Η ψηφιακή αποτύπωση έγινε με χρήση είτε βαθμονομημένης κάμερας, είτε μη βαθμονομημένης, με εφαρμογή στην τελευταία μεθόδων αυτοβαθμονόμησης. Η εργασία αυτή είναι δημοσιευμένη στο [6].
- Η εφαρμογή στατιστικών μοντέλων στο πρόβλημα αντιστοίχισης, τροποποιώντας κλασικές μορφές συναρτήσεων ενέργειας βασισμένων στη φωτεινότητα της εικόνας σε πιο αξιόπιστες μορφές βασισμένες σε μονογενή φίλτρα. Επιπρόσθετα, επισημαίνεται η υλοποίηση πλατφόρμας λογισμικού για την εύκολη προσθήκη εικόνων στο πλαίσιο τριδιάστατης ψηφιακής αποκατάστασης από αλληλουχία εικόνων.
- Συγκριτική αποτίμηση των αποτεβεσμάτων χρησιμοποιώντας χάρτες ανομοιότητας όπως υπολογίζονται με κλασικές τεχνικές και με το προτεινόμενο μέτρο σύγκρισης, [7]. Η διαδικασία αυτή έχει ιδιαίτερο ερευνητικό ενδιαφέρον σε εξαιρετικά σύνθετες φυσικές εικόνες, όπως αυτές αρχαιολογικών μνημείων. Στο σημείο αυτό βασική συνεισφορά της παρούσας διατριβής αποτελεί η χρήση του δείκτη δομικής ομοιότητας εικόνων που βασίζεται στην ανθρώπινη αντιληπτική ικανότητα [88]. Ο δείκτης αυτός θα χρησιμοποιηθεί ως μέτρο δομικής ομοιότητας των υπολογιζόμενων χαρτών ανομοιότητας βάθους με τους αντίστοιχους ιδανικούς αναλαμβάνοντας το ρόλο ενός εργαλείου μέτρησης ποιότητας των αποτελεσμάτων.

## 1.4 Δομή της Διατριβής

Η παρούσα διατριβή αποτελείται από δύο βασικά μέρη τα οποία συνιστούν δύο διαφορετικής φιλοσοφίας προσεγγίσεις στο πρόβλημα τριδιάστατης ψηφιακής ανακατασκευής από φυσικές εικόνες. Το πρώτο μέρος αφορά την ανάλυση και εφαρμογή μεθόδων συσχέτισης χαρακτηριστικών είτε αξιοποιώντας την πληροφορία φωτεινότητας στην εικόνα σύμφωνα με την κλασική προσέγγιση είτε μέσω μετασχηματισμών στο χώρο της φάσης με εφαρμογή ενός προτεινόμενου μέτρου συσχέτισης βασισμένου στην τοπική γεωμετρική, δομική και ενεργειακή πληροφορία όπως εξάγονται από την αναπαράσταση της πληροφορίας μέσω του μονογενούς σήματος.

Το πρώτο μέρος της διατριδής αποτελείται από τα κεφάλαια δύο, τρία και τέσσερα. Η δομή και το περιεχόμενο των συγκεκριμένων κεφαλαίων του πρώτου μέρους της διατριδής ακολουθεί:

Στο δεύτερο κεφάλαιο γίνεται μία εισαγωγή στα χρησιμοποιούμενα μοντέλα κάμερας, καθώς και στις κλασικές μεθόδους βαθμονόμησης, ορθοποίησης και διόρθωσης λοξότητας του γεωμετρικού μοντέλου της κάμερας. Στο τρίτο κεφάλαιο πραγματοποιείται μία εκτενής έρευνα και μελέτη μεθόδων τριδιάστατης ψηφιακής ανακατασκευής με κλασικές μεθόδους τριγωνοποίησης. Μελετάται επίσης ο μετασχηματισμός των χαρακτηριστικών της εικόνας με αναλλοίωτες σε κλίμακα και περιστροφή συναρτήσεις περιγραφής (SIFT) [21]. Στο τέταρτο κεφάλαιο παρουσιάζεται το θεωρητικό υπόβαθρο των τεχνικών σύγκρισης με τεχνικές ομοιότητας φάσης και τοπικής αναπαράστασης με μονογενή σήματα. Προτείνεται η χρήση ενός τροποποιημένου μέτρου συσχέτισης χαρακτηριστικών στο πρόβλημα της στερεοσκοπικής αντιστοίχισης, βασισμένο στη μονογενή αναπαράσταση της τοπικής πληροφορίας της εικόνας. Πραγματοποιείται επίσης η υλοποίηση και η συγκριτική αποτίμηση των παραπάνω προσεγγίσεων σε πειραματικές εικόνες εξωτερικού χώρου.

Το δεύτερο μέρος της διατριβής αποτελείται από τα κεφάλαια πέντε, έξι και επτά. Στο πέμπτο κεφάλαιο συντελείται μία ερευνητική μελέτη στις στοχαστικές περιγραφικές μεθόδους επίλυσης του στέρεο ταιριάσματος. Πραγματοποιείται λοιπόν μία βιβλιογραφική έρευνα για τα Μαρκοβιανά τυχαία μοντέλα και τις Μπεϋζιανές μεθόδους βελτιστοποίησης γράφων αναπαράστασης ετικετών. Στο έκτο κεφάλαιο μελετάται και προτείνεται η χρήση πολυδιάστατων διανυσμάτων γεωμετρικών και δομικών χαρακτηριστικών βασισμένων σε μονογενή φίλτρα. Τα διανύσματα αυτά συντελούν στην κατασκευή κατάλληλων συναρτήσεων ενέργειας κόστους αντιστοίχησης. Η ελαχιστοποίηση αυτών εξασφαλίζει τη βέλτιστη απόφαση ταιριάσματος χαρακτηριστικών σε μία στέρεο διάταξη. Η απόδοση της προτεινόμενης συνάρτησης κόστους αξιολογείται συγκριτικά με κλασσικές μεθόδους σε δύο κατηγορίες εικόνων:

- Σε πειραματικές εικόνες αξιολόγησης αλγορίθμων υπολογιστικής όρασης, διαθέσιμες στη διεθνή βιβλιογραφία.
- Σε ένα σύνολο στέρεο εικόνων από πραγματικές λήψεις εξωτερικού χώρου θεμάτων με πολύπλοκη δομή αρχαιολογικού και αρχιτεκτονικού ενδιαφέροντος.

Στο κεφάλαιο αυτό προτείνεται επίσης και ένα εναλλακτικό μέτρο εκτίμησης της ποιότητας της αποκατεστημένης δομικής πληροφορίας. Στο έβδομο κεφάλαιο παρουσιάζονται τα αποτελέσματα της παραπάνω ανάλυσης με αξιοποίηση της πληροφορίας περισσοτέρων των δύο στέρεο εικόνων, υλοποιώντας με τις προτεινόμενες μεθόδους και αλγορίθμους μία πρώτη προσέγγιση στο ζήτημα της πολυοπτικής στερεοσκοπικής αντιστοίχισης. Στο τελευταίο κεφάλαιο συγκεντρώνονται τα συμπεράσματα της διατριβής και καθορίζονται οι δυνατές μελλοντικές επεκτάσεις του παρόντος ερευνητικού έργου.

## Μέρος Ι

# Τεχνικές Αντιστοίχισης Χαρακτηριστικών και Τριδιάστατη Ανακατασκευή σε Ζεύγη Εικόνων

## Κεφάλαιο 2

## Βασικά Μοντέλα και Τεχνικές Βαθμονόμησης Κάμερας

Η λειτουργία μιας κάμερας μοντελοποιείται συνήθως ως μία αντιστοίχιση ανάμεσα στον τριδιάστατο κόσμο (χώρο αντικειμένων) και στη διδιάστατη εικόνα. Η ανάλυση των περισσότερων μοντέλων κάμερας γίνεται με τη βοήθεια εργαλείων προβολικής γεωμετρίας. Έτσι, όπως θα δούμε και στη συνέχεια, γεωμετρικές οντότητες της κάμερας όπως το κέντρο προβολής και το επίπεδο της εικόνας, μπορούν να υπολογιστούν πολύ εύκολα από την αλγεβρική τους αναπαράσταση σε πίνακες. Υπάρχουν δύο βασικές οικογένειες προβολικών καμερών: αυτές με πεπερασμένο κέντρο και αυτές που έχουν το κέντρο στο «άπειρο». Για τις κάμερες «στο άπειρο», οι αφινικές (affine) κάμερες αποτελούν το πιο σημαντικό αντιπρόσωπο αυτού του είδους. Στην παρούσα διατριβή θα ασχοληθούμε με το απλούστερο μοντέλο κάμερας, το μοντέλο οπής (pinhole model). Τα μαθηματικά και αλγεβρικά μοντέλα που θα χρησιμοποιηθούν είναι σχεδιασμένα κυρίως για αισθητήρες τύπου CCD.

### 2.1 Το Βασικό Μοντέλο Οπής

Αρχικά θεωρούμε κεντρική προβολή των σημείων του χώρου στο επίπεδο. Το κέντρο της προβολής ονομάζεται και κέντρο της κάμερας ή οπτικό κέντρο. Η κάθετη ευθεία στο επίπεδο της εικόνας που διέρχεται από το κέντρο της κάμερας ονομάζεται κύριος (ή πρωτεύων) άξονας ή κύρια ακτίνα της κάμερας. Επίσης το σημείο τομής του κύριου άξονα με το επίπεδο της εικόνας ονομάζεται κύριο (ή πρωτεύον) σημείο [49].

Έστω ότι το κέντρο της προβολής βρίσκεται στην αρχή του Ευκλείδειου συστήματος συντεταγμένων και ας θεωρήσουμε το επίπεδο Z = f ως το επίπεδο της εικόνας ή το εστιακό επίπεδο. Σύμφωνα με το μοντέλο της κάμερας οπής,

ένα σημείο **X** στο χώρο με συντεταγμένες  $\mathbf{X} = \begin{pmatrix} X & Y & Z \end{pmatrix}^{\mathbf{T}}$  αντιστοιχίζεται σε ένα σημείο του επιπέδου της εικόνας, το οποίο είναι το σημείο τομής του επιπέδου της εικόνας και της ευθείας που ενώνει το σημείο **X** του χώρου και το κέντρο C της κάμερας. Αυτό φαίνεται στο Σχήμα 2.1. Στο ίδιο σχήμα από όμοια τρίγωνα μπορεί να υπολογιστεί ότι το σημείο  $\begin{pmatrix} X & Y & Z \end{pmatrix}$  του χώρου

αντιστοιχίζεται στο σημείο  $\left(\frac{fX}{Z} \quad \frac{fY}{Z}\right)^T$ , οπότε συνολικά έχουμε:

$$\begin{pmatrix} X & Y & Z \end{pmatrix}^T \mapsto \begin{pmatrix} \frac{fX}{Z} & \frac{fY}{Z} \end{pmatrix}^T$$
 (2.1)

κάτι που περιγράφει γενικά την κεντρική προβολική απεικόνιση από τον κόσμο στις συντεταγμένες της εικόνας.



Σχήμα 2.1: Γεωμετρία Κάμερας Οπής. *C* είναι το κέντρο της κάμερας και *p* το κύριο σημείο. Το κέντρο της κάμερας είναι εδώ τοποθετημένο στην αρχή των αξόνων. Το επίπεδο της εικόνας είναι τοποθετημένο μπροστά από το κέντρο της κάμερας [49].

Στη σχέση (2.1) έχει θεωρηθεί ότι το κέντρο των συντεταγμένων στο επίπεδο της εικόνας βρίσκεται στο πρωτεύων σημείο, όπως φαίνεται στο Σχήμα 2.1. Στην πράξη αυτό δεν ισχύει πάντα. Στη γενική περίπτωση πρέπει να γράψουμε:

$$\begin{pmatrix} X & Y & Z \end{pmatrix}^T \mapsto \left(\frac{fX}{Z} + p_x \quad \frac{fY}{Z} + p_y\right)^T$$
 (2.2)

όπου  $\begin{pmatrix} p_x & p_y \end{pmatrix}^T$  είναι οι συντεταγμένες του πρωτεύοντος σημείου (principal point). Η συγκεκριμένη απεικόνιση μπορεί να εκφραστεί σε ομογενείς συντεταγμένες ως εξής:

$$\begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \mapsto \begin{pmatrix} fX + Zp_x \\ fY + Zp_y \\ Z \end{pmatrix} = \begin{pmatrix} f & p_x & 0 \\ f & p_y & 0 \\ & 1 & 0 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}$$
(2.3)  

$$\textbf{'Eoto} K = \begin{pmatrix} f & p_x \\ f & p_y \\ & 1 \end{pmatrix} \textbf{ tote } \textbf{\eta} \textbf{ oxeon (2.3) yivetal:}$$

$$\mathbf{x} = K \begin{pmatrix} I & \mathbf{0} \end{pmatrix} \mathbf{X}_{cam} \tag{2.4}$$

όπου **x** οι συντεταγμένες της προβολής του σημείου στο πλαίσιο της εικόνας και  $\mathbf{X}_{cam} = (X, Y, Z, 1)^{\mathrm{T}}$  οι συντεταγμένες του σημείου στο χώρο εκφρασμένες στο πλαίσιο αναφοράς της κάμερας.



Σχήμα 2.2: Σύστημα συντεταγμένων εικόνας και κάμερας [49].

### 2.2 Περιστροφή και Μεταφορά Κάμερας

Τα σημεία στο χώρο θα εκφραστούν σύμφωνα με ένα διαφορετικό Ευκλείδειο σύστημα συντεταγμένων, γνωστό ως παγκόσμιο σύστημα συντεταγμένων. Τα δύο συστήματα συντεταγμένων κάμερας - χώρου σχετίζονται μέσω μετασχηματισμών περιστροφής και μεταφοράς. Έστω λοιπόν ένα μη ομογενές τριδιάστατο διάνυσμα  $\hat{\mathbf{X}}$  που αντιπροσωπεύει τις συντεταγμένες ενός σημείου εκφρασμένες στο παγκόσμιο σύστημα συντεταγμένων και  $\hat{\mathbf{X}}_{cam}$  οι συντεταγμένες της προβολής του ίδιου σημείου στο πλαίσιο της κάμερας, με βάση το μοντέλο κάμερας οπής. Τότε μπορούμε να γράψουμε:

 $\hat{\mathbf{X}}_{cam} = R(\hat{\mathbf{X}} - \hat{\mathbf{C}})$ όπου  $\hat{\mathbf{C}}$  αντιπροσωπεύει τις συντεταγμένες του κέντρου της κάμερας στο παγκόσμιο σύστημα συντεταγμένων και R είναι ένας  $3 \times 3$  πίνακας περιστροφής που αντιπροσωπεύει τον προσανατολισμό του συστήματος συντεταγμένων της κάμερας. Αυτή η εξίσωση σε ομογενείς συντεταγμένες γίνεται:

$$\mathbf{X}_{\text{cam}} = \begin{pmatrix} R & -R\hat{\mathbf{C}} \\ 0 & 1 \end{pmatrix} \mathbf{X}$$
(2.5)

Δεδομένων των σχέσεων (2.4) και (2.5) οδηγούμαστε στο ακόλουθο αποτέλεσμα :

$$\mathbf{x} = KR \begin{pmatrix} I & -\hat{\mathbf{C}} \end{pmatrix} \mathbf{X}$$
(2.6)

όπου K είναι ο πίνακας βαθμονόμησης της κάμερας και **X** είναι τώρα εκφρασμένο στο παγκόσμιο σύστημα συντεταγμένων. Η σχέση (2.6) εκφράζει τη γενική απεικόνιση μίας κάμερας οπής με μήτρα προβολής.

$$P = KR \begin{pmatrix} I & -\hat{\mathbf{C}} \end{pmatrix}$$
(2.7)

### 2.3 Κάμερα με αισθητήρα τύπου CCD

Το μοντέλο της κάμερας οπής υποθέτει ότι οι εικονικές συντεταγμένες είναι ευκλείδειες συντεταγμένες με ίδια κλίμακα (scaling) και στις δύο αξονικές κατευθύνσεις. Στην περίπτωση χρήσης CCD κάμερας υπάρχει η πιθανότητα μη τετραγωνικών pixels. Αυτό έχει ως αποτέλεσμα την εισαγωγή άνισων παραγόντων κλίμακας σε κάθε κατεύθυνση. Πιο συγκεκριμένα αν ο αριθμός pixels ανά μονάδα απόστασης μετρημένη σε εικονικές συντεταγμένες είναι  $m_x$  και  $m_y$  κατά μήκος των κατευθύνσεων x, y, τότε ο μετασχηματισμός

από παγκόσμιες συντεταγμένες σε εικονικές (pixels) προκύπτει από τον προπολλαπλασιασμό του πίνακα βαθμονόμησης K με τον επιπλέον διαγώνιο  $3 \times 3$  πίνακα  $diag(m_x, m_y, 1)$ . Άρα η γενική μορφή του πίνακα βαθμονόμησης μίας CCD κάμερας είναι:

$$K = \begin{pmatrix} a_x & s & x_o \\ 0 & a_y & y_o \\ 0 & 0 & 1 \end{pmatrix}$$
(2.8)

όπου  $a_x = fm_x$  και  $a_y = fm_y$  αντιπροσωπεύουν το εστιακό μήκος της κάμερας εκφρασμένο σε διαστάσεις pixels κατά τις κατευθύνσεις x και y αντίστοιχα. Όμοια το πρωτεύων σημείο εκφρασμένο σε διαστάσεις pixels με συντεταγμένες  $x_o = m_x p_x$  και  $y_o = m_y p_y$ . Έτσι μία CCD κάμερα έχει 9 βαθμούς ελευθερίας (εννοώντας τις ανεξάρτητες παραμέτρους που απαιτούνται για την περιγραφή οποιασδήποτε τυχαίας διάταξης). Γενικότερα όμως μία πεπερασμένη προβολική κάμερα έχει άλλη μία παράμετρο που ονομάζεται λοξότητα (skew) και είναι μηδέν για τις περισσότερες κάμερες. Άρα μία κάμερα με πίνακα προβολής  $P = KR (I, -\hat{\mathbf{C}})$  και πίνακα βαθμονόμησης από τη σχέση (2.8) ονομάζεται πεπερασμένη προβολική κάμερα και έχει 11 βαθμούς ελευθερίας.

## 2.4 Βαθμονόμηση Κάμερας

Η βαθμονόμηση της κάμερας αποτελεί απαραίτητο βήμα στη 3Δ υπολογιστική όραση προκειμένου να εξαχθεί μετρική πληροφορία από 2Δ εικόνες. Αρκετός όγκος δουλειάς έχει παρουσιαστεί σε αυτόν τον τομέα. Οι τεχνικές που έχουν αναπτυχθεί μπορούν να χωριστούν σε δύο βασικές κατηγορίες:

- Φωτογραμμετρική βαθμονόμηση: Η βαθμονόμηση της κάμερας επιτυγχάνεται με την παρατήρηση ενός αντικειμένου, του οποίου η γεωμετρία στο 3Δ χώρο είναι γνωστή και μάλιστα με πολύ καλή ακρίβεια [49]. Το αντικείμενο βαθμονόμησης αποτελείται συνήθως από δύο ή τρία επίπεδα κάθετα μεταξύ τους. Υπάρχει η δυνατότητα το αντικείμενο να εκτελεί μία συγκεκριμένη κίνηση από πριν γνωστή με μεγάλη ακρίβεια στο χώρο. Βέβαια τέτοιου τύπου προσεγγίσεις απαιτούν ακριβό εξοπλισμό.
- Αυτο-Βαθμονόμηση (Self-Calibration): Τεχνικές βαθμονόμησης αυτής της κατηγορίας δεν απαιτούν αντικείμενο βαθμονόμησης. Η κίνηση της κάμερας γύρω από στατική σκηνή δημιουργεί δύο περιορισμούς για τις εσωτερικές παραμέτρους της κάμερας [57]. Αυτό σημαίνει ότι

αν οι εικόνες προέρχονται από την ίδια κάμερα με σταθερές εσωτερικές παραμέτρους τότε οι αντιστοιχίες σημείων μεταξύ τριών εικόνων είναι αρκετές για τον υπολογισμό των εσωτερικών και εξωτερικών παραμέτρων της κάμερας οι οποίες μας επιτρέπουν την 3Δ ανακατασκευή της εικονιζόμενης δομής με κάποιο μετασχηματισμό ομοιότητας. Αυτού του τύπου η προσέγγιση στο πρόβλημα της βαθμονόμησης αν και είναι χαρακτηριστικά ευέλικτη έχει πολλά προβλήματα. Επειδή υπάρχουν πολλές παράμετροι να υπολογιστούν δεν παράγουν πάντοτε αξιόπιστα αποτελέσματα. Παρακάτω θα περιγραφούν αναλυτικά τεχνικές βαθμονόμησης καθώς επίσης θα παρουσιαστούν και θα αξιολογηθούν και κάποια πειραματικά αποτελέσματα. Ιδιαίτερη ανάλυση θα ακολουθήσει και για την τεχνική βαθμονόμησης που προτείνει ο Zhang [87] σύμφωνα με την οποία η κάμερα χρειάζεται να παρατηρήσει ένα επίπεδο αντικείμενο βαθμονόμησης σε μερικούς διαφορετικούς προσανατολισμούς (τουλάχιστον δύο). Είτε η κάμερα είτε το αντικείμενο μπορούν ελεύθερα να κινηθούν και η κίνηση δε χρειάζεται να είναι γνωστή.

## 2.5 Βασικές Εξισώσεις

Έστω διδιάστατο σημείο [u, v], ως προς τους άξονες αναφοράς u και  $\nu$ , και ένα 3Δ σημείο [X, Y, Z]. Με βάση τη προηγούμενη περιγραφή τα αντίστοιχα επαυξημένα (ομογενοποιημένα) διανύσματα είναι:  $\mathbf{m} = [u, v, 1]^T$  και  $\mathbf{M} = [X, Y, Z, 1]^T$ . Η κάμερα μοντελοποιείται ως μία κάμερα οπής. Η σχέση μεταξύ ενός 3Δ σημείου **M** και της προβολής του **m** στην εικόνα δίνεται από τη σχέση:

$$s\mathbf{m} = K[R \mathbf{t}]\mathbf{M} \tag{2.9}$$

όπου s ένας αυθαίρετος παράγοντας κλίμακας,  $(\mathbf{R}, \mathbf{t})$  οι εξωτερικές παράμετροι που αποτελούνται από περιστροφή και μεταφορά. Οι εξωτερικές παράμετροι συσχετίζουν το παγκόσμιο σύστημα συντεταγμένων με αυτό της κάμερας. Τέλος K είναι ο πίνακας των εσωτερικών παραμέτρων της κάμερας που δίνεται όπως παρουσιάστηκε και παραπάνω ως:

$$K = \begin{pmatrix} \alpha & \gamma & u_0 \\ 0 & \beta & \nu_0 \\ 0 & 0 & 1 \end{pmatrix}$$
(2.10)

όπου  $(u_0, \nu_0)$  οι συντεταγμένες του πρωτεύοντος σημείου (principal point),  $\alpha$ ,  $\beta$  τα εστιακά μήκη με παράγοντες κλίμακας της εικόνας στην κατεύθυνση των αξόνων αναφοράς u και  $\nu$ . Τέλος η παράμετρος  $\gamma$  περιγράφει τη λοξότητα (skew) των δύο αξόνων της εικόνας.
### 2.5.1 Υπολογισμός Ομοιογραφήματος (homography) μεταξύ Επίπεδου Μοντέλου και της Εικόνας του

Χωρίς απώλεια της γενικότητας, θεωρούμε ότι το επίπεδο μοντέλο είναι στο επίπεδο Z = 0 του παγκόσμιου συστήματος συντεταγμένων. Από τη σχέση (2.9) έχουμε:

$$s\begin{pmatrix} u\\v\\1 \end{pmatrix} = K\left(\begin{array}{ccc}\mathbf{r}_1 & \mathbf{r}_2 & \mathbf{r}_3 & \mathbf{t}\end{array}\right)\begin{pmatrix} X\\Y\\0\\1 \end{pmatrix} = K\left(\begin{array}{ccc}\mathbf{r}_1 & \mathbf{r}_2 & \mathbf{t}\end{array}\right)\begin{pmatrix} X\\Y\\1 \end{pmatrix} \quad (2.11)$$

οπότε η σχέση (2.9) γίνεται ίση με

$$s\overline{\mathbf{m}} = H\overline{\mathbf{M}}, \quad H = K \begin{pmatrix} \mathbf{r}_1 & \mathbf{r}_2 & \mathbf{t} \end{pmatrix}$$
 (2.12)

όπου  $\overline{\mathbf{m}}$  η προβολή ενός σημείου  $\overline{\mathbf{M}}$  του επίπεδου μοντέλου βαθμονόμησης στο επίπεδο της εικόνας. Για τον υπολογισμό του ομοιογραφήματος H(homography) θα εφαρμοστεί μία τεχνική βασισμένη στο κριτήριο της μέγιστης πιθανοφάνειας. Αρχικά χρησιμοποιώντας γραμμική μέθοδο θα έχουμε μία αρχική εκτίμηση του πίνακα H. Επίσης γνωρίζουμε την αντιστοιχία σημείων  $M_i \leftrightarrow \mathbf{m}_i$  μεταξύ διδιάστατων σημείων του αντικειμένου βαθμονόμησης στο παγκόσμιο σύστημα συντεταγμένων και των αντίστοιχων σημείων της εικόνας (pixels), τα οποία πιθανόν έχουν βρεθεί με τη βοήθεια κάποιου αλγόριθμου επεξεργασίας εικόνας. Έτσι, αν  $\mathbf{x}^{\mathrm{T}} = (\overline{\mathbf{h}}_1 \ \overline{\mathbf{h}}_2 \ \overline{\mathbf{h}}_3)$  με  $\overline{\mathbf{h}}_i$  την i<sup>στη</sup> γραμμή του H, από την (2.12) προκύπτει:

$$\begin{pmatrix} \overline{\mathbf{M}}^T & 0^T & -u\overline{\mathbf{M}}^T \\ 0^T & \overline{\mathbf{M}}^T & -\nu\overline{\mathbf{M}}^T \end{pmatrix} \mathbf{x} = \mathbf{0}$$
(2.13)

Αν δοθούν n σημεία, έχουμε n εξισώσεις με τη μορφή της εξίσωσης (2.13), οι οποίες μπορούν να γραφούν ως εξίσωση πινάκων  $L\mathbf{x} = \mathbf{0}$ , όπου  $L : 2n \times 9$ πίνακας. Το σύστημα είναι υπερπλήρες (over determined) και μία λύση του, υποκείμενη στον περιορισμό  $||\mathbf{x}|| = 1$  αποτελεί το ιδιοδιάνυσμα του  $L^T L$ που αντιστοιχεί στη μικρότερη ιδιοτιμή. Στον πίνακα L κάποια στοιχεία είναι σταθερά, κάποια σε pixels, κάποια σε παγκόσμιες συντεταγμένες και κάποια πολλαπλασιασμός και των δύο. Αυτό το γεγονός κάνει τον L μη επαρκώς αριθμητικά ορισμένο. Συνεπώς είναι σκόπιμο πριν την επίλυση του συστήματος να γίνει μία ομαλοποίηση των δεδομένων όπως έχει προταθεί στο [68]. Η συνηθέστερη ομαλοποίηση είναι η εξής: Τα σημεία της εικόνας μεταφέρονται έτσι ώστε το κέντρο μάζας (centroid) να είναι στο 0 και κλιμακούμενα έτσι ώστε η RMS απόσταση από την αρχή των αξόνων να είναι ίση με  $\sqrt{2}$ . Το είδος της ομαλοποίησης που πρέπει να επιβληθεί στα 2Δ σημεία του χώρου αποτελεί ένα κάπως πιο δύσκολο πρόβλημα. Στην περίπτωση που η μεταβολή του βάθους είναι σχετικά ομαλή επιλέγουμε αντίστοιχη ομαλοποίηση με πριν. Έτσι λοιπόν σε αυτήν την περίπτωση το κέντρο μάζας των σημείων μεταφέρεται στην αρχή των αξόνων και οι συντεταγμένες τους κλιμακώνονται έτσι ώστε η RMS απόσταση από την αρχή να είναι ίση με  $\sqrt{3}$ . Η γραμμική εκτίμηση του ομοιογραφήματος αποτελεί την αρχικοποίηση του επαναληπτικού αλγορίθμου (Levenberg-Marquardt) που ελαχιστοποιεί το γεωμετρικό λάθος στον πίνακα *H*.

### 2.5.2 Περιορισμοί στις Εσωτερικές Παραμέτρους της Κάμερας

Dosménns mías eikónas tou antikeímenou babmonómnsmi moreí na upodogioteí o pínakas  $H = \begin{pmatrix} \mathbf{h}_1 & \mathbf{h}_2 & \mathbf{h}_3 \end{pmatrix}$ , opóte:

$$\begin{pmatrix} \mathbf{h}_1 & \mathbf{h}_2 & \mathbf{h}_3 \end{pmatrix} = \lambda K \begin{pmatrix} \mathbf{r}_1 & \mathbf{r}_2 & \mathbf{r}_3 \end{pmatrix}$$

όπου  $\lambda$  είναι μία αυθαίρετη κλίμακα. Χρησιμοποιώντας τη γνώση ότι τα  $\mathbf{r}_1$  και  $\mathbf{r}_2$  είναι ορθοκανονικά, έχουμε:

$$\mathbf{h}_1^T K^{-T} K^{-1} \mathbf{h}_2 = 0 \tag{2.14}$$

$$\mathbf{h}_{1}^{T}K^{-T}K^{-1}\mathbf{h}_{1} = \mathbf{h}_{2}^{T}K^{-T}K^{-1}\mathbf{h}_{2}$$
(2.15)

Οι σχέσεις (2.14) και (2.15) αποτελούν δύο βασικούς περιορισμούς στις εσωτερικές παραμέτρους, δοσμένου ενός ομοιογραφήματος. Επειδή ένα ομοιογράφημα έχει 8 βαθμούς ελευθερίας και υπάρχουν 6 εξωτερικές παράμετροι (3 για περιστροφή και 3 για μεταφορά), είναι δυνατόν να έχουμε 2 περιορισμούς στις εσωτερικές παραμέτρους. Στην περίπτωση που το μοντέλο βαθμονόμησης στη δεύτερη θέση είναι παράλληλο σε σχέση με την πρώτη θέση, τότε το δεύτερο ομοιογράφημα δεν παρέχει επιπλέον γεωμετρικούς περιορισμούς για την επίλυση του προβλήματος και μιλάμε για εκφυλισμένη περίπτωση. Επίσης η ποσότητα  $K^{-T}K^{-1}$  περιγράφει την εικόνα του απόλυτου κώνου, όπως φαίνεται στο Σχήμα 2.3.



Σχήμα 2.3: Μελέτη των εσωτερικών παραμέτρων της κάμερας μέσω του απόλυτου κώνου βαθμονόμησης [49].

# 2.6 Επίλυση του προβλήματος Βαθμονόμησης της Κάμερας

Θεωρούμε

$$\mathbf{B} = \mathbf{K}^{-T} \mathbf{K}^{-1} = \begin{pmatrix} B_{11} & B_{12} & B_{13} \\ B_{21} & B_{22} & B_{23} \\ B_{31} & B_{32} & B_{33} \end{pmatrix} = \begin{pmatrix} \frac{1}{\alpha_2} & -\frac{\gamma}{\alpha^2 \beta} & \frac{\nu_0 \gamma - u_0 \beta}{\alpha^2 \beta} \\ -\frac{\gamma}{\alpha^2 \beta} & \frac{\gamma_2}{\alpha^2 \beta^2} + \frac{1}{\beta^2} & \frac{-\gamma(nu_0 \gamma - u_0 \beta)}{\alpha^2 \beta^2} - \frac{\nu_0}{\beta^2} \\ \frac{\nu_0 \gamma}{-u_0 \beta} & \frac{-\gamma(\nu_0 \gamma - u_0 \beta)}{\alpha^2 \beta^2} - \frac{\nu_0}{\beta^2} & \frac{(\nu_0 \gamma - u_0 \beta)^2}{\alpha^2 \beta^2} + \frac{\nu_0}{\beta^2} + 1 \end{pmatrix}$$
(2.16)

Ο πίνακας **B** είναι συμμετρικός και ορίζεται από ένα 6Δ διάνυσμα

$$\mathbf{b} = \begin{pmatrix} B_{11} & B_{12} & B_{22} & B_{13} & B_{23} & B_{33} \end{pmatrix}^T$$

Έστω το  $i^{\sigma au o}$  διάνυσμα στήλη του **Η** είναι  $\mathbf{h}_i = \begin{pmatrix} h_{i1} & h_{i2} & h_{i3} \end{pmatrix}^T$  τότε έχουμε:

$$\mathbf{h}_i^T \mathbf{B} \mathbf{h}_j = \mathbf{v}_{ij}^T \mathbf{b}$$
 (2.17)

με

**v** $_{ij} = \begin{pmatrix} h_{i1}h_{j1} & h_{i1}h_{j2} + h_{i2}h_{j1} & h_{i2}h_{j2} & h_{i3}h_{j1} + h_{i1}h_{j3} & h_{i3}h_{j2} + h_{i2}h_{j3} & h_{i3}h_{j3} \end{pmatrix}^T$ οπότε οι δύο βασικοί περιορισμοί (2.14) και (2.15) για κάθε ομοιογράφημα μπορούν να εκφραστούν μέσω του ομογενούς συστήματος με άγνωστο το **b**,

δηλαδη:

$$\begin{pmatrix} \mathbf{v}_{12}^T \\ (\mathbf{v}_{12} - \mathbf{v}_{22})^T \end{pmatrix} \mathbf{b} = \mathbf{0}$$
(2.18)

Αν διατεθούν n εικόνες του επιπέδου βαθμονόμησης τότε συγκεντρώνονται n εξισώσεις όπως την (2.18), οπότε:

$$\mathbf{V}\mathbf{b} = \mathbf{0} \tag{2.19}$$

όπου **V** είναι ένας  $2n \times 6$  πίνακας. Αν έχουμε μοναδική λύση **b** είναι ορισμένη σε μία κλίμακα. Αν μπορούμε να προσθέσουμε έναν περιορισμό μη λοξότητας  $\gamma = 0$  δηλαδή μία επιπλέον εξίσωση στη (2.18) τότε μπορούμε να υπολογίσουμε μόνο 2 εσωτερικές παραμέτρους, θεωρώντας για παράδειγμα ότι το κύριο σημείο της κάμερας είναι τοποθετημένο στο κέντρο της κάμερας. Από τη στιγμή που προσδιοριστεί το **b** υπολογίζουμε και όλες τις εσωτερικές παραμέτρους :

$$v_0 = \frac{B_{12}B_{13} - B_{11}B_{23}}{B_{11}B_{22} - B_{12}^2}$$
(2.20)

$$\lambda = B_{33} - \frac{B_{13}^2 + \upsilon_0 (B_{12} B_{13} - B_{11} B_{23})}{B_{11}}$$
(2.21)

$$\alpha = \sqrt{\frac{\lambda}{B_{11}}} \tag{2.22}$$

$$\beta = \sqrt{\frac{\lambda B_{11}}{B_{11}B_{22} - B_{12}^2}} \tag{2.23}$$

$$\gamma = \frac{-B_{12}\alpha^2\beta}{\lambda} \tag{2.24}$$

$$u_0 = \frac{\gamma v_0}{\beta} - \frac{B_{13} \alpha^2}{\lambda}$$
(2.25)

(2.26)

Οι εξωτερικές παράμετροι με δεδομένες τις εσωτερικές μπορούν εύκολα να προσδιοριστούν ως εξής:

$$\mathbf{r}_1 = \lambda K^{-1} \mathbf{h}_1 \tag{2.27}$$

$$\mathbf{r}_2 = \lambda K^{-1} \mathbf{h}_2 \tag{2.28}$$

$$\mathbf{r}_3 = \mathbf{r}_1 \times \mathbf{r}_2 \tag{2.29}$$

$$\mathbf{t} = \lambda K^{-1} \mathbf{h}_3 \tag{2.30}$$

(2.31)

με  $\lambda = \frac{1}{\|K^{-1}\mathbf{h}_1\|} = \frac{1}{\|K^{-1}\mathbf{h}_2\|}$ . Είναι αυτονόητο ότι εξαιτίας του θορύβου στα δεδομένα ο υπολογισμένος πίνακας  $R = (\mathbf{r}_1 \ \mathbf{r}_2 \ \mathbf{r}_3)$  δεν ικανοποιεί γενικά τις ιδιότητες ενός πίνακα περιστροφής. Έτσι με δεδομένο τον πίνακα  $\mathbf{R}$  μπορούμε να βρούμε το βέλτιστο πίνακα περιστροφής  $\mathbf{R}_{opt}$  βρίσκοντας τον πίνακα εκείνον ο οποίος ελαχιστοποιεί τη Frobenius νόρμα της διαφοράς  $\mathbf{R}_{opt} - \mathbf{R}$ , δηλαδή:

 $min_{\mathbf{R}} \| \mathbf{R}_{opt} - \mathbf{R} \|$  δεδομένου ότι  $\mathbf{R}^T \mathbf{R} = \mathbf{I}$ 

Αν εφαρμοστεί SVD στον R έχουμε:  $R = UDV^T$  αποδεικνύεται ότι  $R_{opt} = UV^T$  [27].

## 2.7 Ακτινική Παραμόρφωση

Είναι δυνατόν σε κάποιες κάμερες να έχουμε εμφανή παραμόρφωση φακών και πιο συγκεκριμένα ακτινική παραμόρφωση. Έστω  $[u, \nu]$  οι ιδανικές (μη παρατηρήσιμες) συντεταγμένες των σημείων της εικόνας σε εικονοστοιχεία pixels, και  $[\breve{u}, \breve{\nu}]$  οι πραγματικές παρατηρήσιμες συντεταγμένες των σημείων της εικόνας. Τα ιδανικά σημεία είναι προβολές των σημείων του μοντέλου βαθμονόμησης σύμφωνα με το μοντέλο της κάμερας οπής. Όμοια [x, y] και  $[\breve{x}, \breve{y}]$  είναι οι ιδανικές (χωρίς παραμόρφωση) και οι πραγματικές (με παραμόρφωση) ομαλοποιημένες συντεταγμένες σημείων της εικόνας. Από [28] έχουμε:

$$\breve{x} = x + x[k_1(x^2 + y^2) + k_2(x^2 + y^2)^2]$$
 (2.32)

$$\breve{y} = y + y[k_1(x^2 + y^2) + k_2(x^2 + y^2)^2]$$
 (2.33)

όπου  $k_1$  και  $k_2$  είναι οι συντελεστές της ακτινικής παραμόρφωσης. Το κέντρο της ακτινικής παραμόρφωσης είναι ίδιο με το κύριο σημείο. Θεωρούμε συντεταγμένες του πρωτεύοντος σημείου  $[u_0, \nu_0]$ , οπότε βασισμένοι στις εξισώσεις (2.32) και (2.33) δημιουργούμε τις εξισώσεις που συνδέουν τα σημεία, εκφρασμένα στο πλαίσιο της εικόνας, με ή χωρίς παραμόρφωση:

$$\breve{u} - u_0 = (u - u_0)[k_1(x^2 + y^2) + k_2(x^2 + y^2)^2]$$
 (2.34)

$$\breve{\nu} - \nu_0 = (\nu - \nu_0)[k_1(x^2 + y^2) + k_2(x^2 + y^2)^2]$$
 (2.35)

Αναμένοντας την ακτινική παραμόρφωση μικρή, μία καλή στρατηγική είναι να υπολογιστούν πρώτα οι υπόλοιπες 5 εσωτερικές (εκτός skew) παράμετροι αγνοώντας την παραμόρφωση. Με αυτόν τον τρόπο θα έχουμε τις ιδανικές συντεταγμένες σε pixels των σημείων της εικόνας (u, v) σύμφωνα με τη σχέση (2.12). Άρα για κάθε σημείο της εικόνας έχουμε δύο εξισώσεις με αγνώστους τα  $k_1, k_2$ . Έχοντας *m* σημεία σε *n* εικόνες, μπορούμε να «στοιβάξουμε» όλες τις εξισώσεις μαζί και να έχουμε 2mn εξισώσεις εκφραζόμενες με πίνακες  $D\mathbf{k} = \mathbf{d}$ , όπου  $\mathbf{k} = (k_1, k_2)^{\mathrm{T}}$ , όπου **D** είναι το διάνυσμα με στοιχεία το πρώτο μέλος των εξισώσεων (2.34) και (2.35), και όπου **d** είναι το διάνυσμα με στοιχεία το δεύτερο μέλος των εξισώσεων (2.34) και (2.35). Η λύση ελάχιστων τετραγώνων του συστήματος δίνεται από τη σχέση:  $\mathbf{k} = (D^T D)^{-1} D^T \mathbf{d}$ .



Σχήμα 2.4: .Επίδραση της Ακτινικής Παραμόρφωσης και διόρθωση [62].

# 2.8 Εξοπλισμός

Το πρώτο μέρος των πειραμάτων που εκτελέσαμε περιλαμβάνει τη βαθμονόμηση μίας φωτογραφικής κάμερας με τη μέγιστη δυνατή ακρίβεια. Η κάμερα που χρησιμοποιήθηκε σε όλα τα πειράματα είναι μια Nikon D70s 18-74mm. Η διαδικασία της βαθμονόμησης βασίστηκε στη βαθμονόμηση με επίπεδα αντικείμενα σύμφωνα με τον Zhang [87]. Το αντικείμενο βαθμονόμησης που χρησιμοποιήθηκε ειναι η εικόνα μίας επίπεδης σκακιέρας. Έχει δοθεί ιδιαίτερη προσοχή στον τρόπο φωτογράφισης έτσι ώστε τα επίπεδα της εικόνας να μην είναι παράλληλα.

# 2.9 Αλγόριθμος Βαθμονόμησης

Θεωρούμε την αρχή των αξόνων του παγκόσμιου συστήματος στην πάνω αριστερή γωνία της σκακιέρας. Κάθε τετράγωνο της σκακιέρας έχει διαστάσεις  $2.8 \times 2.8$  cm.

Κατόπιν χρησιμοποιώντας έναν απλό αλγόριθμο εντοπισμού γωνιών βρίσκουμε τις αντίστοιχες συντεταγμένες σε εικονοστοιχεία των γωνιών. Δεδομένης λοιπόν της αντιστοιχίας μεταξύ σημείων (τουλάχιστον 6) στο παγκόσμιο σύστημα συντεταγμένων και σημείων στο σύστημα συντεταγμένων της εικόνας, θα υπολογιστεί το ομοιογράφημα κάμερας και επιπέδου σκακιέρας. Επειδή το αντικείμενο βαθμονόμησης είναι επίπεδο δηλαδή Z = 0 δεν έχουμε τη δυνατότητα να υπολογίσουμε τη μήτρα προβολής της κάμερας αλλά μόνο το ομοιογράφημα (Homography).

Η αρχική εκτίμηση του ομοιογραφήματος χρησιμοποιώντας γραμμική μέθοδο συνίσταται στα εξής:

#### Ομαλοποίηση

Έγινε χρήση ενός μετασχηματισμού ομοιότητας Τ προκειμένου να ομαλοποιηθούν τα σημεία στην εικόνα, ώστε το κέντρο μάζας τους να είναι στο 0 και η RMS απόσταση ίση με  $\sqrt{2}$ . Επομένως ο πίνακας Τ έχει τη μορφή:

$$T = \begin{pmatrix} scale & 0 & -t_x \\ 0 & scale & -t_y \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} \frac{\sqrt{2n}}{\sum_{i=1}^n distance} & 0 & \frac{\sqrt{2n}}{\sum_{i=1}^n distance} centroid_x \\ 0 & \frac{\sqrt{2n}}{\sum_{i=1}^n distance} & \frac{\sqrt{2n}}{\sum_{i=1}^n distance} centroid_y \\ 0 & 0 & 1 \end{pmatrix}$$

$$(2.36)$$

όπου distance =  $||x_i - centroid||_2$  και όπου  $||.||_2$  αντιπροσωπεύει την Ευκλείδεια απόσταση. Επίσης  $centroid_x = \frac{\sum_{i=1}^N x_i}{N}$  και  $centroid_y = \frac{\sum_{i=1}^N y_i}{N}$ , το κέντρο μάζας των σημείων.

Είναι αναγκαία επίσης και η ομαλοποίηση των σημείων του χώρου έτσι ώστε το κέντρο μάζας τους να είναι στο μηδέν και η μέση RMS απόσταση  $\sqrt{3}$ . Ο αντίστοιχος μετασχηματισμός ομοιότητας που χρησιμοποιήθηκε είναι ο

$$U = \begin{pmatrix} scale & 0 & -t_x \\ 0 & scale & -t_y \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} \frac{\sqrt{3n}}{\sum_{i=1}^n distance} & 0 & \frac{\sqrt{3n}}{\sum_{i=1}^n distance} centroid_x \\ 0 & \frac{\sqrt{3n}}{\sum_{i=1}^n distance} & \frac{\sqrt{3n}}{\sum_{i=1}^n distance} centroid_y \\ 0 & 0 & 1 \end{pmatrix}$$

$$(2.37)$$

όπου distance =  $||X_i - centroid||_2$  και  $||.||_2$  αντιπροσωπεύει την Ευκλείδεια απόσταση. Επίσης  $centroid_x = \frac{\sum_{i=1}^N X_i}{N}$  και  $centroid_y = \frac{\sum_{i=1}^N Y_i}{N}$ . Άρα τα

ομαλοποιημένα σημεία της εικόνας δίνονται από τη σχέση:

$$x_i^{\text{norm}} = Tx_i \tag{2.38}$$

ενώ τα ομαλοποιημένα σημεία του χώρου μετρημένα στο παγκόσμιο σύστημα από τη σχέση:

$$X_i^{\text{norm}} = UX_i \tag{2.39}$$

#### Ελαχιστοποίηση του Γεωμετρικού Λάθους

Αφού σχηματίσουμε τον πίνακα L από τις αντιστοιχίες  $M_i^{\text{norm}} \leftrightarrow m_i^{\text{norm}}$ , όπως στην παράγραφο 2.5.1, υπολογίζουμε όπως αναλύθηκε παραπάνω τη λύση ελαχίστων τετραγώνων του ομογενούς συστήματος. Αρχικά λοιπόν έχουμε μία γραμμική εκτίμηση ως σημείο αρχικοποίησης και επακόλουθα ελαχιστοποιούμε το γεωμετρικό λάθος :

$$\sum_{i} d\left(\overline{\mathbf{m}}_{i}, H\overline{\mathbf{M}}_{i}\right)^{2}$$
(2.40)

χρησιμοποιώντας την επαναληπτική μέθοδο Levenberg-Marquardt. Τέλος υπολογίζουμε το πραγματικό ομοιογράφημα ως εξής:

$$H = T^{-1} \widehat{H} U \tag{2.41}$$

όπου H το ομοιογράφημα που έχει προέλθει από αντιστοιχίες ομαλοποιημένων συντεταγμένων τόσο στο πλαίσιο της εικόνας, όσο και στο πλαίσιο του κόσμου.

# 2.10 Βαθμονόμηση: Μετρήσεις και Αποτελέσματα

Οι διαφορετικές φωτογραφίες της αντικειμένου βαθμονόμησης σε μη παράλληλα επίπεδα φαίνονται στο Σχήμα 2.5.

Ως αντικείμενο βαθμονόμησης χρησιμοποιήσαμε τη φωτογραφία μίας σκακιέρας 9 × 7 τετραγώνων προσαρμοσμένη σε επίπεδη βάση. Το πρόγραμμα βαθμονόμησης έχει σαν είσοδο το μήκος της πλευράς κάθε τετραγώνου σε mm (28mm) και το πλήθος των τετραγώνων σε κάθε διεύθυνση (κατά x : 7 τετράγωνα και κατά y: 9 τετράγωνα). Κατόπιν υλοποιούμε τον αλγόριθμο βαθμονόμησης για επίπεδα αντικείμενα βαθμονόμησης [87]. Έτσι υπολογίζουμε τους πίνακες εσωτερικών και εξωτερικών παραμέτρων ελαχιστοποιώντας παράλληλα το γεωμετρικό λάθος με χρήση επαναληπτικών μεθόδων. Τέλος υπολογίζουμε και τις παραμέτρους ακτινικής παραμόρφωσης για τη



Σχήμα 2.5: Διαδοχικές φωτογραφίες (σε κλίμακα του γκρι) του αντικειμένου βαθμονόμησης από διαφορετικές οπτικές γωνίες σε μη παράλληλα επίπεδα.

διόρθωση των αρχικών φωτογραφιών. Έτσι λοιπόν, στην πρώτη φάση των πειραμάτων η κάμερα βαθμονομήθηκε ακολουθώντας τον αλγόριθμο που προτάθηκε από τον Zhang στο [87], για επίπεδα πρότυπα βαθμονόμησης. Όσον αφορά την λύση του προβλήματος αυτοβαθμονόμησης υιοθετήθηκαν οι ακόλουθες παραδοχές. Το πρωτεύον σημείο βρίσκεται στο κέντρο του επιπέδου της εικόνας, απώλεια λοξότητας, και η εστιακή απόσταση ανά μονάδα κλίμακας και στις δύο κατευθύνσεις του επιπέδου της εικόνας ( $\alpha_x$  και  $\alpha_y$ ) έχουν την ίδια τιμή. Συνεπώς ο πίνακας εσωτερικών παραμέτρων παίρνει την ακόλουθη μορφή:

$$K = \begin{pmatrix} \alpha & 0 & w/2 \\ 0 & \alpha & h/2 \\ 0 & 0 & 1 \end{pmatrix}$$
(2.42)

Επιπρόσθετα, η επανόρθωση των εικόνων απλοποιεί την διαδικασία ταιριάσματος, μετασχηματίζοντας τη περιοχή αναζήτησης από διδιάστατη σε μονοδιάστατη. Επομένως οι επιπολικές γραμμές γίνονται παράλληλες και συμπίπτουν με τις γραμμές σάρωσης που χρησιμοποιούνται για την εύρεση αντίστοιχων εικονοστοιχείων. Ο μετασχηματισμός επανόρθωσης βασίζεται στις παραπάνω υποθέσεις που αφορούν τις εσωτερικές παραμέτρους της κάμερας, σύμφωνα με τους Hartley, [68] και Koch et al., [64]. Τα αποτελέσματα επανόρθωσης βελτιώνονται περαιτέρω με την ολοκλήρωση της διαδικασίας ταιριάσματος, χρησιμοποιώντας είτε συσχέτιση φωτεινότητας, είτε μονογενή φίλτρα.

### 2.10.1 Εσωτερικές Παράμετροι μετά την Ελαχιστοποίηση του Γεωμετρικού Λάθους

#### 1. Εστιακό Μήκος:

$$fc = (832.17370 \ 823.73043)$$

με απόκλιση: (16.05139 16.81352)

#### 2. Πρωτεύον Σημείο:

 $cc = (154.00986 \ 126.87237)$ 

με απόκλιση: (16.51879 20.30434)

#### 3. **Λοξότητα:**

 $alpha_c = 0$ 

η γωνία μεταξύ πλευρών τετραγώνων είναι 90.00000 μοίρες

### 2.10.2 Παράμετροι Ακτινικής Παραμόρφωσης

 $\mathbf{kc} = (0.23717 \ 1.37004)$ 

με απόκλιση (0.14923 4.9275) Η διόρθωση της παραμόρφωσης φαίνεται στο σχήμα (2.6). Οι διαφορές είναι ανεπαίσθητες λόγω της καλής ποιότητας του φακού.

### 2.10.3 Εξωτερικές Παράμετροι

Υπολογίσαμε τις εξωτερικές παραμέτρους της κάμερας για τα διαφορετικά επίπεδα βαθμονόμησης. Στο Σχήμα 2.7 φαίνεται η γραφική αναπαράσταση του προσανατολισμού των επίπεδων πρότυπων βαθμονόμησης.

### 2.11 Συμπεράσματα Κεφαλαίου

Στο παρόν κεφάλαιο προηγήθηκε εκτενής ανάλυση της κύριας μοντελοποίησης της κάμερας που χρησιμοποιήθηκε στα πειράματα κατά τη λήψη φωτογραφιών εξωτερικού χώρου. Επίσης παρουσιάστηκε η κύρια μέθοδος βαθμονόμησης της κάμερας, που χρησιμοποιήθηκε. Κύριος σκοπός της βαθμονόμησης αποτελεί ο ακριβής υπολογισμός των εσωτερικών παραμέτρων της



Σχήμα 2.6: Διόρθωση ακτινικής παραμόρφωσης. Δεξιά στήλη: αρχικές εικόνες. Αριστερή στήλη: Μετασχηματισμένες εικόνες χωρίς ακτινική παραμόρφωση.





Σχήμα 2.7: Εξωτερικές παράμετροι κάμερας-εικόνων αντικειμένων βαθμονόμησης.

κάμερας (γεωμετρικά στοιχεία φακού) και των εξωτερικών παραμέτρων (τριδιάστατη θέση και προσανατολισμός κάμερας σε σχέση με το επίπεδο της εικόνας). Η συγκεκριμένη μέθοδος βαθμονόμησης ανακτά τη γεωμετρία της κάμερας αξιοποιώντας επίπεδα πρότυπα μέτρησης, κάτι που διευκολύνει τη διαδικασία της βαθμονόμησης κατά τη διάρκεια λήψης πειραματικών φωτογραφιών. Οι μέθοδοι που εφαρμόστηκαν αποτελούν μέρος της κλασσικής προσέγγισης στο πρόβλημα της τριδιάστατης ανακατασκευής του κόσμου από εικόνες. Η διαδικασία αυτή μοντελοποίησης και βαθμονόμησης της κάμερας συντελεί στη μετρική συσχέτιση των ληφθέντων εικόνων και του πραγματικού κόσμου. Με το τρόπο αυτό επιτυγχάνεται μία πιο ακριβής τριδιάστατη ανακατασκευή, όπως θα φανεί σε επόμενα κεφάλαια.

# Κεφάλαιο 3

# Τριδιάστατη Ανακατασκευή -Βιβλιογραφική Επισκόπηση -Κλασικές και Σύγχρονες Προσεγγίσεις

Η τριδιάστατη ανακατασκευή και η δημιουργία τριδιάστατων μοντέλων από επίπεδες εικόνες (φωτογραφίες) συνιστά έναν από τους πιο απαιτητικούς τομείς στο χώρο της υπολογιστικής όρασης και των γραφικών με υπολογιστές. Τα προηγούμενα χρόνια παρουσιάστηκε σημαντική ανάγκη ανάπτυξης εφαρμογών με έμφαση στην όραση, στην οδήγηση ρομποτικών οχημάτων και γενικά στην εικονική επιτήρηση ευαίσθητων συστημάτων. Σήμερα με την αλματώδη εξέλιξη τεχνικών απεικόνισης, υπάρχει έντονα η ανάγκη δημιουργίας γραφικών τριδιάστατων μοντέλων μεγάλης ακρίβειας. Τέτοια μοντέλα εξυπηρετούν κυρίως εφαρμογές εικονικής πραγματικότητας.

Η ανάπτυξη όμως, τέτοιας ακρίβειας και πιστότητας 3Δ γραφικών απαιτεί ένα μεγάλο όγκο πληροφορίας. Για αυτό το λόγο πολλές φορές χρησιμοποιείται εξειδικευμένος εξοπλισμός όπως laser scanners ή stereo rigs.

Στην παρούσα διατριβή γίνεται προσπάθεια ανάκτησης τριδιάστατων μοντέλων με χρήση μίας συνηθισμένης επαγγελματικής φωτογραφικής μηχανής SLR. Η ανάγκη απόκτησης ικανού όγκου πληροφορίας ικανοποιείται αφενός με τη χρήση πολλών εικόνων του ίδιου θέματος από διαφορετικές οπτικές γωνίες, και αφετέρου με υλοποίηση μεθόδων που μειώνουν δραστικά ως και εξαλείφουν την ανάγκη συνεχούς βαθμονόμησης της κάμερας, κάτι που θα απαιτούσε χρόνο και πολλές φορές εξειδικευμένο εξοπλισμό.

Ο υπολογισμός και ανάκτηση πληροφορίας βάθους πεδίου από δύο εικόνες συνιστά το κλασικό στερεοσκοπικό πρόβλημα. Η επίλυση ενός τέτοιου προβλήματος περιλαμβάνει διάφορα στάδια, τα οποία είναι τα ακόλουθα:

- 1. Επεξεργασία Εικόνων Ανάκτηση Χαρακτηριστικών
- 2. Αντιστοίχιση Κύριων Χαρακτηριστικών
- 3. Βαθμονόμηση Κάμερας, Επανόρθωση Εικόνων
- 4. Βαθμονομημένη Ανακατασκευή

Παρακάτω θα παρουσιαστούν δύο βασικές προσεγγίσεις στα προβλήματα εντοπισμού χαρακτηριστικών και αντιστοίχισης. Αρχικά η κλασική προσέγγιση αναλύει την εικόνα χωρίς να προηγηθεί κανένα στάδιο προεπεξεργασίας της. Η δεύτερη πιο σύγχρονη προσέγγιση βασίζεται στην πυραμιδική ανάλυση κλίμακας-χώρου της εικόνας. Βασικός στόχος αποτελεί η αναζήτηση χαρακτηριστικών αναλλοίωτων σε μεταβολές της κλίμακας.

Τέλος πραγματοποιείται μία επισκόπηση μεθόδων ορθοποίησης εικόνων, αυτόματης βαθμονόμησης κάμερας καθώς επίσης και κλασικών μεθόδων ανακατασκευής τριδιάστατης δομής.

# 3.1 Επεξεργασία Εικόνων - Ανάκτηση Χαρακτηριστικών - Αρχή της Συσχέτισης

### 3.1.1 Εντοπισμός Χαρακτηριστικών Ακμών - Επισκόπηση Κλασσικών Μεθόδων

Κύριος στόχος του εντοπισμού αλλαγών στη φωτεινότητα της εικόνας αποτελεί η παρακολούθηση των σημαντικών αλλαγών του απεικονιζόμενου θέματος. Ασυνέχειες στη φωτεινότητα της εικόνας πιθανόν αντιστοιχούν σε:

- 1. Ασυνέχειες στο βάθος πεδίου
- 2. Ασυνέχειες στον προσανατολισμό της επιφάνειας
- 3. Μεταβολές στην αντανάκλαση της απεικονιζόμενης σκηνής

Στην ιδανική περίπτωση το αποτέλεσμα εφαρμογής ενός εντοπιστή ακμών σε μία εικόνα οδηγεί σε ένα σύνολο καμπυλών που υποδεικνύουν τα όρια των αντικειμένων ή τις περιοχές ασυνέχειας στον προσανατολισμό των επιφανειών. Συνεπώς εφαρμόζοντας ένα φίλτρο εντοπισμού ακμών μειώνεται ο όγκος της πληροφορίας προς επεξεργασία ενώ ταυτόχρονα διατηρούνται σημαντικά δομικά στοιχεία και ιδιότητες των απεικονιζομένων θεμάτων.

Στη βιβλιογραφία συναντάνται πληθώρα μεθόδων εντοπισμού ακμών. Οι περισσότερες από αυτές ομαδοποιούνται σε δύο βασικές κατηγορίες: οι βασισμένες σε αναζήτηση (search based) και οι βασισμένες σε πλήθος διελεύσεων από το μηδέν (zero crossing based). Οι μέθοδοι βασισμένες σε αναζήτηση εντοπίζουν ακμές αφού πρωταρχικά μετράται η *ισχύς* της ακμής ως έκφραση της παραγώγου πρώτης τάξης ή του πλάτους κλίσης. Κατόπιν αναζητώνται τοπικά ακρότατα του πλάτους κλίσης χρησιμοποιώντας την εκτίμηση του τοπικού προσανατολισμού της ακμής, ή την κατεύθυνση της κλίσης. Οι μέθοδοι διέλευσης από το μηδέν αναζητούν διελεύσεις από το μηδέν της παραγώγου δεύτερης τάξης της εικόνας με στόχο των εντοπισμό ακμών. Συχνά αναζητώνται οι διελεύσεις από το μηδέν της Λαπλασιανής συνάρτησης. Ως κύριο βήμα προεπεξεργασίας για τον εντοπισμό ακμών αποτελεί το στάδιο ομαλοποίησης, το οποίο συχνά υλοποιείται εφαρμόζοντας Γκαουσιανό φίλτρο.

Ο Canny, [15], μελετώντας το μαθηματικό πρόβλημα βελτιστοποίησης φίλτρων ομαλοποίησης, κατέληξε στα κριτήρια εντοπισμού, εντοπιότητας και ελαχιστοποίησης των πολλαπλών αποκρίσεων μίας μοναδικής ακμής. Απέδειξε ότι το βέλτιστο φίλτρο δοσμένων συγκεκριμένων υποθέσεων είναι ένα άθροισμα τεσσάρων εκθετικών όρων. Απέδειξε επίσης ότι αυτό το φίλτρο μπορεί να προσεγγιστεί από παραγώγους πρώτης τάξης Γκαουσιανών συναρτήσεων.

Παρόλο το γεγονός ότι η συγκεκριμένη μεθοδολογία εντοπισμού ακμών εισήχθει την πρώιμη εποχή της όρασης υπολογιστών ακόμη και στις σύγχρονες μέρες έχει ικανοποιητική επίδοση σε προβλήματα εντοπισμού ακμών.

Το φίλτρο εντοπισμού ακμών Canny-Deriche, [61] προέκυψε από όμοια κριτήρια με αυτά του φίλτρου εντοπισμού ακμών από τον Canny χρησιμοποιώντας όμως αναδρομικά φίλτρα για τη διαδικασία ομαλοποίησης της εικόνας αντί για εκθετικά ή Γκαουσιανά φίλτρα.

### 3.1.2 Αντιστοίχιση Κύριων Χαρακτηριστικών - Αρχή της Συσχέτισης

Ένα από τα πιο απαιτητικά προβλήματα της όρασης υπολογιστών είναι το πρόβλημα της αντιστοίχισης (correspondence problem). Το συγκεκριμένο πρόβλημα συνοψίζεται στο εξής: έχοντας δεδομένο ένα σημείο στη μία εικόνα αναζητούμε για το αντίστοιχο στην άλλη. Οι αλγόριθμοι που έχουν αναπτυχθεί χωρίζονται σε δύο μεγάλες κατηγορίες: στις τεχνικές συσχέτισης και στις τεχνικές αντιστοίχισης χαρακτηριστικών. Μία πρώτη προσέγγιση στην επίλυση του προβλήματος αποτελεί η τεχνική συσχέτισης σημείων (correlation based method). Τέτοιες μέθοδοι έχουν χρησιμοποιηθεί στην όραση υπολογιστών και η βασική τους αρχή θα περιγραφεί παρακάτω.

Έστω ένα ζευγάρι στέρεο εικόνων. Αναζητούνται οι συντεταγμένες της μίας εικόνας  $(u_2, \nu_2)$  που αντιστοιχούν στο εικονοστοιχείο με συντεταγμένες  $(u_0, \nu_0)$ , όπως φαίνεται στο Σχήμα 3.1. Θεωρούμε λοιπόν ένα τετραγωνικό

παράθυρο αναζήτησης μεγέθους  $(2P + 1) \times (2N + 1)$ , όπου N, P φυσικοί αριθμοί, με κέντρο το σημείο  $(u_0, \nu_0)$  και υπολογίζουμε το μέτρο συσχέτισης φωτεινότητας του με τη δεύτερη εικόνα στη γραμμή  $\nu_2 = \nu_0$ :

Το μέτρο της συσχέτισης δίνεται:

$$C_{12}(\tau) = \frac{1}{K} \sum_{u=-N}^{+N} \sum_{\nu=-P}^{+P} (I_1(u+u_0,\nu+\nu_0) - \overline{I_1}(u_0,\nu_0)) (I_2(u+u_0+\tau,\nu+\nu_0) - \overline{I_2}(u_0+\tau,\nu_0))$$
(3.1)

όπου  $K = (2N+1)(2P+1)\sigma_1(u_0,\nu_0)\sigma_2(u_0+\tau,\nu_0), \overline{I_1}(u_0,\nu_0)$ είναι η μέση τιμή φωτεινότητας και  $\sigma_1(u_0,\nu_0)$  η τυπική απόκλιση της μίας στέρεο εικόνας στο σημείο  $(u_0,\nu_0)$ . Άρα:

$$\overline{I_1}(u_0,\nu_0) = \frac{1}{(2N+1)(2P+1)} \sum_{u=-N}^{+N} \sum_{\nu=-P}^{+P} (I_1(u+u_0,\nu+\nu_0))$$

$$\sigma_1^2(u_0,\nu_0) = \frac{1}{(2N+1)(2P+1)} \sum_{u=-N}^{+N} \sum_{\nu=-P}^{+P} (I_1(u+u_0,\nu+\nu_0)^2 - \overline{I_1}(u_0,\nu_0))^2$$
The provide measurements are served as a served as a

Το σημείο της εικόνας  $\tau = \tau_0$  για το οποίο η καμπύλη  $C_{12}$  έχει μέγιστο, αποτελεί το υποψήφιο σημείο αντιστοίχισης με την άλλη στέρεο εικόνα, Σχήμα 3.1. Η απόφαση για το ποιο είναι το πιο σωστό ταίριασμα λαμβάνεται μετά διαδοχική επανάληψη των εξής βημάτων, στα πλαίσια δυναμικού προγραμματισμού του αλγορίθμου:

 Αξιοπιστία Ταιριάσματος: Η αξιοπιστία αυξάνεται όσο στη γειτονία ενός ζεύγους αντίστοιχων σημείων διατηρείται το υψηλό μέτρο συσχέτισης και η σωστή διάταξη αντίστοιχων σημείων.



Σχήμα 3.1: Βασική αρχή συσχέτισης σημείων σε στέρεο ζευγάρι εικόνων, [49].

2. Συνθήκη Μοναδικότητας: Υλοποίηση κατάλληλων συνθηκών για να διατηρείται η μοναδικότητα του κάθε ταιριάσματος προς κάθε κατεύθυνση, δηλαδή κάθε σημείο στη μία εικόνα έχει μοναδικό αντίστοιχο στη άλλη εικόνα. Σωστό ταίριασμα λοιπόν θεωρείται αυτό που ελαχιστοποιεί τη ενέργεια ταιριασμάτων όπως αυτή ορίστηκε στο [49].

## 3.2 Αναλλοίωτα Χαρακτηριστικά Εικόνας

Στην παρούσα παράγραφο θα ασχοληθούμε με το είδος των χαρακτηριστικών σε μια εικόνα υποψήφιων για ταίριασμα. Αναζητούνται χαρακτηριστικά ανεξάρτητα από τυχαίες μεταβολές της εικόνας στην κλίμακα και στους πιθανούς γεωμετρικούς μετασχηματισμούς της.

Το ταίριασμα εικόνων όπως έχει προαναφερθεί αποτελεί σημαντική πτυχή σε πολλά προβλήματα της όρασης υπολογιστών, όπως αναγνώριση αντικειμένων και σκηνών, υπολογισμό τρισδιάστατης δομής από πολλαπλές εικόνες, στέρεο αντιστοίχιση και παρακολούθηση κίνησης. Αναζητούνται χαρακτηριστικά τα οποία έχουν πληθώρα γεωμετρικών ιδιοτήτων έτσι ώστε να είναι κατάλληλα για ταίριασμα σε διαφορετικές εικόνες αντικειμένων ή σκηνών του φυσικού κόσμου. Βασικό στόχο αποτελεί η εύρεση χαρακτηριστικών που να παραμένουν αμετάβλητα σε μεταβολές περιστροφής και κλίμακας της εικόνας. Τέτοια χαρακτηριστικά είναι και τμηματικά αμετάβλητα σε μεταβολές της φωτεινότητας και του τρισδιάστατου σημείου του κέντρου της κάμερας.

Τα χαρακτηριστικά αυτά παρουσιάζουν μεγάλη σταθερότητα στη θέση τους τόσο σε χωρικές συντεταγμένες όσο και στο χώρο συχνοτήτων. Η χρήση τέτοιων χαρακτηριστικών στην ανάλυση και επεξεργασία εικόνας οδηγεί στη μείωση της πιθανότητας αλλοίωσής τους λόγω θορύβου ή λόγω επικαλυπτόμενων περιοχών στην εικόνα (occlusions). Σήμερα συναντώνται πληθώρα αποδοτικών αλγορίθμων για τον εντοπισμό τέτοιων χαρακτηριστικών που ειδικά σε εφαρμογές τρισδιάστατης ανακατασκευής και στέρεο επεξεργασίας δίνουν μεγάλο αριθμό επιτυχών ταιριασμάτων και συνεπώς πιο αξιόπιστα αποτελέσματα.

Το υπολογιστικό κόστος εύρεσης τέτοιων χαρακτηριστικών ελαχιστοποιείται με τη χρήση φίλτρων σε διαστρωματική παράθεση. Έτσι λοιπόν, διαδικασίες υπολογιστικά απαιτητικές εφαρμόζονται σε περιοχές της εικόνας όπου θεωρείται πιο πιθανή η εύρεση αμετάβλητων χαρακτηριστικών.

Ακολουθούν τα βασικά στάδια εύρεσης και υπολογισμού του συνόλου των παραπάνω χαρακτηριστικών.

 Εντοπισμός ακρότατων σε χώρο-κλίμακα: Στο πρώτο στάδιο υπολογισμού γίνεται αναζήτηση σε όλες τις κλίμακες και θέσεις της εικόνας. Έχει υλοποιηθεί με επιτυχία χρησιμοποιώντας συναρτήσεις Γκαουσιανών διαφορών για τον εντοπισμό πιθανών σημείων στην εικόνα αμετάβλητων σε κλίμακα και προσανατολισμό.

- Εντοπισμός χαρακτηριστικών σημείων κλειδιά, (keypoints): Σε κάθε υποψήφια θέση, προσαρμόζεται ένα αναλυτικό μοντέλο για τον προσδιορισμό θέσης και κλίμακας. Η επιλογή των χαρακτηριστικών σημείων, βασίζεται σε μέτρα ευστάθειας τους.
- 3. Ανάθεση προσανατολισμού: Μία ή περισσότερες τιμές προσανατολισμού ανατίθενται στα χαρακτηριστικά σημεία που βασίζονται στις τιμές κλίσης σε διάφορες κατευθύνσεις. Όλοι οι υπολογισμοί γίνονται στα μετασχηματισμένα χαρακτηριστικά της εικόνας σε σχέση με την κλίμακα, τον προσανατολισμό και τη θέση τους στην εικόνα. Με αυτόν τον τρόπο τα εν λόγω χαρακτηριστικά παραμένουν αμετάβλητα.
- 4. Περιγραφή χαρακτηριστικών σημείων: Οι μετρήσεις κλίσης της εικόνας συντελούνται για συγκεκριμένη τιμή κλίμακας και σε μία περιοχή γύρω από κάθε χαρακτηριστικό σημείο. Αυτές μετασχηματίζονται σε μία αναπαράσταση των χαρακτηριστικών σημείων για διάφορα επίπεδα παραμόρφωσης της εικόνας και αλλαγών στη φωτεινότητα.

Η προσέγγιση αυτή ονομάζεται Μετασχηματισμός Χαρακτηριστικών Αμετάβλητης Κλίμακας (Scale Invariant Feature Transform, SIFT). Μία πολύ σημαντική πτυχή της προσέγγισης αυτής είναι ότι δημιουργεί έναν μεγάλο αριθμό από χαρακτηριστικά που καλύπτει πυκνά όλη την εικόνα με πλήρες εύρος κλίμακας και θέσης. Βέβαια η ποσότητα σταθερών χαρακτηριστικών εξαρτάται πάντα από το περιεχόμενο της εικόνας και από την επιλογή αρκετών κρίσιμων παραμέτρων.

Σε εφαρμογές ταιριάσματος εικόνων και αναγνώρισης, τα SIFT χαρακτηριστικά εξάγονται αρχικά από ένα σύνολο εικόνων αναφοράς και αποθηκεύονται σε βάση δεδομένων. Τα χαρακτηριστικά κάθε καινούργιας εικόνας συγκρίνονται με αυτά της βάσης δεδομένων και παράγονται υποψήφια ταιριασμένα χαρακτηριστικά. Το εν λόγω ταίριασμα βασίζεται στην Ευκλείδια απόσταση των διανυσμάτων χαρακτηριστικών. Ο περιγραφέας των χαρακτηριστικών σημείων είναι αρκετά ευδιάκριτος, γεγονός που έχει ως συνέπεια μοναδικό χαρακτηριστικό να έχει σωστό ταίριασμα με πολύ καλή πιθανότητα μέσα σε μία μεγάλη βάση δεδομένων χαρακτηριστικών. Αντίθετα σε εικόνες με πολλά χαρακτηριστικά υφής, πολλά εντοπισμένα χαρακτηριστικά από το φόντο της εικόνας εμπλέκονται σε λανθασμένα ταιριάσματα με χαρακτηριστικά των υπόλοιπων εικόνων. Τα σωστά ταιριάσματα διαχωρίζονται μέσω της εύρεσης εκείνων των χαρακτηριστικών σημείων στις διαφορετικές εικόνες αναφοράς, τα οποία συμφωνούν σε τιμές προσανατολισμού, κλίμακας και θέσης.

### 3.2.1 Εντοπισμός Ακρότατων σε Χώρο - Κλίμακας

Όπως περιγράφηκε παραπάνω, ο εντοπισμός χαρακτηριστικών σημείων βασίζεται σε μία διαστρωματική προσέγγιση, σύμφωνα με την οποία εντοπίζονται υποψήφιες για ταίριασμα περιοχές στην εικόνα. Το πρώτο στάδιο εντοπισμού χαρακτηριστικών σημείων συνίσταται στον εντοπισμό θέσεων και κλίμακας που επαναλαμβάνονται σε διαφορετικές οπτικές γωνίες του ίδιου αντικειμένου. Ο εντοπισμός περιοχών στην εικόνα που είναι αμετάβλητες στις αλλαγές της κλίμακας μπορεί να επιτευχθεί με την αναζήτηση σταθερών στοιχείων για όλες τις πιθανές κλίμακες, χρησιμοποιώντας μία συνεχή συνάρτηση χώρου-κλίμακας ( scale-space ) [80]. Έχει δειχθεί από τους Koenderink [36] και Lindeberg [37] ότι κάτω από λογικές υποθέσεις ο μόνος πιθανός πυρήνας μετασχηματισμού χώρου-κλίμακας είναι η Γκαουσιανή συνάρτηση. Επομένως, η συνάρτηση σε χώρο-κλίμακας L(x, y, σ) για μία εικόνα παράγεται από τη συνέλιξη από μία μεταβλητής κλίμακας Γκαουσιανή, G(x, y, σ), με την εικόνα εισόδου, I(x, y):

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y)$$

όπου \*ο τελεστής συνέλιξης σε x και y συντεταγμένες στην εικόνα, και

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-(x^2 + y^2)/2\sigma^2}$$

Προκειμένου να εντοπιστούν με αποδοτικό τρόπο οι θέσεις σταθερών χαρακτηριστικών σημείων σε χώρο-κλίμακας, ο Lowe [21] πρότεινε τη χρήση ακρότατων σε χώρο-κλίμακας για συναρτήσεις διαφοράς Γκαουσιανών, συνελισσόμενες με την υπό εξέταση εικόνα:

$$D(x, y, \sigma) = (G(x, y, k\sigma) - G(x, y, \sigma)) * I(x, y) = L(x, y, k\sigma) - L(x, y, \sigma)$$
(3.2)

Παρατηρούμε ότι η  $D(x, y, \sigma)$  μπορεί να υπολογιστεί από τη διαφορά δύο κοντινών στην τιμή κλιμάκων, που διαφέρουν κατά έναν πολλαπλασιαστικό παράγοντα k. Η συνάρτηση  $D(x, y, \sigma)$  είναι ιδιαίτερα εύχρηστη, γιατί καταρχήν προκύπτει από ομαλοποιημένες εικόνες και σε συνδυασμό με το γεγονός ότι η συνάρτηση L απαιτείται να προσδιοριστεί για κάθε περίπτωση χώρου κλίμακας, η D υπολογίζεται με απλή αφαίρεση ομαλοποιημένων εικόνων.

Επιπρόσθετα, η διαφορά Γκαουσιανών συναρτήσεων παρέχει μία κοντινή προσέγγιση στη Λαπλασιανή της Γκαουσιανής με ομαλοποιημένη κλίμακα,  $\sigma^2 \nabla^2 G$ , όπως μελετήθηκε από τον Lindeberg, [37]. Ο Lindeberg απέδειξε

ότι η ομαλοποίηση της Λαπλασιανής με ένα παράγοντα  $\sigma^2$  είναι απαραίτητη προκειμένου να μείνει η κλίμακα αμετάβλητη. Μετά από αναλυτικές πειραματικές συγκρίσεις, προκύπτει ότι τα μέγιστα και τα ελάχιστα του  $\sigma^2 \nabla^2 G$ οδηγούν σε περισσότερο σταθερά και αμετάβλητα χαρακτηριστικά εικόνας σε σχέση με άλλες πιθανές συναρτήσεις εικόνας όπως: η κλίση, η Hessian ή ο εντοπιστής γωνιών κατά Harris. Η σχέση μεταξύ D και  $\sigma^2 \nabla^2 G$  μπορεί να κατανοηθεί από τη συνάρτηση διάχυσης θερμότητας, παραμετροποιημένη ως προς  $\sigma$ :

$$\frac{\partial G}{\partial \sigma} = \sigma \nabla^2 G$$

Είναι φανερό ότι η ποσότητα  $\nabla^2 G$  μπορεί να υπολογιστεί από πεπερασμένες διαφορές της  $\frac{\partial G}{\partial \sigma}$ , χρησιμοποιώντας κοντινές κλίμακες  $k\sigma$  και  $\sigma$ 

$$\sigma \nabla^2 G = \frac{\partial G}{\partial \sigma} \approx \frac{G(x, y, k\sigma) - G(x, y, \sigma)}{k\sigma - \sigma}$$

και επομένως

$$G(x, y, k\sigma) - G(x, y, \sigma) \approx (k - 1)\sigma^2 \nabla^2 G$$
(3.3)

Η παραπάνω εξίσωση αποδεικνύει ότι όταν η διαφορά Γκαουσιανών συναρτήσεων περιέχει κλίμακες που διαφέρουν κατά ένα σταθερό παράγοντα, αυτός συνεισφέρει μαζί με τη  $\sigma^2$  στην ομαλοποίηση κλίμακας που είναι απαραίτητη για την Λαπλασιανή αμετάβλητης κλίμακας. Ο παράγοντας (k-1)στην εξίσωση (3.3) είναι σταθερός για όλες τις κλίμακες και επομένως δεν επηρεάζει τη θέση των ακρότατων. Το λάθος προσέγγισης τείνει στο μηδέν όσο το k τείνει στο 1. Στη πράξη έχει παρατηρηθεί ότι η προσέγγιση δεν έχει επιρροή στην ευστάθεια του εντοπισμού ακρότατων ακόμη και για σημαντικές διαφορές στην κλίμακα, όπως  $k = \sqrt{2}$ . Μία αποδοτική προσέγγιση στην κατασκευή της  $D(x, y, \sigma)$  φαίνεται στο Σχήμα 3.2. Η αρχική εικόνα συνελίσσεται διαδοχικά με Γκαουσιανές συναρτήσεις, για να παραχθούν εικόνες που διαφέρουν κατά σταθερό παράγοντα κλίμακας k σε κλίμακα-χώρο, όπως φαίνονται στοιβαγμένες στην αριστερή στήλη. Συνήθως επιλέγεται ο χωρισμός του χώρου-κλίμακας ανά οκτάβα και ανά διπλασιασμό της σ σε ακέραιο αριθμό από διαστήματα s τέτοια ώστε  $k = 2^{\frac{1}{s}}$ . Με αυτόν τον τρόπο δημιουργείται στοίβα s + 3 ομαλοποιημένων εικόνων για κάθε οκτάβα, με συνέπεια ο εντοπισμός ακρότατων να καλύπτει μία πλήρη οκτάβα. Οι διαφορές από Γκαουσιανές με διαδοχικές κλίμακες, φαίνονται στη δεξιά πλευρά του Σχήματος 3.2



Σχήμα 3.2: Για κάθε οκτάβα της χώρου-κλίμακας, η αρχική εικόνα συνελίσεται επαναλαμβανόμενα με Γκαουσιανές για να παραχθούν τα αποτελέσματα που φαίνονται αριστερά. Διαδοχικές τέτοιες εικόνες αφαιρούνται για να παραχθούν τα αποτελέσματα στη δεξιά πλευρά. Μετά από κάθε οκτάβα, η εικόνα υποδειγματολειπτείται με παράγοντα 2, και η ίδια διαδικασία επαναλαμβάνεται [21].

#### 3.2.2 Εντοπισμός Τοπικών Ακρότατων

Προκειμένου να εντοπιστούν τοπικά ελάχιστα ή μέγιστα της  $D(x, y, \sigma)$ , κάθε εικονοστοιχείο που δειγματοληπτείται συγκρίνεται με τα οκτώ γειτονικά του της τρέχουσας εικόνας και με τα εννέα γειτονικά του στις εικόνες μίας κλίμακας μεγαλύτερης και μικρότερης (Σχήμα 3.3). Το στοιχείο επιλέγεται μόνο εάν έχει μεγαλύτερη ή μικρότερη τιμή από όλα τα γειτονικά του ανάλογα με το είδος του ακρότατου αναζήτησης. Το κόστος αναζήτησης είναι σχετικά μικρό μιας και τα περισσότερα υπό εξέταση σημεία απαλείφονται μετά τις πρώτες αναζητήσεις.

Ένα επιπλέον σημαντικό ζήτημα είναι ο προσδιορισμός της συχνότητας δειγματοληψίας της εικόνας, που χρειάζεται για τον αξιόπιστο εντοπισμό των ακρότατων. Δυστυχώς δεν υπάρχει ελάχιστη απόσταση μεταξύ των δειγμάτων έτσι ώστε να εξασφαλίζεται ο εντοπισμός όλων των ακρότατων, καθώς τα ακρότατα μπορεί να βρίσκονται αρκετά κοντά με τυχαίο τρόπο. Επομένως αναζητείται μία λύση που οδηγεί στο συγκερασμό αποδοτικότητας και συμβατότητας με τα δεδομένα. Πειράματα σε αυτόν τον τομέα έχουν δείξει ότι ακρότατα που βρίσκονται αρκετά κοντά είναι εξαιρετικά ασταθή σε μικρές μεταβολές της εικόνας. Προτείνεται λοιπόν η αναζήτηση ακρότατων σε μεγάλο



Σχήμα 3.3: Μέγιστα και Ελάχιστα των εικόνων που έχουν προκύψει από διαφορά Γκαουσιανών και έχουν εντοπιστεί από τη σύγκριση ενός εικοστοιχείου (συμβολίζεται με X) με τους 26 γειτονικά του σε περιοχές 3×3 στη συγκεκριμένη ή κοντινή κλίμακα (συμβολίζονται με κύκλο) [21].

εύρος συχνοτήτων και η επιλογή εκείνων με την πιο σταθερή συμπεριφορά τους κατά τη διαδικασία του ταιριάσματος. Ο πειραματικός προσδιορισμός της συχνότητας δειγματοληψίας, για την οποία μεγιστοποιείται η ευστάθεια σε χώρο-κλίμακας των ακρότατων ως προς τις τυχαίες μεταβολές των εικόνων, προκύπτει από τη χρήση γραφημάτων ποσοστού επαναληψιμότητας και ποσότητας χαρακτηριστικών σημείων στην εικόνα ως προς τον αριθμό των δειγματοληπτούμενων κλιμάκων ανά οκτάβα. Τα εν λόγω γραφήματα [21] έχουν προκύψει από μία πληθώρα ταιριασμάτων από εικόνες που περιλαμβάνουν εξωτερικές σκηνές, αεροφωτογραφίες, βιομηχανικές φωτογραφίες, ανθρώπινα πρόσωπα κ.α. Γενικά έχει παρατηρηθεί ότι το θέμα της φωτογραφίας δεν έχει επιρροή στα αποτελέσματα. Κάθε εικόνα υφίσταται διάφορους μετασχηματισμούς που περιλαμβάνουν περιστροφή, κλιμακοποίηση, αφινική παραμόρφωση, αλλαγές στην φωτεινότητα, αντίθεση καθώς και προσθήκη θορύβου. Επειδή οι αλλαγές είναι συνθετικές υπάρχει ακριβής πρόβλεψη για τη θέση και τη μορφή των διαφόρων χαρακτηριστικών στη νέα εικόνα μετά το μετασχηματισμό. Το γεγονός αυτό επιτρέπει μετρήσεις που σχετίζονται με την επαναληψημότητα και την ακρίβεια θέσης των χαρακτηριστικών.

Επιπρόσθετα, γίνεται χρήση διαγραμμάτων που αναπαριστούν την επιρροή της μεταβολής των κλιμάκων ανά οκτάβα στην συνάρτηση της εικόνας, η οποία δειγματοληπτείται πριν τη διαδικασία εντοπισμού ακρότατων. Σε αυτή τη περίπτωση η εικόνα δειγματοληπτείται εκ νέου εφαρμόζοντας περιστροφή κατά τυχαία γωνία και κλιμακοποίηση κατά τυχαία ποσότητα μεταξύ 0.2 και 0.9 φορές του αρχικού μεγέθους της εικόνας. Είναι αξιοπρόσεκτο ότι η αύξηση των δειγματοληπτούμενων κλιμάκων δεν οδηγεί και σε ανάλογη αύξηση της επαναληψιμότητας των χαρακτηριστικών. Το γεγονός αυτό συμβαίνει γιατί τα επιπλέον χαρακτηριστικά που εντοπίζονται με την αύξηση των χρησιμοποιούμενων κλιμάκων είναι ασταθή και επομένως είναι λιγότερο πιθανό να εντοπιστούν στη μετασχηματισμένη εικόνα. Αυτό φαίνεται και με γραφικό τρόπο στην εργασία του Lowe [21]. Ο αριθμός των εντοπισμένων χαρακτηριστικών σημείων αυξάνεται όσο αυξάνεται ο αριθμός των δειγματοληπτούμενων κλιμάκων. Συμπερασματικά, πληθώρα πειραμάτων έχουν καταλήξει ότι οι συναρτήσεις χώρου-κλίμακας διαφοράς Γκαουσιανών έχουν μεγάλο αριθμό ακρότατων και δεν είναι αποδοτικός, ο εντοπισμός του συνόλου τους, από υπολογιστική άποψη. Είναι δυνατός όμως ο εντοπισμός των πιο σταθερών και χρήσιμων χαρακτηριστικών ακόμη και στην περίπτωση αραιής δειγματοληψίας των κλιμάκων. Όπως ακριβώς εντοπίστηκε η συχνότητα δειγματοληψίας ανά οκτάβα του χώρου-κλίμακας, πρέπει να προσδιοριστεί η συχνότητα δειγματοληψίας στο χώρο της εικόνας με βάση την κλίμακα ομαλοποίησης. Δοσμένων των ακρότατων τυχαία κοντά στις τιμές τους, είναι δυνατόν να υπάρξει μία συγκριτική μελέτη ανάμεσα στη συχνότητα δειγματοληψίας και του ρυθμού εντοπισμού χαρακτηριστικών. Η ποσότητα ομαλοποίησης, σ, που εφαρμόζεται σε κάθε επίπεδο εικόνας κατά τη δημιουργία κλίμακας-χώρου αναπαράστασης για μία οκτάβα, προσδιορίζεται και αυτή με πειραματικό τρόπο από το γράφημα του ποσοστού επαναληψημότητας των χαρακτηριστικών σημείων, ως προς τη ποσότητα ομαλοποίησης  $\sigma$  ανά οκτάβα.

Είναι προφανές, ότι εάν η εικόνα ομαλοποιηθεί πριν τον εντοπισμό ακρότατων αποκόπτονται αποδοτικά οι υψηλότερες χωρικές συχνότητες. Επίσης στην προσπάθεια να γίνει πλήρη χρήση της εικόνας εισόδου, η εικόνα εισόδου είναι δυνατόν να επεκταθεί με τέτοιο τρόπο ώστε να εντοπίζονται περισσότερα χαρακτηριστικά σε σχέση με την αρχική εικόνα. Είναι δυνατόν λοιπόν να διπλασιαστεί το μέγεθος της αρχικής εικόνας, χρησιμοποιώντας γραμμική παρεμβολή, πριν κατασκευαστεί το πρώτο επίπεδο της πυραμίδας. Θεωρούμε ότι η αρχική εικόνα έχει θόλωση τουλάχιστον  $\sigma = 0.5$  (η ελάχιστη που είναι απαραίτητη για την αποφυγή σημαντικής επικάλυψης), οπότε η επόμενη εικόνα σύμφωνα με την πυραμιδική διάταξη έχει  $\sigma = 1$ . Αυτό σημαίνει ότι η ομαλοποίηση είναι χρήσιμη πριν τη δημιουργία της πρώτης οκτάβας στην αναπαράσταση χώρου-κλίμακας. Από τη στιγμή που το υποψήφιο χαρακτηριστικό σημείο βρεθεί μετά από σύγκριση με τα γειτονικά του εικονοστοιχεία, το επόμενο βήμα είναι η αναλυτική προσαρμογή του χαρακτηριστικού σημείου στα γειτονικά δεδομένα κλίμακας, θέσης και λόγου στοιχειωδών καμπυλοτήτων. Με αυτόν τον τρόπο απορρίπτονται σημεία χαμηλής αντίθεσης (και επομένως ευαίσθητα σε θόρυβο), ή σημεία τοποθετημένα με ασταθή τρόπο κατά μήκος των ακμών της εικόνας.

Πρόσφατα ο Brown και ο Lowe, [10], ανέπτυξαν μία μέθοδο για να προσαρμόζουν μία τρισδιάστατη τετραγωνική συνάρτηση σε τοπικά εικονοστοιχεία προκειμένου να προσδιοριστεί με παρεμβολή η θέση του μέγιστου. Πειράματα έχουν αποδείξει ότι η διαδικασία αυτή συντελεί στην απόκτηση πιο αξιόπιστων και σταθερών ακρότατων. Η εν λόγω προσέγγιση χρησιμοποιεί ανάπτυγμα Taylor της συνάρτησης κλίμακας-χώρου,  $D(x, y, \sigma)$ , μετατοπισμένη έτσι ώστε η αρχή των αξόνων να συμπίπτει με το δείγμα εικοπνοστοι-χείου.

$$D(\mathbf{x}) = D + \frac{\partial D^T}{\partial \mathbf{x}} \mathbf{x} + \frac{1}{2} \mathbf{x}^T \frac{\partial^2 D}{\partial \mathbf{x}^2} \mathbf{x}$$
(3.4)

όπου D και οι παράγωγοι του υπολογίζονται στα σημεία δειγματοληψίας και  $\mathbf{x} = (x, y, \sigma)^T$  αντιπροσωπεύει την απόκλιση από αυτό το σημείο. Η θέση του ακρότατου,  $\hat{\mathbf{x}}$ , προσδιορίζεται μέσω του υπολογισμού της παραγώγου της συνάρτησης D ως προς x, η οποία εξισώνεται κατά τα γνωστά με μηδέν.

$$\widehat{\mathbf{x}} = -\frac{\partial^2 D^{-1}}{\partial \mathbf{x}^2} \frac{\partial D}{\partial \mathbf{x}}$$
(3.5)

Η παράγωγος και η Hessian της D προσεγγίζεται σύμφωνα με τους Brown και Lowe, [10], χρησιμοποιώντας διαφορές γειτονικών δειγμάτων εικονοστοιχείων. Το προκύπτον γραμμικό  $3 \times 3$  σύστημα μπορεί να επιλυθεί με ελάχιστο κόστος. Εάν η απόκλιση  $\hat{\mathbf{x}}$  είναι μεγαλύτερη από 0.5 σε κάθε διάσταση τότε το ακρότατο βρίσκεται εγγύτερα σε άλλο δείγμα εικονοστοιχείου. Η τελική απόκλιση  $\hat{\mathbf{x}}$  προστίθεται στη θέση του δείγματος για να προκύψει η εκτίμηση της θέσης του ακρότατου μέσω παρεμβολής. Η τιμή της συνάρτησης στο ακρότατο,  $D(\hat{\mathbf{x}})$ , είναι χρήσιμη στην απόρριψη ασταθών ακρότατων με χαμηλή αντίθεση. Αυτό επιτυγχάνεται με την υπέρθεση των εξισώσεων (3.4) και (3.5), η οποία καταλήγει στη σχέση:

$$D(\widehat{\mathbf{x}}) = D + \frac{1}{2} \frac{\partial D^T}{\partial \mathbf{x}} \widehat{\mathbf{x}}$$

Το Σχήμα 3.4 αναπαριστά την επίδραση της επιλογής χαρακτηριστικών σημείων σε μία φυσική εικόνα. Προκειμένου να αποφευχθεί ο εντοπισμός σημείων υφής από το φόντο της εικόνας έχει χρησιμοποιηθεί εικόνα χαμηλής ανάλυσης. Τα χαρακτηριστικά σημεία έχουν αναπαρασταθεί ως διανύσματα για τη θέση, την κλίμακα και τον προσανατολισμό. Το Σχήμα 3.4(*a*) δείχνει την αρχική εικόνα, το Σχήμα 3.4(*b*) δείχνει την πληθώρα χαρακτηριστικών σημείων στα εντοπισμένα ακρότατα της συνάρτησης Γκαουσιανών διαφορών. Το Σχήμα 3.4(*c*) δείχνει τα χαρακτηριστικά σημεία που απέμειναν μετά την αποκοπή εκείνων με τιμή  $|D(\hat{\mathbf{x}})|$  μικρότερη από ένα προκαθορισμένο κατώφλι από πριν προσδιορισμένο. Στην επόμενη παράγραφο, θα εξηγηθεί το Σχήμα3.4(d).



Σχήμα 3.4: Το διάγραμμα αυτό αναπαριστά τη σταδιακή επιλογή χαρακτηριστικών σημείων. (a): Η αρχική εικόνα 233 × 189 εικονοστοιχείων. (b): Οι αρχικές θέσεις των 832 χαρακτηριστικών σημείων στα μέγιστα και ελάχιστα της συνάρτηση Γκαουσιανών διαφορών. Τα χαρακτηριστικά σημεία απεικονίζονται ως διανύσματα που αναπαριστούν κλίμακα, προσανατολισμό και θέση. (c): Μετά την εφαρμογή κατωφλιοποίησης για ελάχιστες τιμές αντίθεσης, απομένουν 729 χαρακτηριστικά σημεία. (d): Ο τελικός αριθμός 536 χαρακτηριστικών σημείων υφίσταται επιπλέον κατωφλιοποίηση στο λόγο των κύριων καμπυλοτήτων [21].

### 3.2.3 Αποκρίσεις Ακμών

Για λόγους ευστάθειας δεν είναι απαραίτητο να αποκοπούν μόνο τα χαρακτηριστικά σημεία που εντοπίζονται σε περιοχές της εικόνας με χαμηλή αντίθεση. Η συνάρτηση διαφοράς Γκαουσιανών εμφανίζει ισχυρές αποκρίσεις κατά μήκος των ακμών, ακόμα και αν η θέση στην περιοχή των ακμών είναι ελλιπώς προσδιορισμένη και άρα εξαιρετικά ευαίσθητη σε μικρές ποσότητες θορύβου. Μία ελλιπώς ορισμένη κορυφή της συνάρτησης διαφοράς Γκαουσιανών έχει μεγάλη τιμή καμπυλότητας κατά μήκος των ακμών αλλά μικρή στην κάθετη διεύθυνση. Οι θεμελιώδεις καμπυλότητες μπορούν να υπολογιστούν με τη βοήθεια ενός 2 × 2 πίνακα Hessian, **Η** για τη θέση και την κλίμακα του χαρακτηριστικού σημείου.

$$\mathbf{H} = \begin{pmatrix} D_{xx} & D_{xy} \\ D_{xy} & D_{yy} \end{pmatrix}$$
(3.6)

Οι παράγωγοι υπολογίζονται από τις διαφορές γειτονικών δειγμάτων εικονοστοιχείων. Οι ιδιοτιμές του **H** είναι ανάλογες των θεμελιωδών καμπυλοτήτων της *D*. Σύμφωνα με την προσέγγιση στο [31], είναι δυνατόν να αποφευχθεί ο ακριβής υπολογισμός των ιδιοτιμών, αφού χρειάζεται η γνώση μόνο του λόγου τους. Έστω α η ιδιοτιμή με το μεγαλύτερο πλάτος και β η ιδιοτιμή με το μικρότερο πλάτος, τότε προκύπτει:

$$Tr(\mathbf{H}) = D_{xx} + D_{yy} = \alpha + \beta$$
$$Det(\mathbf{H}) = D_{xx}D_{yy} - (D_{xy})^2 = \alpha\beta$$

Έστω r ο λόγος μεταξύ της μεγαλύτερης σε πλάτος ιδιοτιμής και της αντίστοιχης μικρότερης, δηλαδή:  $\alpha = r\beta$ . Αυτό οδηγεί:

$$\frac{Tr(\mathbf{H})^2}{Det(\mathbf{H})} = \frac{(\alpha + \beta)^2}{\alpha\beta} = \frac{(r+1)^2}{r}$$

Η παραπάνω σχέση εξαρτάται μόνο από το λόγο των ιδιοτιμών και όχι από τις μεμονωμένες τιμές της κάθε ιδιοτιμής. Η ποσότητα  $\frac{(r+1)^2}{r}$  γίνεται ελάχιστη όταν οι δύο ιδιοτιμές είναι ίσες και αυξάνεται με το r. Άρα, προκειμένου να ελεχθεί εάν ο λόγος θεμελιωδών καμπυλοτήτων είναι κάτω από κάποιο κατώφλι r, αρκεί να ελεγχθεί:

$$\frac{Tr(\mathbf{H})^2}{Det(\mathbf{H})} < \frac{(r+1)^2}{r}$$

Η εφαρμογή της παραπάνω κατωφλιοποίησης στο λόγο θεμελιωδών καμπυλοτήτων φαίνεται στο Σχήμα 3.4(d).

### 3.2.4 Συνάρτηση Τοπικής Περιγραφής Εικόνας

Αναθέτοντας μία τιμή προσανατολισμού σε κάθε χαρακτηριστικό σημείο, που να βασίζεται σε τοπικές ιδιότητες μιας εικόνας, επιτυγχάνεται η δημιουργία ενός περιγραφέα χαρακτηριστικού σημείου συναρτήσει του προσανατολισμού. Με αυτόν τον τρόπο ο περιγραφέας αποκτά αμεταβλητότητα στις περιστροφές της εικόνας. Ακολουθώντας τυπικές μεθόδους η κλίμακα του χαρακτηριστικού σημείου χρησιμοποιείται στην επιλογή Γκαουσιανής ομαλοποιημένης εικόνας L με την πιο κοντινή κλίμακα, οπότε όλοι οι υπολογισμοί γίνονται με τρόπο ανεξάρτητο της κλίμακας. Για κάθε δείγμα εικόνας, L(x, y), σε αυτή την κλίμακα, το πλάτος της κλίσης, m(x, y), και ο προσανατολισμός,  $\theta(x, y)$  υπολογίζονται από πριν χρησιμοποιώντας διαφορές εικονοστοιχείων:

$$\begin{split} m(x,y) &= \sqrt{(L(x+1,y) - L(x-1,y))^2 + (L(x,y+1) - L(x,y-1))^2} \\ \theta(x,y) &= \tan^{-1} \left( \frac{L(x,y+1) - L(x,y-1)}{L(x+1,y) - L(x-1,y)} \right) \end{split}$$

Δημιουργείται λοιπόν ένα ιστόγραμμα από τους προσανατολισμούς κλίσης των δειγμάτων στην περιοχή ενός χαρακτηριστικού σημείου. Το ιστόγραμμα προσανατολισμού καλύπτει όλο το εύρος των 360°. Κάθε δείγμα που προστίθεται στο ιστόγραμμα σταθμίζεται από το πλάτος της κλίσης του και από ένα κυκλικό παράθυρο Γκαουσιανής στάθμισης.

Οι προηγούμενες διαδικασίες θέτουν μία χωρική τιμή, μία τιμή κλίμακας, και μία τιμή προσανατολισμού σε κάθε χαρακτηριστικό σημείο. Οι παράμετροι αυτοί δημιουργούν ένα επαναλαμβανόμενο διδιάστατο σύστημα αναφοράς μέσω του οποίου είναι εφικτή η τοπική περιγραφή μιας εικόνας, και επομένως η ανεξαρτησία της περιγραφής από αυτές. Το επόμενο βήμα είναι ο υπολογισμός περιγραφικής αναπαράστασης για την συγκεκριμένη περιοχή της εικόνας η οποία να είναι διακριτή και όσο το δυνατόν ανεξάρτητη στις υπόλοιπες μεταβολές στην εικόνα, όπως αλλαγές στην τρισδιάστατη θέση του σημείου παρατήρησης ή αλλαγές στην φωτεινότητα. Η προφανής προσέγγιση συνίσταται στη τοπική δειγματοληψία φωτεινότητας γύρω από την περιοχή του χαρακτηριστικού σημείου με κατάλληλη κλίμακα και κατόπιν στη διαδικασία ταιριάσματος με ομαλοποιημένο μέτρο ανομοιότητας. Βασικό μειονέκτημα της εν λόγω προσέγγισης αποτελεί το γεγονός ότι το απλό μέτρο ανομοιότητας είναι εξαιρετικά ευαίσθητο σε αλλαγές στο τρισδιάστατο σημείο παρατήρησης καθώς επίσης και σε μη-εύρωστους μετασχηματισμούς. Μία καλύτερη προσέγγιση έχει προταθεί από τους Edelman, Intrator, Poggio, [24]. Η προσέγγιση αυτή βασίζεται σε μοντέλα βιολογικής όρασης. Σύμφωνα με αυτά, γίνεται υπολογισμός της κλίσης σε συγκεκριμένο προσανατολισμό και χωρική συχνότητα. Η θέση της κλίσης δεν προσδιορίζεται με ακρίβεια, αντίθετα είναι δυνατόν να υφίσταται μικρές μετατοπίσεις. Οι Edelman et al πρότειναν ότι η αντιπροσωπευτική συνάρτηση περιγραφής τοπικής πληροφορίας που προκύπτει από τους παραπάνω υπολογισμούς, μπορεί να αξιοποιηθεί στη διαδικασία του ταιριάσματος ή στην αναγνώριση αντικειμένων.

### 3.2.5 Δομές Περιγραφής Χαρακτηριστικών Σημείων

Το Σχήμα 3.5 αναπαριστά τη διαδικασία υπολογισμού της περιγραφικής συνάρτησης ενός χαρακτηριστικού σημείου. Καταρχήν τα πλάτη κλίσης, οι



Σχήμα 3.5: Μία δομή περιγραφής χαρακτηριστικού σημείου βασίζεται σε υπολογισμούς πλάτους και προσανατολισμού κλίσης σε σημεία της υπό εξέταση εικόνας σε μία περιοχή γύρω από τη θέση του υποψήφιου χαρακτηριστικού σημείου, όπως φαίνεται στο αριστερό μέρος. Κατόπιν δημιουργούνται ιστογράμματα προσανατολισμού από τα σημεία αυτά και αναπαριστούν τα περιεχόμενα υποπεριοχών  $4 \times 4$ , όπως φαίνεται στο δεξιό μέρος. Το μήκος κάθε διανύσματος αντιστοιχεί στο άθροισμα των πλατών της κλίσης σε αυτήν την περιοχή. Το διάγραμμα αυτό δείχνει έναν πίνακα περιγραφής  $2 \times 2$ , όπως έχει προκύψει από σύνολο  $8 \times 8$  δειγμάτων [21].

προσανατολισμοί δειγματοληπτούνται σε μία κοντινή περιοχή του χαρακτηριστικού σημείου σε συγκεκριμένη κλίμακα, έτσι ώστε να προσδιοριστεί το επίπεδο Γκαουσιανής ασάφειας της εικόνας. Προκειμένου να επιτευχθεί ανεξαρτησία προσανατολισμού, οι συντεταγμένες του περιγραφέα και οι τιμές προσανατολισμού της κλίσης, περιστρέφονται σε σχέση με τον προσανατολισμό του χαρακτηριστικού σημείου. Λόγοι υπολογιστικής επίδοσης επιβάλουν τον υπολογισμό των κλίσεων εκ των προτέρων και για όλα τα επίπεδα της πυραμίδας. Η κάθε υπολογισμένη κλίση έχει αναπαρασταθεί στο αριστερό μέρος του Σχήματος 3.5 με τη μορφή μικρών διανυσμάτων (βέλη) για κάθε θέση δείγματος. Μία Γκαουσιανή συνάρτηση στάθμισης με  $\sigma$  ίσο με το μισό του πλάτους του παραθύρου του περιγραφέα, χρησιμοποιείται για να σταθμίσει το πλάτος για κάθε ένα δείγμα. Η διαδικασία αυτή φαίνεται στο στρογγυλό παράθυρο του Σχήματος 3.5. Ο σκοπός χρήσης του Γκαουσιανού παραθύρου είναι η αποφυγή απότομων αλλαγών του περιγραφέα μέσω μικρών, ομαλών αλλαγών στη θέση του παραθύρου. Επιπλέον λόγο αποτελεί η λιγότερη έμφαση που πρέπει να δοθεί σε τιμές κλίσης μακριά από το κέντρο του περιγραφέα, μιας και αυτές οι τιμές οδηγούν σε περισσότερα λάθη θέσης των περιγραφέων.

Ο περιγραφέας του χαρακτηριστικού σημείου φαίνεται στο δεξιό μέρος του Σχήματος 3.5. Όπως είναι φανερό επιτρέπει μεγάλες μεταβολές στον

προσανατολισμό, δημιουργώντας ιστογράμματα προσανατολισμού σε 4 × 4 περιοχές δειγμάτων. Το γράφημα παρουσιάζει οχτώ κατευθύνσεις για κάθε ιστόγραμμα προσανατολισμού, με το μήκος κάθε βέλους να αντιστοιχεί στο πλάτος κάθε εισόδου του ιστογράμματος. Κάθε δείγμα κλίσης από το αριστερό μέρος του Σχήματος 3.5 μπορεί να ματαβληθεί κατά τέσσερεις θέσεις και παράλληλα να συνεισφέρει στο ίδιο ιστόγραμμα του δεξιού μέρους. Με αυτόν τον τρόπο ικανοποιείται ο στόχος των περισσότερων τοπικών μετατοπίσεων. Είναι σημαντικό επίσης να αποφευχθούν όλες οι απότομες και τυχαίες αλλαγές στην αναπαράσταση του περιγραφέα στα όρια του παραθύρου, καθώς τα δείγματα μετατοπίζονται από το ένα ιστόγραμμα στο άλλο και από τη μία τιμή προσανατολισμού σε άλλη. Επομένως χρησιμοποιείται τριγραμμική παρεμβολή για να διαχυθεί η τιμή για κάθε δείγμα κλίσης σε γειτονικές ράβδους ιστογραμμάτων. Με άλλα λόγια κάθε εγγραφή στο ιστόγραμμα πολλαπλασιάζεται με ένα βάρος 1-d για κάθε διάσταση, όπου d είναι η απόσταση του δείγματος από την κεντρική τιμή της ράβδου, όπως αυτή μετράται σε μονάδες ράβδων ιστογραμμάτων. Η μαθηματική αναπαράσταση της συνάρτησης περιγραφής αποτελείται από ένα διάνυσμα που περιέχει τιμές προσανατολισμού από όλες τις ράβδους ιστογραμμάτων, που αντιστοιχούν στο μήκος των βελών στη δεξιά πλευρά του Σχήματος 3.5. Το συγκεκριμένο διάγραμμα παρουσιάζει έναν 2 × 2 πίνακα ιστογραμμάτων προσανατολισμού.

Τέλος το διάνυσμα χαρακτηριστικών τιμών μετασχηματίζεται με τέτοιο τρόπο ώστε να μειώνει τις συνέπειες από την επίδραση των μεταβολών φωτεινότητας. Καταρχήν το μήκος ομαλοποιείται σε μοναδιαίο μήκος. Σε κάθε αλλαγή στην αντίθεση της εικόνας, τα εικονοστοιχεία της πολλαπλασιάζονται με μία σταθερά, όπως επίσης πολλαπλασιάζονται με την ίδια σταθερά και οι τιμές κλίσης. Πολλές φορές όμως η ομαλοποιημένη μορφή των διανυσμάτων αποκόπτει τις αλλαγές στην αντίθεση. Επιπρόσθετα, μία ενδεχόμενη αλλαγή στη φωτεινότητα δε θα επηρεάσει τις τιμές κλίσης, αφού αυτές υπολογίζονται από διαφορές εικονοστοιχείων. Επομένως η συνάρτηση περιγραφής είναι ανεξάρτητη από τις αφινικές μεταβολές της φωτεινότητας. Επίσης προκειμένου να μειωθεί η επίδραση μεγάλων πλατών κλίσης χρησιμοποιείται κατωφλιοποίηση των πλατών και εκ νέου ομαλοποίηση των τιμών για μοναδιαίο μήκος διανυσμάτων. Με αυτόν τον τρόπο αποκτά ιδιαίτερη αξία στην κατασκευή συνάρτησης περιγραφής, η διασπορά τιμών προσανατολισμού συγκριτικά με τα πλάτη κλίσεων.

# 3.3 Αυτόματη Βαθμονόμηση Κάμερας, Επανόρθωση Εικόνων

Στο προηγούμενο κεφάλαιο αναλύσαμε τη βαθμονόμηση κάμερας από επίπεδα αντικείμενα βαθμονόμησης. Στην παρούσα παράγραφο θα αναλύσουμε την περίπτωση όπου δεν έχει προηγηθεί βαθμονόμηση αλλά γνωρίζουμε το σύνολο των ζευγαριών αντίστοιχων σημείων. Αναφορικά με την επανόρθωση του ζεύγους των εικόνων, αναζητούμε τον μετασχηματισμό εκείνον ο οποίος τοποθετεί τις επιπολικές γραμμές παράλληλα μεταξύ τους. Έστω  $H_r$  και  $H_l$  οι άγνωστοι πίνακες επανόρθωσης για δεξιά και αριστερή εικόνα αντίστοιχα του στέρεο ζευγαριού. Τα μετασχηματισμένα αντίστοιχα σημεία  $m_r$  και  $m_l$ πρέπει να ικανοποιούν την επιπολική γεωμετρία του επανορθωμένου ζευγαριού εικόνων σύμφωνα με [64] και [68]:

$$(H_r m_{r_j})^T \overline{F}(H_l m_{l_j}) = 0 \tag{3.7}$$

Ο θεμελιώδης πίνακας είναι:  $F = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}$ . Με την κατάλληλη παρα-

μετροποίηση των  $H_r$  και  $H_l$  και θεωρώντας ότι έχουν την ίδια δομή με τη περίπτωση γνωστής βαθμονόμησης έχουμε:

$$H_r = K_{nr} R_r K_{or}^{-1} \qquad H_l = K_{nl} R_l K_{ol}^{-1}$$
(3.8)

Οι παραπάνω ομοιογραφίες αντιπροσωπεύουν αντιστοιχίες σημείων εικόνας που ανήκουν σε επίπεδο τοποθετημένο στο άπειρο (σημεία φυγής). Κάθε πίνακας επανόρθωσης εξαρτάται από πέντε (εσωτερικές) και τρεις (περιστροφή) άγνωστες παραμέτρους. Ο αριθμός των παραμέτρων μπορεί να μειωθεί κάνοντας τις εξής παραδοχές: έλλειψη παραμόρφωσης, κύριο σημείο στο κέντρο της εικόνας και αναλογία εικόνας ίση με 1. Άρα οι άγνωστοι παράμετροι που απομένουν είναι τα εστιακά μήκη α<sub>r</sub> και α<sub>l</sub>. Ακολουθούν οι απλοποιημένοι πίνακες εσωτερικών παραμέτρων για κάθε εικόνα της στέρεο διάταξης. Οι απλοποιημένοι πίνακες εσωτερικών παραμέτρων ακολουθούν:

$$K_{or} = \begin{pmatrix} \alpha_r & 0 & w/2 \\ 0 & \alpha_r & h/2 \\ 0 & 0 & 1 \end{pmatrix} \qquad K_{ol} = \begin{pmatrix} a_l & 0 & w/2 \\ 0 & a_l & h/2 \\ 0 & 0 & 1 \end{pmatrix}$$
(3.9)

Οι μεταβλητές w και h είναι το πλάτος και το ύψος μετρημένες σε pixels της εικόνας. Η βέλτιστη εκτίμηση των παραμέτρων γίνεται με χρήση επαναληπτικής μεθόδου ελαχιστοποίησης του τετραγωνικού λάθους (Levenberg-Marquadt). Η μέθοδος υπολογισμού που ακολουθήθηκε έχει ως εξής: κρατάμε σταθερά τα εστιακά μήκη α<sub>l</sub> και α<sub>r</sub> και υπολογίζουμε τις παραμέτρους περιστροφής, κατόπιν κρατάμε σταθερές τις παραμέτρους περιστροφής  $R_l$  και  $R_r$  και υπολογίζουμε νέες τιμές για εστιακά μήκη. Η διαδικασία αυτή επαναλαμβάνεται μέχρι τη σύγκλιση.

#### 3.3.1 Υπολογισμός Θεμελιώδους Πίνακα

Ο θεμελιώδης πίνακας ορίζεται από την εξίσωση:

$$\mathbf{x}^{T}F\mathbf{x} = 0 \tag{3.10}$$

για κάθε ζευγάρι αντίστοιχων σημείων στις δύο εικόνες. Δεδομένων πολλών αντίστοιχων σημείων  $x_i \leftrightarrow x'_i$  (τουλάχιστον 7) η εξίσωση (3.10) μπορεί να χρησιμοποιηθεί για να μας δώσει τον άγνωστο πίνακα F. Πιο συγκεκριμένα γράφοντας  $\mathbf{x} = \begin{pmatrix} x & y & 1 \end{pmatrix}^T$  και  $\mathbf{x}' = \begin{pmatrix} x' & y' & 1 \end{pmatrix}^T$  σε κάθε αντιστοίχιση σημείων με γνωστές συντεταγμένες  $\mathbf{x}$  και  $\mathbf{x}'$  δημιουργείται μία γραμμική εξίσωση με αγνώστους τις παραμέτρους του F. Η εξίσωση αυτή είναι:

$$x'xf_{11} + x'yf_{12} + x'f_{13} + y'xf_{21} + y'yf_{22} + y'f_{23} + xf_{31} + yf_{32} + f_{33} = 0$$
 (3.11)

Αν θεωρήσουμε **f** το διάνυσμα διάστασης 9 που δημιουργείται από τα στοιχεία του θεμελιώδη πίνακα τότε η παραπάνω εξίσωση γίνεται:

$$(x'x \ x'y \ x' \ y'x \ y'y \ y' \ x \ y \ 1) \mathbf{f} = 0$$
 (3.12)

Από n αντίστοιχα σημεία δημιουργείται μία ομάδα γραμμικών εξισώσεων της μορφής:

Έχουμε λοιπόν μία ομογενή ομάδα εξισώσεων και το διάνυσμα **f** μπορεί να προσδιοριστεί σε μία κλίμακα. Για να υπάρχει λύση ο πίνακας A πρέπει να είναι τάξης τουλάχιστον 8. Αν η τάξη είναι ακριδώς 8 τότε το σύστημα έχει μοναδική λύση. Υπάρχει δυνατότητα όμως η τάξη του πίνακα να είναι μεγαλύτερη από 8 (στην πραγματικότητα 9 αφού έχει 9 στήλες) και να έχουμε λύση ελάχιστων τετραγώνων [68].

# 3.3.2 Υπολογισμός της Μήτρας Προβολής Κάμερας από τον Ουσιώδη Πίνακα

Η μήτρα προβολής της κάμερας παραγοντοποιείται ως  $P = K(R \mathbf{t})$  και έστω  $\mathbf{x} = P\mathbf{X}$  ένα σημείο της εικόνας. Αν ο πίνακας βαθμονόμησης K είναι γνωστός, τότε χρησιμοποιώντας τον αντίστροφό του παίρνουμε το σημείο  $\mathbf{\hat{x}} = K^{-1}\mathbf{x}$ . Τότε  $\mathbf{\hat{x}} = (R \mathbf{t})\mathbf{X}$ , όπου  $\mathbf{\hat{x}}$  το σημείο της εικόνας σε ομαλοποιημένες συντεταγμένες. Μπορεί επίσης να θεωρηθεί ως η εικόνα ενός σημείου  $\mathbf{X}$ στο χώρο όπως φαίνεται από μία κάμερα  $(R \mathbf{t})$  με πίνακα βαθμονόμησης το μοναδιαίο K = I. Τώρα ας θεωρήσουμε τις ομαλοποιημένες μήτρες κάθε κάμερας  $P = (I \mathbf{0})$  και  $P = (R \mathbf{t})$ . Ο θεμελιώδης πίνακας που αντιστοιχεί στις ομαλοποιημένες μήτρες λέγεται ουσιώδης πίνακας (essential matrix) Eκαι έχει τη μορφή:

$$E = [\mathbf{t}]_x R \tag{3.14}$$

Η εξίσωση που ορίζει τον ουσιώδη πίνακα είναι:

$$\mathbf{x}^{T}E\mathbf{x} = 0 \tag{3.15}$$

σε σχέση με τις ομαλοποιημένες συντεταγμένες των αντίστοιχων σημείων  $x_i \leftrightarrow x'_i$ . Οπότε αντικαθιστώντας στην (3.15) τα  $\widehat{\mathbf{x}}$  και  $\widehat{\mathbf{x}}'$  μετά από απλές πράξεις πινάκων προκύπτει η μαθηματική σχέση που συνδέει τον ουσιώδη με τον θεμελιώδη πίνακα:

$$E = K^{T} F K \tag{3.16}$$

Апо́ тη στιγμή поυ ο ουσιώδης пі́νакаς είναι γνωστός, μπορεί να προσδιοριστεί η μήτρα κάθε κάμερας στη βάση μίας κλίμακας σε τέσσερις πιθανούς προσανατολισμούς. Έτσι λοιπόν μπορούμε να θεωρήσουμε τη μήτρα της πρώτης κάμερας ως  $P = (I \ \mathbf{0})$  και προκειμένου να υπολογίσουμε τη μήτρα της δεύτερης κάμερας P' παραγοντοποιούμε τον E σε γινόμενο ενός αντισυμμετρικού (skew symmetric) πίνακα και ενός πίνακα περιστροφής, δηλαδή: E = SR. Αν ο E παραγοντοποιηθεί με SVD (Singular Value Decomposition), έτσι ώστε  $E = Udiag(1, 1, 0)V^T$ , τότε υπάρχουν δύο πιθανές παραγοντοποιήσεις E = SR:

$$S = UZU^T R = UWV^T \tag{3.17}$$

ή

όπου 
$$W = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$
 και  $Z = \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$  (3.18)

Αποδεικνύεται λοιπόν ότι για τη μήτρα της δεύτερης κάμερας υπάρχουν 4 δυνατές επιλογές:

$$P' = \begin{pmatrix} UWV^T & +\mathbf{u}_3 \end{pmatrix} \tag{3.19}$$

ή

$$P' = \begin{pmatrix} UWV^T & -\mathbf{u}_3 \end{pmatrix}$$
(3.20)

ή

$$P' = \begin{pmatrix} UW^T V^T & +\mathbf{u}_3 \end{pmatrix}$$
(3.21)

ή

 $P' = \begin{pmatrix} UW^T V^T & -\mathbf{u}_3 \end{pmatrix}$ (3.22)

Η γεωμετρική αναπαράσταση αυτών των λύσεων φαίνονται στο σχήμα 3.6



Σχήμα 3.6: Οι τέσσερις λύσεις για βαθμονομημένη ανακατασκευή από το πίνακα Ε [68].

Κατά την υλοποίηση της αντιστοίχισης εκτός από τους επιπολικούς περιορισμούς μπορούμε να προσθέσουμε και επιπλέον περιορισμούς όπως: τη σωστή διάταξη των γειτονικών pixels, την αμφίδρομη συνθήκη μοναδικότητας (κάθε σημείο έχει μόνο ένα αντίστοιχο στην άλλη εικόνα) και τον εντοπισμό μη κοινών περιοχών στις δύο εικόνες. Η διαδικασία αυτή οδηγεί σε εξαγωγή πληροφορίας βάθους από περισσότερες των δύο εικόνες και επομένως σε συγκέντρωση μεγαλύτερου όγκου πληροφορίας, συντελώντας στη δημιουργία πυκνών χαρτών βάθους όπως θα δειχθεί σε επόμενα κεφάλαια. Η μέθοδος πυκνής αντιστοίχισης εφαρμόζεται σε επανορθόμενες εικόνες για να εκμεταλλευτούμε τη μονοδιάστατη αναζήτηση σύμφωνα με τα παραπάνω, [43].

Ο υπολογισμός της μετατόπισης των σημείων σε ορθοποιημένες εικόνες μας οδηγεί στην εξαγωγή αντίστοιχων σημείων ανά ζεύγος εικόνων. Από τη στιγμή που διαθέτουμε αυτή την πληροφορία για κάθε ζεύγος εικόνων μπορούμε να συνδυάσουμε τις αντιστοιχίες αυτές στην κατασκευή ενός κοινού τρισδιάστατου μοντέλου [43]. Σε μία αλληλουχία εικόνων πλήθους n, ξεκινάμε από μία οπτική γωνία αναφοράς i με αντιστοιχίες σημείων ανάμεσα σε διαδοχικές εικόνες προς τις 2 κατευθύνσεις  $(i + 1 \quad i + 2 \quad .... n)$  και  $(i - 1 \quad i - 2 \quad .... 1)$  να συνδέονται σε κοινή δομή. Έχουμε δηλαδή μία εκτίμηση βάθους για κάθε αντίστοιχηση στις 2 κατευθύνσεις.

Επομένως για κάθε αντιστοίχιση σημείων  $(m_i, m_j)$  προκύπτει μία εκτίμηση βάθους  $d(x_i, x_j)$  μέσω τριγωνοποίησης, με κάποια αβεβαιότητα η οποία μειώνεται συνεχώς όσο αυξάνονται οι διαφορετικές οπτικές γωνίες.Το αποτέλεσμα μιας τέτοιας προσέγγισης είναι οι πυκνοί χάρτες βάθους πεδίου και η ευέλικτη αντιμετώπιση προβλημάτων επικαλυπτόμενων περιοχών (occluded areas) από εικόνα σε εικόνα.

## 3.4 Συμπεράσματα - Κεφαλαίου

Κλασικό πρόβλημα στο πεδίο της υπολογιστικής όρασης αποτελεί η τριδιάστατη ανακατασκευή του απεικονιζόμενου κόσμου. Προηγήθηκε η αναλυτική παρουσίαση μεθόδων στο εν λόγω πρόβλημα που βασίζονται στη χρήση είτε γραμμικών φίλτρων, είτε πυραμιδικής αναπαράστασης σε χώρο-κλίμακας, αξιοποιώντας αποκλειστικά την πληροφορία φωτεινότητας από τις εικόνες. Οι μέθοδοι αυτοί συνιστούν βασικό μέρος της κλασικής, παραδοσιακής προσέγγισης στο εν λόγω πρόβλημα. Σε επόμενα κεφάλαια θα ακολουθήσει συγκριτική μελέτη των παραπάνω μεθόδων με τεχνικές ομοιότητας φάσης, αναφορικά με την επίδοσή τους σε φυσικές εικόνες εξωτερικού χώρου, ειδικά δε σε εικόνες αρχαιολογικού ενδιαφέροντος.

# Κεφάλαιο 4

# Στερεοσκοπική Αντιστοίχιση και Τριδιάστατη Ανακατασκευή με Χρήση Μεθόδων Ομοιότητας Φάσης

# 4.1 Επεξεργασία Εικόνων με Μοντέλα Τοπικής Ενέργειας

Η εξαγωγή αξιοποιήσιμης και αξιόπιστης πληροφορίας από τις εικόνες αποτελεί το πρώτο και βασικό στάδιο στη διαδικασία του στέρεο ταιριάσματος και επομένως στην τριδιάστατη αποκατάσταση του εικονιζόμενου θέματος. Η αξιόπιστη πληροφορία συνίσταται στον εντοπισμό γεωμετρικών δομών όπως ακμές στην εικόνα με μέγιστη όσο το δυνατόν ακρίβεια στην εύρεση των ορίων τους. Πληθώρα φίλτρων έχουν εφαρμοστεί σε αυτή την κατεύθυνση με αξιόλογα αποτελέσματα. Συχνά όμως η κλασσική προσέγγιση σε εικόνες που απεικονίζουν σύνθετες δομές, όπως μία εξωτερική σκηνή κάτω από τυχαίες συνθήκες φωτισμού, οδηγεί σε εντοπισμό ψευδεπίγραφων χαρακτηριστικών. Για παράδειγμα η ύπαρξη στην εικόνα μίας σκίασης, πολλές φορές εκλαμβάνεται ως ένα μέρος ακμής. Επίσης, ο εντοπισμός σποραδικών σημείων και στοιχείων υφής αποφεύγεται γιατί δεν αποτελούν μέρος της δομής του απεικονιζόμενου θέματος. Η προσέγγιση στο συγκεκριμένο πρόβλημα που δόθηκε από την κλασσική εργασία του Canny στο [15] βασίζεται σε υπολογισμούς κλίσης φωτεινότητας και είναι εξαιρετικά ευαίσθητη σε τυχαίες αλλαγές του φωτισμού από εικόνα σε εικόνα, στα οπτικά χαρακτηριστικά της κάμερας και στα διαφορετικά επίπεδα κατωφλιοποίησης, με συνέπεια το ταίριασμα χαρακτηριστικών να εμπεριέχει μεγάλο βαθμό αβεβαιότητας.

Ήταν αναγκαία λοιπόν η στροφή μας σε μοντέλα τοπικής ενέργειας για τον εντοπισμό χαρακτηριστικών της εικόνας όπως αυτά αναπτύχθηκαν από τους Morrone και Owens [38]. Η διατήρηση της τοπικής πληροφορίας φάσης αποτέλεσε άμεση προτεραιότητα στο σχεδιασμό του φίλτρου. Προτεινόμενη από τη βιβλιογραφία επιλογή προς αυτήν την κατεύθυνση αποτελούν τα γραμμικά ως προς τη φάση φίλτρα με αντιπροσωπευτικό παράδειγμα το ζευγάρι φίλτρων σε τετραγωνισμό (quadrature filters). Η προσέγγιση της παρούσας διατριβής θα βασιστεί στη χρήση Gabor φίλτρων, σύμφωνα με την προσέγγιση των Morlet et al [34]. Η βασική διαφοροποίηση όμως βρίσκεται στη χρήση λογαριθμικών Gabor συναρτήσεων, όπως προτάθηκαν από τον Field στο [25]. Τα εν λόγω φίλτρα έχουν Γκαουσιανή συνάρτηση μεταφοράς σε λογαριθμικό χώρο κλίμακας. Τα λογαριθμικά φίλτρα Gabor συντελούν στη δημιουργία φίλτρων μεγάλου εύρους ζώνης, διατηρώντας την DC συνιστώσα ίση με μηδέν ακόμη και σε περιττής συμμετρίας φίλτρα.

Σύμφωνα με τα συγκεκριμένα μοντέλα, εντοπίζονται τα χαρακτηριστικά εκείνα στην εικόνα για τα οποία οι παράγοντες του μετασχηματισμού Fourier μεγιστοποιούνται στο χώρο της φάσης. Συγκεκριμένα, η μεγιστοποίηση του μέτρου ομοιότητας φάσης σε κάθε γωνία υποδεικνύει ένα χαρακτηριστικό στην εικόνας. Το μέτρο ομοιότητας φάσης είναι ανεξάρτητο από το ολικό πλάτος του σήματος και αμετάβλητο σε σχέση με τις συνθήκες φωτισμού ή την αντίθεση της εικόνας. Αποδεικνύεται επίσης ότι το μέτρο ομοιότητας φάσης είναι μια συνάρτηση του συνημιτόνου της απόκλισης κάθε παράγοντα φάσης από τη μέση τιμή παράγοντα φάσης: Εναλλακτικά, στην εργασία του Morrone [38] αποδεικνύεται ότι τα σημεία με μέγιστο μέτρο φάσης μπορούν να υπολογιστουν μέσω της αναζήτησης τοπικών μεγίστων στην συνάρτηση τουκόρτηση του τοπικής ενέργειας ορίζεται για τη περίπτωση μονοδιάστατης συνάρτησης φωτεινότητας  $I(\mathbf{x})$  ως εξής:

$$E(x) = \sqrt{F^2(x) + H^2(x)}$$
(4.1)

όπου F(x) το σήμα I(x) χωρίς την DC συνιστώσα, και H(x) ο Hilbert μετασχηματισμός της F(x). Προσεγγίσεις των συναρτήσεων F(x) και H(x)υπολογίζονται μετά από τη συνέλιξη του σήματος με ένα ζευγάρι φίλτρων σε τετραγωνισμό (quadrature pair). Το μέτρο ομοιότητας φάσης (phase congruency) σε αυτή την περίπτωση ακολουθεί:

$$PC_1 = \frac{|E(x)|}{\sum_n A_n(x)} \tag{4.2}$$

Βέβαια όπως έχει δειχθεί στα [54] και [56] αυτό το μέτρο ομοιότητας εντοπίζει χαρακτηριστικά με όχι σωστή τοποθέτηση στο χώρο της εικόνας και επίσης είναι αρκετά ευαίσθητο στην ύπαρξη θορύβου. Τα συγκεκριμένα προβλήματα οδήγησαν τον Kovesi [55] να αναπτύξει μία τροποποιημένη έκφραση του
μέτρου ομοιότητας αποτελούμενη από το συνημίτονο μείον το πλάτος του ημιτόνου της απόκλισης της φάσης από τη μέση τιμή καθώς και από κατάλληλη αντιστάθμιση του θορύβου που οδηγεί σε μέτρο με μεγαλύτερη ευαισθησία ακόμη και σε αδρά χαρακτηριστικά (blurred features):

$$PC_{2} = \frac{\sum_{n} W(x) \lfloor A_{n} \bigtriangleup \phi_{n}(x) - T \rfloor}{\sum_{n} A_{n}(x) + \varepsilon}$$
(4.3)

όπου  $\Delta \phi_n(x) = (\cos(\varphi_n(x) - \overline{\varphi_n}(x)) - |\sin(\varphi(x) - \overline{\varphi}(x))|$  και T η εκτιμώμενη τιμή κατωφλιοποίησης θορύβου. Μόνο οι τιμές που υπερβαίνουν το T, υπολογίζονται στο τελικό αποτέλεσμα. Τα σύμβολα  $\lfloor$  και  $\rfloor$  δηλώνουν ότι η ποσότητα μέσα σε αυτά είναι ίση με τον εαυτό της όταν η τιμή της είναι θετική και μηδέν αλλιώς. Στη πράξη η πληροφορία της τοπικής συχνότητας αποκτάται μέσω συστοιχίας Gabor φίλτρων προσανατολισμένων σε διαφορετικές χωρικές συχνότητες. Το κατάλληλο κατώφλι θορύβου T προσδιορίζεται από τα δεδομένα των αποκρίσεων του φίλτρου της εικόνας.

Ο όρος W(x) σταθμίζει την εξάπλωση της συχνότητας. Το μέτρο της ομοιότητας σε αρκετές συχνότητες έχει περισσότερη αξιοπιστία σε σχέση με το αντίστοιχο μέτρο σε χώρο με μικρότερη εξάπλωση αποκρίσεων του φίλτρου όπως για παράδειγμα σε ομαλοποιημένες εικόνες. Ένα μέτρο της εξάπλωσης των αποκρίσεων του φίλτρου μπορεί να δημιουργηθεί από το άθροισμα των πλατών των αποκρίσεων διαιρούμενο με το μέγιστο μεμονωμένο πλάτος απόκρισης. Το αποτέλεσμα κατόπιν ομαλοποιείται από τον αριθμό των κλιμάκων που έχουν χρησιμοποιηθεί. Συνεπώς καταλήγουμε σε ένα κλασματικό μέτρο στάθμισης της εξάπλωσης αποκρίσεων του φίλτρου που κυμαίνεται μεταξύ του 0 και 1:

$$s(x) = \frac{1}{N} \left( \frac{\sum_{n} A_{n}(x)}{\varepsilon + A_{max}(x)} \right)$$
(4.4)

όπου N ο συνολικός αριθμός των κλιμάκων που χρησιμοποιήθηκαν,  $A_{max}$ είναι το μέγιστο πλάτος του ζευγαριού φίλτρων σε τετραγωνισμό, στη θέση x. Με τη βοήθεια του παράγοντα  $\varepsilon$  τόσο στην εξίσωση (4.3) όσο και στην (4.4), αποφεύγεται η διαίρεση με το μηδέν.

Η συνάρτηση στάθμισης μπορεί να κατασκευαστεί εφαρμόζοντας μία σιγμοειδή συνάρτηση στις τιμές εξάπλωσης των αποκρίσεων του φίλτρου στο χώρο της συχνότητας, δηλαδή:

$$W(x) = \frac{1}{1 + e^{\gamma(c - s(x))}}$$
(4.5)

όπου cη τιμή αποκοπής της εξάπλωσης αποκρίσεων του φίλτρου, κάτω από την οποία κάθε τιμή ταύτισης συχνότητας αποκόβεται και  $\gamma$  ένας παράγοντας κέρδους ο οποίος ελέγχει την ομαλότητα της αποκοπής.

## 4.2 Εξαγωγή Χαρακτηριστικών με Χρήση Τεχνικών Ομοιότητας Φάσης

Η ανάπτυξη τοπικών ενεργειακών μοντέλων για τον εντοπισμό χαρακτηριστικών εικόνας, όπως ακμές, καμπύλες, κάτω από τυχαίες συνθήκες φωτισμού παρουσιάστηκε στο [38]. Σύμφωνα με τα συγκεκριμένα μοντέλα και μέσω διδιάστατου μετασχηματισμού Fourier της εικόνας, ανακτώνται η πληροφορία της τοπικής συχνότητας και φάσης. Όπως αναλύθηκε και στην προηγούμενη παράγραφο, ο ορισμός του μέτρου ομοιότητας φάσης, [38], δίνεται από την επέκταση της σειράς Fourier ενός μονοδιάστατου σήματος σε κάποια θέση x:

$$PC(x) = \max_{\overline{\varphi}(x) \in [0,2\pi]} \frac{\sum_{n} A_n(x) \cos(\varphi_n(x) - \overline{\varphi}(x))}{\sum_{n} A_n(x)}$$
(4.6)

Η ποσότητα φ αντιπροσωπεύει την τοπική τιμή φάσης του μετασχηματισμού Fourier σε συγκεκριμένο σημείο της εικόνας. Η ποσότητα  $\overline{\varphi}$  που μεγιστοποιεί την εξίσωση (4.6) είναι η μέση τιμή φάσης όλων των Fourier παραγόντων σε μία τοπική γειτονιά της εικόνας. Στην παρούσα διατριβή χρησιμοποιήθηκαν συναρτήσεις Gabor (συνημιτονικές και ημιτονικές συναρτήσεις διαμορφωμένες με Γκαουσιανή). Προκειμένου να μετρηθούν σε συγκεκριμένη θέση του μετασχηματισμένου κατά Fourier σήματος το πλάτος και η φάση, εφαρμόζονται δύο γραμμικά φίλτρα σε κάθετη διάταξη. Χρησιμοποιήθηκαν λοιπόν φίλτρα Gabor σε συγκεκριμένη κλίμακα και συχνότητα. Η μαθηματική έκφραση του *Gabor* φίλτρου φαίνεται παρακάτω:

$$g(t) = e^{a^2(t-t_0)^2} (\cos(2\pi f(t-t_0) + \phi) + i\sin(2\pi f(t-t_0) + \phi))$$
(4.7)

όπου:

- α είναι η σταθερά της Γκαουσιανής συνάρτησης (πυκνότητας) πιθανότητας, είναι επίσης αντιστρόφως ανάλογη του πλάτους της συνάρτησης.
- 2. t<sub>0</sub> ορίζει το κέντρο της Γκαουσιανής συνάρτησης.
- 3. f η συχνότητα της ταλάντωσης
- φ είναι η φάση της ταλάντωσης και σχετίζεται με το κέντρο της Γκαουσιανής συνάρτησης διαμόρφωσης.

Εναλλακτικά εφαρμόζονται λογαριθμικά Gabor φίλτρα όπως αυτά έχουν παρουσιαστεί στο [25]. Τα λογαριθμικά φίλτρα Gabor έχουν Γκαουσιανή συνάρτηση μεταφοράς και επιτρέπουν τη δημιουργία φίλτρων με μεγάλο εύρος ζώνης με περιττή συμμετρία και όρο DC ίσο με μηδέν. Ο μηδενικός DC όρος δε διατηρείται σε Gabor φίλτρα με εύρος ζώνης μεγαλύτερο από μία οκτάβα. Η συνάρτηση του λογαριθμικού Gabor φίλτρου έχει απόκριση συχνότητας:

$$G(f) = e^{-[log(f/f_0)]^2/2[log(\sigma/f_0)]^2}$$
(4.8)

Η απόκριση συχνότητας ενός λογαριθμικού φίλτρου Gabor είναι ουσιαστικά μία Γκαουσιανή σε λογαριθμικό άξονα συχνοτήτων. Το  $f_0$  ορίζει την κεντρική συχνότητα του ημιτονοειδούς φέροντος και αντιπροσωπεύει τον παράγοντα κλίμακας του φίλτρου. Επίσης σ είναι ο παράγοντας κλίμακας του εύρους ζώνης.

Στην παρούσα διατριβή χρησιμοποιήσαμε την προσέγγιση του Morlet [34], σύμφωνα με την οποία χρησιμοποιούνται μιγαδικές Gabor συναρτήσεις, δηλαδή ημιτονικές και συνημιτονικές καμπύλες διαμορφωμένες από μία Γκαουσιανή καμπύλη. Χρησιμοποιώντας δύο φίλτρα σε τετραγωνισμό, υπάρχει η δυνατότητα του υπολογισμού πλάτους και φάσης του σήματος για συγκεκριμένη κλίμακα-συχνότητα σε συγκεκριμένο σημείο του χώρου. Εναλλακτικά, αντί για Gabor φίλτρα χρησιμοποιούμε λογαριθμικά log-Gabor φίλτρα όπως προτάθηκαν από τον Field, [19]. Η ανάλυση του σήματος γίνεται με τη συνέλιξη του σήματος με καθένα από τα τετραγωνικά μέρη του Gabor φίλτρου, όπως φαίνεται στη σχέση (4.7).

Έστω I το σήμα και  $M_n^e$  και  $M_n^o$  η περιττής συμμετρίας (συνημιτονική) και άρτιας συμμετρίας (ημιτονική) κυματομορφή σε κλίμακα n (4.7).

$$(e_n(x), o_n(x)) = (I(x) * M_n^e, I(x) * M_n^o)$$
(4.9)

όπου  $e_n(x)$  και  $o_n(x)$  τα πραγματικά και φανταστικά μέρη της συχνότητας στους Fourier όρους. Το πλάτος και η φάση του μετασχηματισμένου σήματος στη συγκεκριμένη κλίμακα του Gabor φίλτρου δίνονται:

$$A_n(x) = \sqrt{e_n(x)^2 + o_n(x)^2}$$
(4.10)

και

$$\phi_n(x) = \arctan(e_n(x), o_n(x)) \tag{4.11}$$

για κάθε σημείο x του σήματος.

Έτσι λοιπόν, δημιουργείται ένα διάνυσμα απόκρισης για κάθε κλίμακα του φίλτρου. Ο πίνακας αυτών των διανυσμάτων αποτελεί την τοπική πληροφορία του σήματος. Ο ορισμός του μέτρου ομοιότητας φάσης μέσω της σχέσης (4.6) οδηγεί σε φτωχά αποτελέσματα τοπικής πληροφορίας για ομαλοποιημένα χαρακτηριστικά της εικόνας. Η έκφραση αυτή, καθιστά τις τιμές μέτρου ομοιότητας φάσης ευαίσθητες στο θόρυβο. Αυτά τα ζητήματα οδήγησαν τον Kovesi [55] να αναπτύξει ένα νέο μέτρο ομοιότητας φάσης, με ταυτόχρονη γενίκευσή του σε δύο διαστάσεις όπως τα σήματα εικόνας που εξετάζονται. Οι ιδέες αυτές υιοθετήθηκαν στην παρούσα διατριβή. Η νέα αυτή έκφραση του μέτρου περιλαμβάνει το άθροισμα πλατών των αποκρίσεων των φίλτρων Gabor πολλαπλασιασμένο στο πεδίο της συχνότητας με συνάρτηση διασποράς για όλες τις τιμές προσανατολισμού και κλίμακας σε συγκεκριμένη θέση στην εικόνα. Οι Γκαουσιανές συναρτήσεις μπορούν να χρησιμοποιηθούν ως συναρτήσεις διασποράς σε κατεύθυνση κάθετη προς αυτές του φίλτρου σύμφωνα με την ανάλυση στο [55]. Αυτή η προσέγγιση διατηρεί την πληροφορία της φάσης αδιατάρακτη, δεδομένου ότι οποιαδήποτε συνέλιξη σήματος με Γκαουσιανή διαμορφώνει του όρους του πλάτους και αφήνει αμετάβλητη τη φάση. Έτσι λοιπόν στην παρούσα διατριβή χρησιμοποιήθηκαν διδιάστατα Γκαουσιανά φίλτρα με αυξανόμενη συχνότητα και εύρος ζώνης. Η συνάρτηση μεταφοράς τους για τη γωνία προσανατολισμού στις 2 διαστάσεις είναι:

$$N(\theta) = e^{-\frac{(\theta - \theta_0)^2}{2\sigma_\theta^2}}$$
(4.12)

όπου  $\theta_0$  είναι η γωνία προσανατολισμού του φίλτρου, και  $\sigma_{\theta}$  η τυπική απόκλιση της Γκαουσιανής συνάρτησης για την γωνιακή κατεύθυνση. Η εξίσωση λοιπόν για την ομοιότητα φάσης σε διδιάστατα σήματα όπως οι εικόνες προκύπτει:

$$PC_{2}(\mathbf{x}) = \frac{\sum_{o} \sum_{n} W_{o}(\mathbf{x}) \lfloor A_{no}(x) \bigtriangleup \phi_{no}(\mathbf{x}) - T \rfloor}{\sum_{o} \sum_{n} A_{no}(\mathbf{x}) + \varepsilon}$$
(4.13)

για κάθε διδιάστατη θέση στην εικόνα **x**, όπου  $\Delta \phi_{no}(\mathbf{x}) = \cos{(\varphi_{no}(\mathbf{x}) - \overline{\varphi}(\mathbf{x}))} -$  $|\sin(\varphi_{no}(\mathbf{x}) - \overline{\varphi}(\mathbf{x}))|$ , και οι δείκτες *o*, *n* αναφέρονται στον προσανατολισμό και την κλίμακα του φίλτρου αντίστοιχα. Πρέπει επίσης να σημειωθεί ότι όταν η ποσότητα ανάμεσα στα σύμβολα | . | δεν είναι θετική, τότε το αποτέλεσμα γίνεται μηδενικό. Ο αριθμητής του παραπάνω κλάσματος αντιπροσωπεύει τη συνολική ενέργεια του διδιάστατου σήματος στην τοπική περιοχή της εικόνας. Ο όρος  $W_o(\mathbf{x})$  σταθμίζει τη διασπορά της συχνότητας. Η ποσότητα  $W_o(\mathbf{x})$  επίσης αντισταθμίζει την έλλειψη αξιοπιστίας των μέτρων ομοιότητας φάσης σε περιοχές της εικόνας με μικρή διασπορά συχνότητας, (π.χ. ομαλοποιημένες εικόνες). Ο ρόλος του ε είναι η αποφυγή διαίρεσης με το μηδέν. Τέλος, μόνο οι τιμές οι οποίες είναι μεγαλύτερες από το κατώφλι Τ (προσδοκώμενη επιρροή θορύβου) χρησιμοποιούνται στον υπολογισμό του τελικού αποτελέσματος. Το κατάλληλο κατώφλι Τ για το θόρυβο προσδιορίζεται πειραματικά, σύμφωνα με την απόκριση του φίλτρου μικρότερης κλίμακας για κάθε εικόνα. Η ταξινόμηση του είδους των χαρακτηριστικών στην εικόνα (γωνία, ευθεία), τα οποία αντιστοιχούν στις μέγιστες τιμές ομοιότητας φάσης, βασίζεται στον υπολογισμό κατάλληλων στιγμών.

#### 4.2.1 Ανάκτηση Ομοιότητας Φάσης στις διάφορες κατευθύνσεις

Προκειμένου να συμπεριλάβουμε πληροφορία σχετικά με τον τρόπο που η ομοιότητα φάσης ποικίλει με την κατεύθυνση του φίλτρου αρχικά υπολογίζουμε το μέτρο της ομοιότητας φάσης ανεξάρτητα σε κάθε κανάλι χρησιμοποιώντας την εξίσωση (4.3). Κατόπιν υπολογίζουμε τις στιγμές της ομοιότητας και εξετάζουμε τη μεταβολή των στιγμών ανάλογα με την κατεύθυνση. Ο κύριος άξονας ο οποίος αντιστοιχεί στον άξονα που ελαχιστοποιείται η στιγμή υποδεικνύει την κατεύθυνση του χαρακτηριστικού της εικόνας. Επίσης το πλάτος της μέγιστης στιγμής σε άξονα κάθετο στον κύριο αντιστοιχεί στην αξιοπιστία του εντοπισμένου χαρακτηριστικού της εικόνας.

Ακολουθώντας λοιπόν την ανάλυση στιγμών για κάθε σημείο της εικόνας έχουμε:

$$\alpha = \sum_{\theta} (PC(\theta)cos(\theta))^2 \tag{4.14}$$

$$b = 2\sum_{\theta} (PC(\theta)cos(\theta))(PC(\theta)sin(\theta))$$
(4.15)

$$c = \sum_{\theta} (PC(\theta)sin(\theta))^2$$
(4.16)

όπου  $PC(\theta)$  είναι η ταύτιση φάσης στην κατεύθυνση  $\theta$ , ενώ το άθροισμα γίνεται στο διακριτό χώρο για το σύνολο των κατευθύνσεων. Η γωνία του κύριου άξονα  $\phi$  δίνεται σύμφωνα με την εργασία του Kovesi [55]:

$$\phi = \frac{1}{2}atan2(\frac{b}{\sqrt{b^2 + (\alpha - c)^2}}, \frac{\alpha - c}{\sqrt{b^2 + (\alpha - c)^2}})$$
(4.17)

Οι μέγιστες και ελάχιστες στιγμές Μ και m αντίστοιχα δίνονται:

$$M = \frac{1}{2}(c + \alpha + \sqrt{b^2 + (\alpha - c)^2})$$
(4.18)

$$m = \frac{1}{2}(c + \alpha - \sqrt{b^2 + (\alpha - c)^2})$$
(4.19)

#### 4.2.2 Αντιστοίχιση εικόνων από Τοπική Πληροφορία Συχνότητας

Στη συνέχεια εξετάζουμε μία σύγχρονη προσέγγιση στη σύγκριση και στο ταίριασμα σημάτων εικόνας, για μετρήσεις ανομοιότητας βάθους (disparity), χρησιμοποιώντας εντοπισμένα δεδομένα συχνότητας.

Αναλύοντας το σήμα της εικόνας στο χώρο της συχνότητας επιτυγχάνουμε τη δημιουργία ενός αδιάστατου μέτρου ομοιότητας με ακρίβεια στη διαδικασία εντοπισμού χαρακτηριστικών. Έχει αποδειχθεί ότι η πληροφορία της φάσης μπορεί να χρησιμοποιηθεί με αποδοτικό τρόπο ως οδηγός για την αναζήτηση ενός σημείου στο ένα σήμα που ταιριάζει με βέλτιστο τρόπο σε ένα σημείο του άλλου σήματος. Έχει δειχθεί επίσης ότι η αναπαράσταση ενός σήματος στο χώρο της φάσης καθώς και η πληροφορία του πλάτους σε λογαριθμική κλίμακα συχνότητας μας δίνουν ένα αποτελεσματικό χώρο για τη σύγκριση σημάτων.

Έχοντας αποκτήσει τοπική αναπαράσταση του σήματος στο χώρο της συχνότητας, (4.10), (4.11), χρησιμοποιούμε τα δεδομένα αυτά για τη διαδικασία του ταιριάσματος. Με δοσμένες τις τιμές των σημάτων σε δύο θέσεις  $I_1(x_1)$ και  $I_2(x_2)$ , θέλουμε να εξετάσουμε κατά πόσον οι πίνακες διανυσμάτων των αποκρίσεων σε αυτές τις θέσεις συσχετίζονται. Το μέτρο συσχέτισης είναι το εσωτερικό γινόμενο μεταξύ αντίστοιχων διανυσμάτων απόκρισης σε κάθε κλίμακα. Το μέτρο αυτό μας δίνει το συνημίτονο της απόκλισης φάσης με βάρη ανάλογα του πλάτους των διανυσμάτων.

Έτσι λοιπόν σχηματίζουμε ένα σταθμισμένο άθροισμα από αντίστοιχα ζεύγη εσωτερικών γινομένων των διανυσμάτων απόκρισης σε όλες τις κλίμακες ομαλοποιημένο από το σταθμισμένο άθροισμα όλων των εσωτερικών γινομένων των πλατών των διανυσμάτων απόκρισης. Προκύπτει έτσι ένα αδιάστατο μέτρο της συσχέτισης σήματος:

$$C(x_1, x_2) = \frac{\left|\sum_n (e_n(x_1)e_n(x_2) + o_n(x_1)o_n(x_2) - T)\right|}{\sum_n A_n(x_1)A_n(x_2) + \varepsilon}$$
(4.20)

όπου  $e_n(x_1)$  και  $o_n(x_1)$ : οι έξοδοι των περιττών και άρτιων συμμετρικών φίλτρων σε κλίμακα n στη θέση  $x_1$  του σήματος  $I_1$ , (όμοια στη θέση  $x_2$ ),  $A_n(x_1)$ το πλάτος της απόκρισης του ζεύγους φίλτρων στο σήμα  $I_1$  στη θέση  $x_1$ , και  $\varepsilon$ μία μικρή σταθερά που αποτρέπει τη διαίρεση με το μηδέν όταν το σήμα είναι επίπεδο, [53]. Ο παράγοντας Τ είναι ένας παράγοντας αντιστάθμισης θορύβου που αντιπροσωπεύει την απόκριση μέγιστης συσχέτισης προερχόμενη αποκλειστικά από το σήμα θορύβου. Αυτός ο παράγοντας υπολογίζεται συνδυάζοντάς την επιρροή του θορύβου από καθένα από τα φίλτρα. Αν υποθέσουμε το φάσμα του θορύβου επίπεδο, η μέγιστη επιρροή του θορύβου για κάθε έξοδο του φίλτρου μπορεί να υπολογιστεί από τη μέση απόκριση πλάτους του φίλτρου μικρότερης κλίμακας. Το φίλτρο μικρότερης κλίμακας έχει περιορισμένη περιοχή που αποκρίνεται σε χαρακτηριστικά της εικόνας άρα το μεγαλύτερο μέρος της απόκρισης οφείλεται στο θόρυβο. Η απόκριση θορύβου των φίλτρων μεγαλύτερης κλίμακας υπολογίζεται σε σχέση με το εύρος ζώνης τους συγκριτικά με το εύρος ζώνης του φίλτρου με χαμηλότερη κλίμακα.

Επιπρόσθετα, για κάθε κλίμακα n, η υπολογιζόμενη μετατόπιση για να ταιριάξουν τα σήματα είναι  $\frac{\lambda_n \delta f_n}{2\pi}$ , όπου  $\delta f_n$  είναι η διαφορά φάσης των διανυσμάτων απόκρισης ενός φίλτρου στην κλίμακα n και όπου  $\lambda_n$  το μήκος κύματος της κυματομορφής που ορίζει το φίλτρο. Αυτή η μετατόπιση εμπεριέχει λάθη προερχόμενα από τις διαφορετικές στάθμες θορύβου στις διάφορες κλίμακες. Έτσι λοιπόν σταθμίζουμε τις μετατοπίσεις με τις αποκρίσεις πλάτους και έχουμε μία εκτίμηση του μέτρου ανομοιότητας βάθους (disparity):

$$\Delta(x_1, x_2) = \frac{\sum_n A_n(x_1) A_n(x_2) \lambda_n \delta \phi_n}{2\pi \sum_n A_n(x_1) A_n(x_2)}$$
(4.21)

#### 4.3 Αναλυτικό Σήμα

Το αναλυτικό σήμα αποτελεί μιγαδική αναπαράσταση της πληροφορίας ενός σήματος. Βρίσκει σημαντικές εφαρμογές στην κωδικοποίηση του σήματος μέσω της διαμόρφωσης συχνότητας και φάσης. Σημαντική ιδιότητα του αναλυτικού σήματος αποτελεί επίσης και ο διαχωρισμός της ταυτότητας. Η ιδιότητα αυτή συνίσταται στην εξαγωγή της τοπικής πληροφορίας πλάτους και φάσης ενός σήματος από την πολική του αναπαράσταση. Η τοπική φάση παραμένει αμετάβλητη σε σχέση με την τοπική ενέργεια και μεταβάλλεται καθώς μεταβάλλεται η τοπική δομή. Το τοπικό πλάτος είναι αμετάβλητο σε σχέση με την τοπική δομή και αντιπροσωπεύει την τοπική ενέργεια. Ενέργεια και δομή αποτελούν πληροφορία ενσωματωμένη στο σήμα. Γεγονός είναι ότι τα φίλτρα σε τετραγωνισμό αποτελούν τελεστές που εξάγουν ζωνοπερατή έκφραση της πληροφορίας πλάτους και φάσης. Από τη στιγμή λοιπόν που οποιοδήποτε σήμα μπορεί να αναλυθεί σε στοιχειώδη σήματα χρησιμοποιώντας ζωνοπερατά φίλτρα μικρού εύρους ζώνης, η πολική αναπαράσταση του αναλυτικού σήματος σε μία μικρή περιοχή συχνοτήτων μπορεί να θεωρηθεί ως ένας τετραγωνικός διαχωρισμός της πληροφορίας. Είναι χρήσιμο λοιπόν κάθε πιθανή επέκταση του μονοδιάστατου αναλυτικού σήματος σε δύο διαστάσεις να διατηρεί την δυνατότητα του τετραγωνικού διαχωρισμού της πληροφορίας. Στη διαδικασία της επέκτασης ένα από τα βασικά προβλήματα αποτελεί το γεγονός ότι μονοδιάστατα μέτρα όπως η τοπική φάση δεν μπορεί να κωδικοποιήσει τη διδιάστατη δομή σε μία εικόνα λόγω ελλιπών βαθμών ελευθερίας. Συχνά η επέκταση σε δύο διαστάσεις του αναλυτικού σήματος πραγματοποιείται μέσω του μετασχηματισμού Hilbert ως προς ένα από τους δύο συνήθως άξονες του συστήματος αναφοράς στο πλαίσιο της εικόνας. Η μεθοδολογία αυτή οδηγεί σε τιμές τοπικής φάσης και τοπικού πλάτους με συστηματικό σφάλμα το οποίο εξαρτάται από τη γωνία μεταξύ των αξόνων του πλαισίου αναφοράς της εικόνας και του προσανατολισμού του σήματος. Σύμφωνα με την προσέγγιση

στο [46] διατηρείται το μονοδιάστατο μέτρο φάσης αλλά προστίθεται και η πληροφορία του τοπικού προσανατολισμού μαζί με το τοπικό μέτρο φάσης και πλάτους. Η νέα λοιπόν γενίκευση του αναλυτικού σήματος ονομάζεται μονογενές σήμα (monogenic signal). Η ενσωματωμένη πληροφορία του μονογενούς σήματος θα χρησιμοποιηθεί ως εναλλακτικό στοιχείο σύγκρισης σε σχέση με την πληροφορία φωτεινότητας στα πλαίσια του προβλήματος αντιστοίχισης όμοιων σημείων ή περιοχών σε ζεύγος στέρεο εικόνων.

### 4.4 Μετασχηματισμός Riesz, Μονογενές Σήμα και Στερεοσκοπική Αντιστοίχιση

Ο nD μετασχηματισμός Fourier (n = 1, 2) του  $f(\mathbf{x})$  δίνεται απο τη σχέση

$$F(\mathbf{u}) = \int_{\Re^n} f\left(\mathbf{x}\right) e^{-i2\pi(\mathbf{x},\mathbf{u})} d\mathbf{x}$$

Αν η ανεξάρτητη μεταβλητή **x** αντιπροσωπεύει χρόνο τότε η μεταβλητή **u** αντιπροσωπεύει συχνότητα. Ο μετασχηματισμός Hilbert ορίζεται από την παρακάτω συνάρτηση μεταφοράς

$$H_1\left(u\right) = -i\mathrm{sgn}\left(u\right)$$

όπου sgn είναι η συνάρτηση προσήμου και ορίζεται για κάθε x πραγματικό αριθμό:

$$\operatorname{sgn}(x) = \begin{cases} -1 & \alpha\nu & x < 0\\ 0 & \alpha\nu & x = 0\\ 1 & \alpha\nu & x > 0 \end{cases}$$

Το διδιάστατο αναλυτικό σήμα βασίζεται στη διδιάστατη γενίκευση του Hilbert μετασχηματισμού γνωστού και ως Riesz μετασχηματισμού. Θα χρησιμοποιήσουμε ένα 3Δ δυναμικό *p* για την πραγματοποίηση του Riesz μετασχηματισμού [39], [46]. Δημιουργούμε το εξής αρμονικό δυναμικό πεδίο:

$$\mathbf{g}(\mathbf{x}) = egin{pmatrix} g_1(\mathbf{x}) & g_2(\mathbf{x}) & g_3(\mathbf{x}) \end{pmatrix}^T = 
abla p(\mathbf{x})$$

Το δυναμικό πεδίο για  $\mathbf{x} = (x_1, x_2, x_3)$  με  $x_3 < 0$  είναι αστρόβιλο και σωληνοειδές δηλαδή:

$$\operatorname{rot}(\mathbf{g}(\mathbf{x})) = \nabla \times \mathbf{g}(\mathbf{x}) = 0 \tag{4.22}$$

και

$$\operatorname{div}(\mathbf{g}(\mathbf{x})) = \langle \nabla, \mathbf{g}(\mathbf{x}) \rangle = 0 \tag{4.23}$$

Η σχέση (4.22) υποδεικνύει ότι το g είναι πεδίο αστρόβιλο ενώ η σχέση (4.23) υποδεικνύει ότι είναι αρμονικό πεδίο. Εφαρμόζοντας διδιάστατο μετασχηματισμό Fourier στις παραπάνω εξισώσεις έχουμε:

$$P(u_1, u_2, x_3) = C(u_1, u_2)exp(2\pi q x_3)$$

όπου  $q = \sqrt{u_1^2 + u_2^2}$  και  $C(u_1, u_2)$  ανεξάρτητο από  $x_3$ . Επομένως έχουμε για  $g = \nabla p$ :

$$G_1(u_1, u_2, x_3) = i2\pi u_1 P(u_1, u_2, x_3)$$
(4.24)

$$G_2(u_1, u_2, x_3) = i2\pi u_2 P(u_1, u_2, x_3)$$
(4.25)

$$G_3(u_1, u_2, x_3) = i2\pi q P(u_1, u_2, x_3)$$
(4.26)

Επιπρόσθετα προσδιορίζουμε και τις σχέσεις μεταξύ  $G_1$ ,  $G_2$  και  $G_3$ :

$$G_1(u_1, u_2, x_3) = \frac{iu_1}{q} G_3(u_1, u_2, x_3)$$
(4.27)

$$G_2(u_1, u_2, x_3) = \frac{iu_2}{q} G_3(u_1, u_2, x_3)$$
(4.28)

Για  $x_3 \mapsto 0^-$  οι παραπάνω σχέσεις γίνονται:

$$\mathbf{G}_1(u_1, u_2, 0) = \frac{iu_1}{q} G_3(u_1, u_2, 0) = \frac{iu_1}{q} F(u_1, u_2)$$

και

$$G_2(u_1, u_2, 0) = \frac{iu_2}{q} F(u_1, u_2)$$

Θέτοντας  $\mathbf{F}_R = (G_1(u_1, u_2, 0), G_2(u_1, u_2, 0))^T$  και ορίζοντας  $\mathbf{u} = (u_1, u_2)^T$  έτσι ώστε  $q = |\mathbf{u}|$  έχουμε την μαθηματική έκφραση του μετασχηματισμού Riesz στο χώρο της συχνότητας:

$$\mathbf{F}_{R}(\mathbf{u}) = \frac{i\mathbf{u}}{|\mathbf{u}|} F(\mathbf{u}) \stackrel{def}{=} \mathbf{H}_{2}(\mathbf{u}) F(\mathbf{u})$$
(4.29)

Η συνάρτηση μεταφοράς του Riesz μετασχηματισμού είναι η  $\mathbf{H}_2$  που επίσης αντιπροσωπεύει τη 2Δ γενίκευση του Hilbert μετασχηματισμού. Σύμφωνα με το [46] ο ορισμός του μονογενούς (monogenic) σήματος δίνεται με τη χρήση της ακόλουθης συνάρτησης μεταφοράς:

$$F_M(\mathbf{u}) = G_3(u_1, u_2, 0) - iG_1(u_1, u_2, 0) - jG_2(u_1, u_2, 0) = F(\mathbf{u}) - (i, j)\mathbf{F}_R(\mathbf{u}) = \frac{|\mathbf{u}| + (1, k)\mathbf{u}}{|\mathbf{u}|}F(\mathbf{u}) \quad (4.30)$$

η οποία είναι ισοδύναμη με

$$f_M(\mathbf{x}) = f(\mathbf{x}) - (i, j)\mathbf{f}_R(\mathbf{x})$$
(4.31)

Η φάση  $\varphi$ του μονογενούς σήματος έχει διανυσματική μορφή και δίνεται από:

$$\varphi(\mathbf{x}) = \frac{\mathbf{f}_R(\mathbf{x})}{|\mathbf{f}_R(\mathbf{x})|} \arctan(\frac{|\mathbf{f}_R(\mathbf{x})|}{f(\mathbf{x})})$$
(4.32)

Στην παρούσα διατριβή, η απαιτητική διαδικασία του ταιριάσματος χαρακτηριστικών βασίστηκε στις ιδέες μονογενών φίλτρων όπως αυτές περιγράφηκαν παραπάνω. Έγινε σαφές ότι το διδιάστατο αναλυτικό σήμα βασίζεται στη διδιάστατη γενίκευση του μετασχηματισμού Hilbert, γνωστού και ως μετασχηματισμού Riesz. Η έκφραση του κατά Riesz του μετασχηματισμένου κατά Fourier σήματος  $F(\mathbf{u})$  στο χώρο της συχνότητας γίνεται:

$$\mathbf{F}_{\mathrm{R}}\left(\mathbf{u}\right)=i\frac{\mathbf{u}}{\left|\mathbf{u}\right|}F\left(\mathbf{u}\right)$$

όπου **u** το διδιάστατο διάνυσμα συχνότητας  $(u_1, u_2)$  και  $|\mathbf{u}| = \sqrt{u_1^2 + u_2^2}$ .

Αρχικά υπολογίζεται ο μετασχηματισμός Fourier της κάθε εικόνας  $I_F$ . Το επόμενο στάδιο συνίσταται στην εισαγωγή του λογαριθμικού φίλτρου Gabor σε πολικές συντεταγμένες. Η συνάρτηση μεταφοράς του διδιάστατου log-Gabor φίλτρου στο χώρο της συχνότητας (σε πολικές συντεταγμένες) προκύπτει από το ακόλουθο γινόμενο:

$$G(\mathbf{u}) = G_{\rho} \cdot G_{\omega} \tag{4.33}$$

όπου  $G_{\rho}$  το ακτινικό μέρος του φίλτρου ( $\rho = |\mathbf{u}|$  είναι το πλάτος του διδιάστατου διανύσματος συχνότητας) και  $G_{\omega}$  το γωνιακό μέρος του φίλτρου ( $\omega = \arctan(-u_2, u_1)$  αντιπροσωπεύει πολική συντεταγμένη γωνίας του διανύσματος συχνότητας).

Το ακτινικό μέρος  $G_{\rho}$  προσδιορίζει το εύρος συχνοτήτων στις οποίες αναταποκρίνεται το log-Gabor φίλτρο. Η απόκριση συχνότητας είναι μία Γκαουσιανή συνάρτηση σε λογαριθμική κλίμακα

$$G_{\rho} = e^{-\frac{\left(\log(|\mathbf{u}|/u_0)\right)^2}{2\left(\log(\sigma_u/u_0)\right)^2}}$$
(4.34)

όπου  $u_0$  αποτελεί την κεντρική συχνότητα του φίλτρου, η μεταβλητή  $|\mathbf{u}|$  αντιπροσωπεύει την απόσταση από το κέντρο του πεδίου συχνότητας, και  $\sigma_u$  αποτελεί ένα παράγοντα κλίμακας του εύρους ζώνης. Ο λόγος  $\sigma_u/u_0$  επηρεάζει άμεσα το εύρος ζώνης του φίλτρου. Συνεπώς προκειμένου να διατηρηθούν φίλτρα σταθερού σχήματος, ο λόγος αυτός πρέπει να παραμένει σταθερός.

Αναφορικά με το γωνιακό παράγοντα  $G_{\omega}$  του φίλτρου, ο οποίος ελέγχει την επιλογή προσανατολισμού του φίλτρου, εκφράζεται επίσης με μία Γκαουσιανή συνάρτηση συναρτήσει της πολικής γωνίας  $\omega$ , με κέντρο στη γωνία  $\omega_0$  και τυπική απόκλιση  $\sigma_{\omega}$ . Ακολουθεί η μαθηματική περιγραφή:

$$G_{\omega} = e^{-\frac{(\omega - \omega_0)^2}{2\sigma_{\omega}^2}}$$
(4.35)

όπου  $\omega = \arctan(-u_2, u_1)$  απεικονίζει την γωνιακή πολική συντεταγμένη του διανύσματος συχνότητας  $\mathbf{u} = (u_1, u_2)$ .

Η επακόλουθη πυραμιδική κατασκευή log-Gabor φίλτρων, βασίστηκε στην ακτινική αναπαράσταση της log-Gabor συνάρτησης, σύμφωνα με τη σχέση (4.34). Η χρήση πληθώρας κεντρικών συχνοτήτων  $u_0$  ανάλογα των επιθυμητών τιμών μήκους κύματος του φίλτρου, οδηγεί σε διαδοχικά φίλτρα, σχετιζόμενα με ένα πάραγοντα κλίμακας. Η τεχνική αυτή συνεισφέρει στην κατασκευή ζωνοπερατών εκφράσεων του σήματος  $\mathbf{F}_R$  στο χώρο της συχνότητας.

$$\mathbf{H}_{R} = \mathbf{F}_{R}(\mathbf{u})G\left(\mathbf{u}\right) \tag{4.36}$$

όπου  $\mathbf{H}_{R} = \begin{pmatrix} H_{R}^{1} & H_{R}^{2} \end{pmatrix}^{T}$ . Τα παραπάνω φίλτρα υλοποιούνται στο χώρο της συχνότητας, και κατόπιν της εφαρμογής αντίστροφου μετασχηματισμού Fourier,  $\mathcal{F}^{-1}$ , το πραγματικό μέρος των σημάτων που προκύπτουν έχει ως εξής:

$$I_F(\mathbf{x}) = Re\left[\mathcal{F}^{-1}\left\{F(\mathbf{u})\,G\left(\mathbf{u}\right)\right\}\right] \tag{4.37}$$

$$I_R^1(\mathbf{x}) = Re\left[\mathcal{F}^{-1}\left\{H_R^1(\mathbf{u})\right\}\right]$$
(4.38)

$$I_R^2(\mathbf{x}) = Re\left[\mathcal{F}^{-1}\left\{H_R^2(\mathbf{u})\right\}\right]$$
(4.39)

Τα παραπάνω οδηγούν σε μία έκφραση του διδιάστατου αναλυτικού σήματος, το οποίο έχει ως πραγματικό μέρος το σήμα  $I_F$  (4.37) και φανταστικό μέρος τη μαθηματική έκφραση του μετασχηματισμού Riesz, σύμφωνα με το [46]. Επομένως, το φανταστικό μέρος του διδιάστατου αναλυτικού σήματος αποτελείται από δύο σήματα, το σήμα  $I_R^1(\mathbf{x})$  (4.38) και το σήμα  $I_R^2(\mathbf{x})$  (4.39). Άρα σε κάθε σημείο της εικόνας (x, y), για συγκεκριμένη κλίμακα και ορισμένο προσανατολισμό φίλτρου, ορίζεται ένα τριδιάστατο διάνυσμα  $\mathbf{x}(x, y)$ αποτελούμενο από τα τρία παραπάνω σήματα (4.37), (4.38) και (4.39).

$$\mathbf{I}_M(\mathbf{x}) = \left[ I_F(\mathbf{x}), I_R^1(\mathbf{x}), I_R^2(\mathbf{x}) \right]$$
(4.40)

Επιπλέον, το μέτρο του πλάτους της ενέργειας του σήματος δίνεται:

$$A(\mathbf{x}) = \sqrt{(I_F(\mathbf{x}))^2 + (I_R^1(\mathbf{x}))^2 + (I_R^2(\mathbf{x}))^2}$$
(4.41)

Η τοπική πληροφορία δομής μίας εικόνας είναι δυνατόν να εξαχθεί από το μονογενές σήμα, υπολογίζοντας τις γωνίες προσανατολισμού και φάσης. Η γωνία προσανατολισμού δίνεται από την ακόλουθη σχέση:

$$\theta(\mathbf{x}) = \arctan\left(I_R^2(\mathbf{x}), I_R^1(\mathbf{x})\right) \tag{4.42}$$

Η γωνία φάσης εξαρτάται από τη γωνίας που σχηματίζει το μονογενές διάνυσμα με το επίπεδο  $H_R^1 - H_R^2$ , και κυμαίνεται μεταξύ  $-\frac{\pi}{2}, \frac{\pi}{2}$ :

$$\phi(\mathbf{x}) = \arctan\left(I_F(\mathbf{x}), \sqrt{\left(I_R^1(\mathbf{x})\right)^2 + \left(I_R^2(\mathbf{x})\right)^2}\right)$$
(4.43)

Δοσμένων δύο θέσεων  $\mathbf{x}_1 = (x_1, y_1)$  και  $\mathbf{x}_2 = (x_2, y_2)$  στην πρώτη και δεύτερη εικόνα ( $I_1$  και  $I_2$ ) αντίστοιχα της στέρεο διάταξης, το μέτρο συσχέτισης που χρησιμοποιείται στην παρούσα διατριβή είναι:

$$C_{12}(\mathbf{x}_{1}, \mathbf{x}_{2}) = \frac{\sum_{m=-k}^{+k} \sum_{n=-l}^{+l} \mathbf{I}_{M1}(x_{1}+m, y_{1}+n)}{\sum_{m=-k}^{+k} \sum_{n=-l}^{+l} A_{1}(x_{1}+m, y_{1}+n)} \frac{\mathbf{I}_{M2}(x_{2}+m, y_{2}+n)}{A_{2}(x_{2}+m, y_{2}+n)} \quad (4.44)$$

όπου  $\mathbf{I}_{M1}(\mathbf{x}_1)$  και  $\mathbf{I}_{M2}(\mathbf{x}_2)$  είναι τα τριδιάστατα διανύσματα απόκρισης των μονογενών φίλτρων στα υποψήφια για ταίριασμα σημεία  $(x_1, y_1)$  και  $(x_2, y_2)$ των εικόνων  $\mathbf{I}_1$  και  $\mathbf{I}_2$  αντίστοιχα. Το μέτρο συσχέτισης στη (4.44) υπολογίστηκε ως το εσωτερικό γινόμενο των παραπάνω διανυσμάτων με παράθυρο από -l σε l και από -k σε k στο διδιάστατο επίπεδο εικόνας. Επιπρόσθετα, το μέτρο αυτό ομαλοποιήθηκε από το άθροισμα των αποκρίσεων πλάτους, σύμφωνα με την (4.41). Το ταίριασμα θεωρείται επιτυχημένο για εκείνα τα ζευγάρια σημείων στα οποία το παραπάνω μέτρο συσχέτισης μεγιστοποιείται (π.χ.  $argmax|C_{12}(\mathbf{x}_1, \mathbf{x}_2)|$ ).

# 4.5 Πειραματική Αξιολόγηση - Εντοπισμός Χαρακτηριστικών Δομών - Ταίριασμα Εικόνων - Αυτοβαθμονόμηση

Η κάμερα που χρησιμοποιήθηκε στις μετρήσεις ήταν μία Nikon D70s 18-74mm. Ως πειραματικά δεδομένα χρησιμοποιήθηκαν οι διαδοχικές φωτο-



Σχήμα 4.1: Διαδοχικές φωτογραφίες σε κλίμακα του γκρι του Ερεχθείου στην Ακρόπολη και μίας από τις πλαϊνές πύλες του Ηρώδειου. Οι φωτογραφίες παράχθηκαν μέσω της σχεδόν ευθείας κίνησης της κάμερας

γραφίες δύο μνημείων:

- 1. Πλαϊνή όψη του Ερεχθείου
- 2. Πλαϊνή πύλη του Ηρώδειου

Κατά τη διάρκεια της λήψης η κάμερα κινήθηκε σε σχεδόν ευθεία γραμμή. Η μικρή μετατόπιση της κάμερας συνετέλεσε στην αποφυγή ασυνεχειών μεγάλων περιοχών από φωτογραφία σε φωτογραφία. Η συνολική μετατόπιση της κάμερας ήταν περίπου 120mm. Όλες οι φωτογραφίες έχουν μετασχηματιστεί χρωματικά σε κλίμακα του γκρι δεδομένου ότι η χρωματική πληροφορία δεν συντελεί σε αυτή τη φάση στον υπολογισμό του βάθους. Ενδεικτικά οι φωτογραφίες που χρησιμοποιήθηκαν στην στέρεο διάταξη των πειραμάτων μας φαίνονται στο Σχήμα 4.1.

Η αρχική φάση του στέρεο ταιριάσματος περιλαμβάνει τον εντοπισμό χαρακτηριστικών για κάθε εικόνα. Ακολουθεί η συγκριτική μελέτη των τριών προσεγγίσεων που αναπτύχθηκαν προγενέστερα για το πρόβλημα αυτό. Η πρώτη προσέγγιση συνίσταται στην εφαρμογή κλασσικών φίλτρων βασισμένων στον υπολογισμό κλίσης της φωτεινότητας στην εικόνα. Κύριο παράδειγμα φίλτρου αυτής της κατηγορίας αποτελεί το φίλτρο εντοπισμού ακμών που προτάθηκε από τον Canny [15]. Στη δεύτερη προσέγγιση αξιοποιήθηκε η αναπαράσταση της εικόνας σε χώρο-κλίμακας προκειμένου να εξαχθούν χαρακτηριστικά σημεία (keypoints) της εικόνας. Τα εν λόγω σημεία κλειδιά εμφανίζουν επαναληψιμότητα στις υπό εξέταση εικόνες για διάφορες τιμές κλίμακας, περιστροφής ή συνθηκών φωτισμού. Τα υποψήφια σημεία κλειδιά εντοπίζονται ακολουθώντας τη μεθοδολογία και τον αλγόριθμο όπως αυτά προτάθηκαν από το Lowe, [21]. Στην τρίτη προσέγγιση ο εντοπισμός ακμών στις ορθοποιημένες εικόνες, βασίστηκε στο μέτρο ομοιότητας φάσης (phase congruency), όπως προτείνεται στην παρούσα διατριβή. Στην εν λόγω προσέγγιση, οι ακμές εντοπίζονται στην εικόνα μέσω του υπολογισμού της μέγιστης τιμής των στιγμών της συνδιακύμανσης αναλογίας φάσης. Από την ομάδα των ακμών επιλέγονται οι ακμές εκείνες που το μήκος τους υπερβαίνει ένα προκαθορισμένο κατώφλι. Τα αποτελέσματα εντοπισμού ακμών φαίνονται στο Σχήμα 4.2 για όλες τις παραπάνω προσεγγίσεις.

Κατά τον εντοπισμό ακμών στην εικόνα με υπολογισμούς φάσης στο χώρο της συχνότητας χρησιμοποιήθηκαν log-Gabor συναρτήσεις με Γκαουσιανή συνάρτηση μεταφοράς σε λογαριθμική κλίμακα συχνότητας. Το φίλτρο που χρησιμοποιήθηκε, εφαρμόστηκε σε έξι διαφορετικούς προσανατολισμούς και τέσσερις κλίμακες, με σταθερό εύρος ζώνης μία οκτάβα, ακολουθώντας την ανάλυση που παρουσιάστηκε προγενέστερα. Παρατηρώντας την πρώτη γραμμή με εικόνες του Σχήματος 4.2 συμπεραίνεται ότι η χρήση κλασσικής προσέγγισης στο πρόβλημα εντοπισμού ακμών, όπως τα φίλτρα Canny, για τέτοιες πολύπλοκες στο θέμα εικόνες, οδηγούν σε εικόνες ακμών με αδύναμο και ανακριβή εντοπισμό τους στην αρχική εικόνα, συγκρινόμενες με αυτές της τρίτης σειράς. Στην τρίτη σειρά του Σχήματος 4.2 ο εντοπισμός ακμών έχει βασιστεί στη χρήση μεθόδου ομοιότητας φάσης. Η δεύτερη σειρά του ίδιου σχήματος αναπαριστά τα υποψήφια για ταίριασμα σημεία κλειδιά, όπως αυτά έχουν ανακτηθεί μέσω του μετασχηματισμού SIFT σε ένα συγκεκριμένο επίπεδο της πυραμιδικής κλίμακας-χώρου αναπαράστασης της εικόνας. Στην περίπτωση αυτή είναι προφανές ότι η χρήση τελεστών Γκαουσιανών διαφορών τονίζει χαρακτηριστικά ακμών ακόμη και εκείνα χαμηλής αντίθεσης. Αυτό το είδος χαρακτηριστικών χαμηλής αντίθεσης θα εξαιρεθούν από τα μετασχηματισμένα με SIFT χαρακτηριστικά ως μή - διαχωρίσιμα. Παρόλο που η παραπάνω μεθοδολογία οδηγεί σε μεγάλο αριθμό υποψήφιων σημείων, η διαδικασία του στέρεο ταιριάσματος βασίζεται σε πολύ λιγότερα στοιχεία (υποψήφια σημεία κλειδιά) της εικόνας.

Το επόμενο στάδιο στη διαδικασία εκτίμησης του βάθους αποτελεί το ταίριασμα αντίστοιχων σημείων μεταξύ των διαδοχικών φωτογραφιών στη στέρεο διάταξη. Η ακρίβεια στη λύση του προβλήματος του στέρεο ταιριάσματος έχει άμεσο αντίκτυπο στην ποιότητα της τρισδιάστατης αποκατάστασης της



Σχήμα 4.2: Αποτελέσματα εντοπισμού ακμών από δύο εικόνες. Πρώτη Γραμμή: Εφαρμογή Canny φίλτρου. Δεύτερη Γραμμή: Χρήση Γκαουσιανού φίλτρου. Τρίτη Γραμμή: Εφαρμογή μεθόδων σύγκρισης φάσης.

σκηνής. Ακολουθεί η συγκριτική μελέτη των τριών βασικών μεθόδων που εφαρμόστηκαν.

- Η πρώτη μέθοδος υιοθέτησε την κλασσική προσέγγιση με φίλτρα Canny για τον εντοπισμό των ακμών και μέθοδο συσχέτισης φωτεινότητας εικονοστοιχείων για το ταίριασμα χαρακτηριστικών.
- Η δεύτερη μέθοδος συνίσταται στον εντοπισμό σημείων κλειδιών μέσω μετασχηματισμού SIFT. Ο εντοπισμός σύμφωνα με τη συγκεκριμένη μέθοδο βασίζεται σε μετρήσεις πλάτους και προσανατολισμού κλίσης της εικόνας, οι οποίες οδηγούν στην τοπική αναπαράσταση της πληροφορίας μέσω μετασχηματισμών περιγραφής. για κάθε εικόνα της στέρεο διάταξης. Τέλος η διαδικασία ταιριάσματος βασίζεται στην αντιστοίχιση αυτών των τοπικών μαθηματικών αναπαραστάσεων πλάτους και προσανατολισμού με κριτήριο την Ευκλίδεια απόστασή τους.

Η τρίτη προσέγγιση στο εξεταζόμενο πρόβλημα στηρίχθηκε στην εφαρμογή μονογενών φίλτρων, όπως αυτα περιγράφηκαν παραπάνω. Η συγκεκριμένη και σταθερής φοράς κίνηση της κάμερας οριοθετεί και μια συγκεκριμένη περιοχή αναζήτησης στην εικόνα των αντίστοιχων σημείων, χαρακτηριστικών. Επιπρόσθετα η διαδικασία ορθοποίησης, τοποθετεί τα αντίστοιχα σημεία σε κοινή επιπολική ευθεία (epipolar line). Ενσωματώνοντας όλες αυτές τις παρατηρήσεις στην ανάλυσή μας, η περιοχή αναζήτησης μειώνεται δραστικά και τα εξαγόμενο ζευγάρια αντίστοιχων σημείων ή περιοχών στην εικόνα είναι πολύ πιο αξιόπιστα.

Κατά τη διάρκεια των πειραμάτων παρατηρήθηκε ότι το ταίριασμα σποραδικών σημείων δημιουργεί αρκετά προβλήματα, ειδικά όταν εφαρμόζονται οι κλασσικές μέθοδοι μέτρησης φωτεινότητας. Το γεγονός αυτό συναντάται γιατί οι υπό εξέταση εικόνες είναι εξωτερικού χώρου κάτω από τυχαίες συνθήκες φωτισμού. Αυτό σημαίνει ότι οι τιμές φωτεινότητας των εικονοστοιχείων εμπεριέχουν μεγάλο ποσοστό αβεβαιότητας. Μεταβολές στη σκίαση (σε μία ή περισσότερες φωτογραφίες), επαναλαμβανόμενα μοτίβα στις εικόνες και η ομοιογενής υφή οδηγούν σε τιμές φωτεινότητας που βρίσκονται πολύ κοντά για συγκεκριμένες περιοχές στις εικόνες και άρα σε μία πληθώρα υποψήφιων αντίστοιχων σημείων και χαρακτηριστικών. Μία πρώτη προσπάθεια μείωσης του παράγοντα αβεβαιότητας στο ταίριασμα αποτελεί και η χρήση γεωμετρικών οντοτήτων όπως συγκεκριμένα μοτίβα, σχήματα, ευθείες για συσχέτιση και όχι σποραδικά ή μεμονωμένα σημεία. Η πρώτερη γνώση της κίνησης της κάμερας χρησιμοποιήθηκε για την αναζήτηση αντίστοιχων σημείων σε κατεύθυνση αντίθετη από αυτήν της κάμερας. Η περιοχή αναζήτησης λοιπόν περιορίστηκε σε μία ευθεία στην εικόνα κατ' αντιστοιχία της ευθύγραμμης κίνησης της κάμερας. Το παράθυρο αναζήτησης λοιπόν επιλέχθηκε με πλάτος περίπου ίσο με το μισό πλάτος της εικόνας (οριζόντια διάσταση) και ύψος μερικά εικονοστοιχεία (εγκάρσια διάσταση). Παρόλα αυτά γεωμετρικές ιδιότητες των χαρακτηριστικών όπως μήκος και κατεύθυνση ευθείας αποκλίνουν σημαντικά από εικόνα σε εικόνα. Για παράδειγμα μία ευθεία που εντοπίζεται χωρικά στην μία εικόνα σε μία άλλη διαδοχική μπορεί να «σπάει» σε δύο ή περισσότερα μέρη. Εφαρμόστηκαν λοιπόν μέθοδοι συσχέτισης σημείων στις υποψήφιες για ταίριασμα ευθείες. Η τροποποιημένη αυτή προσέγγιση βελτίωσε αρκετά τα αποτελέσματα. Η βελτίωση της αξιοπιστίας των ταιριασμάτων επιτεύχθηκε με χρήση μεθόδων χαλάρωσης (relaxation methods) σε εικονοστοιχεία που ανήκουν σε γειτονιά του κάθε ζεύγους αντιστοιχίας. Τα μέτρα αξιοπιστίας που εξάγονται από τον έλεγχο ταιριασμάτων στη γειτονιά του υπό εξέταση ζεύγους οδηγούν στην αποδοχή του ή στην απόρριψή του.

#### 4.6 Τριδιάστατη Ψηφιακή Ανακατασκευή

Ένας από τους βασικούς στόχους της παρούσας διατριβής αποτελεί η μελέτη της απόδοσης των μεθόδων σύγκρισης φάσης και συχνοτικής πληροφορίας, σε σύγκριση με κλασσικές μεθόδους συσχέτισης φωτεινότητας ή SIFT αναπαράστασης σήματος και ταιριάσματος, στα πλαίσια ενός προβλήματος τρισδιάστατης ψηφιακής ανακατασκευής-αποτύπωσης βαθμονομημένων εικόνων.

Η διαδικασία στερεοσκοπικής αντιστοίχισης στοιχείων στις δύο στέρεο εικόνες οδηγεί στην απόκτηση αντίστοιχων ζευγών, οπότε ο υπολογισμός του θεμελιώδους πίνακα γίνεται εφικτός, όπως περιγράφηκε στο κεφάλαιο 3. Ο πίνακας βαθμονόμησης υπολογίζεται ακολουθώντας την προσέγγιση του Zhang, [87] (όπως παρουσιάστηκε στο κεφάλαιο 2). Συνεπώς οι πίνακες προβολής για τη στέρεο διάταξη των πειραμάτων μας υπολογίζονται μέσω του ουσιώδη πίνακα (essential matrix) και της γνωστής μαθηματικής αναπαράστασης όπως προτάθηκε από τους Hartley και Zisserman, [62]. Η μετρική πληροφορία της σκηνής ανακτάται με τη χρήση του πίνακα βαθμονόμησης της κάμερας. Ο υπολογισμός βάθους για κάθε ζεύγος αντίστοιχων σημείων επιτυγχάνεται αρχικά με χρήση τριγωνοποίησης και με περαιτέρω βελτίωση, ελαχιστοποιώντας το επαναπροβαλλόμενο λάθος, χρησιμοποιώντας τη μέθοδο Levenberg-Marquadt.

Συγκριτικά αποτελέσματα από τις τρεις μεθόδους τρισδιάστατης αποκατάστασης φαίνονται στα Σχήματα 4.3(α') και Σχήματα 4.3(β'). Η πρώτη γραμμή και στα δύο σχήματα παρουσιάζει τα αποτελέσματα που αποκτώνται με τη χρήση κλασικών μεθόδων για ανίχνευση χαρακτηριστικών στην εικόνα και για το ταίριασμα αυτών. Αυτές είναι το φίλτρο Canny και η συσχέτιση φωτεινότητας. Στη δεύτερη γραμμή των εν λόγω σχημάτων παρουσιάζεται η εφαρμογή του φίλτρου Γκαουσιανών διαφορών για τον εντοπισμό μετασχηματισμένων κατά SIFT χαρακτηριστικών. Το ταίριασμα αυτών επιτυγχάνεται όπως ήδη έχει αναφερθεί με χρήση διανυσματικών εκφράσεων τοπικής αναπαράστασης πληροφορίας γεωμετρίας και πλάτους και ελαχιστοποίηση της Ευκλείδειας απόστασής τους. Οι τοπικοί αυτοί περιγραφείς αναθέτουν την τιμή πλάτους και προσανατολισμού της κλίσης σε κάθε σημείο κλειδί σύμφωνα με την εργασία στο [21]. Το βέλτιστο ταίριασμα τέτοιων χαρακτηριστικών είναι εκείνο το υποψήφιο ζεύγος με την ελάχιστη Ευκλείδεια απόσταση των αμετάβλητων σε κλίμακα και περιστροφή διανυσμάτων περιγραφής.

Στην τρίτη γραμμή παρουσιάζονται τα αποτελέσματα που προκύπτουν από τη χρήση μέτρου σύγκρισης φάσης για τον εντοπισμό ακμών και μονογενή φίλτρα για το αξιόπιστο ταίριασμα των χαρακτηριστικών της εικόνας, όπως προτείνεται στην παρούσα διατριβή και περιγράφηκε στο κεφάλαιο αυτό.

Τα αποτελέσματα των Σχημάτων 4.3(α') και 4.3(β') οργανώνονται επίσης

και σε τρεις στήλες. Η πρώτη στήλη και στα δύο σχήματα αναπαριστά τα αποτελέσματα της τρισδιάστατης αποκατάστασης για τις τρεις προαναφερόμενες τεχνικές. Η δεύτερη στήλη παρουσιάζει τα ίδια αποτελέσματα με την τριδιάστατη δομή να έχει υποστεί μία στροφή μικρής γωνίας (περίπου 20 μοίρες), μία αναπαράσταση που οδηγεί σε πιο άμεση απεικόνιση της διακύμανσης του βάθους. Τέλος η τρίτη στήλη απεικονίζει μία χρωματική απεικόνιση του βάθους (με κόκκινο χρώμα τα εγγύτερα σημεία και με μπλε τα πιο απομακρυσμένα με μεγαλύτερο βάθος). Η αντιστοίχιση χρώματος και απόστασης γίνεται με γραμμικό τρόπο.

Παρατηρώντας τα αποτελέσματα αυτά είναι σαφές ότι η τρίτη γραμμή στην οποία παρατίθεται η απεικόνιση βάθους, η οποία προκύπτει από υπολογισμούς αποκλειστικά στο χώρο της συχνότητας, διαθέτει τη πιο ακριβή τριδιάστατη αναπαράσταση του θέματος με αρκετά ικανοποιητική διακύμανση του βάθους. Η διακύμανση του βάθους φαίνεται ξεκάθαρα στα αποτελέσματα της δεύτερης στήλης των Σχημάτων 4.3(α') και 4.3(β'). Στα αποτελέσματα που παρουσιάζονται στην πρώτη γραμμή έχουμε μία αρκετά αφηρημένη τριδιάστατη ψηφιακή αποτύπωση των θεμάτων των φωτογραφιών. Τέλος, η δεύτερη γραμμή των αποτελεσμάτων αποδεικνύει ότι και η SIFT τοπική περιγραφή περιοχών της εικόνας αμετάβλητων σε γεωμετρικούς μετασχηματισμούς, δεν οδηγεί σε ικανοποιητικά αποτελέσματα εκτίμησης βάθους. Η εικόνα της ανακατασκευής τόσο στη πρώτη, όσο και στη δεύτερη γραμμή των αποτελεσμάτων των σχημάτων 4.3(α') και 4.3(β') στερείται εμφανώς οποιασδήποτε δομικής πληροφορίας. Επιπρόσθετα, ο υπολογισμός συναρτήσεων περιγραφής των χαρακτηριστικών σημείων κλειδιά κατά το μετασχηματισμό SIFT βασίζεται ως γνωστό σε υπολογισμούς κλίσης για διαφορετικές τιμές Γκαουσιανής θόλωσης. Αυτό το είδος μετρήσεων οδηγεί στον εντοπισμό μικρών κορυφών της συνάρτησης Γκαουσιανών διαφορών, οι οποίες θα απορριφθούν σε δεύτερο χρόνο στον προσδιορισμό συναρτήσεων περιγραφής χαρακτηριστικών σημείων. Μία τέτοια προσέγγιση οδηγεί στον εντοπισμό σποραδικών σημείων κλειδιών που δεν αποτυπώνουν συνολικά μία συγκεκριμένη δομή. Άμεσο αποτέλεσμα αποτελεί η περιορισμένη σε τιμές διακύμανση του βάθους χωρίς ξεκάθαρη αίσθηση των δομικών στοιχείων της σκηνής, στη δεύτερη γραμμή και δεύτερη στήλη των σχημάτων 4.3(α') και 4.3(β').

## 4.7 Επανόρθωση Εικόνων - Αυτόματη Βαθμονόμηση

Τα παραπάνω αποτελέσματα χρησιμοποιήθηκαν για την υλοποίηση ενός σεναρίου αυτοβαθμονόμησης κάμερας. Υπολογίστηκε μέσω της διαδικασίας



(α') Εικονική Αποκατάσταση Ερεχθείου



(β') Πύλη Ηρώδειου

Σχήμα 4.3: Τριδιάστατη ψηφιακή ανακατασκευή από ζευγάρι πειραματικών στέρεο εικόνων. Η ταξινόμηση των αποτελεσμάτων στα δύο σχήματα είναι κοινή. Πρώτη Γραμμή: Κλασσικές τεχνικές που αξιοποιούν αποκλειστικά την πληροφορία της φωτεινότητας των εικονοστοιχείων για κάθε εικόνα. Δεύτερη Γραμμή: Ανακατασκευή σκηνής χρησιμοποιώντας το μέτρο της κλίσης για εντοπισμό ακμών και ταίριασμα SIFT σημείων κλειδιών. Τρίτη γραμμή: Οι υπολογισμοί πραγματοποιήθηκαν αποκλειστικά στο χώρο της συχνότητας-φάσης. Πρώτη Στήλη: Αποτελέσματα ανακατασκευής σκηνής χρησιμοποιώντας διαφορετικές μεθόδους εντοπισμού ακμών και ταιριάσματος χαρακτηριστικών. Δεύτερη Στήλη: Αποτελέσματα ανακατασκευής με στροφή μικρής γωνίας. Τρίτη Στήλη: Χρωματική απεικόνιση του βάθους. Το κόκκινο χρώμα αναπαριστά τις εγγύτερες περιοχές ενώ το μπλε τις μακρύτερες. διόρθωσης επιπολικής γεωμετρίας (επανόρθωση) η άγνωστη εστιακή απόσταση κάτω από τις εξής υποθέσεις:

- 1. Θεμελιώδες σημείο στο κέντρο της εικόνας  $(\frac{w}{2}, \frac{h}{2})$ , με w και h το πλάτος και το ύψος της εικόνας αντίστοιχα.
- 2. Έλλειψη λοξότητας.
- 3. Η εστιακή απόσταση ανά μονάδα απόστασης, τόσο στην οριζόντια όσο και στην κάθετη διάσταση, είναι ίση ( $\alpha_x = \alpha_y$ )

Ο πίνακας εσωτερικών παραμέτρων της κάμερας κάτω από τις παραπάνω υποθέσεις, γίνεται:

$$K = \begin{pmatrix} \alpha_x & 0 & w/2 \\ 0 & \alpha_y & h/2 \\ 0 & 0 & 1 \end{pmatrix}$$
(4.45)

Όταν εφαρμόζεται μέθοδος συσχέτισης βασισμένη σε μετρήσεις φωτεινότητας για το ταίριασμα ευθειών, η εστιακή απόσταση στις δύο κατευθύνσεις (x και y) είναι  $\alpha_x = \alpha_y = 437.5199 \pm 1.0158$  αντίστοιχα. Όταν οι υπολογισμοί γίνονται αποκλειστικά στο χώρο φάσης, οι τιμές εστιακής απόστασης γίνονται:  $\alpha_x = \alpha_y = 724.8499 \pm 0.9889$ . Τέλος η υλοποίηση αξιόπιστης βαθμονόμησης με επίπεδα μοτίβα, σύμφωνα με το Zhang, δίνει:  $\alpha_x = \alpha_y = 796.70289 \pm 0.06782$ . Είναι γεγονός λοιπόν ότι οι υπολογισμοί στο χώρο της φάσης παράγουν πιο ακριβή αποτελέσματα στα γεωμετρικά χαρακτηριστικά της κάμερας. Αυτό επιβεβαιώνεται από το γεγονός ότι τα αποτελέσματα στη δεύτερη περίπτωση αυτοβαθμονόμησης είναι πιο κοντά στα αποτελέσματα βαθμονόμησης με μοτίβα.

Η διαδικασία της επανόρθωσης αποτελεί το πρώτο βήμα στη διαδικασία πυκνής τρισδιάστατης εικονικής αποκατάστασης (dense reconstruction). Μέσω της επανόρθωσης ανάγουμε το πρόβλημα της αντιστοίχισης σε μονοδιάστατο πρόβλημα, αφού οι επιπολικές γραμμές σε κάθε φωτογραφία είναι παράλληλες μεταξύ τους και συμπίπτουν με τις εικονικές γραμμές σάρωσης. Δοκιμάσαμε την υλοποίηση της επανόρθωσης χωρίς πρότερη γνώση του πίνακα εσωτερικών παραμέτρων. Τα αποτελέσματα φαίνονται στο Σχήμα 4.4.

## 4.8 Ανακατασκευή Υπάρχουσας Δομής με χρήση Ουσιώδη Πίνακα (Essential Matrix)

Γνωρίζοντας τον πίνακα εσωτερικών παραμέτρων της κάμερας όπως έχει υπολογιστεί με τις δύο μεθόδους βαθμονόμησης και τις αντιστοιχίσεις των σημείων, υπολογίζουμε τους πίνακες προβολής των δύο καμερών με τη βοήθεια



Σχήμα 4.4: Επανόρθωση εικόνων με αυτόματη βαθμονόμηση.

του ουσιώδη πίνακα [62], όπως αναλύθηκε και στο κεφάλαιο 3. Κατόπιν με τη μέθοδο της τριγωνοποίησης υπολογίζονται τα τρισδιάστατα σημεία της δομής. Οι αντιστοιχίσεις σημείων για τον προσδιορισμό του θεμελιώδη και του ουσιώδη πίνακα υπολογίζονται αρχικά με μεθόδους συσχέτισης φωτεινότητας και κατόπιν με χρήση μονογενών φίλτρων (παράγραφος 4.4).

Οι πίνακες προβολής για τις κάμερες στην περίπτωση συσχέτισης φωτεινότητας είναι:

$$P_1^I = \begin{pmatrix} 832.1737 & 0 & 154.0099 & 0\\ 0 & 823.7304 & 26.8724 & 0\\ 0 & 0 & 1.0000 & 0 \end{pmatrix}$$
(4.46)

$$P_2^I = \begin{pmatrix} 835.3174 & 9.2035 & 135.6181 & -846.1462 \\ -6.5389 & 823.4567 & 128.4709 & -28.8908 \\ 0.0221 & -0.0017 & 0.9998 & -0.1654 \end{pmatrix}$$
(4.47)

Αντίστοιχα οι πίνακες προβολής για την περίπτωση υπολογισμών στο χώρο της συχνότητας είναι:

$$P_{1}^{f} = \begin{pmatrix} 838.2450 & 0 & 164.3724 & 0 \\ 0 & 829.1847 & 112.5955 & 0 \\ 0 & 0 & 1.0000 & 0 \end{pmatrix}$$
(4.48)  
$$P_{2}^{f} = \begin{pmatrix} 462.6197 & -15.1966 & 717.9310 & 337.2591 \\ 76.6036 & 833.2596 & -5.9450 & -72.5593 \\ 0.9285 & 0.0471 & -0.3683 & -0.8253 \end{pmatrix}$$
(4.49)

#### 4.9 Συμπεράσματα Κεφαλαίου

Βασικό συμπέρασμα της προσέγγισης που προτάθηκε στο παρών κεφάλαιο αποτελεί το γεγονός ότι η προτεινόμενη τροποποιημένη έκφραση σύγκρισης των εντοπισμένων χαρακτηριστικών, οδηγεί σε περισσότερο αξιόπιστα αποτελέσματα τριδιάστατης ψηφιακής εικονικής αποκατάστασης. Τα αποτελέσματα παρουσιάζουν μια περισσότερο ακριβή ανασύνθεση της μετρικής πληροφορίας της σκηνής. Η δομική πληροφορία αποκαθίσταται επίσης σε μεγαλύτερη έκταση στην εικόνα. Τα μοντέλα επεξεργασίας σήματος στο χώρο της φάσης εμφανίζονται ως αρκετά ικανοποιητική επιλογή στην τριδιάστατη ψηφιακή ανακατασκευή σύνθετων εικόνων, όπως εικόνες εξωτερικού χώρου, αρχαιολογικού ενδιαφέροντος. Επίσης, είναι εμφανής η ανάγκη επέκτασης αυτών των μοντέλων σε όλη την επιφάνεια του επιπέδου της εικόνας και όχι μόνο στις περιοχές των εντοπισμένων χαρακτηριστικών γεωμετρικών δομών στην εικόνα. Στη συνέχεια, στο δεύτερο μέρος της παρούσας διατριβής εφαρμόζοντας στοχαστικές μεθόδους ανάλυσης της ψηφιακής εικόνας στο χώρο της φάσης, ανακτάται πληροφορία μεταβολής του βάθους από συνεκτικές περιοχές της εικόνας και όχι από μεμονωμένα χαρακτηριστικά της. Η προτεινόμενη προσέγγιση που ακολουθεί συμβάλει επιπρόσθετα στην εξαγωγή συμπερασμάτων σχετικά με την ποιότητα της ψηφιακής αποτύπωσης.

## Μέρος ΙΙ

Στοχαστικές Μέθοδοι Αντιστοίχισης και Ανάκτηση Πληροφορίας Δομής σε Ακολουθίες Εικόνων

## Κεφάλαιο 5

# Θεωρητικό Υπόβαθρο Στοχαστικών Μεθόδων Ανάκτησης Πληροφορίας Δομής από Εικόνες

#### 5.1 Τυχαία Μαρκοβιανά Μοντέλα

Η πειραματική επαλήθευση χρήσης μεθόδων συσχέτισης χαρακτηριστικών σε προβλήματα τριδιάστατης ανακατασκευής έφερε στην επιφάνεια δύο υπαρκτά προβλήματα. Το πρώτο συνίσταται στο γεγονός ότι ο υπολογισμός του βάθους πεδίου περιορίζεται μόνο στις εντοπισμένες ακμές στην εικόνα, δημιουργώντας μία αποσπασματική και κατά συνέπεια ασαφή εικόνα της συνολικής τριδιάστατης δομής. Σημαντική επίσης έλλειψη της έως τώρα προσέγγισης αποτελεί η απώλεια βελτιστοποίησης των αποτελεσμάτων με βάση περιορισμούς που προκύπτουν από την ίδια τη διάταξη των απεικονιζόμενων θεμάτων. Το σύνολο λοιπόν αυτών των προβλημάτων απαιτεί μία συνολική αλλαγή στη μοντελοποίηση τόσο του προβλήματος αντιστοίχισης όμοιων χαρακτηριστικών στην εικόνα, όσο και της επιλογής κατάλληλων περιορισμών συναφών με το περιεχόμενο της εικόνας και των διακριτών απεικονιζόμενων περιοχών. Βασική μέριμνα πρέπει να δοθεί στην εύρεση της κατάλληλης συνάρτησης που θα οδηγήσει σε βέλτιστη λύση του συγκεκριμένου προβλήματος υπολογιστικής όρασης. Ο μεγάλος βαθμός αδεδαιότητας στους υπολογισμούς σε προβλήματα όρασης υπολογιστών επιβάλλει τη χρήση μεθόδων βελτιστοποίησης. Συχνά είναι ανέφικτο να βρεθεί η ιδανική λύση στα εξεταζόμενα προβλήματα, οπότε αναζητείται η βέλτιστη λύση σε συνάρτηση με συγκεκριμένους περιορισμούς του προβλήματος.

Απαραίτητη ομάδα περιορισμών που συντελούν στην κατανόηση της απεικονιζόμενης πληροφορίας αποτελούν οι περιορισμοί πλαισίου (contextual constraints). Η πληροφορία πλαισίου εξάγεται από το περιβάλλον των χαρακτηριστικών δομών της εικόνας, ακόμη και σε ένα πρώιμο επίπεδο επεξεργασίας των εικονοστοιχείων.

Η θεωρία Μαρκοβιανών Τυχαίων Πεδίων (Markov Random Fields, MRF) προτείνει ένα συστηματικό τρόπο μοντελοποίησης δομών στην εικόνα εξαρτώμενου από την περιβάλλουσα πληροφορία, όπως εικονοστοιχεία και συσχετισμένα χαρακτηριστικά. Η συγκεκριμένη μεθοδολογία βασίζεται στην αναπαράσταση της αμοιβαίας σχέσης των δομών με πιθανοτικές κατανομές υπό συνθήκη Markov. Οι κατανομές αυτές αναπαριστούν πρότυπα της εικόνας (υφή ή χαρακτηριστικά αντικειμένων) σύμφωνα με το Μπεϋζιανό πλαίσιο ανάλυσης. Η θεωρία Μαρκοβιανών Τυχαίων Πεδίων σε συνδυασμό με θεωρίες στατιστικής απόφασης και εκτίμησης ικανοποιεί την απαίτηση για βέλτιστη επίλυση προβλημάτων υπολογιστικής όρασης. Η μέγιστη εκ των υστέρων (ύστερη) εκτίμηση (Maximum a posteriori, MAP) αποτελεί από τα πιο διαδεδομένα στατιστικά εργαλεία βελτιστοποίησης. Η μοντελοποίηση λοιπόν με Μαρκοβιανά Τυχαία Πεδία σε συνδυασμό με το στατιστικό κριτήριο απόφασης ΜΑΡ συνιστά το μεθοδολογικό πλαίσιο ανάλυσης και βέλτιστης απόφασης MAP-MRF, το οποίο έχει χρησιμοποιηθεί εκτενώς και στα προβλήματα που εξετάζονται στην παρούσα διατριβή. Στη συγκεκριμένη μεθοδολογία η αντιπροσωπευτική συνάρτηση των περιορισμών είναι η εκ των υστέρων πιθανότητα υπόθεσης των ετικετών εικονοστοιχείων, οι οποίες ουσιαστικά απεικονίζουν τιμές συγκεκριμένων παραμέτρων του προβλήματος. Η βέλτιστη συνάρτηση μαζί με τις αντίστοιχες τιμές των παραμέτρων θα προσδιοριστούν με βάση τον κανόνα Bayes και τη πρότερη κατανομή των τιμών ετικετών καθώς επίσης και τη συνάρτηση πιθανοφάνειας των παρατηρούμενων δεδομένων.

Στο παρόν κεφάλαιο θα παρουσιαστεί μία συνοπτική περιγραφή στα μαθηματικά εργαλεία της συγκεκριμένης προσέγγισης και του τρόπου με τον οποίο αυτές οι ιδέες βρίσκουν εφαρμογή σε προβλήματα στερεοσκοπικής αντιστοίχισης και αποκατάστασης πληροφορίας βάθους πεδίου. Θα παρουσιαστούν επίσης θεωρητικά αποτελέσματα της μεθοδολογίας MAP-MRF. Ιδιαίτερο βάρος θα δοθεί τέλος στις τεχνικές βελτιστοποίησης των συναρτήσεων ενέργειας που μοντελοποιούν τους περιορισμούς του προβλήματος αντιστοίχισης.

#### 5.2 Ετικετοποίηση στην Ανάλυση Εικόνας

Πληθώρα ζητημάτων ανάλυσης εικόνας και υπολογιστικής όρασης μπορούν να θεωρηθούν ως προβλήματα ετικετοποίησης των οποίων η λύση αποτελείται

από ένα σύνολο τιμών ετικετών. Οι τιμές του συνόλου ετικετών αντιστοιχούν σε αυτές εικονοστοιχείων ή χαρακτηριστικών της εικόνας. Η ανάθεση ετικετών επίσης αποτελεί το βασικό εργαλείο μελέτης και εφαρμογής Μαρκοβιανών τυχαίων πεδίων.

Ένα πρόβλημα ετικετοποίησης ορίζεται στα πλαίσια ενός συνόλου από κόμβους και ενός συνόλου από ετικέτες. Έστω S ένα διακριτό σύνολο από m κόμβους:

$$\mathcal{S} = \{1, \dots, m\} \tag{5.1}$$

όπου  $1, \ldots, m$  είναι δείκτες. Ένας κόμβος συχνά αντιπροσωπεύει ένα σημείο, μία περιοχή στον Ευκλείδειο χώρο, όπως ένα εικονοστοιχείο ή ένα χαρακτηριστικό εικόνας, όπως μία γωνία, ένα τμήμα ευθείας ή επιφάνειας. Σε μία διδιάστατη εικόνα διάστασης  $n \times n$  μπορεί να αντιστοιχηθεί το ακόλουθο σύνολο κόμβων:

$$S = \{(i, j) \mid 1 \le i, j \le n\}$$
(5.2)

Τα στοιχεία του συνόλου αντιστοιχούν σε θέσεις εικονοστοιχειων της εικόνας και αναπαριστάται πλέον ως ένα πλέγμα. Συνεπώς μία ετικέτα μπορεί να θεωρηθεί ως μία μεταβαλλόμενη τιμή σε ένα κόμβο.

Έστω  $\mathcal{L}$  το σύνολο ετικετών. Ένα σύνολο ετικετών μπορεί να θεωρηθεί ως συνεχές ή διακριτό. Στην περίπτωση συνεχούς συνόλου μπορεί να θεωρηθεί ως ένα υποσύνολο του  $\mathbb{R}$ :

$$\mathcal{L}_r = [X_l, X_h] \subset \mathbb{R} \tag{5.3}$$

όπου  $X_l$  και  $X_h$  το πάνω και κάτω όριο τιμών αντίστοιχα, του συνόλου ετικετών. Ένα παράδειγμα αυτής της κατηγορίας αποτελεί το δυναμικό εύρος αναλογικής φωτεινότητας εικονοστοιχείων. Στη διακριτή περίπτωση, μία ετικέτα θεωρείται ότι παίρνει μία διακριτή τιμή από ένα σύνολο M ετικετών.

$$\mathcal{L}_{d} = \{l_1, \dots, l_M\}$$
(5.4)

ή πιο απλά

$$\mathcal{L}_d = \{1, \dots, M\} \tag{5.5}$$

Το πρόβλημα λοιπόν ετικετοποίησης συνίσταται στην ανάθεση μίας ετικέτας από το σύνολο  $\mathcal{L}$  σε καθένα από τους κόμβους του συνόλου S. Το σύνολο

$$f = \{f_1, \dots, f_m\}$$
(5.6)

ονομάζεται στο εξής μία ετικετοποίηση των κόμβων του συνόλου S όπου κάθε  $f_i$ λαμβάνει τιμές από ένα σύνολο ετικετών  $\mathcal{L}_i$  του κόμβου i (i = 1, ..., m).

Στην περίπτωση που σε όλους τους κόμβους ανατίθενται τιμές ετικετών από το ίδιο σύνολο  $\mathcal{L}$ , το σύνολο όλων των πιθανών ετικετοποιήσεων είναι το καρτεσιανό γινόμενο:

$$\mathbb{F} = \underbrace{\mathcal{L} \times \mathcal{L} \dots \times \mathcal{L}}_{m \text{ times}} = \mathcal{L}^m$$
(5.7)

όπου m το μέγεθος του S. Για ένα πρόβλημα με m κόμβους και M ετικέτες, υπάρχει ένας αριθμός από  $M^m$  πιθανές διατάξεις στο  $\mathbb{F}$ .

## 5.3 Μοντελοποίηση Πληροφορίας Πλαισίου με Τυχαία Μαρκοβιανά Μοντέλα

Κριτήριο στην επιλογή διαδικασιών ανάλυσης χαμηλού επιπέδου υπολογιστικής όρασης αποτελεί η ύπαρξη ή μη, πληροφορίας πλαισίου. Το είδος αυτό της πληροφορίας προκύπτει από τη γειτονική περιοχή της κάθε ετικέτας εικονοστοιχείου.

Η χρήση πληροφορίας περιορισμών πλαισίου (contextual constraints) είναι απαραίτητη σε διαδικασίες κατανόησης εικόνας [51] και αναγνώρισης προτύπων [16]. Η πληροφορία από τη γειτονιά εικονοστοιχείων χρησιμοποιείται στον υπολογισμό πιθανοτήτων υπόθεσης. Οι περιορισμοί πλαισίου εκφράζονται τοπικά με πιθανότητες υπόθεσης  $P(f_i|\{f_{i'}\})$ , όπου  $\{f_{i'}\}$  αποτελεί το σύνολο των ετικετών σε διαφορετικούς κόμβους  $i' \neq i$  ή συνολικά την από κοινού πιθανότητα P(f).

Στην περίπτωση ανεξαρτησίας μεταξύ των ετικετών (μη ύπαρξη πλαισίου), η από κοινού πιθανότητα προκύπτει από το γινόμενο της κάθε τοπικής πιθανότητας για καθεμία τιμή ετικέτας.

$$P(f) = \prod_{i \in S} P(f_i)$$
(5.8)

Αντίθετα, στο ενδεχόμενο ύπαρξης πληροφορίας πλαισίου, οι ετικέτες είναι αμοιβαία εξαρτώμενες, οπότε η παραπάνω σχέση δεν ισχύει. Στην κατεύθυνση λοιπόν εξαγωγής καθολικών συμπερασμάτων από την εικόνα χρησιμοποιώντας τοπική πληροφορία, τα Μαρκοβιανά Τυχαία Πεδία αποτελούν το κατάλληλο μαθηματικό εργαλείο στο πρόβλημα. Στη συνέχεια γίνεται μία συνοπτική παρουσίαση αρχών μαθηματικής θεμελίωσης των Μαρκοβιανών μοντέλων.

Έστω  $F = \{F_1, \ldots, F_m\}$  το σύνολο τυχαίων μεταβλητών, ορισμένο στο σύνολο S στο οποίο κάθε τυχαία μεταβλητή  $F_i$  παίρνει μία τιμή  $f_i$  από το  $\mathcal{L}$ . Το σύνολο F ονομάζεται τυχαίο πεδίο. Χρησιμοποιείται ο συμβολισμός  $F_i = f_i$  για να αναπαραστήσει την απόδοση της τιμής  $f_i$  στο  $F_i$  και ο συμβολισμός  $(F_1 = f_1, \ldots, F_m = f_m)$ ή πιο απλά F = f αναπαριστά το συνδυασμένο ενδεχόμενο, με  $f = \{f_1, \ldots, f_m\}$  μία διάταξη στο  $\mathbb{F}$ .

Η πιθανότητα, η τυχαία μεταβλητή  $F_i$  να πάρει την τιμή  $f_i$  από ένα διακριτό σύνολο ετικετών  $\mathcal{L}$  είναι  $P(F_i) = f_i$  και απλοποιημένα :  $P(f_i)$ , ενώ η από κοινού πιθανότητα (joint probability) συμβολίζεται:  $P(F = f) = P(F_1 = f_1, \ldots, F_m = f_m)$  και απλοποιημένα : P(f).

Το F ονομάζεται τυχαίο Μαρκοβιανό μοντέλο στο S σε μία γειτονιά  $\mathcal{N}$  εάν και μόνο εάν ικανοποιούνται οι ακόλουθες δύο συνθήκες.

$$P(f) > 0, \forall f \in \mathbb{F}$$
(5.9)

$$P(f_i \mid f_{\mathcal{S}-\{i\}}) = P(f_i \mid f_{\mathcal{N}_i})$$

$$(5.10)$$

όπου  $S - \{i\}$  είναι η διαφορά συνόλων, και  $f_{S-\{i\}}$  αναπαριστά το σύνολο ετικετών που αντιστοιχούν στους κόμβους  $S - \{i\}$ , και η παρακάτω σχέση:

$$f_{\mathcal{N}_{i}} = \{ f_{i'} \mid i' \in \mathcal{N}_{i} \}$$
(5.11)

αναπαριστά το σύνολο ετικετών σε κόμβους που γειτνιάζουν με τον κόμβο στη θέση i.

## 5.4 Πλαίσιο Λήψης Απόφασης MAP-MRF - Τεχνικές Βελτιστοποίησης

Η βελτιστοποίηση των διαδικασιών ανάλυσης εικόνας παίζει έναν πολύ σημαντικό ρόλο και αποτελεί βασικό παράγοντα στην επιτυχή εφαρμογή τέτοιων διαδικασιών υπολογιστικής όρασης. Τη χρήση μεθόδων βελτιστοποίησης επιβάλλει η ύπαρξη αδεβαιότητας σε κάθε στάδιο επεξεργασίας με αλγορίθμους υπολογιστικής όρασης. Χαρακτηριστικά παραδείγματα αποτελούν ο θόρυβος και η επικάλυψη περιοχών. Ακριβείς και τέλειες λύσεις στα προβλήματα υπολογιστικής όρασης δεν υπάρχουν. Αντίθετα, προσεγγιστικές λύσεις αλλά βέλτιστες ως προς κάποιο κριτήριο συναντώνται συχνά.

Υπάρχουν τρία βασικά ζητήματα στη βελτιστοποίηση αλγορίθμων υπολογιστικής όρασης: η αναπαράσταση του προβλήματος, ο προσδιορισμός αντιπροσωπευτικής συνάρτησης και ο κατάλληλος αλγόριθμος βελτιστοποίησης.

Αναφορικά με το πρώτο ζήτημα, δύο είναι οι βασικές προσεγγίσεις της αναπαράσταση του προβλήματος: η περιγραφική και η υπολογιστική. Η πρώτη αξιοποιεί ιδιότητες σχήματος των απεικονιζόμενων αντικειμένων, όπως αυτές σχετίζονται με φωτομετρία και γεωμετρία. Μία τέτοια προσέγγιση δε θα αναλυθεί στην παρούσα διατριβή. Η εναλλακτική υπολογιστική προσέγγιση αφορά τον τρόπο αναπαράστασης της λύσης, η οποία σχετίζεται με την επιλογή κόμβων και ετικετών στα πλαίσια ενός προβλήματος ετικετοποίησης.

Το δεύτερο ζήτημα συνίσταται στον προσδιορισμό χαρακτηριστικής συνάρτησης της οποίας η βελτιστοποίηση αντιστοιχεί στη λύση του εξεταζόμενου προβλήματος. Μία τέτοια συνάρτηση αποτελεί βασικό μέτρο της ποιότητας της λύσης στα πλαίσια ελαχιστοποίησης κάποιάς αντιπροσωπευτικής συνάρτησης κόστους. Βασικό ζητούμενο αποτελεί η κωδικοποίηση βασικών ιδιοτήτων εικονοστοιχείων όπως χρώμα, φωτεινότητα, ή πληροφορία πλαισίου.

Το τρίτο ζήτημα αφορά την αναζήτηση τοπικής βέλτιστης λύσης σε κυρτές συναρτήσεις με υπολογιστικά αποδοτικό τρόπο σε χώρο και χρόνο.

Στα προβλήματα αντιστοίχισης που εξετάζονται στην παρούσα διατριβή οι χαρακτηριστικές συναρτήσεις έχουν τη μορφή συνάρτησης ενέργειας. Η συνάρτηση ενέργειας αποτελεί ουσιαστικά ένα ποσοτικό μέτρο του κόστους αντιστοίχισης. Το κόστος υπολογίζεται από την απόσταση των διανυσμάτων τιμών ιδιοτήτων των υποψήφιων για ταίριασμα σημείων ή περιοχών της εικόνας.

Τα δύο βασικά ζητούμενα στη διαμόρφωση της συνάρτησης ενέργειας, είναι η μορφή της και οι χρησιμοποιούμενες παράμετροι. Μορφή και παράμετροι ορίζουν τη συνάρτηση ενέργειας, η οποία με τη σειρά της ορίζει την τοπικά ελάχιστη λύση. Η μορφή εξαρτάται από τις υποθέσεις που έχουν γίνει αναφορικά με τη λύση της τελικής ετικετοποίησης f και τα παρατηρηθέντα δεδομένα d. Αυτό εκφράζεται ως:  $E(f \mid d)$ , όπου  $E(\cdot)$  μία συναρτησιακή μορφή ενέργειας.

Αν οι παράμετροι του προβλήματος εκφράζονται από τη μεταβλητή θ τότε η συνάρτηση ενέργειας επεκτείνεται ως:  $E(f \mid d, \theta)$ . Γενικά για μία δεδομένη συναρτησιακή μορφή  $E(\cdot)$  και διαφορετικά d ή  $\theta$ , προκύπτει και διαφορετική συνάρτηση ενέργειας,  $E(f \mid d, \theta)$  ως προς f με πιθανόν διαφορετικό τοπικό ελάχιστο,  $f^*$ . Από τη στιγμή που οι παράμετροι αποτελούν μέρος του ορισμού της συνάρτησης ενέργειας  $E(f \mid d, \theta)$ , το ελάχιστο  $f^* = \arg minE(f \mid d)$  δεν είναι πλήρως ορισμένο αν δεν έχουν προσδιοριστεί οι παράμετροι ακόμη και για γνωστή μορφή της συνάρτησης ενέργειας. Η διαδικασία εύρεσης καθολικού ελάχιστοι αποτελεί συχνά ανέφικτο στόχο. Έτσι λοιπόν αναζητούνται τοπικά ελάχιστα, συχνά υπό την επίδραση αστάθμητων παραγόντων στην εικόνα, όπως θόρυβος ή τυχαίες διακυμάνσεις στη φωτεινότητα εξαιτίας πιθανού εξωτερικού φωτισμού. Αρχές λοιπόν δανεισμένες από πιθανοτικά μοντέλα και στατιστική, αξιοποιούνται για την ελαχιστοποίηση των εν λόγω συναρτήσεων. Η Μπεΰζιανή στατιστική αποτελεί θεμελιώδη μέθοδο στην εκτίμηση και λήψη αποφάσεων κάτω από τέτοιες συνθήκες.

Σύμφωνα με τη Μπεϋζιανή στατιστική, όταν η εκ των προτέρων (πρότερη)

κατανομή πιθανότητας και η συνάρτηση πιθανοφάνειας ενός χαρακτηριστικού της εικόνας είναι γνωστά, η βελτιστοποίηση επιτυγχάνεται με τη μεγιστοποίηση της εκ των υστέρων (ύστερης) πιθανότητας (maximum a posterior solution, MAP) [71]. Οι ιδέες αυτές πιθανοτικών κατανομών και συναρτήσεων ενέργειας ανήκουν σε πλαίσιο ανάλυσης και αξιολόγησης πληροφορίας εικόνας κατά MAP-MRF.

Σύμφωνα με τη Μπεϋζιανή προσέγγιση, η βελτιστοποίηση επιτυγχάνεται μέσω της ελαχιστοποίησης μίας συνάρτησης ρίσκου ως προς  $f^*$ , η οποία ορίζεται ως εξής:

$$R(f^*) = \int_{f \in \mathbb{F}} C(f^*, f) P(f \mid d) df$$
(5.12)

όπου d είναι η τιμή της παρατήρησης,  $C(f^*, f)$ , είναι η συνάρτηση του κόστους και  $P(f \mid d)$  η ύστερη κατανομή. Η κατανομή αυτή υπολογίζεται χρησιμοποιώντας τον κανόνα Bayes:

$$P(f \mid d) = \frac{p(d \mid f)P(f)}{p(d)}$$
(5.13)

όπου P(f) είναι η εκ των προτέρων (πρότερη) πιθανότητα των ετικετοποιήσεων f, p(d|f) είναι η υπό συνθήκη συνάρτηση πυκνότητας πιθανότητας των παρατηρήσεων d, που ονομάζεται επίσης συνάρτηση πιθανοφάνειας (likelihood) των δεδομένων παρατήρησης d για δεδομένη ετικετοποίηση f.

Η συνάρτηση κόστους  $C(f^*, f)$  προσδιορίζει το κόστος εκτίμησης του fόταν η τιμή του είναι  $f^*$ . Οι συνήθεις επιλογές για τη συνάρτηση κόστους είναι δύο. Η συνάρτηση τετραγωνικού κόστους:

$$C(f^*, f) = \|f^* - f\|^2$$
(5.14)

και η συνάρτηση κόστους  $\delta(0-1)$ :

$$C(f^*, f) = \begin{cases} 0 εάν ||f^* - f|| \le \delta \\ 1 \delta i a φορετικά \end{cases}$$
(5.15)

όπου  $\|\alpha - \beta\|$  είναι η απόσταση μεταξύ  $\alpha$  και  $\beta$  και  $\delta > 0$  μία οποιαδήποτε μικρή σταθερά. Τα διαγράμματα για κάθε συνάρτηση κόστους φαίνονται στο Σχήμα 5.1. Η συνάρτηση ρίσκου κατά Bayes αποτελεί τη διακύμανση μιας εκτίμησης.

$$R(f^*) = \int_{f \in \mathbb{F}} \|f^* - f\|^2 P(f|d) df$$
(5.16)

Για  $\frac{\partial R(f^*)}{\partial f^*} = 0$  καταλήγουμε στη βέλτιστη εκτίμηση της ελάχιστης διακύμανσης (minimum variance):



Σχήμα 5.1: Βασικά είδη συναρτήσεων κόστους. Από αριστερά : συνάρτηση κόστους  $\delta(0-1)$  και συνάρτηση τετραγωνικού κόστους

$$f^* = \int_{f \in \mathbb{F}} fP(f|d)df$$
(5.17)

η οποία αποτελεί τη μέση τιμή της ύστερης πιθανότητας. Στην περίπτωση  $\delta(\cdot)$  συνάρτησης κόστους, η συνάρτηση ρίσκου κατά Bayes, γίνεται:

$$R(f^*) = \int_{f: \|f^* - f\| > \delta} P(f|d) df = 1 - \int_{f: \|f^* - f\| \le \delta} P(f|d) df$$
(5.18)

Όταν  $\delta \rightarrow 0$ , η σχέση (5.18) προσεγγίζεται:

$$R(f^*) = 1 - kP(f|d)$$
(5.19)

όπου k είναι ο όγκος στο χώρο που περιέχει όλα τα σημεία f για τα οποία  $||f^* - f|| \le \delta$ . Η ελαχιστοποίηση της (5.19) ισοδυναμεί με τη μεγιστοποίηση της ύστερης πιθανότητας. Επομένως, η εκτίμηση του ελάχιστου ρίσκου γίνεται:

$$f^* = \underset{f \in \mathbb{F}}{\operatorname{argmax}} P(f|d) \tag{5.20}$$

η οποία αποτελεί τη μέγιστη εκ των υστέρων εκτίμηση (Maximum A Posterior estimate, MAP estimate). Θεωρώντας σταθερή την πιθανότητα p(d) για σταθερή κατανομή d, σύμφωνα με τον κανόνα Bayes, (5.13), η P(f|d) είναι ανάλογη της συνδυαστικής κατανομής.

$$P(f|d) \propto P(f|d) = p(d|f)P(f)$$
(5.21)

Συνεπώς η μέγιστη εκ των υστέρων εκτίμηση, ισοδύναμα γίνεται:

$$f^* = \operatorname*{argmax}_{f \in \mathbb{F}} \{ p(d|f) P(f) \}$$
(5.22)

Εξετάζοντας την περίπτωση μίας MAP-MRF ετικετοποίησης. η P(f|d) αποτελεί την ύστερη κατανομή ενός Μαρκοβιανού Τυχαίου Πεδίου (MRF). Επομένως σημαντικό βήμα στην ετικετοποίηση ενός MRF αποτελεί ο προσδιορισμός της κατανομής P(f|d).

Εξετάζοντας το πρόβλημα στερεοσκοπικής αντιστοίχισης και υιοθετώντας την προσέγγιση στο [76] θεωρούμε το εν λόγω πρόβλημα ως ένα πρόβλημα ενέργειας και διατήρησης θερμικής ισορροπίας. Τα πιθανά ζεύγη σημείων αν είναι πετυχημένα τείνουν να μειώσουν την συνολική ενέργεια του συστήματος για συγκεκριμένη θερμοκρασία και είναι αποδεκτά. Στην αντίθετη περίπτωση τείνουν να αυξήσουν τη συνολική ενέργεια του συστήματος. Ένα σύστημα σε θερμική ισορροπία υπακούει στην κατανομή κατά Boltzman:

$$P(E) = e^{-\frac{E}{T}}$$
(5.23)

όπου E είναι η ενέργεια του συστήματος, P(E) η από κοινού πιθανότητα ενός συστήματος να έχει ενέργεια E με θερμοκρασία συστήματος T. Η κατά Bayes προσέγγιση της παραπάνω φυσικής αναπαράστασης του προβλήματος οδηγεί στην από κοινού πρότερη κατανομή μίας ετικετοποίησης f:

$$P(f) \propto e^{-U(f)} \tag{5.24}$$

όπου U(f) αποτελεί την πρότερη ενέργεια του προβλήματος. Κατά συνέπεια η ύστερη πιθανότητα είναι:

$$P(f|d) \propto e^{-U(f|d)} \tag{5.25}$$

Σύμφωνα με τον κανόνα Bayes, η ύστερη ενέργεια του συστήματος μπορεί επομένως να γραφεί ως:

$$U(f|d) = U(d|f) + U(f)$$
(5.26)

για σταθερή κατανομή παρατηρήσεων U(d), μιας και έχουμε σταθερό σύνολο d. Στην παραπάνω σχέση (5.26), η U(f) εκφράζει την πρότερη ενέργεια, ενώ η συνάρτηση U(d|f) εκφράζει την ενέργεια των δεδομένων παρατήρησης (ενέργεια πιθανοφάνειας). Αρκεί λοιπόν να προσδιοριστεί η ύστερη ενέργεια U(f|d), για να βρεθεί στη συνέχεια η μέγιστη ύστερη πιθανότητα εκτίμησης μέσω της ελάχιστης ύστερης ενέργειας (MAP-MRF solution):

$$f^* = \arg\min_{f} U(f|d) \tag{5.27}$$

Συνεπώς το πρόβλημα στέρεο αντιστοίχισης μπορεί να αντιμετωπιστεί α-ναζητώντας την κατά MAP-MRF ετικετοποίηση f τέτοια ώστε για  $S \to \mathcal{L}$  να

ελαχιστοποιείται η συνάρτησή ύστερης ενέργειας, (5.26). Σύμφωνα με αυτή την προσέγγιση ο όρος που εκφράζει την ενέργεια πιθανοφάνειας των δεδομένων ( $U(f \| d)$ ) μπορεί να ορισθεί με τέτοιο τρόπο ώστε να αντιπροσωπεύει τους περιορισμούς στερεοσκοπικής συνέπειας, ενώ ο όρος που εκφράζει την ενέργεια πρότερης κατανομής πιθανότητας (U(f)) μπορεί να ορισθεί με τέτοιο τρόπο ώστε να αντικατροπτίζει τους περιορισμούς ομαλότητας.

Ο περιορισμός στερεοσκοπικής συνέπειας μπορεί να περιγραφεί ως εξής: εάν δύο εικονοστοιχεία προέρχονται από τη προβολή του ίδιου σημείου της τριδιάτατης επιφάνειας, τότε έχουν κοντινές τιμές φωτεινότητας. Αν η φωτεινότητα του εικονοστοιχείου *i* είναι I(i), τότε ορίζεται μία συνάρτηση απόκλισης φωτεινότητας των δύο εικονοστοιχείων:

$$D(i,i') = |I(i) - I(i')|^2$$
(5.28)

Επομένως, στην περίπτωση αυτή, ο περιορισμός στερεοσκοπικής συνέπειας μπορεί να εκφρασθεί μέσω της ακόλουθης συνάρτησης ενέργειας πιθανοφάνειας (δεδομένων):

$$U(d|f) = \sum_{(i,i') \in S} D(i,i')$$
(5.29)

Ο περιορισμός ομαλότητας ορίζεται ως το άθροισμα των γειτονικών αποκλίσεων (γραμμικών ή τετραγωνικών) στις τιμές βάθους. Ο στόχος είναι η διατήρηση ομαλής μεταβολής του βάθους από περιοχή σε περιοχή του αντικειμένου.

$$U_{s}(f) = \sum_{(i,i') \in \mathcal{N}} \min\{|f_{i} - f_{i'}|, \gamma\}$$
(5.30)

όπου  $\gamma > 0$  μία σταθερά, και  $\mathcal{N}$  μία σχέση γειτνίασης εικονοστοιχείων. Έτσι λοιπόν διαμορφώνεται η συνάρτηση συνολικής ενέργειας, η ελαχιστοποίηση της οποίας οδηγεί στην εύρεση των αντίστοιχων εικονοστοιχείων από εικόνα σε εικόνα. Η μαθηματική έκφραση της συνολικής ενέργειας σύμφωνα με τη σχέση (5.26), ακολουθεί:

$$E_{total} = \sum_{(i,i')\in\mathcal{S}} D(i,i') + \sum_{(i,i')\in\mathcal{N}} \min\{|f_i - f_{i'}|,\gamma\}$$
(5.31)

Στις επόμενες παραγράφους θα παρουσιαστούν και θα υλοποιηθούν αλγόριθμοι ελαχιστοποίησης της παραπάνω συνάρτησης. Επίσης θα διερευνηθούν εναλλακτικές εκφράσεις στη συνάρτηση στερεοσκοπικής συνέπειας που βασίζονται σε πιο αξιόπιστα μέτρα αναπαράστασης της ψηφιακής πληροφορίας.

## 5.5 Συναρτήσεις Ενέργειας - Βέλτιστες Μέθοδοι-Ελαχιστοποίησης

Πληθώρα αλγορίθμων στην υπολογιστική όραση περιλαμβάνουν τον χαρακτηρισμό των στοιχειωδών εικονοστοιχείων μιας εικόνας (pixel) μέσω της ανάθεσης της τιμής μιας παραμέτρου όπως για παράδειγμα της *ανομοιότητας* βά∂ους (disparity). Ένας συνήθης περιορισμός που τηρείται είναι οι τιμές χαρακτηρισμού των εικονοστοιχείων να μεταβάλλονται ομαλά σε όλη την εικόνα διατηρώντας παράλληλα τις όποιες ασυνέχειες υπάρχουν, όπως για παράδειγμα εκείνες στα όρια των αντικειμένων. Υπάρχει μία ευρεία κλάση από στάθμες ενέργειας με διάφορα είδη περιορισμών και κριτηρίων στην ομαλότητα των τιμών. Αναζητούμε ένα καθολικό ελάχιστο τέτοιων συναρτήσεων ενέργειας, το οποίο αποτελεί ένα πρόβλημα NP ακόμη και στην απλούστερη συνθήκη διατήρησης τοπικής ασυνέχειας. Παρακάτω θα παρουσιαστεί ένας αλγόριθμος βασισμένος σε αποκοπή γράφων (graph cuts) σύμφωνα με την εργασία των Βογκον et al. [86] για την εύρεση τοπικού ελάχιστου με χρήση ευρείας κίνησης στο χώρο των εικονοστοιχείων (pixels).

Πληθώρα προβλημάτων πρώιμης όρασης απαιτούν τον υπολογισμό χωρικά μεταβαλλόμενων ποσοτήτων (φωτεινότητα ή ανομοιότητα) σε θορυβώδες περιβάλλον. Τέτοιες ποσότητες τείνουν να μεταβάλλονται με ομαλό τρόπο στην επιφάνεια ενός αντικειμένου σε αντίθεση με τα όρια του αντικειμένου, όπου μεταβάλλονται ταχύτατα. Σε κάθε εικονοστοιχείο (pixel)  $p \in \mathcal{P}$  ανατίθεται μία τιμή ετικέτας από ένα πεπερασμένο σύνολο  $\mathcal{L}$ , με  $\mathcal{P}$  το σύνολο εικονοστοιχείων. Ο στόχος της παρούσας διατριβής είναι η εύρεση ενός συνόλου ετικετοποίησης (labeling), όπου f είναι ταυτόχρονα τμηματικά ομαλή και συνεπής με τα δεδομένα, έτσι ώστε σε κάθε εικονοστοιχείο  $p \in \mathcal{P}$  να ανατίθεται μία ετικέτα  $f_p \in \mathcal{L}$ . Τέτοιας φύσεως προβλήματα υπολογιστικής όρασης μπορούν να αντιμετωπιστούν ως πρόβλημα ελαχιστοποίησης κάποιας συνάρτησης ενέργειας. Αναζητείται λοιπόν, σε αυτό το πλαίσιο, ένα σύνολο ετικετών f έτσι ώστε να ελαχιστοποιείται η ενέργεια:

$$E(f) = E_{\text{smooth}}(f) + E_{\text{data}}(f)$$
(5.32)

Η ποσότητα  $E_{\text{smooth}}$  εκφράζει την ενέργεια που σχετίζεται με την ομαλότητα της συγκεκριμένης ετικετοποίησης f, και αποτελεί ουσιαστικά ένα μέτρο της περιοχής της εικόνας στην οποία το σύνολο ετικετών f δεν είναι τμηματικά ομαλό. Η ποσότητα  $E_{\text{δατα}}$  εκφράζει την ενεργεια (πιθανοφάνειας) των δεδομένων, και αποτελεί ουσιαστικά ένα μέτρο μέτρο της απόκλισης μεταξύ f και μετρήσεων.

Στη πρόσφατη βιβλιογραφία έχουν προταθεί πολλές διαφορετικές συναρτήσεις ενέργειας. Η ακολουθούμενη μορφή ενέργειας δεδομένων στη συγκεκριμένη διατριβή δίνεται από τη σχέση:

$$E_{data}(f) = \sum_{p \in \mathcal{P}} D_p(f)$$
(5.33)

Η συνάρτηση  $D_p$  αντιπροσωπεύει το μέτρο ταιριάσματος ετικέτας f και πραγματικής τιμής στο εικονοστοιχείο p. Για παράδειγμα, σε ένα πρόβλημα στέρεο αντιστοίχισης το μέτρο  $D_p(f)$  ισούται με  $(f - I_p)^2$ , όπου  $I_p$  η μετρούμενη φωτεινότητα στο εικονοστοιχείο p.

Η επιλογή επίσης της συνάρτησης  $E_{smooth}$  αποτελεί ζήτημα κρίσιμης σημασίας. Οι συναρτήσεις που έχουν προταθεί στη βιβλιογραφία είναι πολλές. Για παράδειγμα, σε κάποιες υλοποιήσεις [86] και [77] έχει γίνει χρήση  $E_{smooth}$  που διατηρεί την ετικετοποίηση f παντού ομαλή. Μία τέτοια αντιμετώπιση στο πρόβλημα έχει φτωχά αποτελέσματα στα περιγράμματα αντικειμένων.

Οι συναρτήσεις που δεν προκαλούν την καθολική ομαλοποίηση των τιμών της ετικέτας ονομάζονται συναρτήσεις διατήρησης ασυνέχειας. Μεγάλη ποικιλία από τέτοιες συναρτήσεις έχουν προταθεί στα [86], [23] και [22]. Σημαντική δυσκολία που συναντάται σε προβλήματα ελαχιστοποίησης της ενέργειας είναι το τεράστιο υπολογιστικό κόστος. Τέτοιου είδους συναρτήσεις ενέργειας διαθέτουν μεγάλο αριθμό τοπικών ελαχίστων και ακόμη περισσότερο ο χώρος όλων των δυνατών τιμών ετικέτας έχει διάσταση |P|, κάτι το οποίο οδηγεί σε πληθώρα λύσεων.

Το είδος των συναρτήσεων ενέργειας που χρησιμοποιήσαμε στην παρούσα διατριβή προκύπτουν από τη Μπεϋζιανή μοντελοποίηση του συνόλου ετικετών σε Μαρκοβιανά τυχαία πεδία όπως αναλύθηκε εκτενώς σε προηγούμενες παραγράφους. Θεωρούμε, ενέργειες της μορφής:

$$E(f) = \sum_{(p,q) \in \mathcal{N}} V_{p,q}(f_p, f_q) + \sum_{p \in P} D_p(f_p)$$
(5.34)

όπου  $\mathcal{N}$  το σύνολο των αλληλεπιδρόντων ζευγαριών εικονοστοιχείων. Συνήθως το σύνολο  $\mathcal{N}$  αποτελείται από γειτονικά εικονοστοιχεία αλλά υπάρχει η δυνατότητα να είναι και αυθαιρέτως επιλεγμένα. Αξίζει να σημειωθεί ότι κάθε ζευγάρι εικονοστοιχείων (p,q) μπορεί να αντιστοιχεί σε διαφορετική τιμή αποκοπής  $V_{p,q}$ . Στο εξής θα απλοποιήσουμε τον παραπάνω συμβολισμό σε V αντί για  $V_{p,q}$ . Χρησιμοποιήθηκαν αλγόριθμοι από την εργασία των Boykov et al, [86], για την προσεγγιστική ελαχιστοποίηση της ενέργειας E(f)για ένα τυχαίο πεπερασμένο σύνολο ετικετών  $\mathcal{L}$  κάτω από την αλληλεπίδραση δύο κλάσεων συναρτήσεων αποκοπής V: μετρικών και ημιμετρικών. Η V
ονομάζεται μετρική όταν σε ένα χώρο ετικετών  $\mathcal{L}$ , ικανοποιεί:

$$V(\alpha,\beta) = 0 \iff \alpha = \beta, \tag{5.35}$$

$$V(\alpha,\beta) = V(\beta,\alpha) \ge 0,$$
(5.36)

$$V(\alpha,\beta) \le V(\alpha,\gamma) + V(\gamma,\beta) \tag{5.37}$$

για κάθε  $\alpha, \beta, \gamma \in \mathcal{L}$ . Αν το V ικανοποιεί μόνο τις σχέσεις (5.35) και (5.36) τότε ονομάζεται ημιμετρικό (semimetric). Αξίζει να σημειωθεί ότι τόσο οι μετρικές συναρτήσεις όσο και οι ημιμετρικές περιλαμβάνουν σημαντικές περιπτώσεις διατήρησης ασυνέχειας. Συμπληρωματικά μία μαθηματική έκφραση διατήρησης ασυνέχειας πρέπει να έχει όριο στο μεγαλύτερο δυνατό κατώφλι αποκοπής. Με αυτό το τρόπο αποφεύγεται μεγάλη αποκοπή τιμών, μεταξύ των αποκλίσεων στις γειτονικές ετικέτες των εικονοστοιχείων. Παρακάτω θα παρατεθούν κάποια βασικά παραδείγματα από τη βιβλιογραφία, από συναρτήσεις που ικανοποιούν την ιδιότητα διατήρησης ασυνέχειας. Θεωρώντας μονοδιάστατο σύνολο ετικετών  $\mathcal{L}$  θα παρουσιάσουμε την συνάρτηση αποκοπής δευτέρου βαθμού  $V(\alpha, \beta) = min(K, |\alpha - \beta|^2)$  (ημιμετρική) και τη συνάρτηση απόλυτης απόλυτη τιμή  $|\cdot|$  από μία νόρμα.

Μία άλλη σημαντική συνάρτηση διατήρησης ασυνέχειας δίνεται από το μοντέλο Potts, [66]:  $V(\alpha, \beta) = K \cdot T(\alpha \neq \beta)$ , όπου  $T(\cdot)$  είναι 1 αν η συνθήκη που περιέχει είναι δυαδικά αληθής, σε διαφορετική περίπτωση 0. Ένα τέτοιο μοντέλο οδηγεί στη δημιουργία ομοιογενών περιοχών στις οποίες τα εικονοστοιχεία έχουν ίδια τιμή στη ετικέτα. Τέτοια μοντέλα ονομάζονται και τμηματικά αμετάβλητα. Οι Geman et al, [17] ήταν οι πρώτοι που χρησιμοποίησαν το μοντέλο αυτό στην όραση υπολογιστών. Ασυνέχειες μεταξύ κάθε ζευγαριού ετικετών αποκόβονται ισοδύναμα. Από μία άποψη αυτό το μοντέλο αποτελεί την απλούστερη περίπτωση μοντέλου διατήρησης ασυνέχειας, και είναι ιδιαίτερα χρήσιμο όταν οι ετικέτες είναι μη ταξινομημένες ή ο αριθμός τους μικρός.

Το είδος των συναρτήσεων που θα αναλυθούν δίνονται από τη σχέση (5.32). Οι μέθοδοι βασισμένοι στην ενέργεια οδηγούν στη μοντελοποίηση καθολικών ιδιοτήτων της εικόνας που δεν είναι προσβάσιμες από τεχνικές συσχέτισης. Ένα βασικό μειονέκτημα σε αυτή τη προσέγγιση είναι η εγγενής δυσκολία να βρεθεί καθολικό ελάχιστο στις πιο χρήσιμες και ενδιαφέρουσες εκφράσεις ενέργειας. Εξαιτίας του μεγάλου υπολογιστικού κόστους κατά την εύρεση καθολικού ελαχίστου, πολλοί ερευνητές επιλέγουν ως λύση ένα τοπικό ελάχιστο. Γενικά όμως το τοπικό ελάχιστο μπορεί να βρίσκεται αυθαίρετα μακριά από το ολικό ελάχιστο. Επιπρόσθετα η επιλογή του τοπικού ελαχίστου οδηγεί σε επιπλέον προβλήματα. Για παράδειγμα, σε περίπτωση αποτυχίας εύρεσης τοπικού ελαχίστου δεν είναι δυνατή η εύρεση της αιτίας για αυτό. Τόσο η μεγάλη απόσταση του τοπικού ελαχίστου από το καθολικό όσο και η λανθασμένη επιλογή της συνάρτησης ενέργειας αποτελούν εν δυνάμει αιτίες του γεγονότος αυτού. Ένα άλλο κοινό πρόβλημα είναι το γεγονός ότι οι τεχνικές εύρεσης τοπικού ελαχίστου είναι ιδιαίτερα ευαίσθητες στις διαφορετικές αρχικοποιήσεις του αλγόριθμου. Γενικά, ένα σύνολο ετικετών f αντιστοιχεί σε τοπικό ελάχιστο της ενέργειας E αν ισχύει, για κάθε f' κοντά στο f:

$$E(f) \le E(f') \tag{5.38}$$

Στην περίπτωση διακριτής ετικετοποίησης, το σύνολο ετικετών κοντά στο f είναι εκείνο που προκύπτει μετά την εφαρμογή μονής αλλαγής τιμής ετικέτας στο f. Πληθώρα τεχνικών τοπικής ελαχιστοποίησης χρησιμοποιούν αυτό που ονομάζεται «τυπικές κινήσεις», σύμφωνα με τις οποίες κάθε στιγμή αλλάζει η τιμή ενός μόνο εικονοστοιχείου. Για την περίπτωση λοιπόν τυπικών κινήσεων η σχέση (5.38) υποδεικνύει το γεγονός ότι εαν κατά τη διαδικασία ελαχιστοποίησης βρεθεί ένα τοπικό ελάχιστο με βάση τυπικές κινήσεις τότε η ενέργεια δεν μπορεί να μειωθεί περαιτέρω από την αλλαγή της τιμής ετικέτας ενός μοναδικού εικονοστοιχείου. Το γεγονός αυτό αποτελεί μία εξαιρετικά ασθενή συνθήκη και οδηγεί σε λύσεις χαμηλής αξιοπιστίας. Ένα παράδειγμα τοπικής μεθόδου χρησιμοποιώντας «τυπικές κινήσεις» αποτελούν οι υπο-συνθήκη επαναληπτικοί τρόποι (Iretated Conditional Modes (ICM)) όπως παρουσιάστηκε στο [86]. Σύμφωνα με τη μέθοδο αυτή αναζητείται επαναλαμβανόμενα για κάθε εικονοστοιχείο εκείνη η τιμή ετικέτας για την οποία συμβαίνει η μεγαλύτερη μείωση μέχρι τη σύγκλιση σε τοπικό ελάχιστο. Ένα άλλο παράδειγμα αλγορίθμου που χρησιμοποιεί «τυπικές κινήσεις» είναι η προσομοιωμένη ανόπτηση (simulated annealing). Η τεχνική αυτή στην όραση υπολογιστών [71], έγινε ιδιαίτερα δημοφιλής γιατί θεωρητικά μπορεί να βελτιστοποιήσει οποιαδήποτε αρχική συνάρτηση ενέργειας, με πολύ εύκολη υλοποίηση. Δυστυχώς όμως χρειάζεται εκθετικό χρόνο για να συγκλίνει, εκτός βέβαια αν εκτελεστεί για μεγάλο χρονικό διάστημα μπορεί θεωρητικά να συγκλίνει στο καθολικό ελάχιστο. Στο [18] καταδεικνύεται ότι ο αλγόριθμος προσομοιωμένης ανόπτησης δίνει αποτελέσματα αρκετά μακριά από το καθολικό ελάχιστο ακόμη και στην απλή περίπτωση συνόλου ετικετών δυαδικών τιμών. Στην προσπάθεια για βελτίωση του ρυθμού σύγκλισης της προσομοιωμένης ανόπτησης στα [75], [67], αναπτύχθηκαν αλγόριθμοι δειγματοληψίας για τα μοντέλα Potts έτσι ώστε να γίνονται μεγαλύτερες κινήσεις κατά τη διάρκεια της βελτιστοποίησης.

Εάν το πρόβλημα εκφραστεί με συνεχή μαθηματικά υπάρχει επίσης η δυνατότητα χρήσης λογισμού των μεταβολών. Μια τέτοια προσέγγιση πα-

ρουσιάστηκε στο [9]. Οι τεχνικές λογισμού των μεταβολών χρησιμοποιούν εξισώσεις Euler οι οποίες εγγυώνται την εύρεση τοπικού ελαχίστου. Η εφαρμογή τέτοιων μεθόδων στην επεξεργασία και ανάλυση εικόνων απαιτεί την διακριτοποίηση των εξισώσεων. Μία εναλλακτική λύση αποτελεί η χρήση διακριτών μεθόδων χαλάρωσης τιμών ετικετών, που έχει ακολουθηθεί από πληθώρα ερευνητών, [52], [70]. Κατά την εφαρμογή τέτοιων μαθηματικών μεθόδων, η συνδυαστική (combinatorial) βελτιστοποίηση μετασχηματίζεται σε συνεχή βελτιστοποίηση με γραμμικές αρχικές συνθήκες. Υπάρχουν μερικές εκφράσεις καθόδου κλίσης (gradient descent), που δίνουν ικανοποιητικές λύσεις με δεδομένες τις συγκεκριμένες αρχικές συνθήκες. Οι τεχνικές χαλάρωσης των τιμών ετικετών είναι πιο γενικές από τις μεθόδους ελαχιστοποίησης ενέργειας, [42].

Επίσης υπάρχουν μέθοδοι που εγγυώνται βελτιστοποίηση σε συγκεκριμένες περιπτώσεις. Οι συνεχείς μέθοδοι, όπως η βαθμωτή κυρτότητα [2], περιλαμβάνουν την προσέγγιση μη-κυρτών συναρτήσεων ενέργειας με μία σειρά κυρτών συναρτήσεων ενέργειας. Υπό συνθήκες αυτές οι μέθοδοι μπορούν να καταλήξουν σε βέλτιστη λύση [2]. Οι συνεχείς μέθοδοι μπορούν να εφαρμοστούν σε πληθώρα συναρτήσεων ενέργειας, αλλά γενικά δεν είναι γνωστή η απόδοση τους, εκτός από τις συγκεκριμένες συνθήκες που προαναφέρθηκαν.

Ο δυναμικός προγραμματισμός επίσης, μπορεί να δώσει καθολικό ελάχιστο για μία ομάδα συναρτήσεων ενέργειας [8]. Παρόλο αυτά ο δυναμικός προγραμματισμός περιορίζεται απαραίτητα σε μονοδιάστατα πλαίσια αναφοράς, κάτι που περιλαμβάνει σημαντικές περιπτώσεις όπως για παράδειγμα τα snakes, [40]. Γενικά στην ανάλυση εικόνας και σε προβλήματα πρώιμης υπολογιστικής όρασης συναντώνται διδιάστατες συναρτήσεις ενέργειας σε προβλήματα βελτιστοποίησης. Ο δυναμικός προγραμματισμός λοιπόν δε μας δίνει λύσεις με αποδοτικό τρόπο.

Οι τεχνικές αποκοπής γράφων μέσω της συνδυαστικής βελτιστοποίησης μπορούν να εντοπίσουν ένα καθολικό ελάχιστο για κάποιες πολυδιάστατες συναρτήσεις ενέργειας. Όταν υπάρχουν δύο μόνο ετικέτες τότε η εξίσωση (5.34) αποτελεί μία ειδική περίπτωση του μοντέλου Ising, [72]. Οι Greig, Porteous, και Seheult [18] έχουν προτείνει τον υπολογισμό καθολικού ελαχίστου με χρήση αποκοπής γράφου δύο ετικετών. Αξίζει να παρατηρηθεί σε αυτό το σημείο ότι το μοντέλο Potts αποτελεί μία γενίκευση του μοντέλου Ising για την περίπτωση ύπαρξης παραπάνω από δύο ετικετών. Πρόσφατα παρουσιάστηκε στα [74], [30] και [84], η χρήση αποκοπής γράφων για την εύρεση ολικού ελάχιστου για συγκεκριμένο τύπο συναρτήσεων ενέργειας. Οι μέθοδοι αυτές βρίσκουν εφαρμογή μόνο σε χρήση ετικετών μιας διάστασης. Ακόμη πιο σημαντικό είναι το γεγονός ότι τέτοιες μέθοδοι απαιτούν κυρτές συναρτήσεις *V* και επομένως δε διαθέτουν την ιδιότητα διατήρησης ασυνέχειας. Η αποκοπή γράφων χρησιμοποιήθηκε εκτενώς σε κατάτμηση εικόνων βασισμένη

σε ομαδοποίηση σε περιοχές στην εικόνα (clustering). Όμοια και σε αυτό το είδος ομαδοποίησης αξιοποιείται μια συνάρτηση αποκοπής δεδομένων  $D_p(\cdot)$  η οποία αναθέτει τιμές ετικετών (για παράδειγμα, τιμές φωτεινότητας ή ανομοιότητας βάθους) σε ορισμένες ομάδες εικονοστοιχείων.

Στην παρούσα διατριβή υιοθετήθηκε μία νέα προσέγγιση, [86] για ελαχιστοποίηση πολυδιάστατων συναρτήσεων ενέργειας χρησιμοποιώντας αποκοπή γράφων επαναληπτικά. Γενικεύοντας τα παραπάνω αποτελέσματα χρησιμοποιήθηκαν αυθαίρετα σύνολα ετικετών, ποικίλες συναρτήσεις στερεοσκοπικής συνέπειας δεδομένων  $D_p$  και μία ευρεία κλάση συναρτήσεων αποκοπής V οι οποίες διαθέτουν ιδιότητες διατήρησης ασυνέχειας.

#### 5.5.1 Κατάτμηση και Χώροι Κίνησης

Κάθε ετικετοποίηση f μπορεί να αντιπροσωπευθεί από μία κατάτμηση των εικονοστοιχείων της εικόνας  $\mathbf{P} = \{P_l | l \in \mathcal{L}\}$ όπου  $P_l = \{p \in \mathcal{P} | f_p = l\}$ είναι ένα υποσύνολο εικονοστοιχείων με τιμή ετικέτας l. Από τη στιγμή που υπάρχει ένα προς ένα αντιστοιχία ανάμεσα στα σύνολα ετικετών f και στην κατάτμηση  $\mathbf{P}$ , οι συμβολισμοί στο εξής θα χρησιμοποιηθούν ισοδύναμα.

Έστω ζευγάρι ετικετών με τιμές  $\alpha$ ,  $\beta$ . Μία κίνηση από την κατάτμηση **P** (σύνολο ετικετών f) σε μία νέα κατάτμηση P' (σύνολο ετικετών f') ονομάζεται  $\alpha$ -β αντιμετάθεση, εάν  $\mathcal{P}_l = \mathcal{P}'_l$  για κάθε ετικέτα  $l \neq \alpha, \beta$ . Αυτό σημαίνει ότι η μόνη διαφορά μεταξύ **P** και **P**' είναι ότι μερικά εικονοστοιχεία τα οποία έχουν τιμή ετικέτας  $\alpha$  στο **P** τώρα έχουν τιμή ετικέτας  $\beta$  στο **P**' και το αντίστροφο. Ένα παράδειγμα της κίνησης τύπου α-β αντιμετάθεση φαίνεται στο διάγραμμα (γ) του Σχήματος 5.2. Δοσμένης μίας ετικέτας α, μία κίνηση από την κατάτμηση  $\mathbf{P}$  (σύνολο ετικετών f) σε μία νέα κατάτμηση  $\mathbf{P}'$  (ετικετοποίηση f') ονομάζεται κίνηση τύπου α-διαστοβή εάν  $\mathcal{P}_{\alpha} \subset \mathcal{P}'_{\alpha}$  και  $\mathcal{P}'_{l} \subset \mathcal{P}_{l}$ για κάθε ετικέτα  $l \neq \alpha$ . Με άλλα λόγια, η κίνηση τύπου  $\alpha$ -διαστολή αλλάζει την τιμή της ετικέτας σε κάθε σύνολο εικονοστοιχείων σε  $\alpha$ . Ένα παράδειγμα της κίνησης α-διαστολή φαίνεται στο διάγραμμα (δ) του Σχήματος 5.2. Αξίζει να επαναληφθεί ότι οι μέθοδοι ΙCM και προσομοιωμένη ανόπτηση χρησιμοποιούν τυπικές κινήσεις κατά τις οποίες η τιμή ετικέτας αλλάζει για ένα εικονοστοιχείο τη φορά. Ένα παράδειγμα τυπικής κίνησης φαίνεται στο διάγραμμα (β) του Σχήματος 5.2. Παρατηρούμε ότι η τυπική κίνηση είναι μία ειδική περίπτωση τόσο της  $\alpha$ -διαστολής όσο και της  $\alpha$ - $\beta$  αντιμετάθεσης.

Εξαιτίας των εγγενών δυσκολιών που σχετίζονται με την εύρεση ολικού ελαχίστου, όπως έχουμε ήδη προαναφέρει, στην παρούσα διατριβή ακολουθήσαμε διαδικασίες εύρεσης τοπικών ελαχίστων με χρήση μεγάλου εύρους κινήσεων. Θα αποδειχθεί ότι η χρήση αυτής της προσέγγισης βοήθησε αποτελεσματικά στην αντιμετώπιση πολλών από τα προβλήματα που πηγάζουν από τη χρήση τοπικών ελαχίστων. Οι αλγόριθμοι που θα παρουσιαστούν σε



Σχήμα 5.2: Παραδείγματα τυπικής και ευρείας κίνησης, από ένα αρχικό σύνολο ετικετών (α). Ο αριθμός των ετικετών είναι  $|\mathcal{L}| = 3$ . Μία τυπική κίνηση (β) αλλάζει την τιμή μίας ετικέτας ενός μοναδικού εικονοστοιχείου (περιοχή εντός του κύκλου). Κινήσεις όπως στα (γ) και (δ) επιτρέπουν την ταυτόχρονη αλλαγή των τιμών ετικετών σε μεγάλο αριθμό εικονοστοιχείων [86].

αυτό το κεφάλαιο δημιουργούν ένα σύνολο ετικετοποίησης το οποίο είναι τοπικό ελάχιστο της ενέργειας όπως δόθηκε στην (5.32) για τους δύο βασικούς τύπους ευρείας κίνησης: α-διαστολή και  $\alpha - \beta$ -αντιμετάθεση. Σε αντίθεση με το είδος των τυπικών κινήσεων αυτές δίνουν τη δυνατότητα σε μια μεγάλη ομάδα εικονοστοιχείων να αλλάζουν τις τιμές των ετικετών τους ταυτόχρονα. Το γεγονός αυτό οδηγεί το σύνολο ετικετών, ύστερα από μονή κίνηση, σε τοπική ελαχιστοποίηση και με μέγεθος εκθετικά μεγάλο. Ταυτόχρονα η συνθήκη (5.38) είναι αρκετά απαιτητική.

### 5.5.2 Βελτιστοποίηση με Αποκοπή Γράφων

Με δεδομένο ένα σύνολο ετικετών f, υπάρχει εκθετικός αριθμός κινήσεων αντιμετάθεσης ή διαστολής. Συνεπώς όταν αναζητούμε τοπικό ελάχιστο χρειάζεται εκθετικός χρόνος αναζήτησης, σε αντίθεση με τη χρήση τυπικών κινήσεων για τις οποίες υπάρχει γραμμικός αριθμός από τυπικές κινήσεις για οποιαδήποτε σύνολο ετικετών f.

Επιβάλλεται λοιπόν να δανειστούμε αποδοτικές μεθόδους, βασισμένες σε γράφους [86]. Είναι δυνατή λοιπόν η εύρεση μίας βέλτιστης κίνησης είτε διαστολής, είτε αντιμετάθεσης που να μας οδηγεί ταχύτατα σε τοπικό ελάχιστο. Ο ψευδοκώδικας των χρησιμοποιούμενων αλγορίθμων φαίνονται στο Σχήμα5.3. Οι δύο αλγόριθμοι εμφανίζουν αρκετές ομοιότητες στη δομή τους. Στο εξής την μοναδική εκτέλεση των βημάτων 3.1 και 3.2 θα ονομάσουμε *επαυάβηψη* και την εκτέλεση των βημάτων 2 εώς 4 κύκλο. Σε κάθε κύκλο, ο αλγόριθμος εκτελεί μία επανάληψη για κάθε ετικέτα (αλγόριθμος διαστολής) ή για κάθε ζευγάρι ετικετών (αλγόριθμος αντιμετάθεσης) με συγκεκριμένη ή τυχαία σειρά. Ένας κύκλος θεωρείται επιτυχημένος εάν βρεθεί μία ετικετοποίηση σε

```
    Start with an arbitrary labeling f
    Set success := 0
    For each pair of labels {α, β} ⊂ L
    3.1. Find f̂ = arg min E(f') among f' within one α-β swap of f
    3.2. If E(f̂) < E(f), set f := f̂ and success := 1</li>
    If success = 1 goto 2
    Return f
    Start with an arbitrary labeling f
    Set success := 0
    For each label α ∈ L
    3.1. Find f̂ = arg min E(f') among f' within one α-expansion of f
    2. If E(f̂) < E(f), set f := f̂ and success := 1</li>
```

Σχήμα 5.3: Κορυφή: Ψευδοκώδικας κίνησης τύπου αντιμετάθεσης. Βάση: Ψευδοκώδικας κίνησης τύπου διαστολής.

κάθε επανάληψη που αντιστοιχεί σε μικρότερη τιμή ενέργειας.

4. If success = 1 goto 2

5. Return f

Οι αλγόριθμοι σταματούν ύστερα από τον πρώτο ανεπιτυχή κύκλο, αφού οποιαδήποτε περαιτέρω βελτίωση δεν είναι δυνατή. Κάθε κύκλος στον αλγόριθμο αντιμετάθεσης χρειάζεται  $|\mathcal{L}|^2$  επαναλήψεις, σε αντίθεση με τον αλγόριθμο διαστολής που χρειάζεται  $|\mathcal{L}|$  επαναλήψεις. Οι αλγόριθμοι αυτοί εγγυώνται τον τερματισμό τους σε πεπερασμένο αριθμό κύκλων. Επίσης με δεδομένες τις υποθέσεις ότι οι συναρτήσεις V και  $D_p$  στην εξίσωση (5.32) είναι σταθερές και ανεξάρτητες από το μέγεθος της εικόνας  $\mathcal{P}$ , εύκολα αποδεικνύεται η ολοκλήρωση επιτυγχάνεται σε  $O(|\mathcal{P}|)$  κύκλους [50].

Στα πειράματά μας ο μεγάλος όγκος των βελτιώσεων συντελείται κατά τη διάρκεια του πρώτου κύκλου. Χρησιμοποιήθηκε η μέθοδος αποκοπής γράφων για τον αποδοτικό υπολογισμό του  $\hat{f}$  στο βήμα 3.1. Σε κάθε επανάληψη ο γράφος διαθέτει  $O(|\mathcal{P}|)$  εικονοστοιχεία. Ο ακριβής αριθμός εικονοστοιχείων, η τοπολογία του γράφου καθώς και τα βάρη των πλευρών του ποικίλλουν από επανάληψη σε επανάληψη. Συμπερασματικά οι λεπτομέρειες του γράφου για τον κάθε αλγόριθμο της διαστολής και αντιμετάθεσης, διαφέρουν αρκετά μεταξύ τους.

Το βήμα 3.1 στο Σχήμα 5.3 παίζει ένα ρόλο κλειδί στην υλοποίηση των αλγορίθμων τόσο διαστολής όσο και αντιμετάθεσης. Πριν αναλυθεί με λεπτομέρεια το συγκεκριμένο βήμα θα παρουσιαστεί η τεχνική αποκοπής γράφων.

**Ορισμός 1.** Έστω μη κατευθυνόμενος σταθμισμένος γράφος  $\mathcal{G} = \langle \mathcal{V}, \mathcal{E} \rangle$ , με  $\mathcal{V}$  το σύνολο των κόμβων και  $\mathcal{E}$  το σύνολο των ακμών. Θεωρούμε ένα υποσύνολο

S του V. Μία αποκοπή  $\delta(S)$  ορίζεται ως το υποσύνοβο των ακμών  $(i,j) \in \mathcal{E}$ τέτοιο ώστε  $|\{i,j\} \cap S| = 1$ 

Аυτό σημαίνει ότι η αποκοπή  $\delta(S)$  αποτελείται από όλες εκείνες τις ακμές με ακριβώς ένα κόμβο από το σύνολο S. Με δεδομένο ένα μη κατευθυνόμενο γράφο  $\mathcal{G} = \langle \mathcal{V}, \mathcal{E} \rangle$  με δύο διακριτούς κόμβους που ονομάζονται τερματικοί κόμβοι (τερματικά) και μία μη αρνητική τιμή κόστους (ή χωρητικότητας)  $c_e$ για κάθε ακμή  $e \in \mathcal{E}$ , το κόστος μίας αποκοπής  $\delta(S)$ , είναι το άθροισμα των τιμών κόστους των ακμών που ανήκουν στη αποκοπή:

$$c(\delta(\mathcal{S})) = \sum_{e \in \delta(\mathcal{S})} c_e \tag{5.39}$$

Μία αποκοπή  $\mathcal{C}$  υποσύνολο του  $\mathcal{E}$  ονομάζεται ελάχιστη:

- 1. Όταν οι τερματικοί κόμβοι διαχωρίζονται στον προκύπτοντα γράφο  $\mathcal{G}(\mathcal{C}) = \langle \mathcal{V}, \mathcal{E} \mathcal{C} \rangle$ .
- 2. Όταν δεν υπάρχει υποσύνολο της ελάχιστης αποκοπής C που να διαχωρίζει τους τερματικούς κόμβους στο γράφο  $\mathcal{G}(C)$ .

Το συνολικό κόστος μίας αποκοπής C συμβολίζεται ως |C| και ισούται με το άθροισμα των τιμών κόστους των ακμών του γράφου  $\mathcal{G}(C)$ .

Ισοδύναμα, το πρόβλημα ελάχιστης αποκοπής συνίσταται στην ανεύρεση αποκοπής με το ελάχιστο κόστος. Στην περίπτωση κατά την οποία όλες οι τιμές κόστους έχουν τιμή ίση με 1, τότε το πρόβλημα μετασχηματίζεται σε πρόβλημα ανεύρεσης μίας αποκοπής γράφου με όσο το δυνατόν λιγότερες ακμές.

Η εύρεση ελάχιστης αποκοπής βασίζεται στο θεώρημα μέγιστης ροής τερματικών κόμβων των Ford και Fulkerson. Οι Boykov et al, [83], απέδειξαν την ισοδυναμία των προβλημάτων μεταξύ ελάχιστης αποκοπής και μέγιστης ροής. Η τιμή της μέγιστης ροής συμπίπτει με την τιμή του κόστους μίας ελάχιστης αποκοπής. Με βάση και τις επόμενες παραγράφους προκύπτει ότι το βήμα 3.1 στο Σχήμα 5.3 αντιπροσωπεύει ουσιαστικά την επίλυση ενός προβλήματος ελάχιστης αποκοπής ενός κατάλληλα ορισμένου γράφου δύο τερματικών κόμβων.

## 5.5.3 Αναζήτηση Βέλτιστης Κίνησης $\alpha - \beta$ Αντιμετάθεσης

Με είσοδο ένα σύνολο ετικετών f (κατάτμηση **P**) και ένα ζευγάρι ετικετών με τιμές ( $\alpha$ , $\beta$ ), αναζητείται η κατάλληλη ανάθεση τιμών ετικέτας από το σύνολο  $\widehat{f} = \{\alpha, \beta\}$  για το οποίο ελαχιστοποιείται η συνάρτηση ενέργειας E(f) για τη συγκεκριμένη ανάθεση τιμών, κατά τη διάρκεια μίας  $\alpha$ - $\beta$  αντιμετάθεσης.

Η διαδικασία αυτή αποτελεί κρίσιμο σημασίας στάδιο για τον αλγόριθμο με κινήσεις τύπου αντιμετάθεσης, όπως φαίνεται στο ψευδοκώδικα του Σχήματος 5.3. Η τεχνική που χρησιμοποιήθηκε βασίζεται στον υπολογισμό ενός συνόλου ετικετών που να αντιστοιχεί στην ελάχιστη αποκοπή του γράφου  $\mathcal{G}_{\alpha\beta} = \langle \mathcal{V}_{\alpha\beta}, \mathcal{E}_{\alpha\beta} \rangle$ . Η δομή του γράφου προσδιορίζεται δυναμικά από την τρέχουσα κατάτμηση **P** και από την κατάλληλη ανάθεση τιμών  $\alpha$  ή  $\beta$  στα εικονοστοιχεία. Η δομή του γράφου φαίνεται στο Σχήμα 5.4. Για λόγους απλότητας το εν λόγω σχήμα αναπαριστά την περίπτωση μονοδιάστατης εικόνας.



Σχήμα 5.4: Παράδειγμα γράφου  $\mathcal{G}_{\alpha\beta}$  για 1Δ εικόνα. Το σύνολο των εικονοστοιχείων στην εικόνα είναι  $\mathcal{P}_{\alpha\beta} = \mathcal{P}_{\alpha} \cup P_{\beta}$ , όπου  $\mathcal{P}_{\alpha} = \{p, r, s\}$  και  $\mathcal{P}_{\beta} = \{q, \dots, w\}$ 

Για κάθε εικόνα η δομή του  $\mathcal{G}_{\alpha\beta}$  έχει ως εξής: Το σύνολο των κόμβων περιλαμβάνει τους τερματικούς κόμβους  $\alpha$ ,  $\beta$ , όπως και τα εικονοστοιχεία pστα σύνολα  $\mathcal{P}_{\alpha}$  και  $\mathcal{P}_{\beta}$  (αυτό σημαίνει  $f_p \in {\alpha, \beta}$ ). Επομένως το σύνολο των κόμβων  $\mathcal{V}_{\alpha\beta}$  περιλαμβάνει τα  $\alpha$ ,  $\beta$  και  $\mathcal{P}_{\alpha\beta} = \mathcal{P}_{\alpha} \cup \mathcal{P}_{\beta}$ . Κάθε εικονοστοιχείο  $p \in \mathcal{P}_{\alpha\beta}$  είναι συνδεδεμένο στους τερματικούς κόμβους  $\alpha$  και  $\beta$  μέσω των ακμών  $t_p^{\alpha}$  και  $t_p^{\beta}$  αντίστοιχα. Στο εξής θα αναφερόμαστε στις συγκεκριμένες ακμές με την ονομασία t-σύνδεσμοι (t-links, terminal links). Επίσης κάθε ζευγάρι εικονοστοιχείων  $\{p,q\} \subset \mathcal{P}_{\alpha\beta}$  που βρίσκεται σε γειτνίαση ( $(p,q) \in \mathcal{N}$ ) συνδέεται με την ακμή  $e_{\{p,q\}}$ , η οποία στο εξής θα καλείται n-σύνδεσμος (n-link, neighbor link).

Το σύνολο των ακμών  $\mathcal{E}_{\alpha\beta}$  επομένως αποτελείται από  $\bigcup_{p\in\mathcal{P}_{\alpha\beta}} \{t_p^{\alpha}, t_p^{\beta}\}$  (tσύνδεσμοι) και  $\bigcup_{\substack{p,q\in\mathcal{N}\\p,q\in\mathcal{P}_{\alpha\beta}}} e_{\{p,q\}}$  (n-σύνδεσμοι). Τα βάρη που ανατίθενται στις

ακμές φαίνονται στον Πίνακα 5.1.

Κάθε ελάχιστη αποκοπή C στο  $\mathcal{G}_{\alpha\beta}$  πρέπει να περιλαμβάνει ένα ακριβώς tσύνδεσμο για κάθε εικονοστοιχείο  $p \in \mathcal{P}_{\alpha\beta}$ . Διαφορετικά δεν ικανοποιούνται οι συνθήκες για να είναι ελάχιστη η αποκοπή C. Πιο συγκεκριμένα, εάν

ακμή	βάρος	πεδίο ορισμού
$t_p^{lpha}$	$D_p(\alpha) + \sum V(\alpha, f_q)$	$p \in \mathcal{P}_{\alpha\beta}$
-	$q \in \mathcal{N}_p \ q \nexists \mathcal{P}_{lpha eta}$	
$t_p^{\beta}$	$D_p(\beta) + \sum V(\beta, f_q)$	$p \in \mathcal{P}_{\alpha\beta}$
	$q \in \mathcal{N}_p \ q  otin \mathcal{P}_{lphaeta}$	
R( )	$V(\alpha, \beta)$	$\{p,q\} \in \mathcal{N}$
$c_{\{p,q\}}$	$V(\alpha,\beta)$	$p, q \in \mathcal{P}_{\alpha\beta}$

Πίνακας 5.1: Πίνακας ανάθεσης βαρών στις πλευρές ενός γράφου, [86]

κανένας εκ των t-συνδέσμων δεν υπάρχει στην αποκοπή C, τότε υπάρχει ένα μονοπάτι μεταξύ των τερματικών κόμβων. Επίσης, αν και οι δύο t-σύνδεσμοι ανήκουν στην αποκοπή C, τότε ένα κανονικό υποσύνολο της C θα αποτελεί αποκοπή.

Το γεγονός αυτό ορίζει ένα φυσικό τρόπο κατανομής τιμών ετικετών  $f^C$  που αντιστοιχεί σε μία αποκοπή C στο γράφο  $\mathcal{G}_{\alpha\beta}$ :

$$f_p^{\mathcal{C}} = \left\{ \begin{array}{l} \alpha \, \operatorname{sáv} t_p^{\alpha} \in \mathcal{C} \, \operatorname{yia} \, p \in \mathcal{P}_{\alpha\beta} \\ \beta \, \operatorname{sáv} t_p^{\beta} \in \mathcal{C} \, \operatorname{yia} \, p \in \mathcal{P}_{\alpha\beta} \\ f_p \, \operatorname{yia} \, p \in \mathcal{P}, p \notin \mathcal{P}_{\alpha\beta} \end{array} \right.$$

Εάν ένα εικονοστοιχείο p, ανήκει στο  $\mathcal{P}_{\alpha\beta}$ , τότε σε αυτό ανατίθεται ετικέτα με τιμή  $\alpha$ , όταν η αποκοπή  $\mathcal{C}$  χωρίζει το p από το τερματικό  $\alpha$ . Αντίστοιχα στο p τίθεται η τιμή  $\beta$  όταν η αποκοπή  $\mathcal{C}$  χωρίζει το p από το τερματικό  $\beta$ . Αν το p δεν ανήκει στο  $\mathcal{P}_{\alpha\beta}$ , τότε διατηρείται η αρχική ετικέτα  $f_p$ .

Το γεγονός αυτό οδηγεί στο ακόλουθο λήμμα:

**Λήμμα 5.5.1.** Ένα σύνοβο ετικετών  $f^C$  που αντιστοιχεί σε μία αποκοπή C του  $\mathcal{G}_{\alpha\beta}$  αποτεβεί μία αντιμετάθεση  $\alpha$ -β σε σχέση με την αρχική ετικετοποίηση f.

Μπορεί να αποδειχθεί ότι μία αποκοπή C του  $\mathcal{G}_{\alpha\beta}$  διαθέτει έναν n-σύνδεσμο  $e_{\{p,q\}}$  μεταξύ των γειτονικών εικονοστοιχείων στον γράφο  $\mathcal{G}_{\alpha\beta}$ , εάν και μόνο εάν η αποκοπή C αφήνει συνδεδεμένα τα εικονοστοιχεία p και q σε διαφορετικούς τερματικούς κόμβους. Συνεπώς για κάθε αποκοπή C και για κάθε n-σύνδεσμο  $e_{\{p,q\}}$  ισχύουν οι ακόλουθες ιδιότητες:

Ιδιότητα 5.5.2. (a) Εάν  $t^{lpha}_p, t^{lpha}_q \in \mathcal{C}$  τότε  $e_{\{p,q\}} \notin \mathcal{C}$ 

- (β) Е $\dot{a}v t_p^{\beta}, t_q^{\beta} \in \mathcal{C}$  то́те  $e_{\{p,q\}} \notin \mathcal{C}$
- (y) Eáv  $t_p^lpha, t_q^eta \in \mathcal{C}$  tóte  $e_{\{p,q\}} \in \mathcal{C}$

Οι ιδιότητες ( $\alpha$ ) και ( $\beta$ ) ακολουθούν την απαίτηση ότι κανένα κανονικό υποσύνολο της αποκοπής C δεν μπορεί να χωρίζει τους τερματικούς κόμβους. Η ιδιότητα ( $\gamma$ ) βασίζεται στο γεγονός ότι μία αποκοπή πρέπει να διαχωρίζει τους τερματικούς κόμβους. Η αναπαράσταση των ιδιοτήτων αυτών φαίνονται στο Σχήμα 5.5.



Σχήμα 5.5: Αναπαράσταση ιδιοτήτων αποκοπής (cut) C σε ένα γράφο  $\mathcal{G}_{\alpha}$  για δύο εικονοστοιχεία  $p, q \in \mathcal{N}$  συνδεδεμένα από ένα n-σύνδεσμο  $e_{\{p,q\}}$ . Οι διακεκομμένες γραμμές στα διαγράμματα αναπαριστούν την αποκοπή ακμών C. Οι συνεχόμενες γραμμές αναπαριστούν τις ακμές που απομένουν στο νέο γράφο  $\mathcal{G}_{\mathcal{C}} = \langle \mathcal{V}, \mathcal{E} - \mathcal{C} \rangle$ . Τα πρώτα δύο διαγράμματα από αριστερά αναπαριστούν τις ιδιότητες ( $\alpha$ ) και ( $\beta$ ), αντίστοιχα, ενώ το τρίτο διάγραμμα, αναπαριστά τις ιδιότητες ( $\gamma$ ) και ( $\delta$ ), [86].

## 5.5.4 Βέλτιστη Κίνηση Διαστολής

Με δεδομένο το αρχικό σύνολο ετικετών f (κατάτμηση **P**) και μία τιμή ετικέτας  $\alpha$ , αναζητείται η εύρεση ενός συνόλου ετικετών  $\hat{f}$  το οποίο ελαχιστοποιεί τη συνάρτηση ενέργειας E(f) για κάθε σύνολο ετικετών, κατά τη διάρκεια μίας κίνησης τύπου  $\alpha$ -διαστολής της f. Η αναζήτηση αυτή αποτελεί κρίσιμο θέμα για τον αλγόριθμο διαστολής (ο αλγόριθμος φαίνεται στο κάτω μέρος του Σχήματος 5.3 με τη μορφή ψευδοκώδικα). Στην παρούσα παράγραφο θα περιγραφούν τεχνικές επίλυσης του παραπάνω προβλήματος θεωρώντας ότι κάθε V ικανοποιεί την τριγωνική ανισότητα (μετρικό). Η τεχνική που θα παρουσιαστεί, βασίζεται στον υπολογισμό ενός συνόλου ετικετών που αντιστοιχεί στην ελάχιστη αποκοπή γράφου  $\mathcal{G}_{\alpha} = \langle \mathcal{V}_{\alpha}, \mathcal{E}_{\alpha} \rangle$ . Η δομή του γράφου προσδιορίζεται από την τρέχουσα κατάτμηση **P** και από την ετικέτα  $\alpha$ . Όπως και στην περίπτωση κινήσεων τύπου  $\alpha$ - $\beta$  αντιμετάθεσης, ο γράφος μεταβάλλεται δυναμικά, σε κάθε επανάληψη.

Η δομή του γράφου φαίνεται στο Σχήμα 5.6. Για λόγους απλότητας στην

περιγραφή, το εν λόγω σχήμα αναπαριστά την περίπτωση μονοδιάστατης εικόνας. Το σύνολο των κόμβων περιλαμβάνει τους δύο τερματικούς κόμβους  $\alpha$ και  $\overline{\alpha}$ , όπως επίσης και όλα τα εικονοστοιχεία  $p \in \mathcal{P}$ . Για  $p \notin \mathcal{P}_{\alpha}$  ο τερματικός κόμβος  $\overline{\alpha}$  αντιπροσωπεύει τις ετικέτες που τίθενται στα εικονοστοιχεία από το αρχικό σύνολο ετικετών f.

Για κάθε ζευγάρι γειτονικών εικονοστοιχείων  $\{p,q\} \in \mathcal{N}$  που δεν ανήκει στην κατάτμηση **P**  $(f_p \neq f_q)$ , δημιουργείται ένας επικουρικός κόμβος  $\gamma_{\{p,q\}}$ . Οι επικουρικοί κόμβοι εισάγονται στα όρια των συνόλων κατατμήσεων  $\mathcal{P}_l$  για  $l \in \mathcal{L}$ . Άρα το σύνολο κόμβων είναι:



$$\mathcal{V}_{\alpha} = \{\alpha, \overline{\alpha}, \mathcal{P}, \bigcup_{\substack{\{p, q\} \in \mathcal{N} \\ f_p \neq f_q}} \gamma_{\{p, q\}}\}$$

Σχήμα 5.6: Παράδειγμα γράφου  $\mathcal{G}_{\alpha}$  για 1Δ εικόνα. Το σύνολο των εικονοστοιχείων στην εικόνα είναι  $\mathcal{P} = \{p, q, r, s\}$  και η συγκεκριμένη κατάτμηση  $\mathbf{P} = \{\mathcal{P}_1, \mathcal{P}_2, \mathcal{P}_{\alpha}\}$ , όπου  $\mathcal{P}_1 = \{p\}, \mathcal{P}_2 = \{q, r\}, \mathcal{P}_{\alpha} = \{s\}$ . Οι δύο επικουρικοί κόμβοι  $\alpha = \alpha_{\{p,q\}}$  και  $\beta = \alpha_{\{r,s\}}$ , προστίθενται στα όρια του συνόλου κατάτμησης  $\mathcal{P}_l$ 

Κάθε εικονοστοιχείο  $p \in \mathcal{P}$  είναι συνδεδεμένο στους τερματικούς κόμβους  $\alpha$  και  $\overline{\alpha}$  με τους t-συνδέσμους  $t_p^{\alpha}$  και  $t_p^{\overline{\alpha}}$  αντίστοιχα. Οι δείκτες μπροστά στο συμβολισμό t αντιπροσωπεύουν τους κόμβους του γράφου που ενώνει o t σύνδεσμος. Κάθε ζευγάρι γειτονικών εικονοστοιχείων  $\{q, r\} \in \mathcal{N}$  το οποίο ανήκει σε κοινή κατάτμηση **P** με  $f_q = f_r$ , είναι συνδεδεμένο με έναν nσύνδεσμο  $e_{\{q,r\}}$ , όπως φαίνεται στο Σχήμα 5.6.

Για κάθε ζευγάρι γειτονικών εικονοστοιχείων  $\{p,q\} \in \mathcal{N}$  τέτοιο ώστε  $f_p \neq f_q$ , είναι δυνατόν να δημιουργηθεί μία τριάδα ακμών  $\mathcal{E}_{\{p,q\}} = \{e_{\{p,\gamma\}}, e_{\{\gamma,q\}}, t_{\gamma}^{\overline{\alpha}}\}$ , όπου  $\gamma = \gamma_{\{p,q\}}$  είναι ο αντίστοιχος επικουρικός κόμβος, Σχ5.6. Οι ακμές

ακμή	βάρος	πεδίο ορισμού
$t_p^{\overline{\alpha}}$	$\infty$	$p \in \mathcal{P}_{\alpha}$
$t_p^{\beta}$	$D_p(f_p)$	$p \nexists \mathcal{P}_{lpha}$
$t_p^{\alpha}$	$D_p^{\alpha}$	$p \in \mathcal{P}$
$e_{\{p,\gamma\}}$	$V(f_p, \alpha)$	$\{p,q\} \in \mathcal{N}, f_p \neq f_q$
$e_{\{\gamma,q\}}$	$V(\alpha, f_q)$	$\{p,q\} \in \mathcal{N}, f_p \neq f_q$
$t_{\gamma}^{\overline{\alpha}}$	$V(f_p, f_q)$	$\{p,q\} \in \mathcal{N}, f_p \neq f_q$
$e_{\{p,q\}}$	$V(f_p, \alpha)$	$\{p,q\} \in \mathcal{N}, f_p = f_q$

Πίνακας 5.2: Πίνακας ανάθεσης βαρών στις πλευρές ενός γράφου [86]

 $e_{\{p,\gamma\}}$  και  $e_{\{\gamma,q\}}$  συνδέουν τα εικονοστοιχεία p,q στον επικουρικό κόμβο  $\gamma_{\{p,q\}}$ . Αντίστοιχα, ο t-σύνδεσμος  $t^{\overline{\alpha}}_{\gamma}$  συνδέει τον επικουρικό κόμβο  $\gamma_{\{p,q\}}$  με τον τερματικό κόμβο  $\overline{\alpha}$ . Το σύνολο λοιπόν όλων των ακμών μπορεί να γραφτεί ως εξής:

$$\mathcal{E}_{\alpha} = \{\bigcup_{p \in \mathcal{P}} \{t_p^{\alpha}, t_p^{\overline{\alpha}}\}, \bigcup_{\substack{\{p,q\} \in \mathcal{N} \\ f_p \neq f_q}} \mathcal{E}_{\{p,q\}}, \bigcup_{\substack{\{p,q\} \in \mathcal{N} \\ f_p = f_q}} e_{\{p,q\}}\}$$

Τα βάρη που σταθμίζονται στις πλευρές του γράφου φαίνονται στο πίνακα του Σχήματος 5.2.

Παρόμοια με την προηγούμενη παράγραφο κάθε αποκοπή  $\mathcal{C}$  του γράφου  $\mathcal{G}_{\alpha}$  πρέπει να εξυπηρετεί ακριδώς ένα t-σύνδεσμο για κάθε εικονοστοιχείο  $p \in \mathcal{P}$ . Το γεγονός αυτό ορίζει ένα φυσικό τρόπο κατανομής ετικετών  $f^{\mathcal{C}}$  που αντιστοιχεί σε μία αποκοπή  $\mathcal{C}$  στο γράφο  $\mathcal{G}_{\alpha}$ :

$$f_{p}^{\mathcal{C}} = \begin{cases} \alpha \, \operatorname{\epsilon \acute{a}v} \, t_{p}^{\alpha} \in \mathcal{C} \\ f_{p} \, \operatorname{\epsilon \acute{a}v} \, t_{p}^{\overline{\alpha}} \in \mathcal{C} \end{cases} \, \forall p \in \mathcal{P} \tag{5.40}$$

Τίθεται σε ένα εικονοστοιχείο p η ετικέτα  $\alpha$  εάν η αποκοπή C διαχωρίζει το εικονοστοιχείο p από τον τερματικό κόμβο  $\alpha$ . Αντίθετα, στο p ανατίθεται η παλιά ετικέτα  $f_p$  εάν η αποκοπή C διαχωρίζει το p από το  $\overline{\alpha}$ . Γενικά αποδεικνύεται το παρακάτω λήμμα [86]:

**Λήμμα 5.5.3.** Μια ετικετοποίηση  $f^{C}$  που αντιστοιχεί σε μία αποκοπή C στο γράφο  $\mathcal{G}_{\alpha}$  αποτεβεί μία απόκβιση από την αρχική ετικετοποίηση μετά από μία κίνηση τύπου  $\alpha$ -διαστοβής.

Апобεικνύεται επίσης ότι μία αποκοπή  $\mathcal{C}$  εξυπηρετεί έναν n-σύνδεσμο  $e_{\{p,q\}}$  μεταξύ γειτονικών εικονοστοιχείων  $\{p,q\} \in \mathcal{N}$ , τέτοιων ώστε  $f_p = f_q$  εάν και μονό εάν η αποκοπή  $\mathcal{C}$  αφήνει τα εικονοστοιχεία p και q συνδεδεμένα σε διαφορετικούς τερματικούς κόμβους, με άλλα λόγια η ιδιότητα 5.5.2 ισχύει αρκεί να αντικατασταθεί με  $\overline{\alpha}$  το  $\beta$ . Έτσι η ιδιότητα μπορεί να συμβολίζεται

με 5.5.2( $\overline{\alpha}$ ). Σε αυτή την περίπτωση υπάρχουν πολλοί τρόποι να αποκοπούν οι ακμές, ακόμη και αν το ζευγάρι των εξυπηρετούμενων *t*-συνδέσμων στα εικονοστοιχεία *p* και *q* είναι σταθερό. Η ελάχιστη αποκοπή όμως *C*, στο γράφο  $\mathcal{G}_{\alpha}$  εγγυάται να περιλαμβάνει ακμές από το σύνολο  $\mathcal{E}_{\{p,q\}}$ . Η εύρεση της εξαρτάται από το ποιοι *t*-σύνδεσμοι αποκόπτονται στα εικονοστοιχεία *p* και *q*. Ένας κανόνας για αυτή την περίπτωση περιγράφεται από την ιδιότητα 5.5.4 παρακάτω.

**Ιδιότητα 5.5.4.** Εάν  $\{p,q\} \in \mathcal{N}$  και  $f_p \neq f_q$  τότε η ελάχιστη αποκοπή  $\mathcal{C}$  στο γράφο  $\mathcal{G}_{\alpha}$  ικανοποιεί:

- (a) Eáv  $t_p^lpha, t_q^lpha \in \mathcal{C}$  tóte  $\mathcal{C} \cap \mathcal{E}_{\{p,q\}} = \emptyset$
- (В) Еа́ν  $t_p^{\overline{lpha}}, t_q^{\overline{lpha}} \in \mathcal{C}$  то́те  $\mathcal{C} \cap \mathcal{E}_{\{p,q\}} = t_{\gamma}^{\overline{lpha}}$
- (y) Eáv  $t_p^{\overline{lpha}}, t_q^{lpha} \in \mathcal{C}$  tóte  $\mathcal{C} \cap \mathcal{E}_{\{p,q\}} = e_{\{p,\gamma\}}$
- (б) Еа́ν  $t^{lpha}_p, t^{\overline{lpha}}_q \in \mathcal{C}$  то́те  $\mathcal{C} \cap \mathcal{E}_{\{p,q\}} = e_{\{\gamma,q\}}$

Η ιδιότητα 5.5.4(a) προκύπτει από το γεγονός ότι δεν υπάρχει υποσύνολο του C που να είναι αποκοπή [86]. Οι υπόλοιπες προκύπτουν από την ελαχιστοποίηση του |C| και από το γεγονός ότι τα  $|e_{\{p,\alpha\}}|$ ,  $|e_{\{\alpha,q\}}|$  και  $|t_{\alpha}^{\overline{\alpha}}|$  ικανοποιούν την τριγωνική ανισότητα οπότε η αποκοπή του καθενός ξεχωριστά έχει μικρότερο κόστος σε σχέση με την αποκοπή των άλλων δύο μαζί. Οι ιδιότητες αυτές φαίνονται στο Σχήμα5.7.

# 5.6 Συμπεράσματα Κεφαλαίου

Προηγήθηκε αναλυτική περιγραφή της διαδικασίας στοχαστικής μοντελοποίησης του προβλήματος αντιστοίχισης χαρακτηριστικών σε αλληλουχία εικόνων με Μαρκοβιανά Τυχαία Πεδία. Η διαδικασία αυτή οδήγησε στη μοντελοποίηση αρχικά του προβλήματος εύρεσης αντίστοιχων ζευγαριών εικονοστοιχείων σε αλληλουχία εικόνων μέσω της συνάρτησης ενέργειας στερεοσκοπικής συνέπειας,  $E_{data}$ . Επιπρόσθετα η ανάγκη για τμηματικά ομαλή λύση σε περιοχές της εικόνας μοντελοποιήθηκε με τη συνάρτηση ενέργειας ομαλότητας δεδομένων,  $E_{smooth}$ . Ακολούθησε αναλυτική περιγραφή και διερεύνηση των μεθόδων βελτιστοποίησης των εν λόγω συναρτήσεων ενέργειας ως ένα πρόβλημα επιλογής ελάχιστης αποκοπής ενός σταθμισμένου γράφου. Στα επόμενα κεφάλαια θα επιχειρηθεί η δημιουργία κατάλληλων συναρτήσεων ενέργειας στην εύρεση στερεοσκοπικών αντιστοιχίσεων. Η προσέγγιση μας θα στηριχθεί στην ανάλυση εικόνας στο χώρο της φάσης καθώς και στην



Σχήμα 5.7: Αναπαράσταση ιδιοτήτων ελάχιστης αποκοπής (cut)  $\mathcal{C}$  σε ένα γράφο  $\mathcal{G}_{\alpha\beta}$ για δύο εικονοστοιχεία  $p,q \in \mathcal{N}$  τέτοια ώστε  $f_p \neq f_q$ . Οι διακεκομμένες γραμμές στα διαγράμματα αναπαριστούν την αποκοπή ακμών  $\mathcal{C}$ . Οι συνεχόμενες γραμμές αναπαριστούν τις ακμές που απομένουν στο νέο γράφο  $\mathcal{G}_{\mathcal{C}} = \langle \mathcal{V}, \mathcal{E} - \mathcal{C} \rangle$ . Τα πρώτα δύο διαγράμματα από αριστερά αναπαριστούν τις ιδιότητες ( $\alpha$ ) και ( $\beta$ ), αντίστοιχα. Ενώ το τρίτο διάγραμμα, αναπαριστά τις ιδιότητες ( $\gamma$ ) και ( $\delta$ ), [86].

τοπική αναπαράσταση της πληροφορίας μέσω μονογενών σημάτων. Η επακόλουθη μέθοδος βελτιστοποίησης θα είναι η αποκοπή γράφων με κινήσεις τύπου  $\alpha$ -διαστολής. Η επιλογή αυτή έγινε δεδομένου του πλεονεκτήματος σε ταχύτητα εκτέλεσης κινήσεων αυτού του τύπου, σε σύγκριση με κινήσεις τύπου  $\alpha - \beta$ -αντιμετάθεσης.

# Κεφάλαιο 6

# Βέλτιστη Αντιστοίχιση Εικόνων με Χρήση Πολυδιάστατου Διανύσματος Δομής

# 6.1 Εισαγωγή

Η αναπαράσταση του προβλήματος στερεοσκοπικής αντιστοίχισης με συναρτήσεις ενέργειας καθώς επίσης και η αξιοποίηση της γεωμετρικής πληροφορίας του μονογενούς σήματος αποτελεί το επόμενο βήμα στην εξέλιξη των μεθόδων της παρούσας διατριβής. Επιπλέον στόχο αποτελεί η εφαρμογή της προτεινόμενης μεθοδολογίας αντιστοίχισης στο σύνολο των εικονοστοιχείων της εικόνας και όχι μόνο σε συγκεκριμένα εντοπισμένα χαρακτηριστικά. Παράλληλα επιχειρείται πειραματική αξιολόγηση των προτεινόμενων μεθόδων. Υπολογίζονται χάρτες ανομοιότητας βάθους τόσο για ομάδες πειραματικών εικόνων δανεισμένες από τη βιβλιογραφία, όσο και σε εικόνες εξωτερικού χώρου με αξιοσημείωτη πολυπλοκότητα στη δομή κάτω από τυχαίες συνθήκες φωτισμού.

# 6.2 Γενικεύσεις Αναλυτικού Σήματος

Το αναλυτικό σήμα αποτελεί μία σημαντική αναπαράσταση της επεξεργασίας σήματος, το οποίο χρησιμοποιείται σε πληθώρα εφαρμογών όπως αναγνώριση φωνής, εντοπισμός αντικειμένων μέσω radar κ.α. Επιπλέον η τοπική συχνότητα του αναλυτικού σήματος, μπορεί να θεωρηθεί μέτρο της τοπικής κλίμακας. Η τοπική φάση και το τοπικό πλάτος έχουν χρησιμοποιηθεί σύμφωνα με τη βιβλιογραφία για τον εντοπισμό γεωμετρικών στοιχείων δομής σε μια εικόνα, όπως ακμές καθώς και για την εκτίμηση της τοπικής ενέργειας, αντίστοιχα. Επιπρόσθετα, οι συγκεκριμένες ιδιότητες έχουν χρησιμοποιηθεί και για την στερεοσκοπική επεξεργασία εικόνων στην εκτίμηση βάθους αλλά και στον υπολογισμό οπτικής ροής για αλληλουχίες εικόνων.

Από τη σκοπιά της επεξεργασίας σήματος εικόνας, η θεμελιώδης ιδιότητα του αναλυτικού σήματος είναι ο διαχωρισμός της ταυτότητας [46]. Αυτό σημαίνει ότι στην πολική αναπαράσταση του αναλυτικού σήματος το μέτρο του μιγαδικού σήματος αποτελεί το τοπικό μέτρο του πλάτους, ενώ το όρισμα του σήματος αποτελεί τοπικό μέτρο της φάσης. Η τοπική φάση είναι ανεξάρτητη της τοπικής ενέργειας του σήματος και αλλάζει καθώς η δομή μεταβάλλεται. Το τοπικό πλάτος είναι ανεξάρτητο από τις μεταβολές της δομής αλλά αντιπροσωπεύει την τοπική ενέργεια του σήματος.

Ενέργεια και δομή αποτελούν ανεξάρτητη πληροφορία που βρίσκεται ενσωματωμένη στο σήμα, εκτός αν το σήμα είναι ένας συνδυασμός μερικών σημάτων με διαφορετικές τιμές τοπικής φάσης στις διάφορες κλίμακες. Το ζωνοπερατό φιλτράρισμα αποτελεί τη πιο γνωστή μέθοδο απαλλαγής από τα μερικά σήματα. Τα φίλτρα σε τετραγωνισμό (quadrature filters) συντελούν προς αυτήν την κατεύθυνση παράγοντας ζωνοπερατή έκφραση του τοπικού πλάτους και φάσης. Από τη στιγμή που είναι δυνατόν να διαχωριστεί ένα σήμα, έστω και προσεγγιστικά, στα μερικά του σήματα χρησιμοποιώντας φίλτρα μικρότερου εύρους ζώνης, τότε η πολική αναπαράσταση του αναλυτικού σήματος σε μικρού εύρους ζώνης φίλτρα μπορεί να θεωρηθεί ως τετραγωνικός μετασχηματισμός της πληροφορίας. Στη συνέχεια λοιπόν θα χρησιμοποιηθούν όροι όπως η ενεργειακή και δομική πληροφορία του σήματος. Το avaλυτικό σήμα έχει μελετηθεί σε βάθος για σήματα μίας διάστασης, και από την παραπάνω εισαγωγή προκύπτει ότι σε οποιαδήποτε προσπάθεια για γενίκευση του αναλυτικού σήματος σε δύο διαστάσεις θα πρέπει να διατηρείται η ιδέα της τετραγωνικής ανάλυσης της πληροφορίας. Ένα από τα βασικά προβλήματα που συναντώνται είναι ότι μέτρα για τη μονοδιάστατη εκδοχή ενός σήματος, όπως η τοπική φάση, δεν μπορούν κωδικοποιήσουν τη διδιάστατη δομή γιατί δεν έχουν αρκετούς βαθμούς ελευθερίας.

Παρακάτω θα βασιστούμε στην προσέγγιση των [46] και [48] (βλ. παράγραφο 4.4) στη γενίκευση του αναλυτικού σήματος και στη δημιουργία μονογενούς σήματος. Σύμφωνα με αυτή, διατηρείται η μονοδιάστατη έκφραση της τοπικής φάσης και προστίθεται ένα επιπλέον χαρακτηριστικό που συνιστά τον τοπικό προσανατολισμό της φάσης στην έκφραση του σήματος. Μία τέτοια επιπρόσθετη πληροφορία συνιστά τη γεωμετρική αναπαράσταση του σήματος. Η γεωμετρική αναπαράσταση του μονογενούς σήματος, οι ιδιότητές του καθώς και η συμβολή της εν λόγω προσέγγισης στη στερεοσκοπική αντιστοίχιση εικόνων θα περιγραφούν παρακάτω.

# 6.3 Γεωμετρική Πληροφορία Μονογενούς Σήμα-τος: Θέση Προσανατολισμός και Φάση

Σε αυτήν την παράγραφο θα περιγραφεί η χρήση της πληροφορίας που προκύπτει από την εφαρμογή μονογενών φίλτρων που βασίζονται στην αποδιαμόρφωση της πληροφορίας σε προσανατολισμό, φάση και πλάτος, κατά τη στερεοσκοπική ανάλυση φωτογραφιών, όπως παρουσιάστηκε στην παράγραφο 4.3.

Ως ένα βασικό βήμα επεξεργασίας βασισμένη σε μονογενή φίλτρα, αποτελεί η συστηματική περιγραφή της δομικής και γεωμετρικής πληροφορίας μίας εικόνας σε κλίμακα του γκρι. Όπως έχει προαναφερθεί το μονογενές σήμα επιτελεί διάσπαση ταυτότητας, δηλαδή με άλλα λόγια διαιρεί το σήμα στην ενεργειακή πληροφορία που υποδεικνύει την πιθανότητα παρουσίας δομής, στον προσανατολισμό θ και στη μεταφορά αντίθεσης (εκφραζόμενη από τη φάση φ). Τα χαρακτηριστικά της εικόνας εξάγονται στα μέγιστα της ενέργειας σε τοπικά χωρία της εικόνας όπου η θέση τους παραμετροποιείται από ένα διάνυσμα **x**. Η φάση μπορεί να χρησιμοποιηθεί για να αναπαραστήσει το είδος της μεταφοράς στην αντίθεση σε αυτό το μέγιστο, για παράδειγμα μία τιμή φάσης  $\frac{\pi}{2}$  αντιστοιχεί σε μία σκοτεινή ακμή σε φωτεινό φόντο, ενώ τιμή φάσης 0 αντιστοιχεί σε μία φωτεινή γραμμή σε σκοτεινό φόντο. Ο χώρος των τιμών μεταφοράς αντίθεσης σε ένα μονοδιάστατο σήμα μπορεί να εκφραστεί από το χώρο των φάσεων.

Η διασπορά των φάσεων σε φυσικές εικόνες έχει μελετηθεί στο [26]. Υπάρχουν καθαρές κορυφές για  $\phi = \frac{\pi}{2}$  και  $-\frac{\pi}{2}$  κάτι που υποδεικνύει ότι οι ακμές είναι επικρατέστερο στοιχείο σε σχέση με τις γραμμικές δομές. Προκειμένου να συμπεριληφθεί στην ανάλυση και η πληροφορία του χρώματος υπολογίζεται ένας μέσος όρος στο RGB χώρο χρώματος, σε τμήματα ακμών στη δεξιά και αριστερή εικόνα. Με αυτόν τον τρόπο προκύπτουν δύο διανύσματα  $\mathbf{c}_r = (c_r^r, c_g^r, c_b^r)$  και  $\mathbf{c}_l = (c_r^l, c_g^l, c_b^l)$  που αντιπροσωπεύουν τις κόκκινες, πράσινες και μπλε τιμές μίας ακμής στην αριστερή και δεξιά εικόνα του στέρεο ζευγαριού. Οι τιμές μετασχηματίζονται με τέτοιο τρόπο ώστε  $c_R + c_G + c_B = 1$ .

Μπορεί να θεωρηθεί λοιπόν το πολυδιάστατο διάνυσμα δομής **e** που έχει τη μορφή:

$$\mathbf{e} = (\theta, \phi, \mathbf{c}) \tag{6.1}$$

Εξάγεται η πληροφορία προσανατολισμού  $\theta$ , του μονογενούς σήματος, η οποία παίρνει τιμές στο διάστημα  $[0, \pi)$ . Όταν προστίθεται η δομική πληροφορία στην εικόνα σε μία τοπική ακμή, υπάρχει η δυνατότητα επέκτασης του προσανατολισμού στην ιδέα της κατεύθυνσης που παραμετροποιείται ως  $d \in [0, 2\pi]$ .



Σχήμα 6.1: Σχηματική αναπαράσταση του βασικού διανύσματος χαρακτηριστικών σε χωρίο εικόνας γύρω από μία ακμή [48].

Στο Σχήμα 6.1 φαίνεται η γραφική αναπαράσταση του προσανατολισμού  $\theta$ (ή κατεύθυνσης d) της φάσης  $\phi$  (ή p εάν η έννοια της κατεύθυνσης d προστεθεί)

Στην περίπτωση στερεοσκοπικής επεξεργασίας, η συσχέτιση κατεύθυνσης σε δύο διδιάστατα χαρακτηριστικά υποδηλώνει την ύπαρξη μίας τρισδιάστατης κατεύθυνσης και αντίστροφα μία τριδιάστατη κατεύθυνση υποδηλώνει διδιάστατες κατευθύνσεις της προβολής της. Η ιδιότητα αυτή καλείται συνθήκη μοναδικότητας.

Η σχέση (6.1) μπορεί να επεκταθεί στο διάνυσμα  $\mathbf{e} = (d, p, \mathbf{c})$ , όπου  $d \in [0, 2\pi]$  αντιπροσωπεύει την κατεύθυνση του σήματος, και  $p \in [-\pi, \pi)$  αντιπροσωπεύει τη φάση όπως αυτή σχετίζεται με την τοπικά μετρούμενη φάση  $\phi$  ( $p = \phi$  εάν  $d = \theta$  και  $p = -\phi$  εάν  $d = \theta + \pi$ ).

Η ιδέα της κατεύθυνσης παίζει σημαντικό ρόλο στη δομική πληροφορία της εικόνας, δεδομένου το δομικό κομμάτι της εξίσωσης αλλάζει σε μία περιστροφή  $\pi$ , όπως επίσης και τα διανύσματα ( $\mathbf{e}_l, \mathbf{e}_r$ ). Η έννοια της κατεύθυνσης παίζει σημαντικό ρόλο, και στον ορισμό της συνάρτησης στερεοσκοπικής ομοιότητας. Για παράδειγμα τα χωρία της εικόνας στο Σχήμα 6.2(a) εμπεριέχουν κοινή πληροφορία χρώματος ή φωτεινότητας αλλά αντίθετη κατεύθυνση, κάτι που πιθανόν σημαίνει ότι συγκρίνονται διαφορετικά μέρη της υπό εξέταση ακμής. Αντίθετα, τα χωρία του Σχήματος 6.2(b)(ii) έχουν την ίδια κατεύθυνση αλλά διαφορετική πληροφορία χρώματος όπως φαίνεται από τη



Σχήμα 6.2: a) Δοσμένης της φάσης φ και με δεδομένο το εξαγόμενο προσανατολισμό θ, το ίδιο χωρίο της εικόνας μπορεί να εκφραστεί με όρους κατεύθυνσης d και φάσης p ( $d = \theta + \pi, p = -\phi$ ) (αριστερά) ή ( $d = \theta, p = \phi$ ) (δεξιά), b) Η ομοιότητα στη δομή και τον προσανατολισμό στα δύο χωρία των εικόνων (στέρεο διάταξη) μεταβάλλεται με την εισαγωγή της έννοιας της κατεύθυνσης. c) Θεωρητικά πιθανές αναπαραστάσεις της κατεύθυνσης στο αριστερό και δεξιό χωρίο της εικόνας. Είναι αξιοπρόσεκτο ότι οι περιπτώσεις *ii*) και *iii*) δε συμφωνούν με τους γεωμετρικούς περιορισμούς της στέρεο διάταξης [48].

σκίαση του κυκλικού χωρίου της εικόνας.

Στις εικόνες ο προσανατολισμός είναι τοπικά μετρήσιμος, ενώ το δομικό μέρος μεταβάλλεται για κάθε αλλαγή της κατεύθυνσης. Προκειμένου να συγκριθούν δύο διανύσματα χαρακτηριστικών  $(d^l, p^l, \mathbf{c}^l)$  και  $(d^r, p^r, \mathbf{c}^r)$  στη συνάρτηση στέρεο ομοιότητας πρέπει να ερευνηθούν όλες οι δυνατές τιμές κατεύθυνσης στο αριστερό και δεξιό χωρίο της εικόνας. Συναντώνται λοιπόν οι εξής περιπτώσεις:

- 1. Ο μετρήσιμος προσανατολισμός και στις δύο εικόνες ισούται με την κατεύθυνση:  $d^l = \theta^l, d^r = \theta^r$ , Σχήμα 6.2(c)(i).
- 2. Ο προσανατολισμός που μετράται στην αριστερή εικόνα ισούται με την τιμή κατεύθυνσης, ενώ στη δεξιά εικόνα του στέρεο συστήματος, προσανατολισμός και κατεύθυνση σχετίζονται με τη σχέση  $d^r = \theta^r + \pi$ . Το γεγονός αυτό υποδηλώνει ότι η φάση έχει αντίθετο πρόσημο με την τοπικά μετρούμενη φάση στη δεξιά εικόνα ( $p^r = -\phi^r$ ) και εναλλαγή των διανυσμάτων χρώματος, Σχήμα 6.2(c)(ii).

- 3. Ο μετρούμενος προσανατολισμός στην αριστερή εικόνα συνδέεται με την κατεύθυνση ( $d^l = \theta^l + \pi$ ). Η παραπάνω σχέση επίσης υποδηλώνει ότι η φάση έχει αντίθετο πρόσημο με την τοπικά μετρούμενη φάση ( $p^l = -\phi^l$ ), όπως επίσης και τα διανύσματα χρώματος εναλλάσσονται Στη δεξιά εικόνα του στέρεο συστήματος, κατεύθυνση και προσανατολισμός συμπίπτουν, Σχήμα 6.2(c)(iii).
- 4. Ο μετρούμενος προσανατολισμός στην αριστερή εικόνα συνδέεται με τη διεύθυνση ως εξής:  $d^l = \theta^l + \pi$ . Αντίστοιχα για τη δεξιά εικόνα ισχύει η σχέση:  $d^r = \theta^r + \pi$ . Οι τιμές των φάσεων επίσης αντιστρέφονται, Σχήμα 6.2(c)(iv).

# 6.4 Μέθοδοι Βελτιστοποίησης σε Στέρεο Αντιστοίχιση

Στη συνέχεια παρουσιάζεται η εφαρμογή των μεθόδων βελτιστοποίησης συνόλου ετικετών, σε προβλήματα τριδιάστατης αποκατάστασης, μέσω στέρεο διάταξης εικόνων. Χρησιμοποιήθηκαν τόσο τυπικές εικόνες αξιολόγησης αλγορίθμων υπολογιστικής όρασης, όσο και πειραματικές φυσικές εικόνες εξωτερικού περιβάλλοντος που συναντήσαμε και στο κεφάλαιο 4. Οι εν λόγω εικόνες απεικονίζουν σύνθετες δομές κάτω από τυχαίο εξωτερικό φωτισμό με πολλά στοιχεία υφής. Οι κλασσικές τεχνικές ανάκτησης δομής, όπως έχει φανεί σε προηγούμενα κεφάλαια (κεφάλαιο 4), δεν παρέχουν καθόλου ικανοποιητικά αποτελέσματα. Συνεπώς, στην προσπάθεια επέκτασης των υπολογιστικών μεθόδων στο χώρο φάσης, σε όλη την επιφάνεια της εικόνας και σε όλο το πλήθος των εικονοστοιχείων, εφαρμόστηκαν Μπεϋζιανά μοντέλα απόφασης και βελτιστοποίησης (MAP-MRF) ετικετών και βέλτιστης αναπαράστασης σε γράφους.

Πληθώρα προβλημάτων στο πεδίο της πρώιμης υπολογιστικής όρασης περιλαμβάνουν την ανάθεση ετικέτας σε κάθε εικονοστοιχείο. Οι ετικέτες αντιπροσωπεύουν κάποιο τοπικό χαρακτηριστικό, όπως για παράδειγμα την ανομοιότητα βάθους (disparity). Αυτό το είδος προβλημάτων ετικετοποίησης αντιμετωπίζεται υπό το πρίσμα ελαχιστοποίησης κάποιας συνάρτησης ενέργειας.

#### 6.4.1 Ενεργειακό Μοντέλο

Ορίζεται ένα πρόβλημα ανάθεσης ετικετών και σε κάθε εικονοστοιχείο στη θέση του διδιάστατου σήματος εικόνας **x** τίθεται μία ετικέτα με τιμή ανομοιότητας βάθους, η οποία στο εξής συμβολίζεται ως  $f_x$ . Η συλλογή όλων των

αναθέσεων συμβολίζεται με f, ο αριθμός εικονοστοιχείων με n, ο αριθμός των ετικετών με m. Η συνάρτηση της ενέργειας αποτελείται από δύο όρους. Ο ένας όρος αποκόβει τις μη συμβατές λύσεις με τα παρατηρούμενα δεδομένα, και ο άλλος όρος ενισχύει την τμηματική ομαλότητα των λύσεων. Η συνάρτηση ενέργειας E, μπορεί να θεωρηθεί επίσης και μία εκ των υστέρων πιθανοτική κατανομή ενός Μαρκοβιανού Τυχαίου Πεδίου και αποτελείται από την ενέργεια των δεδομένων  $E_{data}$ , από την ενέργεια ομαλοποίησης  $E_{smooth}$ , καθώς επίσης και από την παράμετρο σχετικής τους επιρροής στη τελική λύση  $\lambda$ .

$$E = E_{data} + \lambda E_{smooth} \tag{6.2}$$

Η ενέργεια δεδομένων μπορεί να ορισθεί ως το συνολικό κόστος ταιριάσματος κάθε εικονοστοιχείου,  $D_{\mathbf{x}}(f)$ 

$$E_{data} = \sum_{\mathbf{x}} D_{\mathbf{x}}(f) \tag{6.3}$$

Κατά τη διαδικασία εξαγωγής από την εικόνα γεωμετρικών και δομικών χαρακτηριστικών χρησιμοποιήθηκε η αναπαράσταση του διδιάστατου σήματος της εικόνας μέσω του διδιάστατου αναλυτικού σήματος όπως αυτή έχει περιγραφεί στην παράγραφο 6.3. Η επακόλουθη δημιουργία μονογενούς σήματος βασισμένου στην υπό εξέταση εικόνα, οδηγεί στην εξαγωγή της τοπικής πληροφορίας για τη φάση, τον προσανατολισμό και το πλάτος του σήματος. Με αυτόν τον τρόπο λοιπόν μας δίνεται η δυνατότητα αξιοποίησης δομικών και γεωμετρικών χαρακτηριστικών της εικόνας στην κατασκευή της συνάρτησης ενέργειας μέσω της χρήσης νέας προτεινόμενης συνάρτησης απόστασης (ομοιότητας) χαρακτηριστικών βασισμένης στο [48]. Ακολουθώντας την ορολογία όπως παρουσιάστηκε στην παράγραφο 6.3, υπολογίστηκαν οι τοπικές τιμές φάσης, και κατεύθυνσης, p, d αντίστοιχα. Δόθηκαν ίσες τιμές βάρους στην εξαγόμενη γεωμετρική, δομική και χρωματική πληροφορία. Διαμορφώνεται λοιπόν ένα νέο πολυδιάστατο διάνυσμα δομής για κάθε εικόνα στη θέση **x** του διδιάστατου σήματος εικόνας:

$$\mathbf{e}(\mathbf{x}) = [d, p, \mathbf{c}, A] (\mathbf{x}) \tag{6.4}$$

Οι ποσότητες **c**, A αντιπροσωπεύουν αφενός τη χρωματική πληροφορία μέσω του διανύσματος **c** που στην περίπτωση που εξετάζουμε είναι μονοδιάστατο γιατί οι εικόνες μελετώνται σε κλίμακα του γκρι. Αφετέρου η μεταβλητή A αντιπροσωπεύει το πλάτος του μονογενούς σήματος  $f_M(\mathbf{x})$  σε μία θέση **x** του διδιάστατου σήματος εικόνας, όπως φαίνεται στις εξισώσεις (4.31) και (4.41). Στις εικόνες της στέρεο διάταξης υπολογίζονται τα πολυδιάστατα διανύσματα  $\mathbf{e}^{l}(\mathbf{x})$  και  $\mathbf{e}^{r}(\mathbf{x})$  για την αριστερή και δεξιά εικόνα της διάταξης αντίστοιχα. Η αναζήτηση αντίστοιχων σημείων πραγματοποιείται στο πλαίσιο μίας στέρεο διάταξης. Αυτό σημαίνει σύμφωνα με την επιπολική γεωμετρία ότι ένα σημείο της αριστερής εικόνας αντιστοιχεί σε ένα πλήθος υποψήφιων προς αντιστοίχηση εικόνων στην δεξιά εικόνα της διάταξης. Το τετράγωνο της απόστασης των αντίστοιχων πολυδιάστατων διανυσμάτων αντιπροσωπεύει την ενέργεια κόστους αντιστοίχισης. Η ενέργεια κόστους αντιστοίχισης ή στερεοσκοπικής συνέπειας γίνεται:

$$D_{\mathbf{x}}(f) = \|\mathbf{e}^{\mathsf{I}}(\mathbf{x}) - \mathbf{e}^{\mathsf{r}}(\mathbf{x} + \Delta_{\mathbf{x}}(f))\|^2$$
(6.5)

Όπου f αποτελεί το σύνολο της ετικετοποίησης με τιμές ανομοιότητας βάθους. Η συνάρτηση  $\Delta_{\mathbf{x}}(f)$ ) αντιπροσωπεύει τη διαδικασία της στερεοσκοπικής αναζήτησης του βέλτιστου αντίστοιχου σημείου στην επιπολική γραμμή της δεξιάς εικόνας της διάταξης. Στη συνέχεια αντικαθιστώντας τα διανύσματα της Σχέσης (6.5) σύμφωνα με τη Σχέση (6.4), η τελική μορφή της συνάρτησης ενέργειας στερεοσκοπικής συνέπειας γίνεται:

$$D_{\mathbf{x}}(f) = (\triangle d)^2 + (\triangle p)^2 + (\triangle c)^2 + (\triangle A)^2$$
(6.6)

όπου  $\Delta d$  αντιπροσωπεύει την μεταβολή της κατεύθυνσης, και  $\Delta p$  τη διαφορά φάσης, όπως αυτά έχουν οριστεί στο [48]. Η ποσότητα  $\Delta c$  αντιπροσωπεύει την απόσταση τιμών φωτεινότητας στα συγκρινόμενα εικονοστοιχεία. Τέλος η ποσότητα  $\Delta A$  αντιπροσωπεύει τη μεταβολή του τοπικού πλάτους του μονογενούς σήματος όπως αυτό προκύπτει από τη διδιάστατη αναπαράσταση του αναλυτικού σήματος. Όλες οι τιμές  $\Delta d$ ,  $\Delta p$ ,  $\Delta c$  και  $\Delta A$  κανονικοποιούνται κατάλληλα ώστε να έχουν συγκρίσιμη μέση τιμή και συγκρίσιμη τυπική απόκλιση. Συνεπώς η αρχική έκφραση της συνάρτησης ενέργειας δεδομένων όπως αυτή εκφράζεται στη σχέση (6.3) βασισμένη αποκλειστικά στις τιμές φωτεινότητας των υπό εξέταση εικονοστοιχείων εμπλουτίζεται με δομική και γεωμετρική τοπική πληροφορία της εικόνας. Η συγκεκριμένη πληροφορία ανακτάται με τη χρήση μονογενούς σήματος, στην κωδικοποίηση των εικόνων του στέρεο ζεύγους.

Η εύρεση της ελάχιστης απόστασης συμπίπτει με την εύρεση ενός βέλτιστου ταιριάσματος εικονοστοιχείων. Η επίλυση του προβλήματος βασίζεται στη χρήση του αλγορίθμου ελάχιστης αποκοπής γράφων με χρήση ευρέων κινήσεων, είτε διαστολής, είτε α-β αντιμετάθεσης όπως έχουν παρουσιαστεί αναλυτικά σε προηγούμενο κεφάλαιο.

Σε ένα Μαρκοδιανό Τυχαίο Πεδίο η ενέργεια δεδομένων προέρχεται από κατανομές πιθανότητας υπό συνθήκη (όπως αναλύθηκε στο κεφάλαιο 5). Θεωρούμε ότι τα εικονοστοιχεία συνθέτουν ένα διδιάστατο πλέγμα, έτσι ώστε καθένα από αυτά να μπορεί να εκφραστεί στις συντεταγμένες του p = (i, j). Έχει χρησιμοποιηθεί ένα τυπικό σύστημα 4 συνδεδεμένων γειτόνων, έτσι ώστε η ενέργεια ομαλοποίησης  $V_{pq}(f)$  να συντίθεται από το άθροισμα του κόστους γειτονικών ζευγών ομαλοποίησης σε κάθετη και οριζόντια διεύθυνση, όπου εάν p = (i, j) και q = (s, t), τότε |i - s| + |j - t| = 1. Έστω  $\mathcal{N}$  η σχέση γειτνίασης ζευγών εικονοστοιχείων. Η ενέργεια ομαλοποίησης είναι:

$$E_{smooth} = \sum_{(p,q)\in\mathcal{N}} V_{pq}(f)$$
(6.7)

Το ζεύγος εικονοστοιχείων (p,q) στην εξίσωση (6.7) υποδεικνύει αταξινόμητο σύνολο στοιχείων με τιμές ετικέτας  $f_p$  και  $f_q$  αντίστοιχα, οπότε το άθροισμα συντελείται πάνω σε αταξινόμητα ζεύγη γειτονικών εικονοστοιχείων. Στην παρούσα διατριβή θεωρούμε μία γενική μορφή κόστους ομαλοποίησης, κατά την οποία διαφορετικά ζεύγη γειτονικών ετικετών οδηγούν σε διαφορετικές τιμές κόστους. Μία πιο αυστηρή μαθηματική έκφραση της ενέργειας ομαλοποίησης είναι η ακόλουθη:

$$E_{smooth} = \sum_{(p,q)\in\mathcal{N}} w_{pq} V(f_p - f_q)$$
(6.8)

Οι όροι ομαλοποίησης προκύπτουν από το γινόμενο των μεταβαλλόμενων στο χώρο, και από πριν ορισμένων τιμών στάθμης των εξεταζόμενων ζευγών,  $w_{pq}$ , με την αύξουσα συνάρτηση απόλυτης διαφοράς των τιμών ετικετών. Τέτοιες οικογένειες συναρτήσεων συμβάλουν στην προσέγγιση προβλημάτων αποθορυβοποίησης και στερεοσκοπικής αντιστοίχισης. Η συνάρτηση V αναπαριστάται με μία απλή μορφή, αποφεύγοντας την παραμετροποίηση της V:

$$V(\Delta f) = min(|\Delta f|^k, V_{max})$$
(6.9)

 $\mu \epsilon \ k \in \{1, 2\}.$ 

Εάν τεθεί  $V_{max} = 1.0$ , προκύπτει το μοντέλο Potts,  $V(\Delta f) = 1 - \delta(\Delta f)$ σύμφωνα με το οποίο αποκόπτονται όλα τα ζεύγη με διαφορετικές ετικέτες ( $\delta(\cdot)$  είναι η μοναδιαία κρουστική συνάρτηση). Αν και δεν αποτελεί από τους κύριους στόχους της παρούσας διατριβής, υπάρχει ένας αριθμός ειδικών περιπτώσεων, οι οποίες έχουν ταχείς ακριβείς αλγορίθμους. Εάν υπάρχουν μόνο δύο ετικέτες, το φυσικό μοντέλο Potts (στην συγκεκριμένη περίπτωση ονομάζεται Ising μοντέλο, [72] ) για τα κόστη ομαλοποίησης, επιλύεται ακριβώς με τη χρήση αποκοπής γράφων [69]. Εάν οι ετικέτες έχουν γραμμική διάταξη (για παράδειγμα διαδοχικοί ακέραιοι), και η συνάρτηση του κόστους ομαλοποίησης είναι τυχαία κυρτή συνάρτηση, τότε σύμφωνα με το [29] ο αλγόριθμος αποκοπής γράφων οδηγεί σε ακριβή λύση. Η αναπαράσταση της πληροφορίας εικόνας σε γράφους συνίσταται στην ασεικόνιση των εικονοστοιχείων ως κόμβων του γράφου και των τιμών κόστους που προκύπτουν από τις ενέργειες στερεοσκοπικής συνέχειας και τμηματικής ομαλότητας των τιμών ανομοιότητας βάθους στην εικόνα, ως ακμών του. Η αποκοπή τμήματος γράφου με το ελάχιστο κατά περίπτωση κόστος, αντιστοιχεί σε ομάδα αξιόπιστων ταιριασμάτων των εικονοστοιχείων ή περιοχών της εικόνας.

#### 6.4.2 Αποκοπή Γράφων

Όπως έχει ήδη προαναφερθεί οι δύο πιο γνωστές μεθοδολογίες αποκοπής γράφων βασίζονται σε κινήσεις τύπου διαστολής και αντιμετάθεσης. Η λειτουργία των αλγορίθμων που ανήκουν και στις δύο κατηγορίες βασίζεται στον επαναλαμβανόμενο υπολογισμό του ολικού ελάχιστου ενός δυαδικού προβλήματος κατανομής ετικετών στους βρόχους του σχηματιζόμενου γράφου. Η διαδικασία αυτή συγκλίνει πολύ γρήγορα και οδηγεί σε σθεναρό τοπικό ελάχιστο.

Για ένα ζευγάρι ετικετών  $\alpha$  και  $\beta$ , μία κίνηση τύπου αντιμετάθεσης παίρνει ένα υποσύνολο σημείων που τη δεδομένη στιγμή έχουν ετικέτα τιμής  $\alpha$  και αναθέτει σε αυτά την ετικέτα  $\beta$  ή αντίστροφα. Ο αλγόριθμος κίνησης αντιμετάθεσης βρίσκει ένα τοπικό ελάχιστο τέτοιο ώστε να μην υπάρχει κίνηση αντιμετάθεσης για κάθε ζεύγος ετικετών  $\alpha$  και  $\beta$  που να οδηγεί σε σύνολο ετικετών χαμηλότερης ενέργειας.

Με ανάλογο τρόπο ορίστηκε σε προηγούμενο κεφάλαιο η κίνηση τύπου διαστολής κατά την οποία αυξάνεται ο αριθμός του συνόλου εικονοστοιχείων με μία δεδομένη ετικέτα  $\alpha$  ή  $\beta$ . Ο αλγόριθμος ελαχιστοποίησης με κινήσεις διαστολής βασίζεται στην ανεύρεση τοπικού ελάχιστου, τέτοιου ώστε να μην υπάρχει κίνηση διαστολής για κάθε ετικέτα  $\alpha$  και  $\beta$ , που να οδηγεί σε χαμηλότερη ενέργεια.

Τα κριτήρια για την εύρεση τοπικού ελάχιστου κάτω από την επίδραση κινήσεων διαστολής ή αντιμετάθεσης είναι τόσο αυστηρά ώστε υπάρχουν πολύ λιγότερα ελάχιστα σε χώρους υψηλής διάστασης σε σύγκριση με τις τυπικές κινήσεις (κατά τις οποίες σε κάθε κύκλο αλλάζει η τιμή μίας μόνο ετικέτας). Ακολουθώντας την ορολογία από το [58] οι αλγόριθμοι κινήσεων αντιμετάθεσης ή διαστολής χρησιμοποιούν ευρείας γειτονιάς τεχνικές αναζήτησης. Στην πρωτότυπη εργασία των Boykov et al. [83] έχει δειχθεί ότι οι κινήσεις τύπου διαστολής εφαρμόζονται σε κάθε ενέργεια όπου η  $V_{pq}$  είναι μετρική, ενώ οι κινήσεις τύπου αντιμετάθεσης σε κάθε ενέργεια όπου η  $V_{pq}$  είναι ημιμετρική (που σημαίνει ότι υπάρχει περίπτωση να μην ικανοποιείται η τριγωνική ανισότητα για τη συνάρτηση ενέργειας). Σύμφωνα με την εργασία που παρουσιάστηκε στο [78] οι συνθήκες αυτές μπορούν να ελαστικοποιηθούν. Αποδεικνύεται ότι ο αλγόριθμός κίνησης διαστολής μπορεί να χρησιμοποιηθεί για όλες τις ετικέτες  $\alpha, \beta$  και  $\gamma$  που ισχύει:

$$V_{pq}(\alpha, \alpha) + V_{pq}(\beta, \gamma) \le V_{pq}(\alpha, \gamma) + V_{pq}(\beta, \alpha)$$
(6.10)

Από την άλλη ο αλγόριθμος με κίνηση αντιμετάθεσης μπορεί να χρησιμοποιηθεί για όλες τις ετικέτες  $\alpha, \beta$  που ισχύει:

$$V_{pq}(\alpha, \alpha) + V_{pq}(\beta, \beta) \le V_{pq}(\alpha, \beta) + V_{pq}(\beta, \alpha)$$
(6.11)

Εάν κάποιοι όροι της συνάρτησης ενέργειας δεν ακολουθούν αυτούς τους περιορισμούς η μέθοδος αποκοπής γράφων μπορεί ακόμα να εφαρμοστεί αποκόπτοντας τους όρους αυτούς [69]. Βέβαια σε αυτήν την περίπτωση δεν υπάρχει εγγύηση για την εύρεση βέλτιστης ετικετοποίησης κατά τις κινήσεις διαστολής ή αντιμετάθεσης. Στις συναρτήσεις ενέργειας που χρησιμοποιούνται στα πειράματα που ακολουθούν μόνο στην περίπτωση κίνησης διαστολής χρειάζεται η αποκοπή τέτοιων όρων. Η υλοποίηση του αλγορίθμου αποκοπής γράφων και η εύρεση της ελάχιστης αποκοπής βασίστηκε στις βιβλιοθήκες λογισμικού από το [69].

# 6.5 Στέρεο Αντιστοίχιση - Πειραματική Αξιολόγηση

### 6.5.1 Εισαγωγή

Η πειραματική αξιολόγηση των αποτελεσμάτων βασίστηκε στην συγκριτική μελέτη μεθόδων υπολογισμού χαρτών ανομοιότητας βάθους αρχικά σε ομάδες πειραματικών εικόνων για την αξιολόγηση της επίδοσης των χρησιμοποιούμενων αλγορίθμων και στη συνέχεια σε σύνθετες εικόνες εξωτερικού χώρου απεικόνισης αρχαιολογικών μνημείων. Εξετάζεται συγκριτικά η επίδοση δύο διαφορετικών συναρτήσεων ενέργειας κόστους αντιστοίχισης εικονοστοιχείων. Η κλασσική προσέγγιση συνίσταται σε χρήση συναρτήσεων ενέργειας ταιριάσματος που βασίζονται αποκλειστικά σε μετρήσεις φωτεινότητας στην εικόνα. Η νέα προτεινόμενη συνάρτηση ενέργειας αξιοποιεί τη γεωμετρική πληροφορία από το μονογενές σήμα στη διαδικασία αντιστοίχισης.

Η μέτρηση της επίδοσης των διαφορετικών προσεγγίσεων στον υπολογισμό των συναρτήσεων ενέργειας θα βασιστεί σε μετρήσεις σφάλματος απόκλισης των χαρτών ανομοιότητας βάθους από τους αντίστοιχους ιδανικούς για συγκεκριμένες πειραματικές εικόνες διαθέσιμες στη βιβλιογραφία. Επιπρόσθετα η επίδοση της προτεινόμενης προσέγγισης αξιολογείται με βάση τις τιμές που παίρνουν δείκτες δομικής ομοιότητας (Structural SIMilarity (SSIM) index) στις εικόνες χαρτών ανομοιότητας βάθους όπως αυτοί υπολογίζονται με τις εξεταζόμενες μεθοδολογίες.

# 6.5.2 Πειραματικά Δεδομένα - Αποτελέσματα - Αξιολόγηση

Η προτεινόμενη διαδικασία του στέρεο ταιριάσματος βασισμένη σε ιδέες που παρουσιάστηκαν στα [86] και [85] εφαρμόστηκε αρχικά σε πειραματικές εικόνες (http://vision.middlebury.edu/stereo/data/). Οι εικόνες που χρησιμοποιήθηκαν στην παρούσα διατριβή φαίνονται στο Σχήμα 6.3. Οι ετικέτες παίρνουν τιμές ανομοιότητας βάθους, ενώ το κόστος των δεδομένων προκύπτει από τις απόλυτες τετραγωνικές διαφορές τιμών συγκεκριμένων χαρακτηριστικών των υπό εξέταση εικονοστοιχείων για τον υπολογισμό της ανομοιότητας βάθους. Επίσης έχουν χρησιμοποιηθεί διαφορετικές συναρτήσεις ενέργειας ομαλοποίησης για διαφορετικά ζευγάρια εικόνων, όπως επίσης και διαφορετικό βάρος  $\lambda$  της ενέργειας ομαλοποίησης στη συνάρτηση συνολικής ενέργειας όπως περιγράφεται από τη Σχέση (6.2). Για την εικόνα Tsukuba με m = 16 ετικέτες τιμών ανομοιότητας, χρησιμοποιείται κατωφλιοποιημένο γραμμικό κόστος σύμφωνα με τη σχέση (6.9) ( $k = 1, V_{max} = 2$ ) με  $\lambda = 20$ . Για τη Venus με m = 20 ετικέτες, χρησιμοποιείται κατωφλιοποιημένο τετραγωνικό κόστος ( $k = 2, V_{max} = 7$ ) με  $\lambda = 50$ . Από τη στιγμή που ο όρος ομαλοποίησης δεν είναι μετρικός χρησιμοποιείται κίνηση τύπου διαστολής με εφαρμογή κατωφλιομένου κόστους αντιστοίχισης. Για τη Teddy με m = 60ετικέτες χρησιμοποιήθηκε το μοντέλο Potts ( $k = 1, V_{max} = 1$ ) με  $\lambda = 10$ . Για την ομάδα εικόνων Aloe με m = 60 ετικέτες ανομοιότητας βάθους, εφαρμόστηκε συνάρτηση ομαλοποίησης με παραμέτρους σύμφωνα με τη Σχέση (6.9),  $k = 2, V_{max} = 2$  και  $\lambda = 20$ . Για την ομάδα εικόνων map με m = 30ετικέτες ανομοιότητας βάθους, εφαρμόστηκε συνάρτηση ομαλοποίησης με παραμέτρους,  $k = 2, V_{max} = 2$  και  $\lambda = 20$ . Κοινοί παράμετροι στη συνάρτηση ενέργειας ομαλότητας, χρησιμοποιήθηκαν για τις ομάδες εικόνων map, sawtooth και rocks με πλήθος τιμών ετικέτας, m = 30, m = 40 και m = 60αντίστοιχα. Η τυπική τιμή στάθμης τοπικής ομαλότητας είναι  $w_{pq} = 1$ για όλα τα εικονοστοιχεία.

Στο Σχήμα 6.3 παρουσιάζονται οι εικόνες που χρησιμοποιήθηκαν για την αξιολόγηση των αλγορίθμων στέρεο που παρουσιάζονται στην παρούσα διατριβή. Τα αποτελέσματα εφαρμογής του αλγορίθμου αποκοπής γράφων με τις αντίστοιχες συναρτήσεις ενέργειας και ομαλοποίησης φαίνονται στο Σχήμα 6.4. Η ελαχιστοποίηση των συναρτήσεων αυτών επιτυγχάνεται με το ελάχιστο αθροιζόμενο τετραγωνικό κόστος ταιριάσματος στη ενέργεια στερεοσκοπικής συνέπειας (6.3) και απόλυτου κόστους στη συνάρτηση τμηματικής ομαλότητας (6.6). Στο Σχήμα 6.4 παρουσιάζονται οι χάρτες ανομοιότητας όπως αυτοί υπολογίστηκαν με βάση τις μετρήσεις φωτεινότητας σε σειρά πειραματικών στέρεο εικόνων. Οι μετρήσεις φωτεινότητας συνετέλεσαν στον υπολογισμό τόσο του κόστους ομαλότητας όσο και του κόστους δεδομένων (συνάρτηση



Σχήμα 6.3: Συλλογή πειραματικών εικόνων για επαλήθευση αλγορίθμων στέρεο ταιριάσματος μαζί με τους ιδανικούς χάρτες ανομοιότητας βάθους (συλλογή Middlebury, http://vision.middlebury.edu/stereo/data/). Γραμμή 1: Tsukuba, Venus Γραμμή 2: Teddy, Aloe, Γραμμή 3: Map, Sawtooth, Γραμμή 4: Rocks, Rocks



(a') Tsukuba

(β') Venus

(y') Teddy



(δ') Map

(ɛ') Rocks



(f') Aloe

( $\zeta$ ) Sawtooth



στερεοσκοπικής συνέπειας).

Στο ίδιο σύνολο πειραματικών στερεοσκοπικών εικόνων εφαρμόστηκε και η προτεινόμενη μέθοδος, βάση της οποίας οι χάρτες ανομοιότητας βάθους υπολογίζονται χρησιμοποιώντας στην ενεργειακή βελτιστοποίηση τη νέα έκφραση συνάρτησης ομοιότητας, όπως αυτή περιγράφεται στη Σχέση (6.6). Οι χάρτες ανομοιότητας βάθους που προκύπτουν από την εν λόγω προσέγγιση παρουσιάζονται στο Σχήμα 6.5. Με απλή επισκόπηση των αποτελεσμάτων η προτεινόμενη προσέγγιση οδηγεί σε πιο ομαλές λύσεις λιγότερο θορυβώδεις,



(α') Tsukuba

(β') Venus





(δ') Μαρ

(ε') Rocks



(F') Aloe

(ζ) Sawtooth

Σχήμα 6.5: Αποτελέσματα στις πειραματικές εικόνες (tsukuba, venus, map, aloe, sawtooth, rocks και teddy) με χρήση της προτεινόμενης συνάρτησης ενέργειας στερεοσκοπικής συνέπειας

όπως άλλωστε θα δειχθεί και παρακάτω αξιοποιώντας ποσοτικά μέτρα αξιολόγησης.

Για λόγους αριθμητικής ευστάθειας στους υπολογισμούς της βέλτιστης λύσης, οι τιμές ετερογενών ποσοτήτων (μονογενή φάση, πλάτος και φωτεινότητα) έχουν ομαλοποιηθεί με κοινή μέση τιμή και τυπική απόκλιση. Συνολικά, μία τέτοια προσέγγιση παρέχει τη δυνατότητα εκτίμησης του βάθους και της τριδιάστατης αποκατάστασης σε όλη την εικόνα και όχι μόνο στην περιοχή των εντοπισμένων χαρακτηριστικών (γεωμετρικές οντότητες όπως ευθείες, ακμές κτλ).

Εφαρμόστηκαν λοιπόν δύο βασικές προσεγγίσεις στην κατασκευή συναρτήσεων ενέργειας και ειδικά στην διαμόρφωση της ενέργειας στερεοσκοπικής συνέπειας ή της ενέργειας δεδομένων όπως συχνά ονομάζεται στη βιβλιογραφία. Η κλασσική έκφραση της εν λόγω συνάρτησης ενέργειας όπως αυτή περιγράφεται από την εξίσωση (5.26), βασίζεται αποκλειστικά σε μετρήσεις φωτεινότητας της εικόνας. Η προτεινόμενη προσέγγιση βασίζεται επιπρόσθετα σε μετρήσεις των γεωμετρικών και ενεργειακών χαρακτηριστικών του μονογενούς σήματος. Η τελευταία διαδικασία οδηγεί σε μία αναπαράσταση της τοπικής πληροφορίας και ενέργειας του σήματος, αμετάβλητης σε πιθανές και τυχαίες μεταβολές του φωτισμού ή αντανάκλασης.

Τα συνολικά αποτελέσματα υπολογισμού χαρτών ανομοιότητας παρουσιάζονται για τις κλασσικές πειραματικές εικόνες επαλήθευσης σε δύο σχήματα. Στο Σχήμα 6.4 παρουσιάζονται οι χάρτες ανομοιότητας βάθους που προκύπτουν από την εφαρμογή περιορισμών συνέπειας δεδομένων αξιοποιώντας αποκλειστικά την πληροφορία φωτεινότητας. Το Σχήμα 6.5 περιέχει χάρτες ανομοιότητας βάθους που προκύπτουν από την ελαχιστοποίηση της προτεινόμενης συνάρτησης ενέργειας δεδομένων. Η αρχική σύγκριση των χαρτών, οδηγεί στα ακόλουθα συμπεράσματα:

- Λιγότερο θορυβώδεις χάρτες στην εφαρμογή της προτεινόμενης προσέγγισης.
- Εμφάνιση περιοχών με αδιευκρίνιστες τιμές ανομοιότητας στην κλασική προσέγγιση (για παράδειγμα η εικόνα του χάρτη ανομοιότητας βάθους Teddy).
- Ομαλή μεταβολή των τιμών ανομοιότητας βάθους στα όρια των απεικονιζόμενων αντικειμένων, ως αποτέλεσμα της προτεινόμενης προσέγγισης.

Στον πίνακα 6.1 αναπαρίσταται η ποσοτική εκτίμηση της ποιότητας των αποτελεσμάτων για τις δύο μεθόδους, που προκύπτει μέσω του υπολογισμού του μέσου τετραγωνικού σφάλματος (RMS error) μεταξύ του υπολογιζόμενου χάρτη ανομοιότητας βάθους και του αντίστοιχου ιδανικού (ground truth), χωρίς τη χρήση τιμών αποκοπής.

Πέρα από τις δοκιμαστικές τεχνητές εικόνες για τη μέτρηση επίδοσης των αλγορίθμων, εκτελέστηκαν πειράματα με ομάδα φυσικών εικόνων αρχιτεκτονικής και αρχαιολογικής θεματολογίας. Οι εικόνες απεικονίζουν την πλάγια όψη του Ερεχθείου και μία πλαϊνή πύλη του Ηρώδειου, όπως απεικονίζονται στα Σχήματα 6.6.

Συναντώνται κοινές δυσκολίες με προηγούμενες προσεγγίσεις στην ανάλυση εικόνων όμοιας θεματολογίας. Η επαναλαμβανόμενη γεωμετρία στην



(α') Ερεχθείο, αριστερή εικόνα



(β') Ερεχθείο, δεξιά εικόνα



(γ') Ηρώδειο, αριστερή εικόνα



(δ') Ηρώδειο, δεξιά εικόνα

Σχήμα 6.6: Πειραματικές φυσικές εικόνες.

εικόνα των κιόνων του Ναού και η ομοιόμορφη φωτεινότητα στην εικόνα υφής της πύλης του θεάτρου, δημιουργούν τιμές ενέργειας πολύ κοντινές με συνέπεια την ύπαρξη κοντινών τοπικών ελαχίστων. Το αποτέλεσμα είναι, είτε η απώλεια δομικών χαρακτηριστικών στην εικόνα, είτε η ύπαρξη θορυβωδών περιοχών. Χρησιμοποιώντας κλασικές μεθόδους στον υπολογισμό ανομοιότητας με αποκλειστική χρήση της φωτεινότητας της εικόνας, οι χάρτες ανομοιότητας βάθους όπως αυτοί φαίνονται στα Σχήμα6.7(α') και Σχήμα6.7(γ') διαθέτουν περισσότερες περιοχές αδιευκρίνιστου βάθους συγκρινόμενοι με αυτούς που προκύπτουν από τη χρήση του νέου μέτρου ομοιότητας και της συνεπακόλουθης συνάρτησης ενέργειας. Επιπλέον αποκαθίστανται γραφικά περισσότερα δομικά στοιχεία από τη χρήση της νέας προτεινόμενης προσέγγισης, όπως φαίνεται στα Σχήματα 6.7(β') και 6.7(δ').

Στέρεο Εικόνες	RMS (μονογενη φίλτρα)	RMS (σύγκριση τιμών φωτεινότητας)		
Tsukuba	1.7	1.8		
Venus	4.76	5.07		
Teddy	10.15	10.21		
Мар	7.35	7.99		
Aloe	13.84	6.59		
Sawtooth	3.66	6.15		
Rocks	5.59	6.39		

Πίνακας 6.1: Συγκριτικός πίνακας RMS σφάλματος σε χάρτες ανομοιότητας που προκύπτουν είτε με αντιστοίχιση τιμών φωτεινότητας, είτε αξιοποιώντας τα μονογενή φίλτρα

Κατά τη διάρκεια της βελτιστοποίησης η εξέταση αθροιζόμενων τιμών κόστους σε συγκεκριμένες περιοχές της εικόνας, συνιστά την τοπική προσέγγιση του προβλήματος ελαχιστοποίησης καθολικού κόστους. Η διαδικασία αυτή πολλές φορές απαιτεί αποκοπή τιμών που υπερβαίνουν ένα προκαθορισμένο κατώφλι τιμών ανομοιότητας βάθους. Δεδομένου ότι οι χάρτες ανομοιότητας βάθους βαθμονομούνται σε κλίμακα φωτεινότητας γκρι 0-255, η τιμή αποκοπής 255 ισοδυναμεί με αποδοχή όλων των τιμών κόστους κατά τη βελτιστοποίηση.

Η διαδικασία της αποκοπής είναι απαραίτητη σε περιοχές της εικόνας κοντά στα άκρα της και αποκόβει λανθασμένα ζευγάρια αντίστοιχων εικονοστοιχείων από εικόνα σε εικόνα. Είναι λοιπόν σαφές ότι το επίπεδο κατωφλιοποίησης επηρεάζει άμεσα την ποιότητα των ανακτηθέντων χαρτών ανομοιότητας. Μελετήθηκε λοιπόν η μεταβολή του RMS λάθους ανομοιότητας βάθους συναρτήσει της τιμής αποκοπής τιμών κόστους, και τα αποτελέσματα φαίνονται στα διαγράμματα του Σχήματος 6.8. Οι τιμές αποκοπής κυμαίνον-







(β) Ερεχθείο, Προτεινόμενη προσέγγιση



(γ) Πύλη θεάτρου, Κλασσική προσέγγιση



(δ') Πύλη θεάτρου, Προτεινόμενη προσέγγιση

Σχήμα 6.7: Συγκριτική εφαρμογή νέας συνάρτησης ενέργειας σε πειραματικές εικόνες με αρχιτεκτονικά θέματα αρχαιολογικού ενδιαφέροντος. Πρώτη στήλη: χάρτες ανομοιότητας με χρήση κλασικών συναρτήσεων ενέργειας βασισμένων αποκλειστικά στην πληροφορία φωτεινότητας. Δεύτερη στήλη: χάρτες ανομοιότητας όπως αυτοί έχουν υπολογιστεί με χρήση ελάχιστων αποκοπών γράφων, καθώς επίσης και χρήση τοπικής πληροφορίας μέσω μονογενών φίλτρων. ται μεταξύ 1 και 255. Είναι αξιοπρόσεκτη η ποικιλία διαφοροποίησης των αποτελεσμάτων ανάλογα με την τιμή αποκοπής (truncation value).

Πολλές φορές η χρήση τιμής αποκοπής ως κατώφλι απαιτείται για την επίτευξη αξιόπιστων αποτελεσμάτων μέσω της αποκοπής λανθασμένων ζευγαριών αντιστοιχήσεων. Η χρήση επίσης διαδικασίας αποκοπής τιμών αποσκοπεί στην ομαλότητα της λύσης σε περιοχές κοντά στα όρια της εικόνας. Τα διαγράμματα αυτά αποτελούν απαραίτητο εργαλείο για την ανάδειξη της βέλτιστης τιμής αποκοπής ως προς το αντίστοιχο ελάχιστο RMS σφάλμα. Τα διαγράμματα μεταβολής του RMS σφάλματος σε σχέση με τις τιμές αποκοπής παρουσιάζουν μεγάλη ποικιλία στην γενική τους μορφή. Η τιμή αποκοπής για την οποία ελαχιστοποιείται το τετραγωνικό σφάλμα ποικίλει από εικόνα σε εικόνα και έχει άμεση εξάρτηση από συγκεκριμένα χαρακτηριστικά της κάθε εικόνας (υφή, γεωμετρία, αντίθεση). Γενικά παρατηρείται βελτίωση των αποτελεσμάτων για τιμές αποκοπής γύρω στο 20 με 50. Με απλή επισκόπηση των εικόνων ανομοιότητας βάθους συμπεραίνεται η αισθητή μείωση περιοχών αδιευκρίνιστου βάθους καθώς και η συνολική αποθορυβοποίηση της εικόνας με ευδιάκριτα τα όρια των αντικειμένων με συνολικά αξιόπιστη μεταβολή των τιμών ανομοιότητας βάθους.

Η χρήση του τετραγωνικού σφάλματος ως αποκλειστικού μέτρου απόφασης του βέλτιστου χάρτη ανομοιότητας πολλές φορές οδηγεί σε μη αποδεκτά αποτελέσματα, με ταυτόχρονη απώλεια συγκεκριμένης δομής του απεικονιζόμενου θέματος και περιορισμένη διακύμανση βάθους στην εικόνα ανομοιότητας. Στο Σχήμα 6.9 φαίνονται δύο παραδείγματα βέλτιστων χαρτών ανομοιότητας βάθους (Aloe, Rocks) ως προς το RMS σφάλμα από τους αντίστοιχους ιδανικούς χάρτες (ground truth). Οι αρχικές πειραματικές στέρεο εικόνες απεικονίζουν θέματα με πολύ κοντινές τιμές φωτεινότητας, γεγονός που δυσχεραίνει την εύρεση σωστών ταιριασμάτων. Η αναπαράσταση του βάθους στο αποτέλεσμα δεν υποδεικνύει κάποια συγκεκριμένη δομή του απεικονιζόμενου θέματος. Η έννοια της δομής αφορά κυρίως το βαθμό ανάκτησης των ορίων των αντικειμένων στην εικόνα.

Ανακύπτει λοιπόν η ανάγκη για πιο αξιόπιστα μέτρα αξιολόγησης των μεθόδων υπολογισμού χαρτών ανομοιότητας βάθους. Η ανακατασκευή μίας διδιάστατης απεικονιζόμενης σκηνής που ταυτόχρονα γίνεται αντιληπτή από την ανθρώπινη αίσθηση είναι ιδιαιτέρως αξιοποιήσιμη και επιθυμητή. Στην επόμενη παράγραφο θα παρουσιαστεί ένα μέτρο αξιολόγησης που συμπεριλαμβάνει και ποσοτικοποιεί το μέγεθος ανάκτησης της πληροφορίας δομής στην εικόνα και κατά συνέπεια και στους χάρτες ανομοιότητας βάθους.



Σχήμα 6.8: Διαγράμματα μεταβολής τετραγωνικού RMS σφάλματος ανομοιότητας με την τιμή αποκοπής (truncation) και κατωφλιοποίησης κόστους. Είναι εμφανής η μεγάλη εξάρτηση της εικόνας τόσο ως προς τη μορφή της γραφικής παράστασης όσο και ως προς την τιμή της αποκοπής με ελάχιστο σφάλμα ανομοιότητας βάθους. Το RMS σφάλμα υπολογίζεται από τους πρότυπους χάρτες ανομοιότητας όπως διατίθενται στη βιβλιογραφική βάση δεδομένων. Η μεταβολή του σφάλματος μελετάται συγκριτικά για τις δύο προσεγγίσεις στον υπολογισμό της ενέργειας στερεοσκοπικής συνέπειας: (α) αξιοποίηση γεωμετρικής πληροφορίας μονογενών φίλτρων (monogenic) (β) αποκλειστική χρήση πληροφορίας φωτεινότητας (intensity).



Σχήμα 6.9: Δυο χαρακτηριστικά παραδείγματα πειραματικών στέρεο εικόνων (Aloe, Rocks). Εξετάζοντας τις εικόνες σε κλίμακα του γκρι οι τιμές φωτεινότητας δεν παρουσιάζουν ιδιαίτερες μεταβολές. Αυτό αποτελεί ενδεικτικό παράδειγμα όπου παρά τα μικρά σφάλματα ως προς τις τιμές φωτεινότητας, το απεικονιζόμενο θέμα στους προκύπτοντες χάρτες ανομοιότητας παρουσιάζει εμφανώς ασαφείς δομές και απώλεια σημαντικής δομικής πληροφορίας (απώλεια η οποία δεν μπορεί να ερμηνευθεί/ποσοτικοποιηθεί με απλά μέτρα σφαλμάτων φωτεινότητας)
#### 6.5.3 Δείκτης Δομικής Ομοιότητας

Η μέτρηση της ποιότητας των χαρτών ανομοιότητας βάθους μέχρι τώρα βασίστηκε αποκλειστικά στη μέτρηση ορατών λαθών από τους αντίστοιχους ιδανικούς χάρτες βάθους μέσω αριθμητικών διαφορών. Παρατηρήθηκε όμως, ότι η αποκλειστική χρήση του τετραγωνικού σφάλματος για την εκτίμηση της ποιότητας υπολογισμού χαρτών ανομοιότητας βάθους πολλές φορές οδηγεί σε αποτελέσματα μακριά από τη εκλαβανόμενη ποιότητα της ανθρώπινης αίσθησης. Επίσης η ανθρώπινη όραση είναι εξοικειωμένη στον εντοπισμό της διαβάθμισης και της μεταβολής της δομικής πληροφορίας στην εικόνα.

Ορίζεται ως δομική πληροφορία σε μία εικόνα οι γεωμετρικές δομές της απεικονιζόμενης σκηνής ανεξάρτητα από τη μέση αντίθεση και φωτεινότητα. Στις παρακάτω εξισώσεις θα χρησιμοποιηθούν η μέση φωτεινότητα και αντίθεση αφού τα εν λόγω μεγέθη μεταβάλλονται τυχαία στην απεικονιζόμενη σκηνή. Το διάγραμμα ενός συστήματος εκτίμησης ομοιότητας της δομικής πληροφορίας δύο σημάτων φαίνεται στο Σχήμα 6.10.



Σχήμα 6.10: Υπολογιστικό σύστημα μέτρησης δείκτη δομικής ομοιότητας

Έστω **x** και **y** δύο μη-αρνητικά σήματα εικόνας. Θεωρούμε το ένα σήμα με άριστη ποιότητα, οπότε το χρησιμοποιούμενο μέτρο ομοιότητας θα χρησιμοποιηθεί ως ποσοτικό μέτρο της ποιότητας του έτερου σήματος. Το σύστημα χωρίζει τη λειτουργία μέτρησης δομικής ομοιότητας σε τρεις συγκρίσεις, ως προς φωτεινότητα, αντίθεση και δομή. Αρχικά η σύγκριση συντελείται ως προς τη φωτεινότητα. Θεωρώντας τα σήματα διακριτά αυτή υπολογίζεται ως τη μέση φωτεινότητα:

$$\mu_x = \frac{1}{N} \sum_{i=1}^{N} x_i$$
 (6.12)

Οπότε, η συνάρτηση σύγκρισης φωτεινότητας l(x, y) είναι μία συνάρτηση των  $\mu_x$  και  $\mu_y$ . Κατόπιν αφαιρείται η μέση φωτεινότητα από το σήμα και το προκύπτον σήμα  $x - \mu_x$  χρησιμοποιείται για τον υπολογισμό της τυπικής απόκλισης του σήματος ως μία έκφραση της αντίθεσης του σήματος. Στη διακριτή περίπτωση λοιπόν έχουμε:

$$\sigma_x = \left(\frac{1}{N-1} \sum_{i=1}^{N} (x_i - \mu_x)^2\right)^{1/2}$$
(6.13)

Επομένως, η σύγκριση αντίθεσης είναι ουσιαστικά σύγκριση των  $\sigma_x$  και  $\sigma_y$ . Τέλος το σήμα ομαλοποιείται από την τυπική του απόκλιση οπότε τα δύο υπό σύγκριση σήματα έχουν μοναδιαία τυπική απόκλιση. Η σύγκριση δομής λοιπόν  $s(\mathbf{x}, \mathbf{y})$  συντελείται σε αυτά τα ομαλοποιημένα σήματα  $\frac{(\mathbf{x} - \mu_x)}{\sigma_x}$  και  $\frac{(\mathbf{y} - \mu_y)}{\sigma_y}$ . Οι τρεις παραπάνω συναρτήσεις συνδυάζονται για να προκύψει το συνολικό μέτρο σύγκρισης:

$$S(\mathbf{x}, \mathbf{y}) = f(l(\mathbf{x}, \mathbf{y}), c(\mathbf{x}, \mathbf{y}), s(\mathbf{x}, \mathbf{y}))$$
(6.14)

Είναι σημαντικό το γεγονός ότι οι τρεις συναρτήσεις είναι ανεξάρτητες. Με άλλα λόγια αλλαγές σε φωτεινότητα ή αντίθεση στην εικόνα δεν επηρεάζουν τη δομή των εικόνων. Η ολοκλήρωση του ορισμού για την εξίσωση (6.14) απαιτεί τον προσδιορισμό των συναρτήσεων  $l(\mathbf{x}, \mathbf{y})$ ,  $c(\mathbf{x}, \mathbf{y})$  και  $s(\mathbf{x}, \mathbf{y})$ , όπως επίσης και το συνδυασμό τους ως συνάρτηση  $f(\cdot)$ .

Έτσι λοιπόν, για τη σύγκριση φωτεινότητας ορίζεται η συνάρτηση:

$$l(\mathbf{x}, \mathbf{y}) = \frac{2\mu_x \mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1}$$
(6.15)

Η σταθερά  $C_1$  έχει ένα σταθεροποιητικό ρόλο στην παραπάνω συνάρτηση προκειμένου να αποφευχθεί η διαίρεση με το μηδέν.

Η συνάρτηση σύγκρισης αντίθεσης παίρνει μία αντίστοιχη μορφή:

$$c(\mathbf{x}, \mathbf{y}) = \frac{2\sigma_x \sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2}$$
(6.16)

Η σταθερά  $C_2$  και σε αυτή την περίπτωση παίζει ένα ρόλο ρυθμιστικό για την αποφυγή διαίρεσης με το μηδέν.

Η σύγκριση δομής συντελείται μετά την αφαίρεση πληροφορίας φωτεινότητας και την ομαλοποίηση αντίθεσης, όπως έχει ήδη προαναφερθεί. Ουσιαστικά συσχετίζονται τα μοναδιαία διανύσματα  $\frac{(\mathbf{x} - \mu_x)}{\sigma_x}$  και  $\frac{(\mathbf{y} - \mu_y)}{\sigma_y}$  με τη

δομή των δύο σημάτων εικόνας. Η μεταξύ τους συσχέτιση (εκφραζόμενη με το εσωτερικό γινόμενο) αποτελεί ένα αποδοτικό και απλό μέτρο της δομικής ομοιότητάς τους. Η συσχέτιση μεταξύ των  $\frac{(\mathbf{x} - \mu_x)}{\sigma_x}$  και  $\frac{(\mathbf{y} - \mu_y)}{\sigma_y}$ ισοδυναμεί με τη συσχέτιση μεταξύ **x** και **y**. Η συνάρτηση σύγκρισης δομής ακολουθεί:

$$s(\mathbf{x}, \mathbf{y}) = \frac{\sigma_{xy} + C_3}{\sigma_x \sigma_y + C_3}$$
(6.17)

Η σταθερά  $C_3$  και σε αυτή την περίπτωση έχει όμοια ρυθμιστικό ρόλο με τις προηγούμενες συναρτήσεις σύγκρισης.

Τέλος, συνδυάζοντας τις τρεις συναρτήσεις σύγκρισης (6.15), (6.16) και (6.17) το μέτρο ομοιότητας που προκύπτει ονομάζεται Δείκτης Δομικής Ομοιότητας (Structure SIMilarity index, SSIM) [88], και ορίζεται ως ακολούθως:

$$SSIM(\mathbf{x}, \mathbf{y}) = [l(\mathbf{x}, \mathbf{y})]^{\alpha} [c(\mathbf{x}, \mathbf{y})]^{\beta} [s(\mathbf{x}, \mathbf{y})]^{\gamma}$$
(6.18)

όπου  $\alpha > 0$ ,  $\beta > 0$  και  $\gamma > 0$  παράμετροι που ρυθμίζουν τη σχετική βαρύτητα στον υπολογισμό των τριών συναρτήσεων σύγκρισης. Προκειμένου να απλοποιηθεί η παραπάνω έκφραση, τίθεται  $\alpha = \beta = \gamma = 1$  και  $C_3 = C_2/2$ . Η μαθηματική έκφραση λοιπόν του δείκτη ομοιότητας γίνεται:

$$SSIM(\mathbf{x}, \mathbf{y}) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}$$
(6.19)

Η εκτίμηση δομικής ομοιότητας σε μία εικόνα πολλές φορές πραγματοποιείται τοπικά παρά καθολικά. Πολλά χαρακτηριστικά της εικόνας έχουν μεγάλη χωρική εξάπλωση στο επίπεδο της εικόνας με αρκετές μεταβολές σε παραμέτρους τους. Επιπρόσθετα, η εκτίμηση τοπικά της δομικής ομοιότητας της εικόνας συντελεί σε εκτίμηση συνολικά της μεταβολής της δομικής ομοιότητας από περιοχή σε περιοχή της εικόνας.

Τα στατιστικά μέτρα  $\mu_x$ ,  $\sigma_x$  και  $\sigma_{xy}$  υπολογίζονται στα πλαίσια ενός τετραγωνικού παραθύρου διάστασης  $8 \times 8$  το οποίο κινείται από εικονοστοιχείο σε εικονοστοιχείο σε ολόκληρη την εικόνα. Σε κάθε βήμα υπολογίζονται τα στατιστικά μέτρα και κατά συνέπεια ο δείκτης δομικής ομοιότητας (SSIM).

Παρόλη την αποτελεσματικότητα της τοπικής προσέγγισης, η εκτίμηση του μέτρου δομικής ποιότητας στο σύνολο της εικόνας παραμένει χρήσιμη σε πολλές εφαρμογές. Χρησιμοποιήθηκε λοιπόν προς αυτήν την κατεύθυνση ο μέσος δείκτης δομικής ομοιότητας δύο σημάτων (Mean Structural SIMilarity, MSSIM):

$$MSSIM(\mathbf{X}, \mathbf{Y}) = \frac{1}{M} \sum_{i=1}^{M} SSIM(\mathbf{x}_i, \mathbf{y}_i)$$
(6.20)

όπου **X** το σήμα αναφοράς (ιδανικής ποιότητας) και **Y** το πειραματικά υπολογισμένο σήμα. Στις πειραματικές εφαρμογές της παρούσας διατριβής το σήμα αναφοράς αντιπροσωπεύει τον ιδανικό χάρτη ανομοιότητας βάθους (ground truth), ενώ το άλλο σήμα, του οποίου η δομική ομοιότητα με το αντίστοιχο αναφοράς εξετάζεται, αντιπροσωπεύεται από την εικόνα του χάρτη ανομοιότητας βάθους όπως υπολογίστηκε από τους προτεινόμενους ή κλασσικούς αλγορίθμους.

### 6.5.4 Πειραματική Αξιολόγηση βασισμένη στο Δείκτη Δομικής Ποιότητας

Ο δείκτης δομικής ομοιότητας (SSIM) υιοθετήθηκε στη διαδικασία συγκριτικής μελέτης των χαρτών ανομοιότητας βάθους όπως αυτοί υπολογίστηκαν είτε μέσω της σύγκρισης τιμών φωτεινότητας κατά την κλασική προσέγγιση, είτε αξιοποιώντας τα γεωμετρικά χαρακτηριστικά της διδιάστατης αναπαράστασης του σήματος εικόνας από το μονογενές σήμα. Ο δείκτης δομικής ομοιότητας υπολογίζεται με τη σύγκριση δύο σημάτων εικόνας: της εικόνας του χάρτη ανομοιότητας βάθους που υπολογίστηκε με την προτεινόμενη συνάρτηση στερεοσκοπικής συνέπειας και της εικόνας του ιδανικού χάρτη ανομοιότητας βάθους για τις πειραματικές τεχνητές εικόνες που χρησιμοποιήθηκαν. Συνεπώς, από τη στιγμή που ο δείκτης υπολογίζεται με βάση τους ιδανικούς χάρτες ανομοιότητας βάθους, αποτελεί μέτρο ποιότητας των αντίστοιχων που υπολογίστηκαν με την προτεινόμενη μεθοδολογία αντιστοίχισης. Με δεδομένο αυτό στο εξής θα αναφερόμαστε στο εν λόγω δείκτη ως *δείκτη δομικής ποιότητας*.

Υπολογίζεται λοιπόν ο δείκτης δομικής ποιότητας για διάφορες τιμές αποκοπής διαφορών κόστους σύγκρισης (matching cost truncation values). Η εύρεση της βέλτιστης αποκοπής για τις διάφορες τιμές του δείκτη δομικής ποιότητας αντιστοιχεί όπως και στην περίπτωση τετραγωνικού λάθους σε ένα περιορισμένου εύρους αποδεκτό πλήθος τιμών κόστους σύγκρισης. Αυτό οδηγεί σε αξιόπιστα μοντέλα βελτιστοποίησης της στερεοσκοπικά συνεπούς και τμηματικά ομαλής συνάρτησης ενέργειας. Τα συνολικά αποτελέσματα παρουσιάζονται στο Σχήμα 6.11.

Στο συγκεκριμένο σχήμα παρουσιάζονται συγκριτικά διαγράμματα μεταβολής δείκτη δομικής ποιότητας σε συνάρτηση με τις διάφορες τιμές αποκοπής κόστους σύγκρισης. Τίθενται υπό σύγκριση δύο εκφράσεις της συνάρτησης ενέργειας που αντιπροσωπεύει τη συνθήκη στερεοσκοπικής συνέπειας. Η μία έκφραση βασίζεται αποκλειστικά σε συγκρίσεις τιμών φωτεινότητας με σκοπό την εύρεση κοινών εικονοστοιχείων από εικόνα σε εικόνα. Η άλλη έκφραση αξιοποιεί το προτεινόμενο μέτρο ομοιότητας που βασίζεται στη μονογενή αναπαράσταση του σήματος εικόνας, αξιοποιώντας επιπρόσθετα τη γεωμετρική και ενεργειακή πληροφορία του μονογενές σήματος. Με δεδομένη τη βέλτιστη τιμή του δείκτη ομοιοτητας ίση με ένα, διαπιστώνεται η σαφώς ανώτερη δομική ποιότητα της προτεινόμενης προσέγγισης με τιμές του δείκτη δομικής ποιότητας κοντά στη μονάδα σχεδόν για κάθε τιμή αποκοπής.



Σχήμα 6.11: Διαγράμματα μεταδολής δείκτη δομικής ποιότητας (SSIM) σε συνάρτηση με τις τιμές αποκοπής του αθροιζόμενου κόστους σύγκρισης. Χρησιμοποιήθηκαν πειραματικές στέρεο εικόνες για την εκτίμηση της επίδοσης της προτεινόμενης συνάρτησης ενέργειας. Καλύτερη δομική ποιότητα στην εικόνα ανομοιότητας βάθους (τιμές δείκτη SSIM πιο κοντά στη μονάδα) σημαίνει μεγαλύτερος βαθμός αντιληπτικότητας της εικόνας από την ανθρώπινη αίσθηση.

Λαμβάνοντας υπόψιν τα συγκριτικά διαγράμματα του Σχήματος 6.11 είναι δυνατόν να εντοπιστεί ο βέλτιστος χάρτης ανομοιότητας βάθους σε σχέση με το δείκτη δομικής ποιότητας (SSIM). Κατ' αυτόν τον τρόπο στο Σχήμα 6.12 παρουσιάζονται οι χάρτες με το μέγιστο δείκτη ποιότητας.

Οι παραπάνω μετρήσεις υλοποιήθηκαν στους χάρτες ανομοιότητας βάθους όπως αυτοί προέκυψαν από την ελαχιστοποίηση της προτεινόμενης συνάρτησης ενέργειας. Η επιλογή του χάρτη ανομοιότητας βάθους με βάση το δείκτη δομικής ποιότητας οδηγεί σε χάρτες λιγότερο θορυβώδεις με ομαλότερη διακύμανση των τιμών ανομοιότητας σε σύγκριση με τους αντίστοιχους του Σχήματος 6.5.

### 6.6 Συμπεράσματα Κεφαλαίου

Στο κεφάλαιο αυτό πραγματοποιήθηκε η μελέτη και η κωδικοποίηση της τοπικής πληροφορίας μίας εικόνας, η οποία συνίσταται σε τιμές κατεύθυνσης, φάσης και χρώματος. Χρησιμοποιήθηκε τόσο η γεωμετρική όσο και η δομική πληροφορία της εικόνας σε ένα πρόβλημα στέρεο αντιστοίχισης για τη δημιουργία ενός εμπλουτισμένου μέτρου ομοιότητας συναφών περιοχών ή ακόμη και μεμονωμένων σημείων από εικόνα σε εικόνα. Το μέτρο αυτό αξιοποιεί την διδιάστατη τοπική αναπαράσταση της εικόνας μέσω του αναλυτικού σήματος. Η στερεοσκοπική σύγκριση λοιπόν βασίζεται σε εύρωστα μέτρα ομοιότητας που βασίζονται σε ενεργειακή και γεωμετρική πληροφορία, η οποία είναι αμετάβλητη σε τυχαίες μεταβολές φωτισμού ή αντανάκλασης στην εικόνα, ειδικά σε εξωτερικές συνθήκες φωτογράφησης.

Σε μία στέρεο διάταξη επιχειρήθηκε να πραγματοποιηθεί ένα πιο αξιόπιστο ταίριασμα χαρακτηριστικών από εικόνα σε εικόνα. Το γεγονός αυτό θα οδηγήσει σε μία ακριβέστερη τριδιάστατη αποκατάσταση.

Μελετήθηκε λοιπόν η εφαρμογή μίας νέας προτεινόμενης μορφής συνάρτησης ενέργειας δεδομένων σε πειραματικά δεδομένα που αφορούν είτε πειραματικές εικόνες δανεισμένες από τη βιβλιογραφία είτε σύνθετες εικόνες αρχιτεκτονικού και αρχαιολογικού ενδιαφέροντος εξωτερικού χώρου κάτω από τυχαίες συνθήκες φωτισμού. Η ενέργεια δεδομένων αντιπροσωπεύει τη βασική προϋπόθεση στερεοσκοπικής συνέπειας στα προβλήματα ανάκτησης βάθους από εικόνες. Σε πρώτη φάση, η προτεινόμενη μέθοδος εφαρμόστηκε σε πειραματικές εικόνες με την ταυτόχρονη σύγκριση των αποτελεσμάτων με τους ιδανικούς χάρτες ανομοιότητας βάθους των εικόνων. Τα πειραματικά αποτελέσματα που προέκυψαν επιβεβαιώνουν την ανάκτηση χαρτών ανομοιότητας βάθους μεγαλύτερης ακρίβειας με καλύτερη διατήρηση γεωμετρικών χαρακτηριστικών στην τελική εικόνα του χάρτη βάθους. Το γεγονός αυτό επιβεβαιώνεται σε εικόνες σύνθετων δομικών γεωμετρικών χαρακτηριστικών



(a') Map, (truncation value=50, SSIM=0.87)



( $\beta$ ) teddy, (truncation value=255, SSIM=0.34)



( $\gamma$ ) Aloe, (truncation value=50, SSIM=0.71)



(δ) sawtooth, (truncation value=50, SSIM=0.86)



(ε') rocks, (truncation value=50, SSIM=0.82)



(F) venus, (truncation value=50, SSIM=0.84)



(ζ) tsukuba, (truncation value=50, SSIM=0.74)

Σχήμα 6.12: Χάρτες ανομοιότητας βάθους πειραματικών ομάδων στέρεο εικόνων Aloe, Tsukuba, Sawtooth, Venus, Map, Rocks και Teddy με μέγιστο δείκτη δομικής ποιότητας SSIM.

όπως αυτές αρχαιολογικού και αρχιτεκτονικού ενδιαφέροντος. Παρατηρείται ότι η κλασσική προσέγγιση στη διαδικασία ανάκτησης του βάθους, οδηγεί στην απώλεια λεπτομερειών στη δομή αρχιτεκτονικών και αρχαιολογικών ευρημάτων.

Η ποσοτική αξιολόγηση των αποτελεσμάτων βασίστηκε αρχικά στην ύπαρξη βελτιωμένων ποσοστών απόκλισης από τις ιδανικές τιμές. Στη συνέχεια και προς την κατεύθυνση της πιο αντικειμενικής αξιολόγησης της ποιότητας των αποτελεσμάτων όσο αφορά την επίδοση μεθόδων ανάκτησης βάθους χρησιμοποιήσαμε ένα δείκτη μέτρησης δομικής ποιότητας (SSIM) ο οποίος προσομοιάζει τον τρόπο εκτίμησης των δομικών χαρακτηριστικών από την ανθρώπινη αίσθηση. Μελετήθηκε η μεταβολή του στις διάφορες τιμές αποκοπής κόστους, για να επιβεβαιωθεί γενικά η διατήρηση της ανώτερης δομικής ποιότητας για το σύνολο των επιπέδων κατωφλιοποίησης του κόστους αντιστοίχισης που εκφράζεται από τη συνάρτηση ενέργειας δεδομένων. Εικόνες χαρτών με ελάχιστο τετραγωνικό σφάλμα απόκλισης αδυνατούσαν να απεικονίσουν με σαφήνεια δομικά στοιχεία της αποκατεστημένης σκηνής. Οδηγηθήκαμε λοιπόν στη χρήση αυτού του δείκτη προκειμένου να αντιμετωπιστούν ανακρίβειες που σχετίζονται με την αξιολόγηση τετραγωνικών σφαλμάτων απόκλισης.

## Κεφάλαιο 7

# Πολυοπτικό Ταίριασμα Εικόνων -Ανάκτηση Δομής από Κίνηση

### 7.1 Εισαγωγή

Η τριδιάστατη ανακατασκευή απεικονιζομένων αντικειμένων από σειρά εικόνων αποτελεί κλασικό πρόβλημα στην όραση υπολογιστών μαζί με τη στέρεο ανάλυση εικόνων. Το πρόβλημα μπορεί να θεωρηθεί ως μία φυσική επέκταση της στέρεο ανακατασκευής, αν και συχνά εξελίσσεται ως περισσότερο πολύπλοκο. Σε αυτό το κεφάλαιο, θα επιχειρηθεί μία πρώτη απόπειρα επέκτασης των προτεινόμενων αλγορίθμων αντιστοίχισης σε πλήθος εικόνων περισσοτέρων των δύο. Βασική αιτία στην αυξημένη πολυπλοκότητα του πολυοπτικού ταιριάσματος αποτελεί ο μη κοινός αριθμός απεικονιζομένων αντικειμένων από εικόνα σε εικόνα. Στα πλαίσια ενός κλασικού προβλήματος στέρεο ταιριάσματος από ζευγάρι εικόνων, τα περισσότερα στοιχεία της απεικονιζόμενης σκηνής είναι ορατά και από τις δύο στέρεο εικόνες, οπότε η έννοια του βαθμού ορατότητας αντικειμένων από εικόνα σε εικόνα συνήθως δε λαμβάνεται υπόψη. Αντίθετα κατά την επεξεργασία πολλαπλών σκηνών λίγα στοιχεία παραμένουν απεικονιζόμενα και κατά συνέπεια η ανάγκη προσδιορισμού του βαθμού ορατότητας τους είναι υπαρκτή.

Το πρόβλημα πολυοπτικού ταιριάσματος και κατασκευής χάρτη ανομοιότητας βάθους θα αναπαρασταθεί ως ένα πρόβλημα ελαχιστοποίησης μίας συνάρτησης ενέργειας. Η διαδικασία ελαχιστοποίησης θα βασιστεί σε ταχείς αλγορίθμους ελάχιστης αποκοπής γράφων, όμοια με τη περίπτωση ζεύγους εικόνων, όπως παρουσιάστηκε στο κεφάλαιο 6.

Στη συνέχεια, θα προσδιοριστεί η συνάρτηση ενέργειας που αντιπροσωπεύει το πρόβλημα πολυοπτικής αντιστοίχισης εικόνων. Επιπρόσθετα, θα μελετηθεί η εισαγωγή ενός νέου περιορισμού πλαισίου, του περιορισμού ορατότητας, στο πρόβλημα της πολυοπτικής αντιστοίχισης. Η ελαχιστοποίηση της συνάρτηση ενέργειας που αφορά το πρόβλημα πολυοπτικού ταιριάσματος, ικανοποιεί τις ακόλουθες ιδιότητες:

- Αποδοτική διαχείριση των γειτονικών πλαισίων εικόνας που χρησιμοποιούνται στη διαδικασία αντιστοίχισης.
- Υπολογισμός ετικετοποίησης τμηματικά ομαλής με ταυτόχρονη δυνατότητα διατήρησης ασυνεχειών, ειδικότερα στα όρια των αντικειμένων όπου οι τιμές ετικετών εμφανίζουν μεγάλες μεταβολές.
- Κατάλληλη μαθηματική αναπαράσταση του περιορισμού ορατότητας στο πλήθος των εξεταζόμενων πλαισίων εικόνας.

## 7.2 Συνάρτηση Ενέργειας Πολυοπτικής Αντιστοίχισης

Στη συγκεκριμένη παράγραφο θα τεθεί το μαθηματικό πλαίσιο στο πρόβλημα πολυοπτικής αντιστοίχισης από στερεοσκοπικές εικόνες. Έστω n βαθμονομημένες εικόνες που αφορούν την απεικόνιση της ίδιας σκηνής από διαφορετικές οπτικές γωνίες (ή διαφορετικές χρονικές στιγμές). Το σύνολο των εικονοστοιχείων στην εικόνα από την οπτική γωνία (κάμερα) i συμβολίζεται  $\mathcal{P}_i$ και  $\mathcal{P} = \mathcal{P}_1 \cup \mathcal{P}_2 \cup \cdots \cup \mathcal{P}_n$  το σύνολο όλων των εικονοστοιχείων από τις εικόνες που χρησιμοποιούνται στον υπολογισμό της τριδιάστατης δομής της σκηνής. Ένα εικονοστοιχείο  $p \in \mathcal{P}$  αντιστοιχεί σε μία οπτική ακτίνα στον τριδιάστατο χώρο. Στη συνέχεια ας θεωρηθεί ένα σημείο τομής της ακτίνας αυτής με ένα αντικείμενο της απεικονιζόμενης σκηνής. Βασικός στόχος αποτελεί η εύρεση του βάθους του σημείου αυτού για κάθε αντίστοιχό του εικονοστοιχείο στο εξεταζόμενο πλαίσιο εικόνας. Αναζητείται λοιπόν ένα σύνολο ετικετοποίησης τέτοιο ώστε:  $f : \mathcal{P} \to \mathcal{L}$  όπου  $\mathcal{L}$  το διακριτό σύνολο τιμών ετικετών που αντιστοιχεί σε διαφορετικές τιμές βάθους. Στη συγκεκριμένη υλοποίηση οι τιμές ετικέτας αντιπροσωπεύουν τις τιμές βάθους από ένα σταθερό πλαίσιο εικόνας αξιοποιώντας την πληροφορία προερχόμενη από τα γειτονικά πλαίσια εικόνων.

Ένα ζευγάρι  $\langle p, l \rangle$ , όπου  $p \in \mathcal{P}$ , απεικονίζει ένα σημείο του τριδιάστατου χώρου. Στο εξής θα αναφέρουμε τέτοια ζεύγη ως τριδιάστατα σημεία. Σύμφωνα με τη προσέγγιση των Kolmogorov και Zabih στο [79], υιοθετείται η υπόθεση ότι η θέση της κάμερα συμπίπτει με τη θέση ενός επιπέδου στο χώρο και έχει ορατή περιοχή ένα επίπεδο που ανήκει στο τριδιάστατο αντικείμενο. Η ανατιθέμενη τιμή ετικέτας l αντιστοιχεί σε ένα επίπεδο που ανήκει στο τριδιάστατο αντικείμενο, και ένα τριδιάστατο σημείο  $\langle p, l \rangle$ , είναι το σημείο τομής της οπτικής ακτίνας, που διέρχεται από το εικονοστοιχείο p, με το επίπεδο που χαρακτηρίζεται από την τιμή της ετικέτας l.

Στη συνέχεια εισάγεται το σύνολο I το οποίο περιλαμβάνει ζεύγη τριδιάστατων σημείων  $\langle p_1, l_1 \rangle, \langle p_1, l_2 \rangle$  που είναι μεταξύ τους σε κοντινή απόσταση στο τριδιάστατο χώρο και ικανοποιούν τον ακόλουθο περιορισμό:

• Τα τριδιάστατα σημεία που ανήκουν στο σύνολο I βρίσκονται στο ίδιο βάθος, για παράδειγμα εάν  $\{\langle p_1, l_1 \rangle, \langle p_1, l_2 \rangle\} \in I$  τότε  $l_1 = l_2$ 

Στον παραπάνω ορισμό του συνόλου I αναφερθήκαμε στην έννοια της κοντινής απόστασης των τριδιάστατων σημείων. Είναι ανάγκη λοιπόν να προσδιοριστεί το κριτήριο εγγύτητας των σημείων. Το συγκεκριμένο κριτήριο έχει ως ακολούθως: Εάν p είναι ένα εικονοστοιχείο στην εικόνα i και q ένα εικονοστοιχείο σε γειτονικό πλαίσιο εικόνας j, τότε τα τριδιάστατα σημεία  $\langle p, l \rangle$ ,  $\langle q, l \rangle$  ικανοποιούν το εν λόγω κριτήριο εάν πιο κοντινό εικονοστοιχείο στην προβολή του  $\langle p, l \rangle$  στην εικόνα j είναι το εικονοστοιχείο q.

Στη συνέχεια θα περιγραφεί η συνάρτηση ενέργειας που μοντελοποιεί το πρόβλημα πολυοπτικής αντιστοίχισης. Η ενέργεια συνοψίζεται σε τρεις όρους:

$$E(f) = E_{data}(f) + E_{smoothness}(f) + E_{visibility}(f)$$
(7.1)

Ο όρος  $E_{data}$  αντιπροσωπεύει τους περιορισμούς φωτογραφικής συνέπειας και προσδιορίζεται ως εξής:

$$E_{data} = \sum_{\langle p, f(p) \rangle, \langle q, f(q) \rangle \in I} D(p, q)$$
(7.2)

Η μη αρνητική συνάρτηση D(p,q) εξαρτάται από τις τιμές φωτεινότητας των εικονοστοιχείων p και q. Σύμφωνα όμως με την προσέγγιση της παρούσας διατριβής η πληροφορία τιμών φωτεινότητας εμπλουτίζεται με τη διανυσματική αναπαράσταση της γεωμετρικής πληροφορίας του διδιάστατου αναλυτικού σήματος και εκφράζεται ως άθροισμα κόστους αντιστοίχισης στο σύνολο των υποψήφιων εικονοστοιχείων, όπως παρουσιάστηκε στη Σχέση (6.6).

Ο όρος της ομαλότητας στην εξίσωση συνάρτησης ενέργειας αφενός συντελεί στην ύπαρξη ομαλών λύσεων στο πρόβλημα, και αφετέρου εισάγει την έννοια της γειτνίασης εικονοστοιχείων. Ο ορισμός του συστήματος γειτνίασης εικονοστοιχείων ακολουθεί:

$$\mathcal{N} \subset \{\{p,q\} | p,q \in \mathcal{P}\}$$
(7.3)

Ο παραπάνω ορισμός αντιπροσωπεύει εν<br/> δυνάμει ένα σύστημα 4 γειτόνων. Τα εικονοστοιχεί<br/>α $p=(p_x,p_y)$ και  $q=(q_x,q_y)$ είναι γειτονικά εάν ανήκουν στην

ίδια εικόνα και ισχύει  $|p_x - q_x| + |p_y - q_y| = 1$ . Συνεπώς, ακολουθώντας τη μεθοδολογία και την ορολογία που ακολουθήθηκε για τη στέρεο αντιστοίχιση, ο όρος ομαλοποίησης γράφεται:

$$E_{smoothness}(f) = \sum_{\{p,q\} \in \mathcal{N}} V_{\{p,q\}}(f(p), f(q))$$
(7.4)

Η συνάρτηση  $V_{\{p,q\}}$  απαιτείται να είναι μετρική. Οδηγούμαστε λοιπόν σε τμηματικά ομαλή λύση ετικετοποίησης ενώ ταυτόχρονα διατηρούνται τοπικές ασυνέχειες χρημοποιώντας το κατάλληλο εύρωστο μετρικό. Για παράδειγμα:

$$V(l_1, l_2) = min(|l_1 - l_2|, K)$$
(7.5)

όπου Κ μία σταθερά.

Στη διαδικασία πολυοπτικής ανάκτησης χαρτών ανομοιότητας βάθους εισήχθη στη μορφή της συνάρτησης ενέργειας ένας επιπλέον περιορισμός πλαισίου: ο περιορισμός τήρησης ορατότητας στοιχείων στις διαδοχικές εικόνες  $(E_{visibility})$ . Η μαθηματική έκφραση του εν λογώ περιορισμού παίρνει την τιμή μηδέν εάν ικανοποιείται και την τιμή άπειρο σε αντίθετη περίπτωση. Χρησιμοποιήθηκε ως σύνολο αλληλεπιδρόντων τριδιάστατων σημείων το  $I_{vis}$  το οποίο περιέχει εκείνα τα ζευγάρια τριδιάστατων σημείων της απεικονιζόμενης σκηνής που δεν ικανοποιούν τον περιορισμό.

$$E_{visibility}(f) = \sum_{\langle p, f(p) \rangle, \langle q, f(q) \rangle \in I_{vis}} \infty$$
(7.6)

Το σύνολο  $I_{vis}$  ικανοποιεί τον ακόλουθο περιορισμό:

• Το σύνολο  $I_{vis}$  περιλαμβάνει τα τριδιάστατα σημεία με διαφορετικές τιμές βάθους. Για παράδειγμα εάν  $\{\langle p_1, l_1 \rangle, \langle p_1, l_2 \rangle\} \in I_{vis}$  τότε  $l_1 \neq l_2$ .

Ο περιορισμός ομαλότητας εξασφαλίζει το γεγονός ότι εάν ένα τριδιάστατο σημείο  $\langle p, l \rangle$  ανήκει σε μία ετικετοποίηση f (l = f(p)) τότε δεν είναι ορατό από τις υπόλοιπες κάμερες ενός πειράματος πολυοπτικής αντιστοίχισης. Πιο συγκεκριμένα, εάν μία οπτική ακτίνα που αντιστοιχεί στο εικονοστοιχείο q ικανοποιεί το κριτήριο εγγύτητας με το τριδιάστατο σημείο  $\langle p, l \rangle$ , τότε η μέγιστη τιμή βάθους είναι l.

Με βάση την παραπάνω παρατήρηση μπορούμε να προχωρήσουμε σε έναν εναλλακτικό ορισμό για το σύνολο  $I_{vis}$ , αξιοποιώντας το σύνολο I. Συνεπώς, το σύνολο  $I_{vis}$  περιέχει όλα εκείνα τα ζεύγη τριδιάστατων σημείων  $\{\langle p, l \rangle, \langle q, l' \rangle\}$ τέτοια ώστε τα  $\langle p, l \rangle$  και  $\langle q, l \rangle$  να ανήκουν στο σύνολο I και l' > l με δεδομένο ότι οι τιμές ετικετών αντιστοιχούν σε αυξανόμενο βάθος από το σταθερό πλαίσιο εικόνας στο οποίο γίνονται οι υπολογισμοί αντιστοίχισης. Στο Σχήμα 7.1 φαίνεται μία γραφική αναπαράσταση του περιορισμού ορατότητας που χρησιμοποιήθηκε στο πρόβλημα πολυοπτικής στέρεο αντιστοίχισης. Εξετάζονται στέρεο εικόνες από δύο κάμερες ( $C_1$  και  $C_2$ ). Έχουν επίσης σχεδιαστεί πέντε κάθετες γραμμές που αντιπροσωπεύουν πέντε διαφορετικές τιμές ετικέτας βάθους. Οι τιμές κατανέμονται με αυξανόμενη απόσταση από τις σταθερές κάμερες. Στο σχήμα φαίνονται δύο εικονοστοιχεία:

- Το p το οποίο είναι ορατό από την κάμερα  $C_1$  με το αντίστοιχο τριδιάστατο σημείο  $\langle p, 2 \rangle$  (στρογγυλό μαύρο σημείο στην κάθετο i = 2) σε βάθος 2.
- Το q το οποίο είναι ορατό από την κάμερα  $C_2$  με το αντίστοιχο επίσης τριδιάστατο σημείο  $\langle q, 2 \rangle$  (στρογγυλό κόκκινο σημείο στην κάθετο i = 2) σε ίδια στάθμη βάθους 2.

Ta σημεία αυτά βρίσκονται σε ίδιο βάθος και αποτελούν μέρη του συνόλου I,  $\{\langle p, 2 \rangle, \langle q, 2 \rangle\} \in I$ , ικανοποιώντας ταυτόχρονα τη συνθήκη φωτογραφικής συνέπειας. Το στρογγυλό πράσινο σημείο στην κάθετο i = 3 αντιπροσωπεύει το τριδιάστατο σημείο  $\langle q, 3 \rangle$  εικονοστοιχείου με διαφορετική τιμή ετικέτας (μεγαλύτερο βάθος) και βρίσκεται πίσω από το κόκκινο σημείο. Συνεπώς το ζευγάρι τριδιάστατων σημείων  $\{\langle p, 2 \rangle, \langle q, 3 \rangle\}$  ανήκουν στο σύνολο  $I_{vis}$ . Αυτό σημαίνει ότι εάν η οπτική ακτίνα p από την κάμερα  $C_1$  συναντάει το κόκκινο στρογγυλό σημείο  $\langle p, 2 \rangle$ , για την οπτική ακτίνα q από την κάμερα  $C_2$ , δεν είναι ορατό το πράσινο σημείο  $\langle q, 3 \rangle$ .

## 7.3 Πειραματική Αξιολόγηση Πολυοπτικής Αποκατάστασης Ανομοιότητας Βάθους

Η παραπάνω συνάρτηση ενέργειας με ενσωματωμένο το περιορισμό ορατότητας, όπως φαίνεται στη Σχέση (7.1), εφαρμόστηκε σε ομάδες τριών πειραματικών στέρεο εικόνων διατεταγμένων σε σειρά αυξανόμενου βάθους. Οι πειραματικές εικόνες είναι δανεισμένες από την βάση δεδομένων του Middlebury, όπως φαίνονται στο Σχήμα 6.3. Η συνάρτηση ενέργειας πέρα από την ενσωματωμένη συνάρτηση ορατότητας, αποτελείται από τη συνάρτηση ενέργειας ομαλότητας η οποία είναι κοινή με τα πειράματα στέρεο αντιστοίχισης που παρουσιάστηκαν στο προηγούμενο κεφάλαιο (παράγραφος 6.5.2). Η συνάρτηση ενέργειας δεδομένων ή φωτογραφικής συνέπειας  $E_{data}$  που χρησιμοποιήθηκε είναι η προτεινόμενη εκδοχή συνάρτησης αξιοποιώντας τα γεωμετρικά χαρακτηριστικά του διδιάστατου αναλυτικού σήματος, την κατεύθυνση, προσανατολισμό και τοπική ενέργεια σήματος πληροφορίας, όπως παρουσιάστηκε εκτενώς στο κεφάλαιο 6.



Σχήμα 7.1: Γραφική αναπαράσταση του περιορισμού ορατότητας στην περίπτωση ανάκτησης χάρτη ανομοιότητας βάθους από πολυοπτικό στέρεο ταίριασμα χαρακτηριστικών εικόνας [79].

Τα αποτελέσματα φαίνονται στο Σχήμα 7.3. Στο συγκεκριμένο σχήμα φαίνονται οι χάρτες ανομοιότητας βάθους των εικόνων, όπως αυτοί υπολογίστηκαν από την ελαχιστοποίηση της συνάρτησης ενέργειας της Σχέσης (7.1) του προβλήματος πολυοπτικής αντιστοίχισης.

Η διαδικασία αξιολόγησης των αποτελεσμάτων οδήγησε στην υιοθέτηση δύο βασικών μοντέλων. Αρχικά χρησιμοποιήθηκε το RMS σφάλμα ως ποσοτικό μέτρο απόκλισης από τους ιδανικούς χάρτες. Συγκεντρωτικά αποτελέσματα παρουσιάζονται στον πίνακα 7.1. Στο συγκεκριμένο πίνακα παρουσιάζονται συγκριτικά δύο ομάδες αποτελεσμάτων. Στην πρώτη στήλη οι τιμές σφάλματος απόκλισης των χαρτών ανομοιότητας βάθους όπως αυτοί υπολογίστηκαν από πολυοπτική αντιστοίχιση και αξιοποιώντας την προτεινόμενη συνάρτηση ενέργειας δεδομένων. Στη δεύτερη στήλη τα αντίστοιχα αποτελέσματα όπως προκύπτουν από τη διαδικασία ανάκτησης βάθους στην εκδοχή στέρεο αντιστοίχισης, όπως παρουσιάστηκε στο προηγούμενο κεφάλαιο. Η βελτίωση των αποτελεσμάτων στην περίπτωση πολυοπτικής αντιστοίχισης με τη χρήση αφενός του προτεινόμενου μέτρου συσχέτησης, σε συνδυασμό με τη προσθήκη του νέου περιορισμού ορατότητας, επιβεβαιώνεται από τα χαμηλότερα επίπεδα τιμών του RMS σφάλματος.

Η εναλλακτική προσέγγιση στο ζήτημα της αξιολόγησης των αποτελεσμάτων βασίστηκε στη μέτρηση του δείκτη δομικής ποιότητας (SSIM) για κάθε

Πίνακας 7.1: Συγκριτικός πίνακας RMS σφάλματος χαρτών ανομοιότητας βάθους με πολυοπτική αντιστοίχιση, χρησιμοποιώντας την προτεινόμενη μαθηματική έκφραση στη συνάρτηση ενέργειας δεδομένων, *E*<sub>data</sub>

Στέρεο Εικόνες	RMS (πολυοπτική προσέγγιση)	RMS (στέρεο προσέγγιση)
Tsukuba	1.28	1.7
Venus	3.29	4.76
Teddy	6.18	10.15
Aloe	7.03	13.84
Sawtooth	3.54	3.66
Rocks2	6.02	7.07

τιμή κατωφλιοποίησης του αθροιζόμενου κόστους σύγκρισης στη συνάρτηση ενέργειας τήρησης φωτογραφικής συνέπειας. Επαληθεύτηκε η βελτίωση της αξιοπιστίας των αποτελεσμάτων με τη χρήση αφενώς επιπρόσθετης εικόνας, και αφενός της εισαγωγής ενός επιπλέον περιορισμού πλαισίου (συνάρτηση ενέργειας ορατότητας) στη διαδικασία αντιστοίχισης χαρακτηριστικών.

Στο Σχήμα 7.2 παρουσιάζονται τα συγκριτικά γραφήματα του δείκτη δομικής ποιότητας σε συνάρτηση με τις τιμές κατωφλιοποίησης κόστους (τιμές αποκοπής) τόσο στην περίπτωση συσχέτισης στέρεο εικόνων, όσο και στην περίπτωση πολυοπτικού ταιριάσματος διανυσμάτων χαρακτηριστικών. Η ιδανική τιμή του εν λόγω δείκτη είναι η τιμή 1 και αντιστοιχεί ουσιαστικά στην ταύτιση του υπολογιζόμενου χάρτη ανομοιότητας βάθους με τον αντίστοιχο ιδανικό χάρτη. Παρατηρείται η προσέγγιση τιμών του δείκτη σε τιμές εγγύτερα στην ιδανική ίση με ένα για την περίπτωση πολυοπτικού υπολογισμού χαρτών ανομοιότητας.

Επιπρόσθετα επιβεβαιώνεται και η σαφής βελτίωση στις εικόνες ανομοιότητας βάθους μέγιστου δείκτη δομικής ποιότητας, του Σχήματος 7.3 σε σχέση με τους αντίστοιχους του Σχήματος 6.12 οι οποίοι έχουν προκύψει αξιοποιώντας ένα μόνο ζεύγος στέρεο εικόνων.

## 7.4 Συμπεράσματα Κεφαλαίου

Η προσθήκη επιπλέον εικόνων στη διαδικασία ανάκτησης της πληροφορίας βάθους οδηγεί στην αξιοποίηση μεγαλύτερου όγκου πληροφορίας που οδηγεί σε πληρέστερα και βελτιωμένα αποτελέσματα αναφορικά τόσο με τη δομική ποιότητα της εικόνας σύμφωνα με τον αντίστοιχο δείκτη μέτρησης ποιότητας (SSIM), όσο και με ποσοτική μέτρηση της απόκλισης με την εκτίμηση του RMS σφάλματος από τους αντίστοιχους ιδανικούς. Ταυτόχρονα προκύπτει και η ανάγκη αποδοτικής διαχείρισης του όγκου πληροφορίας. Έτσι λοιπόν,



Σχήμα 7.2: Συγκριτικά γραφήματα μεταδολής του δείκτη δομικής ποιότητας σε συνάρτηση με τιμές αποκοπής και κατωφλιοποίησης κόστους σύγκρισης όμοιων χαρακτηριστικών από εικόνα σε εικόνα. Δύο βασικές ομάδες μετρήσεων εξετάζονται συγκριτικά: αρχικά η δομική ποιότητα χαρτών ανομοιότητας βάθους όπως αυτοί έχουν υπολογιστεί ζευγάρι στέρεο εικόνων με συνάρτηση κόστους βασισμένη στο προτεινόμενο μέτρο στερεοσκοπικής συνέπειας και στη συνέχεια το μέτρο ποιότητας για χάρτες ανομοιότητας βάθους από πολυοπτική επεγεργασία αντίστοιχων χαρακτηριστικών από εικόνα σε εικόνα μιας στέρεο ομάδας τριών εικόνων αυξανόμενου βάθους, αξιοποιώντας επίσης το προτεινόμενο μέτρο συσχέτισης σε συνδυασμό με την ελαχιστοποίηση της ενέργειας ορατότητας.







(α') Tsukuba, (truncation value=50, SSIM=0.67)

( $\beta$ ) Aloe, (truncation value=50, SSIM=0.69)

( $\gamma$ ) Rocks, (truncation value=50, SSIM=0.73)



( $\delta$ ) Teddy, (truncation value=50, SSIM=0.48)



(ε) Venus, (truncation value=50, SSIM=0.85)



(F) Sawtooth,(truncation value=50,SSIM=0.73)

Σχήμα 7.3: Χάρτες ανομοιότητας βάθους πειραματικών στέρεο εικόνων Aloe, Tsukuba, Sawtooth, Venus, Map, Rocks και Teddy με μέγιστο δείκτη δομικής ποιότητας SSIM. Οι χάρτες έχουν υπολογιστεί με πολυοπτικό ταίριασμα χαρακτηριστικών σε τρεις διαδοχικές στέρεο εικόνες, αξιοποιώντας την προτεινόμενη συνάρτηση ενέργειας στερεοσκοπικής συνέπειας. εισάγεται στη μαθηματική αναπαράσταση του προβλήματος ένας επιπλέον περιορισμός. Ο περιορισμός αυτός ορίζεται στο σύνολο διαφορετικών τριδιάστατων σημείων με διαφορετικές τιμές βάθους και η προβολή τους είναι ορατή στις διαδοχικές εικόνες. Εισάγεται με τη μορφή μίας συνάρτησης ενέργειας, της οποίας η ελαχιστοποίηση αντιστοιχεί στην ελαχιστοποίηση των περιοχών διαφορετικών τριδιάστατων σημείων με κοινές τιμές βάθους. Οι περιοχές αποκόπτονται από την συνολική ανάλυση ως μη ταυτόχρονα ορατές στις διαδοχικές στέρεο εικόνες. Η διαδικασία αυτή οδηγεί σε ακριβέστερους χάρτες βάθους όπως επιβεβαιώνεται και από τα ποσοτικά μέτρα αξιολόγησης (RMS) σφάλματος και δείκτη δομικής ποιότητας.

# Κεφάλαιο 8

# Σύνοψη Διατριβής

### 8.1 Συμπεράσματα

Η παρούσα διατριβή εντάσσεται στον επιστημονικό κλάδο που σχετίζεται με την αξιοποίηση πληροφορίας από αλληλουχία εικόνων με σκοπό την τριδιάστατη ψηφιακή αποτύπωση των απεικονιζόμενων θεμάτων. Ειδικότερα στα πλαίσια της παρούσας διατριβής επιχειρήθηκε η διεξοδική μελέτη, εφαρμογή καθώς επίσης και η πρόταση καινοτόμων μεθόδων στην κατεύθυνση της ψηφιακής αποτύπωσης και ανακατασκευής πειραματικών εικόνων δανεισμένων από την κοινότητα της υπολογιστικής όρασης υπολογιστών, αλλά και φυσικών εικόνων με θέματα αρχαιολογικού ενδιαφέροντος. Ο τρόπος συλλογής πειραματικών δεδομένων στη δεύτερη περίπτωση βασίστηκε αποκλειστικά στη στερεοσκοπική φωτογράφηση των θεμάτων.

Στο πρώτο στάδιο της ερευνητικής εργασίας της παρούσας διατριβής υπήρξε άμεση η ανάγκη εξαντλητικής μελέτης και βιβλιογραφικής έρευνας των βασικών τεχνικών μοντελοποίησης του συστήματος κάμερας-εικόνας και βαθμονόμησης αυτόματης ή μη της κάμερας. Στα πλαίσια ένταξης της παρούσας διατριβής στο θεωρητικό πλαίσιο του προβλήματος της ψηφιακής τριδιάστατης αποτύπωσης πραγματοποιήθηκε εκτενής έρευνα στην κλασσική αντιμετώπιση του προβλήματος τριδιάστατης ανακατασκευής. Κύρια διαδικασία της ψηφιακής ανακατασκευής αποτελεί η αντιστοίχιση χαρακτηριστικών από εικόνα σε εικόνα. Προς αυτήν την κατεύθυνση στη βιβλιογραφία έχουν προταθεί διάφορες μέθοδοι. Αρχικά συναντώνται μέθοδοι που βασίζοναι στην αρχή της συσχέτισης εικονοστοιχείων γεωμετρικών δομών στην εικόνα όπως οι ακμές. Στη συνέχεια στην παρούσα διατριβή αναλύθηκαν μέθοδοι αντιστοίχισης που βασίζονται στην πυραμιδική αναπαράσταση της πληροφορίας της εικόνας με στόχο τον εντοπισμό αναλλοίωτων χαρακτηριστικών σε μεταβολές της κλίμακας της εικόνας ή σε πιθανούς γεωμετρικούς μετασχηματισμούς της (αλγόριθμος SIFT). Τέλος, αναπτύχθηκαν και υλοποιήθηκαν τεχνικές αντιστοίχισης και εντοπισμού χαρακτηριστικών που βασίζονται σε μέτρα συσχέτισης φάσης. Η συγκριτική πειραματική μελέτη που ακολούθησε με φυσικές εικόνες αρχαιολογικού ενδιαφέροντος με πολύπλοκη απεικονιζόμενη δομή επιβεβαιώνει τη σαφώς αξιόπιστη συμπεριφορά φίλτρων εξαγωγής τοπικής πληροφορίας στο χώρο της φάσης.

Βασική συνεισφορά της παρούσας διατριδής προς αυτήν την κατεύθυνση αποτελεί η εισαγωγή ενός τροποποιημένου μέτρου συσχέτισης. Το μέτρο αυτό βασίζεται στην αναπαράσταση της τοπικής πληροφορίας αξιοποιώντας το διδιάστατο αναλυτικό σήμα και κατ' επέκταση το μονογενές σήμα. Επιπρόσθετα, προτείνεται μία ολοκληρωμένη μέθοδος μοντελοποίησης και τριδιάστατης αποτύπωσης από το στάδιο της φωτογράφησης έως τον υπολογισμό σχετικών ή απόλυτων τιμών βάθους του απεικονιζομένου θέματος. Η εν λόγω μέθοδος/διαδικασία περιλαμβάνει τα στάδια βαθμονόμησης (camera calibration), ορθοποίησης εικόνων (image rectification) και ανακατασκευής βάθους πεδίου ( 3D reconstrucion).

Βασική διαπίστωση που προκύπτει από τη συγκριτική μελέτη του τροποποιημένου μέτρου σε σχέση με τη κλασική προσέγγιση συσχέτισης τιμών φωτεινότητας αποτελεί η αισθητή ποιοτική βελτίωση της τριδιάστατης ανακατασκευής και αναπαράστασης γεωμετρικών δομών της εικόνας. Το γεγονός αυτό αποδεικνύεται με το βαθμό πληρότητας των τριδιάστατων μοντέλων στη αποτύπωση του τριδιάστατου βάθους από διαφορετικές γωνίες παρατήρησης. Αυτό όμως δεν είναι αρκετό αφενός γιατί με την αρχική επεξεργασία των εικόνων εντοπίζονται στην καλύτερη περίπτωση μη θορυβώδεις ακμές στις οποίες εφαρμόζεται μεμονωμένα η εικονική ανακατασκευή τους στον τριδιάστατο χώρο. Το επακόλουθο τριδιάστατο μοντέλο όμως εξακολουθεί να είναι αποσπασματικό και ελλιπές χωρίς σαφή εικόνα της συνολικής δομής του θέματος. Αφετέρου εξακολουθούν να υπάρχουν προβλήματα στη διαδικασία απόφασης του βέλτιστου ταιριάσματος διαδοχικών σημείων, όπως για παράδειγμα αυτά που ανήκουν στην ίδια ακμή. Οι τιμές των μέτρων συσχέτισης είναι αρκετά κοντινές σε γειτονικά σημεία με συνέπεια η επιλογή απλά της μέγιστης τιμής μέτρου δε συνιστά πάντα την αξιόπιστη απόφαση.

Στη συνέχεια στο δεύτερο μέρος της παρούσας διατριβής εφαρμόστηκαν σύγχρονες μέθοδοι βέλτιστης αναπαράστασης του προβλήματος με συναρτήσεις ενέργειας σφαλμάτων, βασισμένες στο θεωρητικό πλαίσιο των Μαρκοβιανών Τυχαίων Πεδίων και των Μπεϋζιανών μοντέλων απόφασης. Η ελαχιστοποίηση των συναρτήσεων αυτών εξασφαλίζει επιτυχή και βέλτιστη στερεοσκοπική αντιστοίχιση. Σε αυτό το σημείο, η κύρια συμβολή της παρούσας διατριβής συνίσταται στην εισαγωγή ενός νέου μέτρου ομοιότητας εικονοστοιχείων στη συνάρτηση ενέργειας που αντιπροσωπεύει τον περιορισμό φωτογραφικής συνέπειας. Η συγκεκριμένη συνάρτηση μαζί με τη συνάρτηση ενέργειας που αντιπροσωπεύει την απαίτηση για τμηματικά ομαλή λύση, διαμορφώνουν τη συνάρτηση ενέργειας του προβλήματος στερεοσκοπικής αντιστοίχισης. Η αντιστοίχιση σύμφωνα με τη σύγχρονη θεώρηση πραγματοποιείται σε όλη την επιφάνεια της εικόνας και δεν επικεντρώνεται στην αντιστοίχηση μεμονωμένων γεωμετρικών δομών της εικόνας, όπως για παράδειγμα σε ακμές.

Η ελαχιστοποίηση της εν λόγω συνάρτησης ενέργειας αντιστοιχεί σε βέλτιστους χάρτες ανομοιότητας βάθους. Η διαδικασία ελαχιστοποίησης βασίστηκε στην αναπαράσταση του προβλήματος μέσω ελάχιστων αποκοπών γράφων. Στην παρούσα διατριβή, κατά τη διαδικασία αξιολόγησης των χαρτών ανομοιότητας βάθους συγκριτικά με τους αντίστοιχους ιδανικούς, προτείνεται η προσαρμογή μέτρων υπολογισμού σφάλματος που αναπαριστούν την ανθρώπινη αντιληπτικότητα στην αξιολόγηση της εικόνας.

Προς αυτήν την κατεύθυνση, μελετήθηκε εκτενώς ο δείκτης δομικής ομοιότητας (Structure SIMilarity (SSIM) index), ο οποίος έχει δειχτεί ότι αντικατοπτρίζει καλύτερα αντιληπτικά χαρακτηριστικά της ανθρώπινης όρασης και μεγιστοποιείται σε εικόνες με σαφή αποκατάσταση της δομής των απεικονιζόμενων αντικειμένων. Στα πειράματα που πραγματοποιήθηκαν, ο εν λόγω δείκτης ενσωματώθηκε ως μέτρο ποιότητας των αποτελεσμάτων. Η δομική ποιότητα των εικόνων αξιολογήθηκε με τη χρήση του δείκτη δομικής ομοιότητας των υπολογιζόμενων χαρτών ανομοιότητας σε σχέση με τους αντίστοιχους ιδανικούς (ground truth maps). Η συγκριτική μελέτη και αξιολόγηση των χαρτών ανομοιότητας βάθους, επιδεδαίωσε τόσο την ποιοτική και ποσοτική (μείωση σφάλματος απόκλισης RMS) βελτίωση των αποτελεσμάτων με τη χρήση της προτεινόμενης συνάρτηση στερεοσκοπικής συνέπειας, όσο και τη βελτίωση της δομικής ποιότητας των εικόνων ανομοιότητας βάθους.

Η πολυοπτική αντιστοίχιση αποτελεί ένα επιπλέον πρόβλημα που μελετήθηκε και αξιολογήθηκε στην παρούσα διατριβή, προτείνοντας παράλληλα την εισαγωγή της τροποποιημένης συνάρτησης ενέργειας στερεοσκοπικής συνέπειας μαζί με ένα επιπλέον περιορισμό πλαισίου στην τελική μορφή συνάρτησης ενέργειας. Αυτός ο περιορισμός ονομάζεται περιορισμός ορατότητας και η ελαχιστοποίησή του οδηγεί στην εξαίρεση από την ανάλυση αλληλοκαλυπτόμενων περιοχών από εικόνα σε εικόνα. Η περαιτέρω βελτίωση στην ακρίβεια των χαρτών ανομοιότητας βάθους επιβεβαιώθηκε με τα χαμηλότερα επίπεδα σφάλματος (RMS), είτε με τον υπολογισμό του δείκτη δομικής ομοιότητας, ο οποίος όπως προσαρμόστηκε στην παρούσα διατριβή θεωρείται ως ένα μέτρο αξιολόγησης ποιότητας. Η βελτίωση των αποτελεσμάτων με τη χρήση της προτεινόμενης συνάρτησης ενέργειας, αποδεικνύεται με το μικρότερο ποσοστό απόκλισης του δείκτη από την ιδανική τιμή του για την περίπτωση υπολογισμών στον ιδανικό χάρτη ανομοιότητας βάθους (ground truth).

### 8.2 Μελλοντικές Κατευθύνσεις

Με δεδομένη την αξιόπιστη συμπεριφορά σε προβλήματα ανακατασκευής, την αναγωγή των υπολογισμών στο χώρο της φάσης, την πολυοπτική επέκταση της αντιστοίχισης και την αξιοποίηση των μονογενών σημάτων, η έρευνα μας εξελίσσεται στις ακόλουθες κατευθύνσεις:

#### 1. Μέθοδοι αντιστοίχισης με στατιστικά μοντέλα

Έχοντας μελετήσει ιδέες και προτάσεις από τα [5], [3], [4] και [63] που σχετίζονται με υπολογισμούς οπτικής ροής και αναγνώριση προτύπων με στατιστικά μοντέλα, αλυσίδες Markov, θα επιχειρηθεί η επέκταση των συγκεκριμένων ιδεών στο παρών πρόβλημα.

#### 2. Πολυοπτική ανακατασκευή

Μελετώντας αλγόριθμους πολυοπτικής ανακατασκευής από τα [64] και [44] θα επιχειρηθεί καταρχήν η βελτίωση της ήδη υπάρχουσας ποιότητας και ακρίβειας των χαρτών ανομοιότητας βάθους, αξιοποιώντας δεδομένα είτε από τη γεωμετρία της σκηνής και τη γεωμετρική αναπαράσταση της κάμερας, είτε με περισσότερο αποδοτικές υπολογιστικά μεθόδους πολυοπτικής αντιστοίχισης από μεγαλύτερο πλήθος εικόνων. Κατά δεύτερον θα επικεντρωθεί η περαιτέρω έρευνα στη συγκριτική έρευνα μεθόδων και αλγορίθμων βελτιστοποίησης των συναρτήσεων ενέργειας του προβλήματος.

#### 3. Αξιολόγηση ποιότητας αποτελεσμάτων

Η υιοθέτηση του μέσου τετραγωνικού σφάλματος απόκλισης (RMS) από τους ιδανικούς χάρτες ανομοιότητας (ground truth) ως αποκλειστικό μέτρου αξιολόγησης της ποιότητας των αποτελεσμάτων πολλές φορές οδηγεί σε εσφαλμένα συμπεράσματα. Εικόνες με κοινό RMS λάθος παρουσιάζουν αισθητές διαφορές στην ανάκτηση της δομικής πληροφορίας. Συνεπώς μία επιπλέον κατεύθυνση της έρευνας μας αποτελεί η περαιτέρω αναζήτηση αξιόπιστων μέτρων αξιολόγησης της ποιότητας της αποκατεστημένης δομικής πληροφορίας της σκηνής.

#### 4. Μοντέλα μερικών διαφορικών εξισώσεων PDEs

Τέλος ένα κομμάτι της έρευνάς μας θα κατευθυνθεί σε αναπαράσταση του προβλήματος με μοντέλα μερικών διαφορικών εξισώσεων PDEs, [35], για την προσέγγιση ενός δεδομένου νέφους τριδιάστατων σημείων όπως προκύπτει από τη στερεοσκοπική συσχέτιση πολλαπλών πλαισίων εικόνας, με γραφική αναπαράσταση συνεχών επιφανειών. Η συγκεκριμένη προσέγγιση θα μας οδηγήσει στην απόκτηση ενιαίων, συνεχών μοντέλων με παράλληλη δυνατότητα υλοποίησης τεχνικών φωτισμού και απεικόνισης που θα οδηγήσουν σε μοντέλα ρεαλιστικής απεικόνισης. Η χρήση PDEs συντελεί επίσης, στη μοντελοποίηση ελλιπών περιοχών αξιοποιώντας πληροφορίες πλαισίου της εικόνας και στατιστικών μεθόδων απόφασης [47], με χρήση πρότερης γνώσης σχήματος σε συνδυασμό με πιθανές υποθέσεις αρχιτεκτονικής μελέτης για απεικονιζόμενα θέματα αρχαιολογικού ενδιαφέροντος.

## Βιβλιογραφία

- [1] L. Falkenhagen. Hierarchical Block-Based Disparity Estimation Considering Neighbourhood Constraints. In *Proc. International Workshop on SNHC and 3D Imaging*, 1997.
- [2] A. Blake and A. Zisserman. Visual Reconstruction. MIT press, 1987.
- [3] A. S. Ogale and Y. Aloimonos. A roadmap to the integration of early visual modules. *International Journal of Computer Vision: Special Issue on Early Cognitive Vision, in press*, 65(1), 2005.
- [4] A. S. Ogale and Y. Aloimonos. Robust contrast invariant stereo correspondence. In *Proc. IEEE Conf. on Robotics and Automation (ICRA)*, 2005.
- [5] A. S. Ogale and Y. Aloimonos. Shape and the stereo correspondence problem. *International Journal of Computer Vision*, 65(1), 2005.
- [6] Alifragis M. and Tzafestas C. S. Stereo pair matching of archaeological scenes using phase domain methods. In *International Joint Conference on Computer Vision Imaging and Computer Graphics Theory and Applications (VISIGRAPP'09)*, Lisbon, 2009.
- [7] Alifragis M. and Tzafestas C. S. Stereo Analysis of Archaelogical Scenes using Monogenic Signal Representation. In Computer Vision, Imaging and Computer Graphics. Theory and Applications. Series: Communications in Computer and Information Science, Springer Berlin Heidelberg, 2010, Volume 68, Part 3, 133–145, 2009.
- [8] A. Amini, T. Weymouth, and R. Jain. Using dynamic programming for solving variational problems in vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(9):855–867, 1990.
- [9] P. Horn and B. Schunk. Determining optical flow. *Artificial Intelligence*, 17:185–203, 1981.

- [10] M. Brown and D. Lowe. Invariant features from interest point groups. In *British Machine Vision Conference*, 2002.
- [11] C. Harris and M. Stephens. A combined corner and edge detector. In Fourth Alvey Vision Conference, pages 147–151, 1988.
- [12] C. Schmid and A. Zisserman. Automatic Line Matching across Views. In IEEE Conference on Computer Vision and Pattern Recognition, IEEE Computer Soc. Press, pp 666–671, 1997.
- [13] C. Schmid, R. Mohr and C. Bauckhage. Comparing and Evaluating Interest Points. In International Conference on Computer Vision, Narosa Publishing House, pp 230–235, 1998.
- [14] C. Tomasi and T. Kanade. Shape and motion from image streams under orthography: A factorization approach. *International Journal of Computer Vision*, 9(2):137–154, 1992.
- [15] F. Canny A computational approach to edge detection. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 8:112–131, 1986.
- [16] C. K. Chow A recognition method using neighbor dependence. IRE Transactions on Electronic Computer, 11:683–690, 1962.
- [17] D. Geman, S. Geman, C. Graffigne, and P. Dong. Boundary detection by constrained optimization. *IEEE Transactions on Pattern Analysis* and Machine Intelligence, 12(7):609[628, 1990.
- [18] D. Greig, B. Porteous, and A. Seheult. Exact maximum a posteriori estimation for binary images. *Journal of the Royal Statistical Society*, Series B, 51(2):271–279, 1989.
- [19] D. J. Field, A. D. Jepson and M. Jenkin. Phase-based disparity measurement. In Z. Pylyshyn editor, Computational Processes in human Vision., 1988.
- [20] D. Marr and T. Poggio. A Computational Theory of Human Stereo Vision. In *Proc. Royal Society of London*, volume 204 of B, pages 301–328, 1979.
- [21] D. G. Lowe. Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision, 60:91–110, 2004.

- [22] D. Lee and T. Pavlidis. One dimensional regularization with discontinuities. *Transactions on Pattern Analysis and Machine Intelligence*, 10(6):822–829, 1988.
- [23] D. Terzopoulos. Regularization of inverse visual problems involving discontinuities. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(4):413-424, 1986.
- [24] S. Edelman, N. Intrator and T. Poggio Complex cells and object recognition. In *Neural Information Processing Systems (NIPS) Foundation*, 1997.
- [25] D. J. Field Relations between the statistics of natural images and the response properties of cortical cells. *Journal of The Optical Society of America A*, 4(12):2379–2394., December 1987.
- [26] F. Worgotter. Multi-modal estimation of collinearity and parallelism in natural image sequences. In *Natural Image Sequences. Network: Computation in Neural Systems*, pages 553–576, 2001.
- [27] G. Golub and C. van Loan. *Matrix Computations*. John Hopkins University Press, 1996.
- [28] Guo-Qing Wei and Song De Ma. Implicit and explicit camera calibration:Theory and experiments. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(5), 1994.
- [29] H. Ishikawa. Exact Optimization for Markov Random Fields with Convex Priors. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 25(10):1333–1336, Oct 2003.
- [30] H. Ishikawa and D. Geiger. Occlusions, discontinuities, and epipolar lines in stereo. In *European Conference on Computer Vision*, 1998.
- [31] C. Harris, and M. Stephens A combined corner and edge detector. In Fourth Alvey Vision Conference, pages 147–151, Manchester, UK, 1988.
- [32] I. Cox, S. Hingorani and S. Rao. A Maximum Likelihood Stereo Algorithm. *Computer Vision and Image Understanding*, 63(3), 1996.
- [33] Iasonas Kokkinos and Georgios Evangelopoulos and Petros Maragos. Texture Analysis and Segmentation Using Modulation Features, Generative Models and Weighted Curve Evolution. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(1), pp. 142–157, 2009.

- [34] J. Morlet and G. Arens and E. Fourgeau and D. Giard. Wave propagation and sampling theory - Part II: Sampling theory and complex waves. *Geophysics*, 47:222–236, February 1982.
- [35] J.A. Sethian. Level Set Methods and Fast Marching Methods: Evolving Interfaces in Computation Geometry, Fluid Mechanics, Computer Vision and Material Science. Cambridge University Press, 2nd edition, 1999.
- [36] J.J. Koenderink The structure of images. *Biological Cybernetics*, 50:363–396, 1984.
- [37] T. Lindeberg Scale-space theory: A basic tool for analysing structures at different scales. *Journal of Applied Statistics*, 21(2):224–270, 1994.
- [38] M. C. Morrone and R. A. Owens. Feature detection from local energy. *Patern Recognition Letters*,vol 6, 1987.
- [39] M. Felsberg and G. Sommer. Processing for Estimating Local Properties and Detecting Features. In *DAGM Symposium 2000, Kiel*, 1997.
- [40] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active contour models. *International Journal of Computer Vision*, 1(4):321–331, 1987.
- [41] M. Okutomi and T. Kanade. A Locally Adaptive Window for Signal Processing. International Journal of Computer Vision, 7:143–162, 1992.
- [42] M. Pelillo. The dynamics of nonlinear relaxation labeling processes. Journal of Mathematical Imaging and Vision, 7:309–323, 1997.
- [43] M. Pollefeys. Self-Calibration and Metric 3D Reconstruction from Uncalibrated Images Sequences. PhD thesis, Katholieke Universiteit Leuven, 1999.
- [44] M. Pollefeys and R. Koch and L. Van Gool. Self-Calibration and Metric Reconstruction in spite of Varying and Unknown Internal Camera Parameters. *International Journal of Computer Vision*, 32(1):7–25, 1999.
- [45] M. Spetsakis and J. Aloimonos. Structure from motion using line correspondences. *International Journal of Computer Vision*, 4(3):171– 183, 1990.
- [46] Michael Felsberg and Gerald Sommer. The monogenic signal. *IEEE Transactions on Signal Processing*, 49(12):3136–3144, 2001.

- [47] N. Paragios and M. Rousson. Shape priors for level set representation. 2001.
- [48] N. Kruger and M. Felsberg and C. Gebken and M. Porksen. An Explicit and Compact Coding of Geometric and Structural Information Applied to Stereo Processing.
- [49] O. Faugeras. Three-Dimensional Computer Vision: a Geometric Viewpoint. MIT press, 1993.
- [50] O. Veksler. Efficient Graph-based Energy Minimization Methods in Computer Vision, PhD thesis. Cornell University, July 1999.
- [51] T. Pavlidis. A critical survey of image analysis methods. In *International Conference of Pattern Recognition*, 1986.
- [52] P.B. Chou and C.M. Brown. The theory and practice of Bayesian image labeling. *International Journal of Computer Vision*, 4(3):185– 210, 1990.
- [53] P.D. Kovesi. Invariant Measures of Image Features from Phase Information. PhD thesis, The University of Western Australia, 1996.
- [54] P. Kovesi. Symmetry and Asymmetry From Local Phase. In Al'97, Tenth Australian Joint Conference on Artificial Intelligence, 1997.
- [55] P. Kovesi. Image Features From Phase Congruency. Videre: A Journal of Computer Vision Research. MIT Press, 1(3), 1999.
- [56] P. Kovesi. Phase Congruency Detects Corners and Edges. In *The Australian Pattern Recognition Society Conference: DICTA 2003*, 2003.
- [57] Q.T. Luong and O. Faugeras. Self Calibration of a moving camera from point correspondenses and fundamental matrices. *International Journal of Computer Vision*, 22(3):261–289, 1997.
- [58] R. Ahuja, O. Ergun, J. Orlin, and A. Punnen. A Survey of Very Large-Scale Neighborhood Search Techniques. *Discrete Applied Math*, 123:75–102, 2002.
- [59] R. Deriche and G. Giraudon, 1(2):167-187. A computational approach for corner and vertex detection. *International Journal of Computer Vision*, 1(2):167-187, 1993.

- [60] R. Deriche and O. Faugeras. Tracking Line Segments. Computer Vision-ECCV'90, Lecture Notes in Computer Science, Springer-Verlag, vol 427, 1990.
- [61] R. Deriche. Optimal edge detection using recursive filtering. pages 501–505, 1987.
- [62] R. Hartley and A. Zisserman. *Multiview Geometry in Computer Vision*. Cambridge University Press, 2000.
- [63] R. Hartley and A. Zisserman. Multiview Geometry in computer vision. International Journal of Computer Vision, 70(1), 2000.
- [64] R. Koch and M. Pollefeys and L. Van Gool. 3D Surface Reconstruction from Uncalibrated Image Sequences. in B, Jahne and H. Haussbecker, 2000.
- [65] R. Koch, M. Pollefeys and L. Van Gool. Multi Viewpoint Stereo from Uncalibrated Video Sequences. In Proc. European Conference on Computer Vision, παγες 55–71, Freiburg, Germany, 1998.
- [66] R. Potts. Some generalized order-disorder transformation. In Proceedings of the Cambridge Philosophical Society, vol 48, παγε 106–109, 1952.
- [67] R. H. Swendson and J. Wang. Nonuniversal critical dynamics in monte carlo simulations. *Physical Review Letters*, 58(2):86–88, 1987.
- [68] R. I. Hartley. In Defence of the eight-point algorithm. IEEE Transactions on Pattern Analysis and Machine Intelligence, 19(6):580–593, 1997.
- [69] Richard Szeliski, Ramin Zabih, Daniel Scharstein, Olga Veksler, Vladimir Kolmogorov, Aseem Agarwala, Marshall Tappen, Carsten Rother. A Comparative Study of Energy Minimization Methods for Markov Random Fields with Smoothness-Based Priors. *IEEE Transactions* on Pattern Analysis and Machine Intelligence, 30(6):1068–1080, June 2008.
- [70] R.S. Szeliski. Bayesian modeling of uncertainty in low-level vision. *International Journal of Computer Vision*, 5(3):271–302, December 1990.
- [71] S. Geman and D. Geman. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6:721–741, 1984.

- [72] S. Li. Markov Random Field Modeling in Computer Vision. Springer, 1995.
- [73] S. Pollard, J. Mayhew and J. Frisby. PMF: A Stereo Correspondence Algorithm Using a Disparity Gradient Limit. *Perception*, 14(4):449– 470, 1985.
- [74] S. Roy and I. Cox. A maximum-flow formulation of the n-camera stereo correspondence problem. In *International Conference on Computer Vision*, 1998.
- [75] S.A. Barker and P.J.W. Rayner. Unsupervised image segmentation. In IEEE International Conference on Acoustics, Speech and Signal Processing, 1998.
- [76] S. Z. Li. *Markov random field modeling in image analysis*. Springer-Verlag New York, Inc., 2001.
- [77] T. Poggio, V. Torre, and C. Koch. Computational vision and regularization theory. *Nature*, 317:314–319, 1985.
- [78] V. Kolmogorov and R. Zabih. What Energy Functions can be Minimized via Graph Cuts. IEEE Trans. Pattern Analysis and Machine Intelligence, 26(2):147–159, Feb. 2004.
- [79] V. Kolmogorov and R. Zabih. Multi-camera Scene Reconstruction via Graph Cuts. In European Conference on Computer Vision, pages 82– 96, 2002.
- [80] A.P Witkin. Scale-space filtering. In International Joint Conference on Artificial Intelligence, 1983.
- [81] Y. Boykov and V. Kolmogorov. An Experimental Comparison of Min-Cut/Max-Flow Algorithms for Energy Minimization in Vision. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 26(9):1124–1137, Sept. 2004.
- [82] Y. Ohta and T. Kanade. Stereo by Intra- and Inter-scanline Search Using Dynamic Programming. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 7(2):139–154, 1985.
- [83] Yuri Boykov and Vladimir Kolmogorov. An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. In International Workshop on Energy Minimization Methods in Computer Vision and Pattern Recognition (EMMCVPR), 2001.

- [84] Yuri Boykov, Olga Veksler, and Ramin Zabih. Markov random fields with efficient approximations. In *IEEE Conference on Computer Vision and Pattern Recognition*, 1998.
- [85] Yuri Boykov, Olga Veksler, and Ramin Zabih. Markov random fields with efficient approximations. In *IEEE Conference on Computer Vision and Pattern Recognition*, 1998.
- [86] Yuri Boykov, Olga Veksler and Ramin Zabih. Fast Approximate Energy Minimization via Graph Cuts. *IEEE Transactions on PAMI*, 23:1222– 1239, 2001.
- [87] Z. Zhang. A flexible new technique for camera calibration. In International Conference on Computer Vision, Kerkyra, September 1999.
- [88] Zhou Wang, Member, IEEE, Alan C. Bovik, Fellow, IEEE, Hamid R. Sheikh, Student Member, IEEE, and Eero P. Simoncelli, Member, IEE-E. Image Quality Assessment: From Error Measurement to Structural Similarity. *IEEE Transactions on Image Processing*, 13(1), 2004.