



Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών
Τομέας Σημάτων, Ελέγχου και Ρομποτικής

**Αναγνώριση Χειρονομιών της Αλφαβήτου της
Ελληνικής Νοηματικής Γλώσσας με Έμφαση
στην Εξαγωγή Περιστροφικά Ανεξάρτητων
Οπτικών Χαρακτηριστικών**

Διπλωματική Εργασία

Γεώργιος Κ. Ρετσινάς

Επιβλέπων : Πέτρος Μαραγκός
Καθηγητής Ε.Μ.Π.

Αθήνα,
Μάρτιος 2014



Εθνικό Μετσόβιο Πολυτεχνείο

Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών

Τομέας Σημάτων, Ελέγχου και Ρομποτικής

**Αναγνώριση Χειρονομιών της Αλφαβήτου της
Ελληνικής Νοηματικής Γλώσσας με Έμφαση στην
Εξαγωγή Περιτροφικά Ανεξάρτητων Οπτικών
Χαρακτηριστικών**

Διπλωματική Εργασία

Γεώργιος Κ. Ρετσινάς

Επιβλέπων : Πέτρος Μαραγκός
Καθηγητής Ε.Μ.Π.

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 18η Μαρτίου 2013.

.....
Πέτρος Μαραγκός
Καθηγητής Ε.Μ.Π.

.....
Κωνσταντίνος Τζαφέστας
Επίκουρος Καθηγητής
Ε.Μ.Π.

.....
Γεράσιμος Ποταμάνος
Αναπληρωτής Καθηγητής
Παν/μίου Θεσσαλίας

Αθήνα,
Μάρτιος 2014

.....

Γεώργιος Κ. Ρετσινάς

Διπλωματούχος Ηλεκτρολόγος Μηχανικός και Μηχανικών Υπολογιστών Ε.Μ.Π.

Copyright © Γεώργιος Κ. Ρετσινάς, 2013.

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 τη χρήση της εργασίας για κερδοσκοπικό σκοπό πρέπει να απευθύνονται προς τον συγγραφέα.

Ευχαριστίες

Θα ήθελα καταρχήν να ευχαριστήσω τον κ. Πέτρο Μαραγκό και τον Σταύρο Θεοδωράκη για την επίβλεψη της διπλωματικής μου εργασίας και την παραχώρηση του Kinect για την συλλογή πειραματικών δεδομένων. Επιπλέον θα ήθελα να ευχαριστήσω θερμά τους συμφοιτητές μου Βαγγέλη, Γιώργο, Γρηγόρη και Μάγια που δημιούργησαν την βάση δεδομένων με την οποία εργάστηκα και έδωσαν ώθηση στο έργο μου. Τέλος, θα ήθελα να ευχαριστήσω τους γονείς μου για την στήριξη που μου προσέφεραν όλα αυτά τα χρόνια.

Περίληψη

Η παρούσα διπλωματική επικεντρώνεται στις μεθόδους εξαγωγής οπτικών χαρακτηριστικών στα πλαίσια του προβλήματος της αναγνώρισης χειρονομιών της αλφαβήτου της Ελληνικής Νοηματικής Γλώσσας για στατικές χειρομορφές (fingerspelling). Συγκεκριμένα γίνεται εκτενής μελέτη διαφορετικών προσεγγίσεων για την εξαγωγή χαρακτηριστικών με ιδιαίτερη έμφαση στην μέθοδο HOG (Histogram of Oriented Gradients), η οποία ουσιαστικά αποτελεί μια πυκνή κατανομή τοπικών περιγραφών (ένα ιστογράμματα κατευθύνσεων για κάθε υποπεριοχή). Για την καλύτερη αξιοποίηση της υπάρχουσας οπτικής πληροφορίας προτείνονται διάφορες παραλλαγές των HOG ως προς την κατεύθυνση της βελτιστοποίησης της χωρικής κατανομής των τοπικών περιγραφών. Επιπλέον, παρουσιάζεται μια απλοποίηση των HOG και των προτεινόμενων παραλλαγών του για την εφαρμογή τους σε δυαδικές εικόνες, μειώνοντας σε μεγάλο βαθμό το υπολογιστικό κόστος και διατηρώντας την αποτελεσματικότητα των HOG περιγραφών. Μεγάλο μέρος της παρούσας μελέτης καταλαμβάνει η εξερεύνηση της δυνατότητας εξαγωγής περιγραφών ανεξάρτητων σε περιστροφές της εικόνας αλλά και της δυνατότητας εκτίμησης της εκάστοτε περιστροφής. Ως προς αυτή την κατεύθυνση αναπτύχθηκαν δυο διαφορετικές προσεγγίσεις με κοινό δομικό στοιχείο την μέθοδο HOG. Η πρώτη βασίζεται σε μια αρχική εκτίμηση της περιστροφής, η οποία βασίζεται στο περίγραμμα της χειρομορφής και αξιοποιεί την μέθοδο Shape Context. Η δεύτερη πρόκειται ουσιαστικά για μια γενίκευση των HOG και αποτελεί μια πιο συμπαγή μέθοδος με πολύ ενθαρρυντικά αποτελέσματα. Τέλος, γίνεται μελέτη της ικανότητας μοντελοποίησης και κατηγοριοποίησης των περιγραφών που προκύπτουν από τις προτεινόμενες μεθόδους με χρήση απλών τεχνικών αναγνώρισης προτύπων.

Λέξεις Κλειδιά

Εξαγωγή Οπτικών Χαρακτηριστικών, Αναγνώριση Προτύπων, fingerspelling, HOG, Περιγραφείς Σχήματος, Shape Context, Περιστροφική Ανεξαρτησία, Εκτίμηση Περιστροφής

Abstract

This thesis focuses on feature extraction techniques regarding the problem of fingerspelling recognition for the Greek Sign Language. In particular, several feature extraction techniques had been explored with emphasis on HOG (Histogram of Oriented Gradients) technique, which extracts dense distributed local histograms of gradient direction. In order to achieve better performance in recognition, some spatial variations of HOG are proposed, which try to make better use of simple shape metrics. In addition, a simplified version of the HOG technique and its variations is proposed for binary images, which reduces computational complexity while maintaining the initial HOG efficiency. Rotation invariant descriptors as well as rotation estimation are studied extensively in this thesis. In this context, two techniques are proposed with HOG as the common base. The first one makes an initial estimation of the rotation based on the shape of the hand with the use of the Shape Context technique. The latter is a generalization of the simple HOG technique and constitutes a more concrete technique with very promising results. Lastly, the efficiency in recognition of the proposed descriptors is examined using popular pattern recognition techniques with emphasis on generative models.

Keywords

Feature Extraction, Pattern Recognition, fingerspelling, HOG, Shape Descriptors, Shape Context, Rotation Invariance, Rotation Estimation

Περιεχόμενα

1	Εισαγωγή	1
1.1	Όραση Υπολογιστών	2
1.2	Αναγνώριση Προτύπων	2
1.2.1	Σύστημα Αναγνώρισης Νοηματικής Γλώσσας	3
1.2.2	Αναγνώριση Στατικών Χειρομορφών - Fingerspelling	4
1.3	Ερευνητικοί Στόχοι - Προσφορά	7
1.3.1	Παραλλαγές HOG	8
1.3.2	Ανεξαρτησία στην Περιστροφή	8
1.3.3	Χώρος Περιγραφών - Αξιολόγηση	8
1.4	Σχετική Έρευνα	8
2	Εξαγωγή Οπτικών Χαρακτηριστικών Σχήματος	11
2.1	Ροπές – Moments	11
2.1.1	Hu Invariant Moments	12
2.2	Fourier Descriptors	12
2.3	Χρήση Εναλλακτικών Περιγραφών του Περιγράμματος	14
2.4	Shape Context	15
3	Εξαγωγή Οπτικών Χαρακτηριστικών Εμφάνισης	21
3.1	HOG	22
3.2	SIFT	24
3.3	Χειρομορφές ως Αντικείμενα Αναγνώρισης	27
4	Τεχνικές Αναγνώρισης Προτύπων - Ταξινομητές	29
4.1	Μείωση Αριθμού Διαστάσεων	29
4.1.1	Ανάλυση σε Κύριες Συνιστώσες - PCA	30
4.1.2	Ανάλυση Γραμμικής Διάκρισης - LDA	31
4.1.3	Σύνοψη - Σύγκριση	33
4.2	Ταξινομητές	35
4.2.1	Θεωρία Αποφάσεων του Bayes	35
4.3	Ταξινομητές Ελάχιστης Απόστασης	36
4.3.1	Naive Bayes Ταξινομητής	37
4.3.2	Παραμετρική Εκτίμηση - Εκτίμηση Μέγιστης Πιθανοφάνειας	38
4.3.3	Ταξινομητής Πλησιέστερων Γειτόνων	38

5	Προτεινόμενες Παραλλαγές των HOG	39
5.1	Centralized HOG	40
5.2	Κυκλική Χωρική Διαμέριση - RHOG	43
5.3	Ελλειπτική Χωρική Διαμέριση - ERHOG	44
5.4	Binary HOG (BHOG) - Εφαρμογή των HOG σε δυαδικές εικόνες	45
5.5	Προσαρμογή των Παραλλαγών των HOG για Δυαδικές Εικόνες	48
6	Δημιουργία Περιστροφικά Ανεξάρτητων Περιγραφών βασισμένων σε HOG	49
6.1	Μια Πρώτη Προσέγγιση για Ανεξαρτησία στην Περιστροφή	49
6.1.1	Παραλλαγή της Μεθόδου Shape Context	50
6.1.1.1	Δημιουργία Ανεξάρτητου Διανύσματος Χαρακτηριστικών	53
6.1.2	Χρήση Shape Context μαζί με HOG	53
6.2	Περιστροφικά Ανεξάρτητα HOG	57
6.2.1	Κυκλική Διαμέριση του Χώρου	58
6.2.2	Ανεξαρτησία Τοπικών Περιγραφών ως προς το Κέντρο της Εικόνας	59
6.2.3	Εύρεση Περιστροφής και Δημιουργία Ενιαίων Ανεξάρτητων Περιγραφών	68
7	Πειραματικό Μέρος	71
7.1	Περιγραφή της Βάσης Δεδομένων	71
7.2	Επιλογή Συνόλου Εκπαίδευσης	74
7.3	Εξαγωγή Χαρακτηριστικών	76
7.4	PCA & LDA	77
7.5	Ταξινομητές	79
7.6	Ο Χώρος των HOG Περιγραφών	81
7.7	Επιλογή Βέλτιστων Παραμέτρων	84
7.7.1	HOG	86
7.7.2	BHOG	89
7.7.3	CHOG & CBHOG	90
7.7.4	RHOG , ERHOG , RBHOG & ERBHOG	92
7.7.5	Παραλλαγή των Shape Context	97
7.8	Σύγκριση των Τεχνικών που Παρουσιάστηκαν	98
7.8.1	Σύγκριση Τεχνικών Περιγράμματος	98
7.8.2	Σύγκριση Τεχνικών Εμφάνισης Βασισμένων σε HOG	98
7.8.3	Συνολική Σύγκριση	101
7.9	Τεχνικές για Ανεξαρτησία στην Περιστροφή	102
7.9.1	Ακρίβεια Εκτίμησης Περιστροφής	102
7.9.2	RBHOG & ERBHOG	105
7.9.3	RHOG & ERHOG	106
7.9.4	Μέθοδος Περιγραμμάτων Αναφοράς	109
7.9.5	Σύγκριση Τεχνικών Περιγράμματος	110
7.9.6	Σύγκριση Τεχνικών Εμφάνισης	114
7.9.7	Αποτελεσματικότητα Αναγνώρισης - Συμπεράσματα	117

8	Συμπεράσματα και Κατευθύνσεις για Μελλοντική Έρευνα	119
8.1	Ανακεφαλαίωση και Συμπεράσματα	119
8.2	Μελλοντική Έρευνα	121
8.2.1	Αναγνώριση Αντικειμένων	121
8.2.2	Αναγνώριση Χειρονομιών	122
A'	Hu Invariant Moments	123
B'	Επιλογή Κέντρου Χειρομορφής	125
Γ'	Περιστροφική Ανεξαρτησία	129
Γ'.1	Ετεροσυσχέτιση (Cross-Correlation)	129
Γ'.2	Fourier Transform Magnitude	130

Κατάλογος σχημάτων

1.1	Χειρομορφές των γραμμάτων της ελληνικής νοηματικής γλώσσας . . .	5
1.2	Προβολή του σκελετού που παράγει το Kinect πάνω στην RGB εικόνα και στην εικόνα Βάθους	6
1.3	Εξαγωγή της περιοχής της χειρομορφής με χρήση της πληροφορίας βάθους	6
1.4	Σύνοψη συστήματος αναγνώρισης	7
2.1	Παράδειγμα εξομάλυνσης με χρήση K Fourier συντελεστών. Παρατηρούμε ότι στην περίπτωση $K = 40$ το προκύπτον περίγραμμα είναι σχεδόν ίδιο με το αρχικό περιέχοντας και μεγάλο ποσοστό του θορύβου που έχει το αρχικό περίγραμμα	13
2.2	Κυκλική Διαμέριση του Χώρου	15
2.3	Παράδειγμα Logpolar διαμέρισης του χώρου για ένα σημείο του περιγράμματος για την κατασκευή του αντιστοίχου ιστογράμματος κατανομής των σημείων του περιγράμματος σε σχέση με το σημείο αναφοράς	16
2.4	Ταίριασμα των σημείων της περιφέρειας των δυο χειρομορφών σύμφωνα με την μέθοδο των shape context και της μεθόδου hungarian . . .	17
2.5	Οπτικοποίηση του αφινικού μετασχηματισμού της δεύτερης εικόνας ως προς την πρώτη και υπερτιθέμενη σε αυτή. Ο αφινικός μετασχηματισμός έχει υπολογιστεί σύμφωνα με το ταίριασμα των σημείων της εικόνας 2.4	18
3.1	Οπτικοποίηση της HOG περιγραφής για κάθε κελί	24
3.2	Παράδειγμα εφαρμογής των SURF. $(\alpha), (\beta)$: η αρχική εικόνα και νέα που δημιουργήθηκε με περιστροφή και αύξηση της κλίμακας. Στις εικόνες αυτές φαίνονται και τα σημεία ενδιαφέροντος, όπως και η κλίμακα τους που είναι ανάλογη με την ακτίνα του κύκλου, (γ) : υπέρθεση των δυο εικόνων και οπτικοποίηση της αντιστοίχισης των σημείων ενδιαφέροντος	26
3.3	Παράδειγμα ομοιότητας τοπικών χαρακτηριστικών για μια χειρομορφή σε διαφορετικές κλίμακες	27
3.4	Παράδειγμα εφαρμογής των SURF. $(\alpha), (\beta)$: δυο εικόνες χειρομορφών της ίδιας κλάσης (αναπαριστάνουν το γράμμα Α). Στις εικόνες αυτές φαίνονται και τα σημεία ενδιαφέροντος, όπως και η κλίμακα τους που είναι ανάλογη με την ακτίνα του κύκλου, (γ) : το αποτέλεσμα του ταίριασματος των σημείων των δυο εικόνων	28
4.1	PCA - το διάνυσμα προβολής είναι παράλληλο στην πράσινη ευθεία . .	33

4.2	LDA - το διάνυσμα προβολής είναι παράλληλο στην πράσινη ευθεία . . .	34
4.3	Περίπτωση μη κανονικής κατανομής	34
4.4	Περίπτωση διακριτικής πληροφορίας στην διακύμανση	34
5.1	Τα βασικά στάδια της μεθόδου των HOG, σύμφωνα με την περιγραφή στο Κεφάλαιο 3. Τα στάδια αυτά είναι 1)Υπολογισμός της παραγώγου 2) Χωρικός διαχωρισμός σε κελία 3)Δημιουργία ιστογραμμάτων κατευθύνσεων για κάθε κελί και 4)Κανονικοποίηση ιστογραμμάτων με ένα επιπλέον στάδιο συνένωσης των τοπικών ιστογραμμάτων σε μια ενιαία περιγραφή	39
5.2	(α) Παράδειγμα κατασκευής του πλέγματος. (β) Αρχική χωρική διαμέριση για την κατασκευή του νέου πλέγματος.	41
5.3	Σύγκριση της μορφής του πλέγματος. Στις εικόνες (α),(β) εμφανίζεται το πλέγμα της απλής μεθόδου και στις (γ),(δ) το παραμορφωμένο πλέγμα της προτεινόμενης	42
5.4	Κυκλική Διαμέριση	44
5.5	Ελλειψοειδής Διαμέριση	44
5.6	Σύνοψη της μεθόδου BHOG. Στο σχήμα αυτό αναλύουμε το βήμα του υπολογισμού της παραγώγου, έτσι ώστε να γίνει εμφανές με ακρίβεια σε ποιο στάδιο γίνεται χρήση των πινάκων Μέτρου και Κατεύθυνσης. . .	47
5.7	Κυκλική και Ελλειπτική Διαμέριση για Δυαδικές Εικόνες Χειρομορφών .	48
6.1	Παράδειγμα κύκλου αναφοράς (με κόκκινο) για μια εικόνα (το περιγράμμα της σημειώνεται με μπλε) σύμφωνα με την παραλλαγή των shape context	51
6.2	Πίνακας τοπικών συσχετίσεων	53
6.3	Παράδειγμα αφινικού ταιριάσματος. (α),(β) οι αρχικές εικόνες μαζί με τα αποδεκτά σημεία του περιγράμματος (γ) η υπέρθεση της μετασχηματισμένης εικόνας με την εικόνα αναφοράς και η αντιστοίχιση των σημείων τους	54
6.4	Παράδειγμα κυκλικού ταιριάσματος. (α),(β) οι αρχικές εικόνες μαζί με τα σημεία αναφοράς (γ) η υπέρθεση της μετασχηματισμένης εικόνας με την εικόνα αναφοράς και η αντιστοίχιση των σημείων τους	55
6.5	Αναπαράσταση του τελικού συστήματος αναγνώρισης σύμφωνα με την προτεινόμενη μέθοδο	56
6.6	Μετασχηματισμός πλέγματος (σύμφωνα με την αντιστοίχιση στο Σχήμα6.4) . (α) πλέγμα αναφοράς (β) μετασχηματισμένο πλέγμα	57
6.7	Μια εικόνα και η περιστροφή της (45°), οπού επισημαίνεται με έναν μπλε κύκλο το ίδιο κελί	59
6.8	Γεωμετρική απεικόνιση της χρήσης του ακτινικού συστήματος συντεταγμένων (α) προβολή του διανύσματος της παραγώγου στο ακτινικό σύστημα συντεταγμένων (β) εναλλακτική περιγραφή ως μέτρο-γωνία της παραγώγου στο ακτινικό σύστημα συντεταγμένων (γ) παράδειγμα περιστροφικής ανεξαρτησίας της περιγραφής της παραγώγου με χρήση ακτινικού συστήματος συντεταγμένων	61
6.9	Επίδραση της περιστροφής στην τοπική περιγραφή - Απλή μέθοδος HOG 62	62

6.10	Επίδραση της περιστροφής στην τοπική περιγραφή - Περιγραφή ως προς το κέντρο του κελιού	63
6.11	Επίδραση της περιστροφής στην τοπική περιγραφή - Περιγραφή ως προς το κέντρο της εικόνας	63
6.12	Επίδραση μικρών μετατοπίσεων στην τοπική περιγραφή - Το κέντρο του κελιού ως σημείο αναφοράς	64
6.13	Επίδραση μικρών μετατοπίσεων στην τοπική περιγραφή - Το κέντρο της εικόνας ως σημείο αναφοράς	64
6.14	Η προσέγγιση του υπολογισμού της ακτινικής παραγώγου ως αριστερή κυκλική ολίσθηση του υπολογισθέντος ιστογράμματος. Διαισθητικά είναι πιο εύκολη η χρήση ενός κυκλικού ιστογράμματος με διαφορετικό αρχικό bin. Για το παράδειγμα του σχήματος χρησιμοποιείται εύρος γωνιών $[0^\circ, 180^\circ)$ (unsigned gradient) και 9 bins.	65
6.15	Παράδειγμα της τελικής προσέγγισης για $\theta < \theta_b$	66
6.16	Ανεξαρτησία τοπικών περιγραφών στην περιστροφή για τις διαφορές προσεγγίσεις που εισήχθησαν	67
6.17	Παραδειγμά της αποτελεσματικότητας της 3 ^{ης} προσέγγισης σε δυαδικές εικόνες	68
6.18	Ανάλυση των δυο εναλλακτικών διαδικασιών για το τελικό στάδιο των κυκλικών περιγραφών: (α) μέθοδος για την εύρεση της περιστροφής μεταξύ δύο εικόνων βασιζόμενη στην ετεροσυσχέτιση των HOG περιγραφών των επιμέρους δακτυλίων για την κατασκευή μιας ενιαίας ετεροσυσχέτισης (β) δημιουργία περιστροφικά ανεξάρτητων ενιαίων περιγραφών	69
7.1	Αναπαράσταση πίνακα ποσοστών εμφάνισης ανά νοηματιστή	72
7.2	Παράδειγμα χειρομορφών της βάσης απο κάθε κλάση	73
7.3	Παράδειγμα ποικιλομορφίας στην εμφάνιση (ποσοστό εμφάνισης του χεριού, φωτεινότητα και οπτική γωνία) για το γράμμα Δ	73
7.4	Αναπαράσταση πίνακα αποστάσεων χωρίς ολική ανεξαρτησία σε περιστροφή	74
7.5	Αναπαράσταση πίνακα αποστάσεων με ολική ανεξαρτησία σε περιστροφή	74
7.6	Επίδραση ποσοστού εκπαίδευσης στο ποσοστό σφάλματος αναγνώρισης	75
7.7	Επίδραση των τεχνικών PCA και LDA στο ποσοστό αναγνώρισης σε σχέση με τον ταξινομητή που χρησιμοποιείται.	78
7.8	confusion matrices : (α) χωρίς την χρήση LDA, (β) με χρήση LDA	79
7.9	Σφαλμά αναγνώρισης για τους ταξινομητές απόστασης σε 4 ενδεικτικές περιγραφές	80
7.10	Σφάλμα αναγνώρισης για τον kNN ταξινομητή σε 4 ενδεικτικές περιγραφές μεταβάλλοντας την τιμή του k	80
7.11	Ευκλείδεια απόσταση των διανυσμάτων κάθε κλάσης από το κέντρο της. Τα διανύσματα αποτελούν τις HOG περιγραφές των εικόνων της βάσης.	81

7.12	Ευκλείδεια απόσταση των σημείων N-διάστατων γκαουσιανών κατανομών από το κέντρο τους. Οι γκαουσιανές του σχήματος έχουν διαγώνιο πίνακα συνδιασποράς (ανεξάρτητα στοιχεία).	82
7.13	Γραφική αναπαράσταση της διαγωνίου του πίνακα συνδιασποράς για κάθε κλάση. Στο Σχήμα (α) φαίνεται η μέση συνδιασπορά για όλες της κλάσεις, όπως προκύπτει και από την τεχνική LDA	85
7.14	Επίδραση του πλήθους των κελιών, μεταβάλλοντας τον αριθμό των διαμερίσεων, ανά άξονα για hog περιγραφές	86
7.15	Επίδραση του πλήθους των bins για hog περιγραφές	87
7.16	Επίδραση του εύρους γωνιών (unsigned gradient vs signed gradient) για hog περιγραφές	88
7.17	Επίδραση του ποσοστού επικάλυψης των κελιών για hog περιγραφές	88
7.18	Επίδραση του τελεστή παραγωγού για hog περιγραφές	89
7.19	Επίδραση του πλήθους των κελιών, μεταβάλλοντας τον αριθμό των διαμερίσεων ανά άξονα, για bhog περιγραφές	90
7.20	Επίδραση του ποσοστού επικάλυψης των κελιών για bhog περιγραφές	90
7.21	Επίδραση του πλήθους των κελιών για chog περιγραφές	91
7.22	Επίδραση του πλήθους των bins για chog περιγραφές	91
7.23	Επίδραση του πλήθους των κελιών για cbhog περιγραφές	91
7.24	Επίδραση του πλήθους των κύκλων αναφοράς διατηρώντας το πλήθος των συνολικών κελιών σταθερό, όπου (α),(β): rhog και (γ),(δ): erhog	92
7.25	Αναλυτική επίδραση των χωρικών παραμέτρων για rhog περιγραφές	93
7.26	Αναλυτική επίδραση των χωρικών παραμέτρων για erhog περιγραφές	94
7.27	Αναλυτική επίδραση των χωρικών παραμέτρων για rbhog περιγραφές	95
7.28	Αναλυτική επίδραση των χωρικών παραμέτρων για erbhog περιγραφές	96
7.29	Επίδραση του πλήθους των σημείων αναφοράς	97
7.30	Επίδραση της ακτίνας του κύκλου αναφοράς	97
7.31	Σύγκριση τεχνικών περιγράμματος	99
7.32	Σύγκριση τεχνικών βασισμένων σε HOG περιγραφές	100
7.33	Συγκρίση των τεχνικών Centralized HOG και Centralized Binary HOG	101
7.34	Confusion Matrices για Centralized HOG και Centralized Binary HOG	101
7.35	Εικόνες αναφοράς και δοκιμών	102
7.36	Σφάλμα εκτίμησης γωνίας περιστροφής για την μειωμένη περιγραφή. (α),(β): ίδια εικόνα αναφοράς (γ),(δ): ως διάνυσμα αναφοράς χρησιμοποιείται το κέντρο της κλάσης	103
7.37	Σφάλμα εκτίμησης της γωνίας περιστροφής για την πλήρη περιγραφή. (α),(β): ίδια εικόνα αναφοράς (γ),(δ): ως διάνυσμα αναφοράς χρησιμοποιείται το κέντρο της κλάσης	104
7.38	Σφάλμα εκτίμησης της γωνίας περιστροφής για διαφορετικό πλήθος σημείων στον χώρο	104
7.39	Επίδραση του πλήθους των τοπικών περιγραφών των δακτυλίων (rhog)	105
7.40	Επίδραση του πλήθους των τοπικών περιγραφών των ελλειπτικών δακτυλίων (erhog)	106
7.41	Ικανότητα αναγνώρισης ανά γωνία περιστροφής (eucl,unseen signers)	107
7.42	Σύγκριση των τεχνικών ανεξαρτησίας των τοπικών περιγραφών στην περιστροφή	108

7.43	Επίδραση της χωρικής παραμέτρου np_{init} (rhog&ftm_8_6_ np_{init})	108
7.44	Επίδραση της χωρικής παραμέτρου np_{init} (erhog&ftm_8_6_ np_{init})	109
7.45	Σύγκριση τεχνικών περιγράμματος ανεξάρτητων σε περιστροφές	112
7.46	Confusion Matrices για τις τεχνικές περιγράμματος	113
7.47	Ιστογράμματα απόκλισης της εκτιμώμενης γωνίας για τις τεχνικές περιγράμματος	113
7.48	Συνδιασμός Τεχνικών Περιγράμματος (eucl,unseen signers)	114
7.49	Ιστογράμματα απόκλισης της εκτιμώμενης γωνίας για τις τεχνικές RHOG, ERHOG και SRef (eucl,unseen signer)	116
7.50	Αποτελεσματικότητα των RHOG, ERHOG στην περιστροφή	117
7.51	Confusion Matrices για τις τεχνικές RHOG, ERHOG, RBHOG και ERBHOG	118
8.1	Προτεινόμενο σύστημα εξαγωγής χαρακτηριστικών για video	122
B'.1	Υπολογισμός του κέντρου της εικόνας ως κέντρο βάρους	126
B'.2	Υπολογισμός του κέντρου της εικόνας ως κέντρο του μέγιστου εγγεγραμμένου κύκλου	126
B'.3	Γραφική απεικόνιση των αποτελεσμάτων του Μετασχηματισμού Απόστασης	127

Κατάλογος πινάκων

5.1	Πιθανές τιμές Διεύθυνσης και Μέτρου	46
7.1	Συνολικά ποσοστά χειρομορφών ανά νοηματιστή	71
7.2	Ποσοστά επιτυχούς αναγνώρισης για mahalanobis ταξινομητή	83
7.3	Ποσοστά επιτυχούς αναγνώρισης για mahalanobis ταξινομητή κοινού πίνακα συνδιασποράς	84
7.4	Ποσοστά επιτυχούς αναγνώρισης για ευκλείδειο ταξινομητή	84
7.5	Αρχικές επιλογές παραμέτρων για τα HOG	86
7.6	Επιλογές βέλτιστων παραμέτρων για τα HOG	89
7.7	Ποσοστά επιτυχούς αναγνώρισης της μεθόδου περιγραμμάτων αναφοράς για seen signers	109
7.8	Ποσοστά επιτυχούς αναγνώρισης της μεθόδου περιγραμμάτων αναφοράς για unseen signers	110
7.9	Ποσοστά επιτυχούς αναγνώρισης της μεθόδου hu	111
7.10	Ποσοστά επιτυχούς αναγνώρισης για τις τεχνικές RHOG, ERHOG και SRef (eucl,unseen signer)	115

Κεφάλαιο 1

Εισαγωγή

Οι χειρονομίες αποτελούν μια πολύ διαδεδομένη και εκφραστική μορφή επικοινωνίας και χρησιμοποιούνται καθημερινά ποικιλοτρόπως, είτε για μια απλή έκφραση είτε ως μια οργανωμένη μορφή επικοινωνίας, όπως η Νοηματική Γλώσσα. Μαζί με την ομιλία, που αδιαμφισβήτητα είναι η κύρια μορφή επικοινωνίας, συνιστούν τον βασικό τρόπο επικοινωνίας του ανθρώπου και επομένως είναι οι φυσικές επιλογές για την αλληλεπίδραση ανθρώπου - υπολογιστή (Human Computer Interaction - HCI). Το πρόβλημα της αναγνώρισης της φωνής (speech recognition) έχει μελετηθεί εκτενώς με μεγάλη επιτυχία, γεγονός που είναι πλέον εμφανές από την πληθώρα εμπορικών εφαρμογών που σχετίζονται με αναγνώριση φωνής. Αντίθετα, το πρόβλημα της αναγνώρισης χειρονομιών (hand gesture recognition) βρίσκεται σε πιο πρώιμο στάδιο λόγω της πολυεπίπεδης δυσκολίας που παρουσιάζει, τόσο λόγω κάποιας ασάφειας στις κινήσεις, που εύκολα ερμηνεύεται από τον άνθρωπο αλλά είναι πολύ δύσκολο να μοντελοποιηθεί, όσο και για την εγγενή δυσκολία προβλημάτων της Όρασης Υπολογιστών.

Γενικά, η αναγνώριση χειρονομιών έχει εφαρμογή σε συστήματα παρακολούθησης, αναγνώρισης κινήσεων, ελέγχου οικιακών και ιατρικών συσκευών ακόμα και σε εφαρμογές ηλεκτρονικών παιχνιδιών, όπου είναι δυνατόν να παρέχει μια πιο ρεαλιστική αλληλεπίδραση με το παιχνίδι. Πέρα από την χρησιμότητα στις παραπάνω εφαρμογές της επικοινωνίας ανθρώπου-μηχανής, η αναγνώριση χειρομορφών, υπό την μορφή της Νοηματικής Γλώσσας, προσφέρει την δυνατότητα επικοινωνίας ανθρώπων με προβλήματα ακοής τόσο με τον υπολογιστή όσο και με άλλα άτομα που δεν γνωρίζουν την νοηματική γλώσσα μέσω του υπολογιστή. Συνεπώς ένα αποτελεσματικό σύστημα αναγνώρισης αλλά και σύνθεσης της νοηματικής γλώσσας θα διευκολύνει σε πολύ μεγάλο βαθμό τέτοια άτομα, αφαιρώντας την ανάγκη για γνώση την νοηματικής γλώσσας από τα άτομα που επικοινωνούν μαζί τους, καθιστώντας δυνατή η συμμετοχή τους σε ένα μεγάλο εύρος δραστηριοτήτων πολύ πιο εύκολα.

1.1 Όραση Υπολογιστών

Για την δημιουργία ενός αποτελεσματικού συστήματος αναγνώρισης χειρονομιών απαιτείται η εξαγωγή ποιοτικών οπτικών χαρακτηριστικών. Το πρόβλημα αυτό αποτελεί αντικείμενο του τομέα της Όρασης Υπολογιστών.

Η Όραση Υπολογιστών είναι το πεδίο το οποίο επιχειρεί να αναπαράγει αλγοριθμικά την αίσθηση της όρασης και έχει εφαρμογή σε πάρα πολλούς τομείς της εποχής μας όπως σε εφαρμογές ασφάλειας-επιτήρησης ή σε εφαρμογές βιοϊατρικής. Ιδιαίτερο ενδιαφέρον παρουσιάζει ο τομέας της Ψηφιακής Επεξεργασίας της Εικόνας για τον σκοπό της αναγνώρισης αντικειμένων (object recognition), όπου γίνεται προσπάθεια για την εύρεση και ομαδοποίηση χαρακτηριστικών κάποιας ψηφιακής εικόνας ή κάποιου μετασχηματισμού της έτσι ώστε να εξαχθεί συστηματικά με συμπαγή τρόπο η αναγκαία πληροφορία για την ερμηνεία της. Η διαδικασία αυτή είναι ανάλογη με αυτή της ανθρώπινης όρασης, καθώς ο άνθρωπος είναι ικανός να αναγνωρίσει επιτυχώς αντικείμενα υπό ένα πολύ μεγάλο εύρος αλλαγών και παραμορφώσεων.

Βέβαια η δυνατότητα ποιοτικής περιγραφής των χειρονομιών δεν εγγυάται την επιτυχία του συστήματος, καθώς απαιτείται και ένα ισχυρό σύστημα αναγνώρισης. Μάλιστα η μέθοδος που θα χρησιμοποιηθεί για τους σκοπούς της αναγνώρισης επιλέγεται είτε αναπτύσσεται σύμφωνα με την προηγηθείσα εξαγωγή χαρακτηριστικών και αντίστροφα, δηλαδή κατά την εξαγωγή χαρακτηριστικών λαμβάνουμε υπόψιν τις δυνατότητες και τις αδυναμίες των διαφόρων τεχνικών αναγνώρισης.

1.2 Αναγνώριση Προτύπων

Ο τομέας της Αναγνώρισης Προτύπων (Pattern Recognition) είναι ένα επιστημονικό πεδίο με στόχο την ανάπτυξη αλγορίθμων για την ταξινόμηση δεδομένων σε προκαθορισμένες κατηγορίες-κλάσεις ή τον διαχωρισμό τους σε ομάδες με βάση κάποια κριτήρια. Ουσιαστικά οι αλγόριθμοι αναγνώρισης προτύπων καλούνται να εκπαιδεύσουν το σύστημα έτσι ώστε να αναγνωρίζει πρότυπα με σκοπό την κατηγοριοποίηση σε κλάσεις.

Για το στάδιο της εκπαίδευσης έχουμε δυο διαφορετικές προσεγγίσεις. Η πρώτη στοχεύει στην δημιουργία ενός μοντέλου για τον καλύτερο διαχωρισμό των κλάσεων μεταξύ τους (discriminative models) με χαρακτηριστικό παράδειγμα αυτό των SVM (Support Vector Machines), ενώ η δεύτερη προσπαθεί να μοντελοποιήσει εσωτερικά κάθε κατηγορία, συνήθως εκτιμώντας τις παραμέτρους κάποιας πιθανοτικής κατανομής (generative models) με χαρακτηριστικά παραδείγματα αυτά των GMM (Gaussian Mixture Models) και HMM (Hidden Markov Models). Η δεύτερη κατηγορία μπορεί να θεωρηθεί πιο γενική, καθώς δίνει μια εκτίμηση της κατανομής των δεδομένων και δεν περιορίζεται στο πρόβλημα της ταξινόμησης. Βέβαια έχει αποδειχτεί ότι η πρώτη κα-

τηγορία είναι ιδιαίτερα ισχυρή σε προβλήματα ταξινόμησης και μπορεί να έχει πολύ καλά αποτελέσματα σε περιπτώσεις που η δεύτερη αδυνατεί ιδίως λόγω της ανάγκης για κάποια υπόθεση κατανομής των δεδομένων.

1.2.1 Σύστημα Αναγνώρισης Νοηματικής Γλώσσας

Ένα πλήρες λειτουργικό σύστημα αναγνώρισης Νοηματικής Γλώσσας συνεχούς «ομιλίας» (video) αποτελεί ένα ιδιαίτερα δύσκολο πρόβλημα καθώς είναι εξαιρετικά δύσκολο να μοντελοποιήσουμε τις μεταβιβάσεις ανάμεσα σε λέξεις. Για τον λόγο αυτό συνήθως γίνεται μελέτη του απλοποιημένου προβλήματος της αναγνώρισης συγκεκριμένων χειρονομιών - λέξεων της Νοηματικής Γλώσσας. Ακόμα και έτσι το πρόβλημα παραμένει δύσκολο και τα σύγχρονα συστήματα κάνουν χρήση πληθώρας διαφορετικών ροών πληροφορίας (multi-channel systems) πέρα των χειρομορφών, όπως η έκφραση του προσώπου ακόμα και πληροφορία φωνής εάν είναι διαθέσιμη.

Ως προς την υλοποίηση η κάθε λέξη-χειρονομία μοντελοποιείται ξεχωριστά. Μια πολύ δημοφιλής τεχνική για τον σκοπό αυτό για εφαρμογές που εμπεριέχουν εξέλιξη στον χρόνο είναι αυτή των HMM, σύμφωνα με την οποία σε κάθε χειρονομία αντιστοιχεί και εκπαιδεύεται ένα HMM που μοιάζει με ένα πιθανοτικό αυτόματο (εκπαιδεύει ένα σύνολο κατανομών σε κάθε state και τις πιθανότητες μεταβίβασης από state σε state). Για την εκπαίδευση του όμως χρειάζεται ένα κατάλληλο σύνολο χαρακτηριστικών για κάθε frame του βίντεο, το οποίο αποτελείται από κάθε δυνατή πληροφορία που μπορεί να εξαχθεί από τα διάφορα κανάλια (κίνηση - θέση χεριών - χειρομορφή - έκφραση προσώπου - φωνή).

Μεγάλη ώθηση σε εφαρμογές επικοινωνίας ανθρώπου μηχανής έχει προσφέρει η κυκλοφορία του Kinect [11] και ειδικότερα στην κατεύθυνση της αναγνώρισης κινήσεων. Συγκεκριμένα το Kinect εκτός από το RGB βίντεο προσφέρει εικόνες βάθους με χρήση ενός αισθητήρα υπερύθρων καθώς και μια τρισδιάστατη εκτίμηση του σκελετού του χρήστη (υλοποιείται από ένα ενσωματωμένο microchip). Επομένως είναι εμφανής η χρησιμότητα του Kinect στις εφαρμογές αναγνώρισης χειρονομιών αφού διευκολύνει σε μεγάλο βαθμό την εκτίμηση της θέσης των κινήσεων του χεριού και του σώματος.

Τέλος αξίζει να αναφέρουμε ότι η πλειονότητα των ως τώρα προτεινόμενων συστημάτων για αναγνώριση χειρονομιών δεν αξιοποιεί πλήρως την οπτική πληροφορία των χειρομορφών λόγω της μεγάλης δυσκολίας της εκτίμησης της τρισδιάστατης πόζας του χεριού. Αντ'αυτού κυρίως βασίζονται στην κίνηση και στην θέση των χεριών συνοδευόμενο από μια κάπως ασαφή πληροφορία για την μορφή του χεριού.

1.2.2 Αναγνώριση Στατικών Χειρομορφών - Fingerspelling

Το πρόβλημα της αναγνώρισης χειρονομιών από βίντεο χωρίς κάποια επιπλέον ροή πληροφορίας (πληροφορία από το πρόσωπο του νοηματιστή είτε ήχο) είναι ένα ιδιαίτερα δύσκολο πρόβλημα Όρασης Υπολογιστών που ουσιαστικά αποτελείται από δυο εσωτερικές ροές πληροφορίας, αυτή της κίνησης και αυτή της εμφάνισης των χειρομορφών. Εμείς θα ασχοληθούμε κυρίως με την δεύτερη ροή πληροφορίας και μάλιστα σε μια πιο απλοποιημένη μορφή της λαμβάνοντας υπόψιν μόνο στατικές εικόνες. Συγκεκριμένα θα ασχοληθούμε με το πρόβλημα της αναγνώρισης των γραμμάτων της ελληνικής αλφαβήτου της Νοηματικής Γλώσσας (fingerspelling), δηλαδή την κατηγοριοποίηση μιας χειρομορφής σε μια από τις 24 διακριτές κλάσεις που αντιστοιχούν στα γράμματα της αλφαβήτου, οι οποίες φαίνονται στο σχήμα 1.1. Όπως είναι εύλογο το πρόβλημα αυτό είναι μια αρκετά απλοποιημένη μορφή του γενικού προβλήματος της αναγνώρισης χειρονομιών, αλλά παρουσιάζει μεγάλο ενδιαφέρον καθώς στην αναγνώριση χειρονομιών συναντάμε μεγάλη δυσκολία στην πλήρη αξιοποίηση της εμφάνισης του χεριού, δηλαδή στην εξαγωγή χρήσιμων οπτικών χαρακτηριστικών.

Το σύστημα αναγνώρισης που χρησιμοποιούμε αποτελείται από τα εξής στάδια, θεωρώντας ως είσοδο βίντεο από το Kinect

- **Εντοπισμός Θέσης Χεριών (Hand Tracking)** Το Kinect πέρα από το βίντεο και την πληροφορία βάθους, μας παρέχει και πληροφορία για τον «σκελετό» του νοηματιστή, υπό την μορφή 15 σημείων κατανεμημένα στο σώμα του, όπως φαίνεται στο σχήμα 1.2. Οι περιοχές των χεριών και συγκεκριμένα το κέντρο τους αποτελούν δυο από τα εξαχθέντα σημεία και συνεπώς έχουμε μια άμεση εκτίμηση για την θέση των χεριών. Αν οι κινήσεις είναι ομαλές, χωρίς κάποια απότομη αλλαγή, η εκτίμηση της θέσης των χεριών μέσω του 'σκελετού' είναι αρκετά ακριβής και για τον λόγο αυτό προτιμήθηκε στην πράξη η αξιοποίηση αυτής της πληροφορίας από την χρήση κάποιας πολύπλοκης τεχνικής εντοπισμού των χεριών.
- **Εύρεση Μάσκας Χεριών (Hand Segmentation)** Σε αυτό το στάδιο, όπως και στο προηγούμενο, γίνεται άμεση αξιοποίηση των δυνατοτήτων του Kinect. Συγκεκριμένα επιθυμούμε να βρούμε την δυαδική μάσκα του χεριού και να αγνοήσουμε το background. Συνήθως κάτι τέτοιο απαιτεί σύνθετες τεχνικές που βασίζονται σε μεγάλο βαθμό στον εντοπισμό περιοχών στο χρώμα του δέρματος (skin detection). Οι τεχνικές αυτές παρότι έχουν χρησιμοποιηθεί με μεγάλη επιτυχία έχουν διάφορους περιορισμούς (π.χ. έντονες αλλαγές φωτεινότητας, το χέρι στο ίδιο ύψος με το πρόσωπο κ.λπ.). Αντίθετα η εξαγωγή της μάσκας του χεριού είναι μια πολύ εύκολη διαδικασία χρησιμοποιώντας την πληροφορία βάθους (depth) που παρέχει το Kinect υπό τον περιορισμό ότι το χέρι του νοηματιστή θα βρίσκεται σε τέτοια απόσταση από το υπόλοιπο σώμα έτσι ώστε να είναι εύκολα διακριτό στις εικόνες βάθους (depth images). Με τον τρόπο αυτό η εξαγωγή της μάσκας του χεριού επιτυγχάνεται με μια απλή κατωφλιοποίηση της πληροφορίας του βάθους. Η διαδικασία περιγράφεται οπτικά στο σχήμα 1.3.

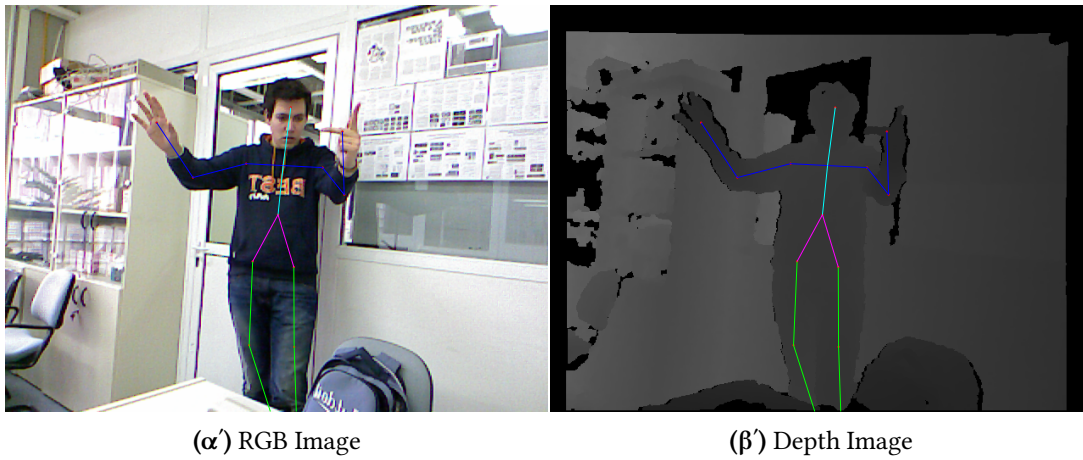


Σχήμα 1.1: Χειρομορφές των γραμμάτων της ελληνικής νοηματικής γλώσσας

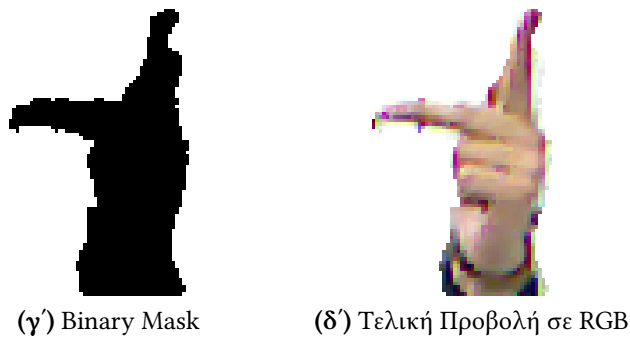
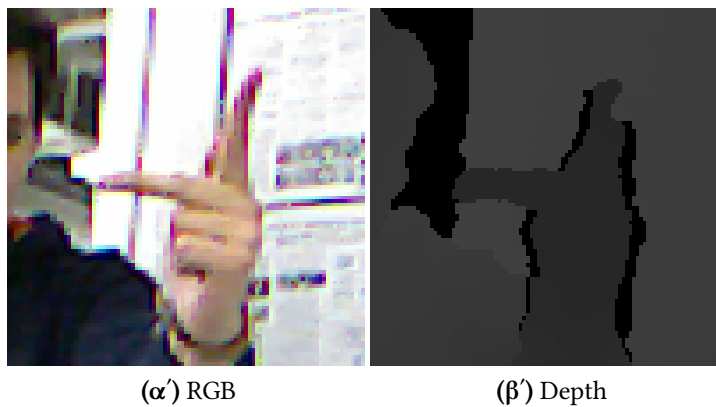
- **Εξαγωγή Χαρακτηριστικών (Feature Extraction)** Η εξαγωγή των οπτικών χαρακτηριστικών της χειρομορφής αποτελεί το πιο δύσκολο και σημαντικό στάδιο στο σύστημα αυτό. Ιδιαίτερη αναφορά στις τεχνικές εξαγωγής χαρακτηριστικών σε εικόνες θα γίνει στην επόμενη ενότητα αλλά και στο μεγαλύτερο μέρος της διπλωματικής.
- **Κατηγοριοποίηση Χειρομορφής (Handshape Classification)** Στο στάδιο αυτό χρησιμοποιούμε κάποια από τις τεχνικές αναγνώρισης προτύπων για να κατηγοριοποιήσουμε την εξαχθείσα χειρομορφή σε μια από τις 24 δυνατές κλάσεις.

Όπως είναι εμφανές από τα παραπάνω, τα δύο πρώτα στάδια είναι ιδιαίτερα απλά με την χρήση του Kinect και για τον λόγο αυτό στο μεγαλύτερο μέρος της διπλωματικής

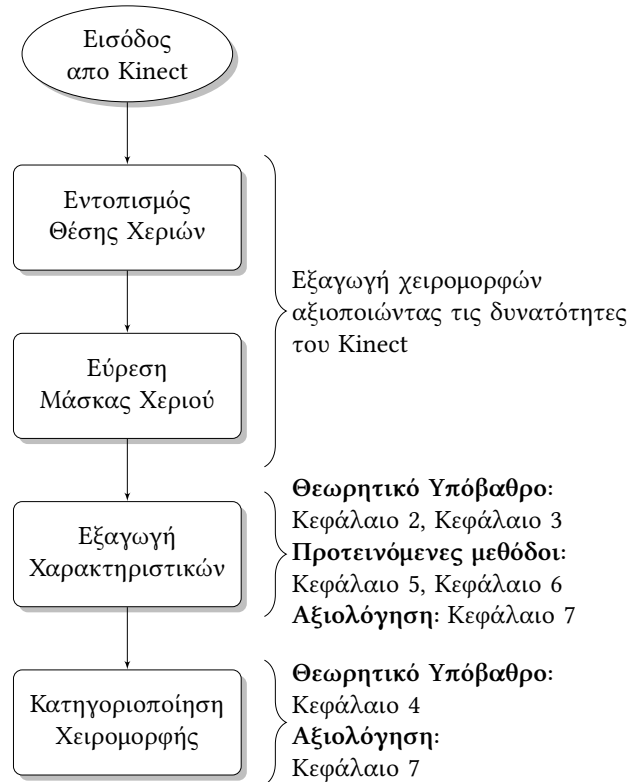
επικεντρωνόμαστε στα δυο τελευταία στάδια, τα οποία παρουσιάζουν και το μεγαλύτερο ενδιαφέρον. Το σύστημα αναγνώρισης απεικονίζεται αφαιρετικά στο σχημα 1.4.



Σχήμα 1.2: Προβολή του σκελετού που παράγει το Kinect πάνω στην RGB εικόνα και στην εικόνα Βάθους



Σχήμα 1.3: Εξαγωγή της περιοχής της χειρομορφής με χρήση της πληροφορίας βάθους



Σχήμα 1.4: Σύνοψη συστήματος αναγνώρισης

1.3 Ερευνητικοί Στόχοι - Προσφορά

Η παρούσα διπλωματική εργασία επικεντρώνεται στο πρόβλημα της αναγνώρισης στατικών χειρομορφών της Νοηματικής Γλώσσας και συγκεκριμένα του ελληνικού αλφαβήτου (fingerspelling). Ιδιαίτερη έμφαση δόθηκε στην εξαγωγή αποδοτικών περιγραφών τόσο περιγράμματος όσο και εμφάνισης με κεντρικό άξονα την μέθοδο Histogram of Oriented Gradients (HOG) [8]. Στόχος της όλης μελέτης είναι οι περιγραφές αυτές να είναι αποδοτικές τόσο ως προς την ικανότητα διάκρισης αλλά και ως προς την συμπίεση της ουσιαδους πληροφορίας. Αναλυτικά η ερευνά μας επικεντρώθηκε στις εξής κατευθύνσεις

- Εξερεύνηση πιθανών παραλλαγών των HOG (Κεφάλαιο 5)
- Εξερεύνηση του προβλήματος της ανεξαρτησίας της εικόνας σε περιστροφές (Κεφάλαιο 6)
- Μελέτη του πολυδιάστατου χώρου των περιγραφών και αξιολόγηση των προτεινόμενων τεχνικών (Κεφάλαιο 7)

1.3.1 Παραλλαγές HOG

Η βασική τεχνική των HOG είναι εξαιρετικά απλή αν και σε μεγάλο βαθμό αποτελεσματική. Για τον λόγο αυτό στο Κεφάλαιο 5 προτείνουμε σχετικά απλές παραλλαγές της μεθόδου για να διατηρήσουμε ως ένα βαθμό την απλότητα και την ταχύτητα της αρχικής μεθόδου, αυξάνοντας παράλληλα την αποτελεσματικότητα στην αναγνώριση. Επιπλέον προτείνουμε μια παραλλαγή των HOG για δυαδικές εικόνες περιγράμματος που βασίζεται στην ευκολία της διαχείρισης των δυαδικών εικόνων και απλοποιεί περαιτέρω την βασική τεχνική.

1.3.2 Ανεξαρτησία στην Περιστροφή

Στο Κεφάλαιο 6 μελετάμε σε βάθος τις επιλογές που πρέπει να γίνουν για να αναπτύξουμε μια τεχνική εξαγωγής οπτικών χαρακτηριστικών ανεξάρτητη σε περιστροφές της εικόνας. Για τον σκοπό αυτό προτείνουμε δυο εναλλακτικές προσεγγίσεις βασισμένες κυρίως σε HOG και παρουσιάζουμε τα προτερήματα και τα πιθανά προβλήματα που μπορεί να ανακύψουν.

1.3.3 Χώρος Περιγραφών - Αξιολόγηση

Το Κεφάλαιο 7 αποτελεί το πειραματικό μέρος της παρούσας μελέτης και μπορεί να χωριστεί σε δυο νοητά τμήματα. Το πρώτο ασχολείται με τον χώρο των περιγραφών που βασίζονται στα HOG και την καταλληλότητα - αποτελεσματικότητα διάφορων τεχνικών αναγνώρισης για τους χώρους αυτούς. Το δεύτερο τμήμα αποτελεί αξιολόγηση των προτεινόμενων τεχνικών αφότου βρεθούν οι παράμετροι που βελτιστοποιούν την απόδοσή τους. Στο τμήμα αυτό μελετάμε τόσο την απλή αναγνώριση χειρομορφών όσο και την αναγνώριση υπό κυκλική περιστροφή.

1.4 Σχετική Έρευνα

Όπως τονίσαμε προηγουμένως, πιθανότατα το πιο απαιτητικό κομμάτι στην αναγνώριση χειρονομιών είναι αυτό της εξαγωγής οπτικών χαρακτηριστικών. Σε αυτή την κατεύθυνση έχουν προταθεί ποικίλες τεχνικές και ξεχωρίζουμε την κατηγοριοποίηση σε τεχνικές περιγράμματος και σε τεχνικές που συνδυάζουν τόσο το περίγραμμα του χεριού όσο και την εμφάνιση του. Οι κατηγορίες αυτές περιγράφονται επαρκώς στο Κεφάλαιο 2 και στο Κεφάλαιο 3 αντίστοιχα.

Στην πρώτη κατηγορία ανήκουν τεχνικές όπως οι ροπές [14] που αξιοποιούν βασικά γεωμετρικά χαρακτηριστικά του χεριού, τεχνικές που βασίζονται στην καμπυλότητα

(curvature) ([33]) του περιγράμματος και δημοφιλείς τεχνικές όπως Fourier Descriptors [7]. Παρόμοιες τεχνικές έχουν χρησιμοποιηθεί κατά κόρον στον πρόβλημα της αναγνώρισης χειρομορφών ([30]), υποθέτοντας ότι κάθε κλάση χειρομορφών είναι δυνατόν να διακριθεί μόνο από το περιγράμματά τους και συγκεκριμένα από τα δαχτυλά, τα οποία προσπαθούν και να εντοπίσουν. Αυτή η υπόθεση δεν ικανοποιείται σε ένα πραγματικό σύστημα, όπου τόσο η πόζα, όσο και εσωτερική αναδιάταξη των δακτύλων δεν μπορούν να μοντελοποιηθούν μόνο με το περίγραμμα.

Η δεύτερη κατηγορία προσφέρει κατά βάση αξιολογικά αποτελέσματα λόγω της καλύτερης αξιοποίησης της υπάρχουσας πληροφορίας. Χαρακτηριστικό παράδειγμα είναι η τεχνική HOG [8], η οποία μελετάται εκτενώς στην παρούσα διπλωματική. Η τεχνική αυτή έχει ήδη εφαρμοστεί στο πρόβλημα της αναγνώρισης χειρομορφών [5], [24], [20] με υποσχόμενα αποτελέσματα, ενώ στο [29] αξιοποιείται η ισχύς του Kinect για το συγκεκριμένο πρόβλημα.

Παρόμοιες τεχνικές, ως προς την κεντρική ιδέα αναπαράστασής της οπτικής πληροφορίας, είναι τα SIFT που εισήγαγε ο Lowe [21] και τα Spatial Pyramids [19], που εφαρμόζουν μια πολυεπίπεδη (πυραμιδωτή) διαμέριση της εικόνας. Η παραλλαγή PCA-SIFT ([15]) χρησιμοποιείται με επιτυχία στο πρόβλημα του *fingerspelling* στο [16].

Ως προς τις τεχνικές αναγνώρισης, τα HOG χρησιμοποιούνται συνήθως σε συνδυασμό με SVM (Support Vector Machines), ενώ έχει δοκιμαστεί με επιτυχία η χρήση gaussian και chi-squared πυρήνων [27] σε αυτά, ενώ στο [13] γίνεται χρήση της τεχνικής LDA (Linear Discriminant Analysis) ως εναλλακτική των SVM για το πρόβλημα της αναγνώρισης αντικειμένων. Η τελευταία παραλλαγή βασίζεται στην υπόθεση γκαουσιανής κατανομής στον χώρο των HOG, μια συνθήκη που θα μελετήσουμε και εμείς για το πρόβλημα της αναγνώρισης χειρομορφών.

Για το πρόβλημα της ανεξαρτησίας των οπτικών χαρακτηριστικών στην περιστροφή της εικόνας έχει γίνει εκτενής μελέτη στην ανεξαρτησία τοπικών χαρακτηριστικών στα [18], [17], ενώ στο [25] παρουσιάζεται μια τεχνική για περιστροφικά ανεξάρτητες τοπικές περιγραφές που βασίζεται σε μια αρχική εξαγωγή χαρακτηριστικών όμοια με αυτή των HOG.

Κεφάλαιο 2

Εξαγωγή Οπτικών Χαρακτηριστικών Σχήματος

Θεωρούμε ότι τα σχήματα αναπαρίστανται ως ένα σύνολο σημείων S ή συνηθέστερα με την ακόλουθη δυαδική συνάρτηση που αντιστοιχεί στο σύνολο αυτό και αποτελεί την δυαδική μάσκα του σχήματος .

$$s(x, y) = \begin{cases} 1, & (x, y) \in S \\ 0, & (x, y) \notin S \end{cases} \quad (2.1)$$

Στις επόμενες παραγράφους περιγράφουμε μερικές από τις βασικότερες και πιο δημοφιλείς μεθόδους για την περιγραφή ενός σχήματος, ξεκινώντας από απλές τεχνικές με ιστορική σημασία και πλήθος απλοϊκών εφαρμογών (Hu moments) και καταλήγοντας σε πιο σύγχρονες και αποτελεσματικές τεχνικές που έχουν χρησιμοποιηθεί με επιτυχία σε αναγνώριση σύνθετων σχημάτων (Shape Context). Κύριος γνώμονας στην ανάλυση κάθε μεθόδου θα είναι η ανεξαρτησία ως προς την μετατόπιση (translation invariant), την αλλαγή κλίμακας (scale invariant) και την περιστροφή (rotation invariant).

2.1 Ροπές – Moments

Μια από τις πιο κλασσικές μεθόδους για την εξαγωγή χαρακτηριστικών ενός σχήματος είναι η χρήση στατιστικών ιδιοτήτων του σχήματος, υπό την μορφή ροπών (moments) που ορίζονται ως εξής, με το n να υποδηλώνει την τάξη των ροπών (n th-order moment):

$$\iint x^k y^{n-k} s(x, y) dx dy, \quad 0 \leq k \leq n, \quad n = 1, 2, 3, \dots \quad (2.2)$$

Η χρησιμότητα τους βασίζεται στην αντιστοιχία τους με απλές γεωμετρικές παρατηρήσεις, όπως το εμβαδόν του σχήματος ($n = 0$), το κέντρο βάρους ($n = 1$), η κατεύθυνση (χρήση σχέσεων για $n = 2$) κλπ.

Στην πράξη γίνεται χρήση των κεντρικών ροπών (central moments) και βέβαια τα ολοκληρώματα της σχέσης 2.2 μετατρέπονται σε αθροίσματα, καθώς η συνάρτηση $s(x, y)$ δεν είναι συνεχής και τα x, y αντιστοιχούν στα pixels. Δεδομένου των συντεταγμένων του κέντρου βάρους \bar{x} και \bar{y} , οι κεντρικές ροπές τάξης n υπολογίζονται με την ακόλουθη σχέση:

$$\mu_{ij} = \sum_x \sum_y (x - \bar{x})^i (y - \bar{y})^j s(x, y), \quad i + j = n \quad (2.3)$$

2.1.1 Hu Invariant Moments

Παρότι οι κεντρικές ροπές είναι ανεξάρτητες σε μετατοπίσεις, δεν ισχύει το ίδιο στην περίπτωση αλλαγής κλίμακας ή περιστροφής. Η ανεξαρτησία στην αλλαγή κλίμακας είναι δυνατόν να επιτευχθεί με κανονικοποίηση, χρησιμοποιώντας τον τύπο:

$$\eta_{ij} = \frac{\mu_{ij}}{\mu_{00}^p}, \quad p = 1 + \frac{i + j}{2}, \quad i + j \geq 2 \quad (2.4)$$

Βασιζόμενος στις παραπάνω κανονικοποιημένες ροπές ο *Hu* εισήγαγε στο [14] ένα σύνολο 7 ροπών (βλ. Α') που προσφέρουν ανεξαρτησία σε μετατόπιση, αλλαγή κλίμακας αλλά και περιστροφή. Οι 7 αυτές τιμές θα αποτελέσουν το διάνυσμα χαρακτηριστικών του εκάστοτε σχήματος, δημιουργώντας έτσι μια απλοϊκή πρώτη περιγραφή.

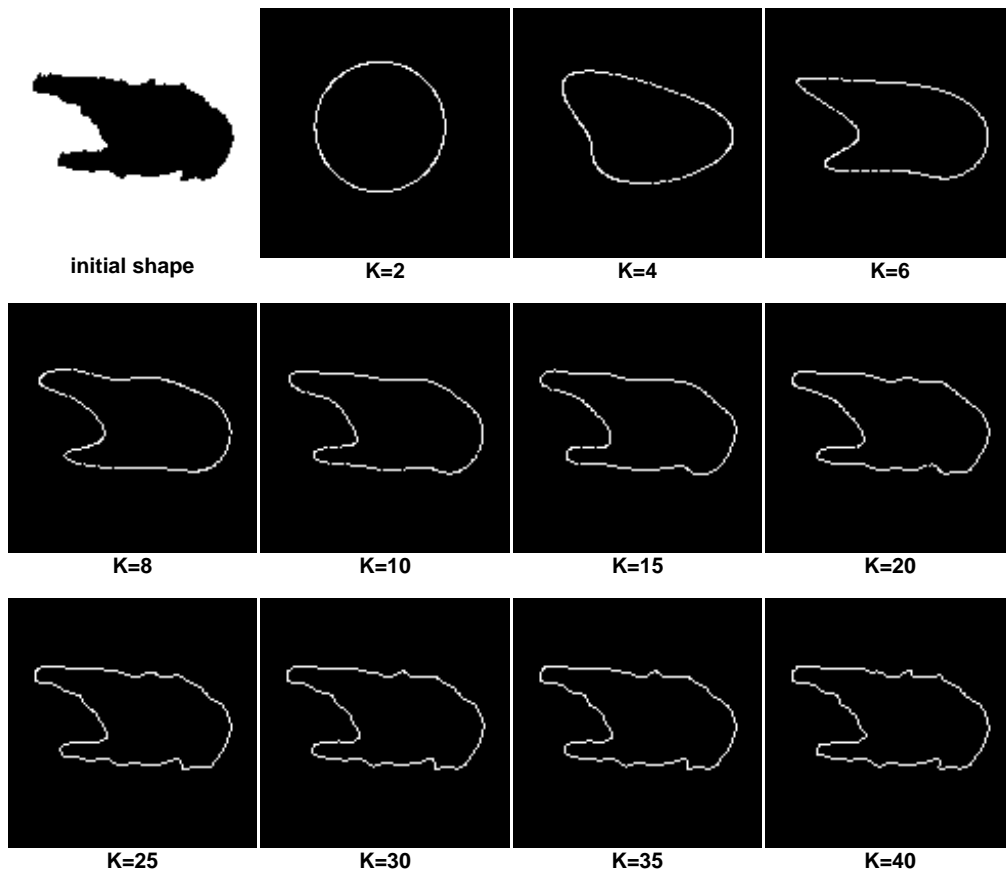
Βέβαια, όπως είναι και αναμενόμενο άλλωστε, η μέθοδος αυτή είναι υπερβολικά απλοϊκή για την αποτελεσματική περιγραφή περίπλοκων σχημάτων όπως είναι οι χειρομορφές, γεγονός που θα επαληθεύει και στο πειραματικό κομμάτι.

2.2 Fourier Descriptors

Μια από τις πιο διαδεδομένες τεχνικές για την περιγραφή ενός σχήματος είναι αυτή των Fourier Descriptors, των οποίων η βασική ιδέα συνοψίζεται στην παράγραφο αυτή. Έχοντας βρει το περίγραμμα του σχήματος ως μια ακολουθία σημείων-συντεταγμένων με μια αυθαίρετη επιλογή για αρχικό σημείο, αρχικά μετατρέπουμε κάθε σημείο (x, y) του περιγράμματος σε έναν μιγαδικό αριθμό $x + iy$ και στην συνέχεια «κωδικοποιούμε» το περίγραμμα, που πλέον περιγράφεται από την προκύπτουσα ακολουθία μιγαδικών, χρησιμοποιώντας τον μετασχηματισμό Fourier. Από την περιγραφή στο πεδίο των συχνοτήτων κρατάμε τους πρώτους k όρους, δημιουργώντας έτσι το τελικό διάνυσμα χαρακτηριστικών.

Εάν διατηρήσουμε όλους τους όρους του μετασχηματισμού, εφαρμόζοντας τον αντίστροφο μετασχηματισμό Fourier θα ανακτήσουμε την αρχική περιγραφή του περιγράμματος. Στην πράξη όμως, όπως ήδη αναφέραμε, κρατάμε μόνο τους πρώτους k όρους

με αποτέλεσμα το προκύπτον περίγραμμα από τον αντίστροφο μετασχηματισμό να είναι απλοποιημένο. Ουσιαστικά η μέθοδος αυτή οδηγεί στην εξομάλυνση του περιγράμματος και ισοδυναμεί με βαθυπερατό φιλτράρισμα (διατήρηση των συχνοτήτων που αντιστοιχούν σε μικρότερες τιμές από το k), άρα και αποκοπή απότομων - υψίσυχνων μεταβολών που είναι πολύ πιθανό να αποτελούν θόρυβο. Βέβαια πρέπει να επισημάνουμε ότι χρειάζεται προσοχή στην επιλογή του k καθώς μια μικρότερη τιμή αντιστοιχεί σε μεγαλύτερη απλοποίηση που εν τέλει μπορεί να οδηγήσει στην απώλεια κρίσιμης πληροφορίας, όπως κάποιο δάχτυλο στην περίπτωση που το αρχικό σχήμα είναι μια χειρομορφή. Η εξομάλυνση αυτή παρουσιάζεται και οπτικά στο σχήμα 2.1 για διαφορετικές τιμές του k .



Σχήμα 2.1: Παράδειγμα εξομάλυνσης με χρήση K Fourier συντελεστών. Παρατηρούμε ότι στην περίπτωση $K = 40$ το προκύπτον περίγραμμα είναι σχεδόν ίδιο με το αρχικό περιέχοντας και μεγάλο ποσοστό του θορύβου που έχει το αρχικό περίγραμμα

Η μέθοδος έτσι όπως παρουσιάστηκε δεν προσφέρει ανεξαρτησία ως προς κάποια από τις ζητούμενες ιδιότητες. Για την κατασκευή ανεξάρτητων - κανονικοποιημένων περιγραφών ακολουθούμε τα παρακάτω βήματα, λαμβάνοντας υπόψιν

- Πρώτον η ανεξαρτησία ως προς την μετατόπιση μπορεί να επιτευχθεί αγνοώντας τον συντελεστή Fourier για μηδενική συχνότητα (dc component) F_0 , καθώς μόνο

ο συντελεστής αυτός περιέχει πληροφορία για την θέση του περιγράμματος στον χώρο.

- Η ανεξαρτησία ως προς την αλλαγή κλίμακας επιτυγχάνεται με κανονικοποίηση. Αυτή μπορεί να υλοποιηθεί είτε πριν τον μετασχηματισμό, διαιρώντας τις συντεταγμένες κάθε σημείου με την μέση απόσταση όλων των σημείων του περιγράμματος από το κέντρο, είτε μετά τον μετασχηματισμό διαιρώντας τους συντελεστές που προκύπτουν με το μέτρο του F_1 αφού αλλαγή της κλίμακας κατά s ισοδυναμεί με πολλαπλασιασμό των συντελεστών Fourier με τον συντελεστή s .
- Τέλος ανεξαρτησία ως προς την περιστροφή, που ισοδυναμεί με ανεξαρτησία στην επιλογή του αρχικού σημείου του περιγράμματος, μπορεί να επιτευχθεί κρατώντας μόνο το μέτρο του μετασχηματισμού Fourier, αφού για περιστροφή κατά ϑ ισοδυναμεί με κυκλική περιστροφή του μιγαδικού διανύσματος περιγραφής του περιγράμματος και όπως είναι γνωστό το μέτρο του Fourier μετασχηματισμού είναι ανεξάρτητο στις (κυκλικές) μετατοπίσεις.¹

2.3 Χρήση Εναλλακτικών Περιγραφών του Περιγράμματος

Η γενική ιδέα της βασικής μεθόδου των των *Fourier Descriptors* μπορεί να διατηρηθεί με σκοπό την εξαγωγή χαρακτηριστικών με τις ζητούμενες ιδιότητες (*translation, scale & rotation invariance*), αλλά χρησιμοποιώντας κάποια διαφορετική αναπαράσταση των σημείων του περιγράμματος. Μια τέτοια αναπαράσταση θα μπορούσε να είναι η απόσταση από το κέντρο συναρτήσει της γωνίας, $r(\vartheta) = \sqrt{(x(\vartheta) - \bar{x})^2 + (y(\vartheta) - \bar{y})^2}$, ή η συνάρτηση καμπυλότητας για την ακολουθία των σημείων του περιγράμματος.

Η πρώτη περίπτωση έχει χρησιμοποιηθεί ουσιαστικά εκτός πλαισίου των *Fourier Descriptors*, βρίσκοντας την μέση απόσταση για κάποιο εύρος γωνιών, αφού χωρίσουμε τον χώρο όπως στο σχήμα 2.2. Στην συνέχεια, στο προκύπτον διάνυσμα χαρακτηριστικών, που έχει μέγεθος όσο και το πλήθος των διαμερίσεων των γωνιών, εφαρμόζεται μετασχηματισμός Fourier και κρατάμε μόνο το μέτρο για περιστροφική ανεξαρτησία, κατά τα γνωστά. Στην περίπτωση αυτή χρησιμοποιούμε όλους τους συντελεστές που

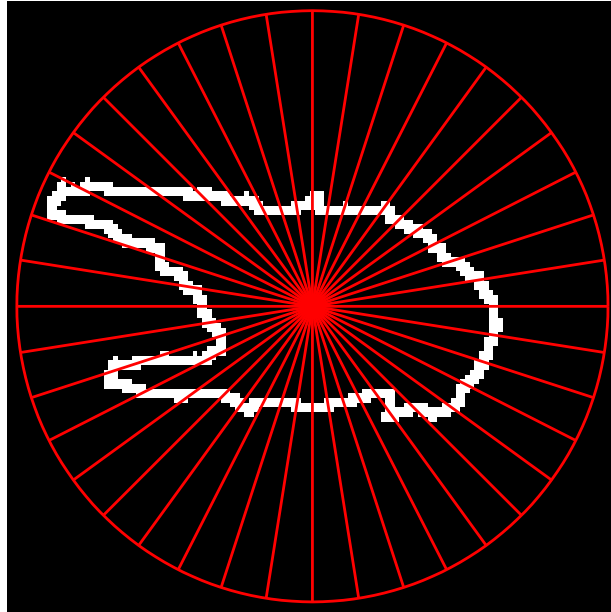
¹ Είναι :

$$\begin{aligned} f(x) &\xrightarrow{\mathcal{F}} F(\omega) \\ f(x-a) &\xrightarrow{\mathcal{F}} e^{-2\pi i a \omega} F(\omega) \\ |e^{-2\pi i a \omega}| &= 1 \end{aligned}$$

Συνεπώς

$$|\mathcal{F}\{f(x-a)\}| = |\mathcal{F}\{f(x)\}|$$

προκύπτουν θεωρώντας ότι έχει επέλθει μια μορφή εξομάλυνσης κατά την εύρεση των μέσων τιμών των αποστάσεων, ενώ για λόγους απλότητας το πλήθος των διαμερίσεων είναι ήδη σχετικά μικρό (της τάξης των 60 συντελεστών).



Σχήμα 2.2: Κυκλική Διαμέριση του Χώρου

2.4 Shape Context

Μια πιο σύγχρονη και αποτελεσματική μέθοδος για την περιγραφή ενός σχήματος εισήχθηκε στο [2]. Για την ακρίβεια η μέθοδος αυτή εφαρμόζεται στο ταίριασμα σχημάτων (*shape matching*) και εμφανίζει πολύ καλή συμπεριφορά ως προς πιθανές παραμορφώσεις. Έχοντας δυο σχήματα προς σύγκριση, η βασική λειτουργία της μεθόδου συνοψίζεται ως εξής :

- 1 Για κάθε σημείο του περιγράμματος δημιουργούμε ένα logpolar ιστόγραμμα που περιγράφει την κατανομή των υπόλοιπων σημείων ως προς αυτό (σημείο αναφοράς - κέντρο του logpolar διαγράμματος). Ο logpolar διαχωρισμός για ένα τυχαίο σημείο της περιφέρειας παριστάνεται γραφικά στο σχήμα 2.3. Συγκεκριμένα για σημείο αναφοράς p_i το ιστόγραμμα δίνεται από την σχέση:

$$h_i[k] = \#\{q \neq p_i : (q - p_i) \in \text{bin}(k)\}$$

Συνεπώς, πλέον έχουμε μια «χοντρική» περιγραφή του σχήματος από την οπτική γωνία του εκάστοτε σημείου αναφοράς.

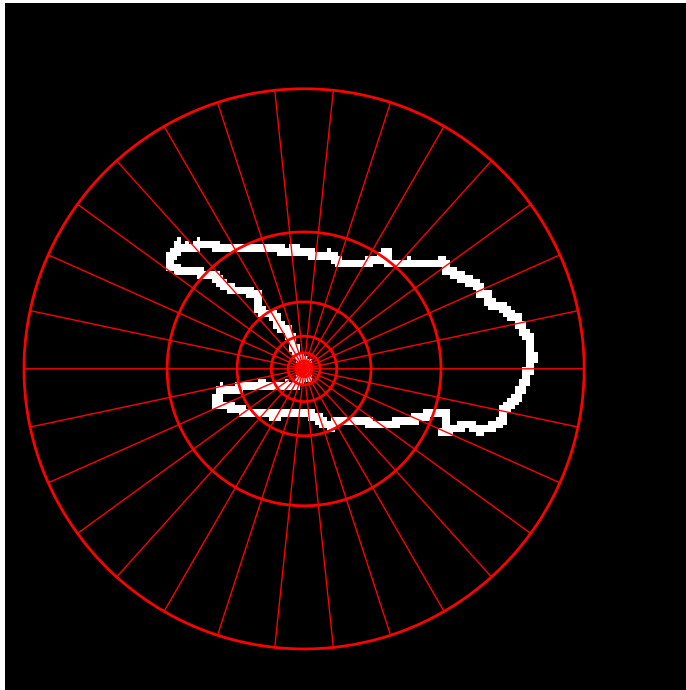
- 2 Έχοντας την περιγραφή για κάθε σημείο από το προηγούμενο βήμα, ακολουθεί η διαδικασία του ταιριάσματος. Στο βήμα αυτό προσπαθούμε να βρούμε την καλύτερη αντιστοίχιση σημείων μεταξύ των δυο σχημάτων σύμφωνα με την απόσταση

των περιγραφών τους. Έστω ότι έχουμε περιορίσει το πλήθος των σημείων αναφοράς σε κάθε σχήμα σε n και δημιουργούμε έναν $n \times n$ πίνακα αποστάσεων των περιγραφών τους, τα στοιχεία του οποίου δίνονται από την σχέση:

$$C_{ij} = \frac{1}{2} \sum_{k=1}^K \frac{[h_i[k] - h_j[k]]^2}{h_i[k] + h_j[k]} \quad (\chi^2 \text{ απόσταση})$$

,όπου h_i, h_j τα ιστογράμματα που προκύπτουν για τα σημεία p_i και p_j αντίστοιχα. Με την προϋπόθεση της 1-1 αντιστοίχισης, απαιτείται χρήση του αλγόριθμου *hungarian* που βρίσκει την αντιστοίχιση με το μικρότερο συνολικό (αθροιστικό) κόστος σύμφωνα με τον πίνακα αποστάσεων.

- 3 Η προαναφερθείσα διαδικασία δεν αρκεί για ένα ποιοτικό ταίριασμα, καθώς είναι πιθανό ένα υποσύνολο των αντιστοιχίσεων να μην ανταποκρίνεται σε λογικές διατάξεις των σημείων του περιγράμματος. Συνεπώς, μετά την αρχική αντιστοίχιση απαιτείται έλεγχος των αποτελεσμάτων έτσι να ικανοποιούνται ορισμένοι περιορισμοί, όπως για παράδειγμα γειτονικά σημεία του περιγράμματος του ενός σχήματος πρέπει να αντιστοιχούν σε κοντινά σημεία στο περίγραμμα του δεύτερου σχήματος.



Σχήμα 2.3: Παράδειγμα Logpolar διαμέρισης του χώρου για ένα σημείο του περιγράμματος για την κατασκευή του αντιστοίχου ιστογράμματος κατανομής των σημείων του περιγράμματος σε σχέση με το σημείο αναφοράς

Όπως είναι εύκολα αντιληπτό η δύναμη της μεθόδου, αλλά και πηγή διάφορων μειονεκτημάτων είναι η διαδικασία του ταίριασματος. Αναλυτικότερα, πέρα από την ίδια την περιγραφή για κάθε σημείο που έχει ένα βαθμό ανεξαρτησίας ως προς τοπικές ιδιαιτερότητες, τα προτερήματα της μεθόδου είναι :

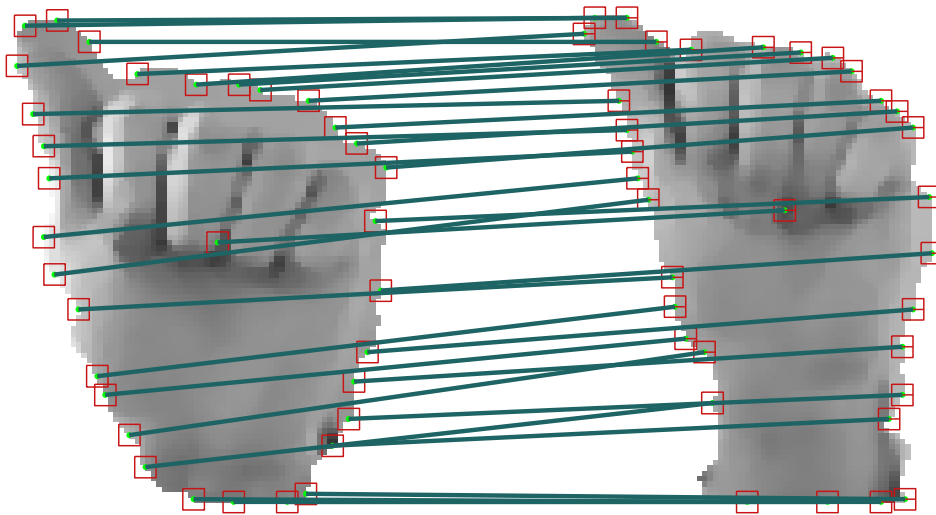
- ανεξαρτησία ως προς την περιστροφή, αφού η διαδικασία του ταιριάσματος δεν επιβάλλει κάποια σειρά αντιστοίχισης των σημείων.
- ανεκτικότητα ως προς παραμορφώσεις

Στον αντίποδα βέβαια η ίδια η διαδικασία του ταιριάσματος έχει τις εξής επιπτώσεις

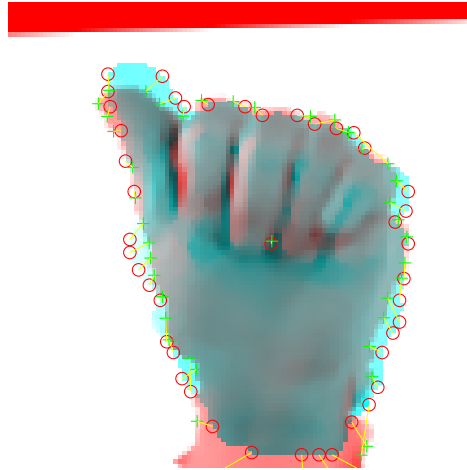
- η μέθοδος, έτσι όπως δίνεται, περιορίζεται για σύγκριση μεταξύ σχημάτων και όχι για την δημιουργία ενός ανεξάρτητου διανύσματος χαρακτηριστικών.
- μεγάλη υπολογιστική πολυπλοκότητα σε σχέση με άλλες μεθόδους. Ο αλγόριθμος *hungarian* έχει πολυπλοκότητα $O(n^3)$, όπου n το πλήθος των σημείων αναφοράς κάθε σχήματος.

Προφανώς στην πράξη δεν θα χρησιμοποιήσουμε όλα τα σημεία του περιγράμματος ως σημεία αναφοράς, αλλά ένα μικρότερο υποσύνολο (στα πειράματα που διεξήχθησαν η επικρατέστερη επιλογή ήταν $n = 40$), το οποίο επιλέχθηκε κρατώντας σημεία του περιγράμματος ανά ίσα διαστήματα. Μάλιστα δοκιμάστηκε ως υποσύνολο, τα σημεία που χαρακτηρίζονται ως γωνίες σύμφωνα με την μέθοδο του *Harris (Harris Detector)*, αλλά τα αποτελέσματα ήταν συγκρίσιμα και για λόγους απλότητας προτιμήθηκε η προηγούμενη τεχνική.

Ουσιαστικά η μέθοδος αυτή μας προσφέρει την δυνατότητα εύρεσης ενός αφινικού μετασχηματισμού μεταξύ των δυο σχημάτων και ένα παράδειγμα της λειτουργίας της παρουσιάζεται στα σχήματα 2.4 και 2.5, όπου γίνεται εμφανής η ικανότητα της μεθόδου να προσαρμόζει ένα σχήμα στο σχήμα αναφοράς.



Σχήμα 2.4: Ταίριασμα των σημείων της περιφέρειας των δυο χειρομορφών σύμφωνα με την μέθοδο των *shape context* και της μεθόδου *hungarian*



Σχήμα 2.5: Οπτικοποίηση του αφινικού μετασχηματισμού της δεύτερης εικόνας ως προς την πρώτη και υπερτιθέμενη σε αυτή. Ο αφινικός μετασχηματισμός έχει υπολογιστεί σύμφωνα με το ταίριασμα των σημείων της εικόνας 2.4

Η εκτίμηση του γεωμετρικού (αφινικού) μετασχηματισμού από την 1η εικόνα στην 2η σύμφωνα με το σύνολο των ζευγών σημείων που προέκυψαν από το βήμα της αντιστοίχισης μπορεί να πραγματοποιηθεί με την χρήση της τεχνικής του μέσου τετραγωνικού σφάλματος (*MSE*).

Ο πίνακας του αφινικού μετασχηματισμού έχει 6 παραμέτρους προς εκτίμηση και γράφεται ως:

$$A = \begin{bmatrix} a_1 & a_2 & a_3 \\ a_4 & a_5 & a_6 \\ 0 & 0 & 1 \end{bmatrix}$$

,σύμφωνα με τον οποίο το σημείο (x, y) μετασχηματίζεται στο (x', y') σύμφωνα με τις ακόλουθες σχέσεις :

$$x' = a_1x + a_2y + a_3, y' = a_4x + a_5y + a_6$$

ή αλλιώς σε πιο συμπαγή μορφή:

$$q = Ap, \quad \text{όπου } p = [x \quad y \quad 1]^T \quad \text{και} \quad q = [x' \quad y' \quad 1]^T$$

Θεωρούμε το σύνολο των σημείων P με κάθε γραμμή να γράφεται ως $p_i = [x_i \quad y_i \quad 1]$ και αντίστοιχα το Q . Το πρόβλημα πλέον γράφεται:

$$PA^T = Q$$

και στην συνέχεια, σύμφωνα με την *MSE* εκτίμηση, ο πίνακας A υπολογίζεται ως:

$$A = (P^+Q)^T$$

,όπου $P^+ = (P^T P)^{-1} P^T$ ο ψευδοαντίστροφος πίνακας του P .

Βέβαια, πρέπει να αναφερθεί ότι στην δημοσίευση, όπου πρωτοεισήχθησαν τα shape context, χρησιμοποιείται ένας μετασχηματισμός που βασίζεται στο μοντέλο *Thin Plate Splines (TSP)* και φαίνεται να είναι πιο αποδοτικός στην περιγραφή μετασχηματισμών σχήματος. Εντούτοις, στην παρούσα διπλωματική δεν χρησιμοποιείται η τεχνική αυτή τόσο για λόγους απλότητας, όσο και για πρακτικούς λόγους. Συγκεκριμένα, σε επόμενο κεφάλαιο θα προσπαθήσουμε να εξάγουμε με χρήση των shape context έναν γεωμετρικό μετασχηματισμό για να τον χρησιμοποιήσουμε μετέπειτα στην ίδια την εικόνα.

Κεφάλαιο 3

Εξαγωγή Οπτικών Χαρακτηριστικών Εμφάνισης

Όπως είναι λογικό, μονό ένα μικρό σύνολο εφαρμογών αναγνώρισης αντικειμένων είναι δυνατόν να αντιμετωπιστεί χρησιμοποιώντας μόνο την πληροφορία του περιγράμματος ενός αντικειμένου. Στην πράξη, η πλειονότητα των προβλημάτων αναγνώρισης αντικειμένων απαιτούν όλη την διαθέσιμη πληροφορία και το περίγραμμα τους αποτελεί μόνο μια επιμέρους ροη πληροφορίας.

Η αξιοποίηση της ολικής πληροφορίας θα επιτευχθεί με την εξαγωγή τοπικών χαρακτηριστικών της εικόνας και πιο συγκεκριμένα θα επικεντρωθούμε σε τεχνικές που βασίζονται σε τοπικές περιγραφές της παραγωγού της εικόνας. Για τους σκοπούς της μελέτης μας μπορούμε να κατηγοριοποιήσουμε τις τεχνικές αυτές σε δυο μεγάλες κατηγορίες :

- *Πυκνοί Περιγραφείς (Dense Descriptors)* , όπου το τελικό διάνυσμα χαρακτηριστικών αποτελείται από την σύνεωση τοπικών περιγραφών, πυκνά κατανεμημένων στην έκταση της εικόνας. Έχουν εφαρμοστεί με επιτυχία σε εφαρμογές ανιχνεύσης αντικειμένων ([8],[10]).
- *Τεχνικές Βασισμένες σε Σημεία Ενδιαφέροντος* , όπου η περιγραφή κάθε εικόνας αποτελείται από ένα σύνολο τοπικών χαρακτηριστικών που εξάγονται γύρω από σημεία ενδιαφέροντος (interest/key points). Τα σημεία αυτά θεωρητικά συγκεντρώνουν γύρω τους την 'σημαντικότερη' πληροφορία για την περιγραφή της εικόνας και αποτελούν χαρακτηριστικά γνωρίσματα αυτής. Οι τεχνικές που βασίζονται σε σημεία ενδιαφέροντος είναι ιδιαίτερα διαδεδομένες, με εφαρμογές σε ένα μεγάλο εύρος προβλημάτων (object recognition, motion tracking, multiview geometry).

3.1 HOG

Όπως αναφέραμε και στην εισαγωγή μια από τις πιο δημοφιλείς σύγχρονες τεχνικές στον τομέα της αναγνώρισης αντικειμένων εισήχθη από τους Dalal και Triggs στο [8] για τον σκοπό της αναγνώρισης ατόμων σε γενικές εικόνες. Η μέθοδος αυτή ονομάστηκε Histogram of Oriented Gradients (HOG), καθώς βασίζεται σε ιστογράμματα βασισμένα στις κατεύθυνσης της πρώτης παραγώγου της εικόνας. Παρά την απλότητα της μεθόδου, τα αποτελέσματα είναι εξαιρετικά καλά σε ένα μεγάλο εύρος εφαρμογών αναγνώρισης και για αυτό τον λόγο η μέθοδος αυτή αποτελεί βασικό δομικό συστατικό σε πολλές πιο σύνθετες σύγχρονες τεχνικές ([10],[19]).

Η βασική μέθοδος για μια εικόνα $I(x, y)$ (intensity image) αποτελείται από τα ακόλουθα βήματα

1 Υπολογισμός Παραγώγου

Αρχικά υπολογίζουμε τις μερικές παραγώγους g_x και g_y της εικόνας $I(x, y)$ (κατευθυντικές παράγωγοι ως προς τον άξονα x και y αντίστοιχα). Ο υπολογισμός αυτός προσεγγίζεται στις εικόνες (λόγω της διακριτής τους φύσης) με την χρήση κάποιου φίλτρου διαφορών (συνέλιξη της εικόνας με κάποιο πυρήνα) που προσεγγίζουν την έννοια της παραγώγου. Οι συνηθέστερες επιλογές είναι

- Τελεστής Κεντρικών Διαφορών (Central Difference Operator)

$$g_x = [-1, 0, 1] * I(x, y) \quad \text{και} \quad g_y = [-1, 0, 1]^T * I(x, y)$$

- Τελεστής Sobel

$$g_x = \begin{bmatrix} -1 & 0 & +1 \\ -2 & 0 & +2 \\ -1 & 0 & +1 \end{bmatrix} * I(x, y) \quad \text{και} \quad g_y = \begin{bmatrix} +1 & +2 & +1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix} * I(x, y)$$

- Τελεστής Prewitt

$$g_x = \begin{bmatrix} -1 & 0 & +1 \\ -1 & 0 & +1 \\ -1 & 0 & +1 \end{bmatrix} * I(x, y) \quad \text{και} \quad g_y = \begin{bmatrix} +1 & +1 & +1 \\ 0 & 0 & 0 \\ -1 & -1 & -1 \end{bmatrix} * I(x, y)$$

- Τελεστής Roberts

$$g_x = \begin{bmatrix} +1 & 0 \\ 0 & -1 \end{bmatrix} * I(x, y) \quad \text{και} \quad g_y = \begin{bmatrix} 0 & +1 \\ -1 & 0 \end{bmatrix} * I(x, y)$$

Στην βασική δημοσίευση των Dalal και Triggs φαίνεται πως καλύτερα αποτελέσματα επιτυγχάνονται χρησιμοποιώντας τον απλό τελεστή των κεντρικών διαφορών για το πρόβλημα αναγνώρισης που πραγματεύονται (pedestrian detection). Παρόμοια αποτελέσματα παρουσιάζουμε και στο πειραματικό κομμάτι για το πρόβλημα των στατικών χειρομορφών.

Η συγκεκριμένη μέθοδος απαιτεί την παράγωγο σε τέτοια μορφή ώστε να γίνει χρήση της πληροφορίας για την κατεύθυνση της (orientation) και για τον λόγο αυτό την υπολογίζουμε σε πολικές συντεταγμένες. Συνεπώς για κάθε σημείο (x, y) της εικόνας υπολογίζουμε το μέτρο $m(x, y)$ και την γωνία $\vartheta(x, y)$ της παραγώγου με τις παρακάτω σχέσεις

$$m(x, y) = \sqrt{g_x^2(x, y) + g_y^2(x, y)} \quad \text{και} \quad \vartheta(x, y) = \arctan \frac{g_y(x, y)}{g_x(x, y)}$$

2 Χωρική Διαμέριση της Εικόνας

Η εικόνα διαμερίζεται σε τμήματα-περιοχές από όπου επιθυμούμε να εξάγουμε κάποια τοπικά χαρακτηριστικά (κατεύθυνση της παραγώγου εν προκειμένω). Ο χωρισμός αυτός της εικόνας για τους σκοπούς της διπλωματικής γίνεται με ισόποση οριζόντια και κάθετη διαμέριση της εικόνας έτσι ώστε να προκύψουν οι ζητούμενες περιοχές ή κελιά (cells) όπως ονομάζονται στην δημοσίευση. Ένα επιπρόσθετο σημαντικό στοιχείο που αποτελεί μια παράμετρο της μεθόδου είναι η επικάλυψη των κελιών. Βέβαια οι Dalal και Triggs χωρίζουν αρχικά την εικόνα σε ισομεγέθη κελιά (6x6 pixels το καθένα) και αυτά τα ομαδοποιούν σε τετραγωνικά είτε κυκλικά επικαλυπτόμενα blocks (R-HOG και C-HOG) για λόγους σταθερότητας των περιγραφών που έχουν να κάνουν με την φύση του προβλήματος που πραγματεύονται (μεγάλες αλλαγές αντίθεσης και φωτεινότητας) και δεν επηρεάζουν σε τόσο μεγάλο βαθμό την παρούσα μελέτη.

3 Δημιουργία Ιστογραμμάτων Κατεύθυνσης

Για κάθε κελί δημιουργούμε ένα ιστόγραμμα κατεύθυνσης χωρίζοντας το εύρος των γωνιών - κατευθύνσεων σε ίσα τμήματα, όπου σε κάθε ομάδα κατεύθυνσης (orientation bin) του ιστογράμματος προσθέτουμε τα pixel που αντιστοιχούν στο εύρος γωνιών της συγκεκριμένης ομάδας με ψήφο ίση με το μέτρο της παραγώγου του εκάστοτε pixel. Το ιστόγραμμα το χωρίζουμε σε ίσες ομάδες κατεύθυνσης - bins σύμφωνα με την γωνία που υπολογίσαμε στο αρχικό στάδιο με το πλήθος των bins να αποτελεί μια από τις παραμέτρους της μεθόδου. Το αρχικό εύρος γωνιών μπορεί να είναι $[0, 180^\circ]$ (unsigned gradient) είτε $[0, 360^\circ]$ (signed gradient).

4 Κανονικοποίηση Ιστογραμμάτων

Η αλλαγές στην φωτεινότητα (illumination) είναι ικανές να προκαλέσουν μεγάλες αλλαγές στην περιγραφή των HOG καθώς επηρεάζουν σημαντικά το μέτρο της παραγώγου. Για τον λόγο αυτό απαιτείται κανονικοποίηση (normalization) των ιστογραμμάτων, η οποία εφαρμόζεται με χρήση της L2 νόρμας. Έστω το αρ-

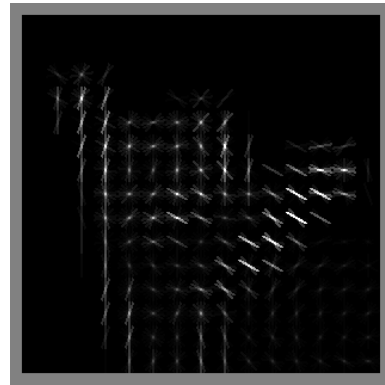
χικό ιστόγραμμα h , τότε το νέο κανονικοποιημένο ιστόγραμμα υπολογίζεται συμφώνα με την σχέση

$$h' = \frac{h}{\sqrt{\|h\|^2 + \epsilon}}, \quad \text{όπου } \epsilon \text{ μια πολύ μικρή σταθερά}$$

Η τελική περιγραφή της εικόνας σύμφωνα με την μέθοδο HOG είναι η συνένωση όλων των ιστογραμμάτων που προέκυψαν από την προαναφερθείσα διαδικασία. Η περιγραφή αυτή έχει τον χαρακτήρα πυκνής δειγματοληψίας των ζητούμενων οπτικών χαρακτηριστικών και είναι ενιαία για την εικόνα. Στο σχήμα 3.1 (β) παρουσιάζεται η οπτικοποίηση των τοπικών ιστογραμμάτων για την εικόνα στο (α).



(α')



(β')

Σχήμα 3.1: Οπτικοποίηση της HOG περιγραφής για κάθε κελί

Η εκπαίδευση και η αναγνώριση ενός αντικειμένου σε μια εικόνα γίνεται με την χρήση γραμμικών Support Vector Machines (SVM), που προσπαθούν να βρουν το βέλτιστο υπερεπίπεδο που διαχωρίζει τις δυο κλάσεις ναι/όχι για το αν υπάρχει το ζητούμενο αντικείμενο. Επειδή η φύση του προβλήματος που μελετάμε είναι διαφορετική, θα χρησιμοποιήσουμε για τους σκοπούς της αναγνώρισης - ταξινόμησης τις τεχνικές που θα περιγράψουμε στο επόμενο κεφάλαιο.

3.2 SIFT

Όπως αναφέραμε ήδη στην αρχή του κεφαλαίου η μέθοδος SIFT είναι η πιο δημοφιλής τεχνική που βασίζεται σε σημεία ενδιαφέροντος και έχει αποτελέσει την κεντρική ιδέα πίσω από πολλές σύγχρονες τεχνικές αναγνώρισης αντικειμένων. Επειδή

στην παρούσα μελέτη δεν γίνεται χρήση τέτοιων τεχνικών θα γίνει μια προσπάθεια παρουσίασης των κυριών συστατικών της μεθόδου, χωρίς να υπεισέλθουμε σε πολλές τεχνικές λεπτομέρειες.

Η δεσπύζουσα ιδέα που εκμεταλλεύονται τα SIFT ([21]) είναι ότι κάθε αντικείμενο έχει κάποια ιδιαίτερα χαρακτηριστικά, τα οποία ορίζονται ως σημεία ενδιαφέροντος στην εικόνα και τοπικές περιγραφές τους χρησιμοποιούνται για την δημιουργία ενός συνόλου χαρακτηριστικών για το εκάστοτε αντικείμενο που το περιγράφουν με αρκετή επιτυχία. Χωρίζουμε λοιπόν την όλη διαδικασία σε τρεις σημαντικές κατηγορίες:

- **Εύρεση Σημείων Ενδιαφέροντος**

Ως σημεία ενδιαφέροντος από άποψης όρασης υπολογιστών μπορούμε να θεωρήσουμε τις ακμές και τις κυρίως γωνίες. Μια κλασική τεχνική υπολογισμού τέτοιων σημείων σε μια εικόνα γίνεται με χρήση του τελεστή Laplacian of Gaussian (LoG), ενώ για εύρεση γωνιών μόνο χρησιμοποιείται εκτενώς ο Harris Corner Detector. Η χρήση της LoG υπολογίζει προσεγγιστικά την δεύτερη παράγωγο της εικόνας για συγκεκριμένη κλίμακα που εξαρτάται από το σ της γκαουσιανής και τα σημεία ενδιαφέροντος προκύπτουν ως τα ακρότατα της αναπαράστασης που προκύπτει μετά τον υπολογισμό της Laplacian of gaussian της εικόνας. Επιπλέον είναι επιθυμητό να βρούμε σημεία ενδιαφέροντος και στις διάφορες κλίμακες (scale-space) και για αυτό επαναλαμβάνουμε την διαδικασία για κάποιο εύρος τιμών του σ της γκαουσιανής ¹.

Συνήθως το παραπάνω βήμα δεν είναι αρκετό για να έχουμε ένα αξιόπιστο σύνολο σημείων ενδιαφέροντος και για αυτό είναι συνετό να ακολουθεί ένα στάδιο ελέγχου των σημείων που προέκυψαν και να απορρίπτονται όσα θεωρούμε ότι δεν προσφέρουν ουσιαστική πληροφορία είτε είναι πιθανό να μας δώσουν παραπλανητική πληροφορία (συνήθως πολύ κοντινά σημεία μεταξύ τους ή μεγάλο σύνολο σημείων σε ακμές και σημείων με χαμηλή αντίθεση). Η επιλογή των 'καλών' σημείων είναι πρωτεύουσας σημασίας για την καλή συμπεριφορά της τελικής περιγραφής.

- **Εξαγωγή Τοπικών Περιγραφών**

Στο βήμα αυτό εξάγονται τα οπτικά χαρακτηριστικά γύρω από τις περιοχές των σημείων ενδιαφέροντος, όπου το εύρος της περιοχής του εκάστοτε σημείου εξαρτάται από την κλίμακα που εξήχθη το σημείο αυτό. Ως τοπική περιγραφή επιλέγουμε την αναπαράσταση σε ιστογράμματα κατευθύνσεων, όπως τα περιγράψαμε στα HOGs. Για ανεξαρτησία στις περιστροφές, που είναι ένα επιπλέον χαρακτηριστικό του SIFT, βρίσκουμε τις κυρίαρχες κατευθύνσεις της παραγώγου για την εκάστοτε περιοχή ως κορυφές του ιστογράμματος κατευθύνσεων (συνήθως επιλέγονται αυτές οι οποίες έχουν τιμή 80% της μεγίστης). Για κάθε μια από τις κυρίαρχες κατευθύνσεις εξάγουμε μια διαφορετική περιγραφή θεωρώντας την κατεύθυνση αυτή ως την κατεύθυνση αναφοράς, δηλαδή αφαιρούμε

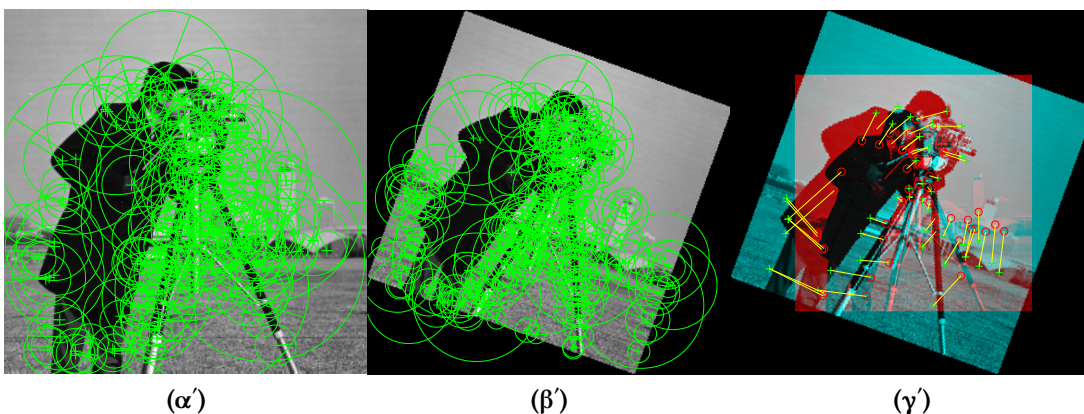
¹Για τον σκοπό αυτό έχουν δημιουργηθεί υπολογιστικά γρήγορες προσέγγισης, όπως Difference of Gaussians (DoG) που χρησιμοποιείται στον SIFT

από τις γωνίες της πολικής περιγραφής της παραγώγου την αντίστοιχη γωνία της κυρίαρχης κατεύθυνσης και υπολογίζουμε εκ νέου το ιστόγραμμα κατευθύνσεων.

- **Ομαδοποίηση Τοπικών Χαρακτηριστικών**

Επιχειρείται ομαδοποίηση των χαρακτηριστικών που εξήχθησαν στο προηγούμενο βήμα για την δημιουργία μιας ενιαίας περιγραφής για το κάθε αντικείμενο. Αυτή η ομαδοποίηση πραγματοποιείται σύμφωνα με το μοντέλο Bag of (Visual) Words - BoW. Το μοντέλο αυτό ουσιαστικά κωδικοποιεί σε ένα ιστόγραμμα την συχνότητα εμφάνισης συγκεκριμένων ομάδων χαρακτηριστικών, οι οποίες προκύπτουν από κάποια τεχνική clustering (π.χ. k-means clustering) στο σύνολο των τοπικών περιγραφών εκπαίδευσης. Αυτή η τεχνική έχει αξιολογήσει αποτελέσματα σε προβλήματα κατηγοριοποίησης αντικειμένων, όπου κάθε αντικείμενο-κλάση αντιπροσωπεύεται από ένα ιστόγραμμα εμφάνισης k διακριτών ομάδων χαρακτηριστικών. Βέβαια το στάδιο αυτό δεν αποτελεί απαραίτητα βήμα μιας μεθόδου που βασίζεται σε σημεία ενδιαφέροντος αλλά σχετίζεται άμεσα με την εκάστοτε εφαρμογή, όπως στην περίπτωση ευθυγράμμισης γειτονικών εικόνων (image alignment) όπου απαιτείται κάποια μορφή αντιστοίχισης μεταξύ των τοπικών χαρακτηριστικών γύρω από τα σημεία ενδιαφέροντος.

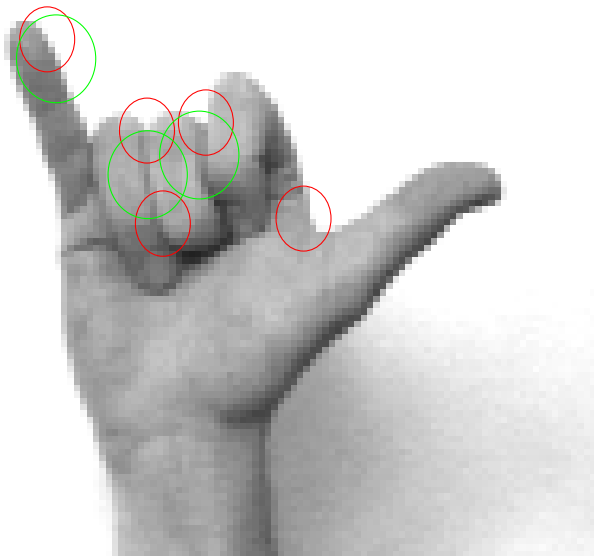
Η μέθοδος αυτή είναι ιδιαίτερα δημοφιλής καθώς προσφέρει ανεξαρτησία τόσο σε αλλαγές κλίμακας και περιστροφής, όπως φαίνεται στο σχήμα 3.2 όπου γίνεται χρήση των SURF, μιας παραλλαγής των SIFT. Επιπλέον έχει αξιολογήσει αποτελέσματα σε περιπτώσεις προβολών (projections) και μερικής επικάλυψη αντικειμένων (occlusions). Η ισχύς της προέρχεται από το γεγονός ότι η αναγνώριση αντικειμένων βασίζεται σε πολύ μεγάλο βαθμό στην αναγνώριση σημαντικών τοπικών χαρακτηριστικών και η ικανότητα αναγνώρισης τοπικών χαρακτηριστικών δεν μειώνεται για παραμορφώσεις της εικόνας (υπό περιορισμούς στον βαθμό παραμόρφωσης).



Σχήμα 3.2: Παράδειγμα εφαρμογής των SURF. (α),(β) : η αρχική εικόνα και νέα που δημιουργήθηκε με περιστροφή και αύξηση της κλίμακας. Στις εικόνες αυτές φαίνονται και τα σημεία ενδιαφέροντος, όπως και η κλίμακα τους που είναι ανάλογη με την ακτίνα του κύκλου, (γ) : υπέρθεση των δυο εικόνων και οπτικοποίηση της αντιστοίχισης των σημείων ενδιαφέροντος

3.3 Χειρομορφές ως Αντικείμενα Αναγνώρισης

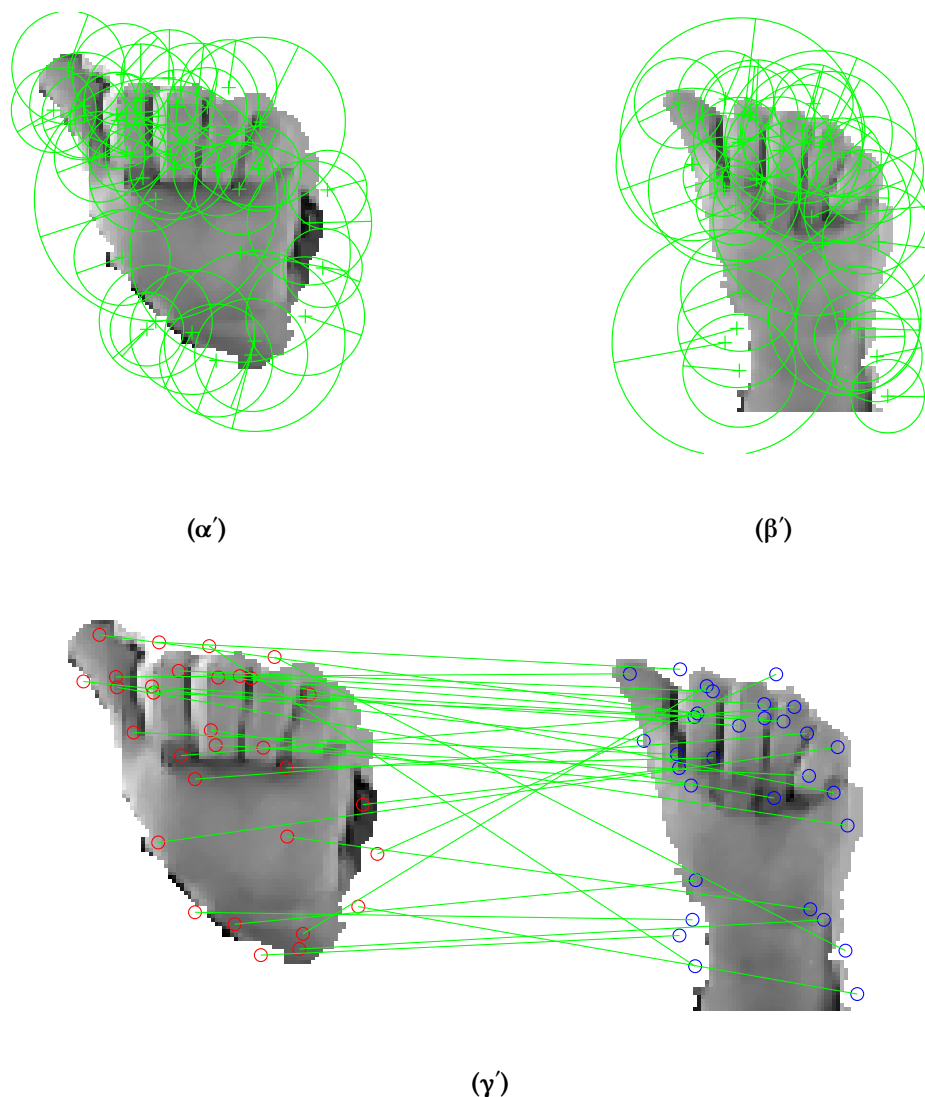
Οι χειρομορφές αποτελούν ένα ιδιόρρυθμο είδος αντικειμένων, καθώς για διαφορετικές κλάσεις χρησιμοποιούμε την ίδια εσωτερική πληροφορία, δηλαδή τα χαρακτηριστικά ενός χεριού, αλλάζοντας κάθε φορά την διάταξη της στον χώρο (θέση δαχτύλων, κατεύθυνση χεριού). Η παραπάνω διαπίστωση καθίστα της τεχνικές που βασίζονται σε σημεία ενδιαφέροντος ιδιαίτερα αναποτελεσματικές, καθώς πλέον όλες οι χειρομορφές μοιράζονται τα ίδια τοπικά χαρακτηριστικά και όλη η ουσιώδης πληροφορία για τον διαχωρισμό των κλάσεων βρίσκεται στο πως κατανέμονται τα χαρακτηριστικά αυτά στον χώρο, δηλαδή οι χωρικές τους συσχετίσεις. Ο συλλογισμός αυτός επιβεβαιώνεται στο σχήμα 3.3, όπως και στο σχήμα 3.4, όπου παρουσιάζονται τα αποτελέσματα της αντιστοίχισης (matching) των σημείων ενδιαφέροντος για δυο χειρομορφές της ίδιας κλάσης. Επιπλέον το μοντέλο Bag of Visual Words που αναφέραμε ως την βασική τεχνική για κατασκευή ενιαίων περιγραφών δεν λαμβάνει υπόψιν καμία χωρική πληροφορία και είναι ακατάλληλο για τέτοιου είδους προβλήματα.



(α')

Σχήμα 3.3: Παράδειγμα ομοιότητας τοπικών χαρακτηριστικών για μια χειρομορφή σε διαφορετικές κλίμακες

Από την άλλη οι πυκνή περιγραφή που προσφέρουν τα HOG είναι ιδανική για τέτοιου είδους προβλήματα, καθώς το τελικό διάνυσμα που προκύπτει έχει άμεση χωρική ερμηνεία που αντιστοιχεί στο πλέγμα που χρησιμοποιήθηκε για την εξαγωγή της τοπικής πληροφορίας. Εντούτοις η τεχνική των HOG είναι αρκετά απλή και για τον λόγο



Σχήμα 3.4: Παράδειγμα εφαρμογής των SURF. (α),(β) : δυο εικόνες χειρομορφών της ίδιας κλάσης (αναπαριστούν το γράμμα Α). Στις εικόνες αυτές φαίνονται και τα σημεία ενδιαφέροντος, όπως και η κλίμακα τους που είναι ανάλογη με την ακτίνα του κύκλου, (γ) : το αποτέλεσμα του ταιριάσματος των σημείων των δυο εικόνων

αυτό είναι ιδιαίτερα ευαίσθητη σε μικρές παραμορφώσεις (όπως περιστροφές), θέτοντας έτσι περιορισμούς στην χρήση τους.

Για όλους τους παραπάνω λόγους στην συνέχεια της διπλωματικής θα ασχοληθούμε με την μέθοδο των HOGs και πιθανές παραλλαγές για πιο σταθερές και ανεξάρτητες περιγραφές βασισμένες σε HOG, απορρίπτοντας αυτές που βασίζονται σε σημεία ενδιαφέροντος ως ακατάλληλες για το συγκεκριμένο πρόβλημα.

Κεφάλαιο 4

Τεχνικές Αναγνώρισης Προτύπων - Ταξινομητές

Στο κεφάλαιο αυτό ασχολούμαστε με την επεξεργασία - μετασχηματισμούς δεδομένων και τις τεχνικές ταξινόμησης που έπονται μετά την εξαγωγή των διανυσμάτων χαρακτηριστικών και αποτελούν ένα μέτρο αξιολόγησης της ικανότητας και της ισχύος της εκάστοτε μεθόδου εξαγωγής χαρακτηριστικών.

Συγκεκριμένα, αρχικά θα αναφερθούμε στις δυο βασικότερες τεχνικές μετασχηματισμού δεδομένων με στόχο της μείωση του αριθμού διαστάσεων των δεδομένων, PCA και LDA, και στην συνέχεια θα περιγράψουμε τους ταξινομητές που θα χρησιμοποιηθούν σε αυτή διπλωματική.

4.1 Μείωση Αριθμού Διαστάσεων

Η εξαγωγή χαρακτηριστικών και ειδικότερα η εξαγωγή πυκνών οπτικών χαρακτηριστικών που πραγματεύεται μεγάλο μέρος της παρούσας διπλωματικής, είναι πολύ πιθανό να δημιουργήσει μεγάλα διανύσματα χαρακτηριστικών με μεγάλο πλεονασμό στην ουσιώδη πληροφορία. Συνεπώς συνίσταται ο μετασχηματισμός των αρχικών δεδομένων τόσο για λόγους υπολογιστικής πολυπλοκότητας, όσο και για λόγους που σχετίζονται άμεσα με την εφαρμοσιμότητα κάποιων τεχνικών ταξινόμησης, όπως θα δούμε και στην συνέχεια.

Στόχος μας είναι λοιπόν να προκύψει ένα νέο σύνολο χαρακτηριστικών που να κωδικοποιούν με τον καλύτερο δυνατό τρόπο τα αρχικά χαρακτηριστικά. Αυτό σημαίνει μείωση του πλήθους των χαρακτηριστικών (Dimensionality Reduction) για συγκεκριμένο πρόβλημα ταξινόμησης, έχοντας βέβαια ως γνώμονα την όσο το δυνατόν μικρότερη απώλεια πληροφορίας.

4.1.1 Ανάλυση σε Κύριες Συνιστώσες - PCA

Η Ανάλυση σε Κύριες Συνιστώσες (Principal Component Analysis - PCA) είναι μια από τις πιο δημοφιλείς τεχνικές μείωσης του αριθμού διαστάσεων. Στόχος του PCA είναι η εύρεση ενός γραμμικού μετασχηματισμού, έτσι ώστε, δεδομένου ενός διανύσματος x , το νέο διάνυσμα χαρακτηριστικών $y = A^T x$ να έχει τα στοιχεία του ασυσχέτιστα (ο μετασχηματισμός αυτός ονομάζεται και Karhunen Loeve Transform). Η εύρεση του μετασχηματισμού αποτελεί το πρώτο στάδιο της μεθόδου, καθώς στο δεύτερο στάδιο θα επιλεγθούν τα περισσότερα σημαντικά από τα στοιχεία αυτά έτσι ώστε να μειωθεί ο αριθμός των χαρακτηριστικών.

Ο αλγόριθμος του PCA για N διανύσματα d στοιχείων - χαρακτηριστικών $x_i, i = 1, 2, \dots, N$ συνοψίζεται ως εξής :

- Αφαίρεση της μέσης τιμής από το σύνολο των δεδομένων

$$\bar{x}_i = x_i - m, \quad i = 1, 2, \dots, N, \quad m = \frac{1}{N} \sum_{i=1}^N x_i$$

- Υπολογισμός του μητρώου αυτοσυσχέτισης R των (\bar{x}_i) , το οποίο ταυτίζεται με το μητρώο συνδιασποράς λόγω της μηδενικής μέσης τιμής (την έχουμε αφαιρέσει στο προηγούμενο βήμα). Ο υπολογισμός γίνεται σύμφωνα με την σχέση:

$$R = \frac{1}{N} \sum_{i=1}^N \bar{x}_i \bar{x}_i^T$$

- Υπολογισμός των ιδιοτιμών λ_i και των ιδιοδιανυσμάτων $a_i, i = 1, 2, \dots, d$, του μητρώου αυτοσυσχέτισης R .
- Ταξινόμηση των ιδιοτιμών λ_i σε φθίνουσα σειρά έτσι ώστε για την νέα διάταξη των ιδιοτιμών να ισχύει $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_d$.
- Επιλογή των μεγαλύτερων m ιδιοτιμών, $\lambda_1, \lambda_2, \dots, \lambda_m$, που ονομάζονται και κύριες συνιστώσες.
- Δημιουργία του γραμμικού μετασχηματισμού ως το σύνολο των ιδιοδιανυσμάτων που αντιστοιχούν στις ιδιοτιμές που επιλέχθηκαν προηγουμένως, δηλαδή

$$A = [a_1 \quad a_2 \quad \dots \quad a_m]$$

- Μετασχηματισμός των αρχικών d -διάστατων διανυσμάτων στον νέο m -διάστατο χώρο. Αγνοώντας την μέση τιμή είναι :

$$\bar{y} = A^T \bar{x}$$

, αλλιώς :

$$y = A^T x = A^T(\bar{x} - m) = \bar{y} - A^T m$$

Ουσιαστικά κάθε στοιχείο i του νέου διανύσματος y είναι η προβολή του αρχικού διανύσματος x πάνω στο ιδιοδιάνυσμα a_i .

Αξιίζει να σημειωθεί το προκύπτον μητρώο αυτοσυσχέτισης $R_y = A^T R_x A$ είναι διαγώνιο καθώς ισχύει :

$$a_i^T R_x a_j = \begin{cases} \lambda_i, & i = j \\ 0, & i \neq j \end{cases}$$

Αυτό σημαίνει πως τα στοιχεία του y είναι ασυσχέτιστα ενώ η διασπορά τους αντιστοιχεί στις ιδιοτιμές του R_x . Συνεπώς, εφόσον η συνολική διασπορά των στοιχείων του x είναι ίση με το άθροισμα των ιδιοτιμών λ_i , η επιλογή των ιδιοδιανυσμάτων που αντιστοιχούν στις μεγαλύτερες ιδιοτιμές προσφέρει μέγιστη δυνατή διασπορά του τελικού διανύσματος για το εκάστοτε m .

4.1.2 Ανάλυση Γραμμικής Διάκρισης - LDA

Η Ανάλυση Γραμμικής Διάκρισης (Linear Discriminant Analysis - LDA) είναι μια τεχνική μείωσης του αριθμού των διαστάσεων υπό επίβλεψη, δηλαδή με γνώση της κλάσης που ανήκει το κάθε δεδομένο, σε αντίθεση με τον PCA που είναι χωρίς επίβλεψη. Στόχος είναι η μείωση της διάστασης διατηρώντας όσο το δυνατόν περισσότερο την διακριτική ικανότητα μεταξύ των τάξεων. Αυτό επιτυγχάνεται με την εύρεση ιδιοδιανυσμάτων (αντίστοιχη λογική με PCA) κατά τις διευθύνσεις των οποίων οι τάξεις διαχωρίζονται καλύτερα και δημιουργώντας έτσι έναν γραμμικό μετασχηματισμό.

Για την περίπτωση δυο κλάσεων επιθυμούμε να βρούμε το διάνυσμα w , έτσι ώστε οι (μονοδιάστατες) προβολές y των διανυσμάτων x ως προς το w να μεγιστοποιούν τον λόγο διάκρισης Fisher (FDR) :

$$FDR = \frac{(\mu_1 - \mu_2)^2}{\sigma_1^2 + \sigma_2^2}$$

όπου μ_1, μ_2 είναι οι μέσες τιμές για κάθε κλάση και σ_1, σ_2 οι αντίστοιχες διασπορές.

Η μεγιστοποίηση του FDR ερμηνεύεται ως προσπάθεια για να απομακρύνουμε όσο το δυνατόν περισσότερο τις δυο κλάσεις και ταυτόχρονα να ελαχιστοποιήσουμε τις διασπορές των κλάσεων αυτών. Συνεπώς λαμβάνονται υπόψιν οι αποστάσεις τόσο ανάμεσα στις κλάσεις όσο και εσωτερικά των κλάσεων.

Για τον σκοπό αυτό ορίζουμε:

- Το μητρώο σκέδασης εντός των κλάσεων (within-class scatter matrix)

$$S_w = \frac{1}{2}(S_1 + S_2)$$

όπου S_1, S_2 είναι τα μητρώα συνδιασποράς των κλάσεων

- Το μητρώο σκέδασης μεταξύ κλάσεων (between class scatter matrix)

$$S_b = \frac{1}{2}(m_1 - m_0)(m_1 - m_0)^T + \frac{1}{2}(m_2 - m_0)(m_2 - m_0)^T$$

όπου m_1, m_2 είναι οι μέσες τιμές των δυο κλάσεων και m_0 η συνολική μέση τιμή των δεδομένων.

Με χρήση των παραπάνω το κριτήριο FDR γράφεται ως εξής (σύμφωνα με τον αρχικό χώρο των δεδομένων) :

$$FDR = \frac{w^T S_b w}{w^T S_w w}$$

Στην περίπτωση των δυο κλάσεων η λύση δίνεται απευθείας από την σχέση

$$w = S_w^{-1}(m_1 - m_2)$$

Στην περίπτωση c κλάσεων το πρόβλημα γενικεύεται βρίσκοντας $c - 1$ διευθύνσεις - διανύσματα προβολής τα οποία σχηματίζουν τον γραμμικό μετασχηματισμό $W = [w_1 \ w_2 \ \dots \ w_{c-1}]$. Πλέον οι πίνακες S_w και S_b ορίζονται ως εξής

$$S_w = \sum_{i=1}^c P_i S_i \quad S_b = \sum_{i=1}^c P_i (m_i - m_0)(m_i - m_0)^T$$

όπου P_i είναι οι αντίστοιχες a priori πιθανότητες για κάθε κλάση (N_i/N).

Το γενικευμένο κριτήριο χρησιμοποιεί τις οριζουσες των within-class και between-class μητρώων των προβολών (\widetilde{S}_w και \widetilde{S}_b), καθώς οι προβολές δεν είναι πλέον βαθμωτά μεγέθη, και ορίζεται ως

$$J = \frac{|\widetilde{S}_b|}{|\widetilde{S}_w|} = \frac{|W^T S_b W|}{|W^T S_w W|}$$

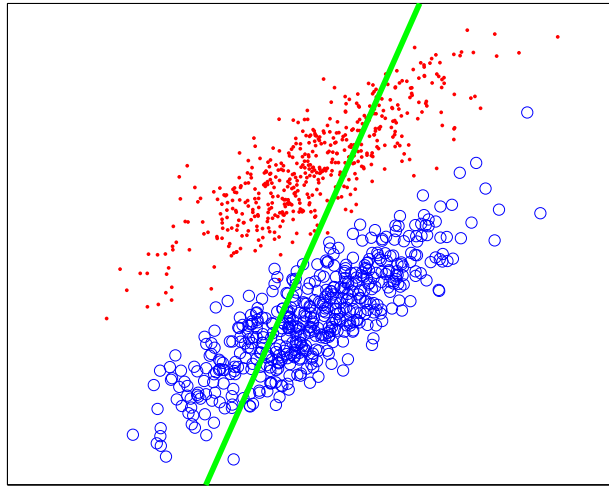
Το παραπάνω κριτήριο μεγιστοποιείται, σύμφωνα με το γενικευμένο πρόβλημα ιδιοτιμών, θέτοντας ως τις ζητούμενες προβολές w_i που απαρτίζουν τον μετασχηματισμό W τα ιδιοδιανύσματα που αντιστοιχούν στις $c - 1$ μη μηδενικές ιδιοτιμές του μητρώου $S_w^{-1} S_b$ (του οποίου η τάξη είναι το πολύ $c - 1$ αφού το S_b είναι τάξης $c - 1$).

Μια εναλλακτική προσέγγιση του LDA μπορεί να πραγματοποιηθεί για την περίπτωση κανονικών κατανομών σε κάθε κλάση με το ίδιο μητρώο συνδιασποράς, εκτιμώντας τις παραμέτρους των κατανομών με την τεχνική της μέγιστης πιθανοφάνειας. Η προσέγγιση αυτή ουσιαστικά αντιστοιχίζει το μητρώο S_w με το κοινό μητρώο συνδιασποράς.

4.1.3 Σύνοψη - Σύγκριση

PCA

- Ιδιοκατεύθυνση μέγιστης μεταβλητότητας (σχήμα 4.1 για την περίπτωση δυο κλάσεων κανονικής κατανομής).



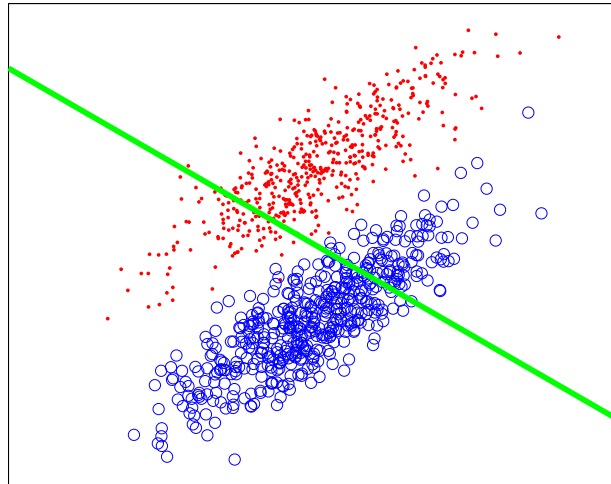
Σχήμα 4.1: PCA - το διάνυσμα προβολής είναι παράλληλο στην πράσινη ευθεία

- Καμία πληροφορία για τι κλάσεις (χωρίς επίβλεψη) - αντιμετωπίζει τα δεδομένα ως ένα ενιαίο σύνολο.

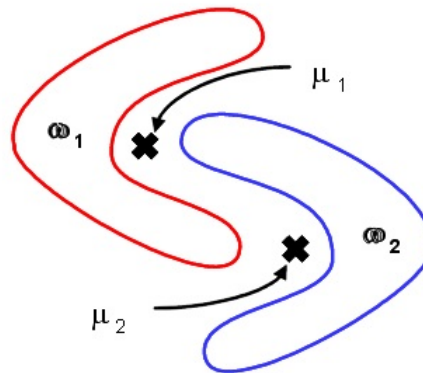
LDA

- Ιδιοκατεύθυνση μεγιστοποίησης διάκρισης μεταξύ των κλάσεων (σχήμα 4.2 για την περίπτωση δυο κλάσεων κανονικής κατανομής) .
- Υπόθεση κανονικής κατανομής (ανεπιτυχής εφαρμογή για άλλες κατανομές όπως στην περίπτωση του σχήματος 4.3).
- Κακή συμπεριφορά σε περιπτώσεις που η ουσιώδης πληροφορία για κάθε κλάση βρίσκεται στην διακύμανση και όχι στην μέση τιμή (παράδειγμα στο σχήμα 4.4).

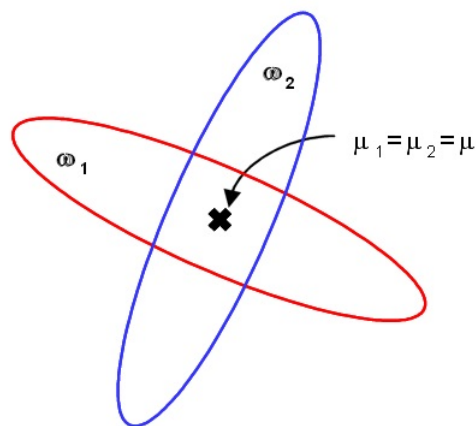
Συμπερασματικά αν ικανοποιούνται οι προϋποθέσεις του LDA (κάθε κλάση περιγράφεται από μια κανονική κατανομή, που ιδανικά έχουν τον ίδιο πίνακα συνδιασποράς, με την ουσιώδη πληροφορία για την διάκριση των κλάσεων να βρίσκεται στην μέση τιμή), γίνεται άμεσα μια πολύ ισχυρή τεχνική για μείωση των διαστάσεων, αρκετά πιο αξιόπιστη από PCA όπως γίνεται εμφανές στα σχήματα 4.1 και 4.2, διατηρώντας όσο τον δυνατόν περισσότερη διακριτική πληροφορία για τις κλάσεις.



Σχήμα 4.2: LDA - το διάνυσμα προβολής είναι παράλληλο στην πράσινη ευθεία



Σχήμα 4.3: Περίπτωση μη κανονικής κατανομής



Σχήμα 4.4: Περίπτωση διακριτικής πληροφορίας στην διακύμανση

4.2 Ταξινομητές

Στόχος της ενότητας αυτής είναι να περιγράψουμε τεχνικές για ταξινόμηση ενός προτύπου - διανύσματος χαρακτηριστικών σε μια από τις δυνατές κλάσεις, σύμφωνα με ένα σύνολο δεδομένων εκπαίδευσης για τα οποία γνωρίζουμε την κλάση που ανήκουν.

4.2.1 Θεωρία Αποφάσεων του Bayes

Δοθέντος ενός d -διάστατου διανύσματος x και ενός συνόλου c κλάσεων ω_i , σύμφωνα με το θεώρημα του Bayes ισχύει

$$P(\omega_i|x)p(x) = p(x|\omega_i)P(\omega_i)$$

και

$$p(x) = \sum_{i=1}^c p(x|\omega_i)P(\omega_i)$$

όπου

- $P(\omega_i)$ είναι η εκ των προτέρων (a priori) πιθανότητα της κλάσης ω_i .
- $P(\omega_i|x)$ είναι η εκ των υστέρων (a posteriori) πιθανότητα της κλάσης ω_i δοθέντος ενός διανύσματος x .
- $p(x)$ είναι η συνάρτηση πυκνότητας πιθανότητας του x .
- $p(x|\omega_i)$ είναι η συνάρτηση της υπό συνθήκη πυκνότητας πιθανότητας του x δοθείσης της κλάσης ω_i (καλείται και πιθανοφάνεια).

Η θεωρία αποφάσεων του Bayes ταξινομεί ένα διάνυσμα x στην κλάση ω_i εφόσον

$$P(\omega_i|x) > P(\omega_j|x), \quad \forall j \neq i \quad (4.1)$$

Η παραπάνω ανισότητα, χρησιμοποιώντας το θεώρημα του Bayes και θεωρώντας ότι η $p(x)$ είναι σταθερή και θετική για όλες τις κλάσεις, γράφεται ως εξής

$$p(x|\omega_i)P(\omega_i) > p(x|\omega_j)P(\omega_j), \quad \forall j \neq i \quad (4.2)$$

Στην πράξη ως συνάρτηση πυκνότητας πιθανότητας χρησιμοποιείται εκτενώς η Gaussian (λόγω της μαθηματικής της απλότητας αλλά και του θεωρήματος του κεντρικού ορίου) και είναι ικανή να μοντελοποιήσει με αρκετή επιτυχία υπό-συνθήκη πυκνότητες πιθανότητας σε πολλά πραγματικά προβλήματα αναγνώρισης - ταξινόμησης προτύπων.

Η gaussian συνάρτηση πυκνότητας πιθανότητας για πολλές διαστάσεις δίνεται από την σχέση

$$p(x) = \frac{1}{\sqrt{2\pi}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(x-m)^T\Sigma^{-1}(x-m)\right)$$

όπου $m = E[x]$ το μέσο διάνυσμα και $\Sigma = E[(x-m)(x-m)^T]$ ο πίνακας συνδιασποράς.

4.3 Ταξινομητές Ελάχιστης Απόστασης

Η πιο απλή τεχνική ταξινόμησης είναι η ανάθεση ενός προτύπου στην κλάση, στην οποία απέχει λιγότερο από το κέντρο της. Η απόσταση μπορεί να είναι η ευκλείδεια ή κάποια άλλη, όπως η histogram intersection ή η mahalanobis.

Είναι ιδιαίτερα ενδιαφέρον το γεγονός ότι ο βέλτιστος bayesian ταξινομητής μπορεί να ταυτιστεί με τους ταξινομητές ελάχιστης απόστασης για την περίπτωση της ευκλείδειας και της mahalanobis απόστασης υπό τις εξής παραδοχές

- Όλες οι κλάσεις έχουν την ίδια a-priori πιθανότητα. Με την υπόθεση αυτή η θεωρία αποφάσεων του Bayes εξαρτάται μόνο από τις υπό-συνθήκη κατανομές και γίνεται:

$$p(x|\omega_i) > p(x|\omega_j), \quad \forall j \neq i \quad (4.3)$$

- Τα δεδομένα σε κάθε κλάση ακολουθούν κανονική κατανομή (gaussian). Συνεπώς κάνοντας και χρήση του αρνητικού λογάριθμου της πιθανοφάνειας (negative log-likelihood) η προηγούμενη σχέση γίνεται

$$(x-m_i)^T\Sigma_i^{-1}(x-m_i) < (x-m_j)^T\Sigma_j^{-1}(x-m_j), \quad \forall j \neq i \quad (4.4)$$

ή αλλιώς

$$\sqrt{(x-m_i)^T\Sigma_i^{-1}(x-m_i)} < \sqrt{(x-m_j)^T\Sigma_j^{-1}(x-m_j)}, \quad \forall j \neq i \quad (4.5)$$

Πλέον ο αρχικός ταξινομητής Bayes μεγίστης a-posteriori πιθανότητας έχει μετατραπεί σε έναν ταξινομητή ελάχιστης απόστασης. Μάλιστα η σχέση 4.5 βασίζεται σε mahalanobis αποστάσεις και είναι μια μορφή του mahalanobis ταξινομητή για κλάσεις με διαφορετικό πίνακα συνδιασποράς.

- Ο πίνακας συνδιασποράς για κάθε κλάση είναι διαγώνιος. Στην περίπτωση αυτή ο ταξινομητής που προκύπτει ταυτίζεται με τον Naive Bayes ταξινομητή που θα περιγράψουμε στην συνέχεια.

- Ο πίνακας συνδιασποράς είναι ο ίδιος για όλες τις κλάσεις και η ταξινόμηση γίνεται με την σχέση

$$\sqrt{(x - m_i)^T \Sigma^{-1} (x - m_i)} < \sqrt{(x - m_j)^T \Sigma^{-1} (x - m_j)}, \quad \forall j \neq i \quad (4.6)$$

Ο mahalanobis ταξινομητής αυτός είναι μια πιο απλοποιημένη έκδοση του προηγούμενου που είχε πιο χαλαρές υποθέσεις αλλά πολλές περισσότερες παραμέτρους (πίνακες συνδιασποράς για κάθε κλάση).

- Περιορίζοντας περαιτέρω τον πίνακα συνδιασποράς έτσι ώστε να είναι διαγώνιος και όλα τα στοιχεία της διαγωνίου να είναι μεταξύ τους ίσα ($\Sigma = \sigma^2 I$) ο αρχικός ταξινομητής εκφυλίζεται στον ταξινομητή ευκλείδειας απόστασης και η σχέση 4.6 γίνεται

$$\|x - m_i\| < \|x - m_j\|, \quad \forall j \neq i \quad (4.7)$$

Αξίζει να σημειωθεί ότι ο ταξινομητής ευκλείδειας απόστασης χρησιμοποιείται πολύ συχνά παρά τις πολλές, συχνά μη ρεαλιστικές, υποθέσεις εξαιτίας της απλότητας του (απαιτεί μόνο τα κέντρα κάθε κλάσης).

4.3.1 Naive Bayes Ταξινομητής

Σύμφωνα με τον Naive Bayes ταξινομητή τα στοιχεία - χαρακτηριστικά ενός d -διάστατου διανύσματος x είναι στατιστικώς ανεξάρτητα και κάθε χαρακτηριστικό ακολουθεί μια μονοδιάστατη γκαουσιανή κατανομή. Συνεπώς η εκτίμηση της υπό συνθήκη σ.π.π. για την εκάστοτε κλάση γίνεται από την σχέση

$$p(x|\omega_i) = \prod_{j=1}^d \frac{1}{\sqrt{2\pi\sigma_{ij}^2}} \exp\left(-\frac{(x_j - m_{ij})^2}{2\sigma_{ij}^2}\right), \quad i = 1, 2, \dots, d \quad (4.8)$$

όπου x_j το j -οστό στοιχείο του x και m_{ij} , σ_{ij} είναι η μέση τιμή και η διασπορά του j χαρακτηριστικού για την i κλάση.

Στην συνέχεια απλά εφαρμόζουμε τον ταξινομητή Bayes και επιλέγουμε για κάποιο πρότυπο x την κλάση με την μεγαλύτερη τιμή της σ.π.π. της σχέσης 4.8, θεωρώντας ίσες a-priori πιθανότητες για κάθε κλάση.

Παρότι η υπόθεση της στατιστικής ανεξαρτησίας είναι μη ρεαλιστική, είναι ιδιαίτερα εξυπηρετική σε χώρους πολλών διαστάσεων, όπου απαιτούνται πολλά δεδομένα εκπαίδευσης για την εκτίμηση μιας πολυδιάστατης σ.π.π., ενώ είναι εύκολο να εξαγάγουμε αξιόπιστες εκτιμήσεις για μονοδιάστατες κατανομές (πολύ λιγότερες παράμετροι) και με σαφώς μικρότερο πλήθος δεδομένων.

4.3.2 Παραμετρική Εκτίμηση - Εκτίμηση Μέγιστης Πιθανοφάνειας

Τόσο στον mahalanobis ταξινομητή όσο και στον Naive Bayes κάνουμε λόγο για μοντελοποίηση σύμφωνα με γκαουσιανές κατανομές, οι οποίες όμως μας είναι άγνωστες, δηλαδή δεν γνωρίζουμε τον μέσο όρο τους και τον πίνακα συνδιασποράς τους που αποτελούν και τις παραμέτρους προς εκτίμηση. Η παραμετρική εκτίμηση των γκαουσιανών κατανομών γίνεται με χρήση της τεχνικής της Μέγιστης Πιθανοφάνειας (Maximum Likelihood). Για N διανύσματα x_i η εκτίμηση για την μέση τιμή είναι

$$m_{ML} = \frac{1}{N} \sum_{i=1}^N x_i$$

και του μητρώου συνδιασποράς είναι

$$\Sigma_{ML,baised} = \frac{1}{N} \sum_{i=1}^N (x_i - m_{ML})(x_i - m_{ML})^T$$

$$\Sigma_{ML,unbiased} = \frac{1}{N-1} \sum_{i=1}^N (x_i - m_{ML})(x_i - m_{ML})^T$$

4.3.3 Ταξινομητής Πλησιέστερων Γειτόνων

Ο ταξινομητής k-πλησιέστερων γειτόνων (k Nearest Neighbors - kNN) υποθέτει ένα σύνολο c κλάσεων $\omega_i, i = 1, 2, \dots, c$ και N d-διάστατων διανυσμάτων αναφοράς - εκπαίδευσης $x_i, i = 1, 2, \dots, N$ για τα οποία γνωρίζουμε την κλάση που ανήκουν και έστω x ένα νέο διάνυσμα προς ταξινόμηση. Ο kNN ταξινομεί τον νέο διάνυσμα ως εξής

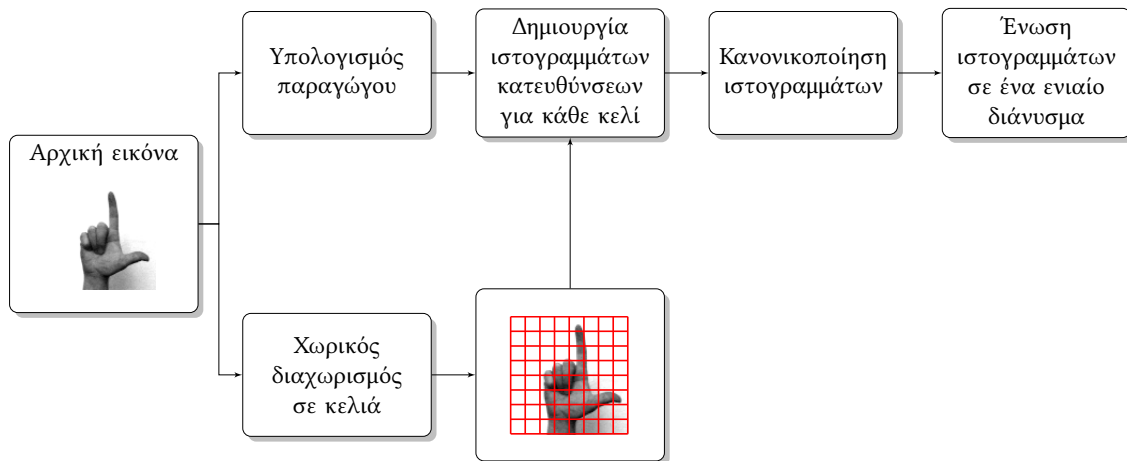
- Αρχικά υπολογίζει τις αποστάσεις όλων των διανυσμάτων αναφοράς από το νέο διάνυσμα. Η απόσταση που χρησιμοποιείται μπορεί να είναι της επιλογής μας με συνηθέστερη επιλογή την ευκλείδεια.
- Βρίσκει τους k κοντινότερους γείτονες του διανύσματός x και προσδιορίζει τις κλάσεις τους
- Το x ανατίθεται στην κλάση στην οποία ανήκει η πλειονότητα των k-πλησιέστερων γειτόνων.

Η τεχνική αυτή είναι ιδιαίτερα δημοφιλής καθώς μπορεί να προσομοιάσει αυθαίρετες κατανομές για την κάθε κλάση, εν τούτοις δεν μπορεί να μοντελοποιήσει τις κατανομές αυτές (που θα ήταν η ιδανική περίπτωση). Επιπλέον το μεγάλο μειονέκτημα του kNN είναι η υπολογιστική πολυπλοκότητα του, καθώς για κάθε νέο διάνυσμα προς ταξινόμηση πρέπει να υπολογιστούν οι αποστάσεις με όλα τα διανύσματα αναφοράς, διαδικασία εξαιρετικά χρονοβόρα (θεωρούμε ότι τα διανύσματα αναφοράς είναι πολλά έτσι ώστε να καλύπτουν επαρκώς τον χώρο).

Κεφάλαιο 5

Προτεινόμενες Παραλλαγές των HOG

Στο κεφάλαιο αυτό γίνεται περιγραφή σχετικά απλών παραλλαγών της βασικής μεθόδου των HOG που συνοψίζεται στο Σχήμα 5.1, όπου παρουσιάζονται τα κύρια βήματα της μεθόδου. Αρχικά θα παρουσιάσουμε ορισμένες προτεινόμενες αλλαγές στο βήμα του χωρικού διαχωρισμού της εικόνας σε κελιά προς την κατεύθυνση αποδοτικότερης ομαδοποίησης της τοπικής πληροφορίας. Συγκεκριμένα στοχεύουμε την βελτιστοποίηση της χωρικής συσχέτισης μεταξύ εικόνων της ίδιας κατηγορίας, με ιδιαίτερη έμφαση στις εικόνες χειρομορφών και στην ανεξαρτησία των τελικών περιγραφών από την εμφάνιση τυπικών παραμορφώσεων, όπως η ύπαρξη ή μη του καρπού. Στην συνέχεια θα παρουσιάσουμε μια αποδοτική προσαρμογή των HOG και των παραλλαγών τους σε δυαδικές εικόνες απλοποιώντας το βήμα του υπολογισμού της παραγώγου.



Σχήμα 5.1: Τα βασικά στάδια της μεθόδου των HOG, σύμφωνα με την περιγραφή στο Κεφάλαιο 3. Τα στάδια αυτά είναι 1)Υπολογισμός της παραγώγου 2) Χωρικός διαχωρισμός σε κελιά 3)Δημιουργία ιστογραμμάτων κατευθύνσεων για κάθε κελί και 4)Κανονικοποίηση ιστογραμμάτων με ένα επιπλέον στάδιο συνένωσης των τοπικών ιστογραμμάτων σε μια ενιαία περιγραφή

5.1 Centralized HOG

Παρά την επιτυχία της βασικής μεθόδου των HOG, είναι αξιοσημείωτο ότι η χωρική κατανομή των κελιών είναι ιδιαίτερα απλοϊκή, κερδίζοντας βέβαια από άποψη υπολογιστικής πολυπλοκότητας. Μια διαφορετική προσέγγιση, που ξεφεύγει όμως από τα πλαίσια αυτής της μελέτης, είναι η χρήση χαρακτηριστικών της εικόνας, όπως εμβασμό, σχήμα άλλα και σημεία ενδιαφέροντος (interest points), ως πληροφορία για αναπροσαρμογή του πλέγματος που ορίζει τα κελία. Με τον τρόπο αυτό προσπαθούμε να επιτύχουμε μια καλύτερη χωρική διαμέριση που να λαμβάνει υπόψιν χαρακτηριστικά της εικόνας, όπως τα προαναφερθέντα, προτείνοντας ουσιαστικά μια ενδιάμεση λύση μεταξύ του απλού πλέγματος και της επιλογής μόνο σημείων ενδιαφέροντος. Η παραπάνω προσέγγιση αυξάνει κατά πολύ την πολυπλοκότητα του προβλήματος και αντιτίθεται σε ένα από τα βασικά χαρακτηριστικά της μεθόδου, την απλότητα. Μια πρώτη εφαρμογή της παραπάνω ιδέας, η οποία διατηρεί την ταχύτητα της αρχικής μεθόδου και ταυτόχρονα προσθέτει μια ιδιαίτερα σημαντική πληροφορία, είναι η χρήση του κέντρου της εικόνας ως σημείο αναφοράς. Αξίζει να σημειωθεί ότι ο βασικός στόχος κατά την ανάπτυξη της μεθόδου ήταν η προσπάθεια μιας χωρικής κατανομής ανεξάρτητης από την παρουσία του καρπού και το ποσοστό εμφάνισης του και κατευθυντήριο ρόλο έπαιξε η παρατήρηση ότι η σημαντική πληροφορία σχετίζεται κατά κύριο λόγο με τα δάχτυλα, που συνήθως βρίσκονται αντιδιαμετρικά του καρπού.

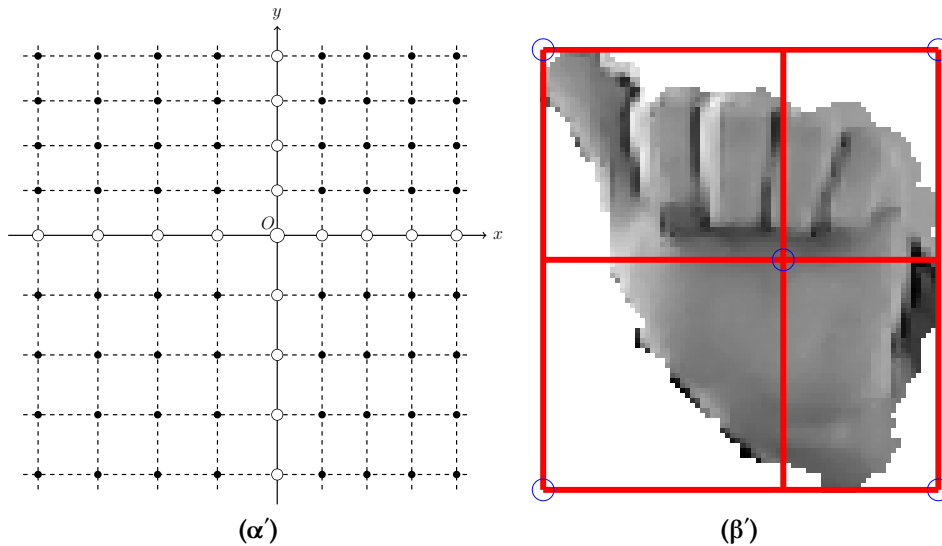
Για διευκόλυνση παριστάνουμε τα κελία με το κέντρο τους, το πλάτος και το ύψος τους, έτσι ώστε να μπορούμε να ορίσουμε πιο εύκολα ένα παραμορφωμένο πλέγμα ως το σύνολο των κέντρων των κελιών, διατηρώντας σταθερό πλάτος και ύψος για όλα τα κελιά. Θεωρώντας ότι έχουμε αλληλοεπικάλυψη κελιών, ένα βασικό στοιχείο για βελτίωση της απόδοσης των HOG ως προς την ανεξαρτησία σε μικρές χωρικές παραμορφώσεις, η επίδραση της μεταβαλλόμενης απόστασης των κέντρων παρά τα σταθερά μεγέθη των κελιών ελαχιστοποιείται (μεταβάλλεται βέβαια σε κάποιο βαθμό το ποσοστό της επικάλυψης, αλλά στην πράξη δεν φαίνεται να επηρεάζει το τελικό αποτέλεσμα). Συνεπώς από εδώ και στο εξής θα αναφερόμαστε στα κελιά και στην διαμέριση της εικόνας ως το πλέγμα (grid) που ορίζεται από τα κέντρα των κελιών.

Ως προς την υλοποίηση, το κέντρο της εικόνας λειτουργεί ως σημείο αναφοράς και για τον λόγο αυτό αντιστοιχίζεται στο κεντρικό σημείο του πλέγματος. Για την κατασκευή του πλέγματος διαμερίζουμε σε ίσα τμήματα τις αποστάσεις των ακραίων στοιχείων από το κέντρο κατά μήκος των αξόνων. Συγκεκριμένα έχουμε 4 διαμερίσεις, μια για κάθε ημιάξονα, χρησιμοποιώντας ίδιο αριθμό στοιχείων για κάθε διαμέριση και στην συνέχεια το τελικό πλέγμα προκύπτει ως ο συνδυασμός των x και y συντεταγμένων των παραπάνω διαμερίσεων, όπως παριστάνεται σχηματικά στο 5.2(α). Το τελικό πλέγμα αποτελείται από τα τέσσερα επιμέρους που δημιουργήθηκαν. Θεωρώντας (x_c, y_c) το κέντρο της εικόνας - χειρομορφής, (x_m, y_m) το μέγεθος της εικόνας και k το πλήθος των διαμερίσεων σε κάθε άξονα (κατά προτίμηση περιττός αριθμός), τα κέντρα των κελιών (x_i, y_j) , $i, j = 1, \dots, k$ υπολογίζονται από τις σχέσεις:

$$x_i = \begin{cases} i[x_c/(\lfloor k/2 \rfloor + 1)], & i = 1, \dots, \lfloor k/2 \rfloor \\ x_c, & i = \lfloor k/2 \rfloor + 1 \\ x_c + (i - \lfloor k/2 \rfloor - 1)[(x_m - x_c)/(\lfloor k/2 \rfloor + 1)], & i = \lfloor k/2 \rfloor + 2, \dots, k \end{cases}$$

$$y_i = \begin{cases} i[y_c/(\lfloor k/2 \rfloor + 1)], & i = 1, \dots, \lfloor k/2 \rfloor \\ y_c, & i = \lfloor k/2 \rfloor + 1 \\ y_c + (i - \lfloor k/2 \rfloor - 1)[(y_m - y_c)/(\lfloor k/2 \rfloor + 1)], & i = \lfloor k/2 \rfloor + 2, \dots, k \end{cases}$$

Ισοδύναμα διαμερίζουμε τον αρχικό χώρο σε τέσσερα ορθογώνια τμήματα που το καθένα ορίζεται από το κέντρο της εικόνας και την αντίστοιχη κορυφή της αρχική εικόνας, όπως στο Σχήμα 5.2(β), και στην συνέχεια για κάθε ορθογώνιο τμήμα της εικόνας δημιουργούμε ένα ισοκατανεμημένο πλέγμα, όπως στην απλή μέθοδο, με το ίδιο πλήθος στοιχείων.

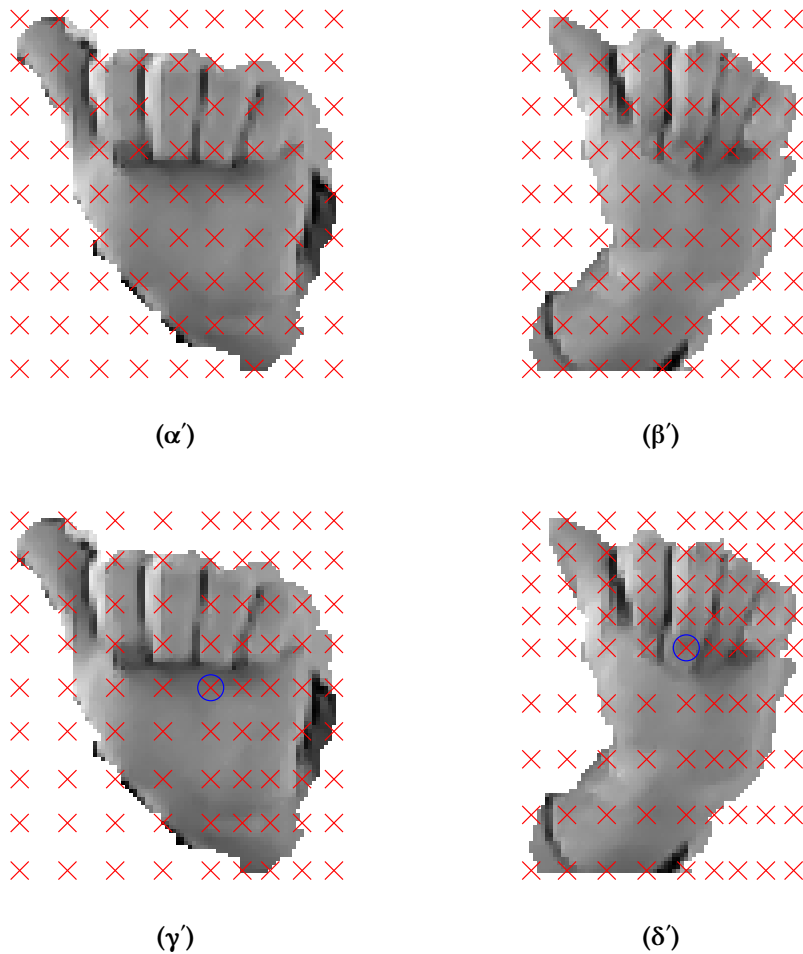


Σχήμα 5.2: (α) Παράδειγμα κατασκευής του πλέγματος. (β) Αρχική χωρική διαμέριση για την κατασκευή του νέου πλέγματος.

Εναλλακτικά μπορούμε να περιγράψουμε την διαδικασία ως μετατροπή του αρχικού ισοκατανεμημένου πλέγματος, που αντιστοιχεί στην βασική μέθοδο, χρησιμοποιώντας το κέντρο της εικόνας ως σημείο έλξης (attractor) για τα υπόλοιπα σημεία του πλέγματος.

Τα αποτελέσματα της προσαρμογής του πλέγματος που προσφέρει η μέθοδος φαίνονται στο Σχήμα 5.3, όπου είναι εμφανής η προοπτική της μεθόδου. Στις εικόνες (α),(β) φαίνονται δυο διαφορετικά στιγμιότυπα της ίδιας χειρομορφής και το απλό πλέγμα που αντιστοιχεί στην βασική μέθοδο HOG. Οι εικόνες αυτές διαφέρουν σημαντικά στο γεγονός ότι η δεύτερη περιέχει ένα μεγάλο τμήμα του καρπού σε αντίθεση με την πρώτη.

Η διαφορά αυτή δεν προβλέπεται από την απλή μέθοδο καθώς υπάρχει προφανής αναντιστοιχία του χώρου που περιγράφουν τα σημεία του πλέγματος. Η παρατηρηθείσα διαφορά είναι δυνατόν να ενσωματωθεί ως έναν βαθμό στην περιγραφή της κλάσης αυτής μόνο με κάποια συνθέτη τεχνική διαχωρισμού των κλάσεων, όπως SVM, και εφόσον υπάρχουν αρκετά δεδομένα για εκπαίδευση. Συνεπώς η προτεινόμενη μέθοδος στοχεύει στην μερική επίλυση της δυσκαμψίας της υπάρχουσας μεθόδου και το δημιουργηθέν πλέγμα για τις εικόνες (α),(β) παρουσιάζεται στις εικόνες (γ),(δ).



Σχήμα 5.3: Σύγκριση της μορφής του πλέγματος. Στις εικόνες (α),(β) εμφανίζεται το πλέγμα της απλής μεθόδου και στις (γ),(δ) το παραμορφωμένο πλέγμα της προτεινόμενης

Παρατηρούμε ότι, παρόλο που και στην περίπτωση αυτή υπάρχει αναντιστοιχία στην πλευρά του καρπού, η αντιστοίχιση στην περιοχή των δακτύλων έχει βελτιωθεί αισθητά που είναι και το ζητούμενο. Με αυτό τον τρόπο μπορούμε να μοντελοποιήσουμε αποδοτικά¹ την εκάστοτε κλάση, αφού η ουσιώδης πληροφορία, δηλαδή τα δάχτυλα, έχει μια πιο σταθερή και ανεξάρτητη χωρική κατανομή.

¹στην αναφερθείσα μοντελοποίηση απαραίτητη είναι και η τεχνική LDA, με την οποία εξάγουμε χαρακτηριστικά που μας ενδιαφέρουν στην αναγνώριση ενώ απορρίπτονται αυτά που δεν έχουν σημαντικό ρόλο

Συμπερασματικά, με το προαναφερθέν επιπρόσθετο στάδιο υπολογισμού ενός προσαρμοζόμενου πλέγματος συμπεριλαμβάνουμε μια εξαιρετικής σημασίας πληροφορία για την χωρική κατανομή της εικόνας. Έτσι ο προκύπτων περιγραφέας *CentHOG* είναι σαφώς πιο ευσταθής σε χωρικές παραμορφώσεις, γεγονός που επιβεβαιώνεται και στο πειραματικό μέρος της μελέτης, ενώ το κόστος του επιπρόσθετου σταδίου, δηλαδή ο υπολογισμός του κέντρου της εικόνας, είναι μικρό, διατηρώντας με αυτό τον τρόπο την ταχύτητα της αρχικής μεθόδου.

5.2 Κυκλική Χωρική Διαμέριση - RHOG

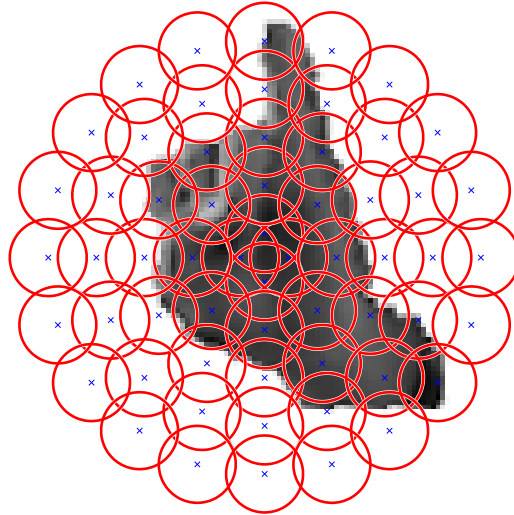
Η κυκλική διαμέριση της εικόνας είναι ένα από τα βασικά συστατικά για την κατασκευή περιγραφών ανεξάρτητων στην περιστροφή της εικόνας, πρόβλημα που θα μελετήσουμε εκτενώς σε επόμενο κεφάλαιο. Πολύ συχνά στην βιβλιογραφία χρησιμοποιούν κυκλικές διαμερίσεις ([25],[18],[17]) για τοπικές περιγραφές (είτε polar είτε logpolar), αλλά δεν έχει μελετηθεί εκτενώς η περίπτωση μιας ενιαίας κυκλικής περιγραφής της εικόνας και όχι θεωρώντας επιμέρους κυκλικές διαμερίσεις (συνήθως σε σημεία ενδιαφέροντος).

Η χωρική διαμέριση της εικόνας είναι ιδιαίτερα σημαντική καθώς ορίζει την διακριτοποίηση της τοπικής πληροφορίας αλλά και την χωρική διάταξη των τοπικών περιγραφών. Στην περίπτωση μας, θέλουμε να έχουμε την ίδια «ποσότητα» πληροφορίας σε κάθε τοπική περιγραφή, δηλαδή ίδιο μέγεθος κελιών και επιπλέον μια κυκλική διάταξη των κελιών αυτών. Οι απαιτήσεις αυτές δεν συμβαδίζουν με μια πιθανή logpolar διαμέριση, που είναι συχνή επιλογή για τέτοιου είδους προβλήματα, αφού η διακριτοποίηση που προσφέρει είναι αρκετά άνιση και όσο απομακρυνόμαστε από το κέντρο η πληροφορία που προσφέρει γίνεται πιο «χοντρική». Για τον λόγο αυτό θεωρούμε κυκλικό πλέγμα ομοκέντρων κύκλων με κέντρο το κέντρο της εικόνας, όπου κάθε κελί έχει το ίδιο μέγεθος και για τον λόγο αυτό κύκλος μεγαλύτερης ακτίνας αντιστοιχεί σε περισσότερα κελιά. Στο Σχήμα 5.4 παρουσιάζεται ένα παράδειγμα της παραπάνω χωρικής διαμέρισης.

Αναλυτικότερα, εαν nc το πλήθος των κύκλων, np_{init} το πλήθος των σημείων που αντιστοιχούν σε τοπικές περιγραφές στον εσωτερικό κύκλο και επιπλέον θεωρήσουμε ότι οι ακτίνες των ομόκεντρων κύκλων δίνονται απο την σχέση $r_i = i r_{init}$, $i = 1, \dots, nc$, όπου r_{init} η ακτίνα του εσωτερικού κύκλου, τότε επειδή αυξάνεται ανάλογα και η περιφέρεια κάθε κύκλου μπορούμε να υποθέσουμε ότι τα σημεία που αντιστοιχούν σε τοπικές περιγραφές στον i -οστό κύκλο είναι $np_i = i np_{init}$ και συνολικά όλα τα σημεία είναι $np = [nc(nc + 1)/2]np_{init}$. Επιπλέον, για τις ανάγκες της γενικής σταθερότητας της περιγραφής θεωρούμε αναγκαία και την επικάλυψη κελιών σε κάποιο βαθμό. Τα κέντρα των κελιών υπολογίζονται σύμφωνα με τις σχέσεις:

$$(x_{ij}, y_{ij}) = (r_i \sin\{j[2\pi/(np_i + 1)]\}, r_i \cos\{j[2\pi/(np_i + 1)]\})$$

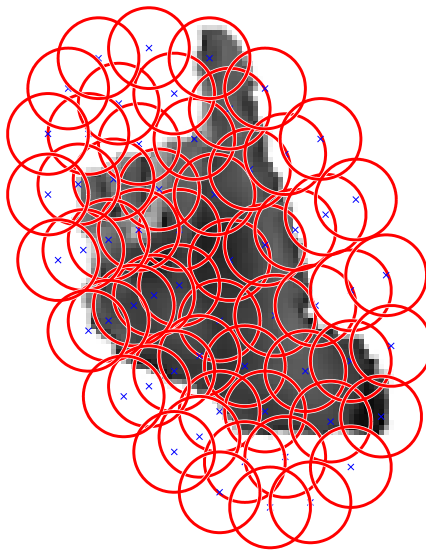
$$i = 1, \dots, nc \quad , \quad j = 1, \dots, np_i$$



Σχήμα 5.4: Κυκλική Διαμέριση

5.3 Ελλειπτική Χωρική Διαμέριση - ERHOG

Η κυκλική διαμέριση είναι πολύ πιθανό να έχει μεγάλο ποσοστό κελιών χωρίς κάποια πληροφορία, καθώς το εύρος του χώρου που εκτείνονται τα κελία ορίζεται από την μέγιστη ακτίνα κατά την οποία υπάρχει πληροφορία, αλλά είναι αρκετά πιθανό προς άλλες κατευθύνσεις η πληροφορία να περιορίζεται από σαφώς μικρότερο εύρος. Στην περίπτωση που γνωρίζουμε το περίγραμμα του αντικείμενου προς περιγραφή (όπως συμβαίνει για τις χειρομορφές που χρησιμοποιούνται στην μελέτη αυτή) μπορούμε να διαμερίσουμε τον χώρο με πιο «συμπαγή» τρόπο από την προηγούμενη κυκλική διαμέριση, υιοθετώντας μια διαμέριση σε σχήμα έλλειψης, όπως φαίνεται και στο παράδειγμα του Σχήματος 5.5.



Σχήμα 5.5: Ελλειψοειδής Διαμέριση

Το τελικό πλέγμα θα αποτελείται, σε αντιστοιχία με την περίπτωση της κυκλικής διαμέρισης, από ομόκεντρες ελλείψεις. Τις παραμέτρους της (εξωτερικής) έλλειψης, τις υπολογίζουμε χρησιμοποιώντας τις κεντρικές ροπές 2^{ης} τάξης με τις οποίες ουσιαστικά κατασκευάζουμε τον πίνακα συνδιασποράς της δυαδικής εικόνας

$$\text{cov}[I(x, y)] = \begin{bmatrix} \mu_{20} & \mu_{11} \\ \mu_{11} & \mu_{02} \end{bmatrix}$$

Τα ιδιοδιανύσματα του παραπάνω πίνακα αντιστοιχούν στους άξονες της έλλειψης και η περιστροφή της προκύπτει ως η γωνία περιστροφής του κύριου άξονα, δηλαδή του ιδιοδιανύσματος που αντιστοιχεί στην μεγαλύτερη ιδιοτιμή.²

Στο σημείο αυτό πρέπει να τονίσουμε ότι η τετραγωνική και η κυκλική διαμέριση μπορούν να χρησιμοποιηθούν γενικά ως παράθυρα αναζήτησης αντικειμένων, ενώ η ελλειψοειδής απαιτεί πρότερη γνώση του σχήματος του αντικειμένου και της διάταξης του στον χώρο, γεγονός που την κάνει δύσχρηστη σε εφαρμογές όπου δεν είναι δυνατό να εξάγουμε άμεσα μια προσέγγιση του σχήματος του αντικειμένου.

5.4 Binary HOG (BHOG) - Εφαρμογή των HOG σε δυαδικές εικόνες

Έχοντας παρουσιάσει τα θετικά της HOG μεθόδου, είναι λογικό να δοκιμάσουμε την επιτυχία της μεθόδου σε δυαδικές εικόνες - μάσκες της αρχικής εικόνας, όπου αποτελούνται από μηδέν στα pixels του background και ένα στα pixel που αντιστοιχούν σε χέρι και ουσιαστικά περιέχουν μόνο πληροφορία για το περίγραμμα. Όπως έχει ήδη αναφερθεί, πολλές τεχνικές για περιγραφή του σχήματος-περιγράμματος (π.χ. Fourier Descriptors) είναι ιδιαίτερα ευαίσθητες σε παρουσία ή απουσία τμημάτων της εικόνας, όπως είναι το σύνηθες πρόβλημα της εμφάνισης του καρπού για την βάση που χρησιμοποιήθηκε. Από την άλλη τα HOG φαίνεται να επηρεάζονται πολύ λιγότερο από τέτοια προβλήματα, καθώς ορίζουν μια πιο ανεξάρτητη χωρική διαμέριση της εικόνας που δεν βασίζεται στο ίδιο το περίγραμμα (στην απλή περίπτωση χωρισμός σε ένα απλό τετραγωνικό πλέγμα).

Αν εφαρμόσουμε την βασική μέθοδο HOG απευθείας σε μια δυαδική εικόνα είναι προφανές ότι υπάρχουν περιττά βήματα στους υπολογισμούς, άρα και επιπρόσθετο υπολογιστικό κόστος χωρίς να είναι ανάγκη. Η βασική διαφορά είναι ότι πλέον οι μερικές παράγωγοι, όρος που χρησιμοποιείται πλέον καταχρηστικά, g_x και g_y (εφαρμόζοντας τα φίλτρα $[-1 \ 0 \ 1]$ και $[-1 \ 0 \ 1]^T$) δεν μπορούν να πάρουν οποιαδήποτε τιμή, αλλά αντ'αυτού περιορίζονται στο σύνολο $\{-1 \ 0 \ 1\}$. Ουσιαστικά πλέον τα g_x και g_y υποδεικνύουν την ύπαρξη αλλαγής στο περίγραμμα για τις διευθύνσεις του άξονα

²Αξίζει να σημειωθεί ότι η συγκεκριμένη διαδικασία είναι ισοδύναμη με την εκτίμηση της δυσδιάστατης γκαουσιανής κατανομής που ικανοποιεί καλύτερα τα σημεία της δυαδικής εικόνας (maximum likelihood estimation).

x και y αντίστοιχα. Συνοψίζοντας οι πιθανές τιμές κατεύθυνσης και μέτρου (angle & magnitude), έτσι όπως ορίζονται στην βασική μέθοδο, καταγράφονται στους επομένους 2 πίνακες που περιγράφουν και τους 9 πιθανούς συνδυασμούς των τιμών των g_x και g_y .

Angle of Binary Gradient				Magnitude of Binary Gradient			
$g_y \backslash g_x$	-1	0	1	$g_y \backslash g_x$	-1	0	1
-1	45°	90°	135°	-1	1.42	1	1.42
0	0°	0°	0°	0	1	0	1
1	135°	-90°	45°	1	1.42	1	1.42

Πίνακας 5.1: Πιθανές τιμές Διεύθυνσης και Μέτρου

Όπως προκύπτει και από τον πίνακα 5.1 τα bins πλέον περιορίζονται σε τέσσερα που πλέον δεν ορίζονται ως παράμετρος από τον χρήστη και αντιστοιχούν στις γωνίες 0°, 45°, 90° και 135°, θεωρώντας μη προσημασμένη παράγωγο σε αντιστοιχία με την βασική μέθοδο. Οι πιθανές τιμές για το μέτρο είναι 0, 1, $\sqrt{2} = 1,42$ και αντιστοιχούν σε ανυπαρξία κατεύθυνσης (μηδενική «παράγωγος», δηλαδή $g_x = 0$ και $g_y = 0$), σε ύπαρξη κατεύθυνσης μόνο ως προς έναν άξονα ($g_x \neq 0$ ή $g_y \neq 0$) και ύπαρξη κατεύθυνσης και στους δυο άξονες ($g_x \neq 0$ και $g_y \neq 0$)³. Στο σημείο αυτό αξίζει να σημειωθεί ότι στην βασική μέθοδο, όπως έχουμε ήδη αναλύσει, τα bins ορίζουν μια διακριτοποίηση στις διευθύνσεις και κάθε bin αντιστοιχεί σε ένα εύρος κατευθύνσεων, ενώ στην περίπτωση μας δεν υπάρχει στην πραγματικότητα κάποιο εύρος διευθύνσεων αλλά μόνο οι προαναφερθείσες τέσσερις διακριτές τιμές (0°, 45°, 90° και 135°). Για τον λόγο αυτό για την κατασκευή του εκάστοτε τοπικού ιστογράμματος αντιστοιχούμε κάθε μια από τις διακριτές τιμές σε ένα bin.

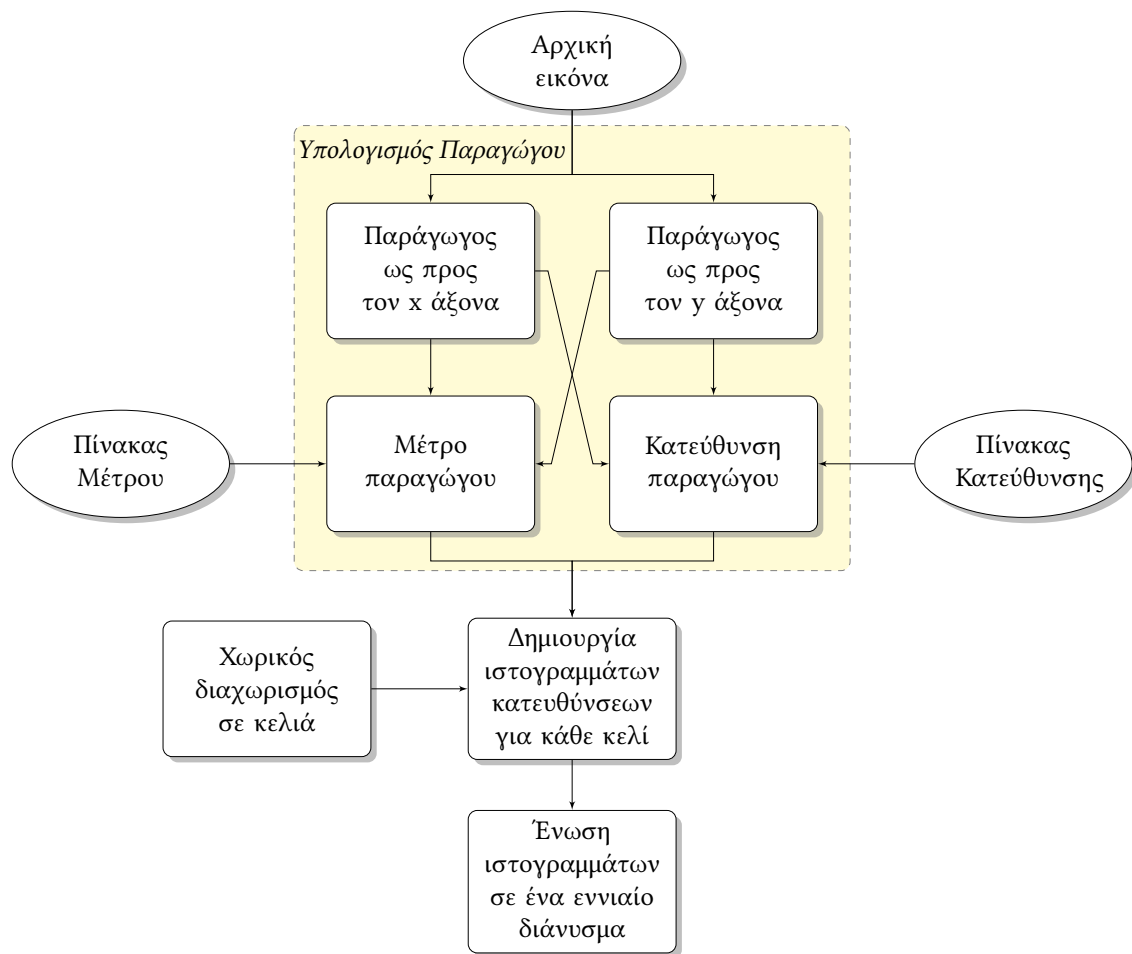
Με την καταγραφή των πιθανών συνδυασμών των μερικών παραγώγων επιτυγχάνουμε την συστηματοποίηση της κατασκευής των ιστογραμμάτων, αφού πλέον υπάρχει η δυνατότητα να χρησιμοποιήσουμε απευθείας τις τιμές των πινάκων κατεύθυνσης και μέτρου για τον υπολογισμό των ιστογραμμάτων. Πλέον η διαδικασία υπολογισμού των διευθύνσεων και του μέτρου για κάθε pixel μετατρέπεται σε απλή χρήση των τιμών των προαναφερθέντων πινάκων, αποφεύγοντας τις υπολογιστικά «κοπιαστικές» συναρτήσεις για κατεύθυνση (αντιστροφή εφαπτομένης ως η γωνία μεταξύ g_x και g_y) και πλάτος (ρίζα αθροίσματος τετραγώνων) της βασικής μεθόδου.

Τέλος, στην βασική μέθοδο κάθε pixel συνεισφέρει στο ιστόγραμμα ανάλογα με την τιμή του μέτρου της παραγώγου στο σημείο αυτό. Βέβαια η τιμή του μέτρου της παραγώγου εξαρτάται σε μεγάλο βαθμό από τις συνθήκες φωτισμού και για τον λόγο αυτό ακολουθεί κανονικοποίηση (normalization) των τοπικών ιστογραμμάτων. Στην παραλα-

³Ο πίνακας του μέτρου ουσιαστικά χρησιμοποιείται για voting στην κατασκευή ιστογραμμάτων και λόγω της φύσης τους προβλήματος μπορούμε να αφαιρέσουμε εντελώς την καταχρηστική έννοια του μέτρου και να προσθέτουμε 1 εφόσον υπάρχει παράγωγος ($g_x \neq 0$ ή $g_y \neq 0$), απλοποιώντας περεταίρω την διαδικασία.

λαγή για δυαδικές εικόνες που παρουσιάσαμε δεν υπάρχει τέτοια ανάγκη για κανονικοποίηση, καθώς η τελική τιμή σε κάθε bin σχετίζεται με το πλήθος των στοιχείων της περιφέρειας στο εκάστοτε κελί και όχι με την ίδια την παράγωγο. Η παραπάνω παρατήρηση συνεισφέρει στην περεταίρω υπολογιστική απλότητα της μεθόδου συγκριτικά με την βασική μέθοδο αφαιρώντας το στάδιο της κανονικοποίησης.

Συνολικά η μέθοδος συνοψίζεται στο Σχήμα 5.6, όπου παρουσιάζονται σχηματικά τα στάδια της παραλλαγής αυτής και η χρήση των πινάκων Μέτρου και Κατεύθυνσης, ενώ αξιοσημείωτη είναι και η απουσία του βήματος της κανονικοποίησης.

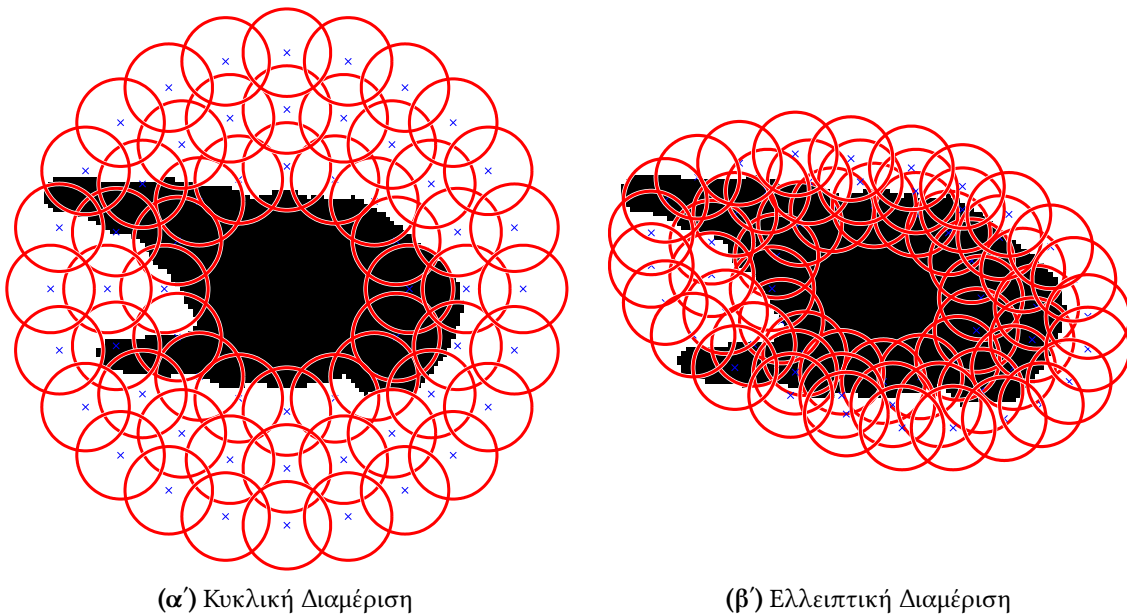


Σχήμα 5.6: Σύνοψη της μεθόδου BHOG. Στο σχήμα αυτό αναλύουμε το βήμα του υπολογισμού της παραγώγου, έτσι ώστε να γίνει εμφανές με ακρίβεια σε ποιο στάδιο γίνεται χρήση των πινάκων Μέτρου και Κατεύθυνσης.

5.5 Προσαρμογή των Παραλλαγών των HOG για Δυαδικές Εικόνες

Συμφώνα με την περιγραφείσα τεχνική των BHOG είναι δυνατόν να προσαρμόσουμε άμεσα τις χωρικές παραλλαγές που προτείναμε στην αρχή του κεφαλαίου για την εφαρμογή τους σε δυαδικές εικόνες. Συγκεκριμένα απλά μεταβάλλουμε το βήμα του υπολογισμού της παραγώγου, όπως περιγράψαμε στην προηγούμενη ενότητα και χωρίς κάποια επιπλέον προσαρμογή καθώς το βήμα της χωρικής διαμέρισης είναι ανεξάρτητο από αυτό του υπολογισμού της παραγώγου. Συνεπώς δημιουργούνται κατ'αντιστοιχεία οι μέθοδοι *CentBHOG*, *RBHOG* και *ERBHOG*.

Βέβαια, παρατηρείται ένας πλεονασμός ανούσιας πληροφορίας αν εφαρμόσουμε αυτούσια την τεχνική για δυαδικές εικόνες σε κυκλική ή ελλειπτική διαμέριση, ο οποίος έγκειται στο γεγονός ότι στην περίπτωση της δυαδικής μάσκας κάποιας χειρομορφής δεν υπάρχει καμία πληροφορία εντός του μεγίστου εγγεγραμμένου κύκλου (που έχει υπολογιστεί κατά τον υπολογισμό του κέντρου της χειρομορφής, βλ. Παράρτημα Α'). Συνεπώς για την περίπτωση των RBHOG, μια λογική κυκλική διαμέριση για δυαδικές εικόνες θα αποτελείται από ομοκέντρους κύκλους, των οποίων η ακτίνα θα κυμαίνεται από την ακτίνα του μεγίστου εγγεγραμμένου κύκλου (r_{min}) έως και την μέγιστη απόσταση του κέντρου από το περίγραμμα της χειρομορφής (r_{max}). Ένα παράδειγμα της προαναφερθείσας διαμέρισης παρουσιάζεται στο Σχήμα 5.7(α). Η ίδια λογική εφαρμόζεται και στην περίπτωση ελλειπτικής διαμέρισης, με την εσωτερική έλλειψη να προσεγγίζει τον μέγιστο εγγεγραμμένο κύκλο του περιγράμματος. Ένα παράδειγμα της ελλειπτικής διαμέρισης για δυαδικές μάσκες χειρομορφών απεικονίζεται στο Σχήμα 5.7(β).



Σχήμα 5.7: Κυκλική και Ελλειπτική Διαμέριση για Δυαδικές Εικόνες Χειρομορφών

Κεφάλαιο 6

Δημιουργία Περιστροφικά Ανεξάρτητων Περιγραφών Βασισμένων σε HOG

Στο κεφάλαιο αυτό προτείνουμε δυο βασικές τεχνικές για ανεξαρτησία στην περιστροφή. Η πρώτη προσέγγιση βασίζεται στην εύρεση μιας κατεύθυνσης της χειρομορφής στην εικόνα χρησιμοποιώντας μεθόδους περιγράμματος βασισμένες σε shape context και στην συνέχεια εφαρμογή της απλής μεθόδου των HOG. Στην πορεία της μελέτης της προτεινόμενης προσέγγισης δημιουργούμε μια ανεξάρτητη τεχνική περιγράμματος που αποτελεί παραλλαγή των shape context. Στην επόμενη ενότητα παρουσιάζουμε μια υποσχόμενη τεχνική που βασίζεται σε κυκλικά HOG (κυκλική χωρική διαμέριση) αλλά και παραλλαγές της ιδέας αυτής, κρατώντας ως κεντρικό άξονα την κυκλική διαμέριση και την χρήση της τεχνικής Fourier Transform Magnitude για περιστροφική ανεξαρτησία.

6.1 Μια Πρώτη Προσέγγιση για Ανεξαρτησία στην Περιστροφή

Μια πρώτη προσέγγιση για την επίτευξη περιστροφικής ανεξαρτησίας βασίζεται στην ιδέα ότι στις χειρομορφές η πληροφορία για την περιστροφή της εικόνας βρίσκεται κατά κύριο λόγο στο σχήμα, καθώς η πληροφορία αυτή σε μεγάλο βαθμό καθορίζεται από την θέση του καρπού και των δαχτύλων. Σύμφωνα με την παραπάνω παρατήρηση μπορούμε να κατασκευάσουμε μια μέθοδο τριών βημάτων:

- 1 Σαν πρώτο βήμα βρίσκουμε μια γενική κατεύθυνση της εικόνας χρησιμοποιώντας μόνο το σχήμα σε σχέση με μια εικόνα αναφοράς. Για τον σκοπό αυτό γίνεται χρήση της μεθόδου Shape Context, η οποία έχει αποδειχτεί στην πράξη ιδιαίτερα

αποτελεσματική. Η περιστροφή της εικόνας θα εξαχθεί από την αντιστοίχιση των σημείων του περιγράμματος της εικόνας με αυτά της εικόνας αναφοράς.

- 2 Στην συνέχεια εφαρμόζουμε τον προκύπτων μετασχηματισμό από την αντιστοίχιση των σημείων του προηγούμενου βήματος. Στο στάδιο αυτό οι δυο εικόνες θα έχουν όσο το δυνατόν ομοιότερο σχήμα, απαλείφοντας εξαρτήσεις από μετατόπιση, αλλαγή κλίμακας, αλλά και περιστροφή.
- 3 Τέλος, στην μορφή που βρίσκεται πλέον η αρχική εικόνα είναι δυνατόν να αξιοποιηθεί η ισχύς της μεθόδου HOG, καθώς έχουμε προνοήσει να εξαλείψουμε πιθανές περιστροφές. Το βήμα αυτό αναλαμβάνει να αξιοποιήσει και την πληροφορία εσωτερικά του περιγράμματος, έτσι ώστε να αυξηθεί η αποτελεσματικότητα της μεθόδου.

Βέβαια η περιγραφείσα μέθοδος μπορεί να εφαρμοστεί μόνο σε αντικείμενα που η γενική τους κατεύθυνση μπορεί ως ένα βαθμό να οριστεί από το Σχήμα τους. Παραδείγματος χάριν εάν έχουμε διαφορετικά αντικείμενα ως προς την εμφάνισή τους (εσωτερική πληροφορία) αλλά όλα έχουν ως περίγραμμα έναν κύκλο, τότε η μέθοδος αυτή δεν είναι ικανή να εντοπίσει ίδιες εικόνες με διαφορετική γωνιά περιστροφής.

6.1.1 Παραλλαγή της Μεθόδου Shape Context

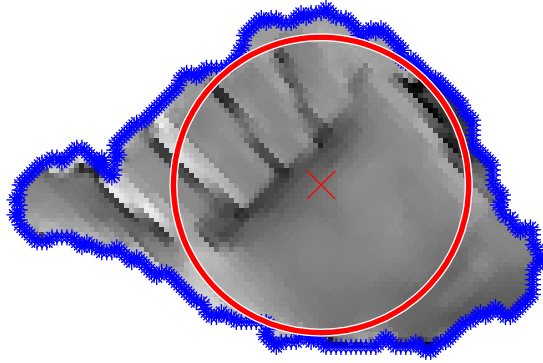
Σε προηγούμενο κεφάλαιο έγινε εκτενής ανάλυση της μεθόδου Shape Context και των προτερημάτων της. Δυο βασικά προβλήματα της μεθόδου είναι

- 1 η ταχύτητα της μεθόδου, καθώς απαιτείται έλεγχος μεταξύ όλων των σημείων (και για την hungarian απαιτείται $O(n^3)$, όπου n το πλήθος των στοιχείων που περιγράφουν κάθε εικόνα)
- 2 ο περιορισμός της χρήσης της μεθόδου για σύγκριση μεταξύ εικόνων και όχι για εξαγωγή ενός διανύσματος χαρακτηριστικών

Οι παρακάτω τροποποιήσεις έγιναν με άξονα την επίλυση των προβλημάτων αυτών στον καλύτερο δυνατό βαθμό και με βασική προτεραιότητα την διατήρηση της ανεξαρτησίας στην περιστροφή. Ξεκινάμε από την παρατήρηση ότι, στην βασική μέθοδο, για κάθε σημείο του περιγράμματος δημιουργείται ένα logpolar ιστόγραμμα της κατανομής των υπόλοιπων σημείων. Συνεπώς γεννάται το εξής ερώτημα : Είναι όλη αυτή η πληροφορία αναγκαία ή είναι δυνατόν να προσεγγιστεί όχι μόνο δειγματοληπτώντας τα σημεία του περιγράμματος που θα χρησιμοποιηθούν ως σημεία αναφοράς, αλλά θέτοντας κάποια άλλα σημεία του χώρου, με στόχο την διευκόλυνση μας, ως σημεία αναφοράς ;

6.1. Μια Πρώτη Προσέγγιση για Ανεξαρτησία στην Περιστροφή

Προκύπτει πειραματικά ότι λόγω του πλήθους της πληροφορίας είναι δυνατόν να χρησιμοποιήσουμε σημεία αναφοράς τέτοια ώστε να μας βολεύουν αρκεί να μην απέχουν πολύ από το περίγραμμα. Τα νέα σημεία αναφοράς γίνονται σημεία ενός κύκλου συγκεκριμένης ακτίνας και έτσι λοιπόν για κάθε τέτοιο σημείο δημιουργείται ένα `logpolar` ιστόγραμμα με την κατανομή των σημείων του περιγράμματος. Το μέγεθος της ακτίνας προτείνεται να είναι κοντά σε αυτό της ακτίνας του μεγίστου εγγεγραμμένου κύκλου, προσδίδοντας ανεξαρτησία στην αλλαγή κλίμακας και μια αρκετά καλή επιλογή για την κατασκευή των ιστογραμμάτων. Η επιλογή του κύκλου έγινε καθαρά για λογούς υπολογιστικής απλότητας και για αποφυγή έντονων και άτοπων παραμορφώσεων που είναι δυνατόν να δημιουργήσει η βασική μέθοδος σε περίπτωση άστοχης συσχέτισης κάποιων σημείων.



Σχήμα 6.1: Παράδειγμα κύκλου αναφοράς (με κόκκινο) για μια εικόνα (το περίγραμμά της σημειώνεται με μπλε) συμφώνα με την παραλλαγή των *shape context*

Είναι πλέον προφανές ότι για σταθερό πλήθος σημείων αναφοράς έχουμε μια περιγραφή πολύ πιο εύκολη στην διαχείριση. Έστω ότι θέλουμε να συγκρίνουμε 2 εικόνες έχοντας K σημεία αναφοράς. Έχουμε λοιπόν για την καθεμία από K ιστογράμματα και η σύγκριση γίνεται μεταξύ των ιστογραμμάτων της εικόνας "αναφοράς" και την κυκλική μετατόπιση των ιστογραμμάτων της άλλης. Δηλαδή αν θεωρήσουμε ότι h_1 και h_2 είναι οι τελικές περιγραφές κάθε εικόνας, που προκύπτουν από την ένωση των επιμέρους K ιστογραμμάτων με πλήθος στοιχείων B το καθένα, τότε η βέλτιστη συσχέτιση τους ως προς την περιστροφή βρίσκεται ως

$$\arg \min_{k=1:K} E[k]$$

όπου

$$E[k] = \sum_{n=1}^{KB} (h_1[((n + kB))_{KB}] - h_2[n])^2$$

Η παραπάνω περιγραφή βασίζεται στην ευκλείδεια απόσταση της κυκλικής μετατόπισης των τοπικών περιγραφών, ενώ εναλλακτικά είναι δυνατό να χρησιμοποιήσουμε *chi-squared* αποστάσεις. Αναλυτικότερα περιγράφουμε το πρόβλημα της εκτίμησης της κυκλικής μετατόπισης στο Παράρτημα Γ' και αναδεικνύουμε την χρησιμότητα της προσέγγισης με *cross-correlation* εφόσον χρησιμοποιήσουμε ευκλείδεια απόσταση.

Ως προς την υλοποίηση λοιπόν έχουμε δυο ισοδύναμες εναλλακτικές που βασίζονται στην μέθοδο της ετεροσυσχέτισης (βλ. Παράρτημα Γ') με την μόνη διαφορά ότι πλέον η κυκλική μετατόπιση είναι ανά τοπικές περιγραφές-ιστογράμματα, δηλαδή ένα σύνολο B στοιχείων, σε αντίθεση με την απλή περίπτωση που είναι ανά στοιχείο. Αρχικά μετασχηματίζουμε τα h_1, h_2 στους πίνακες H_1, H_2 (μεγέθους $K \times B$) όπου κάθε γραμμή αντιστοιχεί σε ένα από τα K ιστογράμματα αναφοράς και στην συνέχεια εφαρμόζουμε απευθείας τον ορισμό του cross-correlation είτε χρησιμοποιούμε την εναλλακτική προσέγγιση μέσα από το πεδίο των συχνοτήτων με την χρήση του μετασχηματισμού Fourier:

- 1 Για την περίπτωση του απευθείας ορισμού ισχύουν οι ακόλουθες σχέσεις

$$L[k] = \sum_{n=1}^{KB} (h_1[((n+kB))_{KB}]h_2[n])$$

ή αλλιώς για τους πίνακες H_1, H_2

$$L[k] = \sum_{n=1}^K \sum_{b=1}^B (H_1[((n+k))_K, b]H_2[n, b])$$

Συμφώνα με την τελευταία σχέση η ετεροσυσχέτιση $L[k]$ προκύπτει ως το άθροισμα των στοιχείων της k -οστής διαγωνίου, με την αρίθμηση του σχήματος 6.2 (θεωρούμε ότι τα όρια του πίνακα «αναδιπλώνονται»), του πίνακα $D = H_1 H_2^T$, μεγέθους $K \times K$. Κάθε στοιχείο $D(i, j)$ αντιστοιχεί στην συσχέτιση (ως το γινόμενο στοιχείου προς στοιχείο) του i -οστού τοπικού ιστογράμματος με το j -οστό.

- 2 Η δεύτερη προσέγγιση βασίζεται στο θεώρημα μετατόπισης του Μετασχηματισμού Fourier και ουσιαστικά πρόκειται για εφαρμογή της μεθόδου phase correlation για την εκτίμηση της σχετικής μετατόπισης δυο εικόνων. Την μέθοδο αυτό την χρησιμοποιούμε στους δισδιάστατους πίνακες H_1 και H_2 θεωρώντας όμως επιτρεπτή μόνο την μετατόπιση κατά τον y άξονα.

Συνολικά για την σύγκριση δυο εικόνων και την συσχέτιση των σημείων τους επιτυγχάνεται ανεξαρτησία :

- 1 στην μετατόπιση που ισοδυναμεί με την μετατόπιση των κέντρων.
- 2 στην αλλαγή κλίμακας που σχετίζεται άμεσα με την ακτίνα του μεγίστου εγγεγραμμένου κύκλου και συνεπώς και των σημείων αναφοράς που επιλέγονται.
- 3 στην περιστροφή της εικόνας, η οποία αντιστοιχεί στην περιστροφή στην σειρά των ιστογραμμάτων περιγραφής, και επιτυγχάνεται όπως περιγράψαμε στην προηγούμενη παράγραφο.

1	2	..	K-1	K
K	1	2	..	K-1
K-1	K	1	2	..
..	K-1	K	1	2
2	..	K-1	K	1

Σχήμα 6.2: Πίνακας τοπικών συσχετίσεων

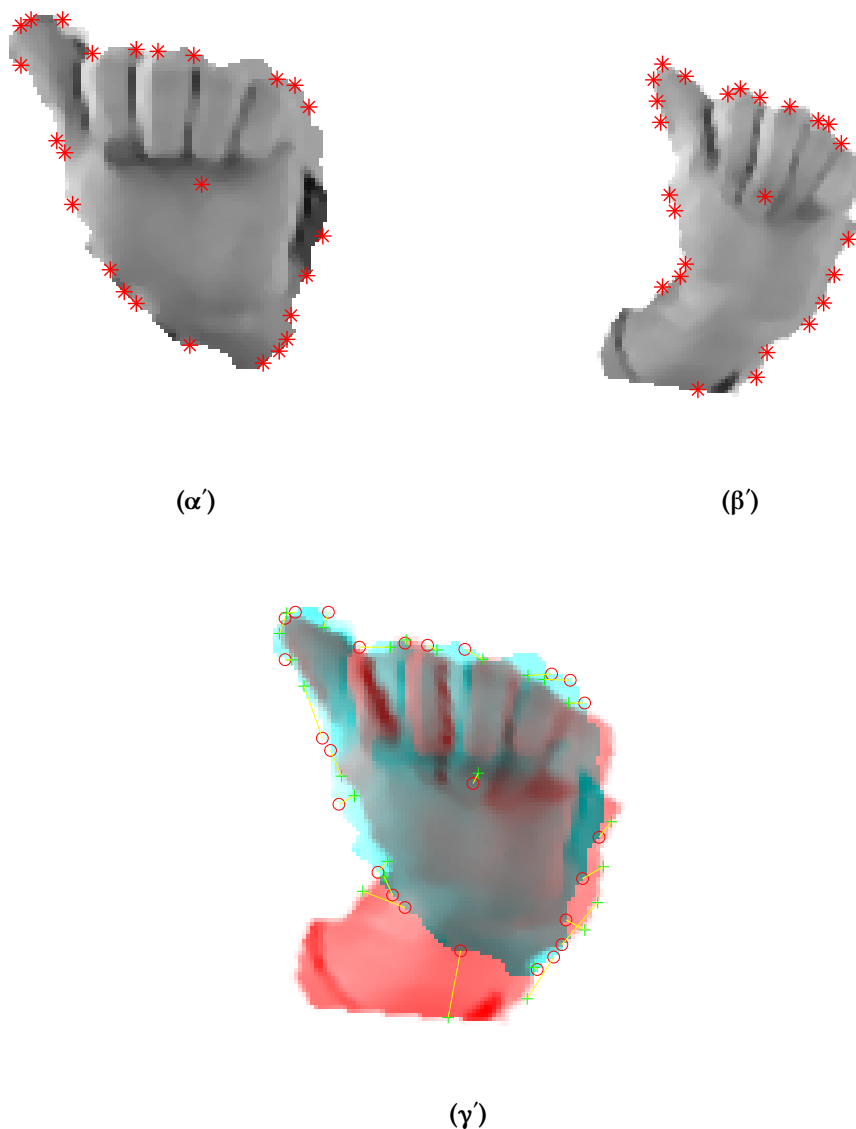
Βέβαια η παραλλαγή αυτή εκφυλίζει τον αφινικό μετασχηματισμό που προκύπτει από την συσχέτιση των σημείων στην βασική μέθοδο σε ομογενή μετασχηματισμό, καθώς η συσχέτιση πλέον διατηρεί μονό τις ιδιότητες της μετατόπισης (translation), της αλλαγής κλίμακας (scale) και της περιστροφής (rotation). Με αυτό τον τρόπο η προτεινόμενη μέθοδος χάνει ως ένα βαθμό την ευχέρεια της βασικής μεθόδου να προσαρμοστεί σε παραμορφώσεις με αντάλλαγμα την ταχύτητα και μια γενική ευστάθεια ως προς την ποιότητα της συσχέτισης των σημείων. Αξίζει να σημειωθεί ότι η μέθοδος HOG είναι ευσταθής ως προς μικρές παραμορφώσεις, οπότε στην συνολική μέθοδο, όπου εφαρμόζεται HOG μετά την περιστροφή με βάση το σχήμα, αντισταθμίζεται ο εκφυλισμός του αφινικού μετασχηματισμού.

6.1.1.1 Δημιουργία Ανεξάρτητου Διανύσματος Χαρακτηριστικών

Είναι δυνατόν να χρησιμοποιήσουμε την παραπάνω παραλλαγή των shape context ως ένα ενιαίο διάνυσμα χαρακτηριστικών αξιοποιώντας την ιδιότητα του μέτρου του μετασχηματισμού Fourier, δηλαδή την μέθοδο Fourier Transform Magnitude (Παράρτημα Γ'). Συγκεκριμένα δεδομένου πίνακα H , με κάθε γραμμή να αποτελεί ένα από τα K τοπικά ιστογράμματα μήκους B , θεωρούμε ότι κάθε στήλη του αντιστοιχεί σε ένα χαρακτηριστικό και επιζητούμε ανεξαρτησία στην κυκλική μετατόπιση στα επιμέρους χαρακτηριστικά. Συνεπώς υπολογίζουμε το μέτρο του FFT για κάθε στήλη του H και η τελική περιγραφή προκύπτει ως η ένωση των επιμέρους αποτελεσμάτων.

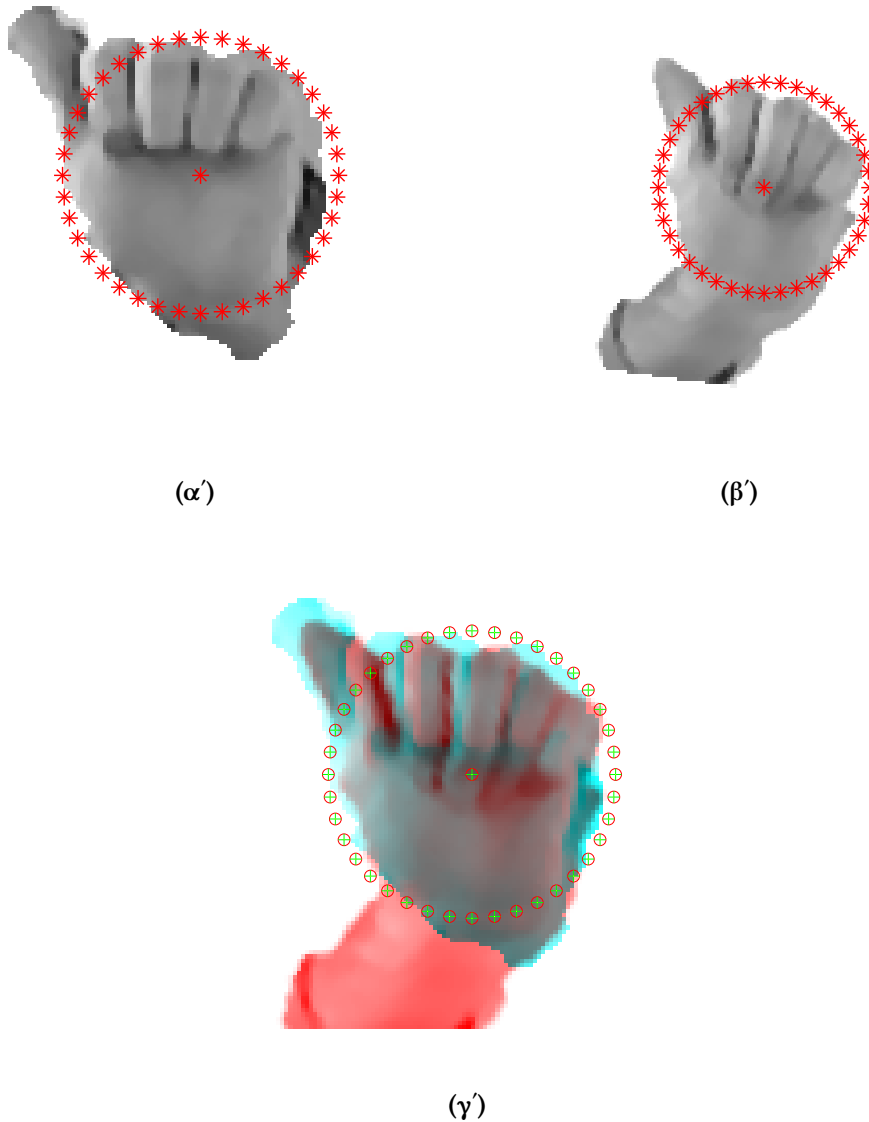
6.1.2 Χρήση Shape Context μαζί με HOG

Έχοντας υπολογίσει έναν μετασχηματισμό μεταξύ της εικόνας προς αναγνώριση και της εικόνας αναφοράς μετασχηματίζουμε την αρχική εικόνα ως προς την εικόνα



Σχήμα 6.3: Παράδειγμα αφινικού ταιριάσματος. (α),(β) οι αρχικές εικόνες μαζί με τα αποδεκτά σημεία του περιγράμματος (γ) η υπέρθεση της μετασχηματισμένης εικόνας με την εικόνα αναφοράς και η αντιστοίχιση των σημείων τους

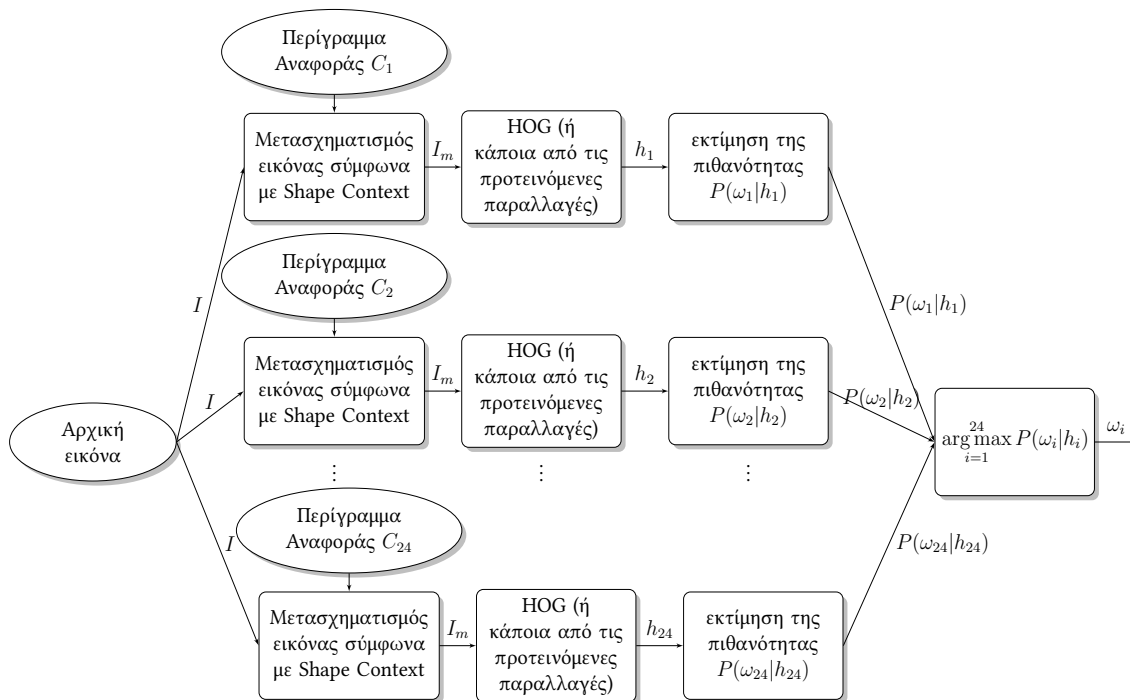
αναφοράς. Για να αυξήσουμε την επιτυχία της μεθόδου μπορούμε να θεωρήσουμε επιτρεπτούς πάνω από έναν μετασχηματισμό με βάση το scoring του ταιριάσματος, τακτική όμως που ανεβάζει και άλλο το κόστος. Τέλος εφαρμόζουμε HOG (ή κάποια παραλλαγή όπως τα CentHOG) στην μετασχηματισμένη εικόνα και χρησιμοποιούμε κάποια από τις γνωστές τακτικές αναγνώρισης, θεωρώντας ότι όλο το σύνολο εκπαίδευσης είναι προσδιορισμένο ως προς τις εικόνες αναφοράς κάθε κλάσης. Ως εικόνες αναφοράς επιλεγούμε το medoid της κάθε κλάσης, που ορίζεται ως η εικόνα της κλάσης της οποίας η περιγραφή απέχει αθροιστικά λιγότερο από όλες τις υπόλοιπες, δηλαδή αποτελεί το κεντρικότερο υπαρκτό σημείο της κλάσης (σε αντίθεση με το κέντρο της κλάσης που είναι ένα κατασκευαζόμενο σημείο).



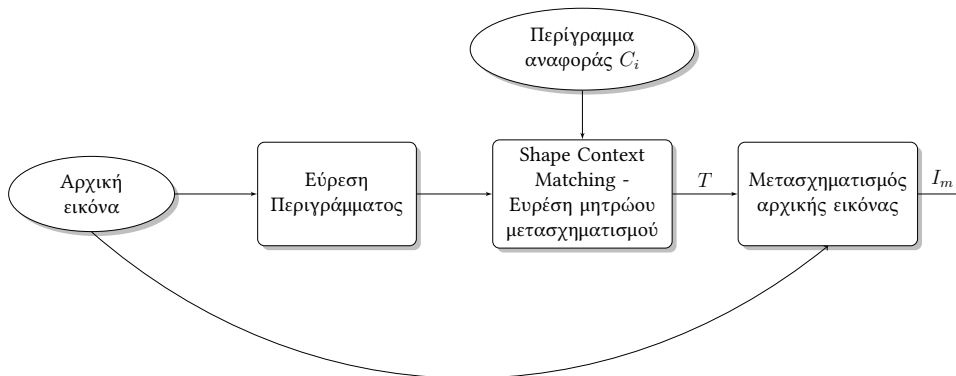
Σχήμα 6.4: Παράδειγμα κυκλικού ταιριάσματος. (α),(β) οι αρχικές εικόνες μαζί με τα σημεία αναφοράς (γ) η υπέρθεση της μετασχηματισμένης εικόνας με την εικόνα αναφοράς και η αντιστοίχιση των σημείων τους

Το σύστημα αναγνώρισης που προκύπτει σύμφωνα με την περιγραφείσα μέθοδο παρουσιάζεται διαγραμματικά στο Σχήμα 6.5. Αναλυτικότερα, η εικόνα προς αναγνώριση μετασχηματίζεται έτσι ώστε το περίγραμμά της να ταιριάζει όσο το δυνατόν καλύτερα με το περίγραμμο αναφοράς για κάθε κλάση. Ο μετασχηματισμός αυτός υπολογίζεται είτε ως αφινικός με την απλή μέθοδο των shape context είτε ως ομογενής με την προτεινόμενη παραλλαγή. Για το τελευταίο στάδιο της ταξινόμησης υπολογίζουμε την πιθανότητα η περιγραφή (με την μέθοδο HOG) της μετασχηματισμένης εικόνας να ανήκει στην κλάση αναφοράς ($P(\omega_i|h_i)$, όπου ω_i η i -οστή κλάση). Η πιθανότητα αυτή υπολογίζεται στην πράξη (negative log-prob) ως η mahalanobis απόσταση για κάθε κλάση είτε απλούστερα ως η ευκλείδεια απόσταση από το κέντρο κάθε κλάσης. Τέλος, η αρ-

χική εικόνα ταξινομείται στην κλάση με την μεγαλύτερη πιθανότητα ή ισοδύναμα με την μικρότερη απόσταση.



(α') Διαγραμματική αναπαράσταση των σημαντικότερων σταδίων της προτεινόμενης μεθόδου για την ταξινόμηση της χειρομορφής εισόδου σε μια από τις 24 πιθανές κλάσεις

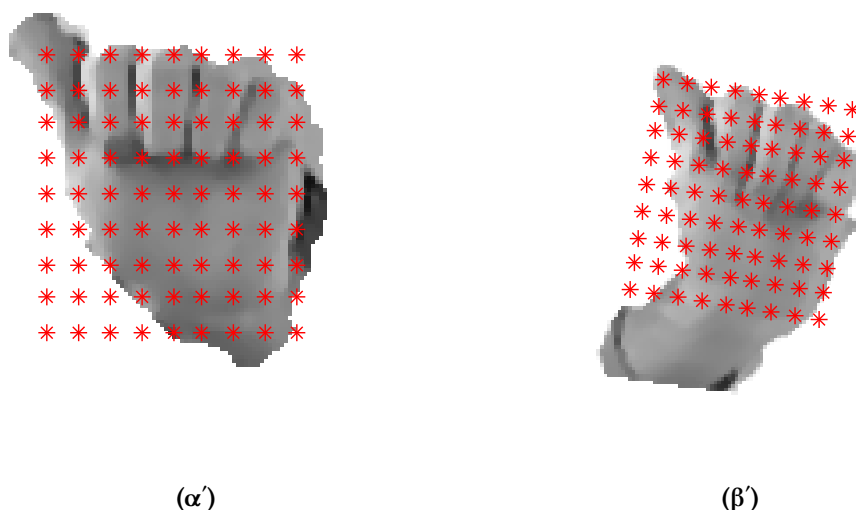


(β') Αναλυτική περιγραφή του block για τον μετασχηματισμό της εικόνας σύμφωνα με την μέθοδο των Shape Context. Το περίγραμμα αναφοράς αντιστοιχεί στο medoid για την εκάστοτε κλάση

Σχήμα 6.5: Αναπαράσταση του τελικού συστήματος αναγνώρισης σύμφωνα με την προτεινόμενη μέθοδο

Η εφαρμογή του μετασχηματισμού που προκύπτει απευθείας στην αρχική εικόνα είναι μια διαδικασία αυξημένου υπολογιστικού κόστους. Αντ' αυτού μπορούμε να εφαρμόσουμε τον αντίστροφο μετασχηματισμό μόνο στο grid της εικόνας αναφοράς, όπως φαίνεται στο παράδειγμα της εικόνας 6.6, και να χρησιμοποιήσουμε το μετασχηματισμένο grid στην αρχική εικόνα για να εξάγουμε την HOG (ή CentHOG) περιγραφή της.

Βέβαια μια τέτοια διαδικασία απαιτεί και την κατάλληλη ανεξαρτησία των κελιών ως προς τον μετασχηματισμό, η οποία επιτυγχάνεται με απλή αλλαγή της κατεύθυνσης της πρώτης παραγώγου της HOG μεθόδου κατά γωνία ίση με την εκτιμώμενη γωνία περιστροφή του προκύπτοντος μετασχηματισμού (είτε με κάποια άλλη τεχνική τοπικής ανεξαρτησία κελιών όπως αυτές που θα περιγράψουμε σε επόμενη ενότητα). Όπως είναι λογικό η προσέγγιση αυτή υπερτερεί τόσο σε υπολογιστικό κόστος, αφού απαιτείται ο μετασχηματισμός πλήθους σημείων ίσο με αυτό των τοπικών περιγραφών, όσο και σε ακρίβεια αφού ο μετασχηματισμός κάποιας εικόνας είναι πολύ πιθανό να εισάγει ανακρίβειες (artifacts) λόγω «τεχνητών» pixels που δημιουργούνται με την μέθοδο της παρεμβολής (interpolation) για την επαναδειγματοληψία της εικόνας μετά τον μετασχηματισμό.



Σχήμα 6.6: Μετασχηματισμός πλέγματος (σύμφωνα με την αντιστοίχιση στο Σχήμα6.4). (α) πλέγμα αναφοράς (β) μετασχηματισμένο πλέγμα

6.2 Περιστροφικά Ανεξάρτητα HOG

Στην ενότητα αυτή θα προτείνουμε μια πιο «συμπαγή» μέθοδο για περιστροφική ανεξαρτησία βασιζόμενοι σε κυκλική δειγματοληψία του χώρου για την εξαγωγή τοπικών περιγραφών. Συγκεκριμένα επικεντρωνόμαστε στην διαδικασία εξαγωγής ενός ενιαίου διανύσματος χαρακτηριστικών, το οποίο να είναι ανεξάρτητο στην περιστροφή της εικόνας.

Η μέθοδος που αναπτύχθηκε αποτελεί ουσιαστικά μια επέκταση της βασικής μεθόδου HOG και παρακάτω θα παρουσιαστούν οι επιμέρους τεχνικές που επιλέχθηκαν. Τα 3 βασικά στάδια για την εξαγωγή ενός τέτοιου διανύσματος χαρακτηριστικών είναι

- 1 Η κατάλληλη χωρική διαμέριση της εικόνας με άξονα την κυκλική ανεξαρτησία.

- 2 Η δημιουργία τοπικών περιγραφών για κάθε κελί (cell) ανεξάρτητων σε περιστροφικές μετατοπίσεις.
- 3 Η κατασκευή ενός ενιαίου διανύσματος με την ιδιότητα της πλήρους ανεξαρτησίας στην περιστροφή.

Εν τέλει στην ενότητα αυτή επικεντρωνόμαστε πρωτίστως στην προσπάθεια δημιουργίας μιας γενικής μεθοδολογίας για περιστροφική ανεξαρτησία, ενώ παράλληλα αναλύουμε τις προϋποθέσεις που πρέπει να ικανοποιούνται για τον σκοπό αυτό. Η μεθοδολογία αυτή μπορεί να εφαρμοστεί άμεσα στις μεθόδους RHOG, ERHOG και στις δυαδικές παραλλαγές τους (θα γίνονται σύντομες αναφορές όπου κρίνεται αναγκαίο), αλλά για λόγους απλότητας θα αναφερόμαστε κυρίως στην απλή κυκλική δομή των RHOG.

6.2.1 Κυκλική Διαμέριση του Χώρου

Στόχος μας είναι να υιοθετήσουμε κάποια κυκλική διάταξη των τοπικών περιγραφών στον χώρο, διατηρώντας βέβαια την ικανότητα διάκρισης της συνολικής περιγραφής που σχετίζεται άμεσα με την πυκνότητα των τοπικών κατανομών. Για τον σκοπό αυτό θεωρούμε κυκλικό πλέγμα ομόκεντρων κύκλων, όπου κάθε κελί έχει το ίδιο μέγεθος και για τον λόγο αυτό κύκλος μεγαλύτερης ακτίνας αντιστοιχεί σε περισσότερα κελιά, όπως ακριβώς αναλύσαμε στο προηγούμενο Κεφάλαιο στην ενότητα των RHOG. Για τις ανάγκες της γενικής σταθερότητας της περιγραφής, ιδιαίτερα ως προς την περιστροφή, θεωρούμε αναγκαία και την επικάλυψη κελιών σε κάποιο σημαντικό βαθμό.

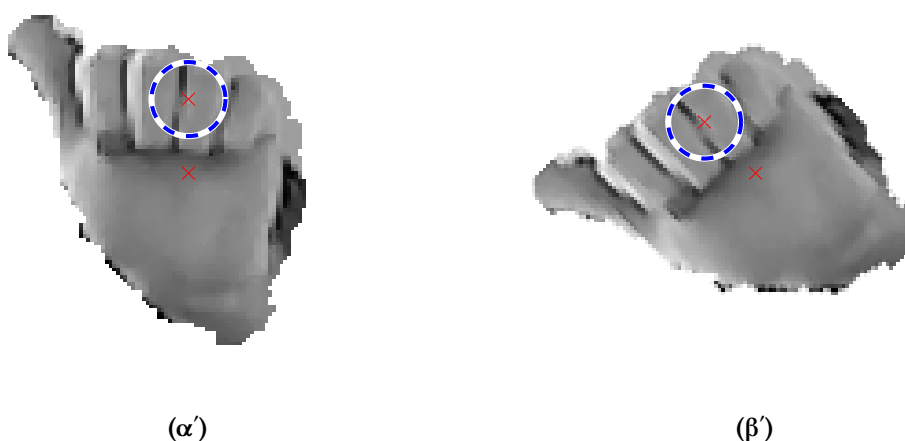
Η επιλογή κυκλικών διατάξεων στον χώρο των τοπικών περιγραφών γίνεται έτσι ώστε μια κυκλική περιστροφή της εικόνας να «μεταφράζεται» σε μετατόπιση των περιγραφών, δηλαδή σε μια πιο απλή και διαχειρίσιμη πράξη, όπου πλέον η κυκλική περιστροφή μπορεί να υπολογισθεί μέσω της δημοφιλούς τεχνικής της ετεροσυσχέτισης. Γενικά η τακτική αυτή είναι ιδιαίτερα διαδεδομένη για polar και logpolar δειγματοληψίες του χώρου ([18],[17]). Εντούτοις, όπως αναφέραμε και στο προηγούμενο κεφάλαιο, για την δημιουργία μια ενιαίας κυκλικής κατανομής των τοπικών περιγραφών, επιθυμούμε το πλήθος των τοπικών περιγραφών να αυξάνεται σε κάθε κύκλο σε αντιστοιχία με το εμβαδό του εκάστοτε δακτυλίου που ορίζει τον χώρο περιγραφής γύρω από κάθε κύκλο.

Συνοψίζοντας τα παραπάνω, ουσιαστικά λοιπόν δειγματολειπούμε κάθε κύκλο-δακτύλιο ανάλογα με την ακτίνα του (και κατ'επέκταση την περίμετρο του κύκλου ή το εμβαδό του δακτυλίου) και η δειγματοληψία αυτή γίνεται ως προς τοπικές περιγραφές και όχι ως προς συγκεκριμένα σημεία. Οι επιλογές αυτές διαφέρουν από την κύρια κατεύθυνση της βιβλιογραφίας (polar κατανομές, δειγματοληψία σημείων για μικρές περιοχές), αλλά διατηρούν την βασική ιδέα της μετατροπής της περιστροφής σε μετατόπιση.

Επιπλέον, όπως είναι λογικό, η περίπτωση των ελλειπτικών χωρικών κατανομών προσφέρει παρόμοιες δυνατότητες διαχείρισης και ικανοποιεί τις προϋποθέσεις για την δημιουργία περιστροφικά ανεξάρτητων περιγραφών, αντιστοιχίζοντας κάθε έλ-λειψη της περιγραφής αυτής σε έναν κύκλο της απλής κυκλικής περιγραφής. Η πα-ραλλαγή αυτή ουσιαστικά προβάλλει την προσεγγιστικά ελλειπτική κατανομή των τοπικών περιγραφών σε μια κυκλική διάταξη.

6.2.2 Ανεξαρτησία Τοπικών Περιγραφών ως προς το Κέντρο της Εικόνας

Η ζητούμενη συμπεριφορά μπορεί να περιγραφεί απλά με την βοήθεια των εικόνων του σχήματος 6.7. Για το κυκλικό κελί (το οποίο σημειώνεται με τον μπλε κύκλο) της εικόνας (α) και το αντίστοιχο της εικόνας (β), η οποία είναι μια περιστροφή της εικόνας (α), είναι λογικό να επιθυμούμε να έχουν την ίδια τοπική περιγραφή, αφού περιέχουν την ίδια πληροφορία (υπό διαφορετική οπτική γωνία). Κάτι τέτοιο δεν είναι δυνατόν να μας το προσφέρει η απλή μέθοδος HOG καθώς η παράγωγος της εικόνας υπολογί-ζεται ως προς τους άξονες x και y . Συνεπώς απαιτείται κάποια άλλη περιγραφή έτσι ώστε οι άξονες οπου προβάλλεται η παράγωγος να είναι ανεξάρτητη ως προς την περι-στροφή της εικόνας. Αυτό για να επιτευχθεί γίνεται χρήση κάποιου σημείου αναφοράς ανεξάρτητου της περιστροφής. Για τον σκοπό αυτό υπάρχουν οι εξής εναλλακτικές: 1) Υπολογισμός της παραγώγου ως προς το κέντρο του κελιού 2) Υπολογισμός της παραγώγου ως προς το κέντρο της εικόνας, το οποίο προφανώς δεν εξαρτάται από περιστροφές.



Σχήμα 6.7: Μια εικόνα και η περιστροφή της (45°), οπου επισημαίνεται με έναν μπλε κύκλο το ίδιο κελί

Επιπρόσθετα, υπάρχει και η εναλλακτική της εύρεσης της κύριας κατεύθυνσης κάθε τοπικής περιγραφής για την επίτευξη της ανεξαρτησίας των κελιών ως προς την περιστροφή, όπως υλοποιείται και στην SIFT μέθοδο. Κάτι τέτοιο, όμως, δεν είναι εφικτό

στην περίπτωση μας, καθώς είναι πολύ πιθανό να μην υπάρχει μόνο μια κύρια κατεύθυνση σε κάθε κελί και η επιλογή της δημιουργίας πολλαπλών περιγραφών για κάθε περιοχή, μια για κάθε κύρια κατεύθυνση, είναι πρακτικά μη εφαρμόσιμη, εφόσον επιθυμούμε μια ενιαία περιγραφή.

Μετασχηματισμός Ακτινικής Παραγώγου - Radial Gradient Transform

Η ζητούμενη περιγραφή προκύπτει σε κάθε περίπτωση βρίσκοντας την ακτινική παράγωγο (radial gradient), που ορίζεται ως η προβολή του αρχικού διανύσματος της παραγώγου στην ακτινική διεύθυνση και στην κάθετη αυτής, ως προς κάποιο σημείο αναφοράς. Η προσέγγιση αυτή χρησιμοποιείται στο [25] για περιστροφικά ανεξάρτητες τοπικές περιγραφές. Συγκεκριμένα, σύμφωνα με το [25], εάν c το κέντρο αναφοράς, p ένα σημείο και R_θ το μητρώο περιστροφής κατά θ , ορίζουμε το νέο τοπικό (ακτινικό) σύστημα συντεταγμένων για το σημείο p ως:

$$r = \frac{p - c}{\|p - c\|} \quad \text{ακτινική κατεύθυνση}, \quad t = R_{\frac{\pi}{2}} r \quad \text{εφαπτομενική κατεύθυνση}$$

Προβάλλοντας την παράγωγο g στο νέο σύστημα συντεταγμένων (r, t) το διάνυσμα της παραγώγου πλέον γράφεται ως $(g^T r, g^T t)$. Εάν περιστρέψουμε το σημείο p και κατ'επέκταση την παράγωγο κατά γωνία θ γύρω από το κέντρο έχουμε τους εξής συμβολισμούς:

$$p' = R_\theta p, \quad g' = R_\theta g, \quad r' = R_\theta r, \quad t' = R_\theta t$$

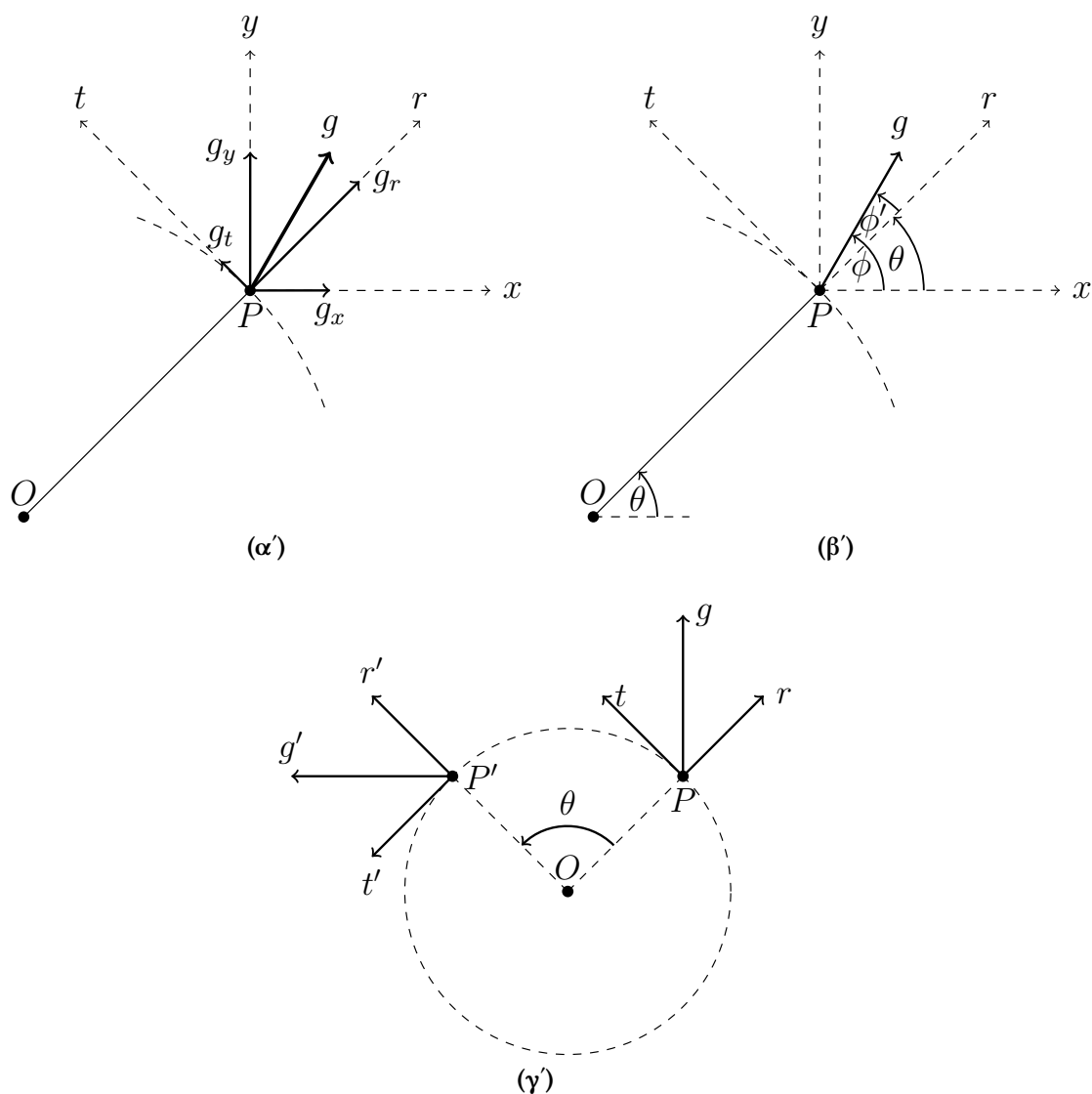
Εν τέλει, η παράγωγος g' στο νέο σύστημα συντεταγμένων (r', t') έχει την ίδια περιγραφή με την g στο (r, t) , όπως αποδεικνύεται παρακάτω:

$$\begin{aligned} (g'^T r', g'^T t') &= ((R_\theta g)^T (R_\theta r), (R_\theta g)^T (R_\theta t)) \\ &= (g^T R_\theta^T R_\theta r, g^T R_\theta^T R_\theta t) \\ &= (g^T r, g^T t) \end{aligned}$$

Η προβολή της παραγώγου στο εκάστοτε τοπικό ακτινικό σύστημα συντεταγμένων απεικονίζεται στο Σχήμα 6.8 (α) και η ανεξαρτησία περιστροφής με χρήση της ακτινικής παραγώγου φαίνεται παραστατικά στο Σχήμα 6.8 (γ).

Στην περίπτωση όπου είναι επιθυμητή μια περιγραφή στο ακτινικό σύστημα συντεταγμένων ως προς μέτρο και κατεύθυνση (γωνία), γίνονται οι ακόλουθες μετατροπές με την διαδικασία να απεικονίζεται στο Σχήμα 6.8 (β):

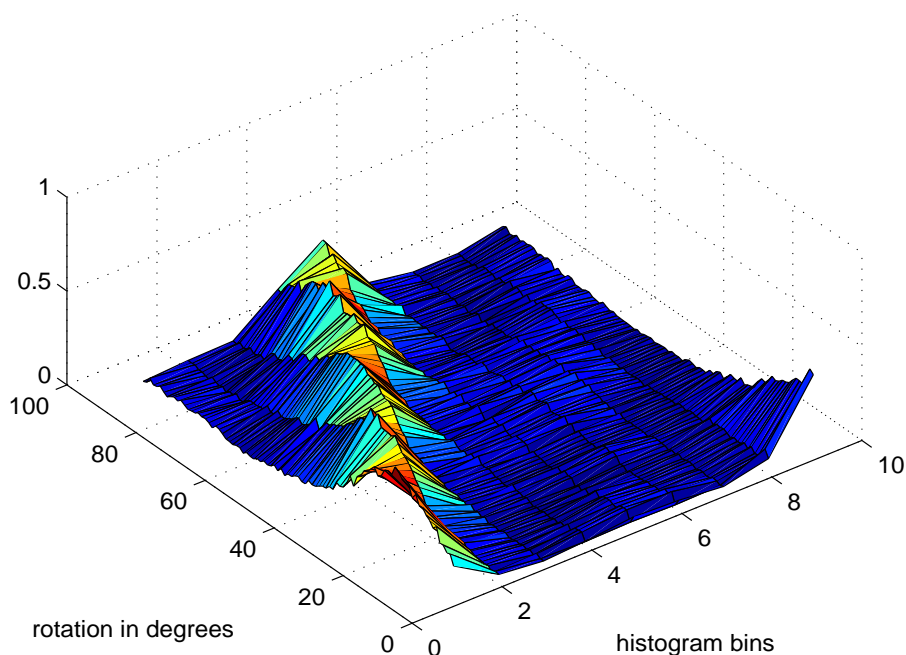
- Το μέτρο της παραγώγου μένει ίδιο, αφού αλλάζουμε μόνο τους άξονες αναφοράς.
- Η γωνία της παραγώγου ϕ' ως προς το ακτινικό σύστημα συντεταγμένων είναι η διαφορά της γωνίας ϕ της παραγώγου ως προς το αρχικό σύστημα συντεταγμένων με την γωνία θ που σχηματίζουν ο άξονας x και η ευθεία που ορίζεται από το κέντρο αναφοράς και το εκάστοτε σημείο p .



Σχήμα 6.8: Γεωμετρική απεικόνιση της χρήσης του ακτινικού συστήματος συντεταγμένων
 (α) προβολή του διανύσματος της παραγώγου στο ακτινικό σύστημα συντεταγμένων
 (β) εναλλακτική περιγραφή ως μέτρο-γωνία της παραγώγου στο ακτινικό σύστημα συντεταγμένων
 (γ) παράδειγμα περιστροφικής ανεξαρτησίας της περιγραφής της παραγώγου με χρήση ακτινικού συστήματος συντεταγμένων

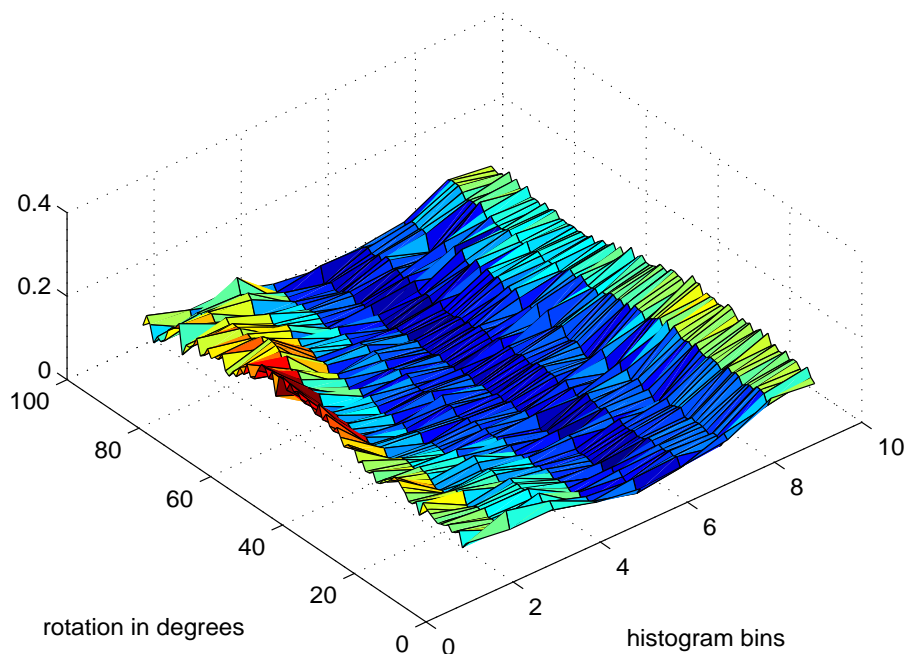
Επιλογή Κέντρου Αναφοράς

Για μια επιλεγμένη περιοχή μιας εικόνας αναπαριστούμε την τοπική περιγραφή ως το ιστόγραμμα κατευθύνσεων και την περιστρέφουμε για να ελέγξουμε την περιστροφική ανεξαρτησία του ακτινικού μετασχηματισμού. Για την βασική μέθοδο των HOG του Σχήματος 6.9, παρατηρείται η αναμενόμενη μετατόπιση του ιστογράμματος. Αντίθετα, στα Σχήματα 6.10 και 6.11 (RGT ως προς το κέντρο του κελιού και ως προς το κέντρο της εικόνας αντίστοιχα) εμφανίζεται η ζητούμενη συμπεριφορά, καθώς τα ιστογράμματα παρουσιάζουν παρόμοια μορφή κατά την περιστροφή της εικόνας.

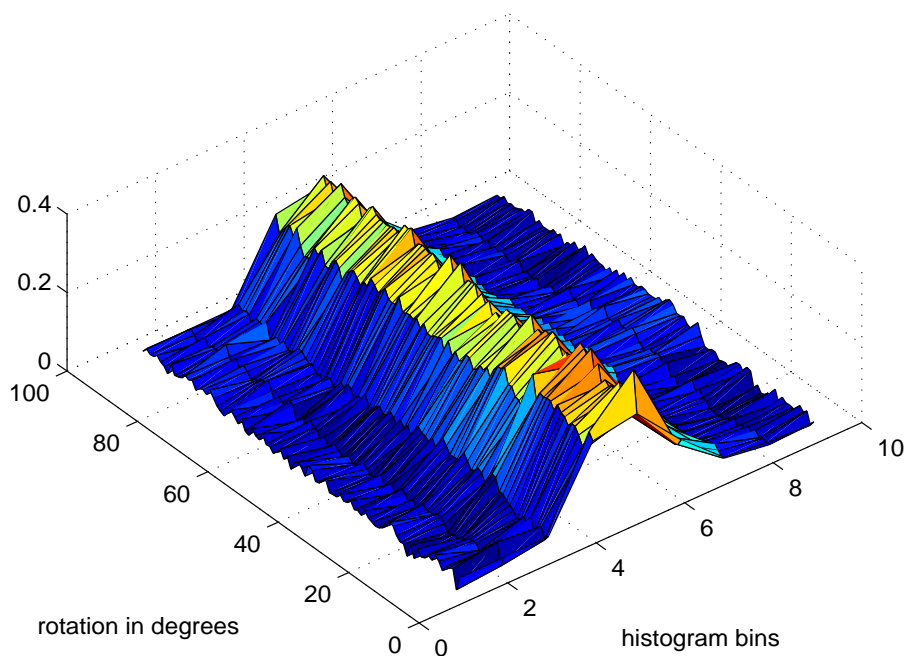


Σχήμα 6.9: Επίδραση της περιστροφής στην τοπική περιγραφή - Απλή μέθοδος HOG

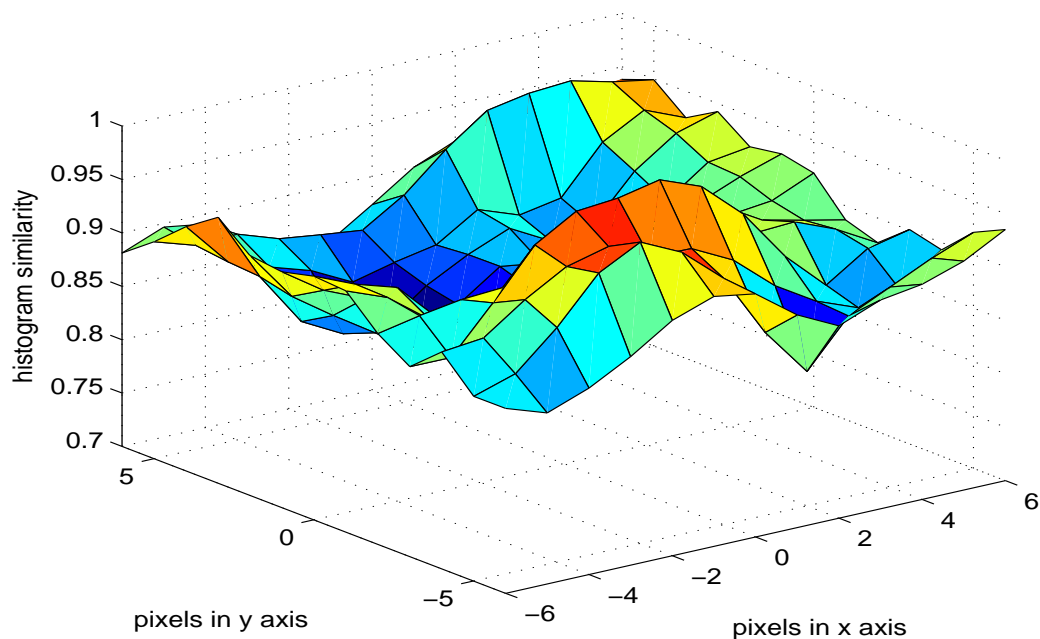
Βέβαια, η ανεξαρτησία ως προς το κέντρο του κελιού έχει ένα σημαντικό μειονέκτημα, καθώς μικρές μετατοπίσεις του κελιού στον χώρο είναι πιθανό να προκαλέσουν σημαντικές αλλαγές στην τοπική περιγραφή. Η παραπάνω παρατήρηση καθιστά την περιγραφή αυτή μια καλή λύση για μεθόδους που βασίζονται σε σημεία ενδιαφέροντος, τα οποία αποτελούν και τα κέντρα των τοπικών περιγραφών. Εντούτοις, στην παρούσα προσέγγιση επιλέγεται πυκνή διαμέριση του χώρου και συνεπώς οι προκύπτουσες τοπικές περιγραφές πρέπει να έχουν κάποιο βαθμό ανεξαρτησίας σε μικρές μετατοπίσεις στον χώρο. Την ιδιότητα αυτή την παρέχει η περιγραφή κάθε κελιού ως προς το κέντρο της εικόνας, αφού μικρές μετατοπίσεις αντιστοιχούν σε μικρές αλλαγές ως προς τις ακτινικές κατευθύνσεις της πρώτης παραγώγου. Τα παραπάνω επιβεβαιώνονται και πειραματικά στα σχήματα 6.12 και 6.13, όπου γίνεται χρήση της απόστασης ομοιότητας συνημίτονου (cosine similarity distance) για την παρουσίαση της μεταβολής των ιστογραμμάτων με την κάθε τεχνική.



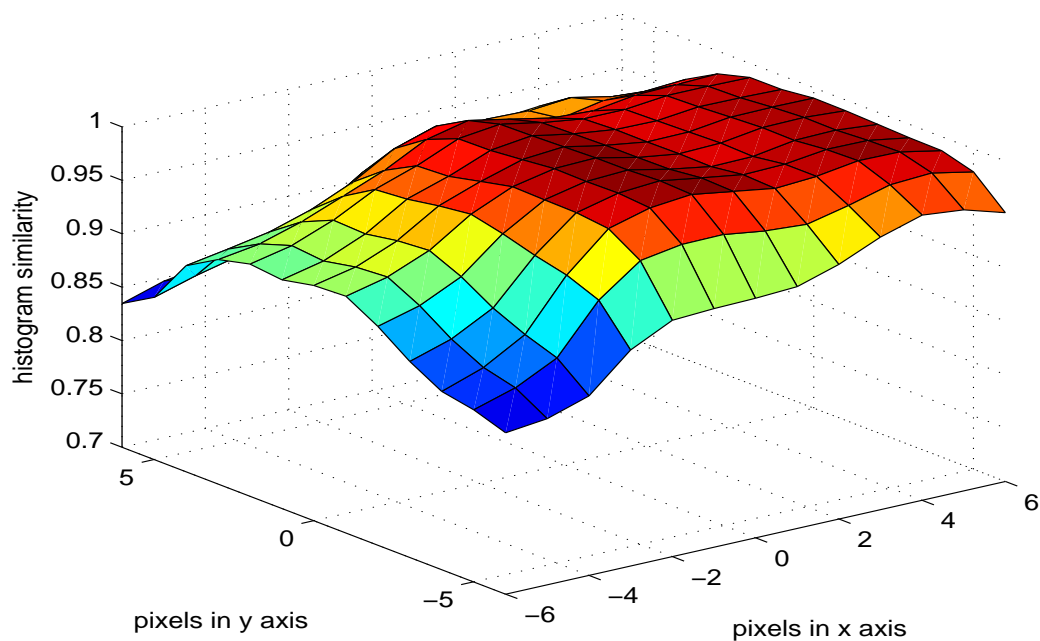
Σχήμα 6.10: Επίδραση της περιστροφής στην τοπική περιγραφή - Περιγραφή ως προς το κέντρο του κελιού



Σχήμα 6.11: Επίδραση της περιστροφής στην τοπική περιγραφή - Περιγραφή ως προς το κέντρο της εικόνας



Σχήμα 6.12: Επίδραση μικρών μετατοπίσεων στην τοπική περιγραφή - Το κέντρο του κελιού ως σημείο αναφοράς



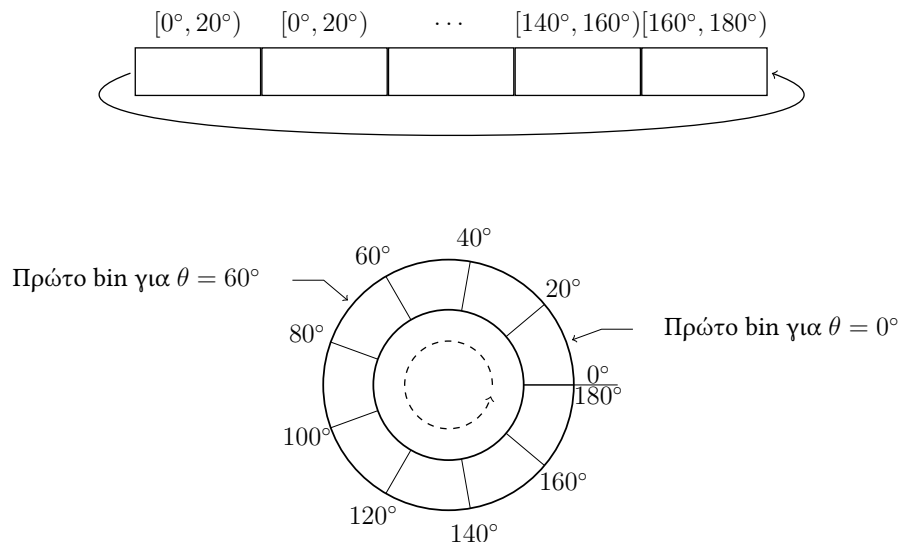
Σχήμα 6.13: Επίδραση μικρών μετατοπίσεων στην τοπική περιγραφή - Το κέντρο της εικόνας ως σημείο αναφοράς

Προσεγγίσεις και Παραλλαγές των Τοπικών Περιγραφών

Ακολουθώς, θα μελετήσουμε παραλλαγές της προαναφερθείσας τεχνικής για ανεξαρτησία τοπικών περιγραφών (ιστογραμμάτων κατεύθυνσης) ως προς το κέντρο της εικόνας, με στόχο μια υπολογιστικά γρήγορη και αποδοτική τεχνική.

- 1 Μια πρώτη προσέγγιση αποτελεί η επιλογή της γωνίας $\bar{\theta}$ που αντιστοιχεί στο κέντρο του κελιού ως κοινή για όλο το κελί, περιορίζοντας το σύνολο των υπολογισμών για τον ακτινικό μετασχηματισμό που αντιστοιχεί σε κάθε pixel του κελιού. Μια τέτοια προσέγγιση είναι ρεαλιστική, εάν το κελί δεν είναι πολύ μεγάλο, ενώ όσο πιο μακριά βρίσκεται από το κέντρο της εικόνας, τόσο καλύτερη είναι και η προσέγγιση.
- 2 Είναι δυνατόν να επιτευχθεί επιπλέον μείωση του υπολογιστικού κόστους, σύμφωνα με την παρατήρηση ότι η αφαίρεση που πραγματοποιείται για να υπολογιστεί η κατεύθυνση της ακτινικής παραγώγου, μετατοπίζει τις κατευθύνσεις γύρω από την γωνία $\bar{\theta}$ σε κατευθύνσεις γύρω από μηδενική γωνία, δηλαδή στο πρώτο bin του κατασκευασθέντος ιστογράμματος και ούτω καθεξής.

Συνεπώς, το νέο ιστόγραμμα μπορεί να προκύψει προσεγγιστικά με κυκλική μετάθεση προς τα αριστερά του αρχικού ιστογράμματος της βασικής μεθόδου, κατά αριθμό bins ίσο με την στρογγυλοποίηση του $\bar{\theta}/\theta_b$, όπου $\theta_b = 180^\circ/\#bins$ και είναι το εύρος γωνιών για κάθε bin. Η παραπάνω λειτουργία παριστάνεται γραφικά στο Σχήμα 6.14.



Σχήμα 6.14: Η προσέγγιση του υπολογισμού της ακτινικής παραγώγου ως αριστερή κυκλική ολίσθηση του υπολογισθέντος ιστογράμματος. Διαισθητικά είναι πιο εύκολη η χρήση ενός κυκλικού ιστογράμματος με διαφορετικό αρχικό bin. Για το παράδειγμα του σχήματος χρησιμοποιείται εύρος γωνιών $[0^\circ, 180^\circ)$ (unsigned gradient) και 9 bins.

Συγκεκριμένα, εάν m_i , ϕ_i είναι το μέτρο και η γωνία της παραγώγου για κάθε

σημείο (pixel) p_i εντός του κελιού, τότε το ιστόγραμμα υπολογίζεται ως:

$$h[k] = \left\{ \sum m_i, \phi_i \in [(k-1)\theta_b, k\theta_b] \right\}, \quad k = 1, \dots, \#bins$$

και για την περίπτωση της ακτινική παραγώγου είναι:

$$h'[k] = \left\{ \sum m_i, (\phi_i - \bar{\theta}) \in [(k-1)\theta_b, k\theta_b] \right\}, \quad k = 1, \dots, \#bins$$

Από την τελευταία σχέση είναι προφανές ότι η αφαίρεση που απαιτείται για τον υπολογισμό της γωνίας της ακτινικής παραγώγου ισοδυναμεί με κυκλική μετάθεση του ιστογράμματος, εφόσον η τιμή της $\bar{\theta}$ είναι πολλαπλάσια της θ_b :

$$\bar{\theta} = m\theta_b, \quad m \in \mathbb{Z} \Rightarrow h'[k] = h[((k+m))_B], \quad B = \#bins \quad 1$$

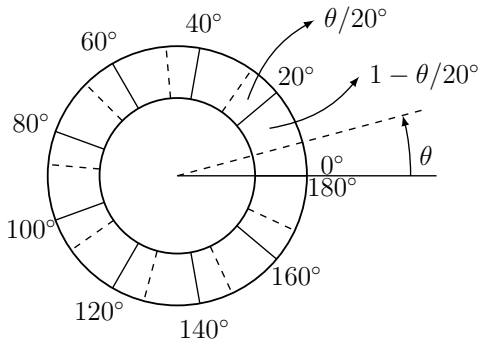
Η λύση αυτή υπερτερεί υπολογιστικά, καθώς πλέον έχουμε μόνο μια κυκλική μετάθεση του ιστογράμματος της βασικής μεθόδου. Εντούτοις, είναι προφανές ότι η ακρίβεια της προσέγγισης περιορίζεται από το εύρος γωνιών των bins και είναι αξιόπιστη για γωνίες κοντά στην τιμή θ_b . Για τον λόγο αυτό η ακρίβεια της μεθόδου αυξάνεται με την αύξηση του πλήθους των bins.

- 3 Το κύριο μειονέκτημα της παραπάνω προσέγγισης είναι η μειωμένη ακρίβεια λόγω της διακριτοποίησης που επιβάλλουν τα bins. Ένας τρόπος για να ξεπεράσουμε το πρόβλημα αυτό, είναι η παραδοχή ότι η κατανομή στα bins είναι ομοιόμορφη. Συνεπώς, είναι δυνατόν να επεκτείνουμε τον ορισμό της κυκλικής μετάθεσης από ακέραιες σε δεκαδικές τιμές, δηλαδή υπάρχει πλέον η δυνατότητα μετατόπισης ενός ποσοστού της πληροφορίας του bin.

Η επέκταση αυτή υλοποιείται με τον ακόλουθο τρόπο: ορίζουμε ως a το ακέραιο μέρος και b το δεκαδικό μέρος της διαίρεσης $\bar{\theta}/\theta_b$ και εάν h το αρχικό ιστόγραμμα, το νέο ιστόγραμμα υπολογίζεται σύμφωνα με την σχέση:

$$h'[k] = (1-b)h[((k+a))_B] + bh[((k+a+1))_B], \quad a = \lfloor \bar{\theta}/\theta_b \rfloor \& b = \bar{\theta}/\theta_b - a$$

Η διαδικασία αυτή απεικονίζεται γραφικά στο Σχήμα 6.15 για την περίπτωση μόνο δεκαδικού μέρους ($\bar{\theta} < \theta_b$ ή $a = 0$).



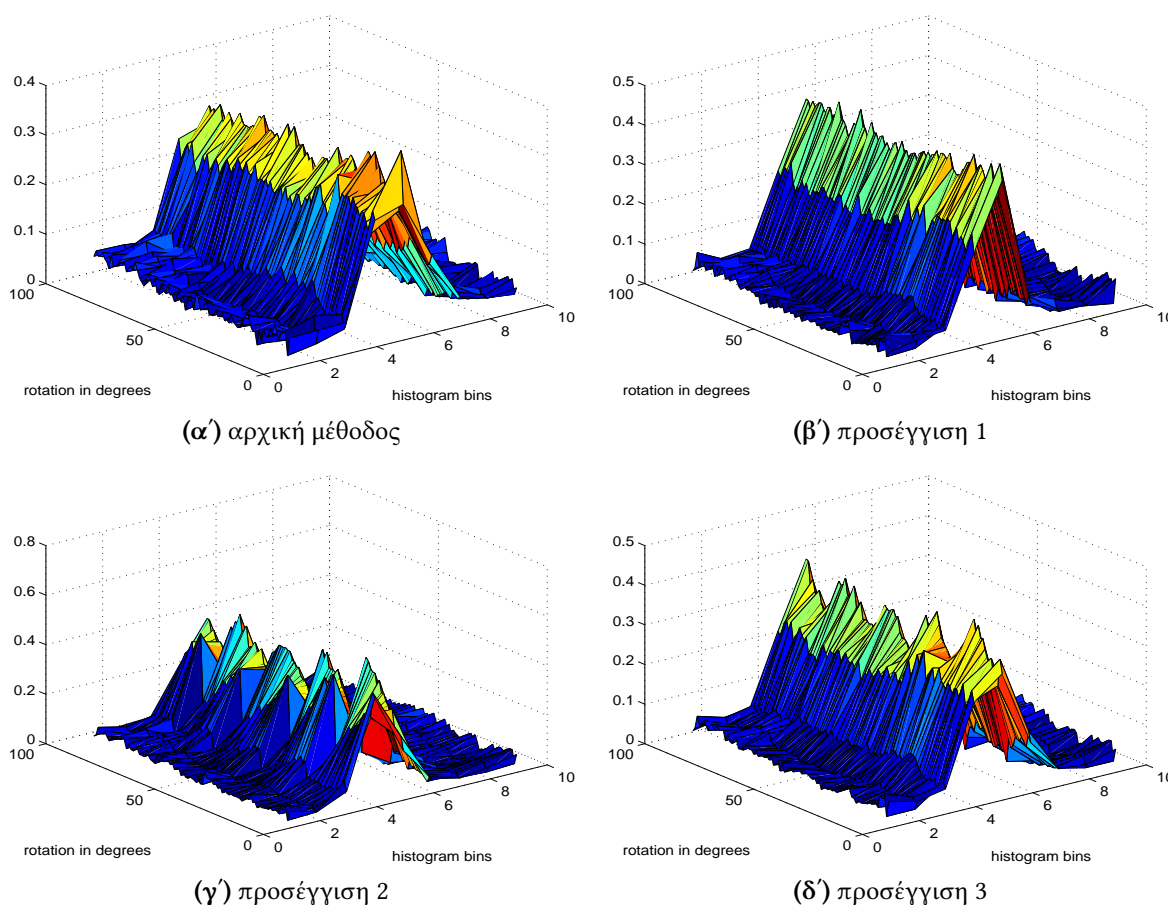
$$h'[k] = \frac{\theta}{20^\circ} h[((k+1))_B] + \left(1 - \frac{\theta}{20^\circ}\right) h[k]$$

Σχήμα 6.15: Παράδειγμα της τελικής προσέγγισης για $\theta < \theta_b$.

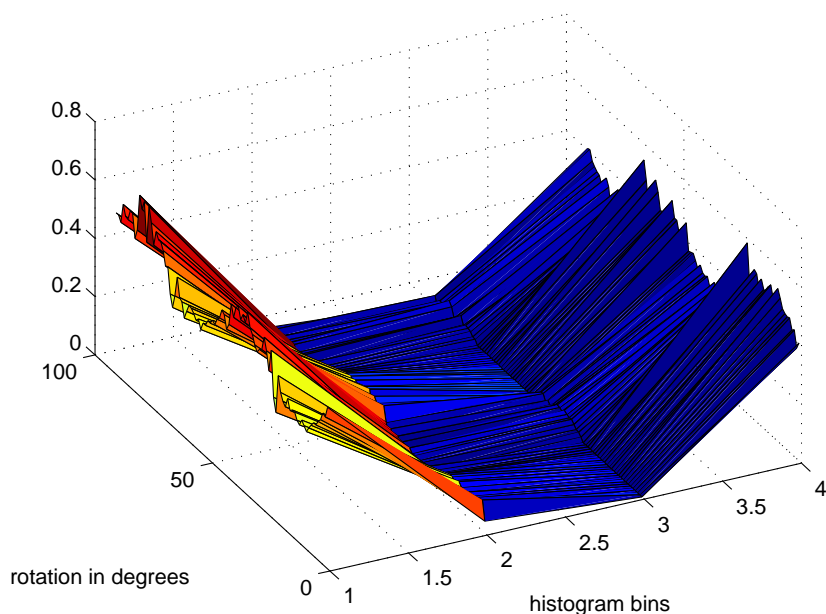
Βέβαια η παραπάνω παραδοχή είναι μη ρεαλιστική, καθώς δεν υπάρχει κάποια συγκεκριμένη λογική που να μπορεί να περιγράψει την κατανομή σε κάθε bin, εντούτοις στην πράξη έχει αρκετά αξιολογήσιμα αποτελέσματα. Η επιτυχία αυτής της προσέγγισης βελτιώνεται λίγο χρησιμοποιώντας αλληλεπικαλυπτόμενα bins, είτε χρησιμοποιώντας μεγαλύτερο αριθμό bins διότι μεγαλύτερος αριθμός bins ισοδυναμεί με μικρότερο εύρος γωνιών άρα και καλύτερη προσέγγιση της ομοιόμορφης κατανομής σε αυτά.

Τέλος, η λογική αυτή είναι εφαρμόσιμη και σε Binary HOGs, όπου τα bins στην πραγματικότητα δεν αντιστοιχούν σε εύρος γωνιών, αλλά σε μια διακριτή τιμή, οπότε είναι δυνατή η υπόθεση μιας μορφής «ομοιομορφίας», ενώ οι αρχικές προσεγγίσεις δεν είναι άμεσα εφαρμόσιμες εκτός αν θεωρήσουμε ότι όντως έχουμε ένα εύρος γωνιών 45° για κάθε bin, αναιρώντας όμως την υπολογιστική απλότητα της μεθόδου.

Στα Σχήματα 6.16 α,β,γ και δ φαίνεται η συμπεριφορά των προσεγγίσεων για ένα συγκεκριμένο κελί κατά την περιστροφή της εικόνας. Επιπλέον στο Σχήμα 6.17 παρουσιάζεται η επιτυχία της προσέγγισης αυτής σε Binary HOG.



Σχήμα 6.16: Ανεξαρτησία τοπικών περιγραφών στην περιστροφή για τις διαφορές προσεγγίσεις που εισήχθησαν



Σχήμα 6.17: Παραδειγμά της αποτελεσματικότητας της 3^{15} προσέγγισης σε δυαδικές εικόνες

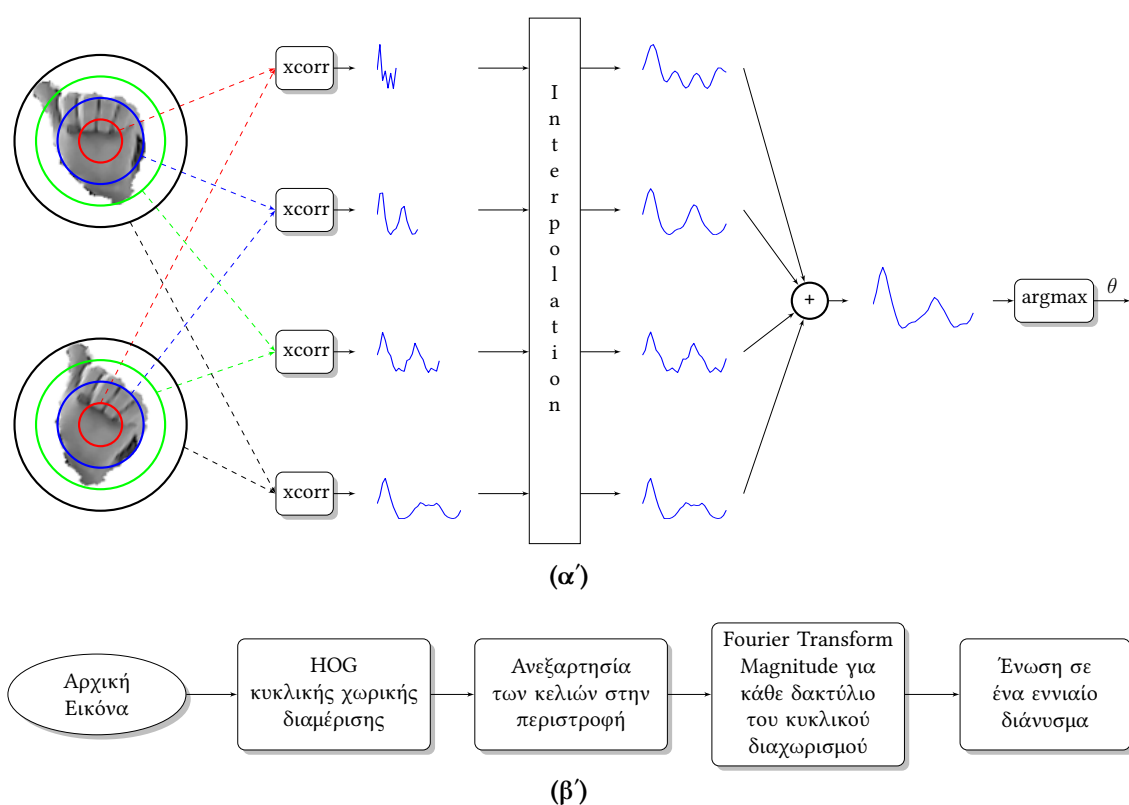
6.2.3 Εύρεση Περιστροφής και Δημιουργία Ενιαίων Ανεξάρτητων Περιγραφών

Έχοντας ως δεδομένα δυο κυκλικές περιγραφές, όπως αυτές που προτείναμε προηγουμένως, από μια εικόνα και την περιστροφή της κατά θ , είναι δυνατόν να βρούμε την περιστροφή της μιας ως προς την άλλη εφαρμόζοντας την μέθοδο cross-correlation, όπως αναπτύχθηκε σε προηγούμενη ενότητα για την παραλλαγή των Shape Context και στο Παράρτημα Γ', σε κάθε δακτύλιο και συνδυάζοντας τα αποτελέσματα.

Βέβαια η μέθοδος αυτή υποθέτει κάποια συσχέτιση μεταξύ των τοπικών περιγραφών των δακτυλίων, η οποία θα μπορούσε να εισαχθεί με την χρήση ίδιου πλήθους τοπικών περιγραφών (polar κατανομή) σε κάθε δακτύλιο ώστε η i -οστή περιγραφή του ενός δακτυλίου να αφορά την ίδια κυκλική μετατόπιση της αντίστοιχης περιγραφής των γειτονικών δακτυλίων και άρα να υπάρχει η αντίστοιχη εξάρτηση μεταξύ των i -στών στοιχείων της ετεροσυσχέτισης κάθε ζεύγους δακτυλίων. Εντούτοις παρόλο που η προτεινόμενη τεχνική έχει μεταβαλλόμενο αριθμό τοπικών περιγραφών για κάθε δακτύλιο, λειτουργεί εξίσου καλά στην πράξη χρησιμοποιώντας την μέθοδο της παρεμβολής (interpolation) για τις ετεροσυσχετίσεις. Συγκεκριμένα χρησιμοποιούμε την μέθοδο της παρεμβολής έτσι ώστε όλες οι προκύπτουσες ετεροσυσχετίσεις να έχουν ίδιο πλήθος στοιχείων με αυτό των τοπικών περιγραφών του εξωτερικού δακτυλίου. Η λογική πίσω από αυτή την κίνηση έγκειται στο γεγονός ότι για μικρές αποστάσεις από το κέντρο η συχνή δειγματοληψία (μεγάλο πλήθος τοπικών περιγραφών) παρουσιάζει πλεονασμό και δεν είναι αναγκαία, οπότε μετέπειτα μπορούμε να την προσομοιάσουμε

με την χρήση της μεθόδου της παρεμβολής. Η παραπάνω διαδικασία συνοψίζεται διαγραμματικά στο Σχήμα 7.36 (α).

Για την δημιουργία ενός ενιαίου διάνυσματος ανεξάρτητου σε κυκλικές περιστροφές γίνεται χρήση του Fourier Transform Magnitude (FTM), μια πολύ δημοφιλής τεχνική για τον σκοπό αυτό έχει χρησιμοποιηθεί συχνά σε παρόμοια προβλήματα ([6],[18],[17]). Για κάθε δακτύλιο θεωρούμε κάθε κοινό bin των ιστογραμμάτων των σημείων (π.χ. το 1ο bin για όλα τα σημεία του δακτυλίου και ούτω καθεξής) ως μια ξεχωριστή ροή πληροφορίας (information channel) και εφαρμόζουμε τον μετασχηματισμό Fourier πάνω σε αυτή, κρατώντας τελικά μόνο το μέτρο του μετασχηματισμού για να αποβάλλουμε της εξαρτήσεις από την περιστροφή, όπως ακριβώς περιγράψαμε και για την παραλλαγή των Shape Context. Το τελικό διάνυσμα είναι η ένωση των περιγραφών των δακτυλίων, οι οποίες με την σειρά τους είναι η ένωση των FTM των επιμέρους «ροών πληροφορίας» κάθε δακτυλίου. Η περιγραφείσα μέθοδος απεικονίζεται στο Σχήμα 7.36 (β).



Σχήμα 6.18: Ανάλυση των δυο εναλλακτικών διαδικασιών για το τελικό στάδιο των κυκλικών περιγραφών:

(α) μέθοδος για την εύρεση της περιστροφής μεταξύ δύο εικόνων βασιζόμενη στην ετεροσυσχέτιση των HOG περιγραφών των επιμέρους δακτυλίων για την κατασκευή μιας ενιαίας ετεροσυσχέτισης

(β) δημιουργία περιστροφικά ανεξάρτητων ενιαίων περιγραφών

Κεφάλαιο 7

Πειραματικό Μέρος

Το κεφάλαιο αυτό περιλαμβάνει όλα τα πειράματα που διεξήχθησαν έτσι ώστε να αξιολογήσουμε την αποτελεσματικότητα των προτεινόμενων μεθόδων.

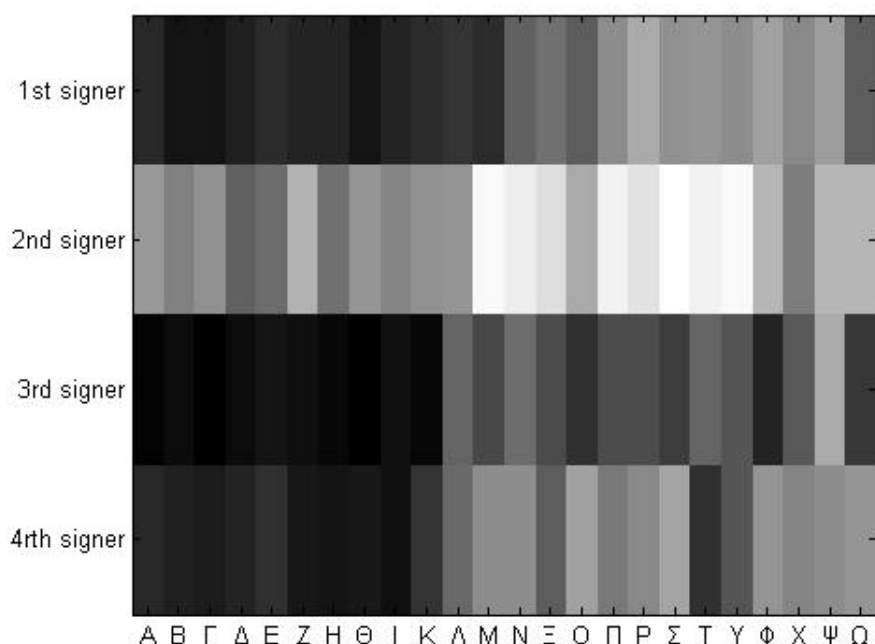
7.1 Περιγραφή της Βάσης Δεδομένων

Η βάση δεδομένων που χρησιμοποιήθηκε στο πειραματικό μέρος αποτελείται από τις χειρομορφές των γραμμάτων της Ελληνικής Νοηματικής Γλώσσας για τέσσερις διαφορετικούς νοηματιστές. Η βάση αυτή δημιουργήθηκε χρησιμοποιώντας το Kinect της Microsoft, όπου η εξαγωγή της χειρομορφής πραγματοποιήθηκε με χρήση της πληροφορίας βάθους που προσφέρει το kinect, απλοποιώντας αρκετά την συγκεκριμένη διαδικασία. Συγκεκριμένα η μάσκα της χειρομορφής αποκτήθηκε χρησιμοποιώντας ένα κατώφλι στην τιμή του βάθους σε ένα παράθυρο γύρω από το σημείο του χεριού, αξιοποιώντας έτσι σε μεγάλο βαθμό την επιπλέον πληροφορία που μας προσφέρει το kinect, όπως η εκτίμηση του σκελετού του νοηματιστή και κατ'επεκτάση μια εκτίμηση της θέσης των χεριών του στον χώρο αλλά και της πληροφορίας βάθους. Βέβαια η ευκολία εξαγωγής οφείλεται και στην απλότητα των λήψεων, όπου η θέση του χεριού ως προς την κάμερα είναι πιο κοντά σε σχέση με το υπόλοιπο σώμα, έτσι ώστε να είναι δυνατή η άμεση αξιοποίηση της πληροφορίας του βάθους. Στο Σχήμα 7.2 παρουσιάζονται ενδεικτικά (grayscale) χειρομορφές από κάθε κλάση-γράμμα της βάσης.

Το ποσοστό εμφάνισης των γραμμάτων στην βάση ανά νοηματιστή παρουσιάζεται στον πίνακα 7.1 και πιο παραστατικά η αναλογία του κάθε γράμματος ανά νοηματιστή στο Σχήμα 7.1.

1 ^{ος} νοηματιστής	2 ^{ος} νοηματιστής	3 ^{ος} νοηματιστής	4 ^{ος} νοηματιστής
21.37 %	42.17 %	14.36 %	22.12 %

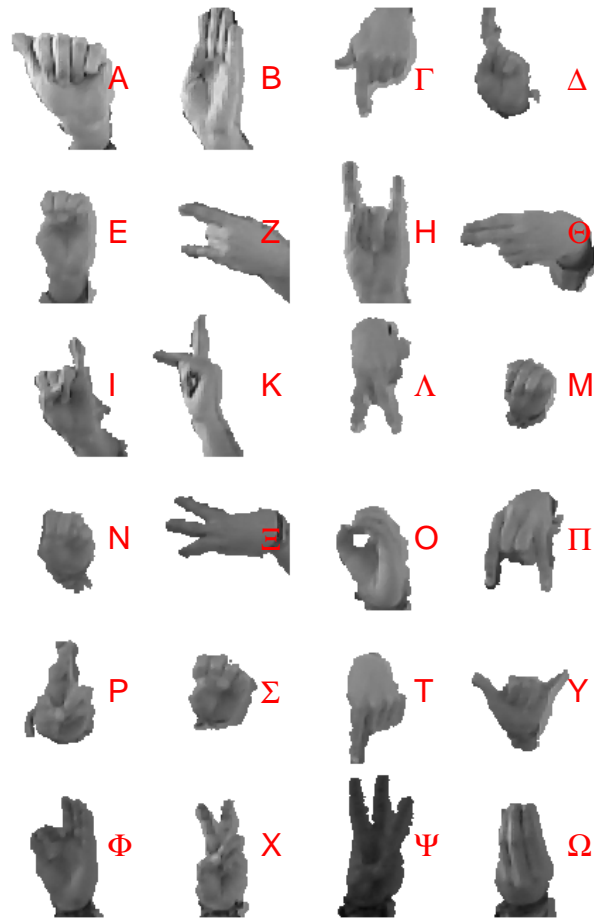
Πίνακας 7.1: Συνολικά ποσοστά χειρομορφών ανά νοηματιστή



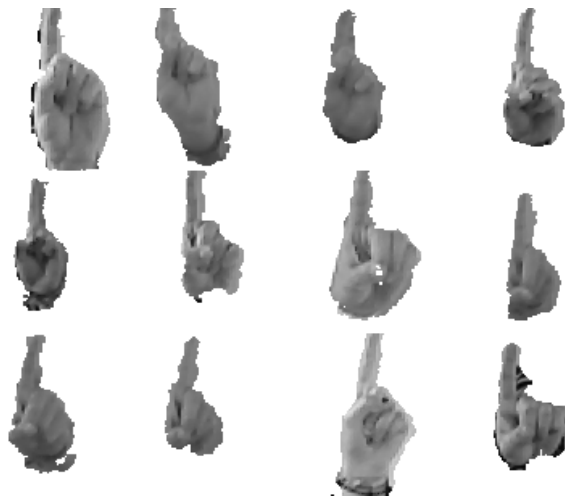
Σχήμα 7.1: Αναπαράσταση πίνακα ποσοτών εμφάνισης ανά νοηματιστή

Οι χειρομορφές της δημιουργηθείσας βάσης παρουσιάζουν αρκετή ποικιλομορφία τόσο σε συνθήκες φωτισμού όσο και στο ποσοστό εμφάνισης του καρπού. Επιπλέον εμφανίζεται σχετική μεταβολή στην γωνία όρασης των χειρομορφών (σχετίζεται με την τρισδιάστατη πόζα του χεριού που πρόκειται για ένα αρκετά σύνθετο πρόβλημα). Τα παραπάνω σε συνδυασμό με την οπτική ομοιότητα κάποιων γραμμάτων της Ελληνικής Νοηματικής Γλώσσας δυσχεραίνουν την διαδικασία αναγνώρισης και προσομοιώνουν την ποικιλομορφία που εμφανίζει ένα πραγματικό σύστημα αναγνώρισης. Στο Σχήμα 7.3 απεικονίζονται κάποιες τυχαία επιλεγμένες χειρομορφές για το γράμμα Δ με σκοπό την ανάδειξη της ποικιλομορφίας των χειρομορφών στην βάση .

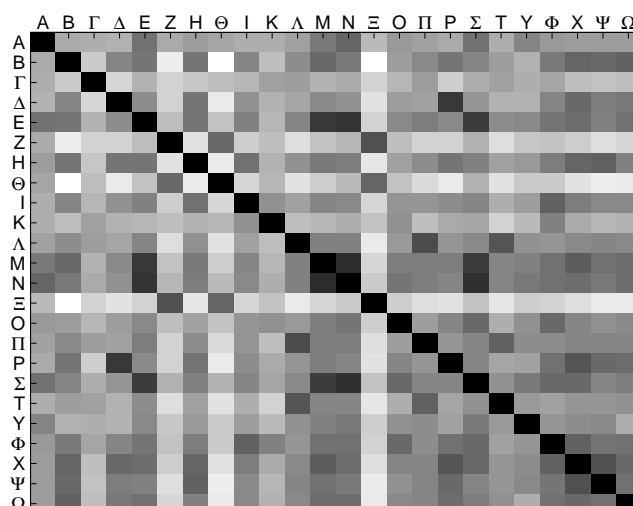
Για την εξαγωγή μιας πρώτης εκτίμησης της συσχέτισης των κλάσεων θα υπολογίσουμε τις αποστάσεις των κέντρων για τις κλάσεις με χρήση δυο διαφορετικών περιγραφών. Στον πίνακα του σχήματος 7.4 παρουσιάζονται οπτικά οι αποστάσεις των γραμμάτων-κλάσεις με χρήση των HOG, από όπου μπορούμε να βγάλουμε κάποια πρώιμα συμπεράσματα για το ποιες κλάσεις παρουσιάζουν μεγάλες ομοιότητες και είναι πιθανότερο να γίνει λάθος αναγνώριση, ενώ στον πίνακα του σχήματος 7.5 παρουσιάζονται οπτικά οι αποστάσεις των γράμματα-κλάσεις σε περίπτωση ολικής ανεξαρτησίας στην περιστροφή με την χρήση των RHOG, που εισήχθησαν στο δεύτερο μισό του προηγούμενου κεφαλαίου. Οι συσχετίσεις των κλάσεων σύμφωνα με τους πίνακες συμβαδίζουν σε μεγάλο βαθμό με αυτές που προκύπτουν από την ανθρώπινη αντίληψη, όπως ότι τα γράμματα Ε,Μ,Ν και Σ είναι όμοια σε μεγάλο βαθμό ή ότι τα Ξ και Ψ φαίνονται σχετικά όμοια για περιγραφές ανεξάρτητες στην περιστροφή. Οι πίνακες αυτοί θα φανούν ιδιαίτερα χρήσιμοι σε μια ποιοτική αξιολόγηση των αποτελεσμάτων των τεχνικών στο πειραματικό κομμάτι που ακολουθεί.



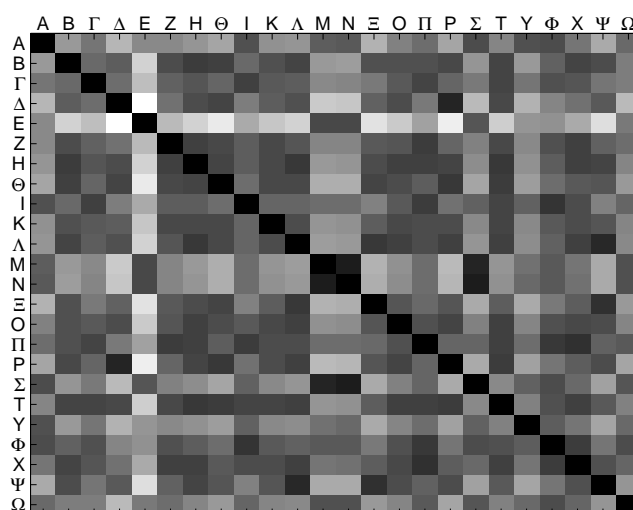
Σχήμα 7.2: Παράδειγμα χειρομορφών της βάσης απο κάθε κλάση



Σχήμα 7.3: Παράδειγμα ποικιλομορφίας στην εμφάνιση (ποσοστό εμφάνισης του χεριού, φωτεινότητα και οπτική γωνία) για το γράμμα Δ



Σχήμα 7.4: Αναπαράσταση πίνακα αποστάσεων χωρίς ολική ανεξαρτησία σε περιστροφή



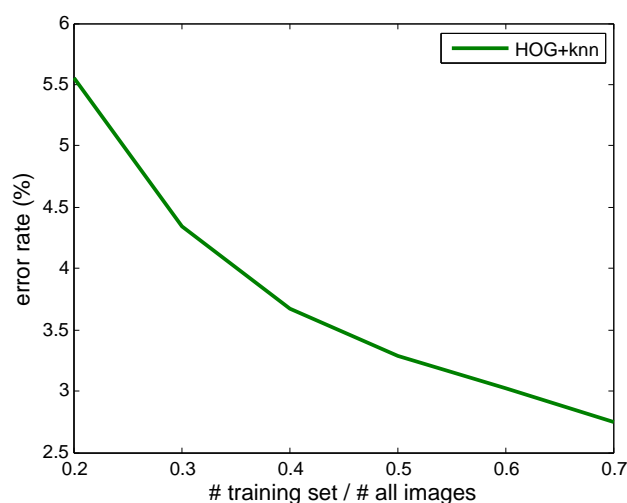
Σχήμα 7.5: Αναπαράσταση πίνακα αποστάσεων με ολική ανεξαρτησία σε περιστροφή

7.2 Επιλογή Συνόλου Εκπαίδευσης

Για την επιλογή του συνόλου εκπαίδευσης (training set) και κατ'επέκταση του συνόλου δοκιμών (testing set) διαμερίζουμε το σύνολο των εικόνων της βάσης δεδομένων είτε τυχαία, είτε με βάση τους νοηματιστές.

Στην περίπτωση της τυχαίας διαμέρισης, ένα τυχαίο σύνολο εικόνων επιλέγεται ως σύνολο εκπαίδευσης και οι υπόλοιπες εικόνες χρησιμοποιούνται ως σύνολο δοκιμών. Για να επιτύχουμε αποτελέσματα ανεξάρτητα κάποιας συγκεκριμένης επιλογής συ-

νόλων, επαναλαμβάνουμε την διαδικασία K φορές και λαμβάνουμε τον μέσο όρο των επιμέρους αποτελεσμάτων. Το ποσοστό των εικόνων της βάσης που χρησιμοποιούνται για εκπαίδευση δίνεται ως παράμετρος και η επίδραση του στην αναγνώριση παρουσιάζεται στο Σχήμα 7.6 για την απλή περίπτωση HOG περιγραφής και knn ταξινομητή. Διαπιστώνουμε ότι, όπως είναι και λογικό, τα ποσοστά λάθους μειώνονται όσο αυξάνεται το σύνολο εκπαίδευσης εις βάρος του συνόλου δοκιμών. Εν τούτοις είναι αξιοσημείωτο ότι όσο και να μειώσουμε το σύνολο εκπαίδευσης, τα ποσοστά αναγνώρισης παραμένουν υψηλά. Για την συνέχεια του πειραματικού μέρους θεωρούμε την παράμετρο αυτή ίση με 0.5, που ισοδυναμεί με ίδιο πλήθος εικόνων εκπαίδευσης και εικόνων δοκιμών.



Σχήμα 7.6: Επίδραση ποσοστού εκπαίδευσης στο ποσοστό σφάλματος αναγνώρισης

Στην δεύτερη περίπτωση χωρίζουμε την βάση έτσι ώστε το σύνολο εκπαίδευσης και το σύνολο δοκιμών να αποτελούνται από διαφορετικούς νοηματιστές. Συγκεκριμένα για κάθε νοηματιστή επιλέγουμε το σύνολο των εικόνων που του αντιστοιχούν ως σύνολο δοκιμών και τις υπόλοιπες εικόνες, δηλαδή αυτές αντιστοιχούν στους υπόλοιπους νοηματιστές, ως σύνολο εκπαίδευσης και υπολογίζουμε το ποσοστό αναγνώρισης για τον συγκεκριμένο διαχωρισμό. Ομοίως με την προηγούμενη περίπτωση, για λόγους ανεξαρτησίας σε «ιδιορρυθμίες» του εκάστοτε διαχωρισμού το τελικό ποσοστό αναγνώρισης υπολογίζεται ως ο μέσος όρος των επιμέρους ποσοστών. Όπως είναι λογικό ο διαχωρισμός σύμφωνα με τους νοηματιστές προσφέρει μια ποιοτικά καλύτερη εκτίμηση της απόδοσης των εξαχθέντων οπτικών χαρακτηριστικών καθώς και την ικανότητα μιας πιο ασφαλούς γενίκευσης σε πραγματικά προβλήματα όπου το σύστημα πιθανότατα δεν θα έχει «ξαναδεί» τον νοηματιστή.

Συνοψίζοντας η πρώτη περίπτωση αποτελεί μια εξιδανίκευση, όπου το σύστημα έχει εκπαιδευτεί με χειρομορφές από όλους τους νοηματιστές, δηλαδή το σύστημα έχει «δει» όλους τους νοηματιστές (seen signers), ενώ στην δεύτερη περίπτωση το σύστημα εκπαιδεύεται μόνο με ορισμένους από τους νοηματιστές, δηλαδή το σύστημα δεν έχει «δει» τους νοηματιστές των δοκιμών (unseen signers).

7.3 Εξαγωγή Χαρακτηριστικών

Στο σημείο αυτό παραθέτουμε τις τεχνικές που θα χρησιμοποιήσουμε για το κομμάτι της εξαγωγής χαρακτηριστικών αλλά και την συντομογραφία τους μαζί με τις βασικές παραμέτρους τους (χωρική διαμέριση, πλήθος bins) για λόγους ευκολίας και ομοιομορφίας για το υπόλοιπο πειραματικό κομμάτι.

hu

Hu Invariant Moments, Κεφάλαιο 2 & Παράρτημα A

fd_{# σημείων περιγράμματος}

Fourier Descriptors, Κεφάλαιο 2

ang_{# σημείων περιγράμματος}

Δειγματοληψία της ακτινικής απόστασης του περιγράμματος από το κέντρο, Κεφάλαιο 2

sc_{# σημείων αναφοράς}

Η προτεινόμενη παραλλαγή των Shape Context για ενιαία περιγραφή, Κεφάλαιο 5

hog_{# bins}_{# διαμερίσεων σε κάθε άξονα}

Η βασική μέθοδος των HOG, Κεφάλαιο 3

chog_{# bins}_{# διαμερίσεων σε κάθε άξονα}

Η παραλλαγή CentHOG που βασίζεται στην εύρεση του κέντρου της χειρομορφής, Κεφάλαιο 5

rhog_{# bins}_{# δακτυλίων}_{# περιγραφών του εσωτερικού δακτυλίου}

Κυκλική Διαμέριση των HOG, Κεφάλαιο 5

erhog_{# bins}_{# ελλειπτικών δακτυλίων}_{# εσωτερικών περιγραφών}

Ελλειψοειδής Διαμέριση των HOG, Κεφάλαιο 5

{bhog|cbhog|rbhog}_{παράμετροι για χωρική διαμέριση}

Παραλλαγή των τεχνικών βασισμένων σε HOG για δυαδικές εικόνες (δεν απαιτείται το πλήθος των bins), Κεφάλαιο 5 & Κεφάλαιο 6

{fd|ang|sc|rhog|erhog|rbhog}&ftm_{παράμετροι της εκάστοτε τεχνικής}

Χρήση Fourier Transform Magnitude (ftm) σε κυκλικές περιγραφές για ανεξαρτησία στην περιστροφή, Κεφάλαιο 2 & Κεφάλαιο 6

{sc|rhog|rbhog}&xcorr_{παράμετροι της εκάστοτε τεχνικής}

Χρήση ετεροσυσχέτισης σε κυκλικές περιγραφές για εύρεση της περιστροφής (μπορεί να χρησιμοποιηθεί και στις υπόλοιπες κυκλικές περιγραφές), Κεφάλαιο 6

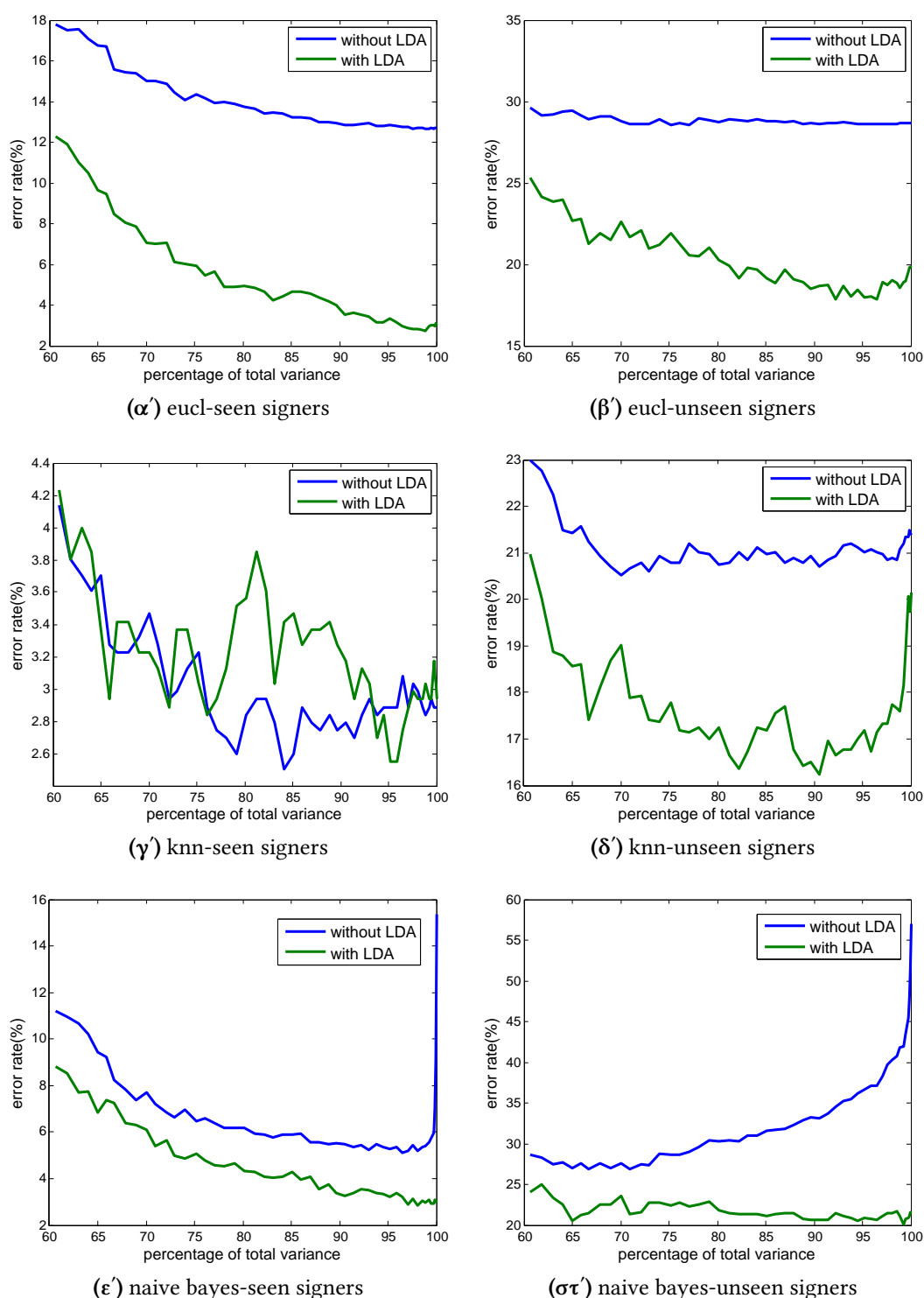
7.4 PCA & LDA

Όπως έχει ήδη αναφερθεί σε προηγούμενο κεφάλαιο, ένα από τα βασικότερα βήματα πριν την εφαρμογή κάποιας τεχνικής αναγνώρισης, είναι η μείωση της διάστασης του διανύσματος χαρακτηριστικών, διατηρώντας όσο το δυνατόν μεγαλύτερο ποσοστό ωφέλιμης πληροφορίας. Για τον σκοπό αυτό χρησιμοποιούμε είτε την ιδιαίτερα δημοφιλή τεχνική PCA είτε την τεχνική LDA.

Σε πολλά προβλήματα η πιο γενική τεχνική PCA φαίνεται να έχει άκρως ικανοποιητικά αποτελέσματα, αφού προβάλλει το αρχικό διάνυσμα στις κατευθύνσεις με την μεγαλύτερη μεταβλητότητα. Βέβαια στην περίπτωση της αναγνώρισης χειρομορφών της ελληνικής αλφαβήτου, όπου υπάρχουν κλάσεις με πολύ μικρές διαφορές (όπως τα γράμματα (M) και (N)) η ιδέα της μεγίστης μεταβλητότητας δεν αντιστοιχεί σε ποιοτικά καλή αναπαράσταση, από άποψη διαχωρισιμότητας, των αρχικών δεδομένων. Αντιθέτως η τεχνική LDA, κάνοντας χρήση της γνώσης των κλάσεων που ανήκει κάθε διάνυσμα (supervised training), μας δίνει την δυνατότητα «ανακάλυψης»-«αξιοποίησης» των χαρακτηριστικών που διαχωρίζουν καλύτερα τις κλάσεις. Στην περίπτωση αυτή όμως είναι δυνατό η μέθοδος να «κολλήσει» και να δώσει βάρος σε αδιάφορη πληροφορία ως προς τον ποιοτικό διαχωρισμό των κλάσεων (π.χ. η συχνή εμφάνιση του καρπού σε μια κλάση σε σχέση με τις υπόλοιπες, μπορεί να οδηγήσει στην αντιστοίχιση της κλάσης αυτής με την εμφάνιση του καρπού). Συνεπώς, σύμφωνα και με τα πειραματικά αποτελέσματα, προτιμάται ένας συνδυασμός των 2 τεχνικών για βέλτιστα αποτελέσματα, έτσι ώστε να διατηρήσουμε ένα σημαντικό μειωμένο πλήθος χαρακτηριστικών μέσω του PCA διώχνοντας έτσι πραγματικά περιττή πληροφορία από το σύνολο των δεδομένων, αλλά αρκετά μεγάλο έτσι ώστε να μπορεί να εφαρμοστεί με επιτυχία η τεχνική LDA στην συνέχεια (που όπως θα επεξηγήσουμε σε επόμενη ενότητα αποτελεί ιδανική επιλογή για το συγκεκριμένο πρόβλημα).

Οι παραπάνω θεωρητικές διαπιστώσεις επιβεβαιώνονται και με τα αντίστοιχα πειράματα, τα αποτελέσματα των οποίων παρουσιάζονται στα σχήματα 7.7, όπου χρησιμοποιείται ο ευκλείδειος ταξινομητής, ο naive bayes ταξινομητής και ο kNN ταξινομητής (περισσότερα για τους ταξινομητές που χρησιμοποιήθηκαν στο πειραματικό μέρος στην επόμενη ενότητα). Συνολικά παρατηρούμε ότι η χρήση των τεχνικών αυτών βελτιώνει σε μεγάλο βαθμό τα αποτελέσματα αναγνώρισης (στην περίπτωση του naive bayes ταξινομητή η επίδραση αυτή φαίνεται πιο έντονα, σχήματα 7.7 (ε),(στ)) και ειδικότερα η χρήση της LDA τεχνικής. Η περίπτωση του απλού κριτηρίου της ευκλείδειας απόστασης από το κέντρο της κλάσης (Σχήμα 7.7 (α),(β)) δίνει και την καλύτερη ποιοτικά εξήγηση για την επίδραση των τεχνικών αυτών, καθώς έχει άμεση και διαισθητική ερμηνεία (κάθε κλάση αντιστοιχεί σε μια υπερσφαίρα) και το μέγιστο ποσοστό αναγνώρισης άρα και ικανότητα διαχωρισιμότητας μεταξύ των κλάσεων ερμηνεύεται ως απομάκρυνση των υπερσφαιρών μεταξύ τους (που είναι και ο στόχος της εφαρμογής της LDA τεχνικής).

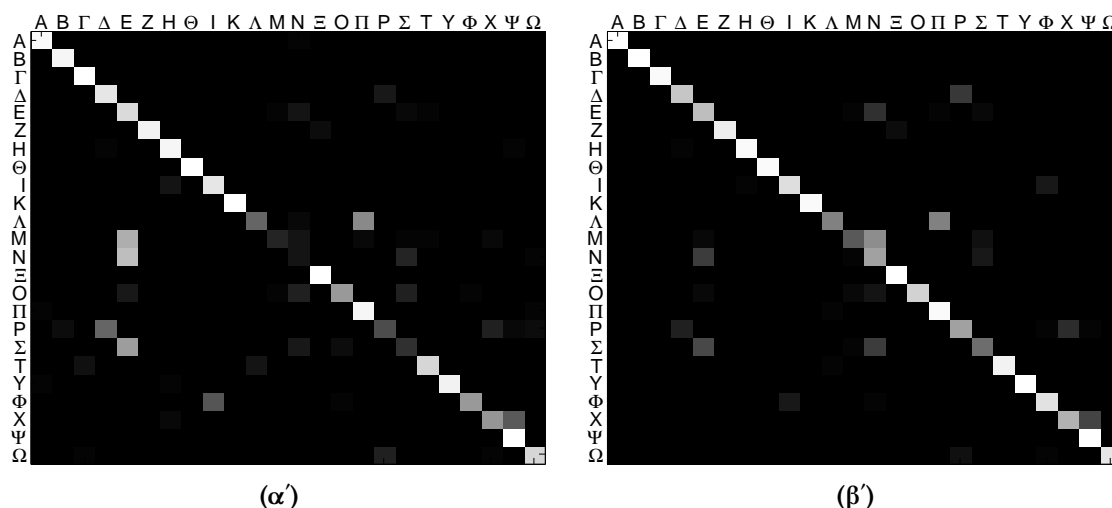
Μια πιο προσεκτική εποπτεία της επίδρασης της τεχνικής LDA μπορεί να πραγματοποιηθεί με σύγκριση των confusion matrices στο Σχήμα 7.8 για τον ταξινομητή



Σχήμα 7.7: Επίδραση των τεχνικών PCA και LDA στο ποσοστό αναγνώρισης σε σχέση με τον ταξινομητή που χρησιμοποιείται.

ευκλείδειας απόστασης (για HOG περιγραφές). Η ισχύς της τεχνικής αυτής είναι ιδιαίτερα εμφανής στα ποσοστά αναγνώρισης των γραμμάτων M και N , τα οποία αυξάνονται σημαντικά με χρήση LDA. Οι χειρομορφές που αντιστοιχούν στα γράμματα αυτά

έχουν πολύ μικρές διαφορές και δεν είναι εφικτό να μοντελοποιηθούν μόνο με PCA (οι διαφορές τους αντιστοιχούν σε χαρακτηριστικά με μικρή μεταβλητότητα). Στον αντίποδα, η τεχνική LDA «εντοπίζει» τα χαρακτηριστικά που διαφοροποιούν τις κλάσεις αυτές και τα αναδεικνύει.

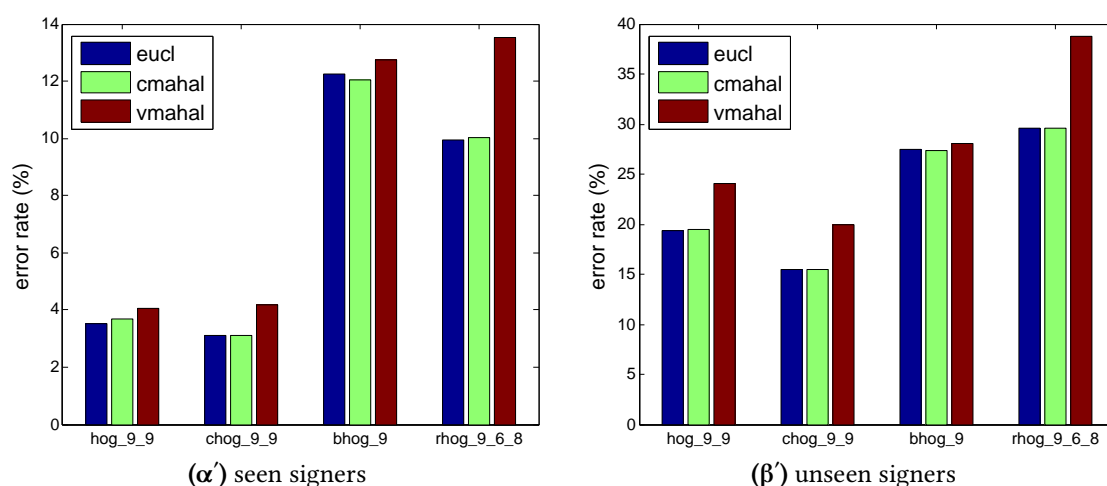


Σχήμα 7.8: confusion matrices : (α) χωρίς την χρήση LDA, (β) με χρήση LDA

7.5 Ταξινομητές

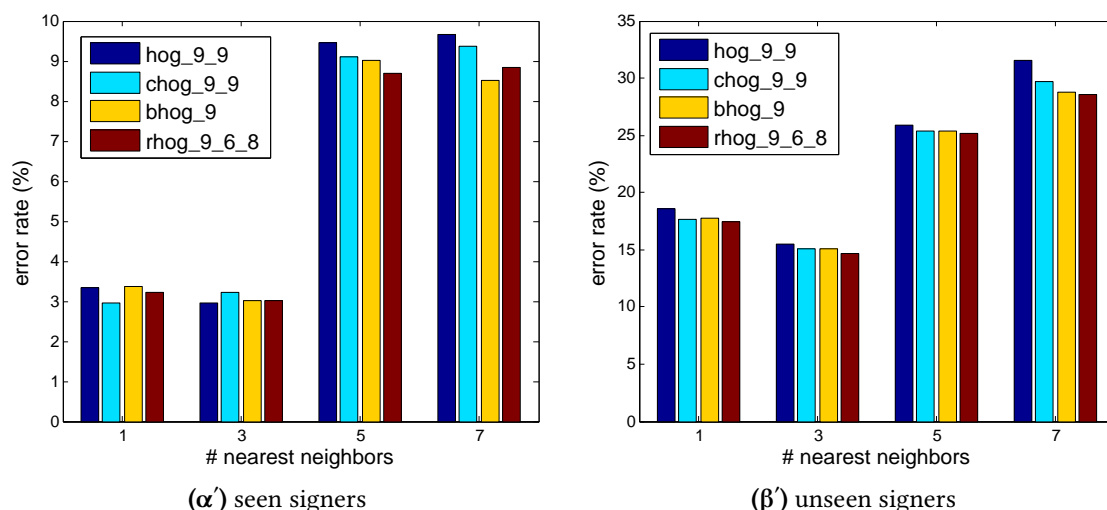
Αρχικά θα συγκρίνουμε την αποτελεσματικότητα των ταξινομητών απόστασης στο πρόβλημα της κατηγοριοποίησης των χειρομορφών για την χρήση της απλής μεθόδου των HOG αλλά και κάποιων βασικών παραλλαγών της. Συγκεκριμένα στο Σχήμα 7.9 συγκρίνουμε τον ευκλείδειο ταξινομητή (eucl), τον mahalanobis ταξινομητή κοινού διαγώνιου πίνακα συνδιασποράς για όλες τις κλάσεις (cmahal) και τον mahalanobis ταξινομητή διαφορετικών διαγώνιων πινάκων συνδιασποράς για κάθε κλάση (vmahal). Η υπόθεση των διαγώνιων πινάκων συνδιασποράς έγινε γιατί θεωρούμε ότι η χρήση PCA και LDA δημιουργεί στατιστικά ασυσχέτιστα χαρακτηριστικά.

Παρατηρούμε ότι η χρήση mahalanobis ταξινομητή απόστασης με κοινό πίνακα συνδιασποράς σε συνδυασμό με τις τεχνικές μείωσης των διατάσεων συνολικά ελαχιστοποιεί το σφάλμα αναγνώρισης, αποτέλεσμα που συνάδει με την θεωρητικό υπόβαθρο της λειτουργίας του LDA (βλ. Κεφάλαιο 4). Επιπλέον το κοινό μητρώο συνδιασποράς μπορεί να υπολογισθεί απευθείας από το μητρώο S_w (within-class scatter matrix) που προκύπτει κατά την εφαρμογή του LDA σύμφωνα με την σχέση $\widetilde{S}_w = W^T S_w W$, όπου W ο προκύπτων μετασχηματισμός. Βέβαια ο ευκλείδειος ταξινομητής έχει ιδιαίτερα ικανοποιητικά αποτελέσματα, σχεδόν το ίδιο ποσοστό σφάλματος με τον mahalanobis κοινού πίνακα συνδιασποράς, και συνυπολογίζοντας την απλότητα του επιλέγεται ως βασικός ταξινομητής απόστασης για την συνέχεια του πειραματικού μέρους.



Σχήμα 7.9: Σφαλμά αναγνώρισης για τους ταξινομητές απόστασης σε 4 ενδεικτικές περιγραφές

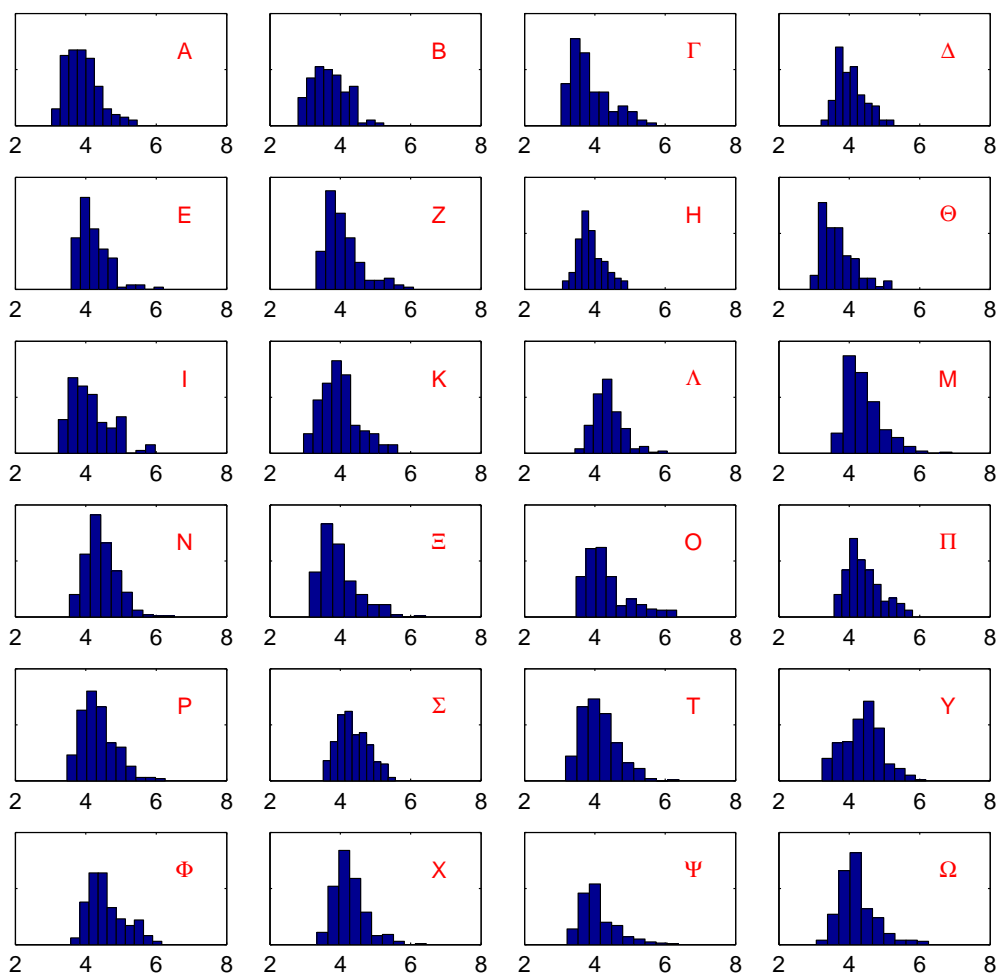
Άλλοι ταξινομητές που θα χρησιμοποιηθούν είναι ο Naive Bayes και ο kNN βασισμένος στην ευκλείδεια απόσταση. Για τον kNN ταξινομητή επιλέγουμε η κατηγοριοποίηση να γίνεται χρησιμοποιώντας τους 7 κοντινότερους γείτονες ($k = 7$), όπως φαίνεται και στο Σχήμα 7.10. Όπως θα δούμε και στην συνέχεια λόγω της φύσης του χώρου των HOG, την οποία θα μελετήσουμε προσεχώς, οι τεχνικές που βασίζονται σε υπόθεση γκαουσιανής κατανομής (ταξινομητές απόστασης, naive bayes) πλησιάζουν σε απόδοση και ανά στιγμές υπερτερούν της kNN, γεγονός που αποδεικνύει ότι είναι δυνατή μια πολύ καλή μοντελοποίηση του χώρου για γρήγορη και εύκολη ταξινόμηση.



Σχήμα 7.10: Σφάλμα αναγνώρισης για τον kNN ταξινομητή σε 4 ενδεικτικές περιγραφές μεταβάλλοντας την τιμή του k

7.6 Ο Χώρος των HOG Περιγραφών

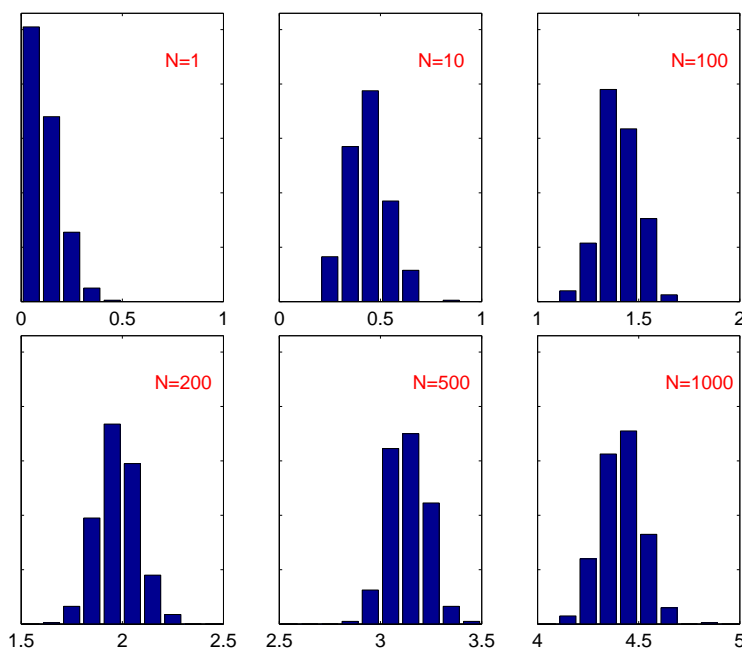
Στην ενότητα αυτή θα προσπαθήσουμε να ερμηνεύσουμε ως ένα βαθμό την συμπεριφορά των HOG περιγραφών, δηλαδή ενός πολυδιάστατου χώρου βασισμένου σε ιστογράμματα, στα πλαίσια του προβλήματος της κατηγοριοποίησης χειρομορφών. Αρχικά θα αποδείξουμε ότι η κάθε κλάση μπορεί να μοντελοποιηθεί ως μια γκαουσιανή κατανομή γραμμικά διαχωρίσιμη από τις υπόλοιπες, υπόθεση που σημαίνει ότι ένας mahalanobis ταξινομητής είναι ικανός για την σωστή ταξινόμηση των χειρομορφών.



Σχήμα 7.11: Ευκλείδεια απόσταση των διανυσμάτων κάθε κλάσης από το κέντρο της. Τα διανύσματα αποτελούν τις HOG περιγραφές των εικόνων της βάσης.

Μια πρώτη εποπτεία της κατανομής των διανυσμάτων χαρακτηριστικών για κάθε κλάση λαμβάνεται με τον υπολογισμό ιστογραμμάτων της απόστασης των διανυσμάτων μιας κλάσης από το κέντρο της. Τα ιστογράμματα αυτά παρουσιάζονται στο Σχήμα 7.11 και παρατηρούμε ότι κάθε κλάση ουσιαστικά περιγράφεται από έναν «φλοιό» μιας υπερσφαίρας (οι διαστάσεις της οποίας είναι ίσες με την διάσταση της περιγραφής), καθώς η ευκλείδεια απόσταση όλων των διανυσμάτων της εκάστοτε κλάσης από το

κέντρο της κυμαίνεται γύρω από κάποια συγκεκριμένη τιμή μεγαλύτερη του μηδενός (στην περίπτωση του σχήματος 7.11, η τιμή αυτή είναι περίπου 4). Η παρατηρηθείσα συμπεριφορά αυτή είναι όμοια με την συμπεριφορά των πολυδιάστατων γκαουσιανών κατανομών, για τις οποίες όσο αυξάνεται το πλήθος των διαστάσεων, αυξάνεται και η απόσταση των σημείων από το κέντρο της, όπως φαίνεται και στο Σχήμα 7.12.



Σχήμα 7.12: Ευκλείδεια απόσταση των σημείων N -διάστατων γκαουσιανών κατανομών από το κέντρο τους. Οι γκαουσιανές του σχήματος έχουν διαγώνιο πίνακα συνδιασποράς (ανεξάρτητα στοιχεία).

Η περιγραφείσα συμπεριφορά έχει την απλή διαισθητική ερμηνεία ότι για πολλές διαστάσεις και θεωρώντας τα στοιχεία των πολυδιάστατων διανυσμάτων ανεξάρτητα, όπου το καθένα περιγράφεται από μια μονοδιάστατη γκαουσιανή, η πιθανότητα όλα τα στοιχεία να βρίσκονται στην μέση τιμή της γκαουσιανής που τα περιγράφει είναι εξαιρετικά μικρή. Συγκεκριμένα λόγω του πλήθους των στοιχείων, πιθανοτικά, ένα σημαντικό πλήθος των στοιχείων θα αποκλίνει από την μέση τιμή των μονοδιάστατων γκαουσιανών που τα περιγράφουν, γεγονός που ισοδυναμεί με απομάκρυνση από το κέντρο της πολυδιάστατης γκαουσιανής. Η παραδοχή της ανεξαρτησίας όλων των σημείων γίνεται για λόγους απλότητας. Η ίδια συμπεριφορά δικαιολογείται και για ανεξαρτησία των στοιχείων κατά ομάδες, περίπτωση που ανταποκρίνεται περισσότερο με τα παραχθέντα ιστογράμματα (συσχετίσεις γειτονικών bins και γειτονικών κελιών).

Βέβαια η ομοιότητα της κατανομής της απόστασης από το κέντρο για κάθε κλάση με αυτή της πολυδιάστατης γκαουσιανής κατανομής αποτελεί μια πρώτη ενθαρρυντική ένδειξη ότι όντως τα δεδομένα μας την προσεγγίζουν, αλλά δεν αποτελεί απόδειξη. Επιπλέον ακόμα και στην περίπτωση που κάθε κλάση προσεγγίζεται από μια γκαουσιανή κατανομή, δεν αρκεί ως πληροφορία για τον σκοπό της ταξινόμησης, καθώς επιθυμούμε επιπρόσθετα οι κλάσεις να είναι γραμμικά διαχωρίσιμες ή αλλιώς οι γκα-

ουσιανές να μην «μπλέκονται».

Μια πιο πρακτική προσέγγιση για την «εξερεύνηση» του χώρου των HOG είναι ο έλεγχος της ικανότητας μοντελοποίησης με γκαουσιανές κατανομές, χρησιμοποιώντας πρώτα LDA, ως προς την επιτυχία αναγνώρισης. Η ικανότητα αυτή μπορεί να αξιολογηθεί ως το ποσοστό της επιτυχούς αναγνώρισης αν εκπαιδεύσουμε το σύστημα μας με όλες τις εικόνες της βάσης και στην συνέχεια τις κατηγοριοποιήσουμε, χρησιμοποιώντας mahalanobis ταξινομητή απόστασης με την υπόθεση ότι κάθε κλάση περιγράφεται από μια πολυδιάστατη γκαουσιανή κατανομή (της όποιος οι παράμετροι εκτιμώνται από το σύνολο εκπαίδευσης με την τεχνική ML κατά τα γνωστά). Τα ποσοστά επιτυχούς αναγνώρισης για 3 ενδεικτικούς περιγραφείς με χρήση του mahalanobis ταξινομητή απόστασης (με διαφορετικό πίνακα συνδιασποράς για κάθε κλάση) παρουσιάζονται στον πίνακα 7.2 και επιβεβαιώνουν τις υποθέσεις για γραμμική διαχωριστικότητα και μοντελοποίηση με γκαουσιανές κατανομές.

hog_9_9	chog_9_9	rhog&ftm_9_6_8
99.8 %	100 %	100 %

Πίνακας 7.2: Ποσοστά επιτυχούς αναγνώρισης για mahalanobis ταξινομητή

Ένα επιπλέον βήμα για την κατανόηση του χώρου μας και της καταλληλότητας των διαθέσιμων τεχνικών αποτελεί η εξερεύνηση της συνθήκης ίδιας συνδιασποράς για όλες τις κλάσεις που συνάδει με την εναλλακτική ερμηνεία του LDA. Εκ πρώτης όψεως κάτι τέτοιο δεν ισχύει, όπως γίνεται αντιληπτό και στην οπτικοποίηση (με χρήση της λογικής των HOG περιγραφών) της διαγωνίου του πίνακα συνδιασποράς για κάθε κλάση στο Σχήμα 7.13. Συγκεκριμένα παρατηρούμε ότι οι ακμές της κάθε κλάσης και κυρίως το περίγραμμα παρουσιάζουν εντονότερες μεταβολές, γεγονός που συντελεί στην ανομοιομορφία των πινάκων συνδιασποράς των κλάσεων. Εν τούτοις η παραπάνω παρατήρηση δεν αποκλείει την χρήση ενός κοινού μητρώου συνδιασποράς που να μπορεί να περιγράψει επαρκώς τις κλάσεις ως προς την ικανότητα διαχωρισμού τους. Την ζητούμενη ιδιότητα την κατέχει ο πίνακας συνδιασποράς S_w του LDA που αποτελεί το μέσο πίνακα συνδιασποράς για όλες τις κλάσεις και η διαγώνιος του έχει την οπτική αναπαράσταση του σχήματος 7.13(α). Τα ποσοστά επιτυχούς αναγνώρισης για 3 ενδεικτικούς περιγραφείς με χρήση του mahalanobis ταξινομητή απόστασης κοινού μητρώου συνδιασποράς και του ευκλείδειου ταξινομητή παρουσιάζονται στους πίνακες 7.3 και 7.4 αντίστοιχα. Η διαφορά μεταξύ των ταξινομητών είναι πρακτικά μηδενική και τα αποτελέσματα αναγνώρισης, στο σύνολο τους, πολύ υψηλά. Επιπλέον η επιτυχία του ευκλείδειου ταξινομητή αναδεικνύει την μορφή των κλάσεων μετά τους αρχικούς μετασχηματισμούς (PCA & LDA) ως περιοχές που περικλείονται σε υπερσφαίρες ίδιας ακτίνας. Το γεγονός αυτό ενισχύει την επιλογή μας του ευκλείδειου ταξινομητή ως βασικού ταξινομητή απόστασης.

Στο σημείο αυτό θα περιγράψουμε μια θεωρητική προσέγγιση για την περιγραφή των *rhog&ftm* ως γκαουσιανές κατανομές, εφόσον τα *hog* περιγράφονται με γκαουσιανές. Αρχικά θεωρούμε ότι τα στοιχεία ενός διανύσματος περιγραφής των HOG είναι στατιστικώς συσχετιζόμενα μόνο σε ομάδες που αποτελούνται από γειτονικά bins (κυ-

hog_9_9	chog_9_9	rhog&ftm_9_6_8
99.2 %	99.3 %	97.9 %

Πίνακας 7.3: Ποσοστά επιτυχούς αναγνώρισης για *mahalanobis* ταξινομητή κοινού πίνακα συνδιασποράς

hog_9_9	chog_9_9	rhog&ftm_9_6_8
99.2 %	99.4 %	97.7 %

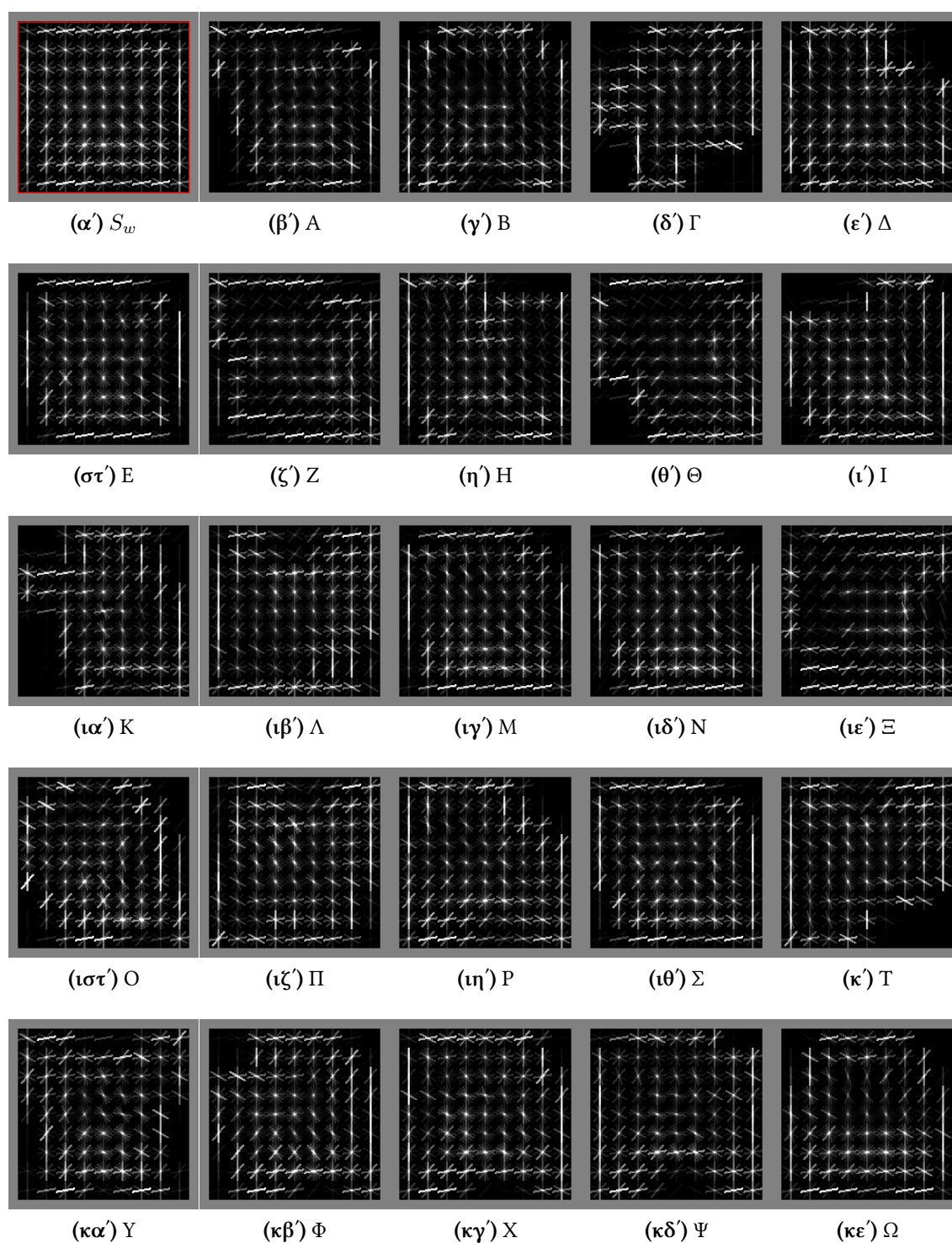
Πίνακας 7.4: Ποσοστά επιτυχούς αναγνώρισης για *ευκλείδειο* ταξινομητή

κλική γειτνίαση - το πρώτο και το τελευταίο bin είναι γειτονικά) και σε ορισμένες περιπτώσεις από bins γειτονικών κελιών. Στην παραδοχή αυτή αναφερθήκαμε και προηγουμένως, καθώς δικαιολογεί την απόσταση των διανυσμάτων απο το κέντρο της κλάσης τους αλλά διευκολύνει και κάποιες θεωρητικές υποθέσεις. Ιδιαίτερα για την περίπτωση των κυκλικών HOG μαζί με Fourier Transform Magnitude, για διευκόλυνση, αποδεχόμαστε την παραδοχή ότι τα ιστογράμματα κάθε κυκλικού δακτυλίου είναι στατιστικώς ανεξάρτητα από αυτά των γειτονικών δακτυλίων. Συνεπώς κάθε δακτύλιος, που αποτελείται από ένα σύνολο κελιών - ιστογραμμάτων, σύμφωνα με τα προηγούμενα μπορεί να περιγραφεί με μια γκαουσιανή άρα και ο μετασχηματισμός Fourier του μπορεί να περιγράφει με γκαουσιανή. Οι παραπάνω παρατηρήσεις και παραδοχές μπορούν να ερμηνεύσουν το τελικό διάνυσμα ως συνένωση στατιστικά ανεξάρτητων ομάδων (δακτυλίων) που περιγράφονται από γκαουσιανές κατανομές και συνεπώς και η συνένωση τους περιγράφεται από μια γκαουσιανή κατανομή.

Εν κατακλείδι συμπεραίνουμε ότι η τεχνική LDA και οι ταξινομητές απόστασης, που βασίζονται σε γκαουσιανές κατανομές, συνάδουν απόλυτα με τον χώρο των ιστογραμμάτων που εισάγουν τα HOG και οι παραλλαγές τους.

7.7 Επιλογή Βέλτιστων Παραμέτρων

Όπως έχουμε περιγράψει στα κεφάλαια όπου παρουσιάζουμε τις διάφορες τεχνικές, υπάρχει πληθώρα παραμέτρων που επηρεάζουν σε μεγάλο βαθμό την συμπεριφορά των περιγραφών και την ικανότητα διάκρισης που μας προσφέρουν. Στην παρούσα ενότητα λοιπόν, με στόχο την εύρεση των βέλτιστων παραμέτρων, παρουσιάζουμε την εξάρτηση της απόδοσης των περιγραφών από τις εκάστοτε παραμέτρους για το πρόβλημα της κατηγοριοποίησης των χειρομορφών. Κυρίως θα μελετήσουμε την επίδραση των παραμέτρων των HOG περιγραφών και τον παραλλαγών τους, δίνοντας ιδιαίτερη έμφαση στην μελέτη των παραμέτρων που σχετίζονται με την χωρική διαμέριση των περιγραφών. Επιπλέον θα γίνει αναφορά και στις παραμέτρους της προτεινόμενης παραλλαγής των Shape Context.



Σχήμα 7.13: Γραφική αναπαράσταση της διαγωνίου του πίνακα συνδιασποράς για κάθε κλάση. Στο Σχήμα (α) φαίνεται η μέση συνδιασπορά για όλες της κλάσεις, όπως προκύπτει και από την τεχνική LDA

7.7.1 HOG

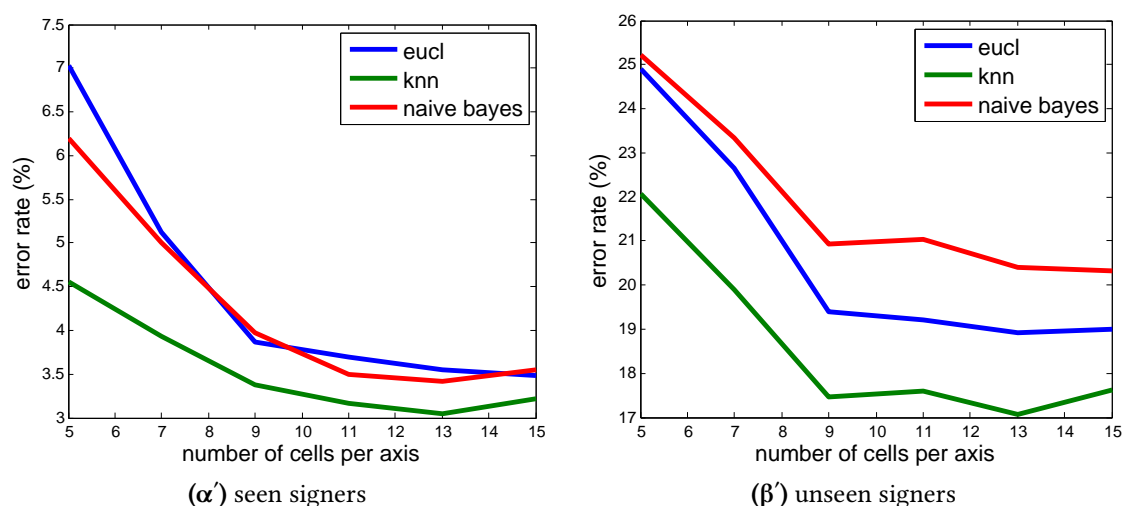
Αρχικά μελετάμε την επίδραση των παραμέτρων της απλής μεθόδου των HOG στο πρόβλημα της αναγνώρισης χειρομορφών. Συγκεκριμένα μεταβάλλουμε τον αριθμό των κελιών της εικόνας, τον αριθμό των bins σε κάθε τοπικό ιστόγραμμα, την επιλογή για signed ($[0, 360^\circ]$) και unsigned ($[0, 180^\circ]$) κατευθύνσεις, την επιλογή του ποσοστού επικάλυψης των κελιών και τέλος την επιλογή του τελεστή πρώτης παραγώγου για την εικόνα. Θα μεταβάλλουμε λοιπόν διαδοχικά μια μια τις παραμέτρους διατηρώντας τις υπόλοιπες σταθερές και ίσες με τις αρχικές τιμές τους. Οι αρχικές τιμές των παραμέτρων, που έχουν επιλεγεί εμπειρικά έτσι ώστε να έχουν καλή απόδοση στην αναγνώριση, είναι:

cells	bins	angle range	overlap	grad operator
9	9	<i>unsigned</i>	0.5	<i>central difference</i>

Πίνακας 7.5: Αρχικές επιλογές παραμέτρων για τα HOG

Πλήθος Κελιών

Η πιο βασική παράμετρος για την χωρική καταμέριση της πληροφορίας είναι το πλήθος των κελιών που θα χρησιμοποιηθούν για να περιγράψουν την εικόνα και ουσιαστικά καθορίζει και το εύρος της τοπικής πληροφορίας. Θεωρούμε ότι συνολικά, για όλες τις εικόνες, δεν υπάρχει χωρικά πυκνότερη κατανομή ως προς κάποιο άξονα από τους δυο ώστε να αυξήσουμε τον αριθμό κελιών σε αυτόν για να αποκτήσουμε μια ποιοτικότερη περιγραφή, οπότε υιοθετούμε τετραγωνικό πλέγμα, δηλαδή ίδιο αριθμό κελιών σε κάθε άξονα.



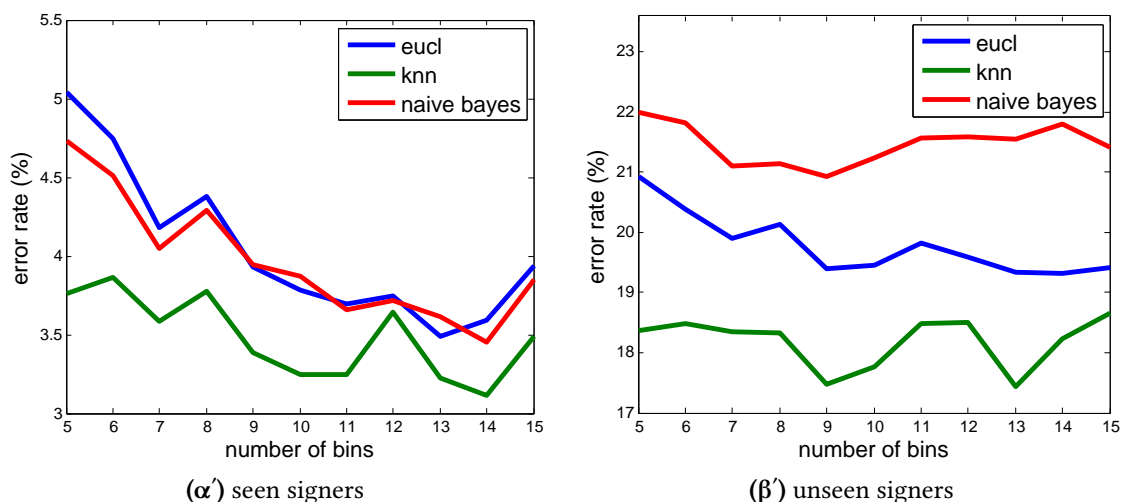
Σχήμα 7.14: Επίδραση του πλήθους των κελιών, μεταβάλλοντας τον αριθμό των διαμερίσεων, ανά άξονα για hog περιγραφές

Τα ποσοστά σφάλματος στην αναγνώριση παρουσιάζονται στο Σχήμα 7.14, από όπου συμπεράνουμε ότι μια καλή επιλογή είναι 9 κελία ανά άξονα (τα διαγράμματα

παρουσιάζουν «γόνατο» για την τιμή αυτή), δηλαδή 81 συνολικά. Λίγο καλύτερα αποτελέσματα εμφανίζονται για την επιλογή 13 κελιών ανά άξονα, δηλαδή 169 συνολικά, αλλά η επιλογή αυτή συνεπάγεται αξιοσημείωτη αύξηση στο διάνυσμα περιγραφής σε σχέση με την βελτίωση στην αναγνώριση.

Διακριτοποίηση Κατεύθυνσης (Πλήθος Bins)

Η επιλογή του πλήθους των bins σχετίζεται άμεσα με την ποιότητα της τοπικής περιγραφής κάθε κελιού και θεωρητικά αύξηση του αριθμού των κελιών ισοδυναμεί με πιο λεπτομερή πληροφορία για την κάθε περιοχή ή αλλιώς μεγαλύτερη ακρίβεια στην κατεύθυνση της παραγώγου (λεπτομερέστερη διακριτοποίηση των κατευθύνσεων). Βέβαια λεπτομερέστερη διακριτοποίηση ισοδυναμεί με μεγαλύτερα τοπικά ιστογράμματα, αρά και μεγαλύτερο τελικό διάνυσμα περιγραφής. Δίνοντας μεγαλύτερο βάρος στην περίπτωση της εξαρτώμενης από τον νοηματιστή εκπαίδευσης, σύμφωνα με το Σχήμα 7.15(β), καταλήγουμε ότι η χρήση 9 bins μας προσφέρει το μικρότερο ποσοστό σφάλματος.

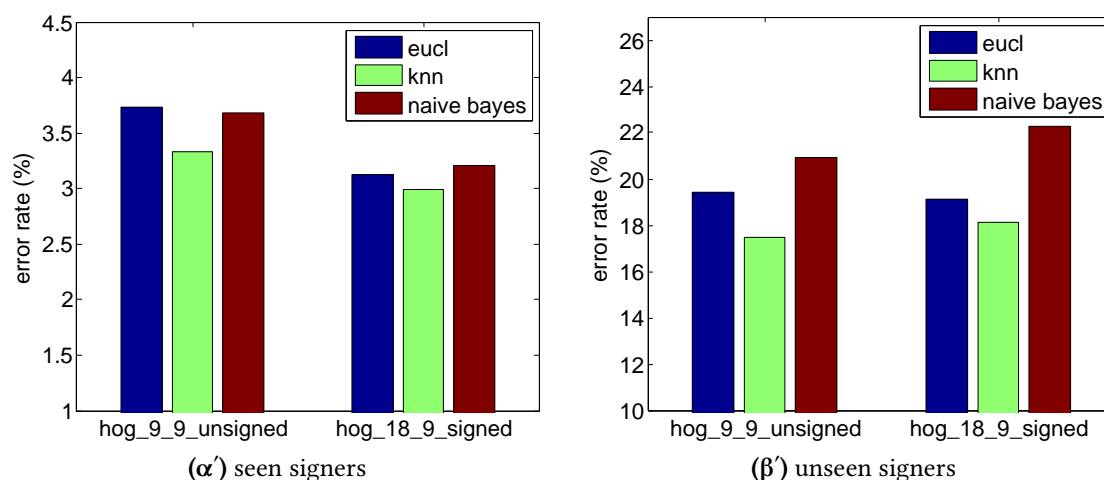


Σχήμα 7.15: Επίδραση του πλήθους των bins για hog περιγραφές

Προσημασμένη ή Μη Προσημασμένη Παράγωγος

Ένα άλλο βασικό χαρακτηριστικό των HOG είναι η επιλογή του εύρους γωνιών στις κατευθύνσεις, που είναι είτε $[0, 360^\circ]$, δηλαδή έχουμε προσημασμένη (signed) παράγωγο (κατεύθυνση), είτε $[0, 180^\circ]$, δηλαδή μη προσημασμένη (unsigned) παράγωγο (ουσιαστικά διεύθυνση). Για να επιτύχουμε ίδια διακριτική ικανότητα του ιστογράμματος στην περίπτωση προσημασμένων κατευθύνσεων πρέπει να διπλασιάσουμε τον αριθμό των bins, αφού διπλασιάζεται και το εύρος των γωνιών, ενώ τα ποσοστά αναγνώρισης των χειρομορφών είναι σχεδόν ίδια σύμφωνα με το Σχήμα 7.16. Συνεπώς, συνυπολογίζοντας και το γεγονός ότι για την επιλογή προσημασμένων κατευθύνσεων διπλασιάζουμε το μέγεθος του διανύσματος χαρακτηριστικών (για ίδιο εύρος γωνιών σε κάθε bin), συμπεραίνουμε ότι η επιλογή μη προσημασμένων κατευθύνσεων αποτελεί ποιοτικότερη λύση τόσο από άποψη ποσοστών αναγνώρισης όσο και υπολογιστικής διευκόλυνσης και για τον λόγο αυτό την θεωρούμε ως δεδομένη και στις παραλλαγές

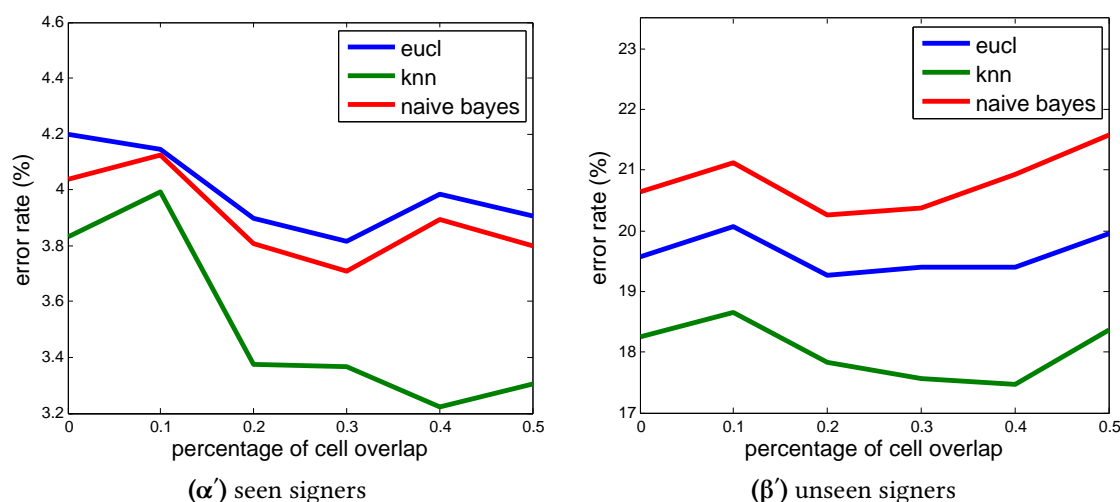
των HOG που ακολουθούν.



Σχήμα 7.16: Επίδραση του εύρους γωνιών (unsigned gradient vs signed gradient) για hog περιγραφές

Ποσοστό Επικάλυψης Γειτονικών Κελιών

Η αλληλοεπικάλυψη κελιών θα βοηθήσει στην δημιουργία περιγραφών λιγότερο ευαίσθητων σε μικρές εσωτερικές αλλαγές της θέσης των ακμών, γεγονός που επαληθεύεται και από το ποσοστό σφάλματος στην αναγνώριση στο Σχήμα 7.17. Σύμφωνα με τα διαγράμματα, μια καλή επιλογή αποτελεί η 20% επικάλυψη μεταξύ των κελιών.

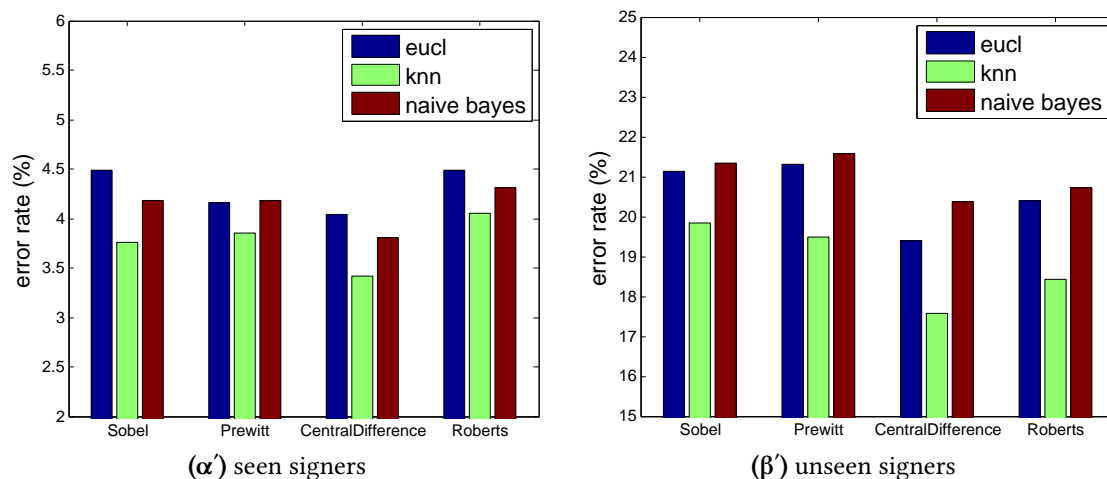


Σχήμα 7.17: Επίδραση του ποσοστού επικάλυψης των κελιών για hog περιγραφές

Επιλογή Τελεστή Παραγωγού

Όπως έχουμε αναφέρει στο Κεφάλαιο 3, ο υπολογισμός των κατευθυντικών παραγώγων της εικόνας g_x και g_y μπορεί να γίνει με φιλτράρισμα της εικόνας με κάποιον

από τους εξής τελεστές : Sobel, Prewitt, Central Difference και Roberts. Σύμφωνα με το Σχήμα 7.18 η πιο αξιόπιστη επιλογή αποτελεί η χρήση του τελεστή Central Difference, δηλαδή φιλτράρισμα της εικόνας με $[-1, 0, 1]$ και $[-1, 0, 1]^T$. Η επιλογή αυτή θα θεωρηθεί ως δεδομένη και σε επόμενες παραλλαγές των HOG.



Σχήμα 7.18: Επίδραση του τελεστή παραγώγου για hog περιγραφές

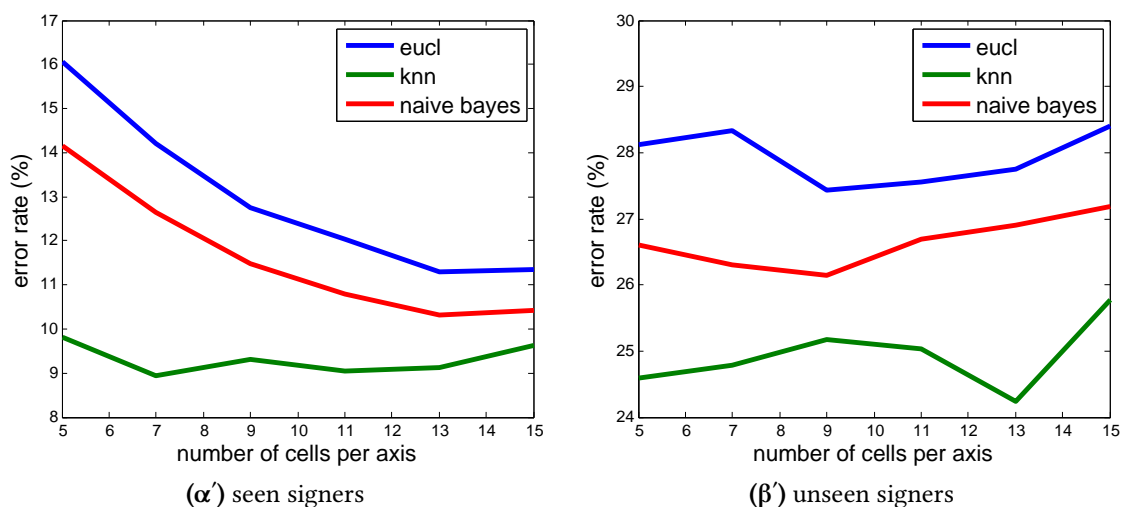
Συμπερασματικά συνοψίζουμε τις τελικές επιλογές των παραμέτρων στον πίνακα 7.6. Πολλές από αυτές τις επιλογές θα τις θεωρήσουμε σταθερές για την παραμετροποίηση των παραλλαγών των HOG (signed gradient, central difference) και δεν θα αναφερθούμε σε αυτές στην συνέχεια.

cells	bins	angle range	overlap	grad operator
9	9	<i>unsigned</i>	0.2	<i>central difference</i>

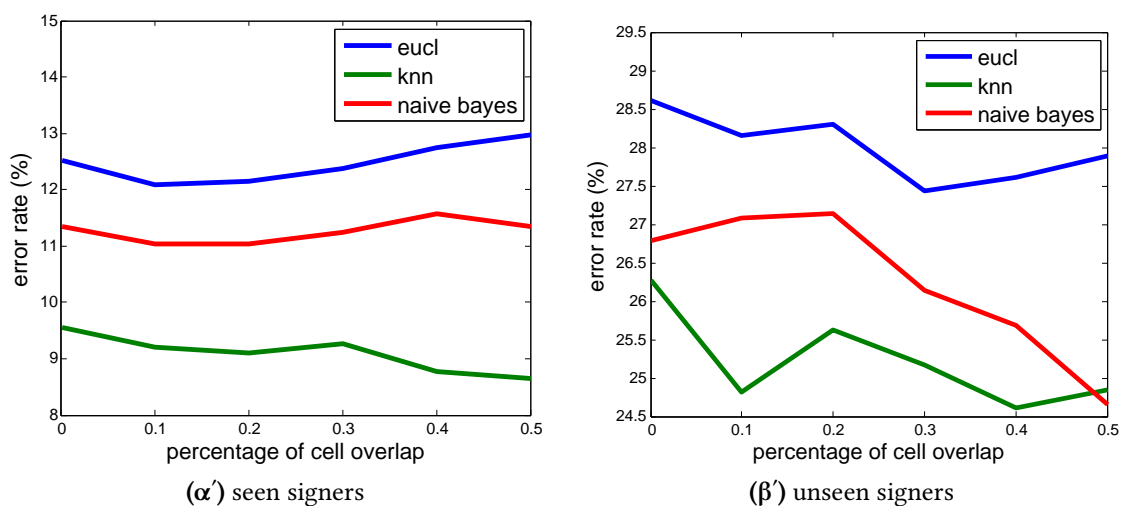
Πίνακας 7.6: Επιλογές βέλτιστων παραμέτρων για τα HOG

7.7.2 BHOG

Για την περίπτωση της παραλλαγής για δυαδικές εικόνες περιγράμματος οι παράμετροι περιορίζονται στην χωρική διαμέριση της εικόνας, καθώς κατά την ανάπτυξη της μεθόδου έχουμε θεωρήσει εύρος γωνιών $[0, 180^\circ]$ (unsigned gradient) και central differences ως τελεστή για την εύρεση της παραγώγου. Στα σχήματα 7.19 και 7.20 παρουσιάζεται η εξάρτηση της απόδοσης των Binary HOG από το πλήθος των κελιών ανά άξονα και το ποσοστό επικάλυψης των κελιών αντίστοιχα. Σύμφωνα με τα διαγράμματα η τελική επιλογή των βέλτιστων παραμέτρων είναι 9 κελία ανά άξονα και 30% επικάλυψη των κελιών μεταξύ τους. Επιπλέον, όπως είναι λογικό, το σφάλμα αναγνώρισης είναι σαφώς μεγαλύτερο από την περίπτωση των HOG καθώς πλέον η μόνη διαθέσιμη πληροφορία είναι το περίγραμμα της χειρομορφής.



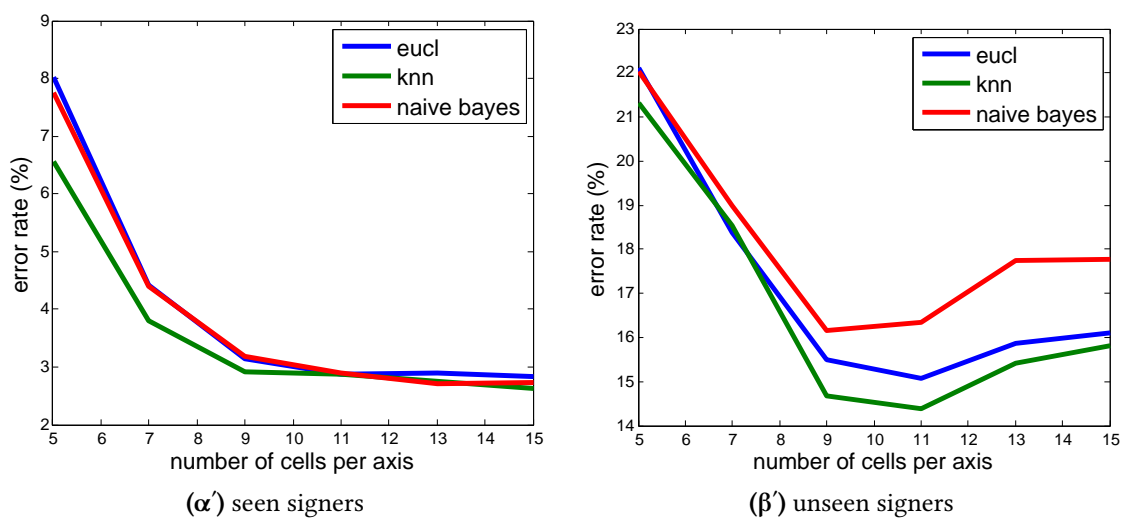
Σχήμα 7.19: Επίδραση του πλήθους των κελιών, μεταβάλλοντας τον αριθμό των διαμερίσεων ανά άξονα, για bhog περιγραφές



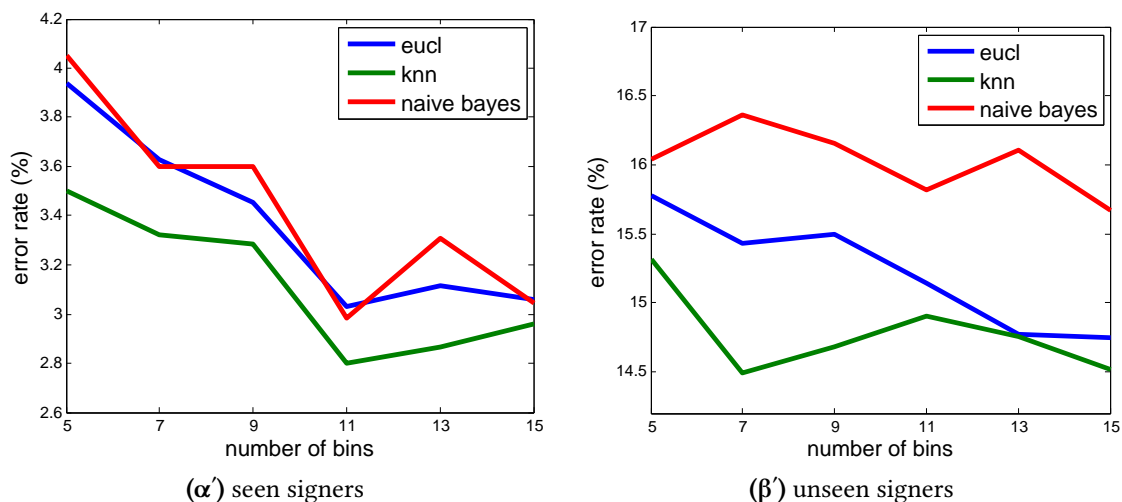
Σχήμα 7.20: Επίδραση του ποσοστού επικάλυψης των κελιών για bhog περιγραφές

7.7.3 CHOG & CBHOG

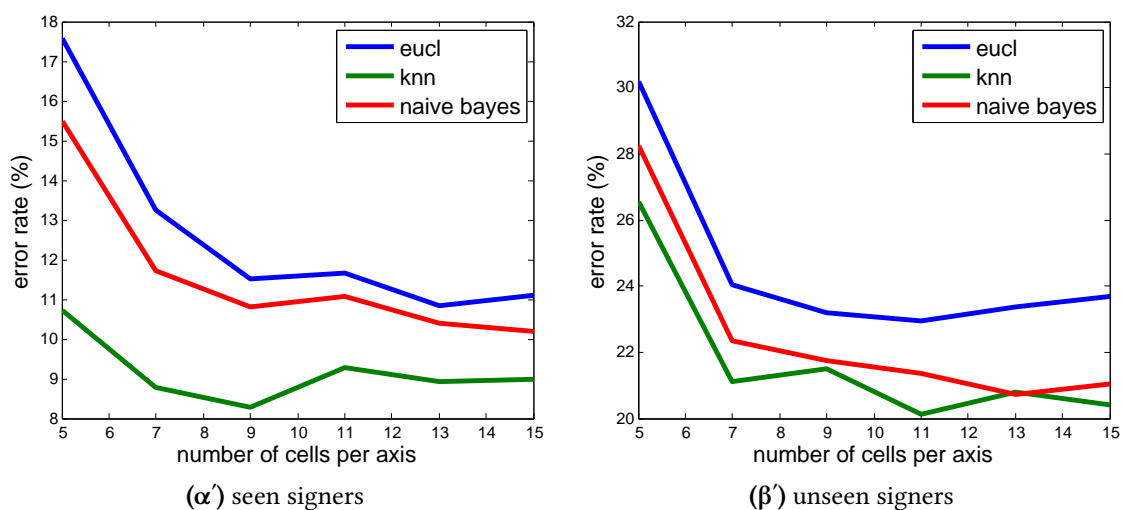
Για την προτεινόμενη μέθοδο των Centralized HOG μεταβάλλουμε τις παραμέτρους της χωρικής διαμέρισης και της διακριτοποίησης της κατεύθυνσης (αριθμός bins). Η επίδραση του πλήθους των κελιών και του πλήθους των bins στο ποσοστό αναγνώρισης παρουσιάζεται στα σχήματα 7.21 και 7.22 αντίστοιχα. Σύμφωνα με τα διαγράμματα η τελική επιλογή των βέλτιστων παραμέτρων είναι, ομοίως με προηγούμενες επιλογές, 9 κελία ανά άξονα και 11 bins. Επιπλέον στο Σχήμα 7.23 παρουσιάζεται η επίδραση του πλήθους των κελιών στην αποτελεσματικότητα της παραλλαγή των CentHOG για δυαδικές εικόνες, από όπου συμπεραίνουμε ότι η επιλογή των 9 κελίων ανά άξονα εξακολουθεί να είναι η επικρατέστερη.



Σχήμα 7.21: Επίδραση του πλήθους των κελιών για χογ περιγραφές



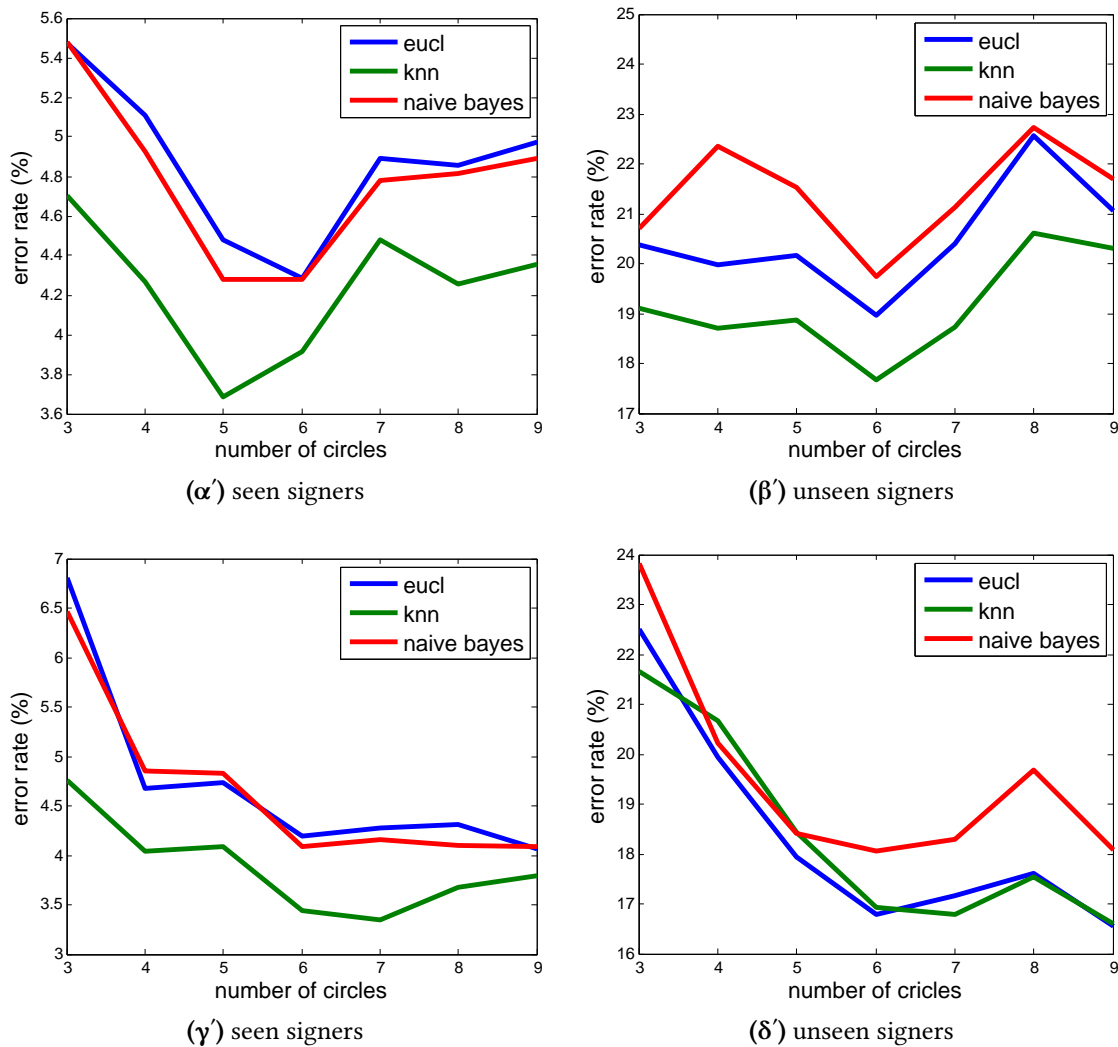
Σχήμα 7.22: Επίδραση του πλήθους των bins για χογ περιγραφές



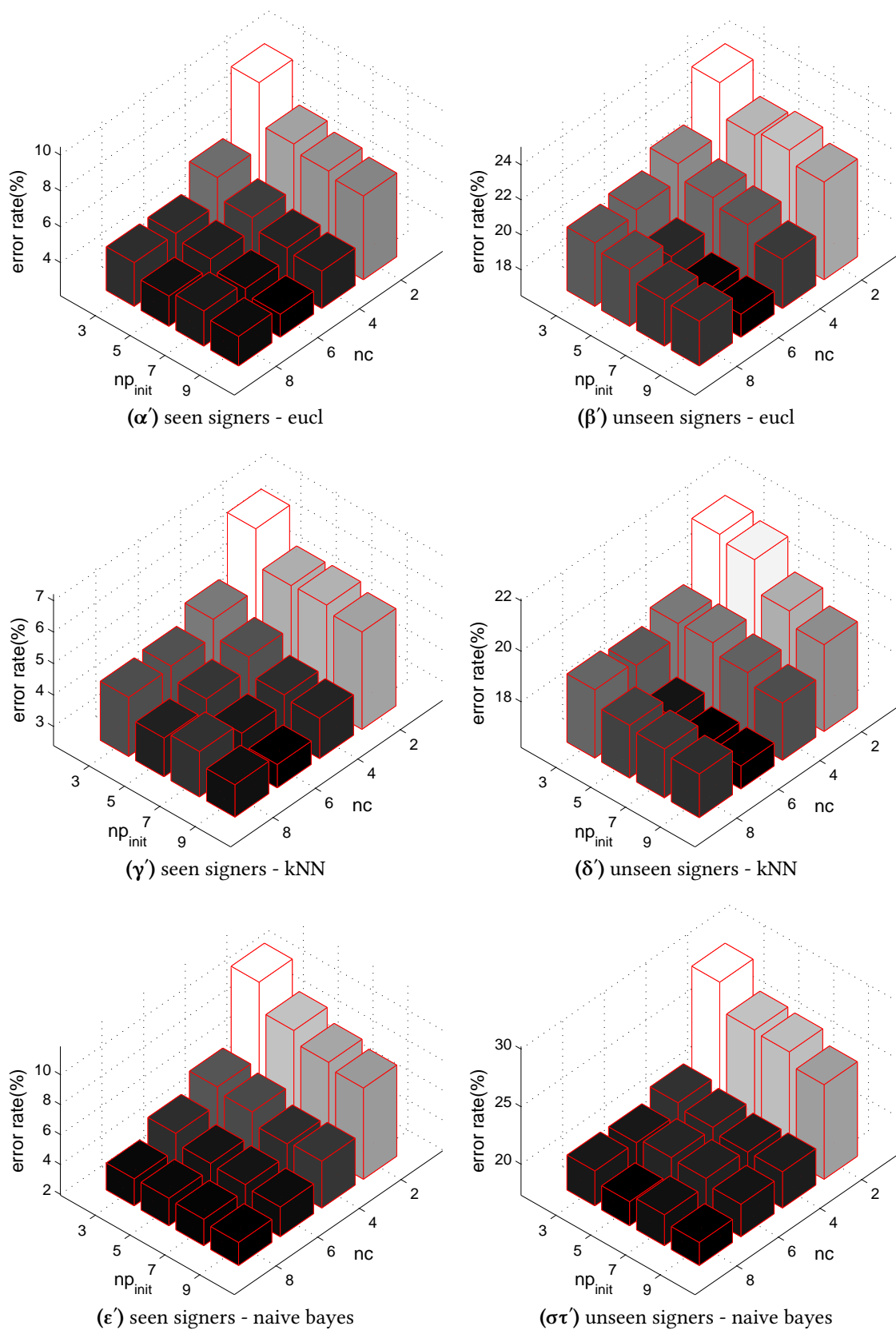
Σχήμα 7.23: Επίδραση του πλήθους των κελιών για cηhog περιγραφές

7.7.4 RHOG , ERHOG , RBHOG & ERBHOG

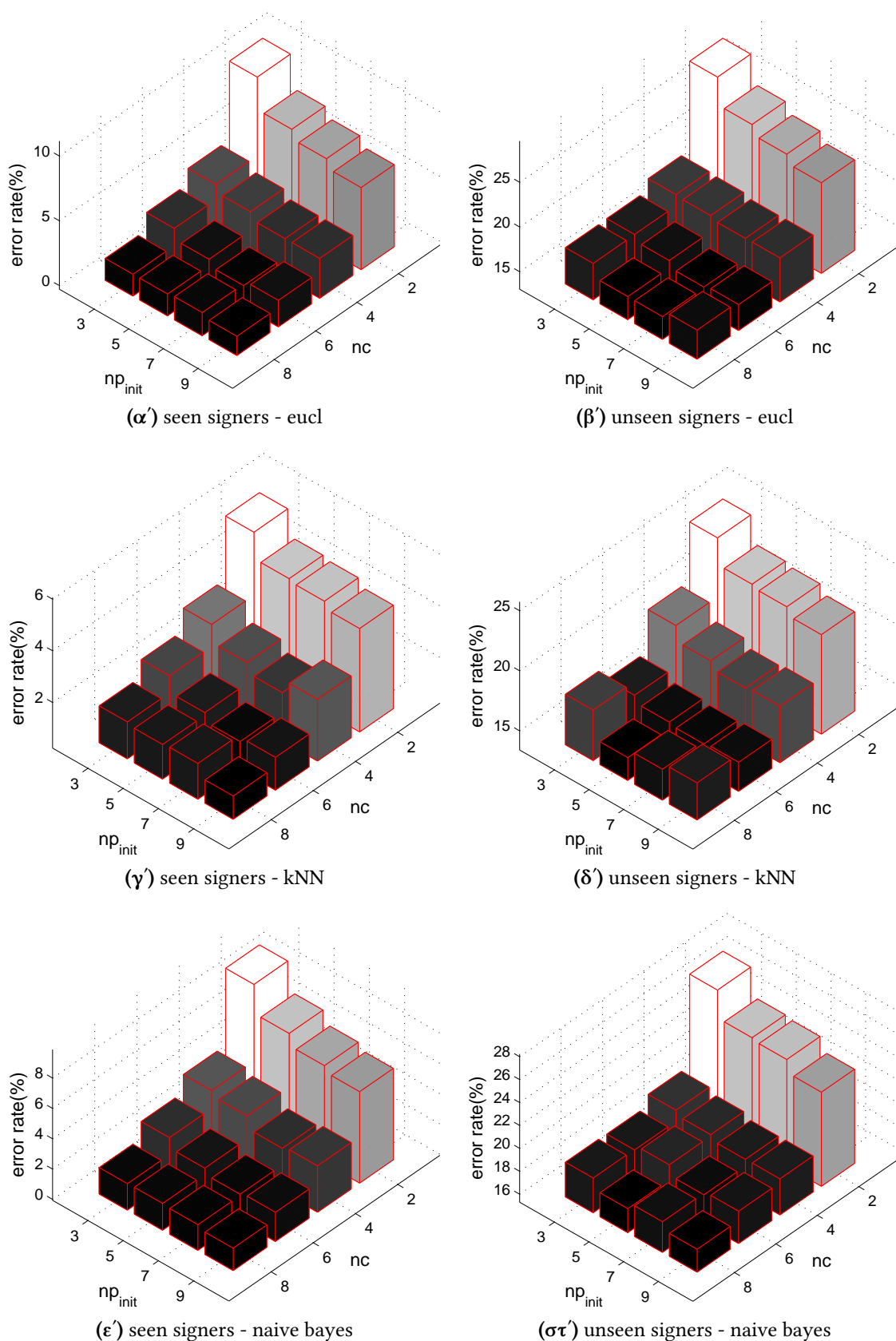
Οι παραλλαγές που βασίζονται σε κυκλική κατανομή των τοπικών περιγραφών εφαρμόζουν μια διαφορετική λογική χωρικής διαμέρισης. Συγκεκριμένα την κυκλική (είτε ελλειπτική) διαμέριση την καθορίζουν 2 παράμετροι, ο αριθμός των κύκλων (nc) και ο αριθμός των σημείων που αντιστοιχούν σε τοπικές περιγραφές στον εσωτερικό κύκλο (np_{init}). Θεωρώντας λοιπόν όλες τις άλλες παραμέτρους σταθερές (9 bins, unsigned gradient, central difference) μελετάμε την επίδραση των χωρικών παραμέτρων είτε διατηρώντας το τελικό διάνυμα σχετικά σταθερό (120 τοπικές περιγραφές) στα σχήματα 7.24, είτε αναλυτικότερα μεταβάλλοντας ανεξάρτητα τις 2 παραμέτρους στα σχήματα 7.25, 7.26 και τέλος για δυαδικές εικόνες στα Σχήματα 7.27, 7.28. Ως προτεινόμενοι περιγραφείς επιλέγουμε τους $rhog_9_6_7$, $erhog_9_6_7$, $rbhog_4_8$ και $erbhog_4_8$.



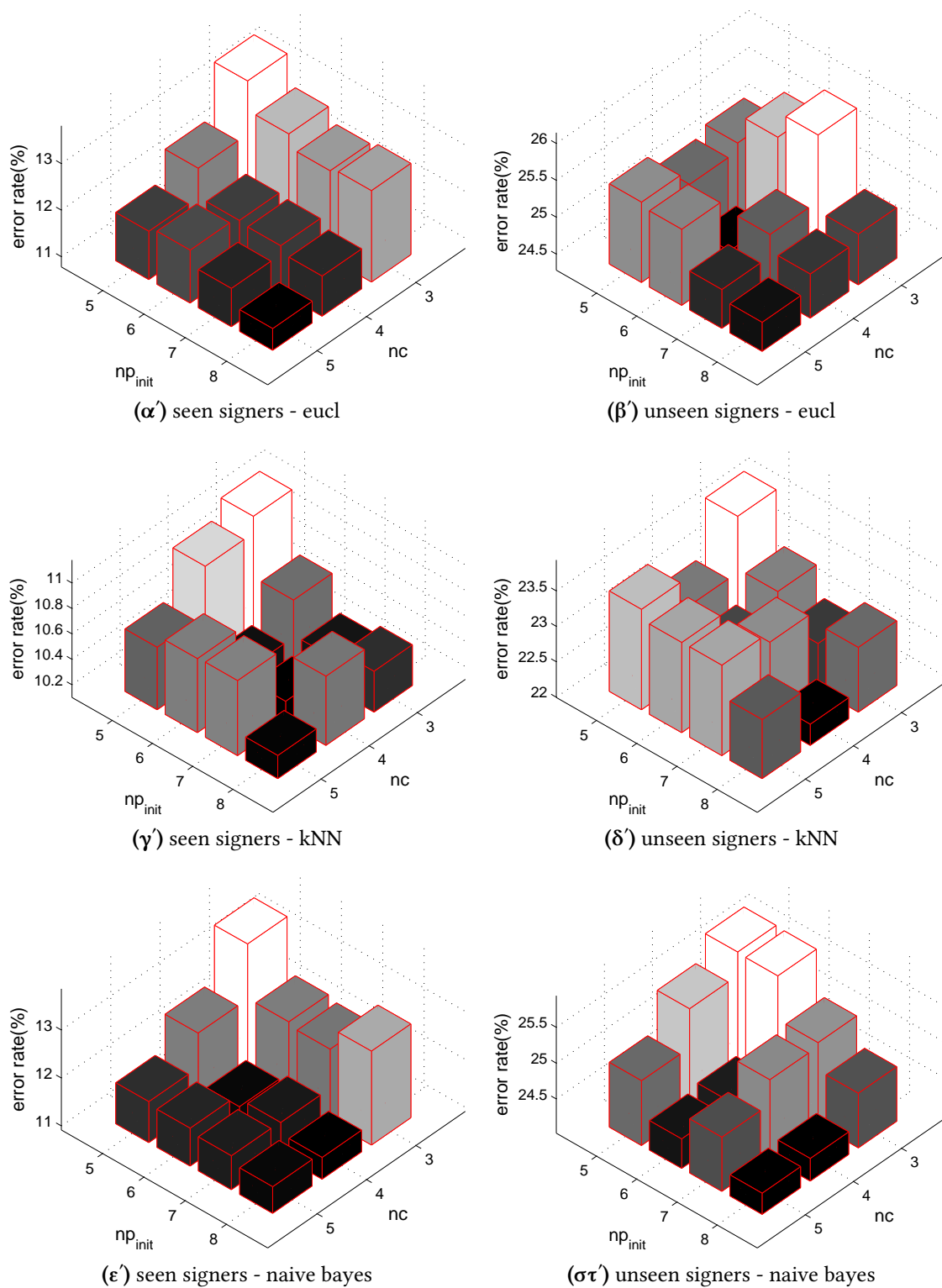
Σχήμα 7.24: Επίδραση του πλήθους των κύκλων αναφοράς διατηρώντας το πλήθος των συνολικών κελιών σταθερό, όπου (α),(β): rhog και (γ),(δ): erhog



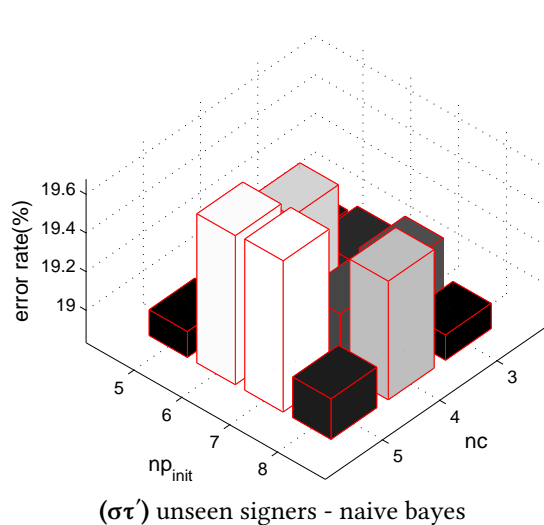
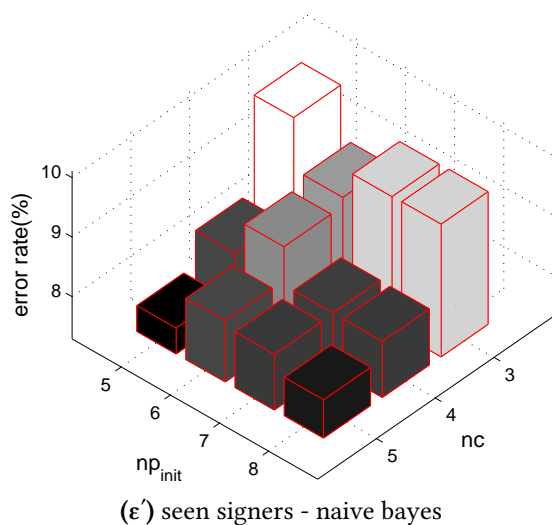
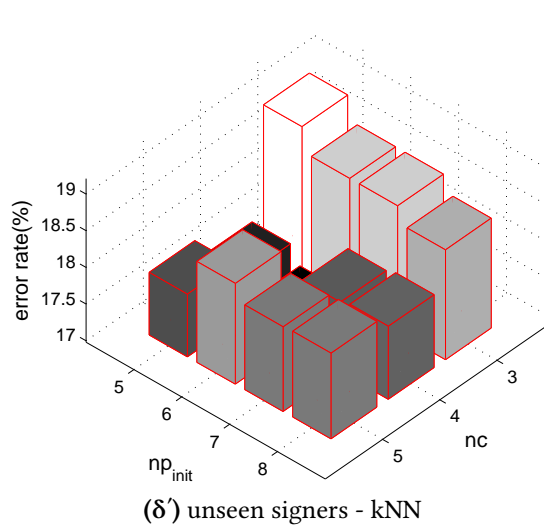
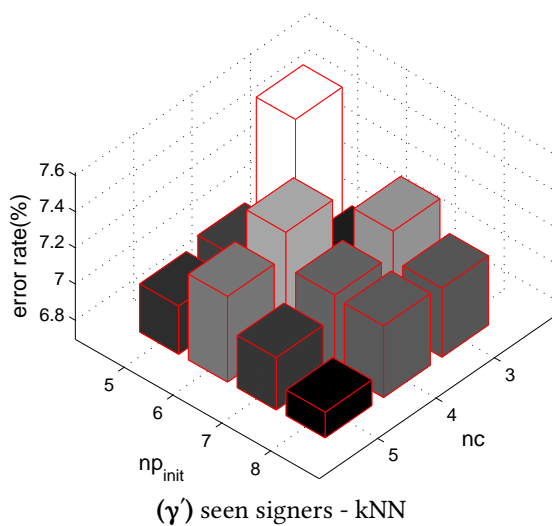
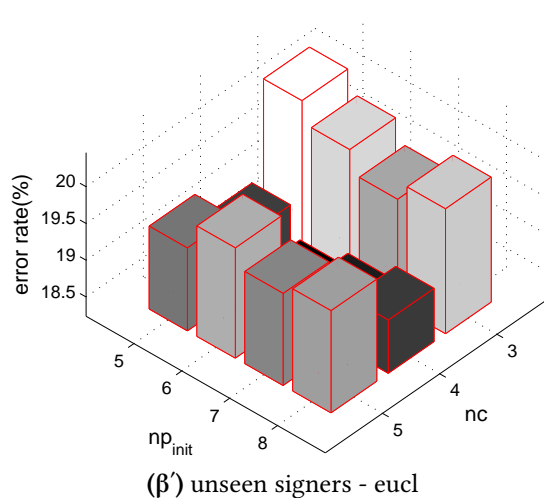
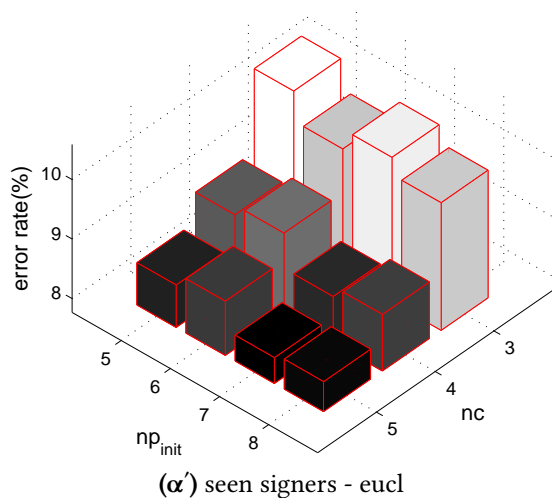
Σχήμα 7.25: Αναλυτική επίδραση των χωρικών παραμέτρων για τησh περιγραφές



Σχήμα 7.26: Αναλυτική επίδραση των χωρικών παραμέτρων για ετηος περιγραφές



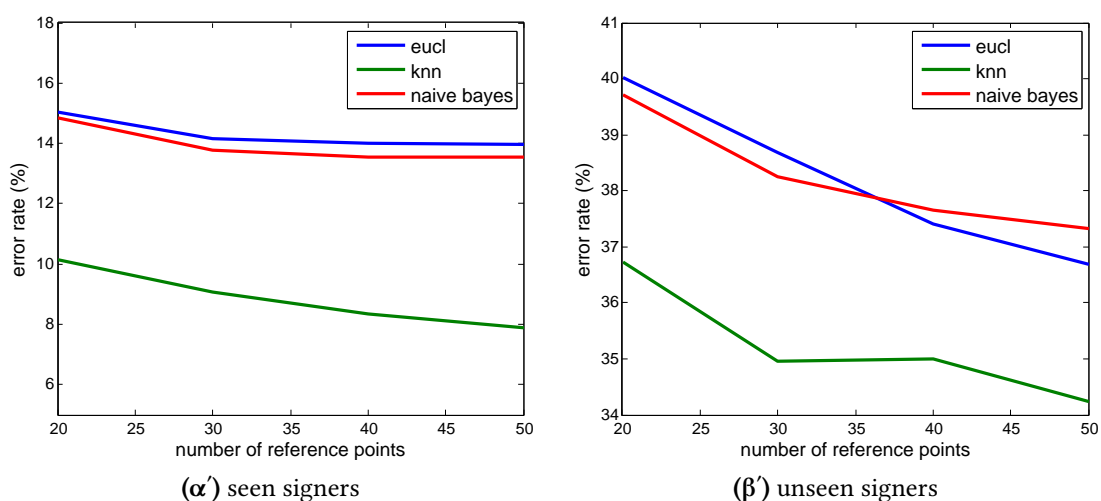
Σχήμα 7.27: Αναλυτική επίδραση των χωρικών παραμέτρων για rthog περιγραφές



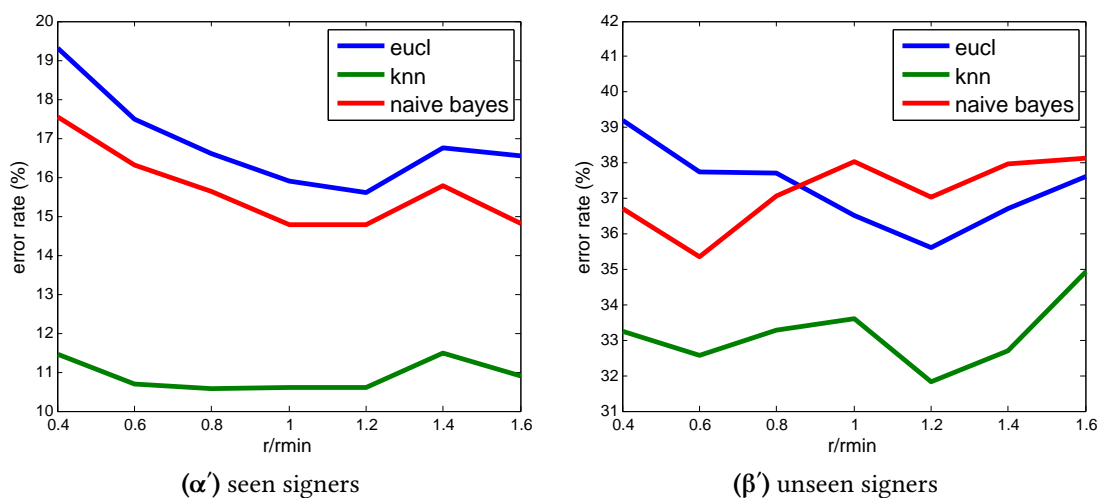
Σχήμα 7.28: Αναλυτική επίδραση των χωρικών παραμέτρων για ερθηος περιγραφές

7.7.5 Παραλλαγή των Shape Context

Στο Κεφάλαιο 6 προτείναμε μια εναλλακτική χρήση των shape context για δημιουργία ενιαίων διανυσμάτων χαρακτηριστικών. Αρχικά παρουσιάζουμε την επίδραση του πλήθους των σημείων αναφοράς στο Σχήμα 7.29, από όπου συμπεραίνουμε ότι μια καλή επιλογή, την οποία θα θεωρούμε δεδομένη για την συνέχεια του πειραματικού μέρους, είναι 40 σημεία αναφοράς ισοκαταναμημένα στον κύκλο αναφοράς. Στη περιγραφή της παραλλαγής αυτής υποθέσαμε αυθαίρετα ότι επιλέγουμε τα σημεία αναφοράς της μεθόδου πάνω σε κύκλο ακτίνας ίσης με τον μέγιστο εγγεγραμμένο κύκλο του περιγράμματος του χεριού. Η επιλογή της ακτίνας είναι μια σημαντική παράμετρος προς μελέτη και στο Σχήμα 7.30 παρουσιάζουμε την επίδραση της στο ποσοστό αναγνώρισης για 40 σημεία αναφοράς. Παρατηρούμε ότι η βέλτιστη επιλογή είναι 1.2 φορές την ακτίνα του μέγιστου εγγεγραμμένου κύκλου.



Σχήμα 7.29: Επίδραση του πλήθους των σημείων αναφοράς



Σχήμα 7.30: Επίδραση της ακτίνας του κύκλου αναφοράς

7.8 Σύγκριση των Τεχνικών που Παρουσιάστηκαν

Επιλέγοντας τις βέλτιστες παραμέτρους για τις τεχνικές που παρουσιάστηκαν, με βάση την ως τώρα μελέτη, συνοψίζουμε τα αποτελέσματα των προτεινόμενων τεχνικών σε σύγκριση και με δημοφιλείς τεχνικές. Στις δυο επόμενες υποενότητες λοιπόν θα παρουσιάσουμε την αποτελεσματικότητά τους, διαχωρίζοντας τις τεχνικές περιγράμματος και αυτές που λαμβάνουν υπόψιν και εσωτερική πληροφορία.

Στις ακόλουθες συγκρίσεις δεν χρησιμοποιούνται τεχνικές που είναι πλήρως ανεξάρτητες στην περιστροφή (HU Moments - παραλλαγές όλων των μεθόδων που βασίζονται σε Fourier Transform Magnitude), καθώς αναμένεται να έχουν μικρότερα ποσοστά αναγνώρισης. Συνεπώς δεν υπάρχει λόγος να παρουσιαστούν στις ακόλουθες συγκρίσεις, ενώ θα γίνει εκτενής αναφορά στο πρόβλημα της ανεξαρτησίας στην περιστροφή σε επόμενη ενότητα.

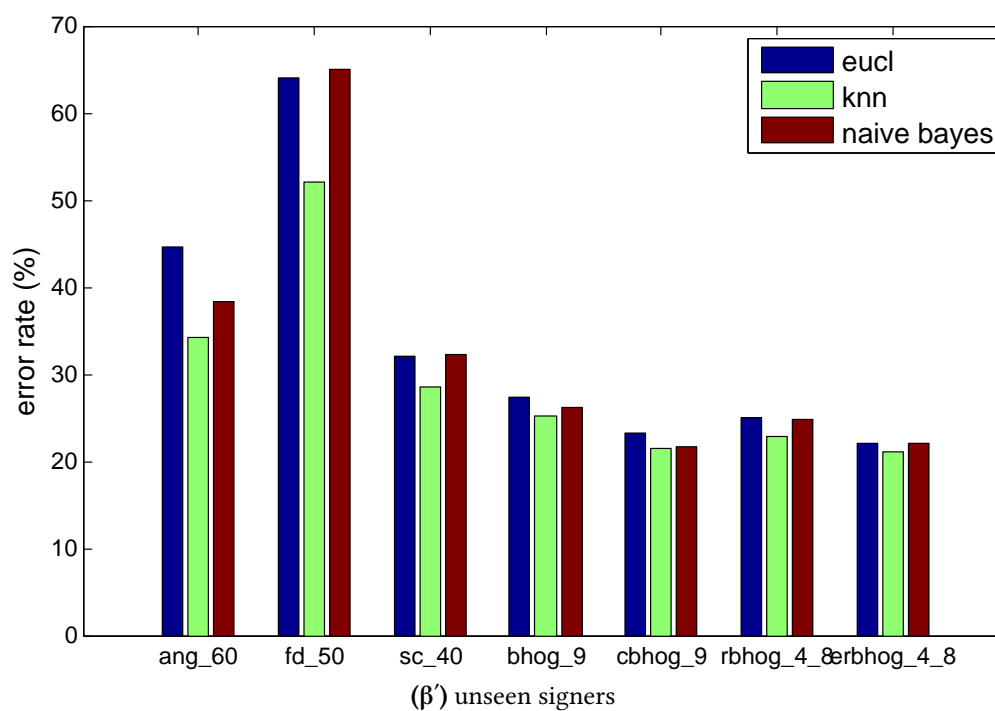
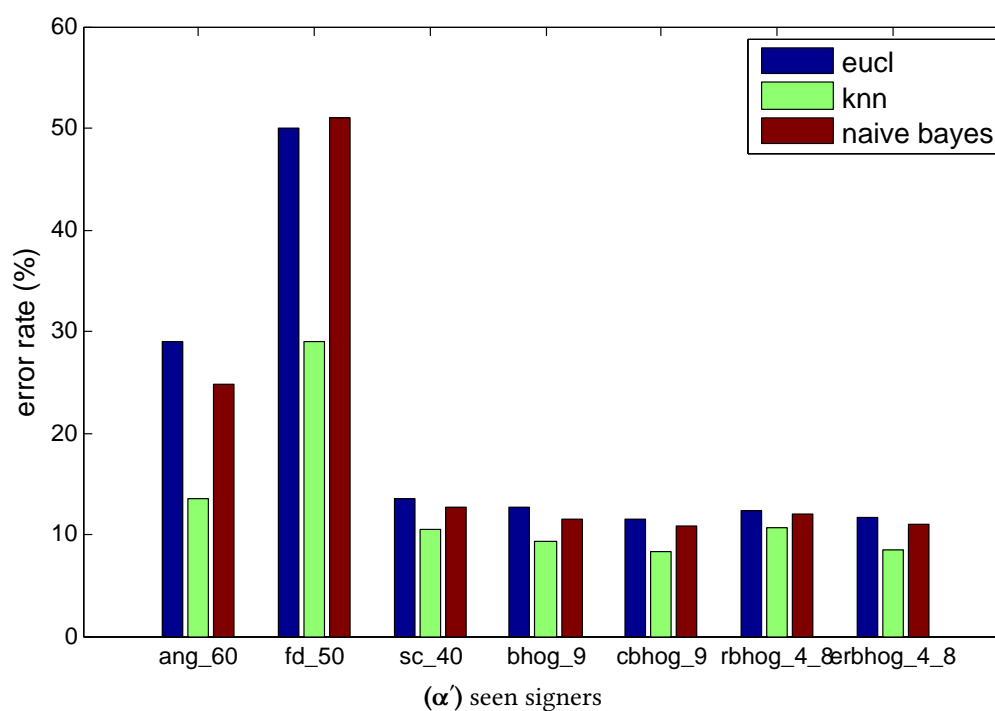
7.8.1 Σύγκριση Τεχνικών Περιγράμματος

Στο Σχήμα 7.31 απεικονίζεται το σφάλμα αναγνώρισης για τις βέλτιστες παραμετροποιήσεις των τεχνικών περιγράμματος. Οι προτεινόμενες τεχνικές είναι αισθητά αποτελεσματικότερες από τις *ang* και *fd*, ενώ οι παραλλαγές των hog-based τεχνικών για δυαδικές εικόνες παρουσιάζουν το μικρότερο σφάλμα αναγνώρισης υπερτερώντας της παραλλαγή των shape context. Συγκεκριμένα οι περιγραφές *cbhog_9* και *erbhog_4_8* παρουσιάζουν την καλύτερη συνολική επίδοση, επιβεβαιώνοντας την αναμενόμενη ισχύ των μεθόδων.

Επιπλέον πρέπει να αναφέρουμε ότι στην περίπτωση των μεθόδων των Fourier Descriptors και της δειγματοληψίας της ακτινικής απόστασης του περιγράμματος (*ang*) ο ταξινομητής kNN έχει αξιοσημείωτα καλύτερα αποτελέσματα από τους υπόλοιπους, οι οποίοι έχουν επιλεγεί για τεχνικές βασισμένες σε HOG.

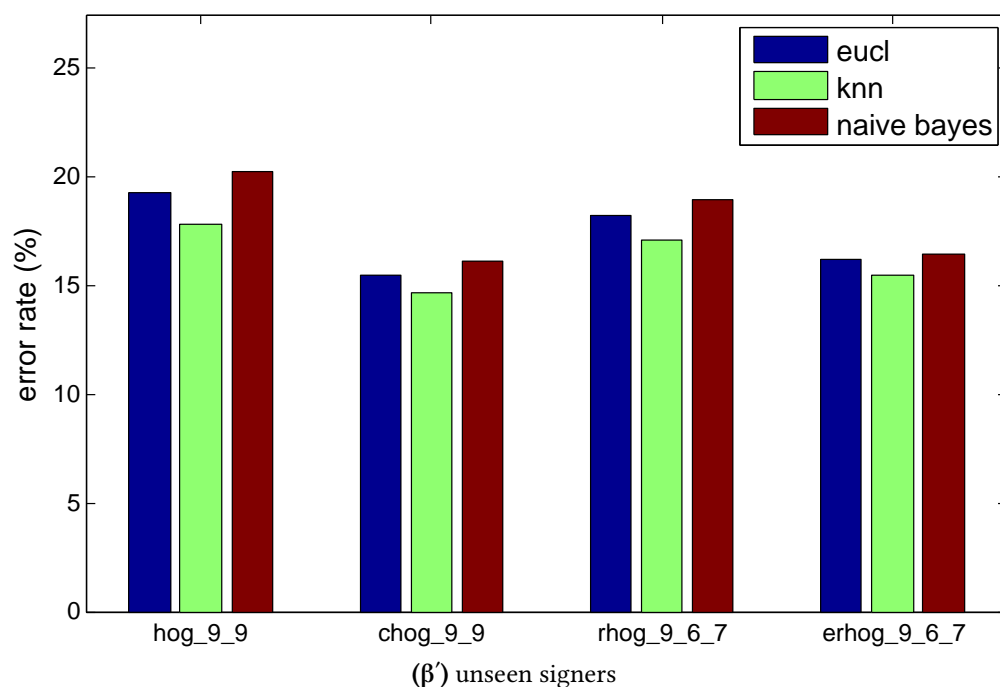
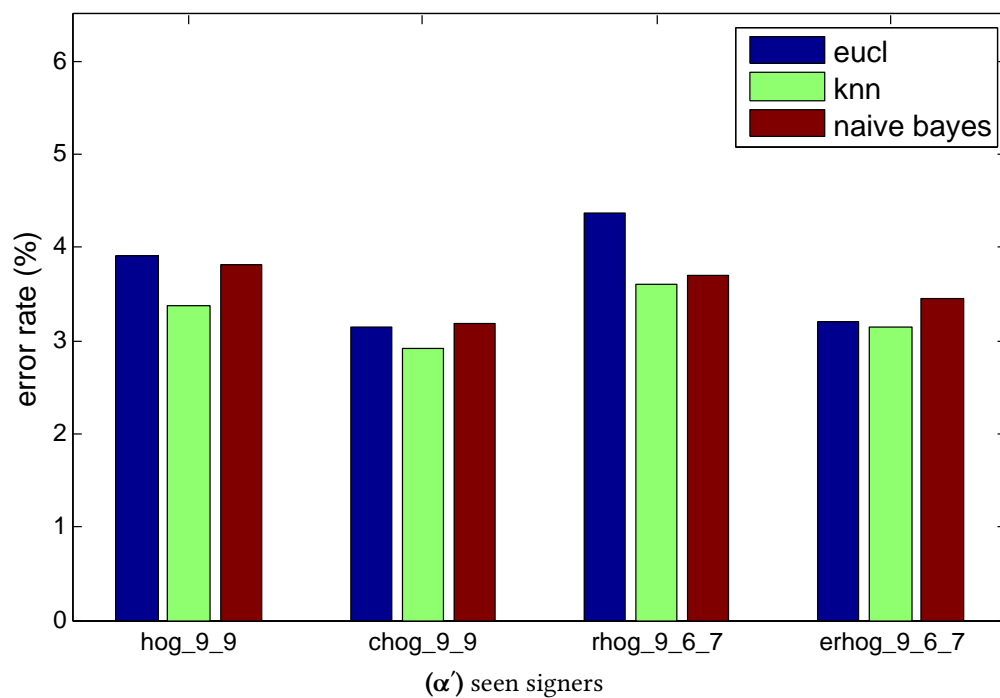
7.8.2 Σύγκριση Τεχνικών Εμφάνισης Βασισμένων σε HOG

Στο Σχήμα 7.32 απεικονίζεται το σφάλμα αναγνώρισης για τις βέλτιστες παραμετροποιήσεις των τεχνικών εμφάνισης που βασίζονται στην βασική μέθοδο των HOG. Αντίστοιχα με τα αποτελέσματα των τεχνικών περιγράμματος η παραλλαγή των Centralized HOG προσφέρει το μικρότερο σφάλμα αναγνώρισης και αποτελεί την βέλτιστη επιλογή, ενώ η παραλλαγή με χρήση ελλειπτικής διαμέρισης του χώρου (*erhog*) υστερεί ελάχιστα σε ικανότητα αναγνώρισης και αποτελεί την δεύτερη καλύτερη επιλογή. Η κυκλική χωρική διάταξη των HOG (*rhog*) έχει παρόμοια ποσοστά σφάλματος με την βασική ορθογώνια διαμέριση (*hog*) και όπως είναι λογικό είναι λιγότερο αποτελεσματικές τεχνικές από τις δυο προηγούμενες, αφού είναι πιο γενικές (σταθερές χωρικές



Σχήμα 7.31: Σύγκριση τεχνικών περιγράμματος

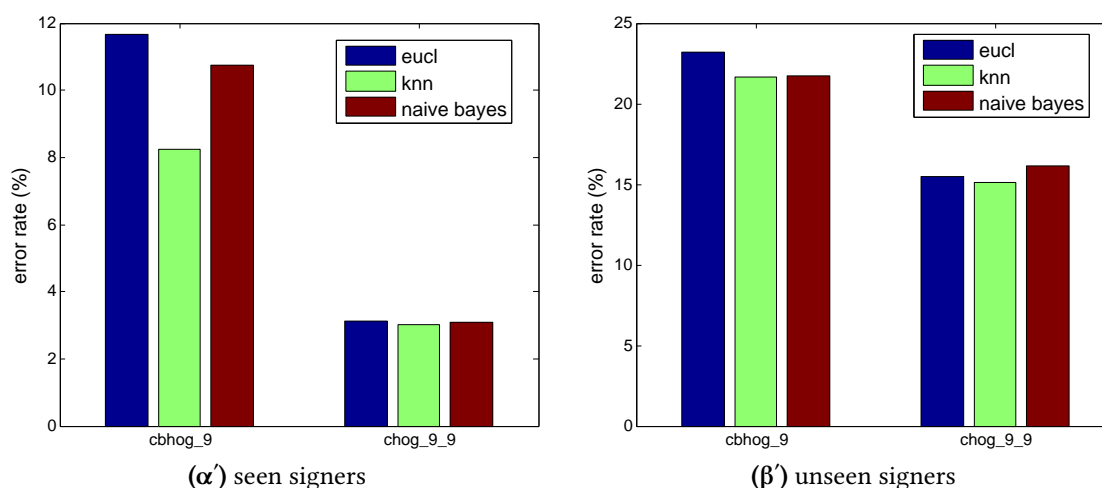
διαμερίσεις) και λιγότερο προσαρμόσιμες στην εκάστοτε περίπτωση όπως οι πρώτες που αξιοποιούν απλοϊκές πληροφορίες για το σχήμα των χειρομορφών.



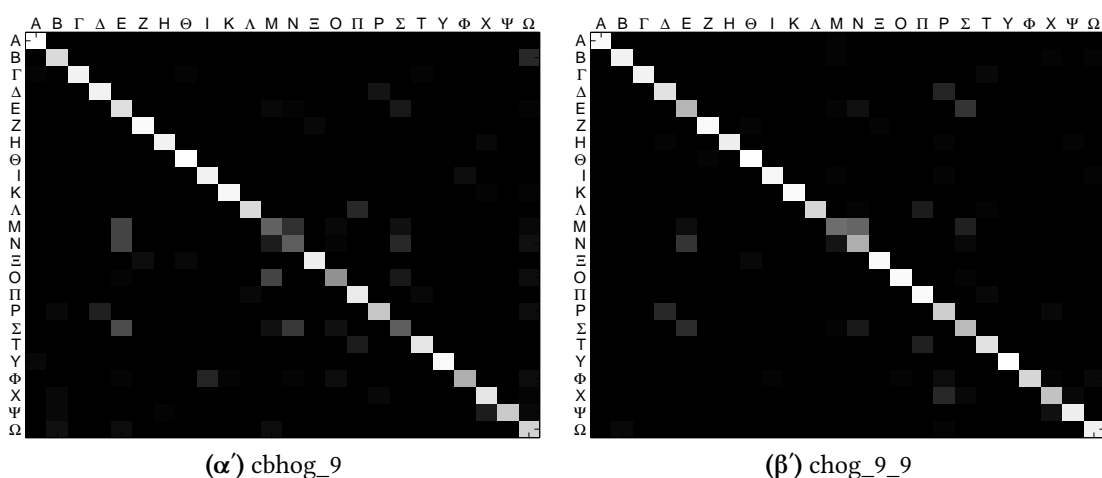
Σχήμα 7.32: Σύγκριση τεχνικών βασισμένων σε HOG περιγραφές

7.8.3 Συνολική Σύγκριση

Στο Σχήμα 7.33 γίνεται μια επιλεκτική παρουσίαση των καλύτερων τεχνικών των δυο προηγούμενων υποενοτήτων (Centralized HOG και Centralized Binary HOG), όπου φαίνεται ξεκάθαρα η ανωτερότητα των τεχνικών εμφάνισης σε σχέση με τις τεχνικές περιγράμματος (>5%). Το προαναφερθέν συμπέρασμα είναι απόλυτα λογικό, καθώς οι τεχνικές εμφάνισης χρησιμοποιούν αρκετά μεγαλύτερο μέρος της διαθέσιμης πληροφορίας, δηλαδή το σύνολο των ακμών στην περίπτωση των τεχνικών που βασίζονται σε HOG. Επιπλέον στο Σχήμα 7.34 παραθέτουμε τους πίνακες σύγχυσης για τις παραπάνω τεχνικές, έτσι ώστε να κατανοήσουμε καλύτερα την διαφορά των τεχνικών στο επίπεδο διαχωρισμού των χειρομορφών. Όπως ήταν αναμενόμενο είναι εμφανής η βελτίωση με χαρακτηριστικότερο το παράδειγμα των κλάσεων E, M και N που έχουν παρόμοιο περίγραμμα και συνεπώς δεν είναι δυνατόν να περιγραφούν ικανοποιητικά με τεχνικές περιγράμματος.



Σχήμα 7.33: Σύγκριση των τεχνικών Centralized HOG και Centralized Binary HOG



Σχήμα 7.34: Confusion Matrices για Centralized HOG και Centralized Binary HOG

7.9 Τεχνικές για Ανεξαρτησία στην Περιστροφή

Πέρα από δημοφιλείς τεχνικές όπως Hu Momentes, Fourier Descriptors και Shape Context (αν και βέβαια στην παρούσα διπλωματική η τεχνική αυτή χρησιμοποιείται παραλλαγμένη) που μας προσφέρουν ανεξαρτησία στην περιστροφή των εικόνων, έχουμε δημιουργήσει ένα πλήθος παραλλαγών των HOG προς αυτή την κατεύθυνση και σκοπός της ενότητας αυτής είναι να αξιολογήσουμε την αποτελεσματικότητά τους. Για την αξιολόγηση ως προς την περιστροφική ανεξαρτησία έχουμε δημιουργήσει μια νέα τεχνική βάση περιστρέφοντας κάθε εικόνα της αρχικής βάσης ανά 5 μοίρες στο διάστημα $[0, 90^\circ]$, θεωρώντας ότι το τεταρτημόριο αυτό καλύπτει όλες τις δυνατές παραμορφώσεις στον χώρο που εισάγει η περιστροφή μιας εικόνας και αναμένοντας την ίδια συμπεριφορά με ελάχιστες αποκλίσεις και στα υπόλοιπα τεταρτημόρια. Ο διαχωρισμός της βάσης σε σύνολα εκπαίδευσης και δοκιμών γίνεται με την ίδια λογική της προηγούμενης ενότητας, με την διαφορά ότι στο σύνολο εκπαίδευσης γίνεται χρήση μόνο των εικόνων χωρίς περιστροφή ενώ το σύνολο δοκιμών προέρχεται από όλες τις δυνατές περιστροφές. Η επιλογή αυτή γίνεται για να αναδείξει την ικανότητα των τεχνικών για πραγματικά ανεξάρτητες περιγραφές.

7.9.1 Ακρίβεια Εκτίμησης Περιστροφής

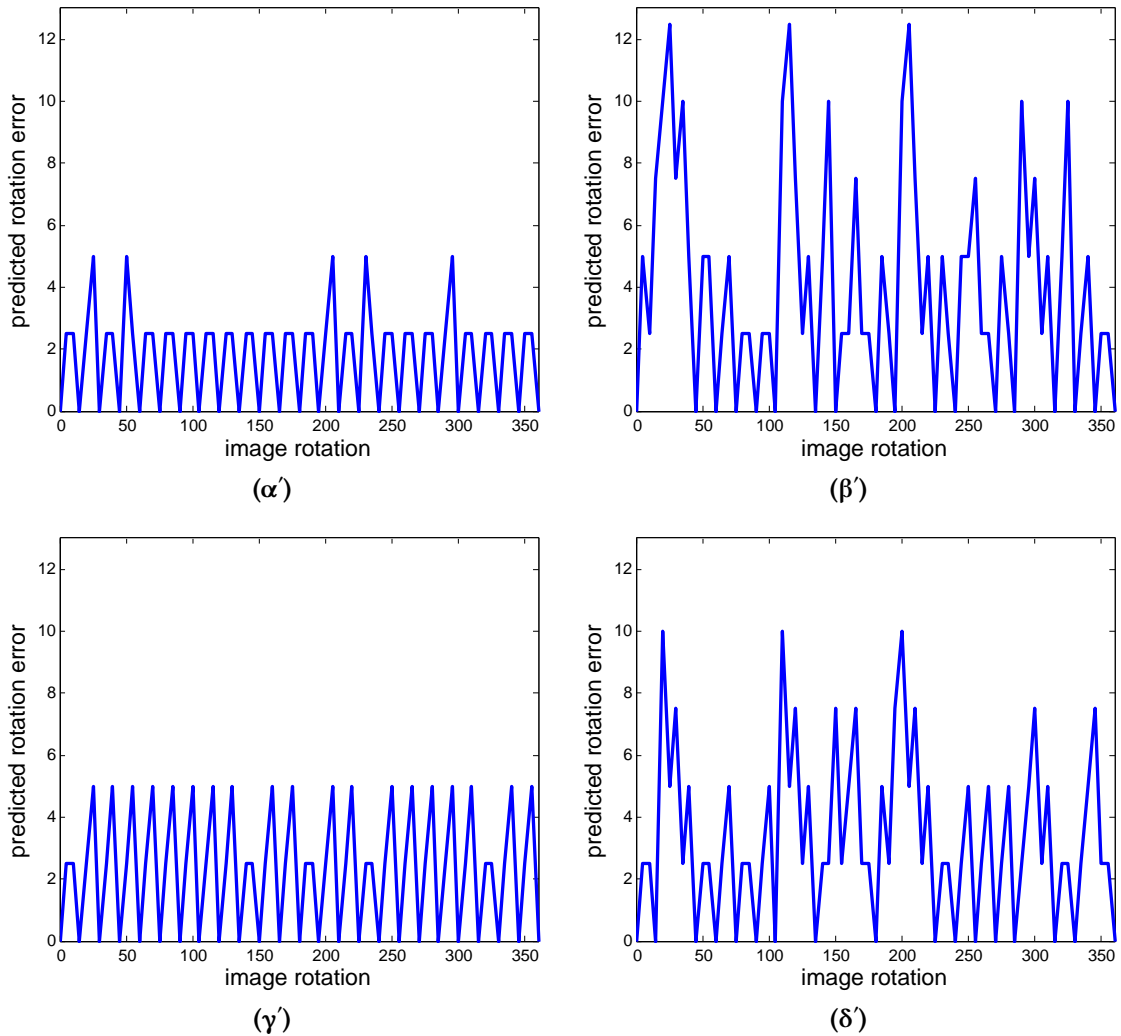
Αρχικά θα μελετήσουμε την ικανότητα διάκρισης της γωνίας περιστροφής με την μέθοδο cross-correlation. Θεωρητικά η εύρεση της γωνίας περιστροφής μπορεί να υπολογιστεί με χρήση της μεθόδου της ετεροσυσχέτισης, χρησιμοποιώντας την περιγραφή των RHOG (ή παραλλαγών της) πριν την εφαρμογή του Fourier Transform Magnitude.

Όπως αναλύσαμε στο προηγούμενο κεφάλαιο προτείνουμε μια μειωμένη περιγραφή σε σύγκριση με την συνηθισμένη τακτική για κυκλικές περιγραφές, δηλαδή πολικές κατανομές, και σκοπός μας είναι επαληθεύσουμε την αποτελεσματικότητας της προτεινόμενης τεχνικής. Συγκεκριμένα για τις δυο χειρομορφές της ίδιας κλάσης του Σχήματος 7.35 (η πρώτη βρίσκεται κοντά στο κέντρο της κλάσης ως προς την περιγραφή της, ενώ η άλλη αποκλίνει) παρουσιάζουμε το σφάλμα εκτίμησης της γωνίας περιστροφής για την μειωμένη μέθοδο στο Σχήμα 7.36 και για την συμβατική στο Σχήμα 7.37.



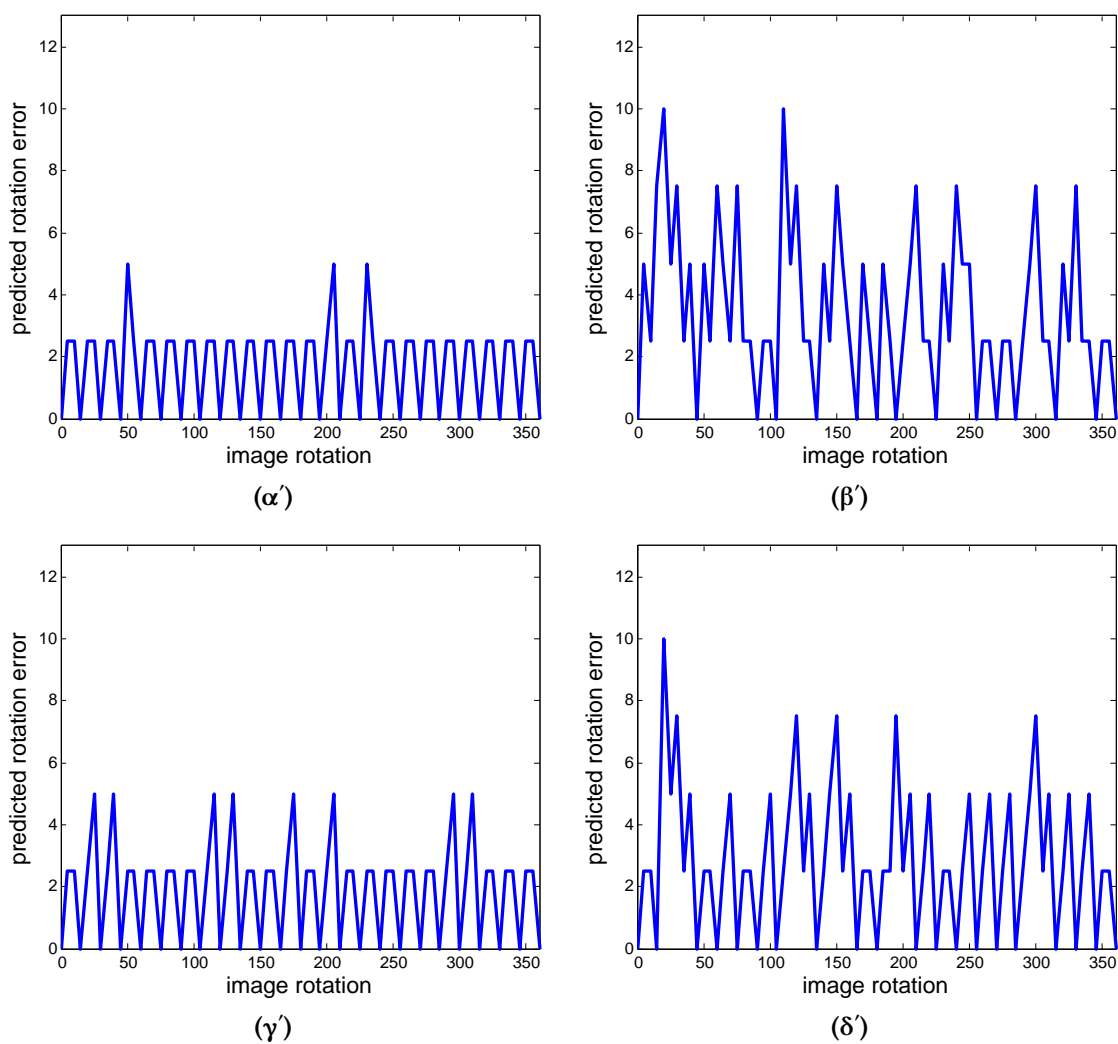
Σχήμα 7.35: Εικόνες αναφοράς και δοκιμών

Παρατηρούμε ότι η κλασική τεχνική (πολική διαμέριση) αποδίδει ανεπαίσθητα καλύτερα αλλά συνυπολογίζοντας την αναλογία στο μέγεθος του τελικού διανύσματος (σχεδόν διπλάσιο από την μειωμένη), μπορούμε να αποφανθούμε ότι η προτεινόμενη τεχνική μειωμένης περιγραφής πρόκειται για μια αποδοτική επιλογή.

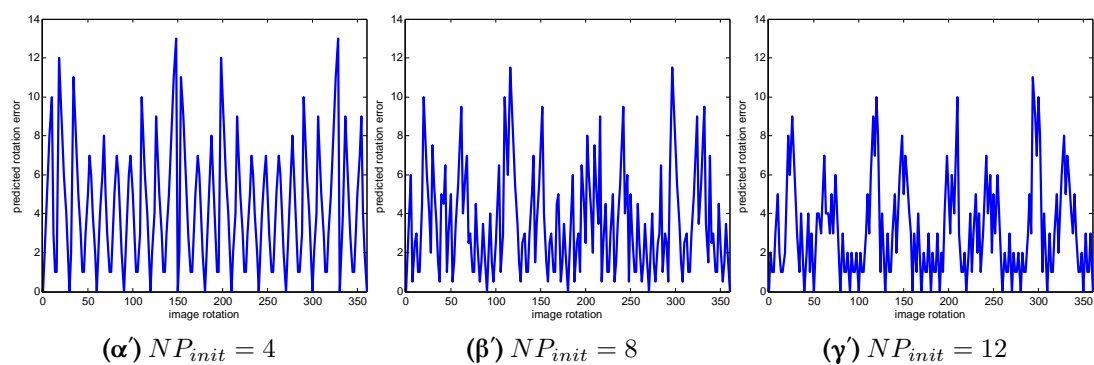


Σχήμα 7.36: Σφάλμα εκτίμησης γωνίας περιστροφής για την μειωμένη περιγραφή. (α),(β): ίδια εικόνα αναφοράς (γ),(δ): ως διάνυσμα αναφοράς χρησιμοποιείται το κέντρο της κλάσης

Αναλυτικότερα για την διακριτική ικανότητα των γωνιών περιστροφής, βλέπουμε ότι έχουμε μέγιστο σφάλμα εκτίμησης 120 στο διάγραμμα 7.36(β), όπου η εικόνα αναφοράς 7.35(α) και η περιστρεφόμενη εικόνα 7.35(β) παρουσιάζουν σημαντικές οπτικές διαφορές. Στην περίπτωση του Σχήματος 7.36(α) όπου η περιστρεφόμενη εικόνα είναι η ίδια με την εικόνα αναφοράς 7.35(α), το μέγιστο σφάλμα εκτίμησης είναι 50. Γενικά καλύτερη συμπεριφορά εμφανίζεται στα διαγράμματα 7.36(γ),(δ), όπου η περιστροφή υπολογίζεται ως προς το μέσο διάνυσμα των περιγραφών της κλάσης. Οι εκτιμήσεις αυτές έχουν γίνει για κυκλική κατανομή 6 κύκλων και 8 αρχικών σημείων στον εσωτερικό κύκλο, το οποίο σημαίνει ότι στον εξωτερικό κύκλο χρησιμοποιούνται 48 σημεία - κελιά, οπότε το πλήθος αυτό ορίζει και την θεωρητική διακριτική ικανότητα της με-



Σχήμα 7.37: Σφάλμα εκτίμησης της γωνίας περιστροφής για την πλήρη περιγραφή. (α),(β): ίδια εικόνα αναφοράς (γ),(δ): ως διάνυσμα αναφοράς χρησιμοποιείται το κέντρο της κλάσης



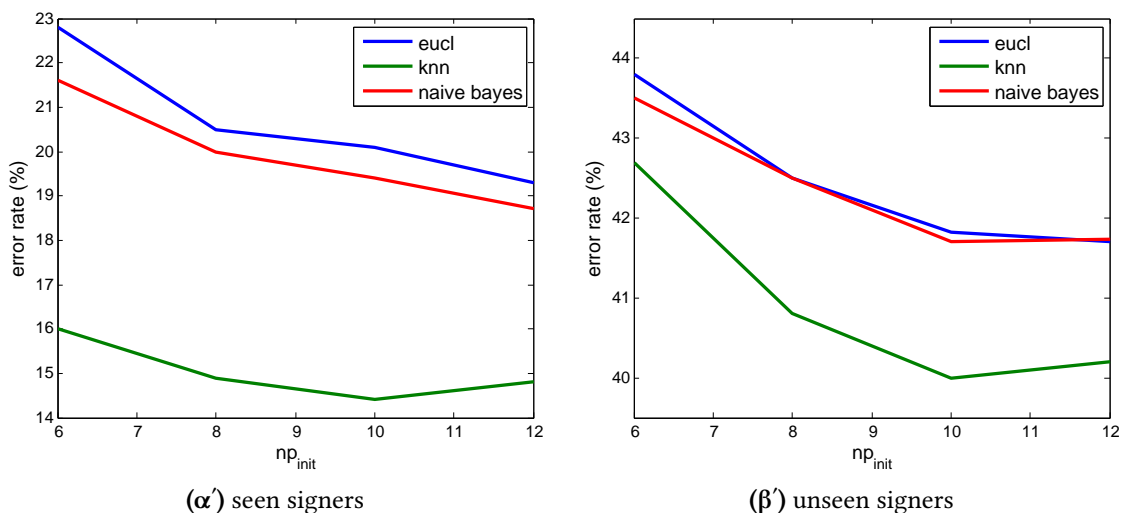
Σχήμα 7.38: Σφάλμα εκτίμησης της γωνίας περιστροφής για διαφορετικό πλήθος σημείων στον χώρο

θόδου που εν προκειμένω είναι $360/48 = 7,5^\circ$, δηλαδή το μέγιστο θεωρητικό σφάλμα είναι $3,75^\circ$. Όπως είναι λογικό στην πράξη θα έχουμε μεγαλύτερο σφάλμα από το θεωρητικό, που στην περίπτωση του Σχήματος 7.36(α) προσεγγίζει πολύ το θεωρητικό αφού έχουμε ίδια εικόνα αναφοράς και περιστροφής, ενώ οι αποκλίσεις στο (β) θεωρούνται αποδεκτές για τις διαφορές που παρουσιάζουν οι εικόνες.

Επιπλέον στο Σχήμα 7.38 παρουσιάζεται το σφάλμα εκτίμησης της γωνίας μεταβάλλοντας το πλήθος των σημείων δειγματοληψίας κάθε δακτυλίου (ορίζοντας διαφορετικό πλήθος για τον εσωτερικό δακτύλιο) για την εξαγωγή τοπικών περιγραφών. Όπως είναι αναμενόμενο, παρατηρούμε ότι αυξάνοντας το πλήθος εξομαλύνονται οι διαφορές στην εκτίμηση της γωνίας περιστροφής.

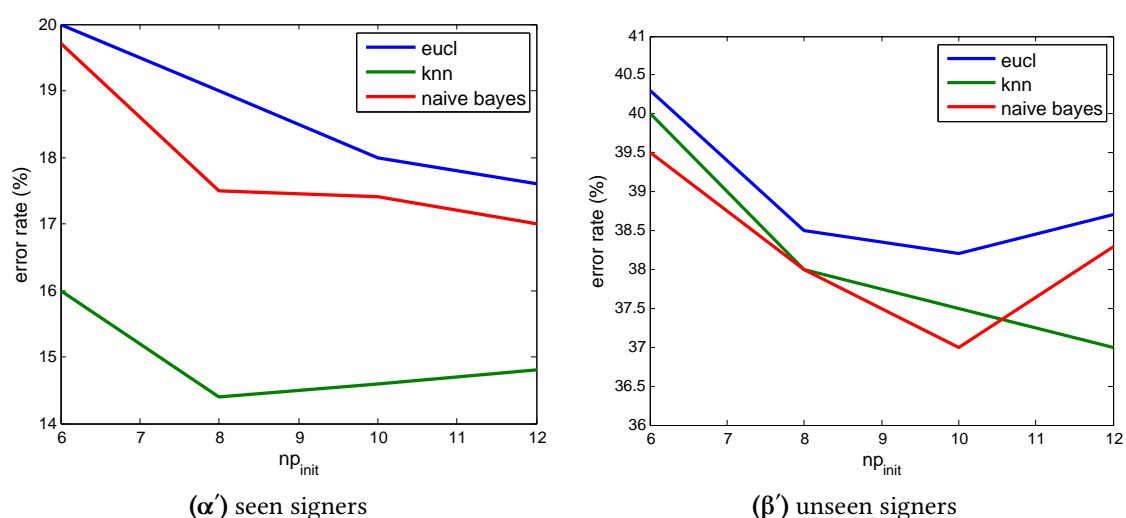
7.9.2 RBHOG & ERBHOG

Η περίπτωση της δυαδικής παραλλαγής των περιστροφικά ανεξάρτητων HOG, λόγω της φύσης των δυαδικών εικόνων, δεν έχει πολλές παραμέτρους ως προς μελέτη. Συγκεκριμένα για την επίτευξη της ανεξαρτησίας των κελιών υπό περιστροφή είναι εφαρμόσιμη μόνο η 3η παραλλαγή, όπως έχουμε αναλύσει εκτενώς στο Κεφάλαιο 6. Θεωρώντας λοιπόν ότι το πλήθος των δακτυλίων της μεθόδου δεν επηρεάζει την απόδοση για περιστρεμμένες εικόνες, μελετάμε την επίδραση του πλήθους των τοπικών περιγραφών ανά δακτύλιο που σχετίζεται άμεσα με την ακρίβεια αναγνώρισης της περιστροφής. Συνεπώς, σύμφωνα με το Σχήμα 7.39, συμπεραίνουμε ότι η βέλτιστη επιλογή για ελαχιστοποίηση του σφάλματος στην αναγνώριση είναι `rbhog&ftm_4_10`.



Σχήμα 7.39: Επίδραση του πλήθους των τοπικών περιγραφών των δακτυλίων (ρ)

Ομοίως για την περίπτωση της ελλειπτικής παραλλαγής ERBHOG, σύμφωνα με το Σχήμα 7.40 καταλήγουμε στην επιλογή της περιγραφής `erbhog&ftm_4_10` ως βέλτιστη.



Σχήμα 7.40: Επίδραση του πλήθους των τοπικών περιγραφών των ελλειπτικών δακτυλίων (*erhog*)

7.9.3 RHOG & ERHOG

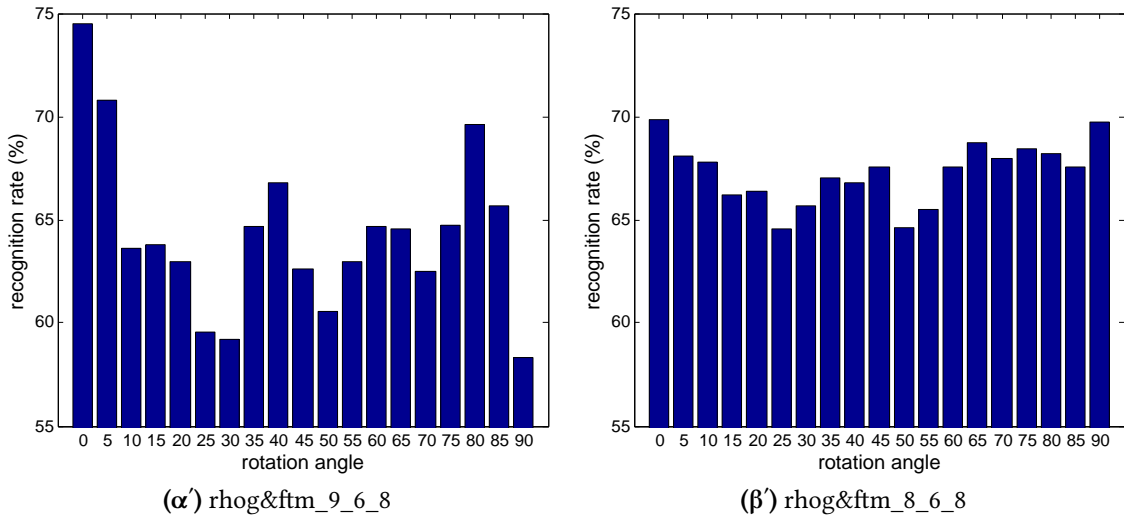
Για την περίπτωση των περιστροφικά ανεξάρτητων HOGs, έχει ιδιαίτερη σημασία η μελέτη των πιθανών τεχνικών ανεξαρτησίας των κελιών στην περιστροφή. Όπως αναφέραμε στο Κεφάλαιο 6, οι τεχνικές αυτές βασίζονται στον Radial Gradient Transform ([25]) και διαδοχικές προσεγγίσεις-απλοποιήσεις του. Συνολικά χρησιμοποιούμε 4 τεχνικές, η αρχική και 3 προσεγγίσεις της, που τις συμβολίζουμε ως ci_0 , ci_1 , ci_2 και ci_3 (Cell Invariance).

Αρχικά θα εστιάσουμε στις 2 τελευταίες προσεγγίσεις, που βασίζονται στην κυκλική ολίσθηση του αρχικού ιστογράμματος και είναι πιθανό να εισάγουν μεγάλες ανακρίβειες λόγω της κβαντισμένης (ανάλογα με το πλήθος των bins) προσέγγισης της γωνίας του κέντρου των κελιών με το κέντρο της εικόνας. Ιδανικά θα θέλαμε να τοποθετήσουμε όσο το δυνατόν περισσότερα κέντρα κελιών σε γωνίες που είναι πολλαπλάσιες του εύρους των bins, έτσι ώστε η πράξη της αφαίρεσης των γωνιών σύμφωνα με την αρχική τεχνική και την 1η προσέγγιση να ισοδυναμεί με κυκλική μετατόπιση του ιστογράμματος.

Παρότι κάτι τέτοιο δεν είναι εύκολο στην γενική περίπτωση, μπορούμε να ορίσουμε ως ένα βαθμό ένα μεγάλο ποσοστό των ρυθμίζοντας κατάλληλα τις χωρικές παραμέτρους σε συνδυασμό με το πλήθος των bins. Συγκεκριμένα η χωρική παράμετρος που μας ενδιαφέρει είναι το πλήθος των τοπικών περιγραφών του εσωτερικού δακτυλίου (np_{init}), που ουσιαστικά ορίζει τις γωνίες για τα κέντρα των εσωτερικών περιγραφών, και επιθυμούμε να έχει όσο το δυνατόν περισσότερους κοινούς πρώτους παράγοντες με τον αριθμό των bins.

Επιβεβαιώνουμε την παραπάνω περιγραφείσα συμπεριφορά παρουσιάζοντας στο

Σχήμα 7.41 τα ποσοστά αναγνώρισης ανά γωνία για τους περιγραφείς $rhog\&ftm_9_6_8$ και $rhog\&ftm_8_6_8$ κάνοντας χρήση του ευκλείδειου ταξινομητή. Είναι εμφανές ότι στο Σχήμα 7.41(α) η περιγραφή είναι σχετικά ασταθής για περιστροφές και χάνεται η ικανότητα ικανοποιητικής γενίκευσης. Αντίθετα στο Σχήμα 7.41(β), όπου οι επιλογές παραμέτρων έχουν γίνει σύμφωνα με τις επιθυμητές προϋποθέσεις, παρατηρούμε πολύ καλύτερη συμπεριφορά ως προς την αναγνώριση υπό περιστροφή.

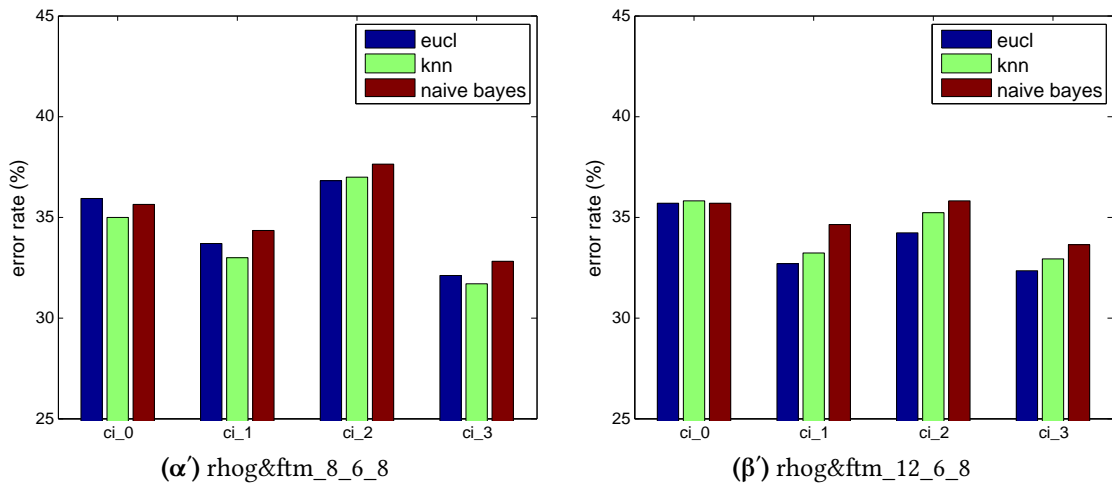


Σχήμα 7.41: Ικανότητα αναγνώρισης ανά γωνία περιστροφής (*eucl,unseen signers*)

Έχοντας κατανοήσει πλέον τις προϋποθέσεις για την ορθή λειτουργία των προσεγγίσεων που βασίζονται σε ολίσθηση ιστογραμμάτων θα εκτιμήσουμε την ικανότητα αναγνώρισης για τις διάφορες προσεγγίσεις για τους περιγραφείς $rhog\&ftm_8_6_8$ και $rhog\&ftm_12_6_8$. Τα αποτελέσματα παρουσιάζονται στο Σχήμα 7.42 και περιοριζόμαστε σε *unseen signers* για λόγους απλότητας. Γενικά η αύξηση των bins δεν φαίνεται να επηρεάζει σημαντικά τις ως προς μελέτη τεχνικές (μάλιστα αυξάνει κατά λίγο το σφάλμα αναγνώρισης), εκτός από την περίπτωση της 2ης προσέγγισης, δηλαδή της απλής ολίσθησης των αρχικών ιστογραμμάτων. Παρατηρούμε ότι συνολικά η καλύτερη επίδοση επιτυγχάνεται με την 3η παραλλαγή, η οποία θεωρητικά έχει και αρκετά καλή υπολογιστική απόδοση (η κυκλική ολίσθηση είναι μια γρήγορη βασική πράξη των υπολογιστικών συστημάτων). Συνεπώς επιλέγουμε την προσέγγιση αυτή ως βασική τεχνική ανεξαρτησίας των τοπικών περιγραφών σε περιστροφές και στην συνέχεια του πειραματικού μέρους θα θεωρούμε την επιλογή αυτή ως δεδομένη.

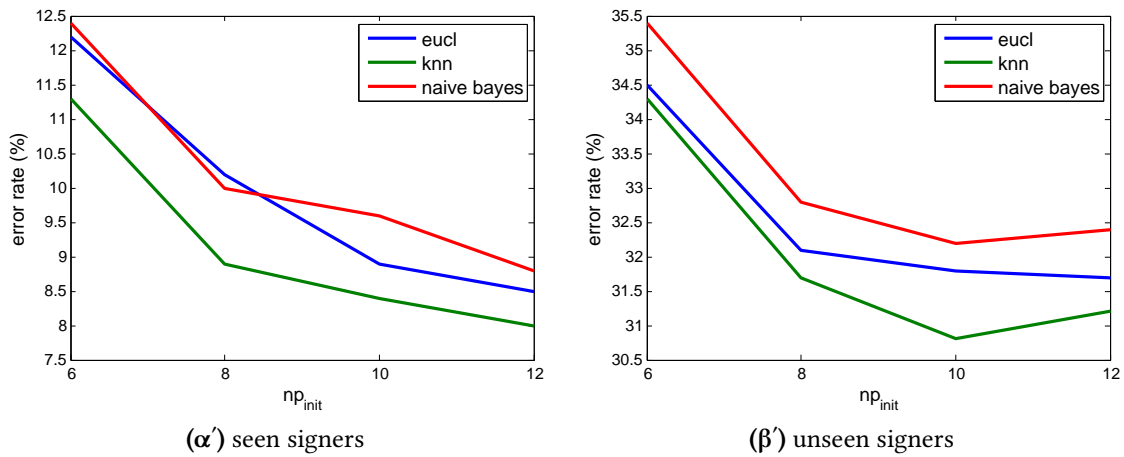
Ως προς την επιλογή των χωρικών παραμέτρων, μεταβάλλουμε την παράμετρο np_{init} , επιλέγοντας παράλληλα κατάλληλη τιμή για τον αριθμό των bins. Τα ποσοστά εσφαλμένης αναγνώρισης απεικονίζονται στο Σχήμα 7.43 διατηρώντας σταθερές τις εξής παραμέτρους: 8 bins, 6 δακτυλίου.

Τα συμπεράσματα για την επιλογή της τεχνικής ανεξαρτησίας των κελιών στην περιστροφή και συνακολούθως για τον αριθμό των bins είναι λογικό να γενικευθούν και για την μέθοδο *ethog*, καθώς μεταβάλλουμε μόνο την χωρική κατανομή των τοπικών περιγραφών. Για τον λόγο αυτό μελετάμε μόνο τις χωρικές παραμέτρους της μεθόδου



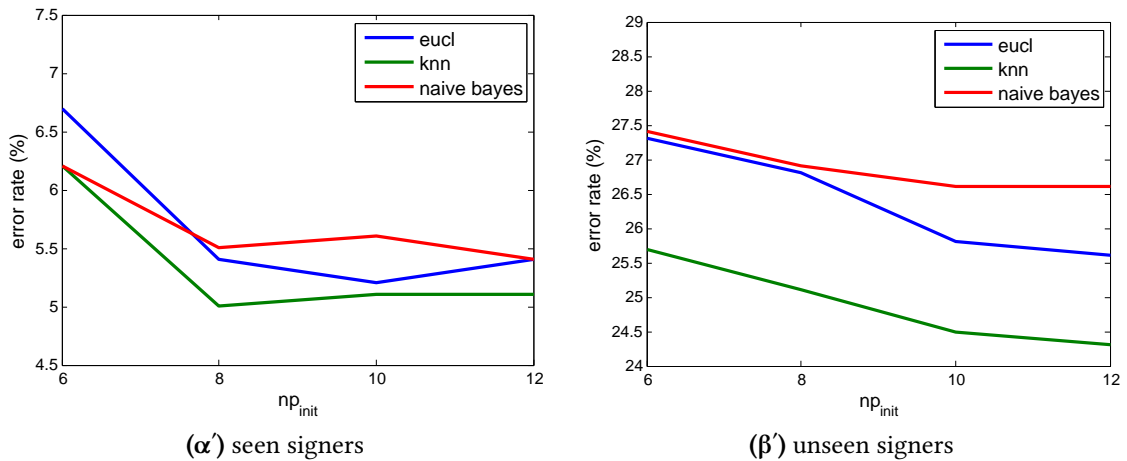
Σχήμα 7.42: Σύγκριση των τεχνικών ανεξαρτησίας των τοπικών περιγραφών στην περιστροφή

και συγκεκριμένα την παράμετρο np_{init} . Η επίδραση της χωρικής παραμέτρου αυτής απεικονίζεται στο Σχήμα 7.44, όπου επίσης έχουμε χρησιμοποιήσει 8 bins και 6 δακτυλίους ως σταθερές παραμέτρους.



Σχήμα 7.43: Επίδραση της χωρικής παραμέτρου np_{init} ($rhog&ftm_8_6_{np_{init}}$)

Σύμφωνα με τα Σχήματα 7.43, 7.44 συμπεραίνουμε ότι η επιλογή $np_{init} = 10$ είναι βέλτιστη τόσο για την RHOG όσο και για την ERHOG μέθοδο. Επιπλέον παρατηρούμε ότι η ERHOG είναι εξαιρετικά αποτελεσματική με εμφανώς μικρότερο σφάλμα αναγνώρισης (6%) από την RHOG. Βέβαια η RHOG μέθοδος είναι πιο γενικεύσιμη, καθώς δεν χρειάζεται να έχουμε επιπλέον πληροφορία για την εφαρμογή της σε αντίθεση με την ERHOG που απαιτεί πληροφορία για το περίγραμμα της χειρομορφής.



Σχήμα 7.44: Επίδραση της χωρικής παραμέτρου np_{init} ($erhog&ftm_8_6_np_{init}$)

7.9.4 Μέθοδος Περιγραμμάτων Αναφοράς

Στο σημείο αυτό θα αξιολογήσουμε την εναλλακτική προσέγγιση του προβλήματος της ανεξαρτησίας σε περιστροφές που εκμεταλλεύεται την επιτυχία της μεθόδου Shape Context ως προς την εύρεση της σχέσης δυο περιγραμμάτων με την μορφή ενός μετασχηματισμού. Η μέθοδος αυτή χρησιμοποιεί περιγράμματα αναφοράς για κάθε κλάση έτσι ώστε να βρει την συσχέτιση τους (και εν τέλει η περιστροφή τους) με κάποια εικόνα εισόδου και για τον λόγο αυτό την ονομάζουμε μέθοδο περιγραμμάτων αναφοράς (Shape Reference Method - SRef).

Στους Πίνακες 7.8, 7.7 παρουσιάζουμε τα ποσοστά επιτυχούς αναγνώρισης μεταβάλλοντας το αρχικό και το τελικό στάδιο της μεθόδου, δηλαδή τα στάδια που σχετίζονται με τα shape context και τα hog αντίστοιχα. Συγκεκριμένα για το αρχικό στάδιο εφαρμόζονται οι δυο αναφερθείσες παραλλαγές στο θεωρητικό κομμάτι, αυτή του αφινικού μετασχηματισμού που σχετίζεται άμεσα με την κλασσική μέθοδο των shape context και αυτή του μετασχηματισμού ομογένειας που προτείνουμε και βασίζεται σε ανεξάρτητα σημεία αναφοράς κατανομημένα σε κύκλο. Στην προτεινόμενη παραλλαγή των shape context θα χρησιμοποιήσουμε 40 σημεία αναφοράς. Επιπλέον για το τελικό στάδιο θα δοκιμάσουμε την κλασσική μέθοδο των HOG, όπως την περιγράψαμε στο Κεφάλαιο 3, αλλά και την προτεινόμενη παραλλαγή CentralizedHOG του Κεφαλαίου 6, η οποία σύμφωνα με προηγούμενα αποτελέσματα υπερτερεί της κλασσικής μεθόδου για το συγκεκριμένο πρόβλημα.

	affine variation	homogenous variation
hog_9_9	75.1 %	78 %
chog_9_9	77.5 %	81.3 %

Πίνακας 7.7: Ποσοστά επιτυχούς αναγνώρισης της μεθόδου περιγραμμάτων αναφοράς για seen signers

	affine variation	homogenous variation
hog_9_9	56.2 %	58.7 %
chog_9_9	59.1 %	61.8 %

Πίνακας 7.8: Ποσοστά επιτυχούς αναγνώρισης της μεθόδου περιγραμμάτων αναφοράς για *unseen signers*

Αρχικά παρατηρούμε πως η χρήση της περιγραφής *chog_9_9* έναντι της *hog_9_9* προσφέρει μια αξιόλογη ώθηση, όπως και ήταν αναμενόμενο σύμφωνα με προηγούμενα πειράματα. Μεγαλύτερο ενδιαφέρον παρουσιάζει η καθολική υπεροχή της προτεινόμενης παραλλαγής των *shape context* έναντι της κλασσικής τεχνικής και του θεωρητικά πιο περιγραφικού αφινικού μετασχηματισμού. Η ικανότητα του αφινικού ταιριάσματος για προσαρμογή των σχημάτων χωρίς αυστηρούς περιορισμούς, που συντελεί σε μεγάλο βαθμό στην επιτυχία της μεθόδου, φαίνεται να αποτελεί και το μεγαλύτερο μειονέκτημα για το συγκεκριμένα πρόβλημα αναγνώρισης. Συγκεκριμένα η μεγάλη ελευθερία στην μορφή του αφινικού μετασχηματισμού (υπάρχουν περιορισμοί, αλλά όχι ιδιαίτερα αυστηροί) οδηγεί σε υπέρμετρη προσαρμογή της αρχικής εικόνας στο περίγραμμα αναφοράς που εν τέλει αποτελεί τροχοπέδη στην επίδοση του συστήματος. Αυτό οφείλεται στο γεγονός ότι εν τέλει η προκύπτουσα περιγραφή (*hog* ή *chog*) έχει πλησιάσει πολύ τις περιγραφές αναφοράς της εκάστοτε κλάσης, χρησιμοποιώντας έναν διαφορετικό αφινικό μετασχηματισμό για την κάθε μια. Το περιγραφέν φαινόμενο μετριάζεται με την προτεινόμενη παραλλαγή, η οποία ουσιαστικά εστιάζει μόνο στην περιστροφή του περιγράμματος αποφεύγοντας έτσι σε κάποιο βαθμό τις άτοπες τελικές περιγραφές.

7.9.5 Σύγκριση Τεχνικών Περιγράμματος

Έως τώρα έχουμε ασχοληθεί με την παραμετροποίηση των προτεινομένων τεχνικών για περιστροφική ανεξαρτησία. Στην υποενότητα αυτή θα συγκρίνουμε την αποτελεσματικότητα των τεχνικών περιγράμματος για την βάση των περιστρεμμένων εικόνων. Συνολικά εκτός από τις προτεινόμενες τεχνικές που ήδη αναφέραμε (*sc&ftm*, *rbhog&ftm*, *erbhog&ftm*) συμπεριλαμβάνουμε τις τεχνικές *fd&ftm* και *ang&ftm*. Η επίδοση των παραπάνω τεχνικών συνοψίζεται στο Σχήμα 7.45 παρουσιάζοντας το ποσοστό εσφαλμένων ταξινομήσεων. Για λόγους πληρότητας στον Πίνακα 7.9 παρουσιάζουμε την ικανότητα αναγνώρισης της τεχνική *hu*, η οποία είναι ακατάλληλη για προβλήματα με τόση μεγάλη ποικιλομορφία περιγράμματος καθώς τα ελάχιστα χαρακτηριστικά (συνολικά 7) που εξάγει η μέθοδος δεν είναι ικανά να προσφέρουν την ποιότητα περιγραφής που επιθυμούμε.

seen signers		
eucl	knn	naive bayes
15.6 %	38.4 %	26.1 %

unseen signers		
eucl	knn	naive bayes
12.7 %	23.2 %	18.2 %

Πίνακας 7.9: Ποσοστά επιτυχούς αναγνώρισης της μεθόδου *hu*

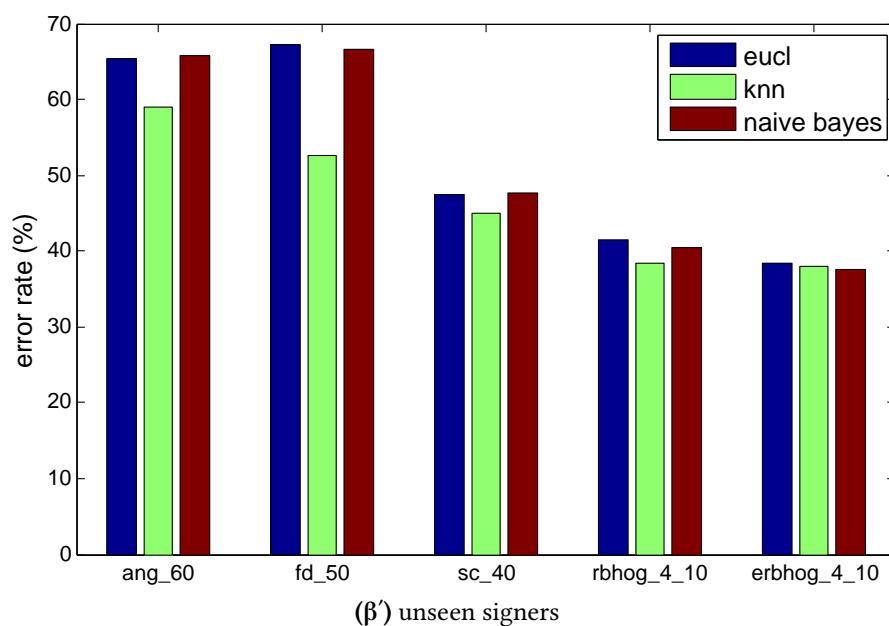
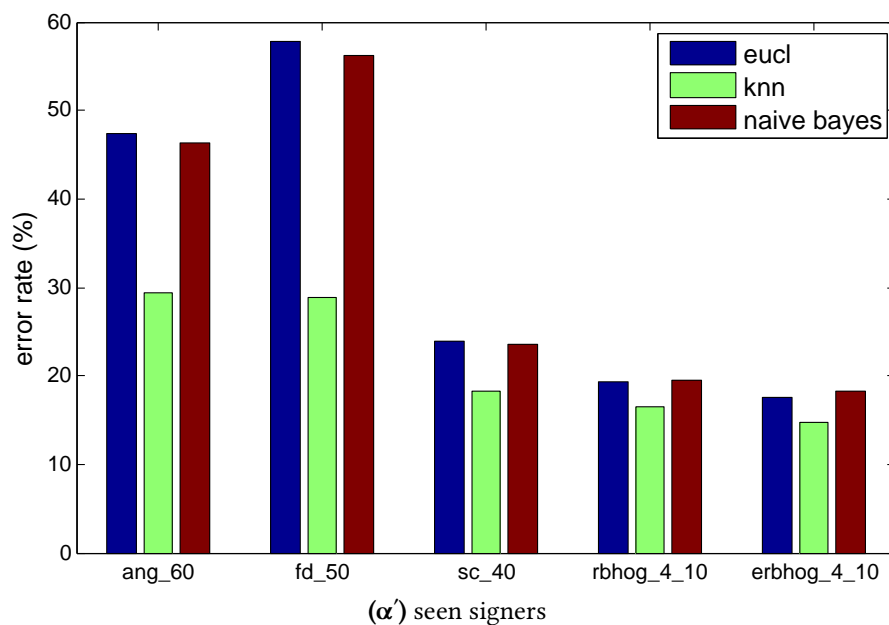
Συμπερασματικά οι προτεινόμενες παραλλαγές των HOG για δυαδικές εικόνες υπερτερεί συνολικά προσφέροντας το μικρότερο σφάλμα αναγνώρισης σε όλους τους ταξινομητές με την *rbhog* να ξεχωρίζει και ακολουθεί η προτεινόμενη παραλλαγή των *shape context*. Οι υπόλοιπες δυο μέθοδοι υστερούν κατά πολύ (>10% σφάλμα για *knn*) ενώ, όπως αναφέρθηκε νωρίτερα αλλά είναι και εμφανές στο Σχήμα 7.45, οι ταξινομητές απόστασης (συμπεριλαμβανομένου του *naive bayes*) δεν είναι κατάλληλοι για τις προκύπτουσες περιγραφές.

Για την περαιτέρω αξιολόγηση των προτεινόμενων τεχνικών παρουσιάζουμε την αποτελεσματικότητά τους ως προς την εύρεση της γωνίας περιστροφής στο Σχήμα 7.47 και τους πίνακες σύγχυσης (*confusion matrices*) στο Σχήμα 7.46¹. Η αποτελεσματικότητα ως προς την ικανότητα αναγνώρισης της γωνίας περιστροφής αναπαριστάται ως ένα ιστόγραμμα του ποσοστού των αποκλίσεων ανά τις γωνίες απόκλισης. Έχοντας κατηγοριοποιήσει μια εικόνα, η εκτιμώμενη γωνία προκύπτει με την εφαρμογή κάποιας παραλλαγής της τεχνικής της εττεροσυσχέτισης (Παράρτημα Γ, Κεφάλαιο 6) στις περιγραφές πριν την εφαρμογή του *ftm* (συγκεκριμένα η περιγραφή της ως προς ταξινόμηση εικόνας και η μέση περιγραφή της προβλεφθείσας κλάσης). Τα σχήματα αυτά έχουν προκύψει για ευκλείδειο ταξινομητή και εκτίμηση αναγνώρισης για *unseen signers*.

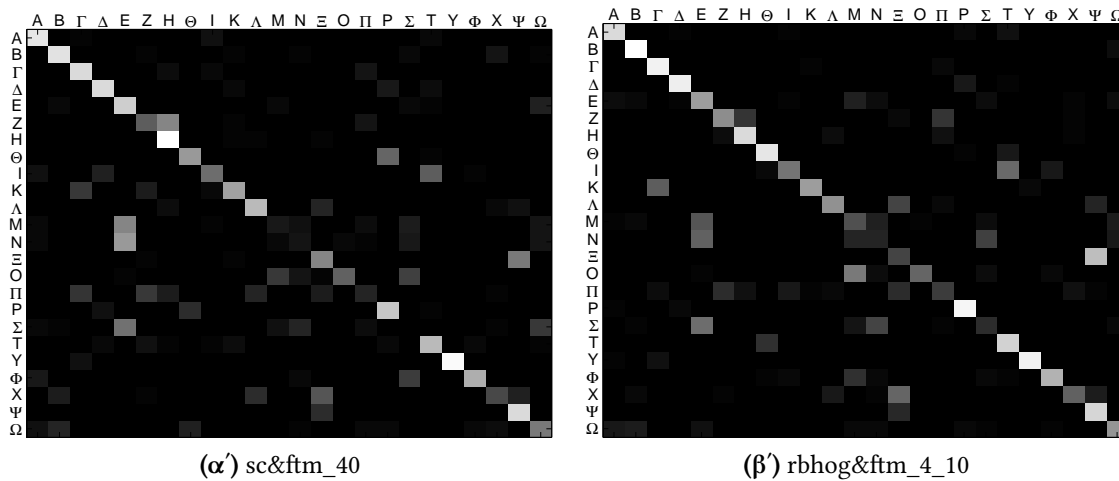
Οι πίνακες του Σχήματος 7.46 προσφέρουν μια πιο αναλυτική παρουσίαση της ικανότητας ταξινόμησης των τεχνικών, από όπου είναι εμφανές ότι τα *rbhog* μειώνουν την συχνότητα σύγχυσης σε πληθώρα ζεύγους γραμμάτων. Επιπρόσθετα τα ιστογράμματα των αποκλίσεων του Σχήματος 7.47 υποδεικνύουν την ικανότητα των τεχνικών αυτών για μια αξιόπιστη εκτίμηση της γωνίας περιστροφής, ενώ τα *rbhog* εμφανίζονται ποιοτικά καλύτερα σε κάποιο βαθμό (μεγαλύτερο ποσοστό αποτελεσμάτων γύρω από την μηδενική απόκλιση).

Εν κατακλείδι θεωρούμε ότι η τεχνική που βασίζεται στα HOG αποτελεί την καταλληλότερη επιλογή για το πρόβλημα, καθώς υπερτερεί σε επιδόσεις τόσο στην αναγνώριση όσο και στην εκτίμηση της γωνίας περιστροφής. Επιπλέον η επιλογή αυτή ενισχύεται από το γεγονός ότι τα *rbhog* είναι αρκετά μικρότερες περιγραφές σε μέγεθος από τα *sc*, γεγονός το οποίο προσφέρει υπολογιστική ταχύτητα.

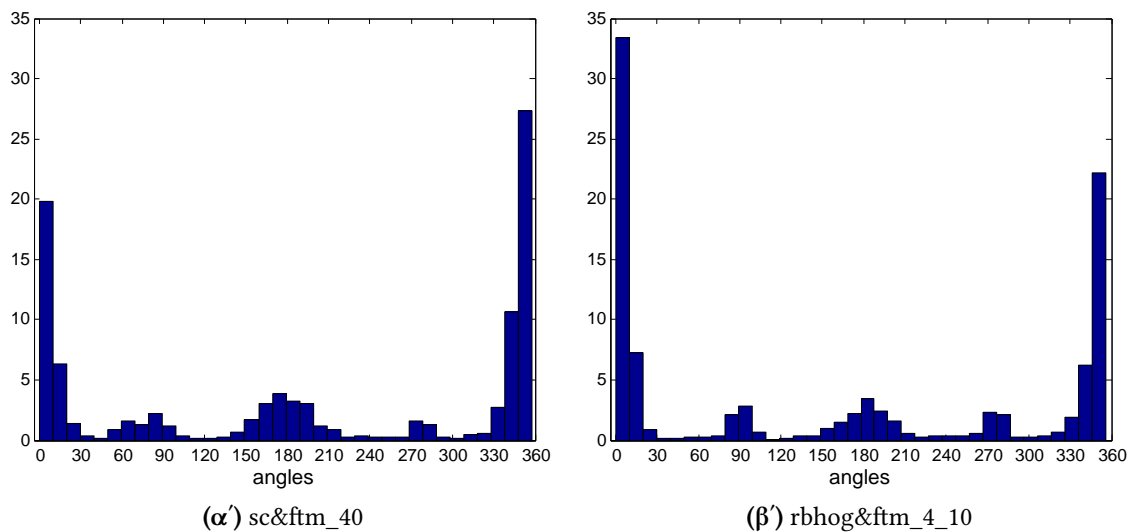
¹Συγκρίνουμε τις περιγραφές *sc&fmt* και *rbhog&fmt* λόγω της κοινής κυκλικής δομής τους



Σχήμα 7.45: Σύγκριση τεχνικών περιγράμματος ανεξάρτητων σε περιστροφές



Σχήμα 7.46: Confusion Matrices για τις τεχνικές περιγράμματος



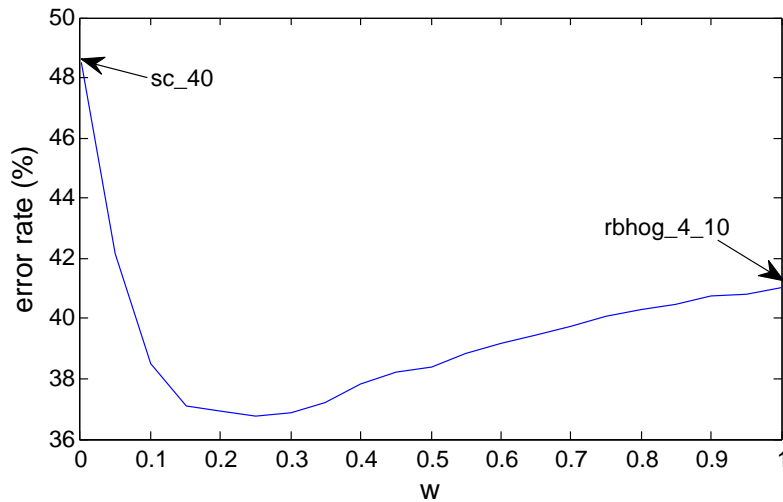
Σχήμα 7.47: Ιστογράμματα απόκλισης της εκτιμώμενης γωνίας για τις τεχνικές περιγράμματος

Στο σημείο αυτό αξίζει να τονίσουμε πως οι τεχνικές shape context και rbhog προσεγγίζουν το ίδιο πρόβλημα από διαφορετική οπτική γωνία, η μεν θεωρώντας κάποια χοντρική περιγραφή του ολικού περιγράμματος ως προς σημεία αναφοράς και η δε θεωρώντας πυκνά κατανομημένες τοπικές περιγραφές του περιγράμματος. Με αλλά λόγια τα shape context έχουν μια γενικότερη εποπτεία του περιγράμματος ενώ τα rbhog εστιάζουν σε ένα σύνολο από μικρές περιοχές. Η βασική αυτή διαφορά μας ώθησε στην αξιολόγηση του συνδυασμού των δυο τεχνικών ως προς την ικανότητα αναγνώρισης. Για τον συνδυασμό των τεχνικών αυτών χρησιμοποιήσαμε έναν συντελεστή βαρύτητας $w \in [0, 1]$ και για τον ευκλείδειο ταξινομητή ορίσαμε την νέα απόσταση ως

$$cd(I, \omega_i) = (1 - w)d(h1, \omega_i) + wd(h2, \omega_i)$$

όπου $cd(I, \omega_i)$ η συνδυασμένη απόσταση της εικόνας εισόδου για την κλάση ω_i και $d(h, \omega_i)$ η απόσταση της περιγραφής h από το κέντρο της κλάσης ω_i για τις αντίστοιχες περιγραφές. Για το πείραμα μας η περιγραφή h_1 αντιστοιχεί στην παραλλαγή των shape context και η h_2 στην παραλλαγή των hog. Ο παραπάνω ταξινομητής είναι ισοδύναμος με τον ακόλουθο (για μια πιο αυστηρή διατύπωση με πιθανότητες):

$$\arg \min_{i=1:24} \{(1 - w) \log P(h_1 | \omega_i) + w \log P(h_2 | \omega_i)\}$$



Σχήμα 7.48: Συνδιασμός Τεχνικών Περιγράμματος (eucl,unseen signers)

Τα πειραματικά αποτελέσματα του συνδυασμού των δυο τεχνικών παρουσιάζονται στο Σχήμα 7.48, όπου παρατηρούμε ότι αυξάνεται αξιόλογα η ικανότητα ταξινόμησης. Μάλιστα το σφάλμα στην αναγνώριση μειώνεται μέχρι και 4% (σε σχέση με τα rbhog που είναι η αποτελεσματικότερη τεχνική από τις δύο) για τιμές της παραμέτρου w κοντά στο 0.3.

7.9.6 Σύγκριση Τεχνικών Εμφάνισης

Έχοντας αξιολογήσει τις τεχνικές περιγράμματος ακολουθούν οι τεχνικές εμφάνισης, στις οποίες συγκαταλέγουμε τις τεχνικές RHOG, ERHOG και SRef (όλες έχουν ως

βασικό δομικό στοιχείο την απλή HOG μέθοδο). Τα ποσοστά επιτυχούς αναγνώρισης των τεχνικών αυτών συνοψίζονται στον Πίνακα 7.10. Στον ακόλουθο πίνακα παραθέτουμε μόνο τα αποτελέσματα του ευκλείδειου ταξινομητή για λόγους συμβατότητας και σύγκρισης αφού είναι ο βασικός ταξινομητής για την μέθοδο περιγραμμάτων αναφοράς (SRef).

RHOG	ERHOG	SRef
68.6 %	74.5 %	61.8 %

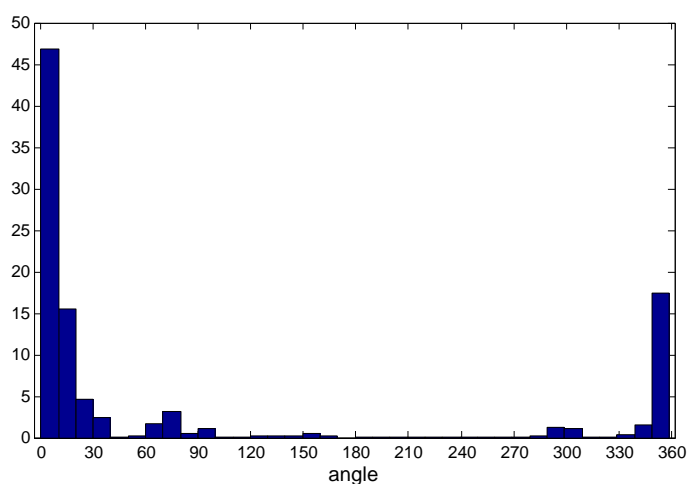
Πίνακας 7.10: Ποσοστά επιτυχούς αναγνώρισης για τις τεχνικές RHOG, ERHOG και SRef (eucl,unseen signer)

Για την περαιτέρω αξιολόγηση των προτεινόμενων τεχνικών παρουσιάζουμε την αποτελεσματικότητά τους ως προς την εύρεση της γωνίας περιστροφής στο Σχήμα 7.49.

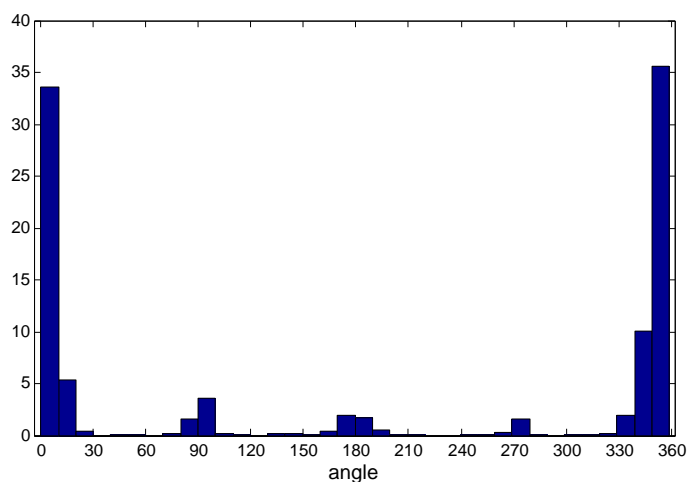
Συνολικά είναι εμφανής η υπεροχή της ERHOG μεθόδου τόσο για τον σκοπό της αναγνώρισης χειρομορφών, όσο και για την εύρεση της γωνίας περιστροφής τους. Η μέθοδος RHOG παρουσιάζεται εξίσου αποτελεσματική για τον σκοπό της εύρεσης της γωνίας περιστροφής αν και υστερεί ως προς την ικανότητα αναγνώρισης. Στον αντίποδα η μέθοδος SRef είναι εξαιρετικά ασταθής και για τα δυο προβλήματα και συνεπώς κρίνεται ακατάλληλη, ειδικά αν συνυπολογίσει κανείς ότι η μέθοδος αυτή είναι αρκετά αργή. Συγκεκριμένα έχει πολύ μεγάλο υπολογιστικό κόστος σε σχέση με τις υπόλοιπες, καθώς συγκρίνει τα περιγράμματα και δημιουργεί μια νέα περιγραφή για κάθε μια από τις κλάσεις. Μάλιστα και η ίδια η χρήση των shape context είναι λιγότερο αποδοτική σε σχέση με τα rhog, erhog.

Ο λόγος της άσχημης επίδοσης της SRef μεθόδου είναι ουσιαστικά ο ίδιος που αναφέραμε για να ερμηνεύσουμε ότι η χρήση του αφινικού μετασχηματισμού υστερεί του μετασχηματισμού ομογένειας. Γενικότερα στην πράξη αποδεικνύεται ότι η προσέγγιση του προβλήματος από την μέθοδο αυτή είναι αναποτελεσματική, δηλαδή βρίσκοντας πρώτα μια εκτίμηση για την περιστροφή και έπειτα βασιζόμενοι «τυφλά» σε αυτή υπολογίζουμε την τελική περιγραφή της εικόνας. Αντίθετα οι μέθοδοι RHOG και ERHOG εκμεταλλεύονται στην ουσία μια αντίστροφη στρατηγική με ιδιαίτερη επιτυχία. Αναλυτικότερα, εκμεταλλεύονται αρχικά όσο το δυνατόν περισσότερη πληροφορία για την δημιουργία πλήρως περιστροφικά ανεξάρτητων περιγραφών για τον σκοπό της ταξινόμησης τους και αφότου ανατεθούν σε κάποια από τις κλάσεις γίνεται ο υπολογισμός της γωνίας περιστροφής.

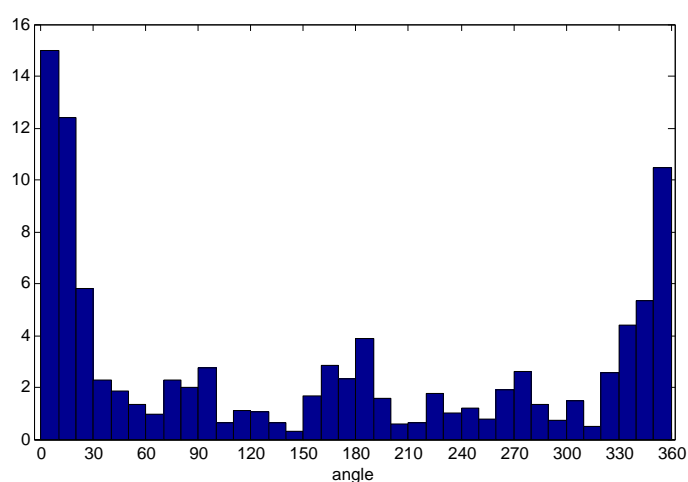
Αξιοσημείωτο είναι το γεγονός ότι το ποσοστό της επιτυχούς αναγνώρισης για την μέθοδο των περιγραμμάτων αναφοράς είναι σχεδόν το ίδιο (60%) με αυτό της rhog μεθόδου που χρησιμοποιεί μόνο την δυαδική μάσκα της αρχικής εικόνας. Η παραπάνω παρατήρηση υποδεικνύει από μια άλλη σκοπιά πόσο αρνητικά δρα το φαινόμενο της έντονης προσαρμογής των εικόνων.



(α') rhog&ftm_8_6_8



(β') erhog&ftm_8_6_8



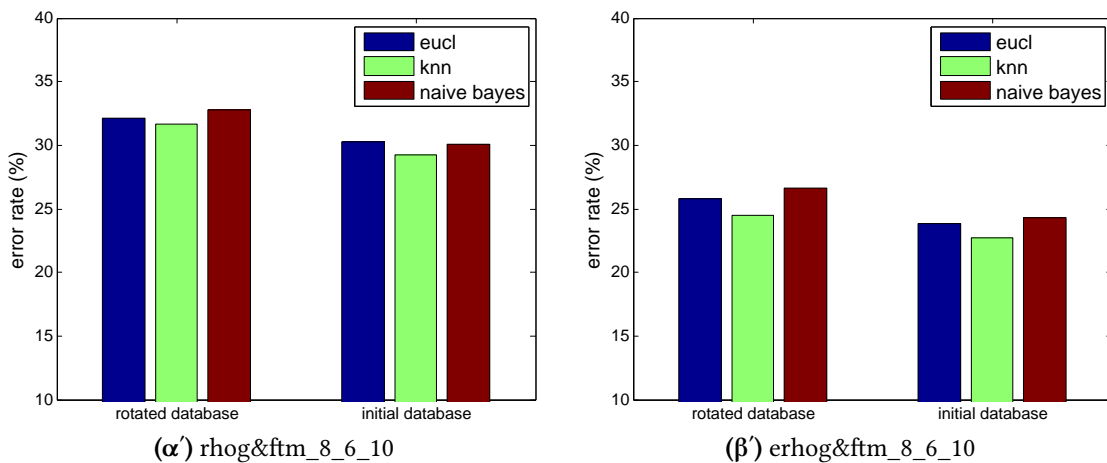
(γ') SRef: sc_40+chog

Σχήμα 7.49: Ιστογράμματα απόκλισης της εκτιμώμενης γωνίας για τις τεχνικές RHOG, ERHOG και SRef (eucl,unseen signer)

7.9.7 Αποτελεσματικότητα Αναγνώρισης - Συμπεράσματα

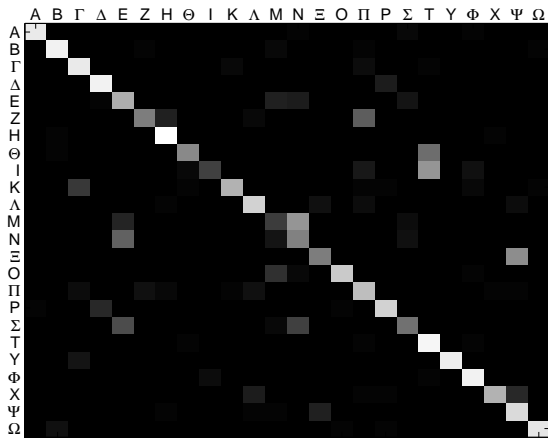
Αφού μελετήσαμε διεξοδικά το πρόβλημα της ανεξαρτησίας σε περιστροφές, αξιολογώντας πληθώρα τεχνικών είναι λογικό να παραθέσουμε μια συνοπτική σύγκριση - ερμηνεία των αποτελεσμάτων για τις ανεξάρτητες στην περιστροφή περιγραφές σε σχέση με τις απλές της προηγούμενης ενότητας. Εκ πρώτης όψεως η διαφορά του σφάλματος αναγνώρισης φαίνεται αρκετά μεγάλη (18% για rhog και 11% για erhog) και εσφαλμένα θα μπορούσε να υποθέσει κάποιος ότι η διαφορά αυτή οφείλεται κυρίως στην ανικανότητα των προτεινόμενων περιγραφών να επιτύχουν πραγματικά αξιολογικό βαθμό ανεξαρτησίας και δευτερευόντως στο πολύ μεγαλύτερο πλήθος της βάσης για τις περιστροφικά ανεξάρτητα περιγραφές. Εντούτοις το μεγαλύτερο μέρος της διαφοράς οφείλεται στην ίδια την ιδέα της ολικής περιστροφικής ανεξαρτησίας, καθώς στο συγκεκριμένο πρόβλημα κάποιες κλάσεις είναι πολύ όμοιες χωρίς τον περιορισμό της κατεύθυνσης του χεριού (π.χ. Z και H, Ξ και Ψ).

Παραδείγματος χάριν για την μέθοδο RHOG και την ERHOG τα ποσοστά εσφαλμένης αναγνώρισης παρουσιάζονται στο Σχήμα 7.50 σε σύγκριση με την εφαρμογή τους στην απλή βάση δίχως περιστροφές. Παρατηρούμε ότι, όπως είναι φυσικό, το ποσοστό αναγνώρισης μειώνεται στην τεχνητή βάση. Εντούτοις η μείωση αυτή είναι ιδιαίτερα μικρή (3%) έως αμελητέα αν αναλογιστούμε ότι το σύνολο δοκιμών έχει εικοσαπλασιαστεί. Συνεπώς μπορούμε να αποφανθούμε ότι η προτεινόμενη rhog μέθοδος (όπως και οι παραλλαγές της erhog,rbhog) είναι ιδιαίτερα αποτελεσματική και προσφέρει περιγραφές πραγματικά ανεξάρτητες στην περιστροφή της εικόνας.

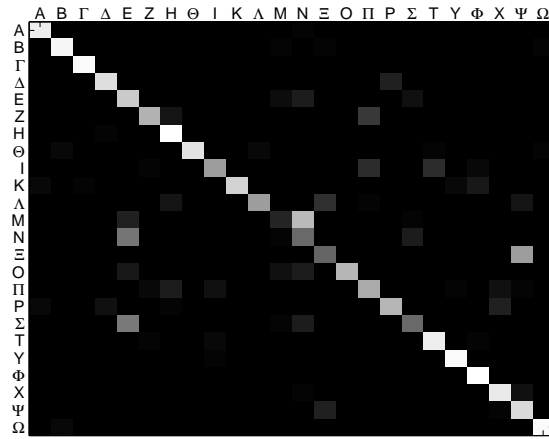


Σχήμα 7.50: Αποτελεσματικότητα των RHOG, ERHOG στην περιστροφή

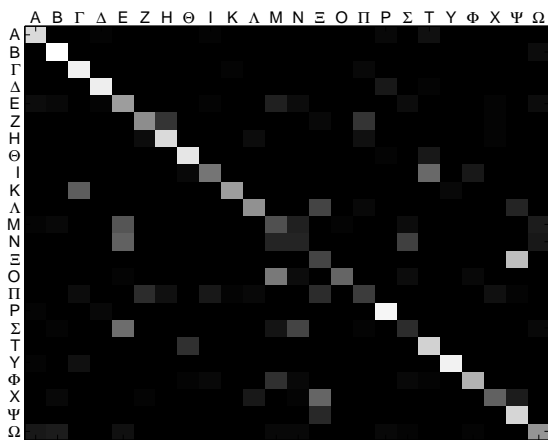
Τέλος, στο Σχήμα 7.51 παραθέτουμε τους πίνακες σύγχυσης για τις τεχνικές RHOG, ERHOG, RBHOG και ERBHOG, έτσι ώστε να γίνει εμφανής η ανωτερότητα των δυο πρώτων στο επίπεδο του διαχωρισμού των χειρομορφών. Όπως ήταν αναμενόμενο φαίνεται να βελτιώνονται αισθητά εξαρτήσεις όπως M και N, Z και H, Ξ και Ψ που έχουν παρόμοιο περίγραμμα (για περιστροφική ανεξαρτησία) και συνεπώς δεν είναι δυνατόν να περιγραφούν ικανοποιητικά με τεχνικές περιγράμματος.



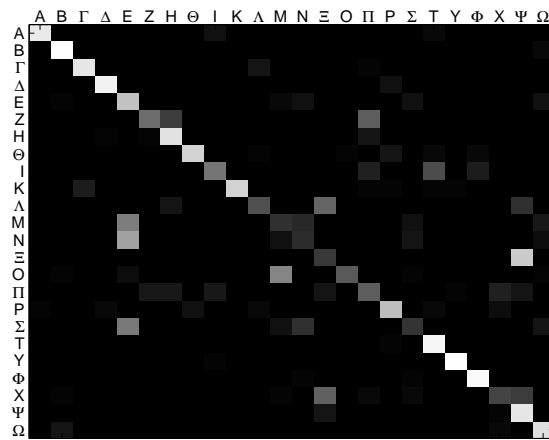
(α') rhog&ftm_8_6_10



(β') erhog&ftm_8_6_10



(γ') rbhog&ftm_4_10



(δ') erbhog&ftm_4_10

Σχήμα 7.51: Confusion Matrices για τις τεχνικές RHOG, ERHOG, RBHOG και ERBHOG

Κεφάλαιο 8

Συμπεράσματα και Κατευθύνσεις για Μελλοντική Έρευνα

Στο κεφάλαιο αυτό θα παρουσιάσουμε συνοπτικά τις μεθόδους που αναπτύξαμε στο πλαίσιο του προβλήματος αναγνώρισης στατικών χειρομορφών και τα συμπεράσματα για την εφαρμοσιμότητα και την αποτελεσματικότητα των μεθόδων σύμφωνα με τα πειραματικά αποτελέσματα. Στην συνέχεια θα παρουσιάσουμε πιθανές κατευθύνσεις για μελλοντική επέκταση της παρούσας ερευνάς.

8.1 Ανακεφαλαίωση και Συμπεράσματα

Στα πλαίσια της μελέτης αυτής έγιναν ποικίλες προσεγγίσεις του προβλήματος της κατηγοριοποίησης στατικών χειρομορφών που απεικονίζουν τα 24 γράμματα της ελληνικής αλφαβήτου. Οι προσεγγίσεις αυτές έχουν ως βασικό δομικό στοιχείο για το στάδιο της εξαγωγής χαρακτηριστικών την δημιουργία ιστογραμμάτων κατεύθυνσης αντιστοιχία με την HOG μέθοδο. Οι τεχνικές εξαγωγής χαρακτηριστικών που υλοποιήθηκαν συνοψίζονται παρακάτω

- Μεταβολή της χωρικής κατανομής των τοπικών ιστογραμμάτων, αξιοποιώντας βασικές πληροφορίες για το σχήμα (κέντρο, ροπές). Οι περιγραφές που προέκυψαν είναι οι CentHOG (προσαρμογή των κελιών έτσι ώστε το κεντρικό να συμπίπτει με το κέντρο της χειρομορφής), RHOG (κυκλική διαμέριση γύρω από το κέντρο) και ERHOG (ελλειπτική διαμέριση).
- Απλοποίηση των μεθόδων που βασίζονται στα HOG για την χρήση τους σε δυαδικές εικόνες (BHOG, CBHOG, RBHOG, ERBHOG), προσφέροντας έτσι υπολογιστική επιτάχυνση. Επιπλέον για την περιγραφή του περιγράμματος εισάγουμε μια παραλλαγή των shape context.

- Δημιουργία δυο κατηγοριών περιστροφικά ανεξάρτητων περιγραφών κάνοντας χρήση μια από τις ακόλουθες στρατηγικές:
 - 1 Για κάθε κλάση εκτιμούμε τον μετασχηματισμό (αφινικό ή ομογένειας) και κατ'έκταση την γωνία περιστροφής βασιζόμενοι στο περίγραμμα της χειρομορφής σε σύγκριση με ένα περίγραμμα αναφοράς. Κατόπιν αφού μετασχηματίσουμε την εικόνα εισόδου έτσι ώστε να προσαρμοστεί στο περίγραμμα αναφοράς χρησιμοποιούμε κάποια HOG τεχνική (π.χ. hog ή chog).
 - 2 Κατασκευάζουμε πλήρως περιστροφικά ανεξάρτητες περιγραφές αξιοποιώντας κυκλικές (ή ελλειπτικές) κατανομές τοπικών περιγραφών και την τεχνική Fourier Magnitude.

Ως προς την κατηγοριοποίηση - αναγνώριση αρχικά μειώνουμε τις διαστάσεις των διανυσμάτων των περιγραφών αξιοποιώντας τις τεχνικές PCA και LDA, έτσι ώστε να αποκτήσουμε πιο εύκολα διαχειρίσιμες περιγραφές, ικανές να μοντελοποιηθούν αποτελεσματικά από το σύνολο εκπαίδευσης. Στην συνέχεια εφαρμόζουμε έναν από τους ακόλουθους ταξινομητές για την τελική αξιολόγηση ως προς την ικανότητα αναγνώρισης της εκάστοτε μεθόδου

- Ταξινομητές Απόστασης. Η απόσταση μπορεί να είναι είτε η mahalanobis, εφόσον έχουμε μοντελοποιήσει κάθε κλάση με μια γκαουσιανή κατανομή. Απλούστερα μπορούμε να χρησιμοποιήσουμε τον naive bayes ταξινομητή ή τον ευκλείδειο ταξινομητή που αποτελούν απλοποιήσεις του mahalanobis κάνοντας (αυθαίρετες) υποθέσεις για την μορφή της γκαουσιανής κατανομής.
- k-Nearest Neighbors (k-NN) ταξινομητής. Ο ταξινομητής αυτός είναι αργός αλλά συνήθως εμφανίζει πολύ καλά αποτελέσματα και για αυτό χρησιμοποιείται ως μέτρο σύγκρισης

Ακολουθώς συνοψίζουμε τα σημαντικότερα συμπεράσματα ως προς την αποτελεσματικότητα των τεχνικών που εφαρμόστηκαν, σύμφωνα με τα πειραματικά αποτελέσματα του προηγούμενου κεφαλαίου.

- Οι προτεινόμενες παραλλαγές των HOG εμφανίζουν πολύ υψηλά ποσοστά αναγνώρισης, με τις τεχνικές CHOG και ERHOG να ξεχωρίζουν.
- Οι δυαδική παραλλαγή των HOG τεχνικών αποδεικνύεται εξίσου αποτελεσματική, καθώς υπερτερεί κλασικών τεχνικών περιγράμματος, όπως Fourier Descriptors.
- Συνολικά οι τεχνικές περιγράμματος δεν είναι ικανές να ανταποκριθούν πλήρως για το συγκεκριμένο πρόβλημα καθώς ορισμένες κλάσεις δεν είναι διαχωρίσιμες μόνο με το περίγραμμα
- Η μέθοδος περιγραμμάτων αναφοράς (SRef) αποδεικνύεται ακατάλληλη και ιδιαίτερα ευαίσθητη λόγω της ελεύθερης προσαρμογής στα περιγράμματα αναφοράς.

- Οι τεχνικές πλήρους περιστροφικής ανεξαρτησίας (rhog,erhog και rbhog) είναι ιδιαίτερα αποτελεσματικές τόσο στην αναγνώριση περιστρεφμένων εικόνων όσο και στην εκτίμηση της περιστροφής. Συγκεκριμένα ξεχωρίζει η μέθοδος ERHOG, καθώς παρουσιάζει πολύ υψηλά ποσοστά επιτυχούς αναγνώρισης δεδομένου της δυσκολίας του προβλήματος.

Συμπερασματικά επιτυγχάνουμε την δημιουργία εξαιρετικά αποτελεσματικών περιγραφών οι οποίες υπερτερούν της απλής HOG μεθόδου για το πρόβλημα της κατηγοριοποίησης χειρομορφών. Επιπλέον μελετήσαμε σε βάθος το πρόβλημα της περιστροφικής ανεξαρτησίας των εικόνων, που αποτελεί και ένα από τα βασικά μειονεκτήματα της HOG μεθόδου, και τελικά δώσαμε λύση σε αυτή προτείνοντας ένα σύνολο περιγραφών με ιδιαίτερη επιτυχία.

Τέλος δώσαμε ιδιαίτερο βάρος στην προσπάθεια μοντελοποίησης των προτεινόμενων τεχνικών και γενικά των τεχνικών που βασίζονται σε ιστογράμματα κατεύθυνσης, πάντα κάτω από το πρίσμα του προβλήματος του fingerspelling. Αποδείξαμε λοιπόν ότι για το συγκεκριμένο πρόβλημα οι περιγραφές ιστογραμμάτων κατεύθυνσης κάθε κλάσης είναι ικανές να μοντελοποιηθούν από μια γκαουσιανή κατανομή. Βέβαια το πλήθος των διαστάσεων των περιγραφών κάνει απαγορευτικό τον απευθείας υπολογισμό των παραμέτρων των γκαουσιανών. Για τον λόγο αυτό θα βασιστούμε σε κάποια τεχνική μείωσης διαστάσεων και συγκεκριμένα στην τεχνική LDA, η οποία υπερτερεί κατά πολύ της PCA στο συγκεκριμένο πρόβλημα (αν και στην πράξη χρησιμοποιείται ο συνδυασμός τους). Εν τέλει η χρήση του LDA φαίνεται να διαχωρίζει αρκετά καλά τις υπάρχουσες κλάσεις σε βαθμό που μπορούν να κατηγοριοποιηθούν με επιτυχία από έναν ευκλείδειο ταξινομητή.

8.2 Μελλοντική Έρευνα

Επιλογικά αναφέρουμε εν συντομία πιθανές κατευθύνσεις για μελλοντική έρευνα ως επέκταση της παρούσας διπλωματικής. Ξεχωρίζουμε λοιπόν δυο σημαντικές διαφορετικές κατευθύνσεις ως λογική συνέχεια της παρούσας μελέτης. Η πρώτη είναι η εφαρμογή των προτεινόμενων τεχνικών σε διαφορετικά προβλήματα (object recognition) και η δεύτερη είναι η επέκταση των τεχνικών αυτών σε προβλήματα αναγνώρισης χειρονομιών, δηλαδή σε βίντεο και όχι στατικές εικόνες.

8.2.1 Αναγνώριση Αντικειμένων

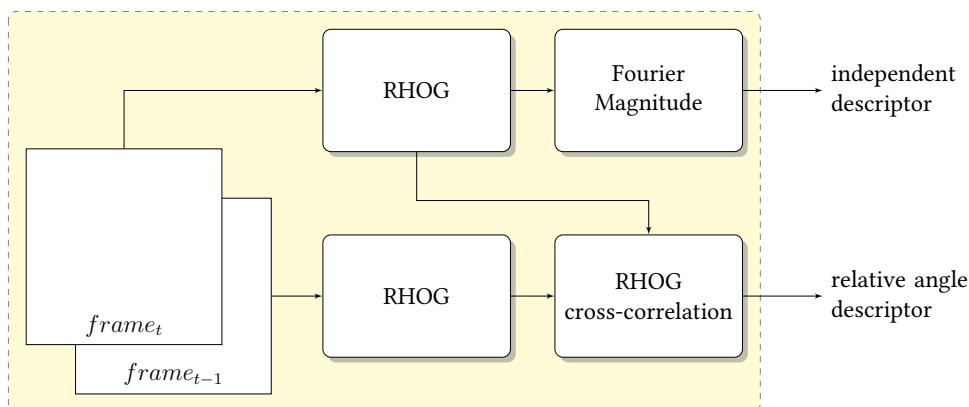
Οι τεχνικές που προτείναμε για περιστροφική ανεξαρτησία λειτουργούν ικανοποιητικά για το παρών πρόβλημα, όπου έχουμε αφαιρέσει το background. Ένα βήμα περαιτέρω για τη ν αξιολόγηση τους θα ήταν η εφαρμογή τους σε ένα γενικό πρόβλημα

αναγνώρισης αντικειμένων σε μια εικόνα, πρόβλημα στο οποίο τεχνικές σημείων ενδιαφέροντος έχουν ιδιαίτερη επιτυχία. Ένα βασικό εμπόδιο για την γενίκευση στο πρόβλημα αναγνώρισης αντικειμένων είναι η αξιοποίηση πληροφορίας περιγράμματος από πολλές από τις προταθείσες τεχνικές. Βέβαια η RHOG μέθοδος είναι εύκολα γενικεύσιμη και είναι δυνατόν να χρησιμοποιηθεί ως ένα κινούμενο κυκλικό παράθυρο κατά αντιστοιχία με την χρήση των HOG στο [8].

8.2.2 Αναγνώριση Χειρονομιών

Το πρόβλημα του fingerspelling είναι μόνο ένα μικρό μέρος του γενικότερου προβλήματος της αναγνώρισης της Νοηματικής Γλώσσας, όπως αναφέρθηκε και στην Εισαγωγή. Μεγαλύτερο ενδιαφέρον αλλά και περισσότερα εμπόδια παρουσιάζει η αναγνώριση χειρονομιών, δηλαδή ακολουθίες χειρομορφών.

Μια προτεινόμενη αξιολόγηση των τεχνικών που παρουσιάστηκαν απεικονίζεται στο Σχήμα 8.1, όπου ουσιαστικά η πληροφορία του κάθε frame διαχωρίζεται σε πληροφορία περιστροφής και εσωτερική πληροφορία εμφάνισης (ανεξάρτητη της περιστροφής), ουσιαστικά δηλαδή εμπλέκεται η εμφάνιση της χειρομορφής από την περιστροφή της. Η περιγραφείσα τεχνική θεωρητικά ενισχύει την δυνατότητα χρήσης hog περιγραφών σε video, καθώς η ευαισθησία τους σε αλλαγές στην περιστροφή δημιουργούσε ένα πολύ μεγάλο πλήθος διαφορετικών περιγραφών και συνεπώς ήταν αδύνατη μια αποτελεσματική χρήση τους με κάποιο μοντέλο όπως τα HMM (απαιτεί μοντελοποίηση των περιγραφών σε κάθε state).



Σχήμα 8.1: Προτεινόμενο σύστημα εξαγωγής χαρακτηριστικών για video

Βέβαια ένα από τα βασικά συστατικά της επιτυχίας των προτεινόμενων μεθόδων στην παρούσα μελέτη ήταν η χρήση της τεχνικής LDA. Κάτι τέτοιο δεν είναι άμεσα εφαρμόσιμο στην περίπτωση του video. Παρολαυτά μια διαφορετική προσέγγιση θα ήταν να επιλέξουμε κάποιες χειρομορφές ως βασικές για την δημιουργία οποιασδήποτε χειρονομίας. Μια τέτοια προσέγγιση θα ήταν αδύνατη να υλοποιηθεί αποδοτικά χωρίς την δυνατότητα περιστροφικής ανεξαρτησίας εξαιτίας της πολύ μεγάλης ποικιλομορφίας της πόζας του χεριού στον χώρο.

Παράρτημα Α΄

Hu Invariant Moments

Οι απλές ροπές (raw moments) ορίζονται ως:

$$m_{ij} = \sum_x \sum_y x^i y^j s(x, y), \quad i + j = n$$

Οι κεντρικές ροπές (central moments) ορίζονται ως:

$$\mu_{ij} = \sum_x \sum_y (x - \bar{x})^i (y - \bar{y})^j s(x, y), \quad i + j = n$$

,όπου \bar{x}, \bar{y} οι συντεταγμένες του κέντρου βάρους (center of mass) και υπολογίζονται με τις σχέσεις:

$$\bar{x} = \frac{m_{10}}{m_{00}} \quad \bar{y} = \frac{m_{01}}{m_{00}}$$

Οι κανονικοποιημένες κεντρικές ροπές ορίζονται ως :

$$\eta_{ij} = \frac{\mu_{ij}}{\mu_{00}^p}, \quad p = 1 + \frac{i + j}{2}, \quad i + j \geq 2$$

Οι ροπές που εισήγαγε ο Hu ως σύνολο περιστροφικά ανεξάρτητων χαρακτηριστικών είναι οι εξής:

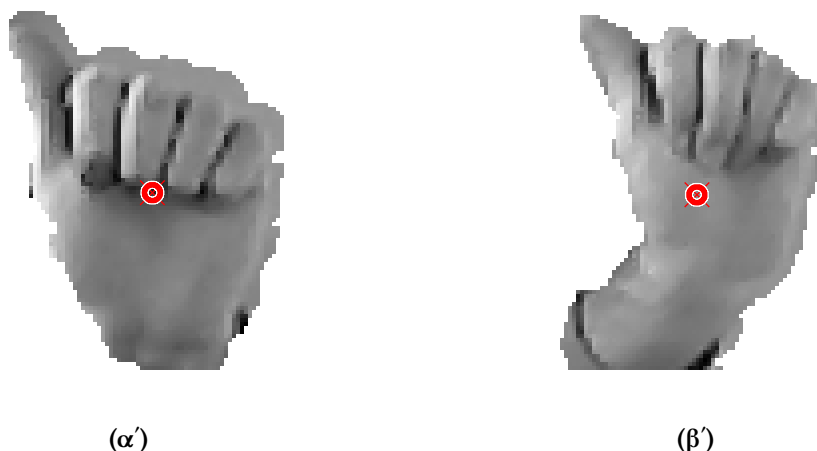
$$\begin{aligned} \phi_1 &= \eta_{20} + \eta_{02} \\ \phi_2 &= (\eta_{20} - \eta_{02})^2 + 4\eta_{11}^2 \\ \phi_3 &= (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2 \\ \phi_4 &= (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2 \\ \phi_5 &= (\eta_{30} - 3\eta_{12})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] \\ &\quad + (3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] \\ \phi_6 &= (\eta_{20} - \eta_{02})[(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] \\ &\quad + 4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{21} + \eta_{03}) \\ \phi_7 &= (3\eta_{21} - \eta_{03})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] \\ &\quad - (\eta_{30} - 3\eta_{12})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] \end{aligned}$$

Παράρτημα Β΄

Επιλογή Κέντρου Χειρομορφής

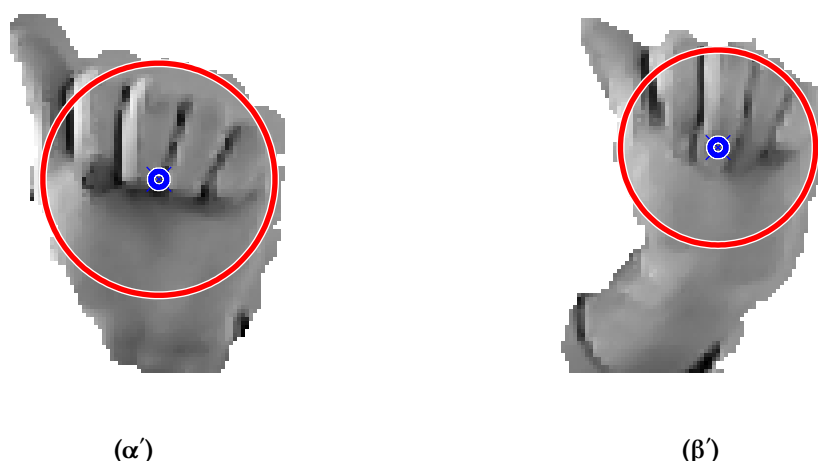
Μια καλή εκτίμηση για το κέντρο της εικόνας είναι ζωτικής σημασίας για την κατανόηση της χωρικής κατανομής του χεριού και συνεπώς για την κατασκευή ενός δυνάμματος χαρακτηριστικών ανεξάρτητο στην κυκλική περιστροφή. Μια απλοϊκή λύση αποτελεί ο υπολογισμός του κέντρου ως το κέντρο βάρους (center of mass) της δυαδικής μάσκας της εικόνας. Μια τέτοια προσέγγιση οδηγεί σε άσχημα αποτελέσματα και περιορίζεται η ιδιότητα της γενίκευσης και της ευστάθειας όσο αφορά την εύρεση ενός «καθολικού» κέντρου ανεξαρτήτως μικρών παραμορφώσεων για την ίδια κλάση χειρομορφών. Τα παραπάνω αποτελούν λογικό επακόλουθο εφόσον το προκύπτον κέντρο εξαρτάται σε πολύ μεγάλο βαθμό από την «έκταση» της χειρομορφής και όχι από την μορφή του περιγράμματος. Για τον λόγο αυτό η θέση του κέντρου επηρεάζεται σημαντικά από την εμφάνιση του καρπού, δηλαδή η απουσία ή η παρουσία τμήματος του καρπού μεταβάλλουν αξιοσημείωτα την θέση του κέντρου. Η προαναφερθείσα εξάρτηση από τον καρπό φαίνεται παραστατικά στο σχήμα Β΄.1, όπου παρουσιάζονται δυο χειρομορφές του γράμματος 'α', η μια σχεδόν χωρίς καθόλου καρπό και η άλλη με ένα μεγάλο μέρος του καρπού και μάλιστα υπό γωνία, και τα υπολογισθέντα κέντρα, τα οποία σημειώνονται με ένα κόκκινο x.

Όπως γίνεται εύκολα αντιληπτό από τα προηγούμενα, η ανεξαρτησία από την μορφή και την εμφάνιση του καρπού αποτελεί βασική προϋπόθεση, ειδικά στην βάση που χρησιμοποιήθηκε, για την εύρεση ενός «καθολικού» κέντρου. Με την προϋπόθεση ότι η δυαδική μάσκα της χειρομορφής έχει μια σχετική ακρίβεια (δηλαδή περιλαμβάνει μόνο δάχτυλα, παλάμη και καρπό και μικρές πιθανές παραμορφώσεις, αλλά όχι μεγάλες περιοχές από το background που αντιστοιχούν σε εσφαλμένη εκτίμηση της δυαδικής μάσκας) το κέντρο της χειρομορφής μπορεί να ορισθεί ως το κέντρο του μέγιστου εγγεγραμμένου κύκλου στην δυαδική μάσκα. Ισοδύναμα ως κέντρο ορίζουμε το σημείο που απέχει περισσότερο από το περίγραμμα προς πάσα κατεύθυνση. Η παραπάνω προσέγγιση στηρίζεται στην παρατήρηση ότι σε φυσικές θέσεις του χεριού, η παλάμη είναι εύκολα οπτικά διαχωρίσιμη από τον καρπό, αφού το χέρι «φαρδαίνει» στην περιοχή της παλάμης και συνεπώς ο αναφερθείς μέγιστος εγγεγραμμένος κύκλος αντιστοιχεί στην περιοχή της παλάμης, ενώ η παρουσία του καρπού, αλλά και το μήκος αυτού,



Σχήμα Β'.1: Υπολογισμός του κέντρου της εικόνας ως κέντρο βάρους

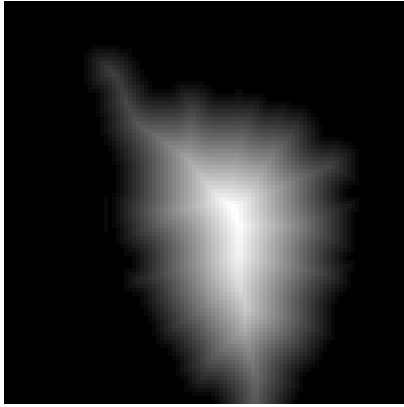
δεν επηρεάζουν καθόλου την θέση και το μέγεθος του εγγεγραμμένου κύκλου αρά και του κέντρου. Συνεπώς επιτυγχάνεται η ζητούμενη ανεξαρτησία ως προς την εμφάνιση του καρπού και τα αποτελέσματα της προσέγγισης παρουσιάζονται στο σχήμα Β'.2 για τις δυο εικόνες του σχήματος Β'.1, όπου είναι και εμφανής η βελτίωση της θέσης του κέντρου σε σχέση με την προηγούμενη προσέγγιση. Επιπλέον για λόγους οπτικής διευκόλυνσης της κατανόησης της μεθόδου παρουσιάζεται και ο μέγιστος εγγεγραμμένος κύκλος στις εικόνες του σχήματος Β'.2.



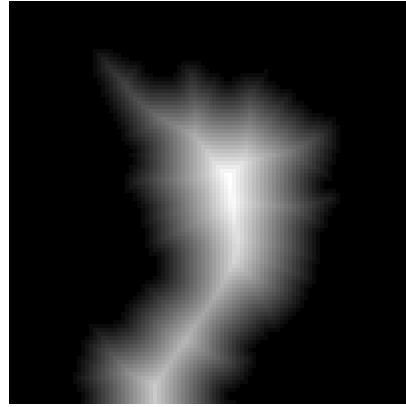
Σχήμα Β'.2: Υπολογισμός του κέντρου της εικόνας ως κέντρο του μέγιστου εγγεγραμμένου κύκλου

Ως προς την υλοποίηση, η διαδικασία που περιγράψαμε αποτελεί ουσιαστικά περιγραφή της λειτουργίας του Μετασχηματισμού Απόστασης ([33]). Ο μετασχηματισμός αντιστοιχίζει σε κάθε pixel του foreground την απόσταση του (συνήθως γίνεται χρήση

της Manhattan distance) από το κοντινότερο pixel του περιγράμματος – background ¹. Έχοντας υπολογίσει τον Μετασχηματισμό Απόστασης, επιλέγουμε ως κέντρο το pixel του μετασχηματισμού με την μεγαλύτερη τιμή. Ο Μετασχηματισμός Απόστασης για τις εικόνες τους σχήματος Β'.2, φαίνεται στο σχήμα Β'.3.



(α')



(β')

Σχήμα Β'.3: Γραφική απεικόνιση των αποτελεσμάτων του Μετασχηματισμού Απόστασης

¹Μαθηματικά ορίζεται ως εξής :

$$DT(p) = \min_{g \in S} \{d(p, g)\}, \quad p \in S^c$$

, όπου $d(x, y)$ η συνάρτηση απόστασης της επιλογής μας μεταξύ των σημείων x, y

Παράρτημα Γ'

Περιστροφική Ανεξαρτησία

Βασικό δομικό στοιχείο για πληθώρα τεχνικών που στοχεύουν στην περιστροφική ανεξαρτησία εικόνων είναι η ερμηνεία της περιστροφής ως κυκλικής μετατόπισης (παραδείγματος χάριν η χρήση - ερμηνεία του περιγράμματος ως κυκλική ακολουθία στην μέθοδο των Fourier Descriptors) και η χρήση του μετασχηματισμού Fourier. Ξεχωρίζουμε τις εξής δυο κατευθύνσεις του προβλήματος της κυκλικής μετατόπισης, τις οποίες αναλύουμε στις ενότητες που ακολουθούν.

- Εύρεση της κυκλικής μετατόπισης (ισοδυναμεί με εύρεση γωνίας περιστροφής) μεταξύ δυο σημάτων - ακολουθιών, όπου χρησιμοποιούμε το ένα από τα δυο ως σήμα αναφοράς. Για τον σκοπό αυτό υπολογίζεται η *ετεροσυσχέτιση* (*cross-correlation*) μεταξύ των δυο σημάτων όπως περιγράφουμε και παρακάτω.
- Κατασκευή πλήρως ανεξάρτητων περιγραφών ως προς την κυκλική μετατόπιση, δηλαδή χωρίς την ανάγκη σήματος αναφοράς, με χρήση του μέτρου του μετασχηματισμού Fourier (*Fourier Transform Magnitude*).

Γ'.1 Ετεροσυσχέτιση (Cross-Correlation)

Θεωρώντας δύο ακολουθίες $f[n]$ και $g[n]$, όπου η g αποτελεί ένα πρότυπο (pattern) που μπορεί να βρεθεί ως μια υποακολουθία στην f . Το πρόβλημα του template matching ([33],[12]) ορίζεται ως η εύρεση της μετατόπισης του g έτσι ώστε να ταιριάζει καλύτερα στην f και μπορεί να υπολογιστεί ως η μετατόπιση k που ελαχιστοποιεί το μέσο τετραγωνικό σφάλμα (MSE)

$$E[k] = \sum_{n \in W} (f[n+k] - g[n])^2$$

όπου W είναι το διάστημα στο \mathbb{N} που ορίζεται η g . Αναπτύσσοντας το τετραγωνικό άθροισμα

$$E[k] = \sum_{n \in W} (f[n+k]^2 + g[n]^2 - 2f[n+k]g[n])$$

Το παραπάνω κριτήριο είναι ισοδύναμο με την μεγιστοποίηση της ετεροσυσχέτισης της f με την g με την προϋπόθεση ότι η f έχει σχεδόν σταθερή τοπική ενέργεια (περιοχές ίσες με το μήκος της g). Η ετεροσυσχέτιση αποτελεί ουσιαστικά ένα μέτρο ομοιότητας μεταξύ των δυο ακολουθιών και για τον λόγο αυτό επιθυμούμε να βρούμε το k για το οποίο μεγιστοποιείται.¹

$$L[k] = \sum_{n \in W} f[n+k]g[n] \quad (\Gamma'.1)$$

Βέβαια το ζητούμενο είναι η εύρεση της κυκλικής μετατόπισης για το καλύτερο ταιριασμα ακολουθιών με το ίδιο πλήθος στοιχείων N , πρόβλημα το οποίο διατυπώνεται μαθηματικά ως

$$L[k] = \sum_{n=1}^N f[(n+k)_N]g[n] \quad (\Gamma'.2)$$

όπου χρησιμοποιούμε τον συμβολισμό $((n)_N)$ για την πράξη ($n \text{ modulo } N$), που αντιστοιχεί σε κυκλική μετατόπιση μιας ακολουθίας N στοιχείων.

Η πράξη της ετεροσυσχέτισης πρόκειται για μια μορφή συνέλιξης και μπορεί να γραφεί μέσω του μετασχηματισμού Fourier και του πεδίου των συχνοτήτων

$$\mathcal{F}^{-1}\{\mathcal{F}^*\{f\}\mathcal{F}\{g\}\}$$

Η χρήση της τελευταίας σχέσης σε συνδυασμό με την αξιοποίηση της ταχύτητας του FFT (Fast Fourier Transform) μας προσφέρει μια υπολογιστικά αποδοτικότερη υλοποίηση για την εύρεση της ετεροσυσχέτισης. Η χρήση του FFT εισάγει κάποια προβλήματα στην υλοποίηση της σχέσης $\Gamma'.1$ καθώς συνδέεται άμεσα με την πράξη της κυκλικής συνέλιξης αλλά ανταποκρίνεται επακριβώς στις απαιτήσεις της σχέσης $\Gamma'.2$. Συγκεκριμένα ως προς την πολυπλοκότητα η απευθείας υλοποίηση της σχέσης $\Gamma'.2$ έχει υπολογιστική πολυπλοκότητα $O(N^2)$, ενώ ο υπολογισμός μέσω του πεδίου των συχνοτήτων έχει πολυπλοκότητα $O(N \log N)$ εκμεταλλευόμενος την ταχύτητα του FFT.

Γ'.2 Fourier Transform Magnitude

Στην προηγούμενη ενότητα περιγράψαμε μια τεχνική για εύρεση της κυκλικής μετατόπισης μεταξύ δυο διανυσμάτων. Στην ενότητα αυτή θα παρουσιάσουμε την συ-

¹Για να έχουμε αξιόπιστα αποτελέσματα με την χρήση της πράξης της ετεροσυσχέτισης είναι συνήθης τακτική να γίνεται ταυτόχρονα και μια κανονικοποίηση των ακολουθιών -διανυσμάτων (normalized cross-correlation). Παρολ'αυτά παραλείπουμε το βήμα αυτό καθώς για τους σκοπούς της διπλωματικής τα χρησιμοποιηθέντα διανύσματα έχουν υποστεί κανονικοποίηση κατά την κατασκευή τους.

νηθέστερη τακτική για δημιουργία διανυσμάτων ανεξάρτητα σε κυκλική μετατόπιση χωρίς την χρήση κάποιου διανύσματος αναφοράς. Η τεχνική αυτή συχνά αναφέρεται ως Fourier Transform Magnitude είτε ως Fourier Transform Modulus ([28],[6]) καθώς κάνει χρήση του μέτρου του μετασχηματισμού Fourier και βασίζεται στις ιδιότητες του. Συγκεκριμένα για τον Discrete Fourier Transform, του οποίου η υλοποίηση είναι ο FFT, ισχύουν τα ακόλουθα :

$$x[n] \xrightarrow{\mathcal{F}} X[k]$$

$$x[((n-a))_N] \xrightarrow{\mathcal{F}} e^{-i(2\pi k/N)a} X[k] = X_a[k]$$

επιπλέον ισχύει $|e^{-i(2\pi k/N)a}| = 1$, συνεπώς για κάθε (κυκλική) μετατόπιση a ισχύει

$$|X_a[k]| = |X[k]|$$

ή αλλιώς (όπου με $\mathcal{F}\{\}$ συμβολίζουμε εν προκειμένω τον μετασχηματισμό DFT σε διακριτά σήματα)

$$|\mathcal{F}\{x[((n-a))_N]\}| = |\mathcal{F}\{x[n]\}|$$

Ουσιαστικά λοιπόν βασιζόμαστε στο γεγονός ότι η πληροφορία για την κυκλική μετατόπιση βρίσκεται εξολοκλήρου στην φάση του μετασχηματισμένου σήματος, οπότε μπορούμε να αποκτήσουμε ανεξαρτησία απορρίπτοντας συνολικά την πληροφορία της φάσης με την χρήση του μέτρου του μετασχηματισμού.

Βιβλιογραφία

- [1] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool. Speeded-up robust features (SURF). *Computer vision and image understanding*, 110(3):346–359, 2008.
- [2] S. Belongie, J. Malik, and J. Puzicha. Shape matching and object recognition using shape contexts. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 24(4):509–522, 2002.
- [3] C. M. Bishop. *Pattern recognition and machine learning*. Springer, New York, 2006.
- [4] A. Bosch, A. Zisserman, and X. Munoz. Image classification using random forests and ferns. In *IEEE International Conference on Computer Vision*, 2007.
- [5] P. Buehler, M. Everingham, and A. Zisserman. Learning sign language by watching TV (using weakly aligned subtitles). In *IEEE Conference on Computer Vision and Pattern Recognition*, 2009.
- [6] D. Casasent and D. Psaltis. Position, rotation, and scale invariant optical correlation. *Applied optics*, 15(7):1795–1799, 1976.
- [7] S. Conseil, S. Bourennane, and L. Martin. Comparison of fourier descriptors and hu moments for hand posture recognition. In *European Signal Processing Conference (EUSIPCO)*. Citeseer, 2007.
- [8] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 1, pages 886–893. IEEE, 2005.
- [9] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern classification*. John Wiley & Sons, 2012.
- [10] P. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan. Object detection with discriminatively trained part-based models. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 32(9):1627–1645, 2010.
- [11] Kinect Sensor for Windows. <http://www.microsoft.com/en-us/kinectforwindows/>.
- [12] R. C. Gonzalez and R. E. Woods. *Digital Image Processing*. Addison-Wesley, 1992.

- [13] B. Hariharan, J. Malik, and D. Ramanan. Discriminative decorrelation for clustering and classification. In *Computer Vision–ECCV 2012*, pages 459–472. Springer, 2012.
- [14] Ming-Kuei Hu. Visual pattern recognition by moment invariants. *Information Theory, IRE Transactions on*, 8(2):179–187, 1962.
- [15] Y. Ke and R. Sukthankar. Pca-sift: A more distinctive representation for local image descriptors. In *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*, volume 2, pages II–506. IEEE, 2004.
- [16] T. Kim, K. Livescu, and G. Shakhnarovich. American sign language fingerspelling recognition with phonological feature-based tandem models. In *Spoken Language Technology Workshop (SLT), 2012 IEEE*, pages 119–124. IEEE, 2012.
- [17] I. Kokkinos, M. Bronstein, and A. Yuille. Dense scale invariant descriptors for images and surfaces. *INRIA Research Report*, 2012.
- [18] I. Kokkinos and A. Yuille. Scale invariance without scale selection. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pages 1–8. IEEE, 2008.
- [19] S. Lazebnik, C. Schmid, and J. Ponce. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, volume 2, pages 2169–2178. IEEE, 2006.
- [20] S. Liwicki and M. Everingham. Automatic recognition of fingerspelled words in british sign language. In *Computer Vision and Pattern Recognition Workshops, 2009. CVPR Workshops 2009. IEEE Computer Society Conference on*, pages 50–57. IEEE, 2009.
- [21] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60(2):91–110, 2004.
- [22] K. Mikolajczyk and C. Schmid. A performance evaluation of local descriptors. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 27(10):1615–1630, 2005.
- [23] A. Misra, T. Abe, and K. Deguchi. Hand gesture recognition using histogram of oriented gradients and partial least squares regression. In *IAPR Conference on Machine Vision Applications*, pages 479–482, 2011.
- [24] K. Shimada, R. Muto, and T. Endo. A combined method based on svm and online learning with hog for hand shape recognition. *JACIII*, 16(6):687–695, 2012.
- [25] G. Takacs, V. Chandrasekhar, S. Tsai, D. Chen, R. Grzeszczuk, and B. Girod. Rotation invariant fast features for large-scale recognition. In *SPIE Optical Engineering and Applications*, pages 84991D–84991D. International Society for Optics and Photonics, 2012.

- [26] S. Theodoridis, A. Pikrakis, K. Koutroubas, and D. Cavouras. *Introduction to Pattern Recognition: A MATLAB Approach*. Elsevier Academic, 2010.
- [27] H. Wang, M. M. Ullah, A. Klaser, I. Laptev, and C. Schmid. Evaluation of local spatio-temporal features for action recognition. In *BMVC 2009-British Machine Vision Conference*, 2009.
- [28] G. Wolberg and S. Zokai. Robust image registration using log-polar transform. In *Image Processing, 2000. Proceedings. 2000 International Conference on*, volume 1, pages 493–496. IEEE, 2000.
- [29] Y. Yamada, T. Matsuo, N. Shimada, and Y. Shirai. Classification of hand shape with similar contour for sign language recognition. *4th Joint Workshop on Machine Perception and Robotics*, 2008.
- [30] E. Yoruk, E. Konukoglu, B. Sankur, and J. Darbon. Shape-based hand recognition. *Image Processing, IEEE Transactions on*, 15(7):1803–1815, 2006.
- [31] D. Zhang and G. Lu. A comparative study of fourier descriptors for shape representation and retrieval. In *Proceedings of the Fifth Asian Conference on Computer Vision*, pages 646–651, 2002.
- [32] Q. Zhu, M.C. Yeh, K.T. Cheng, and S. Avidan. Fast human detection using a cascade of histograms of oriented gradients. In *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, volume 2, pages 1491–1498. IEEE, 2006.
- [33] Π. Μαραγκός. *Ανάλυση Εικόνων και Όρασης Υπολογιστών*. Εκδόσεις ΕΜΠ, 2005.