



Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης
Πολυτεχνική Σχολή
Τμήμα Ηλεκτρολόγων Μηχανικών &
Μηχανικών Υπολογιστών
Τομέας Ηλεκτρονικής και Υπολογιστών
Εργαστήριο Επεξεργασίας Πληροφορίας και Υπολογισμών

Διπλωματική Εργασία

Αναλλοίωτα Χαρακτηριστικά Βασισμένα στην Τρίτης Τάξης Συσχέτιση - Εφαρμογή στην Κατηγοριοποίηση Εικόνων

Εκπόνηση:
Θωμάς Σταθοπούλου
ΑΕΜ: 6507

Επίβλεψη:
Αναπληρωτής Καθηγητής
Ντελόπουλος Αναστάσιος

Θεσσαλονίκη, Ιούλιος 2014

Περίληψη

Στη σημερινή εποχή, ο κλάδος της Μηχανικής Όρασης χρησιμοποιείται όλο και περισσότερο σε πολλές εφαρμογές, τόσο σε προσωπικό όσο και βιομηχανικό επίπεδο. Προκειμένου να μπορεί να γίνεται σωστή και γρήγορη επεξεργασία εικόνων (ή μέρους αυτής) έχουν αναπτυχθεί μέχρι σήμερα πολλοί αλγόριθμοι που υπολογίζουν τοπικούς περιγραφείς/αναπαράστασεις, οι οποίοι να είναι όσο το δυνατόν πιο ανεξάρτητοι από διάφορες μετατροπές που μπορούν να εφαρμοστούν στις εικόνες και να περιέχουν όσο το δυνατόν πιο χρήσιμη πληροφορία. Στα πλαίσια της παρούσας διπλωματικής εργασίας επιχειρείται, με χρήση των τρίτης τάξης συσχετίσεων και των ιδιοτήτων τους, η υλοποίηση αναπαράστασης, ανεξάρτητης της περιστροφής, κλιμάκωσης ή/και μετατόπισης της εικόνας, καθώς και της προσθήκης θορύβου. Η αναπαράσταση αυτή χρησιμοποιείται ως ένας τοπικός περιγραφέας της υψής μιας γειτονιάς της εικόνας και, με χρήση του μοντέλου Bag of Words συντελεί στη δημιουργία ενός ολικού περιγραφέα της εικόνας. Αφού αναλυθούν οι ιδιότητες που χρησιμοποιούνται και ο τρόπος δημιουργίας της αναπαράστασης, εκτελούνται πειράματα σε πραγματικές εικόνες. Στα πλαίσια των πειραμάτων, γίνεται επεξεργασία περίπου δύο χιλιάδων εικόνων, χρησιμοποιώντας την υλοποιημένη προτεινόμενη αναπαράσταση και τον γνωστό αλγόριθμο SIFT και στη συνέχεια εκτελείται ταξινόμηση των εικόνων αυτών σε είκοσι κλάσεις. Τέλος τα αποτελέσματα και των δύο μεθόδων αναλύονται και συγκρίνονται.

Diploma Thesis

Title

Invariant Features Based on Third Order Correlation - Application on Image Classification

Abstract

Computer Vision is being used more and more everyday, in our personal life, as well as in a variety of industrial applications. In order to be able to process an image (or a part of it) accurately and fast, one needs a representation of the image, that does not change when the image does. In light of that, plenty image descriptors are proposed and used nowadays. The current diploma thesis attempts to implement an algorithm that produces an image representation, that is invariant to image transformations, such as rotation, scaling and/or translation, as well as noise. This representation plays the role of a local feature, which through the implementation of the Bag of Words model produces a global feature of the image. Taking advantage of third order correlations and their properties, the representation is created and then tested on real images. A series of experiments are being conducted producing the features of two thousand images, using the proposed representation and the well known and widely used SIFT algorithm. These features are then classified into twenty classes and the results of the classification are analyzed and compared.

Thomai Stathopoulou
Information Processing Laboratory
Electrical & Computer Engineering Department
Aristotle University of Thessaloniki
July, 2014

Περιεχόμενα

1	Εισαγωγή	1
1.1	Διατύπωση του Προβλήματος	1
1.2	Υπάρχουσες Μέθοδοι	2
1.2.1	Μοντέλο Bag of Words	2
1.2.2	SIFT	3
1.2.3	SURF	4
2	Θεωρητικό Υπόβαθρο	5
2.1	Συσχετίσεις	5
2.1.1	Ορισμός	5
2.1.2	Ιδιότητες των Συσχετίσεων	6
2.1.3	Συσχετίσεις & Μετασχηματισμός Fourier	8
3	Προτεινόμενος Περιγραφέας	11
3.1	Θεωρητική Ανάλυση της Αναπαράστασης	11
3.2	Αλγοριθμική Υλοποίηση της Αναπαράστασης	14
3.3	Πολυπλοκότητα	15
3.4	Παραδείγματα	16
3.5	Υλοποίηση Μοντέλου Bag of Words	24
4	Πειράματα	29
4.1	Οργάνωση Πειραμάτων	29
4.1.1	Δεδομένα	29
4.1.2	Υλοποίηση του μοντέλου Bag of Words	30
4.1.3	Ταξινόμηση των εικόνων	32
4.1.4	Αξιολόγηση των αποτελεσμάτων	33
4.2	Αποτελέσματα	34
4.2.1	Προτεινόμενος Περιγραφέας	34
4.2.2	SIFT	41
5	Συμπεράσματα	45

Κατάλογος σχημάτων

1.1	Υπολογισμός ολοκληρωτικής μορφής	4
3.1	Σύστημα Αναπαράστασης Διανυσμάτων με τις γωνίες θ_1, θ_2	13
3.2	Παράδειγμα με τετράγωνο	17
3.3	Παράδειγμα με περίγραμμα τετραγώνου	18
3.4	Παράδειγμα με τρίγωνο	20
3.5	Παράδειγμα με ορθογώνιο	21
3.6	Αναπαράσταση τετραγώνου για διαφορετικό αριθμό επιπέδων κβά- ντισης των $\rho, \phi, \theta_1, \theta_2$	23
3.8	Γκαουσιανή συνάρτηση ως εικόνα	24
3.7	Γράφημα της Γκαουσιανής συνάρτησης στις δύο διαστάσεις	25
3.9	Εικόνα πολλαπλασιασμένη με γκαουσιανή συνάρτηση με διαφορετικά σ_x, σ_y και (x_0, y_0)	26
3.10	Παράδειγμα με πραγματική εικόνα	27
4.1	Πλέγμα σε εικόνα και γειτονιές πριν και μετά την εφαρμογή της Γκα- ουσιανής συνάρτησης	31

Κατάλογος πινάκων

3.1	Χρόνοι παραγωγής της προτεινόμενης αναπαράστασης	15
4.1	Οι κλάσεις που χρησιμοποιήθηκαν και ο αριθμός των εικόνων τους .	30
4.2	MAP κλάσεων με προκαθ. (C, γ) , $Folds = 3$ & $ \theta_1 = \theta_2 = 9$	34
4.3	MAP κλάσεων με προκαθ. (C, γ) , $Folds = 10$ & $ \theta_1 = \theta_2 = 9$	35
4.4	MAP κλάσεων με προκαθ. (C, γ) & $Folds = 50$, $ \theta_1 = \theta_2 = 9$	35
4.5	MAP κλάσεων με προκαθ. (C, γ) , $Folds = 3$ & $ \theta_1 = \theta_2 = 10$	36
4.6	MAP κλάσεων με προκαθ. (C, γ) , $Folds = 10$ & $ \theta_1 = \theta_2 = 10$	36
4.7	MAP κλάσεων με προκαθ. (C, γ) , $Folds = 50$ & $ \theta_1 = \theta_2 = 10$	37
4.8	MAP κλάσεων με βέλτιστα (C, γ) , $Folds = 3$ & $ \theta_1 = \theta_2 = 9$	38
4.9	MAP κλάσεων με βέλτιστα (C, γ) , $Folds = 10$ & $ \theta_1 = \theta_2 = 9$	38
4.10	MAP κλάσεων με βέλτιστα (C, γ) , $Folds = 50$ & $ \theta_1 = \theta_2 = 9$	39
4.11	MAP κλάσεων με βέλτιστα (C, γ) , $Folds = 3$ & $ \theta_1 = \theta_2 = 10$	39
4.12	MAP κλάσεων με βέλτιστα (C, γ) , $Folds = 10$ & $ \theta_1 = \theta_2 = 10$	40
4.13	MAP κλάσεων με βέλτιστα (C, γ) , $Folds = 50$ & $ \theta_1 = \theta_2 = 10$	40
4.14	Συνολικό Mean Average Precision όλων των κλάσεων πριν και μετά την επιλογή των βέλτιστων (C, γ)	41
4.15	SIFT: MAP κλάσεων με βέλτιστα (C, γ) & $Folds = 3$	42
4.16	SIFT: MAP κλάσεων με βέλτιστα (C, γ) & $Folds = 10$	42
4.17	SIFT: MAP κλάσεων με βέλτιστα (C, γ) & $Folds = 50$	43
4.18	Συνολικό Mean Average Precision όλων των κλάσεων με τα βέλτιστα (C, γ)	43

1

Εισαγωγή

1.1 ΔΙΑΤΥΠΩΣΗ ΤΟΥ ΠΡΟΒΛΗΜΑΤΟΣ

Για τον άνθρωπο η αντίληψη του κόσμου μέσω της όρασης είναι κάτι αρκετά εύκολο. Μπορεί να ξεχωρίσει αντικείμενα, σκηνές, πρόσωπα κτλ χωρίς να καταβάλει ιδιαίτερη προσπάθεια. Ακόμα είναι δυνατό να αναγνωρίσει συναισθήματα ή και προθέσεις μέσω των εκφράσεων του προσώπου ή τη στάση του σώματος. Για έναν υπολογιστή, όμως, κάτι αντίστοιχο είναι ένα αρκετά δύσκολο εγχείρημα.

Ήδη από τη δεκαετία του '70 στον τομέα της μηχανικής όρασης γίνονται πολλές προσπάθειες και έχει υπάρξει μεγάλη πρόοδος, σε θέματα επεξεργασίας, αναγνώρισης και κατηγοριοποίησης εικόνων. Πολλές πτυχές αυτού του τομέα χρησιμοποιούνται σήμερα, τόσο στη βιομηχανία, όσο και σε προσωπικές συσκευές. Παρόλα αυτά ο κλάδος της μηχανικής όρασης, θεωρείται πως χρειάζεται πολλή περισσότερη ανάπτυξη, προκειμένου ένας υπολογιστής να αντιλαμβάνεται τον κόσμο μέσω της όρασης όπως ένας άνθρωπος, αν αυτό είναι δυνατό να συμβεί ολοκληρωτικά.

Προκειμένου ένας υπολογιστής να ταξινομήσει σωστά ή να αναγνωρίσει μέρος μιας εικόνας, θα πρέπει τα δεδομένα που παρέχονται να είναι σε μορφή που ευνοεί κάτι τέτοιο. Υπάρχει, λοιπόν, η ανάγκη για αναπαράσταση των εικόνων (ή μέρος αυτών) με τρόπο που δεν επηρεάζεται από τυχόν αλλαγές στην ίδια την εικόνα, όπως είναι περιστροφή, αλλαγή κλίμακας, προσθήκη θορύβου κτλ.

Αυτό επιχειρούν να υλοποιήσουν οι τοπικοί περιγραφείς. Ένας τοπικός περιγραφέας δημιουργεί ψηφιακή αναπαράσταση (συνήθως ένα διάνυσμα) για μια γειτονιά της εικόνας. Στόχος είναι το διάνυσμα αυτό να μένει όσο το δυνατόν ανεπηρέαστο σε αλλαγές που μπορεί να εφαρμοστούν στην εκάστοτε γειτονιά, την οποία περιγράφει. Δημιουργώντας πολλά τέτοια διανύσματα (Local Features) για γειτονιές που καλύπτουν όλη την έκταση της εικόνας, είναι δυνατό, με χρήση του μοντέλου

Bag of Words, να παραχθεί αντιπροσωπευτική αναπαράσταση για ολόκληρη την εικόνα. Δημιουργείται δηλαδή, ένα νέο διάνυσμα, το οποίο λειτουργεί ως ολικός περιγραφέας. Το νέο αυτό διάνυσμα (Feature Vector) μπορεί να τροφοδοτηθεί σε κάποιον ταξινομητή, προκειμένου να γίνει σωστή κατηγοριοποίηση της εικόνας.

Στην παρούσα διπλωματική εργασία επιχειρείται η υλοποίηση μιας αναπαράστασης εικόνας, η οποία να είναι απαλλαγμένη από την επιρροή μεταβολών λόγω περιστροφής, αλλαγής κλίμακας, μετατόπισης και προσθήκης θορύβου. Αυτή η αναπαράσταση λειτουργεί ως τοπικός περιγραφέας υφής, έτσι ώστε μέσω του μοντέλου Bag of Words να παραχθούν ολικοί περιγραφείς εικόνων. Ο τρόπος υλοποίησης του προτεινόμενου περιγραφέα αναλύεται στο Κεφάλαιο 3.

1.2 ΥΠΑΡΧΟΥΣΕΣ ΜΕΘΟΔΟΙ

Για την αναπαράσταση και περιγραφή μιας εικόνας έχουν αναπτυχθεί πολλές μέθοδοι και διάφοροι αλγόριθμοι στον τομέα της τεχνητής όρασης. Παρακάτω περιγράφεται το μοντέλο Bag of Words, το οποίο χρησιμοποιείται στα πλαίσια της εργασίας για την παραγωγή των ολικών περιγραφέων, καθώς και δύο πολύ σημαντικοί τοπικοί περιγραφείς, ο SIFT και ο SURF.

1.2.1 Μοντέλο Bag of Words

Το μοντέλο Bag of Words [SM86] αποτελεί μια απλοποιημένη αναπαράσταση της εικόνας. Προκειμένου να παραχθεί ο περιγραφέας της εικόνας, αυτή αντιμετωπίζεται όπως θα αντιμετωπιζόταν ένα έγγραφο κειμένου. Εξάγονται, δηλαδή οι εικονικές "λέξεις", τις οποίες περιέχει και έτσι τελικά η εικόνα χαρακτηρίζεται από το πόσες και ποιες "λέξεις" περιέχει.

Για τη δημιουργία ολικών περιγραφέων μέσω του μοντέλου Bag of Words, αρχικά εφαρμόζεται στην εικόνα πυκνό πλέγμα σταθερού βήματος. Σε κάθε κόμβο του πλέγματος υπολογίζεται ένας τοπικός περιγραφέας, που περιγράφει μια μικρή γειτονιά γύρω από τον κόμβο. Αυτή η διαδικασία επαναλαμβάνεται για πλήθος εικόνων και έτσι παράγεται ένα σύνολο τοπικών περιγραφέων.

Στη συνέχεια, οι τοπικοί περιγραφείς ομαδοποιούνται σύμφωνα με το περιεχόμενό τους και κάθε ομάδα αντιπροσωπεύεται από ένα διάνυσμα μήκους όσο και τα διανύσματα των τοπικών περιγραφέων. Το σύνολο των διανυσμάτων που αντιπροσωπεύουν τις ομάδες που δημιουργήθηκαν αποτελούν το Codebook, δηλαδή το λεξικό των εικονικών λέξεων που εντοπίστηκαν στις εικόνες, από τις οποίες παράχθηκαν οι τοπικοί περιγραφείς.

Πλέον με βάση το Codebook που έχει δημιουργηθεί, μπορεί μια εικόνα να αναπαρασταθεί σύμφωνα με αυτό. Παράγονται οι τοπικοί περιγραφείς σε κάθε κόμβο του πλέγματος που εφαρμόζεται στην εικόνα, όπως αναφέρθηκε παραπάνω και στη συνέχεια ο κάθε περιγραφέας αντιστοιχίζεται σε μια ομάδα/εικονική λέξη, ανάλογα

με την απόστασή του από το αντίστοιχο διάνυσμα που την αντιπροσωπεύει. Τελικά δημιουργείται ένα ιστόγραμμα που απεικονίζει τη συχνότητα εμφάνισης της κάθε λέξης του Codebook στην εικόνα.

Το ιστόγραμμα που παράγεται αποτελεί τον ολικό περιγραφέα της εικόνας. Πρόκειται, δηλαδή για ένα διάνυσμα, το οποίο έχει τόσα στοιχεία, όσες και οι εικονικές λέξεις του Codebook, όπου το κάθε στοιχείο μας δίνει τη συχνότητα εμφάνισης της αντίστοιχης λέξης. Στο τέλος, το διάνυσμα που προκύπτει για την κάθε εικόνα κανονικοποιείται ώστε τα στοιχεία του να αθροίζουν σε σταθερό αριθμό (συνήθως μονάδα).

1.2.2 SIFT

Ο αλγόριθμος SIFT [Low04] (Scale Invariant Feature Transform) παράγει τοπικούς περιγραφείς, οι οποίοι είναι ανεξάρτητοι από την κλιμάκωση και τον προσανατολισμό της εικόνας.

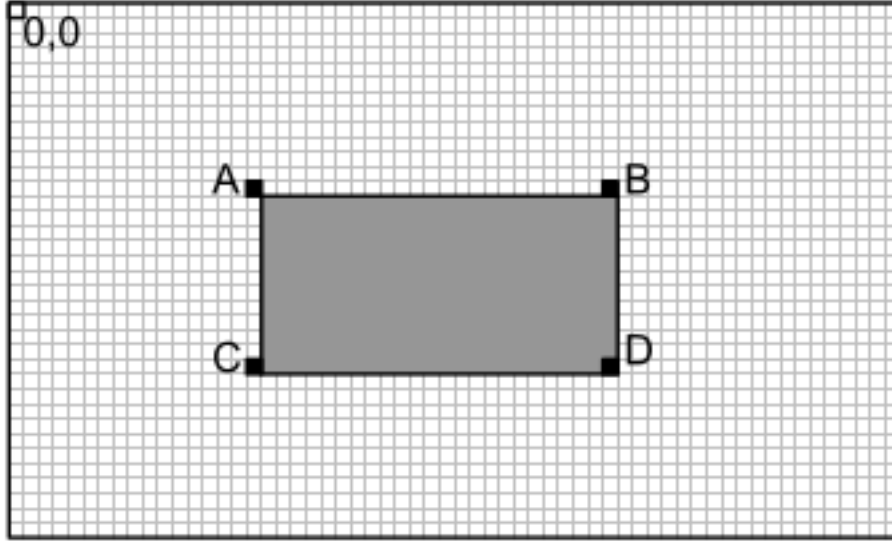
Προκειμένου να υπάρχει ανεξαρτησία από τον προσανατολισμό της εικόνας, αρχικά υπολογίζονται τα παρακάτω δύο μεγέθη σε κάθε σημείο:

$$m(x, y) = \sqrt{(I(x+1, y) - I(x-1, y))^2 + (I(x, y+1) - I(x, y-1))^2} \quad (1.1)$$

$$\theta(x, y) = \tan^{-1}((I(x, y+1) - I(x, y-1)) / (I(x+1, y) - I(x-1, y))), \quad (1.2)$$

όπου $I(x, y)$ η εικόνα εισόδου, $m(x, y)$ το μέγεθος (gradient magnitude) και $\theta(x, y)$ ο προσανατολισμός (gradient orientation) της κλίσης. Στη συνέχεια εφαρμόζεται η γκαουσιανή συνάρτηση με σ ίσο με το μισό του πλάτους της εικόνας, έτσι ώστε όσο μεγαλώνει η απόσταση από το κέντρο, οι αντίστοιχες κλίσεις να έχουν μικρότερη βαρύτητα. Έπειτα, δημιουργούνται ιστογράμματα προσανατολισμού σε υποπεριοχές της εικόνας λαμβάνοντας υπόψιν όλα τα $m(x, y)$ και $\theta(x, y)$ που υπολογίστηκαν. Οι κορυφές αυτών των ιστογραμμάτων αντιστοιχούν στις εντονότερες κατευθύνσεις των τοπικών κλίσεων και αυτές καθορίζουν τον προσανατολισμό της κάθε υποπεριοχής. Τελικά το διάνυσμα του περιγραφέα περιέχει τους εντονότερους προσανατολισμούς των περιοχών. Το διάνυσμα αυτό κανονικοποιείται, ώστε να έχει μέτρο μονάδα. Έπειτα γίνεται έλεγχος στα στοιχεία του, στα οποία εφαρμόζεται ένα ανώφλι 0.2 και, πλέον στο νέο διάνυσμα, γίνεται ξανά κανονικοποίηση του μέτρου στη μονάδα. Αυτές οι ενέργειες οδηγούν στην ανεξαρτησία σε αλλαγές της φωτεινότητας.

Όσον αφορά στην ανεξαρτησία από την κλιμάκωση, υπάρχει η δυνατότητα, μετά από επιλογή της κατάλληλης κλίμακας (ή κλιμάκων), να δημιουργηθούν αντίγραφα της εικόνας, τα οποία κλιμακώνονται με χρήση γκαουσιανού φίλτρου. Πλέον η διαδικασία που περιγράφηκε παραπάνω, επαναλαμβάνεται σε κάθε εκδοχή της εικόνας που έχει παραχθεί.



Σχήμα 1.1 – Υπολογισμός ολοκληρωτικής μορφής

1.2.3 SURF

Ο αλγόριθμος SURF [BETVG08] (Speeded-Up Robust Features), σε αντίθεση με τον SIFT υπολογίζει τον περιγραφέα χρησιμοποιώντας την ολοκληρωτική μορφή της εικόνας (integral image), και όχι αυτούσια την εικόνα. Πρόκειται για έναν πίνακα ίδιας διάστασης με την εικόνα, όπου το (i, j) στοιχείο του είναι το άθροισμα του (i, j) pixel της εικόνας και όλων των pixels που βρίσκονται πάνω και αριστερά από αυτό. Ο τύπος υπολογισμού δίνεται παρακάτω:

$$I_{\Sigma}(x, y) = \sum_{i=0}^x \sum_{j=0}^y I(i, j), \quad (1.3)$$

όπου I είναι η αρχική εικόνα και I_{Σ} η ολοκληρωτική μορφή της. Το ενδιαφέρον με αυτή τη μορφή είναι, πως για να υπολογίσει κανείς την ολοκληρωτική μορφή μιας ορθογώνιας περιοχής της εικόνας, ανεξαρτήτου μεγέθους και θέσης, το μόνο που χρειάζεται είναι οι τιμές τεσσάρων pixel. Ένα παράδειγμα φαίνεται στην εικόνα του Σχήματος 1.1, όπου η ολοκληρωτική μορφή της γκρι περιοχής προκύπτει $\Sigma = I_{\Sigma}(A) + I_{\Sigma}(D) - I_{\Sigma}(C) - I_{\Sigma}(B)$. Αυτή την ιδιότητα εκμεταλλεύεται ο αλγόριθμος SURF για ελάττωση της ταχύτητας των υπολογισμών.

Για ανεξαρτησία από την περιστροφή ανατίθεται στην εικόνα μία "υπερισχύουσα κατεύθυνση" (dominant direction). Στη συνέχεια η εικόνα περιστρέφεται, προκειμένου να αποκτήσει τον προσανατολισμό της υπερσχύουσας κατεύθυνσης που της έχει ανατεθεί. Έπειτα χωρίζεται σε υποπεριοχές, όπου σε κάθε μία εφαρμόζεται φίλτρο Haar. Στα αποτελέσματα του φίλτρου εφαρμόζεται η γκαουσιανή συνάρτηση και αθροίζονται για την κάθε υποπεριοχή σε ένα διάνυσμα. Βάζοντας αυτά τα διανύσματα στη σειρά προκύπτει ο περιγραφέας. Τέλος, γίνεται κανονικοποίηση του μέτρου του διανύσματος στη μονάδα.

2

Θεωρητικό Υπόβαθρο

2.1 ΣΥΣΧΕΤΙΣΕΙΣ

2.1.1 Ορισμός

Η έννοια των συσχετίσεων (correlations) χρησιμοποιείται κατά κόρον στην επεξεργασία σήματος και εικόνας. Έστω $x(\mathbf{t})$ ένα διακριτό σήμα δύο διαστάσεων ($2\mathbf{D}$), όπου $\mathbf{t} \in S = [0, \dots, N-1] \times [0, \dots, N-1]$. Η συσχέτιση δεύτερης και τρίτης τάξης αυτού του σήματος ορίζεται ως:

$$x_2(\boldsymbol{\tau}) \triangleq \sum_{\mathbf{t} \in S} x(\mathbf{t})x(\mathbf{t} + \boldsymbol{\tau}) \quad (2.1)$$

$$x_3(\boldsymbol{\tau}_1, \boldsymbol{\tau}_2) \triangleq \sum_{\mathbf{t} \in S} x(\mathbf{t})x(\mathbf{t} + \boldsymbol{\tau}_1)x(\mathbf{t} + \boldsymbol{\tau}_2) \quad (2.2)$$

αντίστοιχα, όπου $\boldsymbol{\tau}, \boldsymbol{\tau}_1, \boldsymbol{\tau}_2 \in S' = [-(N-1), \dots, N-1] \times [-(N-1), \dots, N-1]$.

Οι παραπάνω εξισώσεις (2.1 και 2.2) αφορούν σε ντετερμινιστικά σήματα. Στην περίπτωση των στοχαστικών σημάτων γίνεται χρήση των εκτιμητριών συναρτήσεων υπολογισμού ροπών:

$$m_2^N(\boldsymbol{\tau}) \triangleq \frac{1}{N} \sum_{\mathbf{t} \in S} x(\mathbf{t})x(\mathbf{t} + \boldsymbol{\tau}) \quad (2.3)$$

$$m_3^N(\boldsymbol{\tau}_1, \boldsymbol{\tau}_2) \triangleq \frac{1}{N^2} \sum_{\mathbf{t} \in S} x(\mathbf{t})x(\mathbf{t} + \boldsymbol{\tau}_1)x(\mathbf{t} + \boldsymbol{\tau}_2) \quad (2.4)$$

όπου και πάλι $\boldsymbol{\tau}, \boldsymbol{\tau}_1, \boldsymbol{\tau}_2 \in S' = [-(N-1), \dots, N-1] \times [-(N-1), \dots, N-1]$.

Οι παραπάνω τύποι υπολογίζουν τις συσχετίσεις διδιάστατων σημάτων. Αντίστοιχα μπορούν να υπολογισθούν οι συσχετίσεις μονοδιάστατων (1D) σημάτων. Έστω $x(t)$ ένα διακριτό μονοδιάστατο σήμα, όπου $t \in S = [0, \dots, N-1]$:

$$x_2(\tau) \triangleq \sum_{t \in S} x(t)x(t+\tau) \quad (2.5)$$

$$x_3(\tau_1, \tau_2) \triangleq \sum_{t \in S} x(t)x(t+\tau_1)x(t+\tau_2) \quad (2.6)$$

$$m_2^N(\tau) \triangleq \frac{1}{N} \sum_{t \in S} x(t)x(t+\tau) \quad (2.7)$$

$$m_3^N(\tau_1, \tau_2) \triangleq \frac{1}{N^2} \sum_{t \in S} x(t)x(t+\tau_1)x(t+\tau_2) \quad (2.8)$$

όπου $\tau, \tau_1, \tau_2 \in S' = [-(N-1), \dots, N-1]$.

Στην περίπτωση, λοιπόν ενός διδιάστατου διακριτού σήματος, δηλαδή μίας ει-κόνας, η συσχέτιση τρίτου βαθμού υπολογίζεται από την εξίσωση 2.2. Αυτό θα χρησιμοποιηθεί στα πλαίσια της παρούσας διπλωματικής εργασίας. Θα πρέπει να σημειωθεί πως το αποτέλεσμα της εξίσωσης 2.2 είναι ένας πίνακας τεσσάρων διαστάσεων (4D), με κάθε διάσταση να έχει μήκος $(2 \times N - 1)$.

Οι παραπάνω σχέσεις ισχύουν για τετραγωνικά¹ και μη σήματα. Για λόγους απλό-τητας όμως, θεωρούμε στο εξής πως τα σήματα με τα οποία ασχολούμαστε είναι τετραγωνικά. Άλλωστε κάτι τέτοιο δεν είναι δύσκολο να ισχύει, αφού στην περί-πτωση που έχουμε ένα μη τετραγωνικό σήμα, αυτό μπορεί εύκολα να μετατραπεί σε τετραγωνικό με επέκταση της μικρότερης διάστασης.

2.1.2 Ιδιότητες των Συσχετίσεων

Σε αυτό το κομμάτι θα αναλυθούν κάποιες ενδιαφέρουσες και πολύ χρήσιμες ιδιότητες των συσχετίσεων. Έστω ένα διδιάστατο διακριτό ντετερμινιστικό σήμα $x(\mathbf{t}), \mathbf{t} \in S = [0, \dots, N-1] \times [0, \dots, N-1]$, διάστασης $N \times N$. Έστω η συσχέτιση δεύτερης και τρίτης τάξης $x_2(\boldsymbol{\tau})$ και $x_3(\boldsymbol{\tau}_1, \boldsymbol{\tau}_2)$, $\boldsymbol{\tau}, \boldsymbol{\tau}_1, \boldsymbol{\tau}_2 \in S' = [-(N-1), \dots, N-1] \times [-(N-1), \dots, N-1]$, που υπολογίσθηκαν από τις εξισώσεις 2.1 και 2.2 αντίστοιχα.

Συμμετρίες Ισχύουν οι παρακάτω σχέσεις συμμετρίας:

$$x_2(\boldsymbol{\tau}) = x_2(-\boldsymbol{\tau}) \quad (2.9)$$

$$\begin{aligned} x_3(\boldsymbol{\tau}_1, \boldsymbol{\tau}_2) &= x_3(\boldsymbol{\tau}_1 - \boldsymbol{\tau}_2, -\boldsymbol{\tau}_2) = x_3(\boldsymbol{\tau}_2 - \boldsymbol{\tau}_1, -\boldsymbol{\tau}_1) \\ &= x_3(\boldsymbol{\tau}_2, \boldsymbol{\tau}_1) = x_3(-\boldsymbol{\tau}_1, \boldsymbol{\tau}_2 - \boldsymbol{\tau}_1) \\ &= x_3(-\boldsymbol{\tau}_2, \boldsymbol{\tau}_1 - \boldsymbol{\tau}_2) \end{aligned} \quad (2.10)$$

¹Τετραγωνικά χαρακτηρίζονται τα σήματα δύο διαστάσεων, των οποίων και οι δύο διαστάσεις έχουν ίδιο μήκος.

Οι συμμετρίες αυτές ισχύουν και για μονοδιάστατα και για στοχαστικά σήματα και είναι πολύ χρήσιμες στην υλοποίηση των συναρτήσεων που υπολογίζουν τις συσχετίσεις, καθώς μπορούν να μειώσουν την υπολογιστική πολυπλοκότητα και τη μνήμη που χρειάζεται για να αποθηκευθούν τα αποτελέσματα.

Αναισθησία στο θόρυβο Έστω $e(t)$ Gaussian θόρυβος με μηδενική μέση τιμή ή γραμμικός και συμμετρικά κατανομημένος θόρυβος. Ισχύει:

$$E\{e_3(\tau_1, \tau_2)\} = 0 \quad (2.11)$$

Επίσης ισχύει:

$$e_3(\tau_1, \tau_2) \rightarrow 0 \text{ αν } N \rightarrow \infty \quad (2.12)$$

Έστω $y(t) = x(t) + e(t)$, το αρχικό διδιάστατο σήμα παραμορφωμένο από θόρυβο. Τότε έχουμε:

$$E\{y_3(\tau_1, \tau_2)\} = x_3(\tau_1, \tau_2), \forall(\tau_1, \tau_2) \quad (2.13)$$

Η αναμενόμενη τιμή της συσχέτισης του παραμορφωμένου σήματος είναι ανεξάρτητη του θορύβου.

Σε γενικές γραμμές, υπό συνθήκες που εξασφαλίζουν την ύπαρξη του $x_3(\tau_1, \tau_2)$, θεωρείται πως $y_3(\tau_1, \tau_2) = x_3(\tau_1, \tau_2)$.

1-1 Αντιστοιχία Δεδομένου ότι $x_3(\tau_1, \tau_2) \neq 0$, δηλαδή **τουλάχιστον ένα** στοιχείο της συσχέτισης δεν είναι μηδενικό, υπάρχει μία 1-1 αντιστοιχία μεταξύ του διδιάστατου σήματος και της συσχέτισής του:

$$x_3(\tau_1, \tau_2) = y_3(\tau_1, \tau_2) \Leftrightarrow x(t) = y(t) \quad (2.14)$$

Αυτό σημαίνει πως, κατά τον υπολογισμό των συσχετίσεων τρίτης τάξης, δεν υπάρχει απώλεια πληροφορίας του σήματος. Μπορούμε, δηλαδή, να αναγνωρίσουμε το σήμα από τη συσχέτισή του.

Μεταφορά - Περιστροφή - Κλιμάκωση Έστω $y(t) = x(\mathbf{T}_{\alpha,\theta}\mathbf{t} + \mathbf{t}_0)$, όπου:

$$\mathbf{T}_{\alpha,\theta} = \alpha \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix} \quad (2.15)$$

πίνακας περιστροφής περί την αρχή των αξόνων κατά γωνία θ και διαστολής / συστολής κατά α και \mathbf{t}_0 διάνυσμα μετατόπισης. Δεδομένου ότι μπορεί να εφαρμοστεί αυτός ο μετασχηματισμός, ισχύει:

$$y_3(\tau_1, \tau_2) = x(\mathbf{T}_{\alpha,\theta}\tau_1, \mathbf{T}_{\alpha,\theta}\tau_2) \quad (2.16)$$

Αυτό σημαίνει πως οι συσχετίσεις δεν επηρεάζονται από τη μετατόπιση του αρχικού σήματος. Όσον αφορά στην περιστροφή και/ή κλιμάκωση, όπως φαίνεται από την εξίσωση 2.16, ένας αντίστοιχος μετασχηματισμός εφαρμόζεται και στα δυο

διανύσματα-δείκτες της συσχέτισης.

Εν ολίγοις αν το αρχικό σήμα περιστραφεί ή/και κλιμακωθεί, τότε το περιεχόμενο της συσχέτισής του δεν αλλάζει, αλλά αλλάζουν θέση τα στοιχεία του. Αν λοιπόν γνωρίζουμε τη γωνία περιστροφής ή/και το λόγο κλιμάκωσης, τότε μπορούμε να γνωρίζουμε τις νέες θέσεις των στοιχείων αυτών.

Σημείωση: Η παραπάνω ιδιότητα ισχύει δεδομένου ότι μετά το μετασχηματισμό του αρχικού σήματος δεν υπάρχει απώλεια πληροφορίας. Μπορούμε, δηλαδή να θεωρήσουμε πως το $x(\mathbf{t})$ είναι αρχικά τοποθετημένο πάνω σε ένα πλαίσιο $P = [0, \dots, M-1] \times [0, \dots, M-1]$, $M > N$. Η απαίτησή μας για να ισχύει η ιδιότητα, είναι, πως μετά τον οποιοδήποτε μετασχηματισμό που θα υποστεί το σήμα, θα βρίσκεται ολόκληρο στο εσωτερικό του P . Για αυτό το λόγο κάνουμε εξ' αρχής την παραδοχή, ότι το S έχει τέτοιο μέγεθος, ώστε να υπάρχει το περιθώριο για το $x(\mathbf{t})$ να μετασχηματιστεί, χωρίς να υποστεί αποκοπή στις άκρες του.

2.1.3 Συσχετίσεις & Μετασχηματισμός Fourier

Η σημαντικότερη, ίσως, ιδιότητα των συσχετίσεων, είναι αυτή που τις συνδέει με το μετασχηματισμό Fourier. Έστω $x(\mathbf{t})$, $x_2(\tau)$ και $x_3(\tau_1, \tau_2)$ το αρχικό σήμα και οι συσχετίσεις του, όπως στην παράγραφο 2.1.2.

Αν $X(\omega)$, $X_2(\omega)$ και $X_3(\omega_1, \omega_2)$ οι μετασχηματισμοί Fourier των $x(\mathbf{t})$, $x_2(\tau)$ και $x_3(\tau_1, \tau_2)$ αντίστοιχα, οι οποίοι υπολογίζονται από τον τύπο του διακριτού μετασχηματισμού Fourier:

$$F(k, l) = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} f(i, j) e^{-i2\pi(\frac{ki}{N} + \frac{lj}{N})} \quad (2.17)$$

τότε, εύκολα με πράξεις, αποδεικνύεται πως ισχύουν οι παρακάτω δύο ιδιότητες:

$$X_2(\omega) = |X(\omega)|^2 = X(\omega)X^*(\omega) = X(\omega)X(-\omega) \quad (2.18)$$

$$X_3(\omega_1, \omega_2) = X(\omega_1)X(\omega_2)X(-\omega_1 - \omega_2) = X(\omega_1)X(\omega_2)X^*(\omega_1 + \omega_2) \quad (2.19)$$

Οι ιδιότητες αυτές μας δείχνουν πως μπορούμε να υπολογίσουμε τους μετασχηματισμούς Fourier των συσχετίσεων, απλά και μόνο γνωρίζοντας το μετασχηματισμό Fourier του σήματος. Μπορούμε, δηλαδή να υπολογίσουμε τις ίδιες τις συσχετίσεις (μέσω του αντίστροφου μετασχηματισμού Fourier) γνωρίζοντας μόνο το μετασχηματισμό Fourier του αρχικού σήματος. Αυτό είναι πάρα πολύ χρήσιμο υπολογιστικά, καθώς μας συμφέρει, από άποψη πολυπλοκότητας κώδικα και από άποψη χρόνου, να βρίσκουμε τις συσχετίσεις, μέσω αυτής της ιδιότητας, με τη χρήση έτοιμων συναρτήσεων υπολογισμού ευθύ και αντίστροφου μετασχηματισμού Fourier, παρά με τη χρήση του ορισμού τους (εξισώσεις 2.1 και 2.2).

Ιδιαίτερα ενδιαφέρον είναι το γεγονός ότι, μέσω της ιδιότητας 2.19, μπορούμε να υπολογίσουμε το αρχικό σήμα γνωρίζοντας την τρίτης τάξης συσχέτισή του (ή το μετασχηματισμό Fourier της). Το ίδιο, όμως, δε συμβαίνει μέσω της ιδιότητας

2.18, αφού η ιδιότητα αυτή συνδέει το μετασχηματισμό Fourier της δεύτερης τάξης συσχέτισης με το μέτρο του μετασχηματισμού Fourier του αρχικού σήματος. Σε αυτή την περίπτωση, δηλαδή, δεν υπάρχει 1-1 αντιστοιχία.

3

Προτεινόμενος Περιγραφέας

Το εγχείρημα της παρούσας διπλωματικής εργασίας είναι η δημιουργία μίας αναπαράστασης εικόνας, η οποία παραμένει αναλλοίωτη σε μετασχηματισμούς της εικόνας, όπως περιστροφή, κλιμάκωση ή/και μετατόπιση και μπορεί να χρησιμοποιηθεί ως τοπικός περιγραφέας υφής. Η προτεινόμενη υλοποίηση σχεδιάστηκε έτσι, ώστε να εκμεταλλεύεται τη δομή και τις ιδιότητες των συσχετίσεων τρίτης τάξης (Κεφ. 2).

3.1 ΘΕΩΡΗΤΙΚΗ ΑΝΑΛΥΣΗ ΤΗΣ ΑΝΑΠΑΡΑΣΤΑΣΗΣ

Έστω διδιάστατο σήμα $x(\mathbf{t})$ και η τρίτης τάξης συσχέτισή του $x_3(\tau_1, \tau_2)$, της οποίας οι δείκτες (τ_1, τ_2) είναι δυο διδιάστατα διανύσματα που εκτείνονται σε όλο το $S' = [-(N-1), \dots, N-1] \times [-(N-1), \dots, N-1]$. Για κάθε (τ_1, τ_2) , μπορεί να οριστεί ένα νοητό τρίγωνο πάνω στο $x(\mathbf{t})$. Οι κορυφές του τριγώνου συμπίπτουν με τρία στοιχεία του σήματος. Αν το τρίγωνο αυτό μετακινηθεί, χωρίς να αλλάξει το μέγεθος και ο προσανατολισμός του, σε όλες τις δυνατές θέσεις πάνω στο σήμα και σε καθεμία διαφορετική θέση (συμπεριλαμβανόμενης και της αρχικής) υπολογισθεί το γινόμενο των εκάστοτε στοιχείων των κορυφών και, στο τέλος, αθροιστούν όλα τα γινόμενα, τότε έχουμε την τιμή $x_3(\tau_1, \tau_2)$.

Ορίζεται, έτσι, το σύνολο $W(\tau_1, \tau_2)$, το οποίο περιέχει όλα τα τρίγωνα που δημιουργήθηκαν από τις επανατοποθετήσεις του αρχικού τριγώνου (τ_1, τ_2) . Ορίζεται, επίσης, το σύνολο $K(\tau_1, \tau_2)$, το οποίο περιέχει όλα τα τρίγωνα που είναι όμοια¹ με αυτά του συνόλου $W(\tau_1, \tau_2)$. Στη συνέχεια, για κάθε σύνολο $K(\tau_1, \tau_2)$ αντιστοιχίζε-

¹Δύο τρίγωνα χαρακτηρίζονται ως όμοια, αν κάθε γωνία του ενός τριγώνου έχει το ίδιο μέτρο με την αντίστοιχη γωνία στο άλλο τρίγωνο, ή, ισοδύναμα, αν ο λόγος των μηκών των πλευρών τους είναι ίδιος και για τις τρεις πλευρές.

ται μία κλάση $C(\tau_1, \tau_2)$ με στοιχεία της συσχέτισης. Κάθε κλάση $C(\tau_1, \tau_2)$ περιέχει όλα εκείνα τα στοιχεία της συσχέτισης, των οποίων οι δείκτες δημιουργούν τρίγωνο όμοιο με το αρχικό τρίγωνο των διανυσμάτων (τ_1, τ_2) .

Έχοντας στην κατοχή μας τα νεοσυσταθέντα σύνολα και κλάση, παρατηρούμε πως, για διαφορετικούς δείκτες (τ_1, τ_2) και (τ_1', τ_2') , μπορούν να δημιουργηθούν πανομοιότυπες κλάσεις $C(\tau_1, \tau_2)$ και $C(\tau_1', \tau_2')$, είτε γιατί τα (τ_1, τ_2) και (τ_1', τ_2') δημιουργούν όμοια αρχικά τρίγωνα, είτε γιατί πολλά στοιχεία της συσχέτισης είναι ίσα, λόγω των συμμετριών που ισχύουν (εξ. 2.10). Προκειμένου να αποφευχθεί αυτή η κατάσταση, δημιουργούμε τις κλάσεις $C(\tau_1, \tau_2)$ κρατώντας σταθερό το τ_1 στη θέση $\tau_1 = [1, 0]$ και επιτρέποντας στο τ_2 να παίρνει τιμές από το υποσύνολο $S_0 \triangleq [1, 0] \times [0, \infty]$.

Αφού έχει εξασφαλισθεί πως οι κλάσεις $C(\tau_1, \tau_2)$ δεν περιέχουν επαναλαμβανόμενα στοιχεία, βλέπουμε πως οποιαδήποτε αλλαγή του $x(t)$ (περιστροφή ή κλιμάκωση) οδηγεί σε επανατοποθέτηση των στοιχείων της κάθε κλάσης $C(\tau_1, \tau_2)$ μέσα στην ίδια κλάση, ενώ οι κλάσεις ως σύνολα δεν αλλάζουν και συνεχίζουν να περιέχουν τα ίδια στοιχεία. Έτσι, ορίζεται η παρακάτω αναπαράσταση για τα στοιχεία της κάθε κλάσης:

$$\tilde{x}_3(\rho, \phi; \tau_1, \tau_2) \triangleq x_3(\mathbf{T}_{\beta, \phi} \tau_1, \mathbf{T}_{\beta, \phi} \tau_2) \quad (3.1)$$

όπου,

$$\mathbf{T}_{\beta, \phi} = \beta \begin{bmatrix} \cos \phi & -\sin \phi \\ \sin \phi & \cos \phi \end{bmatrix} \text{ και } \rho = \log \beta$$

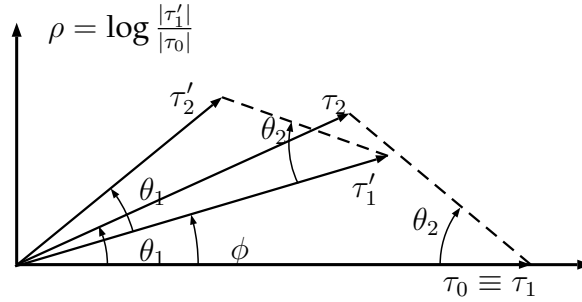
Εισάγονται οι μεταβλητές ρ και ϕ για να αναπαρασταθούν όλα τα τρίγωνα της κλάσης $C(\tau_1, \tau_2)$, τα οποία έχουν υποστεί συστολή/διαστολή β ή/και περιστροφή κατά γωνία ϕ .

Αν θεωρηθεί λοιπόν το σήμα $y(t) = x(\mathbf{T}_{\alpha, \theta} \mathbf{t} + \mathbf{t}_0)$, τότε στο εσωτερικό κάθε κλάσης $C(\tau_1, \tau_2)$ η αναπαράστασή του γίνεται:

$$\tilde{y}_3(\rho, \phi; \tau_1, \tau_2) = \tilde{x}_3(\rho + \log \alpha, \phi + \theta; \tau_1, \tau_2) \quad (3.2)$$

Στη συνέχεια, από την αναπαράσταση του \tilde{x}_3 , υπολογίζεται ο διδιάστατος μετασχηματισμός Fourier ως προς τις μεταβλητές ρ και ϕ , προκειμένου να αποβληθεί η επιρροή του παράγοντα περιστροφής ή/και κλιμάκωσης. Πλέον έχει παραχθεί το $\tilde{X}_3(P, \Phi; \tau_1, \tau_2)$. Παράλληλα, γίνονται κάποιες αλλαγές και ενέργειες, προκειμένου να προκύψει μια αντιπροσωπευτική αναπαράσταση του σήματος, η οποία είναι όσο πιο οικονομική γίνεται, από άποψη όγκου δεδομένων αλλά και από άποψη υπολογιστικών απαιτήσεων για τη δημιουργία της.

Η πρώτη ενέργεια για την επίτευξη των παραπάνω είναι το να συμπεριληφθεί στην τελική αναπαράσταση μόνο το πλάτος του μετασχηματισμού $\tilde{X}_3(P, \Phi; \tau_1, \tau_2)$ και όχι φάση του. Επίσης μία επιπλέον επιλογή είναι να κρατηθεί μόνο η μηδενική



Σχήμα 3.1 – Σύστημα Αναπαράστασης Διανυσμάτων με τις γωνίες θ_1, θ_2 .

συχνότητα του πλάτους. Υπολογίζεται, δηλαδή το $|\tilde{X}_3(0, 0; \tau_1, \tau_2)|$. Στη συνέχεια υιοθετείται μία καινούρια έκφραση για τους δείκτες (τ_1, τ_2) . Αντί για δύο διανύσματα $\tau_1 = [\tau_{1x}, \tau_{1y}]$ και $\tau_2 = [\tau_{2x}, \tau_{2y}]$, μπορεί να χρησιμοποιηθεί ένα από παρακάτω συστήματα:

1. $(\tau_1, \tau_2) = (\tau_0, [k, l]), k \in [0, 1], l \in [0, \infty]$
2. $(\tau_1, \tau_2) = (\tau_0, \tau_2) \rightarrow (\theta_1, \theta_2), \theta_1, \theta_2 \in \left[0, \frac{\pi}{2}\right]$
3. $(\tau_1, \tau_2) = (\tau_0, \tau_2) \rightarrow (\theta_1, \lambda), \theta_1 \in \left[0, \frac{\pi}{2}\right], \lambda \in [0, \infty]$

όπου $\tau_0 = (1, 0)$, οι θ_1, θ_2 είναι οι γωνίες των διανυσμάτων (τ_1, τ_2) και $(\tau_1, \tau_2 - \tau_1)$ αντίστοιχα και λ είναι ο λόγος $\frac{|\tau_1|}{|\tau_2|}$. Σε όλες τις παραπάνω περιπτώσεις ο τετραδιάστατος χώρος, που εκφράζεται από τα (τ_1, τ_2) , μειώνεται αισθητά σε ένα διδιάστατο χώρο, χωρίς καθόλου απώλεια πληροφορίας.

Στην παρούσα διπλωματική εργασία έχει υιοθετηθεί το δεύτερο σύστημα, το οποίο φαίνεται γραφικά στο Σχήμα 3.1.

Είναι σημαντικό να σημειωθεί, ότι κατά τον υπολογισμό των μεταβλητών ρ, ϕ, θ_1 και θ_2 , επειδή οι τιμές τους είναι πολύ πιθανό να είναι πραγματικοί αριθμοί, υπόκεινται σε κβάντιση, προκειμένου να μπορούν να αντιστοιχηθούν ένα διακριτό πλέγμα. Ο βέλτιστος αριθμός επιπέδων κβάντισης για την κάθε μεταβλητή είναι ένας προβληματισμός, ο οποίος πρέπει να μελετηθεί στο πειραματικό στάδιο.

Επίσης σημαντικό είναι, το γεγονός ότι, το $\tilde{x}_3(\rho, \phi; \tau_1, \tau_2)$ δεν υπολογίζεται απευθείας από την εξίσωση 3.1. Αντί να υπολογισθεί το $x_3(\mathbf{T}_{\beta, \phi} \tau_1, \mathbf{T}_{\beta, \phi} \tau_2)$ για κάθε ρ, ϕ , υπολογίζεται πρώτα το $x_3(\tau_1, \tau_2)$ και μετά η τιμή που υπολογίστηκε αντιστοιχίζεται σε κάποια ρ, ϕ . Επίσης η τιμή αντιστοιχίζεται σε κάποια κλάση, ανάλογα με το ποιο από τα τρία προαναφερθέντα συστήματα χρησιμοποιείται. Έτσι, τελικά προστίθεται αυτή η τιμή στο κατάλληλο στοιχείο του \tilde{x}_3 . Επιπλέον, για κάθε $\tau_1 \in S'$,

το τ_2 δε χρειάζεται να παίρνει τιμές από όλο το S' . Για την ακρίβεια, μπορεί να παίρνει τιμές από το:

$$S''(\tau_1) = \{\tau_2 \in S' : 0 \leq \tau_1 \cdot \tau_2 \leq |\tau_1|^2, 0 \leq \tau_1 \cdot \tau_2', (\tau_1 - \tau_2) \in S'\} \quad (3.3)$$

όπου το τ_2' είναι το διάνυσμα τ_2 περιστραμμένο με την ωρολογιακή φορά κατά 90 μοίρες. Το παραπάνω σύνολο $S''(\tau_1)$ προκύπτει αν λάβουμε υπόψιν τις συμμετρίες των συσχετίσεων (προκύπτουν οι δύο πρώτοι περιορισμοί) και το γεγονός ότι το $x_3(\tau_1, \tau_2)$ είναι μηδέν όταν $(\tau_1 - \tau_2) \notin S'$. Ο περιορισμός αυτός του τ_2 απλοποιεί σε πολύ μεγάλο βαθμό την υπολογιστική διαδικασία.

Τελικά το $|\tilde{X}_3(0, 0; \theta_1, \theta_2)|$, αποτελεί τον τοπικό περιγραφέα της γειτονιάς, η οποία έχει τροφοδοτηθεί στον αλγόριθμο.

3.2 ΑΛΓΟΡΙΘΜΙΚΗ ΥΛΟΠΟΙΗΣΗ ΤΗΣ ΑΝΑΠΑΡΑΣΤΑΣΗΣ

Σε αυτή την ενότητα παρατίθενται οι αλγοριθμικές μορφές της υλοποίησης της αναπαράστασης και του υπολογισμού της συσχέτισης τρίτου βαθμού μιας εικόνας.

Αλγόριθμος 3.1 Υπολογισμός αμετάβλητης αναπαράστασης εικόνας.

- 1: variable initialization;
 - 2: calculate $\max(\rho)$, $\max(\phi)$ for quantization;
 - 3: $x_3 \leftarrow 3rdOrderCorrelationOf(x)$;
 - 4: **for** $\tau_1 \in S'$ **do**
 - 5: **for** $\tau_2 \in S''(\tau_1)$ **do**
 - 6: calculate & quantize ρ, ϕ by comparing τ_1 with $\tau_0 = [1, 0]$;
 - 7: calculate & quantize included angles θ_1, θ_2 ;
 - 8: $\tilde{x}_3(\rho, \phi; \theta_1, \theta_2) \leftarrow \tilde{x}_3(\rho, \phi; \theta_1, \theta_2) + x_3(\tau_1, \tau_2)$;
 - 9: **end for**
 - 10: **end for**
 - 11: $\tilde{X}_3(\rho, \phi; \theta_1, \theta_2) \leftarrow FFTAlongFirstTwoDimensionsOf(\tilde{x}_3(\rho, \phi; \theta_1, \theta_2))$
 - 12: $Representation \leftarrow |\tilde{X}_3(0, 0; \theta_1, \theta_2)|$
-

Αλγόριθμος 3.2 Υπολογισμός τρίτης τάξης συσχέτισης εικόνας.

- 1: variable initialization;
 - 2: $X \leftarrow FFT2Of(x)$;
 - 3: **for** $\tau_1 \in S'$ **do**
 - 4: **for** $\tau_2 \in S'$ **do**
 - 5: $X_3(\tau_1, \tau_2) \leftarrow X(\tau_1)X(\tau_2)X^*(\tau_1 + \tau_2)$
 - 6: **end for**
 - 7: **end for**
 - 8: $x_3 \leftarrow iFFTnOf(X_3)$
-

Πίνακας 3.1 – Χρόνοι παραγωγής της προτεινόμενης αναπαράστασης

$ \theta_1 , \theta_2 $	<i>Radius</i>	<i>N</i>	<i>Time(msec)</i>
9	4	9	5.851
9	8	17	85.771
10	4	9	5.8785
10	8	17	86.326

3.3 ΠΟΛΥΠΛΟΚΟΤΗΤΑ

Ένα πολύ σημαντικό ζήτημα που πρέπει να αναλυθεί έχει να κάνει με την πολυπλοκότητα του αλγόριθμου που υλοποιήθηκε καθώς και με το χρόνο που απαιτείται για την παραγωγή μιας αναπαράστασης. Από τους αλγόριθμους της προηγούμενης ενότητας, φαίνεται, πως την πολυπλοκότητα καθορίζουν ο υπολογισμός της συσχέτισης, ο μετασχηματισμός Fourier, καθώς και οι δύο επαναληπτικοί βρόχοι.

Από τον Αλγόριθμο 3.2 μπορεί να υπολογισθεί η πολυπλοκότητα του υπολογισμού της συσχέτισης. Για το μετασχηματισμό Fourier (fft), τόσο τον ευθύ, όσο και τον αντίστροφο, γνωρίζουμε πως έχει πολυπλοκότητα $O(N \log N)$. Έπειτα, είναι οι δύο επαναληπτικοί βρόχοι, προκειμένου τα (τ_1, τ_2) να καλύψουν όλο το $S' = [-(N-1), \dots, N-1] \times [-(N-1), \dots, N-1]$, όπου N , το μέγεθος της εικόνας εισόδου. Στην ουσία ο κάθε επαναληπτικός βρόχος είναι δύο εμφωλευμένοι βρόχοι, όπου αν θεωρήσουμε πως τα (τ_1, τ_2) είναι διανύσματα δύο διαστάσεων, τότε η κάθε συνιστώσα τους εκτείνεται σε εύρος $[-(N-1), \dots, N-1]$. Υπάρχουν δηλαδή, τέσσερις επαναληπτικοί βρόχοι, ο καθένας από τους οποίους έχει $2N-1$ επαναλήψεις. Άρα τελικά, εκτελούνται $(2N-1) \times (2N-1) \times (2N-1) \times (2N-1) = 16N^4 - 32N^3 + 24N^2 - 8N + 1$ επαναλήψεις, δηλαδή μιλάμε για πολυπλοκότητα $O(N^4)$ για τους επαναληπτικούς βρόχους, αλλά και για τον αλγόριθμο υπολογισμού της συσχέτισης συνολικά.

Επιστρέφοντας στον Αλγόριθμο 3.1, φαίνεται πως οι δύο ακόμα επαναληπτικοί βρόχοι δίνουν μια πολυπλοκότητα $O(N^4)$, η οποία υπερσχύει της πολυπλοκότητας του fft $O(N \log N)$. Θα πρέπει να σημειωθεί, βέβαια, πως στην περίπτωση του fft του βήματος (11), το ρόλο του N παίζουν οι παράμετροι (θ_1, θ_2) , αφού υπολογίζοντας το μετασχηματισμό Fourier του $\tilde{x}_3(\rho, \phi, \theta_1, \theta_2)$ για κάθε (ρ, ϕ) , στην ουσία εκτελείται ένας διδιάστατος μετασχηματισμός Fourier σε σήμα διαστάσεων $(\theta_1 + 1) \times (\theta_2 + 1)$. Άρα τελικά για είσοδο μία εικόνα διαστάσεων $(N \times N)$, η πολυπλοκότητα για τη δημιουργία της αναπαράστασής της είναι $O(N^4)$.

Στον Πίνακα 3.1 φαίνονται ενδεικτικά οι μέσοι χρόνοι για την παραγωγή των αναπαραστάσεων για εικόνες με μέγεθος ίδιο με αυτό που επιλέχθηκε για τα πειράματα, καθώς και για τις τιμές $(|\theta_1|, |\theta_2|)$ που χρησιμοποιήθηκαν.

3.4 ΠΑΡΑΔΕΙΓΜΑΤΑ

Σε αυτή την ενότητα παρατίθενται κάποια παραδείγματα, τα οποία, έστω και σε πρωταρχικό στάδιο, επιδεικνύουν τη λειτουργία και την ανεξαρτησία της αναπαράστασης. Συγκεκριμένα χρησιμοποιήθηκαν ως είσοδοι τεχνητές εικόνες που απεικονίζουν κάποια βασικά γεωμετρικά σχήματα. Αφού υπολογισθούν οι αναπαραστάσεις τους, τα σχήματα περιστρέφονται και υπολογίζονται οι καινούριες αναπαραστάσεις. Επίσης έγιναν και κάποιες δοκιμές ως προς τον αριθμό των επιπέδων κβάντισης των τεσσάρων μεταβλητών της αναπαράστασης.

Όπως φαίνεται από τα παραδείγματα των Σχημάτων 3.2-3.5, οι αναπαραστάσεις της κάθε εικόνας και της αντίστοιχης περιστραμμένης είναι σχεδόν ίδιες. Υπάρχουν μικρές διαφοροποιήσεις, οι οποίες οφείλονται στην αλλοίωση που υφίστανται οι εικόνες κατά την περιστροφή τους. Όπως επιβεβαιώνεται, άλλωστε, στην περίπτωση του τριγώνου και του ορθογώνιου, οι αναπαραστάσεις για τις περιστραμμένες κατά 90 μοίρες εικόνες, όπου δεν υπάρχει αλλοίωση, ταυτίζονται απόλυτα.

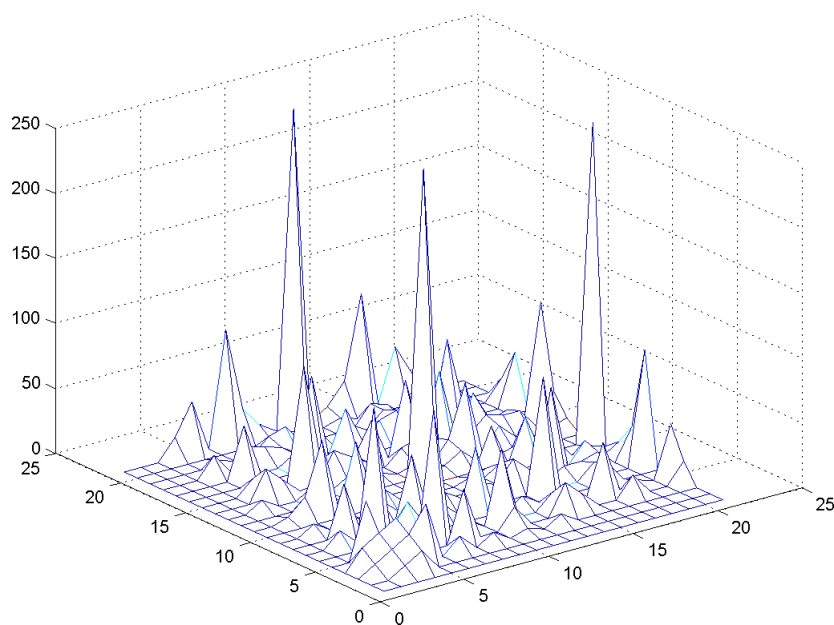
Στη συνέχεια παρατίθενται τρεις αναπαραστάσεις του τετραγώνου (Σχήμα 3.2α') για διαφορετικό αριθμό επιπέδων κβάντισης για τα $\rho, \phi, \theta_1, \theta_2$ (Σχήματα 3.6). Σε αυτές τις αναπαραστάσεις είναι πιο εμφανείς οι διαφορές. Παρόλα αυτά, μπορούμε να παρατηρήσουμε πως υπάρχει μια στοιχειώδης δομή για τη συγκεκριμένη εικόνα εισόδου, η οποία παραμένει σταθερή και στις τρεις περιπτώσεις.



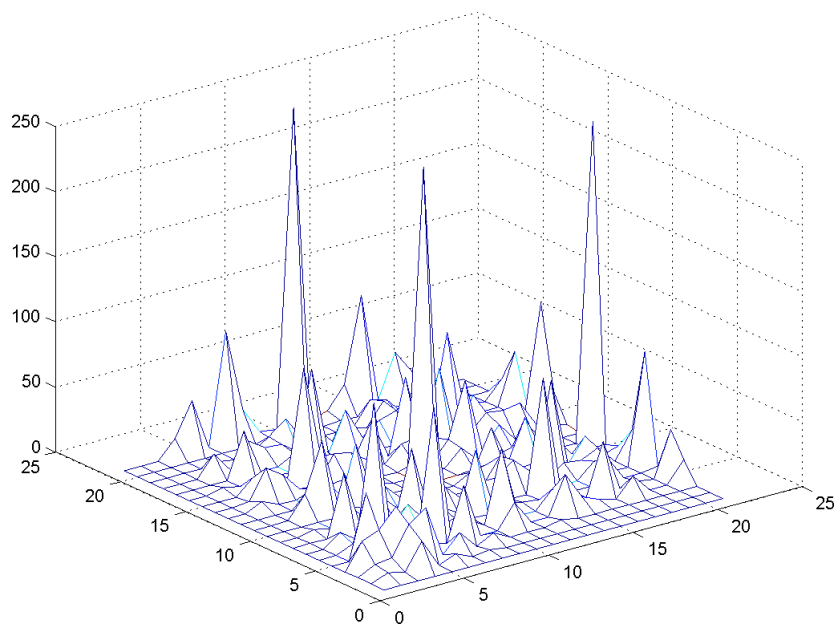
(α') Τετράγωνο



(β') Περιστραμμένο τετράγωνο

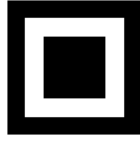


(γ') Αναπαράσταση τετραγώνου με αριθμό επιπέδων κβάντισης: $|\rho| = 15, |\phi| = 15, |\theta_1| = 20, |\theta_2| = 20$

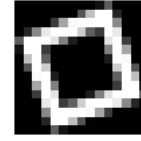


(δ') Αναπαράσταση περιστραμμένου τετραγώνου με αριθμό επιπέδων κβάντισης: $|\rho| = 15, |\phi| = 15, |\theta_1| = 20, |\theta_2| = 20$

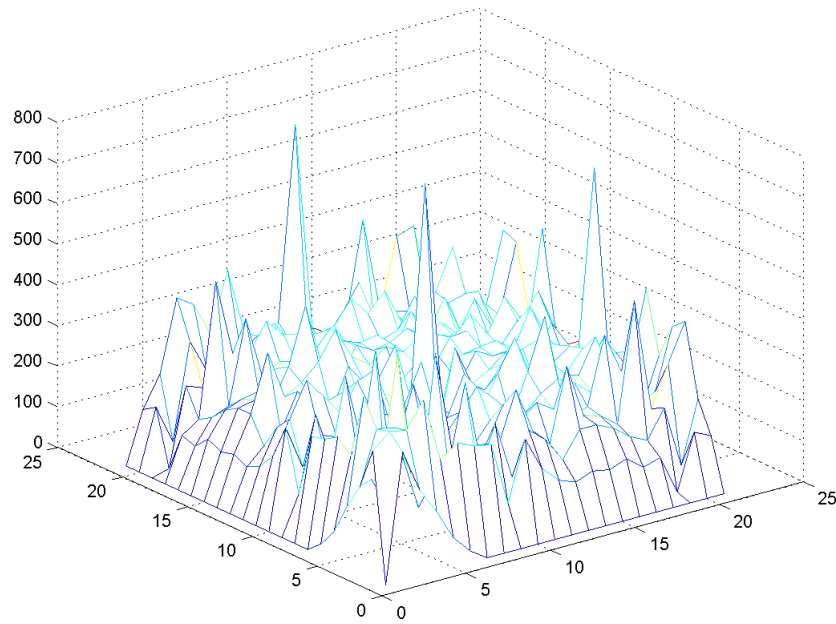
Σχήμα 3.2 – Παράδειγμα με τετράγωνο



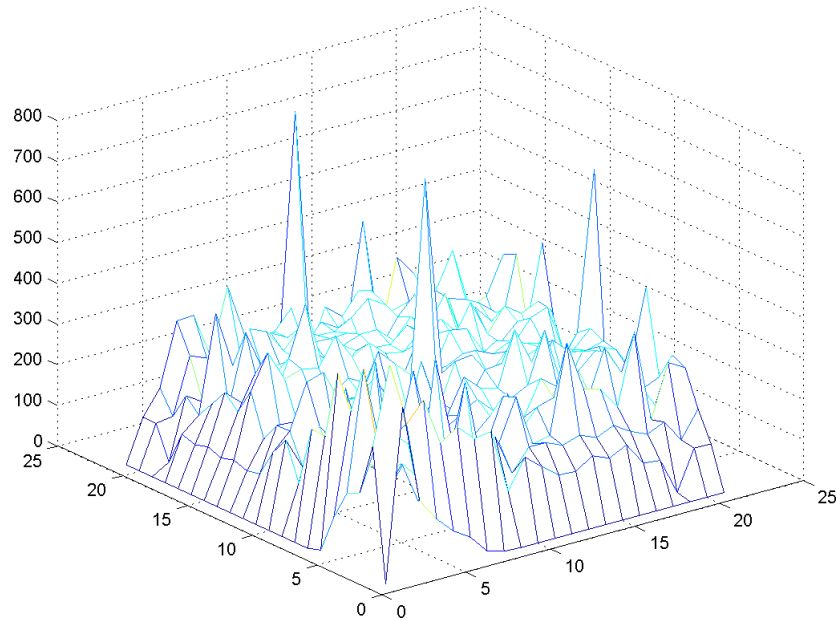
(α') Περίγραμμα τετραγώνου



(β') Περιστραμμένο περίγραμμα τετραγώνου



(γ') Αναπαράσταση περιγράμματος τετραγώνου με αριθμό επιπέδων κβάντισης: $|\rho| = 15, |\phi| = 15, |\theta_1| = 20, |\theta_2| = 20$



(δ') Αναπαράσταση περιστραμμένου περιγράμματος τετραγώνου με αριθμό επιπέδων κβάντισης: $|\rho| = 15, |\phi| = 15, |\theta_1| = 20, |\theta_2| = 20$

Σχήμα 3.3 – Παράδειγμα με περίγραμμα τετραγώνου



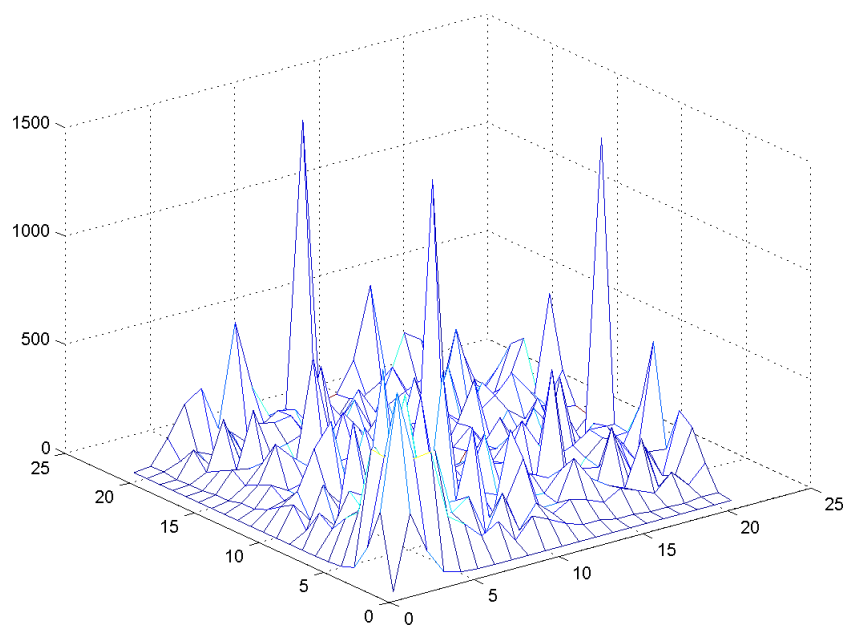
(α') Τρίγωνο



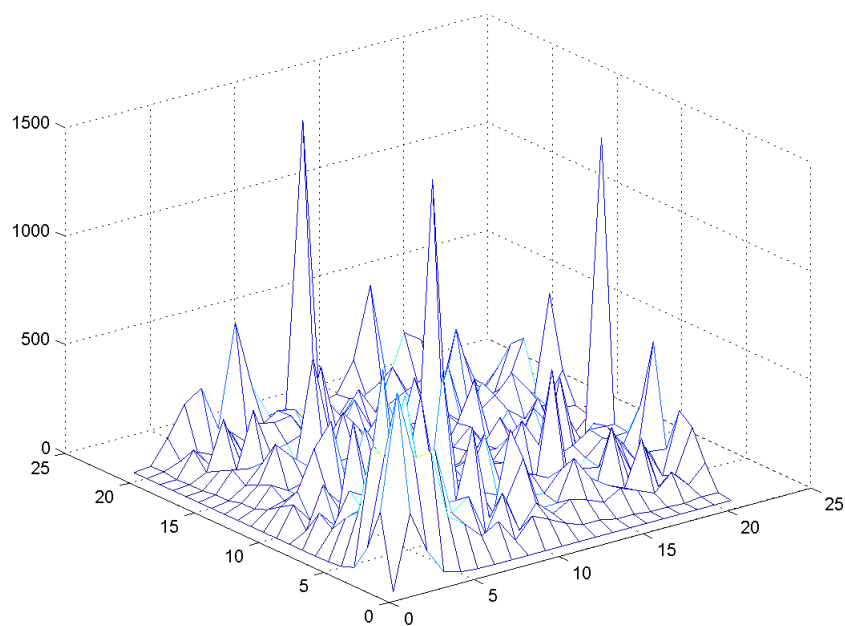
(β') Περιστραμμένο τρίγωνο κατά 90 μοίρες



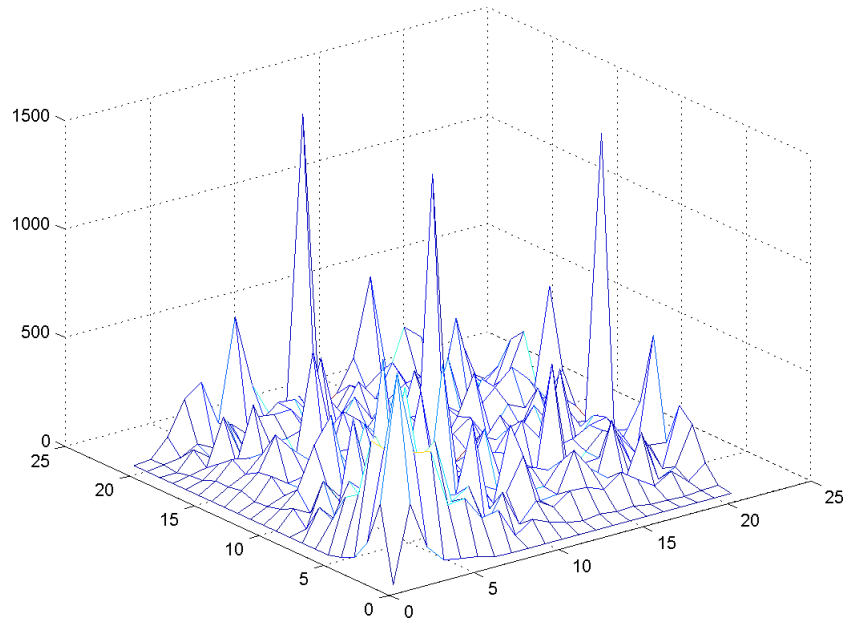
(γ') Περιστραμμένο τρίγωνο



(δ') Αναπαράσταση τριγώνου με αριθμό επιπέδων χβάντισης: $|\rho| = 15, |\phi| = 15, |\theta_1| = 20, |\theta_2| = 20$



(ε') Αναπαράσταση περιστραμμένου κατά 90 μοίρες τριγώνου με αριθμό επιπέδων χβάντισης: $|\rho| = 15, |\phi| = 15, |\theta_1| = 20, |\theta_2| = 20$



(στ') Αναπαράσταση περιστραμμένου τριγώνου με αριθμό επιπέδων κβάντισης: $|\rho| = 15, |\phi| = 15, |\theta_1| = 20, |\theta_2| = 20$

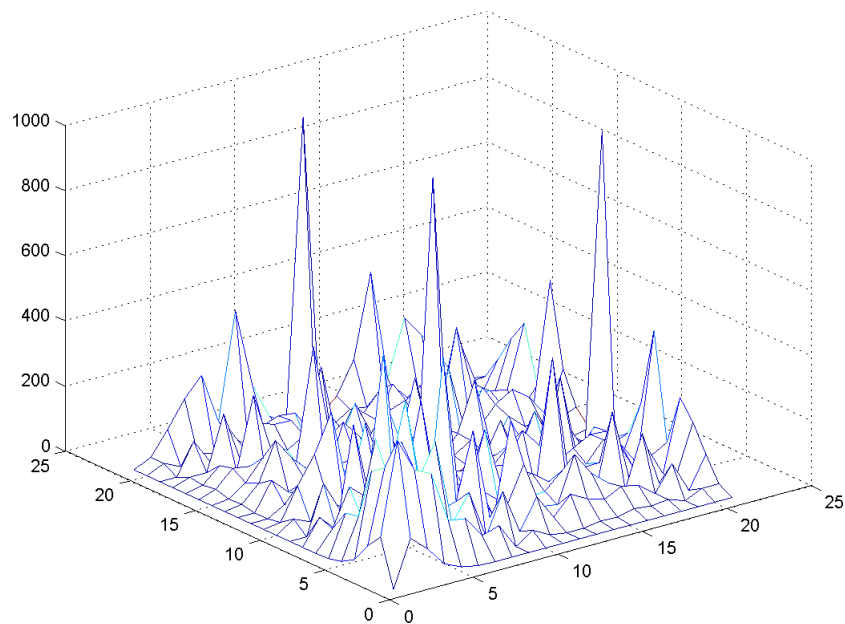
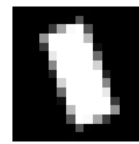
Σχήμα 3.4 – Παράδειγμα με τρίγωνο



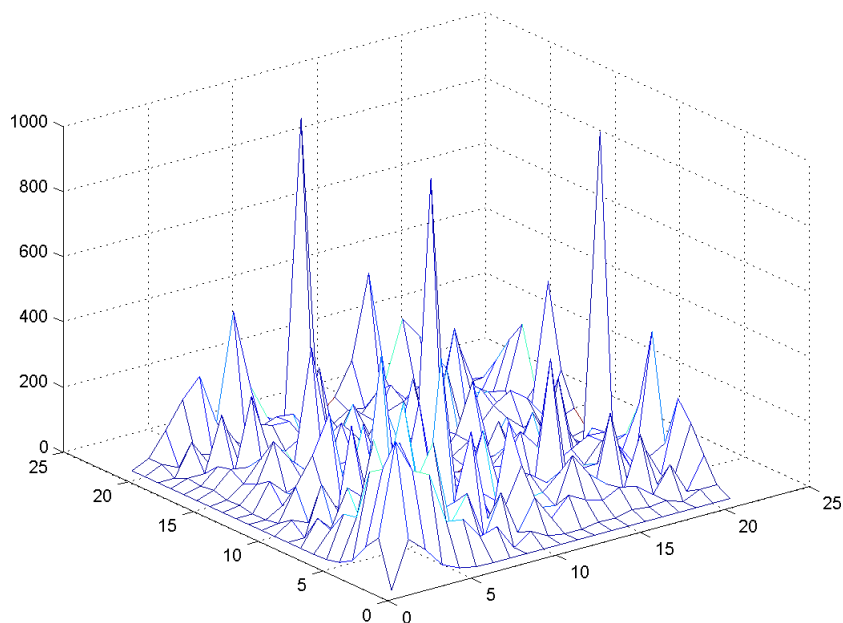
(α') Ορθογώνιο



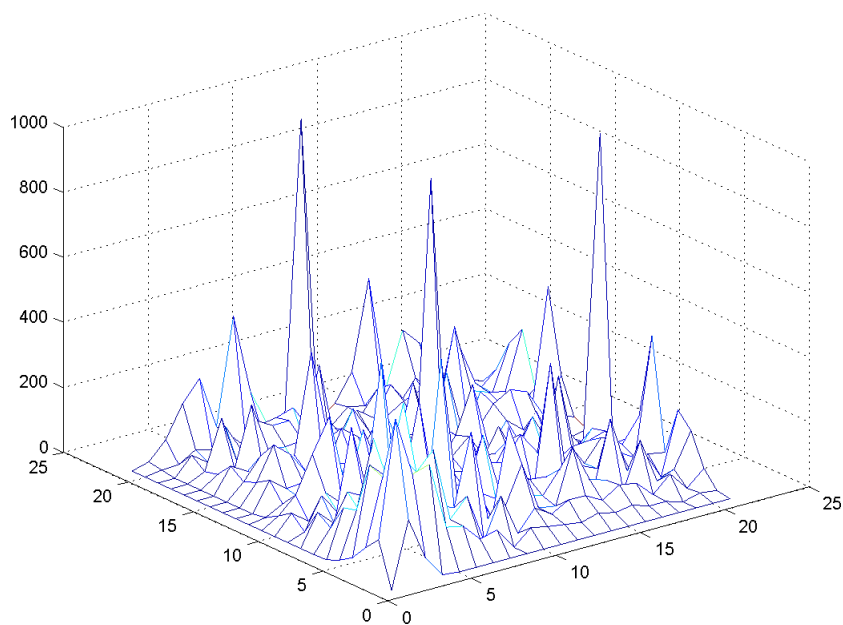
(β') Περιστραμμένο ορθο- (γ') Περιστραμμένο ορθο-
γώνιο κατά 90 μοίρες γώνιο



(δ') Αναπαράσταση ορθογωνίου με αριθμό επιπέδων κβάντισης: $|\rho| = 15, |\phi| = 15, |\theta_1| = 20, |\theta_2| = 20$

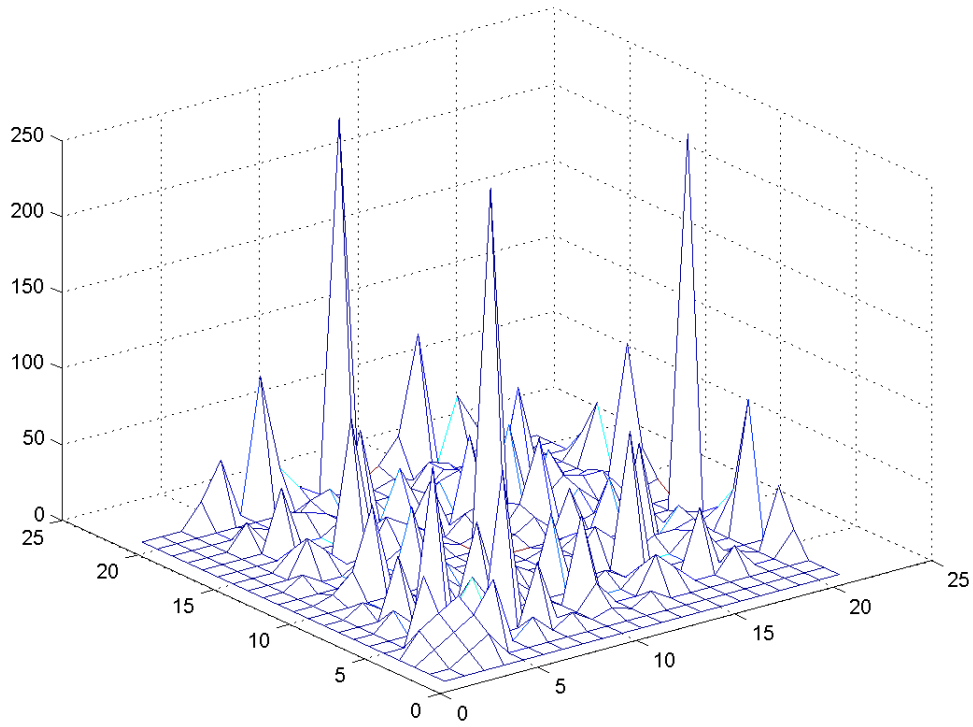


(ε') Αναπαράσταση περιστραμμένου κατά 90 μοίρες ορθογωνίου με αριθμό επιπέδων χβάντισης: $|\rho| = 15, |\phi| = 15, |\theta_1| = 20, |\theta_2| = 20$

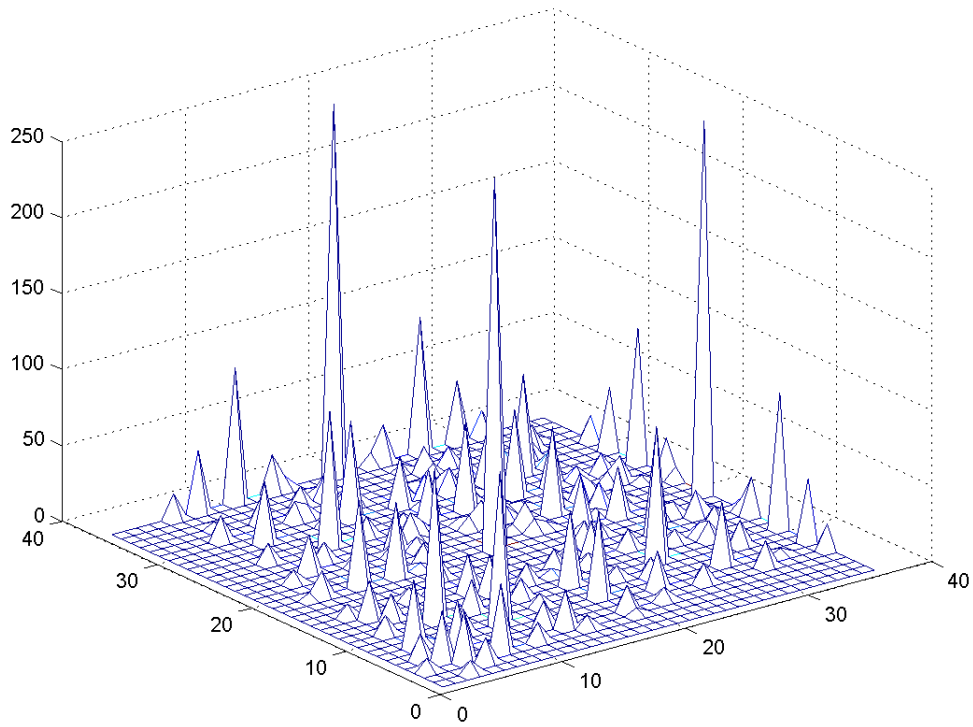


(στ') Αναπαράσταση περιστραμμένου ορθογωνίου με αριθμό επιπέδων χβάντισης: $|\rho| = 15, |\phi| = 15, |\theta_1| = 20, |\theta_2| = 20$

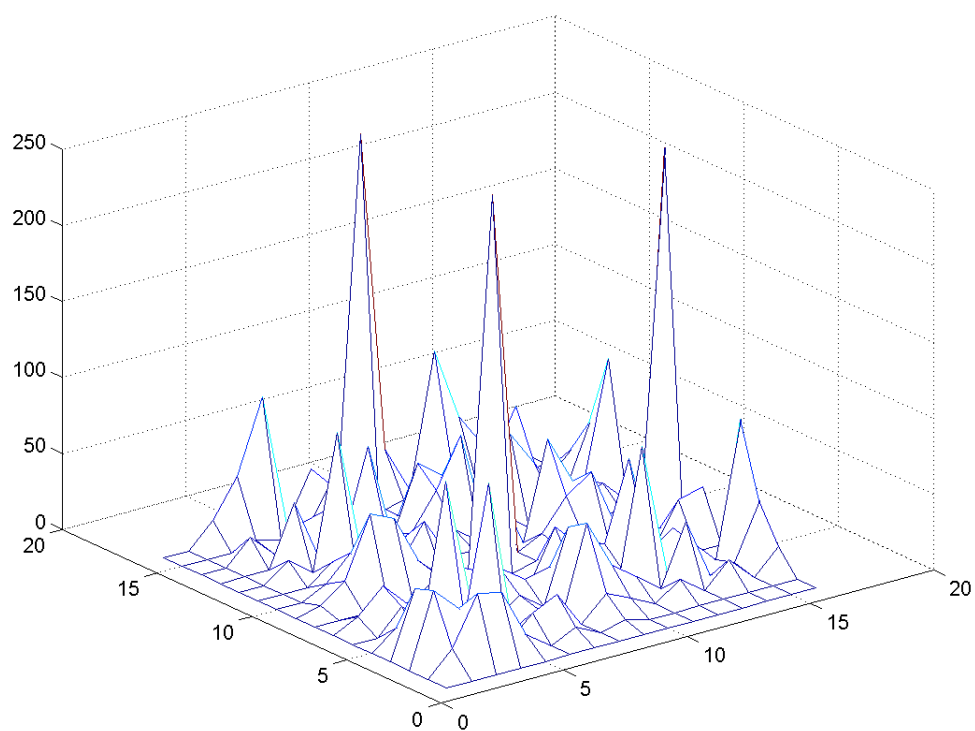
Σχήμα 3.5 – Παράδειγμα με ορθογώνιο



(α') Αναπαράσταση τετραγώνου με αριθμό επιπέδων κβάντισης: $|\rho| = 15, |\phi| = 15, |\theta_1| = 20, |\theta_2| = 20$



(β') Αναπαράσταση τετραγώνου με αριθμό επιπέδων κβάντισης: $|\rho| = 25, |\phi| = 25, |\theta_1| = 35, |\theta_2| = 35$



(γ') Αναπαράσταση τετραγώνου με αριθμό επιπέδων κβάντισης: $|\rho| = 5, |\phi| = 5, |\theta_1| = 15, |\theta_2| = 15$

Σχήμα 3.6 – Αναπαράσταση τετραγώνου για διαφορετικό αριθμό επιπέδων κβάντισης των $\rho, \phi, \theta_1, \theta_2$

3.5 ΥΛΟΠΟΙΗΣΗ ΜΟΝΤΕΛΟΥ BAG OF WORDS

Στην Παράγραφο 2.1.2 σημειώθηκε πως, προκειμένου να ισχύουν οι ιδιότητες των συσχετίσεων, είναι απαραίτητο να μην υπάρχει απώλεια πληροφορίας της εικόνας, κατά το μετασχηματισμό της. Στα παραδείγματα της προηγούμενης παραγράφου, όπου παράχθηκε η αναπαράσταση τεχνητών εικόνων, λήφθηκε υπόψιν αυτή η απαίτηση. Έτσι, οι εικόνες τοποθετήθηκαν σε ένα μαύρο πλαίσιο και, όπως φαίνεται και στα αντίστοιχα σχήματα, κατά την περιστροφή τους δεν υπερέβαιναν αυτό το πλαίσιο και έτσι δεν αποκόπηκαν.

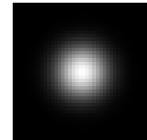
Όταν, όμως, πρόκειται για πραγματικές εικόνες, είναι πολύ σπάνιο, αν όχι αδύνατο, αυτές να περιβάλλονται από πλαίσιο. Ακόμα πιο δύσκολο είναι να περιβάλλεται από μαύρο πλαίσιο κάποια συγκεκριμένη γειτονιά της εικόνας, για την οποία θα πρέπει να παραχθεί τοπικός περιγραφέας. Προκειμένου, λοιπόν να αναπαρασταθεί μια εικόνα μέσω του μοντέλου Bag of Words, χρησιμοποιώντας την προτεινόμενη αναπαράσταση για τη δημιουργία τοπικών περιγραφέων, θα πρέπει το πλαίσιο να προστεθεί στην εικόνα τεχνητά. Αυτό γίνεται με χρήση της γκαουσιανής συνάρτησης, συγκεκριμένα της διδιάστατης γκαουσιανής συνάρτησης.

Ο τύπος για τη διδιάστατη γκαουσιανή συνάρτηση είναι:

$$f(x, y) = \alpha e^{-\left(\frac{(x - x_0)^2}{2\sigma_x^2} + \frac{(y - y_0)^2}{2\sigma_y^2}\right)} \quad (3.4)$$

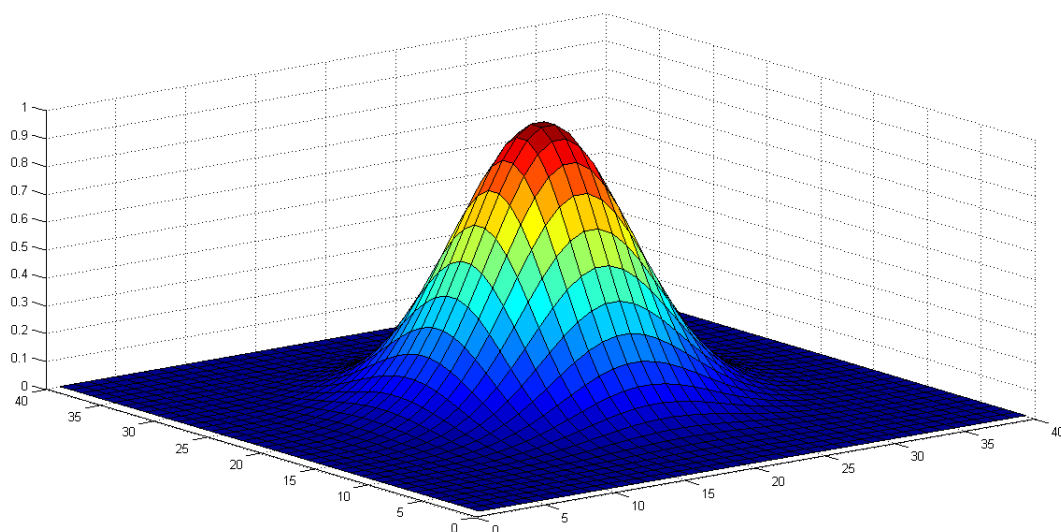
όπου $\alpha, x_0, y_0, \sigma_x, \sigma_y \in \mathbf{R}$. Όπως φαίνεται στο Σχήμα 3.7, το γράφημα της συνάρτησης έχει το σχήμα καμπάνας, είναι δηλαδή συμμετρικό και τείνει προς το μηδέν ασυμπτωτικά. Το α ορίζει το ύψος την επιφάνειας, τα (x_0, y_0) , ορίζουν το κέντρο και τα (σ_x, σ_y) ορίζουν τα πλάτη της καμπάνας ως προς τους άξονες x και y αντίστοιχα. Συγκεκριμένα, όσο πιο μικρό είναι το σ , τόσο πιο γρήγορα φτάνουμε σε σχεδόν μηδενικές τιμές στην αντίστοιχη διάσταση.

Προγραμματιστικά, όταν υπολογίζεται η γκαουσιανή συνάρτηση στις δύο διαστάσεις, στην ουσία αυτό που υπολογίζεται είναι ένας διδιάστατος πίνακας, του οποίου το στοιχείο που αντιστοιχεί στις συντεταγμένες (x_0, y_0) περιέχει την τιμή α και οι τιμές των υπόλοιπων στοιχείων μειώνονται όσο μεγαλώνει η απόσταση από το (x_0, y_0) ακτινικά, έως ότου, σε μεγάλες αποστάσεις, σχεδόν να μηδενιστούν. Αν ο πίνακας αυτός αντιμετωπιστεί ως εικόνα², τότε το αποτέλεσμά είναι μία εικόνα αποχρώσεων του γκρι (grayscale), η οποία στο σημείο (x_0, y_0) έχει τη μέγιστη φωτεινότητα (λευκό χρώμα), η οποία μειώνεται σταδιακά ανάλογα με την απόσταση από το σημείο (x_0, y_0) , μέχρι να φτάσει σε σχεδόν μηδενική φωτεινότητα (μαύρο χρώμα) (Σχ. 3.8). Ο πίνακας που μόλις περιγράφηκε μπορεί να πολλαπλασιαστεί στοιχείο προς στοιχείο



Σχήμα 3.8 – Γκαουσιανή συνάρτηση ως εικόνα

²Ο πίνακας πρέπει να υποστεί κανονικοποίηση στη μονάδα, αν $\alpha \neq 1$.



Σχήμα 3.7 – Γράφημα της Γκαουσιανής συνάρτησης στις δύο διαστάσεις

με τα στοιχεία κάποιας εικόνας (έγχρωμης ή/και grayscale) και, ανάλογα με τα επιλεγμένα σ_x, σ_y και (x_0, y_0) , να μείνει εμφανές κάποιο κομμάτι της εικόνας και γύρω από το κομμάτι αυτό η εικόνα να μαυρίσει (Σχήμα 3.9). Ιδιαίτερη προσοχή χρειάζεται στην περίπτωση έγχρωμων εικόνων, όπου θα πρέπει ο πίνακας της γκαουσιανής να πολλαπλασιαστεί ξεχωριστά με κάθε χρωματική συνιστώσα της εικόνας. Για παράδειγμα, αν χρησιμοποιείται το χρωματικό μοντέλο RGB (Red Green Blue), τότε η εικόνα αναπαρίσταται από έναν πίνακα τριών διαστάσεων. Οι δυο πρώτες διαστάσεις αφορούν στη θέση κάθε σημείου πάνω στην εικόνα και η τρίτη διάσταση (με μήκος τρία) αφορά στη χρωματική συνιστώσα. Είναι δηλαδή στην ουσία, τρεις ξεχωριστοί διδιάστατοι πίνακες ίδιας διάστασης με την εικόνα, όπου ο κάθε πίνακας περιέχει τις τιμές της κάθε ξεχωριστής χρωματικής συνιστώσας. Σε αυτήν την περίπτωση θα πρέπει ο πίνακας της γκαουσιανής συνάρτησης να πολλαπλασιαστεί ξεχωριστά με καθέναν από τους τρεις πίνακες των χρωμάτων και το αποτέλεσμα αυτών των πολλαπλασιασμών θα συνθέτει την τελική εικόνα.

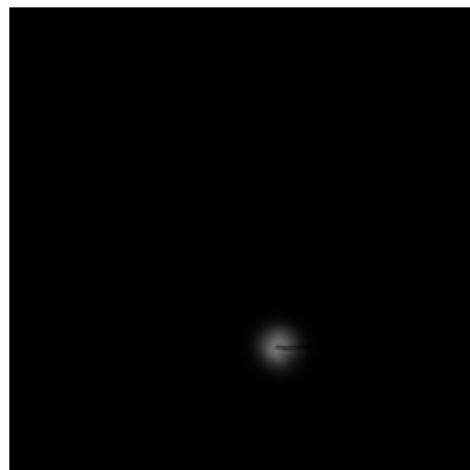
Έτσι, για την παραγωγή των τοπικών περιγραφέων πάνω στο πλέγμα που έχει εφαρμοστεί στην εικόνα, αυτό που γίνεται είναι να πολλαπλασιάζεται η εικόνα με την γκαουσιανή, η οποία είναι κεντραρισμένη κάθε φορά στον κόμβο του οποίου ο περιγραφέας πρέπει να δημιουργηθεί. Τα σ_x, σ_y επιλέγονται κατάλληλα, έτσι ώστε η γειτονιά που μένει εμφανής να έχει το μέγεθος που έχει επιλεγθεί. Στο Σχήμα 3.10 φαίνεται η διαδικασία που περιγράφηκε, όπου το κέντρο της γειτονιάς επιλέχθηκε με γραφικό τρόπο.



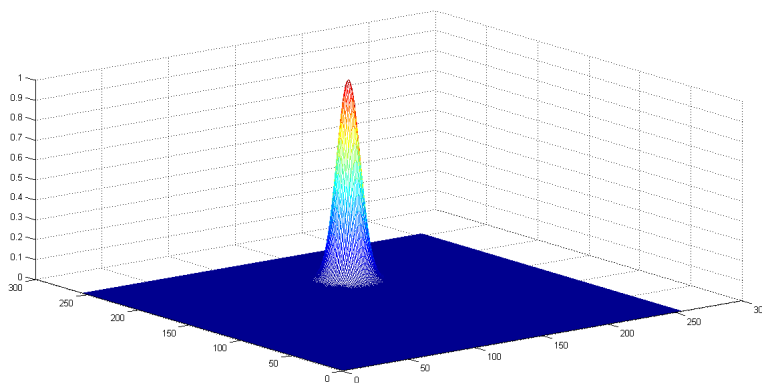
Σχήμα 3.9 – Εικόνα πολλαπλασιασμένη με γκαουσιανή συνάρτηση με διαφορετικά σ_x, σ_y και (x_0, y_0)



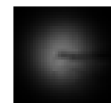
(α) Αρχική εικόνα



(β) Μετά τον πολλαπλασιασμό με τη Γκαουσιανή συνάρτηση



(γ) Γκαουσιανή συνάρτηση κεντραρισμένη στο σημείο ενδιαφέροντος



(δ) Τελική εικόνα

Σχήμα 3.10 – Παράδειγμα με πραγματική εικόνα

4

Πειράματα

Προκειμένου να αξιολογηθεί η απόδοση του προτεινόμενου περιγραφέα, οργανώθηκαν και εκτελέστηκαν πειράματα ταξινόμησης εικόνων. Για την εξαγωγή των τοπικών χαρακτηριστικών έγινε χρήση της προτεινόμενης αναπαράστασης και, προκειμένου να μπορεί να γίνει κάποια σύγκριση, του αλγόριθμου SIFT (Κεφ 1.2.2). Έπειτα υλοποιήθηκε το μοντέλο Bag of Words, προκειμένου να παραχθούν οι περιγραφείς ολόκληρων των εικόνων και να γίνει η ταξινόμηση. Για την αξιολόγηση της ταξινόμησης χρησιμοποιήθηκαν οι μετρικές Average Precision (AP) και Mean Average Precision (MAP).

4.1 ΟΡΓΑΝΩΣΗ ΠΕΙΡΑΜΑΤΩΝ

4.1.1 Δεδομένα

Για τα πειράματα χρησιμοποιήθηκε η **SUN2012** [XHE⁺10], η οποία περιέχει 16.873 έγχρωμες εικόνες, ταξινομημένες σε περισσότερες από χίλιες κλάσεις, ανάλογα με τη σκηνή που περιγράφουν (π.χ. tennis court, cemetery, shore κτλ). Λόγω του πολύ μεγάλου αριθμού εικόνων, καθώς και του ότι πολλές κλάσεις περιέχουν πολύ λίγες εικόνες (υπάρχουν αρκετές κλάσεις με μία μόνο εικόνα), επιλέχθηκαν οι είκοσι πολυπληθέστερες κλάσεις της. Επειδή, όμως και σε αυτές τις κλάσεις υπάρχει μεγάλη ανισορροπία στο πλήθος των εικόνων που περιέχουν (η κατηγορία bedroom περιέχει 998 εικόνες, ενώ η κατηγορία mountain περιέχει 84) από κάθε κλάση επιλέχθηκαν το πολύ εκατό εικόνες. Η τελική λίστα των κλάσεων που χρησιμοποιήθηκαν, καθώς και ο αριθμός των εικόνων που περιέχουν, φαίνονται στον Πίνακα 4.1.

Πίνακας 4.1 – Οι κλάσεις που χρησιμοποιήθηκαν και ο αριθμός των εικόνων τους

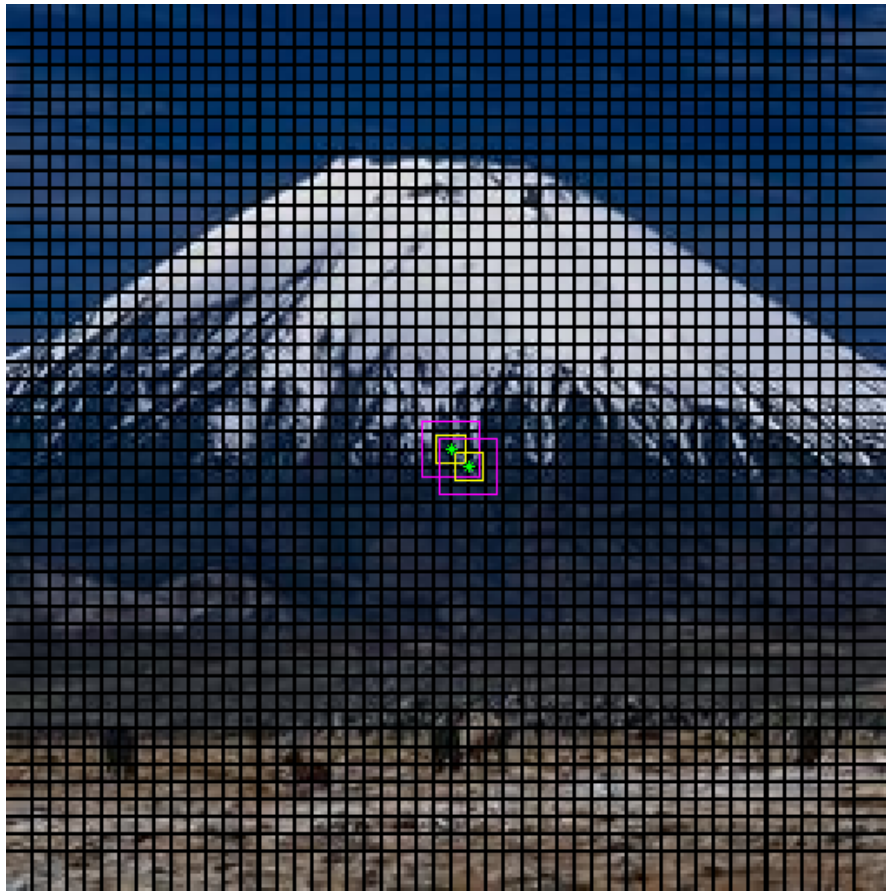
CLASS	No. OF IMAGES
art studio	79
bathroom	100
beach	97
bedroom	100
building facade	100
coast	95
conference room	88
corridor	96
dining room	100
game room	100
highway	100
home office	82
hotel room	100
kitchen	100
living room	100
mountain	84
mountain snowy	100
skyscraper	100
street	100
waiting room	100
Total:	1921

Το επόμενο βήμα είναι η προσαρμογή του μεγέθους των εικόνων. Συγκεκριμένα, οι εικόνες μετασχηματίστηκαν έτσι, ώστε η μικρή τους διάσταση να είναι 254 pixels. Αυτό έγινε προκειμένου να υπάρχει ομοιομορφία στην ανάλυση των εικόνων και να μην υπερτερούν/υστερούν κάποιες κλάσεις.

4.1.2 Υλοποίηση του μοντέλου Bag of Words

Εξαγωγή τοπικών περιγραφών Ακολουθώντας τη διαδικασία του [XHE⁺10], το πλέγμα που εφαρμόζεται στις εικόνες επιλέχθηκε να είναι 5×5 . Επιπλέον για τον τοπικό περιγραφέα που θα παραχθεί σε κάθε κόμβο επιλέγονται δύο γειτονιές διαφορετικού μεγέθους. Με κέντρο τον εκάστοτε κόμβο και ακτίνες $r = 4$ και $r = 8$ παράγονται αναπαραστάσεις για γειτονιές μεγέθους 9×9 και 17×17 pixels αντίστοιχα. Στη συνέχεια αυτές οι αναπαραστάσεις διανυσματοποιούνται και ενώονται και έτσι δημιουργείται ο τοπικός περιγραφέας του κόμβου.

Στο σχήμα 4.1 φαίνεται το πλέγμα μιας εικόνας, στους κόμβους του οποίου υπολογίζονται οι τοπικοί περιγραφείς. Έχουν σημειωθεί δύο διαγωνίως γειτονικοί κόμβοι και οι γειτονιές που αντιστοιχούν στις επιλεγμένες ακτίνες. Στα σχήματα 4.1β' και 4.1δ' φαίνονται οι δύο γειτονιές του πάνω αριστερά κόμβου, όπως έχουν εξαχθεί από την εικόνα και στα σχήματα 4.1γ' και 4.1ε' φαίνονται οι ίδιες γειτονιές μετά



(α') Πλέγμα εικόνας με τις γειτονιές δύο κόμβων



(β')



(γ')



(δ')



(ε')

Σχήμα 4.1 – Πλέγμα σε εικόνα και γειτονιές πριν και μετά την εφαρμογή της Γκαουσιανής συνάρτησης

την εφαρμογή της Γκαουσιανής συνάρτησης. Στην ουσία οι δύο αναπαραστάσεις παράγονται από τις 4.1γ' και 4.1ε' και στη συνέχεια ενώνονται προκειμένου να προκύψει ο τοπικός περιγραφέας του κόμβου.

Για την παραγωγή των τοπικών περιγραφέων του αλγόριθμου SIFT χρησιμοποιήθηκε ο αλγόριθμος ColorDescriptor [vdSGS10] με την επιλογή *opponentsift*, χρησιμοποιώντας βήμα ίσο με 5 pixels για το πλέγμα και δύο διαφορετικά scales (2 και 4).

Το αποτέλεσμα της παραπάνω διαδικασίας είναι η εξαγωγή περίπου 3000 τοπικών περιγραφέων από την κάθε εικόνα.

Δημιουργία των Codebooks Για τη δημιουργία των Codebooks του κάθε περιγραφέα χρησιμοποιήθηκαν 1000 εικόνες διαφορετικές από το σύνολο που αναφέρθηκε στην προηγούμενη ενότητα. Συγκεκριμένα επιλέχθηκε τυχαία μία εικόνα από σχεδόν κάθε κλάση της SUN2012. Αυτό έγινε, προκειμένου οι εικονικές λέξεις που θα προκύψουν να είναι πιο αντικειμενικές και όχι "προσαρμοσμένες" στο σύνολο δεδομένων που θα ταξινομηθεί.

Από τις 1000 εικόνες που επιλέχθηκαν, παράχθηκαν οι τοπικοί περιγραφείς τους. Στη συνέχεια, από αυτούς επιλέχθηκαν τυχαία 250.000, οι οποίοι τροφοδοτήθηκαν στον αλγόριθμο kmeans, με σκοπό την εύρεση ενός αριθμού κέντρων (π.χ. 512 ή 1024). Τα κέντρα αυτά έχουν τόσες διαστάσεις όσες και το μήκος των περιγραφέων και αποτελούν τις εικονικές λέξεις.

Δημιουργία των Feature Vectors Στη συνέχεια, προκειμένου να αναπαρασταθεί η κάθε εικόνα (πλέον από το σύνολο των 1921 εικόνων που επιλέχθηκαν για ταξινόμηση) με βάση τα Codebooks που δημιουργήθηκαν, υπολογίζεται η ευκλείδεια απόσταση του κάθε τοπικού περιγραφέα από τα κέντρα που παράχθηκαν από τον kmeans και ο κάθε περιγραφέας αντιστοιχίζεται στο κοντινότερό του κέντρο. Τελικά τα feature vectors έχουν τόσα στοιχεία όσα και τα κέντρα που παράξε κάθε φορά ο kmeans.

4.1.3 Ταξινόμηση των εικόνων

Τέλος, η ταξινόμηση των εικόνων γίνεται με τον ταξινομητή SVM (Support Vector Machine) με χρήση του λογισμικού LIBSVM [CL11]. Πιο συγκεκριμένα, για κάθε κλάση εκτελείται δυαδική ταξινόμηση (Binary Classification). Αυτό σημαίνει, πως σε κάθε ταξινόμηση, η απόφαση του ταξινομητή έχει να κάνει με το αν μια εικόνα ανήκει ή όχι σε μία συγκεκριμένη κλάση, και δεν έχει να κάνει με την κλάση στην οποία ανήκει η εικόνα. Για την ακρίβεια, το αποτέλεσμα που δίνει ο ταξινομητής για την κάθε εικόνα, είναι η πιθανότητα αυτή να ανήκει στην προς εξέταση κλάση. Με αυτό τον τρόπο εκτελούνται είκοσι ταξινομήσεις, μία για την κάθε κλάση της SUN2012 που έχει επιλεγεί.

Όσον αφορά στον SVM, επιλέγεται να χρησιμοποιεί τον RBF kernel (Radial Basis Function). Αυτό σημαίνει πως πρέπει να επιλεγθούν οι τιμές για δύο παραμέτρους, C και γ . Οι προκαθορισμένες τιμές που χρησιμοποιεί ο LIBSVM είναι $C = 1$ και $\gamma = \frac{1}{\text{NumberOfFeatures}}$. Αυτές είναι και οι τιμές που χρησιμοποιήθηκαν σε πρώτο στάδιο στα πειράματα. Στη συνέχεια έγινε επιλογή των βέλτιστων παραμέτρων C και γ για την κάθε κλάση. Αυτή η διαδικασία γίνεται με 3-fold Cross-Validation για κάθε κλάση. Ουσιαστικά εκτελούνται πολλές ταξινομήσεις για διάφορα C και γ από ένα μεγάλος εύρος τιμών και τελικά επιλέγονται για την κάθε κλάση τα C και γ που επιφέρουν το υψηλότερο Mean Average Precision.

4.1.4 Αξιολόγηση των αποτελεσμάτων

Στον κλάδο της ανάκτησης πληροφορίας (Information Retrieval) χρησιμοποιούνται κάποιες μετρικές αξιολόγησης της απόδοσης ενός ταξινομητή, οι οποίες αναλύονται σύντομα παρακάτω.

Precision (Positive Predictive Value) Πρόκειται για το ποσοστό των δεδομένων που ταξινομήθηκαν σωστά σε σχέση με όλα τα δεδομένα που ταξινομήθηκαν στην εκάστοτε κλάση:

$$Precision = \frac{NumberOfTruePositives}{NumberOfTruePositives + NumberOfFalsePositives} \quad (4.1)$$

Το Precision λαμβάνει υπόψιν όλα τα δεδομένα του dataset. Για την περίπτωση που η ταξινόμηση υλοποιείται πιθανοτικά και τα δεδομένα μπορούν να καταταχθούν σύμφωνα με την πιθανότητά τους να ανήκουν σε κάποια κλάση, το Precision μπορεί να υπολογιστεί για κάποια συγκεκριμένη θέση της κατάταξης (cut-off rank), λαμβάνοντας υπόψιν μόνο τα δεδομένα που είναι ψηλότερα στη λίστα.

Recall (Sensitivity) Είναι το ποσοστό των σωστά ταξινομημένων δεδομένων σε μια κλάση σε σχέση με το σύνολο των δεδομένων που ανήκουν σε αυτή:

$$Recall = \frac{NumberOfTruePositives}{NumberOfTruePositives + NumberOfFalseNegatives} \quad (4.2)$$

Average Precision Οι μετρικές Precision και Recall βασίζονται στο σύνολο των δεδομένων που έχουν ταξινομηθεί. Όταν όμως ένας ταξινομητής κατατάσσει τα δεδομένα σύμφωνα με την πιθανότητα του καθενός να ανήκει σε κάποια κλάση, θα πρέπει να ληφθεί και αυτό το δεδομένο υπόψιν στην αξιολόγηση. Κατατάσσοντας τα δεδομένα σύμφωνα με την πιθανότητα που επιστρέφει ο ταξινομητής και υπολογίζοντας και τις δύο μετρικές σε κάθε θέση της κατάταξης, είναι δυνατό να εκφραστεί το Precision ως συνάρτηση του Recall και να σχεδιαστεί η καμπύλη $P(r)$. Το εμβαδόν κάτω από την καμπύλη είναι η τιμή του Average Precision. Εναλλακτικά μπορεί να υπολογιστεί από τον παρακάτω τύπο:

$$AveragePrecision = \frac{\sum_{k=1}^n (P(k) \times rel(k))}{NumberOfTruePositives + NumberOfFalseNegatives}, \quad (4.3)$$

όπου k είναι η θέση στην κατάταξη των δεδομένων, n ο αριθμός των δεδομένων, $P(k)$ το Precision στη θέση k της κατάταξης και $rel(k)$ μια συνάρτηση, η οποία παίρνει την τιμή 1 αν το δεδομένο της θέσης k ανήκει στην κλάση και την τιμή 0 αν δεν ανήκει.

Mean Average Precision Αν έχουν διεξαχθεί N ταξινομήσεις, τότε το Mean Average Precision είναι ο μέσος όρος των Average Precisions της κάθε ταξινόμησης:

$$MeanAveragePrecision = \frac{\sum_{n=1}^N AveragePrecision(n)}{N} \quad (4.4)$$

4.2 ΑΠΟΤΕΛΕΣΜΑΤΑ

4.2.1 Προτεινόμενος Περιγραφέας

Παρακάτω παρατίθενται τα αποτελέσματα των πειραμάτων που εκτελέστηκαν με τις προκαθορισμένες τιμές C και γ , καθώς και μετά την επιλογή των βέλτιστων. Έγιναν δοκιμές για διαφορετικούς αριθμούς κέντρων για τον kmeans (256, 512 & 1024), δηλαδή για διαφορετικά μήκη των Feature Vectors, για διαφορετικό αριθμό folds για το cross-validation (3, 10 & 50) και για διαφορετικό αριθμό επιπέδων κβάντισης των γωνιών θ_1 και θ_2 (9 & 10).

Στους Πίνακες 4.2 - 4.13 φαίνεται το Mean Average Precision (MAP) για την κάθε κλάση, καθώς και η prior probablity της κάθε κλάσης. Συγκεκριμένα, οι Πίνακες 4.2 - 4.7, έχουν τις τιμές των Mean Average Precisions για τις προκαθορισμένες τιμές $C = 1$ και $\gamma = \frac{1}{1921} = 0.000521$, ενώ οι Πίνακες 4.8 - 4.13 έχουν τις αντίστοιχες τιμές για τα βέλτιστα (C, γ) που επιλέχθηκαν για κάθε κλάση. Στον Πίνακα 4.14 δίνεται το συνολικό Mean Average Precision της κάθε ταξινόμησης (μέσος όρος των Mean Average Precisions της κάθε κλάσης).

Πίνακας 4.2 – MAP κλάσεων με προκαθορισμένα (C, γ) , $Folds = 3$ & $|\theta_1| = |\theta_2| = 9$

Class	$k = 256$	$k = 512$	$k = 1024$	Prior
Art Studio	0.063894	0.081614	0.062726	0.041124
Bathroom	0.127658	0.172253	0.176499	0.052057
Beach	0.171324	0.186015	0.196122	0.050495
Bedroom	0.184648	0.207515	0.220907	0.052056
Building Facade	0.129723	0.092083	0.102008	0.052057
Coast	0.092556	0.123573	0.120666	0.049453
Conference Room	0.058287	0.0752	0.090511	0.04581
Corridor	0.097436	0.160207	0.172728	0.049974
Dining Room	0.136785	0.152057	0.165016	0.052057
Game Room	0.102144	0.104751	0.117927	0.052056
Highway	0.365122	0.436296	0.56936	0.052057
Home Office	0.097321	0.095466	0.085475	0.042686
Hotel Room	0.100107	0.084862	0.113705	0.052056
Kitchen	0.091571	0.09959	0.148255	0.052057
Living Room	0.10587	0.172773	0.118368	0.052057
Mountain	0.122079	0.143805	0.167993	0.043727
Mountain Snowy	0.120966	0.155292	0.188036	0.052056
Skyscraper	0.108786	0.075136	0.074188	0.052057
Street	0.227143	0.229103	0.313941	0.052057
Waiting Room	0.124251	0.167801	0.164805	0.052056

Πίνακας 4.3 – MAP κλάσεων με προκαθορισμένα (C, γ) & $Folds = 10, |\theta_1| = |\theta_2| = 9$

Class	$k = 256$	$k = 512$	$k = 1024$	Prior
Art Studio	0.075893	0.095064	0.082852	0.041124
Bathroom	0.146942	0.205189	0.209388	0.052056
Beach	0.222928	0.203886	0.26532	0.050494
Bedroom	0.205259	0.255747	0.238645	0.052056
Building Facade	0.146719	0.111306	0.134056	0.052056
Coast	0.127882	0.131593	0.154347	0.049452
Conference Room	0.080869	0.088459	0.08354	0.045812
Corridor	0.119455	0.157869	0.24166	0.049973
Dining Room	0.142322	0.181586	0.213156	0.052056
Game Room	0.109764	0.106966	0.140139	0.052056
Highway	0.415427	0.450306	0.628165	0.052056
Home Office	0.10405	0.116504	0.126226	0.042687
Hotel Room	0.098973	0.109223	0.150421	0.052056
Kitchen	0.128514	0.122073	0.17325	0.052056
Living Room	0.164139	0.186187	0.209028	0.052056
Mountain	0.145021	0.156785	0.224321	0.043726
Mountain Snowy	0.168448	0.12648	0.242898	0.052056
Skyscraper	0.135744	0.100235	0.130347	0.052056
Street	0.247276	0.275991	0.364562	0.052056
Waiting Room	0.158255	0.20499	0.193534	0.052056

Πίνακας 4.4 – MAP κλάσεων με προκαθορισμένα (C, γ) & $Folds = 50, |\theta_1| = |\theta_2| = 9$

Class	$k = 256$	$k = 512$	$k = 1024$	Prior
Art Studio	0.130312	0.18446	0.121163	0.041012
Bathroom	0.266897	0.310006	0.286251	0.052065
Beach	0.354134	0.327172	0.376761	0.050486
Bedroom	0.284592	0.403264	0.389215	0.052065
Building Facade	0.203416	0.174679	0.214375	0.052065
Coast	0.234959	0.250111	0.253457	0.049433
Conference Room	0.122138	0.150524	0.168396	0.045911
Corridor	0.20066	0.291181	0.361927	0.050013
Dining Room	0.296275	0.286858	0.333027	0.052065
Game Room	0.20529	0.207475	0.258443	0.052065
Highway	0.569267	0.56918	0.660474	0.052065
Home Office	0.175708	0.244824	0.219133	0.042659
Hotel Room	0.204079	0.211529	0.228224	0.052065
Kitchen	0.202199	0.207997	0.291128	0.052065
Living Room	0.240273	0.307667	0.317422	0.052065
Mountain	0.270273	0.254283	0.303503	0.043644
Mountain Snowy	0.252137	0.265486	0.326909	0.052065
Skyscraper	0.220027	0.156197	0.214107	0.052065
Street	0.408652	0.41957	0.483664	0.052065
Waiting Room	0.274978	0.279101	0.341361	0.052065

ΚΕΦΑΛΑΙΟ 4. ΠΕΙΡΑΜΑΤΑ

Πίνακας 4.5 – MAP κλάσεων με προκαθορισμένα (C, γ) , $Folds = 3$ & $|\theta_1| = |\theta_2| = 10$

Class	$k = 256$	$k = 512$	$k = 1024$	Prior
Art Studio	0.079997	0.080263	0.073313	0.041124
Bathroom	0.137565	0.139438	0.157811	0.052057
Beach	0.248308	0.15636	0.199223	0.050495
Bedroom	0.214903	0.21657	0.261672	0.052056
Building Facade	0.112144	0.116477	0.161065	0.052057
Coast	0.116706	0.12779	0.119791	0.049453
Conference Room	0.078096	0.077753	0.086691	0.04581
Corridor	0.085352	0.119937	0.175487	0.049974
Dining Room	0.161409	0.149748	0.24039	0.052057
Game Room	0.099058	0.125394	0.089614	0.052056
Highway	0.388948	0.376431	0.542788	0.052057
Home Office	0.071432	0.117242	0.093314	0.042686
Hotel Room	0.088796	0.118469	0.10698	0.052056
Kitchen	0.146953	0.161449	0.176027	0.052057
Living Room	0.076906	0.117852	0.153996	0.052057
Mountain	0.070386	0.081114	0.120512	0.043727
Mountain Snowy	0.116827	0.170581	0.152396	0.052056
Skyscraper	0.082739	0.07459	0.105786	0.052057
Street	0.209066	0.220444	0.296925	0.052057
Waiting Room	0.08409	0.079781	0.123116	0.052056

Πίνακας 4.6 – MAP κλάσεων με προκαθορισμένα (C, γ) , $Folds = 10$ & $|\theta_1| = |\theta_2| = 10$

Class	$k = 256$	$k = 512$	$k = 1024$	Prior
Art Studio	0.105937	0.098698	0.100495	0.041124
Bathroom	0.190204	0.177112	0.156943	0.052056
Beach	0.259469	0.158145	0.222484	0.050494
Bedroom	0.231609	0.274107	0.303034	0.052056
Building Facade	0.13816	0.157389	0.183472	0.052056
Coast	0.155485	0.173137	0.129701	0.049452
Conference Room	0.13393	0.103576	0.117282	0.045812
Corridor	0.096897	0.143852	0.203655	0.049973
Dining Room	0.205306	0.175017	0.2979	0.052056
Game Room	0.139024	0.16479	0.134975	0.052056
Highway	0.425514	0.529444	0.530649	0.052056
Home Office	0.090303	0.133913	0.123546	0.042687
Hotel Room	0.120336	0.129741	0.141975	0.052056
Kitchen	0.196764	0.202736	0.200472	0.052056
Living Room	0.116226	0.12239	0.193498	0.052056
Mountain	0.131725	0.154989	0.144307	0.043726
Mountain Snowy	0.183736	0.212613	0.177229	0.052056
Skyscraper	0.104785	0.125049	0.11931	0.052056
Street	0.292208	0.286966	0.327008	0.052056
Waiting Room	0.138026	0.100611	0.163523	0.052056

Πίνακας 4.7 – MAP κλάσεων με προκαθορισμένα (C, γ) , $Folds = 50$ & $|\theta_1| = |\theta_2| = 10$

Class	$k = 256$	$k = 512$	$k = 1024$	Prior
Art Studio	0.205115	0.173588	0.188372	0.041012
Bathroom	0.272554	0.240376	0.279552	0.052065
Beach	0.391903	0.29667	0.334888	0.050486
Bedroom	0.336184	0.404155	0.434131	0.052065
Building Facade	0.233229	0.260713	0.248968	0.052065
Coast	0.180767	0.304528	0.229332	0.049433
Conference Room	0.197017	0.180601	0.203306	0.045911
Corridor	0.185167	0.216838	0.26654	0.050013
Dining Room	0.32361	0.320754	0.37075	0.052065
Game Room	0.22486	0.266379	0.231751	0.052065
Highway	0.534109	0.604833	0.650643	0.052065
Home Office	0.189517	0.239148	0.237753	0.042659
Hotel Room	0.202306	0.19976	0.209218	0.052065
Kitchen	0.300351	0.334994	0.325579	0.052065
Living Room	0.193113	0.196667	0.291978	0.052065
Mountain	0.20063	0.273075	0.277964	0.043644
Mountain Snowy	0.232507	0.29957	0.333334	0.052065
Skyscraper	0.217296	0.222932	0.194985	0.052065
Street	0.428378	0.451346	0.48494	0.052065
Waiting Room	0.226862	0.183279	0.277651	0.052065

Τα παραπάνω αποτελέσματα, δείχνουν πως, ακόμα και σε πρωταρχικό στάδιο, ο προτεινόμενος περιγραφέας έχει καλή απόδοση. Το σημαντικότερο είναι πως, ακόμα και για τα χαμηλότερα MAPs που εμφανίζονται, δεν πέφτουν ποτέ κάτω από την prior probability της κάθε κλάσης, που είναι το ποσοστό των εικόνων της εκάστοτε κλάσης σε σχέση με όλο το σύνολο των εικόνων. Αυτή θα ήταν η απόδοση που θα είχε ένας τυχαίος ταξινομητής. Επιπλέον, είναι εμφανές πως καθώς αυξάνεται ο αριθμός των folds, αυξάνεται και η απόδοση της ταξινόμησης. Επίσης η αύξηση των εικονικών λέξεων, σε γενικές γραμμές προκαλεί αύξηση του MAP της κάθε κλάσης, όμως η αύξηση/μείωση των παραμέτρων θ_1, θ_2 δεν εμφανίζει ξεκάθαρη αλλαγή, αλλά η κάθε κλάση επηρεάζεται διαφορετικά.

Πίνακας 4.8 – MAP κλάσεων με βέλτιστα (C, γ) , $Folds = 3$ & $|\theta_1| = |\theta_2| = 9$

Class	$k = 256$	$k = 512$	$k = 1024$	Prior
Art Studio	1	1	1	0.041124
Bathroom	0.362161	0.362161	0.362161	0.052057
Beach	0.27993	0.230228	0.251213	0.050495
Bedroom	0.150613	0.150177	0.150177	0.052056
Building Facade	0.09927	0.102852	0.090378	0.052057
Coast	0.106875	0.107269	0.116797	0.049453
Conference Room	0.081763	0.087463	0.119224	0.04581
Corridor	0.160346	0.186411	0.256522	0.049974
Dining Room	0.192957	0.245521	0.239706	0.052057
Game Room	0.167939	0.133436	0.142271	0.052056
Highway	0.523908	0.52981	0.60895	0.052057
Home Office	0.162221	0.142854	0.148534	0.042686
Hotel Room	0.134094	0.123842	0.15608	0.052056
Kitchen	0.10016	0.245099	0.114369	0.052057
Living Room	0.18249	0.207072	0.205403	0.052057
Mountain	0.143669	0.154363	0.135092	0.043727
Mountain Snowy	0.169751	0.18275	0.213666	0.052056
Skyscraper	0.131846	0.119361	0.131357	0.052057
Street	0.432261	0.493008	0.482712	0.052057
Waiting Room	0.174738	0.184992	0.181842	0.052056

Πίνακας 4.9 – MAP κλάσεων με βέλτιστα (C, γ) , $Folds = 10$ & $|\theta_1| = |\theta_2| = 9$

Class	$k = 256$	$k = 512$	$k = 1024$	Prior
Art Studio	1	1	1	0.041124
Bathroom	0.381111	0.381111	0.381111	0.052056
Beach	0.300283	0.307213	0.274325	0.050494
Bedroom	0.159832	0.157614	0.157563	0.052056
Building Facade	0.116547	0.131857	0.14746	0.052056
Coast	0.139279	0.12674	0.145841	0.049452
Conference Room	0.104225	0.121341	0.145573	0.045812
Corridor	0.235302	0.245513	0.321208	0.049973
Dining Room	0.274234	0.315362	0.32545	0.052056
Game Room	0.208783	0.171839	0.157295	0.052056
Highway	0.570014	0.553496	0.609362	0.052056
Home Office	0.181871	0.177636	0.177327	0.042687
Hotel Room	0.174673	0.184601	0.179643	0.052056
Kitchen	0.173119	0.162082	0.138157	0.052056
Living Room	0.217057	0.251561	0.279407	0.052056
Mountain	0.183148	0.222426	0.209991	0.043726
Mountain Snowy	0.207899	0.207675	0.258425	0.052056
Skyscraper	0.130135	0.141833	0.119368	0.052056
Street	0.502054	0.490369	0.49149	0.052056
Waiting Room	0.21356	0.22363	0.197682	0.052056

Πίνακας 4.10 – MAP κλάσεων με βέλτιστα (C, γ) , $Folds = 50$ & $|\theta_1| = |\theta_2| = 9$

Class	$k = 256$	$k = 512$	$k = 1024$	Prior
Art Studio	1	1	1	0.041012
Bathroom	0.486667	0.486667	0.486667	0.052065
Beach	0.343087	0.346298	0.333929	0.050486
Bedroom	0.210714	0.206508	0.208294	0.052065
Building Facade	0.193925	0.19873	0.253202	0.052065
Coast	0.23059	0.218194	0.264315	0.049433
Conference Room	0.221471	0.191171	0.291819	0.045911
Corridor	0.349622	0.355057	0.413583	0.050013
Dining Room	0.367021	0.421257	0.393269	0.052065
Game Room	0.314294	0.251372	0.265178	0.052065
Highway	0.628059	0.671587	0.702652	0.052065
Home Office	0.270956	0.315666	0.276838	0.042659
Hotel Room	0.290959	0.28689	0.285223	0.052065
Kitchen	0.27347	0.272288	0.240684	0.052065
Living Room	0.32919	0.364879	0.365865	0.052065
Mountain	0.280085	0.385309	0.295985	0.043644
Mountain Snowy	0.286989	0.342161	0.362631	0.052065
Skyscraper	0.20935	0.230934	0.239688	0.052065
Street	0.574799	0.587352	0.592235	0.052065
Waiting Room	0.294129	0.320549	0.331991	0.052065

Πίνακας 4.11 – MAP κλάσεων με βέλτιστα (C, γ) , $Folds = 3$ & $|\theta_1| = |\theta_2| = 10$

Class	$k = 256$	$k = 512$	$k = 1024$	Prior
Art Studio	1	1	1	0.041124
Bathroom	0.362161	0.362161	0.362161	0.052057
Beach	0.37884	0.24723	0.204693	0.050495
Bedroom	0.280157	0.268146	0.150613	0.052056
Building Facade	0.191962	0.145636	0.181169	0.052057
Coast	0.154578	0.112198	0.167403	0.049453
Conference Room	0.123167	0.134798	0.173129	0.04581
Corridor	0.22359	0.192761	0.204919	0.049974
Dining Room	0.228999	0.21241	0.294331	0.052057
Game Room	0.195765	0.173597	0.143085	0.052056
Highway	0.458761	0.507344	0.534104	0.052057
Home Office	0.15669	0.130915	0.143395	0.042686
Hotel Room	0.148666	0.150314	0.141126	0.052056
Kitchen	0.245099	0.175479	0.167039	0.052057
Living Room	0.163175	0.146913	0.198079	0.052057
Mountain	0.099141	0.172204	0.171781	0.043727
Mountain Snowy	0.194156	0.162975	0.171309	0.052056
Skyscraper	0.110494	0.093246	0.143801	0.052057
Street	0.464944	0.37122	0.482376	0.052057
Waiting Room	0.173	0.140655	0.183369	0.052056

Πίνακας 4.12 – MAP κλάσεων με βέλτιστα (C, γ) , $Folds = 10$ & $|\theta_1| = |\theta_2| = 10$

Class	$k = 256$	$k = 512$	$k = 1024$	Prior
Art Studio	1	1	1	0.041124
Bathroom	0.381111	0.381111	0.381111	0.052056
Beach	0.408942	0.291675	0.223036	0.050494
Bedroom	0.31652	0.323731	0.199734	0.052056
Building Facade	0.232113	0.191482	0.212034	0.052056
Coast	0.179139	0.147699	0.172902	0.049452
Conference Room	0.154772	0.169595	0.228102	0.045812
Corridor	0.258336	0.224534	0.267527	0.049973
Dining Room	0.270572	0.277383	0.315536	0.052056
Game Room	0.210978	0.223081	0.174936	0.052056
Highway	0.519722	0.555625	0.583406	0.052056
Home Office	0.193889	0.231479	0.20664	0.042687
Hotel Room	0.151765	0.190381	0.20488	0.052056
Kitchen	0.278759	0.220858	0.255483	0.052056
Living Room	0.196798	0.211152	0.237256	0.052056
Mountain	0.192967	0.236081	0.254337	0.043726
Mountain Snowy	0.247365	0.247448	0.253723	0.052056
Skyscraper	0.167189	0.125805	0.184506	0.052056
Street	0.534700	0.455385	0.463495	0.052056
Waiting Room	0.197733	0.181173	0.209666	0.052056

Πίνακας 4.13 – MAP κλάσεων με βέλτιστα (C, γ) , $Folds = 50$ & $|\theta_1| = |\theta_2| = 10$

Class	$k = 256$	$k = 512$	$k = 1024$	Prior
Art Studio	1	1	1	0.041012
Bathroom	0.486667	0.486667	0.486667	0.052065
Beach	0.476747	0.37869	0.295119	0.050486
Bedroom	0.42885	0.445628	0.34497	0.052065
Building Facade	0.31614	0.319771	0.29346	0.052065
Coast	0.26742	0.236684	0.27615	0.049433
Conference Room	0.270082	0.256155	0.354509	0.045911
Corridor	0.362493	0.323694	0.426763	0.050013
Dining Room	0.359935	0.383781	0.41658	0.052065
Game Room	0.343965	0.328865	0.276656	0.052065
Highway	0.620645	0.604079	0.652215	0.052065
Home Office	0.360796	0.325602	0.353525	0.042659
Hotel Room	0.266287	0.267304	0.301485	0.052065
Kitchen	0.406613	0.342176	0.35959	0.052065
Living Room	0.318234	0.32887	0.342775	0.052065
Mountain	0.350542	0.324016	0.327796	0.043644
Mountain Snowy	0.350402	0.354578	0.360849	0.052065
Skyscraper	0.257661	0.226321	0.281355	0.052065
Street	0.609053	0.57364	0.590187	0.052065
Waiting Room	0.325521	0.311873	0.307022	0.052065

Πίνακας 4.14 – Συνολικό Mean Average Precision όλων των κλάσεων πριν και μετά την επιλογή των βέλτιστων (C, γ)

<i>Folds</i>	<i>k</i>	$ \theta_1 , \theta_2 $	<i>MAP</i>	<i>MAP</i>
			Default (C, γ)	Best (C, γ)
3	256	9	0.13138	0.2378
3	512	9	0.15077	0.2451
3	1024	9	0.16846	0.2553
10	256	9	0.15719	0.2737
10	512	9	0.16932	0.2787
10	1024	9	0.21029	0.2858
50	256	9	0.25581	0.3578
50	512	9	0.27508	0.3726
50	1024	9	0.30745	0.3802
3	256	10	0.13348	0.2677
3	512	10	0.14138	0.2450
3	1024	10	0.17184	0.2609
10	256	10	0.17278	0.30467
10	512	10	0.18121	0.29428
10	1024	10	0.19857	0.30142
50	256	10	0.26377	0.4089
50	512	10	0.28351	0.39092
50	1024	10	0.30358	0.40238

Μετά τη σωστή παραμετροποίηση του SVM για την κάθε κλάση, είναι έντονη η αύξηση της απόδοσης, τόσο στα MAPs των συνολικών πειραμάτων, όσο και στα MAPs της κάθε κλάσης χωριστά. Ιδιαίτερο ενδιαφέρον παρουσιάζει η κλάση Art Studio, η οποία, ενώ με τις προκαθορισμένες παραμέτρους του SVM παρουσίαζε ένα από τα χαμηλότερα ποσοστά, με σωστή παραμετροποίηση φθάνει $MAP = 1$. Επίσης ενδιαφέρον είναι πως οι κλάσεις που είχαν τα υψηλότερα ποσοστά με τη χρήση των προκαθορισμένων παραμέτρων (π.χ. η κλάση Highway με $MAP = 0.65$ και η κλάση Street με $MAP = 0.48$), μετά τη σωστή παραμετροποίηση παρουσιάζουν μικρή ή και καθόλου αύξηση. Το μέγιστο MAP των πειραμάτων είναι 0.41.

4.2.2 SIFT

Στη συνέχεια παρατίθενται τα αποτελέσματα των πειραμάτων που εκτελέστηκαν με τον αλγόριθμο SIFT. Κατ' αντιστοιχία με την προηγούμενη ενότητα, στους Πίνακες 4.15 - 4.17 δίνονται τα Mean Average Precisions της κάθε κλάσης (οι prior probabilities των κλάσεων, ταυτίζονται με αυτές των Πινάκων 4.8 - 4.13), χρησιμοποιώντας τα βέλτιστα (C, γ) για την κάθε κλάση (ακολουθώντας την ίδια διαδικασία με την προηγούμενη ενότητα). Τέλος, στον Πίνακα 4.18 δίνεται το συνολικό Mean Average Precision της κάθε ταξινόμησης.

Πίνακας 4.15 – SIFT: MAP κλάσεων με βέλτιστα (C, γ) & $Folds = 3$

Class	$k = 256$	$k = 512$	$k = 1024$
hline Art Studio	0.426123	0.456716	0.52434
Bathroom	0.263388	0.301686	0.369081
Beach	0.644619	0.647937	0.687587
Bedroom	0.204364	0.207081	0.203401
Building Facade	0.714735	0.782336	0.768889
Coast	0.643764	0.5564	0.523162
Conference Room	0.419644	0.430917	0.384667
Corridor	0.781111	0.81173	0.835663
Dining Room	0.571273	0.611594	0.561492
Game Room	0.249506	0.230939	0.20605
Highway	0.74971	0.803464	0.855276
Home Office	0.214506	0.213001	0.20864
Hotel Room	0.33015	0.341192	0.404058
Kitchen	0.278814	0.284851	0.269503
Living Room	0.213476	0.254344	0.250156
Mountain	0.558141	0.65293	0.569195
Mountain Snowy	0.504263	0.590986	0.59294
Skyscraper	0.646912	0.696458	0.714188
Street	0.818266	0.786876	0.782615
Waiting Room	0.354517	0.365792	0.410784

Πίνακας 4.16 – SIFT: MAP κλάσεων με βέλτιστα (C, γ) & $Folds = 10$

Class	$k = 256$	$k = 512$	$k = 1024$
hline Art Studio	0.50234	0.55195	0.538948
Bathroom	0.357008	0.416617	0.409677
Beach	0.673515	0.668013	0.72832
Bedroom	0.235785	0.270086	0.245082
Building Facade	0.763746	0.785514	0.801628
Coast	0.626774	0.634321	0.600144
Conference Room	0.501077	0.445713	0.420554
Corridor	0.825307	0.835938	0.854668
Dining Room	0.616943	0.617722	0.633741
Game Room	0.230346	0.254915	0.279225
Highway	0.791088	0.821244	0.892586
Home Office	0.252781	0.282448	0.249099
Hotel Room	0.467479	0.411481	0.393872
Kitchen	0.372117	0.366793	0.354158
Living Room	0.256042	0.310206	0.307545
Mountain	0.674763	0.645572	0.650232
Mountain Snowy	0.664383	0.673839	0.632967
Skyscraper	0.722957	0.698035	0.751292
Street	0.856614	0.819843	0.844512
Waiting Room	0.414722	0.405812	0.471004

Πίνακας 4.17 – SIFT: MAP κλάσεων με βέλτιστα (C, γ) & $Folds = 50$

Class	$k = 256$	$k = 512$	$k = 1024$
hline Art Studio	0.599898	0.589308	0.643892
Bathroom	0.475573	0.514365	0.512358
Beach	0.765803	0.777188	0.785925
Bedroom	0.36769	0.369514	0.352263
Building Facade	0.817592	0.834062	0.872762
Coast	0.713922	0.721346	0.708312
Conference Room	0.534279	0.505616	0.536288
Corridor	0.849223	0.874072	0.871616
Dining Room	0.705192	0.690891	0.73967
Game Room	0.357473	0.349606	0.415885
Highway	0.832915	0.84288	0.8863
Home Office	0.42415	0.422096	0.378307
Hotel Room	0.547673	0.519059	0.515062
Kitchen	0.47314	0.46999	0.48034
Living Room	0.393034	0.419152	0.40329
Mountain	0.816537	0.762426	0.778772
Mountain Snowy	0.742652	0.744524	0.769682
Skyscraper	0.746328	0.79423	0.779273
Street	0.884889	0.870881	0.854873
Waiting Room	0.526949	0.53433	0.554043

Πίνακας 4.18 – Συνολικό Mean Average Precision όλων των κλάσεων με τα βέλτιστα (C, γ)

$Folds$	k	MAP
3	256	0.47936
3	512	0.5014
3	1024	0.50608
10	256	0.54029
10	512	0.5458
10	1024	0.55296
50	256	0.62875
50	512	0.63028
50	1024	0.6419

Τα αποτελέσματα των πειραμάτων με χρήση του αλγόριθμου SIFT παρουσιάζουν παρόμοια συμπεριφορά με αυτή των αποτελεσμάτων της προηγούμενης ενότητας. Και εδώ η αύξηση των folds προκαλεί αύξηση των MAPs των πειραμάτων, καθώς επίσης και η αύξηση του αριθμού των εικονικών λέξεων. Επίσης φαίνεται πως ορισμένες κλάσεις παρουσιάζουν αυξομειώσεις στις αποδόσεις τους, κάτι που συμβαίνει και στα αποτελέσματα του προτεινόμενου περιγραφέα.

Επιπλέον, τα αποτελέσματα που παρατίθενται σε αυτή την ενότητα καθιστούν ξεκάθαρο πως ο SIFT αποδίδει καλύτερα στο σύνολο του κάθε πειράματος σε σχέση με την προτεινόμενη αναπαράσταση, παρόλο που κάτι τέτοιο δεν ισχύει απόλυτα για την απόδοση της κάθε κλάσης ξεχωριστά.

5

Συμπεράσματα

Συνοψίζοντας, στα πλαίσια της παρούσας εργασίας επιχειρήθηκε η υλοποίηση αναπαράστασης εικόνας, η οποία να παραμένει αναλλοίωτη στις διάφορες μεταβολές που μπορεί να υποστεί η εικόνα, όπως περιστροφή, κλιμάκωση ή μετατόπιση με σκοπό να χρησιμοποιηθεί ως τοπικός περιγραφέας, έτσι ώστε με τη χρήση του μοντέλου Bag of Words, να είναι δυνατή η δημιουργία περιγραφέων ολόκληρης εικόνας. Από δοκιμές που έγιναν σε τεχνητές εικόνες προκύπτει, πως η αναπαράσταση παραμένει σταθερή, εμφανίζοντας μικρές διαφοροποιήσεις, οι οποίες αποδίδονται στην αλλοίωση που υφίσταται η δομή της εικόνας, κατά την εφαρμογή των μεταβολών. Συγκεκριμένα, αν η μεταβολή δεν επιφέρει καμία αλλοίωση στην εικόνα (π.χ. περιστροφή κατά 90°), η αναπαράσταση παραμένει ακριβώς ίδια, ενώ στην περίπτωση που η εικόνα αλλοιώνεται, οι διαφοροποιήσεις που προκύπτουν στις αναπαραστάσεις μεταβάλλονται, ανάλογα με το μέγεθος της εικόνας και τις τιμές που έχουν επιλεχθεί για τις παραμέτρους (θ_1, θ_2).

Παρόλο που τόσο στα παραδείγματα με τις τεχνητές εικόνες, όσο και στα πειράματα που εκτελέστηκαν, ο προτεινόμενος περιγραφέας σε συνδυασμό με τη χρήση του μοντέλου Bag of Words απέδωσε πολύ καλά, είναι εμφανές από την εκτέλεση και απόδοση των πειραμάτων με χρήση του αλγόριθμου SIFT πως υπάρχουν περιθώρια σημαντικής βελτίωσης. Αυτό αφορά τόσο στην προγραμματιστική υλοποίηση των τοπικών περιγραφέων, όσο και στην απόδοση της ταξινόμησης. Συγκεκριμένα, οι χρόνοι που απαιτούνται για τη δημιουργία των τοπικών περιγραφέων με την προτεινόμενη υλοποίηση είναι τάξεις μεγέθους μεγαλύτεροι από τους χρόνους που απαιτούνται για τη δημιουργία των τοπικών περιγραφέων με τον αλγόριθμο SIFT. Οι χρόνοι αυτοί θα μπορούσαν σίγουρα να μειωθούν αισθητά, αν παραλληλοποιηθεί η διαδικασία παραγωγής τοπικών περιγραφέων όλης της εικόνας, αφού σε κάθε κόμβο εκτελούνται ακριβώς τα ίδια βήματα. Επιπλέον μείωση μπορεί να προκύψει και με βελτιστοποίηση των αλγορίθμων υπολογισμού της συσχέτισης και δημιουργίας της αναπαράστασης.

Προκειμένου να βελτιωθεί η απόδοση στις ταξινομήσεις, υπάρχουν πολλές δοκιμές που μπορούν να γίνουν προκειμένου να βρεθεί η βέλτιστη αναπαράσταση. Αυτό έχει να κάνει κατ' αρχάς με την πληροφορία που περιέχεται στην αναπαράσταση. Όπως αναφέρθηκε στο Κεφ. 3.1 στα πλαίσια της παρούσας εργασίας χρησιμοποιείται για την αναπαράσταση μόνο η DC συνιστώσα του μετασχηματισμού Fourier του \tilde{x}_3 . Θα μπορούσαν να γίνουν δοκιμές κρατώντας περισσότερη πληροφορία του αρχικού σήματος για την τελική αναπαράσταση, ή ακόμα ρυθμίζοντας τα επίπεδα κβάντισης των γωνιών θ_1, θ_2 . Επιπλέον χρειάζεται να εξακριβωθούν οι συνθήκες στις οποίες λειτουργεί βέλτιστα η αναπαράσταση αναφορικά με το μοντέλο Bag of Words. Αυτό σημαίνει πως μπορούν να γίνουν δοκιμές με το μέγεθος της γειτονιάς για την οποία παράγεται ο τοπικός περιγραφέας, ή ακόμα και το βήμα του πλέγματος που εφαρμόζεται στην εικόνα (π.χ. ο αλγόριθμος ColorDescriptor χρησιμοποιεί πλέγμα, στο οποίο οι κόμβοι σε κάθε δεύτερη ξεκινούν κατά $step/2$ πιο μέσα, προκειμένου να ελαχιστοποιηθεί η επικάλυψη των γειτονιών που χρησιμοποιούνται για την παραγωγή του local feature).

Βιβλιογραφία

- [ADK93] Yannis Avrithis, Anastasios Delopoulos, and Stefanos Kollias. An efficient scheme for invariant optical character recognition using triple correlations. Technical report, National Technical University of Athens, 1993.
- [BETVG08] Herbert Bay, Andreas Ess, Tinne Tuytelaars, and Luc Van Gool. Surf: Speeded up robust features. *Computer Vision and Image Understanding (CVIU)*, 110(3):346–359, 2008.
- [Bri01] David R. Brillinger. *Time Series: Data Analysis and Theory*. SIAM, 2001.
- [CL11] Chih-Chung Chang and Chih-Jen Lin. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2:27:1–27:27, 2011. Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.
- [DB91] A. V. Dandawate and Giannakis G. B. Ergodicity and asymptotic normality for cumulant estimators of nonstationary signal. pages 976–983. 25th Conference on Info Sciences and Systems, March 1991.
- [DTK94] Anastasios Delopoulos, Andreas Tirakis, and Stefanos Kollias. Invariant image classification using triple-correlation-based neural networks. *IEEE Transactions on Neural Networks*, 5(3):392–408, May 1994.
- [Hay99] Monson H. Hayes. *Digital Signal Processing*. The McGraw-Hill Companies, Inc, 1999.
- [HCL10] Chih-Wei Hsu, Chih-Chung Chang, and Chih-Jen Lin. A practical guide to support vector classification, April 2010.
- [Low04] David G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004.
- [Men91] Jerry M. Mendel. Tutorial on higher-order statistics (spectra) in signal processing and system theory: theoretical results and some applications. 79(3):278–305, March 1991.
- [ROD12] Ives Rey Otero and Mauricio Delbracio. The anatomy of the sift method. *Image Processing On Line*, 2012.

BIBΛΙΟΓΡΑΦΙΑ

- [SKFP] Jan Svab, Tomas Krajník, Jan Faigl, and Libor Preucil. Fpga base speeded up robust features.
- [SM86] Gerard Salton and Michael J. McGill. *Introduction to Modern Information Retrieval*. McGraw-Hill, Inc., New York, NY, USA, 1986.
- [Sze10] Richard Szeliski. *Computer Vision: Algorithms and Applications*. Springer, 2010. <http://szeliski.org/Book/>.
- [vdSGS10] K. E. A. van de Sande, T. Gevers, and C. G. M. Snoek. Evaluating color descriptors for object and scene recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(9):1582–1596, 2010. <http://www.science.uva.nl/research/publications/2010/vandeSandeTPAMI2010>.
- [XHE⁺10] Jianxiong Xiao, James Hays, Krista A. Ehinger, Aude Oliva, and Antonio Torralba. Sun database: Large-scale scene recognition from abbey to zoo. 2010.