



Τεχνολογικό Εκπαιδευτικό Ίδρυμα Κρήτης

**Σχολή Τεχνολογικών Εφαρμογών
Τμήμα Εφαρμοσμένης Πληροφορικής & Πολυμέσων**



Πτυχιακή Εργασία

Αυτόματη Αναγνώριση Αντικειμένων Εικόνας

Καλλιατάκης Γρηγόριος (AM 2522)

Επιβλέπων καθηγητής : Τριανταφυλλίδης Γεώργιος

**Επιτροπή αξιολόγησης : Τριανταφυλλίδης Γεώργιος
Βιδάκης Νικόλαος
Κουναλάκης Τσαμπίκος**

Ηράκλειο - Ιούνιος 2012

Abstract

This paper presents an automated object recognition application in images, using Graph Based Visual Saliency (GBVS) and Scale-Invariant Feature Transform (SIFT), aiming at simple image classification of well-known touristic monuments in the geographic area of Heraklion, Crete. The basic idea of the application is the preprocessing of the originally taken images using Saliency in order to minimize the time needed for the classification method, which is proposed in this paper.

To this goal, the system initially uses Graph Based Visual Saliency algorithm on original images taken from various points of interest (monuments). Then, SIFT algorithm is used in two stages of the process. Initially for extracting descriptors and then for calculating matches between two images which are represented by these descriptors. The application is then able to classify these images, saving enough time to be referred as real-time application.

Σύνοψη

Στην παρούσα εργασία παρουσιάζεται μια εφαρμογή αναγνώρισης αντικειμένων σε εικόνες, χρησιμοποιώντας την τεχνική των προεξεχόντων οπτικών σημείων βασισμένα σε γράφους (Graph Based Visual Saliency, GBVS) και τον μετασχηματισμό κλιμακωτά – αναλλοίωτων χαρακτηριστικών (Scale-Invariant Feature Transform , SIFT), στοχεύει στην απλή κατηγοριοποίηση εικόνων γνωστών τουριστικών μνημείων στη γεωγραφική περιοχή του Ηρακλείου Κρήτης. Βασική ιδέα της εφαρμογής θεωρείται η προεπεξεργασία των αρχικών εικόνων, χρησιμοποιώντας τις Προεξοχές προκειμένου να ελαχιστοποιηθεί ο χρόνος που απαιτείται για την μέθοδο κατηγοριοποίησης των εικόνων που θα παρουσιαστεί σε αυτή την εργασία.

Για την επίτευξη αυτού του στόχου, το σύστημα αρχικά χρησιμοποιεί τον αλγόριθμο ανίχνευσης προεξοχών βασισμένο σε γράφους, στις αρχικές εικόνων των περιοχών που ερευνούμε (μνημεία). Στη συνέχεια ο αλγόριθμος του μετασχηματισμού κλιμακωτά – αναλλοίωτων χαρακτηριστικών χρησιμοποιείται σε δυο στάδια της εφαρμογής, Αρχικά για την εξαγωγή περιγραφών της εικόνας, και στη συνέχεια για τον υπολογισμό των αντιστοιχιών μεταξύ δυο εικόνων που αντιπροσωπεύονται από τους εξαγόμενους περιγραφείς τους. Η εφαρμογή στη συνέχεια είναι σε θέση να κατηγοριοποιήσει αυτές τις εικόνες, εξοικονομώντας αρκετό χρόνο ώστε η εφαρμογή να θεωρείται πραγματικού χρόνου.

ΚΑΤΑΛΟΓΟΣ ΠΕΡΙΕΧΟΜΕΝΩΝ

1. ΕΙΣΑΓΩΓΗ	1
1.1 ΠΕΡΙΛΗΨΗ	1
1.2 ΚΙΝΗΤΡΟ ΓΙΑ ΤΗΝ ΔΙΕΞΑΓΩΓΗ ΤΗΣ ΕΡΓΑΣΙΑΣ	1
1.3 ΣΚΟΠΟΣ ΚΑΙ ΣΤΟΧΟΙ ΤΗΣ ΕΡΓΑΣΙΑΣ.....	2
1.4 ΔΟΜΗ ΤΗΣ ΕΡΓΑΣΙΑΣ	2
1.5 ΘΕΩΡΗΤΙΚΟ ΥΠΟΒΑΘΡΟ ΚΑΙ ΕΠΙΣΚΟΠΗΣΗ ΕΡΕΥΝΗΤΙΚΟΥ ΠΕΔΙΟΥ	2
1.5.1 <i>Τεχνητή Όραση.....</i>	3
1.5.2 <i>Ψηφιακή Επεξεργασία Εικόνας</i>	5
1.5.3 <i>Τεχνητή όραση εναντίον Ψηφιακής επεξεργασίας εικόνας.....</i>	7
1.5.4 <i>Προεξέχοντα οπτικά σημεία εικόνας (Visual Saliency).....</i>	8
2. ΜΕΘΟΔΟΙ ΕΞΑΓΩΓΗΣ ΧΑΡΑΚΤΗΡΙΣΤΙΚΩΝ ΕΙΚΟΝΑΣ.....	9
2.1 ΜΕΤΑΣΧΗΜΑΤΙΣΜΟΣ ΚΛΙΜΑΚΩΤΑ – ΑΝΑΛΛΟΙΩΤΩΝ ΧΑΡΑΚΤΗΡΙΣΤΙΚΩΝ (SIFT).....	10
2.1.1 <i>Επισκόπηση της μεθόδου</i>	10
2.1.2 <i>Η Μέθοδος του David Lowe</i>	11
2.1.3 <i>Βασικά στάδια.....</i>	11
2.2 ΕΠΙΤΑΧΥΝΘΕΝΤΑ ΙΣΧΥΡΑ ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ (SPEEDED UP ROBUST FEATURE)	15
2.2.1 <i>Επισκόπηση της μεθόδου</i>	16
2.2.2 <i>Βασικά στάδια.....</i>	16
2.2.3 <i>SIFT εναντίον SURF</i>	21
2.3 ΣΥΣΧΕΤΙΣΜΕΝΟΣ ΜΕΤΑΣΧΗΜΑΤΙΣΜΟΣ ΚΛΙΜΑΚΩΤΑ – ΑΝΑΛΛΟΙΩΤΩΝ ΧΑΡΑΚΤΗΡΙΣΤΙΚΩΝ (ASIFT)....	25
2.3.1 <i>Επισκόπηση της μεθόδου</i>	25
2.3.2 <i>Βασικά στάδια.....</i>	26
2.3.3 <i>SIFT εναντίον ASIFT (Affine SIFT).....</i>	27
2.4 ΙΣΤΟΓΡΑΜΜΑ ΠΡΟΣΑΝΑΤΟΛΙΣΜΕΝΩΝ ΚΛΙΣΕΩΝ (HOG)	29
2.4.1 <i>Θεωρία της μεθόδου.....</i>	29
2.4.2 <i>Βασικά στάδια.....</i>	30
2.5 ΣΥΝΟΠΤΙΚΟΣ ΠΙΝΑΚΑΣ ΜΕΘΟΔΩΝ ΕΞΑΓΩΓΗΣ ΧΑΡΑΚΤΗΡΙΣΤΙΚΩΝ	32
3. ΑΛΓΟΡΙΘΜΟΙ ΚΑΤΗΓΟΡΙΟΠΟΙΗΣΗΣ ΕΙΚΟΝΩΝ	33
3.1 ΤΕΧΝΗΤΑ ΝΕΥΡΩΝΙΚΑ ΔΙΚΤΥΑ.....	34
3.1.1 <i>Γενικά για τα Νευρωνικά Δίκτυα.....</i>	34
3.1.2 <i>Ορισμός Τεχνητών Νευρωνικών Δικτύων.....</i>	34
3.1.3 <i>Ταξινόμηση εικόνων με χρήση Τεχνητών Νευρωνικών Δικτύων</i>	35
3.2 ΜΗΧΑΝΕΣ ΥΠΟΣΤΗΡΙΞΗΣ ΔΙΑΝΥΣΜΑΤΩΝ (SVM).....	36
3.2.1 <i>Ταξινόμηση εικόνων με χρήση Μηχανών Υποστήριξης Διανυσμάτων.....</i>	37
3.3 ΑΣΑΦΗ ΜΕΤΡΑ	38
3.3.1 <i>Ταξινόμηση εικόνων με χρήση Ασαφών Μέτρων.....</i>	38
3.4 ΑΛΓΟΡΙΘΜΟΣ K-ΚΟΝΤΙΝΟΤΕΡΟΥ ΓΕΙΤΟΝΑ	40
3.4.1 <i>Υποθέσεις του KNN αλγορίθμου.....</i>	40
3.4.2 <i>Ταξινόμηση εικόνων με χρήση του αλγορίθμου k-κοντινότερου γείτονα.....</i>	41
3.4.2.1 <i>Πρώτη Περίπτωση για k=1.....</i>	41
3.4.2.2 <i>Δεύτερη Περίπτωση για k=K</i>	41
3.4.3 <i>Βασικές Παρατηρήσεις πάνω στον Αλγόριθμο KNN</i>	42
3.4.4 <i>Πλεονεκτήματα & Μειονεκτήματα του Αλγορίθμου KNN.....</i>	42

4.	ΑΝΑΓΝΩΡΙΣΗ ΑΝΤΙΚΕΙΜΕΝΩΝ.....	43
4.1	ΕΙΣΑΓΩΓΗ ΣΤΗΝ ΕΝΝΟΙΑ ΤΗΣ ΑΝΑΓΝΩΡΙΣΗΣ ΑΝΤΙΚΕΙΜΕΝΩΝ	43
4.2	ΕΠΙΣΚΟΠΗΣΗ ΤΕΧΝΙΚΩΝ ΑΝΑΓΝΩΡΙΣΗΣ ΑΝΤΙΚΕΙΜΕΝΩΝ	44
4.2.1	<i>Αναγνώριση μεμονωμένων αντικειμένων μέσω μοντέλων και αντιστοίχισης.....</i>	<i>45</i>
4.2.2	<i>Αναγνώριση αντικειμένων με τεχνικές αναγνώρισης προτύπων & νευρωνικών δικτύων.....</i>	<i>45</i>
4.2.3	<i>Αναγνώριση αντικειμένων με τμηματικές αναπαραστάσεις.....</i>	<i>46</i>
4.2.4	<i>Αναγνώριση αντικειμένων με παραμορφώσιμα μοντέλα.....</i>	<i>47</i>
5.	ΚΥΡΙΟ ΜΕΡΟΣ ΠΤΥΧΙΑΚΗΣ ΕΡΓΑΣΙΑΣ.....	48
5.1	ΠΡΟΕΞΕΧΟΝΤΑ ΟΠΤΙΚΑ ΣΗΜΕΙΑ ΒΑΣΙΣΜΕΝΑ ΣΕ ΓΡΑΦΟΥΣ (GRAPH BASED VISUAL SALIENCY)	48
5.2	ΥΛΟΠΟΙΗΣΗ ΤΗΣ ΕΦΑΡΜΟΓΗΣ ΓΙΑ ΑΥΤΟΜΑΤΗ ΑΝΑΓΝΩΡΙΣΗ ΑΝΤΙΚΕΙΜΕΝΩΝ ΕΙΚΟΝΑΣ.....	52
5.2.1	<i>Εισαγωγή</i>	<i>52</i>
5.2.2	<i>Αναλυτική Παρουσίαση Διαδικασίας που ακολουθήθηκε</i>	<i>53</i>
5.2.3	<i>Προσδιορισμός Απαιτήσεων.....</i>	<i>56</i>
5.2.4	<i>Πειραματικά αποτελέσματα.....</i>	<i>57</i>
5.2.5	<i>Συμπεράσματα και Μελλοντική Εργασία.....</i>	<i>60</i>
	ΒΙΒΛΙΟΓΡΑΦΙΑ	61

ΚΑΤΑΛΟΓΟΣ ΕΙΚΟΝΩΝ

ΕΙΚΟΝΑ 1: ΙΕΡΑΡΧΙΚΗ ΔΙΑΚΡΙΣΗ ΛΕΙΤΟΥΡΓΙΩΝ ΤΕΧΝΗΤΗΣ ΟΡΑΣΗΣ	3
ΕΙΚΟΝΑ 2 : ΣΧΕΣΗ ΤΕΧΝΗΤΗΣ ΟΡΑΣΗΣ ΜΕ ΆΛΛΕΣ ΕΠΙΣΤΗΜΕΣ	4
ΕΙΚΟΝΑ 3: ΤΑ ΒΑΣΙΚΑ ΣΤΑΔΙΑ ΤΗΣ ΨΗΦΙΑΚΗΣ ΕΠΕΞΕΡΓΑΣΙΑΣ ΕΙΚΟΝΑΣ	6
ΕΙΚΟΝΑ 4: ΕΦΑΡΜΟΓΕΣ ΤΗΣ ΨΗΦΙΑΚΗΣ ΕΠΕΞΕΡΓΑΣΙΑΣ ΕΙΚΟΝΑΣ	7
ΕΙΚΟΝΑ 5: ΤΕΧΝΗΤΗ ΟΡΑΣΗ ΕΝΑΝΤΙΟΝ ΨΗΦ.ΕΠΕΞΕΡΓΑΣΙΑΣ ΕΙΚΟΝΑΣ	7
ΕΙΚΟΝΑ 6: ΠΑΡΑΔΕΙΓΜΑΤΑ ΑΝΤΙΚΕΙΜΕΝΩΝ ΠΟΥ ΠΡΟΕΞΕΧΟΥΝ ΑΠΟ ΤΑ ΥΠΟΛΟΙΠΑ	8
ΕΙΚΟΝΑ 7: ΑΝΙΧΝΕΥΣΗ ΧΑΡΑΚΤΗΡΙΣΤΙΚΩΝ ΧΑΜΗΛΟΥ ΕΠΙΠΕΔΟΥ	10
ΕΙΚΟΝΑ 8: ΑΝΑΠΑΡΑΣΤΑΣΗ ΣΥΝΕΛΙΞΗΣ & ΑΦΑΙΡΕΣΗΣ ΚΑΘΕ ΟΚΤΑΒΑΣ ΓΙΑ ΤΗ ΔΗΜΙΟΥΡΓΙΑ ΤΗΣ GAUSSIAN ΔΙΑΦΟΡΑΣ	12
ΕΙΚΟΝΑ 9: Ο ΥΠΟΛΟΓΙΣΜΟΣ ΤΟΥ ΜΕΓΙΣΤΟΥ ΚΑΙ ΕΛΑΧΙΣΤΟΥ ΤΗΣ GAUSSIAN ΔΙΑΦΟΡΑΣ ΣΕ ΕΝΑ PIXEL	12
ΕΙΚΟΝΑ 10: ΤΑ ΑΠΟΤΕΛΕΣΜΑΤΑ ΤΗΣ ΕΠΙΛΟΓΗΣ ΣΗΜΕΙΩΝ ΚΛΕΙΔΙΩΝ ΣΕ ΜΙΑ ΕΙΚΟΝΑ	13
ΕΙΚΟΝΑ 11: ΠΑΡΑΔΕΙΓΜΑ ΑΝΤΙΣΤΟΙΧΙΣΗΣ SIFT ΧΑΡΑΚΤΗΡΙΣΤΙΚΩΝ	14
ΕΙΚΟΝΑ 12: ΠΑΡΑΔΕΙΓΜΑ ΥΠΟΛΟΓΙΣΜΟΥ SIFT ΠΕΡΙΓΡΑΦΕΩΝ (DESCRIPTORS) ΣΕ ΠΕΡΙΟΧΗ 2x2	15
ΕΙΚΟΝΑ 13: ΤΥΠΟΣ ΤΟΥ ΠΑΡΑΓΟΝΤΑ DoH	17
ΕΙΚΟΝΑ 14: Η ΛΕΝΑ ΚΑΤΑ ΤΗΝ ΕΦΑΡΜΟΓΗ ΤΗΣ ΑΚΡΙΒΗΣ ΠΡΟΣΕΓΓΙΣΗΣ ΤΟΥ DoH ΠΑΡΑΓΟΝΤΑ	17
ΕΙΚΟΝΑ 15: ΤΟ ΟΡΙΟ ΓΙΑ ΤΗΝ ΕΠΙΛΟΓΗ ΧΑΡΑΚΤΗΡΙΣΤΙΚΩΝ ΣΤΟΝ SURF	17
ΕΙΚΟΝΑ 16: ΕΠΙΛΟΓΗ ΧΑΡΑΚΤΗΡΙΣΤΙΚΩΝ SURF ΣΥΝΟΛΙΚΑ 757 ΣΗΜΕΙΑ ΕΝΔΙΑΦΕΡΟΝΤΟΣ	18
ΕΙΚΟΝΑ 17: ΕΠΙΛΟΓΗ ΧΑΡΑΚΤΗΡΙΣΤΙΚΩΝ SURF, ΜΕ ΒΕΛΤΙΩΣΗ, ΣΥΝΟΛΙΚΑ 656 ΣΗΜΕΙΑ ΕΝΔΙΑΦΕΡΟΝΤΟΣ	19
ΕΙΚΟΝΑ 18: ΜΕΡΙΚΑ ΕΝΤΟΠΙΣΜΕΝΑ SURF ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ ΜΑΖΙ ΜΕ ΤΙΣ ΚΥΡΙΑΡΧΕΣ ΚΑΤΕΥΘΥΝΣΕΙΣ ΤΟΥΣ	20
ΕΙΚΟΝΑ 19: ΠΡΟΣΑΝΑΤΟΛΙΣΜΕΝΟΣ SURF ΠΕΡΙΓΡΑΦΕΑΣ	20
ΕΙΚΟΝΑ 20: ΔΙΑΦΟΡΑ ΤΟΥ GAUSSIAN ΦΙΛΤΡΟΥ (DoG) ΤΟ ΟΠΟΙΟ ΧΡΗΣΙΜΟΠΟΙΕΙΤΑΙ ΑΠΟ ΤΟΝ SIFT	22
ΕΙΚΟΝΑ 21: ΦΙΛΤΡΑ ΠΑΡΑΘΥΡΟΥ SURF ΓΙΑ ΤΗΝ ΠΡΟΣΕΓΓΙΣΗ ΔΕΥΤΕΡΗΣ ΤΑΞΗΣ GAUSSIAN ΠΑΡΑΓΩΓΩΝ	22
ΕΙΚΟΝΑ 22: ΠΑΡΑΔΕΙΓΜΑ ΑΝΤΙΣΤΟΙΧΙΩΝ ΣΕ ΙΔΙΟ ΖΕΥΓΑΡΙ ΕΙΚΟΝΩΝ ΜΕ SURF & SIFT	23
ΕΙΚΟΝΑ 23: ΑΠΟΤΥΧΗΜΕΝΗ ΠΡΟΣΠΑΘΕΙΑ ΑΝΙΧΝΕΥΣΗΣ ΧΑΡΑΚΤΗΡΙΣΤΙΚΩΝ ΜΕ SURF	24
ΕΙΚΟΝΑ 24: ΤΟ ΠΡΟΒΛΗΜΑ ΤΗΣ ΠΡΟΟΠΤΙΚΗΣ ΠΑΡΑΜΟΡΦΩΣΗΣ ΜΕ SURF & SIFT	24
ΕΙΚΟΝΑ 25: ΕΠΙΣΚΟΠΗΣΗ ΤΟΥ ASIFT ΑΛΓΟΡΙΘΜΟΥ	26
ΕΙΚΟΝΑ 26: ΠΑΡΑΔΕΙΓΜΑ ΥΨΗΛΗΣ ΚΛΙΣΗΣ ΜΕΤΑΒΑΣΗΣ	27
ΕΙΚΟΝΑ 27: ΣΥΓΚΡΙΣΗ ASIFT & SIFT ΣΤΟ ΙΔΙΟ ΖΕΥΓΟΣ ΕΙΚΟΝΩΝ	28
ΕΙΚΟΝΑ 28: ΣΥΓΚΡΙΣΗ ASIFT & SIFT ΣΕ ΠΑΡΑΜΟΡΦΩΜΕΝΟ ΖΕΥΓΟΣ ΕΙΚΟΝΩΝ	28
ΕΙΚΟΝΑ 29: ΒΑΣΙΚΑ ΣΤΑΔΙΑ ΤΟΥ ΑΛΓΟΡΙΘΜΟΥ HOG	30
ΕΙΚΟΝΑ 30: ΔΗΜΙΟΥΡΓΙΑ ΙΣΤΟΓΡΑΜΜΑΤΩΝ ΕΝΤΟΣ ΤΩΝ ΚΕΛΙΩΝ	30
ΕΙΚΟΝΑ 31: ΓΕΩΜΕΤΡΙΚΑ ΤΜΗΜΑΤΑ ΠΕΡΙΓΡΑΦΕΩΝ HOG	31
ΕΙΚΟΝΑ 32: ΕΝΝΟΙΟΛΟΓΙΚΗ ΑΠΕΙΚΟΝΙΣΗ ΤΗΣ ΔΙΑΔΙΚΑΣΙΑΣ ΤΑΞΙΝΟΜΗΣΗΣ ΕΙΚΟΝΩΝ	33
ΕΙΚΟΝΑ 33: ΚΛΑΣΣΙΚΗ ΜΟΡΦΗ ΤΕΧΝΗΤΟΥ ΝΕΥΡΩΝΙΚΟΥ ΔΙΚΤΥΟΥ	34
ΕΙΚΟΝΑ 34: Η ΑΡΧΗ ΤΩΝ ΜΗΧΑΝΩΝ ΥΠΟΣΤΗΡΙΞΗΣ ΔΙΑΝΥΣΜΑΤΩΝ (SVM)	36
ΕΙΚΟΝΑ 35: ΔΙΑΧΩΡΙΣΙΜΑ ΚΑΙ ΜΗ ΔΙΑΧΩΡΙΣΙΜΑ ΔΕΙΓΜΑΤΑ	37
ΕΙΚΟΝΑ 36: ΑΛΓΟΡΙΘΜΟΣ ΒΑΣΙΣΜΕΝΟΣ ΣΕ ΣΤΙΓΜΙΟΤΥΠΑ	40
ΕΙΚΟΝΑ 37: ΠΑΡΑΔΕΙΓΜΑ ΑΝΑΓΝΩΡΙΣΗΣ ΑΝΤΙΚΕΙΜΕΝΩΝ ΣΕ ΕΙΚΟΝΑ	44
ΕΙΚΟΝΑ 38: ΑΝΑΓΝΩΡΙΣΗ ΑΝΤΙΚΕΙΜΕΝΩΝ ΜΕΣΩ ΑΝΤΙΣΤΟΙΧΗΣΗΣ ΜΟΝΤΕΛΩΝ	45
ΕΙΚΟΝΑ 39: ΑΝΑΓΝΩΡΙΣΗ ΑΝΤΙΚΕΙΜΕΝΩΝ ΜΕΣΩ ΤΜΗΜΑΤΙΚΩΝ ΑΝΑΠΑΡΑΣΤΑΣΕΩΝ	46
ΕΙΚΟΝΑ 40: ΑΝΑΓΝΩΡΙΣΗ ΜΕΣΩ ΠΑΡΑΜΟΡΦΩΣΙΜΩΝ ΜΟΝΤΕΛΩΝ ΤΩΝ ΑΝΤΙΚΕΙΜΕΝΩΝ	47
ΕΙΚΟΝΑ 41: ΒΑΣΙΚΑ ΒΗΜΑΤΑ ΓΙΑ ΤΗΝ ΠΑΡΑΓΩΓΗ ΕΝΟΣ ΧΑΡΤΗ ΠΡΟΕΞΕΧΟΝΤΩΝ ΣΗΜΕΙΩΝ ΣΕ ΜΙΑ ΕΙΚΟΝΑ	49
ΕΙΚΟΝΑ 42: ΚΟΜΒΟΣ ΓΙΑ ΤΟΠΟΘΕΣΙΑ ΣΕ ΕΝΑ ΧΑΡΤΗ ΕΙΣΟΔΟΥ	49
ΕΙΚΟΝΑ 43: ΚΑΤΕΥΘΥΝΣΕΙΣ ΤΩΝ ΑΚΜΩΝ ΜΕΤΑΞΥ ΚΟΜΒΩΝ	50
ΕΙΚΟΝΑ 44: ΠΡΟΣΘΗΚΗ ΒΑΡΩΝ ΣΤΙΣ ΑΚΜΕΣ	50
ΕΙΚΟΝΑ 45: ΣΥΓΚΡΙΣΗ ΜΕΘΟΔΩΝ ΑΠΟΚΤΗΣΗΣ ΧΑΡΤΩΝ ΠΡΟΕΞΕΧΟΝΤΩΝ ΣΗΜΕΙΩΝ	51
ΕΙΚΟΝΑ 46: ΤΑ ΤΡΙΑ ΒΑΣΙΚΑ ΒΗΜΑΤΑ ΠΟΥ ΕΓΙΝΑΝ ΚΑΤΑ ΤΗΝ ΥΛΟΠΟΙΗΣΗ ΤΗΣ ΕΦΑΡΜΟΓΗΣ	53
ΕΙΚΟΝΑ 47: ΧΡΗΣΗ ΤΟΥ ΑΛΓΟΡΙΘΜΟΥ "ΠΡΟΕΞΕΧΟΝΤΩΝ ΟΠΤΙΚΩΝ ΣΗΜΕΙΩΝ ΒΑΣΙΣΜΕΝΑ ΣΕ ΓΡΑΦΟΥΣ"	54

ΕΙΚΟΝΑ 48: ΣΥΝΟΛΙΚΗ ΕΠΙΣΚΟΠΗΣΗ ΤΗΣ ΔΙΑΔΙΚΑΣΙΑΣ ΠΟΥ ΑΚΟΛΟΥΘΗΘΗΚΕ	55
ΕΙΚΟΝΑ 49: ΤΑ ΤΕΣΣΕΡΑ ΤΟΥΡΙΣΤΙΚΑ ΜΝΗΜΕΙΑ ΤΟΥ ΗΡΑΚΛΕΙΟΥ, ΤΑ ΟΠΟΙΑ ΧΡΗΣΙΜΟΠΟΙΗΘΗΚΑΝ ΣΤΗΝ ΕΦΑΡΜΟΓΗ	56
ΕΙΚΟΝΑ 50: ΣΗΜΕΙΑ ΚΛΕΙΔΙΑ ΣΤΗΝ ΑΡΧΙΚΗ ΕΙΚΟΝΑ	57
ΕΙΚΟΝΑ 51: ΣΗΜΕΙΑ ΚΛΕΙΔΙΑ ΣΤΗΝ ΕΠΕΞΕΡΓΑΣΜΕΝΗ ΜΕ GBVS ΕΙΚΟΝΑ	57
ΕΙΚΟΝΑ 52: ΣΥΝΟΛΟ ΑΝΤΙΣΤΟΙΧΙΩΝ ΜΕΤΑΞΥ ΔΥΟ ΑΡΧΙΚΩΝ ΕΙΚΟΝΩΝ	58
ΕΙΚΟΝΑ 53: ΣΥΝΟΛΟ ΑΝΤΙΣΤΟΙΧΙΩΝ ΜΕΤΑΞΥ ΑΡΧΙΚΗΣ ΚΑΙ ΠΡΟ-ΕΠΕΞΕΡΓΑΣΜΕΝΗΣ ΕΙΚΟΝΑΣ	58
ΕΙΚΟΝΑ 54: ΣΥΝΟΛΟ ΑΝΤΙΣΤΟΙΧΙΩΝ ΜΕΤΑΞΥ ΔΥΟ ΠΡΟ-ΕΠΕΞΕΡΓΑΣΜΕΝΩΝ ΕΙΚΟΝΩΝ	59

ΕΥΡΕΤΗΡΙΟ ΠΙΝΑΚΩΝ

ΠΙΝΑΚΑΣ 1: ΣΥΝΟΛΙΚΗ ΠΑΡΟΥΣΙΑΣΗ ΤΩΝ ΜΕΘΟΔΩΝ ΕΞΑΓΩΓΗΣ ΧΑΡΑΚΤΗΡΙΣΤΙΚΩΝ.....	32
ΠΙΝΑΚΑΣ 2:ΣΥΝΟΛΙΚΗ ΠΑΡΟΥΣΙΑΣΗ ΤΩΝ ΑΛΓΟΡΙΘΜΩΝ ΓΙΑ ΤΑΞΙΝΟΜΗΣΗ ΕΙΚΟΝΩΝ.....	39
ΠΙΝΑΚΑΣ 3: ΣΥΓΚΕΝΤΡΩΤΙΚΑ ΠΕΙΡΑΜΑΤΙΚΑ ΑΠΟΤΕΛΕΣΜΑΤΑ ΤΑΞΙΝΟΜΗΣΕΩΝ & ΕΞΟΙΚΟΝΟΜΗΣΗΣ ΧΡΟΝΟΥ	60

1. Εισαγωγή

Η πτυχιακή αυτή εργασία αφορά την περιγραφή και δημιουργία μιας προσωπικά υλοποιημένης εφαρμογής για αυτόματη αναγνώριση αντικειμένων εικόνας, συνδυάζοντας πρωτοποριακές υλοποιήσεις αλγορίθμων στην επιστημονική περιοχή της Τεχνητής Όρασης (Computer Vision) και της Ψηφιακής Επεξεργασίας Εικόνας (Digital Image Processing).

Η ιδέα αυτής της εργασίας βασίστηκε στην προσωπική ενασχόληση με τις δυο παραπάνω επιστημονικές περιοχές καθώς και η φιλοδοξία δημιουργίας μιας εφαρμογής η οποία μελλοντικά θα είναι σε θέση να λειτουργήσει τόσο σε ερευνητικές όσο και σε πραγματικές συνθήκες χρήσης.

1.1 Περίληψη

Με τη συνεχόμενη ανάπτυξη της Επιστήμης Υπολογιστών [1] τα τελευταία χρόνια , αναπτύχθηκαν παράλληλα και αρκετοί νέοι κλάδοι οι οποίοι περιγράφουν και εξομοιώνουν διάφορες βιολογικές λειτουργίες του ανθρώπινου οργανισμού, όπως η όραση και η ακοή. Το βιολογικό σύστημα της όρασης ουσιαστικά αποτελείται από τον συνδυασμό ενός οπτικού συστήματος και ενός συστήματος ερμηνείας της πληροφορίας.

Ως άνθρωποι αντιλαμβανόμαστε την τρισδιάστατη δομή του κόσμου γύρω μας με εμφανή ευκολία. Σκεφτείτε πόσο έντονα αντιλαμβανόμαστε την τρισδιάστατη υπόσταση ενός βάζου με λουλούδια, πάνω στο τραπέζι. Μπορούμε με άνεση να πούμε το σχήμα και την διαύγεια κάθε πετάλου μέσω των λεπτών σχεδίων του φωτός και των σκιών που εμφανίζονται σε ολόκληρη την επιφάνεια του λουλουδιού. Αντιληπτικοί ψυχολόγοι έχουν περάσει δεκαετίες προσπαθώντας να κατανοήσουν πως λειτουργεί το οπτικό σύστημα του ανθρώπινου οργανισμού. Γιατί όμως η όραση είναι τόσο δύσκολο στο να την αναπαράγουμε μέσω υπολογιστή; Εν μέρει αυτό συμβαίνει επειδή η όραση είναι ένα αντίστροφο πρόβλημα, στο οποίο επιδιώκουμε να ανακτήσουμε κάτι άγνωστο, δεδομένης ανεπαρκούς πληροφορίας, για να προσδιορίσουμε μια λύση. Πρέπει λοιπόν να καταφύγουμε σε βασισμένα σε φυσική και πιθανότητες μοντέλα για να αποσαφηνίσουμε τις πιθανές λύσεις. Ωστόσο η μοντελοποίηση του οπτικού κόσμου σε όλη του την πολυπλοκότητα είναι πολύ πιο δύσκολο από ότι ας πούμε η μοντελοποίηση της φωνητικής οδού που παράγει ήχους.

Στην παρούσα εργασία γίνεται προσπάθεια για την εξομοίωση της βιολογικής όρασης αλλά και την παραγωγή γνώσεων και αποφάσεων μέσω αυτής. Συγκεκριμένα γίνεται μια προσπάθεια για αυτόματη αναγνώριση αντικειμένων εικόνας. Πάνω σε αυτό το πλαίσιο, αναπτύχθηκε μια εφαρμογή σκοπός της οποίας είναι η παραγωγή γνώσης (γενικές πληροφορίες, όνομα, τύπος μνημείου κτλ) για τουριστικές περιοχές. Πιο συγκεκριμένα η εφαρμογή, η οποία θα αναπτυχθεί λεπτομερώς στο [Κεφάλαιο 5](#), στοχεύει στην προεπεξεργασία πραγματικών εικόνων από διάφορα μνημεία της πόλης του Ηρακλείου. Αυτή η προεπεξεργασία θα βελτιώσει στη συνέχεια τον απαιτούμενο χρόνο για τη σύγκριση μεταξύ φωτογραφιών από τη βάση δεδομένων και αυτών που έχουν ληφθεί σε πραγματικό, μέσω συγκεκριμένων αλγορίθμων, καταλήγοντας στην παραγωγή των απαραίτητων πληροφοριών που αφορούν το μνημείο που ο χρήστης θα έχει μπροστά του.

1.2 Κίνητρο για την διεξαγωγή της εργασίας

Το κίνητρο για τη διεξαγωγή αυτής της εργασίας αρχικά ήταν η δημιουργία μιας εφαρμογής η οποία θα ήταν σε θέση να χρησιμοποιηθεί εύκολα από τον μέσο άνθρωπο χωρίς ιδιαίτερες γνώσεις στην επιστήμη των υπολογιστών. Πιο συγκεκριμένα η κατασκευή αυτής καθαυτής της εφαρμογής είχε απώτερο στόχο την προσπάθεια συμβολής σε αυτή την αναπτυσσόμενη επιστημονική περιοχή έρευνας, προσφέροντας ταυτόχρονα σημαντική εμπειρία και εμπλουτισμό γνώσεων στον συγγραφέα.

1.3 Σκοπός και στόχοι της εργασίας

Η παρούσα εργασία έχει σαν σκοπό την ανάδειξη καινοτόμων τεχνολογιών στους επιστημονικούς τομείς της τεχνητής όρασης, της ψηφιακής επεξεργασίας εικόνας και της τεχνητής νοημοσύνης. Προς αυτή την κατεύθυνση, έγινε προσπάθεια για τη δημιουργία μιας προσωπικής εφαρμογής η οποία στηρίζεται στην σύμπραξη των τριών τομέων που αναφέρθηκαν παραπάνω. Απώτερος σκοπός κατά τη σύλληψη της συγκεκριμένης ιδέας ήταν η έστω και μικρή εξέλιξη των ήδη υπαρχουσών συστημάτων στον τομέα της τεχνητής όρασης. Άλλος ένας στόχος που είχε τεθεί εξαρχής ήταν το σύστημα να βασιστεί σε πραγματικές λειτουργίες και να δοκιμαστεί υπό πραγματικές συνθήκες. Η χρήση της συγκεκριμένης εφαρμογής χωρίς να συντελεστούν σημαντικές αλλαγές, τουλάχιστον στο κομμάτι της αρχικής ιδέας, καθώς οποιαδήποτε εφαρμογή δύναται να καλυτερεύει με το πέρασμα του χρόνου, θα μπορούσε εύκολα να χρησιμοποιηθεί ως ένας τουριστικός οδηγός ακόμη και με τη χρήση κινητών τηλεφώνων.

1.4 Δομή της εργασίας

Αυτή η εργασία χωρίστηκε σε τμήματα στοχεύοντας στην ευκολότερη κατανόηση και εκτίμηση της. Επίσης ακολουθώντας μια συγκεκριμένη τέτοια πορεία, ο αναγνώστης διευκολύνεται στην παρακολούθηση του κάθε κεφαλαίου αλλά και στο σύνολο αυτής της εργασίας. Κλείνοντας λοιπόν τα εισαγωγικά, ακολουθεί μια σύντομη αναφορά στα κεφάλαια που θα ακολουθήσουν και τα θέματα που θα αναλυθούν σε αυτά.

Στο δεύτερο κεφάλαιο θα παρουσιαστούν οι σημαντικότερες, τουλάχιστον μέχρι τη στιγμή συγγραφής αυτής της εργασίας, μέθοδοι για την εξαγωγή χαρακτηριστικών σε εικόνες, όπως αναφέρθηκαν στον ορισμό τους παραπάνω. Με αυτή την παρουσίαση θα γίνουν γνωστά στον αναγνώστη τόσο η διαδικασία που ακολουθεί η κάθε μέθοδος όσο και ομοιότητες και διαφορές μεταξύ τους οι οποίες παίζουν σημαντικό ρόλο για την τελική επιλογή της μεθόδου που θα χρησιμοποιηθεί.

Στα επόμενα δυο κεφάλαια θα γίνει λόγος για μια πολύ σημαντική διαδικασία της τεχνητής όρασης, την κατηγοριοποίηση (3^ο Κεφάλαιο) και αναγνώριση αντικειμένων (4^ο Κεφάλαιο) στα πλαίσια μιας εικόνας. Με το τέλος αυτών των κεφαλαίων, ο αναγνώστης θα έχει καταλάβει πως λειτουργούν οι διάφοροι μέθοδοι αναγνώρισης και κατηγοριοποίησης, ενώ θα έχει λάβει και το θεωρητικό υπόβαθρο για την εφαρμογή που θα ακολουθήσει.

Στο πέμπτο και τελευταίο κεφάλαιο θα γίνει μια λεπτομερής παρουσίαση της εφαρμογής που υλοποιήθηκε. Ο αναγνώστης θα είναι σε θέση να κατανοήσει τον τρόπο λειτουργίας της αλλά και των μελλοντικών εφαρμογών που δύναται να συναντήσει στα επιστημονικά πεδία της τεχνητής όρασης και επεξεργασίας εικόνας.

1.5 Θεωρητικό Υπόβαθρο και Επισκόπηση Ερευνητικού Πεδίου

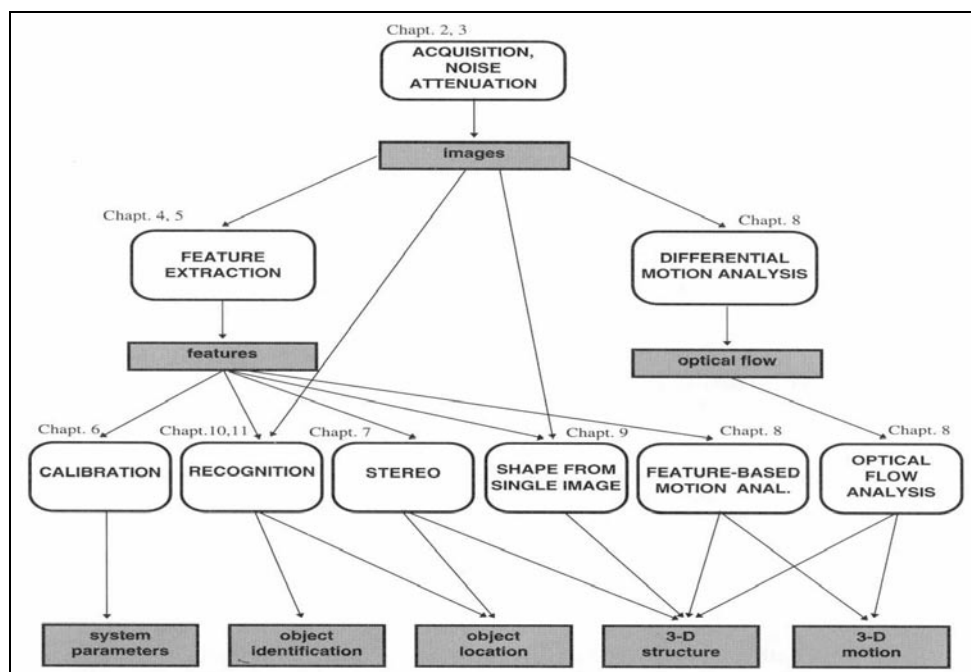
Κρίνεται σκόπιμο να αναφερθούμε στην τεχνητή όραση, ένα ανερχόμενο επιστημονικό πεδίο στο χώρο της επιστήμης υπολογιστών. Επίσης θα γίνει εκτενής αναφορά και στην επιστημονική περιοχή της Ψηφιακής Επεξεργασίας Εικόνας, στοχεύοντας στο να καταδείξουμε την άρρηκτα συνδεδεμένη σχέση των δυο περιοχών αλλά και για να γίνουν κατανοητά τα πλαίσια στα οποία βασίστηκε η εργασία αυτή. Επιπλέον θα αναφερθούμε σε μια ικανότητα του ανθρώπινου οργανισμού η οποία γνωρίζει μεγάλη ανάπτυξη στην περιοχή της επιστήμης υπολογιστών, τα λεγόμενα προεξέχοντα οπτικά σημεία εικόνας. Αυτές οι αναφορές δεν στοχεύουν απλά στην παροχή γνώσεων

για τον αναγνώστη αλλά και στην προσπάθεια να γίνει καλύτερα αντιληπτό το πώς χρησιμοποιήθηκαν στην εφαρμογή που ακολουθεί.

1.5.1 Τεχνητή Όραση

Τεχνητή όραση [2], γνωστή στη βιβλιογραφία και με τον όρο υπολογιστική όραση, καλείται το επιστημονικό πεδίο της επιστήμης υπολογιστών, το οποίο περιλαμβάνει μεθόδους για απόκτηση, επεξεργασία, ανάλυση και κατανόηση εικόνων του πραγματικού κόσμου στοχεύοντας στην παραγωγή αριθμητικής ή συμβολικής πληροφορίας. Με άλλα λόγια, τεχνητή όραση θεωρείται η επιστήμη και τεχνολογία δημιουργίας συστημάτων ανάλυσης και επεξεργασίας πληροφορίας από οπτικά σήματα, λαμβανόμενα από μηχανές, στοχεύοντας στην πρακτική εφαρμογή τους. Συμπερασματικά μπορούμε να πούμε ότι η τεχνητή όραση είναι κατά κάποιον τρόπο το αντίστροφο της γραφικής υπολογιστών [3]. Ενώ η γραφική παράγει δεδομένα εικόνες από τρισδιάστατα μοντέλα, η τεχνητή όραση συνήθως παράγει τρισδιάστατα μοντέλα από δεδομένα μιας εικόνας. Έναυσμα για την ανάπτυξη αυτού του τομέα θεωρείται η ηλεκτρονική προσπάθεια προσομοίωσης των ικανοτήτων του ανθρώπινου συστήματος όρασης, για την αντίληψη και κατανόηση μιας εικόνας. Εν μέρη τεχνητή όραση είναι ανάλογη με τη μετατροπή της οπτικής αίσθησης στην οπτική αντίληψη της βιολογικής όρασης.

Η ιεραρχική διάκριση των λειτουργιών της τεχνητής όρασης είναι αρκετά σημαντική. Συγκεκριμένα, η διάκριση αυτή μπορεί να γίνει ως εξής: α)Χαμηλού επιπέδου όραση, κατά την οποία γίνεται η εξαγωγή χαρακτηριστικών από την εικόνα. β)Ενδιάμεσο επίπεδο όρασης, κατά την οποία αναγνωρίζονται αντικείμενα και ερμηνεύεται η τρισδιάστατη εικόνα χρησιμοποιώντας τα χαρακτηριστικά που προέρχονται από το χαμηλότερο επίπεδο λειτουργιών. γ)Υψηλού επιπέδου όραση, κατά την οποία πραγματοποιείται η ερμηνεία της εξελισσόμενης πληροφορίας η οποία παρέχεται από το ενδιάμεσο επίπεδο όρασης. Η ερμηνεία αυτή, μπορεί να περιλαμβάνει την εννοιολογική περιγραφή μιας σκηνής, όπως την δραστηριότητα, την πρόθεση και τη συμπεριφορά.

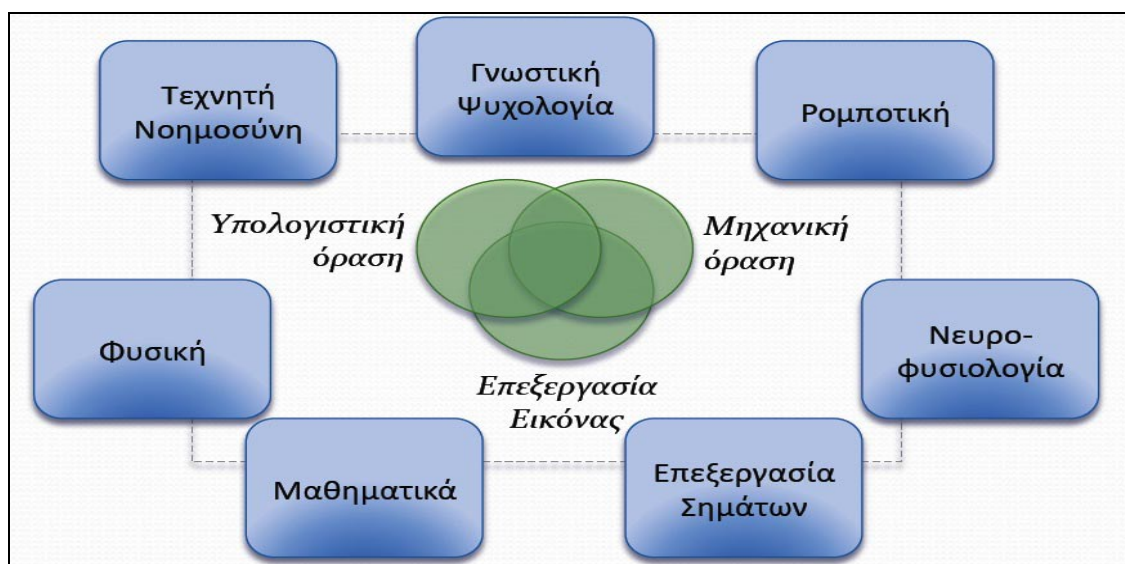


Εικόνα 1: Ιεραρχική διάκριση λειτουργιών Τεχνητής όρασης

Η τεχνητή όραση ως επιστημονικό πεδίο ,ξεκίνησε τη δράση της κατά την δεκαετία του 70 και από τότε συνεχώς εξελίσσεται. Επικεντρώνοντας το ενδιαφέρον μας στην τελευταία δεκαετία, μπορούμε να πούμε ότι η τεχνητή όραση συνέχισε να εμβαθύνει στην αλληλεπίδραση μεταξύ της όρασης και των πεδίων της γραφικής. Ειδικότερα, πολλά από τα θέματα που εισάγονται κάτω από το γενικό σύνθημα εικόνα με βάση την απόδοση, όπως η συρραφή εικόνων και η καταγραφή βάσης του φωτός.

Μια δεύτερη σημαντική τάση της τελευταίας δεκαετίας, η οποία έχει και άμεση σχέση με την παρούσα εργασία, θεωρείται η εμφάνιση τεχνικών βασισμένες σε χαρακτηριστικά των αντικειμένων στοχεύοντας στην αναγνώριση τους. Τέτοιου είδους τεχνικές κυριαρχούν και σε άλλες εργασίες αναγνώρισης, όπως η αναγνώριση σκηνής, πανοραμική θέαση και αναγνώριση τοποθεσίας. Παρόλα αυτά και ενώ το σημείο στο οποίο επικεντρώνεται η παρούσα έρευνα είναι η αναγνώριση βάσης χαρακτηριστικών, κάποιες ομάδες ερευνητών επιδιώκουν την αναγνώριση η οποία βασίζεται σε καμπύλες και κατάτμηση περιοχών.

Η χρήση της τεχνητής όρασης είναι συχνό φαινόμενο ιδίως σε υψηλής τεχνολογίας εφαρμογές όπως στην βιομηχανία, την ιατρική ακόμα και στην εξερεύνηση του διαστήματος. Έτσι αναπόφευκτα έχει άμεση σχέση με άλλους τομείς των επιστημών όπως η επεξεργασία σημάτων, φυσική, τα μαθηματικά, η τεχνητή νοημοσύνη και η νευρο-βιολογία.



Εικόνα 2 : Σχέση Τεχνητής όρασης με άλλες επιστήμες

Κοιτάζοντας σε βάθος τη σχέση της τεχνητής όρασης με άλλους τομείς μπορεί κανείς εύκολα να βρει σημαντικές αλληλεξαρτήσεις. Προς αυτή την κατεύθυνση οι περιοχές της τεχνητής νοημοσύνης που ασχολούνται με τον αυτόνομο σχεδιασμό και τα ρομποτικά συστήματα, συνδέονται άμεσα με την τεχνητή όραση η οποία μπορεί να παρέχει ένα σύστημα που ενεργεί ως αισθητήρας ο οποίος είναι σε θέση να παρέχει πληροφορίες υψηλού επιπέδου τόσο για το περιβάλλον όσο και για το ίδιο το ρομπότ. Επιπλέον η τεχνητή όραση και η τεχνητή νοημοσύνη, μοιράζονται και άλλα κοινά θέματα όπως η αναγνώριση προτύπων και τεχνικές εκμάθησης. Τα πεδία που συνδέονται πιο στενά με την τεχνητή όραση θεωρούνται η επεξεργασία και ανάλυση εικόνας, καθώς και η μηχανική όραση[4]. Παρακάτω θα γίνει ειδική αναφορά στο πεδίο της ψηφιακής επεξεργασίας εικόνας.

Πριν αναφερθούμε στις διάφορες εφαρμογές της τεχνητής όρασης, θεωρείται σκόπιμο να γίνει αναφορά στους παράγοντες που κάνουν τόσο δύσκολη τη διαδικασία για έναν Η/Υ να μιμηθεί τον τρόπο όρασης των ανθρώπων.

Πιο συγκεκριμένα θα παρουσιάσουμε 6 λόγους οι οποίοι αποδεικνύουν τα παραπάνω:

- Απώλεια των πληροφοριών κατά τη μετατροπή της εικόνας από 3 σε 2 διαστάσεις, λόγω του μετασχηματισμού προοπτικής.
- Μια ποικιλία από επιφάνειες με διαφορετικό υλικό και γεωμετρικές ιδιότητες, πιθανότατα κάτω από διαφορετικές συνθήκες φωτισμού θα μπορούσε να οδηγήσει σε όμοιες εικόνες.
- Η μετρούμενη φωτεινότητα δίνεται από ένα περίπλοκο φυσικό σχηματισμό της εικόνας, άρα η αντίστροφη διαδικασία είναι αρκετά δύσκολη.
- Η παρουσία θορύβου υπάρχει πάντα όπως και σε κάθε μέτρηση του πραγματικού κόσμου.
- Πληθώρα δεδομένων, καθώς ένα φύλλο μεγέθους A4 με ανάλυση 300 dots per pixel θα καταλαμβάνει 8.5 Mbytes αποθηκευτικού χώρου.
- Απαιτείται ερμηνεία για να παραχθεί η τελική απόφαση.

Ακολουθούν οι σημαντικότερες εφαρμογές στις οποίες συναντάμε την τεχνητή όραση.

- Συστήματα ελέγχου της παραγωγικής διαδικασίας, συναντιούνται κυρίως σε ρομποτικές εφαρμογές σε βιομηχανικό πλαίσιο και στα αυτόνομα οχήματα.
- Συστήματα ποιοτικού ελέγχου σε γραμμές παραγωγής.
- Συστήματα παρακολούθησης, χρησιμοποιώντας κάμερες ασφαλείας για την καταμέτρηση αντικειμένων.
- Συστήματα προηγμένης διάδρασης με εφαρμογές επαυξημένης πραγματικότητας [5].
- Οργάνωση πληροφορίας με τη μέθοδο της δεικτοδότησης βάσεων εικόνων.
- Συστήματα αναγνώρισης, χρησιμοποιούνται κατά κόρον σε ιατρικές και τοπογραφικές εφαρμογές, καθώς επίσης και για την αναγνώριση χαρακτήρων και ιδιαίτερα προσώπων.

Συνοπτικά μπορούμε να πούμε ότι η οργάνωση ενός συστήματος τεχνητής όρασης έχει άμεση σχέση με την εφαρμογή που υλοποιείται. Έτσι ορισμένα συστήματα χρησιμοποιούν αυτόνομες εφαρμογές που λύνουν ένα συγκεκριμένο πρόβλημα, ενώ άλλες εφαρμογές αποτελούν ένα υπό-σύστημα ενός μεγαλύτερου συστήματος. Οποιαδήποτε και αν είναι η εφαρμογή, υπάρχουν κάποιες βασικές λειτουργίες οι οποίες συναντώνται σε όλα τα συστήματα τεχνητής όρασης.

Αρχικά απαιτείται να αποκτηθεί η εικόνα, η οποία ανάλογα με την τεχνική που χρησιμοποιείται προκύπτει ως δισδιάστατη ή τρισδιάστατη. Η εικόνα αποτελεί την κύρια πληροφορία για ένα σύστημα τεχνητής όρασης. Επόμενο χρονικά στάδιο είναι αυτό της προ-επεξεργασίας της εικόνας. Τα δεδομένα της εικόνας επεξεργάζονται πριν την καθεαυτή χρήση τους από το σύστημα της τεχνητής όρασης, για να πληρούν κάποιες προϋποθέσεις. Σημαντικό βήμα στη συνέχεια είναι η εξαγωγή χαρακτηριστικών της εικόνας, δηλαδή γραμμές, περιγράμματα και γωνίες (αναλύεται λεπτομερώς στο Κεφάλαιο 2 της παρούσας εργασίας). Ως επόμενο στάδιο μπορεί να θεωρηθεί αυτό της κατάτμησης της εικόνας για να ακολουθήσει το τελευταίο στάδιο, κατά το οποίο παράγεται η απόφαση του συστήματος τεχνητής όρασης και αποφασίζεται η αντίδραση της εφαρμογής σε αυτή την απόφαση.

1.5.2 Ψηφιακή Επεξεργασία Εικόνας

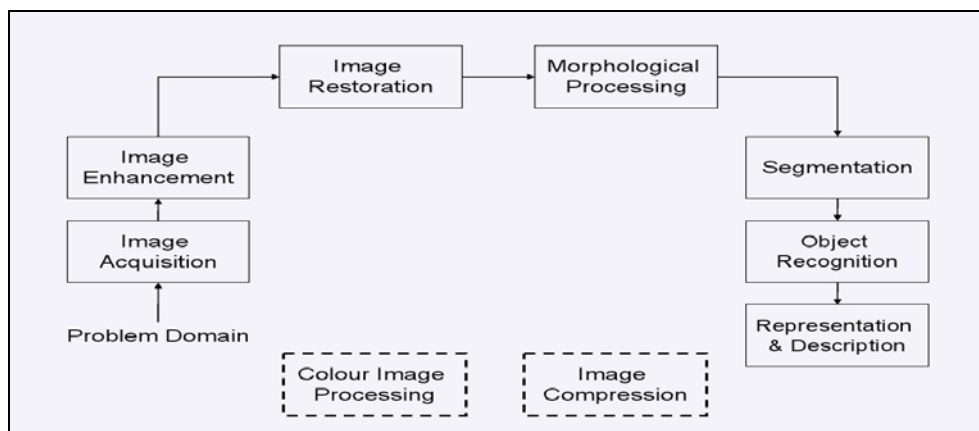
Επεξεργασία εικόνας [6] θεωρείται οποιαδήποτε μορφή επεξεργασίας σήματος για την οποία είσοδος είναι μια εικόνα, όπως για παράδειγμα φωτογραφίες ή το πλαίσιο ενός βίντεο. Η έξοδος αυτής της επεξεργασίας μπορεί να είναι είτε μια εικόνα είτε ένα σύνολο χαρακτηριστικών τα οποία σχετίζονται με την εικόνα. Οι περισσότερες τεχνικές επεξεργασίας εικόνας περιλαμβάνουν τον χειρισμό της εικόνας ως δισδιάστατο σήμα, εφαρμόζοντας σε αυτό πρότυπες τεχνικές επεξεργασίας σήματος. Η επεξεργασία εικόνας αναφέρεται συνήθως στην Ψηφιακή Επεξεργασία Εικόνας, χωρίς να αποκλείονται παρόλα αυτά η οπτική και η αναλογική επεξεργασία εικόνας.

Ψηφιακή εικόνα θεωρείται η αναπαράσταση μιας δισδιάστατης εικόνας ως ένα πεπερασμένο σύνολο από ψηφιακές τιμές που ονομάζονται pixels. Η ψηφιοποίηση συνεπάγεται ότι μια ψηφιακή εικόνα είναι προσέγγιση μιας πραγματικής σκηνής.

Ψηφιακή Επεξεργασία Εικόνας [7] ονομάζεται η χρήση κατάλληλων υπολογιστικών αλγορίθμων για την εκτέλεση της διαδικασίας που ονομάζεται επεξεργασία, πάνω σε ψηφιακές εικόνες. Η Ψηφιακή επεξεργασία εικόνας θεωρείται ένα υποσύνολο του ηλεκτρονικού τομέα όπου η εικόνα μετατρέπεται σε μια σειρά μικρών ακεραίων, τα λεγόμενα pixels, που αντιπροσωπεύουν μια φυσική ποσότητα όπως η σκηνή, η ακτινοβολία και αποθηκεύονται σε ψηφιακή μνήμη για να

υποβληθούν στη συνέχεια σε επεξεργασία από τον υπολογιστή ή αλλά ψηφιακά μέσα. Είτε ως ενίσχυση για τους ανθρώπινους παρατηρητές είτε για εκτέλεση αυτόνομων αναλύσεων η ψηφιακή επεξεργασία εικόνας προσφέρει πλεονεκτήματα στο κόστος, στην ταχύτητα, στην ευελιξία και με τη ραγδαία αύξηση της απόδοσης των προσωπικών υπολογιστών θεωρείται η κυρίαρχη μέθοδος πλέον.

Με την έλευση της ψηφιακής φωτογραφίας η ψηφιακή εικόνα γίνεται πλέον κτήμα του ευρύτερου κοινού. Η Ψηφιακή Επεξεργασία Εικόνων, με μια ιστορία περίπου σαράντα ετών, θεωρείται ήδη μια ώριμη επιστημονική περιοχή, η οποία όπως αναφέρθηκε νωρίτερα είναι άρρηκτα συνδεδεμένη με την τεχνητή όραση. Ως αντικείμενο έχει εικόνες οποιασδήποτε προέλευση, προσιτές ή μη στην ανθρώπινη όραση. Περί τα μέσα της δεκαετίας του 1960, η εξερεύνηση του διαστήματος και η ιατρική απεικόνιση απέφεραν τη ζήτηση για επεξεργασία εικόνων. Τα ζητήματα που ετίθεντο αφορούσαν στη βελτίωση και αποκατάσταση των εικόνων διορθώνοντας αστοχίες και αδυναμίες των συστημάτων κατά τη διάρκεια της λήψης. Έκτοτε οι εφαρμογές βαίνουν αυξανόμενες και πολλαπλές ενώ τα προβλήματα γίνονται πιο σύνθετα, ενώ παράλληλα αυξάνονται και οι δυνατότητες καθώς η υπολογιστική ισχύς μεγαλώνει αδιάκοπα.



Εικόνα 3: Τα βασικά στάδια της Ψηφιακής Επεξεργασίας Εικόνας

Γενικά μπορούμε να πούμε ότι η Ψηφιακή Επεξεργασία Εικόνας αναπτύχθηκε για να αντιμετωπίσει τα ακόλουθα κύρια προβλήματα: α)Τη ψηφιοποίηση, κωδικοποίηση των εικόνων με στόχο την αποθήκευση και μετάδοση τους β)Τη βελτιστοποίηση και αποκατάσταση με στόχο την καλύτερη απεικόνιση και κατανόηση τους γ)Την τμηματοποίηση [8] και την περιγραφή εικόνων δ)Την ανάλυση και κατανόηση των εικόνων. Από την παραπάνω ανάλυση μπορούμε να συμπεράνουμε ότι τα θέματα που αντιμετωπίζει η Ψηφιακή Επεξεργασία Εικόνας αφορούν αφενός την αξιοποίηση των μέσων και αφετέρου την κατανόησης του περιεχομένου των εικόνων με απώτερο στόχο την προσέγγιση της ανθρώπινης όρασης όπως αναφέρθηκε και παραπάνω. Με την έννοια αυτή η Ψηφιακή Επεξεργασία Εικόνας ταυτίζεται με θέματα ρομποτικής όρασης, αναγνώρισης προτύπων, και τεχνητής νοημοσύνης.

Στο σημείο αυτό θα αναφερθούμε στις εφαρμογές της Ψηφιακής Επεξεργασίας Εικόνας που ολοένα και αυξάνονται με την πάροδο του χρόνου.

- Αποθορυβοποίηση [9]: Ο λεγόμενος λευκός θόρυβος αποτελεί τον συχνότερα εμφανιζόμενο θόρυβο στην επεξεργασία εικόνας, σε αυτό ακριβώς επικεντρώνεται η συγκεκριμένη εφαρμογή, στοχεύοντας στην ανάκτηση μιας εικόνας η οποία έχει εκτεθεί σε τέτοιου είδους θόρυβο.
- Κατάτμηση εικόνας: Είναι η διαδικασία τεμαχισμού μιας ψηφιακής εικόνας σε πολλά τμήματα. Στόχος αυτής της εφαρμογής είναι η απλοποίηση και η αλλαγή της αναπαράστασης σε κάτι πιο ουσιαστικό και εύκολο να αναλυθεί.
- Ανίχνευση προσώπου [10]: Είναι μια τεχνολογία που καθορίζει τις θέσεις και τα μεγέθη των ανθρώπινων προσώπων σε μια αυθαίρετη ψηφιακή εικόνα, εντοπίζοντας τα χαρακτηριστικά του προσώπου και αγνοώντας οτιδήποτε άλλο.
- Αναγνώριση χαρακτήρων: Είναι η μηχανική ή ηλεκτρονική μετατροπή των σκαναρισμένων

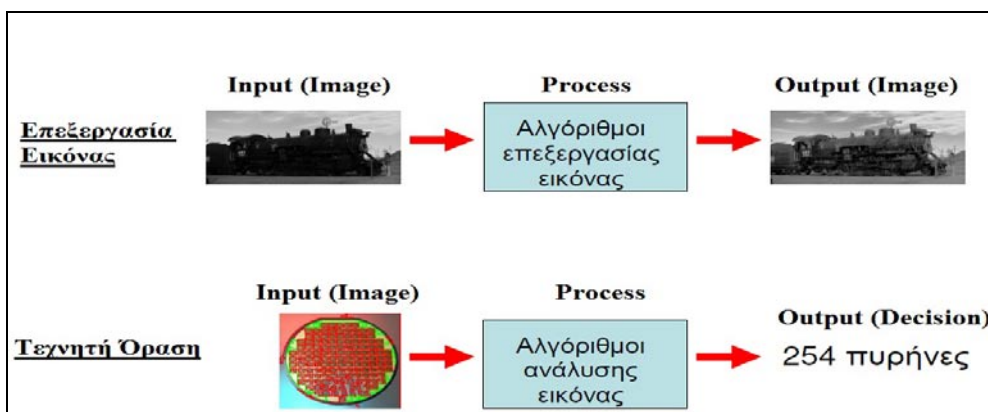
- ο χειρόγραφων ή τυπωμένων κειμένων σε κωδικοποιημένο κείμενο μηχανής.
- ο Ταυτοποίηση δακτυλικών αποτυπωμάτων: Αναφέρεται στην αυτοματοποιημένη μέθοδο εξακρίβωσης της αντιστοιχίας μεταξύ δυο ανθρωπίνων δακτυλικών αποτυπωμάτων.



Εικόνα 4: Εφαρμογές της Ψηφιακής Επεξεργασίας Εικόνας

1.5.3 Τεχνητή όραση εναντίον Ψηφιακής επεξεργασίας εικόνας

Η επεξεργασία εικόνας μελετά την μετατροπή από εικόνα σε εικόνα. Τόσο τα δεδομένα εισόδου όσο και οι πληροφορίες που εξάγονται πρόκειται για εικόνες. Οι τυπικές διαδικασίες της επεξεργασίας εικόνας περιλαμβάνουν συμπίεση, αποκατάσταση και ενίσχυση εικόνων. Αντιθέτως, τεχνητή όραση είναι η καταγραφή ρητά ουσιαστικών περιγραφών των φυσικών αντικειμένων που προέρχονται από τις εικόνες τους. Το αποτέλεσμα της τεχνητής όρασης είναι μια περιγραφή ή μια ερμηνεία ή μερικές ποσοτικές μετρήσεις της δομής σε μια σκηνή τριών διαστάσεων. Η επεξεργασία εικόνας και η αναγνώριση προτύπων βρίσκονται ανάμεσα στις πολλές τεχνικές τις οποίες χρησιμοποιεί η τεχνητή όραση για να επιτύχει τους στόχους της.



Εικόνα 5: Τεχνητή όραση εναντίον Ψηφ.Επεξεργασίας Εικόνας

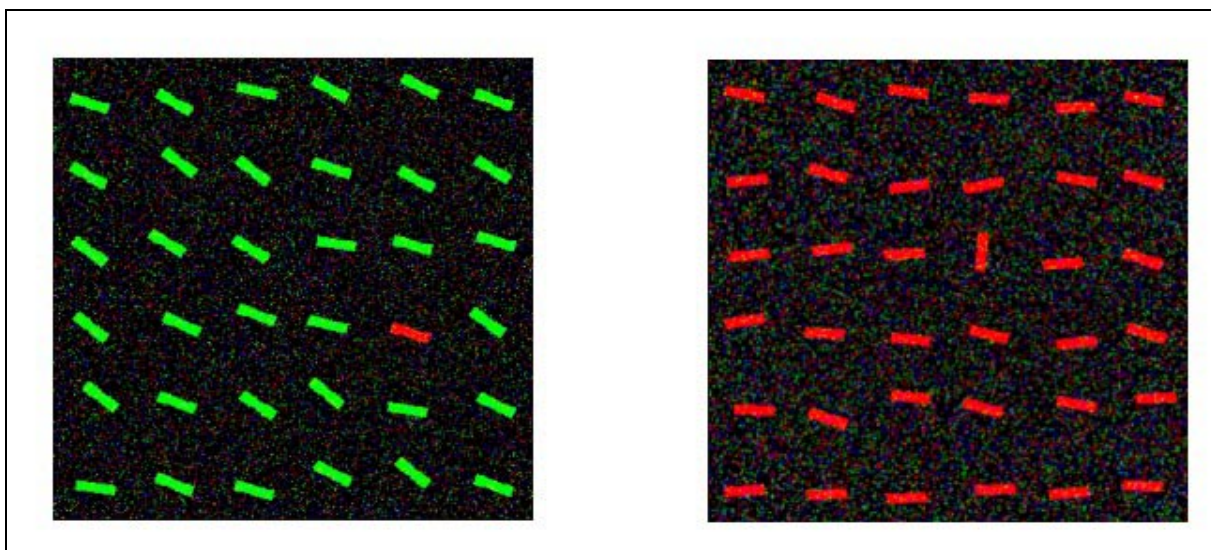
1.5.4 Προεξέχοντα οπτικά σημεία εικόνας (Visual Saliency)

Σε αυτή την παράγραφο γίνεται η προσπάθεια κατανόησης αυτής της σημαντικής ιδιότητας του ανθρώπινου οργανισμού, η οποία χρησιμοποιήθηκε στη δημιουργία της εφαρμογής που θα αναλύσουμε παρακάτω.

Τα περισσότερα σπονδυλωτά, συμπεριλαμβανομένου και του ανθρώπου, μπορούν να κινήσουν τα μάτια τους. Χρησιμοποιούν μάλιστα αυτή την ικανότητα για να πάρουν λεπτομερή δείγματα των πιο σημαντικών χαρακτηριστικών μια σκηνής, τη στιγμή που σπαταλούν περιορισμένους επεξεργαστικούς πόρους σε οτιδήποτε άλλο. Η περίοπτη θέση ενός στοιχείου [11], είτε πρόκειται για ένα αντικείμενο, ένα πρόσωπο, ένα pixel κτλ., είναι η κατάσταση ή ποιότητα, με την οποία ξεχωρίζει σε σχέση με τους γείτονες του. Η ικανότητα ανίχνευσης προεξέχων στοιχείων, θεωρείται βασικός μηχανισμός προσοχής ο οποίος διευκολύνει τη μάθηση και την επιβίωση, επιτρέποντας στους οργανισμούς να εστιάσουν την περιορισμένη αντίληψη τους και τους γνωστικούς πόρους στα πιο σχετικά υποσύνολα των διαθέσιμων αισθητικών δεδομένων.

Προεξοχή ενός σημείου προκύπτει συνήθως από τις αντιθέσεις μεταξύ των στοιχείων και της γειτονιάς τους, όπως μια κόκκινη κουκκίδα που περιβάλλεται από λευκές τελείες, μια ένδειξη μηνύματος ενός τηλεφωνητή που τρεμοπαίζει, ή ένα δυνατό θόρυβο σε ένα κατά τα άλλα ήσυχο περιβάλλον. Ο συγκεκριμένος μηχανισμός συχνά μελετάται στο πλαίσιο του οπτικού συστήματος, αλλά παρόμοιοι μηχανισμοί λειτουργούν και σε άλλα αισθητήρια συστήματα. Όταν η ανάπτυξη προσοχής οδηγείται από εξέχοντα ερεθίσματα, θεωρείται ότι είναι από κάτω προς τα πάνω, χωρίς μνήμη και αντιδραστική. Η ανάπτυξη της προσοχής μπορεί επίσης να καθοδηγείται από πάνω προς τα κάτω, η οποία εξαρτάται από τη μνήμη, τους προληπτικούς μηχανισμούς, όπως για παράδειγμα κοιτάζοντας μπροστά από κινούμενα αντικείμενα. Οι άνθρωποι και άλλα ζώα έχουν δυσκολία στο να δίνουν προσοχή σε περισσότερα από ένα αντικείμενα ταυτόχρονα, έτσι βρίσκονται αντιμέτωποι με την πρόκληση του να εντάσσουν και να δίνουν συνεχώς προτεραιότητα σε διαφορετικές επιδράσεις του περιβάλλοντος.

Είναι σημαντικό να θυμόμαστε ότι τα προεξέχοντα αντικείμενα θεωρούνται το αποτέλεσμα της αλληλεπίδρασης ενός ερεθίσματος με άλλα ερεθίσματα και με το οπτικό (βιολογικό ή τεχνητό). Ως απλό παράδειγμα, ένας άνθρωπος με αχρωματοψία θα έχει μια εντελώς διαφορετική εμπειρία της οπτικής αυτής ικανότητας, από ένα άτομο με φυσιολογική αντίληψη των χρωμάτων, ακόμα και αν κοιτούν ακριβώς την ίδια φυσική σκηνή (Εικόνα 6).



Εικόνα 6: Παραδείγματα αντικειμένων που προεξέχουν από τα υπόλοιπα

Η ουσία της περίοπτης θέσης των αντικειμένων βρίσκεται στην ενίσχυση της νευρικής και αντιληπτικής αναπαράστασης των χώρων στους οποίους τα τοπικά οπτικά χαρακτηριστικά διαφέρουν σημαντικά από τα γενικά που υπάρχουν γύρω από την εικόνα με κάποιο συναφή τρόπο.

Τα προεξέχοντα αντικείμενα έχουν γίνει θέμα του πρόσφατου τεχνολογικού ενδιαφέροντος. Στην πραγματικότητα, μέχρι πρόσφατα οι περισσότεροι αλγόριθμοι υπολογιστικής όρασης είχαν στηριχθεί στην ωμή βία (brute force), συστηματική σάρωση των εικόνων από αριστερά προς τα δεξιά και από πάνω προς τα κάτω στην προσπάθεια εντοπισμού αντικείμενα του ενδιαφέροντος. Το συγκεκριμένο επιστημονικό πεδίο παρέχει ένα σχετικά φθινό και γρήγορο μηχανισμό για την επιλογή μερικών υποψηφίων στοιχείων και εξάλειψη άλλων.

Στο σημείο αυτό θα γίνει αναφορά στις τεχνολογικές εφαρμογές που συναντάμε την ιδιότητα των προεξέχον αντικειμένων.

- Αυτόματος εντοπισμός στόχου, για στρατιωτικά οχήματα.
- Ρομποτική, ορίζοντας τα σημαντικότερα αντικείμενα του περιβάλλοντος για ασφαλή πλοήγηση.
- Συμπύεση εικόνας & βίντεο, δίνοντας υψηλότερη ποιότητα σε σημαντικότερα αντικείμενα.
- Αυτόματη περικοπή, κεντράρισμα της εικόνας για προβολή σε μικρές φορητές οθόνες.
- Εύρεση όγκων σε μαστογραφίες.

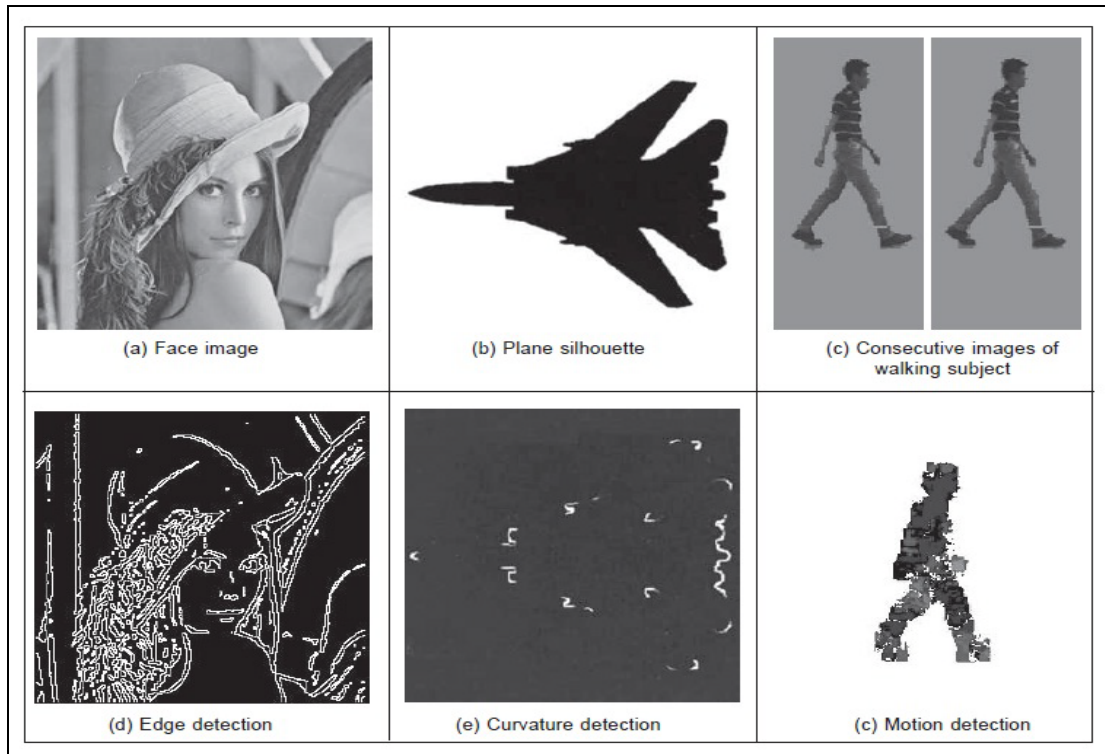
2. Μέθοδοι Εξαγωγής Χαρακτηριστικών Εικόνας

Εξαγωγή χαρακτηριστικών [12] σε μια εικόνα, ονομάζεται η διαδικασία εντοπισμού ενός σημαντικού τμήματος, ποιότητας ή χαρακτηριστικού αυτής της εικόνας. Με άλλα λόγια, εξαγωγή χαρακτηριστικών ορίζεται ως ο εντοπισμός των εικονοστοιχείων (pixels) σε μια εικόνα, τα οποία έχουν κάποια χαρακτηριστικά γνωρίσματα. Συνήθως τέτοιο γνώρισμα θεωρείται κάποια ανομοιογένεια των τοπικών ιδιοτήτων της εικόνας. Για παράδειγμα οι ακμές θεωρούνται ανομοιογένειες της έντασης ή της κλίμακας μιας εικόνας.

Γενικότερα για αυτή τη μέθοδο μπορούμε να πούμε ότι χρησιμοποιείται όταν τα δεδομένα εισόδου σε έναν αλγόριθμο είναι πολύ μεγάλα για να υποβληθούν σε επεξεργασία και υπάρχουν υπόνοιες ότι είναι εμφανώς περιττά, δηλαδή πολλά δεδομένα αλλά όχι αρκετή πληροφορία, τότε τα δεδομένα εισόδου θα πρέπει να μετατραπούν σε μια μειωμένη αναπαράσταση του συνόλου των χαρακτηριστικών, τα οποία ονομάζονται και διανύσματα χαρακτηριστικών. Αυτή λοιπόν η μετατροπή των δεδομένων εισόδου στο σύνολο των χαρακτηριστικών ονομάζεται εξαγωγή χαρακτηριστικών. Εάν τα εξαγόμενα χαρακτηριστικά είναι προσεκτικά επιλεγμένα, αναμένεται ότι το σύνολο τους θα εξάγει τις σχετικές πληροφορίες από τα δεδομένα εισόδου προκειμένου να εκτελεστεί η επιθυμητή εργασία με αυτή τη μειωμένη εκπροσώπηση αντί για το πλήρες μέγεθος των δεδομένων εισόδου.

Η εξαγωγή χαρακτηριστικών είναι ένας γενικός όρος για τις μεθόδους κατασκευής των συνδυασμών των μεταβλητών για να επιλύσουν διάφορα προβλήματα, ενώ εξακολουθούν να περιγράφουν τα δεδομένα με επαρκή ακρίβεια.

Παρακάτω θα συνεχίσουμε με την παρουσίαση των σημαντικότερων μεθόδων που χρησιμοποιούνται για την εξαγωγή τέτοιων χαρακτηριστικών. Αύτη η παρουσίαση έχει σαν στόχο την βήμα προς βήμα παρακολούθηση των συγκεκριμένων ενεργειών οι οποίες χρησιμοποιήθηκαν κατά την υλοποίηση της τελικής εφαρμογής. Επίσης θα γίνουν κατανοητές ομοιότητες αλλά και διαφορές μεταξύ των σημαντικότερων τέτοιων μεθόδων που έχουν επικρατήσει στην επιστημονική κοινότητα.



Εικόνα 7: Ανίχνευση Χαρακτηριστικών Χαμηλού Επιπέδου

2.1 Μετασχηματισμός κλιμακωτά – αναλλοίωτων χαρακτηριστικών (SIFT)

Ο μετασχηματισμός κλιμακωτά αναλλοίωτων χαρακτηριστικών [13] [14], είναι ένας αλγόριθμος της τεχνητής όρασης για τον εντοπισμό και την περιγραφή τοπικών χαρακτηριστικών σε εικόνες. Ο αλγόριθμος δόθηκε στην δημοσιότητα το 1999 από τον David Lowe και έχει κατοχυρωμένο δίπλωμα ευρεσιτεχνίας στις ΗΠΑ υπό την ιδιοκτησία του Πανεπιστημίου της Βρετανικής Κολούμπια, παρέχεται όμως για χρήση σε συστήματα ερευνητικού περιεχομένου. Οι εφαρμογές που έχει χρησιμοποιηθεί είναι αναγνώριση αντικείμενων, παραγωγή χαρτών και πλοήγησης ρομπότ, “ράνιμο” εικόνων, παραγωγή τρισδιάστατων μοντέλων, αναγνώριση χειρονομιών, ανίχνευση αντικείμενων σε βίντεο και χρήση εφέ σε βίντεο

2.1.1 Επισκόπηση της μεθόδου

Για κάθε αντικείμενο στην εικόνα, μπορούν να εξαχθούν ενδιαφέροντα σημεία σχετικά με το αντικείμενο, για να παρέχουν τη λεγόμενη “περιγραφή χαρακτηριστικού” (feature description). Αυτή η περιγραφή η οποία προέρχεται από μια εικόνα εκπαίδευσης, μπορεί στη συνέχεια να χρησιμοποιηθεί για τον εντοπισμό του αντικειμένου, σε μια δοκιμαστική εικόνα η οποία περιέχει πολλά άλλα αντικείμενα. Για να εκτελεστεί μια αξιόπιστη αναγνώριση, είναι σημαντικό τα χαρακτηριστικά που εξάγονται από την εικόνα εκπαίδευσης να ανιχνεύονται ακόμα και μετά από αλλαγές στην κλίμακα της εικόνας, του θορύβου και του φωτισμού. Αυτά τα σημεία βρίσκονται συνήθως σε περιοχές με μεγάλες αντιθέσεις μέσα στην εικόνα, τις λεγόμενες ακμές. Μια άλλη σημαντική ιδιότητα αυτών των χαρακτηριστικών είναι ότι οι σχετικές θέσεις μεταξύ τους στην αρχική σκηνή δεν πρέπει να αλλάζουν από τη μια εικόνα στην άλλη. Ωστόσο στην πράξη ο αλγόριθμος εντοπίζει και χρησιμοποιεί ένα πολύ

μεγαλύτερο αριθμό χαρακτηριστικών από τις εικόνες, μειώνοντας έτσι τη συμβολή των σφαλμάτων που προκαλούνται από τις τοπικές διακυμάνσεις του μέσου σφάλματος από όλα τα λάθη ταιριάσματος στα χαρακτηριστικά.

Η πατενταρισμένη μέθοδος του Lowe μπορεί να προσδιορίσει με αυστηρότητα, αντικείμενα τα οποία βρίσκονται ακόμη και ανακατεμένα ή υπό την επίδραση μερικών αλλαγών. Αυτό συμβαίνει διότι οι περιγραφείς που εξάγει ο αλγόριθμος παραμένουν αναλλοίωτοι σε ενιαία κλίμακα, προσανατολισμό και εν μέρη αμετάβλητοι σε συσχετισμένες στρεβλώσεις και αλλαγές φωτισμού.

2.1.2 Η Μέθοδος του David Lowe

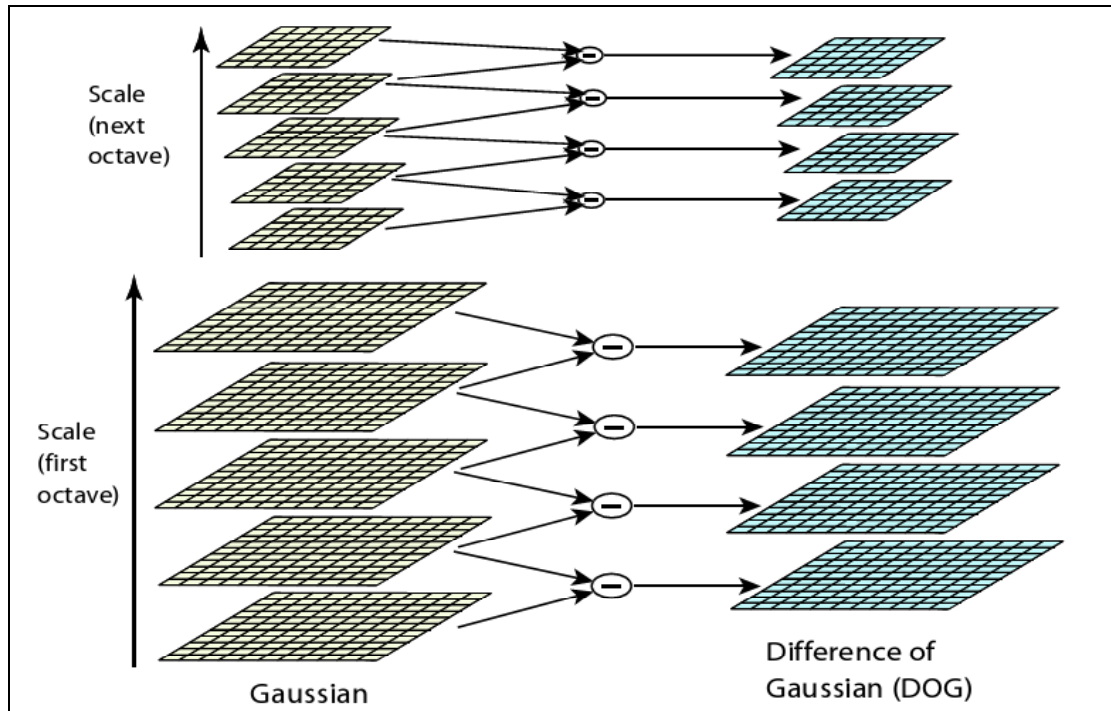
Τα κύρια σημεία των αντικειμένων αρχικά εξάγονται από ένα σύνολο εικόνων αναφοράς και αποθηκεύονται σε μια βάση δεδομένων. Ένα αντικείμενο αναγνωρίζεται σε μια καινούργια εικόνα συγκρίνοντας ατομικά κάθε στοιχείο από την νέα εικόνα με αυτό από τη βάση δεδομένων, βρίσκοντας υποψήφια χαρακτηριστικά τα οποία ταιριάζουν βασισμένα στην Ευκλείδεια απόσταση από τα διανύσματα των χαρακτηριστικών της εικόνας. Στη συνέχεια από το σύνολο των στοιχείων που ταιριάζουν, τα υποσύνολα των κύριων σημείων που συμφωνούν με το αντικείμενο και τη θέση του, την κλίμακα και τον προσανατολισμό στη νέα εικόνα, εντοπίζονται και φιλτράρονται ως καλά ορισμένα ταιριάσματα. Ο προσδιορισμός των σταθερών συμπλεγμάτων γίνεται γρήγορα μια αποτελεσματική εφαρμογή του πίνακα κατακερματισμού (hash table) του γενικευμένου μετασχηματισμού Hough. Κάθε σύμπλεγμα των τριών ή περισσότερων χαρακτηριστικών τα οποία συμφωνούν σε ένα αντικείμενο στη συνέχεια υπόκεινται σε περαιτέρω λεπτομερή έλεγχο και στη συνέχεια οι ακραίες τιμές απορρίπτονται.

Τέλος η πιθανότητα ένα συγκεκριμένο σύνολο χαρακτηριστικών να υποδηλώνει την παρουσία ενός αντικειμένου, υπολογίζεται δεδομένης της ακρίβειας του ταιριάσματος και τον αριθμό των πιθανών εσφαλμένων ταιριασμάτων. Ταιριάσματα μεταξύ των αντικειμένων που έχουν περάσει όλα τα παραπάνω στάδια, μπορούν να αναγνωριστούν ως ορθά με μεγάλη αξιοπιστία.

2.1.3 Βασικά στάδια

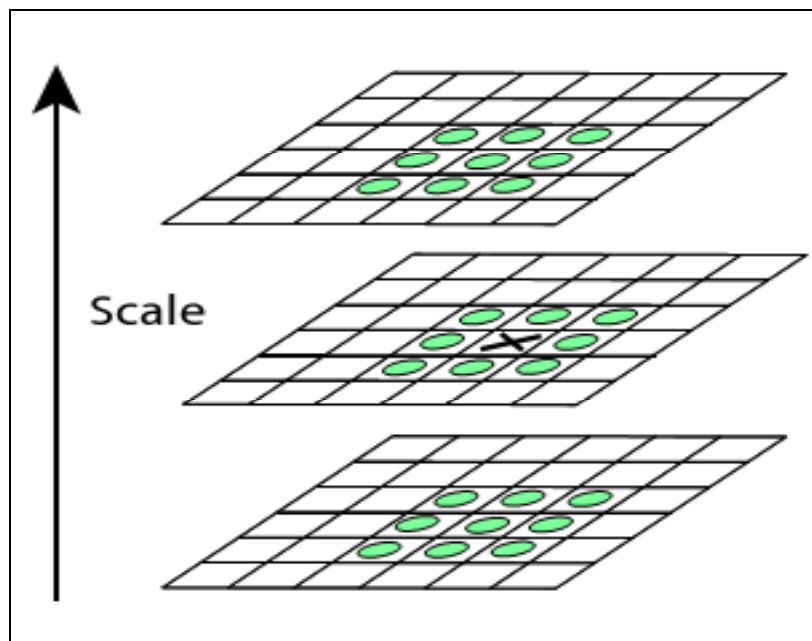
ο Ανίχνευση αμετάβλητων σε κλίμακα χαρακτηριστικών: Η μέθοδος του Lowe για την παραγωγή χαρακτηριστικών εικόνας μετατρέπει την εικόνα σε μια μεγάλη συλλογή από διανύσματα χαρακτηριστικών γνωρισμάτων κάθε ένα από τα οποία είναι αναλλοίωτα στην μετατόπιση, στην κλιμάκωση, στην περιστροφή, εν μέρει αμετάβλητα στις αλλαγές φωτισμού και ανθεκτικά στις τοπικές γεωμετρικές παραμορφώσεις. Τα χαρακτηριστικά αυτά έχουν παρόμοιες ιδιότητες με τους νευρώνες στο κάτω κροταφικό φλοιό που χρησιμοποιείται για την αναγνώριση αντικειμένων στην πρωτεύουσα όραση. Οι βασικές περιοχές ορίζονται ως μέγιστες και ελάχιστες της Gaussian συνάρτησης που εφαρμόζονται στην κλίμακα χώρου σε μια σειρά από εξομαλυμένες και επαναδειγματοληφθέντες εικόνες. Υποψήφια σημεία χαμηλής αντίθεσης και σημεία απόκρισης από άκρη σε άκρη απορρίπτονται. Οι κυρίαρχες κατευθύνσεις ανατίθενται σε εντοπισμένα κύρια σημεία (keypoints). Αυτά τα βήματα διασφαλίζουν ότι τα κύρια σημεία είναι πιο σταθερά για αντιστοίχιση και αναγνώριση. Οι SIFT περιγραφείς ανθεκτικοί στις τοπικά συσχετισμένες στρεβλώσεις προκύπτουν λαμβάνοντας υπόψη τα pixels γύρω από την ακτίνα μιας νευραλγικής θέσης, θολώνοντας και παίρνοντας δείγματα των τοπικών επιπέδων προσανατολισμού της εικόνας.

Η συνάρτηση της Gaussian διαφοράς είναι η εξής: $G(x,y,k\sigma) - G(x,y,\sigma) \approx (k-1) \sigma^2 (L_{xx} + L_{yy})$
Όπου (k-1) θεωρείται μια σταθερά και μπορεί να παραληφθεί.



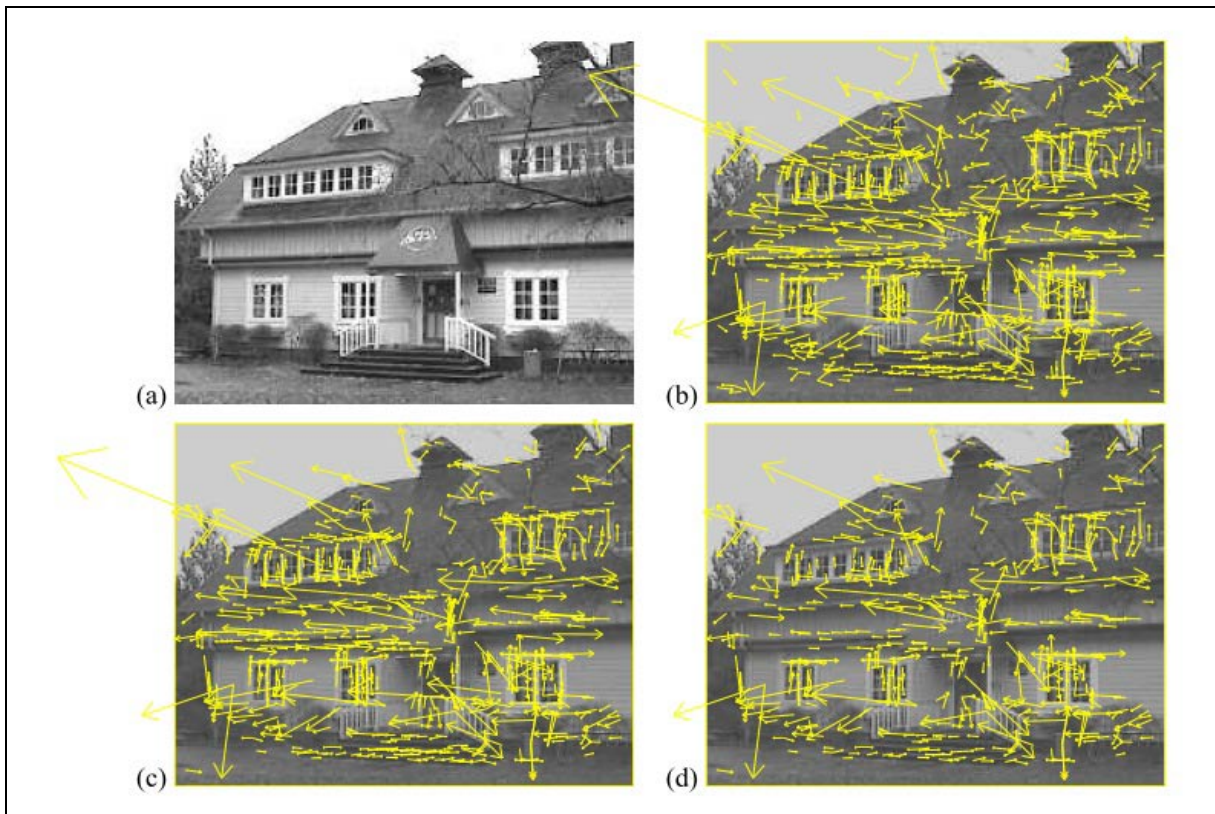
Εικόνα 8: Αναπαράσταση συνέλιξης & αφαιρέσης κάθε οκτάβας για τη δημιουργία της Gaussian διαφοράς

Κάθε pixel στην Gaussian διαφορά συγκρίνεται με τους γείτονες του, στην δικιά του και των διπλανών του pixels κλίμακα. Ένα pixel επιλέγεται μόνο αν είναι πιο σκούρο ή πιο φωτεινό από όλους τους γείτονες του. Δεν είναι όλα τα ακρότατα χρήσιμα στην κλίμακα του χώρου. Ειδικά σημεία με χαμηλή αντίθεση είναι πολύ ευαίσθητα στο θόρυβο και δεν πρέπει να χρησιμοποιούνται. Η Gaussian διαφορά είναι πολύ ευαίσθητη στις ακμές ακόμη και αν η θέση του σημείου κατά μήκος της ακμής είναι ανεπαρκώς εντοπισμένη.



Εικόνα 9: Ο υπολογισμός του μέγιστου και ελαχίστου της Gaussian διαφοράς σε ένα pixel

Η παρακάτω εικόνα απεικονίζει τα στάδια επιλογής των σημείων κλειδιών σε μια εικόνα. Συγκεκριμένα στο παράθυρο (a) Φαίνεται η 233x189 pixel πρωτότυπη εικόνα. (b) Οι αρχικές τοποθεσίες των 832 σημείων κλειδιά, στα μέγιστα και ελάχιστα της Gaussian διαφοράς. Τα σημεία κλειδιά εμφανίζονται ως διανύσματα υποδεικνύοντας την κλίμακα, τον προσανατολισμό και την τοποθεσία. (c) Μετά την εφαρμογή ενός καταωφλίου ελάχιστης αντίθεσης, παραμένουν 729 σημεία κλειδιά. (d) Τα τελικά 536 σημεία κλειδιά τα οποία παραμένουν, αφού έχουν υποστεί την εφαρμογή ενός επιπλέον καταωφλίου που υπολογίζει το λόγο των κύριων καμπυλοτήτων.



Εικόνα 10: Τα αποτελέσματα της επιλογής σημείων κλειδιών σε μια εικόνα

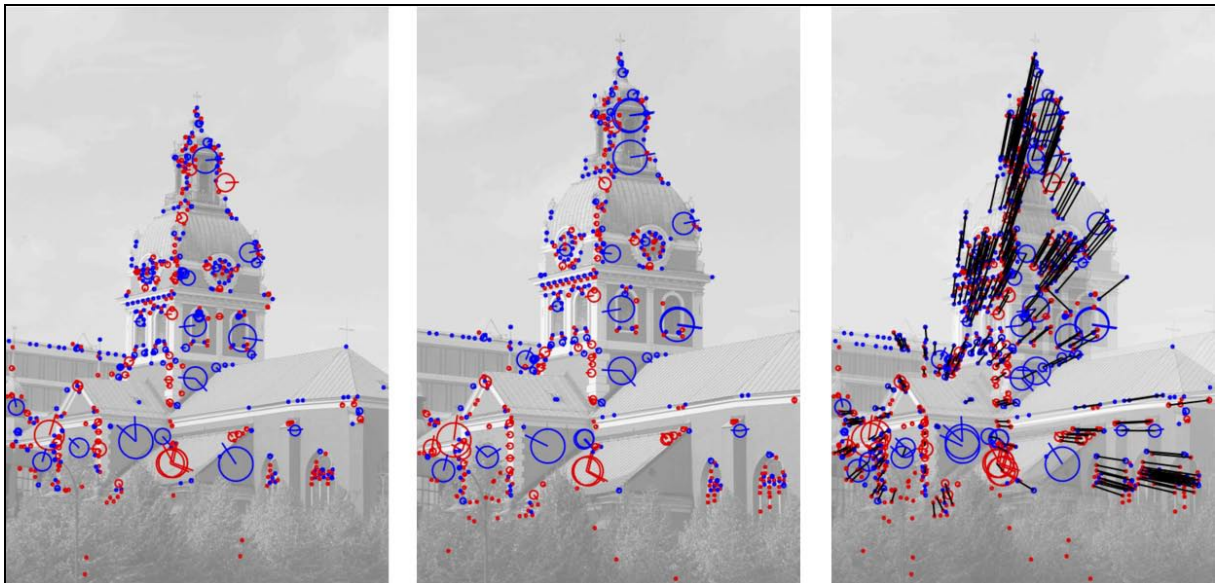
Στην επιστήμη των υπολογιστών ένα δένδρο k-d θεωρείται μια δομή δεδομένων η οποία διαχωρίζει χωρικά και χρησιμοποιείται για την οργάνωση των σημείων σε ένα k-διάστατο χώρο. Τα k-d δένδρα είναι μια ειδική περίπτωση των δυαδικών δένδρων διαχωρισμού του χώρου. Θεωρούνται μια χρήσιμη δομή δεδομένων για διάφορες εφαρμογές όπως έρευνες που αφορούν τον κοντινότερο γείτονα, όπως θα δούμε παρακάτω.

Best bin first είναι ένας αλγόριθμος αναζήτησης ο οποίος έχει σχεδιαστεί για να βρίσκει αποτελεσματικά μια κατά προσέγγιση λύση στο πρόβλημα της αναζήτησης του κοντινότερου γείτονα για χώρους με πολύ μεγάλες διαστάσεις. Ο αλγόριθμος βασίζεται σε μια παραλλαγή του k-d δένδρου αναζήτησης ο οποίος επιτρέπει τη δημιουργία ευρητήριου υψηλότερων χωρικών διαστάσεων.

ο **Αντιστοίχιση χαρακτηριστικών και προσθήκη σε ευρετήριο:** Η προσθήκη στο ευρετήριο αποτελείται από αποθήκευση SIFT χαρακτηριστικών και τον εντοπισμό όσων ταιριάζουν με αυτά της νέας εικόνας. Ο Lowe χρησιμοποίησε μια τροποποιημένη μορφή του k-d δένδρου [15] που ονομάζεται Best-bin-first μέθοδος αναζήτησης [16] η οποία μπορεί να εντοπίσει την πλησιέστερη γειτονιά (nearest neighbor) με υψηλή πιθανότητα, χρησιμοποιώντας μόνο ένα περιορισμένο ποσοστό υπολογιστικής ισχύς.

Η παραπάνω μέθοδος αναζήτησης χρησιμοποιεί μια μορφοποιημένη ταξινόμηση του αλγορίθμου k-d δένδρου έτσι ώστε τα δοχεία αποθήκευσης των χαρακτηριστικών να αναζητούνται βάση της πλησιέστερης θέσης τους από το αρχικό σημείο ενδιαφέροντος. Αυτή η σειρά αναζήτησης απαιτεί τη χρήση ενός σωρού βασισμένο στην ουρά προτεραιότητας για τον αποτελεσματικό καθορισμό της διάταξης για την αναζήτηση. Η καλύτερη αντιστοίχιση για κάθε χαρακτηριστικό βρίσκεται εντοπίζοντας την κοντινότερη γειτονιά του στη βάση δεδομένων των χαρακτηριστικών από τις εικόνες που εκπαιδεύονται. Ως κοντινότερες γειτονιές ορίζονται τα κύρια σημεία με ελάχιστη Ευκλείδεια απόσταση από το δοσμένο διάνυσμα περιγραφέα. Η πιθανότητα ότι μια αντιστοίχιση είναι σωστή μπορεί να καθοριστεί λαμβάνοντας υπόψη την αναλογία της απόστασης από το πλησιέστερο γείτονα, με την απόσταση από το δεύτερο κοντινότερο γείτονα. Η υλοποίηση του Lowe απορρίπτει όλες τις πιθανές αντιστοιχίες των οποίων η αναλογία είναι μεγαλύτερη από 0.8. Με αυτό τον τρόπο εξαλείφεται το 90 % των λανθασμένων αντιστοιχιών ενώ παράλληλα απορρίπτεται μονάχα το 5% των σωστών. Για την περαιτέρω βελτίωση της αποτελεσματικότητας του αλγορίθμου, η αναζήτηση κόβεται μετά τον έλεγχο των πρώτων 200 υποψήφιων πλησιέστερων γειτόνων. Έτσι για παράδειγμα, για μια βάση δεδομένων 100.000 σημείων με τη συγκεκριμένη υλοποίηση παρέχεται μια επιτάχυνση, πάνω ακριβώς από την αναζήτηση πλησιέστερου γείτονα, δυο τάξεων μεγέθους, αλλά με αποτελέσματα απώλειας σωστών αντιστοιχίσεων λιγότερο από 5% .

Στην παρακάτω εικόνα φαίνονται τα σημεία ενδιαφέροντος τα οποία ανιχνεύονται από δυο διαφορετικές εικόνες της ίδιας σκηνής, με τα υπολογισμένα ταιριάσματα της εικόνας να ζωγραφίζονται με μαύρες γραμμές μεταξύ των αντίστοιχων σημείων ενδιαφέροντος. Τα μπλε και τα κόκκινα βέλη στο κέντρο των κύκλων, δείχνουν τον εκτιμώμενο προσανατολισμό, ο οποίος έχει αποκτηθεί από τις τοπικές κορυφές του ιστογράμματος προσανατολισμού γύρω από τα σημεία ενδιαφέροντος.

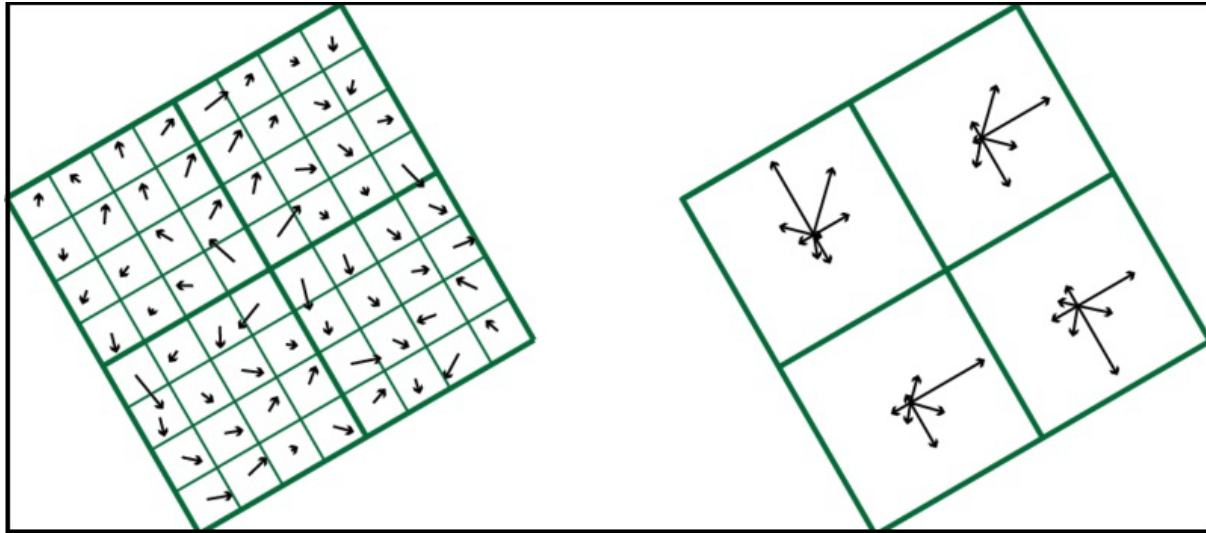


Εικόνα 11: Παράδειγμα αντιστοίχισης SIFT χαρακτηριστικών

ο **Εκχώρηση Προσανατολισμού:** Για την επίτευξη της αμεταβλητότητας στην περιστροφή της εικόνας, σε κάθε σημείο κλειδί ανατίθενται μια ή περισσότερες κατευθύνσεις. Ένα ιστογράμμα προσανατολισμού σχηματίζεται από τους προσανατολισμούς κλίσης των pixels γύρω από ένα σημείο κλειδί. Κάθε δείγμα σταθμίζεται από μέγεθος της κλίσης και ένα Gaussian παράθυρο. Το μέγιστο αυτό του ιστογράμματος και οποιοδήποτε άλλο τοπικό μέγιστο στο πλαίσιο του 80 %, χρησιμοποιείται για την παραγωγή ενός σημείου κλειδί. Για να αυξηθεί η ακρίβεια των τοπικών ιστογραμμάτων χρησιμοποιείται τριγραμμική παρεμβολή για τη διανομή των σταθμισμένων προσαναυξήσεων για τις μετρήσεις του δείγματος εικόνας σε διπλανούς κάδους ιστογράμματος. Με

άλλα λόγια κάθε καταχώρηση στον κάδο πολλαπλασιάζεται με ένα επιπλέον βάρος 1-d, όπου d είναι η απόσταση του δείγματος και της κεντρικής θέσης του κάδου, εκφρασμένο σε μονάδες απόστασης στο ιστόγραμμα.

Στην παρακάτω εικόνα απεικονίζεται το πώς οι SIFT περιγραφείς υπολογίζονται από την δειγματοληψία τιμών του προσανατολισμού της κλίσης και το μέγεθος της, πάνω σε τοπικά προσαρμοσμένο πλέγμα γύρω από κάθε σημείο ενδιαφέροντος. Η εικόνα αυτή δείχνει τον περιγραφέα που υπολογίστηκε σε μια περιοχή 2x2 ενώ συνήθως ο SIFT περιγραφέας υπολογίζεται για περιοχές μεγέθους 4x4.



Εικόνα 12: Παράδειγμα υπολογισμού SIFT περιγραφέων (descriptors) σε περιοχή 2x2

Συνοψίζοντας μπορούμε να πούμε ότι ο αλγόριθμος αυτός ενδείκνυται για αναγνώριση αντικειμένων, εκτελώντας το ταίριασμα ανάμεσα σε κάθε σημείο κλειδί με μια βάση δεδομένων από γνωστά σημεία κλειδιά που έχουν εξαχθεί από διαφορετικές εικόνες. Παρόλα αυτά ο κλασικός τρόπος ταίριασματος με τη χρήση του αλγόριθμου κοντινότερων γειτονιών, μπορεί να παράγει ζεύγη τα οποία δεν είναι σωστά, κάτι το οποίο οφείλεται σε ασαφή χαρακτηριστικά ή σε χαρακτηριστικά που παράγονται από το περιβάλλον της εικόνας και όχι από το κυρίαρχο αντικείμενο της σκηνής. Για το λόγο αυτό για την υλοποίηση της τελικής εφαρμογής για την αυτόματη αναγνώριση αντικειμένων εικόνας, χρησιμοποιήθηκε ο μετασχηματισμός κλιμακωτά αναλλοίωτων χαρακτηριστικών αφού πρώτα έχει γίνει η επεξεργασία των εικόνων με τη μέθοδο εξαγωγής χαρακτηριστικών που αναπτύχθηκε παραπάνω.

2.2 Επιταχυνθέντα ισχυρά χαρακτηριστικά (Speeded up Robust Feature)

Τα επιταχυνθέντα ισχυρά χαρακτηριστικά (SURF) [17] [18] θεωρείται ένας εύρωστος ανιχνευτής τοπικών χαρακτηριστικών, ο οποίος παρουσιάστηκε για πρώτη φορά από τον Herbert Bay το 2006 στο "ECCV 2006" συνέδριο στο Graz της Αυστρίας. Ο συγκεκριμένος ανιχνευτής μπορεί να χρησιμοποιηθεί σε διάφορες εργασίες της τεχνητής όρασης, όπως η αναγνώριση αντικειμένων ή η ανακατασκευή τριών διαστάσεων. Είναι εν μέρει εμπνευσμένος από τον SIFT περιγραφέα, παρόλα αυτά η βασική έκδοση του SURF είναι πολλές φορές γρηγορότερη από τη μέθοδο του SIFT και θεωρείται από τους δημιουργούς του, ότι είναι πιο ανθεκτικός στις διάφορες μεταμορφώσεις της εικόνας από ότι ο SIFT. Η εφαρμογή του αλγορίθμου είναι κατοχυρωμένη με δίπλωμα ευρεσιτεχνίας στις ΗΠΑ. Γενικά μπορούμε να πούμε ότι ο SURF βασίζεται στο ποσό των αποκρίσεων των 2D κυματιδίων Haar και κάνει αποτελεσματική χρήση των ενσωματωμένων εικόνων.

2.2.1 Επισκόπηση της μεθόδου

Οι αντιστοιχίες σημείων ενδιαφέροντος είναι θεμελιώδους σημασίας για πολλές εφαρμογές της τεχνητής όρασης, όπως η βαθμονόμηση κάμερας, καταγραφή εικόνας, αναγνώριση και χαρακτηρισμό αντικειμένων, ανάκτηση εικόνας βάση περιεχομένου. Πολύ συχνά ο χρόνος που απαιτείται για την επεξεργασία είναι κρίσιμο ζήτημα. Αυτό συμβαίνει είτε επειδή υπάρχουν περιορισμοί πραγματικού χρόνου είτε επειδή τεράστιες ποσότητες δεδομένων πρέπει να αντιμετωπιστούν. Ως εκ τούτου αναπτύχθηκε αυτό το σύστημα ανίχνευση σημείων ενδιαφέροντος, το οποίο είναι γρήγορο και αποδοτικό, ταυτόχρονα σεβόμενο την ταυτοποίηση, περιγραφή και αντιστοίχιση των σημείων ενδιαφέροντος. Το σκεπτικό πίσω από την προσέγγιση αυτή, εμπνεύστηκε από την επιτυχία του SIFT, ο οποίος αποδείχθηκε εκπληκτικά εύρωστος κατά των μετασχηματισμών της εικόνας, βάση της οποίας δεν έπρεπε καν να είναι αμετάβλητος. Σε σύγκριση με τις συσχετισμένες περιοχές, η μείωση της πολυπλοκότητας φαίνεται να αντισταθμίζει τη θεωρητική εξειδίκευση. Η αναζήτηση για διακριτά σημεία αντιστοιχίας στην εικόνα μπορεί να διαιρεθεί σε τρία βασικά βήματα.

Πρώτα τα σημεία ενδιαφέροντος επιλέγονται σε διακριτές τοποθεσίες της εικόνας, όπως γωνίες και T-κόμβους. Το σημαντικότερο στοιχείο ενός ανιχνευτή σημείων ενδιαφέροντος, θεωρείται η επαναληψιμότητα. Η επαναληψιμότητα εκφράζει την αξιοπιστία του ανιχνευτή για την εύρεση των ιδίων σημείων ενδιαφέροντος κάτω από διαφορετικές συνθήκες.

Στη συνέχεια, η γειτονιά κάθε σημείου ενδιαφέροντος αντιπροσωπεύεται από ένα διάνυσμα χαρακτηριστικών. Ο περιγραφέας πρέπει να είναι ταυτόχρονα διακριτός και ισχυρός απέναντι στο θόρυβο, στις μετατοπίσεις ανίχνευσης και γεωμετρικές αλλά και φωτομετρικές παραμορφώσεις.

Τέλος τα διανύσματα των περιγραφών δοκιμάζονται αν ταιριάζουν μεταξύ διαφορετικών εικόνων. Η αντιστοίχιση βασίζεται σε μια απόσταση μεταξύ των διανυσμάτων, για παράδειγμα Mahalanobis ή Ευκλείδεια απόσταση. Η διάσταση του περιγραφέα έχει άμεσες επιπτώσεις στον απαιτούμενο χρόνο. Έτσι μικρότερες διαστάσεις είναι επιθυμητές για γρήγορη αντιστοίχιση μεταξύ σημεία ενδιαφέροντος. Ωστόσο η μείωση στις διαστάσεις των διανυσμάτων προσφέρει μικρότερο διακριτό χαρακτήρα από ότι προσφέρουν οι μεγαλύτερες διαστάσεις των διανυσμάτων.

2.2.2 Βασικά στάδια

ο Ανίχνευση σημείων ενδιαφέροντος: Η προσέγγιση του SURF για την ανίχνευση σημείων ενδιαφέροντος χρησιμοποιεί μια πολύ βασική προσέγγιση του Εσσιανού πίνακα (Hessian matrix). Αυτή προσφέρεται για τη χρήση αναπόσπαστων εικόνων, μειώνοντας δραστικά το χρόνο υπολογισμού που απαιτείται. Στο βήμα της ανίχνευσης τα τοπικά μέγιστα του Εσσιανού καθοριστικού παράγοντα, ο οποίος εφαρμόζεται στην κλίμακα χώρου, υπολογίζονται για την επιλογή υποψηφίων σημείων ενδιαφέροντος. Τα εν λόγω υποψήφια σημεία, θα επικυρωθούν αν η απάντηση είναι πάνω από ένα ορισμένο όριο. Τόσο η κλίμακα όσο και η θέση αυτών των υποψηφίων, στη συνέχεια εκκαθαρίζονται χρησιμοποιώντας μια επαναλαμβανόμενη διαδικασία για να ταιριάζει μια τετραγωνική συνάρτηση. Συνήθως λίγες εκατοντάδες από σημεία ενδιαφέροντος ανιχνεύονται σε μια ψηφιακή εικόνα της τάξεως του ενός Mega-Pixel.

Τρία βήματα εμπλέκονται στη διαδικασία ανίχνευσης σημείων ενδιαφέροντος.

- i. *Ανίχνευση χαρακτηριστικών:* Όπως υποστηρίζεται, τα αμετάβλητα σε κλίμακα χαρακτηριστικά μπορούν να ανιχνευτούν χρησιμοποιώντας κλιμακωτά κανονικοποιημένη δεύτερου βαθμού παράγωγο στην χωρική κλίμακα αναπαράστασης της δοσμένης εικόνας. Αυτά τα χαρακτηριστικά αντιστοιχούν σε άκρες ή γωνίες. Σε αντίθεση με τον SIFT, στον οποίο ο Lowe προσεγγίζει τον κανονικοποιημένο Laplace φορέα με τη Gaussian διαφορά, ο SURF προσεγγίζει τον κανονικοποιημένο παράγοντα της Hessian κλίμακας ως εξής:

$$\begin{aligned} \tilde{\text{DoH}}^{(\sigma)}(u) &:= \det(\tilde{H}^{(\sigma)} * u) := \frac{(\tilde{D}_{xx}^{(\sigma)} * u) \times (\tilde{D}_{yy}^{(\sigma)} * u) - (r(\sigma) \tilde{D}_{xy}^{(\sigma)} * u)^2}{l^4} \\ &= l^4 \left\{ \left(\frac{\tilde{D}_{xx}^{(\sigma)}}{l^4} * u \right) \times \left(\frac{\tilde{D}_{yy}^{(\sigma)}}{l^4} * u \right) - \left(r(\sigma) \left(\frac{\tilde{D}_{xy}^{(\sigma)}}{l^4} * u \right) \right)^2 \right\} \end{aligned}$$

Εικόνα 13: Τύπος του παράγοντα DoH

Ακολουθεί ένα παράδειγμα μιας εικόνας στην οποία έχει εφαρμοστεί ο παράγοντας DoH με $l=33$ δηλαδή $\sigma = 13.2$.



Εικόνα 14: Η Lena κατά την εφαρμογή της ακριβής προσέγγισης του DoH παράγοντα

ii. *Επιλογή χαρακτηριστικών* : Ως σημεία ενδιαφέροντος ορίζονται τα σημεία της κλίμακας του χώρου τα οποία αλλοιώνονται στον ίδιο βαθμό κατά τη διάρκεια μιας ενδεχόμενης μετατροπής ομοιότητας και αντιστοιχούν σε τοπικά μέγιστα του εν λόγω παράγοντα DoH, ο οποίος εφαρμόζεται στην κλίμακα αναπαράστασης της εικόνας. Αυτά τα σημεία επιλέγονται εξετάζοντας μια $3 \times 3 \times 3$ γειτονιά, και εκτελώντας μια εξαντλητική σύγκριση του κάθε σημείου της κλίμακας με τους 26 πλησιέστερους γείτονες.

Ο αλγόριθμος, για να αποκτήσει μια συμπαγή αναπαράσταση της εικόνας αλλά και για να αντιμετωπίσει το θόρυβο, επιλέγει τα πλέον εξέχοντα σημεία από το σύνολο των τοπικών μεγίστων. Αυτό επιτυγχάνεται με τη χρήση ενός ορίου (threshold t_H) στην απόκριση του DoH παράγοντα, για κάθε ένα σημείο ενδιαφέροντος (x,y) σε κλίμακα σ .

$$|\tilde{\text{DoH}}^{(\sigma)}(u)(x,y)| > t_H$$

Εικόνα 15: Το όριο για την επιλογή χαρακτηριστικών στον SURF

Παρακάτω φαίνεται το σύνολο το σημείων ενδιαφέροντος, τα οποία εντοπίζονται ως μέγιστα στην τοπική κλίμακα χώρου του DoH παράγοντα και επιλέγονται μετά την εφαρμογή του ορίου. Οι ακτίνες των κύκλων αντιστοιχούν στην κλίμακα των σημείων ενδιαφέροντος.



Εικόνα 16: Επιλογή χαρακτηριστικών SURF συνολικά 757 σημεία ενδιαφέροντος

iii. *Βελτίωση θέσης στην κλίμακα του χώρου:* Για κάθε τοπικό μέγιστο του παράγοντα DoH, ο εντοπισμός του αντιστοιχίου σημείου ενδιαφέροντος M , με συντεταγμένες (x,y,σ) στην κλίμακα-χώρο, μπορεί να τελειοποιηθεί χρησιμοποιώντας μια παρεμβολή δεύτερης σειράς. Συγκεκριμένα βασίζεται σε μια τοποθέτηση σε κλίμακα χώρου με τον εξής τύπο:

$$M' = M + \delta \text{ where } \delta = \begin{pmatrix} \delta_x \\ \delta_y \\ \delta_\sigma \end{pmatrix} = -\mathcal{H}^{-1} \left(D\tilde{o}H^{(\sigma)}(u)(x,y) \right) \times \nabla \left(D\tilde{o}H^{(\sigma)}(u)(x,y) \right)$$

Μπορεί να συμβεί το σημείο M' να μην ανήκει στη γειτονιά του M . Για να αποφευχθεί αυτό το πρόβλημα, ακολουθώντας και τον SIFT, το σημείο M ενημερώνεται ως ο κοντινότερος γείτονας του M' στην $3 \times 3 \times 3$ γειτονιά και η διαδικασία βελτίωσης επαναλαμβάνεται με τον παραπάνω τύπο. Εάν συνολικά αυτή η διαδικασία επαναληφθεί περισσότερες από 5 φορές, το σημείο ενδιαφέροντος απορρίπτεται. Τελικά για κάθε σημείο ενδιαφέροντος, πέρα από τις συντεταγμένες του (x,y) και σ , υποθηκεύεται επίσης και το Laplacian πρόσημο για το στάδιο της αντιστοίχισης, προκειμένου να επιταχυνθεί η σύγκριση του SURF.

Στην εικόνα που ακολουθεί φαίνεται η επιλογή των χαρακτηριστικών αφού έχει γίνει η βελτίωση που αναφέρθηκε παραπάνω.



Εικόνα 17: Επιλογή χαρακτηριστικών SURF, με βελτίωση, συνολικά 656 σημεία ενδιαφέροντος

ο Τοπικοί Περιγραφείς: Από την προηγούμενη φάση, λαμβάνουμε μια σειρά από N σημεία ενδιαφέροντος στην κλίμακα-χώρου, τα οποία επιτρέπουν την κωδικοποίηση αμετάβλητων σε κλίμακα χαρακτηριστικών από μια εικόνα. Παρ' όλα αυτά, πριν από την κατασκευή τέτοιων τοπικών περιγραφέων, πρέπει πρώτα να καθοριστεί για κάθε σημείο ενδιαφέροντος, ο κυρίαρχος προσανατολισμός, ακολουθώντας τη διαδικασία που περιγράφεται παρακάτω, προκειμένου να επιτευχθεί η πλήρης αμεταβλητότητα κατά την περιστροφή.

- i. *Προσανατολισμός του σημείου ενδιαφέροντος:* Όμοια με τον SIFT, ο τοπικός προσανατολισμός ενός χαρακτηριστικού σε κλίμακα-χώρο υπολογίζεται από την τοπική εξάπλωση του προσανατολισμού της κλίσης. Ως εκ τούτου για κάθε σημείο ενδιαφέροντος M_i , θεωρούμε ότι η γειτονιά $B_{6\sigma_i}(x_i, y_i)$, ορίζεται ως ο δίσκος ακτίνας $6\sigma_i$ με κέντρο (x_i, y_i) . Ο υπολογισμός της κλίσης σε αυτή την κλίμακα αυτής της γειτονιάς, επιτυγχάνεται με τη χρήση συνέλιξης με φίλτρα παραθύρου. Για να αποφεύγονται οι ανεπιθύμητες ενέργειες, όλα αυτά τα δείγματα κλίσης σταθμίζονται με τη χρήση ενός Gaussian πυρήνα μια τυπικής απόκλισης ίση με $2\sigma_i$ και η οποία εξαρτάται από την Ευκλείδεια απόσταση του δείγματος από το σημείο ενδιαφέροντος.

Σε αντίθεση με την προσέγγιση του SIFT στην οποία ένα ιστόγραμμα φτιάχνεται για να εκτιμηθεί ο κυρίαρχος προσανατολισμός, ο SURF αλγόριθμος υπολογίζει το μέγιστο της ακόλουθης συνάρτησης ανάλογα με τον προσανατολισμό θ .

$$\Psi(\theta) = \begin{pmatrix} \Psi_x(\theta) \\ \Psi_y(\theta) \end{pmatrix} = \sum_{j \in J(\theta)} \phi_j \quad \text{where } J(\theta) := \{j, \angle \phi_j \in [\theta - \frac{\pi}{6}, \theta + \frac{\pi}{6}] \text{ and } (x_j, y_j) \in B_{6\sigma_i}(x_i, y_i)\}$$

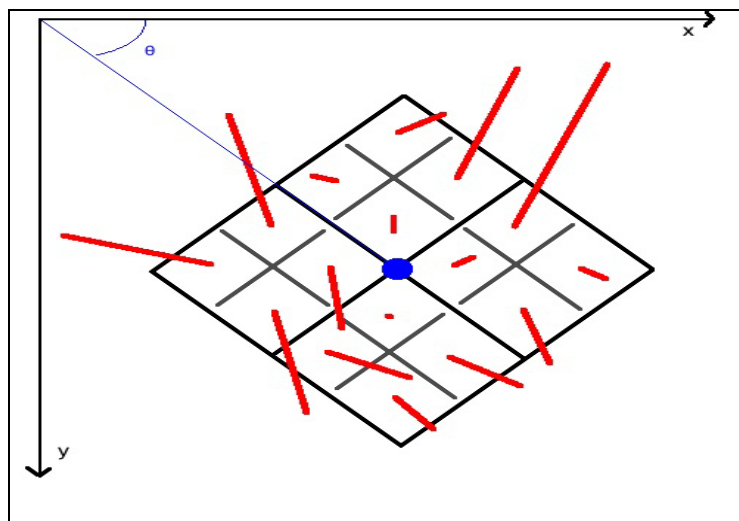
Είναι το άθροισμα των κλίσεων από την εξεταζόμενη περιοχή που έχει περίπου τον προσανατολισμό θ

Για την εξοικονόμηση υπολογιστικού χρόνου, και για να υπάρχουν πάντα οι ίδιοι αριθμοί δειγμάτων, όπως υποστηρίζεται από τον αλγόριθμο, δεν επεξεργάζονται όλα τα pixels στην περιοχή $B_{6\sigma_i}(x_i, y_i)$, αλλά λαμβάνονται υπόψη δείγματα μόνο τα οποία απέχουν σ_i . Ένα παράδειγμα δίνεται στην ακόλουθη εικόνα, στην οποία οι κυρίαρχες κατευθύνσεις εκπροσωπούνται από ένα τμήμα.



Εικόνα 18: Μερικά εντοπισμένα SURF χαρακτηριστικά μαζί με τις κυρίαρχες κατευθύνσεις τους

- ii. *Κατασκευή τοπικών περιγραφέων:* Ένας SURF περιγραφέας είναι ένα διάνυσμα 16×4 το οποίο αντιπροσωπεύει την κανονικοποιημένη στατιστική κλίση (μέση τιμή και απόλυτη τιμή) που εξάγεται από ένα χωροταξικό πλέγμα το οποίο χωρίζεται σε 4×4 κελιά. Για ένα δοσμένο σημείο ενδιαφέροντος $(x_i, y_i, \sigma_i, \theta_i)$, όπως φαίνεται παρακάτω, το αντίστοιχο τετραγωνικό πλέγμα είναι κεντραρισμένο στο (x_i, y_i) , και ευθυγραμμίζεται στο θ_i με μέγεθος $20 \sigma_i$.



Εικόνα 19: Προσανατολισμένος SURF περιγραφέας

Ο SURF περιγραφέας προέρχεται από τη συνένωση των 16 υπολογισμένων διανυσμάτων, για κάθε υποπεριοχή και την ομαλοποίηση του αντίστοιχου 64 διαστάσεων διανύσματος. Αυτό συμβαίνει έτσι ώστε η Ευκλείδεια νόρμα να είναι ενιαία καθιστώντας έτσι τον περιγραφέα αμετάβλητο σε συσχετισμένες αλλαγές αντίθεσης.

ο Αντιστοίχιση Χαρακτηριστικών: Από τα προηγούμενα βήματα, ένα ζευγάρι εικόνων προς αντιστοίχιση, αντιπροσωπεύεται από δυο σημεία ενδιαφέροντος μαζί με τους αντίστοιχους SURF περιγραφείς. Το βήμα της αντιστοίχισης εδώ απλώς εκτελείται ως μια εξαντλητική σύγκριση αυτών των διανυσμάτων σε συνδυασμό με μια τεχνική κατωφλίωσης, την λεγόμενη απόσταση αναλογίας του πλησιέστερου γείτονα (Nearest Neighbor Distance Ratio), η οποία προτάθηκε από τον SIFT και είναι λιγότερο εξαρτημένη από τη δυναμική της εικόνας σε σχέση με το όριο ταιριάσματος που προτείνει ο SURF.

Άλλωστε για την επιτάχυνση της διαδικασίας αντιστοίχισης, συγκρίνονται το πρόσημο του Laplacian μεταξύ δυο σημείων ενδιαφέροντος. Υπολογίζεται η Ευκλείδεια απόσταση μεταξύ του περιγραφέα ερώτησης της πρώτης εικόνας και κάθε περιγραφέα της δεύτερης εικόνας. Μονάχα οι πιο κοντινοί και οι δεύτεροι πιο κοντινοί γείτονες λαμβάνονται υπόψη. Εάν η αναλογία αυτών των δυο αποστάσεων είναι κάτω από 0.8 η αντιστοίχιση μεταξύ του εικόνας-ερωτήματος και του κοντινότερου υποψήφιου γείτονα θεωρείται έγκυρη. Μια τέτοια κατωφλίωση επιτρέπει την απόρριψη πολλών λανθασμένων αντιστοιχιών.

2.2.3 SIFT εναντίον SURF

Καθώς αυτές οι δυο υλοποιήσεις θεωρούνται οι πλέον γνωστές και επιτυχημένες στην επιστημονική κοινότητα, θεωρήθηκε σκόπιμο να παρουσιαστούν κάποια αποτελέσματα από την σύγκριση τους. Έτσι στην ενότητα αυτή θα γίνει λόγος για τις διαφορές των δυο αλγορίθμων που αναπτύχθηκαν παραπάνω και μέσω παραδειγμάτων θα προσπαθήσουμε να καταδείξουμε τα πλεονεκτήματα του καθενός σε διαφορετικά προβλήματα.

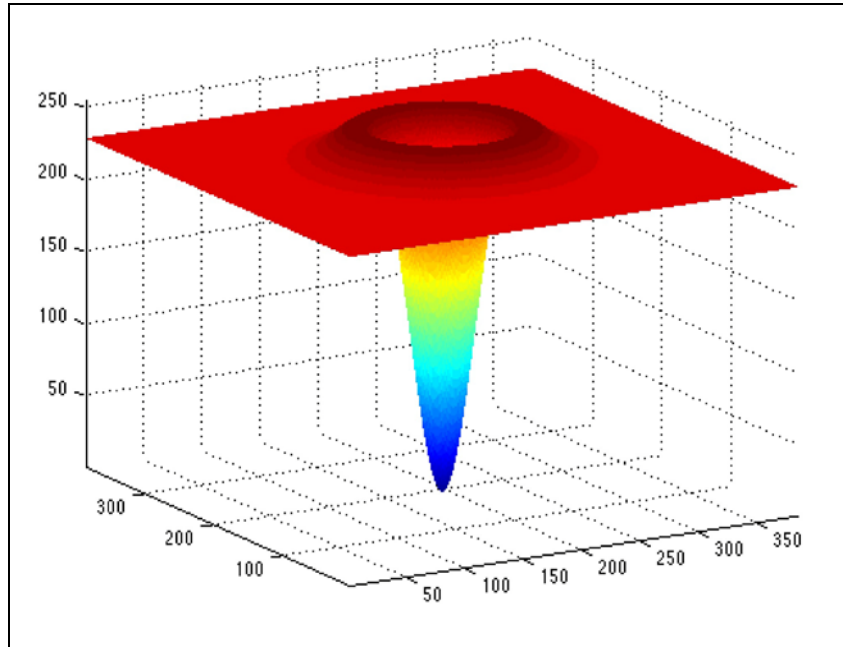
Οι δυο αυτοί αλγόριθμοι χρησιμοποιούν ελαφρώς διαφορετικούς τρόπους για την ανίχνευση χαρακτηριστικών. Ο SIFT κατασκευάζει μια πυραμίδα εικόνας, φιλτράροντας κάθε στρώμα με αυξανόμενες σίγμα τιμές και παίρνει τη διαφορά. Αντίθετα, ο SURF δημιουργεί μια στοίβα με 2:1 μείωση της δειγματοληψίας για τα υψηλότερα επίπεδα της δειγματοληψίας, με αποτέλεσμα εικόνες ίδιας ανάλυσης. Λόγω της χρήσης των αναπόσπαστων εικόνων, ο SURF φιλτράρει τη στοίβα χρησιμοποιώντας ένα πλαίσιο προσέγγισης της δεύτερης τάξης του Gaussian μερικού παράγοντα, δεδομένου ότι οι αναπόσπαστες εικόνες επιτρέπουν τον υπολογισμό των ορθογώνιων φίλτρων πλαισίου σε συνεχή χρόνο.

Ο SURF βασίζεται στις έννοιες του SIFT αλλά εισάγει πιο ριζοσπαστικές προσεγγίσεις προκειμένου να επιταχύνει τη διαδικασία ανίχνευσης. Λόγω της χρήσης των αναπόσπαστων εικόνων, η πολυπλοκότητα του SURF μειώνεται σημαντικά, όμως συχνά επιτυγχάνει μεγαλύτερη απόδοση από τον προκάτοχο του. Αντί του Laplacian φορέα, ο SURF χρησιμοποιεί τον Hessian προσδιοριστικό παράγοντα για την ανίχνευση χαρακτηριστικών στην κλίμακα του χώρου.

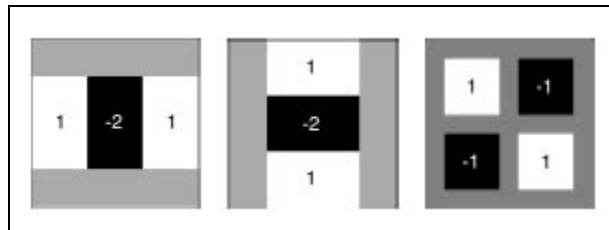
Ο SIFT περιγραφέας κατασκευάζεται από μια τετράγωνη γειτονιά μήκους 12 pixels, όπου είναι η κλίμακα του χαρακτηριστικού. Αυτή η γειτονιά είναι ευθυγραμμισμένη με την κυρίαρχη τοπική κατεύθυνση κλίσης. Για κάθε μια από τις 16 υπό-περιοχές δημιουργείται ένα σταθμισμένο ιστόγραμμα κλίσεων. Λόγω της προσαρμοστικής σε κλίμακα και περιστροφή δημιουργικής διαδικασίας, οι SIFT περιγραφείς είναι ως επί το πλείστον αμετάβλητες σε μέτριους γεωμετρικούς μετασχηματισμούς.

Ο SURF χρησιμοποιεί μια διάταξη περιγραφέα, η οποία επίσης βασίζεται σε μια τετράγωνη περιοχή γύρω από το χαρακτηριστικό σημείο, το οποίο είναι ευθυγραμμισμένο με την κυρίαρχη κλίση και υποδιαιρείται σε 16 υπό-περιοχές.

Οι μοναδικές διαφορές είναι ότι η γειτονιά που επιλέγεται έχει πλάτος 20 pixels και ότι κάθε υπό-περιοχή συμβάλλει μόνο με 4 εγγραφές περιγραφών. Αντί του ιστογράμματος τιμών, το άθροισμα και το απόλυτο ποσό της κλίσης των x και y συστατικών χρησιμοποιείται. Συνολικά ο SURF περιγραφέας περιλαμβάνει ως εκ τούτου 64 συμμετοχές.



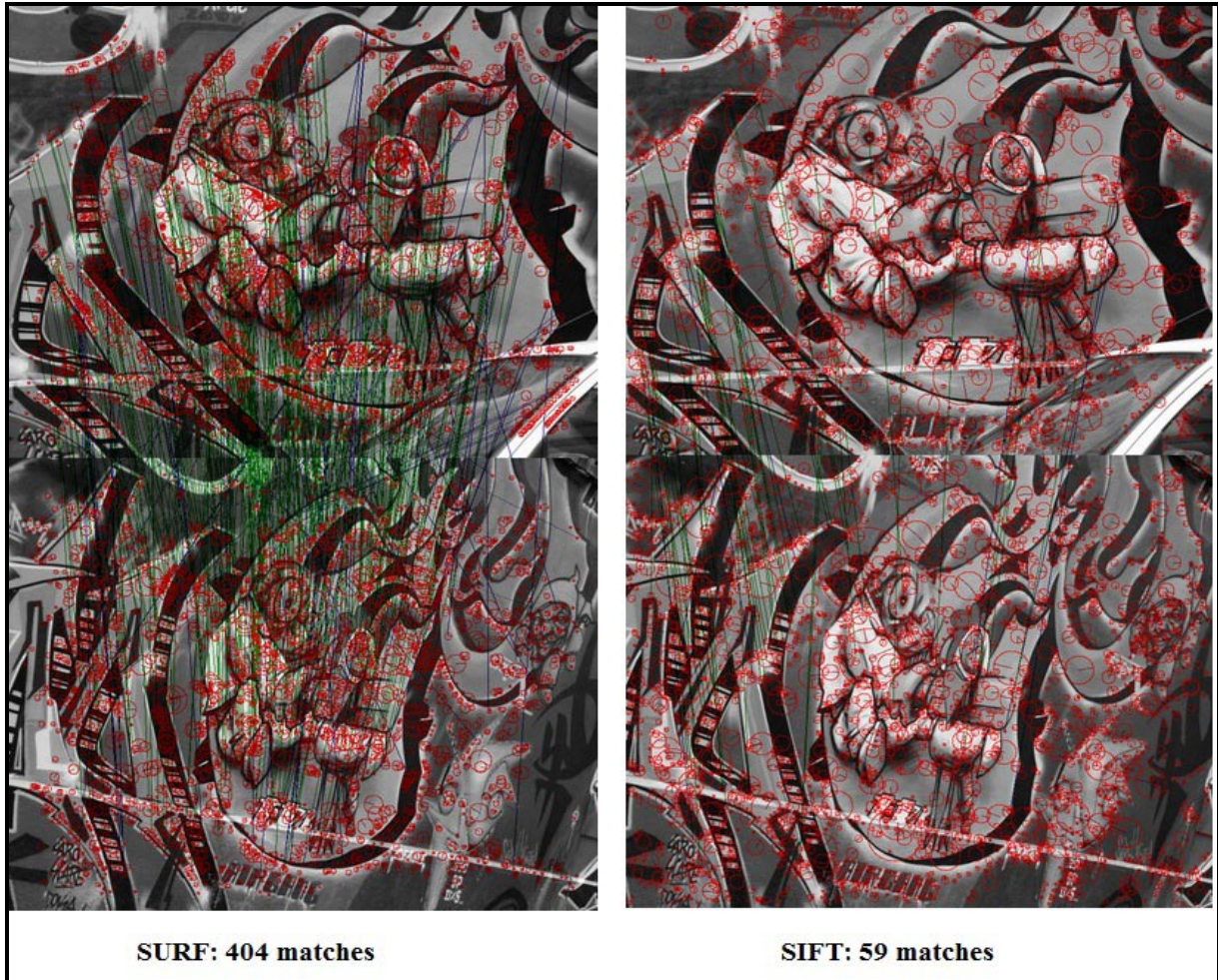
Εικόνα 20: Διαφορά του Gaussian φίλτρου (DoG) το οποίο χρησιμοποιείται από τον SIFT



Εικόνα 21: Φίλτρα παραθύρου SURF για την προσέγγιση δεύτερης τάξης Gaussian παραγώγων

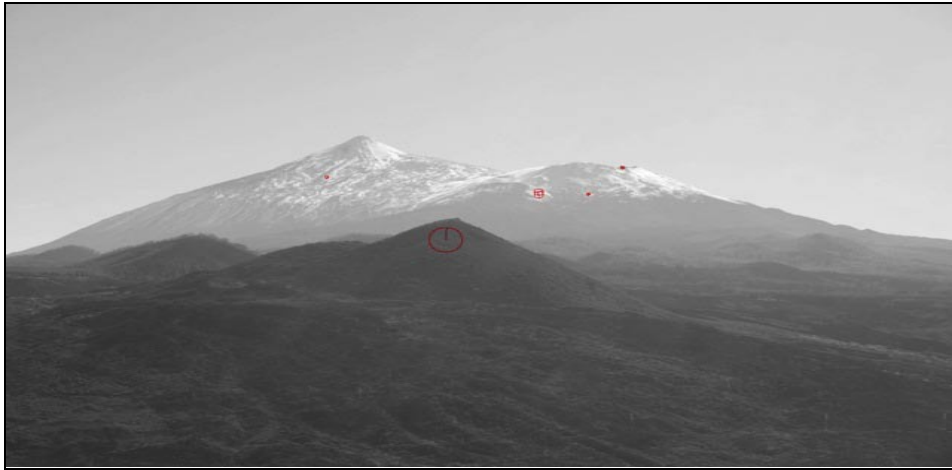
Στις εικόνες που ακολουθούν [18], φαίνονται κάποια παραδείγματα χρήσης των δυο αλγορίθμων και οι διαφορές στις αντιστοιχίες τις οποίες ανιχνεύουν τελικά. Στην Εικόνα 22 ο αλγόριθμος SURF υπερτερεί έναντι του SIFT. Στην Εικόνα 23 φαίνεται μια αποτυχημένη περίπτωση χρήσης του SURF αλγορίθμου.

Συγκεκριμένα στην παρακάτω εικόνα, ο SURF αντιστοιχεί 404 σημεία από το σύνολο των 3026 (για την πάνω αριστερή εικόνα) και 3124 (κάτω αριστερή εικόνα) χαρακτηριστικών. Αντίθετα ο SIFT αντιστοιχεί 59 σημεία από το σύνολο των 2675 (για την πάνω δεξιά εικόνα) και 3179 (κάτω δεξιά εικόνα) χαρακτηριστικών. Τελικά ο SURF δουλεύει αποδοτικότερα και γρηγορότερα σε σχέση με τον SIFT, για το συγκεκριμένο ζεύγος εικόνων.



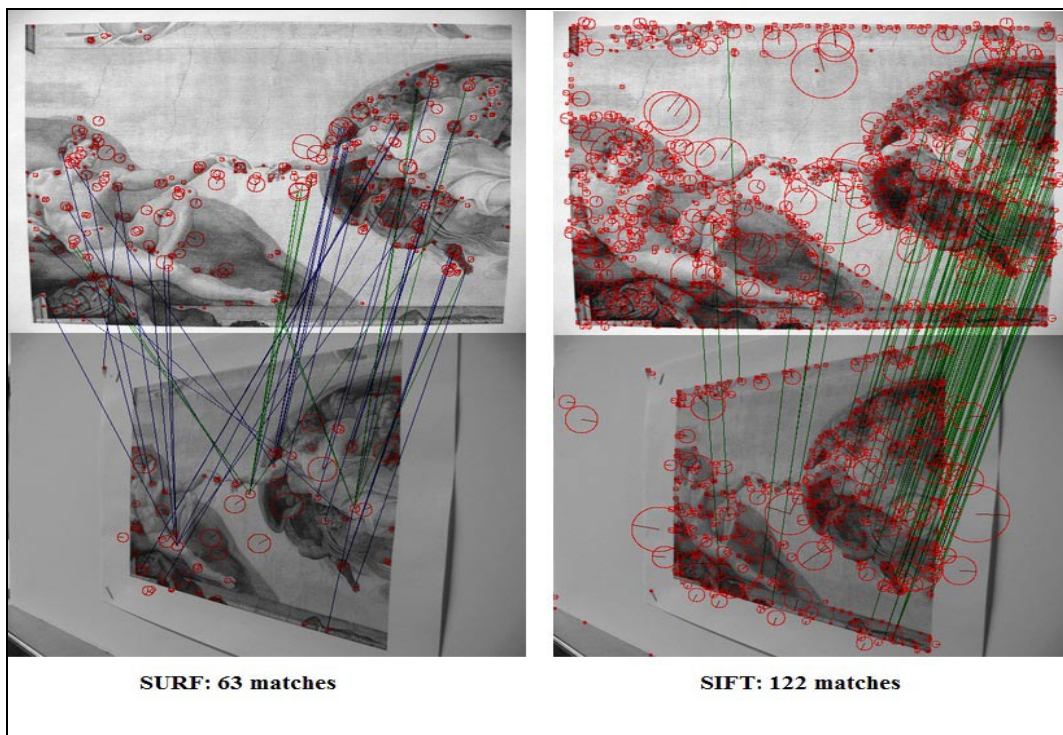
Εικόνα 22: Παράδειγμα αντιστοιχιών σε ίδιο ζευγάρι εικόνων με SURF & SIFT

Στην παρακάτω εικόνα φαίνεται μια αποτυχημένη απόπειρα ανίχνευσης χαρακτηριστικών σε μια εικόνα από τον SURF. Αυτό συμβαίνει διότι πολύ λίγα σημεία ενδιαφέροντος ανιχνεύονται σε μικρής αντίθεσης περιοχές, στην πραγματικότητα ανιχνεύονται πολύ λιγότερα από ότι με τη χρήση του SIFT αλγορίθμου.



Εικόνα 23: Αποτυχημένη προσπάθεια ανίχνευσης χαρακτηριστικών με SURF

Στην παρακάτω εικόνα παρουσιάζεται το πρόβλημα της προοπτικής παραμόρφωσης σε μια εικόνα. Όπως αναφέρθηκε και στην περιγραφή, ο SURF δεν είναι αμετάβλητος στην επίδραση της προοπτικής σε μια εικόνα, όπως για παράδειγμα η ομογραφία). Πρακτικά μπορούμε να πούμε ότι είναι λιγότερο εύρωστος σε τέτοια παραμόρφωση από τον SIFT, όπως φαίνεται και παρακάτω.



Εικόνα 24: Το πρόβλημα της προοπτικής παραμόρφωσης με SURF & SIFT

2.3 Συσχετισμένος Μετασχηματισμός κλιμακωτά – αναλλοίωτων χαρακτηριστικών (ASIFT)

Θα αναφερθούμε σε μια πλήρως αμετάβλητη στις συσχετίσεις μέθοδο σύγκρισης εικόνας, τον λεγόμενο συσχετισμένο μετασχηματισμό κλιμακωτά αναλλοίωτων χαρακτηριστικών (ASIFT) [19]. Ενώ ο SIFT αλγόριθμος, που εξετάστηκε παραπάνω, είναι πλήρως αμετάβλητος σε σχέση μόνο με τέσσερις παραμέτρους και συγκεκριμένα το ζουμ, περιστροφή, και μετακίνηση, η μέθοδος που εξετάζεται εδώ χειρίζεται την υπολειπόμενη παράμετρο, δηλαδή τις γωνίες που καθορίζουν τον προσανατολισμό του άξονα της κάμερας.

Ενάντια σε κάθε πρόγνωση, η προσομοίωση όλων των προβολών βασισμένη σε αυτές τις παραμέτρους είναι εφικτή. Η μέθοδος επιτρέπει τον προσδιορισμό αξιόπιστων χαρακτηριστικών που έχουν υποστεί πολύ μεγάλες στρεβλώσεις, οι οποίες μετρώνται από μια νέα παράμετρο, την λεγόμενη κλίση μετάβασης. Οι πλέον γνωστές μέθοδοι σπανίως υπερβαίνουν κλίσεις μετάβασης του 2 (SIFT), 2.5 (Harris-Affine & Hessian-Affine) και 10 (MSER). Αντίθετα ο ASIFT μπορεί να χειριστεί κλίσεις μεταβάσεων μέχρι 36 και υψηλότερες.

Γίνεται εύκολα αντιληπτό λοιπόν ότι πρόκειται για μια μέθοδο η οποία βασιζόμενη στην εξαιρετική επιτυχία που γνώρισε ο SIFT παλαιότερα, προσπαθεί να εισάγει μια νέα παράμετρο η οποία θα επιτρέπει την καλύτερη λειτουργία του αλγορίθμου κάτω από δύσκολες συνθήκες.

2.3.1 Επισκόπηση της μεθόδου

Εάν ένα φυσικό αντικείμενο διαθέτει ένα ομαλό η τμηματικά ομαλό σύνορο, οι εικόνες που λαμβάνονται, για αυτό το αντικείμενο, από κάμερες σε διαφορετικές θέσεις, υποβάλλονται σε εμφανής ομαλές παραμορφώσεις. Αυτές οι παραμορφώσεις τοπικά προσεγγίζονται αρκετά καλά από συσχετιζόμενους μετασχηματισμούς στο επίπεδο της εικόνας.

Κατά συνέπεια τι πρόβλημα της αναγνώρισης στερεών αντικειμένων έχει καθυστερηθεί από την προσπάθεια υπολογισμού αναλλοίωτα συσχετισμένων τοπικών χαρακτηριστικών σε εικόνες. Τέτοιου είδους χαρακτηριστικά μπορούν να ληφθούν με μεθόδους εξομάλυνσης αλλά προς το παρόν δεν υπάρχουν πλήρως συσχετιζόμενοι μέθοδοι εξομάλυνσης. Ωστόσο με την προσομοίωση της σμίκρυνσης στις δυο εικόνες και με την ομαλοποίηση της περιστροφής, η μέθοδος SIFT καταφέρνει να είναι πλήρως αμετάβλητη σε τέσσερις από τις έξι μεταβλητές ενός συσχετισμένου μετασχηματισμού.

Η μέθοδος του ASIFT προσομοιώνει ένα σύνολο όψεων της αρχικής εικόνας, οι οποίες μπορούν να ληφθούν με τη διαφοροποίηση των δυο καμερών και συγκεκριμένα των δυο παραμέτρων του άξονα προσανατολισμού τους. Δηλαδή αλλάζοντας το γεωγραφικό πλάτος και μήκος των γωνιών, οι οποίες δεν μπορούν να αντιμετωπιστούν με την τυπική μέθοδο SIFT. Στην συνέχεια εφαρμόζεται η μέθοδος SIFT καθαυτή σε όλες τις εικόνες που δημιουργούνται. Έτσι ο ASIFT καλύπτει αποτελεσματικά και τις έξι παραμέτρους του συσχετιζόμενου μετασχηματισμού.

Όπως έχει φανεί από πειράματα, οι περισσότερες σκηνές με αμελητέα αλλαγή στην γωνία θέασης της κάμερας αντιστοιχίζονται σωστά τόσο με τον ASIFT όσο και με τον SIFT, με λιγότερα σημεία αντιστοίχισης βέβαια). Αλλά όταν η αλλαγή της οπτικής γωνίας αλλάζει σε σημαντικό βαθμό ενώ όλες οι άλλες μέθοδοι αποτυγχάνουν, ο ASIFT συνεχίζει να δουλεύει αποτελεσματικά όπως θα διαπιστώσουμε και παρακάτω.

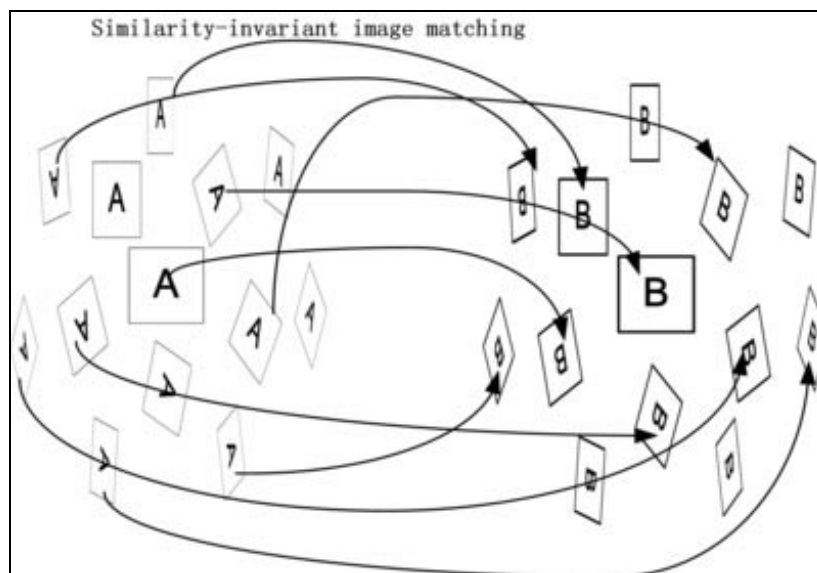
2.3.2 Βασικά στάδια

1. Κάθε εικόνα μετασχηματίζεται προσομοιώνοντας όλες τις πιθανές συσχετισμένες στρεβλώσεις, οι οποίες προκαλούνται από την αλλαγή στον προσανατολισμού του οπτικού άξονα της κάμερας από μια μετωπική θέση. Αυτές οι στρεβλώσεις εξαρτώνται από δυο παραμέτρους το γεωγραφικό μήκος φ και το γεωγραφικό πλάτος θ . Οι εικόνες υποβάλλονται σε περιστροφές γωνίας φ ακολουθημένες από κλίσεις με παράμετρο $t = 1 / |\cos \theta|$. Για ψηφιακές εικόνες, η κλίση πραγματοποιείται από μια κατευθυνόμενη t υπό-δειγματοληψία. Κατά συνέπεια, απαιτείται η προηγούμενη εφαρμογή ενός φίλτρου εξομάλυνσης. Αυτές οι περιστροφές και κλίσεις εκτελούνται για ένα μικρό και πεπερασμένο αριθμό γωνιών γεωγραφικού πλάτους και μήκους. Τα βήματα δειγματοληψίας αυτών των παραμέτρων εξασφαλίζουν ότι οι προσομοιωμένες εικόνες, κρατούνται κοντά σε οποιαδήποτε άλλη δυνατή γωνία θέασης που παράγεται από άλλες τιμές των φ και θ .

2. Όλες οι εικόνες προσομοίωσης συγκρίνονται από τον SIFT αλγόριθμο, ο οποίος μπορεί εύκολα να αντικατασταθεί από οποιοδήποτε άλλη μέθοδο αμετάβλητης αντιστοιχίσης.

3. Η μέθοδος SIFT έχει το δικό της κριτήριο αποβολής εσφαλμένων αντιστοιχιών. Παρόλα αυτά αφήνει κατά κανόνα εσφαλμένες αντιστοιχίες ακόμη και σε ζευγάρια εικόνων που δεν αντιστοιχούν στην ίδια σκηνή. Ο ASIFT συγκρίνοντας πολλά ζευγάρια μπορεί να συσσωρεύσει πολλές λανθασμένες αντιστοιχίες. Είναι σημαντικό λοιπόν να φιλτράρονται τέτοιες αντιστοιχίες. Το κριτήριο που χρησιμοποιείται είναι ότι οι αντιστοιχίες που διατηρούνται πρέπει να είναι συμβατές με επιπολική γεωμετρία. Σε κάποιες περιπτώσεις μπορεί να εμφανιστεί ότι δεν υπάρχουν ASIFT αντιστοιχίες μεταξύ δυο εικόνων, αυτό μπορεί να συμβαίνει διότι οι αντιστοιχίες έχουν εξαλειφθεί λόγω ασυμβατότητας τους με την επιπολική γεωμετρία.

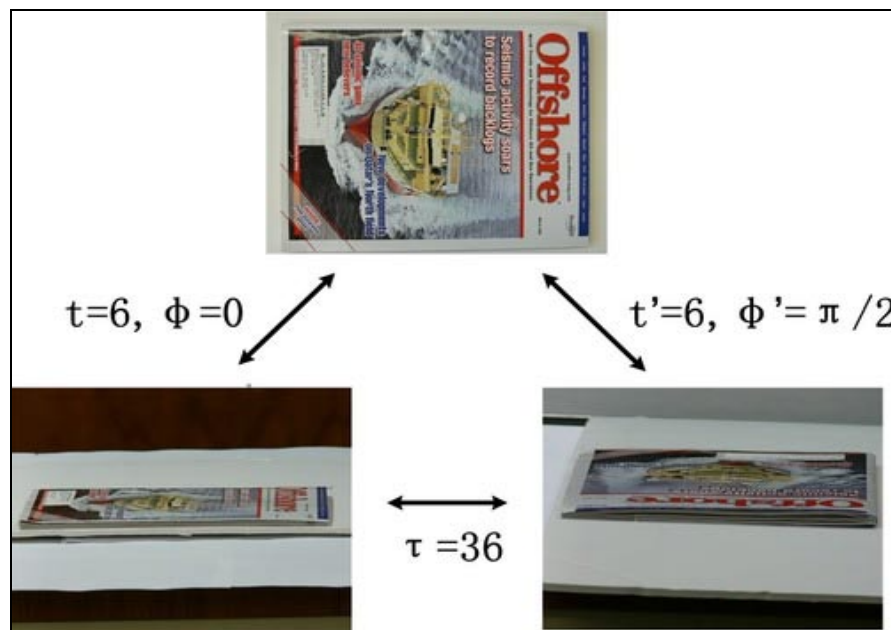
Επιπολική γεωμετρία καλείται η γεωμετρία της στερεοσκοπικής όρασης. Όταν δυο κάμερες βλέπουν μια τρισδιάστατη σκηνή από δυο διαφορετικές θέσεις, υπάρχει μια σειρά από γεωμετρικές σχέσεις μεταξύ των τρισδιάστατων σημείων και των προβολών τους πάνω σε δυσδιάστατες εικόνες, που οδηγούν σε περιορισμούς μεταξύ των σημείων της εικόνας.



Εικόνα 25: Επισκόπηση του ASIFT αλγορίθμου

ο **Κλίσεις μετάβασης του αλγορίθμου**: Η παράμετρος που ονομάζεται κλίση μετάβασης είναι ουσιαστικά η απόλυτη κλίση της πρόσθιας και της λοξής όψης. Σε πραγματικές εφαρμογές οι δυο εικόνες που συγκρίνονται είναι συνήθως σε λοξή προβολή. Η κλίση μετάβασης έχει σχεδιαστεί για να υπολογίζει το ποσό της κλίσης μεταξύ δυο τέτοιων εικόνων. Ωστόσο η κλίση μετάβασης μπορεί να είναι πολύ υψηλότερη από την απόλυτη κλίση. Ως εκ τούτου είναι σημαντικό για αλγορίθμους αντιστοιχίας εικόνων να είναι αμετάβλητοι σε υψηλές κλίσεις μετατόπισης.

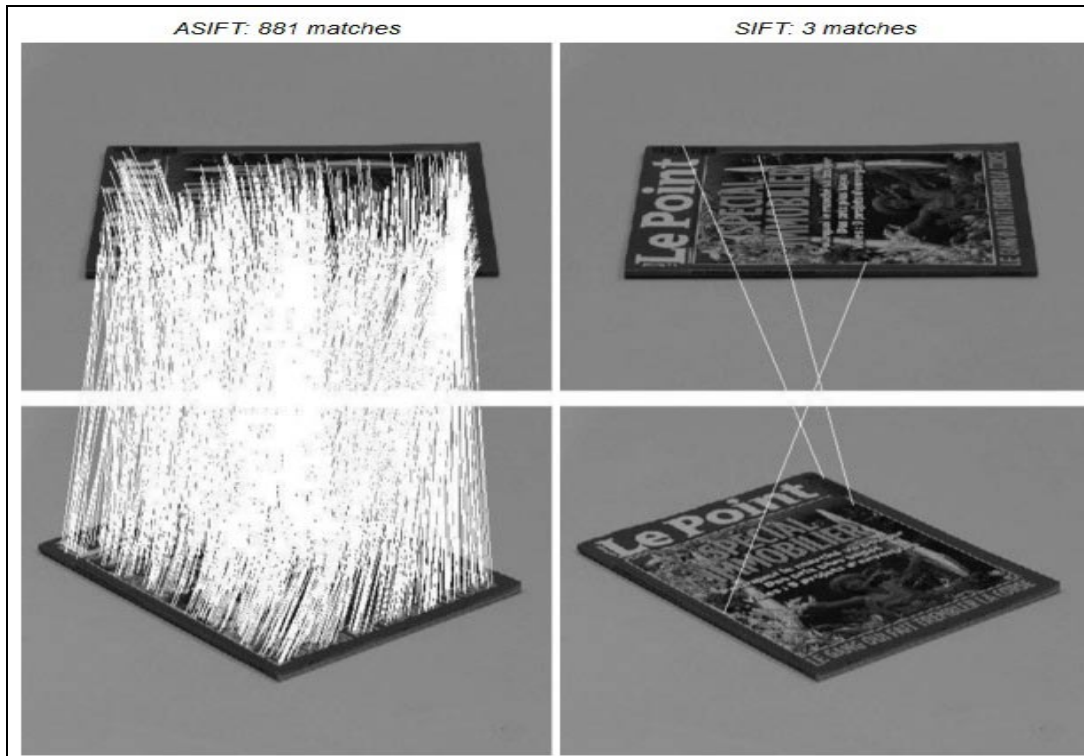
Η παρακάτω εικόνα δείχνει ένα παράδειγμα μεγάλης κλίσης μετάβασης. Η πάνω πρόσθια εικόνα συμπίπτει προς τη μια κατεύθυνση στην αριστερή εικόνα από μια λοξή προβολή και στριμώχεται σε μια ορθογώνια κατεύθυνση από μια άλλη λοξή προβολή. Η απόλυτη κλίση, δηλαδή ο συντελεστής συμπίεσης είναι περίπου 6 σε κάθε όψη. Ο προκύπτων παράγοντας συμπίεσης από αριστερά στα δεξιά είναι $6 \times 6 = 36$.



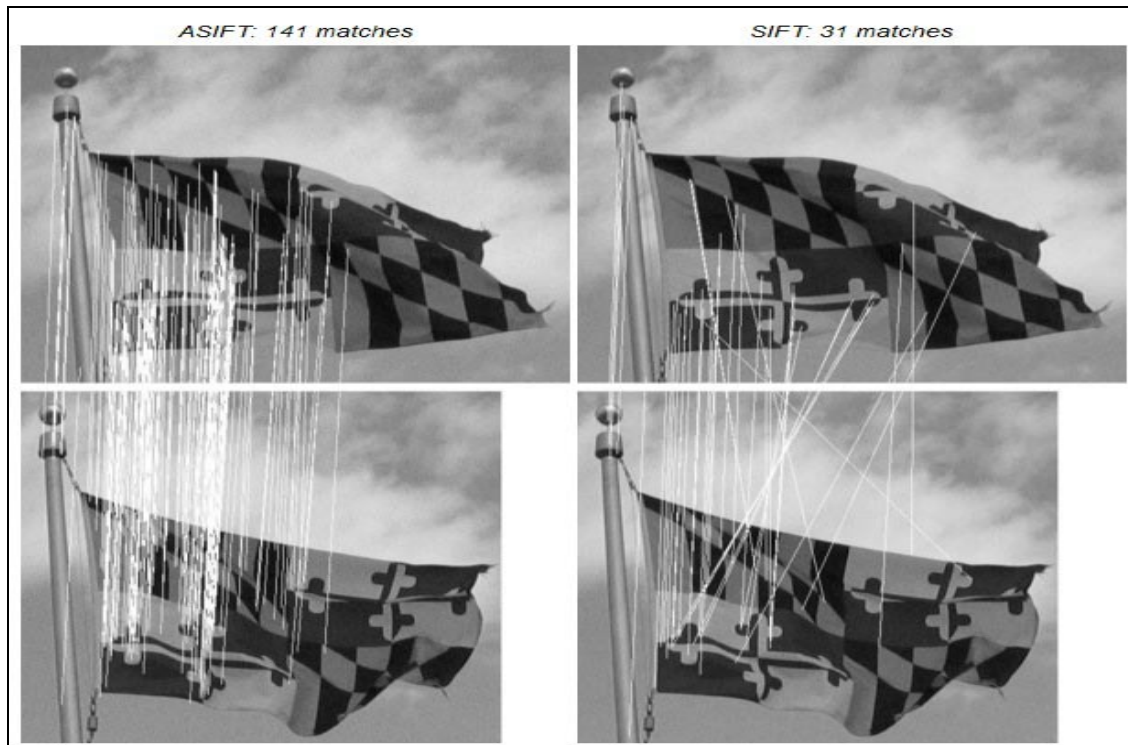
Εικόνα 26: Παράδειγμα υψηλής κλίσης μετάβασης

2.3.3 SIFT εναντίον ASIFT (Affine SIFT)

Η πολυπλοκότητα του υπολογισμού των ASIFT χαρακτηριστικών είναι ανάλογη με την περιοχή της υπό δοκιμή εικόνας. Ο ASIFT προσομοιώνει 13.5 φορές την περιοχή των αρχικών εικόνων, παράγει περίπου 13.5 φορές περισσότερα χαρακτηριστικά τόσο στην εικόνα που παίζει το ρόλο του ερωτήματος όσο και στην εικόνα αναζήτησης. Η πολυπλοκότητα για τη σύγκριση των ASIFT χαρακτηριστικών είναι επομένως 13.52, δηλαδή 180 φορές όσο αυτή του SIFT. Να σημειωθεί εδώ ότι σε τυπικές εικόνες ο υπολογισμός των ASIFT χαρακτηριστικών κυριαρχεί της υπολογιστικής πολυπλοκότητας όσον αφορά τη σύγκριση των χαρακτηριστικών. Εάν αντίστοιχα το πρόβλημα είναι να συγκριθεί μια εικόνα με μια βάση δεδομένων, αυτή η πολυπλοκότητα δεν είναι πλέον αμελητέα, και χρειάζεται να εκτελεστούν 180 περισσότερες συγκρίσεις σε ένα σοβαρό περιορισμό.



Εικόνα 27: Σύγκριση ASIFT & SIFT στο ίδιο ζεύγος εικόνων



Εικόνα 28: Σύγκριση ASIFT & SIFT σε παραμορφωμένο ζεύγος εικόνων

2.4 Ιστόγραμμα Προσανατολισμένων Κλίσεων (HOG)

Το ιστόγραμμα προσανατολισμένων κλίσεων [20] είναι περιγραφείς χαρακτηριστικών οι οποίοι χρησιμοποιούνται στην [τεχνητή όραση](#) και στην [επεξεργασία εικόνας](#), στοχεύοντας στην ανίχνευση αντικειμένων σε μια σκηνή. Αυτή η τεχνική μετράει εμφανίσεις των κλίσεων προσανατολισμού σε εντοπισμένα τμήματα μιας εικόνας. Αυτή η μέθοδος είναι παρόμοια με εκείνη του ιστογράμματος προσανατολισμού ακμών (edge orientation histograms), [μετασχηματισμός κλιμακωτά αναλλοίωτων χαρακτηριστικών](#), περιβάλλοντα σχήματα, αλλά διαφέρει στο ότι υπολογίζεται σε ένα πυκνό πλέγμα από ομοιόμορφα καταναμημένα κελιά και χρησιμοποιεί επικάλυψη των τοπικών αντιθέσεων για μεγαλύτερη ακρίβεια.

Η Navneet Dalal και ο Bill Triggs, ερευνητές του Γαλλικού Εθνικού Ινστιτούτου έρευνας στην επιστήμη των υπολογιστών & Ελέγχου, πρώτοι παρουσίασαν αυτή τη μέθοδο σε μια εργασία τους για το CVPR συνέδριο στην τεχνητή όραση και αναγνώριση προτύπων. Σε αυτή την εργασία ο αλγόριθμος τους επικεντρώθηκε στο πρόβλημα της ανίχνευσης πεζών σε στατικές εικόνες, αν και από τότε επέκτειναν τις δοκιμές τους για να συμπεριλάβουν την ανίχνευση ανθρώπων σε ταινίες και βίντεο καθώς και σε μια ποικιλία κοινών ζώων και οχημάτων σε στατικές εικόνες.

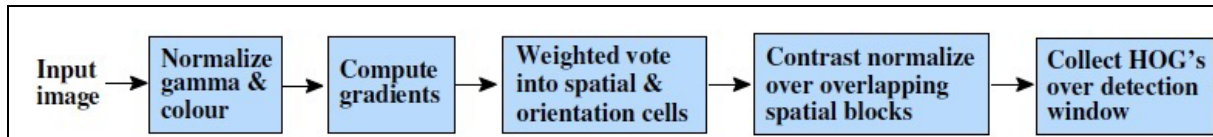
2.4.1 Θεωρία της μεθόδου

Η βασική σκέψη πίσω από αυτή τη μέθοδο είναι ότι η εμφάνιση και το σχήμα ενός τοπικού αντικειμένου μέσα σε μια εικόνα μπορεί να περιγραφεί από την κατανομή των εντάσεων των κλίσεων ή των ακμών κατεύθυνσης. Η εφαρμογή αυτών των περιγραφέων μπορεί να επιτευχθεί διαιρώντας την εικόνα σε μικρές περιοχές οι οποίες συνδέονται μεταξύ τους, τα λεγόμενα κελιά. Για κάθε τέτοιο κελί καταρτίζεται ένα ιστόγραμμα των κλίσεων ή των ακμών κατεύθυνσης για τα pixels μέσα σε αυτό το κελί. Ο συνδυασμός αυτών των ιστογραμμάτων στη συνέχεια θα αποτελέσει των περιγραφέα. Για βελτιωμένη ακρίβεια, τα τοπικά ιστογράμματα μπορούν να κανονικοποιηθούν βάση της αντίθεσης. Αυτό μπορεί να επιτευχθεί υπολογίζοντας ένα μέτρο της έντασης σε μια μεγαλύτερη περιοχή της εικόνας, το λεγόμενο μπλοκ, και στη συνέχεια χρησιμοποιώντας αυτή την τιμή κανονικοποιούνται όλα τα κελιά σε αυτό το μπλοκ. Αυτή η ομαλοποίηση επιτρέπει μεγαλύτερη αμεταβλητότητα σε αλλαγές φωτισμού και σκίασης.

Ο HOG περιγραφέας διατηρεί μερικά βασικά πλεονεκτήματα έναντι άλλων μεθόδων. Δεδομένου ότι ο HOG περιγραφέας λειτουργεί σε εντοπισμένα κελιά, η μέθοδος υποστηρίζει αμεταβλητότητα σε γεωμετρικούς και φωτομετρικούς μετασχηματισμούς εκτός από τον προσανατολισμό του αντικειμένου. Τέτοιες αλλαγές θα εμφανιστούν μόνο σε μεγαλύτερες χωρικά περιοχές.

Επιπλέον, όπως ανακάλυψαν οι ερευνητές του, χονδροειδή χωρική δειγματοληψία, λεπτή δειγματοληψία προσανατολισμού και ισχυρή τοπική φωτομετρική κανονικοποίηση επιτρέπει στην ατομική κίνηση των σώματος των πεζών να αγνοηθεί, εφόσον διατηρούν περίπου όρθια θέση. Συνεπώς ο περιγραφέας HOG θεωρείται ιδιαίτερα κατάλληλος για ανίχνευση ανθρώπων σε εικόνες.

2.4.2 Βασικά στάδια



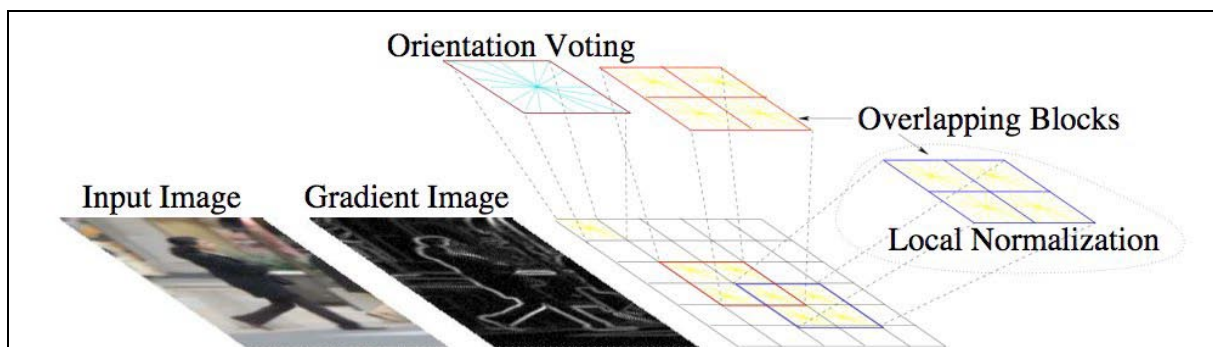
Εικόνα 29: Βασικά στάδια του αλγορίθμου HOG

ο **Υπολογισμός της κλίσης**: Το πρώτο υπολογιστικό βήμα σε πολλούς ανιχνευτές χαρακτηριστικών στην προ-επεξεργασία εικόνας είναι να εξασφαλίσει κανονικοποιημένες χρωματικές και γάμμα τιμές. Όπως προτείνουν οι δημιουργοί του συγκεκριμένου αλγορίθμου, αυτό το βήμα μπορεί να παραληφθεί κατά τον υπολογισμό HOG περιγραφέων. Αυτό συμβαίνει διότι η ακόλουθη εξομάλυνση περιγραφέων επιτυγχάνει ουσιαστικά το ίδιο αποτέλεσμα. Αντί αυτού ως πρώτο υπολογιστικό βήμα ορίζεται ο υπολογισμός των τιμών της κλίσης. Η συνηθέστερη μέθοδος για αυτό είναι η εφαρμογή μιας κεντραρισμένης μάσκας μιας διάστασης, η οποία είναι παράγωγος διακριτών σημείων, σε οριζόντια ή κάθετη κατεύθυνση ή και στις δυο. Συγκεκριμένα αυτή η μέθοδος το φιλτράρισμα των δεδομένων του χρώματος ή της έντασης της εικόνας με τους ακόλουθους πυρήνες φίλτρου.

$$[-1, 0, 1] \text{ and } [-1, 0, 1]^T.$$

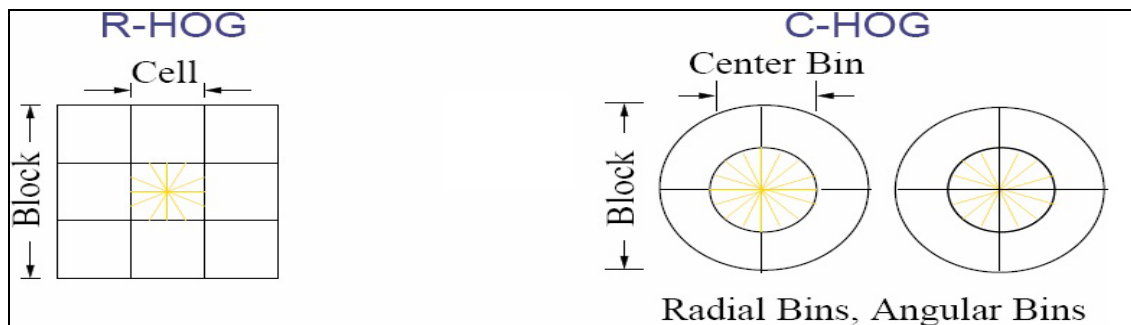
Οι δημιουργοί του αλγορίθμου χρησιμοποίησαν και πολυπλοκότερες μάσκες όπως η 3x3 Sobel μάσκα ή διαγώνιες μάσκες. Επίσης δοκίμασαν την Gaussian εξομάλυνση αλλά κατέληξαν στο ίδιο συμπέρασμα για όλες αυτές τις μάσκες, παρουσίαζαν χαμηλότερη απόδοση στα πειράματα για ανίχνευση ανθρώπων σε εικόνες.

ο **Δημιουργία Ιστογραμμάτων των κελιών**: Το δεύτερο βήμα του υπολογισμού περιλαμβάνει τη δημιουργία ιστογραμμάτων των κελιών. Κάθε pixel στο εσωτερικό του κελιού δημιουργεί μια σταθμισμένη ψήφο, για ένα βασισμένο στον προσανατολισμό κανάλι ιστογράμματος, η οποία βασίζεται στις τιμές που βρέθηκαν στον υπολογισμό κλίσεων του προηγούμενο βήματος. Τα ίδια τα κελιά μπορεί να είναι είτε ορθογώνια είτε ακτινωτά σε σχήμα. Τα κανάλια ιστογράμματος κατανέμονται ομοιόμορφα από 0 μέχρι 180 μοίρες ή από 0 μέχρι 360 ανάλογα με το αν η κλίση είναι προσημασμένη ή χωρίς πρόσημο. Όσον αφορά το βάρος των ψήφων, η συμβολή κάθε pixel μπορεί να είναι είτε το μέγεθος της κλίσης είτε κάποια λειτουργία του μεγέθους. Ωστόσο σε πραγματικές δοκιμές το μέγεθος της κλίσης παράγει καλύτερα αποτελέσματα. Άλλες επιλογές για την κατανομή του βάρους θα μπορούσαν να περιλαμβάνουν την τετραγωνική ρίζα ή το τετράγωνο μεγέθους της κλίσης ή κάποια εκδοχή του μεγέθους που έχει περικοπεί.



Εικόνα 30: Δημιουργία Ιστογραμμάτων εντός των κελιών

3 **Τμήματα Περιγραφών:** Προκειμένου να ληφθούν υπόψη οι αλλαγές του φωτισμού και της αντίθεσης, οι δυνάμεις της κλίσης πρέπει να είναι τοπικά κανονικοποιημένες, το οποίο απαιτεί την ομαδοποίηση των κελιών σε μεγαλύτερα και χωρικά συνδεδεμένα τμήματα. Ο HOG περιγραφέας θεωρείται τότε το διάνυσμα των συστατικών των κανονικοποιημένων κελιών ιστογράμματος από όλες τις περιοχές των τμημάτων. Αυτά τα τμήματα συνήθως επικαλύπτονται, γεγονός που σημαίνει ότι κάθε κελί συνεισφέρει περισσότερο από μια φορά στον τελικό περιγραφέα. Υπάρχουν δυο βασικές γεωμετρικές τμημάτων, τα ορθογώνια τμήματα HOG (R-HOG) και τα κυκλικά τμήματα HOG (C-HOG). Τα R-HOG τμήματα είναι γενικά τετράγωνο πλέγματα που εκπροσωπούνται από τρεις παραμέτρους, τον αριθμό των κελιών, τον αριθμό των pixels ανά κελί και ο αριθμός των καναλιών ανά κελί ιστογράμματος.



Εικόνα 31: Γεωμετρικά τμήματα περιγραφών HOG

Τα R-HOG τμήματα εμφανίζουν αρκετές ομοιότητες με τους SIFT περιγραφείς. Ωστόσο παρά τον παρόμοιο μετασχηματισμό τους, τα R-HOG τμήματα υπολογίζονται σε πυκνά πλέγματα σε κάποια ενιαία κλίμακα χωρίς προσανατολισμένη ευθυγράμμιση. Αντίθετα οι SIFT περιγραφείς υπολογίζονται σε αραιά, αμετάβλητα σε κλίμακα σημεία κλειδιά της εικόνας, και περιστρέφονται έτσι ώστε να ευθυγραμμιστούν βάση προσανατολισμού. Επιπλέον τα R-HOG τμήματα χρησιμοποιούνται σε συνδυασμό για την κωδικοποίηση της μορφής των χωρικών πληροφοριών, ενώ οι SIFT περιγραφείς χρησιμοποιούνται μεμονωμένα.

Τα κυκλικά τμήματα HOG (C-HOG) μπορούν να βρεθούν σε δυο παραλλαγές εκείνα με ένα ενιαίο κεντρικό κελί και αυτά με ένα διαγώνια διαιρούμενο κεντρικό κελί. Επιπλέον αυτά τα κυκλικά τμήματα μπορούν να περιγραφούν με τέσσερις παραμέτρους: τον αριθμό των γωνιακών και ακτινωτών δοχείων (bins), την ακτίνα του κεντρικού δοχείου και τον παράγοντα επέκτασης της ακτίνας.

Το τελικό βήμα στην αναγνώριση αντικειμένων με τη χρήση HOG περιγραφών, είναι να τροφοδοτήσει αυτούς τους περιγραφείς σε κάποιο σύστημα αναγνώρισης το οποίο βασίζεται στην μάθηση με πλήρη επίβλεψη. Ένα σύστημα ταξινόμησης SVM [] είναι ένας δυαδικός ταξινομητής που αναζητά τη βέλτιστη λύση. Ένα τέτοιο σύστημα αφού εκπαιδευτεί σε εικόνες που περιέχουν κάποιο συγκεκριμένο αντικείμενο, είναι σε θέση να παίρνει αποφάσεις σχετικά με την παρουσία ενός αντικειμένου, σε πρόσθετες εικόνες δοκιμής.

Στον παρακάτω πίνακα φαίνεται μια συνολική επισκόπηση των μεθόδων που αναφέρθηκαν στο κεφάλαιο καθώς σε τι υπερτερούν αλλά και σε τι μειονεκτούν σε σχέση με τις άλλες μεθόδους.

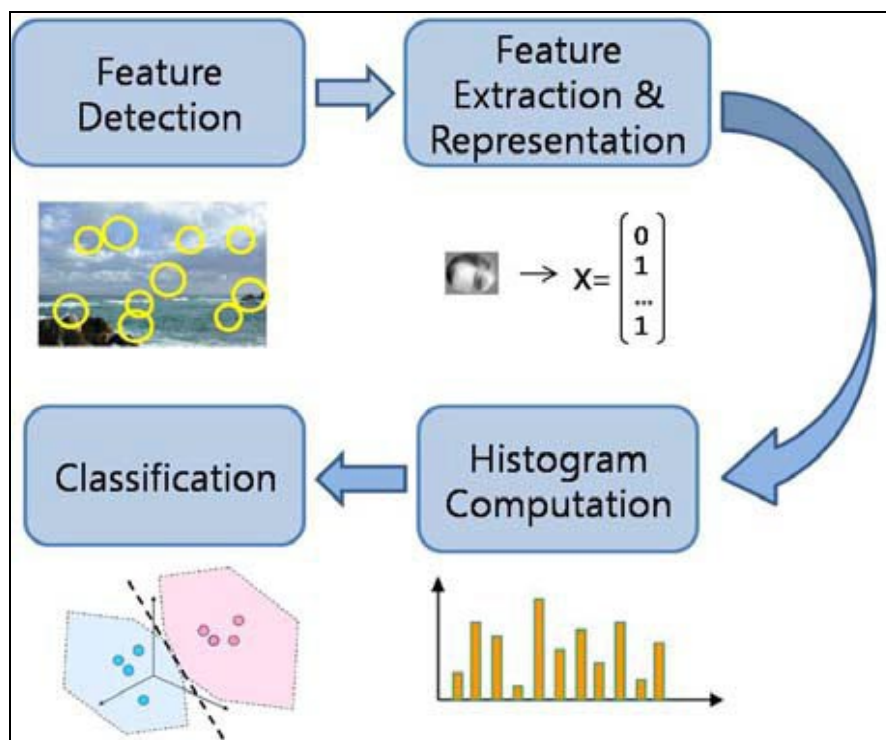
2.5 Συνοπτικός πίνακας μεθόδων Εξαγωγής Χαρακτηριστικών

Πίνακας 1: Συνολική Παρουσίαση των Μεθόδων εξαγωγής χαρακτηριστικών

ΟΝΟΜΑΣΙΑ	ΧΡΟΝΟΛΟΓΙΑ ΕΚΔΟΣΗΣ	ΣΥΝΤΟΜΗ ΠΕΡΙΓΡΑΦΗ ΜΕΘΟΔΟΛΟΓΙΑΣ	ΘΕΤΙΚΑ ΣΤΟΙΧΕΙΑ	ΑΡΝΗΤΙΚΑ ΣΤΟΙΧΕΙΑ
SIFT	1999	<ol style="list-style-type: none"> 1. ΑΝΙΧΝΕΥΣΗ ΑΜΕΤΑΒΛΗΤΩΝ ΣΕ ΚΛΙΜΑΚΑ ΧΑΡΑΚΤΗΡΗΣΤΙΚΩΝ 2. ΑΝΤΙΣΤΟΙΧΗΣΗ ΧΑΡΑΚΤΗΡΗΣΤΙΚΩΝ & ΠΡΣΟΘΗΚΗ ΣΕ ΕΥΡΕΤΗΡΙΟ 3. ΕΚΧΩΡΗΣΗ ΠΡΟΣΑΝΑΤΟΛΙΣΜΟΥ 	ΕΥΡΩΣΤΟΣ ΣΕ ΠΑΡΑΜΟΡΦΩΣΕΙΣ ΠΡΟΣΑΝΑΤΟΛΙΣΜΟΥ ΦΩΤΙΣΜΟΥ & ΣΥΣΧΕΤΙΣΜΕΝΕΣ ΣΤΡΕΒΛΩΣΕΙΣ	ΣΧΕΤΙΚΑ ΜΕΓΑΛΟΣ ΣΥΝΟΛΙΚΟΣ ΧΡΟΝΟΣ ΥΠΟΛΟΓΙΣΜΟΥ
SURF	2006	<ol style="list-style-type: none"> 1. ΑΝΙΧΝΕΥΣΗ ΣΗΜΕΙΩΝ ΕΝΔΙΑΦΕΡΟΝΤΟΣ 2. ΤΟΠΙΚΟΙ ΠΕΡΙΓΡΑΦΕΙΣ 3. ΑΝΤΙΣΟΙΧΗΣΗ ΧΑΡΑΚΤΗΡΗΣΤΙΚΩΝ 	ΜΙΚΡΟΤΕΡΟΣ ΣΥΝΟΛΙΚΟΣ ΧΡΟΝΟΣ ΥΠΟΛΙΣΜΟΥ (ΕΠΙΤΑΧΥΝΕΙ ΤΗΝ ΑΝΙΧΝΕΥΣΗ)	<ol style="list-style-type: none"> 1. ΛΙΓΑ ΣΗΜΕΙΑ ΧΑΡΑΚΤΗΡΗΣΤΙΚΩΝ ΣΕ ΠΕΡΙΟΧΕΣ ΜΙΚΡΗΣ ΑΝΤΙΘΕΣΗΣ 2. ΠΡΟΒΛΗΜΑ ΣΤΗΝ ΠΡΟΟΠΤΙΚΗ ΠΑΡΑΜΟΡΦΩΣΗ
ASIFT	2011	<ol style="list-style-type: none"> 1. ΠΡΟΣΟΜΟΙΩΝΟΝΤΑΙ ΟΛΕΣ ΟΙ ΠΙΘΑΝΕΣ ΣΤΡΕΒΛΩΣΕΙΣ 2. ΣΥΓΚΡΙΣΗ ΤΩΝ ΕΙΚΟΝΩΝ ΠΡΟΣΟΜΟΙΩΣΗΣ ΜΕ SIFT 3. ΑΠΟΒΟΛΗ/ΦΙΛΤΡΑΡΙΣΜΑ ΕΣΦΑΛΜΕΝΩΝ ΑΝΤΙΣΟΙΧΙΩΝ 	ΠΛΗΡΩΣ ΑΜΕΤΑΒΛΗΤΟΣ ΣΕ 6 ΣΥΝΟΛΙΚΑ ΠΑΡΑΜΕΤΡΟΥΣ (ΣΕ ΣΧΕΣΗ ΜΕ ΤΙΣ 4 ΤΟΥ SIFT)	ΑΥΞΑΝΕΤΑΙ Ο ΧΡΟΝΟΣ ΥΠΟΛΟΓΙΣΜΟΥ (ΠΑΡΑΓΕΙ 13.5 ΦΟΡΕΣ ΠΕΡΙΣΣΟΤΕΡΑ ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ ΣΕ ΣΧΕΣΗ ΜΕ SIFT)
HOG	2005	<ol style="list-style-type: none"> 1. ΥΠΟΛΟΓΙΣΜΟΣ ΤΗΣ ΚΛΙΣΗΣ 2. ΔΗΜΙΟΥΡΓΙΑ ΙΣΤΟΓΡΑΜΜΑΤΟΣ ΤΩΝ ΚΕΛΙΩΝ 3. BLOCKS ΠΕΡΙΓΡΑΦΕΩΝ 	ΜΠΟΡΕΙ ΝΑ ΑΝΤΙΠΡΟΣΩΠΕΥΣΕΙ ΤΟ ΣΚΛΗΡΟ ΣΧΗΜΑ ΤΟΥ ΑΝΤΙΚΕΙΜΕΝΟΥ	ΟΧΙ ΕΥΡΩΣΤΟΣ ΣΕ ΔΙΑΦΟΡΕΣ ΠΑΡΑΜΟΡΦΩΣΕΙΣ

3. Αλγόριθμοι Κατηγοριοποίησης Εικόνων

Η ικανότητα του ανθρώπου να αναλύει και να ταξινομεί αντικείμενα και σκηνές με ταχύτητα και ακρίβεια είναι κάτι το οποίο θεωρείται ιδιαίτερα χρήσιμο στην καθημερινότητα μας. Διαπιστώθηκε ότι οι άνθρωποι είναι σε θέση να κατηγοριοποιήσουν πολύπλοκες φυσικές σκηνές πολύ γρήγορα. Ο άνθρωπος για να κατανοήσει ένα σύνθετο σκηνικό, το πρώτο βήμα είναι η αναγνώριση των αντικειμένων και στη συνέχεια η αναγνώριση της κατηγορίας στην οποία ανήκει η σκηνή. Για να επιτευχθεί αυτό στην τεχνητή όραση, χρησιμοποιούνται διάφοροι ταξινομητές οι οποίοι έχουν διαφορετικά χαρακτηριστικά και δυνατότητες. Στο παρελθόν έχουν αναπτυχθεί πολλοί τέτοιοι ταξινομητές από διάφορους ερευνητές. Αυτές οι μέθοδοι περιλαμβάνουν αφελείς ταξινομητές, μηχανές διανυσμάτων υποστήριξης, k-πλησιέστεροι γείτονες, Gaussian μοντέλο μείγματος, δένδρο αποφάσεων και ακτινωτή συνάρτηση βάσης. Αυτοί οι ταξινομητές χρησιμοποιούνται σε αλγορίθμους που περιλαμβάνουν αναγνώριση αντικειμένων.



Εικόνα 32: Εννοιολογική απεικόνιση της διαδικασίας ταξινόμησης εικόνων

Ωστόσο η αναγνώριση αντικειμένων [21] σε μια σκηνή παραμένει πρόκληση για διάφορους λόγους. Ο πρώτος και πιο προφανής λόγος είναι ότι υπάρχουν περίπου 10.000 έως 30.000 διαφορετικές κατηγορίες αντικειμένων. Ο δεύτερος λόγος είναι η διακύμανση στις απόψεις, όπου πολλά αντικείμενα μπορούν να φαίνονται διαφορετικά από διαφορετικές οπτικές γωνίες. Ο τρίτος λόγος είναι ο φωτισμός, ο οποίος κάνει τα ίδια αντικείμενα να μοιάζουν με διαφορετικά αντικείμενα. Ο τέταρτος λόγος θεωρείται η ύπαρξη πολλών μη χρήσιμων πληροφοριών στο φόντο της εικόνας, πράγμα το οποίο ο ταξινομητής δεν είναι πολλές φορές σε θέση να ξεχωρίσει από το αντικείμενο καθεαυτό.

Άλλες προκλήσεις περιλαμβάνουν την παραμόρφωση της κλίμακας, σύγκλιση και ένδο-ταξική παραλλαγή. Οι εφαρμογές για την κατηγοριοποίηση εικόνων στην τεχνητή όραση περιλαμβάνουν την υπολογιστική φωτογραφία, παρακολούθηση για λόγους ασφαλείας και βοηθητική οδήγηση.

Παρακάτω γίνεται μια αναφορά στις σημαντικότερες τεχνικές που χρησιμοποιούνται στην ταξινόμηση εικόνων.

3.1 Τεχνητά Νευρωνικά Δίκτυα

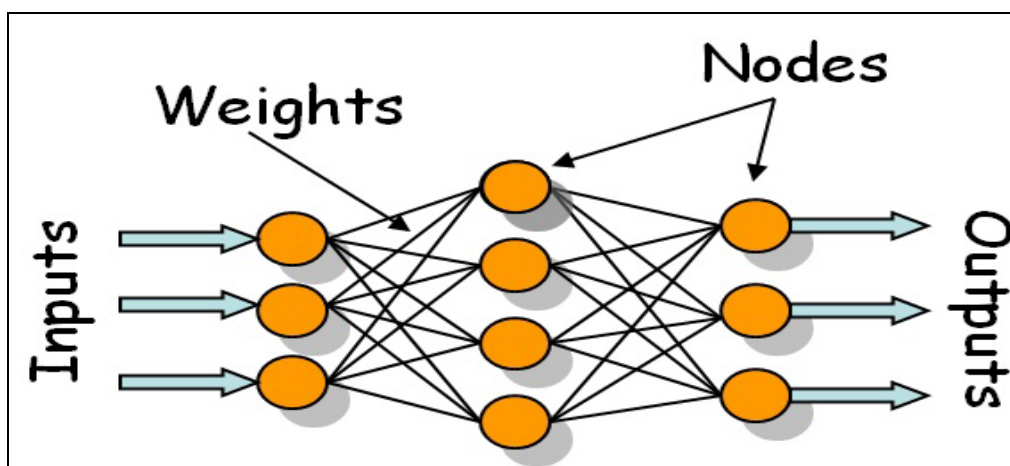
3.1.1 Γενικά για τα Νευρωνικά Δίκτυα

Ο όρος νευρωνικό δίκτυο [22] παραδοσιακά χρησιμοποιείται για την αναφορά σε ένα δίκτυο ή κύκλωμα των βιολογικών νευρώνων [23]. Η σύγχρονη χρήση του όρου αναφέρεται συχνά σε τεχνητά νευρωνικά δίκτυα τα οποία αποτελούνται από τεχνητούς νευρώνες ή κόμβους. Έτσι ο όρος έχει δυο διαφορετικές χρήσεις: α) Βιολογικά νευρωνικά δίκτυα, τα οποία είναι φτιαγμένα από πραγματικούς βιολογικούς νευρώνες οι οποίοι είναι συνδεδεμένοι σε ένα νευρωνικό σύστημα β) τεχνητά νευρωνικά δίκτυα τα οποία αποτελούνται από διασυνδεδεμένους τεχνητούς νευρώνες. Εμείς στη συνέχεια θα επικεντρωθούμε μονάχα στα τεχνητά νευρωνικά δίκτυα που έχουν σχέση με το αντικείμενο της παρούσας εργασίας.

3.1.2 Ορισμός Τεχνητών Νευρωνικών Δικτύων

Ένα τεχνητό νευρωνικό δίκτυο (ANN), συνήθως ονομάζεται απλά νευρωνικό δίκτυο, είναι ένα μαθηματικό ή υπολογιστικό μοντέλο το οποίο είναι εμπνευσμένο από τη δομή και τις λειτουργικές πτυχές των βιολογικών νευρωνικών δικτύων. Ένα τέτοιο δίκτυο, αποτελείται από μια ομάδα διασυνδεδεμένων τεχνητών νευρώνων, και επεξεργάζεται τις πληροφορίες χρησιμοποιώντας μια συνδυαστική προσέγγιση υπολογισμού.

Στις περισσότερες περιπτώσεις ένα τεχνητό νευρωνικό δίκτυο θεωρείται ένα προσαρμοσμένο σύστημα το οποίο αλλάζει τη δομή του, βασισμένο σε εξωτερικές ή εσωτερικές πληροφορίες που ρέουν μέσα από το δίκτυο κατά τη φάση της μάθησης. Τα σύγχρονα νευρωνικά δίκτυα είναι μη γραμμικά στατιστικά εργαλεία μοντελοποίησης δεδομένων και χρησιμοποιούνται σε μοντελοποιήσεις πολύπλοκων σχέσεων μεταξύ εισροών και εκροών ή για βρίσκει πρότυπα δεδομένων. Συνοπτικά λοιπόν μπορούμε να πούμε ότι ο στόχος των νευρωνικών δικτύων είναι η μετατροπή των εισροών σε εκροές με νόημα.



Εικόνα 33: Κλασική μορφή τεχνητού νευρωνικού δικτύου

Ένα τεχνητό νευρωνικό δίκτυο είναι με άλλα λόγια ένας παράλληλα καταναμημένος επεξεργαστής ο οποίος έχει μια φυσική τάση για την αποθήκευση βιωματικής γνώσης. Τέτοια δίκτυα μπορούν να παρέχουν τις κατάλληλες λύσεις για προβλήματα τα οποία γενικά χαρακτηρίζονται από μη γραμμικές σχέσεις, υψηλή διάσταση θορύβου, περίπλοκες ανακρίβειες και ατελείς ή επιρρεπείς σε λάθη αισθητήρες. Ένα βασικό πλεονέκτημα τους είναι ότι το μοντέλο του συστήματος μπορεί να κατασκευαστεί από τα διαθέσιμα δεδομένα.

3.1.3 Ταξινόμηση εικόνων με χρήση Τεχνητών Νευρωνικών Δικτύων

Η ταξινόμηση εικόνων με χρήση νευρωνικών δικτύων πραγματοποιείται εξάγοντας χαρακτηριστικά υφής και στη συνέχεια εφαρμόζοντας τον αλγόριθμο backpropagation [24]. Η υφή χαρακτηρίζεται από τη χωρική διανομή των επιπέδων του γκρι σε μια γειτονιά. Στην ταξινόμηση υφής, ως στόχος θεωρείται ο ορισμός μιας άγνωστης εικόνας σε ένα από τα γνωστά σύνολα των κλάσεων υφής. Επίσης τα χαρακτηριστικά υφής είναι μονοδιάστατοι αριθμοί, διακριτά ιστογράμματα ή εμπειρικές διανομές. Στο σχεδιασμό θεωρούνται τέσσερα χαρακτηριστικά υφής, δηλαδή η γωνιακή δεύτερη στιγμή, συσχέτιση αντίθεσης και διακύμανση. Η υφή και ο τόνος έχουν μια ανέκφραστη σχέση μεταξύ τους. Για να καταγραφεί η χωρική εξάρτηση των τιμών των επιπέδων του γκρι, οι οποίες συμβάλλουν στην αντίληψη της υφής, χρησιμοποιείται ένας δισδιάστατος πίνακας εξάρτησης και ανάλυσης υφής. Δεδομένου ότι η υφή δείχνει τα χαρακτηριστικά της εξίσου από τα pixels αλλά και τις τιμές αυτών, υπάρχουν πολλές προσεγγίσεις για την κατηγοριοποίηση της υφής.

Για την εύρεση χαρακτηριστικών υφής, για κάθε pixel της εικόνας κάθε ένα pixel θεωρείται ως κέντρο και ακολουθείται από ένα 5x5 παράθυρο για εκείνο το κεντρικό pixel. Ο πίνακας σε γκρι τόνο για το συγκεκριμένο παράθυρο υπολογίζεται και κανονικοποιείται. Η τυπική απόκλιση και ο μέσος όρος λαμβάνονται στη συνέχεια για κάθε έναν από αυτούς τους πίνακες και στη συνέχεια υπολογίζονται τα χαρακτηριστικά υφής. Μετά την εξαγωγή των χαρακτηριστικών αυτών, το δίκτυο εκπαιδεύεται από τον αλγόριθμο backpropagation.

Εν συντομία, η λειτουργία του backpropagation αλγορίθμου είναι η εξής:

1. Αρχικοποίηση βάρους σε μικρές τυχαίες τιμές.
2. Τροφοδοτεί τα διανύσματα εισόδου X_0, X_1, \dots, X_6 μέσα στο δίκτυο και υπολογίζει το άθροισμα των βαρών τα οποία έρχονται στη μονάδα και στη συνέχεια εφαρμόζεται η ακόλουθη συνάρτηση:

$$f(x) = \frac{1}{1 + e^{-x}}$$

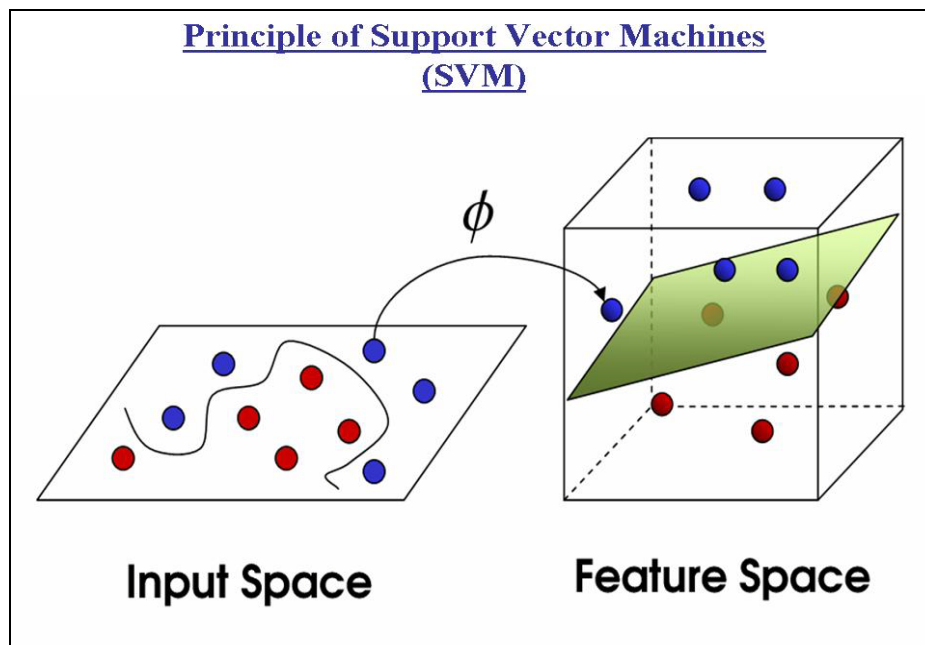
Επίσης θέτει όλα τα επιθυμητά αποτελέσματα d_0, d_1, \dots, d_5 τυπικά στο μηδέν εκτός από αυτό που αντιστοιχεί στην κλάση από την οποία προέρχεται η είσοδος.

3. Υπολογίζει τον όρο σφάλμα για κάθε μονάδα.
4. Προσθέτει τους συντελεστές δέλτα για κάθε μονάδα.
5. Όλα τα βήματα εκτός από το πρώτο επαναλαμβάνονται μέχρι το λάθος να θεωρηθεί εντός λογικών ορίων και στη συνέχεια τα προσαρμοσμένα βάρη αποθηκεύονται για αναφορά στον αλγόριθμο αναγνώρισης.

3.2 Μηχανές Υποστήριξης Διανυσμάτων (SVM)

Η μηχανή υποστήριξης διανυσμάτων [25] θεωρείται μια έννοια στον τομέα της στατιστικής και της επιστήμης υπολογιστών για ένα σύνολο σχετικών εποπτευομένων μεθόδων μάθησης, οι οποίες αναλύουν δεδομένα και αναγνωρίζουν πρότυπα, και χρησιμοποιούνται για ταξινόμηση και ανάλυση παλινδρόμησης. Ο αρχικός αλγόριθμος SVM εφευρέθηκε από τον Vladimir Vapnik και η ισχύουσα πρότυπη υλοποίηση, προτάθηκε το 1995. Το κίνητρο για την ανάπτυξη αυτής της μεθόδου βρίσκεται στο ότι η ταξινόμηση δεδομένων είναι μια κοινή εργασία της μηχανικής μάθησης. Η αρχή των μηχανών υποστήριξης διανυσμάτων έγκειται στο εξής: αν υποθέσουμε ότι κάποια σημεία δεδομένων ανήκουν σε μια κλάση από δυο κλάσεις συνολικά. Στόχος είναι να αποφασισθεί σε ποια από αυτές τις κλάσεις ένα νέο σημείο δεδομένων θα υπάγεται. Στις περιπτώσεις των SVM, ένα σημείο δεδομένων θεωρείται ως ένα p -διαστάσεων διάνυσμα (μια λίστα από p αριθμούς) και θέλουμε να γνωρίζουμε αν μπορούμε να χωρίσουμε αυτά τα σημεία σε επίπεδα.

Μια τυποποιημένη μηχανή υποστήριξης διανυσμάτων, παίρνει ένα σύνολο δεδομένων εισόδου και προβλέπει για καθένα από αυτά, ποιες από τις δυο πιθανές κατηγορίες αποτελεί την είσοδο, κάνοντας τη μηχανή αυτή έναν μη πιθανολογικό δυαδικό γραμμικό ταξινομητή. Λαμβάνοντας υπόψη ένα σύνολο παραδειγμάτων εκπαίδευσης, κάθε ένα χαρακτηρίζεται ότι ανήκει σε μια από τις δυο κατηγορίες, ένας αλγόριθμος εκπαίδευσης SVM δημιουργεί ένα μοντέλο το οποίο αποδίδει νέα παραδείγματα στη μια κατηγορία ή στην άλλη. Ένα SVM μοντέλο είναι μια αναπαράσταση των παραδειγμάτων, όπως τα σημεία στο χώρο, χαρτογραφημένα έτσι ώστε τα παραδείγματα των επιμέρους κατηγοριών να χωρίζονται από ένα σαφές κενό, όσο το δυνατόν ευρύτερο γίνεται. Στη συνέχεια νέα παραδείγματα χαρτογραφούνται στον ίδιο χώρο και αναμένεται να ανήκουν σε μια κατηγορία, με βάση σε ποια πλευρά του χάσματος υπάγονται.



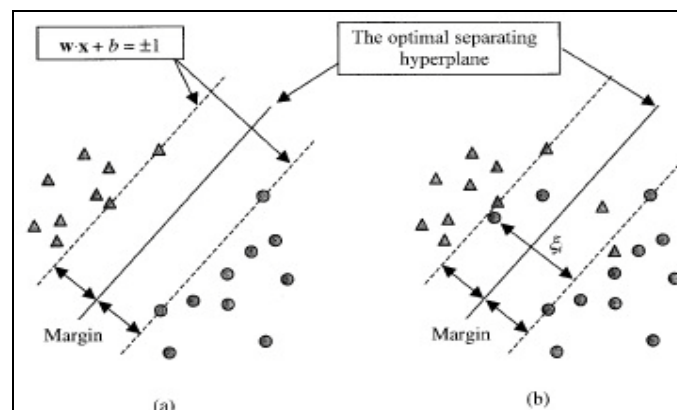
Εικόνα 34: Η αρχή των μηχανών υποστήριξης διανυσμάτων (SVM)

3.2.1 Ταξινόμηση εικόνων με χρήση Μηχανών Υποστήριξης Διανυσμάτων

Η εφαρμοσιμότητα των SVM για την ταξινόμηση εικόνων θα παρουσιαστεί παρακάτω. Οι μηχανές υποστήριξης διανυσμάτων θεωρούνται ανώτερες από όλους τους αλγορίθμους μηχανικής μάθησης. Υποστηρίζουν αλγορίθμους βελτιστοποίησης για τον εντοπισμό των βέλτιστων ορίων μεταξύ των τάξεων. Τα βέλτιστα όρια θα πρέπει να είναι γενικευμένα σε πρωτόγνωρα δείγματα με ελάχιστα λάθη μεταξύ όλων των δυνατών ορίων διαχωρισμού των κλάσεων. Ως εκ τούτου να ελαχιστοποιούνται οι συγχύσεις μεταξύ των κλάσεων.

Η επαγωγική αρχή πίσω από τα SVM θεωρείται η ελαχιστοποίηση του κινδύνου που αφορά τη διάρθρωση [26]. Ο κίνδυνος για μηχανή μάθησης (R) οριοθετείται από το άθροισμα των εμπειρικών κινδύνων που υπολογίζονται από δείγματα εκπαίδευσης (R_{emp}) και ένα διάστημα εμπιστοσύνης (ψ) : $R \leq R_{emp} + \psi$. Η στρατηγική της ελαχιστοποίησης κινδύνου είναι να κρατήσει τον εμπειρικό κίνδυνο καθορισμένο και να ελαχιστοποιηθεί το διάστημα εμπιστοσύνης ή να μεγιστοποιήσει το περιθώριο μεταξύ ενός διαχωριστικού υπερεπιπέδου και των πλησιέστερων σημείων δεδομένων. Ένα διαχωριστικό υπερεπίπεδο θεωρείται μια γραμμή σε έναν πολυδιάστατο χώρο που χωρίζει τα δείγματα των δεδομένων σε δυο κλάσεις. Επί του παρόντος ένα ταξινομητής SVM είναι σε θέση να διαχωρίσει μόνο δυο κλάσεις. Στρατηγικές ενσωμάτωσης απαιτούνται για την επέκταση αυτής της μεθόδου, για την ταξινόμηση πολλαπλών κλάσεων.

Μια σημαντική υπόθεση ωστόσο θεωρείται ότι τα δεδομένα μπορούν να διαχωριστούν στο χώρο των χαρακτηριστικών μιας εικόνας. Είναι εύκολο να ελεγχτεί ότι δεν υπάρχει βέλτιστη λύση εάν τα δεδομένα δεν μπορούν να διαχωριστούν χωρίς σφάλματα. Για να επιλυθεί αυτό το πρόβλημα εισάγεται μια τιμή ποινής για λάθος ταξινόμηση και πιθανές υπολειπόμενες μεταβλητές.



Εικόνα 35: Διαχωρίσιμα και μη διαχωρίσιμα δείγματα

Γενικεύοντας τη μέθοδο για μη γραμμικές συναρτήσεις απόφασης, μπορούμε να πούμε ότι η μηχανή υποστήριξης διανυσμάτων, εφαρμόζει την ακόλουθη μέθοδο: αντιστοιχίζει το διάνυσμα εισόδου x σε ένα χώρο χαρακτηριστικών πολλών διαστάσεων H και βέλτιστα χωρίζει αυτόν το χώρο.

Με τη συστηματική ανάπτυξη των SVM, η λειτουργία του πυρήνα παίζει ένα πολύ σημαντικό ρόλο στον εντοπισμό πολύπλοκων ορίων απόφασης μεταξύ των κλάσεων. Με το να υπάρχουν τα δεδομένα εισόδου σε διαστατό χώρο, η λειτουργία του πυρήνα μετατρέπει τα μη γραμμικά όρια των αρχικών δεδομένων χώρου, σε γραμμικά στον υψηλών διαστάσεων χώρο. Για αυτό το λόγο λοιπόν η απόδοση του SVM επηρεάζεται από την επιλογή της λειτουργίας του πυρήνα και των κατάλληλων τιμών για τις αντίστοιχες παραμέτρους του πυρήνα.

3.3 Ασαφή μέτρα

Η θεωρία των ασαφών μέτρων [27] θεωρεί μια σειρά από ειδικές κλάσεις μέτρων καθεμία από τις οποίες χαρακτηρίζεται από μια ιδιαίτερη ιδιότητα. Ορισμένα από τα μέτρα που χρησιμοποιούνται σε αυτή τη θεωρία είναι η αληθοφάνεια και πίστη των μέτρων, η ασάφεια ορίζει τη συνάρτηση των μελών και τα κλασσικά μέτρα πιθανοτήτων. Στη θεωρία της ασαφούς μέτρησης, οι όροι είναι ακριβείς, αλλά οι πληροφορίες σχετικά με ένα μοναχικό στοιχείο δεν επαρκούν για να καθορίσουν ποιες ειδικές κλάσεις του μέτρου πρέπει να χρησιμοποιηθούν. Η κεντρική ιδέα της ασαφούς θεωρίας θεωρείται το ασαφές μέτρο, έννοια η οποία εισήχθη από τον Choquet το 1953 και ορίστηκε ανεξάρτητα από τον Sugeno το 1974 στο πλαίσιο των ασαφών ολοκληρωμάτων.

Στα ασαφή μέτρα, διαφορετικές στοχαστικές σχέσεις εντοπίζονται για την περιγραφή των ιδιοτήτων μιας εικόνας. Οι διαφορετικοί τύποι στοχαστικών σχέσεων που συγκεντρώνονται είναι ένα σύνολο ιδιοτήτων τα μέλη του οποίου είναι ασαφείς στη συμβολή τους. Η μέθοδος των ασαφών μέτρων δίνει τη δυνατότητα περιγραφής διαφορετικών τύπων στοχαστικών ιδιοτήτων στην ίδια μορφή. Εάν η ασαφής ιδιότητα σχετίζεται περισσότερο με μια περιοχή, τότε ένα ασαφές μέτρο χρησιμοποιείται. Η ασαφής συνάρτηση χρησιμοποιείται όταν η στοχαστική ιδιότητα πρέπει να περιγραφεί από μια συγκεκριμένη διανομή των τιμών του γκρι. Η συγχώνευση αυτών των δυο στοχαστικών ιδιοτήτων παρουσιάζεται ως ασαφές μέτρο και η ασαφής συνάρτηση ορίζει μια περιοχή η οποία επιτυγχάνεται από ένα ασαφές συστατικό. Το αποτέλεσμα αυτού του συστατικού θεωρείται ένα νέο ασαφές μέτρο.

3.3.1 Ταξινόμηση εικόνων με χρήση Ασαφών Μέτρων

Για την εξαγωγή των στοχαστικών ιδιοτήτων χρησιμοποιούνται διάφοροι μέθοδοι, επειδή οι στοχαστικές ιδιότητες είναι σύνθετες και η διάσπαση τους είναι δυνατή μόνο αν οι ιδιότητες που τις αποτελούν υπολογίζονται από διαφορετικές πλευρές. Οι διαφορετικές ιδιότητες που συντίθενται, θα πρέπει να χαρτογραφηθούν μεμονωμένα σε διαφορετικούς χώρους. Το φιλτράρισμα είναι προσαρμοσμένο για τη στοχαστική αλλαγή των τιμών του γκρι που σχετίζονται με τα γειτονικά pixels. Οι περιοχές με διαφορετικές στοχαστικές διακυμάνσεις επιλέγονται σωστά από την εφαρμογή μετασχηματισμού κυματιδίων με επιλεγμένες σταθερές. Η δυναμική της στοχαστικής ιδιότητας μεταξύ των pixel περιγράφει πολλές σχετικές ιδιότητες των υφών.

Οι τιμές για τις πιθανές περιοχές των υφών αντιπροσωπεύονται από ένα ασαφές μέτρο, συνδέονται με τις τιμές στα pixels που αντιπροσωπεύουν τη δύναμη των ιδιοτήτων για μια υφή. Οι παραπάνω τιμές περιγράφονται από μια ασαφή λειτουργία όπου οι τιμές κανονικοποιούνται σε 1. Το άθροισμα όλων των συνδυασμών των ασαφών μέτρων με ασαφές λειτουργίες, σιγουρεύεται ότι όλες οι πιθανές ιδιότητες σε όλους τους συνδυασμούς που μπορούν να θεωρηθούν, χρησιμοποιούνται. Με τέτοιο τρόπο μια εικόνα προκύπτει όπου οι γκρι τιμές αποτελούν ένα μέτρο για την ένταξη στην υφή. Με τον τρόπο αυτό διαφορετικές στοχαστικές ιδιότητες συνδυάζονται με πολλούς τρόπους για την εξαγωγή των σχετικών πληροφοριών.

Στον παρακάτω πίνακα γίνεται σύγκριση των διαφορετικών αλγορίθμων εκπαίδευσης μηχανών που αναπτύχθηκαν παραπάνω [28]

Πίνακας 2:Συνολική Παρουσίαση των Αλγορίθμων για ταξινόμηση εικόνων

ΑΛΓΟΡΙΘΜΟΣ	ΠΛΕΟΝΕΚΤΗΜΑΤΑ	ΠΕΡΙΟΡΙΣΜΟΙ
ΝΕΥΡΩΝΙΚΑ ΔΙΚΤΥΑ	<ul style="list-style-type: none"> Μπορεί να χρησιμοποιηθεί για ταξινόμηση ή παλινδρόμηση Ικανός να αναπαραστήσει Boolean συναρτήσεις (AND,OR,NOT) Ανοχή στους θορύβους στα δεδομένα εισόδου Παραδείγματα μπορούν να ταξινομηθούν με περισσότερα από μια έξοδο. 	<ul style="list-style-type: none"> Δυσκολία κατανόησης της δομής του αλγορίθμου Η βέλτιστη δομή του δικτύου μπορεί να καθοριστεί μόνο με τη διεξαγωγή πειραμάτων
ΜΗΧΑΝΕΣ ΥΠΟΣΗΡΙΞΗΣ ΔΙΑΝΥΣΜΑΤΩΝ	<ul style="list-style-type: none"> Μοντελοποιεί μη γραμμικά όρια κλάσεων Η υπερτοποθέτηση είναι απίθανο να συμβεί Η υπολογιστική πολυπλοκότητα μειώνεται στο πρόβλημα τετραγωνικής βελτιστοποίησης Εύκολη στον έλεγχο πολυπλοκότητα του κανόνα αποφάσεως και συχνότητας λαθών 	<ul style="list-style-type: none"> Η εκμάθηση είναι αργή σε σχέση με τα Δένδρα απόφασης Δυσκολία καθορισμού βέλτιστων παραμέτρων όταν τα δεδομένα εκμάθησης δεν είναι γραμμικώς ανεξάρτητα
ΑΣΑΦΗ ΜΕΤΡΑ	<ul style="list-style-type: none"> Διαφορετικές στοχαστικές θέσεις μπορούν να περιγράψουν ιδιότητες 	<ul style="list-style-type: none"> Εκ των προτέρων γνώση είναι πολύ σημαντική για καλά αποτελέσματα Ακριβείς λύσεις δεν προκύπτουν εάν η κατεύθυνση της απόφασης δεν είναι σαφής
ΓΕΝΕΤΙΚΟΣ ΑΛΓΟΡΙΘΜΟΣ	<ul style="list-style-type: none"> Μπορεί να χρησιμοποιηθεί στην ταξινόμηση και επιλογή χαρακτηριστικών Χρησιμοποιείται κυρίως για βελτιστοποίηση, βρίσκει πάντα μια καλή λύση Μπορεί να χειριστεί μεγάλο, σύνθετο, μη διαχωρίσιμο και πολυτροπικό χώρο 	<ul style="list-style-type: none"> Ο υπολογισμός ή η ανάπτυξη της βαθμολογικής λειτουργίας δεν είναι τετριμμένη Δεν είναι η αποτελεσματικότερη μέθοδος για την εύρεση της βέλτιστης λύσης

3.4 Αλγόριθμος k-κοντινότερου γείτονα

Ο αλγόριθμος kNN (k nearest neighbor = k κοντινότερου γείτονα) [28] είναι επίσης ένας αλγόριθμος που χρησιμοποιείται για την ταξινόμηση εικόνων. Είναι ένας αλγόριθμος μάθησης βασισμένος σε στιγμιότυπα (instance-based), δηλαδή η διαδικασία μάθησης αφορά απλά την αποθήκευση των δεδομένων εκπαίδευσης και θεωρείται ένας από τους πιο απλούς αλγόριθμους μάθησης. Θεωρείται εκπληκτικά ευέλικτος αλγόριθμος και οι εφαρμογές του εκτείνονται από υπολογιστική όραση σε υπολογιστική γεωμετρία σε γράφους και ούτω καθεξής. Τα δεδομένα εκπαίδευσης τυγχάνουν επεξεργασίας όταν εμφανιστεί ένα νέο instance για αυτό και ονομάζεται Lazy Learning [29]. Κάθε φορά που ένα νέο instance πρόκειται να ταξινομηθεί, υπολογίζεται η ομοιότητα του με κάθε ένα από τα αποθηκευμένα δεδομένα εκπαίδευσης. Αυτό σημαίνει ότι η φάση της εκπαίδευσης είναι αρκετά γρήγορη. Η έλλειψη γενίκευσης σημαίνει ότι ο αλγόριθμος διατηρεί όλα τα δεδομένα εκπαίδευσης. Αυτό έρχεται σε αντίθεση με άλλες τεχνικές, όπως Μηχανές Υποστήριξης Διανυσμάτων (SVM) [25] η οποία μπορεί να απορρίψει όλα τα μη υποστηριζόμενα διανύσματα χωρίς πρόβλημα. Επίσης χαρακτηρίζεται ως μη παραμετρικός αλγόριθμος [30]. Αυτό πρακτικά σημαίνει ότι ο αλγόριθμος δεν κάνει υποθέσεις σχετικά με το υποκείμενο της διανομής δεδομένων. Αυτό είναι αρκετά χρήσιμο στον πραγματικό κόσμο καθώς τα περισσότερα από τα πρακτικά δεδομένα δεν υπακούουν τις τυπικές θεωρητικές παραδοχές που έχουν γίνει.



Εικόνα 36: Αλγόριθμος βασισμένος σε στιγμιότυπα

3.4.1 Υποθέσεις του KNN αλγορίθμου

Ο αλγόριθμος KNN βασίζεται σε μια συνάρτηση απόστασης όπως είναι η Ευκλείδεια απόσταση και η απόσταση συννημίτονου, μεταξύ κάθε στοιχείου εκπαίδευσης και του στοιχείου που πρόκειται να ταξινομηθεί. Η τεχνική KNN προϋποθέτει επίσης ότι το σύνολο εκπαίδευσης δεν περιλαμβάνει μόνο τα δεδομένα αλλά επίσης και την επιθυμητή κατηγοριοποίηση για κάθε στοιχείο. Αυτό έχει σαν αποτέλεσμα τα δεδομένα εκπαίδευσης να αποτελούν το μοντέλο κατηγοριοποίησης όταν γίνεται κατηγοριοποίηση ενός νέου στοιχείου, πρέπει να καθοριστεί η απόσταση του από το στοιχείο του συνόλου εκπαίδευσης.

Κάθε ένα από τα στοιχεία κατάρτισης αποτελείται από ένα σύνολο διανυσμάτων και μια ετικέτα κλάσης που συνδέεται με κάθε διάνυσμα. Στην απλούστερη περίπτωση θα είναι είτε + είτε – για θετικές και αρνητικές τάξεις. Ωστόσο ο αλγόριθμος μπορεί να λειτουργήσει εξίσου καλά και με αυθαίρετο αριθμό τάξεων.

Δίνεται επίσης ένας μονός αριθμός "k", ο οποίος αποφασίζει πόσοι γείτονες (ο γείτονας ορίζεται βάση της μετρικής απόστασης) επηρεάζουν την ταξινόμηση. Αυτός ο αριθμός είναι συνήθως μονός όταν ο αριθμός των κλάσεων είναι δυο. Εάν το k είναι 1, τότε ο αλγόριθμος καλείται απλά αλγόριθμος του πλησιέστερου γείτονα.

3.4.2 Ταξινόμηση εικόνων με χρήση του αλγορίθμου k-κοντινότερου γείτονα

Στη συνέχεια θα δούμε πως μπορούμε να χρησιμοποιήσουμε τον KNN αλγόριθμο για ταξινόμηση [31]. Στην περίπτωση αυτή, δίνονται κάποια σημεία δεδομένων για την εκπαίδευση και επίσης νέα δεδομένα χωρίς να χαρακτηρίζονται από κάποια ετικέτα. Στόχος είναι να βρεθεί η κατάλληλη ετικέτα κλάσης για το νέο σημείο. Ο αλγόριθμος έχει διαφορετική συμπεριφορά ανάλογα με τη σταθερά k.

3.4.2.1 Πρώτη Περίπτωση για k=1

Πρόκειται για το απλούστερο σενάριο. Έστω x το σημείο που πρέπει να φέρει την ένδειξη. Βρίσκει το σημείο που βρίσκεται πλησιέστερα του x, ας πούμε το y. Τώρα ο κανόνας του πλησιέστερου γείτονα ζητά την ανάθεση της ετικέτας του y στο x. Αυτό φαίνεται πολύ απλοϊκό. Στο σημείο αυτό υπάρχει η αίσθηση ότι αυτή η διαδικασία θα έχει ως αποτέλεσμα ένα τεράστιο λάθος. Αυτή η αίσθηση ανταποκρίνεται στην πραγματικότητα αλλά μόνο όταν ο αριθμός των δεδομένων σημείων δεν είναι πολύ μεγάλος. Εάν ο αριθμός αυτός είναι πολύ μεγάλος, τότε υπάρχει μια μεγάλη πιθανότητα η ετικέτα του x και του y να είναι οι ίδιες.

Ας υποθέσουμε ότι όλα τα σημεία βρίσκονται σε ένα πεδίο D διαστάσεων. Ο αριθμός των σημείων είναι αρκετά μεγάλος. Αυτό σημαίνει ότι η πυκνότητα του επιπέδου σε οποιοδήποτε σημείο θα είναι μεγάλη. Με άλλα λόγια σε κάθε υπό-χώρο υπάρχει επαρκής αριθμός σημείων. Θεωρούμε ένα σημείο x στον υπό-χώρο το οποίο έχει πολλούς γείτονες, με y τον κοντινότερο από αυτούς. Εάν τα x και y είναι αρκετά κοντά μεταξύ τους, τότε μπορούμε να υποθέσουμε ότι η πιθανότητα το x και το y να ανήκουν στην ίδια κλάση είναι περίπου η ίδια. Έπειτα από την θεωρία των αποφάσεων τα x και τα y έχουν την ίδια κλάση.

3.4.2.2 Δεύτερη Περίπτωση για k=K

Αυτό που συμβαίνει σε αυτή την περίπτωση είναι η προσπάθεια αναζήτησης του k κοντινού γείτονα και γίνεται μια ψηφοφορία με ειδική πλειοψηφία. Εάν για παράδειγμα, k=5 και υπάρχουν 3 στιγμιότυπα του C1 και 2 στιγμιότυπα του C2, σε αυτή την περίπτωση ο αλγόριθμος υποστηρίζει ότι το νέο σημείο πρέπει να επισημανθεί ως C1 καθώς αποτελεί την πλειοψηφία.

Μια επέκταση βέβαια είναι να μην πάρουν όλοι οι γείτονες 1 ψήφο. Μια πολύ συχνή εναλλαγή είναι η εφαρμογή σταθμισμένου KNN, όπου κάθε σημείο έχει ένα βάρος το οποίο συνήθως υπολογίζεται βάση της απόστασης του. Για παράδειγμα αν γίνει στάθμιση σύμφωνα με την αντίστροφη απόσταση, θα καταλήξει στο συμπέρασμα ότι οι γειτονικές μονάδες έχουν υψηλότερη ψήφο από τα πιο απομακρυσμένα σημεία. Είναι προφανές ότι η ακρίβεια μπορεί να αυξηθεί με την αύξηση του k αλλά ταυτόχρονα αυξάνεται και το κόστος του υπολογισμού.

3.4.3 Βασικές Παρατηρήσεις πάνω στον Αλγόριθμο KNN

- Εάν υποθέσουμε ότι τα σημεία είναι d-διαστάσεων, τότε η άμεση εφαρμογή της εύρεσης του k πλησιέστερου γείτονα παίρνει $O(dn)$ χρόνο
- Μπορούμε να σκεφτούμε τον αλγόριθμο με δυο τρόπους. Ο πρώτος είναι ότι ο KNN προσπαθεί να εκτιμήσει την πιθανότητα επισήμανσης του σημείου. Ο εναλλακτικός τρόπος είναι ότι υπολογίζει την επιφάνεια της απόφασης (άμεσα ή έμμεσα) και στη συνέχεια την χρησιμοποιεί για να αποφασίσει για την κλάση των νέων σημείων.
- Ακόμη και αν αυτή η μέθοδος παίρνει $O(dn)$ χρόνο είναι πολύ δύσκολο να γίνει κάτι καλύτερο, εκτός και αν γίνουν κάποιες άλλες υποθέσεις. Υπάρχουν μερικές αποτελεσματικές δομές δεδομένων όπως τα KD-δένδρα που μπορούν να μειώσουν την πολυπλοκότητα του χρόνου, αλλά το πετυχαίνουν εις βάρος του χρόνου εκπαίδευσης και της πολυπλοκότητας.
- Πρέπει να υπογραμμιστεί στο σημείο αυτό ότι η τεχνική KNN είναι υπερβολικά ευαίσθητη στην τιμή $\hat{\epsilon}$ δηλαδή στο πόσοι κοντινότεροι γείτονες χρησιμοποιούνται για την κατηγοριοποίηση. Σύμφωνα με μια εμπειρική μέθοδο πρέπει να ισχύει ότι :

$$\hat{\epsilon} \leq \sqrt{\text{αριθμός_στοιχείων_εκπαίδευσης}}$$

3.4.4 Πλεονεκτήματα & Μειονεκτήματα του Αλγορίθμου KNN

Το πλεονέκτημα του αλγορίθμου αυτού είναι ότι είναι εύκολο να κατανοηθεί και να υλοποιηθεί. Επίσης, έχει αποδειχθεί ότι το λάθος του kNN περιορίζεται στο μισό του λάθους του Bayes [32] κάτω από συγκεκριμένες υποθέσεις. Ο αλγόριθμος kNN δουλεύει καλά σε περιπτώσεις multi-modal κλάσεων και σε εφαρμογές όπου κάποιο αντικείμενο μπορεί να ανήκει σε περισσότερο από μια κλάση. Επίσης, είναι πολύ αποδοτικός σε περιπτώσεις όπου τα δεδομένα εκπαίδευσης περιέχουν θόρυβο (noisy) και σε περιπτώσεις όπου τα δεδομένα εκπαίδευσης είναι πολλά.

Το σημαντικότερο μειονέκτημα του είναι ότι είναι instance-based αλγόριθμος μάθησης, οπότε δεν γίνεται οποιαδήποτε εκπαίδευση, μέχρι να φτάσει κάποιο έγγραφο για ταξινόμηση και επίσης έχει μεγάλο κόστος υπολογισμού γιατί πρέπει να υπολογίσει την απόσταση κάθε όρου του κειμένου που θα ταξινομηθεί με όλα τα έγγραφα εκπαίδευσης. Ένα άλλο σημαντικό μειονέκτημα είναι ότι χρειάζεται να καθοριστεί κάποια τιμή για το k. Το k παίζει αρκετά σημαντικό ρόλο στην αποδοτικότητα του ταξινομητή και είναι δύσκολο να προσδιοριστεί. Αν είναι πολύ μικρό, το αποτέλεσμα μπορεί να είναι ευαίσθητο σε θορυβώδη δεδομένα. Αν είναι πολύ μεγάλο, το αποτέλεσμα των κοντινότερων γειτόνων μπορεί να περιέχει πολλά έγγραφα από άλλες κατηγορίες. Το k μπορεί να οριστεί χρησιμοποιώντας διάφορες τεχνικές οι οποίες χρησιμοποιούν ευρετικά. Τέλος, πρέπει να καθοριστεί μια συνάρτηση απόστασης πρέπει να εφαρμοστεί για να προκύψουν τα καλύτερα αποτελέσματα

4. Αναγνώριση Αντικειμένων

Στο κεφάλαιο αυτό θα γίνει μια προσπάθεια για την κατανόηση του όρου αναγνώριση αντικειμένων σε μια εικόνα, καθώς και μια σύντομη επισκόπηση τέτοιων τεχνικών. Η αναγνώριση αντικειμένων στα πλαίσια μια εικόνας είναι ο απώτερος σκοπός της εφαρμογής που υλοποιήθηκε σε αυτή την εργασία και θα παρουσιαστεί σε επόμενο κεφάλαιο.

4.1 Εισαγωγή στην έννοια της Αναγνώρισης Αντικειμένων

Τα δυο κύρια προβλήματα της όρασης υψηλού επιπέδου είναι: η ανίχνευση αντικειμένων (object detection) [33], στόχος της οποίας είναι να ληφθεί η απόφαση για το αν υπάρχει ένα αντικείμενο σε μια εικόνα, και η αναγνώριση αντικειμένων (object recognition) [34] στόχος της οποίας είναι η λήψη μιας απόφασης βάση παρατηρήσεων για ένα από τα πολλαπλά αντικείμενα της εικόνας. Επιπλέον με τον όρο αναγνώριση αντικειμένων, συνήθως εννοείται και όλο το φάσμα των προβλημάτων της όρασης υψηλού επιπέδου. Δυο συμπληρωματικές πτυχές του προβλήματος που εισάγουν επιπλέον απαιτήσεις είναι : α) η χρήση κατηγοριών αντικειμένων αντί συγκεκριμένων αντικειμένων, όπως αναφέρθηκε στο [Κεφάλαιο 3](#), οπότε πρέπει να ληφθεί υπόψη η ποικιλία που υπάρχει μέσα στην ίδια κατηγορία β) ο εντοπισμός αντικειμένων (object localization), όπου στόχος είναι ,πέραν της λήψης απόφαση περί κατηγοριοποίησης/ύπαρξης του αντικειμένου, να βρεθεί στην εικόνα η τοποθεσία του.

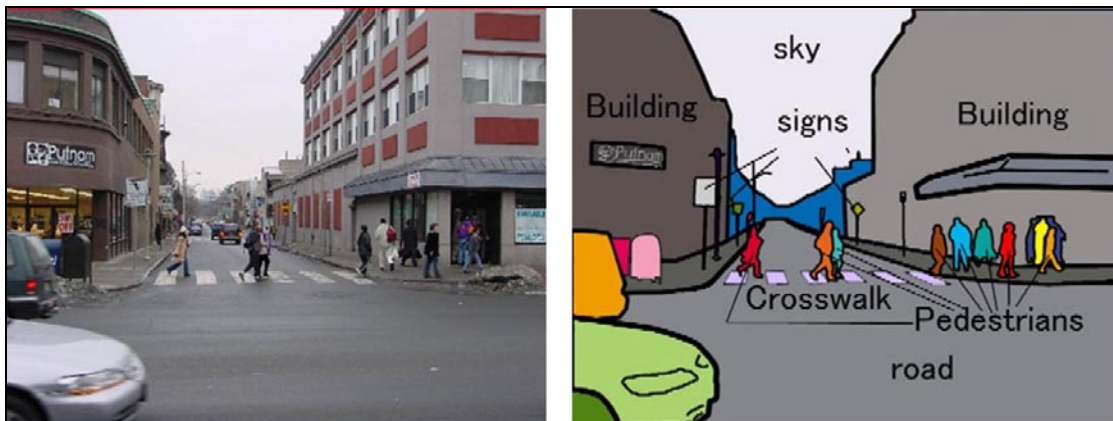
Γενικότερα αυτό που έχει κανείς κατά νου όταν αναφέρεται στον όρο "αναγνώριση αντικειμένων" μπορεί να εκφραστεί με βάση τους παραπάνω όρους ως ο εντοπισμός των θεωρούμενων κατηγοριών σε μια εικόνα και η αναγνώριση των μεμονωμένων αντικειμένων των κατηγοριών στις οποίες εστιάζουμε.

Στην αναγνώριση αντικειμένων εικόνας, η δυσκολία ενός προβλήματος είναι αντιστρόφως ανάλογη της εξειδίκευσης του. Για παράδειγμα, ενώ χρησιμοποιώντας απλοϊκές μεθόδους, ο εντοπισμός ελαττωματικών εργοστασιακών εξαρτημάτων θεωρείται εφικτός, η δημιουργία μεθόδων για να αποφαίνονται για την ύπαρξη ανθρώπου σε μια εικόνα θεωρείται μεγάλη πρόκληση ακόμη και τώρα. Οι παράγοντες που δυσχεραίνουν την επίλυση του προβλήματος συνοψίζονται στους παρακάτω:

- Ποικιλία της εμφάνισης εντός της ίδιας κατηγορίας
- Μεταβολή του φωτισμού
- Εσφαλμένες ανιχνεύσεις λόγω περίπλοκης δομής (clutter)
- Αλληλοεπικαλύψεις με διαφορετικά αντικείμενα
- Πολλαπλές κλίμακες/κατευθύνσεις
- Αλλαγές της εμφάνισης λόγω περιστροφής του αντικειμένου

Ο κύριος όγκος της έρευνας στην αναγνώριση αντικειμένων επικεντρώνεται στα τέσσερα πρώτα προβλήματα, θεωρώντας ότι τα δυο τελευταία μπορούν να αντιμετωπιστούν μέσω εξαντλητικής αναζήτησης και τη χρήση ξεχωριστών μοντέλων για διαφορετικές όψεις, κάτι που συνεπάγεται και μεγαλύτερες υπολογιστικές απαιτήσεις.

Στην παρακάτω εικόνα, φαίνεται ένα παράδειγμα χρήσης της αναγνώρισης αντικειμένων σε πραγματικές συνθήκες. Αριστερά φαίνεται η αρχική εικόνα ενώ δεξιά βλέπουμε την εικόνα η οποία περιλαμβάνει και ενδείξεις για το κάθε αντικείμενο το οποίο έχει αναγνωριστεί επιτυχώς.



Εικόνα 37: Παράδειγμα Αναγνώρισης Αντικειμένων σε εικόνα

4.2 Επισκόπηση Τεχνικών Αναγνώρισης Αντικειμένων

Τις δυο πρώτες δεκαετίες της έρευνας πάνω στο πεδίο της αναγνώρισης αντικειμένων [35], υπήρχαν δυο κυρίαρχα ρεύματα, τα άνωθεν και κάτωθεν μοντέλα τα οποία αντιστοιχούν στην άνωθεν και κάτωθεν προσέγγιση της όρασης. Τα κάτωθεν μοντέλα θεωρούν ότι μια εικόνα μπορεί να αναλυθεί σε πρωτογενή σχήματα (κύλινδροι, σφαίρες) και αναζητούν τρόπους για τη σύνδεση των ανιχνευόμενων σχημάτων μέσω περιπλοκότερων μοντέλων για αντικείμενα. Τα άνωθεν μοντέλα χρησιμοποιώντας τρισδιάστατες αναπαραστάσεις των αντικειμένων, ανάγουν το πρόβλημα της αναγνώρισης στην εκτίμηση των παραμέτρων της προβολής του μοντέλου στην εικόνα.

Στις αρχές τις προηγούμενης δεκαετίας, καθώς οι περιορισμοί των παραπάνω προσεγγίσεων έγιναν αντιληπτοί, ξεκίνησαν οι πρώτες προσπάθειες αφενός για να ληφθεί υπόψη η ποικιλία της εμφάνισης που ενδέχεται να παρουσιάζουν τα αντικείμενα μιας κατηγορίας και αφετέρου για την κατασκευή τέτοιων μοντέλων με μικρότερη ανθρώπινη παρέμβαση. Σύντομα η αναγνώριση αντικειμένων στράφηκε προς τα στατιστικά μοντέλα για αναπαράσταση αντικειμένων, αντίστοιχα αυτών που έχουν επικρατήσει στην τεχνολογία φωνής και τεχνικές εκμάθησης μηχανών για την αυτόματη κατασκευή τους από δεδομένα εκπαίδευσης.

Οι τεχνικές αναγνώρισης αντικειμένων θα μπορούσαν να ταξινομηθούν βάση του εύρους του προβλήματος που καλύπτουν, των μοντέλων των αντικειμένων που χρησιμοποιούν και των μαθηματικών/αλγοριθμικών τεχνικών στις οποίες βασίζονται.

4.2.1 Αναγνώριση μεμονωμένων αντικειμένων μέσω μοντέλων και αντιστοίχισης

Βάση αυτής της προσέγγισης, τα αντικείμενα που βρίσκονται σε μια δισδιάστατη εικόνα αναγνωρίζονται χρησιμοποιώντας αποθηκευμένες αναπαραστάσεις τρισδιάστατων αντικειμένων ή συνδυασμούς δισδιάστατων όψεων τους. Οι βασικές συνιστώσες του προβλήματος είναι η κατασκευή των αναπαραστάσεων, η παραγωγή ενός περιορισμένου πλήθους υποθέσεων δεδομένης μιας εικόνας και η εκτίμηση της πιστότητας των παρατηρήσεων στις προβλέψεις του μοντέλου.

Για την αναζήτηση της θέσης του αντικειμένου μπορούν να χρησιμοποιηθούν άνωθεν τεχνικές ή κάτωθεν τεχνικές όπως ο μετασχηματισμός Hough [36]. Έχοντας κάποιες υποψήφιες τοποθεσίες, η εκτίμηση της πιστότητας των παρατηρήσεων επιτυγχάνεται βάσει κάποιου κατάλληλου μέτρου της απόστασης μεταξύ των χαρακτηριστικών των παρατηρήσεων και του μοντέλου.

Αν και οι προσεγγίσεις αυτές έχουν δώσει ικανοποιητικά αποτελέσματα σε πρακτικές εφαρμογές (π.χ. επισκόπηση εξαρτημάτων σε βιομηχανικές γραμμές συναρμολόγησης, ανάκληση αντικειμένων από μία βάση δεδομένων), το πεδίο εφαρμογών τους είναι περιορισμένο σε συγκεκριμένα αντικείμενα αντί για κατηγορίες αντικειμένων. Όμως, η εμπειρία που έχει συσσωρευτεί έχει αποδειχθεί ωφέλιμη για την διατύπωση και επίλυση νέων προβλημάτων που προκύπτουν σε σύγχρονες τεχνικές αναγνώρισης αντικειμένων.



Εικόνα 38: Αναγνώριση αντικειμένων μέσω αντιστοίχισης μοντέλων

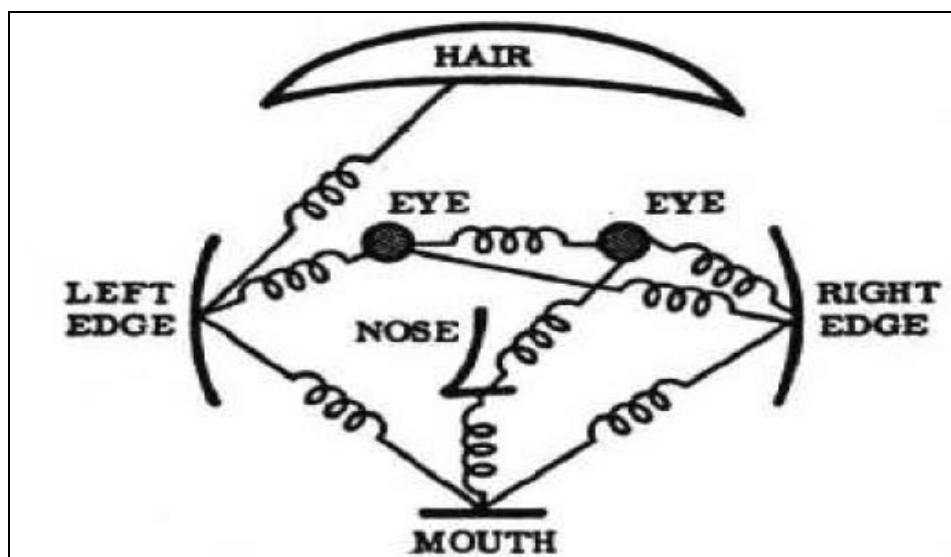
4.2.2 Αναγνώριση αντικειμένων με τεχνικές αναγνώρισης προτύπων & νευρωνικών δικτύων

Οι τεχνικές αυτές ξεκίνησαν από την έρευνα στο χώρο των νευρωνικών δικτύων, όπου το πρόβλημα της ανίχνευσης αντιμετωπίζεται μέσω της παράλληλης κατανεμημένης επεξεργασίας της πληροφορίας. Ξεκινώντας από την εξαγωγή ενός συνόλου χαρακτηριστικών, τα οποία είναι αναλλοίωτα ως προς ανεπιθύμητες πηγές ποικιλίας στην εμφάνιση, όπως ο φωτισμός, η ανίχνευση ενός αντικειμένου, εκφράζεται ως ένα πρόβλημα αναγνώρισης προτύπων. Με άλλα λόγια δεδομένου ενός συνόλου χαρακτηριστικών μέσα στο εξεταζόμενο παράθυρο, ζητείται να ληφθεί η απόφαση για το εάν υπάρχει ένα αντικείμενο. Τα προβλήματα που καλούνται να αντιμετωπίσουν τέτοιες τεχνικές, αφορούν την εξαγωγή κατάλληλων χαρακτηριστικών για τα αντικείμενα, την αποδοτική υλοποίηση των αλγορίθμων ανίχνευσης καθώς και όλα τα προβλήματα του πεδίου της αναγνώρισης προτύπων.

Από τα βασικά πλεονεκτήματα των μεθόδων αυτών θεωρούνται η ωριμότητα των διαθέσιμων τεχνικών από την περιοχή της μηχανικής μάθησης (SVM) καθώς και η ταχύτητα με την οποία μπορεί να γίνει η διαδικασία της ανίχνευσης. Ωστόσο υπάρχουν δυο κύρια προβλήματα. Το βασικότερο είναι ότι για την εκμάθηση τους χρειάζεται συνήθως ένα σύνολο εκπαίδευσης της τάξης των μερικών χιλιάδων εικόνων, καθώς περιλαμβάνεται ένα μεγάλο πλήθος παραμέτρων που πρέπει να εκμαθευτούν. Αντίθετα εμείς ως άνθρωποι αρκούμαστε σε μια εικόνα για να επιτύχουμε ικανοποιητικά αποτελέσματα. Επίσης μια τέτοια προσέγγιση δεν είναι εύκολο να αντιμετωπίσει επικαλύψεις με άλλα αντικείμενα και γενικότερα προβλήματα που δεν συμπεριλαμβάνονται στο σύνολο εκπαίδευσης.

4.2.3 Αναγνώριση αντικειμένων με τμηματικές αναπαραστάσεις

Τα μοντέλα που ανήκουν στην κατηγορία των τμηματικών ταξινομητών βασίζονται στην ανίχνευση των αντικειμένων τους στην ανίχνευση χαρακτηριστικών σημείων (keypoints) των αντικειμένων. Θεωρείται εύκολο να χρησιμοποιηθεί στατιστική πληροφορία για τη χωρική διάταξη των τμημάτων, ώστε να εισαχθεί και γεωμετρική γνώση στη διαδικασία ανίχνευσης. Η προσέγγιση αυτή συνδυάζει την αποδοτικότητα του κάτωθεν μοντέλου της όρασης, στο οποίο από επιμέρους τμήματα αναγνωρίζονται περιπλοκότερα αντικείμενα, με τους περιορισμούς που εισάγει η πρότερη γνώση των άνωθεν μοντέλων απλοποιώντας την αναζήτηση σε συνδυασμούς που θα μπορούσαν να έχουν προκύψει από το αντικείμενο. Το βασικό της μειονέκτημα είναι ότι αντί για το αντικείμενο καθεαυτό μοντελοποιούνται τα τμήματα του και οι αλληλεξαρτήσεις τους που θεωρούνται χρήσιμες για την ανίχνευση τους, παρέχοντας έτσι ένα μοντέλο περιορισμένων δυνατοτήτων για άλλες λειτουργίες, όπως για παράδειγμα την αναγνώριση της ταυτότητας ενός αντικειμένου.

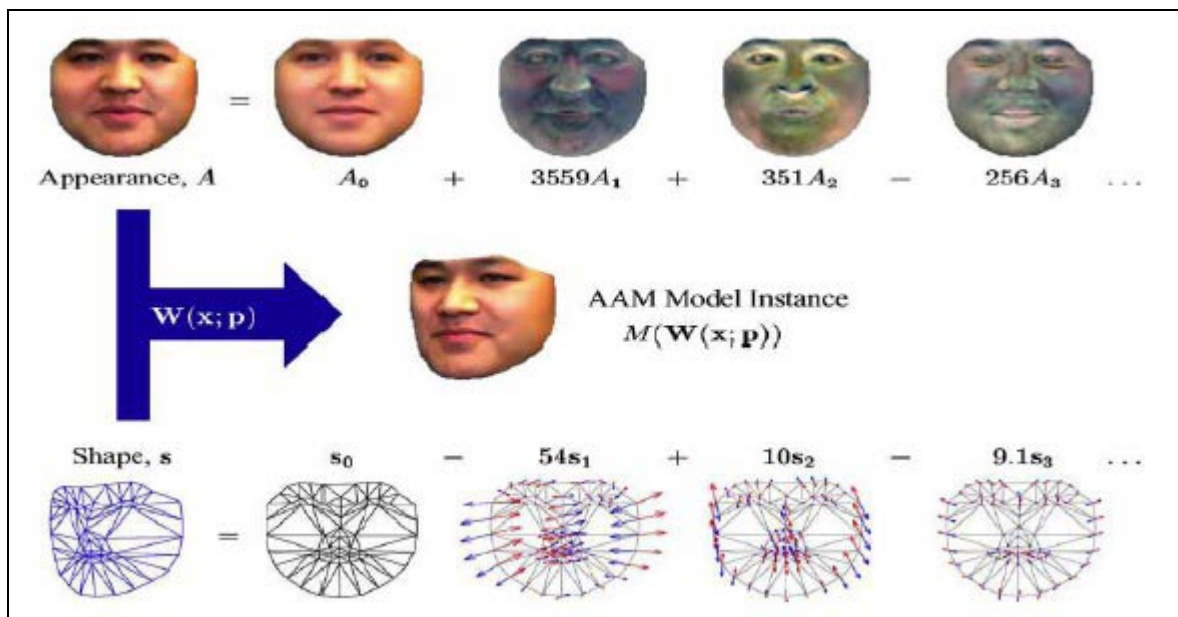


Εικόνα 39: Αναγνώριση Αντικειμένων μέσω τμηματικών αναπαραστάσεων

4.2.4 Αναγνώριση αντικειμένων με παραμορφώσιμα μοντέλα

Αυτή η προσέγγιση βασίζεται στην χρήση ενός παραμετρικού μοντέλου των παραμορφώσεων της εμφάνισης και γεωμετρίας ενός πρωτότυπου αντικειμένου, το οποίο καλύπτει όλη την ποικιλία που χαρακτηρίζει μια κατηγορία αντικειμένων. Δεδομένης μίας νέας εικόνας και μίας αρχικοποίησης, το πρόβλημα της αναγνώρισης/ανίχνευσης ανάγεται στην μεταβολή των παραμέτρων του μοντέλου ώστε αυτό να αναπαράγει τις παρατηρήσεις. Οι δύο βασικές συνιστώσες της προσέγγισης αυτής είναι η μοντελοποίηση της παραμόρφωσης και η εκτίμηση των παραμέτρων των μοντέλων. Ως πρώτες προσπάθειες στην κατεύθυνση της μοντελοποίησης της παραμόρφωσης του σχήματος μπορούν να θεωρηθούν αφενός τεχνικές όπως τα παραμορφώσιμα πρότυπα (Deformable Templates) αφετέρου η αναγνώριση μέσω συνδυασμού των όψεων.

Τα μοντέλα αυτά θεωρούν ότι οι παρατηρήσεις που προέρχονται από το αντικείμενο μπορούν να μοντελοποιηθούν από την χωρική παραμόρφωση του αντικειμένου ακολουθούμενη από την παραμόρφωση της φωτεινότητας του. Το βασικό πρόβλημα με αυτά τα μοντέλα είναι ότι καθώς χρησιμοποιείται ένας επαναληπτικός αλγόριθμος δεν είναι εφικτό να ερευνηθεί κανείς για όλες τις θέσεις, κλίμακες τις εικόνες. Συνεπώς τέτοιες μέθοδοι δεν είναι κατάλληλες για την εξαρχής ανίχνευσης αντικειμένων, αλλά μπορούν να χρησιμοποιηθούν για τη λήψη της τελικής απόφασης. Για αυτό χρησιμοποιούνται κυρίως για αναγνώριση αντικειμένων θεωρώντας γνωστό ότι στη θέση που αρχικοποιούνται υπάρχει ένα αντικείμενο κάποιας συγκεκριμένης κατηγορίας. Μια άλλη εφαρμογή τους είναι η καταγραφή της εξέλιξης αντικειμένων (tracking), κατά την οποία χρησιμοποιείται η πρόβλεψη βάσει της προηγούμενης θέσης του μοντέλου για να ξεκινήσει η αναζήτηση της τρέχουσας.



Εικόνα 40: Αναγνώριση μέσω παραμορφώσιμων μοντέλων των αντικειμένων

5. Κύριο μέρος Πτυχιακής Εργασίας

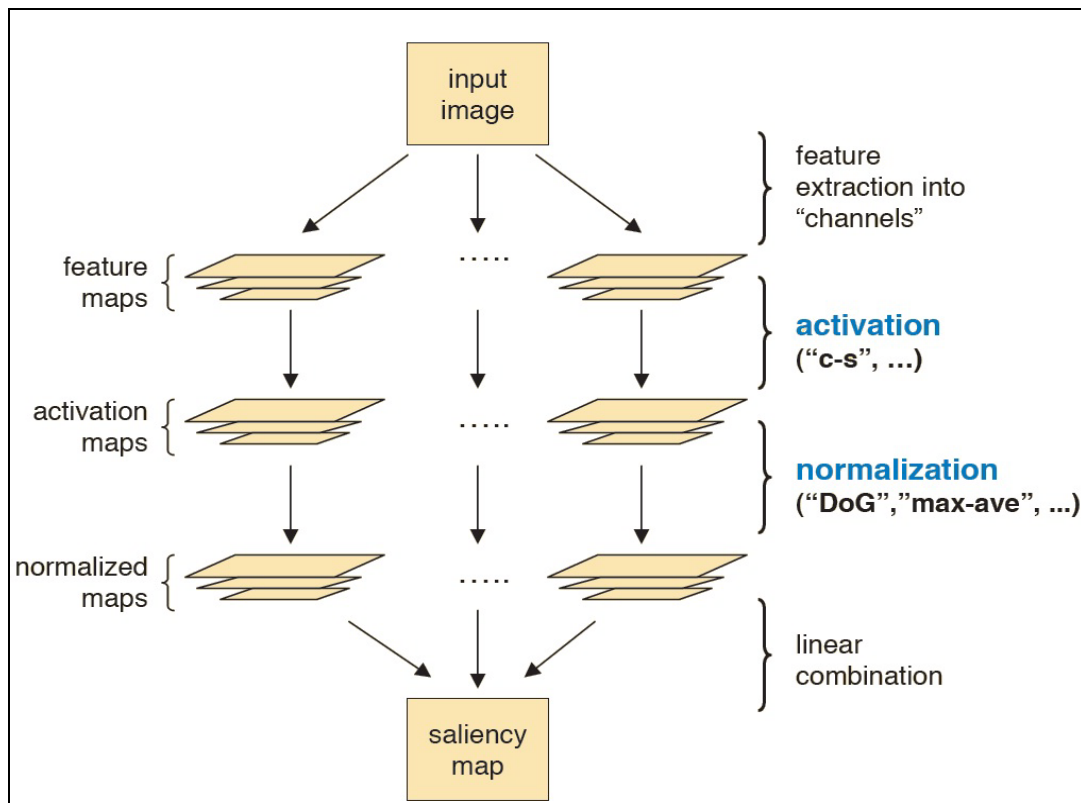
Σε αυτό το κεφάλαιο, περιγράφεται αναλυτικά η ένωση των παραπάνω κεφαλαίων για την υλοποίηση της τελικής εφαρμογής. Πιο συγκεκριμένα, χρησιμοποιήθηκαν κάποιες τεχνικές από τα μοντέλα τα οποία αναφέρθηκαν παραπάνω, σε ένα κοινό σύστημα στοχεύοντας στην αυτόματη αναγνώριση αντικειμένων εικόνας. Στο κεφάλαιο αυτό θα γίνει μια πολύ σύντομη αναφορά στην τεχνική "Προεξέχοντα Οπτικά Σημεία Βασισμένα σε Γράφους" η οποία ανήκει στην επιστημονική περιοχή των [Προεξέχοντων Οπτικών Σημείων Εικόνας](#) η οποία αναφέρθηκε σε προηγούμενο κεφάλαιο. Στη συνέχεια του κεφαλαίου, θα αναπτυχθούν τα θέματα της διαδικασίας που ακολουθήθηκε βήμα προς βήμα, τα ενδιάμεσα αποτελέσματα σε κάθε στάδιο, τα τελικά αποτελέσματα και πως αυτά ερμηνεύονται καθώς και η συμβολή της τελικής εφαρμογής στην επιστημονική κοινότητα. Τέλος θα γίνει αναφορά σε ιδέες οι οποίες ενδέχεται να βελτιώσουν την αποτελεσματικότητα και τη λειτουργικότητα του συστήματος και πρόκειται να εξεταστούν μελλοντικά.

5.1 Προεξέχοντα Οπτικά Σημεία Βασισμένα σε Γράφους (Graph Based Visual Saliency)

Η προεξοχή των οπτικών σημείων πρέπει να οριστεί ως η διακριτικότητα των χαρακτηριστικών σε μια εικόνα. Οι χάρτες προεξέχοντων σημείων περιέχουν πληροφορίες σχετικά με το που μπορούν να βρεθούν ενδιαφέρουσες πληροφορίες στα πλαίσια μιας εικόνας. Αυτές οι περιοχές αντιστοιχούν σε χαρακτηριστικά και θεωρούνται ως σπάνιες ή ενημερωτικές ανάλογα με τον ορισμό της προεξοχής. Συγκεκριμένα, περιοχές υψηλού επιπέδου προεξοχής, αντιστοιχούν σε αντικείμενα η περιοχές που είναι πιθανότερο να βρεθούν, ενώ οι χαμηλότερες προεξοχές συνδέονται με το φόντο μιας εικόνας. Στο πλαίσιο αυτό προτείνεται αυτή η τεχνική GBVS [37].

Αποτελείται από δυο στάδια: πρώτα σχηματίζονται οι χάρτες ενεργοποίησης σε συγκεκριμένα κανάλια χαρακτηριστικών και στη συνέχεια γίνεται η κανονικοποίηση τους με ένα τρόπο που αναδεικνύει την ευδιακριτότητα και δέχεται συνδυασμό με άλλους χάρτες. Το μοντέλο θεωρείται αρκετά απλό και βιολογικά εφικτό στο βαθμό που παραλληλοποιείται φυσικά. Με άλλα λόγια, δεδομένης μιας εικόνας θέλουμε να επισημάνουμε μια σειρά από σημαντικές τοποθεσίες, στις οποίες η εικόνα θεωρείται ενημερωτική σύμφωνα με κάποιο κριτήριο. Η διαδικασία αυτή προϋποθέτει αρχικά τον υπολογισμό των χαρτών οι οποίοι αποτελούνται από χαρακτηριστικά, για παράδειγμα με γραμμικό φιλτράρισμα ακολουθημένο από κάποια στοιχεία γραμμικότητας.

Παρακάτω θα γίνει αναλυτικότερη παρουσίαση της συγκεκριμένης μεθόδου, η οποία χρησιμοποιήθηκε στην τελική εφαρμογή, καθώς και των διαφορών της με τις τυπικές μεθόδους που υπάρχουν ήδη για την παραγωγή χαρτών. Στην παρακάτω εικόνα παρουσιάζονται τα κύρια βήματα που ακολουθούνται σε μια τυπική διαδικασία παραγωγής χάρτη προεξέχοντων σημείων εικόνας. Στη συνέχεια θα παρουσιάσουμε τις εναλλακτικές προσεγγίσεις τις οποίες ακολουθεί η μέθοδος GBVS.

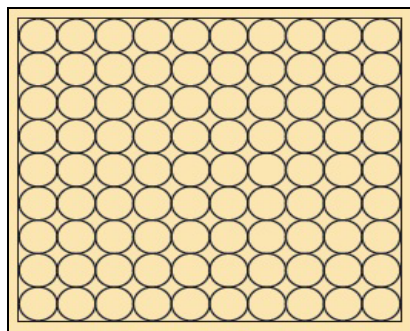


Εικόνα 41: Βασικά βήματα για την παραγωγή ενός χάρτη προεξοχόντων σημείων σε μια εικόνα

Η προσέγγιση της μεθόδου GBVS είναι η ακόλουθη. Προτείνεται μια κατανεμημένη, βασισμένη σε γράφο λύση, η οποία χρησιμοποιεί τοπικό υπολογισμό για την παραγωγή ενός χάρτη προεξοχόντων ο οποίος εξαρτάται παντού από τις γενικές πληροφορίες εντός της εικόνας. Τόσο για την ενεργοποίηση όσο και για την ομαλοποίηση, κατασκευάζεται ένα κατευθυνόμενο γράφημα με βάρη στις ακμές, τα οποία λαμβάνονται από τον χάρτη εισόδου, τον αντιμετωπίζουν ως μια αλυσίδα Markov [38] και υπολογίζεται στη συνέχεια η κατανομή ισορροπίας.

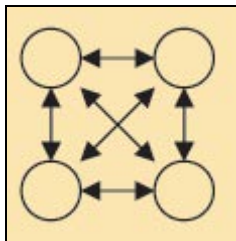
Η κατασκευή του γράφου γίνεται ακολουθώντας την παρακάτω διαδικασία.

1. Δημιουργείται ένας κόμβος για κάθε τοποθεσία σε ένα χάρτη εισόδου (χαρακτηριστικών ή ενεργοποίησης)



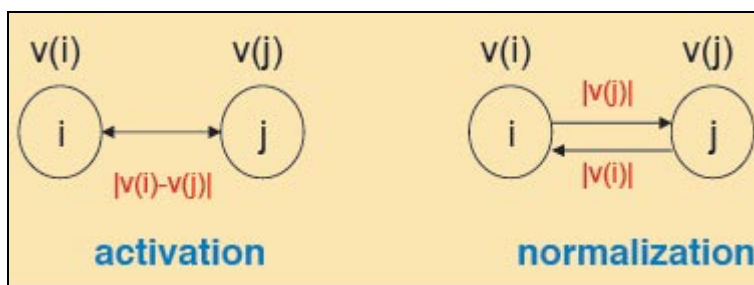
Εικόνα 42: Κόμβος για τοποθεσία σε ένα χάρτη εισόδου

2. Γίνεται εισαγωγή των κατευθύνσεων στις ακμές προς τις δυο κατευθύνσεις μεταξύ κάθε ζευγαριού κόμβων.



Εικόνα 43: Κατευθύνσεις των ακμών μεταξύ κόμβων

3. Καθορίζονται τα βάρη των ακμών ως εξής:



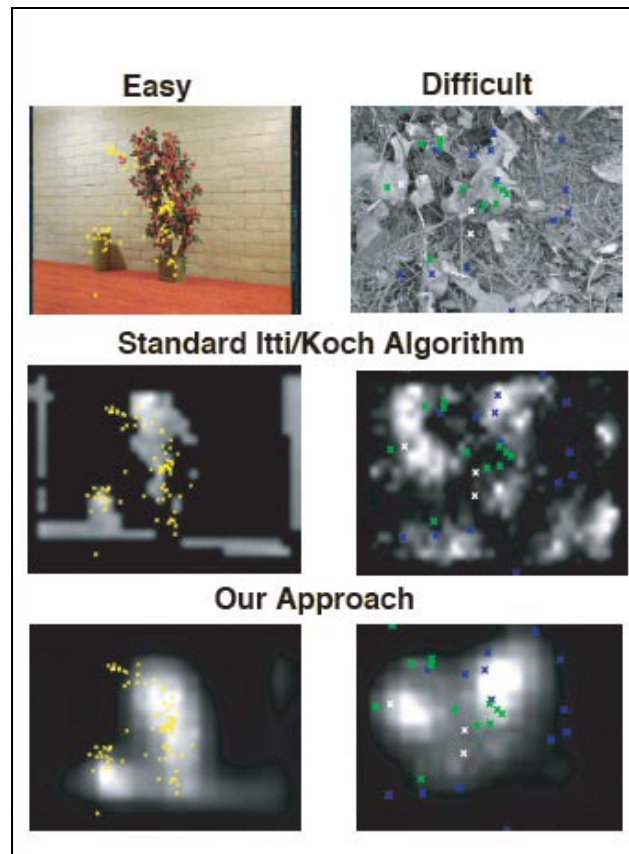
Εικόνα 44: Προσθήκη βαρών στις ακμές

Τα βάρη των ακμών πολλαπλασιάζονται με μια Gaussian ποινή απόστασης, έτσι ώστε οι κόμβοι οι οποίοι απέχουν μόνο ασθενώς να αλληλεπιδρούν. Η προσέγγιση αυτή επεκτείνεται σε πολλαπλές χωρικές κλίμακες με την εισαγωγή κόμβων σε κάθε θέση σε κάθε κλίμακα. Επίσης καθορίζονται οι ακμές και τα βάρη τους το ίδιο με πριν, με τον κατάλληλο ορισμό της απόστασης μεταξύ κόμβων σε διαφορετικές κλίμακες.

Οι κόμβοι αντιμετωπίζονται ως αναφορές και τα βάρη των ακμών ως πιθανότητες μετάβασης. Στη συνέχεια υπολογίζεται η κατανομή ισορροπίας της αλυσίδας Markov. Εάν πρόκειται για έναν χάρτη εισόδου $N \times N$ μεγέθους, αυτό θα έχει ως αποτέλεσμα χρονική πολυπλοκότητα $O(N^2 k)$ όπου $k \ll N$ σε μικρό αριθμό επαναλήψεων που απαιτούνται για την επίτευξη της ισορροπίας.

Για κάθε εικόνα σε κάποιο σώμα των δεδομένων, δημιουργείται μια συλλογή από χάρτες προεξοχών, συνενώνοντας διάφορες διαδικασίες ενεργοποίησης και κανονικοποίησης μεταγενέστερες του ίδιου ακριβώς σταδίου εξαγωγής χαρακτηριστικών. Στη συνέχεια συγκρίνεται η συνέπεια του κάθε χάρτη προεξοχών με τα δεδομένα σύνδεσης.

Στις παρακάτω εικόνες συγκρίνονται τα αποτελέσματα από χάρτες προεξοχόντων σημείων, οι οποίοι έχουν δημιουργηθεί από τις κλασσικές μεθόδους και από την μέθοδο GBVS. Η λειτουργία της τελευταίας, δείχνει ότι είναι πολύ πιο αποτελεσματική σε σχέση με τις άλλες μεθόδους ιδιαίτερα σε δύσκολα αναγνωρίσιμα προεξέχοντα οπτικά σημεία σε μια εικόνα. Αριστερά βρίσκονται θεωρητικά εύκολες εικόνες για ανίχνευση προεξοχόντων σημείων ενώ στα δεξιά βρίσκονται εικόνες δυσκολότερες στην ανίχνευση τέτοιων σημείων.



Εικόνα 45: Σύγκριση μεθόδων απόκτησης χαρτών προεξεχόντων σημείων

Ερμηνεύοντας αυτή την τεχνική μπορούμε να καταλήξουμε στο ότι ο αλγόριθμος αφήνει τη μάζα των προεξεχόντων σημείων μακριά από τα σύνορα του αντικειμένου με ένα μη τετριμμένο τρόπο ο οποίος δεν μπορεί να επιτευχθεί από την μονάχα από την εξομάλυνση. Επίσης η συγκεκριμένη προσέγγιση βρίσκει προεξέχουσες τιμές για κάθε θέση, οι οποίες εξαρτώνται από όλο το επίπεδο της εικόνας. Αυτή είναι μια σημαντική διαφορά από τις περισσότερες προσεγγίσεις οι οποίες βασίζονται στις τοπικές πληροφορίες.

5.2 Υλοποίηση της Εφαρμογής για Αυτόματη Αναγνώριση Αντικειμένων Εικόνας

Στο κεφάλαιο αυτό θα γίνει η αναλυτική παρουσίαση της εφαρμογής που υλοποιήθηκε καθώς και η παρουσίαση των αποτελεσμάτων της. Επίσης θα συζητηθεί το πώς συμβάλλει στον επιστημονικό τομέα της Τεχνητής Όρασης και της Ψηφιακής Επεξεργασίας Εικόνας, όπως αυτές παρουσιάστηκαν σε προηγούμενο κεφάλαιο/

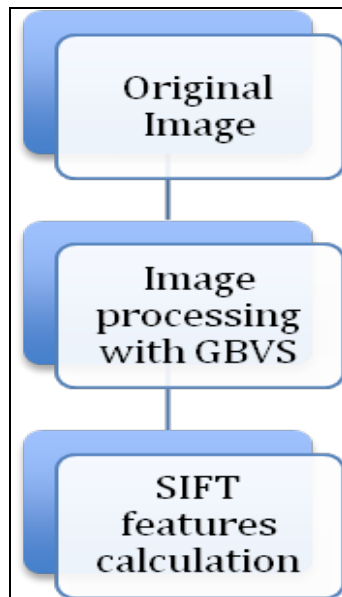
5.2.1 Εισαγωγή

Η βασική ιδέα της εφαρμογής έχει σαν στόχο την αναγνώριση αντικειμένων στα πλαίσια μιας εικόνας. Πιο συγκεκριμένα υλοποιήθηκε μια εφαρμογή στην οποία χρησιμοποιούνται φωτογραφίες γνωστών μνημείων της πόλης του Ηρακλείου, Κρήτης. Αρχικά ο χρήστης φωτογραφίζει κάποιο μνημείο από το κινητό του ή από κάποια φωτογραφική μηχανή. Η λειτουργία της εφαρμογής έχει να κάνει με την παροχή πληροφορίας για αυτό το μνημείο. Συγκεκριμένα η εφαρμογή είναι σε θέση να συγκρίνει την φωτογραφία του χρήστη με διαφορετικές, ήδη υπάρχουσες στη βάση δεδομένων της, και να προσφέρει την πληροφορία για το όνομα του μνημείου το οποίο φωτογράφησε ο εκάστοτε χρήστης ακόμη και την κατηγορία αντικειμένων στην οποία ανήκει το συγκεκριμένο μνημείο.

Η σύλληψη της ιδέας αυτής έχει να κάνει με την προσπάθεια δημιουργίας ενός έξυπνου συστήματος αναγνώρισης αντικειμένων, κύριο πλεονέκτημα του οποίου θεωρείται η χρήση του σε πραγματικές συνθήκες και σε πραγματικό χρόνο (real time). Με άλλα λόγια η διαδικασία σύγκρισης, της παρεχόμενης εικόνας από τον χρήστη με τις ήδη υπάρχουσες εικόνες στη βάση δεδομένων, πρέπει να ολοκληρώνεται σε όσο το δυνατόν μικρότερο χρονικό διάστημα. Προς την επίτευξη αυτού του στόχου εισάγεται η έννοια της προ-επεξεργασίας εικόνων. Αρχικός στόχος του συγγραφέα ήταν να καταδείξει το γεγονός ότι μπορεί να εξοικονομηθεί σημαντικός χρόνος με την προ-επεξεργασία των εικόνων της βάσης δεδομένων ιδιαίτερα σε τέτοιου είδους εφαρμογές στις οποίες τον σημαντικότερο ρόλο παίζει ο απαιτούμενος χρόνος υπολογισμού και το σωστό αποτέλεσμα.

Στόχος αυτής της επεξεργασίας εικόνων είναι η απόρριψη ασήμαντων, για τη διαδικασία ταξινόμησης, πληροφοριών οι οποίες υπάρχουν στην εικόνα. Η συγκεκριμένη διαδικασία αν και θεωρείται αυτονόητη και πολύ εύκολη για έναν άνθρωπο, παραμένει υπολογιστικά δύσκολη για ένα σύστημα τεχνητής όρασης. Αυτό συμβαίνει λόγω της ύπαρξης πολλών και διαφορετικών παραμέτρων οι οποίες επηρεάζουν το τελικό αποτέλεσμα. Με απλά λόγια επιθυμητή είναι η σύγκριση εικόνων οι οποίες περιέχουν όσο το δυνατόν λιγότερες πληροφορίες φόντου. Αυτή η προ-επεξεργασία στην συγκεκριμένη εργασία πραγματοποιείται με την χρήση της τεχνικής των «Προεξεχόντων Οπτικών Σημείων Βασισμένα σε Γράφους» η οποία παρουσιάστηκε στην αρχή αυτού του κεφαλαίου.

Στην παρακάτω εικόνα φαίνεται σχηματικά η διαδικασία η οποία ακολουθήθηκε σε αυτή την εφαρμογή.



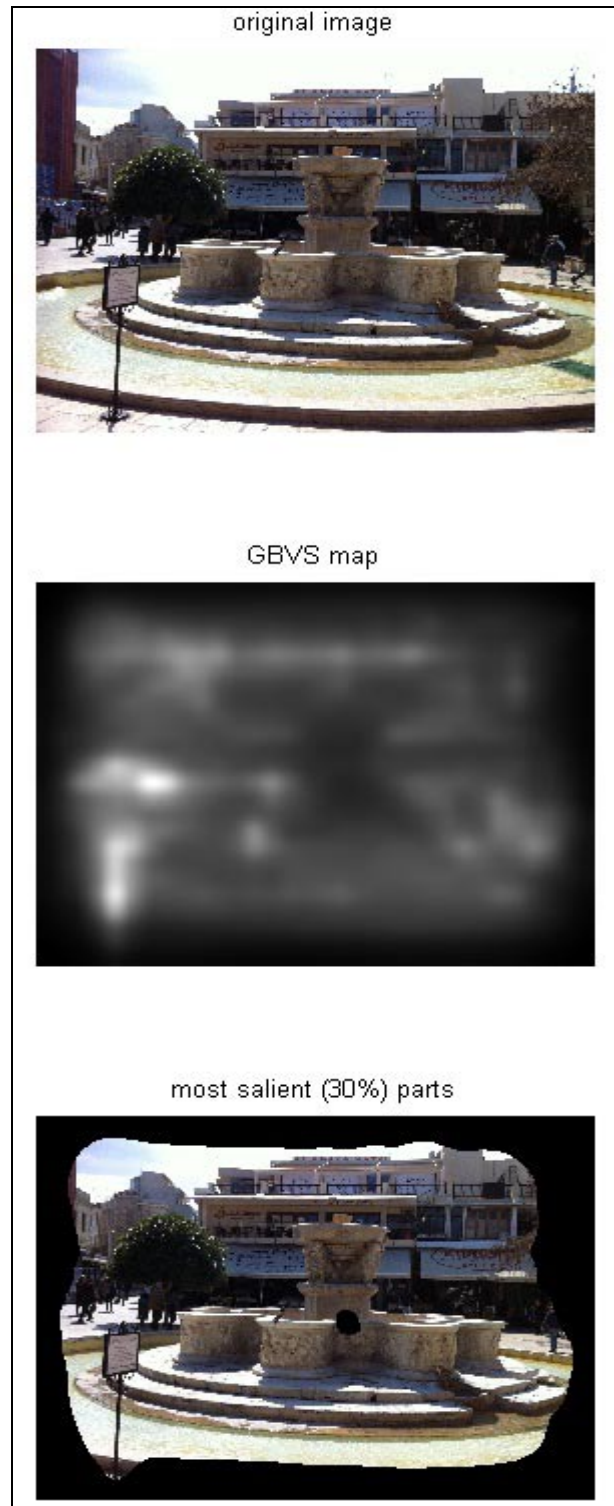
Εικόνα 46: Τα τρία βασικά βήματα που έγιναν κατά την υλοποίηση της εφαρμογής

5.2.2 Αναλυτική Παρουσίαση Διαδικασίας που ακολουθήθηκε

Θα παρουσιαστούν παρακάτω με αναλυτικό τρόπο τα βήματα που ακολουθήθηκαν για την υλοποίηση της τελικής εφαρμογής και των αποτελεσμάτων.

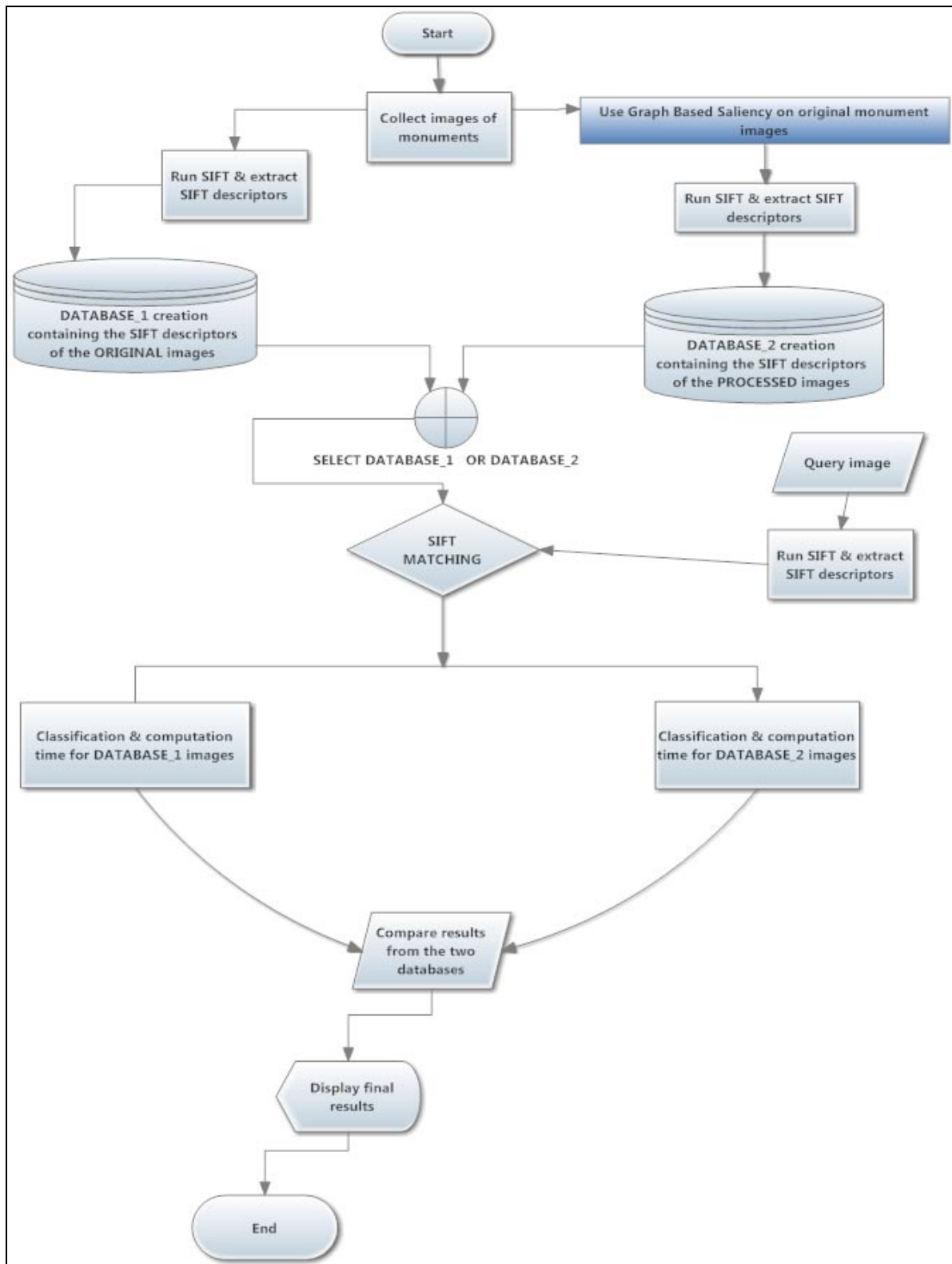
1. Σαν αρχικό βήμα της εφαρμογής θεωρείται η απόκτηση των φωτογραφιών από τα διάφορα μνημεία του Ηρακλείου.
2. Σε αυτό το βήμα γίνεται η χρήση της τεχνικής «Προεξεχόντων Οπτικών Σημείων Βασισμένα σε Γράφους» στις αρχικές εικόνες των μνημείων.
3. Στη συνέχεια χρησιμοποιείται ο SIFT αλγόριθμος για την εξαγωγή περιγραφών τόσο για τις αρχικές όσο και για τις προ-επεξεργασμένες εικόνες των μνημείων.
4. Αυτό το βήμα περιλαμβάνει τη δημιουργία των δυο διαφορετικών βάσεων δεδομένων. Η πρώτη περιέχει τους περιγραφείς που έχουν εξαχθεί, κατά την προηγούμενη φάση, από τις αρχικές εικόνες, ενώ η δεύτερη βάση δεδομένων περιέχει τους περιγραφείς που έχουν εξαχθεί από τις προ-επεξεργασμένες εικόνες.
5. Στη συνέχεια εισάγεται στο σύστημα η εικόνα που έχει τραβήξει ο χρήστης. Ουσιαστικά χρησιμοποιείται και σε αυτήν ο SIFT αλγόριθμος για να γίνει η απαραίτητη εξαγωγή χαρακτηριστικών περιγραφών.
6. Στο βήμα αυτό επιλέγεται μια από τις δυο διαθέσιμες βάσεις δεδομένων που δημιουργήθηκαν στο βήμα 4 για να συγκριθούν με την εικόνα τους χρήστη.
7. Στη συνέχεια γίνεται ο υπολογισμός για το αν έχει ταξινομηθεί σωστά το αντικείμενο από την εικόνα του χρήστη. Επίσης γίνεται υπολογισμός του χρόνου που χρειάστηκε για κάθε μια από τις δυο βάσεις δεδομένων.
8. Συγκρίνονται τα αποτελέσματα. Ουσιαστικά γίνεται έλεγχος για το πόσο καλά έχει ταξινομηθεί το αντικείμενο της εικόνας του χρήστη με τη προ-επεξεργασία που έχει υποστεί η βάση δεδομένων.
9. Τέλος παρουσιάζονται τα τελικά αποτελέσματα της σύγκρισης με μορφή πίνακα.

Στην παρακάτω φωτογραφία φαίνεται ένα παράδειγμα χρήσης του αλγορίθμου Προεξεχόντων Οπτικών Σημείων Βασισμένα σε Γράφους. Πάνω φαίνεται η αρχική εικόνα του μνημείου, στη μέση διακρίνεται ο χάρτης που παράγεται από τη μέθοδο GBVS και κάτω φαίνεται η αρχική εικόνα μαυρισμένη στα σημεία, το 30% της αρχικής, που υπάρχει η μικρότερη συγκέντρωση προεξεχόντων οπτικών σημείων, βάση του αλγορίθμου.



Εικόνα 47:Χρήση του αλγορίθμου "Προεξεχόντων Οπτικών Σημείων Βασισμένα σε Γράφους"

Στο σχεδιάγραμμα της παρακάτω εικόνας φαίνονται αναλυτικά τα βήματα τα οποία ακολουθήθηκαν για την παραγωγή του τελικού αποτελέσματος.



Εικόνα 48:Συνολική Επισκόπηση της Διαδικασίας που ακολουθήθηκε

5.2.3 Προσδιορισμός Απαιτήσεων

Θεωρείται σκόπιμο να παρουσιαστούν οι μεταβλητές που συμμετείχαν στο πείραμα και πως αυτές διαμορφώθηκαν κατά τη διάρκεια ανάπτυξης της εφαρμογής.

Πιο συγκεκριμένα, επιλέχθηκαν 4 ιστορικά μνημεία της πόλης του Ηρακλείου, τα οποία φαίνονται παρακάτω. Οι φωτογραφίες τραβήχτηκαν όλες από την ίδια ψηφιακή κάμερα ενός iPhone4 με 5 MP ανάλυση.



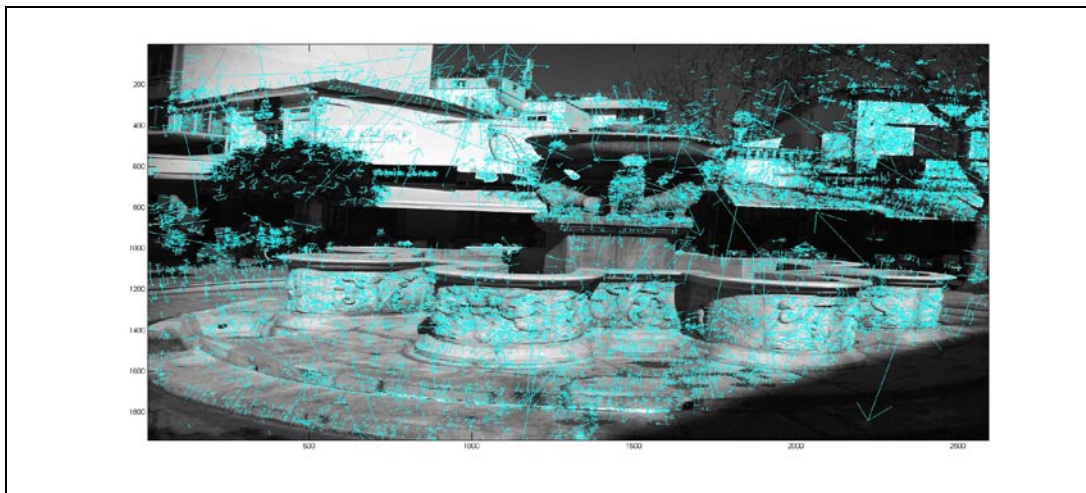
Εικόνα 49: Τα τέσσερα τουριστικά μνημεία του Ηρακλείου, τα οποία χρησιμοποιήθηκαν στην εφαρμογή

Επίσης πρέπει να σημειωθεί στο σημείο αυτό ο υπολογιστικός χρόνος που απαιτείται για την προεπεξεργασία των εικόνων με τη χρήση της GBVS μεθόδου, ότι κυμαίνεται γύρω στα 0.85 – 1 sec. Επιπλέον οι εικόνες που τέθηκαν προς ταξινόμηση ήταν τρεις διαφορετικές μεταξύ τους για το κάθε ένα μνημείο, δηλαδή με διαφορετικό φωτισμό, γωνία λήψης και απόσταση. Συνολικά λοιπόν για την παραγωγή των πειραματικών αποτελεσμάτων, χρησιμοποιήθηκαν 4 διαφορετικές κλάσεις (η κάθε μια αντιπροσωπεύει και ένα μνημείο) με 5 διαφορετικές μεταξύ τους εικόνες. Επίσης πρέπει να τονίσουμε το γεγονός ότι οι εικόνες προς ταξινόμηση είναι διαφορετικές από αυτές που ήδη υπάρχουν στην αντίστοιχη κλάση. Τέλος γίνεται η δοκιμή των εικόνων προς ταξινόμηση, αφού έχουν υποστεί και αυτές προ-επεξεργασία με GBVS για την σύγκριση του υπολογιστικού χρόνου που απαιτείται.

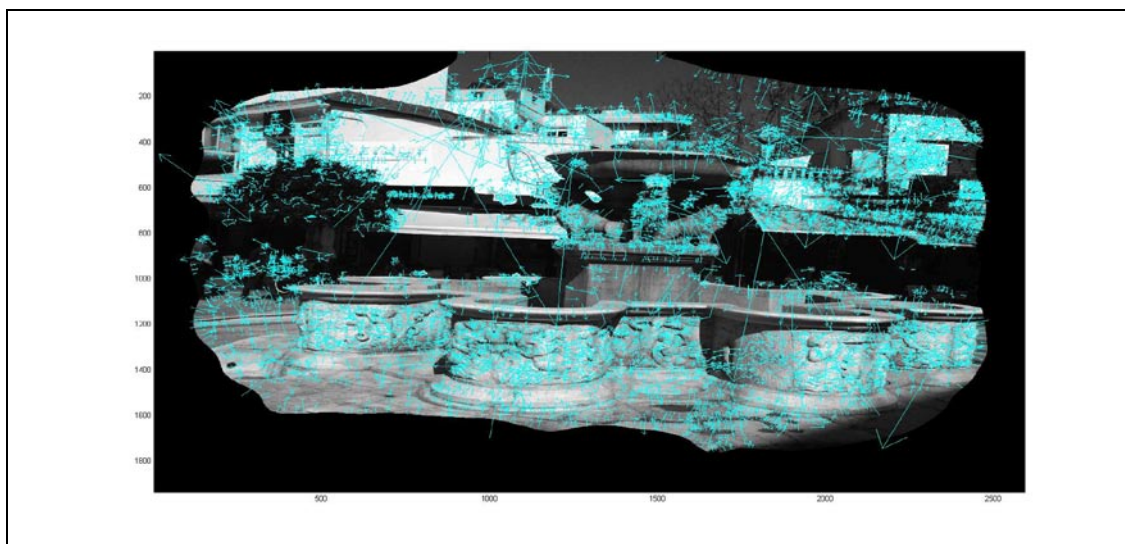
Όσον αφορά την μέθοδο ταξινόμησης που χρησιμοποιήθηκε, κύριος στόχος ήταν η προσπάθεια σωστής ταξινόμησης των εικόνων χωρίς να χρειαστεί να χρησιμοποιήσουμε κάποιο σύστημα εκπαίδευσης, όπως αυτά που αναφέρθηκαν νωρίτερα. Αυτό γίνεται για την αποφυγή του μεγάλου υπολογιστικού χρόνου που απαιτείται καθώς η εφαρμογή μας πρόκειται να λειτουργήσει ιδανικά σε πραγματικό χρόνο.

5.2.4 Πειραματικά αποτελέσματα

Για να γίνει πιο κατανοητό το σκεπτικό της όλης εφαρμογής, παρακάτω παρουσιάζεται ένα μνημείο, όπως αυτό αποθηκεύεται στις δυο διαφορετικές βάσεις δεδομένων τις οποίες χρησιμοποιήσαμε. Με άλλα λόγια εμφανίζονται οι περιγραφείς, των σημείων κλειδιών που έχουν εξαχθεί με τον SIFT αλγόριθμο, πάνω στην εικόνα. Είναι εύκολα αντιληπτό από τη σύγκριση των δυο εικόνων ότι ο συνολικό αριθμός των περιγραφέων για την προ-επεξεργασμένη εικόνα μειώνεται σημαντικά, χωρίς ωστόσο να παραλείπονται οι σημαντικές πληροφορίες του κυρίαρχου αντικειμένου μέσα στην εικόνα, αλλά μειώνοντας τους περιγραφείς στις περιοχές του φόντου.

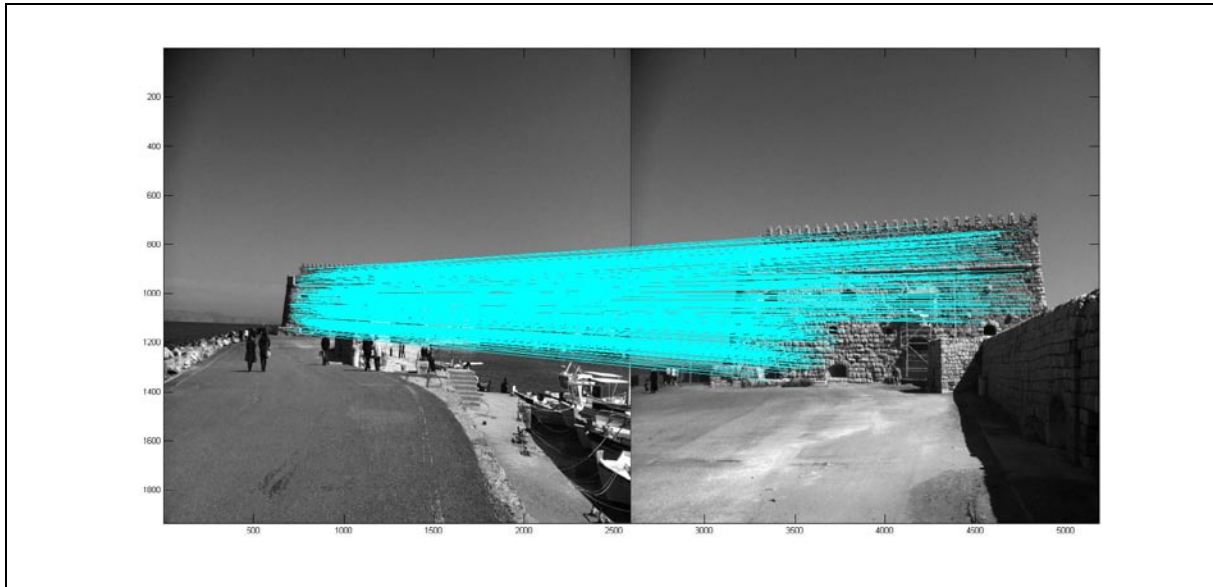


Εικόνα 50: Σημεία Κλειδιά στην αρχική εικόνα

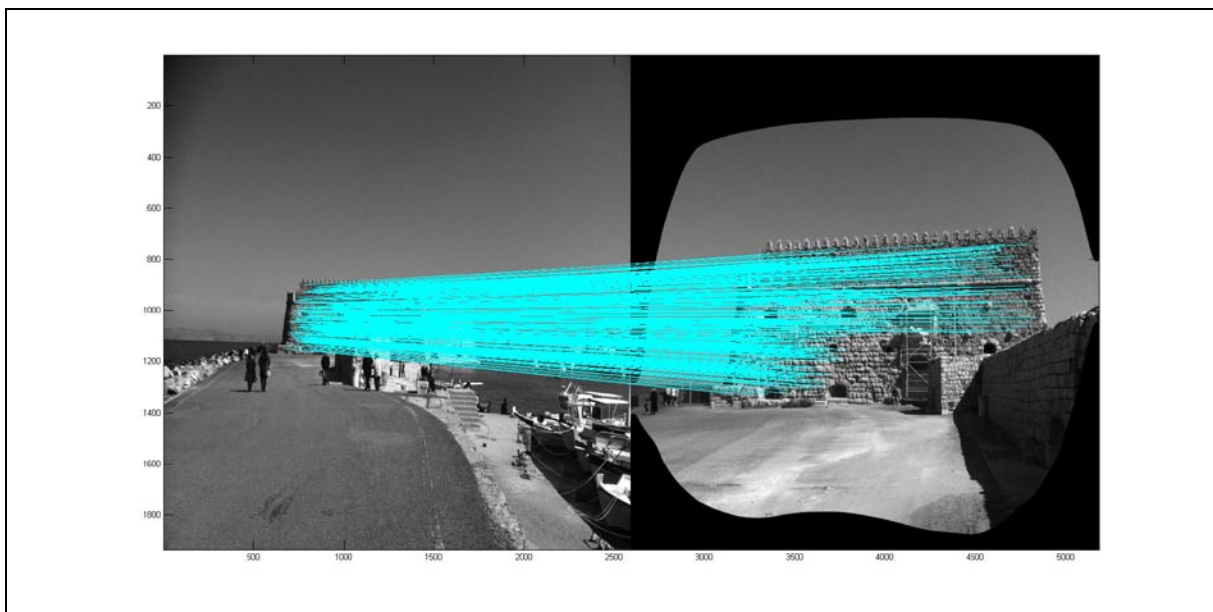


Εικόνα 51: Σημεία Κλειδιά στην επεξεργασμένη με GBVS εικόνα

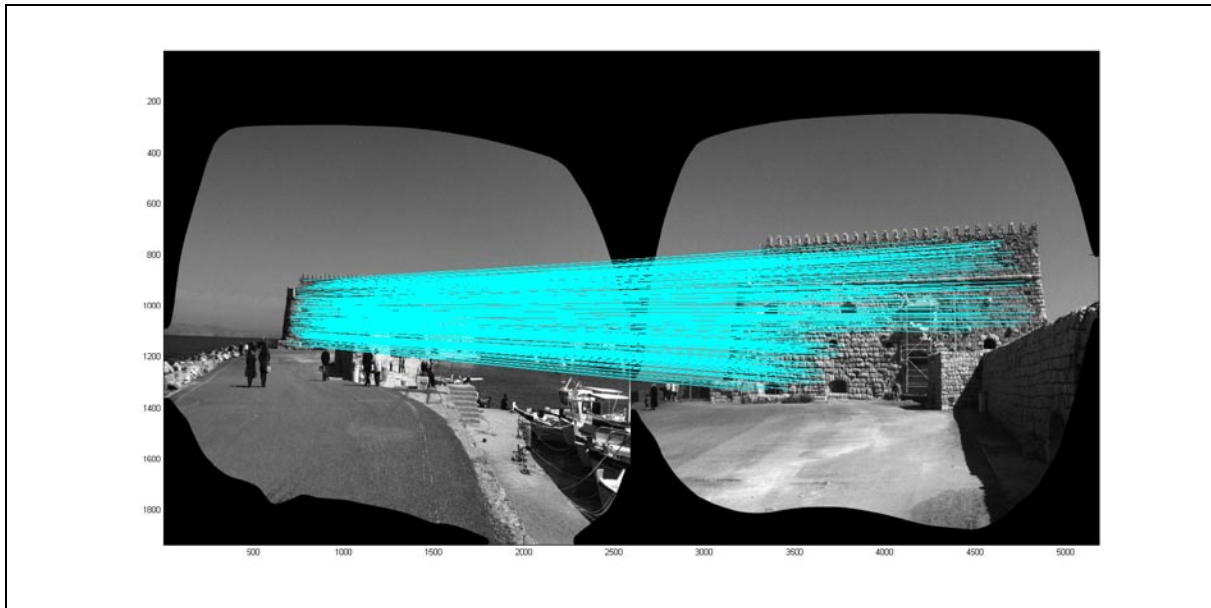
Επίσης πολύ σημαντικό ρόλο, για τη σωστή ταξινόμηση των αντικειμένων που θα ακολουθήσει, διαδραματίζει το σύνολο των αντιστοιχίσεων που θα εμφανίζονται να είναι πρακτικά σωστό. Με άλλα λόγια εξετάζουμε εάν η προ-επεξεργασία στην οποία υποβάλλουμε τις αρχικές εικόνες διατηρεί εξίσου καλά αποτελέσματα κατά τη φάση της αντιστοίχισης μεταξύ της εικόνας προς κατηγοριοποίηση και αυτής της βάσης δεδομένων. Τα αποτελέσματα όπως φαίνονται και στις παρακάτω εικόνες θεωρούνται πολύ ικανοποιητικά, καθώς η διαδικασία της προ-επεξεργασίας δεν επηρεάζει αρνητικά τις αντιστοιχίες που υπάρχουν μεταξύ ενός ζεύγους εικόνων, ενώ παράλληλα μειώνεται ο υπολογιστικός χρόνος που απαιτείται διότι λαμβάνονται υπόψη πολύ λιγότερα σημεία κλειδιά κατά τη χρήση του SIFT αλγορίθμου.



Εικόνα 52: Σύνολο Αντιστοιχιών μεταξύ δυο αρχικών εικόνων



Εικόνα 53: Σύνολο Αντιστοιχιών μεταξύ αρχικής και προ-επεξεργασμένης εικόνας



Εικόνα 54: Σύνολο Αντιστοιχιών μεταξύ δυο προ-επεξεργασμένων εικόνων

Στο σημείο αυτό θα αναφερθούμε στην μέθοδο ταξινόμησης η οποία χρησιμοποιήθηκε στην εφαρμογή. Για τους λόγους που αναφέρθηκαν και παραπάνω, η ταξινόμηση που επιλέχτηκε είναι θεωρητικά απλή, για την εξοικονόμηση χρόνου και για να παρέχει ένα εργαλείο ελέγχου στο κατά πόσο η αρχική ιδέα τεκμηριώνεται και πειραματικά. Με άλλα λόγια κύριος στόχος ήταν παρουσίαση του υπολογιστικού χρόνου που εξοικονομείται χωρίς ωστόσο να επηρεάζονται τα αποτελέσματα ταξινόμησης.

Συγκεκριμένα, κάθε μια εικόνα προς ταξινόμηση συγκρινόταν με κάθε μια εικόνα από όλες τις κλάσεις. Καταγράφαμε τον αριθμό των αντιστοιχιών της δοκιμαστικής εικόνας με την εικόνα από τη βάση δεδομένων. Στη συνέχεια αθροίζαμε αυτές τις αντιστοιχίες. Κατά αυτό τον τρόπο κάθε κλάση έπαιρνε μια ετικέτα ανάλογη με τον αριθμό του αθροίσματος αυτού. Στη συνέχεια ελέγχαμε σε ποια κλάση εμφανιζόταν ο μεγαλύτερος συνολικός αριθμός και έτσι το σύστημα ήταν σε θέση να μας πληροφορήσει για την κλάση στην οποία πιθανολογείται να ανήκει η δοκιμαστική εικόνα.

Στον παρακάτω συγκεντρωτικό πίνακα φαίνονται αναλυτικά τα πειραματικά αποτελέσματα στα οποία καταλήξαμε. Με έντονη γραμματοσειρά παρουσιάζεται ο χρόνος που εξοικονομείται σε κάθε περίπτωση. Επίσης με το σύμβολο \checkmark σημειώνονται οι σωστά ταξινομημένες εικόνες μνημείων.

Καταλήγουμε στα εξής συμπεράσματα:

- Ο χρόνος που απαιτείται για την ολοκλήρωση της ταξινόμησης της κάθε εικόνας, μειώνεται σε όλες τις περιπτώσεις στις οποίες χρησιμοποιείται προ-επεξεργασία με GBVS.
- Σε όλες τις περιπτώσεις γίνεται σωστή ταξινόμηση των αρχικών εικόνων με τη μέθοδο κατηγοριοποίησης που χρησιμοποιήσαμε.
- Επίσης σε όλες τις περιπτώσεις η σωστή ταξινόμηση διατηρείται και για τις προ-επεξεργασμένες εικόνες με GBVS.
- Σε περίπτωση που γίνει προ-επεξεργασία τόσο στις δοκιμαστικές εικόνες, όσο και στις εικόνες της βάσης δεδομένων ο χρόνος που εξοικονομείται αυξάνεται δραστικά. Αυτό σημαίνει ότι με αυτή τη μέθοδο πλησιάζουμε πολύ περισσότερο τον απώτερο στόχο μας χωρίς ωστόσο να σπαταλιέται σημαντικά επιπλέον χρόνος.

Πίνακας 3: Συγκεντρωτικά πειραματικά αποτελέσματα ταξινομήσεων & εξοικονόμησης χρόνου

Query Image	ORIGINAL	SALIENCY ON DB IMAGES ONLY	TIME SAVED	CORRECT CLASS	SALIENCY ON BOTH DB & QUERY IMAGES	TIME SAVED	CORRECT CLASS
Church_1	103 sec 1394 matches	97 sec 1250 matches	6 sec	√	82 sec 1188 matches	20 sec	√
Church_2	248 sec 205 matches	233 sec 287 matches	15 sec	√	182 sec 183 matches	65 sec	√
Church_3	206 sec 415 matches	193 sec 414 matches	13 sec	√	172 sec 409 matches	33 sec	√
Fortress_1	213 sec 1891 matches	197 sec 1921 matches	16 sec	√	193 sec 1904 matches	19 sec	√
Fortress_2	273 sec 1634 matches	259 sec 1620 matches	14 sec	√	244 sec 1599 matches	28 sec	√
Fortress_3	212 sec 361 matches	176 sec 370 matches	36 sec	√	111 sec 218 matches	100 sec	√
Fountain_1	268 sec 1757 matches	221 sec 1626 matches	47 sec	√	211 sec 1603 matches	56 sec	√
Fountain_2	300 sec 424 matches	238 sec 421 matches	62 sec	√	211 sec 408 matches	88 sec	√
Fountain_3	348 sec 546 matches	299 sec 500 matches	49 sec	√	260 sec 400 matches	87 sec	√
Loggia_1	166 sec 167 matches	145 sec 157 matches	21 sec	√	91 sec 146 matches	74 sec	√
Loggia_2	197 sec 326 matches	167 sec 305 matches	30 sec	√	155 sec 288 matches	41 sec	√
Loggia_3	178 sec 297 matches	145 sec 270 matches	33 sec	√	109 sec 195 matches	68 sec	√

5.2.5 Συμπεράσματα και Μελλοντική Εργασία

Τα αποτελέσματα αυτής της πρώτης «προσπάθειας» συνδυασμού δυο μεθόδων για την αυτόματη αναγνώριση αντικειμένων θεωρούνται ιδιαίτερα ενθαρρυντικά. Όπως δείξαμε παραπάνω η χρήση του αλγορίθμου «Προεξεχόντων Οπτικών Σημείων Βασισμένα σε Γράφους» ως ένα στάδιο προ-επεξεργασίας, μειώνει το χρόνο που απαιτείται σε μια διαδικασία αντιστοίχισης η οποία βασίζεται στο αλγόριθμο « Μετασχηματισμού κλιμακωτά αναλλοίωτων χαρακτηριστικών – SIFT » Επίσης αποδείχθηκε ότι η χρήση του GBVS αλγορίθμου τόσο στις εικόνες προς ταξινόμηση όσο και σε αυτές της βάσης δεδομένων, μειώνει ακόμη περισσότερο τον απαιτούμενο χρόνο για την ταξινόμηση των εικόνων.

Μελλοντικά οι επεκτάσεις της εφαρμογής μπορούν να συντελεστούν σε πολλούς διαφορετικούς τομείς. Πιο συγκεκριμένα θα γίνει προσπάθεια για τη βελτίωση του συνολικού χρόνου που απαιτείται για τη μέθοδο ταξινόμησης των εικόνων. Αυτό μπορεί να επιτευχθεί αποφεύγοντας την απλή σύγκριση της εικόνας προς ταξινόμηση με κάθε εικόνα της βάσης δεδομένων. Αντί αυτού μπορούν να χρησιμοποιηθούν πιο εξελιγμένες μέθοδοι ταξινόμησης, όπως αυτοί που περιγράφηκαν παραπάνω αλλά και άλλοι. Επιπλέον μια ιεραρχική κατάταξη των SIFT σημείων κλειδιών, βασιζόμενη στο GBVS θα διερευνηθεί.

Βιβλιογραφία

1. Wikipedia, [Επιστήμη Υπολογιστών](#)
2. Wikipedia, [Computer Vision](#)
3. Wikipedia, [Computer Graphics](#)
4. Wikipedia, [Machine Vision](#)
5. Wikipedia, [Augmented Reality](#)
6. Wikipedia, [Image Processing](#)
7. Wikipedia, [Digital Image Processing](#)
8. Wikipedia, [Segmentation](#)
9. Wikipedia, [Image Denoising](#)
10. Wikipedia, [Face Detection](#)
11. Wikipedia, [Saliency \(neuroscience\)](#)
12. Wikipedia, [Feature Extraction](#)
13. Wikipedia, [SIFT Algorithm](#)
14. Lowe, D. G., "Distinctive Image Features from Scale-Invariant Keypoints", International Journal of Computer Vision, 60, 2, pp. 91-110, 2004.
15. Wikipedia, [k-d Tree](#)
16. Wikipedia, [Best-bin-first search](#)
17. Wikipedia, [SURF Algorithm](#)
18. [Image Processing on Line – Speeded-Up Robust Features](#) Edouard Oyallon, ENS Cachan
19. [Image Processing on Line – ASIFT](#) Guoshen Yu and Jean-Michel Morel
20. Wikipedia, [Histogram of oriented gradients](#)
21. JINHO KIM "Comparing Image Classification Methods: K-Nearest-Neighbor and Support-Vector-Machines"
22. Wikipedia, [Neural network](#)
23. Wikipedia, [Neuron](#)
24. Wikipedia, [Backpropagation](#)
25. Wikipedia, [Support Vector Machine - SVM](#)
26. Wikipedia, [Structural risk minimization](#)
27. Wikipedia, [Fuzzy measure theory](#)
28. M.Seetha, I.V.Muralikrinshka, "Artificial neural networks and other methods of image classification"
29. Wikipedia, [K-nearest neighbor theory](#)
30. Wikipedia, [Lazy learning](#)
31. Wikipedia, [Non-parametric statistics](#)
32. [A Detailed Intro to K-Nearest Neighbor Algorithm](#)
33. Wikipedia, [Bayes' Theorem](#)
34. Wikipedia, [Object Detection](#)
35. Wikipedia, [Object Recognition](#)
36. Ιάσοντας Κόκκινος Εθνικό Μετσόβιο Πολυτεχνείο " Συνέργεια Κατάτμησης Εικόνων και Αναγνώριση Αντικειμένων με Γεωμετρικές και Στατιστικές Τεχνικές της Όρασης Υπολογιστών"
37. Wikipedia, [Hough Transformation](#)
38. Jonathan Harel, Christof Koch , Pietro Perona California Institute of Technology Pasadena, CA 91125 "Graph Based Visual Saliency"
39. Wikipedia, [Markov Chain](#)