



Τεχνολογικό Εκπαιδευτικό Ίδρυμα Κρήτης

Σχολή Τεχνολογικών Εφαρμογών
Τμήμα Εφαρμοσμένης Πληροφορικής & Πολυμέσων



Πτυχιακή εργασία

Τίτλος: Σχηματική και χρωματική αναγνώριση προτύπων

ΓΙΟΥΡΟΥΚΑΚΗΣ ΓΕΩΡΓΙΟΣ (ΑΜ: 1388)

Επιβλέπων καθηγητής: ΤΡΙΑΝΤΑΦΥΛΛΙΔΗΣ ΓΕΩΡΓΙΟΣ

Επιτροπή Αξιολόγησης:

Ημερομηνία παρουσίασης: 18/10/2011

ΠΕΡΙΕΧΟΜΕΝΑ

1. Εισαγωγή	5
1.1. Ψηφιακή εικόνα	5
1.2. Χρησιμότητα ψηφιακής εικόνας	7
1.3. Διάφορες Εφαρμογές	8
2. Αναγνώριση προτύπων	12
2.1. Αναγνώριση προτύπων	12
2.2. Ιστορική αναδρομή - Μεθοδολογίες αναγνώρισης προτύπων	18
2.3. Σύστημα αναγνώρισης προτύπων	28
3. MATLAB	455
3.1 Τι είναι το Matlab	45
3.2 Πως δουλεύει το Matlab	46
3.3 Εργαλειοθήκη αναγνώρισης προτύπων του Matlab	50
4. Εφαρμογές	53
4.1 1 ^η Εφαρμογή και αποτελέσματα	64
4.2 2 ^η Εφαρμογή και αποτελέσματα	69
4.3 3 ^η Εφαρμογή και αποτελέσματα	73
5. Αποτελέσματα της μελέτης και της έρευνας	77
6. Βιβλιογραφία	78

ΠΙΝΑΚΑΣ ΕΙΚΟΝΩΝ

Εικόνα 1 Παράδειγμα ανίχνευσης αντικειμένου.....	9
Εικόνα 2 Παράδειγμα εντοπιστή της τροχιάς ενός κινούμενου ανθρώπου.....	11
Εικόνα 3 Το πρόβλημα της αναγνώρισης προτύπων (14).....	13
Εικόνα 4 Σύνολο εικόνων που απεικονίζουν τους αριθμούς 0 έως 9.....	15
Εικόνα 5 Κατηγορίες χαρακτηριστικών.....	16
Εικόνα 6 Κανονικό μοντέλο ταξινομητή.....	16
Εικόνα 7 Ένα σύνολο δεδομένων δύο κλάσεων διαχωρίζεται από μία γραμμική συνάρτηση.....	17
Εικόνα 8 Το διάγραμμα βαθμίδων ενός συστήματος αναγνώρισης προτύπων.....	17
Εικόνα 9 Πρόβλημα δυαδικής ταξινόμησης.....	20
Εικόνα 10 Περιθώριο σε δύο διαφορετικούς διαχωρισμούς των ίδιων δεδομένων.....	20
Εικόνα 11 Διανύσματα υποστήριξης.....	21
Εικόνα 12 Τα υπερεπίπεδα και τα ημιεπίπεδα του χώρου στον οποίο εφαρμόζεται ο SVM.....	21
Εικόνα 13 Παράδειγμα εφαρμογής του SVM και δύο σημείων.....	22
Εικόνα 14 Παράδειγμα εφαρμογής του μη γραμμικού μοντέλου SVM.....	23
Εικόνα 15 Μετασχηματισμός μη γραμμικά διαχωρίσιμων δεδομένων σε νέο χώρο όπου είναι γραμμικά διαχωρίσιμα.....	24
Εικόνα 16 Πίνακα kernel.....	24
Εικόνα 17 Εικόνες από τη βάση δεδομένων με οχήματα.....	28
Εικόνα 18 Εικόνες από τη βάση δεδομένων με οχήματα.....	29
Εικόνα 19 Εικόνα 20.....	30
Εικόνα 21 Εικόνα 22.....	30
Εικόνα 23.....	30
Εικόνα 24.....	30
Εικόνα 25:Οκτάβες και επίπεδα θολώματος.....	32
Εικόνα 26: Διαγορά Gaussian εικόνων.....	33
Εικόνα 27: Εφαρμογή DoG.....	33
Εικόνα 28: Γειτονικά pixel.....	34
Εικόνα 29: Προσδιορισμός διανυσματικού μεγέθους και κατεύθυνσης.....	37
Εικόνα 30:Ιστόγραμμα κατεύθυνσης.....	37
Εικόνα 31: Δημιουργία παραθύρων γύρω από ένα keypoint.....	38
Εικόνα 32: Για την δημιουργία του keypoint descriptor υπολογίζουμε το διανυσματικό μέγεθος και την κατεύθυνση των στοιχείων σε κάθε σημείο γύρω από το σημείο κλειδί. Αυτά υπολογίζονται από ένα Gaussian παράθυρο, ορισμένο από τον επικαλυπτόμενο κύκλο. Στην συνέχεια το ιστόγραμμα κατευθύνσεων υπολογίζεται συνοψίζοντας τα στοιχεία σε 4X4 υποπεριοχές, όπως φαίνεται στο δεξιό μέρος της εικόνας. Το μήκος κάθε βέλους είναι ίσο με το άθροισμα των διανυσματικών μεγεθών των στοιχείων κοντά σε αυτή την κατεύθυνση μέσα στην περιοχή.....	39
Εικόνα 33 Υποσύστημα σχηματικής αναγνώρισης προτύπων.....	43
Εικόνα 34 Το περιβάλλον του MATLAB.....	46
Εικόνα 35 Παράδειγμα εκτέλεσης εντολής στο MATLAB.....	46
Εικόνα 36 Παράδειγμα εκτέλεσης εντολής στο MATLAB.....	47
Εικόνα 37 Παράδειγμα δημιουργίας μεταβλητών στο MATLAB.....	47
Εικόνα 38 Παράδειγμα δημιουργίας πίνακα στο MATLAB.....	48
Εικόνα 39 Άνοιγμα του σημειωματαρίου του MATLAB.....	48
Εικόνα 40 Σημειωματάριο του MATLAB (Editor).....	49

Εικόνα 41 Thumbnails των εικόνων της βάσης που χρησιμοποιήθηκαν στα πειράματα	63
Εικόνα 42 Εικόνες ελέγχου της 1 ^{ης} εφαρμογής.....	64
Εικόνα 43 Απεικόνιση των 964 σημείων SIFT σε μία από τις εικόνες ελέγχου της 1 ^{ης} εφαρμογής....	65
Εικόνα 44 Απεικόνιση των 426 σημείων SIFT σε μία από τις εικόνες ελέγχου της 1 ^{ης} εφαρμογής....	65
Εικόνα 45 Απεικόνιση των 5.361 σημείων SIFT σε μία από τις εικόνες ελέγχου της 1 ^{ης} εφαρμογής.	66
Εικόνα 46 Οι ταμπέλες που αντιστοιχούν στα τρία οχήματα ελέγχου της 1 ^{ης} εφαρμογής.....	68
Εικόνα 47 Εικόνες ελέγχου της 2 ^{ης} εφαρμογής.....	69
Εικόνα 48 Οι ταμπέλες που αντιστοιχούν στα τρία οχήματα ελέγχου της 2 ^{ης} εφαρμογής.....	72
Εικόνα 49 Εικόνες ελέγχου της 3 ^{ης} εφαρμογής.....	73
Εικόνα 50 Απεικόνιση των 1.618 σημείων SIFT σε μία από τις εικόνες ελέγχου της 3 ^{ης} εφαρμογής.	74
Εικόνα 51 Απεικόνιση των 1.760 σημείων SIFT σε μία από τις εικόνες ελέγχου της 3 ^{ης} εφαρμογής.	74
Εικόνα 52 Απεικόνιση των 4.161 σημείων SIFT σε μία από τις εικόνες ελέγχου της 3 ^{ης} εφαρμογής.	75
Εικόνα 53 Απεικόνιση των 5.233 σημείων SIFT σε μία από τις εικόνες ελέγχου της 3 ^{ης} εφαρμογής.	75
Εικόνα 54 Απεικόνιση των 5.631 σημείων SIFT σε μία από τις εικόνες ελέγχου της 3 ^{ης} εφαρμογής.	76
Εικόνα 55 Απεικόνιση των 774 σημείων SIFT σε μία από τις εικόνες ελέγχου της 3 ^{ης} εφαρμογής....	76

ΠΙΝΑΚΑΣ ΠΙΝΑΚΩΝ

Πίνακας 1 Χρώμα pixel με βάση το χρωματικό μοντέλο RGB	44
Πίνακας 2 Αποτελέσματα 1 ^{ης} εφαρμογής για διάφορες παραμέτρους της πολυωνομικής kernel συνάρτησης	68
Πίνακας 3 Αποτελέσματα 2 ^{ης} εφαρμογής για διάφορες παραμέτρους της πολυωνομικής kernel συνάρτησης	71
Πίνακας 4 Αποτελέσματα 2 ^{ης} εφαρμογής για διάφορες τιμές της διασποράς της γκαουσιανής kernel συνάρτησης	72

1. Εισαγωγή

1.1. Ψηφιακή εικόνα

Μία εικόνα του πραγματικού κόσμου είναι μία συνάρτηση $f(x, y)$ δύο πραγματικών χωρικών μεταβλητών. Για οποιαδήποτε ζεύγος τιμών (x, y) η συνάρτηση f έχει μία τιμή, η οποία δίνει τη φωτεινότητα της εικόνας στο συγκεκριμένο σημείο. Στην περίπτωση που έχουμε μία γκρι εικόνα η φωτεινότητα κάθε σημείου είναι μία απόχρωση του γκρι η οποία εκτείνεται από το άσπρο μέχρι το μαύρο.

Η εικόνα που περιγράψαμε είναι η αναλογική εικόνα. Εικόνες όπου το πεδίο ορισμού και το πεδίο τιμών της συνάρτησης f δεν είναι όλοι οι πραγματικοί αριθμοί αλλά πεπερασμένοι ακέραιοι αριθμοί τότε η εικόνα ονομάζεται ψηφιακή. Σύμφωνα με τον (1) μία ψηφιακή εικόνα προκύπτει με τη διακριτοποίηση του πεδίου τιμών της αντίστοιχης αναλογικής εικόνας. Η διαδικασία μετατροπής μίας αναλογικής εικόνας σε ψηφιακή ονομάζεται ψηφιοποίηση και περιλαμβάνει τα στάδια της δειγματοληψίας και της κβάντισης.

Δειγματοληψία

Δεδομένου ότι η εικόνα απεικονίζεται από μία συνάρτηση δύο μεταβλητών είναι ουσιαστικά ένα διδιάστατο σήμα. Για να γίνει ευκολότερα κατανοητή η δειγματοληψία της εικόνας θα ορίσουμε πρώτα την δειγματοληψία ενός μονοδιάστατου σήματος. Η δειγματοληψία ενός μονοδιάστατου σήματος $f(x)$ είναι η επιλογή συγκεκριμένων τιμών της αρχικής συνάρτησης και το σύνολο αυτών των διακριτών τιμών αποτελούν το νέο διακριτό σήμα. Πιο συγκεκριμένα, έχουμε την ομοιόμορφη δειγματοληψία η οποία είναι η επιλογή τιμών ενός σήματος $f(x)$ ανά συγκεκριμένα διαστήματα τιμών του x . Έχουμε δηλαδή την επιλογή δειγμάτων τιμών του σήματος με συγκεκριμένη συχνότητα f_s , η οποία ονομάζεται συχνότητα δειγματοληψίας.

Στην περίπτωση ενός διδιάστατου σήματος, όπως αυτό της εικόνας έχουμε δύο ανεξάρτητες μεταβλητές x και y οι οποίες έχουν να κάνουν με τον χώρο και δίνουν ένα σημείο στο χώρο. Σε αυτό το σημείο η συνάρτηση έχει τιμή που απεικονίζει την φωτεινότητα της εικόνας.

Για να επεκτείνουμε στο διδιάστατο σήμα την έννοια της δειγματοληψίας όπως την περιγράψαμε για το μονοδιάστατο σήμα, αυτό που συμβαίνει είναι ότι το σήμα σαρώνεται στο χώρο και επιλέγονται κάποιες τιμές με σταθερή συχνότητα δειγματοληψίας. Το αποτέλεσμα είναι ο συμβολισμός της εικόνας με ένα διδιάστατο σήμα διακριτών τιμών. Συνήθως για την καλύτερη κατανόηση της ψηφιακής εικόνας αυτή συμβολίζεται με ένα ορθογώνιο πλέγμα όπου κάθε τετράγωνο ονομάζεται εικονοστοιχείο-pixel. Η θέση του εικονοστοιχείου δίνεται από τις ακέραιες τιμές των x και y και η φωτεινότητα από την τιμή της συνάρτησης $f(x, y)$. Η μαθηματική αναπαράσταση μίας μονόχρωμης ψηφιακής εικόνας I είναι ένας διδιάστατος πίνακας που περιλαμβάνει μη αρνητικά στοιχεία p_{ij} , $0 \leq i < m, 0 \leq j < n$, των οποίων οι τιμές είναι ίσες με αυτές των αντίστοιχων εικονοστοιχείων:

$$I = \begin{pmatrix} p_{00} & \cdots & p_{0(m-1)} \\ \vdots & \ddots & \vdots \\ p_{(n-1)0} & \cdots & p_{(n-1)(m-1)} \end{pmatrix}$$

Η δειγματοληψία χρησιμοποιείται για τη μετατροπή της αναλογική εικόνας σε ψηφιακή. Για να μπορέσουμε να κάνουμε την αντίστροφη διαδικασία, δηλαδή έχοντας την ψηφιακή εικόνα να πάρουμε την αναλογική θα πρέπει να ικανοποιείται το θεώρημα Nyquist. Σύμφωνα με αυτό, η συχνότητα δειγματοληψίας πρέπει να είναι τουλάχιστον διπλάσια από την μέγιστη συχνότητα που αναμένεται να ενυπάρχει στο σήμα, προκειμένου να είναι εφικτή η επανασύνθεση του αρχικού αναλογικού σήματος από το ψηφιακό σήμα χωρίς απώλειες.

Κβάντιση

Η κβάντιση είναι μία διαδικασία που γίνεται στην ψηφιακή εικόνα. Λαμβάνοντας το πεδίο τιμών της εικόνας δηλαδή τις φωτεινότητες της εικόνας, χωρίζουμε το πεδίο σε ίσα διαστήματα τα οποία καθορίζονται από ισαπέχουσες τιμές. Αυτές οι τιμές ονομάζονται στάθμες κβάντισης. Όπως αναφέρθηκε, κάθε εικονοστοιχείο έχει μία τιμή η οποία είναι μέσα στο πεδίο τιμών της συνάρτησης $f(x, y)$. Αυτή η τιμή κβαντίζεται, δηλαδή στρογγυλοποιείται στην κοντινότερη στάθμη κβάντισης. Με αυτό τον τρόπο αναπόφευκτα εισάγονται σφάλματα, δεδομένου ότι η πραγματική τιμή του εικονοστοιχείου γίνεται μικρότερη ή μεγαλύτερη. Τα σφάλματα μειώνονται με την αύξηση των σταθμών κβάντισης. Αυτές οι στάθμες κβάντισης συμβολίζονται με ένα πλήθος δυαδικών ψηφίων. Μεγαλύτερος αριθμός δυαδικών ψηφίων προκύπτει από περισσότερες στάθμες κβάντισης.

Οι περιγραφές που προηγήθηκαν αναφέρονται σε εικόνες της κλίμακας του γκρι ή όπως είναι εσφαλμένα γνωστές ως ασπρόμαυρες εικόνες. Παρακάτω θα δώσουμε τον ακριβή ορισμό των ασπρόμαυρων εικόνων. Οι έγχρωμες εικόνες προκύπτουν από τον συνδυασμό τριών ή παραπάνω χρωμάτων ανάλογα με το χρωματικό μοντέλο. Η δειγματοληψία και η κβάντιση επεκτείνονται εύκολα στις έγχρωμες εικόνες χωρίς περαιτέρω περιορισμούς λόγω χρώματος.

Στην περίπτωση που κβαντίζουμε μία ασπρόμαυρη εικόνα το πλήθος των δυαδικών ψηφίων που απεικονίζουν τις τιμές των pixel καθορίζουν τον αριθμό των διαφορετικών αποχρώσεων του γκρι, ο οποίος ονομάζεται βάθος χρώματος της εικόνας.

Το ελάχιστο δυνατό βάθος χρώματος που μπορεί να υπάρξει σε μία εικόνα είναι δύο, δηλαδή 1 bit. Είναι η περίπτωση της ασπρόμαυρης εικόνας όπου κάθε εικονοστοιχείο μπορεί να είναι ή άσπρο ή μαύρο χωρίς ενδιάμεσες αποχρώσεις του γκρι. Το συνηθέστερο βάθος χρώματος είναι 256, όπου υπάρχουν 8 bits ενώ σε εξαιρετικές περιπτώσεις το βάθος χρώματος μπορεί να φτάσει τα 65536 (16 bits).

Στην περίπτωση που έχουμε έγχρωμη εικόνα και το χρωματικό μοντέλο είναι RGB (Red Green Blue) σημαίνει ότι κάθε εικονοστοιχείο έχει αντί για μία τιμή φωτεινότητας, τρεις τιμές, οι οποίες αντιστοιχούν στα τρία χρώματα και τότε οι τιμές για το βάθος χρώματος που αναφέραμε τριπλασιάζονται.

1.2. Χρησιμότητα ψηφιακής εικόνας

Η δημιουργία ενός ψηφιακού κειμένου, μίας ψηφιακής εικόνας, ενός ψηφιακού ήχου ή ενός ψηφιακού βίντεο περιλαμβάνει στον όρο ψηφιοποίηση έργου. Η ψηφιοποίηση περιλαμβάνει μία σειρά από στάδια από το αρχικό μέχρι το τελικό αποτέλεσμα: σχεδιασμός του έργου ψηφιοποίησης, επιλογή περιεχομένου, προετοιμασία για ψηφιοποίηση, μεταχείριση των πρωτοτύπων, ψηφιοποίηση, διατήρηση του ψηφιακού περιεχομένου, μεταδεδομένα, ενέργειες ανάδειξης – προβολής, πνευματικά δικαιώματα και διαχείριση έργων ψηφιοποίησης.

Το πιο βασικό στάδιο είναι αυτό της ψηφιοποίησης που περιλαμβάνει τη σάρωση, την ψηφιακή φωτογράφιση και την ψηφιακή αποτύπωση των πρωτοτύπων συμπεριλαμβάνοντας και την οποιαδήποτε ψηφιακή επεξεργασία που μπορεί να έχει γίνει.

Τα πρότυπα της εικόνας και των μεταδεδομένων είναι μία κρίσιμη απόφαση και δεν πρέπει να προσπερνιέται. Τα τελευταία χρόνια έχουν επικρατήσει συγκεκριμένα πρότυπα τα οποία περιγράφονται με συντομία στη συνέχεια. Όταν ο χρήστης έχει στην κατοχή του μία εικόνα ενός συγκεκριμένου προτύπου μπορεί να τη μετατρέψει σε άλλο πρότυπο με το κατάλληλο λογισμικό επεξεργασίας εικόνας.

TIFF (Tagged Image File Format)

Οι εικόνες τύπου TIFF είναι υψηλής ποιότητας και κατά τη συμπίεση δεν υπάρχει απώλεια πληροφορίας ή δε γίνεται συμπίεση. Για το λόγο αυτό έχουν μεγάλη χωρητικότητα. Οι σαρωτές και οι ψηφιακές φωτογραφικές μηχανές μπορούν να παράγουν εικόνες με το πρότυπο TIFF. Αυτό το πρότυπο είναι κατάλληλο όταν ο χρήστης επιθυμεί να αποθηκεύσει ψηφιακά αντίγραφα.

JPEG (Joint Photographic Experts Group)

Το JPEG χρησιμοποιείται ευρέως και έχει γίνει βάση ISO. Χρησιμοποιείται για τη μεταφορά και την παρουσίαση εικόνων μέσω δικτύων τα οποία έχουν μικρό εύρος ζώνης δεδομένου ότι οι εικόνες προτύπου JPEG καταλαμβάνουν μικρό όγκο. Σε αντίθεση με το πρότυπο TIFF κατά τη συμπίεση με βάση το JPEG υπάρχει απώλεια όγκου η οποία μειώνει τη χωρητικότητα της εκάστοτε εικόνας. Για το λόγο αυτό όλοι οι φυλλομετρητές ιστού καθώς και οι περισσότερες εφαρμογές του ηλεκτρονικού υπολογιστή αναγνωρίζουν τις εικόνες αυτού του προτύπου.

GIF (Graphics Interchange Format)

Το πρότυπο αυτό χρησιμοποιείται επίσης για τη μεταφορά και την παρουσίαση των εικόνων μέσω δικτύων με μικρό εύρος ζώνης. Η συμπίεση κατά το πρότυπο GIF γίνεται χωρίς απώλεια πληροφορίας αλλά μειώνει τον όγκο της εικόνας. Ένα μειονέκτημα είναι ότι αυτό το πρότυπο υποστηρίζει μόλις 256 χρώματα. Το GIF χρησιμοποιείται σε γραφικά και το JPEG σε εικόνες με μεγάλη ποσότητα πληροφορίας. Οι φυλλομετρητές ιστού και η πλειοψηφία των εφαρμογών ηλεκτρονικού υπολογιστή υποστηρίζουν το συγκεκριμένο πρότυπο.

PNG (Portable Network Graphics)

Το πρότυπο PNG είναι ανοικτό και ο στόχος ανάπτυξής του ήταν να αντικατασταθεί το πρότυπο GIF. Κατά τη συμπίεση ο όγκος μειώνεται αλλά η πληροφορία δε χάνεται και συγκριτικά με το GIF το μέγεθος των εικόνων είναι μικρότερο.

Η χρησιμότητα της ψηφιακής εικόνας φαίνεται και μέσα από τις διάφορες εφαρμογές της όπως κάποιες από αυτές θα περιγραφούν παρακάτω.

1.3. Διάφορες Εφαρμογές

Συμπίεση εικόνας

Μία από τις πρώτες εφαρμογές της ψηφιακής εικόνας ήταν στη βιομηχανία των εφημερίδων και έχει να κάνει με την συμπίεση και τη μετάδοση της εικόνας. Οι πρώτες ψηφιακές εικόνες που μεταδόθηκαν από μία γεωγραφική περιοχή σε μία άλλη με ηλεκτρονικό τρόπο ήταν μέσω υποβρύχιου καλωδίου από το Λονδίνο στη Νέα Υόρκη. Το 1920 αναπτύχθηκε ένα σύστημα μετάδοσης εικόνας μέσω καλωδίων γνωστό ως Bartlane το οποίο βελτίωσε σε πολύ μεγάλο βαθμό το χρόνο μετάδοσης μιας εικόνας. Πιο συγκεκριμένα, μία εικόνα που χρειαζόταν περισσότερο από μία εβδομάδα για να μεταδοθεί μέσω του Ατλαντικού, μπορούσε πλέον να φτάσει στα χέρια του παραλήπτη μέσα σε τρεις ώρες από τη στιγμή αναχώρησής της από τον αποστολέα.

Τα πρώτα συστήματα Bartlane μπορούσαν να κωδικοποιήσουν τις εικόνες σε πέντε διακριτά επίπεδα του γκρι και αργότερα σε 15 επίπεδα (1929). Οι κωδικοποιημένες εικόνες εκείνης της εποχής που αναφέραμε εντάσσονται στην κατηγορία των ψηφιακών εικόνων αλλά οι διαδικασίες που αναφέραμε δεν ανήκουν στον τομέα της ψηφιακής επεξεργασίας εικόνας δεδομένου ότι δεν υπήρχε η συμβολή των ηλεκτρονικών υπολογιστών. Η ψηφιακή επεξεργασία εικόνας σχετίζεται άμεσα με την τεχνολογία και την εξέλιξη των υπολογιστών καθώς έχει να κάνει με την δυνατότητα αποθήκευσης και την υπολογιστική ισχύ. Τα βασικά κομμάτια είναι η αποθήκευση δεδομένων, η απεικόνιση και η μετάδοση.

Η συμπίεση εικόνας είναι απαραίτητη εκτός από την ανάγκη μετάδοσης της εικόνας και στην περίπτωση που χρειάζεται μία προεπισκόπηση εικόνων που ανήκουν σε μία μεγάλη βάση δεδομένων (2). Η αρχή της συμπίεσης της εικόνας είναι να επιλεγούν οι πληροφορίες που θέλουμε να κρατήσουμε και να τις κωδικοποιήσουμε χωρίς απώλειες.

Επομένως το πρώτο στάδιο κατά τη διαδικασία της συμπίεσης είναι να βρεθεί η ελάχιστη ομάδα από κβαντισμένους συντελεστές που θα δώσουν μία εικόνα με υψηλή ποιότητα. Οι πληροφορίες που χάνονται δε μπορούν να ανακτηθούν για αυτό το λόγο αν δεν επιθυμούμε να έχουμε καθόλου απώλειες θα συμπιέσουμε και τον θόρυβο.

Κατά την ανάκτηση της αρχικής εικόνας με τη χρήση της συμπιεσμένης είναι δυνατόν να μη χρειαστεί ολόκληρη η συμπιεσμένη εικόνα. Αυτό συμβαίνει γιατί η αποσυμπίεση γίνεται κατά κλίμακες ξεκινώντας από χαμηλή ανάλυση και αυξάνοντας τη σιγά σιγά ανάλογα με την επιθυμία του παραλήπτη. Αν ο παραλήπτης επιθυμεί απλά να δει το θέμα της εικόνας δεν είναι απαραίτητο να αποσυμπιέσει την εικόνα στην αρχική της ανάλυση.

Όσον αφορά στο θόρυβο που ενυπάρχει στην εικόνα υπάρχουν διάφορες επιλογές ανάλογα με τις ανάγκες της κάθε περίπτωσης. Για παράδειγμα όταν οι εικόνες είναι εικόνες αστρονομίας κρίνεται αναγκαίος ο διαχωρισμός του σήματος θορύβου από το σήμα της εικόνας ώστε να συμπιεστεί μόνο το τελευταίο. Ο θόρυβος δεν είναι απλά περιττός αλλά και ενοχλητικός σε αυτό τον τομέα. Ένα αντίθετο παράδειγμα είναι τα συστήματα προεπισκόπησης εικόνων όπου ο θόρυβος δεν επηρεάζει αισθητά τις ανάγκες ενός τέτοιου συστήματος οπότε δεν αξίζει να αναπτυχθεί μία διαδικασία αφαίρεσής του.

Ανίχνευση αντικειμένων και ομαδοποίηση σημείων

Η ανίχνευση των αντικειμένων είναι μία τεχνολογία των ηλεκτρονικών υπολογιστών που σχετίζεται με την τεχνητή όραση και την ψηφιακή επεξεργασία εικόνας. Αυτή η τεχνολογία ασχολείται με την ανίχνευση αντικειμένων σε ψηφιακές εικόνες και βίντεο. Χαρακτηριστικά παραδείγματα τέτοιων εφαρμογών είναι συστήματα ανίχνευσης προσώπων και ανθρώπων γενικότερα. Τέτοιου είδους συστήματα βρίσκουν εφαρμογή σε διάφορους τομείς όπως η παρακολούθηση μέσω βίντεο.



Εικόνα 1 Παράδειγμα ανίχνευσης αντικειμένου

Οι διαφορετικοί τρόποι απεικόνισης της ψηφιακής εικόνας εξυπηρετούν σε διάφορες εφαρμογές καθώς κάνουν πιο εύκολη την εξαγωγή πληροφοριών μιας εικόνας ανάλογα με την εφαρμογή. Υπάρχει η έννοια των σημασιολογικών χαρακτηριστικών της εικόνας τα οποία μπορούν να βρεθούν όταν η απεικόνιση της εικόνας έχει μια συγκεκριμένη δομή και μπορούν να δώσουν μία ερμηνεία της εικόνας.

Σε εικόνες οι οποίες έχουν μεγάλα αντικείμενα χρησιμοποιήθηκαν συγκεκριμένες τεχνικές ώστε να ανιχνευτούν από διάφορους επιστήμονες όπως (3), (4), (5). Υπάρχουν όμως και οι περιπτώσεις όπου είναι απαραίτητη η ανίχνευση πολύ μικρών αντικειμένων σε μία εικόνα όπως αστρονομικά αντικείμενα.

Για τον εντοπισμό και την ανάλυση συγκεκριμένων αντικειμένων χρησιμοποιούνται μετασχηματισμοί πολλαπλής ανάλυσης της εικόνας. Υπάρχουν χάρτες των αντικειμένων της εικόνας ή χρησιμοποιείται μία δυαδική απεικόνιση της εικόνας η οποία μπορεί να επεξεργαστεί περαιτέρω ώστε να απορρίφθουν τυχόν λανθασμένες ανιχνεύσεις αντικειμένων. Κάποιες από τις τεχνικές που χρησιμοποιούνται σε τέτοιες περιπτώσεις βασίζονται σε διαχειρισή τους με βάση τη μαθηματική μορφολογία Minkowski (6).

Έχουν γίνει πολλές μελέτες για να αποδείξουν ότι το περιεχόμενο της εικόνας παίζει βασικό ρόλο στην ανίχνευση των αντικειμένων που αυτή απεικονίζει (7), (8). Το περιεχόμενο της εικόνας είναι ένας όρος που χρησιμοποιείται συχνά στην τεχνητή όραση χωρίς να υπάρχει ένας σαφής ορισμός. Ουσιαστικά αναφέρεται σε κάθε πληροφορία της εικόνας που μπορεί να επηρεάσει τη σκηνή και τα αντικείμενα που απεικονίζονται.

Μία έξυπνη εφαρμογή – εκμετάλευση της τοπικής πληροφορίας, δηλαδή του περιεχομένου συγκεκριμένων pixel χρησιμοποιείται στην ανίχνευση αντικειμένων όπου αρχικά τοποθετείται ένα παράθυρο σε μία μικρή περιοχή του αντικειμένου και στη συνέχεια αυτό μεγαλώνει σιγά σιγά περιλαμβάνοντας γειτονικά pixel τα οποία έχουν παρόμοιο περιεχόμενο (9), (10).

Η τοπική πληροφορία του περιεχομένου των pixel είναι μία τεχνική που χρησιμοποιείται στην κατάτμηση εικόνων, στην εξαγωγή των ορίων των αντικειμένων και σε μοντέλα αντικειμένων/περιγραμμάτων. Το περιεχόμενο μιας εικόνας χωρίζεται σε κάποιες κατηγορίες. Κάποιες από αυτές περιγράφονται παρακάτω και περιλαμβάνουν το τριδιάστατο περιεχόμενο, το σημασιολογικό, το φωτογραφικό, το περιεχόμενο του φωτός, το καιρικό, το γεωγραφικό, το χρονικό και το πολιτισμικό (11).

Το γεωμετρικό περιεχόμενο μία εικόνας βοηθάει στην τριδιάστατη αναπαράσταση της σκηνής που απεικονίζεται ώστε να αποκαλυφθεί η διάταξη των επιφανειών και των διαφορετικών αντικειμένων δίνοντας απαντήσεις σε προβλήματα που μπορεί να προκληθούν για παράδειγμα από την αλληλοεπικάλυψη των αντικειμένων (12).

Το σημασιολογικό περιεχόμενο μιας εικόνας μπορεί να αποκαλύψει πληροφορίες για τη δραστηριότητα και το γεγονός που απεικονίζονται. Κατ' επέκταση οι τεχνικές αυτές χρησιμοποιούνται σε ψηφιακά βίντεο τα οποία ουσιαστικά είναι διαδοχή ψηφιακών εικόνων. Έτσι έχουν γίνει συστήματα ώστε να ανιχνεύονται τυχόν ύποπτες κινήσεις από εγκληματίες σε μεγάλους χώρους, όπως σε παρκιν αυτοκινήτων, τράπεζες κ.λ.π..

Το περιεχόμενο του φωτισμού μπορεί να δώσει πληροφορίες για την εικόνα σχετικά με τη θέση του ήλιου, την πιθανή παρουσία συννεφιάς και τις σκιές που υπάρχουν. Το περιεχόμενο για τον καιρό έχει να κάνει με τη θερμοκρασία, τους ανέμους, την υγρασία και την ομίχλη.

Το γεωγραφικό περιεχόμενο μπορεί να αποκαλύψει την πραγματική θέση του αντικειμένου ή και τον τύπο γης που αποτυπώνεται σε μία εικόνα, όπως θάλασσα ή έρημος.

Όσον αφορά στο πολιτισμικό περιεχόμενο είναι μία διαφορούμενη έννοια. Κυρίως έχει να κάνει με τον τρόπο που λαμβάνεται η εικόνα, ο τρόπος που επιλέγονται οι διάφορες βάσεις δεδομένων ακόμα και με το όνομα τους.

Σε συστήματα ανίχνευσης αντικειμένων η χρήση του περιεχομένου της εικόνας μπορεί να παίζει καθοριστικό ρόλο. Ανάλογα με τις ανάγκες και το σκοπό της κάθε εφαρμογής ανίχνευσης αντικειμένων η πληροφορία περιεχομένου μπορεί να είναι χρήσιμη. Πιο συγκεκριμένα, η παρουσία και η εμφάνιση του αντικειμένου, το μέγεθός του, η τοποθεσία του και ο πιθανός χώρος που καταλαμβάνει είναι πληροφορίες ώστε να ανιχνευτεί μέσω ενός υπολογιστικού αλγορίθμου.

Πολλά αντικείμενα έχουν συγκεκριμένο περιβάλλον, όπως για παράδειγμα μία τοστιέρα συνήθως συναντάται σε κουζίνες και όχι κάπου αλλού (π.χ. κρεβατοκάμαρα). Επομένως γνωρίζοντας το περιβάλλον που απεικονίζεται σε μία εικόνα ή σε ένα βίντεο μπορεί να προβλεφθεί η παρουσία συγκεκριμένων αντικειμένων μέσα σε αυτό. Όσον αφορά στο γεωγραφικό περιεχόμενο υποδεικνύει ότι κάποια αντικείμενα υπάρχουν ή δεν υπάρχουν σε συγκεκριμένα μέρη, όπως μια καμινάδα δε μπορεί να έχει φωτογραφηθεί στην Αθήνα. Επίσης κάποια αντικείμενα φωτογραφίζονται από άλλα γεγονόσ που έχει να κάνει με το πολιτισμικό περιεχόμενο.

Η φωτεινότητα και η σκίαση ενός αντικειμένου εξαρτώνται από το περιεχόμενο που έχει να κάνει με τον καιρό και με τις συνθήκες φωτισμού. Υπάρχουν βέβαια και οι ρυθμίσεις της κάμερας οι οποίες επηρεάζουν την εμφάνιση του αντικειμένου.

Συμπερασματικά η ανάπτυξη συστημάτων για την αυτόματη ανίχνευση αντικειμένων σε μία εικόνα και κατ' επέκταση σε ένα βίντεο είναι ένας βασικός τομέας που ερευνείται για να βελτιωθεί και να εξελιχθεί από επιστήμονες της τεχνητής όρασης και της ψηφιακής επεξεργασίας εικόνας.

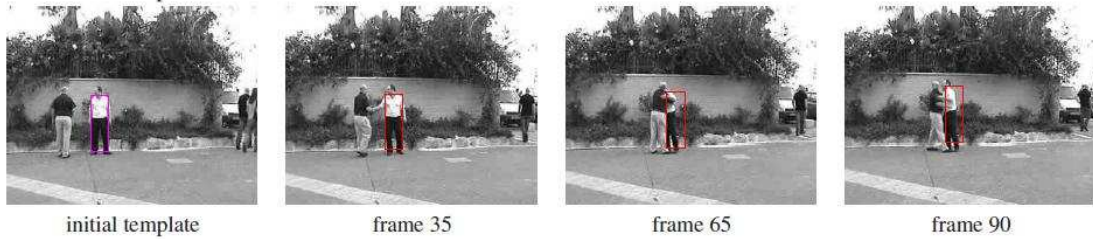
Εντοπισμός της τροχιάς κινούμενων αντικειμένων (ψηφιακό βίντεο)

Μία έννοια στενά συνδεδεμένη με την ανίχνευση αντικειμένων σε εικόνες είναι ο εντοπισμός της τροχιάς κινούμενων αντικειμένων που απεικονίζονται σε ένα ψηφιακό βίντεο. Πρόκειται για ένα πεδίο το οποίο αναπτύχθηκε τα τελευταία χρόνια στον τομέα της τεχνητής όρασης. Οι ερευνητές και οι επιστήμονες αναπτύσσουν αλγορίθμους ώστε να αναπτύξουν εύρωστα συστήματα εντοπισμού που μπορούν να χρησιμοποιηθούν σε συγκεκριμένες εφαρμογές ή ακόμα και σε ένα ευρύ σύνολο διαφορετικών εφαρμογών.

Η παρακολούθηση της τροχιάς ενός κινούμενου αντικειμένου μέσα από ένα ψηφιακό βίντεο είναι μία διαδικασία που συλλαμβάνει την τροχιά ενός ή πολλών αντικειμένων από το πρώτο πλάνο του βίντεο μέχρι το τελευταίο. Ο ορισμός του μπορεί να αποκαλύπτει ένα απλό σύστημα γιατί είναι εύκολο να το συλλάβει ο ανθρώπινος νους αλλά στην πραγματικότητα περιλαμβάνει πολλές διαφορετικές τεχνικές και κατηγορίες βασισμένες σε κριτήρια που έχουν να κάνουν με κάθε σύστημα και τις ανάγκες της εκάστοτε εφαρμογής.

Η ανάπτυξη ενός εύρωστου τέτοιου συστήματος που θα μπορούσε να χρησιμοποιηθεί σε οποιαδήποτε εφαρμογή και υπό διάφορες συνθήκες είναι η μεγαλύτερη πρόκληση για επιστήμονες εξειδικευμένους σε αυτό τον τομέα. Υπάρχουν περιπτώσεις όπου ένα τέτοιο σύστημα μπορεί να αποτύχει εάν κατά την ανάπτυξή του δεν ελήφθησαν υπόψη όλες οι πιθανές περιπτώσεις.

Για το λόγο αυτό μέχρι σήμερα τα συστήματα που αναπτύσσονται για την ανίχνευση της κίνησης αντικειμένων δημιουργούνται με τη χρήση συγκεκριμένων περιορισμών που ανταποκρίνονται σε συγκεκριμένα είδη βίντεο. Αυτό σημαίνει ότι ένας τέτοιος αλγόριθμος αν χρησιμοποιηθεί σε διαφορετική κατηγορία βίντεο θα αποτύχει. Για παράδειγμα, έχουν αναπτυχθεί πολλά τέτοια συστήματα που ακολουθούν τα πρόσωπα που εμφανίζονται σε ένα βίντεο. Αλλά άμα χρησιμοποιηθούν για να ακολουθήσουν την τροχιά των αυτοκινήτων σε έναν δρόμο τότε θα αποτύχουν (13).



Εικόνα 2 Παράδειγμα εντοπιστή της τροχιάς ενός κινούμενου ανθρώπου

Η αρχικοποίηση ενός τέτοιου συστήματος γίνεται είτε χειροκίνητα από τον χρήστη είτε αυτόματα με τη χρήση ενός κατάλληλου ανιχνευτή αντικειμένου σε συγκεκριμένο πλάνο του βίντεο.

Ο ρόλος αυτών των αλγορίθμων είναι καθοριστικός και απαραίτητος σε πολλές εφαρμογές του τομέα τεχνητής όρασης. Χρησιμοποιούνται σε συστήματα παρακολούθησης με τη χρήση βίντεο που είναι τοποθετημένα σε χώρους ευαίσθητους σε εγκληματικές ενέργειες. Επίσης έχουν χρησιμοποιηθεί τέτοιοι αλγόριθμοι στον έλεγχο του κυκλοφοριακού προβλήματος σε αυτοκινητόδρομους, στη συμπίεση βίντεο, σε υπολογιστικά συστήματα διαδραστικά με τον άνθρωπο καθώς και στην εικονική πραγματικότητα.

Κατά την ανάπτυξη ενός συστήματος εντοπισμού της τροχιάς κινούμενων αντικειμένων υπάρχουν κάποιες προδιαγραφές και περιορισμοί που λαμβάνονται υπόψη. Αυτά σχετίζονται με αλλαγές στο φωτισμό της σκηνής, γρήγορες, ξαφνικές και πολύπλοκες κινήσεις και αλλαγές στη θέση των αντικειμένων, αλληλοεπικαλύψεις αντικειμένων καθώς και άλλα. Όταν ένα τέτοιο σύστημα αντιμετωπίσει συνθήκες διαφορετικές από αυτές για τις οποίες προγραμματίστηκε τότε υπάρχει πιθανότητα αποτυχίας του.

2. Αναγνώριση προτύπων

2.1. Αναγνώριση προτύπων

Όταν στην καθημερινή ζωή χρησιμοποιούμε τη λέξη αναγνώριση εννοούμε τον προσδιορισμό της ταυτότητας ενός αντικειμένου είτε έμψυχου είτε υλικού όπως ενός ανθρώπου, ενός ζώου, μια εικόνας, ενός φρούτου. Ο όρος αυτός χρησιμοποιείται στις περιπτώσεις που δεν είναι άμεσα αντιληπτή η ταυτότητα.

Η ταυτότητα έχει να κάνει με μία συγκεκριμένη ιδιότητα του αντικειμένου που το ξεχωρίζει από άλλα αντικείμενα και γενικότερα το εντάσσει σε μία συγκεκριμένη κατηγορία. Για παράδειγμα, όταν τα αντικείμενά μας είναι κτίσματα τότε καθένα από αυτά μπορεί να ανήκει στην κατηγορία των πολυκατοικιών ή των μονοκατοικιών το οποίο σχετίζεται με τον αριθμό ορόφων που έχει.

Σε περιπτώσεις λοιπόν που η ταυτότητα του αντικειμένου δεν είναι προφανής είναι αναγκαία η εύρεση συγκεκριμένων χαρακτηριστικών του αντικειμένου (όπως το χρώμα, π.χ. το φρούτο είναι πορτοκαλί). Η αναγνώριση είναι μία βασική λειτουργία της νοημοσύνης των έμβιων όντων. Οι επιστήμονες προσπαθούν μέσα από τις μελέτες και τις έρευνές τους να προσδώσουν αυτή τη λειτουργία και στις μηχανές.

Η αναγνώριση προτύπων είναι ο επιστημονικός κλάδος που αναπτύσσει μεθόδους και συστήματα για την ταξινόμηση των προτύπων σε κατηγορίες, οι οποίες είναι γνωστές ως κλάσεις ή τάξεις. Στη δικιά μας περίπτωση τα πρότυπα μπορεί να είναι εικόνες ή βίντεο. Ο στόχος είναι η ταξινόμηση ενός συνόλου χαρακτηριστικών (τραχύτητα, μέγεθος, καμπυλότητα, φωτεινότητα, ακμές, γωνίες) σε ένα σύνολο από κλάσεις (κλειδί αυτοκινήτου, σπιτιού, συρταριέρας) με τη χρήση ταξινομητών. Ο όρος πρότυπο προκύπτει από το γεγονός ότι αυτό το σύνολο χαρακτηριστικών αναπαρίσταται σε αυτά τα συστήματα με ένα διάνυσμα. Αυτό το διάνυσμα είναι το πρότυπο. Τα πρότυπα αναπαραστούν τα όντα-αντικείμενα.

Με τον όρο σύλληψη εννοούμε το σύνολο όμοιων προτύπων όχι ίδιων. Τα έμβια όντα, όπως ο άνθρωπος από μικρή ηλικία παρατηρούν με την όραση τα πρότυπα του πραγματικού κόσμου και με τη νοημοσύνη του δημιουργούν για αυτά τα πρότυπα συλλήψεις. Θα μπορούσαμε να πούμε ότι ένα μικρό παιδί βλέποντας για πρώτη φορά ένα αυτοκίνητο δημιουργεί με τη φαντασία του στο μυαλό του εικόνες παρόμοιες αυτοκινήτων ώστε όταν θα δει ένα διαφορετικό αυτοκίνητο θα αναγνωρίσει ότι ανήκει στην ίδια κατηγορία με το πρώτο.

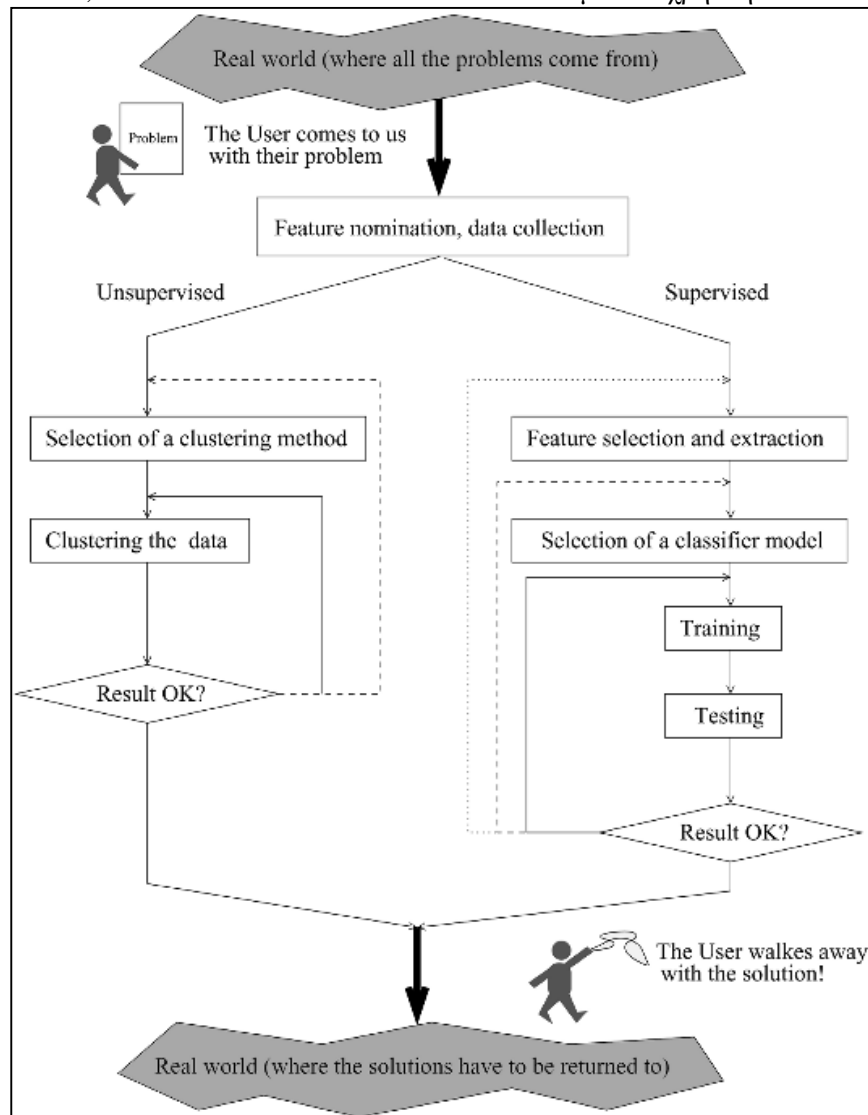
Η ελληνική γλώσσα αποτελείται από 24 γράμματα καθένα από τα οποία είναι και ένα πρότυπο άρα υπάρχουν 24 συλλήψεις. Κάθε άνθρωπος γράφει το ΒΗΤΑ και κάθε γράμμα με διαφορετικό μοναδικό τρόπο. Υπάρχουν διαφορές στο μέγεθος, στο αν είναι κεφαλαίο ή μικρό, στο μελάνι και στο χαρτί. Αυτό δε σημαίνει ότι ένας άνθρωπος που δεν έχει δει ποτέ τα γράμματά ενός άλλου δε θα μπορέσει να διαβάσει ένα κείμενό του. Φυσικά το ίδιο ισχύει και κατά τη γραφή ενός κειμένου με τον υπολογιστή αφού και ο υπολογιστής δίνει τη δυνατότητα για διαφορετικό μέγεθος, χρώμα, γραμματοσειρά και άλλα χαρακτηριστικά των γραμμάτων. Επομένως κάθε διαφορετική σύλληψη έχει χαρακτηριστικά τα οποία διαφέρουν από τις άλλες συλλήψεις. Ο προσδιορισμός αυτών των χαρακτηριστικών είναι το α και το ω για την περιγραφή και την αναγνώριση των προτύπων.

Η ανάγκη για να προσπεραστούν επίπονες ανθρώπινες εργασίες, για να βελτιωθεί το εργασιακό περιβάλλον και να αυξηθεί η παραγωγικότητα έχει οδηγήσει τους επιστήμονες τις τελευταίες δεκαετίες στην ανάπτυξη συστημάτων αναγνώρισης προτύπων από τις μηχανές. Ένα τέτοιο παράδειγμα είναι η μετατροπή ενός τυπωμένου εγγράφου σε ηλεκτρονική μορφή όπου αντιπαραθέτουμε την εργασία που απαιτείται από έναν άνθρωπο και την αντίστοιχη εργασία του υπολογιστή. Ο άνθρωπος θα πρέπει να διαβάζει το κείμενο, να αναγνωρίζει του χαρακτήρες και να τους πληκτρολογεί. Από την άλλη ο ηλεκτρονικός υπολογιστής από την άλλη θα σαρώσει ψηφιακά το έγγραφο οπότε θα προκύψει μία ψηφιακή εικόνα η οποία θα τοποθετηθεί στη μνήμη του. Στη συνέχεια με τη χρήση κατάλληλου προγράμματος αναγνώρισης χαρακτήρων (OCR: Optical Character

Recognition) θα αναγνωρίζεται η ψηφιακή μορφή κάθε χαρακτήρα που αντιστοιχεί στην σύλληψή του και θα αντιστοιχίζεται στον κατάλληλο κωδικό.

Η αναγνώριση προτύπων λοιπόν είναι μία διαδικασία αντιστοίχισης ετικετών (labels) σε αντικείμενα (objects). Τα αντικείμενα απεικονίζονται από διανύσματα μετρήσεων, τα χαρακτηριστικά. Τα θεμέλια αυτού του επιστημονικού κλάδου αναπτύχθηκαν τις δεκαετίες του 1960 και 1970. Πολλά βιβλία γράφτηκαν γύρω από αυτό το θέμα τα οποία μεταφράστηκαν σε διάφορες γλώσσες. Παρ' όλα αυτά μέχρι και σήμερα είναι ένας τομέας όπου συνεχώς αναπτύσσονται νέες μέθοδοι και θεωρίες.

Η αναγνώριση προτύπων καλείται να λύσει προβλήματα που συναντώνται στη καθημερινότητα. Η διαδικασία ενός προβλήματος αναγνώρισης προτύπων δίνεται στην Εικόνα 35. Σύμφωνα με αυτή το πρόβλημα ξεκινάει από το χρήστη ο οποίος ορίζει το πρόβλημα και το σύνολο των δεδομένων. Ο στόχος του επιστήμονα/ειδικού είναι να προσδιορίσει το πρόβλημα, να το μεταφράσει σε πρόβλημα αναγνώρισης προτύπων, να το επιλύσει και να δώσει το αποτέλεσμα στο χρήστη.



Εικόνα 3 Το πρόβλημα της αναγνώρισης προτύπων (14)

Ο αλγόριθμος που χρησιμοποιείται για το σχεδιασμό και την ανάπτυξη ενός συστήματος που θα αναγνωρίζει πρότυπα ονομάζεται ταξινομητής. Σε κάθε τέτοιο σύστημα υπάρχουν δύο βασικά στάδια, η εκπαίδευση και η ταξινόμηση.

Στο στάδιο της εκπαίδευσης καθορίζονται οι κλάσεις και οι συγκεντρώσεις των προτύπων. Κατά την εκπαίδευση προσδιορίζονται επίσης οι κατάλληλες παράμετροι του συστήματος ταξινόμησης οι

οποίες διαφέρουν από σύστημα σε σύστημα και είναι αυτές που δίνουν το μοντέλο του ταξινομητή που είναι ικανό να κατηγοριοποιήσει τα πρότυπα στις σωστές κλάσεις. Δηλαδή κάθε πρότυπο αντιστοιχίζεται στην κλάση που ανήκει. Υπάρχουν περιπτώσεις όπου στο στάδιο της εκπαίδευσης ορίζονται αρχικά για τα πρότυπα κάποιες υποθετικές κλάσεις και στη συνέχεια μέσα από την ταξινόμηση προσδιορίζονται οι σωστές κλάσεις σύμφωνα με κάποιο κριτήριο του εκάστοτε αλγορίθμου.

Υπάρχουν δύο μεγάλες κατηγορίες προβλημάτων αναγνώρισης προτύπων: με επόπτη και χωρίς επόπτη. Στα συστήματα εκπαίδευσης με επόπτη το πρόβλημα είναι να βρει το σύστημα τη δομή των δεδομένων δηλαδή να καθορίσει τις διάφορες συγκεντρώσεις τους. Για παράδειγμα, σε ένα σύστημα που δίνεται ως είσοδος ένα σύνολο από δέκα εικόνες από τις οποίες οι τρεις απεικονίζουν φρούτα, οι δύο απεικονίζουν οχήματα και οι υπόλοιπες ανθρώπους αλλά χωρίς να είναι γνωστές οι κατηγορίες και σε ποια κατηγορία ανήκει η κάθε εικόνα τότε το σύστημα θα προσδιορίζει τη δομή του. Η δομή αυτού του συστήματος είναι ότι οι δέκα εικόνες είναι χωρισμένες σε τρεις ομάδες-συγκεντρώσεις, όπου στη μία ομάδα ανήκουν τρεις εικόνες, στη δεύτερη ανήκουν δύο εικόνες και στην τρίτη ανήκουν πέντε.

Στα προβλήματα χωρίς επόπτη το σύστημα εκτός από τις ομάδες που καλείται να εντοπίσει, πρέπει να καθορίσει και τα χαρακτηριστικά που κάνουν τα αντικείμενα κάθε ομάδας όμοια μεταξύ τους και διαφορετικά με τις υπόλοιπες ομάδες. Διαφορετικοί αλγόριθμοι μπορεί να δώσουν διαφορετικές δομές για το ίδιο σύνολο δεδομένων. Δυστυχώς ή ευτυχώς δεν υπάρχει κάποιος τρόπος να συγκριθούν οι διαφορετικές δομές και να επιλεγεί η πιο σωστή αλλά είναι υποκειμενικό.

Από την άλλη πλευρά έχουμε τα συστήματα εκπαίδευσης με επόπτη όπου έχουμε το σύνολο δεδομένων όπου για κάθε αντικείμενο είναι γνωστή η κλάση στην οποία ανήκει και δίνεται με μία ετικέτα. Ο στόχος είναι το σύστημα-μηχανή να εκπαιδευτεί ώστε να μπορέσει να κάνει αυτόματα την ταξινόμηση των αντικειμένων στις διάφορες κατηγορίες. Αυτό που συμβαίνει δηλαδή είναι ότι επιλέγουμε έναν τύπο ταξινομητή που θεωρούμε ότι είναι κατάλληλος για το σύστημά μας και τον εκπαιδεύουμε με ένα σύνολο δεδομένων και τις αντίστοιχες ετικέτες τους. Με αυτό τον τρόπο κατασκευάζουμε ένα μοντέλο ταξινομητή για τα δεδομένα πρότυπα το οποίο θα μπορεί να ταξινομήσει όμοια πρότυπα αλλά όχι πανομοιότυπα με αυτά της εκπαίδευσης.

Εκπαιδεύεται επομένως ο ταξινομητής – μηχανή με ένα σύνολο από πρότυπα ώστε στη συνέχεια να μπορεί να αναγνωρίσει τις κλάσεις ή τις συγκεντρώσεις προτύπων διαφορετικών συλλήψεων δηλαδή διαφορετικά από αυτά στα οποία εκπαιδεύτηκε. Είναι επιθυμητό η εκπαίδευση να μη γίνει μόνο μία φορά αλλά να επαναληφθεί ώστε ο ταξινομητής να μπορεί να προσαρμοστεί στις διάφορες αλλαγές των προτύπων.

Κατά την ανάπτυξη συστημάτων αναγνώρισης προτύπων υπάρχουν στάδια που σχετίζονται με την εξαγωγή χαρακτηριστικών, την εκτίμηση λάθους, τη στατιστική και συντακτική αναγνώριση προτύπων. Οι εφαρμογές τέτοιων συστημάτων έχουν να κάνουν με την αναγνώριση χαρακτήρων, την ανάλυση φωνής, την αναγνώριση προσώπων, την επικοινωνία του ανθρώπου με τον ηλεκτρονικό υπολογιστή κ.α..

Στην επόμενη ενότητα θα αναφερθούμε και θα περιγράψουμε κάποιες από τις μεθοδολογίες αναγνώρισης προτύπων που χρησιμοποιούνται σε τέτοια συστήματα. Τα συστήματα αυτά είναι συστήματα ανάκτησης εικόνων βασισμένα στο περιεχόμενο (CBIR) και ανάκτησης οπτικής πληροφορίας (VIRS). Ο όρος μεθοδολογίες αναγνώρισης προτύπων έχει να κάνει με την εξαγωγή χαρακτηριστικών, την ομαδοποίηση χαρακτηριστικών, τη γένεση δεικτών σε μία βάση δεδομένων και η απόφαση για ομοιότητες στο περιεχόμενο του ερωτήματος και των στοιχείων της βάσης δεδομένων.

Όσον αφορά τα ποικίλα χαρακτηριστικά των εικόνων και η χρησιμότητά τους στην ανάκτηση εικόνων έχει γίνει μία σχετική μελέτη που περιγράφεται από τον (15). Τα χαρακτηριστικά μπορεί να σχετίζονται με το χρώμα, τα σχήματα ή ακόμα και την υφή που απεικονίζονται σε μία εικόνα.

Στις παραδοσιακές μεθόδους αναγνώρισης προτύπων η γεωμετρία, το σχήμα, η υφή και το χρώμα των αντικειμένων ήταν γνωστή εξαρχής, ήταν δεδομένη. Αλλά στην περίπτωση που έχουμε μία πραγματική εικόνα και μία βάση δεδομένων από βίντεο τότε δεν είναι γνωστά από την αρχή τα χαρακτηριστικά της εικόνας. Για το λόγο αυτό έπρεπε να αναπτυχθούν πιο αυτοματοποιημένες τεχνικές ανάκτησης εικόνων.

Εκπαίδευση μηχανών και αναγνώριση προτύπων

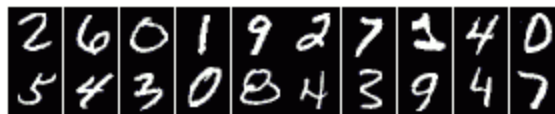
Προκειμένου να γίνει κατανοητή η λειτουργία του δικού μας συστήματος θα περιγράψουμε τον τρόπο λειτουργίας όλων των συστημάτων αναγνώρισης προτύπων.

Εφόσον ο ταξινομητής έχει εκπαιδευτεί στη συνέχεια του δίνουμε το σύνολο δεδομένων εκπαίδευσης χωρίς τις ετικέτες του και ως έξοδο μας δίνει τις ετικέτες. Συγκρίνοντας τις γνωστές ετικέτες με το αποτέλεσμα του ταξινομητή έχουν το ποσοστό ακρίβειας. Αν αυτό είναι ικανοποιητικό τότε το μοντέλο είναι δεκτό και μπορούμε να συνεχίσουμε στην αναγνώριση άλλων δεδομένων. Αν δεν γίνει δεκτό το ποσοστό επιτυχίας τότε κάνουμε πειράματα με διαφορετικό ταξινομητή μέχρι να βρούμε τον κατάλληλο για το συγκεκριμένο σύστημα.

Κάθε αντικείμενο απεικονίζεται από ένα διάνυσμα χαρακτηριστικών. Αυτά τα χαρακτηριστικά δεν είναι σχετικά με το αντικείμενο ισάξια. Κάποια από τα χαρακτηριστικά είναι χρήσιμα σε συνδυασμό με κάποια άλλα ενώ μερικά μπορεί να είναι θόρυβος. Η επιλογή και η εξαγωγή των χαρακτηριστικών είναι σημαντικές καθώς σχετίζονται άμεσα με την ποιότητα της περιγραφής των αντικειμένων.

Η επιλογή, η εκπαίδευση και ο έλεγχος του ταξινομητή αποτελούν τα θεμέλια κάθε συστήματος αναγνώρισης προτύπων με επόπτη. Μπορεί να γίνουν επαναλαμβανόμενα πειράματα όπου να αλλάζουμε εξ' ολοκλήρου τον τύπο του ταξινομητή ή να κρατάμε τον ίδιο τύπο και να αλλάζουμε τις παραμέτρους του ή να κάνουμε διαφορετική επιλογή και εξαγωγή χαρακτηριστικών. Όλα μαζί συνθέτουν την ορθή λειτουργία του συστήματος οπότε αλλαγές και διαφορετικοί συνδυασμοί όλων μπορεί να οδηγήσουν σε βέλτιστο αποτέλεσμα. Εφόσον έχουμε καταλήξει σε ένα επιθυμητό αποτέλεσμα τότε το σύστημα είναι έτοιμο να ελεγχθεί και να εφαρμοστεί από τον χρήστη σε τύπους δεδομένων για τους οποίους σχεδιάστηκε.

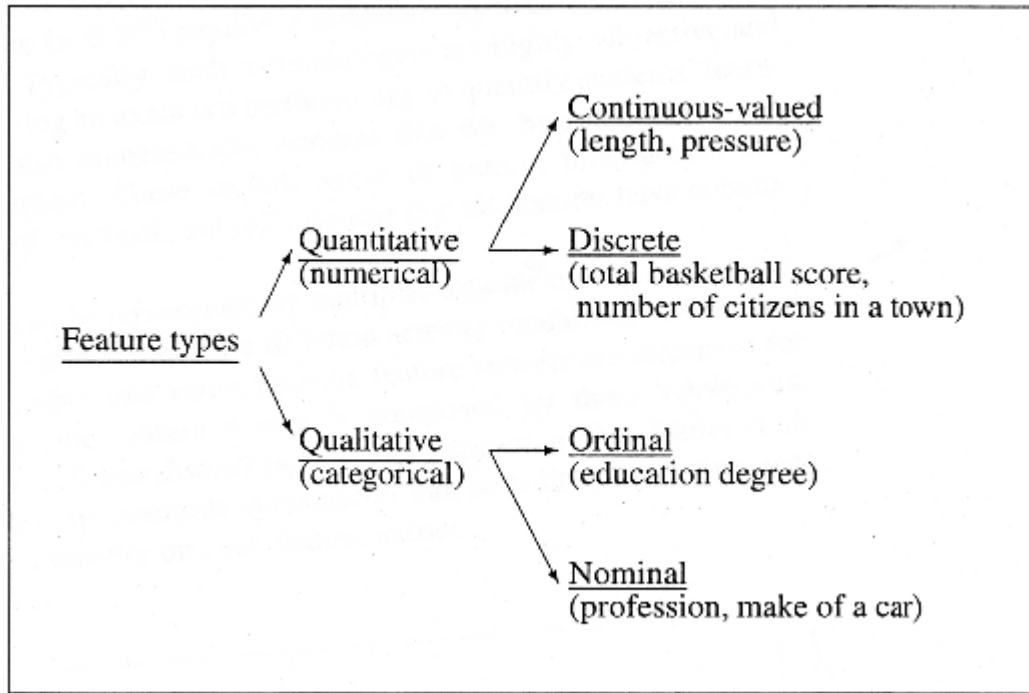
Οι κλάσεις και οι ετικέτες κλάσεων διαφέρουν από πρόβλημα σε πρόβλημα. Για παράδειγμα, σε ένα πρόβλημα όπου το σύνολο δεδομένων σχετίζεται με φωτογραφίες που απεικονίζουν η καθεμία έναν αριθμό από το 0 έως το 9, ο αριθμός των κλάσεων είναι δέκα, μία για κάθε αριθμό και κάθε φωτογραφία θα έχει μία ετικέτα η οποία θα έχει τιμή από 0 έως 9 ανάλογα με τον αριθμό που απεικονίζει (16).



Εικόνα 4 Σύνολο εικόνων που απεικονίζουν τους αριθμούς 0 έως 9

Τα χαρακτηριστικά που επιλέγονται και εξάγονται από τα αντικείμενα-πρότυπα μπορεί να είναι ποσοτικά ή ποιοτικά (Εικόνα 5). Διακριτά χαρακτηριστικά που μπορεί να έχουν πολλές πιθανές τιμές θεωρούνται ποσοτικά. Αντίθετα, τα ποιοτικά χαρακτηριστικά έχουν μικρό αριθμό πιθανών τιμών. Στη στατιστική αναγνώριση προτύπων τα χαρακτηριστικά είναι αριθμητικά.

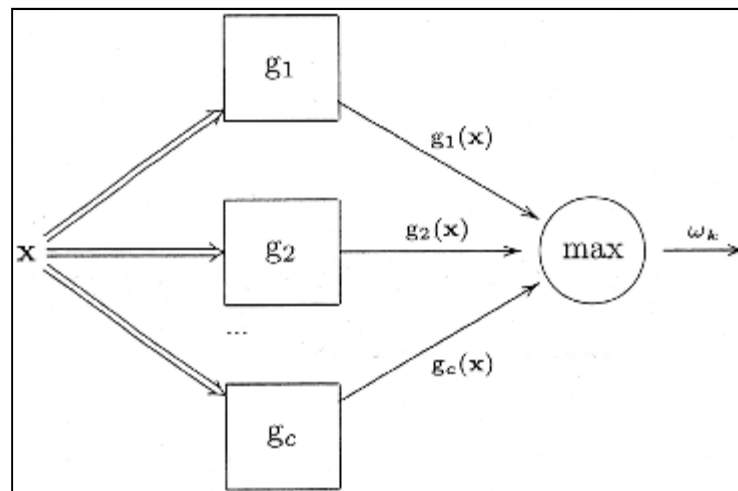
Υπάρχουν περιπτώσεις όπου ένα αντικείμενο μπορεί να αντιπροσωπεύεται από διαφορετικές κατηγορίες χαρακτηριστικών. Για παράδειγμα σε ένα βίντεο μπορεί να εξαχθούν χαρακτηριστικά που έχουν να κάνουν με κάποια χαρακτηριστικά σημεία (σημεία ενδιαφέροντος) της οπτικής πληροφορίας και χαρακτηριστικά σημεία που έχουν να κάνουν με τον ήχο. Αυτές οι δύο κατηγορίες συνθέτουν ένα διάνυσμα χαρακτηριστικών σημείων για το αντικείμενο. Σε κάποιες περιπτώσεις χρησιμοποιούνται διαφορετικοί ταξινομητές για καθένα από τα διαφορετικά διανύσματα χαρακτηριστικών και ο συνδυασμός των αποτελεσμάτων συνθέτουν το τελικό αποτέλεσμα.



Εικόνα 5 Κατηγορίες χαρακτηριστικών

Το σύνολο των δεδομένων είναι τα διανύσματα χαρακτηριστικών των αντικειμένων μαζί με τις ετικέτες τους. Το σύνολο των διανυσμάτων χαρακτηριστικών είναι η είσοδος για τον ταξινομητή. Εάν τα αντικείμενα είναι εικόνες με χειρόγραφους αριθμούς 0 έως 9 τότε σαν χαρακτηριστικά μπορεί να επιλεγούν ο αριθμός των κύκλων της εικόνας, των κάθετων και των οριζόντιων γραμμών. Τα χαρακτηριστικά όπως έχουμε πει είναι πολύ σημαντικά καθώς, μία καλή επιλογή τους θα δώσει ένα σύνολο δεδομένων για την εκπαίδευση του ταξινομητή που θα οδηγήσει στην επιτυχή δημιουργία του μοντέλου ταξινόμησης.

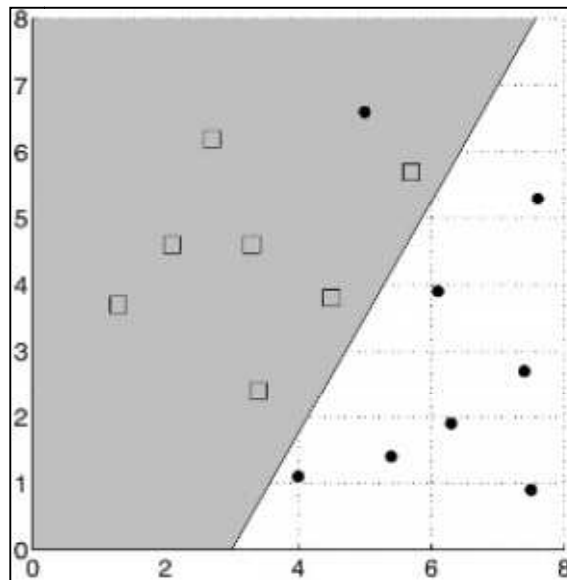
Ένας ταξινομητής είναι ουσιαστικά μία συνάρτηση ($D: R^n \rightarrow \Omega$). Στο κανονικοποιημένο μοντέλο ενός ταξινομητή(17) χρησιμοποιούνται τόσες διακρίνουσες συναρτήσεις όσες και οι κλάσεις των αντικειμένων. Το σύνολο των διακρίνουσων συναρτήσεων $G = \{g_1(x), \dots, g_c(x)\}$, όπου c ο αριθμός των κλάσεων και $g_i: R^n \rightarrow R, i = 1, \dots, c$, όπου η συνάρτηση i έχει μία τιμή που αντιστοιχεί στην κλάση i . Σε αυτή την περίπτωση το αντικείμενο x αντιστοιχεί στην κλάση της οποίας η συνάρτηση $g_i(x)$ έχει τη μεγαλύτερη τιμή από όλες (Εικόνα 6).



Εικόνα 6 Κανονικό μοντέλο ταξινομητή

Στην περίπτωση που οι κλάσεις του συνόλου δεδομένων μπορούν να διαχωριστούν μεταξύ τους με ένα υπερεπίπεδο τότε τα δεδομένα είναι γραμμικώς διαχωρίσιμα. Το υπερεπίπεδο στο χώρο των πραγματικών αριθμών είναι ένα σημείο, στο διδιάστατο χώρο είναι μία γραμμή και στο τριδιάστατο επίπεδο είναι ένα επίπεδο.

Στην Εικόνα 7 φαίνεται ένα σύνολο δεδομένων στο διδιάστατο χώρο που ανήκουν σε δύο διαφορετικές κατηγορίες. Τα διαφορετικά αντικείμενα της μίας κλάσης σημειώνονται με μαύρους κύκλους στην εικόνα ενώ της άλλης κλάσης με τετράγωνα. Σε αυτό το σύνολο δεδομένων έγινε προσπάθεια να διαχωριστούν γραμμικά με μία ευθεία αλλά όπως φαίνεται δεν είναι γραμμικώς διαχωρίσιμα καθώς ένα σημείο της μίας κλάσης ανήκει στο ημιεπίπεδο της άλλης κλάσης.

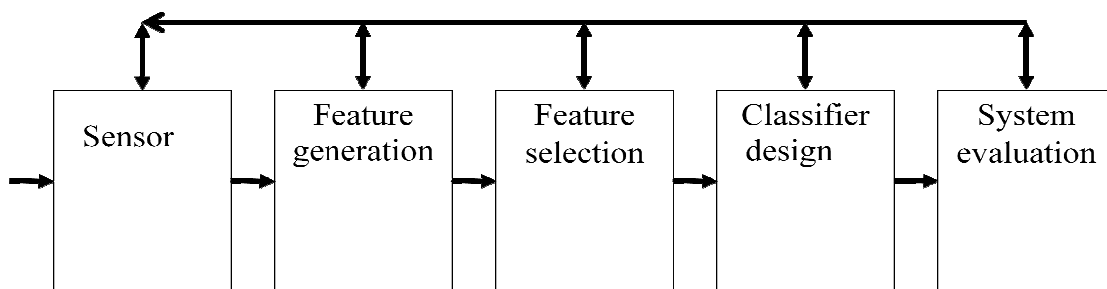


Εικόνα 7 Ένα σύνολο δεδομένων δύο κλάσεων διαχωρίζεται από μία γραμμική συνάρτηση

Το λάθος και η ακρίβεια ταξινόμησης είναι δύο μετρικές για την επίδοση του ταξινομητή. Είναι πολύ σημαντικό να γνωρίζουμε την επίδοση του ταξινομητή μας η οποία χαρακτηρίζεται κυρίως από την ακρίβεια του. Εάν μπορούσαμε να κάνουμε πειράματα για όλα τα πιθανά αντικείμενα εισόδου στον ταξινομητή τότε θα ξέραμε ακριβώς την επίδοσή του. Επειδή όμως αυτό δεν είναι δυνατό και με αυτή τη λογική άλλωστε δημιουργούμε αυτά τα συστήματα, γίνεται μία εκτίμηση της επίδοσής του.

Συμπερασματικά ένα σύστημα αναγνώρισης προτύπων φαίνεται στο διάγραμμα βαθμίδων της Εικόνα 8. Το βήματα είναι τα εξής:

1. Απόκτηση του συνόλου αντικειμένων.
2. Εξαγωγή χαρακτηριστικών.
3. Επιλογή χαρακτηριστικών.
4. Σχεδιασμός ταξινομητή.
5. Επίδοση συστήματος.



Εικόνα 8 Το διάγραμμα βαθμίδων ενός συστήματος αναγνώρισης προτύπων

Σε τέτοια συστήματα η είσοδος ένα ένα ή πολλά πολυμεσικά ή και ψηφιακά αντικείμενα είτε ήχος είτε εικόνα είτε βίντεο είτε κείμενο. Στη δική μας περίπτωση τα αντικείμενα είναι ψηφιακές εικόνες. Ο άνθρωπος βλέποντας μία εικόνα αναγνωρίζει αμέσως τι απεικονίζεται σε αυτή. Πιο συγκεκριμένα, αναγνωρίζει τα χρώματα, τα σχήματα, τα περιγράμματα, τα αντικείμενα που βρίσκονται μέσα σε αυτή ακόμα και τη διάταξή τους. Και αυτή η αναγνώριση γίνεται μέσα σε δευτερόλεπτα ίσως και κλάσματα του δευτερολέπτου.

Για τον ηλεκτρονικό υπολογιστή μια τέτοια διαδικασία δεν είναι τόσο απλή. Δεν αρκεί να δείξουμε μία εικόνα στον υπολογιστή και να την αναγνωρίσει και να την περιγράψει χωρίς να έχει προγραμματιστεί για την αναγνώριση της. Υπάρχει λοιπόν η ανάγκη η εικόνα καθώς και οι διαδικασίες που θα πρέπει να ακολουθήσει ο ηλεκτρονικός υπολογιστής για την αναγνώρισή της να μεταφραστούν σε γλώσσα μηχανής και σε μαθηματικά.

2.2. Ιστορική αναδρομή - Μεθοδολογίες αναγνώρισης προτύπων

Πολλά συστήματα πολυμέσων έχουν αναπτυχθεί κατά καιρούς με αξιοσημείωτο ποσο στο επιτυχίας. Στη συνέχεια θα περιγραφού κάποια από αυτά τα συστήματα μαζί με τις μεθόδους αναγνώρισης προτύπων που χρησιμοποιούν (18).

Εξαγωγή χαρακτηριστικών

Κατά την προ-επεξεργασία των τεχνικών αναγνώρισης προτύπων θα πρέπει να γίνει εξαγωγή χαρακτηριστικών των αντικειμένων. Υπάρχουν πολλές διαφορετικές κατηγορίες και μεθοδολογίες εξαγωγής χαρακτηριστικών. Κάθε εφαρμογή μπορεί να απαιτεί και διαφορετικά χαρακτηριστικά. Για αυτό η εκλογή του είναι πολύ σημαντική για τη μετέπειτα ορθή λειτουργία του συστήματος. Πολλές μελέτες αποκαλύπτουν το σημαντικό ρόλο των χαρακτηριστικών όπως για παράδειγμα σε εφαρμογές ρομποτικής όρασης(19).

Τα χαρακτηριστικά που είναι δυνατό να εξαχθούν χωρίζονται σε δύο κατηγορίες, σε αυτά που είναι χαμηλού επιπέδου και και στα υψηλού επιπέδου. Τα χαμηλού επιπέδου χαρακτηριστικά είναι αυτά που εξάγονται απευθείας από το αντικείμενο χωρίς να απαιτείται κάποια πληροφορία για το σχήμα του στην περίπτωση που έχουμε να κάνουμε με αντικείμενα-εικόνες. Τέτοια χαρακτηριστικά έχουν να κάνουν με τις καμπύλες και τις ακμές που παρατηρούνται σε μία εικόνα. Κάποιοι ανιχνευτές για την εξαγωγή καμπυλών και ακμών είναι οι Sobel, Harris και Canny (20), (21).

Όταν χρησιμοποιούνται τεχνικές ανίχνευσης γεωμετρικών σχημάτων στις εικόνες τότε απαιτείται η εξαγωγή χαρακτηριστικών υψηλού επιπέδου (22). Σε αυτές τις περιπτώσεις η πληροφορία είναι χωρική και αρκετές φορές χρησιμοποιείται ο μετασχηματισμός Hough. Ο απλός Hough εξάγει απλές γραμμές ενώ ο γενικευμένος μπορεί να εξάγει και κύκλους και ελλείψεις.

Στην πραγματικότητα τα πράγματα δεν είναι τόσο απλά. Στο φυσικό κόσμο οι σκηνές που απεικονίζονται σε μία φωτογραφία είναι πολύπλοκες και μπορεί να μεταβάλλονται ανάλογα με το φωτισμό και την αλληλοεπικάλυψη των αντικειμένων. Σε τέτοιες περιπτώσεις προφανώς τα χαρακτηριστικά θα είναι υψηλού επιπέδου και θα πρέπει να είναι ανθεκτικά και ανεξάρτητα σε αλλαγές του φωτισμού της σκηνής, σε περιστροφές, σε κλιμακώσεις, σε μετακινήσεις και γενικότερα σε γεωμετρικές αλλαγές. Παρακάτω αναφέρουμε ενδεικτικά έναν αλγόριθμο εξαγωγής τέτοιων χαρακτηριστικών.

Αλγόριθμος SURF (Speeded-Up Robust Features): Πρόκειται για μία τεχνική ανίχνευσης σημείων ενδιαφέροντος σε μία ψηφιακή εικόνα. Ο συγκεκριμένος αλγόριθμος βρίσκει αρχικά τα σημεία ενδιαφέροντος. Τα σημεία μοναδικότητας δε μεταβάλλονται με την αλλαγή μεγέθους ή προσανατολισμού της εικόνας και φωτισμού της σκηνής ή με την προσθήκη θορύβου.

Έχοντας λοιπόν αυτά τα σημεία με βάση τη γειτονιά τους δημιουργείται ένα χαρακτηριστικό διάλυμα για κάθε σημείο. Στη συνέχεια αυτά τα διανύσματα εξυπηρετούν στην ταύτιση κοινών

σημείων ανάμεσα σε δύο ή περισσότερες εικόνες ώστε να γίνει αναγνώριση και ταξινόμηση αντικειμένων.

Στο πρώτο βήμα ο ανιχνευτής σημείων ενδιαφέροντος SURF χρησιμοποιεί τον γρήγορο εσσιανό πίνακα (Fast Hessian Matrix) σύμφωνα με τον οποίο για κάθε σημείο της εικόνας υπολογίζεται ο εσσιανός του πίνακας (23). Με βάση τον πίνακα εξάγεται η τοποθεσία καθώς και η κλιμάκωση της εικόνας. Στη συνέχεια χρησιμοποιούνται φίλτρα 9x9, τα οποία προσεγγίζουν την παράγωγο δεύτερης τάξης της συνάρτησης Gauss και ουσιαστικά αντιπροσωπεύουν την μικρότερη δυνατή κλίμακα μεγέθους.

Τέλος, με βάση τη γειτονιά των σημείων ενδιαφέροντος και τη σύγκριση των αθροισμάτων των αποκρίσεων που δίνονται από τα παράθυρα των φίλτρων προσδιορίζεται ο ιδανικός προσανατολισμός ως το μεγαλύτερο άθροισμα. Ο προσανατολισμός αυτός χρησιμοποιείται στη συνέχεια για την τελική εξαγωγή χαρακτηριστικών η οποία προκύπτει από μια τετραγωνική περιοχή γύρω από τα σημεία ενδιαφέροντος και με ευθυγράμμιση με τον προσανατολισμό. Το διάνυσμα χαρακτηριστικών για κάθε αντικείμενο που δίνεται από τον αλγόριθμο SURF είναι 64 στοιχεία.

Ταξινομητές

Σε αυτή την ενότητα θα αναφέρουμε ορισμένους από τους ταξινομητές που χρησιμοποιούνται στα συστήματα αναγνώρισης προτύπων.

OMC (Optimal Margin Classifier) ταξινομητής

Αυτός ο ταξινομητής είναι μία μέθοδος για την κατασκευή κανόνων διαχωρισμού προβλημάτων δύο κλάσεων. Τα πρότυπα εισόδου μετασχηματίζονται αυτόματα κατά τέτοιο τρόπο ώστε να μπορούν να διαχωριστούν γραμμικά στο νέο χώρο.

Στο νέο χώρο μόνο κάποια από τα αρχικά πρότυπα είναι απαραίτητα για να καθορίσουν τα όρια απόφασης. Αυτά είναι γνωστά ως τα πρότυπα στήριξης. Μόνο αυτά τα πρότυπα αποθηκεύονται, επομένως ο OMC ταξινομητής απαιτεί λιγότερη μνήμη από άλλους ταξινομητές που απαιτούν την αποθήκευση όλων των προτύπων εκπαίδευσης. Ο ταξινομητής αυτός δημιουργήθηκε αρχικά από τους Boser, Guyon και Vapnik (24) και κάνει επιτυχή ταξινόμηση μόνο στην περίπτωση που τα δεδομένα είναι γραμμικώς διαχωρίσιμα στο νέο χώρο που προκύπτει μετά το μετασχηματισμό τους (25).

SVM Support Vector Machine (Μηχανή Διανυσματικής Στήριξης)

Οι Μηχανές Διανυσμάτων Υποστήριξης (Support Vector Machines-SVM) αποτελούν μία σύγχρονη αποτελεσματική προσέγγιση της επίλυσης ζητημάτων κατηγοριοποίησης. Η μέθοδος αναφέρεται στην εκμάθηση μηχανών. Κατά την διαδικασία της εκμάθησης (learning) σκοπός μας είναι να «μάθουμε» το σύστημα μας να σκέφτεται, δηλαδή να ξεχωρίζει, να αναγνωρίζει, να ομαδοποιεί, να κατηγοροποιεί δεδομένα. Η μεθοδολογία αυτή προέκυψε από τη βαθύτερη ανάλυση της στατιστικής θεωρίας μάθησης (statistical learning theory) που εφευρέθηκε από τον Vladimir Vapnik (26).

Οι Μηχανές Διανυσμάτων Υποστήριξης αποτελούν ένα από τα ισχυρότερα εργαλεία για την επίλυση προβλημάτων κατηγοριοποίησης και προσφέρουν μια από τις πιο ακριβείς μεθόδους. Έχουν καλό θεωρητικό υπόβαθρο, απαιτούν λίγα δεδομένα εκπαίδευσης, και δεν επηρεάζονται από το πρόβλημα των πολλών διαστάσεων. Τα SVM μπορούν να λύσουν πληθώρα προβλημάτων στην αναγνώριση προτύπων, στη βιοπληροφορική και στην εξόρυξη γνώσης από δεδομένα.

Τα SVM χαρακτηρίζεται ως μηχανή μάθησης (learning machines) συνεπώς χαρακτηρίζεται από τα ακόλουθα στάδια:

Εκπαίδευση: Στη φάση αυτή γίνονται οι υπολογισμοί των παραμέτρων του μοντέλου μάθησης με χρήση κατάλληλου συνόλου δεδομένων μάθησης.

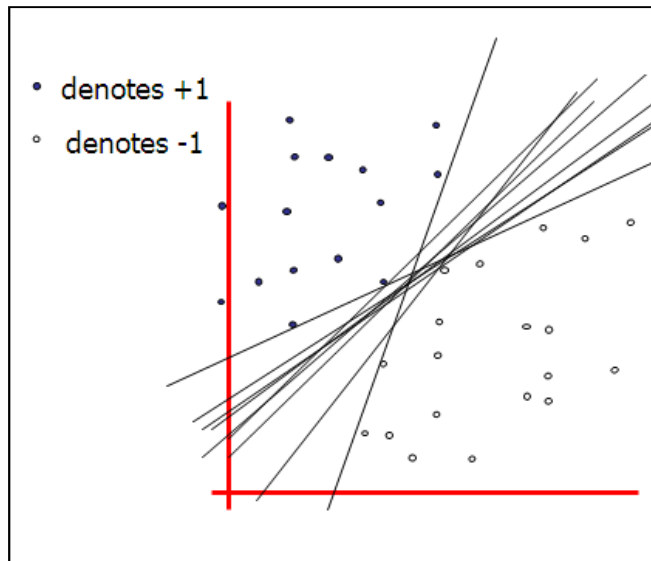
Δοκιμή: Το μοντέλο παραμέτρων (support vectors) που υπολογίστηκε δοκιμάζεται για τη δυνατότητα επιτυχημένης εκτίμησης αποτελέσματος σε ένα σετ δεδομένων που δεν έχει εκπαιδευθεί.

Εκτίμηση επίδοσης: Υπολογίζονται κατάλληλοι δείκτες επίδοσης του μοντέλου, κυρίως του ποσοστού σφάλματος, με στόχο την ιερέυνση της δυνατότητας γενίκευσης του μοντέλου.

Η μάθηση στα SVM γίνεται αφού δοθεί ολόκληρο το σύνολο εκπαίδευσης. Ο τρόπος που γίνεται η μάθηση στο SVM αποτελεί ένα από τα μειονεκτήματα του γιατί αφού τελειώσει η εκπαίδευση και βρεθούν ακόμη μερικά δείγματα εκπαίδευσης, δεν είναι δυνατόν να προστεθεί η νέα αυτή γνώση στο δίκτυο. Αυτό το μειονέκτημα, όμως, αντισταθμίζεται από τα πλεονεκτήματα των SVM και την μεγάλη ικανότητα γενίκευσής τους.

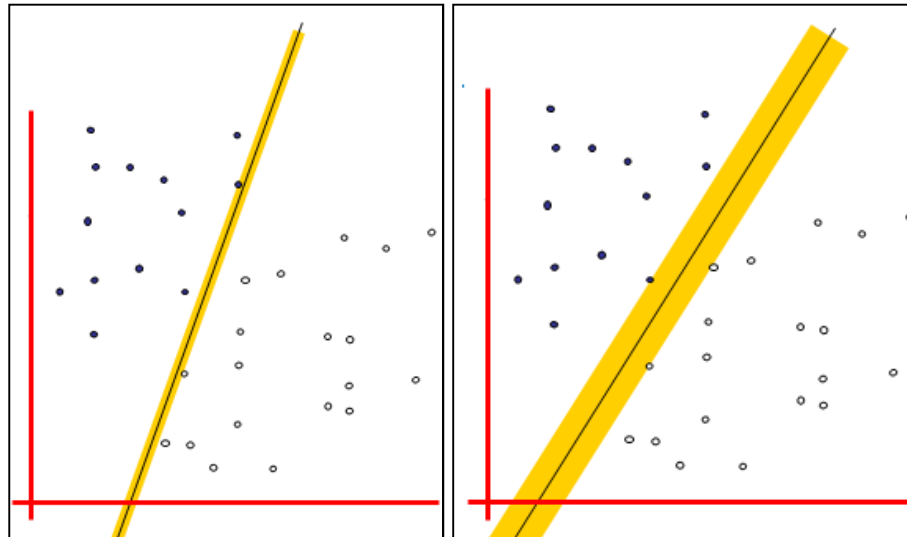
Διαχωρισμός δύο κλάσεων- Γραμμικό μοντέλο

Στη πιο απλή τους μορφή τα SVM επιλύουν γραμμικά προβλήματα δυαδικής ταξινόμησης δηλαδή προβλήματα δύο κλάσεων. Στην Εικόνα 9 παρουσιάζεται ένα πρόβλημα τέτοιας μορφής. Τα στοιχεία εκπαίδευσης είναι:



Εικόνα 9 Πρόβλημα δυαδικής ταξινόμησης

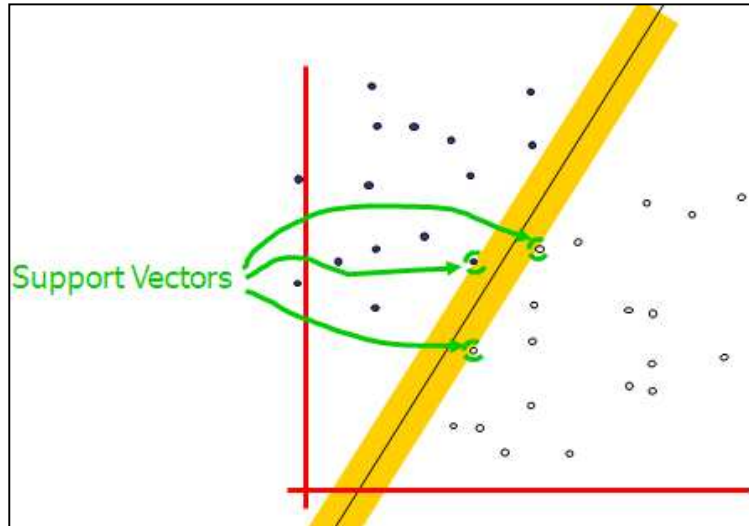
Σκοπός του SVM είναι η δημιουργία ενός υπερεπίπεδου που να χωρίζει το σύνολο των δεδομένων έτσι ώστε τα διανύσματα που ανήκουν στην ίδια κατηγορία να τοποθετούνται και στο ίδιο ημιεπίπεδο. Στο παράδειγμα της εικόνας βλέπουμε ότι τα υπερεπίπεδα που χωρίζουν τους δύο πληθυσμούς μπορεί να είναι άπειρα. Σύμφωνα με την μεθοδολογία των SVM στόχος είναι να αυξηθεί το περιθώριο (margin) ανάμεσα στα δείγματα των δύο κλάσεων όσο το δυνατόν περισσότερο.



Εικόνα 10 Περιθώριο σε δύο διαφορετικούς διαχωρισμούς των ίδιων δεδομένων

Στην Εικόνα 10 βλέπουμε το περιθώριο που δημιουργείται από δύο διαφορετικά υπερεπίπεδα. Στην συνέχεια, ο καλύτερος γραμμικός ταξινομητής είναι το υπερεπίπεδο στη μέση του περιθωρίου.

Τα σημεία επομένως που καθορίζουν το υπερεπίπεδο, είναι τα σημεία που εφάπτονται στο περιθώριο και ονομάζονται διανύσματα στήριξης (Support Vectors-Εικόνα 11). Τα σημεία αυτά είναι και δυσκολότερο να ταξινομηθούν, αφού βρίσκονται πιο κοντά στο επίπεδο απόφασης, και σε περίπτωση που μετακινηθούν, η λύση του προβλήματος αλλάζει αφού οδηγεί σε άλλο βέλτιστο υπερεπίπεδο.



Εικόνα 11 Διανύσματα υποστήριξης

Ο γραμμικός classifier που θα χρησιμοποιηθεί θα είναι του τύπου:

$$f(\bar{x}) = \bar{w}^T \bar{x} + b (= w_1 x_1 + w_2 x_2 + w_3 x_3 + \dots + w_n x_n + b)$$

Όπου \bar{w} το διάνυσμα βαρών, και b το bias.

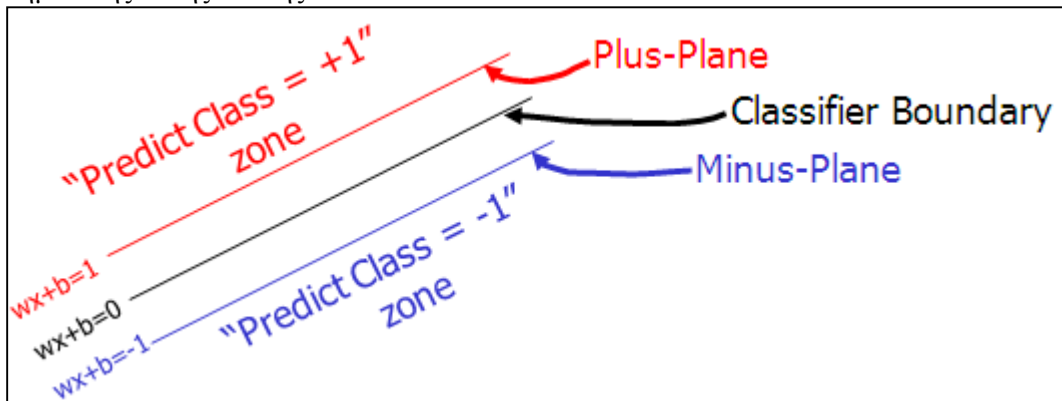
Τότε η ταξινόμηση γίνεται με βάση τον τύπο:

$$\hat{y} = \text{sign}(f(x)) = \begin{cases} +1 & f(x) > 0 \\ -1 & f(x) < 0 \end{cases}$$

Για διαφορετικά w και b μπορούμε να έχουμε και διαφορετικές λύσεις. Πιο συγκεκριμένα, μπορούμε να χρησιμοποιήσουμε διάφορα a έτσι ώστε να παράγονται πολλές διαφορετικές λύσεις:

$$\hat{y} = \text{sign}(a(\bar{w}^T \bar{x} + b)) = \text{sign}(\bar{w}^T \bar{x} + b)$$

Ο SVM δημιουργεί δύο ημιεπίπεδα, στο ένα θα βρίσκονται τα σημεία της μίας κλάσης και στο άλλο τα σημεία της άλλης κλάσης.



Εικόνα 12 Τα υπερεπίπεδα και τα ημιεπίπεδα του χώρου στον οποίο εφαρμόζεται ο SVM

Όπως βλέπουμε και στην Εικόνα 12 έχουμε δύο επίπεδα, το θετικό και το αρνητικό. Για τα σημεία που βρίσκονται πάνω στο θετικό επίπεδο ισχύει:

$$x:w.x + b = +1 \quad (1)$$

Αντίστοιχα για τα σημεία που βρίσκονται στο αρνητικό επίπεδο ισχύει:

$$x:w.x + b = -1 \quad (2)$$

Για τα άλλα σημεία ισχύει:

$$+1 \text{ Αν } w.x + b \geq 1$$

$$-1 \text{ Αν } w.x + b \leq -1$$

Έτσι το βέλτιστο υπερεπίπεδο περιγράφεται από την σχέση:

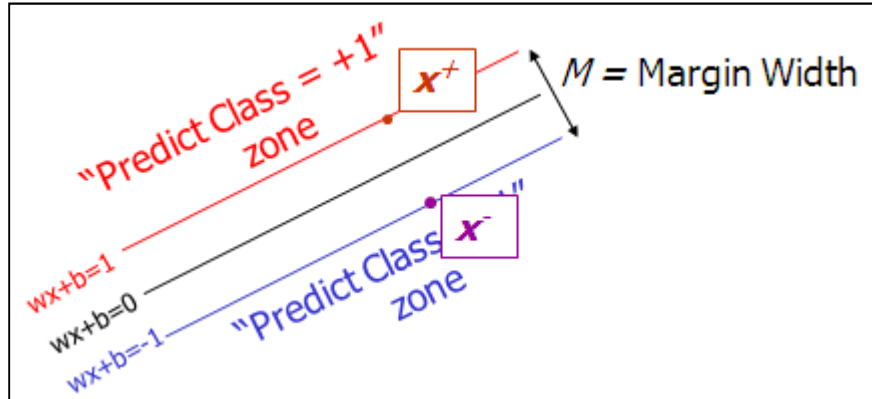
$$\bar{w}^T \bar{x} + b = 0$$

Στην συνέχεια θα δούμε πως μπορούμε να υπολογίσουμε το βέλτιστο μήκος περιθωρίου συναρτήση του w και του b , λαμβάνοντας υπόψιν ότι το w είναι μια διανυσματική κάθετος στο υπερεπίπεδο.

Έστω ότι έχουμε δύο σημεία το x^- και το x^+ . Το x^- είναι ένα σημείο που βρίσκεται πάνω στο αρνητικό επίπεδο και x^+ είναι ένα σημείο που βρίσκεται στο θετικό επίπεδο, τότε ισχύει η παρακάτω σχέση:

$$x^+ = x^- + \lambda w \quad (3)$$

Αυτό ισχύει, γιατί όπως είπαμε το w είναι μια διανυσματική κάθετος στο υπερεπίπεδο, οπότε για να φτάσουμε από το x^- στο x^+ τότε διανύουμε κάποια απόσταση στην κατεύθυνση του w .



Εικόνα 13 Παράδειγμα εφαρμογής του SVM και δύο σημείων

Επιπλέον ισχύει ότι:

$$|x^+ - x^-| = M \quad (4)$$

Όπου M το περιθώριο.

Συνδυάζοντας τις συναρτήσεις (1) και (2) έχουμε

$$w \cdot (x^- + \lambda w) + b = w \cdot x^- + b + \lambda w \cdot w = 1$$

$$\Rightarrow -1 + \lambda w \cdot w = 1 \Rightarrow \lambda = \frac{2}{w \cdot w} \quad (5)$$

$$M = |x^+ - x^-| = |\lambda w| = \lambda |w| = \lambda \sqrt{w \cdot w} = \frac{2\sqrt{w \cdot w}}{w \cdot w} = \frac{2}{w \cdot w}$$

Η μεγιστοποίηση του περιθωρίου είναι ισοδύναμη με την ελαχιστοποίηση $w \cdot w$ (η σταθερά 2 αγνοείται). Το τελικό πρόβλημα είναι να ελαχιστοποιήσουμε το $w \cdot w$, αλλά και ταυτόχρονα να ελαχιστοποιήσουμε την απόσταση των σημείων που δεν βρίσκονται στην σωστή τους θέση, από την σωστή τους θέση.

Έστω ότι έχουμε $R(x_k, y_k)$ σημεία, όπου $y_k = \pm 1$

Συγκεκριμένα οι περιορισμοί ορίζονται ως:

$$w \cdot x_k + b \geq 1 - \epsilon_k \text{ αν } y_k = 1$$

$$w \cdot x_k + b \leq 1 - \epsilon_k \text{ αν } y_k = -1$$

Οι δύο αυτοί περιορισμοί μπορούν να γραφτούν ως ένας, ο οποίος θα έχει την παρακάτω μορφή.
 $y_k(w^T x_k + b) \geq 1$

Η lagrangian συνάρτηση που χρησιμοποιείται για την επίλυση του προβλήματος είναι :

$$L(w, b, \alpha) = \frac{1}{2} w^T w - \sum_{i=1}^N \alpha_i [y_i (w^T x_i + b) - 1], \text{ με } \alpha_i \geq 0, \forall i$$

Όπου $\alpha_1, \alpha_2, \dots, \alpha_n$ είναι οι Lagrange πολλαπλασιαστές και $\alpha = [\alpha_1, \alpha_2, \dots, \alpha_n]^T$.

Για τους πολλαπλασιαστές Lagrange ισχύει επιπλέον:

Εάν

$y_i(w^T x_i + b) > 1$, η τιμή του α_i που μεγιστοποιεί την $L(w, b, \alpha)$ είναι $\alpha_i = 0$.

Εάν $y_i(w^T x_i + b) < 1$, η τιμή του α_i που μεγιστοποιεί την $L(w, b, \alpha)$ είναι $\alpha_i = +\infty$.

Αφού w και b προσπαθούν να ελαχιστοποιήσουν την $L(w, b, \alpha)$, θα πρέπει να αλλάξουν με τέτοιο τρόπο ώστε να κάνουν το $y_i(w^T x_i + b)$ τουλάχιστον ίσο με +1.

Συνθήκες βελτιστοποίησης:

$$\frac{\partial L}{\partial w_j} = 0, \quad \forall j$$

$$\frac{\partial L}{\partial \alpha_i} = 0, \quad \forall i$$

Απαραίτητες Συνθήκες

$$\mathbf{w} = \sum_{i=1}^N \alpha_i y_i \mathbf{x}_i$$

$$\sum_{i=1}^N \alpha_i y_i = 0$$

Το δυικό πρόβλημα βελτιστοποίησης μπορεί να εκφραστεί ως εξής:

$$\sum_i \alpha_i - \frac{1}{2} \sum_i \sum_j \alpha_i \alpha_j y_i y_j \mathbf{x}_i^T \mathbf{x}_j$$

Με την συνθήκη:

$$\sum_{i=1}^N \alpha_i y_i = 0, \quad \alpha_i \geq 0.$$

Τελική πρόβλεψη:

Δεδομένων των τιμών $\alpha_1, \alpha_2, \dots, \alpha_n$ που λαμβάνονται από την επίλυση του δυικού προβλήματος, ο τελικός SVM predictor μπορεί να εκφραστεί ως:

$$f(\mathbf{x}) = \mathbf{w}^T \mathbf{x}_i + b = \sum_{i=1}^n \alpha_i y_i \mathbf{x}_i^T \mathbf{x} + b$$

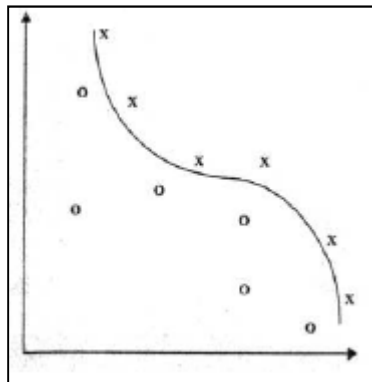
Όπου

$$b = \frac{1}{|I_{\text{support}}|} \sum_{i \in I_{\text{support}}} \left(y_i - \sum_j \alpha_j y_j \mathbf{x}_j^T \mathbf{x}_i \right)$$

και I_{support} είναι το σύνολο των διανυσμάτων στήριξης.

Μη Γραμμικό Μοντέλο SVM

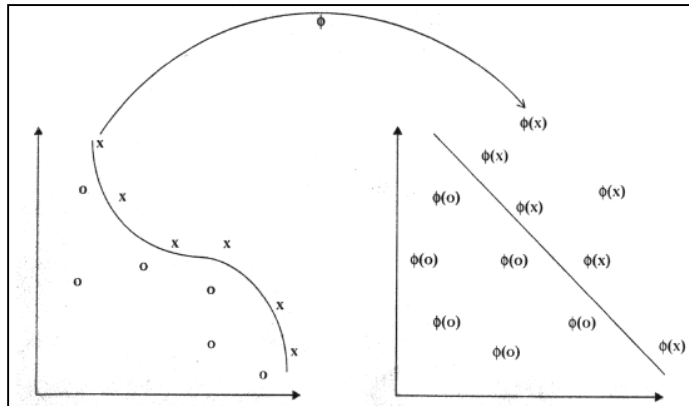
Στην προηγούμενη ενότητα είδαμε πως εφαρμόζεται η τεχνική SVM σε δεδομένα τα οποία είναι γραμμικά ευδιαχώριστα. Τις περισσότερες φορές σε πραγματικές εφαρμογές, είναι πολύ σπάνιο να έχουμε ευδιαχώριστα δεδομένα. Τα περισσότερα δεδομένα έχουν την μορφή της εικόνας ταδε:



Εικόνα 14 Παράδειγμα εφαρμογής του μη γραμμικού μοντέλου SVM

Kernel-Trick

Σε αυτό το τμήμα θα δούμε συναρτήσεις kernel με βάση τις οποίες είμαστε σε θέση να λύσουμε μη γραμμικά προβλήματα. Μέσα από kernel συναρτήσεις τα δεδομένα μεταφέρονται σε ένα νέο χώρο διαφορετικής διάστασης, όπου εκεί τα δεδομένα είναι γραμμικά διαχωρίσιμα, και έτσι μπορούμε να λύσουμε μη γραμμικά προβλήματα. Ένας τέτοιος μετασχηματισμός απεικονίζεται στην Εικόνα 15.



Εικόνα 15 Μετασχηματισμός μη γραμμικά διαχωρίσιμων δεδομένων σε νέο χώρο όπου είναι γραμμικά διαχωρίσιμα

Μετά την εφαρμογή του kernel μετασχηματισμού, τα πράγματα είναι απλά, αφού πλέον έχουμε να κάνουμε με ένα κλασικό ζήτημα γραμμικής παρεμβολής. Όπως περιγράψαμε και στο προηγούμενο κεφάλαιο, το πρόβλημα μπορεί να λυθεί βρίσκοντας την εξίσωση του υπερεπιπέδου που διαχωρίζει τα δεδομένα. Το υπερεπίπεδο που θα κατασκευαστεί, θα είναι μια γραμμική συνάρτηση, όχι πια των αρχικών δεδομένων, αλλά των εικόνων τους στον νέο χώρο.

Συναρτήσεις kernel

Στη συνέχεια κάθε συνάρτηση kernel θα συμβολίζεται με k , η οποία για τις συνιστώσες x, z που ανήκουν στον αρχικό χώρο, υπολογίζει το εσωτερικά γινόμενο των μετασχηματισμένων πλέον δεδομένων $\varphi(x), \varphi(z)$ τα οποία είναι απαραίτητα για την υλοποίηση του αλγορίθμου.

$$k(x, z) = \langle \varphi(x), \varphi(z) \rangle$$

Πίνακας Kernel (27)

Στις περισσότερες περιπτώσεις τα x, z που είναι τα δεδομένα είναι πίνακες με αριθμό γραμμών ίσο με το πλήθος των δειγμάτων και αριθμό στηλών ίσο με τις συνιστώσες τους (αφού πρόκειται για ανύσματα). Επομένως υπολογίζοντας τα εσωτερικά γινόμενα για όλους τους δυνατούς συνδυασμούς των ανυσμάτων όλων των δειγμάτων δημιουργείται ένας νέος πίνακας που ονομάζουμε πίνακα kernel (kernel matrix).

K	1	2	...	l
1	$k(x_1, x_1)$	$k(x_1, x_2)$...	$k(x_1, x_l)$
2	$k(x_2, x_1)$	$k(x_2, x_2)$...	$k(x_2, x_l)$
.
.
.
l	$k(x_l, x_1)$	$k(x_l, x_2)$...	$k(x_l, x_l)$

Εικόνα 16 Πίνακα kernel

Το πρώτο στοιχείο του πίνακα kernel (Εικόνα 16) $k(x_1, x_1)$ προέρχεται από τον υπολογισμό του εσωτερικού γινομένου των μετασχηματισμένων δεδομένων $\varphi(x_1), \varphi(x_1)$. Παρόμοια υπολογίζονται και τα υπόλοιπα στοιχεία του πίνακα.

Αυτός ο πίνακας παίζει σημαντικότατο ρόλο και θα μπορούσαμε να τον αποκαλέσουμε πυρήνα των μεθόδων kernel καθώς «συνδέει» τα δεδομένα εισόδου με τους αλγόριθμους και επιπλέον περιέχει όλες τις απαιτούμενες πληροφορίες για τον αλγόριθμο.

Άλλη σημαντική ιδιότητά του είναι ότι είναι θετικά ημι-ορισμένος. Αυτό από τη μια εξασφαλίζει την ύπαρξη κάποιου χώρου απεικόνισης όπου είναι δυνατός ο γραμμικός διαχωρισμός των δεδομένων. Από την άλλη η παραπάνω ιδιότητα επιτρέπει να χρησιμοποιούμε ως δεδομένα εισόδου πέρα από πραγματικά διανύσματα, σειρές χαρακτήρων (strings), εικόνες, χρονικές ακολουθίες κ.α. Και αυτό γιατί ο πίνακας kernel που θα υπολογιστεί με τα νέα δεδομένα θα εξακολουθεί να είναι θετικά ημι-ορισμένος. Και σε αυτές τις περιπτώσεις τα αρχικά δείγματα προβάλλονται πρώτα σε νέο χώρο όπου έχουν διαχωριστεί μεταξύ τους.

Ένα πολύ σημαντικό χαρακτηριστικό των συναρτήσεων kernel είναι ότι δεν χρειάζεται να γνωρίζουμε την φ για τον υπολογισμό των εσωτερικών δεδομένων. Για παράδειγμα έστω ότι έχουμε ένα διδιάστατο διάνυσμα της μορφής $x = (x_1, x_2)$. Στην συνέχεια θα εφαρμόσουμε σε αυτό το διάνυσμα μια συνάρτηση απεικόνισης φ . Στην περίπτωση που η συνάρτηση kernel είναι πολυωνυμική τότε μπορεί να έχει την παρακάτω μορφή:

$$\varphi: X = (x_1, x_2) \rightarrow \varphi(x) = (x_1^2, x_2^2, \sqrt{2}x_1x_2) \in F = R^3$$

Με την εφαρμογή της παραπάνω συνάρτησης μετασχηματισμού, τα δεδομένα μεταφέρονται από τον διδιάστατο χώρο στον τρισδιάστατο. Σύμφωνα με τις συναρτήσεις kernel, υπολογίζουμε το εσωτερικό γινόμενο των μετασχηματισμένων δεδομένων, δηλαδή $\langle \varphi(x)\varphi(z) \rangle$, όπου x και z είναι δύο τυχαία σημεία του αρχικού χώρου. Στην συνέχεια θα δούμε με τι ισούται αυτό το εσωτερικό γινόμενο.

$$\begin{aligned} \langle \varphi(x), \varphi(z) \rangle &= \langle (x_1^2, x_2^2, \sqrt{2}x_1x_2), (z_1^2, z_2^2, \sqrt{2}z_1z_2) \rangle = \\ &= x_1^2z_1^2 + x_2^2z_2^2 + 2x_1x_2z_1z_2 = \\ &= (x_1z_1 + x_2z_2)^2 = \\ &= \langle x, z \rangle^2 \end{aligned}$$

Συμπεραίνουμε λοιπόν ότι δεν χρειάζεται να ξέρουμε την συνάρτηση απεικόνισης φ , αλλά μόνο τον υπολογισμό του εσωτερικού γινομένου. Άλλες μορφές kernel με βάση τις συναρτήσεις που μπορούν να χρησιμοποιηθούν εκτός από την πολυωνυμική είναι η εκθετική, η Gaussian, η γραμμική, η rbf και η σιγμοειδής.

Συστήματα ανάκτησης εικόνων

Content based retrieval engine (CORE)

Ένα σύστημα πολυμεσικών πληροφοριών αποτελείται από πολυμεσικά αντικείμενα τα και καθορίζεται από τη δημιουργία, τη διαχείριση, την ανάκτηση, την επεξεργασία, την παρουσίαση και τις συναρτήσεις χρήσης. Ένα πολυμεσικό αντικείμενο είναι πολύπλοκο και έχει ποικίλες ερμηνείες.

Το σύστημα CORE βασίζεται στο περιεχόμενο της εικόνας και ως σύστημα ανάκτησης υπάρχουν τέσσερις διακριτές κατηγορίες. Η ανάκτηση μπορεί να γίνει μέσω της οπτικής πληροφορίας, μέσω της σύγκρισης και της ομοιότητας, με βάση την γλωσσική πληροφορία ή με τη χρήση της ασαφούς λογικής. Και οι τέσσερις αυτές κατηγορίες συναντώνται στην αρχιτεκτονική του CORE. Έχει αναπτυχθεί έτσι ώστε να χρησιμοποιεί χρωματικά χαρακτηριστικά και στην περίπτωση ήχου λεξικά και φωνητικά χαρακτηριστικά για την ανάκτηση των αντικειμένων.

Επιπλέον το σύστημα αυτό υποστηρίζει και ένα βρόχο ανάδρασης ώστε να γίνεται ταίριασμα με βάση την ασαφή λογική. Αυτό το επιπλέον κομμάτι της τεχνική αυτής βελτιώνει την αξιοπιστία του συστήματος. Δύο μεγάλα συστήματα έχουν βασιστεί στην τεχνική CORE, το Computer Aided Facial Image Inference and Retrieval (CAFIIR) και το System for Trademark Archival and Retrieval (STAR) (28).

Webseek

Το Webseek είναι μία πρωτότυπη μηχανή αναζήτησης εικόνων και βίντεο από το διαδίκτυο τις οποίες τις κατηγοριοποιεί. Η αναζήτηση και η προεπισκόπηση των αντικειμένων γίνεται με τη χρήση ποικίλων μεθόδων ανάκτησης που χρησιμοποιούν το χρώμα, την υφή και άλλες ιδιότητες. Τα ερωτήματα του χρήστη μπορεί να γίνονται με κείμενο ή και με εικόνα. Και σε αυτή την τεχνική όπως και στην CORE που περιγράφηκε παραπάνω χρησιμοποιούνται τεχνικές ανάδρασης ώστε να ενισχυθεί η αποτελεσματικότητα του συστήματος.

VideoQ

Είναι επίσης ένα σύστημα αναζήτησης στο διαδίκτυο και συγκεκριμένα αναζήτησης βίντεο. Οι χρήστες μπορούν να εντοπίσουν συμπεσιμένα βίντεο του ενδιαφέροντός τους με το σύστημα VideoQ του οποίου η τεχνική βασίζεται στη χρήση οπτικών και χρονικών χαρακτηριστικών και πιο συγκεκριμένα, χρώματος, υφής, σχήματος και κίνησης.

Blobworld

Πρόκειται για ένα σύστημα ανάκτησης εικόνων εντοπίζοντας εικόνες όμοιες με δεδομένα αντικείμενα. Αυτό το σύστημα ψάχνει όχι συνολικά μια εικόνα αλλά τμήματά της. Δηλαδή η εικόνα αντιμετωπίζεται ως ένα σύνολο από τμήματα όπου κάθε τμήμα αποτελείται από σημεία με όμοιο χρώμα και υφή.

Στο σύστημα Blobworld κάθε εικονοστοιχείο περιγράφεται από ένα διάνυσμα οχτώ χαρακτηριστικών τα οποία έχουν να κάνουν με το χρώμα στο $L^*a^*b^*$ χρωματικό μοντέλο, την υφή, την ανισοτροπία, την πολικότητα και τη θέση του (x, y) . Η εικόνα αντιμετωπίζεται ως ένα μείγμα από Γκαουσιανές και γίνεται κατάτμησή της σε περιοχές χρησιμοποιώντας τον αλγόριθμο expectation-minimization. Επομένως ο χρήστης μπορεί να χρησιμοποιήσει απλά ερωτήματα εντοπισμού εικόνων ομοίων με ένα επιλεγμένο τμήμα εικόνας ή και πιο πολύπλοκα ερωτήματα.

MARS (Multimedia analysis and retrieval system) (29)

Το MARS επιτρέπει στο χρήστη να κάνει ερωτήματα για να ανάκτηση εικόνων τα οποία σχετίζονται με το χρώμα, την υφή, το σχήμα καθώς και χαρακτηριστικά διάταξης. Για τα χρωματικά χαρακτηριστικά χρησιμοποιείται διδιάστατο ιστόγραμμα για το χρωματικό μοντέλο HSV. Η τραχύτητα, η αντίθεση και κατευθυνση των pixel χρησιμοποιούνται για να δώσουν την πληροφορία της υφής, ενώ οι τροποποιημένοι περιγραφείς fourier (MFD) δίνουν τα χαρακτηριστικά του σχήματος. Όλα τα παραπάνω χαρακτηριστικά της εικόνας είναι απαραίτητα για τα ερωτήματα διάταξης.

Μία από τις μεθόδους ανάκτησης στις οποίες βασίζεται το MARS είναι η Fuzzy Boolean Retrieval, όπου η απόσταση ανάμεσα στην εικόνα του ερωτήματος και στις εικόνες τις βάσεις δεδομένων σχετίζεται με το βαθμό του ασαφούς συνόλου εικόνων που ταιριάζουν με το δεδομένο χαρακτηριστικό. Από την άλλη υπάρχει και η Probabilistic Boolean Retrieval μέθοδος ανάκτησης από το MARS το οποίο συγκρίνει την εικόνα που δίνεται ως αποτέλεσμα από τη βάση δεδομένων με την εικόνα ερωτήματος και δίνει την πιθανότητα αυτές οι δύο να ταιριάζουν. Το MARS περιλαμβάνει τεχνικές ανάδρασης ώστε να δίνει καλύτερα αποτελέσματα.

PicToSeek(30)

Το PicToSeek σύστημα κάνει ανάκτηση εικόνων βασισμένη στο περιεχόμενο. Το διάνυσμα χαρακτηριστικών βασίζεται στην χρωματική μεταβλητότητα, στον κορεσμό, στη δυνατότητα αλλαγών και στο φόντο. Ο σχεδιασμός και η ανάπτυξη του PicToSeek βασίστηκε στο γεγονός ότι τα παλαιότερα συστήματα αυτής της κατηγορίας που βασίζονται στη χρωματική πληροφορία των εικόνων δε λάμβαναν υπόψη τη θέση της κάμερας, τη γεωμετρία και τις συνθήκες φωτισμού.

Όσον αφορά στη δύναμη αλλαγής του χρώματος αυτή καθορίζεται από τον αριθμό αποχρώσεων στην εικόνα. Το χρωματικό μοντέλο που χρησιμοποιείται είναι το HSV επειδή είναι κατάλληλο για απεικόνιση της έντασης. Τεχνικές ειδώλων και φωτομετρικών χρωματικών χαρακτηριστικών βοηθούν στην ανάκτηση εικόνων που απεικονίζουν παρόμοια αντικείμενα υπό διαφορετικές συνθήκες φωτισμού.

Πειράματα που έγιναν στο PicToSeek έδειξαν ότι η χρήση ετεροσυσχετίσεων ιστογραμμάτων έδωσαν τα καλύτερα αποτελέσματα. Από την άλλη για το ταίριασμα τμημάτων των εικόνων το σύστημα παρέχει τεχνικές για ενεργά χρωματικά περιγράμματα (snakes). Συμπερασματικά, το σύστημα αυτό παρέχει ένα συνδυασμό χαρακτηριστικών χρώματος, υφής και σχήματος για την ανάκτηση εικόνων.

C-BIRD (Content-based image retrieval in digital libraries)

Το C-BIRD σύστημα παρουσιάζει μία προσέγγιση ανάκτησης εικόνων προσπερνώντας το πρόβλημα των αλλαγών φωτισμού βασισμένο στην κανονικοποίηση των χρωματικών καναλιών. Η πληροφορία χρώματος η οποία είναι τριών επιπέδων λόγω των τριών καναλιών R, G και B μειώνεται σε ένα μοναδικό επίπεδο λαμβάνοντας τις μοναδιαίες τιμές πάνω στους τρεις χρωματικούς άξονες.

Αν και αυτός ο μετασχηματισμός δε μεταβάλλει την πληροφορία του φωτισμού, αποτυγχάνει να δώσει τις χρωματικές αλλαγές που γίνονται λόγω φωτισμού. Αυτό επιτυγχάνεται χρησιμοποιώντας

την L2 κανονικοποίηση στα R, G και B διανύσματα και χρησιμοποιώντας τις τιμές και των τριών συνιστωσών για κάθε εικονοστοιχείο.

Το ιστόγραμμα χρησιμοποιείται για το ταίριασμα των εικόνων και για να μειωθεί το μέγεθος του χρησιμοποιείται ο διακριτός μετασχηματισμός συνημιτόνου. Για την υφή χρησιμοποιούνται οι ακμές και η κατεύθυνση σημείων ενώ για το σχήμα γίνεται χρήση του γενικευμένου μετασχηματισμού Hough.

QBIC (Query-by-image-content) (31)

Το QBIC αναπτύχθηκε από την IBM και δίνει τη δυνατότητα στους χρήστες να ψάχνουν, να βλέπουν και να ανακτούν εικόνες, γραφικά και βίντεο από μεγάλες διαδικτυακές συλλογές. Το σύστημα λαμβάνει το ερώτημα του χρήστη και εξάγει οπτικά χαρακτηριστικά σε σχέση με το χρώμα, τη διάταξη και την υφή τα οποία προσπαθεί να ταιριάξει με μία βάση δεδομένων από χαρακτηριστικά που έχουν υπολογιστεί εξαρχής. Ένα ιδιαίτερο χαρακτηριστικό που χρησιμοποιεί αυτό το σύστημα είναι η χρήση εικόνων της κλίμακας του γκρι όπου η χωρική κατανομή του γκρι στην υφή και στις ακμές χρησιμοποιούνται. Το διάνυσμα των χαρακτηριστικών έχει 100 τιμές δηλαδή είναι διάστασης 100.

Infomedial digital video library

Το πρόγραμμα Infomedial digital video library αναπτύσσεται στο Carnegie Mellon University και είναι μία βιβλιοθήκη ψηφιακού βίντεο που είναι σε εξέλιξη. Οι σχεδιαστές και κατασκευαστές αυτής της βιβλιοθήκης ανέπτυξαν μεθόδους ώστε από κάθε βίντεο να εξάγουν μία σύνοψή του που να το χαρακτηρίζει.

Όσον αφορά στην ηχητική πληροφορία και πιο συγκεκριμένα στη λεκτική, είναι δυνατόν να εξαχθούν από το βίντεο κάποιες λέξεις κλειδιά. Στη συνέχεια κάθε βίντεο της βάση δεδομένων απεικονίζεται από ένα σύνολο αντιπροσωπευτικών πλάνων του αρχικού βίντεο τα οποία κρίνονται σημαντικής δραστηριότητας. Τέτοια πλάνα μπορεί να είναι αυτά όπου κόβεται από τομα μία σκηνή ή όταν αλλάζει η σκηνή ή όταν κινείται γρήγορα η κάμερα ή πλάνα όπου ακούγονται οι λέξεις κλειδιά.

2.3. Σύστημα αναγνώρισης προτύπων

Το σύστημα που έπρεπε να σχεδιάσουμε και να αναπτύξουμε εμείς σχετίζεται με τις ανάγκες και την τιμολόγηση στα διόδια με αυτόματο τρόπο. Ο στόχος ήταν να δημιουργηθεί ένα λογισμικό το οποίο να λαμβάνει ως είσοδο εικόνες σχετικά με τα οχήματα που εισέρχονται στο σταθμό των διοδίων και με βάση το είδος του οχήματος και της ταμπέλας του να δίνει ως έξοδο το χρηματικό ποσό που θα πρέπει να καταβάλλει ο οδηγός.

Με ένα τέτοιο σύστημα δεν θα είναι απαραίτητη η ύπαρξη υπαλλήλων στα διόδια που να φροντίζουν για την ορθή πληρωμή των οδηγών ώστε να περάσουν από τα διόδια. Θα υπάρχει μηχανισμός ο οποίος θα έχει κάμερες ώστε να φωτογραφίζει τα οχήματα και με τη χρήση του λογισμικού που αναπτύξαμε θα μπορεί να ελέγχει αν το χρηματικό ποσό που καταβάλλει ο χρήστης είναι το νόμιμο ώστε να του δοθούν πιθανώς ρέστα και να ανοίξει η ράμπα ώστε να συνεχίσει ο οδηγός τη διαδρομή του.

Ακολουθεί αναλυτική περιγραφή του αλγορίθμου και του λογισμικού που αναπτύχθηκε γνωρίζοντας πλέον από τις προηγούμενες ενότητες τη γενική μεθοδολογία ενός συστήματος αναγνώρισης προτύπων.

Είσοδος

Η βασική είσοδος του συστήματος είναι η βάση δεδομένων με τις εικόνες από τα οχήματα και τις ταμπέλες. Έχουμε ουσιαστικά δύο φακέλους με εικόνες:

1. Ο ένας φάκελος έχει εικόνες οχημάτων και συγκεκριμένα αυτοκινήτων, μοτοσυκλετών, λεωφορείων και φορτηγών οι οποίες είναι οργανωμένες σε τέσσερις φακέλους ανάλογα με το είδος του οχήματος που απεικονίζουν (Εικόνα 17).
2. Ο άλλος φάκελος περιέχει εικόνες με ταμπέλες οχημάτων που μπορεί να είναι χρώματος λευκό, κόκκινο ή κίτρινο. Οι εικόνες αυτές είναι οργανωμένες σε τρεις φακέλους ανάλογα με το

χρώμα της ταμπέλα που απεικονίζουν (



3. Εικόνα 18).



Εικόνα 17 Εικόνες από τη βάση δεδομένων με οχήματα



Εικόνα 18 Εικόνες από τη βάση δεδομένων με οχήματα

Σχεδιασμός συστήματος

Το γεγονός ότι η βάση δεδομένων είναι ουσιαστικά σπασμένη σε δύο κομμάτια, τα οχήματα και τις ταμπέλες δεν είναι τυχαίο. Το σύστημα διαχειρίζεται με διαφορετικό τρόπο τις δύο περιπτώσεις.

Στην περίπτωση των οχημάτων η διαφορά τους είναι στο σχήμα του αντικειμένου, δηλαδή άλλο σχήμα έχει η μοτοσυκλέτα, άλλο το αυτοκίνητο, άλλο το λεωφορείο και άλλο το φορτηγό. Το ίδιο όμως δεν ισχύει για τις ταμπέλες. Οι ταμπέλες διαφέρουν ως προς το χρώμα. Επομένως στη πρώτη περίπτωση μας ενδιαφέρει μέσα από την πληροφορία του σχήματος του αντικειμένου που απεικονίζεται σε μία εικόνα να βρούμε το είδος του οχήματος, ενώ στη δεύτερη περίπτωση χρειάζεται η χρωματική πληροφορία της εικόνας ώστε να την κατατάξει στην κατάλληλη κλάση. Επομένως αναπτύξαμε δύο υποσυστήματα αναγνώρισης προτύπων, για τα οχήματα και για τις ταμπέλες που θα περιγραφούν ξεχωριστά.

Υποσύστημα σχηματικής αναγνώρισης προτύπων (για τα οχήματα)

Σε αυτό το σύστημα για να αναπτυχθεί η μηχανή μάθησης τα διανύσματα χαρακτηριστικών των αντικειμένων εκπαίδευσης προέκυψαν από τον συνδυασμό των περιγραφών που προκύπτουν σύμφωνα με τον αλγόριθμο SIFT και του αλγόριθμου k-means λόγω του μεγάλου όγκου των δεδομένων. Στη συνέχεια τα χαρακτηριστικά αυτά μετασχηματίζονται με τη μέθοδο KPCA και LDA σε άλλη διάσταση και το αποτέλεσμα είναι η είσοδος του ταξινομητή. Ο ταξινομητής που χρησιμοποιείται είναι ο ταξινομητής πλησιέστερου γείτονα (k-nearest neighbor). Ακολουθεί αναλυτική θεωρητική περιγραφή των μεθόδων SIFT, k-means, KPCA+LDA και k-nn που προαναφέραμε.

SIFT (32), (33), (34)

Κάθε εικόνα αποτελείται από αντικείμενα. Για την περιγραφή αυτών των αντικειμένων εξάγουμε κάποια χαρακτηριστικά, τα οποία περιγράφουν αυτά τα αντικείμενα, έτσι ώστε να μπορούμε να τα εντοπίσουμε, ακόμα και όταν βρίσκονται σε εικόνες που περιέχουν πολλά άλλα αντικείμενα. Για την εφαρμογή μιας μεθόδου που θα παρέχει αξιόπιστη αναγνώριση αντικειμένων, είναι σημαντικό τα

χαρακτηριστικά που εξάγονται από τις εικόνες (train), να είναι εντοπίσιμα ακόμα και αν έχουμε κάποιες αλλαγές στις εικόνες όπως αλλαγή μεγέθους, περιστροφή, ακόμα και προσθήκη θορύβου στην εικόνα. Χαρακτηριστικά σημεία που είναι αμετάβλητα σε τέτοιου είδους 'επιθέσεις' είναι συνήθως υψηλής αντίθεσης περιοχές όπως οι ακμές του αντικείμενου. Ένα άλλο χαρακτηριστικό αυτών των χαρακτηριστικών σημείων, είναι ότι η σχετική τους θέση παραμένει αμετάβλητη. Για παράδειγμα αν οι τέσσερις γωνίες μιας πόρτας χρησιμοποιηθούν ως χαρακτηριστικά σημεία τότε η σχετική θέση μεταξύ τους παραμένει σταθερή ανεξάρτητα από την θέση της πόρτας μέσα στην εικόνα.

Η μέθοδος Sift (Scale Invariant Feature Transform) είναι ένας μέθοδος που χρησιμοποιείται για να εντοπίζει τοπικά χαρακτηριστικά (σημεία κλειδιά) εικόνων, τα οποία είναι αμετέβλητα σε αλλαγές του μεγέθους της εικόνας (εξού και το όνομα της), καθώς επίσης και σε περιστροφή, σε αφινική παραμόρφωση, και μερικώς αμετάβλητη σε αλλαγές φωτισμού. Για τον χαρακτηρισμό μιας εικόνας χρησιμοποιούνται αυτά τα τοπικά χαρακτηριστικά για την δημιουργία των λεγόμενων «keypoint descriptors», όπως θα δούμε και στην συνέχεια οι οποίοι αποτελούνται από τα σημαντικότερα χαρακτηριστικά των σημείων κλειδιά. Ο αλγόριθμος SIFT δημοσιεύτηκε από τον David Lowe το 1999.

Ο αλγόριθμος Sift χρησιμοποιείται κυρίως για αναγνώριση εικόνων. Εξάγονται τα Sift χαρακτηριστικά από ένα σύνολο εικόνων και τα αντίστοιχα διανύσματα χαρακτηριστικών (feature vector) αποθηκεύονται σε μία βάση δεδομένων. Στην συνέχεια, για μια νέα εικόνα, εξάγεται το διάνυσμα χαρακτηριστικών της, και συγκρίνεται με τα διανύσματα της βάσης δεδομένων, έτσι ώστε να γίνει αντιστοίχιση εικόνων.

Για παράδειγμα, έστω ότι ψάχνουμε για τις παρακάτω εικόνες



Εικόνα 19



Εικόνα 20



Εικόνα 21



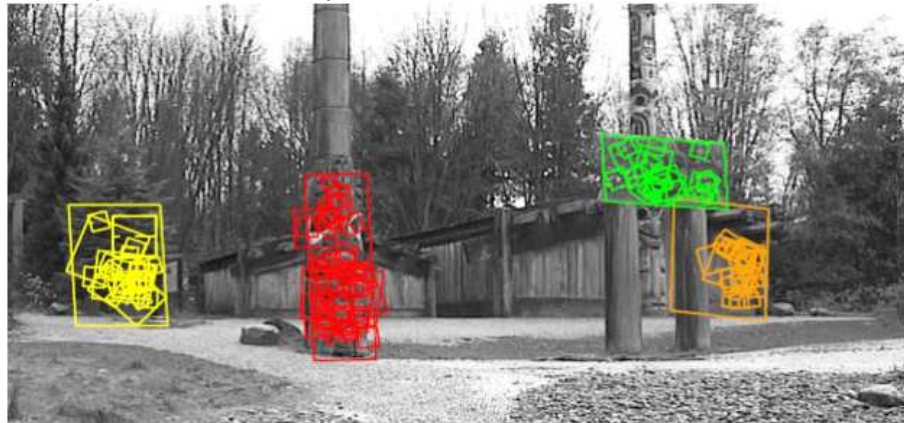
Εικόνα 22

Και η εικόνα στην οποία ψάχνουμε τα παραπάνω αντικείμενα είναι :



Εικόνα 23

Τότε τα αποτελέσματα θα είναι τα παρακάτω.



Εικόνα 24

Αλγόριθμος Sift

Ο αλγόριθμος Sift βήμα βήμα

Κατασκευή του scale space (Space Externa Detection)

Στο πρώτο στάδιο ψάχνουμε να βρούμε περιοχές της εικόνας οι οποίες εντοπίζονται σε διαφορετικές όψεις της εικόνας σε διαφορετικά μεγέθη. Για να βρούμε τέτοιες περιοχές οι οποίες να είναι αμετάβλητες στην αλλαγή μεγέθους της εικόνας, βρίσκουμε κάποια σταθερά χαρακτηριστικά σημεία (keypoints) που δεν αλλάζουν όταν αλλάζει το μέγεθος της εικόνας. Για αυτό τον σκοπό υπολογίζεται η συνέλιξη της εικόνας με Gaussian φίλτρα σε διάφορα μεγέθη.

Gaussian:

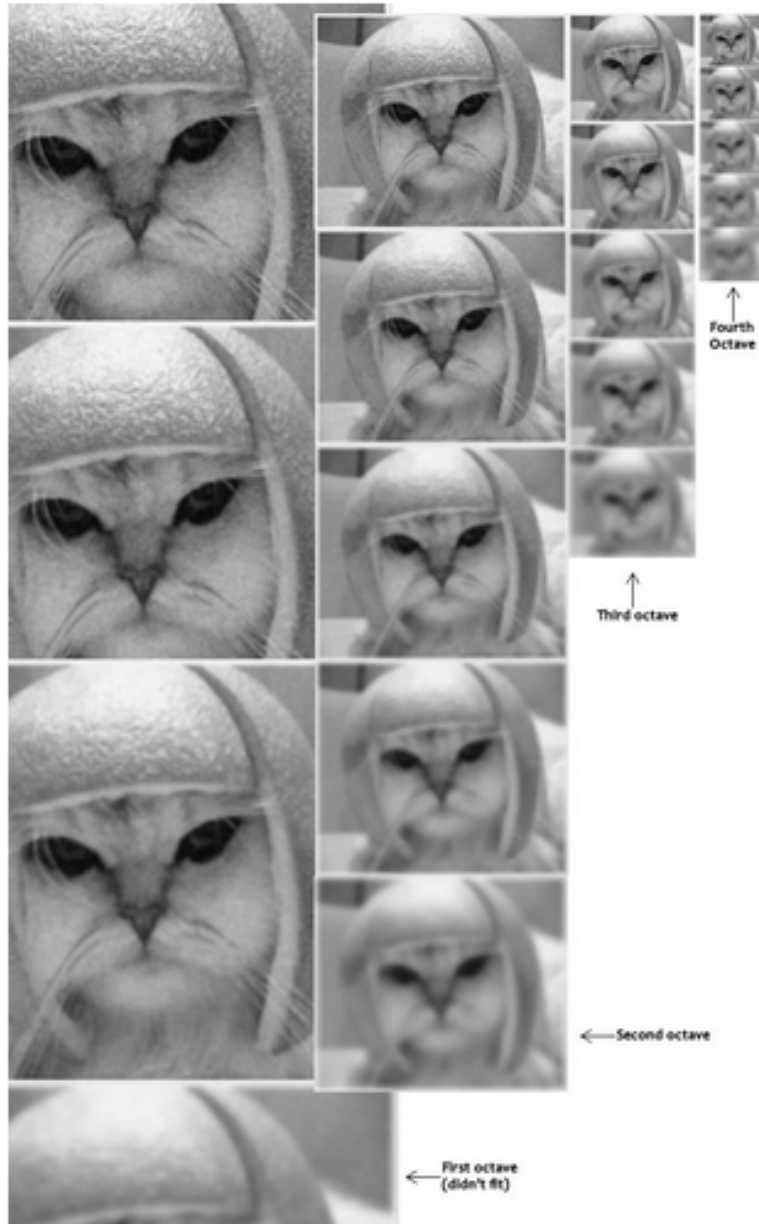
Αρχική εικόνα $I(x, y)$

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

Scale Space απεικόνιση:

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

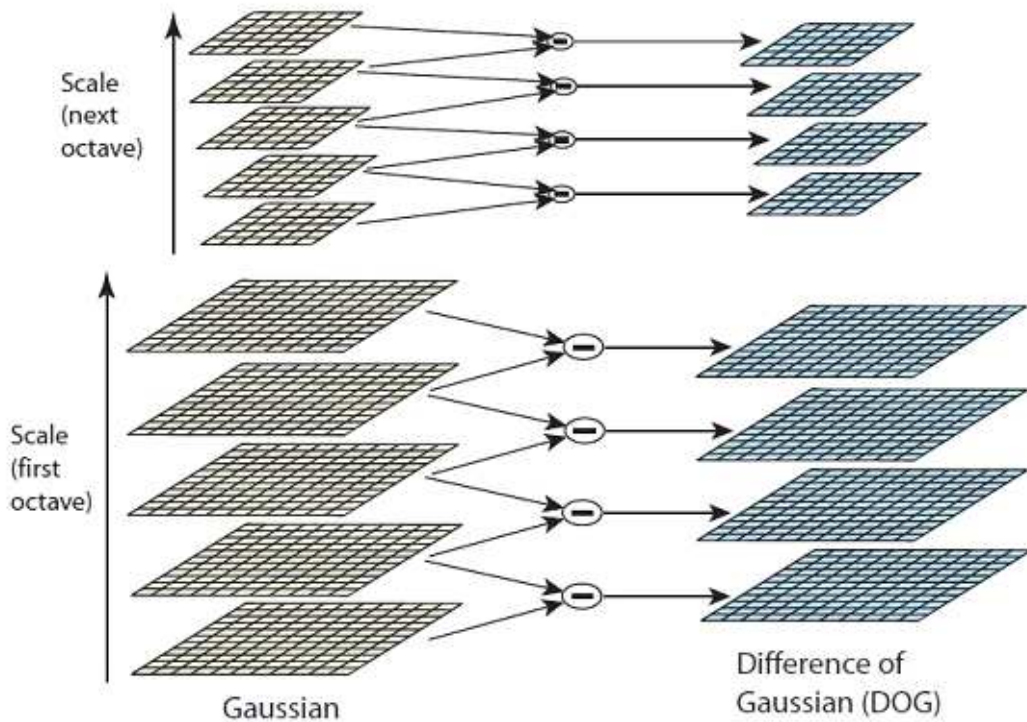
Πιο συγκεκριμένα παίρνουμε την αρχική εικόνα και δημιουργούμε θολά αντίγραφα της, μέσα από Γκαουσιανές συναρτήσεις. Στην συνέχεια η αρχική εικόνα μειώνεται στο μισό της και ξαναπαράγουμε τα αντίστοιχα θολά αντίγραφα. Η ίδια διαδικασία συνεχίζεται. Το πόσα επίπεδα θολώματος θα έχουμε για κάθε μέγεθος εικόνα, καθώς και το πόσα μεγέθη εικόνας θα εξετάσουμε, είναι δύο παράμετροι που πρέπει να οριστούν. Το σετ εικόνων για ένα συγκεκριμένο μέγεθος ονομάζεται οκτάβα (octaves). Στο παρακάτω παράδειγμα έχουμε δημιουργήσει 4 οκτάβες με 5 επίπεδα θολώματος για κάθε οκτάβα.



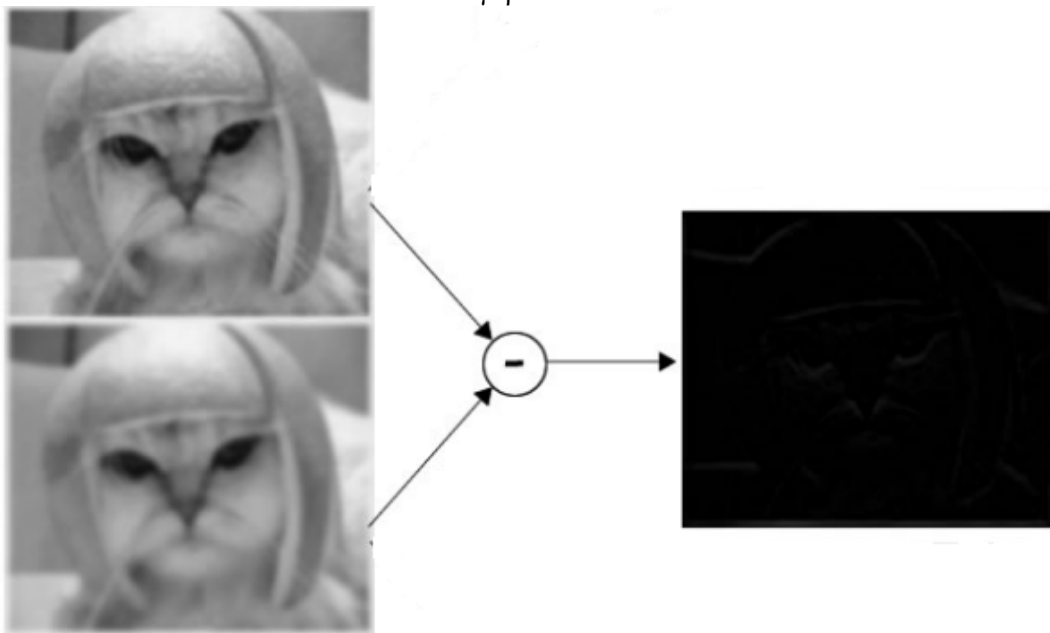
Εικόνα 25:Οκτάβες και επίπεδα θολώματος

Διαφορά Gaussian εικόνων (Difference of Gaussian)

Σε αυτό το στάδιο χρησιμοποιούμε τις προηγούμενες θολές εικόνες για να υπολογίσουμε την διαφορά των Gaussian εικόνων (Difference of Gaussian (DoG)). Υπολογίζουμε την διαφορά μεταξύ δύο διαδοχικών εικόνων, όπως υπολογίστηκαν στο προηγούμενο βήμα, κατά την δημιουργία του scale space όπως φαίνεται στην παρακάτω εικόνα.



Εικόνα 26: Διαφορά Gaussian εικόνων



Εικόνα 27: Εφαρμογή DoG

Εύρεση των χαρακτηριστικών σημείων

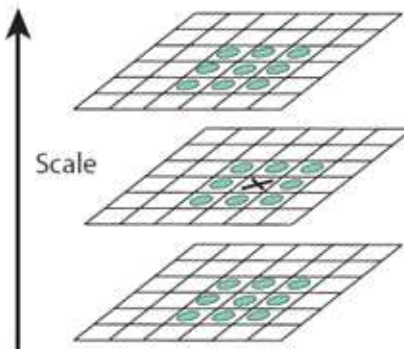
Μετά την δημιουργία των DoG εικόνων, αναγνωρίζονται τα σημεία κλειδιά (keypoints) ως τοπικά ελάχιστα/μέγιστα των Gaussian Διαφορών (DoG) των εικόνων στα διάφορα μεγέθη. Πιο συγκεκριμένα τα δύο παρακάτω βήματα εφαρμόζονται για τον προσδιορισμό των χαρακτηριστικών σημείων (keypoints).

- Να βρούμε τα μέγιστα/ελάχιστα pixel των DoG εικόνων του προηγούμενου βήματος
- Να βρούμε τα αντίστοιχα ελάχιστα μέγιστα subpixel

Μέγιστα/ελάχιστα pixel των DoG εικόνων

Για να βρούμε τα μέγιστα και τα ελάχιστα pixel των DoG εικόνων, απλά υπολογίζουμε για κάθε pixel την τιμή των γειτονικών τους και βλέπουμε αν έχει την μικρότερη ή την μεγαλύτερη τιμή. Τα

γειτονικά ενός pixel, δεν βρίσκονται μόνο στην τρέχουσα DoG εικόνα αλλά και στην προηγούμενη και στην επόμενη της, όπως φαίνεται στην παρακάτω εικόνα:



Εικόνα 28: Γειτονικά pixel

Το x pixel της εικόνας αντιπροσωπεύει το τρέχων pixel. Οι πράσινοι κύκλοι αντιπροσωπεύουν τα pixel-γείτονες. Το x σημειώνεται ως keypoint αν είναι μεγαλύτερο ή μικρότερο από τα υπόλοιπα 26 pixel-γείτονες.

Μετά από αυτή την διαδικασία τα κατά προσέγγιση μέγιστα και ελάχιστα pixel έχουν υπολογιστεί. Είναι κατά προσέγγιση γιατί σπάνια το μέγιστο/ελάχιστο βρίσκεται πάνω ακριβώς σε ένα pixel. Φυσικά δεν έχουμε πρόσβαση μεταξύ των pixel για αυτό εφαρμόζουμε μια μαθηματική προσέγγιση για τον προσδιορισμό των subpixel.

Προσδιορισμός των μέγιστων/ελάχιστων

Αρχικά για κάθε υποψήφιο σημείο κλειδί χρησιμοποιείται η συνάρτηση της παρεμβολής χρησιμοποιώντας τα γειτονικά δεδομένα, για ακριβή προσδιορισμό της θέσης αυτού. Η αρχική προσέγγιση ήταν απλά ο εντοπισμός κάθε σημείου κλειδιού στην θέση και το μέγεθος του υποψηφίου σημείου κλειδιού. Η νέα προσέγγιση υπολογίζει την παρεμβάουσα θέση του μεγίστου, το οποίο με την σειρά του βελτιώνει την αντιστοίχιση και την σταθερότητα των σημείων. Η παρεμβολή γίνεται χρησιμοποιώντας το δεύτερο ανάπτυγμα Taylor της Gaussian Διαφοράς (DoG) scale-space συνάρτηση, $D(x,y,\sigma)$ με υποψήφιο κλειδί ως αρχική τιμή. Το εν λόγω ανάπτυγμα δίνεται από την συνάρτηση:

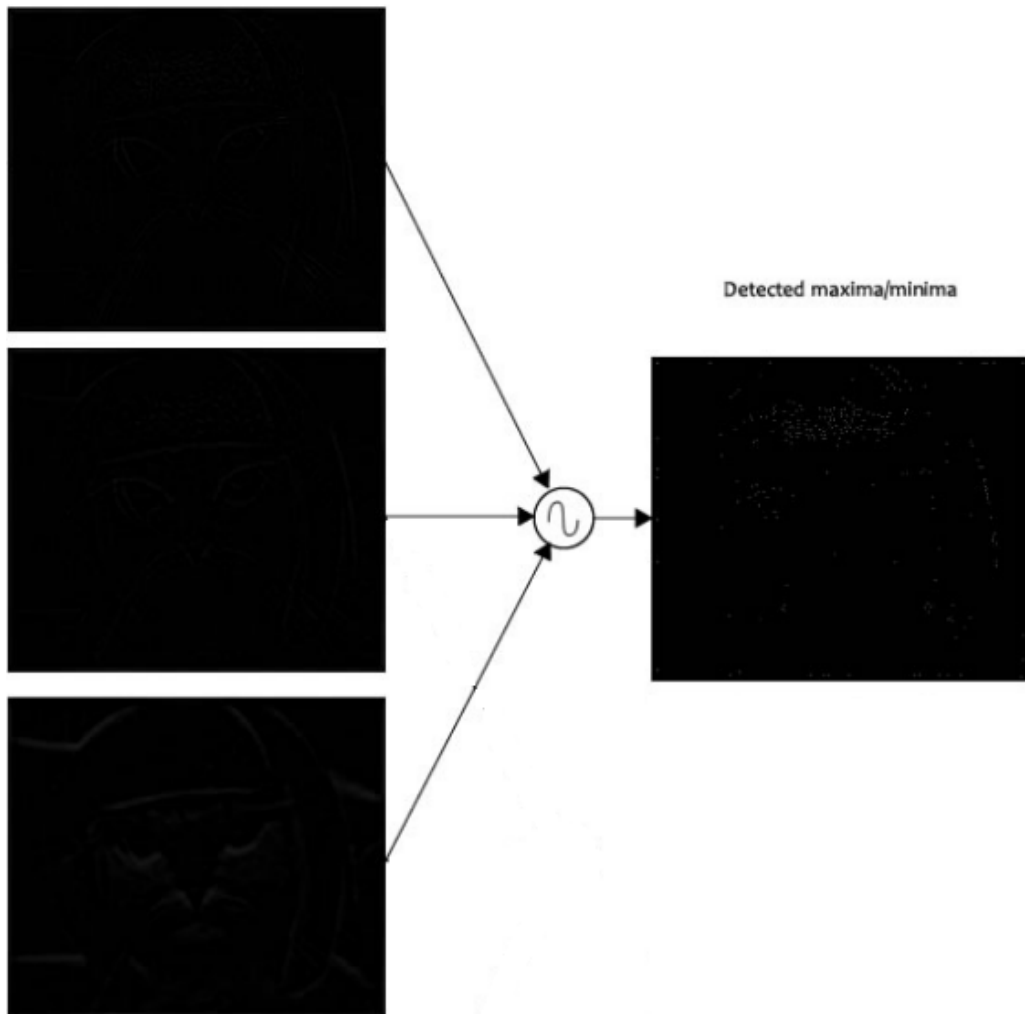
$$D(\mathbf{x}) = D + \frac{\partial D}{\partial \mathbf{x}} \mathbf{x} + \frac{1}{2} \mathbf{x}^T \frac{\partial^2 D}{\partial \mathbf{x}^2} \mathbf{x}$$

Όπου D και οι παράγωγοι του υπολογίζονται στο υποψήφιο σημείο κλειδί και $\mathbf{x} = (x,y,\sigma)^T$ είναι η απόσταση του σημείου κλειδιού από το υπό εξέταση σημείο, με άλλα λόγια η απόσταση (offset) αυτού του σημείου από το keypoint. Η θέση του ακρότατου $\hat{\mathbf{x}}$, καθορίζεται βρίσκοντας την παράγωγο της D και θέτοντας την ίση με μηδέν:

$$\hat{\mathbf{x}} = - \frac{\partial^2 D}{\partial \mathbf{x}^2}^{-1} \frac{\partial D}{\partial \mathbf{x}}$$

Αν το offset $\hat{\mathbf{x}}$ είναι μεγαλύτερο από 0.5 σε κάθε διάσταση, τότε αυτό είναι μια ένδειξη ότι το ακρότατο βρίσκεται πιο κοντά σε κάποιο άλλο υποψήφιο σημείο κλειδί. Σε αυτή την περίπτωση, το υποψήφιο σημείο κλειδί αλλάζει και εκτελείται η παρεμβολή αντί αυτού του σημείου. Σε άλλη περίπτωση το offset προστίθεται στο υποψήφιο κλειδί για να πάρουμε την παρεμβολή εκτίμηση για την θέση ακρότατου.

Στο παρακάτω παράδειγμα συνδυάζουμε 3 DoG εικόνες για να βρούμε τα μέγιστα και τα ελάχιστα pixels.



Διαγραφή ακμών και χαρακτηριστικών χαμηλής αντίθεσης

Το πρώτο στάδιο του αλγορίθμου παράγει πολλά υποψήφια σημεία κλειδιά (keypoints). Σε αυτό το σημείο απορρίπτονται σημεία που έχουν χαμηλή αντίθεση (contrast), γιατί είναι ευαίσθητα στην παρουσία θορύβου. Επιπλέον απορρίπτονται τα σημεία που βρίσκονται πάνω σε κάποια ακμή αφού δεν είναι χρήσιμα ως χαρακτηριστικά

Διαγραφή χαρακτηριστικών χαμηλής αντίθεσης.

Για την διαγραφή των σημείων κλειδιών με χαμηλή αντίθεση (contrast) υπολογίζεται το δεύτερης τάξης ανάπτυγμα του Taylor $D(x)$ στο offset x :

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

Αν η τιμή είναι μικρότερη από 0.03, το υποψήφιο σημείο κλειδί διαγράφεται. Αλλιώς το κρατάμε με τελική θέση $y+\hat{x}$ και μέγεθος σ , όπου y είναι η πραγματική θέση του σημείου κλειδί (keypoint) με μέγεθος σ .

Διαγραφή ακμών

Η ιδέα είναι να υπολογίσουμε δύο κλίσεις (gradients) σε κάθε keypoint, οι οποίες θα είναι κάθετες μεταξύ τους. Η εικόνα γύρω από κάθε keypoint μπορεί να είναι:

- Επίπεδη περιοχή

Σε αυτή την περίπτωση και οι δύο κλίσεις θα είναι μικρές.

- Ακμή

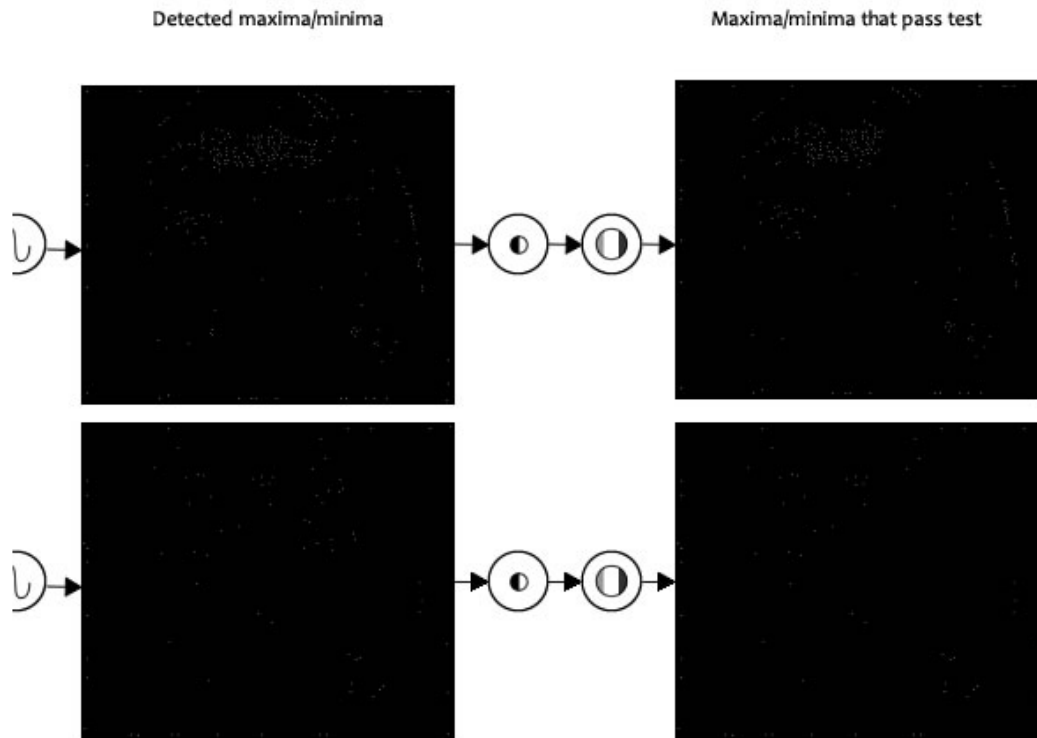
Σε αυτή την περίπτωση η μία κλίση θα είναι μεγάλη και η άλλη θα είναι μικρή.

- Γωνία

Σε αυτή την περίπτωση και οι δύο κλίσεις θα είναι μεγάλες.

Η γωνία είναι ένα πολύ καλό χαρακτηριστικό το οποίο και επιθυμούμε να κρατήσουμε. Σε οποιαδήποτε από τις δύο άλλες περιπτώσεις το keypoint διαγράφεται. Μαθηματικά αυτό μπορεί να υλοποιηθεί με τον Hessian Matrix. Χρησιμοποιώντας αυτόν τον πίνακα ελέγχουμε αν το σημείο είναι γωνία ή όχι.

Στην παρακάτω εικόνα βλέπουμε τι συμβαίνει σε αυτό το στάδιο. Παρατηρούμε ότι μόνο μερικά από τα key points που δημιουργήθηκαν στο προηγούμενο βήμα κρατιούνται.



Προσδιορισμός της κατεύθυνσης των χαρακτηριστικών σημείων.

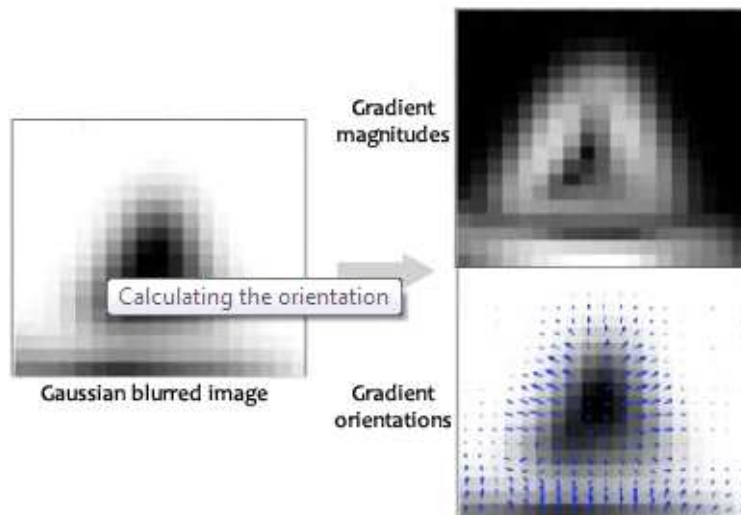
Έχουμε ελέγξει ότι τα keypoint είναι σταθερά όταν αλλάζει το μέγεθος της εικόνας. Σε αυτό το βήμα θα δώσουμε μία κατεύθυνση στα keypoints έτσι ώστε να μπορούμε να έχουμε αμεταβλητότητα και σε περίπτωση περιστροφής.

Αρχικά παίρνεται η Λεία Gaussian (Gaussian-Smooth) εικόνα $L(x,y,\sigma)$ σε περιοχή μεγέθους σ σε κάθε keypoint, έτσι ώστε όλοι οι υπολογισμοί να πραγματοποιηθούν με μια σταθερούς μεγέθους μέθοδο. Για ένα δείγμα εικόνας $L(x,y)$ μεγέθους σ , το διανυσματικό μέτρο (magnitude) των στοιχείων, $m(x,y)$ και η κατεύθυνση (orientation), $\theta(x,y)$, προϋπολογίζονται χρησιμοποιώντας διαφορές μεταξύ των pixel:

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

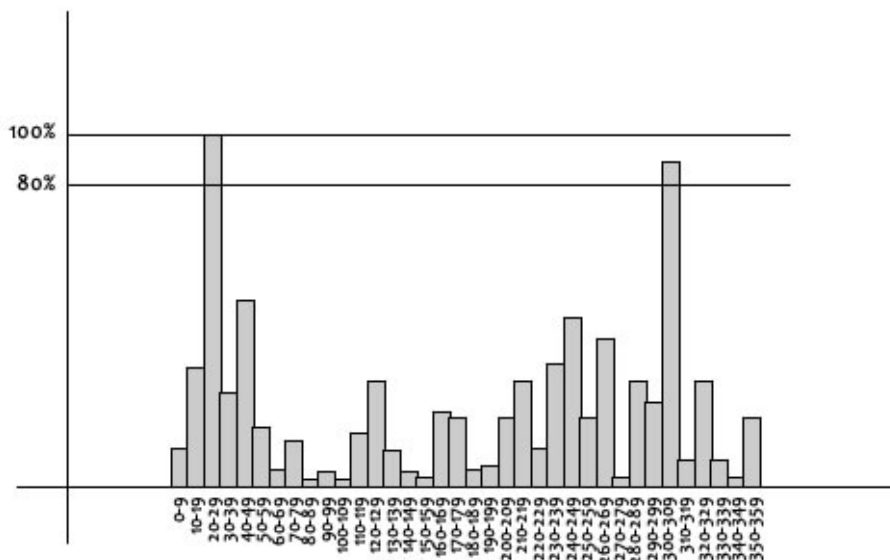
$$\theta(x,y) = \tan^{-1}((L(x,y+1) - L(x,y-1))/(L(x+1,y) - L(x-1,y)))$$

Ο υπολογισμός του διανυσματικού μέτρου (magnitude) και της κατεύθυνσης των στοιχείων γίνονται για κάθε pixel σε μια γειτονική περιοχή γύρω από το σημείο κλειδί (keypoint) στην Gaussian-Blur εικόνα L .



Εικόνα 29: Προσδιορισμός διανυσματικού μεγέθους και κατεύθυνσης

Το μέγεθος και οι κατευθύνσεις υπολογίζονται για όλα τα pixels γύρω από κάθε keypoint. Στην συνέχεια κατασκευάζεται ένα ιστόγραμμα κατεύθυνσης, το οποίο περιέχει 36 στοιχεία (bins), καθένα από τα οποία καλύπτει εύρος 10 μοιρών . (Εικόνα 30)



Εικόνα 30: Ιστόγραμμα κατεύθυνσης

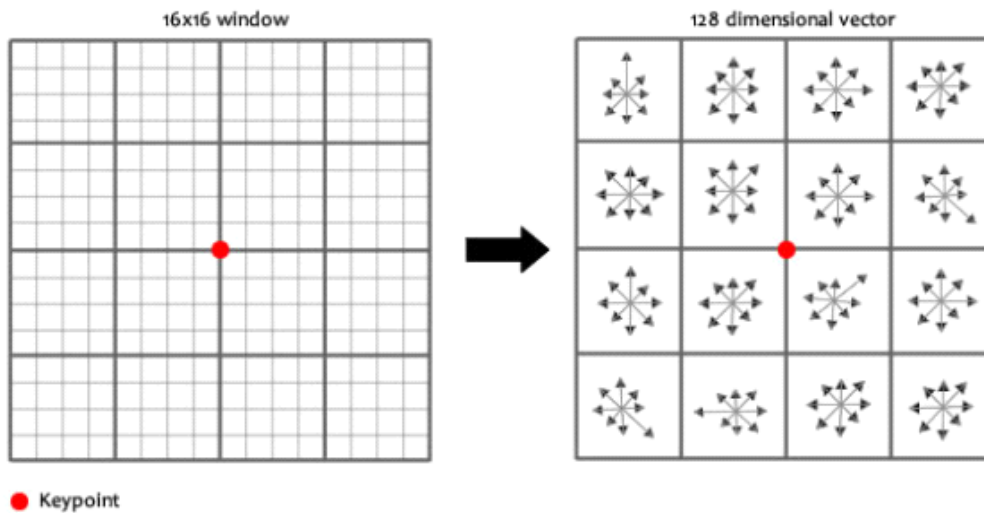
Κάθε δείγμα στο γειτονικό παράθυρο προστίθεται σε ένα ιστόγραμμα στοιχείων που καθορίζεται από το διανυσματικό μέτρο (magnitude) των στοιχείων του και από ένα ισοζυγισμένο Gaussian κυκλικό παράθυρο με σ 1.5 φορά μεγαλύτερο του μεγέθους του σημείου κλειδιού (keypoint). Τα peaks σε αυτό το ιστόγραμμα αναφέρονται στις κυρίαρχες κατευθύνσεις. Από την στιγμή που τι ιστόγραμμα γεμίσει, οι κατευθύνσεις που αναφέρονται στο υψηλότερο peak και τοπικά peaks που βρίσκονται μεταξύ 80% των υψηλότερων peaks ανατίθενται στο σημείο κλειδί. Στην περίπτωση που έχουν ανατεθεί πολλές κατευθύνσεις, δημιουργείται ένα πρόσθετο keypoint με το ίδιο μέγεθος και θέση του πρωτότυπου keypoint για κάθε πρόσθετη κατεύθυνση.

Δημιουργία του SIFT descriptor

Στα προηγούμενα στάδια βρήκαμε τις θέσεις των σημείων κλειδιών (keypoints) σε συγκεκριμένα μεγέθη και αναθέσαμε σε αυτά κατευθύνσεις. Με αυτόν τον τρόπο μπορέσαμε να εξασφαλίσουμε σταθερότητα σε αλλαγές στο μέγεθος της εικόνας, καθώς και σε περιστροφή της εικόνας. Σε αυτό το

στάδιο θα υπολογίσουμε τα διανύσματα του descriptor στα σημεία κλειδιά, για να επιτύχουμε σταθερότητα των descriptor στα εναπομείναντα στοιχεία που μπορούν να προκαλέσουν μεταβλητότητα στα αποτελέσματα όπως η αλλαγή φωτισμού, η αλλαγή 3D οπτικής κ.α.

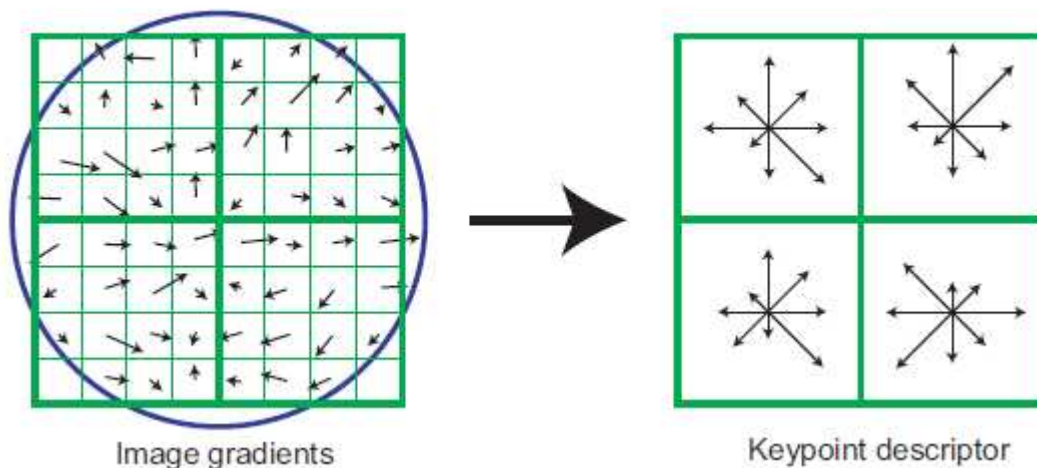
Αυτό που κάνουμε σε αυτό το τελευταίο στάδιο της δημιουργίας των SIFT descriptor είναι δημιουργήσουμε ένα διάνυσμα το οποίο θα χαρακτηρίζει μοναδικά το κάθε σημείο κλειδί. Για την διαδικασία αυτή αρχικά παίρνουμε ένα παράθυρο μεγέθους 16x16 γύρω από κάθε χαρακτηριστικό σημείο (σημείο κλειδί). Αυτό το παράθυρο χωρίζεται επιπλέον σε δεκαέξι μικρότερα παράθυρα μεγέθους 4x4 όπως φαίνεται στην Εικόνα 31. Για κάθε ένα από αυτά τα παράθυρα υπολογίζονται τα διανυσματικά μεγέθη (magnitudes) και οι κατευθύνσεις.



Εικόνα 31: Δημιουργία παραθύρων γύρω από ένα keypoint

Ο Feature Descriptor («Περιγραφέας» Χαρακτηριστικών) υπολογίζεται ως ένα σύνολο από ιστογράμματα κατεύθυνσης με 4x4 σύγκριση γειτονικών pixel. Τα ιστογράμματα κατεύθυνσης σχετίζονται με την κατεύθυνση των keypoints και την κατεύθυνση των δεδομένων τα οποία μας δίνει η Gaussian εικόνα που είναι πιο κοντά σε μέγεθος με το μέγεθος των keypoints. Η συνεισφορά κάθε pixel καθορίζεται από το διανυσματικό μέγεθος κάθε στοιχείου (magnitude), και από μια Gaussian με σ ίσο με 1.5 φορές το μέγεθος του keypoint. Τα ιστογράμματα περιέχουν 8 bins το καθένα και κάθε descriptor περιέχει έναν 4x4 πίνακα από 16 ιστογράμματα γύρω από το σημείο κλειδί. Οπότε ένα SIFT διάνυσμα περιέχει $4*4*8= 128$ στοιχεία.

Αν και η διάσταση του descriptor φαίνεται να είναι αρκετά υψηλή, μελέτες έχουν δείξει ότι descriptors με μικρότερη διάσταση δεν αποδίδουν τόσο καλά. Επιπλέον η BBF (Best-Bin-First) μέθοδος χρησιμοποιείται για την εύρεση του κοντινότερου γείτονα. Descriptors με μεγαλύτερο μέγεθος δεν παρουσιάζουν ιδιαίτερη βελτίωση.



Εικόνα 32: Για την δημιουργία του keypoint descriptor υπολογίζουμε το διανυσματικό μέγεθος και την κατεύθυνση των στοιχείων σε κάθε σημείο γύρω από το σημείο κλειδί. Αυτά υπολογίζονται από ένα Gaussian παράθυρο, ορισμένο από τον επικαλυπτόμενο κύκλο. Στην συνέχεια το ιστόγραμμα κατευθύνσεων υπολογίζεται συνοψίζοντας τα στοιχεία σε 4X4 υποπεριοχές, όπως φαίνεται στο δεξιό μέρος της εικόνας. Το μήκος κάθε βέλους είναι ίσο με το άθροισμα των διανυσματικών μεγεθών των στοιχείων κοντά σε αυτή την κατεύθυνση μέσα στην περιοχή.

k-means

Ο αλγόριθμος k-means (35) είναι ένας από τους πιο γνωστούς αλγορίθμους που χρησιμοποιούνται για την επίλυση προβλημάτων clustering. Ο αλγόριθμος αυτός ανήκει στην κατηγορία του unsupervised learning.

Ο αλγόριθμος δέχεται ως είσοδο ένα συγκεκριμένο αριθμό από k κέντρα των clusters, και αυτός ο αριθμός συμβολίζει και τον αριθμό των τελικών clusters που θα έχουμε ως έξοδο του αλγορίθμου. .

Το επόμενο βήμα που κάνει ο αλγόριθμος είναι να αναθέσει κάθε δεδομένο (datum) στο κοντινότερο του centroid. Όταν ανατεθούν όλα τα δεδομένα, ένα πρώτο clustering έχει γίνει. Σε αυτό το σημείο επαναυπολογίζουμε τα centroids με βάση τα καινούργια clusters που έχουν δημιουργηθεί και τα τοποθετούμε έτσι ώστε να κατοπτρίζουν το κέντρο των δεδομένων που ανήκουν στο cluster τους. Έτσι όταν έχουμε k καινούργια centroids επαναυπολογίζουμε τις θέσεις των δεδομένων και τα εναποθέτουμε στο κοντινότερο σε αυτά κέντρο. Με τον τρόπο αυτό βλέπουμε πως έχει δημιουργηθεί ένας βρόχος ο οποίος τερματίζεται όταν πλέον τα κέντρα δεν κουνιούνται από την θέση τους.

Τέλος, ένας από τους βασικούς στόχους του αλγορίθμου είναι να καταφέρει να ελαχιστοποιήσει την παρακάτω συνάρτηση τετραγωνικού λάθους, όπου είναι η απόσταση μεταξύ της θέσης ενός δεδομένου x και ενός κέντρου c .

Ο αλγόριθμος βήμα βήμα

Αρχική επιλογή του προσδοκώμενου η επιθυμητού αριθμού ομάδων K . Αυθαίρετη επιλογή των αρχικών K κέντρων

Ομαδοποίηση των δεδομένων με χρήση της Ευκλείδειας απόστασης (ή κάποιας άλλης μετρικής απόστασης)

Επαναπροσδιορισμός των κέντρων χρησιμοποιώντας μόνο τα σημεία του κάθε cluster.

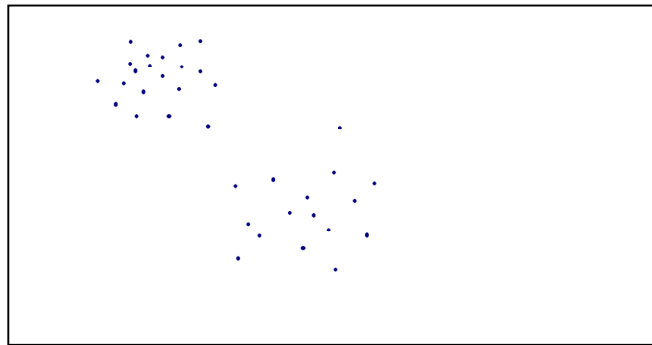
Ομαδοποίηση ξανά όλων των σημείων

Επαναλαμβάνουμε τα βήματα 3 και 4 μέχρι να μην μετακινηθεί κανένα άλλο σημείο από ένα cluster σε άλλο ή μέχρι ο αλγόριθμος να συγκλίνει ικανοποιώντας κάποιο άλλο κριτήριο σύγκλισης

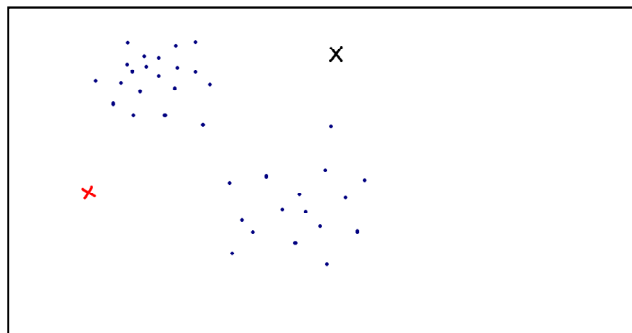
Το κύριο μειονέκτημα του αλγορίθμου είναι ότι αν και τερματίζει πάντα, δεν δίνει πάντα την βέλτιστη λύση. Η αρχικοποίηση του αριθμού των κέντρων από τον χρήστη συντελεί έναν πολύ βασικό παράγοντα. Μια αρχική επιλογή αριθμού κέντρων που δεν είναι και τόσο πετυχημένη μπορεί να οδηγήσει σε μια λάθος συσταδοποίηση.

Παράδειγμα

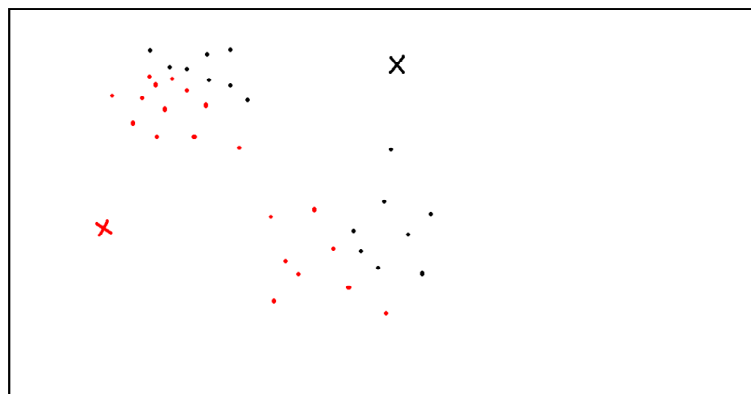
Όπως βλέπουμε στην παρακάτω εικόνα έχουμε ένα σύνολο δεδομένων, τα οποία και θα συσταδοποιήσουμε.



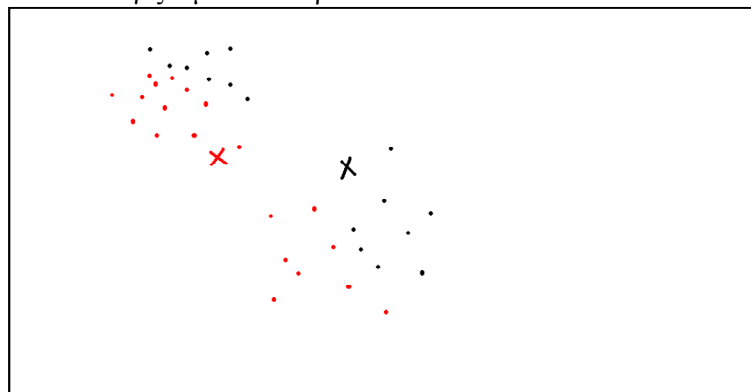
Αρχικά τοποθετούμε 2 σημεία στον χώρο των δεδομένων. Αυτά τα σημεία αντιπροσωπεύουν τα αρχικά κέντρα των clusters.



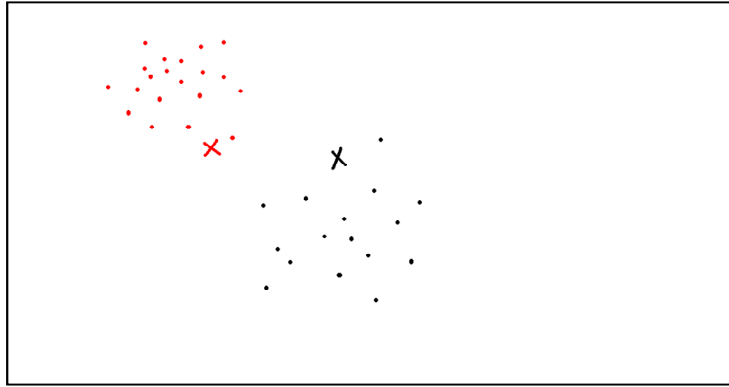
Έπειτα αναθέτουμε κάθε δεδομένο στην συστάδα που έχει το κοντινότερο κέντρο, όπως φαίνεται στην εικόνα.



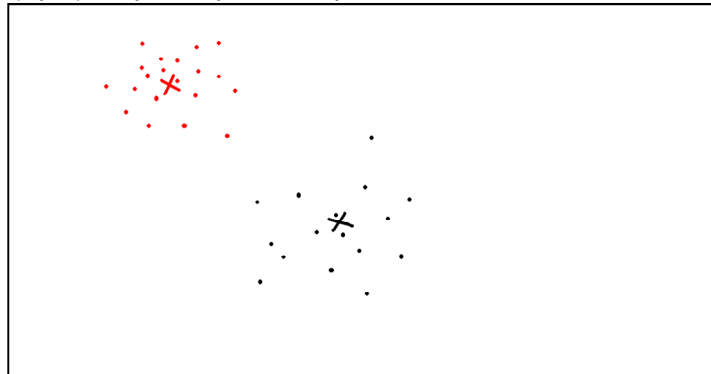
Στην συνέχεια επαναυπολογίζουμε τα κέντρα των συστάδων.



Έπειτα αναθέτουμε κάθε δεδομένο στην συστάδα που έχει το κοντινότερο κέντρο.



Τέλος επαναυπολογίζουμε τις θέσεις των κέντρων.



Όσες φορές και να εκτελεστεί ο αλγόριθμος από εδώ και πέρα τα κέντρα δεν θα αλλάζουν θέση. Δηλαδή ο αλγόριθμος τερματίζει.

KPCA+LDA

Σε πολλές περιπτώσεις επειδή το αποτέλεσμα της εξαγωγής χαρακτηριστικών είναι μεγάλο σε όγκο και μπορεί να κάνει δύσκολη ή και αδύνατη τη διαδικασία ταξινόμησης των αντικειμένων είναι αναγκαίο να βρεθεί τρόπος ώστε να μειωθούν αυτά τα δεδομένα. Ένας προφανής τρόπος θα ήταν να χρησιμοποιηθούν λιγότερα αντικείμενα εκπαίδευσης αλλά αυτό μπορεί να έχει αρνητικές συνέπειες στην επίδοση του ταξινομητή. Για το λόγο αυτό έχουν αναπτυχθεί κατά καιρούς διάφορες τεχνικές μείωσης της διάστασης των χαρακτηριστικών.

Έχει παρατηρηθεί ότι όλα τα χαρακτηριστικά μπορεί να μην είναι απαραίτητα για την εκπαίδευση και τη μάθηση μίας μηχανής και υπάρχουν τρόποι ώστε να απεικονισθεί η ίδια πληροφορία με χαρακτηριστικά μικρότερης διάστασης. Έχουν αναπτυχθεί διάφορες μέθοδοι όπως PCA, KPCA, LDA, KFD, GDA και άλλες (36).

Η μέθοδος LDA (Linear Discriminant Analysis) είναι μία παραδοσιακή στατιστική τεχνική, η οποία έχει αποδειχθεί επιτυχής στη μείωση διάστασης δεδομένων σε πολλές πραγματικές εφαρμογές. Το μειονέκτημά της είναι ότι είναι περιορισμένη για γραμμικά προβλήματα για αυτό και δε μπορεί να εφαρμοστεί σε μη γραμμικά προβλήματα. Για να ξεπεραστεί αυτό το πρόβλημα έχει αναπτυχθεί η ανάλυση KFD (kernel Fischer discriminant) το οποίο περιλαμβάνει δύο στάδια. Ο KFD είναι γνωστός αλλιώς ως ο αλγόριθμος KPCA plus LDA και περιλαμβάνει δύο βήματα. Πριν περιγράψουμε αυτά τα δύο βήματα θα περιγράψουμε τη μέθοδο μείωσης διαστάσεων PCA (Principal Component Analysis).

Αυτή η μέθοδος παίρνει ως είσοδο τον πίνακα χαρακτηριστικών όλων των αντικειμένων εκπαίδευσης. Σε αυτό τον πίνακα κάθε γραμμή αντιστοιχεί στο διάνυσμα χαρακτηριστικών ενός συγκεκριμένου αντικειμένου. Η PCA κάνει ιδιοανάλυση στον πίνακα αυτό και ταξινομεί τα ιδιοδιανύσματα με βάση τις αντίστοιχες ιδιοτιμές. Τελικά ως πίνακας χαρακτηριστικών επιλέγεται το πλήθος m ιδιοδιανυσμάτων που αντιστοιχούν στις μεγαλύτερες ιδιοτιμές. Ο αριθμός m εξαρτάται από το εκάστοτε σύστημα και είναι ευθύνη του σχεδιαστή του συστήματος να το ορίζει. Έχει αποδειχθεί ότι η πληροφορία που περιέχεται σε αυτά τα ιδιοδιανύσματα είναι αρκετή και ισάξια με τον αρχικό πίνακα χαρακτηριστικών για τη μάθηση της μηχανής.

Η μέθοδος KPCA εφαρμόζει τη μέθοδο PCA όχι στον πίνακα χαρακτηριστικών αλλά στον πίνακα kernel. Ο πίνακας kernel έχει περιγραφεί στην ενότητα 2.2. Έχοντας περιγράψει τις μεθόδους KPCA και LDA μπορούμε να περιγράψουμε τον αλγόριθμο KFD (37).

Βήμα 1: Αν θεωρήσουμε ότι τα διανύσματα χαρακτηριστικών είναι n διάστασης δηλαδή ο χώρος εισόδου είναι ο R^n τότε κατασκευάζεται ο πίνακας kernel. Στη συνέχεια γίνεται σε αυτό τον πίνακα ιδιοανάλυση και λαμβάνονται τα ορθογώνια ιδιοδιανύσματα που αντιστοιχούν στις m μεγαλύτερες μη μηδενικές ιδιοτιμές. Οπότε τα διανύσματα χαρακτηριστικών πλέον είναι στο χώρο R^m .

Βήμα 2: Εφαρμογή του αλγορίθμου LDA στο KPCA-μετασχηματισμένο χώρο R^m .

k-nearest neighbor

Ο ταξινομητής πλησιέστερου γείτονα ταξινομεί τα αντικείμενα με βάση τα κοντινότερα αντικείμενα εκπαίδευσης στο χώρο χαρακτηριστικών. Ανήκει στην κατηγορία της «τεμπέλας» εκπαίδευσης ή αλλιώς εκπαίδευσης με βάση παραδείγματα όπου η συνάρτηση διαχωρισμού προσεγγίζεται τοπικά και όλοι οι υπολογισμοί αναβάλλονται μέχρι την ταξινόμηση.

Πρόκειται για έναν από τους πιο απλούς αλγορίθμους μηχανών μάθησης. Ένα αντικείμενο ταξινομείται με βάση την πλειψηφία των κλάσεων στις οποίες ανήκουν τα γειτονικά αντικείμενα και συγκεκριμένα k γειτονικά αντικείμενα. Για παράδειγμα στην περίπτωση που $k = 1$ τότε το αντικείμενο ανήκει στην ίδια κλάση που ανήκει και το κοντινότερο σε αυτό αντικείμενο.

Προκειμένου να βρεθούν οι κοντινότεροι γείτονες σε ένα αντικείμενο χρησιμοποιείται η ευκλείδεια απόσταση συνήθων. Θεωρώντας ότι τα διανύσματα χαρακτηριστικών δύο αντικειμένων είναι n διάστασης τότε μπορούμε να τα συμβολίσουμε ως:

$$fv1 = (x_1, x_2, \dots, x_n) \text{ και}$$

$$fv2 = (y_1, y_2, \dots, y_n)$$

Η μεταξύ τους ευκλείδεια απόσταση δίνεται από τον τύπο:

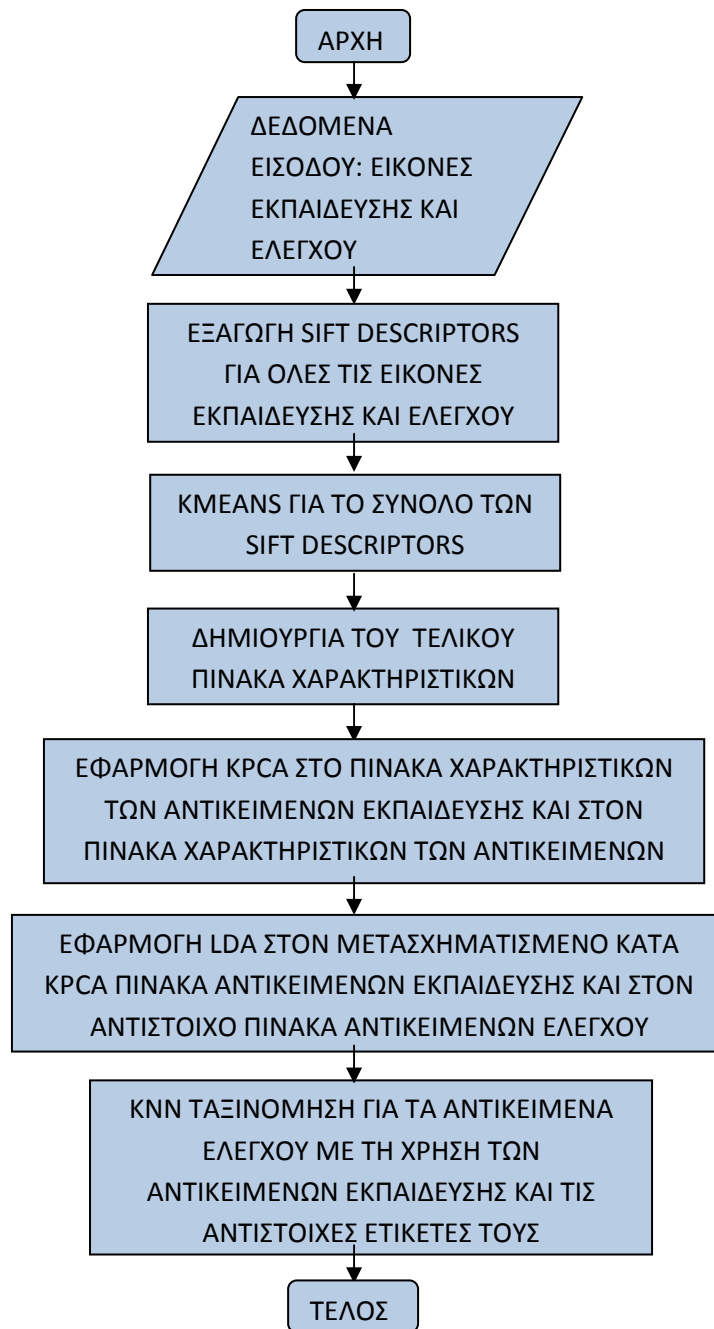
$$|fv1 - fv2| = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_n - y_n)^2}$$

Έχοντας περιγράψει όλη τη θεωρητική βάση της μηχανής μάθησης του υποσυστήματος σχηματικής αναγνώρισης προτύπων δίνουμε σχηματικά τον αλγόριθμο

Εικόνα 33.

Όσον αφορά στα χαρακτηριστικά είναι καλό να ξεκαθαρίσουμε ποια ακριβώς είναι η διαδικασία ώστε να προκύψουν μέσα από το συνδυασμό των αλγορίθμων SIFT και KMEANS. Τα βήματα έχουν ως εξής:

1. Εξάγονται οι περιγραφείς SIFT για όλες τις εικόνες. Σε κάθε εικόνα αντιστοιχεί ένας διδιάστατος πίνακας που έχει τόσες γραμμές όσα και τα σημεία ενδιαφέροντος που βρέθηκαν με τον SIFT ενώ οι στήλες είναι 128.
2. Οι πίνακες που προκύπτουν από κάθε πίνακα ενώνονται για να δώσουν έναν πίνακα με 128 στήλες και γραμμές όσα και τα σημεία ενδιαφέροντος όλων των εικόνων.
3. Σε αυτό τον πίνακα εφαρμόζεται ο αλγόριθμος k-means για k κέντρα και προκύπτει ένας πίνακας στήλη με τόσα στοιχεία όσες και οι γραμμές του πίνακα στον οποίο εφαρμόστηκε.
4. Στη συνέχεια κατασκευάζεται ένας πίνακας με k στήλες και γραμμές ίσες με τον αριθμό των εικόνων όπου το στοιχείο της i γραμμής και της j στήλης δίνει τον αριθμό των σημείων ενδιαφέροντος της εικόνας i που ανήκουν στο j κέντρο, όπου κέντρο ένα από τα k κέντρα του αλγορίθμου k-means.



Εικόνα 33 Υποσύστημα σχηματικής αναγνώρισης προτύπων

Υποσύστημα χρωματικής αναγνώρισης (για τις ταμπέλες)

Αυτό το σύστημα είναι πολύ πιο απλό καθώς το μόνο που χρειάζεται να βρούμε είναι το πλήθος των pixel της εικόνας που έχουν χρώμα κόκκινο, κίτρινο και άσπρο. Ύστερα από πειράματα αποδείχθηκε το αυτονόητο, δηλαδή ανάλογα με το μεγαλύτερο αριθμό Pixel των τριών χρωμάτων καταλήγουμε στο χρώμα της ταμπέλας. Για να αναγνωρισθεί το χρώμα ενός συγκεκριμένου pixel παίρνουμε τις συνιστώσες του στα τρία χρωματικά κανάλια R,G και B, δηλαδή τις φωτεινότητες οι οποίες έχουν τιμή από 0 έως 255.

	Pixel		
Συνιστώσα R	>200	>200	>200
Συνιστώσα G	<100	>200	>200
Συνιστώσα B	<100	<100	>200
ΧΡΩΜΑ	ΚΟΚΚΙΝΟ	ΚΙΤΡΙΝΟ	ΑΣΠΡΟ

Πίνακας 1 Χρώμα pixel με βάση το χρωματικό μοντέλο RGB

3. MATLAB

3.1. Τι είναι το Matlab

Το MATLAB (MATrixLABoratory) είναι ένα υπολογιστικό περιβάλλον λογισμικού στους ηλεκτρονικούς υπολογιστές που χρησιμοποιείται ευρέως από επιστήμονες που αναπτύσσουν συστήματα λογισμικών βασισμένα στα μαθηματικά. Το MATLAB ανήκει στις γλώσσες προγραμματισμού τέταρτης γενιάς.

Αναπτύχθηκε από την MathWorks και πρόκειται για ένα πρόγραμμα λογισμικού που η ανάπτυξή του ξεκίνησε ως ένα εργαστήριο πινάκων παρέχοντας μεγάλη διευκόλυνση σε υπολογισμούς μεταξύ πινάκων αλλά και στην αποθήκευση μεγάλων δεδομένων σε πίνακες κάνοντας εύκολη της οργάνωσή τους, την προσπέλαση τους και την επεξεργασία τους μαζί.

Θεωρείται ένα από τα πιο ισχυρά εργαλεία σε πολλούς τομείς όπως αυτοί της οπτικοποίησης, του προγραμματισμού, της έρευνας, των μηχανικών συστημάτων και των επικοινωνιών. Στα πλαίσια του MATLAB έχουν αναπτυχθεί μοντέρνοι αλγόριθμοι και ισχυρά προγραμματιστικά εργαλεία. Το MATLAB αποτελείται από το βασικό πακέτο συναρτήσεων, βιβλιοθηκών και εργαλείων καθώς και από ανεξάρτητες βιβλιοθήκες και εργαλειοθήκες που έχουν αναπτυχθεί και αναπτύσσονται και παρέχονται ξεχωριστά.

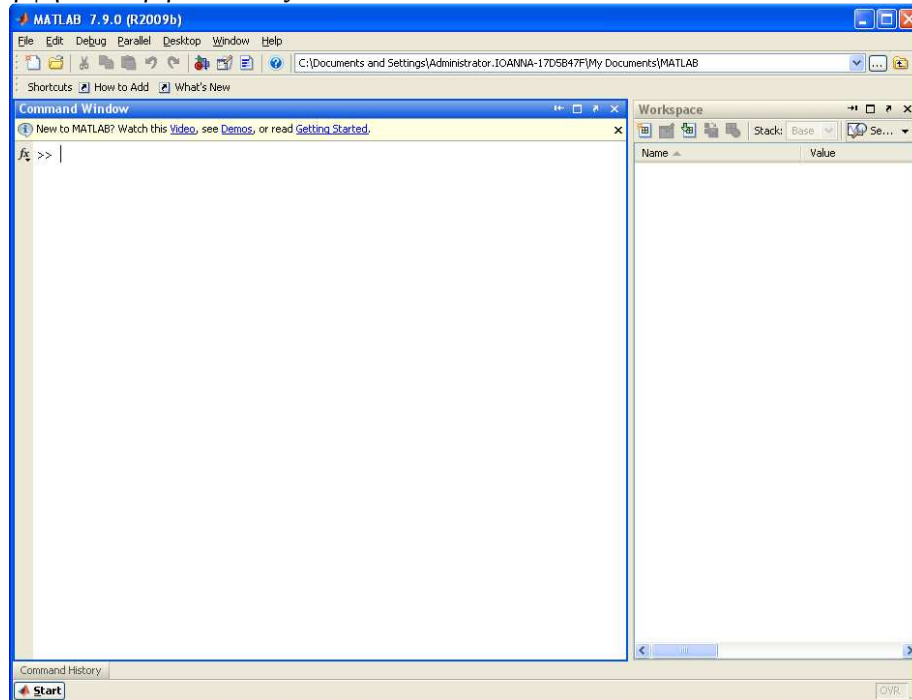
Κάποιος χρήσης ανάλογα με το σύστημα που θέλει να δημιουργήσει χρησιμοποιώντας το MATLAB θα μπορέσει να διαχειριστεί πίνακες, να κάνει διαγράμματα συναρτήσεων και δεδομένων, να υλοποιήσει αλγόριθμους, να δημιουργήσει διεπαφές χρήστη καθώς και να συνδέσει προγράμματα ανεπτυγμένα στο MATLAB με προγράμματα άλλων γλωσσών προγραμματισμού (C, C++, Java και Fortran).

Ιστορία: Το MATLAB αναπτύχθηκε λίγο πριν το 1980 στο Πανεπιστήμιο του Νέου Μεξικού από τον καθηγητή Cleve Moler ο οποίος ανήκε στο τμήμα της πληροφορικής. Ο Moler ήθελε να βρει έναν τρόπο να χρησιμοποιήσουν οι μαθητές του κάποιες βιβλιοθήκες της FORTRAN χωρίς να χρειαστεί να μάθουν τη συγκεκριμένη γλώσσα προγραμματισμού. Αφού το κατάφερε πολλά πανεπιστήμια το χρησιμοποίησαν στη συνέχεια. Μέσα σε αυτά και το Stanford όπου ο μηχανικός Jack Little μαζί με τον Moler και τον Bangert ίδρυσαν το MathWorks και ανέπτυξαν το MATLAB από την αρχή με τη γλώσσα προγραμματισμού C.

Από τότε άρχισαν να το χρησιμοποιούν επιστήμονες των μηχανικών επιστημών αλλά πολλοί γρήγορα διαδόθηκε και άρχισε να χρησιμοποιείται και σε άλλους τομείς. Καθηγητές της γραμμακής άλγεβρας και της αριθμητικής ανάλυσης το διδάσκουν και είναι ιδιαίτερα διαδεδομένο στον τομέα της ψηφιακής επεξεργασίας εικόνας.

3.2. Πώς δουλεύει το Matlab

Περιβάλλον: Όταν ο χρήστης ανοίγει το πρόγραμμα το παράθυρο είναι της μορφής που δίνεται στην Εικόνα 34. Βέβαια το παράθυρο αυτό μπορεί να διαφέρει στη μορφή ανάλογα με τις ρυθμίσεις που κάνει ο κάθε χρήστης. Μεταξύ των διαφορετικών εκδόσεων του MATLAB δεν υπάρχουν ριζικές αλλαγές στη μορφή του περιβάλλοντος.

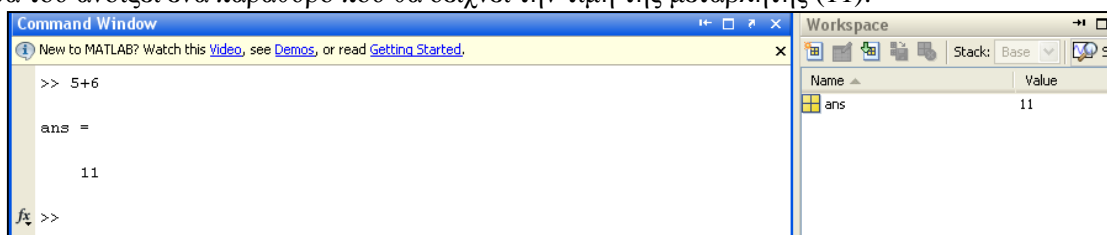


Εικόνα 34 Το περιβάλλον του MATLAB

Ένα από τα βασικά κομμάτια του MATLAB είναι το μεγαλύτερο ορθογώνιο τμήμα που φαίνεται και στην εικόνα και ονομάζεται command window. Είναι το παράθυρο εντολών του προγράμματος. Το MATLAB δουλεύει με συγκεκριμένες εντολές οι οποίες ουσιαστικά είναι μαθηματικές. Αυτές οι εντολές μπορούν να πληκτρολογηθούν στο παράθυρο εντολών του MATLAB.

Ένα απλό παράδειγμα είναι ο χρήστης να θελήσει να πάρει σαν αποτέλεσμα το άθροισμα του 5 και του 6. Τότε θα πληκτρολογήσει στο παράθυρο εντολών $5+6$ και θα του δοθεί το αποτέλεσμα 11 (Εικόνα 35).

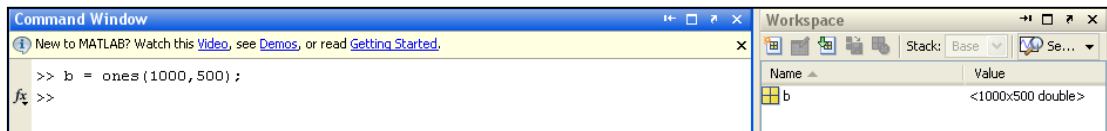
Αν την εντολή την γράψει με ελληνικό ερωτηματικό στο τέλος ($5+6;$) τότε το αποτέλεσμα δε θα εκτυπωθεί στο παράθυρο εντολών. Αλλά όλα τα αποτελέσματα από τις διάφορες εντολές αποθηκεύονται σε μεταβλητές που ορίζει ο χρήστης ή αν δεν ορίσει μεταβλητές αποθηκεύεται μόνο το αποτέλεσμα της τελευταίας εντολής του σαν μεταβλητή ans. Οι μεταβλητές φαίνονται στο κομμάτι του περιβάλλοντος που ονομάζεται workspace στα δεξιά. Αν ο χρήστης κάνει διπλό κλικ πάνω στο ans θα του ανοίξει ένα παράθυρο που θα δείχνει την τιμή της μεταβλητής (11).



Εικόνα 35 Παράδειγμα εκτέλεσης εντολής στο MATLAB

Βέβαια αυτό είναι απαραίτητο στην περίπτωση που η μεταβλητή είναι ένας μεγάλος πίνακας και οι τιμές του δε φαίνονται στο workspace. Στο πρώτο μας παράδειγμα η τιμή είναι μία το 11 και φαίνεται στο workspace (Εικόνα 35).

Ενώ στο δεύτερο παράδειγμα στην Εικόνα 36 δημιουργήσαμε έναν πίνακα με 1000 γραμμές και 500 στήλες τα στοιχεία του οποίου τα θέσαμε όλα ίσα με τη μονάδα. Αυτό τον πίνακα τον αποθηκεύσαμε στο b και δεν τον εκυλώσαμε (b=ones(1000,500)). Με την εκτέλεση αυτή της εντολής η μεταβλητή b φαίνεται στα δεξιά στο παράθυρο του workspace όπου βλέπουμε το όνομα της μεταβλητής (b) και στη στήλη της τιμής δεδομένου ότι δε χωράει όλος ο πίνακας δίνει της διαστάσεις του πίνακα (<1000x500 double>). Ο όρος double σημαίνει ότι ο πίνακας b έχει τιμές αριθμούς τύπου double.

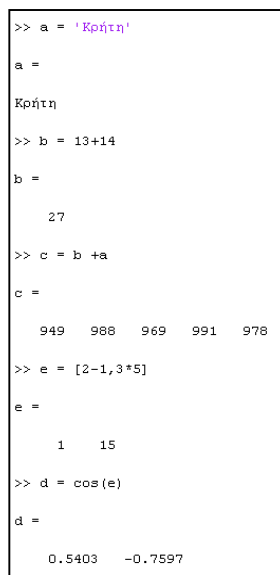


Εικόνα 36 Παράδειγμα εκτέλεσης εντολής στο MATLAB

Έχοντας δώσει δύο χαρακτηριστικά παραδείγματα παρακάτω δίνονται πιο σαφείς και γενικοί ορισμοί.

Μεταβλητές μίας τιμής:

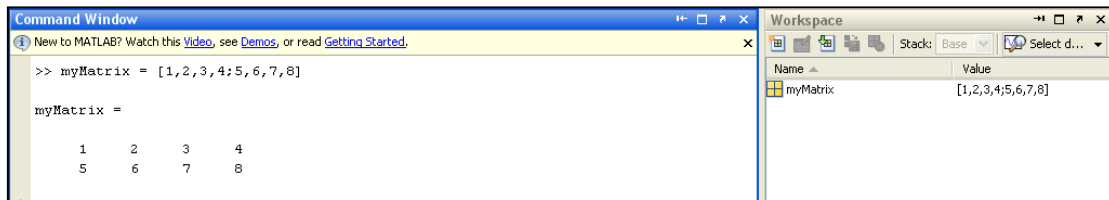
Οι μεταβλητές καθορίζονται με τον τελεστή «=». Το MATLAB δεν είναι μία δυναμική γλώσσα προγραμματισμού για αυτό και οι διάφορες μεταβλητές δε χρειάζεται να οριστούν πριν χρησιμοποιηθούν τι τύπου θα είναι όπως στη C, C++, C#, Java, κ.λ.π.. Οι μεταβλητές μπορούν να προκύψουν είτε από σταθερές, είτε από μαθηματικούς υπολογισμούς με τη χρήση άλλων μεταβλητών ή ακόμα και ως έξοδος μίας συνάρτησης. Δύο τέτοια παραδείγματα είναι οι μεταβλητές a και b που φαίνονται στην Εικόνα 37.



Εικόνα 37 Παράδειγμα δημιουργίας μεταβλητών στο MATLAB

Μεταβλητές πινάκων

Η έννοια του πίνακα στο MATLAB ανταποκρίνεται στην έννοια του πίνακα από τον τομέα της γραμμικής άλγεβρας. Ως πίνακα ορίζουμε ένα μέγεθος το οποίο έχει γραμμές και στήλες και είναι συμπληρωμένος με στοιχεία του ίδιου τύπου, π.χ. αριθμούς ή αλφαριθμητικά. Κάθε στοιχείο του χαρακτηρίζεται από την γραμμή και την στήλη στην οποία ανήκει. Στην δίνεται η εντολή που θα πληκτρολογήσω για να δημιουργήσω έναν πίνακα με δύο γραμμές και τέσσερις στήλες όπου στην πρώτη γραμμή θα έχει με τη σειρά τα στοιχεία 1, 2, 3 και 4 και στη δεύτερη γραμμή τα στοιχεία 5, 6, 7 και 8 και αυτό τον πίνακα να τον αποθηκεύσω σε μία μεταβλητή με όνομα myMatrix.



Εικόνα 38 Παράδειγμα δημιουργίας πίνακα στο MATLAB

Υπάρχουν διάφοροι τελεστές και συναρτήσεις που χρησιμοποιούνται σε πίνακες ή σε μεταβλητές μίας τιμής. Για παράδειγμα αν θέλω να βρω το μέγιστο στοιχείο του πίνακα myMatrix και να το αποθηκεύσω σε μία μεταβλητή maxMyMatrix τότε θα εκτελέσω στο παράθυρο εντολών την εντολή «maxMyMatrix = max(max(myMatrix))», όπου max είναι μία από τις μαθηματικές συναρτήσεις του MATLAB.

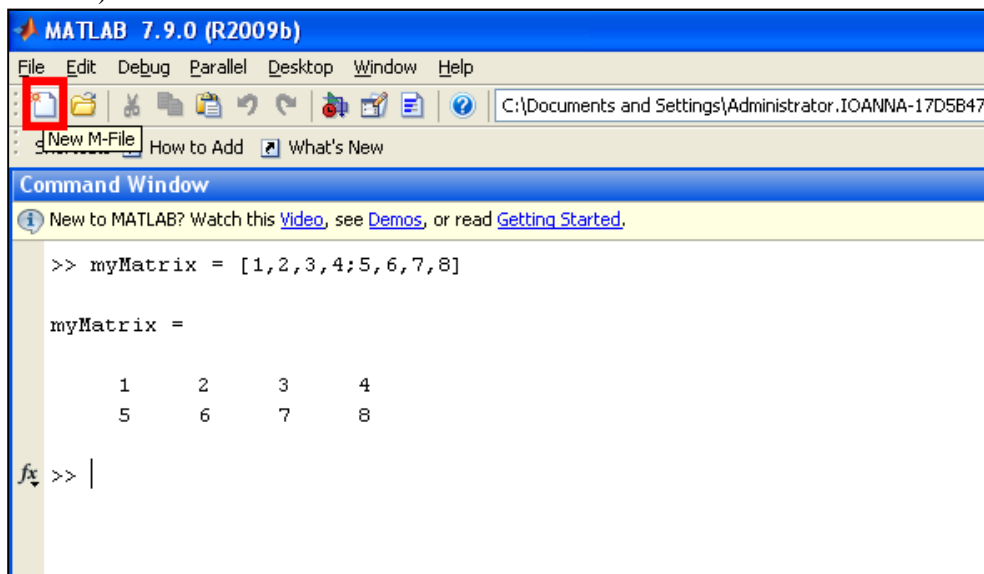
Βασικές επεκτάσεις αρχείων που προκύπτουν από το MATLAB

Αρχεία τύπου .fig:

Αυτά τα αρχεία χρησιμοποιούνται για MATLAB διαγράμματα. Όταν ο χρήστης κάνει το διάγραμμα κάποιων δεδομένων ή κάποιας συνάρτησης μπορεί να το αποθηκεύσει και η μορφή του αρχείου θα είναι .fig.

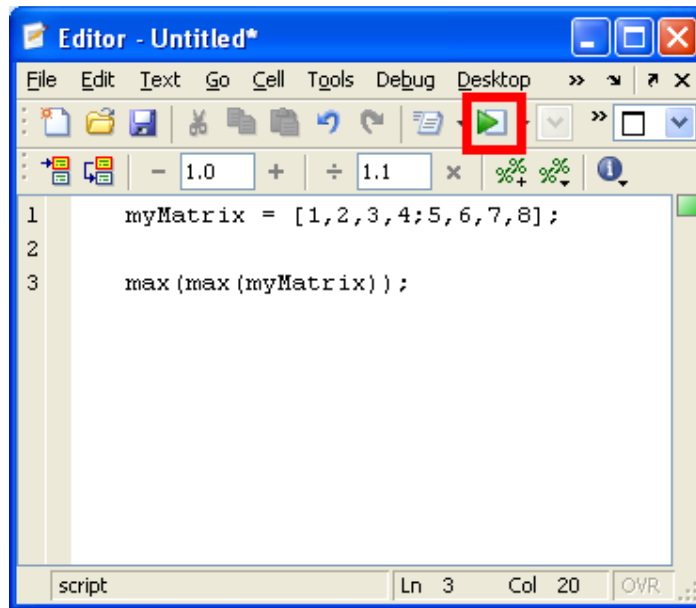
Αρχεία τύπου .m :

Είναι βασικά αρχεία που είναι για την αποθήκευση MATLAB συναρτήσεων, script ή ακόμα και κλάσεων. Όταν ο χρήστης θέλει να εκτελέσει μία σειρά από πολλές εντολές μπορεί να τις γράψει σε ένα αρχείο τύπου .m είτε με τη χρήση του σημειωματάριου των windows και αλλάζοντας την κατάληξη. Αλλά και το Matlab έχει το δικό του σημειωματάριο το οποίο ανοίγει από το κουμπί «New M-File» (Εικόνα 39).



Εικόνα 39 Άνοιγμα του σημειωματαρίου του MATLAB

Στην Εικόνα 40 φαίνεται αυτό το σημειωματάριο που ονομάζεται editor του MATLAB και ο χρήστης μπορεί να πληκτρολογήσει τις εντολές που επιθυμεί την μία κάτω από την άλλη και να το αποθηκεύει σαν αρχείο τύπου .m. Αν θέλει να εκτελέσει όλες τις εντολές πατάει το κουμπί που φαίνεται με κόκκινο στην εικόνα και τα αποτελέσματα φαίνονται στο παράθυρο εντολών και στο workspace του προγράμματος.



Εικόνα 40 Σημειωματάριο του MATLAB (Editor)

Για τη συγκεκριμένη πτυχιακή εργασία όλο το σύστημα βρίσκεται σε αρχεία τύπου m. Δεν εκτελούσαμε ξεχωριστές εντολές μέσα από το παράθυρο εντολών. Με αυτό τον τρόπο μπορέσαμε να κάνουμε πειράματα με διαφορετικά δεδομένα αλλά και να έχουμε αποθηκευμένο το σύστημά μας. Στην περίπτωση που θα εκτελούσαμε τις εντολές μέσα από το παράθυρο εντολών το σύστημα θα χανόταν μόλις θα κλείναμε το MATLAB. Επιπλέον θα ήταν χρονοβόρο.

Αρχεία τύπου .mat

Πρόκειται για MATLAB δυαδικά αρχεία για την αποθήκευση μεταβλητών. Αυτά τα αρχεία χρησιμοποιούνται συνήθως για να αποθηκεύσουμε μεγάλους πίνακες που προκύπτουν σαν αποτέλεσμα. Δεδομένου ότι αν κλείσουμε το Matlab οι μεταβλητές χάνονται έχουμε τη δυνατότητα να τις αποθηκεύσουμε ώστε να αποφύγουμε να ξαναεκτελέσουμε τις ίδιες εντολές. Εξυπηρετεί στις περιπτώσεις που τα αποτελέσματα απαιτούν μεγάλο όγκο δεδομένων και ίσως πολλές εντολές και κατ' επέκταση χρόνο για να προκύψουν. Στη συγκεκριμένη πτυχιακή εργασία τα αρχεία τύπου mat χρησιμοποιήθηκαν σε μεγάλο βαθμό αλλά θα περιγραφεί σε επόμενη ενότητα.

3.3. Εργαλειοθήκη αναγνώρισης προτύπων του Matlab

Χρησιμοποιήθηκε μία εργαλειοθήκη του MATLAB που ονομάζεται `drtoolbox` (dimension reduction toolbox) και παρέχει πληθώρα από μεθόδους για την μείωση των διαστάσεων των δεδομένων, όπως η KPCA και LDA που χρειαστήκαμε εμείς για την ανάπτυξη του ζητούμενου συστήματος. Η συνάρτησεις που χρησιμοποιήσαμε και στις οποίες βασιστήκαμε βρίσκεται μέσα στο φάκελο `techniques` του φακέλου `drtoolbox` και ονομάζονται `kernel_pca.m` και `lda.m`.

Η συνάρτηση `kernel_pca` τροποποιήθηκε ελάχιστα ώστε να δίνει ως έξοδο επιπλέον στοιχεία σε σχέση με αυτά που ήταν προγραμματισμένη να δίνει. Η συνάρτηση αυτή χρησιμοποιήθηκε δίνοντας ως όρισμα τον πίνακα δεδομένων των δεδομένων εκπαίδευσης ώστε να δώσει ως έξοδο τον πίνακα `kernel` πολλαπλασιασμένο με τα ιδιοδιανύσματά του, καθώς και τα ιδιοδιανύσματα. Όσον αφορά τις εισόδους ισχύει:

1. `X`: ο πίνακας των δεδομένων οποίος σε κάθε γραμμή έχει ένα διαφορετικό δεδομένο.
2. `no_dims`: Μπορεί να πάρει ακέραια τιμή μικρότερη ή ίση από την αρχική διάσταση των δεδομένων.
3. `varargin`: Αντιπροσωπεύει δύο ή τρεις παραμέτρους και έχει να κάνει με την συνάρτηση `kernel` που επιθυμούμε να χρησιμοποιηθεί. Η πρώτη παράμετρος παίρνει την τιμή 'gauss' ή 'poly'. Αν πάρει την τιμή `gauss` τότε η συνάρτηση που χρησιμοποιείται ως `kernel` είναι η συνάρτηση `gauss` και με την δεύτερη παράμετρο δίνεται η επιθυμητή τιμή της διασποράς της συνάρτησης. Αν πάρει την τιμή `poly` τότε η συνάρτηση `kernel` είναι πολυωνυμική και με τη δεύτερη παράμετρο δίνεται η σταθερά του πολωνύμου ενώ με την τρίτη ο βαθμός του.

Η συνάρτηση `kernel_pca_test` προέκυψε από αλλαγές που έγιναν πάνω στην `kernel_pca.m` ώστε να μετασχηματίζει τα δεδομένα ελέγχου στην ίδια διάσταση με τα δεδομένα εκπαίδευσης όπως αυτό έγινε με την κλήση της συνάρτησης `kernel_pca`. Η συνάρτηση αυτή δίνει ως έξοδο τον πίνακα `kernel` που προέκυψε από τον συνδυασμό των πινάκων δεδομένων εκπαίδευσης και ελέγχου πολλαπλασιασμένο με τα ιδιοδιανύσματά του πίνακα `kernel` των δεδομένων εκπαίδευσης όπως αυτά προέκυψαν από την `kernel_pca`. Όσον αφορά τις εισόδους ισχύει:

1. `Xtrain`: ο πίνακας των δεδομένων εκπαίδευσης οποίος σε κάθε γραμμή έχει ένα διαφορετικό δεδομένο.
2. `Xtest`: ο πίνακας των δεδομένων ελέγχου οποίος σε κάθε γραμμή έχει ένα διαφορετικό δεδομένο.
3. `Vtrain_KPCA`: Μπορεί να πάρει ακέραια τιμή μικρότερη ή ίση από τον αριθμό των κλάσεων μείον 1.
4. `Mean_rows`: διάνυσμα στήλης το οποίο επιστράφηκε από τη `kernel_pca`. Κάθε στοιχείο είναι το μέσο κάθε γραμμής του πίνακα `kernel` της `kernel_pca`.
5. `varargin`: Ισχύει ότι και στην `kernel_pca`.

Η συνάρτηση `lda` καλείται για να μειώσει περισσότερο τις διαστάσεις των δεδομένων εκπαίδευσης. Ο μέγιστος αριθμός διαστάσεων στον οποίο μπορεί να μειώσει είναι ίσος με τον αριθμό των κλάσεων των δεδομένων μειωμένο κατά 1. Από τις εξόδους μας ενδιαφέρει ο πίνακας δεδομένων εκπαίδευσης μετασχηματισμένος στο νέο χώρο διαστάσεων και ένα πλήθος ιδιοδιανυσμάτων που επιστρέφονται. Όσον αφορά τις εισόδους ισχύει:

1. `X`: ο πίνακας `kernel` των δεδομένων εκπαίδευσης όπως προέκυψε και επιστράφηκε από την `kernel_pca`.
2. `labels`: οι ετικέτες που δείχνουν τις κλάσεις στις οποίες ανήκουν τα δεδομένα εκπαίδευσης.
3. `no_dims`: η διάσταση στην οποία επιθυμούμε να μειωθούν τα δεδομένα.

Στη συνέχεια ο μετασχηματισμένος πίνακας `kernel` που επιστράφηκε από την `kernel_pca_test` πολλαπλασιάζεται με τα ιδιοδιανύσματα που επιστρέφονται από την `lda` ώστε να λάβουμε τον πίνακα δεδομένων ελέγχου μετασχηματισμένα στη μειωμένη διάσταση της `lda` όπως έγινε με τα δεδομένα εκπαίδευσης.

Έχοντας λοιπόν τα κατάλληλα αποτελέσματα η κατηγοριοποίηση γίνεται με τη συνάρτηση του MATLAB `knnclassify`.

3.4. Επιπλέον λογισμικό Matlab για το σύστημά μας

Προκειμένου να κάνουμε τα διάφορα πειράματα με βάση όσα είπαμε παραπάνω αναπτύχθηκε ο κώδικας στο Matlab γιατί εκτός από την εργαλειοθήκη η οποία μας βοήθησε με τις μεθόδους KPCA plus LDA και το `knnclassify`, εκτός από την έτοιμη συνάρτηση του Matlab `kmeans` που υλοποιεί τον ομώνυμο αλγόριθμο θα έπρεπε να κωδικοποιήσουμε και όλες τις υπόλοιπες διαδικασίες ώστε να κατασκευάσουμε το σύστημά μας.

Υποσύστημα της σχηματικής αναγνώρισης SIFT λογισμικό

Για το υποσύστημα της σχηματικής αναγνώρισης βρέθηκε ένας κώδικας από το διαδίκτυο (<http://www.cs.ubc.ca/~lowe/keypoints/>) για να εξάγει τους περιγραφείς SIFT από τις εικόνες (38). Αυτός ο κώδικας είναι μέσα στο φάκελο SOFTWARE\VehiclesCode\1_sift_descriptor και εκτελείται μέσω του αρχείου `main.m`.

Οι εικόνες για τις οποίες βρίσκει τα σημεία ενδιαφέροντος SIFT είναι αυτές που είναι μέσα στο φάκελο `imagesInput` οπότε αν ο χρήστης θέλει να εκτελέσει τον αλγόριθμο για άλλες εικόνες θα πρέπει να τις τοποθετήσει στους αντίστοιχους φακέλους. Η έξοδος του αλγορίθμου είναι δύο αρχεία τύπου `mat` που είναι ουσιαστικά δύο πίνακες του Matlab. Ο ένας πίνακας έχει όλα τα σημεία SIFT για όλες τις εικόνες το ένα κάτω από το άλλο και ο άλλος πίνακας έχει μία στήλη όπου κάθε στοιχείο δίνει τον αριθμό των σημείων κάθε εικόνας.

Για παράδειγμα αν δώσουμε σαν είσοδο δύο εικόνες από τις οποίες η πρώτη έχει 1000 σημεία SIFT και η δεύτερη 216 τότε ο πρώτος πίνακας εξόδου του αλγορίθμου θα έχει στις πρώτες 1000 γραμμές τα σημεία SIFT της πρώτης εικόνας και στις επόμενες 216 της δεύτερης εικόνας. Οι στήλες του πίνακα θα είναι 128 καθώς είπαμε κατά την περιγραφή του αλγορίθμου SIFT ότι κάθε σημείο ενδιαφέροντος έχει 128 τιμές. Ο δεύτερος πίνακας εξόδου θα είναι μία στήλη με δύο στοιχεία όπου το πρώτο στοιχείο θα έχει την τιμή 1000 και το δεύτερο στοιχείο που θα αντιστοιχεί στη δεύτερη εικόνα θα έχει την τιμή 216.

k-means - KPCA plus LDA – knnclassify λογισμικό

Έχοντας τον πίνακα με τα σημεία ενδιαφέροντος SIFT που προέκυψαν από το λογισμικό SIFT και ξεχωρίζοντας τα δεδομένα εκπαίδευσης από τα δεδομένα ελέγχου και έχοντας και τις ετικέτες τους μπορούμε να εκτελέσουμε όλες τις υπόλοιπες διαδικασίες μέσα από το `main.m` του φακέλου SOFTWARE\VehiclesCode\2_KPCA_LDACode. Μέσα από αυτό το αρχείο μπορούμε να καθορίσουμε την συνάρτηση `kernel` που θα χρησιμοποιηθεί, αν θα είναι `gauss`, `polynomial`, `rbf`, `sigmoid` καθώς και να καθορίσουμε τις παραμέτρους για κάθε μία από αυτές όπως για παράδειγμα το βαθμό της `polynomial` συνάρτησης αν την επιλέξουμε. Η έξοδος είναι οι ετικέτες των αντικειμένων ελέγχου, οι κατηγορίες δηλαδή που ανήκουν τα οχήματα των εικόνων που θέλουμε να αναγνωρίσουμε.

Υποσύστημα της χρωματικής αναγνώρισης

Μέσα στο φάκελο SOFTWARE\PlateCode υπάρχει αρχείο `main.m` που υλοποιεί ουσιαστικά τις ιδιότητες του Πίνακα 1. Αφού ο χρήστης επιλέξει την εικόνα που επιθυμεί να αναγνωρίσει το σύστημα μέσα από το φάκελο SOFTWARE\PlateCode\CARPLATES στο αρχείο `main.m` θα πρέπει να αλλάξει τις γραμμές κώδικα

```
folderName = 'RED';  
fileName = 'images3.jpg';
```

και να δίνει στη μεταβλητή `folderName` μία από τις τρεις τιμές 'RED', 'YELLOW' και 'WHITE' ανάλογα με το φάκελο στον οποίο είναι η εικόνα του. Στην μεταβλητή `filename` θα πρέπει να δίνει την τιμή της εικόνας. Η έξοδος του συστήματος είναι το χρώμα της πινακίδας.

Υποσύστημα τιμολόγησης

Έχοντας τα αποτελέσματα των δύο προηγούμενων συστημάτων, δηλαδή της κατηγορίας του οχήματος και του χρώματος της ταμπέλας έχει δημιουργηθεί μία συνάρτηση στο Matlab η οποία παίρνει αυτές τις δύο εξόδους ως είσοδο και δίνει την τιμή που αντιστοιχεί στο συγκεκριμένο όχημα. Η συνάρτηση αυτή είναι η `timologisi.m` του φακέλου `SOFTWARE\PriceCode`. Οι τιμές που υπάρχουν μέσα στη συνάρτηση αυτή τη στιγμή είναι ενδεικτικές και μπορούν να αλλάξουν από τον χρήστη μέσα από τον κώδικα.

4. Εφαρμογές

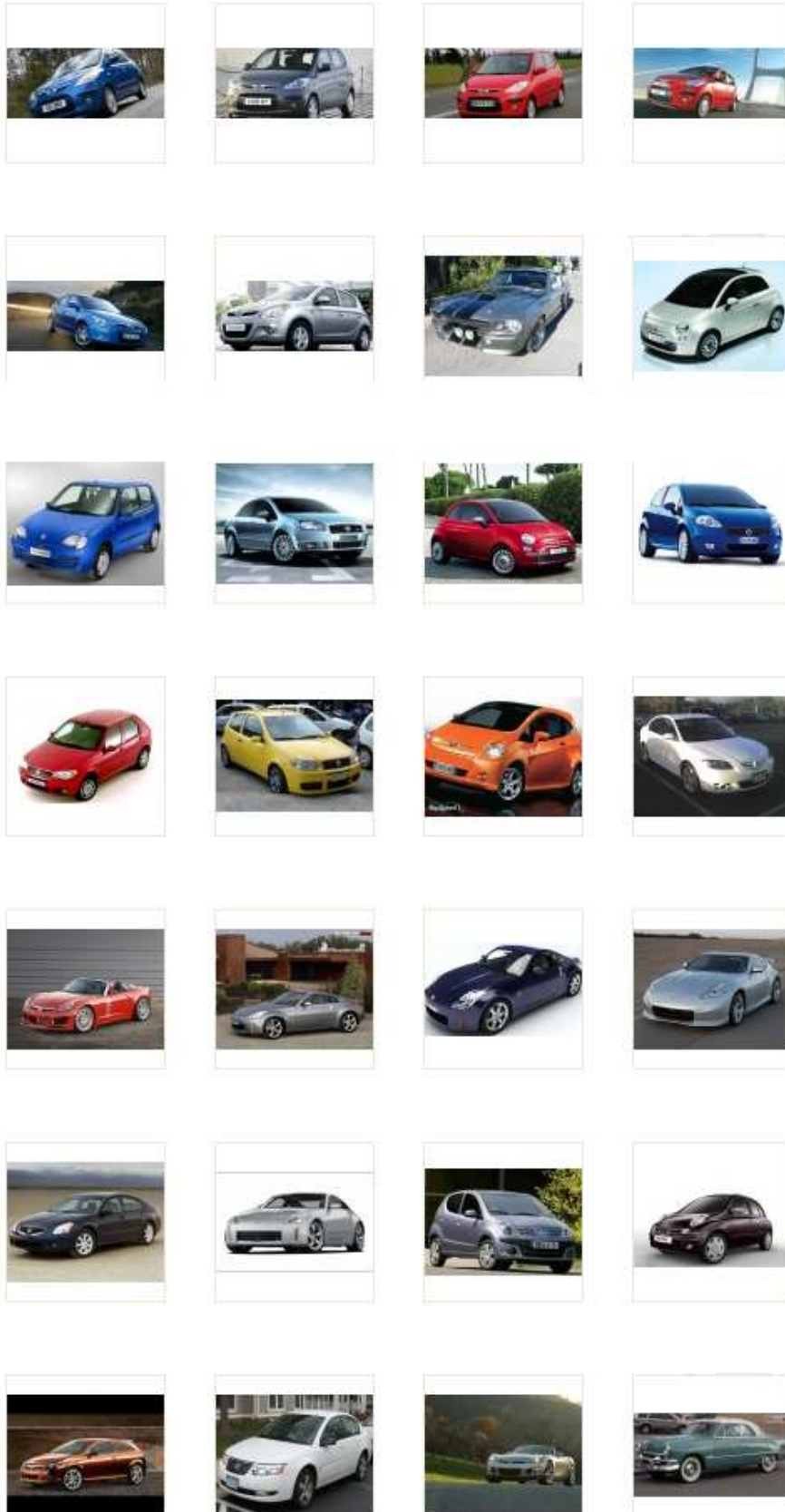
Στις επόμενες σελίδες ακολουθούν μικρογραφίες των εικόνων της βάσης που χρησιμοποιήθηκαν στα πειράματά μας για τα δεδομένα εκπαίδευσης και ελέγχου.





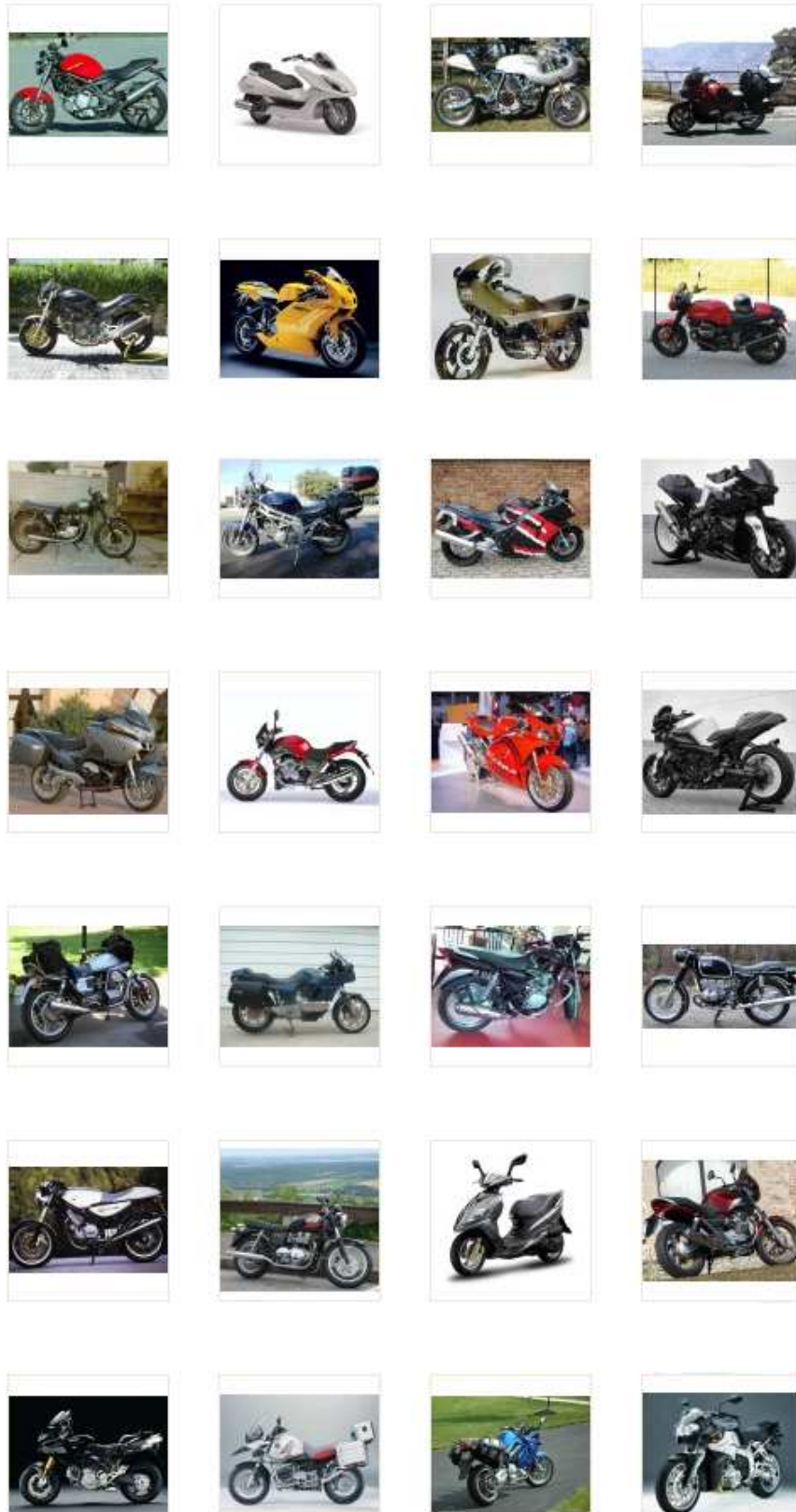


















Εικόνα 41 Thumbnails των εικόνων της βάσης που χρησιμοποιήθηκαν στα πειράματα

5.1. 1η Εφαρμογή και αποτελέσματα

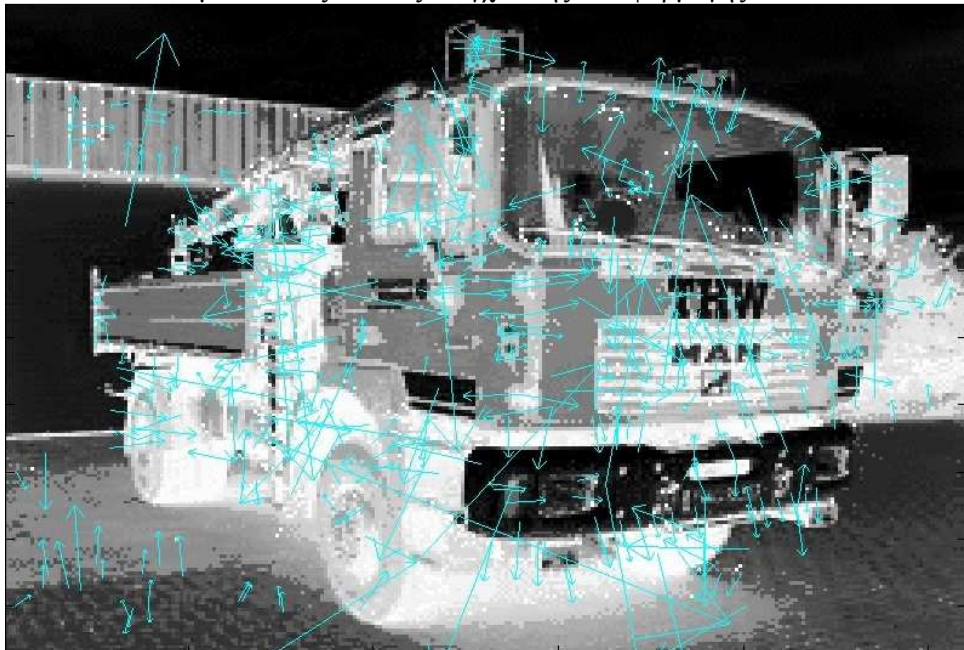
Σε ένα από τα πειράματα που υλοποιήσαμε και παρουσιάζουμε ως πρώτη εφαρμογή χρησιμοποιήσαμε ως δεδομένα εκπαίδευσης τα δύο τρίτα των εικόνων της βάσης δεδομένων για το υποσύστημα σχηματικής αναγνώρισης και δώσαμε ως εισόδους τις τρεις εικόνες που φαίνονται στην Εικόνα 42. Στην Εικόνα 43, στην Εικόνα 44 και στην Εικόνα 45 απεικονίζονται τα σημεία SIFT που προέκυψαν για καθεμία από τις εικόνες ελέγχου. Κάθε σημείο απεικονίζεται με ένα βέλος που δείχνει τον προσανατολισμό του.



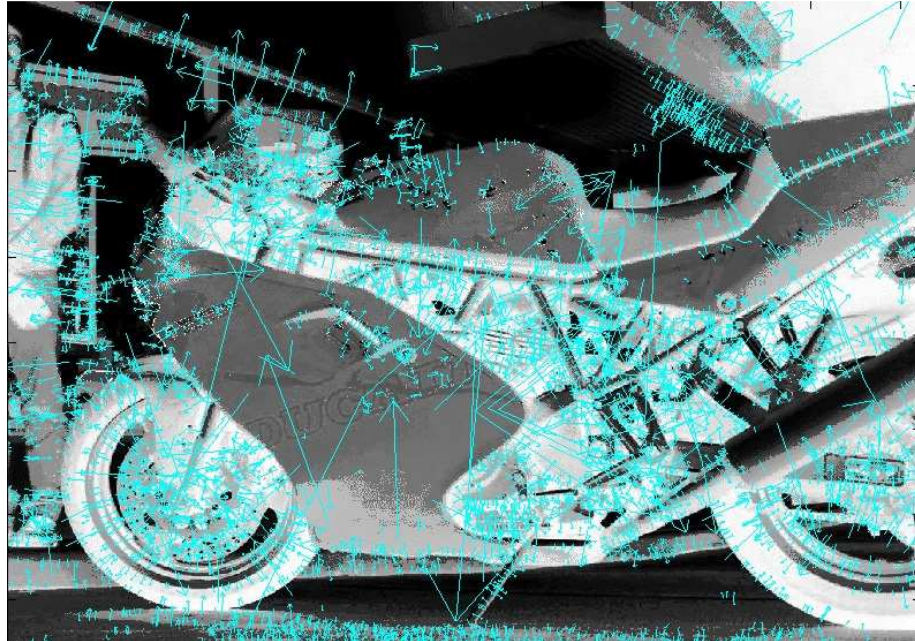
Εικόνα 42 Εικόνες ελέγχου της 1^{ης} εφαρμογής



Εικόνα 43 Απεικόνιση των 964 σημείων SIFT σε μία από τις εικόνες ελέγχου της 1^{ης} εφαρμογής



Εικόνα 44 Απεικόνιση των 426 σημείων SIFT σε μία από τις εικόνες ελέγχου της 1^{ης} εφαρμογής



Εικόνα 45 Απεικόνιση των 5.361 σημείων SIFT σε μία από τις εικόνες ελέγχου της 1^{ης} εφαρμογής

Το k-means έγινε για 200 κέντρα καθώς για λιγότερα η επίδοση του ταξινομητή ήταν χαμηλή ενώ για περισσότερα που ίσως και να μην είναι απαραίτητο υπήρχε αρκετή χρονική καθυστέρηση.

Έγιναν πειράματα για δύο συναρτήσεις kernel , την πολυωνυμική και την γκαουσιανή. Στην πολυωνυμική συνάρτηση αλλάζαμε το βαθμό του πολυωνύμου και τη σταθερά του για να δούμε για τις διάφορες περιπτώσεις την επίδοση του ταξινομητή ενώ στην γκαουσιανή αλλάζαμε τη διασπορά της συνάρτησης. Η γκαουσιανή συνάρτηση αναγνώριζε πλήρως τα οχήματα, είχε 100% επιτυχία. Το ίδιο δεν ισχύει και για την πολυωνυμική συνάρτηση.

Ο Πίνακας 2 δίνει τα αποτελέσματα στην περίπτωση της πολυωνυμικής συνάρτησης. Αυτό που παρατηρούμε από τα ποσοστά επιτυχούς αναγνώρισης των δεδομένων εκπαίδευσης τα οποία κυμαίνονται από ~97% έως και 100% είναι ότι έγινε ορθή μάθηση της μηχανής. Στη συνέχεια από την τέταρτη στήλη, όπου καλείται να γίνει αναγνώριση των τριών οχημάτων της Εικόνα 42, και έχοντας παρατηρήσει και τις ετικέτες οι οποίες δε δίνονται στον πίνακα το συμπέρασμά μας είναι ότι πάντα αναγνωρίζει το αυτοκίνητο για αυτό και το ποσοστό επιτυχίας είναι τουλάχιστον 33,333333%. Το πρόβλημα είναι ότι μπερδεύει τα λεωφορεία με τα φορτηγά. Στις περιπτώσεις όπου το ποσοστό είναι 66.666667% σημαίνει ότι ή το λεωφορείο ή το φορτηγό το αναγνώρισε ως φορτηγό ή λεωφορείο αντίστοιχα. Τα πειράματα όπου έχει γίνει ορθή αναγνώριση και των τριών οχημάτων σημειώνονται με πορτοκαλί στον πίνακα.

Βαθμός πολυωνυμικής συνάρτησης	Σταθερά πολυωνυμικής συνάρτησης	Ποσοστό επιτυχίας αναγνώρισης των δεδομένων εκπαίδευσης	Ποσοστό επιτυχίας αναγνώρισης των δεδομένων εκπαίδευσης
2	0	100	100
3	0	100	66,66666667
9	0	100	33,33333333
10	0	100	66,66666667
11	0	100	66,66666667
12	0	100	33,33333333
13	0	100	33,33333333
14	0	97,26027397	33,33333333
15	0	97,26027397	0
2	200	100	50
3	200	100	66,66666667
4	200	100	33,33333333
7	200	100	33,33333333
10	200	100	33,33333333
11	200	100	66,66666667
12	200	100	0
13	200	100	100
14	200	98,63013699	50
15	200	93,15068493	33,33333333
2	400	100	50
3	400	100	66,66666667
4	400	100	33,33333333
9	400	100	33,33333333
10	400	100	66,66666667
11	400	100	66,66666667
12	400	100	33,33333333
13	400	100	33,33333333
14	400	97,26027397	50
15	400	95,89041096	100
2	600	100	50
3	600	100	66,66666667
4	600	100	66,66666667
5	600	100	66,66666667
6	600	100	33,33333333
7	600	100	66,66666667
8	600	100	66,66666667
9	600	100	33,33333333
10	600	100	33,33333333
11	600	100	66,66666667

12	600	100	33,33333333
13	600	100	33,33333333
14	600	98,63013699	66,66666667
15	600	93,15068493	33,33333333
2	800	100	50
3	800	100	66,66666667
4	800	100	66,66666667
5	800	100	66,66666667
6	800	100	33,33333333
7	800	100	66,66666667
8	800	100	66,66666667
9	800	100	33,33333333
10	800	100	33,33333333
11	800	100	66,66666667
14	800	97,26027397	33,33333333
15	800	95,89041096	100
3	1000	100	66,66666667
4	1000	100	66,66666667
7	1000	100	33,33333333
8	1000	100	66,66666667
9	1000	100	33,33333333
10	1000	100	66,66666667
11	1000	100	66,66666667
14	1000	97,26027397	33,33333333
15	1000	95,89041096	100

Πίνακας 2 Αποτελέσματα 1^{ης} εφαρμογής για διάφορες παραμέτρους της πολυωνμικής kernel συνάρτησης

Όσον αφορά στις τρεις ταμπέλες των οχημάτων και οι τρεις είναι λευκές και τοποθετώντας τις εικόνες που αντιστοιχούν στα τρία οχήματα μας δίνει επιτυχώς το αποτέλεσμα του χρώματός τους (Εικόνα 46).



Εικόνα 46 Οι ταμπέλες που αντιστοιχούν στα τρία οχήματα ελέγχου της 1^{ης} εφαρμογής

4.1. 2η Εφαρμογή και αποτελέσματα

Σε άλλο πείραμά μας χρησιμοποιήσαμε ως δεδομένα εκπαίδευσης τα δύο τρίτα των εικόνων της βάσης δεδομένων αλλά επιλέξαμε κάποιες να διαφέρουν από τις αντίστοιχες εικόνες εκπαίδευσης της 1^{ης} εφαρμογής για το υποσύστημα σχηματικής αναγνώρισης και δώσαμε ως εισόδους τις τρεις εικόνες που φαίνονται στην Εικόνα 42.

Το k-means έγινε για 200 κέντρα καθώς για λιγότερα η επίδοση του ταξινομητή ήταν χαμηλή ενώ για περισσότερα που ίσως και να μην είναι απαραίτητο υπήρχε αρκετή χρονική καθυστέρηση.

Έγιναν πειράματα για δύο συναρτήσεις kernel , την πολυωνυμική και την γκαουσιανή. Στην πολυωνυμική συνάρτηση αλλάζαμε το βαθμό του πολυωνύμου και τη σταθερά του για να δούμε για τις διάφορες περιπτώσεις την επίδοση του ταξινομητή ενώ στην γκαουσιανή αλλάζαμε τη διασπορά της συνάρτησης.



Εικόνα 47 Εικόνες ελέγχου της 2^{ης} εφαρμογής

Ο Πίνακας 3 δίνει τα αποτελέσματα στην περίπτωση της πολυωνυμικής συνάρτησης. Αυτό που παρατηρούμε από τα ποσοστά επιτυχούς αναγνώρισης των δεδομένων εκπαίδευσης τα οποία κυμαίνονται από ~75% έως και 100% και στις περισσότερες περιπτώσεις πάνω από ~93% είναι ότι έγινε ορθή μάθηση της μηχανής. Το γενικό συμπέρασμα από αυτά τα πειράματα είναι ότι η αναγνώριση είναι πιθανότερο να είναι ορθή όταν ο βαθμός της πολυωνυμικής συνάρτησης είναι μικρός.

Βαθμός πολυωνυμικής συνάρτησης	Σταθερά πολυωνυμικής συνάρτησης	Ποσοστό επιτυχής αναγνώρισης των δεδομένων εκπαίδευσης	Ποσοστό επιτυχής αναγνώρισης των δεδομένων εκπαίδευσης
3	0	100	100
4	0	100	100
5	0	100	100
6	0	100	100
7	0	100	66,66666667
8	0	100	100
9	0	100	66,66666667
10	0	100	0
11	0	100	0
12	0	98,63013699	33,33333333
13	0	97,26027397	100
14	0	93,15068493	33,33333333
15	0	95,89041096	33,33333333
2	200	100	100
3	200	100	100
4	200	100	100
5	200	100	100
6	200	100	100
7	200	100	66,66666667
8	200	100	100
9	200	100	100
10	200	100	100
11	200	100	50
12	200	100	0
13	200	97,26027397	50
14	200	93,15068493	33,33333333
15	200	75,34246575	33,33333333
2	400	100	100
3	400	100	100
4	400	100	100
5	400	100	100
6	400	100	100
7	400	100	66,66666667
8	400	100	66,66666667
9	400	100	66,66666667
13	400	97,26027397	66,66666667
14	400	97,26027397	0
15	400	80,82191781	0
2	600	100	100
3	600	100	100

4	600	100	100
5	600	100	100
6	600	100	100
7	600	100	66,66666667
8	600	100	100
9	600	100	100
11	600	100	100
12	600	100	100
13	600	97,26027397	33,33333333
14	600	95,89041096	66,66666667
15	600	78,08219178	0
3	800	100	100
4	800	100	100
5	800	100	100
6	800	100	100
7	800	100	66,66666667
8	800	100	66,66666667
9	800	100	100
10	800	100	100
11	800	100	100
12	800	100	0
13	800	97,26027397	33,33333333
14	800	95,89041096	33,33333333
15	800	75,34246575	0
2	1000	100	100
3	1000	100	66,66666667
4	1000	100	100
5	1000	100	100

Πίνακας 3 Αποτελέσματα 2^{ης} εφαρμογής για διάφορες παραμέτρους της πολυωνμικής kernel συνάρτησης

Ο Πίνακας 4 δίνει τα αποτελέσματα στην περίπτωση της γκαουσιανής συνάρτησης. Αυτό που παρατηρούμε από τα ποσοστά επιτυχούς αναγνώρισης των δεδομένων εκπαίδευσης είναι ότι έγινε ορθή μάθηση της μηχανής. Το συμπέρασμα από αυτά τα πειράματα είναι ότι η αναγνώριση στα δεδομένα ελέγχου γίνεται απολύτως ορθά για τιμές τις διασποράς πάνω από 5.

Διασπορά γκουουσιανής συνάρτησης	Ποσοστό επιτυχής αναγνώρισης των δεδομένων εκπαίδευσης	Ποσοστό επιτυχής αναγνώρισης των δεδομένων εκπαίδευσης
3	100	33,33333
3,5	100	33,33333
4	100	33,33333
4,5	100	33,33333
5	100	66,66667
5,5	100	66,66667
6	100	66,66667
6,5	100	66,66667
7	100	66,66667
7,5	100	66,66667
8	100	100
8,5	100	100
9	100	100
9,5	100	100
10	100	100
10,5	100	100
15	100	100

Πίνακας 4 Αποτελέσματα 2^{ης} εφαρμογής για διάφορες τιμές της διασποράς της γκουουσιανής kernel συνάρτησης

Όσον αφορά στις τρεις ταμπέλες των οχημάτων μπορούμε να θεωρήσουμε ότι είναι και των τριών διαφορετικών χρωμάτων τοποθετώντας τις εικόνες στον αλγόριθμο αναγνώρισής τους μας δίνει επιτυχώς το αποτέλεσμα του χρώματός τους (Εικόνα 48).



Εικόνα 48 Οι ταμπέλες που αντιστοιχούν στα τρία οχήματα ελέγχου της 2^{ης} εφαρμογής

4.2. 3η Εφαρμογή και αποτελέσματα

Σαν 3^η εφαρμογή θα παρουσιάσουμε την περίπτωση όπου έχουμε ισάξιο ποσοστό εικόνων δεδομένων και από τις τέσσερις κατηγορίες και σαν αντικείμενα ελέγχου δίνουμε έξι μοτοσυκλέτες (Εικόνα 49).

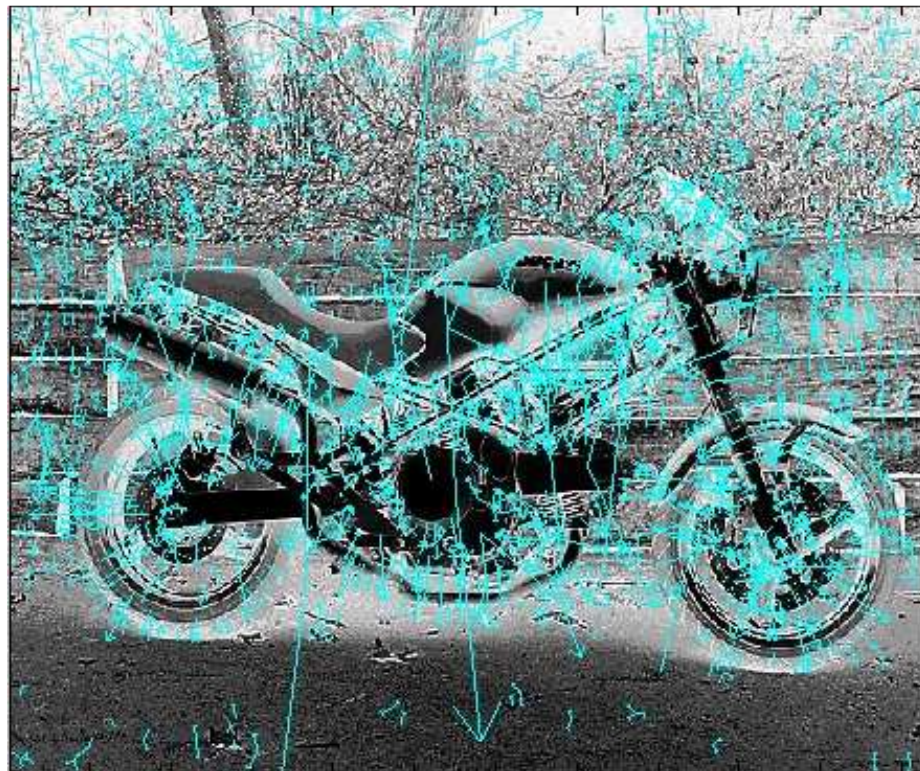


Εικόνα 49 Εικόνες ελέγχου της 3^{ης} εφαρμογής

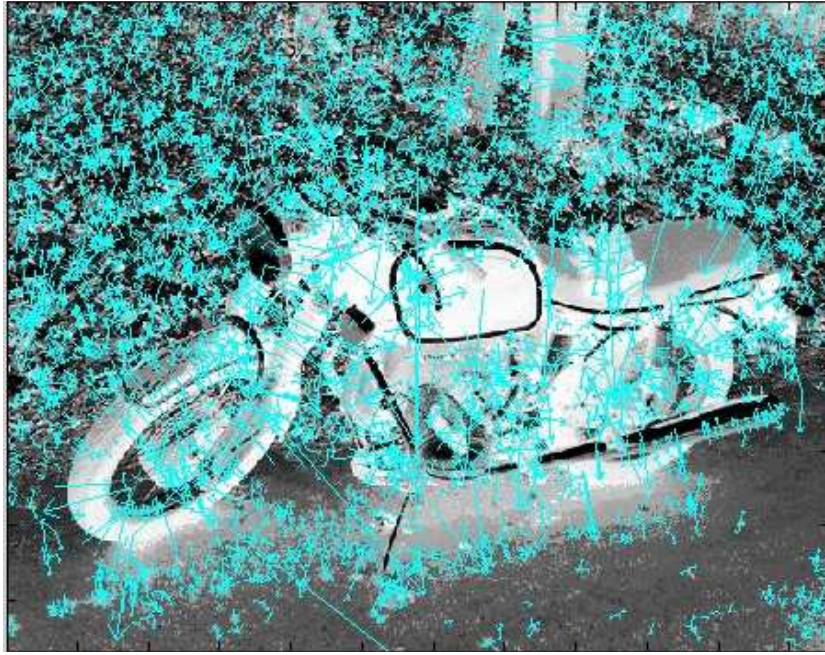
Παρακάτω δίνονται οι απεικονίσεις των σημείων SIFT για τις έξι μηχανές. Σε αυτή την εφαρμογή μας τα πειράματα έδειξαν και για γκαουσιανές συναρτήσεις kernel και για πολυωνυμικές ότι η αναγνώριση γίνεται με επιτυχία πάνω από ~92% για διάφορες παραμέτρους των συναρτήσεων. Για τις ταμπέλες ισχύει ότι και στις άλλες δύο εφαρμογές, αναγνωρίζονται απόλυτα.



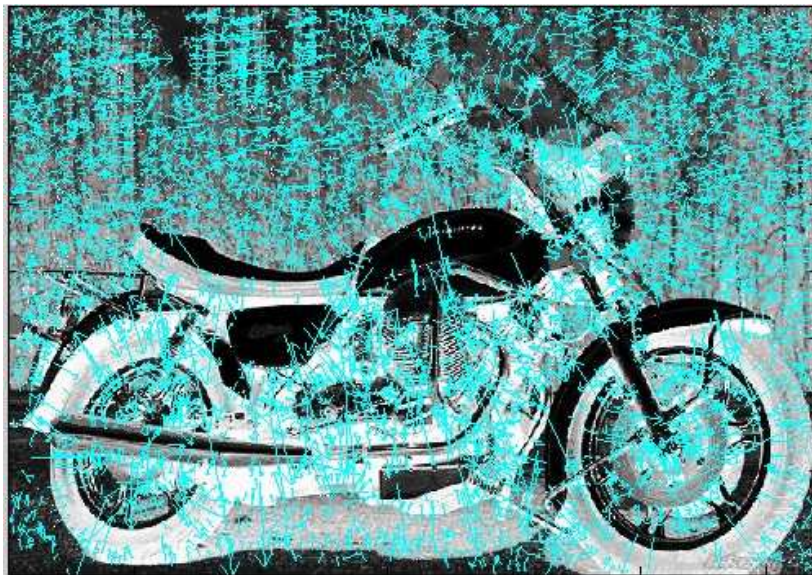
Εικόνα 50 Απεικόνιση των 1.618 σημείων SIFT σε μία από τις εικόνες ελέγχου της 3^{ης} εφαρμογής



Εικόνα 51 Απεικόνιση των 1.760 σημείων SIFT σε μία από τις εικόνες ελέγχου της 3^{ης} εφαρμογής



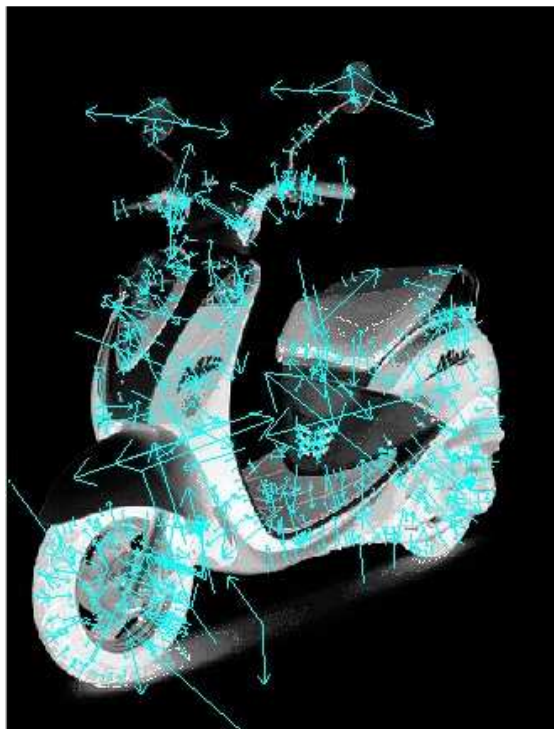
Εικόνα 52 Απεικόνιση των 4.161 σημείων SIFT σε μία από τις εικόνες ελέγχου της 3^{ης} εφαρμογής



Εικόνα 53 Απεικόνιση των 5.233 σημείων SIFT σε μία από τις εικόνες ελέγχου της 3^{ης} εφαρμογής



Εικόνα 54 Απεικόνιση των 5.631 σημείων SIFT σε μία από τις εικόνες ελέγχου της 3^{ης} εφαρμογής



Εικόνα 55 Απεικόνιση των 774 σημείων SIFT σε μία από τις εικόνες ελέγχου της 3^{ης} εφαρμογής

5. Αποτελέσματα της μελέτης και της έρευνας

5.1. Συμπεράσματα

Με βάση το σύστημα που αναπτύξαμε το συμπέρασμά μας για τον αλγόριθμο SIFT είναι ότι είναι κατάλληλος για την εξαγωγή χαρακτηριστικών σε αντικείμενα όπου μας ενδιαφέρει το σχήμα των αντικειμένων που απεικονίζονται. Αυτό που ισχύει για τις διαφορετικές κατηγορίες των οχημάτων είναι ότι διαφέρουν ως προς το σχήμα και αυτό μπορεί να καθοριστεί με τους περιγραφείς SIFT οι οποίοι είναι ανεξάρτητοι από τον φωτισμό της σκηνής την περιστροφή ή το μέγεθος της εικόνας.

Η ανεξαρτησία του SIFT σε τέτοιες συνθήκες είναι απαραίτητες για την περίπτωση μας καθώς οι φωτογραφίες γίνονται σε εξωτερικό χώρο καθ' όλη τη διάρκεια της ημέρας και τα οχήματα μπορεί να φωτογραφηθούν σε διαφορετική απόσταση από την κάμερα οπότε να υπάρχουν διαφορετικά μεγέθη καθώς και διαφορετικοί προσανατολισμοί των οχημάτων. Αυτές οι διαφορές λοιπόν στις διάφορες φωτογραφίες ίδιων οχημάτων ξεπερνιούνται με τα σημεία ενδιαφέροντος SIFT.

Έχοντας λοιπόν κάνει ορθή επιλογή των χαρακτηριστικών επειδή η διάστασή τους είναι υψηλή και άρα υπάρχει μεγάλος όγκος δεδομένων που σημαίνει μεγάλες απαιτήσεις σε μνήμη του υπολογιστή και χρονοβόρα διαδικασία έπρεπε αυτή τη διάσταση να τη μειώσουμε. Η μέθοδος KPCA plus LDA ήταν απαραίτητη στο σύστημά μας αλλιώς θα ήταν αδύνατη η υλοποίηση του συστήματος και ο έλεγχός του από έναν προσωπικό ηλεκτρονικό υπολογιστή.

Η kernel συνάρτηση που χρησιμοποιήθηκε στην KPCA μέσα από τα πειράματα ήταν η πολυωνυμική και η γκαουσιανή. Τα πειράματα έδειξαν ότι στις περισσότερες περιπτώσεις είναι πιο σίγουρη η επιλογή της γκαουσιανής συνάρτησης ώστε να δώσει ορθά αποτελέσματα.

Ο ταξινομητής πλησιέστερου γείτονα αποδείχθηκε αποδοτικός για το σύστημά μας για αυτό δεν κάναμε περαιτέρω πειράματα για άλλους ταξινομητές.

5.2. Προτάσεις επέκτασης της εφαρμογής και εμπορικής εκμετάλλευσης

Η συγκεκριμένη εφαρμογή μπορεί να βελτιωθεί κάνοντας πειράματα και με άλλες παραμέτρους ή και μεθόδους μείωσης διάστασης ώστε να βελτιωθεί ο χρόνος εκτέλεσής της και το υπολογιστικό κόστος που θα απαιτεί.

Πρόκειται για μία έξυπνη εφαρμογή που αν τελειοποιηθεί θα μπορεί να αποτελέσει πολύ καλή εμπορική εκμετάλλευση καθώς θα αυξηθεί η ταχύτητα εξυπηρέτησης των οδηγών στα διόδια. Επίσης το κόστος για τη λειτουργία των διοδίων θα είναι πολύ μικρότερο αφού με τη χρήση ενός τέτοιου συστήματος η εξυπηρέτηση θα γίνεται αυτόματα.

6. Βιβλιογραφία

1. **Rafael C. Gonzalez, Richard E. Woods.** *Digital Image Processing.* s.l. : Addison-Wesley, 1992.
2. **Starck, Jean-Luc.** *Image processing and data analysis: The multiscale approach.*
3. **Bijaoui, A.** *Wavelets and the analysis of astronomical objects,*. s.l. : Large-Scale Structures in Nonlinear Physics, Springer-Verlag, 1991.
4. —. *Algorithms de la transformation en ondelettes.* s.l. : Ondelettes et Paquets d'Ondes, Cours CEA/EdF/INRIA, 1991.
5. **Bijaoui.** *Wavelets and astronomical image analysis, Wavelets, Fractals and Fourier Transforms: New Developments and New Applications.* Oxford, : M. Farge, J.C.R. Hunt and J.C. Vassilicos, Oxford University Press, 1993.
6. **A. J. Ahumada, Jr., A. M. Rohaly, A. B. Watson.** *Image Discrimination Models Predict Object Detection in Natural Backgrounds.* s.l. : Investigative Ophthalmology and Visual Science, 1995. Vol. 36.
7. **P. Carbonetto, N. de Freitas, and K. Barnard.** *A statistical model for general contextual object recognition.* s.l. : Proc. ECCV, 2004.
8. **Koller, G. Heitz and D.** *Learning spatial context: Using stuff to find things.* s.l. : Proc. ECCV, 2008.
9. **Triggs, N. Dalal and B.** *Histograms of oriented gradients for human detection.* s.l. : Proc CVPR, 2005.
10. **S.Bileschi, L. Wolf and.** *A critical view of context.* s.l. : IJVC, 2006.
11. **Santosh K. Divvala, Derek Hoiem, James H. Hays, Alexei A. Efros, Martial Hebert.** *An Empirical Study of Context in Object Detection.*
12. **D. Hoiem, A. Efros, and M. Hebert.** *Recovering occlusion boundaries from a single image.* ICCV : s.n., 2007.
13. **Hua, Ming Yang and Ying Wu and Gang.** *Context-Aware Visual Tracking.* s.l. : IEEE Transactions on Pattern Analysis and Machine Intelligence, 2009.
14. **Kuncheva, Ludmila I.** *Combining Pattern Classifiers.* s.l. : John Wiley and Sons, 2004.
15. **A. Di Lecce, V. Guerriero.** *An evaluation of the effectiveness of image features for image retrieval.* s.l. : Visual Commun. Image Representation 10, 1999.
16. <http://yann.lecun.com/exdb/mnist/>. [Online]
17. **Hart., R. O. Duda and P. E.** *Pattern Classification and Scene Analysis.* . s.l. : John Wiley & Sons, 1973.

18. **Sameer Antani, Rangachar Kasturi, Ramesh Jain.** *A survey on the use of pattern recognition methods for abstraction, indexing and retrieval of images and video.* s.l. : Pattern Recognition, 2002.
19. **Liao M.Z.W., Ling W. and Chen W.F.** *A Novel Affine Invariant Feature Extraction for Optical Recognition.* s.l. : International Conference on Machine Learning and Cybernetics, 2007.
20. **F., Orguner U. and Gustafsson.** *Statistical Characteristics of Harris Corner Detector.* s.l. : Processing of 14th Workshop on Statistical Signal, SSP '07, IEEE/SP, 2007.
21. **Russ, J.C.** *The Image Processing Handbook.* s.l. : Fifth Edition, CRC Press, 2007.
22. **M., Vincze.** *Robust Tracking of Ellipses at Frame Rate.* s.l. : Pattern Recognition, 2001.
23. **Bay H., Tuytelaars T., and Van Gool L.** *Surf: Speeded up robust features.* Austria : Proceedings of European Conference on Computer Vision, Graz, 2006.
24. **Bose, B., Guyon, I. and Vapnik, V.** *A training algorithm for optimal margin classifiers.* Baltimore : Fifth Annual Workshop on Computational Learning Theory, 1992.
25. **Leon Botton, Corinna Cortes, John S. Denker, Harris Drucker, Isabelle Guyon, L.D. Jackel.** *Comparison of classifier methods: A case study in handwritten digit recognition.*
26. **Vapnik, V.** *The Nature of Statistical Learning Theory.* New York : Springer-Verlag, 1995.
27. **Steven C. H. Hoi, Rong Jin and Michael R. Lyu.** *Learning Nonparametric Kernel Matrices from Pairwise Constraints.*
28. **J.K. Wu, P. Lam, B.M. Mehtre, Y.J. Gao, A.D. Narasimhalu.** *Content based retrieval for trademark registration.* s.l. : Multimedia Tools Appl., 1996.
29. **T. Gevers, A.W.M. Smeulders.** *Pictoseek: combining color and shape invariant features for image retrieval.* s.l. : IEEE Trans. Image Process, 2000.
30. —. *Pictoseek: combining color and shape invariant features for image retrieval.* s.l. : IEEE International Conference on Multimedia Computing Systems, 1999.
31. **W. Niblack, X. Zhu, J.L. Hafner, T. Breuel.** *Updates to the QBIC system, Proceedings of IS&T/SPIE.* 1997.
32. **Lindeberg T., and Romeny T. H.** *Linear scale-space in Geometry-Driven Diffusion.* s.l. : . s.l. : Kluwer Academic Publishers.
33. **C, Mikolajczyk K. and Schmid.** *A performance evaluation of local descriptors.* s.l. : IEEE Transactions on Pattern Analysis & Machine Intelligence, 2005.
34. **Bose, B., Guyon, I. and Vapnik, V.** *A training algorithm for optimal margin classifiers.* s.l. : Fifth Annual Workshop on Computational Learning Theory, 1992.
35. **MacQueen, J. B.** *Some Methods for classification and Analysis of Multivariate Observations.* s.l. : University of California Press, 1967.

36. **Jian Yang, Zhong Jina, Jing-yu Yang, David Zhang, Alejandro F. Frangi.** *Essence of kernel Fisher discriminant: KPCA plus LDA.* s.l. : Pattern Recognition, 2003.
37. **Jian Yang, Alejandro F. Frangi, Jing-yu Yang, David Zhang, Senior Member, IEEE, and Zhong Jin.** *KPCA Plus LDA: A Complete Kernel Fisher Discriminant Framework for Feature Extraction and Recognition.* 2005 : IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE.
38. <http://www.aishack.in/2010/07/implementing-sift-in-opencv/>. [Online]
39. **D. Hoiem, A. A. Efros, and M. Hebert.** *Putting objects in perspective.* s.l. : IJCV, 2008.
40. **G., Lowe D.** *Distinctive Image Features from Scale-Invariant Keypoints.* s.l. : International Journal of Computer Vision, 2004.
41. **Lindeberg T., and Romeny T. H.** *Linear scale-space in Geometry-Driven Diffusion.* s.l. : Kluwer Academic Publishers, 1194.
42. **C, Mikolajczyk K. and Schmid.** *A performance evaluation of local descriptors.* s.l. : IEEE Transactions on Pattern Analysis & Machine Intelligence, 2005.