ΤΑΞΙΝΟΜΗΣΗ ΕΙΚΟΝΩΝ ΜΕ BAG OF WORDS LBP DESCRIPTORS

Η ΜΕΤΑΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ ΕΞΕΙΔΙΚΕΥΣΗΣ

υποβάλλεται στην

ορισθείσα από την Γενική Συνέλευση Ειδικής Σύνθεσης
του τμήματος Μηχανικών Η/Υ και Πληροφορικής
Εξεταστική Επιτροπή

από την

Χαρίση Ίνα

ως μέρος των Υποχρεώσεων

για τη λήψη
του

ΜΕΤΑΠΤΥΧΙΑΚΟΥ ΔΙΠΛΩΜΑΤΟΣ ΣΤΗΝ ΠΛΗΡΟΦΟΡΙΚΗ

ΜΕ ΕΞΕΙΔΙΚΕΥΣΗ ΣΤΙΣ ΤΕΧΝΟΛΟΓΙΕΣ-ΕΦΑΡΜΟΓΕΣ

Ιούλιος,2015

ΑΦΙΕΡΩΣΗ

Αφιερωμένο στην οικογένεια μου, στον Αντώνη και και σε όλους τους φίλους μου , που με στήριξαν πολύ σε όλη αυτή τη διαδικασία.

ΕΥΧΑΡΙΣΤΙΕΣ

Ευχαριστώ θερμά τον Καθηγητή μου Κ. Κωνσταντίνο Μπλέκα για την αμέριστη υποστήριξη και συμπαράσταση του.

ΠΕΡΙΕΧΟΜΕΝΑ

ΕΥΡΕΤΗΡΙΟ ΣΧΗΜΑΤΩΝ

ΕΥΡΕΤΗΡΙΟ ΠΙΝΑΚΩΝ

ΠΕΡΙΛΗΨΗ

Χαρίση Ίνα του Σπύρου και της Καλλιόπης.
MSc, Τμήμα Μηχανικών Η/Υ και Πληροφορικής, Πανεπιστήμιο Ιωαννίνων, Ιούλιος, 2015.
Ταξινόμηση εικόνων με Bag of Words LBP Descriptors
Επιβλέποντας: Κωνσταντίνος Μπλέκας.

Η εργασία αυτή διερευνά μια νέα μέθοδο ταξηνόμησης και επεξεργασίας εικόνων που βασίζεται σε ένα LBP πρότυπο με χρήση του μοντέλου Bag of Words για να βελτιώσει την απόδοση ταξινόμησης διάφορων κατηγόριων αντικείμενων εικόνων.

Αρχικά όλες οι εικόνες φιλτράρονται από πρότυπα Lbp,όπως θα δούμε παρακάτω. Έτσι αντί να χρησιμοποιούμε τις κανονικές εικόνες, γίνεται χρήση των Lbp εικόνων. Τα Lbp είναι μέθοδοι ανάλυσης υφής, αποτυπώνουν και κωδικοποιούν τοπικά χαρακτηριστικά των εικόνων, τη γειτονιά, δηλαδή, γύρω από κάθε εικονοστοιχείο. Η κατανομή τους χαρακτηρίζει τις εικόνες. Το γεγονός ότι οι διαβαθμίσεις του γκρι στις περισσότερες εικόνες περιέχουν κάποιο είδος θορύβου τις καθιστά χρήσιμες, καθώς παραμένουν αναλλοίωτες οι αποχρώσεις του γκρι. Υπαρχουν πολλές παραλλαγές τις βασικής μεθόδου LBP.

Έπειτα κατασκευάζεται ο Bag of Word Descriptor,οπού συγκεντρώνονται κάποιες περιοχές των φιλτραρισμένων εικόνων από LBP μεθόδους, που έχουν εξαχθεί στην αρχική φάση. Ακολουθεί Δειγματοληψία, δηλαδή επιλέγονται κάποιες περιοχές για κάθε LBP εικόνα. Οι περιοχές αυτές θα πρέπει να ομαδοποιηθούν με βάση τα κοινά χαρακτηριστικά του, αυτό γίνεται με χρήση του Αλγορίθμου K-Means. Το στάδιο αυτό είναι το στάδιο της εξαγωγής χαρακτηριστικών. Άρα ο BoWL Descriptor είναι μια δομή με σημαντικά χαρακτηριστικά για τις εικόνες

που θα ταξινομηθούν. Τα νέα αυτά χαρακτηριστικά αποτελούν κωδικο-λέξεις με σημαντική πληροφορία και Bow Descriptor είναι στην ουσία ένα λεξικό.

Σε τρίτη φάση, έχουμε την αναπαράσταση των εικόνων ως προς τα χαρακτηριστικά στο λεξικό, για κάθε εικονοστοιχείο (Quantization). Κάθε εικόνα θα πρέπει να έχει τα εικονοστοιχεία της χαρακτηρισμένα ως προς κάποια λέξη. Αναγράφονται νέες τιμές σε κάθε εικόνα, για κάθε pixel, προκύπτουν εκ νέου άλλες εικόνες. Ακολουθεί η εξαγωγή ιστογραμμάτων για κάθε εικόνα. Τέλος αυτό που απομένει να γίνει είναι η ταξινόμηση στα ιστογράμματα που έχουν προκύψει. Γίνεται χρήση κάποιων γνωστών αλγορίθμων ταξινόμησης όπως Knn,Naive Bayes κτλ.

Για κάθε λειτουργιά LBP και αλγόριθμο ταξινόμησης έχει γίνει μια σειρά από πειράματα με διάφορες αλλαγές εσωτερικών χαρακτηριστικών, των αλγορίθμων. Εδώ θα κριθεί η αποτελεσματικότητα και αν ο στόχος της προτεινόμενης μεθοδολογίας, επιτευχθεί. Προβλήματα που προέκυψαν και σημεία για το πως θα μπορούσαμε να έχουμε καλύτερη απόδοση.

EXTENDED ABSTRACT IN ENGLISH

Charisi, Ina,S. MSc, Computer

Science Department, University of Ioannina, Greece, July, 2015

Image Classification with Bag of Words LBP Descriptors

Supervisor:Konstantinos Blekas.

This work investigates a new method of image classification and processing, that is based on Local Binary Patterns (LBP) using Bag of Words (BoW) method to improve the efficiency of the classification of various categories of image items.

Initially all images are filtered by LBP filters. This way, LBP images are used instead of normal images. The Lbp are texture analysis methods, which capture and encode local features of images, neighborhood, i.e., around each pixel. Their distribution characterizes images. The fact that the grayscale in most images contain a noise type makes them useful as they remain unchanged to shades of gray. There are many variations of the basic LBP method.

In a second stage The Bag of Word Descriptor (BoW) is constructed, where some areas of the filtered pictures that were extracted, in initial phase are being concentrated. As a result BoWL Descriptor is a structure with features of high significance for the pictures that are to be classified. Sampling is following , some areas for each LBP image are selected. These areas should be grouped based on common characteristics, it is done by Algorithm K-Means. This step is the step of extracting features. So the BoWL Descriptor is a structure with important features for the images. These new features can called code-words with significant information and Bow Descriptor is essentially a dictionary.

In a thirred phase, representation of the images according to features in the BoWL Lexicon or Descriptor, for each pixel (Quantization) is to be done. Each pixel of an image, must labeled with a word from the lexicon. New values in each image, for each pixel, are entered, resulting again to different images, so the histograms for each image, are exported. Finally, what remains to be done is to classify the histograms which have arisen. Known sorting algorithms as Knn, Naive Bayes etc are used.

For each LBP function and classification algorithm series of experiments have been done. Here the and if the aim of the proposed methodology, achieved,are judged. Problems and solution for methodology improvement are discussed.

ΚΕΦΑΛΑΙΟ 1

ΕΙΣΑΓΩΓΗ

1.1 Ταξινόμηση εικόνων

Η ταξινόμηση εικόνων, η αναζήτηση και η ανάκτηση είναι μια ταχέως αναπτυσσόμενη περιοχή έρευνας. Ο μεγάλος όγκος των ψηφιακών εικόνων που λαμβάνονται κάθε χρόνο σε όλο τον κόσμο, απαιτεί την ανάπτυξη ενός αυτοματοποιημένου συστήματος ταξινόμησης. Εκτός από την ταξινόμηση μεγάλου όγκου εικόνων που δεν έχουν κατηγοριοποιηθεί, η αναγνώριση εικόνας έχει μια μεγάλη ποικιλία χρήσεων , όπως στην πρόβλεψη του καιρού,σε ιατρικές διαγνώσεις και ρομποτική όραση.

Επίσης θα θέλαμε ένα σύστημα ταξινομητών που μπορεί να χειριστεί όλες τις κλίμακες και τα σχήματα των δεδομένων που προκύπτουν από μεγάλα διεθνή καταστήματα και πειραματικά κέντρα.

Η ευκολία με την οποία αντιλαμβανόμαστε τα αντικείμενα γύρω μας είναι ουσιαστικά διαδικασίες αναγνώρισης προτύπων. Οι άνθρωποι εκπαιδεύουν τα κύτταρα του εγκεφάλου τους ώστε να μπορούν να επιλύουν προβλήματα τέτοιου είδους προβλήματα. Στην περιοχή της ψηφιακής επεξεργασίας εικόνων το πρόβλημα της αναγνώρισης προτύπων ανάγεται στην ταξινόμηση ενός συνόλου κατάλληλων χαρακτηριστικών σε κλάσεις με τη βοήθεια κατάλληλων ταξινομητών. Το σύνολο των χαρακτηριστικών που χρησιμοποιείται σε ένα σύστημα ταξινόμησης ονομάζεται διάνυσμα χαρακτηριστικών ή πρότυπο. Οι ταξινομητές είναι ουσιαστικά αλγόριθμοι που χρησιμοποιούνται για να επιτύχουμε την ταξινόμηση των προτύπων ομοιογενείς από άποψη χαρακτηριστικών (όσο είναι αυτό δυνατό) κλάσεις.

Εάν κάποιος προσπαθήσει να ταξινομήσει τα χαρακτηριστικά μιας εικόνας, θα χρησιμοποιήσει κάποια στοιχεία οπτικής ερμηνείας για τον εντοπισμό ομοιογενών ομάδων pixels που αντιπροσωπεύουν διάφορα χαρακτηριστικά ή κατηγορίες ενδιαφέροντος.

Γενικά, ως "χαρακτηριστικό" μπορεί να θεωρηθεί οποιοδήποτε μετρήσιμο μέγεθος που εξάγεται από μια εικόνα. Τα χαρακτηριστικά χρησιμοποιούνται για να περιγράψουν αντικείμενα που περιέχονται στις εικόνες και πρέπει να επιλέγονται κατάλληλα και ανάλογα με τις ιδιαιτερότητες της κάθε εφαρμογής. Ουσιαστικά, η περιγραφή των αντικειμένων με χαρακτηριστικά απεικονίζει τα αντικείμενα στο χώρο των χαρακτηριστικών, με αποτέλεσμα,η αναγνώρισή τους να ισοδυναμεί με τη μέτρηση της ομοιότητας μεταξύ των χαρακτηριστικών των αντικειμένων. Τα χαρακτηριστικά, γενικά, μπορούν να κατηγοριοποιηθούν ως χαρακτηριστικά χώρου και χαρακτηριστικά από μετασχηματισμό. Τα χαρακτηριστικά χώρου μπορούμε να τα διακρίνουμε σε γεωμετρικά χαρακτηριστικά, στατιστικά χαρακτηριστικά και χαρακτηριστικά υφής. Τα γεωμετρικά χαρακτηριστικά περιγράφουν την υφή με καλά ορισμένα θεμελιώδη στοιχεία (primitives) και μια ιεραρχία της χωρικής κατανομής αυτών των στοιχείων. Η επιλογή ενός θεμελιώδους στοιχείου και η πιθανότητα το επιλεγμένο στοι-

χείο να τοποθετηθεί σε μια συγκεκριμένη θέση μπορεί να είναι συνάρτηση της θέσης ή των στοιχείων κοντά στη θέση. Το πλεονέκτημα της δομικής προσέγγισης είναι ότι παρέχει καλή συμβολική περιγραφή της εικόνας. Ωστόσο, το χαρακτηριστικό αυτό είναι πιο χρήσιμο σε εφαρμογές σύνθεσης παρά ανάλυσης. Τα στατιστικά χαρακτηριστικά, σε αντίθεση με τα γεωμετρικά, περιγράφουν την υφή μέσα από τις στατιστικές ιδιότητες της κατανομής των φωτεινοτήτων των εικονοστοιχείων μέσα στην εικόνα. Τέλος τα χαρακτηριστικά υφής προσπαθούν να περιγράψουν την υφή από επαναλαμβανόμενες δομές που οφείλονται σε διακυμάνσεις των φωτεινοτήτων της εικόνας, τα οποία για τη συγκεκριμένη ευκρίνεια της εικόνας, είναι τόσο λεπτά που δεν μπορούν να διακριθούν ως ξεχωριστά αντικείμενα.

Ένας τυπικός ορισμός της ταξινόμησης εικόνας σε κατηγορίες με βάση το περιεχόμενο δίνεται παρακάτω. Έστω $I = \{I_1, I_2, I_3, ...I_N\}$ ένα σύνολο από N εικόνες και έστω $C = \{C_1, C_2, C_3, ...C_M\}$ ένα σύνολο από M προκαθορισμένες κατηγορίες. Στη διαδικασία της ταξινόμησης σε κάθε ζεύγος (I_i, C_i) IxC, ανατίθεται μια λογική τιμή (True/False). Η ανάθεση της τιμής True (Αληθής) σημαίνει ότι η εικόνα I_i ανήκει στην κατηγορία C_i, ενώ η ανάθεση της τιμής False(Ψευδής) σημαίνει ότι η εικόνα I_i δεν ανήκει στην κατηγορία C_i. Ουσιαστικά είναι σαν να έχουμε μία συνάρτηση απεικόνισης $\Phi : \{I * C\} \rightarrow \{T, F\}$ η οποία ονομάζεται ταξινομητής (classifier),για την προσέγγιση μιας άγνωστης συνάρτησης στόχου (target function) $\{\Phi\prime\} : \{I * C\} \rightarrow \{T, F\}$, η οποία περιγράφει τη σωστή ταξινόμηση για τις εικόνες. Φυσικά είναι επιθυμητό η συνάρτηση στόχος και ο ταξινομητής να συμπίπτουν όσο το δυνατόν περισσότερο.

Για παράδειγμα, αν το σύνολο εικόνων $I = \{Im_1, Im_2, Im_3, Im_4, Im_5, Im_6\}$, που φαίνονται στο σχήμα 1.1 πρέπει να ταξινομηθεί στο σύνολο κατηγοριών. $C = \{aeroplane, car, bicycle\}$.

ΣΧΗΜΑ 1.1: Δεδομένα για Ταξινόμηση

1.2 Υπολογιστική Όραση

Η τεχνητή όραση ως επιστημονικό πεδίο ,ξεκίνησε τη δράση της κατά τη δεκαετία του 70 και από τότε συνεχώς εξελίσσετε. Επικεντρώνοντας το ενδιαφέρον μας στην τελευταία δεκαετία, μπορούμε να πούμε ότι η τεχνητή όραση συνέχισε να εμβαθύνει στην αλληλεπίδραση μεταξύ της όρασης και των πεδίων της γραφικής.

Μια δεύτερη σημαντική τάση της τελευταίας δεκαετίας, η οποία έχει και άμεση σχέση με την παρούσα εργασία, θεωρείται η εμφάνιση τεχνικών βασισμένες σε χαρακτηριστικά των αντικειμένων στοχεύοντας στην αναγνώριση τους. Τέτοιου είδους τεχνικές κυριαρχούν και σε άλλες εργασίες αναγνώρισης, όπως η αναγνώριση σκηνής, πανοραμική θέαση και αναγνώριση τοποθεσίας.Η χρήση της τεχνητής όρασης είναι συχνό φαινόμενο ιδίως σε υψηλής τεχνολογίας εφαρμογές όπως στην βιομηχανία, την ιατρική ακόμα και στην εξερεύνηση του διαστήματος. Έτσι αναπόφευκτα έχει άμεση σχέση με άλλους τομείς των επιστημών όπως η επεξεργασία σημάτων, φυσική, τα μαθηματικά, η τεχνητή νοημοσύνη και η νεύρο-βιολογία.

1.3 Προβλήματα Ταξινόμησης

Έκτος από το άμεσο πρόβλημα της ταξινόμησης εικόνων σε κατηγορίες, υπάρχουν προβλήματα όπως τα παρακάτω,ανάγονται σε προβλήματα ταξινόμησης εικόνας με βάση το περιεχόμενο. Τέτοιου τύπου προβλήματα είναι τα παρακάτω.

- Ανάκτηση εικόνας με βάση το περιεχόμενο. (Content based image Retrieval). Έχει να κάνει με την ανάκτηση μιας εικόνας από μια βάση δεδομένων με εικόνες. Σε αυτή τη βάση δεδομένων κάθε εικόνα έχει καταχωρηθεί σαν ένα διάνυσμα μεγάλης διάστασης (feature vector), με βάση τα οπτικά της περιεχόμενα (visual contents), δηλαδή το χρώμα, την υφή, το σχήμα των αντικειμένων που απεικονίζονται σε αυτή. Ουσιαστικά, η βάση δεδομένων δεν περιέχει τις εικόνες αποθηκευμένες στη φυσική τους μορφή, αλλά περιέχει κωδικοποιημένες εικόνες, με τη βοήθεια των διανυσμάτων που αντιστοιχούν σε αυτές (feature database). Για την ανάκτηση εικόνας, ο χρήστης παρέχει στο

  σύστημα ανάκτησης μια εικόνα-παράδειγμα, αντιπροσωπευτική της κατηγορίας εικόνων που τον ενδιαφέρει. Για παράδειγμα, αν κάποιος επιθυμεί να ανακτήσει όλες τις εικόνες που απεικονίζουν ποδήλατο,τότε πρέπει να τροφοδοτήσει τη βάση δεδομένων με μια εικόνα - παράδειγμα, που στη συγκεκριμένη περίπτωση απεικονίζει κούπα. Στη συνέχεια, το σύστημα ανάκτησης μετατρέπει την εικόνα-παράδειγμα σε διάνυσμα και υπολογίζει την απόσταση αυτού του διανύσματος από τα διανύσματα των εικόνων που έχουν καταχωρηθεί στη βάση δεδομένων, προκειμένου να υπολογίσει τις ομοιότητες / διαφορές. Από αυτόν τον υπολογισμό προκύπτουν τα αποτελέσματα της ανάκτησης, μεταξύ των οποίων συγκαταλέγονται εικόνες που εμφανίζουν μεγάλη ομοιότητα με την εικόνα-παράδειγμα.

- Εντοπισμός αντικειμένου(Object Detection. Αυτό το πρόβλημα σχετίζεται με τον εντοπισμό σε μια εικόνα ενός αντικειμένου μιας δεδομένης κατηγορίας. Ειδικότερα τα τελευταία χρόνια η ερευνητική προσπάθεια εστιάζεται σε εντοπισμό ανθρώπινου προσώπου (face detection), αυτοκινήτων, πεζών κ.λ.π.

- Αναγνώριση αντικειμένου (Object Recognition). Η αναγνώριση αντικειμένου στην τε-

χνητή όραση είναι η διαδικασία εντοπισμού ενός δεδομένου αντικειμένου σε μια εικόνα ή βίντεο.

1.4 Εφαρμογές στην ταξινόμηση

Η ταξινόμηση εικόνας σε κατηγορία με βάση το περιεχόμενο μπορεί να φανεί χρήσιμη σε διάφορες εφαρμογές, όπως αυτές που περιγράφονται παρακάτω.

- Αναζήτηση εικόνας (Image search). Η αναζήτηση εικόνας είναι η πιο άμεση εφαρμογή, όταν οι άνθρωποι μιλάνε για ταξινόμηση εικόνας σε κατηγορία με βάση το περιεχόμενο. Με αυτήν την έννοια, μπορεί κανείς να σκεφτεί την αναζήτηση εικόνας στη μεγαλύτερη βάση δεδομένων του κόσμου, με την κατασκευή μηχανών αναζήτησης εικόνων στο διαδίκτυο ή απλώς τη δημιουργία εφαρμογών για αναζήτηση εικόνων σε ένα προσωπικό υπολογιστή. Στις μέρες μας, η κακή επίδοση που παρουσιάζουν οι μηχανές αναζήτησης στο διαδίκτυο οφείλεται στο ότι πραγματοποιούν αναζήτηση όχι με βάση το περιεχόμενο της εικόνας, αλλά χρησιμοποιούν πληροφορίες, όπως το όνομα του αρχείου ή τον κώδικα HTML που πλαισιώνει την εικόνα. Ωστόσο, ο φυσικός τρόπος εντοπισμού μιας εικόνας είναι η οπτική αναζήτηση, και αυτός είναι ο στόχος των μεθόδων τεχνητής όρασης.

- Αναζήτηση βίντεο (Video Search). Τα τελευταία χρόνια έχουν παραχθεί πολλές διαφημίσεις και δεδομένα βίντεο. Τα αρχεία βίντεο συνηθίζεται να αποθηκεύονται σε βάσεις δεδομένων με πληροφορίες μεταδεδομένων (metadata). Θα ήταν χρήσιμο να μπορούσε κανείς να ανακτήσει ένα αρχείο βίντεο με βάση το περιεχόμενό του. Επίσης, οι παραγωγοί ή οι σκηνοθέτες ταινιών θα ενδιαφερόταν να ανακτήσουν πλάνα ταινιών στα οποία πρωταγωνιστεί ένας συγκεκριμένος ηθοποιός ή που εξελίσσονται σε ένα συγκεκριμένο τόπο.

- Ιατρικές εφαρμογές (Medical Applications). Στον τομέα της ιατρικής καθημερινά πα-

ράγονται πολλές εικόνες (ακτινογραφίες, μαστογραφίες κ.λ.π). Θα ήταν χρήσιμο για τους γιατρούς να έχουν στη διάθεσή τους εργαλεία για να ταξινομούν τις ιατρικές εικόνες σε κατηγορία (π.χ εικόνες αξονικής τομογραφίας με καρκίνο, εικόνες αξονικής τομογραφίας χωρίς καρκίνο) και να μην τις εξετάζουν ανά περίπτωση, όπως κάνουν.

- Συμπίεση Βίντεο (Video Compression). Λόγω του περιορισμένου εύρου ζώνης σημαντικών καναλιών επικοινωνίας (π.χ ασύρματων, υποθαλάσσιων κ.λ.π) η μετάδοση βίντεο σε αυτά τα κανάλια απαιτεί ουσιαστική συμπίεση των δεδομένων που μεταδίδονται. Μια πολλά υποσχόμενη προσέγγιση θα ήταν η συμπίεση των δεδομένων με βάση το περιεχόμενο της σκηνής, με την εφαρμογή διαφορετικών επιπέδων συμπίεσης ανάλογα με το πόσο σημαντικά είναι τα αντικείμενα της σκηνής. Για παράδειγμα, οι ηθοποιοί που είναι αντικείμενα μεγάλης σημασίας θα διατηρούσαν υψηλή οπτική ποιότητα, ενώ αντικείμενα του φόντου θα δεχόταν υψηλότερη κωδικοποίηση με σκοπό να καταλαμβάνουν λιγότερα bytes. Σε αυτή την εφαρμογή, η τεχνητή όραση θα συνεισέφερε στο να γίνει αρχικά κατάτμηση του πλάνου στα αντικείμενα που το απαρτίζουν και στη συνέχεια ταξινόμησή τους σε κατηγορίες.

- Παρακολούθηση (Surveillance). Τα συστήματα παρακολούθησης δεν έχουν εξελιχθεί ακόμη αρκετά και απαιτείται η παρουσία του ανθρώπινου παράγοντα για να εντοπίσει ύποπτους ανθρώπους και ασυνήθιστα περιστατικά. Τα προχωρημένα συστήματα επιδιώκουν να ανιχνεύσουν αυτόματα τέτοιου είδους γεγονότα, όπως π.χ ένα διαπληκτισμό σε ένα ποδοσφαιρικό αγώνα που διεξάγεται σε στάδιο.

- Εναέριες εικόνες (Aerial images). Κάθε χρόνο δαπανώνται χιλιάδες ευρώ από τις εθνικές υπηρεσίες χαρτογραφίας, προκειμένου να ενημερωθούν οι χάρτες. Η διαδικασία της ανανέωσης είναι συχνά χρονοβόρα και απαιτεί κάποιο άτομο για τη σύγκριση του τρέχοντος χάρτη με την πιο πρόσφατη δορυφορική εικόνα υψηλής ανάλυσης. Τεχνικές για ταξινόμηση εικόνας θα βοηθούσαν στον εντοπισμό των αλλαγών στο τοπίο, χωρίς την ανθρώπινη παρέμβαση.

- Ρομποτική (Robotics). Η παροχή όρασης σε ένα ρομπότ είναι ίσως ένας από τους πιο

φιλόδοξους στόχους στο πεδίο της τεχνητής όρασης. Με αυτόν τον τρόπο ένα εντελώς αυτόνομο ρομπότ θα μπορούσε να χρησιμοποιηθεί για να εντοπίσει συγκεκριμένα αντικείμενα ενδιαφέροντος και να αντικαταστήσει τον ανθρώπινο παράγοντα σε επικίνδυνες καταστάσεις (πυρκαγιά, υποθαλάσσια εξερεύνηση κ.λ.π).

1.5 Δυσκολίες στην ταξινόμηση

Το πρόβλημα της ταξινόμησης εικόνας σε κατηγορία με βάση το περιεχόμενο, το πρόβλημα θεωρείται ένα από τα πιο δύσκολα στο πεδίο της τεχνητής όρασης, γιατί παρουσιάζει ορισμένες ιδιαιτερότητες, που σχετίζονται με μεταβολές στην εμφάνιση αντικειμένων της ίδιας κατηγορίας. Για να είναι αποτελεσματική μια μέθοδος ταξινόμησης θα πρέπει να μπορεί να χειρίζεται ικανοποιητικά τις μεταβολές αυτές. Κάποιες δυσκολίες αναφέρονται παρακάτω.

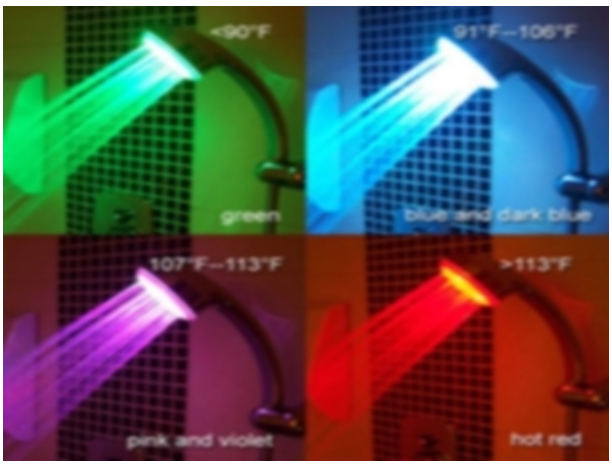
- Μεταβολές που οφείλονται στην οπτική γωνία λήψης. (Camera viewpoint variations). Αφού η εικόνα είναι η προβολή του τρισδιάστατου κόσμου (με τρισδιάστατα αντικείμενα) σε ένα χώρο δύο διαστάσεων, η οπτική γωνία φωτογράφησης του αντικειμένου μπορεί να επηρεάσει αισθητά την εμφάνισή του στην εικόνα, παρόλο που τα χαρακτηριστικά του παραμένουν αμετάβλητα. Στο σχήμα 1.2 το ίδιο αντικείμενο έχει φωτογραφηθεί από διαφορετική οπτική γωνία, ωστόσο και οι τρεις εικόνες θα πρέπει να ταξινομηθούν στην ίδια κατηγορία.

ΣΧΗΜΑ 1.2: Διαφορετικές γωνίες λήψης

- Μεταβολές που σχετίζονται με εσωτερικά χαρακτηριστικά αντικειμένων διαφορετικών κατηγοριών (Interclass variations). Μια σημαντική πρόκληση του προβλήματος είναι ότι είναι πιθανόν, εικόνες από διαφορετικές εικόνες να μοιάζουν τόσο ώστε να συγχέονται. Για παράδειγμα οι εικόνες στο σχήμα 1.3 απεικονίζουν αντίστοιχα τη Marilyn Monroe και άλλες δυο γυναίκες με μεγάλη ομοιότητα στην εξωτερική τους εμφάνιση, παρόλα αυτά ανήκουν σε διαφορετικές κατηγορίες και το σύστημα θα πρέπει να τις κατατάσσει σε 3 κατηγορίες.

ΣΧΗΜΑ 1.3: Διαφορετικά εσωτερικά χαρακτηριστικά

- Συνθήκες φωτισμού (Illumination). Μια σημαντική παράμετρος του προβλήματος είναι οι συνθήκες φωτισμού στις εικόνες. Για παράδειγμα στο σχήμα 1.4 υπάρχουν ομάδες εικόνων, η καθεμιά από το ίδιο τοπίο, το οποίο έχει φωτογραφηθεί κάτω από διαφορετικές συνθήκες. Το σύστημα ταξινόμησης θα πρέπει να αναθέτει αυτές τις εικόνες στην ίδια κατηγορία.

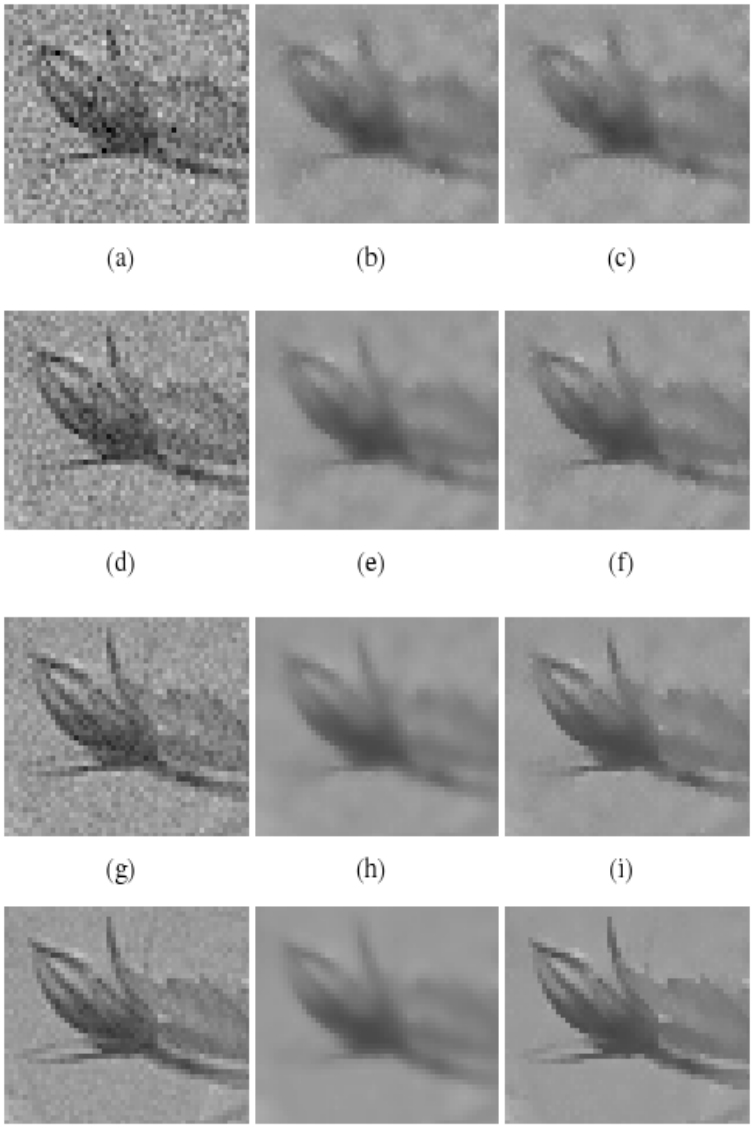


ΣΧΗΜΑ 1.4: Αλλαγή Φωτισμού

- Μεταβολές στην κλίμακα (Scaling). Αυτή είναι μια σημαντική παράμετρος που πρέπει να ληφθεί υπόψη στο πρόβλημα της ταξινόμησης. Μπορούμε να έχουμε μια εικόνα με ένα βουνό μπροστά μας ή εικόνες με ένα βουνό στο φόντο της εικόνας, ωστόσο και οι δύο εικόνες θα πρέπει να ταξινομηθούν στην ίδια κατηγορία. Το σχήμα 1.5 δείχνει εικόνες της ίδιας κούκλας σε διαφορετικές κλίμακες.

ΣΧΗΜΑ 1.5: Παράδειγμα αλλαγής κλίμακας

- Θόρυβος. Είναι συχνό αλλά στο πίσω μέρος της σκηνής να υπάρχουν και άλλα αντικείμενα τα οποία δε θέλουμε να κάνουμε ταξινόμηση πάνω σε αυτά και επιβαρύνουν την "καθαρότητα" της εικόνας, ως προς το αντικείμενο που ταξινομείται(background clutter). Επίσης είναι εξίσου συχνό να παρεμβάλλονται διάφορα αντικείμενα ανάμεσα στο αντικείμενο που θέλουμε να επικεντρωθούμε για ταξινόμηση, επισκιάζοντας ή καλύπτοντας τμήμα του αντικειμένου (occlusion, shadows) ή να έχουμε μη καθαρές εικόνες του. Τα παραπάνω φαινόμενα θεωρούνται,εν γένει,θόρυβος (noise). Στην κατασκευή μιας μεθόδου ταξινόμησης ο θόρυβος είναι μια ανεπιθύμητη κατάσταση, ωστόσο ορισμένες φορές είναι αναπόφευκτος. Στο σχήμα 1.6 βλέπουμε ένα παράδειγμα θορύβου.



ΣΧΗΜΑ 1.6: Εικόνες θορύβου

1.6 Η Δομή της εργασίας

Πιο αναλυτικά, στο Κεφάλαιο 1 αναλύεται η έννοια και η χρησιμότητα της ταξινόμησης εικόνων. Στο Κεφάλαιο 2 παρουσιάζεται το τοπικό δυαδικό πρότυπο Lbp, η λειτουργιά του και η παραλλαγές του. Στο Κεφάλαιο 3 έχουμε τη διαδικασία που ακολουθήσαμε για τον Descriptor BoWL για να καταλήξουμε στην κατασκευή του επιθυμητού λεξικού. και το πως γίνεται η ταξινόμηση και κάποιοι γνωστοί Αλγόριθμοι ταξινόμησης. Στο Κεφάλαιο 4 θα δούμε την πειραματική μελέτη της μεθόδου, ενώ στο 5 εξετάζονται πλεονεκτήματα και μειονεκτήματα και συμπεράσματα.

ΚΕΦΑΛΑΙΟ 2

Ο ΜΕΤΑΣΧΗΜΑΤΙΣΜΟΣ ΕΙΚΟΝΑΣ LBP

2.1 Μέθοδοι LBP

Τα Τοπικά Δυαδικά Πρότυπα ή πιο γνωστά Local Binary Patterns (LBP) αναφέρθηκαν για πρώτη φορά από τον Harwood και στη συνέχεια παρουσιάστηκαν επισήμως ως μέθοδος

ανάλυσης υφής το 1996 από τους Ojala κ.α. Έκτοτε έχουν χρησιμοποιηθεί ευρύτατα σε πολλές εφαρμογές ως ισχυρά και αποτελεσματικά χαρακτηριστικά υφής χάριν της διακριτικής τους ικανότητας και της χαμηλής υπολογιστικής πολυπλοκότητας. Ένας LBP τελεστής είναι απλοϊκός στον υπολογισμό του και πολύ αποδοτικός. Επίσης, ανταποκρίνεται άριστα σε μεθόδους αναγνώρισης προτύπων στο ανθρώπινο οπτικό σύστημα.

Τα LBP ουσιαστικά αποτυπώνουν και κωδικοποιούν τοπικά χαρακτηριστικά των εικόνων, τη γειτονιά, δηλαδή, γύρω από κάθε εικονοστοιχείο, και η κατανομή τους χαρακτηρίζει τις εικόνες. Το γεγονός ότι οι διαβαθμίσεις του γκρι στις περισσότερες εικόνες περιέχουν κάποιο είδος θορύβου τις καθιστά χρήσιμες για απόκρυψη πληροφορίας χωρίς να γίνεται αντιληπτή. Για το λόγο αυτό μπορούν να χρησιμοποιηθούν κατηγοριοποιητές υφής (texture classifiers). Ο LBP ορίζεται ως ένα μέτρο σύγκρισης υφής που αφήνει αναλλοίωτες τις αποχρώσεις του γκρι και οι οποίες προέρχονται από την εξέταση της υφής σε μία "τοπική" γειτονιά .Ο κύριος λόγος για τις ιδιότητες των LBP χαρακτηριστικών είναι η ανοχή στη μονοτονική αλλαγή φωτεινότητας και η υπολογιστική απλότητα.

Το LBP είναι μια μη παραμετρική μέθοδος, όπου συνοψίζει την τοπική δομή μιας εικόνας αποδοτικά, έχει δοθεί όλο και συνεχώς αυξανόμενο ενδιαφέρον στην εφαρμογή του για ταυτοποίηση προσώπων, τα τελευταία χρονιά, αν και αρχικά είχε προταθεί για αναγνώριση υφών. Έχει απήχηση σε πολλές εφαρμογές, όπως ανάκτηση εικόνας/βίντεο, ανάλυση εναερίων εικόνων και στην οπτική επίβλεψη.

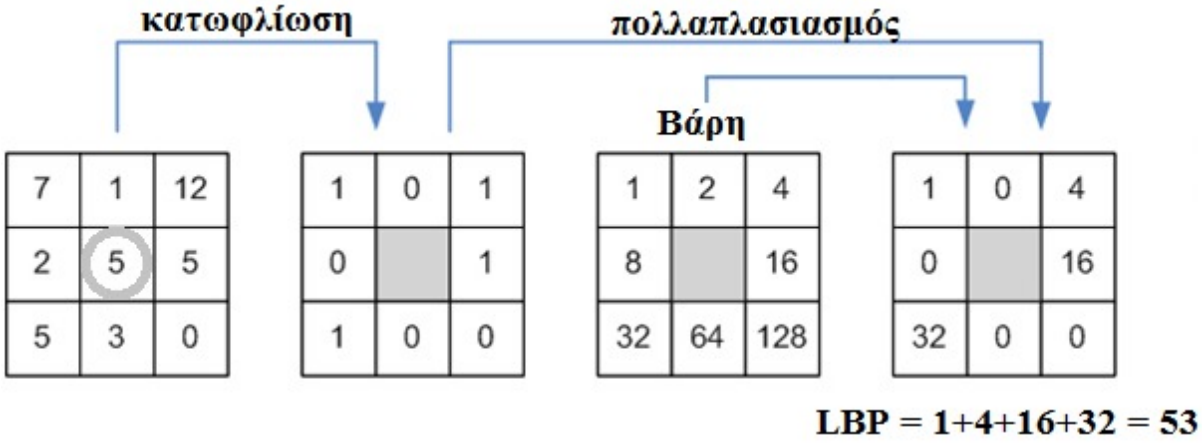
Τα τελευταία χρονιά τα LBP χρησιμοποιήθηκαν ευρέως για την επεξεργασία, αναγνώριση , ανίχνευση προσώπων, ανίχνευση έκφρασης προσώπου και για την ταξινόμηση ηλικίας /φύλλου. Για την αναγνώριση προσώπων χρησιμοποιούνται να φίλτρα Gabor,πολύ ακριβά υπολογιστικά, για την εξαγωγή συντελεστών σε πολλαπλές κλίμακες και προσανατολισμούς, κάτι που δεν απαιτείται με τα LBP.

Η επιλογή των πιο κρίσιμων LBP Χαρακτηριστικών είναι ένα επιπρόσθετο πρόβλημα προς επίλυση

Αρχικά θα εξετάσουμε το βασικό τελεστή LBP, υπάρχουν αρκετές παραλλαγές LBP για καλύτερη αποτελεσματικότητα . Παρακάτω θα δούμε αναλυτικά κάποιες από τις παραλλαγές που χρησιμοποιήθηκαν.

2.2 LBP 3x3

Ο αρχικός τελεστής LBP βάζει τιμές στα pixels μιας εικόνας μεσώ της κατωφλίωσης μιας 3x3 γειτονιάς pixels του κάθε pixel με την τιμή του κεντρικού pixel και θεωρεί το αποτέλεσμα ως ένα δυαδικό αριθμό, από τον οποίο προκύπτει ο αντίστοιχος δεκαδικός αριθμός που χρησιμοποιείται για την επισήμανση , όπως βλέπουμε στο σχήμα 2.1 . για ένα συγκεκριμένο παράδειγμα. Οι προερχόμενοι δυαδικοί αριθμοί Τοπικά Δυαδικά Μοτίβα (LBP) ή LBP κωδικοί.



ΣΧΗΜΑ 2.1: Παράδειγμά του απλού τελεστή LBP

Επομένως για ένα pixel (x_c, y_c), το προκύπτον LBP μπορεί να είναι εκφράζεται σε δεκαδική μορφή ως :

$$LBP(x_c, y_c) = \sum_{n=0}^{7} s(i_n - i_c) * 2^n \tag{2.1}$$

όπου το n δεικτοδοτεί τους γείτονες του κεντρικού pixel, ic και in είναι οι τιμές γκρι επιπέδου

του κεντρικού pixel και του γειτονικού αντίστοιχα. Η συνάρτηση s(x) ορίζεται ως:

$$s(x) = \begin{cases} 1 & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases} \tag{2.2}$$

Το ιστόγραμμα των LBP τιμών που υπολογίζονται για μια περιοχή ή εικόνα μπορεί να αξιοποιηθεί για την περιγραφή υφών.

Ο περιορισμός του βασικού τελεστή LBP είναι ότι οι μικρές 3x3 γειτονίες δεν μπορούν να εξάγουν τα κυρίαρχα χαρακτηριστικά σε δομές με μεγάλες κλίμακες. Ως αποτέλεσμα, για την αντιμετώπιση υφών σε διαφορετικές κλίμακες, ο τελεστής επεκτάθηκε σε γειτονιές με διαφορετικά μεγέθη και άλλους τρόπους.

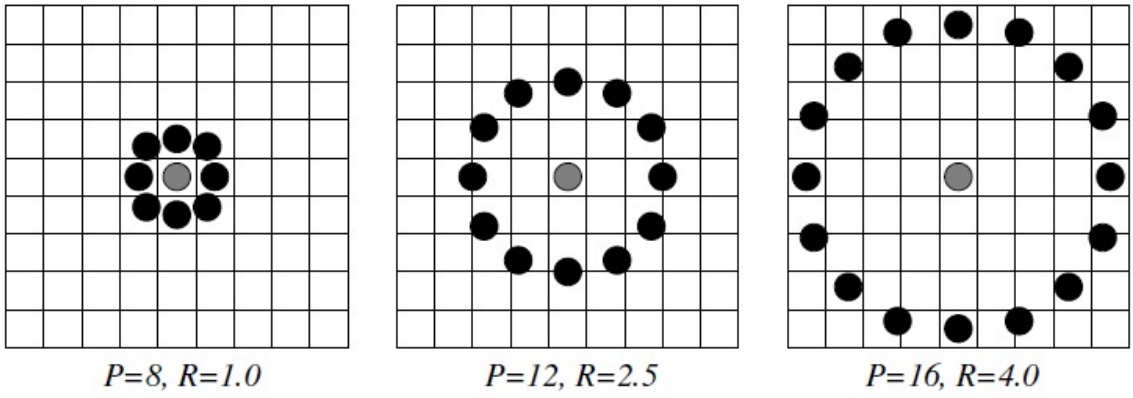
2.3 Circular LBP

Το 2002 οι Ojala κ.α. πρότειναν μια γενίκευση των LBP ώστε να αποτελούν έναν απλό αλλά αποτελεσματικό τρόπο πολυδιακριτικής ανάλυσης. Σύμφωνα με τη γενίκευση αυτή, ο καθορισμός των τιμών των LBP δε γίνεται σε μια γειτονιά 3x3 αλλά πάνω στην περιφέρεια ενός κύκλου (P,R). R είναι η ακτίνα του κύκλου από το κεντρικό εικονοστοιχείο και P είναι το πλήθος των γειτονικών σημείων που κατανέμονται ομοιόμορφα στην περιφέρεια του κύκλου. Τρία χαρακτηριστικά παραδείγματα από γειτονιές διαφορετικής ακτίνας και πλήθους γειτόνων δίνονται στην σχήμα 2.2 όπου οι συντεταγμένες των γειτονικών σημείων (x_p, y_p) προκύπτουν από τις σχέσεις

$$x_p = x_c + R * \cos(\frac{2\pi p}{P}) \tag{2.3}$$

$$y_p = y_c + R * \cos(\frac{2\pi p}{P}) \tag{2.4}$$

όπου (x_c, y_c) είναι οι συντεταγμένες του κεντρικού εικονοστοιχείου. Όταν οι συντεταγμένες των γειτόνων δε συμπίπτουν ακριβώς με το κέντρο ενός εικονοστοιχείου τότε οι τιμές των



ΣΧΗΜΑ 2.2: Κυκλικές συμμετρικές γειτονιές για διάφορες τιμές P και R

φωτεινοτήτων προκύπτουν από παρεμβολή.

Ο υπολογισμός των νέων τιμών των τοπικών χαρακτηριστικών LBP μπορεί να εκφραστεί από την παρακάτω σχέση.

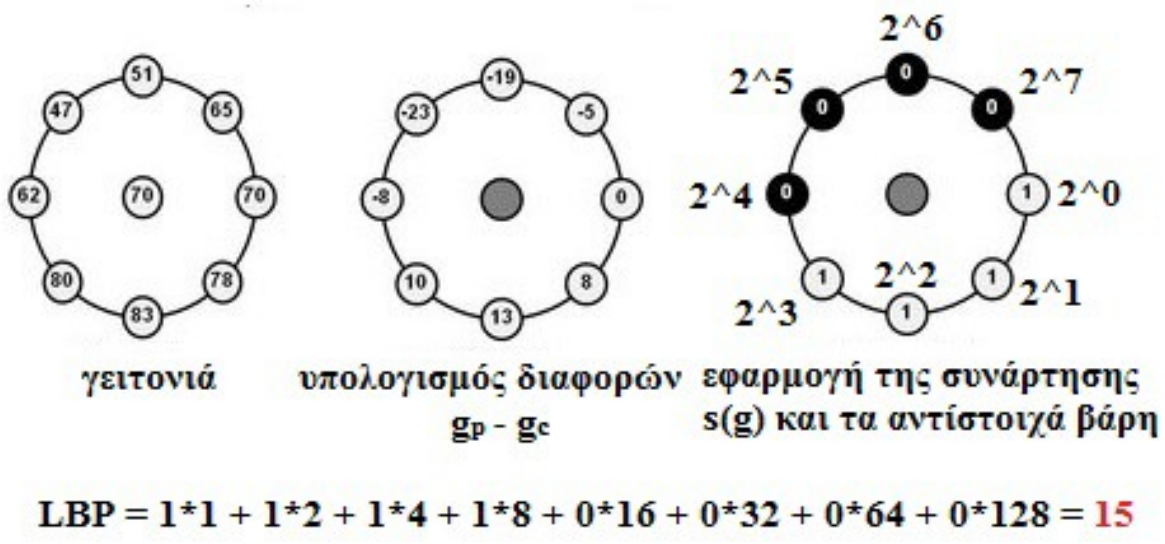
$$LBP_{P,R} = \sum_{p=0}^{P-1} s(g_p - g_c) * 2^p \tag{2.5}$$

όπου g_c είναι η φωτεινότητα του κεντρικού εικονοστοιχείου, g_p είναι οι φωτεινότητες των γειτονικών εικονοστοιχείων που κατανέμονται συμμετρικά στην περιφέρεια του κύκλου και συνήθως το p=0 αντιστοιχεί στο εικονοστοιχείο που βρίσκεται δεξιά από το κεντρικό, και

$$s(g) = \begin{cases} 1 & \text{if } g \geq 0 \\ 0 & \text{if } g < 0 \end{cases} \tag{2.6}$$

η οποία εκφράζει την κατωφλίωση όπως περιγράφηκε στην προηγούμενη παράγραφο. Στο σχήμα 2.3 φαίνεται ο υπολογισμός του $LBP_{P,R}$ για γειτονιά (8,R).

Είναι φανερό ότι για συγκεκριμένα P και R εξάγονται 2^P διαφορετικές τιμές χαρακτηριστικών $LBP_{P,R}$ που αντιστοιχούν στους 2^P διαφορετικούς δυαδικούς αριθμούς που προκύπτουν. Με άλλα λόγια, ένα σύνολο από 2^P διαφορετικά πρότυπα μπορεί να προκύψει. Υπογραμμίζεται ότι ο υπολογισμός των $LBP_{P,R}$ γίνεται μόνο για εικονοστοιχεία που απέχουν R ή περισσότερο από τα άκρα της εικόνας. Τα υπόλοιπα εικονοστοιχεία που δεν μπορούν να έχουν ένα πλήρες κύκλο R γύρω τους αμελούνται. Να σημειωθεί ότι το $LBP_{8,1}$ είναι



ΣΧΗΜΑ 2.3: Υπολογισμός $LBP_{8,R}$

παρόμοιο με το αρχικό LBP που προκύπτει από γειτονιά 3x3 με δύο όμως διαφορές: τα εικονοστοιχεία στο $LBP_{8,1}$ σχηματίζουν μια κυκλική αλυσίδα και οι τιμές των φωτεινοτήτων των διαγώνιων εικονοστοιχείων υπολογίζονται με παρεμβολή. Και οι δυο αυτές αλλαγές είναι απαραίτητες ώστε να έχουμε κυκλικά συμμετρική γειτονιά που δύναται να οδηγήσει στην εξαγωγή LBP χαρακτηριστικών ανεξάρτητων περιστροφής τα οποία θα περιγραφούν στην επόμενη παράγραφο.

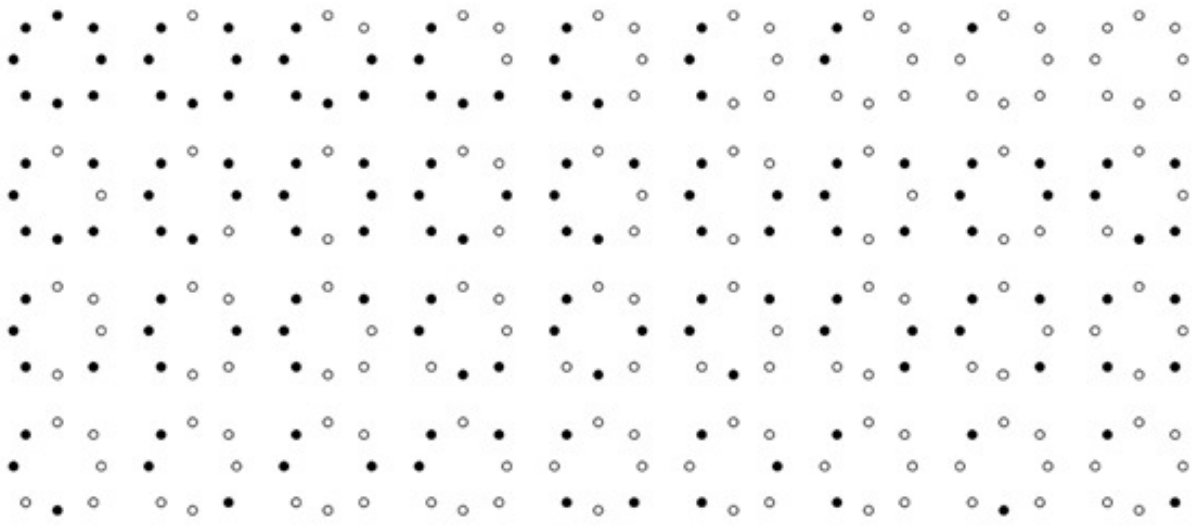
Uniform LBP

Ένα σημαντικό μειονέκτημα των LBP όπως παρουσιάστηκαν μέχρι στιγμής είναι η μη ανθεκτικότητά τους στην περιστροφή της εικόνας ή στην σειρά με την οποία ορίζονται τα βάρη. Με άλλα λόγια, εάν θεωρήσουμε σταθερή την αρχική θέση από τα βάρη (π.χ. p=0 πάντα στο εικονοστοιχείο δεξιά του κεντρικού) τότε όταν η εικόνα περιστραφεί, οι τιμές της έντασης g_p θα μετακινηθούν αντίστοιχα κατά μήκος του κύκλου γύρω από το g_0 . Αφού το g_0 περιέχει πάντα την τιμή της έντασης του στοιχείου $(0, R)$ στα δεξιά του g_c , η περιστροφή ενός συγκεκριμένου δυαδικού προτύπου θα οδηγήσει σε διαφορετική τιμή $LBP_{P,R}$. Το ίδιο ισχύει αντίστοιχα όταν θεωρήσουμε σταθερή την εικόνα και αλλάξουμε τη σειρά από τα βάρη (π.χ. p=0 να αντιστοιχεί στο εικονοστοιχείο πάνω από το κεντρικό). Φυσικά, αυτό το μειονέκτημα

δεν ισχύει για πρότυπα που αποτελούνται αποκλειστικά από 1 ή αποκλειστικά από 0, αφού όπως και να περιστραφούν το τελικό αποτέλεσμα θα είναι το ίδιο. Για να απαλειφθεί το πρόβλημα που προκύπτει σε περίπτωση περιστροφής της εικόνας ή διαφορετικής σειράς των βαρών και να δοθεί μια μοναδική ταυτότητα σε κάθε δυαδικό πρότυπο (ανεξάρτητα της κατεύθυνσης) ορίζονται τα LBP ανεξάρτητα περιστροφής ως εξής:

$$LBP^{ri}_{P,R} = min\{ROR(LBP_{P,R}, i)|i = 0, 1, 2, ..., P - 1\} \tag{2.7}$$

όπου η συνάρτηση $ROR(x, i)$ πραγματοποιεί κυκλική ολίσθηση του δυαδικού αριθμού x (P ψηφίων) προς τα δεξιά i φορές. Με λίγα λόγια, ο κώδικας για ανεξαρτησία περιστροφής παράγεται από ολίσθηση του δυαδικού αριθμού που παράγεται από τα κλασικά LBP μέχρι την εύρεση της ελάχιστης τιμής του. Αυτό σημαίνει ότι όλες οι 2^P διαφορετικές τιμές ενός συγκεκριμένου μοτίβου αντιστοιχίζονται σε μια τιμή αναφοράς, την ελάχιστη. Με τον τρόπο αυτό ομαδοποιούνται όλες οι «περιστρεμμένες» εκδοχές ενός δυαδικού αριθμού σε μία, και από σχηματικής σκοπιάς ομαδοποιούνται πολλά πρότυπα σε λιγότερα πρότυπα ανεξάρτητα της περιστροφής, που αντιστοιχούν σε συγκεκριμένα μικρο-χαρακτηριστικά της εικόνας. Σε



ΣΧΗΜΑ 2.4: Τα 36 μοναδικά ανεξάρτητα περιστροφής πρότυπα σε γειτονιά (8,R). Οι μαύρες βούλες αντιστοιχούν σε 0 και οι άσπρες σε 1

μια γειτονιά (8, R) τα 2^8 = 256 διαφορετικά πρότυπα που προκύπτουν από τον απλό τελεστή

LBP, μειώνονται σε 36 μοναδικά ανεξάρτητα περιστροφής πρότυπα με τη χρήση του τελεστή $LBP^{ri}_{8,R}$ (σχήμα 2.4).

Πολύ ενθαρρυντικά τα αποτελέσματα από τα ανεξάρτητα περιστροφής LBP καθώς μειώνουν σημαντικά το πλήθος των διαφορετικών προτύπων άρα και το μέγεθος του ιστογράμματος. Παρόλα αυτά, στην πράξη διαπιστώθηκε ότι απαιτείται περαιτέρω μείωση του χαρακτηριστικού διανύσματος χωρίς όμως να μειώσουμε την αναλυτικότητα (αριθμός των γειτόνων). Έτσι λοιπόν παρατηρήθηκε ότι συγκεκριμένα τοπικά δυαδικά πρότυπα αποτελούν θεμελιώδεις ιδιότητες της υφής και καταλαμβάνουν τη συντριπτική πλειοψηφία των προτύπων που εμφανίζονται. Αυτά τα θεμελιώδη πρότυπα ονομάστηκαν "ομοιόμορφα" γιατί έχουν ως κοινό στοιχείο μια ομοιόμορφη κυκλική δομή. Ως ομοιόμορφη ορίζεται η δομή η οποία περιέχει λίγες μεταβάσεις από 0 σε 1 και αντίστροφα.

Επομένως, η έννοια της ομοιομορφίας στην περίπτωση των δυαδικών προτύπων U ορίζεται ως το πλήθος των μεταβάσεων από 0 σε 1 και από 1 σε 0 . Για παράδειγμα, το πρότυπο 11111111_2 έχει τιμή U=0 ενώ το πρότυπο 11111110_2 έχει U=2 (δεν πρέπει να ξεχνάμε ότι τα πρότυπα είναι κυκλικά, ανεξάρτητα περιστροφής, δηλαδή το 11111110_2 είναι το ίδιο με το 11101111_2). Γενικά, για να θεωρείται ένα πρότυπο ομοιόμορφο πρέπει να έχει τιμή U μικρότερη ή ίση με 2. Ο μαθηματικός ορισμός της ομοιομορφίας σε μια γειτονιά g δίνεται από τον παρακάτω τύπο.

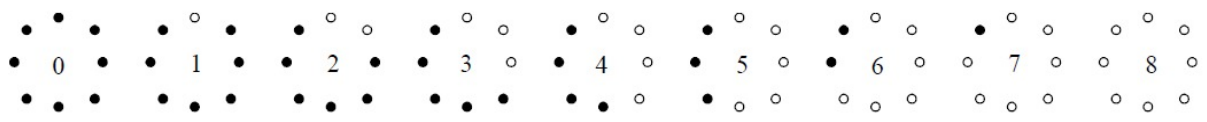
$$U(G) = |s(g_{P-1} - g_c) - s(g_0 - g_c)| + \sum_{p=1}^{P-1} |s(g_p - g_c) - s(g_{p-1} - g_c)| \tag{2.8}$$

Τα ομοιόμορφα ανεξάρτητα περιστροφής $LBP(LBP_P, R)$ ορίζονται ως εξής:

$$LBP^{riu2}_{P,R}(x) = \begin{cases} \sum_{p=0}^{P-1} s(g_p - g_c) & \text{εάν } U(LBP_{P,R}) \geq 2 \\ P+1 & \text{διαφορετικά} \end{cases} \tag{2.9}$$

Επομένως, το σύνολο των ομοιόμορφων ανεξάρτητων περιστροφής προτύπων που μπορεί να εμφανιστούν σε μια γειτονιά (P,R) είναι P+1. Στο σχήμα 2.5 παρουσιάζονται τα ομοιόμορφα

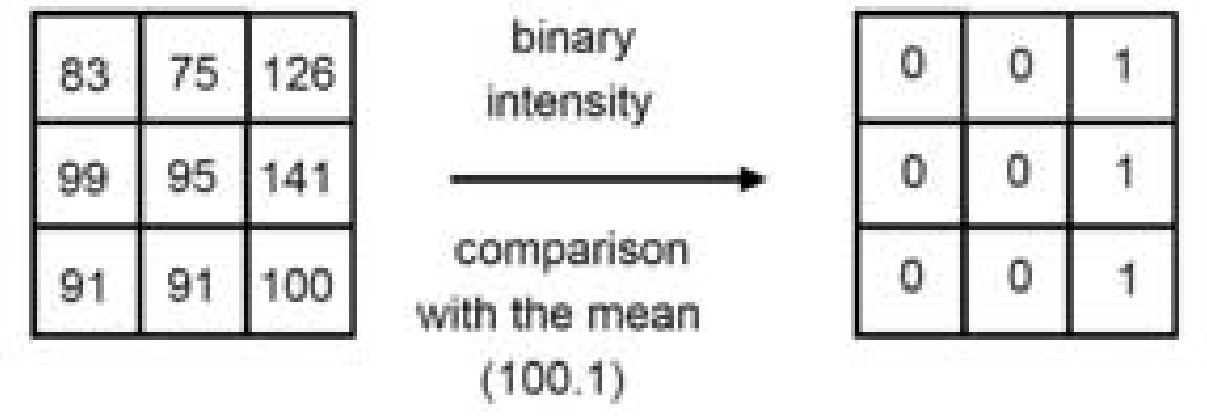
ανεξάρτητα περιστροφής πρότυπα μιας γειτονιάς (8,R) και οι αντίστοιχες τιμές τους.



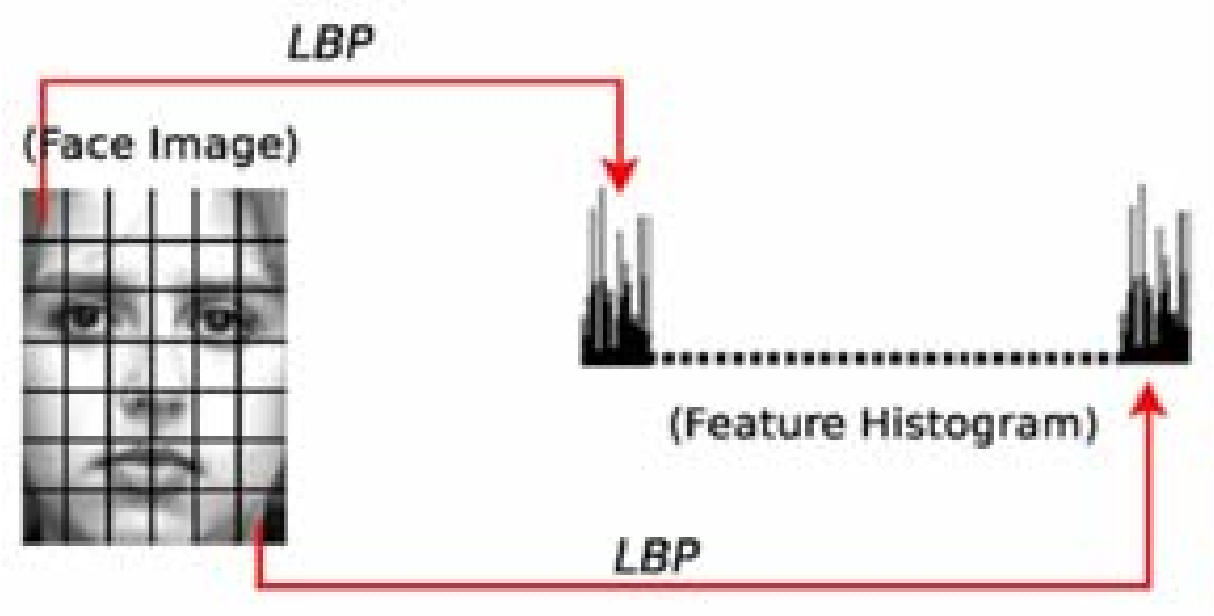
ΣΧΗΜΑ 2.5: Ομοιόμορφα ανεξάρτητα περιστροφής πρότυπα μιας γειτονιάς (8,R). Οι μαύρες βούλες αντιστοιχούν σε 0 και οι άσπρες σε 1

2.4 Βελτιωμένο LBP (ILBP)

Ο Jin κ.α επισήμαναν ότι το LBP θα μπορούν να παραλείπει δομές για τοπικές πληροφορίες δομή κάτω από ορισμένες συνθήκες. Για παράδειγμα ο τελεστής $LBP_{8,1}$ μπορεί να πάρει τιμές 256 (2^8) των συνολικών 511 μοτίβων ($2^9 - 1$, καθώς όλα τα μηδενικά και όλοι οι άσσοι είναι τα ίδια) για μια 3x3 γειτονιά, καθώς το κεντρικό pixel δε συμπεριλαμβάνεται. Προκειμένου να κρατηθεί περαιτέρω πληροφορία, πρότειναν ένα Βελτιωμένο LBP τελεστή (ILBP), ο οποίος συγκρίνει ολα τα pixels (συμπεριλαμβανόμενου και του κεντρικού) με τη μέση τιμή όλων των Pixels στον πυρήνα.(Σχήμα 2.6). Μεταγενέστερα το ILBP επεκτάθηκε και σε γειτονίες διαφόρων μεγεθών.



ΣΧΗΜΑ 2.6: Παράδειγμα του τελεστή ILBP



ΣΧΗΜΑ 2.7: Μια εικόνα προσώπου χωρίζεται σε μικρές περιοχές από τις οποίες τα ιστογράμματα LBP εξάγονται

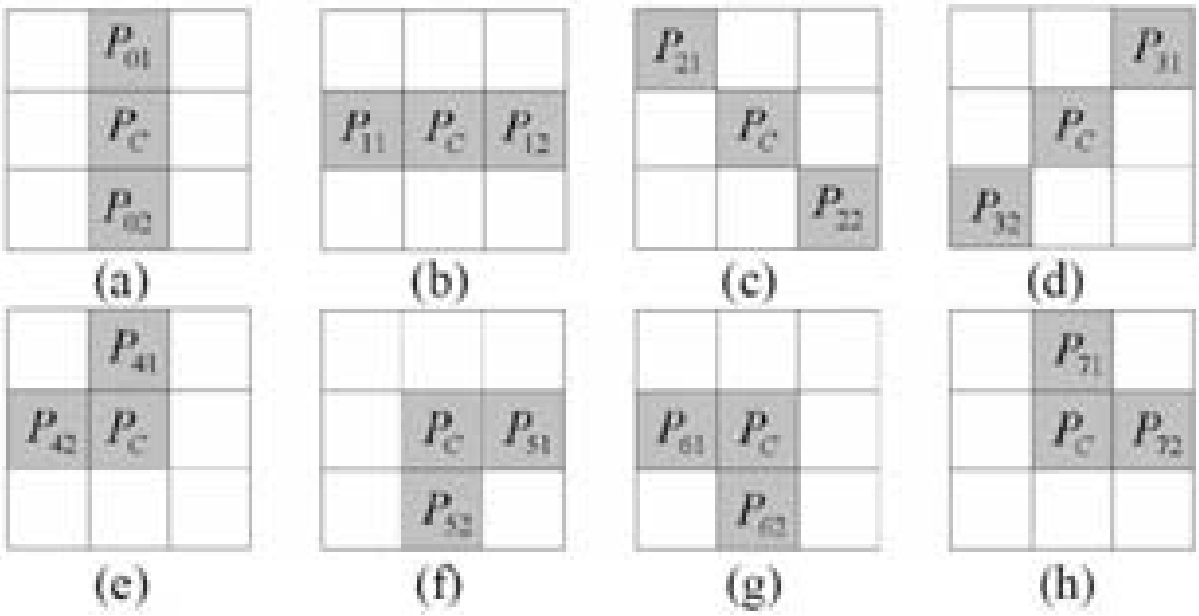
2.5 Τροποποιημένο LBP (MLBP)

Τα ανθρώπινα πρόσωπα διαμορφώνονται από τα χαρακτηριστικά του προσώπου τους. Λαμβάνοντας υπόψη τα χαρακτηριστικά του προσώπου, όπως φρύδια, μάτια, κόρες ματιών, μύτη και το όριο του προσώπου,(8 βασικά χωρικά πρότυπα), για να διατηρηθεί η πληροφορία της μορφής των συνιστωσών του προσώπου, Εισήγαγαν το τροποποιημένο LBP (MLBP) με την προσθήκη των χωρικών προτύπων, αντί σύγκριση του κεντρικού pixel με όλα τα γειτονικά εικονοστοιχεία, στο MLBP, δύο εικονοστοιχεία στη γειτονιά μπορούν να θεωρηθούν ως ένα ζεύγος (P_{i1}, P_{i2}), συμφωνά με αυτό το χωρικό πρότ Κάθε ένα από τα χωρικά πρότυπα αντιστοιχούν σε ένα δυαδικό ψηφίο, έτσι Ο τελεστής MLBP παράγει 256 διαφορετικές τιμές . Ο MLBP υπολογίζεται ως εξής:

$$MLBP = \sum_{i=0}^{7} s_i(x) 2^i \tag{2.10}$$

όπου

$$s_i(x) = \begin{cases} 1 & \text{εάν}(p_c > p_{i1}) \text{ και } (p_c > p_{i2}) \\ 0 & \text{διαφορετικά} \end{cases} \tag{2.11}$$



ΣΧΗΜΑ 2.8: 8 χωρικά πρότυπα στον τελεστή MLBP

2.6 SFLB

Πρόκειται για μια μέθοδο που υλοποιήθηκε στα πλαίσια της εργασίας. Είναι το βασικό πρότυπο LBP για τη γειτονιά 3x3, τώρα όμως αντί να έχουμε σύγκριση του γειτονικού pixel με το κεντρικό ,έχουμε μια τιμή που προκύπτει από τη συνάρτηση:

$$s(x) = e^{\frac{-(i_n - i_c)^2}{2\sigma^2}} \tag{2.12}$$

ή από τη συνάρτηση:

$$s(x) = 1 - e^{\frac{-(i_n - i_c)^2}{2\sigma^2}} \tag{2.13}$$

Έπειτα γίνεται η κατωφλίωση, να αντιστοιχίζεται δηλαδή κάθε pixel της γειτονιάς σε τιμές 0 ή 1 συμφωνά με τη συνάρτηση:

$$g(x) = \begin{cases} 1 & \text{if } s(x) < T \\ 0 & \text{if } s(x) \geq T \end{cases} \tag{2.14}$$

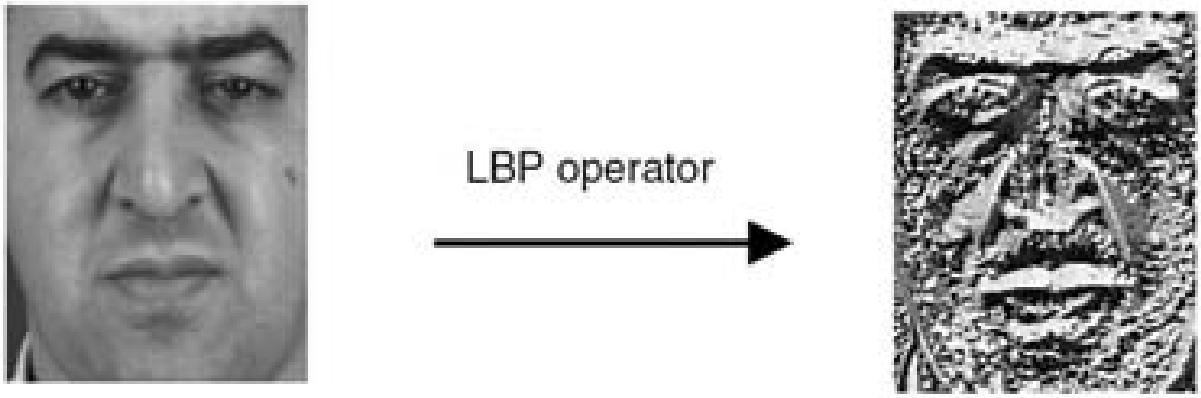
όπου T<1 είναι μια τιμή που καθορίζεται από το χρήστη. Έτσι η SFLB τιμή για το κεντρικό pixel δίνεται όπως και για το βασικό LBP από τη σχέση:

$$SFLBP = \sum_{i=0}^{7} g(x)2^i \tag{2.15}$$

Η μέθοδος μπορεί να επεκταθεί εκτός από την κλασική γειτονιά και για γειτονιά μεγαλύτερης ακτίνας.

2.7 Μερικά παραδείγματα

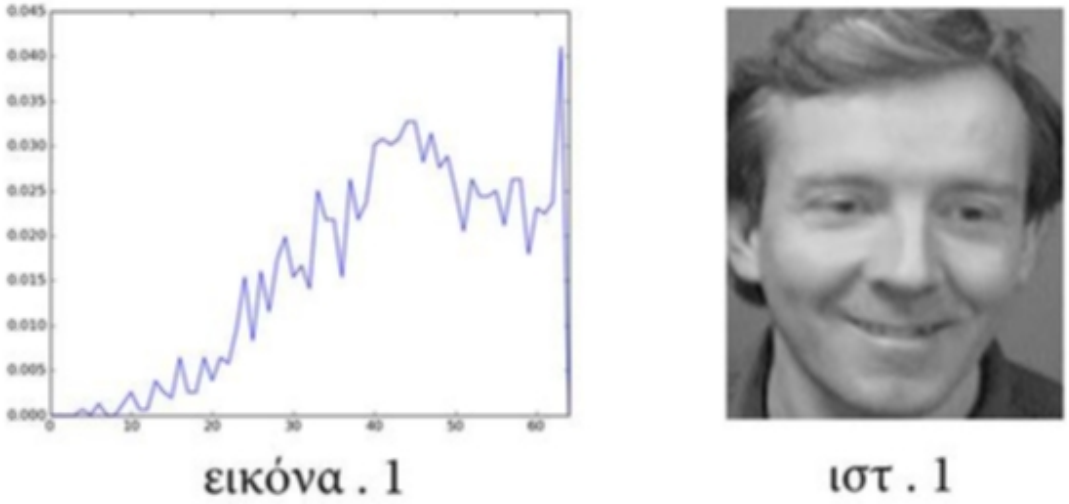
Έχοντας μια αρχική εικόνα, θα δούμε το αποτέλεσμα των τελεστών LBP που έχουν υλοποιηθεί και μελετηθεί στην παρουσία εργασία.



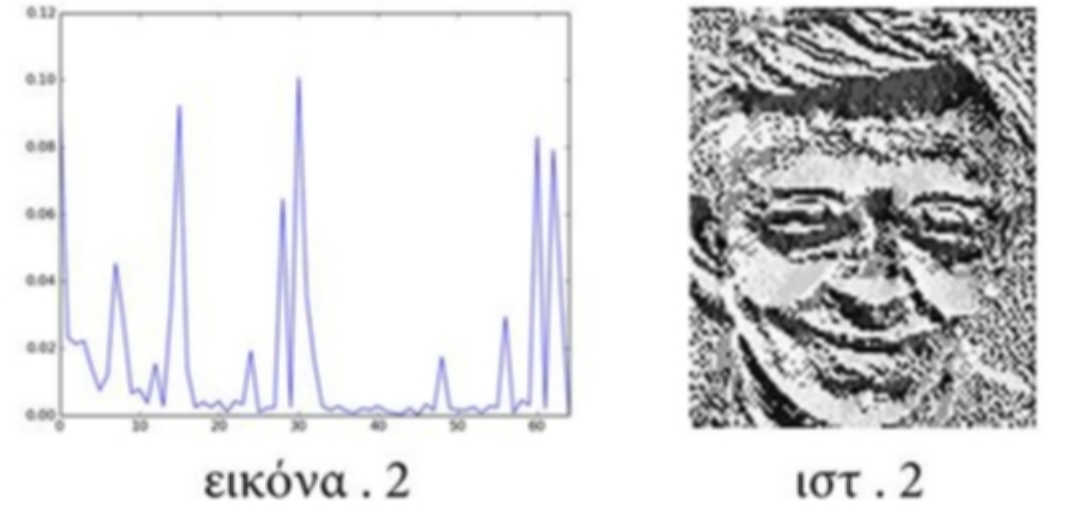
ΣΧΗΜΑ 2.9: Μετατροπή σε LBP

Τελεστές LBP, εικόνες και ιστογράμματα των LBP εικόνων.

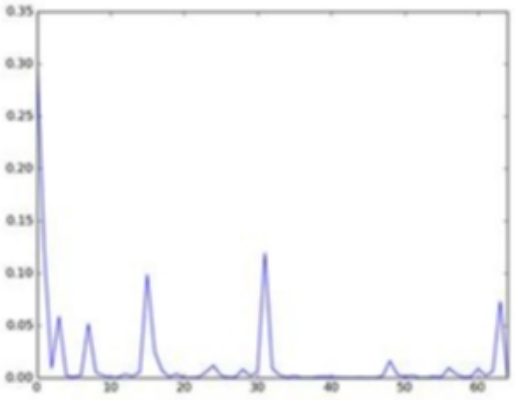
Τα ιστογράμματα εχουν έχουν υπολογιστεί για bins=64 και είναι κανονικοποιημένα,τιμή που δίνεται από το χρήστη.



ΣΧΗΜΑ 2.10: Εικονά1:Η αρχική εικόνα. Ιστόγραμμα1:Το ιστόγραμμα της απλής εικόνας



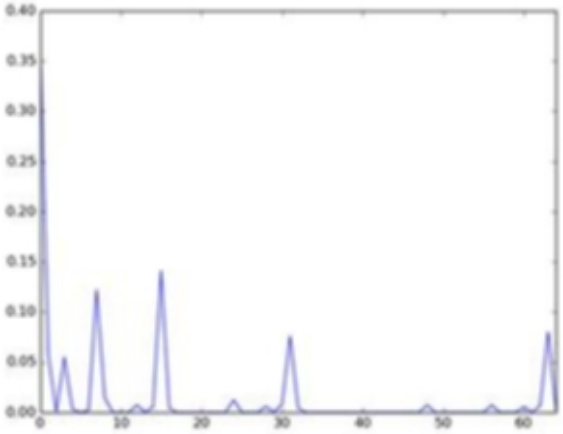
ΣΧΗΜΑ 2.11: Εικονα2:Ο κλασικός τελεστής LBP με τη γειτονιά 3x3. Ιστόγραμμα2:Το ιστόγραμμα της κλασικής LBP εικόνας για τη γειτονιά 3x3



εικόνα . 4

ιστ . 4

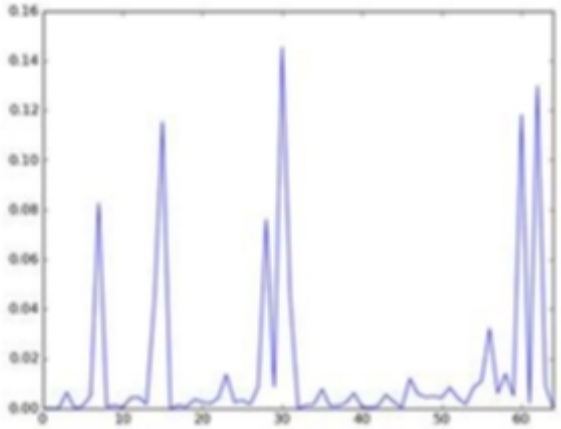
ΣΧΗΜΑ 2.12: Εικονά4:Ο τελεστής Circular LBP με P=16 γειτονιά R=2 ακτίνα. Ιστόγραμμα4:Το ιστόγραμμα της Circular LBP εικόνας με P=16 γειτονιά R=2 ακτίνα



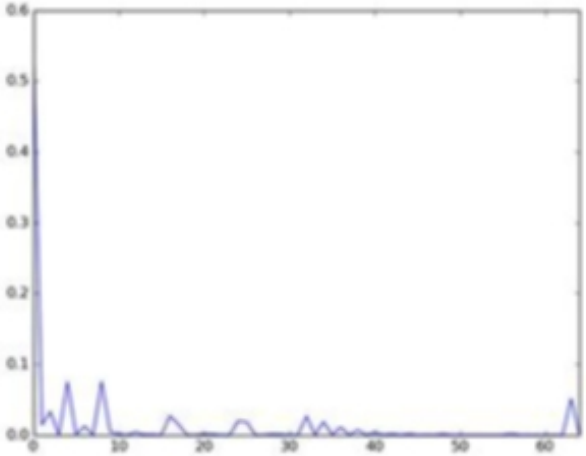
εικόνα . 5

ιστ . 5

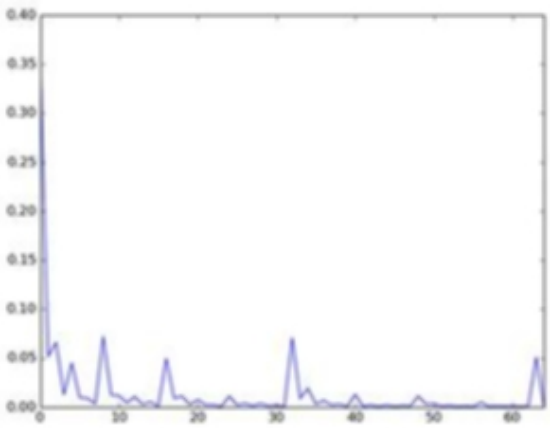
ΣΧΗΜΑ 2.13: Εικονα5:Ο τελεστής Uniform LBP με P=16 γειτονιά R=2 ακτίνα. Ιστόγραμμα5:Το ιστόγραμμα της Uniform LBP εικόνας με P=16 γειτονιά R=2 ακτίνα.



ΣΧΗΜΑ 2.14: Εικονά3:Ο τελεστής Improved LBP. Ιστόγραμμα3:Το ιστόγραμμα της Improved Lbp εικόνας



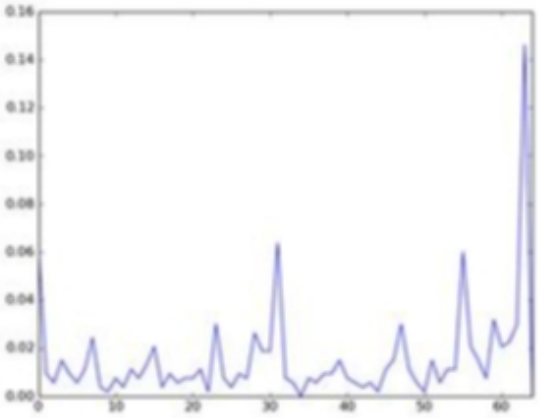
ΣΧΗΜΑ 2.15: Εικονά8:Ο τελεστής Modified LBP. Ιστόγραμμα8:Το ιστόγραμμα της Modified LBP εικόνας



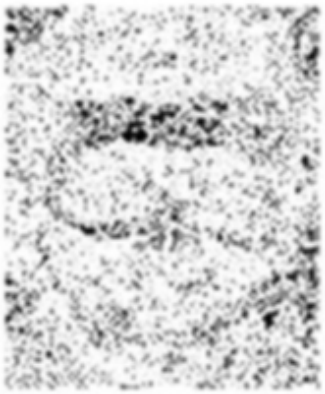
εικόνα . 7

ιστ . 7

ΣΧΗΜΑ 2.16: Εικόνα7:Ο τελεστής SFLBP για τη συναρτηση $s(x) = e^{\frac{-(i_n - i_c)^2}{2\sigma^2}}$ και T=0.8 Ιστόγραμμα7: Το ιστόγραμμα της SFLBP

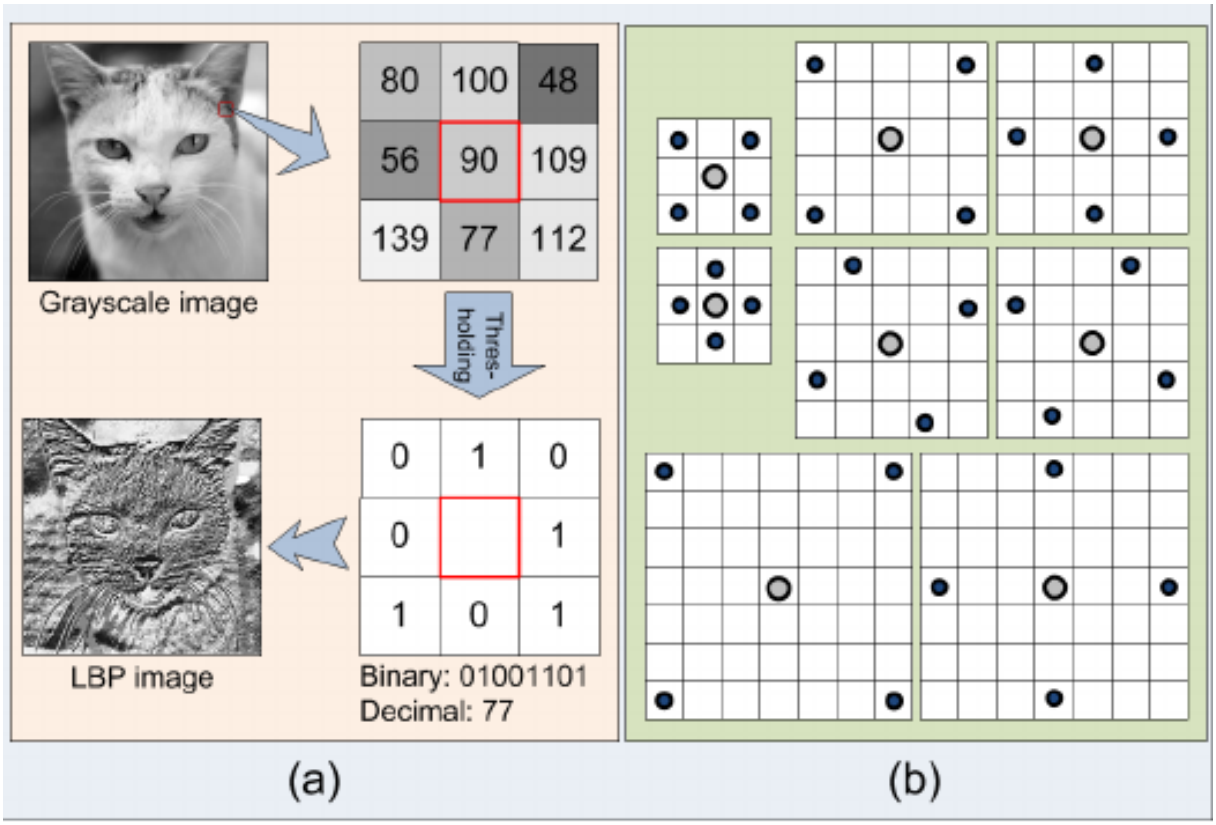


εικόνα . 6



ιστ . 6

ΣΧΗΜΑ 2.17: Εικονά6:Ο τελεστής SFLBP για τη συναρτηση $s(x) = 1 - e^{\frac{-(i_n - i_c)^2}{2\sigma^2}}$ και T=0.8 Ιστόγραμμα6: Το ιστόγραμμα της SFLBP



ΣΧΗΜΑ 2.18: Παραδείγματα LBP 3x3 και Circular LBP

ΚΕΦΑΛΑΙΟ 3

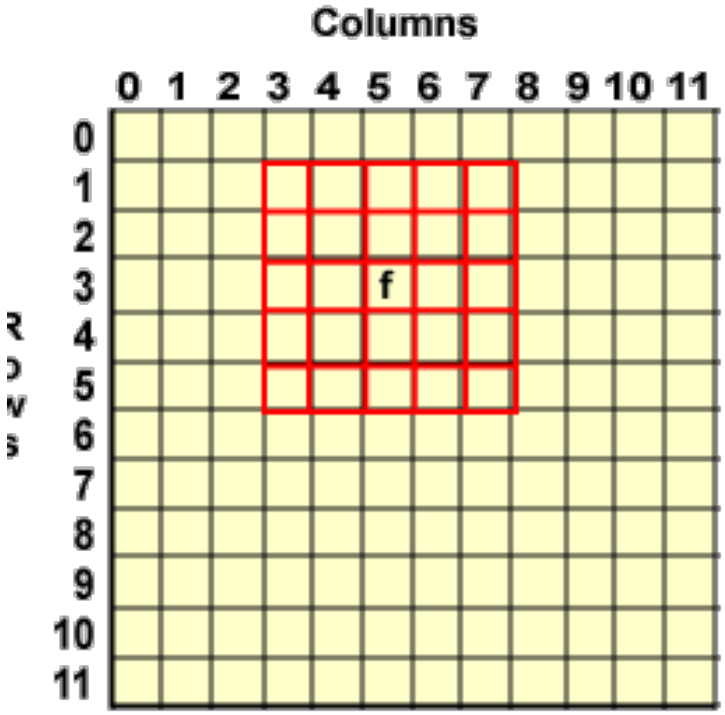
Η ΠΡΟΤΕΙΝΟΜΕΝΗ ΜΕΘΟΔΟΛΟΓΙΑ

3.1 Κατασκευή Λεξικού

Η μέθοδος που παρουσιάζεται στην εργασία αυτή βασίζεται στο μοντέλο Bag of Words. Το μοντέλο Bag of Words είναι μια από τις πιο δημοφιλείς μεθόδους αναπαράστασης για την

κατηγοριοποίηση αντικείμενων. Η κεντρική ιδέα της μεθόδου αυτής είναι ο τεμαχισμός (κβαντισμός) κάποιων εξαγόμενων δεδομένων ως "οπτικές λέξεις". Ο τρόπος για την απόσπαση αυτών των δεδομένων διαφοροποιεί των αλγόριθμο, γνωστοί αλγόριθμοι για αυτό είναι ο SIFT,SURF κ.λ.π. Στην παρούσα εργασία θα ασχοληθούμε με έναν διαφορετικό τρόπο εξαγωγής χαρακτηριστικών και δημιουργίας του Bag of Wors λεξικού. Για να εξάγουμε σωστά λέξεις θα πρέπει να χρησιμοποιήσουμε έναν αλγόριθμο ομαδοποίησης, εδώ χρησιμοποιείται ο KMEANS που θα δούμε αναλυτικά παρακάτω.Παρά το γεγονός ότι μια σειρά από μελέτες έχουν δείξει ενθαρρυντικά αποτελέσματα για τη αναπαράσταση Bag of Words στην κατηγοριοποίηση αντικείμενων, θεωρητικές μελέτες πάνω σε καλύτερες ιδιότητες του μοντέλου είναι απλησίαστο θέμα και δύσκολο λόγω της χρήσης ευρετικών διαδικασιών ομαδοποίησης. Εμπνευσμένοι από την κατηγοριοποίηση κειμένου, η BoW γίνεται μέθοδος αναπαράσταση περιεχομένου εικόνας ,ενώ έχει εφαρμοστεί με επιτυχία στη κατηγοριοποίηση αντικείμενων. Σε μια τυπική Bag of Words αναπαράσταση, "ενδιαφέροντα" τοπικά "αποκόμματα" από μια εικόνα για πρώτη φορά ταυτοποιήθηκαν, είτε με πυκνή δειγματοληψία (Nowak κ.α. , 2006,Winn κ.α 2005) ή από έναν ανιχνευτή σημείων ενδιαφέροντος (Lowe, 2004). Αυτα τα αποκόμματα μπορούμε από δω και στο εξής να τα καλούμε λέξεις ή σημεία κλειδιά, αναπαρίστανται προγραμματιστικά ως διανύσματα πολλών διαστάσεων. Στην παρούσα εργασία θα δουμε το συνδυασμο των τελεστων LBP με τον Bag of Words τρόπο, δηλαδή όλοι οι υπολογισμοί γίνονται πάνω σε LBP εικόνες. Το πρώτο βήμα για τον υπολογισμό του BoW είναι η δειγματοληψία.Θα πρέπει να επιλεχθούν περιοχές ενδιαφέροντος σε κάθε εικόνα. Κάποιοι εξαγωγείς περιοχών ενδιαφέροντος, όπως ο SIFT χρησιμοποιούν ανιχνευτές λέξεων πολλαπλής κλίμακας, αλλά η πυκνή ή τυχαία δειγματοληψία υπερτερεί των μεθόδων δειγματοληψίας.Στη μέθοδο που προτείνεται εδώ η δειγματοληψία γίνεται ως εξής: επιλέγουμε από κάθε εικόνα τυχαία pixel και μια γειτονία σε αποστάσεις ακτίνας R=W/2 από το pixel,οπού W είναι παράμετρος του χρήστη, όποτε προκύπτει μια WxW γειτονία Τέτοιες τυχαίες γειτονιές είναι υποψήφιες ως λέξεις.

Δεν επιτρέπεται η επιλογή οριακών pixel, γιατί δε μπορούμε να έχουμε γειτονία και έχουμε



ΣΧΗΜΑ 3.1: Επιλογή W=4,4x4 υποψήφιας γειτονιάς-λέξης με R=2,ακτίνα

μόνο μια γειτονιά κάθε φορά, δεν υπάρχουν επικαλυπτόμενες γειτονίες. Για κάθε εικόνα που

θέλουμε να ταξινομηθεί επιλέγουμε τέτοιες περιοχές. Ο αριθμός περιοχών για κάθε εικόνα ποικίλει. Είναι μια παράμετρος L που δίνεται από το χρήστη. Για τις εικόνες που συμμετέχουν στη διαδικασία, έχει γίνει φιλτράρισμα από LBP τελεστή. Όπως προαναφέρθηκε για να

γίνουν χρήσιμα γνωρίσματα οι περιοχές δειγματοληψίας και να προκύψουν από τις περιοχές των εικόνων σημαίνουσες λέξεις για τη δημιουργία του οπτικού λεξικού, χρησιμοποιείται ο αλγόριθμος ομαδοποίησης KMEANS. Το μέγεθος του λεξικού που συνήθως χρησιμοποιούν οι ερευνητές κυμαίνεται από μερικές εκατοντάδες σε χιλιάδες και δεκάδες χιλιάδες λέξεις. Πειράματα διεξήχθησαν για τα BOWL γνωρίσματα και βρέθηκε ότι το βέλτιστο λεξικό αποτελείται από 1000 λέξεις. Έπειτα από την κατασκευή του λεξικού,η κάθε εικόνα του σύνολου

εκπαίδευσης και το συνόλου επαλήθευσης αντιστοιχίζεται σε λέξεις του BOWL.Κάθε pixel της εικόνας αντιπροσωπεύει μια λέξη. Έτσι μπορεί να αναπαρασταθεί σαν ένα ιστόγραμμα οπτικών λέξεων.

ΣΧΗΜΑ 3.2: Επιλογή για υποψήφιες λέξεις

3.2 Περιγραφή Αλγορίθμου

Όταν θέλουμε να κάνουμε ταξινόμηση χωρίζουμε τα δεδομένα μας σε δυο σύνολα, το σύνολο εκπαίδευσης (train set) και το σύνολο επαλήθευσης (test set). Το σύνολο εκπαίδευσης είναι αυτό με το οποίο ο ταξινομητής μαθαίνει. Το σύνολο επαλήθευσης επαληθεύει τα αποτελέσματα, οπού θα εξεταστούν αν είναι τα σωστά. Βλέπουμε αν έχει γίνει σωστά η εκπαίδευση. Από το σύνολο επαλήθευσης εξαρτάται η αποδοτικότητα της μεθόδου. Χωρίζουμε τα δεδομένα μας σε δυο σύνολα. Αφού γίνει αυτό μπορούμε να ξεκινήσουμε τον BOWL. Τα βήματα είναι τα εξής:

1. Φιλτράρουμε τις εικόνες του σύνολου,δε δουλεύουμε πάνω στις εικόνες που έχουμε εξ' αρχής, άλλα εξάγουμε τις LBP εικόνες. Δοκιμάζουμε όποιον τελεστή θέλουμε.

2. Ίδιες εικόνες είναι τοποθετημένες διαδοχικά και ονομαστικά αποθηκευμένες έτσι ώστε να μπορούμε να ξέρουμε από την αρχή τις κατηγορίες,ώστε να ξέρουμε αν κάνουμε σωστή επαλήθευση. Οι κατηγορίες για κάθε σύνολο αποθηκεύονται.

3. Για κάθε εικόνα,κάνουμε δειγματοληψία. Επιλέγουμε τυχαία pixel, ο αριθμός τους L δίνεται ως παράμετρος από το χρήστη. Για κάθε pixel επιλέγουμε μια WxW περιοχή πάνω στην εικόνα. Κάθε pixel, επιλέγεται μόνο μια φόρα,οριακά pixel δεν επιλέγονται. Αυτές οι L, WxW περιοχές, για κάθε εικόνα θα είναι τα δείγματα μας.

4. Καλείται ο K-means για να δημιουργηθούν οι λέξεις και να ολοκληρωθεί η διαδικασία κατασκευές του λεξικού.

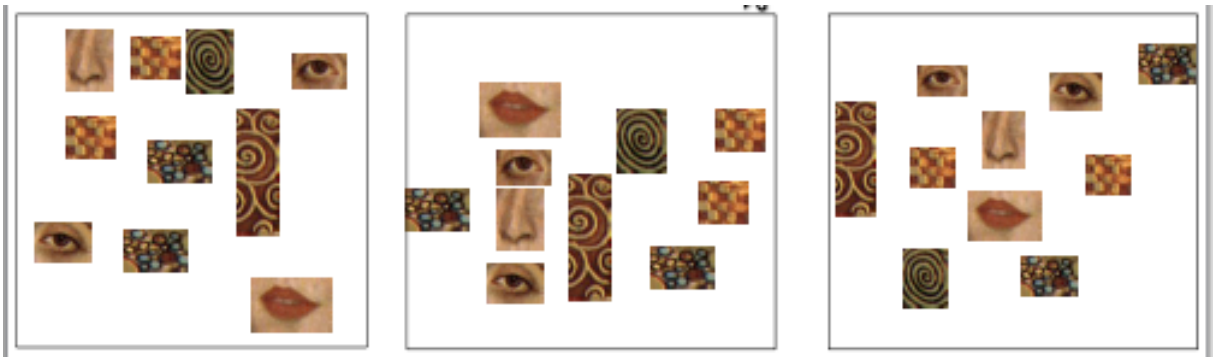
5. Κάθε pixel μιας εικόνας αντιστοιχίζεται σε μια λέξη-κέντρο, συνήθως στο κέντρο με τη μικρότερη απόσταση από τη διανυσματική έκφραση του pixel και της γειτονιάς. Η απόσταση αυτή μπορεί να είναι η ευκλείδεια απόσταση η και άλλοι τρόποι.

6. Εξάγονται για κάθε εικόνα, που τώρα αποτελεί έκφρασή του οπτικού λεξικού, τα ιστογράμματα.

7. Τα ιστογράμματα άφου κανονικοποιηθούν εισάγονται σε αλγορίθμους ταξινόμησης και οι κατηγορίες. Γνωστοί αλγόριθμοι ταξινόμησης είναι ο SVM,KNN κ.α που θα δούμε παρακάτω.

8. Εξετάζουμε για το σύνολό επαλήθευσης, Οι τιμές που επιστρέφει ο ταξινομημένες είναι οι που τιμές αναμέναμε,αν ταιριάζουν. Έχουμε μια εικόνα για την την αποδοτικότητα του αλγορίθμου.

Τροποποιημένο LBP (MLBP) Τα βήματα 1-6 αφορούν και τα 2 σύνολα.



ΣΧΗΜΑ 3.3: Απλουστευμένα παραδείγματα λεξικών

3.3 Ο Αλγορίθμος ομαδοποίησης των K-Μέσων

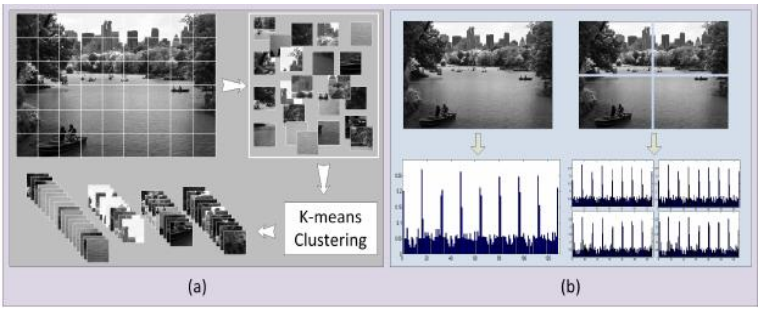
Ο αλγόριθμος k-means (k-μέσων)είναι ένας αλγόριθμος(MacQueen, 1967)που ομαδοποιεί αντικείμενα βάσει των χαρακτηριστικών των k μεριδίων.Αποτελεί μεταβλητή του αλγόριθμου μεγιστοποίησης αναμονής (expectation-maximization algorithm-EM) όπου σκοπός είναι να οριστεί ο k-means δεδομένων που προήλθαν από Gaussian κατανομές. Ο αλγόριθμος υποθέτει ότι τα χαρακτηριστικά του αντικειμένου δημιουργούν ένα χώρο διανυσμάτων και ο σκοπός του είναι να ελαχιστοποιήσει τη συνολική διακύμανση της ομάδας ή τη συνάρτηση τετραγωνικού σφάλματος:

$$V = \sum_{i=1}^{k} \sum_{x_j \exists S_i} |x_j - \mu_i|^2 \tag{3.1}$$

όπου υπάρχουν k ομάδες S_i, i= 1,2,...,k και i είναι το κεντροειδές ή το μεσαίο σημείο από όλα τα σημεία. Τα βασικά βήματα του αλγόριθμου είναι τα εξής:

1. Επιλογή του αριθμού των ομάδων.
2. Τυχαία δημιουργία k ομάδων και ορισμός των κεντροειδών των ομάδων.
3. Μεταβίβαση του κάθε σημείου στο κεντροειδές της κοντινότερης ομάδας.
4. Υπολογισμός των νέων κεντροειδών των ομάδων.
5. Επανάληψη μέχρι να συγκλίνει ο αλγόριθμος σε κάποιο κριτήριο. Ο αλγόριθμος ξεκινά διαχωρίζοντας τα αρχικά σημεία σε k αρχικά σύνολα είτε τυχαία είτε χρησιμοποιώντας ευρετικά δεδομένα. Στη συνέχεια υπολογίζει το μεσαίο ή το κεντροειδές του κάθε συνόλου, υλοποιεί νέο διαχωρισμό ώστε το κάθε σημείο να σχετίζεται με το κοντινότερο κεντροειδές. Έπειτα τα κεντροειδή ξαναϋπολογίζονται για τις νέες ομάδες, ο αλγόριθμος επαναλαμβάνει τα δυο βήματα ωσότου τα σημεία δεν μπορούν να αλλάξουν ομάδες (ή εναλλακτικά τα κεντροειδή παραμένουν αμετάβλητα).

   Ο αλγόριθμος αυτός παραμένει διάσημος επειδή τείνει σε κάποιο όριο πολύ γρήγορα. Όσον αφορά την απόδοση ο αλγόριθμος δεν εγγυάται ότι θα αγγίξει το βέλτιστο. Η ποιότητα της τελική λύσης εξαρτάται πολύ από το αρχικό σύνολο ομάδων και μπορεί να είναι πολύ χαμηλότερη από το συνολικό βέλτιστο. Επίσης ένα άλλο μειονέκτημα του αλγόριθμου είναι ότι ο αριθμός των ομάδων πρέπει να οριστεί εξαρχής.



ΣΧΗΜΑ 3.4: Απλουστευμένα παραδείγματα λεξικών

3.4 Τεχνικές ταξινόμησης

Όταν έχουν εξαχθεί τα ιστογράμματα, τα οποία είναι μια διανυσματική έκφραση των οτικοποιημένων σε λέξεις εικόνων, αυτά εισάγονται σε έναν αλγόριθμο ταξινόμησης. Η κατηγοριοποίηση είναι μία τεχνική της εξόρυξης δεδομένων, κατά την οποία ένα στοιχείο ανατίθεται σε ένα προκαθορισμένο σύνολο κατηγοριών. Ο όρος κατηγοριοποίηση συναντάται στην βιβλιογραφία και ως ταξινόμηση. Γενικότερα, ο στόχος της διαδικασίας αυτής είναι η ανάπτυξη ενός μοντέλου, το οποίο αργότερα θα μπορεί να χρησιμοποιηθεί για την κατηγοριοποίηση μελλοντικών δεδομένων. Τέτοια παραδείγματα είναι ο διαχωρισμός των emails με βάση την επικεφαλίδα τους ή το περιεχόμενό τους, η πρόβλεψη καρκινικών κυττάρων χαρακτηρίζοντας τα ως καλοήθη ή κακοήθη, η κατηγοριοποίηση πελατών μιας τράπεζας ανάλογα με την πιστωτική τους ικανότητα, κατηγοριοποίηση δευτερευόντων δομών πρωτεΐνης ως alpha-helix, beta-sheet, ή random coil, Χαρακτηρισμός ειδήσεων ως οικονομικές, αθλητικές, πολιτιστικές, πρόβλεψης καιρού, κ.λ.π. Γενικά κάθε αλγόριθμος ταξινομήμησης

περιλαμβάνει τα εξής βήματα:

- Είσοδος: συλλογή από εγγραφές Κάθε εγγραφή περιέχει ένα σύνολο από γνωρίσματα (attributes)

  Ένα από τα γνωρίσματα είναι η κλάση (class)

- Συνήθως το σύνολο δεδομένων εισόδου χωρίζεται σε: ένα σύνολο εκπαίδευσης (training set) και ένα σύνολο ελέγχου (test test).

- Το σύνολο εκπαίδευσης χρησιμοποιείται για να κατασκευαστεί το μοντέλο και το σύνολο ελέγχου για να το επικυρώσει.

- Έξοδος: ένα μοντέλο (model) για το γνώρισμα κλάση ως μια συνάρτηση των τιμών των άλλων γνωρισμάτων.

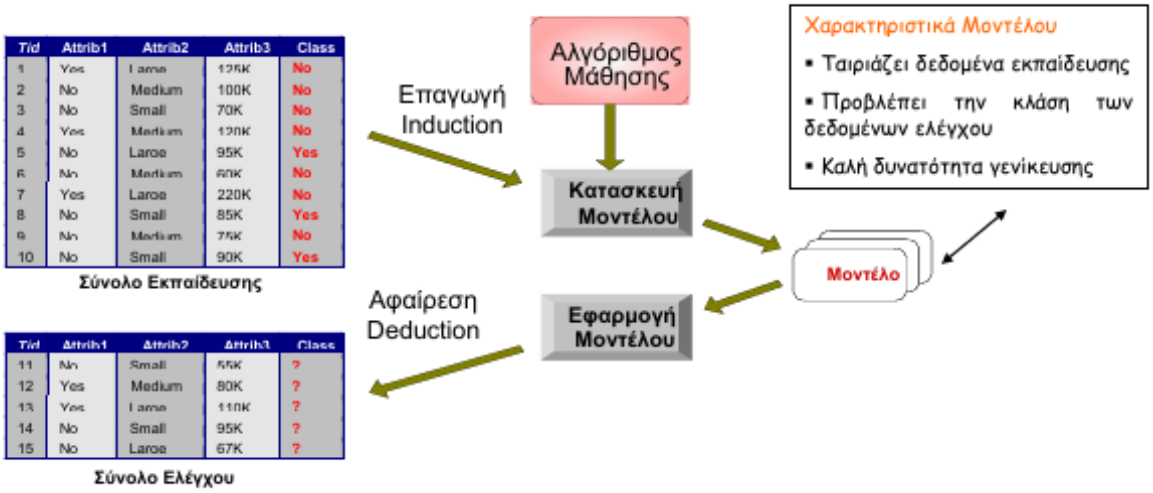
Βήματα Ταξινόμησης.

1. Κατασκευή Μοντέλου

   Χρησιμοποιώντας το σύνολο εκπαίδευσης (στις εγγραφές του το γνώρισμα της κλάσης

είναι προκαθορισμένο) Το μοντέλο μπορεί να είναι ένα δέντρο ταξινόμησης, κανόνες, μαθηματικοί τύποι κλπ).

2. Εφαρμογή Μοντέλου για την ταξινόμηση μελλοντικών ή άγνωστων αντικειμένων Εκτίμηση της ακρίβειας του μοντέλου με χρήση συνόλου ελέγχου Accuracy rate: το ποσοστό των εγγραφών του συνόλου ελέγχου που ταξινομούνται σωστά από το μοντέλο.



ΣΧΗΜΑ 3.5: Ταξινόμηση

3.5 Η Μέθοδος Support Vector Machines (SVM)

Κάποιοι αλγόριθμοι είναι ευρέως διαδεδομένοι,παρακάτω γίνεται μια υπενθύμιση πολύ δημοφιλών μεθόδων που χρησιμοποιήθηκαν στην παρούσα εργασία. Τα Support Vector Machines

(SVMs) είναι ένα σύνολο μεθόδων εκμάθησης που χρησιμοποιούνται για προβλήματα ταξινόμησης και παλινδρομικής ανάλυσης. Η κύρια ιδέα των SVM είναι να κατασκευαστεί ένα υπερεπίπεδο, έτσι ώστε η απόσταση του διαχωρισμού μεταξύ των θετικών και αρνητικών παραδειγμάτων να μεγιστοποιείται. Τα διανύσματα των πιο κοντινών στοιχείων στο υπερεπίπεδο αυτό είναι τα υποστηρικτικά διανύσματα (support vectors). Αυτή η επιθυμητή ιδιότητα

επιτυγχάνεται ακολουθώντας την αρχή της Ελαχιστοποίηση του Δομικού Ρίσκου (Structural

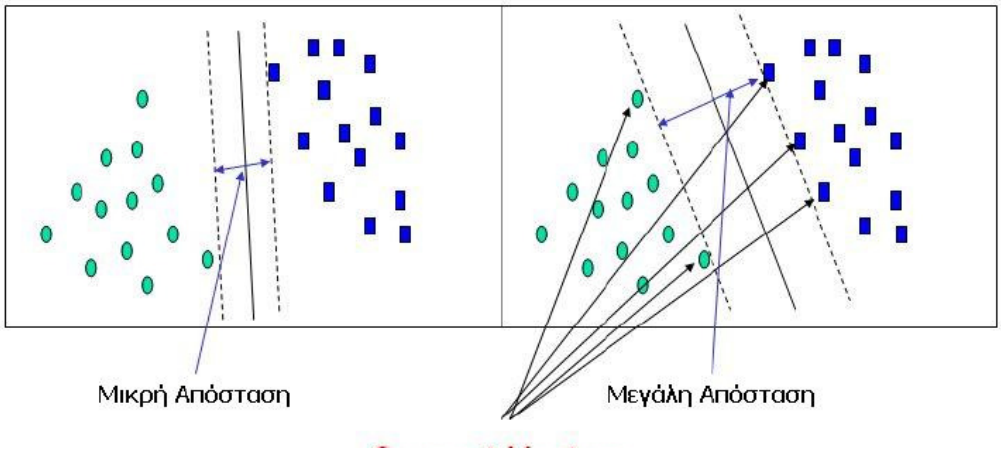
Risk Minimization) από τη θεωρία της μηχανικής μάθησης. Η ιδέα της ελαχιστοποίησης του δομικού ρίσκου είναι να βρεθεί μια υπόθεση h για την οποία μπορούμε να εγγυηθούμε το χαμηλότερο πραγματικό σφάλμα. Το πραγματικό σφάλμα της h είναι η πιθανότητα της h να κάνει λάθος σε ένα τυχαία επιλεγμένο παράδειγμα το οποίο δεν έχει δει στο παρελθόν. Το πλεονέκτημα της τεχνικής αυτής είναι ότι επιτυγχάνονται καλές επιδόσεις στα προβλήματα ταξινόμησης χωρίς να ενσωματώνεται γνώση από τον τομέα του προβλήματος. Βλέποντας

τα δεδομένα εισόδου σαν δύο σύνολα διανυσμάτων σε ένα ν-διάστατο χώρο, το SVM θα κατασκευάσει ένα διαχωριστικό υπερεπίπεδο σε αυτόν το χώρο, που θα μεγιστοποιεί την απόσταση μεταξύ των δύο συνόλων. Για τον υπολογισμό της απόστασης αυτής, κατασκευάζονται δύο παράλληλα υπερεπίπεδα, ένα σε κάθε πλευρά του διαχωριστικού υπερεπιπέδου, τα οποία "σπρώχνονται" πάνω στα δύο σύνολα δεδομένων. Διαισθητικά, ένας καλός διαχωρισμός επιτυγχάνεται από το υπερεπίπεδο που έχει τη μεγαλύτερη απόσταση από τα γειτονικά σημεία δεδομένων και των δύο συνόλων, δεδομένου ότι σε γενικές γραμμές όσο μεγαλύτερη είναι η απόσταση τόσο καλύτερο είναι το λάθος γενίκευσης του ταξινομητή. Η ταξινόμηση των δεδομένων είναι μια κοινή ανάγκη στο πεδίο της μηχανικής μάθησης. Ας υποθέσουμε ότι δίνονται κάποια σημεία δεδομένων που ανήκουν στα δύο σύνολα, και ο στόχος είναι να αποφασίσουμε σε ποιο σύνολο θα μπει ένα νέο σημείο δεδομένων. Στην περίπτωση των SVM, ένα σημείο δεδομένων θεωρείται σαν ένα διάνυσμα p-διαστάσεων, και θέλουμε να ξέρουμε αν μπορούμε να χωρίσουμε αυτά τα σημεία με ένα $p-1$-διάστατο υπερεπίπεδο. Αυτό ονομάζεται γραμμικός ταξινομητής. Υπάρχουν πολλά υπερεπίπεδα που θα μπορούσαν να ταξινομήσουν τα δεδομένα. Ωστόσο, ενδιαφερόμαστε επιπλέον να διαπιστώσουμε εάν μπορούμε να πετύχουμε το μέγιστο διαχωρισμό (απόσταση) μεταξύ των δύο κλάσεων. Με αυτό εννοούμε ότι διαλέγουμε το υπερεπίπεδο, έτσι ώστε η απόσταση από το υπερεπίπεδο στο πλησιέστερο σημείο δεδομένων να μεγιστοποιείται. Αυτό σημαίνει ότι η κοντινότερη απόσταση ανάμεσα σε ένα σημείο στο ένα διαχωρισμένο υπερεπίπεδο και σε ένα σημείο στο άλλο διαχωρισμένο υπερεπίπεδο μεγιστοποιείται. Αν υπάρχει ένα τέτοιο υπερεπίπεδο, είναι γνωστό ως το υπερεπίπεδο μέγιστου-διαχωρισμού, και ένας τέτοιος γραμμικός ταξινομητής είναι γνωστός ως ένας ταξινομητής μέγιστου-διαχωρισμού. Τα SVMs ανήκουν στην κατηγο-

ρία των γενικευμένων γραμμικών ταξινομητών. Μια ειδική διότητά τους είναι ότι ταυτόχρονα ελαχιστοποιούν το εμπειρικό σφάλμα ταξινόμησης και μεγιστοποιούν τη γεωμετρική απόσταση. Ως εκ τούτου, είναι ταξινομητές μέγιστου- διαχωρισμού. Μία αξιοσημείωτη ιδιότητα

των SVMs είναι ότι η ικανότητά τους να μαθαίνουν είναι ανεξάρτητη από τις διαστάσεις του χώρου χαρακτηριστικών. Τα SVMs μετράνε την πολυπλοκότητα των υποθέσεων με βάση την απόσταση που μπορούν να διαχωρίσουν τα στοιχεία, και όχι με βάση τον αριθμό των χαρακτηριστικών. Αυτό σημαίνει ότι μπορούμε να γενικεύσουμε ακόμη και με την παρουσία πάρα πολλών χαρακτηριστικών, αν τα στοιχεία μας μπορούν να διαχωριστούν με ένα ευρύ περιθώριο χρησιμοποιώντας συναρτήσεις από το χώρο υποθέσεων Τα Support Vector Machines

έχουν πολλά ελκυστικά χαρακτηριστικά. Είναι ένα σπάνιο παράδειγμα μεθοδολογίας όπου συνδυάζονται η γεωμετρική διαίσθηση, τα κομψά μαθηματικά,οι θεωρητικές εγγυήσεις και οι πρακτικοί αλγόριθμοι. Μπορούν να εφαρμοστούν αποτελεσματικά σε ένα ευρύ φάσμα προβλημάτων ταξινόμησης. Κλιμακώνονται σε τεράστια σύνολα δεδομένων και είναι ανεξάρτητα του τομέα του προβλήματος. Επιπλέον, μπορούν να αναπτυχθούν αποτελεσματικές συναρτήσεις πυρήνα για κάθε συγκεκριμένο πρόβλημα, προκειμένου να επιτευχθούν ακόμα καλύτερα αποτελέσματα.



ΣΧΗΜΑ 3.6: SVM και διαχωριστικά υπερεπίπεδα

3.6 Η Μέθοδος Naive Bayes

Πρόκειται για μια απλουστευμένη έκδοση του βασικού Bayes αλγορίθμου, η οποία χρησιμοποιείται για την ταξινόμηση των στιγμιότυπων ενός προβλήματος στις προκαθορισμένες κλάσεις του προβλήματος.

Η λειτουργία του Naive Bayes ταξινομητή συνοψίζεται στα ακόλουθα:

- Κάθε στιγμιότυπο X του προβλήματος αποτελείται από ένα σύνολο γνωρισμάτω $x_1, x_2, ...x_n$, δηλαδή $X =< x_1, x_2, ...x_n >$

- Έστω ότι το πρόβλημα έχει m κλάσεις, $c_1, c_2, ...c_m$. Δοθέντος ενός άγνωστου στιγμιότυπου X του προβλήματος, για το οποίο δε γνωρίζουμε σε ποια κλάση ανήκει, ο ταξινομητής προβλέπει ότι το X ανήκει στην κλάση με τη μεγαλύτερη posteriori πιθανότητα. Ο ταξινομητής αναθέτει ένα άγνωστο στιγμιότυπο X του προβλήματος στην κλάση C_i αν και μόνο αν

$$P(C_i|X) > P(C_j|X)1 \ j \ m, j \neq i \tag{3.2}$$

  Έτσι μεγιστοποιείται η πιθανότητα $P(C_i|X)$, η οποία βάσει του θεωρήματος του Bayes δίνεται από τη σχέση:

$$P(C_i|X) = P(X|C_i) * P(C_i)/P(X) \tag{3.3}$$

- Επειδή στον παραπάνω τύπο το P(X) είναι σταθερό για όλα τα στιγμιότυπα, το μόνο που χρειάζεται να μεγιστοποιηθεί είναι η έκφραση $P(X|C_i) * P(C_i)$.

- Ο υπολογισμός του $P(X|C_i)$είναι εξαιρετικά δαπανηρός και προκειμένου να μειώσουμε το υπολογιστικό κόστος υποθέτουμε ότι υπάρχει μια ανεξαρτησία ως προς την κατανομή των κλάσεων. Αυτό σημαίνει πως δοθείσας της κλάσης κάποιου στιγμιότυπου, οι τιμές των γνωρισμάτων του στιγμιότυπου είναι ανεξάρτητες μεταξύ τους, δηλαδή δεν

υπάρχει εξάρτηση μεταξύ των γνωρισμάτων. Έτσι

$$= P(X|C_i) = \prod_{k=1}^{n} P(x_k|C_i) \tag{3.4}$$

Οι πιθανότητες $P(x_k|C_i)$ μπορούν να υπολογιστούν κατευθείαν από τα στιγμιότυπα εκπαίδευσης ως εξής:

- Αν το γνώρισμα A_k είναι κατηγορηματικό (categorical), τότε $P(x_k|C_i) = s_{ik}/s_i$, όπου s_{ik} είναι το πλήθος των στιγμιότυπων εκπαίδευσης της κλάσης C_i με τιμή $A_k = x_k$ και s_i είναι το πλήθος των στιγμιότυπων εκπαίδευσης που ανήκουν στην κλάση C_i.

- Αν το γνώρισμα A_k είναι συνεχές, τότε υποθέτουμε ότι οι τιμές του ακολουθούν τη Gaussian κατανομή

$$P(x_k|C_i) = g(x_k, \mu_c, \sigma_c) = (1/\sqrt{2\pi\pi_{C_i}}) * e^{-(x_k-\mu_{C_i})/2\sigma^2_{C_i}} \tag{3.5}$$

όπου η συνάρτηση $g(x_k, \mu_c, \sigma_c)$είναι η Gaussian συνάρτηση πυκνότητας για το γνώρισμα A_k .

- Προκειμένου να ταξινομήσουμε ένα νέο στιγμιότυπο X, υπολογίζουμε την πιθανότητα $P(X|C_i) * P(C_i)$ για κάθε κλάση Ci του προβλήματος. Το στιγμιότυπο X ανατίθεται στην κλάση C_i αν και μόνο αν

$$P(X|C_i) * P(C_i) > P(X|C_j) * P(C_j) \text{ γιά } 1 \ j \ m, j \neq i \tag{3.6}$$

3.7 Ο Αλγόριθμος KNN

Εντάσσεται στην Κατηγοριοποίηση με χρήση απόστασης. Πρέπει να προσδιορίσουμε την απόσταση μεταξύ ενός στοιχείου και μιας κλάσης. Κάθε κλάση μπορεί να αναπαρασταθεί με

- Κέντρο βάρους (Centroid): η κεντρική τιμή της κλάσης.

- Κεντρικό στοιχείο (Medoid): ένα αντιπροσωπευτικό σημείο – μέλος της.
- Σύνολο από ενδεικτικά σημεία

Η προσέγγιση KNN

- Το σύνολο δεδομένων εκπαίδευσης περιλαμβάνει τις κλάσεις.
- Για να αναθέσουμε ένα νέο στοιχείο σε μια κλάση εξετάζουμε τα K πλησιέστερα σ' αυτό σημεία.
- Τοποθετούμε το νέο στοιχείο στην κλάση που έχει την πλειοψηφία μέσα στα κοντινά στοιχεία.
- Πολυπλοκότητα O(q) για κάθε νέο στοιχείο (q είναι το μέγεθος του συνόλου δεδομένων εκπαίδευσης).

3.8 Δέντρα Αποφάσεων

Τα δέντρα αποφάσεων είναι εποπτευόμενοι αλγόριθμοι που χωρίζουν τα δεδομένα αναδρομικά, βασισμένοι στα χαρακτηριστικά των δεδομένων , μέχρι να ικανοποιηθεί μια τερματική συνθήκη. Ο ταξινομητής δένδρου απόφασης είναι μια από τις πιθανές προσεγγίσεις για πολυεπίπεδη λήψη αποφάσεων. Το πιο σημαντικό στοιχείο του ταξινομητή δένδρου απόφασης είναι η ικανότητα διάσπασης της διαδικασίας απόφασης σε μια συλλογή από πιο απλές αποφάσεις, με τέτοιο τρόπο, έτσι ώστε να παρέχουν μια λύση που είναι συνήθως πιο εύκολο να μεταφραστεί.
Το μοντέλο που δημιουργείται είναι ένα δέντρο. Η λειτουργία του αλγορίθμου βασίζεται στην τεχνική «διαίρει και βασίλευε» για διαίρεση του χώρου αναζήτησης σε υποσύνολα (ορθογώνιες περιοχές) Ένα παράδειγμα κατηγοριοποιείται με βάση την περιοχή στην οποία ανήκει.

Ένα Δέντρο Απόφασης(ΔΑ)ή Δέντρο Κατηγοριοποίησης είναι ένα δέντρο με τις ακόλουθες ιδιότητες:

- Κάθε εσωτερικός κόμβος ονοματίζεται με το όνομα ενός χαρακτηριστικού X_i.
- Κάθε κλαδί/σύνδεση ονοματίζεται με ένα κατηγόρημα που μπορεί να εφαρμοστεί στο χαρακτηριστικό που αποτελεί το όνομα του κόμβου- πατέρα.
- Κάθε φύλλο ονοματίζεται με το όνομα μιας κλάσης.

Χαρακτηριστικές έννοιες.

- Χαρακτηριστικά διάσπασης (splitting attributes) Τα χαρακτηριστικά των παραδειγμάτων στη βάση D που χρησιμο- ποιούνται σαν ονόματα κόμβων του δέντρου, δηλ. επιλέχτηκαν ως καλύτερα χαρακτηριστικά.
- Χαρακτηριστικό στόχου (target attribute).Το χαρακτηριστικό που οι τιμές του αντιπροσωπεύουν τις κλάσεις κατηγοριοποίησης.
- κατηγοριοποίησης. Κατηγορήματα διάσπασης (splitting predicates).Τα κατηγορήματα που χρησιμοποιούνται σαν ονόματα των κλαδιών/ συνδέσεων του δέντρου.
- Κριτήριο διάσπασης (splitting criterion) Το κριτήριο με βάση το οποίο επιλέγεται το καλύτερο χαρακτηριστικό διάσπασης κάθε φορά.
- Κριτήριο τερματισμού (stopping criterion). Το κριτήριο με βάση το οποίο τερματίζεται ο αλγόριθμος.

Παραλλαγές των δύο αυτών κριτηρίων δημιουργούν μια ποικιλία αλγορίθμων.

Βασικά θέματα.

- Επιλογή χαρακτηριστικών διάσπασης

  – Διαφορετικά σύνολα χαρακτηριστικών διάσπασης έχουν σαν αποτέλεσμα διαφορετικά ΔΑ με διαφορετική απόδοση.

  – Η επιλογή τους στηρίζεται όχι μόνο στο σύνολο εκπαίδευσης, αλλά και στη γνώμη του εμπειρογνώμονα.

- Διάταξη των χαρακτηριστικών διάσπασης-Διασπάσεις

  – Η σειρά επιλογής των χαρακτηριστικών διάσπασης παίζει σημαντικό ρόλο στην απόδοση ενός ΔΑ.

  – Ο αριθμός διασπάσεων συνδέεται με τη διάταξη των χαρακτηριστικών διάσπασης. Ο αριθμός διασπάσεων μπορεί εύκολα να προσδιοριστεί όταν το πεδίο είναι μικρό (λίγα χαρακτηριστικά, λίγες και διακριτές τιμές), αλλιώς (πολλά χαρακτηριστικά ή πολλές/συνεχείς τιμές) τα πράγματα δυσκολεύουν.

- Η Δομή του δέντρου.

  – Επιθυμητό είναι να δημιουργούνται δέντρα που είναι ισορροπημένα και με τα λιγότερα επίπεδα (μικρότερο βάθος). Αυτό όμως δεν είναι πάντα εφικτό ούτε το υπολογιστικά φτηνότερο.

  – Μερικοί αλγόριθμοι δημιουργούν μόνο δυαδικά δέντρα.

- Κριτήρια τερματισμού

  – Η δημιουργία ενός δέντρου σταματά οπωσδήποτε όταν όλα τα δεδομένα του (εναπομείναντος) συνόλου εκπαίδευσης κατηγοριοποιούνται πλήρως.

  – Μπορεί όμως να είναι απαραίτητο να σταματήσει νωρίτερα για να αποφευχθούν π.χ. μεγάλα δέντρα. Το πότε ή πού θα σταματήσει είναι θέμα συναλλαγής (trade-off) μεταξύ ακρίβειας (accuracy) και απόδοσης (performance) του αλγορίθμου.

  – Επίσης, πρώιμος τερματισμός μπορεί να γίνει για αποφυγή του φαινομένου της υπερπροσαρμογής (overfitting).

  – Τέλος, μπορεί να προχωρήσει σε μεγαλύτερα δέντρα αν είναι γνωστό ότι υπάρχουν κατηγορίες δεδομένων που δεν αντιπροσωπεύονται στο σύνολο εκπαίδευσης

- Δεδομένα εκπαίδευσης.

  – Η δομή ενός ΔΑ εξαρτάται από τα δεδομένα εκπαίδευσης. Αν το σύνολο εκπαίδευσης είναι πολύ μικρό, τότε το δέντρο μπορεί να μην είναι τόσο λεπτομερές, ώστε να ταξινομεί γενικότερα δεδομένα. Αν είναι πολύ μεγάλο, το δέντρο πιθανόν να υπερπροσαρμόζεται (overfits).

- Κλάδεμα (Pruning).

  – Μετά τη δημιουργία ενός ΔΑ μπορεί να χρειάζονται τροποποιήσεις για να βελτιώσουν την απόδοσή του, όπως π.χ. το κλάδεμα πλεοναζόντων συγκρίσεων ή υποδέντρων.

ΚΕΦΑΛΑΙΟ 4

ΠΕΙΡΑΜΑΤΙΚΗ ΜΕΛΕΤΗ ΤΗΣ ΜΕΘΟΔΟΥ

4.1 Πειράματα και ποσοστά επιτυχίας

Πραγματοποιήθηκαν πειράματα για κάθε συνδυασμό τελεστή LBP και ταξινομητή και παραμέτρων τους, που μελετάμε. Μπορούμε να κρίνουμε το αποτέλεσμα της ταξινόμησης για το σύνολο έλεγχου και να δούμε το ποσοστό επιτυχίας του BOWL Descriptor. Χρησιμοποιήθηκαν διαφορετικά σύνολα δεδομένων, όπως σύνολα πουλιών, λουλουδιών, οχημάτων κ.α Θα παραθέσουμε τα αποτελέσματα για ένα σύνολο οχημάτων και το διαχωρισμό τους στις κατηγορίες αεροπλάνο=0, αυτοκίνητο=1, ποδήλατο=2). Κάθε αντικείμενο αντιστοιχίζεται σε έναν αριθμό.

http://www.gti.ssr.upm.es/data/Vehicle_database.html

ΣΧΗΜΑ 4.1: Σετ δεδομένων

Παρατίθενται τα αποτελέσματα των συνδυασμών των αλγορίθμων και οι πινάκες αντίστοιχα.

- Αποτελεσμάτα χωρίς χρήση LBP.

ΠΙΝΑΚΑΣ 4.1: NOLBP

Ταξινομητής	Mean	Min	Std	Max
SVM	0.34	0.2	0.17	0.8
KNN	0.45	0.2	0.09	0.6
NB	0.38	0.06	0.16	0.6
DT	0.42	0.06	0.14	0.6

- Αποτελεσμάτα LBP

ΠΙΝΑΚΑΣ 4.2: LBP

Ταξινομητής	Mean	Min	Std	Max
SVM	0.35	0.3	0.09	0.6
KNN	0.46	0.2	0.04	0.7
NB	0.44	0.2	0.13	0.6
DT	0.33	0.2	0.07	0.6

- Αποτελεσμάτα CirLBP για R=2 και P=16

ΠΙΝΑΚΑΣ 4.3: CircLBP

Ταξινομητής	Mean	Min	Std	Max
SVM	0.52	0.2	0.13	0.6
KNN	0.48	0.2	0.07	0.5
NB	0.56	0.2	0.06	0.46
DT	0.43	0.2	0.07	0.5

- Αποτελεσμάτα Uniform LBP για R=2 και P=16

ΠΙΝΑΚΑΣ 4.4: UniLBP

Ταξινομητής	Mean	Min	Std	Max
SVM	0.54	0.2	0.14	0.6
KNN	0.50	0.2	0.09	0.7
NB	0.47	0.2	0.07	0.8
DT	0.46	0.2	0.18	0.5

- Αποτελεσμάτα ILBP

ΠΙΝΑΚΑΣ 4.5: IBLP

Ταξινομητής	Mean	Min	Std	Max
SVM	0.40	0.2	0.04	0.6
KNN	0.56	0.2	0.13	0.5
NB	0.54	0.2	0.13	0.6
DT	0.30	0.2	0.07	0.5

- Αποτελεσμάτα MLBP

Ταξινομητής	Mean	Min	Std	Max
SVM	0.40	0.3	0.9	0.5
KNN	0.56	0.1	0.11	0.7
NB	0.46	0.1	0.14	0.6
DT	0.48	0.1	0.11	0.5

ΠΙΝΑΚΑΣ 4.6: MBLP

- Αποτελεσμάτα SFLBP, για τη συνάρτηση: $s(x) = e^{\frac{-(i_n - i_c)^2}{2\sigma^2}}$

Ταξινομητής	Mean	Min	Std	Max
SVM	0.30	0.3	0.08	0.56
KNN	0.55	0.3	0.12	0.6
NB	0.52	0.3	0.14	0.7
DT	0.52	0.2	0.11	0.6

ΠΙΝΑΚΑΣ 4.7: SFLBP1

- Αποτελεσμάτα SFLBP, για τη συνάρτηση: $1 - s(x) = e^{\frac{-(i_n - i_c)^2}{2\sigma^2}}$

Ταξινομητής	Mean	Min	Std	Max
SVM	0.32	0.2	0.08	0.53
KNN	0.35	0.2	0.12	0.53
NB	0.30	0.0	0.14	0.4
DT	0.37	0.1	0.11	0.6

ΠΙΝΑΚΑΣ 4.8: SFLBP2

Ο αλγόριθμος SVM χρησιμοποιήθηκε με kernel='rbf' και ο KNN με αριθμό γειτόνων=1.

ΚΕΦΑΛΑΙΟ 5

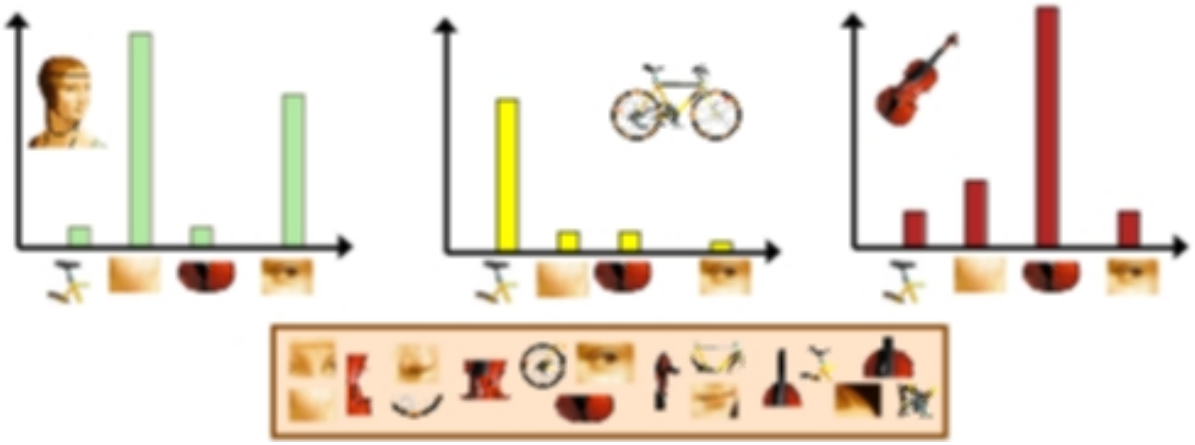
ΣΥΜΠΕΡΑΣΜΑΤΑ

5.1 Μειονεκτήματα

Η ταξινόμηση αποτελεί ένα από τα πιο κλασικά προβλήματα στην Επιστήμη των υπολογιστών Ο περιγραφέας BOWL, για να υλοποιηθεί έχει αρκετά στάδια, από τη δειγματοληψία μέχρι τη χρήση αλγορίθμων ταξινόμηση. Κάθε ένα από τα στάδια αυτά, είναι αμφίβολο ότι είναι αποδοτικό πλήρως, καθώς τα αποτελέσματα που λαμβάνουμε δεν είναι ξεκάθαρο, ότι είναι και τα σωστά, όποτε στο τελικό αποτέλεσμα μετράμε άπλα ένα ποσοστό επιτυχίας. Σε κάθε στάδιο θα δούμε τα προβλήματα που αντιμετωπίζουμε κάθε φορά:

- Δειγματοληψία.

  Κάθε pixel που επιλέγεται με τη μέθοδο έχει την ίδια πιθανότητα επιλογής, κάτι που δυσκολεύει την κατάσταση. Θα ήταν χρήσιμο να βρεθεί μια διαδικασία που να επι-



ΣΧΗΜΑ 5.1: Κατασκεύη Λεξικού

λεγεί χρήσιμα pixel, δηλαδή αυτά που ανήκουν πάνω σε γνωρίσματα κρίσιμα για την ταυτοποίηση. Η τοπική πληροφορία είναι πολύ χρήσιμη για τη δειγματοληψία, άλλα δεν υπάρχει συσχέτιση της επιλογής του pixel και της περιοχές υποψηφίας λέξης με την τοπικότητα της εικόνας. Ένας σημαντικός παράγοντας είναι και το πόσα σημεία θα πάρουμε για δειγματοληψία και το μέγεθος του λεξικού.

- Κατασκευή Λεξικού.

  Στον K-means,η ποιότητα της τελική λύσης εξαρτάται πολύ από το αρχικό σύνολο ομαδών, δειγμάτων και είναι πολύ χαμηλότερη από την πραγματική λύση.Μια κακή δειγματοληψία μεταφέρει το πρόβλημα και στον K-means. Επίσης ένα

- Επίσης οι αλγόριθμοι ταξινόμησης για να επιστρέψουν τη σωστή λύση χρειάζονται ως είσοδο σημαντικά γνωρίσματα στα ιστογράμματα. Τα δυο παραπάνω προβλήματα μεταδίδονται και εδώ. Οι αλγόριθμοι αυτοί γενικά παρουσιάζουν ένα ποσοστό επιτυχίας στην ταξινόμηση.Η σύσταση των σύνολων εκπαίδευσης και έλεγχου καθορίζουν τη συμπεριφορά τους.

5.2 Πλεονεκτήματα

Όπως είδαμε και στα πειράματα η μέθοδος έχει αρκετά ικανοποιητικά αποτελέσματα, με ποσοστό επιτυχίας πάνω από 50%,σε κάποιες περιπτώσεις. Αυτό συμβαίνει λόγω της χρήσης των LBP φίλτρων, οπού πια χρήσιμα γνωρίσματα τονίζονται στις εικόνας. Όπως έχει προαναφερθεί τα LBP παραμένουν αμετάβλητα σε αλλαγές της κλίμακας του γκρί. Σημαντική βοήθεια παρέχει και η αντιστοιχία των εικόνων σε κωδικολέξεις, για να έχουμε συχνότητες εμφάνισης γνωρισμάτων,έτσι ο ταξινομητής ενισχύεται με σημαντικούς συντελεστές.Τα ποσοστά επιτυχίας είναι στην πλειοψηφία χαμηλότερα χωρίς τη χρήση LBP στον BOW Descriptor.

5.3 Προγράμματα και Εργαλεία για την Υλοποίηση του Αλγορίθμου

Για την υλοποίηση της εργασίας:

- Η εργασία υλοποιήθηκε σε LINUX MINT 17 λειτουργικό σύστημα.
- Η συγγραφή των κείμενων έγινε με τη χρήση του εργαλείου XeTeX. Το XeTex είναι ένα πρόγραμμα ηλεκτρονικής στοιχειοθεσίας κειμένου, σε IDE Kile.
- Ολοι οι αλγοριθμοι , εκτος απο αυτον που αντιστοιχιζει την εικονα σε οπτικες λεχεις για καθε pixel της,ο οποιος γράφτηκε σε περιβάλλον αριθμητικής υπολογιστικής και Matlab, γραφτηκαν σε γλωσσα προγραμματισμού Python 2.7. Χρησιμοποιηθηκαν οι βιβλιοθηκες:
  - OpenCV2.4.9: Η OpenCV είναι μια βιβλιοθήκη μηχανικής όρασης ανοιχτού λογισμικού. Ξεκινησε από την Intel και περιεχει πανω απο 500 συναρτησεις για την επεξεργασία εικόνας. Υποστηριζει πολλά λειτουργικά συστήματα και έχει διεπαφές με πολλές γλώσσες προγραμματισμού. Χρησιμοποιήθηκε εδώ, για την αναπαράσταση λειτουργιών εικόνας, επίσης παρέχεται και ο K-means. Δουλέψαμε με τη διεπαφή της για Python.

- Η βιβλιοθήκη NumPy (Numeric Python) η οποία παρέχει τον τύπο ndarray για τη διανυσματική έκφραση σε πινάκες των εικόνων και των γνωρισμάτων που μελετάμε.
- Sklearn (Machine Learning in Python), είναι ένα απλό και αποτελεσματικό εργαλείο για εξόρυξη δεδομένων και ανάλυση. Οι αλγόριθμοι KNN, Naive Bayes, Desicion Tree κ.α παρέχονται από τη βιβλιοθήκη.
- Η Pyplot βιβλιοθήκη γραφικών για την εμφάνιση ιστογραμμάτων κ.α.
- Οι εικόνες αντλήθηκαν από τις βάσεις δεδομένων:
  * http://www-cvr.ai.uiuc.edu/ponce_grp/data/
  * http://cvcl.mit.edu/database.htm
  * http://vision.stanford.edu/lijiali/event_dataset

- : Ο προγραμματισμός έγινε σε IDE Eclipse 3.8

ΒΙΒΛΙΟΓΡΑΦΙΑ

[1] Ojala, T., Pietikainen, M., Harwood, D.: A comparative study of texture measures with clas- sification based on feature distributions. Pattern Recognition 29(1) , 51–59 (1996)

[2] Banerji, S., Verma, A., Liu, C.: Novel color LBP descriptors for scene and image texture classification. In: 15th International Conference on Image Processing, Computer Vision, and Pattern Recognition, Las Vegas, Nevada, July 18-21, pp. 537–543 (2011)

[3] Yang, J., Jiang, Y., Hauptmann, A., Ngo, C.: Evaluating bag-of-visual-words representations in scene classification. In: Multimedia Information Retrieval, pp. 197–206 (2007)

[4] Lowe, D.: Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision 60(2), 91–110 (2004)

[5] Nowak, E., Jurie, F., Triggs, B.: Sampling strategies for bag-of-features image classification. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006. LNCS, vol. 3954, pp. 490–503. Springer, Heidelberg (2006)

[6] Sugata Banerji Atreyee Sinha, and Chengjun Liu :A New Bag of Words LBP (BoWL) Descriptor for Scene Image Classification, Department of Computer Science, New Jersey Institute of Technology, Newark, NJ 07102, USA

[7] Di Huang, Caifeng Shan, Mohsen Ardebilian, Liming Chen: Facial Image Analysis Based on Local Binary Patterns: A Survey

[8] Springer-Verlag, 2004. C. Shan, “Linear subspace learning for facial expression analysis,” in Machine Learning, I-Tech Education and Publishing, 2009.

[9] N. Sun, W. Zheng, C. Sun, C. Zou, and L, Zhao, “Gender classification based on boosting local binary pattern,” in Proc. Int. Symposium on Neural Networks (ISNN), 2006, pp. II: 194-201.

[10] 606-609. O. Lahdenoja, M. Laiho, and A. Paasio, “Reducing the feature vector length in local binary pattern based face recognition,” in Proc. IEEE Int. Conf. Image Processing (ICIP), 2005, pp. II: 914-917.

[11] Conf. Image Processing (ICIP), 2005, pp. II: 914-917. 0] C. Shan, S. Gong, and P. McOwan, “Conditional mutual information based boosting for facial expression recognition,” in Proc. British Machine Vision Conference (BMVC), 2005.

[12] D. Zhao, Z. Lin, and X. Tang, “Contextual distance for data perception,” in Proc. IEEE Int. Conf. Computer Vision (ICCV), 2007.

[13] Ευαγγελία Πιτουρά: Διαφάνειες μαθήματος εξόρυξη Δεδομένων, Πανεπιστήμιο Ιωαννίνων, Τμήμα Μηχανικών Η/Υ και Πληροφορικής

[14] P.-N. Tan, M.Steinbach, V. Kumar, «Introduction to Data Mining», Addison Wesley, 2006

[15] T. Ojala, M. Pietikäinen, and T. Maenpaa, “Multiresolution gray-scale and rotation invariant texture classification with local binary patterns,” IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 24, no. 7, pp. 971–987, 2002.

[16] T. Ahonen, A. Hadid, and M. Pietikäinen, “Face description with local binary patterns: application to face recognition”, IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 28, no. 12, pp. 2037–2041, 2006. C. Shan, S. Gong, and P. W. McOwan, “Facial expression recognition based on local binary patterns: a comprehensive study,” Image and Vision Computing, 2009.

[17] G. Zhao and M. Pietikäinen, “Experiments with facial expression recognition using spatiotemporal local binary patterns,” in Proc. Int. Conf. Multimedia and Expo (ICME), 2007, pp. 1091-1094.

ΒΙΟΓΡΑΦΙΚΟ

ΠΡΟΣΩΠΙΚΑ ΣΤΟΙΧΕΙΑ Ονοματεπώνυμο: Χαρίση Ίνα

Διεύθυνση Κατοικίας: Αγίας Μαρίνας 64Β, 45221, Ιωάννινα

Τηλέφωνο: 6978707533

Ηλεκτρονικό Ταχυδρομείο: inacharisi@hotmail.gr

ΕΚΠΑΙΔΕΥΣΗ

- Πτυχίο Πληροφορικής ,Τμήμα Πληροφορικής, Σχολή Θετικών Επιστημών, Πανεπιστήμιο Ιωαννίνων

ΞΕΝΕΣ ΓΛΩΣΣΕΣ

- Αγγλικά (FCE)

ΕΡΕΥΝΗΤΙΚΑ ΕΝΔΙΑΦΕΡΟΝΤΑ

- Επεξεργασία Εικόνας και Υπολογιστική Όραση
- Υπολογιστική Γεωμετρία

ΕΠΑΓΓΕΛΜΑΤΙΚΗ ΕΜΠΕΙΡΙΑ

- Natech Integrated IT Solutions A.E

- Plantech Consulting Services Ε.Π.Ε
- ΕΛ.ΣΤΑΤ (Ελληνική Στατιστική Αρχή)
- ΙΕΚ Ιωαννίνων
- Φροντιστήρια Γνώμονας
- Μασούτης Super Market
- Deltapost Α.Ε

ΣΕΜΙΝΑΡΙΑ

- Ε.Κ.Ε.Φ.Ε Δημόκριτος, Θερινό Σχολείο
- ΕΕΛ/ΛΑΚ (Ελεύθερο Λογισμικό/Λογισμικό Ανοιχτού Κώδικα)
- ANIMART Βιωματικό Σχολείο Τεχνών
- Μαθήματα Ζωγραφικής