Πανεπιστήμιο Ιωαννίνων
Σχολή Θετικών Επιστημών
Τμήμα Πληροφορικής

Μεταπτυχιακή Εργασία Ειδίκευσης

ΣΤΑΤΙΣΤΙΚΕΣ ΠΡΟΣΕΓΓΙΣΕΙΣ ΓΙΑ ΤΗΝ ΑΝΑΚΤΗΣΗ ΕΙΚΟΝΩΝ

ΚΩΝΣΤΑΝΤΙΝΟΣ ΒΑΛΑΣΟΥΛΗΣ

Ιωάννινα, Σεπτέμβριος 2004

Πρόλογος

Η παρούσα εργασία ασχολείται με Στατιστικές Προσεγγίσεις για Ανάκτηση Εικόνων. Στόχος της εργασίας είναι η μελέτη της απόδοσης που παρουσιάζουν διάφοροι στατιστικοί αλγόριθμοι σε προβλήματα 1. *ευρετηριοποίησης (indexing)* μεγάλων συνόλων εικόνων και 2. *ανάκτησης (image retrival και image search)* εικόνων από μεγάλες βάσεις.

Η παρούσα εργασία δεν θα είχε εκπονηθεί χωρίς την πολύτιμη βοήθεια του καθηγητή κ. Α. Λύκα τον οποίο θα ήθελα να ευχαριστήσω ιδιαιτέρως θερμά. Οι συμβουλές του, η κριτική του καθώς και η συνεχής καθοδήγησή του καθ' όλη τη διάρκεια της εκπόνησης της εργασίας υπήρξαν και παραμένουν πολύτιμο εφόδιο για μένα.

Επίσης θα ήθελα να ευχαριστήσω θερμά τον φίλο μου Κωνσταντίνο Κωνσταντινόπουλο για την πολύτιμη επιστημονική του συμβολή στην επίλυση διαφόρων προβλημάτων που αντιμετωπίστηκαν κατά τη διάρκεια της εκπόνησης της εργασίας. Τα σχόλια, οι παρατηρήσεις αλλά και η ηθική στήριξη που μου παρείχε έπαιξαν καταλυτικό ρόλο στην ολοκλήρωση της εργασίας.

Η εργασία αφιερώνεται στους αγαπημένους μου γονείς, Δημήτρη και Κατερίνα, για τη διαρκή τους στήριξη και κατανόηση.

Ιωάννινα, Σεπτέμβριος 2004
Κωνσταντίνος Βαλασούλης

Περιεχόμενα

Κεφάλαιο 1

Εισαγωγή

1.1 Εκτίμηση Συνάρτησης Πυκνότητας Πιθανότητας

Στο πεδίο της Μηχανικής Μάθησης έχουν προταθεί κατά καιρούς διάφοροι αλγόριθμοι μάθησης. Οι αλγόριθμοι αυτοί ταξινομούνται σε τρείς μεγάλες κατηγορίες: (α) Τεχνικές Μάθησης με Επίβλεψη (supervised learning) (b) Τεχνικές Μάθησης χωρίς Επίβλεψη (unsupervised learning) και (γ) Τεχνικές Ενισχυτικής Μάθησης (reinforcement learning). Στην παρούσα εργασία θα μας απασχολήσουν αποκλειστικά αλγόριθμοι που ταξινομούνται στην δεύτερη κατηγορία (Μαθηση χωρίς Επίβλεψη).

Κατά την μάθηση χωρίς επίβλεψη έχουμε να αντιμετωπίσουμε το εξής πρόβλημα. Δοθέντος ενός συνόλου δεδομένων X επιδιώκουμε να κατασκευάσουμε ένα μοντέλο περιγραφής αυτών. Οι πληροφορίες που αναζητούμε μέσω μιας τέτοιας περιγραφής αφορούν την μορφή ή την δομή των δεδομένων στο αντίστοιχο χώρο. Η κατανομή $p(x)$ από την οποία έχουν προέρθει τα δεδομένα X δίνει πλήρη περιγραφή του X για αυτό και η εκτίμηση συνάρτησης πυκνότητας πιθανότητας θεωρείται ως η πιο γενική τεχνική μάθησης χωρίς επίβλεψη.

Οι γνωστές μέθοδοι εκτίμησης συνάρτησης πυκνότητας πιθανότητας διακρίνονται σε δύο μεγάλες κατηγορίες: στις παραμετρικές και στις μη παραμετρικές [20,7]. Η βασική διαφορά μεταξύ των δύο είναι ότι οι πρώτες υποθέτουν ένα παραμετρικό μοντέλο για την άγνωστη κατανομή ενώ οι δεύτερες δεν υποθέτουν κάτι τέτοιο αλλά προσπαθούν να εκφράσουν την άγνωστη κατανομή απευθείας από τα δεδομένα. Θα αναφερθούμε μόνο στις παραμετρικές μεθόδους αφού μόνο αυτές θα μας απασχολήσουν στην συνέχεια της εργασίας.

1.1.1 Παραμετρικά Μοντέλα

Ένας τρόπος προσέγγισης του προβλήματος εκτίμησης άγνωστης κατανομής είναι να υποθέσουμε ότι η άγνωστη κατανομή είναι μια συγκεκριμένη συνάρτηση εξαρτώμενη

από ένα διάνυσμα παραμέτρων. Η εκτίμηση σ' αυτή την περίπτωση δεν είναι τίποτε άλλο παρά η εύρεση εκείνων των παραμέτρων[1] έτσι ώστε η συνάρτηση να "ταιριάξει" όσο το δυνατό καλύτερα στην κατανομή των δεδομένων. Επομένως υποθέτουμε ότι η $p(x)$ εξαρτάται από ένα διάνυσμα παραμέτρων Θ για το λόγο αυτό και θα την γράφουμε ως $p(x|\Theta)$.

Υπόθεση απλής κατανομής. Για την μορφή της $p(x|\Theta)$ μπορούμε μια υποθέσουμε μια από τις γνωστές συναρτήσεις πυκνότητας πιθανότητας. Η παραμετρική συνάρτηση στην οποία έχει δοθεί η μεγαλύτερη προσοχή από κάθε άλλη είναι η κανονική (ή Gaussian) κατανομή. Η δημοτικότητα της οφείλεται κυρίως στις καλές αναλυτικές και στατιστικές ιδιότητες που διαθέτει. Η κανονική κατανομή στην γενική μορφή έχει την ακόλουθη μορφή:

$$p(x|\mu,\Sigma) = \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right\}, \tag{1.1}$$

όπου μ είναι ένα d-διάστατο διάνυσμα που αναπαριστά το μέσο της κατανομής και Σ είναι ο $d \times d$ πίνακας συμμεταβλητότητας. Ο παράγοντας μπροστά στο εκθετικό μέρος της συνάρτησης εγγυάται ότι ισχύει $\int p(x|\mu,\Sigma)dx = 1$. Το μέσο μ και ο πίνακας Σ ορίζονται από τις σχέσεις:

$$\mu = E[x] = \int x p(x|\mu_j,\Sigma)dx, \tag{1.2}$$

$$\Sigma = E[(x-\mu)(x-\mu)^T] = \int (x-\mu)(x-\mu)^T p(x|\mu_j,\Sigma)dx. \tag{1.3}$$

Κάθε συνιστώσα μ_i του μέσου καθώς και κάθε στοιχείο σ_{ij} του Σ ορίζονται από τις σχέσεις:

$$\mu_i = E[x_i] = \int x_i p(x|\mu_j,\Sigma)dx, \tag{1.4}$$

$$\sigma_{ij} = E[(x_i-\mu_i)(x_j-\mu_j)^T] = \int (x_i-\mu_i)(x_j-\mu_j)^T p(x|\mu_j,\Sigma)dx. \tag{1.5}$$

Αν στο στοιχείο σ_{ij} το $i = j$, τότε η παράμετρος αναπαριστά διακύμανση της συνιστώσας i, ενώ διαφορετικά αναπαριστά συμμεταβλητότητα της συνιστώσας i με την j. Ένας συνηθισμένος συμβολισμός της κανονικής κατανομής είναι $N(\mu,\Sigma)$, ενώ προκειμένου να δηλώσουμε ότι η μεταβλητή x ακολουθεί την προηγούμενη κανονική κατανομή γράφουμε $x \sim N(\mu,\Sigma)$ και λέμε ότι η x ακολουθεί κανονική κατανομή με μέση τιμή μ και πίνακα συμμεταβλητότητας Σ.

[1] Η Μπεϋσιανή μάθηση δεν βρίσκει απλά τιμές παραμέτρων, άλλα μια κατανομή ως προς τις παραμέτρους που εκφράζει το πόσο καλά αναπαριστά τα δεδομένα η κάθε τιμή παραμέτρων.

Από τη σχέση (1.3) προκύπτει ότι ο Σ είναι πάντα συμμετρικός και θετικά ημιορισμένος πίνακας. Λόγω της συμμετρίας μπορεί να περιγραφεί με $d(d+1)/2$ ανεξάρτητους παραμέτρους, οπότε συμπεριλαμβανομένων και των d παραμέτρων για το μέσο η συνάρτηση καθορίζεται πλήρως από $d+d(d+1)/2$ παραμέτρους. Η παρακάτω ποσότητα

$$\Delta^2 = (x-\mu)^T\Sigma^{-1}(x-\mu), \tag{1.6}$$

η οποία εμφανίζεται στο εκθετικό μέρος της (1.1) ονομάζεται απόσταση Mahalanobis μεταξύ x και μ. Προκειμένου να κατανοήσουμε πώς κατανέμονται τα δείγματα μιας κανονικής κατανομής ας σκεφτούμε το εξής: Για σταθερή απόσταση Δ^2, όλα τα x ανήκουν σε μια υπερελλειψοειδή επιφάνεια που έχει ως κέντρο το μ ενώ το σχήμα της καθορίζεται από τον πίνακα Σ. Προφανώς για όλα τα x μιας τέτοιας επιφάνειας η τιμή της συνάρτησης είναι σταθερή, πράγμα που σημαίνει ότι τα πρότυπα που παράγονται με βάση την κατανομή (1.1) ομαδοποιούνται έτσι ώστε να σχηματίζουν υπερελλειψοειδείς πυρήνες.

Μερικές φορές είναι βολικότερο να χρησιμοποιήσουμε μια απλούστερη μορφή της πολυδιάστατης κανονικής κατανομής. Αν για παράδειγμα υποθέσουμε ότι δεν υπάρχει συμμεταβλητότητα μεταξύ των συνιστωσών του x, δηλαδή $\sigma_{ij}=0$ για κάθε $i \neq j$, τότε ο πίνακας συμμεταβλητότητας μετατρέπεται σε έναν διαγώνιο $\Sigma = diag(\sigma_1^2 \ldots \sigma_d^2)$. Με βάση την υπόθεση αυτή η κατανομή παίρνει τη μορφή

$$p(x|\mu,\Sigma) = \frac{1}{(2\pi)^{d/2}\sigma_1 \ldots \sigma_d} \exp\left\{-\frac{(x_1-\mu_1)^2}{2\sigma_1^2} \ldots - \frac{(x_d-\mu_d)^2}{2\sigma_d^2}\right\}, \tag{1.7}$$

ή

$$p(x|\mu,\Sigma) = \prod_{i=1}^{d} \frac{1}{(2\pi)^{1/2}\sigma_i} \exp\left\{-\frac{(x_i-\mu_i)^2}{2\sigma_i^2}\right\}. \tag{1.8}$$

Η παραπάνω μορφή καθορίζεται από $2d$ ανεξάρτητους παραμέτρους. Υποθέτοντας ότι οι διακυμάνσεις της κάθε συνιστώσας είναι ίσες μεταξύ τους, δηλαδή ισχύει $\sigma_i^2 = \sigma^2$ για όλα i, καταλήγουμε σε μια επιπλέον απλούστευση της σχέσης (1.8). Σε αυτή την απλουστευμένη μορφή της κατανομής ο αριθμός των παραμέτρων έχει μειωθεί στις $d+1$ και η κανονική κατανομή γράφεται στην μορφή:

$$p(x|\mu,\sigma^2) = \frac{1}{(2\pi\sigma)^{d/2}} \exp\left\{-\frac{||x-\mu||^2}{2\sigma^2}\right\}, \tag{1.9}$$

όπου με την νόρμα $||x-\mu||$ συμβολίζουμε την ευκλείδεια απόσταση των διανυσμάτων x και μ. Σε αυτή την περίπτωση για σταθερή απόσταση Mahalanobis τα διανύσματα x με ίσες τιμές πιθανότητας $p(x)$ ορίζουν μια υπερσφαίρα στο d-διάστατο χώρο, επομένως πρότυπα κατανεμημένα με βάση την σχέση (1.9) ομαδοποιούνται έτσι ώστε να σχηματίζουν υπερσφαίρες. Η απλουστευμένη αυτή μορφή της κανονικής κατανομής έχει τις λιγότερες παραμέτρους αλλά υστερεί προφανώς σε γενικότητα.

Μικτές Κατανομές. Μια μικτή κατανομή [21] ορίζεται ως μια ειδική περίπτωση γραμμικού συνδυασμού ενός πεπερασμένου αριθμού συναρτήσεων πυκνότητας πιθανότητας. Δηλαδή η πιθανότητα μιας τυχαίας μεταβλητής x που ακολουθεί μικτή κατανομή γράφεται ως άθροισμα συναρτήσεων πυκνότητας πιθανότητας με βάρη και στην γενική περίπτωση των M τέτοιων συναρτήσεων δίνεται από την ακόλουθη σχέση:

$$p(x|\Theta) = \sum_{j=1}^{M} \pi_j p(x|j, \theta_j). \tag{1.10}$$

Τον πυρήνα j τον ονομάζουμε συστατικό πυρήνα ή απλώς πυρήνα, ενώ την αντίστοιχη κατανομή $p(x|j, \theta_j)$ (που εξαρτάται από ένα διάνυσμα παραμέτρων θ_j) του μικτού μοντέλου την ονομάζουμε συστατική συνάρτηση πυκνότητας πιθανότητας της ολικής κατανομής. Το βάρος π_j αποτελεί παράμετρο που εκφράζει την εκ των προτέρων πιθανότητα σύμφωνα με την οποία η παραγωγή ενός δεδομένου οφείλεται στον συστατικό πυρήνα j. Το σύνολο των παραμέτρων της μικτής κατανομής είναι προφανώς $\Theta = \{(\pi_j, \theta_j), j = 1, \ldots, M\}$. Οι παράμετροι π_j δεν μπορούν να λάβουν αρνητικές τιμές και υπόκεινται στον εξής περιορισμό:

$$\sum_{j=1}^{M} \pi_j = 1. \tag{1.11}$$

Η συνάρτηση $p(x|j, \theta_j)$ εκφράζει την δεσμευμένη κατανομή βάσει της οποίας ο πυρήνας j παράγει το δεδομένο x. Προκειμένου να παράγουμε ένα πρότυπο που ακολουθεί μικτή κατανομή της μορφής (1.10) επιλέγουμε, καταρχήν, έναν πυρήνα j από το σύνολο των M πυρήνων με πιθανότητα π_j και στην συνέχεια παράγουμε το πρότυπο με βάση την συστατική κατανομή $p(x|j, \theta_j)$.

Είναι δυνατόν να υποθέσουμε μια μικτή κατανομή για την άγνωστη συνάρτηση πυκνότητας πιθανότητας ενός συνόλου δεδομένων. Όπως είδαμε στην περίπτωση των παραμετρικών μεθόδων η υπόθεση ήταν ότι το σύνολο των δεδομένων έχει παραχθεί από μια εκ των γνωστών συναρτήσεων πυκνότητας (π.χ. την κανονική κατανομή). Στην περίπτωση ενός μικτού μοντέλου η υπόθεση είναι πιο γενική λόγω του ότι το σύνολο δεδομένων λαμβάνεται ως ένα μίγμα συστατικών πληθυσμών ο καθένας εκ των οποίων σχετίζεται με μια συστατική κατανομή και την αντίστοιχη εκ των προτέρων πιθανότητα.

Είναι αξιοσημείωτο ότι στο μικτό μοντέλο η έννοια της εκ των προτέρων πιθανότητας και της συστατικής κατανομής ενός πυρήνα χρησιμοποιείται ακριβώς ανάλογα με την έννοια της εκ των προτέρων πιθανότητας και της δεσμευμένης κατανομής της κατηγορίας στο πρόβλημα ταξινόμησης. Ωστόσο υπάρχει μια σημαντική διαφορά που αφορά την φύση του προβλήματος[2]. Στο πρόβλημα ταξινόμησης τα πρότυπα είναι "χα-

[2]Επιπλέον της προφανούς διαφοράς, δηλαδή ότι η εκτίμηση πυκνότητας και η ταξινόμηση- ως τεχνικές μάθησης- έχουν διαφορετικούς στόχους.

ραχτηρισμένα" ως προς την κατηγορία που ανήκουν, πράγμα που αποτελεί σημαντικό πλεονέχτημα κατά την διαδικασία μάθησης. Όπως αναφέρθηκε προηγουμένως, μπορούμε να διαχωρίσουμε το σύνολο δεδομένων εκπαίδευσης σε τόσα υποσύνολα όσες είναι και οι κατηγορίες και στη συνέχεια να εκτιμήσουμε την υπό συνθήκη κατανομή της κάθε κατηγορίας χρησιμοποιώντας μόνο τα δεδομένα της. Αντιθέτως κατά την εκτίμηση πυκνότητας πιθανότητας με μικτό μοντέλο δεν γνωρίζουμε σε ποιον πυρήνα ανήκει κάθε δεδομένο και επομένως έχουμε ένα επιπρόσθετο πρόβλημα σχετικά με την αντιστοίχηση δεδομένων και πυρήνων.

Μια σημαντική ιδιότητα των μικτών μοντέλων είναι ότι με κατάλληλες επιλογές συστατικών συναρτήσεων κατανομής μπορούν να προσεγγίσουν οποιαδήποτε συνεχή κατανομή με οσοδήποτε ακρίβεια εφόσον χρησιμοποιηθεί επαρκής αριθμός πυρήνων [21].

Είναι ενδιαφέρον να δούμε τις πληροφορίες ομαδοποίησης που μπορεί να μας εξασφαλίσει η εκτίμηση συνάρτησης πυκνότητας πιθανότητας με μικτές κατανομές. Ας υποθέσουμε ότι με κάποια διαδικασία μάθησης έχουν καθοριστεί όλοι οι παράμετροι της μικτής κατανομής. Καταρχήν, η εκ των προτέρων πιθανότητα ενός πυρήνα εκφράζει την αναλογία του αριθμού των δεδομένων που παράγονται από τον πυρήνα αυτό σε σχέση με το σύνολο των δεδομένων. Επιπλέον μέσω των συστατικών κατανομών παίρνουμε πληροφορίες σχετικά με τα χαρακτηριστικά της κάθε ομάδας (π.χ. κέντρο, διακύμανση). Τέλος για ένα οποιοδήποτε δεδομένο x μπορούμε να υπολογίσουμε την εκ των υστέρων πιθανότητα να ανήκει σε ένα πυρήνα j κάνοντας χρήση του κανόνα του Bayes:

$$P(j|x,\Theta) = \frac{\pi_j p(x|j,\theta_j)}{\sum_{i=1}^{M} \pi_i p(x|i,\theta_i)}. \tag{1.12}$$

Οι εκ των υστέρων πιθανότητες ικανοποιούν την σχέση

$$\sum_{j=1}^{M} P(j|x;\Theta) = 1. \tag{1.13}$$

1.2 Μέθοδοι Εκτίμησης Παραμέτρων

1.2.1 Μέγιστη Πιθανοφάνεια

Στην παρούσα ενότητα θα παρουσιάσουμε μια μέθοδο εύρεσης κατάλληλων τιμών παραμέτρων για τα παραμετρικά μοντέλα και θα δούμε πώς εφαρμόζεται στην περίπτωση της κανονικής κατανομής.

Έστω ότι έχουμε αποφασίσει για το ποια θα είναι η παραμετρική συνάρτηση που θα χρησιμοποιήσουμε, αυτό που απομένει είναι να ορίσουμε τρόπους με τους οποίους θα βρίσκουμε κατάλληλες τιμές για τις παραμέτρους. Μια από τις πιο ευρέως χρησιμοποιούμενες μεθόδους είναι αυτή της μέγιστης πιθανοφάνειας. Με βάση την μέθοδο

της μέγιστης πιθανοφάνειας αναζητούμε εκείνες τις τιμές των παραμέτρων οι οποίες μεγιστοποιούν μια συγκεκριμένη συνάρτηση, την οποία ονομάζουμε συνάρτηση πιθανοφάνειας.

Υποθέτουμε ότι έχουμε στην διάθεσή μας ένα σύνολο δειγμάτων X, όπου κάθε στοιχείο $x \in X$ ανήκει στον d-διάστατο χώρο. Επιπλέον υποθέτουμε ότι τα στοιχεία του X έχουν παραχθεί ανεξάρτητα το ένα από το άλλο με βάση την κατανομή $p(x|\Theta)$. Η από κοινού συνάρτηση πυκνότητας πιθανότητας των δεδομένων δίνεται από την σχέση:

$$P(X|\Theta) = \prod_{x \in X} p(x|\Theta). \tag{1.14}$$

Η $P(X|\Theta)$ όταν λαμβάνεται ως συνάρτηση των παραμέτρων Θ ονομάζεται πιθανοφάνεια του συνόλου δεδομένων X. Ο εκτιμητής μέγιστης πιθανοφάνειας είναι εξ ορισμού εκείνο το διάνυσμα παραμέτρων $\hat{\Theta}$ για το οποίο μεγιστοποιείται η πιθανοφάνεια. Μεγιστοποιώντας την ποσότητα (1.14) φαίνεται λογικό ότι η $p(x|\Theta)$ θα ταιριάξει όσο το δυνατό καλύτερα στην άγνωστη κατανομή των δεδομένων (ακριβέστερα στα δεδομένα X). Για αναλυτικούς κυρίως λόγους προκειμένου να βρούμε τον εκτιμητή μέγιστης πιθανοφάνειας είναι βολικότερο να εργαστούμε με το λογάριθμο της σχέσης (1.14). Λόγω του ότι ο λογάριθμος είναι γνησίως μονότονη (αύξουσα) συνάρτηση το μέγιστο της λογαριθμικής πιθανοφάνειας είναι συγχρόνως και το μέγιστο της πιθανόφανειας. Η λογαριθμική πιθανοφάνεια έχει την παρακάτω μορφή:

$$L(\Theta) = \log P(X|\Theta) = \sum_{x \in X} \log p(x|\Theta). \tag{1.15}$$

Εφόσον η $L(\Theta)$ είναι παραγωγίσιμη συνάρτηση ως προς το διάνυσμα Θ ο εκτιμητής μέγιστης πιθανοφάνειας $\hat{\Theta}$ πρέπει να είναι στάσιμο σημείο της (1.15), δηλαδή να αποτελεί λύση της εξίσωσης

$$\nabla_{\Theta} L(\hat{\Theta}) = 0, \tag{1.16}$$

όπου με ∇_{Θ} συμβολίζουμε τον τελεστή παραγώγισης ως προς το διάνυσμα Θ. Στις περισσότερες των περιπτώσεων την παραπάνω εξίσωση δεν μπορούμε να την λύσουμε αναλυτικά πράγμα που σημαίνει ότι απαιτείται να καταφύγουμε σε κάποια μέθοδο βελτιστοποίησης προκειμένου να προσεγγίσουμε τον εκτιμητή μέγιστης πιθανοφάνειας. Ωστόσο επιλέγοντας ένα πλήθος κατανομών για την $p(x|\Theta)$ όπως, π.χ. αυτές που ανήκουν στην οικογένεια των εκθετικών κατανομών [20] μπορούμε να βρούμε τον εκτιμητή μέγιστης πιθανοφάνειας με ένα άμεσο τρόπο. Παρακάτω δείχνουμε μια τέτοια περίπτωση όπου η $p(x|\Theta)$ είναι η κανονική κατανομή.

Εφαρμογή σε μικτές κατανομές. Προηγουμένως παρουσιάσαμε τη μέθοδο της μέγιστης πιθανοφάνειας στην περίπτωση της κανονικής κατανομής. Η μέθοδος χρησιμοποιείται και στην περίπτωση των μικτών μοντέλων, ωστόσο δεν είναι δυνατόν να

επιτευχθεί μια άμεση αναλυτική λύση όπως στην περίπτωση των διάφορων απλών κατανομών.

Υποθέτουμε ότι έχουμε μια μικτή κανονική κατανομή $p(x|\Theta)$ η οποία ορίζεται από την σχέση (1.10) ενώ οι συστατικοί της πυρήνες προς το παρόν υποθέτουμε ότι μπορούν να έχουν οποιαδήποτε μορφή. Η λογαριθμική πιθανοφάνεια έχει την παρακάτω μορφή:

$$L(\Theta) = \log \prod_{x \in X} \sum_{j=1}^{M} \pi_j p(x|j, \theta_j) \sum_{x \in X} \log \sum_{j=1}^{M} \pi_j p(x|j, \theta_j). \tag{1.17}$$

Η $L(\Theta)$ για το συγκεκριμένο σύνολο X αποτελεί μια συνάρτηση του διανύσματος Θ. Η μεγιστοποίηση της (1.17) δεν είναι μια απλή διαδικασία όπως είναι στην περίπτωση των παραμετρικών μεθόδων. Η βασική δυσκολία συνίσταται στο ότι η συνάρτηση έχει υψηλή μη γραμμικότητα (λόγω του αθροίσματος μέσα στο λογάριθμο), και διαθέτει πολλά τοπικά μέγιστα πράγμα που σημαίνει ότι, αναζητώντας τον εκτιμητή μέγιστης πιθανοφάνειας μέσω ενός αλγορίθμου βελτιστοποίησης, είναι εύκολο να εγκλωβιστούμε σε ένα τοπικό μέγιστο. Εκτός των παραπάνω για την λογαριθμική πιθανοφάνεια μικτών κατανομών υπάρχουν διάφορα θεωρητικά ζητήματα σχετικά με την μοναδικότητα του εκτιμητή μέγιστης πιθανοφάνειας. Ειδικότερα λόγω των πολλών τοπικών μεγίστων ενδεχομένως το ολικό μέγιστο να προκύπτει για πολλά διαφορετικά διανύσματα παραμέτρων (που ορίζουν διαφορετικά μοντέλα), οπότε το βέλτιστο διάνυσμα δεν ορίζεται μοναδικά. Επίσης το πρόβλημα ύπαρξης μοναδικής λύσης προέρχεται και από την ίδια την κατανομή για το λόγο ότι μπορεί να μην είναι ταυτοποιήσιμη συνάρτηση[3]. Για τέτοιου είδους θεωρητικά ζητήματα ο αναγνώστης μπορεί να ανατρέξει στο [24].

Ο εκτιμητής μέγιστης πιθανοφάνειας αντιστοιχεί σε κάποιο από τα στάσιμα σημεία της συνάρτησης πιθανοφάνειας. Επομένως ως μια πρώτη προσέγγιση στο πρόβλημα καθορισμού του εκτιμητή μέγιστης πιθανοφάνειας μπορούμε να βρούμε το σύστημα εξισώσεων που ικανοποιεί. Όπως θα δούμε η μορφή των εξισώσεων αυτών δεν επιτρέπει μια άμεση λύση. Αν $\hat{\Theta}$ είναι στάσιμο σημείο της (1.17), τότε ικανοποιεί τις εξισώσεις:

$$\sum_{x \in X} P(j|x, \hat{\Theta}) \nabla_{\theta_j} \log p(x|j, \hat{\theta}_j) = 0, \tag{1.18}$$

για κάθε διάνυσμα θ_j και

$$\hat{\pi}_{jk} = \frac{1}{|X|} \sum_{x \in X} P(j|x, \hat{\Theta}). \tag{1.19}$$

για κάθε εκ των προτέρων πιθανότητα π_j. Από την παραπάνω μορφή των εξισώσεων είναι φανερό ότι δεν μπορεί να βρεθεί αναλυτική λύση για το διάνυσμα παραμέτρων. Η μορφή της εξίσωσης (1.18) εξαρτάται κάθε φορά από την επιλογή της συνάρτησης

[3]Μια παραμετρική κατανομή $p(x|\Theta)$ (ή και γενικότερα ένα παραμετρικό μοντέλο) είναι ταυτοποιήσιμη συνάρτηση αν για κάθε $\Theta_1 \neq \Theta_2$ υπάρχει τουλάχιστον ένα x τέτοιο ώστε $p(x|\Theta_1) \neq p(x|\Theta_2)$.

πυρήνα $p(x|j,\theta_j)$. Παρακάτω εμφανίζονται οι εξισώσεις που παίρνουμε από την (1.18) στη περίπτωση που η $p(x|j,\theta_j)$ είναι η κανονική κατανομή:

$$\hat{\mu}_j = \frac{\sum_{x \in X} P(j|x,\hat{\Theta})x}{\sum_{x \in X} P(j|x,\hat{\Theta})}, \tag{1.20}$$

$$\hat{\Sigma}_j = \frac{\sum_{x \in X} P(j|x,\hat{\Theta})(x-\hat{\mu}_j)(x-\hat{\mu}_j)^T}{\sum_{x \in X} P(j|x,\hat{\Theta})}. \tag{1.21}$$

Σημειωτέον ότι οι δεύτερες παράγωγοι της λογαριθμικής πιθανοφάνειας ως προς τις εκ των προτέρων πιθανότητες π_j δεν μπορούν να είναι θετικές:

$$\nabla_{\pi_j,\pi_i} L(\Theta) = -\frac{1}{\pi_j \pi_i} \sum_{x \in X} P(j|x,\Theta) P(i|x,\Theta) \leq 0. \tag{1.22}$$

Για τον λόγο αυτό ο Εισιανός πίνακας έχει αρνητικούς αριθμούς στη κύρια διαγώνιο και κατά συνέπεια δεν μπορεί να είναι θετικά ορισμένος. Αυτό έχει ως αποτέλεσμα να μη υπάρχει κανένα στάσιμο σημείο της λογαριθμικής πιθανοφάνειας που να είναι ελάχιστο, πράγμα που αποτελεί μια γενική ιδιότητα των μικτών κατανομών [25].

1.2.2 Μπεϋζιανή Μάθηση

Με την μέθοδο της μέγιστης πιθανοφάνειας αναζητούμε μια μοναδική λύση για το διάνυσμα παραμέτρων. Ωστόσο είναι δυνατόν τα δεδομένα X να αναπαριστώνται εξίσου καλά από διάφορες τιμές παραμέτρων και οι διαφορετικές τιμές παραμέτρων να δίνουν εναλλακτικές πιθανές ερμηνείες για την προέλευση των δεδομένων του X. Επομένως μια γενικότερη προσέγγιση εκτίμησης παραμέτρων είναι να βρούμε μια κατανομή εξαρτώμενη από τα δεδομένα, που να εκφράζει την καταλληλότητα της κάθε δυνατής τιμής των παραμέτρων. Κάτι τέτοιο επιτυγχάνεται με την Μπεϋζιανή μάθηση.

Στη Μπεϋζιανή μάθηση υποθέτουμε ότι η άγνωστη συνάρτηση πυκνότητας πιθανότητας έχει μια γνωστή παραμετρική μορφή $p(x|\Theta)$ όπου το διάνυσμα παραμέτρων Θ θεωρείται άγνωστο, και συγχρόνως αποτελεί τυχαία μεταβλητή. Μέρος της πληροφορίας μας για τις τιμές παραμέτρων Θ εκφράζεται μέσω μιας εκ των προτέρων κατανομής $P(\Theta)$, ενώ το υπόλοιπο μέρος της πληροφορίας προέρχεται από το σύνολο X (στη μέγιστη πιθανοφάνεια η πληροφορία προέρχονταν μόνο από το X) που υποτίθεται ότι έχει παραχθεί από την $p(x|\Theta)$. Αν τα δεδομένα X έχουν παραχθεί ανεξάρτητα μεταξύ τους, τότε η εκ των υστέρων κατανομή $p(\Theta|X)$ δίνεται από τον κανόνα του Bayes:

$$p(\Theta|X) = \frac{P(X|\Theta)p(\Theta)}{\int P(X|\Theta)p(\Theta)d\Theta} \tag{1.23}$$

Εφόσον έχει υπολογιστεί η εκ των υστέρων κατανομή, η εκτιμούμενη κατανομή θα δίνεται με βάση την σχέση

$$p(x|X) = \int p(x|\Theta)p(\Theta|X)d\Theta. \tag{1.24}$$

Η παραπάνω εκτίμηση εξαρτάται από την μορφή της κατανομής $p(\Theta|X)$. Όσο ο αριθμός των δεδομένων αυξάνει ο όρος της πιθανοφάνειας στη σχέση (1.23) γίνεται ισχυρότερος, ενώ καθώς ο αριθμός των δεδομένων τείνει στο άπειρο η $p(\Theta|X)$ είναι ανάλογη της $P(X|\Theta)$. Αν διαθέτουμε λίγα δεδομένα ο όρος $p(\Theta)$ επηρεάζει σημαντικά την λύση και τότε η Μπεϋζιανή μέθοδος ενδεχομένως να δώσει τελείως διαφορετική λύση από την μέγιστη πιθανοφάνεια. Λόγω του ότι η εφαρμογή της Μπεϋζιανής μεθόδου είναι δύσκολη υπολογιστικά (αφού απαιτεί την ολοκλήρωση ως προς Θ), πολλές φορές χρησιμοποιείται η προσέγγιση $p(x|X) \approx p(x|\Theta_{MAP})$, όπου Θ_{MAP} μεγιστοποιεί την $p(\Theta|X)$. Μια διαφορετική προσέγγιση υπολογισμού του ολοκληρώματος (1.24) είναι μέσω της μεθόδου ολοκλήρωσης Monte Carlo η οποία πρϋποθέτει την δειγματοληψία με βάση την κατανομή $p(\Theta|X)$ [26].

1.3 Βελτιστοποίηση Μέσω του Αλγορίθμου ΕΜ

Ο αλγόριθμος ΕΜ (Expectation-Maximization) [27] ορίζεται ως μια γενική διαδικασία μεγιστοποίησης λογαριθμικών πιθανοφανειών σε προβλήματα όπου κάποιες μεταβλητές δεν έχουν παρατηρηθεί (μη παρατηρήσιμες ή κρυμμένες μεταβλητές). Θα δώσουμε, καταρχήν, ένα γενικό ορισμό του αλγορίθμου και εν συνεχεία θα δούμε την μορφή που παίρνει στο πρόβλημα εκτίμησης πυκνότητας πιθανότητας υποθέτοντας μικτή κατανομή.

Η λειτουργία του ΕΜ βασίζεται στην σχέση μεταξύ δύο συνόλων. Το πρώτο σύνολο το ονομάζουμε ελλιπές σύνολο (incomplete set) και το δεύτερο πλήρες σύνολο (complete set). Ελλιπή σύνολα δεδομένων είναι συνήθως δείγματα δεδομένων που παίρνουμε από πειράματα ή στατιστικές μετρήσεις, για αυτό το λόγο και τέτοιου είδους σύνολα αποτελούν πραγματικά δεδομένα. Αντιθέτως πλήρη σύνολα δεδομένων είναι συνήθως υποθετικά σύνολα και εκφράζουν την μορφή που θα θέλαμε να έχουν τα δεδομένα μας σε ένα πείραμα. Ωστόσο στην πράξη μια τέτοια μορφή δεν είναι διαθέσιμη, δηλαδή τα σύνολα αυτά είναι μη παρατηρήσιμα.

Υποθέτουμε ότι έχουμε ένα ελλιπές σύνολο προτύπων X για το οποίο ορίζεται η από κοινού κατανομή $P(X|\Theta)$ η οποία εξαρτάται από το άγνωστο διάνυσμα παραμέτρων Θ. Υποθέτουμε επίσης ένα πλήρες σύνολο $Y = (X, Z)$ όπου Z είναι ένα σύνολο μη παρατηρήσιμων μεταβλητών. Η κατανομή $P(Y|\Theta)$ εξαρτάται από το ίδιο διάνυσμα παραμέτρων Θ. Οι δύο κατανομές, δηλαδή του ελλειπούς και του πλήρους συνόλου δεδομένων συνδέονται με την σχέση:

$$P(X|\Theta) = \int_Z P(X, Z|\Theta)dZ \tag{1.25}$$

Επίσης οι λογαριθμικές πιθανοφάνειες των δύο συνόλων είναι $L(\Theta) = \log P(X|\Theta)$ και $L_C(\Theta) = \log P(Y|\Theta)$, αντίστοιχα.

Το πρόβλημα μας είναι να βρούμε εκείνο το διάνυσμα παραμέτρων για το οποίο μεγιστοποιείται η λογαριθμική πιθανοφάνεια του ελλιπούς συνόλου. Ο αλγόριθμος ΕΜ προσπαθεί να μεγιστοποιήσει την ποσότητα αυτή (την $L(\Theta)$) αναδεικνύοντας την σχέση μεταξύ των δύο συνόλων. Συγκεκριμένα ο ΕΜ προσεγγίζει το πρόβλημα μεγιστοποίησης έμμεσα εφαρμόζοντας μια επαναληπτική διαδικασία για την λογαριθμική πιθανοφάνεια $L_C(\Theta)$ του πλήρους συνόλου. Επειδή όμως το σύνολο Y (συγκεκριμένα το Z) είναι μη παρατηρήσιμο και επομένως η λογαριθμική πιθανοφάνεια $L_C(\Theta)$ είναι ακαθόριστη, ο ΕΜ την λαμβάνει ως τυχαία μεταβλητή και υπολογίζει την αναμενόμενη τιμή της ως προς την κατανομή $P(Z|X,\Theta)$, όπου Θ λαμβάνει την τρέχουσα τιμή των παραμέτρων. Ειδικότερα εάν βρισκόμαστε στην $t+1$ επανάληψη του αλγορίθμου και το τρέχον διάνυσμα είναι το $\Theta^{(t)}$ η προηγούμενη ποσότητα ορίζεται ως εξής:

$$Q(\Theta;\Theta^{(t)}) = E[L_C(\Theta)|X,\Theta^{(t)}] = \int_Z L_C(\Theta)P(Z|X,\Theta^{(t)}), \tag{1.26}$$

όπου

$$P(Z|X,\Theta) = \frac{P(X,Z|\Theta)}{P(X|\Theta)}. \tag{1.27}$$

Κάθε επανάληψη του αλγόριθμου αποτελείται από δύο βήματα: το E-βήμα (Expectation-step) στο οποίο καθορίζεται η $Q(\Theta;\Theta^{(t)})$ και το M-βήμα
(Maximization-step) στο οποίο μεγιστοποιείται η ποσότητα αυτή ως προς το διάνυσμα παραμέτρων. Πιο συγκεκριμένα τα βήματα στην $t+1$ επανάληψη ορίζονται ως εξής:

E-βήμα: Υπολογισμός της ποσότητας $Q(\Theta;\Theta^{(t)})$.
M-βήμα: $\Theta^{(t+1)} = \arg\max_\Theta Q(\Theta;\Theta^{(t)})$.

Σύμφωνα με τις ιδιότητες του αλγορίθμου η λογαριθμική πιθανοφάνεια του ελλιπούς συνόλου δεν μειώνεται μετά από μια επανάληψη του αλγορίθμου, δηλαδή ισχύει:

$$L(\Theta^{(t+1)}) \geq L(\Theta^{(t)}). \tag{1.28}$$

Από τον τρόπο που ορίζεται ο αλγόριθμος δεν είναι ξεκάθαρο για το πώς ορίζεται το σύνολο των μη παρατηρήσιμων μεταβλητών Z και γιατί η μεγιστοποίηση της ποσότητας $Q(\Theta;\Theta^{(t)})$ σε κάθε επανάληψη έχει ως αποτέλεσμα την αύξηση της $L(\Theta)$. Για αυτά τα ζητήματα ο αναγνώστης μπορεί να ανατρέξει στο άρθρο εισαγωγής του αλγορίθμου [27] ή στο βιβλίο των McLachlan και Krishnan [23] το οποίο αναφέρεται αποκλειστικά στον αλγόριθμο αυτόν.

1.4 Βελτιστοποίηση Μέσω του Αλγορίθμου Greedy EM

Έχει αποδειχθεί θεωρητικά στο [19] ότι, υπό προϋποθέσεις, η εκπαίδευση ενός μικτού μοντέλου μεγιστοποιώντας την πιθανοφάνεια μπορεί να επιτευχθεί με ένα αυξητικό τρόπο, προσθέτοντας διαδοχικά πυρήνες στο μοντέλο. Συγκεκριμένα, υποθέτουμε ότι ένας νέος πυρήνας $\phi(x;\theta)$ προστίθεται σε ένα μικτό μοντέλο $f_k(x)$ με k πυρήνες για τη δημιουργία του μοντέλου με $k+1$ πυρήνες:

$$f_{k+1}(x) = (1-\alpha) f_k(x) + \alpha\phi(x;\theta) \tag{1.29}$$

όπου $\alpha \in (0,1)$. Αν για κάθε k, δοθέντος του $f_k(x)$, το βάρος α και το διάνυσμα παραμέτρων θ του $\phi(x;\theta)$ επιλέγονται βέλτιστα έτσι ώστε η νέα λογαριθμική πιθανοφάνεια

$$L_{k+1} = \sum_{i=1}^{n} \log f_{k+1}(x_i) = \sum_{i=1}^{n} \log[(1-\alpha) f_k(x_i) + \alpha\phi(x-I;\theta)] \tag{1.30}$$

να μεγιστοποιείται, τότε για μεγάλο k το τελικό μοντέλο έχει λογαριθμική πιθανοφάνεια σχεδόν τουλάχιστον τόσο μεγάλη όσο κάθε μικτή κατανομή του τύπου

$$f_k(x) = \sum_{j=1}^{n} \pi_j \phi(x;\theta). \tag{1.31}$$

Δηλαδή για κάθε μικτή κατανομή και σύνολο δεδομένων, υπάρχει ένας αριθμός C τέτοιος ώστε η λογαριθμική πιθανοφάνεια που επιτυγχάνεται με τον αυξητικό αλγόριθμο είναι το πολύ C/k μικρότερη από την λογαριθμική πιθανοφάνεια της μικτής κατανομής, όπως αποδεικνύεται στο [19]. Επιπλέον μια αξιοσημείωτη ιδιότητα αυτής της τεχνικής μεγιστοποίησης είναι ότι οι παράμετροι του $f_k(x)$ παραμένουν σταθερές κατά την μεγιστοποίηση της L_{k+1}.

Η σπουδαιότητα αυτού του αποτελέσματος είναι το ότι η μεγιστοποίηση της πιθανοφάνειας ενός Gaussian μικτού μοντέλου μπορεί να αντικατασταθεί από την επαναληπτική εκπαίδευση ενός μικτού μοντέλου f_{k+1} δύο στοιχείων, όπου το πρώτο στοιχείο είναι το παλιό μοντέλο f_k και το δεύτερο είναι ένας gaussian πυρήνας $\phi(x;\theta)$ όπου $\theta = [\mu, \Sigma]$ ο μέσος και ο πίνακας συμμεταβλητότητας του αντίστοιχα. Αυτό αποτελεί πλεονέκτημα από πρακτική άποψη, αφού ένα μικτό μοντέλο με δύο στοιχεία είναι πιο εύκολο να εκπαιδευτεί από ότι ένα πιο γενικό μοντέλο. Παρ' όλα αυτά χρειάζονται κατάλληλες τεχνικές αναζήτησης προκειμένου να προσδιοριστούν οι βέλτιστες παράμετροι α, μ, Σ,που μεγιστοποιούν την L_{k+1}.

Μια αποδοτική τεχνική που αντιμετωπίζει το παραπάνω πρόβλημα έχει προταθεί στο [4]. Η μέθοδος χρησιμοποιεί ένα συνδυασμό τοπικής και καθολικής αναζήτησης,

κάθε φορά που προστίθεται ένας νέος πυρήνας στο μικτό μοντέλο. Εφόσον πρέπει να εκπαιδευτεί ένα μικτό μοντέλο με δύο στοιχεία, γίνεται τοπική αναζήτηση με τον αλγόριθμο EM για να βρεθεί ένα μέγιστο της L_{k+1} ως προς τα α, μ και Σ, ενώ οι παράμετροι του $f_k(x)$ παραμένουν σταθεροί. Προκειμένου να εφαρμοστεί ο EM οι παράμετροι α, μ και Σ αρχικοποιούνται υλοποιώντας μια καθολική αναζήτηση στο χώρο των παραμέτρων. Από τη στιγμή που εκτιμηθούν οι παράμετροι του νέου πυρήνα και το βάρος α, εφαρμόζεται και πάλι ο EM για να μεγιστοποιηθεί η L_{k+1} ως προς όλες τις παραμέτρους του μοντέλου.

Συνοπτικά ο αλγόριθμος Greedy EM για εκπαίδευση Gaussian μικτά μοντέλα είναι ο εξής:

1. Αρχικοποίηση με έναν μόνο πυρήνα με $\mu = E[x]$ και $\Sigma = Cov(x)$.
2. Εφάσρμοσε τον αλγόριθμο EM με κριτήριο τερματισμού $|L_k^t / L_k^{t-1} - 1| < 1e-6$.
3. Αναζήτησε κατάλληλο καινούριο πυρήνα και αρχικοποίησε τον κατάλληλα.
4. Εκτέλεσε *Μερικό* EM.
5. Αν $L_{k+1} \leq L_k$ τερμάτισε τη διαδικασία, αλλιώς δέσμευσε τον νέο πυρήνα και πήγαινε στο βήμα 2.

Μιας και στην παρούσα εργασία ο αλγόριθμος Greedy EM για Gaussian Μικτές Κατανομές έχει χρησιμοποιηθεί κατά κόρον, μια πιο λεπτομερής περιγραφή του αλγορίθμου βρίσκεται στο Παράρτημα Α. Για επιπλέον λεπτομέρειες, ο αναγνώστης μπορεί να ανατρέξει στο άρθρο πρότασης του αλγορίθμου [4].

1.5 Ανασκόπηση της Εργασίας

Στο δεύτερο κεφάλαιο της εργασίας γίνεται μια περιγραφή των προβλημάτων που αποτέλεσαν αφορμή αυτής της εργασίας. Τα προβλήματα αυτά είναι η *Ευρετηριοποίηση (Indexing)* σε μεγάλες βάσεις με εικόνες καθώς και η αποδοτική *Αναζήτηση/Ανάχτηση (Searching-Retrieval)* εικόνων στις βάσεις αυτές. Επίσης στο επόμενο κεφάλαιο γίνεται λόγος για τις προσεγγίσεις που έχουν προταθεί μέχρι τώρα, καθώς και για την σκοπιά από την οποία είδαμε το πρόβλημα στην παρούσα εργασία.

Το τρίτο κεφάλαιο είναι αφιερωμένο στην λεπτομερή περιγραφή της διαδικασίας μοντελοποίησης των εικόνων, την οποία προτείνουμε και χρησιμοποιούμε στις πειραματικές εφαρμογές. Στο τέταρτο κεφάλαιο, γίνεται λόγος για αποστάσεις μεταξύ κατανομών και αποστάσεις μεταξύ μικτών κατανομών, ενώ στο πέμπτο κεφάλαιο περιγράφονται πειραματικές εφαρμογές της μεθόδου που πραγματοποιήθηκαν και τα αποτελέσματα αυτών.

Ανακεφαλαιώνοντας την προτεινόμενη μέθοδο, το έκτο και τελευταίο κεφάλαιο ολοκληρώνεται με προτάσεις για μελλοντική έρευνα.

Κεφάλαιο 2

Ευρετηριοποίηση και Ανάκτηση Εικόνων

Κατά την περίοδο άνθισης της τεχνολογίας των πολυμέσων, την προηγούμενη δεκαετία, ο κύριος όγκος της έρευνας και ανάπτυξης στο χώρο αφορούσε κυρίως προβλήματα επικοινωνίας και παρουσίασης των πολυμέσων. Τα τελευταία χρόνια, λόγω της τεράστιας αύξησης του όγκου πληροφορίας των πολυμέσων, έχουν προκύψει νέα προβλήματα που αφορούν κυρίως την αποδοτική οργάνωση αυτών σε μεγάλες βάσεις καθώς και την αποτελεσματική ανάκτησή τους από τις βάσεις αυτές. Παρόλο που τα παραπάνω προβλήματα υφίστανται για όλες τις μορφές μέσων (εικόνες, video, ήχο), το μεγαλύτερο κομμάτι της έρευνας που πραγματοποιείται σήμερα, εστιάζεται κυρίως στην επίλυση των παραπάνω προβλημάτων για εικόνες, και αυτό γιατί (α) ο μεγαλύτερος όγκος πληροφορίας σε πολυμέσα σήμερα αφορά εικόνες και (β) η επίλυση των παραπάνω προβλημάτων για εικόνες εικάζουμε ότι θα μπορέσουν να αναχθούν σε λύση για άλλες μορφές μέσων όπως το video.

2.1 Αναπαράσταση Εικόνων

Οι εικόνες αποτελούν μια *οπτική αναπαράσταση ενός ή περισσοτέρων αντικειμένων, μιας σκηνής, ενός ατόμου ή μιας αφηρημένης έννοιας σε μια επιφάνεια* [31]. Από τον παραπάνω ορισμό γίνεται αντιληπτό πως μια εικόνα μπορεί να εμπεριέχει ιδιαίτερα μεγάλη ποσότητα πληροφορίας η οποία είναι πολύ δύσκολο να ποσοτικοποιηθεί και να διαχειριστεί. Προκειμένου να οργανώσουμε αποδοτικά εικόνες σε μεγάλες βάσεις, απαραίτητη προϋπόθεση αποτελεί η αναπαράσταση αυτών με μοντέλα. Ένα μοντέλο δεν είναι παρά ένας τρόπος έκφρασης ενός μέρους της πληροφορίας που περικλείεται σε μια εικόνα, και προτιμάται διότι μπορεί κάποιος να τη διαχειριστεί πολύ πιο εύκολα. Σχεδόν πάντα, τα μοντέλα "χτίζονται" με βάση κάποια χαρακτηριστικά τα οποία εξά-

γονται από την εικόνα και εσωκλείουν πληροφορία για το περιεχόμενο της εικόνας. Συνηθισμένα χαρακτηριστικά αφορούν το *χρώμα*, *το σχήμα-περίγραμμα* , η *υφή* της εικόνας, τα αντικείμενα τα οποία περιέχονται στην εικόνα και άλλα. Αναπαριστώντας τις εικόνες με μοντέλα, μπορεί να "χάνεται" πληροφορία, διότι από κάτι πολύ σύνθετο (εικόνα) περνάμε σε κάτι απλοϊκό (μοντέλο), επιτυγχάνεται όμως η ποσοτικοποίηση της πληροφορίας, η οποία είναι απαραίτητη για την αντιμετώπιση δύο βασικών προβλημάτων που περιγράφονται στις επόμενες ενότητες.

2.1.1 Ευρετηριοποίηση σε Βάσεις με Εικόνες

Ως *ευρετηριοποίηση* (indexing) σε μια βάση δεδομένων ορίζεται *η διαδικασία κατηγοριοποίησης των καταχωρήσεων με απώτερο σκοπό την πιο εύκολη ανάκτησή τους*. Η μεγαλύτερη δυσκολία στην ευρετηριοποίηση μιας βάσης με εικόνες, εγγυάται στην *σαφή* περιγραφή του περιεχομένου των εικόνων της βάσης. Έχοντας αντιστοιχίσει ένα μοντέλο σε κάθε εικόνα της βάσης, ουσιαστικά έχει γίνει το πρώτο βήμα για τη δημιουργία ενός ευρετηρίου. Το μόνο που απομένει είναι να βρεθεί ένα μέτρο απόστασης με το οποίο θα επιτυγχάνεται σαφής διαχωρισμός μεταξύ των μοντέλων και κατ' επέκταση μεταξύ των εικόνων της βάσης. Συνήθως, τα μοντέλα τα οποία χρησιμοποιούνται στην πράξη είναι μαθηματικά μοντέλα για τα οποία εύκολα μπορούν να οριστούν μέτρα απόστασης. Με τη μοντελοποίηση λοιπόν των εικόνων το πρόβλημα της ευρετηριοποίησης μεγάλων βάσεων με εικόνες φαίνεται πως είναι δυνατόν να αντιμετωπιστεί.

2.1.2 Αναζήτηση και Ανάκτηση Εικόνων

Επιτυγχάνοντας την ευρετηριοποίηση βάσεων με εικόνες, μπορούμε να αντιμετωπίσουμε και παρεμφερή προβλήματα. Η ανάκτηση και αναζήτηση εικόνων σε "αποθήκες εικόνων" με βάση το περιεχόμενο κάποιας εικόνας-ερώτησης είναι ένα τέτοιο πρόβλημα το οποίο μπορεί να αντιμετωπιστεί. Με άλλα λόγια αυτό που επιθυμούμε είναι ένα μηχανισμό ο οποίος θα επιτρέπει στον χρήστη να ελέγχει αν μια εικόνα-ερώτηση υπάρχει σε μια βάση ή να βρίσκει άμεσα (γρήγορα) τις όμοιες, ως προς το περιεχόμενο, εικόνες που υπάρχουν στη βάση. Το πρόβλημα της ανάκτησης/αναζήτησης εικόνων σε μεγάλες βάσεις είναι ιδιαίτερα σύνθετο όσο και κρίσιμο. Εφαρμογές στις οποίες η ανάγκη για αποδοτική αναζήτηση/ανάκτηση εικόνων είναι επιτακτική περιλαμβάνουν από ιατρικά ψηφιακά αρχεία και εικόνες δορυφόρων έως λογότυπα και δεσμευμένα σήματα προϊόντων (trademarks).

2.2 Μερικές Υπάρχουσες Προσεγγίσεις

Όπως αναφέρθηκε και προηγουμένως, για τη μοντελοποίηση μιας εικόνας εξάγονται χαρακτηριστικά τα οποία εμπεριέχουν πληροφορία για το περιεχόμενο της εικόνας. Όμως, το περιεχόμενο μιας εικόνας μπορεί να είναι ιδιαιτέρως σύνθετο, με αποτέλεσμα ο τεράστιος όγκος της πληροφορίας που εξάγεται από την εικόνα να είναι αδύνατον να μοντελοποιηθεί αποδοτικά. Για τον παραπάνω λόγο, σχεδόν όλες οι μεθοδολογίες που έχουν προταθεί μέχρι σήμερα, δεν αντιμετωπίζουν το πρόβλημα γενικά αλλά προσπαθούν να το εξειδικεύσουν, περιορίζοντας το περιεχόμενο της πληροφορίας που εξάγουν από τις εικόνες. Για παράδειγμα, αντί να εξάγουν πληροφορία για το χρώμα, την υφή και το σχήμα-περίγραμμα της εικόνας, εξάγουν χαρακτηριστικά μόνο για ένα ή δύο από αυτά. Με τον παραπάνω τρόπο, η μοντελοποίηση των εικόνων γίνεται πιο αποδοτική με αποτέλεσμα και οι επιδόσεις των μεθόδων στα δύο παραπάνω προβλήματα να είναι μεγαλύτερες.

Από τις αρκετές προσεγγίσεις που έχουν προταθεί κατά καιρούς, θα εστιάσουμε το ενδιαφέρον μας σε αυτές που αντιμετωπίζουν το πρόβλημα μόνο ως προς το σχήμα-περίγραμμα των εικόνων μιας και έτσι αντιμετωπίσαμε το πρόβλημα και εμείς στην παρούσα εργασία.

Οι Adoram και Lew [12] έκαναν χρήση *snakes* για την περιγραφή καμπύλων που παρουσιάζουν αντικείμενα σε εικόνες προκειμένου να φτιάξουν ένα σύστημα ανάκτησης εικόνων με βάση το σχήμα-περίγραμμα. Η προσέγγισή τους μειονεκτεί όσο αφορά την περιγραφή κοιλοτήτων (concavities) στα σχήματα των εικόνων. Οι Gnsel και Tekalp [13] ορίζουν ένα μέτρο ομοιότητας μεταξύ σχημάτων το οποίο βασίζεται στα στοιχεία του πίνακα mismatch που προκύπτει από την ιδιοσχηματική ανάλυση (eigenshape decomposition) των εικόνων. Στο [14] προτείνονται πεπλεγμένα πολυώνυμα για την αποτελεσματική αναπαράσταση γεωμετρικών σχημάτων σε εικόνες. Τα πεπλεγμένα πολυώνυμα είναι εύρωστα και αναλλοίωτα χαρακτηριστικά για την περιγραφή ενός σχήματος. Η μέθοδος βασίζεται στην παρεμβολή πολυωνύμων στις καμπύλες που απαρτίζουν το σχήμα της εικόνας και κατόπιν το διάνυσμα των παραμέτρων των πολυωνύμων χρησιμοποιείται ως μέτρο απόστασης μεταξύ των εικόνων. Οι Alferez και Wang [15] πρότειναν μια μέθοδο για ευρετηριοποίηση εικόνων με βάση το περίγραμμα, αρκετά εύρωστη σε διάφορους μετασχηματισμούς η οποία βασίζεται στη χρήση splines και wavelets για την περιγραφή του σχήματος των αντικειμένων. Οι Petrakis και Milios [16] πρότειναν μια προσέγγιση για αναγνώριση σχημάτων σε διάφορα επίπεδα ανάλυσης με χρήση δυναμικού προγραμματισμού. Οι Sharvit et al [17] προτείνουν μια μέθοδο βασισμένη σε συμμετρίες για την αναπαράσταση σχημάτων σε εικόνες και χρησιμοποιούν γράφους συσχέτισης για να πραγματοποιούν συγκρίσεις μεταξύ εικόνων. Οι Rui et al [18] όρισαν έναν νέο Fourier Descriptor, ο οποίος αποτελεί μια παρεμβαλλόμενη κανονικοποιημένη μορφή των συντελεστών fourier χαμηλής συχνότητας, και τον χρησιμοποίησαν για περιγραφή του σχήματος των εικόνων. Έπειτα για το

"ταίριασμα" των εικόνων χρησιμοποίησαν διάφορες αποστάσεις (Euclidean, Chamfer και Hausdorff). Οι Huet και Hancock [1,9] πρότειναν την κατασκευή ιστογραμμάτων βασισμένη σε τοπικά χαρακτηριστικά που προκύπτουν από τη συσχέτιση ακμών που περιγράφουν το σχήμα μιας εικόνας, ενώ αργότερα εξέτασαν και τις επιδόσεις σχεσιακών γραφημάτων στην μοντελοποίηση των εικόνων [2,3]. Τέλος, ο P. Suganthan [6] προτείνει τη χρήση Self-Organizing Maps (SOMs) για την ευρετηριοποίηση (με βάση το σχήμα) εικόνων.

Κεφάλαιο 3

Μοντελοποίηση Εικόνας με Βάση τα Περιγράμματα

Όπως αναφέραμε και στο προηγούμενο κεφάλαιο, όλες οι μεθοδολογίες επίλυσης των προβλημάτων Ευρετηριοποίησης και Ανάκτησης Εικόνων βασίζονται στην *μοντελοποίηση* των εικόνων ή κάποιων βασικών χαρακτηριστικών τους. Στο κεφάλαιο αυτό περιγράφονται αναλυτικά τα βήματα της διαδικασίας που ακολουθείται ώστε η πληροφορία που αφορά το *σχήμα* μιας εικόνας να μοντελοποιηθεί από μια μικτή gaussian κατανομή (**Gaussian Mmixture Model**, GMM).

3.1 Προεπεξεργασία Εικόνας

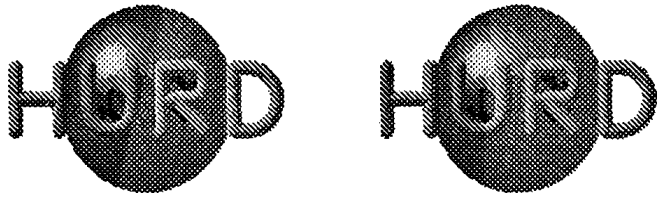
Έστω μια εικόνα **I** την οποία επιθυμούμε να μοντελοποιήσουμε. Η εικόνα **I** αρχικά υποβάλλεται σε προεπεξεργασία ώστε να εξαλειφθεί οποιαδήποτε άσχετη ως προς το *σχήμα* της εικόνας πληροφορία:

3.1.1 Μετατροπή από RGB σε Grayscale

Με χρήση της ρουτίνας **rgb2gray** [30] απομακρύνεται οποιαδήποτε πληροφορία σχετική με χροιά και κορεσμό (hue και saturation) ενώ διατηρείται η πληροφορία φωτεινότητας (luminance) που υπάρχει στην εικόνα. Στο Σχήμα 3.1 φαίνεται το αποτέλεσμα της εφαρμογής της ρουτίνας **rgb2gray** σε μια έγχρωμη εικόνα.

3.1.2 Εντοπισμός ακμών

Στη συνέχεια από τη grayscale εικόνα εξάγουμε τις ακμές που περιγράφουν το περίγραμμα της εικόνας (shape). Αυτό επιτυγχάνεται με την εφαρμογή της ρουτίνας



Σχήμα 3.1: (α) RGB εικόνα (β) Αποτέλεσμα της ρουτίνας rgb2gray στην εικόνα (α)

εντοπισμού ακμών **canny** [30]. Μετά την εφαρμογή της ρουτίνας εντοπισμού ακμών η εικόνα μας έχει μετατραπεί σε *δυαδική* (binary) εικόνα (Σχήμα 3.2).

Το επόμενο βήμα είναι να εφαρμόσουμε μια απλή ρουτίνα συνένωσης ακμών (edge linking). Αυτό κρίνεται απαραίτητο διότι (α) Ο αλγόριθμος canny δεν είναι τέλειος και σε αρκετές περιπτώσεις μπορεί σε κάποιο σημείο να εμφανίσει δύο ακμές οι οποίες αρχικά θα αναμέναμε να ήταν μία, (β) Όπως θα αναφέρουμε και στη συνέχεια, γενικά θα θέλαμε το πλήθος των ακμών να μην είναι υπερβολικά μεγάλο. Η ρουτίνα συνένωσης γραμμών που χρησιμοποιείται είναι η **edgelink** [29] και στο Σχήμα 3.3 φαίνεται το αποτέλεσμα της εφαρμογής της στην εικόνα του Σχήματος 3.2.

3.2 Εξαγωγή Χαρακτηριστικών

Με την παραπάνω διαδικασία προεπεξεργασίας σε μια εικόνα **I** έχουμε αντιστοιχίσει μια λίστα ακμών **L** οι οποίες περιγράφουν το περίγραμμα της εικόνας. Στη συνέχεια χρησιμοποιούμε τις ακμές της λίστας **L** για να παράγουμε ένα σύνολο από χαρακτηριστικά διανύσματα **S** τα οποία θα χαρακτηρίζουν την αρχική μας εικόνα. Τα εν λόγω διανύσματα χαρακτηριστικών προκύπτουν με την εξής διαδικασία [1,6]:

Για ένα αυθαίρετο ζεύγος ακμών στη λίστα **L** (Σχήμα 3.4) υπολογίζουμε το σημείο τομής τους **"i"** . Έπειτα ονομάζουμε το άκρο της πρώτης ακμής (η ακμή της οποίας το μέσο βρίσκεται πιο κοντά στο σημείο τομής **"i"**) το οποίο βρίσκεται πλησιέστερα στο σημείο τομής **"i"** ως **"a"** . Το άλλο άκρο της πρώτης ακμής το ονομάζουμε **"b"** και αντίστοιχα ονομάζουμε τα σημεία **"c"** και **"d"**. Μετά την παραπάνω διαδικασία οι εξής πέντε χαρακτηριστικές ποσότητες μπορούν να υπολογιστούν:

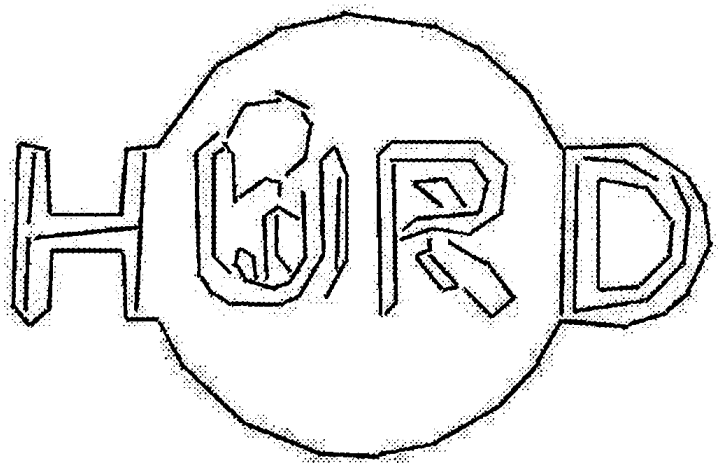
- $f_0 = \theta_{ab,cd} = arccos\{\frac{\vec{ab}\cdot\vec{cd}}{|\vec{ab}|\cdot|\vec{cd}|}\}$
- Λόγος σχετικής θέσης ακμών: $f_1 = \frac{1}{\frac{1}{2}+\frac{l_{il}}{l_{ab}}}$

Σχήμα 3.2: Αποτέλεσμα της ρουτίνας εντοπισμού ακμών canny.

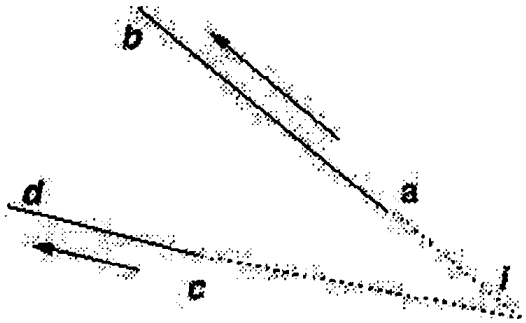
- Λόγος των μηκών των ακμών: $f_2 = \frac{\min(l_{ab},l_{cd})}{\max(l_{ab},l_{cd})}$
- Λόγος των άκρων των ακμών: $f_3 = \frac{\min(l_{ac},l_{bd})}{\max(l_{ac},l_{bd})}$
- Λόγος των σταυρωτών άκρων των ακμών: $f_4 = \frac{\min(l_{ad},l_{bc})}{\max(l_{ad},l_{bc})}$

Οι παραπάνω πέντε χαρακτηριστικές ποσότητες διαθέτουν μία πολύ σημαντική ιδιότητα. Παραμένουν αναλλοίωτες κατά την *περιστροφή, μετατόπιση* και *κλιμάκωση* των ακμών ab και cd [1,6]. Η γωνία $\theta_{ab,cd}$ παίρνει τιμές στο διάστημα $[0,\pi]$. Παρόλα αυτά, αν θεωρήσουμε την περιστροφή από το $\vec{ab}$ στο $\vec{cd}$ ως π.χ. σύμφωνη με αυτή των δεικτών του ρολογιού τότε υπολογίζοντας το διανυσματικό γινόμενο μεταξύ $\vec{ab}$ και $\vec{cd}$ η γωνία αυτομάτως παίρνει τιμές στο $[-\pi,\pi]$. Αυτό το κάνουμε προκειμένου να βελτιώσουμε την δυνατότητα διακριτοποίησης του χαρακτηριστικού f_0.

Εφαρμόζοντας την παραπάνω διαδικασία για κάθε πιθανό ζεύγος ακμών που προκύπτει από τη λίστα **L** , μπορούμε να παράγουμε ένα σύνολο **S** από διανύσματα χαρακτηριστικών τα οποία θα αντιστοιχούν στη λίστα ακμών **L** και κατ' επέκταση στην εικόνα **I**. Επειδή το περίγραμμα σε αρκετές εικόνες μπορεί να είναι ιδιαίτερα σύνθετο, και κατ' επέκταση η λίστα ακμών **L** μεγάλη, υπάρχει κίνδυνος το σύνολο των διανυσμάτων χαρακτηριστικών για κάποιες εικόνες να είναι τεράστιο. Αυτό γενικά δεν είναι επιθυμητό



Σχήμα 3.3: Οι ακμές που βρέθηκαν μετά το edge-linking.



Σχήμα 3.4: Αυθαίρετο ζεύγος ακμών και οι ονομασίες των άκρων.

μιας και όπως θα δούμε στη συνέχεια η μοντελοποίηση υπερβολικά μεγάλων συνόλων δεδομένων είναι πολύ απαιτητική σε χρόνο. Για να αποφύγουμε λοιπόν την παραπάνω δυσκολία, θεωρούμε μόνο εκείνα τα διανύσματα χαρακτηριστικών που αντιστοιχούν σε ζεύγη ακμών που προκύπτουν από τις *έξι* κοντινότερες ακμές ως προς κάθε ακμή, όπως προτείνεται στο [3]. Με αυτό τον τρόπο, σε μια εικόνα **I** αντιστοιχίζεται ένα σύνολο **S** 5D διανυσμάτων χαρακτηριστικών, για το οποίο ισχύει $|S| = 6 \cdot |L|$.

3.3 Μοντελοποίηση των χαρακτηριστικών με Gaussian Mixture Model

Επόμενο βήμα είναι να μοντελοποιήσουμε το σύνολο διανυσμάτων χαρακτηριστικών **S** που αντιστοιχεί στην εικόνα **I** με τη χρήση μιας μικτής gaussian κατανομής (Gaussian Mixture Model, **GMM**). Όπως είδαμε και στο πρώτο κεφάλαιο αν η κατανομή μιας τυχαίας μεταβλητής $x \in R^d$ είναι μια μίξη από k gaussians τότε η συνάρτηση πυκνότητάς της είναι:

$$f(x|\theta) = \sum_{j=1}^{k} \alpha_j \frac{1}{\sqrt{(2\pi)^d |\Sigma|}} exp\left\{-0.5(x-\mu_j)^T \Sigma_j^{-1}(x-\mu_j)\right\} \tag{3.1}$$

όπου για τις παραμέτρους $\theta = \{\alpha_j, \mu_j, \Sigma_j\}_{j=1}^k$ ισχύει ότι $\alpha_j > 0$, $\sum_{j=1}^k \alpha_j = 1$, $\mu_j \in R^d$ και Σ_j είναι ένας $d \times d$ θετικά ορισμένος πίνακας συμμεταβλητότητας. Στην περίπτωσή μας ισχύει ότι $d = 5$.

Δοθέντος του συνόλου διανυσμάτων χαρακτηριστικών **S** η *Maximum Likelihood* (**ML**) εκτίμηση των παραμέτρων θ θα είναι:

$$\theta_{ML} = arg\max_{\theta} L(S|\theta) \tag{3.2}$$

όπου $L(S|\theta) = \sum_{x_j \in S} \log f(x_j|\theta)$. Μια πρώτη προσέγγιση για την εύρεση του θ_{ML} θα ήταν να χρησιμοποιήσουμε τον αλγόριθμο *EM* που περιγράφηκε στην Παράγραφο 1.5. Παρόλα αυτά όπως είδαμε ο αλγόριθμος *EM* έχει ένα σημαντικό μειονέκτημα, την εξάρτησή του από τις αρχικές τιμές των παραμέτρων θ, το οποίο στην περίπτωσή μας είναι πολύ σημαντικό μιας και θα θέλαμε η μέθοδος εκπαίδευσης που χρησιμοποιούμε να δίνει πάντα το ίδιο **GMM** για ένα συγκεκριμένο σύνολο δεδομένων **S**.

Μια εναλλακτική λύση είναι να χρησιμοποιήσουμε τον αλγόριθμο *Greedy EM* για Gaussian Mixture Learning [4], στον οποίο έγινε αναφορά στο Κεφάλαιο 1 και που περιγράφεται λεπτομερώς στο Παράρτημα Α. Ο αλγόριθμος Greedy EM είναι αυξητικός, δηλαδή ξεκινά με ένα πυρήνα, και σταδιακά προσθέτει έναν πυρήνα σε κάθε επανάληψη μέχρι η μικτή κατανομή να φτάσει να αποτελείται από *k-max* πυρήνες. Σε αντίθεση με τον αλγόριθμο *EM* δεν εξαρτάται από τις αρχικές τιμές των παραμέτρων και δίνει σημαντικά καλύτερες λύσεις.

3.4 Μοντελοποίηση βάσης εικόνων με GMMs

Στις προηγούμενες ενότητες του κεφαλαίου είδαμε βήμα-βήμα τη διαδικασία μοντελοποίησης μιας εικόνας από ένα GMM. Για να μοντελοποιήσουμε μια εικόνα **I** την προεπεξεργαζόμαστε και αντιστοιχίζουμε σε αυτή μια λίστα από ακμές **L** οι οποίες περιγράφουν το *περίγραμμα* της. Στη συνέχεια χρησιμοποιούμε τη λίστα **L** για να παράγουμε ένα σύνολο από διανύσματα χαρακτηριστικών **S** τα οποία θα αντιστοιχούν στη λίστα **L** και κατ' επέκταση στην εικόνα I. Τέλος το σύνολο **S** μοντελοποιείται από ένα **GMM** το οποίο θεωρούμε ότι αντιστοιχεί στην εικόνα **I** από την οποία και ξεκινήσαμε. Δοθείσης τώρα μιας βάσης από **N** εικόνες, χρησιμοποιώντας την παραπάνω διαδικασία για κάθε εικόνα ξεχωριστά, μπορούμε να αντιστοιχίσουμε ένα GMM_i σε κάθε εικόνα $I_i, i = 1, ..., N$ που ανήκει στη βάση.

Κεφάλαιο 4

Αποστάσεις Μεταξύ Κατανομών

Έχοντας αντιστοιχίσει ένα **GMM** σε κάθε μια εικόνα στη βάση, μπορούμε να θεωρήσουμε πως η απόσταση μεταξύ δύο εικόνων αντιστοιχεί στην απόσταση μεταξύ δύο **GMMs** που μοντελοποιούν τις εικόνες αυτές. Επόμενο βήμα λοιπόν είναι να μελετήσουμε αποστάσεις μεταξύ κατανομών και πιο συγκεκριμένα την απόσταση μεταξύ **GMMs**. Στις ενότητες που ακολουθούν δίνεται ο ορισμός διαφόρων αποστάσεων μεταξύ μικτών κανονικών κατανομών οι οποίες χρησιμοποιήθηκαν κατά τη διάρκεια των εφαρμογών.

Για τους ορισμούς που δίνονται στις παρακάτω ενότητες ας θεωρήσουμε I_1 και I_2 δύο εικόνες και $S_1 = \{x_{11}, ..., x_{1n_1}\}, S_2 = \{x_{21}, ..., x_{2n_2}\}$ τα σύνολα διανυσμάτων χαρακτηριστικών (διάστασης d) που αντιστοιχούν σε αυτές. Επίσης, έστω f_1 και f_2 δύο συναρτήσεις πυκνότητας πιθανότητας οι οποίες αντιστοιχούν στα S_1 και S_2 αντίστοιχα.

4.1 Μέση Λογαριθμική Πιθανοφάνεια

Ορίζουμε ως απόσταση τη *Μη Συμμετρική Μέση Λογαριθμική Πιθανοφάνεια* (d_{ns}):

$$d_{ns}(S_1||f_2) = \frac{1}{n_1}\sum_{t=1}^{n_1} log f_2(x_{1t}) \tag{4.1}$$

4.1.1 Συμμετρική Μέση Λογαριθμική Πιθανοφάνεια

Ορίζουμε ως απόσταση τη *Συμμετρική Μέση Λογαριθμική Πιθανοφάνεια* (D_S):

$$D_S(f_1||f_2) = \frac{1}{n_1}\sum_{t=1}^{n_1} log f_2(x_{1t}) + \frac{1}{n_2}\sum_{t=1}^{n_2} log f_1(x_{2t}) \tag{4.2}$$

4.2 Απόσταση Kullback-Liebler

Από το [11] η *Μη Συμμετρική* εκδοχή της *Kullback-Liebler Απόστασης* (d_{KL}) έχει οριστεί ως εξής:

$$d_{KL}(f_1||f_2) = E_{f_1}(log\frac{f_1(x)}{f_2(x)}) \tag{4.3}$$

Η παραπάνω ποσότητα δεν είναι δυνατόν να υπολογιστεί αναλυτικά για Gaussian Mixtures αλλά μόνο προσεγγιστικά μέσω διαδικασίας *Monte-Carlo.*

4.2.1 Συμμετρική KL

Και πάλι από το [11] η *Συμμετρική* εκδοχή της *Kullback-Liebler Απόστασης* (D_{KL}) έχει οριστεί ως εξής:

$$\begin{aligned} D_{KL}(f_1, f_2) &= \frac{1}{2}(d_{KL}(f_1||f_2) + d_{KL}(f_2||f_1)) \\ &\cong \frac{1}{n_1}\sum_{t=1}^{n_1} log\frac{f_1(x_{1t})}{f_2(x_{1t})} + \frac{1}{n_2}\sum_{t=1}^{n_2} log\frac{f_2(x_{2t})}{f_1(x_{2t})} \end{aligned} \tag{4.4}$$

4.3 Απόσταση Chernoff

Από το [7] ως *Απόσταση Chernoff* μεταξύ δύο κανονικών κατανομών $N_1(M_1, \Sigma_1)$ και $N_2(M_2, \Sigma_2)$ ορίζεται η ποσότητα:

$$\begin{aligned} \mu(s) = \frac{s(1-s)}{2}(M_2 - M_1)^T[s\Sigma_1 &+ (1-s)\Sigma_2]^{-1}(M_2 - M_1) \\ &+ \frac{1}{2}\ln\frac{|s\Sigma_1 + (1-s)\Sigma_2|}{|\Sigma_1|^s|\Sigma_2|^{1-s}} \end{aligned} \tag{4.5}$$

με την παράμετρο s να παίρνει τιμές στο $[0, 1]$.

4.3.1 Απόσταση Battacharyya

Αν στην απόσταση *Chernoff* θέσουμε $s = 0.5$ τότε η απόσταση παίρνει την παρακάτω μορφή η οποία είναι ευρύτερα γνωστή ως απόσταση *Bhattacharyya.*

$$\mu(s) = \frac{1}{8}(M_2 - M_1)^T[\frac{\Sigma_1 + \Sigma_2}{2}]^{-1}(M_2 - M_1) + \frac{1}{2}\ln\frac{|\frac{\Sigma_1+\Sigma_2}{2}|}{\sqrt{|\Sigma_1||\Sigma_2|}} \tag{4.6}$$

Δυστυχώς η απόσταση *Chernoff* δεν μπορεί να υπολογιστεί αναλυτικά για Μικτές Κατανομές.

4.4 Κανονικοποιημένη Τετραγωνική Απόσταση μεταξύ δυο GMMs

Από το [28] η ποσότητα:

$$D_Q = 1 - \frac{\int_x (f_1 - f_2)^2 dx}{\int_x (f_1^2 + f_2^2) dx} = \frac{2\int_x (f_1 f_2) dx}{\int_x (f_1^2 + f_2^2) dx} \tag{4.7}$$

αποτελεί την κανονικοποιημένη τετραγωνική απόσταση μεταξύ δύο μικτών κανονικών κατανομών. Παίρνοντας τον αρνητικό λογάριθμο της παραπάνω ποσότητας έχουμε:

$$\begin{aligned}
D_Q &= -\log \frac{2\int_x (f_1 f_2) dx}{\int_x (f_1^2 + f_2^2) dx} && (4.8)\\
&= \log \int_x f_1^2 + f_2^2 dx - \log 2 \int_x (f_1 f_2) dx && (4.9)\\
&= \log\{\sum_{i=1}^{N}\sum_{j=1}^{N} \pi_i \pi_j \int_x \phi_{1i}\phi_{1j} dx + \sum_{i=1}^{K}\sum_{j=1}^{K} w_i w_j \int_x \phi_{2i}\phi_{2j} dx\}\\
&- \log 2 \sum_{i=1}^{N}\sum_{j=1}^{K} \pi_i w_j \int_x \phi_{1i}\phi_{2j} dx && (4.10)
\end{aligned}$$

όπου N και K το πλήθος των συνιστωσών στην πρώτη και τη δεύτερη μίξη αντίστοιχα, π_i και w_i το βάρος τη i-οστής συνιστώσας στην πρώτη και τη δεύτερη μίξη αντίστοιχα και ϕ_{1i} και ϕ_{2i} η i-οστή συνιστώσα συνιστώσα στην πρώτη και τη δεύτερη μίξη αντίστοιχα. Τα ολοκληρώματα στην Σχέση 4.10 μπορούν να υπολογιστούν αναλυτικά:

$$\begin{aligned}
\int_x \phi_{1i}\phi_{1j} dx &= (2\pi)^{-d/2} |\Sigma_i + \Sigma_j|^{-1/2}\\
&\exp \{-\frac{1}{2}(\mu_i - \mu_j)^T [\Sigma_i + \Sigma_j]^{-1} (\mu_i - \mu_j)\} && (4.11)\\
\int_x \phi_{2i}\phi_{2j} dx &= (2\pi)^{-d/2} |S_i + S_j|^{-1/2}\\
&\exp \{-\frac{1}{2}(m_i - m_j)^T [S_i + S_j]^{-1} (m_i - m_j)\} && (4.12)\\
\int_x \phi_{1i}\phi_{2j} dx &= (2\pi)^{-d/2} |\Sigma_i + S_j|^{-1/2}\\
&\exp \{-\frac{1}{2}(\mu_i - m_j)^T [\Sigma_i + S_j]^{-1} (\mu_i - m_j)\} && (4.13)
\end{aligned}$$

όπου μ_i, Σ_i και m_j, S_j το μέσο και ο πίνακας συμμεταβλητότητας της i-οστής συνιστώσας στην πρώτη και τη δεύτερη μίξη αντίστοιχα. Επομένως, με αντικατάσταση στη Σχέση 4.10 η απόσταση D_Q υπολογίζεται αναλυτικά. Είναι πολύ σημαντικό το ότι η απόσταση D_Q είναι *συμμετρική*, μπορεί να υπολογιστεί αναλυτικά χωρίς ιδιαίτερο υπολογιστικό κόστος καθώς επίσης και το ότι για τον υπολογισμό της δεν απαιτούνται τα δεδομένα των εικόνων S_1 και S_2.

4.5 Ο Ρόλος των Αποστάσεων στην Μέθοδο

Χρησιμοποιώντας τις παραπάνω ποσότητες ως μέτρο απόστασης μεταξύ των μικτών κατανομών που μοντελοποιούν εικόνες, επιτυγχάνουμε να συγκρίνουμε μικτές κατανομές αντί να συγκρίνουμε τις εικόνες μεταξύ τους. Αν για παράδειγμα, έχουμε μοντελοποιήσει όλες τις εικόνες μιας βάσης με *GMMs* και επιθυμούμε να ελέγξουμε εάν μια εικόνα-ερώτηση υπάρχει στη βάση μας ή αναζητούμε τη πιο όμοια προς αυτή στη βάση, τότε αρκεί να μοντελοποιήσουμε την εικόνα-ερώτηση με μια μικτή κατανομή και να υπολογίσουμε τις αποστάσεις του GMM της από όλα τα GMMs της βάσης. Η εικόνα της βάσης που αντιστοιχεί στο GMM που δίνει τη μικρότερη απόσταση από το GMM της εικόνας-ερώτησης αντιστοιχεί στην *πλησιέστερη*, ως προς το περίγραμμα, εικόνα της βάσης.

Μιας και η φιλοσοφία των αποστάσεων που ορίστηκαν στις προηγούμενες ενότητες δεν είναι πάντα η ίδια, έχει ιδιαίτερο ενδιαφέρον να μελετηθεί η απόδοση τους σε διάφορα προβλήματα. Επίσης, δεδομένου ότι στην πράξη το πρόβλημα της αναζήτησης μιας εικόνας σε μια βάση μπορεί να απαιτεί χιλιάδες συγκρίσεις εικόνων (υπολογισμούς αποστάσεων), η ανάγκη για ένα μέτρο σύγκρισης το οποίο δεν έχει ιδιαίτερες υπολογιστικές απαιτήσεις γίνεται επιτακτική.

Κεφάλαιο 5

Εφαρμογές

Στο κεφάλαιο αυτό παρουσιάζονται πειραματικές εφαρμογές που πραγματοποιήθηκαν προκειμένου να μελετηθεί η απόδοση της προτεινόμενης μεθόδου. Το κεφάλαιο χωρίζεται σε τρεις βασικές ενότητες. Στην Ενότητα 5.1 παρουσιάζεται λεπτομερώς η διαδικασία εφαρμογής και τα αποτελέσματα της μεθόδου σε *πραγματικές εικόνες* (ένα σύνολο από λογότυπα και trademarks) που βρέθηκαν στο διαδίκτυο, ενώ στις Ενότητες 5.2 και 5.3 παρουσιάζεται η αντίστοιχη διαδικασία για δύο βάσεις από εικόνες που κατασκευάστηκαν τεχνητά καθώς και τα αντίστοιχα πειραματικά αποτελέσματα.

5.1 Εφαρμογές σε Πραγματικές Εικόνες

5.1.1 Εικόνες που χρησιμοποιήθηκαν

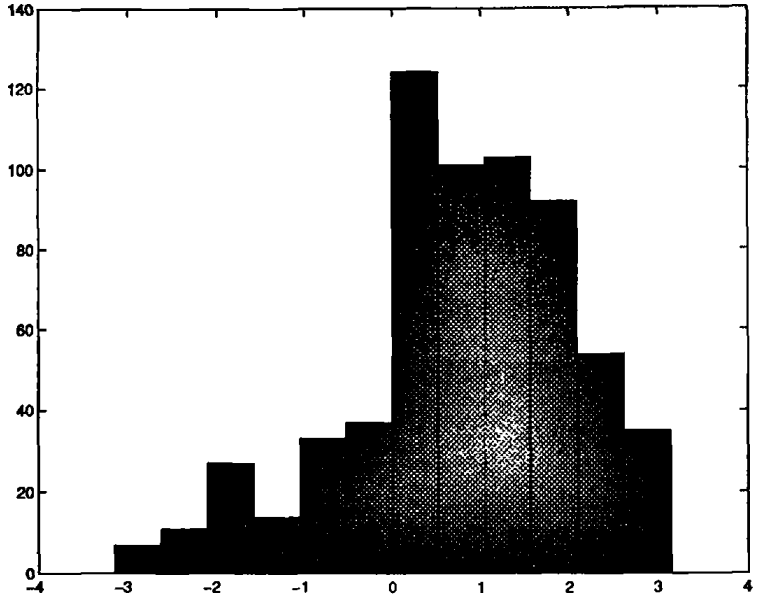
Για τις δύο εφαρμογές σε πραγματικές εικόνες χρησιμοποιήθηκε μια βάση από λογότυπα και trademarks που βρέθηκε στο διαδίκτυο [32], και έχει χρησιμοποιηθεί για την μελέτη της απόδοσης αντίστοιχων μεθόδων [1,6,9]. Η βάση αποτελείται συνολικά από 978 λογότυπα και trademarks δείγμα των οποίων φαίνεται στο Σχήμα 5.1. Στην πρώτη εφαρμογή χρησιμοποιήθηκαν 182 από τις 978 εικόνες της βάσης ενώ στη δεύτερη εφαρμογή και οι 978 εικόνες.

5.1.2 Διαδικασία και Παράμετροι Μοντελοποίησης

Εφαρμογή 1

Η διαδικασία μοντελοποίησης στα πείραμα με τις 182 εικόνες ήταν ακριβώς ίδια με αυτή που περιγράφτηκε στο Κεφάλαιο 3. Για κάθε εικόνα έγινε προεπεξεργασία και κατόπιν εξαγωγή ενός συνόλου διανυσμάτων χαρακτηριστικών (5D) τα οποία και μοντελοποιήθηκε με ένα Gaussian Mixture Model. Η εκπαίδευση του GMM έγινε με τον αλγόριθμο Greedy EM στον οποίο ο μέγιστος αριθμός παραγόντων ορίστηκε να είναι ίσος

Σχήμα 5.1: Μερικές από τα logos και trademarks που περιέχονται στη βάση που χρησιμοποιήθηκε.

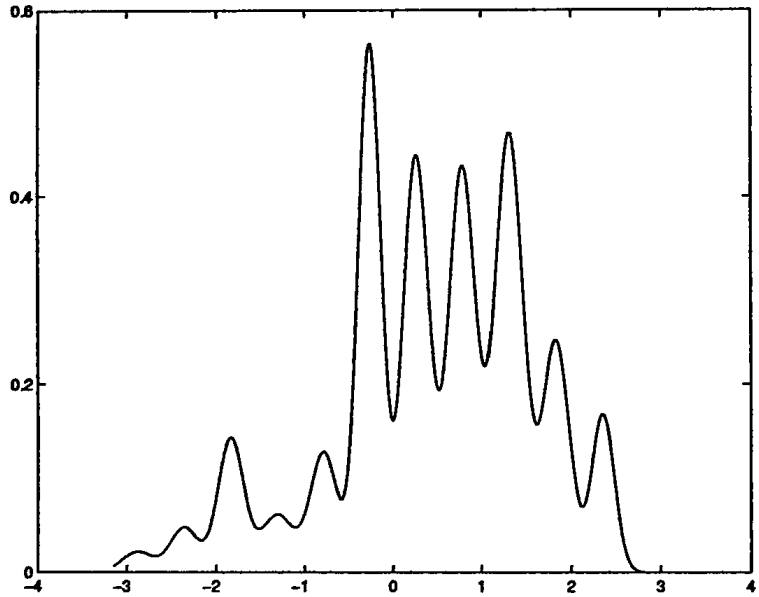


Σχήμα 5.2α: Ιστόγραμμα του χαρακτηριστικού f_0 της εικόνας 3.1α.

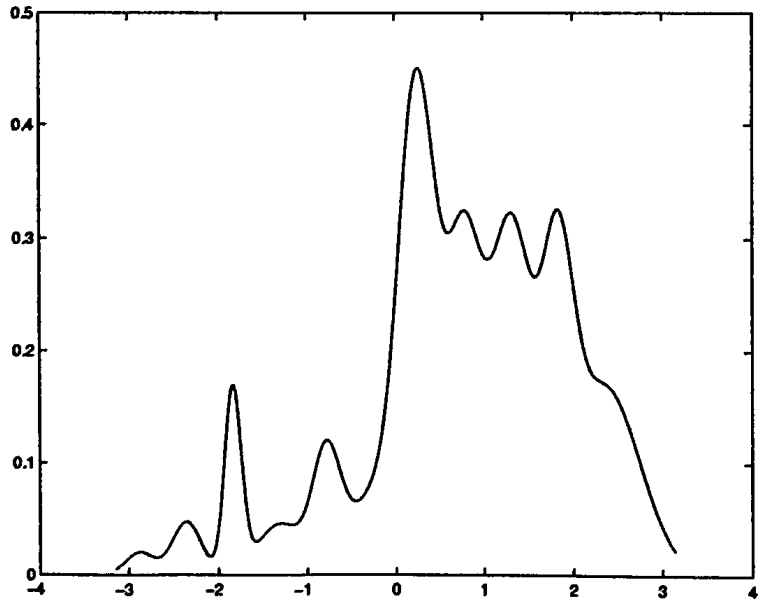
(kmax=6) και οι πίνακες συμμεταβλητότητας των πυρήνων διαγώνιοι.

Εφαρμογή 2

Στα πείραματα με τις 978 εικόνες πραγματοποιήθηκε εξαγωγή ενός μόνο χαρακτηριστικού, της γωνίας $\theta_{ab,cd} = arccos\{\frac{\vec{ab}\cdot\vec{cd}}{|\vec{ab}|\cdot|\vec{cd}|}\}$ (το σύνολο S_i αποτελείται από διανύσματα μήκους ένα). Κάθε εικόνα μοντελοποιήθηκε και πάλι με ένα GMM, η εκπαίδευση του οποίου έγινε με τον αλγόριθμο EM. Το κάθε GMM διαθέτει 12 ή 36 πυρήνες η αρχικοποίηση των οποίων έγινε με βάση το ιστόγραμμα του χαρακτηριστικού. Πιο συγκεκριμένα, για την αρχικοποίηση του EM με 12 πυρήνες κατασκευάζεται το ιστόγραμμα του χαρακτηριστικού με 12 bins και το μέσο και η διακύμανση του κάθε πυρήνα αρχικοποιείται στο μέσο του κάθε bin και τη διακύμανση των τιμών του S_i που αντιστοιχούν στο κάθε bin του ιστογράμματος αντίστοιχα. Επιπλέον, ως αρχική τιμή του βάρους κάθε πυρήνα ορίζεται το ύψος του αντίστοιχου bin, δηλαδή το πλήθος των τιμών του χαρακτηριστικού στο S_i που αντιστοιχούν στο bin προς το συνολικό πλήθος του συνόλου S_i. Στο Σχήμα 5.2α φαίνεται το ιστόγραμμα (με 12 bins) του συνόλου S της Εικόνας 3.1α και το Σχήμα 5.2β η σχηματική αναπαράσταση του GMM πριν την εφαρμογή του EM. Τέλος, στο Σχήμα 5.2γ φαίνεται η τελική μορφή που παίρνει το GMM μετά την εφαρμογή του EM.



Σχήμα 5.2β: Σχηματική αναπαράσταση του αρχικοποιημένου (με βάση το ιστόγραμμα) GMM πριν την εφαρμογή του EM.



Σχήμα 5.2γ: Σχηματική αναπαράσταση του GMM που μοντελοποιεί την εικόνα 3.1α.

44-1610.png

Σχήμα 5.3α: Εικόνα ερώτηση.

5.1.3 Μετρήσεις

Μετά την μοντελοποίηση των εικόνων και στις δύο εφαρμογές δοκιμάσαμε να μετρήσουμε την απόδοση της μεθόδου όσο αφορά την ευρετηριοποίηση των εικόνων κάθε βάσης.

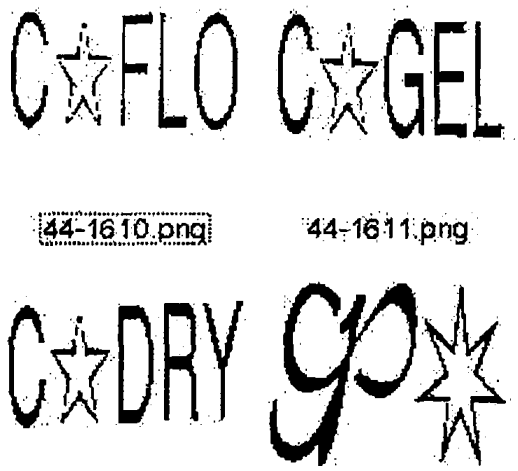
Εφαρμογή 1
Στην πρώτη εφαρμογή εκτελέστηκαν 182 διαφορετικές αναζητήσεις χρησιμοποιώντας κάθε φορά ως εικόνα-ερώτηση μια εικόνα που υπήρχε στη βάση. Σε όλες τις αναζητήσεις το σύστημα επέστρεψε πάντα ως πιο πιθανή τη "σωστή" εικόνα, ανεξάρτητα από το μέτρο απόστασης που χρησιμοποιήθηκε (d_{ns}, D_{KL}, D_S, D_Q). Στο Σχήμα 5.3α φαίνεται μια εικόνα-ερώτηση και στο Σχήμα 5.3β οι τέσσερις πιο "κοντινές" εικόνες που υπάρχουν στη βάση, σύμφωνα με το σύστημα.

Εφαρμογή 2
Αντίστοιχα στη δεύτερη εφαρμογή εκτελέστηκαν 978 διαφορετικές αναζητήσεις χρησιμοποιώντας κάθε φορά ως εικόνα-ερώτηση μια εικόνα που υπήρχε στη βάση. Και στις 978 αναζητήσεις το σύστημα επέστρεψε πάντα ως πιο πιθανή τη "σωστή" εικόνα, ανεξάρτητα από το μέτρο απόστασης που χρησιμοποιήθηκε (d_{ns}, D_{KL}, D_S, D_Q).

Ανοχή της Μεθόδου σε διάφορες μορφές Θορύβου

Προκειμένου να μελετήσουμε την απόδοση της προτεινόμενης μεθόδου σε περιπτώσεις που η εικόνα-ερώτηση δεν είναι "ακριβώς" ίδια με καμία εικόνα της βάσης, χρησιμοποιήσαμε τις εξής μορφές θορύβου [1,2,3,9]:

- Επιπλέον ευθύγραμμα τμήματα: Προσθήκη ευθύγραμμων τμημάτων τυχαίου μήκους με τυχαίες γωνίες ως προς τον οριζόντιο άξονα σε τυχαίες θέσεις στην αρχική εικόνα.



Σχήμα 5.3β: Οι τέσσερις πιο "κοντινές" εικόνες (ως προς την εικόνα 5.3α) που υπάρχουν στη βάση.

- Λιγότερα ευθύγραμμα τμήματα: Τυχαία επιλογή και απομάκρυνση κάποιων ευθύγραμμων τμημάτων που υπάρχουν σε μια εικόνα.
- Σφάλματα κατάτμησης: Μετατόπιση των άκρων κάποιων ευθύγραμμων τμημάτων.
- Μίξη από τις παραπάνω μορφές θορύβου.

Χρησιμοποιώντας τις παραπάνω μορφές θορύβου, για κάθε εικόνα των δύο βάσεων κατασκευάστηκε μια σειρά από εικόνες η κάθε μια από τις οποίες έχει συγκεκριμένο τύπο και ποσοστό θορύβου. Στη συνέχεια, μελετήσαμε την απόδοση της προτεινόμενης μεθόδου όταν ως εικόνα-ερώτηση χρησιμοποιείται μία από τις εικόνες με θόρυβο. Στον Πίνακα 5.1 παρουσιάζονται τα σφάλματα του συστήματος σε 182 διαφορετικές αναζητήσεις για ποσοστά 5%-20% θορύβου. Στο πείραμα αυτό κατά τη μοντελοποίηση των εικόνων χρησιμοποιήθηκε ο αλγόριθμος Greedy EM Οι στήλες **A** δείχνουν τον αριθμό των λαθών όταν "λάθος" σε μια αναζήτηση ορίζεται η αποτυχία του συστήματος να επιστρέψει ως **πιο πιθανή** τη σωστή εικόνα, ενώ οι στήλες **B** δείχνουν τον αριθμό των λαθών όταν το "λάθος" σε μια αναζήτηση ορίζεται ως η απουσία της σωστής εικόνας από τις **πέντε** πιο πιθανές εικόνες που επιστρέφει το σύστημα.

Στον Πίνακα 5.2 παρουσιάζονται τα σφάλματα του συστήματος σε 182 διαφορετικές αναζητήσεις για ποσοστά θορύβου 10%-50%. Στις στήλες "GMMs Hist-init-EM" τα αποτελέσματα προέκυψαν από μοντελοποίηση με GMMs για την εκπαίδευση των

		d_{ns}		D_{KL}		D_S	
Type	**%**	A	B	A	B	A	B
Missing lines	5%	6	**1**	4	**1**	4	**0**
	10%	50	**20**	34	**15**	27	**5**
	15%	81	**48**	71	**42**	58	**24**
	20%	113	**74**	97	**63**	83	**43**
Segment errors	5%	1	**1**	1	**0**	0	**0**
	10%	17	**3**	7	**1**	4	**0**
	15%	40	**12**	24	**5**	12	**2**
	20%	81	**30**	52	**16**	29	**7**
Extra lines	5%	6	**1**	4	**1**	2	**1**
	10%	52	**20**	38	**12**	24	**2**
	15%	111	**50**	97	**32**	62	**6**
	20%	150	**73**	139	**68**	92	**23**
Mixed	15%	64	**30**	43	**21**	28	**8**

Πίνακας 5.1 : Σφάλματα σε ανακτήσεις εικόνων κατά την Εφαρμογή 1.

οποίων χρησιμοποιήθηκε ο αλγόριθμος EM ο οποίος αρχικοποιήθηκε με βάση το ιστόγραμμα ενός χαρακτηριστικού και το μέτρο απόστασης μεταξύ των μικτών κατανομών ήταν η απόσταση D_Q. Οι στήλες "A" και "B" υποδηλώνουν την έννοια σου "σφάλματος" όπως στον Πίνακα 5.1. Στις στήλες "Relational Histograms" τα αποτελέσματα αφορούν την απόδοση της μεθόδου που προτείνουν οι Huet και Hancock στο [9]. Στη συγκεκριμένη μέθοδο οι συγγραφείς κατασκευάζουν το κανονικοποιημένο ιστόγραμμα του χαρακτηριστικού για κάθε εικόνα και πραγματοποιούν συγκρίσεις μεταξύ των ιστογραμμάτων. Το μέτρο απόστασης που χρησιμοποιούν είναι η απόσταση Bhattacharyya μεταξύ κανονικοποιημένων ιστογραμμάτων, η οποία υπολογίζεται αναλυτικά. Η συγκεκριμένη μέθοδος αποτελεί μια από τις πιο απλές και αποδοτικές μεθόδους που έχουν προταθεί μέχρι σήμερα για την επίλυση του συγκεκριμένου προβλήματος. Για επιπλέον λεπτομέρειες όσο αφορά τη μέθοδο των Huet και Hancock ο αναγνώστης παραπέμπεται στο [9].

Στον Πίνακα 5.3 παρουσιάζονται τα σφάλματα του συστήματος σε 978 διαφορετικές αναζητήσεις για ποσοστά θορύβου 10%-30%. Στις στήλες "GMMs Hist-init-EM" τα αποτελέσματα προέκυψαν από μοντελοποίηση με GMMs για την εκπαίδευση των οποίων χρησιμοποιήθηκε ο αλγόριθμος EM ο οποίος αρχικοποιήθηκε με βάση το ιστόγραμμα ενός χαρακτηριστικού και το μέτρο απόστασης μεταξύ των μικτών κατανομών ήταν η απόσταση D_Q. Στις στήλες "Relational Histograms" τα αποτελέσματα αφορούν την απόδοση της μεθόδου που προτείνουν οι Huet και Hancock στο [9]. Και πάλι οι στήλες

		GMMs Hist-Init-EM				Relational Histograms			
Number of bins/components		12		36		12		36	
Type	%	A	B	A	B	A	B	A	B
Missing lines	10%	0	**0**	0	**0**	0	**0**	0	**0**
	20%	16	**0**	2	**0**	12	**1**	0	**0**
	30%	59	**17**	7	**1**	48	**9**	10	**0**
	40%	91	**36**	38	**11**	83	**29**	32	**10**
	50%	124	**75**	95	**40**	122	**71**	72	**35**
Segment errors	10%	0	**0**	0	**0**	1	**0**	0	**0**
	20%	0	**0**	0	**0**	1	**0**	0	**0**
	30%	4	**0**	2	**1**	5	**0**	0	**0**
	40%	13	**3**	10	**0**	11	**1**	2	**0**
	50%	95	**37**	60	**11**	64	**22**	38	**8**
Extra lines	10%	1	**0**	2	**0**	3	**0**	0	**0**
	20%	36	**7**	7	**1**	35	**14**	11	**1**
	30%	88	**43**	42	**13**	71	**31**	33	**14**
	40%	122	**71**	80	**33**	107	**53**	60	**29**
	50%	139	**88**	104	**68**	124	**81**	92	**48**

Πίνακας 5.2 : Σφάλματα σε 182 ανακτήσεις εικόνων παρουσία θορύβου στην προτεινόμενη μέθοδο και στη μέθοδο [9].

		GMMs Hist-Init-EM				Relational Histograms			
Number of bins/components		12		36		12		36	
Type	**%**	A	B	A	B	A	B	A	B
Missing lines	10%	3	**0**	0	**0**	4	**0**	0	**0**
	20%	106	**18**	6	**1**	42	**4**	0	**0**
	30%	363	**141**	90	**29**	199	**52**	23	**6**
Segment errors	10%	5	**0**	2	**0**	3	**0**	0	**0**
	20%	66	**5**	15	**1**	23	**2**	2	**0**
	30%	181	**36**	79	**9**	136	**13**	34	**3**
Extra lines	10%	81	**10**	10	**0**	24	**4**	1	**0**
	20%	464	**219**	153	**47**	259	**76**	55	**16**
	30%	741	**511**	489	**263**	574	**313**	301	**135**

Πίνακας 5.3 : Σφάλματα σε 978 ανακτήσεις εικόνων παρουσία θορύβου στην προτεινόμενη μέθοδο και στη μέθοδο [9].

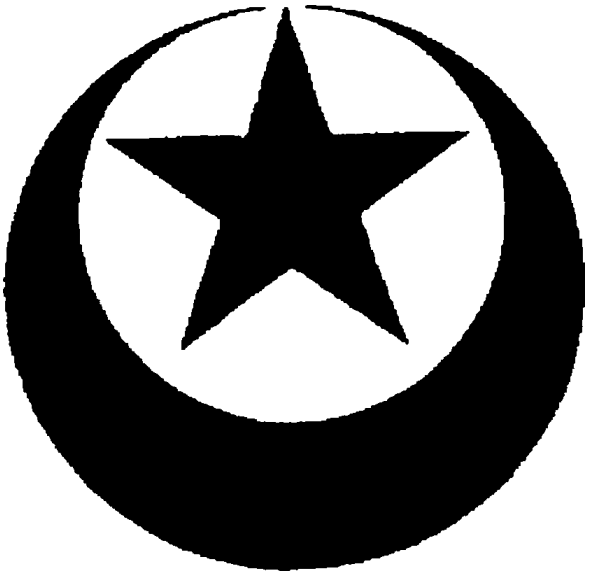
	GMMs Hist-Init-EM		Relational Histograms	
Type	A	B	A	B
crop 25% in	375	**230**	328	**212**
crop 25% out	367	**224**	366	**243**

Πίνακας 5.4 : Σφάλματα σε 978 ανακτήσεις εικόνων παρουσία "αποκοπής" στην προτεινόμενη μέθοδο και στη μέθοδο [9].

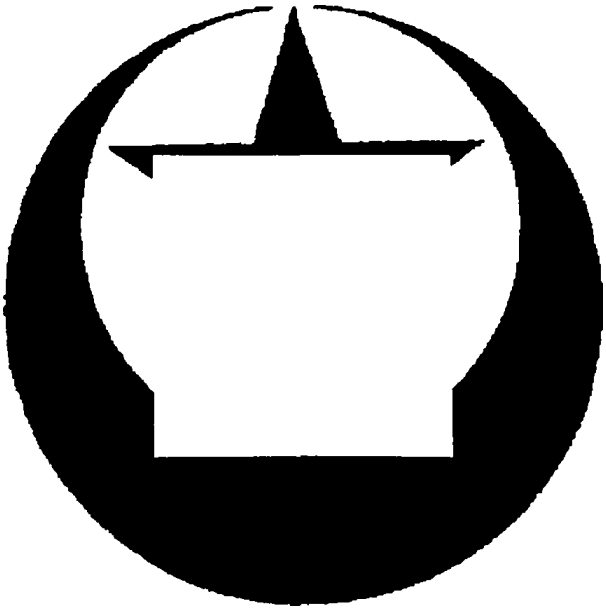
"A" και "B" υποδηλώνουν την έννοια σου "σφάλματος" όπως στον Πίνακα 5.1.

Τέλος μελετήσαμε την απόδοση της προτεινόμενης μεθόδου όταν στην εικόνα-ερώτηση έχει εφαρμοστεί *αποκοπή* του 25% της αρχικής εικόνας *εσωτερικά ή εξωτερικά*. Για παράδειγμα, από την εικόνα στο Σχήμα 5.4a προέκυψαν οι εικόνες-ερωτήσεις που φαίνονται στα Σχήματα 5.4β και 5.4γ.

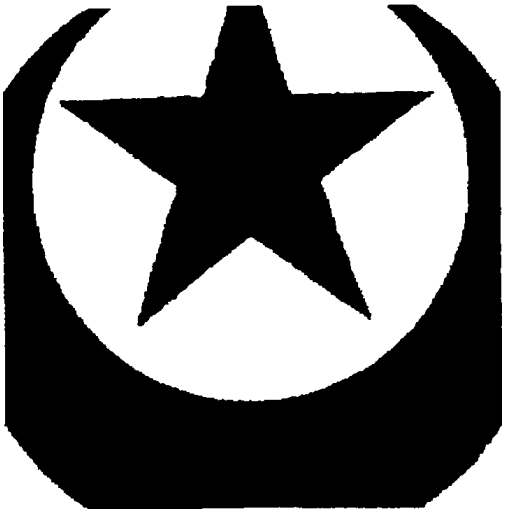
Κατασκευάσαμε 978 εικόνες-ερωτήσεις για κάθε περίπτωση και εκτελέσαμε 978 ερωτήσεις στο σύστημά μας ανά περίπτωση. Η μοντελοποίηση έγινε και πάλι με GMMs (με 36 πυρήνες) τα οποία εκπαιδεύτηκαν με απλό EM ο οποίος αρχικοποιήθηκε με βάση το ιστόγραμμα του χαρακτηριστικού. Το μέτρο απόστασης μεταξύ των GMMs ήταν και πάλι η απόσταση D_Q. Στον Πίνακα 5.4 εμφανίζονται τα σφάλματα τόσο της προτεινόμενης μεθόδου όσο και της μεθόδου [9] (με χρήση 36 bins) στο συγκεκριμένο



Σχήμα 5.4α: Εικόνα που υπάρχει στη βάση.



Σχήμα 5.4β: Η εικόνα-ερώτηση που προέκυψε από την εικόνα στο Σχήμα 5.4α με αποκοπή του 25% του εσωτερικού της.



Σχήμα 5.4γ: Η εικόνα-ερώτηση που προέκυψε από την εικόνα στο Σχήμα 5.4α με αποκοπή του 25% του εξωτερικού της.

πείραμα.

5.2 Εφαρμογές της μεθόδου σε Τεχνητές Εικόνες: Περίπτωση 1

5.2.1 Εικόνες που χρησιμοποιήθηκαν

Κατασκευάσαμε 2000 διαφορετικές εικόνες η κάθε μία από τις οποίες απεικονίζει μια συμβολοσειρά 6 χαρακτήρων από το λατινικό αλφάβητο. Στο Σχήμα 5.6 απεικονίζεται μία από τις 2000 διαφορετικές εικόνες.

5.2.2 Διαδικασία και Παράμετροι Μοντελοποίησης

Μοντελοποιήσαμε τις εικόνες με χρήση GMMs τα οποία αρχικοποιήθηκαν με βάση το ιστόγραμμα του μοναδικού χαρακτηριστικού που εξάχθηκε από κάθε εικόνα και εκπαιδεύτηκαν με απλό EM. Για τη μοντελοποίηση μία φορά επιλέξαμε κάθε GMM να αποτελείται από 12 πυρήνες και μία από 36 πυρήνες. Και στις δύο περιπτώσεις η ευρετηριοποίηση που επιτεύχθηκε ήταν τέλεια. Σε 2000 διαφορετικές αναζητήσεις

Σχήμα 5.6: Εικόνα που υπάρχει στη βάση.

Σχήμα 5.7: Η εικόνα-ερώτηση που προέκυψε από την εικόνα στο Σχήμα 5.6 με αποκοπή ενός γράμματος.

ανά περίπτωση, χωρίς παρουσία θορύβου στην εικόνα-ερώτηση, το σύστημα πάντα επέστρεφε ως πιο "πιθανή" τη σωστή εικόνα.

5.2.3 Ανοχή της Μεθόδου σε Θόρυβο

Για να μελετήσουμε την απόδοση της μεθόδου παρουσία θορύβου στην εικόνα-ερώτηση κάναμε το εξής: Για κάθε εικόνα στην τεχνητή βάση δημιουργήσαμε μία νέα η οποία διαφέρει από την πρωτότυπη μόνο σε ένα χαρακτήρα. Πιο συγκεκριμένα ένας από τους έξι χαρακτήρες της αρχικής εικόνας αντικαταστάθηκε με τον κενό χαρακτήρα. Για παράδειγμα από την εικόνα στο Σχήμα 5.6 προέκυψε η εικόνα στο Σχήμα 5.7. Κάθε νέα εικόνα χρησιμοποιήθηκε ως εικόνα-ερώτηση στο σύστημά μας (συνολικά 2000 ερωτήσεις). Στον Πίνακα 5.5 παρουσιάζονται τα σφάλματα της προτεινόμενης μεθόδου και της μεθόδου [9] κατά αντιστοιχία με τους προηγούμενους πίνακες. Θα πρέπει να σημειωθεί ότι κάθε εικόνα-ερώτηση που δημιουργήθηκε και χρησιμοποιήθηκε στο παραπάνω πείραμα αντιστοιχεί μόνο σε μία εικόνα της βάσης. Με άλλα λόγια οι εικόνες της βάσης διαφέρουν μεταξύ τους τουλάχιστον σε δύο χαρακτήρες.

		GMMs Hist-Init-EM		Relational Histograms	
Type	**#bins/components**	A	B	A	B
one empty character	12	1063	**695**	1045	**671**
	36	246	**78**	428	**294**

Πίνακας 5.5 : Σφάλματα σε 2000 αναζητήσεις εικόνων παρουσία "αποκοπής" ενός χαρακτήρα στην εκάστοτε εικόνα-ερώτηση, στην προτεινόμενη μέθοδο και στη μέθοδο [9].

5.3 Εφαρμογές της μεθόδου σε Τεχνητές Εικόνες: Περίπτωση 2

5.3.1 Εικόνες που χρησιμοποιήθηκαν

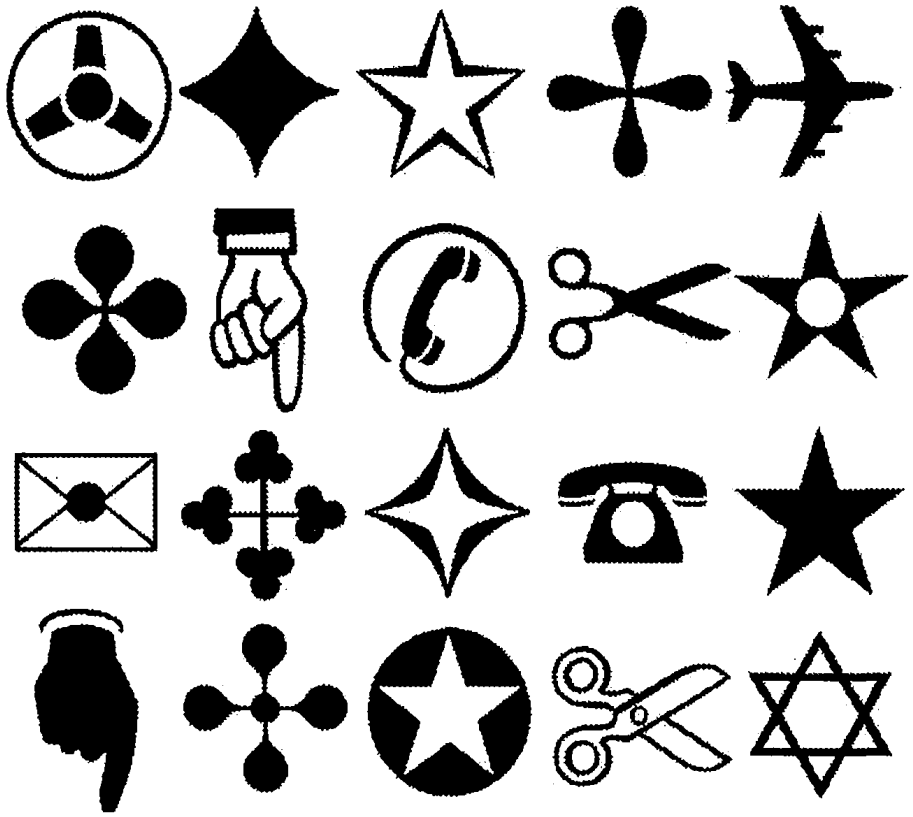
Κατασκευάσαμε 1500 διαφορετικές εικόνες η κάθε μία από τις οποίες απεικονίζει μια μήτρα 3×3 από σύμβολα. Στο Σχήμα 5.8 φαίνονται τα 20 σύμβολα που χρησιμοποιήθηκαν, ενώ στις εικόνες 5.9 και 5.10 απεικονίζονται δύο από τις 1500 διαφορετικές παραγόμενες εικόνες της βάσης.

5.3.2 Διαδικασία και Παράμετροι Μοντελοποίησης

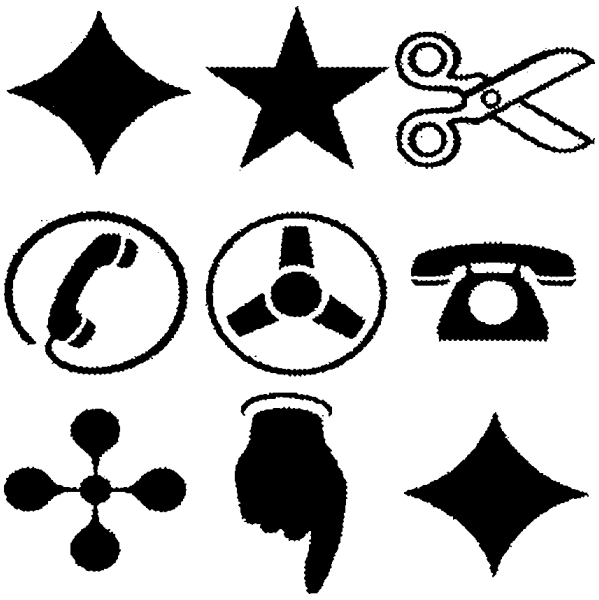
Μοντελοποιήσαμε τις εικόνες με χρήση GMMs τα οποία αρχικοποιήθηκαν με βάση το ιστόγραμμα του μοναδικού χαρακτηριστικού (f_0) που εξάχθηκε από κάθε εικόνα και εκπαιδεύτηκαν με απλό EM. Για τη μοντελοποίηση μία φορά επιλέξαμε κάθε GMM να αποτελείται από 12 πυρήνες και μία από 36 πυρήνες. Και στις δύο περιπτώσεις η ευρετηριοποίηση που επιτεύχθηκε ήταν τέλεια. Σε 1500 διαφορετικές αναζητήσεις ανά περίπτωση, χωρίς παρουσία θορύβου στην εικόνα-ερώτηση, το σύστημα πάντα επέστρεφε ως πιο "πιθανή" τη σωστή εικόνα.

5.3.3 Ανοχή της Μεθόδου σε Θόρυβο

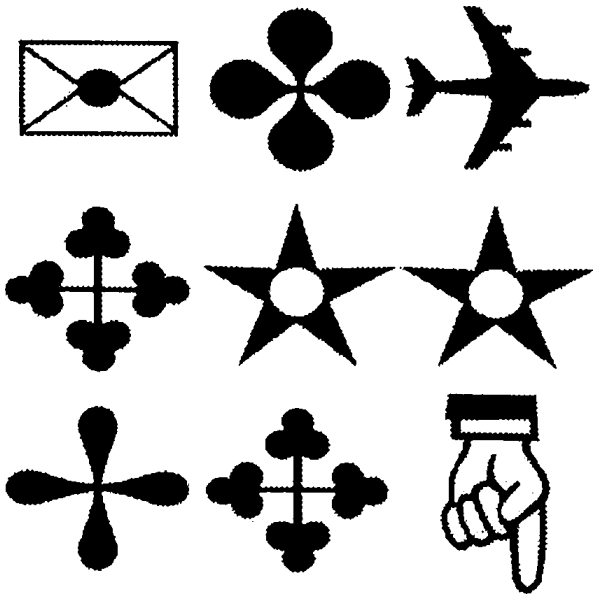
Για να μελετήσουμε την απόδοση της μεθόδου παρουσία θορύβου στην εικόνα-ερώτηση χάναμε το εξής: Για κάθε εικόνα στην τεχνητή βάση δημιουργήσαμε μία νέα η οποία διαφέρει από την πρωτότυπη μόνο σε ένα σύμβολο. Πιο συγκεκριμένα ένα από τα εννέα σύμβολα της αρχικής εικόνας αποκόπηκε. Για παράδειγμα από την εικόνα στο Σχήμα 5.10 προέκυψε η εικόνα στο Σχήμα 5.11. Κάθε νέα εικόνα χρησιμοποιήθηκε ως εικόνα-ερώτηση στο σύστημά μας (συνολικά 1500 ερωτήσεις). Στον Πίνακα 5.6 παρουσιάζονται τα σφάλματα της προτεινόμενης μεθόδου και της μεθόδου [9] κατά



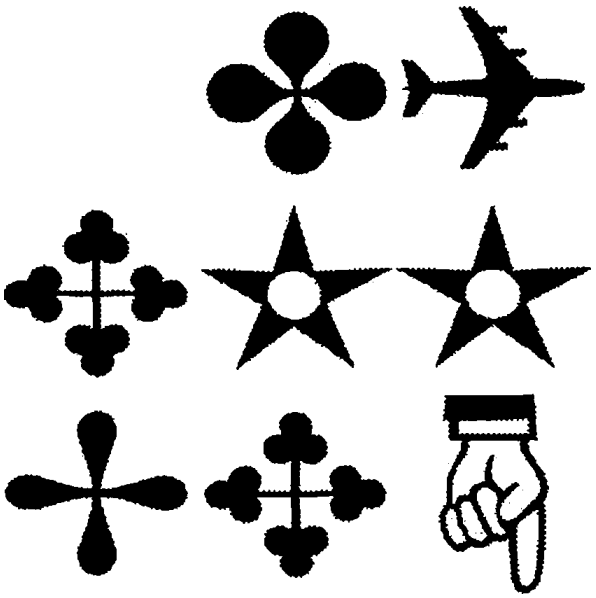
Σχήμα 5.8: Τα 20 σύμβολα που χρησιμοποιήθηκαν για την κατασκευή της βάσης.



Σχήμα 5.9: Μια από τις 1500 διαφορετικές εικόνες της βάσης.



Σχήμα 5.10: Μια από τις 1500 διαφορετικές εικόνες της βάσης.



Σχήμα 5.11: Η εικόνα-ερώτηση που προέκυψε από την εικόνα στο Σχήμα 5.10 με αποκοπή ενός συμβόλου.

αντιστοιχία με τους προηγούμενους πίνακες. Και πάλι θα πρέπει να σημειωθεί ότι κάθε εικόνα-ερώτηση που δημιουργήθηκε και χρησιμοποιήθηκε στο παραπάνω πείραμα αντιστοιχεί μόνο σε μία εικόνα της βάσης. Με άλλα λόγια οι εικόνες της βάσης διαφέρουν μεταξύ τους τουλάχιστον σε δύο σύμβολα.

5.4 Σχόλια-Παρατηρήσεις

Από τα αποτελέσματα των μετρήσεων που πραγματοποιήθηκαν στις εφαρμογές με πραγματικές εικόνες, εύκολα μπορεί κανείς να συμπεράνει πως η προτεινόμενη μέθοδος είναι αρκετά αποδοτική. Για ποσοστά θορύβου έως 15% η μέθοδος έχει καλή συμπεριφορά κάτι που είναι ιδιαιτέρως ικανοποιητικό αν σκεφτεί κανείς ότι η παρουσία υψηλού ποσοστού θορύβου (άνω του 20%) σε μια εικόνα-ερώτηση καθιστά το πρόβλημα της αναζήτησης ιδιαιτέρως δύσκολο ακόμη και για το ανθρώπινο μάτι.

Όσο αφορά τις επιδόσεις των διαφόρων μέτρων απόστασης μεταξύ μικτών κανονικών κατανομών, από τον Πίνακα 5.1 φαίνεται καθαρά πως μεταξύ αυτών που ο υπολογισμός τους απαιτεί τη χρήση των δεδομένων (d_{ns}, D_S, D_{KL}) η Συμμετρική Μέση Λογαριθμική Πιθανοφάνεια D_S έχει την καλύτερη απόδοση. Παρ' όλα αυτά, οι εν λόγω αποστάσεις χρινονται γενικότερα ασύμφορες μιας και οι υπολογιστικές (και κατ' επέκταση χρονικές) απαιτήσεις που έχουν είναι ιδιαιτέρως υψηλές για ένα πρόβλημα όπως

		GMMs Hist-Init-EM		Relational Histograms	
Type	**#bins/components**	A	B	A	B
one empty symbol	12	498	**322**	546	**348**
	36	222	**126**	186	**122**

Πίνακας 5.6 : Σφάλματα σε 1500 αναζητήσεις εικόνων παρουσία "αποκοπής" ενός συμβόλου στην εκάστοτε εικόνα-ερώτηση, στην προτεινόμενη μέθοδο και στη μέθοδο [9].

η αναζήτηση εικόνων όπου ο χρόνος απόκρισης θα πρέπει να είναι ελάχιστος. Αντίθετα, η απόσταση D_Q παρουσιάζει ιδιαίτερα καλές επιδόσεις, τόσο από πλευράς χρόνου όσο και από πλευράς αποτελεσμάτων.

Από την άλλη, από την Εφαρμογή 2 είναι φανερό ότι το πιο σημαντικό χαρακτηριστικό (από τα πέντε) που εξάγεται από κάθε εικόνα είναι το f_0. Σε αυτό καταλήγουν και οι Huet και Hancock στο [9] όπου σημειώνουν ότι οι περιπτώσεις κατά τις οποίες τα υπόλοιπα χαρακτηριστικά βοηθούν σημαντικά στη διαδικασία ανάκτησης είναι πολύ λίγες.

Τέλος, σε σύγκριση με την μέθοδο [9], η οποία αποτελεί μια από τις πιο επιτυχημένες προσεγγίσεις που έχουν προταθεί για το πρόβλημα, η προτεινόμενη μέθοδος δείχνει να τα πάει αρκετά καλά. Για σχετικά "λογικά" ποσοστά θορύβου στην εικόνα-ερώτηση, οι δύο μέθοδοι παρουσιάζουν την ίδια σχεδόν επίδοση ενώ από τα αποτελέσματα της εφαρμογής στην βάση με τις 2000 εικόνες η προτεινόμενη προσέγγιση φαίνεται να έχει καλύτερη συμπεριφορά.

Κεφάλαιο 6

Επίλογος

6.1 Συμπεράσματα

Στην παρούσα εργασία προτάθηκε μια μέθοδος για *ευρετηριοποίηση* μεγάλων βάσεων με εικόνες και αποδοτική *αναζήτηση* εικόνων από αυτή. Η ευρετηριοποίηση και η αναζήτηση των εικόνων γίνεται με βάση το περιεχόμενο της εικόνας και πιο συγκεκριμένα το σχήμα/περίγραμμα της εικόνας. Η μέθοδος βασίζεται στη χρήση στατιστικών μοντέλων (μικτών κανονικών κατανομών) τα οποία χρησιμοποιούνται για τη μοντελοποίηση της πληροφορίας του σχήματος/περιγράμματος που εξάγεται από κάθε εικόνα. Με τη μοντελοποίηση του σχήματος κάθε εικόνας ο βαθμός ομοιότητας μεταξύ δύο εικόνων αντιστοιχίζεται στην απόσταση μεταξύ των μικτών κατανομών που μοντελοποιούν το περιεχόμενο των δύο εικόνων.

Η αποδοτική εκπαίδευση των μικτών κανονικών κατανομών αποτελεί ένα πάγιο πρόβλημα στο πεδίο της στατιστικής μιας και οι ιδιαιτερότητες που παρουσιάζει κάθε πρόβλημα δεν καθιστά ικανή την εφαρμογή μιας απόλυτης μεθόδου εκπαίδευσης. Αυτό μας απασχόλησε έντονα και στην παρούσα εργασία μιας και η επιτυχία της μεθόδου εξαρτάται κυρίως από την ποιότητα της μοντελοποίησης που θα πραγματοποιηθεί στις εικόνες. Το πλήθος των χαρακτηριστικών που εξάγονται από κάθε εικόνα καθώς και η συμπεριφορά τους σε διάφορες μορφές θορύβου είναι οι σημαντικότερες ιδιαιτερότητες του προβλήματος που αντιμετωπίστηκε. Από τα αποτελέσματα των εφαρμογών συμπεραίνουμε πως από τα πέντε χαρακτηριστικά που προτείνονται για την περιγραφή του σχήματος μιας εικόνας, τα τέσσερα σπάνια βοηθούν σημαντικά στην αποδοτική μοντελοποίηση. Επίσης, παρουσία υψηλού ποσοστού θορύβου στην εικόνα τα χαρακτηριστικά αλλοιώνονται σημαντικά κάτι που καθιστά το πρόβλημα ιδιαίτερα δύσκολο.

Ένα άλλο πρόβλημα που αντιμετωπίστηκε είναι η επιλογή του μοντέλου που θα χρησιμοποιηθεί. Η χρήση του Greedy EM ώστε να απαλλαγούμε από το πρόβλημα της αρχικοποίησης του μοντέλου αλλά και την επιλογή του αριθμού των πυρήνων αποδεί-

χτηκε αρκετά ικανοποιητική στις περισσότερες περιπτώσεις. Παρόλα αυτά, υπήρχαν και περιπτώσεις όπου ο αλγόριθμος παρουσίασε αδυναμίες, όπως η τάση για προσθήκη πάρα πολλών πυρήνων σε σημεία όπου υπάρχει μεγάλη συγκέντρωση πληροφορίας.

Τέλος, όσο αφορά τα μέτρα απόστασης μεταξύ μικτών κανονικών κατανομών, τα αποτελέσματα των πειραματικών εφαρμογών καθιστούν την κανονικοποιημένη τετραγωνική απόσταση (D_Q) ως την ιδανικότερη επιλογή. Εκτός της ικανότητας της για ακριβή αποτελέσματα, η απόσταση D_Q μπορεί να υπολογιστεί αναλυτικά και χωρίς ιδιαίτερα υψηλό επεξεργαστικό (και επομένως χρονικό) κόστος. Το τελευταίο χαρακτηριστικό της είναι ιδιαίτερα σημαντικό αν σκεφτεί κανείς πως για μια και μόνο αναζήτηση σε μια βάση μπορούν να απαιτούνται χιλιάδες υπολογισμοί αποστάσεων.

6.2 Μελλοντική Έρευνα

Η παρούσα εργασία αποτέλεσε το πρώτο μας βήμα στην χρήση στατιστικών μεθόδων για την επίλυση προβλημάτων στον τομέα των πολυμέσων. Πρωταρχικός μας στόχος είναι η επέκταση της μεθόδου ώστε να επιτύχουμε αποδοτική ευρετηριοποίηση και αναζήτηση εικόνων με βάση άλλα χαρακτηριστικά του περιεχομένου τους όπως χρώμα και υφή. Επίσης, ιδιαίτερο ενδιαφέρον παρουσιάζει η ιδέα του συνδυασμού χαρακτηριστικών (χρώμα, σχήμα και υφή) των εικόνων ώστε να αντιμετωπιστεί το πρόβλημα της ευρετηριοποίησης γενικά.

Από την άλλη, όπως επισημάνθηκε και στο Κεφάλαιο 2, τα προβλήματα στο πεδίο είναι πολλά και ενδιαφέροντα. Τα αποτελέσματα της παρούσας εργασίας μας ωθούν στην μελέτη και αντιμετώπιση (πάντα με χρήση στατιστικών μεθόδων) και άλλων προβλημάτων όπως η αναζήτηση αντικειμένων σε βάσεις με εικόνες ή video, ομαδοποίηση εικόνων ή αρχείων ήχου, ταξινόμηση εικόνων, video ή αρχείων ήχου με βάση το περιεχόμενό τους και πολλά άλλα. Τέλος, ιδιαίτερο ενδιαφέρον παρουσιάζει η ιδέα της εισαγωγής του ανθρώπινου παράγοντα ως καθοδηγητή στη διαδικασία εκπαίδευσης του συστήματος (relevance feedback).

Βιβλιογραφία

[1] Benoit Huet and Edwin R. Hancock, *Line Pattern Retrieval Using Relational Histograms*, IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 21, No. 12, page 1363-1370, December 1999.pp. 314-45, 1999.

[2] B. Huet, A.D.J. Cross and E.R. Hancock, *Shape Recognition from Large Image Libraries by Inexact Graph Matching*, Pattern Recognition in Practice VI, June 2-4 1999, Vlieland, The Netherlands. Appeared in a special issue of Pattern Recognition Letters, No. 20, page 1259-1269, Dec 1999.

[3] B. Huet and E. R. Hancock, *Relational Object Recognition from Large Structural Libraries*, Pattern Recognition, Vol. 35, No. 9, page 1895-1915, Sept 2002.

[4] Nikos Vlassis and Aristidis Likas, *A Greedy EM Algorithm for Gaussian Mixture Learning*, Neural Processing Letters 15:77-87,2002.

[5] C. M. Bishop *Neural Networks for Pattern Recognition*, Clarendon Press - Oxford, 1997.

[6] P. N. Suganthan, *Shape indexing using self-organising maps*, IEEE Transactions on Neural Networks, Vol. 13, No. 4, pp. 835-840, July 2002.

[7] Keinosuke Fukunaga, *Introduction to Statistical Pattern Recognition*, Academic Press, 1990.

[8] Jeffrey S. Beis and Daivid G. Lowe, *Shape Indexing Using Approximate Nearest-Neighbour Search in High-Dimensional Spaces*, Conference on Computer Vision and Pattern Recognition, pp. 1000-1006, Puerto Rico, June 1997.

[9] B. Huet and E.R. Hancock, *Relational Histograms for Shape Indexing*, IEEE International Conference on Computer Vision (ICCV98), Mumbai India, pp. 563-569, Jan 1998.

[10] Hayit Greenspan, Shiri Gordon and Jacob Goldberger, *Probabilistic models for generating, modelling and matching image categories*, International Conference on Pattern Recognition (ICPR), 2002.

[11] Hayit Greenspan, Jacob Goldberger and Lenny Riddel, *A continuous probabilistic framework for image matching*, Journal of Computer Vision and Image Understanding, Vol. 84, pp. 384-406, 2001.

[12] Adoram M, Lew MS, *IRUS: Image Retrieval Using Shape*, IEEE International Conference on Multimedia Computing Systems, Vol. 2. 1999; 597-602.

[13] Gnsel B, Tekalp AM, *Shape similarity matching for query-by-example*, Pattern Recognition. 1998; 31(7):931-944.

[14] Lei Z, Tasdizen T, Cooper D. *Object signature curve and invariant shape patches for geometric indexing into pictorial databases*, Proceedings of IS&T/SPIE Conference on Multimedia Storage and Archiving Systems. 1997; 3229:232-243.

[15] Alferez R, Wang YF, *Image Indexing and Retrieval Using Image-Derived, Geometrically and Illumination Invariant Features*, IEEE International Conference on Multimedia Computing Systems, vol. 1. 1999; 177-182.

[16] Petrakis EGM, Milios E, *Efficient retrieval by shape content*, IEEE International Conference on Multimedia Computing Systems, Vol. 2. 1999; 616-621.

[17] Sharvit D, Chan J, Tek H, Kimia BB, *Symmetry-based indexing of image databases*, Journal of Visual Communications and Image Representation, 1998; 9(4):366-380.

[18] Rui Y, Huang TS, Mehrotra S, Ortega M, *Automatic matching tool selection using relevance feedback in MARS*, Second International Conference on Visual Information Systems (VISUAL'97). 1997; 109-116.

[19] J. Q. Li and A. R. Barron, *Mixture Density Estimation*, Advances in Neural Information Processing Systems 12, The MIT Press, 2000.

[20] R. O. Duda and P. E. Hart, *Pattern Classification and Scene Analysis*, Wiley, 1973.

[21] D. M. Titterington, A. F. Smith and U. E. Makov, *Statistical Analysis of Finite Mixture Distributions*, Wiley, 1985.

[22] G. McLachlan, D. Peel, *Finite Mixture Models*, Wiley, 2000.

[23] G. McLachlan and T. Krishnan, *The EM Algorithm and Extensions*, Marcel Dekker, 1997.

[24] R. Redner and H. Walker, *Mixture Densities, Maximum Likelihood and the EM ALgorithm*, SIAMM Review, vol. 26, no. 2, pp. 195-239, 1984.

[25] M. A. Carreira-Perpinan and S. Renals, *Practical Identifiability of Finite Mixtures of Bernouli Distributions*, Neural Computation, vol. 12, no. 1, pp. 141-152, Jan 2000.

[26] W. R. Gilks, S. Richardson and D. J. Spiegelhalter, *Marcov Chain Monte Carlo in Practice*, London, Chapman & Hall.

[27] A. P. Dempster, N. M. Laird and D. B. Rubin, *Maximum Likelihood Estimation from Incomplete Data via the EM Algorithm*, Journal of the Royal Statistical Society B, vol. 39, pp. 1-38, 1977.

[28] Ray Surajit, *PhD Dissertation: Distance-based Model-selection with application to Analysis of Gene-expression Data*, Department of Statistics, The Pennsylvania State University, Dec. 2003.

[29] Peter Kovesi, MATLAB Functions for Computer Vision and Image Analysis, School of Computer Science and Software Engineering, The University of Western Australia, Crawley, Western Autralia.

[30] Matlab User's Guide for UNIX, Mathworks Inc.

[31] WordNet Princeton University.

[32] http://www.eurecom.fr/~huet/

Παράρτημα

Greedy EM για εκπαίδευση GMMs [4]

Αν $\phi(x;\theta_j)$ είναι το j-ιοστό component ενός μικτού μοντέλου και θ_j οι παράμετροι αυτού, τότε η πυκνότητα της μίξης για ένα τυχαίο διάνυσμα x θεωρώντας ότι η μίξη αποτελείται από k πυρήνες είναι:

$$f_k(x) = \sum_{j=1}^{k} \pi_j \phi(x;\theta_j)$$

όπου π_j είναι το βάρος του πυρήνα j στη μίξη με $\pi_1 + \cdots + \pi_k = 1, \pi_j \geq 0$, και γενικά για components στις d διαστάσεις

$$\phi(x;\theta_j) = (2\pi)^{-d/2}|S_j|^{-1/2}\exp\{-0.5(x-m_j)^T S_j^{-1}(x-m_j)\}$$

όπου το διάνυσμα μέσου m_j και ο πίνακας συμμεταβλητότητας S_j της κατανομής εμπεριέχονται στο διάνυσμα παραμέτρων θ_j. Θεωρούμε το σύνολο εκπαίδευσης $\{x_1, \cdots, x_n\}$ από iid σημεία τα οποία έχουν ληφθεί από από τη μίξη και στόχος μας είναι να εκτιμήσουμε τις παραμέτρους π_j, m_j, S_j των k πυρήνων που μεγιστοποιούν την λογαριθμική πιθανοφάνεια:

$$L_k = \sum_{i=1}^{k} \log f_k(x_i).$$

Αυτό μπορεί να επιτευχθεί με τη χρήση του αλγόριθμου EM ο οποίος συνοψίζεται στην επαναληπτική εκτέλεση των παρακάτω εξισώσεων για κάθε component j, $j = 1, \cdots, k$

$$\begin{aligned} P(j|x_i) &= \frac{\pi_j \phi(x_i;\theta_j)}{f_k(x_i)}, \\ \pi_j &:= \frac{1}{n}\sum_{i=1}^{n} P(j|x_i), \end{aligned}$$

$$
\begin{aligned}
m_j &:= \frac{\sum_{i=1}^{n} P(j|x_i)x_i}{\sum_{i=1}^{n} P(j|x_i)}, \\
S_j &:= \frac{\sum_{i=1}^{n} P(j|x_i)(x_i - m_j)(x_i - m_j)^T}{\sum_{i=1}^{n} P(j|x_i)}.
\end{aligned}
$$

Έχει αποδειχθεί ότι μετά από κάθε βήμα του ΕΜ η λογαριθμική πιθανοφάνεια δεν μπορεί να έχει μειωθεί.

Ας θεωρήσουμε τώρα ότι σε ένα μοντέλο με k συνιστώσες f_k προσθέτουμε έναν νέο πυρήνα ϕ. Για το παραγόμενο μοντέλο f_{k+1} θα ισχύει:

$$
f_{k+1} = (1-\alpha) f_k(x) + \alpha\phi(x;\theta),
$$

με $\alpha \in (0,1)$. Αν για κάθε k, δοθένος του $f_k(x)$ το βάρος α και το διάνυσμα παραμέτρων θ του εκάστοτε νέου πυρήνα $\phi(x;\theta)$ επιλέγονται βέλτιστα ώστε η νέα λογαριθμική πιθανοφάνεια

$$
L_{k+1} = \sum_{i=1}^{n} \log f_{k+1}(x_i) = \sum_{i=1}^{n} \log[(1-\alpha) f_k(x_i) + \alpha\phi(x_i;\theta)]
$$

να μεγιστοποιείται, τότε μπορούμε να ξεκινήσουμε από έναν πυρήνα και σε κάθε βήμα να προσθέτουμε έναν νέο ξέροντας ότι πάντοτε η λογαριθμική πιθανοφάνεια του μοντέλου μας θα αυξάνει πάντα. Με αυτόν τον τρόπο το πρόβλημά μας ανάγεται στην εκπαίδευση μιας μίξης με δύο συνιστώσες.

Τοπική αναζήτηση για προσθήκη συνιστώσας

Στο βήμα k θα έχουμε ένα μοντέλο $f_k(x)$ του οποίου οι παράμετροι θα παραμένουν σταθερές και έναν υπό εισαγωγή πυρήνα $\phi(x;\theta)$ με παραμέτρους $\theta = \{\alpha, m, S\}$. Εκτελώντας μερικό ΕΜ αναζητούμε τοπικά ένα μέγιστο της L_{k+1} ως προς α, m και S. Τα βήματα του μερικού ΕΜ είναι τα εξής:

$$
\begin{aligned}
P(k+1|x_i) &= \frac{\alpha\phi(x_i;m,S)}{(1-\alpha) f_k(x_i) + \alpha\phi(x_i;m,S)}, \\
\alpha &:= \frac{1}{n}\sum_{i=1}^{n} P(k+1|x_i), \\
m &:= \frac{\sum_{i=1}^{n} P(k+1|x_i)x_i}{\sum_{i=1}^{n} P(k+1|x_i)}, \\
S &:= \frac{\sum_{i=1}^{n} P(k+1|x_i)(x_i - m)(x_i - m)^T}{\sum_{i=1}^{n} P(k+1|x_i)}.
\end{aligned}
$$

Από τη στιγμή που ενημερώνονται μόνο οι παράμετροι του νέου πυρήνα, ο μερικός ΕΜ είναι πολύ γρήγορος αλλά και απλός στην υλοποίησή του. Το μόνο που μένει είναι η αρχικοποίηση των τιμών α, m και S πριν την εφαρμογή του μερικού ΕΜ. Για το λόγο αυτό προτείνεται ολική αναζήτηση στο χώρο των παραμέτρων.

Ολική αναζήτηση για προσθήκη συνιστώσας

Προκειμένου να διευκολυνθούμε αντικαθιστούμε τη συνάρτηση λογαριθμικής πιθανοφάνειας με μια προσέγγιση κατά Taylor γύρω από ένα σημείο $\alpha = \alpha_0$, και στη συνέχεια χρησιμοποιούμε την παραγόμενη προσέγγιση για να αναζητήσουμε βέλτιστες τιμές για τα m και S:

$$\hat{L}_{k+1} = L_k(\alpha_0) - \frac{[L'_{k+1}(\alpha_0)]^2}{2L''_{k+1}(\alpha_0)}$$

όπου L'_{k+1} και L''_{k+1} είναι η πρώτη και δεύτερη παράγωγος της L_{k+1} ως προς α. Αν ορίσουμε

$$\delta(x, \theta) = \frac{f_k(x) - \phi(x;\theta)}{f_k(x) + \phi(x;\theta)}$$

τότε ένα τοπικό μέγιστο της L_{k+1} κοντά στο $\alpha_0 = 0.5$ μπορεί να δειχθεί πως είναι το:

$$\hat{L}_{k+1} = \sum_{i=1}^{n} \log \frac{f_k(x_i) + \phi(x_i;\theta)}{2} + \frac{1}{2}\frac{[\sum_{i=1}^{n}\delta(x_i,\theta)]^2}{\sum_{i=1}^{n}\delta^2(x_i,\theta)}$$

και βρίσκεται για α ίσο με:

$$\hat{\alpha} = \frac{1}{2} - \frac{1}{2}\frac{\sum_{i=1}^{n}\delta(x_i,\theta)}{\sum_{i=1}^{n}\delta^2(x_i,\theta)}.$$

Αν η τιμή $\hat{\alpha}$ είναι εκτός του $(0,1)$ τότε αρχικοποιούμε τον μερικό ΕΜ με $\hat{\alpha} = 0.5$ για $k = 1$ και $\hat{\alpha} = 2/(k+1)$ για $k \geq 2$.

Για την αρχικοποίηση των m, S, παρατηρούμε ότι η $\hat{L}_{k+1}$ εξαρτάται μόνο από το $\phi(x_i; m, S)$, το οποίο, για σταθερό πίνακα συμμεταβλητότητας $S = \sigma^2 I$, είναι η ευκλείδεια απόσταση του διανύσματος m από το σημείο x_i. Για το λόγο αυτό, κατά την αρχικοποίηση του αλγορίθμου κατασκευάζουμε ένα πίνακα $\mathbf{K}$ με στοιχεία:

$$k_{ij} = (2\pi\sigma^2)^{-d/2} exp[-0.5\|x_i - x_j\|^2/\sigma^2]$$

για κατάλληλο σ, και χρησιμοποιούμε τις τιμές k_{ij} κάθε φορά που προσθέτουμε έναν νέο πυρήνα στη μίξη μας. Η επιλογή του σ εξαρτάται από τη διάσταση d των δεδομένων μας και το πλήθος n του συνόλου εκπαίδευσης. Η προτεινόμενη τιμή για το σ είναι:

$$\sigma = \beta[\frac{4}{(d+2)n}]^{1/(d+4)}$$

με β σταθερό.

Ο Αλγόριθμος Greedy EM [4]

Συνοψίζοντας τα παραπάνω τα βήματα του αλγορίθμου είναι τα εξής:

1. Αρχικοποίηση με τη χρήση ενός component με $m = E[x]$ και $S = cov(x)$. Υπολόγισε το σ θέτοντας όπου β ίσο με το μισό της μέγιστης singular τιμής του S. Αρχικοποίησε τον πίνακα K.

2. Εφάρμοσε τα βήματα του EM μέχρι $|L_k^t / L_k^{t-1}| < 1e{-}6$ ή αν ήδη έχουμε μέγιστος αριθμός πυρήνων.

3. Αναζήτησε σε όλα τα x_j υποψήφιες θέσεις για τον νέο πυρήνα. Θέσε το m σε εκείνο το x_j που μεγιστοποιεί τη νέα πιθανοφάνεια $\hat{L}_{k+1}$ με χρήση των στοιχείων k_{ij} στη θέση του $\phi(x_i; x_j, \sigma^2 I)$.

4. Αρχικοποίησε τον μερικό EM με τις τιμές των $m, S = \sigma^2 I$ και $\hat{\alpha}$ που βρέθηκαν.

5. Εφάρμοσε μερικό EM μέχρι να επιτευχθεί σύγκλιση όπως στο βήμα 2.

6. Αν $L_{k+1} \leq L_k$ τότε τερμάτισε, αλλιώς δέσμευσε το νέο πυρήνα και πήγαινε στο βήμα 2.