ΣΤΑΤΙΣΤΙΚΕΣ ΜΕΘΟΔΟΙ ΓΙΑ ΑΝΑΚΤΗΣΗ ΕΙΚΟΝΑΣ ΜΕ ΒΑΣΗ ΤΟ ΠΕΡΙΕΧΟΜΕΝΟ

Η
ΜΕΤΑΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ ΕΞΕΙΔΙΚΕΥΣΗΣ

Υποβάλλεται στην

ορισθείσα από την Γενική Συνέλευση Ειδικής Σύνθεσης
του Τμήματος Πληροφορικής
Εξεταστική Επιτροπή

από τον

Γεώργιο Σφήκα του Ανδρέα-Φανουρίου

ως μέρος των Υποχρεώσεων

για τη λήψη

του

ΜΕΤΑΠΤΥΧΙΑΚΟΥ ΔΙΠΛΩΜΑΤΟΣ ΣΤΗΝ ΠΛΗΡΟΦΟΡΙΚΗ

ΜΕ ΕΞΕΙΔΙΚΕΥΣΗ ΣΤΙΣ ΤΕΧΝΟΛΟΓΙΕΣ-ΕΦΑΡΜΟΓΕΣ

Δεκέμβριος 2006

ΠΕΡΙΕΧΟΜΕΝΑ

ΕΥΡΕΤΗΡΙΟ ΠΙΝΑΚΩΝ

Πίνακας

Σελ

.

ΕΥΡΕΤΗΡΙΟ ΣΧΗΜΑΤΩΝ

ΠΕΡΙΛΗΨΗ

Γεώργιος Σφήκας του Ανδρέα-Φανουρίου και της Ευρυδίκης. MSc, Τμήμα Πληροφορικής, Πανεπιστήμιο Ιωαννίνων, Δεκέμβριος 2006. Στατιστικές μέθοδοι για ανάκτηση εικόνας με βάση το περιεχόμενο. Επιβλέποντας: Νικόλαος Π. Γαλατσάνος.

Εξετάζουμε τεχνικές που χρησιμοποιούνται για ανάκτηση εικόνας, με βάση το περιεχόμενο της. Δηλαδή την αναζήτηση σε βάση εικόνων όχι με λέξεις-κλειδιά στην –ενδεχόμενη- γραπτή περιγραφή της εικόνας, αλλά αξιοποιώντας το αυτό καθαυτό περιεχόμενο της εικόνας. Θα εξετάσουμε μεθόδους από αυτή του χρωματικού ιστογράμματος (1991) μέχρι πιο σύγχρονες μεθόδους (2003) που κάνουν χρήση μοντέλου μίξης κανονικών κατανομών (Gaussian Mixture Models) για δεδομένα χρώματος, υφής και τοπολογίας και κατάτμηση της εικόνας. Επίσης θα παρουσιασθούν και αξιολογηθούν καινοτομίες και επεκτάσεις πάνω σε υπάρχοντα συστήματα ανάκτησης εικόνας.

EXTENDED ABSTRACT IN ENGLISH

Sfikas, Giorgos A.F.. MSc, Computer Science Department, University of Ioannina, Greece. December 2006. “Statistical methods for content-based image retrieval”. Thesis Supervisor: Nikolaos P. Galatsanos.

Content-based image retrieval (“CBIR”) is the subject of the thesis. Interest in this field has been fuelled by the increasing size of image - and multimedia in general - databases in the past few years, which leads to the problem of effectively querying them. Since careful indexing of a large image database can be very hard, alternatively querying must rely straight to the image data itself; hence the term “Content-based”.
There is a rich bibliography concerning CBIR. We present methods as simple as the color histogram [1] (1991) and make our way towards more up-to-date techniques, such as the Blobworld framework [4] (2003).
In chapter 2 we examine some image features that are typically used to describe the image: color, texture, x-y coordinates. For color, the choices range from the well-known RGB and HSV models to uniform space models, like the L*u*v* space. In order to describe texture, the polarity – anisotropy – contrast features are presented, used in [4].
In chapter 3 we present a small introduction in the Bayesian framework and Maximum likelihood estimation, used later in chapter 4. These machine learning techniques will allow us to build a more reliable form of the histogram, the Gaussian mixture model. While no closed-form formula exists to train such models, the EM is an elegant method to solve this problem.
Finally in chapter 4 we are ready to construct image descriptors, using the features examined in chapter 2, and both the histogram and mixture model methods examined in chapter 3. An outline of this procedure is presented here.
Remaining still is the last piece of the CBIR framework, which is how to query the images using the image descriptors constructed. That is shown in chapter 5. Simple and compound queries are presented, along with various descriptor distance metrics. One of them has been presented lately by us in [21]. The metrics and descriptors used in Blobworld are examined. Also, we consider modifications in the Blobworld framework.
The most important techniques described in this Thesis are used in experiments, the results of which are presented in chapter 6. It can be seen that our novel approaches fare rather well, compared to existing CBIR systems.

ΚΕΦΑΛΑΙΟ 1. ΕΙΣΑΓΩΓΗ

1.1. Τι είναι ανάκτηση εικόνας;

Ανάκτηση[1] εικόνας (**Image Retrieval**) λέμε την αναζήτηση για κάποια εικόνα ή εικόνες, μεταξύ κάποιου συνόλου εικόνων.

Η χρησιμότητα ενός συστήματος ανάκτησης εικόνας είναι ευρεία. Μπορεί να θέλουμε να βρούμε εικόνες που περιέχουν κάποιο αντικείμενο, ή είναι να ζητάμε σχέδια με κάποια συγκεκριμένη τεχνοτροπία. Σε πιο ειδικό πλαίσιο, ανάκτηση εικόνας θέλουμε να κάνουμε για παράδειγμα στην ιατρική, όταν ζητάμε να βρούμε ακτινογραφίες συγκεκριμένων οστών, ή ακόμα να ελέγξουμε φωτογραφίες σε ένα οικογενειακό άλμπουμ για ένα συγκεκριμένο πρόσωπο. Μπορούμε να σκεφτούμε εύκολα πολλά παραδείγματα από την καθημερινότητα και όχι μόνο, που η αυτοματοποιημένη ανάκτηση εικόνας θα βοηθούσε.

Ας πάρουμε τα πράγματα από την αρχή.

---

[1] Τυγχαίνει να συγχέεται με την *ανακατασκευή* εικόνας (image restoration) , με την οποία δεν έχει σχέση.

1.2. Ανάκτηση πληροφορίας γενικά

Ανάκτηση πληροφορίας λέγεται η άντληση μιας πληροφορίας, από μια βάση δεδομένων. Λέγοντας πληροφορία, μπορεί να είναι οτιδήποτε από αλφαριθμητικά όπως είναι ένα όνομα ή αριθμός τηλεφώνου, μέχρι εικόνες και βίντεο.
Στις δομημένες βάσεις δεδομένων η ανάκτηση πληροφορίας γίνεται με γλώσσες ερωτήσεων (query languages). Ερωτήσεις είναι μικρά κομμάτια κώδικα, που περιγράφουν τι θέλει να μάθει ο χρήστης από την βάση δεδομένων. Ο κώδικας αυτός γράφεται σε ειδική γλώσσα για αυτό το σκοπό, η δημοφιλέστερη εκ των οποίων είναι η SQL. Για να δώσουμε ένα παράδειγμα, έστω ότι έχουμε ένα πελατολόγιο και θέλουμε να μάθουμε το τηλέφωνο του πελάτη John Smith. Τότε η ερώτηση SQL θα είναι:

```
SELECT phone_number
FROM clients
WHERE full_name = ‘John Smith’
```

Σε περίπτωση τώρα που έχουμε ένα σύνολο δεδομένων, τα οποίο για τον ένα ή τον άλλο λόγο δεν τα έχουμε σε δομημένη μορφή, τα πράγματα αλλάζουν για την ανάκτηση πληροφορίας. Ένα διάσημο μη-δομημένο σύνολο δεδομένων είναι ο παγκόσμιος ιστός (World Wide Web) και γενικότερα το Internet. Παρά τον τεράστιο όγκο πληροφορίας που περιέχει, η άναρχη δόμηση του κάνει την ανάκτηση πληροφορίας δύσκολη υπόθεση. Η γιγάντωση του παγκόσμιου ιστού – ειδικά τα τελευταία χρόνια – τόνισε την ανάγκη για συστήματα ανάκτησης πληροφορίας σε μη-δομημένα σύνολα.

1.3. Ανάκτηση εικόνας

Η ανάκτηση εικόνας είναι ένα πεδίο που επιστρατεύει γνώσεις από αρκετούς διαφορετικούς επιστημονικούς τομείς. Η ανάκτηση πληροφορίας, η μοντελοποίηση δεδομένων, ανάλυση και επεξεργασία εικόνας, αναγνώριση προτύπων και

επικοινωνία ανθρώπου-μηχανής, είναι οι σημαντικότεροι, αλλά μόνο μερικοί από αυτούς τους τομείς.

Όπως και η ανάκτηση πληροφορίας γενικά, η ανάκτηση εικόνας γίνεται όλο και πιο σημαντική όσο αυξάνεται ο όγκος των αποθηκευμένων ψηφιοποιημένων εικόνων, κάτι που ενθαρρύνουν τα φθηνότερα και γρηγορότερα μέσα αποθήκευσης (CD-ROM, σκληροί δίσκοι). Προς αυτή την κατεύθυνση βοηθάνε και οι γρηγορότερες ταχύτητες στα δίκτυα (Internet) και η έντονη αύξηση της ανταλλαγής αρχείων, με την έλευση των peer-to-peer δικτύων.

Το ζητούμενο στην ανάκτηση εικόνας, παρόμοια με την ανάκτηση πληροφορίας γενικά, είναι να ανακτηθούν εικόνες σχετικές με μια ερώτηση (query) που θέλουμε. Τα δεδομένα που μπορούμε να πάρουμε με τις ερωτήσεις μας, μπορούμε να τα χωρίσουμε σε 2 κατηγορίες, ανάλογα με το είδος τους [18].

- Δεδομένα με έμμεση σχέση με το περιεχόμενο της εικόνας. Τέτοια δεδομένα ονομάζονται μεταδεδομένα ανεξάρτητα από το περιεχόμενο (content-independent metadata). Τέτοια δεδομένα μπορεί να είναι το όνομα του δημιουργού της εικόνας, ημερομηνία, τόπος κλπ.
- Δεδομένα με άμεση σχέση με το περιεχόμενο της εικόνας. Αυτά μπορούμε να τα διακρίνουμε σε δύο γενικές υποκατηγορίες.
    - Χαμηλού-μέσου επιπέδου χαρακτηριστικά της εικόνας, όπως χρώμα, υφή, σχήμα, κλπ. Αυτά ονομάζονται μεταδεδομένα εξαρτημένα από το περιεχόμενο.(content-dependent metadata) Αυτά τα δεδομένα είναι άμεσα και συνειδητά αντιληπτά από τις ανθρώπινες αισθήσεις. Ακριβώς λόγω της σχετικής αμεσότητας τους τα λέμε χαμηλού επιπέδου.
    - Υψηλού επιπέδου χαρακτηριστικά της εικόνας, όπως για ποιο πραγματικά βουνό απεικονίζει μια εικόνα, από ποια πόλη είναι μια φωτογραφία, ή ακόμη τι συναισθήματα προκαλεί. Αυτά τα δεδομένα ονομάζονται μεταδεδομένα περιγράφοντα το περιεχόμενο. (content-descriptive metadata).

Ο τύπος των δεδομένων τα οποία ζητάμε με την ερώτηση μας, επηρεάζει και το τρόπο λειτουργίας του συστήματος ανάκτησης εικόνας. Για τον κάθε τύπο δεδομένων, πρέπει να χρησιμοποιηθεί επομένως και διαφορετικό σύστημα και τεχνικές για την ανάκτηση.

Τέλος να πούμε ότι για το υπόλοιπο αυτής της εργασίας, εικόνες για εμάς θα είναι συγκεκριμένα ορθογώνιες ψηφιοποιημένες ακίνητες εικόνες (2D stills). Η βολικότητα μιας τέτοιας σύμβασης είναι προφανής, εάν σκεφτούμε ότι θα ασχοληθούμε με υπολογιστική ανάλυση και επεξεργασία.

1.4. Προσεγγίσεις στην ανάκτηση εικόνας

1.4.1. Χρησιμοποιώντας τα δεδομένα έμμεσης σχέσης με το περιεχόμενο της εικόνας

Οι πρώτες απόπειρες στην λύση του προβλήματος της ανάκτησης εικόνας είχαν να κάνουν με ερωτήσεις για τα δεδομένα που ονομάσαμε προηγουμένως μεταδεδομένα ανεξάρτητα από το περιεχόμενο. Η ανάκτηση εικόνας γίνεται με την προϋπόθεση ότι οι εικόνες μας έχουν κάποια λεκτική περιγραφή του περιεχομένου τους, όπως το τι απεικονίζει η εικόνα, ποιος την έφτιαξε, πότε και που, ή οτιδήποτε άλλο θα μπορούσε να είναι σχετικό με την εικόνα.

Αφού έχουμε να επεξεργαστούμε κείμενο πλέον, η ανάκτηση εικόνας ουσιαστικά ανάγεται σε ανάκτηση κειμένου. Εάν έχουμε για παράδειγμα μια ερώτηση του τύπου «Σε ποιες εικόνες απεικονίζεται ένα βουνό», φτάνει να αναζητηθεί η λέξη-κλειδί «βουνό» (ή έστω «όρος»). Ερωτήσεις τέτοιου τύπου, δηλαδή για λεκτική πληροφορία, θα μπορούσαν να γίνουν με μηχανές αναζήτησης κειμένου όπως αυτές που χρησιμοποιούνται για τον WWW, ή/ και με γλώσσα ερωτήσεων SQL.

Υπάρχουν όμως κάποια σημαντικά μειονεκτήματα, όταν βασιζόμαστε στις λεκτικές περιγραφές. Κατ' αρχάς είναι πολύ πιθανό να υπάρχει κάποια λεπτομέρεια εύκολα ορατή στην εικόνα που να μην υπάρχει όμως μέσα στην λεκτική περιγραφή της. Αν ισχύει αυτό, τότε η ανάκτηση εικόνας έχει αποτύχει.

Θα μπορούσε κανείς να πει ότι θα ήτανε μια λύση να κρατάμε αναλυτική περιγραφή της κάθε εικόνας. Αλλά εδώ εμφανίζονται πάλι δυσκολίες: Δεν είναι πάντα εύκολο να έχουμε καλές περιγραφές, και ειδικά όταν έχουμε να επεξεργαστούμε μεγάλο αριθμό εικόνων, είναι χρονοβόρο και κουραστικό να γράψουμε καλές περιγραφές για όλες μας τις εικόνες. Ακόμα χειρότερα, σε πολλές περιπτώσεις οι εικόνες που

θέλουμε να περιεργαστούμε μπορεί να παράγονται αυτόματα, όπως για παράδειγμα εικόνες από κάμερες ασφάλειας.

Μια άλλη, βασική, αδυναμία του κειμένου, είναι ότι δεν μπορεί να μας βοηθήσει όταν θέλουμε να εξετάσουμε ομοιότητες (similarities) μεταξύ εικόνων. Δυο εικόνες που είναι φερ' ειπείν αγιογραφίες μπορεί να μοιάζουν μεταξύ τους, αλλά το αν θα φαίνεται αυτό από την λεκτική περιγραφή είναι απίθανο.

Τέλος, οι λεκτικές περιγραφές είναι καταδικασμένες στην υποκειμενικότητα του συγγραφέα τους, που σημαίνει ότι δεν μπορούν να είναι πάντα αξιόπιστες.

Δεν μπορούμε επομένως να βασιστούμε σε λεκτικές περιγραφές των εικόνων, για σωστή ανάκτηση εικόνας. Ενώ είναι μια εύκολη λύση για ορισμένες περιπτώσεις, θα προτιμήσουμε πιο προχωρημένες και πετυχημένες μεθόδους και δεν θα ασχοληθούμε με λεκτικές περιγραφές στη συνέχεια της εργασίας.

1.4.2. Χρησιμοποιώντας τα δεδομένα άμεσης σχέσης με το περιεχόμενο της εικόνας (CBIR)

Με την μέθοδο που θα αναφέρουμε δεν χρησιμοποιούμε λεκτικές περιγραφές, αλλά τώρα υποθέτουμε ότι έχει γίνει κάποια προεπεξεργασία στις εικόνες μας έτσι ώστε να καθοριστούν τα 'χαμηλού επιπέδου' χαρακτηριστικά τους, όπως είναι δηλαδή το χρώμα, το σχήμα και η υφή. Ο ακριβής τρόπος, ή καλύτερα κάποιοι από τους τρόπους που μπορεί να γίνει αυτή η προεπεξεργασία, θα εξηγηθούν καλύτερα σε επόμενο κεφάλαιο.

Το επόμενο βήμα είναι να γίνει η σύγκριση μεταξύ της ερώτησης μας, και των χαρακτηριστικών που υπολογίσαμε για κάθε εικόνα. Η ερώτηση τώρα πρέπει να γίνει δίνοντας στον χρήστη μια επιλογή από οπτικά παραδείγματα, και αυτός καλείται να επιλέξει αυτό που έχει περισσότερο στο νου του. Αυτά τα οπτικά παραδείγματα μπορούν να είναι είτε δείγματα από πραγματικές εικόνες και φωτογραφίες, ή να έχουν δημιουργηθεί επίτηδες για να καλύψουν τον σκοπό της ανάκτησης εικόνας. Για να φέρουμε ένα παράδειγμα, θα μπορούσε ο χρήστης να θέλει να ζητήσει εικόνες που απεικονίζουν μια φουρτουνιασμένη θάλασσα. Οπότε επιλέγει μια εικόνα από το δείγμα, όσο το δυνατόν πιο κοντά στην εικόνα που έχει στο μυαλό του. Αν

υποθέσουμε επίσης ότι το σύστημα ανάκτησης εικόνας δουλεύει συγκρίνοντας χρώματα και υφές (colour, texture), αυτό που θα πρέπει να γίνει είναι να υπολογιστεί το χρώμα της εικόνας του δείγματος, και να συγκριθεί με τους υπολογισμούς που έχουμε για όλες τις εικόνες της βάσης μας. Όσες εικόνες βρίσκονται πιο κοντά στο δείγμα, επιλέγονται και δίνονται σαν απάντηση.

Βεβαίως, είναι πιθανό να μην έχουμε τέλεια επιτυχία στην ανάκτηση. Από τις πέντε φερ' ειπείν φωτογραφίες που θα δοθούν σαν αποτέλεσμα της ανάκτησης, οι τέσσερις μεν μπορεί να απεικονίζουν μια φουρτουνιασμένη θάλασσα, αλλά η μία να απεικονίζει ένα συννεφιασμένο ουρανό – που όντως σαν χρώμα και υφή παρουσιάζει αρκετή ομοιότητα με την φουρτουνιασμένη θάλασσα. Για να εκμεταλλευθούμε το λάθος του συστήματος ανάκτησης εικόνας προς όφελος μας, υπάρχει μια τεχνική που ονομάζεται relevance feedback. Αυτό σημαίνει ότι αν κάποιο αποτέλεσμα της ανάκτησης εικόνας ήταν λάθος, ο χρήστης το δηλώνει στο σύστημα, δίνοντας την ευκαιρία σε αυτό να βελτιώσει το σύστημα σύγκρισης του.

Πάλι θα μπορούσαμε να πούμε ότι ενώ οι ερωτήσεις χρησιμοποιώντας εικόνες-παραδείγματα να έχουν μια σχετικά ικανοποιητική ευελιξία όσον αφορά τα χαμηλού επιπέδου χαρακτηριστικά της εικόνας, ίσως να μην μπορούμε να πούμε το ίδιο όταν έχουμε να κάνουμε με πιο υψηλού επιπέδου χαρακτηριστικά, όπως το συναίσθημα που γεννάει μια εικόνα, ή από ποια χώρα είναι η εικόνα που βλέπουμε. Σε αυτή την περίπτωση οι ερωτήσεις γίνονται μέσω πάλι λεκτικής περιγραφής, όπως «Βρες εικόνες από τοπία στην Υεμένη». Σε αυτές τις περιπτώσεις χρειάζεται κάτι παραπάνω από την πληροφορία που παίρνουμε από το περιεχόμενο της εικόνας αυτό καθ' αυτό – αυτό που χρειάζεται είναι *εμπειρία* του «πραγματικού» κόσμου. Με υψηλού επιπέδου χαρακτηριστικά δεν θα ασχοληθούμε περαιτέρω σε αυτή την εργασία.

1.5. Γενικό πλαίσιο εργασίας για ένα σύστημα CBIR

Η γενική διαδικασία που ακολουθείται σε ένα σύστημα ανάκτησης εικόνας είναι:

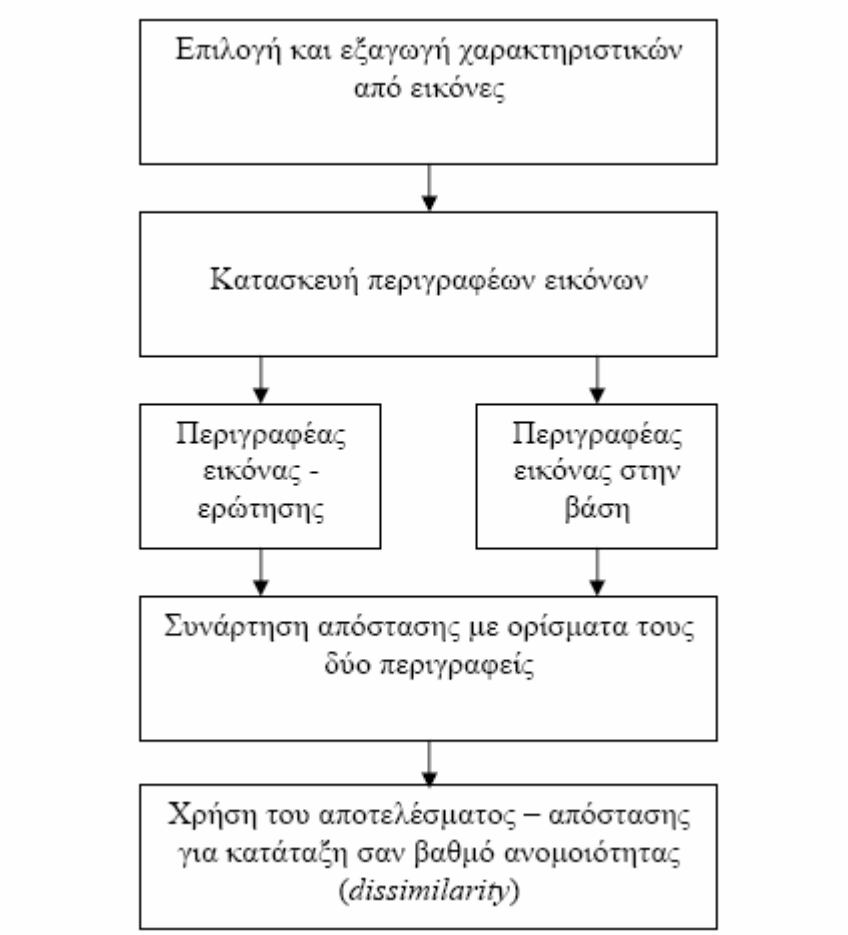
1. «Διάβασε» την ερώτηση από τον χρήστη. Η ερώτηση είναι μια εικόνα ή τμήμα εικόνας.
2. Φτιάξε κατάλληλο περιγραφέα για την εικόνα-ερώτηση. Θα εξηγήσουμε τι εννοούμε «περιγραφέα» αργότερα

3. Σύγκρινε τον περιγραφέα της εικόνας-ερώτηση με τους περιγραφείς *κάθε* εικόνας στην βάση δεδομένων μας. Οι περιγραφείς των εικόνων στην βάση μπορεί να υπολογίζονται κατά τη διάρκεια της ερώτησης (on-line) ή μπορεί να έχουν υπολογιστεί εκ των προτέρων (off-line).
4. Αντιστοίχισε μια απόσταση (πραγματικό αριθμό) για κάθε ζεύγος περιγραφέων <εικόνα ερώτηση – εικόνα στη βάση>.
5. Όσες εικόνες η απόσταση από την εικόνα-ερώτηση είναι μικρότερη από κάποιο προκαθορισμένο κατώφλι, τις επιστρέφουμε σαν απάντηση στην ερώτηση του χρήστη.
6. Επιστροφή στο βήμα 1 για επόμενη ανάκτηση ή τέλος.

Μάλλον χρειάζονται κάποιες εξηγήσεις για τα παραπάνω.

Με τον όρο περιγραφέα λέμε μία δομή αντιπροσωπευτική του περιεχόμενου της εικόνας. Ένα τυπικό παράδειγμα περιγραφέα είναι το χρωματικό ιστόγραμμα [1].

Σύγκριση περιγραφέων λέμε την χρήση μιας συνάρτησης απόστασης μεταξύ δύο περιγραφέων εικόνας – η συνάρτηση αυτή θα πρέπει να δίνει σαν αποτέλεσμα έναν βαθμωτό πραγματικό, μη-αρνητικό αριθμό. Πιο αναλυτικά θα δούμε αυτές τις έννοιες παρακάτω στην εργασία.



Σχήμα 1.1 Η διαδικασία για σύγκριση δύο εικόνων. Τυπικά αυτές θα είναι η εικόνα-ερώτηση με μία-μία εικόνα στη βάση. Ανάλογα γίνεται και η σύγκριση μεταξύ τμημάτων εικόνων.

Συγκεκριμένα, η διάρθρωση της παρούσης εργασίας είναι ως εξής:

Στο κεφάλαιο 2 βλέπουμε τι χαρακτηριστικά μπορούμε να χρησιμοποιήσουμε για να κατασκευάσουμε ένα περιγραφέα για την εικόνα. (feature selection & extraction). Θα δούμε περιγραφείς βασισμένους στο χρώμα, σε υφή και σε τοπολογικά χαρακτηριστικά.

Στο κεφάλαιο 3 κάνουμε μια εισαγωγή σε έννοιες και τεχνικές που θα είναι χρήσιμες για κάνουμε ανάκτηση με στατιστικές – στοχαστικές μεθόδους. Μεταξύ άλλων θα δούμε την έννοια της λύσης μέγιστης πιθανοφάνειας, και τον αλγόριθμο ΕΜ.

Στο κεφάλαιο 4 θα δούμε μεθόδους κατασκευής περιγραφέων εικόνας. Γενικά παρουσιάζουμε δύο κατηγορίες περιγραφέων, τους μεν βασισμένους σε ιστόγραμμα

χαρακτηριστικών της εικόνας, τους δε βασισμένους σε κατασκευή στοχαστικού μοντέλου. Στο τελευταίο βρίσκουν εφαρμογή όσα βλέπουμε στο κεφάλαιο 3.

Στο κεφάλαιο 5 βλέπουμε διάφορες συναρτήσεις απόστασης που έχουν προταθεί, και τρεις γενικούς τρόπους για να κάνουμε ερώτηση για εικόνα. Ο πρώτος είναι να ρωτήσουμε για εικόνες παρόμοιες με ολόκληρο το περιεχόμενο της εικόνας. Ο δεύτερος, είναι να ρωτήσουμε για εικόνες που περιέχουν κάποιο ή κάποια τμήματα της εικόνας ερώτησης. Για παράδειγμα, σε μια εικόνα που απεικονίζει ένα ηλιοβασίλεμα σε θάλασσα, ένα τμήμα θα ήταν ο βυθισμένος ήλιος, ένα ο ουρανός, και ένα η ίδια η θάλασσα. Για να παράγουμε τμήματα εικόνας υπάρχουν διάφορες τεχνικές κατάτμησης (segmentation), μία εκ των οποίων θα δούμε στο ίδιο κεφάλαιο. Τέλος, θα δούμε πως μπορούμε να ρωτήσουμε για τμήμα εικόνας καθορισμένο από τον χρήστη. Αυτό είναι χρήσιμο σε περίπτωση που η αυτόματη κατάτμηση δεν είναι ικανοποιητική.

Στο κεφάλαιο 6 θα δούμε κάποια πειραματικά αποτελέσματα σχετικά με την αποτελεσματικότητα των μεθόδων που παρουσιάζονται στην εργασία, καθώς και σύγκριση μεταξύ των.

ΚΕΦΑΛΑΙΟ 2. ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ ΕΙΚΟΝΑΣ

2.1. Εισαγωγικά

Έχουμε πει ότι θα χρησιμοποιήσουμε το περιεχόμενο της εικόνας για να κάνουμε ανάκτηση εικόνας. Πρώτα όμως πρέπει να βρούμε τρόπους να πάρουμε ωφέλιμη πληροφορία από το περιεχόμενο’ αυτό ακριβώς θα δούμε σε αυτό το κεφάλαιο.

Το χρώμα στην εικόνα είναι η πληροφορία που έχουμε πιο άμεσα, από την σκοπιά της υπολογιστικής επεξεργασίας. Πράγματι, τυπικά όταν λέμε ότι έχουμε αποθηκεύσει μια εικόνα σε ψηφιακό μέσο, αυτό που έχουμε στα χέρια μας είναι το χρώμα για κάθε pixel της εικόνας.

Εικόνα όμως δεν σημαίνει μόνο χρώμα για τον άνθρωπο. Ο ανθρώπινος εγκέφαλος διακρίνει σε μια εικόνα σχήματα και δομές που με τη σειρά τους οδηγούν την σκέψη πιθανώς σε συναισθήματα ή αναμνήσεις. Πέρα από το χρώμα δηλαδή, μπορούμε και πρέπει να εξάγουμε και χαρακτηριστικά όπως υφή, σχήμα, τοπολογία.

Αλλά ας τα δούμε αυτά πιο συγκεκριμένα.

2.2. Χρώμα

2.2.1 Γενικά

Ο απλούστερος τρόπος να αναπαρασταθεί το χρώμα ενός pixel στην ψηφιακή επεξεργασία είναι με έναν φυσικό αριθμό, από 0 ως N όπου N ο αριθμός χρωμάτων που χρησιμοποιείται στην παλέτα της εκάστοτε εικόνας. Υπάρχουν πάντως εναλλακτικά μοντέλα που μπορούν να δώσουν μια πιο ακριβή περιγραφή του χρωματικού ερεθίσματος (color stimulae) [18]:

Χρωματομετρικά μοντέλα. Μετράνε τον βαθμό της απορρόφησης / αντανάκλασης της προσπίπτουσας φωτεινής ακτινοβολίας. Σ' αυτή την κατηγορία υπάγονται διαγράμματα που παρουσιάζουν φάσμα με βαθμούς αντανάκλασης, όπως το CIE chromaticity diagram.

Νευροφυσιολογικά μοντέλα. Βασίζονται σε έρευνες πάνω σε νευροφυσιολογία, συγκεκριμένα με τον τρόπο που λειτουργεί το ανθρώπινο μάτι. Σε αυτό υπάρχουν οι λεγόμενοι 'κώνοι', κύτταρα που ερεθίζονται από ακτινοβολία διαφορετικής φασματικής ο καθένας. Κατά προσέγγιση μπορούμε να τους χωρίσουμε σε *τρεις* κατηγορίες ανάλογα με το χρώμα στο οποίο αντιδρούν. Ένα παράδειγμα τέτοιου μοντέλου είναι το RGB (Red-Green-Blue) μοντέλο.

Ψυχολογικά μοντέλα. Βασίζονται στις αντιδράσεις που προκαλούν τα χρώματα ψυχολογικά στον άνθρωπο. Ένα τέτοιο μοντέλο είναι το μοντέλο hue-saturation-brightness/value (τόνος – ζωντάνια - φωτεινότητα)

2.2.2 RGB χρωματικός χώρος

Η RGB απεικόνιση είναι ο πιο συνηθισμένος τρόπος να εκφραστεί ένα χρώμα σε ψηφιακές εικόνες. Διατηρείται με αυτό συμβατότητα με άλλες συσκευές που κάνουν απεικόνιση εικόνας πέραν του υπολογιστή όπως η τηλεόραση, και είναι ένα μοντέλο

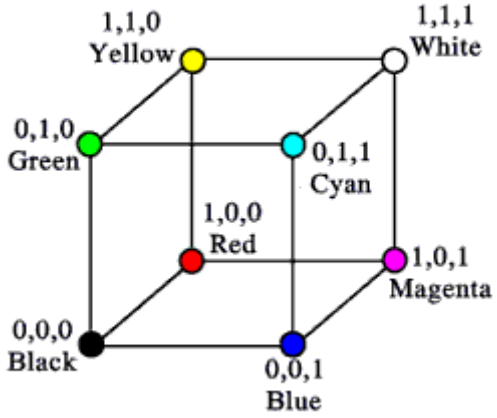
που έχει την βάση του στην φυσιολογία του αμφιβληστροειδούς χιτώνα του ανθρώπου. Στο RGB βλέπουμε το κάθε χρώμα σαν σύνθεση των τριών βασικών του χρωμάτων. Αυτά τα βασικά χρώματα αντιστοιχούν στις περιοχές που κορυφώνεται ο ερεθισμός του εκάστοτε ‘κώνου’ του ματιού.

Μπορούμε να απεικονίσουμε την χρωματική τιμή σε έναν *χρωματικό χώρο*. Αυτός είναι ένας κύβος ακμής μίας μονάδας. Η κάθε διάσταση του κύβου αντιστοιχεί στην τιμή που έχει η κάθε χρωματική συνιστώσα, δηλαδή μία διάσταση για το Κόκκινο, μία για το Πράσινο, μία για το Μπλε. Οι τιμές για την κάθε συνιστώσα κυμαίνονται από 0..1 αφού ο κύβος έχει ακμή μίας μονάδας. Δηλαδή η κάθε χρωματική τιμή αναπαρίσταται σαν ένα διάνυσμα,

$$[\text{ Red value, Green value, Blue Value }]^T$$

Για παράδειγμα ένα χρώμα θα μπορούσε να είναι το $[0.3\ 0.913\ 1]^T$. Στο σχήμα 2.1 μπορούμε να δούμε μια οπτικοποίηση του χρωματικού κύβου.

Ο RGB κύβος έχει κάποιες ενδιαφέρουσες ιδιότητες. Οι κορυφές του κύβου αντιστοιχούν σε χρώματα όπου μεγιστοποιείται η χρωματική ζωντάνια (saturation), και αυτά είναι τα κύρια χρώματα μαζί με τους συνδυασμούς τους, που είναι τα Κυανό, Πορφυρό, Κίτρινο. Επίσης, στη διαγώνιο [0,0,0]-[1,1,1] βρίσκονται τα χρώματα απόχρωσης του γκρίζου (gray scale).



Σχήμα 2.1 Ο Κύβος για τον RGB Χώρο.

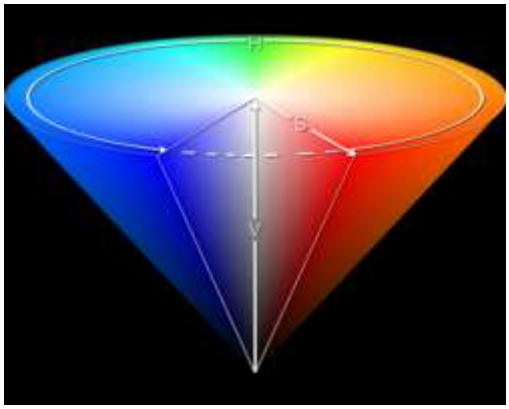
Το μοντέλο CMY (Cyan, Magenta, Yellow) είναι εντελώς αντίστοιχο του RGB. Η διαφορά είναι ότι ενώ στο RGB το χρώμα (0,0,0) αντιπροσωπεύει το μαύρο, ή πλήρης απορρόφηση του φωτός και (1,1,1) αντιπροσωπεύει το λευκό ή πλήρης αντανάκλαση του φωτός, στο CMY μοντέλο ισχύει το αντίστροφο. Κατά μία έννοια δηλαδή αντί να προσθέτουμε φως στο Μαύρο (δηλ 0,0,0) τώρα μετράμε πόσο χρώμα αφαιρούμε από το Λευκό.(δηλ CMY 1,1,1 απορροφάται όλο το χρώμα άρα 1,1,1 είναι το μαύρο). Οι τρεις συνιστώσες αντιστοιχούν στα χρώματα Κυανό (Cyan), Πορφυρό (Magenta) και Κίτρινο (Yellow).
Για μια προσεγγιστική μετατροπή από το ένα μοντέλο στο άλλο υπάρχουν οι απλές εξισώσεις

$$R = 1 - C$$
$$G = 1 - M$$
$$B = 1 - Y$$

Το CMY μοντέλο χρησιμοποιείται περισσότερο στους εκτυπωτές, οι οποίοι κατασκευάζουν το κάθε χρώμα σαν μείξη των Κυανό, Πορφυρό, Κίτρινο.

2.2.3 HSV χρωματικός χώρος

Ο HSV έρχεται σαν εναλλακτική αναπαράσταση του χρώματος. Αντί να δούμε το χρώμα σαν συνδυασμό βασικών χρωμάτων, μετράμε κάποια χαρακτηριστικά του πιο ‘γνώριμα’ στις αισθήσεις. Αυτά είναι το Hue (Τόνος ή απόχρωση), Saturation (Ζωντάνια), Value (Τιμή φωτεινότητας).



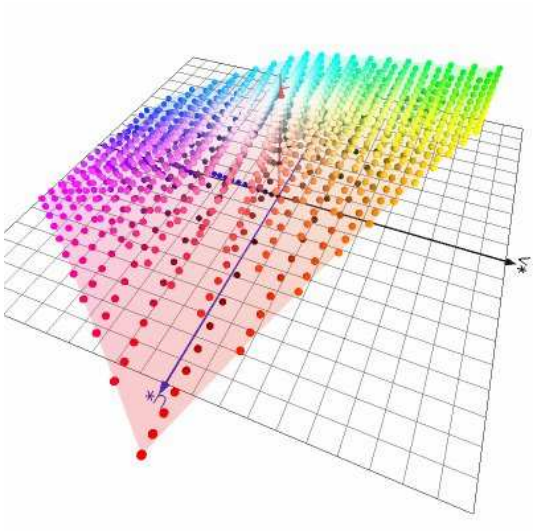
Σχήμα 2.2 Ο HSV Χώρος.

Το μοντέλο HSV ανήκει στην κατηγορία των μη-γραμμικών αναπαραστάσεων [14], σε αντιδιαστολή με το RGB που ανήκει στις γραμμικές αναπαραστάσεις. Αν θέλουμε να δούμε τα χρώματα του HSV σαν σημεία σε γεωμετρικό χώρο, το Hue θα μετράει γωνία σε σχέση με κάποιο χρώμα-σημείο που χρησιμοποιείται σαν αναφορά, το Saturation απόσταση, και το Value ύψος.

2.2.4 Ομοιόμορφοι χώροι

Αυτοί οι χρωματικοί χώροι έχουν μια πολύ ενδιαφέρουσα ιδιότητα: Η Ευκλείδεια απόσταση δύο χρωμάτων εκφρασμένα με κάποιον τρόπο που υπάγεται σε αυτήν την κατηγορία, προσεγγίζει την χρωματική απόσταση όπως την αντιλαμβάνεται ο ανθρώπινος εγκέφαλος [14, 18]. Τέτοιοι χώροι είναι ο L*u*v*, L*a*b*, L*ch. Και αυτοί ανήκουν στις μη-γραμμικές αναπαραστάσεις.

Για παράδειγμα έχουμε τον L*u*v* χώρο. Η L* εκφράζει την φωτεινότητα, όχι με αντικειμενικά μέτρα, αλλά βασιζόμενοι στην ανθρώπινη αντίληψη αυτής της ποιότητας. Τα u*, v* είναι χρωματικές συντεταγμένες. Αντίστοιχα τα a*, b* στον L*a*b*. Μπορούμε να δούμε τον L*u*v* χώρο στο σχήμα 2.3.



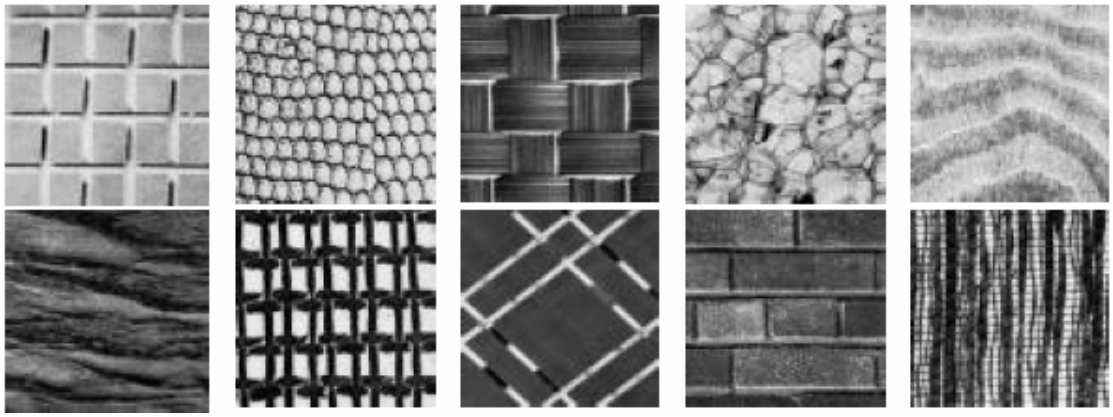
Σχήμα 2.3 Ο L*u*v* ομοιόμορφος χώρος.

Ωστόσο γενικά ο πιο δημοφιλής [14] ομοιόμορφος χώρος είναι ο L*a*b* (CIE 1976 L*a*b*). Το L* κυμαίνεται στο 0..100 και τα a*,b* στα –127..128.
Ο L*a*b* και ο RGB είναι τρόποι αναπαράστασης που συνήθως χρησιμοποιούνται ειδικότερα στο πλαίσιο ανάκτησης εικόνας [1,4,6,8,9,12,14]. Θα τους χρησιμοποιήσουμε κι εμείς στην παρούσα εργασία.

2.3. Υφή

2.3.1 Γενικά

Η υφή (*texture*) είναι μια έννοια που είναι δύσκολο να της δοθεί ένας καλός ορισμός. Το καλύτερο που μπορούμε να κάνουμε είναι να την περιγράψουμε μέσω ιδιοτήτων που έχει, και μέσω παραδειγμάτων υφής. Χαρακτηριστικά βλέπουμε στο σχήμα 2.4 μερικές υφές, από τη συλλογή Brodatz [19].
Συλλογές πολλών όμοιων ή σχεδόν όμοιων αντικειμένων μπορούν να ιδωθούν σαν υφή. Για παράδειγμα το γρασίδι, τρίχες βούρτσας, βότσαλα, μαλλιά, τούβλα.



Σχήμα 2.4. Μερικά παραδείγματα υφών.

Επιπλέον, σαν υφή μπορούν να ιδωθούν επιφάνειες με επαναλαμβανόμενα μοτίβα που *μοιάζουν* με συλλογές όμοιων ή σχεδόν αντικειμένων. Για παράδειγμα η επιφάνεια ενός κορμού ξύλου, το δέρμα ενός ανθρώπου ή ζώου, οι ρίγες στο σώμα μιας ζέβρας, οι βούλες στο σώμα μιας λεοπάρδαλης, τα φτερά μιας πεταλούδας.

Να παρατηρήσουμε ότι η υφή χαρακτηρίζει μια γειτονιά γύρω από ένα pixel, σε αντίθεση με το χρώμα που χαρακτηρίζει ένα ακριβώς pixel. Φερ' ειπείν για να επιβεβαιώσουμε ότι ένα pixel είναι μέρος μιας ρίγας μιας ζέβρας, πρέπει να 'κοιτάξουμε' και γύρω από αυτό.

Παρακάτω θα δούμε μια ομάδα τριών χαρακτηριστικών που μπορούν να χρησιμοποιηθούν για να περιγράψουν υφή. Αυτά είναι τα *polarity*, *anisotropy* και *contrast*, όπως περιγράφονται στο [4]. Βεβαίως δεν είναι οι μόνοι τρόποι για περιγραφή υφής, ούτε και κατηγορηματικά οι καλύτεροι. Βλέπε [14, 18] για μια επισκόπηση στο πρόβλημα περιγραφής υφής.

2.3.2 Polarity & επιλογή κλίμακας

Έστω εικόνα mxn διάστασης. Το L^* συστατικό της ως προς χώρο $L^*a^*b^*$ αποτελεί πραγματικό πίνακα mxn, έστω το όνομα του L. Κάθε στοιχείο είναι στο εύρος 0..100.

Έστω *J*, πίνακας mxn. Κάθε στοιχείο του ορίζεται σαν η Ιακωβιανή του *L* στο αντίστοιχο στοιχείο. Στην πράξη χρησιμοποιούμε προσέγγιση με πρώτες διαφορές. Δηλαδή

$$J_{i,j} \triangleq \begin{bmatrix} \left.\frac{\partial L}{\partial x}\right|_{i,j} \\ \left.\frac{\partial L}{\partial y}\right|_{i,j} \end{bmatrix} \approx \begin{bmatrix} L_{i,j+1} - L_{i,j} \\ L_{i+1,j} - L_{i,j} \end{bmatrix}$$

Έστω *H*, πίνακας mxn. Κάθε στοιχείο του ορίζεται σαν το εξωτερικό γινόμενο

$$H_{i,j} \triangleq J_{i,j} J_{i,j}^T$$

Δηλαδή κάθε στοιχείο του *J* είναι πραγματικό διάνυσμα $2x1$ και κάθε στοιχείο του *H* είναι συμμετρικός πραγματικός πίνακας $2x2$.

Ορίζουμε

$$M_\sigma = G_\sigma * H \quad (2.1)$$

όπου ο αστερίσκος σημαίνει πράξη συνέλιξης. G_σ είναι προσέγγιση σε *Gaussian smoothing* πυρήνα 2D (κανονικός πυρήνας 2D), κανονικής απόκλισης *σ* κατά x και κατά y. Με άλλα λόγια ο M_σ είναι ένας εξομαλυσμένος *H*. Επιπλέον, το σ θα δούμε ότι στην πράξη είναι μεταβλητό από σημείο σε σημείο (οπότε η πράξη δεν είναι συνέλιξη με την αυστηρή έννοια).

Από τον πίνακα M_σ μπορούμε να περάσουμε σε υπολογισμό των χαρακτηριστικών υφής που ζητάμε (polarity, anisotropy, contrast). Πρώτα πρέπει να υπολογίσουμε την σωστή κλίμακα *σ*. Αυτή αντιστοιχεί στην γειτονιά γύρω από κάθε σημείο της εικόνας, την οποία θα χρησιμοποιήσουμε για να εκτιμήσουμε χαρακτηριστικά υφής, όπως αναφέραμε και στην παράγραφο γενικά για τις υφές.

Πρώτα πρέπει να εισάγουμε τον ορισμό του polarity. Για κάθε σημείο i, j της εικόνας έχουμε

$$p_{i,j} = \frac{\left|\mathrm{E}_{i,j,+} - \mathrm{E}_{i,j,-}\right|}{\mathrm{E}_{i,j,+} + \mathrm{E}_{i,j,-}} \quad (2.2)$$

όπου

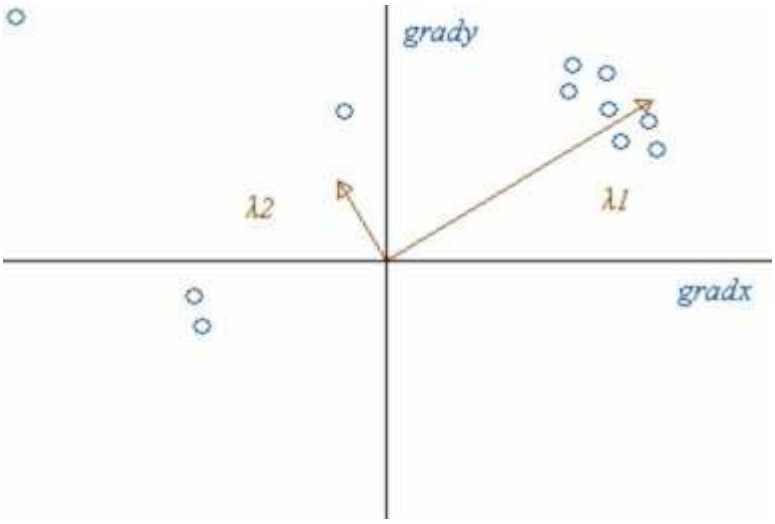
$$E_{i,j,+} = G_{i,j} * \left[J^{\phi_{i,j}} \right]_+ \text{ (συνέλιξη 2-D)}$$

$$E_{i,j,-} = G_{i,j} * \left[J^{\phi_{i,j}} \right]_- \text{ (συνέλιξη 2-D)}$$

,κάθε σημείο (x,y) του $J^{\phi_{i,j}}$ ορίζεται

$$J^{\phi_{i,j}}_{x,y} = J^{T}_{x,y}\phi_{i,j},$$

$G_{i,j}$ είναι κανονικός πυρήνας 2D με κανονική απόκλιση $\sigma_{i,j}$, $[A]_{+}$ μηδενίζει τα μη-θετικά στοιχεία και $[A]_{-}$ μηδενίζει τα μη-αρνητικά στοιχεία πίνακα A (θετική / αρνητική ανόρθωση). $\phi_{i,j}$ ορίζουμε το κυρίαρχο ιδιοδιάνυσμα του στοιχείου *i,j* του M_{σ} (θυμηθείτε ότι τα στοιχεία του M_{σ} είναι προϊόν συνέλιξης με Gaussian 2D πυρήνα, επομένως ζυγισμένο άθροισμα πινάκων $2x2$). Κυρίαρχο ιδιοδιάνυσμα λέμε το ιδιοδιάνυσμα που αντιστοιχεί στην μεγαλύτερη ιδιοτιμή.
Στο σχήμα 2.5 μπορούμε να δούμε ένα παράδειγμα για το πώς θα είναι η ιδιοδομή $H_{i,j}$ για κάποιο pixel. Υποθέτουμε μόνο 10 γειτονικά pixel και αγνοούμε το βάρος που δίνεται στην καθεμία Ιακωβιανή, για απλότητα.



Σχήμα 2.5. Κάθε γαλάζιος κύκλος παριστάνει μια από τις ιακωβιανές στο Gaussian παράθυρο. Οι ευθείες παριστάνουν τα ιδιοδιανύσματα του $H_{i,j}$. Το μέτρο τους είναι ανάλογο των αντίστοιχων ιδιοτιμών. Η polarity εδώ είναι γύρω στο 0.4.

Η polarity μεταβάλλεται συναρτήσει της κλίμακας σ, και παίρνει τιμές στο [0..1]. Ο τρόπος της μεταβολής μπορεί να μας δώσει πληροφορία για την γειτονιά κάθε pixel:

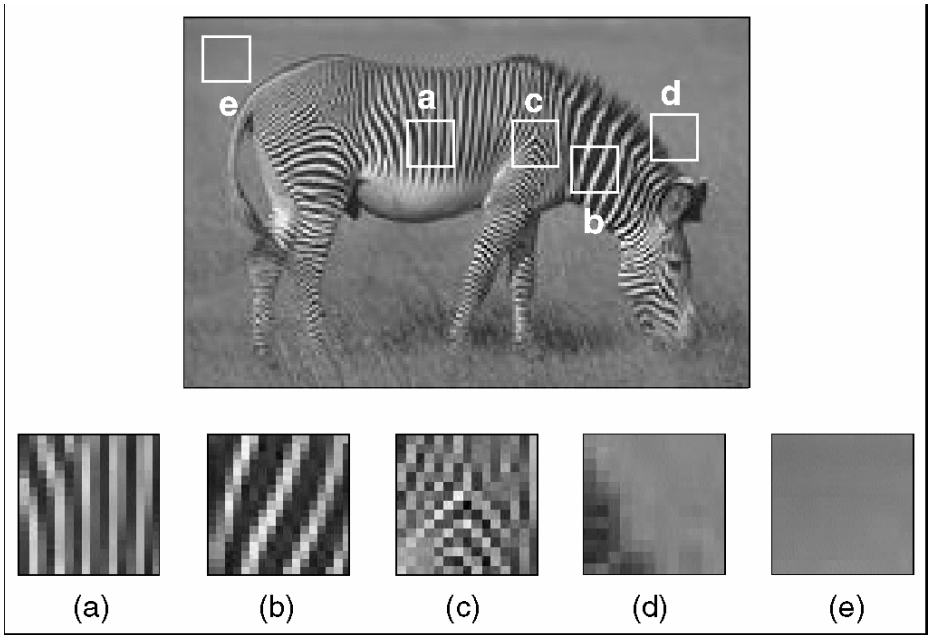
- Σε ακμή: Η polarity παίρνει τιμές κοντά στη μονάδα για κάθε σ, όταν το pixel είναι πάνω σε ακμή.
- Σε υφή: Η polarity μειώνεται όσο αυξάνει η κλίμακα σ. Αυξάνοντας το *σ*, προσμετρώνται ιακωβιανές (στοιχεία του πίνακα ιακωβιανών *J*) σε διάφορες κατευθύνσεις, και οι κυριαρχία της μιας μόνο κατεύθυνσης συνεπώς φθίνει.

- Σε ομοιόμορφη περιοχή: Δεν υπάρχει κυρίαρχη κατεύθυνση. Οι ιδιοτιμές παίρνουν μικρές τιμές, αφού δεν υπάρχουν ακμές.

Οι παραπάνω παρατηρήσεις οδηγούν σε ένα τρόπο επιλογής της σωστής κλίμακας. Συγκεκριμένα, υπολογίζουμε την polarity για κάθε pixel στην εικόνα, και για κάθε κλίμακα στο εύρος

$$\{0, \frac{1}{2}, 1, \frac{3}{2}, 2, \frac{5}{2}, 3, \frac{7}{2}\}$$

Κάθε ένα από αυτούς τους πίνακες polarity τους εξομαλύνουμε με κανονικό πυρήνα κανονικής απόκλισης δυο φορές όσο η αντίστοιχη κλίμακα.



Σχήμα 2.6. Χαρακτηριστικά δείγματα περιοχών σε εικόνα με υφή. Οι κλίμακες υφής: (a) σ = 1.5 (b) σ = 2.5 (c) σ = 1.5 (d) Ακμή, σ = 0 (e) Ομοιόμορφη περιοχή, σ = 0. (*Σχήμα παρμένο από* [4])

Οπότε έχουμε τώρα για κάθε pixel μια ακολουθία 8 τιμών polarity, έστω $\{p_k\}_{k=0}^{7}$. Αρχίζοντας από $k = 0$, επαναληπτικά ως $k = 7$

επιλέγουμε σαν σωστή κλίμακα $\sigma = \frac{k}{2}$ και σταματάμε, όταν $\frac{p_k - p_{k-1}}{p_k} < 2\%$

Το αυθαίρετο της επιλογής να αναζητήσουμε κλίμακες μέχρι $\frac{7}{2}$ pixels, σημαίνει ότι αυτός ο αλγόριθμος θα αποτυχαίνει για εικόνες μεγάλες ανάλυσης, όπου η σωστή

κλίμακα μπορεί να είναι αρκετά pixels μεγάλη. Ωστόσο δεν μπορούμε από την άλλη μεριά να αναζητάμε οσοδήποτε μεγάλες κλίμακες, αφού η υφή χαρακτηρίζει εξ ορισμού μια γειτονιά pixels.

Οπότε έχουμε επιλέξει κλίμακα για κάθε pixel, $\sigma^*(x, y)$ αφού δεν είναι πλέον σταθερή για όλα τα pixel. Θα ονομάσουμε M^* τον πίνακα για τον υπολογισμό του οποίου χρησιμοποιήσαμε τον (2.1) και την *μεταβλητή* τώρα κλίμακα $\sigma^*(x, y)$. Αντίστοιχα ορίζουμε πίνακα ιακωβιανών J^* και πίνακα εξωτερικών γινομένων H^*.

Χρησιμοποιώντας λοιπόν τον J^* για τον τύπο (2.2), υπολογίζουμε πλέον και κρατάμε την polarity με τη σωστή κλίμακα.

2.3.3 Anisotropy και Contrast

Αυτά τα χαρακτηριστικά παράγονται εύκολα από τον πίνακα H^*. Έστω $\boldsymbol{\lambda}^1, \boldsymbol{\lambda}^2$ πίνακες που τα στοιχεία τους είναι αντίστοιχα το μεγαλύτερο και το μικρότερο ιδιοδιάνυσμα του αντίστοιχου στοιχείου του H^*.

Τότε η anisotropy για το pixel στο i, j ορίζεται σαν

$$a_{ij} = 1 - \frac{\boldsymbol{\lambda}_{ij}^2}{\boldsymbol{\lambda}_{ij}^1} \quad (2.3)$$

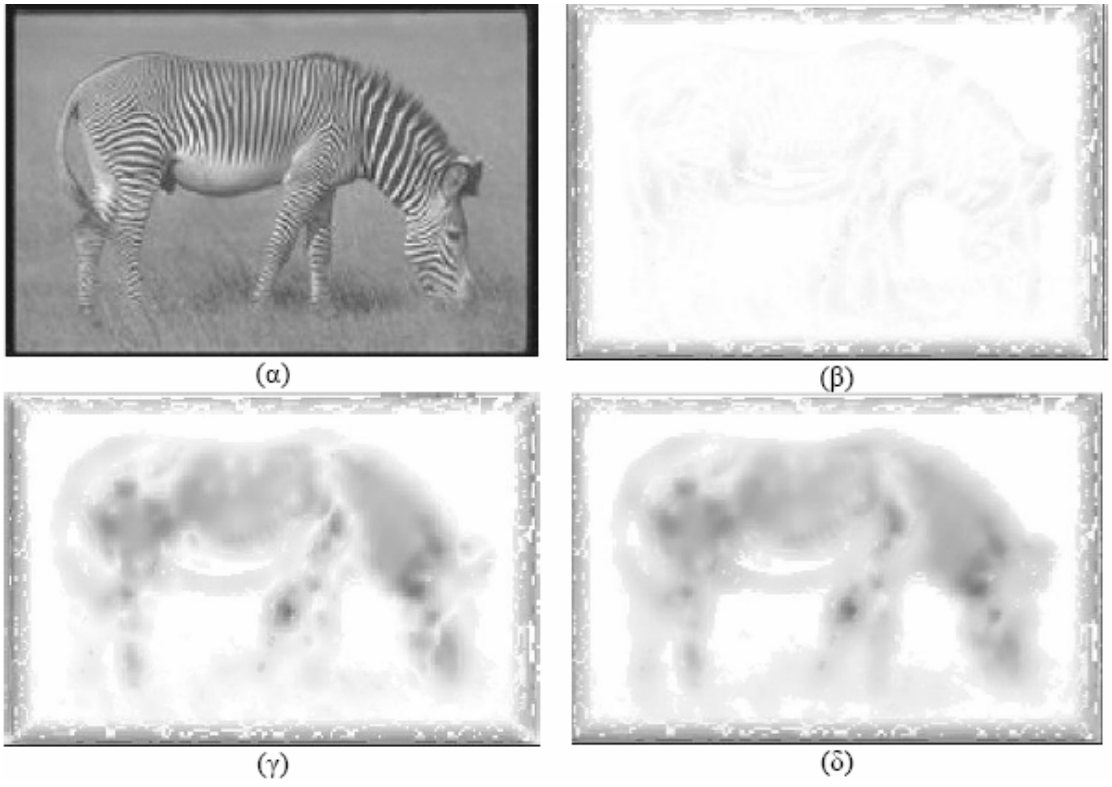
Το anisotropy θα παίρνει μεγάλες τιμές σε γειτονιές pixel όπου επαναλαμβάνονται ακμές προς μία κατεύθυνση – όπως συμβαίνει στη μέση της ζέβρας, [σχήμα 2.7(α)], όπου μία κατακόρυφη ρίγα διαδέχεται την άλλη. Στις περιοχές όπου αλλάζει η κατεύθυνση που προχωρούν οι ρίγες, η anisotropy μειώνεται, αφού εκεί θα υπάρχουν ακμές προς πολλαπλές κατευθύνσεις.

Το contrast ορίζεται σαν

$$c_{ij} = 2\sqrt{\boldsymbol{\lambda}_{ij}^1 + \boldsymbol{\lambda}_{ij}^2} \quad (2.4)$$

Το contrast θα παίρνει μεγάλες τιμές σε μη-ομοιόμορφες περιοχές. Σε ομοιόμορφες περιοχές οι διαφορές προφανώς είναι μικρές, επομένως και οι ιδιοτιμές $\boldsymbol{\lambda}^1, \boldsymbol{\lambda}^2$ και το contrast θα είναι μικρά. Η τετραγωνική ρίζα και ο παράγοντας 2 στον τύπο 2.4 υπάρχουν για λόγους κανονικοποίησης.

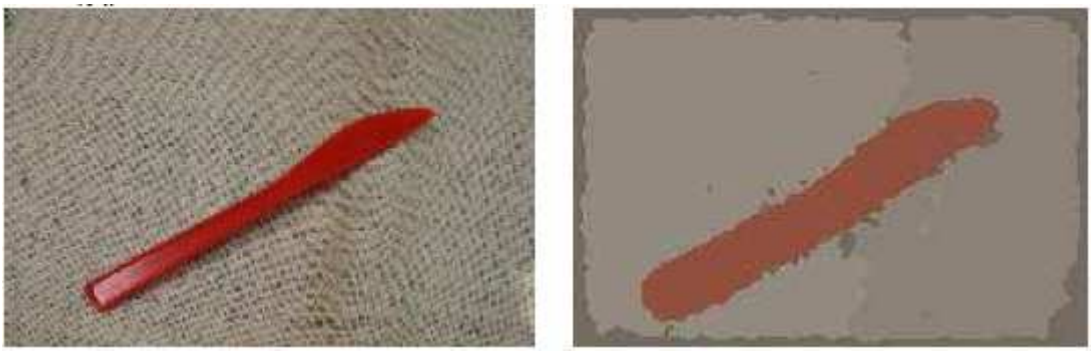
Η παραπάνω παρατηρήσεις για τα anisotropy και contrast μπορούν να επιβεβαιωθούν στο σχήμα 2.7.



Σχήμα 2.7. Χαρακτηριστικά υφής. (α) Το κανάλι L*της εικόνας προς επεξεργασία. Υφή φαίνεται εύκολα στο δέρμα της ζέβρας και λιγότερο στο γρασίδι. (β) Εικόνα polarity (γ) Εικόνα anisotropy (δ) Εικόνα contrast. Η φωτεινότητα σε κάθε pixel είναι αντιστρόφως ανάλογη του μέτρου κάθε χαρακτηριστικού (*Σχήματα παρμένα από* [4])

2.4. Τοπολογικά χαρακτηριστικά

Είναι κοινή πρακτική να συμπεριλαμβάνονται στα χαρακτηριστικά της εικόνας, και οι συντεταγμένες κάθε pixel [4, 9, 14]. Αυτός είναι ένας τρόπος να βάζουμε την απαίτηση ότι γειτονικά στον χώρο pixels πρέπει να είναι παρόμοια, ή όπως αλλιώς θα λέγαμε, είναι πολύ πιθανό να ανήκουν στο ίδιο segment-τμήμα της εικόνας. (βλέπε κεφάλαιο 5 για περισσότερα περί segmentation). Ωστόσο, είναι σύνηθες φαινόμενο να οδηγούμαστε σε υπέρ-κατακερματισμό της εικόνας όταν χρησιμοποιούνται οι συντεταγμένες pixel σαν χαρακτηριστικό, όπως για παράδειγμα στο σχήμα 2.8.



Σχήμα 2.8. Παράδειγμα υπερκατακερματισμού, λόγω χρήσης χαρακτηριστικών τοπολογίας στην

ΚΕΦΑΛΑΙΟ 3. ΕΚΤΙΜΗΣΗ ΠΥΚΝΟΤΗΤΑΣ ΠΙΘΑΝΟΤΗΤΑΣ ΜΕ ΜΕΓΙΣΤΗ ΠΙΘΑΝΟΦΑΝΕΙΑ

3.1. Εισαγωγικά

Αυτό το κεφάλαιο είναι εμβόλιμο ανάμεσα στο κεφάλαιο για την εξαγωγή χαρακτηριστικών, και τους τρόπους περιγραφής εικόνας’ αυτά τα δύο τελευταία είναι μεταξύ τους συνέχεια το ένα του άλλου. Όμως θα πρέπει να σταθούμε για λίγο πριν συνεχίσουμε, και να δούμε κάποιες σημαντικές ιδέες, που θα φανούν χρήσιμες έως απαραίτητες για τα επόμενα κεφάλαια.

3.2. Εκτίμηση μέγιστης πιθανοφάνειας

Έστω συνάρτηση πυκνότητας πιθανότητας

$$p(x;\Psi)$$

με x την τυχαία μεταβλητή και Ψ η παράμετρος που καθορίζει ποια μορφή έχει η p. Για παράδειγμα, p είναι η κατανομή Poisson και Ψ η αναμενόμενη τιμή λ. Υπενθυμίζουμε ότι οι απαιτήσεις για είναι μια συνάρτηση p κατανομή πιθανότητας, είναι όταν αυτή ορίζεται σε διακριτό δειγματικό χώρο Ω:

- $0 \le p(x) \le 1, \ \forall x \in \Omega$
- $\sum_{x \in \Omega} p(x) = 1$

Και όταν ορίζεται σε συνεχή δειγματικό χώρο Ω:

- $p(x) \ge 0, \ \forall x \in \Omega$
- $\int_{x \in \Omega} p(x)dx = 1$

Παρακάτω τα αποτελέσματα που δίνουμε είναι γενικά, είτε για τη συνεχή είτε για τη διακριτή περίπτωση. Αν χρειαστεί να είμαστε πιο ειδικοί, θα το δηλώσουμε ρητά.

Έστω τώρα ότι παράγονται n δείγματα από την κατανομή p. Δηλαδή η πιθανότητα εμφάνισης τους είναι

$$p(x_1, x_2, ..., x_N; \Psi)$$

ή για συντομία γράφουμε

$$p(X;\Psi) \quad (3.1)$$

με $X = [x_1, x_2, ..., x_N]$.

Την $p(X;\Psi)$ θα μπορούσαμε να την υπολογίσουμε αν ξέρουμε την ακριβή μορφή της κατανομής της πηγής πληροφορίας' με άλλα λόγια αν ξέραμε την παράμετρο Ψ.

Αυτό που μπορούμε να κάνουμε, είναι να μαντέψουμε ότι η πραγματική τιμή του Ψ πρέπει να είναι αυτή για την οποία η πιθανότητα εμφάνισης των δεδομένων μας είναι η καλύτερη, δηλαδή η μέγιστη. Πιο ακριβές μάλλον θα ήταν να πούμε ότι αυτή θα είναι η πιθανότερη να είναι η πραγματική τιμή του Ψ.

Αυτή η εκτίμηση λέγεται **εκτίμηση Μέγιστης Πιθανοφάνειας**, ή **Maximum Likelihood Estimate** (ML estimate, MLE), και ορίζεται σαν

$$\hat{\Psi}_{MLE} \triangleq \arg\max_{\Psi} p(\mathrm{X};\Psi)$$

Για να τονίσουμε το γεγονός ότι η μεγιστοποιητέα μεταβλητή είναι η Ψ, ενώ το X είναι κάποια συγκεκριμένα δεδομένα, γράφουμε ισοδύναμα $\hat{\Psi}_{MLE} = \arg\max_{\Psi} L(\Psi; X)$ με $L(\Psi; X) \triangleq p(X;\Psi)$, ή ακόμα παραλείποντας τα δεδομένα X από τον τύπο του L, τα οποία υπονοούνται. Δηλαδή $L(\Psi) \triangleq p(X;\Psi)$. Πολλές φορές επίσης βολεύει να μεγιστοποιήσουμε την $\log p(X;\Psi)$ ισοδύναμα (αφού η log είναι γνησίως αύξουσα). Οπότε γράφουμε ολοκληρωμένα

$$\hat{\Psi}_{MLE} \triangleq \arg\max_{\Psi} p(X;\Psi) = \arg\max_{\Psi} \log p(X;\Psi) = \arg\max_{\Psi} L(\Psi) \quad (3.2)$$

και τελικά κρατάμε για την συνάρτηση L τον ορισμό

$$L(\Psi) \triangleq \log p(X;\Psi) \quad (3.3)$$

Η συνάρτηση L ονομάζεται likelihood (πιθανοφάνεια) ή log-likelihood. Για τη συνέχεια αυτής της εργασίας θα χρησιμοποιούμε τον ορισμό (3.3), και σε αυτόν θα αναφερόμαστε είτε με τον ένα είτε τον άλλο όρο, προς αποφυγή σύγχυσης.

Όσο για την παράμετρο Ψ, είναι σαφές ότι δεν είναι ανάγκη να είναι πάντα ένα μόνο βαθμωτό μέγεθος. Γενικά θα θεωρούμε με Ψ ένα πεπερασμένο σύνολο παραμέτρων $[\psi_1, \psi_2, ..., \psi_m]$.

3.3. Μέγιστη εκ των υστέρων εκτίμηση

Το σύνολο παραμέτρων Ψ μπορεί εναλλακτικά να αντιμετωπιστεί σαν τυχαία μεταβλητή και αυτό, δηλαδή να υπάρχει κάποια έκφραση $p(\Psi)$. Τότε εφαρμόζοντας το νόμο του Bayes, μπορούμε να υπολογίσουμε την μορφή $p(\Psi \mid X)$:

$$p(\Psi \mid X) = \frac{p(X \mid \Psi)\, p(\Psi)}{p(X)} \quad (3.4)$$

Η έκφραση $p(\Psi)$ ονομάζεται και **prior** density ή πρότερη πυκνότητα της Ψ, εκφράζοντας έτσι την πληροφορία που έχουμε για την Ψ *χωρίς* (ή «πριν») να έχουμε γνώση των δεδομένων X. Η έκφραση $p(\Psi \mid X)$ ονομάζεται συμμετρικά, **posterior** density ή ύστερη πυκνότητα της Ψ, δηλαδή η πληροφορία που έχουμε για την Ψ *με* (ή «ύστερα από») γνώση των X.

Αυτό που μας λέει η $p(\Psi = \psi \mid X)$, είναι: «δεδομένου του X, η ψ έχει τόση πιθανότητα να είναι η πραγματική τιμή της παραμέτρου». Οπότε θα ήταν λογικό να εκτιμήσουμε την Ψ σαν

$$\hat{\Psi}_{MAP} \triangleq \arg\max_{\Psi} p(\Psi \mid X)$$

Κάνοντας χρήση του τύπου (3.4), απορρίπτοντας τον ανεξάρτητο από το Ψ όρο $p(X)$ και λογαριθμίζοντας, έχουμε ισοδύναμα

$$\hat{\Psi}_{MAP} \triangleq \arg\max_{\Psi} \{\log p(X \mid \Psi) + \log p(\Psi)\} \quad (3.4)$$

Αυτή η εκτίμηση ονομάζεται εκτίμηση εκ των υστέρων ή Maximum A-Posteriori estimate (MAP estimate).

Βλέπουμε ότι ο τύπος που παίρνουμε είναι ανάλογος του τύπου για την μέγιστη πιθανοφάνεια, (3.2). Στον (3.4) η Ψ είναι τυχαία μεταβλητή και υπάρχει και ο επιπλέον όρος $\log p(\Psi)$. Αυτός ο όρος λειτουργεί σαν όρος ποινής, αφού για εκ των προτέρων πιθανά Ψ ($p(\Psi) \to +\infty$) συνεισφέρει θετικά, ενώ για εκ των προτέρων απίθανα Ψ ($p(\Psi) \to 0$) συνεισφέρει αρνητικά. Σε αναλογία με την εκτίμηση μέγιστης πιθανοφάνειας, μπορούμε να ονομάσουμε τον όρο $\log p(X \mid \Psi)$ πιθανοφάνεια, επίσης [10].

Στην περίπτωση που δεν έχουμε καμμία πρότερη γνώση για το Ψ, απορρίπτουμε και τον όρο $\log p(\Psi)$, αφού θα είναι σταθερός για όλες τις δυνατές τιμές του Ψ. Οπότε η εκτίμηση που παίρνουμε είναι $\arg\max_{\Psi} \log p(X \mid \Psi)$, που συμπίπτει με την εκτίμηση μέγιστης πιθανοφάνειας.

Επομένως, εάν έχουμε εκ των προτέρων γνώση για τις εκτιμητέες παραμέτρους, μπορούμε και καλύτερο είναι να χρησιμοποιήσουμε τον MAP εκτιμητή. Βεβαίως αυτό όταν η prior $p(\Psi)$ που διαθέτουμε είναι όντως σωστή και δεν κάνουμε λάθος

υποθέσεις για την μορφή της! Σε ένα τέτοιο ενδεχόμενο μιας κακής prior, τα αποτελέσματα μπορεί να είναι καταστροφικά, δίνοντας εξίσου κακή εκτίμηση.

3.4. Εκπαίδευση του μοντέλου

Έστω λοιπόν έχουμε σύνολο δεδομένων X. Κάνουμε την υπόθεση ότι αυτά παράγονται από πηγή οι οποία γεννάει δεδομένα, των οποίων η εμφάνιση ακολουθεί *κάποια* πυκνότητα πιθανότητας $p(X;\Psi)$. Η διαδικασία ακριβώς της εκτίμησης των παραμέτρων μιας πυκνότητας πιθανότητας (στο εξής pdf), δοθέντος συνόλου δεδομένων X, λέγεται εκπαίδευση του **μοντέλου**. Με τον όρο μοντέλο εννοούμε την γενική μορφή της pdf που έχουμε, χωρίς δηλαδή να συγκεκριμενοποιήσουμε τις παραμέτρους. Αν ας πούμε

$$p(x;\lambda) = e^{-\lambda}\frac{\lambda^x}{x!},\ x \in \mathbb{R}$$

έχουμε να κάνουμε μοντέλο Poisson. Σε αυτή την περίπτωση δηλαδή έχουμε ένα μόνο δεδομένο, την βαθμωτή τιμή x.

Πριν προχωρήσουμε, να κάνουμε κάποιες διευκρινίσεις. Στις μεθόδους που θα παρουσιαστούν στα επόμενα κεφάλαια της εργασίας, δηλαδή τις εφαρμογές στην ανάκτηση εικόνας, δεν υποθέτουμε καμμία πρότερη γνώση (prior) για τις παραμέτρους προς εκτίμηση. Φυσικά αυτό δεν σημαίνει ότι δεν μπορούμε στην πράξη να κάνουμε τέτοιες υποθέσεις, π.χ. [12]. Οπότε λέγοντας εκτίμηση παραμέτρων, θα υπονοούμε εκτίμηση μέγιστης πιθανοφάνειας, που είδαμε στην παράγραφο 3.2.

Η δεύτερη διευκρίνιση αφορά σε μια υπόθεση που κάνουμε για τα δεδομένα. Κατά κανόνα υποθέτουμε ότι τα δεδομένα που έχουμε

- είναι στοχαστικά ανεξάρτητα το ένα από το άλλο. (independent) Δηλαδή αν γνωρίζουμε την τιμή του ενός από τα δεδομένα, δεν κερδίζουμε κάποια παραπάνω γνώση για το ποιες θα πρέπει να είναι οι τιμές των υπόλοιπων δεδομένων (αυτό μπορεί να φανεί κάπως αντιφατικό με τα προηγούμενα,

αφού πάντα υποθέταμε ότι είναι γνωστό από την ‘αρχή’ ένα ολόκληρο data set. Ας φανταστούμε εδώ την περίπτωση να έχουμε σειριακή άφιξη των δεδομένων, δηλαδή να λαμβάνουμε γνώση των δεδομένων το ένα μετά το άλλο.)

- ακολουθούν όλα την ίδια πυκνότητα πιθανότητας. (identically distributed).

Λέμε με άλλα λόγια ότι είναι ανεξάρτητα, ομοίως κατανεμημένα (independent, identically distributed) ή πιο σύντομα **iid**. Άλλες φορές πράγματι ισχύει μια τέτοια υπόθεση, όπως στην περίπτωση ρίψεων του ίδιου νομίσματος (δοκιμές Bernoulli). Άλλες φορές δεν ισχύει απόλυτα, ή δεν είμαστε σίγουροι αν ισχύει. Όπως και να έχει, κάνοντας iid υπόθεση διευκολύνεται η διαδικασία της εκτίμησης, και αυτός είναι ο λόγος που την κάνουμε.

Ξαναγυρνάμε στο μοντέλο Poisson για ένα ακόμα παράδειγμα. Έστω τώρα data set $X = [x_1, x_2, ..., x_n]$, με $x_i \in \mathbb{R}, \forall i \in [1, n]$, και iid με $x_i \sim Poisson(\lambda)$.Τότε

$$p(X;\lambda) = p(x_1, x_2, ..., x_n; \lambda) = \prod_{i=1}^{n} p(x_i;\lambda) \quad (3.5)$$

και πιο αναλυτικά

$$p(X;\lambda) = \prod_{i=1}^{n} e^{-\lambda} \frac{\lambda^{x_i}}{x_i!} = e^{-\lambda n} \prod_{i=1}^{n} \frac{\lambda^{x_i}}{x_i!}$$

Παίρνουμε τώρα λογαρίθμους, μεγιστοποιούμε ως προς λ μηδενίζοντας την πρώτη παράγωγο, και έχουμε την εκτίμηση μέγιστης πιθανοφάνειας

$$\hat{\lambda}_{MLE} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

3.5. Επιλογή του μοντέλου

3.5.1 Επιλογή του μοντέλου, (α)

Μιλήσαμε για εκτίμηση παραμέτρων ενός μοντέλου, όμως αποσιωπήσαμε τον τρόπο με τον οποίο επιλέγουμε να χρησιμοποιήσουμε το ένα ή το άλλο μοντέλο.

Στην πράξη, όταν έχουμε ένα πρόβλημα εκτίμησης πυκνότητας, δεν μας είναι μόνο γνωστό ένα σύνολο δεδομένων X, αλλά έχουμε και κάποια πληροφορία για το είδος της πηγής των δεδομένων. Επιπλέον, στατιστικές παρατηρήσεις μας βοηθάνε να ξέρουμε ποιο μοντέλο ταιριάζει να χρησιμοποιηθεί για το εκάστοτε φαινόμενο. Μερικά παραδείγματα ακολουθούν:

- *Ρίψεις νομίσματος*. Ο αριθμός ρίψεων δύο ζαριών, μέχρι την πρώτη φορά που θα φέρουμε εξάρες, ακολουθεί Γεωμετρική κατανομή.
- *Αποσύνθεση ραδιενεργούς υλικού*. Ένα ραδιενεργό υλικό εκπέμπει σωμάτια άλφα. Ο αριθμός των σωματίων άλφα που φθάνουν σε συγκεκριμένο σημείο του χώρου μέσα σε χρόνο t, ακολουθεί κατανομή Poisson. (Αυτό ισχύει για μικρά περιθώρια χρόνου, αφού μακροπρόθεσμα η πυκνότητα των άλφα σωματίων θα φθίνει). [13, vol. I]
- *Αφίξεις δεδομένων σε server*. Ο αριθμός πακέτων που λαμβάνει ένας server σε σταθερά χρονικά διαστήματα, ακολουθεί κατανομή Poisson.
- *Βομβαρδισμός του Λονδίνου*. Τα τελευταία χρόνια του Β' Παγκοσμίου πολέμου, το Λονδίνο δεχόταν βομβαρδισμούς από γερμανικές ρουκέτες V-1. Ο αριθμός ρουκετών που έπεσαν σε μια τυχαία περιοχή σταθερού εμβαδού, αν χωρίσουμε την πόλη σε N ίσες περιοχές, ακολουθεί κατανομή Poisson[2]. [13, vol. I]
- *Πυρηνική Φυσική*. Ο τρόπος που κατανέμεται ο αριθμός ενός συνόλου σωματιδίων σε ένα πλήθος ενεργειακών καταστάσεων, ακολουθούν κατανομές Bose-Einstein ή Fermi-Dirac, αναλόγως το είδος των σωματιδίων. Π.χ. τα φωτόνια ακολουθούν Bose-Einstein, τα ηλεκτρόνια ακολουθούν Fermi-Dirac. [13, vol. I].
- *Ανακατασκευή τομογραφίας*. Ο αριθμός των σωματιδίων που λαμβάνονται από ένα ανιχνευτή που σαρώνει το σώμα ασθενούς, κατά τομογραφία PET (Positron Emission Tomography) ή SPECT (Single Photon Emission

[2] Ειρήσθω εν παρόδω, αυτό είναι ένδειξη ότι δεν επλήγησαν συγκεκριμένες μόνο περιοχές 'στρατηγικής σημασίας' στο Λονδίνο, αλλά ολόκληρη η πόλη ομοιογενώς.

Computerized Tomography), μοντελοποιείται σαν άθροισμα Poisson. [10] (λεγόμενη σύνθετη ή compound Poisson)

- *Θεωρία θορύβου*. Ο θόρυβος σε ένα σήμα, συχνά μοντελοποιείται με Gaussian (κανονική) κατανομή.
- *Άθροισμα τυχαίων μεταβλητών*. Σύμφωνα με το κεντρικό οριακό θεώρημα, το άθροισμα τυχαίων μεταβλητών θα τείνει να κατανέμεται (υπό συνθήκες) με Gaussian κατανομή, όσο αυξάνεται ο αριθμός των μεταβλητών που προστίθενται [13, vol. II]. Επομένως στην πράξη ένα πεπερασμένο άθροισμα αρκετά μεγάλου αριθμού τυχαίων μεταβλητών, μπορεί να μοντελοποιείται σαν Gaussian.

Αυτά τα παραδείγματα θα πρέπει να είναι αρκετά.

3.5.2 Επιλογή του μοντέλου, (β)

Τι γίνεται όμως όταν δεν έχουμε τόσο στέρεα αντίληψη για την μορφή του μοντέλου; Το μόνο σίγουρο είναι ότι δεν υπάρχει κάποιο μοντέλο που να είναι η καλύτερη επιλογή – πανάκεια, εκ των προτέρων για κάθε data set. Ένας τρόπος σκέψης θα ήταν να βρούμε μοντέλο τέτοιο, ώστε η μέγιστη πιθανοφάνεια να είναι μεγαλύτερη από τη μέγιστη πιθανοφάνεια κάθε άλλου μοντέλου (μιλώντας πάντα για το ίδιο data set).
Στην πραγματικότητα υπάρχει πάντα μοντέλο για το οποίο να μπορεί να αυξάνεται η πιθανοφάνεια απεριόριστα. Ένα τέτοιο μοντέλο είναι το παρακάτω:

$$p(X;\boldsymbol{\pi},\boldsymbol{\mu},\Sigma)=\sum_{i=1}^{N}\pi_i Normal(\mu_i,\Sigma) \quad (3.6)$$

για $X=[x_1,x_2,...,x_N]$ με $x_i\in\mathbb{R}^d$, το οποίο είναι μια μορφή μίξης κανονικών κατανομών [11]. Αυτές θα τις δούμε πιο αναλυτικά σε επόμενη παράγραφο, από όπου θα δανειστούμε και κάποια ορολογία.

Παρατηρούμε στον (3.6) ότι έχουμε τόσους αθροιστέους όσους και δεδομένα, και ότι οι μήτρες συμμεταβλητότητας Σ είναι όλες ίσες. Τώρα, αν θέσουμε

$$\mu_i=x_i\ ,\ \pi_i=N^{-1},\forall i \text{ και } \Sigma=\varepsilon I,\varepsilon\in\mathbb{R}^+$$

η πιθανοφάνεια θα είναι

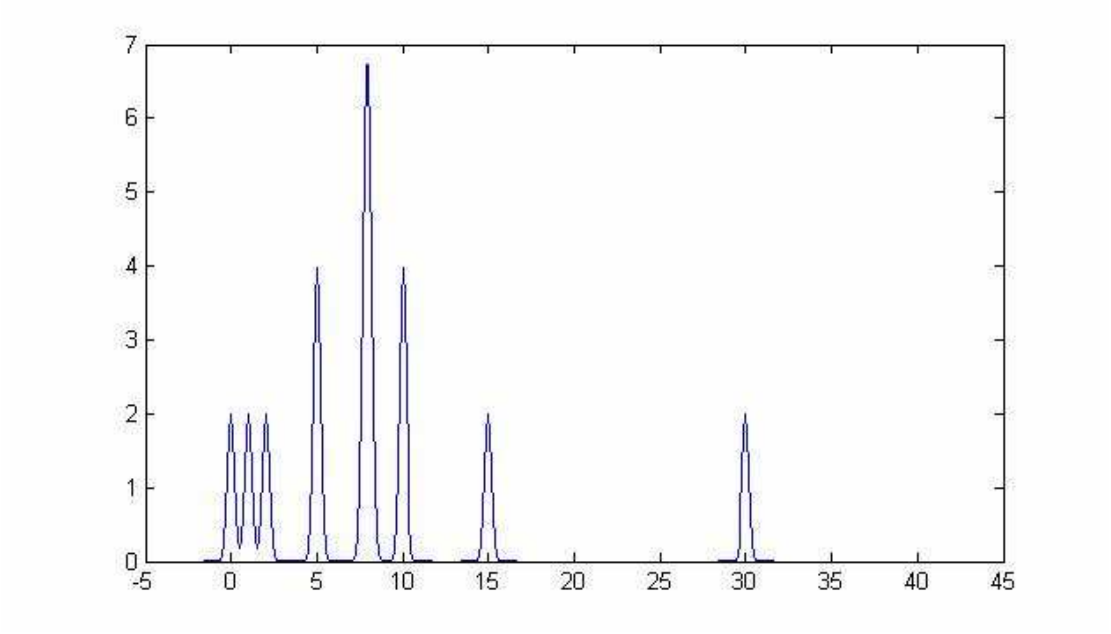
$$L = \log \sum_{i=1}^{N} \frac{1}{N} (2\pi)^{-\frac{d}{2}} |\Sigma|^{-\frac{1}{2}} = \log |\Sigma|^{-\frac{1}{2}} + const.$$

οπότε απορρίπτοντας τους σταθερούς όρους, τελικά έχουμε να μεγιστοποιήσουμε

$$L = -\frac{1}{2} \log \varepsilon$$

Εύκολα βλέπουμε ότι για $\varepsilon \to 0$, θα έχουμε $L \to +\infty$.

Για μια οπτικοποίηση αυτού του αποτελέσματος, θεωρήστε data set βαθμωτών πραγματικών (d = 1), X = [0, 1, 2, 5, 5, 7.8, 7.9, 8, 8.1, 10, 10, 15, 30], και $\varepsilon = 0.2$. Η πυκνότητα πιθανότητας (3.6) με εκτίμηση μέγιστης πιθανοφάνειας δίνεται στο σχήμα 3.1.



Σχήμα 3.1. ‘Εκφυλισμένο’ ταίριασμα σε δεδομένα.

3.5.3 Επιλογή του μοντέλου, (γ): Ικανότητα γενίκευσης

Πράγματι πετύχαμε με το προηγούμενο μοντέλο ένα εξαιρετικό *ταίριασμα* στα δεδομένα. Η πιθανοφάνεια είναι ένας δείκτης για την ποιότητα του ταιριάσματος, την οποία μπορέσαμε να αυξήσουμε οσοδήποτε.

Δυστυχώς όμως, το ταίριασμα του σχήματος 3.1 είναι ένα ταίριασμα σε ένα συγκεκριμένο σύνολο δεδομένων – είναι *υπερβολικά* καλό!(λεγόμενη περίπτωση "overfitting") Στην πραγματικότητα χρειαζόμαστε κάτι παραπάνω, το ταίριασμα να είναι καλό *και για περαιτέρω άγνωστα δεδομένα.* Η πυκνότητα που εκτιμήσαμε και φαίνεται στο σχήμα 3.1, ουσιαστικά μας λέει: «Κάθε datum που είναι ως τώρα γνωστό ότι παράχθηκε από την πηγή δεδομένων, έχει πιθανότητα να παράγεται. Κάθε άλλο datum είναι αδύνατον (ή τουλάχιστον εξαιρετικά απίθανο) να παραχθεί.». Βέβαια αυτή είναι η ακραία περίπτωση μηδενικής γενίκευσης, αλλά το παράδειγμα είναι πιστεύουμε παραστατικό.

Σε ένα πραγματικό σενάριο όπου τα δεδομένα αντιστοιχούν σε φυσικά μεγέθη, κατά κανόνα μια τέτοια εκτίμηση δεν είναι επιθυμητή. Χρειαζόμαστε αντίθετα ένα μοντέλο που να μας δίνει εκτίμηση πυκνότητας με *ικανότητα γενίκευσης* όπως λέγεται [3]. Οπότε ικανότητα γενίκευσης, λέγεται η ικανότητα ενός μοντέλου να δίνει καλό ταίριασμα για τα άγνωστα δεδομένα. Μπορούμε να πούμε ότι η γενίκευση θα ήταν ανάλογη με την μέγιστη πιθανοφάνεια σε ένα σενάριο όπου μπορούμε να έχουμε αριθμό δεδομένων οσοδήποτε μεγάλο, κάτι που βέβαια δεν ισχύει στην πράξη.

3.5.4 Επιλογή του μοντέλου (δ): Κριτήρια επιλογής μοντέλου

Μία παράμετρος που σχετίζεται με την ικανότητα γενίκευσης είναι ο αριθμός των ελεύθερων μεταβλητών σε ένα μοντέλο. Αριθμό ελεύθερων μεταβλητών εννοούμε τον αριθμό των ανεξάρτητων βαθμωτών παραμέτρων σε ένα μοντέλο' για παράδειγμα, το μοντέλο Poisson της (3.5) έχει μία ελεύθερη μεταβλητή, την λ. Το μοντέλο της (3.6) έχει N ελεύθερες μεταβλητές για τα π_i, N για τα μ_i, και

$\frac{d(d+1)}{2}$ για την μήτρα Σ (είναι συμμετρική), σύνολο $2N+\frac{d(d+1)}{2}$, κ.ο.κ. Κατά κανόνα, όσο αυξάνονται οι ελεύθερες μεταβλητές, τόσο πιο πολύ μειώνεται η ικανότητα γενίκευσης, αφού μοντέλα με πολλές παραμέτρους τείνουν να υπερταιριάζουν (overfit) στα δεδομένα.

Θέλουμε λοιπόν ένα κριτήριο με το οποίο να επιλέξουμε ένα βέλτιστο μοντέλο από ένα πλήθος μοντέλων, που να μας δίνει όσο το δυνατόν καλύτερο άνω όριο πιθανοφάνειας και όσο το δυνατόν καλύτερη ικανότητα γενίκευσης. Για να απλοποιήσουμε το πρόβλημα περιορίζουμε το πρόβλημα στην επιλογή από πεπερασμένο πλήθος μοντέλων (εκ των προτέρων γνωστών).

Η πρώτη προσέγγιση είναι να επιλέξουμε το μοντέλο όχι που δίνει την μεγαλύτερη μέγιστη πιθανοφάνεια, αλλά που μεγιστοποιεί

$$Likelihood - p$$

όπου p ο αριθμός των ελεύθερων μεταβλητών. Αυτό το κριτήριο ονομάζεται “An Information Criterion” (AIC) [11, 14]. Ο αριθμός ελεύθερων μεταβλητών δηλαδή γίνεται άμεσα όρος ποινής. Ωστόσο, το AIC κριτήριο δεν περιέχει όρο ποινής σχετικό με το πλήθος δεδομένων. Αυτό είναι αρνητικό αφού οι τιμές που παίρνει η πιθανοφάνεια εξαρτώνται από το πλήθος των δεδομένων (γενικά θα παίρνει μεγαλύτερες τιμές όσο αυξάνονται τα δεδομένα), και ο όρος ποινής p όχι.

Ένα κριτήριο που ικανοποιεί αυτή την απαίτηση είναι ο Minimum Descriptor Length Principle (MDL) [4, 9, 11, 14]. Με αυτό ζητάμε να μεγιστοποιήσουμε

$$Likelihood - \frac{p}{2}\log N$$

όπου N ο αριθμός των δεδομένων. Αυτό το κριτήριο θα χρησιμοποιήσουμε και παρακάτω στην παρούσα εργασία, όταν θα χρειαστεί να επιλέξουμε το βέλτιστο από μια οικογένεια μοντέλων.

Πιο φορμαλιστικά λοιπόν, έστω σύνολο γνωστών δεδομένων X, σύνολο παραμέτρων, οικογένεια μοντέλων $\Delta = [\mathrm{M}_1, \mathrm{M}_2, ..., \mathrm{M}_k]$, και θ_i η εκτίμηση μέγιστης πιθανοφάνειας για το μοντέλο M_i. p_i ο αριθμός ελεύθερων μεταβλητών του μοντέλου M_i, και L_i η συνάρτηση πιθανοφάνειας για το μοντέλο M_i. Σύμφωνα με το κριτήριο MDL, επιλέγουμε το μοντέλο M_ξ με ξ να δίνεται από

$$\xi = \arg\max_i \{L_i(\theta_i) - \frac{p_i}{2}\log N\} \quad (3.7)$$

Βεβαίως αυτά δεν είναι τα μοναδικά κριτήρια επιλογής που μπορούν να χρησιμοποιηθούν. Για μια ανασκόπηση, και εφαρμογή ειδικά σε οικογένειες μικτών μοντέλων (που θα δούμε παρακάτω), βλέπε [11].

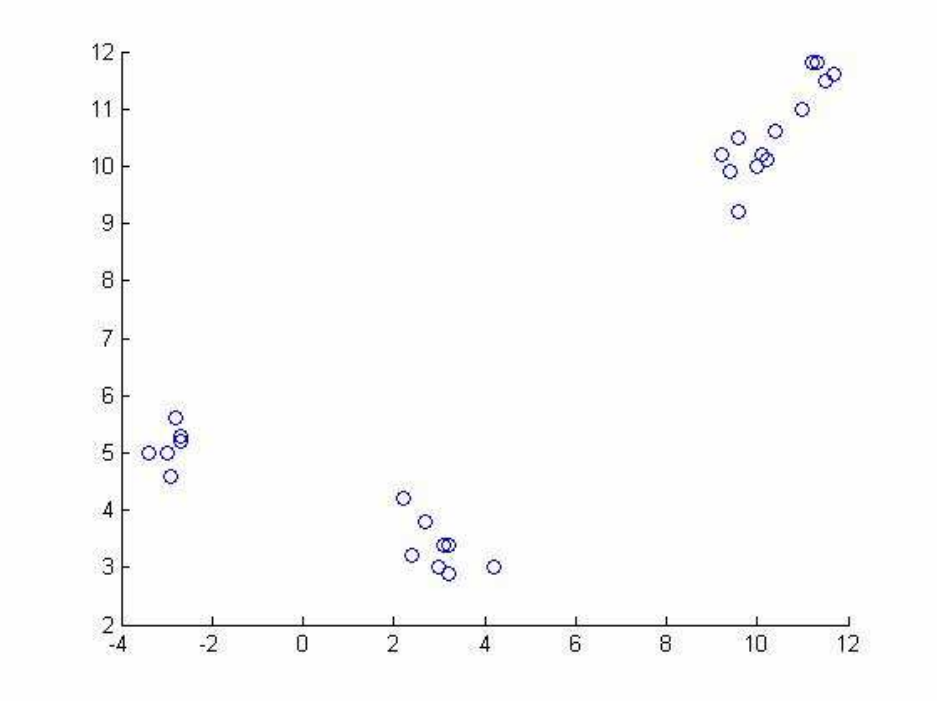
3.6. Μίξεις κανονικών κατανομών

3.6.1 Γενικά. Μίξεις κατανομών, μίξεις κανονικών κατανομών.

Μοντέλο μίξης κατανομών (Finite mixture model) λέγεται η πυκνότητα πιθανότητας της μορφής

$$p(x;\mathbf{\Psi}) = \sum_{i=1}^{K} \pi_i f(x;\theta_i) \quad (3.8)$$

όπου f πυκνότητα πιθανότητας με παράμετρο Ψ_i, Κ σταθερός θετικός ακέραιος, $\mathbf{\Psi}$ σύνολο των παραμέτρων, $\Psi = \left[\Psi_1, \Psi_2, ..., \Psi_K\right]$, με κάθε $\Psi_i = [\pi_i, \theta_i]$. Κάθε π_i παίρνει τιμές στο $[0,1]$, και ισχύει πάντα $\sum_{i=1}^{K} \pi_i = 1$. Κάθε $f(x;\Psi_i)$ την ονομάζουμε *πυρήνα* ή *συνιστώσα* του μοντέλου. Το αντίστοιχο π_i το ονομάζουμε *εκ των προτέρων πιθανότητα* ή *βάρος* του πυρήνα. Εύκολα φαίνεται ότι η μορφή (3.8) πληροί τις προϋποθέσεις για να είναι πυκνότητα πιθανότητας.



Σχήμα 3.2: Νέφη (clusters) δεδομένων στον $\mathbb{R}^2$. Κάθε κύκλος παριστά ένα datum.

Ένα τέτοιο μοντέλο χρησιμοποιείται περισσότερο σε περιπτώσεις όπου έχουμε να μοντελοποιήσουμε δεδομένα που σχηματίζουν μεταξύ τους ένα πλήθος ετερογενών ομάδων ή όπως αναφέρονται συνήθως στην βιβλιογραφία (π.χ. [3]), *clusters*. Στο σχήμα 3.2 μπορούμε να δούμε σαν παράδειγμα, δεδομένα που σχηματίζουν 3 clusters (ή ίσως 4 ή 5, θα μπορούσε να πει κάποιος τρίτος παρατηρητής. Αυτό είναι ενδεικτικό της ασάφειας σχετικά με την έννοια του cluster). Η πηγή μιας τέτοιας ετερογένειας στην πράξη, ποικίλει, και εξαρτάται βέβαια από την φυσική ερμηνεία των δεδομένων. Αν τα δεδομένα που έχουμε μετρούσανε κάποιο βιολογικό-ιατρικό χαρακτηριστικό, ο λόγος της ετερογένειας θα μπορούσε να είναι το φύλο, το είδος, η ηλικία, κ.ο.κ. Αν είχαμε χρώματα pixels μιας εικόνας, θα ήταν ο αριθμός των κυρίως εμφανιζόμενων χρωμάτων στην εικόνα (περισσότερα σχετικά με αυτή την περίπτωση θα δούμε αργότερα).

Η περίπτωση όπου οι πυρήνες f είναι κανονικές κατανομές είναι μια από τις πιο σημαντικές. Μίξη κανονικών κατανομών (Gaussian Mixture model, GMM) λέγεται η πυκνότητα πιθανότητας της μορφής

$$p(x;\boldsymbol{\pi},\boldsymbol{\mu},\boldsymbol{\Sigma}) \triangleq \sum_{i=1}^{K} \pi_i Normal(\mu_i,\Sigma_i) = \sum_{i=1}^{K} \pi_i (2\pi)^{-d/2} \left|\Sigma_i\right|^{-1/2} \exp\left\{-\frac{1}{2}(x-\mu_i)^T \Sigma_i^{-1}(x-\mu_i)\right\}$$
(3.9)

όπου d η διάσταση της τυχαίας μεταβλητής x, K σταθερός θετικός ακέραιος. Κάθε μ_i είναι πραγματικό διάνυσμα dx1 και κάθε Σ_i πραγματικός συμμετρικός πίνακας συμμεταβλητότητας, dxd.

Εν γένει σε προβλήματα εκτίμησης πυκνότητας βολεύει για διάφορους λόγους να χρησιμοποιούμε κανονικές κατανομές στην θέση των πυρήνων. Μερικοί από αυτούς είναι:

- Απλές αναλυτικές ιδιότητες, όπως ότι κάθε ροπή μπορεί να οριστεί συναρτήσει του μέσου της και της μήτρας συμμεταβλητότητας.
- Σύμφωνα με το κεντρικό οριακό θεώρημα, το άθροισμα πλήθους τυχαίων μεταβλητών τείνει να κατανέμεται κανονικά, όσο το πλήθος αυξάνεται (υπό συγκεκριμένες συνθήκες, βλέπε π.χ. [13, vol. II]).
- Μετά από γραμμικό μετασχηματισμό, η κατανομή παραμένει κανονική.
- Οι πυκνότητες περιθωρίου και υπό συνθήκη, είναι πάλι κανονικές (με διαφορετικές παραμέτρους).Δηλαδή

$$(x,y) \sim Normal \Rightarrow x \sim Normal, y \sim Normal$$

και

$$(x,y) \sim Normal \Rightarrow x \mid y \sim Normal, y \mid x \sim Normal$$

  όπου x,y τυχαίες μεταβλητές που παίρνουν τιμές στον $\mathbb{R}$.
- Υπάρχει γραμμικός μετασχηματισμός που διαγωνιοποιεί την μήτρα συμμεταβλητότητας (whitening transform). Έτσι οι μετασχηματισμένες τυχαίες μεταβλητές είναι στοχαστικά ανεξάρτητες, απλοποιώντας τον τύπο της κανονικής κατανομής,
- Η κανονική κατανομή είναι η κατανομή με την μέγιστη εντροπία, δοθέντων κάποιου μέσου και μήτρας συμμεταβλητότητας.

Μπορούμε να δούμε στο σχήμα 3.3 μια διμεταβλητή κανονική κατανομή, που περιγράφεται από $N([10 \quad 10]^T, I)$ (η μήτρα συμμεταβλητότητας είναι διαγώνια, και οι μεταβλητές είναι ανεξάρτητες), και μια μίξη μονομεταβλητών κανονικών κατανομών, που περιγράφεται από $\frac{1}{4}N(-3,1)+\frac{1}{2}N(0,1)+\frac{1}{4}N(4,4)$.

Γενικά κάθε πυρήνας αντιστοιχεί σε ένα μέγιστο ή ‘καμπούρα’ στο γράφημα της κατανομής, χωρίς αυτό να είναι απόλυτο: π.χ. $\frac{2}{3}N(0,2)+\frac{1}{3}N(0,1)$ ή $\frac{3}{4}N(0,1)+\frac{1}{4}N(1,1)$, που έχουν μόνο μία ‘καμπούρα’

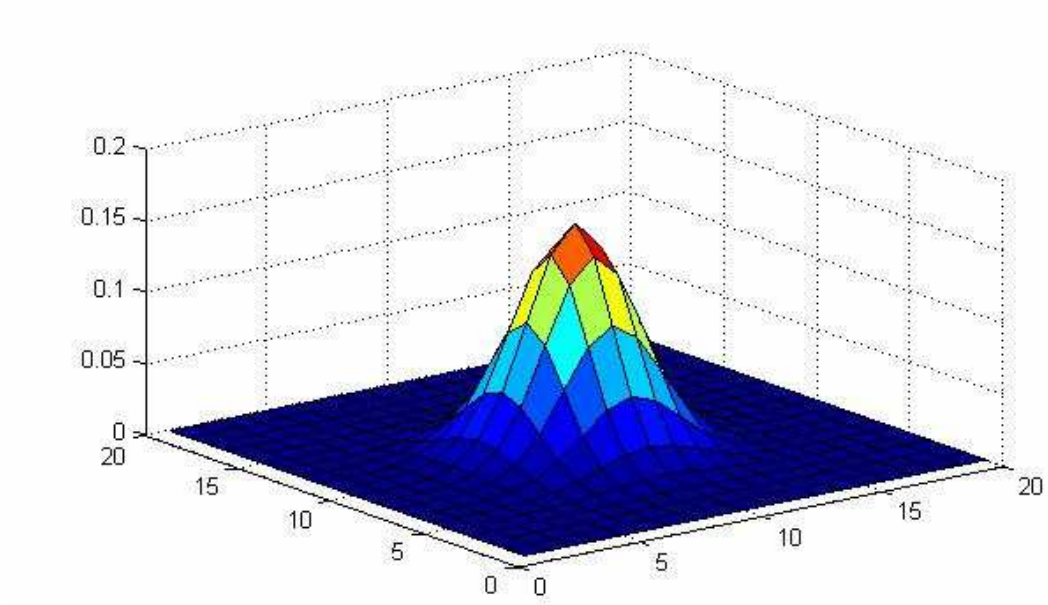
Το μοντέλο μίξης κανονικών κατανομών έχει χρησιμοποιηθεί ευρέως για εκτίμηση πυκνότητας. Μερικές από τις εφαρμογές του συμπεριλαμβάνουν δεδομένα από γενετική και ιατρική μέχρι αστρονομία και επεξεργασία εικόνας. Βλέπε [11] για μια εκτενή καταγραφή.

Παρακάτω θα ξαναδούμε κάποια πράγματα σχετικά με την εκτίμηση μέγιστης πιθανοφάνειας, και πως μπορούμε να την βρούμε για ένα τόσο περίπλοκο μοντέλο όσο το μοντέλο μίξης κατανομών.

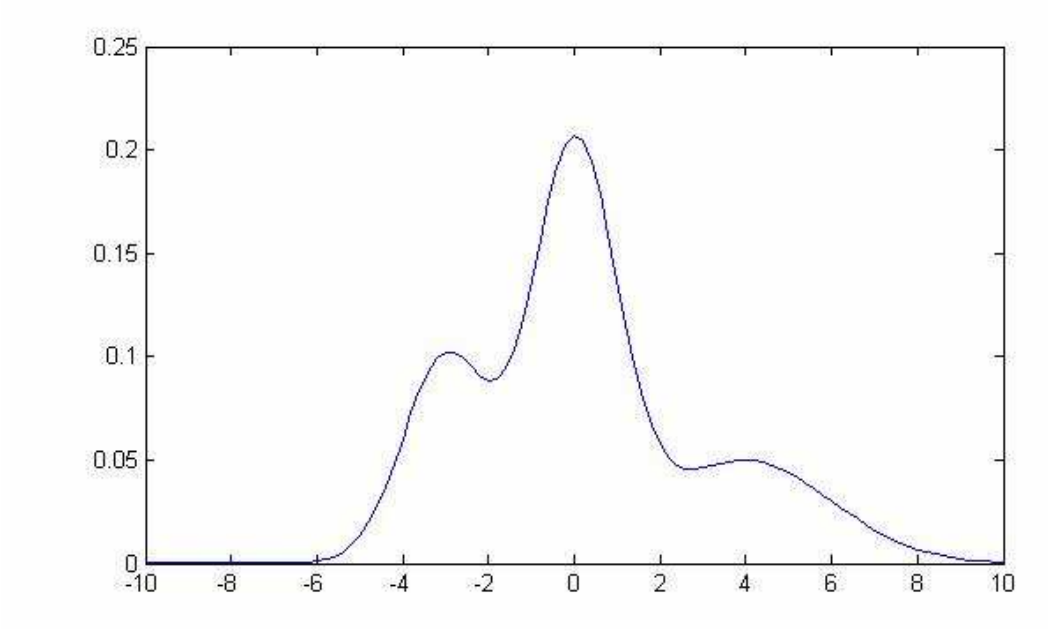
3.7. Μεγιστοποιώντας πιθανοφάνεια με ΕΜ

3.7.1 Ένα πρόβλημα βελτιστοποίησης.

Είδαμε ότι η εκτίμηση μέγιστης πιθανοφάνειας απαιτεί να μεγιστοποιήσουμε μια έκφραση $L(\Psi)$. Αυτό όμως μπορεί να είναι από τετριμμένο, έως εξαιρετικά δύσκολο. Η απλή περίπτωση είναι η πιθανοφάνεια L να εξαρτάται από μία βαθμωτή παράμετρο, να είναι παντού συνεχής και δις παραγωγίσιμη, και η πρώτη παράγωγος της να έχει ρίζες που μπορούν να βρεθούν με κλειστό τύπο. Αυτό συνέβη με το παράδειγμα μοντέλου Poisson, που είδαμε στην παράγραφο 3.4. Όταν τα πράγματα δεν είναι τόσο απλά, πρέπει να καταφύγουμε σε πιο προχωρημένες μεθόδους βελτιστοποίησης.



(α)



(β)

Σχήμα 3.3: (α) Διμεταβλητή κανονική κατανομή. (β) Μίξη μονομεταβλητών κανονικών κατανομών.

Αναφορικά μόνο να πούμε ότι μερικές από τις πιο γενικές και πιο επιτυχημένες είναι οι Nelder-Mead, Newton-Raphson (απαιτεί ύπαρξη δεύτερων παραγώγων), steepest

descent (απαιτεί ύπαρξη πρώτων παραγώγων), μέθοδοι Conjugate Gradients και Quasi-Newton όπως BFGS (απαιτεί ύπαρξη πρώτων παραγώγων) που είναι ίσως η πιο δημοφιλής [15]. Αυτές οι αριθμητικές μέθοδοι λειτουργούν διορθώνοντας επαναληπτικά μια αρχική εκτίμηση της λύσης, μέχρι αυτή να φτάσει αρκετά κοντά στην πραγματική λύση.

Υπάρχουν ωστόσο άλλες πιο ειδικές μέθοδοι που μπορούν να εφαρμοστούν όταν πληρούνται κάποιες πιο εξεζητημένες (εννοώντας σε σχέση με φερ' ειπείν την απαίτηση για ύπαρξη παραγώγων) προϋποθέσεις από την *αντικειμενική* συνάρτηση (όπως λέγεται σε ορολογία βελτιστοποίησης η συνάρτηση που θέλουμε να ελαχιστοποιήσουμε ή να μεγιστοποιήσουμε). Τέτοιος είναι ο αλγόριθμος MM [16]. Περισσότερο μας ενδιαφέρει η έκδοση του MM όταν θέλουμε να βελτιστοποιήσουμε συνάρτηση πιθανοφάνειας, που είναι ο αλγόριθμος EM [10, 11, 3, 14, 16]. Τον MM και τον EM ο οποίος μας ενδιαφέρει προφανώς περισσότερο, θα τους δούμε στη συνέχεια.

3.7.2 Ο αλγόριθμος MM.

3.7.2.1 Γενικά

Ο MM, όπως και ο EM, δεν είναι τόσο αλγόριθμοι στην πραγματικότητα, όσο γενικές μεθοδολογίες για βελτιστοποίηση συναρτήσεων. Πρώτα θα δούμε τον MM, που είναι πιο γενικός.

Τα αρχικά 'MM' έχουν διπλή σημασία. Σημαίνουν είτε 'Majorization-Minimization', είτε 'Minorization-Maximization' αναλόγως αν θέλουμε να βρούμε ελάχιστα ή μέγιστα. Κεντρικό ρόλο στην μέθοδο έχει η έννοια της Majorant (ή Majorizing) και της Minorant (ή Minorizing) συνάρτησης.

Μια συνάρτηση $g(x;x_k)$ λέγεται majorant μιας συνάρτησης $f(x)$ όταν ισχύουν, για οποιαδήποτε τιμή της παραμέτρου x_k στο πεδίο ορισμού της *f*::

- $f(x_k) = g(x_k;x_k)$ (3.10.1)
- $f(x) \le g(x;x_k)$ για $x \ne x_k$ (3.10.2)

Με άλλα λόγια δηλαδή η *g* εφάπτεται στην *f* στο σημείο $(x, f(x_k))$, και παντού αλλού παίρνει μεγαλύτερες -ή ίσες, δηλαδή μπορεί πάλι να εφάπτεται- τιμές από την *f*. Αυτό συνεπάγεται την σημαντική ιδιότητα

$$g(x_{k+1}; x_k) \le g(x_k; x_k) \Rightarrow f(x_{k+1}) \le f(x_k) \quad (3.11)$$

για αυθαίρετα x_k, x_{k+1}. Στην πράξη τελικά μια τιμή που ελαχιστοποιεί, ή τουλάχιστον μειώνει, την $g(x; x_k)$, θα μειώνει αναγκαστικά και την *f*. Βελτιώνοντας συνεχώς τις εκτιμήσεις x_k της λύσης με αυτό τον τρόπο, *γενικά* θα συγκλίνουν σε **τοπικό** ελάχιστο. Η σύγκλιση μπορεί να αποδειχθεί [16] υπό προϋποθέσεις για την *f*. Υπάρχουν πάντως περιπτώσεις όπου υπάρχει μεν σύγκλιση, αλλά όχι σε ελάχιστο ([10], στο πλαίσιο του ΕΜ).

Τα βήματα του αλγόριθμου ΜΜ για εύρεση ελάχιστου είναι λοιπόν:

1. Έστω $k = 0$. Έστω x_k αρχική εκτίμηση του ελάχιστου της *f*. Έστω *g* μια majorant της *f*. Η *g* εξαρτάται από παράμετρο όπως είδαμε.
2. **Majorization** : Υπολόγισε την g για παράμετρο x_k. Δηλαδή την $g(x; x_k)$.
3. **Minimization** : Βρες σημείο $\tilde{x}_k$, ολικό ελάχιστο της $g(x; x_k)$. Στην πράξη αρκεί σημείο που να μειώνει την τιμή της $g(x; x_k)$, σε σχέση με την τιμή της στο x_k.
4. Θέτω $x_{k+1} = \tilde{x}_k$. Αν η διαφορά με x_k είναι αρκετά μικρή, υπάρχει σύγκλιση και ο αλγόριθμος τερματίζει με λύση $x^* = x_{k+1}$.Αλλιώς αύξησε το *k* κατά 1 και επιστροφή στο βήμα 2.

Η περίπτωση minorization είναι εντελώς ανάλογη και τα βήμα 2,3 θα είναι αντίστοιχα Minorization – Maximization. Η φορά της ανισότητας απλά αντιστρέφεται στην (3.10.2), και πλέον μιλάμε για εύρεση μεγίστου όπου μιλάγαμε στην majorization περίπτωση για ελάχιστο.

Τα παραπάνω θα ξεκαθαρίσουν ελπίζουμε, αν δεν είναι ήδη σαφή, με ένα παράδειγμα.

3.7.2.2 Ένα παράδειγμα ΜΜ

Έστω αντικειμενική συνάρτηση

$$f(x) = 2e^{\sin x} + e^{\cos x} \quad (3.12).$$

Κατ' αρχάς παρατηρούμε ότι είναι περιοδική με $T = 2\pi$, οπότε θα την μελετήσουμε για $x \in [-\pi, +\pi)$.

Ψάχνουμε το μέγιστο αυτής, οπότε χρειαζόμαστε μια minorant συνάρτηση *g*. Θα μπορούσαμε να μηδενίσουμε την πρώτη παράγωγο της *f* ώστε να βρούμε στάσιμα σημεία, αλλά αυτό μας οδηγεί σε μορφή που δεν μπορούμε να λύσουμε, τουλάχιστον με κλειστό τύπο. Οπότε θέλουμε να βρούμε μια συνάρτηση *g* που να είναι *και* minorant αλλά *και* να μεγιστοποιείται με τρόπο πιο εύκολο απ' ότι η *f*. Αν η τελευταία απαίτηση δεν ισχύει, μπορούμε να εφαρμόσουμε ΜΜ, αλλά δεν θα έχει καμμία πρακτική αξία! Αυτό είναι γενικά και το κίνητρο για την χρήση του ΜΜ.

Στη συγκεκριμένη περίπτωση, υπάρχει τουλάχιστον μια τέτοια *g*. Εμείς θα διαλέξουμε

$$g(x; x_k) = \exp\{\frac{\ln\{m_1(x_k)\, 2\exp\{\sin x\}\}}{m_1(x_k)} + \frac{\ln\{m_2(x_k)\exp\{\cos x\}\}}{m_2(x_k)}\}$$

$$\text{με } m_1(x) = \frac{2\exp\{\sin x\} + \exp\{\cos x\}}{2\exp\{\sin x\}}$$

$$\text{και } m_2(x) = \frac{2\exp\{\sin x\} + \exp\{\cos x\}}{\exp\{\cos x\}}$$

(Για λεπτομέρειες σχετικά με μεθόδους για να βρίσκουμε minorant συναρτήσεις όπως η *g*, βλέπε [16, σελ. 121]). Σε πρώτη ματιά η *g* φαίνεται πιο πολύπλοκη από την *f*, αλλά αρκεί να μηδενίσουμε την πρώτη παράγωγο και να μερικές πράξεις για να βρούμε τελικά το μέγιστο για την $g(x; x_k)$. Αυτό[3] θα χρησιμοποιούμε και σαν επόμενη εκτίμηση σε κάθε επανάληψη του ΜΜ, οπότε καταλήγουμε

[3] Για την ακρίβεια η arctan θα δώσει στο [-π,π) δύο λύσεις, με απόσταση π μεταξύ τους. Εμείς επιλέγουμε από αυτές τις δύο, την λύση που δίνει μεγαλύτερη τιμή στην

$$x_{k+1} = \arctan\left(2\exp\{\sin x_k - \cos x_k\}\right)$$

Για αρχική εκτίμηση $x_k = 0$ τα αποτελέσματα φαίνονται στον πίνακα 3.1. Όπως βλέπουμε μέσα σε λίγες επαναλήψεις η σύγκλιση είναι ικανοποιητική.

Αριθμός επανάληψης k	x_k	$f(x_k)$
0	0	4.7183
1	0.6343	5.8552
2	1.0168	6.3736
3	1.2239	6.5271
4	1.3031	6.5491
5	1.3274	6.5512
6	1.3342	6.5513
7	1.3361	6.5514
8	1.3366	6.5514
9	1.3368	6.5514
10	1.3368	6.5514

Πίνακας 3.1. Αποτελέσματα ΜΜ για f(x)=2exp{sinx}+exp{cosx}.

Την οπτικοποίηση του αποτελέσματος μπορούμε να δούμε με τα γραφήματα στο σχήμα 3.4. Εκεί φαίνεται πιο ξεκάθαρα η λειτουργία του ΜΜ.

Πριν συνεχίσουμε με τον ΕΜ, να κάνουμε μερικές παρατηρήσεις. Πρώτον, είμαστε εντελώς ελεύθεροι στην επιλογή της majorant ή minorant g αρκεί βέβαια να ισχύουν οι προϋποθέσεις (3.10) που αναφέραμε. Αυτό σημαίνει ότι το ίδιο πρόβλημα μπορεί να λυθεί με ΜΜ, αλλά με πολλούς διαφορετικούς τρόπους αναλόγως την επιλογή της g. Δεύτερον, όπως και οι άλλες αριθμητικές μέθοδοι βελτιστοποίησης (βλέπε 3.7.1) η

αντικειμενική συνάρτηση. Στο συγκεκριμένο παράδειγμα το ένα σημείο αντιστοιχεί σε ολικό ελάχιστο και το άλλο σε ολικό μέγιστο.

λύση που θα βρούμε δεν υπάρχει εγγύηση ότι θα είναι ολική, παρά κατά κανόνα συγκλίνουμε σε τοπικό ακρότατο. Η αρχική εκτίμηση της λύσης είναι καθοριστική για τα αποτελέσματα που θα πάρουμε.

3.7.3 Ο αλγόριθμος ΕΜ

Ο αλγόριθμος ΕΜ (Expectation – Maximization), είναι στην πραγματικότητα ειδική περίπτωση του ΜΜ (για Minorization – Maximization)‘ έχει όλες τις ιδιότητες του ΜΜ που προαναφέραμε. Το ειδικό έγκειται στο ότι μπορεί να εφαρμοστεί μόνο όταν δουλεύουμε με συναρτήσεις που εξαρτώνται από τυχαίες μεταβλητές, ουσιαστικά πιθανοφάνειες, και ότι ορίζει συγκεκριμένη συνάρτηση minorant. Ας δούμε όμως τον αλγόριθμο πιο συγκεκριμένα.

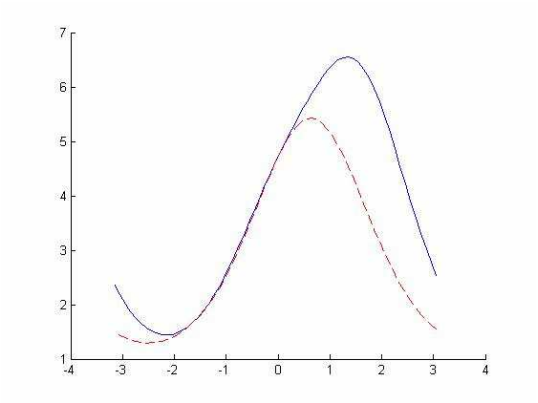
Έστω σύνολο δεδομένων X , και συνάρτηση πιθανοφάνειας

$$L(\Psi) \text{ ή } \ln p(X;\Psi) \quad (3.13),$$

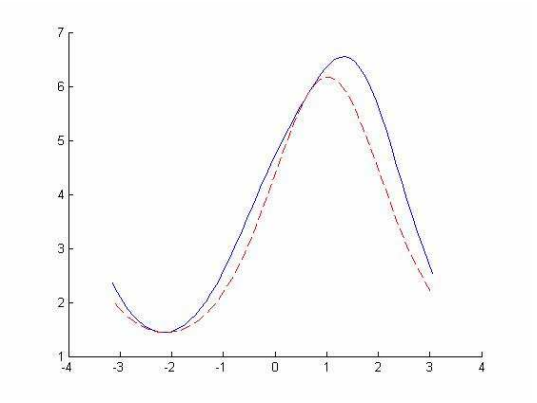
δηλαδή εξαρτάται από κάποιο σύνολο παραμέτρων Ψ. Ζητάμε να βρούμε τη λύση μέγιστης πιθανοφάνειας’ ως εδώ δεν υπάρχει κάτι καινούριο. Εισάγουμε τώρα σύνολο δεδομένων *Y*, ονόματι *πλήρες*, και σε αντιδιαστολή με αυτό ονομάζουμε τώρα το σύνολο δεδομένων X *ελλιπές*, και την πιθανοφάνεια $L(\Psi)$ επίσης *ελλιπή*. Αντίστοιχα ορίζεται συνάρτηση πιθανοφάνειας για τα *Y*, που αναφέρεται στην λύση του ίδιου προβλήματος, και λέγεται *πλήρης πιθανοφάνεια*, η

$$L_c(\Psi) \text{ ή } \ln p(Y;\Psi) \quad (3.14).$$

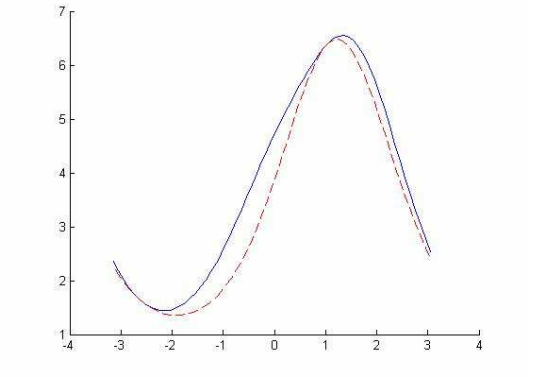
Η ιδέα είναι ότι μπορούμε να ορίσουμε συνάρτηση από κάθε Y σε μοναδικό X, όχι αντίστροφα όμως. Επιπλέον τα Y εξαρτούνται στοχαστικά από τα X, και μάλιστα γνωρίζουμε και την *πυκνότητα εξάρτησης*



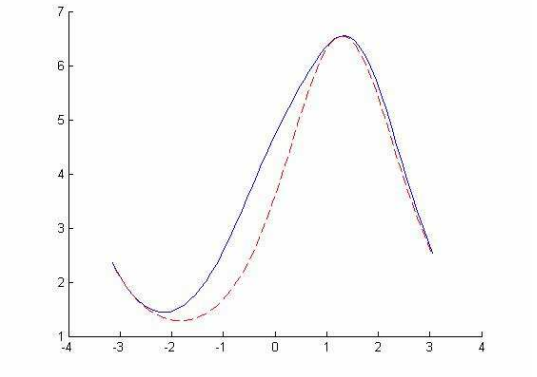
(α) Αρχικοποίηση, $\kappa = 0$.



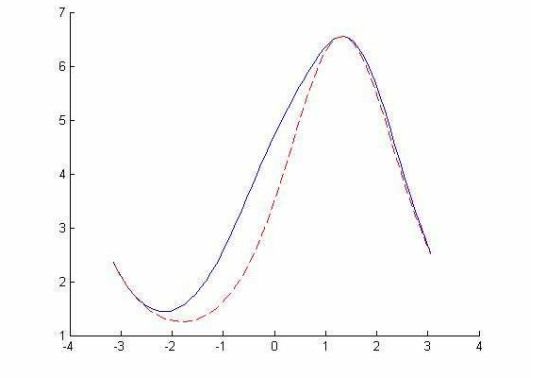
(β) Επανάληψη $\kappa = 1$



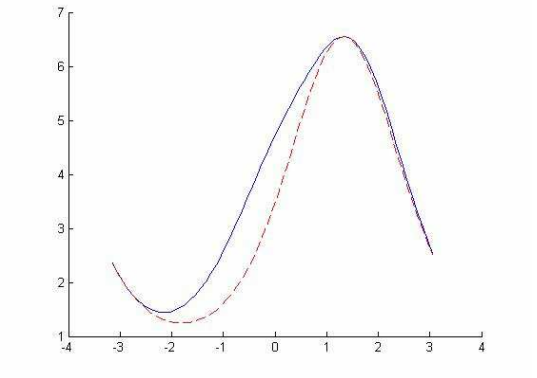
(γ) Επανάληψη $\kappa = 2$



(δ) Επανάληψη $\kappa = 3$



(ε) Επανάληψη $\kappa = 4$



(στ) Επανάληψη $\kappa = 5$

Σχήμα 3.4. Μεγιστοποίηση με ΜΜ. Η αντικειμενική συνάρτηση f (3.12) φαίνεται με μπλε χρώμα. Με διακεκομμένο κόκκινο φαίνεται η minorant g για διαφορετικές τιμές της παραμέτρου x_k, που συμπίπτει με σημείο επαφής των f, g. Αυτές είναι (α) $x_0 = 0$ (β) $x_1 = 0.6343$ (γ) $x_2 = 1.0168$ (δ) $x_3 = 1.2239$ (ε) $x_4 = 1.3031$ (στ) $x_5 = 1.3274$. Μέγιστο στο $x^* = 1.3368$, $f(x^*) = 6.5514$.

$$p(Y \mid X;\Psi) \quad (3.15).$$

Αυτό εξηγεί και την φιλοσοφία της ονοματοδοσίας στα *X, Y* (ελλιπές, πλήρες) που υπονοεί ότι τα *X* είναι δεδομένα παρατηρούμενα, ενώ τα *Y* είναι τα «πραγματικά» δεδομένα που δεν μπορούμε να γνωρίζουμε.

Θα μπορούσαμε να δοκιμάσουμε να βρούμε μέγιστη πιθανοφάνεια για $L_c(\Psi)$ και να λύσουμε το πρόβλημα. Δυστυχώς, αυτή εξαρτάται από τα πλήρη δεδομένα που είναι άγνωστα. Μπορούμε όμως να ξεπεράσουμε αυτό το εμπόδιο' σε αυτό το σημείο ένας ακόμα ορισμός είναι απαραίτητος. Έστω

$$Q(\Psi;\Psi_k) \triangleq \left\langle L_c(\Psi) \mid X;\Psi_k \right\rangle_Y \quad (3.16)$$

δηλαδή η αναμενόμενη τιμή της πλήρους πιθανοφάνειας, με Ψ_k αυθαίρετο σύνολο παραμέτρων. Αλλιώς γράφουμε

$$Q(\Psi;\Psi_k) = \int_{Y(X)} L_c(\Psi)\, p(Y \mid X;\Psi_k) dY$$

$$= \int_{Y(X)} \ln\{p(Y;\Psi)\}\, p(Y \mid X;\Psi_k) dY$$

$$= \int_{Y(X)} \ln\{p(Y,X;\Psi)\}\, p(Y \mid X;\Psi_k) dY \text{ αφού συγκεκριμένο Y συνεπάγεται μοναδικό X,}$$

$$= \int_{Y(X)} \ln\{p(Y \mid X;\Psi)\}\, p(Y \mid X;\Psi_k) dY + \int_{Y(X)} \ln\{p(X;\Psi)\}\, p(Y \mid X;\Psi_k) dY$$

Στον δεύτερο όρο, το $\ln p(X;\Psi)$ βγαίνει έξω από το ολοκλήρωμα αφού δεν εξαρτάται από τον όρο ολοκλήρωσης *Y*. Οπότε

$$\int_{Y(X)} \ln\{p(X;\Psi)\}\, p(Y \mid X;\Psi_k) dY = \ln\{p(X;\Psi)\} \int_{Y(X)} p(Y \mid X;\Psi_k) dY = \ln p(X;\Psi) = L(\Psi)$$

και ορίζοντας συνάρτηση

$$H(\Psi;\Psi_k) \triangleq \int_{Y(X)} \ln\{p(Y \mid X;\Psi)\}\, p(Y \mid X;\Psi_k) dY \quad (3.17)$$

έχουμε τελικά την πιο κομψή σχέση

$$Q(\Psi;\Psi_k) = L(\Psi) + \mathrm{H}(\Psi;\Psi_k) \quad (3.18)$$

Ωστόσο εύκολα αποδεικνύεται ότι η H παίρνει μέγιστο για $\Psi = \Psi_k$. Πράγματι, χρησιμοποιώντας την ανισότητα του Jensen:

$$H(\Psi;\Psi_k) - H(\Psi_k;\Psi_k) = \int_{Y(X)} \ln\{p(Y \mid X;\Psi)/\, p(Y \mid X;\Psi_k)\}\, p(Y \mid X;\Psi_k) dY \leq$$

$$\leq \ln \int_{Y(X)} p(Y \mid X;\Psi)\, p(Y \mid X;\Psi_k)/\, p(Y \mid X;\Psi_k) dY = \ln 1 = 0$$

Αυτό συνεπάγεται

$$L(\Psi) \geq Q(\Psi;\Psi_k) - H(\Psi_k;\Psi_k) \quad (3.19)$$

Εύκολα επιβεβαιώνουμε ότι ισχύει ισότητα στην (3.19) όταν $\Psi = \Psi_k$. Επομένως όπως είδαμε στην προηγούμενη ενότητα για ΜΜ, η συνάρτηση

$$Q(\Psi;\Psi_k) - H(\Psi_k;\Psi_k) \quad (3.20)$$

είναι minorant της L(Ψ)! Οπότε την χρησιμοποιούμε για να φτιάξουμε ένα επαναληπτικό αλγόριθμο που να μεγιστοποιεί την $L(\Psi)$.Αυτό ακριβώς είναι ο ΕΜ.
Άρα με δυο λόγια, ο ΕΜ είναι εφαρμογή του ΜΜ για μεγιστοποίηση της (ελλιπούς) πιθανοφάνειας $L(\Psi)$, χρησιμοποιώντας σαν minorant συνάρτηση την (3.20). Παρακάτω ακολουθούν συνοπτικά τα βήματα του ΕΜ:

1. Έστω $k = 0$. Έστω Ψ_k ή Ψ_0 αρχική εκτίμηση του μέγιστου της πιθανοφάνειας $L(\Psi)$.
2. **Expectation**: Υπολόγισε την $Q(\Psi;\Psi_k)$. Αυτό είναι υπολογισμός της αναμενόμενης τιμής (3.16), εξ' ου και το όνομα expectation.
3. **Maximization**: Υπολόγισε το ολικό μέγιστο της $Q(\Psi;\Psi_k)$, έστω αυτό $\tilde{\Psi}_k$. Αυτό είναι ισοδύναμο με υπολογισμό του ολικού μεγίστου της minorant $Q(\Psi;\Psi_k) - H(\Psi_k;\Psi_k)$, αφού οι δύο εκφράσεις διαφέρουν κατά σταθερό όρο. Επομένως δεν χρειάζεται να υπολογίζουμε το $H(\Psi_k;\Psi_k)$ στην πράξη.
4. Θέσε $\Psi_{k+1} = \tilde{\Psi}_k$. Αν η σύγκλιση της Ψ_k είναι ικανοποιητική, τερμάτισε τον αλγόριθμο και δώσε Ψ_{k+1} σαν τελική λύση (μέγιστο της $L(\Psi)$). Αλλιώς αύξησε το *k* κατά 1 και επέστρεψε στο βήμα 2.

Θα κλείσουμε την ενότητα με μερικές παρατηρήσεις.

Κατ' αρχάς προκειμένου να εφαρμοστεί ο αλγόριθμος ΕΜ, απαραίτητη προϋπόθεση είναι γνώση των μορφών (3.13-3.15) δηλαδή της ελλιπούς, της πλήρους πιθανοφάνειας και της στοχαστικής εξάρτησης των πλήρων από τα ελλιπή δεδομένα. Παρατηρούμε επίσης ότι οι (3.13-3.15) όλες εξαρτούνται από το ίδιες παραμέτρους Ψ.

Το ποια δεδομένα θα ονομάσουμε 'πλήρη' δεν είναι κάτι μοναδικό για το ίδιο πρόβλημα. Η μορφή των πλήρων δεδομένων μπορεί να είναι οποιαδήποτε, αρκεί βεβαίως να γνωρίζουμε και αντίστοιχη πυκνότητα εξάρτησης $p(Y \mid X; \Psi)$.

Στην πράξη ο ΕΜ δεν δίνει ολικό, αλλά τοπικό μέγιστο της πιθανοφάνειας. Αυτό είναι κάτι που δεν μπορούμε να αποφύγουμε, όπως και με τον ΜΜ, και τις αριθμητικές μεθόδους βελτιστοποίησης. Παρομοίως συμβαίνει και με το πρόβλημα της εξάρτησης της λύσης στην οποία συγκλίνουμε, από την αρχική εκτίμηση Ψ_0. Ακόμα χειρότερα, είναι πιθανό η σύγκλιση να γίνει σε σαγματικό ή τοπικά ελάχιστο σημείο [10].

Τέλος ένα ενδιαφέρον σημείο είναι ότι ο όρος $-H(\Psi_k; \Psi_k)$ στον minorant (3.20) είναι η *εντροπία* της πυκνότητας εξάρτησης για παράμετρο Ψ_k, $p(Y \mid X; \Psi_k)$.

3.8. Εφαρμογή του ΕΜ για μοντέλο μίξης κανονικών κατανομών

3.8.1 Εφαρμογή σε γενικό μικτό μοντέλο

Έστω μοντέλο, θεωρώντας iid δεδομένα,

$$p(X; \Psi) = p(X; \boldsymbol{\pi}, \boldsymbol{\theta}) = \prod_{i=1}^{N} p(x_i; \boldsymbol{\pi}, \boldsymbol{\theta}) \quad (3.21)$$

όπου, όπως είδαμε και στην 3.6,

$$p(x_i;\boldsymbol{\pi},\boldsymbol{\theta}) = \sum_{j=1}^{K} \pi_j f(x_i;\theta_j) \quad (3.22)$$

με σύνολο δεδομένων $X = \{x_1, x_2, ..., x_N\}$, παραμέτρους $\Psi = \{\boldsymbol{\pi}, \boldsymbol{\theta}\}$, $\boldsymbol{\pi} = \{\pi_1, \pi_2, ..., \pi_K\}$, $\boldsymbol{\theta} = \{\theta_1, \theta_2, ..., \theta_K\}$. Το $\boldsymbol{\pi}$ είναι σύνολο Κ θετικών βαθμωτών, με $\sum_{j=1}^{K} \pi_i = 1$ (τα βάρη του μοντέλου). Η *f* είναι κάποια πυκνότητα πιθανότητας με παράμετρο θ_j ' δηλαδή έχουμε μίξη Κ πυκνοτήτων *f* με παράμετρο θ_j η καθεμία.

Οπότε ορίζεται η (ελλιπής) πιθανοφάνεια του μοντέλου (3.21), ως

$$L(\Psi) = \sum_{i=1}^{N} \ln \sum_{j=1}^{K} \pi_j f(x_i;\theta_j) \quad (3.23)$$

Έστω τώρα πλήρη δεδομένα $Y = \{X, Z\}$, με $Z = \{z_1, z_2, ..., z_N\}$. Το κάθε z_i ορίζουμε σαν *ετικέτα* του δεδομένου *i,* δηλαδή ο αριθμός 1..*K* από την οποία παρήχθη το δεδομένο. Δεν χρειάζεται να μπούμε στη διαδικασία να σκεφτούμε αν έχει πάντα κάποια φυσική έννοια αυτός ο ορισμός, ή αν «πράγματι» τα δεδομένα μας παράγονται επιλέγοντας μια κατανομή και μετά με δειγματοληψία από την αντίστοιχη συνιστώσα, ή αν κάθε δεδομένο έχει από μια τέτοια ετικέτα. Αρκεί να το αντιμετωπίσουμε σαν ένα τέχνασμα για να μπορέσουμε να κάνουμε ΕΜ.

Έχοντας γνώση πλήρων δεδομένων *Y,* μπορούμε να ορίσουμε την πλήρη πιθανοφάνεια:

$$L_C(\Psi) = \sum_{i=1}^{N} \ln \pi_{z_i} f(x_i;\theta_{z_i}) \quad (3.24)$$

Και πυκνότητα εξάρτησης ελλιπών από πλήρη δεδομένα:

$$p(Y \mid X;\Psi) = \prod_{i=1}^{N} p(y_i \mid X;\Psi) = \prod_{i=1}^{N} p(z_i, x_i \mid x_i;\Psi) = \prod_{i=1}^{N} p(z_i \mid x_i;\Psi)$$

όπου

$$p(z_i \mid x_i;\Psi) = \frac{p(x_i \mid z_i;\Psi)\, p(z_i;\Psi)}{p(x_i;\Psi)} = \frac{\pi_{z_i} f(x_i;\theta_{z_i})}{\sum_{j=1}^{K} \pi_j f(x_i;\theta_j)} \quad (3.25)$$

Τώρα μπορούμε να υπολογίσουμε την $Q(\Psi;\Psi_k)$ που χρειάζεται στο *Expectation* βήμα. Είναι

$$Q(\Psi;\Psi_k) = \sum_{Y} L_C(\Psi)\, p(Y \mid X;\Psi_k) = \sum_{Y} \sum_{i=1}^{N} \ln \pi_{z_i} f(x_i;\theta_{z_i})\, p(Y \mid X;\Psi_k) =$$

$$= \sum_{z_1=1}^{M}\sum_{z_2=1}^{M}\ldots\sum_{z_N=1}^{M}\sum_{i=1}^{N}\ln \pi_{z_i} f(x_i;\theta_{z_i})\prod_{j=1}^{N} p(z_j \mid x_j;\Psi_k) =$$

$$= \sum_{z_1=1}^{M}\sum_{z_2=1}^{M}\ldots\sum_{z_N=1}^{M}\sum_{i=1}^{N}\sum_{l=1}^{M}\delta(z_i,l)\ln \pi_l f(x_i;\theta_l)\prod_{j=1}^{N} p(z_j \mid x_j;\Psi_k) =$$

$$= \sum_{i=1}^{N}\sum_{l=1}^{M}\ln \pi_l f(x_i;\theta_l)\sum_{z_1=1}^{M}\sum_{z_2=1}^{M}\ldots\sum_{z_N=1}^{M}\delta(z_i,l)\prod_{j=1}^{N} p(z_j \mid x_j;\Psi_k)$$

Όπου δ είναι η συνάρτηση Kronecker. Παρατηρούμε ότι

$$\sum_{z_1=1}^{M}\sum_{z_2=1}^{M}\ldots\sum_{z_N=1}^{M}\delta(z_i,l)\prod_{j=1}^{N} p(z_j \mid x_j;\Psi_k) =$$

$$\left(\sum_{z_1=1}^{M}\sum_{z_2=1}^{M}\ldots\sum_{z_{i-1}=1}^{M}\sum_{z_{i+1}=1}^{M}\ldots\sum_{z_N=1}^{M}\prod_{j=1,\, j\neq i}^{N} p(z_j \mid x_j;\Psi_k)\right) p(l \mid x_i;\Psi_k) =$$

$$\left(\prod_{j=1,\, j\neq i}^{N}\sum_{z_j=1}^{M} p(z_j \mid x_j;\Psi_k)\right) p(l \mid x_i;\Psi_k) = p(l \mid x_i;\Psi_k)$$

Οπότε τελικά

$$Q(\Psi;\Psi_k) = \sum_{i=1}^{N}\sum_{j=1}^{M} p(j \mid x_i;\Psi_k)\ln \pi_j f(x_i;\theta_j) \quad (3.26)$$

Προσέχουμε στον (3.26) να μην γίνει παρανόηση: Οι μεταβλητές (το Ψ) είναι τα π_j, θ_j. Τα υπόλοιπα είναι σταθερές ή παράμετροι.

Σε αυτή τη μορφή μπορούμε να μεγιστοποιήσουμε το $Q(\Psi;\Psi_k)$ ως προς τα βάρη π_j. Μηδενίζοντας την πρώτη παράγωγο και εισάγοντας τον περιορισμό τα βάρη να αθροίζονται σε μονάδα, παίρνουμε εύκολα τον κλειστό τύπο

$$\pi_j = \frac{1}{N}\sum_{i=1}^{N} p(j \mid x_i;\Psi_k) \ , \forall j \in [1, M] \quad (3.27)$$

Ο τρόπος που θα γίνει η βελτιστοποίηση ως προς τα θ_j, εξαρτάται από την μορφή των πυρήνων $f(x;\theta_j)$. Παρακάτω θα δούμε την περίπτωση κανονικών πυρήνων.

3.8.2 Εφαρμογή σε μίξη κανονικών κατανομών

Όσον αφορά το *Expectation* βήμα, ισχύουν τα ίδια με την γενική περίπτωση μικτού μοντέλου που είδαμε στην προηγούμενη ενότητα. Παρομοίως για τον υπολογισμό

των νέων βαρών στο *Maximization* βήμα. Μένει ο υπολογισμός των νέων *θ*, δηλαδή τα νέα μέσα και πίνακες συμμεταβλητότητας των πυρήνων.

Θέτω λοιπόν

$$f(x;\mu_j,\Sigma_j)=(2\pi)^{-d/2}\left|\Sigma_j\right|^{-1/2}\exp\{-\frac{1}{2}(x-\mu_j)^T\Sigma_j^{-1}(x-\mu_j)\}$$

και αντικαθιστώντας στον (3.26),

$$Q(\boldsymbol{\mu},\boldsymbol{\Sigma};\Psi_k)=\sum_{j=1}^{M}\sum_{i=1}^{N}p(j\mid x_i;\Psi_k)\ln p(x_i;\mu_j,\Sigma_j)+const.=$$

$$=-\frac{1}{2}\sum_{j=1}^{M}\sum_{i=1}^{N}\left(\ln\left|\Sigma_j\right|+(x-\mu_j)^T\Sigma_j^{-1}(x-\mu_j)\right)p(j\mid x_i;\Psi_k)+const. \quad (3.28)$$

όπου αντιμετωπίσαμε τα βάρη π_j σαν σταθερές, αφού έχουμε ήδη υπολογίσει τις βέλτιστες τιμές για αυτά στον (3.27).

Τώρα, μηδενίζοντας την πρώτη παράγωγο του (3.28) ως προς μ_j, έχω

$$\frac{\partial}{\partial\mu_j}\sum_{i=1}^{N}(x_i-\mu_j)^T\Sigma^{-1}(x_i-\mu_j)\,p(j\mid x_i;\Psi_k)=0$$

και χρησιμοποιώντας κάποια αποτελέσματα από γραμμική άλγεβρα (βλέπε Παράρτημα Α):

$$\sum_{i=1}^{N}\Sigma_j^{-1}(x_i-\mu_j)\,p(j\mid x_i;\Psi_k)=0\Rightarrow$$

$$\mu_j=\frac{\sum_{i=1}^{N}p(j\mid x_i;\Psi_k)x_i}{\sum_{i=1}^{N}p(j\mid x_i;\Psi_k)} \quad (3.29)$$

Μένει ο υπολογισμός των Σ_j. Μηδενίζοντας πρώτη παράγωγο τώρα ως προς Σ_j^{-1}, (είναι ισοδύναμο με μηδενισμό παραγώγου ως προς Σ_j, και βολεύει στους υπολογισμούς):

$$\frac{\partial}{\partial\Sigma_j^{-1}}\sum_{i=1}^{N}p(j\mid x_i;\Psi_k)\left(tr[\Sigma_j^{-1}(x_i-\mu_j)(x_i-\mu_j)^T]-\log\left|\Sigma_j^{-1}\right|\right)=0\Rightarrow$$

$$\sum_{i=1}^{N}p(j\mid x_i;\Psi_k)\left(2\{(x_i-\mu_j)(x_i-\mu_j)^T-\Sigma_j\}-diag\{(x_i-\mu_j)(x_i-\mu_j)^T-\Sigma_j\}\right)=0\Rightarrow$$

$$2\sum_{i=1}^{N}p(j\mid x_i;\Psi_k)\left((x_i-\mu_j)(x_i-\mu_j)^T-\Sigma_j\right)=diag\left(\sum_{i=1}^{N}p(j\mid x_i;\Psi_k)\left((x_i-\mu_j)(x_i-\mu_j)^T-\Sigma_j\right)\right)$$

(3.30)

(Βλέπε πάλι παράρτημα Α για παραγώγους συναρτήσεων πίνακα). Αυτό που έχουμε εδώ είναι ουσιαστικά $2A = diag(A)$ όπου Α το άθροισμα που εμφανίζεται και στα δύο μέλη της (3.30), και είναι πίνακας. Αλλά αυτό μπορεί να ισχύει μόνο όταν $A = 0$. Επομένως

$$\sum_{i=1}^{N} p(j \mid x_i; \Psi_k)\left((x_i - \mu_j)(x_i - \mu_j)^T - \Sigma_j\right) = 0 \Rightarrow$$

$$\Sigma_j = \frac{\sum_{i=1}^{N} p(j \mid x_i; \Psi_k)(x_i - \mu_j)(x_i - \mu_j)^T}{\sum_{i=1}^{N} p(j \mid x_i; \Psi_k)} \quad (3.31)$$

Οπότε τακτοποιώντας τα αποτελέσματα μας οι εξισώσεις για ΕΜ είναι, όπου $\pi_{j(k)}, \mu_{j(k)}, \Sigma_{j(k)}$ εκτιμήσεις στην k επανάληψη, και $\forall j \in [1, M]$:

$$\pi_{j(k+1)} = \frac{1}{N}\sum_{i=1}^{N} p(j \mid x_i; \pi_{(k)}, \mu_{(k)}, \Sigma_{(k)})$$

$$\mu_{j(k+1)} = \frac{\sum_{i=1}^{N} p(j \mid x_i; \pi_{(k)}, \mu_{(k)}, \Sigma_{(k)})x_i}{\sum_{i=1}^{N} p(j \mid x_i; \pi_{(k)}, \mu_{(k)}, \Sigma_{(k)})}$$

$$\Sigma_{j(k+1)} = \frac{\sum_{i=1}^{N} p(j \mid x_i; \pi_{(k)}, \mu_{(k)}, \Sigma_{(k)})(x_i - \mu_{j(k+1)})(x_i - \mu_{j(k+1)})^T}{\sum_{i=1}^{N} p(j \mid x_i; \pi_{(k)}, \mu_{(k)}, \Sigma_{(k)})}$$

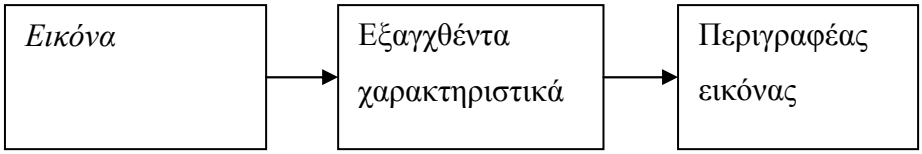
όπου $p(j \mid x_i; \pi_{(k)}, \mu_{(k)}, \Sigma_{(k)})$ υπολογίζεται με κανόνα του Bayes, (3.25).

ΚΕΦΑΛΑΙΟ 4. ΜΕΘΟΔΟΙ ΠΕΡΙΓΡΑΦΗΣ ΕΙΚΟΝΑΣ

4.1. Εισαγωγικά.

Όπως είπαμε και στο εισαγωγικό κεφάλαιο, αντιμετωπίζουμε την εικόνα αρχικά σαν μια δισδιάστατη διάταξη χρωματικών τιμών. Δηλαδή μπορούμε να την δούμε σαν ένα πίνακα P_{mn}. Σε κάθε στοιχείο αντιστοιχεί η χρωματική τιμή, ή πιο γενικά ένα διάνυσμα με τιμές χρώματος σε *RGB* ή οποιουδήποτε άλλου χρωματικού χώρου. Για απλότητα και χωρίς να χάνουμε σε γενικότητα θα κρατήσουμε στο μυαλό μας ότι κάθε στοιχείο είναι *RGB*. Είδαμε στη συνέχεια ότι από εκεί μπορούμε να εξάγουμε διάφορα χαρακτηριστικά πέραν του χρώματος.

Οπότε φθάνουμε στο επόμενο στάδιο επεξεργασίας της εικόνας, που είναι να βρούμε ένα τρόπο να περιγράψουμε το περιεχόμενο της συνολικά, με ένα τέτοιο τρόπο ώστε να μπορεί να συγκριθεί άμεσα με το περιεχόμενο άλλων εικόνων. Αυτή είναι η ποιοτική διαφορά με το προηγούμενο στάδιο, και ο λόγος που εξετάζουμε τα στάδια εξαγωγής χαρακτηριστικών και περιγραφής της εικόνας ξεχωριστά. Η διαφορά θα γίνει πιο εμφανής προχωρώντας.



Σχήμα 4.1.

4.2. Ιστογράμματα

4.2.1 Εισαγωγικά.

Οι Swain και Ballard πρότειναν στο [1] την χρήση ιστογραμμάτων χρώματος σαν περιγραφείς εικόνας. Αυτή η μέθοδος δουλεύει ως εξής:
Έστω εικόνα που περιγράφεται από πίνακα P. Σε κάθε στοιχείο του πίνακα αντιστοιχεί ένα pixel και το καθένα περιγράφεται σε έναν τρισδιάστατο χρωματικό χώρο από ένα διάνυσμα $\begin{bmatrix} X & Y & Z \end{bmatrix}^T$, 3x1. Ένα τέτοιο διάνυσμα παίρνει τιμές

$$A \triangleq [X_{\min}..X_{\max}]x[Y_{\min}..Y_{\max}]x[Z_{\min}..Z_{\max}] \in \mathbb{Z}^3$$

με κάτω και άνω άκρα παντού 0 και 255 αντίστοιχα, όταν μιλάμε για χρωματικό χώρο RGB.
Έστω τώρα συνάρτηση *histogram*, δηλαδή το ιστόγραμμα, και ορίζεται σε

$$histogram : A \to \mathbb{N}$$

Κάθε ένα από τα σημεία του πεδίου ορισμού της *histogram* τα ονομάζουμε κάδους. Ο κάθε κάδος ορίζει γύρω του έναν κύβο, ακμών παράλληλων στους άξονες και μήκους

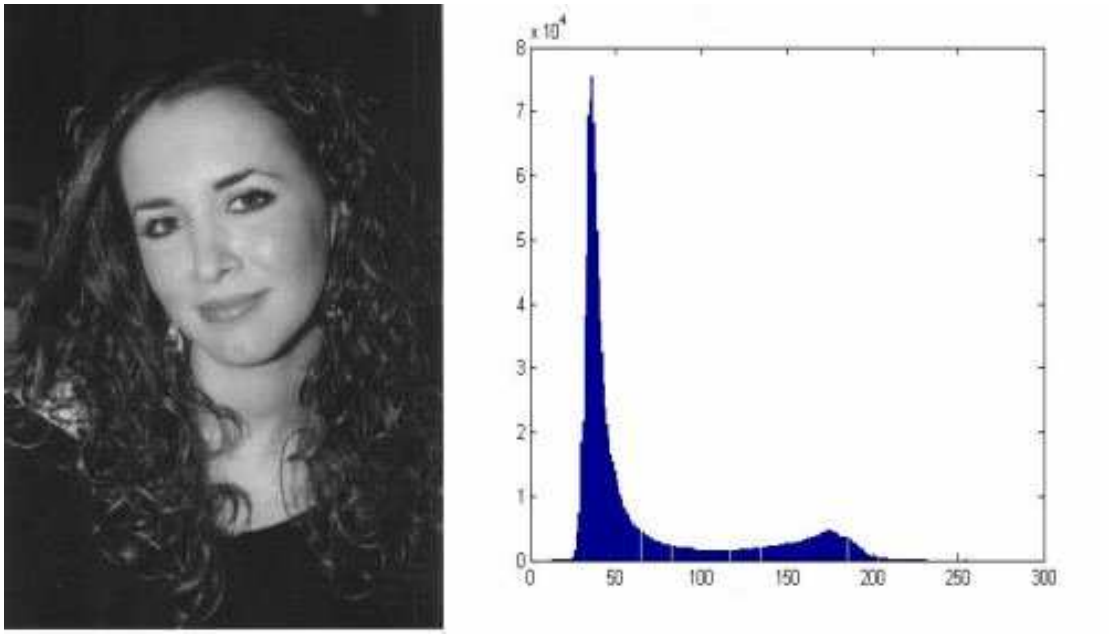
ίσου με μονάδα. Όταν δύο κάδοι έχουν αντίστοιχους κύβους με κοινή έδρα, τότε λέμε ότι αυτοί οι κάδοι είναι γειτονικοί.

Βλέπουμε ότι κάθε πιθανή χρωματική τιμή αντιστοιχεί σε ένα κάδο’ Ο κάθε κάδος ορίζεται τελικά να έχει τιμή όση και το *πλήθος* των αντίστοιχων χρωματικών τιμών που εμφανίζονται στην εικόνα. Ο αλγόριθμος κατασκευής του ιστογράμματος οπότε είναι απλά

1. Αρχικοποίησε συνάρτηση *histogram* , μηδενίζοντας όλους τους κάδους.
2. Σάρωσε pixel-pixel την εικόνα. Για κάθε pixel, βρες τον κάδο που αντιστοιχεί στην χρωματική τιμή του pixel. Αύξησε την τιμή σε αυτόν τον κάδο κατά 1.
3. Επανέλαβε για όλα τα pixels.
4. Τέλος.

Η όλη διαδικασία γενικεύεται εύκολα για οποιαδήποτε διάσταση έχουν οι χρωματικές τιμές. Επίσης, μία άλλη στρατηγική είναι να χρησιμοποιήσουμε λιγότερους κάδους απ’ ότι οι επιτρεπτές χρωματικές τιμές’ το γιατί θα θέλαμε να το κάνουμε αυτό θα το δούμε στην επόμενη ενότητα.

Στο σχήμα 4.2 για παράδειγμα βλέπουμε μια gray-scale εικόνα. Δίπλα της είναι το ιστόγραμμα της με 256 κάδους, και δεξιά το ιστόγραμμα της αυτή τη φορά με 16 κάδους. Το χρώμα περιγράφεται από βαθμωτό μέγεθος, την φωτεινότητα, επομένως το ιστόγραμμα είναι δισδιάστατο γράφημα. Στο σχήμα 4.3 βλέπουμε μία έγχρωμη εικόνα και το RGB ιστόγραμμα της. Κάθε τετράγωνο αντιστοιχεί σε ένα κάδο, και μεγάλο μέγεθος αντιστοιχεί σε υψηλή τιμή σε αυτόν τον κάδο.



Σχήμα 4.2. Grey-scale φωτογραφία κοπέλας, διαστάσεων 933x1329, και ιστόγραμμα φωτεινότητας.

Για να κάνουμε πιο φυσική τη σύνδεση με τους επόμενους παραγράφους που αφορούν στοχαστικές μεθόδους, κάνουμε την εξής παρατήρηση. Αν διαιρέσουμε κάθε στοιχείο της συνάρτησης histogram με το πλήθος των pixels της εικόνας, η συνάρτηση που παίρνουμε σαν αποτέλεσμα, έστω αυτή *histp*, θα είναι μια διακριτή συνάρτηση πυκνότητας πιθανότητας. Όντως ισχύουν για όλους τους κάδους x

$$0 \le histp(x) \le 1$$

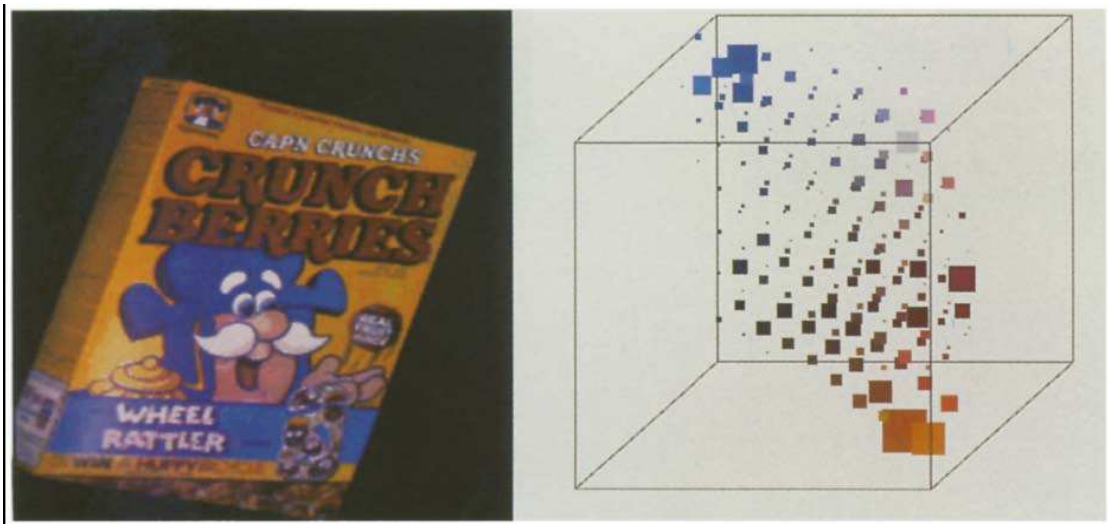
$$\sum histp(x) = 1 \qquad (4.1)$$

Η τελευταία σχέση ισχύει, αφού θα αυξήσουμε την τιμή κάποιου κάδου στο ιστόγραμμα όσες φορές επαναλάβουμε το αντίστοιχο βήμα στον αλγόριθμο. Όμως θα επαναληφθεί τόσες φορές όσα είναι ακριβώς και τα pixels της εικόνας. Άρα

$$\sum histogram(x) = |pixels|$$

που επιβεβαιώνει το (4.1).

Επιπλέον, το κανονικοποιημένο ιστόγραμμα *histp* έχει και μια φυσικά σημασία: Αν επιλέξω τυχαία ένα pixel από την εικόνα, η πιθανότητα να έχει χρωματική τιμή x δίνεται από την τιμή στο $histp(x)$.



Σχήμα 4.3. Φωτογραφία κουτιού δημητριακών και το αντίστοιχο RGB ιστόγραμμα.

Να παρατηρήσουμε ότι δεν μας εμποδίζει τίποτα να γενικεύσουμε την ιδέα του ιστογράμματος ακόμα περισσότερο και να χρησιμοποιήσουμε εκτός από χρώμα οποιαδήποτε χαρακτηριστικά. Μπορούμε δηλαδή να κατασκευάσουμε ιστόγραμμα χρησιμοποιώντας φέρ' ειπείν σαν δεδομένα, διανύσματα μορφής

$$\left[Col_1 \quad Col_2 \quad Col_3 \quad Tex_1 \quad Tex_2 \quad Tex_3 \right]^T$$

όπου χρησιμοποιήσαμε τρεις περιγραφείς για χρώμα και τρεις περιγραφείς για υφή.

4.2.2 Υπέρ και κατά του ιστογράμματος. Binning και Curse of dimensionality.

Τα σημεία που αναφέρονται συνήθως στην βιβλιογραφία [π.χ. 1] σαν τα δυνατά των ιστογραμμάτων χρώματος, είναι μη-μεταβλητότητα σε περιστροφή ή μετακίνηση της εικόνας. Πιο σωστό μάλλον θα ήταν πάντως να πούμε ότι αυτά είναι πλεονεκτήματα στην χρήση του χαρακτηριστικού, του χρώματος, σαν περιγραφέα μιας εικόνας, και όχι τόσο στην χρήση ιστογράμματος. Οι στοχαστικές μέθοδοι που θα δούμε

παρακάτω, διατηρούν αυτά τα πλεονεκτήματα όπως θα δούμε, όταν χρησιμοποιούμε χρώμα σαν χαρακτηριστικό.

Η βασική αδυναμία του ιστογράμματος είναι ότι *δεν είναι ανθεκτικό σε θόρυβο*. Μια μικρή προσθήκη θορύβου σε μια εικόνα, έχει σαν αποτέλεσμα να εξάγεται εντελώς διαφορετικό ιστόγραμμα' με άλλα λόγια εικόνες που για έναν άνθρωπο-παρατηρητή είναι ουσιαστικά ίδιες, θα έχουν ριζικά διαφορετική αναπαράσταση. Προφανώς αυτό δεν είναι επιθυμητό.

Το πρόβλημα πάντως, γνωστό και ως **binning**, λύνεται μερικώς. Υποθέτοντας ότι επιδρά μικρός θόρυβος σε μια αρχική εικόνα χωρίς θόρυβο, η επίδραση στο ιστόγραμμα θα είναι να μετακινηθούν πληθυσμοί από κάθε κάδο σε γειτονικούς του. Οπότε η 'καταστροφή' του ιστογράμματος από τον θόρυβο, αντισταθμίζεται με κατάλληλη επιλογή συνάρτησης απόστασης [5, 6], που να μετράει και να συγκρίνει όχι μόνο πληθυσμούς στους ίδιους κάδους, αλλά και πληθυσμούς μεταξύ παραπλήσιων κάδων. Κινούμενη πάνω στην ίδια ιδέα, μια παραλλαγή στην έννοια του ιστογράμματος προτείνεται στο [7] ("fuzzy color histograms").

Η εναλλακτική στρατηγική είναι να μειώσουμε τον αριθμό τον κάδων, ορίζοντας κάθε κάδο να αντιστοιχεί σε περισσότερες χρωματικές (ή οτιδήποτε άλλο) τιμές. Έτσι κάτω από θόρυβο, οι πληθυσμοί των κάδων υπόκεινται σε μικρότερη μεταβολή. Αυτό όμως έχει και το τίμημα, ότι όσο λιγοστεύουν οι κάδοι, τόσο λιγότερο περιγραφικό γίνεται το ιστόγραμμα, και αναπαριστά όλο και μεγαλύτερο εύρος-ποικιλία εικόνων, τόσο που τελικά ίσως να μην είναι το επιθυμητό. Σκεφτείτε για παράδειγμα την οριακή περίπτωση όπου έχουμε ένα χρωματικό ιστόγραμμα με μόνο 2 κάδους – θα μπορούσε να αντιστοιχεί στο ίδιο ιστόγραμμα κυριολεκτικά οτιδήποτε. Στο άλλο άκρο, βεβαίως βρίσκεται η χρήση υπερβολικού αριθμού κάδων που είναι ευπαθής σε θόρυβο. Η καλύτερη λύση βρίσκεται ανάμεσα στα δύο άκρα, αλλά πού; Η απάντηση δεν είναι προφανής, δεν μπορούμε να την ξέρουμε εκ των προτέρων, και εξαρτάται από τις εικόνες με τις οποίες δουλεύουμε.

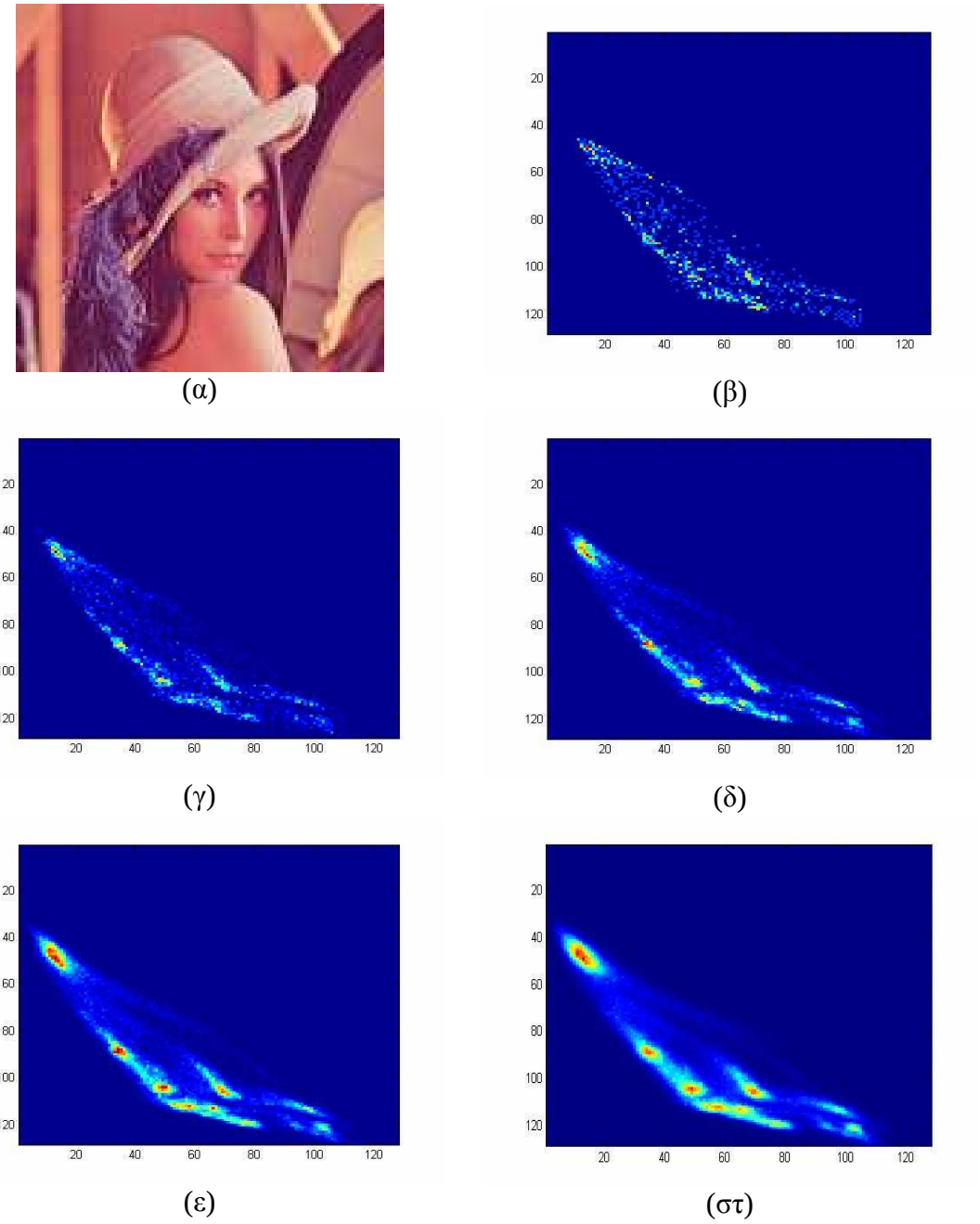
Το binning πρόβλημα γίνεται πιο έντονο όταν έχουμε λίγα δεδομένα σε σχέση με την διάσταση των δεδομένων μας. Είδαμε προηγουμένως το χρωματικό ιστόγραμμα, όπου τα δεδομένα είναι για έγχρωμες εικόνες 3-διάστατα. Όμως στην πράξη πλέον δεν χρησιμοποιείται μόνο το χρώμα [4, 8, 9], κινούμενος πάντως σε διαστάσεις μικρότερες του 10. Παρατηρούμε ότι αν αποφασίσουμε να χωρίσουμε κάθε διάσταση σε M μέρη, θα ορίζονται τελικά στο ιστόγραμμα M^d κάδοι, όπου d ο αριθμός των

χρησιμοποιούμενων χαρακτηριστικών. Δηλαδή οι κάδοι θα αυξάνονται εκθετικά σε σχέση με την διάσταση του προβλήματος, με αποτέλεσμα όταν τα δεδομένα είναι σχετικά με τους κάδους λίγα, να παράγονται αραιά ιστογράμματα.

Για παράδειγμα, έστω εικόνα με 10,000 pixels, και ιστόγραμμα με M = 20, d = 8. Οχτώ χαρακτηριστικά για κάθε pixel είναι αρκετά ρεαλιστικό σενάριο [4]. Τότε έχουμε $20^8 \cong 10^{10}$ κάδους, οπότε θα καταλαμβάνονται το πολύ μόνο 0,0001% των κάδων! Ένα τέτοιο ιστόγραμμα όμως, με τους περισσότερους κάδους ίσους με μηδέν, δεν είναι χρήσιμο σαν αναπαράσταση, κάτι που φαίνεται όταν επιχειρήσουμε να συγκρίνουμε δύο τέτοια ιστογράμματα. Με όποιον τρόπο και να το κάνουμε, οριακά αυτό θα είναι ανώφελο. Η εκθετική αύξηση σημαίνει ότι ακόμα και για μικρό d θα εμφανίζεται αυτός ο εκφυλισμός, που λέγεται **curse of dimensionality** [3].

4.2.3 Ένα ακόμα παράδειγμα σχετικά με την curse of dimensionality

Έστω η εικόνα του σχήματος 4.4(α). Φτιάχνουμε ιστογράμματα (κανονικοποιημένα ως προς τον αριθμό των δειγμάτων) χρησιμοποιώντας σαν χαρακτηριστικά τις εντάσεις του κόκκινου και του πράσινου στον RGB χώρο. Στο ιστόγραμμα 4(β) χρησιμοποιήσαμε 512^2 δείγματα από την εικόνα, στο (γ) 256^2, στο (δ) 128^2, στο (ε) 64^2, στο (στ) 32^2. Αυτά αντιστοιχούν σε subsampled εκδόσεις της ίδιας εικόνας. Έχουμε σύμφωνα με τη σημειολογία της προηγούμενης παραγράφου, M = 128, d = 2, δηλαδή σύνολο 128^2 κάδους.



Σχήμα 4.4. (α)Φωτογραφία προς επεξεργασία (Lenna). Ακολουθούν τα ιστογράμματα, με ένταση Κόκκινου στον οριζόντιο άξονα, και ένταση Πράσινου στον κατακόρυφο. Στα ιστογράμματα, κόκκινο παριστάνει υψηλή τιμή και μπλε χαμηλή. (β) 1,024 δείγματα (γ) 4,096 δείγματα (δ) 16,384 δείγματα (ε) 65,536 δείγματα (στ) 262,144 δείγματα.

Παρατηρούμε προχωρώντας από το ιστόγραμμα με τα λιγότερα δείγματα προς αυτό με τα περισσότερα, ότι το ιστόγραμμα τείνει σε μια σταθερή μορφή. Επιπλέον, όσο

περισσότερα δείγματα χρησιμοποιούμε, θα πρέπει να εξαφανίζεται και η curse of dimensionality, αφού αυτή εξαρτάται από τον λόγο

$$R_{cod} \triangleq \frac{N}{M^d}$$

όπου N ο αριθμός δειγμάτων. Το πρόβλημα είναι τόσο εμφανές όσο μικρότερος είναι αυτός ο λόγος. Πράγματι, το ιστόγραμμα με τα λιγότερα δείγματα παρουσιάζει πολλές ασυνέχειες και άδειους κάδους σε σχέση με τα περισσότερο ακριβή ιστογράμματα ($R_{cod} = 0.0625$). Στον αντίποδα, το ιστόγραμμα με τα περισσότερα δείγματα είναι καθαρά πιο συνεχές και ομαλό ($R_{cod} = 16$). Επιπλέον, υπάρχουν εμφανείς συγκεντρώσεις γύρω από κάποιες χρωματικές τιμές. Μπορούμε να βρούμε με το μάτι τουλάχιστον 8-9 τέτοια κέντρα, γύρω από τα οποία η τιμή του ιστογράμματος μειώνεται όσο απομακρυνόμαστε.

Η ερμηνεία αυτού του φαινομένου είναι ότι τα κέντρα αυτά αντιστοιχούν σε κυρίαρχα χρώματα στην εικόνα, χρώματα που παρατηρούνται περισσότερο. Και επειδή η φωτογραφία που εξετάζουμε είναι *πραγματική* (δηλ. όχι κάποιο ανθρώπινο κατασκεύασμα, όπως λογότυπο εταιρίας ή ένα γεωμετρικό σχήμα), είναι λογικό να εμφανίζονται πολύ και οι αποχρώσεις αυτών των κυρίαρχων χρωμάτων. Οι αποχρώσεις εκφράζονται με τις μικρότερες τιμές γύρω από τα κέντρα στο ιστόγραμμα.

4.3. Στοχαστικές μέθοδοι

4.3.1 Κίνητρο για χρήση στοχαστικών μεθόδων.

Είδαμε ότι όσο αυξάνεται ο αριθμός των δειγμάτων N και ο λόγος R_{cod}, τόσο καλύτερο ιστόγραμμα παίρνουμε, με την έννοια ότι εξαφανίζεται η curse of

dimensionality. Μάλιστα, *εικάζουμε* ότι για $N \to +\infty$, το ιστόγραμμα θα τείνει σε μια οριακή μορφή.

Τώρα βάζουμε την απαίτηση $M \to +\infty$ (αριθμός κάδων), που συνεπάγεται ότι το μέγεθος του κάθε κάδου $\to 0$. Σε συνδυασμό με την απαίτηση να αυξάνονται τα M και N με τέτοιον τρόπο ώστε $R_{cod} \to +\infty$, το αποτέλεσμα είναι να τείνει το ιστόγραμμα (κανονικοποιημένο πάντα ως προς τον αριθμό των δειγμάτων), σε μια συνεχή μορφή, ταυτόχρονα απαλλαγμένη από την curse of dimensionality.

Στην πράξη βεβαίως δεν μπορούμε να το κάνουμε αυτό. Οι εικόνες που έχουμε στην διάθεση μας είναι πάντα σε πεπερασμένη διάσταση, οπότε δεν γίνεται να αυξάνεται ο αριθμός των δειγμάτων απεριόριστα. Επίσης και η χρωματική τιμή είναι ένα κβαντισμένο μέγεθος, (καθώς και άλλα πιθανά χρησιμοποιούμενα χαρακτηριστικά βεβαίως), οπότε και ο αριθμός των κάδων δεν γίνεται να αυξάνεται απεριόριστα.

Το ερώτημα είναι, γίνεται να υπολογίσουμε αυτήν την ιδανική οριακή και συνεχή μορφή του ιστογράμματος, έχοντας στη διάθεση μας πεπερασμένο αριθμό δειγμάτων; Ή ακόμα, έχοντας λίγα δείγματα σε σχέση με την διάσταση του προβλήματος; Η απάντηση είναι, ότι μπορούμε να την *εκτιμήσουμε*.

4.3.2 Η προσέγγιση με εκτίμηση μίξης κανονικών κατανομών

Είδαμε ότι κανονικοποιώντας το ιστόγραμμα ως προς τον αριθμό των pixels, το κάνουμε συνάρτηση πυκνότητας πιθανότητας. Αυτή η ιδιότητα θα ισχύει βεβαίως και για την οριακή συνεχή μορφή του ιστογράμματος.

Οπότε με άλλα λόγια ζητάμε να εκτιμήσουμε μια συνεχή συνάρτηση πυκνότητας πιθανότητας (στο εξής *pdf*). Αυτό το γεγονός είναι βολικό, αφού για προβλήματα εκτίμησης *pdf* έχουμε στη διάθεση μας ένα σημαντικό οπλοστάσιο μεθόδων, μέρος των οποίων είδαμε στο κεφάλαιο 3 αυτής της εργασίας.

Η διαδικασία είναι η εξής.

Έστω εικόνα με $\mu x \nu$ pixels. Κάθε ένα pixel αντιστοιχεί σε ένα δεδομένο σε data set *X*. Δηλαδή $X = \{x_1, x_2, ..., x_N\}$, όπου $N = \mu * \nu$. Κάθε δεδομένο x_i έχει την διανυσματική δομή

$$x_i = [\, feat_{i1}, feat_{i2}, ..., feat_{id} \,]^T$$

όπου $feat_{ij}$ η τιμή του χαρακτηριστικού *j*, για το pixel *i*. Ως εδώ δεν έχουμε πει κάτι καινούριο.

Τώρα, χρησιμοποιούμε το data set *X* για να εκπαιδεύσουμε μοντέλο *pdf*. Τυπικά κάνουμε iid υπόθεση για τα δεδομένα μας, και διαλέγουμε μοντέλο μίξης κανονικών κατανομών. Αυτή η συνταγή έχει χρησιμοποιηθεί εκτενώς στην βιβλιογραφία, όσον αφορά κατανόηση, περιγραφή και ανάκτηση εικόνας [4, 8, 9, 14]. Επιπλέον, ένα επιπλέον κίνητρο για χρήση μικτού μοντέλου δίνεται από την ίδια τη μορφή που τείνουν να έχουν τα δεδομένα, όπως είδαμε στην παράγραφο 4.2.3.

Φορμαλιστικά λοιπόν, έχουμε μοντέλο

$$p(X; \Psi) = \sum_{i=1}^{N} \sum_{j=1}^{K} \pi_j \, Normal(x_i; \mu_j, \Sigma_j)$$

όπου παράμετροι $\Psi = \{\boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\Sigma}\}$, με $\boldsymbol{\pi} = \{\pi_1, \pi_2, ..., \pi_K\}$, $\boldsymbol{\mu} = \{\mu_1, \mu_2, ...\mu_K\}$, $\boldsymbol{\Sigma} = \{\Sigma_1, \Sigma_2, ..., \Sigma_K\}$. Ισχύουν $\pi_j \geq 0, \sum_{j=1}^{K} \pi_j = 1$.

Ο αριθμός πυρήνων K, μπορεί να πάρει κάποια αυθαίρετη τιμή, φερ' ειπείν 5 ή 6. Αυτό βέβαια εξαρτάται και από το τι εικόνα έχουμε να μεταχειριστούμε. Πάντως αντί να μπούμε σε μια τέτοια διαδικασία «εμπειρικής» επιλογής, καλύτερο είναι να χρησιμοποιήσουμε ένα κριτήριο, όπως AIC, MDL (βλ. κεφάλαιο 3), Laplace Empirical, κλπ.

Εμείς θα χρησιμοποιήσουμε το MDL στα πειράματα που θα δούμε στο κεφάλαιο 6. Δίνει καλά αποτελέσματα [11], είναι εύκολο στην υλοποίηση, και έχει χρησιμοποιηθεί στο πλαίσιο ανάκτησης εικόνας ευρέως, π.χ. [8, 9, 4].

Όσον αφορά την εκπαίδευση του μικτού μοντέλου, τη συζητήσαμε αρκετά στο κεφάλαιο 3, και ο ενδιαφερόμενος μπορεί να ανατρέξει εκεί.

Αφού γίνει λοιπόν η εκπαίδευση, κρατάμε την λύση μέγιστης πιθανοφάνειας Ψ^*' αυτή αρκεί για να περιγράψει το μοντέλο. Συγκεκριμένα δηλαδή κρατάμε *K* βάρη

(βαθμωτούς), K μέσα διάστασης d, δηλαδή Kd βαθμωτούς, και K πίνακες συμμεταβλητότητας δηλαδή $\frac{Kd(d+1)}{2}$ βαθμωτούς.

ΚΕΦΑΛΑΙΟ 5. ΡΩΤΩΝΤΑΣ ΓΙΑ ΕΙΚΟΝΑ ΚΑΙ ΓΙΑ ΤΜΗΜΑΤΑ ΕΙΚΟΝΑΣ

5.1. Εισαγωγικά

Στο παρόν κεφάλαιο φτάνουμε στον αρχικό στόχο μας, που ήταν να κάνουμε ερώτηση για εικόνα, με βάση το περιεχόμενο. Έχοντας στα χέρια μας κατάλληλη δομή που περιγράφει το περιεχόμενο κάθε εικόνας στην βάση (βλ. Κεφάλαιο 4), τώρα μένει να πούμε πως θα αξιοποιήσουμε αυτή την πληροφορία.

Κεντρικό ρόλο παίζει η έννοια της απόστασης περιγραφέων εικόνας. Πρώτα θα δούμε λοιπόν κάποιες συναρτήσεις απόστασης. Στη συνέχεια θα περάσουμε στις λεπτομέρειες της διαδικασίας της ερώτησης καθεαυτής.

5.2. Συναρτήσεις απόστασης.

5.2.1 Γενικά

Έστω X μη κενό σύνολο. Συνάρτηση απόστασης (*metric*) στο X τυπικά ονομάζεται μια πραγματική συνάρτηση d, διατεταγμένων ζευγών στοιχείων του Χ, που ικανοποιεί τις παρακάτω τρεις απαιτήσεις [18]:

1. $d(x,y) \geq 0$, και $d(x,y) = 0 \Leftrightarrow x = y$,
2. $d(x,y) = d(y,x)$ (συμμετρικότητα)
3. $d(x,y) \leq d(x,z) + d(z,y)$ (τριγωνική ανισότητα) (5.1)

Όπου $x, y, z \in X$. Στο πλαίσιο της ανάκτησης εικόνας, αυτά θα είναι περιγραφείς εικόνων, και X το σύνολο όλων των δυνατών περιγραφέων.

Προκειμένου να γίνει ερώτηση και ανάκτηση, πρέπει να έχουμε ένα τρόπο για να πούμε «αυτή η εικόνα *μοιάζει* περισσότερο με αυτή περισσότερη από εκείνη». Την ποσοτικοποίηση του κατά πόσο μοιάζουν δυο εικόνες, έρχεται να πραγματοποιήσει η συνάρτηση απόστασης μεταξύ περιγραφέων. Επιπλέον, ενώ οι απαιτήσεις (5.1) θα ήταν βολικές για μια συνάρτηση απόστασης, εν γένει δεν ισχύουν για τις συναρτήσεις που θα δούμε. Το κύριο ζητούμενο από μια συνάρτηση απόστασης στο πλαίσιο ανάκτησης εικόνας, είναι η απόσταση που δίνει για δύο οποιεσδήποτε εικόνες να συμπίπτει με την αντιληπτική απόσταση τους.

Δεν υπάρχει κάποια συνάρτηση απόστασης – «πανάκεια» γενικά. Έχουν προταθεί διάφορες επιλογές για συνάρτηση απόστασης. Εμείς θα δούμε μερικές από τις πιο ενδεικτικές και τις πιο πετυχημένες, όσον αφορά την ποιότητα των αποτελεσμάτων τους σε πραγματικά σενάρια ανάκτησης.

Χωρίζουμε τις αποστάσεις σε δύο κατηγορίες σε αυτή την εργασία. Στη μία έχουμε αποστάσεις για ιστογράμματα, στην άλλη αποστάσεις για πυκνότητες πιθανότητας. Αυτά τα δύο μοντέλα περιγραφέων εικόνας τα είδαμε στο προηγούμενο κεφάλαιο. Ειδικά στις αποστάσεις πυκνοτήτων μας ενδιαφέρουν αποστάσεις για μοντέλα μίξης κανονικών κατανομών.

5.2.2 Αποστάσεις για ιστόγραμμα

5.2.2.1 Τομή ιστογραμμάτων

Στο [1] προτείνεται η απόσταση “Histogram Intersection” (τομή ιστογραμμάτων), που ορίζεται

$$d_I(P,Q) \triangleq \frac{\sum_{i=1}^{bins} \min(P_i, Q_i)}{\sum_{i=1}^{bins} P_i} \quad (5.2)$$

όπου *P, Q* ιστογράμματα με *bins* αριθμό κάδων, και P_i, Q_i το πλήθος στον *i*-οστό κάδο. Η απόσταση παίρνει τιμές στο [0..1], και ισχύει η συμμετρική ιδιότητα.

5.2.2.2 Ευκλείδεια και Τετραγωνικής μορφής απόσταση

Ας δούμε το ιστόγραμμα σαν διάνυσμα διάστασης *binsx*1, στον $\mathbb{R}^{bins}$ χώρο. Τότε μια φυσική απόσταση είναι η ευκλείδεια, δηλαδή

$$d_E(P,Q) \triangleq \|P-Q\|_2 \quad (5.3)$$

Το πρόβλημα με αυτή την απόσταση είναι ότι μετράει ομοιότητα μόνο για αντίστοιχους κάδους (βλέπε και κεφάλαιο 4, σχετικά με το *binning* πρόβλημα). Αυτό μπορεί να διορθωθεί με μια τετραγωνικής μορφής απόσταση [5],

$$d_{Qd}(P,Q) \triangleq \sqrt{(P-Q)^T A (P-Q)} \quad (5.4)$$

όπου *A* πίνακας συσχέτισης, *binsxbins*. Για $A = I$ η (5.4) γίνεται η (5.3)’ με κατάλληλο *A* μπορούμε να συμπεριλάβουμε και αποστάσεις γειτονικών κάδων. Στο [4] προτείνεται ένας *A* που δίνει βάρος 1 για τον ίδιο κάδο, και 0.5 για γείτονες κάδους. Για $bins = 5$ δηλαδή, υποθέτοντας ότι ικανή και αναγκαία συνθήκη για είναι οι κάδοι γειτονικοί είναι να έχουν διαδοχική αρίθμηση (π.χ. 4 και 5), ο *A* θα είναι

$$A = \begin{bmatrix} 1 & 0.5 & 0 & 0 & 0 \\ 0.5 & 1 & 0.5 & 0 & 0 \\ 0 & 0.5 & 1 & 0.5 & 0 \\ 0 & 0 & 0.5 & 1 & 0.5 \\ 0 & 0 & 0 & 0.5 & 1 \end{bmatrix}$$

Πάνω στην ίδια συλλογιστική μπορούμε να διαλέξουμε να θεωρήσουμε βάρος για όλο και πιο μακρινούς κάδους' αυτονόητο είναι βέβαια ότι θα πρέπει να φθίνει όσο πιο μακρινοί είναι οι κάδοι. Στο [20] προτείνεται Gaussian βάρος συναρτήσει της απόστασης των κάδων. Σε μοντέλο ισοδύναμο με την τετραγωνικής μορφής απόσταση καταλήγουν και οι Han και Ma στο [7].

5.2.3. Αποστάσεις για συναρτήσεις πυκνότητας πιθανότητας

5.2.3.1 Συμμετρική Kullback – Leibler

Η προέλευση της απόκλισης Kullback-Leibler (*KL divergence*) είναι από θεωρία πληροφορίας [17]. Για πυκνότητες p, q, η απόκλιση ορίζεται

$$d_{KL}(p \| q) = \int p(x) \ln \frac{p(x)}{q(x)} dx \quad (5.5)$$

Όπου το ολοκλήρωμα είναι για ολόκληρο το πεδίο ορισμού των πυκνοτήτων. Μπορεί να αποδειχθεί ότι παίρνει πάντα μη-αρνητικές τιμές. Δεν είναι συμμετρική, ούτε και ισχύει η τριγωνική ανισότητα.

Μπορούμε να κάνουμε την απόκλιση KL συμμετρική πολύ απλά:

$$d_{SKL}(p,q) \triangleq d_{KL}(p \| q) + d_{KL}(q \| p) = \int p(x) \ln \frac{p(x)}{q(x)} dx + \int q(x) \ln \frac{q(x)}{p(x)} dx \quad (5.6)$$

Την οποία μορφή και θα χρησιμοποιήσουμε στη συνέχεια (κεφ. 6). Στην πράξη χρησιμοποιούμε προσομοίωση Monte-Carlo για να υπολογίσουμε τον 5.6, δηλαδή

$$d_{SKL}(p,q) \approx \frac{1}{N}\left(\sum_{x_i \in p} \ln \frac{p(x_i)}{q(x_i)} + \sum_{x_j \in q} \ln \frac{q(x_j)}{p(x_j)} \right)$$

όπου γίνεται δειγματοληψία N φορές, στο πρώτο άθροισμα από πυκνότητα *p*, στο δεύτερο από πυκνότητα *q*. Η πραγματική απόσταση προσεγγίζεται καλύτερα όσο $N \to +\infty$.

5.2.3.2 Bhattacharyya-based αποστάσεις

Η γενική μορφή της απόστασης Bhattacharyya [2] είναι

$$d_{Bh}\left(p, \tilde{p} \right) = -\int \left[p(x)\, \tilde{p}(x) \right]^{1/2} dx$$

όπου το ολοκλήρωμα είναι για όλο το πεδίο ορισμού του x.

Για δύο κανονικές πυκνότητες $p, \tilde{p}$ με μέσα $\mu, \tilde{\mu}$ και μήτρες συμμεταβλητότητας $\Sigma, \tilde{\Sigma}$ αντίστοιχα, η Bhattacharyya απόσταση έχει μορφή

$$d_{Bh}\left(p, \tilde{p} \right) = \frac{1}{8}(\mu - \tilde{\mu})^T \left(\frac{\Sigma + \tilde{\Sigma}}{2} \right)^{-1} (\mu - \tilde{\mu}) + \frac{1}{2} \ln \left(\frac{\left| \frac{\Sigma + \tilde{\Sigma}}{2} \right|}{2\sqrt{|\Sigma||\tilde{\Sigma}|}} \right) \quad (5.7)$$

Για μίξεις κανονικών κατανομών, μπορούμε να ορίσουμε την απόσταση

$$d_{BhGMM}\left(p, \tilde{p} \right) = \sum_{i=1}^{N} \sum_{j=1}^{M} \pi_i\, \tilde{\pi}_j\, d_{Bh}\left(p_i, \tilde{p}_j \right) \quad (5.8)$$

όπου $p, \tilde{p}$ μίξεις με N και M πυρήνες αντίστοιχα, $\pi_i, \tilde{\pi}_i$ τα αντίστοιχα βάρη, και $p_i, \tilde{p}_i$ οι αντίστοιχοι πυρήνες. Θα αναφερόμαστε σε αυτή την απόσταση σαν Bhattacharyya-GMM. Για αυτήν ισχύει η συμμετρική ιδιότητα. Επίσης προσέξτε ότι ενώ είναι πάντα μη αρνητική, δεν ισχύει η ανακλαστική ιδιότητα και *δεν* ισχύει απαραίτητα $d_{BhGMM}(p,p) \le d_{BhGMM}(p,q)$ για κάθε πυκνότητα *q*.

5.2.3.3 L_2 απόσταση

Στο [21] εισάγαμε πειραματικά την απόσταση

$$d_{L2}\left(p,\tilde{p}\right)=-\ln\left(\frac{2\int p(x)\,\tilde{p}(x)dx}{\int p(x)^2+\tilde{p}(x)^2dx}\right) \quad (5.9)$$

η οποία έχει κλειστή μορφή για την περίπτωση που οι πυκνότητες είναι κανονικές μίξεις. Είναι

$$d_{L2}\left(p,\tilde{p}\right)=-\ln\left(\frac{2\sum_{i=1}^{N}\sum_{j=1}^{M}\pi_i\,\tilde{\pi}_j\,\xi\left(p_i,\tilde{p}_j\right)^{-1/2}}{\sum_{i=1}^{N}\sum_{j=1}^{N}\pi_i\pi_j\xi\left(p_i,p_j\right)^{-1/2}+\sum_{i=1}^{M}\sum_{j=1}^{M}\tilde{\pi}_i\,\tilde{\pi}_j\,\xi\left(\tilde{p}_i,\tilde{p}_j\right)^{-1/2}}\right) \quad (5.10)$$

όπου

$$\xi(g,\tilde{g})=\left|\Sigma+\tilde{\Sigma}\right|\exp\left\{\mu^{\mathrm{T}}\Sigma^{-1}(\mu-m)+\tilde{\mu}^{\mathrm{T}}\tilde{\Sigma}^{-1}(\tilde{\mu}-m)\right\}$$

$$\text{με } m^T=\left(\mu^{\mathrm{T}}\Sigma^{-1}+\tilde{\mu}^{\mathrm{T}}\tilde{\Sigma}^{-1}\right)\left(\Sigma^{-1}+\tilde{\Sigma}^{-1}\right)^{-1}$$

και $g,\tilde{g}$ κανονικές κατανομές με μέσα και συμμεταβλητότητα μ,Σ και $\tilde{\mu},\tilde{\Sigma}$ αντίστοιχα. Βλέπε παράρτημα Β για τις πράξεις που χρειάζονται για να φτάσουμε στον (5.10) από (5.9).

5.2.3.4 EMD απόσταση (Earth Mover's distance)

Η Earth mover's distance [8] ορίζεται για $p,\tilde{p}$ κανονικές μίξεις, ως

$$d_{EMD}\left(p,\tilde{p}\right)=\frac{\sum_{i=1}^{N}\sum_{j=1}^{M}f_{ij}d_G\left(p_i,\tilde{p}_j\right)}{\sum_{i=1}^{N}\sum_{j=1}^{M}f_{ij}} \quad (5.11)$$

όπου $d_G\left(p_i, \tilde{p}_j\right)$ κάποια γνωστή απόσταση μεταξύ κανονικών πυρήνων ("ground distance"). Οι όροι f_{ij} διαλέγονται έτσι ώστε η ποσότητα (5.11) να ελαχιστοποιείται ως προς τους περιορισμούς

$$\sum_{j=1}^{M} f_{ij} \le \pi_i \forall i, \sum_{i=1}^{N} f_{ij} \le \tilde{\pi}_j \ \forall j$$

$$\sum_{i=1}^{N}\sum_{j=1}^{M} f_{ij} = \min\left(\sum_{i=1}^{N}\pi_i, \sum_{j=1}^{M}\tilde{\pi}_j\right) \quad (5.12)$$

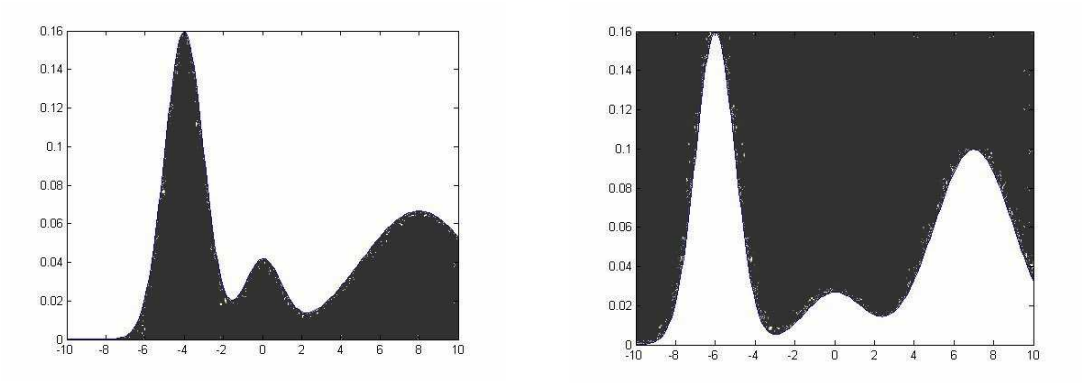
Αυτό είναι ένα πρόβλημα γραμμικού προγραμματισμού, που μπορεί να λυθεί φερ' ειπείν με μέθοδο Simplex ή Karmarkar κλπ. [15]. Όπως και στο [8] σαν ground distance θα χρησιμοποιήσουμε την απόσταση Frechet μεταξύ κανονικών κατανομών $g, \tilde{g}$:

$$d_{Frechet}\left(g, \tilde{g}\right) = \sqrt{\left\|\mu - \tilde{\mu}\right\|_2^2 + tr\left[\Sigma + \tilde{\Sigma} - 2\sqrt{\Sigma\tilde{\Sigma}}\right]} \quad (5.13)$$

η οποία είναι λύση σε κλειστή μορφή της (5.11) για κανονικές κατανομές.

Το όνομα και η ιδέα πίσω από την απόσταση είναι ενδιαφέροντα, και μας δίνουν μια καλύτερη κατανόηση της έννοιας της EMD.

Ας θεωρήσουμε την μάζα της μίας πυκνότητας σαν «χώμα» (earth) και της άλλης σαν «λάκκο» ή καλούπι. Στόχος είναι να μετακινήσουμε το χώμα της πρώτης κατανομής ώστε να ταιριάξει με το καλούπι της δεύτερης. Στην GMM περίπτωση μάλιστα, κάθε πυρήνας αντιστοιχεί αναλόγως σε ένα «λόφο» από χώμα ή «λόφο» στο καλούπι. Το έργο λοιπόν που θα απαιτηθεί για αυτή τη διαδικασία, αν γίνει με τον βέλτιστο δυνατό τρόπο, μας το δίνει η (5.11). Αντίστοιχα το βάρος του πυρήνα αντιστοιχεί σε ποσοστό μάζας και οι όροι f_{ij} αφορούν το ποσοστό χώματος που μετακινείται από τον λόφο *i* στον λόφο *j*. Οι περιορισμοί (5.12) τέλος εκφράζουν τη λογική απαίτηση ότι από ένα λόφο από χώμα δεν μπορούμε να μετακινήσουμε περισσότερο από το χώμα που έχει, και σε ένα λόφο στο καλούπι δεν μπορούμε να βάλουμε περισσότερο χώμα από όσο μπορεί να δεχτεί.



Σχήμα 5.1. Earth mover's distance: Η πυκνότητα αριστερά είναι το «χώμα» και η δεξιά το «καλούπι».

Να πούμε ότι, όπως υπονοείται στον περιορισμό (5.12), η EMD μπορεί να χρησιμοποιηθεί και για θετικές συναρτήσεις που έχουν μάζα μικρότερη από μονάδα. Αυτό επιτρέπει μερικό ταίριασμα ανάμεσα σε τμήματα εικόνων.

5.3. Ρωτώντας για εικόνες (Image querying)

5.3.1 Ρωτώντας για ολόκληρη εικόνα

Φθάσαμε λοιπόν στο αρχικό ζητούμενο της ανάκτησης εικόνας, την ερώτηση για εικόνα. Τυπικά υποθέτουμε ότι έχουμε στην διάθεση μας σύνολο από N εικόνες, έστω αυτό E. Δηλαδή

$$\mathbf{E} = \{E_i\}_{i=1}^{N}$$

όπου E_i οι εικόνες στη βάση μας. Θέλουμε να αναζητήσουμε εικόνες με παρόμοιο περιεχόμενο με κάποια εικόνα-ερώτηση, έστω αυτή E_q. Δεν έχει σημασία αν $E_q \in \mathbf{E}$ (αν και στα πειράματα που θα δούμε στο κεφάλαιο 6, ισχύει $E_q \in \mathbf{E}$).

Έστω D_i ο περιγραφέας για την εικόνα E_i. Είναι αυτονόητο ότι όλες τις εικόνες τις αντιστοιχούμε με ίδιου τύπου περιγραφέα, π.χ. για όλες φτιάχνουμε ιστόγραμμα για συγκεκριμένα χαρακτηριστικά, ή για όλες μικτό μοντέλα για ίδια χαρακτηριστικά. Παρόμοια και για την εικόνα-ερώτηση E_q, κατασκευάζουμε D_q.

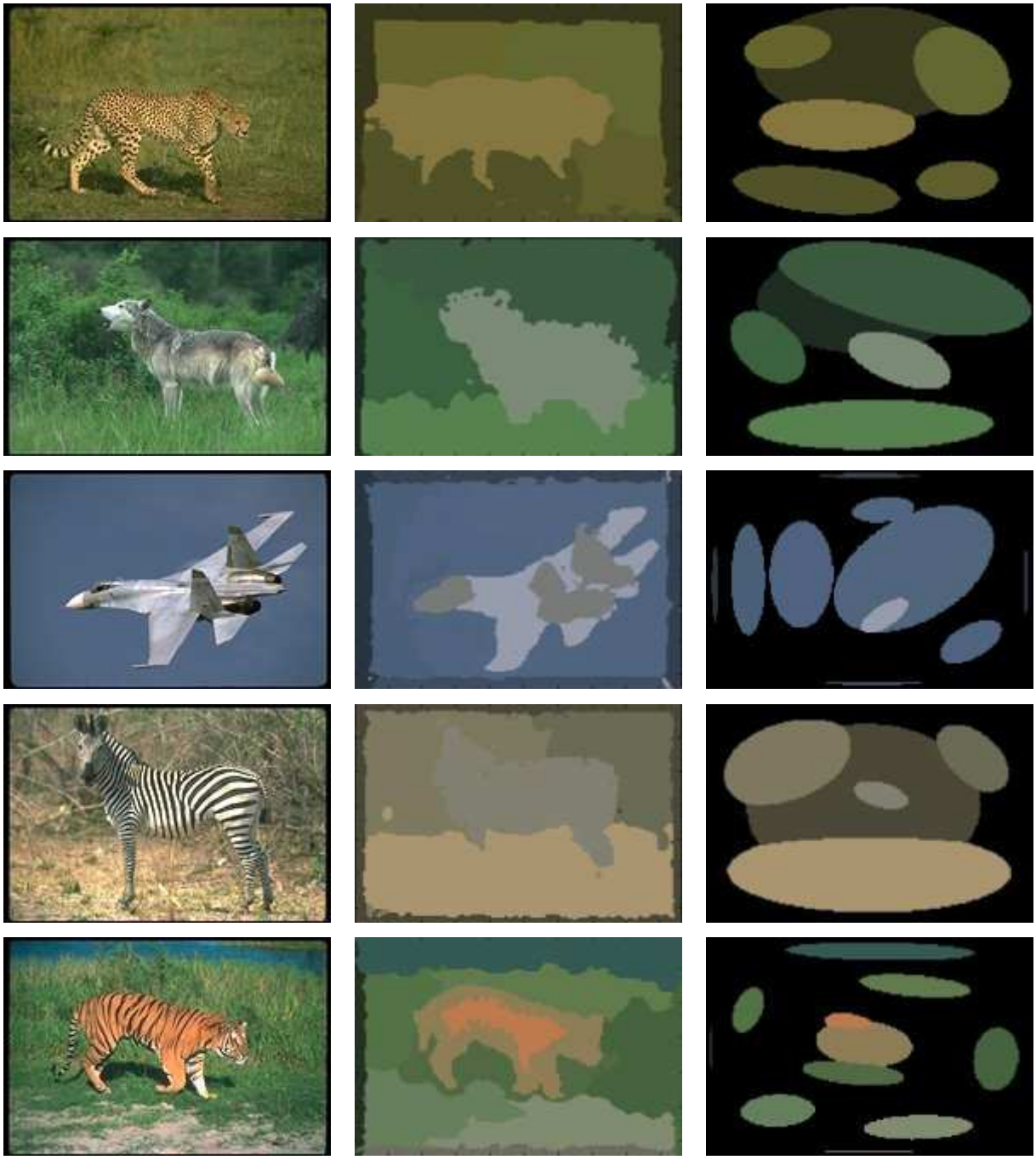
Διαλέγουμε κάποια συνάρτηση απόστασης *d*, και υπολογίζουμε τώρα $d(D_i, D_q)$, για κάθε $i \in [1, N]$. Το αποτέλεσμα της ερώτησης είναι οι εικόνες με την μικρότερη απόσταση – ο αριθμός αυτός εξαρτάται από τον χρήστη, και στην πράξη βέβαια θα είναι μικρός, της τάξης του 10 ή 20 (θα δούμε ότι μετακινώντας το κατώφλι του αριθμού εικόνων που δίνουμε σαν αποτέλεσμα, μπορούμε να φτιάξουμε τις χρήσιμες *Precision-Recall* καμπύλες).

Σχήμα 5.2. Παράδειγμα ανάκτησης εικόνας. Η εικόνα επάνω είναι η ερώτηση, και στην κάτω σειρά είναι τα 5 καλύτερα αποτελέσματα, από αριστερά προς τα δεξιά. Τυπικά θα υπάρχουν και λανθασμένα αποτελέσματα (αντίθετα προς την κοινή αντίληψη), όπως ο βράχος στην 4η από αριστερά εικόνα.

5.3.2 Ρωτώντας για τμήμα εικόνας

5.3.2.1 Γενικά

Συνήθως ρωτώντας για εικόνα, θέλουμε να ψάξουμε για κάποιο συγκεκριμένο χαρακτηριστικό ή αντικείμενο μέσα στην εικόνα. Ας πούμε στο 5.2 είναι πιθανό ο χρήστης να ζητούσε από την αρχή εικόνες που να περιέχουν μια κερασιά, και όχι απαραίτητα κερασιά με φόντο ουρανό και δέντρα.



Σχήμα 5.3. Κατάτμηση εικόνων με μέθοδο Blobworld. Στην πρώτη στήλη είναι οι εικόνες προς κατάτμηση, στην δεύτερη οι κατατμήσεις, με τα τμήματα χρωματισμένα ανάλογα με το μέσο του πυρήνα που προέρχονται. Στην τρίτη στήλη βλέπουμε τις προβολές των κανονικών πυρήνων στις διαστάσεις x-y.

Για να κάνουμε αναζήτηση για τμήμα (“*segment*”) εικόνας, πρέπει να ξέρουμε με κάποιο τρόπο τι τμήματα απαρτίζουν τις εικόνες στη βάση μας. Για παράδειγμα, στην εικόνα ερώτηση του 5.2 ένα τμήμα είναι η κερασιά, ένα τμήμα είναι ο ουρανός και ένα τα δέντρα πίσω από την κερασιά. Βέβαια αυτός ο διαχωρισμός δεν είναι απόλυτος. Θα μπορούσε να πει κανείς ότι τα δέντρα στο φόντο κάνουν από ένα τμήμα το καθένα, κ.ο.κ.

Εμείς για να κάνουμε αναζήτηση για τμήμα εικόνας, θα βασιστούμε σε πρώτο στάδιο στο πλαίσιο που εισάγεται στο [4]. Αυτό προβλέπει να γίνει κατάτμηση για όλες τις εικόνες στην βάση, πριν συνεχίσουμε.

5.3.2.2 Κατάτμηση εικόνας (a la Blobworld)

Στο Blobworld [4] προβλέπεται εκπαίδευση μικτού μοντέλου κανονικών κατανομών, παίρνοντας χαρακτηριστικά (διάνυσμα χαρακτηριστικών) για κάθε pixel

$$\left[L^* \quad a^* \quad b^* \quad p \quad n \quad c \quad x \quad y \right]^T$$

όπου τα πρώτα 3 χαρακτηριστικά είναι τα γνωστά κανάλια του CIELAB, αφού υποστούν *smoothing* με 2D Gaussian κλίμακας που δίνεται με τον τρόπο που περιγράφεται στην ενότητα, τα επόμενα 3 είναι περιγραφείς για υφή (polarity, anisotropy, contrast) και τέλος οι συντεταγμένες στην εικόνα του pixel (βλ. κεφάλαιο 2 για τα παραπάνω). Ο αριθμός των πυρήνων του μοντέλου διαλέγεται σύμφωνα με MDL[4] (βλ. κεφάλαιο 3), έστω ο αριθμός τους K.

Το πρώτο βήμα είναι, να αντιστοιχίσουμε κάθε pixel σε ένα πυρήνα. Αυτή η αντιστοίχιση γίνεται με τον κανόνα

[4] Στην πραγματικότητα οι συγγραφείς του Blobworld δεν χρησιμοποιούν *ακριβώς* κριτήριο MDL, τουλάχιστον όπως φαίνεται στον κώδικα MATLAB που δημοσιεύουν [22], και σε αντίθεση με ό,τι λένε στο [4]. Πολλαπλασιάζουν τον όρο ποινής του MDL με μια αυθαίρετη σταθερά της τάξης 30~50, οπότε ευνοούν την επιλογή μοντέλων με λιγότερους πυρήνες.

$$\arg\max_{j} p(j \mid x)$$

όπου *x* το αντίστοιχο διάνυσμα χαρακτηριστικών του pixel, και η πιθανότητα είναι η εκ των υστέρων πιθανότητα να «προήλθε» το datum *x* από τον πυρήνα *j*. Αυτή η διαδικασία παράγει μια *K-level* εικόνα, έστω αυτή *R*, με $\geq K$ πλήθος συνδεδεμένων περιοχών.

Το επόμενο είναι βήμα προεπεξεργασίας στην *R*.

1. Βρες το ιστόγραμμα κάθε περιοχής, σύμφωνα με τα χρώματα της αρχικής εικόνας.
2. Για κάθε pixel (στο bin αριθμημένο *i)* σε σύνορο μεταξύ δύο ή περισσότερων περιοχών, ανακατέταξε το στην περιοχή όπου η τιμή του ιστογράμματος *i* είναι μεγαλύτερη[5]. Ανανέωσε την *R*.

Επανέλαβε αυτά τα βήματα 4 φορές. (εμπειρικό νούμερο). Αυτή η διαδικασία διορθώνει τα σύνορα της *R* ώστε να ταιριάζουν σύνορα αντιληπτικά πραγματικών περιοχών στην πρωτότυπη εικόνα.

Τέλος, στην *R* εφαρμόζουμε ένα maximum-vote φίλτρο 3x3. Αυτό σημαίνει ότι για κάθε pixel κοιτάμε τους άμεσους γείτονες του, και το ανακατατάσσουμε στην περιοχή που ανήκουν οι περισσότεροι γείτονες. Για παράδειγμα οι γείτονες

3 3 3
1 1 3
1 1 3

τότε το κεντρικό pixel θα πρέπει να ανακαταταχθεί στην περιοχή 3. Η *R* είναι τώρα η κατάτμηση που ζητάμε (βλ. σχήμα 5.3, δεύτερη στήλη).

5.3.2.3 Ερώτηση για ένα μόνο τμήμα

Στο Blobworld, γίνεται κατάτμηση σε κάθε εικόνα στην βάση εικόνων που διαθέτουμε. Αντιστοιχίζεται ένας περιγραφέας για κάθε ένα τμήμα, και από ‘κει και

[5] Στον κώδικα που δημοσιεύουν οι συγγραφείς [22], η απαίτηση για αυτό είναι η υποψήφια νέα περιοχή να έχει τουλάχιστον διπλάσια ιστογραμματική τιμή για το bin *ι*, από αυτήν που έχει η τρέχουσα περιοχή που ανήκει το pixel.

πέρα η ερώτηση πραγματοποιείται όπως είπαμε στην ενότητα 5.3.1(«Ρωτώντας για ολόκληρη εικόνα») συγκρίνοντας πλέον περιγραφείς τμημάτων αντί για περιγραφείς εικόνων.

Ο τρόπος περιγραφής κάθε τμήματος, είναι ιστόγραμμα για το χρώμα (CIELAB) με 5 κάδους για κανάλι L^*, 10 κάδους για κανάλι a^*, 10 κάδους για κανάλι b^*. Μαζί με το ιστόγραμμα «επισυνάπτεται» και μια επιπλέον πληροφορία για υφή – ένας κάδος που κρατάει το μέσο contrast και το μέσο anisotropy *x* contrast της περιοχής (το anisotropy δεν έχει νόημα σε περιοχές χαμηλού contrast). Το polarity δεν συμπεριλαμβάνεται γιατί παίρνει μεγάλες τιμές κυρίως μόνο επάνω σε ακμές, και επομένως δεν χρησιμεύει σαν χαρακτηριστικό για σύγκριση τμήματος με άλλο τμήμα. Σαν συνάρτηση απόστασης χρησιμοποιείται η *quadratic form*, με βάρος 0.5 για γειτονικούς κάδους.

Μπορούμε πάντως να ξεφύγουμε από αυτή τη συνταγή. Κάνουμε την υπόθεση, ότι χοντρικά θα πρέπει ένα τμήμα στην εικόνα να αντιστοιχεί σε ένα πυρήνα του μοντέλου GMM. (Στην πραγματικότητα για κάθε πυρήνα αντιστοιχούν ένα ή περισσότερα από ένα τμήματα).

Υπολογίζουμε για κάθε τμήμα εικόνας, αντί ιστογράμματος, το μέσο και τη συμμεταβλητότητα για το χαρακτηριστικό διάνυσμα

$$\left[L^* \quad a^* \quad b^* \quad c \quad n*c \right]^T$$

όπου c - contrast, n – anisotropy. Αυτό είναι ανάλογο με τα χαρακτηριστικά που χρησιμοποιούνται για περιγραφείς στο Blobworld. Το μέσο και η συμμεταβλητότητα θεωρούμε ότι χαρακτηρίζουν κάποιον κανονικό πυρήνα, όπως είπαμε, και ένας τέτοιος θα περιγράφει κάθε τμήμα.

Οπότε τώρα μπορούμε να κάνουμε ερώτηση για τμήμα, χρησιμοποιώντας κάποια από τις συναρτήσεις απόστασης για μικτά μοντέλα κανονικών κατανομών. Στην περίπτωση αυτή βεβαίως έχουμε να συγκρίνουμε μονήρεις κανονικούς πυρήνες - αυτό δεν μας εμποδίζει. Ονομαστικά δηλαδή μπορούμε να χρησιμοποιήσουμε τις αποστάσεις Symmetric Kullback-Leibler, L2, Bhattacharyya, EMD. Τα αποτελέσματα δεν είναι άσχημα σε σχέση με την μέθοδο του ιστογράμματος του Blobworld (βλ. κεφάλαιο 6 για αποτελέσματα).

5.3.2.4 Ρωτώντας για πολλαπλά τμήματα (compound query)

Μπορούμε να κάνουμε ερωτήσεις για πολλαπλά τμήματα ταυτόχρονα, χρησιμοποιώντας λογικούς τελεστές στην ερώτηση. Για παράδειγμα, ψάχνω για το

τμήμα #1 ***και*** *για το (τμήμα #2* ***ή*** *το τμήμα #3)* (5.14)

Τέτοιου είδους ερώτηση την ονομάζουμε σύνθετη ερώτηση. Αυτή αντιμετωπίζεται ως εξής. Έστω ερώτηση μορφής

$$\phi(S) \quad (5.15)$$

όπου S σύνολο n τμημάτων-ερωτήσεων σε μια εικόνα, και φ λογική συνάρτηση μεταξύ των τμημάτων-ερωτήσεων, όπως είδαμε δηλαδή στο παράδειγμα (5.14), με τελεστές «και» και «ή». Δεν είναι απαραίτητο βέβαια στην ερώτηση να συμμετέχουν όλα τα τμήματα της εικόνας.

Για κάθε εικόνα j στην βάση, υπολογίζουμε για κάθε $s_i \in S$, την απόσταση του s_i με κάθε τμήμα της j – από αυτές κρατάμε την μικρότερη σε κάθε εικόνα, έστω αυτή d_{ij}.

Τώρα, η απόσταση που θέτουμε για την εικόνα j, μας δίνεται από την συνάρτηση φ, αν αντικαταστήσουμε κάθε s_i με d_{ij} ,κάθε λογικό τελεστή «και» με *max* και κάθε λογικό τελεστή «ή» με *min*. Δηλαδή η αντίστοιχη απόσταση για το (5.14) είναι

$$\max\{d_{1j}, \min\{d_{2j}, d_{3j}\}\}$$

Αφού αντιστοιχίσουμε λοιπόν μία απόσταση με κάθε εικόνα στη βάση, επιστρέφουμε σαν απάντηση στην ερώτηση τις εικόνες με ταξινομημένη σειρά ως προς την απόσταση, όπως ακριβώς δηλαδή κάναμε στην περίπτωση ερώτησης για ολόκληρη εικόνα.

Να σημειώσουμε ότι δεν υπάρχει κανένας περιορισμός στην συνάρτηση απόστασης και το είδος των περιγραφέων που πρέπει να χρησιμοποιήσουμε, και δεν είναι ανάγκη να μείνουμε στο πλαίσιο ιστόγραμμα/quadratic form distance , όπως είδαμε στην προηγούμενη παράγραφο (5.3.2.3).

5.3.2.5. Ρωτώντας για region of interest

Ένα άλλο είδος ερώτησης που μπορούμε να κάνουμε, και δεν συμπεριλαμβάνεται στο [4], είναι ερώτηση για μια περιοχή στην εικόνα οριοθετημένη από τον χρήστη.

Βεβαίως περιγράψαμε έναν τρόπο να γίνεται αυτόματα κατάτμηση κάθε εικόνας’ όμως συμβαίνει συχνά το αποτέλεσμα της κατάτμησης να μην είναι ακριβώς ή ακόμα να απέχει αρκετά από αυτό που ζητάμε. Για παράδειγμα, στο σχήμα 5.3 μπορεί να θέλουμε να κάνουμε αναζήτηση για όλη τη ζέβρα (τα πόδια της λανθασμένα δεν συμπεριλαμβάνονται στο ίδιο τμήμα), ή μπορεί ακόμη να θέλουμε να κάνουμε αναζήτηση για το κεφάλι μόνο του λύκου (όλος ο λύκος αποτελεί ένα τμήμα).

Εναλλακτικά, αυτό που μπορούμε να κάνουμε είναι να ζητήσουμε από τον χρήστη να οριοθετήσει μια περιοχή πάνω στην αρχική εικόνα (*region of interest*). Στη συνέχεια κατασκευάζουμε περιγραφέα όπως ακριβώς θα κάναμε αν είχαμε να κάνουμε με τμήμα που υπολογίσαμε αυτόματα. Όμοια είναι και η συνέχεια (βλ.5.3.2.3).

Το αρνητικό είναι ότι, ενώ ο χρήστης μπορεί να διαλέξει μόνος του την περιοχή που θέλει για ερώτηση, δεν μπορεί να γίνει το ίδιο με τα τμήματα εικόνων (λόγω μεγάλου πλήθους) με τα οποία θα γίνει η σύγκριση. Αναγκαστικά βασιζόμαστε σε αυτόματη κατάτμηση για αυτά.

ΚΕΦΑΛΑΙΟ 6. ΠΕΙΡΑΜΑΤΑ ΚΑΙ ΥΛΟΠΟΙΗΣΗ

6.1. Γενικά

Έχουμε υλοποιήσει με κώδικα MATLAB την πλειονότητα των μεθοδολογιών που παρουσιάστηκαν σε αυτή την εργασία. Προκειμένου να εκτιμήσουμε την αποτελεσματικότητα τους, και να τις συγκρίνουμε μεταξύ τους, τρέξαμε κάποια πειράματα τα οποία και θα δούμε στη συνέχεια.

Χρησιμοποιήσαμε για αυτά τα πειράματα δύο ξεχωριστές βάσεις δεδομένων. Η πρώτη περιέχει 200 εικόνες, χωρισμένες σε 5 θεματικές κατηγορίες. Η δεύτερη περιέχει 4330 εικόνες, χωρισμένες σε 23 θεματικές κατηγορίες. Θα αναφερόμαστε σε αυτές αντίστοιχα σαν βάση Α και βάση Β. Περισσότερα σχετικά με τις βάσεις αυτές μπορούμε να δούμε στο παράρτημα Γ.

Πρώτα όμως ας περάσουμε στις μεθόδους που χρησιμοποιούμε για αξιολόγηση των αποτελεσμάτων μας.

6.2. Μέθοδοι αξιολόγησης

6.2.1 Καμπύλη Precision – Recall

Η πρώτη μέθοδος αξιολόγησης, είναι η καμπύλη *Precision – Recall* [18] (*PR Curve*). Περιγράφει την ορθότητα στα αποτελέσματα για μία ακριβώς ερώτηση. Αυτή η καμπύλη είναι η συνάρτηση του *Precision* προς το *Recall*, τα οποία ορίζονται ως εξής:

$$Recall \triangleq \frac{\text{Σχετικές με την ερώτηση ανεκτηθέντες εικόνες}}{\text{Συνολικός αριθμός } \textbf{σχετικών με την ερώτηση εικόνων στη βάση}}$$

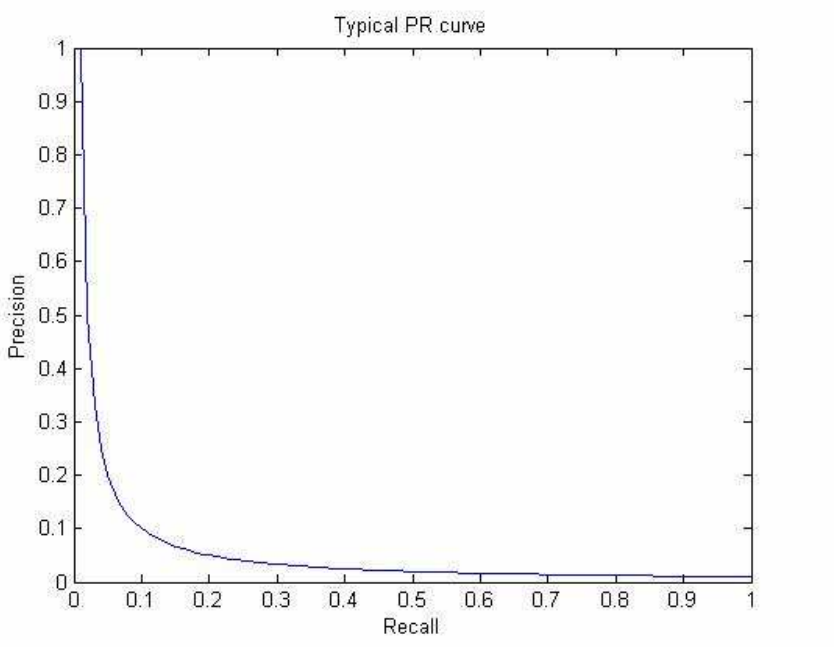
$$Precision \triangleq \frac{\text{Σχετικές με την ερώτηση ανακτηθέντες εικόνες}}{\text{Συνολικός αριθμός } \textbf{ανακτηθέντων εικόνων}}$$

Precision και *Recall* παίρνουν προφανώς πραγματικές τιμές από 0 έως 1, και εξαρτώνται από τον συνολικό αριθμό ανακτηθέντων εικόνων $\boldsymbol{\tau}$. Σε ένα πραγματικό σενάριο ανάκτησης εικόνας, αυτό το κατώφλι εξαρτάται από τον χρήστη. Αφού γίνει η ερώτηση, το σύστημα ανάκτησης εικόνας πρέπει να ταξινομήσει όλες τις εικόνες στην βάση ανάλογα με τον βαθμό ομοιότητας (ή ισοδύναμα, ανομοιότητας) με την εικόνα ερώτηση. Από αυτές επιστρέφονται σαν απάντηση τα $\boldsymbol{\tau}$ καλύτερα ταιριάσματα.

Παρατηρούμε ότι αυξάνοντας το κατώφλι από 1 μέχρι max εικόνες στην βάση, αναγκαστικά αυξάνεται (όχι γνησίως) το *Recall*. Κάθε τιμή του *Recall* αντιστοιχεί σε έναν αριθμό από τιμές κατωφλίου, και σε κάθε τιμή του κατωφλίου αντιστοιχεί μία τιμή του *Precision*. Συμβατικά θα κρατήσουμε την μικρότερη από τις τιμές κατωφλίου για το ίδιο *Recall*, ώστε να μπορέσουμε να αντιστοιχίσουμε επομένως κάθε τιμή του *Recall* σε μια μοναδική τιμή *Precision*.

Στο σχήμα 6.1 βλέπουμε μια τυπική καμπύλη Precision – Recall. Όσο πιο μεγάλες τιμές παίρνει η καμπύλη, σημαίνει τόσο καλύτερα αποτελέσματα ανάκτησης έχουμε. Επίσης για να συνοψίσουμε καλύτερα την πληροφορία που μας δίνει η καμπύλη,

μπορούμε να μετρήσουμε το εμβαδό κάτω από την καμπύλη, για διάφορα διαστήματα *Recall*. Εμείς διαλέξαμε να διαμερίσουμε το πεδίο ορισμού του Recall σε 5 ίσα μέρη, δηλαδή 0-20%, 20%-40%, 40%-60%, 60%-80%, 80%-100%.



Σχήμα 6.1. Τυπική καμπύλη Precision – Recall.

Να πούμε βέβαια ότι από μία ερώτηση δεν μπορούμε να αξιολογήσουμε ένα σύστημα ανάκτησης εικόνας' μπορεί να έτυχε να έχει πολύ κακή, ή πολύ καλή απόκριση για τη συγκεκριμένη ερώτηση. Η λύση είναι να χρησιμοποιήσουμε πληροφορία για πολλές ερωτήσεις, και να φτιάξουμε την μέση καμπύλη για αυτές. Όσο περισσότερες καμπύλες χρησιμοποιούμε, τόσο καλύτερη αίσθηση έχουμε αν δουλεύει καλά το σύστημα μας.

6.2.2 Αποστάσεις μεταξύ κατηγοριών εικόνων

Έστω ότι μπορούμε να χωρίσουμε τις εικόνες στην βάση μας σε *M* ομάδες. Οι ομάδες είναι τέτοιες ώστε κάθε εικόνα στην ίδια ομάδα είναι σχετική με τις άλλες εικόνες

στην ίδια ομάδα, και μάλιστα *μόνο* με αυτές. Ορίζουμε την απόσταση ομάδας με ομάδα σαν

$$\mathbf{d}_{class}(P,Q) \triangleq \frac{1}{|P|}\frac{1}{|Q|}\sum_{\forall x \in P}\sum_{\forall y \in Q} d(x,y)$$

όπου P, Q δύο ομάδες εικόνων, και *d* η συνάρτηση απόστασης που χρησιμοποιούμε για τους περιγραφείς των εικόνων.

Τώρα, σύμφωνα με την διαίσθηση μας, θα πρέπει η απόσταση μιας ομάδας με τον εαυτό της να είναι μικρότερη από την απόσταση με κάθε άλλη ομάδα εικόνων, αν βεβαίως το σύστημα ανάκτησης δουλεύει σωστά. Μάλιστα, όσο μικρότερη είναι η απόσταση της ομάδας από τον εαυτό της, και όσο μεγαλύτερη είναι από τις άλλες ομάδες, τόσο το καλύτερο.

Θα χρησιμοποιήσουμε και τις δύο παραπάνω μεθόδους αξιολόγησης στη συνέχεια. Να παρατηρήσουμε ότι και στις δύο υποθέσαμε σιωπηρά ότι έχουμε στα χέρια μας την λεγόμενη *ground truth*, δηλαδή ότι η βάση που χρησιμοποιούμε για να κάνουμε τις αξιολογήσεις είναι ήδη διαχωρισμένη σε ένα αριθμό από ομάδες εικόνων. Αυτό βέβαια δεν ισχύει σε πραγματικό σενάριο ανάκτησης εικόνας (δεν θα είχε νόημα η ανάκτηση!), αλλά γίνεται μόνο χάριν της αξιολόγησης του συστήματος ανάκτησης.

6.3. Πειράματα

6.3.1 Βάση A

Υπολογίσαμε τις αποστάσεις μεταξύ ομάδων εικόνων στην βάση A, αφού εκπαιδεύσαμε μοντέλα GMM για όλες τις εικόνες, χρησιμοποιώντας εμπειρικά 5 πυρήνες για το καθένα, διάνυσμα χαρακτηριστικών RGB (*3x1*, χρώμα), και τρεις διαφορετικές συναρτήσεις απόστασης, τις Bhattacharyya-GMM, L_2, και Symmetric Kullback-Liebler. Τα αποτελέσματα φαίνονται στον πίνακα 6.1. Οι χρόνοι για τους υπολογισμούς φαίνονται στον πίνακα 6.3. Ο μεγάλος χρόνος που χρειάζεται η

	Cherries	Arborgreens	Football	Cannonbeach	Campus
Cherries	1	1,12	1,12	1,43	1,67
Arborgreens	2,84	1	1,87	2,57	2,94
Football	4,96	3,26	1	6,98	3,87
Cannonbeach	1,88	1,32	2,07	1	2,35
Campus	2,94	2,03	1,54	3,15	1

(α)

	Cherries	Arborgreens	Football	Cannonbeach	Campus
Cherries	1	1,55	1,08	1,28	1,91
Arborgreens	1,89	1,05	1	2,78	1,92
Football	1,66	1,25	1	2,34	1,76
Cannonbeach	1	1,77	1,19	1,02	1,87
Campus	1,65	1,36	1	2,08	1,36

(β)

	Cherries	Arborgreens	Football	Cannonbeach	Campus
Cherries	1	1,64	1,69	1,45	1,5
Arborgreens	1,75	1	1,91	2,15	1,42
Football	2,28	2,43	1	2,67	1,82
Cannonbeach	1,12	1,56	1,52	1	1,5
Campus	1,57	1,39	1,4	2,02	1

(γ)

Πίνακας 6.1. Αποστάσεις μεταξύ ομάδων για βάση Α, για διαφορετικές συναρτήσεις απόστασης: (α) Για SKL, (β) Για Bh-GMM, (γ) Για L_2.

Symmetric Kullback-Liebler, οφείλεται στο ότι χρησιμοποιούμε Monte-Carlo προσομοίωση για να την υπολογίσουμε, «τραβώντας» τα δείγματα μας από τα pixels

των πραγματικών εικόνων. Παρά ότι χρησιμοποιούνται 4096 δείγματα από κάθε εικόνα (1% των pixels μιας 700x500 εικόνας) είναι αρκετά αργή. Προσέξτε ότι οι αποστάσεις είναι κανονικοποιημένες, ως εξής: Διαιρέσαμε τα στοιχεία κάθε γραμμής με την μικρότερη τιμή της ίδιας γραμμής (αυτός είναι και ο λόγος που οι πίνακες δεν είναι συμμετρικοί), επομένως το στοιχείο με την μικρότερη τιμή θα έχει κανονικοποιηθεί σε μονάδα. Επομένως, όποτε οι μονάδες είναι στην κύρια διαγώνιο, το σύστημα ανάκτησης με αυτή τη σύνθεση και απόσταση δουλεύει ικανοποιητικά.
Στη συνέχεια για να αξιολογήσουμε την ανθεκτικότητα του συστήματος ανάκτησης, συγκρίνουμε κάθε ομάδα εικόνων με τις ίδιες εικόνες σε μικρότερη ανάλυση

	Cherries	Arborgreens	Football	Cannonbeach	Campus
S-Cherries	$3{,}6^{\varepsilon}16$	$2{,}8^{\varepsilon}19$	1	1,21	$5{,}7^{\varepsilon}19$
S-Arborgr.	2,14	1,21	1	2,87	2,14
S-Football	1,86	1,38	1	2,45	1,94
S-Cannonb.	1	2,12	1,15	1	2,86
S-Campus	1,8	1,56	1	2,12	1,7

(α)

	Cherries	Arborgreens	Football	Cannonbeach	Campus
S-Cherries	1	1,56	1,66	1,39	1,52
S-Arborgr.	1,5	1	1,62	1,68	1,27
S-Football	2,17	2,41	1	2,22	1,82
S-Cannonb.	1,23	1,74	1,67	1	1,69
S-Campus	1,47	1,39	1,31	1,67	1

(β)

Πίνακας 6.2. Αποστάσεις μεταξύ ομάδων εικόνων και ομάδων subsampled εικόνων για βάση A, για διαφορετικές συναρτήσεις απόστασης: (α) Για Bh-GMM, (β) Για L_2. Το πρόθεμα S- δείχνει ομάδων subsampled εικόνων.

(subsampled), συγκεκριμένα με μισό πλάτος και μισό ύψος. Πάλι το ζητούμενο είναι η μικρότερη απόσταση να δίνεται όταν συγκρίνεται κάθε ομάδα με την subsampled

εκδοχή της. Τα αποτελέσματα φαίνονται στον πίνακα 6.2, και εδώ δεν έχουμε συμπεριλάβει την SKL απόσταση λόγω της κακής ταχύτητας της στους προηγούμενους υπολογισμούς.
Οι υπολογισμοί έγιναν σε PC Pentium 2.4 GHz. Το συμπέρασμα που βγαίνει από αυτή την σειρά πειραμάτων, είναι ότι οι L_2 και SKL αποστάσεις έχουν εξαιρετικά αποτελέσματα, με μέτρια αποτελέσματα για την Bh-GMM. Η SKL έχει το μειονέκτημα ότι είναι εξαιρετικά αργή, ειδικά με όσο μεγαλύτερες εικόνες έχουμε να κάνουμε.

6.3.2 Βάση Β.

Η δεύτερη βάση είναι αρκετά μεγαλύτερη από την πρώτη, και ένας υπολογισμός όλων των αποστάσεων μεταξύ αποστάσεων θα ήταν εξαιρετικά χρονοβόρος. Γι' αυτό εδώ διαλέξαμε να κατασκευάσουμε καμπύλες PR.

	SKL	Bh-GMM	L_2
Χρόνος (sec)	33,161	154	674

Πίνακας 6.3. Χρόνοι υπολογισμών αποτελεσμάτων του πίνακα 6.1.

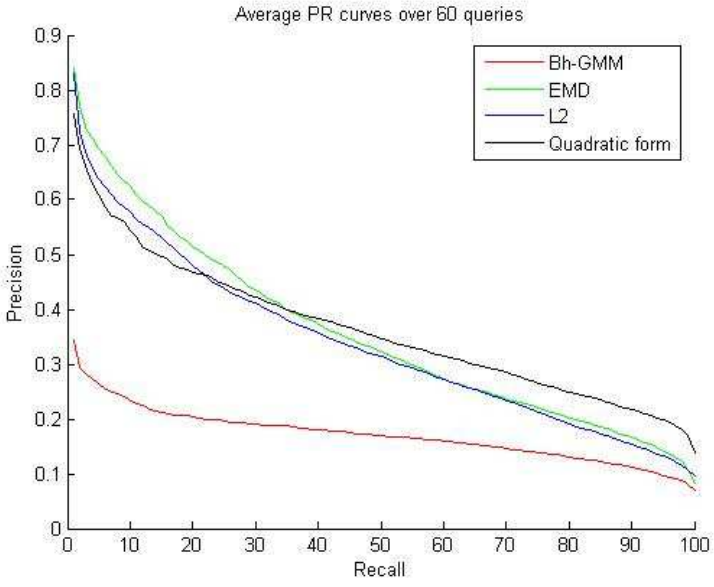
Για όλες τις εικόνες εκπαιδεύσαμε μοντέλα GMM, επιλέγοντας αριθμό πυρήνων σύμφωνα με το κριτήριο MDL (με αυξημένο βάρος στον όρο ποινής, βλέπε κεφάλαιο 5, και [22, 4]), και με διάνυσμα χαρακτηριστικών[6] smoothed CIELAB + polarity, anisotropy, contrast + x, y (χρώμα, υφή, τοπολογία, *8x1).* Χρησιμοποιήσαμε τέσσερις διαφορετικές συναρτήσεις απόστασης: Bh-GMM, EMD, L_2, Quadratic form (όπως στο Blobworld, θεωρώντας όλη την εικόνα ένα τμήμα, και βάρη για γειτονικούς

[6] Αυτά τα χαρακτηριστικά έχουν κανονικοποιηθεί ώστε το καθένα να έχει μηδενικό μέσο και μοναδιαία κανονική απόκλιση. Το μέσο και η απόκλιση υπολογίστηκαν ως προς όλες τις εικόνες της βάσης.

κάδους ίσα με 0.5). Την Symmetric KL δεν τη συμπεριλαμβάνουμε επειδή είναι χρονοβόρα –ιδιαίτερα τώρα που θέλουμε να δουλέψουμε με μεγαλύτερη βάση- όπως διαπιστώσαμε προηγουμένως (βλ. και πίνακα 6.4).

Για κάθε επιλογή συνάρτησης απόστασης, κατασκευάσαμε μια καμπύλη PR. Καθεμία είναι στην πραγματικότητα ο μέσος όρος 60 καμπύλων για ισάριθμες τυχαίες ερώτησεις στη βάση.

Η καμπύλες, μαζί με συγκριτικά στοιχεία για το εμβαδό υπό των καμπυλών, φαίνονται αντίστοιχα στο σχήμα 6.2 και στον πίνακα 6.4. Η Quadratic form απόσταση βέβαια δεν χρησιμοποιείται ποτέ για όλη την εικόνα στο Blobworld’ όμως μπορούμε να βγάλουμε κάποια συμπεράσματα από τις καμπύλες. Παρά την απλοϊκότητα της, η Quadratic form δίνει γενικά καλά αποτελέσματα, και οι αποστάσεις για GMM (πέραν της Bh-GMM) φαίνεται ότι είναι ελαφρώς καλύτερες μόνο στα μικρότερα επίπεδα Recall, 0-20%. Η απόσταση Bh-GMM φαίνεται να είναι μάλλον αποτυχία.



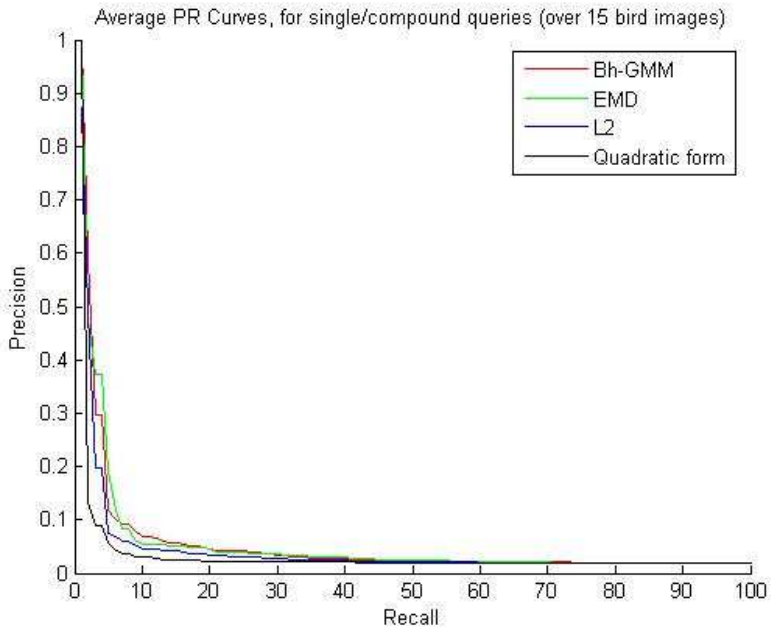
Σχήμα 6.2. Σύγκριση καμπύλων PR για διάφορες συναρτήσεις απόστασης.

Προκειμένου να αξιολογήσουμε την αποτελεσματικότητα ερώτησης για τμήμα ή τμήματα της εικόνας, ετοιμάσαμε και ένα δεύτερο σετ πειραμάτων. Για τρεις κατηγορίες εικόνων, ονομαστικά τις birds, cars και cows, κάνουμε ερώτηση με τις

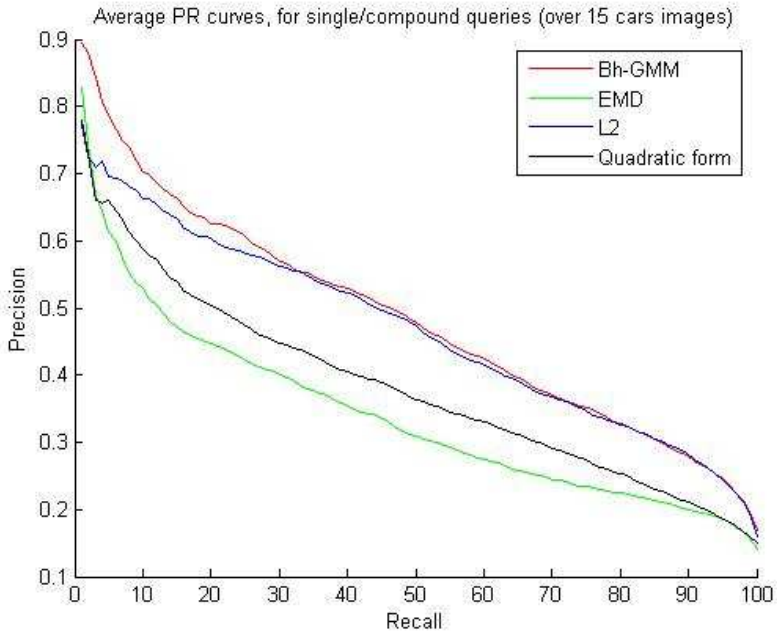
ίδιες αποστάσεις πάλι. Ρωτάμε για το τμήμα, ή αν έχει υπερκατατμηθεί, τα τμήματα που απαρτίζουν το αντικείμενο στο προσκήνιο (foreground). Αντίστοιχα για τις ομάδες εικόνων, αυτά είναι ένα πουλί, ένα αυτοκίνητο, και μια αγελάδα. Από κάθε ομάδα εικόνων διαλέγουμε 15 εικόνες για ερώτηση, και από αυτές φτιάχνουμε αντίστοιχα μέσες καμπύλες PR. Τις καμπύλες μπορούμε να δούμε στο σχήμα 6.3.

	0%-20%	20%-40%	40%-60%	60%-80%	80%-100%	Σύνολο
Bh-GMM	0.048	0.038	0.039	0.029	0.021	0.170
EMD	0.126	0.087	0.064	0.047	0.031	0.355
L_2	0.117	0.081	0.062	0.046	0.029	0.337
Quadratic form	0.111	0.084	0.069	0.056	0.042	0.362

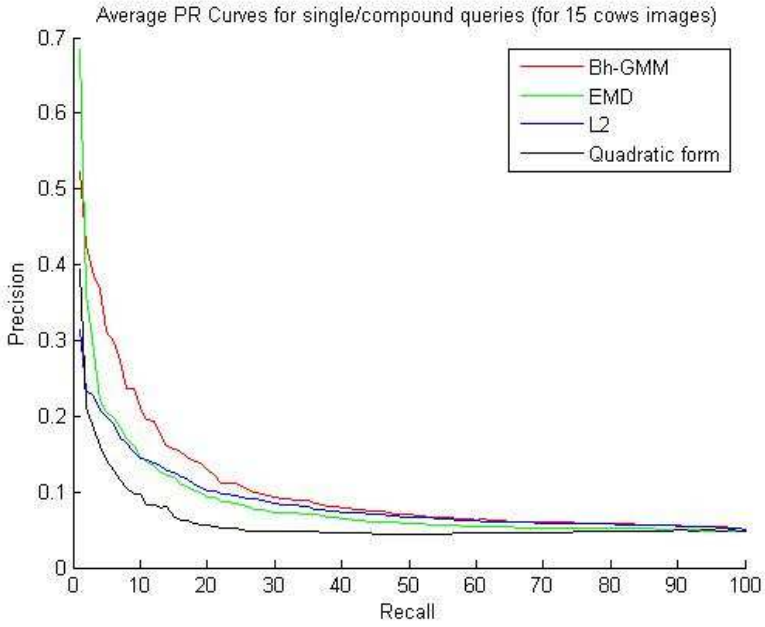
Πίνακας 6.4. Εμβαδά καμπύλων Precision – Recall, για CIELAB (smoothed) + polarity-anisotropy-contrast + x-y χαρακτηριστικά.



(α)


Average PR curves, for single/compound queries (over 15 cars images)
Bh-GMM
EMD
L2
Quadratic form
Precision
Recall
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
10
20
30
40
50
60
70
80
90
100

(β)


Average PR Curves for single/compound queries (for 15 cows images)
Bh-GMM
EMD
L2
Quadratic form
Precision
Recall
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
10
20
30
40
50
60
70
80
90
100

(γ)

Σχήμα 6.3. Σύγκριση καμπύλων PR, για ερώτηση ολόκληρης εικόνας και εικόνας κατά τμήματα. Οι ερωτήσεις έγιναν για εικόνες από (α) Πουλιά (β) Αυτοκίνητα (γ) Αγελάδες.

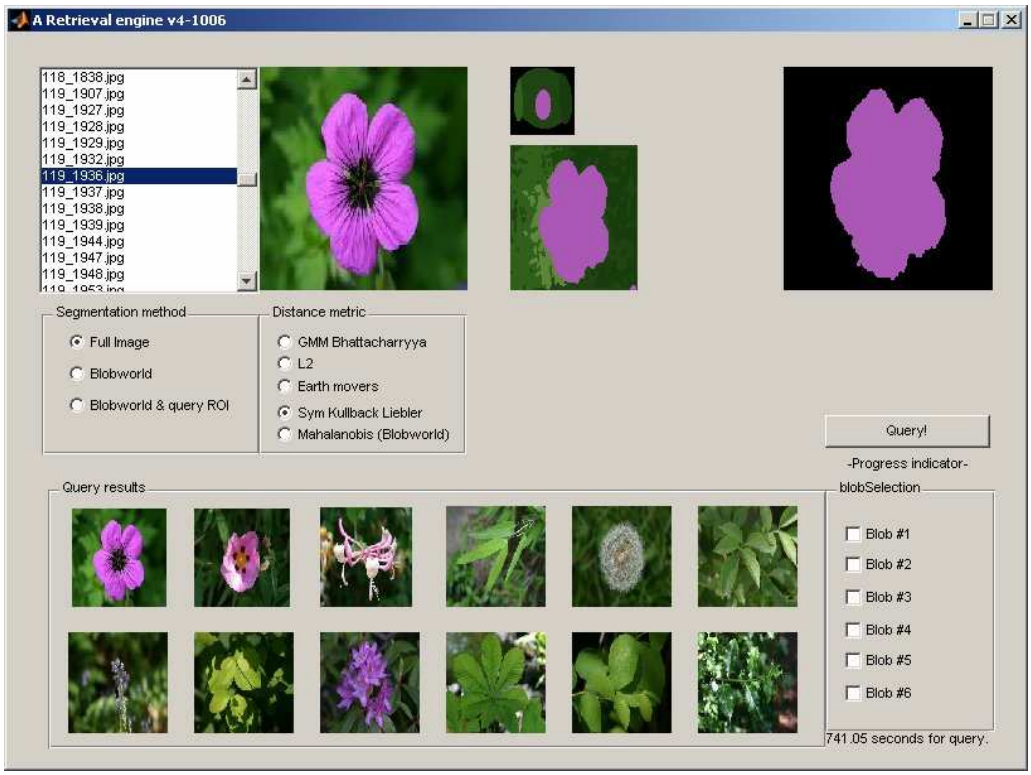
Να πούμε ότι αφού έχουμε να κάνουμε με αποστάσεις κανονικών κατανομών εδώ (όχι μίξεων), η Bh-GMM είναι ουσιαστικά η απόσταση Bhattacharryya, και η EMD είναι η απόσταση Frechet. Γενικά αυτό που βλέπουμε και στα τρία σετ καμπύλων είναι ότι η Bhattacharryya, και λιγότερο η L_2, αποδίδουν καλύτερα της Quadratic form.

6.4. Interface για ανάκτηση εικόνας

Στα πλαίσια της εργασίας αναπτύχθηκε interface σε MATLAB για ανάκτηση εικόνας. Μπορούμε να χρησιμοποιήσουμε τις περισσότερες από τις τεχνικές που παρουσιάστηκαν στα προηγούμενα κεφάλαια.

Συγκεκριμένα, μπορούμε να κάνουμε ερώτηση για όλη την εικόνα, για ένα ή περισσότερα τμήματα, ή περιοχή ενδιαφέροντος (Region of Interest). Σε συνδυασμό με αυτά μπορούμε να χρησιμοποιήσουμε οποιαδήποτε από τις εξής συναρτήσεις απόστασης: Bhattacharyya-GMM, L2, Earth mover's distance, Symmetric Kullback-Liebler, Quadratic form (εμφανίζεται σαν «Mahalanobis») distance.

Ο μηχανισμός είναι πιστεύουμε πολύ απλός. Κάνοντας κλικ στο όνομα μιας εικόνας στον browser πάνω αριστερά, το πρώτο που εμφανίζεται είναι η ίδια η εικόνα, οι προβολές των γκαουσιανών πυρήνων σε x-y στο μικρότερο παράθυρο επάνω, και κάτω η κατάτμηση της εικόνας. Δεξιά φαίνονται ποια τμήματα επιλέγουμε σε περίπτωση που θέλουμε να ρωτήσουμε για τμήμα ή τμήματα της εικόνας. Αφού επιλεχθούν η μέθοδος, η συνάρτηση απόστασης που θέλουμε και τα τμήματα για τα οποία ρωτάμε (αν ρωτάμε για τμήματα), κάνουμε κλικ στο κουμπί 'Query' και εμφανίζονται τα αποτελέσματα.



Σχήμα 6.4. Φωτογραφία του interface που αναπτύχθηκε για την παρούσα εργασία.

ΑΝΑΦΟΡΕΣ

[1] Swain M.J., and Ballard D.H. "**Color indexing**". *International Journal of Computer Vision*, 1991, Vol.7(1), 1132.

[2] Fukunaga K., "**Statistical Pattern Recognition**", *Academic press*, London, 1972.

[3] Bishop C.M., "**Neural Networks for Pattern Recognition"**, *Oxford University press*, 1995

[4] Carson C., Belongie S., Greenspan H., Malik J. "**Blobworld: Image segmentation using Expectation-Maximization and its application to image querying**", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Volume 24(8), pp. 1026 – 1038, August 2002.

[5] Hafner J., Sawhney H., Equitz W., Flickner M., and Niblack W., "**Efficient Color Histogram Indexing for Quadratic Form Distance Functions**", *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 17, no. 7, pp. 729-736, July 1995.

[6] Flickner M.; Sawhney H.; Niblack W.; Ashley J.; Qian Huang; Dom B.; Gorkani M.; Hafner J.; Lee D.; Petkovic D.; Steele D.; Yanker P.: "**Query by image and video content: the QBIC system**" *IEEE transactions in Comput*. Volume 28, Issue 9, Sept. 1995 Page(s):23 – 32

[7] Han J.; Ma K.K.; "**Fuzzy color Histogram and its use in color image retrieval**", *IEEE Transactions on Image Processing*, Vol 11(8), pp 944-952.

[8] Greenspan H., Dvir G. and Rubner Y., **"Context-dependent segmentation and matching in image databases"**, *Computer Vision and Image Understanding*, Vol 93(1) January 2004, pp. 86-109.

[9] Greenspan H., Goldberger J., Ridel L., "**A continuous probabilistic framework for image matching**", *Comput. Vis. Image Understanding 84*, December 2001, pp. 384-406.

[10] McLachlan G.J., Krishnan T., "**The EM algorithm and its extensions**", *John Wiley & sons*, New York, 1996.

[11] McLachlan G.J., Peel D., "**Finite Mixture models**", *John Wiley & sons*, New York, 2000.

[12] Blekas K., Likas A., Galatsanos N.P., Lagaris I.E., "**A spatially constrained mixture model for image segmentation",** *IEEE transactions on Neural Networks*, Volume 16(2), pp. 494-498, March 2005.

[13] Feller W. , "**An introduction to probability theory and its applications**", *John Wiley & sons*, Volume I New York 1968, Volume II New York 1971.

[14] Forsyth D.A., Ponce J.: "**Computer Vision: A modern approach**", Prentice Hall, New Jersey 2003.

[15] Nocedal J., Wright S.J.: "**Numerical Optimization**", Springer-Verlag, New York 1999.

[16] Lange K.: "**Optimization**", Springer-Verlag, New York 2004.

[17] Kullback S., and Leibler R.A., "**On information and sufficiency**", *Annals of Mathematical Statistics 22*: 79-86, 1951.

[18] Del Bimbo A.: "**Visual information Retrieval**", *Morgan Kaufmann publishers*, San Francisco, 1999

[19] Brodatz P.: "**Textures: A Photographic Album for Artists and Designers**." New York: Dover, 1966.

[20] Wang Liwei, Zhang Y., Feng J., "**On the Euclidean Distance of Images**", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27(8), pp.1334-1339, August 2005.

[21] Sfikas G., Constantinopoulos C., Likas A., Galatsanos N.P., "**An Analytic Distance Metric for Gaussian Mixture Models with Application in Image Retrieval**", *Proc. 15th International Conference on Artificial Neural Networks*, vol. 2, pp. 835-840, 2005.

[22] Website URL: "http://elib.cs.berkeley.edu/src/blobworld/"

ΠΑΡΑΡΤΗΜΑ Α – ΜΕΡΙΚΑ ΑΠΟΤΕΛΕΣΜΑΤΑ ΑΠΟ ΛΟΓΙΣΜΟ ΠΙΝΑΚΩΝ

Έστω Α πραγματικός τετραγωνικός πίνακας, dxd . Τότε ορίζονται,

$tr(A)$: Ίχνος του πίνακα, ήτοι άθροισμα των στοιχείων της διαγωνίου. Είναι ίσο με το άθροισμα των ιδιοτιμών του A.

$diag(A)$: Πίνακας με όλα τα στοιχεία ίσα με 0, εκτός από αυτά της διαγωνίου. Αυτά συμπίπτουν με τα στοιχεία της διαγωνίου του A.

Και ισχύουν, για αυθαίρετο διάνυσμα $x, dx1$, και για πίνακες B, C επίσης πραγματικούς τετραγωνικούς dxd :

$$x^T Ax = tr\left(Axx^T\right) \quad \text{(A.1)}$$

$\frac{\partial A}{\partial C} = \frac{\partial A}{\partial B}\frac{\partial B}{\partial C}$ όταν A συνάρτηση του B, B συνάρτηση του C. (A.2)

$$\frac{\partial AB}{\partial C} = \frac{\partial A}{\partial C}B + A\frac{\partial B}{\partial C} \quad \text{(A.3)}$$

$$\frac{\partial Ax}{\partial x} = A \quad \text{(A.4)}$$

$$\frac{\partial x^T Ax}{\partial x} = (A + A^T)x \quad \text{(A.5)}$$

$$\frac{\partial tr(AB)}{\partial A} = B \quad \text{(A.6)}$$

$$\frac{\partial \log|A|}{\partial A} = A^{-T} \quad \text{(A.7)}$$

$\frac{\partial tr(AB)}{\partial A} = B + B^T - diag(B)$, όταν ο A περιορίζεται συμμετρικός. (A.8)

$$\frac{\partial \log|A|}{\partial A} = 2A^{-1} - diag(A^{-1})$$, όταν ο A περιορίζεται συμμετρικός. (Α.9)

Σημειώστε ότι οι (Α.8),(Α.9) *δεν* είναι ειδικές περιπτώσεις των (Α.6),(Α.7).

ΠΑΡΑΡΤΗΜΑ Β – ΥΠΟΛΟΓΙΣΜΟΣ ΤΗΣ ΣΥΝΑΡΤΗΣΗΣ ΑΠΟΣΤΑΣΗΣ L_2

Θέλουμε να υπολογίσουμε την ποσότητα (τα όρια της ολοκλήρωσης είναι όλο το πεδίο ορισμού του *x*):

$$d_{L2}\left(p, \tilde{p}\right) = -\ln\left(\frac{2\int p(x)\,\tilde{p}(x)dx}{\int p(x)^2 + \tilde{p}(x)^2\,dx}\right) \quad \text{(B.1)}$$

όταν οι πυκνότητες $p, \tilde{p}$ είναι μίξεις κανονικών κατανομών. Δηλαδή

$$p(x) = \sum_{i=1}^{K} \pi_i \left(2\pi\right)^{-d/2} \left|\Sigma_i\right|^{-1/2} \exp\left\{-\frac{1}{2}(x-\mu_i)^{\mathrm{T}}\Sigma_i^{-1}(x-\mu_i)\right\}$$

$$\tilde{p}(x) = \sum_{i=1}^{\tilde{K}} \tilde{\pi}_i \left(2\pi\right)^{-d/2} \left|\tilde{\Sigma}_i\right|^{-1/2} \exp\left\{-\frac{1}{2}(x-\tilde{\mu}_i)^{\mathrm{T}}\tilde{\Sigma}_i^{-1}(x-\tilde{\mu}_i)\right\}$$

όπου $K, \tilde{K}$ σταθεροί θετικοί ακέραιοι, *x* τυχαία μεταβλητή στο $\mathbb{R}^d$ (διάνυσμα), π_i μη-αρνητικά και αθροίζονται στη μονάδα, μ_i πραγματικά διανύσματα *dx1* (μέσα), Σ_i πραγματικοί συμμετρικοί πίνακες *dxd* (μήτρες συμμεταβλητότητας). Παρομοίως για τα π_i, μ_i, Σ_i.

Πρώτα πρέπει να υπολογίσουμε την ποσότητα

$$\int p(x)\,\tilde{p}(x)dx \quad \text{(B.2)}$$

που περιλαμβάνει τον όρο (παραλείψαμε τους δείκτες για απλότητα)

$$\exp\left\{-\frac{1}{2}(x-\mu)^{\mathrm{T}}\Sigma^{-1}(x-\mu) - \frac{1}{2}(x-\tilde{\mu})^{\mathrm{T}}\tilde{\Sigma}^{-1}(x-\tilde{\mu})\right\} \quad \text{(B.3)}$$

Αν εξισώσουμε τον εκθέτη του (B.3) με κάποια ποσότητα $-\frac{1}{2}(x-m)^T S^{-1}(x-m) - \frac{1}{2}k$, παίρνουμε:

$$x^T S^{-1} x = x^T \Sigma^{-1} x + x^T \tilde{\Sigma}^{-1} x \Rightarrow$$

$$S^{-1} = \Sigma^{-1} + \tilde{\Sigma}^{-1} \quad \text{(B.4)}$$

(εξισώνοντας τους παράγοντες του x^2)

$$2x^T\Sigma^{-1}\mu + 2x^T\,\tilde{\Sigma}^{-1}\,\tilde{\mu} = 2x^T S^{-1} m \Rightarrow$$

$$m = S(\Sigma^{-1}\mu + \tilde{\Sigma}^{-1}\,\tilde{\mu}) \Rightarrow \text{ (χρησιμοποιούμε B.4)}$$

$$m = (\Sigma^{-1} + \tilde{\Sigma}^{-1})^{-1}(\Sigma^{-1}\mu + \tilde{\Sigma}^{-1}\,\tilde{\mu}) \text{ (B.5)}$$

(εξισώνοντας τους παράγοντες του x)

$$\mu^{\mathrm{T}}\Sigma^{-1}\mu + \tilde{\mu}^{\mathrm{T}}\,\tilde{\Sigma}^{-1}\,\tilde{\mu} = m^T S^{-1} m + k \Rightarrow \text{ (χρησιμοποιούμε B.4,B.5)}$$

$$k = \mu^{\mathrm{T}}\Sigma^{-1}(\mu - m) + \tilde{\mu}^{\mathrm{T}}\,\tilde{\Sigma}^{-1}(\tilde{\mu} - m) \text{ (B.6)}$$

(εξισώνοντας τους σταθερούς παράγοντες)

Οπότε η (B.2) γίνεται

$$\int p(x)\,\tilde{p}(x)dx = (2\pi)^{-d}\sum_{i=1}^{N}\sum_{j=1}^{M}\pi_i\,\tilde{\pi}_j\left[\xi(p_i, \tilde{p}_j)\right]^{-1/2} \quad \text{(B.7)}$$

όπου $p_i, \tilde{p}_j$ κανονικοί πυρήνες των μίξεων, και

$$\xi(g,\tilde{g})^{-1/2} \triangleq \left|\Sigma\tilde{\Sigma}\right|^{-1/2} \exp\{-\frac{1}{2}k\}(2\pi)^{d/2}\int \exp\{-\frac{1}{2}(x-m)^T S^{-1}(x-m)\} \Rightarrow$$

$$\xi(g,\tilde{g})^{-1/2} = \left|\Sigma\tilde{\Sigma}\right|^{-1/2} |S|^{1/2} \exp\{-\frac{1}{2}k\} \Rightarrow$$

$$\xi(g,\tilde{g})^{-1/2} = \left|\Sigma^{-1} + \tilde{\Sigma}^{-1}\right|^{-1/2} \left|\Sigma\tilde{\Sigma}\right|^{-1/2} \exp\{-\frac{1}{2}k\} \Rightarrow$$

$$\xi(g,\tilde{g})^{-1/2} = \left|I + \Sigma\tilde{\Sigma}^{-1}\right|^{-1/2} \left|\tilde{\Sigma}\right|^{-1/2} \exp\{-\frac{1}{2}k\} \Rightarrow$$

$$\xi(g,\tilde{g})^{-1/2} = \left|\tilde{\Sigma} + \Sigma\right|^{-1/2} \exp\{-\frac{1}{2}k\} \Rightarrow$$

$$\xi(g,\tilde{g}) = \left(\left|\tilde{\Sigma} + \Sigma\right| \mathrm{e}^k\right)^{-1/2} \quad \text{(B.8)}$$

με k να δίνεται από (B.6), και $g, \tilde{g}$ κανονικές κατανομές (πυρήνες) με μέσα και συμμεταβλητότητα μ, Σ και $\tilde{\mu}, \tilde{\Sigma}$ αντίστοιχα.

Χρησιμοποιώντας (B.1),(B.7),(B.8), καταλήγουμε σε

$$d_{L2}\left(p,\tilde{p}\right)=-\ln\left(\frac{2\sum_{i=1}^{N}\sum_{j=1}^{M}\pi_i\,\tilde{\pi}_j\,\xi\left(p_i,\tilde{p}_j\right)^{-1/2}}{\sum_{i=1}^{N}\sum_{j=1}^{N}\pi_i\pi_j\xi\left(p_i,p_j\right)^{-1/2}+\sum_{i=1}^{M}\sum_{j=1}^{M}\tilde{\pi}_i\,\tilde{\pi}_j\,\xi\left(\tilde{p}_i,\tilde{p}_j\right)^{-1/2}}\right) \quad \text{(B.9)}$$

όπου

$$\xi(g,\tilde{g})=\left|\Sigma+\tilde{\Sigma}\right|\exp\left\{\mu^{\mathrm{T}}\Sigma^{-1}(\mu-m)+\tilde{\mu}^{\mathrm{T}}\,\tilde{\Sigma}^{-1}(\tilde{\mu}-m)\right\}$$

$$\text{με } m^{T}=\left(\mu^{\mathrm{T}}\Sigma^{-1}+\tilde{\mu}^{\mathrm{T}}\,\tilde{\Sigma}^{-1}\right)\left(\Sigma^{-1}+\tilde{\Sigma}^{-1}\right)^{-1}$$

με $g,\tilde{g}$ κανονικές κατανομές με μέσα και συμμεταβλητότητα μ,Σ και $\tilde{\mu},\tilde{\Sigma}$.

ΠΑΡΑΡΤΗΜΑ Γ – ΧΡΗΣΙΜΟΠΟΙΟΥΜΕΝΕΣ ΒΑΣΕΙΣ ΕΙΚΟΝΩΝ

Για να κάνουμε πειράματα αξιολόγησης των μεθόδων που παρουσιάζονται σε αυτή την εργασία, χρησιμοποιήσαμε δύο βάσεις εικόνων. Για ευκολία θα αναφερόμαστε σε αυτές σαν «βάση Α» και «βάση Β». Όλες οι εικόνες είναι έγχρωμες (24 bits per pixel), και αποθηκευμένες σε μορφή jpeg.

α/α	Όνομα	Περιγραφή	Πλήθος εικόνων
1	Cherries	Κερασιές	~40
2	Arborgreens	Φυτά και βλάστηση γενικά	~40
3	Cannonbeach	Παραθαλάσσιο χωριό, συννεφ. ουρανός	~40
4	Campus in fall	Μια πανεπιστημιούπολη, φθινοπωρινό τοπίο.	~40
5	Football	Αγώνας ράγκμπυ	~40

Πίνακας Γ.1 . Περιγραφές εικόνων στην βάση Α.

α/α	Όνομα	Περιγραφή	Πλήθος εικόνων
1	Aeroplanes	Αεροπλάνα	59
2	Benches_chairs	Καρέκλες και παγκάκια	69
3	Bicycles	Ποδήλατα	273
4	Birds	Πτηνά	73
5	Buildings	Κτίρια	147
6	Cars	Αυτοκίνητα	496
7	Chimneys	Καμινάδες	266
8	Clouds	Σύννεφα	430
9	Countryside	Εικόνες από εξοχή	185

10	Cows	Αγελάδες σε λιβάδι	183
11	Doors	Διαφόρων ειδών πόρτες	167
12	Flowers	Λουλούδια	167
13	Forks	Πηρούνια	37
14	Knives	Μαχαίρια	66
15	Leaves	Φύλλα	119
16	Miscellaneous	Διάφορες εικόνες από την καθημερινότητα	189
17	Office	Γραφεία κάποιας επιχείρισης	69
18	Sheep	Πρόβατα σε λιβάδι	191
19	Signs	Πινακίδες	166
20	Spoons	Κουτάλια	75
21	Trees	Δέντρα	218
22	Urban	Δρόμοι σε κάποια κωμόπολη	37
23	Windows	Διαφόρων ειδών παράθυρα	653

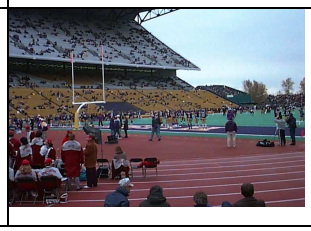
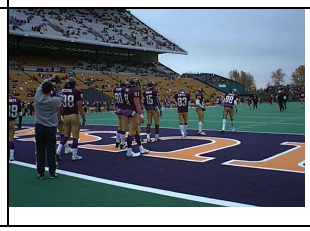
Πίνακας Γ.2. Περιγραφές εικόνων στην βάση Β.

Στη βάση Α έχουμε συνολικά 200 εικόνες, χωρισμένες σε 5 κατηγορίες με 40 περίπου εικόνες στην καθεμία. Κάθε εικόνα είναι διαστάσεων 700x500.

Μπορούμε να δούμε μια περιγραφή της κάθε κατηγορίας εικόνων στην βάση Α, στον πίνακα Γ.1. Στον πίνακα Γ.3 βλέπουμε χαρακτηριστικά δείγματα από κάθε κατηγορία. Στη βάση Β έχουμε συνολικά 4.335 εικόνες, χωρισμένες σε 23 κατηγορίες. Κάθε εικόνα είναι διαστάσεων 192x128 ή 128x192.

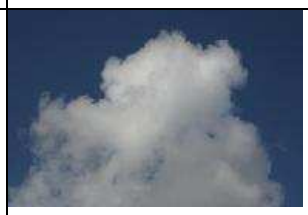
Περιγραφές των εικόνων και κατηγοριών στην βάση Β, μαζί με χαρακτηριστικά δείγματα εικόνων από κάθε κατηγορία μπορούμε να δούμε στους πίνακες Γ.2 και Γ.4 αντίστοιχα.

α	Ονομα	Δείγματα		
1	Cherries			

2	Arborgreens			
3	Cannonbeach			
4	Campus in fall			
5	Football			

Πίνακας Γ.3. Δείγματα από την βάση Α.

α	Ονομα	Δείγματα		
1	Aeroplanes			

2	Benches_chairs			
3	Bicycles			
4	Birds			
5	Buildings			
6	Cars			
7	Chimneys			
8	Clouds			
9	Countryside			

10	Cows			
11	Doors			
12	Flowers			
13	Forks			
14	Knives			
15	Leaves			
16	Miscellaneous			
17	Office			

18	Sheep			
19	Signs			
20	Spoons			
21	Trees			
22	Urban			
23	Windows			

Πίνακας Γ.4. Δείγματα από την βάση Β.

ΔΗΜΟΣΙΕΥΣΕΙΣ ΣΥΓΓΡΑΦΕΑ

Sfikas G., Constantinopoulos C., Likas A., Galatsanos N.P., "**An Analytic Distance Metric for Gaussian Mixture Models with Application in Image Retrieval**", *Proc. 15th International Conference on Artificial Neural Networks,* vol. 2, pp. 835-840, 2005.

ΣΥΝΤΟΜΟ ΒΙΟΓΡΑΦΙΚΟ

Ο Γεώργιος Σφήκας του Ανδρέα-Φανουρίου και της Ευρυδίκης γεννήθηκε στον Χολαργό Αττικής το 1982. Αποφοίτησε από το 51^{ο} Λύκειο Κολωνού Αθηνών το 1999. Σπούδασε στο Τμήμα Πληροφορικής του Πανεπιστημίου Ιωαννίνων την περίοδο 1999-2004 , από όπου αποφοίτησε και έλαβε πτυχίο με βαθμό 7.1 «λίαν καλώς». Στο διάστημα 2004 με 2006 παρακολούθησε το μεταπτυχιακό πρόγραμμα του ιδίου τμήματος.