ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

ΤΟΜΕΑΣ ΤΕΧΝΟΛΟΓΙΑΣ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΥΠΟΛΟΓΙΣΤΩΝ

Ανάκτηση Εικόνων Μόδας βασισμένη στην χρήση Οπτικών Περιγραφέων

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

της

ΚΟΥΛΟΥΡΙΑ ΜΑΡΙΑΣ

Επιβλέπων : Στέφανος Κόλλιας
Καθηγητής Ε.Μ.Π.

Αθήνα, Μάρτιος 2016

ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ
ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ
ΤΟΜΕΑΣ ΤΕΧΝΟΛΟΓΙΑΣ ΠΛΗΡΟΦΟΡΙΚΗΣ
ΚΑΙ ΥΠΟΛΟΓΙΣΤΩΝ

Ανάκτηση Εικόνων Μόδας βασισμένη στην χρήση Οπτικών Περιγραφέων

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

της

ΚΟΥΛΟΥΡΙΑ ΜΑΡΙΑΣ

Επιβλέπων : Στέφανος Κόλλιας
Καθηγητής Ε.Μ.Π.

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 26η Φεβρουαρίου 2016.

.....	.....	.....
Σ. Κόλλιας	Γ. Στάμου	Α.-Γ. Σταφυλοπάτης
Καθηγητής Ε.Μ.Π.	Επ. Καθηγητής Ε.Μ.Π.	Καθηγητής Ε.Μ.Π.

Αθήνα, Μάρτιος 2016

.....

ΚΟΥΛΟΥΡΙΑ ΜΑΡΙΑ

Διπλωματούχος Ηλεκτρολόγος Μηχανικός και Μηχανικός Υπολογιστών Ε.Μ.Π.

Ευχαριστίες

Η παρούσα διπλωματική εργασία εκπονήθηκε στη σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών στον τομέα Τεχνολογίας Πληροφορικής και Υπολογιστών.

Θα ήθελα να ευχαριστήσω τον επιβλέποντα καθηγητή μου κ.Στέφανο Κόλλια που μου εμπιστεύθηκε την εργασία αυτή και μου έδωσε την ευκαιρία να ασχοληθώ με ένα θέμα που με ενδιαφέρει ιδιαίτερα.

Επίσης, θα ήθελα να ευχαριστήσω θερμά τον Υποψήφιο Διδάκτορα κ.Κωνσταντίνο Ραπαντζίκο για την πολύτιμη καθοδήγηση και τις χρήσιμες υποδείξεις που μου παρείχε σε όλη την διάρκεια εκπόνησης της διπλωματικής, καθώς και για τον χρόνο που διέθεσε για τον σκοπό αυτό.

Τέλος, θα ήθελα να ευχαριστήσω την οικογένεια και τους φίλους μου που με στήριξαν όλα τα χρόνια των σπουδών μου, και ιδιαίτερα την μητέρα μου και την φίλη μου Εύη.

Κουλούρια Μαρία,
Μάρτιος 2016

Περίληψη

Ο σκοπός της παρούσας διπλωματικής εργασίας είναι η ανάπτυξη μεθοδολογίας για την περιγραφή και την ανάκτηση εικόνων παρόμοιας ενδυμασίας με χρήση οπτικών περιγραφέων. Για το λόγο αυτό έγινε χρήση ενός συνόλου εικόνων από επιδείξεις μόδας. Τα αποτελέσματα που προέκυψαν κατέδειξαν την σημασία της χρωματικής πληροφορίας έναντι άλλων χαρακτηριστικών.

Συγκεκριμένα, έγινε χρήση του περιγραφέα SIFT σε συνδυασμό με χρωματικά ιστογράμματα για την εξαγωγή χρωματικής πληροφορίας σε σημεία-κλειδιά των εικόνων, καθώς και χρωματικά ιστογράμματα, χρωματικές ροπές και χρωματικές ετικέτες για την καταγραφή χρωματικής πληροφορίας ξεχωριστά. Επίσης, για την απομόνωση μόνο του τμήματος της εικόνας που έχει για εμάς σημασία, δηλαδή την ενδυμασία, χρησιμοποιήθηκαν εκτίμηση πόζας αλλά και μάσκες στην περιοχή ενδιαφέροντος.

Στην παρούσα μελέτη αποδείχτηκε μέσω πειραμάτων ότι η περιγραφή των εικόνων με βάση το χρώμα μπορεί να οδηγήσει σε εύρωστη ανάκτηση παρόμοιων ρούχων όταν αυτά απεικονίζονται φορεμένα σε εικόνες, και άρα δεν έχουν σταθερή μορφή, και να βοηθήσει στην ανάπτυξη αυτόματων συστημάτων που παρέχουν προτάσεις για παρόμοιες στυλιστικές επιλογές.

Λέξεις Κλειδιά: περιγραφή χαρακτηριστικών εικόνας, ανάκτηση εικόνων, οπτικοί περιγραφείς, ανάκτηση βασισμένη στο χρώμα, SIFT, χρωματικό SIFT, χρωματικά ιστογράμματα, χρωματικές ετικέτες, χρώματικες ροπές, εκτίμηση πόζας, ROI μάσκες

Abstract

The purpose of this thesis is to develop a methodology for describing and retrieving images with similar attire based on visual descriptors. For this reason, we made use of a set of images from fashion shows. The results obtained showed the importance of color information over other image characteristics.

Specifically, we used the SIFT descriptor combined with color histograms to extract color information at key points of images, and color histograms, color moments and color names for recording color information separately. Furthermore, for the isolation of the parts within an image that have significance to us, which is the attire, we used pose estimation and masks at the area of interest.

In the present study we proved through experiments that a description of images based on color can lead to a more robust retrieval of similar clothes when those appear worn in real world images and therefore do not have a fixed shape, and assist in further development of automated systems that provide suggestions for similar stylistic choices.

Keywords: Image feature description, image retrieval, visual descriptors, color based retrieval, SIFT, color SIFT, color histograms, color names, color moments, pose estimation, ROI masks

Πίνακας περιεχομένων

Ευρετήριο Εικόνων

Ευρετήριο Πινάκων

1 Εισαγωγή

1.1 Περιγραφή και Ανάκτηση Εικόνων Μόδας

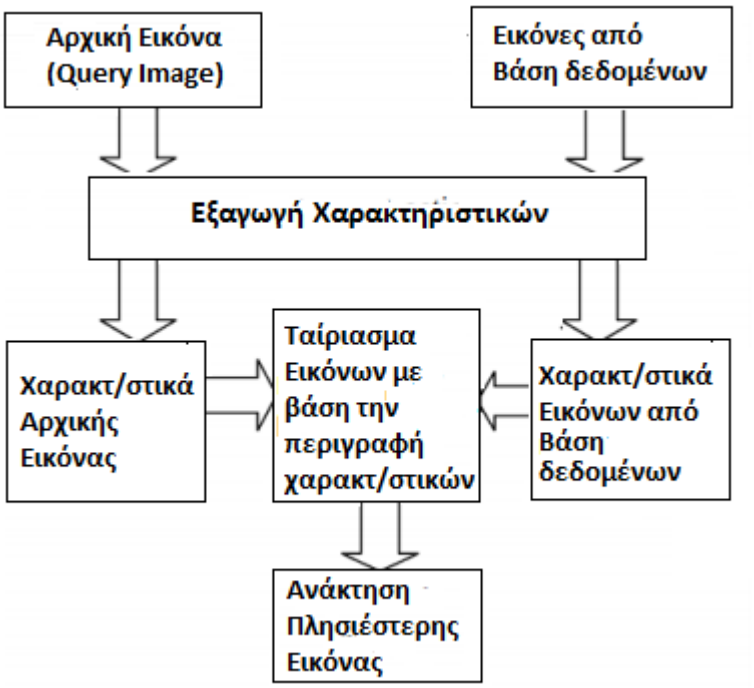
Το ενδιαφέρον που συγκεντρώνει ο χώρος της μόδας στο διαδίκτυο είναι ταχέως εξελισσόμενο τα τελευταία χρόνια και επικεντρώνεται κυρίως γύρω από ιστοτόπους σχετικούς με στυλιστικές εμφανίσεις διασημοτήτων ή ακόμα και τις στυλιστικές επιλογές περαστικών στο δρόμο. Απώτερος σκοπός τους είναι να παρουσιάσουν παρόμοιες ενδυματολογικά εικόνες από τις οποίες μπορεί κανείς να εμπνευστεί και να δημιουργήσει το δικό του στυλ ή ακόμα και να προτείνουν συγκεκριμένα προϊόντα ένδυσης και από που μπορεί κανείς να τα προμηθευτεί. Οι εικόνες που χρησιμοποιούνται για αυτό το σκοπό συνήθως συνοδεύονται από μεταδεδομένα περιγραφής τους, μια διαδικασία αρκετά χρονοβόρα. Ιδανικά θέλουμε η πληροφορία αυτή να προέρχεται αυτόματα από την ανάλυση της εικόνας και ανάλογα με την οπτική περιγραφή της να γίνεται η ανάκτηση παρόμοιων εικόνων από μια βάση δεδομένων. Για το λόγο αυτό, είναι σημαντική η κατάλληλη περιγραφή των σημείων της εικόνας που φαίνεται να υπάρχει ενδυμασία στην οποία θα βασιστεί η μετέπειτα ανάκτηση.

1.2 Αντικείμενο διπλωματικής

Η διαδικασία που περιγράψαμε προηγουμένως, δηλαδή η περιγραφή και ανάκτηση εικόνων μόδας, παρουσιάζεται συνοπτικά στην **Εικόνα 1.1**. Αρχικά, κάνοντας χρήση ενός συνόλου εικόνων μόδας γίνεται εξαγωγή και αποθήκευση των χαρακτηριστικών τους. Όταν επιλέγουμε μια αρχική εικόνα (Query Image) για την οποία αναζητούμε πλησιέστερη, τα χαρακτηριστικά της

συγκρίνονται με αυτά των υπόλοιπων εικόνων και ανακτάται αυτή με την οποία ταιριάζει περισσότερο.

Σκοπός της διπλωματικής είναι να βοηθήσει στην επιλογή **κατάλληλων χαρακτηριστικών** στα **σημεία ενδιαφέροντος** μιας εικόνας μόδας. Με τον όρο κατάλληλα χαρακτηριστικά, αναφερόμαστε σε εκείνα που είναι εύρωστα σε αλλαγές όπως αλλαγές φωτισμού, περιστροφών κλπ. ενώ τα σημεία ενδιαφέροντος είναι αυτά που εντοπίζονται στην περιοχή της ενδυμασίας. Επομένως, τα προβλήματα που πρέπει να αντιμετωπίσουμε είναι δύο, να απομονώσουμε την σωστή περιοχή του ρούχου και να εντοπίσουμε ποια χαρακτηριστικά ενός ρούχου μπορούν να περιγραφούν σωστά. Αυτό θα συντελέσει στην σωστή εξαγωγή χαρακτηριστικών των εικόνων μόδας. Το γεγονός ότι οι εικόνες που χρησιμοποιήσαμε προέρχονται από επιδείξεις, δηλαδή περιλαμβάνουν φορεμένα ρούχα, σημαίνει αλλαγές στην μορφή τους, καθώς και ότι επεξεργαζόμαστε εικόνες πραγματικού κόσμου και όχι εικόνες με ένα καθαρό περιβάλλον γύρω από το ρούχο.



Εικόνα 1.1 : *Περιγραφή και ανάκτηση εικόνων*

Για τον εντοπισμό κατάλληλων χαρακτηριστικών :

1. Μελετήσαμε τον οπτικό περιγραφέα SIFT
2. Μελετήσαμε τους χρωματικούς χώρους, χρωματικά ιστογράμματα, χρωματικές ροπές και χρωματικές ετικέτες με σκοπό την εξαγωγή χρωματικής πληροφορίας

3. Υλοποιήσαμε τον περιγραφέα SIFT σε συνδυασμό με χρωματική πληροφορία
4. Υλοποιήσαμε χρωματική περιγραφή ξεχωριστά

Για τον εντοπισμό των περιοχών ενδιαφέροντος:

5. Υλοποιήσαμε εκτίμηση πόζας
6. Χρησιμοποιήσαμε μάσκες στην περιοχή ενδιαφέροντος

Η αξιολόγηση των αποτελεσμάτων κατέδειξε την χρησιμότητα της χρωματικής πληροφορίας της ενδυμασίας και την σπουδαιότητα της στην ανάκτηση παρόμοιων ενδυμάτων.

1.3 Οργάνωση κειμένου

Η παρούσα διπλωματική εργασία διακρίνεται στο θεωρητικό κομμάτι, που αναλύεται στο **Κεφάλαιο 2** και στην ανάλυση του πειραματικού μέρους που παρουσιάζεται στα **Κεφάλαια 3** και **4.**

Στην **Ενότητα 2.1** του **Κεφαλαίου 2** παρουσιάζεται ο περιγραφέας SIFT ενώ στην **Ενότητα 2.2** παρέχεται το θεωρητικό υπόβαθρο για την χρωματική περιγραφή με χρωματικά ιστογράμματα, χρωματικές ετικέτες, χρωματικές ροπές. Στην **Ενότητα 2.3** γίνεται περιγραφή της μεθόδου εκτίμησης πόζας.

Στο **Κεφάλαιο 3** περιγράφεται η μεθοδολογία που ακολουθήσαμε κάνοντας χρήση του MATLAB με σκοπό την εξαγωγή σωστής πληροφορίας, ενώ στο **Κεφάλαιο 4** παρουσιάζονται αναλυτικά τα αποτελέσματα της ανάκτησης ανάλογα με το είδος περιγραφής που επιλέχθηκε (π.χ. ιστογράμματα, ετικέτες κλπ) και ανάλογα με την μέθοδο απομόνωσης του ρουχισμού (εκτίμηση πόζας ή μασκών). Στο κεφάλαιο αυτό αποδεικνύεται αναλυτικά η υπεροχή της χρωματικής περιγραφής των εικόνων με χρήση μασκών.

Τέλος, στο **Κεφάλαιο 5** προτείνονται βελτιώσεις των μεθόδων που υλοποιήθηκαν στο **Κεφάλαιο 4** και παρουσιάζονται νέες κατευθύνσεις που μπορούν μελλοντικά να βοηθήσουν στον τομέα της περιγραφής και ανάκτησης εικόνων μόδας.

2 Θεωρητικό υπόβαθρο

Στην ενότητα αυτή εισάγονται οι βασικές θεωρητικές έννοιες που θα χρειαστούν για την μετέπειτα κατανόηση μεθοδολογιών και τεχνικών με σκοπό την εξαγωγή χαρακτηριστικών από μια εικόνα και την κατάλληλη περιγραφή της. Αυτές οι τεχνικές θα αξιολογηθούν στο πειραματικό μέρος της εργασίας (**Κεφάλαιο 4**) ανάλογα με το ποια περιγραφή εξασφαλίζει καλύτερα αποτελέσματα ανάκτησης.

2.1 Τοπικός Ανιχνευτής και Περιγραφέας SIFT

Το ταίριασμα των εικόνων με βάση τα χαρακτηριστικά τους αποτελεί μια εύκολη διαδικασία για τον ανθρώπινο οφθαλμό. Ωστόσο, για τα υπολογιστικά συστήματα η εξαγωγή χρήσιμης και ουσιώδους πληροφορίας είναι πιο δύσκολη υπόθεση. Για το σκοπό αυτό απαιτείται ο εντοπισμός των κατάλληλων χαρακτηριστικών (features) της εικόνας και η μετέπειτα περιγραφή τους με τέτοιο τρόπο, ώστε να διευκολύνει το ταίριασμα με άλλες εικόνες. Με τον όρο κατάλληλα χαρακτηριστικά αναφερόμαστε σε χαρακτηριστικά της εικόνας που παραμένουν αναλλοίωτα σε αλλαγές κλίμακας και περιστροφές και ως ένα βαθμό δεν επηρεάζονται από αλλαγές φωτισμού. Πρόκειται για «σημεία κλειδιά» όπως τα ονομάζουμε της εικόνας τα οποία εντοπίζονται κυρίως σε χαμηλές συχνότητες προς αποφυγή θορύβου.

2.1.1 Περιγραφέας SIFT (Scale Invariant Feature Transform)

Για εντοπισμό τέτοιων «σημείων-κλειδιών» χρησιμοποιείται ο αλγόριθμος SIFT (Scale Invariant Feature Transform). Πρόκειται για έναν αλγόριθμο της όρασης υπολογιστών με σκοπό να εντοπίσει και να περιγράψει τοπικά χαρακτηριστικά σε εικόνες. Ο αλγόριθμος δόθηκε στη δημοσιότητα από τον David Lowe, το 1999. Τα βασικά βήματα παραγωγής των σημείων κλειδιών είναι τα ακόλουθα:

1. Ανίχνευση ακρότατων σε χώρο-κλίμακα (scale-space detection)

Στο πρώτο αυτό βήμα αναζητούνται τα βασικά χαρακτηριστικά στο χώρο και σε διαφορετικές κλίμακες κάνοντας χρήση των γκαουσιανών διαφορών, στις οποίες αναφερόμαστε αναλυτικά στην συνέχεια της Ενότητας 2.1.1, με σκοπό τα χαρακτηριστικά που θα βρεθούν να είναι αναλλοίωτα σε αλλαγές κλίμακας-περιστροφών.

2. Εντοπισμός σημείων-κλειδιών

Σε κάθε τοποθεσία που θεωρείται υποψήφια ως σημείο-κλειδί προσδιορίζεται η θέση και η κλίμακα και ανάλογα με την ευστάθεια των σημείων αυτών επιλέγονται τα τελικά σημεία-κλειδιά.

3. Ανάθεση προσανατολισμού

Στο βήμα αυτό επιδιώκεται αναλλοίωτη συμπεριφορά σε αλλαγές προσανατολισμού. Σε κάθε σημείο κλειδί αποδίδεται μία ή περισσότερες κατευθύνσεις με βάση τις κατευθύνσεις των παραγώγων της εικόνας. Η προσδιορισμένη αυτή κατεύθυνση θα αποτελέσει την βάση για όλους τους υπολογισμούς που θα ακολουθήσουν στα στοιχεία της εικόνας.

4. Περιγραφή σημείων κλειδιών

Γύρω από κάθε σημείο κλειδί και στην κλίμακα που αυτό έχει εντοπιστεί, υπολογίζονται οι παράγωγοι και μετατρέπονται σε μια κατάλληλη αναπαράσταση που επιτρέπει σημαντική παραμόρφωση και αλλαγές φωτισμού.

Στην συνέχεια ακολουθεί αναλυτική μελέτη των βημάτων του αλγορίθμου SIFT:

1. Ανίχνευση ακρότατων σε χώρο-κλίμακα

- Κατασκευή χώρου-κλίμακας (scale-space)

Για να ορίσουμε ένα σημείο ως **«σημείο-κλειδί»**, δηλαδή ένα καίριο σημείο της εικόνας, θα πρέπει αυτό να διατηρείται για διαφορετικές κλίμακές της. Για το λόγο αυτό, ορίζουμε μια συνάρτηση αναπαράστασης της αρχικής εικόνας σε διαφορετικά επίπεδα ανάλυσης. Αυτό γίνεται με τη δημιουργία ενός χώρου-κλίμακα. Κάθε εικόνα αναπαριστάνεται πλέον ως μια οικογένεια εικόνων που έχουν υποστεί εξομάλυνση, δηλαδή εικόνων σε διαφορετικές κλίμακες. Η αναπαράσταση αυτή, ονομάζεται **scale-space** αναπαράσταση και έχει ως παράμετρο μια μεταβλητή που καθορίζει την κάθε κλίμακα. Η

μεταβλητή κλίμακας χρησιμοποιείται για την καταστολή των πιο ασήμαντων δομών της εικόνας, με σκοπό να παραμείνουν τα πιο ευσταθή σε αλλαγές κλίμακας σημεία της. Αυτά θα είναι και τα υποψήφια σημεία κλειδιά.

Έχει αποδειχθεί από τους Koenderink (1984) και Lindeberg (1994) ότι ο καλύτερος δυνατός τύπος χώρου-κλίμακας είναι η γραμμική (Gaussian) χώρο-κλίμακα. Σε αυτή η μεταβλητή κλίμακας ορίζεται ως η τυπική απόκλιση της γκαουσιανής. Αναλυτικά, έστω I(x,y) η αρχική εικόνα, η γραμμική scale-space αναπαράσταση της είναι μια οικογένεια σημάτων L (x,y,t) που συμβολίζουν το σύνολο των εικόνων που προκύπτουν με την αλλαγή της κλίμακας. Αυτή ορίζεται ως η συνέλιξη της I (x, y) με μια γκαουσιανή συνάρτηση, έστω G(x,y,σ) τυπικής απόκλισης σ (**Εξίσωση 2.1 - 2.2**). Όσο μεγαλύτερη η τιμή σ τόσο μεγαλύτερη η θαμπάδα (blur) της εικόνας και άρα η καταστολή ασήμαντων λεπτομερειών.

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y) \qquad (2.1)$$

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2} \qquad (2.2)$$

Στην συνέχεια υπολογίζονται οι παράγωγοι δεύτερης τάξης (**Εξίσωση 2.3**) της συνάρτησης L. Η νέα συνάρτηση, γνωστή και ως "Laplacian of Gaussian" ή LoG, χρησιμοποιείται για να ανιχνεύσει ακμές, γωνίες της εικόνας και γενικότερα σημεία που θα μπορούσαν να αποτελέσουν σημεία -κλειδιά.

$$\nabla^2 L = L_{xx} + L_{yy} \qquad (2.3)$$

Ωστόσο η LoG εμφανίζει τα εξής μειονεκτήματα. Αρχικά ο υπολογισμός των παραγώγων έχει μεγάλο υπολογιστικό κόστος. Δεύτερον, η συνάρτηση LoG δεν είναι αναλλοίωτη σε αλλαγές κλίμακας (scale-invariant), εξαιτίας του σ^2 στον παρονομαστή της Εξίσωσης 2.2. Ο Lindeberg έδειξε ότι η κανονικοποίηση της LoG με τον παράγοντα σ^2 είναι απαραίτητη για την ανεξαρτησία από την κλίμακα. Μετά από πειραματικές συγκρίσεις, ο Mikolajczyk (2002) διαπίστωσε, ότι από τα μέγιστα και ελάχιστα της κανονικοποιημένης $\sigma^2\nabla^2 G$ προκύπτουν πιο σταθερά χαρακτηριστικά της εικόνας σε σχέση με άλλες συναρτήσεις, όπως η Harris συνάρτηση γωνιών. Προκειμένου να περιοριστεί το υπολογιστικό κόστος και να διατηρηθεί ανεξαρτησία ως προς την κλίμακα προτείνεται μια προσέγγιση της κανονικοποιημένης LoG, τις **διαφορές γκαουσιανών** όπου η κανονικοποίηση με τον όρο σ^2 επιτυγχάνεται αυτόματα.

- Διαφορές Γκαουσιανών (Difference of Gaussians)

Οι διαφορές γκαουσιανών ή εν συντομία DoG είναι η διαφορά των γκαουσιανών σε διαφορετικές κλίμακες kσ και σ, G(x,y,kσ) και G(x,y,σ). Η συνέλιξη τους με την εικόνα μας δίνει την συνάρτηση D(x,y,σ) στην οποία θα αναζητήσουμε μέγιστα-ελάχιστα.

$$D(x,y,\sigma) = \big(G(x,y,k\sigma) - G(x,y,\sigma)\big) * I(x,y)$$
$$= L(x,y,k\sigma) - L(x,y,\sigma)$$

Η προσέγγιση της DoG από την LoG προκύπτει κάνοντας χρήση της εξίσωσης διάχυσης θερμότητας, όπου την παράμετρο θερμικής διαχυτικότητας παίζει η τυπική απόκλιση(**Εξίσωση 2.4**).

$$\frac{\partial G}{\partial \sigma} = \sigma \nabla^2 G \quad (2.4)$$

Εάν γράψουμε την παράγωγο $\frac{\partial G}{\partial \sigma}$ προσεγγιστικά ως διαφορά των γειτονικών κλιμάκων σ, kσ προκύπτει:

$$\sigma \nabla^2 G = \frac{\partial G}{\partial \sigma} \approx \frac{G(x,y,k\sigma) - G(x,y,\sigma)}{k\sigma - \sigma}$$

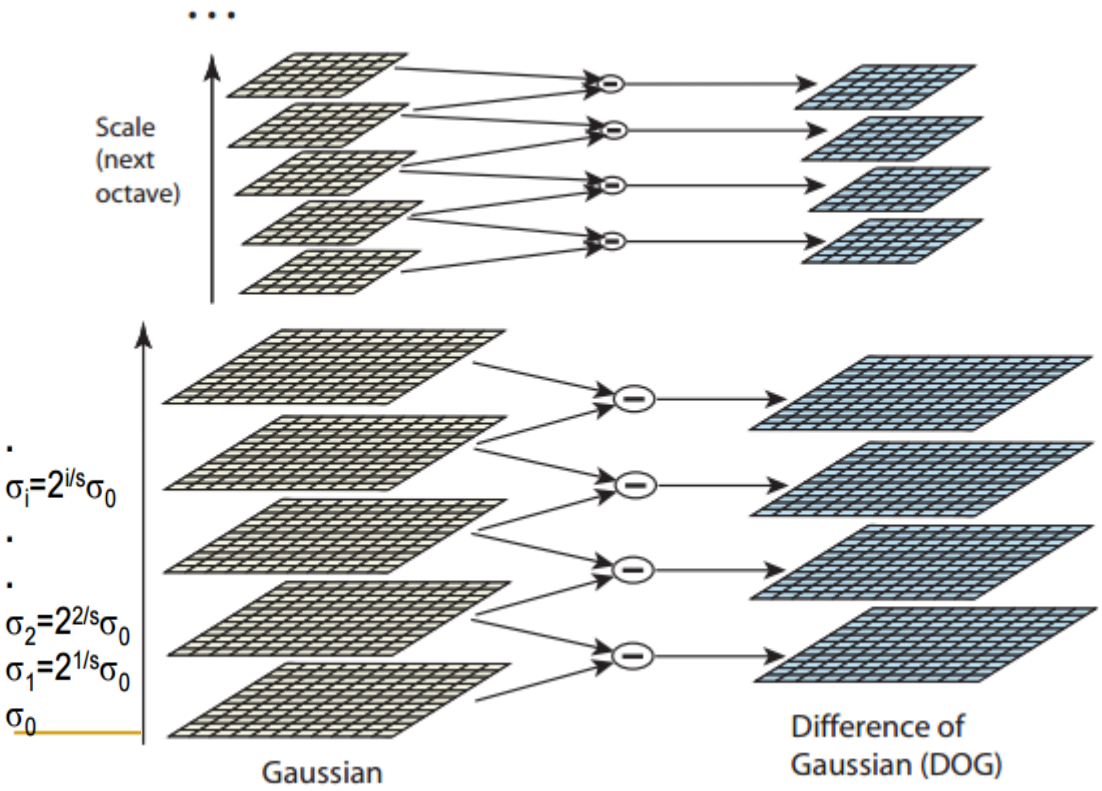
$$G(x,y,k\sigma) - G(x,y,\sigma) \approx (k-1)\sigma^2 \nabla^2 G \quad (2.5)$$

Από την **Εξίσωση 2.5** είναι προφανές το πλεονέκτημα της προσέγγισης αυτής, καθώς η συνάρτηση DoG είναι ήδη πολλαπλασιασμένη με τον παράγοντα σ^2 ανιχνεύοντας έτσι πολύ καλύτερα σημεία. Ακόμα, προκύπτει από απλή αφαίρεση άρα είναι γρήγορη και αποτελεσματική. Ωστόσο, έχει ως παράγοντα τον όρο k-1. Αυτό αποτελεί μικρό μειονέκτημα εφόσον είναι μια σταθερά πάνω από όλες τις κλίμακες και άρα δεν επηρεάζει την ανίχνευση μέγιστων τιμών, μόνο τις τιμές που λαμβάνουν στις θέσεις αυτές.

Η διαδικασία που περιγράψαμε, δηλαδή η δημιουργία χωροκλίμακας και η χρήση γκαουσιανών διαφορών με σκοπό την εύρεση σημείων ενδιαφέροντος, παρουσιάζεται στην **Εικόνα 2.1**. Αυτή αναλύεται ως εξής:

Αρχικά χωρίζεται την χώρο-κλίμακα σε οκτάβες. Αύξηση της κλίμακας κατά μία οκτάβα σημαίνει διπλασιασμό του μεγέθους του σ. Για παράδειγμα, η πρώτη οκτάβα χρησιμοποιεί κλίμακα σ, η δεύτερη οκτάβα χρησιμοποιεί κλίμακα 2σ, κλπ .

Σε κάθε οκτάβα, πραγματοποιείται συνέλιξη της αρχικής εικόνα επανειλημμένα με γκαουσιανές, ώστε να παράγει ένα σύνολο εικόνων χώρου- κλίμακας οι οποίες φαίνονται στοιβαγμένες στο αριστερό μέρος της **Εικόνας 2.1**



Εικόνα 2.1: *Χώρο-κλίμακες και δημιουργία διαφοράς Γκαουσιανών (DoG)*

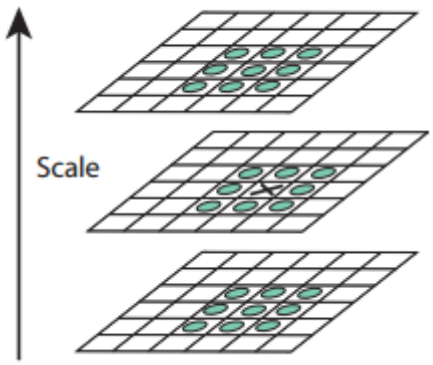
Διαιρούμε την κάθε οκτάβα γκαουσιανών σε έναν ακέραιο αριθμό επιπέδων s τέτοια ώστε η κλίμακα κάθε γκαουσιανής να διαφέρει σταθερά από την επόμενη κατά $k = 2^{1/s}$. Με βάση τα αποτελέσματα του Lowe μια καλή τιμή για την εύρεση σταθερών σημείων είναι s=3. Ο λόγος των υποδιαιρέσεων κάθε οκτάβας είναι ότι χωρίς αυτά τα υποεπίπεδα η πυραμίδα που θα προέκυπτε θα ήταν απότομη και θα ανταποκρινόταν μόνο σε δομές της εικόνας με κλίμακες που είναι δυνάμεις του δύο.

Επιπλέον, παρατηρούμε πως μετά από κάθε οκτάβα η εικόνα υφίσταται υποδειγματοληψία κατά έναν παράγοντα του 2, δηλαδή επιλέγεται κάθε δεύτερο εικονοστοιχείο κάθε γραμμής και στήλης της εικόνας που βρίσκεται στην κορυφή της προηγούμενης οκτάβας για να παράγει μια εικόνα με μέγεθος 1/4 του αρχικού της. Δεδομένου ότι όσο μεγαλώνει η κλίμακα τόσο οι εικόνες θολώνουν, άρα περιορίζεται η συχνότητα εμφάνισης χαρακτηριστικών τους, δεν χρειάζεται να διατηρηθούν όλα τα εικονοστοιχεία της εικόνας. Για αυτό και γίνεται σταδιακά μείωση του αριθμού τους όσο περισσότερο αυξάνεται η κλίμακα.

Τέλος, οι γειτονικές γκαουσιανές κάθε οκτάβας αφαιρούνται για να παράγουν τις συναρτήσεις DoG που φαίνονται στην δεξιά στήλη και στις οποίες και θα αναζητηθούν τα σημεία κλειδιά.

2. Εντοπισμός σημείων-κλειδιών

Προκειμένου να εντοπιστούν τα σημεία-κλειδιά της εικόνα αναζητούμε τα μέγιστα-ελάχιστα στην χωρό-κλίμακα συγκρίνοντας τα εικονοστοιχεία κάθε εικόνας D(x,y,σ) με τα οκτώ γειτονικά του στον χώρο και με κάθε έναν από τους εννέα γείτονες του στις κλίμακες εκατέρωθεν της δικής του. Επιλέγεται μόνο εάν είναι μεγαλύτερο ή μικρότερο από όλους τους γείτονες του.



Εικόνα 2.2: *Το σημείο X επιλέγεται ως υποψήφιο σημείο κλειδί μόνο αν είναι μεγαλύτερο από τα 26 γειτονικά του*

Για τον εντοπισμό σημείων-κλειδιών το υπολογιστικό κόστος δεν είναι ιδιαίτερα μεγάλο διότι για τα μη μέγιστα ή ελάχιστα σημεία αρκούν μερικές πρώτες συγκρίσεις για να απορριφθούν. Ένα ζήτημα που προκύπτει από την δειγματοληψία της εικόνας, τόσο στο χώρο αλλά και στην κλίμακα, είναι ότι μπορεί να χαθούν σημεία ενδιαφέροντος (μέγιστα-ελάχιστα) που βρίσκονται πολύ κοντά ανάλογα με την συχνότητα δειγματοληψίας που χρησιμοποιείται. Μετά από πειράματα που πραγματοποίησε ο Lowe, διαπίστωσε πως ένας καλός αριθμός συχνότητας σε επίπεδο κλίμακας είναι τρεις εικόνες ανά

οκτάβα, καθώς τότε εμφανίζεται η μεγαλύτερη επαναληψιμότητα των σημείων και στις τρεις οκτάβες, πράγμα που τα καθιστά ιδιαιτέρως πιθανά σημεία-κλειδιά. Αν και μπορούν να προστεθούν παραπάνω κλίμακες, το κόστος υπολογισμού θα αυξηθεί και όσο μεγαλώνει η κλίμακα τα μέγιστα ή ελάχιστα δεν θα είναι τόσο ευσταθή. Επίσης, στον χώρο προτείνεται τυπική απόκλιση σ_0=1.6.

Προηγουμένως αναφέραμε πως ένα εικονοστοιχείο X σημειώνεται ως σημείο κλειδί, εάν είναι μεγαλύτερο-μικρότερο από τα γειτονικά του. Όμως τα σημεία X αποτελούν προσεγγιστικά τα σημεία ακροτάτων, καθώς ένα ακρότατο σπάνια βρίσκεται ακριβώς στην θέση ενός εικονοστοιχείου. Για να ληφθεί πιο ακριβής τοποθεσία των ακροτάτων, χρησιμοποιούμε τη σειρά Taylor, όπου x είναι η μετατόπιση από το σημείο της δειγματοληψίας:

$$D(x) = D + \frac{\partial D^T}{\partial x} x + \frac{1}{2} x^T \frac{\partial^2 D}{\partial x^2} x \qquad (2.6)$$

Από την **Εξίσωση2.6** υπολογίζοντας την παράγωγο και θέτοντας την ίση με μηδέν, βρίσκουμε τις υποδιαιρέσεις των εικονοστοιχείων που είναι οι θέσεις των σημείων-κλειδιών:

$$\hat{x} = -\frac{\partial^2 D^{-1}}{\partial x^2} \frac{\partial D}{\partial x} \qquad (2.7)$$

Βρίσκοντας τις τιμές αυτές, αυξάνονται οι πιθανότητες ταιριάσματος των σημείων και η σταθερότητα του αλγορίθμου.

- <u>Απόρριψη «κακών» σημείων-κλειδιών</u>

Το προηγούμενο στάδιο παράγει πληθώρα βασικών σημείων. Μερικά από αυτά δεν είναι χρήσιμα σαν χαρακτηριστικά καθώς εντοπίζονται κατά μήκος μιας ακμής ή δεν έχουν αρκετή αντίθεση. Σε αμφότερες τις περιπτώσεις, θα πρέπει να απορριφθούν. Αυτό βοηθά στην αύξηση της αποτελεσματικότητας και την ευρωστία του αλγορίθμου.

1. Όταν τα σημεία παρουσιάζουν **χαμηλή αντίθεση** δηλαδή το μέτρο D(x) του τρέχοντος εικονοστοιχείου μιας εικόνας DoG είναι μικρότερο από κάποια τιμή αναφοράς τότε απορρίπτεται. Αντικαθιστώντας την Εξίσωση 2.7 στην επέκταση Taylor (Εξίσωση 2.6) για να λάβουμε την τιμή έντασης στο σημείο έχουμε:

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

Εάν το μέγεθος αυτό είναι μικρότερο από 0.03 το σημείο απορρίπτεται.

2. Η ιδέα που εφαρμόζεται για την απόρριψη σημείων **πάνω σε ακμές** βασίζεται στην γνώση, από τον ανιχνευτή γωνιών Harris, ότι για τις ακμές η μία ιδιοτιμή είναι σημαντικά μεγαλύτερη από την άλλη. Έτσι, χρησιμοποιώντας τις ιδιοτιμές του πίνακα Hessian, έστω H, μπορεί να καθοριστεί αν πρόκειται για σημείο μιας γωνίας, ακμής ή επίπεδης επιφάνειας.

$$H = \begin{bmatrix} D_{xx} & D_{xy} \\ D_{xy} & D_{yy} \end{bmatrix}$$

Οι παράγωγοι του πίνακα υπολογίζονται ως οι διαφορές γειτονικών σημείων.

Μια επίπεδη περιοχή θα έχει μικρές και τις δύο ιδιοτιμές του πίνακα ενώ μια γωνία θα έχει μεγάλες και τις δύο ιδιοτιμές. Τα σημεία που βρίσκονται πάνω σε ακμή θα έχουν τη μια ιδιοτιμή αρκετά μεγαλύτερη από την άλλη. Κάνοντας χρήση της προσέγγισης των Harris και Stephens (1988) προς αποφυγή υπολογισμού των ιδιοτιμών καθεαυτών, εφόσον μας ενδιαφέρει μόνο ο λόγος τους, και θέτοντας α την ιδιοτιμή με τη μεγαλύτερη τιμή και β με την μικρότερη, έχουμε από τον υπολογισμό του ίχνους και της ορίζουσας του πίνακα Hessian :

$$Tr(H) = D_{xx} + D_{yy} = a + \beta$$
$$Det(H) = D_{xx}D_{yy} - (D_{xy})^2 = a\beta$$

Εάν τεθεί α=rβ όπου r ο λόγος των ιδιοτιμών και κάνοντας χρήση των προηγούμενων σχέσεων προκύπτει η **Εξίσωση 2.8**

$$\frac{Tr(H)^2}{Det(H)} = \frac{(a+\beta)^2}{a\beta} = \frac{(r\beta+\beta)^2}{r\beta^2} = \frac{(r+1)^2}{r} \qquad (2.8)$$

Αυτή εξαρτάται μόνο από τον λόγο των ιδιοτιμών. Συγκεκριμένα, αν οι δυο ιδιοτιμές δεν διαφέρουν σημαντικά, και άρα ο λόγος τους είναι μικρός, η ποσότητα $\frac{(r+1)^2}{r}$ ελαττώνεται. Αντίθετα, αυξάνεται με την αύξηση του r. Άρα ο λόγος r επαρκεί για τον έλεγχο ύπαρξης σημείων κλειδιών σε ακμή. Το r στο δεύτερο μέρος της εξίσωσης παίρνει μια σταθερή τιμή ελέγχου, έστω r=10 (Lowe).

$$\frac{Tr(H)^2}{Det(H)} < \frac{(r+1)^2}{r}$$

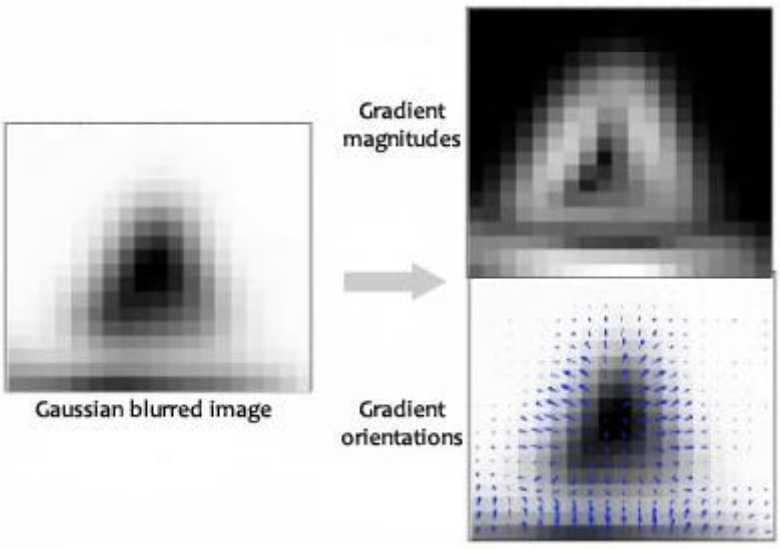
3. Ανάθεση προσανατολισμού

Από το προηγούμενο βήμα έχουν εξασφαλίστει ευσταθή σημεία-κλειδιά. Σκοπός του βήματος αυτού είναι να επιτευχθεί αναλλοίωτη συμπεριφορά σε αλλαγές κατεύθυνσης. Αυτό γίνεται μέσω της ανάθεσης προσανατολισμού σε κάθε σημείο-κλειδί ξεχωριστά.

Αναλυτικά, αφού έχει προσδιορίστει την κλίμακα που βρέθηκε το σημείο κλειδί επιλέγεται η εικόνα L(x,y,σ) , που έχει εξομαλυνθεί με γκαουσιανή συνάρτηση της οποίας η κλίμακα πλησιάζει την κλίμακα του σημείου-κλειδιού μιας εικόνας διαφορών DoG. Σε αυτήν την εικόνα εξετάζεται η γειτονιά γύρω από τη θέση του σημείου-κλειδιού ανάλογα με την κλίμακα που εντοπίστηκε. Όσο μεγαλύτερη η κλίμακα, τόσο μεγαλύτερη η περιοχή συλλογής. Στην γειτονία αυτή υπολογίζεται το μέτρο και η διεύθυνση των κλίσεων (gradient) από τους τύπους :

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

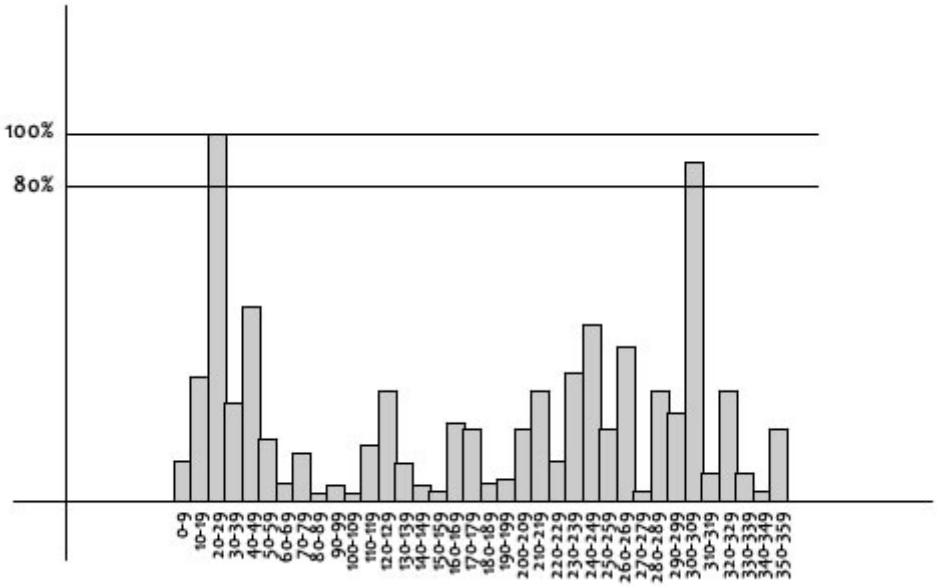
$$\theta(x,y) = tan^{-1}((L(x,y+1) - L(x,y-1))/(L(x+1,y) - L(x-1,y)))$$



Εικόνα 2.3 : *Παράδειγμα διάκρισης του μέτρου (gradient magnitude) και του προσανατολισμού των παραγώγων (gradient orientation), που σημειώνεται με μπλε χρώμα, μιας εικόνας που έχει εξομαλυνθεί με γκαουσιανή(Gaussian blurred image)*

Εν συνεχεία, δημιουργείται ένα ιστόγραμμα προσανατολισμού με 36 bins που καλύπτουν 360 μοίρες. Το ποσό που προστίθεται στον κάθε bins είναι ανάλογο

με το μέτρο της κλίσης (gradient magnitude) στις γειτονικές περιοχές του σημείου. Η ποσότητα που προστίθεται εξαρτάται επίσης από την απόσταση από το σημείο κλειδί. Συγκεκριμένα, το μέτρο των κλίσεων των γειτονικών σημείων που βρίσκονται πιο μακριά από το σημείο κλειδί θα πρέπει να συμβάλλουν λιγότερο στο ιστόγραμμα. Αυτό γίνεται χρησιμοποιώντας ένα γκαουσιανό παράθυρο πάνω από το σημείο κλειδί με σ 1,5 φορές μεγαλύτερο από την κλίμακα του σημείου-κλειδιού. Η υψηλότερη κορυφή στο ιστόγραμμα και κάθε κορυφή με ύψος πάνω από 80% της μεγαλύτερης κορυφής αποτελούν τις κυρίαρχες κατευθύνσεις που θα καθορίσουν τον προσανατολισμό του σημείου-κλειδιού. Έτσι, δημιουργείται ένα νέο σημείο - κλειδί με τις αντίστοιχες κατευθύνσεις του. Όλοι οι μετέπειτα υπολογισμοί γίνονται συγκριτικά με αυτόν τον προσανατολισμό. Εάν υπάρχουν πολλές κορυφές πάνω από το 80%, μετατρέπονται όλες σε νέα σημεία-κλειδιά με τις αντίστοιχες κατευθύνσεις τους.



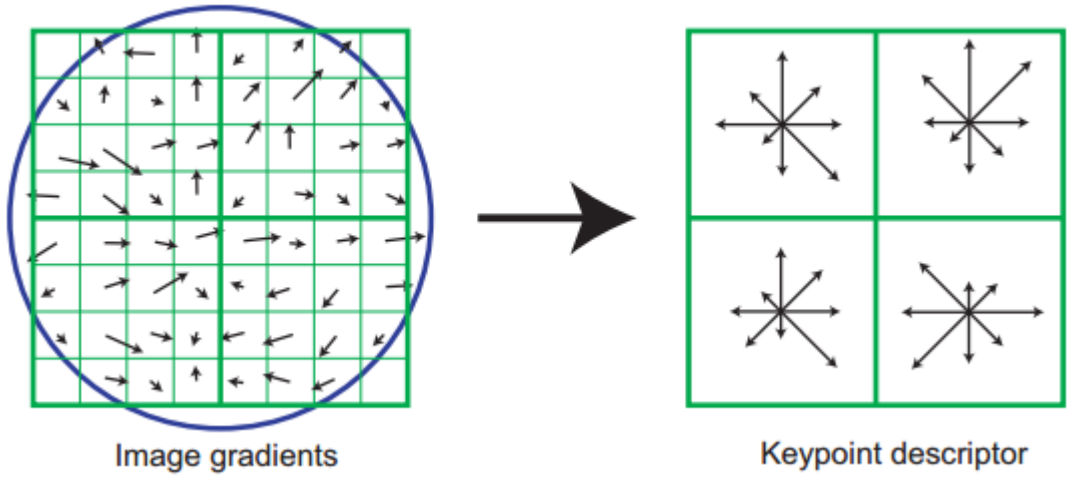
Εικόνα 2.4 *: Περίπτωση πολλαπλών κορυφών ιστογράμματος*

4. Περιγραφή σημείων-κλειδιών

Στο τελικό στάδιο, αφού έχουν εντοπιστεί τα σημεία κλειδιά, σκοπός είναι να γίνει η κατάλληλη περιγραφή τους, δηλαδή να δοθεί ένα μοναδικό αποτύπωμα που θα μας βοηθά να τα ξεχωρίσουμε.

Συγκεκριμένα, το αποτύπωμα αυτό είναι ένα διάνυσμα 1x128 που λαμβάνεται με τον εξής τρόπο. Αφού λάβουμε τη γειτονική περιοχή ενός σημείου-κλειδιού κάνοντας χρήση ενός παραθύρου διαστάσεων 16x16 γύρω από το σημείο-

κλειδί, το οποίο και χωρίζουμε σε δεκαέξι παράθυρα διαστάσεων 4x4, υπολογίζουμε το μέτρο και την κατεύθυνση των παραγώγων σε κάθε τέτοιο παράθυρο 4x4.



Εικόνα 2.5 : *Δημιουργία ιστογραμμάτων του περιγραφέα (Keypoint descriptor) βασισμένων στο μέτρο και τον προσανατολισμό των παραγώγων της εικόνας (Image Gradients) για κάθε παράθυρο 4x4 μιας γειτονιάς που έχει υποστεί εξομάλυνση με γκαουσιανή (μπλε κύκλος)*

Οι κατευθύνσεις θα διαιρεθούν σε ένα ιστόγραμμα των 8 bins, ενώ το ποσό που προστίθεται στο κάθε bin εξαρτάται από την απόσταση από το σημείο-κλειδί με βάση μια γκαουσιανή με *σ=1,5*κλίμακα*. Η διαδικασία επαναλαμβάνεται για το σύνολο των δεκαέξι γειτονιών 4x4 και άρα επιστρέφεται για κάθε γειτονιά ένα ιστόγραμμα των 8 bins, δηλαδή ένα διάνυσμα 1x128 για κάθε σημείο-κλειδί. Αυτό είναι το χαρακτηριστικό διάνυσμα το οποίο κανονικοποιείται και περιγράφει το σημείο-κλειδί μοναδικά. Η αναλλοίωτη συμπεριφορά σε αλλαγές προσανατολισμού επιτυγχάνεται μέσω του προσανατολισμού του σημείου που υπολογίσαμε στο βήμα 3. Είναι προφανές ότι η κλίση του 1x128 διανύσματος περιγραφής θα αλλάξει αν η εικόνα περιστραφεί. Για το λόγο αυτό, πρέπει να αφαιρεθεί από το διάνυσμα ο προσανατολισμός του σημείου κλειδιού του βήματος 4 που λειτουργεί ως προσανατολισμός αναφοράς.

Ένα ζήτημα που προκύπτει κατά την δημιουργία των ιστογραμμάτων είναι ότι κάποιο γειτονικό σημείο μπορεί να βρίσκεται στα όρια των 4x4 περιοχών και άρα να συμβάλλει σε περισσότερα από ένα ιστογράμματα. Προκειμένου η συμβολή του σημείου να γίνει δίκαια σε κάθε ιστόγραμμα, κάθε τιμή που εισάγεται σε κάποιο bin πολλαπλασιάζεται με ένα βάρος 1-d όπου d είναι η απόσταση του δείγματος από την κεντρική τιμή του ιστογράμματος.

Τέλος, προκειμένου αλλαγές του φωτισμού να μην επηρεάσουν τον περιγραφέα SIFT διακρίνουμε δύο περιπτώσεις που χρήζουν αντιμετώπισης. Στην περίπτωση αφινικών αλλαγών φωτισμού (δηλαδή αλλαγές φωτισμού της μορφής B =>αB+β, όπου B ο φωτισμός) πρέπει το διάνυσμα περιγραφής να κανονικοποιηθεί. Έτσι, στην περίπτωση που κάθε εικονοστοιχείο πολλαπλασιαστεί κατά μια σταθερά α, την ίδια επίπτωση θα υποστούν και οι κλίσεις (gradients), άρα η κανονικοποίηση θα ακυρώσει την κοινή σταθερά α. Αν πάλι η φωτεινότητα αλλάξει κατά μια σταθερά β, η επίδραση της σταθεράς θα ακυρωθεί λόγω των παραγώγων, εφόσον αυτές υπολογίζονται ως διαφορές εικονοστοιχείων. Στην περίπτωση μη γραμμικών αλλαγών φωτισμού, οι αλλαγές αυτές θα επηρεάσουν κυρίως τα μέτρα και όχι των προσανατολισμό των μερικών παραγώγων. Για να ακυρώσουμε επομένως την επίδραση τέτοιων μεγάλων μέτρων, χρησιμοποιείται ένα κατώφλι το οποίο αν υπερβούν οι τιμές του διανύσματος περιγραφής, τίθενται ίσες με αυτό και επαναλαμβάνεται εκ νέου κανονικοποίηση του διανύσματος. Η τιμή κατωφλίου προέκυψε από πειράματα (βλέπε Lowe) ίση με 0.2.

2.2 Χρωματική Περιγραφή

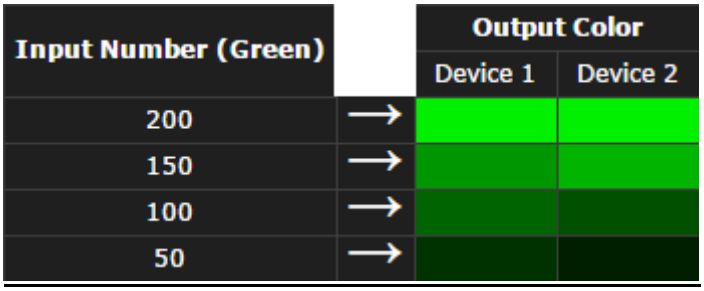
Στην συνέχεια του **Κεφαλαίου 2** εισάγονται οι βασικές έννοιες που αφορούν το χρώμα των εικόνων και θα χρησιμεύσουν ιδιαίτερα στο πειραματικό μέρος της εργασίας που αφορά στην εξαγωγή χρωματικής πληροφορίας αποκλειστικά.

2.2.1 Χρωματικοί Χώροι (Color Spaces)

Προτού αναλύσουμε του χρωματικούς χώρους δηλαδή τα χαρακτηριστικά και τις διαφορές τους, είναι θεμιτό να παρουσιάσουμε τους λόγους που αναπτύχθηκαν και σε τι εξυπηρετούν. Για το λόγο αυτό, ορίζουμε την διαδικασία «Διαχείρισης χρωμάτων». Με τον όρο αυτό αναφερόμαστε στην διαδικασία κατά την οποία κάνοντας χρήση των χαρακτηριστικών των χρωμάτων για κάθε συσκευή σε μια αλυσίδα απεικόνισης, όπως της **Εικόνας 2.6**, αναπαράγουμε όσο ποιο σωστά γίνεται τα χρώματα.

Εικόνα 2.6: *Αλυσίδα απεικόνισης χρωμάτων*

Στην **Εικόνα 2.6** τα χρώματα του αρχικού τοπίου που απεικονίζεται θα μεταφραστούν και θα αποθηκευτούν με διαφορετικό τρόπο από την κάμερα και εν συνεχεία από οποιαδήποτε άλλη συσκευή αναπαράστασης τους. Σκοπός της διαχείρισης είναι να εξασφαλιστεί ότι η τελική συσκευή απεικόνισης θα αποδώσει το αρχικό χρώμα όσο πιο σωστά γίνεται.



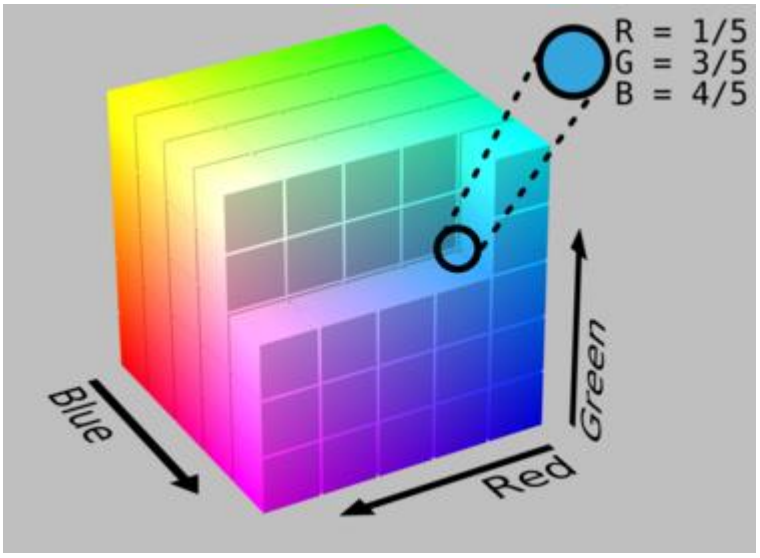
Εικόνα 2.7: *Διαφορά απόδοσης χρωμάτων από συσκευή 1 και 2*

Για να το πετύχουμε, χρειάζεται να δημιουργηθεί ένα προφίλ χρώματος της κάθε συσκευής που θα υποδεικνύει πώς μια τιμή εισόδου μεταφράζεται σε χρώμα εξόδου (**Εικόνα 2.7**). Οι χρωματικοί χώροι αποτελούν χρήσιμα εργαλεία για το σκοπό αυτό, δηλαδή την κατανόηση της συμβατότητας χρώματος μεταξύ δύο διαφορετικών συσκευών. Μπορούν να θεωρηθούν μία καλά οργανωμένη ψηφιακή παλέτα χρωμάτων που δείχνουν τις δυνατότητες μας όταν προσπαθούμε να αναπαράγουμε το χρώμα σε μια άλλη συσκευή, δηλαδή ποιες λεπτομέρειες, όπως σκιές ή κορεσμοί χρώματος, μπορούν να διατηρηθούν. Πρόκειται για τρισδιάστατα αντικείμενα που περιέχουν όλους τους χρωματικούς συνδυασμούς, ενώ κάθε διάσταση τους αντιπροσωπεύει κάποια πτυχή του χρώματος, όπως φωτεινότητα, κορεσμός ή απόχρωση.

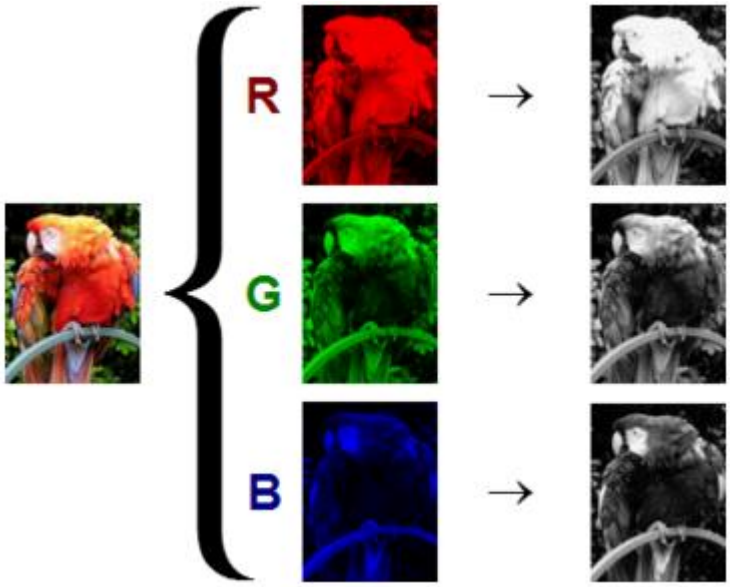
Στην συνέχεια της ενότητας παρουσιάζονται συνοπτικά οι χρωματικοί χώροι που θα φανούν χρήσιμοι στις μετέπειτα ενότητες και κυρίως στο πειραματικό μέρος.

Τύποι χρωματικών χώρων

- **RGB**: Πρόκειται για **γραμμικό** χρωματικό μοντέλο στο οποίο προστίθενται κόκκινο, πράσινο και μπλε φως σε διαφορετικούς συνδυασμούς, έτσι ώστε να αναπαραχθεί ένα ευρύ φάσμα χρωμάτων. Η ιδέα αυτή αναπαράστασης του χρώματος ως διάνυσμα τριών πρωταρχικών συνιστωσών και μάλιστα του κόκκινου, πράσινου και μπλε προήλθε από τη διαπίστωση ότι τα κωνία του ανθρώπινου οφθαλμού είναι ιδιαίτερα ευαίσθητα σε αυτά τα μήκη κύματος, δηλαδή στην κόκκινη, πράσινη και μπλε ακτινοβολία αντίστοιχα. Το RGB είναι ένα μοντέλο χρώματος **εξαρτώμενο από τη συσκευή**, δηλαδή διαφορετικές συσκευές ανιχνεύουν ή αναπαράγουν μια δεδομένη τιμή RGB διαφορετικά. Έτσι, μια τιμή RGB δεν καθορίζει το ίδιο χρώμα σε όλες τις συσκευές χωρίς κάποιο είδος διαχείρισης χρωμάτων.



Εικόνα 2.8: *Κύβος αναπαράστασης RGB χρωματικού μοντέλου*



Εικόνα 2.9: *Ανάλυση εικόνας στα τρία κανάλια RGB*

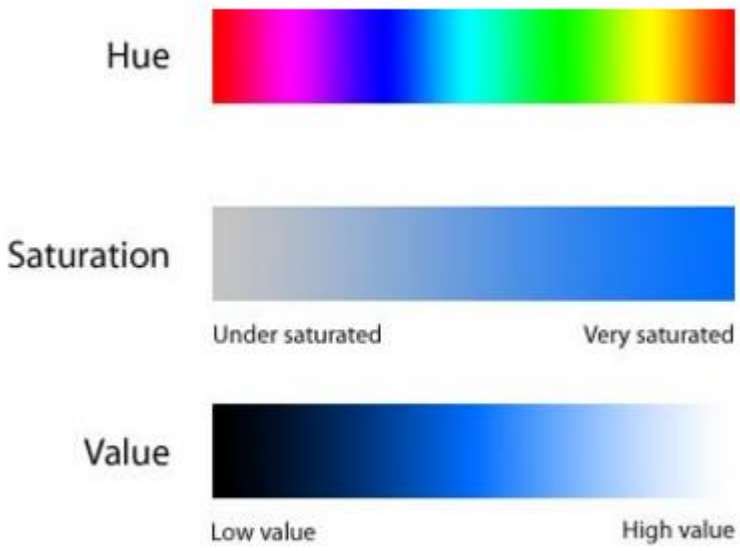
Μια ψηφιακή εικόνα που τα χρώματα της αντιστοιχούν στο RGB χρωματικό μοντέλο μπορεί να αναλυθεί στα τρία κανάλια, δηλαδή σε τρεις ξεχωριστές εικόνες greyscale που αντιπροσωπεύουν την ένταση του κάθε καναλιού χρώματος (**Εικόνα 2.9**).

- **HSV**: Τα αρχικά αυτού του μοντέλου προέρχονται από τις λέξεις **hue-saturation-value**, δηλαδή απόχρωση, κορεσμός, τιμή. Με τον όρο απόχρωση χρώματος αναφερόμαστε στο καθαρό χρώμα στο οποίο εκείνο

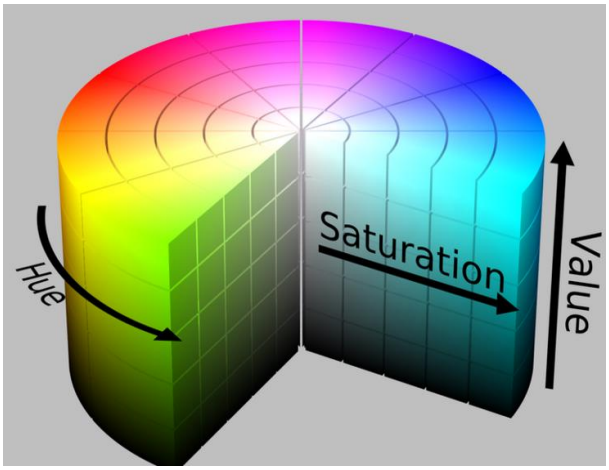
μοιάζει. Για παράδειγμα, όλοι οι διαφορετικοί τόνοι κόκκινου έχουν την ίδια απόχρωση. Η απόχρωση δηλαδή μεταβάλλεται καθώς μεταβαίνουμε από το ένα χρώμα στο άλλο. Ο κορεσμός του χρώματος περιγράφει πόσο άσπρο περιέχει το χρώμα. Ένα καθαρό κόκκινο είναι πλήρως κορεσμένο, με κορεσμό 1 ενώ το λευκό έχει κορεσμό 0. Με τον όρο τιμή αναφερόμαστε στην φωτεινότητα (brightness) του χρώματος. Όσο μειώνεται η τιμή, τόσο πιο σκοτεινό γίνεται το χρώμα. Το HSV μοντέλο παρέχει μια διαισθητικά καλύτερη αντίληψη του χρώματος επειδή είναι συχνά πιο φυσικό να σκεφτούμε ένα χρώμα από την άποψη της απόχρωσης και κορεσμού αντί ενός συνόλου στοιχείων που προστίθενται (βλέπε RGB). Ο HSV είναι ένας **μη γραμμικός** μετασχηματισμός του χρωματικού χώρου RGB σύμφωνα με τις σχέσεις:

$$H = cos^{-1}\left\{\frac{\frac{1}{2}[(R-G)+(R-B)]}{\sqrt{(R-G)^2+(R-B)(G-B)}}\right\}$$

$$S = 1 - \frac{3}{R+G+B}[\min(R,G,B)], \quad V = \frac{1}{3}(R+G+B)$$

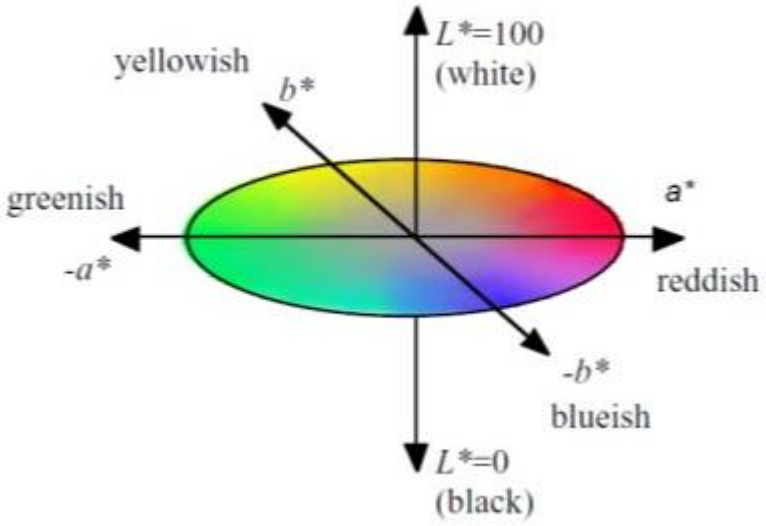


Εικόνα 2.10: *Κλίμακα απόχρωσης (Hue) του μοντέλου HSV. Αλλαγή κορεσμού (Saturation). Από δεξιά προς αριστερά βλέπουμε τα αποτελέσματα μείωσης του κορεσμού. Αλλαγή τιμής (Value). Αύξηση φωτεινότητας από αριστερά προς τα δεξιά.*



Εικόνα 2.11: *Αναπαράσταση HSV χρωματικού χώρου*

- **CIELAB (L*a*b*):** Το χρωματικό μοντέλο CIELAB έχει σχεδιαστεί για να προσεγγίσει την ανθρώπινη όραση και το πώς αυτή λειτουργεί. Σχεδιάστηκε με βάση εργαστηριακά πειράματα υπό ελεγχόμενες συνθήκες και προσπαθεί να εξασφαλίσει **γραμμική αντίληψη** του χρώματος. Αυτό σημαίνει πως η απόσταση δυο σημείων στο χρωματικό χώρο θα αντιστοιχούν σε ανάλογη διαφορά για το ανθρώπινο μάτι (perceptual linearity). Η συνιστώσα L αναφέρεται στη φωτεινότητα, με L=0 να αντιστοιχεί στο πιο σκούρο μαύρο και L=1 στο πιο φωτεινό λευκό. Στην **Εικόνα 2.12**, στον άξονα a αναπαριστάνεται η διαφορά μεταξύ πράσινου-κόκκινου χρώματος με το πράσινο χρώμα να παίρνει τιμές στον αρνητικό ημιάξονα και το κόκκινο στον θετικό, ενώ στον άξονα b αναπαριστάνεται η διαφορά κίτρινου-μπλε, με το μπλε στον αρνητικό ημιάξονα και τον κίτρινο στον θετικό.



Εικόνα 2.12: *Αναπαράσταση LAB χώρου*

Οι δύο άξονες χρώματος, a* και b*, βασίζονται στην θεωρία των **ανταγωνιστικών χρωμάτων** (opponent colors). Η θεωρία των ανταγωνιστικών χρωμάτων, που υποστηρίχθηκε για πρώτη φορά από τον Ewald Hering, περίπου το 1870, βασίζεται στο ότι οι τρεις τύποι των κωνίων του ανθρώπινου οφθαλμού έχουν κάποια επικάλυψη στα μήκη κύματος του φωτός και αυτό καθιστά πιο αποτελεσματικό για το οπτικό σύστημα να καταγράφει διαφορές μεταξύ των αποκρίσεων των κωνίων, αντί για τις ξεχωριστές αποκρίσεις τους. Στη θεωρία των ανταγωνιστικών χρωμάτων υπάρχουν τρία ανταγωνιστικά ζεύγη: το κόκκινο σε σχέση με πράσινο, το μπλε σε σχέση με το κίτρινο και ένα μη χρωματικό ζεύγος, το μαύρο έναντι του λευκού, που ανιχνεύει τις αλλαγές φωτεινότητας. Μια σημαντική ιδιότητα του χώρου είναι η ανεξαρτησία του από συσκευές (**device independence**).

Αυτοί είναι οι βασικοί χρωματικοί χώροι στους οποίους θα βασιστούν όλοι όσοι θα αναφερθούμε μετέπειτα.

2.2.2 Φυσική του χρώματος

Στην ενότητα αυτή μελετώνται οι φυσικοί κανόνες που διέπουν το χρώμα με σκοπό την καλύτερη κατανόηση του και την ανάπτυξη κατάλληλων οπτικών περιγραφέων χρώματος, δηλαδή διανυσμάτων που θα εμπεριέχουν όσο το δυνατό πιο εύρωστη χρωματική πληροφορία.

Η γενική ιδέα όσον αφορά το χρώμα είναι η εξής: Το χρώμα είναι ανακλώμενο φως. Τα περισσότερα αντικείμενα ανακλούν ένα μέρος του φωτός και το υπόλοιπο το απορροφούν. Το χρώμα προκύπτει από τα μήκη κύματος που ανακλώνται. Η εμφάνιση ενός έγχρωμου αντικειμένου εξαρτάται από τις ιδιότητες του αλλά και το είδος του φωτισμού που χρησιμοποιείται.

Η ποσότητα του φωτός που ανακλάται από ένα αντικείμενο, και το πώς αυτό ανακλάται, εξαρτάται σε μεγάλο βαθμό από το είδος της επιφάνειας δηλαδή την ομαλότητα ή την υφή της. Η ανάκλαση μπορεί να χωρισθεί σε δύο κατηγορίες, την **κατοπτρική ανάκλαση** (specular reflection) και την **διάχυτη ανάκλαση** (diffuse reflection). Η κατοπτρική ανάκλαση μπορεί να οριστεί ως το φως που ανακλάται από μία λεία επιφάνεια σε μια ορισμένη γωνία, ίση με τη γωνία του προσπίπτοντος φωτός, ενώ η διάχυτη ορίζεται ως η ανάκλαση που παράγεται από τραχείες επιφάνειες που τείνουν να αντανακλούν το φως προς όλες τις κατευθύνσεις. Οι περισσότερες επιφάνειες είναι ένας συνδυασμός των κατοπτρικών και διάχυτων ανακλάσεων.

Δεδομένου του ότι το χρώμα ενός υλικού καθορίζεται από το βαθμό στον οποίο ένα υλικό αντανακλά το φως των διαφορετικών μηκών κύματος συνειδητοποιούμε ότι ένα υλικό μπορεί να έχει δύο διαφορετικά χρώματα, ένα διάχυτο χρώμα που υποδεικνύει τον τρόπο που το υλικό αντανακλά το φως διάχυτα και ένα κατοπτρικό χρώμα που υποδεικνύει πώς να αντανακλά το φως κατοπτρικά. Σύμφωνα με το διχρωματικό μοντέλο ανάκλασης του Shafer (Dichromatic reflection model) το χρώμα ενός αντικειμένου μπορεί να περιγραφεί σαν συνδυασμός των δύο αυτών ανακλάσεων.

Όμως, όταν ένα χρωματιστό αντικείμενο έχει τόσο διάχυτη όσο και κατοπτρική ανάκλαση, συνήθως μόνο η διάχυτη συνιστώσα είναι χρωματισμένη. Άρα μια απλουστευμένη μορφή απόδοσης του χρώματος μπορεί να περιλαμβάνει μόνο τη διάχυτη συνιστώσα. Για το λόγο αυτό ορίζουμε τη Lambertian ανάκλαση. Με τον όρο Lambertian ανάκλαση αναφερόμαστε στην ανάκλαση μιας Lambertian επιφάνειας δηλαδή μια επιφάνειας διάχυτης ανάκλασης. Υποθέτοντας μια τέτοια επιφάνεια, οι μετρούμενες τιμές παρατήρησης της εικόνας f = {R,G,B}, μπορούν να αναπαρασταθούν από την παρακάτω εξίσωση:

$$f(x) = \int_{\omega} e(\lambda)\rho_k(\lambda)s(x,\lambda)d\lambda \quad (2.9)$$

,όπου x είναι η χωρική μεταβλητή της εικόνας και ω το ορατό φάσμα. Η πηγή φωτός μοντελοποιείται ως μια ενιαία πηγή με χρώμα e (λ), όπου λ είναι το μήκος κύματος. Ο όρος s (x,λ) είναι η ανάκλαση της Lambertian επιφάνειας και ρ_k (λ) είναι η συνάρτηση ευαισθησίας της κάμερας (k ∈{R, G, B}).

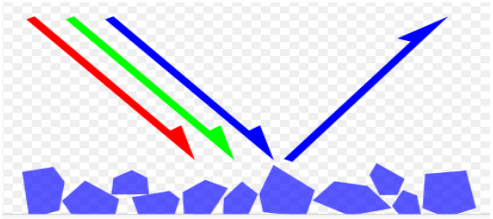
Στο μοντέλο της εξίσωσης κρίνεται σκόπιμο να προστεθεί ένας επιπλέον όρος που αντιστοιχεί στο **διάχυτο φως** του περιβάλλοντα χώρου, το οποίο θεωρείται ότι έχει μικρότερη ένταση και προέρχεται από όλες τις κατευθύνσεις σε ίσες ποσότητες. Ο όρος αυτός, ωστόσο, περιλαμβάνει ένα ευρύτερο φάσμα πιθανών αιτιών πέρα από το διάχυτο φως, όπως η ευαισθησία υπερύθρων του αισθητήρα της κάμερας, η σκέδαση στο μέσο ή φακό κλπ. Στην **Εξίσωση 2.9** άρα προστίθεται ένας επιπλέον όρος, A(λ):

$$f(x) = \int_{\omega} e(\lambda)\rho_k(\lambda)s(x,\lambda)d\lambda + \int_{\omega} A(\lambda)\rho_k(\lambda)d\lambda$$

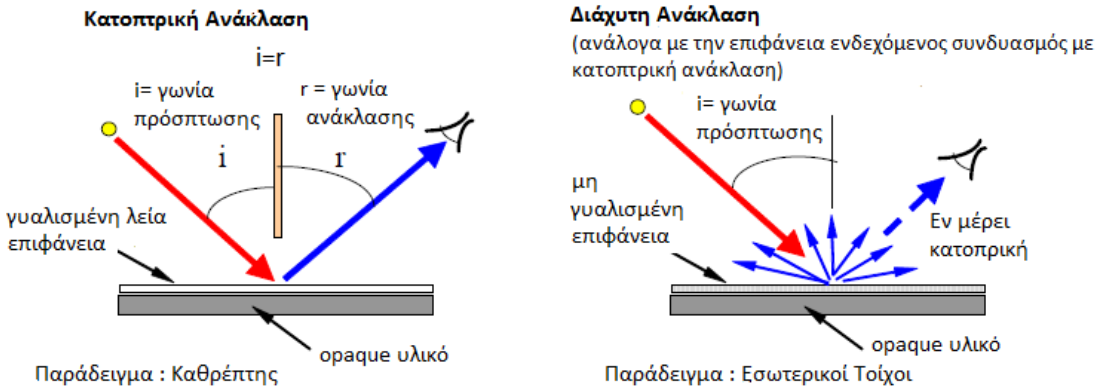
Εάν υπολογίσουμε την χωρική παράγωγο της εικόνας $f_{\chi,\sigma}$(x) στην κλίμακα σ έχουμε :

$$f_{x,\sigma}(x) = \int_{\omega} e(\lambda)\,\rho_k(\lambda)s_{x,\sigma}(x,\lambda)d\lambda$$

δηλαδή η επίδραση του όρου A(λ) του διάχυτου φωτός καταργείται, πράγμα που σημαίνει πως η παράγωγος εξασφαλίζει αναλλοίωτη συμπεριφορά στο διάχυτο φως.



Εικόνα 2.13: *Παράδειγμα ανάκλασης από μια επιφάνεια. Το χρώμα που βλέπουμε είναι αυτό που ανακλάται από αυτήν*



Εικόνα 2.14: *Παράδειγμα κατοπτρικής (αριστερά) και διάχυτης ανάκλασης (δεξιά).*

Από την προηγούμενη ανάλυση είναι σαφές ότι αλλαγές στον φωτισμό (θέση και είδος προσπίπτουσας ακτινοβολίας) οδηγούν σε αλλαγές της ανακλώμενης ακτινοβολίας και άρα σε αλλαγές χρώματος των αντικειμένων. Το ανθρώπινο νευρικό σύστημα πετυχαίνει να προσαρμόζεται σε αλλαγές φωτισμού και κατορθώνει να διατηρεί την εμφάνιση των χρωμάτων ενός αντικειμένου. Έτσι, παρά την ευρεία διακύμανση του φωτός που μπορεί να αντανακλάται από ένα αντικείμενο, τα αντικείμενα εμφανίζουν σταθερό χρώμα.

Για να επιτευχθεί χρωματική προσαρμογή από ψηφιακά συστήματα όπως ψηφιακές κάμερες, αναπτύχθηκε από τον von Kries το **διαγώνιο μοντέλο**. Η μέθοδος που χρησιμοποιεί είναι να παρομοιάσει το ανθρώπινο νευρικό σύστημα και να εφαρμόσει ένα κέρδος για αποκρίσεις ευαισθησίας καθενός από τα κωνία έτσι ώστε να κρατήσει το προσαρμοσμένο το χρώμα. Η Εξίσωση

2.9 μπορεί να αναπαρασταθεί από **το διαγώνιο μοντέλο** ή **van Kries** μοντέλο ως εξής:

$$f^c = D^{u,c} f^u$$

$$\begin{pmatrix} R^c \\ G^c \\ B^c \end{pmatrix} = \begin{pmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & c \end{pmatrix} \begin{pmatrix} R^u \\ G^u \\ B^u \end{pmatrix} \quad (2.10)$$

Με τον όρο $f^u=(R^u,G^u,B^u)$ αναφερόμαστε στην εικόνα που τραβήχτηκε κάτω από άγνωστη πηγή φωτός, ενώ $f^c=(R^c,G^c,B^c)$ είναι η ίδια εικόνα μεταμορφωμένη, έτσι ώστε να φαίνεται σαν να είχε ληφθεί κάτω από το φωτισμό αναφοράς c και τον οποίο ονομάζουμε κανονική πηγή φωτός. Η $D^{u,c}$ είναι ένας διαγώνιος πίνακας που αντιστοιχεί τα χρώματα που έχουν ληφθεί κάτω από μια άγνωστη πηγή φωτός u με το αντίστοιχα χρώματα τους κάτω από την κανονική πηγή φωτός c. Εάν προσθέσουμε και έναν επιπλέον όρο για τον διάχυτο φωτισμό του περιβάλλοντα χώρου, η εξίσωση μετατρέπεται σε :

$$\begin{pmatrix} R^c \\ G^c \\ B^c \end{pmatrix} = \begin{pmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & c \end{pmatrix} \begin{pmatrix} R^u \\ G^u \\ B^u \end{pmatrix} + \begin{pmatrix} o_1 \\ o_2 \\ o_3 \end{pmatrix}$$

,όπου το διάνυσμα $o=(o_1,o_2,o_3)$ παριστάνει το επιπλέον διάχυτο φως σε κάθε κανάλι.

Σκοπός της ανάλυσης που προηγήθηκε είναι να αναπτυχθούν οι κατάλληλοι περιγραφείς χρώματος που θα παρουσιάζουν αμεταβλητότητα (invariance), πράγμα μου σημαίνει πως οποιαδήποτε αλλαγή των συνθηκών φωτισμού δεν θα επηρεάζουν την περιγραφή του χρώματος και οι περιγραφείς θα μένουν αμετάβλητοι.

Σύμφωνα με την μελέτη των K. van de Sande, T. Gevers, C. G. M. Snoek «**Evaluating Color Descriptors for Object and Scene Recognition»**, θεωρούμε πέντε τύπους αλλαγών που μπορεί να υποστεί η εικόνα f(x):

1. **Αλλαγές στην ένταση του φωτισμού:** Η αλλαγή αυτή περιλαμβάνει τόσο αλλαγές στην ένταση της πηγής φωτισμού όσο και αλλαγές λόγω σκίασης και μεταφράζεται σε πολλαπλασιασμό κάθε καναλιού με την ίδια μεταβλητή α=b=c. Στην περίπτωση αυτή ο περιγραφέας που παραμένει αμετάβλητος σε τέτοιες αλλαγές λέμε πως παρουσιάζει **scale invariance** όσον αφορά την ένταση, δηλαδή παραμένει αναλλοίωτος όσον αφορά αλλαγές κλίμακας της έντασης του φωτισμού.

$$\begin{pmatrix} R^c \\ G^c \\ B^c \end{pmatrix} = \begin{pmatrix} a & 0 & 0 \\ 0 & a & 0 \\ 0 & 0 & a \end{pmatrix} \begin{pmatrix} R^u \\ G^u \\ B^u \end{pmatrix}$$

2. **Μετατοπίσεις της έντασης του φωτισμού:** Αυτές οφείλονται σε διάχυτο φωτισμό εξαιτίας για παράδειγμα της σκέδασης μιας πηγής λευκού φωτός, την ευαισθησία υπερύθρων του αισθητήρα της κάμερας κλπ. Όταν ένας περιγραφέας είναι αναλλοίωτος σε μια ελαφριά μετατόπιση έντασης, τότε παρουσιάζει **shift-invariance** όσο αφορά την ένταση του φωτός, δηλαδή παραμένει αναλλοίωτος σε μετατοπίσεις της έντασης φωτισμού. Στην περίπτωση αυτή έχουμε $o_1=o_2=o_3$, δηλαδή θεωρούμε πως η μετατόπιση γίνεται ομοιόμορφα στα τρία κανάλια.

3. **Συνδυασμό**ς αλλαγών κλίμακας και μετατόπισης της έντασης φωτισμού

$$\begin{pmatrix} R^c \\ G^c \\ B^c \end{pmatrix} = \begin{pmatrix} a & 0 & 0 \\ 0 & a & 0 \\ 0 & 0 & a \end{pmatrix} \begin{pmatrix} R^u \\ G^u \\ B^u \end{pmatrix} + \begin{pmatrix} o_1 \\ o_1 \\ o_1 \end{pmatrix}$$

Ο περιγραφέας στην περίπτωση αυτή είναι **scale και shift invariant** όσον αφορά την ένταση του φωτισμού.

4. **Αλλαγές χρώματος** του φωτός: Οι αλλαγές αυτές περιλαμβάνουν αλλαγές στο χρώμα της πηγής αναφοράς ή σκέδασης του φωτός και παριστάνονται από την **Εξίσωση 2.10** με $\alpha \neq b \neq c$.

5. **Μετατόπιση του χρώματος** της πηγής: Οι αλλαγές αυτές χρώματος εκφράζονται με $o_1 \neq o_2 \neq o_3$ πέρα από $\alpha \neq b \neq c$.

2.2.3 Χρωματικοί Περιγραφείς (Color Descriptors)

Με τον όρο χρωματικοί περιγραφείς αναφερόμαστε σε διανύσματα τα οποία συνοδεύουν τις εικόνες προκειμένου να τις διαφοροποιούν με βάση τα χαρακτηριστικά των χρωμάτων τους. Εφόσον τα διανύσματα αυτά υπολογιστούν για ένα σύνολο εικόνων και καταχωρηθούν σε μια βάση

δεδομένων, μπορεί να επιτευχθεί ανάκτηση εικόνων με βάση την ομοιότητα των χρωματικών χαρακτηριστικών. Στην ενότητα αυτή παρουσιάζονται οι βασικοί περιγραφείς χρώματος και εξετάζεται η ικανότητα τους να παραμένουν αμετάβλητοι στις αλλαγές φωτισμού που περιγράψαμε στην προηγούμενη ενότητα.

1. Χρωματικά ιστογράμματα

Παρακάτω παρουσιάζονται οι περιγραφείς χρώματος που βασίζονται σε ιστογράμματα. Τα ιστογράμματα εξετάζουν τις ιδιότητες των εικονοστοιχείων και δεν περιέχουν χωρική πληροφορία, δηλαδή η περιγραφή χρώματος γίνεται γενικά για όλη την εικόνα.

RGB ιστόγραμμα: Πρόκειται για ένα συνδυασμό των ιστογραμμάτων του κάθε καναλιού του χρωματικού χώρου RGB. Δεν διαθέτει invariance ιδιότητες σε αλλαγές του φωτισμού.

Opponent ιστόγραμμα: Αρχικά ορίζουμε τον Opponent χώρο σε σχέση με τον χρωματικό χώρο RGB:

$$\begin{pmatrix} O_1 \\ O_2 \\ O_3 \end{pmatrix} = \begin{pmatrix} \frac{R-G}{\sqrt{2}} \\ \frac{R+G-2B}{\sqrt{6}} \\ \frac{R+G+B}{\sqrt{3}} \end{pmatrix}$$

Στο κανάλι O_3 έχουμε την πληροφορία έντασης-φωτεινότητας, και στα κανάλια O_1,O_2 την χρωματική πληροφορία. Από τον παραπάνω ορισμό προκύπτει ότι τα κανάλια O_1,O_2 είναι αναλλοίωτα όσον αφορά την ένταση του φωτός, αφού αν προσθέσουμε έναν επιπλέον άγνωστο φωτισμό ισχύει:

$$\begin{pmatrix} O_1 \\ O_2 \end{pmatrix} = \begin{pmatrix} \frac{R^c-G^c}{\sqrt{2}} \\ \frac{R^c+G^c-2B^c}{\sqrt{6}} \end{pmatrix} = \begin{pmatrix} \frac{(R^u+o_1)-(G^u+o_1)}{\sqrt{2}} \\ \frac{(R^u+o_1)+(G^u+o_1)-2(B^u+o_1)}{\sqrt{6}} \end{pmatrix} = \begin{pmatrix} \frac{R^c-G^c}{\sqrt{2}} \\ \frac{R^c+G^c-2B^c}{\sqrt{6}} \end{pmatrix}$$

Ωστόσο σε οποιαδήποτε άλλη αλλαγή φωτισμού ή χρώματος, τα κανάλια δεν μένουν αναλλοίωτα.

Hue ιστόγραμμα: Πρόκειται για το ιστόγραμμα του καναλιού απόχρωσης (hue) του HSV χώρου που δίνεται από την σχέση:

$$hue = \arctan\left(\frac{O_1}{O_2}\right) = arctan\left(\frac{\sqrt{3}(R-G)}{(R-G-2B)}\right)$$

Έχει παρατηρηθεί πως η απόχρωση παρουσιάζει αστάθεια κοντά στον γκρι άξονα. Για το λόγο αυτό εφαρμόζεται η μέθοδος που προτείνεται από τους Joost van de Weijer και Cordelia Schmid στο σύγγραμμα **«Coloring Local Feature Extraction»**, προκειμένου να επιτευχθεί μεγαλύτερη ευρωστία. Συγκεκριμένα, με βάση την ανάλυση λάθους (error analysis) της απόχρωσης έχουμε :

$$(\partial hue)^2 = (\frac{\partial hue}{\partial O_1}\partial O_1)^2 + (\frac{\partial hue}{\partial O_2}\partial O_2)^2 = \frac{1}{{O_1}^2 + {O_2}^2} = \frac{1}{saturation^2}$$

Προκύπτει ότι η βεβαιότητα της απόχρωσης είναι αντιστρόφως ανάλογη του κορεσμού, δηλαδή μεγαλύτερος κορεσμός συνεπάγεται μικρότερη αβεβαιότητα της απόχρωσης. Αυτή την πληροφορία την χρησιμοποιούμε στο πειραματικό κομμάτι για να χτίσουμε εύρωστα ιστογράμματα απόχρωσης.

Το ιστόγραμμα παρουσιάζει **scale και shift invariance** όσον αφορά την ένταση του φωτισμού.

rg ιστόγραμμα: Πρόκειται για ένα ιστόγραμμα RGB που έχει υποστεί κανονικοποίηση, δηλαδή γράφεται ως :

$$\begin{pmatrix} r \\ g \\ b \end{pmatrix} = \begin{pmatrix} \frac{R}{R+G+B} \\ \frac{G}{R+G+B} \\ \frac{B}{R+G+B} \end{pmatrix}$$

Η χρωματική πληροφορία εμπεριέχεται στα κανάλια r,g αφού το κανάλι b είναι περιττό. Τα κανάλια r,g είναι αναλλοίωτα σε αλλαγές κλίμακας της έντασης φωτισμού **(scale invariance)**. Αναλυτικά,

$$\binom{r}{g} = \begin{pmatrix} \frac{R^c}{R^c + G^c + B^c} \\ \frac{G^c}{R^c + G^c + B^c} \end{pmatrix} = \begin{pmatrix} \frac{aR^u}{aR^u + aG^u + aB^u} \\ \frac{aG^u}{aR^u + aG^u + aB^u} \end{pmatrix} = \begin{pmatrix} \frac{R^c}{R^c + G^c + B^c} \\ \frac{G^c}{R^c + G^c + B^c} \end{pmatrix}$$

Ιστόγραμμα Μεταμορφωμένη κατανομή χρώματος-Transformed color distribution: Σκοπός της κατανομής αυτής είναι να εξασφαλιστεί το αναλλοίωτο σε αλλαγές των συνθηκών φωτισμού που δεν παρουσιάζει το απλό rgb ιστόγραμμα αλλά και το αναλλοίωτο σε αλλαγές χρώματος του φωτισμού και το πετυχαίνει με κανονικοποίηση κάθε καναλιού ξεχωριστά. Τα κανάλια ορίζονται ως

$$\begin{pmatrix} R' \\ G' \\ B' \end{pmatrix} = \begin{pmatrix} \frac{R - \mu_R}{\sigma_R} \\ \frac{G - \mu_G}{\sigma_G} \\ \frac{B - \mu_B}{\sigma_B} \end{pmatrix}$$

όπου μ_κ ο μέσος όρος και σ_κ η τυπική απόκλιση του κάθε καναλιού με κ=R,G,B. Έτσι, για κάθε κανάλι εξασφαλίζεται κατανομή χρώματος με μ=0 και σ=1.

2. Χρωματικές Ροπές (Color Moments)

Στα μαθηματικά, οι ροπές αποτελούν ειδικά ποσοτικά μέτρα του σχήματος ενός συνόλου σημείων. Εάν τα σημεία αντιπροσωπεύουν μια κατανομή πιθανότητας, η μηδενική ροπή είναι η συνολική πιθανότητα , η πρώτη ροπή είναι η **μέση τιμή**, η δεύτερη ροπή αντιστοιχεί στη **διακύμανση**, η τρίτη ροπή είναι η **ασυμμετρία**, και η τέταρτη η **κύρτωση**.

Η βασική ιδέα πίσω από τις **χρωματικές ροπές** είναι να θεωρηθεί ότι η κατανομή του χρώματος σε μια εικόνα μπορεί να ερμηνευτεί ως μια κατανομή πιθανοτήτων και άρα να υπολογιστούν οι ροπές που αντιστοιχούν σε αυτήν

και να λειτουργήσουν ως τα χαρακτηριστικά (features) που θα προσδιορίζουν την εικόνα.

Μια έγχρωμη εικόνα αντιστοιχεί σε μια συνάρτηση I που καθορίζει τις τρεις τιμές RGB που αντιστοιχούν στις θέσεις της εικόνας

$$I(x,y) \rightarrow (R(x,y), G(x,y), B(x,y))$$

Από Mindru et al, στη μελέτη «**Moment invariants for recognition under changing viewpoint and illumination»** ορίζονται οι γενικευμένες χρωματικές ροπές τάξης p+q και βαθμού a+b+c ως εξής:

$$M_{pq}^{abc} = \iint x^p \, y^q [I_R(x,y)]^a [I_G(x,y)]^b [I_B(x,y)]^c dxdy$$

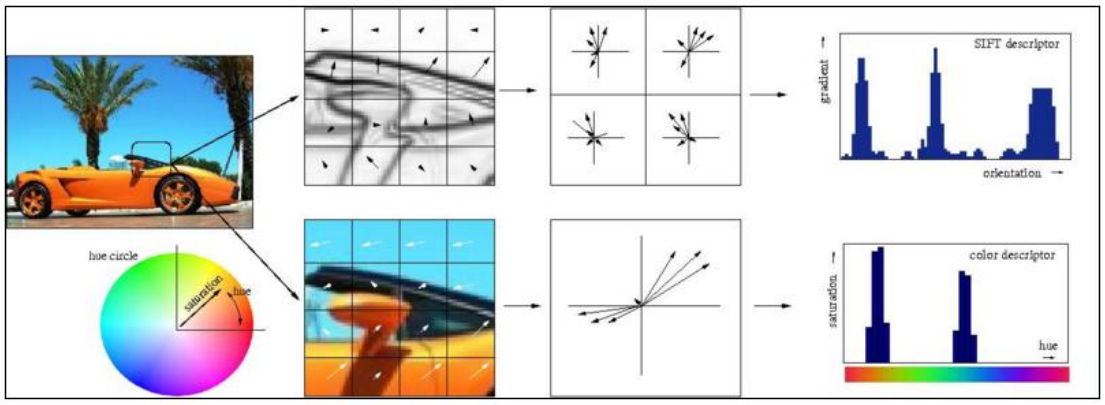
Τυπικά, χρησιμοποιούνται ροπές μέχρι πρώτης τάξης και δευτέρου βαθμού. Αυτό οδηγεί σε εννιά πιθανούς συνδυασμούς τάξης, εάν εξαιρεθεί η σταθερή ροπή $M^{000}{}_p$ και σε τρεις συνδυασμούς βαθμών, άρα σε ένα διάνυσμα περιγραφής 27 διαστάσεων. Οι ροπές αυτές είναι **shift invariant** ως προς την ένταση του φωτισμού. Προκειμένου να επιτευχθεί αμετάβλητη συμπεριφορά και σε άλλες συνθήκες φωτισμού, προτείνεται ένας συνδυασμός χρωματικών ροπών που ονομάζονται **color moments invariants**.

3. Χρώμα και SIFT

Πρόκειται για συνδυασμό των ιστογραμμάτων με τον περιγραφέα SIFT. Αμφότερες οι μέθοδοι έχουν αναλυθεί σε προηγούμενες ενότητες.

HSV SIFT: Στην περίπτωση του HSV SIFT, βρίσκουμε τους SIFT περιγραφείς σε κάθε κανάλι του χώρου HSV, δηλαδή ένα διάνυσμα 1x128 για κάθε κανάλι (3x128 συνολικά). Το μοντέλο αυτό δε διαθέτει αναλλοίωτη συμπεριφορά για οποιαδήποτε μεταβολή του φωτισμού, διότι συνδυάζει το κανάλι Hue με τα κανάλια S,V που επηρεάζονται από τέτοιες αλλαγές.

Hue SIFT: Απομονώνοντας το hue κανάλι και εφαρμόζοντας τη μέθοδο που προτείνεται για μείωση της αστάθειας του γκρι άξονα, ο περιγραφέας αυτός συνενώνει τον SIFT περιγραφέα με το hue ιστόγραμμα σε ένα ενιαίο διάνυσμα. Ο περιγραφέας αυτός είναι αναλλοίωτος σε μεταβολές κλίμακας και μετατόπισης του φωτισμού.



Εικόνα 2.15: *Εφαρμογή του Hue SIFT περιγραφέα*

Στην **Εικόνα 2.15** παρατηρούμε πως προκύπτουν τα δύο ιστογράμματα, το ένα με βάση τα SIFT χαρακτηριστικά της εικόνας και το άλλο με χρωματικά κριτήρια. Το πρώτο ιστόγραμμα έχει αναλυθεί στην Ενότητα 2.2.1 που αφορά τον περιγραφέα SIFT. Το δεύτερο προκύπτει με βάση την απόχρωση κάθε 4x4 υποδιαίρεσης του σημείου-κλειδιού που εξετάζουμε. Η θέση κάθε ψήφου δηλαδή στο ιστόγραμμα εξαρτάται από την απόχρωση ενώ το μέτρο της ψήφου ζυγίζεται ανάλογα με τον κορεσμό στην θέση αυτή, προκειμένου να επιτευχθεί ευστάθεια του γκρι άξονα.

Opponent SIFT: Αρχικά μετατρέπουμε τον χρωματικό χώρο σε Opponent και υπολογίζουμε τους SIFT περιγραφείς σε κάθε κανάλι, παρόμοια με τον περιγραφέα HSV SIFT. Το αποτέλεσμα δίνει έναν περιγραφέα αναλλοίωτο σε αλλαγές κλίμακας και μετατόπισης. Παρόλο που το Opponent ιστόγραμμα είναι αναλλοίωτο στα κανάλια O_1,O_2 μόνο για μεταβολές μετατόπισης, λόγω του SIFT περιγραφέα (βλέπε **Ενότητα 2.2.1**) εξασφαλίζεται το αναλλοίωτο τόσο σε κλίμακα όσο και σε μετατόπιση φωτισμού (**scale and shift invariance**).

C SIFT : Ο περιγραφέας αυτός δημιουργήθηκε για να εξασφαλίσει αναλλοίωτη συμπεριφορά σε αλλαγές έντασης στα κανάλια O_1,O_2 του Opponent χώρου. Αυτό επιτυγχάνεται κανονικοποιώντας τον χώρο Opponent βάση των σχέσεων $\frac{O_1}{O_3}$ και $\frac{O_2}{O_3}$. Άρα οι όροι α=b=c της **Εξίσωσης 2.10** που εκφράζουν αλλαγές στην ένταση του φωτισμού θα διαγραφούν (**scale invariance**). Ωστόσο ο περιγραφέας δεν είναι αναλλοίωτος σε αλλαγές μετατόπισης (shift invariance).

Rg SIFT: Για τον περιγραφέα αυτόν υπολογίζουμε τους περιγραφείς SIFT στον κανονικοποιημένο χρωματικό χώρο rg. Ο περιγραφέας έχει τις ίδιες αναλλοίωτες ιδιότητες με το rg ιστόγραμμα.

Transformed Color SIFT: Οι SIFT περιγραφείς υπολογίζονται σε κάθε κανάλι ξεχωριστά του Transformed χρωματικού χώρου. Ισχύουν οι ίδιες αναλλοίωτες ιδιότητες όπως και στο αντίστοιχο ιστόγραμμα.

4. Χρωματικές Ετικέτες (Color Naming)

Με τον όρο χρωματικές ετικέτες εννοούμε τη διαδικασία ανάθεσης χρώματος στα εικονοστοιχεία της εικόνας με τον ίδιο τρόπο που θα περιγράφαμε το χρώμα στην καθημερινή μας ζωή, δηλαδή με λέξεις (πχ.κόκκινο φόρεμα, πράσινο σακάκι). Η μέθοδος αυτή παρουσιάζεται στο σύγγραμμα των *Weijer, Schmid, Verbeek, Larlus,* ***«Color naming for real world applications»***. Ο τρόπος αυτός περιγραφής χρώματος εμφανίζει τα εξής πλεονεκτήματα σε σχέση με τους χρωματικούς περιγραφείς :

1. Με τις χρωματικές ετικέτες τόσο η **περιγραφή,** όσο και ο **έλεγχος** ορθότητάς τους γίνεται **πιο απλά** και κατανοητά για τον άνθρωπο από ότι με τους περιγραφείς ή τα ιστογράμματα που επιστρέφουν ένα διάνυσμα του οποίου το περιεχόμενο δεν γίνεται άμεσα αντιληπτό. Με το RGB ιστόγραμμα, για παράδειγμα, έχουμε ένα διάνυσμα που εκφράζει την κατανομή φωτεινότητας του κάθε καναλιού χρώματος. Με τις χρωματικές ετικέτες κάνουμε αντιστοίχιση των RGB τιμών σε απλά ονόματα χρωμάτων όπως χρησιμοποιούνται στην καθομιλουμένη. Έτσι, έχουμε ένα διάνυσμα κατανομής χρωμάτων, δηλαδή 15 κόκκινα εικονοστοιχεία, 42 κίτρινα κλπ.

Εικόνα 2.16*: Διάνυσμα χρωματικών ετικετών*

Όμως, και ο έλεγχος ορθότητας των ετικετών είναι πολύ πιο απλός. Στην **Εικόνα 2.17** εύκολα μπορούμε να διαπιστώσουμε ότι στο μεγαλύτερο μέρος του φορέματος έχει δοθεί ορθά η χρωματική ετικέτα μπλε.

***Εικόνα 2.17**: Εικόνα μετά την υλοποίηση χρωματικών ετικετών*

2. Οι χρωματικές ετικέτες είναι πιο εύρωστες σε φωτομετρικές αλλαγές καθώς η εκπαίδευση του μοντέλου έχει γίνει σε εικόνες πραγματικού κόσμου (**real world images**). Με τον όρο αυτό εννοούμε εικόνες χωρίς τον ιδανικό φωτισμό ενός εργαστηρίου ή ένα ιδανικό λευκό φόντο. Αν και υπήρξαν αρκετές προγενέστερες μελέτες για την εκμάθηση χρωμάτων από color chips (**Εικόνα 2.18**), στη μελέτη **«*Color naming for real world applications»*** προτείνεται η εκμάθηση να γίνεται από εικόνες πραγματικού κόσμου οι οποίες παρουσιάζουν ιδιαιτερότητες όπως σκιές, σφάλματα συμπίεσης κλπ. Το γεγονός ότι το μοντέλο έχει εκπαιδευτεί κατ'αυτόν τον τρόπο είναι ιδιαίτερα χρήσιμη στην παρούσα εργασία καθώς οι εικόνες μόδας που προέρχονται από επιδείξεις εμφανίζουν τέτοιες ιδιαιτερότητες, όπως σκιές ή διαφορετικές πηγές φωτισμού.

***Εικόνα 2.18**: Παραδείγματα Color Chips*

Βασικά Χρώματα

Προτού περιγράψουμε τον τρόπο εκμάθησης των χρωματικών ετικετών είναι σημαντικό να προσδιορίσουμε τον όρο βασικά χρώματα. **Βασικά χρώματα** ονομάζονται εκείνα στα οποία όταν αναφερόμαστε, είναι ξεκάθαρο ποια είναι χωρίς καμία ασάφεια. Αναγνωρίζονται από την πλειοψηφία των ανθρώπων , ενώ συναντώνται σε διάφορες γλώσσες και πάντα για να περιγράψουν το ίδιο χρώμα. Για παράδειγμα, το λιλά ή το μπορντώ δεν είναι σίγουρο ότι θα είναι ευκόλως αναγνωρίσιμα σε σύγκριση με το κόκκινο ή το μπλε.

Κάθε γλώσσα και κουλτούρα εκφράζει διαφορετικά τον τρόπο που αντιλαμβάνεται την πραγματικότητα. Το ίδιο συμβαίνει και με τα χρώματα καθώς οι λαοί χωρίζουν διαφορετικά τον χρωματικό χώρο και αναθέτουν διαφορετικά χρωματικά ονόματα. Η ρωσική γλώσσα, για παράδειγμα , είναι μία από τις γλώσσες που έχει 12 βασικούς όρους χρώματος. Ο όρος χρώματος μπλε χωρίζεται σε δύο χρωματικούς όρους: goluboi (goluboiˇ), και siniy (siniiˇ).

Στην μελέτη τους οι Berlin και Kay ***«Basic Color Terms: Their Universality and Evolution»,*** ωστόσο, αναλύοντας ενενήντα οκτώ γλώσσες, διαπίστωσαν ότι όλες οι γλώσσες, αν και διαφέρουν μεταξύ τους, συγκλίνουν σε έντεκα χρωματικές λέξεις. Τα έντεκα αυτά χρώματα είναι τα βασικά χρώματα της αγγλικής γλώσσας: μαύρο, άσπρο, γκρι, καφέ, μπλε, πράσινο, κόκκινο, ροζ, πορτοκαλί, κίτρινο και μωβ

Περιγραφή Εικόνων Εκμάθησης (Training Dataset)

Όπως αναφέραμε παραπάνω η μελέτη, **«*Color naming for real world applications»,*** διαφοροποιείται από τις προηγούμενες μελέτες πάνω στην ανάθεση χρώματος, χρησιμοποιώντας εικόνες πραγματικού κόσμου για την εκμάθηση του μοντέλου και μάλιστα εικόνες προερχόμενες από το Google Image Search δίνοντας ως λέξη προς αναζήτηση το χρώμα που μας ενδιαφέρει και τη λέξη χρώμα πχ. «κόκκινο χρώμα». Χρησιμοποιώντας το όνομα αρχείου της εικόνας και τα μεταδεδομένα της ιστοσελίδας από την οποία προέρχεται, γίνεται ανάκτηση των εικόνων με βάση το χρώμα. Τα χρώματα που αναζητούνται είναι τα έντεκα βασικά.

Η ανάθεση χρώματος αρχικά γινόταν σε εργαστηριακές συνθήκες.Για το σκοπό αυτό, διάφοροι άνθρωποι ρωτήθηκαν σχετικά με το τι χρωματική ονομασία θα

έδιναν σε ένα σύνολο από color chips. Σταδιακά, για την αποφυγή ανάθεσης χρώματος με το χέρι, χρησιμοποιήθηκε το διαδίκτυο, όπου ζητήθηκε από χρήστες να αναθέσουν την καλύτερη χρωματική ονομασία.

Με βάση τα παραπάνω, είναι εύκολα αντιληπτό πως η χρησιμότητα της εκμάθησης εικόνων μέσω του Google Image Search είναι διττή. Πρώτον, η ανάθεση γίνεται αυτόματα από την μηχανή αναζήτησης χωρίς να χρειάζεται να ερωτηθεί κανείς και δεύτερον γιατί οι εικόνες που επιστρέφονται περιλαμβάνουν διάφορες αποχρώσεις, προέρχονται από διαφορετικές κάμερες και έχουν αποθηκευτεί με διάφορες μεθόδους συμπίεσης, άρα η εκπαίδευση ενός μοντέλου εκμάθησης χρώματος από τέτοιες εικόνες εξασφαλίζει καλύτερη απόδοση σε σφάλματα που παρουσιάζονται υπό πραγματικές συνθήκες. Επιπλέον, διασφαλίζεται η ευελιξία προσθήκης επιπλέον χρωμάτων πέρα από των βασικών, όπως το μπεζ, με απλή αναζήτηση στην μηχανή Google «μπεζ χρώμα». Αντίθετα, με τις μεθόδους εκμάθησης βασισμένες σε chips αυτό συνεπάγεται την επανάληψη όλης της διαδικασίας, δηλαδή της ανθρώπινης επισήμανσης του χρώματος για όλα τα chips από την αρχή.

Ωστόσο, ανάμεσα στο πλήθος των εικόνων που επιστρέφονται είναι και κάποιες που δεν αντιστοιχούν στο χρώμα που αναζητείται (false positives). Επιπλέον, δεν υπάρχει καμία πληροφορία σχετικά με το σημείο της εικόνας που εντοπίζεται το συγκεκριμένο χρώμα (ασθενής ανάθεση). Πολλές φορές μόνο ένα μικρό κομμάτι της εικόνας αντιστοιχεί στο χρώμα αναζήτησης.

Εικόνα 2.19: *Εικόνες που επιστρέφονται μετά από αναζήτηση «κόκκινο χρώμα» στο Google Image Search. Παρατηρούμε ότι η πρώτη αντιστοιχεί σωστά στο χρώμα, η δεύτερη λανθασμένα (false postive) και η τρίτη σωστά για ένα μικρό κομμάτι της εικόνας(ασθενής ανάθεση).*

Προεπεξεργασία

Τα ασθενώς επονομαζόμενα στοιχεία που περιλαμβάνονται στο σύνολο των εικόνων που λαμβάνονται από την μηχανή αναζήτησης επιβάλουν προεπεξεργασία με σκοπό να απομονωθεί το τμήμα της εικόνας που μας αφορά. Για παράδειγμα, στην **Εικόνα 2.19** θέλουμε να κρατήσουμε μόνο το κόκκινο αυτοκίνητο. Για την αφαίρεση κάποιων από τα εικονοστοιχεία τα οποία είναι πιθανό να μην υποδεικνύονται ορθά από την ετικέτα της εικόνας, αφαιρούμε το υπόβαθρο από τις εικόνες της Google μέσω της επαναληπτικής αφαίρεσης εικονοστοιχείων τα οποία διαθέτουν το ίδιο χρώμα με τα σύνορα (borders) της εικόνας. Επιπλέον, δεδομένου ότι η ετικέτα χρώματος συχνά αναφέρεται σε ένα αντικείμενο που βρίσκεται στο κέντρο της εικόνας, γίνεται περικοπή αυτής έτσι ώστε να είναι στο 70% του αρχικού πλάτους και ύψους της.

Επιπλέον οι εικόνες στο στάδιο της προεπεξεργασίας υφίστανται διόρθωση γάμμα (gamma correction) και γίνεται μετατροπή τους στο χρωματικό χώρο L*a*b* που περιγράψαμε στην **Ενότητα. 2.2.1**. Ο λόγος αυτής της μετατροπής είναι η ιδιότητα της γραμμικής αντίληψης του L*a*b* χώρου (perceptual linearity) περισσότερο από άλλους χρωματικούς χώρους. Γραμμικότητα αντίληψης σημαίνει ότι η αλλαγή του ίδιου ποσού σε μια τιμή χρώματος θα πρέπει να παράγει μια μεταβολή περίπου ίδιας οπτικής σημασίας για τον άνθρωπο. Στη συνέχεια οι εικόνες, θα αναπαρασταθούν με ιστογράμματα γεγονός που επιβεβαιώνει τη σημασία της γραμμικότητας του L*a*b* χώρου καθώς καθιστά πιο εύκολη την δημιουργία ενιαίων bins για το ιστόγραμμα.

Εκμάθηση ετικετών

Στο στάδιο εκπαίδευσης στόχος είναι η εκμάθηση του μοντέλου ώστε να αναγνωρίζει τα χρώματα χρησιμοποιώντας ως εικόνες εκπαίδευσης τις εικόνες από το Google Image Search. Με βάση αυτές θα χτίσει ένα οπτικό λεξιλόγιο χρωμάτων, δηλαδή θα αποδίδει για κάθε χρώμα την αντίστοιχη λέξη-ετικέτα.

Με βάση την παραπάνω περιγραφή είναι εμφανής η ανάγκη ύπαρξης ενός **''μοντέλου κρυμμένων μεταβλητών''** ή **latent variable model**. Ένα τέτοιο μοντέλο είναι ένα στατιστικό μοντέλο που συσχετίζει ένα σύνολο μεταβλητών που είναι εμφανείς (manifest variables) με ένα σύνολο κρυμμένων μεταβλητών οι οποίες δεν είναι άμεσα μετρήσιμες (latent variables). Στην περίπτωση μας, εμφανείς μεταβλητές είναι τα εικονοστοιχεία της εικόνας και κρυμμένες μεταβλητές το χρώμα.

Τα μοντέλα αυτά χρησιμοποιούνται ευρέως στην ανάλυση κειμένου όπου σκοπός είναι η εξαγωγή κρυμμένων σημασιολογικών θεμάτων (semantic topics) από έγγραφα λαμβάνοντας υπόψιν τις λέξεις του εγγράφου. Στο κομμάτι της όρασης υπολογιστών, μπορούμε να αντιστοιχήσουμε τα έγγραφα με εικόνες, τις λέξεις με τις RGB τιμές των εικονοστοιχείων και τα θέματα από σημασιολογικά σε οπτικά. Συγκεκριμένα για τον σκοπό αυτής της εργασίας το θέμα μας (topic) είναι το χρώμα. Στον πίνακα που ακολουθεί παρουσιάζεται αυτή η αντιστοίχηση:

	Σημασιολογικά Στοιχεία (Semantic)	**Οπτικά** Στοιχεία (Visual)
Εμφανείς μεταβλητές (Manifest variables)	Έγγραφα	Εικόνες
	Λέξεις	RGB τιμές
Κρυμμένες μεταβλητές (Latent variables)	Θέμα Κειμένου	Χρώμα

Πίνακας 2.1: *Αντιστοίχιση στοιχείων σημασιολογικού και οπτικού περιεχομένου*

Πιθανοτική Λανθάνουσα Σημασιολογική Ανάλυση (PLSA)

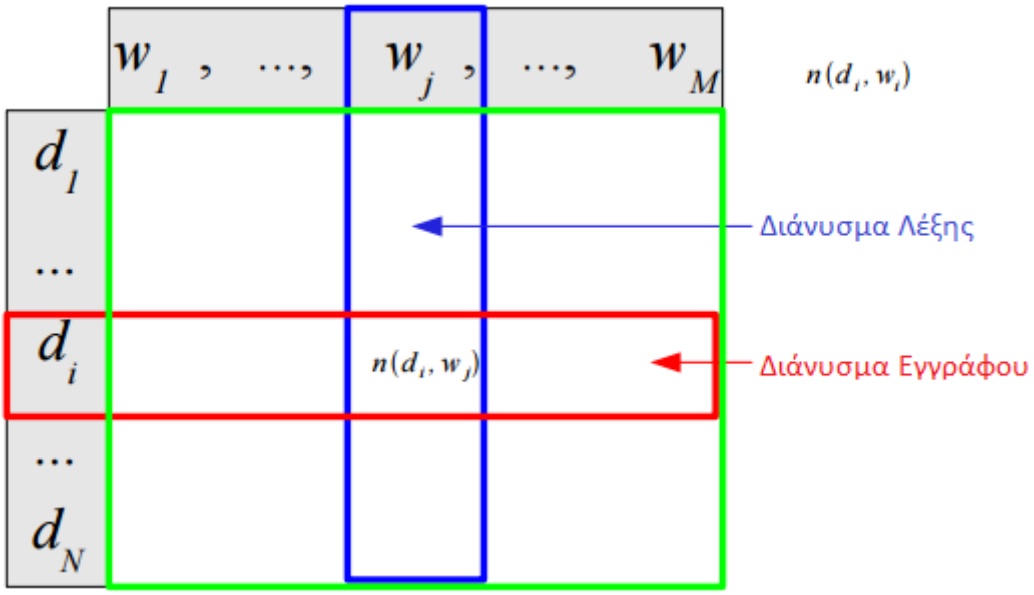
Από τα διάφορα latent μοντέλα που έχουν αναπτυχθεί, επιλέγεται η Πιθανοτική Λανθάνουσα Σημασιολογική Ανάλυση - **Probabilistic Latent Semantic Analysis (pLSA)** που αναπτύχθηκε το 1999 από τον Th. Hofmann. Κύριος στόχος του μοντέλου είναι να διαμορφώσει πληροφορίες συνεμφάνισης (co-occurrence information) κάτω από ένα πιθανολογικό πλαίσιο, προκειμένου να ανακαλύψει την υποβόσκουσα σημασιολογική δομή των στοιχείων.

Στο PLSA διακρίνουμε τα δεδομένα μας σε τρία σύνολα μεταβλητών:

- **Έγγραφα**: $d \in D = \{d1, \cdots, dN\}$, έστω N ο αριθμός τους,
- **Λέξεις**: $w \in W = \{w1, \cdots, wM\}$, έστω M είναι ο αριθμός τους
- **Θέματα**: $z \in Z = \{z1, \cdots, ZK\}$, έστω K αριθμός τους (στην περίπτωση μας K=11)

Τα δεδομένα εκπαίδευσης (training data),δηλαδή το σύνολο των εγγράφων (εικόνων από Google Search), μπορεί να αναπαρασταθεί με έναν πίνακα συνεμφάνισης (co-occurrence matrix) που υποδεικνύει τον αριθμό των φορών που κάθε λέξη εμφανίζεται σε κάθε έγγραφο. Στόχος του pLSA είναι κάνοντας χρήση αυτού του πίνακα, να εξάγει το θέμα (χρώμα) και να ερμηνεύσει τα

έγγραφα (εικόνες) ως μείγμα θεμάτων. Ο πίνακας αυτός έχει την ακόλουθη μορφή:

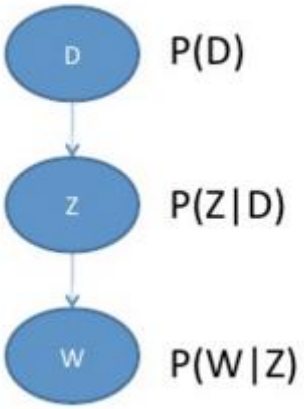


***Εικόνα 2.20**: Πίνακας συνεμφάνισης (Co-occurrence matrix)*

Κάθε σειρά αντιστοιχεί σε ένα διάνυσμα έγγραφο d_i και κάθε στήλη σε ένα διάνυσμα λέξη w_j. Για το ζεύγος παρατήρηση (d,w), ορίζουμε ένα παραγωγικό μοντέλο (generative model), δηλαδή ένα μοντέλο που εκφράζει το πώς τα παρατηρούμενα δεδομένα έχουν δημιουργηθεί με βάση τις κρυφές παραμέτρους. Η κοινή κατανομή (joint distibution) του ζεύγους παρατηρήσεων P(w,d) μπορεί να προκύψει από τα ακόλουθα παραγωγικά μοντέλα που είναι στατιστικά ισοδύναμα:

1. Παραγωγή ξεκινώντας από το έγγραφο

Το γραφικό μοντέλο της **Εικόνας 2.21** υποδεικνύει τα βήματα που ακολουθούνται. Αρχικά, επιλέγεται ένα έγγραφο d_n με πιθανότητα P(d), στην συνέχεια από αυτό το έγγραφο ένα θέμα με πιθανότητα P(z|d) και τέλος μια λέξη από αυτό το θέμα P(w|z).



Εικόνα 2.21*: Γραφική αναπαράσταση του πρώτου μοντέλου παραγωγής*

Με βάση τον νόμο αλυσίδα (chain rule), γράφουμε:

$$P(d_i, w_j) = \sum_{z \in Z} P(d_i)P(z_k|d_i)P(w_j|z_k) \quad (2.11)$$

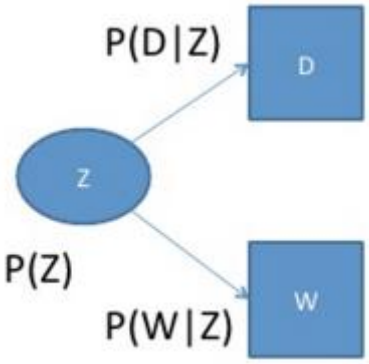
ή

$$P(d_i, w_j) = P(d_i)P(w_j|d_i) \quad (2.12)$$,όπου

$$P(w_j|d_i) = \sum_{z \in Z} P(w_j, z_k|d_i) = \sum_{z \in Z} P(w_j|z_k)\, P\,(z_k|d_i)$$

2.Παραγωγή ξεκινώντας από το θέμα

Αρχικά επιλέγουμε το θέμα d_n με πιθανότητα P(d), στην συνέχεια ένα έγγραφο, δεδομένου του θέματος, με πιθανότητα P(d|z) και μια λέξη από αυτό το θέμα P(w|z).



Εικόνα 2.22: *Γραφική αναπαράσταση του δεύτερου μοντέλου παραγωγής*

Με βάση τον νόμο αλυσίδα (chain rule) προκύπτει:

$$P(d_i, w_j) = \sum_{z \in Z} P(z_k) P(w_j | z_k) P(d_i | z_k)$$

Ας υποθέσουμε ότι επιλέγουμε τον πρώτο τρόπο παραγωγής. Τότε από την **Εξίσωση 2.12** έχουμε

$$P(w_j | d_i) = \frac{n(d_i, w_j)}{\sum n(d_i, w_j)} = \sum_{z \in Z} P(w_j | z_k) \, P(z_k | d_i)$$

Στόχος μας είναι να βρούμε τις παραμέτρους του μοντέλου που μεγιστοποιούν την πιθανότητα καταγραφής,(log likelihood), δηλαδή να μεγιστοποιήσει την μέση προβλεπόμενη πιθανότητα για τις παρατηρούμενες εμφανίσεις λέξεων. Με άλλα λόγια, σκοπός μας είναι να μάθουμε τις μη παρατηρήσιμες πιθανότητες P(w|z), P(d|z) χρησιμοποιώντας ως μέτρο σύγκρισης τις παρατηρούμενες P(w|d). Ορίζουμε την **πιθανοφάνεια (likelihood)** ως

$$L = \prod_{d \in D} \prod_{w \in W} P(d_i, w_j)^{n(d_i, w_j)}$$

και την **λογαριθμική πιθανοφάνεια (log-likelihood)** ως

$$L = \sum_{d \in D} \sum_{w \in W} n(d_i, w_j) \log p(d_i, w_j)$$

Για τον προσδιορισμό των P(w|z), P(d|z) χρησιμοποιείται η μέθοδος Εκτίμησης-Μεγιστοποίησης (Expectation-Maximization). Ο αλγόριθμος ΕΜ αποτελείται από δύο βήματα:

1. Ε-βήμα: Υπολογισμός προσδοκίας (= εκ των υστέρων πιθανότητες) για τις λανθάνουσες μεταβλητές λαμβάνοντας υπόψιν τις παρατηρήσεις χρησιμοποιώντας τις τρέχουσες εκτιμήσεις των παραμέτρων

$$P(z_k | d_i, w_j) = \frac{P(w_j, z_k | d_i)}{P(w_j | d_i)} = \frac{P(w_j | z_k, d_i) P(z_k | d_i)}{P(w_j | d_i)}$$

$$= \frac{P(w_j | z_k) P(z_k | d_i)}{\sum_{z \in Z} P(w_j | z_k) P(z_k | d_i)}$$

2. Μ-βήμα: Ενημέρωση παραμέτρων P(w|z) ,P(d|z) έτσι ώστε η καταγραφή πιθανοφάνειας (**log L**) να αυξάνεται χρησιμοποιώντας τις εκ των υστέρων πιθανότητες του Ε-βήματος. Τα P(d), n(d), n(d,w) μπορούν άμεσα να εκτιμηθούν από τα δεδομένα.

$$P(w_j|z_k) = \frac{\sum_{i=1}^{N} n(d_i, w_j) P(z_k|d_i, w_j)}{\sum_{m=1}^{M} \sum_{i=1}^{N} n(d_i, w_m) P(z_k|d_i, w_m)}$$

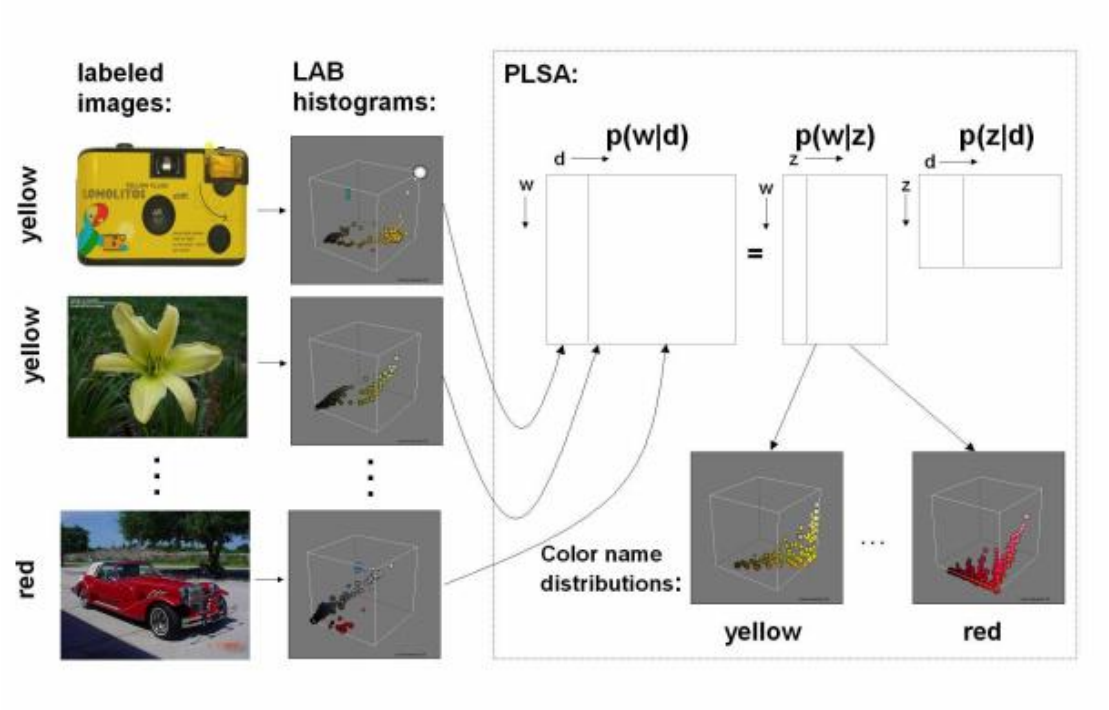
$$P(z_k|d_i) = \frac{\sum_{i=1}^{M} n(d_i, w_j) P(z_k|d_i, w_j)}{n(d_i)}$$

Τα δύο αυτά βήματα επαναλαμβάνονται μέχρι να υπάρξει σύγκλιση ,δηλαδή η λογαριθμική πιθανοφάνεια σταματά να αλλάζει ή αλλάζει λίγο.

Διαφοροποιήσεις του μοντέλου στην επεξεργασία των εικόνων

Έχουμε αναφέρει πως στο πλαίσιο της επεξεργασίας εικόνων, γίνεται αντιστοίχιση των **λέξεων** της επεξεργασίας κειμένου και των **RGB** τιμών μιας εικόνας. Ωστόσο, η εικόνα κατα την εκμάθηση χρωματικών ετικετών μετατρέπεται στον χρωματικό χώρο L*a*b*. Συγκεριμένα οι τιμές της εικόνας διακριτοποιούνται σε ένα λεξιλόγιο μέσω της ανάθεσης με κυβική παρεμβολή σε ένα 10x20x20 πλέγμα στον χώρο L*a*b*. Στη συνέχεια μια εικόνα αντιπροσωπεύεται από ένα ιστόγραμμα που δείχνει πόσα εικονοστοιχεία έχουν εκχωρηθεί σε κάθε bin του ιστογράμματος. Έτσι, με τον όρο **λέξη** εννοούνται τα **bins** του ιστογράμματος.

Στην **Εικόνα 2.23** παρατηρούμε για τρεις διαφορετικές εικόνες πως γίνεται η ανάλυση pLSA. Στην δεύτερη στήλη παρατηρούμε τα ιστογράμματα στον L*a*b * χώρο. Ανάλογα με τον αριθμό των εικονοστοιχείων που αντιστοιχούνται σε κάθε λέξη δηλαδή bin του ιστογράμματος, μεταβάλλεται και το μέγεθος του bin στην κυβική αναπαράσταση. Τέλος, παρατηρούμε την ανάλυση του πίνακα p(w|d) στις επιμέρους παραμέτρους p(w|z), p(z|d) οι οποίες προσδιορίζονται από τον αλγόριθμο EM. Για τον σκοπό της ανάκτησης χρωματικών ετικετών είναι προφανές πως στο στάδιο της εκπαίδευσης μας ενδιαφέρει η πιθανότητα p(w|z) που εκφράζει ακριβώς αυτή την κατανομή της χρωματικής ετικέτας στον L*a*b* χώρο, με άλλα λόγια ποια κομμάτια του χώρου αναγνωρίζονται ως κίτρινα, κόκκινα κλπ.



Εικόνα 2.23*: Παρουσίαση PLSA μοντέλου στην επεξεργασία εικόνων*

Στο **«*Color naming for real world applications*»** προτείνονται δύο σημαντικές βελτιώσεις του μοντέλου:

1.Χρήση των επιγραφών των εικόνων

Το γεγονός ότι οι εικόνες προέρχονται από αναζήτηση στο Google Image Search εξασφαλίζει πως οι εικόνες συνοδεύονται από κάποια **επιγραφή (label).** Το χρώμα που αντιστοιχεί στην επιγραφή της εικόνας μπορεί να χρησιμοποιηθεί ως a-priori πληροφορία, υποθέτοντας ότι έχει υψηλότερη συχνότητα σε σχέση με άλλα χρώματα. Το κλασικό μοντέλο pLSA δεν εκμεταλλεύεται τις ετικέτες των εικόνων με αποτέλεσμα η σύγκλιση να μην είναι βέβαιη. Με την χρήση των επιγραφών των εικόνων αυξάνονται τα περιθώρια σύγκλισης.

Συγκεκριμένα, οι επιγραφές καθορίζουν εκ των προτέρων την κατανομή που σχετίζεται με τη συχνότητα των θεμάτων (ονόματα χρωμάτων) στα έγγραφα δηλ. το μέγεθος p(z|d). Η πιθανότητα p(z|d) θεωρείται πως έχει προκύψει από **κατανομή Dirichlet** με παράμετρο α_{ld} όπου ld είναι η επιγραφή του εγγράφου d (document label) . Το α_{ld} είναι ένα διάνυσμα με μήκος K (αριθμός χρωμάτων) που φέρει τιμή μεγαλύτερη της μονάδας για κάθε χρώμα που περιλαμβάνεται στην επιγραφή της εικόνας και μονάδα για όσα χρώματα δεν περιλαμβάνονται στην επιγραφή, δηλαδή:

$$\alpha_{ld}(z) = c \geq 1, \text{ για } z = ld$$
$$\alpha_{ld}(z) = 1, \text{ αλλιώς}$$

Μεταβάλλοντας την τιμή c , ελέγχεται το μέγεθος επιρροής της επιγραφής της εικόνας πάνω στην κατανομή P (z|d). Η αλλαγή που επέρχεται στο pLSA μοντέλο μετά την επιρροή των επιγραφών είναι:

$$p(z|d) \propto (a_{ld}(z) - 1) + \sum_{w} n(d,w)p(z|d,w)$$

Προφανώς για $\alpha_{ld}(z) =1$ έχουμε το κλασσικό pLSA μοντέλο.

2.Επικράτηση ενός κυρίαρχου χρώματος

Σκοπός της δεύτερης αλλαγής του μοντέλου είναι η προσαρμογή του ώστε να επιβάλλει **μονοκόρυφη κατανομή πιθανότητας** (δηλαδή μια μέγιστη κορυφή) για την εκτιμούμενη p(z|w). Αυτό σημαίνει πως δοσμένης μιας λέξης w, θέλουμε μόνο ένα κυρίαρχο χρώμα z. Αυτό γίνεται εφικτό μέσω της μεθόδου **greyscale reconstruction**, ένα ιδιαίτερα χρήσιμο εργαλείο που παρέχεται από την επιστήμη της μορφολογίας και ανήκει στους γεωδαιτικούς τελεστές (geodesic operators). Εάν υποθέσουμε πως έχουμε δυο εικόνες greyscale, την εικόνα δείκτη (marker) και την εικόνα μάσκα (mask) η ανασυγκρότηση εξάγει τις κορυφές της μάσκας οι οποίες υποδεικνύονται από την εικόνα δείκτη. Αυτό γίνεται με συνεχόμενες διαστολές (dilations) της εικόνας δείκτη κάτω από την εικόνα μάσκα μέχρι να υπάρξει σταθερότητα. Μια ανάλογη διαδικασία εφαρμόζεται για τις κατανομές χρώματος p(z|w), για τις οποίες υπολογίζουμε την μονοκόρυφη εκδοχή τους, έστω ρ_{ζ}^{mz}, και επιβάλλουμε η διαφορά τους να είναι αρκούντως μικρή προσθέτοντας έναν ακόμη όρο στην λογαριθμική πιθανοφάνεια ο οποίος λειτουργεί ως όρος κανονικοποίησης.

$$L = \sum_{d\in D}\sum_{w\in W} n(d_i, w_j) \log p(d_i, w_j) - \gamma \sum_{z\in Z}\sum_{w\in W} (p(z_k|w_j) - \rho_z^{mz}(w_j))^2$$

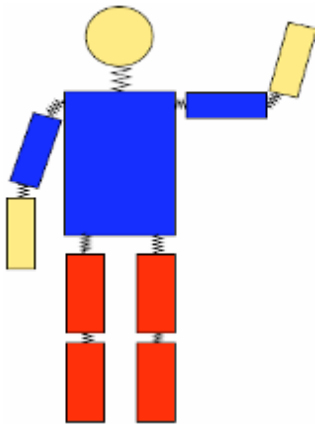
Ο τρόπος ανακατασκευής της εικόνας γίνεται στον τρισδιάστατο L*a*b * χώρο χρησιμοποιώντας στοιχεία συνεκτικότητας 26 συνδέσεων (26-connected structuring element). Αυτό σημαίνει ότι δύο λέξεις (bins) του τρισδιάστατου ιστογράμματος θεωρούνται συνδεδεμένες εάν ακουμπούν κάποιες από τις γωνίες, ακμές ή πλευρές τους. Τα στοιχεία αυτά λειτουργούν ως δείκτες (markers) για το πού πρέπει να αναζητηθεί μέγιστο και για αυτό τοποθετούνται στο κέντρο μάζας της κατανομής p(z|w). Μετά την ανακατασκευή, επιστρέφεται η κατανομή με μια κορυφή, πράγμα που σημαίνει πως δοσμένης μιας λέξης υπάρχει ένα χρώμα που αντιστοιχεί με μέγιστη πιθανότητα σε αυτή .

Μετά το τέλος της εκπαίδευσης έχουμε ένα μοντέλο που έχει διδαχθεί να αναγνωρίζει τα χρώματα και να τους αποδίδει ετικέτες. Δεν μένει παρά να δοκιμάσουμε την απόδοση του μοντέλου στο σύνολο των εικόνων που εξετάζουμε. Αυτό γίνεται στο πειραματικό μέρος της εργασίας.

2.3 Εκτίμηση Πόζας (Pose Estimation)

Οι προηγούμενες ενότητες είχαν ως σκοπό να εισάγουν έννοιες και να παρουσιάσουν μεθόδους σχετικές με την περιγραφή χαρακτηριστικών των εικόνων. Η οποιαδήποτε περιγραφή όμως θα πρέπει να περιορίζεται σε ένα συγκεκριμένο κομμάτι της εικόνας, σε αυτό που μας ενδιαφέρει ως στοιχείο ανάκτησης, στην περίπτωση μας τα ρούχα των μοντέλων. Άρα, κρίνεται πρωταρχικής σημασίας ο εντοπισμός τόσο της γενικής θέσης του μοντέλου όσο και ιδανικά των ξεχωριστών τμημάτων του σώματος, όπως κορμός και πόδια, ώστε η περιγραφή να γίνεται για κάθε κομμάτι ρουχισμού ξεχωριστά.

Σκοπός της ενότητας αυτής είναι να παρουσιάσει τη **μέθοδο εκτίμησης πόζας** των Yang και Ramanan που παρουσιάζεται στο σύγγραμμα «**Articulated Human Detection with Flexible Mixtures-of-Parts»** και αποτελεί ένα μοντέλο προσέγγισης των αρθρώσεων του σώματος. Η τεχνική που ακολουθούν βασίζεται στα **pictorial strusture μοντέλα** που εισήγαγαν οι Fischler και



Εικόνα 2.24: *Αναπαράσταση μοντέλου παραμορφώσιμης διάταξης*

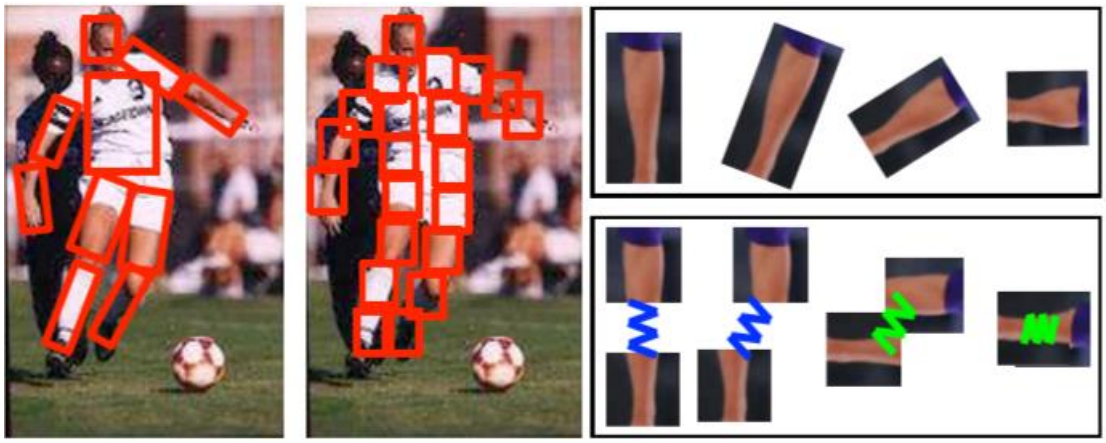
Elschlager. Η βασική ιδέα αυτών των μοντέλων είναι να γίνεται αναπαράσταση ενός αντικειμένου από ένα σύνολο τμημάτων διατεταγμένα σε παραμορφώσιμη διάταξη. Η εμφάνιση κάθε τμήματος διαμορφώνεται

χωριστά, και η παραμορφώσιμη διάταξη αναπαριστάνεται ως συνδέσεις ελατηρίων μεταξύ των τμημάτων.

Η δυσκολία εντοπισμού της πόζας σώματος , δηλαδή των διαφορετικών μερών που το απαρτίζουν, κρίνεται ιδιαίτερα δύσκολη εάν αναλογιστούμε πως μεταβάλλεται η εμφάνιση των άκρων του σώματος ανάλογα με τα ρούχα καθώς και λόγω αλλαγών οπτικής γωνίας. Μερικές από τις δυσκολίες της εκτίμησης πόζας παρουσιάζονται στην παρακάτω εικόνα.

Εικόνα 2.25: *Δυσκολίες κατά την εκτίμησης πόζας*

Η διαφορά της μεθόδου των Yang-Ramanan με σκοπό την καλύτερη αντιμετώπιση τέτοιων θεμάτων παρουσιάζεται στην **Εικόνα 2.26**. Εάν θεωρήσουμε πως η πόζα σωμάτων προσεγγίζεται από μια οικογένεια προτύπων (templates) που αλλάζουν μορφή, τα οποία διακρίνονται με κόκκινο χρώμα, οι Yang και Ramanan πρότειναν ένα μείγμα μικρότερων και μη προσανατολισμένων τέτοιων προτύπων.



Εικόνα 2.26: *Σύγκριση μεθόδου των Yang-Ramanan με προγενέστερες.*
και μείγμα προτύπων που συνδέονται με ελατήρια.

Εάν υποθέσουμε ότι ένα πρότυπο μπορεί να υποστεί μόνο γραμμικές παραμορφώσεις, δεδομένου της θέσης ενός εικονοστοιχείου x σε αυτό το πρότυπο η νέα θέση μπορεί να γραφτεί ως

$$w(x) = (I + \Delta A)x+b$$

όπου $A = I + \Delta A$ η γραμμική παραμόρφωση και b οποιαδήποτε μετατόπιση. Εάν $s(x) = w(x) - x$ η αλλαγή θέσης του x μετά την παραμόρφωση και $x + \Delta x$ ένα γειτονικό σημείο η αλλαγή της δικής του θέσης γράφεται

$$\begin{aligned} s(x + \Delta x) &= w(x + \Delta x) - (x + \Delta x) \\ &= (I + \Delta A)(x + \Delta x) + b - x - \Delta x = s(x) + \Delta A \Delta x \end{aligned}$$

Εάν το γινόμενο $\Delta A \Delta x$ είναι αμελητέο, τότε τα σημεία μπορούν να θεωρηθούν ότι ανήκουν στο ίδιο τμήμα. Αυτό συμβαίνει όταν η νόρμα του Δx είναι αμελητέα ή όταν η ορίζουσα ΔA είναι μικρή. Στην ουσία, μικρή ορίζουσα ΔA σημαίνει πως το μοντέλο αρθρώσεων θα αποτελείται από πρότυπα με πολύ μικρό εύρος μεταβολών προσανατολισμού και προοπτικής. Οι Yang και Ramanan ωστόσο επιλέγουν να διατηρήσουν μικρές τιμές Δx, να χρησιμοποιήσουν δηλαδή μικρότερα σε μέγεθος πρότυπα, αντί να κρατήσουν μικρό το εύρος των παραμορφώσεων.

Το μοντέλο μπορεί να θεωρηθεί ως ένα πρόβλημα ονομασίας γραφήματος (graph labeling) όπου κόμβοι είναι τα διάφορα τμήματα και ακμές οι συνδέσεις μεταξύ τους. Για κάθε απόδοση 'ονόματος' στους κόμβους ,δηλαδή την διαμόρφωσης των τμημάτων που αποδίδεται, υπάρχει κάποιο σκορ που δίνεται από την **Εξίσωση 2.12.**

$$S(t) = \sum_{i \in V} {b_i}^{t_i} + \sum_{i,j \in E} b_{ij}^{t_i, t_j} \quad (2.12)$$

όπου $li = (x, y) \in \{1, \dots L\}$, η θέση του τμήματος i όπου $i \in \{1, \dots K\}$ και $ti \in \{1, \dots T\}$ είναι ο αποκαλούμενος «τύπος» του μέρους i, και αφορά έναν συνδυασμό ιδιοτήτων όπως για παράδειγμα ο προσανατολισμούς ενός τμήματος (π.χ, κάθετο έναντι οριζόντιου προσανατολισμένου χεριού) ή σημασιολογικές κλάσεις (ανοιχτό έναντι κλειστού χεριού). Ο δείκτης t δηλώνει το σύνολο τέτοιων τύπων για κάθε τμήμα i , $t = \{t1, \dots tK\}$. Το πρώτο μέρος του αθροίσματος, ανάλογα από την τιμή του, ευνοεί ή όχι έναν συγκεκριμένο τύπο για κάθε τμήμα i ενώ ο δεύτερος όρος την συνύπαρξη δύο τμημάτων με

τύπους ti, tj. Τελικό μοντέλο θα είναι εκείνο με τις ονομασίες των τμημάτων για τις οποίες προκύπτει μεγαλύτερο σκορ. Η σχέση μπορεί να γραφτεί ως:

$$S(I, l, t) = S(t) + \sum_{i \in V} w_i^{t_i} \cdot \varphi(I, l_i) + \sum_{ij \epsilon E} w_{ij}^{t_i, t_j} \cdot \psi(l_i - l_j) \quad (2.13)$$

Ο πρώτος όρος εκφράζει την a priori πιθανότητα των τύπων, ενώ ο δεύτερος όρος αποτελεί ένα μοντέλο εμφάνισης (appearance model) που υπολογίζει τοπικά την βαθμολογία τοποθέτησης ενός προτύπου (template), έστω w με τύπο t_i στην θέση l_i ,όπου έχουμε υπολογίσει ένα διάνυσμα χαρακτηριστικών $\varphi(I, l_i)$ (συγκεκριμένα ένας περιγραφέας HOG). Ο τρίτος όρος εκφράζει ένα μοντέλο παραμόρφωσης που ελέγχει τι αποτέλεσμα θα φέρει η σχετική τοποθέτηση του τμήματος i και του j δοκιμάζοντας διαφορετικά ελατήρια μεταξύ τους. Ο όρος $\psi\ (li - lj)\ =\ [dx\ dy\ dx^2 dy^2\ ]$, εκφράζει την σχετική θέση των μερών, ενώ ο όρος $w_{i,j}$ παραμετροποιεί την μορφή των ελατηρίων μεταξύ τους. Οι Yang και Ramanan πρότειναν η Εξίσωση 2.13 να απλοποιηθεί σύμφωνα με τη σχέση:

$$w_{ij}^{t_i, t_j} = w_{ij}^{t_i}$$

Δηλαδή, οι σχετικές θέσεις του κάθε τμήματος σε σχέση με το γονέα του να εξαρτάται από τον τύπο του τμήματος, αλλά όχι από τον τύπο του γονέα του. Τα συμπεράσματα θα προκύψουν μετά από την μεγιστοποίηση της σχέσης 2.13. Υποθέτοντας επιβλεπόμενη μάθηση (supervised learning) το μοντέλο εκπαιδεύεται δοσμένου ενός συνόλου θετικών και αρνητικών παραδειγμάτων Συγκεκριμένα, χρησιμοποιείται ένα μοντέλο Structured SVM για την πρόβλεψη δομημένων αντικειμένων. Στην περίπτωση μας η δομημένη περιοχή παραγωγής είναι το σύνολο όλων των πιθανών δέντρων που μπορεί να προκύψουν από τη σύνδεση των τμημάτων της πόζας.

3 Πειραματικό Μέρος

Στο κεφάλαιο αυτό θα παρουσιαστούν πειράματα με σκοπό την αξιολόγηση των μεθόδων περιγραφής των εικόνων μόδας και τον εντοπισμό σημείων ενδιαφέροντος σε αυτές καθώς και τα ποσοστά επιτυχίας της ανάκτησης ανάλογα με τον αριθμό των εικόνων που ανακτήθηκαν.

3.1 Οργάνωση πειραμάτων

Προκειμένου να επιτευχθεί όσο το δυνατόν καλύτερα η σωστή περιγραφή και η ανάκτηση εικόνων μόδας μέσω των εργαλείων που μελετήθηκαν στο θεωρητικό κομμάτι της εργασίας, γίνεται χρήση ενός συνόλου 240 εικόνων από επιδείξεις μόδας (catwalks). Στις εικόνες αυτές έχουμε διάφορες συνθήκες φωτισμού. Συχνά μάλιστα εντοπίζονται σκιές (shadowing). Επιπλέον οι εικόνες προέρχονται από διαφορετικές κάμερες. Σκοπός μας είναι τέτοιες διαφοροποιήσεις να μην επηρεάσουν το τελικό αποτέλεσμα και να επιτευχθεί σωστό ταίριασμα των εικόνων.

Αρχικά, όλες οι περιγραφές γίνονται στο σύνολο της εικόνας (**Ενότητα 3.3**) ενώ σταδιακά επιδιώκεται η απομόνωση μόνο των σημείων ενδιαφέροντος δηλαδή της ενδυμασίας του μοντέλου (**Ενότητα 3.4**). Τα πειράματα συνοψίζονται στην **Ενότητα 3.5** με κάποια παραδείγματα επιτυχούς ανάκτησης.

3.2 Προγραμματιστικά εργαλεία

Για την υλοποίηση των πειραμάτων γίνεται χρήση MATLAB καθώς και της βιβλιοθήκης VLFEAT που υλοποιεί αλγορίθμους της όρασης υπολογιστών.

3.3 Πειράματα στο σύνολο της εικόνας

Τα πειράματα που ακολουθούν πραγματοποιούνται στο σύνολο της εικόνας και δεν επιδιώκεται ο εντοπισμός του μοντέλου και της ενδυμασίας του αποκλειστικά, ούτε η κατάτμηση του ρουχισμού στα επιμέρους τμήματα του (σακάκι,παντελόνι) και άρα η εύρεση εικόνων που ταιριάζουν ανά τμήμα αλλά εικόνες επιδείξεων που παρουσιάζουν ομοιότητες στο σύνολο τους.

3.3.1 Τοπική ανάκτηση με μορφολογικούς περιγραφείς SIFT

Ο όρος τοπική ανάκτηση σημαίνει πως η ανάκτηση παρόμοιων εικόνων θα βασιστεί σε χαρακτηριστικά σημείων που εντοπίζονται τοπικά στην εικόνα. Αυτά τα σημεία στην περίπτωση του SIFT είναι τα σημεία-κλειδια που περιγράψαμε στην **Ενότητα 2.1.1**.

Αρχικά, χρησιμοποιούμε την βιβλιοθήκη VLFEAT μέσω της οποίας είναι διαθέσιμος ο περιγραφέας SIFT με της εντολή vl_sift. Η εντολή αυτή επιστρέφει τα σημεία-κλειδιά και την αντίστοιχη περιγραφή τους.

Ας υποθέσουμε πως θέλουμε να ταιριάξουμε την **Εικόνα 3.1** ,η οποία αποτελεί την αρχική εικόνα αναζήτησης (query image), με κάποια από τις υπόλοιπες 239 του συνόλου των εικόνων μας.

Εικόνα 3.1: *Αρχική Εικόνα(Query Image)*

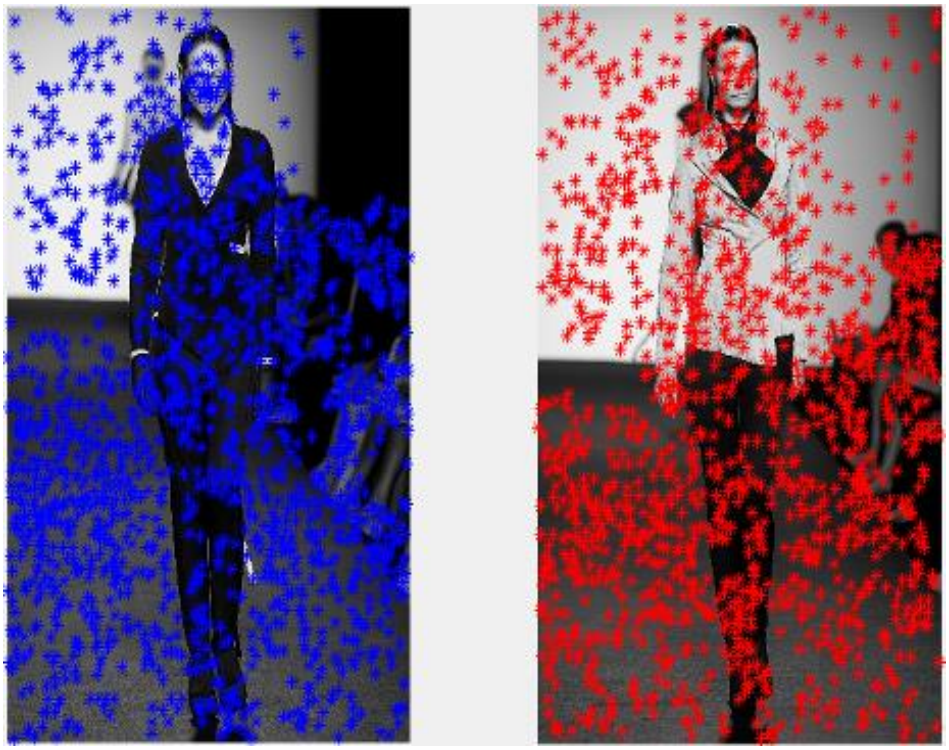
Αφού γίνει περιγραφή όλων των σημείων κλειδιών των υπόλοιπων εικόνων, οι περιγραφές αυτές συγκρίνονται με τις περιγραφές των σημείων της αρχικής εικόνας query. Εικόνα ταίρι θα θεωρηθεί αυτή για την οποία βρέθηκαν περισσότερα σημεία όμοια με αυτά της εικόνας query. Αυτή η διαδικασία γίνεται με την εντολή vl_ubcmatch που βρίσκει τα σημεία-κλειδιά μεταξύ των δύο εικόνων που ταιριάζουν καθώς και την ευκλείδεια απόσταση τους. Η συνάρτηση αυτή απορρίπτει ένα ταίριασμα το οποίο θεωρείται αμφίβολο με βάση τον τρόπο που περιγράφεται από τον Lowe. Εάν υποθέσουμε πως ψάχνουμε για ένα σημείο της εικόνας query το πλησιέστερο σε κάποια άλλη εικόνα , δεν αρκούμαστε στο να βρούμε το πρώτο κοντινότερο σημείο με την

άλλη εικόνα, αλλά και το δεύτερο. Μόνο αν η απόσταση του πρώτου και του δεύτερου πλησιέστερου σημείου είναι αρκούντως μικρή, θεωρείται το ταίριασμα των σημείων αποδεκτό, διαφορετικά απορρίπτεται. Συγκεκριμένα, ένας περιγραφέας D1 θα ταιριάξει με ένα περιγραφέα D2 μόνο αν η απόσταση d (D1, D2) πολλαπλασιαζόμενη επί την τιμή κατωφλίου(στο πείραμα μας ίση με 1.5) δεν είναι μεγαλύτερη από την απόσταση του D1 με όλους τους άλλους περιγραφείς

***Εικόνα 3.2**: Πλησιέστερη Εικόνα*

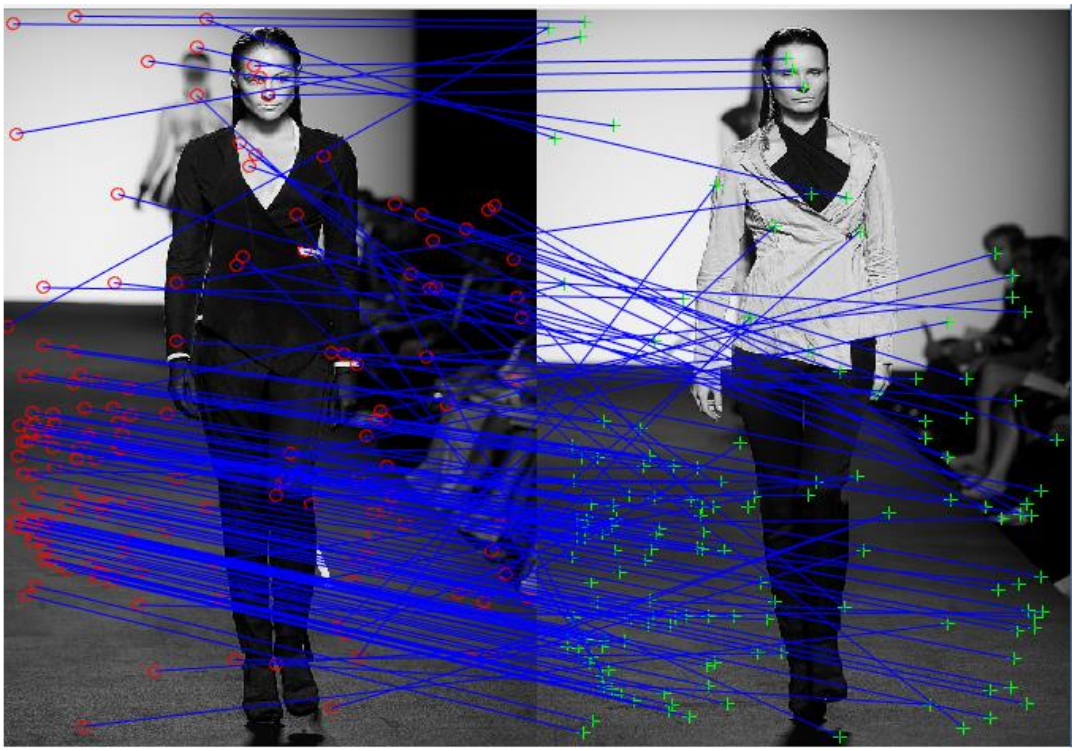
Με βάση αυτή τη μέθοδο προκύπτει η **Εικόνα 3.2** ως πρώτη πλησιέστερη.

Αν και το παραπάνω αποτέλεσμα φαίνεται ικανοποιητικό μια διερεύνηση των σημείων που τελικά ταίριαξε ο αλγόριθμος προκειμένου να προβεί στο παραπάνω συμπέρασμα δείχνουν το αντίθετο. Στην **Εικόνα 3.3** , με μπλε και κόκκινο χρώμα παρατηρούμε τα σημεία-κλειδιά που επέστρεψε ο SIFT για τις δύο εικόνες. Από το σύνολο αυτών των σημείων τελικά βρίσκουν ταίρι μόνο αυτά της **Εικόνας 3.4**. Παρατηρούμε πως αυτά εντοπίζονται κυρίως στο φόντο (background) της εικόνας. Άρα, η φαινομενικά σωστή επιλογή δεν έγινε με βάση τα χαρακτηριστικά του ρουχισμού. Επιλέον το γεγονός ότι οι δύο εικόνες προέρχονται από την ίδια επίδειξη μόδας, με τα ίδια σκηνικό ως φόντο εξηγεί που βασίστηκε το ταίριασμα.



Εικόνα 3.3: *Σημεία-Κλειδιά που εντοπίζει ο SIFT*

Εικόνα 3.4: *Σημεία- Κλειδιά που βρίσκουν πλησιέστερο ταίρι*



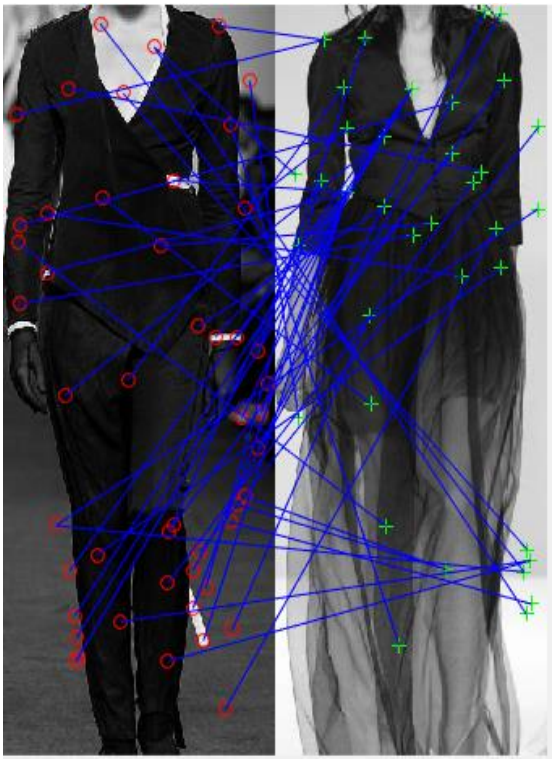
Εικόνα 3.5: *Αντιστοίχιση σημείων-κλειδιών που βρίσκουν ταίρι*

Για να περιορίσουμε όσο γίνεται την αναζήτηση σημείων-κλειδιών στην περιοχή που εντοπίζεται ο ρουχισμός του μοντέλου κόβουμε ένα κομμάτι του φόντου. Η διαδικασία αυτή γίνεται βασιζόμενη στην υπόθεση ότι το μοντέλο και άρα τα ρούχα βρίσκονται περίπου στο κέντρο της εικόνας. Αργότερα ,θα μελετηθούν αυτόματοι τρόποι αναζήτησης του σώματος και άρα του ρούχου. Για το σκοπό αυτού του πειράματος ,όμως, η εντολή imcrop του MATLAB αρκεί για να δημιουργήσει ένα νέο σύνολο 240 εικόνων στις οποίες έχει αφαιρεθεί το ίδιο υπόβαθρο εκατέρωθεν του μοντέλου (**Εικόνα 3.6**).Τα σημεία που απομένουν, ωστόσο, δεν παρέχουν αρκετή πληροφορία για να πραγματοποιηθεί σωστή αντιστοίχιση.

Στην **Εικόνα 3.7** ,παρατηρούμε ότι μετά την αφαίρεση τμήματος των αρχικών εικόνων, η πλησιέστερη εικόνα που εντοπίζεται αλλάζει. Ωστόσο, ούτε αυτό το ταίριασμα βασίζεται σε σημεία του ρουχισμό που φαίνεται να ταιριάζουν μεταξύ τους. Ιδανικά θα θέλαμε η αντιστοίχιση να είχε γίνει με βάση τα δύο όμοια άνω κομμάτια του ρουχισμού δηλαδή το σακάκι , δηλαδή τα άνω σημεία στο πρώτο σακάκι να είχαν αντιστοιχηθεί με τα άνω σημεία στο άλλο σακάκι της εικόνας.

Με βάση τα παραπάνω φανερώνεται έτσι το πρόβλημα περιγραφής εικόνων μόδας με βάση την μορφολογία τους, καθώς ως αντικείμενα τα ρούχα είναι ιδιαιτέρως ευμετάβλητα δεν έχουν δηλαδή μια σταθερή δομή. Για παράδειγμα ακόμα και το ίδιο φόρεμα μπορεί να αλλάξει εντελώς από άποψη σχήματος ανάλογα από το πως ποζάρει το μοντέλο που το φορά.

Εικόνα 3.6: *Αρχική και νέα εικόνα μετά την αφαίρεση φόντου*



Εικόνα 3.7: *Ανάκτηση πλησιέστερης εικόνας μετά την αφαίρεση φόντου*

Για το λόγο αυτό συνειδητοποιούμε πως τα μορφολογικά χαρακτηριστικά δεν αρκούν και προσανατολιζόμαστε στην αναζήτηση επιπλέον χαρακτηριστικών που μπορούν να βοηθήσουν στο ταίριασμα των εικόνων όπως είναι το χρώμα.

3.3.2 Τοπική Ανάκτηση με περιγραφείς χρώματος(Local Retrieval with color descriptors)

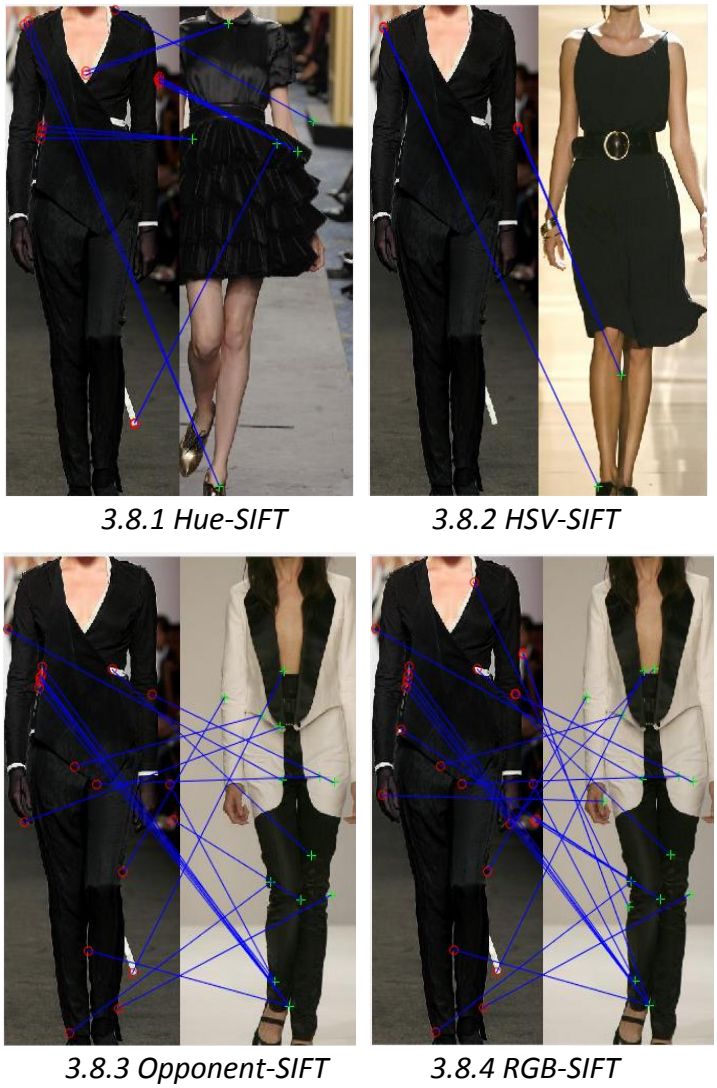
Οι περιγραφείς χρώματος κάνουν χρήση και του περιγραφέα SIFT αλλά και της πληροφορίας χρώματος που εμπεριέχεται στην εικόνα. Αναλυτικά ο κάθε περιγραφέας έχει παρουσιαστεί στην **Ενότητα 2.2.4**. Για τα πειράματα χρησιμοποιήθηκε το λογισμικό χρωματικών περιγραφέων που παρέχεται από τον Koen van de Sande. Αρχικά, χρησιμοποιείται ο ανιχνευτής σημείων Harris-Laplace για τον εντοπισμό πιθανών σημείων αμετάβλητης κλίμακας (scale-invariant). Στη συνέχεια επιλέγεται το υποσύνολο των σημείων για τα οποία οι συναρτήσεις LoG παρουσιάζουν μέγιστο σε μια κλίμακα. Οι περιγραφείς χρώματος της ενότητας υπολογίζονται στην περιοχή γύρω από τα σημεία. Το μέγεθος αυτής της περιοχής εξαρτάται από τη μέγιστη κλίμακα των συναρτήσεων LoG. Έτσι, επιστρέφεται για κάθε εικόνα ένα σύνολο περιγραφέων για τα σημεία της ,τους οποίους περιγραφείς συγκρίνουμε με την εντολή vl_ubcmatch του VLFEAT.Για την αρχική εικόνα αναζήτησης ,οι πλησιέστεροι γείτονες που επιστρέφονται καθώς και τα σημεία που βρήκαν ταίρι παρουσιάζονται στην **Εικόνα 3.8.** Από αυτή είναι ξεκάθαρο ότι τα σημεία που επιστρέφονται δεν αρκούν και πολλές φορές αντιστοιχούν σε σημεία εκτός ρουχισμού. Άρα, το οποιοδήποτε ταίριασμα δεν βασίζεται σε επαρκή και έγκυρα σημεία.

3.3.3 Τοπική Ανάκτηση με χρωματικές ροπές (Local Retrieval with color moments)

Ο Koen van de Sande προτείνει την χρήση χρωματικών ροπών με βάση το σύγγραμα του Mindru, δηλαδή γενικευμένες χρωματικές ροπές όπως αυτά ορίστηκαν στην **Ενότητα 2.2.3**.

Οι χρωματικές ροπές υπολογίζονται στα σημεία που εντοπίζει ο ανιχνευτής Harris-Laplace. Η ομοιότητα των σημείων καθορίζεται από την ευκλείδεια απόσταση των διανυσμάτων τους και το σύνολο των σημείων-ταίρι προκύπτουν από την εντολή vl_ubcmatch. Ανάλογα με το σύνολο αυτό, δηλαδή πόσο μεγάλο είναι, θα προκύψει και η πλησιέστερη εικόνα. Ωστόσο το αποτέλεσμα δεν είναι ικανοποιητικό και δεν επαρκεί για την εξαγωγή συμπερασμάτων. Αποδεικνύεται ότι η απόδοση του ταιριάσματος όταν οι περιγραφείς περιορίζονται σε συγκεκριμένα σημεία (σημεία που επιστρέφει ο

ανιχνευτής Harris) δεν είναι τόσο ικανοποιητική. Στο επόμενο στάδιο προσανατολιζόμαστε σε εξαγωγή αποκλειστικά χρωματικής πληροφορίας από το σύνολο της εικόνας. Η ανάκτηση εικόνων που ακολουθεί θα είναι δηλαδή ολική και θα γίνεται με βάση χρωματικά κριτήρια χωρίς να περιορίζεται σε σημεία κλειδιά.



3.8.1 Hue-SIFT *3.8.2 HSV-SIFT*

3.8.3 Opponent-SIFT *3.8.4 RGB-SIFT*

***Εικόνα 3.8** : Αντιστοίχηση σημείων χρωματικών περιγραφέων*

3.3.4 Ολική Ανάκτηση με χρωματικές ροπές (Global Retrieval with color moments)

Στο πείραμα αυτό οι ροπές χρώματος δεν υπολογίζονται σε συγκεκριμένα σημεία αλλά στο σύνολο της εικόνας. Συγκεκριμένα, χρησιμοποιούνται οι πρώτες τρεις κεντρικές ροπές (central moments) δηλαδή η μέση τιμή, τυπική απόκλιση και η ασυμμετρία. Αυτές υπολογίζονται και για τα τρία κανάλια αναπαράστασης του χρώματος. Εδώ θα περιοριστούμε στην RGB και HSV αναπαράσταση. Κάθε εικόνα χαρακτηρίζεται δηλαδή από εννέα τιμές, τις τρεις τιμές ροπών και για τα τρία κανάλια.

ΡΟΠΗ 1	$E_i = \sum_{j=1}^{N} \left(\frac{1}{N}\right) p_{ij}$	Ο μέσος όρος χρώματος της εικόνας
ΡΟΠΗ 2	$\sigma_i = \sqrt{\left(\frac{1}{N}\right)\left(\sum_{j=1}^{N} \left(p_{ij} - E_i\right)^2\right)}$	Η τυπική απόκλιση δηλαδή η τετραγωνική ρίζα της διακύμανσης της κατανομής
ΡΟΠΗ 3	$s_i = \sqrt[3]{\left(\frac{1}{N}\sum_{j=1}^{N} \left(p_{ij} - E_i\right)^3\right)}$	Το μέτρο του βαθμού της ασυμμετρίας στην κατανομή.

Πίνακας 3.1: *Πίνακας Ροπών*

Για κάθε εικόνα για την οποία θέλουμε να ανακτήσουμε όμοιες με αυτή υπολογίζουμε τις ροπές για αυτήν και για όλες τις εικόνες της βάσης δεδομένων και μετά από αυτό χρησιμοποιείται η ακόλουθη συνάρτηση για να υπολογιστεί ένα σκορ ομοιότητας μεταξύ της εικόνας ενδιαφέροντος με τις υπόλοιπες:

$$d_{nom}(H, I) = \sum_{i=1}^{r} w_{i1} \left|E_i^1 - E_i^2\right| + w_{i2} \left|\sigma_i^1 - \sigma_i^2\right| + w_{i3} \left|s_i^1 - s_i^2\right|$$

όπου H, I :οι χρωματικές κατανομές των δύο εικόνων που συγκρίνονται

i , r : ο δείκτης του καναλιού και r ο συνολικός αριθμός των καναλιών

$E_i^1, E_i^2, \sigma_i^1, \sigma_i^2, s_i^1, s_i^2$: οι πρώτες , δεύτερες και τρίτες ροπές αντίστοιχα για την πρώτη και δεύτερη εικόνα που συγκρίνονται

w_{i1}, w_{i2}, w_{i3}: οι συντελεστές-βάρη που καθορίζονται ανάλογα με το κάθε κανάλι χρώματος

Οι συντελεστές-βάρη καθορίζουν την σημασία που θέλουμε να προσδώσουμε σε κάθε κανάλι. Στα πειράματα, επειδή κυρίως μας ενδιαφέρει το χρώμα όταν υπολογίζονται οι χρωματικές ροπές στον RGB χώρο χρησιμοποιούμε τα ίδια βάρη για κάθε κανάλι ενώ για τον HSV χρωματικό χώρο δίνεται μεγαλύτερη βαρύτητα στο κανάλι που περιέχει την βασική χρωματική πληροφορία, δηλαδή το κανάλι Hue.

Με πιο απλά λόγια κάθε 3x3 πίνακας ροπών μιας εικόνας πολλαπλασιάζεται με έναν πίνακα συντελεστών :

$$w_{RGB} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \qquad w_{HSV} = \begin{bmatrix} 2 & 1 & 1 \\ 2 & 1 & 1 \\ 2 & 1 & 1 \end{bmatrix}$$

Ο συντελεστής 2 που χρησιμοποιείται στο κανάλι Hue λειτουργεί ως ποινή εφόσον αυξάνει την απόσταση δύο εικόνων όσο μεγαλύτερη είναι η διαφορά των ροπών του καναλιού απόχρωσης. Η παραπάνω διαδικασία υλοποιείται στο MATLAB για τα χρωματικά κανάλια RGB και HSV. Προκειμένου να υπάρξει ένα μέτρο αξιολόγησης των αποτελεσμάτων, δηλαδή των εικόνων που επιστρέφονται ως πλησιέστερες με βάση την απόσταση των ροπών, δημιουργήσαμε έναν πίνακα αληθείας (groundtruth table) μεγέθους 240x240 όπου για κάθε εικόνα προσδιορίζεται η χρωματική της ομοιότητα με τις υπόλοιπες ως εξής:

Εάν το μεγαλύτερο ποσοστό χρωμάτων του ρουχισμού που περιέχει αυτή η εικόνα, έστω i, είναι όμοιο με το μεγαλύτερο ποσοστό χρωμάτων μιας εικόνας, έστω j, δίνεται ένας συντελεστής 2 στο κελί (i, j) του πίνακα. Αυτό σημαίνει πως τα ρούχα των δύο εικόνων μοιάζουν χρωματικά πάρα πολύ ή απόλυτα. Εάν τα ρούχα των δύο εικόνων είναι χρωματικά όμοια κατά ένα ποσοστό της τάξης του 50% δίνεται η τιμή 1. Διαφορετικά, εάν τα ρούχα δεν μοιάζουν καθόλου χρωματικά ή ελάχιστα δίνεται η τιμή 0. Ο πίνακας αυτός που εκφράζει την χρωματική ομοιότητα όπως την αντιλαμβανόμαστε οπτικά θα χρησιμοποιηθεί για να αξιολογήσει τις εικόνες που επιστρέφονται. Συγκεκριμένα, ορίζονται τα **Κριτήρια ομοιότητας 2 και 1** όπου το **Κριτήριο 2** προϋποθέτει μεγάλη ομοιότητα μεταξύ των εικόνων (τιμή πίνακα αληθείας 2), ενώ το **Κριτήριο 1** σχετική ομοιότητα (τιμή πίνακα αληθείας 1)

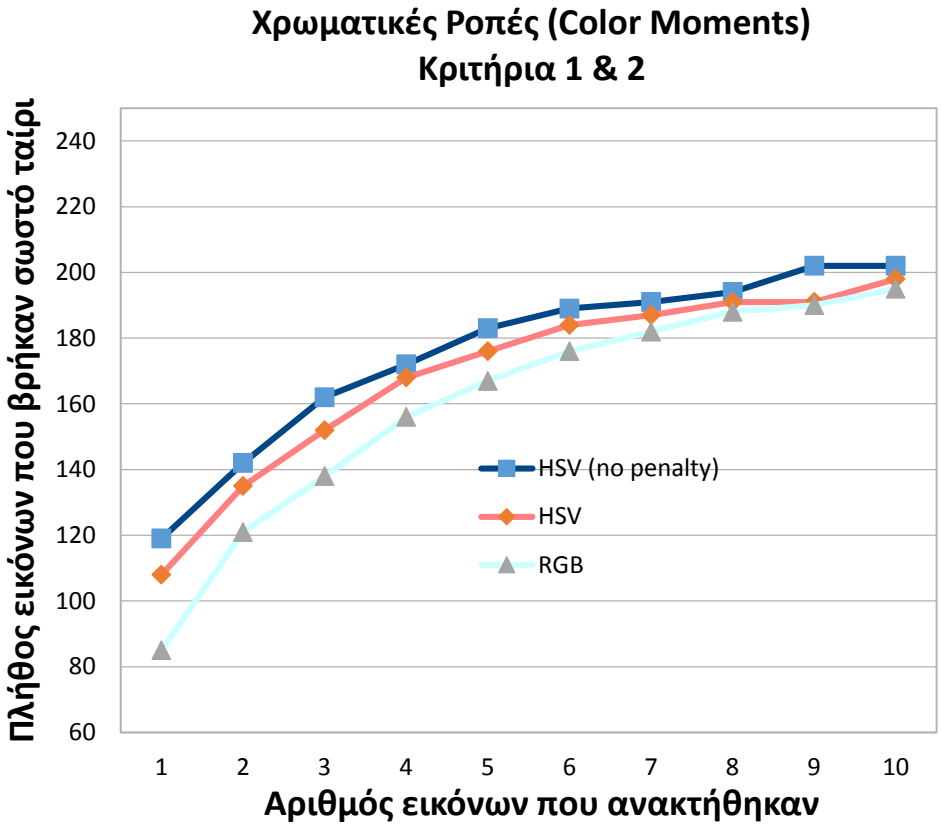
Τα αποτελέσματα για τις χρωματικές ροπές ανάλογα με το κανάλι χρώματος παρουσιάζονται στα διαγράμματα των **Εικόνων 3.9** και **3.10**

Στο διάγραμμα 3.9 απεικονίζεται ο αριθμός των εικόνων που βρήκαν σωστό ταίρι σε σχέση με τον αριθμό των εικόνων που ανακτήθηκαν. Η διαδικασία έχει ως εξής: Απαιτώντας τις πλησιέστερες χρωματικά εικόνες, θεωρούμε πως η ανάκτηση πέτυχε, εάν έστω και μία από τις εικόνες που ανακτήθηκαν ταιριάζει χρωματικά με την αρχική, δηλαδή τηρείται το Κριτήριο 2 και 1 με βάση τον

πίνακα αληθείας. Εάν θέλουμε να είμαστε πιο αυστηροί και θεωρούμε ότι η ανάκτηση πέτυχε μόνο εάν κάποια από τις εικόνες που ανακτήθηκαν έχουν ομοιότητα τάξης 2 και όχι 1 (Κριτήριο 2), τότε τα αποτελέσματα παρουσιάζονται στην **Εικόνα 3.10.**

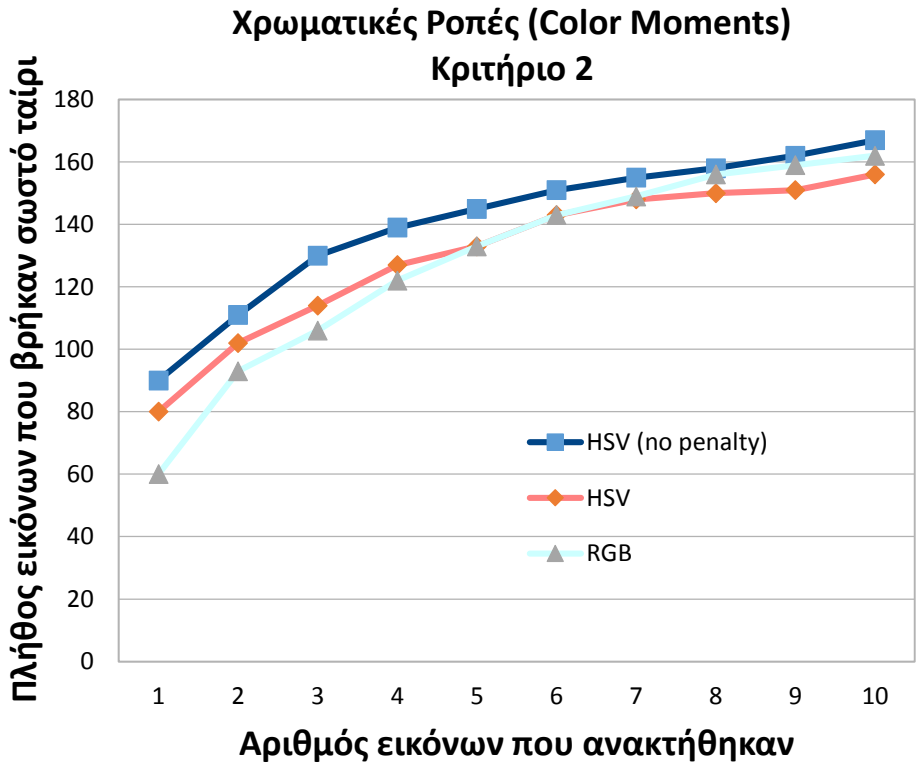
Συμπερασματικά, βλέπουμε πως καλύτερη απόδοση και στις δύο περιπτώσεις παρουσιάζει η ανάκτηση με βάση χρωματικές ροπές στο κανάλι HSV χωρίς κάποιο συντελεστή-ποινή. Η ανωτερότητα αυτής της μεθόδου φαίνεται ξεκάθαρα στην **Εικόνα 3.10**, όπου παρατηρούμε πως για ανάκτηση με βάση αυστηρά χρωματικά κριτήρια, η καμπύλη HSV χωρίς ποινή έχει μεγαλύτερη διαφορά από τις υπόλοιπες για μικρότερο αριθμό εικόνων που ανακτώνται.

Με ανάκτηση των 10 πλησιέστερων εικόνων η μέθοδος αυτή έχει επιτυχία ανάκτησης 202 εικόνων από τις 240.



***Εικόνα 3.9** : Απόδοση ανάκτησης με χρήση χρωματικών ροπών στα χρωματικά κανάλια HSV,RGB.*

Η απόδοση της ανάκτησης, όταν η περιγραφή γίνεται ολικά (όχι σε σημεία - κλειδιά) με χρήση αποκλειστικά χρωματικής πληροφορίας βελτιώνεται καθοριστικά. Για το λόγο αυτό και στην πλειοψηφία των πειραμάτων της εργασίας αυτής προσανατολιζόμαστε προς αυτή την κατεύθυνση.

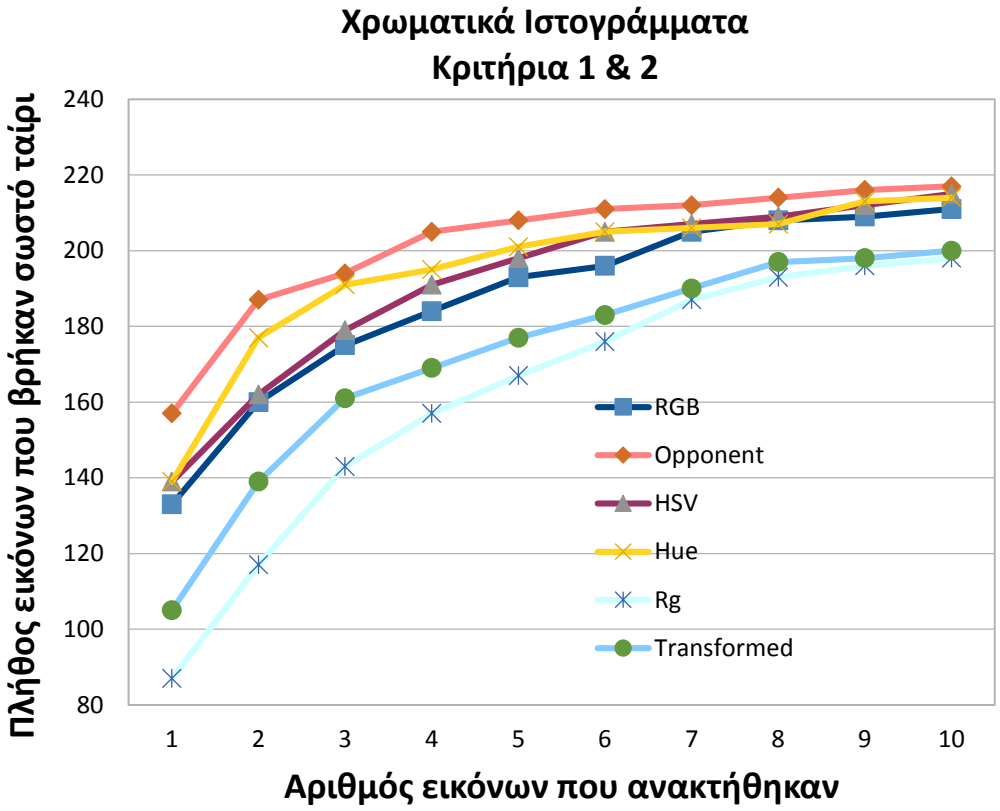


***Εικόνα 3.10** : Απόδοση ανάκτησης με χρήση χρωματικών ροπών στα χρωματικά κανάλια HSV, RGB με με απαίτηση να τηρείται το Κριτήριο 2*

3.3.5 Ολική Ανάκτηση με χρωματικά ιστογράμματα (Global Retrieval with color histograms)

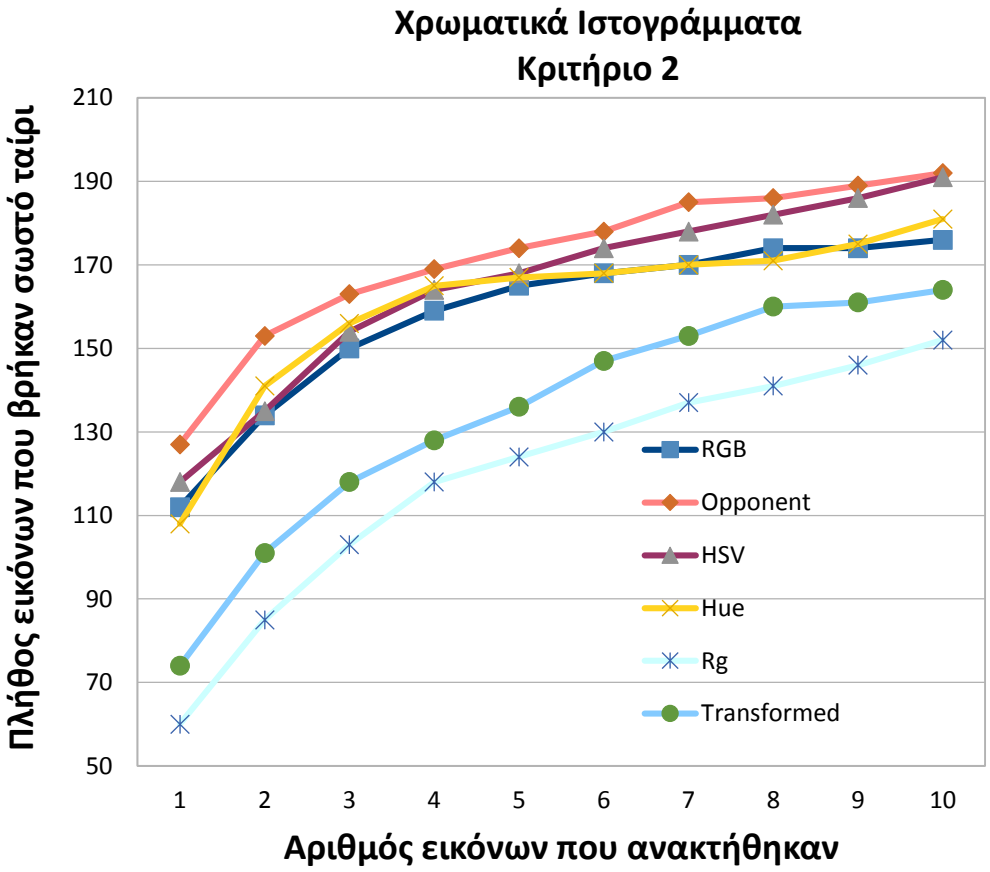
Στο πείραμα αυτό ελέγχεται η απόδοση των ιστογραμμάτων διάφορων χρωματικών χώρων όπως έχουν αναλυθεί στην ενότητα 2.2.2. Τα ιστογράμματα RGB, Opponent, Rg, Transformed διαθέτουν 256 bins για το ιστόγραμμα του κάθε ξεχωριστού καναλιού, δηλαδή για το RGB ιστόγραμμα έχουμε ένα διάνυσμα 3*256=768 διαστάσεων. Η κατασκευή ιστογράμματος με 256 bins στην ουσία σημαίνει πως καταμετρώνται οι φορές που εμφανίζεται κάθε τιμή από 0-255 ενός εικονοστοιχείου. Για το ιστόγραμμα HSV δίνεται προτεραιότητα στην απόχρωση H, το ιστόγραμμα της οποίας χωρίζεται σε 16 bins έναντι 4 bins για τα S,V κανάλια. Τέλος, η κατασκευή του ιστογράμματος Hue βασίζεται στην μέθοδο που προτείνει στο σύγγραμα του ο Van de Weijer προκειμένου να αντιμετωπισθεί η αστάθεια του γύρω από τον γκρι άξονα. Συγκεκριμένα, ο κορεσμός λειτουργεί ως βάρος για κάθε τιμή hue που εισάγεται στο ιστόγραμμα. Στη συνέχεια το ιστόγραμμα εξομαλύνεται με φίλτρο [-1 2 1] και κανονικοποιείται με την συνολική απόχρωση συν την

ποσότητα γκρι της εικόνας. Τα αποτελέσματα που προκύπτουν υπολογίζονται στις εικόνες της βάσης δεδομένων στις οποίες έχει αφαιρεθεί τμήμα του υποβάθρου με το χέρι (**Εικόνα 3.6**).



***Εικόνα 3.11** : Απόδοση ανάκτησης με χρήση χρωματικών ιστογραμμάτων*

Και από τα δύο ανωτέρω διαγράμματα προκύπτει ότι το Opponent ιστόγραμμα έχει καλύτερες επιδόσεις. Στην **Εικόνα 3.11** παρουσιάζεται το πλήθος των εικόνων που βρίσκουν σωστό ταίρι μετά την ανάκτηση σε σχέση με τον αριθμό των εικόνων που ανακτώνται. Επιτυχία θεωρείται το σωστό ταίριασμα της αρχικής εικόνας με έστω και μία από τις εικόνες που ανακτώνται. Σωστό ταίριασμα σημαίνει πως στον πίνακα αληθείας έχουν βαθμό ομοιότητας 1 ή 2. Εάν θέλουμε να είμαστε πιο αυστηροί στα κριτήρια που χρησιμοποιούμε και θωρούμε πως η αρχική εικόνα βρήκε ταίρι μεσα στις ανακτώμενες μόνο εάν αυτές μοιάζουν χρωματικά πάρα πολύ, τότε κρατάμε μόνο αυτές με τις οποίες που πληρούν το Κριτήριο 2 που απεικονίζονται στην **Εικόνα 3.12.**

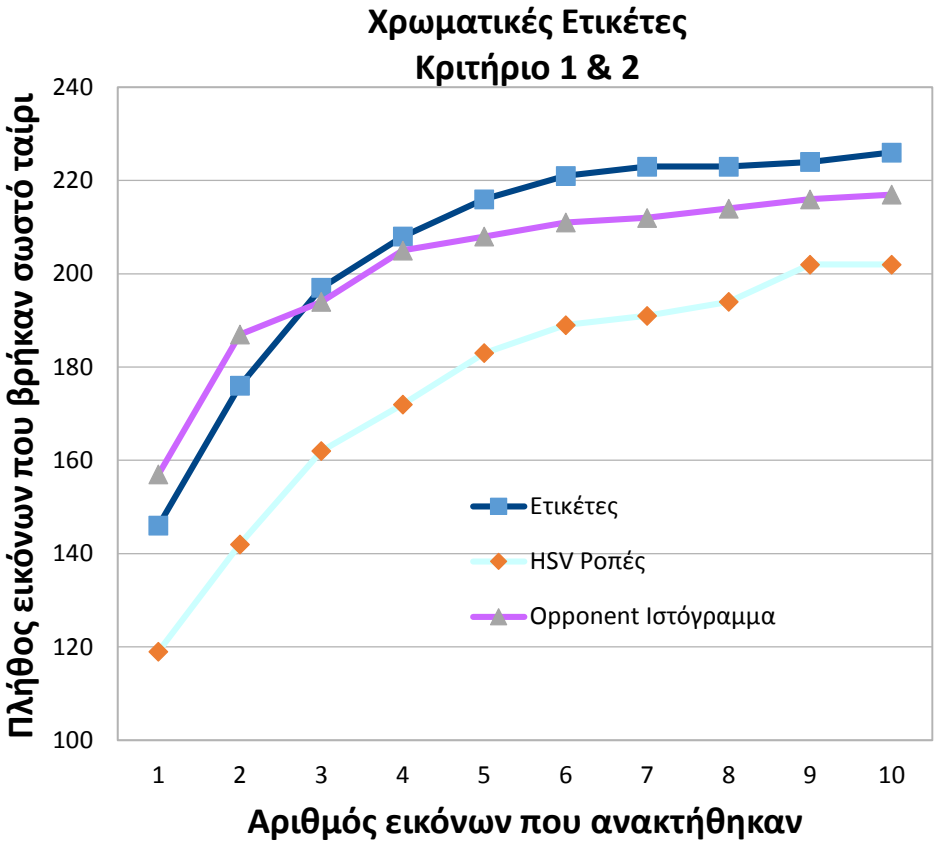


***Εικόνα 3.12** : Απόδοση ανάκτησης μετά την χρήση χρωματικών ιστογραμμάτων με απαίτηση να τηρείται το Κριτήριο 2*

3.3.6 Ολική Ανάκτηση με χρωματικές ετικέτες (Global Retrieval with color naming)

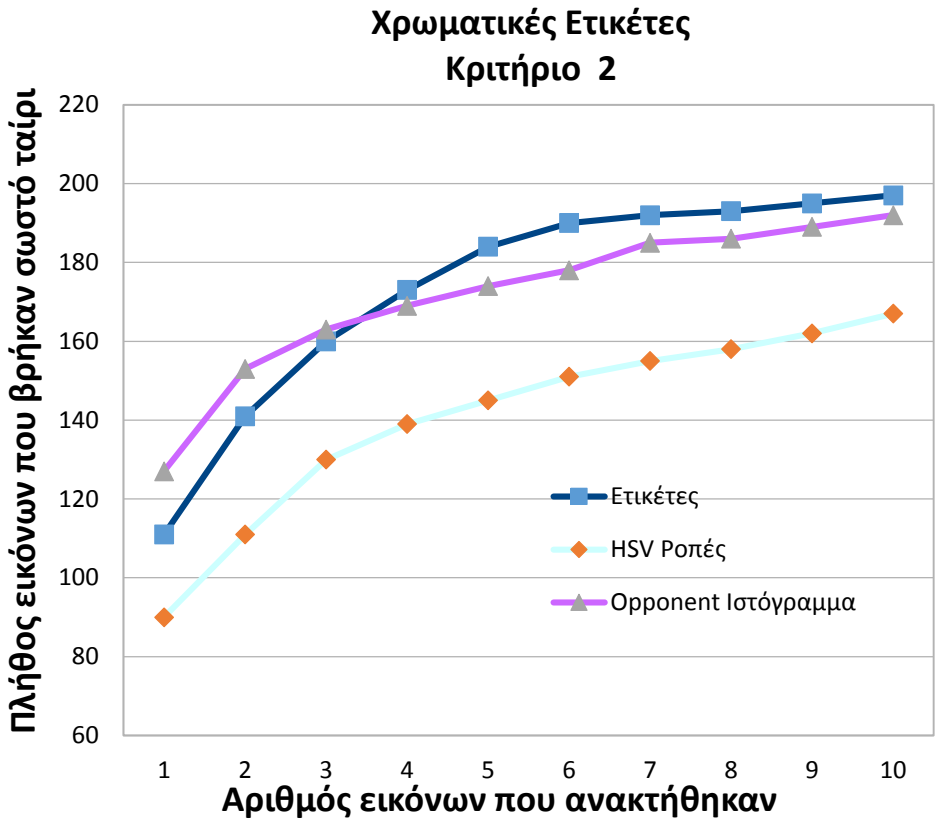
Με βάση την αντιστοίχιση λέξεων των εικόνων από την μηχανή αναζήτησης της Google με χρώματα που πραγματοποιείται στην εκπαίδευση του μοντέλου και περιγράψαμε στην **Ενότητα 2**, μπορούμε να κάνουμε τώρα, στο κομμάτι του testing, δηλαδή αντιστοίχηση των εικονοστοιχείων οποιασδήποτε εικόνας της βάσης δεδομένων με χρώματα. Αυτό γίνεται μέσω του πίνακα w2c (words to colors) που δίνεται από την εκπαίδευση του μοντέλου Color Naming, και εκφράζει για κάθε λέξη από τις 32768 που αντιστοιχήθηκαν στην εκπαίδευση ποια είναι η πιθανότητα να είναι κάποιο από τα 11 χρώματα. Χωρίζοντας σε λέξεις την κάθε εικόνα της βάσης δεδομένων και βρίσκοντας για κάθε μια ποιο χρώμα έχει την μεγαλύτερη πιθανοφάνεια παίρνουμε το τελικό αποτέλεσμα. Η ομοιότητα των εικόνων που ανακτώνται με βάση την αρχική γίνεται μέσω

σύγκρισης ενός διανύσματος 11 θέσεων που περιλαμβάνει το πλήθος των λέξεων της εικόνας που βρέθηκαν να αντιστοιχούν για κάθε χρώμα. Το διάνυσμα αυτό στην συνέχεια κανονικοποιείται με βάση το μέγεθος κάθε εικόνας. Κοντινότερη εικόνα θεωρείται αυτή με το διάνυσμα που διαθέτει την μικρότερη ευκλείδεια απόσταση από το διάνυσμα της αρχικής.



Εικόνα 3.13 : *Απόδοση ανάκτησης με χρήση χρωματικών ετικετών*

Με βάση τα διαγράμματα των **Εικόνων 3.13 και 3.14** είναι εμφανές πως οι χρωματικές ετικέτες υπερισχύουν έναντι των άλλων μεθόδων μετά από ανάκτηση τριών και άνω εικόνων. Υπάρχει όμως ένα βασικό πρόβλημα στην ενότητα αυτή το οποίο επιδιώκεται να λυθεί στην επόμενη ενότητα,το γεγονός ότι οι εικόνες που χρησιμοποιούμε πρόερχονται από κοινές κολεξιόν αρά έχουν κοινό υπόβαθρο. Σκοπός μας είναι η ανάκτηση να γίνεται με βάση τα σωστά κριτήρια, δηλαδή μόνο τις ομοιότητες της ενδυμασίας.



Εικόνα 3.14 : *Απόδοση ανάκτησης με χρήση χρωματικών ετικετών με απαίτηση με απαίτηση να τηρείται το Κριτήριο 2*

3.4 Πειράματα σε τμήματα της εικόνας

Τα προηγούμενα πειράματα εφαρμόστηκαν όχι στο σύνολο των αρχικών εικόνων αλλά στις εικόνες αφότου έχει αφαιρεθεί περιμετρικά ένα κομμάτι από το υπόβαθρο το οποίο σε καμία περίπτωση δεν σχετίζεται με το μέρος της εικόνας που μας ενδιαφέρει, δηλαδή τα ρούχα. Ωστόσο, η νέα εικόνα που προκύπτει εμπεριέχει και πάλι ένα τμήμα από το υπόβαθρο. Σε συνδυασμό με το γεγονός ότι πολλές από τις εικόνες μόδας που χρησιμοποιούμε στη βάση δεδομένων προέρχονται από την ίδια συλλογή (fashion collection), άρα μοιράζονται παρόμοιο υπόβαθρο καθώς και το γεγονός ότι τα ρούχα της ίδιας κολεξιόν πολλές φορές μοιάζουν χρωματικά μεταξύ τους, ένα μέρος των εικόνων που βρήκαν ταίρι δεν βασίστηκαν μόνο στην απλή ομοιότητα των ρούχων τους αλλά και στο παρόμοιο υπόβαθρο ή πολλές φορές κυρίως σε αυτό.

Προκύπτει λοιπόν η ανάγκη να απομονώσουμε μόνο το τμήμα εκείνο της εικόνας που εμπεριέχει κομμάτι ρούχου. Στα πειράματα της ενότητας αυτής

εφαρμόζονται οι ίδιοι χρωματικοί περιγραφείς που παρατηρήσαμε ότι λειτουργούν καλύτερα και συγκεκριμένα τα Hue,Opponent,HSV ιστογράμματα και τις χρωματικές ετικέτες στα σημεία της εικόνας που ανιχνεύεται ανθρώπινο σώμα. Για το σκοπό αυτό χρησιμοποιείται η εκτίμηση πόζας που αναλύθηκε στην **Ενότητα 2** αλλά και μάσκες που εντοπίζουν το σημείο ενδιαφέροντος (ROI masks).

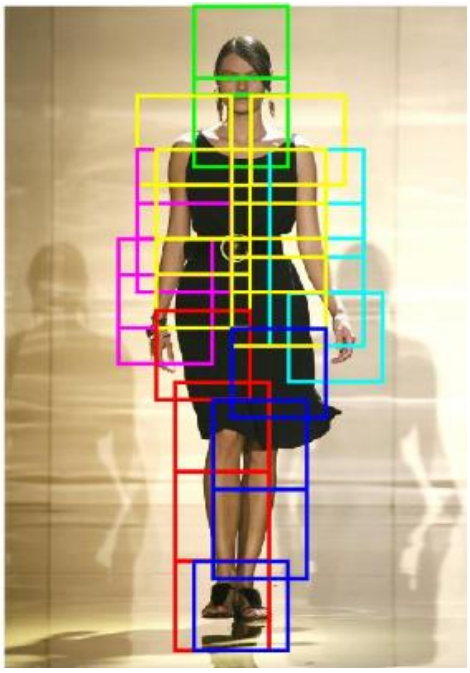
3.4.1 Εκτίμηση Πόζας (Pose Estimation)

Η ανίχνευση ανθρώπων σε εικόνες είναι ένα δύσκολο έργο εξαιτίας του ευρέως φάσματος από πόζες που μπορούν να υιοθετήσουν. Προηγούμενες εργασίες σχετικές με την εκτίμηση πόζας προϋποθέτουν ότι οι άνθρωποι είναι εντοπισμένοι από πριν με έναν ανιχνευτή που παρέχει πληροφορία σχετικά με την θέση και την κλίμακα που εντοπίζεται κάθε άτομο. Το μοντέλο των Yang και Ramanan που θα χρησιμοπιήσουμε όμως εμείς στην ενότητα αυτή, δεν απαιτεί τέτοια προ-επεξεργασία καθώς αναζητά πάνω σε όλες τις τοποθεσίες και τις κλίμακες για τον εντοπισμό των ανθρώπων. Το εκπαιδευμένο μοντέλο για τον εντοπισμού όλου του σώματος απαιτεί περίπου 45 δευτερόλεπτα για την κάθε εικόνα. Ο εντοπισμός ανθρώπων διαφορετικών μεγεθών γίνεται με αναζήτηση πάνω σε μια εικόνα πυραμίδα.

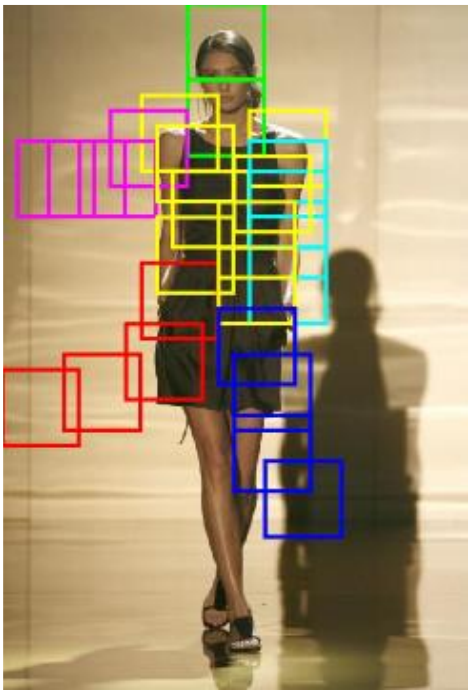
Η εκτίμηση πόζας επιστρέφει κουτιά(bounding boxes) γύρω από το κάθε μέρος του σώματος. Παρατηρήσαμε ότι ο εντοπισμός του κορμού (upper body), δηλαδή τα κουτιά με κίτρινο εντοπίζονται πολύ πιο εύστοχα από τα υπόλοιπα σημεία του σώματος ιδίως τα χέρια. Για το λόγο αυτό τα απομονώσαμε ,για να υπολογίσουμε μόνο σε αυτά τα ιστογράμματα και τις χρωματικές ετικέτες.Το γεγονός ότι πολλά από αυτά έχουν επικάλυψη (overlap) και άρα τα τελικά διανύσματα χρωματικής περιγραφής της εικόνας θα περιέχουν παραπάνω από μια φορά την χρωματική πληροφορία για κάποια εικονοστοιχεία δεν μας απασχολεί ιδιαίτερα, εφόσον παρατηρήθηκε ότι οι περιοχές για τις οποίες έχουμε overlap είναι περιοχές που είναι πιθανότερο να υπάρχει ρούχο, δηλαδή στο κέντρο του κορμού άρα η πολλαπλή επίδραση αυτών των περιοχών λειτουργεί ως συντελεστής βαρύτητας που υποδηλώνει την σημασία αυτών των εικονοστοιχείων εφόσον έχουν μεγαλύτερη πιθανότητα να εμπεριέχουν ρουχισμό από ότι υπόβαθρο με την προϋπόθεση ότι η εκτίμηση πόζας έχει πραγματοποιηθεί ικανοποιητικά (**Εικόνα 3.15**).

Τα αποτελέσματα που προκύπτουν εμφανίζονται στις **Εικόνες 3.17** και **3.18.** Στην **Εικόνα 3.17** παρουσιάζονται οι επιδόσεις των ιστογράμματων και των ετικετών όταν αυτές υπολογίζονται στον κορμό της εκτιμούμενης πόζας. Επειδή το ιστόγραμμα Hue φαίνεται να υπερέχει έναντι των άλλων μεθόδων ,

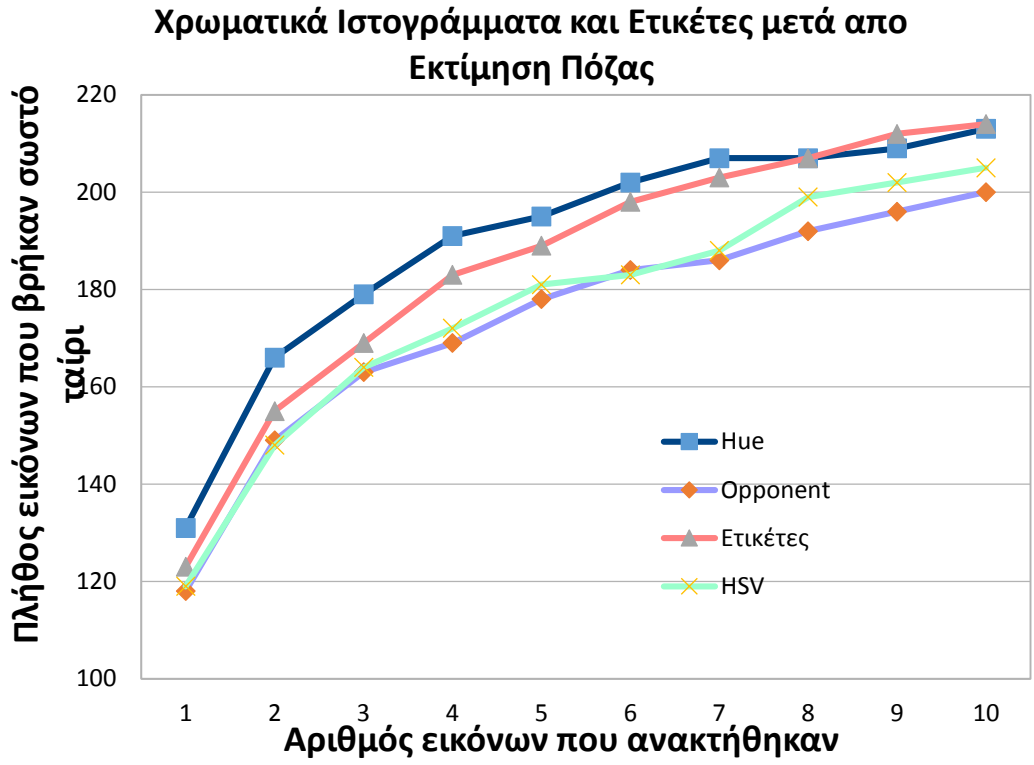
εξετάζουμε και την περίπτωση που περιλαμάνονται και τα bounding boxes των ποδιών και συγκεκριμένα μόνο τα άνω κομμάτια των ποδιών αφού αυτά εντοπίζονται με μεγαλύτερη ευρωστία σε σχέση με το κάτω μέρος (**Εικόνα 3.16**).



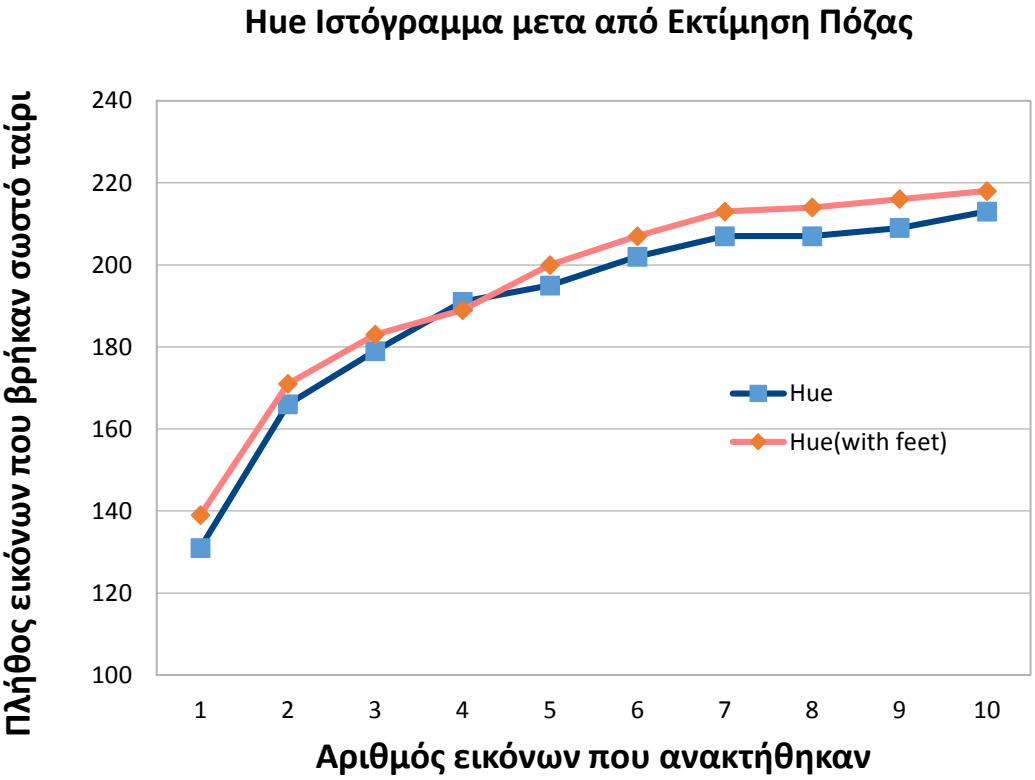
Εικόνα 3.15 : *Παράδειγμα σωστής εκτίμησης πόζας*



Εικόνα 3.16 : *Παράδειγμα λανθασμένης εκτίμησης πόζας*



Εικόνα 3.17: *Απόδοση ανάκτησης με χρήση χρωματικών ετικετών και ιστογραμμάτων μετα την εκτίμηση πόζας για κορμό*



Εικόνα 3.18 : *Απόδοση ιστογραμμάτος Hue μετα την εκτίμηση πόζας για κορμό και άνω μέρος ποδιών*

ΑΡΙΘΜΟΣ ΑΝΑΚΤΟΥΜΕΝΩΝ ΕΙΚΟΝΩΝ	ΠΟΣΟΣΤΟ ΕΠΙΤΥΧΙΑΣ **HUE**	ΠΟΣΟΣΤΟ ΕΠΙΤΥΧΙΑΣ **ΕΤΙΚΕΤΩΝ**	ΠΟΣΟΣΤΟ ΕΠΙΤΥΧΙΑΣ HUE (**ΜΕ ΠΟΔΙΑ**)
	ΟΜΟΙΟΤΗΤΑ 1 & 2	ΟΜΟΙΟΤΗΤΑ 1 & 2	ΟΜΟΙΟΤΗΤΑ 1&2
1	54,60%	51,25%	57,92%
2	69,17%	64,58%	71,25%
3	74,58%	70,42%	76,25%
4	79,58%	76,25%	78,75%
5	81,25%	78,75%	83,33%
6	84,17%	82,50%	86,25%
7	86,25%	84,58%	88,75%
8	86,25%	86,25%	89,17%
9	87,08%	88,33%	90,00%
10	88,75%	89,17%	90,83%

Πίνακας 3.2: *Πίνακας ποσοστών επιτυχίας της ανάκτησης βασισμένη σε ιστόγραμμα Hue μετα την εκτίμηση πόζας για κορμό μόνο καθώς και με το άνω μέρος ποδιών*

Αν και η ανάκτηση βασισμένη στην περιγραφή με Hue ιστογράμματα στα σημεία που προκύπτουν από την εκτίμηση πόζας φαίνεται να φτάνει σε αρκετά καλά ποσοστά,συγκεριμένα μέχρι και 90,8 % (**Πίνακας 3.2**) συμπεριλαμβανομένου και του άνω μέρους των ποδιών, ο χρόνος που απαιτεί ειναι απαγορευτικός για ανάκτηση πραγματικού χρόνου. Ακόμα, λόγω εκπαίδευσης του μοντέλου σε περαστικούς, εμφανίζει δυσκολία για εικόνες προερχόμενες από επιδείξεις μόδας. Τα άκρα έχουν δυσκολία να εντοπισθούν σωστά. Αυτός είναι και ο λόγος άλλωστε που περιοριστήκαμε στον κορμό και το άνω μέρος των ποδιών. Στην επόμενη ενότητα εξετάζεται η χρήση μάσκας για την αντιμετώπιση τέτοιων φαινομένων κατα τον εντοπισμό των σημείων ενδιαφέροντος ,δηλαδή της ενδυμασίας.

3.4.2 Μάσκες στην περιοχή ενδιαφέροντος

Με τον όρο μάσκα περιοχής ενδιαφέροντος εννοούμε μια εικόνα με εικονοστοιχεία μαύρου και άσπρου χρώματος, δηλαδή στην ουσία έναν πίνακα με τιμές 0 για τα σημεία που δεν θέλουμε να λάβουμε υπόψιν και 1 για τα σημεία που θέλουμε να εντοπίσουμε. Απομονώνοντας το μεγαλύτερο συνεκτικό κομμάτι από τιμές 1 της μάσκας λαμβάνουμε το καλυπτόμενο από ρούχα σώμα του μοντέλου, ιδανικά χωρίς τα τμήματα του δέρματος ή των μαλλιών. Ωστόσο υπάρχουν εξαιρέσεις που δυσχεραίνουν την σωστή ανάκτηση (βλέπε **Εικόνα 3.21**).

Το σύνολο από μάσκες που χρησιμοποιήθηκαν στην ενότητα αυτή προέκυψαν από την εργασία του Κωνσταντίνου Ραπαντζίκου. Αρχικά, εντοπίζεται το πρόσωπο του μοντέλου και με βάση το χρώμα του δέρματος σε αυτό επιδιώκεται η απομόνωση παρόμοιων χρωματικών τμημάτων της εικόνας με σκοπό την αφαίρεση του δέρματος που δρα παραπλανητικά στον προσδιορισμό του χρώματος των ρούχων. Παράλληλα αφαιρείται κομμάτι από το υπόβαθρου.

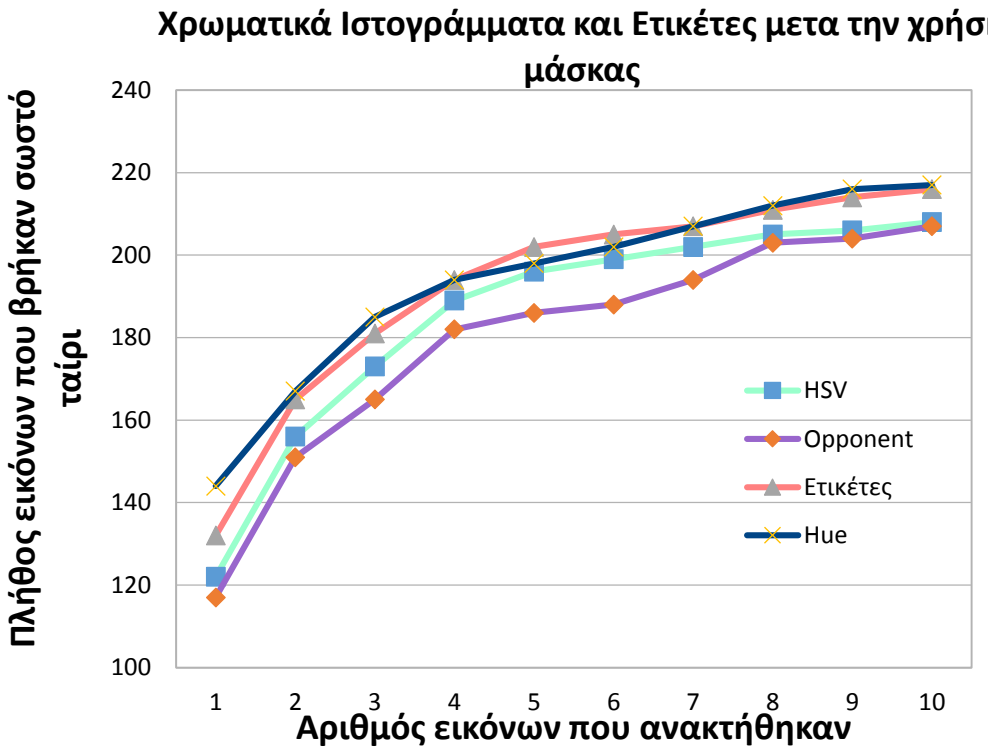
Η χρωματική περιγραφή γίνεται πλέον στην νέα εικόνα που έχει προκύψει μετά την εφαρμογή της μάσκας (**Εικόνα 3.20)**. Σε αυτή υπολογίζονται τα χρωματικά ιστογράμματα και οι χρωματικές ετικέτες που αποδείχτηκαν πιο αποδοτικά στα προηγούμενα πειράματα. Συγκεκριμένα, χρησιμοποιούνται τα ιστογράμματα Hue,Opponent,HSV.

Εικόνα 3.20: *Αρχική εικόνα ,αντίστοιχη μάσκα και εικόνα μετά την εφαρμογή της μάσκας*

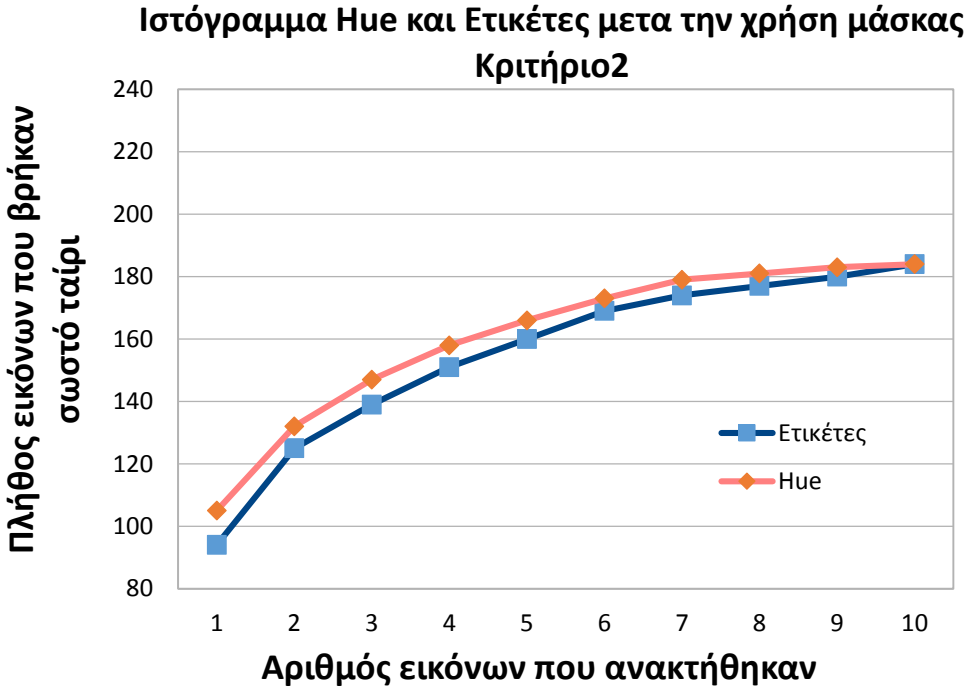
Εικόνα 3.21: *Αρχική εικόνα και αντίστοιχη μάσκα*

Στην **Εικόνα 3.21** παρουσίαζεται μια μάσκα η οποία έχει εντοπίσει λάθος περιοχή και όχι το κομμάτι που περιλαμβάνει ενδυμασία. Αν και υπάρχουν κάποια τέτοια παραδείγματα στο σύνολο των μασκών τα αποτελέσματα σωστής ανάκτησης φτάνουν πολύ καλά ποσοστά ιδιαίτερα καθώς αυξάνεται ο αριθμός ανακτούμενων εικόνων.

Για ακόμα μια φόρα οι χρωματικές ετικέτες αλλά και το ιστόγραμμα Hue υπερισχύουν φτάνωντας σε ποσοστά της τάξης 90,4% για 10 ανακτώμενες εικόνες.(**Εικόνα 3.22** και **Πίνακας 3.3**). Ωστόσο, το ιστόγραμμα Hue φαίνεται να έχει καλύτερη απόδοση για ανάκτηση με βάση το Κριτήριο 2 (**Εικόνα 3.23**).Αυτό σημαίνει πως η αρχική εικόνα για την οποία αναζητάμε κάποια παρόμοια χρωματικά εικόνα, έχει μεγαλύτερη πιθανότητα να βρει μέσα στις ανακτώμενες κάποια στην οποία μοιάζει πάρα πολύ (Κριτήριο 2). Βέβαια, οι διαφορές απόδοσης σε σχέση με τις χρωματικές ετικέτες δεν διαφέρει σημαντικά και άρα και οι δύο μέθοδοι δίνουν ικανοποιητικά αποτελέσματα.



***Εικόνα 3.22**: Απόδοση ανάκτησης με χρήση χρωματικών ετικετών και ιστογραμμάτων μετα την εφαρμογή μάσκας*



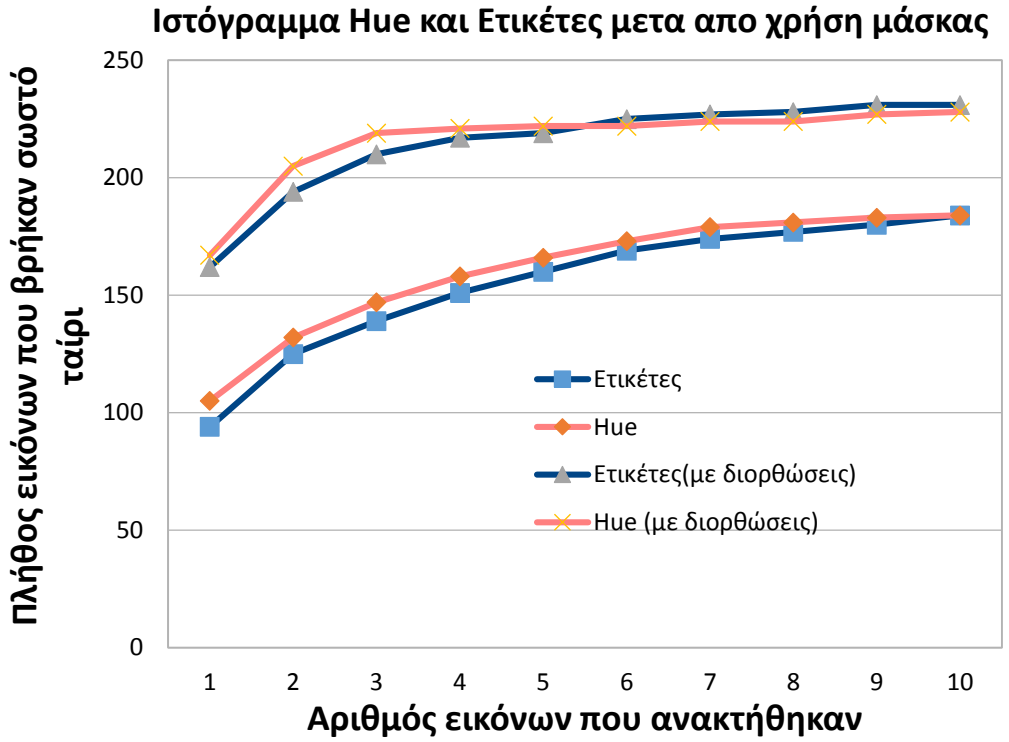
***Εικόνα 3.23:** Απόδοση Hue ιστογράμματος και χρωματικών ετικετών μετα την εφαρμογή μάσκας με απαίτηση να τηρείται το Κριτήριο 2*

ΑΡΙΘΜΟΣ ΑΝΑΚΤΟΥΜΕΝΩΝ ΕΙΚΟΝΩΝ	ΠΟΣΟΣΤΟ ΕΠΙΤΥΧΙΑΣ **HUE**		ΠΟΣΟΣΤΟ ΕΠΙΤΥΧΙΑΣ **ΕΤΙΚΕΤΩΝ**	
	ΟΜΟΙΟΤΗΤΑ 1 & 2	ΟΜΟΙΟΤΗΤΑ 2	ΟΜΟΙΟΤΗΤΑ 1 & 2	ΟΜΟΙΟΤΗΤΑ 2
1	60,00%	43,75%	55,00%	39,17%
2	69,58,%	55,00%	68,75%	52,08%
3	77,08%	61,25%	75,41%	57,92%
4	80,83%	65,83%	80,83%	62,92%
5	82,50%	69,17%	84,17%	66,67%
6	84,17%	72,08%	85,42%	70,42%
7	86,25%	74,58%	86,25%	72,50%
8	88,33%	75,42%	87,92%	73,75%
9	90,00%	76,25%	89,17%	75,00%
10	90,42%	76,67%	90,00%	76,67%

Πίνακας 3.3: *Πίνακας ποσοστών επιτυχίας της ανάκτησης βασισμένη σε ιστόγραμμα Hue και χρωματικές ετικέτες μετα την εφαρμογή μάσκας*

3.4.3 Μάσκες στην περιοχή ενδιαφέροντος με διορθώσεις

Στο κομμάτι αυτό της μελέτης εξετάζουμε τα περιθώρια βελτίωσης των ποσοστών ανάκτησης που προκύπτουν μετά από διορθώσεις στις μάσκες του πειράματος 3.3.1. Οι διορθώσεις γίνονται με τη εντολή του matlab impoly και createmask. Η βελτίωση των αποτελεσμάτων παρουσιάζονται στον **Πίνακα 3.4**



Εικόνα 3.24: *Απόδοση ανάκτησης με χρήση ιστογράμματος Hue και χρωματικές ετικέτες μετά την εφαρμογή διορθωμένων μασκών*

ΑΡΙΘΜΟΣ ΑΝΑΚΤΟΥΜΕΝΩΝ ΕΙΚΟΝΩΝ	ΠΟΣΟΣΤΟ ΕΠΙΤΥΧΙΑΣ **HUE**		ΠΟΣΟΣΤΟ ΕΠΙΤΥΧΙΑΣ **ΕΤΙΚΕΤΩΝ**	
	ΟΜΟΙΟΤΗΤΑ 1 & 2	ΟΜΟΙΟΤΗΤΑ 2	ΟΜΟΙΟΤΗΤΑ 1 & 2	ΟΜΟΙΟΤΗΤΑ 2
1	69,58%	59,58%	67,50%	47,92%
2	85,42%	71,67%	80,83%	59,58%
3	91,25%	77,50%	87,50%	65,42%
4	92,08%	80,00%	90,42%	70,00%
5	92,50%	81,67%	91,25%	73,33%
6	92,50%	82,08%	93,75%	76,67%
7	93,33%	84,17%	94,58%	79,58%
8	93,33%	85,00%	95,00%	80,83%
9	94,58%	86,25%	96,25%	82,50%
10	95,00%	86,67%	96,25%	83,33%

Πίνακας 3.4: *Πίνακας ποσοστών επιτυχίας της ανάκτησης βασισμένη σε ιστόγραμμα Hue και χρωματικές ετικέτες μετά την εφαρμογή διορθωμένων μασκών*

3.4.4 Παραδείγματα σωστής ανάκτησης

Στην ενότητα αυτή θέλουμε να δείξουμε μερικά παραδείγματα σωστής ανάκτησης με βάση χρωματικά χαρακτηριστικά. Η χρωματική περιγραφή έχει γίνει με βάση τις μάσκες της ενότητας 3.4.2 και το χρωματικό ιστόγραμμα απόχρωσης (hue) αλλά και τις χρωματικές ετικέτες.

(α)

(β)

(γ)

(δ)

(ε)

(ζ)

(η) *(θ)* *(ι)*

Εικόνα 3.25: *Οι εικόνες της πρώτης στήλης (α),(δ),(η) είναι οι αρχικές εικόνες για τις οποίες αναζητούμε χρωματικά όμοιες ,οι εικόνες της δεύτερης στήλης (β),(ε),(θ) είναι αυτές που ανακτήθηκαν πρώτες και οι (γ),(ζ),(ι) της τρίτης στήλης αυτές που ανακτήθηκαν δεύτερες με βάση το ιστόγραμμα Hue.*

(α) *(β)* *(γ)*

(δ) *(ε)* *(ζ)*

(η) *(ϑ)* *(ι)*

Εικόνα 3.26: *Οι εικόνες της πρώτης στήλης (α),(δ),(η) είναι οι αρχικές εικόνες για τις οποίες αναζητούμε χρωματικά όμοιες ,οι εικόνες της δεύτερης στήλης (β),(ε),(ϑ) είναι αυτές που ανακτήθηκαν πρώτες και οι (γ),(ζ),(ι) της τρίτης στήλης αυτές που ανακτήθηκαν δεύτερες με βάση τις χρωματικές ετικέτες.*

3.4.5 Κατάτμηση της εικόνας με μάσκα σε μέρη (upper-lower part segmentation of masked image)

Τα πειράματα που προηγήθηκαν εξετάζουν την χρωματική ομοιότητα βασιζόμενη στα χρώματα της ενδυμασίας ενιαία , όχι κατά τμήματα δηλαδή το άνω μέρος του ρουχισμού μιας εικόνας (μπλούζα ,σακάκι ,κλπ) να ταιριάζει χρωματικά με το άνω της άλλης και το κάτω (παντελόνι,φούστα) με το κάτω αντίστοιχα.

Με την προϋπόθεση ότι οι μάσκες έχουν εντοπίσει σωστά το σύνολο της ενδυμασίας ή ότι η εκτίμηση πόζας έχει εντοπίσει και έχει τοποθετήσει ορθά τα μέρη του σώματος μπορούμε να κάνουμε ανάκτηση εικόνων με βάση πιο εξειδικευμένες απαιτήσεις χρώματος. Για παράδειγμα , να ανακτώνται εικόνες με παρόμοια χρώματα ενδυμασίας όχι σαν σύνολο αλλά για κάθε τμήμα της ενδυμασίας (άνω-κάτω) ξεχωριστά ή να ανακτώνται εικόνες που έχουν μόνο ένα τμήματα όμοιο (πχ ενδιαφερόμαστε για εικόνες με ίδιο χρώμα μπλούζας-σακάκι άρα ομοιότητα μόνο στο άνω μέρος). Στις παρακάτω εικόνες δίνονται κάποια τέτοια παραδείγματα. Σε αυτά έχει γίνει κατάτμηση της μάσκας σε άνω και κάτω μέρος και έχουν υπολογιστεί ξεχωριστά οι χρωματικές περιγραφές για το καθένα. Στο τέλος σχηματίζεται ένα ενιαίο διάνυσμα μετά από την συνένωση του άνω και κάτω μέρους για το οποίο υπολογίζεται η ευκλείδεια απόσταση με τα υπόλοιπα.

Εικόνα 3.27.1 και 3.27.2 : *Αρχική εικόνα και εικόνα που ανακτήθηκε*

Χωρίς διάκριση άνω-κάτω μέρους η **Εικόνα 3.27.2** ανακτάται ως πέμπτη πλησιέστερη της αρχικής (**Εικόνα 3.27.1**) ενώ με διάκριση ως πρώτη πλησιέστερη.

Ακόμη ένα παράδειγμα εμφανίζεται στις **Εικόνες *3.28***. Χωρίς διάκριση άνω-κάτω μέρους η **Εικόνα 3.28.2** ανακτάται ως όγδοη πλησιέστερη ενώ με διάκριση ως πρώτη πλησιέστερη. Άρα με διάκριση άνω και κάτω μέρους η ανάκτηση καθίσταται πιο αυστηρή.

Εικόνα 3.28.1 και 3.28.2: *Αρχική εικόνα και εικόνα που ανακτήθηκε*

4 Επίλογος

Στο Κεφάλαιο 4 συνοψίζουμε τα αποτελέσματα της διπλωματικής και προτείνονται πιθανές μελλοντικές επεκτάσεις που θα μπορούσαν να συντελέσουν σε καλύτερη περιγραφή, και άρα ανάκτηση εικόνων μόδας

4.1 Σύνοψη και συμπεράσματα

Με βάση τα αποτελέσματα του Πειραματικού Μέρους απορρέει πως η καλύτερη μέθοδος για την απομόνωση του τμήματος ενδυμασίας μιας εικόνας μόδας είναι η χρήση μασκών. Ωστόσο, και η εκτίμηση πόζας έδωσε ικανοποιητικά αποτελέσματα παρά τον ανασταλτικό παράγοντα του χρόνου εκτέλεσης. Όσον αφορά τα χαρακτηριστικά των τμημάτων ρουχισμού, αποδείχτηκε ότι το χρώμα αποτελεί το πλέον εύρωστο για περιγραφή ενδυμασίας κυρίως μέσω της χρήσης ιστογράμματος απόχρωσης ή χρωματικών ετικετών.

4.2 Μελλοντικές επεκτάσεις

Οι μελλοντικές βελτιώσεις πρέπει να κινηθούν προς την κατεύθυνση του καλύτερου εντοπισμού του σημείου ενδιαφέροντος και τον αποκλεισμό παραπλανητικής πληροφορίας κατά την περιγραφή, στην περίπτωση μας το υπόβαθρο ή το δέρμα του μοντέλου. Η εκτίμηση πόζας θα μπορούσε να εκπαιδευτεί με ένα σύνολο εικόνων από επιδείξεις μόδας για την τοποθέτηση των κουτιών (bounding boxes) πιο εύστοχα, καθώς όπως αναφέραμε έχει εκπαιδευτεί από εικόνες στον δρόμο και όχι σε επιδείξεις μόδας. Για την μείωση του χρόνου που απαιτείται από την εκτίμηση πόζας, προτείνεται να εξεταστεί η μελέτη των P.Felzenszwalb, R.Girshick, D.McAllester**, «Cascade Object Detection with Deformable Part Models»** με σκοπό να επιτευχθούν επιδόσεις σε πραγματικό χρόνο με περαιτέρω επιταχύνσεις μέσω κλιμακούμενων υλοποιήσεων.

Αν και το δέρμα αντιμετωπίστηκε ως ανεπιθύμητη πληροφορία στα πειράματα και οι μάσκες ιδανικά πρέπει να αποκλείουν τέτοια πληροφορία, θα είχε ενδιαφέρον να εξεταστεί ως περαιτέρω πληροφορία ομοιότητας μεταξύ των ρούχων. Για παράδειγμα, δυο εικόνες θα θεωρούνταν όμοιες όχι μόνο στο πλαίσιο του χρώματος πχ. κόκκινο ρούχο αλλά και του δέρματος που καλύπτουν ή όχι, πχ. κοντό κόκκινο ρούχο. Έτσι, οι μάσκες θα

χρησιμοποιούσαν την πληροφορία δέρματος που θα προέκυπταν από τον εντοπισμό προσώπου (face detection), για τον εντοπισμό παρόμοιου χρώματος δέρματος στο σώμα και ανάλογα με το ποσοστό εκτεθειμένου δέρματος να κρίνεται η ομοιότητα των εικόνων.

Επιπλέον, η ανάκτηση θα μπορούσε να επεκταθεί και να μην βασίζεται μόνο στο χρώμα αλλά και στην υφή των ρούχων έτσι ώστε να εντοπίζονται ρούχα με παρόμοια patterns όπως ριγέ ή καρό.

Όλη αυτή η πληροφορία θα μπορούσε να συνδυαστεί για την δημιουργία ενός πλήρως αυτοματοποιημένου συστήματος το οποίο είναι ικανό να παράγει μια λίστα σημασιολογικών όρων που χαρακτηρίζουν τα ρούχα, όπως γίνεται στο σύγγραμμα των H. Chen, A. Gallagher, and B. Girod, **«Describing Clothing by Semantic Attributes»** και την μετέπειτα σύγκριση των όρων για να προταθούν παρόμοιες εικόνες. Αυτή η τεχνική έχει το πλεονέκτημα της αυτόματης ανάθεσης σημασιολογικού περιεχομένου βασισμένη στην οπτική περιγραφή της εικόνας και δύναται να αποτελέσει ιδιαίτερα χρήσιμη εφαρμογή για στυλιστικές προτάσεις .

5 Βιβλιογραφία

1. D. G. Lowe: Distinctive Image Features from Scale-Invariant Keypoints, Jan. 2004

2. Koen E. A. van de Sande, Theo Gevers, and Cees G. M. Snoek: Evaluating Color Descriptors for Object and Scene Recognition, *IEEE Transactions on Pattern Analysis and Machine* Intelligence Volume 32 (9), pages 1582-1596, 2010

3. Joost van de Weijer and Cordelia Schmid: Coloring Local Feature Extraction

4. J. van de Weijer, Cordelia Schmid, Jakob Verbeek, Diane Larlus: Learning Color Names for Real-World Applications. IEEE TIP 2009.

5. Yi Yang and Deva Ramanan, Articulated Human Detection with Flexible Mixtures-of-Parts

6. D.Oneata: "Probabilistic Latent Semantic Analysis", available at http://homepages.inf.ed.ac.uk/rbf/CVonline/LOCAL_COPIES/AV1011/oneata.pdf

7. http://www.slideshare.net/NYCPredictiveAnalytics/introduction-to-probabilistic-latent-semantic-analysis

8. http://www.multimedia-computing.de/mediawiki/images/3/3d/SS08_BN-Lec11-pLSA.pdf

9. http://www2.vincent-net.com/luc/papers/93ieeeip_recons.pdf

10. http://www.mathworks.com/help/images/morphological-reconstruction.html

11. https://commons.wikimedia.org/wiki/File:RGB_channels_separation.png

12. https://www.fhwa.dot.gov/publications/research/safety/13018/002.cfm

13. https://en.wikipedia.org/wiki/Pigment#/media/File:Simple_reflectance.svg

14. http://energy-models.com/lighting

15. http://www.uniformnatural.com/blog/?cat=23

16. https://www.fhwa.dot.gov/publications/research/safety/13018/002.cfm

17. http://www.webdesignerdepot.com/2009/12/how-to-get-a-professional-look-with-color/

18. http://homepages.inf.ed.ac.uk/rbf/CVonline/LOCAL_COPIES/AV0405/KEEN/avas2nkeen.pdf

19. http://lear.inrialpes.fr/people/vandeweijer/color_descriptors.html

20. https://en.wikipedia.org/wiki/HSL_and_HSV

21. http://www.cambridgeincolour.com/tutorials/color-management1.htm

22. http://aishack.in/tutorials/sift-scale-invariant-feature-transform-scale-space/