



**ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ**  
**ΣΧΟΛΗ ΕΦΑΡΜΟΣΜΕΝΩΝ ΜΑΘΗΜΑΤΙΚΩΝ**  
**ΚΑΙ ΦΥΣΙΚΩΝ ΕΠΙΣΤΗΜΩΝ**

ΔΙΑΤΜΗΜΑΤΙΚΟ ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ

**«ΜΑΘΗΜΑΤΙΚΗ ΠΡΟΤΥΠΟΠΟΙΗΣΗ**  
**σε ΣΥΓΧΡΟΝΕΣ ΤΕΧΝΟΛΟΓΙΕΣ και την ΟΙΚΟΝΟΜΙΑ»**

Σχεδιασμός, Ανάπτυξη και Αξιολόγηση Μοντέλων Εμφάνισης  
με Εφαρμογή στην Οπτική Παρακολούθηση

ΤΣΙΡΩΝΗΣ ΒΑΣΙΛΕΙΟΣ  
ΑΡΙΘΜΟΣ ΜΗΤΡΩΟΥ: 09315035

ΕΠΙΒΛΕΠΩΝ ΚΑΘΗΓΗΤΗΣ: ΚΩΝΣΤΑΝΤΙΝΟΣ ΚΑΡΑΝΤΖΑΛΟΣ

ΑΘΗΝΑ, Οκτώβριος 2017

-Η σελίδα έχει αφαιρεθεί σκοπίμως κενή-



**NATIONAL TECHNICAL UNIVERSITY OF ATHENS**  
**SCHOOL OF APPLIED MATHEMATICAL AND**  
**PHYSICAL SCIENCES**

MASTER THESIS

**MSc in «MATHEMATICAL MODELING IN MODERN TECHNOLOGIES AND THE ECONOMICS»**

Design, Implementation and Validation of Appearance Models  
for Visual Tracking

TSIRONIS VASILEIOS  
STUDENT ID: 09315035

SUPERVISOR PROFESSOR: KONSTANTINOS KARANTZALOS

ATHENS, October 2017

-Η σελίδα έχει αφαιρεθεί σκοπίμως κενή-



**ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ**  
**ΣΧΟΛΗ ΕΦΑΡΜΟΣΜΕΝΩΝ ΜΑΘΗΜΑΤΙΚΩΝ**  
**ΚΑΙ ΦΥΣΙΚΩΝ ΕΠΙΣΤΗΜΩΝ**

ΔΙΑΤΜΗΜΑΤΙΚΟ ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ

**«ΜΑΘΗΜΑΤΙΚΗ ΠΡΟΤΥΠΟΠΟΙΗΣΗ σε ΣΥΓΧΡΟΝΕΣ ΤΕΧΝΟΛΟΓΙΕΣ  
και την ΟΙΚΟΝΟΜΙΑ»**

Σχεδιασμός, Ανάπτυξη και Αξιολόγηση Μοντέλων εμφάνισης  
με Εφαρμογή στην Οπτική Παρακολούθηση

ΤΣΙΡΩΝΗΣ ΒΑΣΙΛΕΙΟΣ  
ΑΡΙΘΜΟΣ ΜΗΤΡΩΟΥ: 09315035

ΤΡΙΜΕΛΗΣ ΕΞΕΤΑΣΤΙΚΗ ΕΠΙΤΡΟΠΗ:

.....  
ΚΑΡΑΝΤΖΑΛΟΣ  
ΚΩΝΣΤΑΝΤΙΝΟΣ  
Σ.Α.Τ.Μ. Ε.Μ.Π.

.....  
ΑΡΓΙΑΛΑΣ  
ΔΗΜΗΤΡΙΟΣ  
Σ.Α.Τ.Μ. Ε.Μ.Π.

.....  
ΚΑΡΡΑΣ  
ΓΕΩΡΓΙΟΣ  
Σ.Α.Τ.Μ. Ε.Μ.Π.

ΑΘΗΝΑ, Οκτώβριος 2017

..... Βασίλειος Κ. Τσιρώνης  
Διπλωματούχος Αγρονόμος και Τοπογράφος Μηχανικός Ε.Μ.Π.

Copyright © Βασίλειος Κ. Τσιρώνης, 2017

Με επιφύλαξη παντός δικαιώματος. All rights reserved.

Απαγορεύεται η αντιγραφή, αποθήκευση και διανομή της παρούσας εργασίας, εξ ολοκλήρου ή τμήματος αυτής, για εμπορικό σκοπό. Επιτρέπεται η ανατύπωση, αποθήκευση και διανομή για σκοπό μη κερδοσκοπικό, εκπαιδευτικής ή ερευνητικής φύσης, υπό την προϋπόθεση να αναφέρεται η πηγή προέλευσης και να διατηρείται το παρόν μήνυμα. Ερωτήματα που αφορούν τη χρήση εργασίας για κερδοσκοπικό σκοπό πρέπει να απευθύνονται προς το συγγραφέα.

Οι απόψεις και τα συμπεράσματα που περιέχονται σε αυτό το έγγραφο εκφράζουν τον συγγραφέα και δεν πρέπει να ερμηνευθεί ότι αντιπροσωπεύουν τις επίσημες θέσεις του Εθνικού Μετσόβιου Πολυτεχνείου.

*“There are only two things in the world:  
nothing and semantics.”*  
-Werner Erhard-

## Ευχαριστίες

Σε αυτό το σημείο ειλικρινά θα ήθελα να ευχαριστήσω θερμά τον επιβλέποντα καθηγητή μου κ. Κ. Καράντζαλο για την καθοριστική συμβολή του όχι μόνο για την εκπόνηση της παρούσας μεταπτυχιακής εργασίας αλλά και για την γενικότερη καθοδήγηση που μου παρείχε. Επίσης, θα ήθελα να ευχαριστήσω τον κ. Γ. Καρρά και την κ. Ε. Πέτσα, που μοιράστηκαν μαζί μου διάφορες ανησυχίες τους, τόσο περί όρασης υπολογιστών όσο και γενικότερες. Τέλος, θα ήθελα να ευχαριστήσω την κοπέλα μου Αθηνά, για τις κοινές ανησυχίες γύρω από τον επιστημονικό κλάδο της εργασίας, και την οικογένειά μου, καθώς και όλα τα άτομα που ήταν κοντά μου αυτό το διάστημα, για την αμέριστη στήριξή τους στη διάρκεια των μεταπτυχιακών σπουδών μου.



## ΠΕΡΙΛΗΨΗ

Τα τελευταία χρόνια στην επιστήμη της Όρασης Υπολογιστών έχει πραγματοποιηθεί αξιοσημείωτη πρόοδος σε πολλά επιμέρους πεδία, επιτρέποντας την επιτυχημένη υλοποίηση πολλών εφαρμογών με σημαντική πολυπλοκότητα. Σε πολλές από αυτές τις εφαρμογές χρησιμοποιούνται μοντέλα που περιγράφουν χωρικά και φασματικά την εμφάνιση των αντικειμένων ενδιαφέροντος, όπως μοντέλα εμφάνισης ανθρώπων. Η παρούσα εργασία επικεντρώθηκε αρχικά στην βιβλιογραφική έρευνα για τα μοντέλα εμφάνισης, τη χρησιμότητα και αξιοποίησή τους, καθώς και την ανάπτυξη και αξιολόγηση πρωτότυπων μοντέλων εμφάνισης για την περιγραφή ανθρώπων ή αντικειμένων.

Όσον αφορά το πρώτο σκέλος, γίνεται ρητή αναφορά σε πολλαπλές δυνατές κατηγοριοποιήσεις των μοντέλων εμφάνισης τόσο βάσει της διαχωριστικής τους ικανότητας, όσο και βάσει της μεθοδολογίας που ακολουθούν για να περιγράψουν την εκάστοτε εμφάνιση. Επιπλέον, παρουσιάζονται διάφορες και ποικίλες εφαρμογές των μοντέλων εμφάνισης, όπως αυτές συναντώνται στη βιβλιογραφία, ενώ αναπτύσσεται ένα στιβαρό πλαίσιο τόσο για το θεωρητικό υπόβαθρο για τις μεθοδολογίες που αξιοποιεί πιθανώς ένα μοντέλο, όσο και για την αξιολόγηση των μοντέλων εμφάνισης.

Στο δεύτερο σκέλος παρουσιάζονται τρία πρωτότυπα μοντέλα εμφάνισης. Ως εφαρμογή αναφοράς των συγκεκριμένων μοντέλων επιλέγεται η οπτική παρακολούθηση (visual tracking), γεγονός που απαιτεί από τα αναπτυχθέντα μοντέλα να παρουσιάζουν υψηλή διακριτική ικανότητα μεταξύ ομοειδών αντικειμένων. Το κύριο μοντέλο που αναπτύσσεται βασίζεται στην περιγραφή της εμφάνισης του ανθρώπου (ή του εκάστοτε αντικειμένου) τόσο με τη χρήση της χρωματικής πληροφορίας, όσο και με την ενσωμάτωση χωρικών στοιχείων στην περιγραφή. Η αναπαράσταση της χρωματικής πληροφορίας στον κατάλληλο χώρο χρώματος (CIE Lab) είναι κρίσιμη καθώς απαιτείται, θεωρητικά, η ομοιομορφία των χρωματικότητων. Η χωρική πληροφορία ενσωματώνεται στην περιγραφή υπό την έννοια της γραμμικής χωρικής αναφοράς που δίνεται σε κάθε στοιχείο ομοιόμορφης χρωματικής πληροφορίας. Τελικά, το αποτέλεσμα της άνωθεν περιγραφής δεν είναι παρά ένα 3D νέφος σημείων σε ένα 3D χώρο που δημιουργείται από 2 χρωματικές και μία χωρική συνιστώσα.

Τα έτερα δύο μοντέλα που αναπτύχθηκαν αποτελούν παραλλαγές του αρχικού ώστε να προκύπτει τελικά ένας διανυσματικός περιγραφέας. Το μεν χρησιμοποιεί κατάλληλη χωρική διαμέριση και την έννοια του χρωματικού μέσου για να παραγάγει ένα σχετικά περιορισμένου μήκους διάνυσμα περιγραφής, ενώ το δε στηρίζεται στη διαμέριση του πλήρους 3D χώρου, με διαφορετικά εύρη κβάντισης, όμως, για τη χωρική και τις χρωματικές συνιστώσες. Τελικά και στις δύο περιπτώσεις προκύπτει μία διανυσματική περιγραφή της εμφάνισης του εκάστοτε ανθρώπου, γεγονός που καθιστά τα δύο αυτά μοντέλα ως άμεσα αξιοποιήσιμα και εφαρμόσιμα σε αλγορίθμους οπτικής παρακολούθησης, σε αντίθεση με το πρώτο (μη διανυσματικό) μοντέλο που αναπτύχθηκε.

Το τελευταίο κομμάτι της παρούσας μεταπτυχιακής εργασίας αποτελούν η αξιολόγηση των μοντέλων και η εξαγωγή ειδικότερων και γενικότερων συμπερασμάτων. Όσον αφορά την αξιολόγηση ακολουθείται μία ιεραρχική δομή πειραμάτων με κλιμακούμενη πολυπλοκότητα. Τελικά τα μοντέλα δοκιμάζονται σε ένα πραγματικό σύστημα οπτικής παρακολούθησης πολλαπλών αντικειμένων με τα αποτελέσματα να αναδεικνύουν τη σημαντικότητα των μοντέλων. Η μεθοδολογία της αξιολόγησης στηρίζεται σε ευρέως χρησιμοποιούμενες μετρικές σφάλματος για την περίπτωση των διανυσματικών περιγραφέων, ενώ για το πρωταρχικό μοντέλο αναπτύσσεται μία σειρά μετρικών ώστε να αντιμετωπισθεί η (μη διανυσματική) φύση του ως 3D σημειοσυνόλου. Στο τελευταίο κεφάλαιο εξάγονται συμπεράσματα τόσο για την άμεση χρησιμότητα των μοντέλων εμφάνισης όσο και για τις μελλοντικές κατευθύνσεις του κλάδου.

## ABSTRACT

---

### Master Thesis

*“Design, Implementation and Validation of appearance models  
for visual tracking”*

Vasileios K. Tsironis

National Technical University of Athens

School of Applied Mathematical and Physical Sciences

---

Appearance modelling constitutes a critical part of many modern Computer Vision (CV) algorithms. Appearance Models can be found in various forms in the literature and are usually the main submodule of various CV procedures and algorithms, e.g. in Object Recognition and Detection algorithms or in Visual Tracking systems (of both single and multiple objects). This thesis discusses two main topics; first a research part about the types and the integration of the appearance models in a variety of algorithms and applications. Second, the design, development and evaluation of prototype models for human (in most cases) appearance description.

In the first part of the thesis, several possible categorizations of appearance models are presented. Two main types of categorizations are dominant; the one based on the discriminative ability of the models and the other one based on the principal features upon which each model is based. Furthermore, a selection of applications using diverse types of appearance models is presented based on bibliography. Finally, the theoretical background of both methodologies behind the appearance models and their respective evaluation criteria are thoroughly presented.

The second part of the thesis, on the other hand, discusses the design, implementation and evaluation of three prototype appearance models. The reference application for these models is visual tracking with emphasis on multiple pedestrian tracking that requires a high intra-object discriminative ability for the models. The main model, namely “ab-Y”, is based on modelling both the spectral and spatial features of an object. The spectral representation is implemented on the CIELab color space mainly due to its uniformity abilities. The resulting descriptor consists of a 3D point cloud where each point is described using two spectral and one spatial component. That

descriptor, however, is not easily applicable due to its non-vectorized nature.

Thus, two more appearance models were developed to address this issue. The first one, namely “Levels”, is based on the quantization of the spatial component followed by an aggregation method for the remaining two spectral components. The resulting description vector is characterized by its limited length and high information compression. The second, namely “Voxels”, uses a binary description vector that models the binary density of the “ab-Y” point cloud quantizing both the spectral and spatial components. Both models are far easier to reclaim and could potentially be used as a “plug-and-play” component in most visual tracking systems.

In the final two chapters, all three developed appearance models are evaluated in a series of tests of escalating difficulty and complexity using both qualitative and quantitative criteria. The final test is their integration in a real-world multiple object tracking system where their real potential and advantages-disadvantages are discovered. From these tests certain assumptions are extracted and extensively discussed. Finally, a discussion about further work and, in general, further future scientific progress takes place.

## Πίνακας περιεχομένων

ΠΕΡΙΛΗΨΗ .....	- 3 -
ABSTRACT .....	- 5 -
Πίνακας περιεχομένων .....	- 7 -
<b>ΚΕΦΑΛΑΙΟ 1.</b>	
<b>ΕΙΣΑΓΩΓΗ .....</b>	<b>- 11 -</b>
1) Η ανάγκη για σημασιολογική περιγραφή.....	- 11 -
2) Αντικείμενο, κίνητρο και στόχος .....	- 13 -
3) Δομή της εργασίας.....	- 14 -
<b>ΚΕΦΑΛΑΙΟ 2.</b>	
<b>ΘΕΩΡΗΤΙΚΟ ΥΠΟΒΑΘΡΟ.....</b>	<b>- 15 -</b>
1) Ψηφιακή Εικόνα και Βίντεο .....	- 15 -
2) Μοντελοποίηση χρώματος – Χώροι χρώματος .....	- 18 -
a) Ορισμός χρώματος.....	- 18 -
b) Χρωματομετρία .....	- 20 -
c) Θεωρίες χρωματικής αντίληψης και χρωματική συνταύτιση.....	- 20 -
d) Γραμμικοί χώροι χρώματος.....	- 25 -
(i) Ο χώρος RGB.....	- 25 -
(ii) Ο χώρος XYZ.....	- 26 -
(iii) Ο χώρος CMY(K) .....	- 28 -
e) Χώροι χρώματος με χρωματικότητες .....	- 29 -
(i) Ο χώρος Ycber .....	- 29 -
(ii) Χώροι με πολικές χρωματικές συνιστώσες (HSV/HSI/HSL).....	- 30 -
(iii) Ομοιόμορφοι χώροι χρώματος.....	- 32 -
3) Υπερ-εικονοστοιχεία (super-pixels).....	- 36 -
a) Ορισμός υπερεικονοστοιχείων .....	- 36 -
b) Simple Linear Iterative Clustering (SLIC).....	- 36 -
(i) Ο ρόλος της απόστασης .....	- 40 -

### ΚΕΦΑΛΑΙΟ 3.

<b>ΜΟΝΤΕΛΑ ΕΜΦΑΝΙΣΗΣ.....</b>	<b>- 42 -</b>
1) Κατηγοριοποίηση.....	- 42 -
2) Μοντέλα βασισμένα σε ακμές/κλίσεις.....	- 43 -
3) Μοντέλα βασισμένα σε keypoints.....	- 45 -
4) Μοντέλα βασισμένα σε patches.....	- 46 -
5) Μοντέλα βασισμένα στη χρωματική πληροφορία .....	- 47 -
6) Βιβλιογραφική αναφορά αξιοποίησης μοντέλων εμφάνισης και εφαρμογές .....	- 49 -

### ΚΕΦΑΛΑΙΟ 4.

<b>ΑΝΑΠΤΥΧΘΕΝΤΑ ΜΟΝΤΕΛΑ .....</b>	<b>- 55 -</b>
1) Κίνητρα - Στόχευση.....	- 55 -
2) Επιλογή χώρου χρώματος.....	- 57 -
3) Μοντέλο Εμφάνισης “ab-Y” .....	- 59 -
a) Περιγραφή / Μεθοδολογία.....	- 59 -
b) Χώρος προτύπων.....	- 61 -
c) Γιατί Y;.....	- 62 -
d) Περιγραφέας ως 3Δ νέφος και το πρόβλημα συν-προσαρμογής.....	- 63 -
4) Μοντέλο Εμφάνισης “Levels” .....	- 64 -
a) Περιγραφή / Μεθοδολογία.....	- 64 -
b) Χώρος Προτύπων .....	- 64 -
c) Διανυσματικός περιγραφέας.....	- 65 -
d) Σχέση με “ab-Y” .....	- 65 -
5) Μοντέλο Εμφάνισης “Voxels” .....	- 66 -
a) Περιγραφή / Μεθοδολογία.....	- 66 -
b) Χώρος Προτύπων .....	- 67 -
c) Διανυσματικός περιγραφέας.....	- 67 -
d) Σχέση με “ab-Y” .....	- 67 -
6) Επισκόπηση και ζητήματα επίδοσης .....	- 68 -

## ΚΕΦΑΛΑΙΟ 5.

### ΠΕΙΡΑΜΑΤΙΚΑ ΑΠΟΤΕΛΕΣΜΑΤΑ ΚΑΙ ΑΞΙΟΛΟΓΗΣΗ ..... - 70 -

- 1) Αλγόριθμοι και μετρικές σφάλματος ..... - 70 -
  - a) Αλγόριθμοι συν-προσαρμογής νεφών “ab-Y” ..... - 71 -
    - (i) ICP-Z..... - 71 -
    - (ii) Εγγύτερου σημείου ..... - 72 -
  - b) Μετρικές σφάλματος ..... - 72 -
    - (i) Μέσο τετραγωνικό σφάλμα..... - 72 -
    - (i) Αμφίδρομο ενδιάμεσο σφάλμα ..... - 73 -
    - (ii)  $Lp - νόρμες$ ..... - 74 -
- 2) Αναγνώριση ενός στόχου σε ένα καρέ ..... - 76 -
  - a) abY πρότυπο ..... - 76 -
    - (i) Closest Point algorithm και  
δ्वιπλό ενδιάμεσο σφάλμα..... - 76 -
    - (ii) ICP-Z και Μέσο Τετραγωνικό σφάλμα ..... - 77 -
  - b) Levels πρότυπο..... - 78 -
    - (i) Ευκλείδεια μετρική ..... - 78 -
    - (ii) City block μετρική ..... - 79 -
  - c) Voxels πρότυπο ..... - 80 -
    - (i) Ευκλείδεια μετρική ..... - 80 -
    - (ii) City block μετρική ..... - 81 -
- 3) Αντιστοίχιση ανιχνεύσεων και στόχων  
από καρέ σε καρέ ..... - 82 -
- 4) Αντιστοίχιση μοντέλων εμφάνισης  
με τις ανιχνεύσεις ενός καρέ ..... - 87 -
- 5) Εφαρμογή μοντέλων εμφάνισης  
σε σύστημα οπτικής παρακολούθησης ..... - 90 -

**ΚΕΦΑΛΑΙΟ 6.**

<b>ΣΥΜΠΕΡΑΣΜΑΤΑ ΚΑΙ ΠΡΟΟΠΤΙΚΕΣ</b> .....	- 92 -
1) Συμπεράσματα.....	- 94 -
2) Μελλοντικές προεκτάσεις.....	- 96 -
<b>ΒΙΒΛΙΟΓΡΑΦΙΑ</b> .....	- 98 -



# ΚΕΦΑΛΑΙΟ 1.

## ΕΙΣΑΓΩΓΗ

Η επιστήμη της Όρασης Υπολογιστών έχει σημειώσει τεράστια πρόοδο κατά τη διάρκεια των τελευταίων χρόνων. Σημαντικό ρόλο σε αυτό έχουν επιτελέσει τόσο η άνοδος της υπολογιστικής ικανότητας των Η/Υ, γεγονός που επιτρέπει την ανάπτυξη ολοένα και πιο πολύπλοκων αλγορίθμων, όσο και η άνθιση της Μηχανικής Μάθησης (Machine Learning). Ειδικά η τελευταία έχει επιτρέψει τη χρήση αλγορίθμων που «μαθαίνουν» σε ένα ευρύ φάσμα εφαρμογών της Όρασης Υπολογιστών.

Στο παρόν Κεφάλαιο θα επισημανθεί η διαρκώς αυξανόμενη ανάγκη της Όρασης Υπολογιστών για εξαγωγή σημασιολογικής πληροφορίας και θα προκύψουν τόσο τα κίνητρα που οδήγησαν στην παρούσα εργασία όσο και ο στόχος της. Τέλος, θα γίνει αναφορά στο επιστημονικό πλαίσιο στο οποίο εντάσσεται η μεταπτυχιακή εργασία.

### 1) Η ανάγκη για σημασιολογική περιγραφή

Ένας πιθανός ορισμός του αντικείμενου της Όρασης Υπολογιστών θα μπορούσε σε γενικές γραμμές να είναι ο εξής: Ως Όραση Υπολογιστών ορίζεται ο κλάδος της επιστήμης που ασχολείται με την εξαγωγή μετρικής και σημασιολογικής πληροφορίας από μία ή περισσότερες εικόνες. Από τον παραπάνω ορισμό γίνεται ήδη φανερό ότι η σημασιολογική περιγραφή της εικόνας αποτελεί σημαντικό κεφάλαιο της επιστήμης.

Ιστορικά, τα πρώτα προβλήματα που αντιμετωπίστηκαν δεν ήταν παρά τα προβλήματα μέτρησης. Σήμερα η διαδικασία μέτρησης μέσω των εικόνων (είτε μέσω τεχνικών Φωτογραμμετρίας ή με χρήση πιο γενικών μεθόδων) έχει επιτύχει έναν υψηλό βαθμό ωρίμανσης με τη νέα έρευνα να εστιάζεται σε βελτιστοποιήσεις των υπάρχουσών τεχνικών (π.χ. S.L.A.M., M.V.S., κοκ.) και όχι σε βασική έρευνα σε επίπεδο βασικής θεωρίας (π.χ. στη γεωμετρία).

Η περίπτωση της σημασιολογικής περιγραφής μιας εικόνας διαφέρει άρδην, όσον αφορά το ποσοστό ωρίμανσης ως κλάδος, από την αντίστοιχη της μέτρησης. Η έννοια της σημασιολογίας (semantics) σε μία εικόνα αποτελεί ένα πολύ γενικά ορισμένο αντικείμενο. Για παράδειγμα, η ταξινόμηση

εικόνων ανάλογα με το περιεχόμενό τους αποτελεί μία άμεση εφαρμογή εξαγωγής σημασιολογικής πληροφορίας. Η περίπτωση της χρήσης ενός περιγραφέα σε κάποιο σημείο ενδιαφέροντος με στόχο την αλληλο-προσαρμογή δύο εικόνων (co-registration), αν και τελικά αποτελεί μία μετρητική διαδικασία, εμπεριέχει σε κάποιο βαθμό κάποια σημασιολογική περιγραφή έστω και αν βρίσκεται σε πεπλεγμένη μορφή στον περιγραφέα. Το παραπάνω γίνεται φανερό όταν ο ίδιος περιγραφέας χρησιμοποιηθεί, π.χ., σε ένα σύστημα ταξινόμησης εικόνων με τη χρήση αναπαράστασης Bag-of-Visual-Words και SVM ταξινομητών.

Παρ' όλη τη χρησιμότητα της σημασιολογικής περιγραφής σε διεργασίες της Όρασης Υπολογιστών, η χρησιμοποίηση πραγματικής και «καθαρής» σημασιολογικής πληροφορίας αποτελεί ένα σχετικά πρόσφατο φαινόμενο της τελευταίας πενταετίας περίπου. Ο λόγος που συμβαίνει αυτό δεν είναι παρά η τεράστια δυσκολία των χαρακτηριστικών που έχουν σχεδιαστεί από τον άνθρωπο και με γνώμονα τη δυνατότητα γενίκευσης ώστε να αποτελέσουν την πηγή άντλησης της πραγματικής σημασιολογικής πληροφορίας. Αν και, όπως προαναφέρθηκε, η σημασιολογική πληροφορία, κατά κανόνα, εμπεριέχεται σε πεπλεγμένη μορφή (πιθανώς σε συνιστώσες κάποιου διανύσματος) στα περισσότερα χαρακτηριστικά (μεσαίου επιπέδου) μιας εικόνας, δεν είναι προφανής η μεθοδολογία μετάβασης από την πεπλεγμένη στην «καθαρή» μορφή της πληροφορίας.

Λύση στο παραπάνω πρόβλημα «μετάβασης» πληροφορίας δίνουν, μερικώς, οι αλγόριθμοι μάθησης της Μηχανικής Μάθησης, με χαρακτηριστικότερο παράδειγμα τους ταξινομητές SVM. Παρόμοιες προσεγγίσεις υιοθετούνται από την ερευνητική κοινότητα εδώ και δεκαετίες με καλά αποτελέσματα, όμως συνεχίζουν να υπάρχουν σαφείς περιορισμοί στις μεθοδολογίες. Την τελευταία πενταετία η ανάπτυξη του υπο-κλάδου της Βαθιάς Μάθησης (Deep Learning) και των Βαθιών Τεχνητών Νευρωνικών Δικτύων έχει οδηγήσει στην «εγκατάλειψη» των χειροκίνητα σχεδιασμένων χαρακτηριστικών προς χάριν χαρακτηριστικών που μαθαίνονται από τα ίδια τα δεδομένα. Τα χαρακτηριστικά αυτά, κυρίως λόγω της μεθοδολογίας εκμάθησης (π.χ. ελαχιστοποίηση του (σημασιολογικού) σφάλματος ταξινόμησης), επιτρέπουν στην σημασιολογική πληροφορία να προκύψει με ομαλό και σταδιακό τρόπο από τα ίδια τα δεδομένα όπως, π.χ., μέσω των στρώσεων ενός Συνελκτικού Νευρωνικού Δικτύου (Convolutional Neural Network - CNN).

## 2) Αντικείμενο, κίνητρο και στόχος

Η παρούσα μεταπτυχιακή εργασία έχει ως αντικείμενό της τα **μοντέλα εμφάνισης αντικειμένων** με έμφαση στα μοντέλα εμφάνισης ανθρώπων. Τα μοντέλα που απασχολούν την παρούσα εργασία διευκολύνουν τη διάκριση μεταξύ ομοειδών αντικειμένων καθιστώντας τα ιδανικά για πληθώρα εφαρμογών της Όρασης Υπολογιστών. Συγκεκριμένα, σε αυτή την εργασία αναπτύσσονται διαφορετικά μοντέλα εμφάνισης ανθρώπων, τα οποία αξιολογούνται με γνώμονα τις επιδόσεις τους ως τμήματος ενός συστήματος παρακολούθησης ενός πλήθους πεζών.

Τα **κίνητρα** για την επιλογή του παραπάνω αντικειμένου ήταν πολλαπλά. Αρχικά, τα μοντέλα εμφάνισης αποτελούν εξ ορισμού μία διαδικασία εξαγωγής πληροφορίας άμεσα συσχετισμένη με σημασιολογικά χαρακτηριστικά της εμφάνισης ενός αντικειμένου. Δεδομένης της προαναφερθείσας βαρύνουσας σημασίας της σημασιολογίας στην Όραση Υπολογιστών, το αντικείμενο της εργασίας καθίσταται «συμβατό» με τις σημερινές (ερευνητικές και επιχειρησιακές) προκλήσεις του κλάδου. Επιπλέον τα μοντέλα εμφάνισης, αν και σπάνια εμφανίζονται αυτόνομα, τυγχάνουν ευρείας εφαρμογής σε πολλές κατηγορίες αλγορίθμων όπως, για παράδειγμα, οι αλγόριθμοι οπτικής παρακολούθησης. Σε επόμενο κεφάλαιο υπάρχει αναφορά σε πληθώρα εφαρμογών των μοντέλων εμφάνισης, όμως στις περισσότερες των περιπτώσεων οι εφαρμογές αυτές αποτελούν ανοιχτά ερευνητικά ζητήματα. Κατά αυτόν τον τρόπο η μελέτη του αντικειμένου των μοντέλων εμφάνισης απασχολεί κατά κόρον την ερευνητική κοινότητα και είναι άξιο αναφοράς το γεγονός ότι δεν υπάρχει συγκεκριμένη μεθοδολογία δημιουργίας ενός μοντέλου εμφάνισης. Το παραπάνω επιτρέπει (ερευνητικό) πειραματισμό επί του αντικειμένου και καθιστά τη δημιουργία ενός μοντέλου εμφάνισης μία ερευνητική πρόκληση, τόσο λόγω της ποικιλότητας των προσεγγίσεων στη διεθνή βιβλιογραφία, όσο και εξαιτίας της μη ύπαρξης «πεπατημένης» οδού με, σχεδόν, κάθε νέα περίπτωση εφαρμογής να απαιτεί μία εκ βάθρων διαφορετική προσέγγιση.

Στόχος της παρούσας εργασίας, τελικά, δεν είναι παρά η εμβάθυνση στη διαδικασία εξαγωγής σημασιολογικής πληροφορίας από μία εικόνα μέσω της μελέτης του υπο-κλάδου των μοντέλων εμφάνισης αντικειμένων και ανθρώπων. Όσον αφορά το πρακτικό σκέλος, κύριο στόχο αποτελεί η ανάπτυξη πρωτότυπων μοντέλων εμφάνισης ανθρώπων και η αξιολόγησή

τους σε πραγματικές συνθήκες μέσω της ένταξής τους σε έναν αλγόριθμο οπτικής παρακολούθησης. Σε καμία περίπτωση η παρούσα εργασία δεν σκοπεύει στην ανάπτυξη “state-of-the-art” μοντέλων εμφάνισης, αλλά στην εμπάθυνση (τόσο βιβλιογραφικά, κυρίως, όσο και «επί του πρακτέου») σε έναν επιμέρους κλάδο της Όρασης Υπολογιστών. Ακόμη, αν και η σύγχρονη πρακτική «επιτάσσει» τη χρήση τεχνικών Βαθιάς Μάθησης για την ανάπτυξη μοντέλων εμφάνισης, η τελευταία δεν υιοθετήθηκε καθώς κρίθηκε ότι τα «με το χέρι» σχεδιασμένα μοντέλα εμφάνισης αναδεικνύουν καλύτερα τις προκλήσεις και τις δυσκολίες που συναντώνται σε κάθε στάδιο ανάπτυξης του μοντέλου. Επίσης, η παραπάνω προσέγγιση επιτρέπει τη λήψη κρίσιμων αποφάσεων κατά την ανάπτυξη των μοντέλων, βασιζόμενων στη θεωρία πίσω από τα μοντέλα εμφάνισης, καθιστώντας τη διαδικασία σημαντικά πιο ωφέλιμη διδακτικά σε σχέση με την «τυφλή» διαδικασία της μάθησης ενός Αλγορίθμου Βαθιάς Μάθησης.

### 3) Δομή της εργασίας

Η παρούσα μεταπτυχιακή εργασία οργανώνεται σε 6 κεφάλαια. Στο πρώτο κεφάλαιο τονίζεται η σημασία της σημασιολογικής πληροφορίας σε διάφορες διεργασίες της Όρασης Υπολογιστών και σκιαγραφείται το αντικείμενο και ο στόχος της εργασίας.

Στο δεύτερο κεφάλαιο αναπτύσσονται μια σειρά από θεωρητικά προαπαιτούμενα στα οποία στηρίζεται μεγάλο μέρος της εργασίας. Στη συνέχεια, στο τρίτο κεφάλαιο περιγράφεται εκτενώς ο κλάδος των μοντέλων εμφάνισης και προβάλλονται διάφορες εφαρμογές τους, όπως αυτές συναντώνται στη διεθνή βιβλιογραφία.

Το τέταρτο κεφάλαιο αφορά το σχεδιασμό και την υλοποίηση των τριών πρωτότυπων μοντέλων εμφάνισης που πραγματεύεται το πρακτικό σκέλος της εργασίας. Στο επόμενο κεφάλαιο τα μοντέλα αυτά υποβάλλονται σε μία πληθώρα πειραμάτων και αξιολογείται η απόδοσή τους. Τέλος, στο έκτο και τελευταίο κεφάλαιο εξάγονται χρήσιμα συμπεράσματα και συζητείται η μελλοντική έρευνα που απομένει τόσο στον κλάδο των μοντέλων εμφάνισης όσο και σε συναφή πεδία.

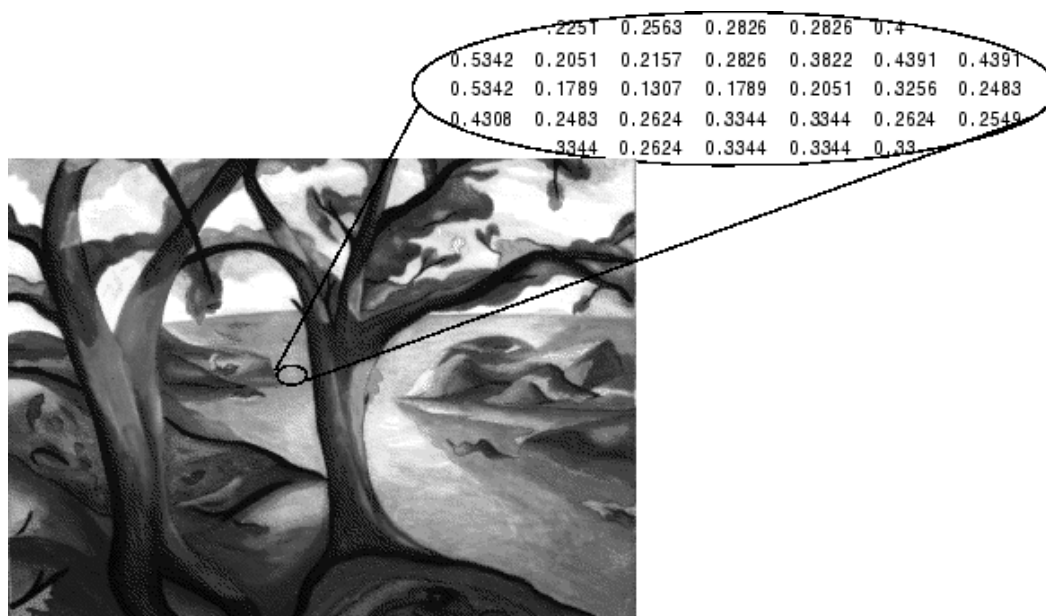
## ΚΕΦΑΛΑΙΟ 2.

### ΘΕΩΡΗΤΙΚΟ ΥΠΟΒΑΘΡΟ

#### 1) Ψηφιακή Εικόνα και Βίντεο

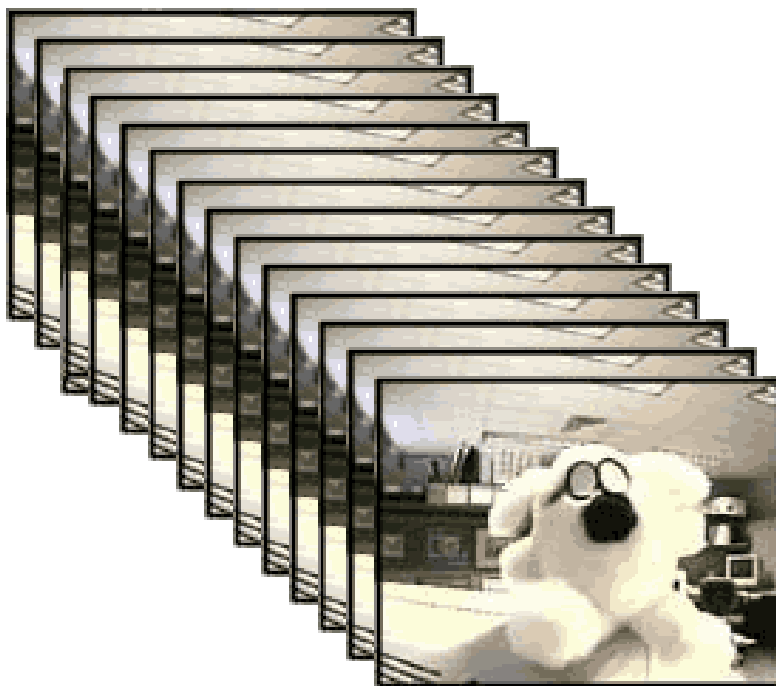
Η ψηφιακή εικόνα αποτελεί τη βάση της επιστήμης της Όρασης Υπολογιστών. Η βασική δομή κάθε ψηφιακής εικόνας είναι το εικονοστοιχείο (picture element - pixel), ενώ ορίζεται και η διάσταση της εικόνας ως  $m \times n$  εικονοστοιχεία, όπου  $m$  ο αριθμός των εικονοστοιχείων κατά την κατακόρυφο διεύθυνση και  $n$  ο αριθμός των εικονοστοιχείων κατά την οριζόντια διεύθυνση.

Αν και όσον αφορά τον άνθρωπο μία εικόνα χαρακτηρίζεται από το περιεχόμενό της και τη σημασιολογική της πληροφορία, στον Η/Υ παρουσιάζεται ως μία ακολουθία αριθμών. Ειδικότερα, για την περίπτωση της «ασπρόμαυρης» (ή καλύτερα της σε κλίμακα του γκρι) εικόνας διαστάσεων  $m \times n$  εικονοστοιχείων, αυτή αναπαρίσταται ως ένας πίνακας  $m$  γραμμών επί  $n$  στηλών. Σε κάθε στοιχείο του πίνακα της εικόνας αντιστοιχεί μία τιμή. Η τιμή αυτή αφορά τον τόνο του γκρι ως ποσοστό αναλογίας ανάμιξης του λευκού έναντι του μαύρου, το εύρος των τιμών όμως εξαρτάται από την ραδιομετρία της εικόνας. Συνοπτικά πρόκειται για την κβάντιση του διαστήματος  $[0,1]$  των ποσοστών ανάμιξης λευκού-μαύρου και την ανάθεση κάθε υπο-διαστήματος σε κάποιο σύνολο τιμών «1-1». Η πιο συχνά χρησιμοποιούμενη ραδιομετρία για γκρι εικόνες είναι η χρήση μη-προσημασμένων ακεραίων 8bit (unsigned integer 8bit - *uint8*), δηλαδή οι τιμές του γκρι κυμαίνονται από το 0 έως το 255 λαμβάνοντας μόνο ακέραιες τιμές. Αυτό συμβαίνει γιατί έγινε κβάντιση 8bit, άρα  $2^8 = 256$  υποδιαστήματα τα οποία ανατίθενται με «1-1» αντιστοιχία σε μη-αρνητικούς (χωρίς πρόσημο) ακέραιους, με τους 256 πρώτους (στον πλήρως διατεταγμένο χώρο των μη-αρνητικών ακεραίων) να είναι οι ακέραιοι αριθμοί από το 0 έως το 255.



Εικόνα 1: Η ψηφιακή εικόνα ως πίνακας ποσοστών ανάμιξης. Πηγή: <http://mathworks.com>

Στην περίπτωση των έγχρωμων εικόνων, όπως αναλύεται και στο επόμενο εδάφιο, υπάρχουν τρεις διαφορετικές γκριζες εικόνες που δημιουργούν την έγχρωμη εικόνα. Οι τιμές κάθε επιμέρους γκριζας εικόνας αναφέρονται στα ποσοστά ανάμιξης του κόκκινου, πράσινου και μπλε χρώματος στον χώρο χρώματος RGB που στηρίζεται στη θεωρία της τριχρωμίας (βλέπε επόμενο εδάφιο περί χρώματος). Όσον αφορά τη ραδιομετρία, η πιο συχνά χρησιμοποιούμενη για την περίπτωση έγχρωμων εικόνων είναι αυτή των μη-προσημασμένων ακεραίων 24bit (*uint24*), η οποία αναφέρεται στην ύπαρξη 3 *uint8* γκριζων εικόνων, μίας για κάθε «κανάλι» χρώματος.



Εικόνα 2: Αναπαράσταση βίντεο ως τανυστή. Πηγή: University of Virginia

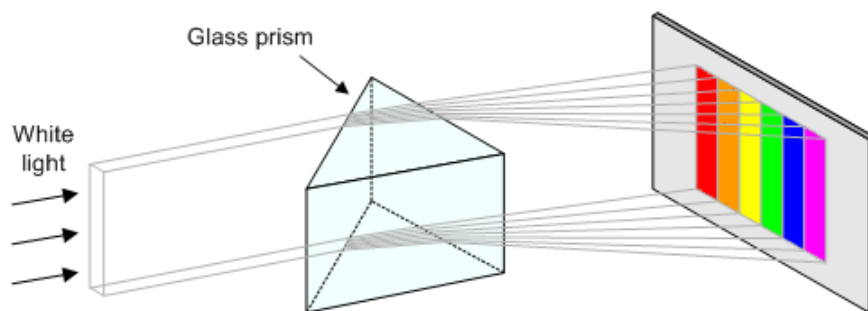
Τα ψηφιακά βίντεο αποτελούν μία ακολουθία (κατά τον χρόνο  $t$ ) εικόνων (γκρίζες ή έγχρωμες) και παριστάνονται με τη χρήση τανυστών (tensors), δηλαδή (άτυπα) πολυδιάστατους πίνακες με διαστάσεις  $m \times n \times (x3) \times t$  όπου  $m \times n \times (x3)$  η διάσταση της κάθε εικόνα της ακολουθίας (καρέ) και  $t$  ο αριθμός των στοιχείων της ακολουθίας (πλήθος καρέ).

## 2) Μοντελοποίηση χρώματος – Χώροι χρώματος

### a) Ορισμός χρώματος

Στην επιστήμη της Φυσικής ως χρώμα ορίζεται μία κατανομή έντασης σε ένα (συνεχές) διάστημα μηκών κύματος του φωτός. Όσον αφορά το «χρώμα», στη διεθνή βιβλιογραφία [30,31] διακρίνεται η έννοια του «χρώματος της πηγής» και του «χρώματος του αντικειμένου». Όσον αφορά το πρώτο, αυτό αναφέρεται στο «φάσμα» (κατανομή μηκών κύματος) που εκπέμπουν οι πηγές, ενώ το δεύτερο αφορά το συνδυασμό του φάσματος της πηγής και των ανακλαστικών ιδιοτήτων του εκάστοτε αντικειμένου. Οι ανακλαστικές ιδιότητες συνδέονται με έννοιες όπως η απορροφητικότητα και η σκέδαση, και μεταβάλλονται συναρτήσει του μήκους κύματος.

Στη βιβλιογραφία συναντάται επίσης και η έννοια του «καθαρού χρώματος» το οποίο ορίζεται ως μία μόνο κρουστική<sup>1</sup> σε ένα μήκος κύματος. Σημαντική συνεισφορά στον παραπάνω ορισμό του «χρώματος» είχαν τα πειράματα του Νεύτωνα, περί το τέλος του 17<sup>ου</sup> αιώνα, στα οποία ανακαλύφθηκε ότι το λευκό φως μπορεί να αποσυντεθεί σε ένα συνεχές φάσμα διαφορετικών καθαρών χρωμάτων και τη σχέση του χρώματος με το φασματικό περιεχόμενο.



Εικόνα 3: Το πείραμα του Νεύτωνα. Πηγή: [www.clivemaxfield.com](http://www.clivemaxfield.com)

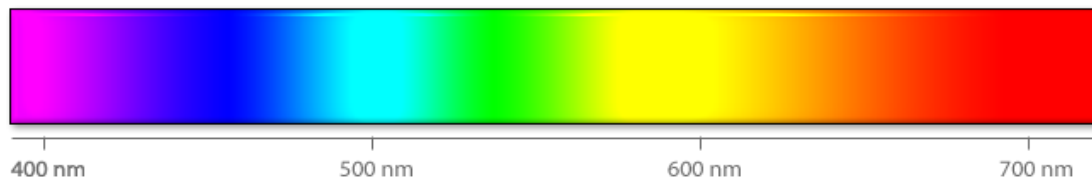
Το χρώμα, πέραν του παραπάνω επιστημονικού ορισμού, μπορεί να χαρακτηριστεί και ως ένα ποιοτικό οπτικό χαρακτηριστικό της όρασης του

---

<sup>1</sup> Ως κρουστική θεωρείται ένα «σύντομο» σήμα εισόδου. Πρακτικά σε συνεχή συστήματα (όπως εν προκειμένω τα μήκη κύματος) μοντελοποιείται από τη συνάρτηση  $\delta$  του Dirac. [https://en.wikipedia.org/wiki/Dirac\\_delta\\_function](https://en.wikipedia.org/wiki/Dirac_delta_function)

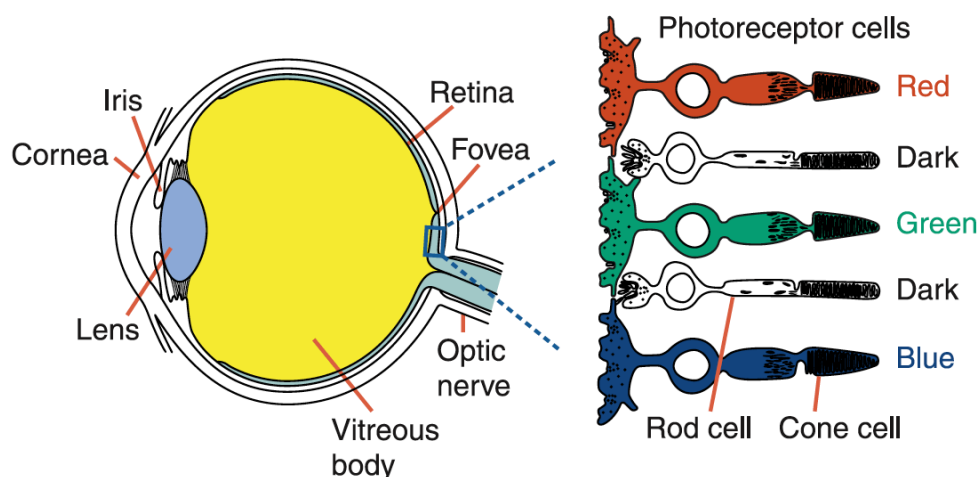


ανθρώπου. Υπό την παραπάνω έννοια ορίζεται ένα σύνολο χρωμάτων, όπως το κόκκινο, το πορτοκαλί, το κίτρινο, το πράσινο, το μπλε κτλ. Το σύνολο των «ορατών» χρωμάτων προκύπτει ότι αντιστοιχεί σε μία ζώνη εύρους 300nm περίπου από τα 400-700nm, η οποία χωρίζεται σε υπο-ζώνες που αντιστοιχούν στις παραπάνω ποιοτικές κατηγοριοποιήσεις.



Εικόνα 4: Εύρος ορατού φάσματος

Ο άνθρωπος αντιλαμβάνεται το χρώμα μέσω κατάλληλων φωτοευαίσθητων βιο-αισθητήρων του οφθαλμού, των κωνίων. Τα κωνία διακρίνονται σε τρεις κατηγορίες ανάλογα με την περιοχή του φάσματος στην οποία τα φωτοευαίσθητα μόρια που διαθέτουν είναι ευαίσθητα και διεγείρονται από εισερχόμενα φωτόνια. Ειδικότερα, υπάρχουν κωνία ευαίσθητα στη φασματική περιοχή του κόκκινου, πράσινου και μπλε χρώματος. Η τελική αίσθηση προκύπτει τόσο από τις αποκρίσεις των κωνίων όσο και από το γειτονικό περιβάλλον (context) που υπεισέρχεται σε μετέπειτα στάδιο της (βιολογικής) επεξεργασίας του συστήματος όρασης. Πρακτικά, η αντίληψη του χρώματος προκύπτει τελικά από τα μήκη κύματος που περιέχονται στο φάσμα τους, τις εντάσεις τους καθώς και το περιβάλλον στο οποίο αυτό παρατηρείται.



Εικόνα 5: Η ανατομία του ανθρώπινου οφθαλμού. Πηγή: [amkrosnar.files.wordpress.com](http://amkrosnar.files.wordpress.com)

Το χρώμα κατέχει ιδιαίτερα σημαντικό ρόλο στη σύγχρονη επιστήμη και στην τεχνολογία με πληθώρα εφαρμογών όπως η φωτογραφία, η εκτύπωση, η οπτικοποίηση δεδομένων, τα γραφικά και η Όραση υπολογιστών. Στις τελευταίες δύο, κατά κύριο λόγο, περιπτώσεις σημαντική είναι η μοντελοποίηση του χρώματος, μία διαδικασία η οποία ανήκει στον κλάδο της Χρωματομετρίας. Να σημειωθεί ότι στον κλάδο της Χρωματομετρίας, σύμφωνα με τον ορισμό της CIE<sup>2</sup>, θεωρείται ότι η αντίληψη του χρώματος είναι ανεξάρτητη του περιβάλλοντος καθώς το χρώμα, θεωρείται ότι, αποτελείται από χώρο/χρονικά αμετάβλητα χαρακτηριστικά.

## **b) Χρωματομετρία**

Ως χρωματομετρία (colorimetry) αναφέρεται η επιστήμη μέτρησης και κατανόησης του χρώματος [31]. Η χρωματομετρία αποτελεί τη γενέτειρα επιστήμη για πολλές υπο-επιστήμες όπως η φωτομετρία, καθώς η «φωτεινότητα» είναι ένα από τα χαρακτηριστικά του χρώματος. Κύριος στόχος αυτής της επιστήμης αποτελεί η δυνατότητα αναπαραγωγής ενός χρώματος με πολύ υψηλή ακρίβεια.

Για τον σκοπό αυτό, η μέτρηση του χρώματος πραγματοποιείται στη μορφή διανύσματος, συνήθως 3 διατάσεων, σε κατάλληλους διανυσματικούς χώρους χρώματος. Οι χώροι χρώματος μπορούν να ταξινομηθούν σε γραμμικούς ή μη ανάλογα με τη (γραμμική ή μη) διαδικασία προβολής από την πρωτογενή μορφή μέτρησης (στον χώρο RGB συνήθως). Η ποικιλότητα που παρατηρείται στους παραπάνω χώρους χρώματος οφείλεται σε μεγάλο βαθμό στις θεωρίες που έχουν κατά καιρούς διατυπωθεί για την αντίληψη του χρώματος.

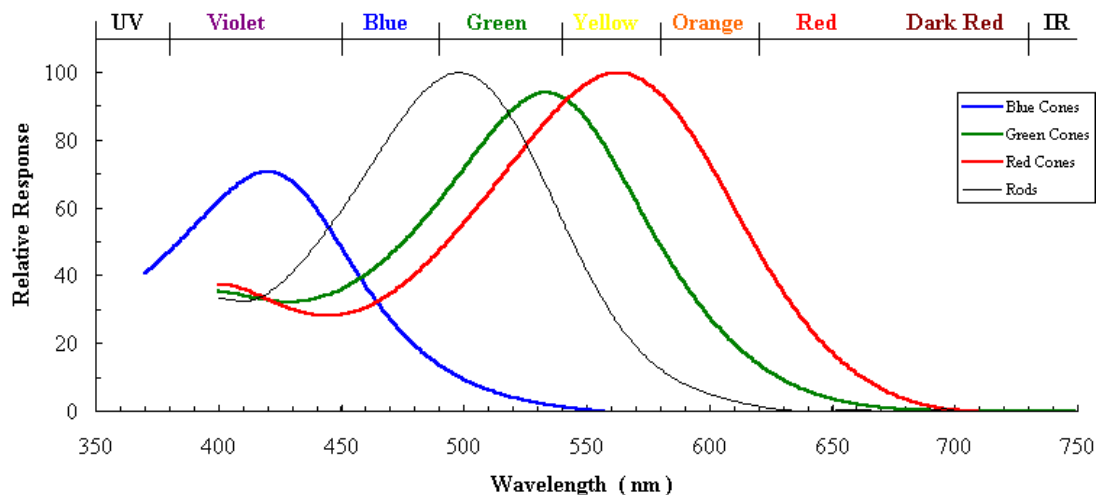
## **c) Θεωρίες χρωματικής αντίληψης και χρωματική συνταύτιση**

Κατά καιρούς έχουν διατυπωθεί διάφορες θεωρίες για την αντίληψη του χρώματος, όμως μόνο δύο έχουν επικρατήσει, κυρίως λόγω και των πειραματικών ευρημάτων που τις συνοδεύουν. Πρόκειται για τη Θεωρία της Τριχρωμίας και τη Θεωρία των ανταγωνιστικών χρωμάτων.

---

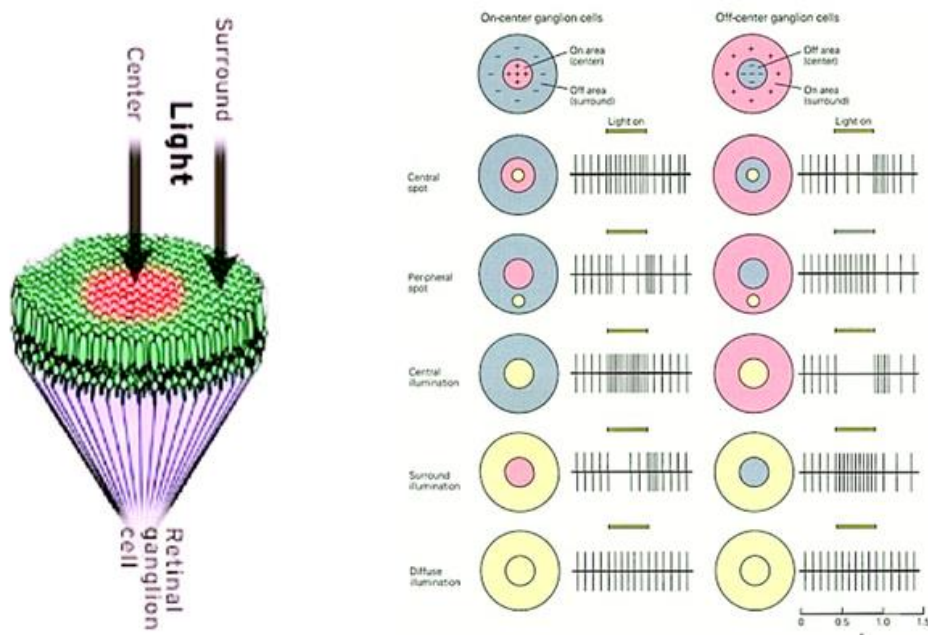
<sup>2</sup> CIE: International Commission on Illumination. Βλ. <http://www.cie.co.at/>

Η **Θεωρία της τριχρωμίας** (trichromacy) αναπτύχθηκε κυρίως από τους Helmholtz - Young - Maxwell και αναφέρει ότι κάθε άγνωστο χρώμα μπορεί να αναπαρασταθεί ως ο γραμμικός συνδυασμός τριών βασικών χρωμάτων (primary colors) τα οποία είναι μεταξύ τους γραμμικά ανεξάρτητα. Η παραπάνω θεωρία ενισχύεται και βιολογικά καθώς τα 3 είδη κωνίων έχουν το καθένα μεγαλύτερη ευαισθησία σε διαφορετικές, αλλά επικαλυπτόμενες, ζώνες στο ορατό φάσμα.



Εικόνα 6: Συναρτήσεις φασματικής ευαισθησίας ραβδίων και κωνίων. Πηγή: <https://www.unm.edu>

Η έτερη θεωρία αντίληψης χρώματος, η **Θεωρία των Ανταγωνιστικών Χρωμάτων** (color opponent process theory), αναπτύχθηκε σε πρώτη φάση από τον Hering (19<sup>ος</sup> αιώνας) και στηρίζεται στην ιδέα ότι στο σύστημα βιολογικής όρασης υπάρχουν διεργασίες οι οποίες μετρούν τη διαφορά έντασης μεταξύ κόκκινου-πράσινου, κίτρινου-μπλε και άσπρου-μαύρου. Σύμφωνα με την παραπάνω θεωρία η αντίληψη του χρώματος βασίζεται στις παραπάνω μετρήσεις διαφορών. Η θεωρία των ανταγωνιστικών χρωμάτων ενισχύθηκε αργότερα από τα πειράματα σταθερότητας χρώματος του Land και τις ανακαλύψεις νευροβιολόγων για βιολογική χωρική επεξεργασία του χρώματος με διαδικασίες τύπου center-surround processing.



Εικόνα 7: Βιολογική επεξεργασία χρώματος τύπου center-surround. Πηγή: <https://qph.ec.quoracdn.net/>

Σύμφωνα με τη θεωρία της τριχρωμίας η επιλογή 3 ανεξάρτητων χρωμάτων  $C_1(\lambda)$ ,  $C_2(\lambda)$ ,  $C_3(\lambda)$  ως κύρια (για παράδειγμα κόκκινο-πράσινο-μπλε) επιτρέπει τον προσδιορισμό οποιουδήποτε άλλου χρώματος  $F(\lambda)$  ως γραμμικού συνδυασμού των 3 κύριων. Δηλαδή:

$$F(\lambda) = a_1 C_1(\lambda) + a_2 C_2(\lambda) + a_3 C_3(\lambda)$$

Οι συντελεστές του γραμμικού συνδυασμού ( $a_1, a_2, a_3$ ) ονομάζονται **ποσοστά μίξης πρωταρχικών χρωμάτων**. Η σχέση ισότητας στον παραπάνω τύπο δεν είναι αριθμητική αλλά αναφέρεται στο «οπτικό ταίριασμα» που έχει διενεργηθεί από ανθρώπινο παρατηρητή. Για το παραπάνω ταίριασμα κομβική είναι η εργασία του Grassman ο οποίος διατύπωσε τους νόμους που το διέπουν.

Το ταίριασμα αυτό μπορεί να διακριθεί σε προσθετικό (additive) και σε αφαιρετικό (subtractive). Η πρώτη περίπτωση αφορά το «ταίριασμα» κατά το οποίο όλα τα ποσοστά ανάμιξης είναι θετικά (μη-αρνητικά), ενώ στη δεύτερη επιτρέπεται κάποια από τα ποσοστά να είναι αρνητικά. Το παραπάνω αντιστοιχεί στην πρόσθεση του υπό «ταιριάσματος» χρώματος  $F$  ενός εκ των κύριων χρωμάτων ή ακόμα και του λευκού.

Το παραπάνω ταίριασμα μπορεί να αναλυθεί και αλγεβρικά, και να προσδιοριστούν οι τρεις συνιστώσες ( $c_1, c_2, c_3$ ) της αντίληψης του χρώματος ενός ανθρώπινου παρατηρητή μέσω της πράξης του εσωτερικού γινομένου

της φασματικής κατανομής του υπό εξέταση χρώματος  $F(\lambda)$  με τις τρεις οπτικές φασματικές συναρτήσεις ευαισθησίας  $(S_1(\lambda), S_2(\lambda), S_3(\lambda))$  (visual spectral sensitivity functions) των τριών τύπων κωνίων της ρέτινας του ανθρώπινου οφθαλμού. Δηλαδή ισχύει η σχέση:

$$F(\lambda) = G(\lambda) \Leftrightarrow c(F) = c(G)$$

$$c(F) = \begin{pmatrix} c_1 \\ c_2 \\ c_3 \end{pmatrix}, c_i = \int_{\lambda_{min}}^{\lambda_{max}} F(\lambda) S_i(\lambda) d\lambda, i = 1, 2, 3$$

όπου  $\lambda_{min}, \lambda_{max}$  τα όρια της ζώνης του ορατού φάσματος [400,700] (nm).

Οι τρεις οπτικές φασματικές συναρτήσεις ευαισθησίας δημιουργούν (μέσω του συνόλου των γραμμικών τους συνδυασμών) και αποτελούν βάση του (υπο)χώρου **Human Visual Subspace (HVSS)**. Τότε για κάθε κατανομή  $F(\lambda)$  το «χρώμα» που αντιλαμβάνεται ο ανθρώπινος παρατηρητής δεν είναι παρά η προβολή  $c(F)$  της κατανομής στον υπόχωρο HVSS και παριστάνεται ως ένα 3Δ διάνυσμα, όση και η διάσταση του HVSS.

Θεωρώντας κατάλληλη διακριτοποίηση του εύρους μηκών κύματος, με κάποιο βήμα δειγματοληψίας  $\lambda_{sample} \sim 5-10\text{nm}$ , οι φασματικές κατανομές  $F(\lambda)$  καθώς και οι συναρτήσεις ευαισθησίας  $S(\lambda)$  αναπαρίστανται ως διανύσματα διατάσεων  $N \times 1$ , όπου  $N = \frac{\lambda_{max} - \lambda_{min}}{\lambda_{sample}} + 1$ . Ουσιαστικά πλέον είναι:

$$F(\lambda) \rightarrow \mathbf{f} = [F(\lambda_1), F(\lambda_2), \dots, F(\lambda_N)]^T$$

$$S(\lambda) = [S_1(\lambda), S_2(\lambda), S_3(\lambda)] \rightarrow \mathbf{S} = [\mathbf{s}_1, \mathbf{s}_2, \mathbf{s}_3]_{N \times 3}$$

Υπό την παραπάνω διακριτοποίηση τα ολοκληρώματα υποκαθίστανται από το εσωτερικό γινόμενο και των πολλαπλασιασμό πινάκων. Οπότε (στη διακριτή περίπτωση) δύο (διακριτές) φασματικές κατανομές  $\mathbf{f}, \mathbf{g}$  ταιριάζουν χρωματικά αν και μόνο αν:

$$\mathbf{f} = \mathbf{g} \Leftrightarrow \mathbf{S}^T \mathbf{f} = \mathbf{S}^T \mathbf{g}$$

Η παραπάνω σχέση υποδηλώνει ότι ο (πεπερασμένων διαστάσεων) χώρος HVSS συμπίπτει με το χώρο στηλών (column space) του πίνακα  $\mathbf{S}$ . Σε συνδυασμό των παραπάνω με τη θεωρία της τριχρωμίας προκύπτει για τις διακριτοποιημένες  $(\mathbf{c}_1, \mathbf{c}_2, \mathbf{c}_3)$  ο πίνακας  $\mathbf{C} = [\mathbf{c}_1, \mathbf{c}_2, \mathbf{c}_3]_{N \times 3}$ . Η συνθήκη της χρωματικής ανεξαρτησίας των τριών πρωταρχικών χρωμάτων πλέον εκφράζεται ως  $\det(\mathbf{S}^T \mathbf{C}) \neq 0$ . Αν ακόμα χρησιμοποιηθούν τα τρία πρωταρχικά

χρώματα ως βάση του HVSS, τότε το τυχόν χρώμα  $\mathbf{f}$  θα αντιστοιχεί στο εξής διάνυσμα  $\mathbf{m}(\mathbf{f})$ :

$$\mathbf{m}(\mathbf{f}) = (\mathbf{S}^T \mathbf{C})^{-1} \mathbf{S}^T \mathbf{f}$$

Από την παραπάνω σχέση ένα χρώμα που αναπαρίσταται (διακριτά) από  $N$  συνιστώσες αναπαρίσταται πλέον με μόλις 3 ( $\ll N$ ). Η αναπαράσταση αυτή  $\mathbf{m}(\mathbf{f}) \in \mathbb{R}^3$  αναφέρεται στη βιβλιογραφία ως **tristimulus** αναπαράσταση.

Έστω τα  $N$ -διάστατα μοναδιαία διανύσματα  $\mathbf{e}_i, i = 1, \dots, N$  τα οποία έχουν 1 στην  $i$  συνιστώσα και 0 αλλού. Τα διανύσματα αυτά είναι μονοχρωματικά φάσματα μοναδιαίας έντασης και αποτελούν διανύσματα (ορθοκανονικής) βάσης για τον χώρο όλων των διακριτών φασμάτων. Η tristimulus ( $\mathbf{u}_i$ ) αναπαράσταση των διανυσμάτων αυτών προκύπτει ως:

$$\mathbf{u}_i = (\mathbf{S}^T \mathbf{C})^{-1} \mathbf{S}^T \mathbf{e}_i$$

Για τα διανύσματα αυτά διαστάσεων  $3 \times 1$  ορίζεται ο **Πίνακας Χρωματικής Συνταύτισης (Color Matching Matrix) M**:

$$\mathbf{M} = [\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_N]^T = [\mathbf{m}_1, \mathbf{m}_2, \mathbf{m}_3]$$

Με τον παραπάνω πίνακα είναι δυνατή η συνταύτιση όλων των μονοχρωματικών μοναδιαίων κατανομών ως  $\mathbf{S}^T \mathbf{I} = \mathbf{S}^T \mathbf{C} \mathbf{M}^T$ , όπου  $\mathbf{I}$  ο μοναδιαίος πίνακας  $N \times N$ . Οι στήλες του πίνακα  $\mathbf{M}$  ( $\mathbf{m}_1, \mathbf{m}_2, \mathbf{m}_3$ ) ονομάζονται **συναρτήσεις χρωματικής συνταύτισης** (color matching functions - CMF). Τελικά ο πίνακας  $\mathbf{M}$  μπορεί να οριστεί και ως:

$$\mathbf{M} = \mathbf{S}(\mathbf{C}^T \mathbf{S})^{-1}$$

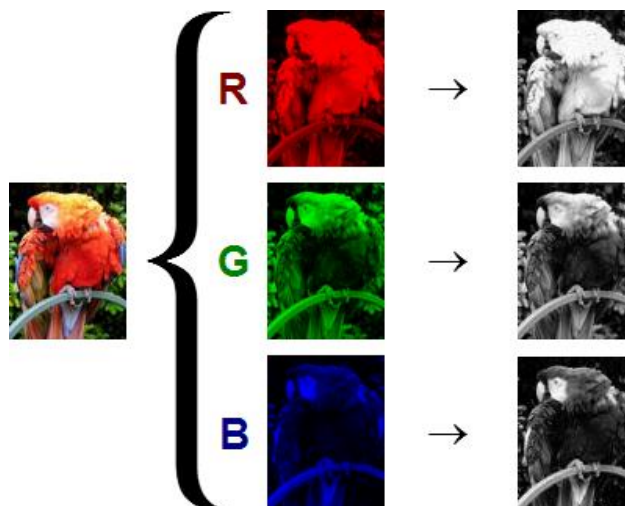
Για τον πίνακα  $\mathbf{M}$  ισχύουν οι εξής θεμελιώδεις ιδιότητες:

- 1) Οι συναρτήσεις χρωματικής συνταύτισης έχουν προκύψει από μη ιδιάζοντα μετασχηματισμό των συναρτήσεων οπτικής χρωματικής ευαισθησίας
- 2) Δύο χρώματα με φάσματα  $\mathbf{f}, \mathbf{g}$  ταιριάζουν αν  $\mathbf{M}^T \mathbf{f} = \mathbf{M}^T \mathbf{g}$
- 3) Ο υπόχωρος HVSS συμπίπτει με το χώρο στηλών του  $\mathbf{M}$

## d) Γραμμικοί χώροι χρώματος

### (i) Ο χώρος RGB

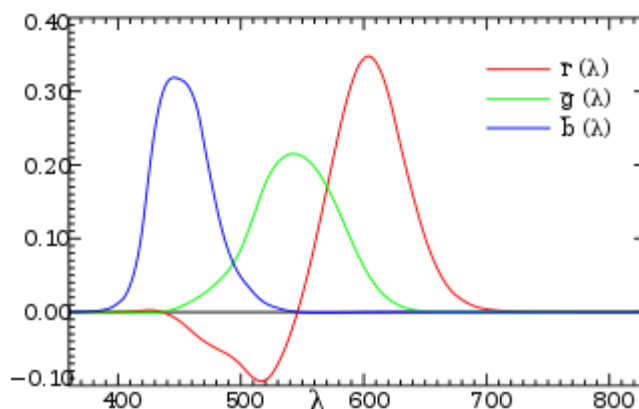
Ο χώρος RGB οφείλει το όνομά του στα τρία πρωταρχικά χρώματα που χρησιμοποιεί, δηλαδή το Κόκκινο το Πράσινο και το Μπλε. Πέραν της άμεσης σχέσης του χώρου με τη Θεωρία της τριχρωμίας, ο χώρος RGB αποτελεί ίσως τον πιο ευρέως διαδεδομένο χώρο χρώματος με πληθώρα εφαρμογών καθώς πέραν όλων των άλλων αποτελεί τον πρωτογενή χώρο καταγραφής χρωματικής πληροφορίας για το 99% των συσκευών καταγραφής εικόνας και βίντεο αλλά και απεικόνισης χρωματικής πληροφορίας για οθόνες CRT, LED, projectors, κτλ.



Εικόνα 8: Αποσύνθεση εικόνας στα τρία RGB κανάλια. Πηγή: <http://wikipedia.org>

Όπως και σε όλους τους χώρους του παρόντος εδαφίου προϋποτίθεται εδώ η ύπαρξη κατάλληλων συναρτήσεων χρωματικής συνταύτισης ώστε να καταστεί δυνατή η μέτρηση των tristimulus τιμών. Για τον χώρο RGB οι συναρτήσεις/φάσματα προέκυψαν έπειτα από πειραματικές μετρήσεις της CIE το 1931 και ονομάστηκαν *μέσες χαρακτηριστικές καμπύλες χρωματικής συνταύτισης*  $\bar{r}(\lambda)$ ,  $\bar{g}(\lambda)$ ,  $\bar{b}(\lambda)$  και αντιστοιχούν σε **πραγματικά** πρωταρχικά χρώματα. Η μέτρηση των tristimulus τιμών στο χώρο RGB γίνεται ως εξής για ένα φάσμα  $F(\lambda)$ :

$$R = \int_0^{\infty} \bar{r}(\lambda)F(\lambda)d\lambda, G = \int_0^{\infty} \bar{g}(\lambda)F(\lambda)d\lambda, B = \int_0^{\infty} \bar{b}(\lambda)F(\lambda)d\lambda$$



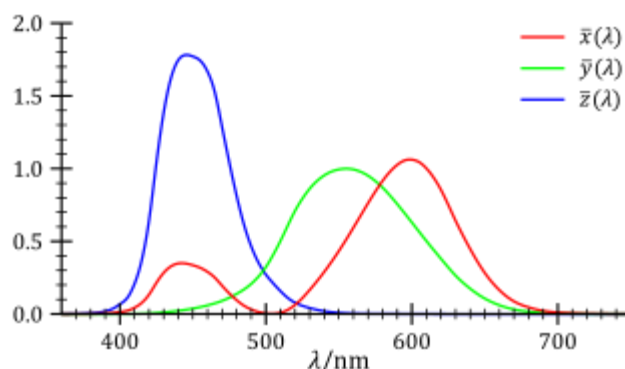
Εικόνα 9: Χαρακτηριστικές καμπύλες χρωματικής συνταύτισης χώρου RGB. Πηγή: <http://wikipedia.org>

## (ii) Ο χώρος XYZ

Παρατηρώντας τις συναρτήσεις χρωματικής συνταύτισης του χώρου RGB διαπιστώνει κανείς ότι η συνάρτηση για το χρώμα κόκκινο λαμβάνει και αρνητικές τιμές. Το πρόβλημα αυτό «διορθώνει» ο χώρος XYZ ο οποίος προκύπτει από τον RGB μέσω ενός γραμμικού μετασχηματισμού. Κατά αυτόν τον τρόπο οι 3 νέες μη-αρνητικές συναρτήσεις χρωματικής συνταύτισης  $\bar{x}(\lambda)$ ,  $\bar{y}(\lambda)$ ,  $\bar{z}(\lambda)$  αντιστοιχούν σε 3 **φανταστικά πρωταρχικά χρώματα**. Ακόμη, έχει επιλεγεί η συνάρτηση  $\bar{y}(\lambda)$  να αντιστοιχεί στη συνάρτηση *φωτεινής αποδοτικότητας* της CIE (Luminous Efficiency Function). Οι tristimulus τιμές για τον χώρο XYZ υπολογίζονται ως:

$$X = \int_0^{\infty} \bar{x}(\lambda)F(\lambda)d\lambda, Y = \int_0^{\infty} \bar{y}(\lambda)F(\lambda)d\lambda, Z = \int_0^{\infty} \bar{z}(\lambda)F(\lambda)d\lambda$$





Εικόνα 10: Χαρακτηριστικές καμπύλες χρωματικής συνταύτισης χώρου XYZ. Πηγή: <http://wikipedia.org>

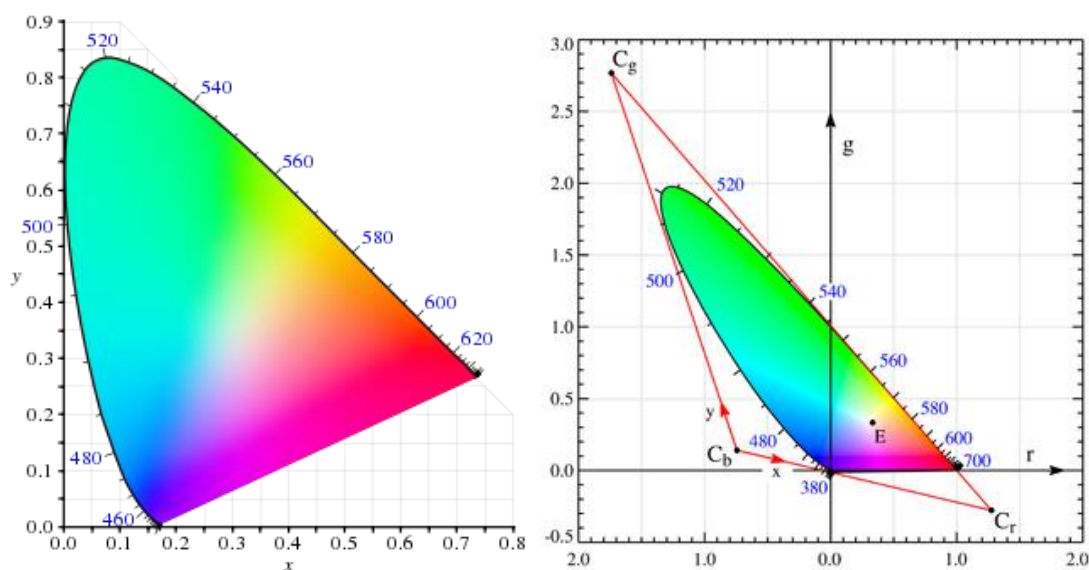
Από τις παραπάνω συνιστώσες η Y ονομάζεται «γκρίζα φωτεινότητα» (luminance). Οι συνιστώσες X,Z προσομοιάζουν περίπου τα χρώματα Κόκκινο και Μπλε. Ο γραμμικός μετασχηματισμός που συνδέει τους χώρους RGB και XYZ είναι ο εξής (Πρότυπο Rec. 709):

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0.412453 & 0.357580 & 0.180423 \\ 0.212671 & 0.715160 & 0.072169 \\ 0.019334 & 0.119193 & 0.950227 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}_{709}$$

Από τον χώρο XYZ προκύπτει το χρωματικό διάγραμμα CIE-xy. Αρχικά οι tristimulus τιμές κανονικοποιούνται ως εξής:

$$(x, y, z) = \left( \frac{X}{X+Y+Z}, \frac{Y}{X+Y+Z}, \frac{Z}{X+Y+Z} \right)$$

Η παραπάνω κανονικοποίηση δεσμεύει τις κανονικοποιημένες συνιστώσες να κείνται επί του επιπέδου  $X + Y + Z = 1$ . Η δυάδα  $(x, y)$  αναφέρεται ως δυάδα *χρωματικότητας* καθώς φέρουν μόνο τη χρωματική πληροφορία, αφού η πληροφορία της έντασης του γκριζου ενσωματώνεται στην Y-συνιστώσα. Αντίστοιχα με το CIE-xy μπορεί να οριστεί το χρωματικό διάγραμμα CIE-rg από τον χώρο RGB. Στα χρωματικά διαγράμματα η χρωματισμένη περιοχή αντιστοιχεί στο σύνολο των χρωματικότητας που είναι αντιληπτές από το ανθρώπινο σύστημα όρασης και αναφέρεται διεθνώς ως **Gamut of Human Vision (GHV)**.



Εικόνα 11: Το φάσμα της ανθρώπινης όρασης στα χρωματικά διαγράμματα  $xy$  και  $rg$ . Πηγή: <http://wikipedia.org>

### (iii) Ο χώρος CMY(K)

Ο χώρος CMY είναι ο συμπληρωματικός του χώρου RGB. Αυτό συμβαίνει διότι τα 3 χρώματα Κυανό (C), Ματζέντα (M) και Κίτρινο (Y) αποτελούν τα συμπληρωματικά των Κόκκινο, Πράσινο και Μπλε, αντίστοιχα. Συνεπώς οι tristimulus τιμές για τον χώρο CMY προκύπτουν από τις αντίστοιχες του RGB ως:

$$(C, M, Y) = (1 - R, 1 - G, 1 - B)$$

Οι εφαρμογές του CMY είναι πολλές και αναφέρονται σε περιπτώσεις «χρωματισμού» μίας λευκής επιφάνειας (αφαιρετικά κύρια χρώματα) σε αντίθεση με τον RGB που έχει ως βάση το «μαύρο» (προσθετικά κύρια χρώματα). Τέτοιες εφαρμογές περιλαμβάνουν την εκτύπωση σε όλες τις μορφές. Για καλύτερη ποιότητα αναπαράστασης χρώματος σε πρακτικές εφαρμογές ο CMY επαυξάνεται με μία μαύρη συνιστώσα (K) και προκύπτει ο χώρος CMYK.

## ε) Χώροι χρώματος με χρωματικότητες

### (i) Ο χώρος Ycber

Ως στόχος του χώρου χρώματος Ycber είναι οι tristimulus τιμές να προκύπτουν από όσο το δυνατόν «ασυσχέτιστες συναρτήσεις συνταύτισης» (στατιστική ορθογωνιότητα) σε αντίθεση με τον βασικό χώρο RGB όπου οι 3 συνιστώσες παρουσιάζουν υψηλή συσχέτιση. Όπως και στον χώρο XYZ η συνιστώσα Y εκφράζει την «γκρίζα φωτεινότητα», ενώ οι άλλες δύο (cb-cr) είναι συνιστώσες χρωματικότητας, δηλαδή εμπεριέχουν μόνο χρωματική πληροφορία. Γενικά, ο μετασχηματισμός από τον RGB στον Ycber είναι κατά κύριο λόγο γραμμικός, όμως οι RGB συνιστώσες έχουν υποστεί προεπεξεργασία κατά μη-γραμμικό τρόπο μέσω της διαδικασίας του gamma-correction<sup>3</sup>. Αρχικά, ο χώρος αυτός χρησιμοποιήθηκε, υπό την ονομασία YPbPr, στο αναλογικό βίντεο με τον κάτωθι μετασχηματισμό:

$$\begin{aligned} Y' &= k_r R' + (1 - k_r - k_b) G' + k_b B' \\ P_b &= (B' - Y') / 2(1 - k_b) \\ P_r &= (R' - Y') / 2(1 - k_r) \end{aligned}$$

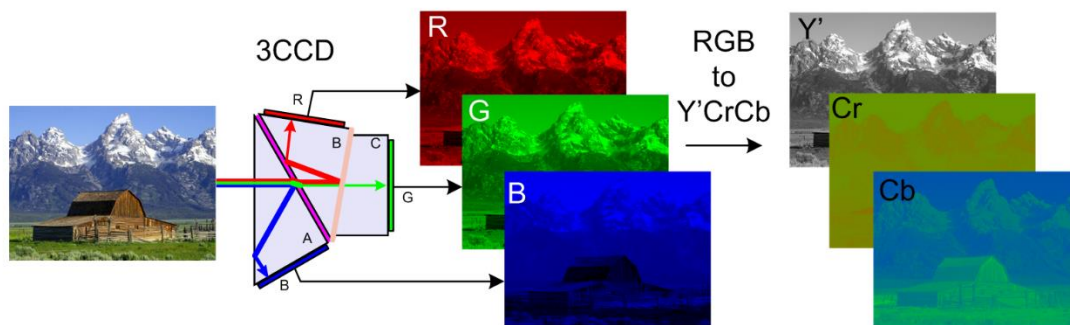
όπου οι τονούμενες συνιστώσες (R',G',B') είναι οι gamma-corrected εκδόσεις των αρχικών με [Y',R',G',B'] ∈ [0,1] και [Pb,Pr] ∈ [-0.5,0.5], με τις (k<sub>r</sub>, k<sub>b</sub>) να είναι σταθερές εξαρτώμενες από την αρχική προτυποποίηση (κατά CIE) του χώρου RGB. Για παράδειγμα, για το πρότυπο Rec.709(HDTV) λαμβάνουν τις τιμές 0,2126 και 0,0722, αντίστοιχα. Ο ψηφιακός χώρος YCbCr προκύπτει από τον YPbPr μέσω κλιμάκωσης και μετάθεσης ως:

$$\begin{aligned} Y &= 16 + 219Y' \\ C_b &= 128 + 224P_b \\ C_r &= 128 + 224P_r \end{aligned}$$

Τελικώς η συνιστώσα Y αναπαριστά την ολική φωτεινότητα, ενώ οι άλλες δύο, C<sub>b</sub> - C<sub>r</sub>, χρωματικές διαφορές. Ο χώρος YCbCr χρησιμοποιείται ευρέως σε συμπιεσμένα φορμάτ εικόνων και βίντεο όπως π.χ. τα JPEG και MPEG.

---

<sup>3</sup> Βλ. [https://en.wikipedia.org/wiki/Gamma\\_correction](https://en.wikipedia.org/wiki/Gamma_correction)



Εικόνα 12: Ο χώρος YCbCr. Πηγή: <http://wikimedia.org>

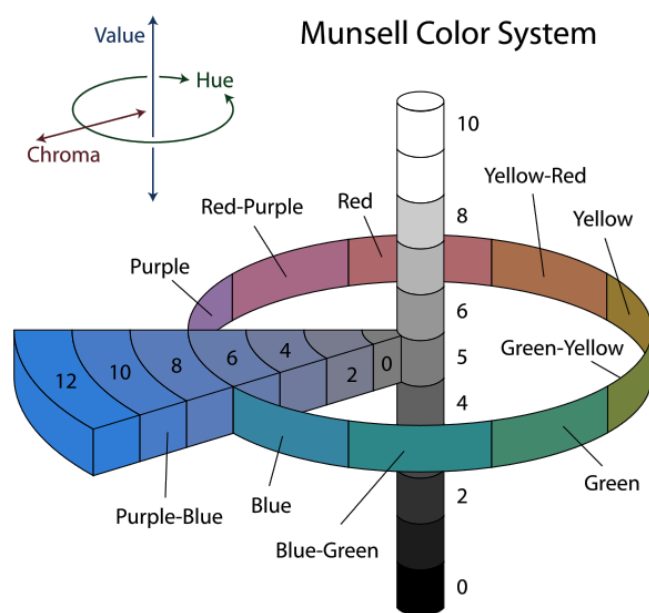
## (ii) Χώροι με πολικές χρωματικές συνιστώσες (HSV/HSI/HSL)

Σύμφωνα με πρόσφατο χρωματικό μοντέλο της CIE, το CIECAM02 (εκδόθηκε το 2002, υιοθετήθηκε το 2008), η αντίληψη του χρώματος από ανθρώπινους παρατηρητές δύναται να ποσοτικοποιηθεί από έξι ψυχολογικά χαρακτηριστικά/αισθήσεις:

1. **Brightness** (Λαμπρότητα): Καθορίζει το πόσο πολύ (ή λίγο) φως «φαίνεται» να εκπέμπει μία επιφάνεια
2. **Lightness** (Φωτεινότητα): Ορίζεται ως η σχετική λαμπρότητα μιας επιφάνειας ως προς μία λευκή επιφάνεια υπό τις ίδιες συνθήκες φωτισμού
3. **Hue** (Απόχρωση): Η ομοιότητα στην εμφάνιση μιας επιφάνειας σε σχέση με κάποιο από τα χρώματα κόκκινο, πράσινο κίτρινο και μπλε ή κάποιον ανά δύο συνδυασμό τους
4. **Colorfulness** (Ζωηροχρωμία): Καθορίζει το πόσο πολύ (ή λίγο) μια επιφάνεια «φαίνεται» γεμάτη από χρώμα
5. **Chroma** (Χρώμα): Ορίζεται ως η σχετική ζωηροχρωμία μιας επιφάνειας ως προς την λαμπρότητα μιας λευκής υπό τις ίδιες συνθήκες φωτισμού
6. **Saturation** (Κορεσμός): Ορίζεται ως η σχετική ζωηροχρωμία μιας επιφάνειας ως προς τη λαμπρότητά της

Οι παραπάνω ιδιότητες ώθησαν την διεθνή κοινότητα στην ανάπτυξη χώρων χρώματος των οποίων (κάποιες) συνιστώσες να έχουν άμεση συσχέτιση με κάποια ή κάποιες εκ των άνωθεν ιδιοτήτων. Η αρχή για τη δημιουργία τέτοιων χώρων έγινε κατά τις αρχές του 20ου αιώνα από τον

Munsell (REF) και τον ομώνυμο χώρο. Ο χώρος χρώματος του Munsell έχει 3 συνιστώσες, με κάθε μία να αντιστοιχεί στις ιδιότητες Hue, Lightness και Chroma. Τέτοιου τύπου χώροι χρώματος δεν εμπίπτουν στα ενδιαφέροντα της παρούσας εργασίας. Για λόγους πληρότητας θα αναφερθούν οι κυριότεροι χώροι χρώματος του είδους, ενώ ο αναγνώστης παραπέμπεται στη διεθνή βιβλιογραφία για περισσότερες λεπτομέρειες επί του (άκρως ενδιαφέροντος) θέματος.

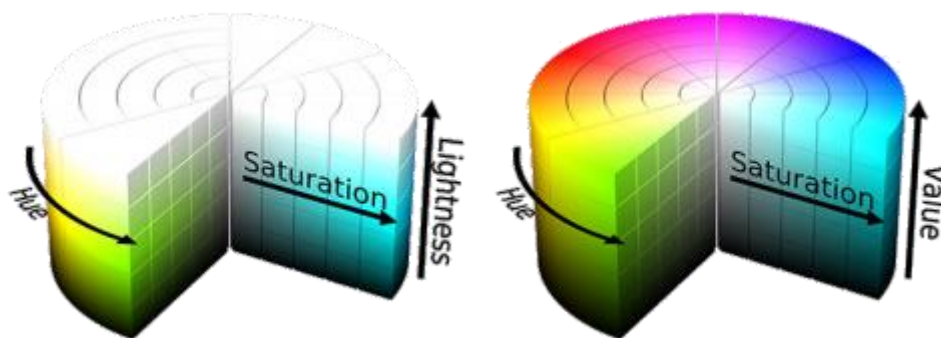


Εικόνα 13: Ο χώρος χρώματος του Munsell. Πηγή: <https://wikipedia.org>

Παρακάτω παρατίθεται οι σημαντικότεροι χώροι χρώματος με πολικές συντεταγμένες:

- HSV – Hue, Saturation, Value (Lightness)
- HSL – Hue, Saturation, Lightness
- HSI – Hue, Saturations, Intensity

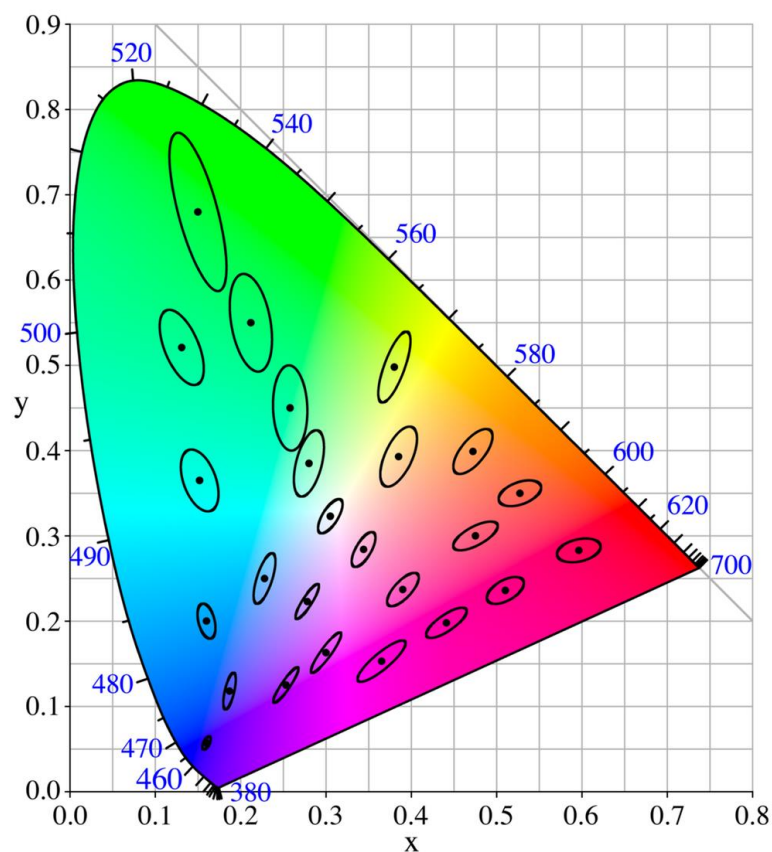
Οι παραπάνω χώροι χαρακτηρίζονται για τις δύο κοινές τους συνιστώσες. Η συνιστώσα H (Απόχρωση) μετριέται σε μονάδες γωνίας (μοίρες/ακτίνια) ενώ η συνιστώσα S (Κορεσμός) είναι βαθμωτό μέγεθος. Η τρίτη συνιστώσα αντιστοιχεί στην ολική φωτεινότητα με μικρές διαφοροποιήσεις. Για περισσότερες πληροφορίες για τους παραπάνω χώρους χρώματος μπορεί κανείς να διαβάσει στο [8].



Εικόνα 14: Οι χώροι HSL και HSV. Πηγή: <http://www.beliefmedia.com>

### (iii) Ομοιόμορφοι χώροι χρώματος

Η έννοια των χρωματικοτήτων που εισήχθη κατά την παρουσίαση του χώρου χρώματος XYZ επιτρέπει τη μέτρηση των χρωματικών «αποστάσεων» στο επίπεδο των χρωματικοτήτων. Δυστυχώς όμως η Ευκλείδεια απόσταση μεταξύ δύο χρωμάτων στο xy-επίπεδο των χρωματικοτήτων δεν αντιστοιχεί στην διαφορά αντίληψης των δύο χρωμάτων από τον άνθρωπο με τρόπο ομοιόμορφο σε όλο το xy-επίπεδο. Προκύπτει λοιπόν η ανάγκη χώρων χρώματος στους οποίους η απόσταση δύο σημείων θα παρέχει ασφαλείς ενδείξεις για την «οπτική» διαφορά αντίληψης των δύο χρωμάτων. Οι χώροι αυτοί ονομάζονται **ομοιόμορφοι** και **ισότροποι** χώροι χρώματος. Σε τέτοιου είδους χώρους ορίζεται η έννοια της «μόλις διακρινόμενης διαφοράς» (Just Noticeable Difference - JND) που πρόκειται για ένα κατώφλι στις αποστάσεις κάτω από το οποίο δύο χρώματα είναι πρακτικά ισοδύναμα στον ανθρώπινο παρατηρητή. Στην ανάπτυξη τέτοιων χώρων συνέβαλε χαρακτηριστικά η έρευνα του McAdam κατά τη δεκαετία του 1940 η οποία οδήγησε στην έννοια της έλλειψης McAdam. Οι πρώτοι ομοιόμορφοι και ισότροποι χώροι χρώματος αναπτύχθηκαν κατά το 1976 και πρόκειται για τους CIE Lab και CIE Luv, οι οποίοι είναι σαφώς πιο ομοιόμορφοι σε σχέση με τον RGB και τον XYZ.



Εικόνα 15: Ελλείψεις MacAdam του χώρου RGB στο xy-χρωματικό διάγραμμα.  
Πηγή: <http://wikipedia.org>

Ο χώρος CIE<sub>Lu</sub>ν προκύπτει έπειτα από έναν μη γραμμικό μετασχηματισμό του χώρου XYZ ως εξής:

$$L^* = 116g\left(\frac{Y}{Y_w}\right) - 16$$

$$u^* = 13L^*(u' - u_w)$$

$$v^* = 13L^*(v' - v_w)$$

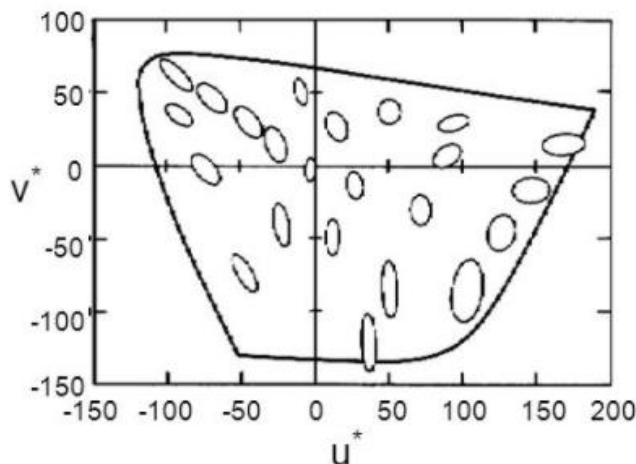
όπου ο υποδείκτης  $w$  δηλώνει το λευκό σημείο ( $R=G=B=1$  στον RGB) και τα τονούμενα τον εξής προοπτικό μετασχηματισμό:

$$u' = \frac{4X}{X + 15Y + 3Z}$$

$$v' = \frac{9X}{X + 15Y + 3Z}$$

Ενώ η συνάρτηση  $g: \mathbb{R} \rightarrow \mathbb{R}$  ορίζεται ως:

$$g(k) = \begin{cases} k^{\frac{1}{3}}, & k > \left(\frac{6}{29}\right)^3 \\ \frac{1}{3} \left(\frac{29}{6}\right)^2 k + \frac{4}{29}, & k \leq \left(\frac{6}{29}\right)^3 \end{cases}$$

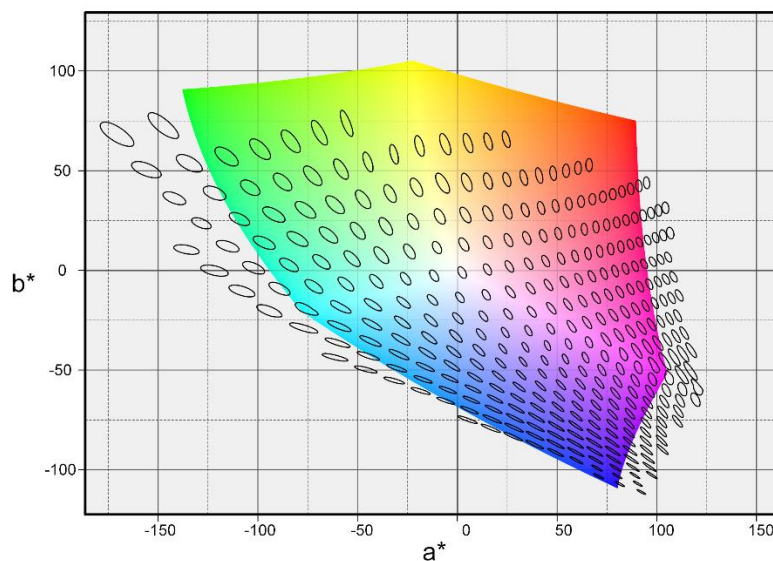


Εικόνα 16: Ελλείψεις McAdam του χώρου CIE<sub>Luv</sub> στο uv-χρωματικό διάγραμμα.  
Πηγή: Από διαφάνεια του Idit Haran

Ο έτερος ομοιόμορφος χώρος είναι ο **CIE Lab**. Πρόκειται για έναν πολύ δημοφιλή χώρο χρώματος με τεράστιες εφαρμογές στην επιστήμη της Όρασης Υπολογιστών. Στην παρούσα μεταπτυχιακή εργασία γίνεται εκτενής χρήση του χώρου τόσο κατά τη δημιουργία των υπερεικονοστοιχείων (βλ. παρακάτω) όσο και ως ο βασικός χώρος χρωματικής αναπαράστασης για το σύνολο των αναπτυχθέντων μοντέλων εμφάνισης ανθρώπινων στόχων. Ο χώρος CIE Lab, όπως και ο CIE<sub>Luv</sub>, προκύπτει από τον XYZ μέσω του ακόλουθου μη-γραμμικού μετασχηματισμού:

$$\begin{aligned} L^* &= 116g\left(\frac{Y}{Y_w}\right) - 16 \\ a^* &= 500 \left[ g\left(\frac{X}{X_w}\right) - g\left(\frac{Y}{Y_w}\right) \right] \\ b^* &= 500 \left[ g\left(\frac{Y}{Y_w}\right) - g\left(\frac{Z}{Z_w}\right) \right] \end{aligned}$$





Εικόνα 17: Ελλείψεις McAdam του χώρου CIE Lab στο  $ab$ -χρωματικό διάγραμμα.

Πηγή: <http://n-colorspace.cool.coocan.jp>

Για τον χώρο CIE Lab οι τιμές της  $L^*$  συνιστώσας (φωτεινότητα) ανήκουν στο διάστημα  $[0,100]$  (μαύρο-άσπρο) με τη μοναδιαία μεταβολή να βρίσκεται στο JND, ενώ οι συνιστώσες χρωματικότητας  $a^*$  και  $b^*$  παριστάνουν τα χρώματα ως εξής:

- $a^* > 0$  δηλώνει κοκκινωπό χρώμα
- $a^* < 0$  δηλώνει πρασινωπό χρώμα
- $b^* > 0$  δηλώνει κιτρινωπό χρώμα
- $b^* < 0$  δηλώνει μπλε χρώμα
- $a^* = 0, b^* = 0$  δηλώνει το αχρωματικό γκριζο

Από τα παραπάνω γίνεται εμφανές ότι ο χώρος CIE Lab στηρίζεται στη Θεωρία των ανταγωνιστικών χρωμάτων καθώς οι χρωματικότητες του χώρου μετρούν χρωματικές διαφορές.

### 3) Υπερ-εικονοστοιχεία (super-pixels)

#### a) Ορισμός υπερεικονοστοιχείων

Ως **υπερεικονοστοιχείο** ονομάζεται ένα σύνολο **συνδεδεμένων**<sup>4</sup> εικονοστοιχείων τα οποία παρουσιάζουν **υψηλή χρωματική ομοιογένεια**. Η διαδικασία της δημιουργίας υπερεικονοστοιχείων θα μπορούσε να θεωρηθεί μια μορφή ομαδοποίησης (clustering) ή υπερκατάτμησης (oversegmentation) της εικόνας. Στη συνέχεια ακολουθεί αναλυτική περιγραφή του δημοφιλούς αλγορίθμου δημιουργίας υπερεικονοστοιχείων, βασισμένη κατά κύριο λόγο στη διπλωματική εργασία της Ψάλτα Αθηνάς, Ε.Μ.Π. 2015 [14]. Για λόγους πληρότητας να σημειωθεί ότι στη βιβλιογραφία αναφέρονται και άλλες μέθοδοι δημιουργίας υπερεικονοστοιχείων όπως ο αλγόριθμος TurboPixels [8].

#### b) Simple Linear Iterative Clustering (SLIC)

Μία μέθοδος δημιουργίας υπερεικονοστοιχείων σε μία εικόνα που έχει αναπτυχθεί πρόσφατα από το πανεπιστήμιο της Λοζάνης (EPFL) αποτελεί ο αλγόριθμος SLIC (Simple Linear Iterative Clustering) υποσχόμενη καλύτερη αποδοτικότητα και λιγότερο υπολογιστικό χρόνο σε σχέση με άλλες συνήθεις μεθόδους κατάτμησης μιας εικόνας σε υπερεικονοστοιχεία [7]. Αυτός ο αλγόριθμος ουσιαστικά αποτελεί μία διαφορετική προσέγγιση του ευρέως διαδεδομένου αλγορίθμου ομαδοποίησης δεδομένων K-means προκειμένου να δημιουργηθούν τα υπερεικονοστοιχεία στην εικόνα με τις εξής, όμως, διαφοροποιήσεις :

- Ο αριθμός των υπολογισμών που πραγματοποιούνται ως προς τις αποστάσεις μεταξύ των εικονοστοιχείων μειώνεται δραματικά σε σχέση με τον K-means, διότι ο SLIC περιορίζει τον χώρο αναζήτησης σε μία περιοχή μεγέθους ανάλογου με το μέγεθος του υπερεικονοστοιχείου. Όπως γίνεται κατανοητό, λοιπόν, η χρονική πολυπλοκότητα

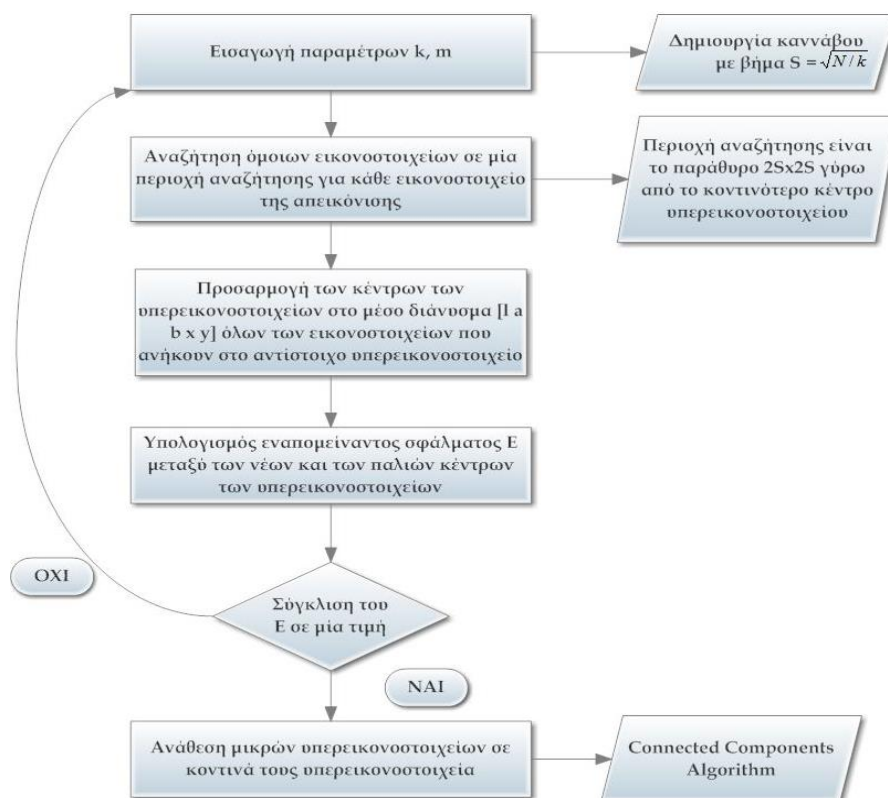
---

<sup>4</sup> Η συνδεσιμότητα των εικονοστοιχείων είναι αρκετά διαδεδομένη ως έννοια στην επεξεργασία δυαδικών εικόνων (bitmaps) και διακρίνεται σε 4- και 8-συνδεσιμότητα (connectivity). Για περισσότερες πληροφορίες βλ. [https://en.wikipedia.org/wiki/Pixel\\_connectivity](https://en.wikipedia.org/wiki/Pixel_connectivity)

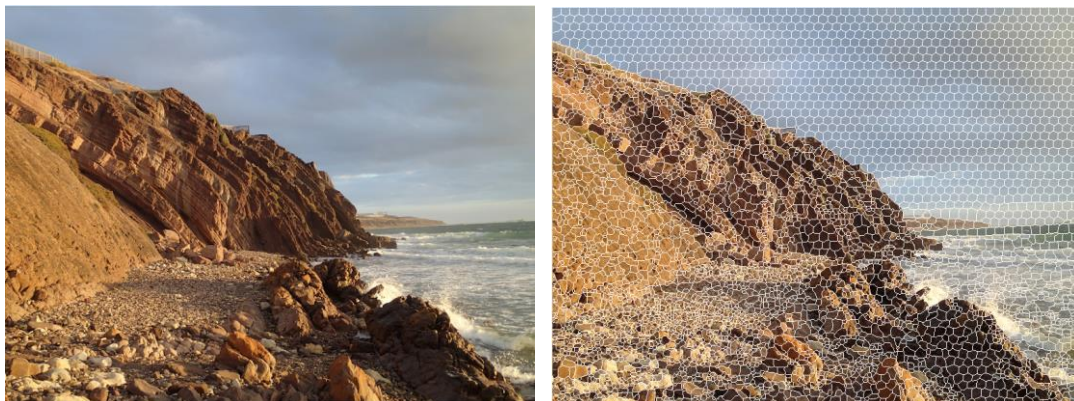
του αλγορίθμου είναι ανεξάρτητη του αριθμού των υπερεικονοστοιχείων που θα δημιουργηθούν στην εικόνα, αλλά αυξάνεται πλέον γραμμικά αναλόγως του αριθμού των εικονοστοιχείων της εικόνας.

- Αντί για την κλασσική έννοια της απόστασης, σε αυτόν τον αλγόριθμο δίνεται η δυνατότητα για τη χρήση μίας σταθμισμένης απόστασης που συνδυάζει τη χωρική και φασματική εγγύτητα, η οποία μπορεί με αυτό τον τρόπο να παρέχει έλεγχο στον χρήστη σχετικά με το μέγεθος και το κατά πόσο είναι συμπαγή τα υπερεικονοστοιχεία που έχουν δημιουργηθεί στην εικόνα.

Ενδεικτικά, τα βήματα του αλγορίθμου εμφανίζονται στη παρακάτω εικόνα :



Εικόνα 18: Τα βασικά βήματα του αλγορίθμου κατάτμησης SLIC. Πηγή: Διπλωματική εργασία Αθηνάς Ψάλτα, 2015, Σ.Α.Τ.Μ. Ε.Μ.Π.

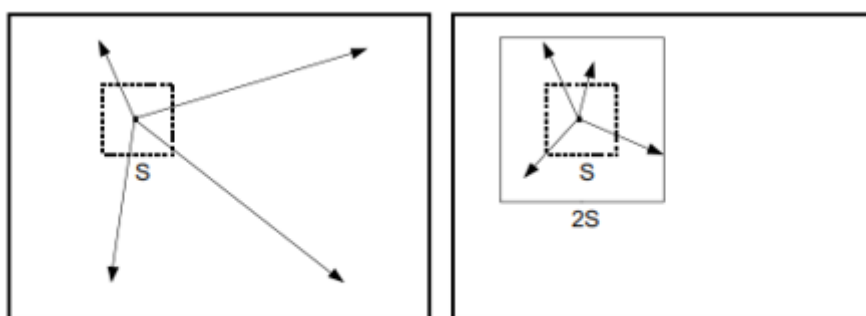


Εικόνα 19: Αποτέλεσμα εφαρμογής του SLIC σε μία εικόνα RGB. Πηγή: Hallett Cove, South Australia)

Η βασική μορφή του αλγορίθμου SLIC εφαρμόζεται σε εικόνες RGB και στο σημείο αυτό πρόκειται να αναλυθεί εκτενώς η μεθοδολογία αυτής της τεχνικής κατάτμησης. Εξ ορισμού, μία από τις δύο παραμέτρους του αλγορίθμου είναι ο κατά προσέγγιση αριθμός  $k$  των υπερεικονοστοιχείων που πρόκειται να δημιουργηθούν στην εικόνα ανάλογα με την επιθυμία του χρήστη. Η βασική μορφή του SLIC πραγματοποιείται στον χρωματικό χώρο CIELAB και η διαδικασία της κατάτμησης της εικόνας ξεκινά με ένα στάδιο αρχικοποίησης, στο οποίο λαμβάνονται δειγματοληπτικά σε πρώτη φάση  $k$  κέντρα των υπερεικονοστοιχείων πάνω σε ένα κάνναβο που έχει σχηματιστεί με βήμα  $S$ . Προκειμένου να δημιουργηθούν περίπου ισομεγέθη υπερεικονοστοιχεία, το βήμα του καννάβου ορίζεται ως  $S = \sqrt{N/k}$  και τα κέντρα των υπερεικονοστοιχείων μετακινούνται εντός ενός παραθύρου  $3 \times 3$  στο σημείο στο οποίο οι κλίσεις είναι ελάχιστες (lowest gradient position). Αυτό συμβαίνει ώστε αφενός να αποφευχθεί η περίπτωση στην οποία το κέντρο ενός υπερεικονοστοιχείου θα τοποθετηθεί σε μία ακμή της εικόνας, αφετέρου για να μειωθεί ο κίνδυνος να συμπεριληφθεί ένα εικονοστοιχείο που εμπεριέχει θόρυβο σε ένα οποιοδήποτε υπερεικονοστοιχείο.

Το επόμενο βήμα του αλγορίθμου είναι το στάδιο της ανάθεσης των εικονοστοιχείων, όπου κάθε εικονοστοιχείο της απεικόνισης σχετίζεται με το κοντινότερο κέντρο υπερεικονοστοιχείου που βρίσκεται εντός μιας περιοχής αναζήτησης, όπως παρουσιάζεται στην Εικόνα. Ο περιορισμός της περιοχής αναζήτησης μειώνει σε σημαντικό βαθμό τον αριθμό των υπολογισμών των αποστάσεων μεταξύ των εικονοστοιχείων και των κέντρων των υπερεικονοστοιχείων, με αποτέλεσμα ο αλγόριθμος να έχει μικρό υπολογιστικό χρόνο σε σχέση με τον συμβατικό αλγόριθμο K-means, στον οποίο οι αποστάσεις αυτές πρέπει να υπολογιστούν για όλα τα εικονοστοιχεία της

απεικόνισης. Εφόσον η προσδοκώμενη χωρική έκταση ενός υπερεικονοστοιχείου στην εικόνα είναι μία περιοχή μεγέθους  $S \times S$ , η αναζήτηση για όμοια εικονοστοιχεία πραγματοποιείται σε μία γειτονιά  $2S \times 2S$  γύρω από το κέντρο του υπερεικονοστοιχείου. Αξίζει να αναφερθεί ότι, περιορίζοντας την αναζήτηση κατά τη διαδικασία της ομαδοποίησης, ο SLIC αποφεύγει να εκτελέσει χιλιάδες πλεονάζοντες υπολογισμούς αποστάσεων, διότι στη πράξη ένα εικονοστοιχείο μπορεί να βρίσκεται εντός μιας γειτονιάς λιγότερων από οκτώ κέντρων υπερεικονοστοιχείων. Αυτό σημαίνει ότι η χρονική πολυπλοκότητα του αλγορίθμου εξαρτάται γραμμικά από τον αριθμό  $N$  των εικονοστοιχείων της εικόνας ανεξαρτήτως του μεγέθους της τιμής της παραμέτρου  $k$  που εισάγει ο χρήστης.



(α) Περιοχή αναζήτησης του *K-means*      (β) Περιοχή αναζήτησης του *SLIC*

Εικόνα 20: Παρουσίαση των περιοχών αναζήτησης υπερεικονοστοιχείων στη περίπτωση (α) του κλασσικού αλγορίθμου *K-means* και (β) του *SLIC*.

Αφού, λοιπόν, κάθε εικονοστοιχείο έχει συσχετιστεί με το κοντινότερο κέντρο ενός υπερεικονοστοιχείου, το επόμενο βήμα του αλγορίθμου αφορά την προσαρμογή αυτών των κέντρων στο μέσο διάνυσμα  $[l \ a \ b \ x \ y]^T$  όλων των εικονοστοιχείων που ανήκουν στο αντίστοιχο υπερεικονοστοιχείο. Για τον υπολογισμό του εναπομείναντος σφάλματος  $E$  μεταξύ των καινούργιων και των προηγούμενων κέντρων των υπερεικονοστοιχείων που έχουν δημιουργηθεί χρησιμοποιείται η  $L2$  νόρμα. Τα βήματα της ανάθεσης των εικονοστοιχείων της απεικόνισης σε κάποιο υπερεικονοστοιχείο και της προσαρμογής των κέντρων των υπερεικονοστοιχείων μπορούν να επαναληφθούν πολλές φορές μέχρι το εναπομένον σφάλμα  $E$  να συγκλίνει σε μία τιμή.

Το τελευταίο στάδιο του αλγορίθμου αφορά την ενίσχυση της συνδεσιμότητας των εικονοστοιχείων μέσω της συνεχόμενης ανάθεσης ορισμένων

εικονοστοιχείων σε κοντινά τους υπερεικονοστοιχεία. Πιο συγκεκριμένα, μετά το τέλος της διαδικασίας της δημιουργίας των υπερεικονοστοιχείων είναι απολύτως αναμενόμενο ορισμένα από τα εικονοστοιχεία της εικόνας να μην ανήκουν σε κάποιο από τα υπερεικονοστοιχεία που έχουν δημιουργηθεί, αλλά το καθένα από αυτά να αποτελεί ένα μοναδικό υπερεικονοστοιχείο. Προκειμένου, λοιπόν, να μην δημιουργούνται τόσο μικρού μεγέθους υπερεικονοστοιχεία, αυτά τα εικονοστοιχεία ανατίθενται στα κοντινότερα κέντρα των υπερεικονοστοιχείων μέσω ενός αλγορίθμου σύνδεσης συνιστωσών (connected components algorithm).

### **(i) Ο ρόλος της απόστασης**

Τα υπερεικονοστοιχεία που δημιουργούνται από τον SLIC αντιστοιχούν σε αντικείμενα της εικόνας που αναφέρονται στον χρωματικό χώρο του CIE Lab με αποτέλεσμα η μέτρηση της απόστασης  $D$  μεταξύ των εικονοστοιχείων της εικόνας και των κέντρων των υπερεικονοστοιχείων να μην είναι άμεσα προφανής. Το εύρος των χρωματικών τιμών ενός εικονοστοιχείου σε αυτό τον χώρο είναι ένα ήδη γνωστό διάστημα, ενώ οι εικονοσυντεταγμένες του μπορούν να έχουν ένα εύρος τιμών ανάλογα με το μέγεθος της εικόνας. Συνεπώς, είναι προφανές, ότι ορίζοντας κανείς την απόσταση  $D$  στον ευκλείδειο χώρο LABXY που έχει 5 διαστάσεις, αναμένει να δημιουργηθούν ασυνέχειες στον τρόπο με τον οποίο πραγματοποιείται η κατάτμηση της εικόνας για διαφορετικά μεγέθη υπερεικονοστοιχείων. Αναλυτικότερα, για μεγάλα υπερεικονοστοιχεία οι χωρικές διαφορές θα υπερτερούν των φασματικών διαφορών δίνοντας έτσι πιο μεγάλη σημασία στη χωρική εγγύτητα σε σχέση με τη φασματική. Ως αποτέλεσμα σε αυτή τη περίπτωση πρόκειται να δημιουργηθούν συμπαγή υπερεικονοστοιχεία που, όμως, δεν θα προσαρμόζονται ικανοποιητικά στα όρια της εικόνας. Προφανώς για τα μικρού μεγέθους υπερεικονοστοιχεία που δημιουργούνται ισχύει ακριβώς το αντίστροφο.

Ο τρόπος με τον οποίο ο αλγόριθμος αντιμετωπίζει αυτό το πρόβλημα είναι η χρήση μιας απόστασης που συνδυάζει τόσο το χωρικό όσο και το φασματικό στοιχείο της εικόνας. Αυτό επιτυγχάνεται μέσω της ομαλοποίησης της χωρικής εγγύτητας  $N_c$  και της φασματικής εγγύτητας  $N_s$  από τις αντίστοιχες μέγιστες αποστάσεις τους μέσα σε ένα υπερεικονοστοιχείο, συνεπώς η απόσταση  $D'$  εκφράζεται ως εξής :

$$\begin{aligned}d_c &= \sqrt{(l_j - l_i)^2 + (a_j - a_i)^2 + (b_j - b_i)^2} \\d_s &= \sqrt{(x_j - x_i)^2 + (y_j - y_i)^2} \\D' &= \sqrt{\left(\frac{d_c}{N_c}\right)^2 + \left(\frac{d_s}{N_s}\right)^2}.\end{aligned}\quad (2.10)$$

Η μέγιστη χωρική απόσταση που αναμένεται μέσα σε ένα υπερεικονοστοιχείο πρέπει να αντιστοιχεί στο διάστημα της δειγματοληψίας  $N_s = N_c = \sqrt{N/K}$ . Ο καθορισμός της μέγιστης φασματικής απόστασης  $N_c$  δεν είναι μια απλή διαδικασία διότι μία τέτοια απόσταση μπορεί να ποικίλλει σημαντικά από το ένα υπερεικονοστοιχείο στο άλλο και από εικόνα σε εικόνα, συνεπώς η (3.1) μπορεί να γραφτεί ως :

$$D = \sqrt{d_c^2 + \left(\frac{d_s}{S}\right)^2} m^2 \quad (2.11)$$

Η σταθερά  $m$  που αναγράφεται στη παραπάνω σχέση εισάγεται από τον χρήστη δίνοντάς του τη δυνατότητα να ορίσει ο ίδιος τη σχετική σημασία μεταξύ χωρικής και φασματικής εγγύτητας και μπορεί να λάβει όλες τις τιμές εντός του διαστήματος [1,40] όταν χρησιμοποιείται ο χρωματικός χώρος του CIELAB. Η χωρική εγγύτητα είναι πιο σημαντική όταν το  $m$  είναι μεγάλο με αποτέλεσμα τα υπερεικονοστοιχεία που δημιουργούνται να είναι αρκετά συμπαγή, ενώ όταν το  $m$  έχει μικρή τιμή τα υπερεικονοστοιχεία έχουν λιγότερο κανονικό μέγεθος και σχήμα, όμως προσαρμόζονται καλύτερα στα όρια της εικόνας.

## ΚΕΦΑΛΑΙΟ 3.

# ΜΟΝΤΕΛΑ ΕΜΦΑΝΙΣΗΣ

Τα μοντέλα εμφάνισης (appearance models) αποτελούν βασική συνιστώσα σε αρκετές μεθόδους και εφαρμογές της Όρασης Υπολογιστών. Ως υποσυστήματα αλγορίθμων τα μοντέλα εμφάνισης σπάνια εμφανίζονται αυτόνομα, όμως ο ρόλος τους είναι, σχεδόν σε κάθε περίπτωση εφαρμογής τους, κρίσιμος για την αποτελεσματικότητα της ανάλυσης των δεδομένων. Στη συνέχεια του Κεφαλαίου θα κατηγοριοποιηθούν τα μοντέλα εμφάνισης σε πιθανές ομάδες και θα παρουσιαστούν διάφορες προσεγγίσεις της βιβλιογραφίας για την περιγραφή της εμφάνισης ενός αντικειμένου.

### 1) Κατηγοριοποίηση

Αρχικά ως **Μοντέλο Εμφάνισης** θα μπορούσε να οριστεί ένα σύνολο **χαρακτηριστικών το οποίο θα μπορεί να περιγράψει επαρκώς και ευδιακρίτως τις ιδιότητες εμφάνισης ενός αντικειμένου**. Από τον ορισμό του μοντέλου εμφάνισης προκύπτει ότι χαρακτηρίζονται από μεγάλη ποικιλότητα τόσο ως προς το πρόβλημα που προσπαθούν να επιλύσουν, όσο και στον τρόπο με τον οποίο το επιτυγχάνουν.

Όσον αφορά τα προβλήματα που καλούνται να επιλύσουν τα μοντέλα εμφάνισης, αυτά θα μπορούσαν χονδρικά να κατηγοριοποιηθούν σε δύο μεγάλες κατηγορίες: το πρόβλημα διάκρισης μεταξύ ετεροειδών αντικειμένων και το πρόβλημα διάκρισης μεταξύ ομοειδών αντικειμένων. Αναλόγως με το πρόβλημα που επιλύει το εκάστοτε μοντέλο, τα μοντέλα εμφάνισης μπορεί να καταταγούν στις εξής τρεις κατηγορίες:

- **Γενικά μοντέλα εμφάνισης:** Είναι τα μοντέλα τα οποία χρησιμοποιούνται για να περιγράψουν ένα είδος αντικειμένου (π.χ. άνθρωπος) και να το καταστήσουν εύκολα διακριτό σε σχέση με άλλα αντικείμενα. Τα μοντέλα αυτά αποτελούν τη συνήθη μορφή μοντέλων εμφάνισης με αρκετά ευρύ φάσμα εφαρμογών.
- **Εξειδικευμένα μοντέλα εμφάνισης:** Πρόκειται για μοντέλα τα οποία περιγράφουν ένα συγκεκριμένο (ή μία ομάδα) αντικείμενο με στόχο να το καταστήσουν εύκολα διακριτό από άλλα ομοειδή αντικείμενα



(π.χ. να διακρίνουν έναν συγκεκριμένο άνθρωπο από ένα σύνολο ανθρώπων). Τέτοια μοντέλα αν και είναι αδύναμα στη διαφοροποίηση του ίδιου του αντικειμένου σε σχέση με ετεροειδή αντικείμενα, τυχάνουν εφαρμογής σε εξειδικευμένες εφαρμογές όπου ενυπάρχει η έννοια της ταυτότητας. Τα εξειδικευμένα μοντέλα είναι το αντικείμενο της παρούσας εργασίας.

- **Μικτά μοντέλα:** Αποτελούν συνδυασμό των έτερων δύο μοντέλων και ο σχεδιασμός τους (αν θα είναι πιο πολύ γενικά ή εξειδικευμένα) εξαρτάται από το πρόβλημα που επιχειρούν να επιλύσουν.

Πέραν όμως της άνωθεν κατηγοριοποίησης τα μοντέλα εμφάνισης μπορούν να καταταγούν και ανάλογα με τον τύπο των πρωτογενών χαρακτηριστικών που χρησιμοποιούν. Έτσι προκύπτουν οι εξής κατηγορίες:

- Μοντέλα βασισμένα σε ακμές/κλίσεις
- Μοντέλα βασισμένα σε σημεία ενδιαφέροντος (keypoints)
- Μοντέλα βασισμένα σε τμήματα εικόνας (patches)
- Μοντέλα βασισμένα στη χρωματική πληροφορία

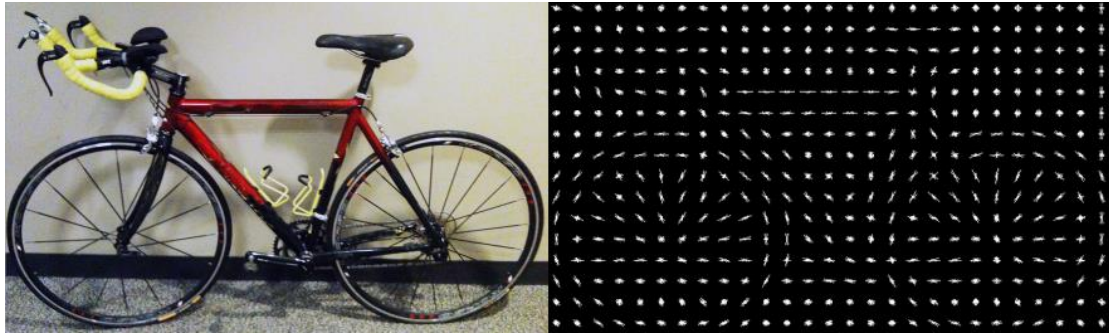
Προφανώς, υπάρχουν και μικτές προσεγγίσεις αλλά οι παραπάνω 4 κατηγορίες αποτελούν τη βάση σχεδιασμού κάθε μοντέλου εμφάνισης.

## 2) Μοντέλα βασισμένα σε ακμές/κλίσεις

Ο υπολογισμός των κλίσεων (gradient) και η εξαγωγή ακμών βρίσκονται στον πυρήνα πολλών αλγορίθμων και διαδικασιών της Όρασης Υπολογιστών. Τα μοντέλα εμφάνισης βασισμένα στις κλίσεις της εικόνας μπορεί να θεωρηθεί ότι αποτελούν την πιο χαρακτηριστικότερη περίπτωση **γενικών** μοντέλων εμφάνισης.

Το χαρακτηριστικό παράδειγμα τέτοιων μοντέλων δεν είναι παρά ο περιγραφέας Histogram of Oriented Gradients (HOG) [9]. Ο συγκεκριμένος περιγραφέας έχει ως βάση του έναν σημειακό περιγραφέα. Ουσιαστικά ο περιγραφέας HOG υπολογίζει τα ιστογράμματα κατανομής (ως προς τον προσανατολισμό) των κλίσεων για κάθε περιοχή μίας διαμέρισης τις εικόνας (σε επικαλυπτόμενες υποπεριοχές).

Οι εφαρμογές του συγκεκριμένου μοντέλου εμφάνισης είναι πολλές και αφορούν κατά κύριο λόγο την αναγνώριση και εντοπισμό αντικειμένων σε εικόνες καθώς και την ταξινόμησή τους.



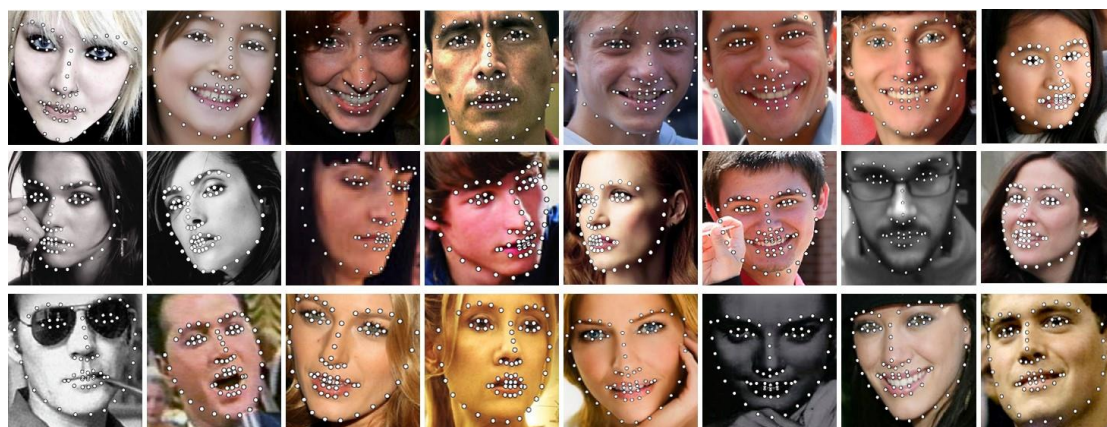
Εικόνα 21: Ο καθολικός περιγραφέας εμφάνισης HOG. Πηγή:<http://mathworks.com>

Αξιζει, επίσης, να σημειωθεί το γεγονός ότι τέτοιου είδους μοντέλα εμφάνισης προκύπτουν «έμμεσα» και στα Συνελικτικά Νευρωνικά Δίκτυα (Convolutional Neural Networks) κυρίως ως αποκρίσεις των κρυφών «στρώσεων» (hidden layers activations).

### 3) Μοντέλα βασισμένα σε keypoints

Τα σημεία ενδιαφέροντος αποτελούν εκείνο τον τύπο χαρακτηριστικών που χρησιμοποιείται ευρέως σε πολυάριθμες εφαρμογές της Όρασης Υπολογιστών. Συνοπτικά, τα σημεία ενδιαφέροντος είναι σημεία της εικόνας/βίντεο τα οποία είναι κρίσιμα για κάποια συνάρτηση κόστους. Για παράδειγμα, οι «γωνίες» είναι σημεία μεγίστου του πίνακα αυτοσυσχέτισης, ενώ τα blobs σχετίζονται άμεσα με τον Hessian πίνακα. Αντίστοιχα μπορεί να προκύψουν σημεία ενδιαφέροντος ως ελάχιστα της συνέλιξης (ή correlation) της εικόνας με κάποιο πυρήνα-πρότυπο. Χαρακτηριστικά παραδείγματα σημειακών χαρακτηριστικών αποτελούν τα χαρακτηριστικά SIFT[10], FAST[11], Harris[12], SURF[13] κτλ.

Τα σημειακά χαρακτηριστικά από μόνα τους, όμως, δεν καθιστούν ένα μοντέλο εμφάνισης δόκιμο. Στη βιβλιογραφία προτείνονται διάφορα μοντέλα εμφάνισης βασισμένα σε σημειακά χαρακτηριστικά, όμως δύο είναι οι κύριες προσεγγίσεις. Αρχικά τα σημεία μπορεί να συνοδεύονται από χωρικές σχέσεις. Είτε πρόκειται για απλές σχέσεις διαδοχής είτε για πιο σύνθετες χωρικές αλληλεπιδράσεις ή ακόμα και για ένα πλήρες χωρικό στατιστικό μοντέλο. Χαρακτηριστικό παράδειγμα αποτελεί η μοντελοποίηση της εμφάνισης του προσώπου, όπου με τη χρήση σημείων ενδιαφέροντος προσαρμόζεται ένα στατιστικό μοντέλο περιγραφής του προσώπου. Από το αποτέλεσμα της προσαρμογής τελικά προκύπτει το μοντέλο εμφάνισης το οποίο μπορεί να είναι είτε γενικό είτε (συνηθέστερα) εξειδικευμένο διαχωρίζοντας και ταυτοποιώντας συγκεκριμένα άτομα.



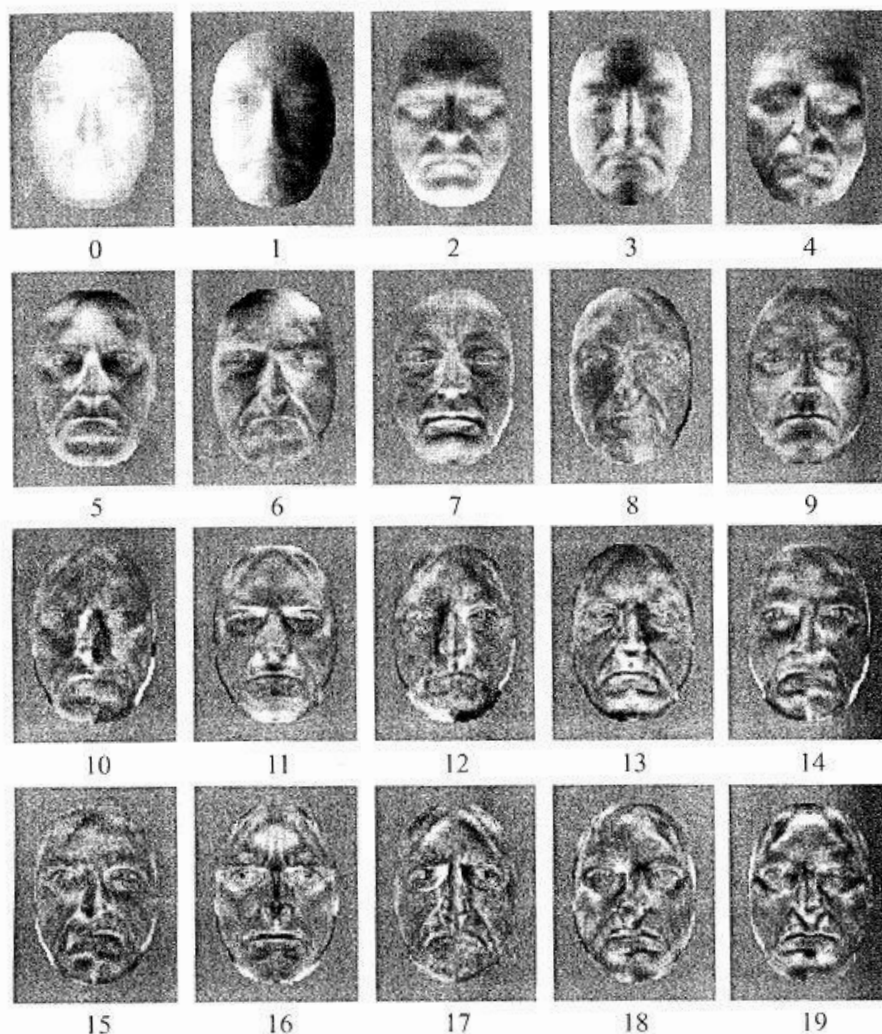
Εικόνα 22: Σημεία ενδιαφέροντος για μοντελοποίηση της εμφάνισης του ανθρώπινου προσώπου.  
Πηγή: <http://mmlab.ie.cuhk.edu.hk>

Η έτερη προσέγγιση αφορά τη δημιουργία ιστογραμμάτων συχνότητας εμφάνισης σημειακών χαρακτηριστικών για την περιγραφή της εμφάνισης. Τέτοια περίπτωση αφορά η, ευρέως χρησιμοποιούμενη στην ταξινόμηση εικόνων, τεχνική του Bag-of-Visual-Words. Σύμφωνα με την τεχνική αυτή, υπολογίζεται και καταγράφεται πόσο συχνά εμφανίζονται σημειακά χαρακτηριστικά συγκεκριμένου τύπου. Το τελικό αποτέλεσμα είναι ένας διανυσματικός περιγραφέας εμφάνισης για το σύνολο της εικόνας, πετυχαίνοντας σε μεγάλο βαθμό στην γενίκευση του μοντέλου.

## 4) Μοντέλα βασισμένα σε patches

Ένας άμεσος τρόπος αναπαράστασης της εμφάνισης ενός αντικειμένου είναι ένα τμήμα εικόνας που το απεικονίζει. Σε αυτήν τη λογική στηρίζονται τα εξειδικευμένα μοντέλα εμφάνισης που χρησιμοποιούν τμήματα εικόνων (patches). Στην πιο απλή περίπτωση το αντικείμενο περιγράφεται από ένα patch και «ταυτοποιείται» μέσω τεχνικών μέτρησης ομοιότητας μεταξύ παραθύρων όπως η συνέλιξη και ο συντελεστής συσχέτισης.

Πέραν όμως του παραπάνω απλοϊκού μοντέλου μπορεί να χρησιμοποιηθούν και πιο σύνθετες περιγραφές βασισμένες σε στατιστικές μεθόδους. Χαρακτηριστική είναι η μέθοδος προσδιορισμού ιδιοεικόνων του αντικειμένου οι οποίες μπορούν μέσω γραμμικού συνδυασμού να «αναπλάσουν» την αρχική εικόνα του αντικειμένου. Κατά αυτόν τον τρόπο οι συντελεστές του γραμμικού συνδυασμού αποτελούν ένα είδος μοντέλου εμφάνισης που βασίζεται σε τμήματα της εικόνας.



Εικόνα 23: Ιδιοπρόσωπα (Eigenfaces). Πηγή: <http://aix1.uottawa.ca>

## 5) Μοντέλα βασισμένα στη χρωματική πληροφορία

Στην περίπτωση που στόχο του μοντέλου εμφάνισης αποτελεί η «ταυτοποίηση» ενός αντικειμένου και ο διαχωρισμός του μεταξύ άλλων ομοειδών του, δηλαδή η ανάπτυξη ενός εξειδικευμένου μοντέλου εμφάνισης, τα χρωματικά χαρακτηριστικά του αντικειμένου μπορούν εν δυνάμει να παρέχουν την απαιτούμενη πληροφορία. Για παράδειγμα, ένα σύνολο χαρακτηριστικών που βασίζονται στη χρωματική πληροφορία θα μπορούσε να αποτελέσει για την περίπτωση του ανθρώπου το χρώμα των μαλλιών, του δέρματος, των ρούχων, κ.ο.κ. Το προαναφερθέν παράδειγμα σκιαγραφεί ένα

εξειδικευμένο μοντέλο εμφάνισης στηριζόμενο σε πολύ υψηλού επιπέδου σημασιολογική πληροφορία που εξάγεται κατά κύριο λόγο από χρωματικά χαρακτηριστικά.

Η πολυπλοκότητα ενός τέτοιου εξειδικευμένου μοντέλου εμφάνισης, πέραν της δυσκολίας στην υλοποίηση και τον επαναλαμβανόμενο υπολογισμό, δεν αποτελεί κατ' ανάγκη χαρακτηριστικό των υπό συζήτηση μοντέλων εμφάνισης. Πιο απλές προσεγγίσεις όπως μετρήσεις της χρωματικής κατανομής ή στοχαστική μοντελοποίηση του ίδιου του χρώματος του αντικειμένου θα μπορούσαν να εφαρμοστούν με καλά αποτελέσματα.

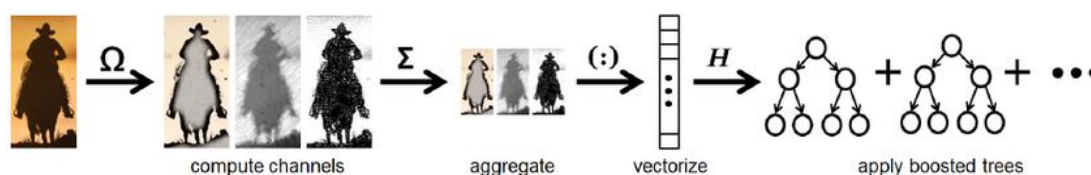
Άξιο αναφοράς κρίνεται το γεγονός ότι στα πλαίσια της παρούσας μεταπτυχιακής εργασίας αναπτύχθηκαν (εξειδικευμένα) μοντέλα εμφάνισης κυρίως βασισμένα στην χρωματική πληροφορία του ανθρώπου-στόχου. Η χρωματική πληροφορία μπορεί να αποτελέσει βάση και για γενικά μοντέλα καθώς, για παράδειγμα, υπάρχουν κατηγορίες αντικειμένων που παρουσιάζουν σταθερή χρωματική κατανομή όπως η μπάλα σε μερικά αθλήματα (τένις -> κίτρινη μπάλα) ή ακόμα και η απόχρωση του δέρματος (πολλαπλές a-priori γνωστές χρωματικές κλάσεις).



Εικόνα 24: Ανίχνευση δέρματος. Πηγή: <http://opencv.org>

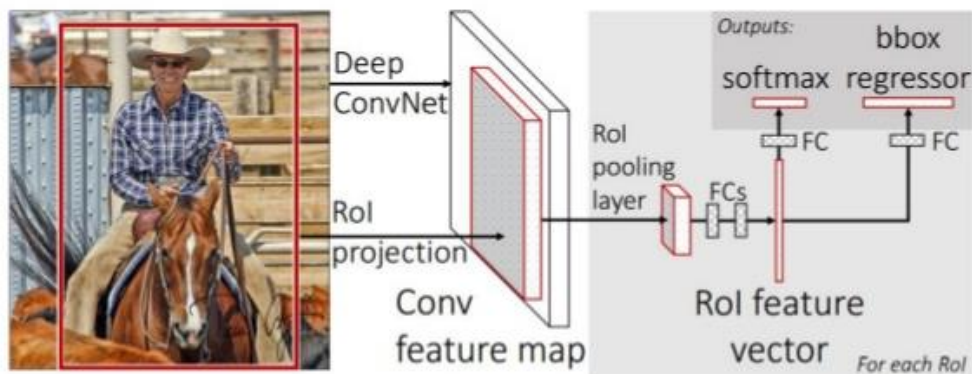
## 6) Ανασκόπηση βιβλιογραφίας

Στο παρόν εδάφιο θα παρουσιαστεί μία σύντομη ανασκόπηση των εφαρμογών στις οποίες χρησιμοποιούνται σε κάποιο στάδιο της διαδικασίας τα μοντέλα εμφάνισης. Αρχικά, ο κύριος κλάδος εφαρμογής μοντέλων εμφάνισης δεν είναι παρά η ανίχνευση αντικειμένων. Ένας από τους πλέον εύρωστους αλγορίθμους, ο ACF Detector [15] χρησιμοποιεί ένα μικτό μοντέλο εμφάνισης που ενσωματώνει τόσο χαρακτηριστικά κλίσεων υπό την μορφή χαρακτηριστικών HOG, όσο και χρωματικά χαρακτηριστικά χρησιμοποιώντας τον ομοιόμορφο χώρο χρώματος CIE<sub>Luv</sub>. Ο συγκεκριμένος ανιχνευτής αφού ενσωματώσει την παραπάνω πληροφορία σε «κανάλια» υπολογίζει μία διανυσματική αθροιστική αναπαράστασή τους. Το διάνυσμα αυτό αποτελεί τελικά το προϊόν του μοντέλου εμφάνισης που υπολογίζει ο αλγόριθμος ACF.



Εικόνα 25: Ο αλγόριθμος Aggregate Channel Features (ACF). Πηγή: Από δημοσίευση ACF

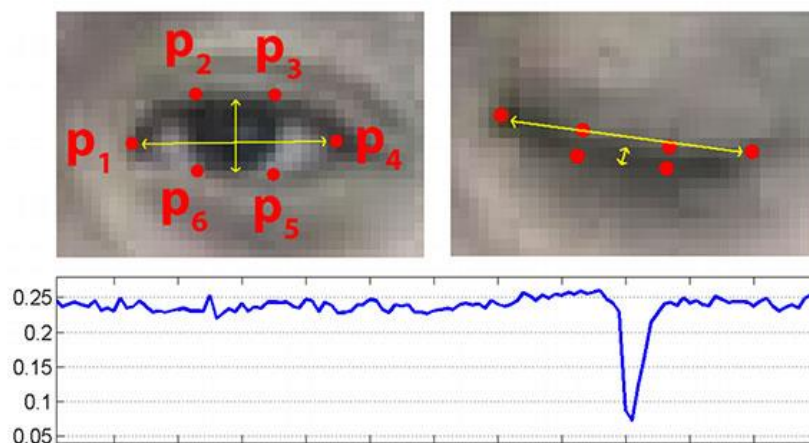
Στον κλάδο της ανίχνευσης αντικειμένων η έννοια του μοντέλου εμφάνισης είναι καθοριστικής σημασίας όπως διαφαίνεται και από τον ανιχνευτή ACF [15]. Ακόμη και πιο μοντέρνες προσεγγίσεις με τη χρήση «ανά περιοχή» Συνελκτικών Νευρωνικών Δικτύων (Region CNN) υπάρχει μοντέλο εμφάνισης το οποίο «μαθαίνεται» από τα δεδομένα αληθείας. Στην συγκεκριμένη περίπτωση για ανιχνευτές όπως ο Fast R-CNN [16] ή ο Faster R-CNN [17] τα προϊόντα του «εκμαθημένου» μοντέλου εμφάνισης ενυπάρχουν «κρυμμένα» ως οι αποκρίσεις νευρώνων στις τελικές ενδιάμεσες στρώσεις του (βαθιού) δικτύου. Τα μοντέλα αυτά θα μπορούσαν να χαρακτηριστούν ως βασισμένα σε patches, όμως ανάλογα με τα φίλτρα που «μαθαίνει» το νευρωνικό δίκτυο δύναται να στηρίζονται τόσο σε πληροφορία κλίσεων όσο και σε χρωματική πληροφορία.



Εικόνα 26:Ο ανιχνευτής FAST R-CNN (Αρχιτεκτονικής δικτύου). Πηγή: Από δημοσίευση

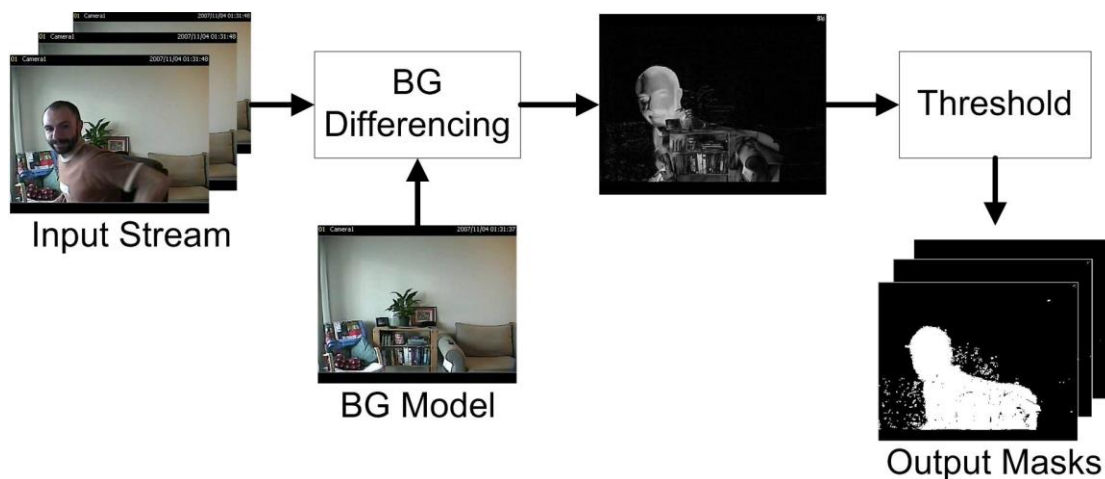
Πέραν των παραπάνω παραδειγμάτων ανίχνευσης αντικειμένων σε εικόνες που ενσωματώνουν γενικά μοντέλα εμφάνισης δύναται να προκύψουν εφαρμογές όπου απαιτείται η προσαρμογή ενός στατιστικού μοντέλου σε ένα πλήθος σημείων ενδιαφέροντος. Σε αυτές τις περιπτώσεις το μοντέλο εμφάνισης μπορεί να είναι μικτό ή ακόμα και εξειδικευμένο. Τέτοιο παράδειγμα μικτού μοντέλου αποτελεί η ανίχνευση προσώπων [18] όπου η μορφή του δικτύου σημείων που σχηματίζεται από τα σημεία ενδιαφέροντος καταδεικνύει αν πρόκειται για πρόσωπο ή όχι με ταυτόχρονη τοπικοποίηση στην εικόνα. Η μορφή του δικτύου, η οποία μπορεί να δηλώνει κάποιο γεωμετρικό μοντέλο ή απλά μετρήσεις χωρικής κατανομής και πυκνότητας, αποτελεί ένα είδος μοντέλου εμφάνισης βασισμένο σε keypoints. Για λόγους πληρότητας να αναφερθεί ότι η ανίχνευση προσώπων μπορεί να επιτευχθεί και με τη χρήση κυματιδίων Haar[19], χαρακτηριστικά που συνδέονται με τις κλίσεις της εικόνας. Παρόμοιες μέθοδοι έχουν εφαρμοστεί και για την ανίχνευση άλλου είδους αντικειμένων με απαιτήσεις για υψηλή ακρίβεια εντοπισμού όπως του ανθρώπινου οφθαλμού[20], ή ακόμα και δράσεων όπως η ανίχνευση «βλεφαρισμού» του ματιού[21].





Εικόνα 27: Ανίχνευση βλεφαρισμού. Πηγή: <https://pyimagesearch.com>

Στην ευρεία κατηγορία εφαρμογών που χρησιμοποιούν μοντέλα εμφάνισης εντάσσεται και η δημοφιλής διαδικασία της μοντελοποίησης του χρώματος μέσω Γκαουσιανών Μοντέλων Ανάμιξης (Gaussian Mixture Models - GMM) [22]. Η παραπάνω μοντελοποίηση αποτελεί μία μορφή μοντέλου εμφάνισης βασισμένου στο χρώμα, με χαρακτηριστικότερες εφαρμογές τόσο την ανίχνευση του δέρματος [23] όσο και της απαλοιφής παρασκηνίου σε βίντεο [24].

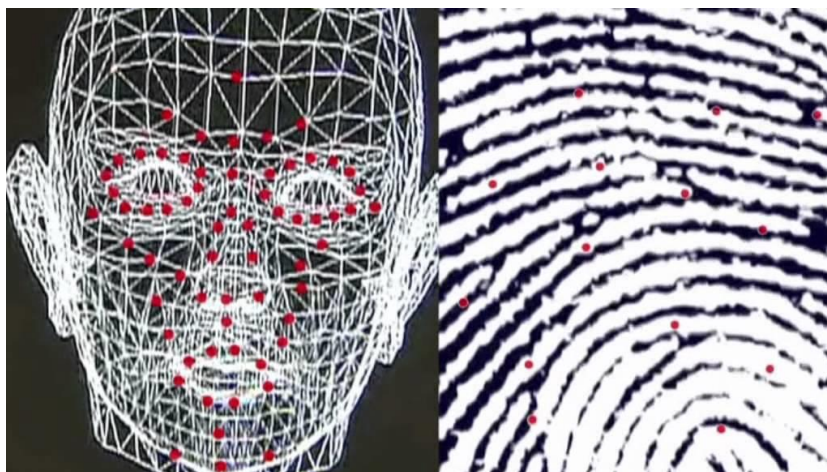


Εικόνα 28: Αφαίρεση παρασκηνίου σε βίντεο με σταθερή κάμερα. Πηγή: <http://wikidot.com>

Κατά τις προηγούμενες παραγράφους διενεργήθηκε μία επισκόπηση στην χρήση των γενικών μοντέλων εμφάνισης σε μία επιλογή διαφόρων διαδικασιών της Όρασης Υπολογιστών. Κατά την παρούσα εργασία, όμως, αναπτύσσονται εξειδικευμένα μοντέλα εμφάνισης. Τα εξειδικευμένα μοντέλα τυγχάνουν και αυτά ευρείας χρήσης σε πληθώρα εφαρμογών. Λόγω

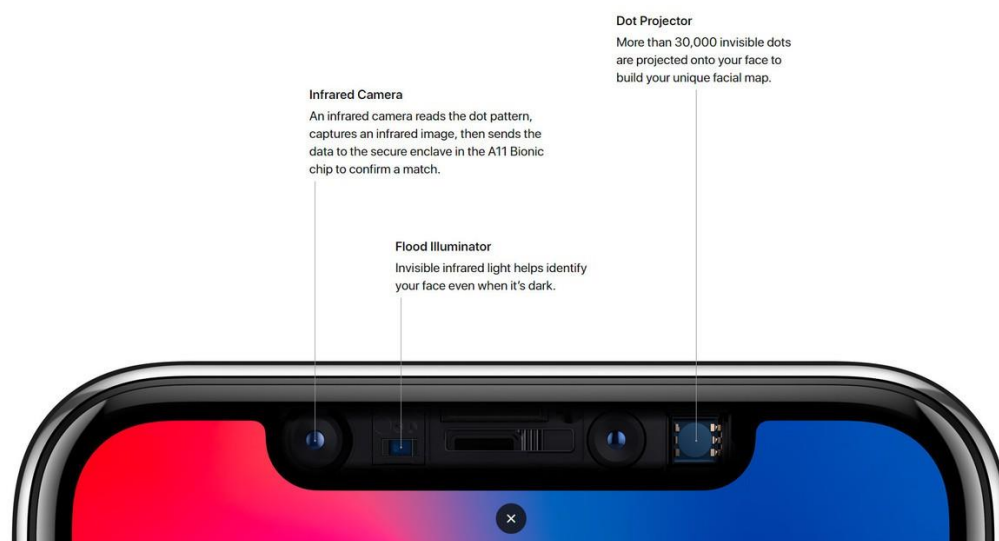
της δυνατότητας διαχωρισμού μεταξύ ομοειδών αντικειμένων πολλές εφαρμογές που τα αξιοποιούν στηρίζονται στη δυνατότητα «ταυτοποίησης» και συνταύτισης αντικειμένων που προσφέρουν.

Μία χαρακτηριστική εφαρμογή τους αποτελεί η βιομετρική ταυτοποίηση. Η πιο απλή έκφρασή της δεν είναι παρά η ταυτοποίηση «δακτυλικών αποτυπωμάτων» [25]. Για τον σκοπό αυτό το κάθε αποτύπωμα αποκτά το δικό του μοντέλο εμφάνισης (βασισμένο σε ακμές και σημεία ενδιαφέροντος) γεγονός που διευκολύνει έναν αλγόριθμο συνταύτισης αποτελεσματικά να το αναγνωρίσει και να το ταυτίσει σε κάποια επόμενη εμφάνισή του. Παρόμοιες εφαρμογές αφορούν την ταυτοποίηση προσώπων ή λοιπών βιομετρικών χαρακτηριστικών.



Εικόνα 29: Εφαρμογές βιομετρικών συστημάτων ασφαλείας. Πηγή: <https://i.ytimg.com>

Ειδικά για τα βιομετρικά συστήματα ασφαλείας, η ωριμότητα των αλγορίθμων ταυτοποίησης επιτρέπει πληθώρα εμπορικών εφαρμογών, όπως η χρήση ταυτοποίησης «δακτυλικών αποτυπωμάτων» σε συσκευές κινητών τηλεφώνων ή Η/Υ με τη χρήση οπτικών αισθητήρων. Τελευταία έχει αρχίσει και η εμπορική διάθεση συστημάτων ταυτοποίησης προσώπων με τη χρήση αισθητήρων εικόνας και βάθους (Structured Light Scanners) σε συσκευές κινητών τηλεφώνων.



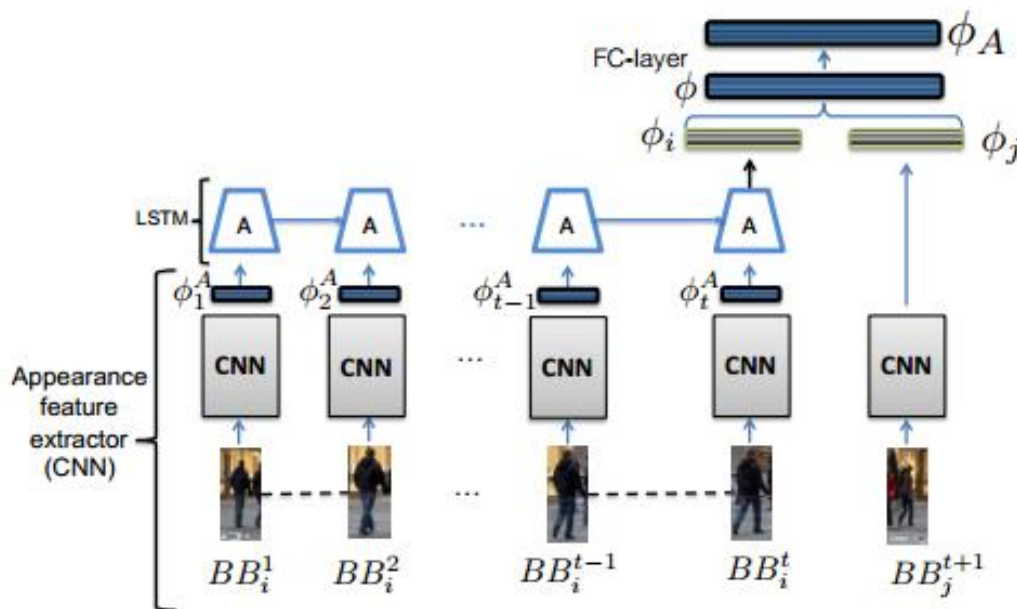
Εικόνα 30: Εμπορικό προϊόν ταυτοποίησης προσώπου. Πηγή: <https://apple.com>

Τέλος, μια δημοφιλής εφαρμογή των εξειδικευμένων μοντέλων εμφάνισης αποτελούν οι μέθοδοι οπτικής παρακολούθησης (visual tracking), εντοπισμού και παρακολούθησης κινούμενων αντικειμένων σε ακολουθίες εικόνων/ βίντεο. Κατά την διαδικασία της παρακολούθησης ο καθορισμός ενός εύρωστου και διαχρονικού μοντέλου εμφάνισης αποτελεί μία από τις πιο συνηθισμένες πρακτικές. Ταυτόχρονα αποτελεί και σημερινή επιστημονική πρόκληση λόγω του σημαντικού βαθμού δυσκολίας στην ανάπτυξη αλγορίθμων που αποτελεσματικά σε διαφορετικές συνθήκες μπορούν να μοντελοποιήσουν τα αντικείμενα ή τους κινούμενους στόχους.

Στη διεθνή βιβλιογραφία έχουν προταθεί διάφορες λύσεις μοντέλων εμφάνισης, άλλοτε βασισμένες σε patches όπως στην περίπτωση του αλγόριθμου παρακολούθησης ενός στόχου TLD [26] ή του αλγόριθμου παρακολούθησης πολλαπλών αντικειμένων MDP [27] και άλλοτε σε μικτά μοντέλα εμφάνισης με τη χρήση, κατά κύριο λόγο τα τελευταία χρόνια, τεχνητών νευρωνικών δικτύων.

Μία τέτοια περίπτωση συναντάται στην σχετική εργασία των Sadeghian et al. [28] όπου συνδυάζεται η χρήση ενός Συνελκτικού Νευρωνικού Δικτύου (CNN) με ένα Δίκτυο Μακρο-Βραχυπρόθεσμης μνήμης (LSTM) για το μοντέλο εμφάνισης κάθε στόχου. Η συγκεκριμένη προσέγγιση παρουσιάζει έντονο ερευνητικό ενδιαφέρον και λόγο πιθανώς να υιοθετηθεί και μελετηθεί σημαντικά από την ερευνητική κοινότητα. Αυτό κυρίως διότι αποδομείται το μοντέλο εμφάνισης σε δύο υπο-διεργασίες: η πρώτη αφορά τα χαρακτηριστικά και τον χώρο προτύπων του μοντέλου

(CNN) ενώ η δεύτερη προσδίδει στο μοντέλο χρονική ευρωστία βελτιώνοντας το τελικό μοντέλο (LSTM).



Εικόνα 31: Μοντελοποίηση εμφάνισης στόχου σε πολλαπλά καρέ με χρήση CNN και LSTM.  
 Πηγή: Από δημοσίευση [28]

Συμπερασματικά, τα μοντέλα εμφάνισης αποτελούν σημαντικό και νευραλγικό τμήμα πολλών διεργασιών της Όρασης Υπολογιστών. Λόγω της ευρύτητας με την οποία έχουν κατά καιρούς οριστεί τα μοντέλα, απαντώνται στη διεθνή βιβλιογραφία με έντονη ποικιλομορφία. Σίγουρα, η βελτίωσή τους αποτελεί κρίσιμη ερευνητική κατεύθυνση για την κοινότητα καθώς η ανάγκη για καλύτερα και λεπτομερέστερα μοντέλα είναι κρίσιμη για πληθώρα ανοιχτών προβλημάτων στην Όραση Υπολογιστών. Σχετικά με πιθανές μελλοντικές κατευθύνσεις προς την ανάπτυξη νέων καινοτόμων μοντέλων εμφάνισης γίνεται εκτενής αναφορά στο τελευταίο κεφάλαιο της εργασίας.

## ΚΕΦΑΛΑΙΟ 4. ΑΝΑΠΤΥΧΘΕΝΤΑ ΜΟΝΤΕΛΑ

### 1) Κίνητρα - Στόχευση

Όπως αναφέρεται και στην εισαγωγή στόχος της παρούσας μεταπτυχιακής εργασίας αποτελεί η ανάπτυξη πρωτότυπων μοντέλων εμφάνισης ανθρωπίνων στόχων. Τα μοντέλα που αναπτύσσονται δεν εστιάζουν στον διαχωρισμό μεταξύ ανθρώπου ή όχι, αλλά αποτελούν εξειδικευμένα μοντέλα τα οποία διαχωρίζουν μεταξύ τους ομοειδή αντικείμενα, εν προκειμένω ανθρώπινους στόχους. Η παραπάνω επιλογή έγινε έχοντας κατά κύριο λόγο υπόψιν εφαρμογές της Όρασης Υπολογιστών όπως η οπτική παρακολούθηση (visual tracking), εφαρμογές οι οποίες αξιοποιούν πληροφορία από τέτοιου είδους εξειδικευμένα μοντέλα.

Υπό τις παραπάνω συνθήκες προκύπτει ότι για να υπάρξει η δυνατότητα ορθής εφαρμογής των αναπτυχθέντων μοντέλων έπεται η *a-priori* γνώση ότι το τμήμα της εικόνας (patch) στο οποίο εφαρμόζεται το μοντέλο περιγραφής απεικονίζει άνθρωπο (και κατ' αντιστοιχία όποια κατηγορία αντικειμένου αναφέρεται στο εκάστοτε εξειδικευμένο μοντέλο). Αν και δεν αποτελεί ικανή και αναγκαία συνθήκη, η γνώση της δυαδικής κατάτμησης του κάθε τμήματος της εικόνας σε άνθρωπο ή όχι είναι επιθυμητή, καθώς μειώνει τον θόρυβο στην προκύπτουσα σημασιολογική περιγραφή. Στην συνέχεια της εργασίας η γνώση της παραπάνω κατάτμησης θα θεωρείται δεδομένη.



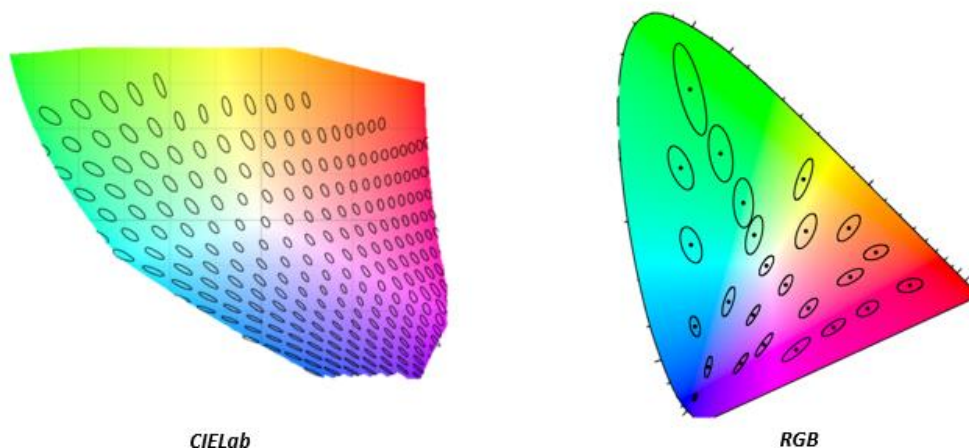
*Εικόνα 32: Διαδική σημασιολογική κατάτμηση patch*

Στη συνέχεια του κεφαλαίου θα παρουσιαστεί ο χώρος χρώματος που χρησιμοποιείται κατά την προτυποποίηση καθώς και τα αίτια που οδηγούν στην επιλογή αυτή. Τέλος, ακολουθεί η παρουσίαση των τριών πρωτότυπων μοντέλων που δημιουργήθηκαν όπως και μία σύντομη συζήτηση περί ζητημάτων (υπολογιστικής) επίδοσης των μοντέλων.

## 2) Επιλογή χώρου χρώματος

Η μοντελοποίηση της εμφάνισης των ανθρωπίνων στόχων, για τα δεδομένα της παρούσας εργασίας, στηρίζεται στα χρωματικά χαρακτηριστικά του καθενός. Θεωρώντας ως δεδομένο ότι η κύρια πηγή άντλησης πληροφορίας είναι τμήματα μίας RGB εικόνας προκύπτει άμεσα η ανάγκη επιλογής ενός χώρου χρώματος ως χώρου αναφοράς. Η επιλογή του χώρου RGB (βλέπε Κεφάλαιο 2) δεν αποτελεί μία καλή επιλογή για δύο κυρίως λόγους. Αφενός η συνιστώσα της φωτεινότητας βρίσκεται πεπλεγμένη στις τρεις συνιστώσες-βάσεις του χώρου (Red - Green - Blue) καθιστώντας τα χαρακτηριστικά «ευάλωτα» σε αλλαγές φωτεινότητας, αφετέρου ο χώρος RGB δεν είναι ομοιόμορφος. Δηλαδή θεωρώντας σταθερή την χρωματική απόσταση (υπό την αντίστοιχη νόρμα του χώρου) δύο χρωματικών σημείων, η αντίληψη της διαφοράς του χρώματος από τον άνθρωπο δεν παραμένει σταθερή σε όλο το χώρο.

### McAdam ελλείψεις σφάλματος



Εικόνα 33: Ελλείψεις McAdam στους χώρους CIE Lab και RGB

Για την επίλυση των παραπάνω δύο ζητημάτων επιλέχθηκε η χρήση του μη γραμμικού χώρου χρώματος CIE Lab ο οποίος αποτελεί μία καλή λύση και για τα δύο. Αρχικά, ο χώρος αυτός αποτελείται από 3 διανύσματα βάσης (όπως όλοι οι χώροι χρώματος που προέρχονται από τον RGB) με τη διαφορά ότι το ένα, η συνιστώσα L, αφορά την φωτεινότητα του «σημείου» ενώ οι άλλες δύο, οι συνιστώσες a και b, είναι συνιστώσες χρωματικότητας (chromaticities). Ακόμη ο χώρος CIE Lab είναι ένας ομοιόμορφος χώρος χρώματος, όπως περιγράφεται στο Κεφάλαιο 2. Στον αντίποδα, η μεταφορά των σημείων στον χώρο CIE Lab από τον πρωτογενή χώρο RGB αποτελεί μία

διαδικασία με σημαντικό υπολογιστικό κόστος δημιουργώντας διάφορα ζητήματα επίδοσης που συζητούνται σε επόμενο εδάφιο. Στη συνέχεια ακολουθεί αναλυτική περιγραφή των τριών μοντέλων εμφάνισης που αναπτύχθηκαν στο πλαίσιο της εργασίας.

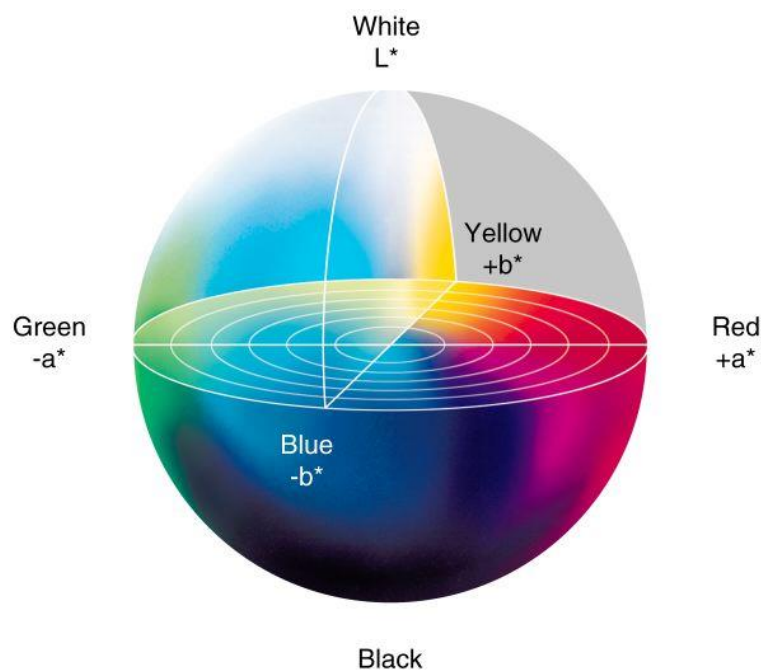


### 3) Μοντέλο Εμφάνισης “ab-Y”

Το πρώτο πρωτότυπο μοντέλο εμφάνισης που παρουσιάζεται είναι το μοντέλο “ab-Y”. Το συγκεκριμένο αποτελεί την βάση και των έτερων δύο μοντέλων εμφάνισης γεγονός που το καθιστά ως το σημαντικότερο των τριών.

#### a) Περιγραφή / Μεθοδολογία

Βάση του μοντέλου “ab-Y” αποτελεί ο χώρος χρώματος CIELab. Με αφετηρία το χώρο αυτό, για κάθε εικονοστοιχείο του τμήματος της εικόνας το οποίο είναι χαρακτηρισμένο ως άνθρωπος (δεδομένης δυαδικής σημασιολογικής κατάτμησης) υπολογίζεται η προβολή του στο χώρο CIELab. Η συνιστώσα της φωτεινότητας του κάθε εικονοστοιχείου, εν συνεχεία, απομονώνεται και απαλείφεται με τις δύο συνιστώσες χρωματικότητας να απομένουν. Τελικά προκύπτει ένα σύνολο σημείων στο επίπεδο a-b του χώρου CIELab που «περιγράφουν» χρωματικά τον άνθρωπο (ή αντικείμενο) στο εκάστοτε τμήμα της εικόνας. Να τονιστεί ότι η περιγραφή αυτή είναι θεωρητικά αμετάβλητη σε αλλαγές φωτεινότητας. Κάτι τέτοιο είναι αναγκαίο καθώς σε πραγματικές συνθήκες, όπως για παράδειγμα ένα βίντεο, έντονες μεταβολές φωτεινότητας δύνανται να προκύψουν εντός διαστήματος λίγων συνεχόμενων καρέ.



Εικόνα 34: Ο χώρος CIE Lab. Βασισμένος στη θεωρία των ανταγωνιστικών χρωμάτων. Πηγή: nazdar.com

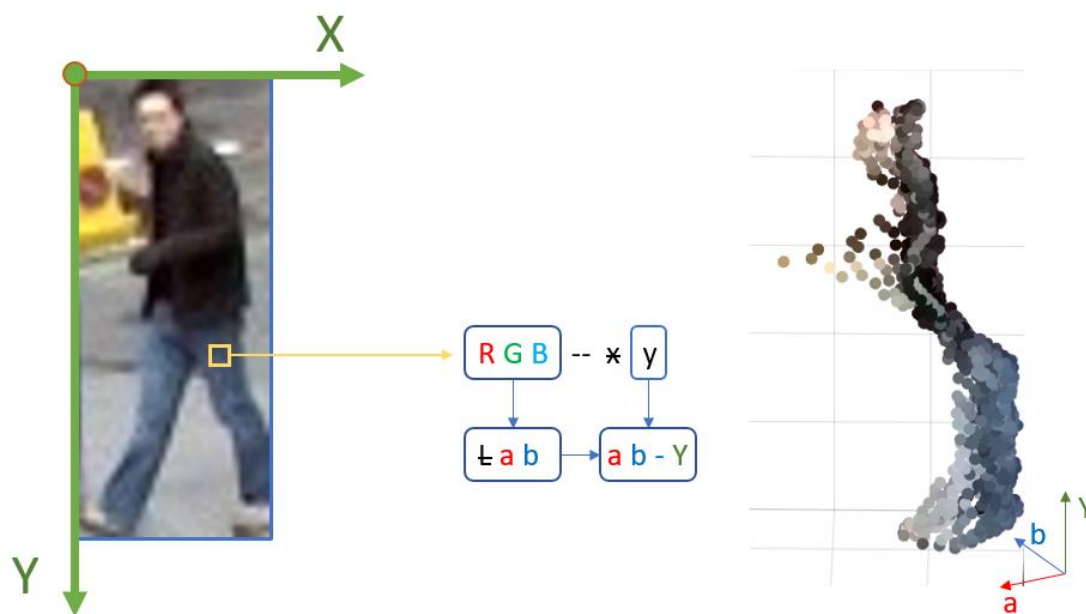
Το πρόβλημα της άνωθεν περιγραφής δεν είναι παρά η παντελής έλλειψη χωρικής αναφοράς. Η προτεινόμενη λύση, που υλοποιήθηκε και αξιολογήθηκε, για το συγκεκριμένο πρόβλημα είναι η ενσωμάτωση στην περιγραφή της συνιστώσας θέσης “Y” του κάθε εικονοστοιχείου του τμήματος της εικόνας. Συγκεκριμένα, θεωρώντας ένα τμήμα της εικόνας ως ένα ορθογώνιο ορίζει κανείς το χωρικό αριστερόστροφο 2Δ σύστημα αναφοράς “X-Y” με κέντρο το άνω αριστερό άκρο του ορθογωνίου. Τότε η συνιστώσα Y αντιστοιχεί στον αριθμό των γραμμών που απέχει το κάθε εικονοστοιχείο από την αρχή του συστήματος ( $Y > 0$ ). Κατ’ αναλογία η συνιστώσα X αναφέρεται στις στήλες του τμήματος της εικόνας. Η συνιστώσα X θεωρήθηκε ως μη χρήσιμη καθώς μεταβάλλεται από πιθανή στροφή του αντικειμένου. Κάτι τέτοιο δεν συμβαίνει για τη συνιστώσα Y αν θεωρηθούν φυσιολογικές συνθήκες λήψης (στροφή «κ» της κάμερας περίπου 0), με την παραδοχή ότι το χαρακτηριστικό αυτό καθίσταται αναλλοίωτο.



Εικόνα 35: Σύστημα αναφοράς patch

## b) Χώρος προτύπων

Τελικώς ο χώρος προτύπων του συγκεκριμένου περιγραφέα είναι ένας νέος σύνθετος χώρος, ο οποίος αποτελείται από τις δύο χρωματικές συνιστώσες του CIE Lab a-b και τη χωρική συνιστώσα Y. Πρόκειται για έναν 3Δ χώρο ικανό πλέον να αποτελέσει τη βάση ενός εξειδικευμένου μοντέλου εμφάνισης. Σε αυτόν τον χώρο προβάλλονται όλα τα εικονοστοιχεία του τμήματος της εικόνας (που έχουν χαρακτηριστεί ως άνθρωπος). Οι τρεις συνιστώσες του χώρου (Y-a-b) έχουν δώσει το όνομα στο παρόν μοντέλο εμφάνισης.



Εικόνα 36: Μεθοδολογία εξαγωγής νέφους  $ab-Y$

### ς) Γιατί $Y$ ;

Η μεγαλύτερη συνεισφορά του συγκεκριμένου μοντέλου δεν είναι παρά η ενσωμάτωση της χωρικής συνιστώσας  $Y$  στο μοντέλο περιγραφής. Η σημασία αυτής της πληροφορίας είναι πολύ ισχυρή διότι ο άνθρωπος διαφοροποιείται σημασιολογικά κατά τη διεύθυνση αυτή. Για παράδειγμα, το κεφάλι, ο κορμός και τα πόδια ενός ανθρώπου διαχωρίζονται πλήρως κατά την  $Y$ -διεύθυνση. Το γεγονός αυτό επιτρέπει τη χρωματική μοντελοποίηση του κάθε επιμέρους τμήματος ξεχωριστά. Πρακτικά, δίνεται η δυνατότητα ξεχωριστής χρωματικής περιγραφής του κάθε τμήματος ενός ανθρωπίνου στόχου, αυξάνοντας ως συνέπεια τη χωρητικότητα της περιγραφής σε σημασιολογική πληροφορία με έμμεσο τρόπο.

Για να καταστεί δυνατή όμως η περιγραφή αντιπροσωπευτική ενός ανθρώπου και όχι ενός τμήματος μιας εικόνας που απεικονίζει έναν άνθρωπο δεν επαρκεί η ενσωμάτωση της απόλυτης τιμής της συνιστώσας  $Y$ . Για αυτό το λόγο η συνιστώσα  $Y$  κανονικοποιείται στο διάστημα  $[0,1]$ . Η κανονικοποίηση διενεργείται με το να τεθεί ως «σημείο 0» η ελάχιστη τιμή της συνιστώσας  $Y$  από το σύνολο των εικονοστοιχείων του τμήματος της εικόνας που έχουν χαρακτηριστεί ως «άνθρωπος» και αντίστοιχα ως «σημείο 1» η μέγιστη τιμή της. Κατά αυτόν τον τρόπο η περιγραφή αναφέρεται σε

ποσοστιαία μετατόπιση επί της εμφάνισης του ανθρώπου, δημιουργώντας πλέον ένα **φραγμένο** πεδίο προτύπων καθώς τόσο η συνιστώσα  $Y$  φράσσεται στο διάστημα  $[0,1]$  όσο και οι χρωματικότητες  $a-b$  είναι «κατά φυσικό τρόπο»<sup>5</sup> φραγμένες.

## **d) Περιγραφέας ως 3Δ νέφος και το πρόβλημα συν-προσαρμογής**

Το τελικό προϊόν του μοντέλου εμφάνισης “ $Y-ab$ ” δεν είναι παρά ένας περιγραφέας (βλέπε Κεφάλαιο 2). Η κύρια διαφοροποίηση σε σχέση με κλασσικούς περιγραφείς της Όρασης Υπολογιστών δεν είναι παρά η μη διανυσματική του φύση. Ο τελικός περιγραφέας είναι ένα σύνολο τριάδων ( $a-b-Y$ ) για κάθε εικονοστοιχείο του τμήματος της εικόνας (προκαθορισμένο ως άνθρωπο ή αντικείμενο ενδιαφέροντος). Το παραπάνω σύνολο θα μπορούσε να ιδωθεί και ως ένα 3Δ σημειοσύνολο, ή αλλιώς ένα 3Δ νέφος σημείων (3D point cloud). Μια τέτοια θεώρηση επιτρέπει τη χρήση μεθόδων της 3Δ Γεωμετρίας για την αξιολόγηση των διαφορών μεταξύ δύο νεφών περιγραφικής εμφάνισης.

Σε πραγματικές συνθήκες, εξαιτίας σφαλμάτων τόσο στη διαδικασία της κατάτμησης όσο και στην μεταβολή της στάσης του ανθρώπου στο χρόνο, η μέτρηση της «απόστασης» δύο νεφών δεν επαρκεί για την ορθή σύγκριση δύο νεφών περιγραφής εμφάνισης. Για να καταστεί κάτι τέτοιο δυνατό απαιτείται η συν-προσαρμογή των δύο νεφών ώστε αφού τα δύο νέφη ταιριάξουν χωρικά βέλτιστα (υπό κάποιο κριτήριο) να καταστεί δυνατή η μέτρηση των μεταξύ τους αποκλίσεων. Ένα λεπτό σημείο αποτελεί η δυνατότητα σχετικής μετακίνησης δύο “ $ab-Y$ ” νεφών μόνο κατά την  $Y$ -διεύθυνσή τους (κατά αντιστοιχία  $Z$ -διεύθυνση ενός « $XYZ$ » νέφους) καθώς τυχόν μετατόπιση ενός νέφους κατά τη διεύθυνση των χρωματικότητων θα οδηγούσε προφανώς σε αλλοίωση της χρωματικής τους πληροφορίας. Στο Κεφάλαιο 5 των αποτελεσμάτων παρουσιάζονται 2 τεχνικές συν-προσαρμογής και μέτρησης αποκλίσεων νεφών “ $ab-Y$ ”.

---

<sup>5</sup> Όλοι οι χώροι χρώματος είναι πρακτικά/φυσικά φραγμένοι ως προς τις χρωματικότητες τους καθώς περιορίζονται από το «φάσμα της ανθρώπινης όρασης» (gamut of human vision)

## 4) Μοντέλο Εμφάνισης “Levels”

Το δεύτερο, κατά σειρά παρουσίασης, πρωτότυπο μοντέλο εμφάνισης που αναπτύχθηκε και αξιολογήθηκε είναι το μοντέλο “Levels”, θεωρητική βάση του οποίου αποτελεί το μοντέλο “ab-Y”

### a) Περιγραφή / Μεθοδολογία

Εκκινώντας από τον χώρο προτύπων του μοντέλου “ab-Y” ορίζεται η έννοια της στάθμης (Level). Για δεδομένο εύρος της συνιστώσας Y [ $Y_{\min}, Y_{\max}$ ] υπολογίζεται ο μέσος όρος των χρωματικότητων a-b για το σύνολο των προβεβλημένων (στο χώρο “ab-Y”) εικονοστοιχείων<sup>6</sup> που βρίσκονται εντός του παραπάνω εύρους. Για κάθε στάθμη αντιστοιχούν δύο τιμές χρωματικότητας. Έπειτα από κατάλληλη διαμέριση<sup>7</sup> του συνολικού εύρους ([0,1]) της Y-διεύθυνσης προκύπτει ένα διάνυσμα περιγραφής που αποτελεί το τελικό προϊόν του μοντέλου. Να σημειωθεί ότι το ζήτημα σφαλμάτων κατά την Y-διεύθυνση υφίσταται και στο μοντέλο “Levels”. Για την επίλυσή του χρησιμοποιείται κατάλληλη (σχετικά υψηλή) επικάλυψη μεταξύ των σταθμών μειώνοντας έτσι την επίδραση των προαναφερθέντων σφαλμάτων.

### b) Χώρος Προτύπων

Ο τελικός χώρος προτύπων του περιγραφέα είναι ο  $n$ -διάστατος χώρος πραγματικών αριθμών, όπου  $n$  ο αριθμός των συνιστωσών του διανύσματος του τελικού περιγραφέα. Εν προκειμένω, η διάσταση  $n$  προκύπτει ως το διπλάσιο (για τις δύο χρωματικότητες) των αριθμών των σταθμών. Γενικά, το μήκος του διανύσματος είναι σχετικά μικρό (μερικές δεκάδες συνιστώσες) σε σχέση με λοιπούς περιγραφείς της Όρασης Υπολογιστών, παρ’ όλα αυτά τα αποτελέσματα (βλ. Κεφάλαιο 5) είναι ικανοποιητικά, γεγονός που υποδηλώνει ότι ο περιγραφέας επιτυγχάνει υψηλή συμπίεση της (σημασιολογικής) πληροφορίας.

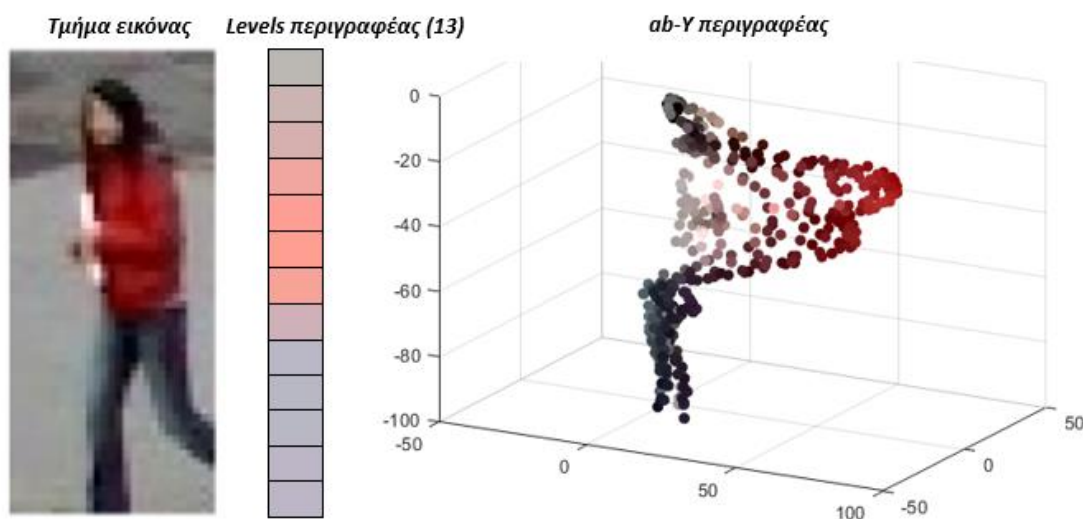
---

<sup>6</sup> Στο εξής στη συνέχεια της εργασίας ως εικονοστοιχεία θα αναφέρονται τα εικονοστοιχεία του τμήματος της εικόνας που έχουν χαρακτηριστεί ως «άνθρωπος» έπειτα από τη διαδικασία της σημασιολογικής κατάτμησης

<sup>7</sup> Διαμέριση συνόλου: Ένα σύνολο υποσυνόλων των οποίων η πράξη της ένωσης ισούται με το αρχικό σύνολο

### c) Διανυσματικός περιγραφέας

Εν αντιθέσει με το μοντέλο “ab-Υ” το “Levels” παράγει έναν διανυσματικό περιγραφέα καθιστώντας το πολύ πιο ευέλικτο. Στον συγκεκριμένο αλγόριθμο κάθε τμήμα της εικόνας αναπαρίσταται ως ένα σημείο σε έναν  $n$ -διάστατο χώρο και όχι ως ένα 3D νέφος, όπως στην περίπτωση του μοντέλου “ab-Υ”. Η χρήση διανυσματικών περιγραφέων τυγχάνει ευρείας εφαρμογής σε διάφορες διαδικασίες της Όρασης Υπολογιστών και της Μηχανικής Μάθησης κυρίως λόγω της ευκολίας μέτρησης των μεταξύ τους «αποστάσεων»/αποκλίσεων ως αποστάσεων (υπό κάποια νόρμα) στον  $n$ -διάστατο χώρο προτύπων που ανήκουν.



Εικόνα 37: Το μοντέλο εμφάνισης Levels και σύγκριση με patch και ab-Υ νέφος.

### d) Σχέση με “ab-Υ”

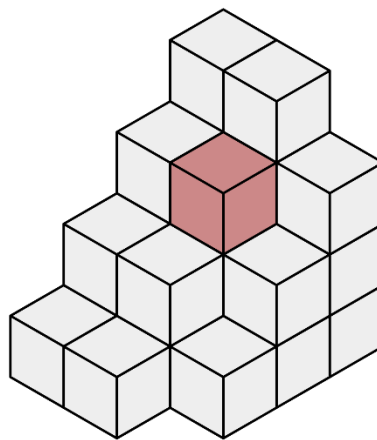
Όπως αναφέρεται και στην αρχική περιγραφή του προτύπου, ο περιγραφέας “Levels” έχει ως θεωρητική του αφετηρία το πρότυπο “ab-Υ”. Ουσιαστικά αποτελεί μία προσπάθεια διανυσματοποίησης του προτύπου “ab-Υ” με ταυτόχρονη συμπίεση της πληροφορίας μέσω της πράξης του μέσου όρου. Η έννοια της Υ-διεύθυνσης υποκαθίσταται από μία κατάλληλη διακριτοποίηση/διαμέρισή της εν είδει των σταθμών (Levels), καθιστώντας δυνατή τη διανυσματική περιγραφή του (πλήρους) χώρου “ab-Υ”. Στην συμπίεση της σημασιολογικής πληροφορίας συνεισφέρει, όπως προαναφέρθηκε, ο μέσος όρος των χρωματικότητων ανά στάθμη. Πρακτικά τα σημεία του νέφους “ab-Υ” που ανήκουν στην εκάστοτε στάθμη υποκαθίστανται από τον διανυσματικό τους μέσο στον υπόχωρο “a-b”.

## 5) Μοντέλο Εμφάνισης “Voxels”

Τρίτο, και τελευταίο κατά σειρά παρουσίασης, μοντέλο εμφάνισης είναι το πρότυπο “Voxels”. Το μοντέλο αυτό βασίζεται, όπως και το πρότυπο “Levels”, σε μεγάλο βαθμό στο πρότυπο “ab-Y” για τη θεωρητική του τεκμηρίωση. Ειδικότερα, ακολουθεί η αναλυτική περιγραφή του μοντέλου.

### a) Περιγραφή / Μεθοδολογία

Περιγράφοντας το πρότυπο “ab-Y” είχε επισημανθεί ότι τελικά το πεδίο τιμών του προτύπου “ab-Y” είναι φραγμένο στον ομώνυμο χώρο. Το γεγονός αυτό επιτρέπει την κβάντιση/διαμέριση του χώρου. Η κβάντιση αυτή προκύπτει από τη διαμέριση της κάθε συνιστώσας (a-b-Y) και τα ορθογώνια (υποσύνολα) προκύπτουν μέσω της πράξης του καρτεσιανού γινομένου. Πρακτικά, ο 3Δ χώρος “ab-Y” χωρίζεται σε ένα σύνολο πανομοιότυπων ορθογωνίων παραλληλεπίπεδων («ογκοστοιχεία»-voxels<sup>8</sup>), οπότε πλέον κάθε στοιχείο του νέφους “ab-Y” αντιστοιχεί κατά μοναδικό τρόπο σε κάποιο voxel.



Εικόνα 38: Οπτικοποίηση ενός ογκοστοιχείου (voxel). Πηγή: <http://wikipedia.org>

Το πρότυπο “Voxels” ουσιαστικά στηρίζεται στη μέτρηση της πυκνότητας στο χώρο “ab-Y”, περιγράφοντας το σημειοσύνολο του προτύπου “ab-Y” ως μία συνάρτηση πυκνότητας στον ομώνυμο χώρο. Για λόγους που αναλύονται παρακάτω η τιμή η οποία ανατίθεται σε κάθε voxel είναι δυαδική, δηλαδή το εκάστοτε voxel λαμβάνει την τιμή 0 αν δεν εμπεριέχει κάποιο σημείο του νέφους “ab-Y”, αλλιώς λαμβάνει την τιμή 1.

---

<sup>8</sup> Βλ. <https://en.wikipedia.org/wiki/Voxel>



## **b) Χώρος Προτύπων**

Ο τελικός χώρος προτύπων στον οποίο ανήκει ο περιγραφέας “Voxels” δεν είναι παρά ο  $\{0,1\}^n$   $n$ -διάστατος δυαδικός χώρος. Ο αριθμός της διάστασης  $n$  ισούται με το σύνολο των ογκοπηφίδων της κβάντισης του (φραγμένου) χώρου “ab-Y”.

## **c) Διανυσματικός περιγραφέας**

Όπως είναι ήδη φανερό ο περιγραφέας του προτύπου “Voxels” αποτελείται από ένα δυαδικό διάνυσμα. Το μήκος του διανύσματος είναι ίσο με τη διάσταση του χώρου προτύπων, γεγονός που σε συνδυασμό με τη λεπτομερή κβάντιση του χώρου οδηγεί σε διανύσματα με πολύ μεγάλο μήκος (δεκάδες χιλιάδες στοιχεία). Γι’ αυτόν το λόγο επιλέχθηκε ο συγκεκριμένος περιγραφέας να είναι δυαδικός, μειώνοντας έτσι την πολυπλοκότητά του και αντισταθμίζοντας το μεγάλο μήκος διανύσματος.

## **d) Σχέση με “ab-Y”**

Το πρότυπο “Voxels” είναι άμεσα συσχετισμένο με τον περιγραφέα “ab-Y” καθώς αποτελεί μία σχεδόν άμεση διανυσματοποίησή του. Συνεπώς τα προβλήματα του τελευταίου συνεχίζουν να υφίστανται, όπως το πρόβλημα της συν-προσαρμογής. Το συγκεκριμένο μπορεί να αντιμετωπισθεί μερικώς με χρήση (όπως και στον “Levels”) κατάλληλης επικάλυψης κατά την Y-διεύθυνση των ογκοπηφίδων.

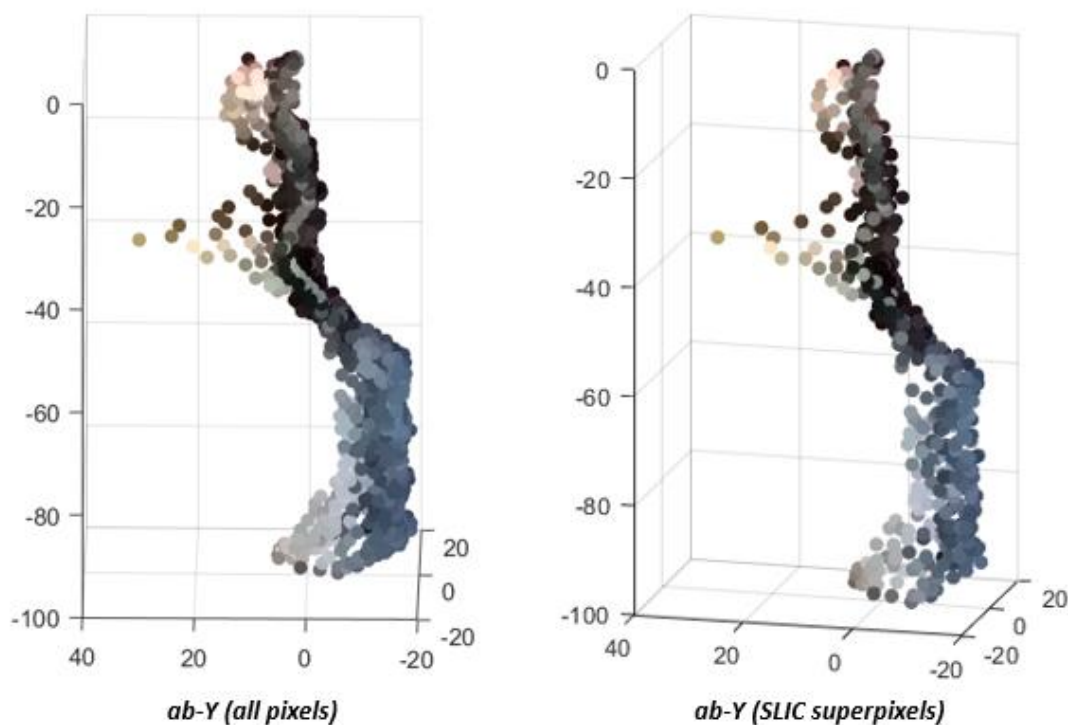
Στο πρότυπο “ab-Y” προκύπτει ότι η χωρική πυκνότητα μπορεί να επηρεάσει το «ταίριασμα» δύο νεφών (βλ. Κεφάλαιο 5). Ο περιγραφέας “Voxels”, αν και περιγραφέας πυκνότητας, δεν αντιμετωπίζει τέτοιο πρόβλημα εξαιτίας της δυαδικότητάς του. Ουσιαστικά μετριέται για το εκάστοτε νέφος “ab-Y” η δυαδική πυκνότητά του στο χώρο “ab-Y”.

## 6) Επισκόπηση και ζητήματα επίδοσης

Συνολικά, στα πλαίσια της παρούσας εργασίας αναπτύχθηκαν 3 μοντέλα εμφάνισης αντικειμένων/ ανθρωπίνων στόχων. Το πρώτο και σημαντικότερο σε θεωρητικό επίπεδο των τριών αποτελεί το μοντέλο “ab-Y”. Το συγκεκριμένο πρότυπο εκφράζει τη θεμελιώδη περιγραφή των εικονοστοιχείων ως ένα 3D νέφος σε ένα χώρο αποτελούμενο από δύο χρωματικές και μία χωρική συνιστώσα. Η παραπάνω σημασιολογική περιγραφή των εικονοστοιχείων αποτελεί τη βάση και των έτερων δύο αναπτυχθέντων μοντέλων εμφάνισης, του μοντέλου “Levels” και του μοντέλου “Voxels”. Τα δύο αυτά μοντέλα απλουστεύουν (άλλο περισσότερο και άλλο λιγότερο) την περιγραφή του νέφους “ab-Y”, όμως το τελικό προϊόν τους, δηλαδή ένα διάγραμμα, είναι υπολογιστικά πολύ πιο διαχειρίσιμο και εύκολο στην ενσωμάτωση σε πολλές, ήδη υπάρχουσες υλοποιήσεις και εφαρμογές, με απαιτήσεις για ανάλυση σε σχεδόν πραγματικό χρόνο, διαδικασίες της Όρασης Υπολογιστών και της Μηχανικής Μάθησης.

Συγκεκριμένα, σημαντικό ρόλο στην καθιέρωση μιας μεθοδολογίας, όπως τα μοντέλα εμφάνισης, ως επιχειρησιακά αξιοποιήσιμη παίζει η υπολογιστική επίδοσή της. Τα αναπτυχθέντα πρωτότυπα μοντέλα δεν αποδίδουν βέλτιστα στον συγκεκριμένο τομέα καθώς χαρακτηρίζονται από το υψηλό υπολογιστικό τους κόστος. Ως ένας γενικός κανόνας της χρονικής αποδοτικότητας των αναπτυχθέντων αλγορίθμων προκύπτει η αναλογία του χρόνου υπολογισμού με τον αριθμό εικονοστοιχείων.

Λαμβάνοντας υπόψιν τον παραπάνω εμπειρικό κανόνα υλοποιήθηκε ένα στάδιο προεπεξεργασίας του εκάστοτε τμήματος της εικόνας που περιλαμβάνει τη δημιουργία υπερεικονοστοιχείων με τη μέθοδο SLIC (βλέπε Κεφάλαιο 2). Κατά αυτόν τον τρόπο μειώνεται ο αριθμός των εικονοστοιχείων κατά τουλάχιστον 1-2 τάξεις μεγέθους βελτιώνοντας κατ’ αναλογία και τη χρονική αποδοτικότητα του αλγορίθμου, χωρίς να επηρεάζεται σημαντικά η «μορφή» του νέφους ab-Y όπως φαίνεται στην παρακάτω εικόνα. Στην περίπτωση αυτή κάθε «νέο» (υπερ)εικονοστοιχείο συνοδεύεται από ένα βάρος ανάλογο του αριθμού των πρωτογενών εικονοστοιχείων που περιέχει και με αυτό συμμετέχει στα μοντέλα εμφάνισης που αναπτύχθηκαν. Άξιο αναφοράς είναι ότι για μια τυπική ανίχνευση του ανιχνευτή ACF [6] ο αριθμός των υπερεικονοστοιχείων που επαρκεί για την ικανοποιητική εφαρμογή του μοντέλου εμφάνισης του ανθρώπινου στόχου δεν υπερβαίνει τα 700 υπερεικονοστοιχεία (από >7000 εικονοστοιχεία).



Εικόνα 39: *ab-Y* νέφος με και χωρίς τη χρήση υπερεικονοστοιχείων. Το σχήμα διατηρείται, η πυκνότητα μειώνεται.

Σημαντική επίδραση στην αποδοτικότητα έχει η (μη γραμμική) μετατροπή από τον χώρο χρώματος RGB στον CIE<sub>Lab</sub>. Η συγκεκριμένη μετατροπή μεταξύ χώρων χρώματος είναι από τις πλέον υπολογιστικά κοστοβόρες και ίσως θα είχε ενδιαφέρον η διερεύνηση διαφορετικών χώρων χρώματος που στηρίζονται στην έννοια των δύο χρωματικοτήτων όπως ο *Luv*, *Ycb-cr*, *HSV/I/L* ή ακόμα και ο (γραμμικός) *XYZ*. Κανένας όμως από τους παραπάνω χώρους δεν είναι ομοιόμορφος και για αυτόν τον λόγο δεν επιχειρήθηκε η παραπάνω διερεύνηση στα πλαίσια της παρούσας εργασίας σε συνδυασμό με το γεγονός ότι η υψηλή επίδοση δεν συμπεριλαμβανόταν στους εξ αρχής στόχους της εργασίας.

## ΚΕΦΑΛΑΙΟ 5. ΠΕΙΡΑΜΑΤΙΚΑ ΑΠΟΤΕΛΕΣΜΑΤΑ ΚΑΙ ΑΞΙΟΛΟΓΗΣΗ

### 1) Αλγόριθμοι και μετρικές σφάλματος

Για τα πειράματα της παρούσας μεταπτυχιακής εργασίας αναπτύχθηκαν και χρησιμοποιήθηκαν οι 3 πρωτότυποι αλγόριθμοι μοντελοποίησης της εμφάνισης ενός ανθρώπου-στόχου που αναλύθηκαν σε πρότερα κεφάλαια της εργασίας. Ειδικότερα, αναπτύχθηκαν οι αλγόριθμοι “ab-Y Cloud”, “Levels Features” και “Voxels Features”. Για κάθε αλγόριθμο επιλέχθηκαν από δύο μεθοδολογίες μέτρησης ομοιότητας μεταξύ των στόχων. Πιο συγκεκριμένα για το πρώτο μοντέλο (“ab-Y”), εξαιτίας της μη διανυσματικής φύσης της περιγραφής που προκύπτει για τον στόχο, απαιτείται για τη σύγκριση μεταξύ δύο στόχων ένα βήμα συνταύτισης των δύο, προς σύγκριση, νεφών περιγραφής. Η πρώτη μεθοδολογία για τον αλγόριθμο αυτό είναι η χρήση του αλγορίθμου του «Εγγύτερου σημείου» (βλ. παρακάτω) για την συνταύτιση των δύο νεφών και την μετρική του αμφίδρομου ενδιάμεσου σφάλματος (βλ. παρακάτω). Όσον αφορά τη δεύτερη μεθοδολογία μέτρησης ομοιότητας/αποκλίσεων χρησιμοποιήθηκε μια απλουστευμένη παραλλαγή του κλασικού αλγορίθμου συν-προσαρμογής (coregistration) 3D νεφών σημείων “ICP”, υπό την ονομασία “ICP-Z”, συνδυσασμένου με τη μετρική του μέσου τετραγωνικού σφάλματος (ΜΤΣ - RMSE).

Οι υπόλοιποι δύο αλγόριθμοι μοντελοποίησης (“Levels” και “Voxels”) δεν απαιτούν κάποιο βήμα προεπεξεργασίας καθώς, όντας διανυσματικοί περιγραφείς, τα διανύσματα περιγραφής της εμφάνισης ανήκουν σε κάποιο πολυδιάστατο μαθηματικό χώρο με νόρμα, καθιστώντας την μέτρηση της ομοιότητάς τους ένα πρόβλημα μέτρησης απόστασης μεταξύ δύο σημείων υπό κάποια νόρμα σε έναν  $n$ -διάστατο χώρο ( $\mathbb{R}^n, \|\cdot\|_p$ ). Το παραπάνω πρόβλημα είναι καλώς ορισμένο και η λύση του απλή. Εν προκειμένω, επιλέχθηκαν οι  $L_p$  νόρμες για  $p = 2$  (επίσης γνωστή ως «Ευκλείδεια») και  $p = 1$  (γνωστή και ως «City-block»).

## a) Αλγόριθμοι συν-προσαρμογής νεφών “ab-Y”

### (i) ICP-Z

Η μέθοδος του ICP (Iterative Closest Point) αποτελεί την πιο συνηθισμένη και αξιόπιστη επιλογή για την συν-προσαρμογή δύο 3D σημειοσυνόλων. [4] Συνοπτικά πρόκειται για μία επαναληπτική διαδικασία η οποία σε κάθε επανάληψη «φέρνει πιο κοντά» τα δύο νέφη ελαχιστοποιώντας (τοπικά όχι ολικά) μία συνάρτηση κόστους βασισμένη σε κάποια μετρική σφάλματος. Στη βιβλιογραφία [5] έχουν ανά καιρούς προταθεί διάφορες παραλλαγές του πρωτότυπου αλγορίθμου.

Στην παρούσα εργασία έχει αναπτυχθεί και χρησιμοποιηθεί μία πρωτότυπη απλουστευμένη παραλλαγή του ICP αλγορίθμου, υπό τον τίτλο ICP-Z. Ουσιαστικά πρόκειται για μία απλούστευση του προβλήματος συν-προσαρμογής δύο σημειοσυνόλων καθώς επιτρέπεται ελευθερία σχετικής κίνησης μόνο κατά τον Z-άξονα. Υπό αυτήν τη συνθήκη το πρόβλημα γίνεται μονοδιάστατο. Σε κάθε επανάληψη του αλγορίθμου υπολογίζεται η μέση κατά Z σημειακή απόσταση μεταξύ των δύο νεφών. Με την θεώρηση ότι το ένα σύνολο είναι σταθερό και το άλλο κινούμενο, μετακινείται το δεύτερο κατά την προηγούμενη μέση απόσταση, κάνοντας τα μοντέλα να πλησιάζουν. Η διαδικασία συνεχίζεται μέχρι η μεταβολή του ΜΤΣ μεταξύ των δύο μοντέλων να μην υπερβαίνει κάποιο κατώφλι μεταξύ δύο συνεχόμενων επαναλήψεων.

Σε πρώτη ανάγνωση μία τέτοια διαδικασία θα ήταν υπερβολική για την επίλυση ενός 1D προβλήματος, όμως εν προκειμένω το πρόβλημα γίνεται δυσκολότερο καθώς δεν υπάρχει καλά ορισμένη η έννοια της «σημειακής ομολογίας» μεταξύ των δύο σημειοσυνόλων. Για παράδειγμα δεν είναι αναγκαίο κάποιο σημείο του σταθερού σημειοσυνόλου να έχει το ίδιο ομόλογο σημείο στο κινούμενο σημειοσύνολο στην πρώτη και στη δεύτερη επανάληψη του αλγορίθμου. Για τον λόγο αυτό αναπτύχθηκε μία καλώς ορισμένη έννοια «ομολογίας» η οποία αναλύεται περαιτέρω σε παρακάτω εδάφιο περί μετρικών σφάλματος και ειδικότερα στο εδάφιο του Μέσου Τετραγωνικού Σφάλματος.

## (ii) Εγγύτερου σημείου

Ο αλγόριθμος του ICP αποτελεί μία πολύ ισχυρή μέθοδο για την συν-προσαρμογή σημειοσυνόλων και η 1Δ παραλλαγή του, ο ICP-Z, μία πολύ δυνατή μέθοδο συν-προσαρμογής “ab-Y” σημειοσυνόλων. Το μεγάλο τους μειονέκτημα αποτελεί η χαμηλή τους αποδοτικότητα, γεγονός που οφείλεται κατά κύριο λόγο στην επαναληπτική τους φύση. Η έννοια της αποδοτικότητας είναι μεγίστης σημασίας σε πολλαπλές εφαρμογές των μοντέλων εμφάνισης ανθρώπινων στόχων όπως η οπτική παρακολούθηση, εφαρμογή η οποία απαιτεί τον υπολογισμό πολλαπλών μοντέλων εμφάνισης (πιθανώς δεκάδες) σε κάθε frame. Το πρόβλημα της αποδοτικότητας προσπαθεί να βελτιώσει ο αλγόριθμος του «Εγγύτερου Σημείου».

Η μέθοδος του «Εγγύτερου Σημείου» στηρίζεται στον ICP-Z, όμως δεν είναι επαναληπτική. Στην πραγματικότητα είναι το αποτέλεσμα της πρώτης επανάληψης του ICP-Z αλγορίθμου. Η συγκεκριμένη επιλογή έγινε διότι, έπειτα από πολλαπλές δοκιμές, προέκυψε ότι ο ICP-Z παρουσιάζει στις πρώτες του επαναλήψεις πολύ υψηλές ταχύτητες σύγκλισης και ήδη από την πρώτη επανάληψη τα δύο μοντέλα είναι αρκετά κοντά για πληθώρα εφαρμογών. Στο γεγονός αυτό της ταχείας σύγκλισης υποβοηθά σημαντικά και το γεγονός ότι ο ICP-Z επιλύει ένα πρακτικά μονοδιάστατο πρόβλημα σε αντίθεση με τον πλήρη ICP ο οποίος αντιμετωπίζει ένα σαφώς πιο πολύπλοκο πρόβλημα τριών διατάσεων.

## b) Μετρικές σφάλματος

### (i) Μέσο τετραγωνικό σφάλμα

Έστω  $Y$  ένα διάνυσμα από  $n$  παρατηρήσεις ενός (αληθούς) μοντέλου και  $Y^*$  το αντίστοιχο διάνυσμα προβλέψεων βάσει ενός προσεγγιστικού μοντέλου (πιθανώς κάποιο βέλτιστο υπό κάποιο κριτήριο προσσαρμοσμένο μοντέλο). Τότε ορίζεται το **Μέσο Τετραγωνικό Σφάλμα** (Mean Squared Error - MSE) ως εξής:

$$MSE = \frac{\sum_{i=1}^n (Y_i - Y_i^*)^2}{n}$$

Στα πειράματα που διενεργήθηκαν το ΜΤΣ χρησιμοποιήθηκε για την μέτρηση της μέσης απόστασης μεταξύ δύο (κατάλληλα συνταυτισμένων)

“ab-Y” νεφών. Προφανώς η παραπάνω σχέση δεν τυγχάνει άμεσης εφαρμογής, αλλά ορίζεται η έννοια της «ομολογίας» μεταξύ των δύο νεφών. Για να ορισθεί η παραπάνω έννοια ένα από τα δύο νέφη θεωρείται ως νέφος αναφοράς (κατά σύμβαση το σταθερό στον ICP-Z) και για κάθε σημείο του ορίζεται ως ομολόγό του το εγγύτερο σε αυτό σημείο του άλλου νέφους. Θεωρώντας ως σφάλμα την μεταξύ τους απόσταση, τότε το ΜΤΣ είναι καλώς ορισμένο για τη σύγκριση δύο νεφών “ab-Y”.

### **(i) Αμφίδρομο ενδιάμεσο σφάλμα**

Το μεγαλύτερο πρόβλημα της προηγούμενης μετρικής σφάλματος, του Μέσου Τετραγωνικού Σφάλματος, δεν είναι παρά η ανάγκη για θεώρηση ενός νέφους «αναφοράς». Κατά αυτόν τον τρόπο μετρείται μονάχα η προσαρμογή του νέφους αναφοράς στο (κινούμενο στον ICP-Z) έτερο νέφος. Για παράδειγμα έστω ότι το νέφος αναφοράς ήταν ένα επίπεδο στον χώρο “ab-Y” και το άλλο νέφος αποτελούνταν από δύο παράλληλα επίπεδα, με το ένα να συμπίπτει (με μικρό σφάλμα) με το επίπεδο του νέφους αναφοράς. Στην περίπτωση αυτή το ΜΤΣ θα λάμβανε εξαιρετικά χαμηλές τιμές, μην αποτυπώνοντας ορθά τη διαφορετικότητα των δύο μοντέλων εμφάνισης. Επιπροσθέτως, επειδή το ΜΤΣ στηρίζεται στην αρχή του μέσου όρου είναι πολύ ευαίσθητο στις ακραίες παρατηρήσεις. Συνεπώς για να είναι αντικειμενικό το ΜΤΣ θα πρέπει τα δύο μοντέλα να έχουν και παρόμοια χωρική πυκνότητα στα σημεία των νεφών τους.

Τα προαναφερθέντα δύο προβλήματα επιλύει το **Αμφίδρομο Ενδιάμεσο Σφάλμα**. Το ΑΕΣ ουσιαστικά υπολογίζεται σε δύο φάσεις. Στην πρώτη θεωρεί κανείς το ένα μοντέλο ως μοντέλο αναφοράς και ορίζοντας την έννοια της ομολογίας με τον ίδιο τρόπο με το ΜΤΣ υπολογίζει την ενδιάμεση τιμή του σφάλματος (median error) για το σύνολο των παρατηρούμενων ομολογιών. Στη δεύτερη φάση επαναλαμβάνεται η ίδια διαδικασία με εναλλαγή των δύο νεφών (το δεύτερο ως αναφοράς πλέον) και το τελικό ΑΕΣ υπολογίζεται ως ο μέσος όρος των δύο ενδιάμεσων σφαλμάτων.

Συνολικά το ΑΕΣ αποτελεί μια πιο ολοκληρωμένη μετρική σφάλματος μεταξύ δύο νεφών “ab-Y” καθώς απαιτεί (θεωρούμενο ως πρόβλημα εύρεσης ελαχίστου) συνολικά στο χώρο “ab-Y” να «πλησιάζουν» τα μοντέλα. Συγχρόνως είναι αρκετά εύρωστο σε (περιστασιακές) διακυμάνσεις μεταξύ των δύο μοντέλων, γεγονός εξαιρετικά χρήσιμο και επιθυμητό αν θεωρηθεί ότι, για παράδειγμα, σε ένα σύστημα οπτικής παρακολούθησης (Visual

Tracking) το μοντέλο εμφάνισης του εκάστοτε στόχου έχει προκύψει, κατά πάσα περίπτωση, από πολλαπλά καρέ, και συνεπώς από συνεκτίμηση της πληροφορίας πολλαπλών “ab-Y’ νεφών.

## (ii) $L^p$ – νόρμες

Έστω ο  $n$ -διάστατος πεπερασμένων διατάσεων πραγματικός διανυσματικός χώρος  $\mathbb{R}^n$ . Μία συνάρτηση  $\|\cdot\|: \mathbb{R}^n \rightarrow \mathbb{R}$  είναι μία **νόρμα** στον  $\mathbb{R}^n$  αν ισχύουν τα παρακάτω  $\forall \mathbf{u}, \mathbf{v} \in \mathbb{R}^n$  και  $\forall a \in \mathbb{R}$ :

- $\|a * \mathbf{u}\| = |a| * \|\mathbf{u}\|$  (απόλυτη ομογένεια)
- $\|\mathbf{u}\| \geq 0$  (μη αρνητικότητα)
- $\|\mathbf{u}\| = 0 \Rightarrow \mathbf{u} = \mathbf{0}_{\mathbb{R}^n}$  (διαχωρίζει τα σημεία)
- $\|\mathbf{u} + \mathbf{v}\| \leq \|\mathbf{u}\| + \|\mathbf{v}\|$  (υπο-προσθετικότητα)

Στους παραπάνω χώρους ορίζεται η **p-νόρμα** ή αλλιώς  **$L^p$ -νόρμα** η συνάρτηση  $\forall p > 0$ :

$$\|\mathbf{u}\|_p = \left( \sum_{i=1}^n |u_i|^p \right)^{\frac{1}{p}} \quad \forall \mathbf{u} \in \mathbb{R}^n$$

Για  $p = \infty$  ορίζεται η νόρμα

$$\|\mathbf{u}\|_\infty = \max_i |u_i| \quad \forall \mathbf{u} \in \mathbb{R}^n$$

Για  $p = 0$  η συνάρτηση που προκύπτει δεν είναι νόρμα και η τιμή της είναι ο αριθμός των μη μηδενικών συνιστωσών σε ένα διάνυσμα του χώρου.

Στη διεθνή βιβλιογραφία έχει επικρατήσει να λέγεται ως «Ευκλείδεια» η νόρμα-2, ενώ συχνά η νόρμα-1 συναντάται και ως μετρική City-Block ή/και Manhattan.

Να σημειωθεί ότι η έννοια της απόστασης και της νόρμας είναι άμεσα συσχετισμένες καθώς η απόσταση δυο σημείων σε έναν χώρο με νόρμα μετρείται υπό την νόρμα αυτού του χώρου ως η τιμή της νόρμας της διαφοράς των δύο σημείων. Η ισοδύναμη:

$$dist(\mathbf{u}, \mathbf{v})_p = \|\mathbf{u} - \mathbf{v}\|_p \quad \forall \mathbf{u}, \mathbf{v} \in \mathbb{R}^n$$

Στα πλαίσια της συγκεκριμένης εργασίας η νόρμα αξιοποιείται ποικιλοτρόπως, με κύρια εφαρμογή της την μέτρηση αποκλίσεων/αποστάσεων



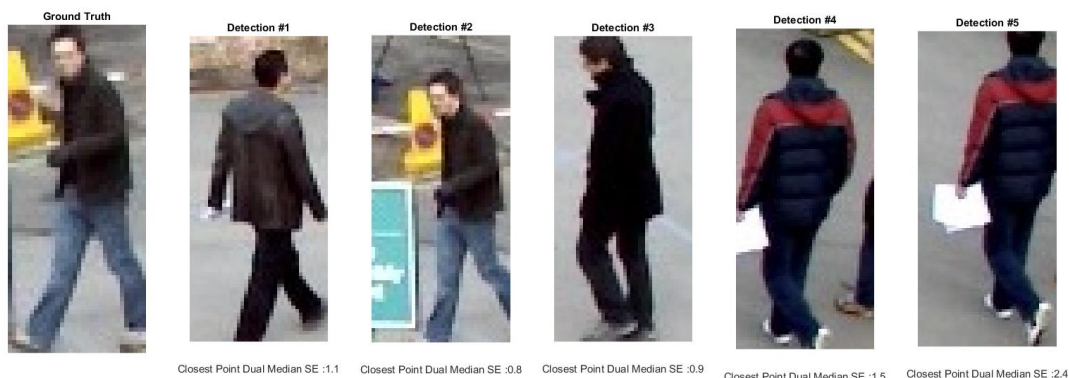
μεταξύ διανυσματικών περιγραφών (π.χ. “Levels” και “Voxels”) και υπό την έννοια του ελαχίστου την ομοιότητα τους

## 2) Αναγνώριση ενός στόχου σε ένα καρτέ

Στο συγκεκριμένο πείραμα έγινε προσπάθεια αντιστοίχισης ενός στόχου που έχει προσδιορισθεί σε κάποιο καρτέ ενός βίντεο σε κάποια από τις ανιχνεύσεις του επόμενου καρτέ. Για τον σκοπό αυτό τα 3 μοντέλα περιγραφής εμφάνισης που αναπτύχθηκαν εφαρμόστηκαν στο patch του εκάστοτε στόχου και η σύγκριση έγινε με τη χρήση 2 μετρικών μέτρησης σφάλματος. Ως καλύτερο «ταίριασμα» σε κάθε δοκιμή θεωρείται η ανίχνευση που πέτυχε τη μικρότερη τιμή στη μετρική σφάλματος. Ειδικότερα για το μοντέλο “ab-Y” χρησιμοποιήθηκαν ως μετρικές σφάλματος μεταξύ των παραγόμενων «νεφών» το «αμφίδρομο» ενδιάμεσο σφάλμα (σε συνδυασμό με τον αλγόριθμο συνταύτισης «Εγγύτερου Σημείου») και το μέσο τετραγωνικό σφάλμα (σε συνδυασμό με τον αλγόριθμο συνταύτισης ICP-Z). Όσον αφορά τα μοντέλα “Levels” και “Voxels” ως μετρικές σφάλματος χρησιμοποιήθηκαν η Ευκλείδεια απόσταση και η City-block ( $L_1$ ) απόσταση μεταξύ των παραγόμενων περιγραφών.

### a) abY πρότυπο

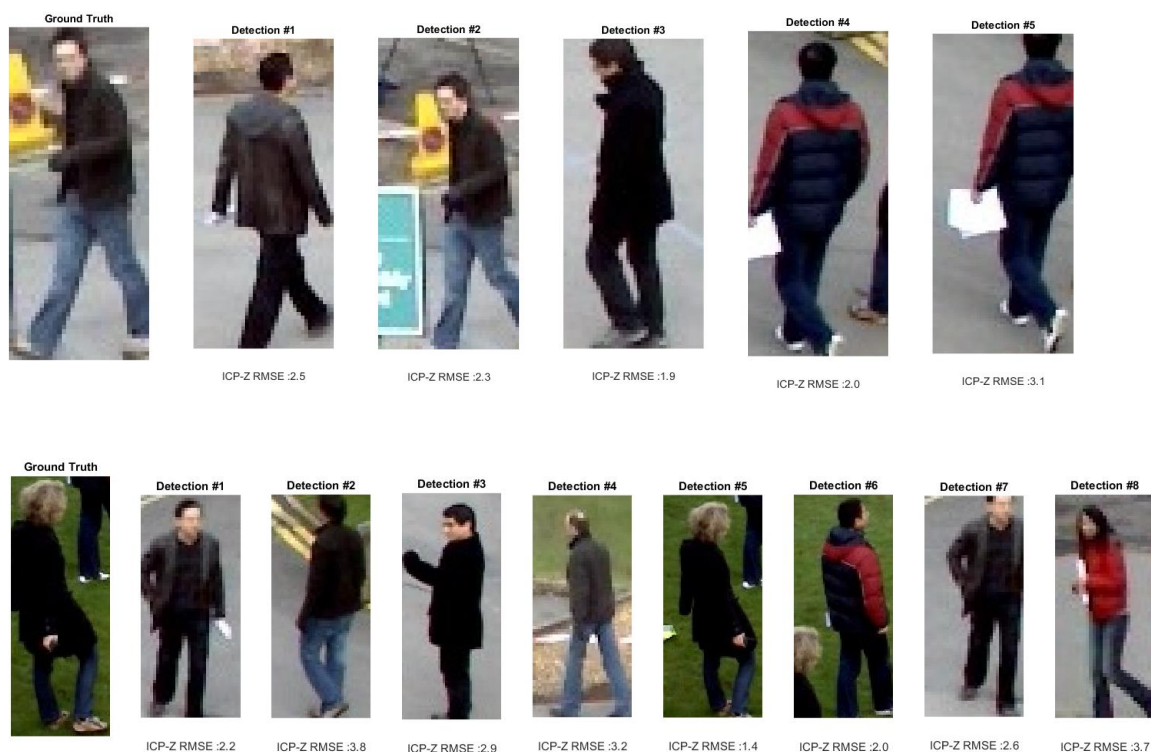
#### (i) *Closest Point algorithm και διπλό ενδιάμεσο σφάλμα*





Παρατηρώντας τα παραπάνω αποτελέσματα διαπιστώνει κανείς ότι ο στόχος ταυτοποιήθηκε με επιτυχία και στις δύο δοκιμές. Παρ' όλα αυτά όμως, είναι φανερό πως η μετρική σφάλματος απέτυχε να διαχωρίσει με μεγάλη σιγουριά τους στόχους. Ειδικότερα στην πρώτη περίπτωση η ελάχιστη τιμή της μετρικής είναι 0,8 με την αμέσως επόμενη στο 0,9, ενώ στη δεύτερη περίπτωση παρατηρούνται ανάλογα φαινόμενα με την ελάχιστη τιμή στο 0,5 και την αμέσως επόμενη στο 0,6.

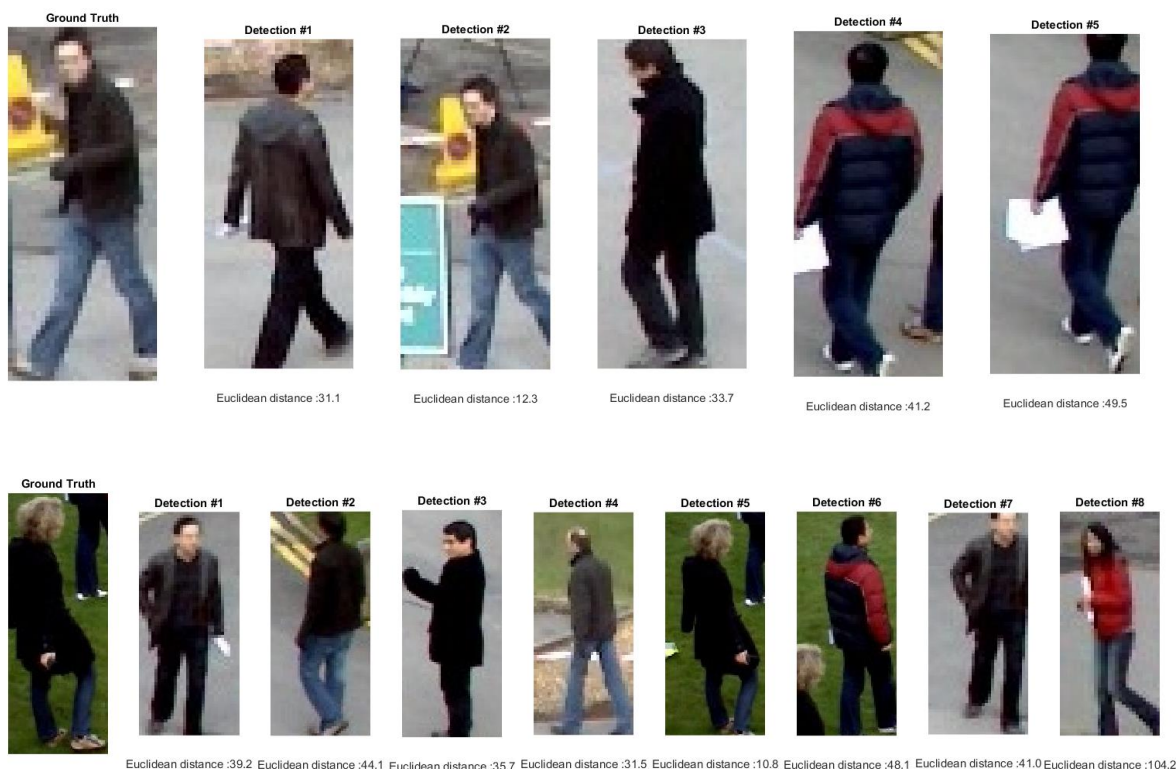
## (ii) ICP-Z και Μέσο Τετραγωνικό σφάλμα



Η συγκεκριμένη επιλογή αλγορίθμου συν-προσαρμογής και μετρικής σφάλματος φαίνεται να αποδίδει χειρότερα από την προηγούμενη. Στο πρώτο πείραμα, ο αλγόριθμος απέτυχε πλήρως να ταυτοποιήσει τον στόχο, καθώς τον κατέταξε ως το 3<sup>ο</sup> πιο πιθανό «ταίριασμα» με τιμή μετρικής σχεδόν 22% μεγαλύτερη από την ελάχιστη (1,9). Αντίθετη εικόνα παρατηρείται στο 2<sup>ο</sup> πείραμα όπου ο στόχος αναγνωρίζεται επιτυχώς, ενώ επίσης εξασφαλίζει αρκετά καλό διαχωρισμό από το 2<sup>ο</sup> καλύτερο «ταίριασμα» της τάξης του 43% (τιμές μετρικής 1<sup>ου</sup> - 2<sup>ου</sup> :1,4 - 2,0).

## b) Levels πρότυπο

### (i) Ευκλείδεια μετρική



Η χρήση διανυσματικών περιγραφών, όπως του περιγραφέα “Levels” εν προκειμένω, φαίνεται, σε πρώτη φάση τουλάχιστον, να παράγει αποτελέσματα σημαντικά καλύτερα σε σχέση με τους περιγραφείς σημειοσυνόλων όπως ο “ab-Y”. Στο συγκεκριμένο πείραμα προκύπτει ότι ο αλγόριθμος ταυτοποίησης λειτουργήσε εξαιρετικά, αναγνωρίζοντας ορθά τους

στόχους στις ανιχνεύσεις και πετυχαίνοντας πολύ υψηλά ποσοστά διαχωρισμού του καλύτερου «ταιριάσματος». Τα ποσοστά αυτά ανέρχονται σε 152% για το πρώτο πείραμα και σε 192% για το δεύτερο.

## (ii) *City block* μετρική



Στο παρόν πείραμα η μετρική σφάλματος City-block αποδίδει καλύτερα σε σχέση με την αντίστοιχη Ευκλείδεια. Πέραν της δεδομένης επιτυχούς ορθής αντιστοίχισης στόχων ανιχνεύσεων, επιτυγχάνονται και υψηλότερα ποσοστά διαχωρισμού πρώτου με δεύτερου «ταιριάσματος». Για την ακρίβεια αυτά ανέρχονται σε 140% και 212% (έναντι 152% και 192% για την Ευκλείδεια μετρική).

## c) Voxels πρότυπο

### (i) Ευκλείδεια μετρική



Οι παρατηρήσεις που διατυπώθηκαν παραπάνω περί επιδόσεων διανυσματικών περιγραφών έναντι των ομολόγων τους που χρησιμοποιούν σημειοσύνολα ισχύουν και για την περίπτωση του περιγραφέα “Voxels”. Με χρήση της Ευκλείδειας απόστασης ως μετρικής σφάλματος ολοκληρώθηκε επιτυχώς η ορθή αντιστοίχιση στόχου - ανιχνεύσεων. Τα ποσοστά διαχωρισμού κυμάνθηκαν σε 19% και 46% για τα δύο πειράματα. Γενικά το μοντέλο εμφάνισης “Levels” απέδωσε καλύτερα στη συγκεκριμένη εφαρμογή σε σχέση με το “Voxels”, με μετρική σφάλματος την Ευκλείδεια απόσταση.

(ii) *City block* μετρική



Παρόμοια εικόνα με προηγουμένως παρατηρείται και για τη μετρική City-block. Όλες οι ταυτοποιήσεις ήταν ορθές και τα ποσοστά διαχωρισμού κυμάνθηκαν σε 41% και 115%, αντίστοιχα. Γενικά διαφαίνεται ότι η διαχωριστική ικανότητα της μετρικής City-block είναι ανώτερη εκείνης της Ευκλείδειας ομολογής της. Το συμπέρασμα αυτό προκύπτει και για τους δύο περιγραφείς μοντέλων εμφάνισης ανθρώπινων στόχων, τα “Levels” και τα “Voxels”.

### 3) Αντιστοίχιση ανιχνεύσεων και στόχων από καρτέ σε καρτέ

Στο συγκεκριμένο πείραμα έγινε προσπάθεια αντιστοίχισης ενός συνόλου στόχων που έχουν προσδιορισθεί σε κάποιο καρτέ ενός βίντεο στο σύνολο των ανιχνεύσεων του επόμενου καρτέ. Οι μετρικές σφάλματος όπως και τα μοντέλα περιγραφής εμφάνισης που χρησιμοποιήθηκαν ήταν τα ίδια με το προηγούμενο πείραμα. Η αντιστοίχιση επιτελέστηκε μέσω της εύρεσης του συνδυασμού (στόχων-ανιχνεύσεων) ελαχίστου κόστους (απόστασης) με τη χρήση του αλγορίθμου αντιστοίχισης “Munkre’s Hungarian Algorithm” [30]. Τα αποτελέσματα που προέκυψαν ήταν τα εξής:



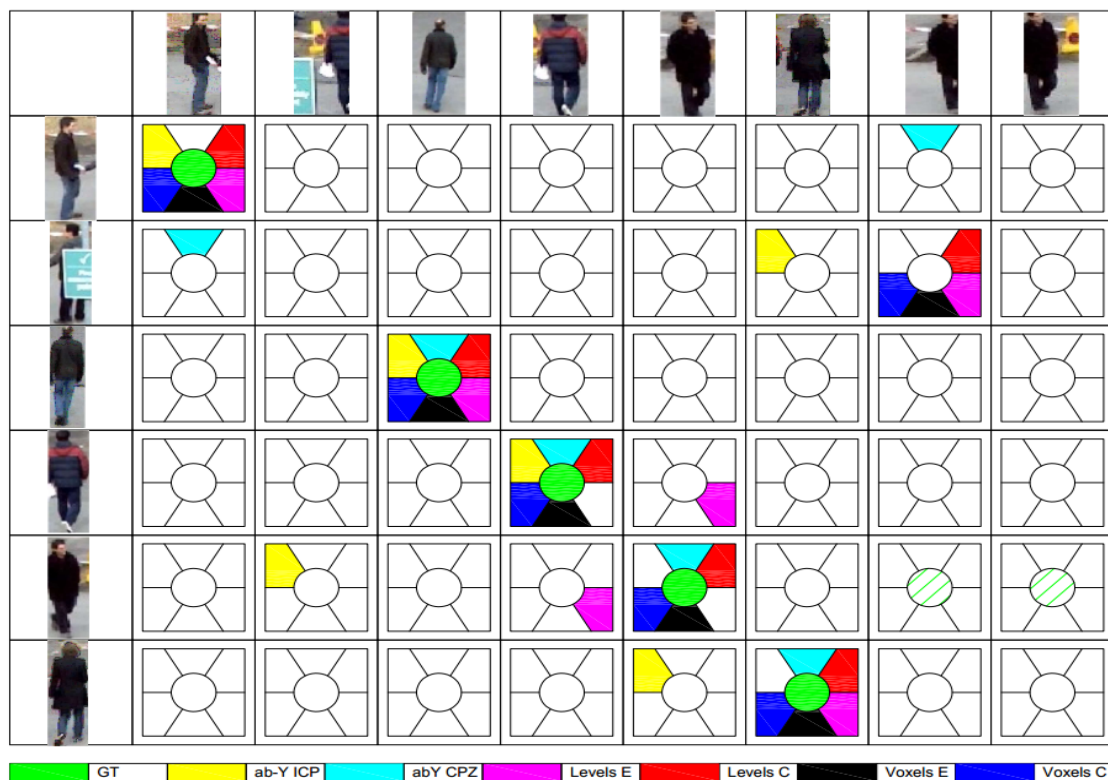
**Καρέ 311 / Βίντεο 'PETS09-S2L1':**

**Frame 311 - Data Association**



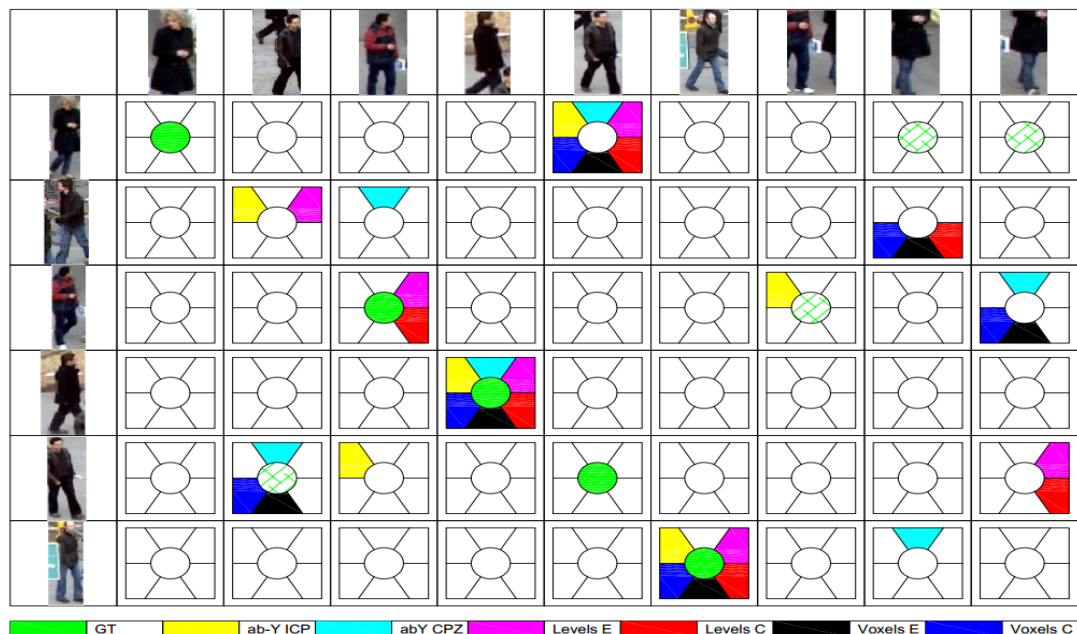
**Καρέ 84 / Βίντεο 'PETS09-S2L1':**

**Frame 84 - Data Association**



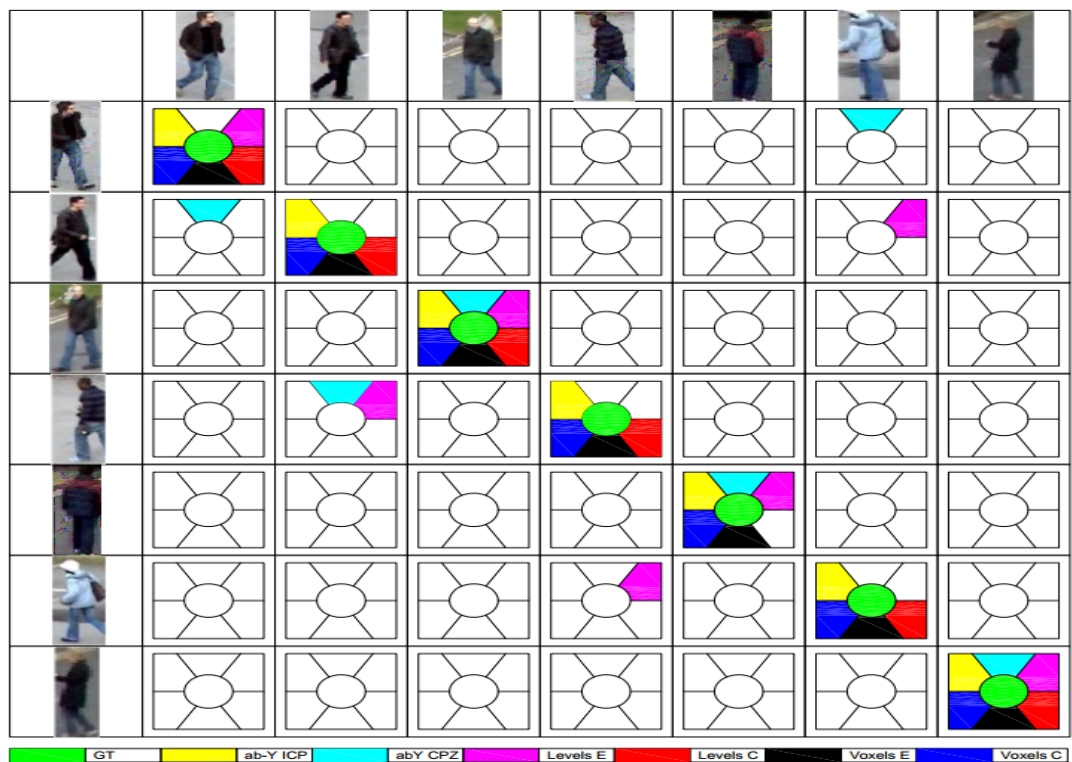
**Καρέ 344 / Βίντεο 'PETS09-S2L1':**

Frame 344 - Data Association



**Καρέ 180 / Βίντεο 'PETS09-S2L1':**

Frame 180 - Data Association



Αναλυτικά τα αποτελέσματα σε πίνακα για όλες τις δοκιμές:

Αλγόριθμος	Καρέ #311	Καρέ #84	Καρέ #344	Καρέ #180	Μ.Ο.
<b>Ab-Y (CPZ)</b>	33%	67%	33%	43%	44%
<b>Ab-Y (ICP-Z)</b>	50%	50%	50%	100%	63%
<b>Levels Ευκλειδεια</b>	33%	50%	50%	57%	48%
<b>Levels City-block</b>	50%	100%	50%	100%	75%
<b>Voxels Ευκλειδεια</b>	33%	100%	50%	100%	71%
<b>Voxels City-block</b>	33%	100%	50%	100%	71%

Παρατηρώντας τα αποτελέσματα του πειράματος βλέπει κανείς ότι ο αλγόριθμος συσχέτισης στόχων-ανιχνεύσεων με χρήση του μοντέλου “ab-Y” και μετρική σφάλματος το Αμφίδρομο Ενδιάμεσο Σφάλμα με συν-προσαρμογή μέσω του αλγορίθμου του Εγγύτερου Σημείου δεν απέδωσε καλά (ab-Y CPZ). Ο μέσος όρος ακρίβειας για τις 4 δοκιμές κυμάνθηκε στο (χαμηλό) 44%. Το γεγονός αυτό οφείλεται στη χαμηλή διαχωριστική ικανότητα του συγκεκριμένου συνδυασμού περιγραφέα-μετρικής σφάλματος, γεγονός που τονίστηκε και στο προηγούμενο πείραμα που περιλάμβανε τον παραπάνω συνδυασμό.

Όσον αφορά τον δεύτερο συνδυασμό, ο μέσος όρος της ακρίβειας κυμαίνεται στο (σχετικά μέτριο) 63%, παρ’ όλα αυτά υπάρχει περίπτωση (καρέ #180) όπου ο συνδυασμός μοντέλου εμφάνισης “ab-Y” και μετρικής σφάλματος το μέσο τετραγωνικό σφάλμα επιτυγχάνει το απόλυτο 100% ποσοστό επιτυχίας αντιστοιχίσεων στόχων-ανιχνεύσεων.

Τα αποτελέσματα της τρίτης σειράς δοκιμών, δηλαδή του μοντέλου “Levels” με την Ευκλείδεια μετρική, δεν είναι στο επιθυμητό επίπεδο καθώς επιτυγχάνεται χαμηλός μέσος όρος ακρίβειας μικρότερος του 50%. Όπως και στον προηγούμενο τύπο πειραμάτων έτσι και εδώ διαφαίνεται μία ανεπάρκεια της Ευκλείδειας μετρικής να μοντελοποιήσει επαρκώς τη διαφοροποίηση μεταξύ των διανυσματικών περιγραφών.

Η χρήση της μετρικής City-block βελτιώνει θεαματικά τα αποτελέσματα αντιστοιχίσης για τον ίδιο περιγραφέα μοντέλου εμφάνισης, τον

“Levels”. Το τελικό αποτέλεσμα παρουσιάζει μέσο όρο ακρίβειας αντιστοίχισης το 75%, ενώ υπάρχουν 2 παραδείγματα (Καρέ #84 και #180) όπου επιτυγχάνεται το απόλυτο 100% στην ακρίβεια αντιστοίχισης. Τελικώς, λαμβάνοντας υπόψη και τις δοκιμές με τον περιγραφέα “Voxels”, ο συνδυασμός “Levels”-City-block επιτυγχάνει τα καλύτερα αποτελέσματα για αυτόν τον τύπο πειράματος.

Τελευταία σειρά δοκιμών για το συγκεκριμένο τύπο πειράματος αποτελούν οι δοκιμές του διανυσματικού περιγραφέα μοντέλων εμφάνισης “Voxels”. Αρχικά, η πρώτη δοκιμή αφορούσε το συνδυασμό του με την Ευκλείδεια μετρική σφάλματος. Τα αποτελέσματα της δοκιμής παρουσίασαν έναν (καλό) μέσο όρο αντιστοίχισης της τάξης του 71%, ενώ για δύο παραδείγματα προέκυψε 100% ακρίβεια αντιστοίχισης. Τα παραπάνω αποτελέσματα είναι σαφώς καλύτερα εκείνων του συνδυασμού περιγραφέα-μετρική “Levels”-Ευκλείδεια απόσταση, ενώ βρίσκονται αρκετά κοντά (<5% κατά μέσο όρο) στα καλύτερα αποτελέσματα της παρούσας σειράς πειραμάτων.

Η τελευταία σειρά δοκιμών του πειράματος παρουσιάζει ακριβώς τα ίδια αποτελέσματα με την αμέσως προηγούμενη. Το γεγονός αυτό την κατατάσσει ως τη δεύτερη καλύτερη επίδοση στον συγκεκριμένο τύπο πειραμάτων. Η ομοιότητα στα αποτελέσματα μπορεί να εξηγηθεί από τη διαχωριστική ικανότητα (στον αντίστοιχο χώρο προτύπων) του δυαδικού διανυσματικού περιγραφέα “Voxels”. Η διαφορά στην μετρική σφάλματος δεν φαίνεται να επηρεάζει έντονα τα αποτελέσματα για τον συγκεκριμένο περιγραφέα, παρ’ όλα αυτά τα προηγούμενα πειράματα έχουν δείξει μια ελαφρά υπεροχή της μετρικής City-block έναντι της Ευκλείδειας.

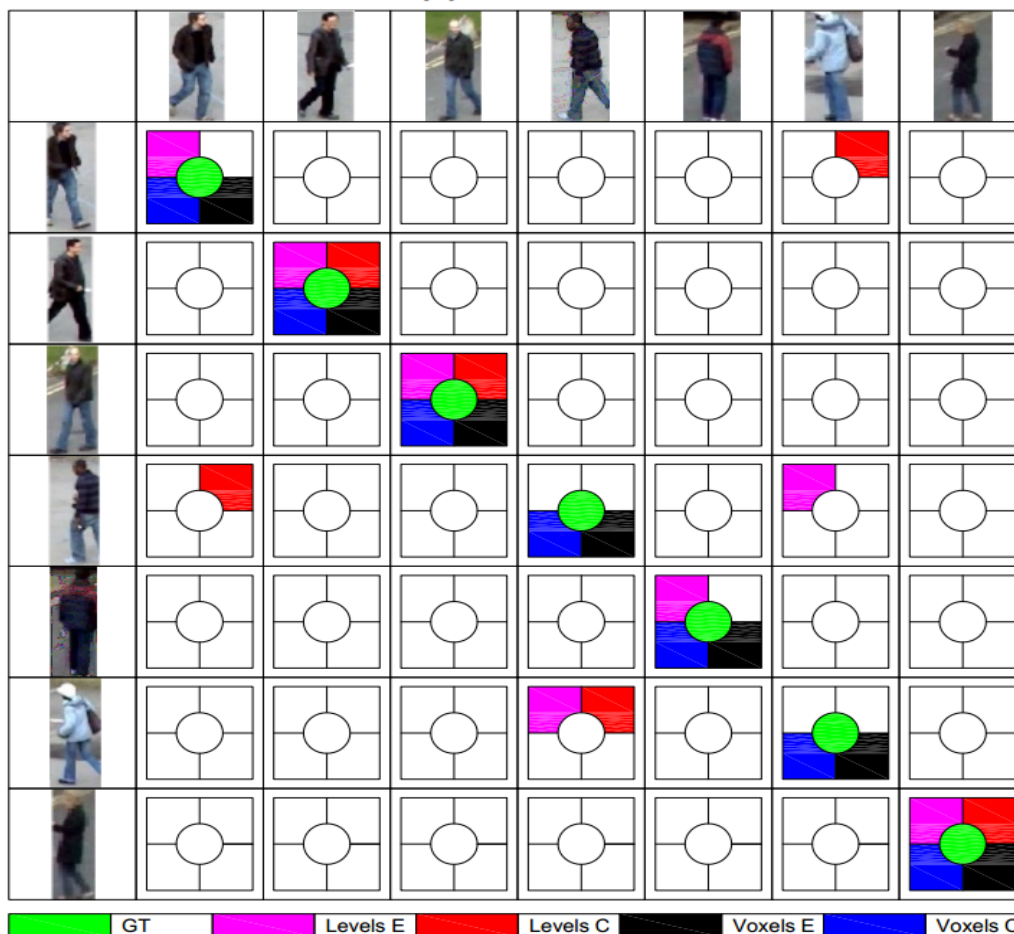
#### 4) Αντιστοίχιση μοντέλων εμφάνισης με τις ανιχνεύσεις ενός καρτέ

Στο συγκεκριμένο πείραμα και το επόμενο έγινε προσπάθεια αντιστοίχισης ενός συνόλου στόχων που έχουν προσδιορισθεί σε κάποιο καρτέ ενός βίντεο στο σύνολο των ανιχνεύσεων του επόμενου καρτέ. Η διαφορά έγκειται στο γεγονός ότι οι στόχοι πλέον προσδιορίζονται όχι από ένα καρτέ, αλλά από ένα πλήρες μοντέλο εμφάνισης που έχει υπολογισθεί από ένα ικανού μεγέθους (~20-30) σύνολο καρτέ ανίχνευσης του στόχου (τροχιά). Ως μοντέλα εμφάνισης επιλέχθηκαν τα μοντέλα "Levels" και "Voxels" καθώς είναι τα δύο που παράγουν διανυσματικούς περιγραφείς. Ο αλγόριθμος αντιστοίχισης παρέμεινε ο "Munkre's Hungarian Algorithm". Ο τρόπος παραγωγής του Γενικού Μοντέλου Εμφάνισης καθορίστηκε ως η κατά συνιστώσα μέση τιμή του συνόλου των περιγραφέων ή ως η κατά συνιστώσα ενδιάμεση τιμή του παραπάνω συνόλου. Τα αποτελέσματα του συγκεκριμένου πειράματος προέκυψαν ως εξής:

Average Appearance Model



### Median Appearance Model



Στον παρακάτω πίνακα αναφέρονται τα ποσοστά ακρίβειας αντιστοίχισης για το σύνολο των δοκιμών:

Πρότυπο-Μετρική		Μέσο Μοντέλο Εμφ.	Ενδιάμεσο Μοντέλο Εμφ.
Levels	Ευκλείδεια	71% (μη ανάθεση)	71%
Levels	City-block	71% (μη ανάθεση)	43%
Voxels	Ευκλείδεια	100%	100%
Voxels	City-block	100%	100%

Τα πειράματα και οι δοκιμές τέτοιου τύπου αποτελούν το πιο σημαντικό κριτήριο αξιολόγησης ενός αλγορίθμου μοντελοποίησης της εμφάνισης ανθρωπίνων στόχων. Αυτό συμβαίνει διότι μία πολύ σημαντική εφαρμογή των μοντέλων εμφάνισης είναι στην επίλυση/υποβοήθηση προβλημάτων οπτικής παρακολούθησης ενός (visual object tracking) ή πολλαπλών αντικειμένων (multiple object tracking).

Εν προκειμένω, η πρώτη δοκιμή, δηλαδή περιγραφείας “Levels” – Ευκλείδεια μετρική – Μέσο μοντέλο εμφάνισης) παρουσιάζει ποσοστό επιτυχίας αντιστοίχισης στόχων – ανίχνευσεων 71%. Το συγκεκριμένο ποσοστό αποτελεί ένα σχετικά χαμηλό ποσοστό ακρίβειας, καθιστώντας το συνδυασμό περιγραφείας-μετρικής-τύπου μοντέλου εμφάνισης ως ανεπαρκή. Να σημειωθεί ότι ένας στόχος δεν κατέστη δυνατό να «αντιστοιχηθεί» σε κάποια ανίχνευση, καθώς ο αλγόριθμος αντιστοίχισης θεώρησε ότι δεν υπήρχε ταιριασμα με αρκετά μικρό «κόστος».

Η αλλαγή της μετρικής σφάλματος για το συγκεκριμένο παράδειγμα δεν ωφέλησε τα αποτελέσματα αντιστοίχισης, τα οποία παρέμειναν ίδια όπως προηγουμένως με ακρίβεια 71% και μία μη-ανάθεση. Γενικά η συμπεριφορά του προτύπου “Levels” δεν αποδίδει αρκετά καλά στο συγκεκριμένο τύπο πειραμάτων.

Το πρότυπο “Voxels” παρουσιάζει καλύτερη εικόνα σε σχέση με το πρότυπο “Levels”, καθώς επιτυγχάνει ποσοστό επιτυχία αντιστοίχισης 100%. Το συγκεκριμένο ποσοστό εμφανίζεται για κάθε συνδυασμό μετρικής-τύπου μοντέλου εμφάνισης και καθιστά το πρότυπο αρκετά αξιόπιστο για τέτοιου είδους πειράματα.

Όσον αφορά το Ενδιάμεσο μοντέλο εμφάνισης (Median Appearance Model), η εικόνα των αποτελεσμάτων είναι παρόμοια όσον αφορά το περιγραφείας “Voxels”, ο οποίος συνεχίζει να επιτυγχάνει ακρίβεια αντιστοίχισης 100%. Τα αποτελέσματα χειροτερεύουν στην περίπτωση του προτύπου “Levels”, όπου με τη χρήση της Ευκλείδειας μετρικής επιτυγχάνεται ακρίβεια αντιστοίχισης 71%, ενώ με τη χρήση της μετρικής City-block προκύπτει μείωση της ακρίβειας αντιστοίχισης στο 43%. Ακόμη, στην περίπτωση του Ενδιάμεσου Μοντέλου Εμφάνισης δεν εμφανίζεται το φαινόμενο της μη-ανάθεσης στόχου σε ανίχνευση όπως στο Μέσο Μοντέλο Εμφάνισης.

## 5) Εφαρμογή μοντέλων εμφάνισης σε σύστημα οπτικής παρακολούθησης

Στο τελευταίο πείραμα εφαρμόστηκαν τα δύο μοντέλα εμφάνισης (“Levels” και “Voxels”) σε συνδυασμό με τις δύο μετρικές σφάλματος (Ευκλείδεια και City-block) ως υποσύστημα σε ένα πλήρες σύστημα οπτικής παρακολούθησης πολλαπλών στόχων (multiple object tracking). Ο αλγόριθμος παρακολούθησης είναι αυτός που περιγράφεται στην μεταπτυχιακή εργασία της Αθηνάς Ψάλτα, Ε.Μ.Π. [29]. Το βίντεο στο οποίο εφαρμόστηκε είναι το “PETS09-S2L1” από τον διαδικτυακό διαγωνισμό οπτικής παρακολούθησης πολλαπλών αντικειμένων 2DMOT2015 motchallenge.net [1], για το οποίο υπάρχει διαθέσιμη η δυαδική σημασιολογική κατάτμηση του κάθε καρέ σε περιοχές με ή χωρίς άνθρωπο, έτσι ώστε να είναι δυνατή η εφαρμογή των μοντέλων εμφάνισης.

Η αξιολόγηση των αποτελεσμάτων βασίζεται στις μετρικές σφάλματος MOTA (multiple object tracking accuracy) και MOTP (multiple object tracking precision)[2] με την πρώτη να αποτελεί την πιο σημαντική και αντιπροσωπευτική της συμπεριφοράς των σχετικών αλγορίθμων. Να σημειωθεί ότι στα πλαίσια του παρόντος πειράματος δεν έχει σημασία η απόλυτη τιμή των μετρικών αλλά οι σχετικές διαφορές τους καθώς στόχος είναι να αξιολογηθούν τα μοντέλα εμφάνισης και όχι ο αλγόριθμος παρακολούθησης. Τα αποτελέσματα για τους 4 συνδυασμούς μοντέλων εμφάνισης και μετρικών σφάλματος προέκυψαν ως εξής:

Μοντέλο Εμφάνισης \ Μετρική	MOTA	MOTP	MOTAL
<b>Levels - Euclidean</b>	60.4%	<b>71.3%</b>	61.6%
<b>Levels - City-Block</b>	59.6%	71.1%	60.7%
<b>Voxels - Euclidean</b>	57.3%	71.0%	58.4%
<b>Voxels - City-Block</b>	<b>61.9%</b>	71.2%	<b>62.6%</b>

Ως καλύτερο αποτέλεσμα προκύπτει αυτό του συνδυασμού του μοντέλου “Voxels” με την μετρική city-block με διαφορά 1,5% (απόλυτη) ή



2,5% (σχετική) συγκριτικά με τον δεύτερο καλύτερο συνδυασμό, δηλαδή “Levels” με Ευκλείδεια μετρική. Ως χειρότερο αποτέλεσμα προκύπτει ο συνδυασμός “Voxels” - Ευκλείδεια μετρική με σχετική διαφορά από τον πρώτο 8% (4,6% απόλυτη). Αν και οι διαφορές αυτές φαντάζουν μικρές στην πραγματικότητα είναι αρκετά σημαντικές καθώς ένα σύστημα παρακολούθησης αποτελείται από μία πληθώρα υποσυστημάτων όπως το μοντέλο κίνησης, ο αλγόριθμος αντιστοίχισης και το μοντέλο χωρικών αλληλεπιδράσεων, γεγονός που καθιστά μία μεγάλη βελτίωση σε ένα υπο-σύστημα να εμφανίζεται ως μικρότερη συνολικά για το τελικό σύστημα.

## ΚΕΦΑΛΑΙΟ 6. ΣΥΜΠΕΡΑΣΜΑΤΑ ΚΑΙ ΠΡΟΟΠΤΙΚΕΣ

Στο πλαίσιο της παρούσας εργασίας αναπτύχθηκαν τρία διαφορετικά μοντέλα εμφάνισης ανθρωπίνων στόχων. Εξαιτίας της φύσης αυτών των μοντέλων ως υποσυστήματα γενικότερων αλγορίθμων, όπως για παράδειγμα συστημάτων παρακολούθησης ενός ή πολλαπλών στόχων, δεν καθίσταται δυνατή η σύγκρισή τους με λοιπές προσεγγίσεις της βιβλιογραφίας αλλά μόνο μεταξύ τους. Υπό το πρίσμα αυτό, θα μπορούσε να ειπωθεί ότι τα πειράματα που διενεργήθηκαν συγκρίνουν με διαφορετικό τρόπο κάθε φορά τα μοντέλα μεταξύ τους και διερευνούν διάφορες δυνατότητες, όπως το «ταίριασμα» των μοντέλων με διάφορες μετρικές σφάλματος, τη διαχωριστική περιγραφική ικανότητα των μοντέλων και τη δυνατότητα διαχρονικής αφομοίωσης της πληροφορίας εμφάνισης για κάθε μοντέλο.

Το σύνολο των αποτελεσμάτων υποδεικνύει ότι δεν προκύπτει κάποιο μοντέλο ως κυρίαρχο στο σύνολο των δοκιμών παρά ότι ανάλογα με το πείραμα αναδεικνύονται προτερήματα και μειονεκτήματα του κάθε μοντέλου. Συνολικά διαφαίνεται ότι και τα τρία μοντέλα έχουν την ικανότητα να περιγράψουν ικανοποιητικά το μοντέλο εμφάνισης ενός ανθρώπου. Από την πρώτη τριάδα πειραμάτων σημαντικότερο προκύπτει εκείνο που εξετάζει τη δυνατότητα των μοντέλων να περιγράψουν έναν στόχο σε σειρά καρτέ, καθώς αυτό αποτελεί τη συνήθη εφαρμογή τέτοιου είδους μοντέλων. Για το πείραμα αυτό ως καλύτερη προσέγγιση προέκυψε το “Μοντέλο Voxels” ανεξάρτητα της επιλογής μετρικής ή τρόπου εκμάθησης.

Το συγκεκριμένο μοντέλο επέδειξε καλές επιδόσεις και στα υπόλοιπα δύο είδη πειραμάτων. Σημαντική είναι η επίπτωση των μετρικών σφάλματος με την μετρική city block να προκύπτει ως η καλύτερη επιλογή στις περισσότερες περιπτώσεις. Η επιλογή τρόπου εκμάθησης δεν φαίνεται να επηρεάζει τα αποτελέσματα, όμως η χρήση της διαμέσου έναντι του μέσου όρου σε πραγματικές συνθήκες πιθανώς να παρουσιάζει καλύτερα αποτελέσματα λόγω της συγκριτικής ευρωστίας της. Όσον αφορά το έτερο μοντέλο, δηλαδή το “Μοντέλο Levels”, παρατηρείται ότι αποδίδει λίγο χειρότερα αποτελέ-

σματα, όχι όμως σε μη αποδεκτά επίπεδα. Τέλος, το “Μοντέλο abY” απορρίπτεται κυρίως λόγω της μη διανυσματικής του μορφής που το καθιστά αρκετά δύσκολα διαχειρίσιμο και εφαρμόσιμο. Παρόλα αυτά, αποτελεί τη θεωρητική βάση των άλλων δύο μοντέλων και για το λόγο αυτό είναι εξαιρετικά χρήσιμη η ύπαρξή του.

Τέλος, σύμφωνα με την εφαρμογή των μοντέλων εμφάνισης ως υποσυστημάτων ενός ολοκληρωμένου αλγορίθμου παρακολούθησης, η υπεροχή των χαρακτηριστικών “Voxels” επιβεβαιώνεται. Επιβεβαιώνεται επίσης και η καλύτερη συμπεριφορά της μετρικής City-block σε πραγματικά δεδομένα με θόρυβο λόγω της σχετικής ευρωστίας της σε σχέση με την αντίστοιχη Ευκλείδεια. Το πρότυπο “Levels” φαίνεται να αποδίδει χειρότερα, όμως είναι λιγότερο ευαίσθητο στην επιλογή μετρικής. Συνολικά, η διαφορά στην τελική επίδοση του αλγορίθμου μεταξύ των δύο μοντέλων κυμάνθηκε στο 1,5%. Το ποσοστό αυτό φαντάζει μικρό, όμως στην πραγματικότητα είναι **αρκετά σημαντικό** αν αναλογιστεί κανείς το σχετικά μικρό τμήμα ενός αλγορίθμου παρακολούθησης που αφορούν τα μοντέλα εμφάνισης.

## 1) Συμπεράσματα

Η χρήση μοντέλων εμφάνισης έχει αποκτήσει τα τελευταία χρόνια υψηλή δυναμική σε διάφορους κλάδους της Όρασης Υπολογιστών με χαρακτηριστικότερο παράδειγμα αυτό της παρακολούθησης ενός ή πολλαπλών αντικειμένων. Τα μοντέλα εμφάνισης βασισμένα σε στοιχεία της εικόνας όπως οι κλίσεις και οι ακμές τυγχάνουν ευρείας εφαρμογής, σχεδόν από τη γένεση του κλάδου της Όρασης Υπολογιστών, σε πληθώρα εφαρμογών όπως οι σημειακοί τελεστές, χαρακτηριστικά τύπου Bag-of-Features και η ανίχνευση και εντοπισμός (localization) αντικειμένων σε εικόνες. Τα μοντέλα αυτά, όμως, ενώ έχουν υψηλή διακριτική ικανότητα μεταξύ ετεροειδών αντικειμένων (π.χ. άνθρωπος-δέντρο-αυτοκίνητο), αδυνατούν να περιγράψουν με αρκετά διαχωρισμό τρόπο τις διαφορές μεταξύ ομοειδών αντικειμένων (π.χ. άντρας με κόκκινο παντελόνι - άντρας με μπλε παντελόνι). Για τον λόγο αυτό πρέπει να αξιοποιηθεί πληροφορία όπως η φασματική και η χρωματική. Τα μοντέλα που αξιοποιούν τέτοιου είδους πληροφορία επαρκούν για την εξαγωγή ειδικής και όχι γενικής σημασιολογικής πληροφορίας. Χαρακτηριστικό παράδειγμα τέτοιου είδους πληροφορίας αποτελεί η πληροφορία ότι ένα τμήμα της εικόνας απεικονίζει έναν άνθρωπο (γενική σημασιολογική πληροφορία) και ότι σε ένα τμήμα της εικόνας που απεικονίζει άνθρωπο εμφανίζεται μία κόκκινη μπλούζα (ειδική πληροφορία).

Η παρούσα μεταπτυχιακή εργασία πραγματεύεται την κατηγορία μοντέλων εμφάνισης βασισμένων κατά κύριο λόγο στη χρωματική πληροφορία. Τα μοντέλα αυτά χαρακτηρίζονται για την υψηλότερου επιπέδου περιγραφή που επιτυγχάνουν, θεωρούμενης ως δεδομένης της *a-priori* κατάλληλης ταξινόμησης του εκάστοτε τμήματος της εικόνας στην θεματική κατηγορία για την οποία έχει αναπτυχθεί το, βασισμένο σε φασματικά χαρακτηριστικά, μοντέλο εμφάνισης. Η εξαγωγή σημασιολογικής πληροφορίας υψηλού επιπέδου και ανάλυσης<sup>9</sup> αποτελεί τον ένα από τους δύο στόχους της Όρασης Υπολογιστών (ο έτερος αφορά την εξαγωγή υψηλής ακρί-

---

<sup>9</sup> Η έννοια της ανάλυσης στη σημασιολογία αφορά την ύπαρξη ιεραρχίας στην πληροφορία. Για παράδειγμα ο άνθρωπος διακρίνεται σε άντρα-γυναίκα, αλλά και σε χρώμα/τύπος μπλούζας/παντελονιού, ψηλός/κοντός ή λεπτός/χοντρός. Το καθένα από τα προηγούμενα χαρακτηριστικά ανήκει σε κάποιο επίπεδο (χαμηλότερο/υψηλότερο), η σύνθεση και η πληρότητά τους όμως αναφέρεται στην ανάλυση.

βειας μετρητικής πληροφορίας). Συνεπώς, η απαίτηση σε μοντέλα/περιγραφείς οι οποίοι θα εξαγάουν εξειδικευμένη πληροφορία αναμένεται να αυξηθεί κατακόρυφα τα επόμενα χρόνια.

Ειδικότερα, τα μοντέλα εμφάνισης όπως αυτά που αναπτύχθηκαν στο πλαίσιο της παρούσας εργασίας αποτελούν αναπόσπαστο κομμάτι πληθώρας διεργασιών όπως η οπτική παρακολούθηση, η ανίχνευση αντικειμένων και η σημασιολογική κατάτμηση εικόνων ή/και 3D νεφών. Οι εφαρμογές στις οποίες αποτελούν βασικό μέρος τους ποικίλλουν από εφαρμογές ασφάλειας (security), αυτόνομης οδήγησης (autonomous driving), ταξινόμησης (classification), ομαδοποίησης (clustering), συνταύτισης (matching) κτλ.

Στο σύνολο των αναπτυσσόμενων μοντέλων, ισχύει η θεώρηση της γνώσης μίας γενικής σημασιολογικής κατάτμησης σε περιοχές με άνθρωπο ή όχι. Η θεώρηση αυτή είναι άμεσα συνυφασμένη με το προαναφερθέν περιεχόμενο της σημασιολογικής πληροφορίας και διάκρισής της σε γενική και ειδική. Η κρισιμότητα του συγκεκριμένου ζητήματος είναι αρκετά μεγάλη καθώς αν δεν διατίθεται αρκετά λεπτομερής κατάτμηση υπάρχει ο κίνδυνος μοντελοποίησης των χρωματικών χαρακτηριστικών και του υποβάθρου λόγω αδυναμίας διαχωρισμού της προέλευσης της πληροφορίας, δηλαδή όπως τονίστηκε παραπάνω η ειδική πληροφορία (χρωματικό μοντέλο) αφορά την γενική (άνθρωπος), η γνώση της οποίας είναι απαραίτητη. Τα τελευταία χρόνια έχει ενταθεί η έρευνα στο κομμάτι της σημασιολογικής κατάτμησης της εικόνας, κυρίως λόγω της ολοένα αυξανόμενης ανάγκης για ποιοτική και λεπτομερή σημασιολογική πληροφορία. Η σύγχρονη θεώρηση στον υπο-κλάδο αυτό προτάσσει τη χρήση Συνελκτικών Τεχνητών Νευρωνικών Δικτύων (CNN) και ταξινόμηση της εικόνας σε επίπεδο εικονοστοιχείου (pixel level classification), με χαρακτηριστικότερο παράδειγμα να αποτελεί το δίκτυο SegNet [3]. Για λόγους πληρότητας, κρίνεται άξιο αναφοράς το γεγονός ότι πλέον υπάρχουν αρκετές υλοποιήσεις (ανοιχτού κώδικα ή όχι) τέτοιων αρχιτεκτονικών με αποτέλεσμα να έχει επιταχυνθεί η έρευνα σε αυτό το κομμάτι.

## 2) Μελλοντικές προεκτάσεις

Αυτή η εργασία εντάσσεται στον ευρύτερο ερευνητικό κλάδο περί εξαγωγής ιεραρχημένης και υψηλού επιπέδου σημασιολογικής πληροφορίας. Η πρόοδος στον τομέα των μοντέλων εμφάνισης θα καθορίσει και θα καθοριστεί από την πρόοδο όμορων επιστημονικών υπο-πεδίων, όπως εκείνο της σημασιολογικής κατάτμησης και της ταξινόμησης. Η χρήση αλγορίθμων της Μηχανικής Μάθησης (Machine Learning) και της Βαθιάς Μάθησης (Deep Learning) βρίσκεται στην αιχμή της έρευνας σε αυτούς τους τομείς καθώς επιτρέπουν την ανάπτυξη αλγορίθμων εξαγωγής ολοένα και υψηλότερου επιπέδου σημασιολογικής πληροφορίας (semantics).

Όπως έχει ήδη αναφερθεί, οι επιδόσεις των μοντέλων εμφάνισης βελτιώνονται δραστικά με την ύπαρξη ορθής και ακριβούς σημασιολογικής κατάτμησης. Οι σημερινοί αλγόριθμοι σημασιολογικής κατάτμησης βασίζονται αποκλειστικά σε μεθοδολογίες ταξινόμησης εικονοστοιχείων εκπαιδευμένων με τη χρήση από-άκρη-σε-άκρη μάθησης (end-to-end learning) και σε τεράστια σετ δεδομένων αληθείας. Θα ήταν, πιθανώς, ωφέλιμο σε τέτοιες περιπτώσεις η μάθηση να υποβοηθείται από αλγορίθμους υπερ-εικονοστοιχείων και από μοντέλα εμφάνισης σε συνδυασμό με κατάλληλη πρότερη αναγνώριση. Ιδεατά, θα μπορούσαν τα μοντέλα εμφάνισης να ενσωματωθούν στον κλάδο της σημασιολογικής κατάτμησης. Ένας τρόπος που θα μπορούσε πιθανόν να συμβεί κάτι τέτοιο θα ήταν η έξοδος ενός συστήματος σημασιολογικής κατάτμησης να μην ήταν ένας απλός «θεματικός χάρτης» αλλά μία πολυεπίπεδη δομή χάρτη η οποία θα συμπεριλάμβανε σε κάθε εικονοστοιχείο (ή ομάδα τέτοιων/υπερ-εικονοστοιχείο) μία ιεραρχημένη (τοπολογικά-ορθή) σημασιολογική πληροφορία υψηλής ανάλυσης. Μία τέτοια προσπάθεια θα απαιτούσε, βέβαια, και αντίστοιχη εργασία για την παραγωγή πολύ-επίπεδων δεδομένων αληθείας και μία αντίστοιχα πολύ-επίπεδη και πολύπλοκη αρχιτεκτονική ΤΝΔ.

Συμπερασματικά, τα μοντέλα εμφάνισης αποτελούν σημαντικό υποσύστημα διάφορων εφαρμογών της όρασης υπολογιστών. Η συνεχώς αυξανόμενη ερευνητική τάση προς εξαγωγή λεπτομερούς σημασιολογικής πληροφορίας καθιστά τα μοντέλα εμφάνισης αρκετά επίκαιρα καθώς πρόκειται εξ ορισμού για δομές σημασιολογικής περιγραφής. Παρ' όλα αυτά είναι πλέον εμφανής η ανάγκη για σημασιολογική πληροφορία όχι πεπλεγμένη

σε ένα διάγραμμα περιγραφής, αλλά για μια περιγραφή σημασιολογικών ιδιοτήτων για το κάθε αντικείμενο. Η παραπάνω κατεύθυνση πιθανώς θα αποτελέσει την εξέλιξη των μοντέλων εμφάνισης στο μέλλον καθώς οι απαιτήσεις σε εξειδικευμένη σημασιολογική ή μη πληροφορία αναμένεται να ενταθούν.

## ΒΙΒΛΙΟΓΡΑΦΙΑ

- [1] Leal-Taixé, Laura, et al. "**Motchallenge 2015: Towards a benchmark for multi-target tracking.**" *arXiv preprint arXiv:1504.01942* (2015).
- [2] Bernardin, Keni, and Rainer Stiefelhagen. "**Evaluating multiple object tracking performance: the CLEAR MOT metrics.**" *EURASIP Journal on Image and Video Processing* 2008.1 (2008): 246309.
- [3] Badrinarayanan, Vijay, Alex Kendall, and Roberto Cipolla. "**Segnet: A deep convolutional encoder-decoder architecture for image segmentation.**" *arXiv preprint arXiv:1511.00561* (2015).
- [4] Besl, Paul J., and Neil D. McKay. "**A method for registration of 3-D shapes.**" *IEEE Transactions on pattern analysis and machine intelligence* 14.2 (1992): 239-256.
- [5] Rusinkiewicz, Szymon, and Marc Levoy. "**Efficient variants of the ICP algorithm.**" *3-D Digital Imaging and Modeling, 2001. Proceedings. Third International Conference on. IEEE, 2001.*
- [6] Dollár, Piotr, et al. "**Fast feature pyramids for object detection.**" *IEEE Transactions on Pattern Analysis and Machine Intelligence* 36.8 (2014): 1532-1545.
- [7] Achanta, Radhakrishna, et al. "**SLIC superpixels compared to state-of-the-art superpixel methods.**" *IEEE transactions on pattern analysis and machine intelligence* 34.11 (2012): 2274-2282.
- [8] Carter, Matthew P. "**Computer graphics: principles and practice.**" (1997): 65-66.
- [9] Dalal, Navneet, and Bill Triggs. "**Histograms of oriented gradients for human detection.**" *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on. Vol. 1. IEEE, 2005.*
- [10] Lowe, David G. "**Object recognition from local scale-invariant features.**" *Computer vision, 1999. The proceedings of the seventh IEEE international conference on. Vol. 2. Ieee, 1999.*
- [11] Rosten, Edward, and Tom Drummond. "**Fusing points and lines for high performance tracking.**" *Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on. Vol. 2. IEEE, 2005.*
- [12] Harris, Chris, and Mike Stephens. "**A combined corner and edge detector.**" *Alvey vision conference. Vol. 15. No. 50. 1988.*



- [13] Bay, Herbert, Tinne Tuytelaars, and Luc Van Gool. "**Surf: Speeded up robust features.**" Computer vision–ECCV 2006 (2006): 404-417.
- [14] Psalta, Athena. "**Development of segmentation algorithms for hyperspectral data. Clustering and Classification.**" Diploma thesis. NTUA, Greece. 2015.
- [15] Dollár, Piotr, et al. "**Fast feature pyramids for object detection.**" IEEE Transactions on Pattern Analysis and Machine Intelligence 36.8 (2014): 1532-1545.
- [16] Girshick, Ross. "**Fast r-cnn.**" Proceedings of the IEEE international conference on computer vision. 2015.
- [17] Ren, Shaoqing, et al. "**Faster R-CNN: Towards real-time object detection with region proposal networks.**" Advances in neural information processing systems. 2015.
- [18] Rodrigues, João, and JM Hans du Buf. "**Multi-scale keypoints in V1 and beyond: object segregation, scale selection, saliency maps and face detection.**" BioSystems 86.1 (2006): 75-90.
- [19] Mita, Takeshi, Toshimitsu Kaneko, and Osamu Hori. "**Joint haar-like features for face detection.**" Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on. Vol. 2. IEEE, 2005.
- [20] Herpers, Rainer, et al. "**Edge and keypoint detection in facial regions.**" Automatic Face and Gesture Recognition, 1996., Proceedings of the Second International Conference on. IEEE, 1996.
- [21] Morris, Tim, Paul Blenkhorn, and Farhan Zaidi. "**Blink detection for real-time eye tracking.**" Journal of Network and Computer Applications 25.2 (2002): 129-143.
- [22] Reynolds, Douglas. "**Gaussian mixture models.**" Encyclopedia of biometrics (2015): 827-832.
- [23] Yang, Ming-Hsuan, and Narendra Ahuja. "**Gaussian mixture model for human skin color and its applications in image and video databases.**" Storage and Retrieval for Image and Video Databases VII. Vol. 3656. International Society for Optics and Photonics, 1998.
- [24] Zivkovic, Zoran. "**Improved adaptive Gaussian mixture model for background subtraction.**" Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on. Vol. 2. IEEE, 2004.
- [25] Jain, Anil K., Yi Chen, and Meltem Demirkus. "**Pores and ridges: High-resolution fingerprint matching using level 3 features.**" IEEE Transactions on Pattern Analysis and Machine Intelligence 29.1 (2007): 15-27.

- [26] Kalal, Zdenek, Krystian Mikolajczyk, and Jiri Matas. "**Tracking-learning-detection.**" IEEE transactions on pattern analysis and machine intelligence 34.7 (2012): 1409-1422.
- [27] Xiang, Yu, Alexandre Alahi, and Silvio Savarese. "**Learning to track: Online multi-object tracking by decision making.**" Proceedings of the IEEE International Conference on Computer Vision. 2015.
- [28] Sadeghian, Amir, Alexandre Alahi, and Silvio Savarese. "**Tracking the untrackable: Learning to track multiple cues with long-term dependencies.**" arXiv preprint arXiv:1701.01909 (2017).
- [29] Psalta, Athena. "**Ανάπτυξη και Αξιολόγηση Μεθοδολογίας για την Παρακολούθηση Πολλαπλών Αντικειμένων σε Ακολουθίες Εικόνων.**" Master thesis. NTUA, Greece. 2017.
- [30] Munkres, James. "**Algorithms for the assignment and transportation problems.**" Journal of the society for industrial and applied mathematics 5.1 (1957): 32-38.
- [31] Logvinenko, Alexander D. "**The geometric structure of color.**" Journal of vision 15.1 (2015): 16-16.
- [32] Schanda, János, ed. "**Colorimetry: understanding the CIE system.**" John Wiley & Sons, 2007.