ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΜΗΧΑΝΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ
Δ.Π.Μ.Σ. ΣΥΣΤΗΜΑΤΑ ΑΥΤΟΜΑΤΙΣΜΟΥ
ΤΟΜΕΑΣ ΤΕΧΝΟΛΟΓΙΑΣ ΤΩΝ ΚΑΤΕΡΓΑΣΙΩΝ

ΕΝΤΟΠΙΣΜΟΣ ΘΕΣΗΣ ΚΥΛΙΝΔΡΙΚΩΝ ΤΕΜΑΧΙΩΝ ΣΕ ΟΡΙΖΟΝΤΙΑ ΕΠΙΦΑΝΕΙΑ ΜΕΣΩ ΨΗΦΙΑΚΗΣ ΕΠΕΞΕΡΓΑΣΙΑΣ ΕΙΚΟΝΑΣ

ΜΕΤΑΠΤΥΧΙΑΚΗ ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Νικόλαος Π. Πεισμάνης

Επίβλεψη : Βοσνιάκος Γεώργιος – Χριστόφορος
Καθηγητής Ε.Μ.Π.

Αθήνα, Δεκέμβριος 2015

ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΜΗΧΑΝΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ
Δ.Π.Μ.Σ. ΣΥΣΤΗΜΑΤΑ ΑΥΤΟΜΑΤΙΣΜΟΥ
ΤΟΜΕΑΣ ΤΕΧΝΟΛΟΓΙΑΣ ΤΩΝ ΚΑΤΕΡΓΑΣΙΩΝ

ΕΝΤΟΠΙΣΜΟΣ ΘΕΣΗΣ ΚΥΛΙΝΔΡΙΚΩΝ ΤΕΜΑΧΙΩΝ ΣΕ ΟΡΙΖΟΝΤΙΑ ΕΠΙΦΑΝΕΙΑ ΜΕΣΩ ΨΗΦΙΑΚΗΣ ΕΠΕΞΕΡΓΑΣΙΑΣ ΕΙΚΟΝΑΣ

ΜΕΤΑΠΤΥΧΙΑΚΗ ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Νικόλαος Π. Πεισμάνης

Επίβλεψη : Βοσνιάκος Γεώργιος – Χριστόφορος
Καθηγητής Ε.Μ.Π.

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 21/12/2015.

.....	.....	.....
Βοσνιάκος Γ.Χ.	Αντωνιάδης Ι.	Μανωλάκος Δ.
Καθηγητής Ε.Μ.Π.	Καθηγητής Ε.Μ.Π.	Καθηγητής Ε.Μ.Π.

Αθήνα, Δεκέμβριος 2015

ΠΕΡΙΛΗΨΗ

Σκοπός της μεταπτυχιακής εργασίας είναι ο εντοπισμός μέσω επεξεργασίας ψηφιακών φωτογραφιών της θέσης κυλινδικών μεταλλικών αντικειμένων πάνω σε επίπεδη επιφάνεια. Ως εντοπισμός θεωρείται η εύρεση των διαστάσεων των αντικειμένων, οι συντεταγμένες του κέντρου τους καθώς και η γωνία περιστροφής τους. Για την επίτευξη του σκοπού αυτού χρησιμοποιούνται μοντέλα τεχνητής όρασης, μηχανικής μάθησης και ψηφιακής επεξεργασίας εικόνας. Η υλοποίηση γίνεται με ανάπτυξη λογισμικού στη γλώσσα προγραμματισμού C++. Το λογισμικό στηρίζεται σε συναρτήσεις της βιβλιοθήκης OpenCV. Η μέθοδος επίλυσης της παρούσας διπλωματικής εργασίας στηρίζεται σε εργαλεία της μηχανικής μάθησης για αναγνώριση και αφαίρεση του παρασκηνίου από τα αντικείμενα του προσκηνίου καθώς και σε ανά pixel μετασχηματισμούς και σε μεθόδους περιγραφής περιγράμματος σχημάτων. Τέλος παρουσιάζονται αποτελέσματα της μεθόδου και ο σχολιασμός τους.

ΛΕΞΕΙΣ ΚΛΕΙΔΙΑ

Εντοπισμός θέσης αντικειμένων, Επεξεργασία εικόνας, Μηχανική μάθηση, Τεχνητή όραση OpenCV, Περιγραφή περιγράμματος σχήματος, Αφαίρεση φόντο

ABSTRACT

The scope of this postgraduate thesis is the localization of cylindrical workpieces on a horizontal flat surface via digital surface image processing. As localization is considered the calculation of objects' dimensions, the coordinates of their center and their angle of rotation. In order to achieve the objective of this thesis there have been used models and techniques of computer vision, machine learning and image processing. The implementation is done through software development in programming language C++. The software is based on functions of OpenCV library. The approach of the current thesis is based on a machine learning technique for identification and subtraction of background scene from elements on the foreground scene as well as on per pixel tranformations and contour descriptor method. Finally the results and the evaluation of the used method are presented.

KEY WORDS

Localization of objects, Image processing, Machine learning, OpenCV, Computer vison, Contour descriptor, Background subtraction

ΕΥΧΑΡΙΣΤΙΕΣ

Η μεταπτυχιακή αυτή διπλωματική εργασία εκπονήθηκε κατά το ακαδημαϊκό έτος 2014-2015 υπό την επίβλεψη του κ. Γ.Χ. Βοσνιάκου, καθηγητή του Ε.Μ.Π. της σχολής Μηχανολόγων Μηχανικών. Θα ήθελα να ευχαριστήσω θερμά τον κ. Βοσνιάκο για την αμέριστη βοήθεια, συμπαράσταση και κατανόησή του. Ακόμα θα ήθελα να ευχαριστήσω θερμά την οικογένειά μου και τα κοντινά μου πρόσωπα για την βοήθειά τους τόσο σε επίπεδο ψυχολογικής συμπαράστασης όσο και σε πρακτικό επίπεδο κατά τη διάρκεια εκπόνησης της παρούσας εργασίας .

ΠΕΡΙΕΧΟΜΕΝΑ

ΚΕΦΑΛΑΙΟ 1: ΕΙΣΑΓΩΓΗ

1.1 Διατύπωση του προβλήματος 1

1.2 Λοιπές τεχνικές και προηγούμενη σχετική έρευνα 2

1.3 Δομή της εργασίας 3

1.4 Βιβλιογραφία 3

ΚΕΦΑΛΑΙΟ 2: ΤΕΧΝΗΤΗΣ ΟΡΑΣΗ ΚΑΙ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

2.1 Εισαγωγή 5

2.2 Πιθανότητες 5

2.2.1 Τυχαίες μεταβλητές 5

2.2.2 Δεσμευμένη πιθανότητα - Κανόνας Bayes 6

2.3 Κοινές κατανομές πιθανοτήτων 8

2.3.1 Κατανομή Bernoulli και Beta 8

2.3.2 Μονοπαραγοντική κανονική κατανομή – Κανονικά κλιμακώμενη αντίστροφη κατανομή gamma 9

2.3.3 Πολυπαραγοντική Κανονική κατανομή- Κανονικά αντίστροφη κατανομή Wishart 10

2.3.4 Συζυγία 12

2.4 Προσαρμογή πιθανοτικών κατανομών 13

2.4.1 Μέγιστο ενδεχόμενο 13

2.4.2 Μεταγενέστερο μέγιστο 13

2.4.2 Bayesian-ή προσέγγιση 14

2.5 Μηχανική Μάθηση (Machine Learning) 15

2.6 Αφαίρεση Παρασκηνίου 17

2.6.1 Μοντελοποίηση παρασκηνίου με κανονική κατανομή 17

2.6.2 Μοντελοποίηση σύνθετων κατανομών δεδομένων – Συνδιασμός Gaussian-ών 18

2.7 Ψηφιακή επεξεργασία εικόνας 26

2.7.1 Γραμμικό φιλτράρισμα στοιχείων εικόνας 27

2.7.2 Εντοπισμός άκρων – Μέθοδος Canny 28

2.8 Πρότυπο ενεργού περιγράμματος 30

2.9 Βιβλιογραφία 33

ΚΕΦΑΛΑΙΟ 3: ΒΙΒΛΙΟΘΗΚΕΣ OpenCV

3.1 Εισαγωγή 35

3.2 Ενότητα core 35

3.2.1 Βασικές δομές 36
3.2.2 Συναρτήσεις σχεδιασμού 38
3.3 Ενότητα highgui 40
3.3.1 Μέθοδοι διεπαφής χρήστη 40
3.3.2 Εγγραφή και ανάγνωση εικόνων και βίντεο 41
3.4 Ενότητα imgproc 42
3.4.1 Εφαρμογές φιλτραρίσματος εικόνας 42
3.4.2 Ανίχνευση χαρακτηιστικών σημείων – Ανιχνευτής άκρων Canny 44
3.4.3 Δομική ανάλυση και περιγραφή σχήματος 45
3.5 Ενότητα video – Διαχωρισμός παρασκηνίου 48
3.6 Βιβλιογραφία 49

ΚΕΦΑΛΑΙΟ 4: ΜΟΝΤΕΛΟ ΓΙΑ ΕΝΤΟΠΙΣΜΟ ΘΕΣΗΣ ΚΥΛΙΝΔΡΙΚΩΝ ΤΕΜΑΧΙΩΝ

4.1 Εξέταση του προβλήματος - Συμβιβασμοί 51
4.2 Περιγραφή και υλοποίηση μοντέλου 51
4.2.1 Αρχιτεκτονική της επίλυσης 52
4.3 Πειραματικές διάτάξεις - Αποτελέσματα 54
4.3.1 Πρώτη πειραματική διάταξη - Αποτελέσματα 55
4.3.2 Δεύτερη πειραματική διάταξη – Αποτελέσματα 63
4.3.3 Σχολιασμός αποτελεσμάτων – Σφάλματα 68
4.4 Βιβλιογραφία 70

ΚΕΦΑΛΑΙΟ 5: ΣΥΜΠΕΡΑΣΜΑΤΑ

6.1 Σύνοψη της εργασίας και των αποτελεσμάτων 71
6.2 Συμπληρώσεις 72

ΠΑΡΑΡΤΗΜΑ Α ΠΗΓΑΙΟΣ ΚΩΔΙΚΑΣ ΕΦΑΡΜΟΓΗΣ 73

ΠΑΡΑΡΤΗΜΑ Β ΑΠΟΤΕΛΕΣΜΑΤΑ ΔΙΑΔΙΚΑΣΙΑΣ ΚΑΛΙΜΠΡΑΡΙΣΜΑΤΟΣ ΚΑΜΕΡΑΣ 77

ΚΕΦΑΛΑΙΟ 1

ΕΙΣΑΓΩΓΗ

1.1 ΔΙΑΤΥΠΩΣΗ ΤΟΥ ΠΡΟΒΛΗΜΑΤΟΣ

Η ανάπτυξη της τεχνολογίας τα τελευταία χρόνια έφερε επανάσταση στο χώρο του αυτομάτου ελέγχου. Η δυνατότητα που προσφέρει η επιστήμη της πληροφορικής για ταχύτατη επεξεργασία μεγάλου όγκου δεδομένων από σύνθετους αλγορίθμους σε συνδιασμό με τα εξελιγμένα ρομποτικά συστήματα άνοιξε το έδαφος για εισαγωγή νέων τεχνικών αυτοματοποίησης στις διαδικασίες κατασκευής και επεξεργασίας σε βιομηχανικό επίπεδο.

Η μεγάλη αυτή πρόοδος απαντάται και στον τομέα της τεχνητής όρασης σε επίπεδο διαχείρησης των μεγάλων όγκων πληροφοριών που ενπεριέχονται στις εικόνες, σε επίπεδο ενσωμάτωσης αλγορίθμων από το χώρο της μηχανικής μάθησης που δίνουν τη σημαντικότατη δυνατότητα εκμάθησης και αναγνώρισης προτύπων και μοτίβων στα περιεχόμενα της εικόνας και τέλος σε επίπεδο εύρεσης έτοιμων προγραμματιστικών μεθόδων τεχνητής όραση με μεγάλη γκάμα και αποτελεσματικότητα. Εκμεταλευόμενοι όλη αυτή την πρόοδο θα γίνει προσπάθεια στην παρούσα διπλωματική εργασία να υλοποιηθεί μοντέλο μέσω λογισμικού για τον οπτικό εντοπισμό θέσης κυλινδρικών μεταλλικών αντικειμένων πάνω σε επίπεδη επιφάνεια μέσω μέσω ψηφιακών φωτογραφιών που θα λαμβάνονται από τα συγκεκριμένα μεταλλικά αντικείμενα. Η διαδικασία έυρεσης της θέσης των αντικειμένων ανήκει σε ένα ευρύτερο πλάνο αυτόματης περισυλλογής των μεταλλικών αντικειμένων από ρομποτικό μηχανισμό στηριζόμενος στα εξαγώμενα δεδομένα του λογισμικού. Σκοπός είναι τα δεδομένα θέσης να είναι ακριβή για την αποφυγή αστοχιών κατά τη συλλογή των αντικειμένων.

Το μέγεθος – χαρακτηριστικό που επιλέχθηκε ως κατάλληλο για την παροχή πληροφοριών για τη θέση των αντικειμένων είναι το περίγραμμά τους κάθως είναι εύκολο στον προσδιορισμό του και μόλις εντοπιστεί μετά με απλές τεχνικές μπορεί να προσδιοριστεί το κέντρο του και η διεύθυνσή του. Σημαντικό γνώρισμα είναι ότι οι τεχνικές εύρεσης περιγραμμάτων που προσφέρει η επιστήμη της τεχνητής όρασης είναι απλές στην εφαρμογή τους, με γρήγορη απόκριση, με πλήθος σχετιζόμενων συναρτήσεων για περαιτέρω επεξεργασία τους. Άλλες τεχνικές της τεχνητής όρασης που στηρίζονται στην εύρεση χαρακτηριστικών σημείων και δημιουργίας προτύπου ήταν δύσκολο και πολύπλοκο να χρησιμοποιηθούν καθώς εκ κατασκευής τα συγκεκριμένα μεταλλικά αντικείμενα δεν έχουν καποια χαρακτηριστικά γνωρίσματα εκτός από το χρώμα τους, που αυτό είναι παρόμοιο σε όλα και επίσης μεταβάλλεται ανάλογα με το φωτισμό και το σχήμα τους το οποίο είναι κοινό σε όλα, κυλινδρικό και παρόμοιων διαστάσεων.

Ένα σημαντικό πρόβλημα στον προσδιορισμό του περιγράμματος είναι ότι το αντικείμενο πρέπει να είναι ξεκάθαρο από το παρασκήνιο καθώς οι μέθοδοι προσδιορισμού περιγράμματος δεν είναι “ευφυή”. Οπότε στόχος του μοντέλου που θα αναπτυχθεί είναι πρώτον η απομόνωση των αντικειμένων από το παρασκήνιο, δεύτερον ο προσδιορισμός των

περιγραμμάτων τους και τρίτον η περαιτέρω επεξεργασία του περιγράμματος για υπολογισμό της θέσης του κέντρου του και της διεύθυνσής του.

1.2 ΛΟΙΠΕΣ ΤΕΧΝΙΚΕΣ ΚΑΙ ΠΡΟΗΓΟΥΜΕΝΗ ΣΧΕΤΙΚΗ ΕΡΕΥΝΑ

Η ανίχνευση αντικειμένων σε μια εικόνα είναι μια υπολογιστική που έχει στόχο να εξάγει πληροφορία σχετικά με τον τύπο των αντικειμένων που υπάρχουν καθώς και τη θέση και τον προσανατολισμό τους (στο επίπεδο ή στον τρισδιάστατο χώρο). Μια εικόνα αποτελείται από pixels (εικονοστοιχεία). Καθένα έχει μια τιμή που αντιστοιχεί στη φωτεινότητα ή στο χρώμα του. Συνεπώς, μπορούν να θεωρηθεί ως διδιάστατο σήμα και άρα να χρησιμοποιηθούν αντίστοιχα εργαλεία για την αναπαράσταση και την επεξεργασία της.

Όπως έχει αναλυθεί σε πολλά σχετικά βιβλία και μονογραφίες στο χώρο της υπολογιστικής ή μηχανικής όρασης [1.1], οι πιο γνωστές σχετικές τεχνικές περιλαμβάνουν το μετασχηματισμό Hough [1.2], το μετασχηματισμό χαρακτηριστικών ανεξάρτητων κλίμακας - SIFT (Scale-Invariant Feature Transform) [1.3], τα ιστογράμματα κατευθύνσεων κλίσης - HOG (Histogrames of Oriented Gradients) [1.4] ιστογράμματα προσανατολισμού ακμών (Edge Orientation Histograms – EOH) [1.5], ανίχνευση περιγράμματος [1.6], μετασχηματισμό κυματιδίων με συναρτήσεις βάσης τύπου Haar [1.7] και ανάλυση ανεξάρτητων συνιστωσών (Independent Component Analysis – ICA). Με αυτούς τους τρόπους περιγράφονται ουσιαστικά πρότυπα προς αναγνώριση στην υπό εξέταση εικόνα, αλλά και αυτή η ίδια.

Για τον έλεγχο αν ένα τμήμα της εικόνας ταυτίζεται με ένα συγκεκριμένο πρότυπο (διαδικασία ταξινόμησης), μπορούν να χρησιμοποιηθούν νευρωνικά δίκτυα, μηχανές διανυσμάτων στήριξης SVM (Support Vector Machines) [1.8], αλγόριθμοι ενδυνάμωσης (boosting), [1.9] κλπ.

Παραδείγματα χρήσης των παραπάνω, καθώς και περαιτέρω εξέλιξης τους, υπάρχουν πάρα πολλά στη βιβλιογραφία των τελευταίων 15 ετών. Για παράδειγμα, ανίχνευση βασίστηκε σε τοπικά χαρακτηριστικά της εικόνας, που εντοπίζονται με χρήση ενός ανιχνευτή περιοχών ενδιαφέροντος και περιγράφονται με κατάλληλο λεξιλόγιο [1.10]. Συνδυασμός τοπικών και γενικών πληροφοριών από εικόνες βάσει δύο ανιχνευτών, έναν τοπικού χαρακτήρα [1.10] και έναν βασιζόμενο σε ανάλυση ανεξάρτητων συνιστωσών έδωσε καλύτερα αποτελέσματα [1.11]. Επίσης, έχει γίνει ανίχνευση περιοχών ενδιαφέροντος σε συνδυασμό με ιστογράμματα έντασης, για παράσταση τμημάτων της εικόνας με βάση τη θέση και την κλίμακά τους [1.12]. Η ανίχνευση συνδυάστηκε, επίσης, με την τμηματοποίηση [1.13] έτσι ώστε ο ανιχνευτής να εκπαιδεύεται σχετικά με τις δομές που μπορεί να εμφανιστούν σε κάθε αντικείμενο και να είναι δυνατή η αναγνώριση με παρουσία επικαλύψεων [1.14]. Ακόμη, το αφηρημένο μοντέλο σχήματος [1.13] συνδυάστηκε με έναν ταξινομητή SVM για μείωση των λανθασμένων θετικών ταξινομήσεων του [1.15]. Παρόμοια, ο μετασχηματισμός Gabor χρησιμοποιήθηκε σε κάθε δυνατή θέση και κλίμακα για αναπαράσταση από τις συναρτήσεις βάσης και στη συνέχεια ένας ταξινομητής SVM έδωσε επιτυχή αποτελέσματα [1.16]. Τέλος, χαρακτηριστικά τύπου Haar χρησιμοποιήθηκαν σε γενικότερο πλαίσιο κυρίως για αναγνώριση ανθρώπων [1.17], αλλά αργότερα η μέθοδος επεκτάθηκε και για γενικότερα αντικείμενα και μάλιστα για αναγνώριση τους σε πραγματικό χρόνο [1.18].

1.3 ΔΟΜΗ ΤΗΣ ΕΡΓΑΣΙΑΣ

Στο **Κεφάλαιο 2** παρουσιάζεται το θεωρητικό υπόβαθρο των μεθόδων της τεχνητής όρασης, της μηχανικής μάθησης και της ψηφιακής επεξεργασίας είκονας που εφαρμόστηκαν στο μοντέλο εύρεσης θέσης μεταλλικών αντικειμένων .

Στο **Κεφάλαιο 3** παρουσιάζονται οι δομές και οι συναρτήσεις της βιβλιοθήκης OpenCV που χρησιμοποιήθηκαν για την υλοποίηση του μοντέλου.

Στο **Κεφάλαιο 4** περιγράφεται η λογική και η δομή του λογισμικού που υλοποιεί το μοντέλο εύρεσης της θέσης των κυλινρικών μεταλλικών αντικειμένων. Περιγράφονται οι πειραματικές διατάξεις στις οποίες δοκιμάστηκε το λογισμικό. Παρουσιάζονται και σχολιάζονται τα αποτελέσματά τους.

Στο **Κεφάλαιο 5** γίνεται μια σύνοψη της εργασίας και παρουσιάζονται τα συμπεράσματα και οι απαραίτητες επεκτάσεις του μοντέλου.

Στο **Παράρτημα Α** βρίσκεται ο πηγαίος κώδικας του λογισμικού.

Στο **Παράρτημα Β** βρίσκονται τα αποτελέσματα της διαδικασίας καλιμπραρίσματος της κάμερας που χρησιμοποιήθηκε.

1.4 ΒΙΒΛΙΟΓΡΑΦΙΑ

[1.1] Χ. Βαρυτιμίδης, Ανίχνευση αντικειμένων και ημιαυτόματος χαρακτηρισμός εικόνων, Διπλωματική Εργασία, Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών ΕΜΠ, Ιούνιος 2008.

[1.2] RO Duda, PE Hart, Use of the Hough transformation to detect lines and curves in pictures, Communications of the ACM, 15.1 (1972): 11-15.

[1.3] DG Lowe, Distinctive Image Features from Scale-Invariant Keypoints, International Journal of Computer Vision, 60.2 (2004): 91-110.

[1.4] N Dalal, B 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.

[1.5] K Levi, Y Weiss, Learning object detection from a small number of examples: The importance of good features, Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on. Vol. 2. IEEE, 2004.

[1.6] J Shotton, A Blake, R Cipolla, Contour-based learning for object detection, Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on. Vol. 1. IEEE, 2005.

[1.7] M Oren, C Papageorgiou, P Sinha, E Osuna, T Poggio, Pedestrian detection using wavelet templates, Computer Vision and Pattern Recognition, 1997. Proceedings., 1997 IEEE Computer Society Conference on. IEEE, 1997.

[1.8] CJC Burges, A Tutorial on Support Vector Machines for Pattern Recognition, Data Mining and Knowledge Discovery, 2.2 (1998): 121-167.

[1.9] Y Freund and R E. Schapire. A decision-theoretic generalization of on-line learning and an application to boosting, Computational learning theory. Springer Berlin Heidelberg, 1995, 23–37.

[1.10] S Agarwal, D Roth, Learning a sparse representation for object detection, Computer Vision—ECCV 2002. Springer Berlin Heidelberg, 2002. 113-127.

[1.11] A Garg, S Agarwal, T Huang, Fusion of global and local information for object detection, Pattern Recognition, 2002. Proceedings. 16th International Conference on. Vol. 3. IEEE, 2002.

[1.12] R Fergus, P Perona, A Zisserman, Object class recognition by unsupervised scale-invariant learning, Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on. Vol. 2. IEEE, 2003.

[1.13] B Leibe, A Leonardis, B Schiele, Combined object categorization and segmentation with an implicit shape model, Workshop on statistical learning in computer vision, ECCV. Vol. 2. No. 5. 2004.

[1.14] B Leibe, A Leonardis, B Schiele, Robust Object Detection with Interleaved Categorization and Segmentation, International journal of computer vision 77.1-3 (2008): 259-289.

[1.15] M Fritz, B Leibe, B Caputo, B Schiele, Integrating Representative and Discriminative Models for Object Category Detection, Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on. Vol. 2. IEEE, 2005.

[1.16] J Mutch, DG Lowe, Multiclass object recognition with sparse localized features, Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on. Vol. 1. IEEE, 2006.

[1.17] CP Papageorgiou, M Oren, T Poggio, A general framework for object detection, Computer vision, 1998. Sixth international conference on. IEEE, 1998.

[1.18] P Viola, M Jones, Rapid object detection using a boosted cascade of simple features, Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on. Vol. 1. IEEE, 2001.

ΚΕΦΑΛΑΙΟ 2

ΤΕΧΝΗΤΗ ΟΡΑΣΗ ΚΑΙ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

2.1 ΕΙΣΑΓΩΓΗ

Στόχος της τεχνητής όρασης (computer vision) είναι η εξαγωγή χρήσιμων πληρφοριών από εικόνες. Μεγάλο πρόβλημα στην ανάπτυξη του συγκεκριμένου τομέα της επιστήμης είναι η πολυπλοκότητα και ο όγκος των οπτικών δεδομένων. Τα τελευταία χρόνια όμως υπάρχει σημαντική πρόοδος στην επιστήμη της τεχνητής όρασης κυρίως λόγω της ραγδαίας ανάπτυξης της ισχύς επεξεργασίας, της μνήμης και της αποθηκευτικής χωρητικότητας των υπολογιστών. Ένας επίσης σημαντικός λόγος είναι η εφαρμογή μεθόδων της μηχανικής μάθησης (machine learning) η οποία όχι μόνο παρέχει χρήσιμα εργαλεία αλλά βοηθάει στην καλύτερη κατανόηση των υπαρχόντων αλγορίθμων υπό νέα οπτική πλευρά.

Το παρόν κεφάλαιο περιλαμβάνει πληροφορίες πάνω στις πιθανότητες οι οποίες είναι μία χρήσιμη γλώσσα για την περιγραφή μοντέλων της τεχνητής όρασης. Περιλαμβάνει επίσης βασικές αρχές της μηχανικής μάθησης που θεμελειώνουν όλες τις μεθόδους εξαγωγής χρήσιμων πληροφοριών από εικόνες καθώς και στατιστικά μοντέλα που συσχετίζουν τα δεδομένα που περιέχονται σε μία εικόνα με την προς ανάκτηση πληροφορία. Επιπλέον περιγράφονται μέθοδοι επεξεργασίας της εικόνας, οι οποίες παίζουν καθοριστικό ρόλο στην αποτελεσματική λειτουργία του στατιστικού μοντέλου.

2.2 ΠΙΘΑΝΟΤΗΤΕΣ

Σχεδόν όλα τα μοντέλα της τεχνητής όρασης μπορούν να ερμηνευτούν μέσα από ένα πιθανοθεωρητικό πλαίσιο. Ο γενικός στόχος της τεχνητής όρασης είναι να χρησιμοποιηθούν τα δυσδιάστατα δεδομένα που προκύπτουν από την προβολή του τρισδιάστου κόσμου πάνω σε μία επιφάνεια ώστε να εξαχθούν συμπεράσματα για τον πραγματικό κόσμο που τα δημιούργησε. Υπάρχουν δύο προβλήματα ως προς την εξαγωγή συμπερασμάτων, το πρώτο είναι ότι η διαδικασία καταγραφής δεδομένων είναι θορυβώδης δηλαδή αυτό που καταγράφεται στον αισθητήρα δεν είναι μόνο η ποσότητα του φωτός που συγκεντρώνεται στην επιφάνειά του αλλά μία θορυβώδης εκτίμηση αυτής της ποσότητας. Η περιγραφή του θορύβου στα δεδομένα γίνεται με χρήση πιθανοτήτων. Δεύτερον η σχέση μεταξύ του πραγματικού κοσμου και των μετρήσεων είναι πολλά προς λίγα, μπορεί δηλαδή να υπάρχουν πολλές πραγματικές καταστάσεις που ταιριάζουν με τα καταγεγραμμένα δεδομένα. Η περίπτωση του να είναι στην πραγματικότητα μία από αυτές τις καταστάσεις βρίσκεται μέσω των πιθανοτήτων.

2.2.1 Τυχαίες μεταβλητές

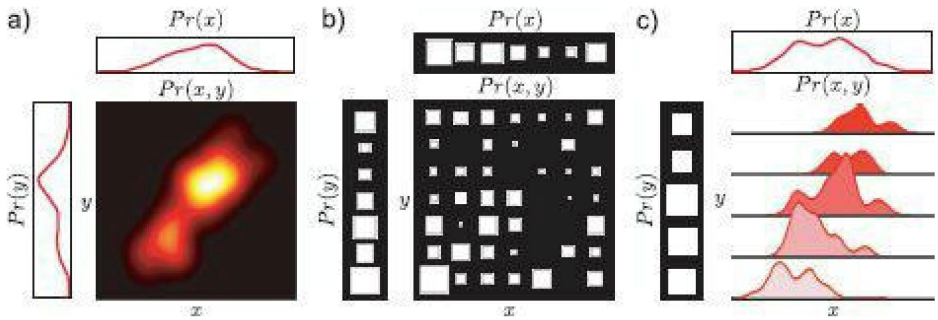
Μία τυχαία μεταβλητή x υποδηλώνει μία ποσότητα η οποία είναι αβέβαιη, όπως μία μεταβλητή κατάσταση του πραγματικού κόσμου π.χ. η θερμοκρασία. Εάν καταγραφούν αρκετές τιμες της μεταβλητής $\{x_i\}_{i=1}^{I}$ μπορεί κάποιες να είναι διαφορετικές μεταξύ τους αλλά και κάποιες να επαναλαμβάνονται. Αυτή η πληροφορία, των επαναλαμβανόμενων τιμών, παρουσιάζεται από την κατανομή πιθανοτήτων Pr(x) της τυχαίας κατανομής. Η τυχαία

μεταβλητή μπορεί να είναι διακριτή ή συνεχόμενη, πεπερασμένη ή άπειρη. Σημαντικό εργαλείο είναι η συνάρτηση πυκνότητας πιθανότητας η οποία αναπαριστά τη σχετική τάση της τυχαίας μεταβλητής να πάρει μία συγκεκριμένη τιμή. Το ολοκλήρωμα της συνάρτησης πυκνότητας πιθανότητας παίρνει πάντα τιμή ένα.

Στην περίπτωση που υπάρχουν δύο τυχαίες μεταβλητές x και y κάποιοι συνδιασμοί των x και y θα προκύπτουν πιό συχνά από άλλους. Αυτή η πληροφορία περιλαμβάνεται στην κοινή κατανομή πιθανοτήτων (joint probability) των x και y, η οποία γράφεται ως Pr(x,y). Η κοινή κατανομή πιθανοτήτων μπορεί να συσχετίζει μεταβλητές που να είναι όλες διακριτές ή όλες συνεχείς ή συνεχείς με διακριτές.

Στην περίπτωση κοινής κατανομής πιθανοτήτων υπάρχει η δυνατότητα του διαχωρισμού (marginalization), δηλαδή του υπολογισμού της κατανομής πιθανότητας καθεμίας μεταβλητής ξεχωριστά από την κοινη κατανομή πιθανότητας είτε ολοκληρώνοντας είτε αθροίζοντας πάνω στις άλλες μεταβλητές (Σχήμα 2.1). Για παράδειγμα εάν x και y είναι και οι δύο συνεχείς μεταβλητές και η Pr(x,y) είναι γνωστή τότε μπορεί να ανακτηθούν οι κατανομές Pr(x) και Pr(y) ως εξής:

$$\Pr(x) = \int \Pr(x, y)\partial y \qquad (2.1)$$
$$\Pr(y) = \int \Pr(x, y)\partial x$$



Σχήμα 2.1 Κοινές και διαχωρισμένες κατανομές πιθανοτήτων. a) x και y συνεχείς κατανομές. b) x και y διακριτές κατανομές. c) x συνεχής και y διακριτή.

2.2.2 Δεσμευμένη πιθανότητα - Κανόνας Bayes

Η δεσμευμένη πιθανότητα της μεταβλητής x δεδομένου ότι η y παίρνει την τιμή y^* δείχνει τη σχετική τάση της τυχαίας μεταβλητής x να λαμβάνει τιμές αποτελεσμάτων δοσμένου ότι η τυχαία μεταβλητή y έχει σταθερή τιμή y^*. Η εξαρτημένη πιθανότητα γράφεται ως $\Pr(x \mid y = y^*)$ και μπορεί να ανακτηθεί από την κοινή κατανομή Pr(x,y) λαμβάνοντας την κατάλληλη τομή $\Pr(x, y = y^*)$ (Σχήμα 2.2). Οι τιμές στην επιλεγμένη τομή εκφράζουν τη σχετική πιθανότητα για την μεταβλητή x να παίρνει ποικίλες τιμές έχοντας δεδομένο ότι $y = y^*$ αλλά η τομή αυτή δεν αποτελεί από μόνη της έγκυρη πιθανοτική κατανομή καθώς, είτε το άθροισμα είτε το ολοκλήρωμα της κατανομής δεν ισούται με μονάδα κάθως είναι μόνο ένα τμήμα της κοινής κατανομής, η οποία συνολικά ισούται με μονάδα. Οπότε για να υπολογιστεί η εξαρτημένη κατανομή πυκνότητας πιθανότητας γίνεται κανονικοποίηση υπό την συνολική πιθανότητα της τομής ως ακολούθως:

$$\Pr(x \mid y = y^*) = \frac{\Pr(x, y = y^*)}{\int \Pr(x, y = y^*)\partial x} = \frac{\Pr(x, y = y^*)}{\Pr(y = y^*)} \qquad (2.2)$$

Είναι σύνηθες να γράφεται η συνάρτηση εξαρτημένης πυκνότητας πιθανότητας χωρίς να προσδιορίζεται ρητά η τιμή $y = y^*$ με σκοπό να περιγραφεται από μία πιό συμπαγής παράσταση ως εξής:

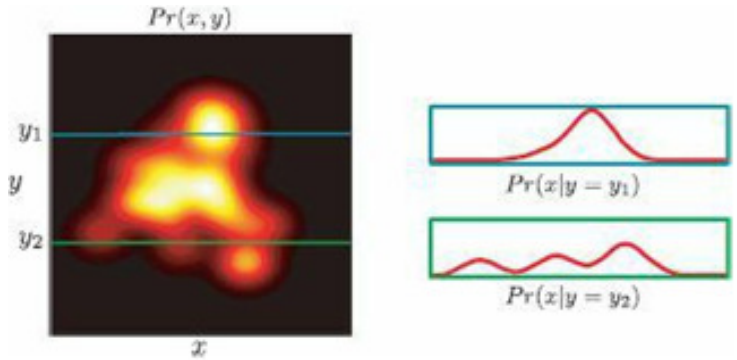
$$\Pr(x \mid y) = \frac{\Pr(x, y)}{\Pr(y)} \tag{2.3}$$

ή αλλιώς

$$\Pr(x, y) = \Pr(x \mid y)\Pr(y) \tag{2.4}$$

και λόγω συμμετρίας ισχύει αντίστοιχα :

$$\Pr(x, y) = \Pr(y \mid x)\Pr(x) \tag{2.5}$$



Σχήμα 2.2 Δεσμευμένη πιθανότητα. Αριστερά η από κοινού συνάρτηση πυκνότητας πιθανότητας των μεταβλητών x και y. Δεξιά δύο κατανομές δεσμευμένης πιθανότητας $\Pr(x \mid y = y_1)$ και $\Pr(x \mid y = y_2)$.

Εξισώνοντας τις δύο σχέσεις 2.4 και 2.5 προκύπει ο μαθηματικός τύπος γνωστός ως κανόνας του Bayes ως ακολούθως:

$$\begin{aligned}\Pr(y \mid x) &= \frac{\Pr(x \mid y)\Pr(y)}{\Pr(x)} \\ &= \frac{\Pr(x \mid y)\Pr(y)}{\int \Pr(x, y)dy} \\ &= \frac{\Pr(x \mid y)\Pr(y)}{\int \Pr(x \mid y)\Pr(y)dy}\end{aligned} \tag{2.6}$$

Ο όρος Pr(y|x) περιέχει την μεταγενέστερη ή “εκ των υστέρων” (posterior) πληροφορία, αναπαριστά δηλαδή τη γνώση που υπάρχει για τη μεταβλητή y δοσμένης της x. Αντιστρόφως ο όρος Pr(y) αναπαριστά την προγενέστερη ή “εκ των προτέρων” (prior) πληροφορία, δηλαδή τι είναι γνωστό ή αλλιώς την πεποίθηση για την y πριν ληφθεί υπόψιν η x. Ο όρος Pr(x|y) είναι το μοντέλο κατανομής πυκνότητας πιθανότητας που ακολουθεί η μεταβλητή x το οποίο παραμετροποιείται από την πληροφορία της μεταβλητής y και ονομάζεται ενδεχόμενο (likelihood). Η συνάρτηση Pr(x) ονομάζεται απόδειξη (evidence) και χρησιμεύει για την κανονικοποίηση της κατανομής. Κατά κύριο λόγο στα προβλήματα

τεχνητής όρασης το ενδιαφέρον είναι στη μεταβλητή y οπότε ο κανόνας του Bayes χρησιμοποιείται για τον υπολογισμό της πιθανότητας Pr(y|x).

2.3 ΚΟΙΝΕΣ ΚΑΤΑΝΟΜΕΣ ΠΙΘΑΝΟΤΗΤΩΝ

Η επιλογή της κατανομής Pr(x) που ακολουθούν τα δεδομένα **x** εξαρτάται από το είδος των δεδομένων που μοντελοποιούνται. Οι πιό συχνά χρησιμοποιούμενες κατανομές για την μοντελοποίηση των οπτικών δεδομένων παρουσιάζονται στον από κάτω πίνακα:

Κατανομές	**Δεδομένα x**	**Μοντελοποίηση παραμέτρων από:**
Bernoulli	$\boxed{x \in \{0,1\}}$	beta
Μονοπαραγοντική Κανονική	$\boxed{x \in \Re}$	Κανονική Αντίστροφη Gamma
Πολυπαραγοντική Κανονική	$\boxed{x \in \Re^k}$	Κανονική Αντίστροφη Wishart

Πίνακας 2.1 Κοινές κατανομές πιθανοτήτων. Στην πρώτη στήλη είναι τα ονόματα των κατανομών, στη δεύτερη το είδος των δεδομένων και στην τρίτη στήλη η κατανομή από την οποία μοντελοποιούνται οι παράμετροι της αντίστοιχης κατανομής της πρώτης στήλης

Όταν προσαρμόζεται μία κατανομή πιθανοτήτων στα δεδομένα χρειάζεται να υπολογίζεται πόσο αβέβαιη είναι αυτή η προσαρμογή. Αυτή η αβεβαιότητα αναπαρίσταται ως πιθανοτική κατανομή πάνω στις παραμέτρους του προς προσαρμογή μοντέλου. Οπότε για κάθε κατανομή που χρησιμοποιείται για μοντελοποίηση δεδομένων υπάρχει μία δεύτερη κατανομή πάνω στις σχετιζόμενες παραμέτρους της οποίας οι παράμετροι ονομάζονται υπερπαράμετροι. Πιό γενικά οι υπερπαράμετροι καθορίζουν το σχήμα της αρχικής κατανομής πάνω στις παραμέτρους.

2.3.1 Κατανομή Bernoulli και Beta

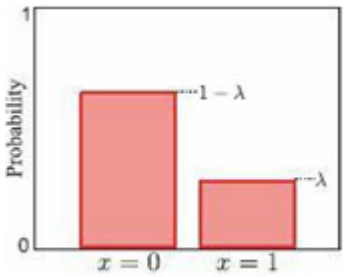
Η κατανομή Bernoulli είναι μία διακριτή κατανομή που μοντελοποιεί δυαδικά δεδομένα, περιγράφει δηλαδή περιπτώσεις όπου υπάρχουν δύο μόνο αποτελέσματα $x \in \{0,1\}$. Η κατανομή Bernoulli στα πλαίσια της μηχανικής μάθησης μπορεί για παράδειγμα να περιγράψει την πιθανότητα ενός pixel να πάρει τιμή έντασης μικρότερη ή μεγαλύτερη μίας σταθεράς ή επίσης να μοντελοποιήσει την πραγματική κατάσταη ως προς το εάν υπάρχει πρόσωπο ή όχι σε μία εικόνα.

Η κατανομή Bernoulli (Σχήμα 2.3) έχει μία μόνο παράμετρο τη $\lambda \in [0,1]$ και καθορίζει την πιθανότητα το x να είναι ένα.

$$\begin{aligned} \Pr(x=0) &= 1-\lambda \\ \Pr(x=1) &= \lambda \end{aligned} \tag{2.7}$$

ή εναλλακτικά

$$\Pr(x) = \lambda^x (1-\lambda)^{1-x} \tag{2.8}$$
$$\Pr(x) = Bern_x[\lambda]$$



Σχήμα 2.3 Η κατανομή Bernoulli είναι διακριτή με δύο πιθανά αποτελέσματα {0,1}, έχει μόνο μία παράμετρο τη λ που εκφράζει την πιθανότητα το αποτέλεσμα να είναι 1.

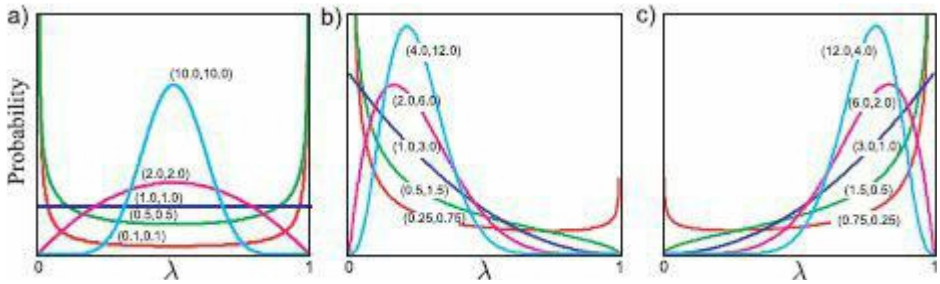
Για την μοντελοποίηση των παραμέτρων της κατανομής Bernoulli χρησιμοποιείται η κατανομή Beta. Η κατανομή Beta είναι μία συνεχής κατανομή με μία μεταβλητή λ όπου $\lambda \in [0,1]$, γεγονός που την καθιστά κατάλληλη για την έκφραση της αβεβαιότητας της παραμέτρου λ της κατανομής Bernoulli.

Η κατανομή Beta έχει δύο παραμέτρους $\alpha, \beta \in [0,\infty]$ οι οποίες επηρεάζουν το σχήμα καμπύλης (Σχήμα 2.4). Ο μαθηματικός τύπος της κατανομής είναι:

$$\Pr(\lambda) = \frac{\Gamma[\alpha+\beta]}{\Gamma[\alpha]\Gamma[\beta]} \lambda^{\alpha-1}(1-\lambda)^{\beta-1} \tag{2.9}$$

$$\Pr(\lambda) = Beta_\lambda(\alpha,\beta)$$

όπου Γ[ ] είναι η συνάρτηση gamma.



Σχήμα 2.4 Κατανομή Beta. Έχει πεδίο ορισμού [0,1] και δύο παραμέτρους (α,β) οι τιμές των οποίων επηρεάζουν τη μέση τιμή και την συγκέντρωση της καμπύλης γύρω από τη μέση τιμή.

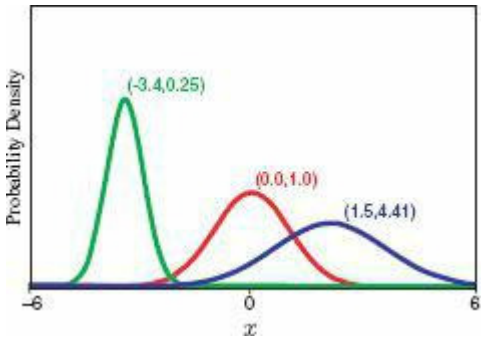
2.3.2 Μονοπαραγοντική κανονική κατανομή – Κανονικά κλιμακώμενη αντίστροφη κατανομή gamma

Κατανομή που χρησιμοποιείται ευρύτατα στη μοντελοποιήση οπτικών δεδομένων είναι η κανονική μονοπαραγοντική κατανομή (univariate normal distribution) ή αλλιώς κατανομή Gauss της οποίας οι παράμετροι μοντελοποιούνται από την κανονικά κλιμακώμενη αντίστροφη gamma κατανομή (normal-scaled inverse gamma distribution) . Η κανονική κατανομή μίας μεταβλητής ή αλλιώς κατανομή Gauss (Σχήμα 2.5) ορίζεται πάνω σε συνεχείς τιμές $x \in [-\infty,\infty]$. Στην τεχνητή όραση αγνοείται το γεγονός ότι η τιμή της έντασης ενός

pixel είναι κβαντισμένη και μοντελοποιείται με συνεχή κανονική κατανομή. Εκτός της τιμής της έντασης ενός pixel η κανονική κατανομή μοντελοποιεί και την κατάσταση του πραγματικού κόσμου, για παράδειγμα η απόσταση από ένα αντικείμενο μπορεί να παρασταθεί με αυτό τον τρόπο.

Η κανονική κατανομή έχει δύο παραμέτρους, τη μέση τιμή μ και τη διακύμανση σ^2. Η παράμετρος μ μπορεί να πάρει οποιαδήποτε τιμή και καθορίζει τη θέση της κορυφής. Η παράμετρος σ^2 παίρνει μόνο θετικές τιμές και καθορίζει το πλάτος της κατανομής. Η κανονική κατανομή ορίζεται ως εξής :

$$\Pr(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-0.5\frac{(x-\mu)^2}{\sigma^2}\right] \tag{2.10}$$



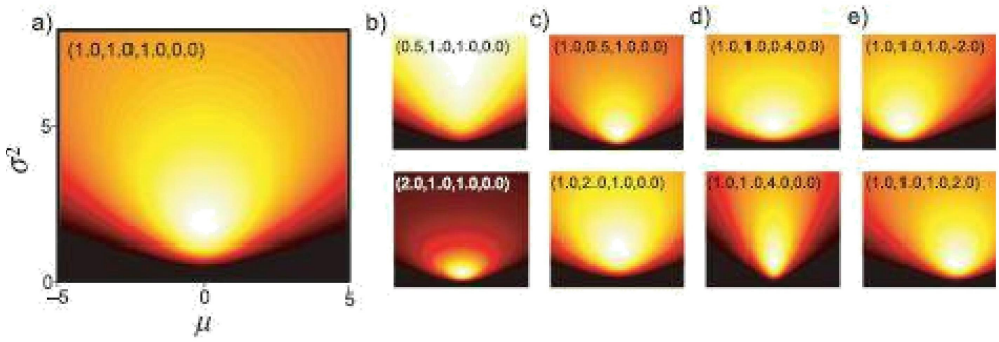
Σχήμα 2.5 Κανονική μονοπαραγοντική κατανομή ή κατανομή Gauss. Το $x \in \Re$ και έχει δύο παραμέτρους $\{\mu, \sigma^2\}$ όπου μ η μέση τιμή και σ^2 η διακύμανση.

Η κανονικά κλιμακώμενη αντίστροφη κατανομή gamma (Σχήμα 2.6) έχει δύο συνεχόμενες μεταβλητές που στη συγκεκριμένη περίπτωση θα συμβολίζονται ως μ και σ^2. Η πρώτη παίρνει οποιαδήποτε τιμή και η δεύτερη οποιαδήποτε θετική τιμή. Η συγκεκριμένη κατανομή έχει τις ιδιότητες για να περιγράψει την αβεβαιότητα πάνω στη μέση τιμή και στη διακύμανση της κανονικής κατανομής. Η συγκεκριμένη κατανομή έχει τέσσερεις παραμέτρους α, β, γ, δ όπου οι α, β, γ είναι πραγματικοί θετικοί αριθμοί ενώ η δ μπορεί να πάρει οποιαδήποτε τιμή. Η συνάρτηση είναι η εξής:

$$\Pr(\mu, \sigma^2) = \frac{\sqrt{\gamma}\beta^{\alpha}}{\sigma\sqrt{2\pi}\Gamma[\alpha]}\left(\frac{1}{\sigma^2}\right)^{\alpha+1} \exp\left[-\frac{2\beta + \gamma(\delta-\mu)^2}{2\sigma^2}\right] \tag{2.11}$$

2.3.3 Πολυπαραγοντική Κανονική κατανομή- Κανονικά αντίστροφη κατανομή Wishart

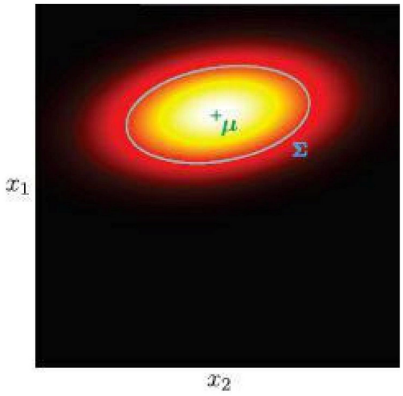
Η πολυπαραγοντική κανονική κατανομή (multivariate normal distribution) ή Gauss κατανομή (Σχήμα 2.7) μοντελοποιεί D-διαστάσεων μεταβλητές **x** όπου κάθε στοιχείο της $x_1 \ldots x_D$ είναι συνεχής στο διάστημα $[-\infty, +\infty]$. Στην τεχνητή όραση η πολυπαραγοντική κανονική κατανομή μπορεί να μοντελοποιήσει τόσο τις εντάσεις D pixels σε ένα τμήμα μίας εικόνας όσο και τη κατάσταση της πραγματικότητας όπως για παράδειγμα η μοντελοποίηση της αβεβαιότητας της τρισδιάστατης θέσης (x,y,z) ενός αντικειμένου στο σκηνικό.



Σχήμα 2.6 Κανονικά κλιμακώμενη αντίστροφη κατανομή gamma. Αποτελεί δυσδιάστατη κατανομή δύο συνεχόμενων μεταβλητών μ και σ^2, έχει τις παραμέτρους α, β, γ, δ οι οποιές παίρνουν διάφορες τιμές στις περιπτώσεις a, b, c, d, e.

Η πολυπαραγοντική κανονική κατανομή έχει δύο παραμέτρους, τον μέσο όρο **μ** και τη συνδιακύμανση Σ. Ο μέσος όρος είναι ένα $D \times 1$ διάνυσμα και η συνδιακύμανση Σ είναι ένας συμμετρικός $D \times D$ θετικά ορισμένος πίνακας. Η συνάρτηση πυκνότητας πιθανότητας έχει την ακόλουθη μορφή:

$$\Pr(x) = \frac{1}{(2\pi)^{D/2}|\Sigma|^{1/2}} \exp\left[-0.5(x-\mu)^{\mathrm{T}} \Sigma^{-1} (x-\mu)\right] \tag{2.12}$$

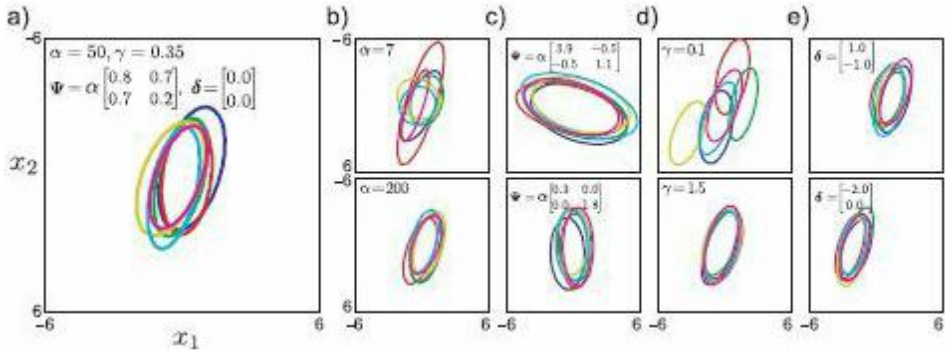


Σχήμα 2.7 Πολυπαραγοντική κανονική κατανομή. Μοντελοποιεί D-διαστάσεων μεταβλητές **x**. Έχει ως παραμέτρους διάνυσμα μεσων τιμών **μ** διαστάσεων Dx1και πίνακα συνδιακυμάνσεων **Σ** διαστάσεων DxD. Τα ισοεπίπεδα σημεία σχηματίζουν έλλειψη, το κέντρο της καθορίζεται από το **μ** και το σχήμα από το **Σ**

Η κανονικά ανίστορφη κατανομή Wishart (normal inverse Wishart distribution) (Σχήμα 2.8) έχει δύο μεταβλητές, ένα $D \times 1$ διάνυσμα που στη συγκεκριμένη περίπτωση ονομάζεται **μ** και σε ένα $D \times D$ θετικά ορισμένο πίνακα **Σ**. Ως εκ τούτου είναι κατάλληλη για την περιγραφή της αβεβαιότητας των παραμέτρων της πολυπαραγοντικής κανονικής κατανομής. Η κανονικά αντίστροφη κατανομή Wishart έχει τέσσερεις παραμέτρους τις α, **Ψ**, γ, **δ**, όπου α και γ είναι θετικοί βαθμωτοί αριθμοί, **δ** είναι ένα $D \times 1$ διάνυσμα και **Ψ** ένας $D \times D$ θετικά ορισμένος πίνακας. Η συνάρτηση πυκνότητας πιθανότητας είναι:

$$\Pr(\mu,\Sigma)=\frac{\gamma^{D/2}|\Psi|^{a/2}|\Sigma|^{-(a+D+2)/2}\exp\left[-0.5(Tr\left[\Psi\Sigma^{-1}\right]+\gamma(\mu-\delta)^{T}\Sigma^{-1}(\mu-\delta)\gamma^{D/2})\right]}{2^{aD/2}(2\pi)^{D/2}\Gamma_{D}[a/2]} \quad (2.13)$$

όπου $\Gamma_D[\ ]$ είναι η πολυπαραγοντική συνάρτηση gamma και το Tr[Ψ] επιστρέφει το ίχνος του πίνακα Ψ.



Σχήμα 2.8 a) Κανονικά αντίστορφη κατανομή Wishart. Δέχεται ως μεταβλητές ένα διάνυσμα Dx1 διαστάσεων και ένα πίνακα DxD διαστάσεων και έχει έξι παραμέτρους α, **Ψ**, γ, **δ** . Στο σχήμα η οπτικοποίηση γίνεται με δυσδιάστατες ελλείψης που περιγράφουν τα ισοεπίπεδα σημεία της πολυπαραγοντικής κανονικής κατανομής. b) Αλλαγή της παραμέτρου α c) Αλλαγή της παραμέτρου **Ψ** d) Αλλαγή της παραμέτρου γ e) Αλλαγή της παραμέτρου **δ**

2.3.4 Συζυγία

Τα ζεύγη κατανομών που παρουσιάστηκαν στις υποπαραγράφους 2.3.1, 2.3.2, 2.3.3, έχουν επιλεγεί με ιδιαίτερη προσοχή καθώς έχουν μία ειδική σχέση, σε κάθε περίπτωση η δεύτερη κατανομή είναι συζυγής της πρώτης, για παράδειγμα η beta είναι συζυγής στην Bernoulli. Όταν πολλαπλασιάζεται μία κατανομή με τη συζυγή της τότε το αποτέλεσμα είναι ανάλογο με μία νέα κατανομή η οποία έχει τη μορφή της συζυγούς. Για παράδειγμα:

$$Bern_{x}[\lambda]Beta_{\lambda}[\alpha,\beta]=\kappa(x,\alpha,\beta)Beta_{\lambda}\left[\tilde{\alpha},\tilde{\beta}\right] \quad (2.14)$$

όπου κ(x,α,β) είναι ένας βαθμωτός παράγοντας που είναι σταθέρος σε σχέση την ενδιαφερόμενη μεταβλητή λ.

Η συζυγής σχέση είναι σημαντική ιδιότητα καθώς στην τεχνητή όραση λαμβάνονται γινόμενα κατανομών τόσο κατά τη διαδικασία εκμάθησης των παραμέτρων (προσαρμογή κατανομών) όσο και κατά την εκτίμηση των νέων δεδομένων που προκύπει από το μοντέλο (υπολογισμός της πιθανότητας νέων δεδομένων να βρίσκονται υπό την προσαρμοσμένη κατανομή). Η συζυγής σχέση σημαίνει ότι τα αναφερθέντα γινόμενα μπορούν να υπολογιστούν απλά, σε κλειστή μορφή.

2.4 ΠΡΟΣΑΡΜΟΓΗ ΠΙΘΑΝΟΤΙΚΩΝ ΚΑΤΑΝΟΜΩΝ

Η διαδικασία προσαρμογής των πιθανοτικών κατανομών στα δεδομένα $\{x_i\}_{i=1}^{I}$ αναφέρεται ως εκμάθηση καθώς μαθαίνονται οι παράμετροι **θ** της κατανομής. Υπάρχουν τρείς μέθοδοι προσαρμογής: το μέγιστο ενδεχόμενο (maximum likelihood), το μεταγενέστερο μέγιστο (maximum a posteriori), η προσέγγιση Bayes.

2.4.1 Μέγιστο ενδεχόμενο

Θεωρώντας ότι κάθε μέτρηση $\{x_i\}_{i=1}^{I}$ είναι ανεξάρτητη από τς υπόλοιπες, η συνάρτηση πιθανότητας $\Pr(x_{1\ldots I} \mid \theta)$ για ένα σύνολο σημείων είναι το γινόμενο των ξεχωριστών ατομικών πιθανοτήτων οπότε η εκτίμηση των παραμέτρων **θ** σύμφωνα με τη μέθοδο μέγιστου ενδεχομένου είναι :

$$\hat{\theta} = \arg\max_{\theta}\left[\Pr(x_{1\ldots I} \mid \theta)\right] \tag{2.15}$$

$$= \arg\max_{\theta} \prod_{i=1}^{I} \Pr(x_i \mid \theta)$$

Όπου η συνάρτηση $\underset{\theta}{\arg\max} f[\theta]$ επιστρέφει τη τιμή **θ** που μεγιστοποιεί την συνάρτηση $f[\theta]$. Για να υπολογιστεί η πιθανότητα ενός νέου δεδομένου x^* να ανήκει στην προσαρμοσμένη κατανομή υπολογίζεται η συνάρτηση πυκνότητας πιθανότητας $\Pr\left(x^* \mid \hat{\theta}\right)$.

Πολύ συχνά για τη διευκόλυνση της επίλυσης χρησιμοποιείται ο λογάριθμος της σχέσης (2.15) (το λογαριθμικό ενδεχόμενο, L). Καθώς ο λογάριθμος είναι μονότονη συνάρτηση, η θέση του μεγίστου στη μετασχηματισμένη συνάρτηση παραμένη η ίδια. Αλγεβρικά ο λογάριθμος μετατρέπει το γινόμενο των ξεχωριστών σημείων σε άθροισμα οπότε και αποσυνδέεται η συνεισφορά του καθενός. Η σχέση (2.15) μετασχηματίζεται ως εξής:

$$\hat{\theta} = \arg\max_{\theta}\left[\sum_{i=1}^{I} \log[\Pr(x_i \mid \theta)]\right] \tag{2.16}$$

2.4.2 Μεταγενέστερο μέγιστο

Στη μέθοδο προσαρμογής παραμέτρων "μεταγενέστερο μέγιστο", εισάγεται η προγενέστερη πληροφορία για την παράμετρο **θ**. Αυτό επιτυγχάνεται κάνοντας εκτίμηση της πιθανοτικής κατανομής των παραμέτρων κυρίως μέσω προηγούμενης εμπειρίας. Όπως και το όνομα υπονοεί, η μέθοδος "μεταγενέστερο μέγιστο" μεγιστοποιεί την συνάρτηση πυκνότητας πιθανότητας $\Pr(\theta \mid x_{1\ldots I})$ πάνω στις παραμέτρους. Η συνάρτηση προς μεγιστοποίηση προκύπτει ως εξής:

$$\hat{\theta} = \arg\max_{\theta}\left[\Pr(\theta \mid x_{i\ldots I})\right] \tag{2.17}$$

$$= \arg\max_{\theta} \left[\frac{\Pr(x_{i...I} \mid \theta) \Pr(\theta)}{\Pr(x_{1...I})} \right]$$

$$= \arg\max_{\theta} \left[\frac{\prod_{i=1}^{I} \Pr(x_i \mid \theta) \Pr(\theta)}{\Pr(x_{1...I})} \right]$$

Όπου χρησιμοποιήθηκε ο κανόνας του Bayes θεωρώντας ότι τα δεδομένα είναι ανεξάρτητα. Ο παρονομαστής μπορεί να διαγραφεί καθώς είναι σταθερός σε σχέση με τις παραμέτρους και έτσι δεν επηρεάζει τη θέση του μέγιστου. Η από πάνω σχέση γίνεται:

$$\hat{\theta} = \arg\max_{\theta} \left[\prod_{i=1}^{I} \Pr(x_i \mid \theta) \Pr(\theta) \right] \tag{2.18}$$

Η από πάνω σχέση είναι όμοια με τη μέθοδο "μέγιστου ενδεχομένου" με τη διαφορά ότι έχει προστεθεί ο όρος Pr(θ), η κατανομή προγενέστερης πληροφορίας πάνω στις παραμέτρους **θ**. Όπως και στην περίπτωση του μέγιστου ενδεχομένου η πιθανότητα ενός νέου δεδομένου x^* να ανήκει στο προσαρμοσμένο μοντέλο υπολογίζεται από τη συνάρτηση πυκνότητας πιθανότητας $\Pr\left(x^* \mid \hat{\theta}\right)$.

2.4.3 Bayesian-ή προσέγγιση

Στην Bayesian-ή προσέγγιση δεν ακολουθείται διαδικασία εύρεσης συγκεκριμένων τιμών των μεταβλητών **θ** καθώς θεωρείται το προφανές, ότι υπάρχουν πολλές τιμές των μεταβλητών **θ** που είναι συμβατές με τα δεδομένα. Υπολογίζεται μία πιθανοτική κατανομή $\Pr(\theta \mid x_{i..I})$ πάνω στις παραμέτρους θ βασισμένη στα δεδομένα $\{x_i\}_{i=1}^{I}$ χρησιμοποιώντας το κανόνα του Bayes ώστε:

$$\Pr\left(\theta \mid x_{1...I}\right) = \left[\frac{\prod_{i=1}^{I} \Pr(x_i \mid \theta) \Pr(\theta)}{\Pr(x_{1...I})} \right] \tag{2.19}$$

Εν συνεχεία ο υπολογισμός της πιθανότητας ενός νέου δεδομένου x^* να ανήκει στην ίδια κατανομή με τα $\{x_i\}_{i=1}^{I}$ γίνεται μέσω μίας πιθανοτικής κατανομής πάνω στα πιθανά μοντέλα η οποία υπολογίζεται ως εξής:

$$\Pr\left(x^* \mid x_{1...I}\right) = \int \Pr(x^* \mid \theta) \Pr(\theta \mid x_{1...I}) \partial\theta \tag{2.20}$$

Το ολοκλήρωμα μπορεί να θεωρηθεί ως ένα σταθμισμένο άθροισμα των προβλέψεων $\Pr\left(x^* \mid \theta\right)$ οι οποίες προκύπτουν από διαφορετικές παραμέτρους **θ** και η στάθμισή τους καθορίζεται από τη μεταγενέστερη πιθανοτική κατανομή $\Pr(\theta \mid x_{i..I})$ πάνω στις

παραμέτρους (εκφράζοντας έτσι την πεποίθηση ότι διαφορτικές τιμές παραμέτρων **θ** μπορεί να είναι εξίσου σωστές).

Η Bayesian-ή προσέγγιση είναι η πιο έγκυρη μέθοδος αλλά οδηγεί σε συγκεκριμένη ορθή λύση μόνο όταν η συνάρτηση προγενέστερης πιθανότητας είναι συζυγής της συνάρτησης ενδεχόμενης πιθανότητας.

2.5 ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ (MACHINE LEARNING)

Στην περίπτωση της μηχανικής μάθησης η όραση αντιμετωπίζεται ως διαδικασία αναγνώρισης προτύπου, μία καινούργια δηλαδή εικόνα ερμηνεύεται βάση προγενέστερης γνώσης από εικόνες που το περιεχόμενό τους είναι γνωστό. Η διαδικασία αυτή χωρίζεται σε δύο μέρη. Πρώτο μέρος είναι η διαδικασία της εκμάθησης όπου μοντελοποιείται η σχέση μεταξύ των δεδομένων της εικόνας **x** και της κατάστασης του πραγματικού κόσμου **w** και υπολογίζονται οι παράμετροι του μοντέλου αυτού με τη χρήση εκπαιδευτικών δεδομένων. Σε πολλές περιπτώσεις δεν είναι απαραίτητα τα εκπαιδευτικα δεδομένα οπότε και ακολουθείτε πιο λυτή μέθοδος εκμάθησης. Δεύτερο μέρος είναι η διαδικασία επαγωγής συμπεράσματος για νέες εικόνες όπου γίνεται εκμετάλευση της προαναφερθείσας μοντελοποιήσης με σκοπό την πρόβλεψη των περιεχομένων των νέων εικόνων. Τα οπτικά δεδομένα **x** που λαμβάνονται μπορεί να είναι συμβατά με πάνω από μία πραγματική καταστάση **w** οπότε για τη συγκεκριμένη ασάφεια η καλύτερη προσέγγιση είναι η χρήση της μεταγενέστερης πιθανοτικής κατανομής $\Pr(w \mid x)$ πάνω στις πιθανές καταστάσεις **w** για την επαγωγή συμπεράσματος.

Για να επιλυθεί ένα πρόβλημα οπτικής μέσω μηχανικής μάθησης χρειάζονται τρία στοιχεία:

1) Ένα μοντέλο που θα σχετίζει τα οπτικά δεδομένα **x** με την πραγματική κατάσταση **w**. Το μοντέλο θα προσδιορίζει ένα σύνολο πιθανών σχέσεων (κατανομών) μεταξύ των **x** και **w** και η σχέση (κατανομή) θα συγκεκριμενοποιείται με τον προσδιορισμό συγκεκριμένων παραμέτρων **θ.**

2) Έναν αλγόριθμο εκμάθησης που θα προσδιορίζει τις παραμέτρους **θ** του μοντέλου χρησιμοποιώντας δείγματα $\{x_i, w_i\}$ όπου είναι γνωστές και οι μετρήσεις **x** και η πραγματικές καταστάσεις **w**.

3) Ένας αλγόριθμος εξαγωγής συμπεράσματος ο οποίος θα λαμβάνει μία νέα παρατήρηση x^* και θα επιστρέφει τη πιθανότητα $\Pr(w \mid x^*, \theta)$ πάνω στις πραγματικές καταστάσεις **w**.

Κάθε μοντέλο που σχετίζει τα δεδομένα **x** με την πραγματική κατάσταση **w** εμπίπτει σε μία από τις παρακάτω δύο κατηγορίες:

1. Μοντελοποίηση της πραγματικής κατάστασης πάνω στα δεδομένα $\Pr(w \mid x)$.
2. Μοντελοποίηση των δεδομένων πάνω στην πραγματική κατάσταση $\Pr(x \mid w)$.

Η πρώτη κατηγορία μοντέλου ονομάζεται διακριτικό (discriminative) και η δεύτερη γεννητικό (generative). Στο διακριτικό μοντέλο για τη διαδικασία εκμάθησης μοντελοποιείται η συνάρτηση πυκνότητας πιθανότητας $\Pr(w \mid x)$. Για τη μοντελοποίηση αυτή επιλέγεται πρώτα μία κατάλληλη κατανομή Pr(w) πάνω στις πραγματικές καταστάσεις **w** και στη συνέχεια οι παραμέτροι της κατανομής γίνονται συνάρτηση των δεδομένων **x**. Για παράδειγμα εάν η πραγματική κατάσταση είναι συνεχής κατανομή η Pr(w) μπορεί να μοντελοποιηθεί με μία κανονική μονοπαραγοντική ή πολυπαραγοντική (Gaussian-ή) κατανομή και ο μέσος όρος **μ** να είναι συνάρτηση των δεδομένων **x**. Η τιμή που αυτή η συνάρτηση επιστρέφει εξαρτάται και από ένα επιπλέον σύνολο παραμέτρων **θ**, οπότε η

μεταγενέστερη κατανομή ορίζεται ως $\Pr(w \mid x, \theta)$. Ο στόχος του αλγορίθμου εκμάθησης είναι να προσαρμόσει τις παραμέτρους χρησιμοποιώντας ζεύγη "εκπαιδευτικών" δεδομένων $\{x_i, w_i\}_{i=1}^{I}$. Αυτό μπορεί να επιτευχθεί με την εφαρμογή μίας εκ των μεθόδων "μέγιστο ενδεχόμενο" (ML), μέγιστο μεταγενέστερο ενδεχόμενο (MAP) και την Bayesian-ή προσέγγιση.

Στην διαδικασία επαγωγής συμπεράσματος ο στόχος είναι , χρησιμοποιώντας τα αποτελέσματα του αλγορίθμου εκμάθησης, να προσδιοριστεί η κατανομή πάνω στις πραγματικές καταστάσεις **w** για μία νέα καταγραφή x^*. Στη συγκεκριμένη κατηγορία η εύρεση μοντέλου κατανομής είναι εύκολη υπόθεση καθώς η έκφραση της μεταγενέστερης κατανομής $\Pr(w \mid x, \theta)$ έχει εξ ορισμού δημιουργηθεί από τη διαιδικασία εκμάθησης και απλώς χρειάζεται ο υπολογισμός της με τα νέα δεδομένα.

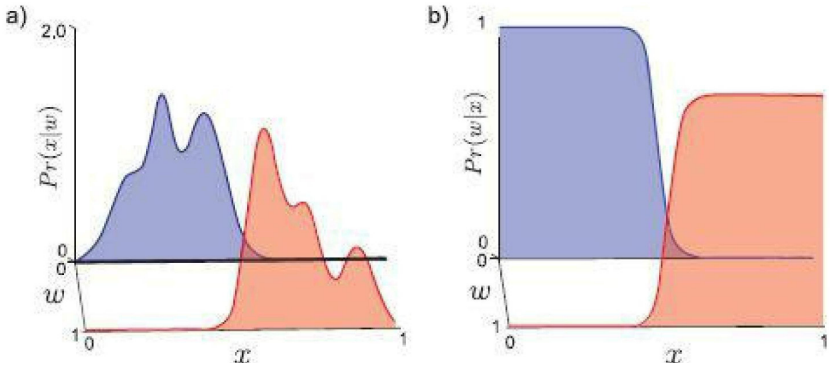
Στο γεννητικό μοντέλο για τη διαδικασία της εκμάθησης μοντελοποιείται η συνάρτηση πυκνότητας πιθανότητας Pr(x|w). Οπότε αρχικά επιλέγεται η επιθυμητή κατανομή Pr(x) πάνω στα δεδομένα και με κάποια εκ των παραμέτρων της **θ** να είναι συνάρτηση της πραγματικής κατάστασης w. Η κατανομή συμβολίζεται εκ νέου ως Pr(x|w,θ) και αναφέρεται ως συνάρτηση ενδεχομένου.Ο στόχος της εκμάθησης και εδώ είναι η προσαρμογή των παραμέτρων **θ** χρησιμοποιώντας ζεύγη "εκπαιδευτικών" δεδομένων $\{x_i, w_i\}_{i=1}^{I}$.

Για την επαγωγή συμπεράσματος στόχος είναι ο υπολογισμός της μεταγενέστερης κατανομής Pr(w|x). Για να επιτευχθεί η επαγωγή συμπεράσματος προσδιορίζεται η προγενέστερη κατανομή Pr(w) πάνω στις πραγματικές κατάστασεις **w** και εν συνεχεία με τη χρήση του κανόνα Bayes υπολογίζεται η μεταγενέστερη κατανομή Pr(w|x) :

$$\Pr(w \mid x) = \frac{\Pr(x \mid w)\Pr(w)}{\int \Pr(x \mid w)\Pr(w)dw} \tag{2.21}$$

Για το ποιό μοντέλο, διακριτικό ή γεννητικό, χρησιμοποιείται σε κάθε περίπτωση δεν υπάρχει ξεκάθαρη απάντηση, παρόλα αυτά υπάρχουν οι εξής θεωρήσεις:

- Η επαγωγή συμπεράσματος είναι πιό απλή διαδικασία στα διακριτικά μοντέλα καθώς μοντελοποείται ήδη από την διαδικασία της εκμάθησης η δεσμευεμένη πιθανοτική κατανομή της πραγματικής κατάστασης $\Pr(w \mid x, \theta)$. Αντίθετα στο γεννητικό μοντέλο η μεταγενέστερη κατανομή $\Pr(w \mid x, \theta)$ υπολογίζεται έμμεσα από τον κανόνα του Bayes χρειάζεται μεγαλύτερη ισχύ επεξεργασίας
- Τα γεννητικά μοντέλα δημιουργούν στη διαδικασία της εκμάθησης πιθανοτικά μοντέλα Pr(x|w) πάνω στα δεδομένα ενώ τα διακριτικά πιθανοτικά μοντέλα Pr(w|x) πάνω στην πραγματική κατάσταση (Σχήμα 2.9). Τα δεδομένα είναι συνήθως αρκετά μεγαλύτερης τάξης διαστάσεων από ότι οι πραγματικές καταστάσεις και η μοντελοποίησή τους είναι "δαπανηρή" διαδικασία.
- Μοντελοποιώντας τη συνάρτηση ενδεχομένου Pr(x|w) αντικατοπτρίζεται ο τρόπος που τα δεδομένα δημιουργήθηκαν, οπότε εάν είναι επιθυμητό να ενπεριέχεται η πληροφορία για τη διαιδικασία δημιουργίας των δεδομένων τότε το γεννητικό μοντέλο είναι το κατάλληλο.
- Μία θεμελειώδης ιδιότητα του γεννητικού μοντέλου είναι ότι επιτρέπει την ενσωμάτωση εξειδικευμένης γνώσης με τη μορφή της προγενέστερης πιθανοτικής κατανομής. Η ενσωμάτωση προγενέστερης γνώσης με δομημένο τρόπο στο διακριτικό μοντέλο είναι αρκετά πιο δύσκολη.



Σχήμα 2.9 a) Γεννητική προσσέγιση: Στη φάση της εκμάθησης μοντελοποιείται η συνάρτηση πυκνότητας πιθανότητας Pr(x|w) πάνω στα δεδομένα για κάθε τάξη πραγματικής κατάστασης. Απαιτείται σύνθετη μοντελοποίηση με πολλές παραμέτρους. b) Διακριτική προσέγγιση: Στη φάση της εκμάθησης μοντελοποιείται η μεταγενέστερη συνάρτηση πυκνότητας πιθανότητας Pr(w|x) πάνω στην πραγματική κατάσταση, παίρνοντας την προγενέστερη κατανομή Pr(w) ως ομοιόμορφη.

2.6 ΑΦΑΙΡΕΣΗ ΠΑΡΑΣΚΗΝΙΟΥ

Μία εφαρμογή του γεννητικού μοντέλου που χρησιμοποιείται και στην παρούσα διπλωματική εργασία είναι η αφαίρεση του παρασκηνίου. Σκοπός είναι η τοποθέτηση δυαδικής ετικέτας $w_n \in \{0,1\}$, που υποδεικνύει εάν το n^{th} pixel μίας εικόνας είναι μέρος γνωστού παρασκηνίου (w=0) ή εάν αποτελεί μέρος αντικειμένου του προσκηνίου (w=1). Το μοντέλο βασίζεται στις RGB μετρήσεις $x_n = \left[x_n^R, x_n^G, x_n^B\right]$ κάθε pixel.

Είναι σύνηθες να χρησιμοποιούνται εκπαιδευτικά δεδομένα $\{x_{in}\}_{i=1,n=1}^{I,N}$ τα οποία είναι ένα πλήθος εικόνων με άδειο σκηνικό που σημαίνει ότι όλα τα pixel ανήκουν στο παρασκήνιο. Στη συγκεκριμένη εφαρμογή συνήθως δεν χρρησιμοποιούνται φωτογραφίες αντικειμένων του προσκηνίου για εκπαίδευση εξαιτίας της μεγάλης ποικιλίας που αυτά μπορεί να έχουν, δεν παρέχεται δηλαδή από τη μέθοδο η δυνατότητα ταξινόμησης αντικειμένων του παρασκηνίου. Ακολουθούν δύο μέθοδοι μοντελοποίησης του παρασκηνίου, η μία μέσω κανονικής κατανομής και η άλλη μέσω συνδιασμού κανονικών κατανομών.

2.6.1 Μοντελοποίηση παρασκηνίου με κανονική κατανομή

Η ταξικά εξαρτώμενη κατανομή του παρασκηνίου $\Pr(x_n \mid w=0)$ μοντελοποιείται ως κανονική κατανομή (Σχέση 2.22), μονοπαραγοντική εάν αφορά κάθε μεμονωμένο pixel ή πολυπαραγοντική εάν πρόκειται για όμαδα pixel σε μία περιοχή της εικόνας :

$$\Pr(x_n \mid w=0) = Norm_{x_n}\left[\mu_{n0}, \Sigma_{n0}\right] \tag{2.22}$$

Η προγενέστερη πιθανοτική κατανομή Pr(w) πάνω στην πραγματική κατάσταση w ορίζεται από την κατανομή Bernouli :

$$\Pr(w) = Bern_w[\lambda] \tag{2.23}$$

Σκοπός της διαδικασίας εκμάθησης είναι να προσδιοριστούν τα μ_0, Σ_0 για w=0. Ο προσδιορισμός των παραμέτρων γίνεται με την προσαρμογή της κανονικής πιθανοτικής κατανομής στα δεδομένα με μία από τις τεχνικές που περιγράφηκαν στην παράγραφο 2.4.

Για την ταξινόμηση ενός νέου δεδομένου x^* ως παρασκήνιο (w=0) ή προσκήνιο (w=1) υπολογίζεται από τον τύπο του Bayes, οπότε η μεταγενέστερη πιθανότητα υπολογίζεται ως εξής:

$$\Pr(w=1 \mid x^*) = \frac{\Pr(x^* \mid w=1)\Pr(w=1)}{\sum_{0}^{1} \Pr(x^* \mid w=k)\Pr(w=k)} \tag{2.24}$$

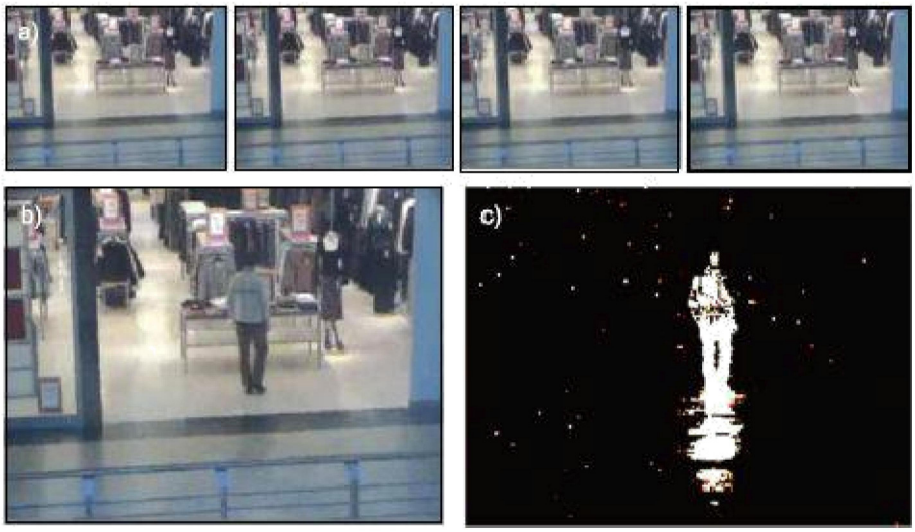
Όπου η ταξικά εξαρτώμενη κατανομή του προσκηνίου μοντελοποιείται με ομοιόμορφη κατανομή ως εξής:

$$\Pr(x^* \mid w=1) = \begin{cases} \dfrac{1}{255^3} & 0 < x^{*R}, x^{*G}, x^{*B} < 255 \\ 0 & \textit{διαφορετικά} \end{cases} \tag{2.25}$$

Οπότε το νέο pixel θεωρείται ότι ανήκει στο προσκήνιο όταν $\Pr(w=1 \mid x^*) > \varepsilon$ όπου ε το επιθυμητό ποσοστό, συνήθως 50% (Σχήμα 2.10).

2.6.2 Μοντελοποίηση σύνθετων κατανομών δεδομένων – Συνδιασμός Gaussian-ών

Στην προηγούμενη παράγραφο περιγράφηκε μοντέλο αφαίρεσης παρασκηνίου που ανήκει στις ταξικά-εξαρτώμενες συναρτήσεις Pr(**x**|w=k) πάνω στα παρατηρήσιμα δεδομένα **x** για κάθε τιμή της πραγματικής κατάστασης **w**. Στην πραγματικότητα όμως δεν είναι ρεαλιστικό να υποτεθεί ότι όλη η πολυπλοκότητα των οπτικών δεδομένων μπορεί να περιγραφεί από μία μόνο κανονική κατανομή καθώς έχει μία μόνο κορυφή και ούτε το προσκήνιο ούτε το παρασκήνιο περιγράφονται ικανοποιητικά από μία συνάρτηση πυκνότητας πιθανότητας με μία κορυφή. Επιπλέον η κανονική κατανομή δεν είναι εύρωστη, μία ακραία τιμή μπορεί να επηρεάσει την εκτίμηση του μέσου όρου και της συνδιακύμανσης. Τέλος η πολυπαραγοντική κανονική κατανομή που αφορά ένα πλήθος pixel έχει πίνακα συνδιακύμανσης με μεγάλο πλήθος παραμέτρων με αποτέλεσμα να χρειάζεται ένας μεγάλος αριθμός εκπαιδευτικών δειγμάτων για τον προσδιορισμό των παραμέτρων.



Σχήμα 2.10 Αφαίρεση παρασκηνίου. a) Εκμάθηση ταξικά εξαρτώμενου μοντέλου Pr(**x**|w) για το παρασκήνιο από εκπαιδευτικά δεδομένα άδειων σκηνικών. Το μοντέλο του προσκηνίου θεωρείται ομοιόμορφο b) Για νέα εικόνα υπολογίζεται η μεταγενέστερη κατανομή από τον κανόνα του Bayes. c) Μεταγενέστερη πιθανότητα ύπαρξης προσκηνίου Pr(w=1|**x**).

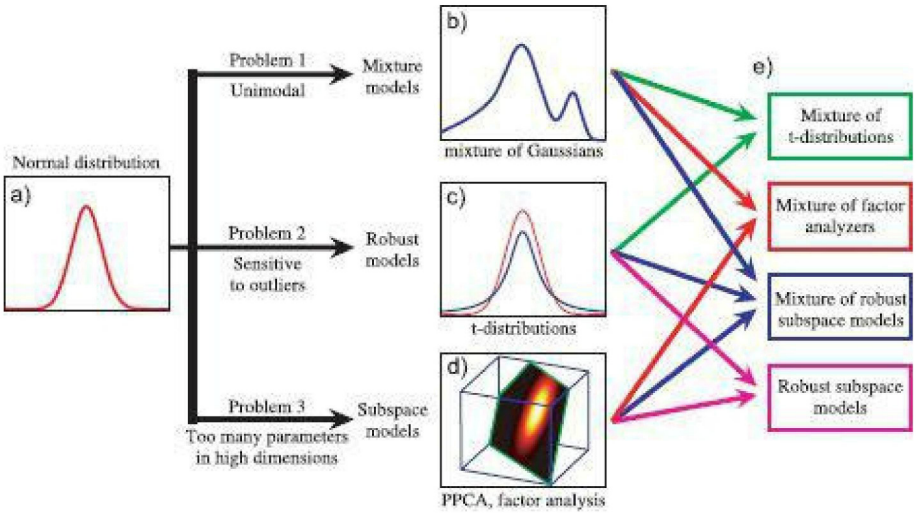
Για τον περιορισμό των αναφερθέντων προβλημάτων χρησιμοποιούνται τρεις μέθοδοι (Σχήμα 2.11), πρώτον ο συνδιασμός κανονικών κατανομών για τη δημιουργία κατανομής με πλήθος κορυφών, δεύτερον η αντικατάσταση της κανονικής κατανομής με μία t-κατανομή για την αύξηση της ευρωστίας και τρίτον μοντέλα του υποχώρου για την εκτίμηση των στοιχείων της συνδιακύμανσης σε μεγάλο αριθμό διαστάσεων.

Η μοντελοποίηση σύνθετων συναρτήσεων πυκνότητας πιθανότητας της μεταβλητής **x** γίνεται με την χρήση κρυμμένων μεταβλητών h_i όπου η μεταβλητή h_i μπορεί να είναι είτε συνεχής είτε διακριτή. Για την κατανόηση της κρυμμένης μεταβλητής, η τελική συνάρτηση πυκνότητας πιθανότητας Pr(**x**) (Σχήμα 2.12) περιγράφεται ως απομόνωση της από κοινού πυκνότητας πιθανότητας Pr(x,h) μεταξύ των x και h σύμφωνα με την ακόλουθη σχέση:

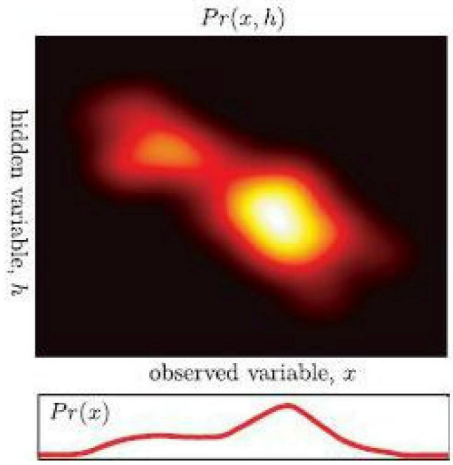
$$\Pr(x) = \int \Pr(x,h)dh \tag{2.26}$$

Η από κοινού συνάρτηση μπορεί να επιλεχθεί ώστε να είναι απλή στην μοντελοποιήση αλλά να προκύπτει ένα σύνολο χαρακτηριστικών κατανομών Pr(x) όταν γίνεται ολοκλήρωση προς x. Οποιαδήποτε μορφή και εάν χρησιμοποιηθεί για την από κοινού κατανομή θα έχει κάποιες παραμέτρους **θ** όπως φαίνεται και στην ακόλουθη εξίσωση:

$$\Pr(x \mid \theta) = \int \Pr(x,h \mid \theta)dh \tag{2.27}$$



Σχήμα 2.11 a)Προβλήματα με την πολυπαραγοντική κανονική κατανομή. b) Συνδιασμός κανονικών κατανομών c) t-κατανομές d) Μείωση παραμέτρων σε πολλές διαστάσεις με χρήση μοντέλων υποχώρου e) Συνδιασμός μεθόδων για δημιουργία υβριδικού μοντέλου



Σχήμα 2.12 Εφαρμογή μεθόδου κρυμμένων μεταβλητών h για μοντελοποίηση σύνθετων κατανομών. Η συνάρτση Pr(x) προκύπτει από την ολοκλήρωση της από κοινού κατανομής Pr(x,h).

Η προσαρμογή του μοντέλου στα εκπαιδευτικά δεδομένα $\{x_i\}_{i=1}^{I}$ γίνεται με την εφαρμογή της λογαριθμικής μεθόδου "μέγιστο ενδεχόμενο":

$$\hat{\theta} = \arg\max_{\theta}\left[L(\theta)\right] \tag{2.28}$$

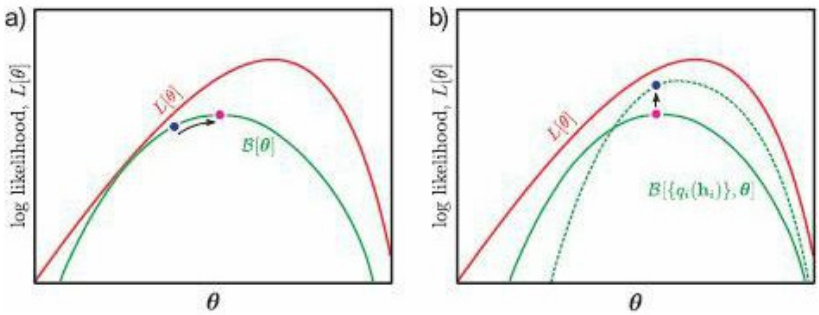
όπου $L(\theta)=\sum_{i=1}^{I}\log\left[\Pr(x_i \mid \theta)\right]$

Με την χρήση των κρυμένων μεταβλητών η σχέση (2.28) γίνεται:

$$\hat{\theta} = \arg\max_{\theta} \sum_{t=1}^{I} \log\left[\int \Pr(x_i, h_i \mid \theta) dh_i\right] \tag{2.29}$$

Στην συγκεκριμένη περίπτωση για την εύρεση των παραμέτρων $\hat{\theta}$ που μεγιστοποιούν τη σχέση (2.29) εφαρμόζεται ο αλγόριθμος γενικού σκοπού "μεγιστοποίηση εμπιστοσύνης" (expectation maximization), ο οποίος εντοπίζει με μεγαλύτερη ευκολία συγκεκριμένη λύση κλειστής μορφής καθώς χρησιμοποιεί τεχνική που μεταφέρει τη λογαριθμική πράξη της σχέσης (2.29) έξω από το ολοκλήρωμα. Ο αλγόριθμος υπολογίζει μία ακολουθία κατώτερων ορίων $\mathrm{B}\left[\{q_i(h_i)\}, \theta\right]$ τα οποία έιναι πάντα μικρότερα ή ίσα από τη λογαριθμική συνάρτηση του ενδεχόμενου L(θ) (Σχήμα 2.13). Το κατώτερο όριο παραμετροποιείται από τη μεταβλητή **θ** και από ένα σύνολο I πιθανοτικών κατανομών $\{q_i(h_i)\}_{i=1}^{I}$ πάνω στις κρυμένες μεταβλητές $\{h_i\}_{i=1}^{I}$. Η σχέση του κατώτερου ορίου δίνεται ως εξής:

$$\begin{aligned} \mathrm{B}\left[\{q_i(h_i)\}, \theta\right] &= \sum_{i=1}^{I} \int q_i(h_i) \log\left[\frac{\Pr(x_i, h_i \mid \theta)}{q_i(h_i)}\right] dh_i \\ &\leq \sum_{i=1}^{I} \log\left[\int \Pr(x_i, h_i \mid \theta) dh_i\right] \end{aligned} \tag{2.30}$$



Σχήμα 2.13 Συνάρτηση κατώτερου ορίου. a) Η λογαριθμική πιθανοτική σχέση ενδεχομένου L(θ) με δεδομένα τα $\{x_i\}_{i=1}^{I}$ φαίνεται στην κόκκινη καμπύλη. Η συνάρτηση B[θ], πράσινη καμπύλη, είναι παντού μικρότερη ή ίση της L[θ]. Ένας τρόπος για βελτίωση της τρέχουσας εκτίμησης, μπλε κουκίδα, είναι η διαχείριση των παραμέτρων θ ωστε να αυξηθεί η B[θ], ροζ κουκίδα. (M-βήμα) b) Το κατώτερο όριο $\mathrm{B}\left[\{q_i(h_i)\}, \theta\right]$ εξαρτάται επίσης από τις κατανομές $\{q_i(h_i)\}_{i=1}^{I}$ όπου η διαχείρισή τους αυξάνει την καμπύλη, μπλέ κουκίδα για τις τρέχουσες παραμέτρους **θ** (E-βήμα)

Ο αλγόριθμος μεγιστοποιήσης εμπιστοσύνης για να αυξήσει το κατώτερο όριο χειρίζεται με διαδοχικές εναλλαγές τις παραμέτρους **θ**, όπου το συγκεκριμένο βήμα ονομάζεται μεγιστοποίηση (maximization) ή M-βήμα και τις κατανομές $\{q_i(h_i)\}_{i=1}^{I}$, που ονομάζεται βήμα της εμπιστοσύνης (expectation step) ή E- βήμα. Ως βέλτιστη επιλογή η κατανομή $q_i(h_i)$ τίθεται ίση με τη μεταγενέστερη κατανομή πυκνότητας πιθανότητας $\Pr(h_i \mid x_i, \theta)$ της κρυμμένης μεταβλητής h_i με δοσμένα τα εκπαιδευτικά δεδομένα x_i και τις παραμέτρους **θ**.

Για παράδειγμα στο E-βήμα της t+1 επανάληψης ο υπολογισμός της $q_i(h_i)$ με εφαρμογή του κανόνα του Bayes γίνεται ως εξής:

$$\hat{q}_i(h_i) = \Pr(h_i \mid x_i, \theta^{[t]}) = \frac{\Pr(x_i \mid h_i, \theta^{[t]}) \Pr(h_i \mid \theta^{[t]})}{\Pr(x_i)} \tag{2.31}$$

Στο M- βήμα μεγιστοποιείται κατευθείαν το κατώτερο όριο (Σχέση 2.30) με διαχείριση των παραμέτρων **θ**. Η σχέση 2.30 εξαλείφωντας τους όρους που δεν εξαρτώνται από τις **θ**, οπότε προκύπτει η συνάρτηση :

$$\hat{\theta}^{[t+1]} = \arg\max_{\theta} \left[\sum_{\iota=1}^{\mathrm{I}} \int q_i(h_i) \log\left[\hat{\Pr}(x_i, h_i \mid \theta)\right] dh_i \right] \tag{2.32}$$

Καθένα από τα βήματα εξασφαλίζει την βελτίωση του κατώτερου ορίου και επαναλαμβάνοντάς τα εναλλάξ εξασφαλίζεται η εύρεση το λιγότερο ενός τοπικού μέγιστου με αναφορά το **θ**.

Μία σηματική εφαρμογή του μοντέλου που περιγράφηκε είναι ο συνδιασμός Gaussian-ών (mixture of Gaussians, MoG) που εφαρμόζεται για την αφαίρεση του παρασκηνίου, τεχνική που χρησιμοποιείται και στην παρούσα διπλωματική εργασία. Στο μοντέλο MoG το μέτρο της πυκνότητας πιθανότητας των παρατηρούμενων δεδομένων **x** δίνεται από σταθμισμένο άθροισμα K κανονικών κατανομών:

$$\Pr(x \mid \theta) = \sum_{\kappa=1}^{\mathrm{K}} \lambda_k Norm_x\left[\mu_k, \Sigma_k\right] \tag{2.33}$$

Όπου $\mu_{1...K}$ και $\Sigma_{1...K}$ είναι οι μέσες τιμές και οι συνδιακυμάνσεις των κανονικών κατανομών και $\lambda_{1...\mathrm{K}}$ είναι θετικες τιμές στάθμισης του αθροίσματος.

Η συνάρτηση (2.33) εκφράζεται με τη μορφή διαχωρίσιμης κοινής συνάρτησης των **x** με τις διακριτές κρυμένες μεταβλητές $h \in \{1...K\}$ σύμφωνα με τη σχέση 2.27 ως εξής:

$$\Pr(x \mid \theta) = \sum_{k=1}^{K} \Pr(x, h = k \mid \theta) \tag{2.34}$$

Όπου εύκολα αποδεικνύεται ορίζοντας τις συναρτήσεις:

$$\begin{aligned} \Pr(x \mid h, \theta) &= Norm_x[\mu_h, \Sigma_h] \\ \Pr(h \mid \theta) &= Cat_h[\lambda] \end{aligned} \tag{2.35}$$

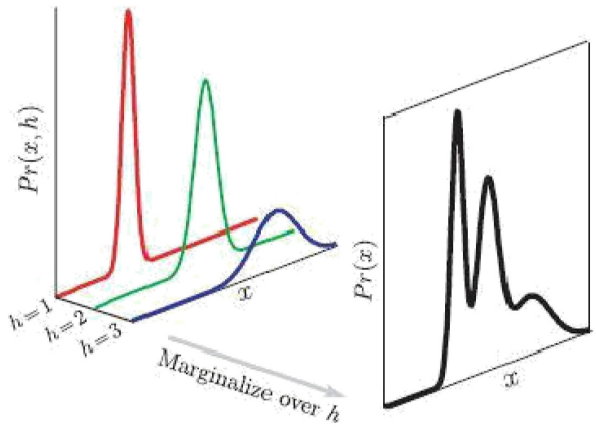
και σύμφωνα με τη σχέση 2.3, ότι:

$$\Pr(x \mid \theta) = \sum_{k=1}^{K} \Pr(x, h = k \mid \theta) \tag{2.36}$$

$$= \sum_{k=1}^{K} \Pr(x \mid h = k, \theta) \Pr(h = k \mid \theta)$$

$$= \sum_{k=1}^{K} \lambda_{\kappa} Norm_x[\mu_k, \Sigma_k]$$

Στη σχέση 2.36 ο παράγοντας Pr(h=k|θ) δίνει το μέτρο της πιθανότητας η k-οστή καταναμομή της h μεταβλητής (Σχήμα 2.14) να ταιριάζει με κάποια από τις παραμέτρους **θ** που έχουν υπολογιστεί από την μέθοδο μεγιστοποιήσης εμπιστοσύνης όπως θα περιγραφεί στη συνέχεια. Οπότε και ανάλογα ο συγκεκριμένος παράγοντας ενισχύει ή εξασθενεί την αντίστοιχη στοιχειώδη k-οστή κατανομή Pr(x|h=k,θ).



Σχήμα 2.14 Το μοντέλο συνδιασμού Gaussian-ών ως διαχωρισμός σε στοιχειώδεις κατανομές h_i

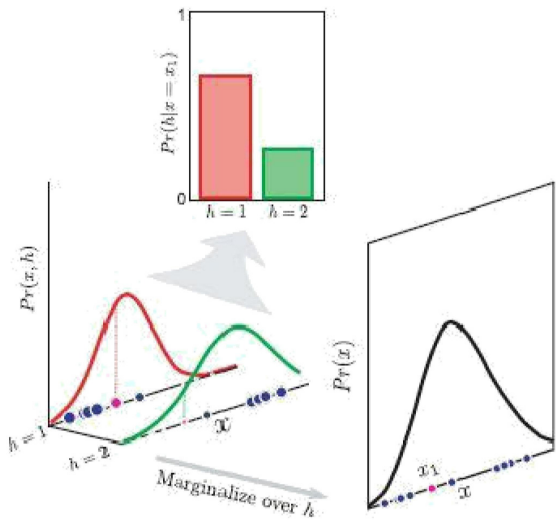
Για την εκμάθηση των παραμέτρων $\theta = \{\mu_k, \Sigma_k, \lambda_k\}_{k=1}^{K}$ από τα εκπαιδευτικά δεδομένα $\{x_i\}_{i=1}^{I}$ εφαρμόζεται ο αλγορίθμος μεγιστοποίησης εμπιστοσύνης. Η αρχικοποίηση των παραμέτρων γίνεται τυχαία και ακολουθούν επαναλαμβανόμενες εναλλαγές μεταξύ του Ε (Σχέση 2.31) και Μ βήματος (Σχέση 2.32). Στο βήμα Ε μεγιστοποιείται το όριο μέσω των κατανομών $q_i(h_i)$ οι οποίες εκφράζουν την μεταγενέστερη πιθανοτική κατανομή $\Pr(h_i \mid x_i)$ για κάθε κρυμμένη μεταβλητή h_i, δοσμένων των εκπαιδευτικών δεδομένων x_i και των τρεχόντων παραμέτρων **θ**. Ο τύπος για τον υπολογισμό του $q_i(h_i)$ είναι ο εξής:

$$q_i(h_i) = \Pr(h_i = k \mid x_i, \theta^{[t]}) = \frac{\Pr(x_i \mid h_i = k, \theta^{[t]}) \Pr(h_i = k, \theta^{[t]})}{\sum_{j=1}^{K} \Pr(x_i \mid h_i = j, \theta^{[t]}) \Pr(h_i = j, \theta^{[t]})} \tag{2.37}$$

$$= \frac{\lambda_{\kappa} Norm_{x_i}[\mu_k, \Sigma_k]}{\sum_{j=1}^{K} \lambda_j Norm_x[\mu_j, \Sigma_j]}$$

$$= r_{ik}$$

Υπολογίζεται δηλαδή η πιθανότητα $\Pr(h_i = k \mid x_i, \theta^{[t]})$ ότι η k-οστή κανονική κατανομή είναι υπεύθυνη για το i-οστό pixel (σχήμα 2.15). Ονομάζεται "υπευθυνότητα" r_{ik} εν συντομία.



Σχήμα 2.15 Ε-βήμα για προσαρμογή συνδιαστκού μοντέλου Gaussian-ών. Για καθένα δεδομένο $x_{1..I}$ υπολογίζεται η μεταγενέστερη κατανομή $\Pr(h_i \mid x_i)$ για κάθε μεταβλητή h_i. Για το δεδομένο x_i, πορφυρό σημείο, η στοιχειώδης κατανομή 1 (κόκκινη καμπύλη) έχει διπλάσιες πιθανότητες να είναι " υπεύθυνη" από ότι η στοιχειώδης κατανομή 2 (πράσινη καμπύλη).

Στο Μ-βήμα (Σχήμα 2.16) μεγιστοποιείται το όριο μέσω παραμέτρων του μοντέλου $\theta = \{\mu_k, \Sigma_k, \lambda_k\}_{k=1}^{K}$, σύμφωνα με τη σχέση 2.32, ώστε:

$$\hat{\theta}^{[t+1]} = \arg\max_{\theta}\left[\sum_{\iota=1}^{\mathrm{I}}\sum_{\kappa=1}^{\mathrm{K}} q_i(h_i = k)\log\left[\hat{\Pr}(x_i, h_i = k \mid \theta)\right]\right] \tag{2.38}$$

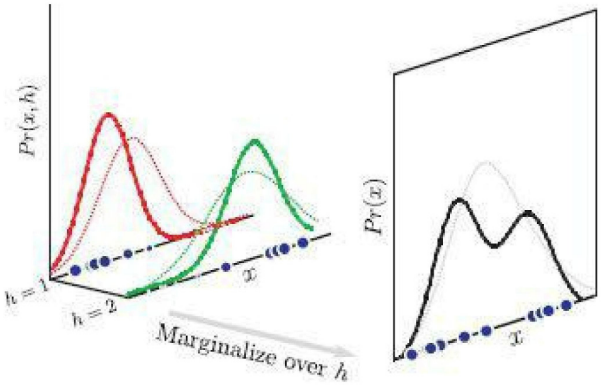
$$= \arg\max_{\theta}\left[\sum_{\iota=1}^{\mathrm{I}}\sum_{k=1}^{K} r_{ik}\log\left[\lambda_k Norm_{x_i}\left[\mu_k, \Sigma_k\right]\right]\right]$$

Η εύρεση του μεγίστου από τη σχέση 2.38 μπορεί να επιτευχθεί λαμβάνωντας τις μερικές παραγώγους τις οποίες εξισώνοντάς τις με μηδέν, αναδιατάσσοντάς τες και διασφαλίζοντας τον περιορισμό $\sum_k \lambda_k = 1$ με την χρήση πολλαπλασιατών Lagrange προκύπτουν οι επαναληπτικές συναρτήσεις:

$$\lambda_k^{[t+1]} = \frac{\sum_{i=1}^{I} r_{ik}}{\sum_{j=1}^{K}\sum_{i=1}^{I} r_{ij}} \tag{2.39}$$

$$\mu_k^{[t+1]} = \frac{\sum_{i=1}^{I} r_{ik} x_i}{\sum_{i=1}^{I} r_{ik}}$$

$$\Sigma_k^{[t+1]} = \frac{\sum_{i=1}^{I} r_{ik}(x_i - \mu_k^{[t+1]})(x_i - \mu_k^{[t+1]})^T}{\sum_{i=1}^{I} r_{ik}}$$



Σχήμα 2.16 Μ-βήμα για προσαρμογή συνδιαστκού μοντέλου Gaussian-ών. Για την k-οστή στοιχειώδη κανονική κατανομή ανανεώνονται οι παράμετροι $\{\mu_k, \Sigma_k, \lambda_k\}$. Το i-οστό σημείο x_i συνεισφέρει σε αυτές τις ανανεώσεις αναλογικά με την "υπευθυνότητά" του r_{ik}. Τα σημεία που συσχετίζονται περισσότερο με τη k-οστή στοιχειώδη κατανομή έχουν μεγαλύτερη επίδραση στις παραμέτρους.

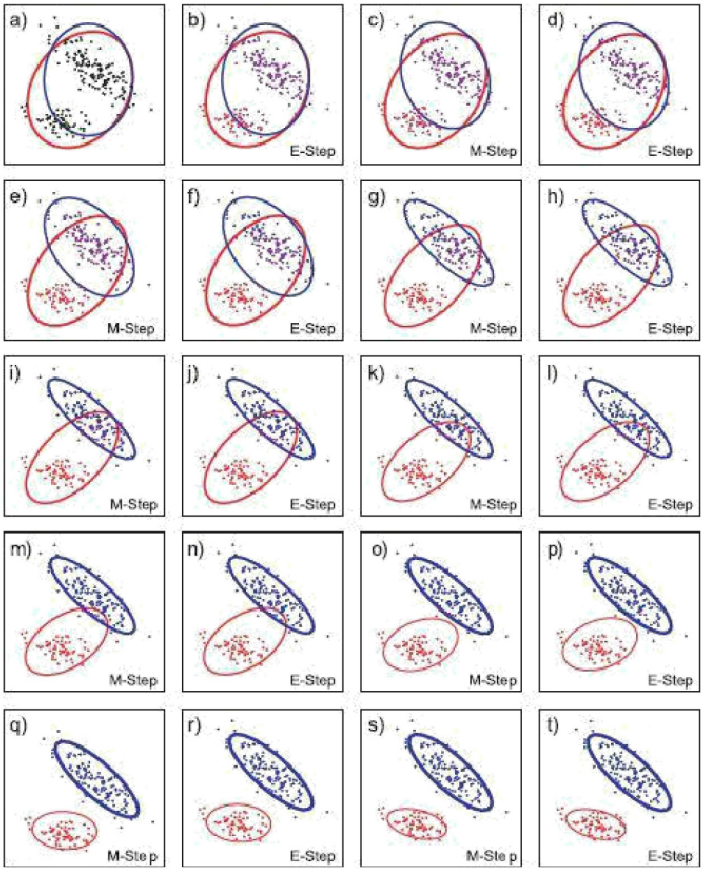
Στην πράξη τα Ε και Μ βήματα εναλλάσονται μέχρι το όριο να μην αυξάνεται άλλο και να μην αλλάζουν οι παράμετροι. Η εναλλαγή Ε και Μ βημάτων για δυσδιάστατο δυσδιάστατα δεδομένα παρουσιάζεται στο σχήμα 2.17.

Η εφαρμογή του αλγορίθμου μεγιστοποίησης εμπιστοσύνης για την εκτίμηση του συνδιαστικού μοντέλου κανονικών κατανομών έχει δύο ελκυστικά χαρακτηριστικά:

1) Και τα δύο βήματα εξίσου του αλγορίθμου μπορούν να υπολογιστούν σε κλειστή μορφή χωρίς την ανάγκη χρήσης διαδικασίας βελτιστοποίησης.
2) Η επίλυση εγγυάται ότι οι περιορισμοί των παραμέτρων διασφαλίζονται, δηλαδή η στάθμιση $\{\lambda_\kappa\}_{k=1}^{K}$ είναι θετική και αθροιστικά ισούται με ένα καθώς και οι μήτρες συνδιακύμανσης είναι θετικά ορισμένες.

Ο αλγόριθμος μεγιστοποίησης εμπιστοσύνης δεν εγγυάται όμως την εύρεση καθολικής λύσης στην προσαρμογή των παραμέτρων οπότε ένας τρόπος περιορισμού του συγκεκριμένου προβλήματος είναι η εκκίνηση σου αλγορίθμου από διαφορετικές αρχικές τιμές επιλέγοντας τελικά αυτή που η λογαριθμική συνάρτηση ενδεχομένου θα έχει τη μεγαλύτερη τιμή.

Η διαδικασία αφαίρεσης παρασκηνίου μοντελοποιείται απο τη μέθοδο συνδιασμού Gaussian-ών κατανομών με δύο κλάσεις ταξινόμησης της πραγματικής κατάστασης όπως ακριβώς περιγράφηκε στο υποκεφάλαιο 2.6.1. Η επαγωγή συμπεράσματος στη μέθοδο συνδιασμού Gaussian-ών ακολουθεί την ίδια διαδικασία του υποκεφαλαίου 2.6.1 με τη διαφορά ότι στην διαδικασία της εκμάθησης η συνάρτηση του ενδεχομένου Pr(x|w,θ) του παρασκηνίου (w=0) περιγράφεται αντί της σχέσης 2.22 από την 2.33 και οι παράμετροί της $\{\mu_k, \Sigma_k, \lambda_k\}$ εκτιμώνται από τη μέθοδο μεγιστοποίησης εμπιστοσύνης.



Σχήμα 2.17 Προσαρμογή συνδιαστικού μοντέλου δύο Gaussian-ών σε δυσδιάστατα δεδομένα. a) Αρχικό μοντέλο b) E-βήμα. Για κάθε σημείο υπολογίζεται η μεταγενέστερη πιθανότητα που προκύπτει από κάθε Gaussian-ή (υποδεικνύεται από το χρώμα) c) M-βήμα . Ο μέσος όρος , η διακύμανση και η στάθμιση κάθε Gaussian-ής ανανεώνεται βασιζόμενη στις μεταγενέστερες πιθανότητες. d-t) Περαιτέρω επαναλήψεις E- M βημάτων.

2.7 ΨΗΦΙΑΚΗ ΕΠΕΞΕΡΓΑΣΙΑ ΕΙΚΟΝΑΣ

Πριν την εφαρμογή μοντέλου μηχανικής μάθησης, είτε διακριτικού είτε γεννητικού που συσχετίζει τις οπτικές μετρήσεις **x** με τις πραγματικές κατάσταση **w** είναι, σχεδόν πάντα, απαραίτητη η προεπεξεργασία των pixel της εικόνας με στόχο να διαμορφωθούν τα κατάλληλα διανύσματα μέτρησης **x**. Οι παράμετροι των μετασχηματισμών στα πλαίσια της προεπεξεργασίας δεν "μαθαίνονται" από εκπαιδευτικά δεδομένα αλλά επιλέγονται με βάση την εμπειρία του "τι δουλεύει καλά". Για παράδειγμα σε μία εφαρμογή εντοπισμού αντικειμένου οι RGB τιμές επιρεάζονται από το κέρδος της κάμερας, το φωτισμό και τη στάση του αντικειμένου, οπότε σκοπός της προεπεξεργασίας είναι να εξαλειφθούν αυτές οι ανεπιθύμητες μεταβολές και να διατηρηθούν όσο το δυνατόν καλύτερα τα στοιχεία που είναι απαραίτητα για την εξαγωγή συμπερασμάτων.

Η προεπεξεργασία μπορεί να γίνει με μεθόδους που αντικαθηστούν την τιμή κάθε pixel της εικόνας με νέα τιμή και με μεθόδους εύρεσης και χαρακτηρισμού άκρων, γωνιών και σημείων ενδιαφέροντος.

2.7.1 Γραμμικό φιλτράρισμα στοιχείων εικόνας

Σημαντική ανά pixel μέθοδος επεξεργασίας είναι το γραμμικό φιλτράρισμα της εικόνας, μέθοδος που χρησιμοποιείται και στην παρούσα διπλωματική. Για την κατανόηση του γραμμικού φιλτραρίσματος θεωρείται ένας αρχικός δυσδιάστατος πίνακας με pixel που συμβολίζεται ως **P** και κάθε pixel χαρακτηρίζεται από μία τιμή p_{ij} που δείχνει την ένταση του σε κλίμακα του γκρι. Οπότε με τη διαδικασία του φιλτραρίσματος επιστρέφεται ένας νέος δυσδιάστατος πίνακας **X** του ίδιου μεγέθους όπου η νέα τιμή x_{ij} κάθε pixel ισούται με ένα σταθμισμένο άθροισμα των εντάσεων του ίδιου και γειτονικών pixel από την αρχική εικόνα **P**. Οι σταθμιστικοί παράγοντες $f_{m,n}$ είναι αποθηκευμένοι σε ένα φίλτρο Kernel **F**. Ποιό συγκεκριμένα η εφαρμογή ενός φίλτου είναι η συνέλιξη του πίνακα **P** με το φίλτρο **F**, όπου η δυσδιάστατη συνέλιξη ορίζεται ως εξής:

$$x_{ij} = \sum_{m=-M}^{M} \sum_{n=-N}^{N} p_{i-m,j-n} f_{m,n} \tag{2.40}$$

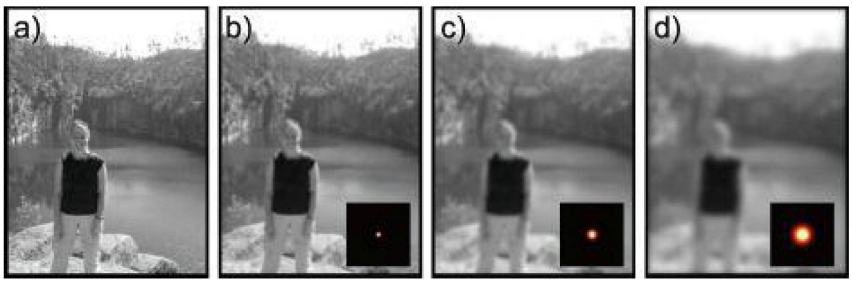
Για τη διαδικασία θόλωσης μίας εικόνας χρησιμοποιείται ένα δυσδιάστατο Gaussian-ό φίλτρο που έχει τη μορφή της εξίσωσης (2.40). Κάθε προκύπτον pixel είναι ένα σταθμισμένο άθροισμα των γειτονικών pixel και οι σταθμιστικοί παράγοντες εξαρτώνται από το προφίλ του Gaussin-ού φίλτρου (Σχέση 2.41). Η θόλωση της εικόνας εξαρτάται από την απόκλιση σ του Gaussian-ού φίλτρου (Σχήμα 2.18). Πρόκειται για μία απλή μέθοδο μείωσης του θορύβου σε εικόνες που ελήφθησαν με ελλειπή φωτισμό.

$$f(m,n) = \frac{1}{2\pi\sigma^2} \exp\left[-\frac{m^2+n^2}{2\sigma^2}\right] \tag{2.41}$$

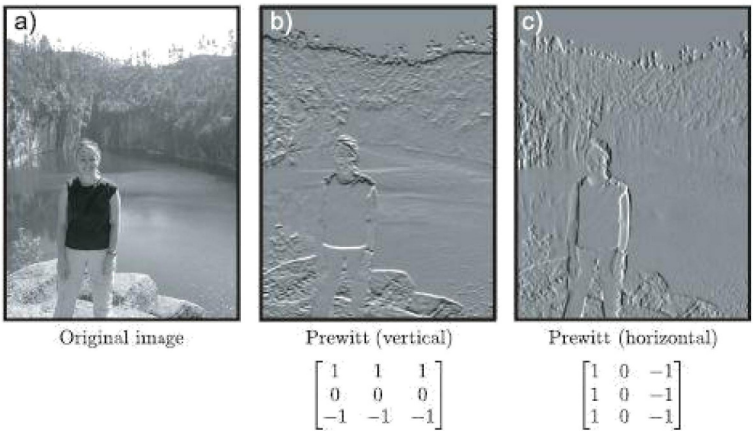
Μία δεύτερη χρήση του φιλτραρίσματος εικόνας είναι ο εντοπισμός περιοχών της εικόνας όπου η ένταση των pixel αλλάζει απότομα. Μία προσέγγιση της συγκεκριμένης λειτουργίας γίνεται υπολογίζοντας τη διαφορά μεταξύ των τιμών δύο διαδοχικών pixel κατα μήκος μίας γραμμής. Αυτό μπορεί να επιτευχθεί με χρήση του φίλτρου $F = [-101]$ η συνέλιξη με το οποίο δίνει μηδενική απόκριση όταν η εικόνα είναι ισοεπίπεδη κατά την οριζόντια διεύθυνση, αρνητική απόκριση όταν οι τιμές των pixel αυξάνονται κατά την οριζόντια διεύθυνση και θετική απόκριση όταν οι τιμές των pixel μειώνονται κατά την οριζόντια διεύθυνση (η συνέλιξη στρίβει το φίλτρο κατά $180°$). Η απόκριση του φίλτρου $F = [-101]$ είναι θορυβώδης λόγο της περιορισμένης χωρικής του έκτασης οπότε στην πράξη για τον εντοπισμό άκρων χρησιμοποιούνται πιο εκλεπτισμένα φίλτρα όπως τα φίλτρα Prewitt τα οποία είναι της μορφής:

$$F_x = \begin{bmatrix} 1 & 0 & -1 \\ 1 & 0 & -1 \\ 1 & 0 & -1 \end{bmatrix}, \; F_y = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ -1 & -1 & -1 \end{bmatrix} \tag{2.42}$$

Όπου F_x φίλτρο με επιλεκτικότητα ως προς τα άκρα στην οριζόντια διεύθυνση και F_y φίλτρο με επιλεκτικότητα ως προς τα άκρα στην κατακόρυφη διεύθυνση (Σχήμα 2.19).



Σχήμα 2.18 Θόλωση εικόνας a) Αρχική εικόνα b) Συνέλιξη αρχικής εικόνας με φίλτρο Gauss που φαίνεται κάτω δεξιά στην εικόνα. c-d) Συνέλιξη εικόνας με φίλτρα με μεγαλύτερη απόκλιση σ.



Σχήμα 2.19 Φιλτράρισμα εικόνας με πρώτης τάξης παράγωγο. a) Αρχική εικόνα b) Συνέλιξη με κάθετο φίλτρο Perwitt c) Συνέλιξη με οριζόντιο φίλτρο Perwitt.

2.7.2 Εντοπισμός άκρων – Μέθοδος Canny

Στη διαδικασία ανίχνευσης άκρων σκοπός είναι η δημιουργία δυαδικής εικόνας όπου οι μη μηδενικές τιμές δηλώνουν την παρουσία άκρων στην εικόνα. Μια μέθοδος για ανίχνευση άκρων είναι η μέθοδος Canny κατά την οποία η εικόνα **P** πρώτα θολώνεται και στη συνέχεια ακολουθείται η διαδικασία της συνέλιξης με φίλτρα όπως τα Prewitt ώστε να δημιουργηθούν εικόνες **H** και **V** που είναι οι παράγωγοι σε οριζόντια και κάθετη διεύθυνση αντίστοιχα. Για ένα pixel στη θέση (i,j) της εικόνας η κατεύθυνση της κλίσης θ_{ij} και το μέτρο a_{ij} της κλίσης υπολογίζεται από τους τύπους:

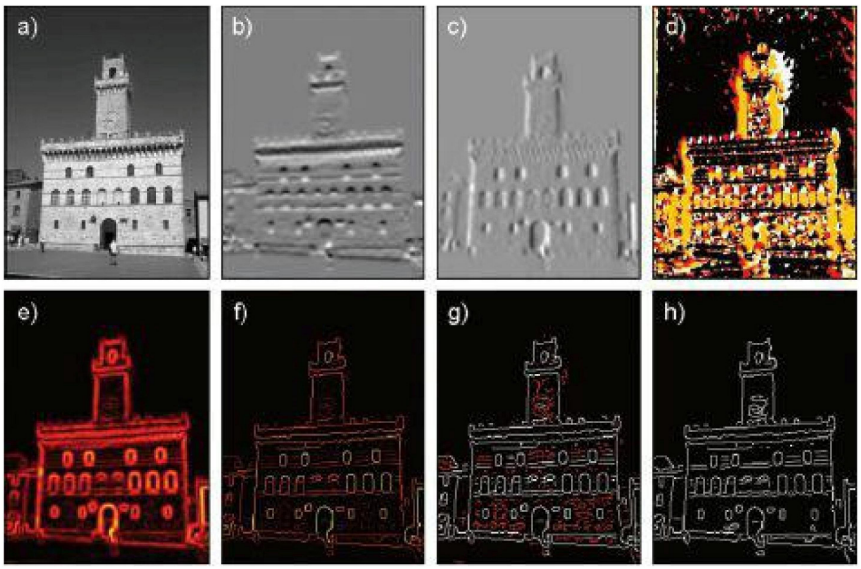
$$\theta_{ij} = \arctan\lfloor v_{ij} / h_{ij} \rfloor \qquad (2.43)$$

$$a_{ij} = \sqrt{h_{ij}^2 + v_{ij}^2}$$

Μία απλή προσέγγιση θα ήταν να καταχωρείται ένα σημείο στη θέση (i,j) ως σημείο ακμής εάν το πλάτος του σύμφωνα με τη σχέση (2.43) υπερβαίνει μία κρίσιμη τιμή. Η μέθοδος αυτή δίνει όμως φτωχά αποτελέσματα καθώς ο τύπος του πλάτους παίρνει μεγάλες τιμές όχι μόνο στις άκρες αλλά και στις γειτονικές τους περιοχές.

Η μέθοδος ανίχνευσης άκρων Canny εξαλείφει τις ανεπιθύμητες αποκρίσεις χρησιμοποιώντας τη μέθοδο "κατάργηση μη μεγίστων". Στη μέθοδο "κατάργηση μη μεγίστων» η κατεύθυνση της κλίσης κβαντοποιείται σε τέσσερεις τιμές $\{0^\circ, 45^\circ, 90^\circ, 135^\circ\}$. Τα pixel που σχετίζονται με κάθε κβαντισμένη τιμή γωνίας διαχειρίζονται ξεχωριστά. Στη συνέχεια η τιμή του μέτρου της κλίσης σε μία θέση (i,j) τίθεται μηδέν εάν τα γειτονικά pixel που βρίσκονται κάθετα στην κατεύθυνση της κλίσης του έχουν μεγαλύτερη τιμή μέτρου κλίσης. Με τον τρόπο αυτό οι κλίσεις με το μέγιστο μέτρο διατηρούνται ενώ αυτές που έχουν χαμηλότερες τιμές μέτρου καταστέλλονται.

Στη συνέχεια μία δυαδική εικόνα άκρων (Σχήμα 2.20) δύναται τώρα να δημιουργηθεί συγκρίνοντας τα εναπομείναντα μέτρα κλίσεων με δύο τιμές κατωφλίου. Χρησιμοποιούνται δύο τιμές για να μειωθούν οι αστοχίες. Όλα τα pixel των οποίων το μέτρο της κλίσης τους είναι πάνω από το μέγιστο κατώφλι επισημαίνονται ως άκρη. Τα pixel που το μέτρο τους είναι πάνω από το χαμηλότερο κατώφλι και κάτω από το μέγιστο και επιπλέον συνδέονται με pixel που ανήκει σε άκρη τότε επισημαίνονται και αυτά ως άκρη. Η τεχνική αυτή ονομάζεται υστέρηση κατωφλίωσης.

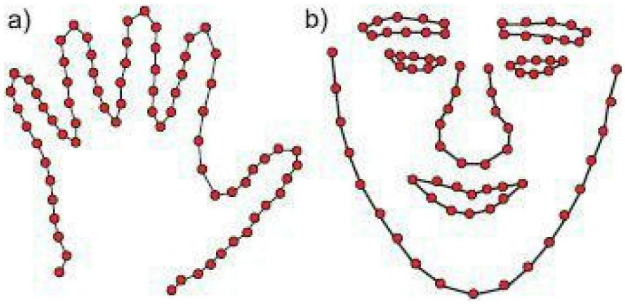


Σχήμα 2.20 Ανίχνευση άκρων με τη μέθοδο Canny. a) Αρχική εικόνα b) Αποτέλεσμα μετά από συνέλιξη με κάθετο φίλτρο Prewitt c) Αποτέλεσμα μετά από συνέλιξη με οριζόντιο φίλτρο Prewitt d) Κβαντισμένος πίνακας κατευθύνσεων θ_{ij} e) Πίνακας μέτρου κλήσεων a_{ij} f) Αποτέλεσμα μετά την εφαρμογή της μεθόδου "κατάργηση μη μεγίστων". g) Εφαρμογή κατοφλίων σε δύο επίπεδα, τα λευκά pixel είναι κάτω από το μεγαλύτερο κατώφλι και τα κόκκινα ανάμεσα στο μέγιστο και ελάχιστο κατώφλι. h) Τελικός πίνακας άκρων μετά την εφαρμογή της διαδικασίας υστέρησης

2.8 ΠΡΟΤΥΠΟ ΕΝΕΡΓΟΥ ΠΕΡΙΓΡΑΜΜΑΤΟΣ

Στην παρούσα παράγραφο αναλύεται μία γενική μέθοδος εντοπισμού περιγράμματος αντικειμένων σε μία εικόνα. Η γενική αυτή μέθοδος είναι η βάση του μοντέλου που χρησιμοποιείται στην παρούσα διπλωματική για την περιγραφή περιγραμμάτων. Ο εντοπισμός του εξωτερικού περιγράμματος παρουσιάζει ιδιαίτερη χρησιμότητα καθώς βοηθάει στην περιγραφή του σχήματος, στον εντοπισμό και στην τμηματοποίηση αντικειμένων μέσα σε μία εικόνα. Για την αναπαράσταση του περιγράμματος χρησιμοποιείται ένα σύνολο διακριτικών σημείων (Σχήμα 2.21), τα οποία μπορεί να σχηματίζουν είτε συνεχόμενη γραμμή είτε να είναι διακριτά μεταξύ τους.

Η μέθοδος ονομάζεται παραμετρικό μοντέλο περιγράμματος το οποίο αρκετές φορές αναφέρεται ως ενεργό μοντέλο περιγράμματος (active contour model) και ως “snakes”. Στο μοντέλο αυτό παρέχεται ελάχιστη προγενέστερη γεωμετρική πληροφορία για το περίγραμμα δηλαδή μόνο ότι η μορφή του περιγράμματος είναι κλειστή ή ανοιχτή και ομαλή. Οπότε είναι κατάλληλη για περιπτώσεις όπου λίγα είναι γνωστά για το περιεχόμενο της εικόνας.



Σχήμα 2.21 Διακριτικά σημεία για αναπαράσταση περιγράμματος. a) Τα διακριτικά σημεία ορίζουν ένα μονό ανοιχτό περίγραμμα που περιγράφει σχήμα χεριού. b) Τα διακριτικά σημεία ενώνονται σε διαφορετικά σύνολα ανοιχτών και κλειστών περιγραμμάτων που περιγράφουν περιοχές προσώπου.

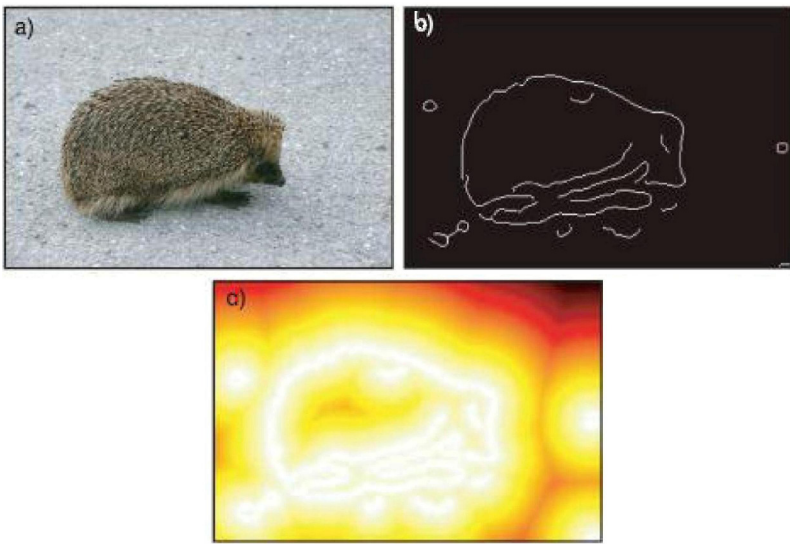
Λαμβάνεται ότι ένα κλειστό περίγραμμα ορίζεται από ένα σύνολο N δυσδιάστατων διακριτικών σημείων **W**=[w1,w2…wn], τα οποία είναι οι άγνωστοι του μοντέλου. Στόχος είναι να βρεθεί η διάταξη των διακριτικών αυτών σημείων που να εξηγεί καλύτερα το περίγραμμα ενός αντικειμένου της εικόνας. Για την εύρεση των σημείων αυτών κατασκευάζεται ένα γεννητικό μοντέλο το οποίο προσδιορίζεται από τη συνάρτηση ενδεχομένου Pr(**x**|**W**), η οποία υποδεικνύει τη συμφωνία της διάταξης των διακριτικών σημείων **W** με τα δεδομένα της εικόνας καθώς και τη συνάρτηση προγενέστερης πιθανότητας Pr(**w**), η οποία περικλύει την προγενέστερη πληροφόρηση για τη συχνότητα που που υπάρχει συμφωνία μεταξύ διακριτικών σημείων και δεδομένων εικόνας.

Στο συγκεκριμένο γεννητικό μοντέλο η συνάρτηση ενδεχομένου Pr(**x**|**W**) πάνω στα δεδομένα **x** της εικόνας δοσμένων των διακριτικών σημείων **W**, δεν χρησιμοποιεί εκπαιδευτικά δεδομένα για προσαρμογή των παραμέτρων της κατανομής όπως αναφέρθηκε στο κεφάλαιο 2.5 αλλά σκοπός της είναι να επιστρέφει τιμή η οποία είναι υψηλή όταν τα διακριρικά σημεία **W** βρίσκονται πάνω ή κοντά σε άκρα (edges) που περιέχονται στην εικόνα και μικρή όταν βρίσκονται σε ισόπεδες περιοχές της εικόνας.

Ένα μοντέλο της συνάρτησης ενδεχομένου προκύπτει χρησιμοποιώτας πρώτα τη μέθοδος ανίχνευσης άκρων Canny στην εικόνα να εφαρμοστεί στη συνέχεια ένας μετασχηματιμός απόστασης όπου σε κάθε pixel να δίνεται μία τιμή σε σχέση με την τετραγωνική απόσταση του από το κοντινότερο pixel άκρου. Το μοντέλο περιγράφεται από την ακόλουθη σχέση:

$$\Pr(x \mid W) \propto \prod_{n=1}^{N} \exp[-(dist[x, w_n])^2] \tag{2.44}$$

Όπου η συνάρτηση $dist[x, w_n]^2$ επιστρέφει την τιμή του μετασχηματισμού απόστασης στην θέση w της εικόνας. Η τιμή που επιστρέφεται από τη σχέση 2.44 είναι μεγάλη όταν όλα τα διακριτικά σημεία πέφτουν κοντά στις άκρες που περιέχονται στην εικόνα (όπου η μετασχηματισμένη απόσταση δίνει χαμηλές τιμές) και μειώνεται ομαλά καθώς η απόσταση μεταξύ των διακριτικών σημείων και των άκρων αυξάνεται (Σχήμα 2.22). Αυτή η προσέγγιση έχει επίσης μία ελκυστική ερμηνεία, η αντικειμενική συνάρτηση της "τετραγωνική απόστασης" είναι ισοδύναμη με το να υποτεθεί ότι η προσεγκιστική μέτρηση της θέσης του άκρου είναι μία θορυβώδης εκτίμηση της πραγματικής θέσης του και ο θόρυβος είναι πρόσθετος και κανονικά κατανεμημένος.



Σχήμα 2.22 Ενδεχόμενο για διακριτικά σημεία. a) Αρχική εικόνα b) Αποτέλεσμα εφαρμογής μεθόδου αντοπισμού άκρων Canny c) Αρνητική απόσταση από κοντινότερο άκρο.

Εάν η εύρεση των διακριτικών σημείων **W** στηριζόταν μόνο στη σχέση 2.44, τότε κάθε σημείο ξεχωριστά θα έλκοταν από ένα έντονο άκρο και το αποτέλεσμα πολύ πιθανόν να μην ήταν ένα συνεκτικό σχήμα, μπορεί ακόμα όλα τα διακριτικά σημεία να συνέκληναν στην ίδια θέση. Για να αποφευχθεί αυτό το πρόβλημα καθορίζεται η συνάρτηση προγενέστερης πιθανότητας Pr(W) να ευνοεί τα ομαλά περιγράμματα με μικρή καμπυλότητα. Ένας τρόπος υλοποίησης είναι η επιλογή της συνάρτησης Pr(W) με δύο όρους:

$$\Pr(W) \propto \prod_{n=1}^{N} \exp[a \cdot space[w, n] + \beta \cdot curve[w, n]] \tag{2.45}$$

Όπου οι βαθμωτές μεταβλητές α και β ελέγχουν την σχετική συνεισφορά αυτών των όρων.

Ο πρώτος όρος space[w,n] ενθαρρύνει τα σημεία γύρω από το περίγραμμα να ισαπέχουν μεταξύ τους και επιστρέφει μεγάλη τιμή εάν η απόσταση μεταξύ του n-οστού σημείου του περιγράμματος wn και των γειτονικών του είναι κοντά στη μέση τιμή απόστασης ανάμεσα στα γειτονικά σημεία κατά μήκος του περιγράμματος. Η συνάρτηση περιγράφεται από την ακόλουθη σχέση όπου το περίγραμμα θεωρείται κλειστής μορφής με w0=wN:

$$space[w,n] = -\left(\frac{\sum_{n=1}^{N} \sqrt{(w_n - w_{n-1})^T (w_n - w_{n-1})}}{N} - \sqrt{(w_n - w_{n-1})^T (w_n - w_{n-1})} \right)^2 \quad (2.46)$$

Ο δεύτερος όρος της συνάρτησης curve[w,n] επιστρέφει μεγάλες τιμές όταν η καμπυλότητα είναι μικρή ώστε να προτρέπει το περίγραμμα να είναι ομαλό. Ορίζεται ως εξής:

$$curve[w,n] = -(w_{n-1} - 2w_n + w_{n+1})^T (w_{n-1} - 2w_n + w_{n+1}) \quad (2.47)$$

Υπάρχουν μόνο δύο παράμετροι σε αυτό το μοντέλο, τα βάρη α και β. Αυτοί μπορεί είτε να "μαθευτούν" από δείγματα προς εκπαίδευση είτε να επιλεχθούν από το χρήστη ώστε να αποφευχθεί μία ειδική διαδικασία εκμάθησης.

Η επαγωγή συμπεράσματος από μία καινούρια εικόνα **x** συνίσταται στην προσαρμογή των σημείων {wi} στο περίγραμμα ώστε να περιγράφουν την εικόνα όσο το δυνατόν καλύτερα. Για το σκοπο αυτό χρησιμοποιείται η μέθοδος προσαρμογής "μεταγενέστερο μέγιστο" ως εξής:

$$\begin{aligned} \hat{W} &= \arg\max_W \left[\Pr(W \mid x)\right] \\ &= \arg\max_W \left[\Pr(x \mid W)\Pr(W)\right] \\ &= \arg\max_W \left[\log\left[\Pr(x \mid W)\right] + \log\left[\Pr(W)\right]\right] \end{aligned} \quad (2.48)$$

Το βέλτιστο της από πάνω αντικειμενικής συνάρτησης δεν μπορεί να βρεθεί σε κλειστή μορφή οπότε εφαρμόζεται μία μη γραμμική διαδικασία βελτιστοποίησης όπως η μέθοδος Newton. Ο αριθμός των αγνώστων είναι διπλάσιος από τον αριθμό των διακριτικών σημείων καθώς κάθε σημείο έχει ένα x και y ως συντεταγμένες.

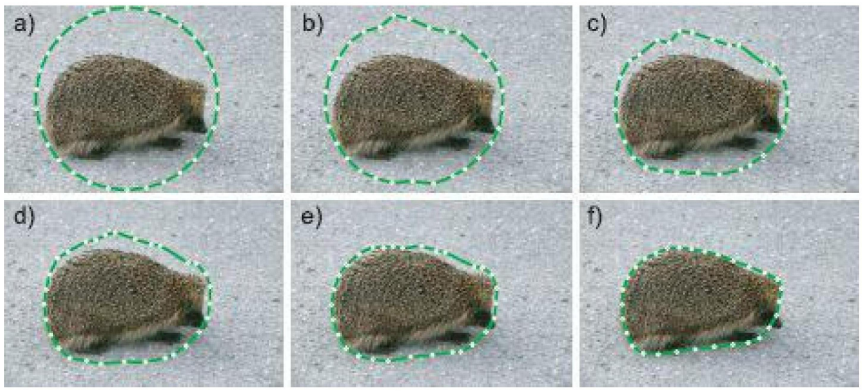
Ένα παράδειγμα της διαδικασίας παρουσιάζεται στο Σχήμα 2.23.Καθώς η ελαχιστοποίηση προχωράει το περίγραμμα κινείται αργά γύρω από την εικόνα αναζητώντας το σύνολο των διακριτικών σημείων με τη μέγιστη μεταγενέστερη πιθανότητα. Γι αυτό το λόγο αυτό το είδος του μοντέλου είναι γνωστό ως "snake" (φίδι) ή ενεργό μοντέλο περιγράμματος.

Το τελικό περίγραμμα του σχήματος έχει σφιχτή εφαρμογή στα εξωτερικά όρια του σκατζόχοιρου. Παρόλα αυτά δεν υπάρχει σωστή οριοθέτηση στην περιοχή της μύτης, η οποία είναι μία περιοχή μεγάλης καμπυλότητας και το μέγεθος της μύτης είναι μικρότερο από την απόσταση μεταξύ των προσαρμοσμένων διακριτικών σημείων. Το πρότυπο μπορεί να βελτειωθεί με την χρήση συνάρτησης ενδεχομένου η οποία να εξαρτάται από θέσεις κατά μήκος του περιγράμματος ανάμεσα στα διακριτικά σημεία. Με αυτό τον τρόπο μπορεί να αναπτυχθεί ένα πρότυπο το οποίο έχει λίγους αγνώστους παραπάνω αλλά να εξαρτάται από την εικόνα με πιό συνεκτικό τρόπο.

Ένα δεύτερο πρόβλημα με το συγκεκριμένο μοντέλο είναι ότι η διαδικασία βελτιστοποίησης μπορεί να κολλήσει σε κάποιο τοπικό μέγιστο. Ένας πιθανός τρόπος αντιμετώπισης του προβλήματος είναι η επανεκκίννηση της διαδικασίας βελτιστοποιήσης από ένα πλήθος διαφορετικών αρχικών συνθηκών και η τελική λύση να είναι αυτή με την μεγαλύτερη τιμή από τη σχέση 2.48. Εναλλακτικά είναι δυνατόν να τροποποιηθεί η συνάρτηση της προγενέστερης πληροφορίας ώστε να γίνει η επαγωγή συμπεράσματος πιο εύκολη, για παράδειγμα η συνάρτηση space μπορεί να ξαναοριστεί ως εξής:

$$space[w,n] = -\left(\mu_S - \sqrt{\left(w_n - w_{n-1}\right)^T \left(w_n - w_{n-1}\right)}\right)^2 \tag{2.49}$$

Τύπος ο οποίος ενθαρρύνει την απόσταση μεταξύ των σημείων να είναι κοντά σε μία προκαθορισμένη τιμή μs οπότε και να βρίσκονται λύσεις σε καθορισμένη κλίμακα.



Σχήμα 2.23 Ενεργό μοντέλο περιγράμματος. a-f) Καθώς η διαδικασία βελτιστοποίησης προχωράει και η επιστρεφόμενη τιμή μεταγενέστερης πιθανότητας αυξάνεται το ενεργό περίγραμμα σύρεται κατά μήκος της εικόνας.

2.9 ΒΙΒΛΙΟΓΡΑΦΙΑ

[2.1] S.J.D. Prince, *Computer Vision Models, Learning, and Inference*. Cambridge University Press, 2012.

[2.2] C.M. Bishop, *Pattern Recognition and Machine Learning*. Springer Science, 2006.

[2.3] Π.Δελλαπόρτας–Π.Τσιαμυρτζής, *Στατιστική κατά Bayes.* Τμήμα Στατιστικής, Οικονομικό Πανεπιστήμιο Αθηνών, 2004.

[2.4] Δικτυακός Ιστότοπος *https://en.wikipedia.org/wiki/Active_contour_model*

ΚΕΦΑΛΑΙΟ 3

ΒΙΒΛΙΟΘΗΚΗ OpenCV

3.1 ΕΙΣΑΓΩΓΗ

Το OpenCV είναι μία ανοιχτού κώδικα βιβλιοθήκη τεχνητής όρασης, μηχανικής μάθησης και ψηφιακής επεξεργασίας εικόνας υπό καθεστώς άδειας BSD και είναι διαθέσιμη από την ιστοσελίδα http://opencv.org . Η βιβλιοθήκη ειναι γραμμένη σε γλώσσα προγραμματισμού C και C++. Ένας από τους κυρίους στόχους του OpenCV είναι να παρέχει εύχρηστες δομές τεχνητής όρασης, μηχανικής μάθησης και επεξεργασίας εικόνας με σκοπό τη γρήγορη ανάπτυξη αρκετά εξελιγμένων λειτουργικών εφαρμογών όρασης από τον χρήστη. Η βιβλιοθήκη OpenCV περιλαμβάνει πάνω από 500 συναρτήσεις που καλύπτουν αρκετά πεδία εφαρμογών όπως η επίβλεψη βιομηχανικής παραγωγής, η ιατρική απεικόνιση, τα θέματα ασφαλείας, η βαθμονόμιση κάμερας, η στερεοσκοπική όραση και η ρομποτική

Η βιβλιοθήκη OpenCV οργανώνεται σε ενότητες. Κάθε συνάρτηση της βιβλιοθήκης ανήκει σε μία συγκεκριένη ενότητα. Οι ενότητες που χρησιμοποιούνται στην παρούσα διπλωματική εργασία είναι οι εξής.

- Η ενότητα Core που περιέχει τις βασικές δομές δεδομένων
- Η ενότητα Highgui που περιέχει συναρτήσεις διεπαφής με το χρήστη οι οποίες μπορούν να χρησιμοποιηθούν για παρουσίαση εικόνων ή την εισαγωγή απλών δεδομένων από τον χρήστη.
- Η ενότητα Imgproc περιέχει μεθόδους επεξεργασίας εικόνας όπως γραμμικές και μη μεθόδους φιλτραρίσματος εικόνας, γεωμετρικούς μετασχηματισμούς, ιστογράμματα κ.λ.π.
- Η ενότητα video περιέχει μεθόδους ανάλυσης βίντεο όπως εκτίμηση κίνησης, αφαίρεση παρασκηνίου, εντοπισμός αντικειμένων.

Ακολουθεί στα επόμενα υποκεφάλαια παρουσίαση των κλάσεων και των συναρτήσεων κάθε ενότητας που χρησιμοποιήθηκαν στην παρούσα διπλωματική εργασία.

3.2 ΕΝΟΤΗΤΑ CORE

Η ενότητα core περιλαμβάνει όλους τους βασικούς τύπους δεδομένων οι οποίοι αποτελούνται από απλά διανύσματα και πίνακες, παράσταση απλών γεωμετρικών δομών όπως σημεία, ορθογώνια κ.λ.π., καθώς και μεγάλες παρατάξεις δεδομένων.

3.2.1 Βασικές δομές

- **Κλάση σημείου -Point**

Η πρότυπη κλάση σημείου δύο διαστάσεων προσδιορίζεται από τις συντεταγμένες του x,y. Υπάρχουν σημεία τύπου ακεραίων και δεκαδικών. Το μεγάλο πλεονέκτημα της κλάσης σημείου είναι ότι είναι απλή στην εφαρμογή της. Για τα αντικείμενα της κλάσης σημείου ισχύει η μεταξύ τους πρόσθεση και αφαίρεση καθώς καθώς και ο πολλαπλασιαμσός τους με βαθμωτό μέγεθος. Τα αντικείμενα της κλάσης σημείου μπορούν να εφαρμοστούν και σε αντικείμενα της κλάσης διανύσματος (vec) και μεγάλης παράταξης δεδομένων (mat), όπως θα παρουσιαστούν στη συνέχεια. Για στιγμιότυπο της κλάσης σημείου δύο διαστάσεων χρησιμοποιείται η έκφραση:

Point2<data_type> p(x,y)

Όπου:
data_type: i για ακέραιες συντεταγμένες και f για δεκαδικές συντεταγμένες
p(x,y): p το όνομα του αντικειμένου και x,y οι τιμές των συντεταγμένων

- **Κλάση διανύσματος - Vec**

Η κλάση διανυσμάτων είναι ουσιαστικά ένας δυσδιάστατος πίνακας με μία στήλη η οποία έχει μεταβλητό μήκος. Αντικείμενα της συγκεκριμένης κλάσης είναι σημαντικά όταν υπάρχει ανάγκη αποθήκευσης πολυδιάστατων δεδομένων αγνώστου πλήθους. Για τη δημιουργία στιγμιοτύπου της κλάσης διανύσματος χρησιμοποιείται η έκφραση:

Vec<dimensions >< data_type > v

Όπου:
dimensions: το πλήθος των διαστάσεων της κάθε καταχώρησης στο διάνυσμα.
data_type: i για ακέραιες συντεταγμένες και f για δεκαδικές συντεταγμένες
v: όνομα μεταβλητής

- **Κλάση βαθμωτού μεγέθους - Scalar**

Η κλάση βαθμωτού μεγέθους είναι στην πραγματικότητα συντόμευση για ένα τετραδιάστατο διάνυσμα διπλής ακρίβειας. Ο τύπος Scalar χρησιμοποιείται ευρέως στο OpenCV για την εισαγωγή τιμών pixel. Για τη δημιουργία στιγμιοτύπου της κλάσης διανύσματος χρησιμοποιείται η έκφραση:

Scalar $s(x_0, x_1, x_2, x_3)$

Όπου:
$s(x_0, x_1, x_2, x_3)$: s το όνομα του αντικειμένου και **x** οι τιμές του τετραδιάστατου ορίσματος

- **Κλάση μεγέθους -Size**

Η πρότυπη κλάση μεγέθους χρησιμοποιείται για τον προσδιορισμό του μεγέθους μίας εικόνας ή ενός ορθογωνίου. Η κλάση περιέχει δύο μέλη, το πλάτος (width) και το ύψος (height). Για στιγμιότυπο της κλάσης μεγέθους χρησιμοποιείται η έκφραση:

Size2< data_type> sz(w,h)

Όπου:
data_type: i για ακέραια ορίσματα και f για δεκαδικά ορίσματα
sz(w,h): sz το όνομα της μεταβλητής και w,h οι τιμές του πλάτους και του ύψους αντίστοιχα.

- **Κλάση ορθογωνίου – Rect**

Πρότυπη κλάση για δύο διαστάσεων ορθογώνια που περιλαμβάνει ως μέλη τις συντεταγμένες της πάνω αριστερά γωνίας του ορθογωνίου καθώς και το πλάτος και ύψος του. Για στιγμιότυπο της κλάσης ορθογωνίου χρησιμοποιείται η έκφραση :

Rect rc(x,y,width,height)

Όπου:
rc(x,y,width,height): rc το όνομα του αντικειμένου, x και y οι συντεταγμένες της πάνω αριστεράς γωνίας του και width και height το πλάτος του και το ύψος του αντίστοιχα.

- **Κλάση περιστραμμένου ορθογωνίου – RotatedRect**

Η κλάση αναπαριστά ορθογώνιο στραμμένο κατά συγκεκριμένη γωνία πάνω σε επίπεδο (Σχήμα 3.1). Κάθε στραμμένο ορθογώνιο χαρακτηρίζεται από το σημείο του κέντρου (κέντρο μάζας), από το μήκος κάθε πλευράς και τη γωνία περιστροφής. Για στιγμιότυπο της κλάσης χρησιμοποιείται η έκφραση:

RotatedRect rr(const Point2f& center, const Size2f& size, float angle)

Όπου :
rr(center, size, angle): rr το όνομα του αντικειμένου, center οι συντεταγμένες σε pixel του κέντρου του, size οι διαστάσεις του, angle η γωνία περιστροφής του

- **Κλάση παράταξης δεδομένων - Mat**

Η κλάση παράταξης δεδομένων δημιουργεί n-διάστατους πυκνούς πίνακες οι οποίοι μπορούν να χρησιμοποιηθούν για αποθήκευση διανυσμάτων, εικόνων, νέφος σημείων και ιστογραμμάτων. Για τη δημιουργία στιγμιοτύπου της κλάσης παράταξης δεδομένων χρησιμοποιείται η έκφραση:

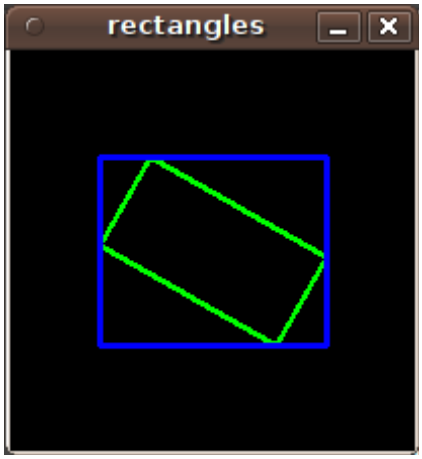
Mat m(rows, columns, data_type)

Όπου:
Mat m(rows, columns, data_types): m το όνομα του αντικειμένου, rows και columns τα ορίσματα του πλήθους των γραμμών και στηλών και data_type η μορφή που θα έχουν τα δεδομένα.

Σημαντικό στοιχείο για τις δομές της κλάσης Mat είναι η δυνατότητα εύρεσης διεύθυνσης των στοιχείων της μέσω της λειτουργίας at η οποία χρησιμοποιείται ως εξής:

<όνομα δομής Mat>.at<data_type>(i,j)

Στις διατάξεις δεδομένων Mat μπορούν αν εφαρμοστούν όλες οι αλγεβρικές πράξεις.



Σχήμα 3.1 Περιστραμμένο ορθογώνιο, με πράσινο χρώμα

3.2.2 Συναρτήσεις σχεδιασμού

Οι συναρτήσεις σχεδιασμού είναι χρήσιμες σε περιπτώσεις που χρειάζεται να σχεδιαστεί σχήμα (γραμμές, τετράγωνα, κύκλοι κ.λ.π.) σε μία υπάρχουσα εικόνα. Στην παρούσα διπλωμα εργασία χρησιμοποιήθηκαν οι ακόλουθες συναρτήσεις σχεδιασμού.

- **Κύκλος –circle**

Η συνάρτηση circle σχεδιάζει ένα κύκλο από δοσμένο κέντρο και ακτίνα. Ορίζεται ως εξής:

Void **circle**(Mat& img, Point center, int radius, const Scalar& color, int thickness, int lineType)

Όπου
Img: Η εικόνα όπου ο κύκλος σχεδιάζεται
Center: Οι συντεταγμένες του κέντρου του κύκλου

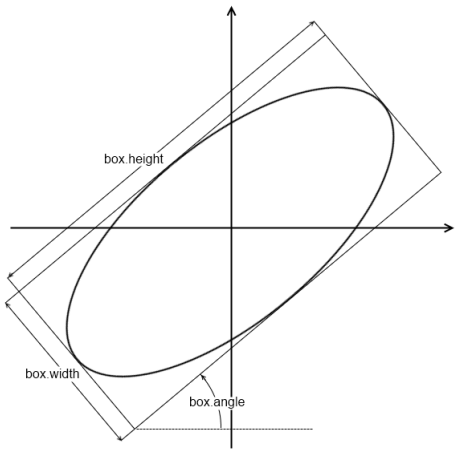
Radius: η ακτίνα του κύκλου
Color: το χρώμα του κύκλου
Thickness: το πόσο λεπτό θα είναι το περίγραμμα του κύκλου
lineType: ο τύπος του ορίου του κύκλου

- **Έλλειψη –Ellipse**

Η συνάρτηση ellipse σχεδιάζει ένα ελλειπτικό τόξο (Σχήμα 3.2). Ορίζεται ως εξής:

Void **ellipse**(Mat& img, const RotatedRect& box, const Scalar& color, int thickness, int lineType)

Όπου:
Img: Η εικόνα όπου η έλλειψη σχεδιάζεται
Box: Έλλειψη σχεδιασμένη μέσω αντικειμένου της κλάσης RotatedRect. Η συνάρτηση εγγράφει έλλειψη σε ορθογώνιο.
Color: Το χρώμα του περιγράμματος της έλλειψης
Thickness: Το πόσο λεπτό θα είναι το περίγραμμα του κύκλου
lineType: Ο τύπος του ορίου του κύκλου



Σχήμα 3.2 Σχεδιασνός έλλειψης εγγεγραμμένης σε ορθογώνιο από συνάρτηση ellipse().

- **fillConvexPoly**

Η συνάρτηση σχεδιάζει ένα εσωτερικά γεμάτο πολύγωνο. Ορίζεται ως εξής:

Void **fillConvexPoly**(Mat& img, const Point* pts, int npts, const Scalar& color, int lineType)

Όπου:
Img: Η εικόνα όπου η επιφάνεια σχεδιάζεται
Pts: Σημεία κορυφών πολυγώνου

Npts: Πλήθος σημείων pts
Color: Το χρώμα του πολυγώνου
lineType: Ο τύπος του ορίου του πολυγώνου.

- **Line**

Η συνάρτηση line σχεδιάζει ένα ευθύγραμμο τμήμα ενώνοντας δύο σημεία. Ορίζεται ως εξής:

Void **line**(Mat& img, Point pt1, Point pt2, const Scalar& color, int thickness, int lineType)

Όπου:
Img: Η εικόνα όπου η γραμμή σχεδιάζεται
Pt1: Σημείο αρχή του ευθύγραμμου τμήματος
Pt2: Σημείο τέλους του ευθύγραμμου τμήματος
Color: το χρώμα του ευθύγραμμου τμήματος
Thickness: το πόσο λεπτό θα είναι το ευθύγραμμο τμήμα

3.3 ΕΝΟΤΗΤΑ HighGUI

Η ενότητα HighGUI προσφέρει μεθόδους διεπαφής χρήστη λογισμικού για γρήγορη δοκιμή λειτουργικότητας και οπτικοποίησης των αποτελεμάτων. Οι μέθοδοι διεπαφής πιό συγκεκριμένα περιλαμβάνουν:

- Δημιουργία και διαχείριση παραθύρου για εμφάνιση εικόνων.
- Τοποθέτηση μπάρας στο παράθυρο, χειρισμό απλών συμβάντων του ποντικιού και χειρισμό εντολών από το πληκτρολόγιο
- Ανάγνωση και εγγραφή εικόνων από και προς το δίσκο
- Ανάγνωση βίντεο από κάμερα ή αρχείο και εγγραφή βίντεο σε αρχείο

Στην παρούσα εργασία χρησιμοποιήθηκαν κλάσεις και συναρτήσεις για δημιουργία και διαχείριση απαραθύρου καθώς και ανάγνωσης και εγγραφής εικόνων.

3.3.1 Μέθοδοι διεπαφής χρήστη

- **imshow**

Η σύναρτηση imshow απεικονίζει μία εικόνα σε συγκερκιμένο παράθυρο και ορίζεται ως εξής:

void **imshow**(const string& winname, InputArray mat)

όπου:
winname: Το όνομα που έχει δώσει ο χρήστης στο παράθυρο

mat: Η εικόνα προς απεικόνιση

- **namedWindow**

Η συνάρτηση namedWindow δημιουργεί παράθυρο το οποίο μπορεί να χρησιμοποιηθεί για απεικόνιση εικόνας και μπάρας χειρισμού παραμέτρων. Τα δημιουργημένα παράθυρα καλούνται με το όνομά τους. Η συνάρτηση ορίζεται ως εξής:

void **namedWindow**(const string& winname, int flags=WINDOW_AUTOSIZE)

όπου:
winname: το όνομα του παραθύρου
flags: Flags του παραθύρου. Τα υποστηριζόμενα είναι:
– WINDOW_NORMAL : Δίνει τη δυνατότητα στο χρήστη να μεταβάλει το μέγεθος του παραθύρου
– WINDOW_AUTOSIZE : Το παράθυρο προσαρμόζεται αυτόματα στο μέγεθος της εικόνας και δεν μπορεί να αλλάξει

3.3.2 Εγγραφή και ανάγνωση εικόνων και βίντεο

- **imread**

Η συνάρτηση imread φορτώνει μία εικόνα από συγκεκριμένο αρχείο και την επιστρέφει. Ορίζεται ως εξής:

Mat **imread**(const string& filename, int flags=1)

Όπου:
Filename: Όνομα αρχείου που θα φορτωθεί
Flags: Flags για τον καθορισμό του είδους χρωματισμού της εικόνας που θα φορτωθεί
– CV_LOAD_IMAGE_ANYDEPTH : Επιστρέφει εικόνα 16-bit/32-bit όταν η εικόνα υποστηρίζει το αντίστοιχο βάθος, αλλιώς την μετατρέπει σε8-bit.
– CV_LOAD_IMAGE_COLOR: Μετατρέπει την εικόνα σε χρωματιστή
– CV_LOAD_IMAGE_GRAYSCALE: Μετατρέπει την εικόνα σε αποχρώσεις του γκρι

- **Imwrite**

Η συνάρτηση imwrite αποθηκεύει την εικόνα σε καθορισμένο φάκελο. Η διαμόρφωση της εικόνας που θα αποθηκευτεί καθορίζεται από την κατάληξη του αρχείου. Ορίζεται ως εξής:

bool **imwrite**(const string& filename, InputArray img)

Όπου:
Filename: Όνομα αρχείου με το οποίο θα αποθηκευτεί η εικόνα
image : Η εικόνα που θα αποθηκευτεί.

3.4 ΕΝΟΤΗΤΑ IMGPROC

Η ενότητα imgproc περιέχει ένα πλήθος συναρτήσεων και κλάσεων σχετικά με την επεξεργασία εικόνας και την τεχνητή όραση. Οι συναρτήσεις και οι κλάσεις αυτές αφορούν διαδικασίες φιλτραρίσματος, γεωμετρικούς και ποικίλους μετασχηματισμούς εικόνας, ιστογράμματα, δομική ανάλυση και περιγραφή σχήματος, ανάλυση κίνησης, ανίχνευση χαρακτηριστικών σημείων και εντοπισμό αντικειμένων. Οι συνάρτήσεις και οι κλάσεις που χρησιμοποιήθηκαν στην παρούσα εργασία αφορούν τις διαδικασίες φιλτραρίσματος, την ανίχνευση χαρακτηριστικών σημείων , τη δομική ανάλυση και περιγραφή σχήματος.

3.4.1 Εφαρμογές φιλτραρίσματος εικόνας

Συναρτήσεις και κλάσεις που περιγράφονται σε αυτό το υποκεφάλαιο χρησιμοποιούνται για ποικίλες γραμμικές και μη γραμμικές λειτουργίες φιλτραρίσματος σε δυσδιάστατες εικόνες. Αυτό σημαίνει ότι για κάθε θέση pixel (x,y) στην πηγαία εικόνα τα γειτονικά pixel χρησιμοποιούνται για τον υπολογισμό της τιμής του pixel στη θέση (x,y) στην τελική εικόνα. Οι συναρτήσεις που εκτελούν διαδικασίες φιλτραρίσματος που χρησιμοποιήθηκαν στην παρούσα εργασία είναι οι ακόλουθες:

- **Dilate και erode**

Οι βασικοί μορφολογικοί μετασχηματισμοί ονομάζονται dilation (διαστολή) και erosion (διάβρωση) και χρησιμοποιούνται κυρίως για αφαίρεση θορύβου, απομόνωση μενονωμένων αντικειμένων και συνένωση ξεχωριστών αντικειμένων αντίστοιχα μέσα σε μία εικόνα.

Dilation είναι η διαδικασία συνέλιξης μίας εικόνας με ένα πίνακα kernel στον οποίο κάθε pixel αντικαθίσταται με το τοπικό μέγιστο από όλες τις τιμές των pixel που βρίσκονται εντός του πίνακα kernel. Γενικά η λειτουργία dilation διευρίνει μία φωτεινή περιοχή (Σχήμα 3.3). Η συνάρτηση ορίζεται ως ακολούθως:

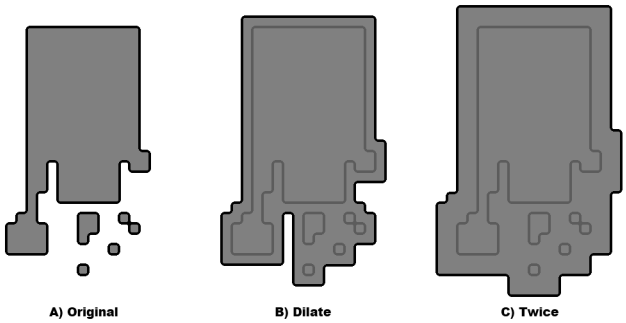
void **dilate**(InputArray src, OutputArray dst, InputArray kernel, Point anchor=Point(-1,-1), int iterations=1)

src: Η εικόνα προς επεξεργασία
dst: Η εικόνα με το τελικό αποτέλεσμα.
kenrel: Ο πίνακας kernel με προεπιλογή 3x3διαστάσεων
anchor: Η θέση του προς αντικατάσταση pixel στον πίνακα kernel, με προεπιλογή το κέντρο του
iterations: Το πλήθος των επαναλήψεων της λειτουργίας dilate.

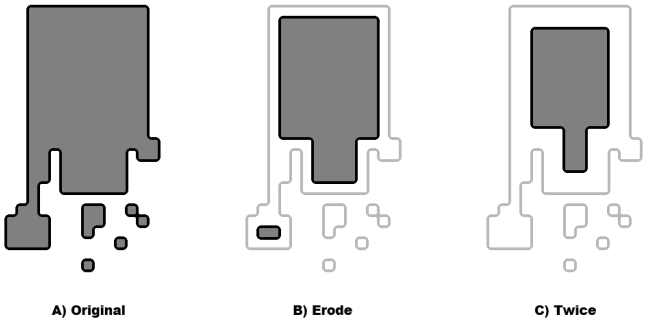
Erosion είναι η αντίστροφη λειτουργία κατά την οποία κάθε pixel αντικαθίσταται από το τοπικό ελάχιστο των τιμών των pixel που βρίσκονται εντός του πίνακα kernel. Γενικά η διαδικασία erosion συρρικνώνει μία φωτεινή περιοχή (Σχήμα 3.4). Η συνάρτηση ορίζεται ως ακολούθως:

void **erode**(InputArray src, OutputArray dst, InputArray kernel, Point anchor=Point(-1,-1), int iterations=1)

Οι παράμετροι είναι όμοιοι με τη λειτουργία dilate.



Σχήμα 3.3 Εφαρμογή μορφολογικής επεξεργασίας dilate A) Αρχική εικόνα B) Εφαρμογή συναρτησης dilate C) Δεύτερη εφαρμογή συνάρτησης dilate



Σχήμα 3.4 Εφαρμογή μορφολογικής επεξεργασίας erode A) Αρχική εικόνα B) Εφαρμογή συναρτησης erode C) Δεύτερη εφαρμογή συνάρτησης erode

- **Γενική συνάρτηση μορφολογίας - morfologyEx**

Το opencv προσφέρει μερικές, πιό προηγμένες λειτουργίες για μορφολογική επεξεργασία εικόνας μέσω της γενικής συνάρτησης morphologyEx, που ως βάση έχουν τις συναρτήσεις dilate και erode. Η συνάρτηση ορίζεται ως εξής:

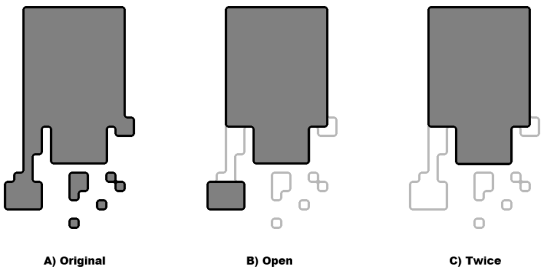
void **morphologyEx**(InputArray src, OutputArray dst, int op, InputArray kernel)

Όπου:
src : Εικόνα προς επεξεργασία
dst : Εικόνα με το τελικό αποτέλεσμα
element: Δομικό στοιχείο από τη συνάρτηση getStructuringElement(). Η συνάρτηση κατασκευάζει δομή Kernel επιθυμητών διαστάσεων και μορφής.
op: Ο τύπος της μορφολογικής λειτουργίας μπορεί να είναι ένας από τους επόμενους
– MORPH_OPEN
– MORPH_CLOSE
– MORPH_GRADIENT
– MORPH_TOPHAT
– MORPH_BLACKHAT

Οι μορφολογικοί μετασχηματισμοί στηρίζονται στις λειτουργίες dilate και erode ως εξής:

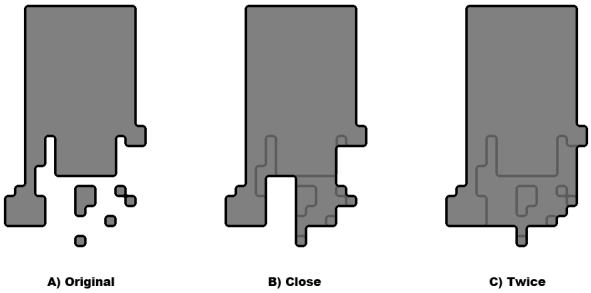
Open: dilate(erode(src,element))
Close: erode(dilate(src,element))
Gradient: dilate(src,element) – erode(src,element)
Tophat: src – open(src,element)
Blackhat: close(src,element) - src

Στην παρούσα εργασία χρησιμοποιούνται οι λειτουργίες opening και closing. Η λειτουργία opening εφαρμόζεται για το διαχωρισμό περιοχών σε μία εικόνα (Σχήμα 3.5). Για παράδειγμα σε μία εικόνα κυττάρων από μικροσκόπιο μπορεί να χρησιμοποιηθεί η λειτουργία opening για διαχωρισμό κυττάρων που βρίσκονται πολύ κοντά μεταξύ τους.



Σχήμα 3.5 Εφαρμογή λειτουργίας open A) Αρχική εικόνα B) Εφαρμογή λειτουργίας open C) Δεύτερη εν συνεχεία εφαρμογή λειτουργίας open

Η λειτουργία closing χρησιμοποιείται για τη μείωση μη θεμιτών ή λόγω θορύβου διαχωρισμών (Σχήμα 3.6).



Σχήμα 3.6 Εφαρμογή λειτουργίας close A) Αρχική εικόνα B) Εφαρμογή λειτουργίας close C) Δεύτερη εν συνεχεία εφαρμογή λειτουργίας close.

3.4.2 Ανίχνευση χαρακτιστικών σημείων – Ανιχνευτής άκρων Canny

Για τον εντοπισμό χαρακτηριστικών σημείων σε μία εικόνα, τα οποία αποτελούν τη βάση για τον εντοπισμό και περιγραφή αντικειμένων της εικόνας, χρησιμοποιούνται μετασχηματισμοί οι οποίοι αλλάζουν την εικόνα προς μία εναλλακτική απεικόνιση των δεδομένων της. Η νέα απεικόνιση παραμένει μία διάταξη με τιμές αλλά αυτές ο τιμές είναι διαφορετικές σε ερμηνεία από ότι οι τιμές της έντασης στην αρχική εικόνα. Σημαντική μέθοδος στην εύρεση χαρακτηριστικών σημείων είναι ο ανιχνευτής άκρων Canny, ο οποίος εντοπίζει τα σημεία στα οποία υπάρχει σημαντική αλλαγή στην ένταση του φωτός (άκρα)

στην αρχική εικόνα και τα επισημάνει την τελική εικόνα. Το πιό σημαντικό χαρακτηριστικό του αλγορίθμου Canny είναι ότι προσπαθεί να συγκεντρώσει τα μεμονωμένα υποψήφια σημεία άκρων σε περιγράμματα (Σχήμα 3.7). Η συνάρτηση Canny ορίζεται ως εξής:

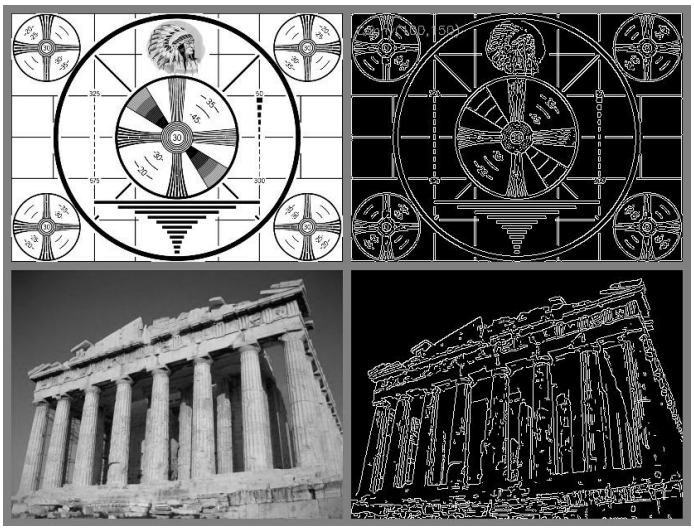
void **Canny**(InputArray image, OutputArray edges, double threshold1, double threshold2)

Όπου :
image: Αρχική εικόνα μονού καναλιού.
edges: Εικόνα όπου θα αποθηκευτεί το αποτέλεσμα.
threshold1: Άνω κατώφλι για αποδοχή ως pixel άκρου.
threshold2: Κάτω κατώφλι για απόρριψη.

3.4.3 Δομική ανάλυση και περιγραφή σχήματος

Συναρτήσεις όπως ο ανιχνευτής άκρων Canny που χρησιμοποιούνται για την εύρεση των pixel άκρων τα οποία άκρα είναι τα όρια που διαχωρίζουν τα διάφορα τμήματα της εικόνας δεν έχουν λειτουργίες που να προσδιορίζουν τις άκρες ως σύνολο. Η συγκέντρωση των άκρων σε περιγράμματα και η περαιτέρω επεξεργασία τους υλοποιείται από συναρτήσεις που ανήκουν στην κατηγορία δομικής ανάλυσης και περιγραφής σχήματος. Τα περιγράμματα είναι ένα σημαντικότατο εργαλείο για περαιτέρω διαδικασίες όπως η ανάλυση σχήματος, ο εντοπισμός και η αναγνώριση αντικειμένων.

Περίγραμμα θεωρείται μία λίστα σημείων που αναπαριστούν καμπύλη γραμμή πάνω σε εικόνα. Το OpenCV για την αποθήκευση των σημείων χρησιμοποιεί αντικείμενα της κλάσης Vec<Point2f> όπου κάθε καταχώρηση στο διάνυσμα κωδικοποιεί πληροφορίες για τη θέση του επόμενου σημείου της καμπύλης. Η συνάρτηση εντοπίζει τα περιγράμματα των δομικών στοιχείων της εικόνας είναι η findContous( ) και ανήκει στην κατηγορία μοντέλων ενεργών περιγραμμάτων.



Σχήμα 3.7 Εφαρμογή συνάρτησης Canny

- **Εντοπισμός περιγραμμάτων - findContours**

Η συνάρτηση findContours υπολογίζει περιγράμματα από δυαδικές εικόνες που δημιουργούνται από συναρτήσεις όπως η μέθοδος εύρεσης άκρων Canny και οι διάφορες μέθοδοι κατωφλίωσης.

C++: void **findContours**(InputOutputArray image, OutputArrayOfArrays contours, OutputArray hierarchy,int mode, int method, Point offset=Point())

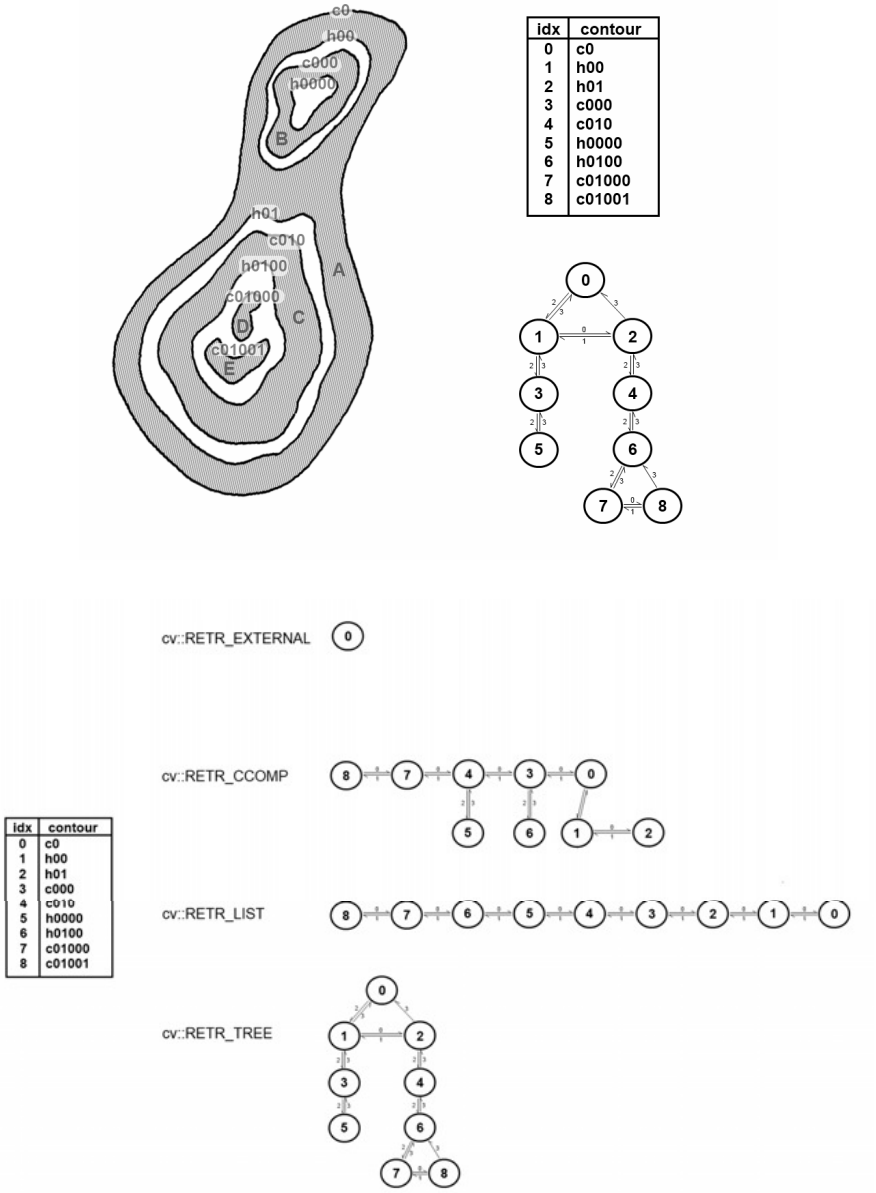
Όπου:
Image: Πηγαία εικόνα ενός καναλιού των 8 bit.
contours: Εντοπισμένα περιγράμματα σε μορφή διάνυσμα διανυσμάτων. Κάθε περίγραμμα αποθηκεύεται ως διάνυσμα σημείων
Hierarchy: Διάνυσμα που λαμβάνεται ως έξοδος και περιέχει πληροφορίες για την τπολογία της εικόνας. Έχει τόσα στοιχεία όσα και το πλήθος των περιγραμμάτων. Για κάθε i-οστό περίγραμμα τα στοιχεία hierarchy[i][0], [i][1], [i][2], [i][3] λαμβάνουν πληροφορίες για το προηγούμενο και επόμενο ίδιας ιεραρχίας περίγραμμα και για τον πρώτο απόγωνο και πρόγωνο αντίστοιχα (Σχήμα 3.8).
Mode: Λειτουργίες ανάκτησης περιγραμμάτων (Σχήμα 3.8)
CV_RETR_EXTERNAL: Ανάκτηση μόνο των εσχάτων εξωτερικών περιγραμμάτων.
CV_RETR_LIST: Ανάκτηση όλων των περιγραμμάτων χωρίς ιεραρχική σχέση
CV_RETR_CCOMP: Ανάκτηση όλων των περιγραμμάτων και οργάνωσή τους σε δύο επίπεδα ιεράρχισης .Στο πρώτο επίπεδο υπάρχουν τα εξωτερικά όρια των αντικειμένων. Στο δεύτερο επίπεδο υπάρχουν τα όρια των οπών. Εάν υπάρχει άλλο εξωτερικό περίγραμμα στο εσωτερικό της οπής τοποθετείται στο πρώτο επίπεδο.
CV_RETR_TREE: Ανάκτηση όλων των περιγραμμάτων και οργάνωσή τους σε πλήρη ένθετη ιεραρχία.
Method: Καθορίζει το πως θα απεικονίζονται τα περιγράμματα.
Offset: Προαιρετικό όρισμα. Κάθε σημείο του περιγράμματος μετατοπίζεται κατά αυτή την τιμή.

- **Σχεδίαση περιγραμμάτων - drawContours**

Η συνάρτηση drawContours χρησιμοποιεί τη λίστα περιγραμμάτων από τη συνάρτηση findContours και σχεδιάζει τα περιγράμματα στην εικόνα. Ορίζεται ως ακολούθως:

void **drawContours**(InputOutputArray image, InputArrayOfArrays contours, int contourIdx, const Scalar& color, int thickness=1, int lineType=8, InputArray hierarchy=noArray(), int maxLevel = INT_MAX, Point offset=Point())

Όπου:
Image: Εικόνα που θα σχεδιαστούν τα περιγράμματα.
Contours: Διανύσματα με τα σημεία των περιγραμμάτων.
contourIdx: Παράμετρος που δείχνει εάν θα σχεδιαστεί ένα συγκεκριμένο περίγραμμα ή όλα (αρνητική τιμή).
Color: Χρώμα περιγραμμάτων.
Thickness: Προσδιορίζει το πάχος της γραμμής του περιγράμματος
lineType: Τύπος γραμμής
hierarchy: Προαιρετική πληροφορία για την ιεραρχία, χρησιμοποιείται μόνο σε περίπτωση σχεδίασης συγκεκριμένων περιγραμμάτων.
maxLevel: Επίπεδο ιεραρχίας μέχρι το οποίο θα σχεδιαστούν τα περιγράμματα.
Offset: Μετατόπιση περιγραμμάτων σύμφωνα με το συγκεκριμένο offset.



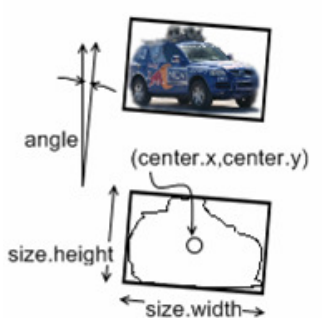
Σχήμα 3.8 Στην πάνω πλευρά της εικόνας φαίνεται δοκιμαστική εικόνα που τίθεται ως όρισμα στη συνάρτηση findContours και περιέχει πέντε περιοχές, A ως E με 9 συνολικά περιγράμματα. Τα εξωτερικά περιγράμματα συμβολίζονται με c και τα εσωτερικά με h (οπές), πίνακας contour. Η συνάρτηση έχει τη δυνατότητα να ξεχωρίζει τα εσωτερικά και εξωτερικά περιγράμματα. Στην κάτω πλευρά παρουσιάζονται οι διάφορες λειτουργίες ανάκτησεις περιγραμμάτων (mode). Κάθε κόμβος είναι ένα περίγραμμα και οι συνδέσεις τους (1 ως 4) είναι οι αντίστοιχες που περιγράφονται από τη διάταξη hierarchy.

- **Προσαρμογή ορθογωνίου πλαισίου - minAreaRect**

Η συνάρτηση minAreaRect αποτελεί αντικείμενο της κλάσης RotatedRect και υπολογίζει και επιστρέφει περιστραμμένο ορθογώνιο πλαίσιο ελάχιστου εμβαδού που εσωκλείει σύνολο σημείων (περίγραμμα) (Σχήμα 3.9). Η συνάρτηση ορίζεται ως εξής:

RotatedRect **minAreaRect**(InputArray points)

Όπου:
Points: Διάνυσμα δυσδιάστατων σημείων.



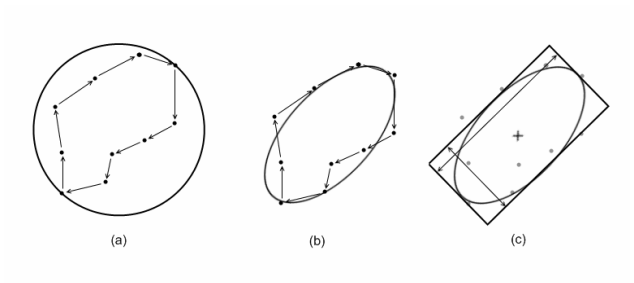
Σχήμα 3.9 Εφαρμογή συνάρτησης minAreaRect για προσαρμογή ορθογωνίου πλαισίου στα σημεία του περιγράμματος

- **Προσαρμογή έλλειψης - fitEllipse**

Η συνάρτηση fitEllipse αποτελεί αντικείμενο της κλάσης RotatedRect και υπολογίζει μία έλλειψη που είναι η καλύτερη προσέγγιση στο σύνολο των σημείων του περιγράμματος. Στην πραγματικότητα επιστρέφει ένα περιστραμμένο ορθογώνιο στο οποίο εγγράφεται η έλλειψη. Η προσαρμογή γίνεται με την εφαρμογή των ελαχίστων τετραγώνων. Ορίζεται ως εξής:

RotatedRect **fitEllipse**(InputArray points)

Όπου:
Points: Σύνολο δυσδιάστατων σημείων.



Σχήμα 3.10 Περίγραμμα δέκα σημείων a) Περιβάλλον κύκλος b) Προσαρμοσμένη έλλειψη c) Χρήση περιστραμμένου ορθογωνίου για αναπαράσταση της έλλεψης.

3.5 ΕΝΟΤΗΤΑ VIDEO - ΔΙΑΧΩΡΙΣΜΟΣ ΠΑΡΑΣΚΗΝΙΟΥ

Η ενότητα video περιέχει κλάσεις και συναρτήσεις που αφορούν ανάλυση βίντεο και εντοπισμό αντικειμένων. Στην παρούσα εργασία παρόλο που δεν χρησιμοποιείται video αλλά φωτογραφίες για επεξεργασία δανίζεται μέθοδος διαχωρισμού παρασκηνίου προσκηνίου που έχει δημιουργηθεί με σκοπό να μαθαίνεται το παρασκήνιο μέσα από τη ροή του βίντεο και

όταν εμφανίζονται καινούργια αντικείμενα αυτά να απομονώνονται θεωρούμενα ως προσκήνιο. Η κλάση της μεθόδου που χρησιμοποιείται ονομάζεται BackgroundSubtractorMOG και υλοποιεί αλγόριθμο που στηρίζεται στη μέθοδο συνδιασμού Gaussin-ών (υποκεφάλαιο 2.6.2) με σκοπό τον εντοπισμό αντικειμένων σε πραγματικό χρόνο (μέσω της διαδικασίας διαχωρισμού προσκηνίου και παρασκηνίου). Τα αντικείμενα της κλάσης ορίζονται ως ακολούθως:

BackgroundSubtractorMOG BSM(int history, int nmixtures, double backgroundRatio, double noiseSigma=0)

Όπου:
BSM: Το όνομα του αντικειμένου.
history: Το σημείο από το οποίο ο αλγόριθμος αλλάζει από τη διαδικασία αρχικοποίησης (μετράται σε καρέ)
Nmixtures: Πλήθος Gaussian-ών που επιλέγεται να χρησιμοποιηθούν στο μοντέλο εκμάθησης
backgroundRatio: Ελάχιστο ποσοστό από τις σταθμισμένες κατανομές Gauss του μοντέλου που επαρκούν για να μοντελοποιήσουν το παρασκήνιο.
noiseSigma: Η παράμετρος θέτει μία αβεβαιότητα σε κάθε νέα κατανομή Gauss που δημιουργείται.

Για τον χειρισμό του αντικειμένου ώστε πρώτον να λάβει μία νέα εικόνα και να υπολογίσει τη μάσκα του προσκηνίου για τη νέα εικόνα και δεύτερον να παράγει το μοντέλο του παρασκηνίου, χρησιμοποιήται η μορφή:

BSM(InputArray image, OutputArray fgmask, double learningRate)
Όπου:

BSM: Το όνομα του αντικειμένου
Image: Η εικόνα προς επεξεργασία
Fgmask: Η μάσκα που παράγεται για τα αντικείμενα του προσκηνίου.
learningRate: Ρυθμός εκμάθησης, 1 εκμάθηση κάθε καρέ, 0 εκμάθηση κανενός καρέ.

3.6 ΒΙΒΛΙΟΓΡΑΦΙΑ

[3.1] E-book "*The OpenCV Reference Manual*" από ηλεκτρονική διεύθυνση http://docs.opencv.org/

[3.2] A. Kaebler , G. Bradski "*Learning OpenCV Computer Vision in C++ with the OpenCV Library*" O'Reilly, 2nd edition, early-release.

ΚΕΦΑΛΑΙΟ 4

ΜΟΝΤΕΛΟ ΓΙΑ ΕΝΤΟΠΙΣΜΟ ΘΕΣΗΣ ΚΥΛΙΝΔΡΙΚΩΝ ΤΕΜΑΧΙΩΝ

Στα κεφάλαια 2 και 3 παρουσιάστηκαν συγκεκριμένα μοντέλα της επιστήμης της τεχνητής όρασης και της μηχανικής μάθησης τόσο σε θεωρητικό επίπεδο όσο και σε επίπεδο εφαρμογών. Τα μοντέλα αυτά θα χρησιμοποιηθούν συνδιαστικά μεταξύ τους, με κατάλληλη παραμετροποίηση ώστε να δομηθεί ένα νέο μοντέλο με χαρακτηριστικά ικανά να οδηγήσουν σε μία όσο το δυνατόν πίο εύρωστη και αποτελεσματική επίλυση του προβλήματος που απασχολεί την παρούσα διπλωματική.

4.1 ΕΞΕΤΑΣΗ ΤΟΥ ΠΡΟΒΛΗΜΑΤΟΣ - ΣΥΜΒΙΒΑΣΜΟΙ

Το πρόβλημα που ορίζεται προς επίλυση είναι ο εντοπισμός της θέσης μεταλλικών κυλινδικών τεμαχίων πάνω σε οριζόντια επιφάνεια με χρήση ψηφιακής επεξεργασίας εικόνων που έχουν ληφθεί από τα αναφερθέντα μεταλλικά αντικείμενα. Ως προσδιορισμός θέσης του αντικειμένου θεωρούνται οι συντεταγμένες του κέντρου βάρους του όγκου του ως προς συγκεκριμένο σύστημα αναφοράς συντεταγμένων στον πραγματικό κόσμο καθώς και η διεύθυνση του κατά μήκους άξονά τους.

Τα κυλινδρικά μεταλλικά αντικείμενα θεωρείται ότι είναι της τάξης των λίγων εκατοντάδων χιλιοστών, έχουν σταθερή διατομή και είναι τοποθετημένα οριζοντίως στην επιφάνεια. Ως επιφάνεια θεωρείται μία οποιαδήποτε επίπεδη επιφάνεια μηδενικής κλήσης οποιουδήποτε χρωματισμού και επαρκώς φωτισμένη. Οι ψηφιακές εικόνες μπορεί να λαμβάνονται από οποιουδήποτε είδους ψηφιακή φωτογραφική μηχανή.

Για τη διαδικασία επίλυσης του προβλήματος γίνονται οι ακόλουθοι συμβιβασμοί. Χρησιμοποιείται μία μόνο φωτογραφική μηχανή η οποία τοποθετείται σε σταθερό ύψος κατακόρυφα από την επιφάνεια οπότε και λαμβάνεται μόνο η κάτοψη των αντικειμένων και του παρασκηνίου. Η πληροφορία που περιέχεται στον τρισδιάστατο κόσμο μετασχηματίζεται και περιορίζεται σε μία δυσδιάστατη οριζόντια προβολή. Επίσης η επεξεργασία γίνεται σε μεμονομένες φωτογραφίες και όχι σε ροή εικόνων (βίντεο) όπου σε κάθε λήψη τα αντικείμενα είναι ακινητοποιημένα και η οποιαδήποτε μεταβολή τους συμβαίνει ανάμεσα στις λήψεις. Τέλος θεωρείται ότι υπάρχει η δυνατότητα λήψης εικόνων του παρασκηνίου χωρίς αντικείμενα.

4.2 ΠΕΡΙΓΡΑΦΗ ΚΑΙ ΥΛΟΠΟΙΗΣΗ ΜΟΝΤΕΛΟΥ

Για την επίλυση του προβλήματος επιλέχθηκε να υλοποιηθεί μοντέλο που θα εντοπίζει και θα επεξεργάζεται τα περιγράμματα των μεταλλικών αντικειμένων καθώς μέσα από τα δεδομένα που προκύπτουν από το περίγραμμα υπάρχει η δυνατότητα άντλησης πληροφοριών για την θέση του κάθε αντικειμένου. Κάθε εικόνα που λαμβάνεται επεξεργάζεται κατάλληλα ώστε να εντυπωθούν με όσο το δυνατόν μεγαλύτερη ακρίβεια τα περιγράμματα των αντικειμένων τα οποία στη συνέχεια προσεγγίζονται απο προσαρμοσμένα ορθογώνια παραλληλόγραμμα πλαίσια. Η επιλογή για την προσαρμογή ορθογωνίων πλαισίων

έγινε γιατί τα περιγράμματα που προκύπτουν έχουν συνήθως αρκετά ανομοιόμορφο σχήμα και οποιαδήποτε επαγωγή συμπεράσματος θα περιείχε έντονο σφάλμα καθώς και γιατί η δυσδιάστατη ιδανική προβολή σε οριζόντιο επίπεδο των κυλινδρικών αντικειμένων είναι μία ορθογώνια παραλληλόγραμμη επιφάνεια. Η ανομοιομορφία του περιγράμματος οφείλεται πρώτον στις μη ιδανικές συνθηκες φωτισμού που προκαλούν σκιάσεις, δεύτερον στις έντονες αντανακλάσεις από την μεταλλική επιφάνεια που αλλοιώνουν σε μεγάλο βαθμό την απεικόνιση των αντικειμένων και τρίτον στο θόρυβο της όλης διαδικασίας παραγωγής της εικόνας.

Καθοριστικό χαρακτηριστικό του μοντέλου είναι η δυνατότητά του να εκπαιδεύεται ώστε να αναγνωρίζει τη μορφή και το περιεχόμενο του παρασκηνίου. Σκοπός της συγκεκριμένης λειτουργίας είναι σε κάθε νέα λήψη να αφαιρείται το παρασκήνιο οπότε και οι δυσλειτουργίες και αστοχίες που προκαλεί το περιεχόμενό του στη διαδικασία προσδιορισμού του περιγράμματος των αντικειμένων.

4.2.1 Αρχιτεκτονική της επίλυσης

Το μοντέλο επίλυσης υλοποιήθηκε στη γλώσσα προγραμματισμού C++ στο ολοκληρωμένο περιβάλλον ανάπτυξης λογισμικού Code::Blocks 10.05 με GCC compiler και GNU make. Στο συγκεκριμένο compiler έχει προστεθεί η βιβλιοθήκη OpenCV 2.4.9. Η ανάπτυξη έγινε σε λειτουργικό σύστημα Windows Vista Home Prenium 32 bit αρχιτεκτονικής.

Το λογισμικό που υλοποιεί το μοντέλο ακολουθεί την ροή:

- Εισαγωγή εικόνων στο σύστημα
- Προεπεξεργασία εικόνων
- Εκμάθηση παρασκηνίου
- Αφαίρεση παρασκηνίου από είκονα με αντικείμενα
- Επεξεργασία παραγώμενης εικόνας
- Εύρεση περιγραμμάτων
- Προσαρμογή ορθογώνιου παραλληλόγραμμου πλαισίου και έλλειψης σε κάθε περίγραμμα
- Παρουσίαση αποτελεσμάτων στην οθόνη.

Στην συνέχεια αναλύεται το κάθε βήμα ξεχωριστά.

- Εισαγωγή εικόνων στο σύστημα

Η επεξεργασία των εικόνων δε γίνεται σε πραγματικό χρόνο. Τα ονόματα των αρχείων των εικόνων με σκέτο το παρασκήνιο και των εικόνων με τα αντικείμενα δηλώνονται στα ορίσματα του λογισμικού. Κατά σειρά, πρώτα οι εικόνες του παρασκηνίου που προορίζονται για την εκπαιδευση του λογισμικού και στη συνέχεια η εικόνα με τα αντικείμενα.

- Προεπεξεργασία εικόνων

Η προεπεξεργασία κάθε είκονας έχει ως σκοπό μιά αρχική μορφοποίηση ώστε να διευκολυνθεί η επεξεργασία της εικόνας στα επόμενα βήματα. Στην αρχή γίνεται αποπαραμόρφωση της εικόνας βάση του πίνακα κάμερας και των συντελεστών παραμόρφωσης όπως προέκυψαν από τη διαδικασία καλιμπραρίσματος της φωτογραφικής μηχανής. Στη συνέχεια εφαρμόζεται η διαδικασία φιλτραρίσματος opening όπως παρουσιάστηκε στο υποκεφάλαιο (3.4.1) ώστε να εξομαλυνθούν οι περιοχές των σωμάτων με έντονη αντανάκλαση.

- Εκμάθηση παρασκηνίου

Η διαδικασία της εκμάθησης υλοποιείται εντός ενός βρόγχου for που περιέχει αντικείμενο της κλάσης BackgroundSubtractorMOG (υποκεφάλαιο 3.5), στο οποίο είναι ενεργοποιημένη η λειτουργία εκμάθησης. Το πλήθος των επαναλήψεων σε αντιστοιχία με τις διαθέσιμες εικόνες για εκμάθηση όρίζεται χειροκίνητα από το χρήστη εντός του λογισμικού.

- Αφαίρεση παρασκηνίου από είκονα με αντικείμενα

Οταν ολοκληρωθεί η διαδικασία της εκμάθησης ακολουθεί η τελευταία επανάληψη του προαναφερθέντος βρόγχου for στην οποία απενεργοποιείται η λειτουργία εκμάθησης του αντικειμένου της κλάσης BackgroundSubtractorMOG και ταυτόχρονα εισέρχεται σε λειτουργία αφαίρεσης του γνωστού πλέον παρασκηνίου από την τελευταία εικόνα των ορισμάτων που περιέχει τα κυλινδρικά αντικείμενα. Ως έξοδος λαμβάνεται δυαδική ασπρόμαυρη εικόνα με μαύρο χρώμα στον εκτιμώμενο ως παρασκήνιο χώρο και άσπρο στον εκτιμώμενο χώρο των νεοεμφανιζόμενων αντικειμένων. Η εικόνα αυτή ονομάζεται μάσκα.

- Επεξεργασία παραγώμενης εικόνας

Στην παραγώμενη εικόνα μάσκας εφαρμόζεται η διαδικασία φιλτραρίσματος closing, υποκεφάλαιο (3.4.1), ώστε να ενωποιηθούν τα τυχόν αποκομμένα τμήματα λόγω αστοχιών που αντιστοιχούν στο χώρο ενός νεοεμφανιζόμενου αντικειμένου. Το μέγεθος του φίλτρου Kernel που χρησιμοποιεί η διαδικασία closing ορίζεται βάση αποτελεσματικότητας από τον χρήστη. Στη συνέχεια γίνεται αντιστροφή των χρωματων της μάσκας ώστε το παρασκήνιο να αντιστοιχεί σε άσπρο χρώμα και ο χώρος των αντικειμένων σε μαύρο.

- Εύρεση περιγραμμάτων

Στη συνέχεια στην εικόνα μάσκας εφαρμόζεται η συνάρτηση findContours (υποκεφάλαιο 3.4.3), η οποία εντοπίζει το περίγραμμα κάθε συνεκτικής μορφής της εικόνας μάσκας και αποθηκεύει τα αποτελέσματα.

- Προσαρμογή ορθογώνιου παραλληλόγραμμου πλαισίου και τόξου έλλειψης σε κάθε περίγραμμα

Ακολούθως κάθε περίγραμμα εισάγεται ως όρισμα στις συναρτήσεις minAreaRect και fitEllipse (υποκεφάλαιο 3.4.3) οι οποίες επιστρέφουν ένα ορθογώνιο παραλληλόγραμμο και ένα τόξο έλλειψης αντίστοιχα, προσαρμοσμένα στο περίγραμμα. Από το αντικείμενο “ορθογώνιο” που επιστέφει η συνάρτηση minAreaRect επιλέγουμε το μέλος του “.center” το οποίο δίνει τις συντεταγμένες του κέντρου του ορθογωνίου. Εάν το λαμβάνον ορθογώνιο είναι σφιχτά προσαρμοσμένο στην απεικόνιση του μεταλλικού κυλινδρικού αντικειμένου τότε ταυτίζονται τα κέντρα τους (στο επίπεδο της δυσδιάστατης προβολής του αντικειμένου). Η εύρεση της διεύθυνσης γίνεται μέσα από το μέλος “.angle” του αντικειμένου “έλλειψης” που επιστέφει η συνάρτηση fitEllipse καθώς έτσι ορίζεται μοναδικά η γωνία ως προς τον μέγιστο άξονα της προσαρμοσμένης έλλειψης σε αντίθεση με το μέλος “.angle” του αντικειμένου “ ορθογωνίου” το οποίο επιστρέφει ως γωνία διεύθυνσης είτε τη γωνία προς τον κατά μήκος άξονα του ορθογωνίου είτε προς τον κατά πλάτος άξονα, οπότε δεν υπάρχει μονοσήμαντη εκτίμηση της γωνίας του μεταλικού αντικειμένου.

- Παρουσίαση αποτελεσμάτων στην οθόνη.

Το οπτικό αποτέλεσμα κάθε βήματος του λογισμικού εμφανίζεται στην οθόνη σε σύγκριση με την αρχική εικόνα ώστε να ύπαρχει η δυνατότητα οπτικού ελέγχου της διαιδικασίας από τον χρήστη. Τέλος εμφανίζονται στην οθόνη διεπαφής του λογισμικού περιβάλλοντος τα εξής στοιχεία:

- Πλήθος εντοπισμένων αντικειμένων
- Διαστάσεις και εμβαδόν κάθε εντοπισμένου αντικειμένου (σε pixel)
- Συντεταγμένες (σε pixel) του κέντρου κάθε προσαρμοσμένου ορθογωνίου παραλληλογράμου
- Η γωνία κάθε προσαρμοσμένου τόξου έλλειψης προς τον μέγιστο αξονά της.

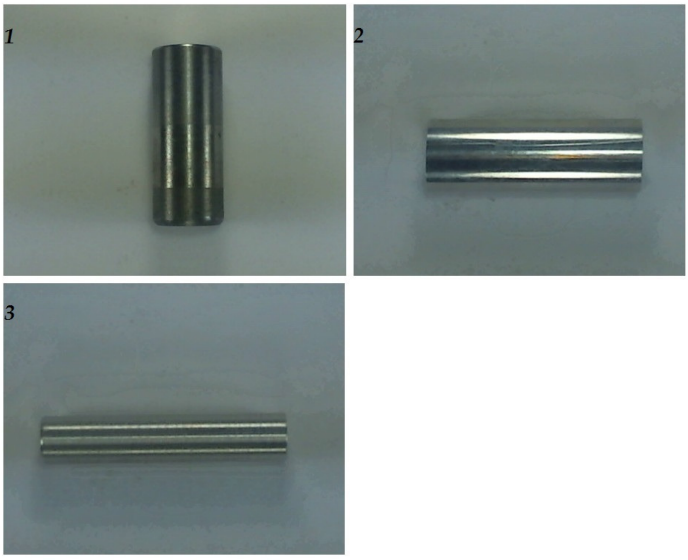
4.3 ΠΕΙΡΑΜΑΤΙΚΕΣ ΔΙΑΤΑΞΕΙΣ - ΑΠΟΤΕΛΕΣΜΑΤΑ

Στα πλαίσια της παρούσας διπλωματικής εργασίας για τη λήψη ψηφιακών φωτογραφιών χρησιμοποιήθηκε η web βιντεοκάμερα Logitech C525 HD (Σχήμα 4.1). Σημαντικό είναι το επίπεδο του φακού της κάμερας να είναι παράλληλο με το επίπεδο στο οποίο είναι τοποθετημένα τα μεταλλικά αντικείμενα καθώς και το κέντρο του φακού να βρίσκεται ακριβώς κάθετα πάνω από το κέντρο της περιοχής που έχει οριστεί για εντοπισμό αντικειμένων.

Η επιλογή των κυλινδρικών μεταλλικών αντικειμένων για τα οποία θα υπολογίζεται η θέση του κέντρου τους και η διεύθυνσή τους έγινε τυχαία με σκοπό να μελετηθεί η ευρωστία και η ακρίβεια των αποτελεσμάτων του λογισμικού. Τα συγκεκριμένα κυλινδρικά αντικείμενα που χρησιμοποιήθηκαν απεικονίζονται στο σχήμα 4.2. Το αντικείμενο στη θέση 1 του σχήματος 4.2 έχει διαστάσεις 90 x 35 mm στη θέση 2 85 x 30 mm και στη θέση 3 100 x 20 mm.

Για τη μελέτη των αποτελεσμάτων του λογισμικού υλοποιήθηκαν δύο πειραματικές διατάξεις οι οποίες παρουσιάζονται στα επόμενα δύο υποκεφάλαια.

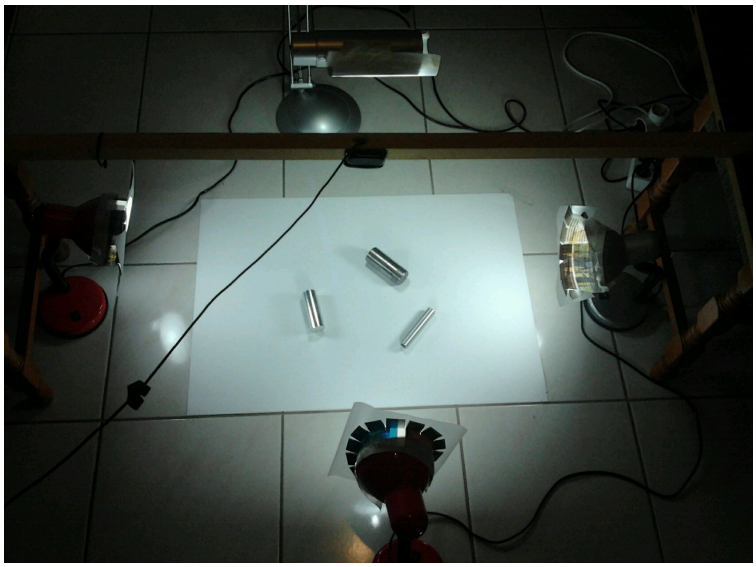
Σχήμα 4.1 web βιντεοκάμερα LogitechC525 HD



Σχήμα 4.2 Απεικόνιση των τριών μεταλλικών αντικειμένων που χρησιμοποιήθηκαν για έλεγχο του λογισμικού 1) Διαστάσεις 90 x 35 mm 2) Διαστάσεις 85 x 30 mm 3) Διαστάσεις 100 x 20 mm

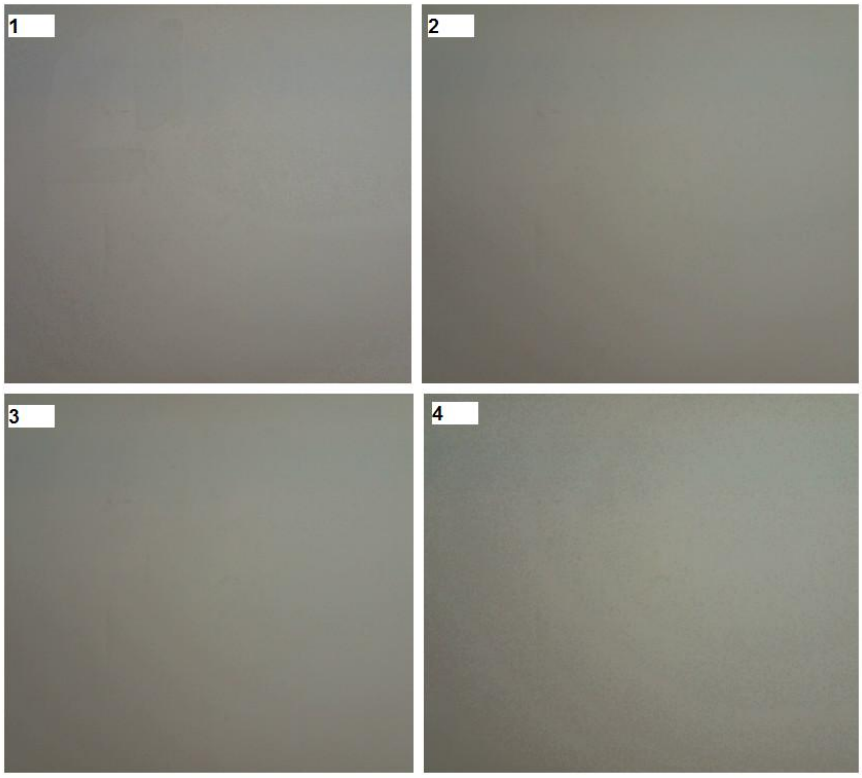
4.3.1 Πρώτη πειραματική διάταξη - Αποτελέσματα

Η πρώτη πειραματική διάταξη που κατασκευάσθηκε αποτελείται από λευκή επίπεδη επιφάνεια διαστάσεων 70 x 50 cm στην οποία τοποθετούνται τυχαιά εντός εμβέλειας κάμερας τα τρία μεταλλικά κυλινδικά αντικείμενα. Η κάμερα τοποθετείται σε οριζόντια θέση παράλληλα στο επίπεδο της επιφάνειας και σε ύψος 65 cm πάνω από το κέντρο της επιφάνειας (Σχήμα 4.3). Η διάταξη φωτίζεται απο τέσσερα φωτιστικά λαμπατερ με 20 W λευκή λάμπα φθορίου το καθένα.



Σχήμα 4.3 Πρώτη πειραματική διάταξη. Λευκή επίπεδη επιφάνεια διαστάσεων 70 x 50 cm στην οποία τοποθετούνται τυχαιά εντός εμβέλειας κάμερας τα τρία αντικείμενα. Η κάμερα έχει τοποθετηθεί οριζόντια σε ύψος 65 cm πάνω από το κέντρο της επιφάνειας. Η διάταξη φωτίζεται απο τέσσερα φωτιστικά λαμπατερ με 20 W λευκή λάμπα φθορίου το καθένα.

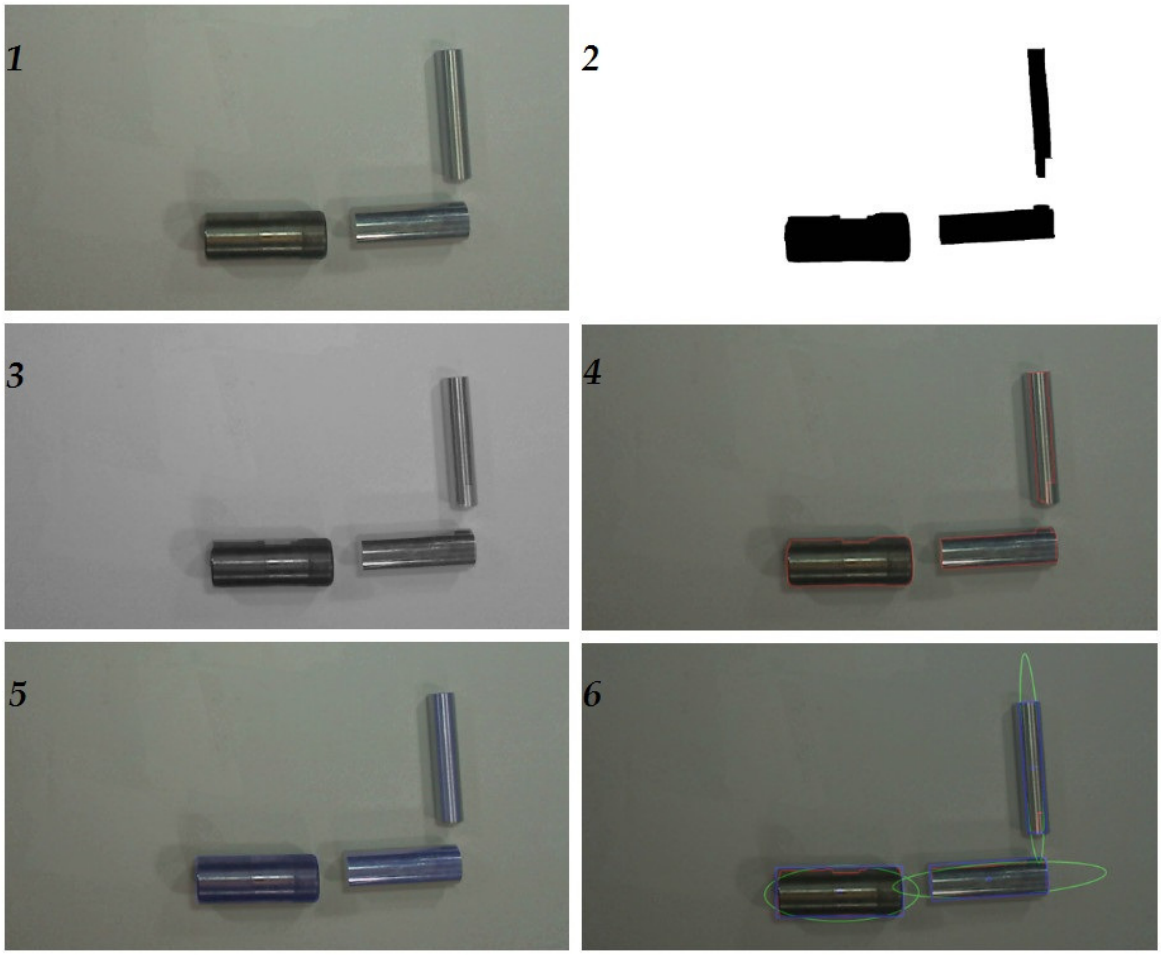
Αρχικά λαμβάνονται λήψεις του παρασκηνίου (Σχήμα 4.4) δηλαδή της λευκής επιφάνειας χωρίς αντικείμενα οι οποίες θα χρησιμοποιηθούν από το λογισμικό ως εκπαιδευτικές για την εκμάθηση του παρασκηνίου. Λόγω της απλότητας και της αμεταβλητότητας του παρασκηνίου παρατηρήθηκε ότι τέσσερεις λήψεις δίνουν ικανοποιητικά αποτελέσματα για την εκμάθησή του.



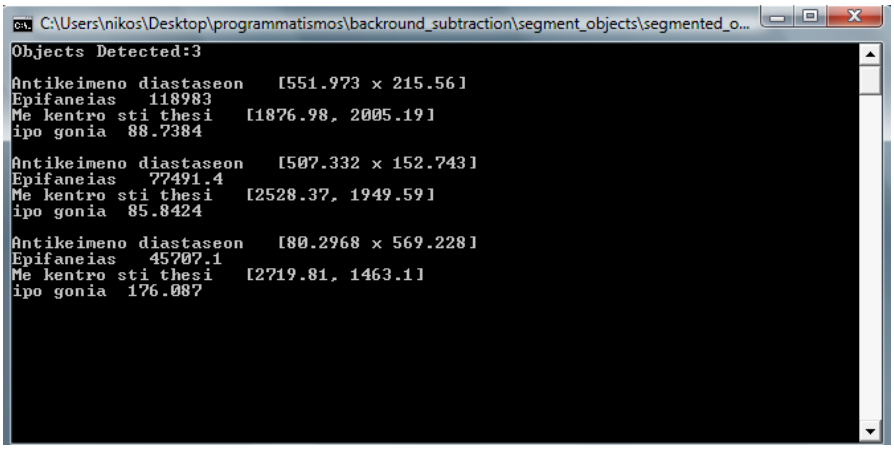
Σχήμα 4.4 1-4)Ξεχωριστές λήψεις παρασκηνίου

Στη συνέχεια τοποθετούνται τα τρία μεταλλικά κυλινδρικά αντικείμενα σε τυχαίες θέσεις και λαμβάνονται φωτογραφίες (Σχήματα 4.5-9 θέση 1). Κάθε λήψη εισάγεται ως όρισμα στο πρόγραμμα. Στο τέλος της επεξεργασίας λαμβάνεται μία σειρά εικόνων που στόχο έχει τον οπτικό έλεγχο των αποτελεσμάτων με σκοπό να εντοπίζεται τυχόν αστοχία και να διευθετείται αναλόγως. Η πρώτη εικόνα που εμφανίζεται στην οθόνη είναι αυτή της μάσκας όπως προκύπτει από τη διαδικασία αφαίρεσης του παρασκηνίου (Σχήματα 4.5-9 θέση 2) και η σύγκρισή της με την αρχική εικόνα (Σχήματα 4.5-9 θέση 3). Ακολουθεί εικόνα που παρουσιάζει τη μορφή του περιγράμματος κάθε ενιαίου σώματος της μάσκας με κόκκινο χρώμα και σε σύγκριση με την αρχική εικόνα (Σχήματα 4.5-9 θέση 4). Στη συνέχεια εμφανίζεται εικόνα που παρουσιάζει με ορθογώνιες παραλληλόγραμμες επιφάνειες χρώματος μπλε το προσαρμοσμένο σε κάθε περίγραμμα ορθογώνιο παραλληλόγραμμο σχήμα. Η παρουσίαση γίνεται σε σχέση με την αρχική εικόνα (Σχήματα 4.5-9 θεση 5). Τέλος ακολουθεί εικόνα που απεικονίζει το προσαρμοσμένο σε κάθε περίγραμμα ορθογώνιο παραλληλόγραμμο πλαίσιο και τη θέση του κέντρου του με μπλέ χρώμα καθώς και το προσαρμοσμένο σε κάθε περίγραμμα τόξο έλλειψης με πράσινο χρώμα (Σχήματα 4.5-9 θέση 6). Στο παράθυρο διεπαφής με το λογισμικό (Σχήματα 4.5-9 θέση 7) εμφανίζονται πληροφορίες που αφορούν το πλήθος των περιγραμμάτων που εντοπίστηκαν, τις διαστάσεις, το εμβαδόν και τις συντεταγμένες του κέντρου (σε pixel) κάθε προσαρμοσμένου ορθογώνιου πλαισίου καθώς και τη διεύθυνση κάθε περιγράμματος όπως προκύπτει από τον μέγιστο άξονα της προσαρμοσμένης έλλειψης. Η διεύθυνση υπολογίζεται από τη γωνία που σχηματίζει διάνυσμα αναφοράς το οποίο είναι κάθετο προς την μεγάλη πλευρά της εικόνας και με φορά προς επάνω (Σχήμα 4.5.8) με την ευθεία του μεγάλου άξονα της έλλειψης

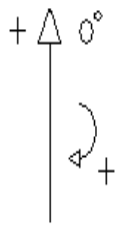
θεωρώντας την δεξιόστροφη φορά ως θετική. Ως σημείο (0,0) του συστήματος αναφοράς της θέσης των pixel θεωρείται το πρώτο πάνω αριστερά pixel της εικόνας.



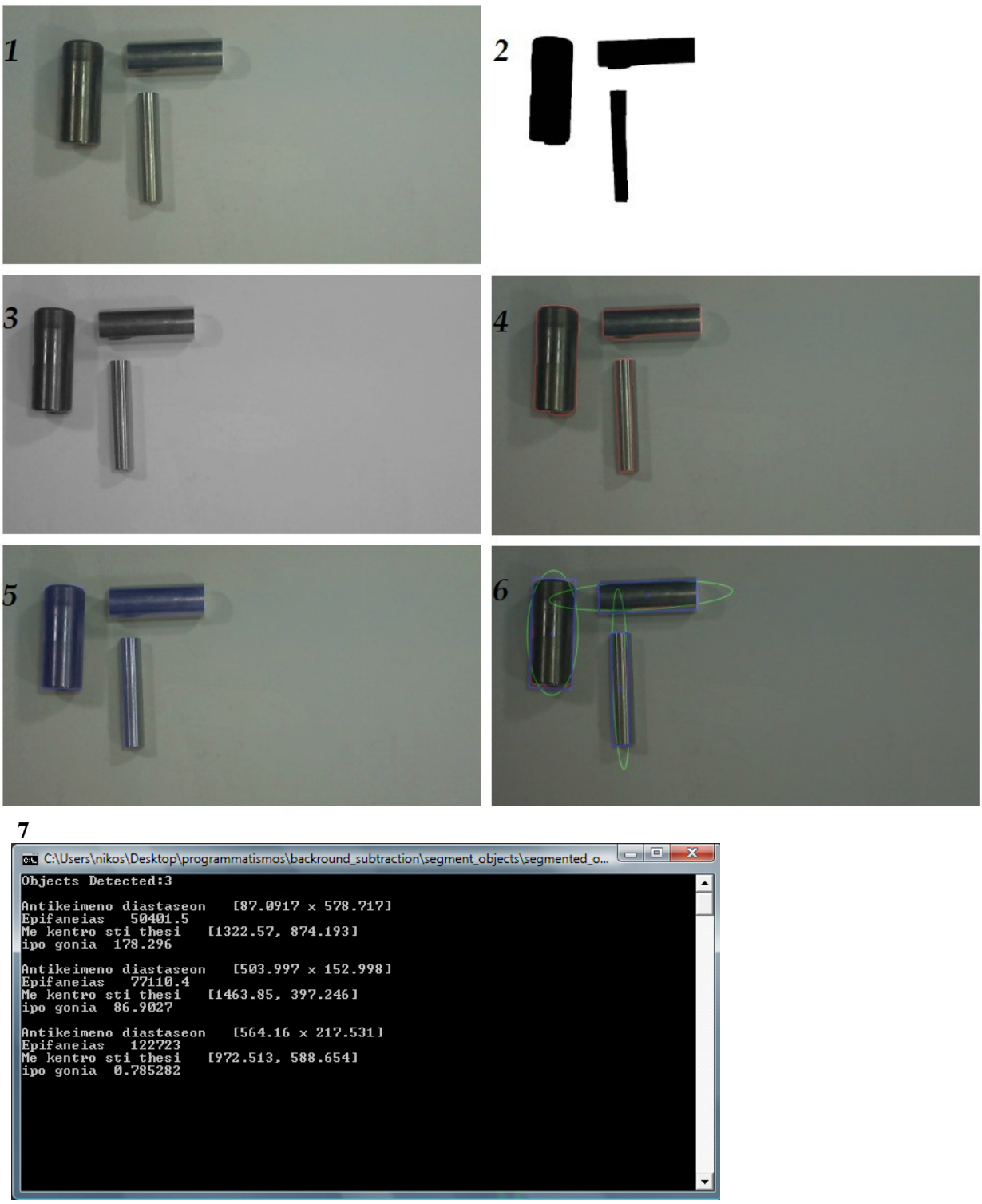
7



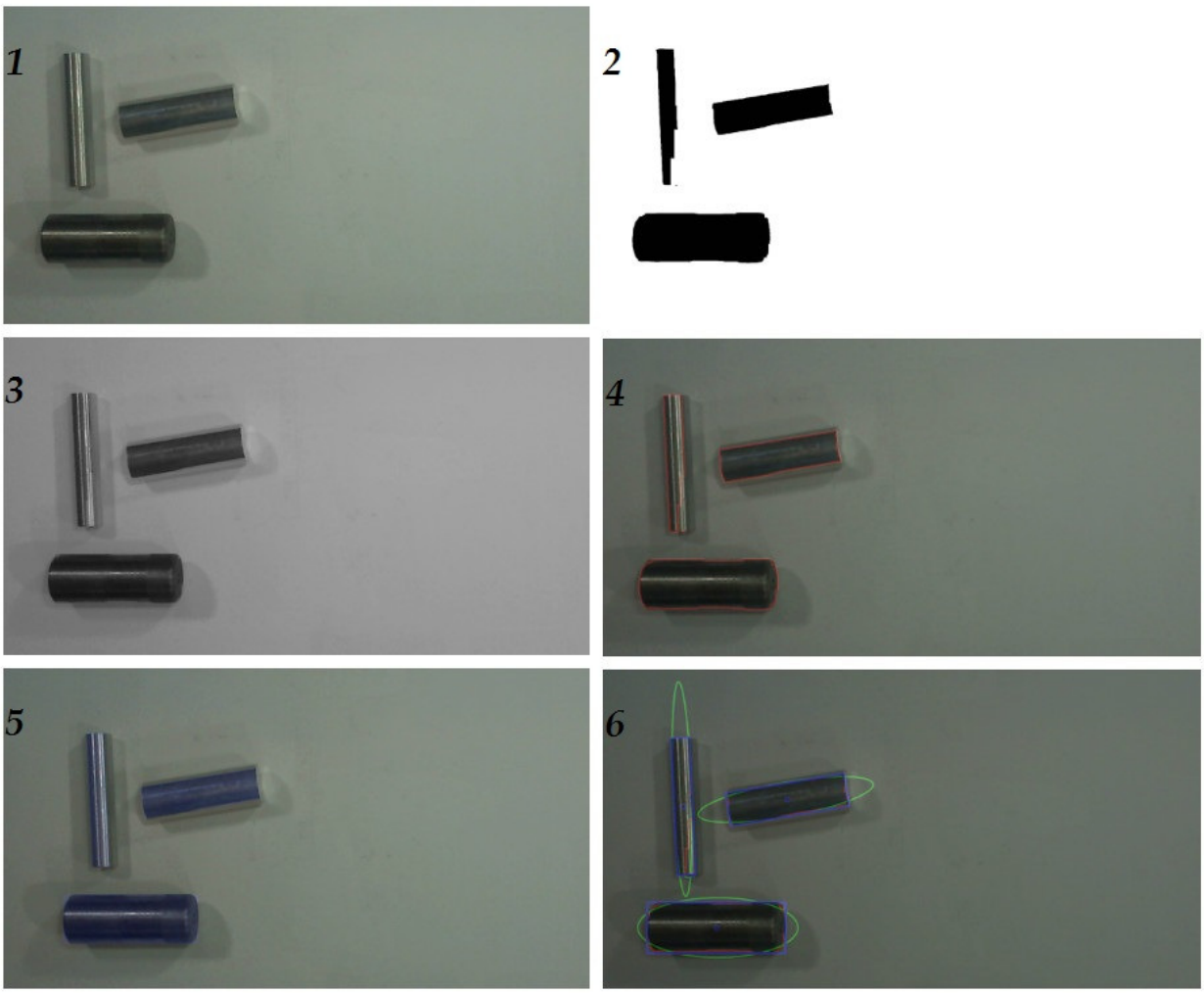
8



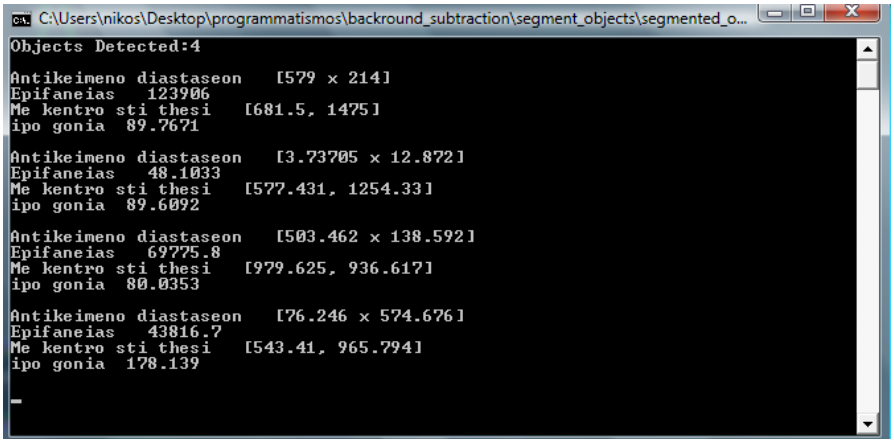
Σχήμα 4.5 Αποτελέσματα του λογισμικού 1)Αρχική εικόνα 2)Εικόνα μάσκας μετά την αφαίρεση του παρασκηνίου 3) Αρχική εικόνα σε συνδιασμό με τη μάσκα 4) Τα περιγράμματα που υπολογίστικαν με κόκκινο χρώμα πάνω στην αρχική εικόνα 5) Επιφάνεια προσαρμοσμένου πλαισίου με μπλέ χρώμα πάνω στην αρχική εικόνα. 6) Προσαρμοσμένο πλαίσιο και θέση του κέντρου του με μπλέ χρώμα, προσαρμοσμένη έλλειψη με πράσινο χρώμα πάνω στην αρχική εικόνα. 7) Πληροφορίες προσαρμοσμένου πλαισίου και ελλείψεως 8) Διάνυσμα που αναπαρηστά τη γωνία αναφοράς 0°



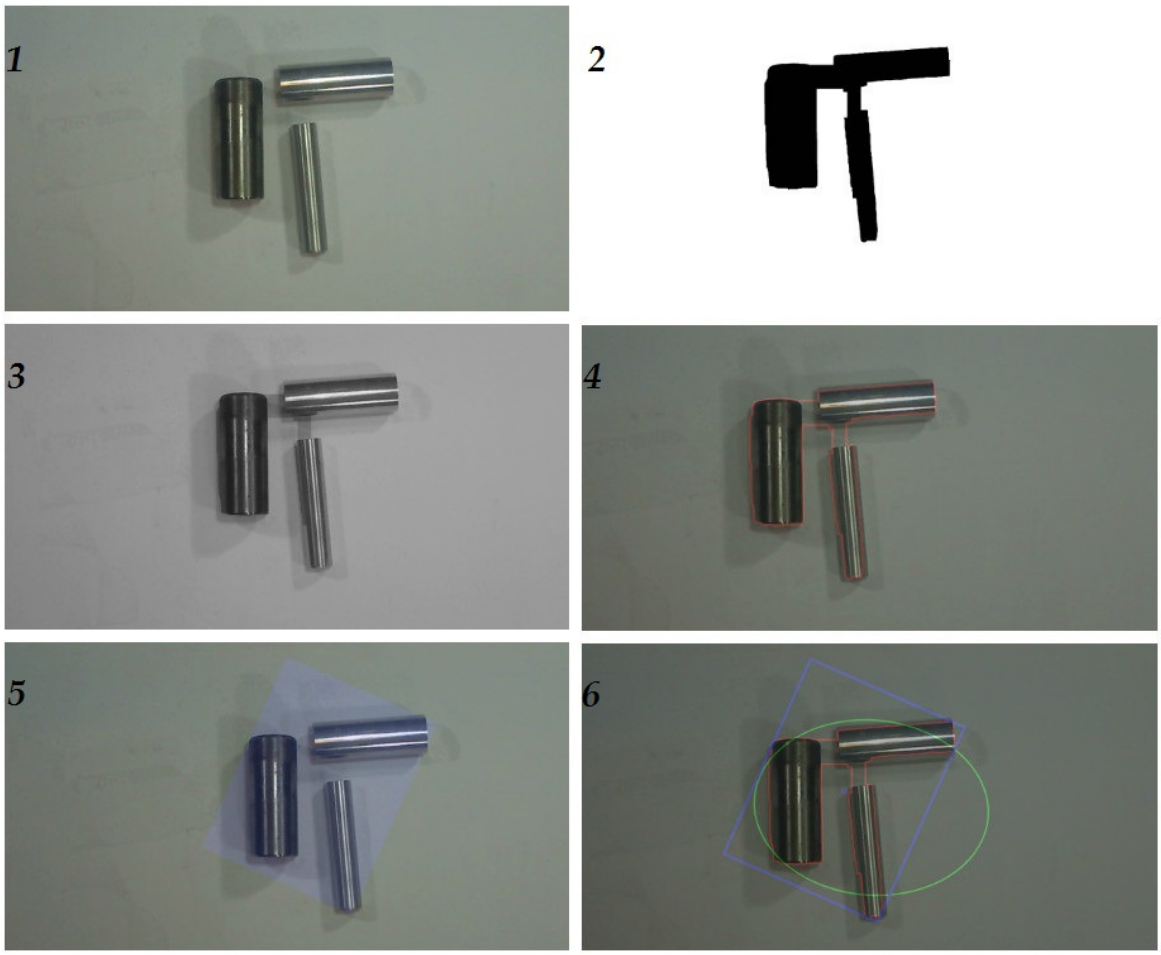
Σχήμα 4.6 Αποτελέσματα του λογισμικού 1)Αρχική εικόνα 2)Εικόνα μάσκας μετά την αφαίρεση του παρασκηνίου 3) Αρχική εικόνα σε συνδιασμό με τη μάσκα 4) Τα περιγράμματα που υπολογίστικαν με κόκκινο χρώμα πάνω στην αρχική εικόνα 5) Επιφάνεια προσαρμοσμένου πλαισίου με μπλέ χρώμα πάνω στην αρχική εικόνα. 6) Προσαρμοσμένο πλαίσιο και θέση του κέντρου του με μπλέ χρώμα, προσαρμοσμένη έλλειψη με πράσινο χρώμα πάνω στην αρχική εικόνα. 7) Πληροφορίες προσαρμοσμένου πλαισίου και ελλείψεως



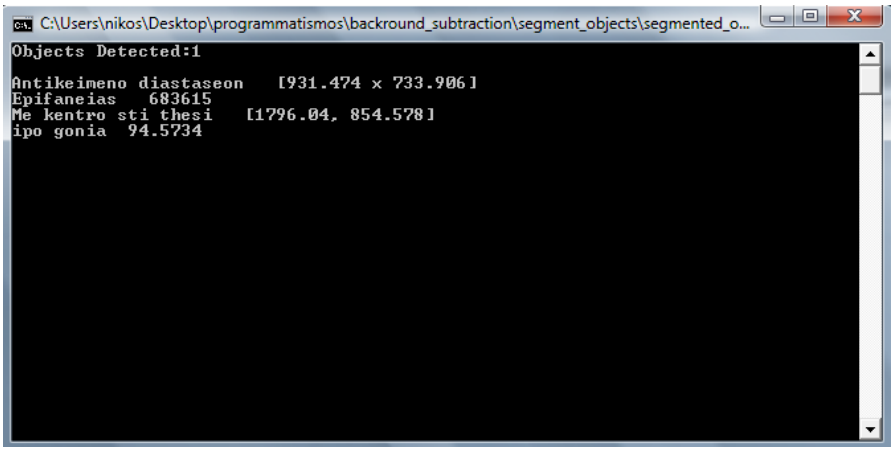
7



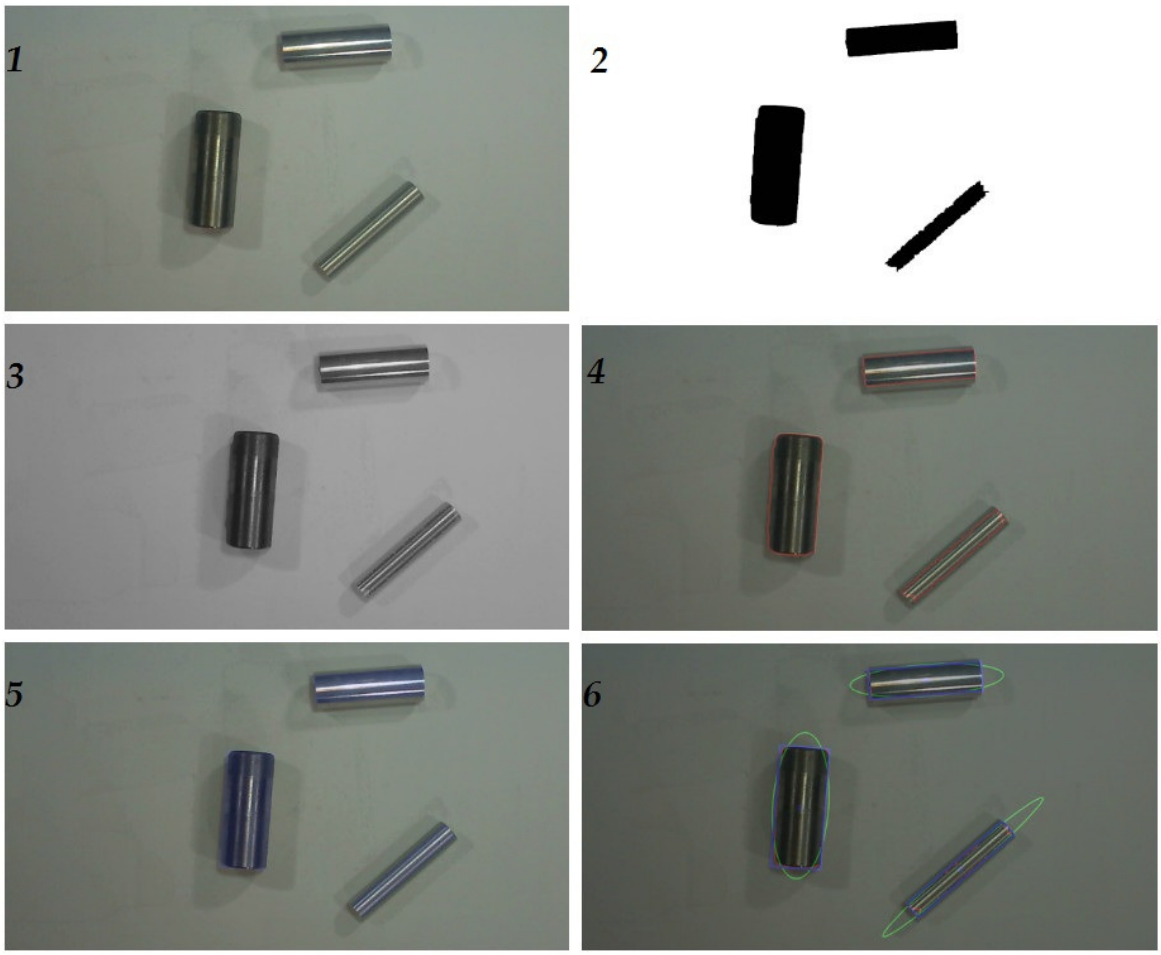
Σχήμα 4.7 Αποτελέσματα του λογισμικού 1)Αρχική εικόνα 2)Εικόνα μάσκας μετά την αφαίρεση του παρασκηνίου 3) Αρχική εικόνα σε συνδιασμό με τη μάσκα 4) Τα περιγράμματα που υπολογίστικαν με κόκκινο χρώμα πάνω στην αρχική εικόνα 5) Επιφάνεια προσαρμοσμένου πλαισίου με μπλέ χρώμα πάνω στην αρχική εικόνα. 6) Προσαρμοσμένο πλαίσιο και θέση του κέντρου του με μπλέ χρώμα, προσαρμοσμένη έλλειψη με πράσινο χρώμα πάνω στην αρχική εικόνα. 7) Πληροφορίες προσαρμοσμένου πλαισίου και ελλείψεως



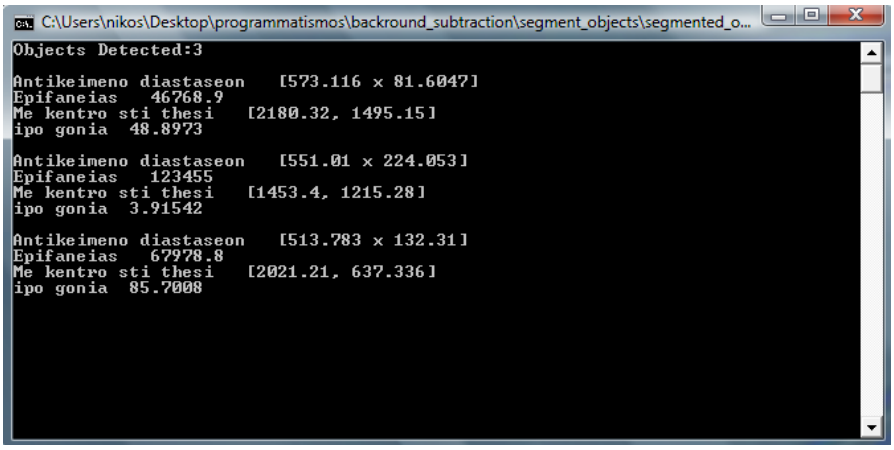
7



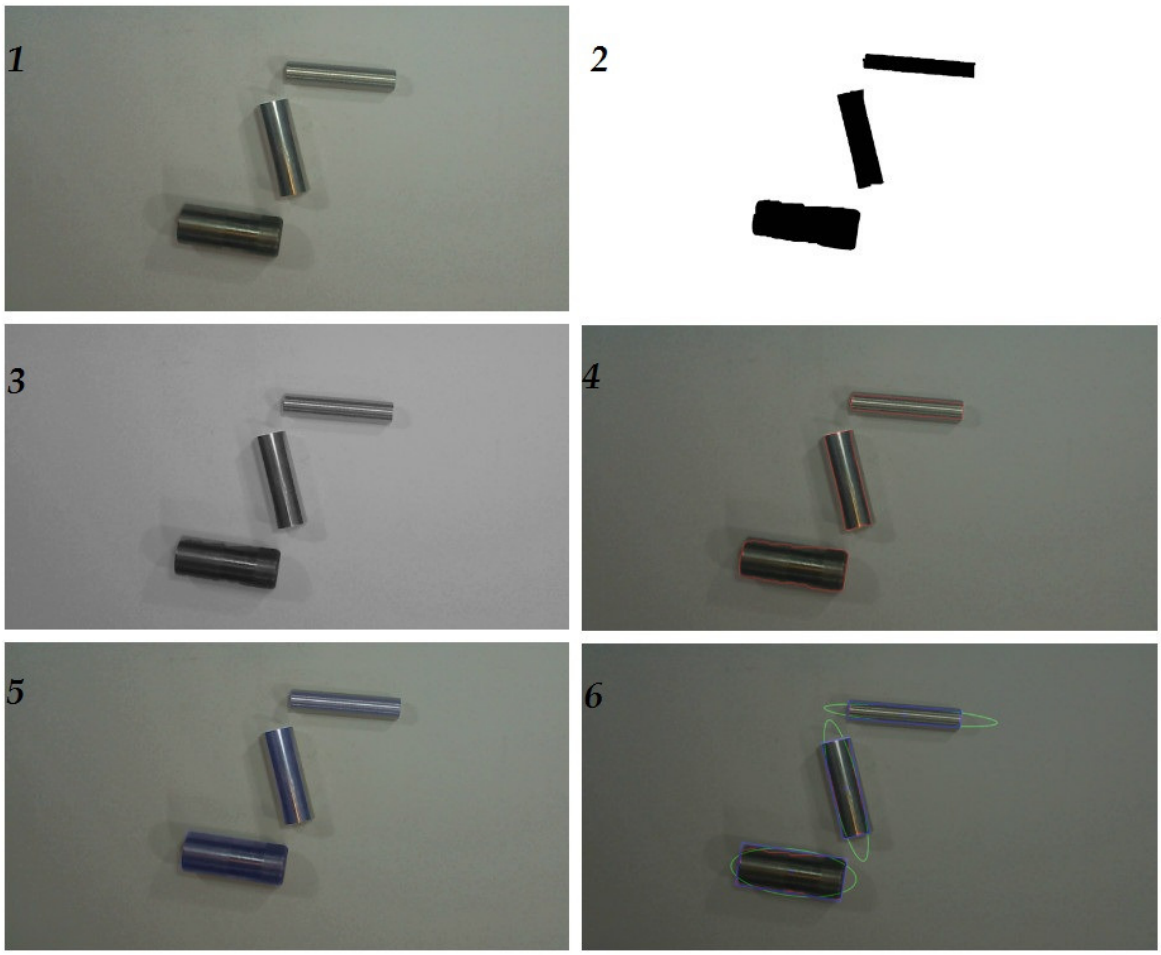
Σχήμα 4.8 Αποτελέσματα του λογισμικού 1)Αρχική εικόνα 2)Εικόνα μάσκας μετά την αφαίρεση του παρασκηνίου 3) Αρχική εικόνα σε συνδιασμό με τη μάσκα 4) Τα περιγράμματα που υπολογίστικαν με κόκκινο χρώμα πάνω στην αρχική εικόνα 5) Επιφάνεια προσαρμοσμένου πλαισίου με μπλέ χρώμα πάνω στην αρχική εικόνα. 6) Προσαρμοσμένο πλαίσιο και θέση του κέντρου του με μπλέ χρώμα, προσαρμοσμένη έλλειψη με πράσινο χρώμα πάνω στην αρχική εικόνα. 7) Πληροφορίες προσαρμοσμένου πλαισίου και ελλείψεως



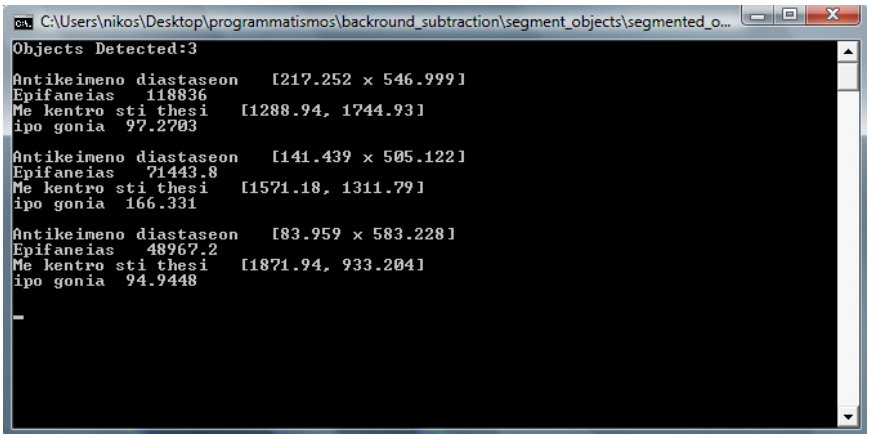
7



Σχήμα 4.9 Αποτελέσματα του λογισμικού 1)Αρχική εικόνα 2)Εικόνα μάσκας μετά την αφαίρεση του παρασκηνίου 3) Αρχική εικόνα σε συνδιασμό με τη μάσκα 4) Τα περιγράμματα που υπολογίστικαν με κόκκινο χρώμα πάνω στην αρχική εικόνα 5) Επιφάνεια προσαρμοσμένου πλαισίου με μπλέ χρώμα πάνω στην αρχική εικόνα. 6) Προσαρμοσμένο πλαίσιο και θέση του κέντρου του με μπλέ χρώμα, προσαρμοσμένη έλλειψη με πράσινο χρώμα πάνω στην αρχική εικόνα. 7) Πληροφορίες προσαρμοσμένου πλαισίου και ελλείψεως



7



Σχήμα 4.10 Αποτελέσματα του λογισμικού 1)Αρχική εικόνα 2)Εικόνα μάσκας μετά την αφαίρεση του παρασκηνίου 3) Αρχική εικόνα σε συνδιασμό με τη μάσκα 4) Τα περιγράμματα που υπολογίστικαν με κόκκινο χρώμα πάνω στην αρχική εικόνα 5) Επιφάνεια προσαρμοσμένου πλαισίου με μπλέ χρώμα πάνω στην αρχική εικόνα. 6) Προσαρμοσμένο πλαίσιο και θέση του κέντρου του με μπλέ χρώμα, προσαρμοσμένη έλλειψη με πράσινο χρώμα πάνω στην αρχική εικόνα. 7) Πληροφορίες προσαρμοσμένου πλαισίου και ελλείψεως

Πίνακας 4.1 Σύγκριση διαστάσεων προσαρμοσμένων πλαισίων που αντιστοιχούν στο ίδιο κυλινδρικό αντικείμενο σε κάθε πείπτωση λήψης των σχημάτων 5.5-10.

	Κύλινδρος 1	Κυλινδρος 2	Κύλινδρος 3
	Μήκος Πλάτος	Μήκος Πλάτος	Μήκος Πλάτος
Σχήμα 4.5	569,228 x 80,2968	507,332 x 152,743	551.973 x 215.56
Σχήμα 4.6	578,717 x 87,0917	503,997 x 152,998	564,16 x 217,531
Σχήμα 4.7	574,676 x 76,246	503,462 x 138,592	579 x 214
Σχήμα 4.8	-	-	-
Σχήμα 4.9	573,116 x 81.6047	513,798 x 132,31	551,01 x 224,053
Σχήμα 4.10	583,228 x 83,959	505,122 x 141.439	546,9 x 217,252

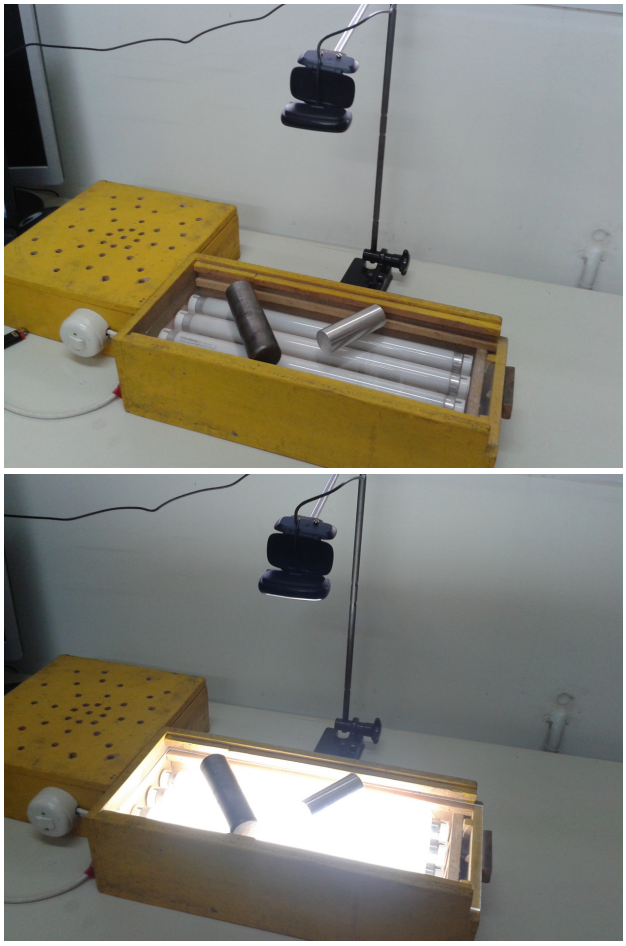
4.3.2 Δεύτερη πειραματική διάταξη – Αποτελέσματα

Η δεύτερη πειραματική διάταξη που χρησιμοποιήθηκε αποτελείται από ξύλινο ορθογώνιο κουτί με την άνω πλευρά του ανοιχτή και καλυμμένη με επιφάνεια γυαλιού διαστάσεων 42 x 15 cm (Σχήμα 4.11). Τα μεταλλικά κυλινδρικά αντικείμενα τοποθετούνται σε τυχαία θέση πάνω στο γυάλινο κάλλυμα του ξύλινου κουτιού. Η κάμερα στηρίζεται από ορθοστάτη σε οριζόντια θέση παράλληλα στο επίπεδο της γυάλινης επιφάνειας και σε ύψος 35 cm πάνω από το κέντρο της επιφάνειας. Ο φωτισμός της διάταξης προέρχεται από τέσσερεις λάμπες φθορίου ισχύος 20 W η κάθε μία οι οποίες είναι τοποθετημένες στο εσωτερικό της ξύλινης κατασκευής οπότε και η διεύθυνση του φωτισμού είναι από την κάτω πλευρά της γυάλινης επιφάνειας και προς τον φακό της κάμερας. Η συγκεκριμένη διάταξη φωτισμού έχει ως αποτέλεσμα ότι στην εικόνα που λαμβάνει η κάμερα ο χώρος γύρω από τα μεταλλικά αντικείμενα είναι έντονα φωτισμένος και τα μεταλλικά αντικείμενα καθώς μπαίνουν εμπόδιο στη διάδοση του φωτός αναπαριστώνται ως σκουρόχρωμες επιφάνειες. Λόγω της συγκεκριμένης έντονης αντίθεσης η σιλουέτα των αντικειμένων περιγράφεται με μεγάλη ευκρίνεια και επιπλεόν περιορίζονται σε μεγάλο βαθμό οι αντανακλάσεις χαρακτηριστικά και τα δύο πολύ σημαντικά για την αποτελεσματικότητα του λογισμικού προγράμματος καθώς ο εντοπισμός των περιγραμμάτων γίνεται με μεγαλύτερη ακρίβεια.

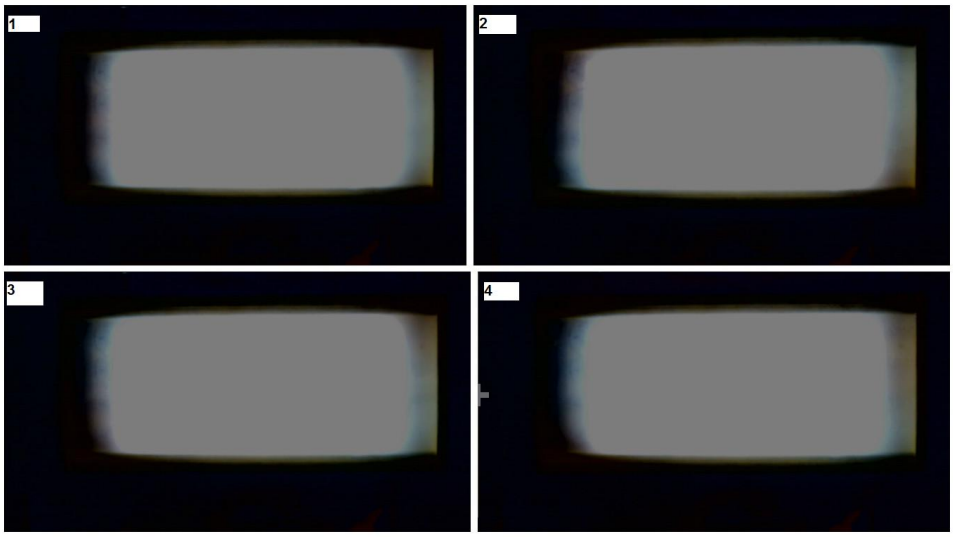
Κατά την ίδια λογική με την πρώτη πειραματική διάταξη αρχικά λαμβάνονται λήψεις του παρασκηνίου (Σχήμα 4.12) ως εκπαιδευτικές για την εκμάθηση του παρασκηνίου. Λόγω της απλότητας και της αμεταβλητότητας και στη συγκεκριμένη διάταξη του παρασκηνίου παρατηρήθηκε ότι τέσσερεις λήψεις δίνουν ικανοποιητικά αποτελέσματα για την εκμάθησή του.

Στη συνέχεια τοποθετούνται τα τρία μεταλλικά κυλινδρικά αντικείμενα σε τυχαίες θέσεις και λαμβάνονται φωτογραφίες (Σχήματα 4.13-16 θέση 1). Κάθε λήψη εισάγεται ως όρισμα στο πρόγραμμα. Στο τέλος της επεξεργασίας λαμβάνεται μία σειρά εικόνων που στόχο έχει τον οπτικό έλεγχο των αποτελεσμάτων με σκοπό να εντοπίζεται τυχόν αστοχία και να διευθετήται αναλόγως. Η πρώτη εικόνα που εμφανίζεται στην οθόνη είναι αυτή της μάσκας όπως προκύπτει από τη διαδικασία αφαίρεσης του παρασκηνίου (Σχήματα 4.13-16 θέση 2). Ακολουθεί εικόνα που παρουσιάζει τη μορφή του περιγράμματος κάθε ενιαίου σώματος της μάσκας με κόκκινο χρώμα και σε σύγκριση με την αρχική εικόνα (Σχήματα 4.13-16 θέση 3). Τέλος ακολουθεί εικόνα που απεικονίζει το προσαρμοσμένο σε κάθε περίγραμμα ορθογώνιο παραλληλόγραμμο πλαίσιο και τη θέση του κέντρου του με μπλέ χρώμα καθώς και την προσαρμοσμένο σε κάθε περίγραμμα τόξο έλλειψης με πράσινο χρώμα

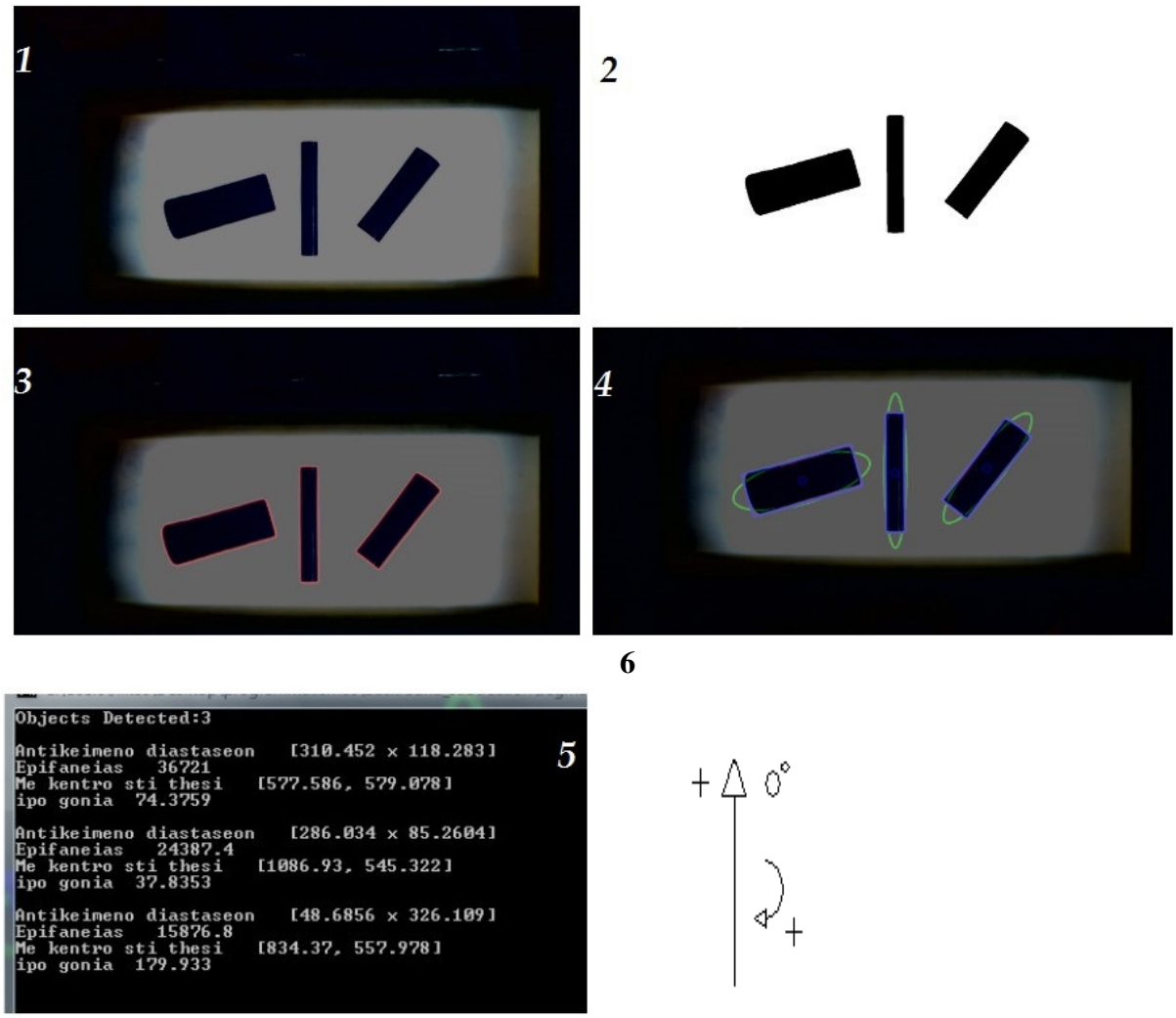
(Σχήματα 4.13-16 θέση 4). Στο παράθυρο διεπαφής με το λογισμικό (Σχήματα 4.13-16 θέση 5) εμφανίζονται πληροφορίες που αφορούν το πλήθος των περιγραμμάτων που εντοπίστηκαν, τις διαστάσεις, το εμβαδόν και τις συντεταγμένες του κέντρου (σε pixel) κάθε προσαρμοσμένου ορθογώνιου πλαισίου καθώς και τη διεύθυνση κάθε περιγράμματος όπως προκύπτει από τον μέγιστο άξονα της προσαρμοσμένης έλλειψης. Η διεύθυνση υπολογίζεται από τη γωνία που σχηματίζει διάνυσμα κάθετο προς την μεγάλη πλευρά της εικόνας και με φορά προς επάνω (Σχήμα 4.13 θέση 6) με την ευθεία του μεγάλου άξονα της έλλειψης θεωρώντας την δεξιόστροφη φορά ως θετική. Ως σημείο (0,0) του συστήματος αναφοράς της θέσης των pixel θεωρείται το πρώτο πάνω αριστερά pixel της εικόνας.



Σχήμα 4.11 Δεύτερη πειραματική διάταξη. Σημαντικό χαρακτηριστικό της είναι ότι η διεύθυνση φωτισμού των μεταλλικών αντικειμένων είναι από την επιφάνεια στην οποία εφάπτονται τα μεταλλικά αντικείμενα και προς την κάμερα.

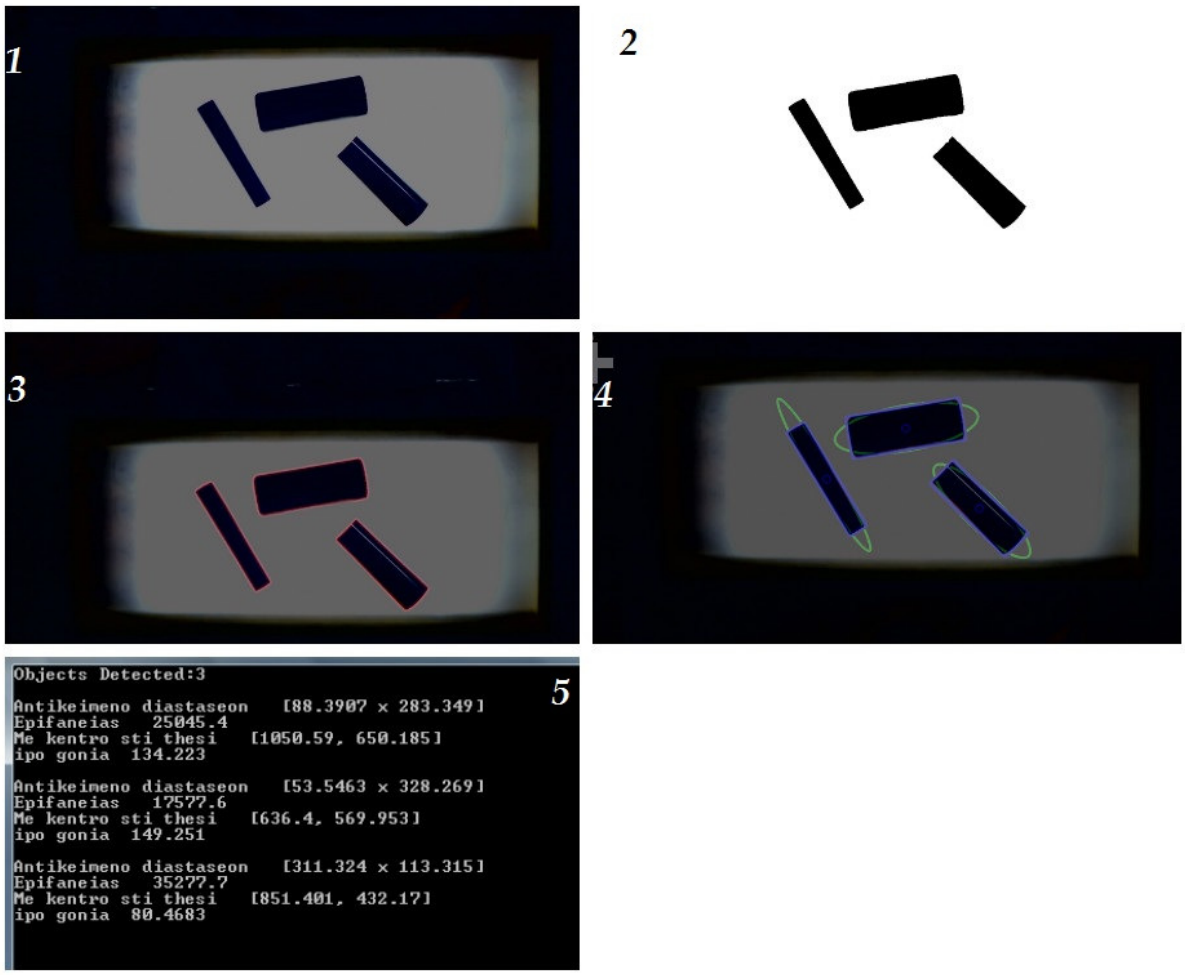


Σχήμα 4.12 1-4)Ξεχωριστές λήψεις παρασκηνίου

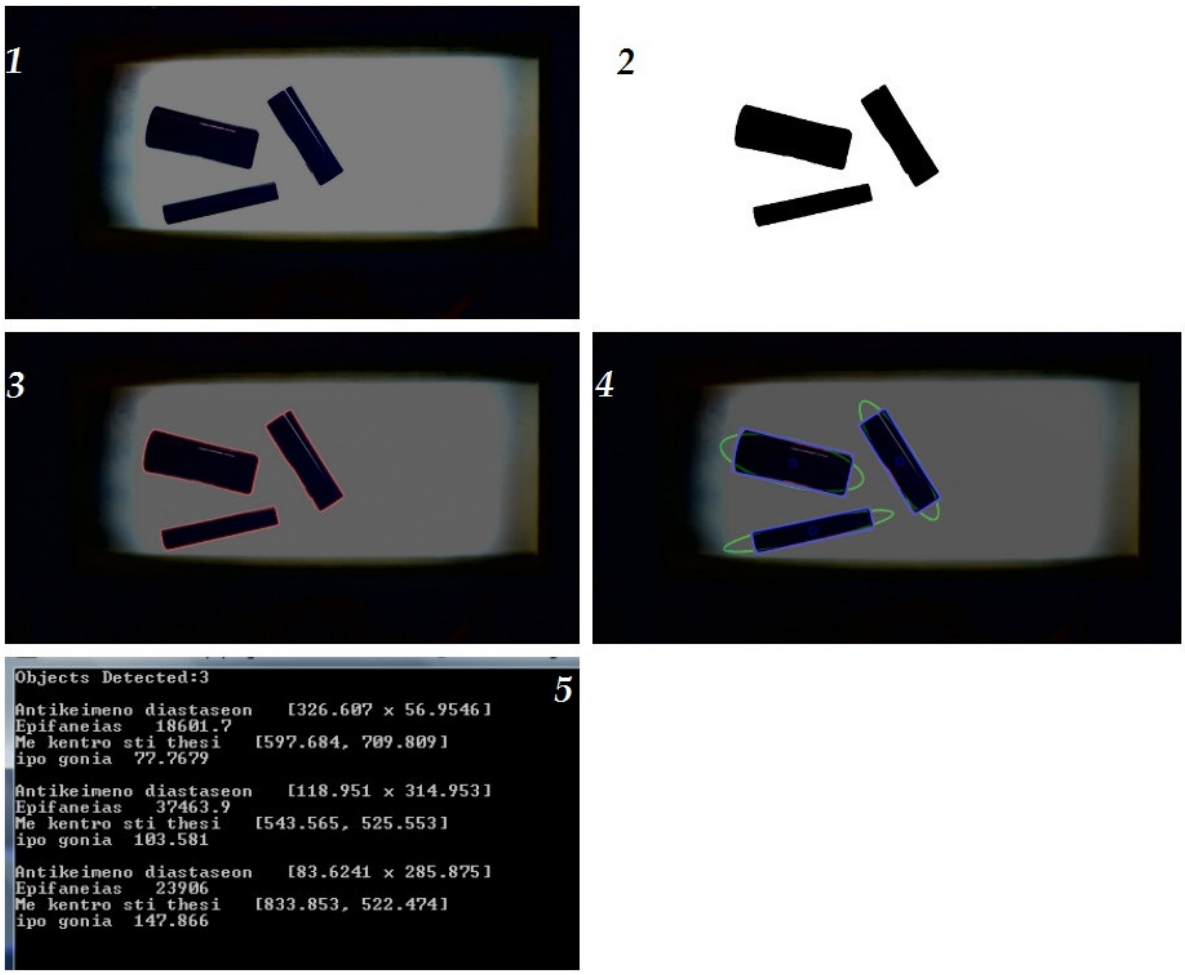


Σχήμα 4.13 Αποτελέσματα του λογισμικού 1)Αρχική εικόνα 2)Εικόνα μάσκας μετά την αφαίρεση του παρασκηνίου 3) Τα περιγράμματα που υπολογίστικαν, με κόκκινο χρώμα πάνω στην αρχική εικόνα 4) Προσαρμοσμένο πλαίσιο και θέση του κέντρου του με μπλέ χρώμα, προσαρμοσμένη έλλειψη

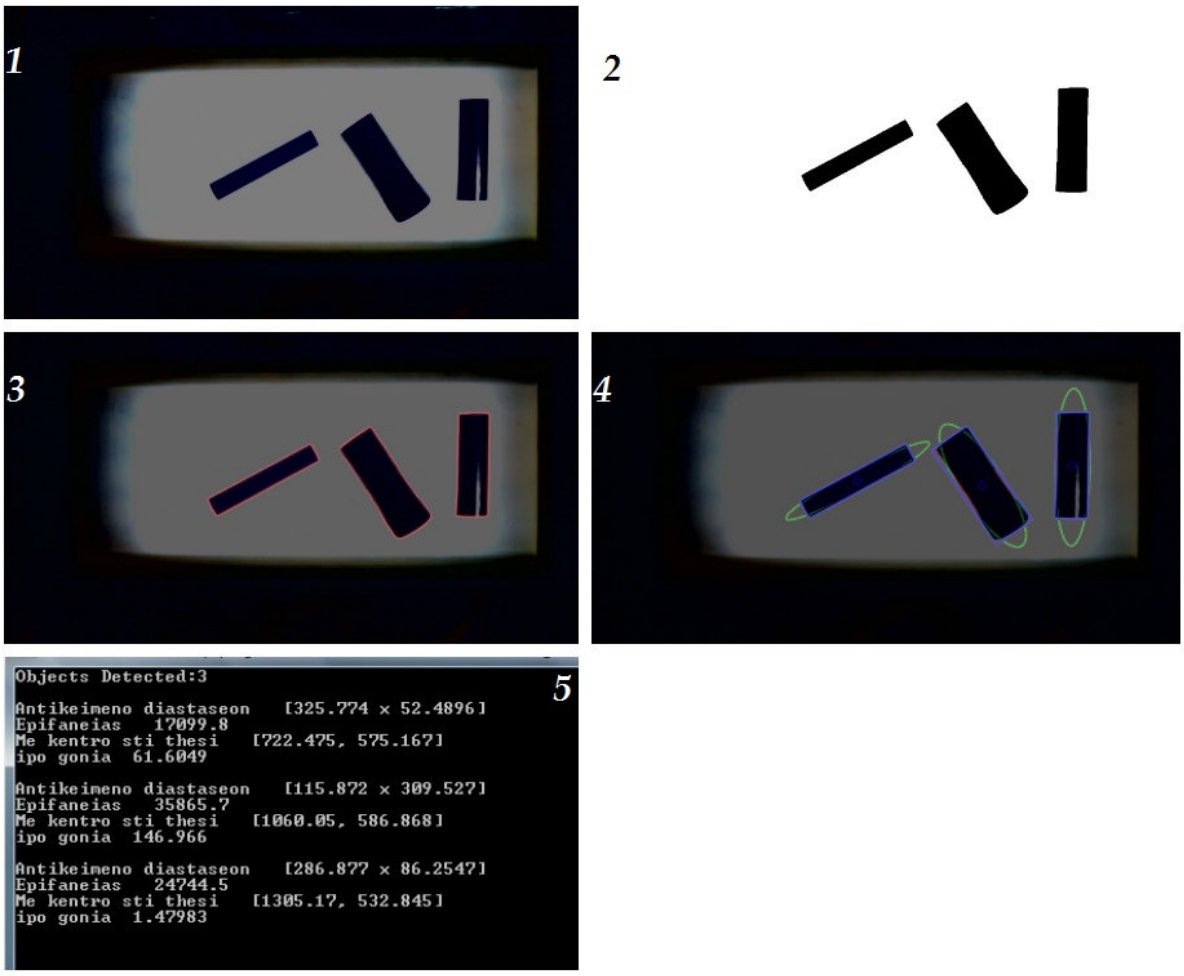
με πράσινο χρώμα πάνω στην αρχική εικόνα. 5) Πληροφορίες προσαρμοσμένων πλαισίων και ελλείψεων 6) Διάνυσμα που αναπαρηστά τη γωνία αναφοράς 0°



Σχήμα 4.14 Αποτελέσματα του λογισμικού 1)Αρχική εικόνα 2)Εικόνα μάσκας μετά την αφαίρεση του παρασκηνίου 3) Τα περιγράμματα που υπολογίστικαν, με κόκκινο χρώμα πάνω στην αρχική εικόνα 4) Προσαρμοσμένο πλαίσιο και θέση του κέντρου του με μπλέ χρώμα, προσαρμοσμένη έλλειψη με πράσινο χρώμα πάνω στην αρχική εικόνα. 5) Πληροφορίες προσαρμοσμένων πλαισίων και ελλείψεων



Σχήμα 4.15 Αποτελέσματα του λογισμικού 1)Αρχική εικόνα 2)Εικόνα μάσκας μετά την αφαίρεση του παρασκηνίου 3) Τα περιγράμματα που υπολογίστικαν, με κόκκινο χρώμα πάνω στην αρχική εικόνα 4) Προσαρμοσμένο πλαίσιο και θέση του κέντρου του με μπλέ χρώμα, προσαρμοσμένη έλλειψη με πράσινο χρώμα πάνω στην αρχική εικόνα. 5) Πληροφορίες προσαρμοσμένων πλαισίων και ελλείψεων



Σχήμα 4.16 Αποτελέσματα του λογισμικού 1)Αρχική εικόνα 2)Εικόνα μάσκας μετά την αφαίρεση του παρασκηνίου 3) Τα περιγράμματα που υπολογίστικαν, με κόκκινο χρώμα πάνω στην αρχική εικόνα 4) Προσαρμοσμένο πλαίσιο και θέση του κέντρου του με μπλέ χρώμα, προσαρμοσμένη έλλειψη με πράσινο χρώμα πάνω στην αρχική εικόνα. 5) Πληροφορίες προσαρμοσμένων πλαισίων και ελλείψεων

Πίνακας 4.2 Σύγκριση διαστάσεων προσαρμοσμένων πλαισίων που αντιστοιχούν στο ίδιο κυλινδρικό αντικείμενο σε κάθε πείπτωση λήψης των σχημάτων 4.13-16.

	Κύλινδρος 1	Κυλινδρος 2	Κύλινδρος 3
Σχήμα 4.13	326,109 x 48,6856	286,034 x 85,2604	310,452 x 118,283
Σχήμα 4.14	328,269 x 53,5463	283,349 x 88,3907	311,324 x 113,315
Σχήμα 4.15	326,607 x 56,9546	285,875 x 83,6241	314,953 x 118,951
Σχήμα 4.16	325,774 x 52,4896	286,877 x 86,2547	309,527 x 115,872

4.3.3 Σχολιασμός αποτελεσμάτων – Σφάλματα

Στην πρώτη πειραματική διάταξη παρατηρείται αρχικά ότι η διαδικασία αφαίρεσης του παρασκηνίου είναι αποτελεσματική με εξαίρεση σημεία έντονων αντανακλάσεων του φωτός και έντονων σκιάσεων. Οι αντανακλάσεις εντοπίζονται με τη μορφή ανοιχτόχρωμων

λωρίδων (σε σχέση με το υπόλοιπο χρώμα του αντικειμένου) κατά μήκος του αντικειμένου. Στην περίπτωση των έντονων αντανακλάσεων υπάρχει η περίπτωση το αντικείμενο να αντιμετωπιστεί ως δύο ή και παραπάνω ξεχωριστά αντικείμενα από το λογισμικό εξαιτίας των έντονων χρωματικών μεταβολών κατά πλάτος του σώματος του (Σχήμα 4.7). Το φαινόμενο αυτό στις ακραίες περιπτώσεις δεν αντιμετωπίζεται ούτε με τη διαδικασία φιλτραρίσματος closing που ενοποιεί τα ασύνδετα μεταξύ τους μέρη ενός σώματος μετά την αφαίρεση του παρασκηνίου.

Στην περίπτωση των σκιάσεων που εμφανίζονται κυρίως γύρω από τα όρια των αντικειμένων υπάρχει η περίπτωση όταν τα αντικείμενα βρίσκονται πλησίον το ένα του άλλου να δημιουργείται "βαθιά" σκιά λόγω της συμβολής των ατομικών σκιάσεων με αποτέλεσμα της δημιουργία σκουρόχρωμης χρωματικής λωρίδας ανάμεσά τους (Σχήμα 4.8). Το φαινόμενο δυσχεραίνεται και με το φιλτράρισμα της διαδικασίας closing οπότε το λογισμικό αναγνωρίζει το συνδιασμό αντικειμένων και έντονων σκιάσεων ως ένα αντικείμενο.

Γενικά σε όλες τις περιπτώσεις της πρώτης πειραματικής διάταξης δημιουργούνται αστοχίες και λάθος εκτιμήσεις λόγω των αντανακλάσεων και των σκιάσεων όπως παρατηρείται και στον πίνακα 4.1 όπου φαίνονται οι σημαντικές αποκλίσεις στις εκτιμήσεις των διαστάσεων των αντικειμένων στις περιπτώσεις των εικόνων 4.5-10. Το ευμεγέθες φίλτρο Kernel (60x60 pixel) της διαδικασίας closing που χρησιμοποιείται στην πρώτη πειραματική διάταξη για την αντιμετώπηση των αστοχιών λόγω αντανακλάσεων και σκιάσεων οδηγεί σε απώλεια ακρίβειας των αποτελεσμάτων. Παρόλα αυτά στην πλειονότητα των περιπτώσεων όπου τα αντικείμενα δεν βρίσκονται αρκετά πλησίον το ένα του άλλου το λογισμικό δίνει ικανοποιητικά αποτελέσματα.

Στην περίπτωση της δεύτερης πειραματικής διάταξης παρατηρείται ότι οι αστοχίες λόγω αντανακλάσεων και σκιάσεων έχουν εξαλειφθεί σε μεγάλο βάθμό. Αυτό οφείλεται στην έντονη αντίθεση μεταξύ των αντικειμένων και του παρασκηνίου. Η εμπόδιση διάδοσης του φωτός από τα σώματα τους προσδίδει κοινή σκούρα χρωματική απόχρωση εξαλείφωντας τις όποιες χρωματικές αλλάγες ειτε κατασκευαστικα ειτε από αντανακλάσεις εξωτερικού φωτός. Επιπλεόν ο έντονος φωτισμός ανάμεσα στα αντικείμενα εξαλείφει το πρόβλημα των σκιάσεων που αντιμετώπιζε η πρώτη πειραματική διάταξη. Στον πίνακα 4.2 παρατηρείται μικρή σχετικά απόκλιση στις εκτιμήσεις των διαστάσεων των αντικειμένων στις περιπτώσεις των εικόνων 4.13-16. Γενικά τα αποτελέσματα του λογισμικού στην πλειονότητα των περιπτώσεων στη δεύτερη πειραματική διάταξη είναι ικάνης ακρίβειας.

Εκτός από τα σφάλματα – αστοχίες που οφείλονται στην επίδραση του φωτισμού υπάρχουν όπως και σε κάθε μετρητική διάταξη σύνολο πηγών σφαλμάτων, η κάθε μία με την ιδιαίτερη συμπεριφορά της. Μία πηγή σφάλματος είναι το μη σωστό κεντράρισμα της κάμερας πάνω απο το κέντρο της ενδιαφερόμενης επιφάνειας γεγονός που μπορεί να οδηγήσει σε λάθος αποτελέσματα λόγω αλλοιωμένης προβολής των αντικειμένων. Επιπλέον λόγω της δυσδιάστατης προβολής τρισδιάστατων αντικειμένων όταν τα αντικείμενα απομακρίνονται από κέντρο προβολής της κάμερας η λήψη περιλαμβάνει και μέρος της πλαινής κυκλικής επιφάνειάς τους οπότε αλλάζει και η εκτιμώμενη επιφάνεια του προσαρμοσμένου πλαισίου.

Τέλος μία σημαντική παρατήρηση είναι ότι η εκτίμηση της γωνίας των αντικειμένων που γίνεται μέσω προσαρμοσμένου σε κάθε περίγραμμα τόξου έλλειψης δεν δείχνει την πιο μικρή γώνία ως προς το σύστημα αναφοράς αλλά τη γωνία που σχηματίζει το διάνυσμα

αναφοράς με τον μέγιστο άξονα της έλλειψης έτσι όπως θα τον "συναντήσει" στρεφόμενο κατά την ωρολογιακή φορά.

4.4 ΒΙΒΛΙΟΓΡΑΦΙΑ

[4.1] Α.Καραγιαννίδης "Μετρηση γωνίας περιστροφής άξονα βιομηχανικού ρομπότ με χρήση κάμερας χαμηλού κόστους". Διπλωματική εργασία, σχολή Μηχανολόγων Μηχανικών Ε.Μ.Π , Αθήνα, Οκτώβριος 2012

ΚΕΦΑΛΑΙΟ 5

ΣΥΜΠΕΡΑΣΜΑΤΑ

5.1 ΣΥΝΟΨΗ ΤΗΣ ΕΡΓΑΣΙΑΣ ΚΑΙ ΤΩΝ ΑΠΟΤΕΛΕΣΜΑΤΩΝ

Στην παρούσα διπλωματική εργασία παρουσιάστηκε ένα μοντέλο οπτικού προσδιορισμού της θέσης κυλινδρικών μεταλλικών αντικειμένων πάνω σε επίπεδη επιφάνεια. Το μοντελο αυτό λειτουργεί εφαρμόζοντας μεθόδους τη τεχνητής όρασης, της μηχανικής μάθησης και της ψηφιακής επεξεργασίας εικόνας στις ψηφιακές φωτογραφίες που λαμβάνονται από τα αντικείμενα. Ο εντοπισμός της θέσης γίνεται από την εύρεση των περιγραμμάτων των αντικειμένων. Στο περίγραμμα μετά τον εντοπισμό του προσαρμόζεται ορθογώνιο παραλληλόγραμο πλαίσιο και τόξο έλλειψης από τα οποία λαμβάνεται η θέση του κέντρου της δυσδιάστατης προβολής του μεταλλικού αντικειμένου, η διεύθυνσή του και οι διαστάσεις του σε pixel. Για τον ακριβή εντοπισμό των περιγραμμάτων εφαρμόζεται διαδικασία αναγνώρισης και απομάκρυνσης του παρασκηνίου ώστε να απομονώνονται τα αντικείμενα.

Το λογισμικό εφαρμόστηκε σε δύο πειραματικές διατάξεις. Στην πρώτη διάταξη τα αντικείμενα ήταν τοποθετημένα πάνω σε λευκή επιφάνεια και φωτιζόντουσαν από φωτιστικά τοποθετημένα περιμετρικά. Από τα αποτελέσματα παρατηρήθηκε οτί υπάρχει σχετική ανακρίβεια στον εντοπισμό των περιγραμμάτων που οφείλεται στις σκιάσεις και στις αντανακλάσεις του φωτός πάνω στα αντικείμενα. Η ανακρίβεια είναι περιορισμένη όταν τα αντικείμενα είναι απομακρυσμένα μεταξύ τους ώστε να μην δημιουργούνται “βαθιές” σκιές και όταν δεν υπάρχει έντονη αντανάκλαση. Σε περιπτώσεις “βαθιάς” σκιάς και έντονων αντανακλάσεων το λογισμικό δίνει λανθασμένα αποτελέσματα.

Στην δεύτερη πειραματική διάταξη τα αντικείμενα είναι τοποθετημένα πάνω σε επίπεδη γυάλινη επιφάνεια κάτω από την οποία υπάρχει σύστημα φωτισμού. Τα αντικείμενα βρίσκονται ανάμεσα στο σύστημα φωτισμού και την κάμερα. Με αυτή τους την τοποθέτηση μπαίνουν εμπόδιο στη διάδοση του φωτός και λαμβάνονται ως σκουρόχρωμες επιφάνειες από την κάμερα σε ένα έντονο φωτισμένο περιβάλλον. Με αυτή τη διάταξη εξαλείφωνται οι χρωματικές παραλλαγές στην επιφάνεια των αντικειμένων καθώς και οι σκιάσεις με αποτέλεσμα τα αποτελέσματα να είναι αυξημένης ακρίβειας.

5.2 ΣΥΜΠΛΗΡΩΣΕΙΣ

Σημαντικές συμπληρώσεις που πρέπει να γίνουν στο λογισμικό είναι οι ακόλουθες:

- Μετασχηματισμός των δεδομένων του αποτελέσματος από το επίπεδο της εικόνας στον πραγματικό κόσμο ως προς σύστημα αναφοράς που έχει οριστεί από τον χρήστη πάνω στην επίπεδη επιφάνεια. Ο μετασχηματισμός αυτός μπορεί να υλοποιηθεί με τη χρήση του πίνακα κάμερας όπως προκύπτει από τη διαδικασία καλιμπραρίσματος.
- Κατηγοριοποίηση ομοίων εντοπισμένων αντικειμένων βάση των αποτελεσμάτων του λογισμικού που αφορούν τις διαστάσεις τους.
- Αποθήκευση αποτελεσμάτων στη μνήμη του υπολογιστή σε μορφή αρχείου xml.

ΠΑΡΑΡΤΗΜΑ Α

ΠΗΓΑΙΟΣ ΚΩΔΙΚΑΣ ΕΦΑΡΜΟΓΗΣ

```
#include <opencv2/opencv.hpp>
#include <stdio.h>
#include <string>
#include <iostream>

using namespace cv;
using namespace std;

// ΟΡΙΣΜΟΣ GLOBAL ΠΑΡΑΜΕΤΡΩΝ

Mat  tmp_frame, tmp_frame_gray, bgmask, out_frame, open_frame, close_bgmask;

void contours();

 int main(int argc, char** argv)
{

// ΜΕΤΑΦΟΡΑ  ΠΙΝΑΚΑ ΚΑΜΕΡΑΣ ΚΑΙ ΣΥΝΤΕΛΕΣΤΩΝ ΠΑΡΑΜΟΡΦΩΣΗΣ ΑΠΟ
ΚΑΛΙΜΠΡΆΡΙΣΜΑ ΚΑΜΕΡΑΣ // ΣΕ ΜΕΤΑΒΛΗΤΕΣ ΤΟΥ ΛΟΓΙΣΜΙΚΟΥ

   string filename="out_camera_data.xml";
   FileStorage fs(filename,FileStorage::READ);
   if(!fs.isOpened())
   return false;
   Mat CameraMatrix,DistCoeffs;
   fs["Camera_Matrix"]>> CameraMatrix;
   fs["Distortion_Coefficients"]>>DistCoeffs;

 // ΕΠΙΛΟΓΗ ΠΑΡΑΜΕΤΡΩΝ ΓΙΑ ΤΑ ΦΙΛΤΡΑ

   Mat element1 = getStructuringElement(0,Size( 41, 41 ), Point( 20, 20) );
   Mat element2 = getStructuringElement(0,Size( 101, 101 ), Point( 50, 50) );

 // ΕΠΙΛΟΓΗ ΠΑΡΑΜΕΤΡΩΝ ΓΙΑ ΤΗΝ ΔΙΑΔΙΚΑΣΙΑ ΑΦΑΙΡΕΣΗΣ ΠΑΡΑΣΚΗΝΙΟΥ ΚΑΙ
ΕΝΕΡΓΟΠΟΙΗΣΗ
 // ΛΕΙΤΟΥΡΓΙΑΣ ΕΚΜΑΘΗΣΗΣ

   BackgroundSubtractorMOG bgsubtractor;
   bgsubtractor.set("noiseSigma", 24);
   bgsubtractor.set("backgroundRatio", 0.9);
   bgsubtractor.set("nmixtures", 5);
   bool update_bg_model = true;

 // ΕΝΑΡΞΗ ΕΠΑΝΑΛΗΤΙΚΗΣ ΔΙΑΔΙΚΑΣΙΑΣ ΕΚΜΑΘΗΣΗΣ ΠΑΡΑΣΚΗΝΙΟΥ
```

```
for(int i=0;i<9;i++) // i<n ΟΠΟΥ n = ΠΛΗΘΟΣ ΕΙΚΟΝΩΝ ΕΚΠΑΙΔΕΥΣΗΣ + ΕΙΚΟΝΑ ΜΕ
                     //ΑΝΤΙΚΕΙΜΕΝΑ
{
   Mat original_frame= imread(argv[i+1],1);

// ΑΠΟΠΑΡΑΜΟΡΦΩΣΗ ΚΑΘΕ ΕΙΚΟΝΑΣ

   undistort(original_frame,tmp_frame,CameraMatrix,DistCoeffs);
   if( !tmp_frame.data )
      break;

// ΕΦΑΡΜΟΓΗ ΦΙΛΤΡΟΥ  ΣΕ ΚΑΘΕ ΕΙΚΟΝΑ

   morphologyEx(tmp_frame, open_frame,2, element1);

   bgsubtractor(tmp_frame, bgmask, update_bg_model ? -1 : 0);

// ΑΠΕΝΕΡΓΟΠΟΙΗΣΗ ΔΙΑΔΙΚΑΣΙΑΣ ΕΚΜΑΘΗΣΗΣ  ΩΣΤΕ ΣΤΗΝ ΤΕΛΕΥΤΑΙΑ ΕΠΑΝΑΛΗΨΗ ΝΑ
//ΑΦΑΙΡΕΘΕΙ ΤΟ ΠΑΡΑΣΚΗΝΙΟ ΠΟΥ ΕΧΕΙ ΜΑΘΕΥΤΕΙ ΚΑΙ ΝΑ ΜΕΙΝΟΥΝ ΤΑ ΑΝΤΙΚΕΙΜΕΝΑ
// ΤΟΥ  ΠΡΟΣΚΗΝΙΟΥ

   if( i>7 ) // i>n-2, ΒΛΕΠΕ ΒΡΟΓΧΟ for
   {
      update_bg_model = false;
   }
}

// ΕΜΦΑΝΙΣΗ ΦΙΛΤΡΑΡΙΣΜΕΝΗΣ ΕΙΚΟΝΑΣ ΜΕ ΑΝΤΙΚΕΙΜΕΝΑ
// ΠΡΙΝ ΤΗΝ ΑΦΑΙΡΕΣΗ ΤΟΥ ΠΑΡΑΣΚΗΝΙΟΥ

namedWindow("filter_1",0);
imshow("filter_1",tmp_frame);
imwrite ("filter_1.jpg",tmp_frame);

// ΕΦΑΡΜΟΓΗ ΔΕΥΤΕΡΟΥ ΦΙΛΤΟΥ ΣΤΗ ΜΑΣΚΑ ΠΟΥ ΠΡΟΚΥΠΤΕΙ
// ΑΠΟ ΤΗ ΔΙΑΔΙΚΑΣΙΑ ΑΦΑΙΡΕΣΗΣ ΠΑΡΑΣΚΗΝΙΟΥ ΚΑΙ ΕΜΦΑΝΙΣΗ ΑΠΟΤΕΛΕΣΜΑΤΟΣ

morphologyEx(bgmask, close_bgmask,3, element2);

Scalar s(255,255,255);
close_bgmask=s - close_bgmask;

namedWindow("mask_filter", 0);
imshow("mask_filter", close_bgmask);
imwrite("mask_filter.jpg",close_bgmask);

// ΟΠΤΙΚΗ ΣΥΓΚΡΙΣΗ ΑΡΧΙΚΗΣ ΕΙΚΟΝΑΣ ΜΕ ΜΑΣΚΑ ΑΠΟ ΔΙΑΔΙΚΑΣΙΑ  ΑΦΑΙΡΕΣΗΣ
ΠΑΡΑΣΚΗΝΙΟΥ

cvtColor(tmp_frame,tmp_frame_gray,CV_BGR2GRAY);
Mat adding;
addWeighted( tmp_frame_gray, 0.9, close_bgmask,0.1, 0.0, adding);
namedWindow("original_with_mask",0);
imshow( "original_with_mask", adding );
imwrite("original_with_mask.jpg",adding);
```

```
// ΚΛΗΣΗ ΣΥΝΑΡΤΗΣΗΣ CONTOUR

  contours();

  fs.release();

  waitKey(0);

  return 0;
}

// ΣΥΝΑΡΤΗΣΗ CONTOUR
// ΛΑΜΒΑΝΕΙ ΑΠΟ ΤΗΝ GLOBAL ΜΕΤΑΒΛΗΤΗ  ΤΗ  ΜΑΣΚΑ ΑΠΟ ΤΗ ΔΙΑΔΙΚΑΣΙΑ  ΑΦΑΙΡΕΣΗΣ
//ΠΑΡΑΣΚΗΝΙΟΥ ΚΑΙ ΕΠΙΣΤΡΕΦΕΙ  ΤΙΣ ΔΙΑΣΤΑΣΕΙΣ, ΤΙΣ ΣΥΝΤΕΤΑΓΜΕΝΕΣ ΚΕΝΤΡΟΥ ΚΑΙ ΤΗ
//ΔΙΕΥΘΥΝΣΗ ΚΑΘΕ ΑΝΤΙΚΕΙΜΕΝΟΥ

void contours()
{

//  ΟΡΙΣΜΟΣ ΔΙΑΝΥΣΜΑΤΩΝ ΓΙΑ ΑΠΟΘΗΚΕΥΣΗ ΣΗΜΕΙΩΝ ΠΕΡΙΓΡΑΜΜΑΤΩΝ

  vector<vector<Point> > contours;
  vector<Vec4i> hierarchy;

// ΕΠΙΛΟΓΗ ΧΡΟΜΑΤΩΝ ΓΙΑ  ΑΠΕΙΚΟΝΙΣΗ ΠΕΡΙΓΡΑΜΜΑΤΟΣ
// ΚΑΙ ΔΙΑΦΟΡΩΝ ΣΗΜΕΙΩΝ ΕΝΔΙΑΦΕΡΟΝΤΟΣ

  Scalar color1 = Scalar( 0, 0, 255 );
  Scalar color2 = Scalar( 0, 255, 0 );
  Scalar color3 = Scalar( 255, 0, 0 );

 // ΕΥΡΕΣΗ ΠΕΡΙΓΡΑΜΜΑΤΩΝ

  findContours( close_bgmask, contours, hierarchy, CV_RETR_TREE,
CV_CHAIN_APPROX_TC89_L1, Point(0, 0) );

  cout<<"Objects Detected:"<< contours.size()-1<< "\n\n";

  // ΟΡΙΣΜΟΣ ΔΙΑΝΥΣΜΑΤΩΝ, ΠΙΝΑΚΩΝ ΚΑΙ  ΣΗΜΕΙΩΝ
 // ΓΙΑ ΑΠΟΘΗΚΕΥΣΗ ΔΕΔΟΜΕΝΩΝ ΠΟΥ ΑΦΟΡΟΥΝ
// ΠΡΟΣΕΓΓΙΣΗ ΠΕΡΙΓΡΑΜΜΑΤΟΣ ΜΕ ΓΕΩΜΕΤΡΙΚΑ ΣΧΗΜΑΤΑ

  vector<RotatedRect> minRect( contours.size() );
  vector<RotatedRect> minEllipse( contours.size() );
  Point pts [contours.size() ][4];
  Point2f center_rect[contours.size()-1];
  double area[contours.size()];
  Mat contour_line = Mat::zeros( close_bgmask.size(), CV_8UC3 );
  Mat final_img = Mat::zeros( close_bgmask.size(), CV_8UC3 );
  Mat areas = Mat::zeros( close_bgmask.size(), CV_8UC3 );

 // ΥΠΟΛΟΓΙΣΜΟΣ ΟΡΘΟΓΩΝΙΟΥ ΚΑΙ ΕΛΛΕΙΨΗΣ ΠΡΟΣΑΡΜΟΣΜΕΝΩΝ ΣΤΑ ΠΕΡΙΓΡΑΜΜΑΤΑ

  for( size_t i = 0; i < contours.size(); i++ )
    {
      minRect[i] = minAreaRect( Mat(contours[i]) );
```

```
      if( contours[i].size() > 5 )
        { minEllipse[i] = fitEllipse( Mat(contours[i]) ); }
     }

// ΥΠΟΛΟΓΙΣΜΟΣ ΕΜΒΑΔΟΥ ΚΑΘΕ ΠΡΟΣΑΡΜΟΣΜΕΝΟΥ ΟΡΘΟΓΩΝΙΟΥ
// ΟΠΤΙΚΟΠΟΙΗΣΗ ΠΕΡΙΓΡΑΜΜΑΤΟΣ, ΠΡΟΣΑΡΜΟΣΜΕΝΟΥ ΟΡΘΟΓΩΝΙΟΥ, ΕΛΛΕΙΨΗΣ
// ΕΞΑΓΩΓΗ ΣΥΝΤΕΤΑΓΜΕΝΩΝ ΚΕΝΤΡΟΥ, ΔΙΑΣΤΑΣΕΩΝ, ΔΙΕΥΘΥΝΣΗΣ

for( size_t i=1; i<contours.size(); i++)
{
   drawContours( contour_line, contours, i, color1, 9, 8, vector<Vec4i>(), 0, Point() );
   drawContours( final_img, contours, i, color1, 5, 8, vector<Vec4i>(), 0, Point() );
   ellipse( final_img, minEllipse[i], color2, 5, 8 );
   circle(final_img,minRect[i].center,10,color3,5,8,0);
   center_rect[i]=minRect[i].center;

  // ΥΠΟΛΟΓΙΣΜΟΣ ΕΜΒΑΔΟΥ
 // ΣΧΕΔΙΑΣΜΟΣ ΠΡΟΣΑΡΜΟΣΜΕΝΟΥ ΟΡΘΟΓΩΝΙΟΥ

   Point2f rect_points[4];
   vector<Point2f>pointsofarea;
   minRect[i].points( rect_points );
   for( int j = 0; j < 4; j++ )
   {
     line( final_img, rect_points[j], rect_points[(j+1)%4], color3, 9, 8 );
     pointsofarea.push_back(rect_points[j]);
     pts[i][j].x=rect_points[j].x;
     pts[i][j].y=rect_points[j].y;
    }
  area[i]= contourArea(pointsofarea);
  fillConvexPoly(areas,pts[i] ,4,color3,8,0);

  cout<<"Antikeimeno diastaseon   "<<minRect[i].size<< " \nEpifaneias   "<< area[i]<<"\nMe
kentro sti thesi   "<< center_rect[i]<< " \nipo gonia  "<< minEllipse[i].angle<<"\n\n";
   }

 //  ΟΠΤΙΚΗ ΕΠΑΛΗΘΕΥΣΗ ΜΕ ΑΡΧΙΚΗ ΕΙΚΟΝΑ //

  addWeighted( tmp_frame, 0.8, contour_line,0.2 , 0.0, contour_line);
  namedWindow("Contours", 0);
  imshow( "Contours", contour_line);
  imwrite("Contours.jpg",contour_line);

  addWeighted( tmp_frame, 0.9, areas,0.1 , 0.0, areas);
  namedWindow("Areas", 0);
  imshow( "Areas", areas);
  imwrite("Areas.jpg",areas);

  addWeighted( tmp_frame, 0.7, final_img,0.3 , 0.0, final_img);
  namedWindow("Final img", 0);
  imshow( "Final img", final_img);
  imwrite("Final_img.jpg",final_img);
}
```

ΠΑΡΑΡΤΗΜΑ Β

ΑΠΟΤΕΛΕΣΜΑΤΑ ΔΙΑΔΙΚΑΣΙΑΣ ΚΑΛΙΜΠΡΑΡΙΣΜΑΤΟΣ ΚΑΜΕΡΑΣ

Τα αποτελέσματα της διαδικασίας καλιμπραρίσματος της κάμερας παράγοται σε μορφή xml και είναι τα ακόλουθα:


```
<?xml version="1.0"?>
<opencv_storage>
<calibration_Time>"03/18/15 22:15:05"</calibration_Time>
<nrOfFrames>9</nrOfFrames>
<image_Width>3264</image_Width>
<image_Height>2448</image_Height>
<board_Width>9</board_Width>
<board_Height>6</board_Height>
<square_Size>25.</square_Size>
<FixAspectRatio>1.</FixAspectRatio>
<!-- flags:  +fix_aspectRatio +fix_principal_point -->
<flagValue>6</flagValue>
<Camera_Matrix type_id="opencv-matrix">
  <rows>3</rows>
  <cols>3</cols>
  <dt>d</dt>
  <data>
    3.6976069895458918e+003 0. 1.6315000000000000e+003 0.
    3.6976069895458918e+003 1.2235000000000000e+003 0. 0. 1.</data></Camera_Matrix>
<Distortion_Coefficients type_id="opencv-matrix">
  <rows>5</rows>
  <cols>1</cols>
  <dt>d</dt>
  <data>
    8.3025271205107408e-003 -1.1984263428909485e-001
    1.6600011096342545e-003 9.8536513211766922e-004
    -1.7596073759780304e-001</data></Distortion_Coefficients>
<Avg_Reprojection_Error>9.9581493980558389e-001</Avg_Reprojection_Error>
<Per_View_Reprojection_Errors type_id="opencv-matrix">
  <rows>9</rows>
  <cols>1</cols>
  <dt>f</dt>
  <data>
    8.60845268e-001 7.36879408e-001 7.97146559e-001 8.54784071e-001
    7.49571741e-001 7.33340263e-001 6.24467492e-001 2.05792236e+000
    7.41641939e-001</data></Per_View_Reprojection_Errors>
<!-- a set of 6-tuples (rotation vector + translation vector) for each view -->
<Extrinsic_Parameters type_id="opencv-matrix">
```

<rows>9</rows>
<cols>6</cols>
<dt>d</dt>
<data>
1.9663992186920234e-003 2.1656265342177777e-002
-1.5598654062848258e+000 -2.5508523442972279e+002
1.7674269908491762e+001 6.5177024332002270e+002
1.1553727945820581e-002 2.6294786381544904e-002
-1.5771021505085938e+000 1.2050194583004992e+002
2.5522781483804383e+001 6.4643166229019516e+002
4.1543361795935990e-001 -2.3761132463568868e-001
-1.6473080886077136e+000 -1.4041696105265643e+002
1.2691670455236755e+002 4.8810062662605952e+002
4.0301060120128677e-002 1.4289548525015922e-002
-2.6962656138951493e-003 -8.7149012222494903e+001
-1.7714867911492595e+002 6.4414277749000814e+002
1.7161731750887803e-002 2.3178259534774679e-002
-1.5656626877466946e+000 -2.5024796371239927e+002
1.8445407020704258e+002 6.5666869856785672e+002
1.6788138640125249e-002 3.2714042207198472e-002
-1.5814640813541563e+000 1.2917211289072577e+002
9.6853521570174422e+001 6.4826540424010261e+002
2.2397289699047650e-002 1.9662679029368982e-002
-1.0039684595021851e-002 -7.5708657074238033e+001
5.7255508788311701e+001 6.5106589173193788e+002
6.7227913199526901e-001 8.0884161724437287e-001
-1.4772921390524074e+000 -6.5087107835048229e+001
1.3056893577276266e+001 5.5594399884368067e+002
1.2967630387030341e-002 2.0482051756772736e-002
-1.5398105096856534e+000 -2.5604466659615059e+002
1.1267773386095860e+002 6.5414602371217632e+002</data></