



ΠΑΝΕΠΙΣΤΗΜΙΟ
ΘΕΣΣΑΛΙΑΣ

ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ

ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ ΜΕ ΕΦΑΡΜΟΓΕΣ ΣΤΗΝ ΒΙΟΪΑΤΡΙΚΗ

**ΜΕΤΡΗΣΕΙΣ ΑΠΟΣΤΑΣΕΩΝ ΜΕΣΩ ΕΙΚΟΝΩΝ
ΚΑΙ ΒΙΝΤΕΟ ΜΕ ΕΦΑΡΜΟΓΕΣ ΣΕ
ΕΝΔΟΣΚΟΠΙΚΑ ΔΕΔΟΜΕΝΑ**

Γεώργιος Δήμας

Επιβλέπων

Δημήτρης Ιακωβίδης

Αναπληρωτής Καθηγητής

ΠΕΡΙΛΗΨΗ

Η ενδοσκόπηση είναι μία ιατρική μέθοδος η οποία χρησιμεύει για την διάγνωση ή ακόμα και την θεραπεία παθήσεων του ανωτέρου και κατωτέρου γαστρεντερικού συστήματος. Η εξέταση της ενδοσκόπησης γίνεται χρησιμοποιώντας ένα ιατρικό εργαλείο που ονομάζεται ενδοσκόπιο. Τα ενδοσκόπια έχουν σηνήθως τη μορφή λεπτών εύκαμπτων σωλήνων (ενσύρματο) ή την μορφή κάψουλας (ασύρματο) και φέρουν στο άκρο τους κάμερα και απαραίτητο εξοπλισμό για σωστή απεικόνιση. Η εξέλιξη της τεχνολογίας στον τομέα αυτό, με την άνοδο της ασύρματης ενδοσκόπησης, έφερε την ανάγκη ανάπτυξης τεχνικών, υπολογισμού της θέσης και της μετατόπισης των ενδοσκοπίων. Μέχρι τώρα η εύρεση της θέσης των ασύρματων ενδοσκοπίων γίνεται με την χρήση ραδιοσυχνοτήτων και εξωτερικών αισθητήρων για την εύρεση της θέσης τους, μέσω τριγωνοποίησης. Αυτή η μέθοδος ωστόσο δεν παρέχει πληροφορία για την θέση του ενδοσκοπίου συναρτήσει της γαστρεντερικής οδού, αλλά βάση του τρισδιάστατου ή δισδιάστατου κοιλιακού χώρου του ασθενούς. Σε αυτήν την εργασία προτείνονται δύο μέθοδοι για τον υπολογισμό της θέσης και της μετατόπισης μιας ενδοσκοπικής κάψουλας, χρησιμοποιώντας αμιγώς οπτική πληροφορία μέσω των ληφθέντων εικόνων και βίντεο. Οι δύο μέθοδοι που χρησιμοποιούνται είναι η κλασσική γεωμετρική προσέγγιση της οπτικής οδομετρίας και μία καινοτόμα μεθοδολογία που βασίζεται στην μηχανική μάθηση. Οι δύο μέθοδοι είχαν ως αποτέλεσμα πολύ μικρά σφάλματα, 7.12 ± 4.01 cm και 2.70 ± 1.62 cm αντίστοιχα.

ABSTRACT

Endoscopy is a medical, diagnostic and treatment procedure for the upper and lower gastrointestinal system. The endoscopy is performed using a medical device that is called endoscope. Endoscopes usually are in the form of thin, elastic pipe or under the form of a capsule, with a mounted camera on the edge and the required equipment for the proper imaging of the gastrointestinal tract. The technological advancements in the field, and the rise of wireless capsule endoscopy, brought up the need for the development of methods for the localization and displacement estimation of the endoscopes. So far, the localization of the endoscopes is performed by radiofrequency triangulation using external sensors. Although, the radiofrequency approach is only capable of locating the capsule with in the 3D or 2D abdominal space of the patient and not with regard to the gastrointestinal tract. In this paper are proposed two different methods of localization and displacement estimation of the endoscope, using solely optical information from the captured images and videos. More specifically, the two methods used are the standard geometric visual odometry approach and a novel methodology that is based on machine learning. The two aforementioned methods resulted in errors of 7.12 ± 4.01 cm and 2.70 ± 1.62 cm respectively.

ΠΕΡΙΕΧΟΜΕΝΑ

1. Εισαγωγή	1
1.1 Ενδοσκόπια	1
1.2 Στόχοι	3
1.3 Επιστημονική Συνεισφορά.....	3
1.4 Δομή.....	6
2. Βιβλιογραφική Ανασκόπηση	9
2.1 Βαθμονόμηση Κάμερας	10
2.2 Γεωμετρική Οπτική Οδομετρία	13
2.3 Γεωμετρική Οπτική Οδομετρία Με Εφαρμογές Στην Ενδοσκόπηση.....	14
2.4 “Εξυπνη” Βαθμονόμηση Κάμερας Και Οπτική Οδομετρία.....	16
3. Μεθοδολογία.....	19
3.1 Γεωμετρική Παραμετρική Προσέγγιση.....	19
3.1.1 Βαθμονόμηση Κάμερας	20
3.1.2 Εξαγωγή Σημείων και Ταίριασμα.....	27
3.1.3 Ανακατασκευή πορείας	39
3.2 “Εξυπνη” Μη Παραμετρική Προσέγγιση	44
3.2.1 Διανύσματα Χαρακτηριστικών Εισόδου	46
3.2.2 Ανάλυση Αρχιτεκτονικής Δικτύου	47
4. Πειραματική Αξιολόγηση.....	52
4.1 Ρομποτικό Πείραμα	52
4.2 Αποτελέσματα.....	55
4.2.1 Γεωμετρική Οπτική Οδομετρία	55
4.2.2 “Εξυπνη” Οπτική Οδομετρία.....	57
4.3 Τεχνικές Λεπτομέρειες Υλοποίησης.....	70
5. Συμπεράσματα.....	72
6. Βιβλιογραφία	82

1. Εισαγωγή

1.1 Ενδοσκόπηση

Ο γενικός ορισμός της ενδοσκόπησης, είναι η εξέταση των εσωτερικών περιοχών των διαφόρων συστημάτων, χρησιμοποιώντας μία συσκευή η οποία ονομάζεται ενδοσκόπιο. Η ενδοσκόπηση έχει εφαρμογές στην εξέταση του εσωτερικού περίπλοκων μηχανολογικών συστημάτων, στην ιατρική, στην εξέταση εκρηκτικών μηχανισμών κ.α.

Στην ιατρική, η ενδοσκόπηση είναι μέθοδος, διαγνωστικής ή θεραπευτικής εξέτασης και απεικόνισης του ανώτερου και κατώτερου γαστρεντερικού συστήματος. Έτσι, με την ενδοσκόπηση, μπορεί να παρατηρηθεί, μέσω απεικόνισης, όλη η εσωτερική επιφάνειά του οισοφάγου, στομάχου, δωδεκαδάκτυλου και του παχέος εντέρου. Τα όργανα που χρησιμοποιούνται για την εξέταση ονομάζονται ενδοσκόπια. Τα ενδοσκόπια του γαστρεντερικού συστήματος, έχουν την μορφή λεπτών, εύκαμπτων σωλήνων ή ασύρματης κάψουλας τα οποία είναι εξοπλισμένα με κάμερα για την απεικόνιση του συστήματος που εξετάζουν. Η ενδοσκόπηση με την χρήση εύκαμπτων ενδοσκοπίων (εικόνα 1.1(α)) πραγματοποιείται, με ή χωρίς την καταστολή του ασθενούς, και επιτρέπει την επισκόπηση του ανώτερου γαστρεντερικού συστήματος (γαστροσκόπηση) και του παχέος εντέρου (κολονοσκόπηση). Η εξέταση του λεπτού εντέρου καθίσταται δυνατή χρησιμοποιώντας είτε πιο εξελιγμένα εύκαμπτα ενδοσκόπια, είτε ασύρματη κάψουλα ενδοσκόπησης την οποία την καταπίνει ο ασθενής και διατρέχει όλο το μήκος του εντέρου (εικόνα 1.1(β)).

Μέσω της ενδοσκόπησης, πέραν της δυνατότητας που προσφέρει για την διάγνωση παθολογιών, επιτρέπει και την θεραπευτική παρέμβαση. Κάποιες από τις θεραπευτικές μεθόδους που καθίστανται δυνατές μέσω της ενδοσκόπησης είναι η αιμόσταση σε περίπτωση αιμοραγίας του πεπτικού συστήματος, η διαστολή χρησιμοποιώντας μπαλόνι σε στενώσεις και η τοποθέτηση ενδοπροθέσεων, η αφαίρεση πολυπόδων κ.α.



(α)



(β)

Εικόνα 1.1. Στην εικόνα (α) απεικονίζεται ένα κλασσικό ενσύρματο ενδοσκόπιο, ενώ στην εικόνα (β) παρουσιάζεται μία ασύρματη κάψουλα ενδοσκόπησης. Όπως μπορεί να παρατηρηθεί στο καλώδιο του ενδοσκοπίου της εικόνας (α) είναι αναγεγραμμένο το μήκος του καλωδίου σε διάφορα σημεία, κάτι που καθιστά τον προσδιορισμό της θέσης που παρατηρείται κάθε στιγμή πολύ πιο εύκολη.

Η ανάγκη για τον υπολογισμό της θέσης και της μετατόπισης του ενδοσκοπίου κατά την διάρκειά της εξέτασης του γαστρεντερικού συστήματος, είναι σημαντική για την προσέγγιση της θέσης που εντοπίζονται πιθανές ανωμαλίες. Η εύρεση της θέσης είναι άκρως σημαντική, για την ακριβή εύρεση της τοποθεσία ανωμαλιών, έτσι ώστε, σε δεύτερο χρόνο, να πραγματοποιηθεί βιοψία ή και θεραπεία, χρησιμοποιώντας άλλες μεθόδους, όπως χειρουργικές επεμβάσεις κ.α (Kita & Ciuti, 2006).

Η συγκεκριμένη έρευνα που παρουσιάζει η παρούσα εργασία, εστιάζει περισσότερο στην εφαρμογή των μετρήσεων, σε ενδοσκοπικά βίντεο που έχουν ληφθεί από ασύρματες κάψουλες ενδοσκόπησης. Ο λόγος είναι ότι εξαιτίας της ασύρματης φύσης των συγκεκριμένων ενδοσκοπίων, ο υπολογισμός της θέσης των πιθανών ανωμαλιών που εντοπίζονται, είναι είναι δύσκολος και όχι ιδιαίτερα ακριβής. Στις κλασσικές μεθόδους ενδοσκόπησης, που χρησιμοποιούν ενσύρματα ενδοσκόπια, η προσέγγιση της θέσης του ενδοσκοπίου είναι πολύ πιο εύκολη. Αυτό συμβαίνει διότι συνήθως, πάνω στο καλώδιο των ασύρματων ενδοσκοπίων, αναγράφεται το μήκος του καλωδίου συναρτήσει του άκρου που βρίσκεται η κάμερα (εικόνα 1.1(α)). Αντιθέτως, στις ασύρματες κάψουλες ενδοσκόπησης δεν έχουμε κάποιο σταθερό σημείο αναφοράς ούτως ώστε να είναι εύκολος ο υπολογισμός της θέσης της. Επίσης, μια

μεθοδολογία για τον εντοπισμό της θέσης του ενδοσκοπίου, που έχει εφαρμογή με ικανοποιητικά αποτελέσματα σε ασύρματες κάψουλες ενδοσκόπησης, είναι πολύ εύκολο να εφαρμοστεί και στα υπόλοιπα είδη ενδοσκοπίων.

1.2 Στόχοι

Η ενδοσκόπηση μέσω ασύρματης κάψουλας ενδοσκόπησης επιτρέπει την απεικόνιση της γαστρεντερικής οδού. Μέσω της απεικόνισης μπορούμε να εντοπίσουμε πιθανές ανωμαλίες και παθολογίες του γαστρεντερικού συστήματος. Παρόλα αυτά, η θέση τους, στο εσωτερικό του συστήματος είναι δύσκολο να τεκμηριωθεί μέσω αυτής της μεθόδου. Σε συνδιασμό λοιπόν με την ανάπτυξη τεχνολογιών αυτοματοποίησης της διαδικασίας ανίχνευσης αλλοιώσεων και ανωμαλιών σε ενδοσκοπικά βίντεο εξετάσεων (Karkanis *et al.*, 2003) (Iakovidis & Koulaouzidis, 2014), προκύπτει η ανάγκη για την εύρεση του σημείου στο οποίο βρίσκονται. Αυτό γιατί, σε δεύτερο χρόνο, πέραν της διάγνωσης, χρειάζονται περαιτέρω διερεύνηση (βιοψία) είτε χρήζουν θεραπείας μέσω άλλων μεθόδων.

Οι στόχοι αυτής της εργασίας μπορούν να συνοψιστούν στα ακόλουθα σημεία:

- i. Διερεύνηση μεθόδων μέτρησης αποστάσεων βάσει οπτικής πληροφορίας.
- ii. Εφαρμογή των μεθόδων του σημείο (i) στο πλαίσιο της οπτικής οδομετρίας.
- iii. Ανάπτυξη και πειραματική αξιολόγηση πρωτότυπης μεθοδολογίας για τον προσδιορισμό της θέσης και ανακατασκευής της πορείας της ενδοσκοπικής κάψουλας από δεδομένα βίντεο.

1.3 Επιστημονική Συνεισφορά

Κατά την διάρκεια της έρευνας για εύρεση λύσης στο πρόβλημα της μέτρησης αποστάσεων βάσει οπτικής πληροφορίας, πραγματοποιήθηκε η ανάπτυξη τριών διαφορετικών μεθοδολογιών, οι οποίες είναι οι εξής:

- i. Η προσαρμογή της κλασσικής γεωμετρικής οπτικής οδομετρίας για την εφαρμογή της σε

ενδοσκοπικές εικόνες/στιγμιότυπα βίντεο.

- ii. Η ανάπτυξη υβριδικής μεθοδολογίας για το πρόβλημα υπολογισμού της θέσης της κάψουλας. Στην μεθοδολογία αυτή, γίνεται συνδιασμός της κλασσικής γεωμετρικής προσέγγισης με ένα τεχνητό νευρωνικό δίκτυο για την βελτιστοποίηση των αποτελεσμάτων.
- iii. Η πρόταση καινοτόμας μεθοδολογίας, βασιζόμενη σε μία αρχιτεκτονική τεχνητού νευρωνικού δικτύου, για τον υπολογισμό της απόστασης που διανύεται από το ασύρματο ενδοσκόπιο.

Ως αποτέλεσμα της έρευνας που πραγματοποιήθηκε στην παρούσα εργασία, δημοσιεύτηκαν οι ακόλουθες εργασίες σε δύο διεθνή επιστημονικά περιοδικά και δύο διεθνή συνέδρια. Επιπλέον, μια πιο πρόσφατη εργασία βρίσκεται υπό αξιολόγηση (V) για την δημοσίευση της σε διεθνές ιατρικό περιοδικό. Οι δημοσιεύσεις παραθέτονται ονομαστικά παρακάτω:

- I. D. K. Iakovidis, G. Dimas, A. Karargyris, G. Ciuti, F. Bianchi, A. Koulaouzidis, E. Toth, ***Robotic validation of visual odometry for wireless capsule endoscopy***, IEEE International Conference of Imaging System and Techniques (IST), 83-87, 2016.
- II. G. Dimas, D. K. Iakovidis, A. Karargyris, G. Ciuti, A. Koulaouzidis, ***Visual localization of wireless capsule endoscopes aided by artificial neural networks***, IEEE International Symposium on Computer-Based Medical Systems, 2017.
- III. G. Dimas, D. K. Iakovidis, A. Karargyris, G. Ciuti, A. Koulaouzidis, ***An artificial neural network architecture for non-parametric visual odometry in wireless capsule endoscopy***, Measurement Science and Technology, 28(9), 2017.
- IV. G. Dimas, E. Spyrou, D. K. Iakovidis, A. Koulaouzidis, ***Intelligent visual localization of wireless capsule endoscopes enhanced by color information***, Computers in Biology and Medicine, Volume 89, Pages 429-440.

- V. A. Koulaouzidis, D. K. Iakovidis, D. Yung, E. Mazomenos, F. Bianchi, A. Karargyris, G. Dimas, D. Stoyanov, H. Thorlacious, E. Toth & G. Ciuti, *Novel experimental and software methods for image reconstruction and localization in capsule endoscopy*, Endoscopy International Open, 2017 [Submitted (accepted with minor revisions)].

Στην εργασία της πρώτης δημοσίευσης (I), εξετάστηκε η δυνατότητα εφαρμογής της γεωμετρικής οπτικής οδομετρίας, στο ερευνητικό πλαίσιο της ενδοσκόπησης, μέσω ασύρματης κάψουλας. Επιπλέον, για την βαθμονόμηση της κάμερας του ασύρματου ενδοσκοπίου, εξετάστηκαν και συγκρίθηκαν δυο διαφορετικές μεθοδολογίες, αυτή του Zhang (1999) και των Kannala και Brandt (2006). Η επαλήθευση της μεθοδολογίας έγινε βάση μιας *ex-vivo* πειραματικής διάταξης υποβοηθούμενης από έναν ρομποτικό βραχίονα. Τα αποτελέσματα έδειξαν ότι η μεθοδολογία βαθμονόμησης των Kannala και Brandt οδήγησε σε μικρότερα σφάλματα στην εφαρμογή της οπτικής οδομετρίας, αλλά συγκρίσιμα με αυτά της μεθοδολογίας του Zhang. Συγκεκριμένα το μικρότερο σφάλμα που επιτεύχθηκε ήταν 7.2 ± 1.4 cm.

Στην επόμενη δημοσίευση (II), προτάθηκε μία υβριδική μεθοδολογία στην οποία συνδιάζονται η κλασσική μέθοδος της οπτικής οδομετρίας με ένα νευρωνικό δίκτυο για την βελτίωση των αποτελεσμάτων της πρώτης. Τα δεδομένα για την επαλήθευση αυτής της υβριδικής προσέγγισης βασίστηκαν στα πειραματικά δεδομένα της (I). Το νευρωνικό δίκτυο εκπαιδεύτηκε έτσι ώστε να υπολογίζει την απόσταση που διανύει η ασύρματη κάψουλα ενδοσκόπησης στον άξονα Z ο οποίος θεωρείται αυτός που είναι παράλληλος της γαστρεντερικής οδού. Εκμετελευόμενοι την γνώση για την απόσταση που έχει διανύσει η κάψουλα στον άξονα Z μπορούμε να υπολογίσουμε τον παράγοντα κλιμάκωσης για να κλιμακώσουμε τα αποτελέσματα της οπτικής οδομετρίας σε φυσικές μονάδες μέτρησης. Το μέσο απόλυτο σφάλμα που καταγράφηκε, για απόσταση 19.6 cm είναι 0.79 ± 0.51 cm. Ωστόσο, το σετ δεδομένων εκπαίδευσης και επαλήθευσης ήταν σχετικά μικρό.

Στην εργασία η οποία δημοσιεύτηκε στο διεθνή περιοδικό Measurement Science and Technology (III), χρησιμοποιήσαμε ένα νευρωνικό δίκτυο, με αρχιτεκτονική εμπρόσθιας τροφοδοσίας, για να εκτελέσουμε οπτική οδομετρία σε δεδομένα που έχουν ληφθεί από μία

ασύρματη κάψουλα ενδοσκόπησης. Ως είσοδοι του δικτύου χρησιμοποιήθηκαν οι συντεταγμένες αντιστοιχισμένων σημείων μεταξύ δύο εικόνων καθώς και η αντίστοιχη ένταση φωτεινότητας των εικονοστοιχείων που αντιστοιχούν. Τα αποτελέσματα μέσω αυτής της μεθόδου ήταν 1.14 ± 0.75 cm.

Στην πιο πρόσφατη εργασία που δημοσιεύθηκε, έγινε μία εκτενής έρευνα για το πώς η χρωματική πληροφορία επηρεάζει τους υπολογισμούς ενός νευρωνικού δικτύου για τον υπολογισμό της θέσης της ασύρματης κάψουλας ενδοσκόπησης. Επίσης διερευνήθηκε και η αποτελεσματικότητα διαφόρων αλγορίθμων αντιστοίχισης σημείων μεταξύ διαδοχικών στιγμιοτύπων ενδοσκοπικού βίντεο. Τα αποτελέσματα έδειξαν ότι ο συνδιασμός των αλγορίθμων KLT-RANSAC για την εύρεση αντιστοιχιών και η χρήση χρωματικής πληροφορίας του χρωματικού χώρου *CIE-Lab* έδωσαν τα μικρότερα σφάλματα. Συγκεκριμένα για αρχιτεκτονική με 5 κρυφούς νευρώνες το μέσο απόλυτο σφάλμα ήταν 2.70 ± 1.62 cm.

Στην τελευταία εργασία (V) εξετάστηκε η εφαρμογή της οπτικής οδομετρίας σε ενδοσκοπικά βίντεο από ασύρματη κάψουλα ενδοσκόπησης με το ελάχιστο σφάλμα που σημειώνεται, να είναι 1.4 ± 0.8 cm. Επίσης ερευνήθηκαν και δύο διαφορετικές τεχνικές τρισδιάστατης ανακατασκευής του πλάνου των εικόνων. Συγκεκριμένα, οι τεχνικές αυτές είναι η μέθοδος Σχήματος από Σκίαση (Shape from Shading) και ο Ταυτόχρονος Εντοπισμός Θέσης και Χαρτογράφησης (ORB-SLAM).

1.4 Δομή

Η συνέχεια της συγκεκριμένης εργασίας, είναι οργανωμένη σε 4 βασικές ενότητες:

- i. Στην ενότητα **2**, η οποία είναι αφιερωμένη στην ανασκόπηση της βιβλιογραφίας σχετικά με τεχνικές οπτικής οδομετρίας και βαθμονόμησης της κάμερας. Η ενότητα **2** χωρίζεται ως εξής:
 - a. Στις μεθοδολογίες βαθμονόμησης κάμερας.
 - b. Σε γενικές εφαρμογές της γεωμετρικής οπτικής οδομετρίας.

- c.* Σε εφαρμογές της γεωμετρικής οπτικής οδομετρίας, στο πλαίσιο της ενδοσκόπησης.
 - d.* Σε προσεγγίσεις στο πρόβλημα της βαθμονόμησης της κάμερας και της οπτικής οδομετρίας χρησιμοποιώντας μεθόδους μηχανικής μάθησης.
- ii.** Στην ενότητα **3**, αναλύονται οι δύο κύριες μεθοδολογίες που προτείνονται στην συγκεκριμένη εργασία. Η ενότητα **3** χωρίζεται ως εξής:
 - a.* Στην ανάλυση της μεθοδολογίας της γεωμετρικής οπτικής οδομετρίας.
 - b.* Στη ανάλυση της καινοτόμας “έξυπνης” προσέγγισης στο πρόβλημα, που βασίζεται σε μέθοδο μηχανικής μάθησης.
- iii.** Στην ενότητα **4** η οποία είναι αφιερωμένη στην ανάλυση του πειράματος στο οποίο εδραιώθηκε η βάση αληθείας και στην αναφορά των αποτελεσμάτων που προσέφερε η κάθε μεθοδολογία. Επίσης, γίνεται σύγκρισή των δύο προτεινόμενων μεθοδολογιών. Η ενότητα **4** χωρίζεται ως έχει:
 - a.* Στην ανάλυση του ρομποτικού πειράματος που χρησιμοποιήθηκε για την εδραίωση της βάσης αληθείας (ground truth) για την αξιολόγηση των μεθοδολογιών που μελετήθηκαν.
 - b.* Στην παρουσίαση των αποτελεσμάτων των δύο προαναφερθέντων μεθοδολογιών που προτείνονται στην παρούσα εργασία.
- iv.** Τέλος, στην ενότητα **5** όπου γίνεται σύγκριση μεταξύ των δύο, και αυτής με τα καλύτερα αποτελέσματα, με άλλες μεθοδολογίες που εστιάζουν στην λύση του προβλήματος μέτρησης αποστάσεων σε ενδοσκοπικά βίντεο. Επίσης στην ενότητα **5** γίνεται και σχολιασμός επί των περιορισμών που είχε η συγκεκριμένη έρευνα καθώς και στα σχέδια για μελλοντικές δουλειές που προέκυψαν από την πορεία της συγκεκριμένης μελέτης.

2. Βιβλιογραφική Ανασκόπηση

Ο τομέας των οπτικών μετρήσεων, μέσω της οπτικής οδομετρίας, ανήκει στο ευρύτερο ερευνητικό πλαίσιο της υπολογιστικής όρασης. Η οπτική οδομετρία χρησιμεύει σε διαδικασίες όπως η πλοήγηση ρομπότ (*M. Maimone et al., 2007*), αμαξιών αλλά και στη τρισδιάστατη ανακατασκευή τοπίων (*M. Faessler et al., 2016*), σε γνωστά ή άγνωστα περιβάλλοντα. Η διαδικασία αυτή επιτυγχάνεται υπολογίζοντας την μετατροπή της θέσης (translation) αλλά και της περιστροφή (orientation) του υποκειμένου που φέρει την κάμερα, χρησιμοποιώντας αμιγώς οπτική πληροφορία (μέσα από στιγμιότυπα βίντεο ή εικόνες). Η συλλογή των εικόνων και των βίντεο γίνεται χρησιμοποιώντας μία (monocular odometry) ή δύο (stereo odometry) βαθμονομημένες (calibrated) κάμερες (σε μερικές περιπτώσεις χρησιμοποιούνται και ευρυγώνιες κάμερες), έτσι ώστε σε δεύτερο χρόνο, να γίνει εφαρμογή της οπτικής οδομετρίας. Συνοπτικά, οι μετρήσεις αποστάσεων γίνονται θεσπίζοντας την σχέση που υπάρχει μεταξύ των συντεταγμένων των συντεταγμένων των εικονοστοιχείων μιας εικόνας (2D image pixel coordinates), με τις αντίστοιχες συντεταγμένες του πραγματικού κόσμου (3D real world coordinates).

Η βαθμονόμηση (calibration) είναι απαραίτητη για τον υπολογισμό των παραμέτρων της κάθε κάμερας, για την εκπόνηση της διαδικασίας της οδομετρίας (*D. Scaramuzza and F. Friedrich, 2011*). Οι παράμετροι της κάμερας χρησιμοποιούνται στην πλειοψηφία των μεθοδολογιών οπτικής οδομετρίας, και χωρίζονται σε δύο κατηγορίες, τις *α)* εσωτερικές παραμέτρους οι οποίες είναι το εστιακό μήκος (focal length), τον παραάγοντα κλιμάκωσης, το πρωτεύον σημείο (principal point) (είναι δηλαδή οι παράμετροι οι οποίες χαρακτηρίζουν το γεωμετρικό μοντέλο της κάμερας), και τις *β)* εξωτερικές παραμέτρους της κάμερας, οι οποίες είναι η θέση της κάμερας και ο τρόπος περιστροφής της.

Χρησιμοποιώντας τις εξωτερικές παραμέτρους της κάμερας, μπορούμε να υπολογίσουμε την θέση της, στον πραγματικό, τρισδιάστατο χώρο. Πέραν των προαναφερθέντων εσωτερικών παραμέτρων, ως εσωτερικές παράμετροι μπορούν να χαρακτηριστούν και η

παραμόρφωση του φακού της κάμερας (lens distortion) αλλά και η κλίση του αισθητήρα, οι οποίες διαφοροποιούνται από κάμερα σε κάμερα. Οι εσωτερικές παράμετροι, για την επιτυχή εφαρμογή της γεωμετρικής οπτικής οδομετρίας, πρέπει να υπολογιστούν μέσω της διαδικασίας της βαθμονόμησης της κάμερας (camera calibration). Μέσω της βαθμονόμησης της κάμερας επιτυγχάνεται η αντιστοίχιση των δισδιάστατων συντεταγμένων των εικονοστοιχείων της εικόνας, στις τρισδιάστατες συντεταγμένες του πραγματικού κόσμου.

Πέραν της κλασσικής μεθοδολογίας της γεωμετρικής οπτικής οδομετρίας, έχουν προταθεί και “έξυπνοι” μέθοδοι για τον υπολογισμό της θέσης και της περιστροφής της κάμερας. Οι μέθοδοι αυτοί χρησιμοποιούν αλγορίθμους μηχανικής και βαθιάς μάθησης. Συγκεκριμένα, τεχνητά νευρωνικά δίκτυα (artificial neural networks, ANN) και συνελκτικά νευρωνικά δίκτυα (convolutional neural networks, CNN).

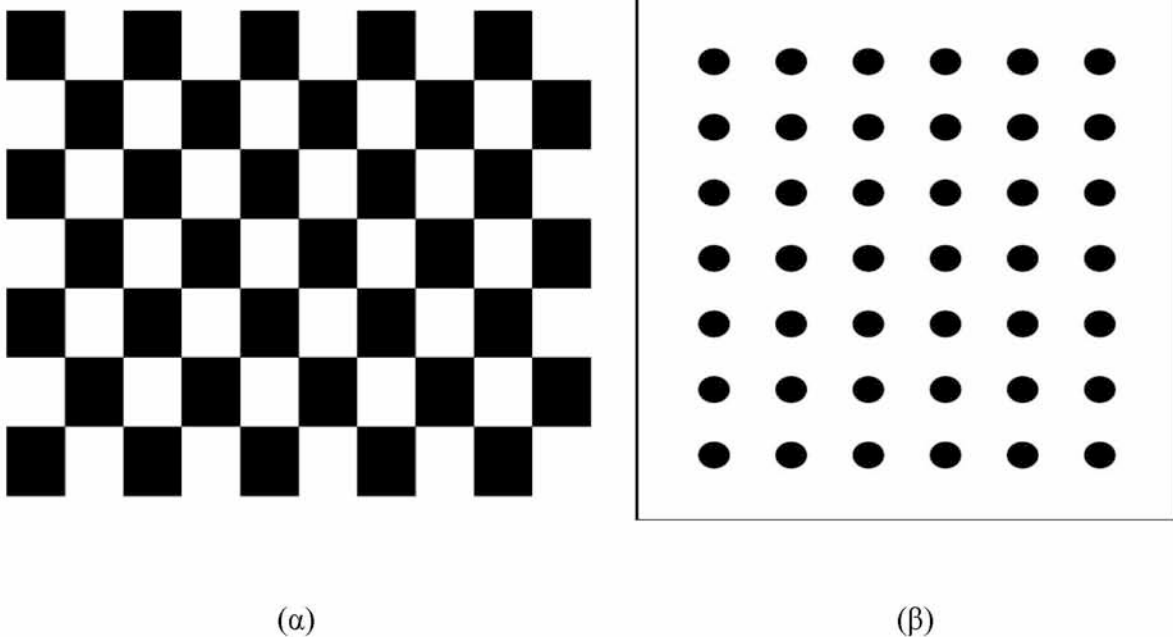
Στην συνέχεια αυτής της ενότητας πραγματοποιείται ανασκόπηση της βιβλιογραφίας που σχετίζεται με την παρούσα εργασία. Ξεκινώντας από την υφιστάμενη κατάσταση (state of the art) της βαθμονόμησης της κάμερας και της μεθοδολογίας της γεωμετρικής οπτικής οδομετρίας και καταλήγοντας σε σύγχρονες προσεγγίσεις, με “έξυπνες” μεθόδους των παραπάνω τεχνολογιών.

2.1 Βαθμονόμηση Κάμερας

Η βαθμονόμηση της κάμερας, στο ερευνητικό πλαίσιο της τρισδιάστατης οπτικής όρασης, ορίζεται ως η διαδικασία μέσω της οποίας υπολογίζονται οι γεωμετρικές και οπτικές παράμετροι που χαρακτηρίζουν την κάμερα. Επίσης, μέσω της βαθμονόμησης της κάμερας μπορούν να υπολογιστούν, η θέση και ο προσανατολισμός της, συναρτήσει κάποιου συστήματος συντεταγμένων (Tsai, 1987). Οι εσωτερικές παράμετροι της κάμερας συνήθως περιλαμβάνουν το εστιακό μήκος, τον παράγοντα κλιμάκωσης και το οπτικό κέντρο (Heikkila, 1997).

Μία από τις πιο διαδεδομένες σε χρήση και πρωτοποριακές μεθοδολογίες βαθμονόμησης είναι αυτή που προτάθηκε από τον Zhang (1999). Στην συγκεκριμένη μεθοδολογία έχει

μοντελοποιηθεί και η παραμόρφωσή που προκαλείται στην εικόνα από τον φακό. Για την υλοποίηση της μεθοδολογίας του Zhang χρειάζεται ένα επίπεδο πρότυπο (planar pattern), γνωστό και ως σκακιέρα λόγω της μορφοποίησης του. Οι διαστάσεις των κουτιών πρέπει να είναι γνωστές. Στην εικόνα 2.1 (α) παρουσιάζεται μία αντιπροσωπευτική εικόνα ενός τέτοιου επιπέδου. Η μεθοδολογία του, τεκμηριώθηκε και από πραγματικά αλλά και από υπολογιστικά παραγμένα δεδομένα.



Εικόνα 2.1. Εικόνα που παρουσιάζει παράδειγμα ενός (α) επιπέδου σκακιέρας που χρησιμοποιείται για την διαδικασία βαθμονόμησης στην μεθοδολογία του Zhang και ενός (β) πρότυπου επιπέδου με κυκλικά στοιχεία που χρησιμοποιείται στις μεθοδολογίες των Heikkila και Kannala και Brandt

Για την σωστή υλοποίηση της μεθοδολογίας του Zhang, πρέπει να ληφθούν τουλάχιστον δύο εικόνες του επιπέδου, από διαφορετικές αποστάσεις αλλά και κλίσεις, χρησιμοποιώντας την κάμερα που θέλουμε να βαθμονομήσουμε. Κατά την λήψη των εικόνων για την βαθμονόμηση της κάμερας, μπορούμε είτε να αλλάζουμε την κλίση και την απόσταση του επιπέδου είτε της κάμερας. Για την επιτυχή βαθμονόμηση θα πρέπει να είναι γνωστό ποιά από τα δύο αντικείμενα μετακινούνται. Δεδομένων των παραπάνω, η συγκεκριμένη μεθοδολογία βαθμονόμησης είναι εξαιρετικά απλή σε σχέση με παλαιότερες προτάσεις, και

αποδίδει πολύ καλά αποτελέσματα (Zhang, 1999). Σε πραγματικά δεδομένα όταν χρησιμοποιήθηκαν 2 εικόνες για την βαθμονόμηση, η μέση ρίζα τετραγώνων (root mean square, RMS) των διαφορών των πραγματικών με των υπολογισμένων τιμών των παραμέτρων ήταν 0.761 ± 0.295 , για 3 εικόνες 0.987 ± 0.393 , για 4 εικόνες 0.927 ± 0.361 και για 5 0.881 ± 0.335 .

Αργότερα ο Heikkila (2000), πρότεινε μία μεθοδολογία βαθμονόμησης κάμερας όπου το μοντέλο του επιπέδου αποτελείται από κυκλικά στοιχεία (εικόνα 2.1 (β)). Στην μεθοδολογία του πραγματοποιεί χαρτογράφηση από τις πραγματικές τρισδιάστατες συντεταγμένες των στοιχείων, στις δισδιάστατες συντεταγμένες εικόνας και αντίστροφα από δισδιάστατες συντεταγμένες εικόνας, σε οπτικές γραμμές ή τρισδιάστατα σημεία του πραγματικού πλάνου. Το μοντέλο κάμερας που χρησιμοποίησε, επιτρέπει την χρήση βελτιστοποίησης ελάχιστων τετραγώνων, για την μείωση της παραμόρφωσης των παραμορφωμένων συντεταγμένων της εικόνας. Η ακρίβεια αυτής της τεχνικής είναι της τάξης του 1/50 του μεγέθους των εικονοστοιχείων. Για την επίτευξη τέτοια ακρίβειας ωστόσο χρειάζεται η εξάλειψη πηγών σφάλματος, όπως η χρωματική εκτροπή και ο θόρυβος της εικόνας.

Έπειτα, ο Zhang (2004), πρότεινε μια μεθοδολογία η οποία αντί να χρησιμοποιείται ένα δισδιάστατο επίπεδο πρότυπο για την βαθμονόμηση της κάμερας, χρησιμοποιεί ένα μονοδιάστατο αντικείμενο. Το αντικείμενο αυτό διαθέτει τρία συγγραμικά σημεία, των οποίων η θέση τους πάνω στο αντικείμενο είναι γνωστή. Παρόλο που έχει δειχθεί ότι η βαθμονόμηση της κάμερας χρησιμοποιώντας ένα μονοδιάστατο αντικείμενο, το οποίο πραγματοποιεί ελεύθερη κίνηση, δεν είναι εφικτή, ο Zhang έδειξε ότι μπορεί να υλοποιηθεί εάν το αντικείμενο κινείται γύρω από ένα σταθερο σημείο. Αυτή η τεχνική επιτρέπει την ταυτόχρονη βαθμονόμηση παραπάνω από μιας κάμερας, οι οποίες είναι διατεταγμένες μακριά η μία από την άλλη. Αυτό είναι εφικτό αφού, λόγω της μονοδιάστατης φύσης του αντικειμένου, τα σημεία του μπορούν να είναι φανερά από διάφορες οπτικές γωνίες. Η διαφορά στις τιμές των υπολογισμένων παραμέτρων, σε σχέση και την μεθοδολογία που χρησιμοποιεί το δισδιάστατο μοντέλο, είναι 2%.

Πιο πρόσφατα, το 2006, οι Kannala και Brandt (2006), πρότειναν μία μέθοδο

βαθμονόμησης που έχει εφαρμογή τόσο στο κλασσικό μοντέλο κάμερας αλλά και σε μή τυπικά μοντέλα, όπως το ευρυγώνιο. Η μεθοδολογία που προτείνουν έχει εξίσου απλή εφαρμογή με αυτή του Zhang και χρησιμοποιεί ένα μοντέλο επιπέδου, είτε με τετράγωνα στοιχεία (εικόνα 2.1), είτε με κυκλικά (εικόνα 2.2). Στην δική τους προσέγγιση, το επίπεδο με τα κυκλικά στοιχεία απέδιδε καλύτερα αποτελέσματα σε σχέση με αυτό με τα τετράγωνα. Για την επιτυχή βαθμονόμηση χρησιμοποιώντας αυτή την προσέγγιση, χρειάζεται ο χρήστης να έχει γνώση επί των παραμέτρων της κάμερας. Δηλαδή, η εστιακή απόσταση καθώς και η γωνία θέασης (angle of view) πρέπει να είναι γνωστές. Η μεθοδολογία τους όταν εξετάστηκε σε κλασσική και ευρυγώνια κάμερα το μικρότερο σφάλμα σε μορφή ρίζας μέσων τετραγώνων ήταν 0.052 και 0.057 αντίστοιχα. Για δύο κάμερες τύπου fish-eye τα καλύτερα αποτελέσματα ήταν 0.089 και 0.137.

2.2 Γεωμετρική Οπτική Οδομετρία

Όπως αναφερθηκε και στην εισαγωγή του κεφαλαίου 2 η μέθοδος της γεωμετρικής οπτικής οδομετρίας έχει εφαρμοστεί σε αρκετούς τομείς. Το 2004 παρουσιάστηκε από τους Nister et al. (2004) ένα σύστημα υπολογισμού της κίνησης μιας στερεοσκοπικής διάταξης καμερών ή μιας μόνο κάμερας, καθώς βρίσκονται σε κίνηση. Ο υπολογισμός βασιζόταν αμιγώς σε οπτική πληροφορία και γίνονταν αντιστοιχίζοντας σημεία ενδιαφέροντος μεταξύ ζευγάρια εικόνων. Μετά την αντιστοίχιση των σημείων, υπολόγιζαν την μετατόπιση στο χρονικό διάστημα που είχαν ληφθεί τα συγκεκριμένα ζευγάρια εικόνων. Τα αποτελέσματα που προβάλλουν στην εργασία τους εστιάζουν στην στερεοσκοπική διαταξη των καμερών που ήταν εγκατεστημένες σε ένα όχημα εδάφους. Η ακρίβεια που πετύχαν ήταν εξαιρετικά μεγάλη για αρκετά μεγάλο διάστημα οδήγησης. Συγκεκριμένα όταν το όχημα εκτελούσε κυκλική κίνηση το σφάλμα που σημειώθηκε ήταν 1.07% επί της συνολικής απόστασης, σε ανοιχτό πεδίο 4.86% και σε δασική περιοχή 1.63%.

Πέραν των οχημάτων εδάφους που κινούνται στην γη, η οπτική οδομετρία έχει λειτουργήσει ως μεσο υπολογισμού της κίνησης των ρομποτικών οχημάτων Mars Exploration Rovers. Οι Maimone et al. (2007) σύνταξαν μια εργασία για την απόδοση της οπτικής

οδομετρίας σε σύγκριση με άλλες μεθόδους υπολογισμού της κίνησης, που εκτελούνται από το υπολογιστικό σύστημα των οχημάτων. Τα συμπεράσματα της μελέτης είναι ότι η οπτική οδομετρία είναι ένα αποτελεσματικό εργαλείο για την διατήρηση της ασφάλειας των οχημάτων. Επίσης λειτουργεί σαν ελεγκτικός μηχανισμός για το εάν το όχημα εκτελεί πορεία και εξασφαλίζει την ορθή επιστημονική απεικόνιση του πεδίου. Παρόλα αυτά χρήζει ανθρώπινης βοήθειας σε περιβάλλοντα φτωχά σε οπτικά χαρακτηριστικά (featureless). Για την πλανητική εξερεύνηση η οπτική οδομετρία παρουσιάζει έναν περιορισμό, την επεξεργασία των εικόνων η οποία καταναλώνει αρκετούς υπολογιστικούς πόρους. Πέραν αυτού του προβλήματος που δημιουργείται λόγω των περιορισμένων υπολογιστικών δυνατοτήτων των υπολογιστικών συστημάτων των οχημάτων, η ακρίβεια των μετρήσεων της κίνησης είναι ικανοποιητικές.

Σε μια πιο σύγχρονη εργασία (Huang *et al.*, 2017), παρουσιάζεται μια πειραματική ανάλυση μιας προσέγγισης αυτόνομης πτήσης χρησιμοποιώντας έναν RGB-D αισθητήρα. Η μεθοδολογία συνδιάζει την τεχνική της οπτικής οδομετρίας όπως προτείνεται στην βιβλιογραφία, με μια προηγούμενη δουλειά των συγγραφέων (Henry *et al.*, 2010) για αυτόνομη πτήση και χαρτογράφηση του χώρου. Το προτεινόμενο σύστημα είναι ικανό να εκτελέσει όλους τους υπολογισμούς που χρειάζονται για την τοπική διαχείριση της θέσης του οχήματος. Χρησιμοποιώντας τον αισθητήρα RGB-D το σύστημα μπορεί να προγραμματίσει περίπλοκες τρισδιάστατες πορείες.

2.3 Γεωμετρική Οπτική Οδομετρία Με Εφαρμογές Στην Ενδοσκόπηση

Στο ερευνητικό πλαίσιο των μετρήσεων αποστάσεων σε ενδοσκοπικές εξετάσεις, μία από τις μεγαλύτερες προκλήσεις έγκειται στην μέθοδο εξέτασης μέσω ασύρματης κάψουλας (capsule endoscope, CE). Η έλλειψη δεδομένων, που να προσφέρουν βάση αληθείας από εξετάσεις ή *in-vivo* πειράματα, κάνει την διερεύνηση της δυνατότητας εφαρμογής της οπτικής οδομετρίας, για τον υπολογισμό της θέσης της κάψουλας, άκρως απαιτητική. Για τον λόγο αυτό, οι μελέτες που διερευνούν την εφαρμογή της οπτικής οδομετρίας στην ασύρματη

ενδοσκόπηση, βασίζονται κυρίως σε υπολογιστικές προσομοιώσεις.

Η δυνατότητα της εφαρμογής της οπτικής οδομετρίας για τον υπολογισμό της θέσης του ενδοσκοπίου στο γαστρεντερικό σύστημα έχει μελετηθεί σε προηγούμενες εργασίες. Έρευνα πάνω στην μέτρηση της περιστροφής, της ασύρματης κάψουλας ενδοσκόπησης, μέσα στο γαστρεντερικό σύστημα, καθώς και στις αλλαγές των παραμέτρων της κλίμακας μεταξύ δύο διαδοχικών εικόνων, έδειξε ότι μπορούν να υπολογιστούν επιτυχώς (*Iakovidis et al.*, 2013). Η διαδικασία αυτή, επιτυγχάνεται, με την ταύτιση σημείων μεταξύ δύο διαδοχικών εικόνων, που αντιστοιχούν στο ίδιο σημείο στο τρισδιάστατο πλάνο που απεικονίζουν. Μπόρεσαν έτσι, να υπολογίσουν τον γεωμετρικό μετασχηματισμό που υφίσταται η εικόνα, κατά την κίνηση της κάψουλας. Όσον αφορά την κλίμακα και την περιστροφή, το σφάλμα μεταξύ των υπολογιζόμενων και πραγματικών τιμών ήταν της τάξης του 10^{-1} . Η χαμηλή τιμή του σφάλματος, μεταξύ των πραγματικών και υπολογιζόμενων τιμών, δείχνει ότι υπάρχει η δυνατότητα πρακτικής εφαρμογής της οπτικής οδομετρίας σε τέτοιου είδους βίντεο. Σε επέκταση της παραπάνω έρευνας (*Spyrou & Iakovidis*, 2013), βασιζόμενοι αποκλειστικά σε βίντεο που είχε καταγραφεί από την κάμερα που φέρει η κάψουλα, κατάφεραν να υπολογίσουν την θέση της. Ωστόσο, η μέτρηση της απόστασης που διένυε η κάψουλα μεταξύ των διαδοχικών εικόνων του βίντεο, δεν υπολογίστηκε σε πραγματικές μονάδες μέτρησης (π.χ cm, mm etc.) αλλά σε σχετική κλίμακα.

Ένας σημαντικός παράγοντας, που επηρεάζει θετικά, την απόδοση της οπτικής οδομετρίας, είναι η σωστή εξαγωγή και ταίριασμα σημείων σε δύο διαδοχικές εικόνες. Στα ενδοσκοπικά βίντεο, η διαδικασία εξαγωγής και ταίριασματος των σημείων, είναι ακόμα πιο απαιτητική λόγω της φύσης της επιφάνειας του εσωτερικού των οργάνων. Για τον λόγο αυτό, διεξάχθηκε μία συγκριτική έρευνα (*Spyrou et al.*, 2015) πάνω σε δημοφιλείς αλγόριθμους εξαγωγής χαρακτηριστικών και εντοπισμού σημείων ενδιαφέροντος. Κάποιοι από τους αλγόριθμους που εξετάστηκαν ήταν ο Speeded Up Robust Features (SURF), Scale Invariant Feature Transform (SIFT), Good Features To Track (GFtT) κ.α. Για την εξέταση των αλγόριθμων, χρησιμοποιήθηκε μια παλαιότερη εργασία τους η οποία εστιάζει στην μέτρηση της μετατόπισης και της περιστροφής της κάψουλας (*Spyrou & Iakovidis*, 2013). Τα αποτελέσματα έδειξαν ότι ο καλύτερος σε απόδοση αλγόριθμος για τις ανάγκες της οπτικής

οδομετρίας είναι ο SIFT. Το μέσο απόλυτο σφάλμα, χρησιμοποιώντας τον SIFT για την μέτρηση της περιστροφής της κάψουλας, ήταν της τάξης $\sim 10^{-4}$, και για την μέτρηση της κίνησης της $\sim 10^{-3}$.

Σημαντική συνεισφορά στο πλαίσιο των *in-vivo* οπτικών μετρήσεων χρησιμοποιώντας ασύρματες κάψουλες ενδοσκόπησης, είναι η εργασία των Bao et al. (2013). Οι συγγραφείς δημιούργησαν ένα τρισδιάστατο γραφικό μοντέλο λεπτού εντέρου και τοποθέτησαν μία ψηφιακή προσομοίωση ασύρματης κάψουλας ενδοσκόπησης. Εν τέλει, “τρέχοντας” την προσομοίωση το μέσο σφάλμα που παρατηρήθηκε ήταν 0.04 cm. Παρόλα αυτά, λόγω της επισωρευτικής φύσης των τεχνικών ανίχνευσης κίνησης, το σφάλμα συνέχισε να αυξάνεται μεταξύ των διαδοχικών εικόνων του βίντεο, φτάνοντας τη μέγιστη τιμή των 4 cm. Η συγκεκριμένη έρευνα έπειτα βελτιώθηκε και επεκτάθηκε από τους Mi et al. (2014). Οι Mi et al. πρότειναν μία μεθοδολογία για να θεσπιστεί σχέση μεταξύ της διαμέτρου του εντέρου και της μαύρης τρύπας που φαίνεται όταν το ενδοσκόπιο βρίσκεται σε παράλληλη τροχιά με το έντερο. Η σχέση αυτή χρησιμεύει για την διόρθωση των αποτελεσμάτων της οπτικής οδομετρίας, έτσι ώστε, να αντικατοπτρίζουν την πραγματική απόσταση που διανύεται. Το μέσο σφάλμα του υπολογισμού της θέσης που σημειώνεται μέσα στην εργασία είναι 0.81 και 1.6. Ωστόσο, υπάρχει επιφύλαξη για τις δύο παραπάνω προσεγγίσεις, διότι δεν βασίζονται σε πραγματικά *in-vivo* ή *ex-vivo* πειράματα, αλλά σε υπολογιστικές προσομοιώσεις.

2.4 “Εξυπνη” Βαθμονόμηση Κάμερας Και Οπτική Οδομετρία

Η επιτυχημένη εφαρμογή της γεωμετρικής οπτικής οδομετρίας, βασίζεται στην σωστή βαθμονόμηση της κάμερας που χρησιμοποιείται κάθε φορά. Παρόλο που ως διαδικασία μπορεί να θεωρηθεί απλή, προϋποθέτει τη γνώση του γεωμετρικού μοντέλου της κάμερας (π.χ αν είναι ευρυγώνιο, τύπου “fisheye” κλπ.), έτσι ώστε να υποστηρίζεται από την μεθοδολογία βαθμονόμησης που χρησιμοποιείται. Η γνώση του γεωμετρικού μοντέλου είναι άκρως σημαντική για την σωστή υλοποίηση της μεθοδολογίας. Παρόλα αυτά, ακόμα και όταν το μοντέλο της κάμερας είναι γνωστό, οι κλασσικές μέθοδοι βαθμονόμησης δεν είναι κατάλληλες για όλα τα όλα τα μοντέλα. Για αυτό τον λόγο καταφεύγουμε σε μεθοδολογίες που βασίζονται

στην μηχανική μάθηση.

Μία ενδιαφέρουσα προσέγγιση, στην διαδικασία βαθμονόμησης της κάμερας, προτάθηκε από τους Ahmed et al. (1999). Για τον υπολογισμό των παραμέτρων της κάμερας, χρησιμοποίησαν ένα σχετικά απλό τεχνητό νευρωνικό δίκτυο. Το δίκτυο αποτελούνταν από 3 επίπεδα (layers), ένα επίπεδο εισόδου (input layer) δεδομένων, ένα κρυφό (hidden layer), ένα επίπεδο εξόδου (output layer) και χρησιμοποιούσε ως συνάρτηση ενεργοποίησης την σιγμοειδή (αρχιτεκτονική MLP). Το δίκτυο ήταν σχεδιασμένο να αντιστοιχεί τα τρισδιάστατα σημεία του πραγματικού κόσμου (επιπέδου) σε δισδιάστατα σημεία του πλάνου της εικόνας. Η αντιστοιχία γινόταν με τρόπο τέτοιο, ώστε τα βάρη του κρυφού επιπέδου να αντιστοιχούν στις εξωτερικές παραμέτρους της κάμερας και τα βάρη του επιπέδου της εξόδου, στις εσωτερικές παραμέτρους της κάμερας. Η συγκεκριμένη δουλειά αναφέρεται αποκλειστικά στο θέμα της βαθμονόμησης της κάμερας. Με την συγκεκριμένη μεθοδολογία δεν χρειάζεται να υπάρχει καμία γνώση επί των παραμέτρων ή του μοντέλου της κάμερας, αφού έδειξαν ότι μπορεί να υπάρξει σύγκλιση ακόμα και όταν γίνεται τυχαία αρχικοποίηση των βαρών του δικτύου. Παρόμοια μέθοδος, χρησιμοποιώντας μια στέρεοσκοπική διάταξη καμερών και αρχιτεκτονική δικτύου MLP, με τρία επίπεδα, προτάθηκε από τους Memon και Khan (2001). Ωστόσο, δεν υπολόγιζαν τις παραμέτρους της κάμερας, αλλά χρησιμοποιούσαν τις δισδιάστατες αντιστοιχίες σημείων μεταξύ των εικόνων, για την τρισδιάστατη ανακατασκευή του χώρου απεικόνισης.

Μία πιο σύγχρονη προσέγγιση βαθμονόμησης, μιας στερεοσκοπικής διάταξης καμερών, μελετήθηκε από τον Besdok (2009). Ο Besdok αντικατέστησε την αρχιτεκτονική του MLP με αυτή της ακτινικής συνάρτησης βάσεως (Radial Basis Function, RBF) και εκπαίδευσε το δίκτυο χρησιμοποιώντας έναν γενετικό αλγόριθμο. Αυτό το δίκτυο χρησιμοποιήθηκε για την αντιστοίχιση των δισδιάστατων σημείων (συντεταγμένων) του πλάνου της εικόνας, σε τρισδιάστατα σημεία του πραγματικού κόσμου. Με αυτόν τον τρόπο κατάφερε την εξαγωγή των παραμέτρων των καμερών της στερεοσκοπικής διάταξης. Όπως και στην προηγούμενη δουλειά που αναφέρθηκε, χρησιμοποιήθηκε αρχιτεκτονική δικτύου με 3 επίπεδα, που λόγω της απλότητάς της, καθιστά την μεθοδολογία υπολογιστικά αποδοτική.

Φυσικά, με την άνοδο της “βαθιάς μηχανικής μάθησης” (deep learning) και την μείωση του χρόνου εκπαίδευσης, μέσω της εκμετάλλευσης της μονάδας επεξεργασίας γραφικών (GPU), εμπνεύστηκαν διάφοροι μέθοδοι οπτικής οδομετρίας. Μία αρχιτεκτονική βαθιάς μάθησης αποτελείται από μια πολυεπίπεδη συστοιχία απλών μονάδων, εκ των οποίων, όλες ή οι περισσότερες υπόκεινται σε εκπαίδευση και αρκετές υπολογίζουν ως έξοδο μη γραμμικές χαρτογραφήσεις δεδομένων κάποιων εισόδων (*LeCun et al., 2015*). Οι Konda και Memisevic (2015) πρότειναν μία μεθοδολογία βασισμένη σε συνελκτικά νευρωνικά δίκτυα (Convolutional Neural Networks (*Krizhevsky et al., 2012*)). Χρησιμοποίησαν ένα ζευγάρι συνελκτικών δικτύων, εκ των οποίων το ένα εκπαιδεύτηκε να υπολογίζει την ταχύτητα του αντικειμένου που έφερε την κάμερα και το άλλο την κατεύθυνσή του. Επαλήθευσαν την μεθοδολογία τους πάνω στο σετ δεδομένων του KITTI (*Menze & Greiger, 2015*), το οποίο είναι σημείο αναφοράς για την οπτική οδομετρία στο πλαίσιο υπολογισμού της κίνησης οχημάτων. Αποτελείται από βίντεο τα οποία έχουν καταγραφεί χρησιμοποιώντας κάμερες σε στέρεο διάταξη τοποθετημένες σε οχήματα εν κινήσει. Χρησιμοποιώντας τα δεδομένα της κατεύθυνσης και της ταχύτητας που υπολογίστηκαν από τα δύο συνελκτικά νευρωνικά δίκτυα, κατάφεραν να ανακατασκευάσουν την πορεία που διένυσαν τα οχήματα, κατά την κίνηση τους.

3. Μεθοδολογία

Για την εξέταση της δυνατότητας μετρήσεων, χρησιμοποιώντας αποκλειστικά οπτική πληροφορία, σε ενδοσκοπικά βίντεο αναπτύχθηκαν τρεις διαφορετικές μεθοδολογίες. Στην παρούσα εργασία θα παρουσιαστούν οι δύο εξ αυτών. Η μία βασισμένη στην κλασσική γεωμετρική, παραμετρική, μέθοδο της οπτικής οδομετρίας, ενώ η δεύτερη σε μία “έξυπνη” προσέγγιση του προβλήματος, χρησιμοποιώντας ένα τεχνητό νευρωνικό δίκτυο. Με την σύγκριση των δύο μεθόδων, εκ των αποτελεσμάτων που απέδωσαν (όπως θα δούμε και αναλυτικά στην ενότητα 4), η τελευταία μέθοδος απέδωσε καλύτερα χωρίς την ανάγκη γνώσης επί των εσωτερικών παραμέτρων και του γεωμετρικού μοντέλου της κάμερας.

3.1 Γεωμετρική Παραμετρική Προσέγγιση

Όπως αναφέραμε και στην ενότητα 2 η γεωμετρική προσέγγιση στο πρόβλημα της οπτικής οδομετρίας είναι ευρέως διαδεδομένη. Σημειώνει αρκετές επιτυχημένες εφαρμογές σε πραγματικές συνθήκες πέραν των ελεγχόμενων πειραματικών δοκιμών. Έχει λειτουργήσει ως μέσω πλοήγησης ρομπότ, drones και οχημάτων. Λαμβάνοντας υπ’ όψιν τα παραπάνω θα ήταν λάθος να μην προσπαθήσουμε να εξετάσουμε την δυνατότητα εφαρμογής αυτής της μεθοδολογίας σε ενδοσκοπικά εργαλεία, για την μέτρηση των αποστάσεων που διανύουν κατά την εκτέλεση της εξέτασης.

Συνοπτικά τα βήματα που πρέπει να ακολουθηθούν για την επιτυχή υλοποίηση της μεθοδολογίας της οπτικής οδομετρίας είναι τα εξής:

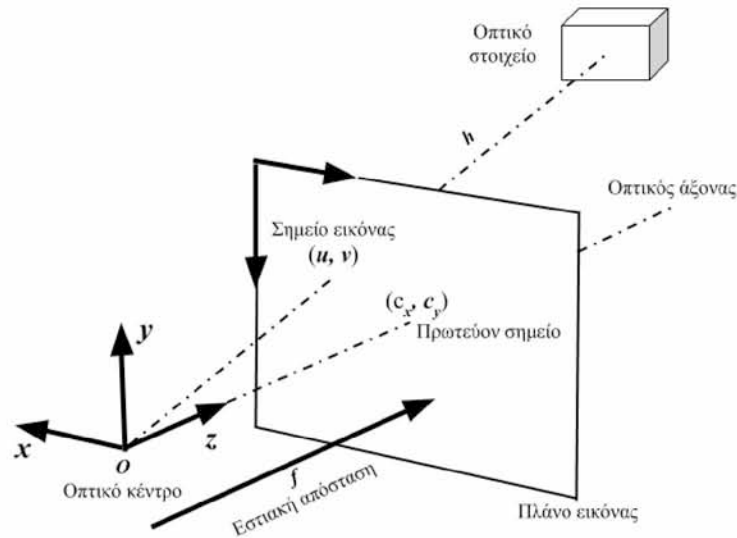
- i. Βαθμονόμηση της κάμερας που χρησιμοποιούμε, για τον υπολογισμό των εσωτερικών και εξωτερικών της παραμέτρων. Εφόσον γίνει επιτυχώς η βαθμονόμηση, μπορούμε να προχωρήσουμε στον υπολογισμό της πορείας της κάμερας, μέσω της αλληλουχίας των εικόνων ή του βίντεο που έχει καταγραφεί.

- ii. Δεδομένης μιας ακολουθίας εικόνων (ή στιγμιότυπων ενός βίντεο), $I_i, I_{i+1}, I_{i+2}, I_{i+3}, \dots, I_{i+n}$, θα πρέπει να εντοπίσουμε σημεία ενδιαφέροντος (points of interest) από δύο διαδοχικές εικόνες I_i και I_{i+1} .
- iii. Έχοντας εντοπίσει τα σημεία, από το βήμα (ii), έχουμε πλέον γνώση επί των συντεταγμένων τους, αλλά και επίσης γνωρίζουμε τα διανύσματα χαρακτηριστικών που τα χαρακτηρίζουν (πχ. χρώμα, εάν αντιστοιχούν σε γωνίες κλπ.). Με τις παραπάνω πληροφορίες ταιριάζουμε τα σημεία χρησιμοποιώντας κάποιο σχετικό αλγόριθμο (πχ RANSAC, KLT, BF Matcher κλπ.) για να αποκτήσουμε αντιστοιχίες δισδιάστατων σημείων, μεταξύ δύο εικόνων, που αναφέρονται στο ίδιο τρισδιάστατο σημείο στον χώρο.
- iv. Εδραιώνοντας τις αντιστοιχίες των σημείων, ταιριάζοντας τα, υπολογίζουμε τους πίνακες μετατροπής (translation matrix) και περιστροφής (rotation matrix), μέσω του επιπολικού πίνακα (essential matrix) (Nistér et al., 2004).
- v. Τέλος για την ανακατασκευή της πορείας, πρέπει να υπολογίσουμε τον παράγοντα κλιμάκωσης, ούτως ώστε, να γίνει κλιμάκωση των αποτελεσμάτων για να αντιστοιχούν σε πραγματικές μονάδες μέτρησης (πχ. cm, mm κλπ.).

3.1.1 Βαθμονόμηση Κάμερας

Η βαθμονόμηση είναι ένα απαραίτητο βήμα για την υλοποίηση της γεωμετρικής οπτικής οδομετρίας. Μέσω της διαδικασίας της βαθμονόμησης υπολογίζονται οι εσωτερικές παράμετροι της κάμερας, που είναι μεγίστης σημασίας για την μετάβαση από το δισδιάστατο επίπεδο, στο τρισδιάστατο φυσικό χώρο. Στην συγκεκριμένη μεθοδολογία υιοθετήθηκε η μέθοδος βαθμονόμησης του Zhang (Zhang, 1999) όπως υλοποιήθηκε από τον Bouget για το περιβάλλον ανάπτυξης Matlab (Bouget, 2004). Παρόλο που έρευνα στα πλαίσια αυτής της εργασίας (Iakovidis et al., 2016) έδειξε ότι η μέθοδος βαθμονόμησης που προτείνεται από τους Kannala και Brandt, δίνει καλύτερα αποτελέσματα, απαιτεί γνώση, εκ των προτέρων, επί των εσωτερικών παραμέτρων της κάμερας. Επειδή αυτές δεν παρέχονται συνήθως από τους

κατασκευαστές, στην προτεινόμενη μεθοδολογία χρησιμοποιούμε την μεθοδολογία του Zhang η οποία αναλύεται παρακάτω.



Εικόνα 3.2. Γραφική αναπαράσταση των βασικών στοιχείων του μοντέλου κάμερας μικροσκοπικής οπής (pinhole).

Ένα δισδιάστατο σημείο ορίζεται ως $\mathbf{m} = [\mathbf{u}, \mathbf{v}]^T$ και το αντίστοιχο τρισδιάστατο ως $\mathbf{M} = [X, Y, Z]^T$. Χρησιμοποιούμε τη σήμανση $\tilde{\mathbf{x}}$ για να ορίσουμε το επαυξημένο διάνυσμα προσθέτοντας 1 ως τελευταίο στοιχείο, οπότε τα \mathbf{m} και \mathbf{M} διαμορφώνονται ως $\tilde{\mathbf{m}} = [\mathbf{u}, \mathbf{v}, 1]^T$ και $\tilde{\mathbf{M}} = [X, Y, Z, 1]^T$. Εάν η κάμερα ακολουθεί το μοντέλο της κάμερας μικροσκοπικής οπής (pinhole camera model) (εικόνα 3.1) η σχέση ανάμεσα στο τρισδιάστατο σημείο \mathbf{M} και της προβολής του, στο πλάνο της εικόνας \mathbf{m} , δίνεται από την εξής σχέση :

$$s\tilde{\mathbf{m}} = A[\mathbf{R} \ \mathbf{t}]\tilde{\mathbf{M}} \quad (3.1)$$

όπου το s είναι ένας αυθαίρετος παράγοντας κλιμάκωσης, τα (\mathbf{R}, \mathbf{t}) ονομάζονται εξωτερικές παράμετροι και συμβολίζουν την περιστροφή και την μετατροπή του σημείου, και ο A συμβολίζει την πίνακα των εσωτερικών παραμέτρων. Ο πίνακας A αναλύεται ως εξής :

$$A = \begin{bmatrix} \alpha & \gamma & u_0 \\ 0 & \beta & v_0 \\ 0 & 0 & 1 \end{bmatrix} \quad (3.2)$$

όπου $(\mathbf{u}_0, \mathbf{v}_0)$ είναι οι συντεταγμένες του προτεύοντος σημείου, \mathbf{a} και \mathbf{b} είναι οι παράγοντες κλιμάκωσης της εικόνας στους \mathbf{u} και \mathbf{v} άξονες και τέλος γ είναι η παράμετρος της στρέβλωσης των αξόνων της εικόνας. Χρησιμοποιείται η συντομογραφία A^{-T} για $(A^{-1})^T$ ή $(A^T)^{-1}$.

Χωρίς να χάνεται η γενικότητα, θεωρούμε ότι το πλάνο του μοντέλου βρίσκεται στον $Z = 0$ άξονα του συστήματος συντεταγμένων του πραγματικού κόσμου. Ας ορίσουμε την i -οστή στήλη του πίνακα περιστροφής \mathbf{R} ως \mathbf{r}_i . Από την σχέση (3.1) έχουμε :

$$s \begin{bmatrix} \mathbf{u} \\ \mathbf{v} \\ 1 \end{bmatrix} = A[\mathbf{r}_1 \ \mathbf{r}_2 \ \mathbf{r}_2 \ t] \begin{bmatrix} X \\ Y \\ 0 \\ 1 \end{bmatrix} = A[\mathbf{r}_1 \ \mathbf{r}_2 \ \mathbf{r}_2 \ t] \begin{bmatrix} X \\ Y \\ 1 \end{bmatrix} \quad (3.3)$$

Με κατάχρηση του συμβολισμού, συνεχίζουμε να χρησιμοποιούμε το \mathbf{M} για να συμβολίσουμε ένα σημείο πάνω στο πλάνο του μοντέλου, αλλά $\mathbf{M} = [X, Y]^T$ αφού το Z είναι συνέχεια ίσο με το 0. Άρα, $\tilde{\mathbf{M}} = [X, Y, 1]^T$. Οπότε, ένα σημείο του μοντέλου \mathbf{M} και της εικόνας \mathbf{m} είναι συσχετικμένα μέσω μιας ομογραφία \mathbf{H} :

$$s\tilde{\mathbf{m}} = \mathbf{H}\tilde{\mathbf{M}}, \text{ με } \mathbf{H} = A[\mathbf{r}_1 \ \mathbf{r}_2 \ t] \quad (3.4)$$

Όπως είναι φανερό, ο πίνακας \mathbf{H} , με διαστάσεις 3×3 ορίζεται όταν κλιμακώνεται από κάποιον παράγοντα κλιμάκωσης.

Δεδομένης μιας εικόνας του πλάνου του μοντέλου που χρησιμοποιούμε, μπορούμε να υπολογίσουμε την ομογραφία. Ας ορίζουμε την ομογραφία ως $\mathbf{H} = [\mathbf{h}_1 \ \mathbf{h}_2 \ \mathbf{h}_3]$. Από την σχέση (3.4) έχουμε :

$$[\mathbf{h}_1 \ \mathbf{h}_2 \ \mathbf{h}_3] = \lambda A[\mathbf{r}_1 \ \mathbf{r}_2 \ t] \quad (3.5)$$

όπου λ είναι μια αυθαίρετη βαθμωτή μεταβλητή. Εφόσον τα \mathbf{r}_1 και \mathbf{r}_2 είναι γνωστό ότι είναι ορθοκανονικά, έχουμε :

$$\mathbf{h}_1^T A^{-T} A^{-1} \mathbf{h}_2 = 0 \quad (3.6)$$

$$\mathbf{h}_1^T \mathbf{A}^{-T} \mathbf{A}^{-1} \mathbf{h}_1 = \mathbf{h}_2^T \mathbf{A}^{-T} \mathbf{A}^{-1} \mathbf{h}_2 \quad (3.7)$$

Οι σχέσεις (3.6) και (3.7) δηλώνουν τους δύο βασικούς περιορισμούς των εσωτερικών παραμέτρων συναρτήσει μιας ομογραφίας. Λόγω του ότι η ομογραφία προσφέρει 8 βαθμούς ελευθερίας και υπάρχουν 6 εξωτερικές παράμετροι (3 για την περιγραφή της περιστροφής και 3 για την περιγραφή της μετατροπής), μπορούμε να αποκτήσουμε μόνο 2 περιορισμούς για τις εσωτερικές παραμέτρους. Πρέπει να σημειωθεί ότι $\mathbf{A}^{-T} \mathbf{A}^{-1}$ (Luong & Faugeras, 1997) περιγράφει την εικόνα της απόλυτης κωνικής (absolute conic).

Επεκτείνοντας τις σχέσεις (3.6) και (3.7) στην απόλυτη κωνική, μπορούμε να πιστοποιήσουμε, ότι το πλάνο του μοντέλου, περιγράφεται στο σύστημα συντεταγμένων της κάμερας από την ακόλουθη εξίσωση:

$$\begin{bmatrix} r_3 \\ r_3^T t \end{bmatrix}^T \begin{bmatrix} x \\ y \\ z \\ -w \end{bmatrix} = 0 \quad (3.8)$$

όπου $w = 0$ για σημεία στο άπειρο και $w = 1$ αλλού. Αυτό το πλάνο τέμνει το πλάνο στο άπειρο σε μία ευθεία. Μπορούμε εύκολα να παρατηρήσουμε ότι $\begin{bmatrix} r_1 \\ 0 \end{bmatrix}$ και $\begin{bmatrix} r_2 \\ 0 \end{bmatrix}$ είναι δύο σημεία πάνω σε αυτή την ευθεία. Κάθε σημείο πάνω της, είναι ένας γραμμικός συνδιασμός αυτών των δύο σημείων, δηλαδή:

$$\mathbf{x}_\infty = a \begin{bmatrix} r_1 \\ 0 \end{bmatrix} + b \begin{bmatrix} r_2 \\ 0 \end{bmatrix} = \begin{bmatrix} ar_1 + br_2 \\ 0 \end{bmatrix} \quad (3.9)$$

Για να υπλογιστεί η τομή της παραπάνω ευθείας με την απόλυτη κωνική, εξ ορισμού, το σημείο \mathbf{x}_∞ , είναι γνωστό ως κυκλικό σημείο και ικανοποιεί την εξής σχέση:

$$\mathbf{x}_\infty^T \mathbf{x}_\infty = 0 \quad (3.10)$$

$$(\mathbf{a}r_1 + \mathbf{b}r_2)^T (\mathbf{a}r_1 + \mathbf{b}r_2) = 0, \text{ ή } a^2 + b^2 = 0 \quad (3.11)$$

Η λύση σχέσης (3.11) είναι $\mathbf{b} = \pm a\mathbf{i}$, όπου $\mathbf{i}^2 = -1$. Έτσι, τα δύο σημεία τομής είναι :

$$\mathbf{x}_\infty = \mathbf{a} \begin{bmatrix} \mathbf{r}_1 \pm i\mathbf{r}_2 \\ \mathbf{0} \end{bmatrix} \quad (3.12)$$

Η προβολή των σημείων στο πλάνο της εικόνας δίνεται, κλιμακώμενη από έναν παράγοντα κλίμακας, από την σχέση :

$$\tilde{\mathbf{m}}_\infty = \mathbf{A}(\mathbf{r}_1 \pm i\mathbf{r}_2) = \mathbf{h}_1 \pm i\mathbf{h}_2 \quad (3.13)$$

Το σημείο $\tilde{\mathbf{m}}_\infty$ βρίσκεται στην εικόνα της απόλυτης κωνικής, η οποία περιγράφεται ως $\mathbf{A}^T \mathbf{A}^{-1}$. Αυτό μας δίνει την εξής σχέση :

$$(\mathbf{h}_1 \pm i\mathbf{h}_2)^T \mathbf{A}^{-T} \mathbf{A}^{-1} (\mathbf{h}_1 \pm i\mathbf{h}_2) = \mathbf{0} \quad (3.14)$$

Απαιτώντας το πραγματικό και το φανταστικό μέρος της σχέσης (3.14) να είναι ίσο με το μηδέν οδηγούμαστε στις σχέσεις (3.6) (3.7).

Για την λύση του προβλήματος της βαθμονόμησης υποθέτουμε ότι :

$$\begin{aligned} \mathbf{B} = \mathbf{A}^{-T} \mathbf{A}^{-1} &= \begin{bmatrix} \mathbf{B}_{11} & \mathbf{B}_{21} & \mathbf{B}_{31} \\ \mathbf{B}_{12} & \mathbf{B}_{22} & \mathbf{B}_{32} \\ \mathbf{B}_{13} & \mathbf{B}_{23} & \mathbf{B}_{33} \end{bmatrix} \\ &= \begin{bmatrix} \frac{1}{\alpha^2} & -\frac{\gamma}{\alpha^2 \beta} & \frac{v_0 \gamma - u_0 \beta}{\alpha^2 \beta} \\ -\frac{1}{\alpha^2 \beta} & \frac{\gamma^2}{\alpha^2 \beta^2} + \frac{1}{\beta^2} & -\frac{\gamma(v_0 \gamma - u_0 \beta)}{\alpha^2 \beta^2} - \frac{v_0}{\beta^2} \\ \frac{v_0 \gamma - u_0 \beta}{\alpha^2 \beta} & -\frac{\gamma(v_0 \gamma - u_0 \beta)}{\alpha^2 \beta^2} & \frac{\gamma(v_0 \gamma - u_0 \beta)}{\alpha^2 \beta^2} + \frac{v_0^2}{\beta^2} + \mathbf{1} \end{bmatrix} \end{aligned} \quad (3.15)$$

Πρέπει να σημειωθεί ότι ο πίνακας \mathbf{B} είναι συμμετρικός και ορίζεται από έναν 6-διάστατο διάνυσμα :

$$\mathbf{b} = [\mathbf{B}_{11}, \mathbf{B}_{12}, \mathbf{B}_{22}, \mathbf{B}_{13}, \mathbf{B}_{23}, \mathbf{B}_{33}]^T \quad (3.16)$$

Ας υποθέσουμε, ότι η i -οστή στήλη του διανύσματος \mathbf{H} είναι $\mathbf{h}_i = [\mathbf{h}_{i1}, \mathbf{h}_{i2}, \mathbf{h}_{i3}]^T$. Τότε προκύπτει ότι :

$$\mathbf{h}_i^T \mathbf{B} \mathbf{h}_j = \mathbf{v}_{ij}^T \mathbf{b} \quad (3.17)$$

όπου v_{ij} :

$$v_{ij} = [h_{i1}h_{j1}, h_{i1}h_{j2} + h_{i2}h_{j1}, h_{i2}h_{j2}, h_{i3}h_{j1} + h_{i1}h_{j3}, h_{i3}h_{j2} + h_{i2}h_{j3}, h_{i3}h_{j3}]^T \quad (3.18)$$

Επομένως οι δύο θεμελιώδεις περιορισμοί που εκφράζονται στις σχέσεις (3.6) (3.7), μπορούν να γραφτούν σαν δύο ομογενείς εξισώσεις στην \mathbf{b} ως εξής :

$$\begin{bmatrix} v_{12}^T \\ (v_{11} - v_{22})^T \end{bmatrix} \mathbf{b} = \mathbf{0} \quad (3.19)$$

Εάν παρατηρηθούν n εικόνες του πλάνου του μοντέλου, βάζοντας τις με τέτοιο τρόπο όπως ορίζει η σχέση (3.19), προκύπτει το εξής :

$$V \mathbf{b} = \mathbf{0} \quad (3.20)$$

Ο V είναι ένας πίνακας με διαστάσεις $2n \times 6$. Εάν το $n \geq 3$, θα έχουμε μία μοναδική λύση του \mathbf{b} η οποία θα ορίζεται από έναν παράγοντα κλιμάκωσης. Εάν το $n = 2$, μπορούμε να θεωρήσουμε ότι ο παράγοντας στρέβλωσης είναι $\gamma = 0$, δηλαδή, $[\mathbf{0}, 1, \mathbf{0}, \mathbf{0}, \mathbf{0}, \mathbf{0}] \mathbf{b} = \mathbf{0}$, η οποία σχέση προσθέτεται σαν επιπλέον εξίσωση στην σχέση (3.20). Η λύση της σχέσης (3.20) είναι γνωστή ως το ιδιοδιάνυσμα του $V^T V$ το οποίο σχετίζεται με την μικρότερη ιδιοτιμή.

Μόλις υπολογιστεί το \mathbf{b} , μπορούμε να υπολογίσουμε όλες τις εσωτερικές παραμέτρους της κάμερας, δηλαδή τα στοιχεία του πίνακα A . Όταν όλα τα στοιχεία του πίνακα A γίνουν γνωστά, οι εξωτερικές παράμετροι της κάμερας είναι έτοιμες να υπολογιστούν. Από την σχέση (3.4) έχουμε :

$$\mathbf{r}_1 = \lambda A^{-1} \mathbf{h}_1 \quad (3.21)$$

$$\mathbf{r}_2 = \lambda A^{-1} \mathbf{h}_2 \quad (3.22)$$

$$\mathbf{r}_3 = \mathbf{r}_1 \times \mathbf{r}_2 \quad (3.23)$$

$$\mathbf{t} = \lambda \mathbf{A}^{-1} \mathbf{h}_3 \quad (3.24)$$

με το λ να είναι ίσο με :

$$\lambda = \frac{1}{\|\mathbf{A}^{-1} \mathbf{h}_1\|} = \frac{1}{\|\mathbf{A}^{-1} \mathbf{h}_2\|} \quad (3.24)$$

Στην μεθοδολογία του ο Zhang συμπεριλαμβάνεται επίσης και η παραμόρφωση που προκύπτει από τον φακό της κάμερας. Για την λύση του προβλήματος αυτού, θεωρούμε ότι οι (\mathbf{u}, \mathbf{v}) είναι οι ιδανικές, μη παραμορφωμένες, συντεταγμένες εικονοστοιχείων της εικόνας και ως $(\tilde{\mathbf{u}}, \tilde{\mathbf{v}})$ τις αντίστοιχες συντεταγμένες οι οποίες έχουν παρατηρηθεί στην εικόνα (παραμορφωμένες). Τα ιδανικά σημεία είναι οι προβολές των σημείων του μοντέλου μικροσκοπικής οπής. Ομοίως, (\mathbf{x}, \mathbf{y}) και $(\tilde{\mathbf{x}}, \tilde{\mathbf{y}})$ είναι οι ιδανικές (δεν παρουσιάζουν παραμόρφωση) και οι πραγματικές (παρουσιάζουν παραμόρφωση (εικόνα 3.6)) κανονικοποιημένες συντεταγμένες της εικόνας αντίστοιχα. Έτσι έχουμε ότι :

$$\tilde{\mathbf{x}} = \mathbf{x} + \mathbf{x}[\mathbf{k}_1(\mathbf{x}^2 + \mathbf{y}^2) + \mathbf{k}_2(\mathbf{x}^2 + \mathbf{y}^2)^2] \quad (3.25)$$

$$\tilde{\mathbf{y}} = \mathbf{y} + \mathbf{y}[\mathbf{k}_1(\mathbf{x}^2 + \mathbf{y}^2) + \mathbf{k}_2(\mathbf{x}^2 + \mathbf{y}^2)^2] \quad (3.26)$$

όπου \mathbf{k}_1 και \mathbf{k}_2 είναι οι συντελεστές της ακτινικής παραμόρφωσης (radial distortion). Οι συντεταγμένες του κέντρου της ακτινικής παραμόρφωσης είναι οι ίδιες με αυτές του προτεύοντος σημείου. Από τις σχέσεις :

$$\tilde{\mathbf{u}} = \mathbf{u}_0 + \alpha \tilde{\mathbf{x}} + \gamma \tilde{\mathbf{y}} \quad (3.27)$$

$$\tilde{\mathbf{v}} = \mathbf{v}_0 + \beta \tilde{\mathbf{y}} \quad (3.28)$$

και υποθέτοντας ότι $\gamma = \mathbf{0}$, προκύπτουν οι εξής σχέσεις :

$$\tilde{\mathbf{u}} = \mathbf{u} + (\mathbf{u} - \mathbf{u}_0)[\mathbf{k}_1(\mathbf{x}^2 + \mathbf{y}^2) + \mathbf{k}_2(\mathbf{x}^2 + \mathbf{y}^2)^2] \quad (3.29)$$

$$\tilde{\mathbf{v}} = \mathbf{v} + (\mathbf{v} - \mathbf{v}_0)[\mathbf{k}_1(\mathbf{x}^2 + \mathbf{y}^2) + \mathbf{k}_2(\mathbf{x}^2 + \mathbf{y}^2)^2] \quad (3.30)$$

3.1.2 Εξαγωγή Σημείων και Ταίριασμα

Για την εξαγωγή των σημείων στην συγκεκριμένη μεθοδολογία, χρησιμοποιείται ο αλγόριθμος SIFT (Lowe, 2004). Η επιλογή του SIFT βασίστηκε σε μία συγκριτική μελέτη, για αλγορίθμους εξαγωγής χαρακτηριστικών και σημείων από εικόνες που έχουν ληφθεί από ασύρματες κάμερες ενδοσκόπηση (Spyrou et al., 2015). Η σύγκριση έγινε με γνώμονα το πως τα σημεία που εξάγει ο κάθε αλγόριθμος επηρεάζουν τα αποτελέσματα της οπτικής οδομετρίας. Η συγκεκριμένη έρευνα έδειξε, ότι ο αλγόριθμος SIFT αποδίδει καλύτερα από τους υπολοίπους στο πλαίσιο της οπτικής οδομετρίας.

Τα χαρακτηριστικά τα οποία στοχεύει να εντοπίσει ο SIFT, είναι σχεδιασμένα έτσι, ούτως ώστε να αντιστοιχούν σε σημεία ενδιαφέροντος, εύρωστα έναντι διάφορων μετασχηματισμών που υφίσταται η εικόνα. Ο συγκεκριμένος αλγόριθμος παρουσιάζει ευρωστία σε μετασχηματισμούς όπως αυτοί της κλιμάκωσης και της περιστροφής. Πέραν της ανθεκτικότητας του SIFT σε μετασχηματισμούς, σαν κριτήριο λαμβάνεται υπ' όψιν και η αντοχή, των σημείων που εξάγει, στον θόρυβο αλλά και στην αλλαγή της φωτεινότητας. Μετά από ένα στάδιο όπου η εικόνα περνάει από μια αλυσίδα φίλτρων και αρκετών συνελίξεων με γκαουσιανά φίλτρα, σε διάφορες κλίμακες, οι εικόνες που αντιπροσωπεύουν διαφορές της γκαουσιανής (Differences of Gaussians, DoGs) ομαδοποιούνται σε οκτάβες. Σε αυτό το σημείο, μετά από διάφορους μετασχηματισμούς κλίμακας κ.α, υπολογίζονται τα τοπικά μέγιστα που παρουσιάζονται στις εικόνες, και αυτά που έχουν χαμηλή αντίθεση ή αντιστοιχούν σε ακμές απορρίπτονται.

Τα διανύσματα χαρακτηριστικών του SIFT, υπολογίζονται σε μια “γειτονιά” κοντά στα σημεία ενδιαφέροντος που εξάγονται. Τα διανύσματα αυτά αποτελούνται από ένα ιστόγραμμα τιμών του προσανατολισμού της κλίσης των ακμών του σημείου, το οποίο ιστόγραμμα είναι κανονικοποιημένο, σε ένα μοναδιαίο διάνυσμα. Με την κανονικοποίηση του ιστογράμματος σε ένα μοναδιαίο διάνυσμα εξασφαλίζεται η ευρωστία των σημείων ενδιαφέροντος σε αλλαγές της φωτεινότητας. Τέλος, στο διάνυσμα χαρακτηριστικών του SIFT, εμπεριέχονται οι επικρατέστεροι προσανατολισμοί της κλίσης των σημείων, καθώς και σε ποια τιμή

κλιμάκωσης εντοπίστηκαν.

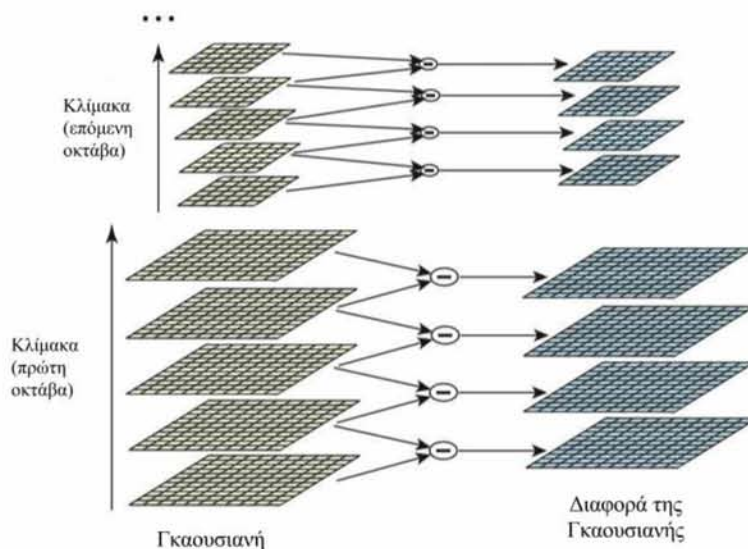
Πιο αναλυτικά, η έναρξη εύρεσης σημείων ενδιαφέροντος ξεκινάει κάνοντας συνέλιξη με γκαουσιανά φίλτρα την εικόνα από την οποία θέλουμε να εξάγουμε τα σημεία, σε διάφορες κλίμακες της εικόνας. Έπειτα, υπολογίζουμε τις διαφορές της γκαουσιανής των θολωμένων (blurred) εικόνων. Τα σημεία ενδιαφέροντος επιλέγονται στις συντεταγμένες της εικόνας που προκύπτουν μέγιστα και ελάχιστα των διαφορών της γκαουσιανής σε διάφορες τιμές κλιμάκωσης. Μία εικόνα διαφορών της γκαουσιανής σημειώνεται ως $D(x, y, \sigma)$ και προκύπτει ως εξής:

$$D(x, y, \sigma) = L(x, y, k_i \sigma) - L(x, y, k_j \sigma) \quad (3.31)$$

Ως $L(x, y, k\sigma)$ σημειώνεται το προϊόν της συνέλιξης της εικόνας $I(x, y)$ με το γκαουσιανό φίλτρο $G(x, y, k\sigma)$, που μπορεί να εκφραστεί τυπικά ως:

$$L(x, y, k\sigma) = G(x, y, k\sigma) * I(x, y) \quad (3.32)$$

Έτσι, οι διαφορές της γκαουσιανής, μεταξύ των παραγόντων κλιμάκωσης $k_i\sigma$ και $k_j\sigma$ είναι απλά οι διαφορές της εικόνας που έχουν συνελιχθεί με το γκαουσιανό φίλτρο σε κλίμακες $k_i\sigma$ και $k_j\sigma$.



Εικόνα 3.2. Γραφική αναπαράσταση του βήματος δημιουργίας οκτάβων σε διάφορες κλίμακες, και τον υπολογισμό της διαφοράς της γκαουσιανής του αλγορίθμου SIFT.

Για την εύρεση των μεγίστων στον χώρο των διάφορων κλιμάκων, όπως προαναφέρθηκε, γίνεται συνέλιξη των κλιμακωμένων εικόνων με γκαουσιανό φίλτρο. Έπειτα, οι εικόνες-προϊόντα αυτών των συνελίξεων, ομαδοποιούνται σε οκτάβες όπου η κάθε οκτάβα αντιστοιχεί στον διαπλασιασμό της μεταβλητής σ . Η μεταβλητή k_i θέτεται έτσι, ώστε να υπάρχει συγκεκριμένος αριθμός εικόνων-προϊόντων συνέλιξης ανά οκτάβα. Τέλος, οι διαφορές της γκαουσιανής υπολογίζονται μεταξύ γειτονικών εικόνων σε μία οκτάβα (εικόνα 3.2). Στο στάδιο αυτό, αφού έχουν υπολογιστεί οι διαφορές της γκαουσιανής, γίνεται η πρώτη επιλογή σημείων ενδιαφέροντος από τον αλγόριθμο SIFT. Τα σημεία ενδιαφέροντος, επιλέγονται βάση των μεγίστων και ελαχίστων της διαφοράς της γκαουσιανής μεταξύ των εικόνων σε διάφορες κλίμακες. Αυτό επιτυγχάνεται, συγκρίνοντας κάθε εικονοστοιχείο των εικόνων της διαφοράς της γκαουσιανής με τα 8 γειτονικά του εικονοστοιχεία (δηλ. τα εικονοστοιχεία περιμετρικά αυτού) της ίδιας κλίμακας, και των 9 εικονοστοιχείων (δηλ. του αντίστοιχού του συν των περιμετρικών αυτού) της εικόνας της αμέσως μεγαλύτερης και μικρότερης κλίμακας. Εάν, η τιμή του εικονοστοιχείου είναι μέγιστη ή ελάχιστη σε σχέση με αυτών που συγκρίθηκε, σημειώνεται ως υποψήφιο σημείο ενδιαφέροντος.

Ο αριθμός των σημείων ενδιαφέροντος που προκύπτει από την εύρεση μεγίστων και ελαχίστων από την παραπάνω διαδικασία είναι πολύ μεγάλος. Για την μείωση του αριθμού των σημείων και την επιλογή αυτών που είναι καλύτερα, γίνεται απόρριψη αυτών που έχουν χαμηλή αντίθεση ή αυτών που βρίσκονται πάνω σε ακμές. Αρχικά, για κάθε υποψήφιο σημείο ενδιαφέροντος στις διάφορες κλίμακες, γίνεται παρεμβολή (interpolation) των κοντινών δεδομένων, για την ακριβή εύρεση του υποψήφιου σημείου, συναρτήσει της αρχικής εικόνας. Αυτό είναι αναγκαίο διότι σε πρώτη φάση, το κάθε μέγιστο και ελάχιστο της διαφοράς της γκαουσιανής υπολογίζεται σε τιμές κλιμάκωσης της αρχικής εικόνας. Η παρεμβολή γίνεται χρησιμοποιώντας το τετραγωνικό ανάπτυγμα Taylor της συνάρτησης κλίμακας-χώρου διαφοράς της γκαουσιανής, $D(x, y, \sigma)$ με το υποψήφιο σημείο ενδιαφέροντος ως σημείο αναφοράς. Αυτό το ανάπτυγμα Taylor ορίζεται ως:

$$D(x) = D + \frac{dD^T}{dx} x + \frac{1}{2} x^T \frac{d^2D}{dx^2} x \quad (3.33)$$

όπου το D και η παράγωγός του μελετούνται στο υποψήφιο σημείο αναφοράς που εξετάζεται κάθε φορά και $x = (x, y, \sigma)^T$ είναι η μετατόπιση (offset) αυτού του σημείου. Η θέση του μεγίστου \hat{x} , αποφασίζεται υπολογίζοντας την παράγωγο της σχέσης (3.3) και θέτοντας την ίση με το 0. Εάν η μετατόπιση του σημείου μεγίστου \hat{x} είναι μεγαλύτερη από 0.5 προς οποιαδήποτε κατεύθυνση, θεωρείται ότι το υποψήφιο σημείο βρίσκεται κοντά σε άλλο μέγιστο. Σε αυτή την περίπτωση το εξεταζόμενο, υποψήφιο, σημείο ενδιαφέροντος απορρίπτεται και συνεχίζεται η αξιολόγηση του επόμενου. Εάν, η μετατόπιση είναι μικρότερη από 0.5 τότε αυτή προστίθεται στο εξεταζόμενο σημείο, και το νέο σημείο που προκύπτει, είναι η αντιστοιχία του υποψήφιου στην αρχική εικόνα.

Για την απόρριψη των σημείων ενδιαφέροντος με χαμηλή αντίθεση, ο αλγόριθμος SIFT υπολογίζει την τιμή του αναπτύγματος Taylor δευτέρου βαθμού, $D(x)$, στην θέση του μεγίστου \hat{x} (μετά από την μετατόπιση). Εάν η τιμή στο σημείο αυτό είναι μικρότερη του 0.03, τότε το υποψήφιο σημείο απορρίπτεται. Εάν η τιμή είναι μεγαλύτερη του 0.03, τότε το υποψήφιο σημείο διατηρείται ως σημείο ενδιαφέροντος.

Λόγω της χρήσης διαφορών της γκαουσιανής σημειώνονται πολλά σημεία ενδιαφέροντος σε σημεία που εμφανίζονται ακμές, ακόμα και αν αυτά τα σημεία δεν παρουσιάζουν ευρωστία στον θόρυβο. Για τον λόγο αυτό, ο SIFT απορρίπτει τα σημεία αυτά, που δεν είναι ανθεκτικά στον θόρυβο και σημειώνονται σε ακμές. Για τον εντοπισμό των λανθασμένα επιλεγμένων σημείων ενδιαφέροντος στις εικόνες διαφορών της γκαουσιανής, θεωρείται ότι η κύρια καμπυλότητα εγκαρσίως των ακμών θα είναι μεγαλύτερη σε σχέση αυτής κατά μήκος των ακμών. Η εύρεση των τιμών της κύριας καμπυλότητας, γίνεται λύνοντας ως προς τις ιδιοτιμές (eigen values) του εσιανού πίνακα, H , δευτέρου βαθμού:

$$H = \begin{bmatrix} D_{xx} & D_{xy} \\ D_{xy} & D_{yy} \end{bmatrix} \quad (3.34)$$

Οι ιδιοτιμές του H είναι ανάλογες τις κύριας καμπυλότητας του D . Συμπεραίνεται ότι, ο λόγος

των δύο ιδιοτιμών του \mathbf{H} , όπου α η πρώτη και β η δεύτερη, ορίζεται ως:

$$\exists \alpha > \beta, r = \frac{\alpha}{\beta} \quad (3.35)$$

Το ίχνος του πίνακα \mathbf{H} , δηλαδή το άθροισμα $D_{xx} + D_{yy}$, δίνει το άθροισμα των δύο ιδιοτιμών α και β , ενώ η ορίζουσα, δηλαδή $D_{xx}D_{yy} - D_{xy}^2$ το γινόμενο. Το πηλίκο,

$$\mathbf{R} = \frac{D_{xx} + D_{yy}}{D_{xx}D_{yy} - D_{xy}^2} = \frac{Tr(\mathbf{H})^2}{Det(\mathbf{H})} \quad (3.36)$$

ισούται με το εξής πηλίκο:

$$\frac{(r+1)^2}{r} \quad (3.37)$$

το οποίο βασίζεται αμιγώς στο πηλίκο των ιδιοτιμών. Η ελάχιστη τιμή του \mathbf{R} , προκύπτει όταν $\alpha = \beta$. Έτσι, όσο μεγαλύτερη είναι η διαφορά των απολύτων τιμών των ιδιοτιμών, τόσο μεγαλύτερη είναι η διαφορά των απολύτων τιμών των κύριων καμπυλοτήτων του \mathbf{D} και μεγαλύτερη τιμή του \mathbf{R} . Εν τέλει, για κάποιο κατώφλι που ορίζεται για το πηλίκο των ιδιοτιμών, $r_{\text{threshold}}$, εαν το \mathbf{R} , για κάποιο υποψήφιο σημείο ενδιαφέροντος, είναι μεγαλύτερο από το πηλίκο της σχέσης (3.7), όπου $r = r_{\text{threshold}}$, το σημείο αυτό απορρίπτεται.

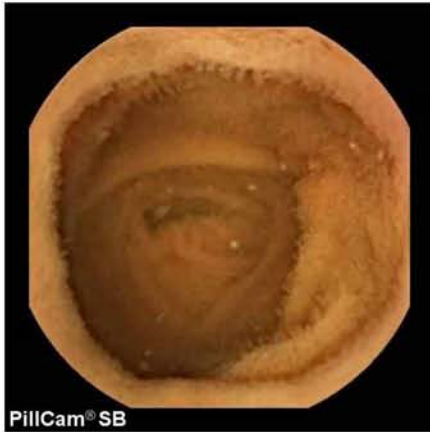
Το τελευταίο βήμα του αλγορίθμου SIFT πριν την τελική εξαγωγή των σημείων ενδιαφέροντος μαζί με τα διανύσματα χαρακτηριστικών που τους αντιστοιχούν, είναι η ανάθεση προσανατολισμού σε καθένα από αυτά τα σημεία. Σε κάθε σημείο αναθέεται ένας ή παραπάνω προσανατολισμοί βάσει των διαφορετικών κατευθύνσεων κλίσης που παρατηρούνται τοπικά. Με αυτό το βήμα επιτυγχάνεται η ευρωστία των σημείων ενδιαφέροντος σε μετατροπές περιστροφής που μπορεί να γίνουν στην εικόνα. Αρχικά, στην εικόνα που έχει εφαρμοστεί το γκαουσιανό φίλτρο, $L(\mathbf{x}, \mathbf{y}, \sigma)$, ο προσανατολισμός του σημείου μελετάται σε μια περιοχή ανάλογου μεγέθους του παράγοντα κλιμάκωσης σ . Έτσι επιτυγχάνεται ευρωστία του σημείου σε μετασχηματισμούς κλίμακας. Για μια εικόνα $L(\mathbf{x}, \mathbf{y})$ οποία είναι κλιμακωμένη βάση ενός παράγοντα κλιμάκωσης σ , το πλάτος της κλίσης $m(\mathbf{x}, \mathbf{y})$, και ο προσανατολισμός $\theta(\mathbf{x}, \mathbf{y})$, υπολογίζονται ως εξής:

$$m(x, y) = \sqrt{(L(x+1, y) - L(x-1, y))^2 + (L(x, y+1) - L(x, y-1))^2} \quad (3.38)$$

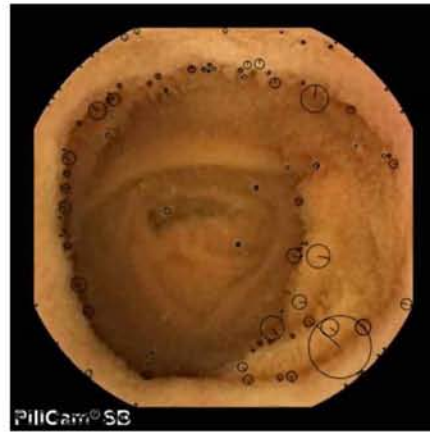
$$\theta(x, y) = \text{atan2}(L(x, y+1) - L(x, y-1), L(x+1, y) - L(x-1, y)) \quad (3.39)$$

Το πλάτος και ο προσανατολισμός της κλίσης του κάθε σημείου ενδιαφέροντος, υπολογίζονται για μια περιοχή, ανάλογη της κλίμακας της εικόνας L , γύρω από το σημείο. Έπειτα οι προσανατολισμοί που υπολογίζονται σχηματίζουν ένα ιστόγραμμα, μήκους 36 bin, όπου το κάθε bin καλύπτει 10° . Οι κορυφές του ιστογράμματος ενός σημείου ενδιαφέροντος, αντιστοιχούν στον επικρατέστερο προσανατολισμό του σημείου αυτού.

Στην παρακάτω εικόνα (εικόνα 3.3) φαίνονται σημεία ενδιαφέροντος που εξάχθηκαν χρησιμοποιώντας τον αλγόριθμο SIFT. Οι εικόνες που φαίνονται έχουν ληφθεί χρησιμοποιώντας μία ασύρματη κάψουλα ενδοσκόπησης και προέρχονται από μια τυπική εξέταση (*in-vivo*) αλλά και από το πείραμα που χρησιμοποιείται για την επαλήθευση της μεθοδολογίας της παρούσας εργασίας (*ex-vivo*).



(α)



(β)



(γ)



(δ)

Εικόνα 3.3. (α) Μία τυπική εικόνα η οποία λήφθηκε από την ασύρματη κάψουλα ενδοσκόπησης κατά την διάρκεια εξέτασης (β) Σημεία ενδιαφέροντος που εξάχθηκαν από την εικόνα (α) χρησιμοποιώντας τον SIFT αλγόριθμο (γ) Μία εικόνα η οποία λήφθηκε χρησιμοποιώντας μία κάψουλα ασύρματης ενδοσκόπησης από το εσωτερικό του τεχνητού εντέρου του πειράματός μας (δ) Σημεία ενδιαφέροντος που εξάχθηκαν από την εικόνα (γ) χρησιμοποιώντας τον SIFT αλγόριθμο. Τα σημεία ενδιαφέροντος του SIFT σημειώνονται από τα κέντρα των κύκλων που φαίνονται στις εικόνες (β) και (δ). Η κλίμακα τους είναι ανάλογη της ακτίνας του κύκλου που τα περιβάλλει και ο προσανατολισμός τους σημειώνεται από την ακτίνα.

Αφού εξάγουμε σημεία κλειδιά από δύο διαδοχικές εικόνες I_i και I_{i+1} καταλήγουμε σε δύο σύνολα συντεταγμένων εικονοστοιχείων που αντιστοιχούν στις δύο εικόνες. Για το

ταίριασμα αυτών των σημείων, σε πρώτο χρόνο χρησιμοποιείται ο KLT ανιχνευτής (*Tomasl & Kanade, 1991*). Ο KLT είναι ένας από τους πιο σημαντικούς και ευρέως διαδεδομένους αλγορίθμους ανίχνευσης χαρακτηριστικών (σημείων-κλειδιών) π.χ σε διαδοχικά στιγμιότυπα βίντεο. Η πρωταρχική εκδοχή του KLT αλγορίθμου είχε προταθεί για την καταχώρηση εικόνων. Ο συγκεκριμένος αλγόριθμος χρησιμοποιεί χωρική πληροφορία, αναφορικά με την ένταση της φωτεινότητας της εικόνας, έτσι ώστε να κατευθύνει την έρευνα για την ταύτιση των σημείων ενδιαφέροντος, προς την κατεύθυνση που μεγιστοποιείται η οπτική ροή (optical flow). Ο KLT είναι ένας γρήγορος και αποδοτικός αλγόριθμος που συνήθως χρησιμοποιείται μαζί με τον GftT (good features to track) (*Shi, 1994*) και αποδίδει ικανοποιητικά αποτελέσματα ακόμα και σε περιπτώσεις όπου το σύνολο των σημείων ενδιαφέροντος είναι μικρό.

Η μεθοδολογία του KLT βασίζεται στην υπόθεση, ότι υπάρχει διατήρηση της φωτεινότητας μεταξύ των στιγμιότυπων ενός βίντεο τα οποία είναι κοντά χρονικά, το ένα με το άλλο. Με την εξαγωγή (ή την επιλογή), σημείων ενδιαφέροντος από ένα στιγμιότυπο και υποθέτοντας ότι η κίνηση μεταξύ του πλάνου ανάμεσα στα στιγμιότυπα του βίντεο είναι σχετικά μικρή, ο KLT στοχεύει στον εντοπισμό των νέων θέσεων των εξαγομένων σημείων-κλειδιών ενός στιγμιότυπου στο επόμενο. Σε αντίθεση με προηγούμενες προσεγγίσεις, ο συγκεκριμένος αλγόριθμος δεν βασίζεται σε ανίχνευση και ταίριασμα προτύπων (pattern matching), αντιθέτως, δίνει μία αναλυτική και γραμμική λύση στο πρόβλημα. Συγκεκριμένα, η προαναφερθείσα υπόθεση, μπορεί να γραφεί τυπικά ως εξής:

$$I(x, y, t + \Delta t) = I(x - \Delta x, y - \Delta y, t) \quad (3.40)$$

Έστω $\mathbf{x} = (x, y)^T$ ένα εικονοστοιχείο (pixel) ενός στιγμιότυπου I_i και ως $\mathbf{d} = (\Delta x, \Delta y)^T$, η μετατόπισή του σε ένα διαδοχικό στιγμιότυπο I_{i+1} . Έπειτα μπορούμε να ορίσουμε αυτό το στιγμιότυπο ως $J(\mathbf{x}) = I(\mathbf{x} - \mathbf{d} + \varepsilon)$. Ο στόχος του αλγορίθμου είναι να υπολογίσει την μετατόπιση του \mathbf{d} , έτσι ώστε να πετύχει ελαχιστοποίηση του σφάλματος ε , που γενικευμένα και εντός ενός “παραθύρου” W μπορεί να εκφραστεί ως:

$$\varepsilon = \int_w w [I(\mathbf{x} - \mathbf{d}) - J(\mathbf{x})]^2 d\mathbf{x} \quad (3.41)$$

Η ανάλυση του KLT λύνει το ακόλουθο γραμμικό σύστημα:

$$\mathbf{Z}\mathbf{d} = \varepsilon \quad (3.42)$$

όπου \mathbf{Z} ένας 2×2 πίνακας που ορίζεται ως εξής:

$$\mathbf{Z} = \iint_w \mathbf{g}(\mathbf{x})\mathbf{g}^T(\mathbf{x})w(\mathbf{x})d\mathbf{x} \quad (3.43)$$

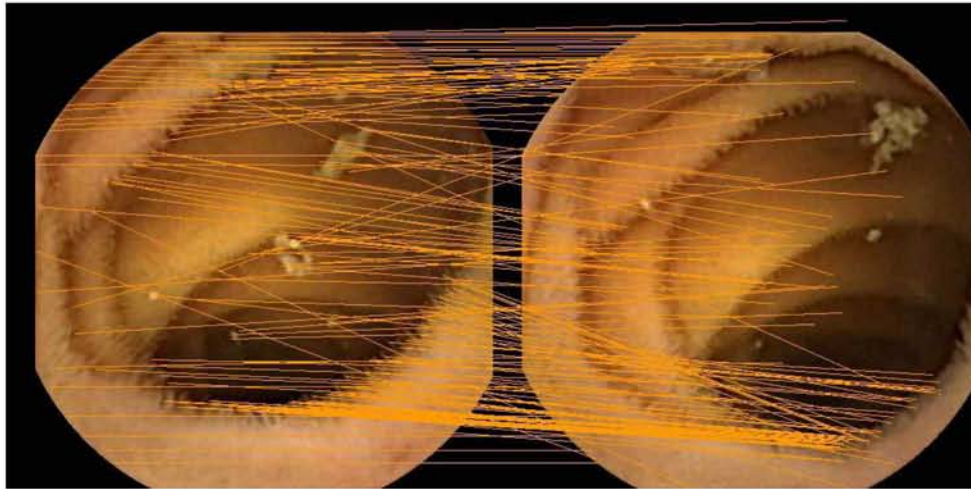
Ενώ το ακόλουθο είναι 2×1 διάνυσμα,

$$\varepsilon = \iint_w [I(\mathbf{x}) - J(\mathbf{x})]\mathbf{g}(\mathbf{x})w(\mathbf{x})d\mathbf{x} \quad (3.44)$$

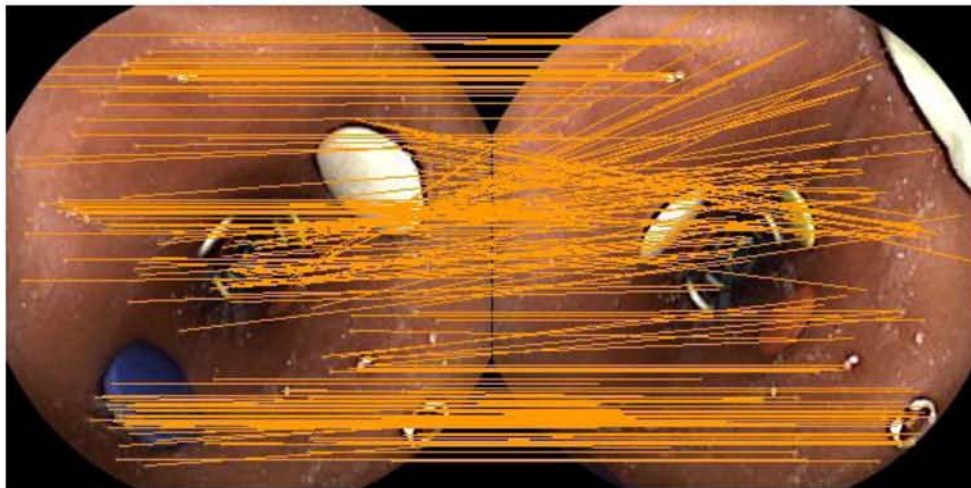
και η συνάρτηση \mathbf{g} ορίζεται ως:

$$\mathbf{g} = \left[\frac{d}{dx}(I + J); \frac{d}{dy}(I + J) \right] \quad (3.45)$$

Ολοκληρώνοντας, πρέπει να σημειωθεί ότι ο KLT χρησιμοποιεί χωρική πληροφορία όσον αφορά την ένταση της φωτεινότητας για να καθοδηγήσει την έρευνα, προς την κατεύθυνση που εξασφαλίζει το βέλτιστο ταίριασμα των σημείων ενδιαφέροντος, μεταξύ των εικόνων/στιγμιότυπων. Επίσης, αυτός ο αλγόριθμος είναι ταχύτερος, σε σχέση με άλλες προσεγγίσεις στο πλαίσιο της οπτικής ροής, αφού ελέγχει για αρκετά λιγότερα πιθανά ταίρια ανάμεσα στα στιγμιότυπα. Στην ακόλουθη εικόνα (εικόνα 3.4) απεικονίζονται τα αποτελέσματα του KLT ανιχνευτή, χρησιμοποιώντας σημεία-κλειδιά, που έχουν εξαχθεί χρησιμοποιώντας τον αλγόριθμο SIFT σε εικόνες που έχουν ληφθεί από ασύρματη κάψουλα ενδοσκόπησης.



(α)



(β)

Εικόνα 3.4. Ανιχνευμένα ταίρια σημείων, χρησιμοποιώντας τον KLT αλγόριθμο, σε διαδοχικά στιγμιότυπα βίντεο παρμένα από μια τυπική εξέταση χρησιμοποιώντας μία ασύρματης κάψουλας ενδοσκόπησης (α) και από το ρομποτικό πείραμα στο οποίο βασίστηκε η επαλήθευση της προτεινόμενης μεθοδολογίας (β). Μπορεί να παρατηρηθεί ότι υπάρχουν αρκετά λανθασμένες αντιστοιχίες σημείων ενδιαφέροντος μεταξύ των δύο εικόνων πράγμα το οποίο οδηγεί σε λανθασμένα αποτελέσματα της γεωμετρικής οπτικής οδομετρίας.

Μπορούμε εύκολα να παρατηρήσουμε στην εικόνα 3.4 ότι χρησιμοποιώντας μόνο τον KLT καταλήγουμε σε αρκετές λανθασμένες αντιστοιχίες σημείων μεταξύ των εικόνων. Τα

λανθασμένα “ταίρια”, όταν θα χρησιμοποιηθούν για να υπολογιστεί ο επιπολικός πίνακας (υποενότητα 3.1.3, σχέση (3.51)), θα αποτελέσουν κύριο παράγοντα απόδοσης εσφαλμένων αποτελεσμάτων σχετικά με την μετατόπιση της κάμερας. Για να λυθεί αυτό το πρόβλημα έγινε εισαγωγή ενός ακόμα βήματος στην διαδικασία της αντιστοιχίας των σημείων ενδιαφέροντος. Για να αφαιρέσουμε τις λανθασμένες αντιστοιχίες, γίνεται εφαρμογή του αλγορίθμου RANSAC πάνω στο σύνολο των αντιστοιχισμένων σημείων από τον KLT.

Ο αλγόριθμος RANSAC είναι σχεδιασμένος έτσι ώστε να επιλέγει το βέλτιστο μοντέλο απόφασης αντιστοιχιών, με την παρουσία λάθος επιλεγμένων σημείων. Ο παράγοντας της αποτελεσματικότητας του μοντέλου είναι φραγμένος από μία πιθανότητα, η οποία επιλέγεται από τον χρήστη, πριν την εφαρμογή του αλγορίθμου. Στο πλαίσιο της συγκεκριμένης έρευνας, ο αλγόριθμος αυτός εφαρμόζεται για να βρεθεί ο ακριβέστερος γεωμετρικός μετασχηματισμός, μεταξύ δύο διαδοχικών στιγμιότυπων του βίντεο και τα σημεία που τον ακολουθούν. Οι αντιστοιχίες αυτές των σημείων οι οποίες βρίσκονται εντός του ορίου λάθους του αλγορίθμου, είναι αντιστοιχίες οι οποίες έχουν εντοπιστεί από τον KLT και ακολουθούν τον μετασχηματισμό που ικανοποιεί το μοντέλο του RANSAC. Αντίθετα, αυτές οι αντιστοιχίες που βρίσκονται εκτός του ορίου λάθους του RANSAC, είναι αυτές που η οπτική πληροφορία που έχουν τα σημεία τους ταυτίζεται, αλλά δεν ακολουθούν τον γεωμετρικό μετασχηματισμό.

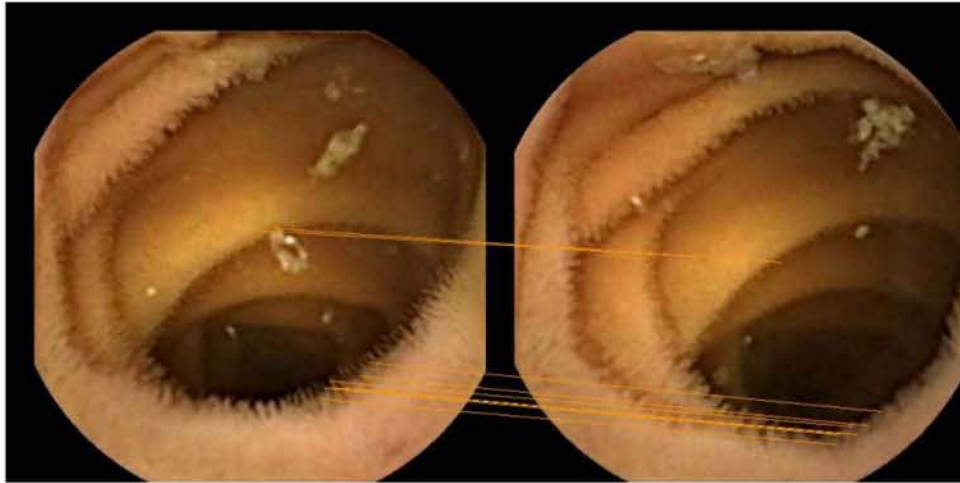
Συγκεκριμένα, έστω T σημειώνουμε ένα συσχετισμένο (affine) μετασχηματισμό που χαρτογραφεί κάθε σημείο ενδιαφέροντος x_i ενός στιγμιότυπου ενός βίντεο στο αντίστοιχο x'_i του διαδοχικού στιγμιότυπου. Δεδομένου των αντιστοιχιών που μας δίνει ως αποτέλεσμα ο KLT, π.χ. όλα τα ζευγάρια $x_i \rightarrow x'_i$ το T μπορεί να χαρακτηριστεί ως:

$$x'_i = T \cdot x_i \quad (3.46)$$

Πρέπει να σημειωθεί ότι η εφαρμογή του RANSAC για την λύση του προβλήματος υπολογισμού του γεωμετρικού μετασχηματισμού, προέρχεται από την διαδικασία της βαθμονόμησης στερεοσκοπικά διατεταγμένων καμερών (Hartley & Zisserman, 2003). Παρομοίως με προηγούμενη δουλειά των Spyrou και Iakovidis, ο RANSAC εφαρμόζεται σε διαδοχικά στιγμιότυπα βίντεο που έχουν καταγραφεί από μία κάμερα η οποία κινείται με αργή

ταχύτητα (δηλαδή ανάμεσα σε δύο στιγμιότυπα το πλάνο απεικονίζεται από διαφορετική προοπτική).

Χρησιμοποιώντας, σαν πρώτο βήμα, τον KLT για το ταίριασμα των σημείων ενδιαφέροντος που έχουν εξαχθεί με τον SIFT, συνήθως αναμένονται λανθασμένα ταίρια σημείων λόγω θορύβου στην εικόνα ή από “παράσιτα” (artifacts) που δημιουργούνται λόγω συμπίεσης. Για αυτόν ακριβώς τον λόγο, επιλέγεται η εφαρμογή του RANSAC στα αποτελέσματα του KLT, αφού έχει την δυνατότητα να υπολογίσει σωστά τον γεωμετρικό μετασχηματισμό που υφίσταται το πλάνο της εικόνας, με την παρουσία εσφαλμένων αντιστοιχιών σημείων. Στην παρακάτω εικόνα (εικόνα 3.5) απεικονίζεται το αποτέλεσμα της εφαρμογής του RANSAC στα ταίρια σημείων της εικόνας 3.4. Είναι φανερό ότι τα εναπομείναντα ταίρια σημείων αντιστοιχούν στα ίδια σημεία του πραγματικού αντικειμένου και εν τέλει, ικανοποιούν το σωστό γεωμετρικό μετασχηματισμό που αντιστοιχεί στην κίνηση της κάμερας.



(α)



(β)

Εικόνα 3.5. Στις παραπάνω εικόνες απεικονίζονται τα αποτελέσματα της εφαρμογής του RANSAC στις αντιστοιχίες σημείων της εικόνας 3.4 για την εξάλειψη των εσφαλμένων αποτελεσμάτων: (α) Τυπικό βίντεο εξέτασης από ασύρματη κάψουλαενδοσκόπησης (β) Βίντεο της ασύρματης κάψουλας από τεχνητό έντερο. Όλες οι αντιστοιχίες που απεικονίζονται είναι οι σωστές αντιστοιχίες γεωμετρικού μετασχηματισμού όπως υπολογίστηκαν από τον RANSAC.

3.1.3 Ανακατασκευή πορείας

Η σχέση μεταξύ των δισδιάστατων (2D) σημείων/εικονοστοιχείων μιας εικόνα ή ενός στιγμιότυπου βίντεο και των αντίστοιχων τρισδιάστατων (3D) συντεταγμένων του

πραγματικού πλάνου που απεικονίζονται περιγράφεται από την εξής εξίσωση:

$$\mathbf{s} \cdot [x \ y \ 1] = [X \ Y \ Z \ 1] \cdot \mathbf{P} \quad (3.47)$$

Η μεταβλητή \mathbf{s} συμβολίζει τον παράγοντα κλιμάκωσης, οι συντεταγμένες (x, y) αντιστοιχούν στις συντεταγμένες των εικονοστοιχείων της εικόνας/στιγμιότυπου, τα (X, Y, Z) συμβολίζουν τις συντεταγμένες του τρισδιάστατου πλάνου και ως \mathbf{P} συμβολίζεται ο πίνακας της κάμερας (camera matrix). Ο πίνακας της κάμερας μπορεί να γραφεί ως εξής:

$$\mathbf{P} = \begin{bmatrix} \mathbf{R} \\ \mathbf{t} \end{bmatrix} \mathbf{K} \quad (3.48)$$

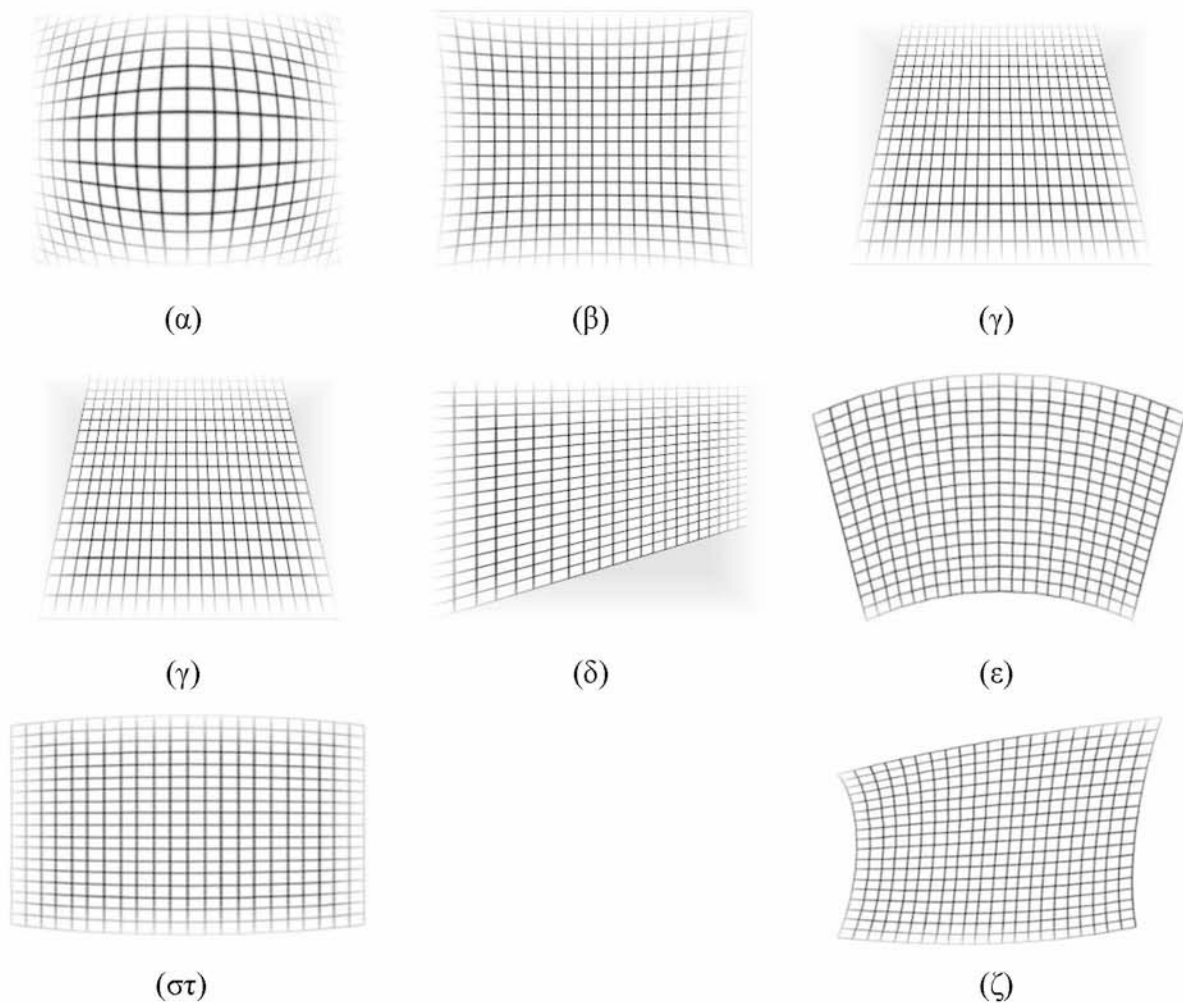
Στην παραπάνω εξίσωση (3.48) ο \mathbf{R} συμβολίζει τον πίνακα περιστροφής (rotation matrix) και ο \mathbf{t} τον πίνακα μετατροπής (translation matrix). Ο \mathbf{K} συμβολίζει τον πίνακα εσωτερικών παραμέτρων της κάμερας και αναλύεται ως:

$$\mathbf{K} = \begin{bmatrix} f_x & 0 & 0 \\ a & f_y & 0 \\ c_x & c_y & 1 \end{bmatrix} \quad (3.49)$$

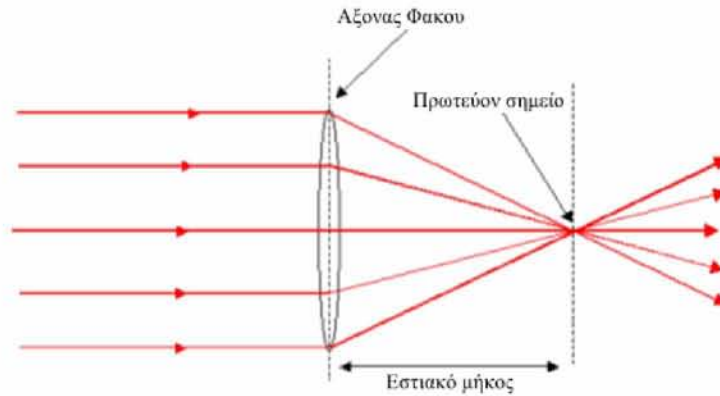
όπου οι μεταβλητές f_x και f_y αντιστοιχούν στην εστιακή απόσταση (focal length) της κάμερας στους x και y άξονες εκφρασμένη σε μονάδες εικονοστοιχείων (pixel units),

$$f_x = \frac{f}{p_x}, f_y = \frac{f}{p_y} \quad (3.50)$$

όπου f είναι η εστιακή απόσταση εκφρασμένη σε mm, τα p_x, p_y αντιστοιχούν στις διαστάσεις των εικονοστοιχείων σε mm, τα c_x, c_y είναι οι συντεταγμένες του πρωτεύοντος σημείου (principal point) σε μονάδες εικονοστοιχείων και η μεταβλητή a είναι ο συντελεστής παραποιημένης στρέβλωσης (skew distortion coefficient) η οποία έχει μη μηδενική τιμή εάν οι άξονες της εικόνας δεν είναι κάθετοι μεταξύ τους.



Εικόνα 3.6. Γραφική αναπαράσταση διαφόρων τύπων παραμόρφωσης που προκύπτουν λόγω του είδους το φακού της κάμερας. Στην εικόνα (α) αναπαρησάται επίπεδο που έχει υποστεί παράμορφωση τύπου βαρελιού (barrel distortion) (β) αναπαρησάται επίπεδο που έχει υποστεί παραμόρφωση τύπου μαξιλαριού (pincushion distortion) (γ) αναπαρησάται επίπεδο που έχει υποστεί παραμόρφωση τύπου προοπτικής (perspective distortion) (δ) αναπαρησάται επίπεδο που έχει υποστεί παραμόρφωση τύπου παραποιημένης στρέβλωσης (skew distortion) (ε) αναπαρησάται επίπεδο που έχει υποστεί παραμόρφωση τύπου καμπυλωμένου ορίζοντα (curved horizon distortion) (στ) αναπαρησάται επίπεδο που έχει υποστεί παραμόρφωση τύπου πανοράματος (panorama distortion) (ζ) αναπαρησάται επίπεδο που έχει υποστεί πολύπλοκη παραμόρφωση (complex distortion).



Εικόνα 3.7. Γραφική αναπαράσταση του εστιακού μήκους (focal length) και του πρωτεύοντος σημείου (principal point) ως προς τον άξονα του φακού (lensa xis).

Όπως μπορεί να παρατηρηθεί από τις παραπάνω σχέσεις για να λύσουμε επιτυχώς την σχέση (3.47) και να μπορέσουμε έπειτα να ανακατασκευάσουμε την πορεία της κάμερας, θα πρέπει πρώτα να υπολογιστούν οι πίνακες \mathbf{K} , \mathbf{R} και \mathbf{t} . Τα στοιχεία του πίνακα των εσωτερικών παραμέτρων της κάμερας \mathbf{K} τα υπολογίζουμε ακολουθώντας την διαδικασία της βαθμονόμησης (υποενότητα 3.1.1). Έπειτα για να υπολογιστούν οι πίνακες περιστροφής \mathbf{R} και μετατροπής \mathbf{t} θα πρέπει πρώτα να υπολογιστεί ο επιπολικός πίνακας (essential matrix).

Για να υπολογίσουμε τον επιπολικό πίνακα, θεωρούμε ένα ζευγάρι ομογενών κανονικοποιημένων συντεταγμένων εικόνας \mathbf{u} και \mathbf{u}' σε διαδοχικά ληφθέντων εικόνων/στιγμιότυπων βίντεο $\mathbf{I}_i, \mathbf{I}_{i+1}$ αντίστοιχα. Τα δύο αυτά σημεία, \mathbf{u} και \mathbf{u}' , αντιστοιχούν στο ίδιο τρισδιάστατο σημείο του πλάνου. Ο επιπολικός πίνακας ορίζεται ως εξής:

$$(\mathbf{u}')^T \mathbf{E} \mathbf{u} = 0 \quad (3.51)$$

Τα ζευγάρια των σημείων, \mathbf{u} και \mathbf{u}' , βρίσκονται ακολουθώντας την μεθοδολογία της υποενότητας (3.1.2). Έχοντας λοιπόν αυτές της αντιστοιχίες σημείων, μπορούμε πλέον να υπολογίσουμε τον επιπολικό πίνακα.

Ο επιπολικός πίνακας μπορεί να εκφραστεί ως:

$$E = R[t]_x \quad (3.52)$$

όπου R είναι ο πίνακας περιστροφής και $[t]_x$ είναι μία αναπαράσταση πίνακα ενός εξωτερικού γινομένου με τον πίνακα t . Επομένως, οι πίνακες R και t μπορούν να ανακτηθούν μέσω της αποσύνθεσης μοναδικής τιμής (singular value decomposition, SVD) ως εξής:

$$E = U\Sigma V^T \quad (3.53)$$

όπου U και V , ορθογώνιοι πίνακες με διαστάσεις 3×3 και Σ , διαγώνιος πίνακας, επίσης με διαστάσεις 3×3 ,

$$\Sigma = \begin{bmatrix} w & 0 & 0 \\ 0 & w & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad (3.54)$$

Τα διαγώνια στοιχεία του Σ αποτελούν τις μοναδικές τιμές του πίνακα E που σύμφωνα με τους εσωτερικούς περιορισμούς του επιπολικού πίνακα, πρέπει να είναι οι δύο τιμές ίσες και η μία μηδενική. Στην συνέχεια ορίζουμε,

$$W = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, W^{-1} = W^T = \begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (3.55)$$

και κάνοντας τη αντίστοιχη προσάρτηση των παραπάνω έχουμε:

$$[t]_x = UW\Sigma U^T \quad (3.56)$$

$$R = UW^{-1}V^T \quad (3.57)$$

Παρόλα αυτά, ο πίνακας Σ δεν ικανοποιεί πλήρως τους περιορισμούς που υπάρχουν όταν μελετώνται πραγματικά δεδομένα. Η εναλλακτική προσέγγιση που βοηθάει σε τέτοιες περιπτώσεις είναι η εξής:

$$[t]_x = UZU^T, Z = \begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad (3.58)$$

Αφού υπολογιστούν οι πίνακες περιστροφής και μετατροπής μπορούμε πλέον να λύσουμε την σχέση (3.47) για $s = I$. Το αποτέλεσμα θα είναι μια σχετική τιμή η οποία ορίζει το σημείο που βρίσκεται η κάμερα, όχι όμως σε φυσικές μονάδες (π.χ mm, cm κλπ.). Για να βρούμε την πραγματική θέση της κάμερας και κατ' επέκταση, την μετατόπισή της, θα πρέπει να κλιμακώσουμε αυτή την τιμή υπολογίζοντας τον παράγοντα κλιμάκωσης s .

Στην προτεινόμενη μεθοδολογία για να βρούμε τον παράγοντα κλιμάκωσης, υπολογίζουμε κατά την διάρκεια της ζυγοστάθμισης της κάμερας, την αντιστοιχία της σχετικής μετατόπισης ως προς την πραγματική. Έτσι εγκαθιδρύεται μία αναλογία μεταξύ της σχετικής και της πραγματικής μετατόπισης και ο λόγος αυτός είναι ο παράγοντας κλιμάκωσης s . Έτσι έχοντας όλες πλέον τις πληροφορίες που ικανοποιούν την σχέση (3.47), την λύνουμε για δύο διαφορετικά αντιστοιχισμένα σημεία διαδοχικών εικόνων/στιγμιότυπων και υπολογίζουμε την μετατόπιση της κάμερας μεταξύ αυτών.

3.2 “Έξυπνη” Μη Παραμετρική Προσέγγιση

Η κύρια καινοτομία της συγκεκριμένης εργασίας είναι η παρουσίαση μιας “έξυπνης”, μη παραμετρικής, προσαρμοζόμενης μεθοδολογίας μετρήσεων αποστάσεων, σε ενδοσκοπικά βίντεο και εικόνες και η σύγκρισή της, ως προς την κλασσική γεωμετρική προσέγγιση (3.1). Η προτεινόμενη “έξυπνη” μέθοδος, δεν απαιτεί γνώση επί των εσωτερικών παραμέτρων της κάμερας. Επιπλέον έχει ολιστικό χαρακτήρα, αφού μπορεί να εφαρμοστεί για οποιοδήποτε τύπο κάμερας, ακόμα και αν οι παράμετροι τους είναι παντελώς άγνωστοι. Επιπλέον, δεν βασίζεται στον κλασσικό τρόπο βαθμονόμησης.

Σε αντίθεση με την κλασσική γεωμετρική, παραμετρική προσέγγιση, η οποία είναι άκρως εξαρτώμενη του γεωμετρικού μοντέλου της κάμερας και απαιτεί βαθμονόμηση της κάμερας, η προτεινόμενη “έξυπνη” μέθοδος, προσαρμόζεται αυτόματα σε οποιοδήποτε μοντέλο κάμερας. Αυτό συμβαίνει διότι μαθαίνει την σχέση μεταξύ των συντεταγμένων του επιπέδου της εικόνας (image plane) και του πραγματικού κόσμου, χρησιμοποιώντας ένα τεχνητό νευρωνικό δίκτυο. Επίσης, χρησιμοποιεί την πληροφορία της “εμφάνισης” των εικονοστοιχείων (π.χ

ένταση του φωτός, χρωματική πληροφορία κλπ.) για την καλύτερη αντίληψη του βάθους. Τα βήματα αυτής της μεθοδολογίας μπορούν να συνοψιστούν ως εξής:

- i. Ας υποθέσουμε ότι I_i και I_{i+1} είναι δύο διαδοχικές εικόνες/στιγμιότυπα βίντεο. Από την κάθε μία εξάγουμε σημεία ενδιαφέροντος χρησιμοποιώντας τον αλγόριθμο SIFT.
- ii. Εφαρμόζουμε τον αλγόριθμο ανίχνευσης οπτικής ροής KLT για να βρούμε αντιστοιχίες σημείων μεταξύ των I_i και I_{i+1} .
- iii. “Φιλτράρισμα” πιθανών εσφαλμένων αντιστοιχιών σημείων (π.χ σημείων που έχουν φαινομενική ομοιότητα αλλά δεν παρουσιάζουν γεωμετρική συνοχή), χρησιμοποιώντας τον αλγόριθμο RANSAC στο σύνολο των προαναφερθέντων αντιστοιχιών του βήματος (ii).
- iv. Ως εισόδους στο νευρωνικό δίκτυο, εισάγονται, οι συντεταγμένες των σημείων που λαμβάνονται ως αποτέλεσμα από το βήμα (iii), καθώς και χαρακτηριστικά χρώματος/φωτεινότητας.
- v. Ως έξοδο το νευρωνικό δίκτυο εξάγει την μετατόπιση της κάμερας, σε φυσικές μονάδες (π.χ mm, cm κλπ), ως προς τους άξονες X (οριζόντιος ως προς το έντερο), Y (κάθετος ως προς το έντερο) και Z (παράλληλος ως προς το έντερο). Έχοντας τις μετατοπίσεις ως προς όλους του άξονες του τρισδιάστατου χώρου μπορούμε εύκολα να υπολογίσουμε την συνολική απόσταση Δd που διένυσε η κάμερας χρησιμοποιώντας τον τρισδιάστατο τύπο της ευκλείδειας απόστασης,

$$\Delta d = \sqrt{\Delta X^2 + \Delta Y^2 + \Delta Z^2} \quad (3.59)$$

Πριν από αυτή την μεθοδολογία, διερευνήσαμε και προτείναμε, μία υβριδική προσέγγιση όπου συνδιάζεται η γεωμετρικής προσέγγιση με ένα τεχνητό νευρωνικό δίκτυο. Μέσω του δικτύου, υπολογιζόταν η μετατόπιση της κάμερας επί του Z άξονα (παράλληλος της γαστρεντερικής οδού) σε χρονικό διάστημα δύο διαδοχικών εικόνων.

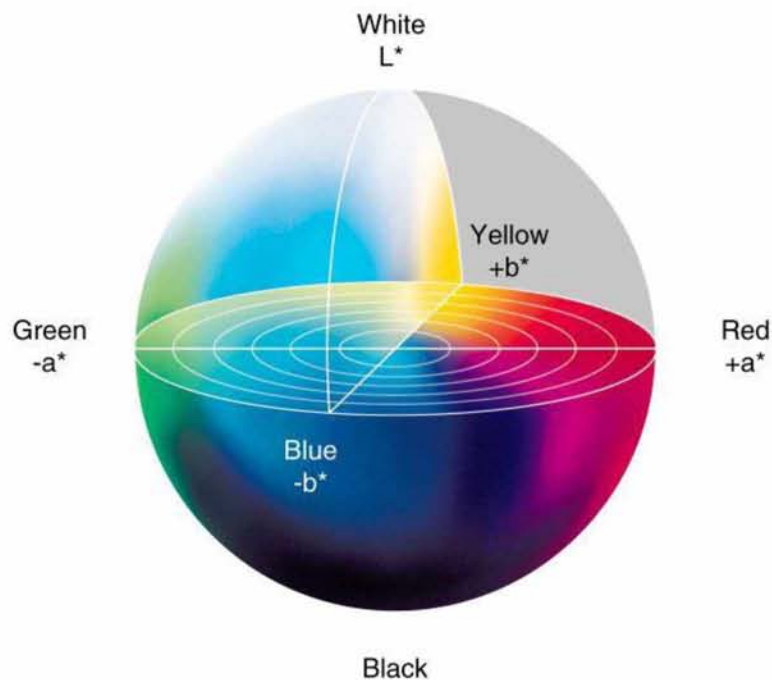
Έχοντας γνωστή την μετατόπιση της κάψουλας, στον άξονα Z , χρησιμοποιούσαμε αυτήν την πληροφορία για να υπολογίσουμε τον παράγοντα κλιμάκωσης. Με τον παράγοντα κλιμάκωσης, μπορούσαμε πλέον να κλιμακώσουμε τα αποτελέσματα της οπτικής οδομετρίας έτσι ώστε να αντιστοιχούν σε φυσικές μονάδες μέτρησης. Αυτή η μεθοδολογία αντιστοιχεί στην δημοσίευση (I), που αναφέρεται στην υποενότητα 1.3. Αυτή η μεθοδολογία δεν αναλύεται στην παρούσα εργασία, διότι λειτούργησε σαν γέφυρα για την μεθοδολογία που παρουσιάζεται στην ενότητα 3.2. Πλέον δεν χρειάζεται ο συνδιασμός των δύο μεθόδων, αφού η συγκεκριμένη αρχιτεκτονική του νευρωνικό δίκτυο μπορεί να εκτελέσει μετρήσει μέσω οπτικής πληροφορίας.

3.2.1 Διανύσματα Χαρακτηριστικών Εισόδου

Όπως προαναφέρθηκε και στην εισαγωγή αυτής της ενότητας (3.2), τα χαρακτηριστικά που δέχεται ως εισόδους το νευρωνικό δίκτυο, είναι οι συντεταγμένες των αντιστοιχιών των σημείων ενδιαφέροντος, μεταξύ δύο διαδοχικών εικόνων καθώς και η πληροφορία χρώματος/φωτεινότητας, των εικονοστοιχείων που αντιστοιχούν. Ως βασικοί είσοδοι του νευρωνικού είναι οι συντεταγμένες των σημείων, καθώς μπορούν να υπάρξουν παραλλαγές στα χαρακτηριστικά χρώματος ή φωτεινότητας. Λόγω της διαφοροποίησης της επιπλέον πληροφορίας, επηρεάζεται, όπως θα δούμε παρακάτω (υποενότητα 3.2.2), και ο αριθμός των νευρώνων εισόδου του δικτύου.

Ο λόγος που επιλέγεται να χρησιμοποιηθούν ως είσοδοι του νευρωνικού, τιμές οι οποίες χαρακτηρίζουν το χρώμα ή/και την ένταση των εικονοστοιχείων, είναι αυτού καθ' αυτού η φύση της ενδοσκοπικής εξέτασης. Το πλάνο που καταγράφεται από το όργανο που χρησιμοποιείται για την ενδοσκοπική εξέταση, φωτίζεται αποκλειστικά από την πηγή φωτός (π.χ light emission diodes, LED) που είναι εγκατεστημένη στο κάθε ενδοσκόπιο. Λόγω της ανάκλασης του φωτός στην βλεννογόνο επιφάνεια, όταν το όργανο με το οποίο εκτελείται η εξέταση της ενδοσκόπησης, βρίσκεται κοντά στην επιφάνεια, το αντανακλώμενο φως γίνεται πιο έντονο. Για να ποσοτικοποιηθεί η ένταση του φωτός, στην παρούσα εργασία χρησιμοποιείται το στοιχείο της έντασης της φωτεινότητας L του χρωματικού χώρου CIE- Lab

(εικόνα 3.8) καθώς και παραλλαγές μαζί με τα στοιχεία a και b .



Εικόνα 3.8. Γραφική αναπαράσταση του εύρος τιμών του χρωματικού χώρου CIE-*Lab*.

Ο χρωματικός χώρος που αναπαριστά ο CIE-*Lab* επιτυγχάνεται μέσω ενός μη γραμμικού μετασχηματισμού, του χρωματικού μοντέλου RGB, το οποίο χρησιμοποιείται σε μεθόδους ενδοσκόπησης όπως στην κάψουλα ασύρματης ενδοσκόπησης (Wyszecki & Stiles, 1982) (Koulaouzidis et al., 2015). Τα στοιχεία a και b του χρωματικού χώρου, αναπαριστούν την χρωματική ποσότητα του κόκκινου ($a > 0$) ή του πράσινου ($-a > 0$), την ποσότητα του κίτρινου ($b > 0$) ή του μπλε ($-b > 0$) ενός εικονοστοιχείου. Η χρήση των χρωματικών στοιχείων προσδίδει ευρωστία στον προσδιορισμό των αντιστοιχιών μεταξύ των εικόνων/στιγμιότυπων, αφού προσεγγιστικά δεν συσχετίζονται με το στοιχείο L .

3.2.2 Ανάλυση Αρχιτεκτονικής Δικτύου

Η αρχιτεκτονική του νευρωνικού δικτύου η οποία υιοθετείται σε αυτή τη μεθοδολογία, είναι αυτή του πολυστρωματικού νευρωνικού δικτύου (multiple layer perceptron, MLP). Τα πολυστρωματικά νευρωνικά δίκτυα είναι μοντέλα καθολικής προσέγγισης, δηλαδή μπορούν να προσεγγίσουν οποιαδήποτε συνεχή συνάρτηση εντός του υποσυνόλου R^n (Haykin &

Network, 2004). Αυτό σημαίνει ότι, όταν τους δοθούν οι κατάλληλες παράμετροι, μπορούν να αναπαραστήσουν (μέσω μιας διαδικασίας εκμάθησης) ένα μεγάλο αριθμό συναρτήσεων.

Η αρχιτεκτονική του συγκεκριμένου δικτύου αποτελείται από 3 επίπεδα, ένα επίπεδο εισόδου, ένα κρυφό και ένα επίπεδο εξόδου. Αυτή η αρχιτεκτονική προτιμήθηκε έναντι κάποιας άλλης με περισσότερα από ένα κρυφά επίπεδα, αφού ακόμα και μια τόσο απλή αρχιτεκτονική είναι ικανή να αποδώσει μια καθολική προσέγγιση (με αρκετά καλή ακρίβεια), όπως έχουν δείξει στην μελέτη τους οι Hornik et al. (1990).

Όπως προαναφέρθηκε και στην υποενότητα 3.2.1, το επίπεδο εισόδου του δικτύου λαμβάνει ως παραμέτρους τις συνταγμένες των αντιστοιχισμένων σημείων ενδιαφέροντος μεταξύ δύο διαδοχικών εικόνων/στιγμιότυπων I_i και I_{i+1} . Οι αντιστοιχισμένες συντεταγμένες μεταξύ των δύο εικόνων μπορούν να γραφτούν ως (x^j_i, y^j_i) και (x^j_{i+1}, y^j_{i+1}) , όπου ο j -οστός όρος αντιστοιχεί στο ζευγάρι των αντιστοιχισμένων σημείων. Ο αριθμός των νευρώνων εισόδου ορίζεται από τον ελάχιστο αριθμό αντιστοιχιών σημείων ενδιαφέροντος m , μεταξύ όλων των διαδοχικών εικόνων/στιγμιότυπων μιας ακολουθίας εικόνων/βίντεο οι οποίες χρησιμοποιούνται στην διαδικασία της εκπαίδευσης του δικτύου. Δηλαδή, τα βήματα (i) και (ii/ii,iii) επαναλαμβάνονται για όλες τις εικόνες πριν αποφασιστεί ο αριθμός των νευρώνων εισόδου. Ο αριθμός των νευρώνων εισόδου N , αποφασίζεται μέσω της σχέσης,

$$N = 6 \cdot m \quad (3.60)$$

Ο αριθμός 6 ανά ζευγάρι αντιστοιχισμένων σημείων, προκύπτει διότι, πέραν των 4 προαναφερθέντων συντεταγμένων των σημείων ανά αντιστοιχισμένων σημείων, συμπεριλαμβάνεται και η αντίστοιχη ένταση της φωτεινότητας του εικονοστοιχείου, L^j_i, L^j_{i+1} , που αντιστοιχούν οι συντεταγμένες. Εάν θέλουμε να συμπεριλάβουμε όμως και τις 3 τιμές του χρωματικού χώρου CIE-*Lab* (L, a, b), αντί μόνο της έντασης του φωτός, ο αριθμός N των νευρώνων εισόδου υπολογίζεται ως,

$$N = 10 \cdot m \quad (3.61)$$

Δεδομένου των παραπάνω μία γενική εξίσωση για τον αριθμό των νευρώνων εισόδου,

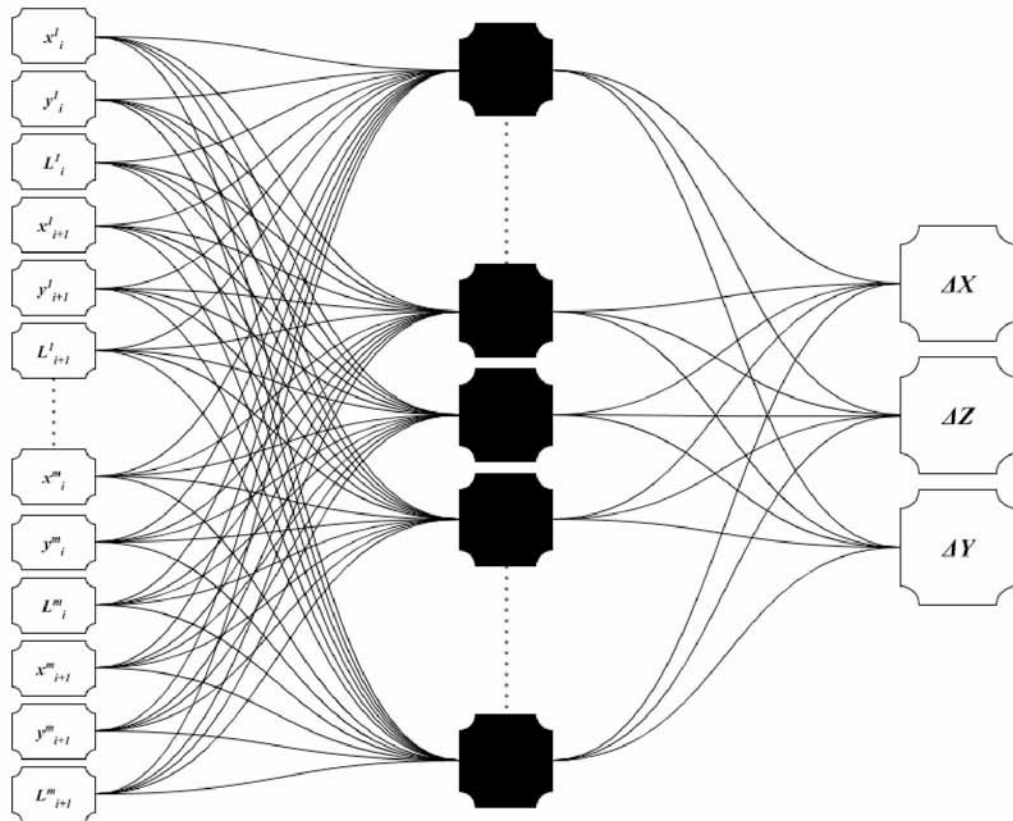
ανάλογο των χαρακτηριστικών πέραν των συντεταγμένων των σημείων, γράφεται ως εξής:

$$N = (4 + 2 \cdot l) \cdot m \quad (3.62)$$

όπου l είναι ο αριθμός των επιπλέον χαρακτηριστικών που έχουν εξαχθεί από κάθε συντεταγμένη των αντιστοιχισμένων σημείων.

Εάν, ο αριθμός των αντιστοιχισμένων σημείων δύο διαδοχικών εικόνων/στιγμιότυπων, I_i και I_{i+1} , είναι μεγαλύτερος του m , οι επιπλέον αντιστοιχίες διαιρούνται σε υποσύνολα από m αντιστοιχίες, και σχηματίζουν αντίστοιχες ομάδες N -διάστατων διανυσμάτων εισόδου. Αυτό γίνεται μέχρι το μέγιστο ακέραιο πολλαπλάσιο του m . Παραδείγματος χάριν, εάν ο αριθμός των αντιστοιχιών είναι μεταξύ 9 και 11, και ο $m = 3$, τότε θα δημιουργηθούν τρία N -διάστατα διανύσματα εισόδου και οι εναπομείναντες αντιστοιχίες θα αγνοηθούν.

Το επίπεδο εξόδου του δικτύου αποτελείται από 3 νευρώνες εξόδου και ο καθένας δίνει σαν έξοδο, υπολογισμούς της απόστασης που διανύθηκε από την κάμερα ως προς τους τρεις άξονες του καρτεσιανού συστήματος συντεταγμένων, μεταξύ δύο διαδοχικών εικόνων/στιγμιότυπων. Έτσι η έξοδος του κάθε νευρώνα μπορεί να γραφτεί ως ΔX , ΔY , ΔZ και η συνολική απόσταση που διανύθηκε από την κάμερα, Δd και υπολογίζεται από την σχέση (3.59). Στην επόμενη εικόνα (εικόνα 3.9) απεικονίζεται μια γραφική αναπαράσταση της αρχιτεκτονικής του δικτύου η οποία δέχεται σαν επιπλέον εισόδους τον συντελεστή L του CIE-Lab.



Εικόνα 3.9. Γραφική αναπαράσταση παραδείγματος της αρχιτεκτονικής του νευρωνικού δικτύου. Στο συγκεκριμένο παράδειγμα ως επιπλέον είσοδοι στο δίκτυο χρησιμοποιείται το στοιχείο L του χρωματικού χώρου CIE-*Lab*.

Το κρυφό επίπεδο του δικτύου αποτελείται από ένα σύνολο “εκπαιδευμένων” παραμέτρων, οι οποίες προκύπτουν από μία διαδικασία εκπαίδευσης η οποία συμπεριλαμβάνει ένα υποσύνολο από βίντεο όπου η κίνηση της κάμερας είναι γνωστή. Όπως έχει ήδη αναφερθεί αυτό το σύνολο των παραμέτρων είναι υπεύθυνο για την μετατροπή των αντιστοιχιών των σημείων στην πραγματική μετατόπιση που πραγματοποιεί η κάμερα μέσα στο γαστρεντερικό σύστημα. Ο αριθμός των νευρώνων του κρυφού επιπέδου του δικτύου αποφασίζεται πειραματικά.

Η λογική πίσω από την χρήση του δικτύου, που περιγράφηκε σε αυτή την ενότητα, είναι η ανάγκη εύρεσης μια συνάρτησης η οποία χαρτογραφεί την κίνηση που παρατηρείται σε δύο

δισδιάστατες διαδοχικές εικόνες/στιγμιότυπα, στην μετατόπιση, στον τρισδιάστατο χώρο, που πραγματοποιείται από την κάμερα, στο αντίστοιχο χρονικό διάστημα. Για αυτό τον λόγο, το δίκτυο εκπαιδεύεται χρησιμοποιώντας συντεταγμένες ζευγαριών αντιστοιχισμένων σημείων μεταξύ δύο διαδοχικών εικόνων στις οποίες είναι γνωστή η ακριβής μετατόπιση της κάμερας. Το σύνολο των εικόνων, στις οποίες υπάρχει βάση αληθείας (ground truth), λήφθηκαν μέσω ρομποτικού πειράματος το οποίο αναλύεται στην επόμενη ενότητα (4.1). Δεδομένης της διαδικασίας ταιριάσματος και “φιλτραρίσματος” των σημείων (βήματα ii-iii), οι αντιστοιχίες των σημείων που λαμβάνονται, κατά κύριο λόγο, θα ανήκουν στην ίδια δομή ιστού (εικόνα 3.5(α)). Επομένως, η κίνηση των σημείων θα είναι αντιπροσωπευτική της κίνησης που παρατηρείται από την κάμερα του ενδοσκοπίου. Επίσης, βασιζόμενοι στην παρατήρηση ότι ο ιστός που βρίσκεται πιο κοντά στην κάμερα, εμφανίζεται διαφορετικά σε σχέση με αυτούς που είναι μακριά (διότι είναι πιο κοντά στην πηγή του φωτός), το χρώμα των αντίστοιχων εικονοστοιχείων θεωρείται σαν επιπλέον είσοδος στο δίκτυο.

Εφόσον οι διαφορές, της εμφάνισης του ιστού, εξαρτώνται από την ένταση της φωτεινότητας στα διάφορα σημεία του, το στοιχείο της ποσοτικοποίησης της φωτεινότητας, L του χρωματικού χώρου *CIE-Lab*, προβλέπεται να επηρεάσει θετικά τις προσεγγίσεις του νευρωνικού δικτύου. Παρόλα αυτά, οι διαφορές των χρωματικών στοιχείων στην εικόνα, προκύπτουν επίσης, από την διάθλαση του φωτός, καθώς και από τις σκιάσεις που δημιουργούνται. Επιπλέον, η χρήση, ως είσοδο, των χρωματικών στοιχείων του χρωματικού χώρου *CIE-Lab*, μπορεί να επιτελέσει ως βοηθητικός παράγοντας απόρριψης εσφαλμένων αντιστοιχιών σημείων, από το ίδιο το δίκτυο, σε περίπτωση που ο συνδυασμός των KLT-RANSAC έχει αποτύχει. Παραδείγματος χάριν, τα εικονοστοιχεία που ανήκουν στην βλεννογόνο επιφάνεια του ιστού του γαστρεντερικού συστήματος, αναμένεται να έχουν κόκκινη απόχρωση, ενώ αυτά που αντιστοιχούν σε ακαθαρσίες συνήθως εμφανίζουν κίτρινη ή πράσινη απόχρωση.

4. Πειραματική Αξιολόγηση

Σε αυτήν την ενότητα αναλύεται η διάταξη του πειράματος μέσω του οποίου θεσπίστηκε βάση αληθείας, καθώς και δεδομένα για την εξέταση της απόδοσης των μεθοδολογιών, για την αξιολόγηση των προαναφερθέντων προσεγγίσεων. Επίσης αναλύονται και τα αποτελέσματα της κάθε μεθοδολογίας στις αντίστοιχες υποενότητες 4.2.1 και 4.2.2.

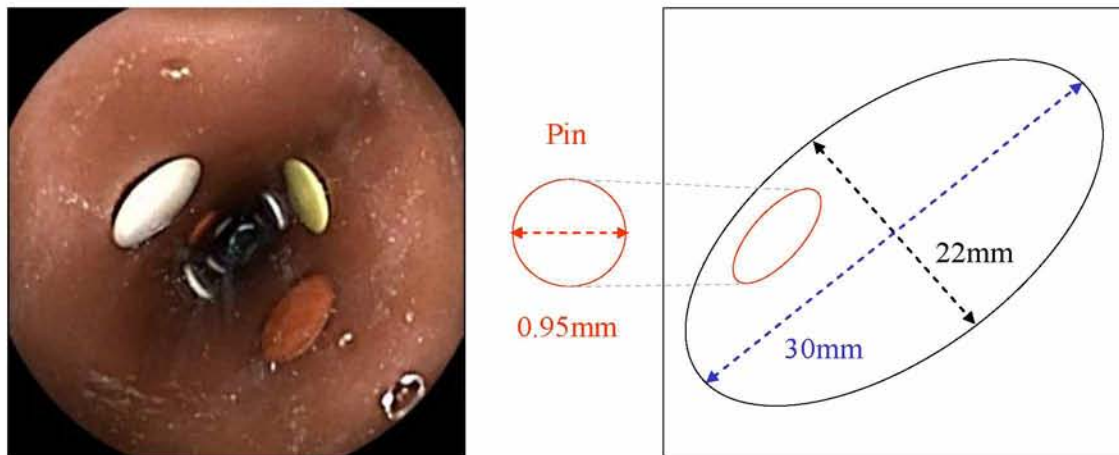
4.1 Ρομποτικό Πείραμα

Η αξιολόγηση της κάθε μεθοδολογίας, έγινε βάσει των δεδομένων του ρομποτικού πειράματος που χρησιμοποιήθηκε σε έρευνα για την εξέταση της εφαρμογής της γεωμετρικής οπτικής οδομετρίας, σε ασύρματες κάψουλες ενδοσκόπησης (*Iakovidis et al.*, 2016). Στο συγκεκριμένο πείραμα, προσομοιώνεται, *ex-vivo*, το έντερο και η κίνηση του ενδοσκοπίου. Ο λόγος που χρησιμοποιήθηκε αυτό το πειραματικό μοντέλο αντί πραγματικών δεδομένων από ενδοσκοπική εξέταση, ήταν ανάγκη για την θέσπιση βάσης αληθείας για την θέση του ενδοσκοπίου. Έτσι τέθηκε δυνατή η λήψη ακριβών μετρήσεων για την πορεία του ενδοσκοπίου εντός του τεχνητού εντέρου.

Το πείραμα αυτό αποτελείται από τέσσερα βασικά μέρη, τα οποία απεικονίζονται στην εικόνα 4.2:

- i. Μία ασύρματη κάψουλα ενδοσκόπησης Pillcam[®]SB3 (Medtronic, USA), με διαστάσεις 11.4×26.2 mm, ανάλυση εικόνας 320×320 pixel, μεταβλητό ρυθμό προβολής (frame-rate) μεταξύ 2 – 6 στιγμιότυπων το δευτερόλεπτο (frames per second, fps) και γωνία προβολής 156° (*Sliker & Ciuti*, 2014) (*G. I. Corporation*, 2014). Πρέπει να σημειωθεί, ότι η εταιρία Medtronic δεν παρήχε καμία πληροφορία σχετικά με την εστιακή απόσταση της κάμερας της κάψουλας.
- ii. Ένα τεχνητό ομοίωμα εντέρου, μήκους 30 cm (Lifelike Biotissue Inc., Ontario, Canada). Στο εσωτερικό του ομοιώματος τοποθετήθηκαν σε γνωστά σημεία, πινέζες διαφορετικών

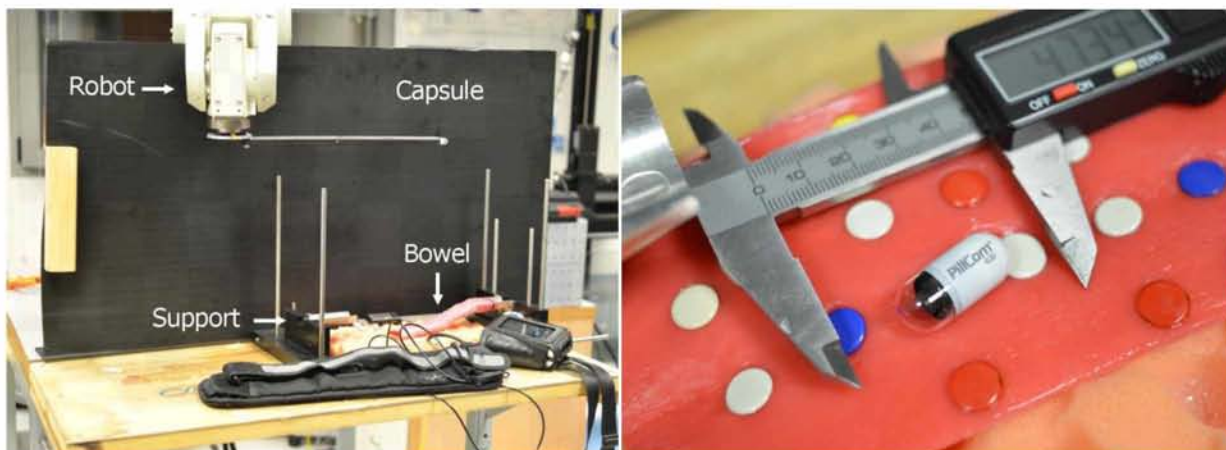
αποχρώσεων (αποχρώσεων κόκκινου, λευκού, κίτρινου και μπλε). Η διάμετρος των κεφαλών ήταν 9.5 mm και τοποθετήθηκαν έτσι ώστε να μην αλλάξει η θέση του κατά την διάρκεια του πειράματος. Πριν την έναρξη του πειράματος, καταγράφηκε η απόσταση τους από τα άκρα του τεχνητού εντέρου καθώς και η απόχρωση στον χρωματικό χώρο RGB με την μεγαλύτερη δυνατή ακρίβεια. Απεικόνιση, του εντέρου καθώς και τον τρόπο τοποθέτησής των πινεζών φαίνεται στις εικόνες 4.2 (α, β) και 4.1.



Εικόνα 4.1. Γραφική απεικόνιση του εσωτερικού του τεχνητού εντέρου, των διαστάσεων του καθώς και των διαστάσεων των κεφαλών των πινεζών.

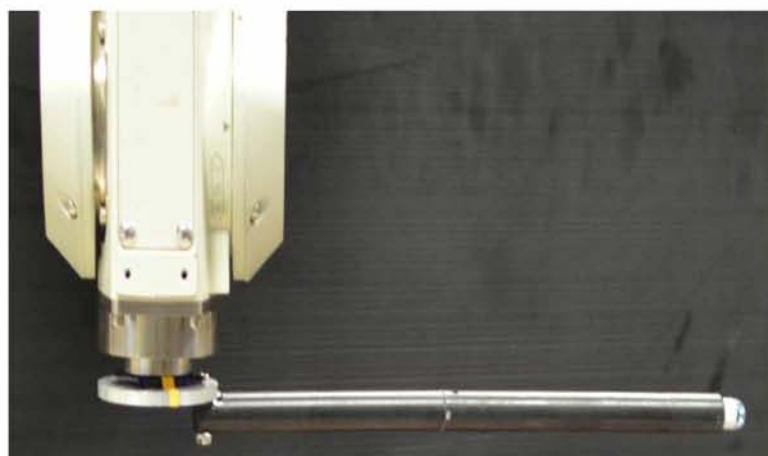
- iii. Σύστημα στήριξης για την ευθυγράμμιση καθώς και την διατήρηση της στάσης του τεχνητού εντέρου καθ' όλη την διάρκειά της διεξαγωγής του πειράματος. Στην εικόνα 4.2(α) απεικονίζεται το σύστημα στήριξης (στην εικόνα 4.2 το αριστερό άκρο του εντέρου δεν είναι τοποθετημένο στο σύστημα διότι η συγκεκριμένη φωτογραφία λήφθηκε πριν την έναρξη του πειράματος). Όταν και τα δύο άκρα του εντέρου ήταν τοποθετημένα στο σύστημα στήριξης, το έντερο ήταν τεντωμένο και σε ευθεία στάση το οποίο είχε σαν αποτέλεσμα να έχει ελλειπτικό σχήμα με διαστάσεις 30×22 mm.
- iv. Ένας υψηλής ακρίβειας, ρομποτικός βραχίονας (RV3SBrobot, Mitsubishi, Tokyo, Japan). Υποβοηθούμενο από τον ρομποτικό βραχίονα, το ενδοσκόπιο κινούταν εντός του τεχνητού εντέρου, με σταθερή ταχύτητα. Η κίνηση του βραχίονα ήταν ελεγχόμενη και είχε την δυνατότητα να κινείται είτε εμπρός είτε πίσω με διάφορες ταχύτητες και επιτάχυνση. Σε αυτή την έρευνα υποθέτουμε ότι η κύρια κίνηση της κάμερας είναι παράλληλη με το Z

άξονα και αναλόγως τοποθετήθηκε στην άκρη του βραχίονα όπως φαίνεται και στην εικόνα 4.2(γ).



(α)

(β)



(γ)

Εικόνα 4.2. Απεικόνιση του πειράματος που χρησιμοποιήθηκε για την συλλογή δεδομένων για την διεξαγωγή της έρευνας και την θέσπιση της βάσης αληθείας. Στην εικόνα (α) φαίνεται η πειραματική διάταξη πριν ξεκινήσει το πείραμα. Στην εικόνα (β) φαίνεται το εσωτερικό του ομοιώματος του εντέρου, οι πινέζες καθώς και η ασύρματη κάψουλα ενδοσκόπησης που χρησιμοποιήθηκε. Στην εικόνα (γ) φαίνεται από κοντινή απόσταση ο ρομποτικός βραχίονας.

Η συγκεκριμένη πειραματική διάταξη, χρησιμοποιήθηκε, για να εκτελεστούν δύο φορές, τριών ειδών βόλτες, με διαφορετική ταχύτητα κάθε φορά. Συγκεκριμένα, οι ταχύτητες που χρησιμοποιήθηκαν ήταν 0.5, 1.0 και 2.0 mm/s. Έτσι, συλλέχθηκαν 364 στιγμιότυπα βίντεο (12

βίντεο) τα οποία λήφθηκαν από την ασύρματη κάψουλα ενδοσκόπησης (i) κατά την κίνησή της εντός του εντέρου. Εφόσον η ταχύτητα με την οποία ταξίδευε το ενδοσκόπιο ήταν γνωστή, τα στιγμιότυπα έφεραν τον χρόνο τον οποίο λήφθηκαν, η βάση αληθείας σχηματίστηκε βάσει της γνωστής εξίσωσης:

$$\Delta x = v \cdot \Delta t \quad (4.1)$$

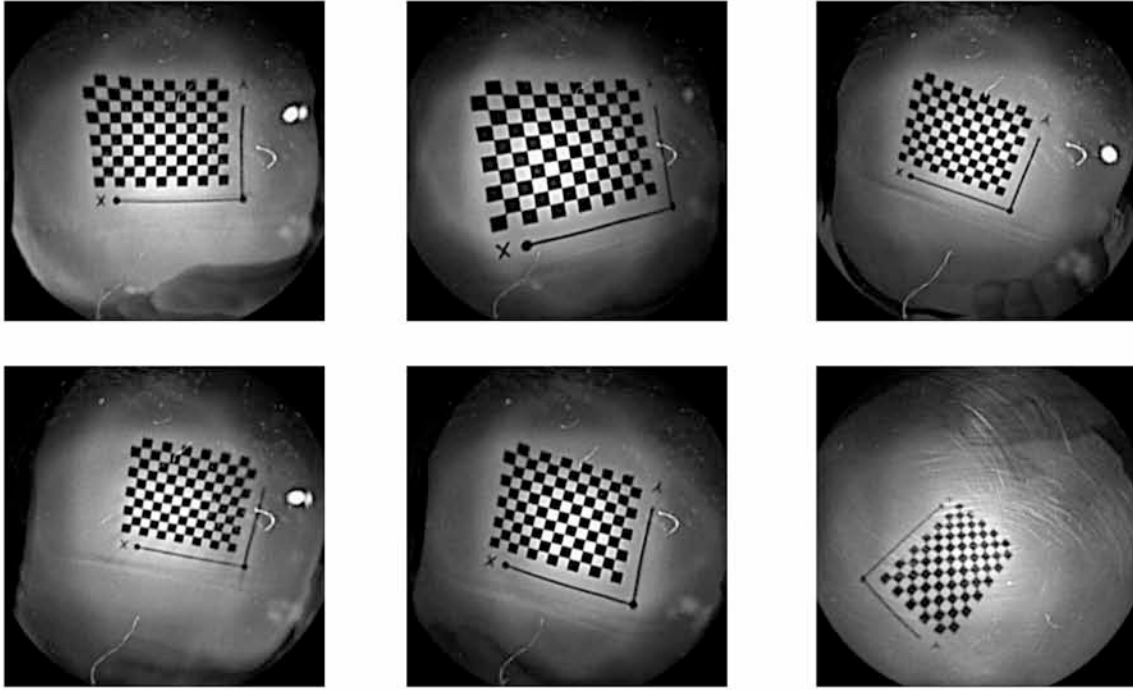
όπου, Δx σημειώνεται η απόσταση που διανύθηκε από το ενδοσκόπιο σε διάστημα δύο διαδοχικών στιγμιότυπων, των οποίων ο χρόνος λήψης τους διαφέρει κατά Δt , θεωρώντας ότι η ταχύτητα, v του ενδοσκοπίου, είναι γνωστή και σταθερή.

4.2 Αποτελέσματα

Σε αυτή την ενότητα αναλύονται και συγκρίνονται τα αποτελέσματα των δύο μεθοδολογιών, της γεωμετρικής και “έξυπνης” οπτικής οδομετρίας.

4.2.1 Γεωμετρική Οπτική Οδομετρία

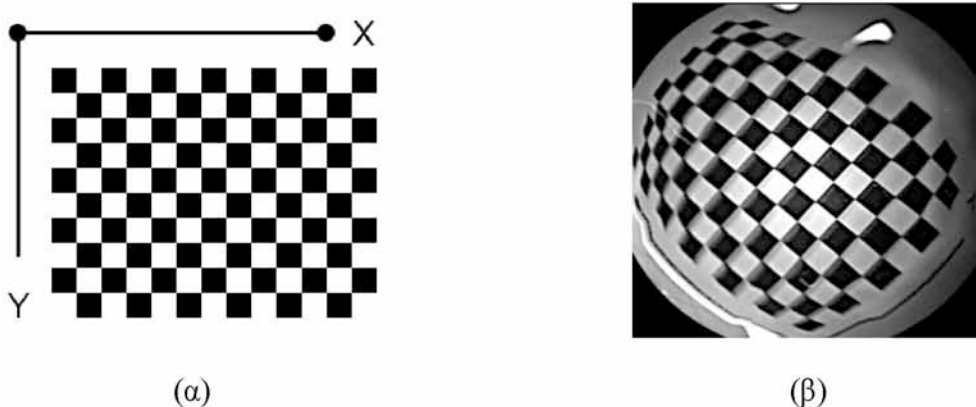
Για την βαθμονόμηση της κάμερας του ενδοσκοπίου εφαρμόστηκε η μεθοδολογία του Zhang (1999), όπως είναι υλοποιημένη από τον Bouget, σε περιβάλλον Matlab (Bouget, 2004). Η διαδικασία βαθμονόμησης που προτείνει ο Zhang, χρησιμοποιεί ένα επίπεδο πρότυπο, τύπου σκακιάρας (εικόνα 4.4 (α)). Για την χρήση του επιπέδου προέκυψε η ανάγκη τροποποίησης του μεγέθους του, λόγω του μικρού μεγέθους της ασύρματης κάψουλας ενδοσκόπησης. Πιο συγκεκριμένα, το μοτίβο που φαίνεται στην εικόνα 4.4 (α), αποτελείται από 10×3 τετράγωνα που το καθένα έχει διαστάσεις 3×3 mm. Για τον ακριβέστερο υπολογισμό των εσωτερικών παραμέτρων της κάμερας, λήφθηκαν 17 εικόνες του επιπέδου της σκακιάρας, από διάφορες οπτικές γωνίες (εικόνα 4.3), κρατώντας σταθερό το επίπεδο και αλλάζοντας θέση της κάψουλας.



Εικόνα 4.3. Εικόνες του επιπέδου με μοτίβο σκακιέρας, όπως έχουν ληφθεί από την κάμερα του ενδοσκοπίου από διάφορες οπτικές γωνίες. Οι εικόνες αυτές χρησιμοποιήθηκαν για την βαθμονόμηση της κάμερας.

Οι τιμές των εσωτερικών παραμέτρων της κάψουλας που προέκυψαν από την εκτέλεση της βαθμονόμησης ήταν για την εστιακή απόσταση (σε μονάδες pixel) $(f_x, f_y) = (148, 146)$ και για το πρωτεύον σημείο $(c_x, c_y) = (146, 150)$. Το αντίστοιχο σφάλμα επαναπροβολής των ακτίνων, ήταν της τάξης 10^{-1} . Με τον υπολογισμό των εσωτερικών παραμέτρων της κάμερας, ακολουθήθηκε η μεθοδολογία της γεωμετρικής οπτικής οδομετρίας όπως περιγράφεται στην ενότητα 3.1.

Οι αποστάσεις που υπολογίστηκαν χρησιμοποιώντας την προσέγγιση αυτή φαίνονται στον πίνακα 4.7. Το μέσο απόλυτο σφάλμα (mean absolute error, MAE) που σημειώθηκε ήταν 7.12 ± 4.01 cm. Οι αντίστοιχες πορείες του ενδοσκοπίου, που υπολογίστηκαν μέσω των μετατοπίσεων που σημειώθηκαν από τη γεωμετρική προσέγγιση σε 4 σετ δεδομένων επαλήθευσης (test datasets) απεικονίζονται στην εικόνα 4.7. Μπορεί εύκολα να παρατηρηθεί ότι οι υπολογισμένες πορείες σημειώνουν μεγάλο σφάλμα (της τάξης 10) ως προς τους άξονες X και Y .



Εικόνα 4.4. (α) Επίπεδο με μοτίβο σκακιέρας που χρησιμοποιήθηκε για την βαθμονόμηση της κάμερας της ασύρματης κάψουλας ενδοσκόπησης. (β) Το επίπεδο της εικόνας όπως λήφθηκε από την κάμερα του ενδοσκοπίου για τις ανάγκες της διαδικασίας της βαθμονόμησης.

4.2.2 “Εξυπνη” Οπτική Οδομετρία

Η πειραματική επαλήθευση της “έξυπνης” οπτικής οδομετρίας πραγματοποιήθηκε χρησιμοποιώντας, συνολικά, 12 βίντεο τα οποία λήφθηκαν μέσω του πειράματος. Από τα 12 βίντεο, τα 8 χρησιμοποιήθηκαν για την εκπαίδευση του νευρωνικού δικτύου ακολουθώντας την μέθοδο της διασταυρωμένης επικύρωσης (cross-validation). Με αυτόν τον τρόπο μπορούν να βρεθούν οι βέλτιστες τιμές βαρών του μοντέλου του νευρωνικού δικτύου για την εφαρμογή της οπτικής οδομετρίας. Τα εναπομείναντα 4 βίντεο, ήταν τα ίδια στα οποία εξετάστηκε η γεωμετρική οπτική οδομετρία, και χρησιμοποιήθηκαν για περαιτέρω εξέταση του δικτύου και την σύγκρισή του με την γεωμετρική προσέγγιση.

Στα 8 βίντεο που χρησιμοποιήθηκαν για τις ανάγκες της εκπαίδευσης, ο ρομποτικός βραχίονας κινούσε το ενδοσκόπιο εμπρός και πίσω, με σταθερές ταχύτητα 0.5 και 1.0mm/s. Στα υπόλοιπα 4 βίντεο που χρησιμοποιήθηκαν, το ενδοσκόπιο κινούταν με σταθερή ταχύτητα 2.0 mm/s ακολουθώντας πίσω και εμπρός κίνηση. Από το κάθε βίντεο, έγινε εξαγωγή αντιστοιχιών σημείων ενδιαφέροντος μεταξύ διαδοχικών στιγμιότυπων ακολουθώντας την μέθοδο που περιγράφεται στην υπενότητα 3.1.2. Πέραν των συντεταγμένων των σημείων ως είσοδοι στο δίκτυο συμπεριλήφθηκαν η ένταση του φωτός (L) καθώς και τα χρωματικά στοιχεία (a , b) του χρωματικού χώρου CIE-Lab. Πρέπει να σημειωθεί ότι οι αντιστοιχίες που

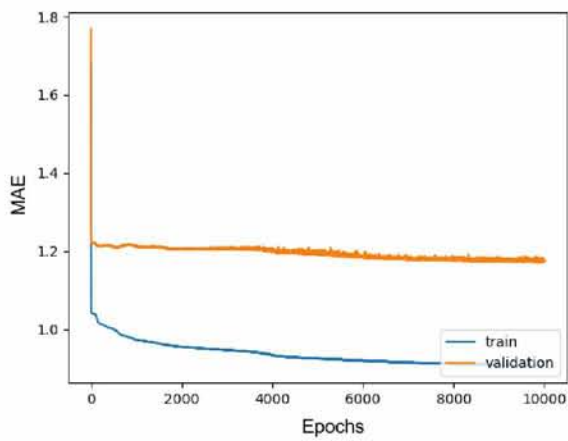
προέκυπταν στις πινέξες, δεδομένου ότι το χρώμα των κεφαλών των πινεζών ήταν γνωστό και διάφορο της επιφάνειας του εντέρου, απορρίπτονταν. Κάθε ένα από τα επιπλέον στοιχεία εξετάστηκαν ξεχωριστά χρησιμοποιώντας διαφορετικές αρχιτεκτονικές δικτύου (διαφορετικό αριθμό νευρώνων επιπέδου εισόδου).

Για την εκπαίδευση του δικτύου, χρησιμοποιήθηκε μέθοδος 5-φορών διασταυρωμένης επικύρωσης (5-fold cross validation). Δηλαδή, το σύνολο των δεδομένων “έσπασε” σε 5 όμοιες, μη επικαλυπτόμενες ομάδες δειγμάτων. Ακολουθώντας την αρχή του Pareto (επίσης γνωστή ως νόμος 80-20), κάθε ομάδα δειγμάτων χωρίστηκε σε υποσύνολα εκπαίδευσης (training) και επαλήθευσης (validation) με αναλογία 80/20 αντιστοίχως. Για την ανανέωση των βαρών του δικτύου εφαρμόστηκε η μέθοδος εκπαίδευσης στοιβάδας. Επίσης στη διαδικασία της εκπαίδευσης εφαρμόστηκε η μέθοδος πρόωρου τερματισμού, βασική τεχνική για την γενίκευση της απόδοσης του δικτύου.

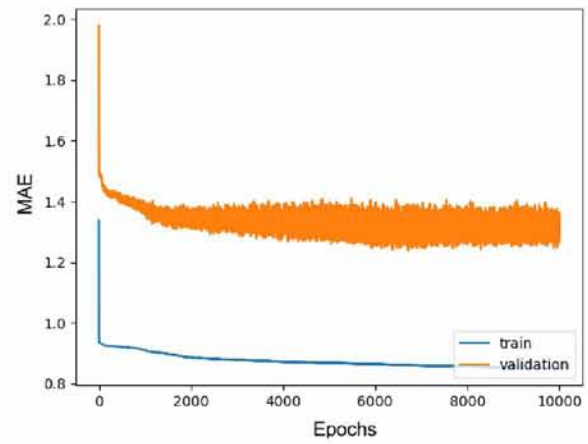
Ο αριθμός των νευρώνων εισόδου του δικτύου, για τα διάφορα σύνολα χαρακτηριστικών (π.χ $[x, y, L]$, $[x, y, L, a, b]$ κλπ.), αποφασίστηκε βάσει της εξίσωσης (3.62). Οι ελάχιστες αντιστοιχίες σημείων-ενδιαφέροντος m μεταξύ όλων των διαδοχικών στιγμιότυπων των δεδομένων εκπαίδευσης, ήταν $m = 3$. Όταν οι τιμές του χρωματικού χώρου CIE-*Lab* θεωρούνται επιπλέον είσοδοι του νευρωνικού μαζί με τις συντεταγμένες των αντιστοιχισμένων σημείων, η μεταβλητή l ισούται με $l = 3$ και το αποτέλεσμα της εξίσωσης (3.62) είναι $N = 30$. Άρα ο αριθμός των νευρώνων εισόδου γίνεται ίσος με 30. Χρησιμοποιώντας μόνο την ένταση του φωτός ως επιπλέον είσοδο στο νευρωνικό ο αριθμός l γίνεται ίσος με $l = 1$ και μέσω της εξίσωσης (3.62) ο αριθμός νευρώνων εισόδου διαμορφώνεται στους $N = 18$. Η χρήση αποκλειστικά και μόνο, των συντεταγμένων των σημείων ενδιαφέροντος έχει ως αποτέλεσμα ο αριθμός νευρώνων εισόδου του δικτύου να είναι $N = 12$, αφού $l = 0$.

Για τον αριθμό των νευρώνων στο κρυφό δίκτυο, εξετάστηκαν διάφορα σενάρια που χρησιμοποιούνται από 3 έως 1,500 κρυφοί νευρώνες. Ο αλγόριθμος που χρησιμοποιήθηκε για την εκπαίδευση του δικτύου ήταν αυτός της αναστροφής διάδοσης σφάλματος (error back propagation) και τα βάρη ανανεώνονταν με ρυθμό στοιβάδας διανυσμάτων εισόδου (batch

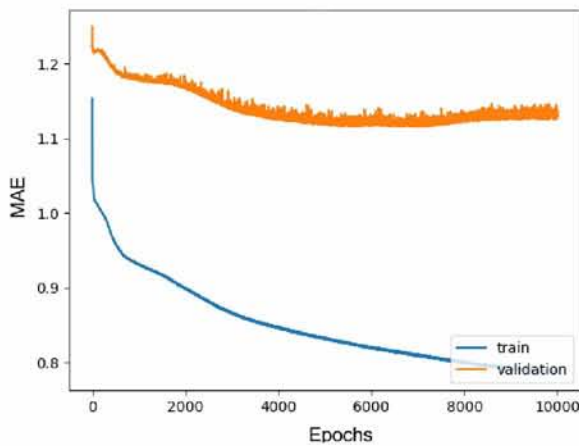
training) και όχι ανά διάνυσμα εισόδου (in-line training) και με ορμή (momentum). Πριν την έναρξη της εκπαίδευσης, μία εμπειρική έρευνα έδειξε ότι τιμές για τον ρυθμό μάθησης (learning rate) ίσο με 0.1, σταθερά ορμής ίση με 0.9 και μέγιστο αριθμό εποχών εκπαίδευσης (epochs) ίσο με 6,000 (εικόνα 4.5) ήταν επαρκείς για να συγκλίνουν όλες οι αρχιτεκτονικές δικτύων, σε ολικό ελάχιστο του σφάλματος εξόδου



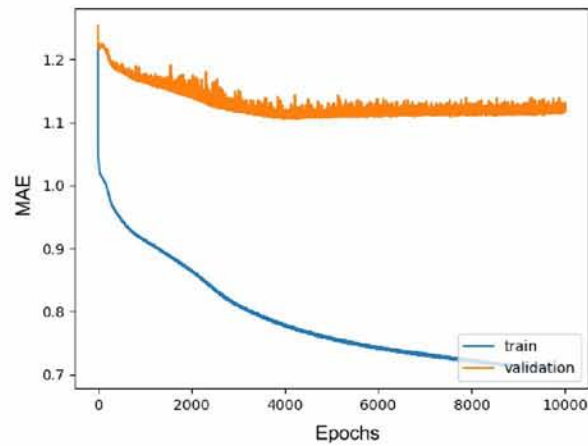
(α)



(β)

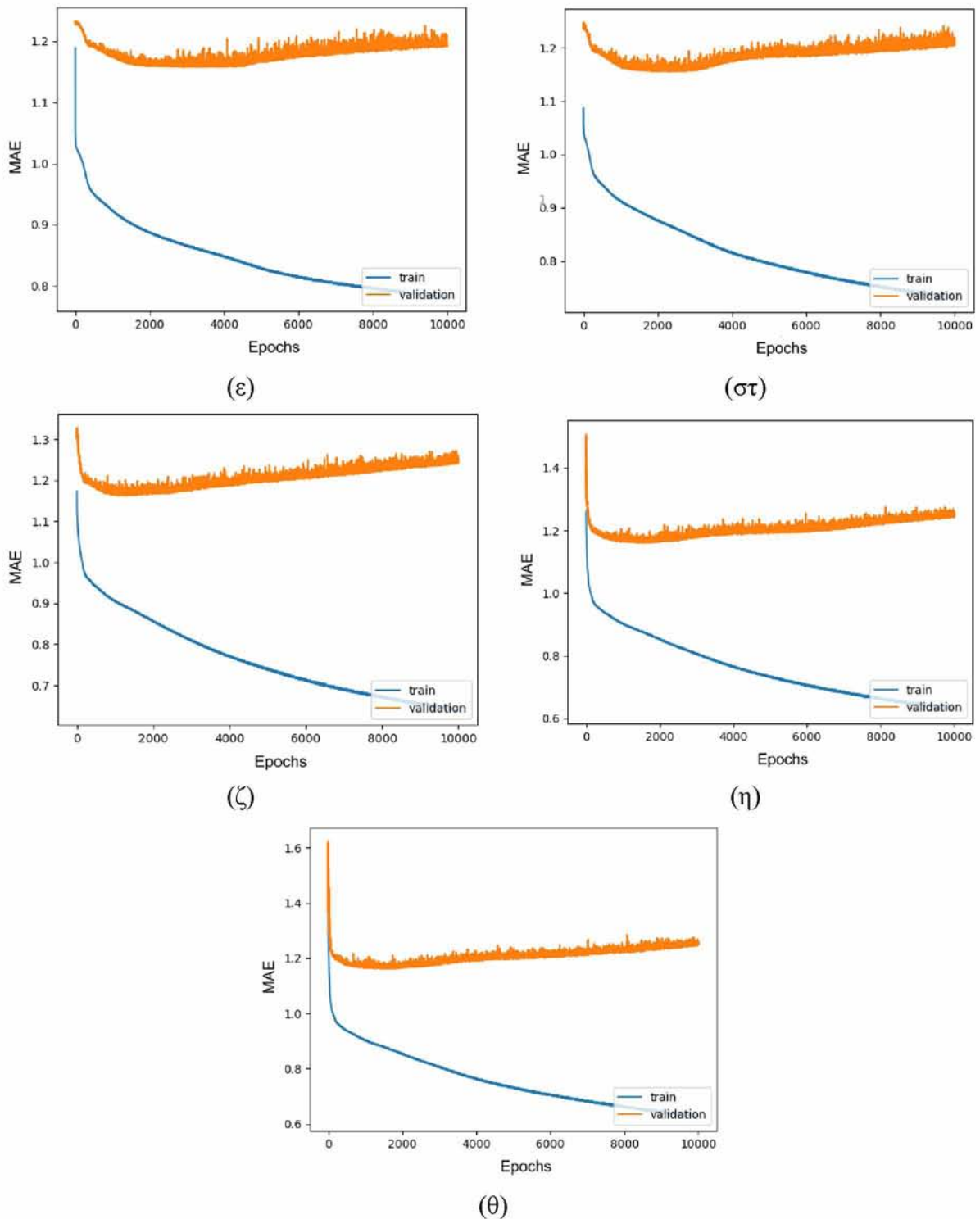


(γ)



(δ)

Εικόνα 4.5. Διαγράμματα εκπαίδευσης των νευρωνικών δικτύων διαφορετικών αρχιτεκτονικών (Συνεχίζεται στην επόμενη σελίδα).



Εικόνα 4.5. (Συνέχεια) Όπως φαίνεται από τα παραπάνω διαγράμματα, 6000 εποχές εκπαίδευσης είναι ικανές για να βρεθεί ολικό ελάχιστο.

Στους πίνακες 4.1 – 4.6 αναγράφονται, συνοπτικά, τα αποτελέσματα διάφορων

αρχιτεκτονικών νευρωνικών δικτύων, χρησιμοποιώντας για την αντιστοιχία των σημείων ενδιαφέροντος, είτε μόνο τον RANSAC ή τον KLT ή τον συνδυασμό των δύο. Επίσης, για τον κάθε συνδυασμό που αναφέρθηκε, εξετάζεται η χρήση διαφορετικών, επιπλέον χαρακτηριστικών. Τα σφάλματα που αναφέρονται είναι μέσα απόλυτα σφάλματα. Το μέσο απόλυτο σφάλμα υπολογίζεται βάσει της συνολικής απόστασης που διενυσε η κάμμουλα, όπως αυτή υπολογίστηκε μέσω του νευρωνικού δοκτύου από τα δεδομένα αξιολόγησης (test set), έναντι της βάσης αληθείας. Γραφική απεικόνιση των αποτελεσμάτων για τις διάφορες αρχιτεκτονικές δικτύων που εξετάστηκαν, φαίνεται στην εικόνα 4.6. Το εύρος των μέσων απολύτων σφαλμάτων που προέκυψαν ήταν μεταξύ 3.72 ± 2.19 και 11.43 ± 1.18 cm.

Πίνακας 4.1. Συνοπτική παρουσίαση αποτελεσμάτων, που λήφθηκαν χρησιμοποιώντας αρχιτεκτονικές με διάφορους αριθμούς κρυφών νευρώνων. Για την αντιστοιχία των σημείων έχει χρησιμοποιηθεί το σχήμα KLT-RANSAC, και σαν είσοδοι στο δίκτυο, πέραν των συντεταγμένων σημείων, όλες οι τιμές του χρωματικού χώρου CIE-Lab.

Αρχιτεκτονική Δικτύων	Αποτελέσματα Στο TestSet (cm)
30-3-3	4.32±2.93
30-5-3	3.72±2.19
30-10-3	3.98±2.75
30-20-3	4.10±2.62
30-50-3	4.52±2.30
30-100-3	4.22±2.09
30-500-3	4.54±2.28
30-1000-3	4.44±2.85
30-1500-3	4.54±2.86

Πίνακας 4.2. Συνοπτική παρουσίαση αποτελεσμάτων, που λήφθηκαν χρησιμοποιώντας αρχιτεκτονικές με διάφορους αριθμούς κρυφών νευρώνων. Για την αντιστοιχία των σημείων έχει χρησιμοποιηθεί ο αλγόριθμος RANSAC, και σαν είσοδοι στο δίκτυο, πέραν των συντεταγμένων σημείων, όλες οι τιμές του χρωματικού χώρου *CIE-Lab*.

Αρχιτεκτονική Δικτύων	Αποτελέσματα Στο TestSet (cm)
30-3-3	7.94±1.58
30-5-3	7.47±1.39
30-10-3	7.19±1.70
30-20-3	6.39±2.40
30-50-3	6.30±2.56
30-100-3	6.77±2.28
30-500-3	6.76±2.07
30-1000-3	6.16±1.68
30-1500-3	6.31±2.37

Πίνακας 4.3. Συνοπτική παρουσίαση αποτελεσμάτων, που λήφθηκαν χρησιμοποιώντας αρχιτεκτονικές μεδιάφορους αριθμούς κρυφών νευρώνων. Για την αντιστοιχία των σημείων έχει χρησιμοποιηθεί το σχήμα KLT-RANSAC, και σαν είσοδοι στο δίκτυο, πέραν των συντεταγμένων σημείων, η τιμή L χρωματικού χώρου CIE-*Lab*.

Αρχιτεκτονική Δικτύων	Αποτελέσματα Στο TestSet (cm)
18-3-3	10.99±3.43
18-5-3	8.43±3.25
18-10-3	5.88±3.90
18-20-3	6.50±3.39
18-50-3	6.86±3.27
18-100-3	5.16±1.87
18-500-3	6.23±2.63
18-1000-3	6.19±2.36
18-1500-3	6.01±2.38

Πίνακας 4.4. Συνοπτική παρουσίαση αποτελεσμάτων, που λήφθηκαν χρησιμοποιώντας αρχιτεκτονικές μεδιάφορους αριθμούς κρυφών νευρώνων. Για την αντιστοιχία των σημείων έχει χρησιμοποιηθεί ο αλγόριθμος RANSAC, και σαν είσοδοι στο δίκτυο, πέραν των συντεταγμένων σημείων, η τιμή L χρωματικού χώρου CIE-*Lab*.

Αρχιτεκτονική Δικτύων	Αποτελέσματα Στο TestSet (cm)
18-3-3	10.05±2.01
18-5-3	10.04±2.01
18-10-3	9.91±1.96
18-20-3	9.12±2.51
18-50-3	10.27±1.43
18-100-3	9.66±2.05
18-500-3	7.54±1.33
18-1000-3	6.97±1.17
18-1500-3	8.52±2.15

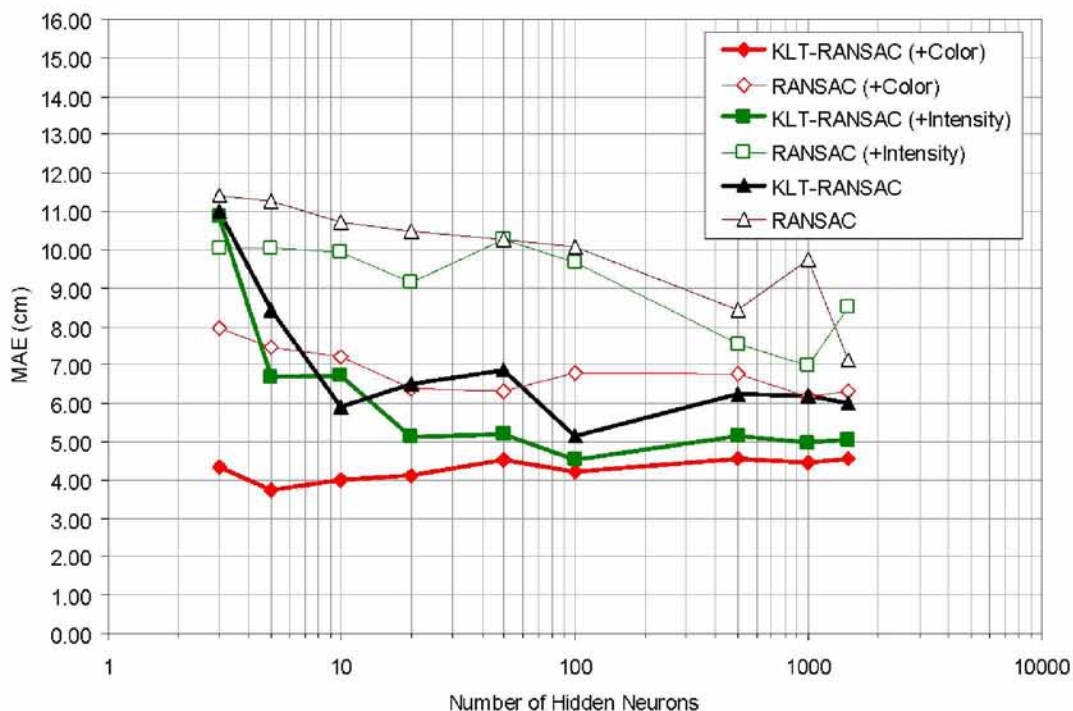
Πίνακας 4.5. Συνοπτική παρουσίαση αποτελεσμάτων, που λήφθηκαν χρησιμοποιώντας αρχιτεκτονικές μεδιάφορους αριθμούς κρυφών νευρώνων. Για την αντιστοιχία των σημείων έχει χρησιμοποιηθεί το σχήμα KLT-RANSAC, και σαν είσοδοι στο δίκτυο μόνο οι συντεταγμένες των αντιστοιχισμένων σημείων.

Αρχιτεκτονική Δικτύων	Αποτελέσματα Στο TestSet (cm)
12-3-3	10.87±3.41
12-5-3	6.66±3.08
12-10-3	6.73±4.90
12-20-3	5.12±3.03
12-50-3	5.18±2.39
12-100-3	4.52±2.78
12-500-3	5.13±3.20
12-1000-3	4.96±2.65
12-1500-3	5.03±2.76

Πίνακας 4.6. Συνοπτική παρουσίαση αποτελεσμάτων, που λήφθηκαν χρησιμοποιώντας αρχιτεκτονικές μεδιάφορους αριθμούς κρυφών νευρώνων. Για την αντιστοιχία των σημείων έχει χρησιμοποιηθεί ο αλγόριθμος RANSAC, και σαν είσοδοι στο δίκτυο οι συντεταγμένες των αντιστοιχισμένων σημείων.

Αρχιτεκτονική Δικτύων	Αποτελέσματα Στο TestSet (cm)
12-3-3	11.43±1.18
12-5-3	11.27±1.14
12-10-3	10.71±1.47
12-20-3	10.47±1.69
12-50-3	10.27±1.73
12-100-3	10.06±2.06
12-500-3	8.44±2.93
12-1000-3	9.75±1.39
12-1500-3	7.14±1.57

Γενικά, τα επίπεδα των σφαλμάτων είναι μικρότερα όταν χρησιμοποιείται ως επιπλέον είσοδος στο δίκτυο η χρωματική πληροφορία. Επίσης, ο συνδυασμός των αλγορίθμων KLT-RANSAC δίνει καλύτερα αποτελέσματα επί του υπολογισμού της θέσης του ενδοσκοπίου σε σχέση με όταν χρησιμοποιείται μόνο ο RANSAC, ανεξαρτήτως των χαρακτηριστικών που χρησιμοποιούνται. Η εικόνα 4.6 ξεκαθαρίζει ότι το δίκτυο αποδίδει καλύτερα όταν χρησιμοποιούνται, ως είσοδοι, χρωματικά χαρακτηριστικά, που εξάγονται μέσω του συνδυασμού KLT-RANSAC. Έτσι, χρησιμοποιώντας το σχήμα KLT-RANSAC και χρωματικά χαρακτηριστικά, το σφάλμα είναι μικρό και η απόδοση του δικτύου σταθερή. Τα καλύτερα αποτελέσματα που σημειώνονται είναι 3.72 ± 2.19 cm χρησιμοποιώντας αρχιτεκτονική με μόνο 5 κρυφούς νευρώνες.



Εικόνα 4.6. Γραφική αναπαράσταση της απόδοσης εύρεσης της θέσης του ενδοσκοπίου σε ως προς τον αριθμό των κρυφών νευρώνων του δικτύου και των διάφορων χαρακτηριστικών και μεθόδων εξαγωγής τους

Παρόλα αυτά, τα αποτελέσματα της αρχιτεκτονικής με 5 κρυφούς νευρώνες, όπου για την εξαγωγή των χαρακτηριστικών, χρησιμοποιείται το σχήμα KLT-RANSAC μπορεί να θεωρηθούν συγκρίσιμα με την αρχιτεκτονική των 3 κρυφών νευρώνων (4.32 ± 2.93 cm). Ωστόσο, ως βέλτιστη επιλογή θεωρούμε την αρχιτεκτονική με τους 5 (30-5-3) νευρώνες επειδή είναι αυτή που δίνει το χαμηλότερο σφάλμα έχοντας την μικρότερη πολυπλοκότητα. Επίσης, μπορεί να παρατηρηθεί ότι όταν χρησιμοποιείται ο συνδυασμός KLT-RANSAC και ο αριθμός των κρυφών νευρώνων αυξάνεται στους 100, τα αποτελέσματα όλων των αρχιτεκτονικών γίνονται συγκρίσιμα για όλα τα εξεταζόμενα χαρακτηριστικά. Η επίδραση που έχουν, η ένταση της φωτεινότητας και τα χρωματικά στοιχεία, επί των υπολογισμών του δικτύου για την θέση του ενδοσκοπίου, είναι θετική ακόμα και για μεγάλο αριθμό κρυφών νευρώνων. Ωστόσο, λαμβάνοντας υπ' όψιν το μέγεθος των σφαλμάτων, η θετική επιρροή που έχει, στην απόδοση του δικτύου, η χρήση επιπλέον χαρακτηριστικών, γίνεται ασήμαντη όσο

περισσότερο αυξάνεται ο αριθμός των νευρώνων στο κρυφό επίπεδο.

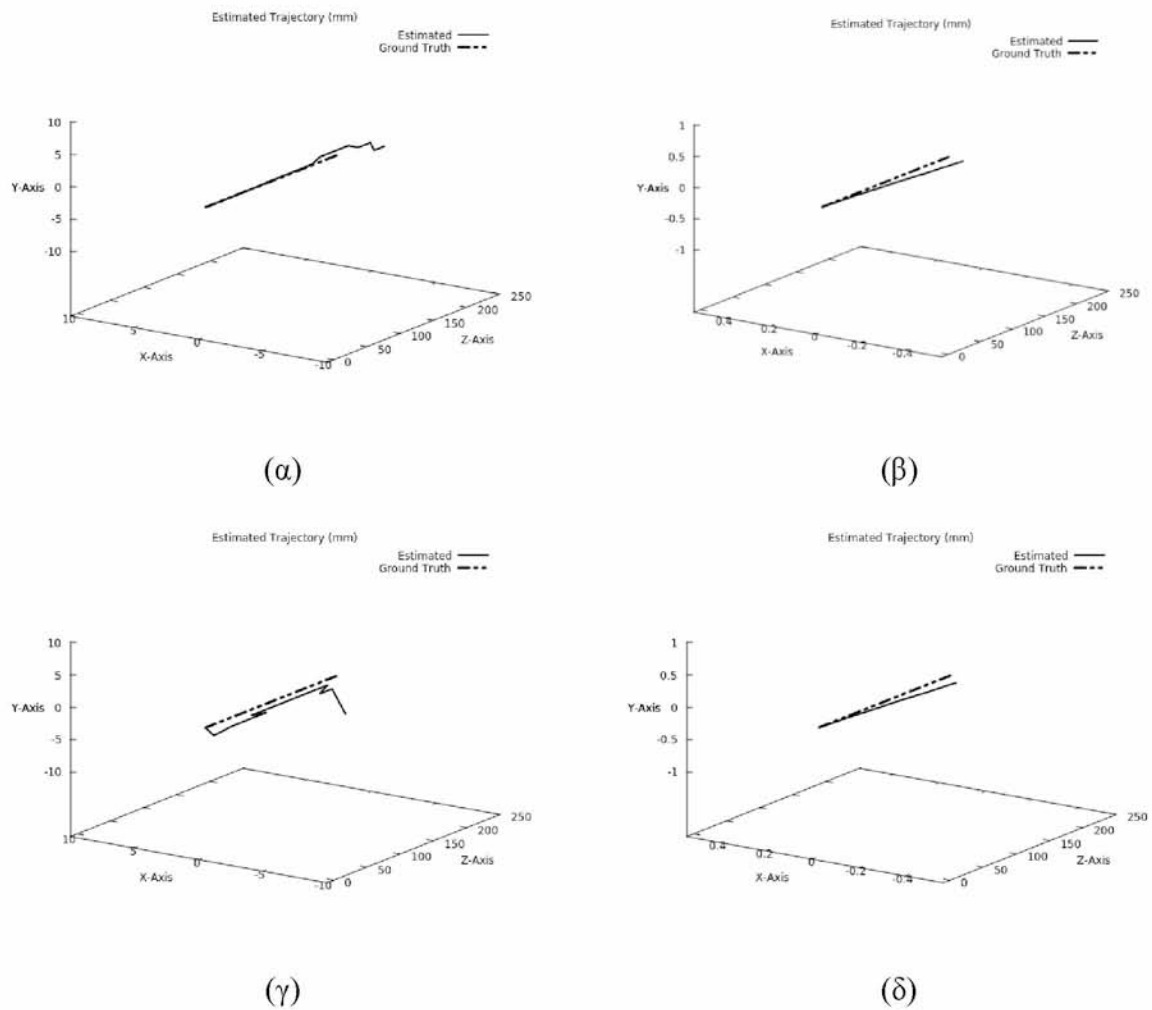
Στην περίπτωση, που για την αντιστοιχία των σημείων ενδιαφέροντος, χρησιμοποιείται μόνο ο αλγόριθμος RANSAC, αντί του σχήματος KLT-RANSAC, η χρήση χρωματικής πληροφορίας ευνοεί τα αποτελέσματα του δικτύου. Αυτό μπορεί να παρατηρηθεί και γραφικά στην εικόνα 4.6. Τα καλύτερα αποτελέσματα που σημειώθηκαν χρησιμοποιώντας τον RANSAC, επιτυγχάνονται χρησιμοποιώντας αρκετά μεγάλο αριθμό κρυφών νευρώνων ($\geq 1,000$) στην αρχιτεκτονική του νευρωνικού δικτύου.

Πίνακας 4.7. Συγκριτικός πίνακας αποτελεσμάτων μεταξύ της μεθόδου του νευρωνικού δικτύου (αρχιτεκτονική 30-5-3, σχήμα KLT-RANSAC) και της γεωμετρικής προσέγγισης. Και οι δύο μέθοδοι έχουν αξιολογηθεί στο ίδιο σύνολο δεδομένων (4 βίντεο αξιολόγησης).

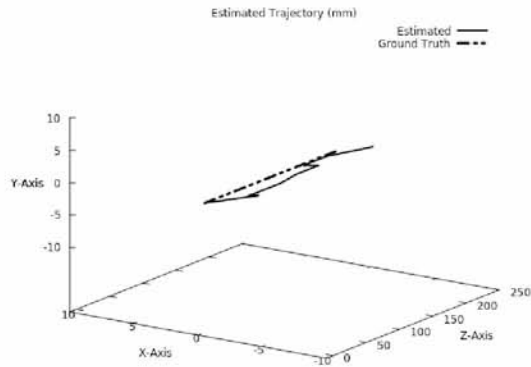
Σύνολα Δεδομένων	Απόλυτα Σφάλματα (cm)	
	<u>Νευρωνικό Δίκτυο (CIE-Lab)</u>	<u>Γεωμετρική Προσέγγιση</u>
Test 1 (εμπρόσθια κίνηση)	2.08	6.32
Test 2 (οπίσθια κίνηση)	0.98	6.22
Test 3 (εμπρόσθια κίνηση)	2.91	12.74
Test 4 (οπίσθια κίνηση)	4.83	3.2
Μέσο Απόλυτο Σφάλμα	2.70±1.62	7.12±4.01

Ακολουθώντας τον πειραματικό κύκλο της διασταυρωμένης επικύρωσης, η αρχιτεκτονική που απέδωσε καλύτερα αποτελέσματα, εξετάστηκε πάνω στα 4 βίντεο, που όπως προαναφέρθηκε, δεν συμπεριλήφθηκαν στο σύνολο των δεδομένων της εκπαίδευσης. Το ελάχιστο σφάλμα που σημειώθηκε από την βέλτιστη αρχιτεκτονική (αυτή που παίρνει ως εισόδους τις συντεταγμένες σημείων και τα στοιχεία του χρωματικού χώρου CIE-Lab), πάνω στα 4 βίντεο ήταν μόλις 2.70 ± 1.62 cm. Το αντίστοιχο ελάχιστο σφάλμα που αποδόθηκε χρησιμοποιώντας την γεωμετρική μέθοδο, ήταν 7.12 ± 4.01 cm (πίνακας 4.7). Οι αντίστοιχες γραφικές αναπαραστάσεις των πορειών του ενδοσκοπίου που ανακατασκευάστηκαν

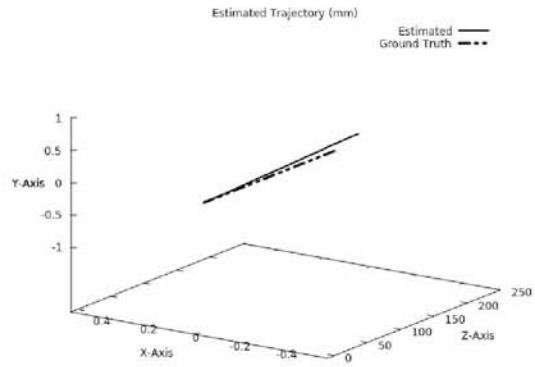
χρησιμοποιώντας τις δύο μεθόδους φαίνονται στην εικόνα 4.5. Μπορεί να παρατηρηθεί ότι χρησιμοποιώντας την “έξυπνη” προσέγγιση, το σφάλμα επί των αξόνων X και Y είναι εξαιρετικά μικρό (της τάξεως του 10^{-4}). Το σφάλμα που προκύπτει από τον υπολογισμό της θέσης μέσω του νευρωνικού δικτύου, συγκεντρώνεται κυρίως στον άξονα Z (εικόνα 4.7 (β, δ, στ, η)).



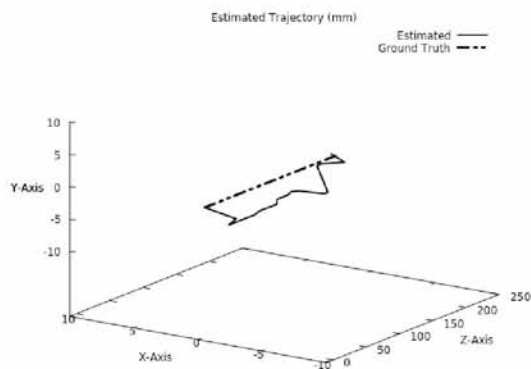
Εικόνα 4.7. Τρισδιάστατη ανακατασκευή πορείας του ενδοσκοπίου εντός του τεχνητού εντέρου έναντι της βάσης αληθείας. Αριστερά, παρουσιάζονται οι ανακατασκευασμένες πορείες χρησιμοποιώντας την κλασική μέθοδο γεωμετρικής οπτικής οδομετρίας (Συνεχίζεται στην επόμενη σελίδα).



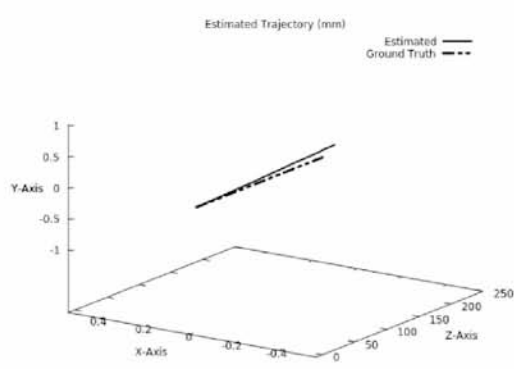
(ε)



(στ)



(ζ)



(η)

Εικόνα 4.7. (Συνέχεια) Δεξιά, παρουσιάζονται οι ανακατασκευασμένες πορείες χρησιμοποιώντας το τεχνητό νευρωνικό δίκτυο. Στις αναπαραστάσεις (α, β, γ, δ) η πορεία του ενδοσκοπίου είναι εμπρόςθια ενώ στις (ε, στ, ζ, η) η πορεία είναι προς τα πίσω.

4.3 Τεχνικές Λεπτομέρειες Υλοποίησης

Το νευρωνικό δίκτυο καθώς και το πειραματικό framework που περιγράφηκαν, υλοποιήθηκαν σε γλώσσα προγραμματισμού Python χρησιμοποιώντας την βιβλιοθήκη Keras και το TensorFlow ως backend του Keras. Τα πειράματα εκτελέστηκαν χρησιμοποιώντας επεξεργαστή Inter[®] 4×Core[™] i5-5200U CPU @ 2.20 GHz με 8048 MB RAM εκ των οποίων χρησιμοποιήθηκαν τα 5939 MB. Οι χρόνοι της εκπαίδευσης και της εξέτασης ήταν 20 min και

6.50 msec αντίστοιχα χρησιμοποιώντας μόνο τον RANSAC. Χρησιμοποιώντας το σχήμα οι χρόνοι εκπαίδευσης και αξιολόγησης ήταν 1.5 h και 53 msec αντίστοιχα. Οι χρόνοι εκπαίδευσης υπολογίστηκαν ως οι μέσοι χρόνοι εκτέλεσης του 5-fold cross validation, ενώ της αξιολόγησης ως ο μέσος χρόνος υπολογισμού κάθε στιβάδας.

5. Συμπεράσματα

Στην παρούσα εργασία προτείνεται και επαληθεύεται, η δυνατότητα εφαρμογής οπτικών μετρήσεων σε ενδοσκοπικά βίντεο και εικόνες. Συγκεκριμένα, παρουσιάζεται στο ερευνητικό πλαίσιο της ενδοσκόπησης, η καινοτόμα, διαδικασία της “έξυπνης” οπτικής οδομετρικής προσέγγισης για τον εντοπισμό του ενδοσκοπίου και συγκρίνεται με την κλασσική γεωμετρική μέθοδο. Η επαλήθευση επιτυγχάθηκε, μέσω ενός *ex-novo* ρομποτικού πειράματος, το οποίο προσέφερε βάση αληθείας, σε φυσικές μονάδες, για την κίνηση του ενδοσκοπίου εντός του τεχνητού εντέρου (π.χ mm, cm κλπ).

Η φυσική απόρροια της συλλογιστικής που αναπτύχθηκε, είναι ότι η προσέγγιση του προβλήματος, χρησιμοποιώντας μηχανική μάθηση, παρέχει βελτιωμένα αποτελέσματα συγκριτικά με την γεωμετρική προσέγγιση. Τα σφάλματα που αναφέρονται στη βιβλιογραφία σχετικά με γεωμετρικές προσεγγίσεις οπτικής οδομετρίας, κυμαίνονται μεταξύ 2.7 και 7.2 cm χρησιμοποιώντας διάφορες πειραματικές διατάξεις (Iakovidis *et al.*, 2016) (Bao *et al.*, 2014).

Έχουν επίσης προταθεί και άλλες μέθοδοι εντοπισμού του ενδοσκοπίου (κυρίως της ασύρματη κάμουλας ενδοσκόπησης), οι οποίες σημειώνουν απόδοση συγκρίσιμη ή και καλύτερη σε σχέση με την προτεινόμενη εργασία. Για παράδειγμα οι Geng & Pahlavan (2016), ακολούθησαν μια υβριδική προσέγγιση για την αντιμετώπιση του προβλήματος εντοπισμού του ενδοσκοπίου, συνδυάζοντας, μεθόδους βασισμένες σε ραδιοσυχνότητες, με μεθοδολογία βασισμένη σε οπτική πληροφορία. Στην εν λόγω εργασία, αντικείμενο μελέτης αποτέλεσαν δύο μέθοδοι βασισμένες σε ραδιοσυχνότητες, ονομαστικά, η Ισχύς του Ληφθέντος Σήματος (Receiving Signal Strength, RSS) και ο Χρόνος Άφιξης (Time Of Arrival, TOA). Μέσω της υβριδικής προσέγγισης, χρησιμοποιώντας τη μέθοδο ισχύος του ληφθέντος σήματος, επιτεύχθηκε ακρίβεια μετρήσεων 3.2 cm, καθώς με τη μέθοδο χρόνου άφιξης η αντίστοιχη ακρίβεια ήταν 1.8 cm. Τα παραπάνω αποτελέσματα διαμορφώθηκαν έτσι χρησιμοποιώντας ως μέσο εύρεσης περιθωρίου βελτίωσης, την μεθοδολογία Posterior Cramer-Rao Lower Bound (PCRLB).

Τα αποτελέσματα αυτά, υπολογίστηκαν βάσει πειραμάτων που αναπτύχθηκαν σε υπολογιστικές προσομοιώσεις. Αντιθέτως, η πειραματική αξιολόγηση της προτεινόμενης μεθοδολογίας βασίζεται σε πραγματικά δεδομένα που έχουν ληφθεί από μία *ex-vivo* πειραματική διάταξη. Συγκρίνοντας τα αποτελέσματα της μελέτης που παρουσιάζονται σε αυτή την εργασία με αυτά της εργασίας των Geng και Pahlavan (2016), φαίνεται ότι είναι συγκρίσιμα. Ωστόσο, θα πρέπει να υπογραμμιστεί, ότι οι μεθοδολογίες οι οποίες βασίζονται σε ραδιοσυχνότητες, απαιτούν πολλούς αισθητήρες τοποθετημένους στην κοιλιακή περιοχή του ασθενούς. Οι αισθητήρες αυτοί αποτελούν πηγή ενόχλησης και δεν επιτρέπουν στον ασθενή να ακολουθήσει πλήρως τις καθημερινές του δραστηριότητες. Η “έξυπνη” προσέγγισή της οπτικής οδομετρίας, βασίζεται αποκλειστικά σε οπτική πληροφορία και καθίσταται δυνατό να χρησιμοποιηθεί ως μια εναλλακτική λύση στη γεωμετρική προσέγγιση της οπτικής οδομετρίας, για την περαιτέρω βελτίωση της ακρίβειας εντοπισμού του ενδοσκοπίου.

Μία ακόμα αξιοσημείωτη έρευνα, παρουσιάστηκε από τους Pourhomayoun et al. (2012), που αργότερα επεκτάθηκε από τους ίδιους (*Pourhomayoun et al.*, 2014). Στην εν λόγω μελέτη προτάθηκε μια μεθοδολογία που χρησιμοποιεί τόσο την Ισχύ Ληφθέντος Σήματος όσο και το Χρόνο Αφίξης, ώστε να αναπτυχθεί μια αποτελεσματική μέθοδος που θα μπορούσε να εφαρμοστεί για τον εντοπισμό του ενδοσκοπίου, βασιζόμενη στην αραιότητα του εκπεμπόμενου σήματος της κάψουλας, στον τρισδιάστατο χώρο. Αυτή η μεθοδολογία είναι ανθεκτική στον θόρυβο και λειτουργεί καλά με μικρά δείγματα σημάτων από τον πομπό σε συνδιασμό με μικρό αριθμό δεκτών. Επίσης, η μέθοδος αυτή είναι αποδοτική ακόμα και όταν το σήμα του πομπού (ασύρματη κάψουλα ενδοσκόπησης) είναι χαμηλής ισχύος, κάτι που την καθιστά ικανή να προσαρμόζεται στα διάφορα είδη ιστών που υπάρχουν στο σώμα. Το συνολικό σφάλμα εντοπισμού που επιτεύχθηκε ήταν λιγότερο από 8.8 mm όταν η σύσταση του ιστού ήταν γνωστή και μικρότερο από 7.5 mm (στη χειρότερη περίπτωση) όταν η σύσταση του ιστού ήταν απροσδιόριστη. Τα σφάλματα που αναφέρονται σε αυτό το έργο είναι σημαντικά χαμηλότερα από τις τιμές που προέκυψαν με την δικιά μας προσέγγιση. Ωστόσο, τα αποτελέσματα της έρευνας των Pourhomayoun et al. αναφέρονται στον γενικό τρισδιάστατο χώρο της κοιλιακής περιοχής, ενώ η παρούσα μελέτη διερευνά τον εντοπισμό του ενδοσκοπίου συναρτήσει του εσωτερικού του γαστρεντερικού συστήματος. Επιπλέον, όπως και στις

προηγούμενες μελέτες που αναφέρθηκαν τα αποτελέσματα αυτά, επιτεύχθηκαν με τη χρήση υπολογιστικών προσομοιώσεων. Επίσης, λόγω του ότι αυτή η μεθοδολογία βασίζεται σε ραδιοσυχνότητες, καθίσταται εξαρτώμενη από εξωτερικούς αισθητήρες που θα πρέπει να τοποθετηθούν στο σώμα του ασθενούς. Για αυτόν τον λόγο, αυτού του είδους οι μέθοδοι μπορεί να είναι ευαίσθητες στην αποδυνάμωση του μεταδιδόμενου σήματος η οποία προκύπτει λόγω της φύσεως των ιστών των οργάνων (υπάρχουν μεταβολές στο πάχος, στη σύσταση, κ.λπ.). Η παριστάμενη μεθοδολογία, βασίζεται εξ ολοκλήρου σε οπτική πληροφορία που λαμβάνεται από την κάμερα που φέρει το κάθε ενδοσκόπιο.

Σε μια άλλη, συγκριτική μελέτη, από τους Ye et al. (2014), η αποτελεσματικότητα των βασισμένων σε ραδιοσυχνότητες μεθόδων, χρόνος άφιξης και ισχύς ληφθέντος σήματος, αποτέλεσε κεντρικό σημείο μελέτης. Στο έργο αυτό, πραγματοποιήθηκε πείραμα βασισμένο σε υπολογιστική προσομοίωση και η μέθοδος του χρόνου άφιξης φάνηκε να είναι πιο ακριβής από αυτή της Ισχύς του ληφθέντος σήματος, σε θορυβώδες περιβάλλον, με τη χρήση δύο αισθητήρων τύπου δέκτη. Τα σφάλματα που προέκυψαν χρησιμοποιώντας τη μέθοδο του χρόνου άφιξης ήταν της τάξεως χιλιοστών, που στη χειρότερη περίπτωση φθάναν τα 3.5 cm. Από την άλλη πλευρά, η μέθοδος της ισχύος ληφθέντος σήματος παρήγαγε σφάλματα της τάξεως εκατοστών, με την υψηλότερη τιμή σφάλματος να είναι 5.1 cm. Παρά την μεγάλη ακρίβεια, η προσέγγιση του Χρόνου Άφιξης μπορεί εύκολα να διαταραχθεί από απλές σωματικές κινήσεις. Κατά αυτόν τον τρόπο, προκύπτει άμεσα, ότι εάν η θέση των αισθητήρων, που είναι τοποθετημένοι στον ασθενή, αλλάξει με απλές σωματικές κινήσεις, τότε προκύπτουν πολύ μεγαλύτερα σφάλματα στη μέτρηση της απόστασης, μεταξύ των αισθητήρων και του ενδοσκοπίου.

Σε σύγκριση με προηγούμενα *in-vivo* έργα, μία εκ των καινοτομιών της προτεινόμενης μεθοδολογίας βασίζεται στην πειραματική διάταξη από την οποία λήφθηκε η βάση αληθείας. Το συγκεκριμένο πείραμα παρήχε βάση αληθείας πάνω σε πραγματικά δεδομένα και όχι μέσω υπολογιστικών προσομοιώσεων. Με αυτό τον τρόπο, τα αποτελέσματα αναφέρονται σε όρους φυσικών μονάδων και έτσι είναι εφικτή η αξιολόγηση της χρησιμότητας της, σε εφαρμογές όπως ο εντοπισμός θέσης αναγνωρισμένων αλλοιώσεων στην οδό του γαστρεντερικού

συστήματος.

Παρόλο που τα αποτελέσματα βασίζονται σε *ex-vivo* πειραματική διάταξη, θα πρέπει να σημειώσει κανείς ότι:

- i. Η επιφάνεια του τεχνητού εντέρου, κατά την διάρκεια διεξαγωγής του πειράματος, ήταν υγρή, όπως και του πραγματικού εντέρου.
- ii. Οι αντιστοιχίες των σημείων ενδιαφέροντος που εντοπίστηκαν στις χρωματιστές πινέζες που χρησιμοποιήθηκαν για την θέσπιση της βάσης αληθείας, αποκλείονται από το σύνολο της διαδικασίας εντοπισμού της θέσης του ενδοσκοπίου.
- iii. Σε αντίθεση με την πραγματική δραστηριότητα του έντερο, το τεχνητό έντερο, δεν παρουσιάζει καθόλου κινητικότητα (όπως π.χ συσπάσεις).
- iv. Η υφή (texture) του πραγματικού εντέρου είναι πιο πλούσια σε σχέση με εκείνη του τεχνητού, εξαιτίας της παρουσίας των εντερικών λαχνών. Επομένως, τα σημεία ενδιαφέροντος που προκύπτουν, μπορούν να οδηγήσουν σε περισσότερες σωστές αντιστοιχίες.
- v. Παρότι οι επιπλέον ουσίες (υπολείμματα) δεν έχουν συμπεριληφθεί στο πείραμά μας, σε βίντεο πραγματικών ενδοσκοπικών εξετάσεων, μπορούν να μειωθούν σημαντικά με μια καλή προετοιμασία των ασθενών χρησιμοποιώντας καθαρτικά φάρμακα ή/και κατάλληλη διατροφή (Koulaouzidis *et al.*, 2015). Τα υπολείμματα αυτά, τείνουν να έχουν διαφορετικό σχήμα μεταξύ διαδοχικών στιγμιότυπων, και έτσι οι αντιστοιχίες που εντοπίζονται μεταξύ τους είναι σπάνιες. Ακόμη όμως και στην περίπτωση που υπάρξει αντιστοιχία, όπως αναφέρεται και στην υποενότητα 3.4, η χρήση της χρωματικής πληροφορίας, μπορεί να παρέχει συμπληρωματική ευρωστία στην μεθοδολογία, υπό την παρουσία των υπολειμμάτων.

Ένα σημαντικό πλεονέκτημα των μεθόδων υπολογισμού θέσης, χρησιμοποιώντας την μέθοδο της οπτικής οδομετρίας έναντι άλλων, συμβατικών, σύγχρονων μεθόδων που έχουν

προταθεί στο πλαίσιο της ενδοσκόπησης (κυρίως της ασύρματης) είναι ότι δεν απαιτούν κανένα εξωτερικό αισθητήρα (πομπό ή δέκτη). Έτσι, προκύπτει το συμπέρασμα ότι η αποτελεσματική εφαρμογή της προτεινόμενης μεθοδολογίας, μπορεί να οδηγήσει στην ανάπτυξη ενδοσκοπίων με βελτιωμένη χρηστικότητα, τόσο για τους γιατρούς και όσο και για τους ασθενείς. Προσφέροντας μεγαλύτερη άνεση στους ασθενείς, και βελτιώνοντας τον λόγο κόστους - αποτελέσματος, διότι δεν χρειάζονται επιπλέον έξοδα για τους εξωτερικούς αισθητήρες και τη συντήρησή τους.

Καλό είναι να επισημανθεί, ότι η προσέγγιση της οπτικής οδομετρίας, επιτρέπει τον εντοπισμό του ενδοσκοπίου συναρτήσει του εσωτερικού του γαστρεντερικού συστήματος και όχι στον δισδιάστατο ή τρισδιάστατο κοιλιακό χώρο, όπως συμβαίνει με προσεγγίσεις που χρησιμοποιούν αισθητήρες. Σε αντίθεση με τις προσεγγίσεις οι οποίες βασίζονται σε ραδιοσυχνότητες, η ακρίβεια της οπτικής οδομετρίας, δεν εξαρτάται από τις μεταβλητές που επηρεάζουν την απόδοση των αισθητήρων, όπως είναι ο κοιλιακός όγκος, οι ιστοί των οργάνων και η κίνηση του σώματος. Ωστόσο, οι υβριδικές προσεγγίσεις εντοπισμού θέσης του ενδοσκοπίου που συνδυάζουν την γεωμετρική οπτική οδομετρία με μεθοδολογίες που χρησιμοποιούν εξωτερικούς αισθητήρες, επιτυγχάνουν αποτελέσματα μεγάλης ακρίβειας. Εκτιμώντας τα αποτελέσματα όσων έχουν ήδη προταθεί από την πλευρά της οπτικής οδομετρίας, η ενσωμάτωσή της σε ένα υβριδικό σύστημα, χρησιμοποιώντας επιπλέον σύγχρονους αισθητήρες μπορεί να οδηγήσει σε περαιτέρω βελτίωση της ακρίβειας εντοπισμού του ενδοσκοπίου.

Η κρίση αυτής της μελέτης θα ήταν ελλιπής αν δεν υπογραμμιστεί ότι οι περισσότερες από τις σύγχρονες έρευνες που αναφέρονται εδώ, δείχνουν αρκετά ελπιδοφόρα αποτελέσματα για το μέλλον του συγκεκριμένου ερευνητικού πλαισίου. Ωστόσο, οι περισσότερες από αυτές δείχνουν ότι οι τεχνολογίες εντοπισμού της θέσης των ενδοσκοπίων βρίσκονται ακόμα σε πρώιμο στάδιο, καθώς τα αποτελέσματα βασίζονται, σχεδόν αποκλειστικά, σε υπολογιστικές προσομοιώσεις. Σε αυτή την ερευνητική προσπάθεια, γίνεται ένα βήμα μπροστά, καθώς πέρα από την επαλήθευση της δυνατότητας εφαρμογής της γεωμετρικής προσέγγισης σε ενδοσκοπικές εικόνες/βίντεο, επικυρώνεται η δυνατότητα εφαρμογής “έξυπνων” μεθόδων σε

ex-novo πειραματικές συνθήκες, επιτρέποντας έτσι την επαλήθευση των θεωρητικών υποθέσεων, σε φυσικές μονάδες μέτρησης.

Αντίθετα από τις προηγούμενες προσεγγίσεις οπτικής οδομετρίας που προτείνονται στο πλαίσιο της ενδοσκοπησης και βασίζονται σε γεωμετρικά μοντέλα, η παρούσα “έξυπνη” μέθοδος είναι ικανή να προσαρμοστεί σε οποιοδήποτε γεωμετρικό μοντέλο κάμερας. Επομένως, μπορεί να εφαρμοστεί σε οποιοδήποτε ενδοσκόπιο και να προσαρμοστεί τόσο στο γεωμετρικό όσο και στο χρωματικό μοντέλο (δηλαδή, τον τρόπο με τον οποίο ο αισθητήρας της κάμερας ανταποκρίνεται στο συχνοτικό ορατό φάσμα του φωτός). Έτσι, προσφέρετε το πλεονέκτημα ότι δεν είναι απαραίτητο να προηγηθεί καμία υπόθεση όσον αφορά το γεωμετρικό και το χρωματικό μοντέλο της κάμερας. Επιπλέον, δεν απαιτείται καμία γνώση σχετικά με τις εσωτερικές παραμέτρους της κάμερας του ενδοσκοπίου. Αυτό καθίσταται μείζονος σημασίας, δεδομένου ότι πληροφορίες επί των εσωτερικών παραμέτρων της κάμερας, δεν παρέχονται συνήθως από τους κατασκευαστές τους. Ο τρόπος με τον οποίο υπολογίζονται οι εσωτερικές παράμετροι της κάμερα, είναι η μέθοδος της βαθμονόμησης (υποενότητα 2.1). Ωστόσο, μια τέτοια διαδικασία δεν μπορεί να θεωρηθεί ρουτίνα, και όπως έχει δειχθεί, λάθος υποθέσεις ως προς το μοντέλο της κάμερας, συνεπάγεται εσφαλμένο υπολογισμό των εσωτερικών παραμέτρων και οδηγεί σε σημαντικά μεγαλύτερα σφάλματα του υπολογισμού της θέσης της κάμερας/ενδοσκοπίου.

Καθίσταται σαφές ότι το προτεινόμενο “ευφύες” σχήμα οπτικής οδομετρίας σε ενδοσκοπικά βίντεο λειτουργούσε βάσει των εξής περιορισμών:

- i.** Το ενδοσκόπιο, καθ’ όλη την διάρκειά του πειράματος, διατηρούσε ευθεία πορεία.
- ii.** Η ταχύτητα του ενδοσκοπίου, ήταν σταθερή στην περίοδο μιας πλήρους διαδρομής.
- iii.** Το ενδοσκόπιο κινούταν μόνο προς δύο κατευθύνσεις, μπρος και πίσω.

Οι περιορισμοί αυτοί αντιπροσωπεύουν σπάνια παραδείγματα της πραγματικής κίνησης του ενδοσκοπίου εντός της γαστρεντερικής οδού. Γενικά, η γαστρεντερική οδός χαρακτηρίζεται από μαλακές πτυχώσεις και η κίνηση του ενδοσκοπίου είναι μεταβλητή, με

ακαθόριστο τρόπο. Οι ταχύτητες που ερευνήθηκαν προσομοιάζουν την κανονική, αργή κίνηση του λεπτού εντέρου. Ένα πιο απαιτητικό πρόβλημα για τον εντοπισμό της θέσης του ενδοσκοπίου αποτελεί η περίοδος σύσπασης του εντέρου, κατά την οποία η ασύρματη κάψουλα ενδοσκόπησης κινείται γρηγορότερα. Παρόλα αυτά, ο ρυθμός λήψης/προβολής στιγμιοτύπων των τρεχόντων ασύρματων ενδοσκοπίων (όχι όλων) είναι πολύ χαμηλός και δεν επιτρέπει τη λήψη επαρκών στιγμιοτύπων κατά τη διάρκεια της περιόδου σύσπασης. Για παράδειγμα, μια εμπορική κάμερα ασύρματου ενδοσκοπίου όπως αυτή που χρησιμοποιείται στο παρόν πείραμα, έχει μέγιστο ρυθμό λήψης/προβολής στιγμιοτύπων 6 fps, θα είναι σε θέση να συλλάβει ένα μέγιστο αριθμό 3 στιγμιοτύπων για μία απόσταση 2 cm, μετά από ένα περισταλτικό κύμα με πολλαπλάσια ταχύτητα 2 cm/s. Συμπληρωματικά, υπάρχει μία πολύ μικρή πιθανότητα αυτά τα στιγμιότυπα να μοιράζονται σημεία που να αντιστοιχούν στο ίδιο σημείο του πραγματικού 3D πλάνου.

Σύμφωνα με τους Bao et al. (2014) όταν η ταχύτητα της κάψουλας φτάνει έως 10 mm/s, σχεδόν κανένα κοινό τμήμα δεν υπάρχει μεταξύ δύο διαδοχικών στιγμιοτύπων, συνεπώς, κανένα αντίστοιχο σημείο δε μπορεί να ανιχνευθεί και η μέθοδος της οπτικής οδομετρίας, αποτυγχάνει. Προκύπτει με σαφήνεια, ότι ενδοσκόπια με υψηλότερο ρυθμό λήψης/προβολής στιγμιοτύπων χρειάζεται να αναπτυχθούν, προκειμένου να αξιοποιηθεί η μέθοδος της οπτικής οδομετρίας σε συνθήκες *in-vivo* (Iakovidis et al., 2014). Πέρα από τους περιορισμούς της ισχύουσας τεχνολογίας των ενδοσκοπίων, μια πιο ρεαλιστική *ex-vivo* πειραματική διάταξη για την επαλήθευση του εντοπισμού της θέσης των ενδοσκοπίων, θα απαιτούσε την αντιμετώπιση σημαντικών μηχανικών προκλήσεων. Κάποιες από αυτές έχουν να κάνουν με την επινόηση μεθόδων για τον έλεγχο της κίνησης του ενδοσκοπίου, επί την παρουσία μαλακών πτυχώσεων, και μια ρεαλιστική εξομοίωση των συστολών του εντέρου. Ερευνητικά βήματα προς αυτή την κατεύθυνση έχουν ήδη γίνει, π.χ η ανάπτυξη ρεαλιστικών προσομοιώσεων των συνθηκών και της συμπεριφοράς του εντέρου (Slawinski et al., 2015). Ωστόσο, η αξιοποίηση και η εκμετάλλευση αυτών των τεχνολογιών, για τον σκοπό της επικύρωσης της δυνατότητας εφαρμογής της οπτικής οδομετρίας σε ενδοσκοπικές εικόνες/βίντεο, μπορούν να θεωρηθούν ακόμα ως ένα μελλοντικό βήμα. Μία εκτενής συγκριτική μελέτη σχετικά με μεθόδους υπολογισμού της ασύρματης κάψουλας ενδοσκόπησης έχει συνταχθεί από τους Than et al.

(2012)

Στην συγκεκριμένη εργασία, ερευνήθηκε η εφαρμογή της κλασσικής γεωμετρικής προσέγγισης οπτικής οδομετρίας σε ενδοσκοπικά βίντεο και εικόνες για την μέτρηση αποστάσεων που διάνυσε το ενδοσκόπιο. Επίσης, προτάθηκε μία καινοτόμα “έξυπνη” προσέγγιση στο πρόβλημα της μέτρησης αποστάσεων μέσω εικόνων και βίντεο χρησιμοποιώντας τεχνητά νευρωνικά δίκτυα. Η σύγκριση των δύο μεθοδολογιών ως προς την μέτρηση αποστάσεων μέσω εικόνων/βίντεο από μια ασύρματη κάψουλα ενδοσκόπησης, έδειξε ότι η προτεινόμενη μεθοδολογία που χρησιμοποιεί “μηχανική μάθηση” είναι πολύ πιο αποδοτική από αυτή της κλασσικής μέθοδου. Συνοπτικά, η προσφορά αυτής της εργασίας είναι η εξής:

- i. Εκμετάλλευση της εμφάνισης των εικονοστοιχείων των σημείων που ανιχνεύει ο αλγόριθμος SIFT. Δηλαδή, η ένταση της φωτεινότητας και το χρώμα από τα οποία λαμβάνεται περεταίρω πληροφορία για το βάθος του πλάνου μια εικόνας/στιγμιότυπου.
- ii. Η χρήση του αλγορίθμου KLT για την ανίχνευση οπτικής ροής σε συνδυασμό με τον RANSAC για την εξάλειψη εσφαλμένων αντιστοιχιών σημείων ενδιαφέροντος.
- iii. Χρήση ενός τεχνητού νευρωνικού δικτύου, σχεδιασμένο για την εκτέλεση οπτικής οδομετρίας χρησιμοποιώντας μόνο μία κάμερα, με εφαρμογή σε ενδοσκοπικά βίντεο.
- iv. Τα δεδομένα εξέτασης της μεθοδολογίας και εκπαίδευσης του δικτύου, προήλθαν από ένα καινοτόμο ρομποτικό πείραμα, μέσω του οποίου εδραιώθηκε βάση αληθείας για την επαλήθευση της προτεινόμενης μεθοδολογίας σχετικά με την θέση του ενδοσκοπίου.

Τα αποτελέσματα που προέκυψαν από το πείραμα, έδειξαν ότι η προτεινόμενη “έξυπνη” οπτική οδομετρία δίνει σαφώς καλύτερα αποτελέσματα σε σχέση με την κλασσική προσέγγιση. Ένα σημαντικό συμπέρασμα που πηγάζει μέσα από αυτά τα αποτελέσματα, είναι ότι η χρωματική πληροφορία λειτουργεί ευεργετικά για τον υπολογισμό της θέσης του ενδοσκοπίου, χρησιμοποιώντας την προτεινόμενη αρχιτεκτονική δικτύου. Η χρωματική πληροφορία φάνηκε να λειτουργεί ως απλοποιητικός παράγοντας του προβλήματος του εντοπισμού της θέσης. Η

λύση του προβλήματος χρησιμοποιώντας πληροφορία χρώματος, απλοποιεί σημαντικά την αρχιτεκτονική του δικτύου (μικρότερος αριθμός νευρώνων του κρυφού επιπέδου). Αντιθέτως, με την απουσία της χρωματικής πληροφορίας από τα διανύσματα εισόδου του δικτύου και με την χρήση μόνο των συντεταγμένων των αντιστοιχισμένων σημείων, η πολυπλοκότητα του δικτύου για την απόδοση καλών αποτελεσμάτων αυξάνεται αρκετά.

Βασιζόμενοι στα παραπάνω αποτελέσματα, στα σχέδια για μελλοντικές μελέτες είναι η περαιτέρω έρευνα για την βελτιστοποίηση της απόδοσης εύρεσης θέσης των ενδοσκοπίων, χρησιμοποιώντας διαφορετικές αρχιτεκτονικές δικτύων, εστιάζοντας σε αυτά με χαμηλή πολυπλοκότητα (π.χ μερικώς συνδεδεμένα δίκτυα και αλγόριθμοι με λιγότερες παραμέτρους όπως αυτός της ανάστροφης διάδοσης σφάλματος με προσαρμοζόμενο ρυθμό μάθησης).

Η πειραματική διάταξη η οποία χρησιμοποιήθηκε για την επαλήθευση της μεθοδολογίας σε αυτή την έρευνα, μπορεί να χαρακτηριστεί ως υπεραπλοποιημένη σε σχέση με τις πραγματικές συνθήκες του γαστρεντερικού συστήματος. Παρόλα αυτά, τα πειράματα που έγιναν χρησιμοποιώντας την συγκεκριμένη διάταξη μπορούν να δώσουν αποτελέσματα και στοιχεία για την δυνατότητα εφαρμογής της οπτικής οδομετρίας σε πραγματικές συνθήκες. Μία επιτυχημένη εφαρμογή μετρήσεων βασισμένη σε οπτική πληροφορία σε ένα τέτοιο πειραματικό περιβάλλον, σηματοδοτεί “πράσινο φως” για την ανάπτυξη πειραμάτων, μεγαλύτερης πολυπλοκότητας και ρεαλισμού. Οπότε, το συγκεκριμένο πείραμα χρησιμεύει ως μέσο επαλήθευσης της ύπαρξης δυνατότητας οπτικών μετρήσεων σε ενδοσκοπικά βίντεο και ως αποτέλεσμα, εισιτήριο για την επένδυση στην ανάπτυξη πιο περίπλοκων πειραμάτων.

Ανάμεσα στα μελλοντικά πλάνα έρευνας στο πλαίσιο αυτό, είναι η σύσταση πειράματος στο οποίο το ενδοσκόπιο θα κινείται, όχι μόνο ευθεία, αλλά και περιστροφικά μέσα στο έντερο με αλλαγές προσανατολισμού. Αυτό θα οδηγήσει, σε πείραμα το οποίο θα προσομοιώνει καλύτερα τις πραγματικές συνθήκες του γαστρεντερικού συστήματος και της συμπεριφοράς του ενδοσκοπίου εντός αυτού. Πέραν τούτου, λειτουργεί σαν ένα ακόμα βήμα προς μία πλήρως αυτοματοποιημένη διαδικασία ανίχνευσης της θέσης του ενδοσκοπίου εντός του γαστρεντερικού συστήματος. Ωστόσο, μέχρι τη στιγμή της πλήρους αυτοματοποίησης αυτής της διαδικασίας υπάρχουν πολλά εμπόδια να αντιμετωπιστούν, όπως οι μαλακές πτυχώσεις

του εντέρου. Επίσης, στο πλαίσιο της ασύρματης ενδοσκόπησης με κάψουλα, πολύ μεγάλη δυσκολία σηματοδοτεί η σύσπαση του εντέρου, η οποία δυσκολία είναι ανάλογη της δύναμης και της συχνότητάς της σύσπασης. Επίσης, ένα από τα μελλοντικά σχέδια, είναι η έρευνα της δυνατότητας μέτρησης των διαστάσεων ανωμαλιών και παθολογιών σε βίντεο ασύρματης κάψουλας ενδοσκόπησης. Σχετικά με την *in-vivo* εξέταση του προβλήματος, είναι δυνατή η σύσταση πειράματος χρησιμοποιώντας πολλαπλά εργαλεία ιατρικής απεικόνισης. Παραδείγματος χάριν, ο συνδυασμός της ασύρματης ενδοσκόπησης με ραδιολογικά δεδομένα τα οποία μπορούν να παρέχουν βάση αληθείας για επαλήθευση της μεθοδολογίας σε συνθήκες *in-vivo*.

6. Βιβλιογραφία

Ahmed, T. M., Hemayed, E. E. and Farag, A. A., *Neurocalibration: a neural network that can tell camera calibration parameters*, Computer Vision, The Proceedings of the Seventh IEEE International Conference on (Vol. 1, pp. 463-468), 1999.

Bao, G., Mi, L. and Pahlavan, K., *Emulation on motion tracking of endoscopic capsule inside small intestine*, Proceedings of the International Conference on Bioinformatics & Computational Biology (BIOCOMP) (p. 1), The Steering Committee of The World Congress in Computer Science, Computer Engineering and Applied Computing (WorldComp), 2013.

Bao, G., Mi, L., Geng, Y., Zhou, M. & Pahlavan, K., *A video-based speed estimation technique for localizing the wireless capsule endoscope inside gastrointestinal tract*, Engineering in Medicing and Biology Society (EMBC), 36th Annual International Conference of the IEEE, pp. 5615-5618, 2014.

Beşdok, E., *3D Vision by using calibration pattern with inertial sensor and RBF Neural Networks*, Sensors, 9(6), 4572-4585, 2009.

Bouget, J.-Y., *Camera calibration toolbox for Matlab*, 2004.

Faessler, M., Fontana, F., Forster, C., Mueggler, E., Pizzoli, M. and Scaramuzza, D., *Autonomous Vision-based Flight and Live Dense 3D Mapping with a Quadrotor Micro Aerial Vehicle*, Journal of Field Robotics 33(4), 431-450, 2016

G. I. Cooperation, *Pillcam[®] Capsule Endoscopy RAPID[®] V8.0 User Manual*, DOC-2044th-02 ed., 2013.

Geng, Y. & Pahlavan, K., *Design, Implementation, and fundamental limits of image and RF based wireless capsule endoscopy hybrid localization*, Transactions on Mobile Computing, IEEE, 15(8), 1951-1964, 2016.

Hartley, R. & Zisserman, A., *Multiple view geometry in computer vision*, Cambridge university press, 2003.

Haykin, S. & Network, N., *A comprehensive foundation*, Neural Networks, 2(2004), 41, 2004.

Heikkilä, J. & Silven, O., *A four-step camera calibration procedure with implicit image correction*, In Computer Vision and Pattern Recognition, Proceedings 1997, IEEE Computer Society Conference, pp 1106-1112, 1997.

Heikkilä, J., *Geometric camera calibration using circular control points*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 22(1), 1066-1077, 2000.

Henry, P., Krainin, M., Herbst, E., Ren, X. & Fox, D., *RGB-D mapping: Using depth cameras for dense 3D modeling of indoor environments*, In the 12th International Symposium on Experimental Robotics (ISER), 2010.

Hornik, K., Stinchcombe, M. & White, H., *Universal approximation of an unknown mapping and its derivatives using multilayer feedforward networks*, Neural Networks, 3(5), 551-560, 1990.

Huang, A. S., Bachrach, A., Henry, P., Krainin, M., Maturana, D., Fox, D. & Roy, N., *Visual odometry and mapping for autonomous flight using an RGB-D camera*, In Robotics Research, pp. 235-252, Springer International Publishing, 2017.

Iakovidis, D. K. & Koulaouzidis, A., *Automatic lesion detection in capsule endoscopy based*

on color saliency: closer to an essential adjunct for reviewing software, Gastrointestinal endoscopy, 80(5), 877-883, 2014.

Iakovidis, D. K., Dimas, G., Karargyris, A., Ciuti, G., Bianchi, F., Koulaouzidis, A. and Toth, E., *Robotic validation of visual odometry for wireless capsule endoscopy*, Imaging Systems and Techniques (IST), IEEE International Conference on (pp. 83-87), 2016.

Iakovidis, D. K., Sarmiento, R., Silva, J. S., Histace, A., Romain, O., Koulaouzidis, A., ... & Dray, X., *Towards intelligent capsules for robust wireless endoscopic imaging of the gut*, IEEE International Conference on Image Systems and Techniques, pp. 95-100, 2014.

Iakovidis, D. K., Spyrou, E., Diamantis, D. and Tsiopanidis, I., *Capsule endoscope localization based on visual features*, Bioinformatics and Bioengineering, IEEE 13th International Conference on pp. 1-4, 2013.

Kannala, J., Brandt, S. S., *A generic camera model and calibration method for conventional, wide-angle, and fish-eye lenses*, IEEE transactions on pattern analysis and machine intelligence, 28(8), 1335-1340, 2006.

Karkanis, S. A., Iakovidis, D. K., Maroulis, D. E., Karras, D. A, Tzivrass, M., *Computer-aided tumor detection in endoscopic video using color wavelet features*, IEEE transactions on information technology in biomedicine, 7(3), 141-152, 2003.

Kita, H. & Ciuti, G. , *Double-balloon endoscopy for the diagnosis and treatment of small intestinal disease*, Best Practice & Research Clinical Gastroenterology, vol. 20, no. 1, pp. 179-194, 2006.

Konda, K. R. and Memisevic, R., *Learning Visual Odometry with a Convolutional Network*, VISAPP (1) (pp. 486-490), 2001.

Koulaouzidis, A., Iakovidis, D. K., Karargyris, A. & Plevris, J. N., *Optimizing lesion detection in small-bowel capsule endoscopy: from present problems to future solutions*, Expert review of gastroenterology & hepatoenterology, 9(2), 217-235, 2015.

Koulaouzidis, A., Iakovidis, D. K., Karargyris, A. & Rondonotti, E., *Wireless capsule endoscopy in 2020: Will it still be a capsule?*, World Journal of gastroenterology: WJG, 21(17), 5119, 2015.

Krizhevsky, A., Sutskever, I. & Hinton, G. E., *Imagenet classification with deep convolutional neural networks*, In *Advances in neural information processing systems*, pp 1097-1105, 2012.

LeCun, Y., Bengio, Y. & Hinton, G., *Deep Learning*, *Nature*, 521(7553), 436-444, 2015.

Lowe, D. G., *Distinctive image features from scale-invariant keypoints*, International journal of computer vision, 60(2), 91-110, 2004.

Luong, Q.-T. & Faugeras, O., *Self-calibration of a moving camera from point correspondences and fundamental matrices*, The International Journal of Computer Vision, 22(3), 261-289, 1997.

Maimone, M., Yang, C. and Matthies, L., *Two years of visual odometry on the mars exploration rovers*, Journal of Field Robotics 24(3), 169-186, 2007

Memon, Q. and Khan, S., *Camera calibration and three-dimensional world reconstruction of stereo-vision using neural networks*, International Journal of Systems Science, 32(9), 1155-1159, 2001.

Menze, M. and Geiger, A., *Object scene flow for autonomous vehicles*, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 3061-3070), 2015.

Mi, L., Bao, G. and Pahlavan, K., *Geometric estimation of intestinal contraction for motion tracking of video capsule endoscope*, SPIE Medical Images (pp. 90360B-90360B), International Society for Optics and Photonics, 2014.

Nistér, D., Naroditsky, O. and Bergen, J., *Visual odometry*, Computer Vision and Pattern Recognition, Proceedings of the 2004 IEEE Computer Society Conference on, vol. 1, pp. I-652, 2004.

Pourhomayoun, M., Fowler, M. & Jin, Z., *A novel method for medical implant in-body localization*, Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEE, pp. 5757-5760, 2012.

Pourhomayoun, M., Fowler, M. L. & Jin, Z., *Accurate localization of in-body medical implants based on spatial sparsity*, IEEE Transactions on Biomedical Engineering, 61(2), 590-597, 2014.

Scaramuzza, D. and Friedrich, F, *Visual Odometry [tutorial]*, IEEE Robotics and Automation magazine, 18(4), 80-92, 2011.

Shi, J., *Good features to track*, Computer Vision and Pattern Recognition, Proceedings CVPR'94, IEEE Computer Society Conference on pp. 593-600, 1994.

Slawinski, P. R., Oleynikov, D. & Terry, B. S., *Intestinal biomechanics simulator for robotic capsule endoscope validation*, Journal of medical engineering & technology, 39(1), 54-59, 2015.

Sliker, L. J. & Ciuti, G., *Flexible and capsule endoscopy for screening diagnosis and treatment*, Expert review of medical devices, 11(6), 649-666, 2014.

Spyrou E. and Iakovidis, D. K., *Video-based measurements for wireless capsule endoscope tracking*, Measurements Science and Technology, 25(1), 015002, 2013.

Spyrou E., Iakovidis D. K. and Koulaouzidis, A., *Comparative assessment of feature extraction methods for visual odometry in wireless capsule endoscopy*, Computer in biology and medicine, 65, 297-307, (2015).

Than, T. D., Alici, G., Zhou, H. & Li, W., *A review of localization systems for robotic endoscopic capsules*, Biomedical Engineering, IEEE Transactions, 59(9), 2387-2399, 2015.

Tomasi, C. & Kanade, T., *Detection and tracking of point features*, 1991.

Tsai, R. Y., *A versatile camera calibration technique for high accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses*, IEEE Journal of Robotics and Automation, RA-3(4): 323-344, 1987.

Wyszecki, G. & Stiles, W. S., *Color science*, New York: Wiley, Vol. 8, 1982.

Ye, Y., Khan, U., Alsindi, N., Fu, R. & Pahlavan, K., *On the accuracy of RF positioning in multi-capsule for robust wireless endoscopic imaging of the gut*, IEEE International Conference on Imaging Systems and Techniques, pp. 95-100, 2014.

Zhang, Z., *Camera calibration with one-dimensional objects*, IEEE transactions on Pattern Analysis and Machine Intelligence, 26(7), 892-899, 2004.

Zhang, Z., *Flexible camera calibration by viewing a plane from unknown orientations*, The Proceedings of the Seventh IEEE International Conference on (Vol. 1, pp. 666-673), 1999.