

Ατομική Διπλωματική Εργασία

**ΑΦΗΓΗΣΗ ΠΑΡΑΜΥΘΙΩΝ ΣΕ ΜΙΚΤΗ
ΠΡΑΓΜΑΤΙΚΟΤΗΤΑ**

Κυριάκος Ηρακλέους

ΠΑΝΕΠΙΣΤΗΜΙΟ ΚΥΠΡΟΥ



ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ

Μάιος 2011

ΠΑΝΕΠΙΣΤΗΜΙΟ ΚΥΠΡΟΥ

ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ

Αφήγηση Παραμυθιών σε Μικτή Πραγματικότητα

Κυριάκος Ηρακλέους

Επιβλέποντες Καθηγητές

Γιώργος Χρυσάνθου, Χαράλαμπος Πουλλής

Η Ατομική Διπλωματική Εργασία υποβλήθηκε προς μερική εκπλήρωση των απαιτήσεων απόκτησης του πτυχίου Πληροφορικής του Τμήματος Πληροφορικής του Πανεπιστημίου Κύπρου

Μάιος 2011

Ευχαριστίες

Καταρχάς, θα ήθελα να ευχαριστήσω τους επιβλέποντες καθηγητές μου, κ. Γιώργο Χρυσάνθου και κ. Χαράλαμπο Πουλλή, για την πολύτιμη βοήθεια και καθοδήγηση που μου έχουν προσφέρει, κατά τη διάρκεια διεξαγωγής της διπλωματικής μου εργασίας.

Επίσης θα ήθελα να ευχαριστήσω του φίλους μου, για τη βοήθεια τους και τη στήριξη τους. Και τέλος οφείλω ένα μεγάλο ευχαριστώ στην οικογένεια μου, για την όλη υποστήριξη που μου παρείχαν κατά τη διάρκεια της φοίτησης μου.

Περίληψη

Η εργασία αυτή ασχολείται με την μελέτη και ανάπτυξη ενός συστήματος Επαυξημένης Πραγματικότητας (Augmented Reality). Μέσω του συστήματος ο υπολογιστής θα παρακολουθεί τον πραγματικό κόσμο μέσω μίας κάμερας. Όταν μέσα στη σκηνή που θα παρακολουθεί εντοπίζει ένα συγκεκριμένο βιβλίο θα προβάλει στο χρήστη μέσα από την οθόνη, μια εικονική σκηνή, η οποία θα δίνει την αίσθηση ότι βρίσκεται στην επιφάνεια του βιβλίου.

Κάτι τέτοιο αρχικά απαιτεί την αναγνώριση του βιβλίου από τον υπολογιστή, μέσα στη σκηνή που θα καταγράφει η κάμερα. Η αναγνώριση αυτή γίνεται μέσω περιγραφών δηλαδή, δημιουργείται μια περιγραφή για τη σελίδα του βιβλίου καθώς και μία περιγραφή για τη σκηνή που καταγράφει η κάμερα. Στη συνέχεια, αυτές οι περιγραφές συγκρίνονται και βάση της σύγκρισης τους αποφασίζεται αν το αντικείμενο - βιβλίο βρίσκεται μέσα στη σκηνή ή όχι.

Μετά την αναγνώριση του βιβλίου μέσα στη σκηνή, πρέπει να υπολογιστούν κάποια γεωμετρικά στοιχεία αυτού. Τα στοιχεία αυτά είναι ο χώρος που καταλαμβάνει μέσα στη σκηνή καθώς και η περιστροφή γύρω από τον εαυτό του σε σχέση με το σημείο από το οποίο το αντικείμενο παρακολουθείται από τη κάμερα. Τα στοιχεία αυτά στη συνέχεια θα χρησιμοποιηθούν για τη τοποθέτηση της εικονικής σκηνής πάνω στη επιφάνεια του βιβλίου.

Όλα αυτά πρέπει να γίνονται για κάθε καταγραφή της σκηνής από τη κάμερα και να προβάλλονται στο χρήστη σε πραγματικό χρόνο. Η διαδικασία της σύγκρισης των περιγραφών όμως, απαιτεί αρκετό υπολογιστικό χρόνο, προκαλώντας έτσι σε μερικές περιπτώσεις το σύστημα να μη λειτουργεί σε πραγματικό χρόνο. Έτσι, στη προσπάθεια επίλυσης αυτού του προβλήματος προτείνεται η χρήση Οπτικής Ροής (Optical Flow) για παράκαμψη της σύγκρισης των περιγραφών.

Περιεχόμενα

| | |
|---|-----------|
| Κεφάλαιο 1 Εισαγωγή | 1 |
| 1.1 Περιγραφή του συστήματος..... | 1 |
| 1.2 Πεδίο μελέτης | 4 |
| Κεφάλαιο 2 Θεωρητικό Υπόβαθρο | 6 |
| 2.1 Εισαγωγή | 6 |
| 2.2 Σύγκριση βασισμένη σε ανίχνευση και ταίριασμα χαρακτηριστικών | 6 |
| 2.3 Ανίχνευση κίνησης | 12 |
| Κεφάλαιο 3 Περιγραφή Συστήματος | 15 |
| 3.1 Εισαγωγή | 15 |
| 3.2 Διάγραμμα Τμημάτων του συστήματος | 16 |
| 3.3 Περιγραφή Τμημάτων..... | 17 |
| Κεφάλαιο 4 Εξαγωγή Χαρακτηριστικών | 20 |
| 4.1 Εισαγωγή | 20 |
| 4.2 Ανίχνευση Χαρακτηριστικών | 20 |
| 4.3 Περιγραφή Χαρακτηριστικών | 26 |
| 4.4 Σύνοψη..... | 29 |
| Κεφάλαιο 5 Ταύτιση Χαρακτηριστικών..... | 31 |
| 5.1 Εισαγωγή | 31 |
| 5.2 Διαδικασία ταιριάσματος χαρακτηριστικών | 31 |
| 5.3 Σύνοψη..... | 35 |
| Κεφάλαιο 6 Βαθμονόμηση Κάμερας (Camera Calibration)..... | 37 |
| 6.1 Εισαγωγή | 37 |
| 6.2 Βαθμονόμηση κάμερας..... | 38 |
| 6.3 Περιγραφή Διαδικασίας διεξαγωγής Geometric Camera Calibration | 42 |
| 6.4 Εφαρμογή βαθμονόμησης..... | 45 |
| 6.5 Σύνοψη..... | 46 |
| Κεφάλαιο 7 Τοποθέτηση Εικονικού Αντικειμένου | 48 |
| 7.1 Εισαγωγή | 48 |
| 7.2 Ομογραφία (Homography) | 49 |

| | |
|---|-----------|
| 7.3 Εικονική Επαύξηση σκηνής..... | 52 |
| 7.4 Σύνοψη..... | 61 |
| Κεφάλαιο 8 Υπολογισμός νέας θέσης αντικειμένου μέσω της Οπτικής Ροής (Optical Flow)..... | 63 |
| 8.1 Εισαγωγή | 63 |
| 8.2 Οπτική ροή..... | 64 |
| 8.3 Οπτική ροή για επιτάχυνση της παρακολούθησης (tracking) | 67 |
| 8.4 Πειραματικά αποτελέσματα Οπτικής Ροής | 68 |
| 8.5 Σύνοψη..... | 72 |
| Κεφάλαιο 9 Συμπεράσματα και Μελλοντικές Επεκτάσεις | 74 |
| 9.1 Συμπεράσματα | 74 |
| 9.2 Μελλοντικές Επεκτάσεις | 76 |
| Βιβλιογραφία | 78 |

Κεφάλαιο 1

Εισαγωγή

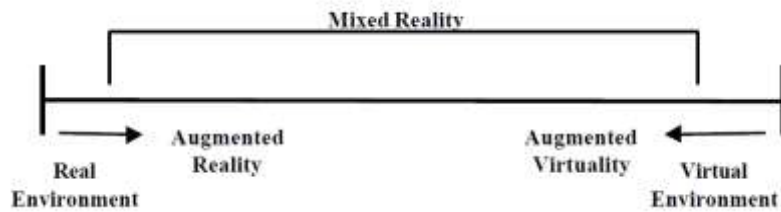
| | |
|-----------------------------------|---|
| 1.1 Περιγραφή του συστήματος..... | 1 |
| 1.2 Πεδίο μελέτης..... | 4 |

1.1 Περιγραφή του συστήματος

Τα τελευταία χρόνια παρατηρείται μεγάλο ενδιαφέρον στη μελέτη νέων ή αναθεωρημένων τρόπων διεξαγωγής διαφόρων καθημερινών δραστηριοτήτων μέσα σε νέα επίπεδα πραγματικότητας. Από τον πραγματικό κόσμο, που αντιλαμβανόμαστε με όλες τις αισθήσεις μας, μέχρι την εικονική πραγματικότητα, όπου μιλάμε για ένα συνθετικό (τεχνητό) κόσμο εμπνευσμένο μέσα από την ανθρώπινη φαντασία και κατασκευασμένο μέσω υπολογιστικών συστημάτων, υπάρχει ένα μεγάλο χάσμα. Μέσα από αυτό το χάσμα αναδύεται μια καινούρια έννοια, ένα νέο είδος πραγματικότητας, η Μικτή Πραγματικότητα (Mixed Reality).

Μικτή Πραγματικότητα είναι μια ενδιάμεση κατάσταση αυτού που αντιλαμβανόμαστε ως πραγματικότητα και της εικονικής πραγματικότητας. Η εργασία αυτή ασχολείται με ένα κόσμο που χαρακτηρίζεται ως Επαυξημένη Πραγματικότητα (Augmented Reality) και είναι ένα μικρότερο κομμάτι αυτού που έχει αναφερθεί νωρίτερα ως Μικτή Πραγματικότητα. Συγκεκριμένα, σε αυτό το επίπεδο πραγματικότητας γίνεται προσθήκη εικονικών στοιχείων στον τρόπο που αντιλαμβανόμαστε τον φυσικό κόσμο, με σκοπό την αύξηση της πληροφορίας που λαμβάνουμε από το περιβάλλον μας. Η

Επαυξημένη Πραγματικότητα μπορεί να βρει εφαρμογές στους τομείς της βιομηχανίας, της ιατρικής, και στα πλαίσια αυτής της εργασίας στους τομείς της μάθησης – εκπαίδευσης καθώς και της ψυχαγωγίας.



Σχήμα 1.1 Θεωρώντας ότι οι δυο ακραίες καταστάσεις που μπορεί ένας άνθρωπος να βιώσει είναι η πραγματικότητα και η εικονική πραγματικότητα, τότε οποιαδήποτε ενδιάμεση κατάσταση μπορεί να θεωρηθεί ως μικτή πραγματικότητα. Αν αυτή η μίξη των πραγματικοτήτων «περιέχει» μεγαλύτερο βαθμό πραγματικότητας και λιγότερο βαθμό εικονικής, τότε μπορεί να θεωρηθεί Επαυξημένη Πραγματικότητα (Augmented Reality), σε αντίθετη περίπτωση μπορεί να θεωρηθεί Επαυξημένη Εικονικότητα (Augmented Virtuality).

Στο Σχήμα 1.1 βλέπουμε την ευθεία οριζόντια γραμμή, όπου στο δεξί και αριστερό άκρο της βρίσκονται ο φυσικός – πραγματικός κόσμος και ο εικονικός κόσμος αντίστοιχα. Ξεκινώντας από το πραγματικό κόσμο και κινούμενοι προς τον εικονικό, εικονικά στοιχεία αρχίζουν να προστίθενται στο περιβάλλον, συνθέτοντας αυτό που αποκαλείται ως Επαυξημένη Πραγματικότητα. Αντίθετα κινούμενοι από τον εικονικό προς τον πραγματικό κόσμο, πραγματικά στοιχεία αρχίζουν να ξεπετάγονται μέσα στο καθόλα αυτό εικονικό περιβάλλον, που θεωρείται πλέον Επαυξημένη Εικονικότητα (Augmented Virtuality). Σε θεωρητικό επίπεδο, στο μέσο αυτής της ευθείας, υπάρχει μια ισορροπία μεταξύ των δύο πραγματικοτήτων όπου το 50% αυτών που κάποιος αντιλαμβάνεται, προέρχονται από την πραγματικότητα, ενώ το άλλο 50% είναι εικονικό.

Στην ανάπτυξη αυτής της νέας επαυξημένης πραγματικότητας καθοριστικό ρόλο παίζει η ανάπτυξη της επιστήμης της πληροφορικής στο ερευνητικό φάσμα της Υπολογιστικής Όρασης (Computer Vision). Η υπολογιστική όραση ασχολείται με τους τρόπους όπου ένα υπολογιστικό σύστημα μπορεί να αντιλαμβάνεται και να εξάγει πληροφορίες οι οποίες προέρχονται μέσα από εικόνες. Πιο απλά να «αντιλαμβάνεται το τη βλέπει», όπως ο ανθρώπινος εγκέφαλος αντιλαμβάνεται τον κόσμο μέσα από τις εικόνες που δέχεται από τα μάτια. Σε κάποιες περιπτώσεις το ενδιαφέρον μπορεί να συγκεντρώνεται στην εξαγωγή πληροφοριών από μια και μόνο εικόνα παραδείγματος

χάριν αναγνώριση ύπαρξης προσώπων μέσα σε μια εικόνα. Ενώ σε άλλες περιπτώσεις το ενδιαφέρον μπορεί να συγκεντρώνεται στη εξαγωγή πληροφοριών από περισσότερες από μία ξεχωριστές εικόνες, παραδείγματος χάριν στην εύρεση ενός μέρους μιας εικόνας μέσα σε μια άλλη, ή ακόμα τη μετακίνηση ενός αντικειμένου μέσα από μια σειρά διαδοχικών εικόνων (βίντεο).

Αυτή η εργασία λοιπόν έχει ως σκοπό τη μελέτη και υλοποίηση ενός συστήματος Επαυξημένης Πραγματικότητας για ανάγνωση παιδικών βιβλίων. Το σύστημα θα αλληλεπιδρά με τον χρήστη μέσω μίας κάμερας, ο χρήστης θα έχει ανοιχτό μπροστά του το βιβλίο ενώ η κάμερα θα είναι τοποθετημένη με τέτοιο τρόπο ώστε να βλέπει την επιφάνεια του βιβλίου. Το σύστημα θα αναγνωρίζει την ανοιχτή σελίδα του βιβλίου και μέσα στην οθόνη θα δημιουργεί μια τρισδιάστατη σκηνή που θα αντιπροσωπεύει το κομμάτι της ιστορίας το οποίο διαδραματίζεται στη σελίδα αυτή, Σχήμα 1.2.



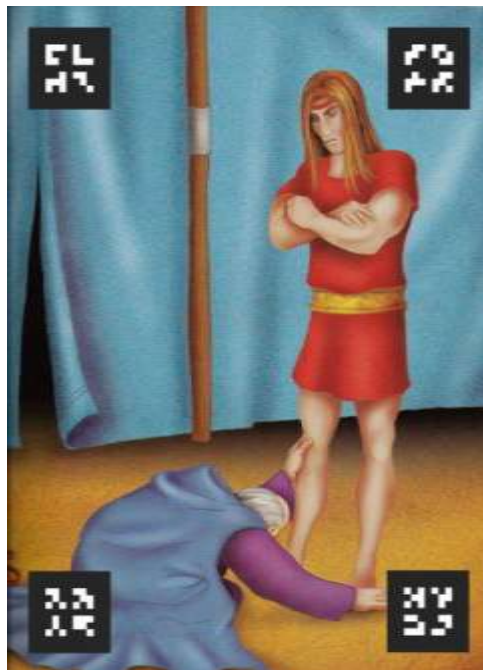
Σχήμα 1.2 Η κάμερα βλέπει το βιβλίο μέσα από τον πραγματικό κόσμο, και στην οθόνη προβάλλεται μια επαυξημένη εκδοχή του πραγματικού κόσμου όπου ένα εικονικό αντικείμενο εμφανίζεται πάνω στη σελίδα του βιβλίου.

Η τρισδιάστατη σκηνή που θα εμφανίζεται θα πρέπει να προσαρμόζεται με τέτοιο τρόπο ώστε να δίνει την εντύπωση στο χρήστη, ότι βρίσκεται επάνω στην επιφάνεια του βιβλίου με αυτό τον τρόπο ο χρήστης θα μπορεί να δει τη σκηνή από διάφορες οπτικές γωνίες, απλά και μόνο με τη μετακίνηση ή περιστροφή του βιβλίου μέσα στο οπτικό πεδίο της κάμερας.

Ένα τέτοιο σύστημα μπορεί να παρουσιάσει ένα νέο τρόπο ανάγνωσης βιβλίων από μικρά παιδιά παρουσιάζοντας τους κάτι πιο δυναμικό από την απλή εικονογράφηση που προσφέρει ένα απλό βιβλίο, χωρίς όμως ταυτόχρονα να την καταργεί. Επίσης, ένα τέτοιο σύστημα δε περιορίζεται μόνο στην απλή εμφάνιση ενός τρισδιάστατου μοντέλου (σκηνής), αλλά μπορεί να προσφέρει τα μέσα για εισαγωγή νέων ιδεών που θα μπορούσαν να κάνουν την ανάγνωση ενός βιβλίου από ένα παιδί πιο ενδιαφέρων και πιο δημιουργική, παραδείγματος χάριν με τη χρήση ηχητικών εφέ.

1.2 Πεδίο μελέτης

Το πεδίο μελέτης της εργασίας αυτής θα εστιάσει κυρίως σε τρόπους και τεχνικές με τις οποίες μπορεί να γίνει αναγνώριση των σελίδων ενός βιβλίου από τον υπολογιστή μέσα στη εικόνα που θα παίρνει από τη κάμερα καθώς και τοποθέτηση τρισδιάστατων μοντέλων πάνω στην εικόνα με τέτοιο τρόπο ώστε να δίνεται η εντύπωση ότι το μοντέλο βρίσκεται πράγματι πάνω στο βιβλίο.



Σχήμα 1.3 Παράδειγμα σελίδας βιβλίου με markers (αναγνωριστικά σημάδια) στα τέσσερα άκρα της σελίδας.

Για την αναγνώριση των σελίδων χρειάζεται να γίνεται αναγνώριση της κάθε ξεχωριστής σελίδας χωρίς τη χρήση ειδικών markers (αναγνωριστικών σημαδιών) στις γωνίες της, Σχήμα 1.3. Δηλαδή το σύστημα, πρέπει να αναγνωρίζει τη κάθε σελίδα ως

ένα αυτόνομο κομμάτι, βάση των διαφόρων χαρακτηριστικών που μπορούν να βρίσκονται διάσπαρτα σε αυτή, σε αντίθεση με τη χρήση markers, όπου γίνεται αρχικά αναγνώριση των τεσσάρων markers και στη συνέχεια ανάλογα με το συνδυασμό τους γίνεται αναγνώριση της κάθε ξεχωριστής σελίδας.

Στη τοποθέτηση των τρισδιάστατων μοντέλων πάνω στην εικόνα θέλουμε το μοντέλο να προσαρμόζεται πάνω στο βιβλίο ανάλογα με το πώς το βλέπει η κάμερα. Με αυτόν τον τρόπο ο χρήστης θα μπορεί για παράδειγμα να δει το μοντέλο από διάφορες πλευρές περιστρέφοντας.

Κεφάλαιο 2

Θεωρητικό Υπόβαθρο

| | |
|---|----|
| 2.1 Εισαγωγή | 6 |
| 2.2 Σύγκριση βασισμένη σε ανίχνευση και ταίριασμα χαρακτηριστικών | 6 |
| 2.3 Ανίχνευση κίνησης..... | 12 |

2.1 Εισαγωγή

Η ανίχνευση ενός αντικειμένου μέσα σε μια εικόνα είναι κάτι που έχει απασχολήσει πολύ τους ερευνητές που ασχολούνται με την υπολογιστική όραση εδώ και αρκετά χρόνια. Σε αυτό το κεφάλαιο θα γίνει μια βασική μελέτη τεχνικών που χρησιμοποιούνται για την επίλυση αυτού του προβλήματος. Επίσης, θα γίνει αναφορά σε τεχνικές που προσφέρουν δυνατότητα αναγνώρισης της μετακίνησης ενός αντικειμένου μέσα στη σκηνή.

2.2 Σύγκριση βασισμένη σε ανίχνευση και ταίριασμα χαρακτηριστικών

Αρχικά για την ανάπτυξη μιας μεθόδου, που θα προσφέρει αναγνώριση ενός αντικειμένου μέσω μιας δοσμένης εικόνας του, μέσα σε μία σκηνή, προϋποθέεται η μελέτη του τι θεωρείται ως αναγνώριση ενός αντικειμένου μέσα σε μία σκηνή. Δηλαδή αν το αντικείμενο βρίσκεται στη σκηνή αντεστραμμένο, με διαφορετικό μέγεθος ή παρουσιάζεται με διαφορετική κατανομή φωτισμού σε σχέση με την αναφορική του εικόνα, τότε το αντικείμενο θεωρείται ότι βρίσκεται μέσα στη σκηνή; Αυτά αποτελούν

τις προδιαγραφές της κάθε τεχνικής. Παράδειγμα αναγνώρισης αντικειμένου μέσα σε σκηνή φαίνεται στο Σχήμα 2.1.



Σχήμα 2.1 Παραδείγματα αναγνώρισης αντικειμένου (αριστερά) μέσα σε σκηνή(δεξιά). Μέσα στη σκηνή το αντικείμενο εμφανίζεται σε διαφορετικό μέγεθος, περιστρεφόμενο και μερικός επικαλυμμένο. Οι κύκλοι στο αντικείμενο αντιπροσωπεύουν χαρακτηριστικά που έχουν ταιριάξει με χαρακτηριστικά στη σκηνή, οι γραμμές υποδεικνύουν την αντιστοιχία μεταξύ των χαρακτηριστικών.

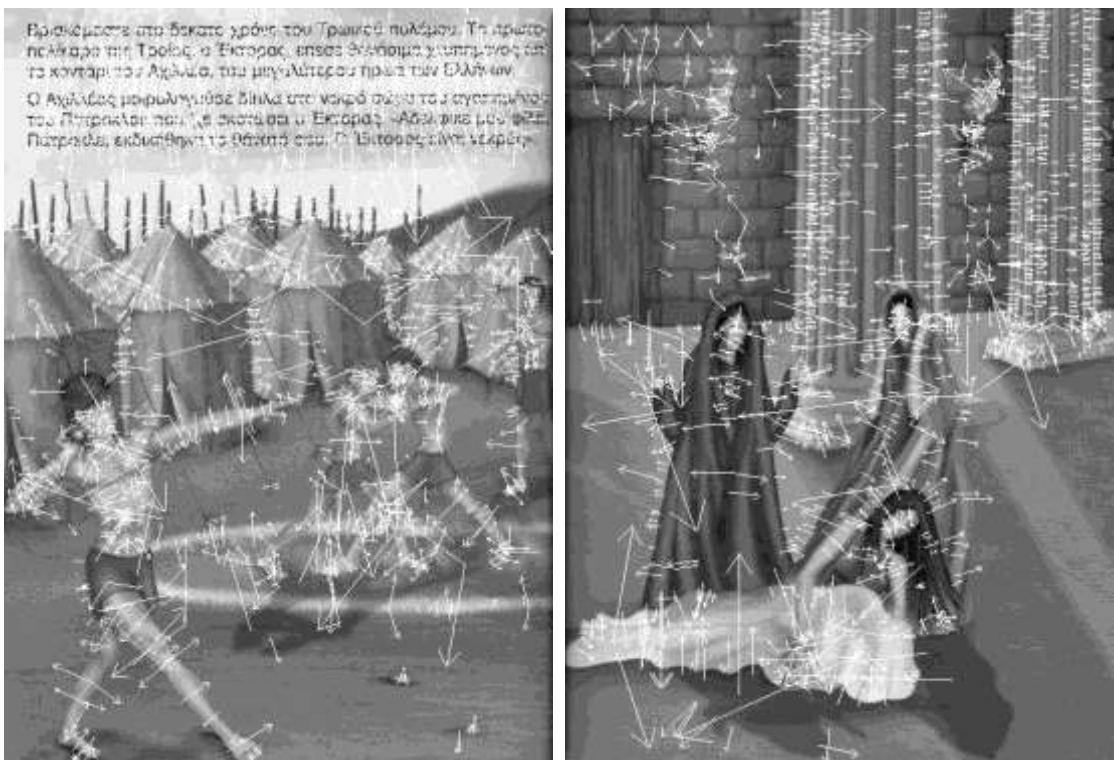
Πιο κάτω θα γίνει αναφορά σε τέτοιες τεχνικές, οι οποίες είναι βασισμένες στην ίδια ιδέα, αλλά η κάθε μια υλοποιείται διαφορετικά. Η βασική ιδέα είναι η ανίχνευση κάποιων χαρακτηριστικών, τα οποία ξεχωρίζουν ιδιαίτερα από οποιαδήποτε άλλα μέσα στην εικόνα του αντικειμένου ή της σκηνής, και στο σύνολο τους προσδιορίζουν τη μορφή της εικόνας. Αφού εντοπιστούν αυτά τα χαρακτηριστικά στη συνέχεια μπορούν να προσδιοριστούν ως μαθηματικά διανύσματα γνωστά ως descriptors.

Η ευκλείδεια απόσταση δύο τέτοιων διανυσμάτων μπορεί να καθορίσει το ποσοστό ομοιότητας των χαρακτηριστικών αυτών. Στο Σχήμα 2.1 παρουσιάζεται παράδειγμα ταύτισης χαρακτηριστικών.

2.2.1 SIFT

Μια από τις πιο γνωστές και σημαντικές μεθόδους που έχουν αναπτυχθεί για το πιο πάνω πρόβλημα είναι η SIFT (Scale-Invariant feature transform) [8]. Η τεχνική SIFT

παρουσιάστηκε το 2004, από τον David G. Lowe ως μια τεχνική για εξαγωγή χαρακτηριστικών, για αξιόπιστη ταύτιση ενός αντικειμένου, μέσα σε εικόνες μιας σκηνης. Τα χαρακτηριστικά που εξάγει δεν εξαρτώνται από τη μεγέθυνση ή τη περιστροφή της εικόνας, δηλαδή το κάθε χαρακτηριστικό μπορεί να βρεθεί και σε αναπαραστάσεις του αντικειμένου σε διαφορετικές κλίμακες καθώς και σε οποιαδήποτε περιστροφή του αντικειμένου. Ακόμη, οι περιγραφές που εξάγονται είναι μέχρι ενός βαθμού ανεξάρτητες από τη φωτεινότητα του αντικειμένου καθώς και από την αλλαγή της θέσης του σημείου παρακολούθησης του. Στο Σχήμα 2.2 παρουσιάζονται σχηματικά παραδείγματα περιγραφής της τεχνικής αυτής.



Σχήμα 2.2 Παραδείγματα περιγραφής σελίδων βιβλίου, με τη χρήση του SIFT. Στις εικόνες τα τόξα προσδιορίζουν την θέση κάθε χαρακτηριστικού (αρχή του τόξου), την κλίμακα που ανιχνεύτηκε (μήκος τόξου) και την κατεύθυνση του. Οι περιγραφές αυτές έχουν γίνει με τη χρήση του SIFT demo (version 4) το οποίο παραχωρεί ο δημιουργός του για ερευνητικούς σκοπούς στη σελίδα: <http://www.cs.ubc.ca/~lowe/keypoints/>.

Για τη περιγραφή μιας εικόνας: αρχικά η τεχνική ψάχνει, μέσα σε διάφορες κλίμακες της, να εντοπίσει, με τη χρήση difference-of-Gaussian, περιοχές ενδιαφέροντος οι οποίες να είναι ανεξάρτητες της κλίμακας (scale) και στις περιστροφές της εικόνας. Στη συνέχεια, μέσα σε αυτές τις περιοχές επιλέγονται τα σημεία που παρουσιάζουν μεγάλη σταθερότητα σχετικά με της μετατροπές που έχουν προαναφερθεί. Σε αυτά τα

σημεία ανατίθενται κατευθύνσεις βάση της υφιστάμενης εικόνας. Τέλος, δημιουργείται ένα ιστόγραμμα με τα τοπικά χαρακτηριστικά της γύρο περιοχής του κάθε σημείου και αποθηκεύεται σε διάνυσμα (vector) 128 διαστάσεων.

Έχοντας ένα σύνολο από χαρακτηριστικά ενός αντικειμένου, και δοσμένης μιας εικόνας που καταγράφει τη σκηνή για την οποία θα αποφασιστεί κατά πόσο το αντικείμενο εμφανίζεται μέσα σε αυτή, αρχικά γίνεται η περιγραφή της σκηνής. Στη συνέχεια, γίνεται ταύτιση των δύο εικόνων μέσω της σύγκρισης κάθε χαρακτηριστικού του αντικειμένου με όλα τα χαρακτηριστικά της σκηνής. Έτσι, για τη σύγκριση κάθε ζεύγους χαρακτηριστικών υπολογίζεται η ευκλείδεια απόσταση των αντίστοιχων διανυσμάτων τους, και τα χαρακτηριστικά με μικρή απόσταση μεταξύ τους θεωρείται ότι ταιριάζουν. Αν ένας μεγάλος αριθμός χαρακτηριστικών ταιριάζει μεταξύ των δύο περιγραφών, τότε θεωρείται ότι το αντικείμενο έχει ανιχνευτεί μέσα στη σκηνή.

Η διαδικασία αυτή, για το ταίριασμα των χαρακτηριστικών, είναι πολύ δαπανηρή έχοντας πολυπλοκότητα της τάξης $O(n^2)$, καθώς όλα τα χαρακτηριστικά του αντικειμένου πρέπει να συγκριθούν με όλα τα χαρακτηριστικά της σκηνής. Επίσης αρκετά δαπανηρή είναι και η κάθε σύγκριση δύο χαρακτηριστικών από μόνη της, καθώς τα διανύσματα έχουν μεγάλο αριθμό διαστάσεων.

2.2.2 PCA-SIFT

Το 2004 προτάθηκε μια παραλλαγή του SIFT, γνωστή ως PCA-SIFT [7], η οποία κάνει χρήση της τεχνικής PCA (Principal Component Analysis) [6]. Η τεχνική PCA έχει την ικανότητα, δεδομένου ενός αριθμού συσχετιζόμενων μεταβλητών, να παράγει ένα σύνολο από νέες μεταβλητές οι οποίες δεν έχουν πλέον εξάρτηση η μια με την άλλη, ενώ το νέο σύνολο περιέχει μικρότερο ή ίσο αριθμό στοιχείων από το αρχικό. Για την ιδιότητα της αυτή η τεχνική χρησιμοποιείται σε πολλές εφαρμογές της υπολογιστικής όρασης.

Έτσι, το PCA-SIFT, κάνοντας χρήση του PCA, καταφέρνει να μειώσει σημαντικά τον αριθμό των διαστάσεων του διανύσματος που περιγράφει ένα σημείο. Πετυχαίνοντας

έτσι μεγάλη βελτίωση στο χρόνο που απαιτείται για την ταύτιση των χαρακτηριστικών, σε σχέση με την αυθεντική εκδοχή του SIFT.

Αν και οι δημιουργοί της τεχνικής αυτής υποστήριζαν ότι μαζί με την μείωση του υπολογιστικού κόστους, δε θα επέφερε και μύωση στις αξιοπιστίας των αποτελεσμάτων της, μετά από πολλές συγκρίσεις αυτό φάνηκε να μην ισχύει.

2.2.3 SURF

Μια άλλη τεχνική, η οποία και χρησιμοποιείται σε αυτή την εργασία, είναι η SURF (Speed Up Robust Features) [1]. Η τεχνική αυτή προτάθηκε το 2006, ως μια υπολογιστικά γρηγορότερη μέθοδος εντοπισμού και περιγραφής χαρακτηριστικών εικόνων σε σχέση με τις, μέχρι τότε, ήδη υπάρχουσες χωρίς να θυσιάζει, βάση των δημιουργών της πάντα, την αξιοπιστία και την εγκυρότητα των χαρακτηριστικών αυτών.

Εκτός από το γρηγορότερο εντοπισμό και περιγραφή των σημείων, η μέθοδος αυτή έρχεται να ανταγωνιστεί τη αντίστοιχη μέθοδο SIFT στο χρόνο ταύτισης, χρησιμοποιώντας μόνο 64 διαστάσεις για την περιγραφή των σημείων.

Η μέθοδος αυτή χρησιμοποιείται στην εργασία για περιγραφή των σελίδων του βιβλίου και για αναγνώριση τους μέσα στη σκηνή. Ο τρόπος λειτουργίας του SURF θα μελετηθεί εκτενέστερα στο Κεφάλαιο 4 της εργασίας αυτής, ενώ στο Κεφάλαιο 5 θα εξηγηθεί η διαδικασία ταύτισης των χαρακτηριστικών του.

2.2.4 Σύγκριση των μεθόδων

Σαφώς, έχοντας διάφορες τεχνικές για την επίλυση του ίδιου προβλήματος, και ειδικότερα όταν αυτό το πρόβλημα είναι ζωτικό σε ένα αναπτυσσόμενο τομέα όπως είναι η υπολογιστική όραση, το ενδιαφέρον για σύγκριση και αξιολόγηση αυτών των τεχνικών είναι μεγάλο.

Ο Mikołajczyk στο [11] παρουσιάζει αποτελέσματα πειραματικών δοκιμών, μεταξύ άλλων, και του SIFT καθώς και της PCA παραλλαγής του. Με ένα σύνολο δοκιμών που περιλάμβανε αλλαγές περιστροφής, κλίμακας, σημείου παρακολούθησης, blurring (θόλωση εικόνας), φωτισμού και συμπίεσης JPEG, καταλήγει με το SIFT να δίνει καλύτερα αποτελέσματα στις περισσότερες δοκιμές σε σχέση με το PCA-SIFT.

Επίσης, στο άρθρο τους [10] οι Luo Juan και Oubong Gwun κάνουν σύγκριση και αξιολόγηση των τριών τεχνικών που έχουν αναφερθεί, μέσα από πρακτικά πειράματα, καταγράφοντας τα δυνατά και αδύναμα σημεία της κάθε μιας. Καταλήγοντας με τον SIFT να έχει ως μεγαλύτερο του μειονέκτημα το ότι είναι πολύ αργός και με βασικές επιδόσεις σε αλλαγές φωτισμού, με μεγάλη σταθερότητα όμως σε μεγέθυνση, περιστροφή, και blurring. Ο SURF αποδεικνύεται ως ο γρηγορότερος και καλύτερος σε αλλαγές φωτισμού, με σχετικά καλή επίδοση σε μεγέθυνση, blurring αλλά με μεγάλη αστάθεια στη περιστροφή. Το PCA-SIFT σε γενικές γραμμές δίνει καλά αποτελέσματα χωρίς να πετυχαίνει όμως άριστα αποτελέσματα σε κάποια από τις δοκιμές. Στο Πίνακα 2.1 παρουσιάζονται τα τελικά αποτελέσματα όλων των δοκιμών.

| Method | Time | Scale | Rotation | Blur | Illumination | Affine |
|----------|--------|--------|----------|--------|--------------|--------|
| SIFT | common | best | best | best | common | good |
| PCA-SIFT | good | common | good | common | good | good |
| SURF | best | good | common | good | best | good |

Πίνακας 2.1 Αποτελέσματα σύγκρισης [10] των τεχνικών SIFT, PCA-SIFT και SURF.

Ενδιαφέρων παρουσιάζει η απόδοση και των τριών στις αλλαγές σημείου παρακολούθησης, όπου το SIFT και το SURF δίνουν αρκετά καλά αποτελέσματα σε περιπτώσεις μικρής μετακίνησης του, ενώ όσο μεγαλώνει η μετακίνηση, η επίδοσή τους μικραίνει μέχρι σε κάποιο σημείο όπου αποτυγχάνουν εντελώς. Ο PCA-SIFT από την άλλη έχει χαμηλά ποσοστά επίδοσης για μικρές αλλαγές, αλλά ο ρυθμός μείωσης του ποσοστού επίδοσης σε σχέση με τη αύξηση της αλλαγής είναι μικρός, με αποτέλεσμα σε πολύ μεγάλες αλλαγές να δίνει καλύτερα αποτελέσματα από τους άλλους δυο.

2.3 Ανίχνευση κίνησης

Ένα άλλο σημείο που έχει απασχολήσει αρκετά τον τομέα της Υπολογιστικής όρασης από την αρχή της ανάπτυξης της είναι η ανίχνευση κίνησης. Δηλαδή έχοντας ένα αντικείμενο σε μια σκηνή, αν αυτό σε μια μεταγενέστερη καταγραφή της σκηνής μετακινηθεί, να εντοπίζεται η νέα του θέση παίρνοντας ταυτόχρονα και πληροφορίες σχετικές με τη κίνηση του. Παραδείγματος χάριν η ταχύτητα της ροής, η οποία αποτελεί μια από τις σημαντικότερες πληροφορίες που μπορούν να εξαχθούν. Οι τεχνικές που έχουν αναπτυχθεί για τη διεκπεραίωση αυτής της ανάγκης ονομάζονται τεχνικές Οπτικής Ροής (Optical Flow). Στο Σχήμα 2.3 αναπαριστάται γραφικά ανίχνευση της οπτικής ροής μεταξύ δύο διαδοχικών εικόνων.

Οι τεχνικές ανίχνευσης της οπτικής ροής μπορούν να διαχωριστούν σε κατηγορίες, βάση των ιδιοτήτων τους, βάση του τρόπου που εντοπίζουν την κίνηση μέσα στις εικόνες ή και βάση των χαρακτηριστικών μιας εικόνας των οποίων βασίζονται για τον εντοπισμό της κίνησης.

Οι S.S. Beauchemin και J. L. Barron στο [2] διαχωρίζουν τις διάφορες αυτές τεχνικές στις εξής κατηγορίες: βασισμένες στη διαφορά έντασης (intensity), πολλαπλών περιορισμών, βασισμένες στη συχνότητα (frequency), πολλαπλής κίνησης, και temporal refinement μέθοδοι. Ομολογώντας όμως ότι λόγω της πολυπλοκότητας όλων αυτών των τεχνικών, τα όρια της κάθε κατηγορίας δε μπορούν να είναι ξεκάθαρα, έχοντας τεχνικές που θα μπορούσαν να ανήκουν σε περισσότερες από μία κατηγορίες.

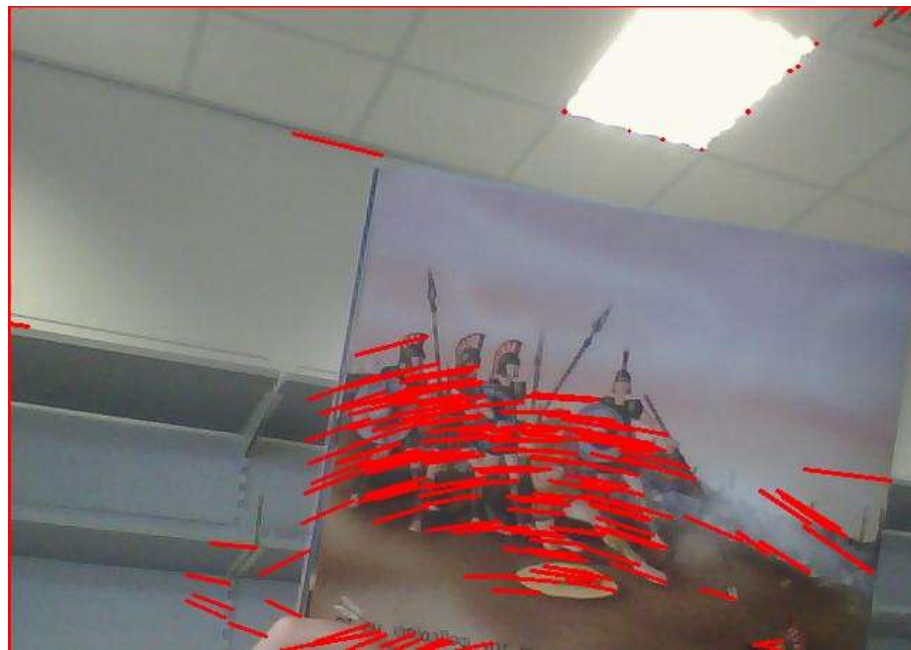
Παρόλο των πολλών τεχνικών που έχουν προταθεί για οπτική ροή, και παρόλο των καλών αποτελεσμάτων που δίνουν μερικές από αυτές, ο υπολογισμός της οπτικής ροής παραμένει ανοιχτό κεφάλαιο στο τομέα της υπολογιστικής όρασης, μιας και καμία τεχνική δε δίνει τόσο ακριβή αποτελέσματα και τα οποία να είναι ανεξάρτητα των διαφόρων συνθηκών που μπορούν να προκύψουν, ώστε να μπορεί να χρησιμοποιηθεί σε εφαρμογές οι οποίες απαιτούν μεγάλη ακρίβεια.



(α)



(β)



(γ)

Σχήμα 2.3 Ανίχνευση Οπτικής ροής. (α) και (β) δύο διαδοχικές εικόνες στις οποίες θα ανιχνευτεί η οπτική ροή. (γ) η δεύτερη διαδοχική εικόνα πάνω στην οποία φαίνονται με κόκκινες γραμμές τα αποτελέσματα της ανίχνευσης της οπτικής ροής, το ένα άκρο (αριστερό) κάθε γραμμής αντιπροσωπεύει τη προηγούμενη θέση του σημείου υπό μελέτη και το άλλο άκρο (δεξί) δείχνει την υφιστάμενη θέση του σημείου.

Μια από της γνωστότερες τεχνικές αποτελεί αυτή των B. D. Lucas και T. Kanade [9], η οποία προτάθηκε το 1981. Ανήκει στη κατηγορία των τεχνικών βασισμένων στη διαφορά της έντασης των εικόνων, είναι σχετικά ανθεκτική στο θόρυβο που μπορεί να παρουσιάζεται σε μία εικόνα. Η τεχνική αυτή υπολογίζει τη ταχύτητα της ροής βάση της διαφοράς δύο διαδοχικών εικόνων και υποθέτει ότι η ροή είναι σταθερή μέσα στη γύρω περιοχή του *pixel* που μελετά, και για αυτό το λόγο δε μπορεί να εντοπίσει ροή σε σημεία που ανήκουν στο εσωτερικό ομοιογενών περιοχών.

Άλλη μέθοδος είναι αυτή των Horn και Schunck [5], είναι εξίσου διαδεδομένη όσο η Lucas and Kanade, είναι και αυτή βασισμένη στη διαφορά μεταξύ των εικόνων μόνο που εδώ η ταχύτητα της ροής υπολογίζεται επαναληπτικά μέσα από μία σειρά εικόνων. Η τεχνική αυτή μπορεί να υπολογίσει μετακίνηση σημείου που βρίσκεται μέσα σε ομοιογενή περιοχή βάσει της μετακίνησης του περιγράμματος της περιοχής αυτής. Η μέθοδος αυτή παρουσιάζει μικρή αδυναμία στο θόρυβο.

Οι δυο αυτές τεχνικές θεωρούνται πλέον κλασικές για υπολογισμό της οπτικής ροής και λόγω των καλών επιδόσεων τους, πολλοί ερευνητές έχουν ασχοληθεί με την επέκταση και την βελτιστοποίηση τους.

Κεφάλαιο 3

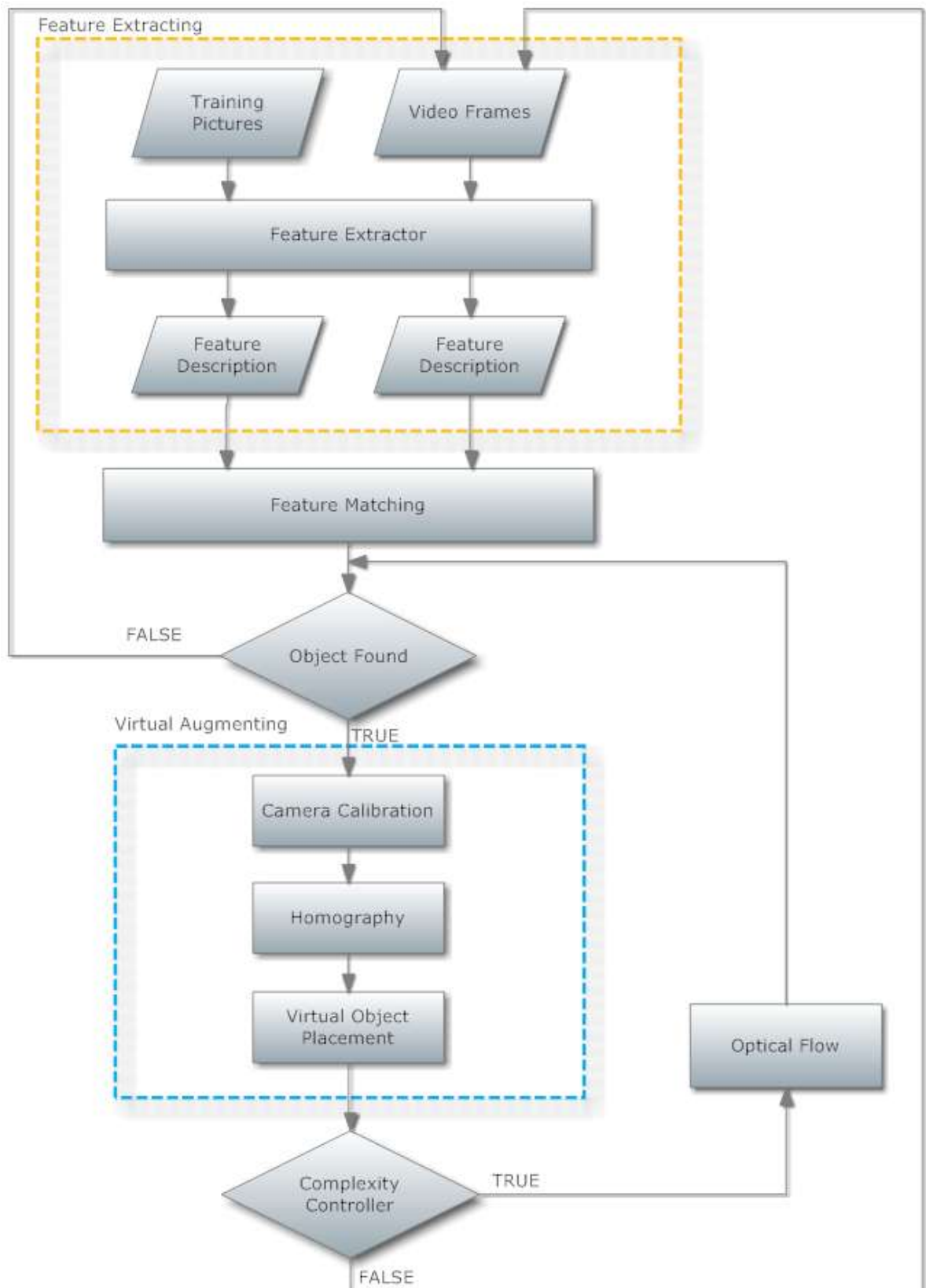
Περιγραφή Συστήματος

| | |
|--|----|
| 3.1 Εισαγωγή | 15 |
| 3.2 Διάγραμμα Τμημάτων του συστήματος..... | 16 |
| 3.3 Περιγραφή Τμημάτων | 17 |

3.1 Εισαγωγή

Σε αυτό το κεφάλαιο θα γίνει διαγραμματική παρουσίαση των βασικών τμημάτων του συστήματος και στη συνέχεια θα ακολουθήσει μια γενική περιγραφή του καθενός από αυτά.

3.2 Διάγραμμα Τμημάτων του συστήματος



3.3 Περιγραφή Τμημάτων

Στο διάγραμμα του Υποκεφαλαίου 3.2 φαίνονται τα βασικά τμήματα του συστήματος: Εξαγωγή χαρακτηριστικών (Feature extracting), Ταύτιση χαρακτηριστικών (Feature matching), Εικονική επαύξηση σκηνής (Virtual Augmenting), Υπολογισμός οπτικής ροής (Optical Flow).

3.3.1 Εξαγωγή χαρακτηριστικών

Εξαγωγή χαρακτηριστικών είναι μια διαδικασία η οποία λαμβάνει μία εικόνα και εξάγει μια περιγραφή αυτής. Η περιγραφή αυτή περιγράφει μοναδικά την εικόνα και επομένως μπορεί να χρησιμοποιηθεί αντιστρόφως για την αναγνώριση της.

Η διαδικασία αρχικά βρίσκει τα χαρακτηριστικά της δοσμένης εικόνας. Αφού τα εντοπίσει, στη συνέχεια υπολογίζει την περιγραφή αυτών των χαρακτηριστικών. Όλες οι περιγραφές των χαρακτηριστικών, της εικόνας, στο σύνολο τους αποτελούν τη γενική περιγραφή της.

Από αυτή τη διαδικασία περνάνε όλες οι εικόνες εκπαίδευσης (Training Picture) καθώς και μερικές από τις εικόνες που λαμβάνονται από τη κάμερα (Video Frames), οι οποίες καταγράφουν τη σκηνή. Εικόνες εκπαίδευσης είναι οι αναφορικές εικόνες των αντικειμένων που πρέπει να ανιχνεύονται στη σκηνή, οι οποίες χρησιμοποιούνται για εκμάθηση της εφαρμογής για να αναγνωρίζει τα αντικείμενα αυτά.

Το ποιες από τις εικόνες που λαμβάνονται από τη κάμερα θα περάσουν από τη διαδικασία αυτή αποφασίζεται από τον Controller. Αν μια εικόνα που λαμβάνεται από τη κάμερα αποφασιστεί να μη περάσει από αυτό το στάδιο, τότε περνάει από τη διαδικασία υπολογισμού της οπτικής ροής, η οποία συζητείται στην Ενότητα 3.3.4 . Ο διαχωρισμός αυτός γίνεται για καλύτερη απόδοση της εφαρμογής, ώστε να μπορεί να τρέχει σε πραγματικό χρόνο, αυτό οφείλεται στο ότι η χρήση της οπτικής ροής μπορεί να μειώσει δραστικά το χρόνο εκτέλεσης της όλης διαδικασίας.

3.3.2 Ταύτιση Χαρακτηριστικών

Η διαδικασία ταύτισης χαρακτηριστικών απαιτεί την εξαγωγή των χαρακτηριστικών των αντικειμένων και της σκηνής. Παίρνοντας ως δεδομένα την περιγραφή ενός αντικειμένου και μίας σκηνής βρίσκει τα χαρακτηριστικά που ταυτίζονται μεταξύ τους, αν υπάρχουν. Αν τα ζεύγη χαρακτηριστικών που έχουν ταυτιστεί είναι αρκετά τότε μπορεί να θεωρηθεί ότι το αντικείμενο έχει εντοπιστεί μέσα στη σκηνή.

Με το πέρας της διαδικασίας αυτής, αν έχει εντοπιστεί το αντικείμενο μέσα στη σκηνή τότε το σύστημα προχωρά στην εικονική επαύξηση της σκηνής. Σε αντίθετη περίπτωση, λαμβάνεται νέα καταγραφή της σκηνής και ακολουθείται από την αρχή η διαδικασία εξαγωγής χαρακτηριστικών και η ταύτιση τους.

3.3.3 Εικονική Επαύξηση σκηνής

Η εικονική επαύξηση της σκηνής χωρίζεται σε δύο κομμάτια. Αρχικά, γίνεται βαθμονόμηση της κάμερας (camera calibration), όπου υπολογίζονται τα χαρακτηριστικά της κάμερας που καταγράφει τη σκηνή. Στη συνέχεια υπολογίζεται η ομογραφία (homography) της σελίδας μέσα στη σκηνή, όπου ουσιαστικά υπολογίζεται μια συσχέτιση κάθε σημείου της αναφορικής εικόνας της σελίδας με το αντίστοιχο σημείο της μέσα στη σκηνή. Τέλος τοποθετείται το εικονικό αντικείμενο.

Από τα χαρακτηριστικά που εξάγονται από τη βαθμονόμηση της κάμερας ο πίνακας περιστροφής της χρησιμοποιείται στο σύστημα αυτό για τη σωστή τοποθέτηση του εικονικού αντικειμένου. Ο πίνακας αυτός περιλαμβάνεται στα εξωγενή χαρακτηριστικά της κάμερας και περιγράφει τον τρόπο με τον οποίο η κάμερα είναι στραμμένη προς το αντικείμενο.

Επίσης, για τη διαδικασία αυτή, υπολογίζεται το μέγεθος του εικονικού αντικειμένου και η θέση που θα τοποθετηθεί σε δύο διαστάσεις, με χρήση της ομογραφίας. Με τον υπολογισμό της περιστροφής, του μεγέθους και της θέσης του εικονικού αντικειμένου, μπορεί να γίνει στη συνέχεια και η τοποθέτηση του μέσα στη σκηνή.

3.3.4 Υπολογισμός Οπτικής Ροής

Η οπτική ροή (optical flow), στο σύστημα αυτό, χρησιμοποιείται για παρακολούθηση (tracking) αντικειμένου μέσα στη σκηνή, έτσι ώστε να εντοπίζεται η καταγραφή του αντικειμένου.

Για χρήση της οπτικής ροής στο σύστημα πρέπει πρώτα να έχει εντοπιστεί το αντικείμενο μέσα στη σκηνή, για το προηγούμενο frame που έχει ληφθεί από τη κάμερα. Έτσι, έχοντας τα σημεία που εντοπίστηκαν τα χαρακτηριστικά του αντικειμένου στο προηγούμενο frame, υπολογίζεται η νέα τους θέση μέσα στο νέο frame.

Η χρήση της οπτικής ροής επιτρέπει τη παράκαμψη της διεξαγωγής των διαδικασιών εξαγωγής και ταύτισης χαρακτηριστικών. Αυτό είναι χρήσιμο διότι η διαδικασία ταύτισης χαρακτηριστικών απαιτεί περισσότερο υπολογιστικό χρόνο από ότι ο υπολογισμός της οπτικής ροής.

Με τον υπολογισμό της οπτικής ροής αν εντοπιστεί η νέα θέση των χαρακτηριστικών τότε ακολουθεί κανονικά η εικονική επαύξηση της σκηνής. Σε αντίθετη περίπτωση το νέο frame που λαμβάνεται στέλνεται για εξαγωγή χαρακτηριστικών και ταύτιση.

Κεφάλαιο 4

Εξαγωγή Χαρακτηριστικών

| | |
|-------------------------------------|----|
| 4.1 Εισαγωγή | 20 |
| 4.2 Ανίχνευση Χαρακτηριστικών | 20 |
| 4.3 Περιγραφή Χαρακτηριστικών | 26 |
| 4.4 Σύνοψη | 29 |

4.1 Εισαγωγή

Εξαγωγή χαρακτηριστικών από μία εικόνα, μπορεί να οριστεί ως η διαδικασία κατά την οποία αρχικά ανιχνεύονται σημαντικά χαρακτηριστικά της και στη συνέχεια η περιγραφή τους. Για την διαδικασία αυτή, όπως έχει ήδη αναφερθεί, θα χρησιμοποιηθεί η τεχνική SURF. Στη συνέχεια του κεφαλαίου αυτού θα γίνει ανάλυση του τρόπου υλοποίησης των εν λόγω διαδικασιών μέσα σε αυτή τη τεχνική.

Σημαντικό είναι να σημειωθεί ότι η τεχνική SURF επεξεργάζεται εικόνες σε ένα μονοχρωματικό κανάλι, δηλαδή ασχολείται με την ένταση (φωτεινότητα) κάθε σημείου της εικόνας, όπως αυτή μπορεί να γίνει αντιληπτή μέσα από μετατροπή της εικόνας σε αποχρώσεις του γκριζου.

4.2 Ανίχνευση Χαρακτηριστικών

Για την διαδικασία ανίχνευσης χαρακτηριστικών, ο σκοπός της τεχνικής είναι η εύρεση σημείων-περιοχών, διαφόρων μεγεθών, μέσα στην εικόνα. Έτσι ώστε η περιγραφή της

εικόνας να είναι ανεξάρτητη της κλίμακας (scale invariant) και της περιστροφής (rotation invariant) της εικόνας.

Για την ανίχνευση χαρακτηριστικών το SURF κάνει χρήση ενδιάμεσων εικόνων [13] (Integral images) και Hessian matrix [4] του οποίου οι τιμές υπολογίζονται με τη χρήση box φίλτρων, προσεγγιστικών παραλλαγών των φίλτρων του Gaussian. Τα box φίλτρα σε συνδυασμό με τις ενδιάμεσες εικόνες προσφέρουν πολύ γρήγορο υπολογισμό του Hessian matrix.

4.2.1 Ενδιάμεσες εικόνες

Η ιδέα των ενδιάμεσων εικόνων είναι η δημιουργία μιας εικόνας ίσων διαστάσεων με την επεξεργαζόμενη, στην οποία κάθε pixel θα παίρνει τιμή ίση με το άθροισμα όλων των pixels που βρίσκονται στην ορθογώνια περιοχή μεταξύ του αντίστοιχου pixel και της αρχής των αξόνων της εικόνας. Έτσι αν $E_{\Sigma}(p)$ η τιμή του pixel $p(x,y)$ στην ενδιάμεση εικόνα, και $E(i,j)$ η τιμή της αρχικής εικόνας στο σημείο (i,j) τότε:

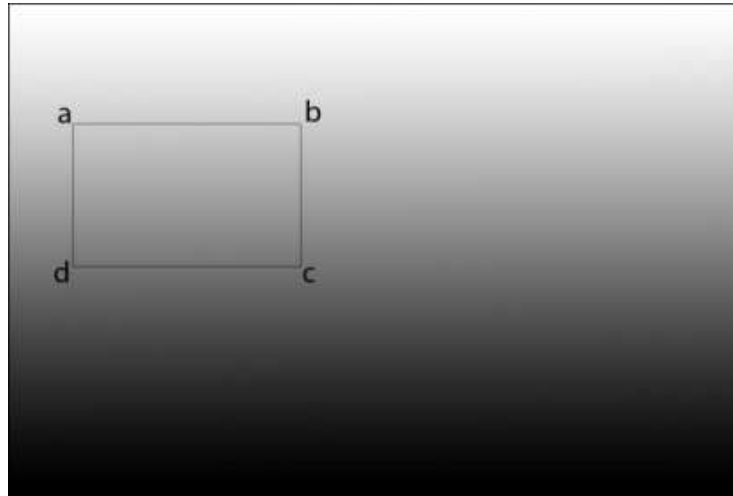
$$E_{\Sigma}(p) = \sum_{i=0}^{i \leq x} \sum_{j=0}^{j \leq y} E(i,j) \quad (4.1)$$

Με τη χρήση της ενδιάμεσης εικόνας ο υπολογισμός του αθροίσματος των τιμών μέσα σε οποιαδήποτε ορθογώνια περιοχή της εικόνας γίνεται σε σταθερό χρόνο, ανεξάρτητο του μεγέθους της, με τρεις μόνο υπολογισμούς, Σχήμα 4.1. Δυο υπολογισμούς για πρόσθεση των τιμών των δυο διαγώνιων ζευγών pixels, της ενδιαφερόμενης περιοχής, και μία αφαίρεση του ενός αθροίσματος από το άλλο, όπως φαίνεται ακολούθως:

$$S(a,b,c,d) = [E(a) + E(c)] - [E(b) + E(d)] \quad (4.2)$$

Όπου $S(a,b,c,d)$ το άθροισμα των εντάσεων μέσα στην ορθογώνια περιοχή που ορίζεται από τα σημεία a,b,c,d και $E(p)$ η ένταση στο σημείο p .

Η συνέλιξη των box φίλτρων, που χρησιμοποιεί η τεχνική, με την εικόνα, μπορεί να υπολογιστεί πολύ γρήγορα με τη βοήθεια της ενδιάμεσης εικόνας.



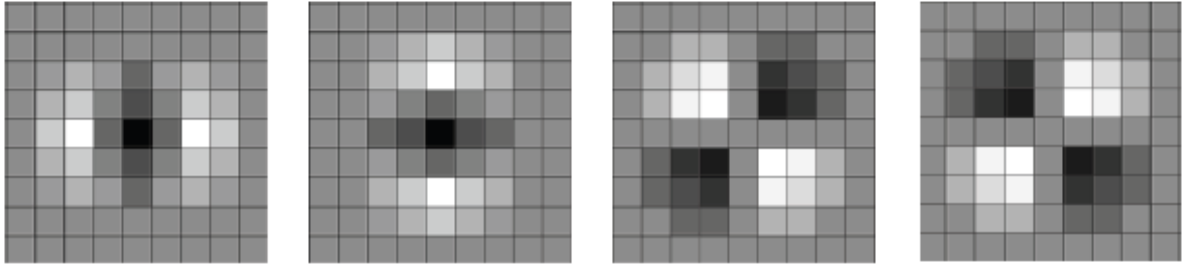
Σχήμα 4.1 Ο υπολογισμός της συνολικής έντασης μέσα στο παράθυρο abcd στην ενδιάμεση εικόνα μπορεί να υπολογιστεί απευθείας βάση της έντασης στα τέσσερα γωνιακά σημεία με τον υπολογισμό $(a+c)-(b+d)$.

4.2.2 Hessian matrix και συνέλιξη εικόνας με φίλτρα

Όπως έχει ήδη αναφερθεί, γίνεται χρήση Hessian matrix, ο οποίος είναι ένας τετραγωνικός πίνακας, σε αυτή τη περίπτωση μεγέθους 2×2 λόγω της δισδιάστατης φύσης των εικόνων. Έτσι, ο πίνακας μπορεί να υπολογιστεί, για κάθε pixel, ως συνάρτηση της θέσης του pixel $p(x,y)$ και της κλίμακας σ .

$$H(p, \sigma) = \begin{bmatrix} L_{xx}(p, \sigma) & L_{xy}(p, \sigma) \\ L_{yx}(p, \sigma) & L_{yy}(p, \sigma) \end{bmatrix} \quad (4.3)$$

Ο συμβολισμός $L_{xx}(p, \sigma)$ αντιπροσωπεύει τη συνέλιξη του pixel p με το φίλτρο της δεύτερης παραγώγου του Gaussian $\frac{d^2 g(\sigma)}{dx^2}$. Τα τέσσερα αυτά φίλτρα έχουν μέγεθος ανάλογο της κλίμακας σ τους, στο Σχήμα 4.2 παρουσιάζονται τα 9×9 φίλτρα που αντιστοιχούν στη κλίμακα $\sigma=1.2$.



Σχήμα 4.2 Φίλτρα δεύτερης παραγώγου του Gaussian. Από αριστερά προς δεξιά: κατεύθυνσης x, y, xy και yx

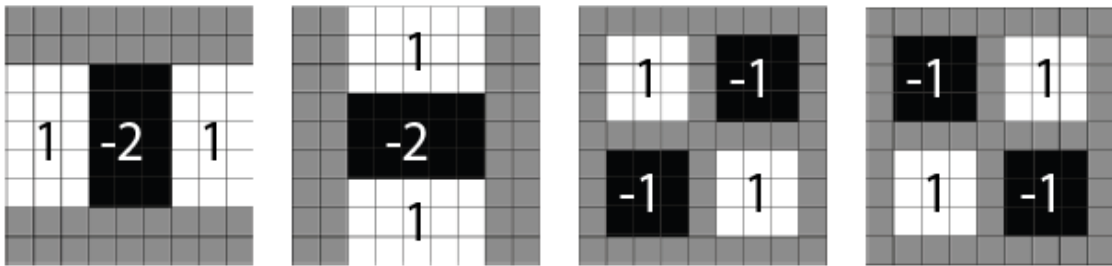
Η συνέλιξη των φίλτρων με το σημείο p της εικόνας γίνεται με τον εξής τρόπο. Το κεντρικό pixel του φίλτρου εφαρμόζει στο pixel p της εικόνας, και η τιμή της συνέλιξης υπολογίζεται ως το άθροισμα των γινομένων των τιμών των αντίστοιχων pixel. Αν $C(x,y)$ η συνάρτηση που επιστρέφει τη τιμή της συνέλιξης ενός φίλτρου και μιας εικόνας στο σημείο x,y , $E(i,j)$ η τιμή του pixel στη θέση i,j , της εικόνας και $f(0,0)$ το κεντρικό pixel του φίλτρου με μέγεθος 9×9 τότε η συνέλιξη υπολογίζεται ως εξής:

$$C(x, y) = \sum_{i=-4}^{i \leq 4} \sum_{j=-4}^{j \leq 4} p(x + i, y + j) f(i, j) \quad (4.4)$$

Αυτή η διαδικασία για ένα φίλτρο διαστάσεων 9×9 χρειάζεται τουλάχιστον 81 προσβάσεις στους πίνακες των εικόνων, ενώ όσο μεγαλώνει το φίλτρο ο αριθμός των προσβάσεων αυξάνεται εκθετικά. Για αυτό το λόγο ο SURF χρησιμοποιεί τέσσερα άλλα φίλτρα γνωστά ως box filters, που αποτελούν προσεγγίσεις των πρώτων, τα φίλτρα αυτά φαίνονται στο Σχήμα 4.3. Η χρήση αυτών των φίλτρων σε συνδυασμό με την ενδιάμεση εικόνα, απαιτεί σημαντικά μικρότερο αριθμό υπολογισμών σε σχέση με τα αρχικά φίλτρα. Επίσης, σημαντικό να αναφερθεί είναι πώς ο χρόνος υπολογισμού της συνέλιξης με τα νέα φίλτρα είναι ανεξάρτητος του μεγέθους των φίλτρων, η ιδιότητα αυτή θα αποδειχθεί πολύ χρήσιμη κατά τη μελέτη της ανεξαρτησίας της περιγραφής στη κλίμακα.

Ο γρήγορος και ανεξάρτητος υπολογισμός της συνέλιξης της εικόνας, που προσφέρουν τα box φίλτρα, έρχεται με το τίμημα του αφού τα εν λόγω φίλτρα αποτελούν προσεγγιστικές παραλλαγές των φίλτρων του Gaussian, προκαλούν μικρή απώλεια ακρίβειας στα αποτελέσματα που δίνουν. Παρόλα αυτά όμως με τη χρήση τους

πετυχαίνεται μεγάλη βελτίωση σε χρόνο υπολογισμού, η οποία αξίζει περισσότερο της μικρής απώλειας ακρίβειας που δημιουργείται.



Σχήμα 4.3 Box φίλτρα. Από αριστερά προς δεξιά: κατεύθυνσης x, y, xy και yx

Αφού υπολογιστεί ο Hessian matrix του pixel για μια κλίμακα, τότε υπολογίζεται η ορίζουσα του πίνακα. Η τιμή αυτής αποτελεί το κριτήριο για το αν το εν λόγω pixel στην εν λόγω κλίμακα αποτελεί σημαντικό χαρακτηριστικό για περιγραφή ή όχι. Η διαδικασία αυτή θα μελετηθεί στη συνέχεια του κεφαλαίου.

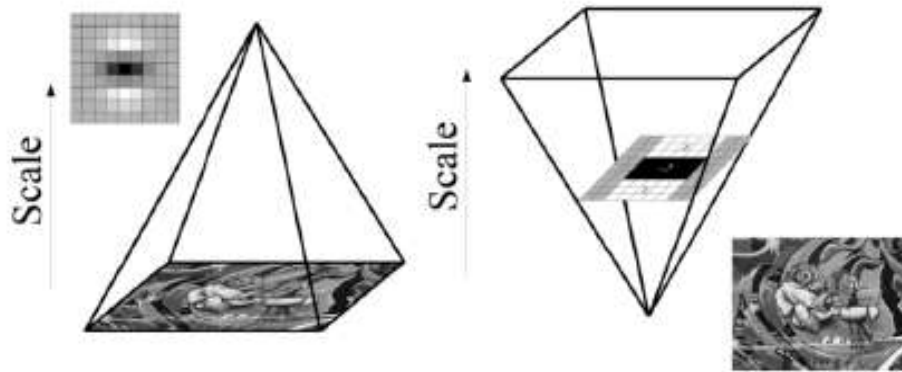
4.2.3 Ανεξαρτησία κλίμακας

Γενικά στο Computer Vision για ανεξαρτητοποίηση της κλίμακας, χρησιμοποιείται μια τεχνική κατά την οποία κατασκευάζεται μια πυραμίδα εικόνων που αντιπροσωπεύουν την εικόνα σε διάφορες κλίμακες της. Η εικόνα υποβάλλεται σε μία διαδικασία λείανσης (smoothing) και στη συνέχεια γίνεται αναπαραγωγή της σε μικρότερη κλίμακα και η διαδικασία της συνέλιξης επαναλαμβάνεται σε διάφορες κλίμακες της εικόνας.

Το SURF κάνει χρήση της ίδιας προσέγγισης με τη διαφορά ότι εδώ αντί να δουλεύει με διάφορες κλίμακες της εικόνας, δουλεύει με μια πυραμίδα από μεγεθύνσεις των φίλτρων, Σχήμα 4.4. Έτσι, αποφεύγεται ο υπολογισμός των σμικρύνσεων της εικόνας μειώνοντας και πάλι το χρόνο που απαιτείται για τον υπολογισμό.

Συγκεκριμένα, η τεχνική δουλεύει σε οκτάβες, όπου κάθε οκτάβα αντιπροσωπεύει μια σειρά από εφαρμογές φίλτρων στην εικόνα. Στη κάθε οκτάβα εφαρμόζονται 4 κλίμακες του κάθε φίλτρου. Για κάθε κλίμακα η αύξηση του φίλτρου είναι σταθερή και ανάλογη με την οκτάβα. Η πρώτη οκτάβα ξεκινά με διαστάσεις φίλτρου 9x9 και συνεχίζει με

τρεις αυξήσεις των φίλτρων κατά 6 pixels και στις δύο διαστάσεις του, 15x15, 21x21 και 27x27. Για κάθε επόμενη οκτάβα η αύξηση διπλασιάζεται, από 6 σε 12 για τη δεύτερη, σε 24, σε 48. Έχοντας έτσι φίλτρα διαστάσεων 15, 27, 39 και 51 για τη δεύτερη οκτάβα, 27, 51, 75 και 99 για την τρίτη και 51, 99, 147, 195 για την τέταρτη.



Σχήμα 4.4 Αναπαράσταση της τεχνικής της πυραμίδας [1]. Στα αριστερά παρουσιάζεται η γενική τεχνική όπου το φίλτρο μένει σταθερό και γίνεται συνέλιξη με την εικόνα σε διάφορες κλίμακες. Δεξιά η παραλλαγή της τεχνικής όπου η συνέλιξη υπολογίζεται με σταθερό μέγεθος της εικόνας και φίλτρο σε διάφορες κλίμακες.

Μέσα σε κάθε οκτάβα γίνεται επιλογή των pixels που θα περιγραφούν, όπως θα περιγραφεί στη συνέχεια. Επίσης, η χρήση περισσότερων των τεσσάρων οκτάβων, πειραματικά έχει αποδειχθεί αχρείαστη, μιας και σε κάθε οκτάβα ο αριθμός των χαρακτηριστικών που ανιχνεύονται μειώνεται δραστικά.

4.2.4 Επιλογή στοιχείων για περιγραφή

Η επιλογή των pixel που θα περιγραφούν, σε μια συγκεκριμένη κλίμακα, καθορίζεται από μια διαδικασία δυο βημάτων, όπου σταδιακά μέσα από κάθε βήμα μειώνεται ο αριθμός των υποψήφιων pixel.

Αρχικά το πρώτο ξεκαθάρισμα γίνεται με τη χρήση ενός threshold, έτσι όποιο pixel κατέχει τιμή μικρότερη από το threshold αγνοείται. Αυτό το σημείο είναι πολύ βασικό καθώς μπορεί να καθορίσει το πλήθος των στοιχείων που τελικά θα επιλεγούν για να περιγραφεί, μεγαλύτερο threshold συνεπάγεται και λιγότερα στοιχεία.

Στη συνέχεια, εφαρμόζεται non-maximum suppression, σε περιοχή $3 \times 3 \times 3$, πάνω στα pixels που έχουν απομείνει. Συγκεκριμένα, κάθε pixel συγκρίνεται με τα 26 γειτονικά του pixels, 8 pixels που το περιτριγυρίζουν στη κλίμακα που έχει μελετηθεί και 9 pixels στην επόμενη και προηγούμενη του κλίμακα. Το pixel επιλέγεται αν έχει τιμή μεγαλύτερη από όλα τα υπόλοιπα με τα οποία έχει συγκριθεί.

4.3 Περιγραφή Χαρακτηριστικών

Η περιγραφή των χαρακτηριστικών, που έχουν ανιχνευτεί όπως έχει περιγραφεί στο Υποκεφάλαιο 4.2, χωρίζεται σε δύο βασικά βήματα. Αρχικά, υπολογίζεται η κατεύθυνση των χαρακτηριστικών και στη συνέχεια υπολογίζεται μια περιγραφή της γύρω περιοχής του σημείου σε σχέση με την κατεύθυνση του. Όπου στο τέλος της διαδικασίας θα δημιουργήσουν ένα διάνυσμα 64 διαστάσεων το οποίο θα περιγράφει το εν λόγω σημείο.

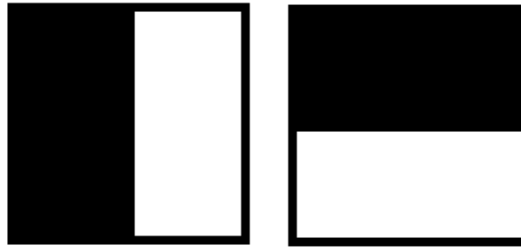
Μαζί με το διάνυσμα περιγραφής αυτό στη μέθοδο SURF φυλάσσεται και μία δυαδική παράμετρος η οποία δείχνει την αντίθεση φωτεινότητας του σημείου σε σχέση με την γύρω περιοχή του. Περιγράφοντας έτσι αν το σημείο αυτό είναι φωτεινό μέσα σε σκοτεινό φόντο ή το αντίθετο. Αυτό αν και μπορεί να φαίνεται ως ασήμαντη λεπτομέρεια, μπορεί να συμβάλει στη γρηγορότερη ταύτιση των χαρακτηριστικών όπως θα εξηγηθεί αργότερα στο Κεφάλαιο 5.

4.3.1 Υπολογισμός κατεύθυνσης

Βάση της ανάγκης για περιγραφή, η οποία να είναι ανεξάρτητη στη περιστροφή της εικόνας, για κάθε χαρακτηριστικό, που έχει ανιχνευτεί υπολογίζεται η κατεύθυνση του. Το SURF υποστηρίζει επίσης και περιγραφή μη ανεξάρτητη στη περιστροφή (U-SURF), για περιπτώσεις που είναι γνωστό από την αρχή ότι δε υπάρξει η ανάγκη για ανίχνευση αντικειμένων τα οποία να έχουν περιστραφεί. Για τέτοια περίπτωση η κατεύθυνση του αντικειμένου δεν υπολογίζεται.

Για τις περιπτώσεις στις οποίες είναι αναγκαία η ανεξαρτησία στη περιστροφή τότε, η κατεύθυνση πρέπει να υπολογιστεί με τέτοιο τρόπο έτσι ώστε να μπορεί να

επαναυπολογιστεί κάτω από διάφορες συνθήκες. Βάση αυτής της κατεύθυνσης θα γίνει στη συνέχεια η περιγραφή του χαρακτηριστικού.



Σχήμα 4.5 Haar Wavelet φίλτρα. Στα δεξιά, φίλτρο για τον x άξονα και αριστερά, για τον y άξονα. Η μαύρη περιοχή κάθε φίλτρου αντιστοιχεί σε βάρος 1 και -1 για την άσπρη περιοχή.

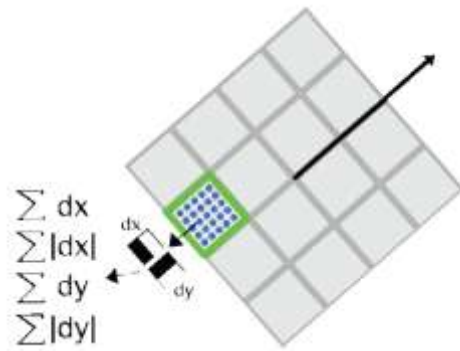
Για την εύρεση της κατεύθυνσης, υπολογίζονται τα Haar wavelet, Σχήμα 4.5, στις κατευθύνσεις των αξόνων x και y , για ένα σύνολο σημείων που βρίσκονται σε περιοχή ακτίνας 6σ γύρω από το pixel, όπου σ η κλίμακα στην οποία το pixel έχει επιλεγθεί. Τα σημεία του συνόλου αυτού δειγματοληπτικά επιλέγονται, μέσα από την περιοχή, με συχνότητα μεγέθους σ . Στη συνέχεια, γίνεται κανονική κατανομή (GAUSS) βαρών μεγέθους 2.5σ , στα αποτελέσματα των Haar wavelet, και αναπαριστούνται αυτά, βάση των βαρών τους, ως σημεία στο χώρο γύρω από το αρχικό pixel. Για την επιλογή του τελικού προσανατολισμού, ένα κυκλικό τμήμα γωνίας $\pi/3$ περιστρέφεται γύρω από το pixel και κάθε φορά υπολογίζεται το διάνυσμα που δημιουργείται από τα σημεία που έχουν υπολογιστεί. Το μεγαλύτερο διάνυσμα που δημιουργείται επιλέγεται ως το διάνυσμα που θα καθορίζει τη κατεύθυνση του εν λόγω σημείου.

4.3.2 Περιγραφή γειτονικής περιοχής

Αφού υπολογιστούν οι κατευθύνσεις των χαρακτηριστικών, τότε δημιουργείται ένα ορθογώνιο παράθυρο περιγραφής γύρω από το pixel, μεγέθους 20σ , όπου και πάλι το σ αντιστοιχεί στη κλίμακα που έχει βρεθεί το pixel. Επίσης, το παράθυρο τοποθετείται προσανατολισμένο στη κατεύθυνση του χαρακτηριστικού, έτσι ώστε όλοι οι μετέπειτα υπολογισμοί για την περιγραφή να είναι συσχετισμένη με την κατεύθυνση του. Στο Σχήμα 4.6 (α) παρουσιάζονται τα παράθυρα περιγραφής που έχουν υπολογιστεί για μια εικόνα. Στη συνέχεια, το παράθυρο χωρίζεται σε 16 (4×4) ίσες περιοχές όπου μέσα σε κάθε μια από αυτές, υπολογίζονται τα Haar wavelets, μεγέθους 2σ , 25 δειγματικών σημείων, κανονικής κατανομής, Σχήμα 4.6 (β).

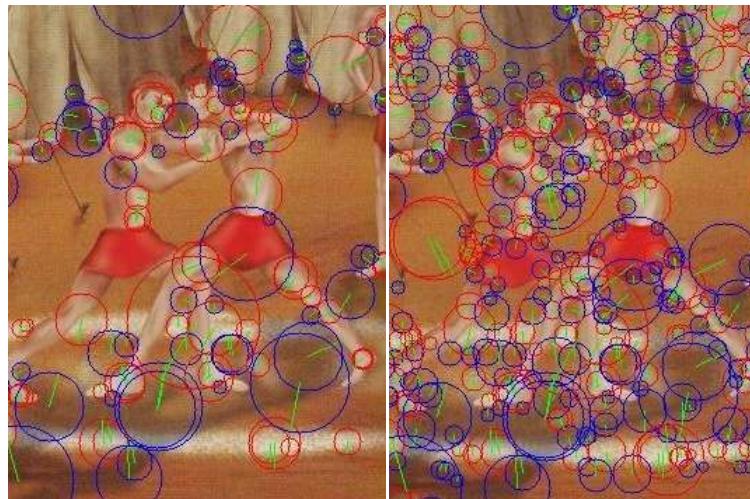


(α)



(β)

Σχήμα 4.6 Παράθυρα περιγραφής [1]. (α) μία εικόνα στο στάδιο της περιγραφής των χαρακτηριστικών. Τα τετράγωνα αντιπροσωπεύουν τα παράθυρα περιγραφής και οι πράσινες γραμμές μέσα σε αυτά δείχνουν τη κατεύθυνση του χαρακτηριστικού. (β) το παράθυρο περιγραφής χωρισμένο σε 4x4 περιοχές κατά τη διαδικασία υπολογισμού των Haar wavelets των 25 δειγματικών σημείων του.



(α)

(β)

Σχήμα 4.7 Αποτελέσματα περιγραφής του SURF με διαφορετικό threshold. (α) περιγραφή με threshold 0.0009 και (β) με 0.0001. Μικρότερο threshold επιτρέπει σε περισσότερα σημεία να επιλεγθούν για να περιγραφούν. Οι κύκλοι αντιπροσωπεύουν χαρακτηριστικά που έχουν περιγραφεί, και το μέγεθος του κάθε κύκλου είναι αντίστοιχο της κλίμακας που έχει ανιχνευτεί. Οι κόκκινοι κύκλοι υποδηλώνουν φωτεινά χαρακτηριστικά σε σκοτεινό φόντο και οι μπλε κύκλοι το αντίθετο.

Για κάθε μια από αυτές τις περιοχές του παραθύρου χρειάζεται να υπολογιστεί το άθροισμα των Haar wavelets για τον x άξονα, και τον y άξονα καθώς και τα αθροίσματα των απόλυτων τιμών τους στο x και y άξονα αντίστοιχα. Καταλήγοντας με τέσσερις τιμές για κάθε περιοχή και 16 περιοχές για κάθε παράθυρο περιγραφής.

Με 64 συνολικά τιμές για κάθε παράθυρο δημιουργείται το διάνυσμα 64 διαστάσεων το οποίο περιγράφει το κάθε χαρακτηριστικό. Με το πέρας της διαδικασίας αυτής ολοκληρώνεται η περιγραφή της εικόνας. Αποτελέσματα της διαδικασίας αυτής παρουσιάζονται σχηματικά, για διαφορετικά thresholds στο Σχήμα 4.7 και για διαφορετικό πλήθος οκτάβων στο Σχήμα 4.8.

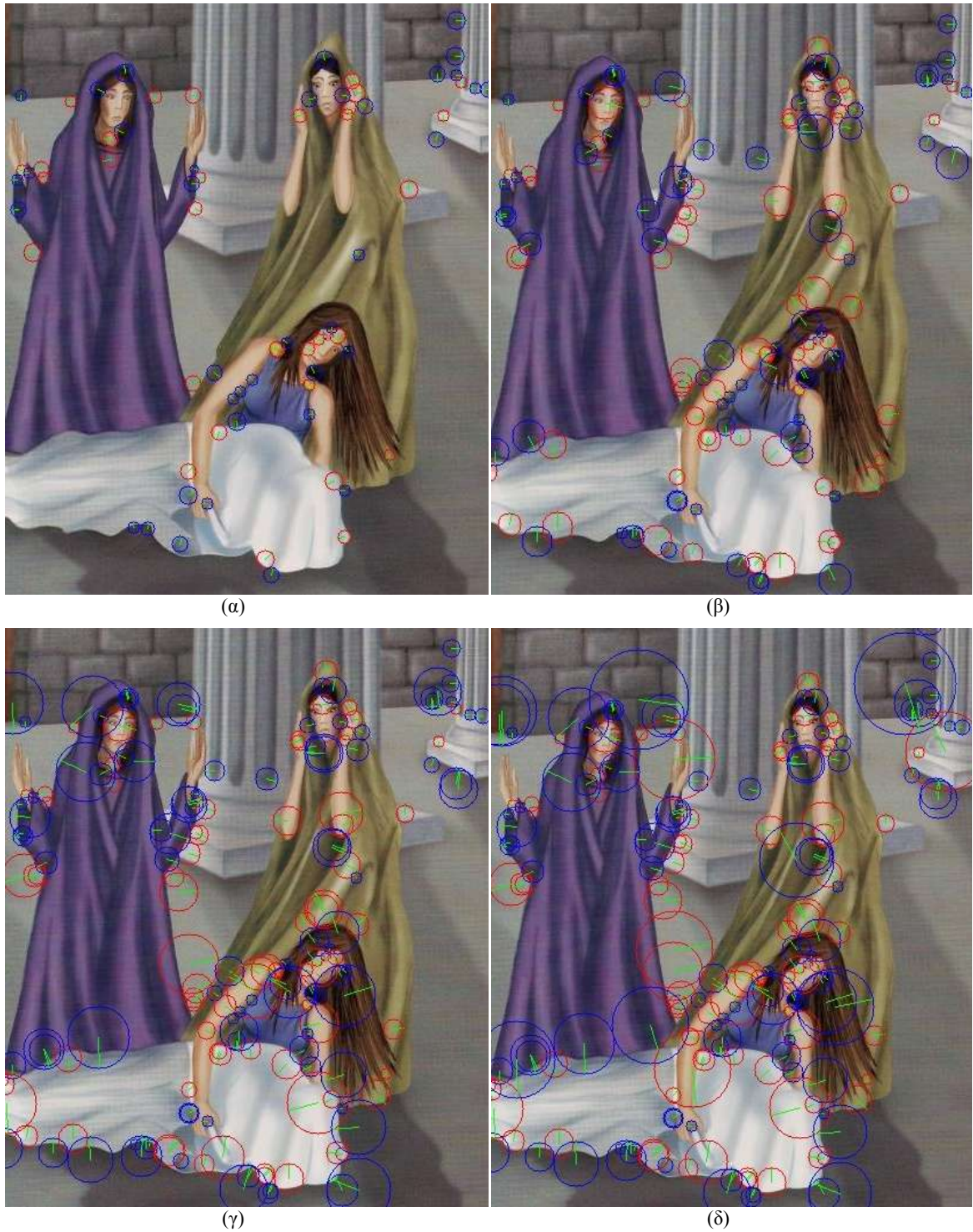
4.4 Σύνοψη

Στο κεφάλαιο αυτό επεξηγήθηκε με λεπτομέρεια το πώς διεξάγονται οι διαδικασίες εντοπισμού και περιγραφής χαρακτηριστικών μιας εικόνας, όπως αυτές υλοποιούνται στη μέθοδο SURF. Οι δύο αυτές διαδικασίες αποτελούν την εξαγωγή των χαρακτηριστικών μιας εικόνας.

Ψάχνοντας αρχικά για αυτά τα χαρακτηριστικά μέσα στην εικόνα, υπολογίζεται η ενδιάμεση εικόνα. Στη συνέχεια γίνεται η συνέλιξη της ενδιάμεσης εικόνας με διάφορα μεγέθη box φίλτρων. Ακολούθως, γίνεται η αρχική επιλογή σημείων με χρήση threshold βάση της τιμής των pixels που έχουν πάρει μετά τη συνέλιξη τους με τα φίλτρα. Η τελική επιλογή των σημείων που θα αποτελούν τα χαρακτηριστικά της εικόνας γίνεται με εφαρμογή non-maximum-suppression πάνω στη σημεία τα οποία έχουν απομείνει.

Με την αναγνώριση αυτών των σημείων ξεκινά η διαδικασία περιγραφής τους. Η περιγραφή πρέπει να γίνει με τέτοιο τρόπο έτσι ώστε να είναι ανεξάρτητη από διάφορες αλλαγές που μπορεί να προκύψουν την εικόνα, παραδείγματος χάριν στη περιστροφή. Για την περιγραφή αυτή υπολογίζεται η κατεύθυνση του χαρακτηριστικού και γίνεται μια περιγραφή μιας μικρής τετραγωνικής περιοχής γύρο από αυτό. Επίσης, η περιγραφή του σημείου περιλαμβάνει ένα στοιχείο το οποίο δηλώνει την αντίθεση φωτεινότητας του σημείου σε σχέση με την περιβαλλόμενη περιοχή του.

Η διαδικασία της εξαγωγής των χαρακτηριστικών έχει ως αποτέλεσμα της, την δημιουργία της περιγραφής της εικόνας. Η οποία είναι απαραίτητη για την ανίχνευση του αντικειμένου μέσα στη σκηνή.



Σχήμα 4.8 Αποτελέσματα περιγραφής του SURF με διαφορετικό αριθμό οκτάβων. (α) περιγραφή με χρήση μόνο μιας οκτάβας, (β) με δύο οκτάβες, (γ) με τρις και (δ) με τέσσερις. Σε κάθε οκτάβα τα χαρακτηριστικά που επιλέγονται, επιλέγονται για μεγαλύτερη κλίμακα επιτρέποντας την ανεξαρτησία στη κλίμακα.

Κεφάλαιο 5

Ταύτιση Χαρακτηριστικών

| | |
|--|----|
| 5.1 Εισαγωγή | 31 |
| 5.2 Διαδικασία ταιριάσματος χαρακτηριστικών..... | 31 |
| 5.3 Σύνοψη..... | 35 |

5.1 Εισαγωγή

Το κεφάλαιο αυτό θα ασχοληθεί με τη διαδικασία ταύτισης χαρακτηριστικών. Συγκεκριμένα, θα εξηγηθεί ο τρόπος με τον οποίο γίνεται η σύγκριση των περιγραφών δυο εικόνων, καθώς και ο τρόπος σύγκρισης δυο χαρακτηριστικών.

Στη διαδικασία αυτή συγκρίνονται η περιγραφή του αντικειμένου και η περιγραφή της σκηνής. Κάθε χαρακτηριστικό της μίας ελέγχεται με τα χαρακτηριστικά της άλλης με σκοπό να βρεθεί κάποιο με το οποίο να ταυτίζεται, δηλαδή να έχουν μεγάλη ομοιότητα.

Με το πέρας της διαδικασίας αυτής θα έχουν βρεθεί όλα τα ζεύγη χαρακτηριστικών που ταυτίζονται. Αν το πλήθος των ζευγών που θα εντοπιστούν είναι μεγαλύτερο από ένα threshold τότε μπορεί να θεωρηθεί ότι το αντικείμενο εντοπίζεται στη σκηνή.

5.2 Διαδικασία ταιριάσματος χαρακτηριστικών

Για να εντοπιστεί ένα αντικείμενο μέσα σε μία σκηνή και αφού έχουν ήδη υπολογιστεί οι περιγραφές και των δύο, όπως έχει μελετηθεί στο Κεφάλαιο 4, αυτό που μένει είναι η

σύγκριση των δύο περιγραφών. Η κάθε περιγραφή, αντικειμένου και σκηνης, αποτελείται από ένα σύνολο διανυσμάτων 64 διαστάσεων, όπου το κάθε διάνυσμα περιγράφει ένα χαρακτηριστικό που έχει ανιχνευτεί.

Η σύγκριση των δύο περιγραφών είναι μια πολύ απλή διαδικασία, όπου το κάθε διάνυσμα χαρακτηριστικού της περιγραφής του αντικειμένου, συγκρίνεται με όλα τα διανύσματα της περιγραφής της σκηνης. Η ανάγκη για σύγκριση του κάθε χαρακτηριστικού της μιας περιγραφής με κάθε χαρακτηριστικό της άλλης συμβάλει στο να γίνεται αυτή η διαδικασία σχετικά αργή.

Για τη σύγκριση των δύο χαρακτηριστικών πρέπει να βρεθεί η απόσταση μεταξύ των δύο διανυσμάτων τους. Πριν όμως υπολογιστεί η απόσταση αυτή γίνεται έλεγχος αν και τα δύο χαρακτηριστικά έχουν την ίδια αντίθεση φωτεινότητας. Η αντίθεση φωτεινότητας του χαρακτηριστικού προσδιορίζει το αν αυτό είναι φωτεινότερο από το χώρο που το περιτριγυρίζει ή σκοτεινότερο. Επομένως δυο χαρακτηριστικά που έχουν διαφορετική αντίθεση δε χρειάζεται να υπολογιστεί η απόσταση των διανυσμάτων τους για να αποφανθεί αν ταυτίζονται, αφού η διαφορά αντίθεσης μπορεί να θεωρηθεί ικανή συνθήκη για τη μη ταύτιση τους.

Η απόσταση μεταξύ των δύο διανυσμάτων που χρησιμοποιείται για τη διαδικασία αυτή είναι η ευκλείδεια απόσταση. Η ευκλείδεια απόσταση αντιπροσωπεύει την απόσταση δύο σημείων μέσα σε ένα n -διάστατο διανυσματικό χώρο. Αν τα δύο σημεία μέσα στο χώρο είναι το x και το y , και x_i και y_i οι θέσεις των σημείων στην i -οστή διάσταση, τότε η ευκλείδεια απόσταση μπορεί να υπολογιστεί ως εξής:

$$\sqrt{(x_1 + y_1)^2 + (x_2 + y_2)^2 + \dots + (x_n + y_n)^2} \quad (5.1)$$

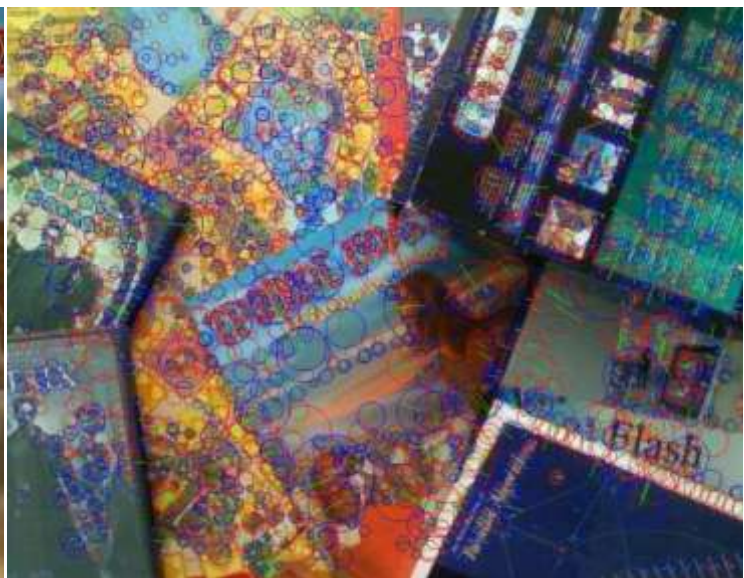
ή

$$\sqrt{\sum_{i=1}^n (x_i + y_i)^2} \quad (5.2)$$

Δεδομένου ότι τα διανύσματα που ελέγχονται σε αυτή τη περίπτωση είναι 64 διαστάσεων, η σύγκριση δύο χαρακτηριστικών απαιτεί αρκετό χρόνο και επομένως συμβάλλει στην περεταίρω αύξηση του χρόνου διεξαγωγής της διαδικασίας ταύτισης ακόμα περισσότερο.



(α)



(β)

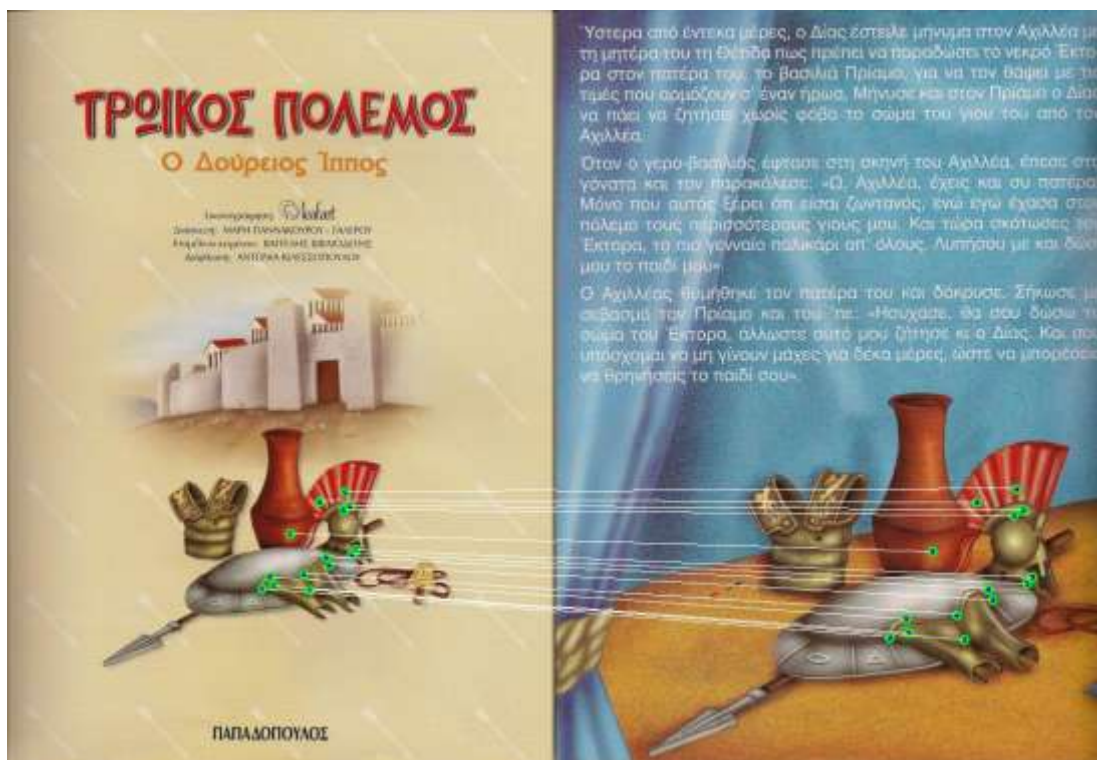


(γ)

Σχήμα 5.1 Ταίριασμα χαρακτηριστικών. Στης (α) και (β) παρουσιάζονται γραφικά οι περιγραφές του αντικειμένου (α) και της σκηνής (β). Στη (γ) παρουσιάζονται τα ζεύγη των χαρακτηριστικών που έχουν ταυριάζει.

Για να ληφθεί η απόφαση αν το αντικείμενο εμφανίζεται στη σκηνή, για κάθε χαρακτηριστικό, στη περιγραφή του αντικειμένου, υπολογίζεται η απόσταση του, με όλα τα χαρακτηριστικά της περιγραφής της σκηνής, και κρατείται η μικρότερη και η δεύτερη μικρότερη απόσταση που έχει ευρεθεί μέχρι στιγμής. Αφού ελεγχθούν όλα τα σημεία της σκηνής με ένα χαρακτηριστικό του αντικειμένου, τότε βάση του λόγου των δύο μικρότερων αποστάσεων, αποφασίζεται αν το χαρακτηριστικό ταιριάζει με αυτό που έχει βρεθεί να έχει τη μικρότερη απόσταση. Συγκεκριμένα, αν ο λόγος των δύο αποστάσεων είναι μικρότερος από 0.65 τότε μπορεί να θεωρηθεί ότι το χαρακτηριστικό του αντικειμένου ταυτίζεται με αυτό που έχει δώσει τη μικρότερη απόσταση από τα χαρακτηριστικά της σκηνής.

Τέλος, αν το πλήθος των χαρακτηριστικών που έχουν ταιριάζει μέσα στη σκηνή είναι πέραν ενός threshold, τότε αποφασίζεται ότι το αντικείμενο εμφανίζεται μέσα στη σκηνή. Τα αποτελέσματα μιας τέτοιας διαδικασίας φαίνονται στο Σχήμα 5.1.



Σχήμα 5.2 Παράδειγμα λανθασμένης ανίχνευσης σελίδας. Η λανθασμένη ανίχνευση στη περίπτωση αυτή οφείλεται στην εμφάνιση ακριβώς του ίδιου σχεδίου και στις δύο σελίδες του βιβλίου. Επομένως, ενώ είναι ορθή η ταύτιση των χαρακτηριστικών γίνεται λανθασμένη ανίχνευση σελίδας.

Σε αυτό το σημείο αξίζει να αναφερθεί ότι βάση του τρόπου που δουλεύει η διαδικασία εξαγωγής χαρακτηριστικών αλλά και η διαδικασία ταιριάσματος μπορούν να παρατηρηθούν προβλήματα λανθασμένης ανίχνευσης. Στο Σχήμα 5.2 παρατηρείται λανθασμένη ανίχνευση της σελίδας. Αυτό όπως είναι προφανές συμβαίνει λόγω της χρήσης του ίδιου σχεδίου και στις δύο σελίδες. Αυτό αφενός οδηγεί σε ορθό ταιρίασμα χαρακτηριστικών, αφού τα χαρακτηριστικά είναι όντως τα ίδια, αφετέρου όμως οδηγεί σε λανθασμένη αναγνώριση σελίδας, εφόσον τα ζεύγη χαρακτηριστικών είναι αρκετά για να θεωρηθεί ως ορθή ανίχνευση.

Αυτό μπορεί να αντιμετωπιστεί επιλέγοντας ως ορθή σελίδα αυτή στην οποία εμφανίζονται τα περισσότερα χαρακτηριστικά που να ταιριάζουν. Ακόμη, μπορεί να αποφευχθεί εφαρμόζοντας κάποιους περιορισμούς κατά τη σχεδίαση του βιβλίου που θα χρησιμοποιηθεί.

5.3 Σύνοψη

Η ταύτιση χαρακτηριστικών είναι η διαδικασία κατά την οποία συγκρίνονται οι περιγραφές της σκηνής και του αντικειμένου. Η σύγκριση των περιγραφών γίνεται με τη σύγκριση κάθε χαρακτηριστικού της μίας με όλα τα χαρακτηριστικά της άλλης, βάση της ευκλείδειας απόστασης τους. Με τον τρόπο αυτό βρίσκονται όλα τα ζεύγη χαρακτηριστικών που ταυτίζονται, αν το πλήθος τους ξεπερνά ένα threshold αυτό μπορεί να οδηγήσει στην αναγνώριση του αντικειμένου μέσα στη σκηνή.

Το μέγεθος του threshold αποτελεί σημαντική παράμετρο στην ακρίβεια των αποτελεσμάτων. Για παράδειγμα, σε αρκετές περιπτώσεις η αναγνώριση τεσσάρων με πέντε ζευγαριών, χαρακτηριστικών που ταυτίζονται, είναι αρκετή ώστε να θεωρηθεί ότι το αντικείμενο εμφανίζεται στη σκηνή. Παρόλα αυτά τέσσερα με πέντε ζεύγη δεν αποτελούν αρκετή πληροφορία για την αναγνώριση της ακριβούς θέσης του αντικειμένου μέσα στη σκηνή.

Η διαδικασία αποτελεί την πιο αργή διαδικασία της εφαρμογής, αφού απαιτεί αρκετό υπολογιστικό χρόνο, έχοντας πολυπλοκότητα της τάξης του $O(n^2)$. Επίσης, εκτελείται για κάθε frame που λαμβάνεται από τη κάμερα προκαλώντας στο πρόγραμμα μεγάλη

καθυστέρηση μέχρι να ζητηθεί από τη κάμερα η επόμενη καταγραφή της σκηνής. Στο Κεφάλαιο 7 θα συζητηθεί τρόπος παράκαμψης της διαδικασίας αυτής για κάποια frames.

Κεφάλαιο 6

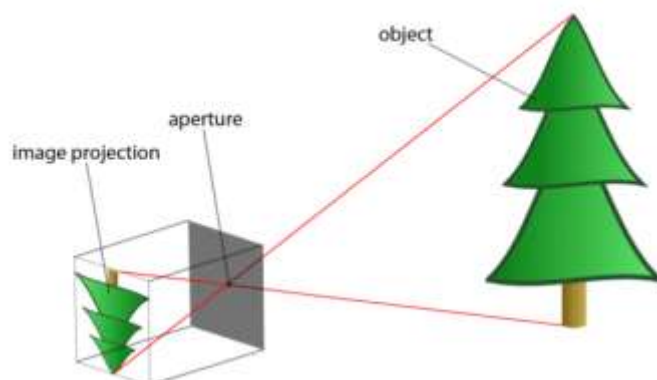
Βαθμονόμηση Κάμερας (Camera Calibration)

| | |
|---|----|
| 6.1 Εισαγωγή | 37 |
| 6.2 Βαθμονόμηση κάμερας | 38 |
| 6.3 Περιγραφή Διαδικασίας διεξαγωγής Geometric Camera Calibration | 42 |
| 6.4 Εφαρμογή βαθμονόμησης | 45 |
| 6.5 Σύνοψη | 46 |

6.1 Εισαγωγή

Μια κάμερα καταγράφει μία φωτογραφία, μέσα από τη σύλληψη φωτός το οποίο περνά από τον φακό της και προβάλλεται πάνω σε ένα ηλεκτρικό αισθητήρα φωτός (photosensitive sensor). Ο αισθητήρας με τη σειρά του μετατρέπει το φώς που λαμβάνει σε ψηφιακή πληροφορία. Έτσι, κάθε pixel αντιστοιχεί σε μια ακτίνα φωτός η οποία έχει περάσει μέσα από το φακό και έχει προβληθεί πάνω στον αισθητήρα. Βάση αυτού η καταγραφή της φωτογραφίας εξαρτάται άμεσα από την κατάσταση του φακού και του αισθητήρα.

Το κεφάλαιο αυτό θα ασχοληθεί με την διαδικασία βαθμονόμησης της κάμερας [3] (camera calibration). Κατά την διαδικασία αυτή υπολογίζονται οι παράμετροι που επηρεάζουν την καταγραφή των φωτογραφιών από μια κάμερα.



Σχήμα 6.1 Λειτουργία κάμερας. Μια κάμερα είναι ένα κλειστό σκοτεινό κουτί, στο οποίο εισέρχεται φως μέσω μιας οπής όπως φαίνεται στο σχήμα. Στην απέναντι μεριά του κουτιού, από αυτή που βρίσκεται η οπή προβάλλεται το είδωλο του αντικειμένου, αντεστραμμένο. Όσο πιο μικρή είναι η οπή τόσο πιο καθαρή είναι η απεικόνιση του ειδώλου, όμως όσο πιο μικρή τόσο λιγότερο το φως που εισέρχεται.

6.2 Βαθμονόμηση κάμερας

Βαθμονόμηση κάμερας είναι μια διαδικασία, κατά την οποία υπολογίζονται οι παράμετροι, που επηρεάζουν την εικόνα, που καταγράφει η κάμερα. Αυτοί οι παράμετροι καθορίζουν το πώς δημιουργείται η εικόνα σε σχέση με τις συνθήκες κατά τις οποίες, καταγράφει η κάμερα της φωτογραφίες, καθώς και κάποιες από της αλλοιώσεις που μπορούν να προκύψουν κατά τη καταγραφή της φωτογραφίας σε σχέση με αυτό που πραγματικά βλέπει η κάμερα.

Η διαδικασία αυτή είναι πολύ σημαντική, όταν οι φωτογραφίες που λαμβάνονται από μια κάμερα, προορίζονται για να εξαχθούν ποσοτικές πληροφορίες σχετικές με τη σκηνή που καταγράφεται. Για αυτό το λόγο η διαδικασία αυτή θεωρείται ως μια από τις πλέον σημαντικές διαδικασίες στην υπολογιστική όραση.

Συγκεκριμένα ο όρος βαθμονόμηση κάμερας μπορεί να αναφερθεί για δύο διαφορετικά είδη βαθμονόμησης, τη γεωμετρική βαθμονόμηση (geometric camera calibration) και τη φωτομετρική (photometric camera calibration). Η χρησιμότητα του κάθε είδους εξαρτάται από της ανάγκες που προκύπτουν, καθώς στη γεωμετρική βαθμονόμηση υπολογίζονται οι παράμετροι που καθορίζουν το πώς καταγράφεται η εικόνα, ενώ στη φωτομετρική βαθμονόμηση υπολογίζονται οι παράμετροι που καθορίζουν το πώς

καταγράφεται ο φωτισμός (radiance) στη σκηνή. Σε κάποιες περιπτώσεις υπάρχει η ανάγκη διεξαγωγής μόνο του ενός από τα δύο είδη, ενώ σε άλλες υπάρχει η ανάγκη διεξαγωγής και των δύο.

Στη γεωμετρική βαθμονόμηση της κάμερας υπολογίζονται οι παράμετροι που επηρεάζουν γεωμετρικά την εικόνα που δίνει μια κάμερα, καθώς και πληροφορίες σχετικά με τη τοποθεσία της κάμερας μέσα σε μια δεδομένη σκηνή. Στη φωτομετρική βαθμονόμηση γίνεται μια συσχέτιση της πραγματικής φωτεινότητας που βλέπει η κάμερα μέσα στη σκηνή, με αυτή που αντιλαμβάνεται και τελικά καταγράφει. Στα πλαίσια αυτής της εργασίας χρησιμότητα βρίσκει μόνο η γεωμετρική βαθμονόμηση.

Η παράμετροι που δίνουν οι πιο πάνω διαδικασίες επηρεάζονται από διάφορες συνθήκες κάτω από τις οποίες λειτουργεί η κάμερα, και για αυτό το λόγο η διαδικασία πρέπει να επαναλαμβάνεται όποτε αυτό απαιτείται, παραδείγματος χάριν η μετακίνηση της κάμερας από ένα σταθερό σημείο σε ένα άλλο ή η αλλαγή εστίασης (focus), απαιτεί και την επανεκτέλεση της διαδικασίας της γεωμετρικής βαθμονόμησης, ενώ η αλλαγή του φωτισμού στο χώρο απαιτεί την επανεκτέλεση της γεωμετρικής βαθμονόμησης.

Geometric Camera Calibration

Η διαδικασία της γεωμετρικής βαθμονόμησης κάμερας υπολογίζει παραμέτρους για τη κάμερα οι οποίες μπορούν να χωριστούν σε δύο ομάδες, στα ενδογενή (intrinsic parameters) και στα εξωγενή χαρακτηριστικά (extrinsic parameters) της κάμερας.

Ενδογενή χαρακτηριστικά είναι παράμετροι που σχετίζονται κυρίως με τον φακό της κάμερας, και δίνονται μέσα από ένα 3×3 πίνακα, γνωστό ως πίνακα ενδογενών χαρακτηριστικών ή *k matrix*, και ένα διάνυσμα πέντε διαστάσεων το οποίο περιέχει του συντελεστές στρέβλωσης.

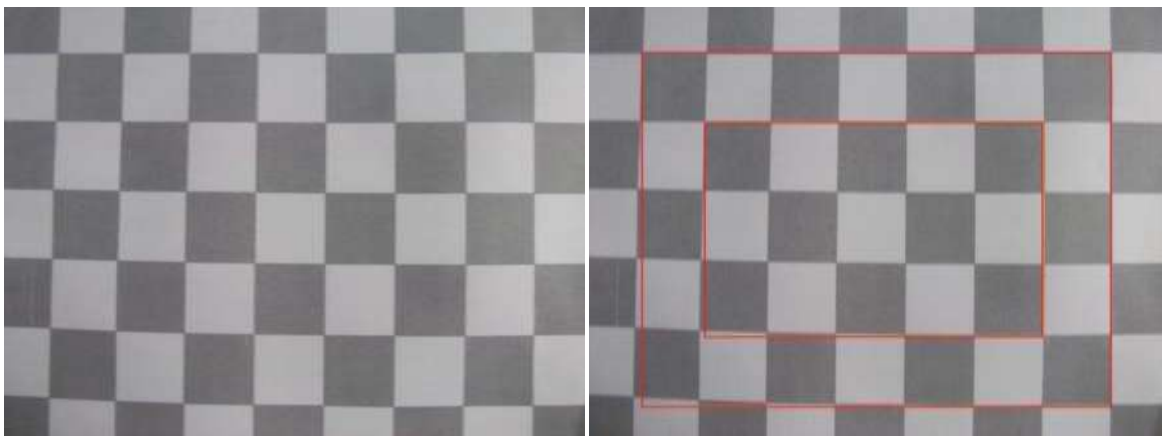
Εξωγενή χαρακτηριστικά σχετίζονται με τη θέση της κάμερας από ένα σημείο της σκηνής το οποίο θεωρείται ως η αρχή του κόσμου, και δίνεται μέσα από ένα πίνακα περιστροφής (rotation matrix) και ένα διάνυσμα μετατόπισης (translation vector).

Ενδογενής παράγοντες

Ο πίνακας ενδογενών χαρακτηριστικών φαίνεται πιο κάτω:

$$k = \begin{pmatrix} f \cdot s_x & f \cdot s\theta & c_x \\ 0 & f \cdot s_y & c_y \\ 0 & 0 & 1 \end{pmatrix} \quad (6.1)$$

Η παράμετρος f , αντιπροσωπεύει την εστιακή απόσταση (focal length) του φακού, η οποία ορίζεται, ως η απόσταση μεταξύ του σημείου εστίασης και του image plane. Τα s_x και s_y είναι παράγοντες κλίμακας (scale factors) οι οποίοι συσχετίζουν τα pixels με την απόσταση στους άξονες x και y αντίστοιχα. Οι παράγοντες κλίμακας είναι διαφορετικοί για τους δύο άξονες στις περιπτώσεις που τα pixels της cameras δεν είναι τετράγωνα. Η παράμετρος $s\theta$ είναι ο συντελεστής λοξότητας και οι παράμετροι c_x και c_y αντιπροσωπεύουν το κέντρο της εικόνας (image center, principal point), το οποίο ιδανικά θα πρέπει να βρίσκεται ακριβώς στο κέντρο της.



(α)

(β)

Σχήμα 6.1 Στρέβλωση εικόνας (image distortion) περίπτωση γνωστή ως barrel distortion . (α) η φωτογραφία ενός μοτίβου σκακιού, όπως έχει καταγραφεί από κάμερα και (β) με κόκκινα παραλληλόγραμμα σχεδιάζεται η κανονική θέση του μοτίβου όπου προς τα κάτω αριστερά παρατηρείται μια μικρή απόκλιση, η οποία προκαλείται από τη στρογγυλότητα του φακού ή άλλα τυχόν προβλήματα του.

Το διάνυσμα πέντε διαστάσεων περιέχει τους συντελεστές στρέβλωσης (distortion coefficients) του φακού, όπου αντιπροσωπεύουν ένα πέμπτου (συνήθως) βαθμού πολυωνυμικό μοντέλο όπου χρησιμοποιούνται μόνο οι περιττές δυνάμεις. Το μοντέλο αυτό αναφέρεται στις γεωμετρικές παραμορφώσεις που συμβαίνουν στη ακτινική και

εφαπτομένη διεύθυνση, και προκαλούνται λόγω της στρογγυλότητας του φακού ή και άλλων βλαβών του, που μπορούν να προκύψουν με τη χρήση της κάμερας. Παράδειγμα στρέβλωσης εικόνας φαίνεται στο Σχήμα 6.1.

Εξωγενής παράγοντες

Εξωγενής παράμετροι ορίζουν τη θέση και την κατεύθυνση του κέντρου της κάμερας σε σχέση με μία δεδομένη θέση στο χώρο η οποία θεωρείται ως το κέντρο του χώρου, προσδιορίζοντας έτσι τη περιστροφή και τη μετατόπιση της κάμερας σε σχέση με αυτό το σημείο. Με τη βοήθεια των παραμέτρων αυτών μπορεί να γίνει μετασχηματισμός συντεταγμένων, από σύστημα συντεταγμένων του χώρου σε σύστημα συντεταγμένων της κάμερας και αντίστροφα. Οι συντεταγμένες ενός σημείου σε σύστημα συντεταγμένων του χώρου, δείχνουν τη θέση του σημείου σε σχέση με το κέντρο του χώρου, και όταν μετατραπούν στο σύστημα συντεταγμένων της κάμερας δείχνουν τη θέση του εν λόγω στοιχείου σε σχέση με το κέντρο της κάμερας.

Οι εξωγενής παράγοντες χωρίζονται σε δύο ομάδες αυτούς που αντιπροσωπεύουν τη περιστροφή και δίνονται μέσα από ένα 3x3 πίνακα, τον πίνακα περιστροφής (rotation matrix) και από ένα διάνυσμα τριών διαστάσεων, το διάνυσμα μετατόπισης (translation vector), όπως φαίνονται πιο κάτω:

$$R = \begin{pmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{pmatrix} \quad (6.2)$$

$$T = \begin{pmatrix} t_x \\ t_y \\ t_z \end{pmatrix} \quad (6.3)$$

Η μετατροπή των συντεταγμένων ενός σημείου, από σύστημα συντεταγμένων χώρου σε σύστημα συντεταγμένων της κάμερας, γίνεται ως εξής:

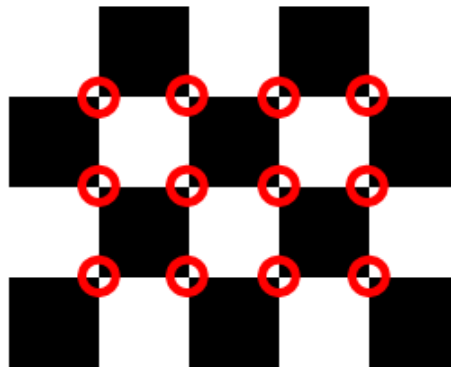
$$p_c = Rp_w + T \quad (6.4)$$

Όπου p_w οι συντεταγμένες (x,y,z) του σημείου p στο χώρο, p_c οι συντεταγμένες (x,y,z) του σημείου σε σχέση με τη κάμερα, και R,T ο πίνακας περιστροφής και το διάνυσμα μετατροπής αντίστοιχα.

6.3 Περιγραφή Διαδικασίας διεξαγωγής Geometric Camera Calibration

Για τη γεωμετρική βαθμονόμηση της κάμερας, απαιτείται η καταγραφή μίας σειράς από φωτογραφίες, στις οποίες η κάμερα βλέπει μια επίπεδη επιφάνεια, σε διαφορετικές θέσεις κάθε φορά. Η επιφάνεια αυτή πρέπει να περιέχει κάποια χαρακτηριστικά, των οποίων η γεωμετρία να είναι ήδη γνωστή. Από την επεξεργασία όλων των φωτογραφιών μπορούν να υπολογιστούν τα χαρακτηριστικά της κάμερας.

Στις περισσότερες περιπτώσεις για τη διαδικασία αυτή χρησιμοποιείται μια επίπεδη επιφάνεια πάνω στην οποία υπάρχει ένα μοτίβο σκακιού. Τα ενδιαφέροντα χαρακτηριστικά του μοτίβου αυτού είναι οι εσωτερικές γωνίες του, όπως φαίνεται στο Σχήμα 6.2. Οι γωνίες αυτές είναι εύκολο να εντοπιστούν λόγω της μεγάλης χρωματικής αντίθεσης του μοτίβου, είναι ευθυγραμμισμένες, και επιπλέον οι αποστάσεις μεταξύ δύο διαδοχικών γωνιών είναι γνωστές και πάντα ίσες.



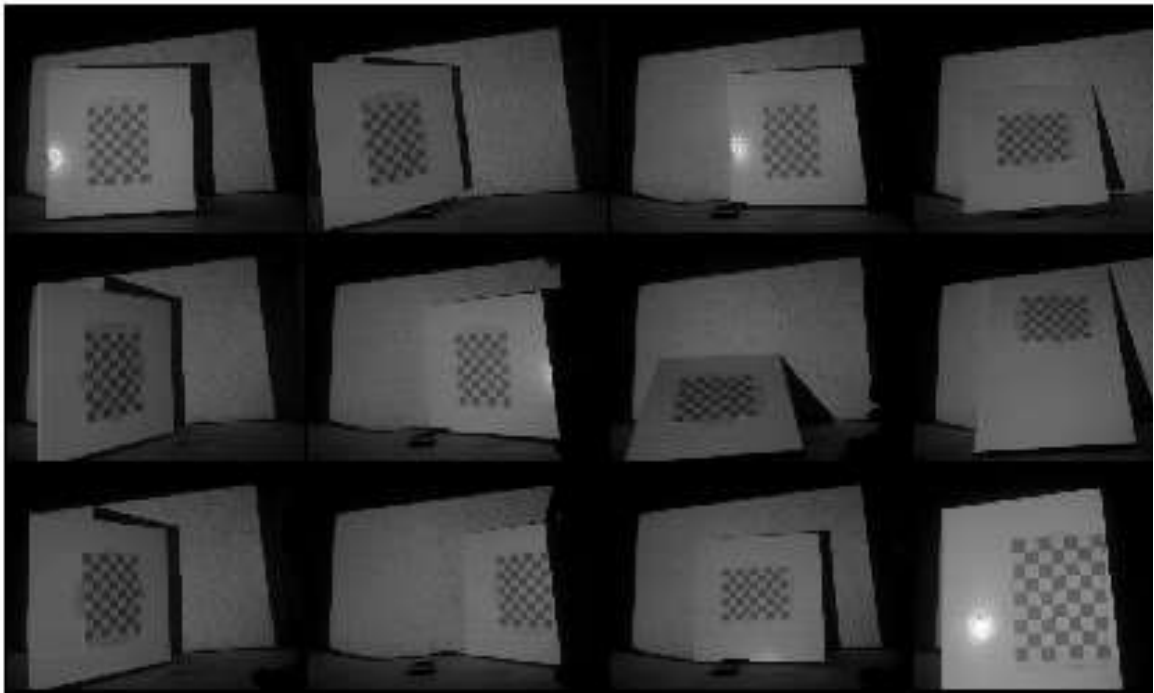
Σχήμα 6.2 Μοτίβο σκακιού που χρησιμοποιείται για γεωμετρική βαθμονόμηση της κάμερας. Με κόκκινους κύκλους είναι σημειωμένες οι εσωτερικές γωνίες του μοτίβου οι οποίες πρέπει να εντοπιστούν μέσα σε κάθε φωτογραφία. Η ανίχνευση των γωνιών είναι σχετικά εύκολη λόγω της μεγάλης αντίθεσης του μοτίβου. Τα τετράγωνα του μοτίβου έχουν όλα το ίδιο μέγεθος, έτσι οι περιγραφές των πιο πάνω χαρακτηριστικών μπορεί αρκεστεί στο οριζόντιο και κάθετο πλήθος των γωνιών και στην απόσταση μεταξύ δύο συνεχόμενων γωνιών.

Η επιφάνεια αυτή φωτογραφίζεται από διάφορες οπτικές γωνίες, έχοντας πάντα σταθερή τη κάμερα και μετακινώντας την επιφάνεια όπως φαίνεται στο Σχήμα 6.3. Για

μεγάλη ακρίβεια στον υπολογισμό των χαρακτηριστικών, συνήθως χρησιμοποιείται μια σειρά από 20 με 30 φωτογραφίες.



(α)

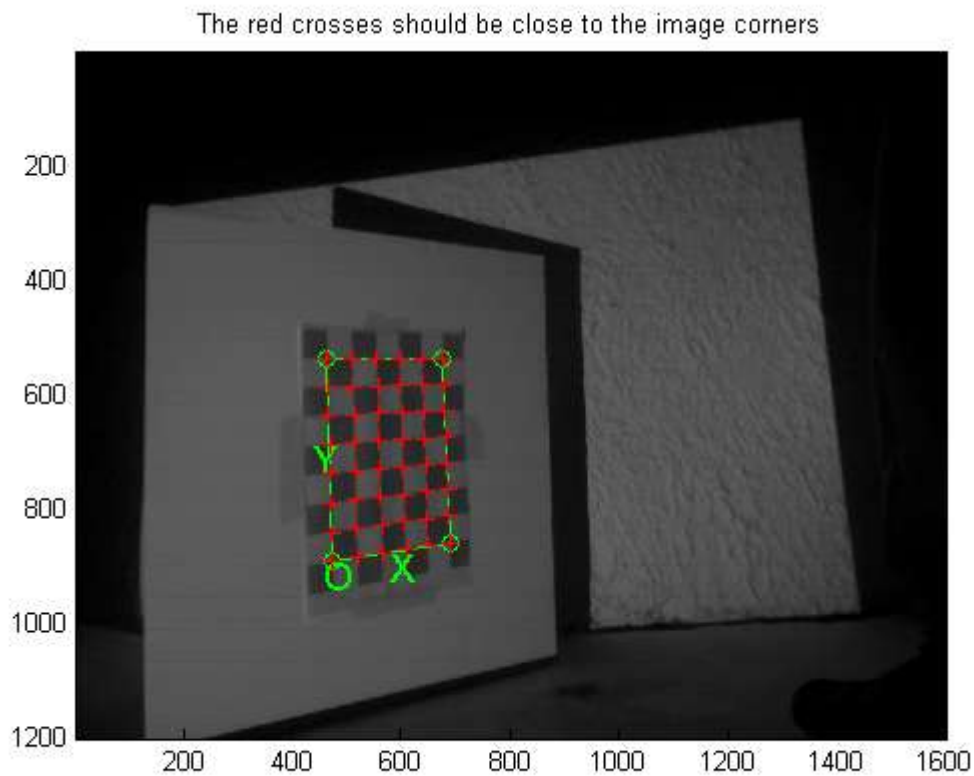


(β)

Σχήμα 6.3 Calibration images. Μια σειρά από φωτογραφίες που έχουν χρησιμοποιηθεί για βαθμονόμηση της κάμερας. (α) Οι τρεις από τις εικόνες του data set σε πραγματικά χρώματα και (β) οι υπόλοιπες εικόνες του data set σε μονοχρωματικό κανάλι, όπως τις μετατρέπει το εργαλείο Camera Calibration Toolbox for Matlab[®] για να τις επεξεργαστεί. Για τη διαδικασία αυτή η κάμερα μένει σταθερή σε ένα σημείο και μετακινείται η επιφάνεια με το μοτίβο σκακιού μέσα στη σκηνή.

Μέσα στις φωτογραφίες που έχουν καταγραφεί, ανιχνεύονται οι εσωτερικές γωνίες του μοτίβου, και κρατούνται οι θέσεις τους για κάθε φωτογραφία, Σχήμα 6.4. Στη συνέχεια βάση των θέσεων των γωνιών αυτών, σε κάθε φωτογραφία και έχοντας ως δεδομένο την πραγματική απόσταση μεταξύ δυο διαδοχικών γωνιών, υπολογίζονται τα ενδογενή και εξωγενή χαρακτηριστικά της κάμερας. Στο Σχήμα 6.5 φαίνεται διάγραμμα σχετικά

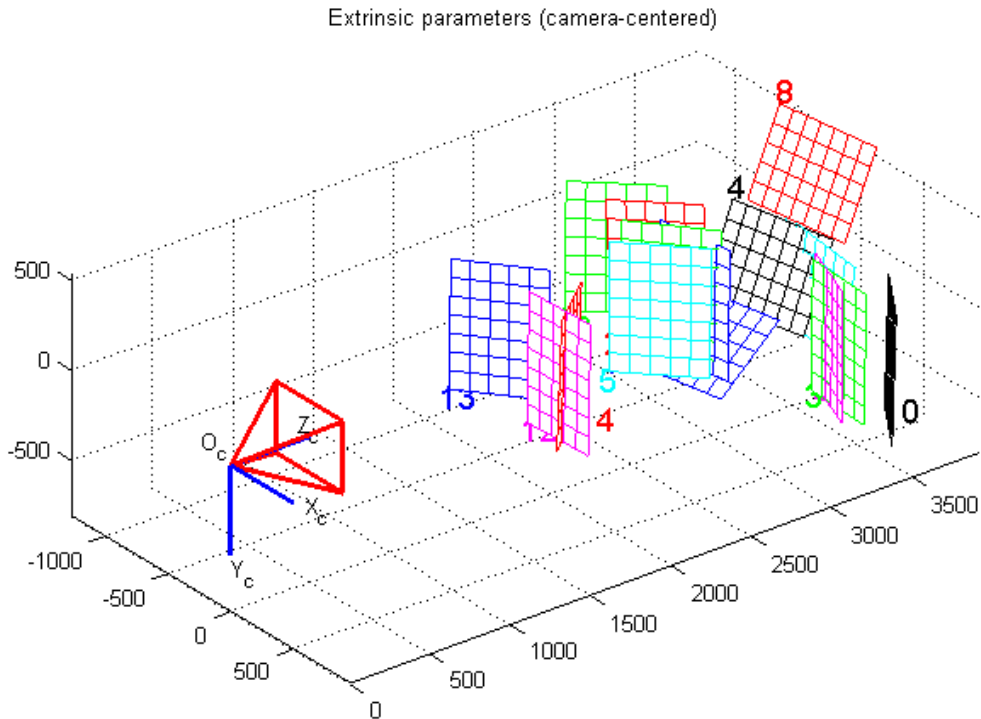
με τις εξωγενή παραμέτρους και στο πίνακα 6.1 φαίνονται οι ενδογενής παράμετροι που έχουν υπολογιστεί για παράδειγμα εκτέλεσης βαθμονόμησης.



Σχήμα 6.4 Αποτελέσματα ανίχνευσης χαρακτηριστικών του μοτίβου μέσα σε φωτογραφία, από το Camera Calibration Toolbox for Matlab[®]. Από το χρήστη ζητείται να κάνει χειροκίνητα την εύρεση των τεσσάρων γωνιακών χαρακτηριστικών, τα οποία φαίνονται με πράσινους κύκλους στην εικόνα, και στη συνέχεια γίνεται αυτόματα η εύρεση όλων των χαρακτηριστικών, τα οποία φαίνονται με κόκκινους σταυρούς.

| | | |
|------------------|-----------|--|
| Focal Length: | fc = | [1498.25041 1505.08169] ± [12.62719 12.51053] |
| Principal point: | cc = | [834.46832 564.58687] ± [14.95720 18.13909] |
| Skew: | alpha_c = | [0.00000] ± [0.00000] |
| Distortion: | kc = | [0.07719 -0.21508 0.00877 -0.00061 0.00000] ± [0.01851 0.05178 0.00420 0.00370 0.00000] |
| Pixel error: | err = | [0.29447 0.25536] |

Πίνακας 6.1 Αποτελέσματα υπολογισμού ενδογενών χαρακτηριστικών από γεωμετρική βαθμονόμηση κάμερας του συνόλου εικόνων από το Σχήμα 6.3, με χρήση του εργαλείου Camera Calibration Toolbox for Matlab[®].



Σχήμα 6.5 Γραφική παρουσίαση εξωγενών χαρακτηριστικών κάμερας για το σύνολο των εικόνων που φαίνονται στο Σχήμα 6.3, από το Camera Calibration Toolbox for Matlab[®]. Η κόκκινη πυραμίδα αντιπροσωπεύει τη κάμερα, μέσα στο πραγματικό κόσμο θεωρώντας ότι το κέντρο της αντιπροσωπεύει το κέντρο του κόσμου. Τα διάφορα χρωμάτων πλέγματα αντιπροσωπεύουν το μοτίβο σκακιού, όπως αυτό εμφανίζεται σε κάθε δεδομένη εικόνα.

6.4 Εφαρμογή βαθμονόμησης

Στην εργασία αυτή γίνεται γεωμετρική βαθμονόμηση της κάμερας, κάθε φορά που με το πέρας της διαδικασίας της ταύτισης των χαρακτηριστικών, Κεφάλαιο 5, βρεθούν αρκετά ζεύγη που έχουν ταυτιστεί. Για να λαμβάνονται όσο το δυνατόν ακριβέστερα αποτελέσματα από τη διαδικασία αυτή, το threshold για το απαιτούμενο πλήθος ζευγών θέτεται ως 15.

Στα πλαίσια της εφαρμογής η οποία απαιτεί επεξεργασία δεδομένων σε πραγματικό χρόνο, η διαδικασία αυτή εκτελείται με μία και μόνο φωτογραφία η οποία αντιστοιχεί στο τρέχων frame που έχει ληφθεί από τη κάμερα. Αν και θεωρητικά θα μπορούσε η διαδικασία αυτή να γίνει μόνο μία φορά κατά την έναρξη της εφαρμογής, με χρήση περισσότερων εικόνων, αυτό πρακτικά δεν είναι εφικτό.

Για να γινόταν αυτή η διαδικασία κατά την έναρξη της εφαρμογής, οι απαιτήσεις που θα είχε η εφαρμογή από το χρήστη θα ήταν περισσότερες, και ίσως να ήταν μεγαλύτερες από αυτές που ένας μέσος χρήστης μπορεί ή είναι διατεθειμένος να προσφέρει για την χρήση της εφαρμογής. Εκτός από τις απαιτήσεις που θα είχε μια τέτοια προσέγγιση από τον χρήστη θα είχε και περιορισμούς, από τη στιγμή που θα ξεκινούσε η εφαρμογή ο χρήστης δε θα μπορούσε να μετακινήσει τη κάμερα. Για τους λόγους αυτούς η διαδικασία γίνεται για κάθε frame στο οποίο έχει εντοπιστεί το βιβλίο, με μόνο μια εικόνα και με χρήση του ίδιου αντικειμένου που ο χρήστης θα χρησιμοποιεί ούτως ή άλλως.

Έτσι τα σημεία που δίνονται στη βαθμονόμηση, αντίστοιχα με τα σημεία του μοτίβου του σκακιού, είναι τα χαρακτηριστικά που έχουν ανιχνευτεί μέσα στο frame καθώς και τα αντίστοιχα σημεία μέσα στη αναφορική εικόνα της σελίδας του βιβλίου. Με λίγα λόγια η σελίδα του βιβλίου λειτουργεί με αντίστοιχο τρόπο που λειτουργεί η επιφάνεια με το μοτίβο σκακιού, όταν διεξάγεται κλασικά η γεωμετρική βαθμονόμηση της κάμερας.

Με αυτό το τρόπο υπολογίζονται τα χαρακτηριστικά της κάμερας, για κάθε frame, από τα οποία χρησιμοποιείται ο πίνακας περιστροφής για τη τοποθέτηση του αντικειμένου, όπως αυτό θα μελετηθεί στο Κεφάλαιο 7.

6.5 Σύνοψη

Βαθμονόμηση κάμερας ονομάζεται μια διαδικασία, η οποία υπολογίζει παραμέτρους οι οποίες επηρεάζουν την καταγραφή των φωτογραφιών μέσω της κάμερας. Υπάρχουν δύο ήδη βαθμονόμησης η γεωμετρική βαθμονόμηση και η φωτομετρική. Στα πλαίσια της εργασίας αυτής χρησιμοποιείται μόνο η γεωμετρική βαθμονόμηση.

Μέσω της γεωμετρικής βαθμονόμησης υπολογίζονται τα γεωμετρικά χαρακτηριστικά της κάμερας, τα οποία αποτελούνται από τα ενδογενή και τα εξωγενή χαρακτηριστικά. Η βαθμονόμηση στην εργασία γίνεται με τη χρήση των ζευγών των χαρακτηριστικών που έχουν ανιχνευτεί κατά τη διαδικασία ταύτισης χαρακτηριστικών. Από τα

αποτελέσματα που λαμβάνονται από τη διαδικασία αυτή χρήση βρίσκουν τα εξωγενή χαρακτηριστικά της κάμερας.

Εξωγενή χαρακτηριστικά είναι τα χαρακτηριστικά τα οποία περιγράφουν τη θέση της κάμερας μέσα στο χώρο, σε σχέση με την σελίδα του βιβλίου. Τα χαρακτηριστικά αυτά θα χρησιμοποιηθούν για τη τοποθέτηση του εικονικού αντικειμένου μέσα στη σκηνή, όπως αυτό θα περιγραφεί στο Κεφάλαιο 7.

Κεφάλαιο 7

Τοποθέτηση Εικονικού Αντικειμένου

| | |
|-----------------------------------|----|
| 7.1 Εισαγωγή | 48 |
| 7.2 Ομογραφία (Homography)..... | 49 |
| 7.3 Εικονική Επαύξηση σκηνής..... | 52 |
| 7.4 Σύνοψη | 61 |

7.1 Εισαγωγή

Με την ανίχνευση μιας σελίδας του βιβλίου μέσα στη σκηνή, το επόμενο στάδιο απαιτεί την τοποθέτηση του εικονικού αντικειμένου μέσα σε αυτή. Το αντικείμενο αυτό πρέπει να τοποθετείται με τέτοιο τρόπο ώστε να δίνεται στον αναγνώστη η εντύπωση ότι, πραγματικά βρίσκεται πάνω στην επιφάνεια του βιβλίου.

Για να γίνει αυτό χρειάζεται να εντοπιστεί η περιοχή που καταλαμβάνει ολόκληρη η επιφάνεια του βιβλίου μέσα στη σκηνή. Αυτό θα βοηθήσει στο να βρεθεί η θέση και το μέγεθος του εικονικού αντικειμένου. Επίσης, χρησιμότητα βρίσκουν και στοιχεία σχετικά με τη περιστροφή της κάμερα που καταγράφει τη σκηνή, όπου βάση αυτών θα γίνει η περιστροφή του αντικειμένου. Για την αναγνώριση της περιοχής που καταλαμβάνει η επιφάνεια του βιβλίου μέσα στη σκηνή θα χρησιμοποιηθεί ομογραφία (homography) ενώ η περιστροφή της κάμερας υπολογίζεται κατά τη βαθμονόμηση της, όπως έχει περιγραφεί στο Κεφάλαιο 6.

Στη συνέχεια του κεφαλαίου αυτού θα επεξηγηθεί ο όρος της ομογραφίας όπως και η χρήση του μέσα στην εργασία. Επίσης, θα σχολιαστεί ο τρόπος με τον οποίο επιτυγχάνεται η σωστή τοποθέτηση του εικονικού αντικειμένου.

7.2 Ομογραφία (Homography)

Ομογραφία είναι ένας όρος της γεωμετρίας, ο οποίος ορίζεται ως η σχέση μεταξύ δύο εικόνων, όπου κάθε στοιχείο της μίας αντιστοιχίζεται με ένα και μόνο στοιχείο της άλλης. Η έννοια αυτή μπορεί να περιγράψει τη προοπτική προβολή μιας σκηνής.

Συγκεκριμένα με τη χρήση της ομογραφίας μπορεί να υπολογιστεί η θέση μέσα στη σκηνή, ενός δεδομένου σημείου της σελίδας. Επομένως, με τον εντοπισμό κάθε γωνιάς της σελίδας μέσα στη σκηνή, μπορεί να αναγνωριστεί και η περιοχή που καταλαμβάνει.

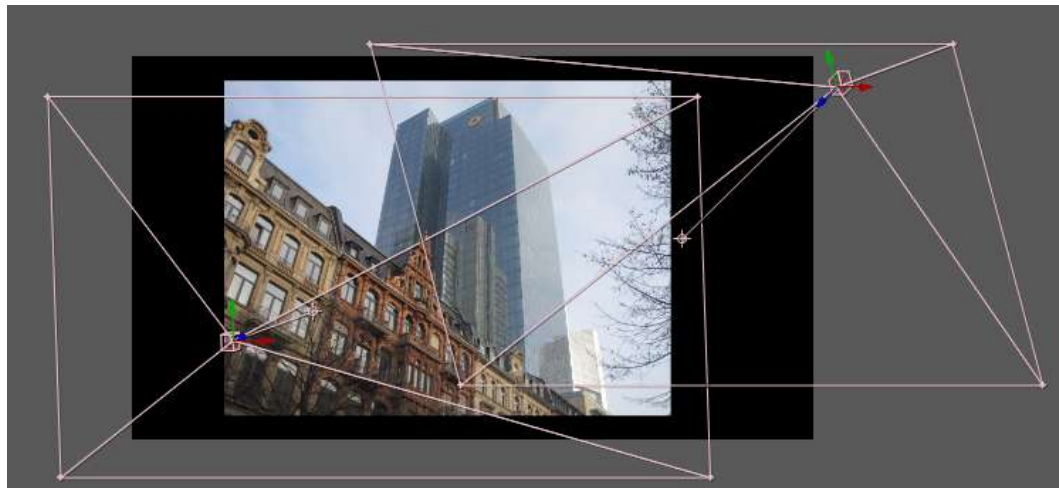
7.2.1 Προοπτική Προβολή

Η προοπτική προβολή αντιστοιχεί στην απεικόνιση της σκηνής όπως αυτή παρατηρείται από ένα σημείο στο χώρο. Στη προβολή αυτή τα τέσσερα βασικά στοιχεία είναι το αντικείμενο, το σημείο παρακολούθησης, οι ακτίνες προβολής και το επίπεδο στο οποίο γίνεται η προβολή.

Το αντικείμενο ή η σκηνή είναι αυτό που παρατηρείται από το σημείο παρακολούθησης. Το σημείο παρακολούθησης αντιπροσωπεύει τη θέση, μέσα στον τρισδιάστατο χώρο, από την οποία παρακολουθείται το αντικείμενο. Οι ακτίνες προβολής είναι αντίστοιχες με τις ακτίνες φωτός που φθάνουν από κάθε ορατό σημείο, του αντικειμένου, στο σημείο παρακολούθησης. Περνώντας από το σημείο παρακολούθησης οι ακτίνες αυτές συνεχίζουν τη πορεία τους μέχρι να φτάσουν στο επίπεδο προβολής. Ο σκοπός του επιπέδου προβολής είναι να φτάσουν σε αυτό οι ακτίνες προβολής και να προβάλουν σε αυτό το αντικείμενο.

Οι ακτίνες προβολής, εφόσον περνούν όλες από το σημείο παρακολούθησης, δεν μπορούν να είναι παράλληλες. Αυτό φέρει ως αποτέλεσμα, στη προβολή της σκηνής να μη διατηρούνται οι γωνίες του αντικειμένου. Επίσης, οι παράλληλες ευθείες του, οι

οποίες δεν είναι παράλληλες με το επίπεδο προβολής, δεν διατηρούνται παράλληλες. Τέλος, το μέγεθος του αντικειμένου, στη προοπτική προβολή, είναι ανάλογο των αποστάσεων μεταξύ αντικειμένου και σημείου παρακολούθησης και μεταξύ σημείου παρακολούθησης και επιπέδου προβολής.



(α)



(β)



(γ)

Σχήμα 7.1 Παρακολούθηση μιας τρισδιάστατης σκηνής από δύο κάμερες, οι οποίες βρίσκονται σε διαφορετικά σημεία στον τρισδιάστατο χώρο. Οι δυο κάμερες βλέπουν την επίπεδη επιφάνεια – φωτογραφία που βρίσκεται στη τρισδιάστατη σκηνή (α). Η καταγραφές της σκηνής από τις δύο κάμερες παρουσιάζονται στη (β) και (γ).

Η σημαντικότητα της προοπτικής προβολής σε σχέση με άλλα είδη προβολής είναι ότι εδώ γίνεται μια ρεαλιστική αναπαράσταση των τρισδιάστατων χαρακτηριστικών του αντικειμένου καθώς αντιπροσωπεύει τον τρόπο που δουλεύει το ανθρώπινο μάτι καθώς και η κάμερα, Σχήμα 7.1.

7.2.2 Χρήση Ομογραφίας στην εργασία

Στην εργασία αυτή χρησιμοποιείται η ομογραφία για να υπολογιστεί ο χώρος που καταλαμβάνει η σελίδα μέσα στη σκηνή. Η ομογραφία υπολογίζεται όταν η διαδικασία τάυτισης χαρακτηριστικών έχει τελειώσει και έχουν βρεθεί αρκετά χαρακτηριστικά που έχουν ταιριάξει ώστε να θεωρείται ότι μια σελίδα έχει ανιχνευτεί με επιτυχία.

Παρόλο που αυτά τα χαρακτηριστικά από μόνα τους δίνουν πληροφορίες σχετικά με το πού βρίσκεται η σελίδα μέσα στη σκηνή, υπάρχει η ανάγκη για μια πιο γενική συσχέτιση της σελίδας με το χώρο που αυτή καταλαμβάνει. Για το λόγο αυτό, βάση αυτών των ζευγαριών, υπολογίζεται η ομογραφία και δημιουργείται ο ομογραφικός πίνακας (homography matrix) της σελίδας, ο οποίος συσχετίζει την αναφορική εικόνα με τη σκηνή.

$$H = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \quad (7.1)$$

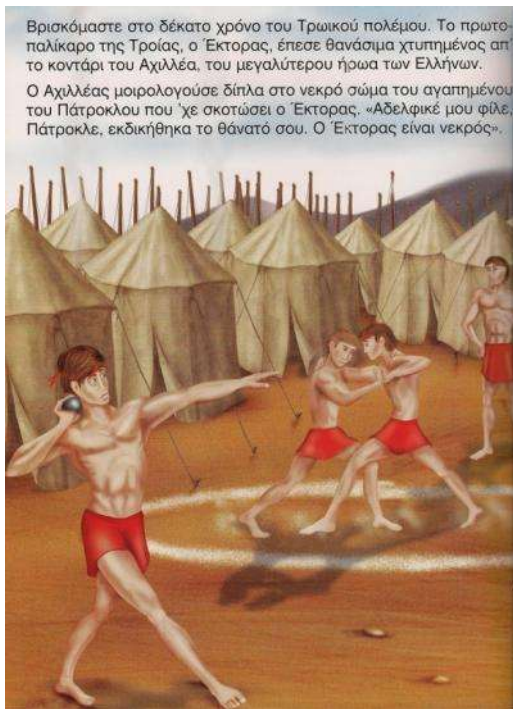
Για τον υπολογισμό του πίνακα αυτού χρειάζονται τουλάχιστον τέσσερα ζευγάρια σημείων, το κάθε ένα από αυτά πρέπει να αποτελεί μια αντιστοίχιση ενός σημείου της αναφορικής εικόνας της σελίδας, και της καταγραφής της σκηνής.

Μέσο του ομογραφικού πίνακα μπορεί να βρεθεί η θέση κάθε σημείου της αναφορικής εικόνας της σελίδας μέσα στη σκηνή. Με τον τρόπο αυτό μπορούν να βρεθούν οι θέσεις των τεσσάρων γωνιών της σελίδας μέσα στη σκηνή, και να αναγνωριστεί ο χώρος που καταλαμβάνει. Ο υπολογισμός της θέσης ενός σημείου της αναφορικής σελίδας του βιβλίου μέσα στη σκηνή μπορεί να γίνει ως εξής:

$$p_s = Hp_r \quad (7.2)$$

Όπου p_r ένα σημείο της σελίδας, σε τρεις διαστάσεις ($x,y,z=1$) και p_s η αντίστοιχη θέση του σημείου p_r μέσα στη σκηνή και πάλι σε τρεις διαστάσεις.

Στο Σχήμα 7.2 παρουσιάζεται παράδειγμα χρήσης της ομογραφίας μέσα στην εφαρμογή όπου βάση των σημείων που έχουν ταυριάζει, υπολογίζεται ο ομογραφικός πίνακας και στη συνέχεια υπολογίζονται τα όρια της σελίδας μέσα στη σκηνή.



(α)



(β)



(γ)

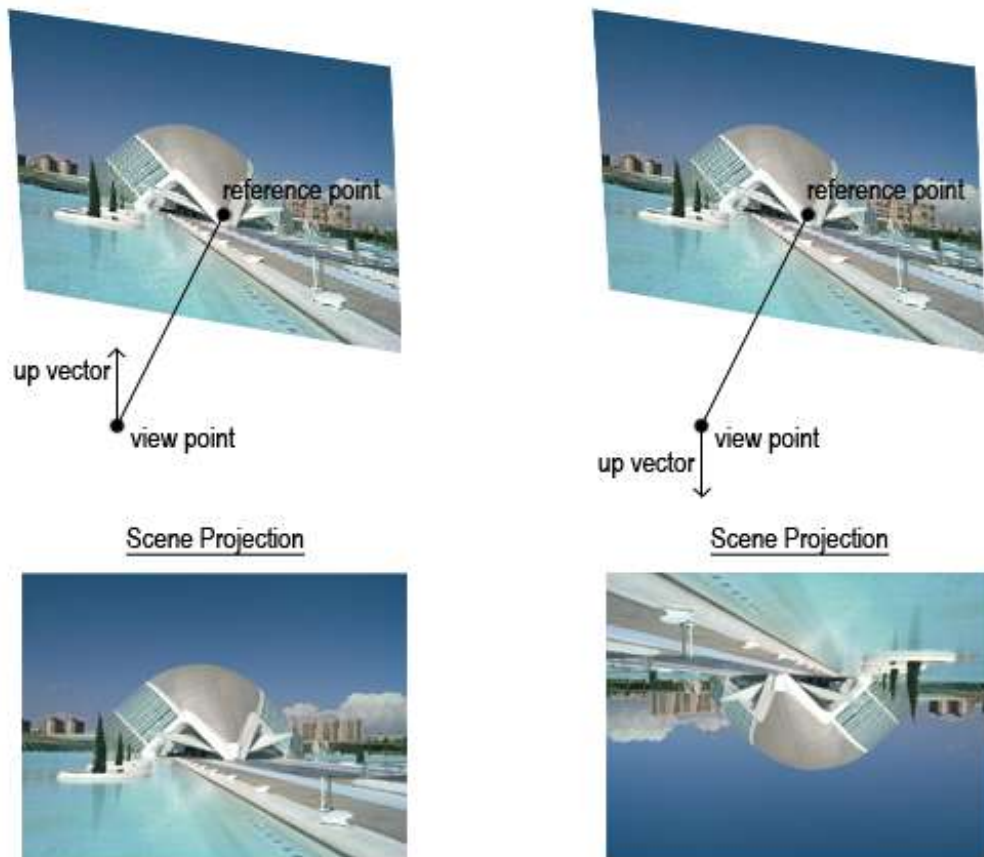
Σχήμα 7.2 Χρήση ομογραφίας στην εφαρμογή. (α) η αναφορική εικόνα μιας σελίδας. (β) καταγραφή σκηνής με σημειωμένα με κύκλους τα σημεία που έχουν ταυτιστεί. (γ) η καταγραφή της σκηνής με σημειωμένο το χώρο που καταλαμβάνει η σελίδα, όπως μπορεί να βρεθεί με την αντιστοίχιση των τεσσάρων γωνιών, της βάση του ομογραφικού πίνακα.

7.3 Εικονική Επαύξηση σκηνής

Μέσα στο σύστημα που έχει αναπτυχθεί στα πλαίσια αυτής της εργασίας, συνυπάρχουν δύο περιβάλλοντα. Το πρώτο είναι ένα δυδιάστατο του οποίου ο χειρισμός γίνεται μέσω της βιβλιοθήκης OpenCV [14], σε αυτό γίνεται η επεξεργασία των εικόνων που καταγράφονται από τη κάμερα. Το δεύτερο περιβάλλον είναι ένας τρισδιάστατος

εικονικός χώρος δημιουργημένος μέσω της OpenGL [12] [15], όπου μέσα σε αυτό γίνεται ο χειρισμός των τρισδιάστατων εικονικών αντικειμένων.

Στην οθόνη του χρήστη, αυτά τα δύο περιβάλλοντα παρουσιάζονται, ως ένα, σαν να υπάρχουν δύο επίπεδα εικόνας με τον δισδιάστατο χώρο να μπαίνει σαν φόντο στο τρισδιάστατο.

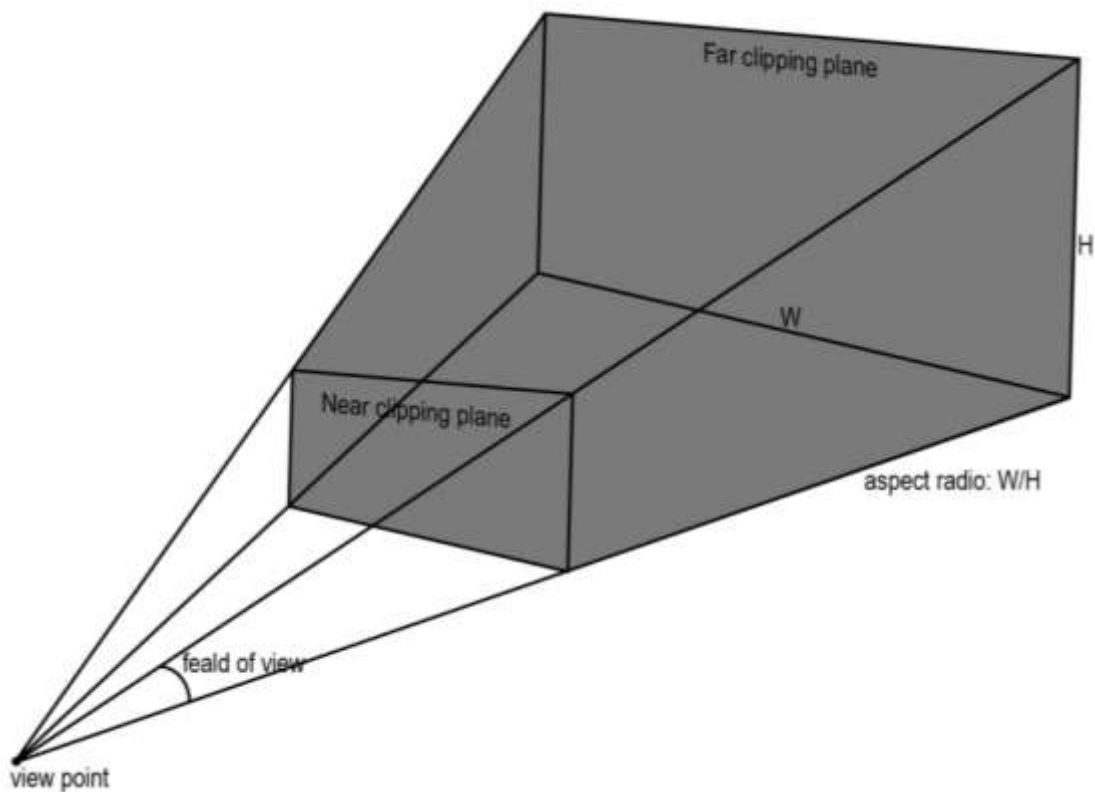


Σχήμα 7.3 Προς τα πάνω κατεύθυνση του σημείου παρακολούθησης. Η προς τα πάνω κατεύθυνση επηρεάζει τη περιστροφή της προβολής της σκηνής.

Για την παρουσίαση στο χρήστη εικόνων από τον εικονικό κόσμο, η OpenGL χρησιμοποιεί σημεία παρακολούθησης (points of view). Ένα τέτοιο σημείο χαρακτηρίζεται από τη θέση του στο χώρο σε τρισδιάστατες συντεταγμένες (x,y,z) , τη θέση του αναφορικού σημείου (reference point), καθώς και από τη προς τα πάνω διεύθυνση του (up vector). Το αναφορικό σημείο καθορίζει το προς ποία κατεύθυνση βλέπει το σημείο παρακολούθησης, και η προς τα πάνω διεύθυνση δηλώνει το πώς

περιστρέφεται το σημείο γύρω από τον άξονα που το ενώνει με το αναφορικό σημείο το οποίο βλέπει, Σχήμα 7.3.

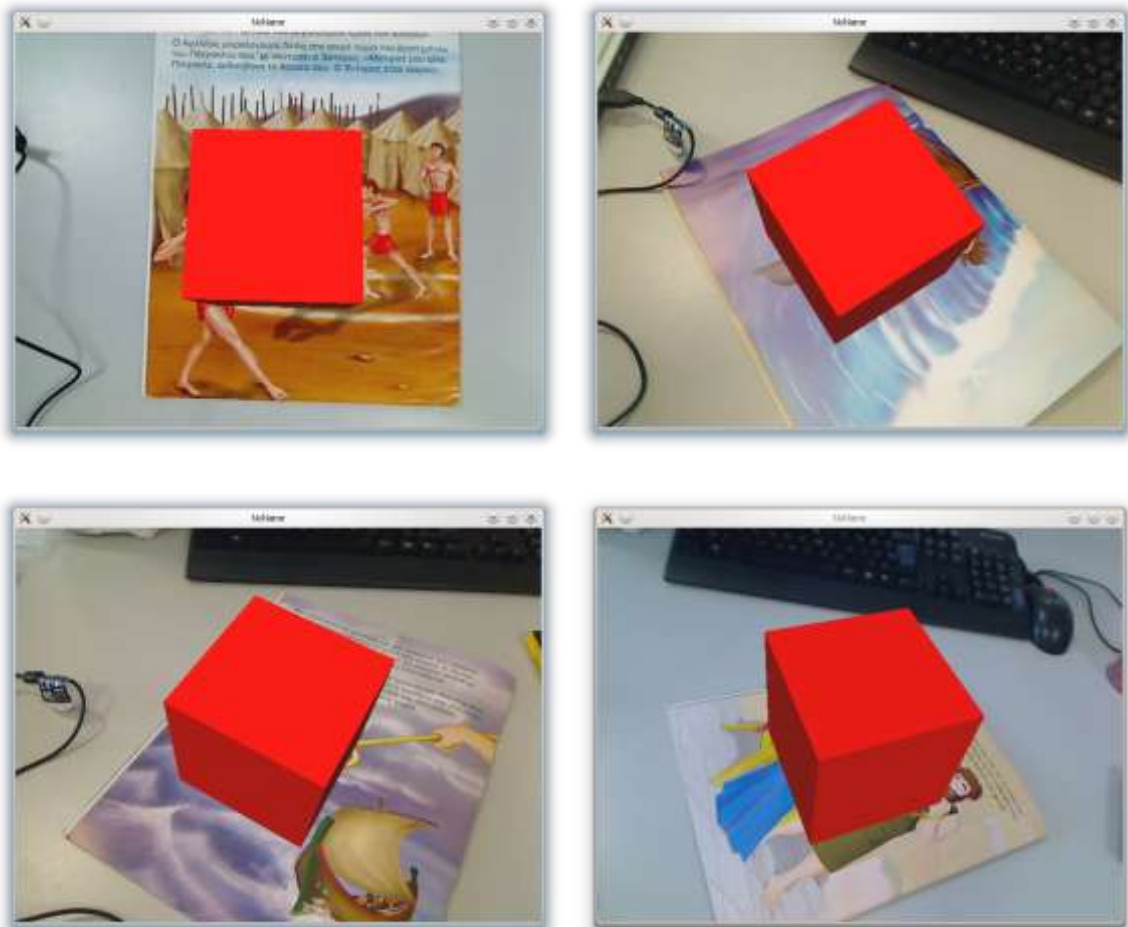
Επίσης στην OpenGL δίνεται η δυνατότητα για επιλογή είδους προβολής. Για τις ανάγκες της εργασίας χρησιμοποιείται η προοπτική προβολή. Για τη χρήση της προοπτικής προβολής, χρειάζεται να δηλωθούν τα χαρακτηριστικά της. Τα χαρακτηριστικά αυτά είναι η γωνία του οπτικού πεδίου (field of view angle) στον κατακόρυφο άξονα, η αναλογία της εικόνας (aspect ratio) που θα καταγράφεται στο επίπεδο προβολής, και η απόσταση του κοντινού και μακρινού επιπέδου αποκοπής από το σημείο παρακολούθησης. Όλες αυτές οι παράμετροι υποδηλώνουν τα όρια του ορατού πεδίου του σημείου παρακολούθησης, όπως φαίνεται στο Σχήμα 7.4.



Σχήμα 7.4 Ορατό οπτικό πεδίο στην OpenGL. Η γκρίζα περιοχή υποδηλώνει το ορατό πεδίο από το σημείο παρακολούθησης. Ότι βρίσκεται μπροστά από το κοντινό επίπεδο αποκοπής και πέραν του μακρινού επιπέδου αποκοπής αγνοείται κατά την απεικόνιση του εικονικού κόσμου.

7.3.1 Περιστροφή εικονικού αντικειμένου γύρω από τον εαυτό του

Πρώτο βήμα για τη τοποθέτηση του αντικειμένου, είναι η περιστροφή αυτού ώστε να ταυτιστεί με τη περιστροφή του βιβλίου μέσα στη σκηνή. Στο Κεφάλαιο 6 έχει περιγραφεί η διαδικασία βαθμονόμησης της κάμερας, στην οποία μεταξύ άλλων υπολογίζεται και ο πίνακας περιστροφής της κάμερας. Ο πίνακας αυτός θα χρησιμοποιηθεί εδώ για να επιτευχθεί η περιστροφή του αντικειμένου.



Σχήμα 7.5 Παραδείγματα περιστροφής του εικονικού αντικειμένου. Το εικονικό αντικείμενο περιστρέφεται με ανάλογο τρόπο, με τον οποίο το βιβλίο φαίνεται να περιστρέφεται μέσα στη σκηνή που καταγράφει η κάμερα. Στις εικόνες αυτές εφαρμόζεται μόνο περιστροφή στο εικονικό αντικείμενο.

Όπως έχει ήδη αναφερθεί η βαθμονόμηση της κάμερας γίνεται βάση του επιπέδου του βιβλίου και ο εν λόγω πίνακας περιστροφής, περιγράφει τη περιστροφή της κάμερας μέσα στο χώρο, βάση του βιβλίου. Αφού ο πίνακας αυτός περιγράφει τη περιστροφή

της κάμερας σε σχέση με το επίπεδο του βιβλίου ο ίδιος πίνακας περιγράφει και τη περιστροφή του επιπέδου σε σχέση με τη κάμερα.

Επομένως, το εικονικό αντικείμενο περιστρέφεται γύρω από τον εαυτό του και ταυτίζεται έτσι με τη περιστροφή του βιβλίου σε σχέση με τη κάμερα. Επομένως η περιστροφή στη προβολή του εικονικού αντικειμένου ταιριάζει με αυτή του βιβλίου, Σχήμα 7.5.

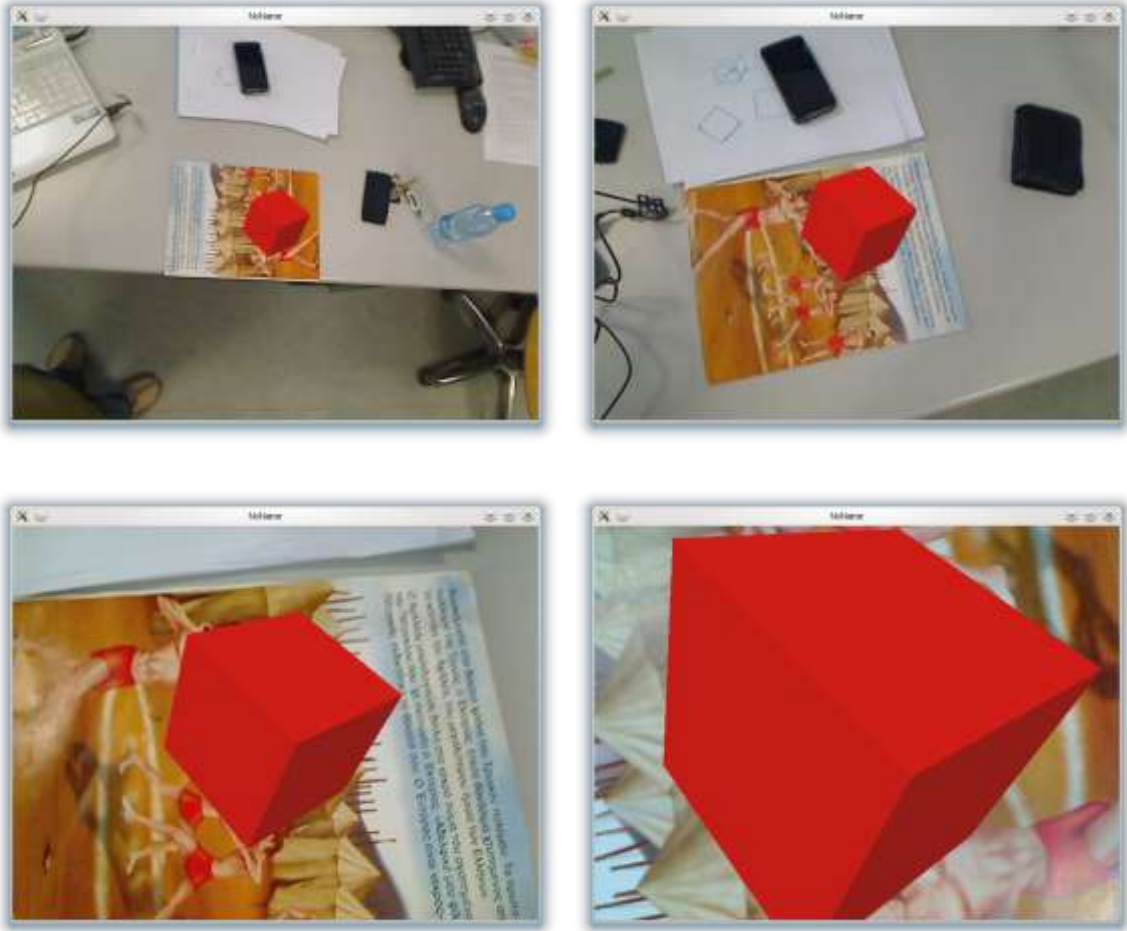
7.3.2 Μέγεθος εικονικού αντικειμένου

Το μέγεθος του αντικειμένου πρέπει να είναι ανάλογο με το μέγεθος της σελίδας, έτσι ώστε να διατηρείται μια ισορροπία μεταξύ των μεγεθών των δύο και να δίνεται στον παρακολουθητή η αίσθηση ότι το εικονικό αντικείμενο συνυπάρχει και ακολουθεί το βιβλίο μέσα σε ένα κοινό χώρο.

Για να επιτευχθεί αυτό, αρχικά, πρέπει να αποφασιστεί το επιθυμητό μέγεθος του αντικειμένου για ένα προεπιλεγμένο (default) μέγεθος της σελίδας. Για να γίνει αυτό απαιτείται πειραματική μελέτη του μεγέθους του εικονικού αντικειμένου σε σχέση με το προεπιλεγμένο μέγεθος της σελίδας.

Κατά την εκτέλεση της εφαρμογής υπολογίζεται το μέγεθος της σελίδας που ανιχνεύτηκε μέσα στη σκηνή βάση των τεσσάρων γωνιών της. Οι τέσσερις γωνίες της σελίδας μέσα στη σκηνή μπορούν να εντοπιστούν με τη χρήση του ομογραφικού πίνακα, όπως αναφέρεται στην ενότητα 7.2.2 αυτού του κεφαλαίου.

Βάση των αποστάσεων των τεσσάρων γωνιών της σελίδας μπορούν να βρεθούν οι διαστάσεις της σελίδας και να υπολογιστεί ο λόγος τους σε σχέση με το προεπιλεγμένο μέγεθος της σελίδας. Στη συνέχεια, ο λόγος αυτός πολλαπλασιάζεται με το μέγεθος του αντικειμένου που έχει αποφασιστεί για το προεπιλεγμένο μέγεθος σελίδας και το αντικείμενο απεικονίζεται στην ανάλογη κλίμακα (scale), Σχήμα 7.6



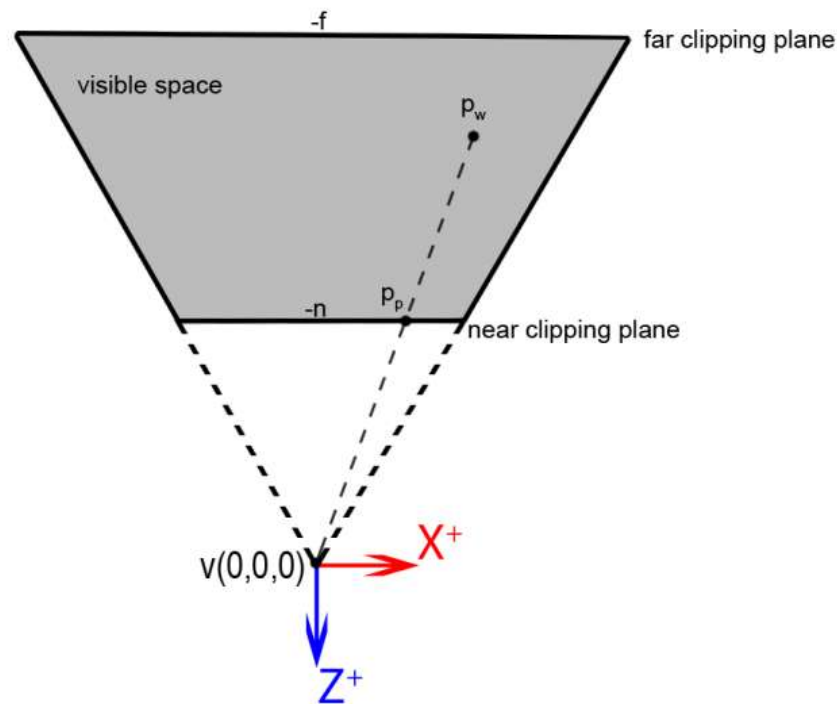
Σχήμα 7.6 Παράδειγμα τοποθέτησης του εικονικού αντικειμένου με μέγεθος ανάλογο του μεγέθους της σελίδας. Στις εικόνες αυτές μόνο η μεγέθυνση εφαρμόζεται στο εικονικό αντικείμενο. Περιστροφή και μετακίνηση δεν εφαρμόζονται.

7.3.3 Θέση εικονικού αντικειμένου

Για την απεικόνιση του εικονικού αντικειμένου στη επιθυμητή θέση, αρχικά πρέπει να βρεθούν οι συντεταγμένες στις οποίες πρέπει αυτό να βρίσκεται μέσα στην σκηνή, βάση της εικόνας της σκηνής την οποία καταγράφει η κάμερα. Οι συντεταγμένες αυτές θα είναι δύο διαστάσεων και θα πρέπει να μετατραπούν σε συντεταγμένες του επιπέδου προβολής του εικονικού κόσμου. Τέλος βάση αυτού υπολογίζεται η μετατόπιση της εικονικής κάμερας.

Ξέροντας τις συντεταγμένες του σημείου, μέσα στην αναφορική εικόνα της σελίδας, στο οποίο πρέπει να εμφανίζεται το εικονικό αντικείμενο, μπορούν να υπολογιστούν η αντίστοιχες συντεταγμένες του σημείου μέσα στη εικόνα που λαμβάνεται από τη

κάμερα, με τη χρήση του ομογραφικού πίνακα. Οι συντεταγμένες αυτές θα είναι σχετικές με την αρχή των αξόνων της εικόνας που καταγράφει η κάμερα. Η αρχή των αξόνων, για αυτές τις εικόνες, βρίσκεται στη πάνω αριστερή γωνία με τις οριζόντιες συντεταγμένες να αυξάνονται προς τα δεξιά και τις κατακόρυφες να αυξάνονται προς τα κάτω. Στη συνέχεια, οι συντεταγμένες αυτές πρέπει να μετατραπούν σε σύστημα συντεταγμένων του επιπέδου προβολής της εικονικής κάμερας. Σε αντίθεση με τις εικόνες που λαμβάνονται από την πραγματική κάμερα, η αρχή των αξόνων για το επίπεδο προβολής στην OpenGL αντιστοιχεί με το κέντρο του επιπέδου, με τις οριζόντιες συντεταγμένες να αυξάνονται προς τα αριστερά και τις κατακόρυφες να αυξάνονται προς τα πάνω.



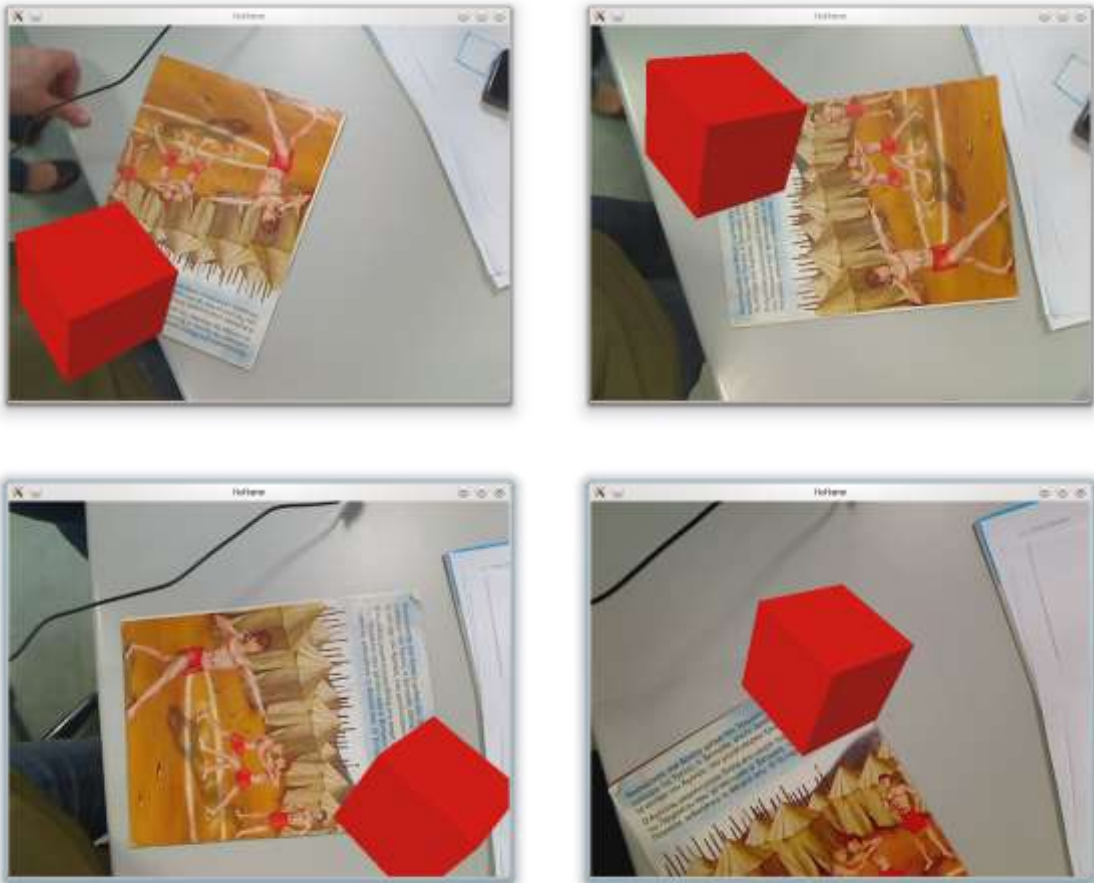
Σχήμα 7.7 Κάτοψη προοπτικής προβολής στην OpenGL. Το επίπεδο προβολής είναι το ίδιο με το επίπεδο κοντινής αποκοπής, το οποίο βρίσκεται σε απόσταση n από το σημείο παρακολούθησης v . Το p_p η προβολή του σημείου p_w πάνω στο επίπεδο προβολής. Η γκριζα περιοχή αντιπροσωπεύει την ορατή περιοχή προς το σημείο παρακολούθησης.

Η μετατροπή των συντεταγμένων μιας εικόνας που λαμβάνεται από τη κάμερα, σε συντεταγμένες στο επίπεδο προβολής μπορεί να γίνει ως εξής:

$$x_p = x_c - h \quad (7.3)$$

$$y_p = y_c - w \quad (7.4)$$

Όπου x_c , y_c η συντεταγμένες του σημείου στη εικόνα που λαμβάνεται από τη κάμερα, x_p , y_p οι συντεταγμένες βάση του συστήματος συντεταγμένων του επιπέδου προβολής και w, h το πλάτος και το ύψος του επιπέδου προβολής αντίστοιχα.



Σχήμα 7.8 Παραδείγματα μετακίνησης του εικονικού αντικειμένου μέσα στη σκηνή. Το αντικείμενο ακολουθεί τη δεξιά γωνία της σελίδας. Στις εικόνες του παραδείγματος μόνο μετακίνηση του αντικειμένου εφαρμόζεται. Περιστροφή και μετακίνηση δεν εφαρμόζονται.

Το επίπεδο προβολής στην OpenGL συμπίπτει με το κοντινό επίπεδο αποκοπής (near clipping plane). Το κοντινό επίπεδο και το μακρινό επίπεδο αποκοπής (far clipping plane) ορίζουν τα όρια του βάθους όπου τα αντικείμενα είναι ορατά, δηλαδή καταγράφονται στο επίπεδο προβολής. Τα επίπεδα αυτά βρίσκονται πάντα μπροστά από το σημείο παρακολούθησης, και ότι βρίσκεται μεταξύ του κοντινού επιπέδου αποκοπής και του σημείου παρακολούθησης δεν καταγράφεται όπως επίσης και ότι

βρίσκεται πέραν του μακρινού επιπέδου. Έτσι το επίπεδο προβολής σε αυτή τη περίπτωση δε βρίσκεται πίσω από το σημείο παρακολούθησης αλλά μπροστά από αυτό.

Επομένως, η θέση της προβολής ενός σημείου στο τρισδιάστατο χώρο, πάνω στο επίπεδο προβολής, εξαρτάται από την απόσταση του κοντινού επιπέδου αποκοπής από το σημείο παρακολούθησης. Καθώς επίσης και από την απόσταση των παράλληλων επιπέδων προς το επίπεδο αποκοπής του σημείου παρακολούθησης και του ενδιαφερόμενου σημείου. Αν το σημείο παρακολούθησης βρίσκεται στο σημείο $(0,0,0)$ με κατεύθυνση τον αρνητικό άξονα των z , με απόσταση από το επίπεδο παρακολούθησης n , η προβολή p_p του σημείου p_w με συντεταγμένες (x_w, y_w, z_w) , όπως φαίνεται στο Σχήμα 7.7, μπορεί να υπολογιστεί ως εξής:

$$x_p = \frac{-n}{z_w} x_w \quad (7.5)$$

$$y_p = \frac{-n}{z_w} y_w \quad (7.6)$$

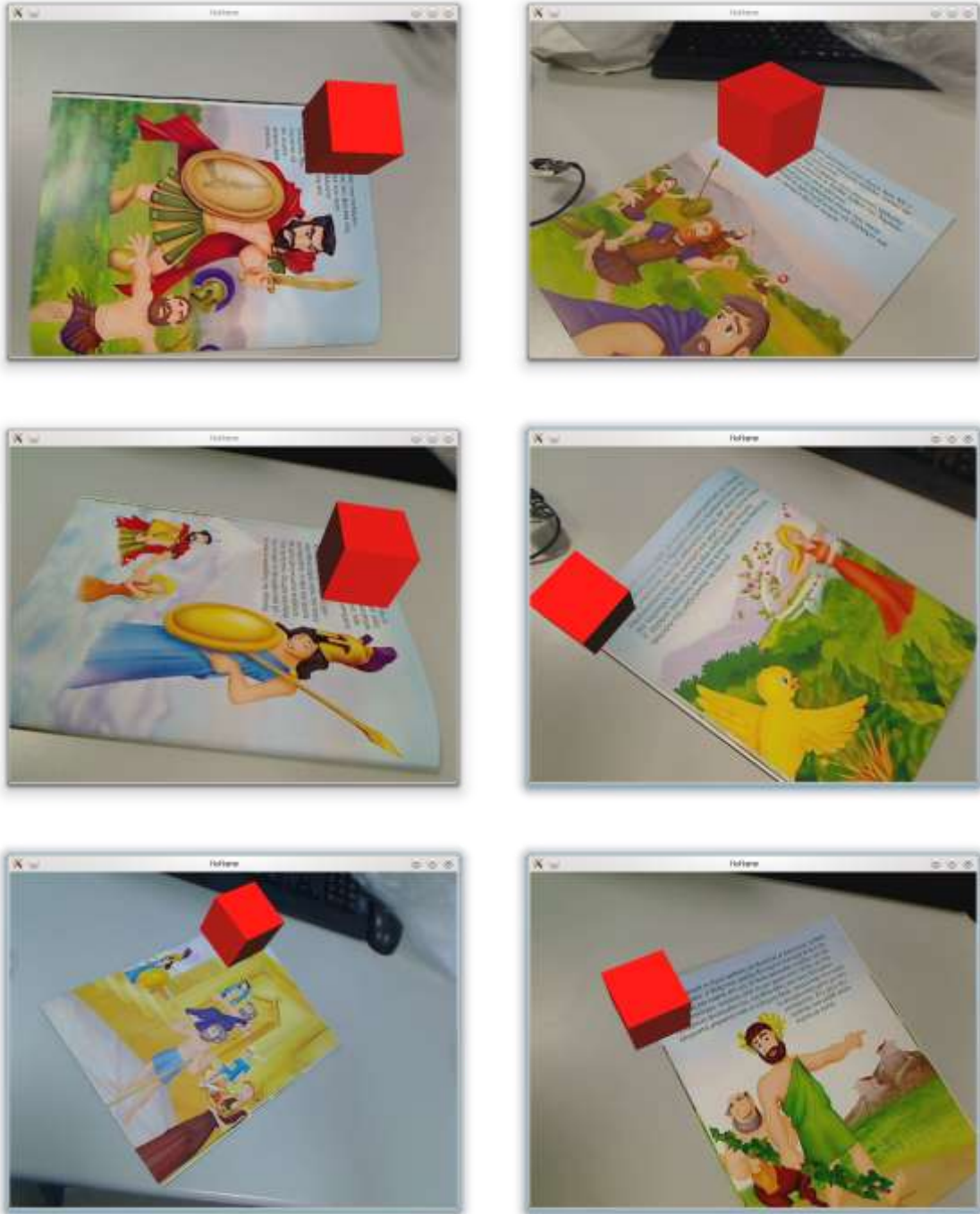
$$z_p = -n \quad (7.7)$$

Οπότε έχοντας ως γνωστή τη θέση που πρέπει να προβληθεί το εικονικό αντικείμενο μέσα στο επίπεδο προβολής καθώς και το βάθος στο οποίο βρίσκεται (θέση κατά το μήκος του άξονα z). Τότε, η βάση των (5) και (6) οι θέση του αντικειμένου στους άξονες x και y μπορεί να υπολογιστεί ως εξής:

$$x_w = \frac{z_{w1}}{-n} x_p \quad (7.8)$$

$$y_w = \frac{z_{w1}}{-n} y_p \quad (7.9)$$

Επομένως, το εικονικό αντικείμενο τοποθετείται στον τρισδιάστατο χώρο στη θέση που έχει υπολογιστεί με τις εξισώσεις (8) και (9), και στη προβολή του στην οθόνη του χρήστη αυτό φαίνεται να βρίσκεται στο επιθυμητό σημείο του βιβλίου, Σχήμα 7.8. Στο Σχήμα 7.9 φαίνονται τα αποτελέσματα της εφαρμογής όταν εφαρμόζονται περιστροφή, μεγέθυνση και μετακίνηση του εικονικού αντικειμένου μαζί.



Σχήμα 7.9 Αποτελέσματα εφαρμογής, όπου το εικονικό αντικείμενο, περιστρέφεται και παίρνει μέγεθος ανάλογα με το βιβλίο και τοποθετείται στη πάνω δεξιά γωνία του,

7.4 Σύνοψη

Η διαχείριση του εικονικού κόσμου μέσα στην εφαρμογή γίνεται μέσω της OpenGL η οποία είναι μια βιβλιοθήκη ανοικτού κώδικα. Η βιβλιοθήκη αυτή παρέχει τη δυνατότητα δημιουργίας και διαχείρισης τρισδιάστατων γραφικών για υπολογιστές (3D

Computer Graphics). Ο τρισδιάστατος αυτός χώρος προβάλλεται στη οθόνη του χρήστη με χρήση προοπτικής προβολής, η οποία προσφέρει ρεαλιστική απεικόνιση των τρισδιάστατων χαρακτηριστικών των αντικειμένων.

Για τη σωστή απεικόνιση του εικονικού αντικειμένου μέσα στην οθόνη που βλέπει ο χρήστης το αντικείμενο πρέπει να περιστραφεί γύρω από τον εαυτό του, έτσι ώστε η περιστροφή του να ταυτίζεται με τη περιστροφή του βιβλίου σε σχέση με το σημείο από το οποίο παρακολουθείται από τη κάμερα. Επίσης, το αντικείμενο πρέπει να απεικονιστεί με ανάλογο μέγεθος με αυτό που έχει το βιβλίο μέσα στη σκηνή, καθώς και στη σωστή θέση πάνω σε αυτό.

Όλα αυτά μπορούν να υπολογιστούν με τη χρήση των αποτελεσμάτων που έχει δώσει η διαδικασία βαθμονόμησης της κάμερας και του πίνακα ομογραφίας. Από τη βαθμονόμηση της κάμερας εξάγονται πληροφορίες σχετικά με την περιστροφή, ενώ με τη χρήση του ομογραφικού πίνακα μπορεί βρεθεί ο χώρος των οποίου καταλαμβάνει το βιβλίο μέσα στη σκηνή, και βάση αυτού να υπολογιστεί το μέγεθος και η θέση που πρέπει να τοποθετηθεί το αντικείμενο.

Κεφάλαιο 8

Υπολογισμός νέας θέσης αντικειμένου μέσω της Οπτικής Ροής (Optical Flow)

| | |
|---|----|
| 8.1 Εισαγωγή | 63 |
| 8.2 Οπτική ροή | 64 |
| 8.3 Οπτική ροή για επιτάχυνση της παρακολούθησης (tracking) | 67 |
| 8.4 Πειραματικά αποτελέσματα Οπτικής Ροής..... | 68 |
| 8.5 Σύνοψη..... | 72 |

8.1 Εισαγωγή

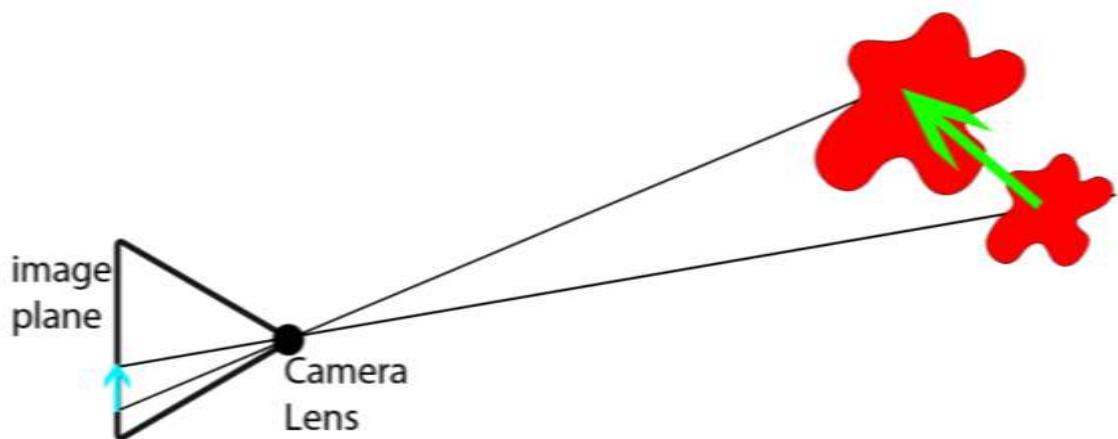
Με πειραματική μελέτη της εφαρμογής έχει παρατηρηθεί ότι σε απλούς φορητούς υπολογιστές που κυκλοφορούν στην αγορά αυτές τις μέρες η εφαρμογή δουλεύει χωρίς ιδιαίτερα προβλήματα. Ωστόσο, λαμβάνοντας υπόψη ότι φορητοί υπολογιστές οι οποίοι κατασκευάζονται για να χρησιμοποιούνται από παιδιά είναι υπολογιστικά ποιο αδύναμοι από ένα μέσο φορητό υπολογιστή. Έτσι, δημιουργείται η ανάγκη για εύρεση τεχνικών για μείωση του απαιτούμενου υπολογιστικού κόστους.

Στα πλαίσια αυτής της προσπάθειας προτείνεται η χρήση οπτικής ροής (optical flow). Η οπτική ροή μπορεί να χρησιμοποιηθεί, στην εφαρμογή, για μερική αντικατάσταση της διαδικασίας ταύτισης χαρακτηριστικών, η οποία είναι η πιο ακριβή, σε χρόνο υπολογισμού, διαδικασία που εκτελείται στην εφαρμογή.

Στη συνέχεια του κεφάλαιο αυτού θα περιγραφεί η λειτουργία της μεθόδου των Lucas και Kanade, για υπολογισμό της οπτικής ροής, και θα εξηγηθεί πώς εφαρμόζεται στα πλαίσια της εργασίας.

8.2 Οπτική ροή

Οπτική ροή είναι ο όρος που χρησιμοποιείται για να περιγράψει τη κίνηση που ανιχνεύεται μέσα σε μια καταγραφή μιας σκηνής, βάση μιας προηγούμενης καταγραφής της. Στην ουσία για την ανίχνευση αυτής της ροής υπολογίζεται η νέα θέση κάθε σημείου, από την χρονικά προγενέστερη καταγραφή στην μεταγενέστερη. Μέσα από αυτή την εύρεση, της μετακίνησης, μπορούν να εξαχθούν διάφορα συμπεράσματα σχετικά με την κίνηση όπως παραδείγματος χάριν η ταχύτητα της.



Σχήμα 8.1 Το κόκκινο αντικείμενο μετακινείται σε τρισδιάστατο χώρο, το πράσινο διάνυσμα δείχνει τη μετακίνηση του. Το μπλε βέλος αντιπροσωπεύει το optical flow vector (διάνυσμα οπτικής ροής) όπου ορίζεται ως τη μετακίνηση της προβολής του αντικειμένου μέσα στην εικόνα που καταγράφει η κάμερα.

Οι αλγόριθμοι υπολογισμού της οπτικής ροής, προσπαθούν να υπολογίσουν την κίνηση που παρατηρείται σε δύο εικόνες, που έχουν καταγραφεί σε χρόνο t_1 και t_2 αντίστοιχα, για κάθε στοιχείο που εμφανίζεται στη σκηνή. Σε μια τέτοια προσέγγιση οι δισδιάστατες εικόνες αποκτούν μια τρίτη διάσταση, αυτή του χρόνου. Έτσι, κάθε στοιχείο μπορεί να περιγραφεί ως (x,y,t) σε ένα χώρο τριών διαστάσεων όπου x και y η οριζόντια και η κάθετη θέση του μέσα στα frames και t η χρονική στιγμή κατά την οποία βρίσκεται στη θέση (x,y) . Έτσι, προκύπτει ότι η φωτεινότητα των δύο αντίστοιχων σημείων, στις δύο εικόνες, είναι ίση, οπότε έχουμε την εξής ισότητα:

$$I(x, y, t) = I(x + dx, y + dy, t + dt) \quad (8.1)$$

Με την πρώτη επέκταση των σειρών Taylor μπορεί να γραφτεί ως:

$$I(x, y, t) = I(x, y, t) + \frac{\delta I}{\delta x} dx + \frac{\delta I}{\delta y} dy + \frac{\delta I}{\delta t} dt \quad (8.2)$$

Και μπορεί να απλοποιηθεί ως:

$$I_x dx + I_y dy + I_t dt = 0 \quad (8.3)$$

διαίρωντας με dt, παίρνουμε:

$$u = \frac{dx}{dt}, v = \frac{dy}{dt} \quad (8.4)$$

και τότε η (8.3) γράφεται:

$$I_x u + I_y v = -I_t \quad (8.5)$$

Στο σύστημα αυτό υπάρχουν δύο άγνωστοι και μία μόνο εξίσωση, πράγμα που καθιστά δύσκολη την ακριβή επίλυση του.

Στη μέθοδο Lucas-Kanade [9] λαμβάνεται η υπόθεση ότι η θέση των περιεχομένων μιας σκηνής, ανάμεσα σε δύο διαδοχικά frames, είναι σχεδόν σταθερή. Έτσι για κάθε σημείο αναμένεται να βρεθεί η νέα του θέση σε μια σχετικά μικρή γειτονική περιοχή.

Βάση αυτής της υπόθεσης για κάθε pixel p δημιουργείται ένα παράθυρο, το οποίο έχει ως κεντρικό του στοιχείο το p και αντιπροσωπεύει τη γειτονική περιοχή αυτού. Μέσα στη περιοχή αυτού του παραθύρου αναμένεται να βρεθεί η νέα θέση του σημείου p.

Η εξίσωση (8.5) θα μπορούσε να γραφτεί και ως:

$$\begin{bmatrix} I_x & I_y \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} = -I_t \quad (8.6)$$

Βάση της υπόθεσης ότι τα u και v είναι σταθερά για τη μικρή περιοχή που αντιπροσωπεύει το παράθυρο τότε μπορεί να χρησιμοποιηθεί το εξής σύστημα:

$$A\vec{u} = B \quad (8.7)$$

όπου

$$A = \begin{bmatrix} I_{x1} & I_{y1} \\ I_{x2} & I_{y2} \\ \cdot & \cdot \\ \cdot & \cdot \\ I_{xn} & I_{yn} \end{bmatrix} \quad (8.8), \quad \vec{u} = \begin{bmatrix} u \\ v \end{bmatrix} \quad (8.9), \quad B = - \begin{bmatrix} I_{t1} \\ I_{t2} \\ \cdot \\ \cdot \\ I_{tn} \end{bmatrix} \quad (8.10)$$

Στο σύστημα αυτό υπάρχουν πολύ περισσότερες εξισώσεις από ότι είναι οι άγνωστοι, με αποτέλεσμα να υπερ-προσδιορίζεται (over determined). Για αυτό το λόγο χρησιμοποιείται προσέγγιση ελαχίστων τετραγώνων (least squares), με το σύστημα να μετατρέπεται ως εξής:

$$A^T A \vec{u} = A^T B \quad (8.11)$$

ή

$$\vec{u} = (A^T A)^{-1} A^T B \quad (8.12)$$

Όπου A^T ο ανάστροφος του A . Η (8.12) υπολογίζεται ως

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^n I_{xi}^2 & \sum_{i=1}^n I_{xi} I_{yi} \\ \sum_{i=1}^n I_{xi} I_{yi} & \sum_{i=1}^n I_{yi}^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum_{i=1}^n I_{xi} I_{ti} \\ \sum_{i=1}^n I_{yi} I_{ti} \end{bmatrix} \quad (8.13)$$

Μέχρι στιγμής, στο σύστημα αυτό δίνεται η ίδια σημαντικότητα σε όλα τα pixel που περιέχονται στο παράθυρο, πρακτικά όμως είναι χρησιμότερο να δίνεται μεγαλύτερη σημαντικότητα στα pixels που βρίσκονται πιο κοντά στο κεντρικό pixel p . Για αυτό το σύστημα μπορεί να γραφτεί ως εξής:

$$\vec{u} = (A^T W A \vec{u})^{-1} A^T W B \quad (8.14)$$

Όπου W ένας πίνακας διαστάσεων $n \times n$ όπου στη διαγώνιο του κρατεί τα βάρη των pixel του παραθύρου, και υπολογίζεται ως

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^n w_i I_{xi}^2 & \sum_{i=1}^n w_i I_{xi} I_{yi} \\ \sum_{i=1}^n w_i I_{xi} I_{yi} & \sum_{i=1}^n w_i I_{yi}^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum_{i=1}^n w_i I_{xi} I_{ti} \\ \sum_{i=1}^n w_i I_{yi} I_{ti} \end{bmatrix} \quad (8.15)$$

Όπου $w_i = W_{ii}$.

Η μέθοδος αυτή, των Lucas and Kanade, παρουσιάζει μερικά προβλήματα, όπως για παράδειγμα δε μπορεί να ανιχνεύσει τη κίνηση σημείων που βρίσκονται στο εσωτερικό ομογενών περιοχών. Είναι όμως πολύ καλή μέθοδος για περιοχές με πληροφορίες υψηλών συχνοτήτων.

8.3 Οπτική ροή για επιτάχυνση της παρακολούθησης (tracking)

Όπως έχει περιγραφεί μέχρι το παρών στάδιο, για κάθε frame που λαμβάνεται από τη κάμερα εξάγονται τα χαρακτηριστικά και στη συνέχεια εκτελείται η διαδικασία ταύτισης των χαρακτηριστικών του frame με τα χαρακτηριστικά της σελίδας. Αν η σελίδα βρεθεί μέσα στη σκηνή τότε τοποθετείται το αντικείμενο.

Η διαδικασία της ταύτισης, είναι μια διαδικασία η οποία απαιτεί χρόνο τάξης $O(n^2)$, αυτό είναι λογικό να μειώνει την απόδοση της εφαρμογής όταν εκτελείται για κάθε frame που λαμβάνεται από τη κάμερα. Έχοντας ως επακόλουθο να καθυστερείται το

αίτημα για νέο frame και επομένως να μειώνεται ο αριθμός των frames, που προβάλλονται στον χρήστη, ανά δευτερόλεπτο.

Για μια εφαρμογή η οποία πρέπει να δέχεται εικόνες από μία κάμερα, να της επεξεργάζεται και να της προβάλλει σε πραγματικό χρόνο, ένα κάτω φράγμα για τον αριθμό των frames που θα προβάλλονται στο χρήστη ανά δευτερόλεπτο μπορούν να θεωρηθούν τα 25 frames. Για μικρότερο αριθμό δε μπορεί να θεωρηθεί ότι η προβολή των frames γίνεται σε πραγματικό χρόνο, καθώς η ανανέωση των frames θα καθυστερεί αρκετά ώστε να γίνεται αντιληπτή από τον χρήστη.

Για να προσπεραστεί αυτό το πρόβλημα, προτείνεται η χρήση οπτικής ροής για παράκαμψη της διαδικασίας ταύτισης χαρακτηριστικών με τη χρήση του SURF, η οποία έχει περιγραφεί στο Κεφάλαιο 5. Συγκεκριμένα, η θέση του βιβλίου μπορεί να εντοπιστεί με χρήση της οπτικής ροής, εφόσον όμως το βιβλίο έχει ήδη ανιχνευτεί στο προηγούμενο frame.

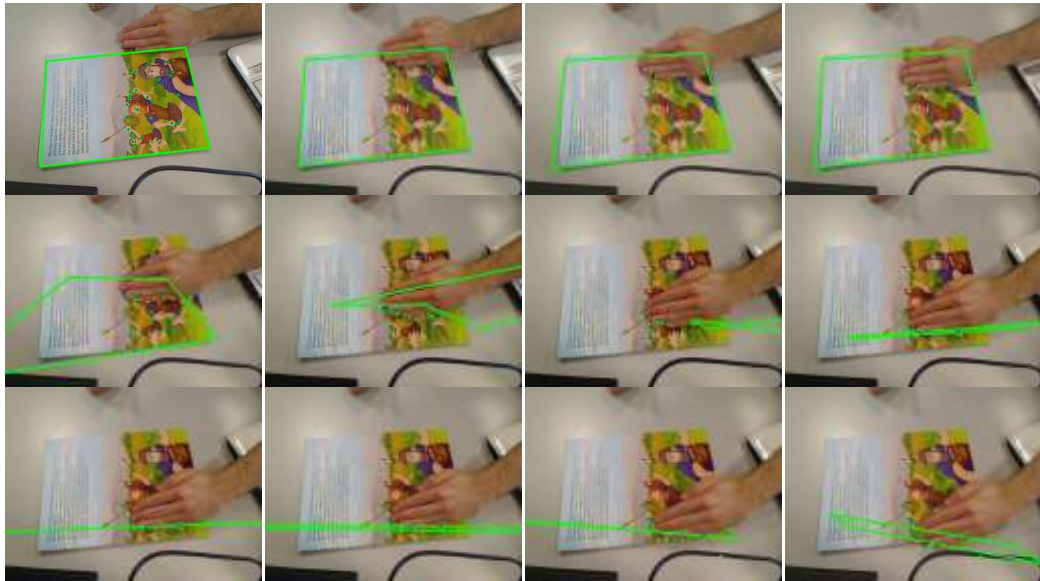
Για ένα frame στο οποίο έχει αναγνωριστεί μια σελίδα του βιβλίου, με χρήση του SURF, τότε στο επόμενο frame δίνονται τα σημεία, που αντιστοιχούν στα χαρακτηριστικά που έχουν ταυτιστεί, για να υπολογιστεί η νέα τους θέση βάσης της οπτικής ροής.

Αν ο εντοπισμός γίνει με επιτυχία, συνεχίζει η διαδικασία τοποθέτησης του εικονικού αντικειμένου μέσα στη σκηνή, και στη συνέχεια τα στοιχεία των οποίων η νέα θέση βρέθηκε με επιτυχία, δίνονται για εντοπισμό της νέας θέσης τους, στο επόμενο frame, με χρήση της οπτικής ροής και πάλι. Αν ο εντοπισμός της νέας θέσης των σημείων αποτύχει (αποτυχία θεωρείται η επιτυχία για πολύ μικρό αριθμό σημείων ή και καθόλου επιτυχία) τότε ο έλεγχος δίνεται και πάλι στο SURF, όπου γίνεται εξαγωγή χαρακτηριστικών από τη σκηνή και ταύτιση τους με αυτών του αντικειμένου.

8.4 Πειραματικά αποτελέσματα Οπτικής Ροής

Η χρήση της οπτικής ροής έδηξε να συμβάλει δραστικά στη μείωση του υπολογιστικού χρόνου που απαιτείται για την επεξεργασία ενός frame, εφόσον όμως έχει

εντοπιστεί πρώτα η σελίδα μέσα στη σκηνή με τη χρήση του SURF σε ένα προηγούμενο frame.



(α)



(β)

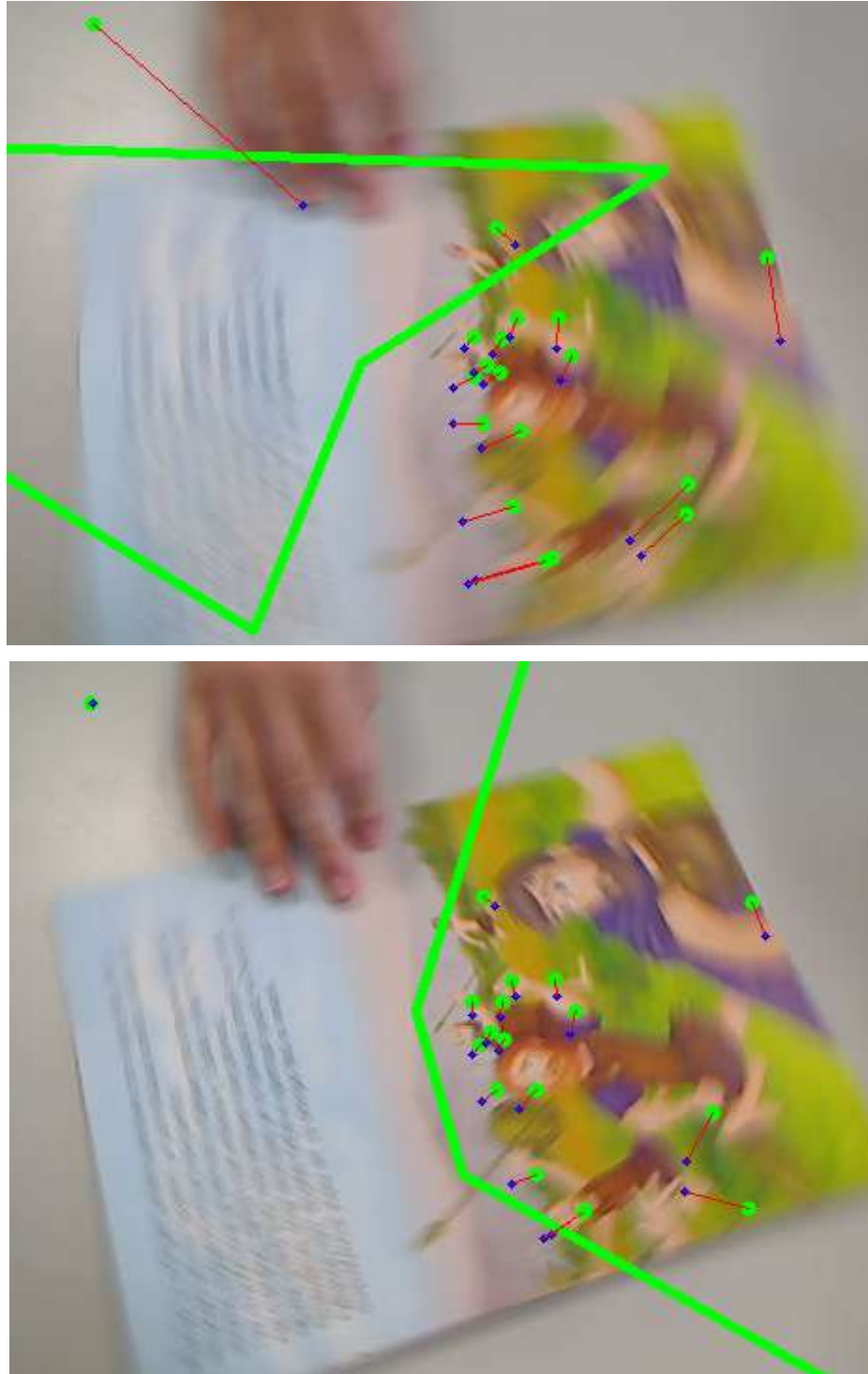
(γ)

Σχήμα 8.2 Παράδειγμα προβλήματος που παρουσιάζει η παρακολούθηση σημείων με χρήση κλασσικής προσέγγισης οπτικής ροής. Τα πράσινα στοιχεία αντιπροσωπεύουν τα σημεία που παρακολουθούνται. (α) Παρουσιάζεται η κίνηση του χεριού μέσα στη σκηνή. Στην (β) φαίνονται τα σημεία στη κανονική τους θέση, πριν την εισδοχή του χεριού μέσα στο χώρο του βιβλίου. Στην (γ) τα σημεία τα οποία έχει περάσει από κοντά τους το χέρι, το ακολουθούν. Με αποτέλεσμα ο υπολογισμός των ορίων της σελίδας μέσα στη σκηνή να είναι λανθασμένος.

Παρόλα αυτά, στην πράξη η χρήση της οπτικής ροής, επιφύλασσε κάποια σοβαρά προβλήματα τα οποία θα έπρεπε να επιλυθούν προκειμένου να μπορεί να χρησιμοποιηθεί αποτελεσματικά η τεχνική αυτή στην εφαρμογή.



Σχήμα 8.3 Ανίχνευση οπτικής ροής σε γρήγορη κίνηση. Δύο διαδοχικές εικόνες στις οποίες τα σημεία που είναι σημειωμένα με πράσινο δείχνουν τη θέση του σημείου στο προηγούμενο frame και με μπλε χρώμα τη θέση που ανιχνεύτηκε στο νέο frame. Συνέχεια διαδοχικών frames στο Σχήμα 8.4.



Σχήμα 8.4 Ανίχνευση οπτικής ροής σε γρήγορη κίνηση. Διαδοχικά frames, με αυτά στο Σχήμα 8.3, στα οποία μπορεί να παρατηρηθεί λανθασμένη ανίχνευση της νέας θέσης κάποιων σημείων, που προκύπτει από τη χρήση της κλασικής προσέγγισης της οπτικής ροής.

Το πρόβλημα που παρουσιάστηκε μέσα από τη πρακτική μελέτη της οπτικής ροής στην εφαρμογή ήταν ότι σε κάποιες περιπτώσεις σημεία υπολογίζονταν σε λάθος θέση. Το φαινόμενο αυτό παρατηρείται κυρίως όταν γίνονται γρήγορες μετακινήσεις του

βιβλίου, Σχήμα 8.3 και Σχήμα 8.4, και σε περιπτώσεις που ένα άλλο αντικείμενο περνούσε μπροστά από τέτοια στοιχεία, Σχήμα 8.2.

Συνεπώς όταν οι νέες θέσεις των σημείων υπολογίζονται λανθασμένα τότε συνεχίζει λανθασμένα και η τοποθέτηση του εικονικού αντικειμένου, και το λάθος μεταφέρεται και στα επόμενα frame μέχρι να λειτουργήσει ξανά το SURF για εντοπισμό της σελίδας στη σκηνή.

Για επίλυση αυτού του προβλήματος χρησιμοποιήθηκε μια προσέγγιση κατά την οποία, η οπτική ροή δεν υπολογίζεται μόνο προς μία χρονική κατεύθυνση αλλά προς δύο. Δηλαδή, υπολογίζεται η ροή από το προηγούμενο στο επόμενο frame και αντίστροφα από το επόμενο στο προηγούμενο. Ορθά αποτελέσματα θεωρούνται μόνο αυτά τα οποία ταυτίζονται στις δύο κατευθύνσεις.

Επομένως για ένα σημείο p με θέση (x,y) στο frame1 εντοπίζεται η νέα θέση του στο μεταγενέστερο frame2, όπου η θέση αυτή θα αντιστοιχεί στο σημείο $(x + dx, y + dy)$. Στη συνέχεια για το σημείο $(x + dx, y + dy)$ του frame2, εντοπίζεται η θέση του στο frame1 η οποία θα αντιστοιχεί στο σημείο $((x + dx) + dx, (y + dy) + dy)$. Έτσι, αν $dx=-dx$ και $dy=dy$ τότε $(x,y) = ((x + dx) + dx, (y + dy) + dy)$, και επομένως ο εντοπισμός της θέσης του σημείου p στο frame2 έχει γίνει με επιτυχία.

Αυτή η προσέγγιση πολλές φορές αναφέρεται ως bidirectional optical flow, και μπορεί να βελτιώσει σε μεγάλο βαθμό τα προβλήματα, που παρατηρούνται κατά χρήση της απλής προσέγγισης της οπτικής ροής, σε σχέση με την παρακολούθηση αντικειμένων.

8.5 Σύνοψη

Στα πλαίσια της προσπάθειας για βελτίωση της χρονικής απόδοσης του συστήματος της εργασίας αυτής, προτείνεται η χρήση της οπτικής ροής. Έτσι, οι διαδικασίες περιγραφής της σκηνής και ταύτισης των χαρακτηριστικών της με τα χαρακτηριστικά του αντικειμένου, μπορούν να παρακαμφθούν με τη χρήση της οπτικής ροής.

Γνωρίζοντας τη θέση κάποιων σημείων του αντικειμένου μέσα στο προηγούμενο frame που έχει ληφθεί από τη κάμερα, τα σημεία αυτά μπορούν να παρακολουθηθούν (tracked) μέσα στο καινούριο frame, με χρήση της οπτικής ροής και να βρεθεί η νέα τους θέση. Βάση της νέας θέσης των σημείων αυτόν μπορεί στη συνέχεια να υπολογιστεί ο νέος χώρος που καταλαμβάνει το αντικείμενο μέσα στη σκηνή και να γίνει η βαθμονόμηση της κάμερας ώστε να γίνει σε επόμενο βήμα η τοποθέτηση του εικονικού αντικειμένου στη σκηνή.

Η χρήση της οπτικής ροής προσφέρει γρηγορότερο εντοπισμό του αντικειμένου μέσα στη σκηνή, βασισμένη στον εντοπισμό της μετακίνησης του από το προηγούμενο στο επόμενο frame, σε σχέση με τον εντοπισμό του με τη χρήση του SURF, όπου καμία πληροφορία για την προηγούμενη του θέση δεν αξιοποιείται.

Η χρήση της οπτικής ροής βάσει της κλασικής προσέγγισης, όπου υπολογίζεται από το προηγούμενο στο επόμενο frame έχει δείξει αδυναμίες στη παρακολούθηση σημείων για γρήγορες κινήσεις του αντικειμένου, όπου η νέα θέση των σημείων πολλές φορές υπολογίζεται λανθασμένα. Επίσης, για περιπτώσεις που ένα άλλο αντικείμενο περάσει πάνω από σημεία που παρακολουθούνται, αυτά μπορούν να παρασυρθούν μαζί του.

Όλα αυτά μπορούν να αποτελέσουν βασικό λόγο στο λανθασμένο εντοπισμό του αντικειμένου με αποτέλεσμα το λάθος να προχωρά και στα επόμενα frames. Για το λόγο αυτό προτείνεται χρήση της bidirectional προσέγγισης της οπτικής ροής. Σε αυτή τη προσέγγιση πιθανών λανθασμένων εντοπισμών της νέας θέσης ενός σημείου οδηγεί στο τερματισμό της παρακολούθησης του εν λόγω σημείου.

Κεφάλαιο 9

Συμπεράσματα και Μελλοντικές Επεκτάσεις

9.1 Συμπεράσματα 74

9.2 Μελλοντικές Επεκτάσεις..... 76

9.1 Συμπεράσματα

Η εργασία αυτή έχει ασχοληθεί με τη μελέτη και υλοποίηση ενός συστήματος επαυξημένης πραγματικότητας για ανάγνωση παιδικών βιβλίων. Ο στόχος του συστήματος ήταν να λαμβάνει εικόνες - frames από τη κάμερα η οποία θα καταγράφει τη σκηνή, να της επεξεργάζεται ώστε να αναγνωρίζει αν μέσα στη εικόνα εμφανίζεται κάποια σελίδα του βιβλίου και να προβάλλει μια εικονική τρισδιάστατη σκηνή μέσα στο frame και τέλος να το προβάλλει στη οθόνη του χρήστη. Επίσης, μέσα στους στόχους ήταν, η αναγνώριση του βιβλίου να γίνεται χωρίς τη χρήση ειδικών αναγνωριστικών σημαδιών (markers) μέσα στη σελίδα του βιβλίου. Ακόμα, όλη αυτή η διαδικασία να διεκπεραιώνεται μέσα σε πραγματικό χρόνο, για κάθε frame που λαμβάνεται από τη κάμερα, ώστε να δίνεται στο χρήστη η αντίληψη πραγματικής ροής μέσα από τις προβαλλόμενες εικόνες και όχι απλά μιας προβολής σταθερών διαδοχικών εικόνων, με αντιληπτά χρονικά διαστήματα καθυστέρησης για την εναλλαγή τους. Για να μπορεί να θεωρείται ότι η διαδικασία γίνεται σε πραγματικό χρόνο, πρέπει να εκτελείται τουλάχιστον 25 φορές μέσα σε ένα δευτερόλεπτο, έτσι ώστε να προβάλλονται στο χρήστη 25 frames ανά δευτερόλεπτο (frames per second - fps)

Πρώτο και βασικό σημείο ήταν ο στόχος για αναγνώριση του βιβλίου μέσα στη σκηνή χωρίς τη χρήση αναγνωριστικών σημαδιών. Αυτό αυτόματα δημιουργεί διάφορους περιορισμούς σχετικά με την αναγνώριση των σελίδων μέσα στη σκηνή καθώς και με

τη φυσική εμφάνιση των σελίδων του βιβλίου. Για να γίνει αυτό πιο ξεκάθαρο, η μη χρήση ιδικών αναγνωριστικών, δεν επιτρέπει τη δυνατότητα χρήσης κάποιων υποθέσεων που μπορούν να βοηθήσουν στη γρήγορη ανίχνευσή τους. Παραδείγματος χάριν, λαμβάνοντας υπόψη ότι τα αναγνωριστικά αυτά περιέχουν μόνο μαύρες και άσπρες περιοχές, ο εντοπισμός τους μπορεί να εστιαστεί μόνο σε περιοχές της σκηνής, με μεγάλη αντίθεση φωτεινότητας (contrast). Επίσης, με τον εντοπισμό μιας περιοχής που πιθανών να αντιστοιχεί σε αναγνωριστικό, μπορεί να χρησιμοποιηθούν τα χαρακτηριστικά μόνο αυτής της περιοχής για αν συγκριθούν με αυτά των αναγνωριστικών που αναμένονται να ανιχνευτούν. Από την μεριά της φυσικής εμφάνισης των σελίδων οι σελίδες πρέπει περιέχουν ανιχνεύσιμα χαρακτηριστικά χωρίς μεγάλες ομοιογενείς περιοχές. Για παράδειγμα, μια άδεια σελίδα δε έχει κανένα ανιχνεύσιμο χαρακτηριστικό. Επίσης, οι σελίδες του ίδιου βιβλίου δε θα πρέπει να περιέχουν εικονογράφηση με χρήση ίδιων σχεδίων, μιας και αυτό μπορεί να προκαλέσει λανθασμένη αναγνώριση σελίδας.

Με τη χρήση της μεθόδου SURF το σύστημα περιγράφει τις σελίδες του βιβλίου και στη συνέχεια προχωρεί στη λήψη εικόνων, καταγραφών της σκηνής, από τη κάμερα. Κάθε εικόνα που λαμβάνεται από τη κάμερα περιγράφεται. Για κάθε περιγραφή αναγνωρίζονται μέσα στην εικόνα σημεία τα οποία έχουν ιδιαίτερο ενδιαφέρον και περιγράφονται ένα - ένα ξεχωριστά. Το σύνολο των περιγραφών των χαρακτηριστικών που έχουν ανιχνευτεί σε μία εικόνα αποτελούν και τη περιγραφή της εικόνας.

Στη συνέχεια, αρχίζει η σύγκριση της περιγραφής της σκηνής με τις περιγραφές των σελίδων του βιβλίου. Η σύγκριση δύο περιγραφών είναι ανάλογη των χαρακτηριστικών που έχουν ανιχνευτεί μέσα σε κάθε εικόνα, έχοντας κάθε χαρακτηριστικό της μίας να συγκρίνεται με κάθε χαρακτηριστικό της άλλης. Η διαδικασία για σύγκριση δύο περιγραφών έχει πολυπλοκότητα της τάξης $O(n^2)$. Σε ένα μέσω υπολογιστή, μια μόνο σύγκριση τρέχει σε αρκετά ικανοποιητικό χρόνο ώστε η όλη επεξεργασία και προβολή κάθε frame να γίνεται σε πραγματικό χρόνο. Έχοντας όμως ένα βιβλίο με 15 σελίδες για παράδειγμα, η ανίχνευση της 15^{ης} σελίδας προϋποθέτει και 15 συγκρίσεις περιγραφών σκηνής – σελίδας. Η σύγκριση μεγάλου αριθμού σελίδων κατά τη διάρκεια επεξεργασίας ενός frame, προκαλεί μεγάλη καθυστέρηση με αποτέλεσμα η προβολή του να μη γίνεται σε πραγματικό χρόνο.

Έτσι, στην εργασία αυτή προτείνεται η χρήση οπτικής ροής, για εύρεση της νέας θέσης ενός αντικειμένου, το οποίο είχε ανιχνευτεί επιτυχώς στο αμέσως προηγούμενο frame. Η χρήση της οπτικής ροής έχει ήδη να βοηθάει κατά πολύ το χρόνο προβολής των διαδοχικών frames καταφέροντας να φτάσει μέχρι και 45 fps. Παρόλα αυτά, για να μπορεί να γίνει η χρήση της οπτικής ροής θα πρέπει το αντικείμενο να έχει ανιχνευτεί πρώτα μια φορά με τη χρήση των περιγραφών του SURF, έτσι μέχρι τη πρώτη αναγνώριση για κάθε frame θα πρέπει να υπολογίζεται η περιγραφή του, και να συγκρίνεται με την περιγραφή κάθε σελίδας. Επίσης, αν η οπτική ροή αποτύχει να εντοπίσει τη νέα θέση του αντικειμένου τότε ο έλεγχος δίνεται και πάλι στο SURF.

Επομένως, στις περιπτώσεις που το σύστημα δε δουλεύει με τη χρήση της οπτικής ροής το πρόβλημα δε λύνεται. Για το λόγο αυτό, στο σύστημα λαμβάνονται οι εξής περιορισμοί: η πρώτη σελίδα που θα αναγνωρίζεται κατά τη εκκίνηση της εφαρμογής θα είναι η πρώτη σελίδα του βιβλίου. Ακόμα, κάθε φορά (εκτός από τη στιγμή της εκκίνησης μέχρι τη πρώτη αναγνώριση) που η αναγνώριση του βιβλίου θα γίνεται με τη χρήση του SURF, θα γίνονται μόνο τρεις συγκρίσεις περιγραφών. Μια σύγκριση για τη σελίδα που είχε αναγνωριστεί την τελευταία φορά, μια για την επόμενη της σελίδα, καθώς και μια σύγκριση για την προηγούμενη σελίδα.

9.2 Μελλοντικές Επεκτάσεις

Λόγο του περιορισμένου χρόνου για την ανάπτυξη του συστήματος, στο παρόν στάδιο η επαύξηση της σκηνής γίνεται με ένα και μόνο εικονικό αντικείμενο. Σε μελλοντικό στάδιο η επαύξηση μπορεί να γίνεται με μια ολόκληρη εικονική σκηνή για κάθε σελίδα. Όπου θα μπορούν να υπάρχουν αντικείμενα που να κινούνται, όπως για παράδειγμα χαρακτήρες ανθρώπων. Επίσης, θα μπορεί να γίνει χρήση ηχητικών εφέ, ή και ακόμα αφήγηση του κομματιού της ιστορία που εξιστορείται στη κάθε σελίδα.

Ακόμη, η αναγνώριση της κατανομής του φωτός μέσα στη σκηνή για εφαρμογή κατάλληλου φωτισμού στα τρισδιάστατα μοντέλα, που επαυξάνουν την σκηνή, θα μπορούσε να εναρμονίσει καλύτερα τα εικονικά αντικείμενα με την καταγραφή του

πραγματικού κόσμου. Δίνοντας έτσι, μια πιο ρεαλιστική άποψη της μικτής πραγματικότητας της εφαρμογής.

Θα μπορούσε επίσης να επεκταθεί ώστε να δίνει τη δυνατότητα στον αναγνώστη – παιδί να παίξει κάποια παιχνίδια, για παράδειγμα να βοηθήσει τον ήρωα του βιβλίου να εκπληρώσει το σκοπό του. Κάνοντας έτσι τη ανάγνωση του βιβλίου πιο διασκεδαστική και ενδιαφέρων, δίνοντας του ταυτόχρονα τη δυνατότητα να εμπεδώσει καλύτερα την ιστορία ή αυτά που προσπαθεί να του διδάξει το βιβλίο.

Αυτή τη στιγμή η επαυξημένη πραγματικότητα έχει αρχίσει να κάνει τα πρώτα βήματα της στην αγορά. Οι εφαρμογές και οι συσκευές που προσφέρουν ψυχαγωγία με χρήση μικτής πραγματικότητας αυξάνονται καθημερινά, και συνεχώς εμφανίζονται νέες και καινοτομικές ιδέες για χρήση της.

Μικτή πραγματικότητα είναι ένας κόσμος στον οποίο συνυπάρχει η φαντασία με την πραγματικότητα. Τα όρια της πραγματικότητας μπορούν να γίνουν εύκολα αντιληπτά από τον καθένα, τα όρια της μικτής πραγματικότητας φτάνουν μέχρι εκεί που σταματά η φαντασία μας. Με φαντασία και δημιουργικότητα λοιπόν οι επεκτάσεις μιας τέτοιας εφαρμογής είναι άπειρες.

Βιβλιογραφία

- [1] Bay, H., Tuytelaars, T., & Van Gool, L. (2008). SURF: Speeded-Up Robust Features. *Computer Vision and Image Understanding*, 346-359.
- [2] Beauchemin, S. S., & Barron, J. L. (1995). The Computation of Optical Flow. *ACM Computing Surveys* , 433-467.
- [3] Bouguet, J.-Y. (2010, July 9). *Camera Calibration Toolbox for Matlab*. Ανάκτηση από Computer Vision at Caltech: http://www.vision.caltech.edu/bouguetj/calib_doc/
- [4] Hiriart-Urruty, J.-B., Strodiot, J.-J., & Nguyen, V. H. (1984). Generalized Hessian matrix and second-order optimality conditions for problems with C. *Applied Mathematics & Optimization* , 43-56.
- [5] Horn, B. K., & Schunck, B. G. (1981). Determining optical flow. *Artificial Intelligence* , 185-203.
- [6] Jolliffe, I. T. (1986). *Principal Component Analysis*. Springer-Verlag.
- [7] Ke, Y., & Sukthankar, R. (2004). A More Distinctive Representation for Local Image Descriptors. *Conference on Computer Vision and Pattern Recognition*, (σσ. 511-517). Washington.
- [8] Lowe, G. D. (2004). Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision* , 2 (60), 91-110.
- [9] Lucas, B. D., & Kanade, T. (1981). An iterative image registration technique with an application to stereo vision. *Proceedings of Imaging Understanding Workshop*, (σσ. 121-130).

- [10] Luo, J., & Oubong, G. (2009). A Comparison of SIFT, PCA-SIFT and SURF. *International Journal of Image Processing* , 143–152.
- [11] Mikolajczyk, K., & Schmid, C. (2005). A performance evaluation of local descriptors. *Pattern Analysis and Machine Intelligence* , 1615-1630.
- [12] *NeHe Productions*. Ανάκτηση από GameDev.net: <http://nehe.gamedev.net/>
- [13] Viola, P., & Jones, M. (2001). Rapid Object Detection using a Boosted Cascade of Simple. *Computer Vision and Pattern Recognition*, (σσ. 511-518).
- [14] <http://opencv.willowgarage.com/wiki/>
- [15] <http://www.opengl.org/>

