



## **Τεχνολογικό Εκπαιδευτικό Ίδρυμα Κρήτης**

**Σχολή Τεχνολογικών Εφαρμογών  
Τμήμα Εφαρμοσμένης Πληροφορικής & Πολυμέσων**

### **Πτυχιακή Εργασία**

**Αναγνώριση αντικείμενων με χρήση χαρτών βάθους και τεχνολογιών  
μάθησης μηχανής**

**Κουναλάκης Τσαμπίκος (ΑΜ 2141)**

**Επιβλέπων καθηγητής : Τριανταφυλλίδης Γεώργιος**

**Επιτροπή αξιολόγησης : Παπαδουράκης Γεώργιος  
Τριανταφυλλίδης Γεώργιος  
Φυσαράκης Κωνσταντίνος**

**Ημερομηνία Παρουσίασης : 11 Απριλίου 2011**

## **Abstract**

This paper presents a novel system that is fusing efficient and state-of-the-art techniques of stereo vision and machine learning, aiming at object detection and recognition. To this goal, the system initially creates depth maps by employing the Graph-Cut technique. Then, the depth information is used for object detection by separating the objects from the whole scene. Next, the Scale-Invariant Feature Transform (SIFT) is used, providing the system with unique features for each extracted object, which are used in training an Artificial Neural Network (ANN). The system is then able to classify and recognize the nature of these object, creating knowledge from the real world.

## Σύνοψη

Αυτή η εργασία παρουσιάζει ένα πρωτοποριακό σύστημα το οποίο ενώνει αποτελεσματικές και καινοτόμες τεχνικές της υπολογιστικής στερεοσκοπικής όρασης και της μάθησης μηχανής, με στόχο την ανίχνευση και αναγνώριση αντικείμενων. Για αυτόν τον σκοπό, το σύστημα αρχικά δημιουργεί χάρτες βάθους με την χρήση τεχνικών τομών γράφων (Graph-Cut). Τότε, η πληροφορία του βάθους χρησιμοποιήτε για την ανίχνευση αντικειμένων με τον διαχωρισμό των αντικειμένων από την υπόλοιπη σκηνή. Εν συνεχεία, χρησιμοποιήτε ο μετασχηματισμός κλιμακωτά – αναλλοίωτων χαρακτηριστικών (Scale-Invariant Feature Transform , SIFT), παρέχοντας στο σύστημα με μοναδικά χαρακτηριστικά για κάθε αντικείμενο, που χρησιμοποιούνται στην εκπαίδευση ενός τεχνητού νευρωνικού δικτύου. Το σύστημα είναι τότε ικανό να κατηγοριοποιήσει και να αναγνωρίσει την φύση των αντικειμένων, παράγοντας γνώση από τον πραγματικό κόσμο.

## Κατάλογος περιεχομένων

1.Εισαγωγή .....	8
1.1 Περίληψη .....	8
1.2 Κίνητρο για την διεξαγωγή της εργασίας .....	10
1.3 Σκοπός και στόχοι της εργασίας .....	11
1.4 Δομή εργασίας.....	11
2.Μεθοδολογία Υλοποίησης.....	13
2.1 Μέθοδος ανάλυσης και ανάπτυξης πτυχιακής.....	13
2.2 Μοντέλα .....	14
2.2.1 Στερεοσκοπική όραση .....	15
2.3 Θεωρίες .....	18
2.3.1 Τεχνητή όραση.....	19
2.3.2 Τεχνητή νοημοσύνη.....	22
2.3.3 Θεωρία των γράφων .....	26
2.4 Αλγόριθμοι .....	30
2.4.1 Υπολογισμός οπτικής αντιστοιχίας μέσω τομών σε γράφους (Graph cuts).....	31
2.4.2 Αλγόριθμος συνδεδεμένων στοιχείων .....	41
2.4.3 Μετασχηματισμός κλιμακωτά – αναλλοίωτων χαρακτηριστικών (SIFT) .....	45
2.4.4 Τεχνητά νευρωνικά δίκτυα .....	56
3.Σχέδιο δράσης για την εκπόνηση της πτυχιακής εργασίας.....	61
3.1 Τεχνολογίες ακμής (state-of-the-art) .....	61
3.2 Σημαντικοί στόχοι εκπόνησης της πτυχιακής εργασίας.....	66
4.Κύριο μέρος Πτυχιακής εργασίας.....	68
4.1 Ανάλυση Προβλημάτων .....	68
4.1.1 Προβλήματα λήψης και επεξεργασίας εικόνας .....	68
4.1.2 Προβλήματα αναγνώρισης και κατηγοριοποίησης αντικειμένων.....	70
4.1.3 Προβλήματα παραγωγής γνώσης και αναγνώρισης.....	71
4.1.4 Απαιτήσεις του συστήματος.....	71
4.2 Σχεδιασμός υλοποίησης.....	72
4.3 Υλοποίηση του συστήματος.....	74
4.3.1 Σχεδιάγραμμα αλγορίθμου συστήματος.....	74
4.3.2 Περιγραφή υλοποίησης συστήματος.....	75
5.Αποτελέσματα.....	82
5.1 Συμπεράσματα και μελλοντική εργασία .....	85

## Κατάλογος εικόνων

Εικόνα 1: Το οπτικό σύστημα, το συνδεδετικό μονοπάτι και ο οπτικός φλοιός .....	8
Εικόνα 2: Σχέσεις μεταξύ τεχνητής όρασης και άλλων επιστημονικών πεδίων .....	9
Εικόνα 3: Παράδειγμα παραγωγής εικόνας από μικρότερα εικονοστοιχεία. ....	15
Εικόνα 4: Διαφορά στην οπτική γωνία που παράγει διαφορετικές εικόνες.....	16
Εικόνα 5: Στερεοσκοπικό ζεύγος εικόνων .....	16
Εικόνα 6: Πίνακας διαφοράς ή χάρτης βάθους σε τρισδιάστατη απεικόνιση .....	17
Εικόνα 7: Εικόνα μέτρησης μεγέθους οργάνου μέσω ενεργού περιγράμματος (snakes).....	19
Εικόνα 8: Ανίχνευση και κατηγοριοποίηση κυττάρων .....	20
Εικόνα 9: Απεικόνιση του NASA's Mars Exploration Rover.....	21
Εικόνα 10: Μέθοδος τρισδιάστατης ανασυγκρότησης σκηνης.....	21
Εικόνα 11: Ακολουθία βημάτων κατασκευής συστήματος τεχνητής όρασης .....	22
Εικόνα 12: Δίκτυο από οντολογίες και οι μεταξύ τους δεσμοί.....	24
Εικόνα 13: Ο ASIMO εφαρμόζει πολλές τεχνικές τεχνητής νοημοσύνης .....	25
Εικόνα 14: Ο Kismet είναι ένα ρομπότ που προσομοιώνει ανθρώπινα συναισθήματα .....	26
Εικόνα 15: Παράδειγμα γράφου .....	27
Εικόνα 16: Το πρόβλημα των 7 γεφυρών της πόλης Königsberg .....	27
Εικόνα 17: Απλό παράδειγμα γράφου του προβλήματος των τεσσάρων χρωμάτων .....	28
Εικόνα 18: Απεικόνιση βάσεων δεδομένων με την βοήθεια γράφου.....	29
Εικόνα 19: Συνάρτηση ενέργειας τομών σε γράφους από τον Greig .....	31
Εικόνα 20: Παράδειγμα κατηγοριοποίησης με γράφους.....	31
Εικόνα 21: Παράδειγμα ελάχιστης τομής σε κατευθυνόμενο γράφο .....	32
Εικόνα 22: Σχέση μεταξύ των προβλημάτων μεγίστης ροής και ελάχιστης τομής σε παράδειγμα κατάτμηση της εικόνας .....	33
Εικόνα 23: Πίνακας Διαφοράς από κλασικό αλγόριθμο.....	34
Εικόνα 24: Η ενέργεια που ελαχιστοποιείτε στο μοντέλο του Pott [19] .....	34
Εικόνα 25: Ενέργεια μοναδικής διαμόρφωσης $f$ .....	35
Εικόνα 26: Τύπος όρου κρυφών σημείων $E_{occ}$ .....	35
Εικόνα 27: Απλός όρος ομαλότητας $E_{smooth}$ .....	36
Εικόνα 28: Εξελιγμένος όρος ομαλότητας $E_{smooth}$ .....	36
Εικόνα 29: Βήματα του αλγορίθμου επέκτασης.....	36
Εικόνα 30: Μεταποίηση μιας εικόνας σε ένα γράφο .....	37
Εικόνα 31: Αρχική ανάθεση κόστους κρυφών σημείων .....	38
Εικόνα 32: Αρχικό κόστος ομαλοποίησης .....	38
Εικόνα 33: Πίνακας βαρών για κάθε διαμόρφωση.....	38
Εικόνα 34: Τύπος κόστους συνδέσμων-t .....	39
Εικόνα 35: Τύπος κόστους συνδέσμων-c.....	39
Εικόνα 36: Ενέργεια κρυφών σημείων για διαμόρφωση $fC$ .....	39
Εικόνα 37: Σύνολο ενέργειας κρυφών σημείων για διαμόρφωση $fC$ .....	39
Εικόνα 38: Γενικό σύνολο εξισώσεων διαμόρφωσης $fC$ .....	39
Εικόνα 39: Τύπος ενέργειας για $\alpha$ - έκταση του $fC$ .....	40
Εικόνα 40: Δεύτερο μέρος τύπου ενέργειας για $\alpha$ - επέκταση του $fC$ .....	40
Εικόνα 41: Κόστη που προκύπτουν από τύπο ενέργειας για $\alpha$ - επέκταση του $fC$ .....	40
Εικόνα 42: Ανάπτυξη μέρους πρώτου εικόνα 41 .....	40
Εικόνα 43: Ανάπτυξη μέρους δεύτερου εικόνας 41 .....	40
Εικόνα 44: Τελική απόδειξη θεωρήματος.....	41
Εικόνα 45: γράφος με τρία συνδεδεμένα στοιχεία .....	41

Εικόνα 46: Επισήμανση των συνδεδεμένων στοιχείων σε 8 σημεία.....	43
Εικόνα 47: Επισήμανση των συνδεδεμένων στοιχείων σε 4 σημεία.....	43
Εικόνα 48: Παράδειγμα επισήμανσης των συνδεδεμένων στοιχείων σε δυαδική εικόνα με δυο περιοχές.....	43
Εικόνα 49: Παράδειγμα αλγορίθμου επισήμανσης των συνδεδεμένων στοιχείων δυο περασμάτων σε δυαδική εικόνα .....	44
Εικόνα 50: Κλιμακωτός χώρος μιας εικόνας $L(x, y, \sigma)$ .....	46
Εικόνα 51: Συνέλιξη ενός μεταβλητής κλίμακας Gaussian, $G(x, y, \sigma)$ .....	47
Εικόνα 52: Ακρότατα κλιμακωτού χώρου σε λειτουργία διαφοράς- του - Gaussian σε συνέλιξη με την εικόνα $D(x, y, \sigma)$ .....	47
Εικόνα 53: Αναπαράσταση συνέλιξης και αφαίρεσης κάθε οκτάβας για την δημιουργία διαφοράς - της - Gaussian .....	47
Εικόνα 54: Οι σχέσεις μεταξύ του $D$ και του $\sigma^2 \square 2G$ .....	48
Εικόνα 55: Υπολογισμός πεπερασμένης διαφορετικής προσέγγισης στο $\theta G/\theta \sigma$ , χρησιμοποιώντας την διαφορά των γύρω κλιμάκων στο $k$ .....	48
Εικόνα 56: Διαφορά- της - Gaussian λειτουργίας οι οποίες διαφέρουν από έναν σταθερό παράγοντα.....	48
Εικόνα 57: Ο υπολογισμός του μέγιστου και ελάχιστου μιας διαφοράς- της - Gaussian που ανιχνεύετε με σύγκριση ενός εικονοστοιχείου .....	49
Εικόνα 58: Η προσέγγιση με επέκταση του Taylor της λειτουργίας κλιμακωτού χώρου , $D(x, y, \sigma)$ .....	51
Εικόνα 59: Η τοποθεσία του μέγιστου , $\hat{x}$ .....	51
Εικόνα 60: Η τιμή της λειτουργίας στο μέγιστο, $D(\hat{x})$ .....	51
Εικόνα 61: Τα αποτελέσματα της επιλογής σημείων κλειδιών σε μια εικόνα.....	52
Εικόνα 62: Υπολογισμός κύριας καμπυλότητας από ένα $2 \times 2$ Hessian πίνακα , $H$ .....	53
Εικόνα 63: Υπολογισμός του σύνολου των ιδιοτιμών από το ίχνος του $H$ και τα προϊόντα του από την ορίζουσα .....	53
Εικόνα 64: Υπολογισμός αναλογίας ανάμεσα στην μεγαλύτερης σημασίας ιδιοτιμή και την μικρότερη σημασίας .....	53
Εικόνα 65: Ο έλεγχος της αναλογίας των κύριων καμπυλοτήτων.....	53
Εικόνα 66: Υπολογισμός κλίμακας κλίσης μεγέθους, $m(x,y)$ , και προσανατολισμού, $\theta(x,y)$ .....	54
Εικόνα 67: Ο υπολογισμός του περιγραφέα σημείων κλειδιών .....	55
Εικόνα 68: Τεχνητό νευρωνικό δίκτυο ενός επιπέδου .....	56
Εικόνα 69: Κλασική μορφή τεχνητού νευρωνικού δικτύου .....	57
Εικόνα 70: Τεχνητό νευρωνικό δίκτυο γράφημα εξάρτησης.....	58
Εικόνα 71: Επαναλαμβανόμενο τεχνητό νευρωνικό δίκτυο γράφημα εξάρτησης .....	59
Εικόνα 72: Ένας από τους δορυφόρους του προγράμματος STEREO της NASA .....	61
Εικόνα 73: Πρώτη εικόνα του προγράμματος STEREO της NASA.....	62
Εικόνα 74: Σύστημα EyeSight από την εταιρία αυτοκινήτων Subaru.....	62
Εικόνα 75: Αποτέλεσμα συστήματος EyeSight.....	63
Εικόνα 76: Ρομπότ το οποίο πλοηγεί στον χώρο χάρης τον αλγόριθμο SIFT .....	63
Εικόνα 77: Τα ίχνη του Mars Rover της NASA στην επιφάνεια του Άρη.....	64
Εικόνα 78: Ράψιμο των εικόνων αεροφωτογραφίας.....	64
Εικόνα 79: Σύννεφο σημείων το οποίο απεικονίζει μια τρισδιάστατη σκηνή.....	65
Εικόνα 80: Ρομπότ Mamoru που βοηθάει ηλικιωμένους .....	65
Εικόνα 81: Εμφάνιση κρυφών σημείων στην εικόνα του πειράματος .....	69
Εικόνα 82: Πρόβλημα σχηματισμού αντικειμένων που άνηκαν σε διαφορετικά επίπεδα .....	70
Εικόνα 83: Δείγμα του επιθυμητού αποτελέσματος από την αναγνώριση των αντικειμένων.....	70
Εικόνα 84: Αρχικό διάγραμμα σχεδιασμού υλοποίησης .....	72

Εικόνα 85: Τελικό διάγραμμα σχεδιασμού υλοποίησης .....	73
Εικόνα 86: Σχεδιάγραμμα αλγορίθμου συστήματος.....	74
Εικόνα 87: Χάρτης βάθους του πειραματικού ζεύγους εικόνων .....	75
Εικόνα 88: Βάθος 1 από επεξεργασμένο χάρτη βάθους.....	76
Εικόνα 89: Βάθος 2 από επεξεργασμένο χάρτη βάθους.....	76
Εικόνα 90: Βάθος 3 από επεξεργασμένο χάρτη βάθους.....	77
Εικόνα 91: Βάθος 4 από επεξεργασμένο χάρτη βάθους.....	77
Εικόνα 92: Βάθος 5 από επεξεργασμένο χάρτη βάθους.....	78
Εικόνα 93: Ενοποίηση βάθους 3 και 4.....	78
Εικόνα 94: Παράδειγμα συλλογής αντικειμένου από επίπεδο βάθους.....	79
Εικόνα 95: Παράδειγμα αντιστοίχισης αντικειμένου στην πραγματική του εικόνα .....	79
Εικόνα 96: Σημεία κλειδιά μετασχηματισμού SIFT αντικείμενο 1 .....	80
Εικόνα 97: Σημεία κλειδιά μετασχηματισμού SIFT αντικείμενο 2.....	80
Εικόνα 98: Σημεία κλειδιά μετασχηματισμού SIFT αντικείμενο 3.....	80
Εικόνα 99: Σημεία κλειδιά μετασχηματισμού SIFT αντικείμενο 4.....	81
Εικόνα 100: Σημεία κλειδιά μετασχηματισμού SIFT αντικείμενο 5.....	81
Εικόνα 101: Σημεία κλειδιά μετασχηματισμού SIFT αντικείμενο 6.....	81
Εικόνα 102: Οπτική αναπαράσταση αποτελεσμάτων του πίνακα 1 .....	82
Εικόνα 103: Τρεις άγνωστες εικόνες από λάμπες που χρησιμοποιήθηκαν στην τελική επαλήθευση των αποτελεσμάτων.....	84

## Ευρετήριο πινάκων

Πίνακας 1: Πίνακας αποτελεσμάτων ταιριάσματος SIFT με την χρήση γρήγορου αλγόριθμου του κοντινότερου γείτονα.....	89
Πίνακας 2: Πίνακας αποτελεσμάτων ταιριάσματος SIFT με την χρήση τεχνητού νευρωνικού δικτύου (πρώτη προσπάθεια).....	90
Πίνακας 3: Πίνακας αποτελεσμάτων ταιριάσματος SIFT με την χρήση τεχνητού νευρωνικού δικτύου (δεύτερη προσπάθεια) .....	90
Πίνακας 4: Πίνακας αποτελεσμάτων δοκιμής με άγνωστα αντικείμενα.....	91
Πίνακας 5: Πίνακας τελικών αποτελεσμάτων .....	92

## 1. Εισαγωγή

Η πτυχιακή αυτή εργασία είναι η περιγραφή ενός πρωτότυπου συστήματος. Η ιδέα της εργασίας αυτής είναι προσωπική και βασίστηκε πάνω στην παρατήρηση φαινομένων του περιβάλλοντος. Η αρχική απορία που έδωσε το έναυσμα για αυτό το σύστημα, ήταν ο τρόπος με τον οποίο οι οργανισμοί είναι σε θέση να αντιληφθούν το περιβάλλον μέσω της λειτουργίας της όρασης. Αργότερα σε αυτό το ερώτημα προστέθηκε και ο τρόπος των οργανισμών να παράγουν συνδέσμους ανάμεσα σε αυτές εικόνες και η μετατροπή τους σε γνώσεις.

Έτσι αποφάσισα την προσομοίωση αυτών των δυο παρατηρήσεων με τις επιστήμες της τεχνητής όρασης και τεχνητής νοημοσύνης. Έτσι σχεδιάζα την χρήση των ποιο πρωτοποριακών αλγορίθμων με στόχο την επίλυση των δυο παραπάνω ερωτημάτων. Αποτέλεσμα θα ήταν η παραγωγή ενός εντελώς πρωτοποριακού συστήματος το οποίο θα συνδύαζε την όραση με την μάθηση σε ένα υπολογιστικό σύστημα.

### 1.1 Περίληψη

Η επιστήμη υπολογιστών γνωρίζει ραγδαία ανάπτυξη τα τελευταία χρόνια. Η ανάγκη των συστημάτων πληροφορικής να γίνουν ποιο ακριβή και να αλληλεπιδρούν με το περιβάλλον των χρηστών ανάγκασε στην προσομοίωση βιολογικών λειτουργιών. Έτσι η επιστήμη υπολογιστών απέκτησε κλάδους οι οποίοι περιγράφουν και εξομοιώνουν βιολογικές λειτουργίες όπως η όραση, η ακοή και η παραγωγή γνώσης.

Σε αυτήν την εργασία γίνεται η προσπάθεια της εξομοίωσης της βιολογικής όρασης αλλά και παραγωγής γνώσης από αυτήν. Το σύστημα το οποίο επιτρέπει την βιολογική όραση είναι ο συνδυασμός ενός οπτικού συστήματος και ενός συστήματος ερμηνείας της οπτικής πληροφορίας.

Το οπτικό σύστημα είναι τμήμα του κεντρικού νευρικού συστήματος και επιτρέπει στους οργανισμούς επεξεργαστούν οπτικές λεπτομέρειες. Οι πληροφορίες αποτελούνται από το ορατό φως για την οικοδόμηση μιας αναπαράστασης του περιβάλλοντος [1]. Τα τρία βασικά στοιχεία ενός οπτικού συστήματος στα θηλαστικά αποτελείται από τον οφθαλμό, το οπτικό νεύρο και το οπτικό φλοιό. Ο οφθαλμός είναι το όργανο το οποίο μετατρέπει το ορατό φως σε πληροφορία η οποία μπορεί να επεξεργαστεί από τον οργανισμό. Το οπτικό νεύρο είναι το κανάλι επικοινωνίας του οφθαλμού και του οπτικού φλοιού, έτσι μεταφέρονται και οι πληροφορίες από το περιβάλλον στο εγκέφαλο.

Ο οπτικός φλοιός είναι ένα κομμάτι του εγκεφάλου το οποίο είναι υπεύθυνο για την επεξεργασία της πληροφορίας. Αυτή η επεξεργασία αποκαλείται οπτική αντίληψη. Η οπτική αντίληψη είναι η ικανότητα τις ερμηνείας των πληροφοριών του περιβάλλοντος από τις επιπτώσεις που προκαλεί το ορατό φως στον οφθαλμό. Η αντίληψη η οποία προκύπτει ονομάζεται όραση [2].





δεν έχουν υλική ύπαρξη από υπηρεσίες στο διαδίκτυο μέχρι υπηρεσίες στην ιατρική. Αλλά και η τεχνητή νοημοσύνη έχει πολλούς κλάδους και αυτός που μας ενδιαφέρει είναι η μηχανική μάθηση.

Η μηχανική μάθηση είναι ένας επιστημονικός κλάδος που ασχολείται με τον σχεδιασμό και την ανάπτυξη αλγορίθμων που επιτρέπουν στους υπολογιστές να εξελίσσουν συμπεριφορές που βασίζονται σε εμπειρικά δεδομένα. Αυτά τα δεδομένα βασίζονται σε δεδομένα αισθητήρων ή βάσης δεδομένων[5]. Στηρίζετε στην θεωρία ότι ένα σύστημα επωφελείται από τα παραδείγματα του προβλήματος τα οποία μαθαίνει μέσω μιας επανάληψης αλλά και μέσω των λαθών του. Άρα το σύστημα από μόνο του σε κάποιο ορισμένο χρόνο αποκτά εμπειρία και είναι σε θέση να διαχωρίσει καταστάσεις ή δεδομένα.

Σε αυτήν την εργασία χρησιμοποιήθηκαν όλες οι παραπάνω τεχνικές για την επίτευξη του στόχου. Έτσι προσομοιώνεται η ανθρώπινη όραση με εικόνες από μια στερεοσκοπική κάμερα. Η στέρεο οπτική είναι η διαδικασία που οδηγεί στην αντίληψη του βάθους από δυο ελαφρώς διαφορετικές προβολές του κόσμου πάνω στον αμφιβληστροειδή των δυο ματιών. Οι διαφορές στις δυο εικόνες ονομάζετε οριζόντια ανισότητα ή ανισότητα του αμφιβληστροειδούς [6]. Η στερεοσκοπική κάμερα χρησιμοποιεί αυτό το φαινόμενο για την αντίληψη του βάθους και δημιουργεί χάρτες βάθους.

Η αντίληψη του βάθους στο σύστημα μου χρησιμοποιείτε για την αντίληψη αντικειμένων στο φυσικό περιβάλλον. Χάρη τους χάρτες βάθους που παράγονται το σύστημα μπορεί εύκολα να ανιχνεύσει αντικείμενα να τα ταξινομήσει ανάλογα με το βάθος που βρίσκονται τα αντικείμενα στην εικόνα. Επίσης το σύστημα μπορεί να κατανοήσει το φυσικό τους μέγεθος το οποίο χρησιμοποιεί αργότερα από τον αλγόριθμο. Οι χάρτες βάθους παράχθηκαν με έναν αλγόριθμο ο οποίος θεωρείτε ο καλύτερος του είδους του. Ο αλγόριθμος αυτός ονομάζετε κοπή γράφων (Graph cuts [7]) και χρησιμοποιείτε στον τομέα της τεχνητής όρασης για μια μεγάλη ποικιλία προβλημάτων χαμηλής όρασης. Το αποτέλεσμα αυτού του αλγορίθμου είναι ένας πολύ ακριβής χάρτης βάθους.

Στην συνέχεια αυτός ο χάρτης δέχεται μια επεξεργασία. Στην επεξεργασία εξετάζετε κάθε επίπεδο βάθους το οποίο παρήγαγε ο προηγούμενος αλγόριθμος. Έτσι επιλέγουμε μέσω ενός συστήματος ψηφοφορίας εάν σε κάθε ένα από τα επίπεδα υπάρχει αντικείμενο. Αυτά τα επίπεδα ονομάζονται υποψήφια επίπεδα. Αργότερα τα επίπεδα αυτά επιλέγονται και ομαδοποιούνται εάν κριθούν ότι περιγράφουν το ίδιο αντικείμενο.

Αφού οι περιοχές των υποψηφίων επιπέδων της εικόνας πλέον έχουν εξαχθεί μέσω του παραπάνω αλγορίθμου είναι σειρά να ανιχνευθούν και τα αντικείμενα. Τα αντικείμενα επιλέγονται μέσω του αλγορίθμου συνδεδεμένων μερών (connected components[8]). Αυτός ο αλγόριθμος ανιχνεύει και ομαδοποιεί τα υποψήφια αντικείμενα σε κάθε υποψήφιο επίπεδο. Αν το μέγεθος του αντικειμένου είναι αρκετό τότε επιλέγεται προς περαιτέρω επεξεργασία.

Αφού είμαστε σίγουροι για τα αντικείμενα μας τότε πρέπει να τα χαρακτηρίσουμε. Ο αλγόριθμος που χρησιμοποιείτε είναι ο μετασχηματισμός κλιματικά - αναλλοίωτων χαρακτηριστικών (SIFT : Scale-Invariant Feature Transform[9]). Αυτός ο αλγόριθμος παράγει μοναδικά χαρακτηριστικά σε ένα αντικείμενο. Έτσι τα αντικείμενα έχουν χαρακτηριστικά που δεν μπορούν να αλλοιωθούν από την κλίση τους, τον φωτισμό και το μέγεθος τους. Αφού πλέον έχουμε τα αντικείμενα μας και τα χαρακτηριστικά τους τα κατηγοριοποιούμε σε μια βάση δεδομένων.

Ακολουθούν δυο μέρη στην κατασκευή του συστήματος η εκπαίδευση και η χρήση. Η εκπαίδευση γίνεται με μια βάση δεδομένων που παρέχει ο χρήστης η οποία περιεχί αντικείμενα τα οποία είναι γνωστά σε αυτόν. Τα ομαδοποιεί και τα κατηγοριοποιεί αναλόγως όπως θα έκανε και στον κανονικό περιβάλλον. Αφού ολοκληρώσει την διαδικασία αυτή τα τοποθετεί σε ένα παράλληλο σύστημα δημιουργίας βάσεων δεδομένων από χαρακτηριστικά. Με αυτόν τον τρόπο η βάση δεδομένων γίνεται η είσοδος ενός τεχνητού νευρωνικού δικτύου με την οποία εκπαιδεύεται με τα χαρακτηριστικά των αντικειμένων.

Η χρήση περιέχει και την εκπαίδευση με έναν διαφορετικό τρόπο. Στην χρήση τα αντικείμενα από την εικόνα του συστήματος μας κάνουν όλη την παραπάνω διαδικασία εκτός από την εκπαίδευση. Αφού έχουμε τα αντικείμενα με τα χαρακτηριστικά τους εισέρχονται στο τεχνητό νευρωνικό δίκτυο το οποίο μας δίνει την πιθανότητα του κάθε αντικειμένου να είναι ένα από αυτά με τα οποία έχει εκπαιδευτεί. Έτσι επιλέγεται το καλύτερο σε πιθανότητες αντικείμενο και αυτό επιστρέφει μια απάντηση στον χρήστη. Αν η απάντηση είναι λανθασμένη τότε το σύστημα θα έχει την δυνατότητα δημιουργίας νέας κατηγορίας στην βάση δεδομένων εάν αυτή δεν υπάρχει και θα επανεκπαιδεύσει το σύστημα. Αν υπάρχει η κατηγορία τότε θα εμπλουτίζετε με τα νέα δεδομένα και το τεχνητό νευρωνικό δίκτυο θα εκπαιδεύεται ξανά.

Με αυτόν τον τρόπο έχουμε πραγματικά εξαιρετικά αποτελέσματα και κατορθώνω τον σκοπό αυτής της εργασίας. Πλέον κάθε αντικείμενο του φυσικού κόσμου θα μπορεί να υποστεί ψηφιοποίηση και να

χρησιμοποιηθεί σαν μάθηση από ένα υπολογιστικό σύστημα.

## 1.2 Κίνητρο για την διεξαγωγή της εργασίας

Το κίνητρο για την διεξαγωγή αυτής της εργασίας αρχικά ήταν η δημιουργία ενός εντελώς πρωτότυπου συστήματος. Ήταν για εμένα μια πρόκληση να μπορώ να κατασκευάσω ένα σύστημα το οποίο θα μπορούσε μέσω στερεοσκοπικών καμερών να ανακτά πληροφορίες από τις εικόνες και να τις εξάγει ως γνώση μηχανής σε ένα υπολογιστικό σύστημα.

Όμως ήταν προφανές οι γνώσεις που είχα έως τότε σαν βάση στους τομείς της τεχνητής όρασης και τεχνητής νοημοσύνης δεν ήταν αρκετές. Το σύστημα αυτό απαιτούσε την χρήση αλγορίθμων τελευταίας τεχνολογίας. Έτσι άλλο ένα κίνητρο που προέκυψε ήταν και ο εμπλουτισμός των γνώσεων μου σε αυτούς τους τομείς της επιστήμης των υπολογιστών.

Επίσης η κατασκευή αυτού καθαυτού του συστήματος λειτούργησε και σαν μια εμπειρία η οποία δεν μου είχε συμβεί έως τότε. Για την κατασκευή αρχικά χρειάστηκε μια πρωτότυπη ιδέα. Οι ανάγκες εκπόνησης ενός πρωτότυπου συστήματος με έβαλε στην λογική της εργαστηριακής έρευνας,. Αρχικά προέκυψε η ανάγκη να βρω τι είχε ήδη γίνει σε αυτό τον τομέα και τι εργαλεία είχα διαθέσιμα για την διεκπεραίωση της εργασίας μου. Αργότερα ήρθε η ανάγκη για έναν αρχικό σχεδιασμό έτσι ώστε να ακολουθηθεί ένα πρόγραμμα κατασκευής. Κατά την διάρκεια της εργασίας άλλαξα αρκετές φορές το αρχικό μου σχέδιο και εμπλούτισα είτε αφαιρέσα κάποια κομμάτια. Έτσι πέτυχα το καλύτερο δυνατό αποτέλεσμα και ανέδειξα τις αρετές και τις εφαρμογές αυτού του συστήματος.

## 1.3 Σκοπός και στόχοι της εργασίας

Η εργασία αυτή είχε σαν σκοπό της την ανάδειξη καινοτόμων τεχνολογιών των τομέων τεχνητής νοημοσύνης και τεχνητής όρασης. Για την ανάδειξη αυτών των τομέων της επιστήμης υπολογιστών έπρεπε να δημιουργηθεί κάτι εντελώς νέο που να εκμεταλλεύεται τις αρετές τους. Έτσι επιτυγχάνετε μια μικρή εξέλιξη στην τεχνητή όραση αφού συνδυάζετε με τον καλύτερο δυνατό τρόπο με άλλες τεχνικές. Το ίδιο θα μπορούσα να πω και για τον τομέα της τεχνητής νοημοσύνης αφού αποκτά μεγαλύτερη σημασία σε ένα τέτοιο σύστημα. Άλλος ένας σκοπός ήταν το σύστημα να βασιστεί σε πραγματικές λειτουργίες. Εάν το σύστημα στηριζόταν σε αυτό θα ήταν καλύτερη η κατανόηση του και περισσότερες οι εφαρμογές του στον τομέα των υπολογιστών. Αποκτώντας τα παραπάνω, η χρήση του πλέον θα μπορούσε να εφαρμοστεί από επιστημονικές έως και σε καθημερινές εμπορικές εφαρμογές.

Οι στόχοι της εργασίας είναι πρωτοτυπία που αποτελεί κομμάτι του στόχου για την εξέλιξη των τομέων της τεχνητής νοημοσύνης και τεχνητής όρασης. Η αναγνωρισιμότητα που προσφέρει μέσω της λειτουργίας του ένα τέτοιο σύστημα στις επιστήμες, στον δημιουργό και το σύστημα το ίδιο. Τέλος, ένας κύριος στόχος ήταν η κατάρτιση με γνώσεις έως τότε άγνωστες σε εμένα. Έτσι μου παρέχετε ένα ακόμα εφόδιο για την συνέχεια της περαιτέρω καριέρας μου.

## 1.4 Δομή εργασίας

Αυτή η εργασία χωρίστηκε σε τμήματα με τέτοιο τρόπο έτσι να είναι καλύτερη η κατανόηση και η εκτίμηση της. Κλείνοντας λοιπόν την εισαγωγή της αυτής εργασίας ακολουθεί μια σύντομη περίληψη των υπολοίπων κεφαλαίων και των θεμάτων τα οποία θα αναλυθούν σε αυτά.

Το επόμενο κεφάλαιο είναι η μεθοδολογία για την κατασκευή του συστήματος. Έτσι σε αυτό το κεφάλαιο θα αναπτύξουμε τους μεθόδους ανάλυσης και ανάπτυξης της πτυχιακής. Στην συνέχεια θα ασχοληθούμε αναλυτικά με τα μοντέλα τα οποία χρησιμοποιήθηκαν για να πάρει την τελική μορφή το σύστημα μας.

Έπειτα θα συνεχίσουμε με τις θεωρίες οι οποίες αποτελούν τα θεμέλια για τους αλγορίθμους τους οποίους χρησιμοποιήθηκαν. Τέλος, το κεφάλαιο θα κλείσει με τους αλγορίθμους που είναι ένα μεγάλο κομμάτι τις εργασίας λόγω της πολυπλοκότητας και του ανώτερου επίπεδου τους.

Το κεφάλαιο που ακολουθεί ονομάζεται σχέδιο δράσης. Σε αυτό το κεφάλαιο θα αναλυθεί με κάθε λεπτομέρεια το τρόπος με τον οποίο εκπονήθηκε η εργασία. Αρχικά θα αναφερθούν άλλες προσπάθειες που έχουν γίνει σε αυτήν την κατεύθυνση που κινείτε η πτυχιακή εργασία. Επίσης θα αναφερθούν διεξοδικά σημαντικοί στόχοι για την ολοκλήρωση της πτυχιακής εργασίας. Κλείνοντας το κεφάλαιο θα αναλυθεί και θα αναρτηθεί ένα αναλυτικό χρονοδιάγραμμα για την εκπόνηση της εργασίας.

Συνέχεια έχει το κύριο μέρος της πτυχιακής. Στο κύριο μέρος πλέον θα αναλυθεί όλη η εργασία η οποία έχει γίνει. Αρχικά θα υπάρξει μια ανάλυση των προβλημάτων που προέκυψαν αρχικά και κατά την διάρκεια της εκπόνησης της πτυχιακής εργασίας. Συνέχεια έχουν οι απαιτήσεις του συστήματος, δηλαδή όλα τα εργαλεία και οι γνώσεις που χρειάζονται για την εκκίνηση της λειτουργίας του συστήματος. Θα ακολουθήσει ένας γενικός σχεδιασμός του προβλήματος και οι προσθήκες που έγιναν στον αρχικό σχεδιασμό. Στο τέλος του κεφαλαίου θα πραγματοποιηθεί η αναλυτική υλοποίηση του συστήματος βήμα προς βήμα.

Τελευταίο άλλα και από τα πιο σημαντικά κεφάλαια αποτελεί το κεφάλαιο με τα αποτελέσματα. Σε αυτό το κεφάλαιο παραδίδονται τα πραγματικά αποτελέσματα. Αφότου γίνει η παρουσίαση των αποτελεσμάτων θα ακολουθήσουν τα συμπεράσματα. Τα συμπεράσματα θα αναλυθούν διεξοδικά με παρατηρήσεις οι οποίες θα μας οδηγήσουν και στο τελικό συμπέρασμα. Τέλος, με το κλείσιμο της εργασίας θα γράφει κεφάλαιο το οποίο έχει σχέση με τις μελλοντικές επεκτάσεις της εργασίας. Με αυτό τον τρόπο επιδιώκετε και η βελτιστοποίηση του συστήματος και των αποτελεσμάτων του.

## 2. Μεθοδολογία Υλοποίησης

Μεθοδολογία είναι ένα σύνολο μεθόδων, άρχων και κανόνων για την ρύθμιση ενός δεδομένου γνωστικού αντικειμένου όπως στις τέχνες και τις επιστήμες [10]. Αρχικά με την σύλληψη της ιδέας γεννήθηκαν αμέσως ερωτήματα για το πως μπορεί να υλοποιηθεί. Η αρχική μέθοδος ήταν να κατακερματιστεί η ιδέα σε περαιτέρω τομείς. Οι τομείς αυτοί έπρεπε να είναι ξεκάθαροι με μορφή βημάτων οπου η εκτέλεση του ενός οδηγούσε στον επόμενο. Οι τομείς αυτοί αφορούσαν διαφορετικές θεωρίες και αλγορίθμους. Έτσι μεσώ ερευνάς η ιδέα κατακερματίστηκε σε στοιχειώδη μέρη τα οποία χειρίστηκα ξεχωριστά.

Αφού το σύστημα είχε πλέον τομείς, ξεκίνησα από την βάση του ιδέας οπου ήταν η όραση. Στην όραση χρησιμοποιήθηκαν αρκετά και διαφορετικά μοντέλα. Στην τεχνητή όραση υπάρχει η δυνατότητα για την χρήση διαφορετικών μοντέλων και τεχνικών για την αποκομιδή ενός αποτελέσματος. Τα μοντέλα που επιλέχθηκαν όντως είχαν διαφορετικές παραμέτρους και διαφορετικά αποτελέσματα. Έτσι ξεκίνησε η έρευνα στο ποιο μοντέλο τεχνητής όρασης ήταν το κατάλληλο για το σύστημα μας.

Με την επιλογή του μοντέλου το οποίο ταίριαζε στο σύστημα μας στην συνέχεια προήλθε η ανάγκη να βρεθεί αλγόριθμος κατάλληλος για την χρήση του. Οι αλγόριθμοι οι οποίοι χρησιμοποιήθηκαν προέρχονται από διαφορετικές θεωρίες. Με αυτό σαν δεδομένο κάθε αλγόριθμος έπρεπε να ερευνηθεί ξεχωριστά έτσι ώστε να προκύψει ένα συμπέρασμα για την συμπεριφορά και τα αποτελέσματα τους. Απαραίτητο σε αυτό το εγχείρημα ήταν και η μελέτη των θεωριών που “κρύβονταν” πίσω από κάθε αλγόριθμο. Μεσώ διεξοδικής ερευνάς έφτασα στο επιθυμητό αποτέλεσμα.

Στην συνέχεια τα αποτελέσματα του παραπάνω αλγορίθμου έπρεπε να μεταβιβαστούν στον επόμενο τομέα ο οποίος είναι υπεύθυνος για την εύρεση τον αντικειμένων στην εικόνα. Έτσι χρησιμοποιήθηκε αλγόριθμος ο οποίος ήταν ικανός να παράγει αποτελέσματα τα οποία μπορούσαν να συνδυαστούν με την αρχική πληροφορία. Μετά τον συνδυασμό των δυο η πληροφορίες που παρήχθησαν στάλθηκαν στον επόμενο τομέα.

Σε αυτόν τον τομέα έπρεπε να γίνει η χρήση αλγορίθμων οι οποίοι να έδιναν μοναδικά και έγκυρα αποτελέσματα. Ήταν το κομμάτι το οποίο χρειάστηκε τον περισσότερο χρόνο στην εργασία αυτή αφού ερευνήθηκαν αρκετοί άλλοι αλγόριθμοι. Χρησιμοποιήθηκαν αλγόριθμοι οι οποίοι ήταν αρκετά απαιτητικοί από την άποψη της χρήσης τους μέσα στο σύστημα. Πρόβλημα επίσης ήταν και η θεωρία που περικλείει αυτούς τους αλγορίθμους αφού ήταν και αυτή απαιτητικοί στην κατανόηση της. Τέλος, προβλήματα υπήρξαν και με την αξιολόγηση των αποτελεσμάτων τους. Τελικά μετά από αρκετή έρευνα και τριβή με όλους τους αλγορίθμους επιλέχθηκε ο βέλτιστος.

Τέλος το σύστημα τερματίζει με μια τελευταία επεξεργασία. Η θεωρία αυτής της επεξεργασίας υπάγεται στον τομέα της τεχνητής νοημοσύνης. Έτσι ο αλγόριθμος επιλέχθηκε μετά από έρευνα ανάμεσα σε άλλους αλγορίθμους με κίνητρο το μέγιστο αποτέλεσμα. Αυτή η επεξεργασία μας δίνει το τελικό αποτέλεσμα το οποίο παρουσιάζετε και στον χρήστη.

Από όλα τα παραπάνω προκύπτουν τέσσερα κεφάλαια. Το πρώτο είναι μια μάτια στην μέθοδο ανάλυσης και ανάπτυξης της πτυχιακής εργασίας δηλαδή μια ποιο αναλυτική μάτια όλων των παραπάνω. Στην συνέχεια ακολουθεί ένα κεφάλαιο με τα μοντέλα τα οποία είτε χρησιμοποιήθηκαν είτε ερευνήθηκαν και οι μεταξύ τους διαφορές, τα μειονεκτήματα και τα πλεονεκτήματα τους. Αμέσως επόμενο ακολουθηθεί ένα κεφάλαιο το οποίο έχει θέμα τις θεωρίες. Οι θεωρίες είναι απαραίτητες για την κατανόηση των αλγορίθμων που προκύπτουν από αυτές. Τέλος, το κεφάλαιο με τους αλγορίθμους ενσωματώνει και τα μοντέλα άλλα και τις

θεωρίες. Επίσης αναφέρονται όλοι οι αλγόριθμοι, οι διαφορές τους και τα πλεονεκτήματα ορισμένων ενάντια σε κάποιους άλλους.

## 2.1 Μέθοδος ανάλυσης και ανάπτυξης πτυχιακής

Η μέθοδος η οποία χρησιμοποιείτε βασίστηκε πάνω στην ίδια την ιδέα του συστήματος. Το σύστημα μας αποτελείται από τρία βασικά κομμάτια. Τα κομμάτια αυτά είναι η είσοδος της εικόνας μέσω τεχνητής όρασης, η επεξεργασία της εικόνας μέσω τεχνητής όρασης και τελικά παραγωγή εμπειρίας και γνώσης μέσω τεχνητής νοημοσύνης. Παρατήρησα ότι τα κομμάτια του συστήματος ήταν σε τέτοια διάταξη όπου έπρεπε να ξεκινήσεις από την αρχή. Δηλαδή το σύστημα δεν σου επιτρέπει να ασχοληθείς με κάτι άλλο πριν από την δημιουργία του προηγούμενου κομματιού όπως ακριβώς και η κατασκευή ενός κτιρίου. Έτσι τα προβλήματα τα οποία προέκυπταν σε κάθε κομμάτι του συστήματος αντιμετωπίζονταν ένα προς ένα χάρις την προηγούμενη μεθοδολογία.

Παίρνοντας το πρώτο κομμάτι, που ασχολείται με την αρχική πληροφορία του συστήματος άρχισε η διερεύνηση του βέλτιστου τρόπου για την εισαγωγή της εικόνας στο σύστημα. Αυτό το επίπεδο σύμφωνα με την τεχνητή όραση ονομάζετε λήψη της εικόνας. Είναι το πρώτο επίπεδο και το χαμηλότερο για την εισαγωγή της εικόνας σε ένα σύστημα. Ασχολείται με τους αισθητήρες αλλά και με τις τεχνικές της λήψης μιας εικόνας από το περιβάλλον. Το σύστημα δεν έχει άμεση σχέση με το είδος ή τα τεχνικά χαρακτηριστικά των αισθητήρων. Όμως έχει άμεση σχέση με τις τεχνικές λήψης. Για αυτό το σύστημα χρησιμοποιήθηκαν μέθοδοι όπως λήψη εικόνας από μια κάμερα, λήψη εικόνας από στερεοσκοπική κάμερα και λήψη εικόνας από πολύ – στερεοσκοπική κάμερα. Για το σύστημα επιλέχθηκε η απλή στερεοσκοπική λήψη εικόνων. Παρακάτω αναφέρετε με λεπτομέρειες η χρήση αυτής της μεθόδου και οι διαφορές της από τις υπόλοιπες που δεν χρησιμοποιήθηκαν.

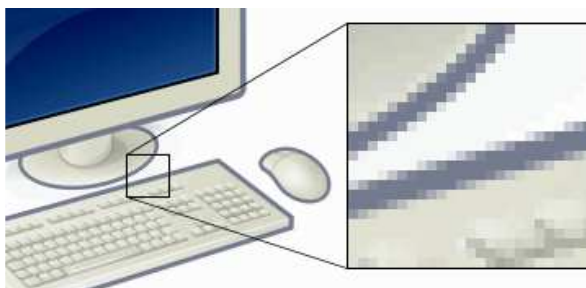
Στο δεύτερο κομμάτι το οποίο είναι και το πιο μακροσκελές σε αυτήν την εργασία, ασχολούμαστε με την επεξεργασία της εικόνας. Αυτή η εικόνα προέρχεται από το προηγούμενο κομμάτι το οποίο χρησιμοποιεί την στερεοσκοπική μέθοδο για την λήψη της εικόνας. Έτσι αρχικά δημιουργείτε η ανάγκη για τον έλεγχο αυτής της μεθόδου. Έτσι ερευνήθηκαν ένας κλασσικός και ένας εξελιγμένος αλγόριθμος στερεοσκοπικής όρασης. Συνέχεια είχε η περαιτέρω επεξεργασία αυτής της πληροφορίας για την εύρεση των αντικειμένων στην εικόνα. Έτσι και εκεί ερευνήσα την χρήση διαφορετικών αλγορίθμων για την κατηγοριοποίηση τους. Αφού πλέον τα αντικείμενα ήταν αντιληπτά από το σύστημα προέκυψε η ανάγκη για την δυνατότητα του συστήματος να μπορεί να τα χαρακτηρίσει. Ο χαρακτηρισμός των αντικειμένων έπρεπε να είναι σαφής και να παρέχει μοναδικά χαρακτηριστικά σε κάθε αντικείμενο. Μέσω χρήσης και έρευνας διαφορετικών αλγορίθμων προέκυψε το τελικό αποτέλεσμα για την χρήση του βέλτιστου. Όλοι η παραπάνω αλγόριθμοι δεν είναι μόνο διαφορετικοί μεταξύ τους στο αποτέλεσμα που παρέχουν. Οι αλγόριθμοι αυτοί έχουν και διαφορετικές λειτουργίες οι οποίες στηρίζονται και σε διαφορετικές θεωρίες. Έτσι στην συνέχεια του κεφαλαίου θα γίνει πλήρη αναφορά στα χαρακτηριστικά των αλγορίθμων αλλά και μια εισαγωγή στις θεωρίες που τους περικλείουν.

Στο τελικό κομμάτι του συστήματος έχουμε την εισαγωγή μιας νέας θεωρίας. Αυτή η θεωρία είναι της τεχνητής νοημοσύνης όπου περικλείει μέσα της διαφορετικούς αλγορίθμους. Οι αλγόριθμοι της τεχνητής νοημοσύνης είναι πολλοί και χρησιμοποιούνται σε διαφορετικές περιπτώσεις. Οι περιπτώσεις αυτές ποικίλουν ανάλογα με τον τύπο των δεδομένων και τον τύπο της επεξεργασίας τους από τους αλγορίθμους. Για αυτόν τον λόγο έγινε έρευνα για την αναδείξει του καλύτερου αλγορίθμου ο οποίος ταιριάζει στα δεδομένα μας. Αναφορά για τα πλήρη χαρακτηριστικά τους γίνεται στην συνέχεια του κεφαλαίου.

Με την επεξήγηση και εκτενή αναφορά στα γνωρίσματα των μοντέλων, των θεωριών και των αλγορίθμων που εσωκλείονται σε αυτούς πετυχαίνουμε την καλύτερη κατανόηση τους. Με την κατανόηση όλων των παραπάνω έχουμε το πλεονέκτημα της γνώσης των χαρακτηριστικών γνωρισμάτων τους και την γνώση της ορθής χρήσης τους. Έτσι στο σύστημα μας επιτυγχάνετε η μέγιστη απόδοση και ομαλή λειτουργία. Επίσης μας παρέχετε και η γνώση για μελλοντική τροποποίηση και αναβάθμιση του συστήματος.

## 2.2 Μοντέλα

Τα μοντέλα τα οποία χρησιμοποιήθηκαν στο σύστημα μας έχουν σχέση με τον τομέα της τεχνητής όρασης. Στην τεχνητή όραση το πρώτο στάδιο ονομάζεται στάδιο απόκτησης της εικόνας. Μια ψηφιακή εικόνα παράγεται από έναν ή περισσότερους αισθητήρες. Η παραδοσιακή εικόνα παράγεται από αισθητήρες οι οποίοι είναι ευαίσθητοι στο φως. Ανάλογα με τον τύπο και την τεχνική που παρέχετε ο κάθε αισθητήρας παράγει διαφορετικό αποτέλεσμα. Έτσι μπορεί να έχουμε μια δισδιάστατη εικόνα, μια τρισδιάστατη εικόνα είτε μια σειρά από εικόνες. Το συνηθισμένο αποτέλεσμα είναι μια ψηφιακή εικόνα η οποία παράγεται σε έναν δισδιάστατο πίνακα και αποτελείται από εικονοστοιχεία. Τα εικονοστοιχεία αυτά είτε αντιστοιχούν σε μια ένταση του φυσικού φωτός είτε έχουν περισσότερες φασματικές ζώνες. Σε μερικές τεχνικές τα εικονοστοιχεία αντιστοιχούν σε βάθος, απορρόφηση μιας ανάκλασης ηχητικών ή ηλεκτρομαγνητικών κυμάτων [3].



*Εικόνα 3: Παράδειγμα παραγωγής εικόνας από μικρότερα εικονοστοιχεία.*

Το σύστημα που περιγράφετε σε αυτήν την εργασία ενδιαφέρετε για το βάθος το οποίο εκφράζετε μια εικόνα. Το βάθος μιας εικόνας είναι ένα μέτρο το οποίο μπορεί να παρέχει τα αντικείμενα στην εικόνα. Αυτό συμβαίνει διότι στον φυσικό κόσμο τα αντικείμενα είναι τοποθετημένα στον χώρο. Η δυνατότητα της αναγνώρισης των τριών διαστάσεων στον χώρο ονομάζεται αντίληψη του βάθους.

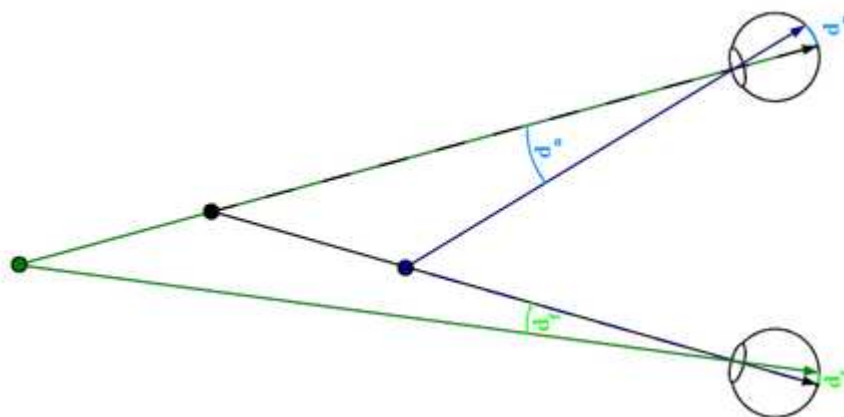
Αντίληψη του βάθους είναι η οπτική ικανότητα να αντιλαμβανόμαστε τον κόσμο σε τρεις διαστάσεις και την απόσταση ενός αντικειμένου. Η αντίληψη βάθους προκύπτει από μια ποικιλία τεχνικών βάθους. Αυτές συνήθως κατατάσσονται στην διοφθαλμική τεχνική όπου απαιτεί στοιχεία από δυο οφθαλμούς ή μόνο - οφθαλμική τεχνική από έναν οφθαλμό. Οι διοφθαλμικές τεχνικές περιλαμβάνουν την στέρεο - οπτική. Οι μόνο - οφθαλμικές τεχνικές περιλαμβάνουν το μέγεθος των αντικειμένων καθώς και την γωνία τους [11]. Επίσης υπάρχουν θεωρητικές οπτικές τεχνικές η οποίες χρησιμοποιούν περισσότερους από έναν οφθαλμούς. Έτσι στο σύστημα μας δίνετε η δυνατότητα χρήσης πολλών τεχνικών απόκτησης της εικόνας. Με αυτό το δεδομένο έγινε έρευνα ανάμεσα σε αυτές τις τεχνικές. Η τεχνική η οποία χρησιμοποιήθηκε είναι η στερεοσκοπική όραση.

### 2.2.1 Στερεοσκοπική όραση

Η στερεοσκοπική όραση είναι η πιο συνηθισμένη τεχνική που συναντάτε στην φύση. Η πορεία της στερεοσκοπικής όρασης είναι σχεδόν συνδεδεμένη με την δημιουργία πολυκύτταρων οργανισμών. Οι μεγαλύτερες ομάδες βιολογικών οργανισμών που την χρησιμοποιούν είναι τα θηλαστικά, τα πτηνά, τα ερπετά και τα ψάρια. Όπως πολλές φορές έχει αποδειχθεί η φύση είναι σοφή, δεν κάνει λάθη σε σχέση με τις τεχνικές που εφαρμόζει πάνω στους οργανισμούς. Στον άνθρωπο αυτή η τεχνική είναι γνωστή και είναι ένα από τα εργαλεία τα οποία των βοήθησαν στην εξέλιξη και την κυριαρχία του πάνω στον πλανήτη.

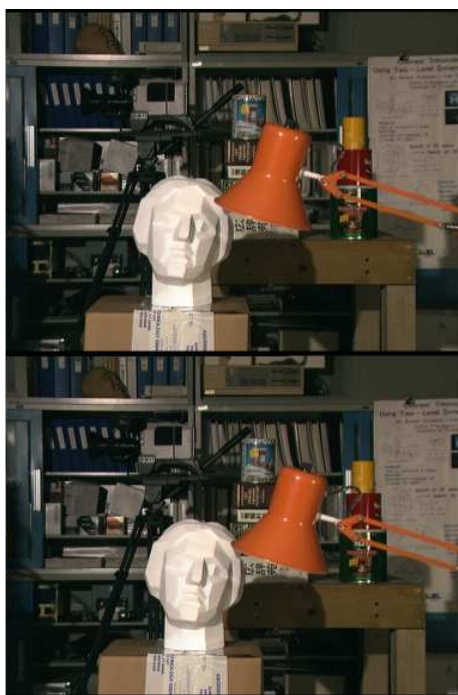
Το στερεοσκοπικό φαινόμενο στηρίζεται στην στέρεο - οπτική. Η γεωμετρικές βάσης της στερεοσκοπικής στηρίζονται στην λειτουργία των οφθαλμών. Όταν ένα πρόσωπο κοιτά ένα αντικείμενο τα δυο μάτια συγκλίνουν ώστε το αντικείμενο να εμφανίζεται στο κέντρο του αμφιβληστροειδούς και στους δυο οφθαλμούς. Τα υπόλοιπα αντικείμενα γύρω από το κύριο αντικείμενο μετατοπίζονται σε σχέση με το κύριο

αντικείμενο. Επειδή κάθε οφθαλμός είναι σε διαφορετική οριζόντια θέση, το κάθε ένα έχει μια ελαφρώς διαφορετική οπτική γωνιά σε μια σκηνή που παράγουν διαφορετικές εικόνες στον αμφιβληστροειδή [6].



Εικόνα 4: Διαφορά στην οπτική γωνιά που παράγει διαφορετικές εικόνες

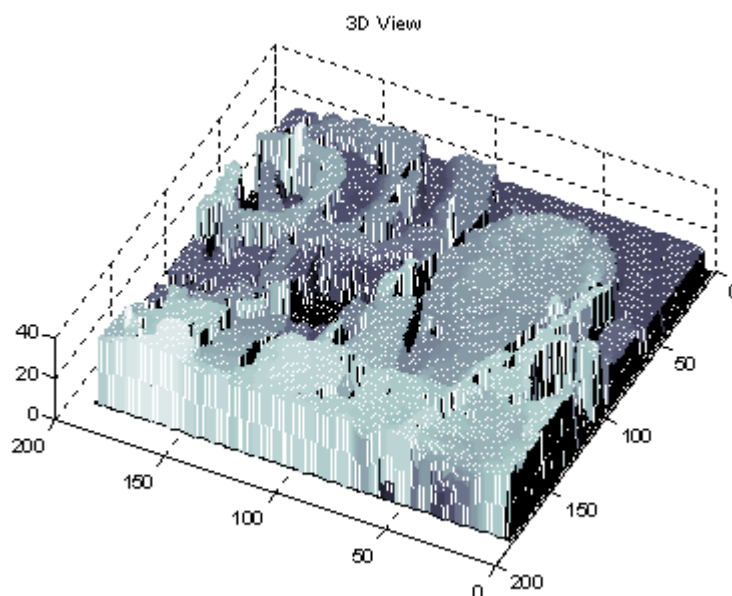
Στην επιστήμη των υπολογιστών η στερεοσκοπική όραση είναι κομμάτι του τομέα της τεχνητής όρασης. Χρησιμοποιήτε ήδη σε ρομπότ για τον εντοπισμό εμποδίων ή την ρομποτική χειρουργική. Σε αυτήν την περίπτωση οι οφθαλμοί αναπαριστώνται από δυο κάμερες η οποίες τραβούν φωτογραφίες της ίδιας σκηνής με την ίδια απόσταση με αυτήν τον ματιών μας. Οι κάμερες είναι απαραίτητο να έχουν τα ίδια τεχνικά χαρακτηριστικά καθώς και την ίδια ανάλυση. Το αποτέλεσμα των δυο καμερών είναι δυο εικόνες. Οι δυο εικόνες αυτές έχουν μια ελαφριά μετατόπιση μεταξύ τους. Φυσικά η μετατόπιση αυτή μεγαλώνει καθώς τα αντικείμενα της εικόνας βρίσκονται πιο μακριά από τις δυο κάμερες. Έτσι το αποτέλεσμα μιας λήψης των δυο καμερών ονομάζετε στερεοσκοπικό ζεύγος εικόνων.



Εικόνα 5: Στερεοσκοπικό ζεύγος εικόνων



Απαραίτητο όμως στην χρήση αυτής της τεχνικής είναι και ένας αλγόριθμος ο οποίος θα διαχειρίζεται και επεξεργάζεται το στερεοσκοπικό ζεύγος εικόνων. Το ζητούμενο αποτέλεσμα είναι ένας πίνακας των διαφορών ανάμεσα στις δυο εικόνες ή διαφορετικά ένας χάρτης βάθους. Ο χάρτης βάθους ή πίνακας διαφοράς είναι ένας πίνακας στο μέγεθος των αρχικών εικόνων. Στον χάρτη βάθους περιγράφεται το επίπεδο βάθους το οποίο κατηγοριοποιείται κάθε εικονοστοιχείο.



*Εικόνα 6: Πίνακας διαφοράς ή χάρτης βάθους σε τρισδιάστατη απεικόνιση*

Σε αυτήν την εργασία χρησιμοποιήθηκαν δυο αλγόριθμοι για την παράγωγή χαρτών βάθους. Ο ένας ήταν ο κλασικός αλγόριθμος που χρησιμοποιείτε συνήθως για την παραγωγή των χαρτών βάθους. Ο άλλος ήταν ένας εξελιγμένος αλγόριθμος ο οποίος είναι πιο ακριβείς, ακυρώνει όλα τα μειονεκτήματα του πρώτου και χρησιμοποιεί την θεωρία των γράφων. Και οι δυο περιγράφονται παρακάτω στο κεφάλαιο των αλγορίθμων μαζί με τις θεωρίες που τους περικλείουν.

Το σύστημα που περιγράφεται είναι αρκετά απαιτητικό σε ότι αφορά την λειτουργία και την ακρίβεια των αποτελεσμάτων τους. Από αυτά προκύπτει η απορία για ποιον λόγο χρησιμοποιήθηκε αυτή η τεχνική έναντι σε άλλες. Εναλλακτικές μέθοδοι οι οποίες ερευνήθηκαν ήταν η χρήση μιας κάμερας και η χρήση πολλαπλών καμερών

Η χρήση μιας κάμερας ήταν η πιο απλή τεχνική από τις άλλες δυο που ερευνήθηκαν. Οι απαιτήσεις αυτής της τεχνικής στο κομμάτι του υλικού του συστήματος είναι ελάχιστες αφού η κάμερα είναι μοναδική και δεν είναι απαραίτητο ο συγχρονισμός της με κάποια άλλη. Η χρήση μιας απλής κάμερας επίσης παρέχει λιγότερες πληροφορίες οι οποίες είναι εύκολες στην διαχείριση και μικρές σε επεξεργαστικό κόστος. Όμως αφού δεν μας παρέχετε παρά ένα μικρο κομμάτι πληροφορίας της εικόνας πρέπει να χρησιμοποιηθούν τεχνικές για την ανάκτηση των αντικειμένων της εικόνας αυτής. Αυτές η τεχνικές πρέπει να εκτελεστούν σταδιακά και με μια συγκεκριμένη σειρά. Αυτό βέβαια έχει μεγάλο υπολογιστικό κόστος στο σύστημα. Παρόλα αυτά η χρήση αυτής της τεχνικής δεν θεωρήθηκε ασφαλής αφού τα αποτελέσματα ήταν ανάλογα με την ποιότητα της αρχικής εικόνας.

Στην συνέχεια ερευνήθηκε και η χρήση πάνω από δυο καμερών στο σύστημα. Αυτή η τεχνική των πολλαπλών καμερών χρησιμοποιείτε συνήθως σε στατικά συστήματα, αυτό κύριος οφείλετε στο μέγεθος των συστημάτων. Έτσι εύκολα είναι φανερό η μεγάλες απαιτήσεις όσο αναφορά τον συντονισμό όλων των καμερών πάνω σε ένα σημείο ή μια σκηνή. Επίσης είναι φανερό ότι σε αυτήν την τεχνική η επεξεργασία των δεδομένων που θα παραχθούν είναι αρκετά μεγάλη. Παρόλα αυτά η ποιότητα είναι σαφώς καλύτερη. Έτσι επιλέχθηκε με μια μικρή θυσία στην ποιότητα αλλά μεγάλο κέρδος στην ταχύτητα η στερεοσκοπική τεχνική.

Τα πλεονεκτήματα της τεχνικής είναι αρκετά. Αρχικά από την χρήση των δυο καμερών και την επεξεργασία των δεδομένων δεν έχουμε μόνο την πληροφορία της εικόνας αλλά και την πληροφορία του βάθους. Η πληροφορία του βάθους είναι η πληροφορία με την οποία το σύστημα ανακαλύπτει διακυμάνσεις στον χώρο. Αυτές τις διακυμάνσεις τις ερευνά για την εύρεση αντικειμένων στο περιβάλλον. Μάλιστα με την διαβάθμιση των καμερών μπορεί να γνωρίζει το μέγεθος των αντικειμένων ακόμα και την απόσταση τους από τις κάμερες. Η εύρεση των αντικειμένων είναι τόσο επιτυχής ώστε το περιβάλλον το οποίο περικλείει τα αντικείμενα να απορρίπτεται από τον αλγόριθμο. Έτσι έχουμε για περαιτέρω χρήση καθαρά τα αντικείμενα μας.

Η τεχνική αυτή έχει και απαιτήσεις. Οι απαιτήσεις προέρχονται αρχικά από το υλικό κομμάτι του συστήματος. Όπως αναφέρθηκε και παραπάνω το υλικό είναι απαραίτητο να έχει τα ίδια τεχνικά χαρακτηριστικά καθώς και την ίδια ανάλυση. Επόμενο πρόβλημα μπορεί να προκύψει από την υπολογιστική ισχύ του συστήματος, αφού μεγάλες εικόνες και η επεξεργασία τους απαιτεί και μεγάλη μνήμη. Έτσι πρέπει να βρεθεί μια ισορροπία ανάμεσα σε αυτούς τους παράγοντες. Πρόβλημα επίσης μπορεί να θεωρηθεί και η απόδοση του αλγορίθμου που χειρίζεται το σύστημα. Η ακρίβεια και ο χρόνος των αποτελεσμάτων είναι σημαντικές παράμετροι και για την επίτευξη τους περιγράφετε παρακάτω στο κεφάλαιο των αλγορίθμων.

Η προσφορά αυτής της τεχνικής στο σύστημα μας είναι πολύ μεγάλη και εξίσου σημαντική όπως και τα επόμενα κομμάτια του συστήματος. Αρχικά δεν θα ήταν δυνατό το σύστημα να έχει ως πεδίο εργασίας του τον πραγματικό τρισδιάστατο χώρο. Επίσης με την χρήση μιας οποιαδήποτε διαφορετικής τεχνικής δεν θα ήταν δυνατό να έχουμε την πληροφορία του βάθους στην κάθε εικόνα και κατ' επέκταση και τον διαχωρισμό των αντικειμένων στον χώρο.

## 2.3 Θεωρίες

Οι θεωρίες είναι το κομμάτι αυτής της πτυχιακής εργασίας στο οποίο στηρίζονται όλοι οι αλγόριθμοι και οι τεχνικές. Οι θεωρίες περιγράφονται σαν ένα γενικό κομμάτι αλλά περικλείουν μέσα τους αρκετούς και διαφορετικούς αλγορίθμους που μερικούς από αυτούς συναντώνται και σε αυτήν την εργασία. Στο σύστημα αυτό έγινε εκτενή χρήση τριών θεωριών.

Η πρώτη είναι η θεωρία της τεχνητής όρασης στην οποία και στηρίχθηκε η πρωταρχική ιδέα για την εκπόνηση αυτής της εργασίας. Η τεχνητή όραση όπως αναφέρετε και παραπάνω είναι κομμάτι του τομέα της επιστήμης υπολογιστών και ασχολείται με την όραση στα υπολογιστικά συστήματα και άλλες μηχανές. Είναι σχετικά νέα ως κλάδος της επιστήμης υπολογιστών και είναι σχετικά ανεπτυγμένη. Η χρήσης αυτής της θεωρίας σήμερα μπορεί να μην είναι τόσο καθημερινή όπως προβλέπετε στο μέλλον αλλά ακόμη και σήμερα χρησιμοποιείτε κατά κόρον για σοβαρούς εργαλεία και σε άλλους κλάδους των επιστημών.

Η επόμενη θεωρία είναι η τεχνητή νοημοσύνη η οποία και αυτή με την σειρά της είναι σημαντικό κομμάτι για την εκτέλεση του συστήματος. Η τεχνητή νοημοσύνη είναι και αυτή κομμάτι του τομέα της επιστήμης υπολογιστών και ασχολείται με την εφαρμογή και την χρήση αλγορίθμων πάνω σε υπολογιστικά συστήματα και άλλες συσκευές. Σαν κλάδος της επιστήμης υπολογιστών έχει μεγάλη άνθηση και διασημότητα ειδικά τα τελευταία χρόνια. Η χρήση των μεθόδων της τεχνητής νοημοσύνης είναι ποιο συχνά και ποιο καθημερινά αφού σήμερα κάθε συσκευή μπορεί να έχει έστω και ένα κομμάτι της στηριγμένο στους αλγόριθμους είτε στην θεωρία της.

Τέλος, είναι η θεωρία των γράφων η οποία είναι μαθηματική θεωρία. Αυτή η θεωρία θα μπορούσε να μην θεωρηθεί έως κομμάτι αυτής της εργασίας. Κάτι τέτοιο όμως θα ήταν λάθος για τον λόγο ότι κάποιοι από τους αλγόριθμους που χρησιμοποιήθηκαν περικλείουν αυτή την θεωρία. Επίσης πλέον στην επιστήμη των υπολογιστών η μαθηματική θεωρία των γράφων είναι γνωστή από την χρήση στην επεξεργασία εικόνων και από την ταξινόμηση διάφορων κομματιών μεταξύ τους.

### 2.3.1 Τεχνητή όραση

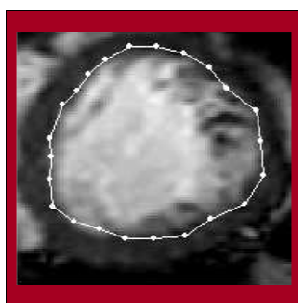
Τεχνητή όραση είναι η επιστήμη και η τεχνολογία των μηχανών που ικανότητα της όρασης για μια μηχανή είτε ένα υπολογιστικό σύστημα σημαίνει ότι η έχουν την δυνατότητα να εξάγουν πληροφορία από μια εικόνα για να λύσουν μια διεργασία. Σαν δομή η τεχνητή όραση στηρίζεται σε συστήματα τεχνητής νοημοσύνης που χρησιμοποιούν εικόνες σαν δεδομένα. Οι εικόνες μπορεί να έχουν πολλές μορφές όπως βίντεο είτε λήψη από πολλαπλές κάμερες είτε πολυδιάστατα δεδομένα όπως μια αξονική τομογραφία[3].

Η τεχνητή όραση είναι ένας σχετικά νέος τομέας στην επιστήμη των υπολογιστών. Στις πρώτες μέρες των υπολογιστών, ήταν δύσκολο να επεξεργαστούν ακόμα και μικρά σύνολα των εικόνων. Αυτό συνεχίστηκε μέχρι την δεκαετία του 70 όπου έγινε μια πιο εστιασμένη έρευνα και μελέτη πάνω στον τομέα. Η τεχνητή όραση καλύπτει ένα ευρύ φάσμα θεμάτων που συνήθως συνδέονται με άλλους κλάδους. Κατά συνέπεια δεν υπάρχει μια τυποποιημένη διατύπωση και αντιμετώπιση των προβλημάτων της τεχνητής όρασης. Αντί αυτού υπάρχει μια πληθώρα από μεθόδους για την λύση διαφόρων σαφώς ορισμένων διεργασιών της τεχνητής όρασης, που αυτές οι μέθοδοι συχνά είναι συγκεκριμένοι και σπάνια γενικοί στα προβλήματα που αντιμετωπίζουν. Πολλοί μέθοδοι είναι ακόμα σε αρχικό στάδιο ανάπτυξης, αλλά όλο και περισσότερες μέθοδοι συναντώνται σε εμπορικά προϊόντα, συνήθως όμως αποτελούν κομμάτι μεγαλύτερων και πιο πολύπλοκων συστημάτων που καλούνται να εκτελούν πολύπλοκες εργασίες[3].

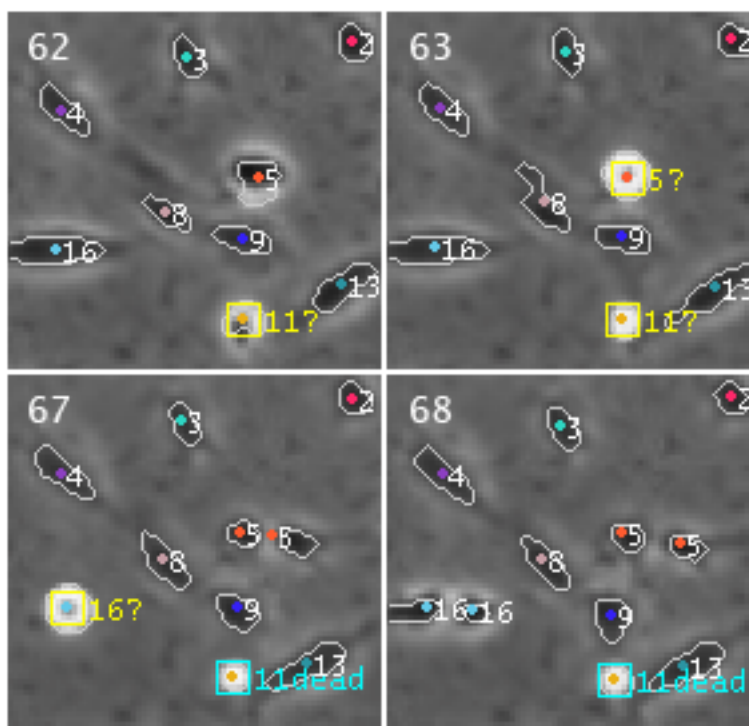
Η τεχνητή όραση προσπαθεί να χρησιμοποιήσει τις θεωρίες, τα μοντέλα της και τις τεχνικές της για την κατασκευή συστημάτων τεχνητής όρασης. Μερικές από αυτές τις εφαρμογές είναι:

- ♣ έλεγχος διεργασιών ( όπως σε ρομπότ και αυτόνομα οχήματα)
- ♣ ανίχνευση γεγονότων ( όπως σε συστήματα παρακολούθησης)
- ♣ οργάνωση πληροφοριών ( όπως σε βάσεις δεδομένων από εικόνες)
- ♣ κατασκευή μοντέλων ή περιβάλλοντος ( όπως βιομηχανικός έλεγχος, ιατρικές εικόνες, τοπογραφία)
- ♣ διαδραστικότητα ( όπως διαδραστικότητα με ανθρώπους)

Ένα από τα πιο σημαντικά πεδία εφαρμογής είναι η ιατρική τεχνητή όραση ή επεξεργασία ιατρικών εικόνων. Το πεδίο αυτό χαρακτηρίζεται από την άντληση πληροφοριών από τα στοιχεία μιας εικόνας για την ιατρική διάγνωση του ασθενούς. Σε γενικές γραμμές, τα δεδομένα της εικόνας είναι στην μορφή εικόνων μικροσκοπίου, εικόνων από ακτίνες X, εικόνες αγγειογραφίας, εικόνες υπερήχων και εικόνες τομογράφων. Παράδειγμα πληροφοριών που μπορούν να εξαχθούν από τέτοιες εικόνες είναι ανίχνευση όγκων, αρτηριοσκλήρωση και άλλες κακόβουλες μεταβολές. Επίσης χρησιμοποιήστε για την μέτρηση οργάνων και ροής του αίματος. Αυτή η περιοχή εφαρμογής υποστηρίζει την ιατρική έρευνα με την παροχή νέων πληροφοριών όπως η δομή του εγκεφάλου ή για την ποιότητα των ιατρικών θεραπειών[3].



Εικόνα 7: Εικόνα μέτρησης μεγέθους οργάνου μέσω ενεργού περιγράμματος (snakes)



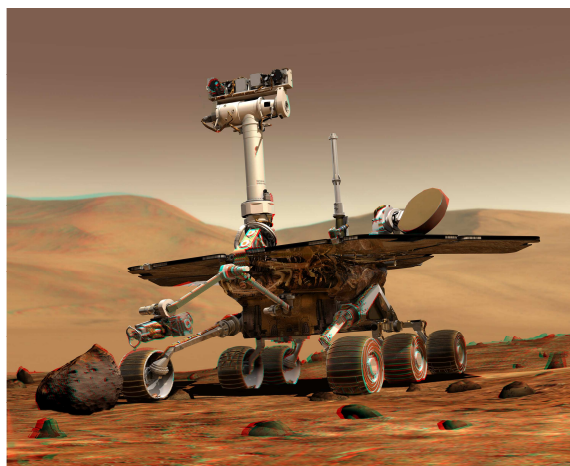
Εικόνα 8: Ανίχνευση και κατηγοριοποίηση κυττάρων

Ένας άλλος τομέας εφαρμογής της τεχνητής όρασης είναι η βιομηχανία, που συχνά ονομάζετε και μηχανική όραση, όπου η πληροφορίες εξάγονται με σκοπό την υποστήριξη μιας διαδικασίας παρασκευής. Ένα παράδειγμα είναι ο ποιοτικός έλεγχος όπου οι λεπτομέρειες των τελικών προϊόντων ελέγχονται αυτόματα για την εύρεση ελαττωμάτων. Ένα άλλο παράδειγμα είναι η μέτρηση της θέσης και του προσανατολισμού των στοιχείων που πρέπει να πάρει ένας ρομποτικός βραχίονας. Η μηχανική όραση επίσης χρησιμοποιήτε βαρέως και σε γεωργικές διαδικασίες για την απομάκρυνση ανεπιθύμητων τροφίμων, η διαδικασία αυτή ονομάζεται οπτική διαλογή[3].

Η στρατιωτικές εφαρμογές είναι ίσως ένας από τους μεγαλύτερους τομείς της τεχνητής όρασης. Τα προφανή παραδείγματα είναι η ανίχνευση αντίπαλων στρατευμάτων ή οχημάτων και η καθοδήγηση πυραύλων. Ποιο προηγμένα συστήματα για καθοδήγηση πυραύλων στέλνουν το βλήμα σε μια περιοχή πάρα σε έναν συγκεκριμένο στόχο και η επιλογή του στόχου γίνεται όταν το βλήμα είναι στην περιοχή από τοπικά αποκτημένες εικόνες. Σύγχρονες στρατιωτικές έννοιες επίσης συμπεριλαμβάνουν αισθητήρες εικόνες για

την παροχή ενός πλούσιου συνόλου πληροφοριών σχετικά με την σκηνή μιας μάχης που μπορεί να χρησιμοποιηθεί για την υποστήριξη στρατηγικών αποφάσεων. Σε αυτήν την περίπτωση χρησιμοποιήτε αυτόματη επεξεργασία δεδομένων για την μείωση της πολυπλοκότητας και τις πληροφορίες ασφάλειας από πολλαπλούς αισθητήρες για την αύξηση της αξιοπιστίας[3].

Μια από τις νεότερες εφαρμογές είναι στα αυτόνομα οχήματα, στα οποία συμπεριλαμβάνονται υποβρύχια, χερσαία οχήματα, εναέρια οχήματα και μη επανδρωμένα εναέρια οχήματα (UAV). Το επίπεδο της αυτονομίας κυμαίνεται από πλήρη αυτόνομα οχήματα είτε σε βασισμένα σε τεχνητή όραση συστήματα για την βοήθεια του οδηγού ή πιλότου σε διάφορες καταστάσεις. Πλήρη αυτόνομα οχήματα χρησιμοποιούν την τεχνητή όραση για πλοήγηση δηλαδή για να αναγνωρίζουν που βρίσκονται, για την κατασκευή χαρτών του περιβάλλοντος (SLAM) είτε για την αναγνώριση εμποδίων. Επίσης χρησιμοποιείτε για την ανίχνευση συγκεκριμένων γεγονότων που έχουν καθήκον να εκτελέσουν. Υπάρχουν παραδείγματα για υποβοήθηση συστημάτων ανίχνευσης εμποδίων σε αυτοκίνητα και συστήματα αυτόματης προσγειώσεως σε αεροσκάφη. Επίσης ορισμένοι κατασκευαστές αυτοκινήτων έχουν κατασκευάσει αυτόνομα αυτοκίνητα που στηρίζονται στην τεχνητή όραση. Τέλος, έχουν κατασκευαστεί και δοκιμάζονται αυτόνομα χερσαία οχήματα για την εξερεύνηση άλλων πλανητών στο ηλιακό μας σύστημα με παράδειγμα το NASA's Mars Exploration Rover και το ESA's ExoMars Rover[3].



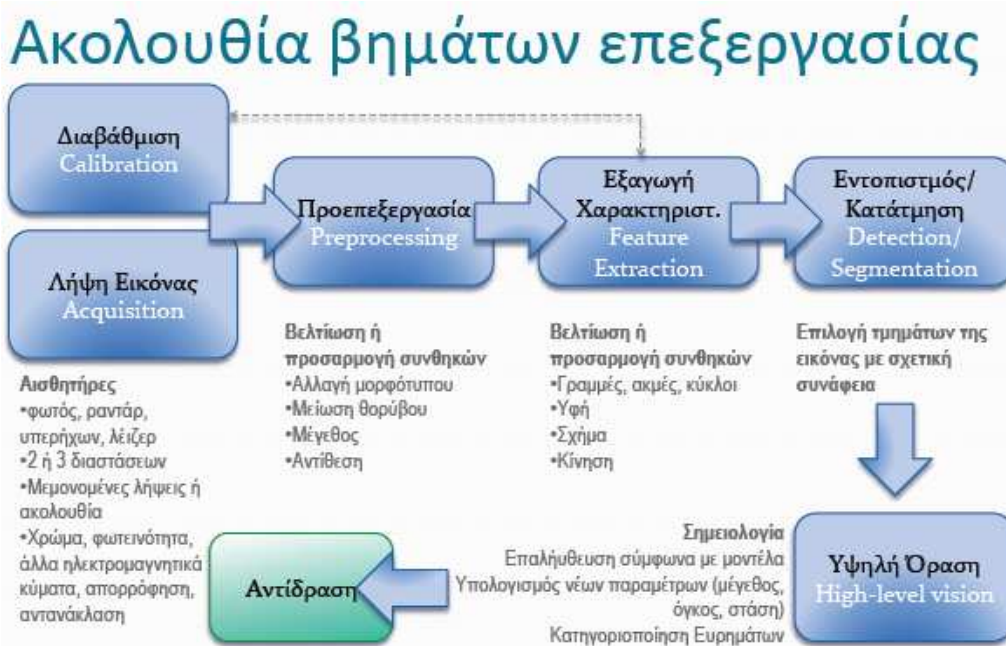
Εικόνα 9: Απεικόνιση του NASA's Mars Exploration Rover

Οι τομείς της εφαρμογής της τεχνητής όρασης είναι πάρα πολλοί. Κάθε εφαρμογή απασχολεί ένα φάσμα καθυκόντων όρασης περισσότερο ή λιγότερο γνωστά προβλήματα μέτρησης είτε προβλήματα επεξεργασίας, που μπορούν να λυθούν με μια ποικιλία μεθόδων. Η αναγνώριση είναι ένα από τα κλασσικά προβλήματα στην τεχνητή όραση, επεξεργασία εικόνας και μηχανική όραση και ασχολείται με το κατά πόσο ή όχι τα δεδομένα της εικόνας περιέχουν ποιο συγκεκριμένα αντικείμενα, χαρακτηριστικά ή δραστηριότητα. Οι μέθοδοι που χρησιμοποιούνται μπορούν μόνο να λύσουν ένα κομμάτι του προβλήματος και μπορούν να αναγνωρίσουν συγκεκριμένα αντικείμενα για συγκεκριμένες καταστάσεις. Μερικές διεργασίες σχετίζονται με την εκτίμηση κίνησης, όπου μια ακολουθία από εικόνες επεξεργάζεται για να παράγει μια εκτίμηση της ταχύτητας είτε σε κάθε σημεία στην εικόνα, είτε σε μια τρισδιάστατη απεικόνιση, είτε ακόμα για την κάμερα που παρέχει τις εικόνες. Μια ακόμα μέθοδος είναι η ανασυγκρότηση μιας σκηνής. Λαμβάνοντας υπόψη μια ή περισσότερες εικόνες από μια σκηνή είτε βίντεο, η ανασυγκρότηση στοχεύει στον υπολογισμό ενός τρισδιάστατου μοντέλου αυτής της σκηνής. Τέλος, μια ακόμα μέθοδος είναι η αποκατάσταση της εικόνας. Ο στόχος της αποκατάστασης μιας εικόνας είναι η αφαίρεση του θορύβου από τις εικόνες. Η απλούστερη προσέγγιση για την αφαίρεση του θορύβου είναι η χρήση διάφορων τύπων φίλτρων. Ποιο εξιδανικευμένες μεθόδους αναλαμβάνουν ένα μοντέλο στο πως είναι η δομή μιας εικόνας και το διακρίνει από τους θορύβους[3].



*Εικόνα 10: Μέθοδος τρισδιάστατης ανασυγκρότησης σκηνής*

Η οργάνωση ενός συστήματος όρασης υπολογιστή εξαρτάται πάρα πολύ από την εφαρμογή. Ορισμένα συστήματα είναι αυτόνομες εφαρμογές που λύνουν ή αναγνωρίζουν το πρόβλημα ενώ άλλες αποτελούν ένα υπό – σύστημα ενός μεγαλύτερου σχεδίου. Οποιαδήποτε η εφαρμογή υπάρχουν κάποιες βασικές λειτουργίες οι οποίες συναντώνται στα συστήματα της τεχνητής όρασης. Πρώτο και σημαντικό είναι η απόκτηση της εικόνας. Η εικόνα αποτελείται από έναν ή περισσότερους αισθητήρες εικόνας και από διαφορετικούς τύπους αισθητήρων. Ανάλογα με τον τύπο και την τεχνική που χρησιμοποιεί κάθε αισθητήρας προκύπτουν δισδιάστατες ή και τρισδιάστατες εικόνες. Έτσι λαμβάνετε η κύρια πληροφορία για κάθε σύστημα τεχνητής όρασης. Το επόμενο στάδιο είναι αυτό της προ - επεξεργασίας. Πριν από την χρήση των αλγορίθμων και των μεθόδων της τεχνητής όρασης είναι σημαντικό να προ επεξεργαστούν τα δεδομένα για να πληρούν τις προσδοκίες και παραδοχές που συνάγονται με την κάθε μέθοδο. Το επόμενο βήμα για ένα σύστημα τεχνητής όρασης είναι η εξαγωγή χαρακτηριστικών από μια εικόνα, όπως για παράδειγμα γραμμές, γωνίες και περιγράμματα. Επόμενο στάδιο είναι ο εντοπισμός και η κατάτμηση όπου το σύστημα της τεχνητής όρασης αποφασίζει για τους τομείς εργασίας του πάνω σε μια εικόνα. Τέλος είναι το κομμάτι το οποίο ονομάζετε υψηλή όραση όπου το αποτέλεσμα όλων των προηγούμενων βημάτων αξιολογείται από το σύστημα και το σύστημα αποφασίζει την αντίδραση του. Ο τελικός σχεδιασμός του συστήματος το οποίο περιγράφετε στην εργασία εκτελεί όλα τα παραπάνω βήματα.



Εικόνα 11: Ακολουθία βημάτων κατασκευής συστήματος τεχνητής όρασης

### 2.3.2 Τεχνητή νοημοσύνη

Τεχνητή νοημοσύνη είναι η νοημοσύνη που αναδεικνύουν οι μηχανές και ο τομέας της επιστήμης υπολογιστών που σκοπεύει να την δημιουργήσει[4]. Ο τομέας της τεχνητής νοημοσύνης ιδρύθηκε με την προσδοκία να καταφέρει την προσομοίωση της νοημοσύνης που βρίσκουμε στους ανθρώπους όπως και στα έμβια όντα. Η νοημοσύνη από παλαιά θεωρείτο ως μοναδικό προνόμιο των ανθρώπων, όπως αποδείχθηκε από την επιστήμη και άλλα πλάσματα έχουν μορφές νοημοσύνης. Η παραδοχή νοημοσύνης στις μηχανές έφερε φιλοσοφικά ερωτήματα και ηθικά διλήμματα για την συμπεριφορά μας απέναντι σε αυτά τα συστήματα. Σήμερα, η τεχνητή νοημοσύνη δεν δείχνει να επηρεάζεται από αυτά τα θέματα και είναι πλέον σημαντικό κομμάτι στην βιομηχανική παραγωγή και βοηθάει στην λύση πολλών προβλημάτων της επιστήμης υπολογιστών.

Στην τεχνητή νοημοσύνη έχουμε πολλά τεχνικά και ειδικά στοιχεία σε ότι αφορά την έρευνα και χωρίζετε σε υπό - πεδία τα οποία ορισμένες φορές δεν έχουν καλή επικοινωνία μεταξύ τους [4]. Τα υπό - πεδία έχουν ξεχωριστές ιδιότητες και αρχές, έχουν κατασκευαστεί από διαφορετικούς επιστήμονες και λύνουν διαφορετικά προβλήματα. Τα υπό - πεδία αυτά προσφέρουν εργαλεία, τεχνικές και αλγόριθμους που δεν είναι πανάκεια για κάθε πρόβλημα. Έτσι οι μέθοδοι μπορεί να μην είναι καθορισμένοι αφού τα προβλήματα μπορεί να μην έχει εξεταστεί στο παρελθόν.

Η πρώτη τεχνητή νοημοσύνη εμφανίζετε σαν ιδέα από τους αρχαίους ελληνικούς μύθους. Σε αυτούς τους μύθους εμφανίζονται σκεπτόμενες μηχανές όπως το μηχανικό μπρούτζινο ρομπότ της Κρήτης ο Τάλως. Ακόμα αναφορές μύθων έχουμε και από άλλους πολιτισμούς όπως οι Αιγύπτιοι που μιλούν για αυτοματοποιημένα ανθρωποειδή. Τον 19ο και 20ο αιώνα, τα τεχνητά οντά μεταπήδησαν από τους αρχαίους μύθους στην μεγάλη οθόνη και την επιστημονική φαντασία. Αώο την επιστημονική της πλευρά η μηχανική ή “επίσημος” συλλογισμός αναπτύχθηκε από φιλοσόφους και μαθηματικούς από την αρχαιότητα. Η μελέτη της λογικής οδηγεί στην εφεύρεση των προγραμματιζόμενων ψηφιακών ηλεκτρονικών υπολογιστών βασισμένα στην δουλειά του μαθηματικού Alan Turing και άλλων. Η θεωρία του Turing υποστήριζε ότι με τα σύμβολα 0 και 1 και οι συνδυασμοί τους είναι ικανές να προσομοιάσουν κάθε δυνατή μαθηματική πράξη. Αυτή η θεωρία μαζί με ανακαλύψεις στην νευρολογία, την θεωρία της πληροφορική και των δικτύων, ενέπνευσε μια ομάδα ερευνητών στον ενδεχόμενο κατασκευής ηλεκτρονικού εγκεφάλου[4].

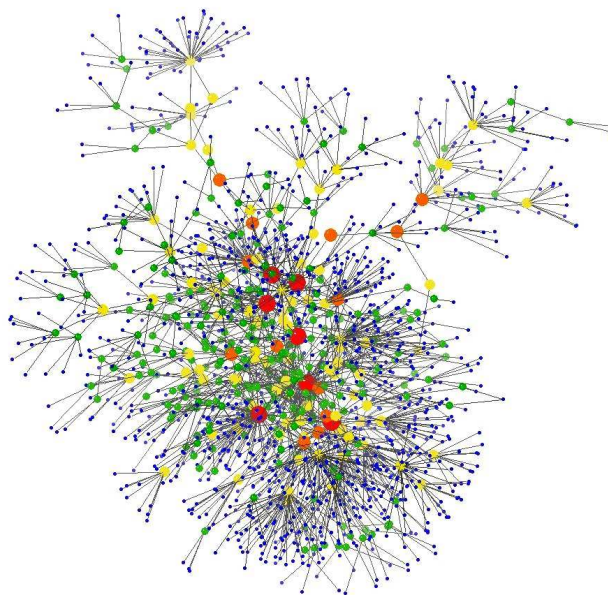
Το πεδίο της τεχνητής νοημοσύνης ιδρύθηκε σε ένα συνέδριο στην πανεπιστημιούπολη του Dartmouth

college το καλοκαίρι του 1956. Οι συμμετέχοντες, συμπεριλαμβανομένων των John McCarthy, Marvin Minsky, Allen Newell και Herbert Simon, που έγιναν επικεφαλής της έρευνας της τεχνητής νοημοσύνης για πολλά χρόνια. Αυτοί και οι μαθητές τους έγραψαν προγράμματα, για τούς περισσότερους ανθρώπους ήταν άπλα εκπληκτικό, το γεγονός ότι οι υπολογιστές ήταν σε θέση να λύνουν προβλήματα άλγεβρας, να αποδεικνύουν θεωρήματα και να μιλούν την αγγλική γλώσσα. Στα μέσα της δεκαετίας του 60' η έρευνα τους χρηματοδοτούνταν από το υπουργείο αμύνης των Ηνωμένων Πολιτειών της Αμερικής. Οι ιδρυτές της τεχνητής νοημοσύνης ήταν βαθύτατα αισιόδοξοι για το μέλλον του πεδίου. Ο Herbert Simon προέβλεψε ότι “οι μηχανές θα είναι σε θέση, μέσα σε είκοσι χρόνια, να κάνουν οποιαδήποτε εργασία ένας άνθρωπος μπορεί να κάνει” και ο Marvin Minsky συμφώνησε λέγοντας ότι σε μια γενιά το πρόβλημα της δημιουργίας τεχνητής νοημοσύνης θα οδηγηθεί στην επίλυση του. Δεν κατάφεραν να αναγνωρίσουν ορισμένα προβλήματα τα οποία αντιμετώπισαν. Το 1974 σε απάντηση του Sir James Lighthill και συνεχή πίεση από το κογκρέσο για την χρηματοδότηση ποιο παραγωγικών έργων, οι κυβερνήσεις των Η.Π.Α και του Ηνωμένου Βασιλείου, σταματήσαν την έρευνα πάνω στην τεχνητή νοημοσύνη. Αυτή η περίοδος ονομάστηκε “χειμώνας της τεχνητής νοημοσύνης”. Στις αρχές του 1980, η τεχνητή νοημοσύνη αναβίωσε από την εμπορική επιτυχία των έμπειρων συστημάτων, μιας μορφής τεχνητής νοημοσύνης που προσομοιώνει γνώση και δεξιότητες ανάλυσης από περισσότερους από έναν ανθρώπινο εμπειρογνώμονα. Στα τέλη του 1985 η αγορά για την τεχνητή νοημοσύνη ξεπέρασε το ένα δισεκατομμύριο δολάρια. Την ίδια στιγμή ο πέμπτης γενιάς Ιαπωνικός υπολογιστής ανάγκασε της κυβερνήσεις των Η.Π.Α και του Ηνωμένου Βασιλείου να επαναφέρουν τα κεφάλαια για την ακαδημαϊκή έρευνα του τομέα. Όμως, με την πτώση της μηχανής Lisp στην αγορά το 1987, η τεχνητή νοημοσύνη έπεσε ξανά για δεύτερη φορά σε “χειμώνα της τεχνητής νοημοσύνης”. Στις αρχές της δεκαετίας του 90' και την αρχή του 21 αιώνα, η τεχνητή νοημοσύνη κατάφερε της μεγαλύτερες επιτυχίες. Η τεχνητή νοημοσύνη χρησιμοποιήθηκε για λογιστικούς σκοπούς, διαγνωστικούς σκοπούς και είχε πολλές ακόμα εφαρμογές στην βιομηχανία της τεχνολογίας. Η επιτυχία της στηρίζεται σε πολλούς παράγοντες όπως η αύξηση της υπολογιστικής δύναμης, μεγαλύτερη έμφαση δόθηκε στην λύση συγκεκριμένων υπό - προβλημάτων, οι δεσμοί μεταξύ της τεχνητής νοημοσύνης και επιστημών που εργάζονταν σε παρόμοια θέματα, και οι προσήλωση των ερευνητών σε μαθηματικούς μεθόδους[4].

Το γενικό πρόβλημα να προσομοιώνουμε ή να παράγουμε γνώση έχει κατακερματιστεί σε έναν αριθμό από υπό - προβλήματα. Το πρώτο ονομάζεται αφαιρετικότητα, λογική και επίλυση προβλημάτων που απασχολεί τους ερευνητές από το ξεκίνημα του τομέα της τεχνητής νοημοσύνης. Οι ερευνητές έχουν κατασκευάσει αλγόριθμους που μιμούνται την λογική σε βήματα για να λύνουν γρίφους είτε παίζουν επιτραπέζια παιχνίδια. Στα τέλη του 80' και αρχές του 90' η έρευνα πάνω στην τεχνητή νοημοσύνη μπόρεσε επιτυχώς να παράγει μεθόδους που μπορούν να ελέγχουν μη σίγουρα ή ανολοκλήρωτα δεδομένα, χρησιμοποιώντας τις πιθανότητες και τα οικονομικά. Για δύσκολα προβλήματα οι περισσότεροι αλγόριθμοι απαιτούν μεγάλα μεγέθη επεξεργαστικής ισχύς και μνήμης και αυτό μπορεί να προκαλέσει προβλήματα στο σύστημα μας. Τα ανθρώπινα οντά όμως δεν χρησιμοποιούν αυτή την μέθοδο για παρόμοιους λογούς έτσι νέοι τρόποι ερευνώνται σε αυτήν την κατεύθυνση[4].

Ένα πρόβλημα ακόμα είναι και η αναπαράσταση γνώσης που πρόκειται για κεντρικό κομμάτι της έρευνας της τεχνητής νοημοσύνης. Πολλά από τα προβλήματα που αναμένονται να λυθούν από τις μηχανές θα χρειαστούν εκτεταμένη γνώση για τον κόσμο. Ανάμεσα σε άλλα πράγματα η τεχνητή νοημοσύνη πρέπει να αναπαριστά αντικείμενα, ιδιότητες, κατηγορίες και σχέσεις μεταξύ των αντικειμένων. Επίσης πρέπει να αναπαρασταθούν γεγονότα, χρονικές στιγμές, καταστάσεις, αιτίες και αποτελέσματα. Μια πλήρη αναπαράσταση “του τι υπάρχει” είναι μια οντολογία μια έννοια δανεισμένη από την φιλοσοφία, εκ των οποίων ονομάζονται άνω οντολογίες. Υπό - προβλήματα της αναπαράστασης γνώσης είναι η κανονική λογική και το πρόβλημα προσόντων. Πολλά από αυτά τα οποία γνωρίζουν οι άνθρωποι προέρχονται από υποθέσεις, ακόμα και η περιγραφή ενός πτηνού μπορεί να γίνει με διαφορετικό τρόπο από διαφορετικούς ανθρώπους έτσι δημιουργείτε πρόβλημα στην ταυτοποίηση στο τι μπορεί να είναι πτηνό και τι όχι. Ένα υπό - πρόβλημα που γενάτε από την αναπαράσταση γνώσης είναι το εύρος της κοινής γνώσης. Η μέση γνώση ενός ανθρώπου αποτελείται από πάρα πολλά πράγματα, για την κατασκευή μιας τέτοιας βάσης δεδομένων από τους ερευνητές είναι αδύνατη. Έτσι προκύπτει η ανάγκη για την δημιουργία συστημάτων που χρησιμοποιούν μόνο τους πηγές και να κατασκευάζουν μονά τους οντολογίες. Τέλος, υπάρχει και το πρόβλημα της υπό - συμβολική μορφή της κοινής γνώσης. Αυτή η γνώση παρέχετε με την εμπειρία και δεν μπορεί πάντα να περιγράφει με την κοινή μορφή, πρόκειται για διαισθήσεις ή τάσεις που εκπροσωπούνται στον εγκέφαλο μη συνειδητά και υπό - συμβολικά[4].





*Εικόνα 12: Δίκτυο από οντολογίες και οι μεταξύ τους δεσμοί*

Ο σχεδιασμός προκύπτει να είναι ακόμα ένα από τα μεγάλα προβλήματα που έχει να αντιμετωπίσει ο τομέας της τεχνητής νοημοσύνης. Τα ευφυή συστήματα πρέπει να είναι σε θέση να θέτουν στόχους και να τούς επιτυγχάνουν. Χρειάζονται έναν τρόπο με τον οποίο να προβλέπουν το μέλλον και να κάνουν επιλογές με στόχο την μεγιστοποίηση του αποτελέσματος. Στον κλασικό σχεδιασμό το σύστημα παίρνει σαν δεδομένα και ενεργεί μόνο του και όλα είναι αποτέλεσμα της πράξης που εκτελεί. Στην πραγματικότητα αυτό δεν ισχύει, το σύστημα πρέπει να είναι σε θέση να υπολογίζει εάν οι πράξεις του κινούνται γύρω από τις προβλέψεις που κάνει και να προσαρμόζετε αν οι συνθήκες το απαιτούν[4].

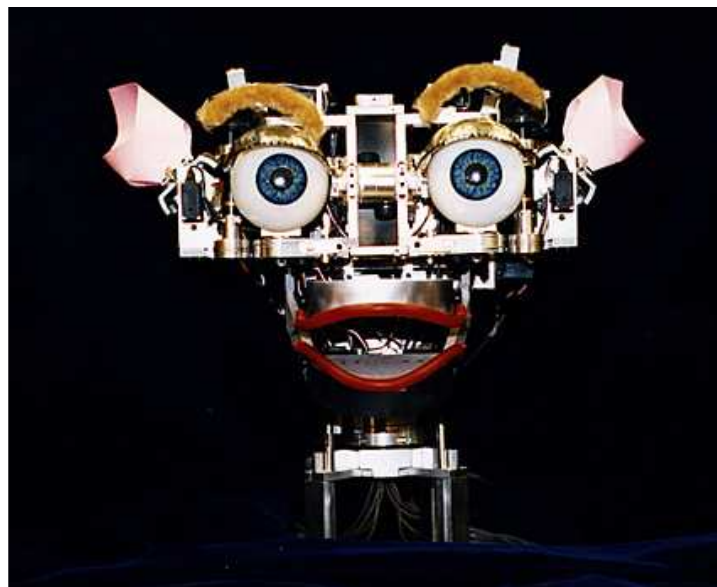
Η μηχανική μάθηση είναι ένα κεντρικό κομμάτι στην τεχνητή νοημοσύνη και απασχολεί ερευνητές από την αρχή της δημιουργίας του κλάδου. Μη επιβλεπόμενη μάθηση είναι η ιδιότητα της εύρεσης προτύπων σε ένα ρεύμα από εισόδους. Η επιβλεπόμενη μάθηση περιλαμβάνει τόσο την ταξινόμηση όσο και την αριθμητική παλινδρόμηση. Η κατηγοριοποίηση χρησιμοποιείται για να καθορίσει σε ποια κατηγορία ανήκει κάτι, μετά από εκπαίδευση με μια σειρά παραδειγμάτων και επαναλήψεων. Η παλινδρόμηση προσπαθεί να πάρει το σύνολο των εσόδων και να ανακαλύψει μια συνάρτηση ανάμεσα σε αυτές και τις τελικές εξόδους. Στην ενισχυμένη μάθηση το σύστημα ανταμείβετε για τα καλά αποτελέσματα και τιμωρείτε για τα κακά. Αυτό μπορούν να αναλυθούν από την άποψη της θεωρίας των αποφάσεων τη χρήση εννοιών όπως το βοηθητικό πρόγραμμα. Η μαθηματική ανάλυση των αλγορίθμων μηχανικής μάθησης και των επιδόσεων τους είναι κλάδος της θεωρητικής επιστήμης υπολογιστών γνωστή και ως υπολογιστική θεωρία μάθησης.

Η επεξεργασία φυσικής γλώσσας δίνει την ικανότητα στις μηχανές να διαβάζουν και να κατανοούν τις φυσικές γλώσσες, δηλαδή τις γλώσσες που μιλούν οι άνθρωποι. Πολλοί ερευνητές πιστεύουν ότι ένα αρκετά ισχυρό φυσικό σύστημα επεξεργασίας της γλώσσας θα είναι σε θέση να αποκτηθεί αυτόματα γνώση από το διαδίκτυο. Μερικές άπλες εφαρμογές επεξεργασίας φυσικής γλώσσας περιλαμβάνουν ανάκτηση πληροφοριών και μηχανική μετάφραση. Ακόμα ένα πρόβλημα της σύγχρονης τεχνητής νοημοσύνης έχει να κάνει σχέση και με έναν άλλον τομέα αυτόν της ρομποτικής. Η κίνηση και η πράξεις χειραγώγησης είναι ένα κομμάτι το οποίο είναι απαραίτητο για την πλοήγηση, τον εντοπισμό και την χαρτογράφηση. Ακόμα ένα πρόβλημα της τεχνητής νοημοσύνης είναι η αντίληψη μηχανής. Η μηχανική αντίληψη είναι η ικανότητα τις χρήσης των στοιχείων που προέρχονται από αισθητήρες για να κατανοήσει πτυχές του φυσικού κόσμου[4].



*Εικόνα 13: Ο ASIMO εφαρμόζει πολλές τεχνικές τεχνητής νοημοσύνης*

Τα αισθήματα και η κοινωνικές δεξιότητες παίζει δυο ρόλους για ένα ευφρές σύστημα. Πρώτον, πρέπει να είναι σε θέση να προβλέπει τις ενέργειες των άλλων, καταλαβαίνοντας τις συναισθηματικές καταστάσεις. Επίσης για την καλή αλληλεπίδραση μεταξύ συστήματος και χρήστη μια έξυπνη μηχανή επίσης απαιτεί την εμφάνιση των αισθημάτων τους. Θα πρέπει τουλάχιστον να πρέπει να εμφανίζεται ευγενικό και ευαίσθητο με τους ανθρώπους που αλληλεπιδρά. Στην καλύτερη περίπτωση, θα πρέπει να έχει το ίδιο φυσιολογικά συναισθήματα[4].



*Εικόνα 14: Ο Kismet είναι ένα ρομπότ που προσομοιώνει ανθρώπινα συναισθήματα*

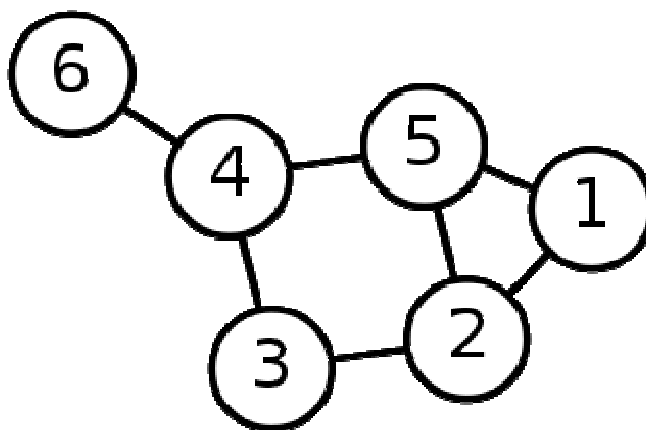
Η δημιουργικότητα είναι άλλο ένα θέμα για τα ευφρή συστήματα που πραγματοποιείτε τόσο θεωρητικά όσο

και πρακτικά και η σχετική περιοχή της υπολογιστικής ερευνάς ονομάζετε τεχνητή διαίσθηση και τεχνητή φαντασία. Τέλος, το μεγαλύτερο πρόβλημα αλλά και αυτό το οποίο απασχολεί πολλούς ερευνητές είναι αυτό της γενικής νοημοσύνης. Η γενική νοημοσύνη είναι ο συνδυασμός όλων των παραπάνω δεξιοτήτων και ανθρώπινων δυνατοτήτων σε ένα κοινό σύστημα. Λίγοι είναι αυτοί που πιστεύουν ότι ανθρωπόμορφα χαρακτηριστικά όπως τεχνητή συνείδηση ή ένας τεχνητός εγκέφαλος είναι ικανά να επιτευχθούν[4].

Οι εφαρμογές των τεχνικών του πεδίου της τεχνητής νοημοσύνης είναι ευρέως διαδεδομένες και είναι πάρα πολλές. Αρχικά οι εφαρμογές δεν ήταν πολύ πολύπλοκες αλλά ήταν εξαιρετικά επιτυχείς, έτσι δειλά δειλά άρχισε η ευρεία χρήση των τεχνικών της τεχνητής νοημοσύνης. Οι πρώτες εφαρμογές χρησιμοποιήθηκαν εκεί όπου υπήρχαν πόροι για επένδυση και ανάγκη για μεγαλύτερη αποτελεσματικότητα, αυτοί οι τομείς ήταν η βιομηχανία, η επιστήμη και άμυνα. Στην βιομηχανία η τεχνητή νοημοσύνη χρησίμευσε στην αύξηση της παραγωγικότητας είτε ελέγχοντας είτε εξελίσσοντας τα συστήματα παραγωγής είτε κάνοντας καλύτερη διαχείριση υλικών και εργασιακών πόρων. Στην επιστήμη η χρήση της τεχνητής νοημοσύνης χρησιμοποιήθηκε από άλλους τομείς όπως τα μαθηματικά και η φυσική. Η χρήση των τεχνικών της έγινε για να παραχθούν αποτελέσματα τα οποία ήταν δύσκολο να υπολογιστούν από απλούς αλγορίθμους. Επίσης, πολλές εφαρμογές υπάρχουν στην αμυντική βιομηχανία, όπως αυτόνομα οχήματα, κατευθυνόμενοι αυτοδιαχειριζόμενοι πύραυλοι και συγχρονισμός αμυντικών εξοπλισμών. Τέλος, η πιο κοινή χρήση αυτών των εφαρμογών πλέον γίνεται στην καθημερινή μας ζωή όπου πολλές από τις συσκευές που χρησιμοποιούμε κάθε μέρα έχουν ενσωματώσει αλγορίθμους τεχνητής νοημοσύνης.

### 2.3.3 Θεωρία των γράφων

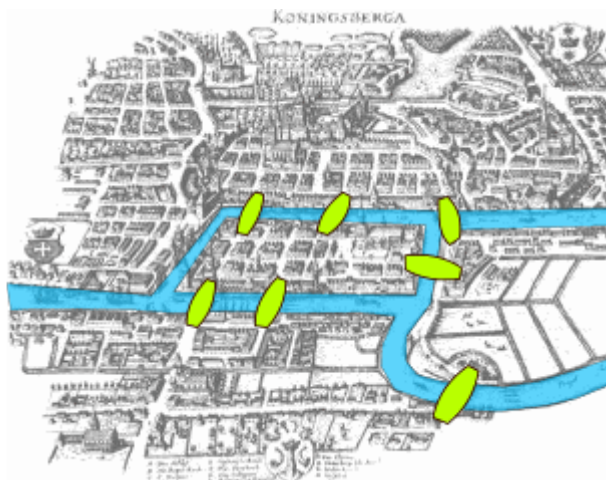
Στα μαθηματικά και την επιστήμη των υπολογιστών, αυτή η θεωρία μελετά τους άγραφους. Οι γράφου είναι μαθηματικές δομές όπου χρησιμοποιούνται για την μοντελοποίηση των σχέσεων ζεύγους μεταξύ αντικειμένων από μια συγκεκριμένη συλλογή. Ένας “γράφος” στο πλαίσιο αυτό αναφέρετε σε μια συλλογή των κορυφών ή “κόμβων” και μια συλλογή από ακμές που ενώνουν τα ζεύγη των κορυφών. Ένα γράφος μπορεί να είναι μη κατευθυνόμενος που σημαίνει ότι δεν υπάρχει καμιά διάκριση μεταξύ των δυο κορυφών που συνδέονται από κάθε ακμή, ή οι ακμές του μπορεί να είναι κατευθυνόμενες από την μια κορυφή στην άλλη. Οι γράφοι που μελετώνται στην θεωρία των γράφων δεν πρέπει να συγχέονται με παραστάσεις από λειτουργίες ή άλλα γραφήματα[12].



Εικόνα 15: Παράδειγμα γράφου

Η θεωρία των γράφων ξεκινά με μια δημοσίευση του Leonhard Euler πάνω στις επτά γέφυρες του Königsberg το 1736. Αυτή η δημοσίευση, καθώς και αυτή που γράφτηκε από τον Vandermonde σχετικά με το πρόβλημα του ιππότη, συνέχισε με την ανάλυση τοποθεσίας που ξεκίνησε ο Leibniz. Ο τύπος του Euler που αφορά τον αριθμό των ακμών, των κορυφών, και αντιμετωπίζει ένα κυρτό πολύεδρο που μελετήθηκε

και γενικεύτηκε από τον Cauchy και τον L'Huilier και είναι η καταγωγή της τοπολογίας[12].



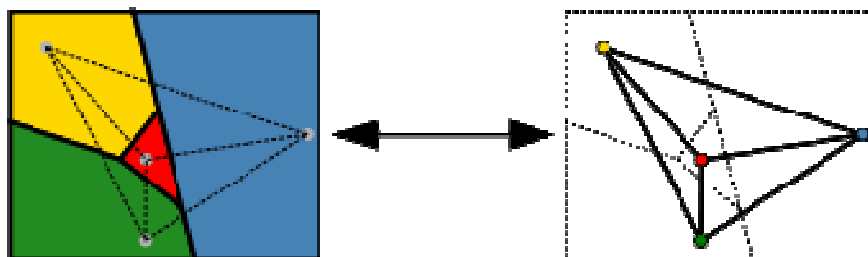
Εικόνα 16: Το πρόβλημα των 7 γεφυρών της πόλης Königsberg.

Πάνω από έναν αιώνα μετά την δημοσίευση του Euler για τις επτά γέφυρες του Königsberg και καθώς ο Listing εισήγαγε την τοπολογία, ο Cayley οδήγησε στην μελέτη συγκεκριμένων αναλυτικών μορφών που προκύπτουν από τον διαφορικό λογισμό για την μελέτη συγκεκριμένης κατηγορίας γράφων, τα δένδρα. Η μελέτη αυτή είχε μεγάλες επιπτώσεις στην θεωρητική χημεία. Η εμπλεκόμενες τεχνικές κυρίως αφορούσαν την απαρίθμηση των γράφων που έχουν ειδικές ιδιότητες. Η θεωρία απαριθμημένων γράφων εξάχθηκε από τα αποτελέσματα του Cayley και τα θεμελιώδη αποτελέσματα που δημοσιεύτηκαν από τον Ρόlya ανάμεσα στο 1935 και 1937 και την γενίκευση τους από τον De Bruijn το 1959. Ο Cayley συνέδεσε τα αποτελέσματα από τα δένδρα με τις σύγχρονες μελέτες των χημικών συνθέσεων. Η σύντηξη των ιδεών που προέρχονταν από τα μαθηματικά με αυτές που προέρχονταν από την χημεία είναι η προέλευση ενός μέρους της βασικής ορολογίας της θεωρίας των γράφων[12].

Ειδικότερα, ο όρος γράφος εισήχθη από τον Sylvester σε μια δημοσίευση του το 1878 στο περιοδικό Nature, όπου επιστά την αναλογία μεταξύ “κβαντικά αναλλοίωτων” και “συν – μεταβλητών” της άλγεβρας και των μοριακών διαγραμμάτων:

“[...] Κάθε αμετάβλητη και συν – μεταβλητή γίνετε εκφραστική μέσω ενός γράφου ακριβώς όμοια με ένα Kekuléan διάγραμμα ή χημειογράφο.[...] δίνω έναν κανόνα για γεωμετρικό πολλαπλασιασμό γράφων, πχ για την κατασκευή ενός γράφου το προϊόν τον  $a$ - και συν-μεταβλητών των οποίων δίνονται οι ξεχωριστές παραστάσεις.[...] “. [12]

Ένα από τα διασημότερα και παραγωγικά προβλήματα της θεωρίας των γράφων είναι το πρόβλημα των τεσσάρων χρωμάτων : “Είναι αλήθεια ότι κάθε χάρτης σχεδιασμένος σε επίπεδο μπορεί να έχει περιφέρειες χρωματισμένες με τέσσερα χρώματα, με τέτοιο τρόπο έτσι ώστε κάθε δυο περιοχές με κοινά σύνορα να έχουν διαφορετικά χρώματα ;”. Αυτό το πρόβλημα που πρώτος έθεσε ο Francis Guthrie το 1852 και καταγράφηκε πρώτη φορά σε ένα γράμμα του De Morgan που έστειλε στον Hamilton την ίδια χρονιά. Πολλές ανακριβείς αποδείξεις είχαν προταθεί, συμπεριλαμβανομένου αυτών από τον Cayley, τον Kempe και άλλους. Η μελέτη αυτού του προβλήματος από τους Tait, Heawood, Ramsey και τον Hadwiger οδήγησε στην μελέτη των χρωματισμών του γράφου ενσωματωμένα σε επιφάνειες με αυθαίρετο γένος. Η αναδιατύπωση του Tait που δημιούργησε μια νέα κατηγορία του προβλήματος, το πρόβλημα της παραγοντοποίησης, που μελετήθηκε ειδικότερα από τον Petersen και τον König. Τα έργα του Ramsey σε χρωματισμούς και ειδικότερα στα αποτελέσματα του Turán το 1941 αποτέλεσε την αφετηρία ενός άλλου κλάδου της θεωρίας των γράφων, την ακραία θεωρία των γράφων[12].



Εικόνα 17: Απλό παράδειγμα γράφου του προβλήματος των τεσσάρων χρωμάτων

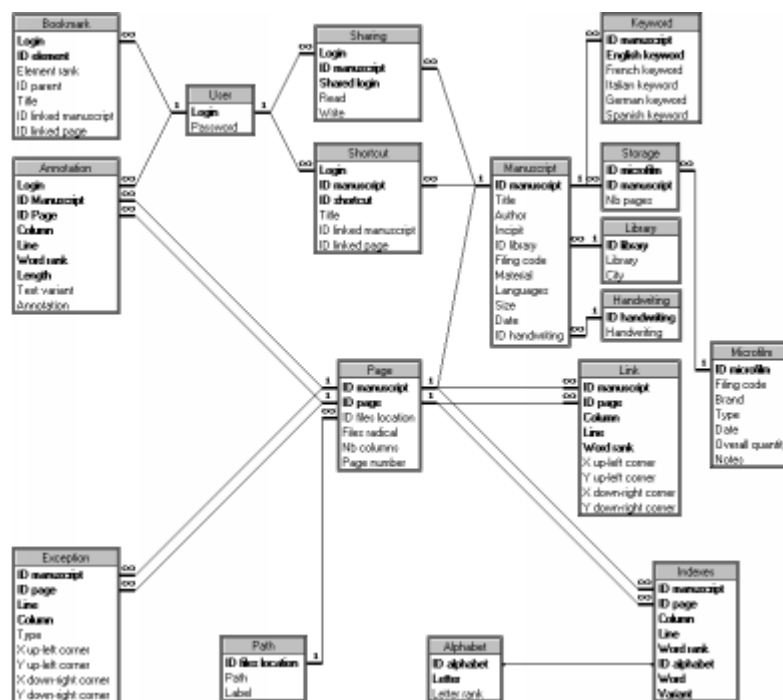
Το πρόβλημα των τεσσάρων χρωμάτων παρέμεινε άλυτο για έναν αιώνα. Το 1969 ο Heinrich Heesch δημοσιεύει μια μέθοδο όπου λύνει το πρόβλημα με την χρήση των υπολογιστών. Μια απόδειξη με την βοήθεια του υπολογιστή παράχθηκε το 1976 από τον Kenneth Appel και τον Wolfgang Haken κάνει θεμελιώδη την χρήση της έννοιας της “απόρριψης” που αναδείχθηκε από τον Heesch. Η απόδειξη απαιτούσε τον έλεγχο χαρακτηριστικών από 1936 συνθέσεων από υπολογιστή, και δεν ήταν πλήρως αποδεκτή την εποχή εκείνη λόγω της πολυπλοκότητας της. Μια πιο απλούστερη απόδειξη λαμβάνοντας υπόψιν μόνο 633 συνθέσεις δόθηκε 20 χρόνια αργότερα από τους Robertson, Seymour, Sanders και Thomas.

Η αυτόνομη ανάπτυξη της τοπολογίας από το 1860 και το 1930 γονιμοποίησε την θεωρία των γράφων μέσα από τα έργα των Jordan, Kuratowski και Whitney. Ένας άλλος σημαντικός παράγοντας της κοινής ανάπτυξης της θεωρίας των γράφων και της τοπολογίας προήλθε από την χρήση των τεχνικών της σύγχρονης άλγεβρας. Το πρώτο παράδειγμα μιας τέτοιας χρήσης προερχόταν από το έργο του μεγάλου φυσικού Gustav Kirchhoff, που δημοσίευσε το 1845 τους νόμους των κυκλωμάτων του Kirchhoff για τον υπολογισμό των τάσεων του ρεύματος στα ηλεκτρικά κυκλώματα[12].

Η εισαγωγή των πιθανοτικών μεθόδων στην θεωρία των γράφων, ιδιαίτερα στην μελέτη των Erdős και Rényi για την ασυμπτωτική πιθανότητα σύνδεσης γράφων, δημιούργησε ακόμα ένα κλάδο, γνωστός ως θεωρία τυχαίων γράφων, όπου είναι μια γόνιμη πηγή από γράφω – θεωρητικά αποτελέσματα[12].

Οι γράφοι έχουν πάρα πολλές εφαρμογές στις επιστήμες, είναι τα πιο συχνά εμφανιζόμενα μοντέλα που περιγράφουν φυσικές και ανθρώπινες κατασκευές. Χρησιμοποιούνται για την μοντελοποίηση πολλών τύπων σχέσεων και δυναμικές διαδικασίες στην φυσική, την βιολογία και σε κοινωνικά συστήματα. Πολλά προβλήματα πρακτικού ενδιαφέροντος μπορούν να αναπαρασταθούν με γράφους.

Στην επιστήμη των υπολογιστών, οι γράφοι χρησιμοποιούνται για την αναπαράσταση δικτύων, οργάνωση δεδομένων, υπολογιστικές συσκευές, την ροή του υπολογισμού, κ.α. Ένα πρακτικό παράδειγμα είναι: η συνδεσμική δομή μιας ιστοσελίδας μπορεί να παρασταθεί με έναν κατευθυνόμενο γράφο. Οι κορυφές είναι οι ιστοσελίδες που είναι διαθέσιμες στην ιστοσελίδα και μια κατευθυνόμενη ακμή από την σελίδα A στην B υπάρχει αν και μόνο αν η A περιέχει έναν σύνδεσμο για την B. Μια παρόμοια προσέγγιση μπορεί να θεωρηθεί για προβλήματα σε ταξίδια, την βιολογία, την σχεδίαση υλικού υπολογιστών, και πολλούς άλλους τομείς. Η ανάπτυξη των αλγορίθμων τον έλεγχο των γράφων είναι για αυτόν τον λόγο πολύ μείζονος ενδιαφέροντος στην επιστήμη των υπολογιστών. Εκεί, η μετατροπή των γράφων είναι συχνά επισημοποιημένη και εκπροσωπεί από συστήματα επανεγγράψιμου γράφου. Είναι είτε συχνά χρησιμοποιούνται άμεσα είτε μελετώνται οι ιδιότητες των επανεγγράψιμων συστημάτων (πχ συμβολή στο σύστημα). Συμπληρωματικά στα συστήματα μετατρεπόμενων γράφων που στηρίζονται σε ρόλους χειραγώγησης της μνήμης των γράφων είναι οι γράφοι βάσεων δεδομένων. Αυτοί οι γράφοι προσανατολίζονται προς συναλλακτικά ασφαλείς τρόπους, με συνεχής και επίμονης αποθήκευσης και αναζήτησης των δομημένων σε γράφους δεδομένων.



Εικόνα 18: Απεικόνιση βάσεων δεδομένων με την βοήθεια γράφων.

Επίσης συναντάμε τους γράφους και σε άλλους τομείς της επιστήμης υπολογιστών. Οι γράφοι στην τεχνητή όραση αποτελούν κομμάτια μιας εικόνας είτε ολόκληρη την εικόνα, έτσι βοηθούν στην κατανόηση της δομής της εικόνας και των περιεχομένων της σε πολύ μικρά επίπεδα. Επίσης η γράφοι συναντώνται και στην τεχνητή νοημοσύνη όπου χρησιμοποιούνται και σαν δενδροειδής δομές. Ιδιότητες και θεωρήματα των γράφων χρησιμοποιούνται για την κατάταξη, την ταξινόμηση και την κατηγοριοποίηση στοιχείων σε κλάσεις. Επίσης, συχνή χρήση τους γίνεται στην αναγνώριση προτύπων και σε τεχνικές νοημοσύνης δικτύων. Μια δομή γράφου μπορεί να επεκταθεί με την ανάθεση ενός βάρους σε κάθε ακμή του κάθε γράφου. Οι γράφοι με βάρη χρησιμοποιούνται για την αναπαράσταση δομών στις οποίες οι συνδέσεις των ζευγών έχουν κάποιες αριθμητικές τιμές. Για παράδειγμα εάν ένας γράφος αναπαριστά το οδικό δίκτυο, τότε τα βάρη είναι η αναπαράσταση του μήκος κάθε δρόμου.

Ένας κατευθυνόμενος γράφος με σταθμισμένες ακμές στο πλαίσιο της θεωρίας των γράφων ονομάζεται δίκτυο. Η ανάλυση δικτύων έχει πολλές πρακτικές εφαρμογές, για παράδειγμα, η μοντελοποίηση και η ανάλυση δικτύων μεταφοράς. Εφαρμογές της ανάλυσης δικτύου χωρίζονται γενικά σε τρεις κατηγορίες. Η πρώτη είναι η ανάλυση για τον περιορισμό δομικών ιδιοτήτων του δικτύου, όπως η διανομή των βαθμών των κορυφών και της διαμέτρου ενός γράφου. Ένας τεράστιος αριθμός μέτρων υπάρχει στους γράφους, και η παραγωγή των χρήσιμων για διάφορους τομείς παραμένει ένας ενεργός τομέας έρευνας. Η δεύτερη είναι η ανάλυση για να βρούμε μια μετρήσιμη ποσότητα μέσα σε ένα δίκτυο, για παράδειγμα, για ένα δίκτυο μεταφοράς, το επίπεδο της ροής των οχημάτων σε οποιοδήποτε τμήμα του. Τέλος, η ανάλυση των δυναμικών ιδιοτήτων των δικτύων.

## 2.4 Αλγόριθμοι

Οι αλγόριθμοι είναι το κομμάτι στο οποίο στηρίζετε η δημιουργία ενός συστήματος. Αφού παραπάνω αναλύθηκαν οι θεωρίες που περικλείουν αυτούς τους αλγορίθμους είναι πλέον εύκολο να κατανοήσουμε τους σκοπούς που θέλουν να επιτύχουν. Στο σύστημα που περιγράφετε σε αυτήν την πτυχιακή η αλγόριθμοι είναι συνήθως κομμάτια από μια ή περισσότερες από τις παραπάνω θεωρίες. Έτσι παρακάτω εμφανίζονται οι αλγόριθμοι με την σειρά που εκτελούνται από το πρόγραμμα.

Πρώτος είναι ο αλγόριθμος υπολογισμού οπτικής αντιστοιχίας μέσω της τόνων των γράφων (Graph cuts). Αυτός ο αλγόριθμος ανήκει στην θεωρία της τεχνητής όρασης και χρησιμοποιεί την θεωρία των γράφων. Αυτός ο αλγόριθμος χρησιμοποιεί σαν εισόδους ένα ζεύγος στερεοσκοπικών εικόνων. Από αυτές τις εικόνες ζητάτε να βρεθεί χάρτης βάθους για περαιτέρω επεξεργασία. Ο χάρτης βάθους με αυτήν την μέθοδο επεξεργάζεται εντελώς διαφορετικά από την κλασσική μέθοδο. Αρχικά η οπτική απόκριση των εικόνων γίνεται μέσω της θεωρίας των γράφων και η ελαχιστοποίηση της ενέργειας των μη εμφανών περιοχών με την βοήθεια των θεωρημάτων της ίδιας θεωρίας. Έτσι με αυτήν την μέθοδο έχουμε την καλύτερη δυνατή οπτική απόκριση και ως αποτέλεσμα έναν ποιο ακριβή χάρτη βάθους.

Ο επόμενος αλγόριθμος είναι αυτός των συνδεδεμένων στοιχείων ο οποίος και αυτός ανήκει στην θεωρία των γράφων αλλά χρησιμοποιείτε από την τεχνητή νοημοσύνη σαν μια ήπια μορφή εύρεσης προτύπων. Αυτή η μέθοδος προσπαθεί να ομαδοποιήσει με τον βέλτιστο τρόπο τους γράφους που του δίνονται. Στο σύστημα χρησιμοποιείτε για κατηγοριοποίηση αντικείμενων από σε μια εικόνα.

Ακόμα ένας αλγόριθμος που χρησιμοποιήθηκε στο σύστημα ήταν ο μετασχηματισμός κλιμακωτά – αναλλοίωτων χαρακτηριστικών ( SIFT ). Σε αυτό τον αλγόριθμο οφείτε και το αποτέλεσμα του συστήματος. Όταν εισάγετε μια εικόνα σε αυτόν τον αλγόριθμο τότε εξάγονται κάποια χαρακτηριστικά. Αυτά τα χαρακτηριστικά κρατούν αναλλοίωτες τις ιδιότητες τους με τέτοιο ώστε να μην επηρεάζονται από την κλίμακα αλλά ούτε και από την κλίση της εικόνας. Ακόμα ένα πολύ σημαντικό γεγονός είναι ότι δεν επηρεάζετε ακόμα και αν η εικόνα είναι ολόκληρη ή ελαφρώς επικαλυμμένη από κάποια άλλη. Επίσης σημαντικό είναι ότι τα χαρακτηριστικά είναι μοναδικά για κάθε εικόνα. Καμιά εικόνα δεν έχει τα ίδια χαρακτηριστικά με κάποια άλλη. Έτσι εξασφαλίζετε ότι τα δεδομένα είναι μοναδικά και χαρακτηρίζουν την εικόνα αυτή.

Τέλος, το κομμάτι που δίνει το τελικό αποτέλεσμα είναι ο αλγόριθμος των τεχνητών νευρωνικών δικτύων. Αυτός ο αλγόριθμος προέρχεται από την θεωρία της τεχνητής νοημοσύνης. Με αυτόν τον αλγόριθμο προσομοιώνεται μια λειτουργία του ανθρώπινου εγκεφάλου. Το αποτέλεσμα του αλγορίθμου είναι ένα δίκτυο από νευρώνες που περιέχουν μέσα τους μαθηματικά μοντέλα και είναι ικανά για την εκτέλεση πράξεων. Τα τεχνητά νευρωνικά δίκτυα έχουν δυο λειτουργίες. Η πρώτη είναι αυτής της αρχικοποίησης και η δεύτερη είναι αυτή της εκτέλεσης. Η λειτουργία της εκπαίδευσης είναι αυτή που δημιουργεί το τεχνητό νευρωνικό δίκτυο και δημιουργεί τους δεσμούς ανάμεσα σε είσοδο και έξοδο. Ενώ η λειτουργία της εκτέλεσης επεξεργάζεται τα δεδομένα σύμφωνα με τα χαρακτηριστικά στα οποία εκπαιδεύτηκε. Τα αποτελέσματα του τεχνητού νευρωνικού δικτύου αποτελούν τα αποτελέσματα του συστήματος που περιγράφετε σε αυτήν την εργασία.

## 2.4.1 Υπολογισμός οπτικής αντιστοιχίας μέσω τομών σε γράφους (Graph cuts)

Στην θεωρία των γράφων, η τομή είναι ένα τμήμα από κορυφές ενός γράφου ανάμεσα σε δυο υποσύνολα. Η συνολική – τομή μιας τομής είναι το σύνολο των άκμων των οποίων τα τελικά σημεία είναι σε διαφορετικό υποσύνολο του διαμερίσματος. Οι ακμές λέγετε πως “περνούν” από την τομή εάν είναι σε συνολική – τομή. Σε έναν μη - βεβαρημένο μη – κατευθυνόμενο γράφο, το μέγεθος και το βάρος μιας τομής είναι ο αριθμός των άκμων που διέρχονται την τομή. Σε έναν βεβαρημένο γράφο, ορίζετε ο ίδιος όρος από το άθροισμα των βαρών και των ακμών του διέρχονται την τομή.[13]

Η τομή των γράφων εφαρμόζετε στον τομέα της τεχνητής όρασης Η τομή των γράφων μπορεί να χρησιμοποιηθεί για να επιλύσει αποτελεσματικά μια ευρεία ποικιλία από χαμηλού επιπέδου προβλήματα της τεχνητής όρασης (αρχική όραση), όπως η εξομάλυνση της εικόνας, υπολογισμός οπτικής απόκρισης στερεοσκοπικών εικόνων, και πολλά άλλα μπορούν να διατυπωθούν ως προς την ελαχιστοποίηση της ενέργειας. Τέτοια προβλήματα ελαχιστοποίησης ενέργειας μπορούν να μειωθούν σε περιπτώσεις του προβλήματος της μεγίστης ροής σε έναν γράφο (και μέσω του θεωρήματος της μεγίστης ροής με μικρότερη τομή, να καθοριστεί μια ελάχιστη τομή σε έναν γράφο). Σύμφωνα με τις περισσότερες διατυπώσεις τέτοιων

προβλημάτων στην τεχνητή όραση, η ελαχιστοποίηση της ενέργειας αντιστοιχεί στην μέγιστη εκ των ύστερων εκτίμηση της λύσης. Παρόλα αυτά πολλοί αλγόριθμοι της τεχνητής όρασης εμπλέκουν μια τομή ενός γράφου (πχ οι εξομαλυνμένες τομές). Ο όρος “τομή γράφων” εφαρμόζετε ειδικά σε μοντέλα που χρησιμοποιούν μια βελτιστοποίηση τύπου της μέγιστης ροής με μικρότερη τομή ( άλλοι αλγόριθμοι τόμων γράφων μπορεί να θεωρηθούν σαν διαμερισμένοι αλγόριθμοι γράφων). [7]

“Δυαδικά” προβλήματα (όπως η αποθορυβοποίηση μιας δυαδικής εικόνας) μπορούν αν λυθούν χρησιμοποιώντας ακριβώς αυτήν την προσέγγιση. Προβλήματα όπου τα εικονοστοιχεία πρέπει να κατηγοριοποιηθούν σε πάνω από δυο κατηγορίες (όπως η στερεοσκοπική οπτική απόκριση, είτε την αποθορυβοποίηση μιας ασπρόμαυρης εικόνας) δεν μπορούν να λυθούν ακριβώς., αλλά λύσεις που δημιουργούνται είναι συνήθως πολύ κοντά στην βέλτιστη λύση.[7]

Ο Greig ήταν ο πρώτος που ανακάλυψε ότι οι δυνατοί αλγόριθμοι της μέγιστης ροής με μικρότερη τομή μπορούν να χρησιμοποιηθούν για να ελαχιστοποιήσουν συγκεκριμένες λειτουργίες στην τεχνητή όραση. Οι ενέργειες που κλήθηκαν αργότερα από τον Greig για τις μεθόδους βασισμένες στους γράφους μπορούν να αναπαρασταθούν όπως είναι εμφανές στην παρακάτω εικόνα[14].

$$E(L) = \sum_{p \in P} D_p(L_p) + \sum_{(p,q) \in N} V_{p,q}(L_p, L_q),$$

Εικόνα 19: Συνάρτηση ενέργειας τομών σε γράφους από τον Greig

όπου  $L = \{L_p \mid P \square P\}$  είναι η κατηγορία μιας εικόνας  $P$ , το  $D_p$  είναι λειτουργία τιμωρίας δεδομένων, το  $V_{p,q}$  είναι η πιθανότητα αλληλεπίδρασης και  $N$  είναι ένα σύνολο από ζεύγη των γειτονικών εικονοστοιχείων.



Εικόνα 20: Παράδειγμα κατηγοριοποίησης με γράφους.

Παράδειγμα φαίνεται στην επόμενη εικόνα.

Στην εικόνα είναι εμφανές ένα παράδειγμα κατηγοριοποίησης. Η εικόνα (a) είναι ένα σύνολο εικονοστοιχείων στο  $P$  που παρατηρούνται εντάσεις  $I_p$  για κάθε  $P \square P$ . Μια κατηγορία  $L$  όπως φαίνεται στο (b) έχουν ανατεθεί κάποιες κατηγορίες  $L_p \square 0,1,2$  για κάθε  $P \square P$ . Τέτοιες κατηγορίες μπορεί να είναι βάθος σε μια στερεοσκοπική εικόνα, κάποιος δείκτης αντικείμενου, πραγματική τιμή και άλλες ιδιότητες των εικονοστοιχείων. Συνήθως, μέθοδοι στηριζόμενες σε γράφους θεωρούν ότι το σύνολο των εφικτών κατηγοριών είναι άπειρο. Η χοντρή γραμμή στην εικόνα (b) δείχνει ότι η κατηγοριοποίηση έχει διακοπεί από γειτονικά εικονοστοιχεία.

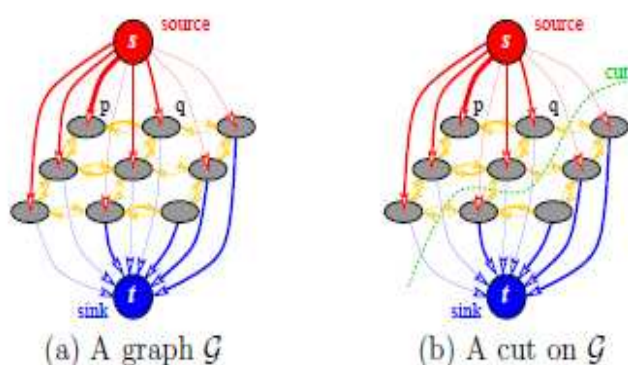
Τυπικά, οι κυρώσεις στα δεδομένα  $D_p$  δείχνουν μεμονωμένες προτιμήσεις κατηγορίας των εικονοστοιχείων που βασίζετε σε παρατηρούμενες εντάσεις και συναρτήσεις προκαθορισμένης πιθανότητας. Οι δυνατότητες αλληλεπίδρασης  $V_{p,q}$  ενθαρρύνουν την χωρική συνοχή με το να τιμωρούν ασυνέχειες ανάμεσα σε γειτονικά εικονοστοιχεία. Βασισμένες σε γράφους μέθοδοι ελαχιστοποίησης της ενέργειας παρέχουν αναμφισβήτητα κάποιες από τις ακριβείς λύσεις για συγκεκριμένες εφαρμογές[14].

Δυστυχώς, η τεχνική του Greig στις τομές των γράφων έμεινε απαρατήρητη για περίπου 10 χρόνια κυρίως επειδή η αποκατάσταση διαδίκων εικόνων φαινόταν τόσο περιορισμένη εφαρμογή. Οι πρώτες προσπάθειες για την χρήση συνδυαστικών αλγορίθμων τομών των γράφων στην τεχνητή όραση ήταν περιορισμένες στην ομαδοποίηση των εικόνων. Στα τέλη του 90' ένας μεγάλος αριθμός από νέες τεχνικές τεχνητής όρασης



εμφανίστηκαν και χρησιμοποίησαν την μέθοδο της μέγιστης ροής με μικρότερη τομή σε γράφους για την επίλυση μη δυαδικών προβλημάτων. Αργότερα, προέκυψε ότι με το σωστό βάρος της ακμής σε γράφο μπορεί κανείς να ελαχιστοποιήσει μια γενική ενεργειακή λειτουργία. Σε μια ακόμα γενική περίπτωση ενεργειών πολλαπλών πηγών η ποινή της αλληλεπίδρασης είναι μετρήσιμη. Ο αλγόριθμος α- επέκτασης πιθανός βρίσκει αρκετά καλά αποτελέσματα εκτελώντας επαναληπτικά τον αλγόριθμο μέγιστης ροής με μικρότερη τομή σε κατάλληλους γράφους[14].

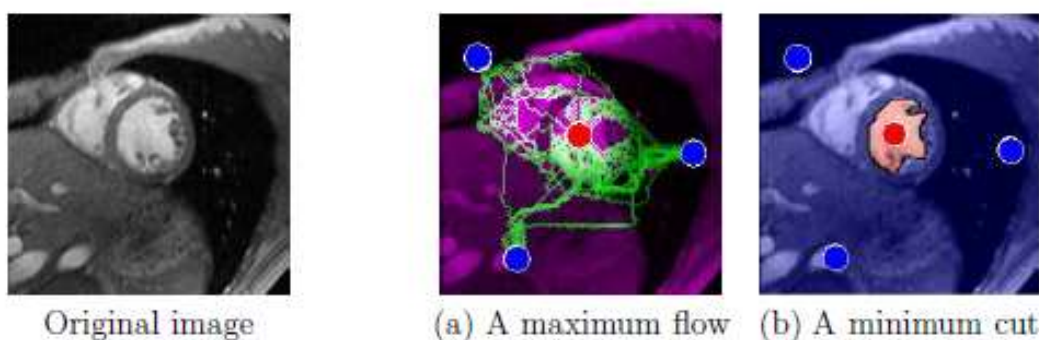
Σε ένα δίκτυο ροής, τομή s-t είναι η τομή που απαιτεί μια πηγή και το βαθύτερο σημείο να είναι σε διαφορετικά υποσύνολα. Στην συνδυαστική βελτιστοποίηση το κόστος μια τομής  $C = \{S, T\}$ , ορίζετε ότι το συνολικό κόστος των “οριακών” ακμών (p,q) όπου  $p \in S$  και  $q \in T$ . Πρέπει να σημειωθεί ότι το κόστος της τομής είναι “κατευθυνόμενο” καθώς συνοψίζει τα βάρη των κατευθυνόμενων ακμών ειδικά από το S στο T. Στο πρόβλημα της ελάχιστης τομής σε έναν γράφο είναι να βρεθεί μια τομή με το ελάχιστο κόστος ανάμεσα σε όλες τις τομές.



Εικόνα 21: Παράδειγμα ελάχιστης τομής σε κατευθυνόμενο γράφο

Στην παραπάνω εικόνα είναι εμφανείς ένα παράδειγμα ένα κατευθυνόμενο γράφο όπου το κόστος των τομών φαίνεται από το πάχος των γραμμών. Ένας παρόμοιος με αυτόν γράφος χρησιμοποιήθηκε πρώτα στην τεχνητή όραση από τον Greig για αποκατάσταση δυαδικών εικόνων.

Ένα από τα θεμελιώδη αποτελέσματα στην συνδυαστική βελτιστοποίηση είναι ότι το πρόβλημα μιας τομής s-t μπορεί να λυθεί βρίσκοντας την μέγιστη ροή από την πηγή s στο βαθύτερο σημείο t. Μιλώντας μεταφορικά μπορούμε να πούμε ότι μέγιστη ροή είναι η μέγιστη “ποσότητα νερού” που μπορεί να κυλίσει από την πηγή στο βαθύτερο σημείο με το να θεωρούμε της ακμές του γράφου ως κατευθυνόμενους “αγωγούς” με χωρητικότητες ίσες με το βάρος των άκμων. Το θεώρημα Ford και Fulkerson [15] δηλώνει ότι μια μέγιστη ροή από το s στο t γεμίζει ένα σύνολο άκμων σε ένα γράφο χωρίζοντας τους κόμβους σε δυο ασυνεχή μέρη {S,T} που αντιστοιχούν σε μια ελάχιστη τομή. Απέδειξαν έτσι ότι τα προβλήματα ελάχιστης τομής και μέγιστης ροής είναι ισοδύναμα. Στην πραγματικότητα, η τιμή της μέγιστης ροής είναι ίση με το κόστος της ελάχιστης τομής. Η διπλή σχέση μεταξύ των προβλημάτων μέγιστης ροής και ελάχιστης τομής είναι εμφανής στην επόμενη εικόνα, όπου στο πλαίσιο της εικόνας (a) έχουμε κατάτμηση της εικόνας από την εμφανιζόμενη μέγιστη ροή και το γέμισμα των “συνοριακών” ακμών με ελάχιστη τομή (b).



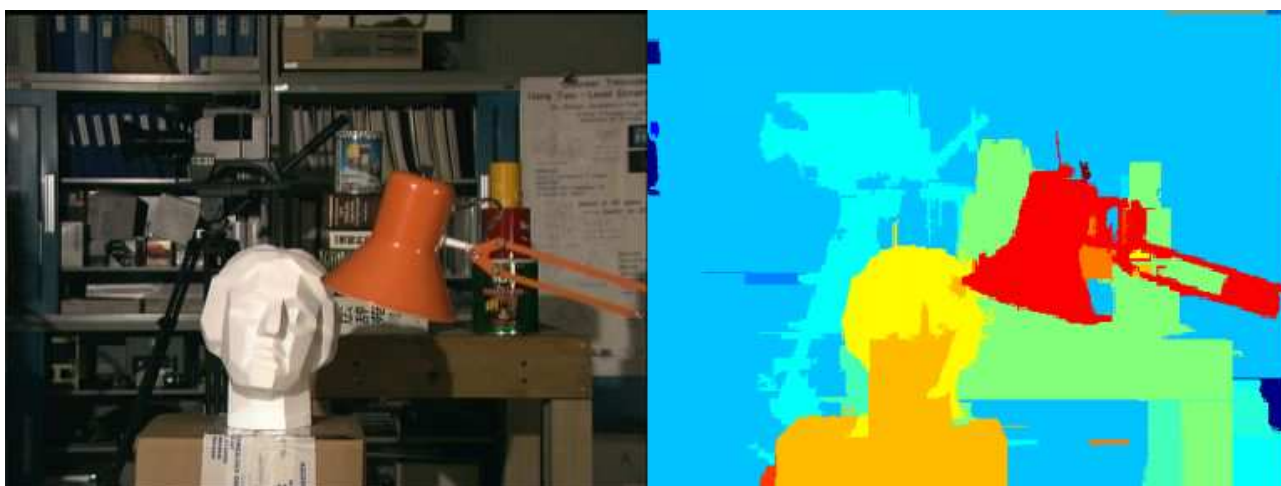
Εικόνα 22: Σχέση μεταξύ των προβλημάτων μεγίστης ροής και ελάχιστης τομής σε παράδειγμα κατάτμησης της εικόνας

Τα κόκκινα αντιστοιχούν στην πηγή  $s$  και μπλε σημεία στα βαθύτερα σημεία  $t$ . Ως συνήθως, τα κόστη των ακμών μεταξύ των εικονοστοιχείων (κόμβων του γράφου) έχουν οριστεί σε χαμηλές τιμές στα σημεία με υψηλή αντίθεση. Έτσι οι τομές στα όρια των αντικειμένων πρέπει να είναι φθηνότερες. Οι αδύναμες ακμές επίσης δουλεύουν ως “σημεία υψηλής συμφόρησης” σε μια ροή. Στην εικόνα (a) φαίνεται η μέγιστη ροή από το  $s$  στο  $t$ . Στην πραγματικότητα, γεμίζει τις ακμές των γράφων που αντιστοιχούν στα όρια μιας ελάχιστης ροής όπως φαίνεται στην (b) [14].

Με την βοήθεια της εικόνας 19 μπορούμε να δείξουμε πως μια ελάχιστη τομή ή μέγιστη ροή σε έναν γράφο μπορεί να λύσει το πρόβλημα της ελαχιστοποίησης της ενέργειας στην ανάθεση ετικετών μιας εικόνας. Αν αυτός ο γράφος είναι μια εικόνα τρία επί τρία κάθε τομή από το  $s$  στο  $t$  χωρίζει τους κόμβους σε ξεχωριστές ομάδες η οποίες περιέχουν ακριβώς ένα τεματικό. Για αυτόν τον λόγο, κάθε τομή αντιστοιχεί σε κάποιες αναθέσεις εικονοστοιχείων (κόμβων) σε ετικέτες (τεματικά). Αν τα βάρη των τομών είναι καταλλήλως τοποθετημένα βασισμένα σε παραμέτρους μιας ενεργείας, η τομή του ελαχίστου κόστους θα αντιστοιχεί σε μια ετικέτα με την ελάχιστη τιμή αυτής της ενεργείας[14].

Στον κλασικό αλγόριθμο υπολογισμού οπτικής αντιστοιχίας ανάμεσα σε στερεοσκοπικό ζεύγος εικόνων, ο υπολογιστής συγκρίνει τις δυο εικόνες καθώς μετατοπίζει την μια πάνω από την άλλη. Έτσι βρίσκει τα τμήματα που ταιριάζουν αλλά και της διαφορές τους [6]. Το πόσο είναι η μετατόπιση των δυο εικόνων ονομάζετε διαφορά. Για παράδειγμα, ένα σημείο που εμφανίζετε στην συντεταγμένη  $(x,t)$  ,μετρημένη σε εικονοστοιχεία, στην αριστερή εικόνα μπορεί να βρίσκετε στις συντεταγμένες  $(x,t-3)$  της δεξιάς εικόνας. Σε αυτήν την περίπτωση η διαφορά είναι 3 εικονοστοιχεία [16]. Όμως αυτή η τεχνική μπορεί να εμφανίσει κρυφά σημεία. Τα κρυφά σημεία είναι τα σημεία αυτά που δεν είναι ορατά και στις δυο εικόνες. Αυτές οι περιοχές είναι σχετικά μικρές αλλά προκαλούν παρεμβολές που δεν μπορούν να ελέγξουν όλοι οι αλγόριθμοι.

Αρχικά στο σύστημα χρησιμοποιήθηκε ο απλός αλγόριθμος του οποίου τα αποτελέσματα είναι ορατά παρακάτω(εικόνα 23). Ο αλγόριθμος αυτός βρέθηκε έτοιμος αναρτημένος στα εργαλεία του Matlab και παράγει έναν πίνακα διαφορών από δυο εικόνες [17]. Τα πλεονεκτήματα αυτού του αλγορίθμου ήταν η απλότητα του και η ταχύτητα της επεξεργασίας του. Θα μπορούσε να χρησιμοποιηθεί για την ανάλυση ενός στερεοσκοπικού βίντεο. Όμως είχε μικρή λεπτομέρεια και ανοχή στην ενέργεια που παράγονται από τα κρυφά σημεία.



Εικόνα 23: Πίνακας Διαφοράς από κλασσικό αλγόριθμο

Έτσι για την επίλυση των παραπάνω προβλημάτων χρησιμοποιήθηκε ο αλγόριθμος match v3.1 του Vladimir Kolmogorov [18]. Έστω,  $L$  είναι το σύνολο των εικονοστοιχείων της αριστερής εικόνας,  $R$  το σύνολο των εικονοστοιχείων της δεξιάς εικόνας, και  $P$  το σύνολο των εικονοστοιχείων όπου:  $P = L \square R$ . Το εικονοστοιχείο  $p$  θα έχει συντεταγμένες  $(p_x, p_y)$ . Στην κλασσική προσέγγιση του υπολογισμού οπτικής αντιστοιχίας ανάμεσα σε στερεοσκοπικό ζεύγος εικόνων, ο σκοπός είναι να υπολογίσει για κάθε εικονοστοιχείο της αριστερής εικόνας μια ετικέτα  $f_p$  που να δωρίζει μια τιμή ανισότητας για ένα εικονοστοιχείο  $p$ . Η ενέργεια που ελαχιστοποιείτε στο μοντέλο του Pott [19] είναι:

$$E(f) = \sum_{p \in \mathcal{L}} D_p(f_p) + \sum_{p, q \in \mathcal{N}} V_{p, q} \cdot T(f_p \neq f_q).$$

Εικόνα 24: Η ενέργεια που ελαχιστοποιείτε στο μοντέλο του Pott [19]

Το μοντέλο του Pott στηρίζετε σε αυτό του Greig όπως περιγράφηκε παραπάνω. Εδώ το  $D_p(f_p)$  είναι τιμωρία για κάθε εικονοστοιχείο το οποίο έχει ανισότητα  $f_p$ , το  $\mathcal{N}$  είναι ένα σύστημα γειτονιών του κάθε φωτοστοιχείου της αριστερής εικόνας και  $T(\square)$  έχει την τιμή ένα εάν το επιχείρημα είναι αληθές αλλιώς η τιμή του είναι μηδέν. Ας θεωρήσουμε ότι μια συγκεκριμένη ανισότητα  $a$  (ή ετικέτα  $a$ ). Μια διαμόρφωση  $f'$  λέγετε ότι βρίσκεται σε μια κίνηση  $a$  - επέκτασης του  $f$  εάν για όλα τα εικονοστοιχεία  $p \square L$  είτε  $f'_p = f_p$  ή  $f'_p = a$ . Τώρα αν θεωρήσουμε ότι ένα ζεύγος ανισοτήτων  $a, \beta, a \neq \beta$ . Μια διαμόρφωση  $f'$  λέγετε ότι βρίσκεται σε μια κίνηση  $a\beta$  - ανταλλαγής του  $f$  εάν για όλα τα εικονοστοιχεία  $p \square L$ , με  $f_p \square \{a, \beta\}$  να συνεπάγεται  $f'_p = f_p$ . Το σημαντικό γεγονός είναι ότι για την συγκεκριμένες διαμορφώσεις του  $f$  είναι δυνατό να βρεθεί αποτελεσματικά ένα τοπικό ελάχιστο τις ενέργειας με ακρίβεια. Ποιο συγκεκριμένα, είναι η ελάχιστη ενέργεια διαμόρφωσης ανάμεσα σε μια απλή κίνηση  $a$  - επέκτασης ή  $a\beta$  - ανταλλαγής του  $f$  αντίστοιχα. Αυτές οι τοπικές εργασίες βελτίωσης στηρίζονται πάνω στην θεωρία των γράφων και στις τομές τους. Ο αλγόριθμος επέκτασης αποτελείτε αποκλειστικά από μια ακολουθία  $a$  - επέκτασης για τοπικά βελτιωμένες διαδικασίες για διαφορετικές ανισότητες  $a$ , μέχρις ότου καμιά  $a$  - επέκταση να μην μπορεί να μειώσει την ενέργεια. Ομοίως, ο αλγόριθμος ανταλλαγής αποτελείτε αποκλειστικά από μια ακολουθία  $a\beta$  - ανταλλαγής που βελτιώνει τοπικά της πράξεις μεταξύ ζευγών ανισοτήτων  $a, \beta$  μέχρις ότου η  $a\beta$  - ανταλλαγή να μην μπορεί να μειώσει την ενέργεια [18].

Αυτή όμως η διατύπωση δεν μπορεί δυστυχώς να αντιμετωπισή τελείως τις ανισότητες. Αρχικά, μπορεί εύκολα να συμβεί σε δυο εικονοστοιχεία τις αριστερής εικόνας να χαρτογραφηθούν στο ίδιο εικονοστοιχείο της δεξιάς εικόνας. Επιπλέον, θεωρεί ότι κάθε εικονοστοιχείο τις αριστερής εικόνας χαρτογραφείτε σε κάποιο εικονοστοιχείο της δεξιάς εικόνας καθώς στην πραγματικότητα κάποια από τα εικονοστοιχεία της

αριστερής εικόνας μπορούν να μην έχουν αντίκρισμα σε κάποιο εικονοστοιχείο της δεξιάς εικόνας[18]. Έστω  $A$  είναι το σύνολο των ζευγών των εικονοστοιχείων (χωρίς διάταξη) που μπορούν πιθανός να έχουν αντιστοιχία. Για ένα στερεοσκοπικό ζεύγος εικόνων από ευθυγραμμισμένες κάμερες, για παράδειγμα, έχουμε  $A = \{ p, q \mid py = qy \text{ και } 0 \leq qx - px < k \}$ . Υποθέτουμε ότι οι ανισότητες βρίσκονται σε κάποιο περιορισμένο εύρος, έτσι ένα εικονοστοιχείο  $L$  μπορεί να ανταποκρίνεται μόνο σε ένα  $k$  πιθανό εικονοστοιχείο στο  $R$  και αντίθετα. Η κατάσταση για την κίνηση είναι παρόμοια, εκτός από το σύνολο των πιθανών ανισοτήτων στις δυο διαστάσεις. Ο σκοπός είναι να βρεθεί υποσύνολο  $A$  το οποίο να περιέχει μόνο ζεύγη εικονοστοιχείων τα οποία να αντιστοιχούν το ένα στο άλλο. Ισοδύναμα, μπορούμε να δώσουμε σε κάθε ανάθεση  $a \in A$  μια τιμή  $f_a$  που είναι ένα εάν τα εικονοστοιχεία  $p$  και  $q$  αντιστοιχούν το ένα στο άλλο αλλιώς η τιμή θα είναι μηδέν. Ας καθοριστούν μοναδικές διαμορφώσεις  $f$ . Θα ονομάσουμε τις αναθέσεις στο  $A$  που έχουν την τιμή ένα έως ενεργές. Έστω λοιπόν ότι  $A(f)$  είναι το σύνολο των ενεργών αναθέσεων σύμφωνα με την διαμόρφωση  $f$ . Έστω ότι  $Np(f)$  είναι το σύνολο των αναθέσεων στην  $f$  που εμπλέκουν το εικονοστοιχείο  $p$ ,  $\pi_x Np(f) = \{ p, q \in A(f) \}$ . Μια διαμόρφωση  $f$  είναι μοναδική εάν το κάθε εικονοστοιχείο εμπλέκεται το πολύ σε μια ενεργή ανάθεση εργασίας,  $\pi_x \cap \pi_y \cap P / DNS(\eta) \leq 1$ . Να σημειωθεί ότι αυτά τα εικονοστοιχεία για τα οποία  $|Np(f)| = 0$  είναι ακριβώς τα κρυφά εικονοστοιχεία [18].

Είναι δυνατό να επεκταθεί η έννοια των  $a$  – επεκτάσεων σε αυτήν την αναπαράσταση. Για ανάθεση εργασίας  $a = \langle p, q \rangle$  και το  $d(a)$  είναι η ανισότητα τότε :  $d(a) = (qx - px, qy - py)$ , και έστω  $A^a$  είναι το σύνολο όλων των εργασιών στο  $A$  με οπτική ανισότητα  $a$ . Μια διαμόρφωση  $f$  βρίσκεται μέσα σε μια κίνηση  $a$  – επέκτασης του  $f$  εάν το  $A(f)$  είναι ένα υποσύνολο του  $A(f) \cap A^a$ . Με άλλα λόγια, κάποιες ενεργές αναθέσεις μπορεί να διαγράφουν, και κάποιες αναθέσεις που έχουν οπτική ανισότητα  $a$  μπορεί να προστεθούν. Επίσης είναι δυνατό να επεκταθεί η έννοια της  $\alpha\beta$  – ανταλλαγής. Μια διαμόρφωση  $f$  λέγεται ότι βρίσκεται σε μια κίνηση  $\alpha\beta$  – ανταλλαγής του  $\eta$  εάν  $A(f) \cap A^{\alpha\beta} = A(f) \cap A^{\beta\alpha}$ , όπου  $A^{\alpha\beta}$  είναι το σύνολο όλων των αναθέσεων στο  $A$  με οπτική ανισότητα  $\alpha$  ή  $\beta$ . Με άλλα λόγια, μόνο οι αλλαγές στο  $f$  μπορούν να προσθέσουν ή να διαγράψουν αναθέσεις σε οπτική ανισότητα  $\alpha$  ή  $\beta$  [18]. κρυφά ακριβώς είναι οποία αυτά ότι σημειωθεί

Τώρα πρέπει να ορίσουμε για μια διαμόρφωση  $f$  την ενέργεια. Για τον επιτυχή χειρισμό των μοναδικών διαμορφώσεων μπορεί να θεωρηθεί ότι για μια μη – μοναδική διαμόρφωση η ενέργεια είναι άπειρη και για μια μοναδική διαμόρφωση η ενέργεια είναι της μορφής:

$$E(f) = E_{data}(f) + E_{occ}(f) + E_{smooth}(f). \quad (2)$$

Εικόνα 25: Ενέργεια μοναδικής διαμόρφωσης  $f$

Οι τρεις όροι που περιλαμβάνει είναι ο όρος των δεδομένων  $E_{data}$ , που είναι τα αποτελέσματα των διάφορων στην ένταση μεταξύ αντίστοιχων εικονοστοιχείων. Ένας όρος κρυφών σημείων  $E_{occ}$ , ο οποίος αναπαριστά μια τιμωρία για την δημιουργία ενός εικονοστοιχείου σε κρυφό σημείο. Και ένας όρος ομαλότητας  $E_{smooth}$ , ο οποίος κάνει τα γειτονικά εικονοστοιχεία στην ίδια εικόνα να έχουν την τάση να έχουν κοινές οπτικές ανισότητες. Ο όρος δεδομένων θα είναι  $E_{data}(f) = \sum_{a \in A(f)} D(a)$ , τυπικά για κάθε ανάθεση  $a = \langle p, q \rangle$ ,  $D(a) = (I(p) - I(q))^2$ , όπου  $I$  δίνει την ένταση ενός εικονοστοιχείου [18]. Ο όρος των κρυφών σημείων επιβάλλει μια τιμωρία  $C_p$  για κάθε εικονοστοιχείο  $p$  που είναι κρυφό σημείο θα γραφτεί σαν:

$$E_{occ}(f) = \sum_{p \in P} C_p \cdot T(|N_p(f)| = 0).$$

Εικόνα 26: Τύπος όρου κρυφών σημείων  $E_{occ}$

Το ποιο μη τετριμμένο κομμάτι εδώ είναι η επιλογή όρου ομαλότητας. Είναι δυνατόν να γράφουν αρκετές εκφράσεις του όρου ομαλότητας. Ο όρος ομαλότητας περιλαμβάνει την έννοια μιας γειτονιάς και θεωρούμε ότι υπάρχει σύστημα γειτονιών σε αναθέσεις όπου  $N \ni \{a1, a2 \mid a1, a2 \in A\}$ . Έτσι η προφανή επιλογή είναι :

$$E_{smooth}(f) = \sum_{\{a1,a2\} \in N, a1,a2 \in A(f)} V_{a1,a2}, \quad (3)$$

Εικόνα 27: Απλός όρος ομαλότητας *Esmooth*

Εκεί όπου το σύστημα της γειτονιάς  $N$  αποτελείται μόνο από ζευγάρια  $\{a1, a2\}$  τέτοια ώστε αναθέσεις  $a1$  και  $a2$  να έχουν διαφορετικές οπτικές ανισότητες. Το  $N$  μπορεί να περιεχί, για παράδειγμα, ζεύγη από αναθέσεις  $\{<p, q>, <p', q'>\}$  για τα οποία είτε  $p$  και  $p'$  είναι γείτονες είτε  $q$  και  $q'$  είναι γείτονες, και  $d(p, q) \neq d(p', q')$ . Έτσι επιβάλετε τιμωρία εάν δυο κοντινές αναθέσεις που έχουν διαφορετικές οπτικές ανισότητες είναι και οι δυο παρούσες στην διαμόρφωση [18].

Τα ποιο ελπιδοφόρα αποτελέσματα λαμβάνονται χρησιμοποιώντας ένα διαφορετικό όρο ομαλότητας, ο οποίος κάνει δυνατή την χρήση τομών γράφου. Για την αποδοτική εύρεση ένα ελάχιστο της ενέργειας ανάμεσα σε όλες τις διαμορφώσεις μέσα σε μια μόνο  $\alpha$  – επέκταση της αρχικής διαμόρφωσης. [18] Ο όρος ομαλότητας είναι ο εξής:

$$E_{smooth}(f) = \sum_{\{a1,a2\} \in N} V_{a1,a2} \cdot T(f(a1) \neq f(a2)).$$

Εικόνα 28: Εξελιγμένος όρος ομαλότητας *Esmooth*

Το σύστημα γειτονιών αποτελείται από μόνο ζευγάρια  $\{a1, a2\}$  τέτοια ώστε αναθέσεις  $a1$  και  $a2$  να έχουν τις ίδιες οπτικές ανισότητες. Μπορεί να περιεχί, για παράδειγμα, ζεύγη από αναθέσεις  $\{<p, q>, <p', q'>\}$  για τα οποία είτε  $p$  και  $p'$  είναι γείτονες είτε  $q$  και  $q'$  είναι γείτονες, και  $d(p, q) = d(p', q')$ . Έτσι επιβάλετε τιμωρία εάν μια ανάθεση είναι παρούσα στην διαμόρφωση, και σε μια κλειστή ανάθεση, έχοντας την ίδια οπτική ανισότητα δεν είναι παρούσα. Παρόλα αυτά η ενεργεία αυτή είναι διαφορετική από την προηγούμενη και επιβάλλει τον ίδιο περιορισμό: εάν οι οπτικές ανισότητες των παρακείμενων εικονοστοιχείων είναι οι ίδιες τότε η τιμωρία της ομαλότητας είναι μηδενική, διαφορετικά έχει κάποια θετική τιμή. Ο λόγος για τον οποίο αυτή η ενεργεία επιτρέπει την χρήση τόμων των γραφών είναι γιατί έχει παρόμοια μορφή με αυτήν της ενέργειας στο μοντέλο του Pott [19]. Ωστόσο, είναι η ενέργεια του μοντέλου του Pott στις αναθέσεις πάρα στα εικονοστοιχεία και σαν συνέπεια, κανένας από τους προηγούμενους αλγόριθμους πάνω στις τομές των γραφών δεν μπορεί να εφαρμοστεί [18].

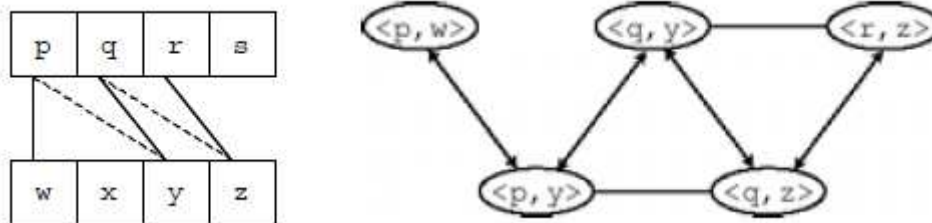
Λόγο της καλύτερης αποδοτικότητας που παρέχει το σύστημα χρησιμοποιεί τον αλγόριθμο επέκτασης, έναντι του αλγόριθμου ανταλλαγής. Το αποτέλεσμα της μεθόδου που χρησιμοποιείτε είναι ένα τοπικό ελάχιστο με έντονη αίσθηση. Συγκεκριμένα, αν θεωρήσουμε ως είσοδο μια διαμόρφωση  $f$  και οπτική ανισότητα  $\alpha$ . Σε μια άλλη διαμόρφωση  $f'$  ορίζετε να είναι μια  $\alpha$  – επέκταση της  $f$  εάν κάποιες αναθέσεις στην  $f$  είναι ανενεργές και κάποιες αναθέσεις με οπτική ανισότητα  $\alpha$  να γίνουν ενεργές [18].

1. Start with an arbitrary unique configuration  $f$
2. Set success := 0
3. For each disparity  $\alpha$ 
  - 3.1. Find  $\hat{f} = \operatorname{argmin} E(f')$  among unique  $f'$  within single  $\alpha$ -expansion of  $f$
  - 3.2. If  $E(\hat{f}) < E(f)$ , set  $f := \hat{f}$  and success := 1
4. If success = 1 goto 2
5. Return  $f$

Εικόνα 29: Βήματα του αλγορίθμου επέκτασης

Στον αλγόριθμο που χρησιμοποιείται (εικόνα 29) απλά επιλέγεται, με συγκεκριμένη σειρά ή τυχαία, μια οπτική ανισότητα  $\alpha$ , και βρίσκει την μοναδική διαμόρφωση μέσα σε μια κίνηση  $\alpha$  – επέκτασης. Εάν αυτό ελαττώσει την ενεργεία, το αποτέλεσμα είναι έτοιμο. Το σημαντικό βήμα σε αυτήν την μέθοδο είναι να υπολογιστεί αποδοτικά η  $\alpha$  – επέκταση με την ελάχιστη ενεργεία [18].

Με την βοήθεια των τομών των γραφών αυτό το πρόβλημα μπορεί να λυθεί. Έστω  $G = V, E$  είναι ένας βεβαρημένος γραφός με δυο διακεκριμένες τερματικές κορυφές  $\{s, t\}$  που ονομάζονται πηγή και βαθύτερο  $t$ . Μια τομή  $C = V^s, V^t$  είναι *χόρισμα* των δυο κορυφών σε δυο σύνολα τα οποία  $s \in V^s$  και  $t \in V^t$ . Το κόστος της τομής, που συμβολίζεται  $|C|$ , είναι ίσο με το σύνολο των βαρών από τις ακμές ανάμεσα στις κορυφές του  $V^s$  και τις κορυφές του  $V^t$ . Το πρόβλημα της ελάχιστης τομής είναι να βρεθεί η τομή με το ελάχιστο κόστος. Αυτό το πρόβλημα μπορεί να λυθεί πολύ αποδοτικά με τον υπολογισμό της μέγιστης ροής ανάμεσα στις κορυφές, σύμφωνα με το θεώρημα του Ford και Fulkerson [15]. Το χειρότερο δυνατό σενάριο πολυπλοκότητας είναι ένα χαμηλής τάξης πολυώνυμο, όμως στην πραγματικότητα ο χρόνος εκτέλεσης είναι σχεδόν γραμμικός για γραφούς με πολλά μικρά μονοπάτια ανάμεσα στην πηγή και το βαθύτερο σημείο [18].



Εικόνα 30: Μεταποίηση μιας εικόνας σε ένα γράφο

Πρώτα κατασκευάζετε γραφός  $G = [V, E]$ , και δίνετε η αντιστοιχία ανάμεσα στις τομές στον  $G$  και στις διαμορφώσεις. Έπειτα εμφανίζεται ότι η ελάχιστη τομή στο  $G$  που αποδίδει την διαμόρφωση που ελαχιστοποιεί την  $E$  ανάμεσα σε μοναδικές διαμορφώσεις μέσα σε μια  $\alpha$  – επέκταση. Στην  $\alpha$  – επέκταση, οι ενεργές αναθέσεις μπορεί να γίνουν ανενεργές, και οι ανενεργές αναθέσεις στις οποίες η οπτική ανισότητα είναι  $\alpha$  μπορεί να γίνουν ενεργές. Έστω ότι ξεκινάμε με μια μοναδική διαμόρφωση  $f^0$ . Οι ενεργές αναθέσεις για μια νέα διαμόρφωση μέσα σε μια  $\alpha$  – επέκταση θα είναι ένα υποσύνολο  $A^\sim = A^0 \sqcup A^\alpha$ , όπου  $A^0 = \{a \in A(f^0) \mid d(a) \neq \alpha\}$  και  $A^\alpha = \{a \in A \mid d(a) = \alpha\}$ . Θα ορίσουμε ως διαμόρφωση  $f^\sim$  από  $A(f^\sim) = A^\sim$ . Να σημειωθεί ότι το  $f^\sim$  δεν είναι μοναδικό. Ο κατευθυνόμενος γράφος  $G$  ο οποίος αποτελείται από κορυφές που αντιστοιχούν σε αναθέσεις. Τα τερματικά θα ονομάζονται  $s$  και  $t$ , και για κάθε ανάθεση  $A^\sim$  θα υπάρχει μια κορυφή. Οι ακμές στον  $G$  έχουν ως εξής. Για κάθε κορυφή  $a \in A^\sim$  θα υπάρχουν ακμές  $(s, a)$  και  $(a, t)$ . Επιπλέον, εάν  $\{a1, a2\} \in N$  όπου θα υπάρχουν ακμές  $(a1, a2)$  και  $(a2, a1)$ . Να σημειωθεί ότι σε αυτήν την περίπτωση, είτε η  $a1$  και  $a2$  είναι στο  $A^0$  ή είναι και οι δυο στην  $A^\alpha$ . Τελικά, θεωρούμε ότι ένα ζεύγος από κορυφές  $a1, a2$  μπαίνουν σε κοινό εικονοστοιχείο  $p$  (πχ όπου  $a1 = [p, q]$  και  $a2 = [p, r]$ ). Να σημειωθεί επίσης ότι σε αυτήν την περίπτωση είτε  $a1 \in A^0, a2 \in A^\alpha$  ή αντίθετα. Θα υπάρχουν ακμές ανάμεσα σε κάθε

τέτοιο ζεύγος αναθέσεων [18].

Τώρα θεωρούμε ότι μια τομή  $C = V^s, V^t$  στον  $G$ . Η διαμόρφωση  $f^C$  η οποία αντιστοιχεί σε αυτήν την τομή ορίζεται ως:

$$\forall a \in \mathcal{A}^0 \quad f_a^C = \begin{cases} 1 & \text{if } a \in V^s \\ 0 & \text{if } a \in V^t \end{cases}$$

$$\begin{aligned} \forall a \in \mathcal{A}^\alpha \quad f_a^C &= \begin{cases} 1 & \text{if } a \in V^s \\ 0 & \text{if } a \in V^t \end{cases} \\ \forall a \notin \bar{A} \quad f_a^C &= 0 \end{aligned}$$

Η παραπάνω έκφραση δημιουργεί τον συλλογισμό ότι είναι τομή στον γράφο  $G$  αν και μόνο αν διαμόρφωση βρίσκεται μέσα σε μια απλή  $\alpha$  επέκταση της αρχικής διαμόρφωσης  $f^0$ . Τώρα δίνουμε τα βάρη στις ακμές στον γράφο  $G$  [18]. Αρχικά, ορίζεται το κόστος των κρυφών σημείων.

$$D_{occ}((p, q)) = D_{occ}(p) + D_{occ}(q),$$

*Εικόνα 31: Αρχική ανάθεση κόστους κρυφών σημείων*

Όπου  $D_{occ}(p) = C_p$  εάν  $A$  έχει μόνο μια ακμή που εισέρχεται στο  $p$ , και 0 σε οποιαδήποτε άλλη περίπτωση [18]. Το κόστος καθορίζεται από τον τύπο:

$$D_{smooth}(a1) = \sum_{\substack{\{a1, a2\} \in \mathcal{N} \\ a2 \in \bar{A}}} V_{a1, a2}.$$

*Εικόνα 32: Αρχικό κόστος ομαλοποίησης*

και τα βάρη έχουν ως εξής:

edge	weight	for
$(s, a)$	$D_{occ}(a)$	$a \in \mathcal{A}^0$
$(a, t)$	$D_{occ}(a)$	$a \in \mathcal{A}^\alpha$
$(a, t)$	$D(a) + D_{smooth}(a)$	$a \in \mathcal{A}^0$
$(s, a)$	$D(a)$	$a \in \mathcal{A}^\alpha$
$(a1, a2)$ $(a2, a1)$	$V_{a1, a2}$	$\{a1, a2\} \in \mathcal{N},$ $a1, a2 \in \mathcal{A}$
$(a1, a2)$	$\infty$	$p \in \mathcal{P}, a1 \in \mathcal{A}^0, a2 \in \mathcal{A}^\alpha$ $a1, a2 \in \mathcal{N}_p(f)$
$(a2, a1)$	$C_p$	$p \in \mathcal{P}, a1 \in \mathcal{A}^0, a2 \in \mathcal{A}^\alpha$ $a1, a2 \in \mathcal{N}_p(f)$

Εικόνα 33: Πινάκας βαρών για κάθε διαμόρφωση

Γίνετε αναφορά στους συνδέσμους με βάρη  $D_{occ}(a)$  όπως φαίνονται και στις δυο πάνω γραμμές του παραπάνω πίνακα ως *συνδέσμους-t*. Επίσης αναφερόμαστε στους συνδέσμους με κόστη  $C_p$  ως *συνδέσμους-c*. Ένα παράδειγμα σε αυτό μπορεί να αποτελέσει η εικόνα 30 όπου είναι εμφανείς με διακεκομμένες γραμμές νέες αναθέσεις. Στην συγκεκριμένη διαμόρφωση, το εικονοστοιχείο  $s$  και  $x$  είναι κρυφά σημεία, και η προτεινόμενη κίνηση επέκτασης δεν θα αλλάξει την κατάσταση τους. Ο αντίστοιχος γράφος στην ίδια εικόνα οι πάνω σειρά με τους τρεις κόμβους σχηματίζουν  $\mathcal{A}^0$  και οι δυο κάτω κόμβοι σχηματίζουν  $\mathcal{A}^\alpha$ . Έτσι σε αυτό το παράδειγμα οι ακμές από το  $[p, w]$  στο  $[p, y]$  έχουν άπειρο βάρος, αφού και οι δυο αυτές αναθέσεις δεν μπορούν να είναι ενεργές ταυτόχρονα [18].

Για την βελτιστοποίηση του αλγορίθμου εάν το  $C$  είναι μια ελάχιστη τομή σε έναν γράφο  $G$ , τότε  $f^C$  είναι η διαμόρφωση που ελαχιστοποιεί την ενέργεια  $E$  πάνω σε μοναδικές διαμορφώσεις. Θεωρείτε ότι το κόστος της τομής είναι  $C$  πεπερασμένο αν και μόνο αν αντιστοιχεί σε διαμόρφωση  $f^C$  που είναι μοναδική. Η αποδείξει σε αυτό το θεώρημα είναι ότι εάν το  $f^C$  δεν είναι μοναδικό τότε υπάρχει ένα εικονοστοιχείο  $p \in \mathcal{P}$  τέτοιο ώστε το ζεύγος των αναθέσεων  $a1, a2 \in \mathcal{N}_p(f^C)$  να είναι και τα δυο  $\mathcal{A}(f^C)$ . Χωρίς την έλλειψη της γενικότητας έστω  $a1 \in \mathcal{A}^0$  και  $a2 \in \mathcal{A}^\alpha$ . Τότε έχουμε  $a1 \in V^s$  και  $a2 \in V^t$ , έτσι ώστε ακμή  $(a1, a2)$ , που έχει άπειρο βάρος, πρέπει να κοπεί. Παρομοίως, εάν το βάρος της  $C$  είναι άπειρο, μια από τις ακμές είναι τομή, και έτσι ορισμένα εικονοστοιχεία  $p$  δεν είναι μοναδικά [18].

Αν θεωρήσουμε ότι έστω το  $f^C$  είναι μοναδική διαμόρφωση, με αντίστοιχη τομή  $C$ . Τότε το κόστος των των συνδέσμων-t συν του συνδέσμου-c στο  $C$  ισούται με  $E_{occ}(f^C)$  συν μια σταθερά. Η απόδειξη είναι ότι το κόστος των συνδέσμων-t είναι :

$$\sum_{a \in \mathcal{A}^0} D_{occ}(a) \cdot T(a \in \mathcal{V}^t) + \sum_{a \in \mathcal{A}^\alpha} D_{occ}(a) \cdot T(a \in \mathcal{V}^s).$$

Εικόνα 34: Τύπος κόστους συνδέσμων-t

και το κόστος των συνδέσμων-c είναι:



$$\sum_{\substack{p \in \mathcal{P}, a1 \in A^0, a2 \in A^a \\ a1, a2 \in N_p(\bar{f})}} C_p \cdot T(a1 \in \mathcal{V}^t \wedge a2 \in \mathcal{V}^s).$$

Εικόνα 35: Τύπος κόστους συνδέσμων- $c$

Επίσης έχουμε :

$$E_{occ}(f^c) = \sum_{p \in \mathcal{P}} C_p \cdot T(|N_p(f^c)| = 0)$$

Εικόνα 36: Ενέργεια κρυφών σημείων για διαμόρφωση  $f^c$

Η οποία είναι μια σταθερά συν:

$$\sum_{\substack{p \in \mathcal{P} \\ |N_p(\bar{f})|=1}} C_p \cdot T(|N_p(f^c)| = 0) + \sum_{\substack{p \in \mathcal{P} \\ |N_p(\bar{f})|=2}} C_p \cdot T(|N_p(f^c)| = 0).$$

Εικόνα 37: Σύνολο ενέργειας κρυφών σημείων για διαμόρφωση  $f^c$

Μπορούμε να πούμε ότι αυτή είναι μια μεταβλητή συν τους τρεις όρους:

$$\begin{aligned} & \sum_{\substack{p \in \mathcal{P} \\ N_p(\bar{f})=\{a\} \subset A^0}} D_{occ}(p) \cdot T(a \notin A(f^c)) \\ & + \sum_{\substack{p \in \mathcal{P} \\ N_p(\bar{f})=\{a\} \subset A^a}} D_{occ}(p) \cdot T(a \notin A(f^c)) \\ & + \sum_{\substack{p \in \mathcal{P}, a1 \in A^0, a2 \in A^a \\ a1, a2 \in N_p(\bar{f})}} C_p \cdot T(a1, a2 \notin A(f^c)). \end{aligned}$$

Εικόνα 38: Γενικό σύνολο εξισώσεων διαμόρφωσης  $f^c$

Που ισούται με το σύνολο των εξισώσεων στην εικόνα 34 και εικόνα 35.

Τέλος, στο θεώρημα ότι έστω  $C$  είναι μια ελάχιστη τομή σε έναν γράφο  $G$  τότε το  $f^c$  είναι η μοναδική διαμόρφωση μέσα σε μια  $a$ -επέκταση του  $f^0$  που ελαχιστοποιούν την ενέργεια  $E$ . Και η απόδειξη όπως μας δείχνει το πρώτο θεώρημα ότι βρίσκετε  $f^{C_{μίσσα}}$  σε μια  $a$ -επέκταση του  $f^0$ . Το δεύτερο θεώρημα δείχνει ότι η ελάχιστη τομή είναι μοναδική, αφού υπάρχουν προφανείς τομές στον με πεπερασμένα  $G$  κόστη, για αυτόν τον λόγο δεν συμπεριλήφθηκαν στην  $C$  σύνδεσμοι με άπειρα κόστη [18]. Σύμφωνα με το τρίτο θεώρημα αυτό που παραμένει είναι να εμφανιστεί ότι το κόστος στην τομή  $C$ , αγνοώντας στους συνδέσμους- $t$  και συνδέσμους- $c$ , είναι ότι  $E_{data}(f^c) + E_{smooth}(f^c)$ , που είναι :

$$\sum_{a \in A(f^c)} D(a) + \sum_{\{a1, a2\} \in \mathcal{N}} V_{a1, a2} \cdot T(f_{a1}^c \neq f_{a2}^c).$$

Εικόνα 39: Τύπος ενεργείας για  $a$  - έκταση του  $f^c$

Το δεύτερο σύνολο θα ξαναγράψει ως:

$$\sum_{\substack{\{a1, a2\} \in \mathcal{N} \\ a1, a2 \in \bar{A}}} V_{a1, a2} \cdot T(f_{a1}^c \neq f_{a2}^c) + \sum_{\substack{\{a1, a2\} \in \mathcal{N} \\ a1 \in \bar{A}, a2 \notin \bar{A}}} V_{a1, a2} \cdot T(f_{a1}^c \neq f_{a2}^c).$$

Εικόνα 40: Δεύτερο μέρος τύπου ενέργειας για  $a$  - επέκταση του  $f^c$

Αγνοώντας τους συνδέσμων- $t$  και του συνδέσμου-  $c$  , τα κόστη μπορούν να γράφουν ως εξής:

$$\begin{aligned} & \sum_{a \in A^0} (D(a) + D_{smooth}(a)) \cdot T(a \in \mathcal{V}^s) \\ & + \sum_{a \in A^a} D(a) \cdot T(a \in \mathcal{V}^t) \\ & + \sum_{\substack{\{a1, a2\} \in \mathcal{N} \\ a1, a2 \in \bar{A}}} V_{a1, a2} \cdot T((a1 \in \mathcal{V}^s, a2 \in \mathcal{V}^t) \vee (a1 \in \mathcal{V}^t, a2 \in \mathcal{V}^s)). \end{aligned}$$

Εικόνα 41: Κόστη που προκύπτουν από τύπο ενέργειας για  $a$  - επέκταση του  $f^c$

Οι δυο πρώτοι οροί είναι:

$$\sum_{a \in A^0 \cap A(f^c)} (D(a) + D_{smooth}(a)) + \sum_{a \in A^a \cap A(f^c)} D(a),$$

Εικόνα 42: Ανάπτυξη μέρους πρώτου εικόνα 41

καθώς και ο τρίτος όρος είναι:

$$\sum_{\substack{\{a1, a2\} \in \mathcal{N} \\ a1, a2 \in \bar{A}}} V_{a1, a2} \cdot T(f_{a1}^c \neq f_{a2}^c).$$

Εικόνα 43: Ανάπτυξη μέρους δεύτερου εικόνας 41

Οι όροι που αφορούν το σύνολο  $D(a)$  στο  $\sum_{a \in A(f^c)} D(a)$ , έτσι αυτό που αρκεί είναι:

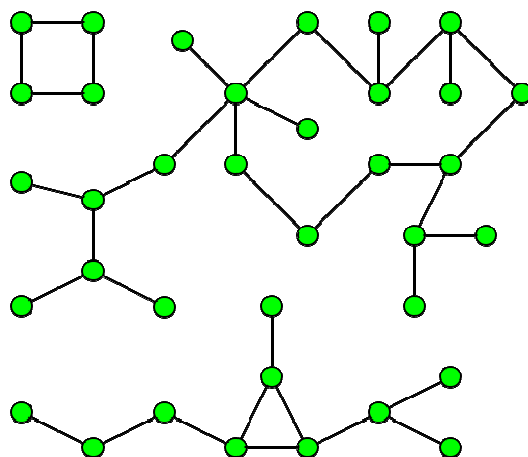
$$\sum_{\substack{\{a1, a2\} \in \mathcal{N} \\ a1 \in \bar{A}, a2 \notin \bar{A}}} V_{a1, a2} \cdot T(f_{a1}^c \neq f_{a2}^c) = \sum_{a \in A^0 \cap A(f^c)} D_{smooth}(a).$$

Εικόνα 44: Τελική απόδειξη θεωρήματος

Στην πρώτη έκφραση,  $a1 \in A^0$ , εφόσον  $a1 \in Aa$  και  $\{a1, a2\} \in N$  συνεπάγεται  $a2 \in Aa \in A^*$ . Η απόδειξη φαίνεται από τον ορισμό του *Dsmooth* [18]. Έτσι πλέον αποδεικνύεται και η ορθή λειτουργία όλου του αλγορίθμου και τεκμηριώνετε η χρήση του στο σύστημα ως αλγορίθμου με εξαιρετικά αποτελέσματα.

## 2.4.2 Αλγόριθμος συνδεδεμένων στοιχείων

Στην θεωρία των γραφών, η σύνδεση στοιχείων σε έναν μη – κατευθυνόμενο γράφο είναι ένας υπό – γραφος στον οποιοδήποτε δυο κορυφές είναι συνδεδεμένες η μια με την άλλη από ένα μονοπάτι, το οποίο δεν είναι συνδεδεμένο με άλλες κορυφές. Όπως φαίνεται και στην εικόνα 45 έχουμε έναν γράφο με τρία συνδεδεμένα στοιχειά. Ένας γράφος ο οποίος είναι από μόνος του συνδεδεμένος έχει ακριβώς ένα συνδεδεμένο στοιχείο, που αποτελεί όλων τον γράφο[8].



Εικόνα 45: γράφος με τρία συνδεδεμένα στοιχειά

Ένας εναλλακτικός τρόπος για να οριστούν τα συνδεδεμένα στοιχειά περιλαμβάνει την ισοδυναμία των κλάσεων από τις ισοδύναμες σχέσεις οι οποίες ορίζονται στις κορυφές των γραφών. Σε έναν μη κατευθυνόμενο γράφο, μια κορυφή  $v$  είναι *προσβάσιμη* από μια κορυφή  $u$  εάν υπάρχει μονοπάτι από την  $u$  στην  $v$ . Σε αυτόν τον ορισμό, μια μοναδική κορυφή μετράτε ως ένα μονοπάτι μηδενικής απόστασης, και η ίδια κορυφή μπορεί να συμβεί παραπάνω από μια φορές μέσα σε ένα μονοπάτι. Η ιδιότητα να φτάσει μια κορυφή την άλλη είναι μια σχέση ισοδυναμίας[8], όταν:

- ♣ Είναι αυτοπαθής: Υπάρχει ένα ασήμαντο μονοπάτι μηδενικού μήκους από κάθε κορυφή στον εαυτό της.
- ♣ Είναι συμμετρικός: Αν υπάρχει μονοπάτι από το  $u$  στο  $v$ , οι ίδιες ακμές αποτελούν ένα μονοπάτι από το  $v$  στο  $u$ .
- ♣ Είναι μεταβατικός: Εάν υπάρχει μονοπάτι από το  $u$  στο  $v$  και ένα μονοπάτι από το  $v$  στο  $w$ , τα δυο μονοπάτια μπορούν να ενωθούν μαζί για να σχηματίσουν ένα μονοπάτι από το  $u$  στο  $w$ .

Τα συνδεδεμένα στοιχειά είναι τότε οι παραγόμενοι γράφοι που σχηματίζονται από τις ισοδύναμες κατηγορίες αυτής της σχέσης. Ο αριθμός των συνδεδεμένων στοιχείων είναι σημαντικός και τοπολογικά αναλλοίωτος για ένα γράφο. Στην θεωρία τοπολογιών γραφών μπορεί να ερμηνευτεί σαν τον σειριακό μηδενικό αριθμό του Betti σε έναν γράφο. Στην αλγεβρική θεωρία των γραφών ο αριθμός αυτός ισούται με την πολλαπλότητα του μηδενός ως μια ιδιότητα ενός πίνακα Laplace . Είναι επίσης ο δείκτης του πρώτου μη μηδενικού συντελεστή του χρωματικού πολυώνυμου σε έναν γράφο. Ο αριθμός των συνδεδεμένων στοιχείων παίζει σημαντικό ρόλο στο θεώρημα του Tutte που χαρακτηρίζει τους γράφους που έχουν τέλει ταίριασμα και τον ορισμό της σκληρότητας των γραφών[8].

Στους αλγορίθμους είναι εύκολο να υπολογίσουμε τα συνδεδεμένα στοιχειά των γραφών σε γραμμικό χρόνο

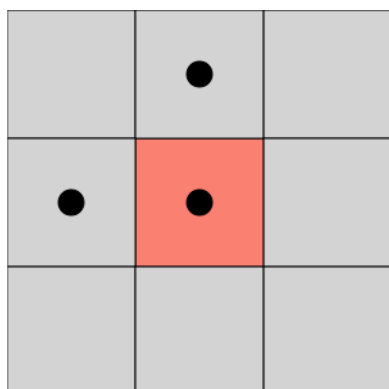
χρησιμοποιώντας είτε αναζήτηση πλάτους – πρώτα ή βάθους πρώτα. Σε κάθε περίπτωση, μια ερευνά που ξεκινά από μια συγκεκριμένη κορυφή  $v$  θα βρει ολόκληρα συνδεδεμένα στοιχεία  $v$ , και όχι περισσότερα, πριν επιστρέψει. Για να βρει άλλα τα συνδεδεμένα στοιχεία σε έναν γράφο, επαναλαμβάνουμε πέρασμα από τις κορυφές, ξεκινώντας μια νέα αναζήτηση πλάτους – πρώτα ή βάθους πρώτα όποτε η επανάληψη φτάσει σε κορυφή που δεν είχε συμπεριληφθεί στα προηγούμενα συνδεδεμένα στοιχεία. Ο Hopcroft και ο Tarjan περιγράφουν κατ' ουσίαν αυτόν τον αλγόριθμο, και δηλώνουν ότι έχει εκείνο το σημείο ήταν “πολύ γνώριμος”[8].

Υπάρχουν επίσης αποδοτικοί αλγόριθμοι για να ανιχνεύουν δυναμικά τα συνδεδεμένα στοιχεία ενός γράφου προστίθενται ακμές και κορυφές, σαν μια απλή εφαρμογή από σύνολο ασύνδετων δομών δεδομένων. Αυτοί οι αλγόριθμοι απαιτούν αναπόσβεστο χρόνο  $O(n)$  για κάθε λειτουργία, όπου  $n$  το να προσθεθεί κορυφές και ακμές και να καθορίζεις συνδεδεμένα στοιχεία στην οποία μια κορυφή εμπίπτει σε δυο λειτουργίες, και  $n$  είναι μια πολύ αργά αυξανόμενη αντίστροφη από μια πολύ γρήγορα αναπτυσσόμενη λειτουργία Ackermann. Ένα παρόμοιο πρόβλημα είναι η ανίχνευση συνδεδεμένων στοιχείων καθώς όλες οι ακμές διαγράφονται μια προς μια από έναν γράφο, ο αλγόριθμος που υπάρχει για την επίλυση του με ένα σταθερό χρονικό διάστημα κάθε εξέτασης, και  $O(|V||E|)$  χρόνος να διατηρήσει την δομή των δεδομένων. Αυτό είναι ένα αποσβεσμένο κόστος του  $O(|V|)$  ανά ανίχνευση ακμής. Για τα δάση γράφων, το κόστος μπορεί να μειωθεί σε  $O(q + |V| \log |V|)$ , ή  $O(\log |V|)$  αποσβεσμένο κόστος ανά ανίχνευση ακμής[8].

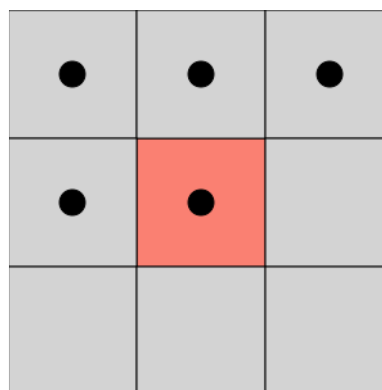
Ερευνητές επίσης έχουν μελετήσει αλγορίθμους για την εύρεση συνδεδεμένων στοιχείων σε ποιο οριακά μοντέλα υπολογισμών, όπως προγράμματα στα οποία η μνήμη είναι περιορισμένη με ένα λογαριθμικό αριθμό από bits. Ο Lewis και ο Παπαδημητρίου (1982) αναρωτήθηκαν το πότε είναι δυνατόν να ελεγχθούν σε λογαριθμικοί χώρου πινάκες εάν δυο κορυφές ανήκουν στο ίδιο συνδεδεμένο στοιχείο ενός μη κατευθυνόμενου γράφου, και όρισαν μια κατηγορία πολυπλοκότητας SL των προβλημάτων συνδεσιμότητα των λογαριθμικοί χώρου πινάκων. Τελικά, ο Reingold (2008) κατάφερε στο να βρει έναν αλγόριθμο για την επίλυση προβλημάτων σε λογαριθμικό χώρο, αποδεικνύοντας ότι  $L=SL$ [8].

Η τεχνική που χρησιμοποιήθηκε στο σύστημα ήταν η επισήμανση των συνδεδεμένων στοιχείων. Ο αλγόριθμος ονομάζεται διαφορετικά και ανάλυση συνδεδεμένων στοιχείων, εξαγωγή σταγόνας, επισήμανση περιοχής, ανακάλυψη σταγόνας ή εξαγωγή περιοχής. Η επισήμανση των συνδεδεμένων στοιχείων είναι ένας αλγοριθμική εφαρμογή της θεωρίας των γράφων, οπότε ένα υποσύνολο από συνδεδεμένα στοιχεία είναι μοναδικά επισημασμένα, και δεν πρέπει σε καμία περίπτωση να συγχέεται με την κατάτμηση των γράφων. Η συνηθισμένη χρήση του αλγορίθμου είναι στην τεχνητή όραση για την ανίχνευση περιοχών σε μια δυαδική ψηφιακή εικόνα, παρόλα αυτά έγχρωμες εικόνες και δεδομένα με περισσότερες διαστάσεις μπορούν να επεξεργαστούν. Όταν χρησιμοποιήσετε σε συστήματα αναγνώριση εικόνας ή σε περιβάλλον αλληλεπίδρασης με ανθρώπους ο αλγόριθμος μπορεί να λειτουργήσει σε μια ποικιλία πληροφοριών. Η εξαγωγή φυσαλίδας γενικά εκτελείτε σε μια δυαδική εικόνα που προκύπτει από ένα βήμα με κατώτατο όριο αποδεκτών στοιχείων. Η φυσαλίδες μπορούν να μετρηθούν, φιλτραριστούν και ανιχνευθούν[20].

Ένας γράφος που αποτελείτε από κορυφές και ακμές, κατασκευάζετε από σχετικά δεδομένα εισόδου. Οι κορυφές περιέχουν πληροφορίες που απαιτείτε από την εφευρετική σύγκρισης, καθώς οι ακμές σηματοδοτούν τους συνδεδεμένους “γείτονες”. Ένας αλγόριθμος ο οποίος διατρέχει των γράφο, επισημαίνοντας τις κορυφές που βασίζονται στην συνδεσιμότητα και συγγενικές τιμές των γειτόνων τους. Η συνδεσιμότητα προσδιορίζετε από το μέσο, όπως για παράδειγμα στους γράφους μιας εικόνας μπορούμε να έχουμε συνδεδεμένους σε 4 σημεία και σε 8 σημεία[20]. Οι αλγόριθμοι μπορούν να γενικευτούν σε αυθαίρετες διαστάσεις, αν και με αυξημένο χρόνο και πολυπλοκότητα του χώρου.

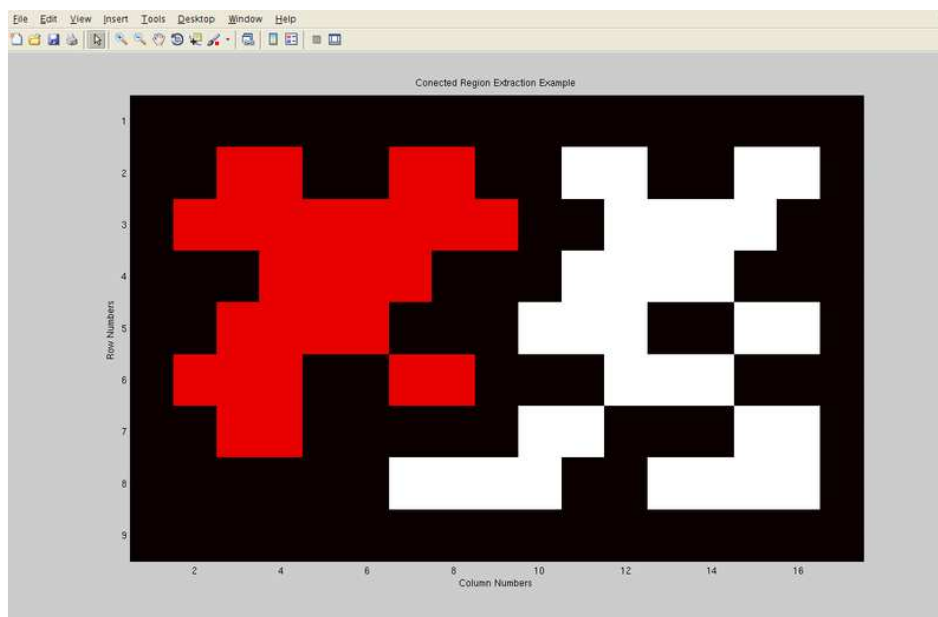


Εικόνα 46: Επισήμανση των συνδεδεμένων στοιχείων σε 4 σημεία.



Εικόνα 47: Επισήμανση των συνδεδεμένων στοιχείων σε 8 σημεία.

Ένας γενικά εύκολος σε εφαρμογή και κατανόηση αλγόριθμος είναι ο αλγόριθμος δυο περασμάτων, ο οποίος ερευνά δισδιάστατα δυαδικά δεδομένα. Ο αλγόριθμος κάνει δυο περάσματα από μια εικόνα, στο πρώτο πέρασμα για να καταγράψει ισοδυναμίες και να αναθέσει προσωρινές επισημάνσεις και στο δεύτερο πέρασμα για να αντικαταστήσει προσωρινές επισημάνσεις από την επισήμανση της ισοδύναμης κατηγορίας της. Τα δεδομένα που δίνονται για επεξεργασία μπορούν να τροποποιηθούν “επί τόπου”, που περιεχέει και έναν κίνδυνο διαφθοράς δεχόμενων, ή οι επισημανθείσες πληροφορίες μπορούν να διατηρηθούν σε μια πρόσθετη δομή δεδομένων[20].



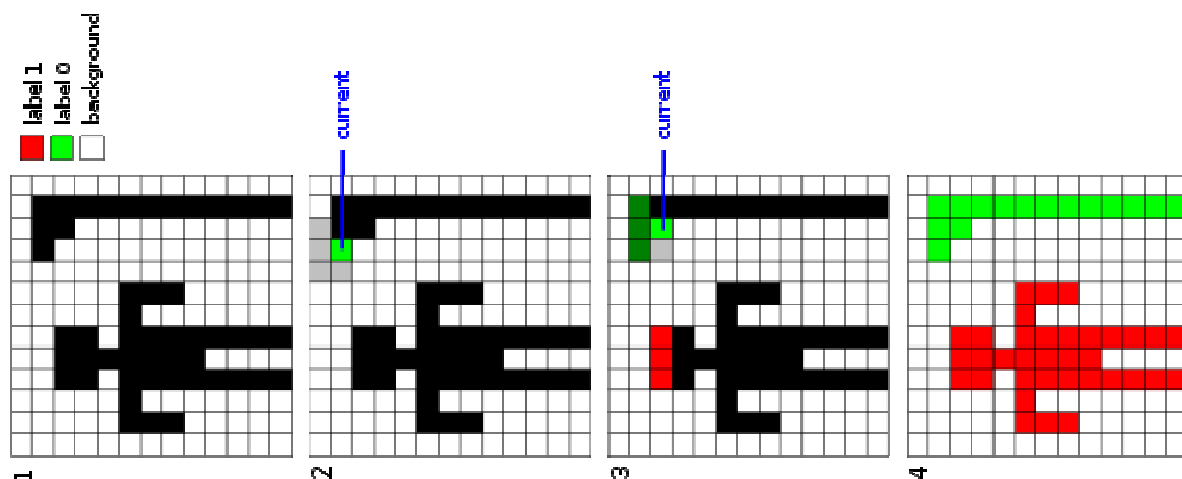
Εικόνα 48: Παράδειγμα επισήμανσης των συνδεδεμένων στοιχείων σε δυαδική εικόνα με δυο περιοχές

Οι Έλεγχοι συνδεσιμότητα εκτελούνται με το να ελέγχονται οι επισημάνσεις κάθε εικονοστοιχείου που βρίσκονται βορειοανατολικά, βόρεια, βορειοδυτικά και δυτικά από το σημείο του συγκεκριμένου εικονοστοιχείου, θεωρώντας ότι έχουμε επισήμανση των συνδεδεμένων στοιχείων σε 8 σημεία.. Επισήμανση των συνδεδεμένων στοιχείων σε 4 σημεία χρησιμοποιεί μόνο τους βορινούς και δυτικούς γείτονες του συγκεκριμένου εικονοστοιχείου[20]. Οι ακόλουθοι όροι ελέγχονται για να προσδιορίσουν την τιμή κάθε επισήμανσης που θα ανατεθεί σε συγκεκριμένο εικονοστοιχείο, θεωρούμε ότι έχουμε επισήμανση

των συνδεδεμένων στοιχείων σε 4 σημεία:  
 Συνθήκες που πρέπει να ελεγχθούν:

- Έχει το εικονοστοιχείο στα αριστερά (δυτικά) την ίδια τιμή με το εξεταζόμενο;  
 Ναι – έχουμε κοινή περιοχή. Ανάθεσε την ίδια επισήμανση στο εικονοστοιχείο.  
 Όχι - έλεγξε την επόμενη συνθήκη.
- Έχουν τα εικονοστοιχεία στον Βορρά και στην Δύση του συγκεκριμένου εικονοστοιχείου την ίδια τιμή αλλά όχι την ίδια επισήμανση;  
 Ναι – Ξέρουμε ότι τα βόρεια και τα δυτικά εικονοστοιχεία ανήκουν στην ίδια περιοχή και πρέπει να ενωθούν. Ανάθεσε στο συγκεκριμένο εικονοστοιχείο την ελάχιστη επισήμανση των βορείων και των δυτικών επισημάνσεων, και κατέγραψε την ισοδύναμη τους σχέση.  
 Όχι - έλεγξε την επόμενη συνθήκη.
- Έχει το αριστερό εικονοστοιχείο (δυτικό) διαφορετική τιμή και αυτό το οποίο βρίσκετε βοριά έχει την ίδια τιμή?  
 Ναι – Ανάθεσε την επισήμανση του βορίου εικονοστοιχείου σε αυτό το εικονοστοιχείο.  
 Όχι - έλεγξε την επόμενη συνθήκη.
- Έχει το εικονοστοιχείο βορινούς και δυτικούς γείτονες με διαφορετικές τιμές;  
 Ναι – Φτιάξε μια ξεχωριστή ταυτότητα επισήμανσης σε αυτό το εικονοστοιχείο.

Ο αλγόριθμος συνεχίζει με αυτόν τον ρυθμό, και δημιουργεί νέες επισημασμένες περιοχές όπου αυτό είναι απαραίτητο. Το κλειδί για ένα γρήγορο αλγόριθμο είναι το πως γίνονται οι ενώσεις ανάμεσα στα εικονοστοιχεία. Ο παραπάνω αλγόριθμος χρησιμοποιεί ενοποιημένη εύρεση δομή δεδομένων όπου παρέχει εξαιρετική απόδοση για την ενημέρωση των ισοδυνάμων σχέσεων. Η ενοποιημένη εύρεση κατ' ουσίαν αποθηκεύει επισημάνσεις οι οποίες αντιστοιχούν στην ίδια περιοχή σε μια συνεχές σύνολο δομών δεδομένων, κάνοντας εύκολη την απομνημόνευση την ισότητα δυο επισυνάψεων με την χρήση μιας μεθόδου διασύνδεσης. Αφού έχει γίνει η αρχική επισήμανση, στο δεύτερο πέρασμα απλώς αντικαθιστώνται κάθε επισημασμένα εικονοστοιχεία με τον ισοδύναμο εκπρόσωπο στοιχείο από το ασυνεχές σύνολο[20].



Εικόνα 49: Παράδειγμα αλγορίθμου επισήμανσης των συνδεδεμένων στοιχείων δυο περασμάτων σε δυαδική εικόνα

Η προσφορά αυτού του αλγορίθμου στο σύστημα είναι αρκετά μεγάλη και χρησιμοποιήθηκε στην εύρεση περιοχών όπου πιθανόν να αντιστοιχούσαν σε αντικείμενα στην εικόνα που εξετάζονταν, με πλεονεκτήματα του να είναι η περιορισμένη απαίτηση του σε υπολογιστική ισχύ και η εύκολη χρήση του.

### 2.4.3 Μετασχηματισμός κλιμακωτά – αναλλοίωτων χαρακτηριστικών (SIFT)

Ο μετασχηματισμός κλιμακωτά – αναλλοίωτων χαρακτηριστικών (SIFT) είναι ένας αλγόριθμος που ανήκει στην τεχνητή όραση που ανακαλύπτει και περιγράφει τοπικά χαρακτηριστικά στις εικόνες. Ο αλγόριθμος έχει δημοσιευθεί από τον David Lowe το 1999. Οι εφαρμογές που έχει χρησιμοποιηθεί είναι αναγνώριση αντικείμενων, παραγωγή χαρτών και πλοήγησης ρομπότ, “ράψιμο” εικόνων, παραγωγή τρισδιάστατων μοντέλων, αναγνώριση χειρονομιών, ανίχνευση αντικείμενων σε βίντεο και χρήση εφέ σε βίντεο. Ο αλγόριθμος έχει καταχωρημένο δίπλωμα ευρεσιτεχνίας στις Ηνωμένες Πολιτείες της Αμερικής υπό την ιδιοκτησία του Πανεπιστημίου της Βρετανικής Κολούμπια. Όμως παρέχετε για την χρήση του σε ερευνητικού περιεχόμενου συστήματα[9].

Το ταίριασμα των εικόνων είναι η θεμελιώδεις πτυχή πολλών προβλημάτων στην τεχνητή όραση, συμπεριλαμβανομένου την αναγνώριση αντικείμενων ή τοπίων, επιλύοντας την τρισδιάστατη δομή πολλών εικόνων, την στερεοσκοπική αντιστοιχία και την ανίχνευση κίνησης. Αυτός ο αλγόριθμος περιγράφει χαρακτηριστικά των εικόνων που έχουν πολλές ιδιότητες που τα κάνουν κατάλληλα για το ταίριασμα διαφορετικών εικόνων του ίδιου αντικείμενου ή σκηνής. Τα χαρακτηριστικά αυτά είναι αναλλοίωτα στην κλιμάκωση της εικόνας και την περιστροφή της, και μερικώς αναλλοίωτα στην αλλαγή στον φωτισμό και την τοποθέτηση της κάμερας. Είναι πολύ καλά τοπικά τοποθετημένα και στον τομέα του χώρου άλλα και των συχνοτήτων, μειώνοντας έτσι την πιθανότητα αποδιοργάνωσης από κρυφά σημεία, ή άλλων θορύβων. Μεγάλοι αριθμοί από χαρακτηριστικά μπορούν να εξαχθούν από τυπικές εικόνες με αποδοτικούς αλγόριθμους. Επιπλέον, τα χαρακτηριστικά έχουν υψηλή διακριτικότητα, που επιτρέπει ένα μόνον χαρακτηριστικό να μπορεί να αντιστοιχεί με μεγάλη πιθανότητα σε μια μεγάλη βάση δεδομένων από χαρακτηριστικά, αποδεικνύοντας τις βάσεις για αναγνώριση αντικείμενων και σκηνών[21].

Το κόστος τις εξαγωγής αυτόν των χαρακτηριστικών μειώνετε λαμβάνοντας μια προσέγγιση φιλτραρίσματος μιας σειράς σύνδεσης, στην οποία η ποιο “ακριβές” πράξεις γίνονται μόνον σε τοποθεσίες που περνούν την αρχική δοκιμασία. Ακολουθώντας τα παρακάτω βήματα επεξεργασίας παράγονται τα σύνολα των χαρακτηριστικών στις εικόνες:

1. Κλιμακωτή ανίχνευση ακρότατων στοιχείων στον χώρο.: Αυτό το πρώτο στάδιο της επεξεργασίας ερευνά όλες τις κλίμακες και τοποθεσίες στις εικόνες. Έχει υλοποιηθεί αποδοτικά με την χρήση μιας λειτουργίας διαφοράς των Gaussians φίλτρων για την ανίχνευση πιθανών σημείων ενδιαφέροντος τα οποία είναι αμετάβλητα στην κλίμακα και τον προσανατολισμό.
2. Εντοπισμός σημείων κλειδιών: Σε κάθε υποψηφία τοποθεσία, ένα λεπτομερές μοντέλο είναι κατάλληλο να καθορίσει την τοποθεσία και την κλίμακα. Τα σημεία κλειδιά επιλέγονται βάση την μέτρηση της σταθερότητας τους.
3. Ανάθεση προσανατολισμού: Ένας ή περισσότεροι προσανατολισμοί αναθέτοντε σε κάθε σημείο κλειδί βασιζόμενο στις τοπικές κατευθύνσεις κλίσης της εικόνας. Όλες οι μελλοντικές εργασίες εκτελούνται στα δεδομένα τις εικόνες που έχει μετασχηματιστεί σχετικά με τον αποδιδόμενο προσανατολισμό, κλίμακα και τοποθεσία για κάθε χαρακτηριστικό, αποδεικνύοντας έτσι τον αναλλοίωτο χαρακτήρα σε αυτές τις μετατροπές.
4. Περιγραφέας σημείων κλειδιών: Οι βαθμίδες των τοπικών εικόνων μετρώνται στην συγκεκριμένη κλίμακα στην περιοχή γύρω από κάθε σημείο κλειδί. Μεταμορφώνονται σε αναπαραστάσεις που επιτρέπουν για σημαντικά επίπεδα διαστρέβλωσης της τοπικής μορφής και αλλαγές στην φωτεινότητα.

Αυτή η προσέγγιση ονομάζεται μετασχηματισμός κλιμακωτά – αναλλοίωτων χαρακτηριστικών (SIFT), καθώς μεταμορφώνει δεδομένα εικόνας σε κλιματικά - αναλλοίωτων συντεταγμένες συγγενικές σε τοπικά χαρακτηριστικά. Μια σημαντική πτυχή αυτής της προσέγγισης είναι ότι παράγει μεγάλες ποσότητες από χαρακτηριστικά τα οποία καλύπτουν πυκνά την εικόνα σε όλο το φάσμα των κλιμάκων και των τοποθεσιών. Μια τυπική εικόνα μεγέθους 500 \* 500 εικονοστοιχείων μπορεί να παράγει περίπου 2000 σταθερά χαρακτηριστικά, παρόλα που το νούμερο αυτό εξαρτάται και από το περιεχόμενο της εικόνας και άλλες επιλογές διάφορων παραμέτρων. Η ποσότητα των χαρακτηριστικών είναι πολύ σημαντική ειδικά για ανάγνωση αντικείμενων, όπου η ικανότητα για την εύρεση μικρών αντικειμένων σε “ακατάστατα” υπόβαθρα απαιτεί ότι τουλάχιστον τρία χαρακτηριστικά για να ταιριάζει σωστά με ένα αντικείμενο και να παρέχει αξιόπιστη ταυτοποίηση[21].

Για την αντιστοίχιση εικόνων και την αναγνώριση, τα χαρακτηριστικά SIFT πρώτα εξάγονται από ένα σύνολο από εικόνες αναφοράς που έχουν αποθηκευτεί σε μια βάση δεδομένων. Στην κλασική προσέγγιση του αλγορίθμου είναι σε μια νέα εικόνα να αντιστοιχείται και να συγκρίνεται μεμονωμένα κάθε χαρακτηριστικό από κάθε νέα εικόνα με την προηγούμενη βάση δεδομένων και να βρίσκετε υποψήφιο χαρακτηριστικό αντιστοιχίας που βασίζετε στην Ευκλείδεια απόσταση των χαρακτηριστικών τους διανυσμάτων. Έτσι ο αλγόριθμος ενσωματώνει και έναν υπό – αλγόριθμο γρήγορης εύρεσης του κοντινότερου γείτονα που μπορεί να ανιχνεύσει χαρακτηριστικά μέσα σε μια μεγάλη βάση δεδομένων. Το συγκεκριμένο κομμάτι δεν χρησιμοποιήθηκε από το σύστημα αντί' αυτού χρησιμοποιήθηκε ένα τεχνητό νευρωνικό δίκτυο το οποίο δεν απαιτούσε από τα χαρακτηριστικά μια διαφορετική ποιο αποτελεσματική προσέγγιση ταιριάσματος[21].

Οι περιγράφεις των σημείων κλειδιών είναι υψηλά διακριτά, αυτό επιτρέπει σε ένα μοναδικό χαρακτηριστικό να βρει το σωστό του ταίρι με μεγάλη πιθανότητα μέσα σε μεγάλες βάσεις δεδομένων. Παρόλα αυτά, σε μια πυκνή εικόνα με πολλά αντικείμενα, πολλά από τα χαρακτηριστικά από το υπόβαθρο δεν θα έχουν κάποιο σωστό ταίριασμα στην βάση δεδομένων, δίνοντας την ικανότητα για πολλά λάθος ταιριάσματα σε αντίθεση με τα σωστά. Η σωστές αντιστοιχίες μπορούν να φιλτραριστούν από ένα πλήρες σύνολο από αντιστοιχίες με τον προσδιορισμό υπό – συνόλων από σημεία κλειδιά τα οποία συμφωνούν στο αντικείμενο και στην τοποθεσία του, την κλίμακα, και τον προσανατολισμό της νέας εικόνας. Η πιθανότητα ότι κάποια χαρακτηριστικά θα συμφωνούν με αυτές τις παραμέτρους από τύχη είναι πολύ μικρότερη από την πιθανότητα κάθε ταίριασμα χαρακτηριστικού να είναι λάθος. Η αποφασιστικότητα αυτών των συνεπών συστάδων μπορεί να εκτελεστούν ταχύτατα με την χρήση ενός αποδοτικού πίνακα κατακερματισμού του γενικευμένου μετασχηματισμού Hough. Παρόλα αυτά το σύστημα που περιγράφετε σε αυτήν την πτυχιακή μπορεί να παρακάμψει αυτόν τον μετασχηματισμό χάρη στην χρήση του τεχνητού νευρωνικού δικτύου[21].

Η ανίχνευση των σημείων κλειδιών χρησιμοποιεί μια αλληλεπικαλυπτόμενη μέθοδο φιλτραρίσματος που χρησιμοποιεί αποδοτικούς αλγορίθμους για την αναγνώριση υποψηφίων τοποθεσιών που εξετάζονται με μεγαλύτερη λεπτομέρεια. Το πρώτο στάδιο για την ανίχνευση είναι η αναγνώριση τοποθεσιών και κλιμάκων που μπορούν επανειλημμένα να ανατεθούν κάτω από διαφορετικές προβολές του ίδιου αντικειμένου. Η αναγνώριση τοποθεσιών που είναι αμετάβλητες στις αλλαγές μιας εικόνας μπορεί να επιτευχθεί με την αναζήτηση για σταθερά χαρακτηριστικά σε όλες τις πιθανές κλίμακες, με την χρήση μιας συνεχείς λειτουργία της κλίμακας γνωστή και ως κλιμάκωση χώρου (Witkin,1983)[21].

Όπως έχει αποδειχθεί από τον Koenderink (1984) και τον Lindeberg (1994) ότι κάτω από μια ποικιλία από λογικές υποθέσεις ο μόνος δυνατός πυρήνας κλίμακας και χώρου είναι η Gaussian λειτουργία. Για αυτόν τον λόγο, ο κλιμακωτός χώρος μιας εικόνας ορίζετε ως  $L(x, y, \sigma)$ , που παράγετε από την συνέλιξη ενός μεταβλητής κλίμακας Gaussian,  $G(x, y, \sigma)$ , με μια εικόνα εισόδου  $I(x, y)$ :

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

*Εικόνα 50: Κλιμακωτός χώρος μιας εικόνας  $L(x, y, \sigma)$ .*



οπού \* είναι η λειτουργία της συνελλιξης στο x και το y, και:

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

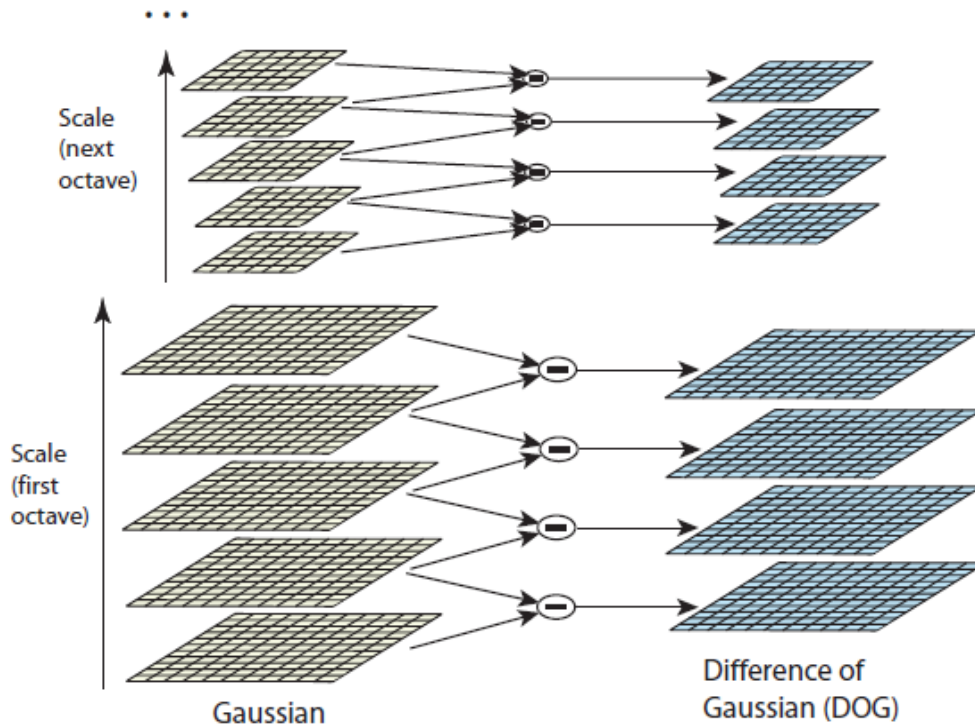
*Εικόνα 51: Συνέλιξη ενός μεταβλητής κλίμακας Gaussian,  $G(x, y, \sigma)$ .*

Για την αποδοτική αναγνώριση σταθερών τοποθεσιών σημείων κλειδιών στον κλιμακωτό χώρο, όπως είχε προταθεί και από τον Lowe το 1999 η χρήση ακρότατων κλιμακωτού χώρου σε λειτουργία διαφοράς- του - Gaussian συνελιχθεί με την εικόνα  $D(x, y, \sigma)$ , που μπορεί να υπολογιστεί από την διαφορά δυο κοντινών κλιμάκων που χωρίζονται από έναν σταθερό πολλαπλασιαστικό παράγοντα k:

$$\begin{aligned} D(x, y, \sigma) &= (G(x, y, k\sigma) - G(x, y, \sigma)) * I(x, y) \\ &= L(x, y, k\sigma) - L(x, y, \sigma). \end{aligned}$$

*Εικόνα 52: Ακρότατα κλιμακωτού χώρου σε λειτουργία διαφοράς- του - Gaussian σε συνέλιξη με την εικόνα  $D(x, y, \sigma)$*

Υπάρχει ένας αριθμός από λόγους για την επιλογή αυτής της λειτουργίας. Πρώτον, είναι ιδιαίτερα αποδοτική λειτουργία για υπολογισμό, καθώς οι εξομαλυμένες εικόνες, L, πρέπει να υπολογιστούν για κάθε περίπτωση περιγραφής χαρακτηριστικών κλιμακωτού χώρου, και για αυτόν τον λόγο το D μπορεί να υπολογιστεί με μια απλή αφαίρεση στην εικόνα[21].



Εικόνα 53: Αναπαράσταση συνέλιξης και αφαίρεσης κάθε οκτάβας για την δημιουργία διαφοράς - της - Gaussian

Όπως φαίνεται στην παραπάνω εικόνα κάθε οκτάβα από κλιμακωτό χώρο, η αρχική εικόνα συνεχώς συνελίσσεται με Gaussian για να παράγει ένα σύνολο από κλιμακωτού χώρου εικόνες στα αριστερά. Διπλές Gaussian εικόνες αφαιρούνται για να παράγουν την διαφορά - της - Gaussian των εικόνων στα δεξιά. Μετά από κάθε οκτάδα, η εικόνα Gaussian υπό - δειγματοληπτείται με συντελεστή δυο, και η διαδικασία συνεχίζεται.

Επιπλέον, η διάφορα - της - Gaussian λειτουργίας παρέχει μια στενή προσέγγιση στην κανονικοποιημένη κλιμακωτά Laplacian του Gaussian,  $\sigma^2 \square^2 G$ , όπως αυτό μελετήθηκε από τον Lindeberg (1994). Ο Lindeberg έχει αποδείξει ότι η κανονικοποίηση της Laplacian με συντελεστή δυο είναι απαραίτητη για να παραμείνει αναλλοίωτη η κλίμακα. Σε λεπτομερή πειραματικές συγκρίσεις, ο Mikolajczyk (2002) βρήκε ότι τα μέγιστα και τα ελάχιστα του  $\sigma^2 \square^2 G$ , παράγουν τα ποιο σταθερά χαρακτηριστικά της εικόνας συγκρινόμενα με ένα εύρος άλλων λειτουργιών στις εικόνες, όπως η κλίσεις, Hessian, ή η λειτουργία εύρεσης γωνιών Harris[21].

Οι σχέσεις μεταξύ του  $D$  και του  $\sigma^2 \square^2 G$  είναι κατανοητά από εξίσωση διάχυσης της θερμότητας, παρατροποποιημένες σε όρους διαφορετικούς από το συνηθισμένο  $t = 2$ :

$$\frac{\partial G}{\partial \sigma} = \sigma \square^2 G.$$

Εικόνα 54: Οι σχέσεις μεταξύ του  $D$  και του  $\sigma^2 \square^2 G$

Από το παραπάνω, βλέπουμε ότι  $\square^2 G$  μπορεί να υπολογιστεί από την πεπερασμένη διαφορική προσέγγιση στο  $\theta G / \theta \sigma$ , χρησιμοποιώντας την διαφορά των γύρω κλιμάκων στο  $k$  και:

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

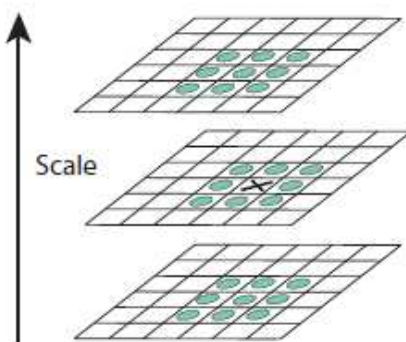
Εικόνα 55: Υπολογισμός πεπερασμένης διαφορετικής προσέγγισης στο  $\theta G/\theta\sigma$ , χρησιμοποιώντας την διαφορά των γύρω κλιμάκων στο  $k$

και για αυτόν τον λόγο, έχουμε:

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

Εικόνα 56: Διαφορά- της - Gaussian λειτουργίας οι οποίες διαφέρουν από έναν σταθερό παράγοντα

Αυτό δείχνει την διάφορα όταν η διάφορα - της - Gaussian λειτουργίας έχει κλίμακες οι οποίες διαφέρουν από έναν σταθερό παράγοντα ήδη ενσωματώνει την κλίμακα ομαλοποίησης του δυο είναι απαραίτητη για την μη μεταβλητότητα της κλίμακας Laplacian.



Εικόνα 57: Ο υπολογισμός του μέγιστου και ελάχιστου μιας διαφοράς- της - Gaussian που ανιχνεύετε με σύγκριση ενός εικονοστοιχείου

Στην παραπάνω εικόνα είναι εμφανείς η διαδικασία στην οποία φαίνεται ο υπολογισμός του μέγιστου και του ελάχιστου μιας διαφοράς- της - Gaussian που ανιχνεύετε με την σύγκριση ενός εικονοστοιχείου (που είναι σημειωμένο με X) με τους 26 γείτονες του σε μια περιοχή 3x3 στις συγκεκριμένες και παρακείμενες κλίμακες (που είναι σημειωμένες με κύκλους)[21].

Ο παράγοντας  $(k - 1)$  στην εξίσωση είναι μια σταθερά πάνω σε κάθε κλίμακα και για αυτόν τον λόγο δεν επηρεάζει την τοποθεσία των μεγίστων. Το λάθος της προσέγγισης θα φτάσει στο μηδέν καθώς το  $k$  πηγαίνει στο 1, αλλά στην πραγματικότητα η προσέγγιση δεν έχει σχεδόν καμιά επιρροή στην σταθερότητα της εύρεσης μεγίστων για ακόμα μεγαλύτερες διαφορές στην κλίμακα, όπως το  $k = \sqrt{2}$ . Μια αποτελεσματική προσέγγιση στην κατασκευή της  $D(x, y, \sigma)$  φαίνεται στην εικόνα 53. Η αρχική εικόνα σταδιακά συνελλίσσετε με τις Gaussians για να παράγει εικόνες που χωρίζονται από έναν σταθερό παράγοντα  $k$  στην κλίμακα του χώρου, που φαίνεται στοιβαγμένοι στην αριστερή στήλη. Έχει επιλεγεί ο διαχωρισμός κάθε οκτάβας της κλίμακας του χώρου σε έναν ακέραιο αριθμό,  $s$ , των διαστημάτων, έτσι το  $k = 21/s$ . Πρέπει να παραχθούν

s+3 εικόνες στην σωρό με τις θολές εικόνες κάθε οκτάβας, έτσι ώστε η τελική ανίχνευση μέγιστον να καλύπτει μια ολόκληρη οκτάβα. Δίπλες Gaussian εικόνες αφαιρούνται για να παράγουν την διάφορα – της - Gaussian των εικόνων στα δεξιά. Όταν μια πλήρη οκτάβα είναι επεξεργασμένη, δηματοληπούμε ξανά την Gaussian εικόνα που έχει διπλή την αρχική τιμή περνώντας κάθε δεύτερο εικονοστοιχείο από κάθε σειρά και στήλη. Η ακρίβεια της δειγματοληψίας δεν είναι διαφορετική από την αρχική της προηγούμενης οκτάβας, ενώ η επεξεργαστική ισχύς που χρειάζεται, μειώνετε σημαντικά[21].

Για να βρούμε τα τοπικά μέγιστα και ελάχιστα του  $D(x, y, \sigma)$ , κάθε σημείο δείγματος συγκρίνεται με τους οκτώ γείτονες του στην συγκεκριμένη εικόνα και με τους εννιά γείτονες του στην παραπάνω και παρακάτω κλίμακα όπως στην εικόνα 57. Επιλέγεται μόνο εάν είναι μεγαλύτερο από άλλους τους γείτονες του ή μικρότερο από αυτούς. Το κόστος αυτού του ελέγχου είναι λογικά χαμηλό καθώς τα περισσότερα δείγματα θα εξαλειφθούν από τους πρώτους ελέγχους. Ένα σοβαρό θέμα είναι ο καθορισμός της συχνότητας δειγματοληψίας της εικόνας και των κλιμακωτών τομέων που είναι απαραίτητα για την αξιόπιστη ανίχνευση μεγίστων. Δύστυχος, όπως είναι εμφανές ότι δεν υπάρχει ελάχιστο διάστημα δειγμάτων που θα ανακαλύψουν όλα τα μέγιστα, καθώς τα μέγιστα μπορεί να είναι αυθαίρετος πολύ κοντά μεταξύ τους. Αυτό μπορεί να γίνει προφανές αν θεωρήσουμε έναν άσπρο κύκλο σε μαύρο περίγραμμα, το οποίο θα έχει μια μοναδική κλίμακα τοπικών μεγίστων όπου η κυκλική θετική κεντρική περιοχή της διαφοράς– της - Gaussian λειτουργίας ταιριάζει με το μέγεθος και την τοποθεσία του κύκλου. Για κάθε επιμήκεις έλλειψη, θα υπάρχουν δυο μέγιστα κοντά στο τέλος της έλλειψης. Καθώς η αναζήτηση περιοχών μεγίστων είναι συνεχής λειτουργία σε μια εικόνα, για κάποιες ελλείψεις με ενδιάμεση επιμήκυνση θα υπάρχει μια μετάβαση από το μόνο μέγιστο σε δυο, με τα μέγιστα αυθαίρετα κοντά το ένα στο άλλο κοντά στην μετάβαση.

Για αυτόν τον λόγο έπρεπε να βρεθεί συμβιβαστική λύση οι οποία ανταλλάσσει την αποδοτικότητα με την πληρότητα των αποτελεσμάτων. Στην πραγματικότητα, όπως αναμένεται τα μέγιστα τα οποία είναι πολύ κοντά και μαζί είναι ασταθή στις μικρές διαταραχές τις εικόνας[21].

Σε πειραματικό προσδιορισμό της συχνότητας της δειγματοληψίας στο φαινόμενο του ποικίλου αριθμού κλιμάκων ανά οκτάβα στην οποία η λειτουργία στην εικόνα δειγματοληπτείται εκ των προτέρων για την ανίχνευση μεγίστων. Σε αυτήν την περίπτωση, κάθε εικόνα δειγματοληπτημένο ξανά ακολουθώντας την περιστροφή σε μια τυχαία γωνία και σε μια τυχαία κλίμακα ανάμεσα σε 0,2 ή 0,9 την αρχική εικόνα. Τα σημεία κλειδιά από τις μειωμένης ανάλυσης εικόνες αντιστοιχίστηκαν με αυτές από την αρχική εικόνα έτσι ώστε οι κλίμακες για όλα τα σημεία κλειδιά να είναι παρών στην αντιστοιχεί εικόνα. Επιπλέον, το 1% του θορύβου τις εικόνας προστέθηκε, για να δείξουμε ότι κάθε τιμή του εικονοστοιχείου έχουν εύρος από [0,1]. όπως έδειξαν τα αποτελέσματα η καλύτερη επαναληψημότητα όταν δειγματοληπτείται 3 κλίμακες ανά οκτάβα. Φαίνεται εκπληκτικό το γεγονός ότι η επαναληψημότητα δεν βελτιώνει καθώς δειγματοληπτούνται περισσότερες κλίμακες. Ο λόγος αυτόν των αποτελεσμάτων είναι ότι ανακαλύπτονται περισσότερα τοπικά μέγιστα, αλλά αυτά τα μέγιστα δεν είναι κατά μέσο όρο ποιο σταθερά και για αυτόν τον λόγο είναι ποιο απίθανο να ανιχνευθούν στην μετασχηματισμένη εικόνα. Ο αριθμός των σημείων κλειδιών αυξάνετε με την αύξηση της δειγματοληψίας των κλιμάκων. Καθώς η επιτυχία της αναγνώρισης αντικείμενων συνήθως εξαρτάται από την ποσότητα των σωστά ταιριασμένων σημείων κλειδιών, καθώς είναι αντίθετο με το ποσοστό του σωστού ταιριάσματος, για πολλές εφαρμογές είναι βέλτιστη η χρήση μεγαλύτερου αριθμού κλιμακωτών δειγμάτων. Παρόλα αυτά, το υπολογιστικό κόστος αυξάνετε και αυτό με των παραπάνω αριθμό[21].

Συνοψίζοντας, τα πειράματα έδειξαν ότι η λειτουργία κλιμακωτού χώρου διαφοράς – της - Gaussian έχει μεγάλο αριθμό από μέγιστα τα οποία θα είχε μεγάλο κόστος η ανίχνευση τους. Ευτυχώς, επιτυγχάνετε η εύρεση των ποιο σταθερών και χρήσιμων υποσυνόλων ακόμα και με τραχεία δειγματοληψία από κλίμακες, Όπως προσδιορίζετε η συχνότητα της δειγματοληψίας ανά οκτάβα του κλιμακωτού χώρου, έτσι πρέπει να καθορίσουμε την συχνότητα της δειγματοληψίας στην περιοχή της εικόνας που είναι συγγενική στην κλίμακα της εξομάλυνσης. Δεδομένης ότι το μέγιστο μπορεί να είναι αυθαίρετα πολύ κοντά μαζί, θα υπάρχει μια ανταλλαγή ανάμεσα στην συχνότητα δειγματοληψίας και στην αναλογία της αναγνώρισης. Φυσικά, εάν οι εικόνες είναι προ – λειασμένες πριν την ανίχνευση μεγίστων, μπορούν να αφαιρεθούν με επιτυχία οι υψηλότερες χωρικές συχνότητες. Για αυτόν τον λόγο, για να κάνουμε πλήρη χρήση των δεδομένων εισόδου, η εικόνα μπορεί να επεκταθεί για την δημιουργία περισσότερων δειγμάτων που θα αναπαρασταθούν στην αρχική εικόνα. Διπλασιάζοντας το μέγεθος της εικόνας εισόδου χρησιμοποιώντας γραμμική παρεμβολή για την κατασκευή του πρώτου επιπέδου της πυραμίδας. Καθώς η ισοδύναμη λειτουργία μπορεί αποτελεσματικά να χρησιμοποιηθεί σε σύνολα από αντισταθμισμένα φίλτρα υπό – εικονοστοιχείων στην

αρχική εικόνα, ο διπλασιασμός της εικόνας οδηγεί σε ποιο αποτελεσματική εφαρμογή. Θεωρώντας ότι η αρχική εικόνα έχει μια θολούρα τουλάχιστον ίση με 0,5 και για αυτόν τον λόγο η διπλάσια εικόνα έχει 1,0 σχετικά με την νέα διάταξη των εικονοστοιχείων. Αυτό σημαίνει ότι μια καλύτερη εξομάλυνση είναι απαραίτητη εκ των προτέρων για την δημιουργία της πρώτης οκτάβας του κλιμακωτού χώρου. Ο διπλασιασμός της εικόνας αυξάνει τον αριθμό των σταθερών σημείων κατά σχεδόν τέσσερις φορές, αλλά δεν παρέχει άλλες βελτιώσεις[21].

Αύου έχει βρεθεί ένα υποψήφιο σημείο κλειδί με την σύγκριση των εικονοστοιχείων με τα γειτονικά, το επόμενο βήμα είναι να αποδίδει λεπτομερή ταίριασμα στα γύρω δεδομένα για τοποθεσία, κλίμακα και αναλογία των κύριων καμπυλοτήτων. Αυτή η πληροφορία επιτρέπει την απόρριψη των σημείων που έχουν χαμηλή αντίθεση ή είναι ανεπαρκώς εντοπισμένη κατά μήκος μιας ακμής.

Η αρχική εφαρμογή αυτής της προσέγγισης άπλα βρίσκουν σημεία κλειδιά σε περιοχές και κλίμακα των κεντρικών δειγμάτων. Παρόλα αυτά, ο Brown έχει εφαρμόσει μια μέθοδο για την συναρμολόγηση μιας τρισδιάστατης τετραγωνικής λειτουργίας για τοπικά δείγματα για να καθορίσει την παρεμβολή της τοποθεσίας του μέγιστου, και τα αποτελέσματα του έδειξαν ότι παρείχε μεγάλη αναβάθμιση στο ταίριασμα και την σταθερότητα. Η προσέγγιση του χρησιμοποιεί την επέκταση του Taylor της λειτουργίας κλιμακωτού χώρου,  $D(x, y, \sigma)$ , μετατοπισμένα έτσι ώστε η προέλευση κάθε εικονοστοιχείου να είναι:

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

*Εικόνα 58: Η προσέγγιση με επέκταση του Taylor της λειτουργίας κλιμακωτού χώρου,  $D(x, y, \sigma)$*

όπου το  $D$  και τα παράγωγα του αξιολογούνται από το δείγμα  $\mathbf{x} = (x, y, \sigma)^T$  είναι το αντιστάθμισμα από κάθε σημείο. Η τοποθεσία του μέγιστου,  $\hat{\mathbf{x}}$ , προσδιορίζετε παίρνοντας τα παράγωγα αυτής της λειτουργίας με σεβασμό στο  $\mathbf{x}$  και θέτοντας το στο μηδέν, δίνοντας:

$$\hat{\mathbf{x}} = - \frac{\partial^2 D}{\partial \mathbf{x}^2}^{-1} \frac{\partial D}{\partial \mathbf{x}}$$

*Εικόνα 59: Η τοποθεσία του μέγιστου,  $\hat{\mathbf{x}}$*

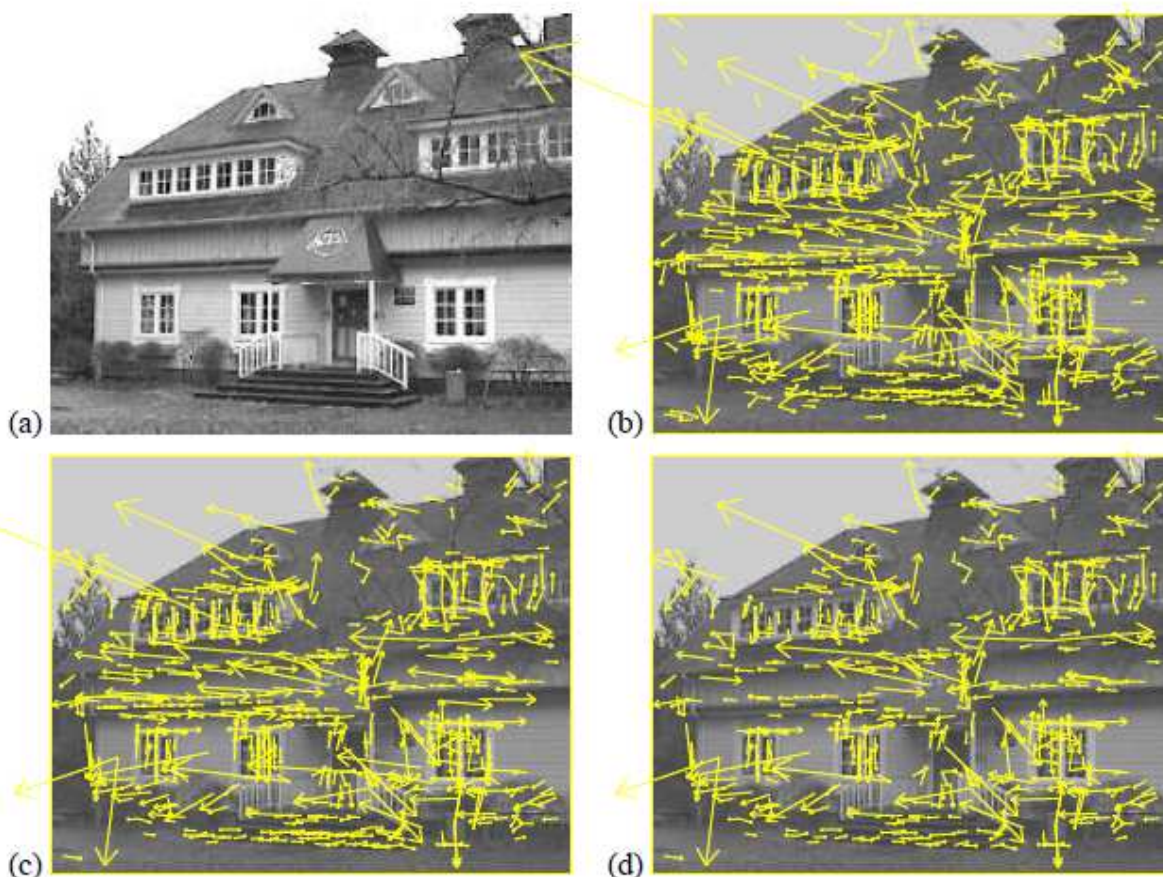
Όπως προτάθηκε από τον Brown, η Brown και τα παράγωγα του  $D$  είναι προσεγγισμένα με την χρήση διαφόρων μεταξύ γειτονικών σημείων. Το γραμμικό  $3 \times 3$  σύστημα που απορρέει μπορεί να επιλυθεί με το ελάχιστο κόστος. Αν το αντιστάθμισμα  $\hat{\mathbf{x}}$  είναι μεγαλύτερο από 0,5 σε κάθε διάσταση, τότε σημαίνει ότι το μέγιστο βρίσκετε πιο κοντά σε ένα διαφορετικό σημείο δείγματος. Σε αυτήν την περίπτωση, το σημείο δειγματοληψίας αλλάζει και η παρεμβολή που δημιουργείτε σε αυτό το σημείο. Το τελικό αντιστάθμισμα  $\hat{\mathbf{x}}$  προστίθεται στην τοποθεσία της τοποθεσίας του δείγματος για να λάβει την εκτίμηση της παρεμβολής για την τοποθεσία του μέγιστου[21].

Η τιμή της λειτουργίας στο μέγιστο,  $D(\hat{\mathbf{x}})$ , είναι χρήσιμο για την απόρριψη μη σταθερών μεγίστων με χαμηλή αντίθεση. Αυτό μπορεί να γίνει αντιληπτό από την αντικατάσταση των παραπάνω εξισώσεων με την παρακάτω:

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

Εικόνα 60: Η τιμή της λειτουργίας στο μέγιστο,  $D(\hat{x})$

όλες οι τιμές του  $|D(\hat{x})|$  που είναι μικρότερες από 0,03 απορρίπτονται.



Εικόνα 61: Τα αποτελέσματα της επιλογής σημείων κλειδιών σε μια εικόνα.

Η εικόνα 61 δείχνει τα αποτελέσματα της επιλογής σημείων κλειδιών σε μια εικόνα. Για την αποφυγή μεγάλης συμφόρησης σημείων, η ανάλυση της εικόνας είναι 233 επί 189 εικονοστοιχεία και τα σημεία κλειδιά φαίνονται ως πίνακες που δίνουν τις συντεταγμένες, την κλίμακα και τον προσανατολισμό κάθε σημείου κλειδιού. Στο κομμάτι (a) είναι η αρχική εικόνα, η οποία εμφανίζεται με χαμηλότερη αντίθεση πίσω από τις μεταγενέστερες εικόνες. Στο κομμάτι (b) εμφανίζονται τα 832 σημεία κλειδιά σε όλα τα ανιχνευμένα μέγιστα και ελάχιστα της λειτουργίας διαφοράς – της - Gaussian, καθώς το κομμάτι (c) δείχνει τα 729 σημεία κλειδιά που απομένουν με την αφαίρεση αυτών με τιμή  $|D(\hat{x})|$  μικρότερη από 0,03. Στο κομμάτι (d) θα εξηγηθεί παρακάτω[21].

Για σταθερότητα των αποτελεσμάτων, δεν είναι επαρκές να απορρίπτονται σημεία κλειδιά με χαμηλή αντίθεση. Η διαφοράς – της - Gaussian λειτουργίας έχει μεγάλη απόκριση ανάμεσα στις ακμές, ακόμα και αν η τοποθεσία κατά μήκος των ακμών αποφασίζεται με αδυναμία και για αυτόν τον λόγο είναι ασταθής σε μικρό μέρος θορύβου. Μια αδυναμία καθορισμένη κορυφή στην διαφορά – της - Gaussian λειτουργίας θα

έχει μεγάλη κύρια καμπυλότητα κατά μήκος της ακμής άλλα ένα μικρο κομμάτι στην κάθετη κατεύθυνση. Η κύρια καμπυλότητα μπορεί να υπολογιστεί από ένα 2x2 Hessian πίνακα ,  $\mathbf{H}$ , που υπολογίζεται στην τοποθεσία και την κλίμακα του σημείου κλειδιού:

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

Εικόνα 62:

Υπολογισμός κύριας  
καμπυλότητας από ένα  
2x2 Hessian πίνακα ,  $\mathbf{H}$

Τα παράγωγα υπολογίζονται με την διαφορά που δημιουργείτε με τα γειτονικά δείγματα. Οι ιδιοτιμές του  $\mathbf{H}$  είναι ανάλογες με την κύρια καμπυλότητα του  $D$ . Δανειζόμενοι από την προσέγγιση από τους Harris και Stephens (1988), μπορούμε να αποφύγουμε ρητά τον υπολογισμό των ιδιοτιμών, καθώς τον ενδιαφέρον έχουν μόνο οι αναλογίες τους. Ας θέσουμε ότι η ιδιοτιμία με τη μεγαλύτερης σημασίας και  $\beta$  να είναι η μικρότερη. Τότε μπορούμε να υπολογίσουμε το σύνολο των ιδιοτιμών από το ίχνος του  $\mathbf{H}$  και τα προϊόντα του από την ορίζουσα[21]:

$$\text{Tr}(\mathbf{H}) = D_{xx} + D_{yy} = \alpha + \beta,$$

$$\text{Det}(\mathbf{H}) = D_{xx}D_{yy} - (D_{xy})^2 = \alpha\beta.$$

Εικόνα 63: Υπολογισμός του συνόλου  
των ιδιοτιμών από το ίχνος του  $\mathbf{H}$  και τα  
προϊόντα του από την ορίζουσα

Στην χειρότερη περίπτωση η ορίζουσα είναι αρνητική, οι καμπυλότητες έχουν διαφορετικά σημεία έτσι ώστε το σημείο να απορρίπτεται σαν μη μέγιστο. Έστω ότι  $r$  είναι η αναλογία ανάμεσα στην μεγαλύτερης σημασίας ιδιοτελή και την μικρότερη, έτσι ώστε  $\alpha = r\beta$ . Τότε :

$$\frac{\text{Tr}(\mathbf{H})^2}{\text{Det}(\mathbf{H})} = \frac{(\alpha + \beta)^2}{\alpha\beta} = \frac{(r\beta + \beta)^2}{r\beta^2} = \frac{(r + 1)^2}{r},$$

Εικόνα 64: Υπολογισμός αναλογίας ανάμεσα  
στην μεγαλύτερης σημασίας ιδιοτιμή και την  
μικρότερη σημασίας

το οποίο εξαρτάται μόνο από την αναλογία των ιδιοτιμών παρά από την μεμονωμένη τιμή τις κάθε μιας. Η ποσότητα  $(r+1)^2/r$  είναι στο ελάχιστο όταν δύο ιδιοτιμές είναι ίσες και αυξάνετε με το  $r$ . Για αυτόν τον λόγο, ο έλεγχος αυτής της αναλογίας των κύριων καμπυλοτήτων είναι μικρότερη από ένα όριο,  $r$ , πρέπει να ελέγχετε :

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

Εικόνα 65: Ο έλεγχος  
της αναλογίας των  
κύριων καμπυλοτήτων

Αυτό μπορεί να υπολογιστεί πολύ εύκολα, με λιγότερους από είκοσι πράξεις με floating point αριθμούς που χρειάζονται για να ελεγχθεί κάθε σημείο κλειδί. Στην εικόνα 61 (d) είναι εμφανής αυτή η λειτουργία όπου το  $r = 10$ .

Με την ανάθεση ενός συνεπή προσανατολισμού σε κάθε σημεία κλειδιά βασιζόμενοι στις ιδιότητες της εικόνας, ο περιγραφέας σημείων κλειδιών μπορεί να παρουσιαστεί σχετικός με τον προσανατολισμό και να πετύχει την ιδιότητα του αναλλοίωτου αποτελέσματος στην περιστροφή της εικόνας. Αυτή η προσέγγιση έρχεται σε αντίθεση με τους αναλλοίωτους περιγραφείς προσανατολισμού του Schmid και Mohr (1997), στους οποίους κάθε ιδιότητα της εικόνας είναι βασισμένη σε περιστροφικά αναλλοίωτο μέτρο. Το μειονέκτημα αυτής της προσέγγισης είναι ότι περιορίζει τους περιγραφείς που μπορούν να χρησιμοποιηθούν και ότι απορρίπτει πληροφορία από την εικόνα με το να μην απαιτεί όλα τα μέτρα να βασίζονται σε μία συνεχή περιστροφή[21].

Η κλίμακα των σημείων κλειδιών χρησιμοποιήτε για την επιλογή της Gaussian εξομαλυσμένης εικόνας,  $L$ , με την κοντινότερη κλίμακα έτσι ώστε όλοι οι υπολογισμοί να πραγματοποιούνται σε κλιμακωτά – αμετάβλητο τρόπο. Για κάθε δείγμα εικόνας,  $L(x,y)$ , σε αυτήν την κλίμακα η κλίση μεγέθους,  $m(x,y)$ , και ο προσανατολισμός,  $\theta(x,y)$ , προϋπολογίζετε με την χρήση των διαφορών των εικονοστοιχείων:

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

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

*Εικόνα 66: Υπολογισμός κλίμακας κλίσης μεγέθους,  $m(x,y)$ , και προσανατολισμού,  $\theta(x,y)$*

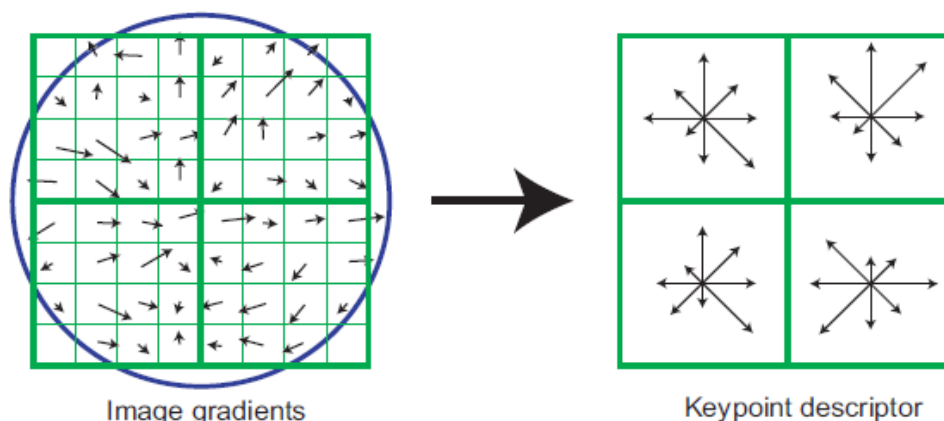
Ένα ιστόγραμμα προσανατολισμού παράγεται από τις κλίσεις του προσανατολισμού των δειγμάτων μέσα σε μία περιοχή γύρω από ένα σημείο κλειδί. Το ιστόγραμμα προσανατολισμού έχει 36 κάδους που καλύπτουν πεδίο προσανατολισμού 360 μοιρών. Κάθε δείγμα που προστίθεται στο ιστόγραμμα σταθμίζετε από την κλίση μεγέθους και από ένα Gaussian – σταθμισμένο κυκλικό παράθυρο με  $\sigma$  1,5 φορές μεγαλύτερο από την κλίμακα του σημείου κλειδιού.

Κορυφές στο ιστόγραμμα προσανατολισμού αντιστοιχούν σε δεσπόζουσες κατευθύνσεις σε τοπικές κλίσεις. Η μεγαλύτερη κορυφή στο ιστόγραμμα ανιχνεύεται, και τότε κάθε άλλη τοπική κορυφή που βρίσκετε στο 80% των μεγαλύτερων κορυφών χρησιμοποιήτε για την δημιουργία σημείων κλειδιών με αυτόν τον προσανατολισμό. Για αυτόν τον λόγο, για πολλές τοποθεσίες με πολλαπλές κορυφές παρόμοιου μεγέθους, θα υπάρχουν πολλαπλά σημεία κλειδιά που δημιουργούνται στο ίδιο σημείο και κλίμακα άλλα σε διαφορετικό προσανατολισμό. Σε μόνο 15% αυτών των σημείων εκχωρούνται πολλαπλοί προσανατολισμοί, αλλά αυτές συμβάλλουν σημαντικά στην σταθερότητα της ταύτισης. Τελικά, μια παραβολή τοποθετήτε σε τρεις τιμές του ιστογράμματος κοντύτερα σε κάθε κορυφή για να παρεισάγουν την τοποθεσία της κορυφής για καλύτερη ακρίβεια[21].

Η προηγούμενες λειτουργίες ανέθεταν μια τοποθεσία στην εικόνα, κλίμακα, και προσανατολισμό σε κάθε σημείο κλειδί. Αυτές οι παράμετροι επιβάλλον ένα επαναλαμβανόμενο δισδιάστατο σύστημα συντεταγμένων το οποίο περιγράφει τις τοπικές περιοχές της εικόνας, και για αυτόν τον λόγο παρήχαν μη μεταβλητότητα σε αυτές τις παραμέτρους. Το επόμενο βήμα είναι ο υπολογισμός ενός περιγραφέα για τοπικές περιοχές μιας εικόνας οι οποίοι είναι υψηλά διακριτοί άλλα και αμετάβλητοι όσο είναι δυνατόν σε υπολειπόμενες παραλλαγές, όπως μια αλλαγή στον φωτισμό ή στην τρισδιάστατη άποψη. Μια προφανή προσέγγιση θα ήταν να δειγματοληπτήσουμε τις εντάσεις τις εικόνας γύρω από σημεία κλειδιά στην κατάλληλη κλίμακα, και να ταιριάζουμε αυτές χρησιμοποιώντας ένα μετρό κανονικοποιημένης συσχέτισης. Παρόλα αυτά, απλές συσχετίσεις από κομμάτια εικόνων είναι πολύ ευαίσθητα στις αλλαγές που προκαλούν χαμένες εγγραφές από δείγματα, όπως συγγενικότητα ή τρισδιάστατες αλλαγές όψης είτε μη – άκαμπτες παραμορφώσεις. Μια καλύτερη προσέγγιση παρουσιάστηκε από τους Edelman, Intrator, and Poggio (1997). Η προτεινόμενη αναπαράσταση βασίστηκε σε ένα μοντέλο της βιολογικής όρασης, συγκεκριμένα σε πολύπλοκους νευρώνες στον κύριο οπτικό λοβό. Αυτοί οι πολύπλοκοι νευρώνες αντιστοιχούν σε μια κλίση ενός συγκεκριμένου προσανατολισμού και χωρικής συχνότητας, αλλά η θέση της κλίσης στον



αμφιβληστροειδή επιτρέπει την αλλαγή από ένα μικρό δεκτικό πεδίο παρά την ακριβή τοποθέτηση. Ο Edelman υπέθεσε ότι αυτή η λειτουργία αυτών των πολυπλοκότεων νευρώνων μπορούσε να επιτρέψει το ταίριασμα και την αναγνώριση των τρισδιάστατων αντικειμένων σε ένα εύρος οπτικών πλευρών. Πειραματίστηκαν με την χρήση λεπτομερών τρισδιάστατων υπολογιστικών μοντέλων από αντικείμενα και σχήματα ζώων τα οποία δείχνουν ότι το ταίριασμα των κλίσεων καθώς για μετατοπίσεις στις θέσεις τους έχουν σαν αποτέλεσμα σε καλύτερο διαχωρισμό κάτω από τρισδιάστατη περιστροφή. Η εφαρμογή του αλγορίθμου εμπνευσμένη από αυτήν την ιδέα επιτρέπει αλλαγή της θέσης χρησιμοποιώντας ένα διαφορετικό υπολογιστικό μηχανισμό[21].



Εικόνα 67: Ο υπολογισμός του περιγραφέα σημείων κλειδιών

Στην εικόνα 67 απεικονίζετε ο υπολογισμός του περιγραφέα σημείων κλειδιών. Πρώτα οι κλίσεις των μεγεθών της εικόνας και οι προσανατολισμοί δειγματοληπτούνται γύρω από μια τοποθεσία σημείου κλειδιού, χρησιμοποιώντας την κλίμακα του σημείου κλειδιού για την επιλογή του επιπέδου της Gaussian θολούρας για την εικόνα. Για να επιτευχθεί το αναλλοίωτο του προσανατολισμού, οι συντεταγμένες του περιγραφέα και οι κλίσεις των προσανατολισμών περιστρέφονται ανάλογα με τον προσανατολισμό των σημείων κλειδιών. Για αποδοτικότητα, οι κλίσεις έχουν προϋπολογιστεί για όλα τα επίπεδα όπως αναφέρθηκε και παραπάνω. Αυτές απεικονίζονται με μικρά βέλη σε κάθε δείγμα στο αριστερό μέρος της εικόνας 67.

Η λειτουργία στάθμισης Gaussian που είναι ίση με το μισό πλάτος του παραθύρου του περιγραφέα χρησιμοποιήστε για την ανάθεση ενός βάρους στο μέγεθος κάθε δείγματος. Αυτό αναπαριστάτε με ένα κυκλικό παράθυρο στο αριστερό μέρος της εικόνας 67, παρόλα αυτά προφανώς το βάρος κατανέμετε ομοιόμορφα. Ο σκοπός αυτού του Gaussian παραθύρου είναι η αποφυγή των ξαφνικών αλλαγών στον περιγραφέα με μικρές αλλαγές στην θέση του παραθύρου, και να δώσει μικρότερη έμφαση στις κλίσεις που είναι μακριά από το κέντρο του περιγραφέα, καθώς είναι η λιγότερο επηρεασμένες από τα λάθη των εγγραφών[21].

Ο περιγραφέας των σημείων κλειδιών φαίνεται στην δεξιά πλευρά της εικόνας 67. Απαιτεί για σημαντική αλλαγή στις θέσεις των κλίσεων με την δημιουργία ιστογραμμάτων πάνω σε 4x4 περιοχές δειγμάτων. Η εικόνα δείχνει οκτώ κατευθύνσεις για κάθε προσανατολισμό ιστογράμματος, με το μέγεθος κάθε βέλους να αντιστοιχεί στο μέγεθος της καταχώρισης του ιστογράμματος. Ένα δείγμα κλίσης στα αριστερά μπορεί να μετακινηθεί σε πάνω από τέσσερις θέσεις δειγμάτων καθώς ακόμα συμβάλλει στο ίδιο ιστογράμμα στα δεξιά, έτσι επιτυγχάνει τον στόχο να επιτρέψει μεγαλύτερες τοπικές μετακινήσεις τοποθεσιών.

Είναι σημαντικό να αποφευχθούν όλες οι επιρροές των συνόρων στις οποίες ο περιγραφέας αλλάζει απότομα καθώς ένα δείγμα αλλάζει ομαλά από το να βρίσκετε μέσα σε ένα ιστογράμμα σε ένα άλλο ή από έναν προσανατολισμό σε έναν άλλο. Για αυτόν τον λόγο, χρησιμοποιήστε η τριγωνική παρεμβολή για να διανέμει την τιμή κάθε κλίσης δείγματος σε παρακείμενους κάδους ιστογράμματος. Με άλλα λόγια, κάθε είσοδος σε έναν κάδο πολλαπλασιάζεται με ένα βάρος 1-d για κάθε διάσταση, όπου d είναι η απόσταση από το δείγμα στην κεντρική τιμή του κάδου όπως μετράτε σε μονάδες απόστασης κάδου ιστογράμματος[21].

Ο περιγραφέας σχηματίζεται από έναν πίνακα που περιέχει τις τιμές όλων των εισόδων προσανατολισμού

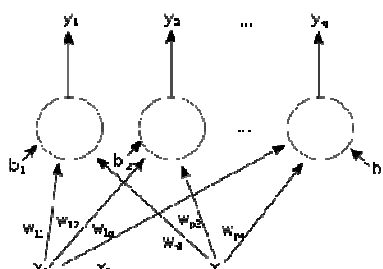
ενός ιστογράμματος., που αντιστοιχεί στα μήκη των βελών που περιέχονται στο δεξί μέρος τις εικόνας 67. Η εικόνα απεικονίζει 2x2 πίνακα ιστογράμματος προσανατολισμού, καθώς στον πραγματικό αλγόριθμο χρησιμοποιήθηκε ένας 4x4 πίνακα ιστογράμματος με οκτώ κάδους προσανατολισμού στο κάθε ένα. Για αυτόν τον λόγο τα αποτελέσματα του αλγορίθμου είναι  $4 \times 4 \times 8 = 128$  στοιχειά πίνακα χαρακτηριστικών για κάθε σημείο κλειδί.

Τελικά ο πίνακας χαρακτηριστικών τροποποιείτε για να μειώσει τα αποτελέσματα της αλλαγής στον φωτισμό. Πρώτα, ο πίνακας κανονικοποιήτε σε μονάδα μήκους. Μια αλλαγή στην αντίθεση της εικόνας στην οποία κάθε τιμή εικονοστοιχείου πολλαπλασιάζετε με μια σταθερά θα πολλαπλασιάσει κλίσεις με την ίδια αντίθεση, έτσι η αλλαγή στην αντίθεση θα ακυρωθεί από ένα κανονικοποιημένο πίνακα. Μια αλλαγή στην φωτεινότητα στην οποία μια σταθερά προστίθεται ,σε κάθε εικονοστοιχείο της εικόνας δεν θα επηρεάσει τις τιμές των κλίσεων, καθώς υπολογίζονται από τις διαφορές των εικονοστοιχείων. Για αυτόν τον λόγο, ο περιγραφέας είναι αμετάβλητος σε αλλαγές στην φωτεινότητα. Όμως, μη γραμμικές αλλαγές στην φωτεινότητα μπορούν να συμβούν από τον κορεσμό της κάμερας ή από αλλαγές φωτεινότητας που μπορεί να επηρεάσουν τις τρισδιάστατες επιφάνειες με διαφοροποίηση των προσανατολισμών με διαφορετικά ποσά. Αυτά τα φαινόμενα μπορούν να προκαλέσουν μεγάλες αλλαγές σε συγγενικά μεγέθη κάποιων κλίσεων, άλλα είναι λιγότερο πιθανόν να επηρεάσουν τους προσανατολισμούς των κλίσεων. Για αυτόν τον λόγο, ελαττώνετε η επιρροή από τα μεγάλα μεγέθη των κλίσεων με την εφαρμογή κατωτάτου κατωφλίου οι τιμές του πίνακα χαρακτηριστικών δεν μπορούν να είναι μεγαλύτερες από 0,2., και τότε κανονικοποιούμε ξανά την μονάδα του μήκους. Αυτό σημαίνει ότι το ταίριασμα των μεγεθών για μεγάλες κλίσεις δεν είναι ποια σημαντικό, και η διανομή των προσανατολισμών έχει μεγαλύτερη έμφαση[21].

Ο αλγόριθμος αυτός ενδείκνυται και για αναγνώριση αντικειμένων όπου εκτελείται ταίριασμα ανάμεσα σε κάθε σημείο κλειδιού ανεξάρτητα με μια βάση δεδομένων από γνωστά σημεία κλειδιά που έχουν εξαχθεί από εικόνες εκπαίδευσης. Ο κλασικός τρόπος ταίριασματος με έναν αλγόριθμο κοντινότερων γειτόνων μπορεί να παράγει ζεύγη τα οποία να είναι αναληθή και αυτό συμβαίνει σε ασαφή χαρακτηριστικά ή σε χαρακτηριστικά που παράγονται από το περιβάλλον. Έτσι χρησιμοποιήθηκε μια διαφορετική προσέγγιση από αυτήν του κλασικού αλγορίθμου. Μπορούμε να εξάγουμε τα πραγματικά αντικείμενα τις εικόνας με στερεοσκοπική μέθοδο όπως αυτή αναλύεται παραπάνω. Τέλος, πλέον δεν χρειαζόμαστε την χρήση του αλγορίθμου κοντινότερων γειτόνων για τον λόγο ότι στο σύστημα αυτό ο διαχωρισμός των χαρακτηριστικών γίνεται μέσω ενός νευρωνικού δικτύου όπως αναφέρετε παρακάτω[21].

## 2.4.4 Τεχνητά νευρωνικά δίκτυα

Τα τεχνητά νευρωνικά δίκτυα (ANN), είναι ένα μαθηματικό μοντέλο ή υπολογιστικό μοντέλο το οποίο είναι εμπνευσμένο από την δομή και την λειτουργία των βιολογικών νευρωνικών δικτύων. Ένα νευρωνικό δίκτυο αποτελείται από ένα εσωτερικά συνδεδεμένη ομάδα από τεχνητούς νευρώνες, και επεξεργάζεται πληροφορίες χρησιμοποιώντας μια συνδεδεμένη προσέγγιση επεξεργασίας. Στις περισσότερες περιπτώσεις ένα ANN είναι ένα προσαρμοστικό σύστημα που αλλάζει την δομή τους βασίζόμενο σε εσωτερικές ή εξωτερικές πληροφορίες οι οποίες ρέουν μέσα στο δίκτυο κατά την διάρκεια της φάσης εκπαίδευσης. Τα μοντέρνα νευρωνικά είναι μη – γραμμικά στατιστικά μοντέλα δεδομένων. Συχνά χρησιμοποιούνται για να μοντελοποιήσουν πολύπλοκες σχέσεις ανάμεσα σε εισόδους και εξόδους ή για την εύρεση προτύπων στα δεδομένα[22].

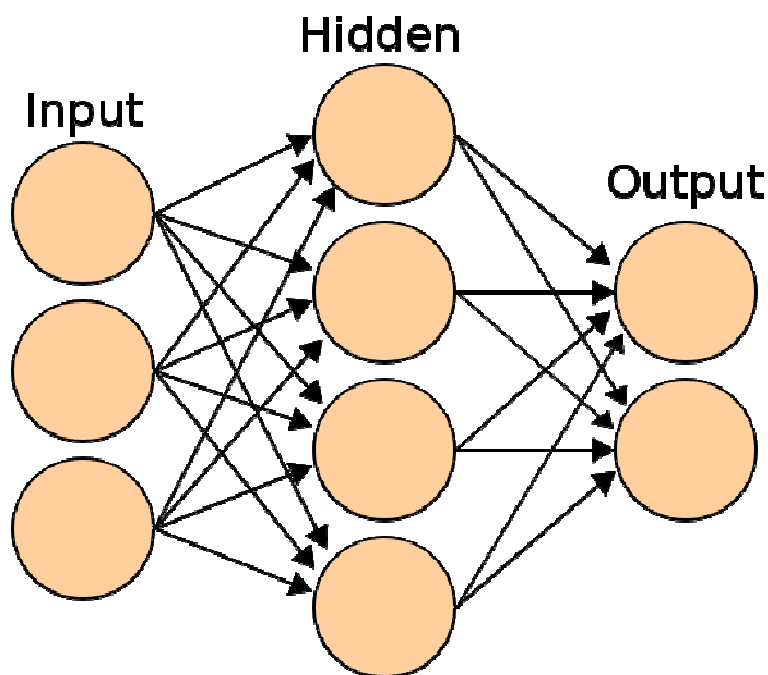


Εικόνα 68: Τεχνητό νευρωνικό δίκτυο ενός επιπέδου

Η αρχική έμπνευση για τον όρο *Τεχνητό Νευρωνικό Δίκτυο* ήρθε για εξέταση ενός κεντρικό νευρωνικό σύστημα και τους νευρώνες, τους άξονες, οι δενδίτες, και οι συνάψεις, οι οποίες αποτελούν τις μονάδες επεξεργασίας ενός βιολογικού νευρωνικού δικτύου, το οποίο ερευνάτε από την νευρολογία. Εάν ένα τεχνητό νευρωνικό δίκτυο έχουμε, απλούς τεχνητούς κόμβους, που αποκαλούνται ευρέως ως “νευρώνες”, “υπολογιστικά στοιχεία” ή “μονάδες”, που είναι συνδεδεμένες μεταξύ τους για να σχηματίσουν ένα δίκτυο από κόμβους που μιμούνται τα βιολογικά νευρωνικά δίκτυα και από αυτό προήλθε και ο όρος “τεχνητά νευρωνικά δίκτυα”[22].

Επειδή η νευρολογία έχει πολλές αναπάντητες ερωτήσεις, και αφού υπάρχουν πολλά επίπεδα αφαιρετικότητας και για αυτόν τον λόγο πολλοί τρόποι για να λάβουμε έμπνευση από τον εγκέφαλο, δεν υπάρχει κανένας τοπικός ορισμός για το τι είναι ένα τεχνητό νευρωνικό δίκτυο. Γενικά, περιέχει ένα δίκτυο από απλά υπολογιστικά στοιχεία που εκθέτουν πολύπλοκη καθολική συμπεριφορά που καθορίζετε από ενώσεις ανάμεσα στα στοιχεία υπολογισμού και στα στοιχεία παραμέτρων. Καθώς ένα τεχνητό νευρωνικό δίκτυο δεν χρειάζεται να είναι προσαρμοστικό, η πρακτική χρήση του προέρχεται από την ιδιότητα του αλγορίθμου να αλλάζει την δύναμη των συνδέσεων του στο δίκτυο και να παράγει μια ποθητή ροή δεδομένων[22].

Αυτά τα δίκτυα είναι επίσης όμοια με τα βιολογικά νευρωνικά δίκτυα σε μια έννοια στις οποίες η λειτουργίες εκτελούνται συλλογικά και παράλληλα σε μονάδες, παρά την καθαρή περιγραφή των δευτερευόντων εργασιών στις οποίες εκχωρούνται πολλές μονάδες. Σήμερα, ο όρος τεχνητά νευρωνικά δίκτυα τείνει να αναφέρετε κυρίως σε μοντέλα νευρωνικών δικτύων που εφαρμόζονται στην στατιστική, στην γνωστική ψυχολογία και την τεχνητή νοημοσύνη. Τα μοντέλα νευρωνικών δικτύων σχεδιάζονται για την άμιλλα του κεντρικού νευρικού συστήματος και θεωρητικά είναι ένα υπό κεφάλαιο της θεωρητικής νευρολογίας και υπολογιστικής νευρολογίας[22].



Εικόνα 69: Κλασική μορφή τεχνητού νευρωνικού δικτύου

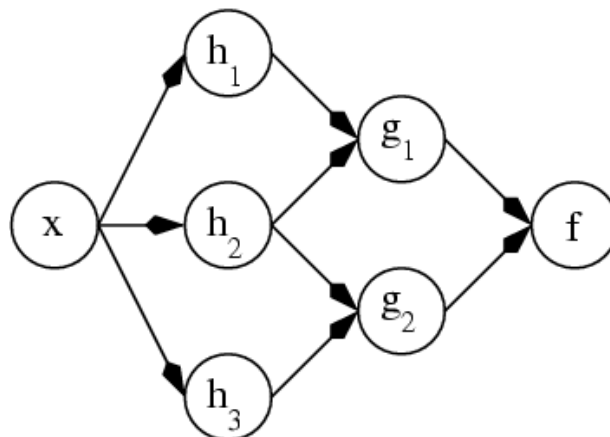
Σε μοντέρνες υλοποιήσεις λογισμικού των τεχνητών νευρωνικών δικτύων, η προσέγγιση που εμπνεύστηκε από την βιολογία έχει εγκαταλειφθεί και μια πιο πρακτική προσέγγιση βασισμένη στην στατιστική και στην επεξεργασία σημάτων. Σε μερικά από αυτά τα συστήματα, τα νευρωνικά δίκτυα ή κομμάτια αυτών χρησιμοποιούνται σαν κομμάτια ενός μεγαλύτερου συστήματος που περιέχει προσαρμοστικά και μη προσαρμοστικά στοιχεία. Καθώς μια πιο γενική προσέγγιση τέτοιων προσαρμοστικών συστημάτων είναι πιο κατάλληλα για να λύνουν προβλήματα του αληθινού κόσμου, γιατί δεν έχει να κάνει με τα

παραδοσιακά συνδεδεμένα μοντέλα τεχνητής νοημοσύνης. Αυτό που έχουν όμως κοινό είναι η αρχή της μη γραμμικότητας, διανομής, παράλληλης τοπικής επεξεργασίας και προσαρμογής[22].

Τα μοντέλα νευρωνικών δικτύων στην τεχνητή νοημοσύνη συνήθως αναφέρονται ως τεχνητά νευρωνικά δίκτυα τα οποία είναι απλά μαθηματικά μοντέλα τα οποία καθορίζουν μια λειτουργία  $f: X \rightarrow Y$  ή μια διανομή από το  $X$  ή και από το  $X$  και το  $Y$ , αλλά μερικές φορές τα μοντέλα είναι τόσο στενά συνδεδεμένα με ένα συγκεκριμένο αλγόριθμο μάθησης ή με ένα κανόνα μάθησης. Η κοινή χρήση της φράσης των μοντέλων τεχνητών νευρωνικών δικτύων σημαίνει τον ορισμό μιας κατηγορίας των εν λόγω λειτουργιών.

Η λέξη *δίκτυο* στον ορισμό “τεχνητό νευρωνικό δίκτυο” αναφέρετε στις εσωτερικές συνδέσεις ανάμεσα στους νευρώνες και στα διαφορετικά επίπεδα κάθε συστήματος. Η πιο βασική μορφή αποτελείται από τρία επίπεδα όπως είναι εμφανές και στην εικόνα 69. Το πρώτο επίπεδο είναι η νευρώνες εισόδου, που στέλνουν τα δεδομένα μέσα από συνάψεις στο δεύτερο επίπεδο, και μετά μέσα από ακόμα περισσότερες συνάψεις καταλήγουν στο τρίτο επίπεδο με του νευρώνες εξόδου. Πολλά πολύπλοκα συστήματα έχουν περισσότερα επίπεδα από νευρώνες με μερικά να έχουν αυξημένα επίπεδα νευρώνων εισόδου και νευρώνων εξόδου. Οι συνάψεις αποθηκεύουν τις παραμέτρους που ονομάζονται ως βάρη τα οποία χειράγουν τα δεδομένα και τους υπολογισμούς τους[22].

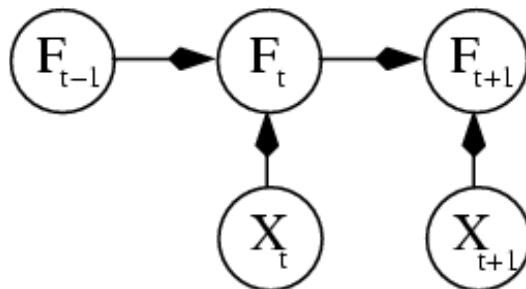
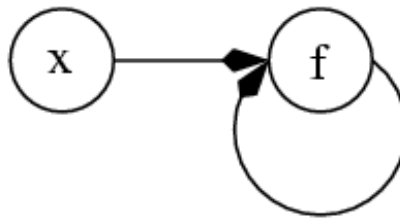
Τα επίπεδα σχηματίζουν ένα δίκτυο μέσω των μαθηματικών και των αλγορίθμων του συστήματος. Η λειτουργία του δικτύου  $f(x)$  καθορίζετε σαν μια σύνθεση από άλλες λειτουργίες  $g^i(x)$ , που μπορούν να οριστούν περαιτέρω με άλλες λειτουργίες. Αυτό μπορεί βολικά να παρασταθεί σαν μια δομή δικτύου, με βέλη τα οποία απεικονίζουν εξαρτήσεις μεταξύ μεταβλητών. Ένας τύπος σύνθεσης με ευρεία χρήση είναι το μη γραμμικά βεβαρημένο σύνολο, όπου  $f(x) = K(\sum_i w_i g_i(x))$ , όπου το  $K$  είναι μια προκαθορισμένη λειτουργία, όπως η υπερβολική εφαπτομένη. Θα ήταν βολικό για τα επόμενα να αναφερθεί η συλλογή από λειτουργίες  $g^i$  σαν ένα απλό πίνακα όπου  $g = (g_1, g_2, \dots, g_n)$ .



Εικόνα 70: Τεχνητό νευρωνικό δίκτυο γράφημα εξάρτησης

Αυτή η εικόνα απεικονίζει μια αποσύνθεση του  $f$ , με εξαρτήσεις ανάμεσα στις μεταβλητές που υποδεικνύονται από τα βέλη. Αυτές μπορούν να ερμηνευτούν με δυο τρόπους.

Ο πρώτος τρόπος είναι μια λειτουργική άποψη όπου: η είσοδος  $x$  μετατρέπεται σε έναν τρισδιάστατο πίνακα  $H$  όπου τότε μετατρέπεται σε έναν δισδιάστατο πίνακα  $G$ , που μετατρέπεται τελικά σε  $f$ . Αυτή η οπτική εμφανίζεται πιο συχνά σε περιεχόμενο βελτιστοποίησης. Η δεύτερη όψη είναι μια πιθανοτική όψη: η τυχαία μεταβλητή  $F=f(G)$  που εξαρτώνται από την τυχαία μεταβλητή  $G=g(H)$ , η οποία εξαρτάται από την  $H=h(X)$ , που εξαρτάται πάνω στην τυχαία μεταβλητή  $X$ . Αυτή η όψη είναι πιο συχνά εμφανίσιμη στα περιεχόμενα γραφικών μοντέλων. Οι δυο όψεις είναι σε μεγάλο βαθμό ισοδύναμες. Σε κάθε περίπτωση, για τις συγκεκριμένες αρχιτεκτονικές δικτύων, τα στοιχεία των μεμονωμένων επιπέδων είναι ανεξάρτητα το ένα από το άλλο. Αυτό φυσικά επιτρέπει ένα βαθμό παραλληλισμού στην εφαρμογή[22].



Εικόνα 71: Επαναλαμβανόμενο τεχνητό νευρωνικό δίκτυο γράφημα εξάρτησης

Δίκτυα όπως το προηγούμενο ονομάζονται κοινός τροφοδότησης προς τα εμπρός, γιατί οι γράφοι τους είναι κατευθυνόμενοι μη κυκλικοί γράφοι. Δίκτυα με κύκλους ονομάζονται επαναλαμβανόμενα. Τέτοια δίκτυα. Τέτοια δίκτυα απεικονίζονται στην παραπάνω εικόνα όπου το  $f$  εμφανίζεται να είναι εξαρτώμενο μόνον από τον εαυτό του. Ωστόσο, δεν εμφανίζετε μια συνεπαγόμενη χρονική εξάρτηση δεν εμφανίζετε[22]. Τα νευρωνικά δίκτυα εξαρτώνται από τρία βασικά κριτήρια:

- ⤴ εσωτερική σύνδεση ανάμεσα στα διαφορετικά επίπεδα των νευρώνων
- ⤴ διαδικασία εκπαίδευσης του τεχνητού νευρωνικού δικτύου
- ⤴ λειτουργία ενεργοποίησης στους νευρώνες

Η εσωτερική σύνδεση δείχνει την σχέση μεταξύ ενός απλού επιπέδου, με πολλαπλά επίπεδα παραμέτρων εισόδου και εξόδου των νευρώνων δείχνει μια σχέση του ενός με τους πολλούς. Αυτό σημαίνει ότι οι ίδιες εισοδοι μπορούν να εκτελέσουν πολλές εξόδους για διαφορετικά επίπεδα αρχιτεκτονικής.

Αυτό το οποίο συγκεντρώνει το μεγαλύτερο ενδιαφέρον στα νευρωνικά δίκτυα είναι η πιθανότητα της μάθησης. Δίνοντας του ένα συγκεκριμένο καθήκον να λύσει, και μια κλάση μιας λειτουργίας  $F$ , η μάθηση σημαίνει την χρήση του συνόλου των παρατηρήσεων για την εύρεση του  $f^* \in F$  το οποίο λύνει το έργο σε κάποια βέλτιστη έννοια. Αυτό συνεπάγεται με τον καθορισμό μιας λειτουργίας κόστους  $C: F \rightarrow \mathbb{R}$  τέτοιο ώστε, για την βέλτιστη λύση  $f^*, \forall f \in F, C(f^*) \leq C(f)$ . Η λειτουργία κόστους  $C$  είναι μια σημαντική έννοια στην μάθηση, σαν μέτρο του πόσο μακριά βρίσκεται η συγκεκριμένη λύση από την βέλτιστη λύση στο πρόβλημα που πρέπει να λυθεί. Οι αλγόριθμοι εκπαίδευσης ψάχνουν τον χώρο επίλυσης για την εύρεση μιας λειτουργίας που έχει το μικρότερο δυνατό κόστος[22].

Για εφαρμογές όπου η λύση είναι εξαρτώμενη από κάποια δεδομένα, το κόστος πρέπει να είναι απαραίτητα μια λειτουργία από παρατηρήσεις, διαφορετικά δεν θα μοντελοποιείται τίποτα συγγενικό με τα δεδομένα. Είναι συχνά καθορισμένο σαν στατιστικό στο οποίο μόνο προσεγγίσεις μπορούν να γίνουν. Ως απλό παράδειγμα, λαμβάνετε το πρόβλημα εύρεσης μοντέλου  $f$ , το οποίο ελαχιστοποιεί  $C = E[(f(x) - y)^2]$ , για ζεύγη δεδομένων  $(x, y)$  που λήφθηκαν από την κατανομή  $D$ . Σε πρακτικές καταστάσεις μπορούμε να έχουμε μόνον  $N$  δείγματα από  $D$ , για το παραπάνω παράδειγμα, θα μπορούσε να ελαχιστοποιηθεί μόνο

$\hat{C} = \frac{1}{N} \sum_{i=1}^N (f(x_i) - y_i)^2$ . Έτσι το κόστος θα ελαχιστοποιούταν από ένα δείγμα δεδομένων πάρα από ένα ολόκληρο σύνολο δεδομένων[22].

Όταν  $N \rightarrow \infty$  κάποια μορφή της μηχανικής μάθησης πρέπει να χρησιμοποιηθεί, όπου το κόστος είναι μερικώς ελαχιστοποιημένο καθώς κάθε νέο παράδειγμα ερευνάτε. Καθώς η μηχανική μάθηση χρησιμοποιήτε συνήθως όταν  $D$  είναι καθορισμένο, είναι το ποιο χρήσιμο σε αυτήν την περίπτωση όπου η διανομή αλλάζει αργά με το πέρασμα του χρόνου. Στην μέθοδο των νευρωνικών δικτύων, κάποιες μορφές από μηχανική μάθηση χρησιμοποιείτε συχνά για πεπερασμένη βάση δεδομένων.

Καθώς είναι δυνατόν να καθοριστεί ορισμένη αυθαιρεσία, για αυτό το κόστος της λειτουργίας, συχνά θα χρησιμοποιηθεί ένα συγκεκριμένο κόστος, είτε επειδή έχει επιθυμητές ιδιότητες ή επειδή ανακύπτει φθισικά από συγκεκριμένη διατύπωση του προβλήματος. Τελικά, το κόστος λειτουργίας θα εξαρτηθεί από την επιθυμητή εργασία. Μια επισκόπηση από τρεις κύριες κατηγορίες από εργασίες εκπαίδευσης εμφανίζονται παρακάτω και αυτές είναι η επιβλεπόμενη εκπαίδευση, μη επιβλεπόμενη εκπαίδευση και ενισχυμένη εκπαίδευση[22].

Στην επιβλεπόμενη εκπαίδευση, μας δίνετε ένα σύνολο από ζεύγη  $(x, y), x \in X, y \in Y$  και ο σκοπός είναι να βρεθεί μια λειτουργία  $f: X \rightarrow Y$  στην επιτρεπόμενη κλάση από λειτουργίες οι οποίες ταιριάζουν στο παράδειγμα. Με άλλα λόγια, επιθυμούμε το συμπεράνουμε την χαρτογράφηση που συνεπάγεται στα δεδομένα, το κόστος της λειτουργίας είναι συγγενικό με το λάθος ταίριασμα ανάμεσα στην χαρτογράφηση και τα δεδομένα και σιωπηρά περιέχει κύρια γνώση για το πεδίο ορισμού του προβλήματος.

Ένα κόστος το οποίο χρησιμοποιήτε συχνά είναι το ελάχιστο – τετραγωνικό σφάλμα, το οποίο προσπαθεί να ελαχιστοποιήσει τον μέσο τετραγωνισμένο λάθος ανάμεσα στις εξόδους του δικτύου,  $f(x)$ , και η τιμή του στόχου  $y$  ανάμεσα σε όλα τα ζεύγη των παραδειγμάτων. Όταν το ένα προσπαθεί να ελαχιστοποιηθεί το κόστος χρησιμοποιεί μια βαθμίδα καθόδου για κάθε κλάση των νευρωνικών δικτύων που ονομάζονται δίκτυα πολλαπλών επιπέδων, το άλλο αποκτά τον πολύ γνωστό αλγόριθμο διάδοσης προς τα πίσω για την εκπαίδευση των νευρωνικών δικτύων. Εργασίες οι οποίες εμπίπτουν στο παράδειγμα της επιβλεπόμενης εκπαίδευσης είναι η αναγνώριση προτύπων και οπισθοδρόμηση διαδικασίας. Το παράδειγμα της επιβλεπόμενης εκπαίδευσης είναι επίσης εφαρμόσιμο σε διαδοχικά δεδομένα. Αυτό μπορεί να αναλογιστεί ως να μαθαίνει με έναν “δάσκαλο”, στην μορφή της λειτουργίας που παρέχει συνεχή ανατροφοδότηση στην ποιότητα των αποτελεσμάτων που αποκτήθηκε μέχρι εκεί. Βασικά η επιβλεπόμενη εκπαίδευση ταξινομείται σε δυο τύπους. Αυτές είναι σφάλμα σύνδεσης καθόδου βαθμίδας, που έχει δυο μεθόδους που ελάχιστο τετραγωνικό λάθος και διάδοσης προς τα πίσω, και στοχαστική[22].

Η εκπαίδευση ενός μοντέλου νευρωνικού δικτύου κατ' ουσίαν σημαίνει ότι επιλέγετε ένα μοντέλο από το σύνολο των επιτρεπόμενων μοντέλων που ελαχιστοποιεί το κόστος των κριτηρίων. Υπάρχουν πολυάριθμοι αλγόριθμοι που είναι διαθέσιμη για την εκπαίδευση ενός μοντέλου νευρωνικού δικτύου, τα περισσότερα από αυτά μπορούν να εμφανιστούν ως μια απλή εφαρμογή της θεωρίας της βελτιστοποίησης και της στατιστικής εκτίμησης. Πρόσφατες εξελίξεις σε αυτόν τον τομέα χρησιμοποιούν την βελτιστοποίηση του νέφους σωματιδίων και άλλες τεχνικές νοημοσύνης νεφών. Περισσότεροι από αυτούς τους αλγορίθμους που χρησιμοποιούνται στην εκπαίδευση τεχνητών νευρωνικών δικτύων απασχολούν κάποια μορφή μεθόδου βαθμίδας. Αυτό είναι δυνατό με το να πάρουμε άπλα τα παράγωγα των λειτουργιών κόστους στις παραμέτρους του δικτύου και τότε να αλλάζει αυτές τις παραμέτρους σε μια κατεύθυνση συγγενική με την βαθμίδα.

Εξελικτικές μέθοδοι, προσομοίωση ανάκτησης, μέθοδοι μεγιστοποίησης της προσδοκίας και μη παραμετρικές μέθοδοι συχνά χρησιμοποιούνται για την εκπαίδευση των νευρωνικών δικτύων. Η χρονικά αντιληπτική μάθηση βασίζετε στην εύρεση χρονικών σχέσεων σε σήματα αισθητήρων. Στο περιβάλλον, τα στατιστικά σημαντικότερες χρονικές συσχέτισης μπορούν να βρεθούν με την παρακολούθηση των στιγμών άφιξης των σημάτων. Και αυτό γίνεται σε αντιληπτικά δίκτυα[22].

### 3. Σχέδιο δράσης για την εκπόνηση της πτυχιακής εργασίας

Στο κεφάλαιο αυτό περιγράφονται οι λόγοι για την χρήση των παραπάνω αλγορίθμων, θεωρειών και μοντέλων. Οι αλγόριθμοι αυτοί δημιουργήθηκαν με σκοπό την επίλυση κάποιων προβλημάτων. Με την χρήση αυτή τα σημερινά συστήματα και λογισμικά μπορούν να εκτελέσουν εξαιρετικά πολύπλοκες εφαρμογές που για εμάς φάνταζαν απίστευτες πριν από λίγο καιρό. Αυτό ήταν και ένα από τα κίνητρα για την κατασκευή του συστήματος που περιγράφετε σε αυτήν την εργασία.

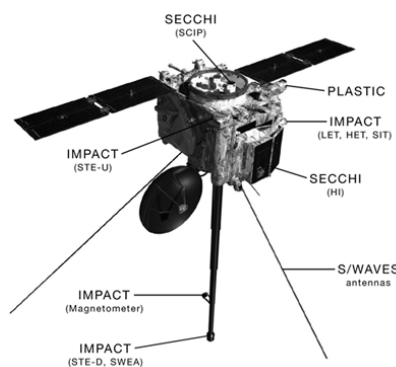
Ο ορισμός τεχνολογίες αιχμής (state-of-the-art) ακούγεται για ότι ποιο νέο και πρωτοποριακό. Αυτές οι τεχνολογίες είναι αυτές που φτιάχνουν καθημερινά νέα όρια για την επιστήμη και την εφαρμογή της. Σήμερα αυτή η τεχνολογία εφαρμόζονται σε συστήματα ρομποτικής, ναυπηγικής, αεροναυπηγικής και οποιασδήποτε κατασκευής ξεφεύγει από έναν παραδοσιακό τρόπο λειτουργίας. Όπως έχει δείξει και η ιστορία πολλές από αυτές τις εφαρμογές που πριν κάμποσα χρόνια ήταν πρωτοποριακές τώρα έχουν εφαρμοστεί σε πολλές πτυχές της καθημερινής μας ζωής κάνοντας την ευκολότερη και καλύτερη.

Η αναζήτηση αυτών των μερικών εφαρμογών γίνεται για να περιγράψει ο τρόπος σκέψης και χρήσης των παραπάνω αλγορίθμων, θεωρειών και μοντέλων. Παρακάτω θα αναδειχθούν κάποια συστήματα τα οποία έχουν συμπαντικές εφαρμογές στην επιστήμη και την καθημερινή ζωή. Αυτά επιλέχθηκαν έτσι ώστε να είναι προφανείς η χρήση των αλγορίθμων, θεωρειών και μοντέλων ιδίων με αυτές που χρησιμοποιήθηκαν και σε αυτές τις εφαρμογές. Έτσι μέσα από αυτές τις εφαρμογές δημιουργήθηκε μια νέα φιλοσοφία και ιδέα για την κατασκευή ενός συστήματος πρωτοποριακού όπως είναι αυτό που περιφραχτεί.

#### 3.1 Τεχνολογίες ακμής (state-of-the-art)

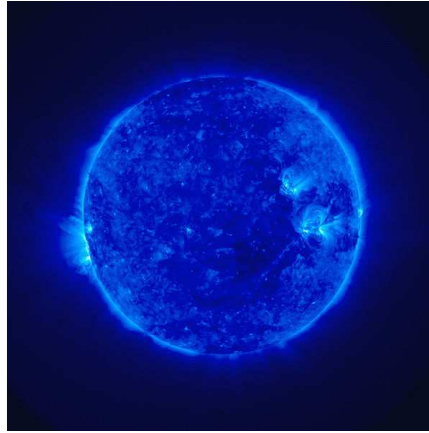
Ο πρώτος αλγόριθμος ο οποίος στηρίζετε το σύστημα είναι η υπολογιστική στερεοσκοπική όραση. Η υπολογιστική όραση βρίσκει πολλές εφαρμογές σε αυτόματα και αυτόνομα συστήματα. Η στερεοσκοπική όραση είναι πολύ σημαντική σε κλάδους όπως η ρομποτική, για την εξαγωγή των πληροφοριών από την θέση των τρισδιάστατων αντικειμένων στην περιοχή των αυτόνομων συστημάτων. Άλλες εφαρμογές της ρομποτικής συμπεριλαμβάνουν αναγνώριση αντικειμένων, όπου το βάθος επιτρέπει στο σύστημα να ξεχωρίζει κρυφά στοιχεία μιας εικόνας, όπως μια καρεκλιά μπροστά από μια άλλη, τις οποίες δεν μπορεί διαφορετικά να ξεχωρίσει με άλλα κριτήρια ένα ρομπότ[23].

Εφαρμογές στην επιστήμη για την ψηφιακή στερεοσκοπική όραση συμπεριλαμβάνουν εξαγωγή πληροφοριών από αεροφωτογραφίες, για τον υπολογισμό των χαρτών περιγράμματος ή ακόμα και την εξαγωγή γεωμετρίας για την τρισδιάστατη απεικόνιση κτιρίων, είτε ακόμα και για την τρισδιάστατη απεικόνιση της επιφάνειας του ήλιου όπως αυτή ελήφθη από το πρόγραμμα της NASA STEREO[23].



Εικόνα 72: Ένας από τους δορυφόρους του προγράμματος STEREO της NASA

Το NASA STEREO (Solar **TE**rrestrial **RE**lations **O**bservatory) είναι μια αποστολή παρατήρησης του ήλιου. Δυο σχεδόν ίδιοι δορυφόροι εκτοξεύτηκαν στις 26/10/2006 και τέθηκαν σε τροχιές οι οποίες τους κατεύθυναν μακρύτερα από την γη. Αυτή η αποστολή επιτρέπει την στερεοσκοπική απεικόνιση του ήλιου και των ηλιακών φαινομένων.



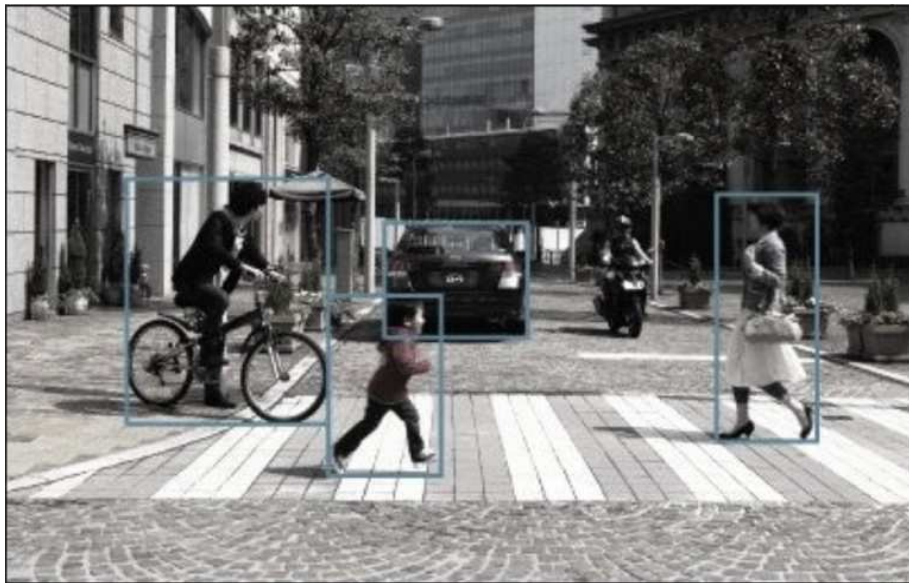
*Εικόνα 73: Πρώτη εικόνα του προγράμματος STEREO της NASA*

Σε διαφορετική φιλοσοφία αλλά με ίδια τεχνική η Subaru προσπαθεί να φέρει και αυτή την επανάσταση στην αυτοκίνηση με το σύστημα “EyeSight”. Με την χρήση στερεοσκοπικών καμερών, όπως φαίνεται παρακάτω, το σύστημα που συνεργάζεται με το αυτοκίνητο υπολογίζει τις αποστάσεις και βοηθά με προειδοποιήσεις στον οδηγό ή ακόμα και διαχείριση του συστήματος πέδησης του αυτοκίνητου. Το σύστημα είναι ήδη έτοιμο για την αναγνώριση πεζών, ποδηλατών, οχημάτων και άλλων χρηστών του δρόμου. Το πρώτο σύστημα EyeSight παρουσιάστηκε στην Ιαπωνία τον Μάιο του 2008.



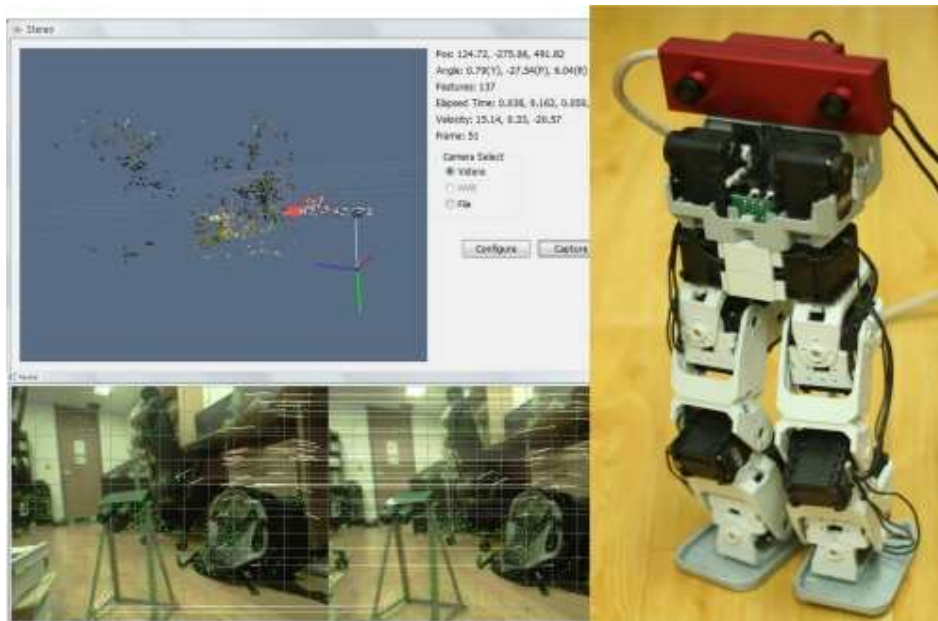
*Εικόνα 74: Σύστημα EyeSight από την εταιρία αυτοκίνητων Subaru*





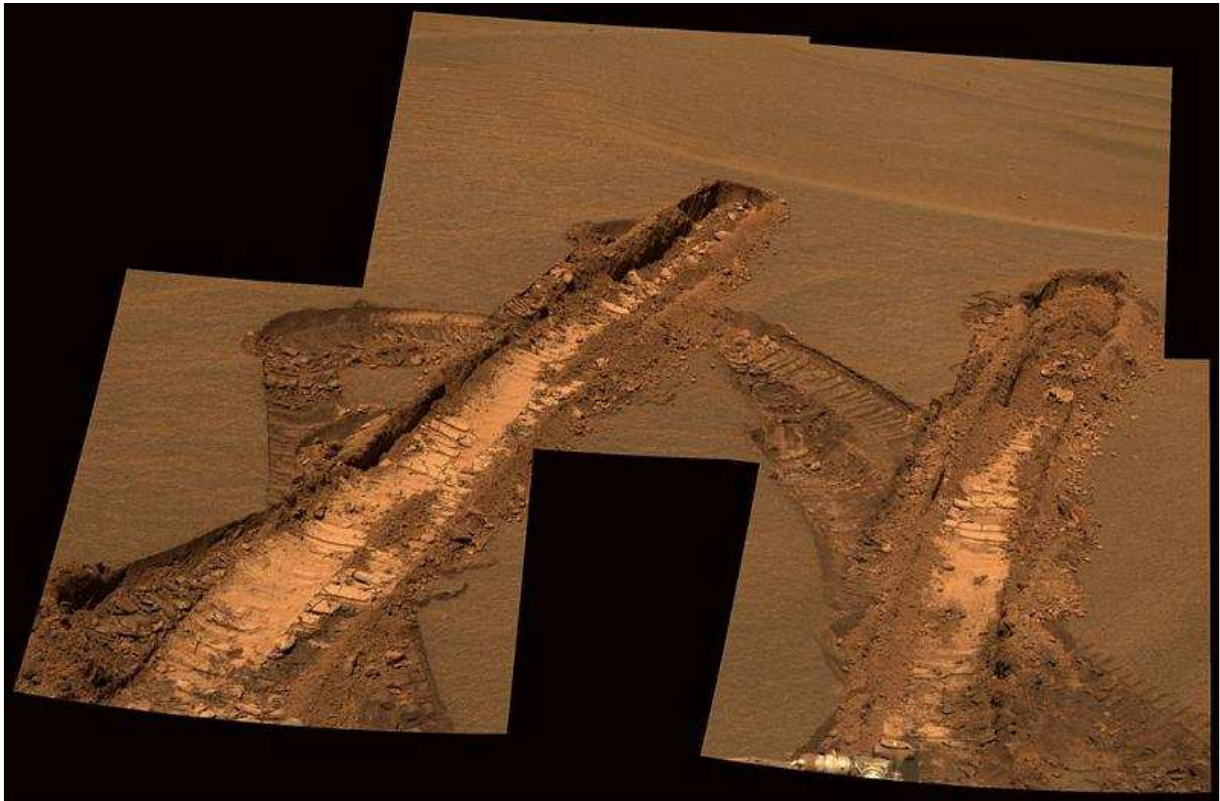
Εικόνα 75: Αποτέλεσμα συστήματος EyeSight

Αλλάζοντας φιλοσοφία και αλγόριθμο ακολουθούν συστήματα τα οποία έχουν να κάνουν με την αναγνώριση των αντικείμενων και την πλοήγηση. Δεδομένου των ιδιοτήτων του αλγορίθμου SIFT όπως αυτές αναφέρθηκαν παραπάνω, ο αλγόριθμος αυτός μπορεί να παράγει σημεία κλειδιά τα οποία είναι αμετάβλητα στις αλλαγές της εικόνας. Στις εφαρμογές ένα τριπλό στερεοσκοπικό σύστημα εκτιμά τις θέσεις των σημείων κλειδιά και κινείται στον χώρο εκτιμώντας τα ζεύγη των χαρακτηριστικών σε ένα υπάρχον τρισδιάστατο χάρτη του περιβάλλοντος. Έτσι το ρομπότ μπορεί να ανανεώνει και να υπολογίζει την θέση του στον χώρο. Παράδειγμα αποτελεί το επόμενο ρομπότ:



Εικόνα 76: Ρομπότ το οποίο πλοηγεί στον χώρο χάρις τον αλγόριθμο SIFT

άλλα και τα ρομπότ τα οποία εξερευνούν άλλους πλανήτες όπως δείξαμε παραπάνω (εικόνα 9) χρησιμοποιούν παρόμοια τεχνική που στηρίζετε στα παραπάνω αλλά και σε ράγνιμο εικόνων μεταξύ τους για να μπορούν να κινηθούν σε άγνωστα και δύσβατα τερέν.



*Εικόνα 77: Τα ίχνη του Mars Rover της NASA στην επιφάνεια του Άρη*

Επίσης μια άλλη εφαρμογή αποτελεί το ράψιμο των εικόνων. Με αυτήν την εφαρμογή μπορούν να φτιαχτούν προβολές από εικόνες η οποίες έχουν μερικώς το ίδιο περιεχόμενο. Τα κοινά χαρακτηριστικά αναγνωρίζονται στις δυο εικόνες και τότε η μια μετακινήτε με τέτοιο τρόπο έτσι ώστε να συμπληρώσει την άλλη εικόνα. Το καλό με αυτήν την μέθοδο είναι ότι επιτρέπει την σύνδεση εικόνων που δεν έχουν ληφθεί από με τον ίδιο τρόπο. Παράδειγμα εφαρμογής είναι η σύνδεση αεροφωτογραφιών για την χρήση τους σαν χάρτες σε πεδία μάχης, ακόμα και καθημερινή σύνδεση πολλαπλών εικόνων για την δημιουργία πανοραμάτων.



*Εικόνα 78: Ράψιμο των εικόνων αεροφωτογραφίας*

Μια άλλη εφαρμογή που χρησιμοποιήτε συχνά ο αλγόριθμος SIFT είναι οι απεικόνιση και μοντελοποίηση τρισδιάστατων σκηνών. Με τα παραγόμενα σημεία κλειδιά που καθορίζουν τον αλγόριθμο από τους πιο επιτυχείς με την συνεργασία πολλαπλών καμερών ή πολλαπλών λήψεων μέσω λογισμικού είναι ικανή η μοντελοποίηση. Οι κάμερες παρέχουν τα στοιχεία του βάθους στην εικόνα και τα βαθύ κάθε σημείου κλειδιού. Επίσης τις παραπάνω τεχνικές του ραψίματος της εικόνας μπορούμε να γνωρίζουμε σε ποια σημεία ακριβώς έγινε κάθε λήψη. Έτσι παράγετε ένα λεγόμενο σύννεφο από σημεία τα οποία μέσω αλγορίθμων τριγωνοποίησης καταφέρνουν την δημιουργία τρισδιάστατων μοντέλων ή σκηνών.



**3D Scene Simulation from Community Photo Collections (CPCs)**

*Εικόνα 79: Σύννεφο σημείων το οποίο απεικονίζει μια τρισδιάστατη σκηνή.*

Σε έναν τομέα στον οποίο υπάρχει αρκετός συναγωνισμός σε αλγόριθμους που αναγνωρίζουν σημεία κλειδιά είναι η αναγνώριση. Η αναγνώριση συνήθως αναφέρετε σε αντικείμενα τα οποία υπάρχουν στον χώρο και το σύστημα πρέπει να βρει τρόπους για να τα κατηγοριοποιήσει. Υπάρχουν συστήματα τα οποία απομνημονεύουν τα αντικείμενα τα οποία βρίσκονται στο περιβάλλον τους και με αυτόν τον τρόπο τους αναθέτουντα i διάφορες εργασίες όπως να θυμούνται την ώρα της φαρμακευτικής αγωγής άλλα και την τοποθεσία των φαρμάκων στο χώρο όπως το ρομπότ του IRT Research Institute που παρουσιάστηκε τον Δεκέμβριο του 2008.



*Εικόνα 80: Ρομπότ Mamoju που βοηθάει ηλικιωμένους*

Εφαρμογές γνωστές είναι και αυτές που σχετίζονται με της κινήσεις των ανθρώπων. Μέσω σημείων πάνω στο σώμα ενός ανθρώπου είναι δυνατόν να αναγνωριστεί η κίνηση του, οι προθέσεις του με χρήση σε συστήματα ασφάλειας, ακόμα και τα συναισθήματα του. Τέλος, εφαρμογές υπάρχουν και στην Ιατρική όπου γίνετε μετρήσεις μορφολογίας με βάση σημεία κλειδιά σε εικόνες από όργανα, οστά ή άλλα μέλη. Επίσης είναι δυνατή και η χρήση τέτοιων μεθόδων στην τρισδιάστατες απεικονίσεις από μαγνητικές τομογραφίες.

### **3.2 Σημαντικοί στόχοι εκπόνησης της πτυχιακής εργασίας**

Οι στόχοι που βάζουμε σε ένα σύστημα μπορούν να το εξελίσουν σε τέτοιο βαθμό έτσι ώστε όχι μόνο να επιτευχθούν αλλά να υπερκαλυφθούν με την ανάδειξη του συστήματος σε ένα πρωτοποριακό σύστημα. Πιστεύοντας σε αυτήν την φιλοσοφία οι στόχοι αυτής της εργασίας φάνταζαν αρχικά πολύ δύσκολοι, όμως με την γνώση που είχα αποκομίσει και την σωστή καθοδήγηση από του επιβλέποντος καθηγητή οι στόχοι έγιναν πραγματικότητα.

Έτσι αρχικά με τις γνώσεις που κατείχα μέχρι την στιγμή της αναζήτησης της πτυχιακής εργασίας, έγινε ένας σχηματισμός μιας ιδέας. Η ιδέα ήταν να παραχθεί ένα σύστημα το οποίο να μπορεί να μιλήσει την δομή της βιολογικής όρασης και αντίληψης του κόσμου. Μεγαλύτερη βαρύτητα από αυτήν την δομή δόθηκε στον τρόπο αντίληψης των αντικειμένων και της κατηγοριοποίησης τους από τον άνθρωπο. Έτσι παρήχθη η ιδέα της κατασκευής ενός υβριδικού συστήματος τεχνίτης όρασης και τεχνητής νοημοσύνης. Η τεχνητή όραση θα μπορούσε να παρέχει τις πληροφορίες από το περιβάλλον με μια προσομοιώσει της βιολογικής όρασης και αντίληψης, ενώ η τεχνητή νοημοσύνη θα εκμεταλλευόταν αυτές τις πληροφορίες για την κατηγοριοποίηση τους σε γνωστές κλάσεις.

Έτσι το πρώτο που συνέβη ήταν η έναρξη της έρευνας τις τεχνολογίας ακμής (state-of-the-art). Αυτό είναι το σημαντικότερο κομμάτι στην καθιέρωση μιας ιδέας και ενός ολοκληρωμένου σχεδίου δράσης για την εκτέλεση της. Ψάχνοντας σε διάφορες πηγές και κυρίως στο διαδίκτυο έπεσα πάνω σε συστήματα τα οποία ήταν ερευνητικά και είχαν μια παρόμοια φιλοσοφία και ιδέα με αυτό που ήθελα και εγώ να εκπληρώσω στην πτυχιακή μου εργασία. Παραπάνω φαίνονται μερικά από αυτά.

Αφού ολοκληρώθηκε και η έρευνα των τεχνολογιών ακμής (state-of-the-art), σειρά είχε ένας σχεδιασμός για την εκπόνηση της πτυχιακής. Μέσα από τον σχεδιασμό ερευνηθήκε η βιολογική όραση και αντίληψη. Επίσης ερευνηθήκε μέθοδοι του κλάδου της επιστήμης των υπολογιστών, της τεχνίτης όρασης. Έγινε μια

σύνδεση ανάμεσα στις βιολογικές λειτουργίες και τις λειτουργίες της τεχνητής όρασης. Τέλος, έγινε έρευνα για την εύρεση της κατάλληλης μεθόδου για την ένωση των δεδομένων από το σύστημα τεχνητής όρασης στο σύστημα της τεχνητής νοημοσύνης.

Με το τέλος του σχεδιασμού του συστήματος της πτυχιακής εργασίας έγινε το σταδιακό στήσιμο του συστήματος. Αρχικά το σύστημα ξεκινά από την εισαγωγή των στερεοσκοπικών εικόνων και την επεξεργασία τους και την εξαγωγή των δεδομένων. Μετά υπάρχει μια επεξεργασία αυτών των δεδομένων για την εξαγωγή χαρακτηριστικών γνωρισμάτων. Αφού και αυτό ολοκληρώνετε τότε το σύστημα πρέπει να κατηγοριοποιήσει τα δεδομένα του σε κλάσεις αντικειμένων. Όμως με την υλοποίηση του πρώτου πιλοτικού συστήματος έγιναν αντιληπτά κάποια προβλήματα τα οποία έχουν να κάνουν σχέση με την απόδοση του.

Με την αξιολόγηση της γνώσεις που έλαβα από το μερικώς αποτυχημένο στήσιμο του πρώτου πιλοτικού συστήματος έδειξαν τα σημεία που επιστούσαν την προσοχή μου. Έτσι έγινε ανάλυση και νέα αναζήτηση για καλύτερους αλγόριθμους που να επίλυαν τα προβλήματα. Η υλοποίηση του δεύτερου πιλοτικού συστήματος ήταν επιτυχής και πλέον τα αποτελέσματα ήταν τα επιθυμητά. Σε εκείνο το σημείο έγινε ένας λεπτομερής έλεγχος στην λειτουργία και την σαφήνεια των αποτελεσμάτων του.

Πλέον με μόνο τα διαδικαστικά της παρουσίασης να αναμένουν τα επόμενα βήματα είναι η ολοκλήρωση της συγγραφής αυτής της αναφοράς, η υποβολή αίτησης αξιολόγησης και προετοιμασία παρουσίασης της πτυχιακής εργασίας. Έτσι τέλος έχουμε μια αναλυτική λίστα με όλα τα περιγραφόμενα η οποία τοποθετείτε και σε χρονοδιάγραμμα όπως φαίνεται παρακάτω.

- ▲ Ολοκλήρωση σύλληψης και σχεδιασμού της ιδέας του συστήματος.
- ▲ Ολοκλήρωση έρευνας των τεχνολογιών αιχμής (state-of-the-art)
- ▲ Ολοκλήρωση σχεδιασμού της εκπόνησης της πτυχιακής εργασίας
- ▲ Ολοκλήρωση πρώτου πιλοτικού συστήματος της πτυχιακής εργασίας
- ▲ Ολοκλήρωση ανάλυση, αξιολόγησης και διόρθωσης προβλημάτων πιλοτικού συστήματος
- ▲ Ολοκλήρωση δεύτερου πιλοτικού συστήματος της πτυχιακής εργασίας
- ▲ Ολοκλήρωση τεχνικού μέρους της πτυχιακής εργασίας
- ▲ Ολοκλήρωση ελέγχου λειτουργίας συστήματος της πτυχιακής εργασίας
- ▲ Ολοκλήρωση συγγραφής αναφοράς και παρουσίασης της πτυχιακής εργασίας
- ▲ Ολοκλήρωση υποβολή αίτησης για την αξιολόγηση της πτυχιακής εργασίας
- ▲ Ολοκλήρωση προετοιμασίας παρουσίασης αναφοράς της πτυχιακής εργασίας
- ▲ Ολοκλήρωση παρουσίαση αναφοράς της πτυχιακής εργασίας

## 4. Κύριο μέρος Πτυχιακής εργασίας

Σε αυτό το κεφάλαιο της πτυχιακής εργασίας περιγράφετε αναλυτικά η ένωση όλων των παραπάνω κεφαλαίων. Στα προηγούμενα κεφάλαια έγινε μια εισαγωγή πάνω στα θέματα που πραγματεύεται το σύστημα που περιγράφετε στην εργασία. Έτσι όλα τα παραπάνω μοντέλα, θεωρίες, αλγόριθμοι αλλά και οι καινοτομίες των προαναφερθέντων συστημάτων συνδέονται σε ένα κοινό σύστημα το οποίο τα αξιοποιεί για των στόχο της αναγνώρισης αντικειμένων μέσω των χαρτών βάθους και μάθησης μηχανής.

Στην συνέχεια του κεφαλαίου θα αναπτυχθούν τα θέματα όπως η ανάλυση των προβλημάτων τα οποία εμφανίστηκαν κατά την εκπόνηση της εργασίας. Τα προβλήματα αυτά κυρίως εμφανίστηκαν ανάμεσα στους συνδυασμούς των πολλαπλών αλγορίθμων αλλά και των ορισμένες φορές πολύπλοκων αποτελεσμάτων τους. Επίσης, γίνεται αναφορά στις απαιτήσεις τις οποίες θα πρέπει να έχει ένα υπολογιστικό σύστημα για την εκτέλεση του περιγραφόμενου συστήματος. Έτσι χρειάστηκαν ριζικές αλλαγές οι οποίες περιγράφονται στο κεφάλαιο του αναλυτικού σχεδιασμού.

Θα ακολουθήσει αναλυτικός σχεδιασμός της υλοποίησης του συστήματος. Σε αυτόν γίνεται η σύνδεση ανάμεσα στα προβλήματα που περιγράφηκαν και στις λύσεις που λήφθηκαν για την διόρθωση και την αποφυγή τους. Έτσι αφού πλέον είναι γνωστά τα προβλήματα αλλά και η λύσεις τους, γίνεται περιγραφή της υλοποίησης του συστήματος.

Στην περιγραφή του συστήματος γίνεται αναλυτική ανάλυση της υλοποιήσεως που είναι στην πραγματικότητα η εκτέλεση του σχεδιασμού. Έτσι σε εκείνο το κομμάτι θα δοθεί ο κώδικας και θα γίνει εκτενής περιγραφή τις κάθε γραμμής και των αποτελεσμάτων που λαμβάνουμε.

### 4.1 Ανάλυση Προβλημάτων

Η ανάλυση του προβλήματος είναι το κομμάτι της εργασίας το οποίο έχει άμεση σχέση με την ιδέα της κατασκευής του. Αυτό είναι απόλυτα κατανοητό και φυσιολογικό, αφού κάθε ιδέα μπορεί να οδηγήσει σε μεγάλα προβλήματα τα οποία δεν είναι εμφανή με την πρώτη ματιά. Αυτά τα προβλήματα αντιμετωπίζονται με εφευρετικές λύσεις που εφευρίσκονται ή ανακαλύπτονται μετά από έρευνα και πειραματισμό.

Η αρχική ιδέα του συστήματος περιγράφετε και στον τίτλο της πτυχιακής αυτής εργασίας και έχει να κάνει με αναγνώριση αντικείμενων με χρήση χαρτών βάθους και τεχνολογιών μάθησης μηχανής. Σκοπός ήταν η κατασκευή ενός συστήματος το οποίο να μπορεί να αναγνωρίζει αντικείμενα στο περιβάλλον (τρισεδιάστατος χώρος) και μέσω μιας διαδικασίας να είναι σε θέση να αναγνωρίζει την κατηγορία που ανήκει χωρίς απαραίτητα να υπάρχει ένα ίδιο αντικείμενο στην βάση δεδομένων του συστήματος. Για την πραγματοποίηση της ιδέας αυτής το σύστημα θα έπρεπε να καταταμηθεί σε τρία κομμάτια. Το πρώτο κομμάτι που θα λάμβανε τις εικόνες και θα τις επεξεργαζόταν. Το δεύτερο κομμάτι που θα έβρισκε τα αντικείμενα και χαρακτηριστικά για κατηγοριοποίηση τους. Και τέλος, το τρίτο κομμάτι το οποίο θα έκανε την κατηγοριοποίηση και θα παρήγαγε την γνώση για το σύστημα. Επίσης, σοβαρό θέμα αποτέλεσαν και οι γενικές απαιτήσεις του υπολογιστικού συστήματος οι οποίες είναι απαραίτητες για την εκτέλεση του σε κάθε σύστημα.

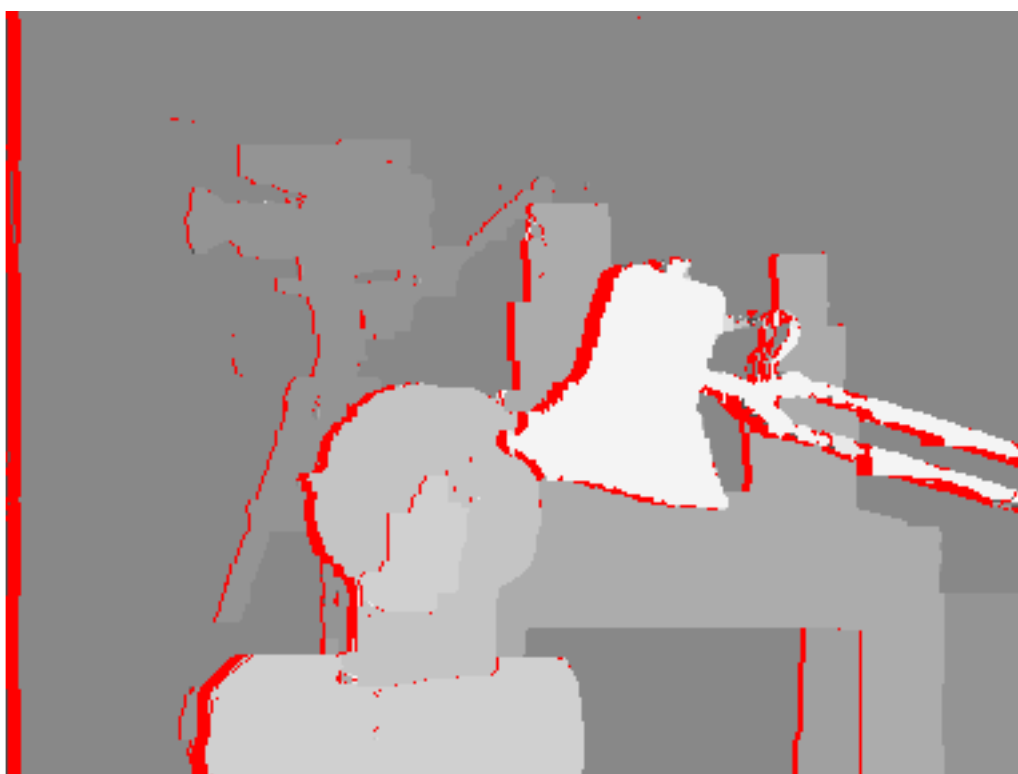
#### 4.1.1 Προβλήματα λήψης και επεξεργασίας εικόνας

Σε αυτό το κεφάλαιο αναλύονται διεξοδικά όλα τα προβλήματα τα οποία συνέβησαν κατά την διάρκεια της διαδικασίας λήψης και επεξεργασίας μιας εικόνας. Η αρχική ιδέα αναφέρει ότι τα αντικείμενα βρίσκονται σε κανονικό περιβάλλον. Ο χώρος γύρω μας περιγράφεται από τα μαθηματικά και την φυσική ως τρισεδιάστατος χώρος. Αυτό σημαίνει ότι κάθε αντικείμενο μέσα σε αυτόν τον χώρο έχει τρεις διαστάσεις το πλάτος, το ύψος και το βάθος του από την γωνιά θέασης. Όπως αναφέρετε και σε παραπάνω κεφάλαια ο τρόπος των ανεπτυγμένων οργανισμών για την ανίχνευση του βάθους που τους βοηθά στην πλοήγηση μέσα στον χώρο

αλλά και στην αναγνώριση αντικειμένων όπως άλλοι οργανισμοί και φυσικά εμπόδια είναι η στερεοσκοπική όραση. Η διαδικασία αυτή είναι γνωστή στην επιστήμη των υπολογιστών και τον τομέα της τεχνητής όρασης. Επιλέχθηκε ως η βέλτιστη λύση του προβλήματος της αναγνώρισης των αντικειμένων στον χώρο.

Η στερεοσκοπική υπολογιστική όραση, όπως αναφέρετε και παραπάνω έχει επιλεχθεί σε μεγάλα επιστημονικά προγράμματα. Η κλασική προσέγγιση αυτής της μεθόδου είναι η χρήση δυο καμερών (ποιο απλή περίπτωση) οι οποίες λαμβάνουν δυο εικόνες ελάχιστα μετατοπισμένες μεταξύ τους όπως φαίνεται στην εικόνα 5. Αυτές οι εικόνες γίνονται δεδομένα επεξεργασίας ενός αλγορίθμου ο οποίος υπολογίζει το βάθος στο οποίο βρίσκετε κάθε εικονοστοιχείο. Αυτό γίνεται με τον υπολογισμό της μετατόπισης κάθε εικονοστοιχείου από την μια εικόνα σε σχέση με την άλλη. Έτσι κάθε εικονοστοιχείο που εμφανίζεται μακριά είναι όλο και πιο μετατοπισμένο από την μια εικόνα στην άλλη. Η ενέργεια που παράγετε ομαλοποιείτε συνήθως από κάποιων αλγόριθμο ο οποίος ομαλοποιεί την τάση κάποιων περιοχών να παράγουν τιμές που χαλάν το αποτέλεσμα. Με αυτόν τον τρόπο παράγονται τελικά οι χάρτες βάθους όπως εμφανίζετε στην εικόνα 6.

Το πρώτο πρόβλημα το οποίο καλέστηκα να αντιμετωπίσω ήταν ακριβώς αυτό το πρόβλημα τις ενέργειας. Αρχικά χρησιμοποίησα τον κλασικό αλγόριθμο με την κλασική προσέγγιση[17]. Παρατήρησα ότι, όπως εμφανίζετε στην εικόνα 23, οι περιοχές που βρίσκονταν στην εικόνα εμφανίζουν μεγάλη ενέργεια από τις ανταποκρίσεις των εικονοστοιχείων. Στην πραγματικότητα πολλά από τα εικονοστοιχεία ήταν λανθασμένος τοποθετημένα σε περιοχές στις οποίες δεν θα έπρεπε να βρίσκονται. Αυτή η παρατήρηση με οδήγησε σε έρευνα των λόγων για το οποίο συμβαίνει αυτό το φαινόμενο.



*Εικόνα 81: Εμφάνιση κρυφών σημείων στην εικόνα του πειράματος*

Η έρευνα κατέληξε σε ένα φαινόμενο το οποίο συμβαίνει λόγω της γεωμετρίας των καμερών. Τα σημεία τα οποία είναι εμφανή σε μια λήψη και μη εμφανή σε μια άλλη, αυτά ονομάζονται ως κρυφά σημεία. Τα κρυφά σημεία, όπως εμφανίζονται στην παραπάνω εικόνα, παράγουν μεγάλη ενέργεια επειδή έχουν μεγάλες αποκλίσεις στις ανταποκρίσεις των εικονοστοιχείων μεταξύ των εικόνων. Έτσι στην εικόνα που χρησιμοποιήθηκε για την έρευνα και το στήσιμο του συστήματος τα αρχικά αποτελέσματα ήταν ικανοποιητικά αλλά όχι απόλυτα ακριβή για την εύρεση των αντικειμένων στον χώρο.

Άλλα προβλήματα τα οποία παράχθηκαν ήταν μεταγενέστερα του παραπάνω προβλήματος αλλά έχουν

άμεση σχέση με αυτό. Οι χάρτες βάθους στην πραγματικότητα είναι πίνακες οι οποίοι περιγράφουν τα επίπεδα βάθους μιας εικόνας. Έτσι κάθε αριθμός μέσα σε κελί του πίνακα βάθους συμβολίζει ένα επίπεδο και τα αντικείμενα τα οποία βρίσκονται σε αυτό. Έτσι κάποια αντικείμενα τα οποία μπορεί να βρίσκοντουσαν σε παραπάνω από ένα βάθη θα έπρεπε με κάποιον τρόπο να εννοηθούν και να σχηματίσουν το αντικείμενο.

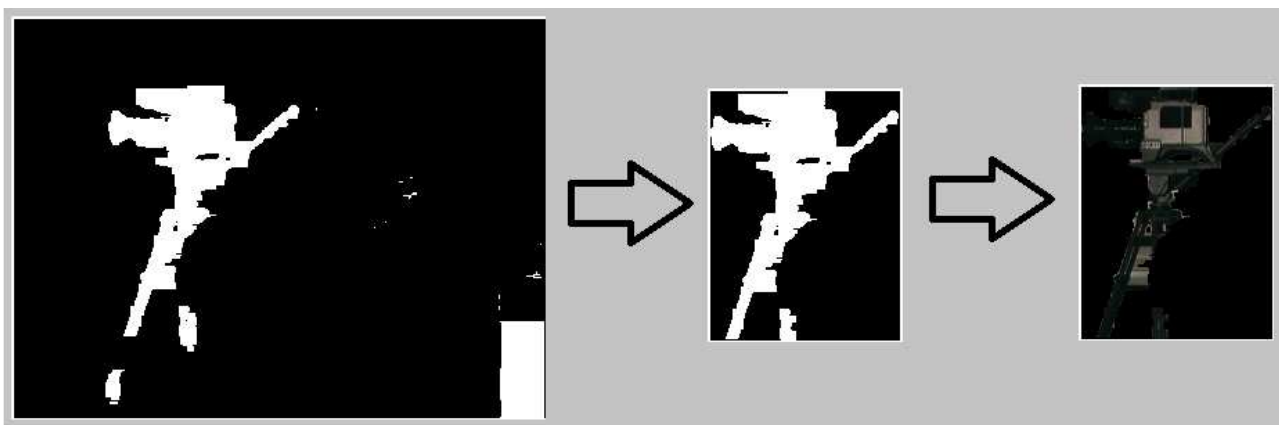


Εικόνα 82: Πρόβλημα σχηματισμού αντικειμένων που άνηκαν σε διαφορετικά επίπεδα

#### 4.1.2 Προβλήματα αναγνώρισης και κατηγοριοποίησης αντικειμένων

Τα προβλήματα τα οποία αντιμετωπίστηκαν κατά την διάρκεια της αναγνώρισης και κατηγοριοποίησης κυρίως αφορούσαν την έλλειψη γνώσεων πάνω στις μεθόδους που χρησιμοποιήθηκαν. Αφού είχε γίνει η επεξεργασία στην εικόνα, επέτρεπε στο επόμενο βήμα όπως αναφέρθηκε παραπάνω στην αρχική ιδέα να βρεθούν τα αντικείμενα μέσα σε αυτή. Αμέσως μετά έπρεπε κάθε αντικείμενο να αποκτήσει τις δικές του ιδιότητες και χαρακτηριστικά για να παραχθούν δεδομένα τα οποία να χρησιμοποιηθούν από έναν αλγόριθμο τεχνητής νοημοσύνης.

Το αποτέλεσμα των εικόνων ήταν χάρτες βάθους στους συμβολίζονταν βάθη τα οποία είχαν συγκέντρωση ενεργών εικονοστοιχείων που οδηγούσαν στην πιθανή ύπαρξη αντικειμένων. Στην επεξεργαστική τους μορφή η χάρτες βάθους ήταν δυαδικοί πίνακες. Αφού είχαν επιλεγεί με σύστημα ψηφοφορίας παραπάνω, αυτό σήμαινε ότι υπήρχαν αντικείμενα σε αυτό το βάθος. Η εξαγωγή τους ήταν αρκετά δύσκολη γιατί απαιτούσε να γίνει ομαδοποίηση των εικονοστοιχείων τα οποία περιέγραφαν το ίδιο αντικείμενο. Επίσης αφού γινόταν αυτό θα έπρεπε να γίνει ταυτοποίηση με κάποιον αξιόπιστο τρόπο έτσι ώστε να υπάρχει σιγουριά ότι πρόκειται για αντικείμενο. Τέλος, αφού γινόταν η ταυτοποίηση έπρεπε να εξαχθεί το αντικείμενο από την κανονική εικόνα με τέτοιο τρόπο ώστε να ήταν ακριβής και να μην υπήρχε λήψη στοιχείων που να επηρεάσουν την απόδοση του αλγορίθμου που θα έβγαζε τα χαρακτηριστικά.



Εικόνα 83: Δείγμα του επιθυμητού αποτελέσματος από την αναγνώριση των αντικειμένων



Με την λύση των παραπάνω φτάνουμε στο σημείο όπου έχει γίνει η εξαγωγή των αντικειμένων από την εικόνα. Από εδώ και πέρα το θέμα βρίσκεται στο πως αυτά τα αντικείμενα θα εξάγουν χαρακτηριστικά ή άλλα δεδομένα από την μορφή τους. Έτσι στην αναγνώριση αντικειμένων στην τεχνητή όραση έχουμε δυο κύριους τρόπους. Ο πρώτος τρόπος, που είναι και ποιο κλασικός, είναι η εξαγωγή συμπερασμάτων για το αντικείμενο μέσω της εμφάνισης του. Σε αυτόν τον τρόπο η επεξεργασία της εικόνας γίνεται ανάλογα με την δομή της εικόνας ή με το πεδίο των συχνοτήτων. Μερικοί από αυτούς τους αλγόριθμους είναι η αναγνώριση ακμών (Canny, Roberts, Sobel), η αναγνώριση γωνιών (Harris), ο μετασχηματισμός Hough και ο μετασχηματισμός Fourier. Αυτοί η αλγόριθμοι παραμένουν καλοί στην επίλυση ορισμένων, συνήθως απλών, προβλημάτων της υπολογιστικής όρασης. Στην πραγματικότητα όμως παραμένουν αρκετά ευαίσθητοι σε θέματα όπως η τοποθέτηση των καμερών και ο φωτισμός.

Ο δεύτερος τρόπος είναι η εξαγωγή χαρακτηριστικών. Αυτή η μέθοδος παίρνει την κάθε εικόνα και μέσω κάποιας επεξεργασίας παράγει σημεία κλειδιά τα οποία σηματοδοτούν και δίνουν στην εικόνα ένα ξεχωριστό χαρακτήρα. Μερικοί από αυτούς τους αλγόριθμους είναι ο μετασχηματισμός εξαγωγής χαρακτηριστικών αμετάβλητων στην κλίμακα (Scale-Invariant Feature Transform, SIFT) , αλγόριθμος γρήγορων και στιβαρών χαρακτηριστικών (Speeded Up Robust Features, SURF), και άλλοι. Παρ' όλη την πολυπλοκότητα του είναι "ανθεκτικοί" στην λήψη της εικόνας, στον φωτισμό και κάποιιοι από αυτούς και στην περιστροφή της και η υπολογιστική ισχύ που απαιτούν δεν είναι μεγάλη. Όμως επειδή μιλάμε για αλγόριθμους αρκετά πολύπλοκους η συμπεριφορά και τα δεδομένα τους έπρεπε να μελετηθούν πάρα πολύ για την εύρεση κατάλληλων αλγόριθμων τεχνητής νοημοσύνης όπου να μπορούν να τους επεξεργαστούν.

### 4.1.3 Προβλήματα παραγωγής γνώσης και αναγνώρισης

Σε αυτό το σημείο όλα τα προαναφερθέντα προβλήματα έχουν πλέον λυθεί και παραμένει μονάχα η χρήση των αλγορίθμων αυτών που ευθύνονται για την παραγωγή των γνώσεων και της τελικής αναγνώρισης. Αυτό το σημείο σύμφωνα με την αρχική ιδέα είναι και σημείο κλειδί στην αποτελεσματικότητα όλου του συστήματος.

Κάθε αντικείμενο τώρα πλέον έχει περάσει από την παραπάνω διαδικασία και έχουν εξαχθεί από αυτό κάποια σημαντικά χαρακτηριστικά τα οποία το ορίζουν. Σύμφωνα με την περιγραφή των χαρακτηριστικών που παράγονται, κάθε χαρακτηριστικό είναι ένας πίνακας στοιχείων με μέγεθος 128 κελιών. Έτσι το πρώτο πράγμα που αντιλαμβανόμαστε είναι η επεξεργαστική ισχύ πρέπει να είναι μεγάλη ή η τεχνική του αλγόριθμου να μπορεί να επεξεργαστεί εύκολα αυτόν τον όγκο των δεδομένων. Επίσης η φύση των περιεχομένων των κελιών είναι ένα θέμα αφού δεν μπορούν να επεξεργαστούν ξεχωριστά αφού έτσι χάνετε το νόημα τους. Επίσης πρέπει να βρεθούν και οι κατάλληλες ρυθμίσεις για να επιτραπεί η χρήση των δεδομένων και να λαμβάνουμε την κατάλληλη φύση των αποτελεσμάτων. Τέλος, ένα πρόβλημα το οποίο προέκυψε ήταν και η επανεκπαίδευση και αναπροσαρμογή του συστήματος σε νέα δεδομένα.

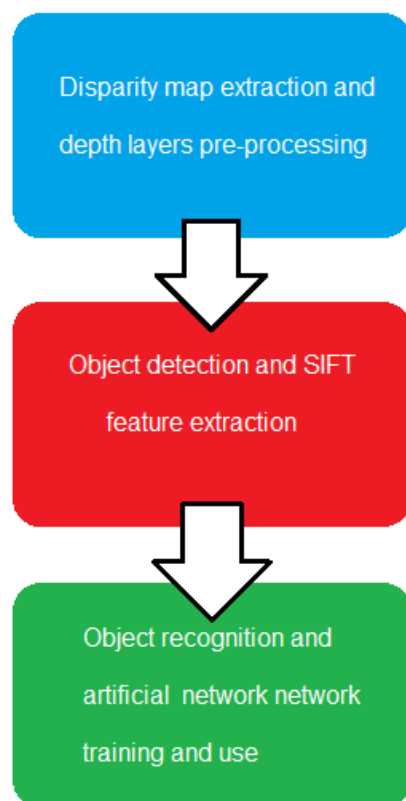
### 4.1.4 Απαιτήσεις του συστήματος

Σε αυτό το κεφάλαιο αναφέρονται αναλυτικά όλες αυτές οι απαιτήσεις τις οποίες πρέπει να καλύπτει το υπολογιστικό σύστημα για να μπορέσει να εκτελέσει το περιγραφόμενο σύστημα. Το σημαντικότερο σε ότι έχει να κάνει με τις απαιτήσεις του συστήματος η έλλειψη μεγάλης επεξεργαστικής ισχύς καθώς και η χρήση του από οποιοδήποτε υπολογιστικό σύστημα. Επειδή το σύστημα που περιγράφετε έχει γίνει πειραματικά, για την δημιουργία του εργάστηκα πάνω στο πρόγραμμα Matlab της Mathworks. Έτσι κάθε υπολογιστικό σύστημα που θέλει να τρέξει αυτό το σύστημα πρέπει να έχει:

- ▲ οποιοδήποτε λειτουργικό σύστημα το οποίο είναι συμβατό με το Matlab
- ▲ εκδοσή του Mathworks Matlab νεότερη από την R2007
- ▲ compiler Microsoft Visual C++ 2008 SP1 ή νεότερο.

## 4.2 Σχεδιασμός υλοποίησης

Ο σχεδιασμός της υλοποίησης του συστήματος είναι ένα από τα πιο σημαντικά κομμάτια της τελικής εργασίας. Ο αρχικός σχεδιασμός έγινε με την φιλοσοφία της απλής κατάτμησης των τομέων του συστήματος όπως αυτοί αναφέρονται παραπάνω. Έτσι προέκυψε το ακόλουθο διάγραμμα:



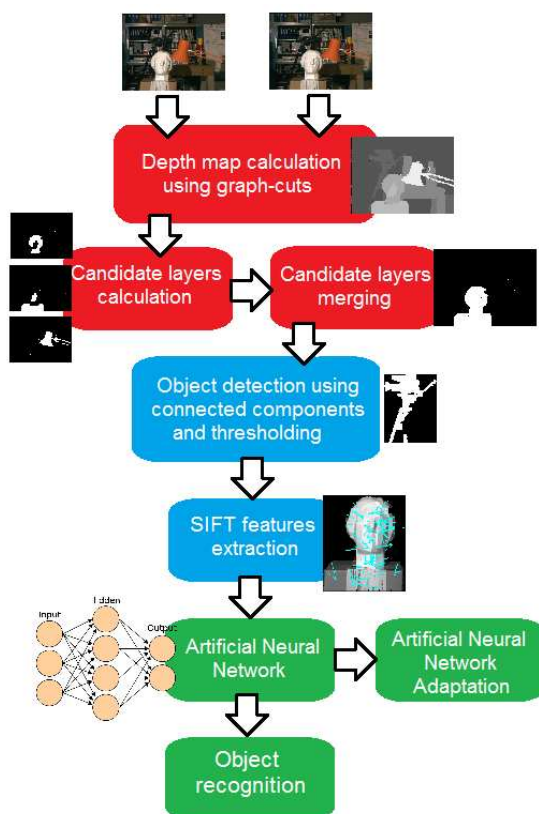
Εικόνα 84: Αρχικό διάγραμμα σχεδιασμού υλοποίησης

Στο διάγραμμα αυτό περιγράφονται και οι τρεις τομείς που έπρεπε να αναλυθούν. Έτσι το πρώτο κομμάτι αποτέλεσε ένα σύστημα από αλγόριθμο που επέτρεπε την χρήση στερεοσκοπικού ζεύγους εικόνων και την εξαγωγή χαρτών βάθους από αυτό. Η χρήση του αλγορίθμου match v3.1 του Vladimir Kolmogorov [18] κρίθηκε ως η καταλληλότερη για τους λόγους που περιγράφηκαν αναλυτικά σε παραπάνω κεφάλαια. Ο αλγόριθμος αυτός λόγω της χρήσης του θεωρήματος μεγίστης ροής και ελάχιστης τομής από την θεωρία των γράφων καταφέρνει να επιτύχει το μέγιστο αποτέλεσμα. Λύνονται με αυτόν τον τρόπο τα προβλήματα με την υψηλή ενέργεια η οποία δημιουργήτε στις κρυφές περιοχές των δυο εικόνων. Έτσι λαμβάνετε με επιτυχία ένας από τους αναλυτικότερους χάρτες βάθους για περαιτέρω ανάλυση. Παρόλα αυτά το σύστημα θέλει να κατηγοριοποιήσει τα επίπεδα τα οποία περιέχουν κάποιο αντικείμενο από αυτά τα οποία δεν έχουν. Έτσι δημιουργήθηκε σύστημα ψηφοφορίας στο οποίο κάθε επίπεδο εξετάζεται για τα ενεργά του εικονοστοιχεία. Αν το επίπεδο έχει αρκετά εικονοστοιχεία θεωρείτε ότι περιέχει αντικείμενα και περνάει για περαιτέρω ανάλυση. Μερικά αντικείμενα περιγράφονται σε ένα ή παραπάνω βάθη έτσι χρειάστηκε και μηχανισμός συγχώνευσης κοντινών επιπέδων βάθους.

Αφού έγιναν όλες οι παραπάνω ενέργειες ο σχεδιασμός του συστήματος περνάει στην δεύτερη φάση όπου γίνεται η αναγνώριση πιθανών αντικειμένων και η χρήση του μετασχηματισμού κλιμακωτά – αναλλοίωτων χαρακτηριστικών (SIFT) για την εξαγωγή χαρακτηριστικών. Η αναγνώριση των αντικειμένων γίνεται σε κάθε επίπεδο βάθους ξεχωριστά. Εκεί με την χρήση του αλγορίθμου συνδεδεμένων στοιχείων είναι δυνατόν να βρεθούν και να κατηγοριοποιηθούν όλα τα υποψήφια αντικείμενα σε κάθε ένα επίπεδο βάθους. Το

πρόβλημα της αναγνώρισης λύνεται με μια απλή μέτρηση της επιφάνειας των αντικειμένων σε κάθε επίπεδο βάθους και με την επιλογή των επικρατέστερων αντικειμένων μέσα σε αυτό. Αφού πλέον έχουν βρεθεί τα αντικείμενα, πρέπει επίσης να βρεθεί και η φυσική τους θέση μέσα στην εικόνα. Έτσι μέσω στατιστικών βρίσκετε η θέση κάθε αντικειμένου στην εικόνα και αντιστοιχίζετε με την φυσική του εικόνα. Αφού πλέον έχουμε το αντικείμενο το μόνο που χρειαζότανε για την αναγνώριση του είναι η εξαγωγή χαρακτηριστικών του. Εδώ παίρνει τα ηνία ο μετασχηματισμός κλιμακωτά – αναλλοίωτων χαρακτηριστικών (SIFT) που είναι υπεύθυνος για αυτήν την δουλειά. Αφού γίνει η επεξεργασία με τον μετασχηματισμό τα στοιχεία του αποθηκεύονται και μεταφέρονται στο τρίτο και τελευταίο κομμάτι του συστήματος.

Το τελευταίο κομμάτι ασχολείται με την αξιοποίηση των παραπάνω χαρακτηριστικών. Ο αλγόριθμος της τεχνητής νοημοσύνης που επιλέχθηκε είναι τα τεχνητά νευρωνικά δίκτυα. Αυτό το μοντέλο ενσωματώνει πλεονεκτήματα όπως σταθερή δομή, προσαρμοστικότητα και γρήγορη επεξεργασία μεγάλου αριθμού δεδομένων. Ο αλγόριθμος αυτός χωρίζετε σε δυο υπό - τμήματα. Το πρώτο είναι η δημιουργία και εκπαίδευση η οποία έχει προηγηθεί της εκτέλεσης του κύριος συστήματος. Δεν απαιτείται να γίνεται συνέχεια αλλά μόνον μια φορά στην αρχή του στησίματος, και μετά κάθε αλλαγή γίνεται με προσαρμογή των αρχικών παραμέτρων. Το δεύτερο κομμάτι είναι αυτό της εκτέλεσης, στο οποίο το σύστημα παρέχει τα παραπάνω δεδομένα που εξήγαγε από τα αντικείμενα για να επεξεργαστούν από το τεχνητό νευρωνικό δίκτυο. Το αποτέλεσμα του τεχνητού νευρωνικού δικτύου γίνεται μέσω ενός αλγορίθμου αποτίμησης των αποτελεσμάτων και εξαγωγής τελικού συμπεράσματος. Στο ολοκληρωμένο σύστημα υπάρχει ήδη νευρωνικό δίκτυο έτοιμο για χρήση με προκαθορισμένες ρυθμίσεις και δεν είναι απαραίτητο το στήσιμο του. Έτσι όλος ο παραπάνω σχεδιασμός καταλήγει σε ένα ποιο εκτενές και λεπτομερές διάγραμμα του σχεδιασμού της υλοποίησης.



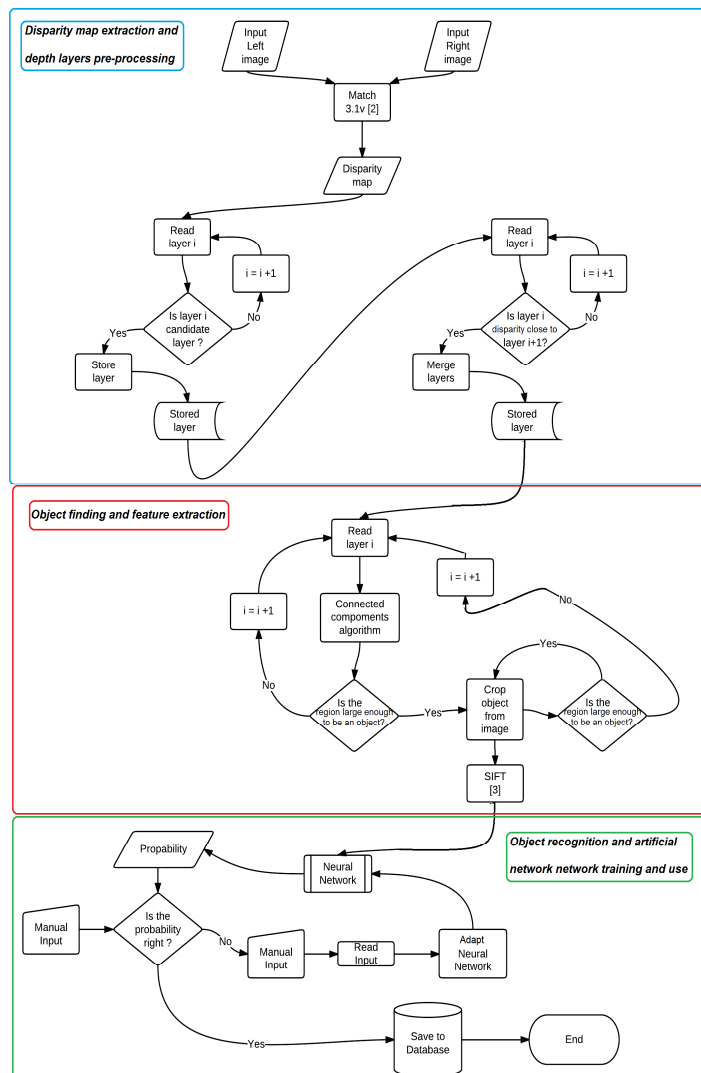
Εικόνα 85: Τελικό διάγραμμα σχεδιασμού υλοποίησης

### 4.3 Υλοποίηση του συστήματος

Σε αυτό το κεφάλαιο είναι εμφανή η απόδειξη της υλοποίησης του συστήματος της πτυχιακής εργασίας. Η υλοποίηση του γίνεται με την αποκόμιση της εμπειρίας του σχεδιασμού της υλοποίησης και την πλήρη κατανόηση των προβλημάτων των οποίων καλέστηκε να επιλύσει. Έτσι παρακάτω παραδίδεται ο αλγόριθμος του τελικού συστήματος τόσο το σχεδιάγραμμα του αλγορίθμου όσο και ο κώδικας ο οποίος τον υλοποιεί. Στην συνέχεια γίνεται λεπτομερής περιγραφή της υλοποίησης κάθε κομματιού ξεχωριστά και τις επίδειξης των αποτελεσμάτων τους.

#### 4.3.1 Σχεδιάγραμμα αλγορίθμου συστήματος

Το σχεδιάγραμμα του ενός αλγορίθμου είναι ένα οπτικό εργαλείο για τον προγραμματισμό. Το σχεδιάγραμμα είναι η οπτική απεικόνιση που κάνει ευκολότερη την κατανόηση του παραπάνω αλγορίθμου. Έτσι κάθε ένας είναι σε θέση με μικρή προσπάθεια να μπει στην φιλοσοφία του αλγορίθμου του συστήματος που περιγράφεται σε αυτήν την πτυχιακή εργασία.



Εικόνα 86: Σχεδιάγραμμα αλγορίθμου συστήματος

### 4.3.2 Περιγραφή υλοποίησης συστήματος

Για να εκτελέσουμε το σύστημα πρέπει πρώτα να γίνουν ορισμένες ενέργειες. Αρχικά πρέπει να είναι σίγουρο ότι το υπολογιστικό σύστημα πλήρη τις προϋποθέσεις όπως αυτές αναφέρονται παραπάνω. Εφόσον έχουν ολοκληρωθεί οι παραπάνω ενέργειες το σύστημα μπορεί να εκτελεστεί με δυο τρόπους. Ο πρώτος τρόπος είναι η χρήση του πρώτου κομματιού κώδικα που αναφέρετε παραπάνω για την παραγωγή ενός νέου τεχνητού νευρωνικού δικτύου για περαιτέρω χρήση στο σύστημα. Ο κώδικας μπορεί να βρεθεί στον φάκελο ANN train στο αρχείο ann\_adv.m. Επίσης ο χρήστης μπορεί να πειραματιστεί και με διάφορες βάσεις δεδομένων που χρησιμοποιήθηκαν κατά την έρευνα των κατάλληλων ρυθμίσεων του τεχνητού νευρωνικού δικτύου. Αν ο χρήστης θέλει μια πιο γρήγορη εκτέλεση δεν χρειάζεται να στήσει ξανά το τεχνητό νευρωνικό δίκτυο.

Ο δεύτερος και πιο απλός τρόπος είναι να εκτελέσουμε απευθείας τον κώδικα δίχως την δημιουργία ενός νέου τεχνητού νευρωνικού δικτύου αλλά να χρησιμοποιηθεί το προκατασκευασμένο. Είτε με την χρήση νέου είτε με την χρήση του προκαθορισμένου τεχνητού νευρωνικού δικτύου πρέπει να γίνει μια αρχικοποίηση πριν τρέξει ο κώδικας. Στις πρώτες γραμμές του κώδικα του αρχείου system\_v\_2\_0.m στον φάκελο Πτυχιακή Τελικό εμφανίζετε μια συνάρτηση η stereomatching( ) όπου είναι προσωπική μετατροπή του κώδικα match v3.1 του Vladimir Kolmogorov [18] που ήταν γραμμένος σε C. Έτσι πρέπει πρώτα να εκτελεστεί το αρχείο Build\_code.m το οποίο αρχικοποιεί τις βιβλιοθήκες και τις συναρτήσεις για την χρήση του συστήματος.

Αφού πλέον το σύστημα μας έχει αρχικοποιηθεί εντελώς μπορούμε να συνεχίσουμε με την εκτέλεση του ανοίγοντας το αρχείο system\_v\_2\_0.m. Αρχικά στην πρώτη γραμμή του κώδικα εμφανίζετε η εντολή imread η οποία διαβάζει την εικόνα την οποία θα χρησιμοποιήσουμε αργότερα μέσα στο σύστημα. Η δεύτερη εντολή είναι η συνάρτηση stereomatching() όπου μας δίνει σαν αποτελέσματα δυο πινάκες βάθους ο ένας είναι χωρίς την εμφάνιση των κρυφών σημείων (εικόνα 87) και ο δεύτερος είναι με την εμφάνιση των κρυφών σημείων (εικόνα 81).

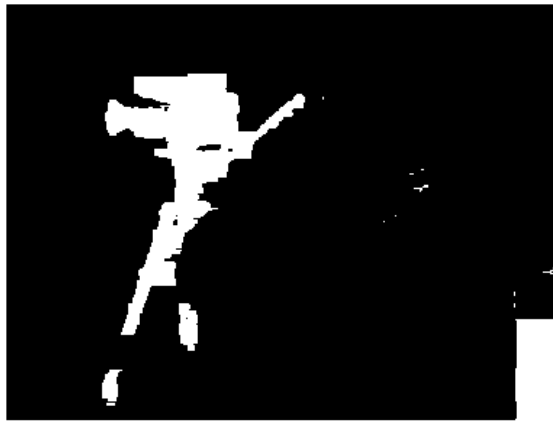


Εικόνα 87: Χάρτης βάθους του πειραματικού ζεύγους εικόνων

Αυτό που μας ενδιαφέρει είναι ο χάρτης βάθους που εμφανίζετε παραπάνω. Αρχικές παρατηρήσεις σε αυτό το επίπεδο είναι ότι με την εξαγωγή των βαθών και την ξεχωριστή τους επεξεργασία μπορεί να αποφέρει τα επιθυμητά αποτελέσματα όσο αφορά την εύρεση των αντικειμένων. Έτσι στην συνέχεια έχουμε ένα κομμάτι εύρεσης των βαθών τα οποία περιέχουν κάποιο αντικείμενο. Έτσι για κάθε βάθος  $L \in D$  όπου  $D$  είναι ο χάρτης βάθους, υπάρχει αριθμός εικονοστοιχείων  $N_L$ :

$$N_L = \sum_i \sum_j p(i, j), \text{ όπου } p(i, j) = \begin{cases} 1 & \text{if } p(i, j) \in L \\ 0 & \text{if } p(i, j) \notin L \end{cases}$$

με  $i$  το πλάτος και  $j$  το ύψος της εικόνας. Έτσι με τον ορισμό ενός ελαχίστου αριθμού εικονοστοιχείων  $T_D$ , ο οποίος είναι απαραίτητος για την απομάκρυνση μη χρήσιμων βαθών, τα βάθη τα οποία παίρνουν την διαδικασία  $N_L > T_D$  ονομάζονται υποψηφία επίπεδα βάθους. Με αυτήν την διαδικασία λάβαμε σαν αποτέλεσμα από την πειραματική εικόνα τα εξής βάθη :



Εικόνα 88: Βάθος 1 από επεξεργασμένο χάρτη βάθους



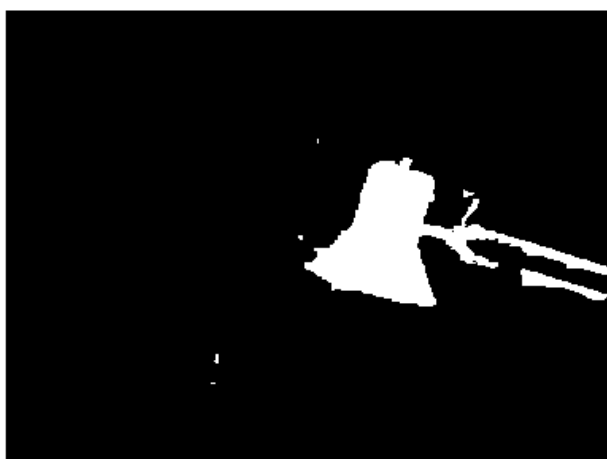
*Εικόνα 89: Βάθος 2 από επεξεργασμένο χάρτη βάθους*



*Εικόνα 90: Βάθος 3 από επεξεργασμένο χάρτη βάθους*



*Εικόνα 91: Βάθος 4 από επεξεργασμένο χάρτη βάθους*



*Εικόνα 92: Βάθος 5 από επεξεργασμένο χάρτη βάθους*

Όπως παρατηρήθηκε από τα αποτελέσματα αυτής της διαδικασίας τα αντικείμενα τώρα βρίσκονται μέσα στα βάθη και περιγράφονται όλα σε ξεχωριστό βάθος. Σε μια περίπτωση αυτή του αγάλματος υπάρχει ένα πρόβλημα. Παρόλο που το αντικείμενο έχει ανιχνευτεί αυτό έχει γίνει σε δυο διαφορετικά βάθη. Έτσι στην συνέχεια κατασκευάζετε το επόμενο κομμάτι κώδικα το οποίο ενώνει τα επίπεδα τα οποία είναι πάρα πολύ κοντά. Το αποτέλεσμα αυτού του μέρους είναι εμφανές μόνον στην εικόνα του αγάλματος, έτσι ως αποτέλεσμα έχουμε :





Εικόνα 93: Ενοποίηση βάθους 3 και 4

Στην συνέχεια του αλγόριθμου γίνεται επεξεργασία του κάθε υποψήφιου επιπέδου για την εύρεση των αντικειμένων μέσα σε αυτό. Έτσι χρησιμοποιείτε ο αλγόριθμος συνδεδεμένων στοιχείων για την εύρεση των αντικειμένων σε κάθε υποψήφιο επίπεδο. Το αποτέλεσμα του αλγόριθμου είναι η ομαδοποίηση περιοχών τις εικόνας που τα ονομάζονται υποψήφια αντικείμενα. Ο αριθμός  $N_o$  που αποτελεί ένα υποψήφιο αντικείμενο καθορίζετε ως :

$$N_o = \sum_i \sum_j p(i, j), \text{ όπου } p(i, j) = \begin{cases} 1 & \text{if } p(i, j) \in O \\ 0 & \text{if } p(i, j) \notin O \end{cases}$$

όπου  $O$  θεωρείτε το αντικείμενο που εξετάζετε. Αυτή η πληροφορία για κάθε υποψήφιο αντικείμενο  $O$  χρησιμοποιείτε από το σύστημα για να καθορίσει εάν το υποψήφιο αντικείμενο είναι όντως πραγματικό συγκρίνοντας το  $N_o$  με ένα προκαθορισμένο όριο  $T_o$ . Αυτό το όριο καθορίζει τον ελάχιστο αριθμό εικονοστοιχείων για να θεωρηθεί το αντικείμενο ως πραγματικό. Όσα από αυτά τα αντικείμενα κατηγοριοποιηθούν ως αληθινά αντικείμενα συλλέγονται από τις συντεταγμένες όπου βρίσκονται με τις πληροφορίες της εντολής `regionprops` και την χρήση της `imcrop`. Ένα παράδειγμα είναι η παρακάτω εικόνα :



*Εικόνα 94: Παράδειγμα συλλογής αντικειμένου από επίπεδο βάθους*

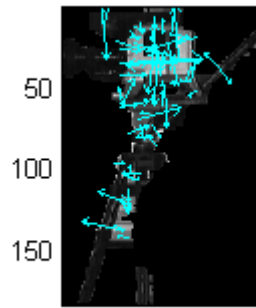
αφού γίνει η συλλογή του αντικειμένου επίσης γίνεται και η αντιστοίχιση του στην κανονική εικόνα που διαβάσαμε στην αρχή του συστήματος, παράδειγμα της λειτουργίας αυτής είναι :



*Εικόνα 95: Παράδειγμα αντιστοίχισης αντικειμένου στην πραγματική του εικόνα*

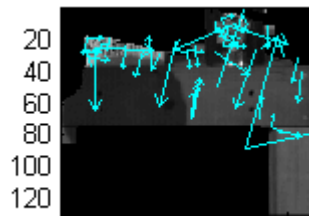
έτσι πλέον έχουμε όλα τα αντικείμενα από την εικόνα και έτοιμα για περαιτέρω ανάλυση.

Στην συνέχεια έχουμε την εξαγωγή των χαρακτηριστικών μέσω του μετασχηματισμού εξαγωγής χαρακτηριστικών αμετάβλητων στην κλίμακα (Scale-Invariant Feature Transform, SIFT). Για την χρήση του αλγορίθμου αυτού κάθε εικόνα θα πρέπει να μετατρέπεται σε ξεχωριστή εικόνα μέσω της εντολής `imwrite`. Έτσι τώρα γίνεται η εισαγωγή του κάθε αντικείμενου, η επεξεργασία του και τέλος η εξαγωγή και αποθήκευση των σημείων κλειδιών καθώς και των τοπικών περιγραφέων τους. Έτσι το αποτέλεσμα από τα αντικείμενα της πειραματικής εικόνας είναι :



2040608010020

*Εικόνα 96: Σημεία κλειδιά μετασχηματισμού SIFT αντικείμενο 1*



50 100 150

*Εικόνα 97: Σημεία κλειδιά μετασχηματισμού SIFT αντικείμενο 2*



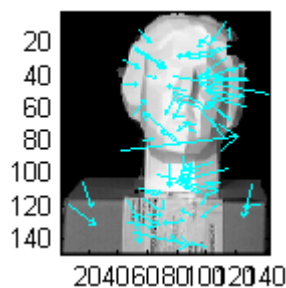
20 40

*Εικόνα 98: Σημεία κλειδιά μετασχηματισμού SIFT αντικείμενο 3*

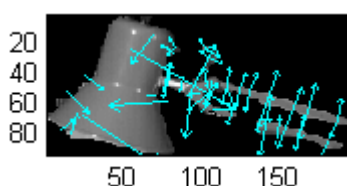


5 10 15 20 25

*Εικόνα 99: Σημεία κλειδιά μετασχηματισμού SIFT αντικείμενο 4*



Εικόνα 100: Σημεία κλειδιά μετασχηματισμού SIFT αντικείμενο 5



Εικόνα 101: Σημεία κλειδιά μετασχηματισμού SIFT αντικείμενο 6

Τώρα πλέον έχουμε μια βάση δεδομένων την οποία μπορεί να επεξεργαστεί το τεχνητό νευρωνικό δίκτυο. Το τεχνητό νευρωνικό δίκτυο είτε πρέπει να εκπαιδευτεί από τον χρήστη όπως αναφέρθηκε παραπάνω είτε να εισαχθεί στο σύστημα με την εντολή load και την διεύθυνση ενός από τα έτοιμα τεχνητά νευρωνικά δίκτυα τα οποία βρίσκονται στον φάκελο ANN train σε μορφή αρχείων .mat. Έτσι αφού πλέον έχουμε την βάση δεδομένων μας και το τεχνητό νευρωνικό δίκτυο μπορούμε να περάσουμε στην χρήση και την επαλήθευση των αποτελεσμάτων. Αλλάζοντας κάθε φορά τον αριθμό x που βρίσκετε στην αρχή του αρχείου eval.m αλλάζουμε και το αντικείμενο το οποίο θέλουμε κάθε φορά να ερευνήσουμε. Έτσι για γνωστή βάση δεδομένων με σημεία κλειδιά του μετασχηματισμού SIFT που ορίζετε ως  $D: \{C_n, n=1,2,3, \dots\}$ , όπου το  $C_n$  είναι οι κατηγορίες των αντικείμενων που χρησιμοποιούνται στην αναγνώριση και  $n$  είναι ο αριθμός των διαφορετικών κατηγοριών. Το αποτέλεσμα του τεχνητού νευρωνικού δικτύου είναι  $\max(Pr(O \cap C_n))$  που είναι η μέγιστη πιθανότητα ταιριάσματος με μια γνωστή κατηγορία. Το τελικό αποτέλεσμα είναι τέσσερις αριθμοί οι οποίοι εξηγούνται αναλυτικά στο επόμενο κεφάλαιο.

## 5. Αποτελέσματα

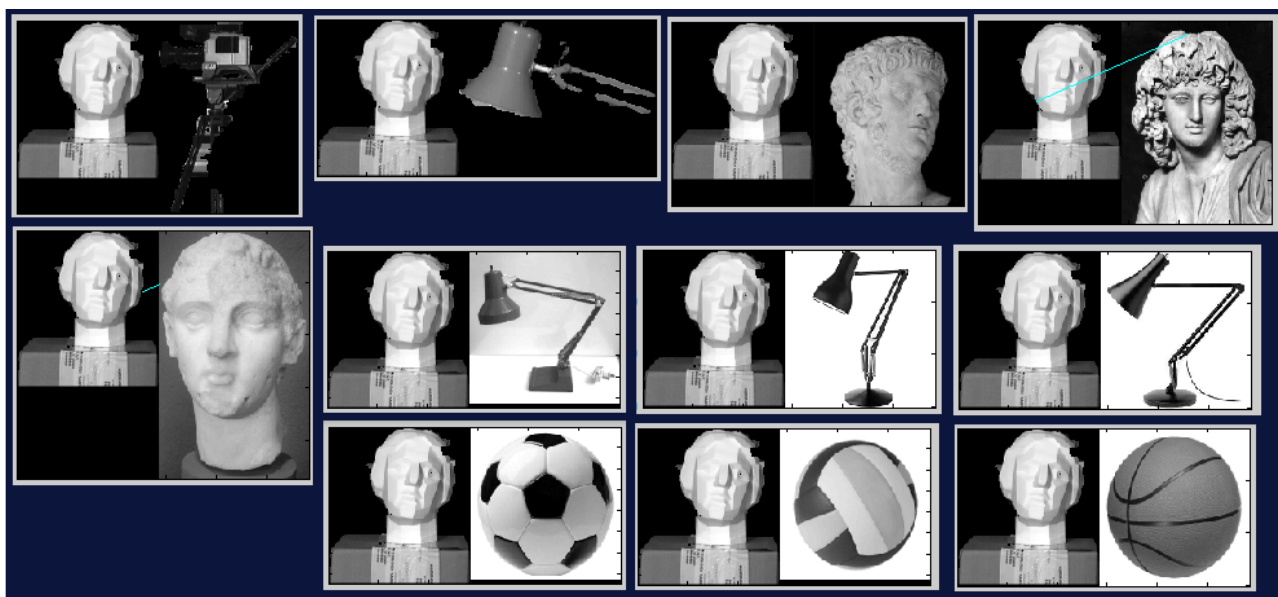
Για την απόδειξη των αποτελεσμάτων του συστήματος αυτού για την ικανότητα του να ταιριάζει άγνωστα αντικείμενα με γνωστές κατηγορίες αντικείμενων δημιουργήθηκε βάση δεδομένων με παρόμοια αλλά όχι ίδια αντικείμενα. Η βάση δεδομένων που χρησιμοποιήθηκε για τον πειραματισμό των αποτελεσμάτων έγινε με την επιλογή των εικόνων και την εξαγωγή των χαρακτηριστικών τους μέσω του μετασχηματισμού SIFT. Επίσης από την χρήση του πειραματικού ζεύγους στερεοσκοπικών εικόνων εξάχθηκαν έξι κύρια αντικείμενα από τα οποία μόνο τα τρία χρησιμοποιήθηκαν για την απόδειξη της αποτελεσματικότητας του συστήματος,

και αυτά είναι η κάμερα, το άγαλμα και η λάμπα.

Αρχικά σε πειραματικό ακόμα στάδιο χρησιμοποιήθηκε ένας έτοιμος αλγόριθμος ο οποίος ήταν προτεινόμενος με την χρήση του μετασχηματισμού SIFT. Αυτός ήταν ένας γρήγορος αλγόριθμος του κοντινότερου γείτονα. Αυτός ο αλγόριθμος υπολόγιζε την απόσταση μεταξύ των τιμών ανάμεσα στα σημεία κλειδιά του ελεγχόμενου αντικειμένου με σημεία κλειδιά μέσα από την βάση δεδομένων. Και το αποτέλεσμα ήταν ένα ταίριασμα μεταξύ των δυο σημείων στην εικόνα. Αυτός ο αλγόριθμος είχε προηγουμένως χρησιμοποιηθεί από τον δημιουργό του μετασχηματισμού SIFT για την αναγνώριση ίδιων αντικειμένων μέσα σε μια εικόνα. Παρόλα αυτά για την απόκλιση κάθε αμφιβολίας παραθέτονται τα αποτελέσματα του αλγορίθμου, για την εύρεση παρομοίων αντικειμένων :

<b>Πίνακας αποτελεσμάτων (1)</b>				
<b>Αντικείμενα</b>	<b>Σημεία κλειδιά που έχουν ταιριάζει</b>	<b>Συνολικός αριθμός σημείων κλειδιών</b>	<b>Ποσοστό ταιριάσματος (%)</b>	<b>Κατηγοριοποίηση ως</b>
Αντικείμενο κάμερα	0	100	0	τίποτα
Αντικείμενο άγαλμα	2	88	1,76	άγαλμα
Αντικείμενο λάμπα	0	48	0	τίποτα

Πίνακας 1: Πινάκας αποτελεσμάτων ταιριάσματος SIFT με την χρήση γρήγορου αλγόριθμου του κοντινότερου γείτονα



Εικόνα 102: Οπτική αναπαράσταση αποτελεσμάτων του πίνακα 1

Με την επιβεβαίωση των παραπάνω αποτελεσμάτων είναι πλέον εμφανές ότι η χρήση του γρήγορου αλγορίθμου του κοντινότερου γείτονα δεν είναι ο κατάλληλος για την χρήση σε παρόμοια αντικείμενα. Έτσι πάρθηκε η απόφαση για την χρήση ενός τεχνητού νευρωνικού δικτύου. Τα πρώτα αποτελέσματα δεν είχαν την αναμενόμενη απόδοση όπως φαίνεται και στον παρακάτω πίνακα.

**Πινάκας αποτελεσμάτων (2)**

Αντικείμενα	Σημεία κλειδιά που έχουν ταιριάζει	Συνολικός αριθμός σημείων κλειδιών	Ποσοστό ταιριάσματος (%)	Κατηγοριοποίηση ως
Αντικείμενο κάμερα	74	100	74	άγαλμα
Αντικείμενο άγαλμα	64	88	72,73	άγαλμα
Αντικείμενο λάμπα	37	48	77,08	άγαλμα

*Πίνακας 2: Πίνακας αποτελεσμάτων ταιριάσματος SIFT με την χρήση τεχνητού νευρωνικού δικτύου (πρώτη προσπάθεια)*

Τα παραπάνω αποτελέσματα έδειξαν ότι τα αντικείμενα δεν είχαν κατηγοριοποιηθεί πλήρως στις σωστές κατηγορίες αντικειμένων. Ο λόγος της λανθασμένης κατηγοριοποίησης της κάμερας ως άγαλμα είναι κατανοητός λόγο του μη αρχικού ορισμού μιας κατηγορίας καμερών. Δηλαδή το σύστημα δεν είχε στην βάση δεδομένων του κάποια κατηγορία με κάμερες, έτσι λόγω της ιδιότητας του τεχνητού νευρωνικού δικτύου κατηγοριοποιήθηκαν λάθος. Ο λόγος που έγινε λανθασμένη κατηγοριοποίηση της λάμπας ήταν το γεγονός ότι τα χαρακτηριστικά που υπήρχαν στην κατηγορία με τις λάμπες δεν ήταν αρκετά και είχαν ασκήσει μικρή επιρροή κατά την εκπαίδευση του συστήματος. Έτσι αποκομίσθηκαν δυο σημαντικά στοιχεία για την περαιτέρω ανάπτυξη του συστήματος. Το πρώτο είναι ότι πρέπει να υπάρχει μηχανισμός επίβλεψης και ανατροφοδότησης από τον χρήστη για να συμβάλει στην εκπαίδευση του συστήματος. Και το δεύτερο είναι ότι το πλήθος των σημείων κλειδιών που αποτελούν μια κατηγορία αντικειμένων είναι σημαντικό κομμάτι στην εκπαίδευση και αποτελεσματικότητα του συστήματος. Έτσι έχοντας τα παραπάνω ως γνώμονα έγινε αρχικά τοποθέτηση μιας νέας κατηγορίας “κάμερα” στην βάση δεδομένων, και μεγέθυνση στην κατηγορία “λάμπα”. Έτσι τα αποτελέσματα διαμορφώθηκαν ως εξής :

<b><u>Πίνακας αποτελεσμάτων (3)</u></b>				
Αντικείμενα	Σημεία κλειδιά που έχουν ταιριάζει	Συνολικός αριθμός σημείων κλειδιών	Ποσοστό ταιριάσματος (%)	Κατηγοριοποίηση ως
Αντικείμενο κάμερα	80	100	80	κάμερα
Αντικείμενο άγαλμα	77	88	87,5	άγαλμα
Αντικείμενο λάμπα	38	48	79,17	λάμπα

*Πίνακας 3: Πίνακας αποτελεσμάτων ταιριάσματος SIFT με την χρήση τεχνητού νευρωνικού δικτύου (δεύτερη προσπάθεια)*

Η τελική απόδειξη της ορθής λειτουργίας του συστήματος έπρεπε να γίνει με άγνωστα στο σύστημα αντικείμενα. Έτσι για χρησιμοποιήθηκαν τρεις άγνωστες εικόνες από λάμπες όπως εμφανίζονται στην παρακάτω εικόνα.



Εικόνα 103: Τρεις άγνωστες εικόνες από λάμπες που χρησιμοποιήθηκαν στην τελική επαλήθευση των αποτελεσμάτων

Σε αυτές εφαρμόστηκε ο μετασχηματισμός SIFT για την εξαγωγή των χαρακτηριστικών τους. Αρχικά χρησιμοποιήθηκαν μόνον οι δυο και τα αποτελέσματα δεν ήταν πλήρως επιτυχή. Η μια μόνο από τις δυο λάμπες κατηγοριοποιήθηκαν σωστά στην κατηγορία “λάμπα”.

<u>Πίνακας αποτελεσμάτων (4)</u>				
Αντικείμενα	Σημεία κλειδιά που έχουν ταιριάζει	Συνολικός αριθμός σημείων κλειδιών	Ποσοστό ταιριάσματος (%)	Κατηγοριοποίηση ως
Άγνωστο αντικείμενο λάμπα 1	64	125	50,79	λάμπα
Άγνωστο αντικείμενο λάμπα 2	58	136	42,65	άγαλμα

Πίνακας 4: Πίνακας αποτελεσμάτων δοκιμής με άγνωστα αντικείμενα

Έτσι έγινε επανεκπαίδευση του συστήματος και επανεκτίμηση των δεδομένων. Τα τελικά αποτελέσματα δικαιώνουν τις παραπάνω εκτίμησης για την χρήση συστήματος ανατροφοδότησης και υποβοηθούμενης από τον χρήστη εκπαίδευσης όπως φαίνεται στον παρακάτω πίνακα.

<b>Πινάκας αποτελεσμάτων (5)</b>				
<b>Αντικείμενα</b>	<b>Σημεία κλειδιά που έχουν ταιριάζει</b>	<b>Συνολικός αριθμός σημείων κλειδιών</b>	<b>Ποσοστό ταιριάσματος (%)</b>	<b>Κατηγοριοποίηση ως</b>
Άγνωστο αντικείμενο λάμπα 1	105	126	83,33	λάμπα
Άγνωστο αντικείμενο λάμπα 2	107	136	78,68	λάμπα
Άγνωστο αντικείμενο λάμπα 3	138	310	44,51	λάμπα
Αντικείμενο κάμερα	70	100	70	κάμερα
Αντικείμενο άγαλμα	81	88	92,5	άγαλμα
Αντικείμενο λάμπα	36	48	75	λάμπα

Πίνακας 5: Πίνακας τελικών αποτελεσμάτων

## 5.1 Συμπεράσματα και μελλοντική εργασία

Τα συμπεράσματα τα οποία εξάγονται από τα παραπάνω αποτελέσματα είναι ότι το σύστημα που περιγράφετε σε αυτήν την πτυχιακή είναι άρτιο και ολοκληρωμένο σε κάθε λειτουργία του. Η χρήση μεθόδων πρωτοποριακών στην επιστήμη υπολογιστών και στα πεδία της τεχνητής όρασης και τεχνητής νοημοσύνης έχει οδηγήσει το σύστημα σε εκπληκτικά αποτελέσματα. Εάν το σύστημα έχει αρκετά μεγάλη βάση δεδομένων και την υποβοηθήσει του χρήστη θα είναι σε θέση, με την συνεχή διεύρυνση των γνώσεων του, να οδηγηθεί σε πλήρες αυτόνομο σύστημα με ελάχιστα λάθη.

Στο μέλλον το επεκτάσεις του συστήματος μπορούν να γίνουν πάνω σε πολλούς τομείς. Αρχικά στον τομέα της βάσης δεδομένων όπου με την χρήση ανεπτυγμένων τεχνικών διαχείρισης το σύστημα να είναι σε θέση να απαντήσει και να αντιδράσει σε συγκεκριμένες απαιτήσεις του χρήστη. Επίσης δουλειά μπορεί να γίνει στην κατασκευή ενός διαδραστικού περιβάλλοντος το οποίο θα κάνει ποιο εύκολη την χρήση του συστήματος. Η έρευνα σε νέες μεθόδους εκπαίδευσης και αλγορίθμων τεχνητής νοημοσύνης μπορούν να οδηγήσουν σε ένα καλύτερο σε αποτελέσματα σύστημα. Αναβάθμιση του συστήματος θα εκπληρωνόταν με την χρήση νέων μεθόδων που είτε θα εξαλειφόταν την χρήση δισδιάστατων χαρακτηριστικών είτε θα καλυτερεύσουν την αποτελεσματικότητά τους. Τέλος, η ποιο μεγάλη αλλαγή θα αποτελούσε η χρήση οντολογιών για την περιγραφή των κατηγοριών και των αντικειμένων που ανήκουν σε αυτές.



## Βιβλιογραφία

- 1: Wikipedia, Visual system, [http://en.wikipedia.org/wiki/Visual\\_system](http://en.wikipedia.org/wiki/Visual_system),
- 2: Wikipedia, Visual perception, [http://en.wikipedia.org/wiki/Visual\\_perception](http://en.wikipedia.org/wiki/Visual_perception),
- 3: Wikipedia, Computer vision, [http://en.wikipedia.org/wiki/Computer\\_vision](http://en.wikipedia.org/wiki/Computer_vision),
- 4: Wikipedia, Artificial intelligence, [http://en.wikipedia.org/wiki/Artificial\\_intelligence](http://en.wikipedia.org/wiki/Artificial_intelligence),
- 5: Wikipedia, Machine learning, [http://en.wikipedia.org/wiki/Machine\\_learning](http://en.wikipedia.org/wiki/Machine_learning),
- 6: Wikipedia, Stereopsis, <http://en.wikipedia.org/wiki/Stereopsis>,
- 7: Wikipedia, Graph cuts in computer vision,  
[http://en.wikipedia.org/wiki/Graph\\_cuts\\_in\\_computer\\_vision](http://en.wikipedia.org/wiki/Graph_cuts_in_computer_vision),
- 8: Wikipedia, Connected component,  
[http://en.wikipedia.org/wiki/Connected\\_component\\_%28graph\\_theory%29](http://en.wikipedia.org/wiki/Connected_component_%28graph_theory%29),
- 9: Wikipedia, Scale Invariant Feature Transform, [http://en.wikipedia.org/wiki/Scale-invariant\\_feature\\_transform](http://en.wikipedia.org/wiki/Scale-invariant_feature_transform),
- 10: Dictionary.com, Methodology, <http://dictionary.reference.com/browse/methodology>,
- 11: Wikipedia, Depth perception, [http://en.wikipedia.org/wiki/Depth\\_perception](http://en.wikipedia.org/wiki/Depth_perception),
- 12: Wikipedia, Graph theory, [http://en.wikipedia.org/wiki/Graph\\_theory](http://en.wikipedia.org/wiki/Graph_theory),
- 13: Wikipedia, Cut (graph theory), [http://en.wikipedia.org/wiki/Cut\\_%28graph\\_theory%29](http://en.wikipedia.org/wiki/Cut_%28graph_theory%29),
- 14: Yuri Boykov and Vladimir Kolmogorov, An Experimental Comparison of Min-Cut/Max-Flow Algorithms for Energy Minimization in Vision, IEEE Transactions on PAMI, Vol. 26, No. 9, pp. 1124-1137, Sept. 2004,
- 15: Wikipedia, Ford–Fulkerson algorithm, [http://en.wikipedia.org/wiki/Ford-Fulkerson\\_algorithm](http://en.wikipedia.org/wiki/Ford-Fulkerson_algorithm),
- 16: Wikipedia, Binocular disparity, [http://en.wikipedia.org/wiki/Binocular\\_disparity](http://en.wikipedia.org/wiki/Binocular_disparity),
- 17: Shawn Lankton, 3D vision and stereo disparity, <http://www.shawnlankton.com/2007/12/3d-vision-with-stereo-disparity/>,
- 18: Vladimir Kolmogorov Ramin Zabih, Computing Visual Correspondence with Occlusions via Graph Cuts,
- 19: Wikipedia, Pott's model, [http://en.wikipedia.org/wiki/Potts\\_model](http://en.wikipedia.org/wiki/Potts_model),
- 20: Wikipedia, Connected component labeling,  
[http://en.wikipedia.org/wiki/Connected\\_component\\_labeling](http://en.wikipedia.org/wiki/Connected_component_labeling),
- 21: David G.Lowe, Distinctive Image Features from Scale-Invariant Keypoints,
- 22: Wikipedia, Artificial neural network, [http://en.wikipedia.org/wiki/Artificial\\_neural\\_network](http://en.wikipedia.org/wiki/Artificial_neural_network),
- 23: Wikipedia, Computer stereo vision, [http://en.wikipedia.org/wiki/Computer\\_stereo\\_vision](http://en.wikipedia.org/wiki/Computer_stereo_vision),