

Ανακατασκευή τρισδιάστατου σχήματος από πολλαπλές προβολές-Σύγκριση μοντέλων ενέργειας

Theocharis Plioukas
Τμήμα Μηχανικών Πληροφορικής,
Υπολογιστών και Τηλεπικοινωνιών
Διεθνές Πανεπιστήμιο Ελλάδος
Σέρρες, Ελλάδα



Abstract – Η εργασία αυτή εμβαθύνει στον τομέα της τρισδιάστατης ανακατασκευής, αξιοποιώντας ένα οπλοστάσιο από ανιχνευτές και περιγραφείς τελευταίας τεχνολογίας, συμπεριλαμβανομένων των Difference of Gaussians (DoG), KeyNet, Corner Harris with Good Features to Track (CornerGFTT), HardNet, SosNet και SIFT. Χρησιμοποιούνται αυτοί οι αλγόριθμοι για την εξαγωγή σημαντικών χαρακτηριστικών από μια συλλογή εικόνων 2D, θέτοντας τις βάσεις για την ακριβή ανακατασκευή τρισδιάστατης σκηνής. Δύο διαφορετικοί αλγόριθμοι απόρριψης ακραίων στοιχείων, ο DISC και ο AdaLAM, διερευνώνται και αξιολογούνται στο πλαίσιο της βιβλιοθήκης COLMAP, διευκολύνοντας μια ολοκληρωμένη σύγκριση. Οι ανιχνευτές και οι περιγραφείς παίζουν καθοριστικό ρόλο στον εντοπισμό και την περιγραφή διακριτών βασικών σημείων στις εικόνες εισόδου. Οι ανιχνευτές DoG, KeyNet και CornerGFTT παρέχουν ποικίλες επιλογές για την ανίχνευση χαρακτηριστικών, ο καθένας με τα μοναδικά του πλεονεκτήματα. Ομοίως, οι περιγραφείς HardNet, SosNet και SIFT προσφέρουν διαφορετικά επίπεδα ευρωστίας και διακριτικής ικανότητας, ανταποκρινόμενοι σε διαφορετικές σκηνές και προκλήσεις. Η καρδιά αυτής της μελέτης έγκειται στη σύγκριση και αξιολόγηση αυτών των ανιχνευτών και των περιγραφέων καθώς γίνονται πειράματα για την εξαγωγή συμπερασμάτων με όλους τους δυνατούς συνδυασμούς αυτών. Τα πειράματα διεξάγονται με 2000 και με 8000 χαρακτηριστικά σημεία. Ο DISK, μια καινοτόμος προσέγγιση που βασίζεται στην ελαχιστοποίηση της ενέργειας, επιδιώκει την ευθυγράμμιση των εικόνων απευθείας στον τρισδιάστατο χώρο, ενώ ο AdaLAM αξιοποιεί τους περιορισμούς τοπικής συνέπειας για την ισχυρή απόρριψη ακραίων τιμών με παραλληλισμένο τρόπο φιλικό προς τις GPU. Αντιπαραβάλλοντας αυτούς τους δύο αλγόριθμους μεταξύ τους, το παρόν έργο ρίχνει φως στα αντίστοιχα πλεονεκτήματα και μειονεκτήματά τους στο πλαίσιο της τρισδιάστατης ανακατασκευής. Η βιβλιοθήκη COLMAP χρησιμεύει ως πεδίο δοκιμών, επιτρέποντας την αυστηρή αξιολόγηση σε διάφορα σύνολα δεδομένων και σεναρία. Τα αποτελέσματα αυτής της μελέτης προσφέρουν πολύτιμες πληροφορίες σχετικά με την απόδοση, την αποδοτικότητα και την καταλληλότητα αυτών των αλγορίθμων για εργασίες τρισδιάστατης ανακατασκευής. Εν τέλει, το έργο αυτό συμβάλλει στην πρόοδο του τομέα της όρασης υπολογιστών, παρέχοντας μια αποχρωματισμένη κατανόηση της αλληλεπίδρασης μεταξύ ανιχνευτών, περιγραφών και μεθόδων απόρριψης ακραίων τιμών.

Εισαγωγή

Η ανακατασκευή τρισδιάστατου σχήματος από πολλαπλές προβολές είναι η διαδικασία δημιουργίας ενός τρισδιάστατου μοντέλου ή μιας τρισδιάστατης σκηνής από δισδιάστατες εικόνες ή βίντεο. Έχει μια πολύ μεγάλη γκάμα εφαρμογών σε τομείς όπως η αρχιτεκτονική, η ρομποτική, η ιατρική, η ψυχαγωγία και πολλούς άλλους. Στο πεδίο της ρομποτικής όρασης έχει ποικίλες εφαρμογές όπως η εικονική πραγματικότητα, τα αυτόνομα οχήματα, η ιατρική απεικόνιση και πολλά άλλα. Οι παραδοσιακές μέθοδοι για την 3D ανακατασκευή απαιτούσαν πολύ χρόνο και δύσκολη χειροκίνητη εργασία όπως τη φωτομετρία και το laser scanning. Ωστόσο με την ανάπτυξη της τεχνητής νοημοσύνης και των νευρωνικών δικτύων και τη χρήση αλγορίθμων Deep Learning, η 3D ανακατασκευή έχει γίνει πιο εύκολη, πιο αυτοματοποιημένη και πιο ακριβής από ποτέ.

Ένα πολύ σημαντικό στοιχείο στη 3D ανακατασκευή που επηρεάζει άμεσα την ποιότητα και την ακρίβεια του

αποτελέσματος είναι τα μοντέλα ενέργειας (energy models) που εφαρμόζονται. Τα μοντέλα ενέργειας είναι ουσιαστικά τρόποι σύγκρισης του 3D μοντέλου που θέλουμε να φτιάξουμε με το μοντέλο αναφοράς. Όσο μικρότερες είναι οι διαφορές μεταξύ του μοντέλου που κατασκευάστηκε και του μοντέλου αναφοράς, τόσο πιο ακριβές είναι το μοντέλο ενέργειας που χρησιμοποιήθηκε. Για την ελαχιστοποίηση των διαφορών γίνεται ελαχιστοποίηση της κατάλληλης συνάρτησης.

Στην παρούσα εργασία συγκρίνονται 3 μοντέλα ενέργειας αξιοποιώντας τις δυνατότητες που μας δίνει η βιβλιοθήκη COLMAP[1]. Χρησιμοποιούνται έτοιμα datasets τα οποία τροφοδοτούνται στα μοντέλα ενέργειας με σκοπό να συγκριθεί και να αναλυθεί η ποιότητα, η αποτελεσματικότητα, η υπολογιστική αποδοτικότητα και η ευρωστία του κάθε μοντέλου.

I. ΙΣΤΟΡΙΚΗ ΑΝΑΔΡΟΜΗ & RELATED WORK

Πριν την ανάπτυξη του deep learning, τη δεκαετία πριν το 2010, η τρισδιάστατη ανακατασκευή βασιζόταν κυρίως σε τεχνικές όπως το Structure from Motion (SfM) και το Multi-View Stereo (MVS). Αισθητήρες βάθους όπως το Kinect έκαναν πιο δημοφιλές το 3D sensing αλλά παρ' όλα αυτά δεν ήταν ιδιαίτερα ενσωματωμένοι με νευρωνικά δίκτυα έτσι ώστε να παρέχουν μεγαλύτερη ακρίβεια.

Στις αρχές του 2010 οι ερευνητές άρχισαν να πειραματίζονται και να αναζητούν τι δυνατότητες θα μπορούσε να έχει το deep learning στη τρισδιάστατη ανακατασκευή και έτσι το 2014 παρουσιάστηκε το "DeepStereo" το οποίο χρησιμοποίησε Convolutional Neural Networks (CNN's) για την τρισδιάστατη ανακατασκευή σκηνής από ζεύγος εικόνων. Στη συνέχεια, το 2016, εισήχθη το '3D-R2N2' το οποίο πρωτοποριακά χρησιμοποίησε Recurrent Neural Networks (RNNs) για την πρόβλεψη τρισδιάστατων σχημάτων από δισδιάστατες εικόνες.

Στα μέσα περίπου του 2017 οι έρευνες στράφηκαν προς το μονοφθαλμική εκτίμηση βάθους με το 'Monodepth'(2017)[2] και τον διάδοχό του το 'Monodepth2'[3] (2018) να δείχνουν σημαντικά αποτελέσματα στην εκτίμηση χαρτών βάθους από μία μόνο εικόνα. Αυτά τα μοντέλα έθεσαν τα θεμέλια για για εφαρμογές όπως η αυτόνομη οδήγηση, τη ρομποτική, την επαυξημένη πραγματικότητα κ.α.

Στα τέλη της δεκαετίας του 2010 'MVSNet' (2018)[4] βελτίωσε την ανακατασκευή με MVS χρησιμοποιώντας βαθιά δίκτυα, καθιστώντας δυνατή την τρισδιάστατη ανακατασκευή σκηνής από ένα σετ εικόνων. Αντίστοιχα το 2015 με το ShapeNet[5] και το 2018 με το Pixel2Mesh[6] παρουσιάστηκαν μέθοδοι για ανακατασκευή 3D σχημάτων από 2D εικόνες, με εφαρμογές στη τρισδιάστατη σχεδίαση μέσω υπολογιστή και την τρισδιάστατη μοντελοποίηση.

Το 2017 το 'DeepVoxels'[7] έφερε επανάσταση στην ογκομετρική παράσταση σκηνών, κάνοντας δυνατή τη σύλληψη πολύπλοκων τρισδιάστατων δομών. Τρία χρόνια αργότερα το 'NeRF'[8] πρωτοπόρησε εισάγοντας τα neural radiance fields για τη σύνθεση νέων οπτικών της σκηνής και των αντικειμένων με εφαρμογές στον εικονικό τουρισμό και την κατασκευή 3D περιεχομένου.

Οι τωρινές έρευνες από το 2020 και μετά συνεχίζουν να στοχεύουν στη βελτίωση της αποτελεσματικότητας, της

δυνατότητας κλιμάκωσης και των real time εφαρμογών. Παράδειγμα είναι η μέθοδος 'PlenOctrees' η οποία συνδυάζει τα δυνατά στοιχεία των voxel-based και των point-based παραστάσεων για υψηλής ποιότητας ανακατασκευή. Επίσης γίνονται έρευνες για ανάπτυξη εφαρμογών πραγματικού χρόνου που να υποστηρίζονται και από κινητές συσκευές έτσι ώστε το πεδίο της τρισδιάστατης ανακατασκευής να γίνει πιο προσιτό στους καταναλωτές.

II. ΑΛΓΟΡΙΘΜΟΙ

Οι αλγόριθμοι που θα χρησιμοποιηθούν χωρίζονται σε 2 κατηγορίες. Οι αλγόριθμοι με τους οποίους θα γίνει η αναγνώριση των χαρακτηριστικών (features) και των σημείων ενδιαφέροντος (keypoints) σε ένα σετ 2D εικόνων (detectors) και οι αλγόριθμοι με τους οποίους θα γίνει το matching των χαρακτηριστικών και των σημείων ενδιαφέροντος που αναγνωρίστηκαν από τους detectors. Μερικοί detectors είναι οι παρακάτω:

1. Interest Point Detectors όπως οι Harris Corner Detector, FAST και SIFT.
2. Keypoint Detectors όπως οι ORB και AKAZE.
3. Object Detectors όπως οι YOLO και ο Faster R-CNN
4. Depth Sensors όπως τα LiDAR

Στη παρούσα μελέτη θα χρησιμοποιηθούν οι detectors KeyNet, DoG και CornerGFTT.

Ενώ κάποια key points για τους descriptors είναι τα παρακάτω:

1. Feature Representation
2. Matching
3. Invariance (μη μεταβλητότητα)
4. Dimensionality

A. Detectors

DoG (Difference of Gaussians)

Ο αλγόριθμος DoG (Difference of Gaussians) είναι μια πολύ βασική έννοια στη ρομποτική όραση και στην επεξεργασία των εικόνων και χρησιμοποιείται για την αναγνώριση και την ενίσχυση ιδιαίτερων χαρακτηριστικών μέσα σε εικόνες. Λειτουργεί υπολογίζοντας τη διαφορά μεταξύ 2 γκαουσιανά φιλτραρισμένων εικόνων. Εφαρμόζοντας γκαουσιανό φίλτρο για εξομάλυνση μειώνεται ο θόρυβος στην εικόνα και αναγνωρίζεται η κλίμακά της. Αφαιρώντας στη συνέχεια 2 διαφορετικά παραλλαγμένες εκδοχές της ίδιας εικόνας, ο αλγόριθμος ενισχύει τις άκρες και τα χαρακτηριστικά που υπάρχουν σε συγκεκριμένες κλίμακες. Αυτό έχει ως αποτέλεσμα τη λεγόμενη 'πυραμίδα' διαφορετικών εικόνων κάθε μία από τις οποίες τονίζει τις δομές με διαφορετικά επίπεδα λεπτομέρειας. Γενικά ο DoG είναι πολύ χρήσιμος γιατί αναγνωρίζει και εξάγει χαρακτηριστικά και δομές οι οποίες είναι αμετάβλητες σε αλλαγές κλίμακας και φωτισμού.

Ο αλγόριθμος DoG αναδεικνύεται ως μια σημαντική τεχνική στην όραση υπολογιστών, ικανή να ανιχνεύει και να εξάγει σημαντικά χαρακτηριστικά από ψηφιακές εικόνες.

Αυτό το κεφάλαιο ξετυλίγει σχολαστικά τις περιπλοκές της ροής εργασίας του, ξετυλίγοντας το κουβάρι των πιθανών εφαρμογών του και αναγνωρίζοντας τις περιπλοκές των πλεονεκτημάτων και των μειονεκτημάτων του.

Ροή εργασίας

Η ουσία της επιτυχίας του αλγορίθμου DoG έγκειται στη σχολαστικά ενορχηστρωμένη ροή εργασίας του:

1. Κατασκευή χώρου κλίμακας: Το ταξίδι του DoG ξεκινά με τη δημιουργία μιας πυραμίδας χώρου κλίμακας - μιας πολυώροφης δομής που αντικατοπτρίζει την εικόνα εισόδου. Κάθε επίπεδο αυτής της πυραμίδας είναι μια εικόνα που έχει υποστεί δύο θεμελιώδεις λειτουργίες: θόλωση Gauss και υποδειγματοληψία. Η πρώτη τονίζει τις λεπτές λεπτομέρειες, ενώ η δεύτερη διατηρεί αποτελεσματικά τις βασικές δομικές πληροφορίες, δημιουργώντας μια σειρά εικόνων σε διαφορετικές κλίμακες.

2. Υπολογισμός διαφορών: Ενσωματωμένες στην πυραμίδα του χώρου κλίμακας είναι οι εικόνες Difference of Gaussians (DoG), οι οποίες περικλείουν τις μεταβολές της έντασης ή της δομής σε παρακείμενες κλίμακες. Υπολογιζόμενες ως η διαφορά μεταξύ δύο επιπέδων κλίμακας, οι εικόνες αυτές διευκρινίζουν περιοχές εντός της εικόνας που στεγάζουν εμφανείς αλλαγές, περιλαμβάνοντας εξέχοντα χαρακτηριστικά όπως ακμές και σημεία κλειδιά.

3. Ανίχνευση σημείων κλειδιών: Οι εικόνες DoG χρησιμεύουν ως γόνιμο έδαφος για πιθανά σημεία-κλειδιά ή ελάχιστα και μέγιστα. Αυτά τα ακρότατα συμβολίζουν τοποθεσίες όπου η δομή της εικόνας αποκλίνει σημαντικά από το περιβάλλον. Η αξιοσημείωτη προσαρμοστικότητα του αλγορίθμου έγκειται στην ικανότητά του να εντοπίζει αυτά τα ακρότατα σε διάφορες κλίμακες, καθιστώντας τον ανθεκτικό στις μεταβολές του μεγέθους του αντικειμένου και της εγγύτητας της κάμερας.

4. Εντοπισμός σημείων κλειδιών: Για την περαιτέρω βελτίωση των εντοπισμένων βασικών σημείων, θεσπίζονται βήματα εντοπισμού. Μέσω μεθόδων όπως η τετραγωνική παρεμβολή, ο αλγόριθμος εντοπίζει τις ακριβείς θέσεις των σημείων κλειδιών εντός της εικόνας - μια κρίσιμη πτυχή για μεταγενέστερες προσπάθειες όπως η αντιστοίχιση χαρακτηριστικών και η αναγνώριση αντικειμένων.

5. Περιγραφή σημείων κλειδιών: Στην καρδιά κάθε εντοπισμένου σημείου κλειδιού βρίσκεται ένας περιγραφέας, που περικλείει τη σχετική τοπική δομή της εικόνας. Συνήθως, αυτοί οι περιγραφείς ενθυλακώνουν πληροφορίες κλίσης κοντά στο σημείο κλειδί. Δύο ζωτικής σημασίας πτυχές καθορίζουν τη δημιουργία περιγραφέων: ανάθεση προσανατολισμού και υπολογισμός διανύσματος χαρακτηριστικών. Ο πρώτος αποδίδει έναν κυρίαρχο προσανατολισμό στο σημείο κλειδί, εξασφαλίζοντας το αναλλοίωτο σε περιστροφή, ενώ ο δεύτερος δημιουργεί ένα διάνυσμα χαρακτηριστικών - που συχνά περιλαμβάνει δεδομένα μεγέθους κλίσης και προσανατολισμού - που ενισχύει τη διακριτική ικανότητα για την αντιστοίχιση.

Εφαρμογές

Ο αλγόριθμος DoG λάμπει σε ποικίλες εφαρμογές:

Ανίχνευση χαρακτηριστικών: Η αποτελεσματικότητά του στον εντοπισμό χαρακτηριστικών της εικόνας, όπως γωνίες, ακμές και κηλίδες, στηρίζει διάφορες εργασίες υπολογιστικής όρασης, που περιλαμβάνουν αναγνώριση αντικειμένων, παρακολούθηση και συρραφή εικόνων.

Αντιστοίχιση χαρακτηριστικών με αναλλοίωτη κλίμακα: Οι εγγενείς ιδιότητες του χώρου κλίμακας της DoG την εφοδιάζουν με την ικανότητα να εντοπίζει και να ταιριάζει χαρακτηριστικά σε πολλαπλές κλίμακες, καθιστώντας την την επιτομή της προσαρμοστικότητας σε σενάρια που περιλαμβάνουν αντικείμενα σε διάφορες αποστάσεις από την κάμερα.

Πλεονεκτήματα

Αναλλοίωτη κλίμακα: Αυτό αποτελεί πλεονέκτημα σε πραγματικές συνθήκες που χαρακτηρίζονται από δυναμική κλιμάκωση αντικειμένων.

Διακριτικότητα: Οι περιγραφείς χαρακτηριστικών που παράγονται από τον αλγόριθμο είναι ιδιαίτερα διακριτικοί, γεγονός που μεταφράζεται σε ακριβή και ισχυρή αντιστοίχιση χαρακτηριστικών.

Μειονεκτήματα

Υπολογιστική πολυπλοκότητα: Η υπολογιστική όρεξη του DoG μπορεί να είναι σημαντική, ιδίως όταν εφαρμόζεται σε εκτεταμένες εικόνες ή συνεχείς ροές βίντεο.

Ευαισθησία παραμέτρων: Η βέλτιστη απόδοση εξαρτάται συχνά από τη συνετή επιλογή παραμέτρων, μια διαδικασία που απαιτεί προσεκτική βαθμονόμηση.

Ο αλγόριθμος Difference of Gaussians (DoG), μέσω της συνετά δομημένης ροής εργασίας του, αναδεικνύεται σε στέλεχος της όρασης υπολογιστών. Η ικανότητά του στην ανίχνευση και περιγραφή χαρακτηριστικών, σε συνδυασμό με την αναλλοίωτη κλίμακα, τον καθιστά απαραίτητο σε μια σειρά εφαρμογών. Ενώ οι υπολογιστικές απαιτήσεις και η ευαισθησία στις παραμέτρους αποτελούν προκλήσεις, η αποτελεσματικότητά της παραμένει σε σενάρια που απαιτούν ανθεκτική ανίχνευση και αντιστοίχιση χαρακτηριστικών.

Good Features to Track (GFTT)

Ο αλγόριθμος GFTT στέκεται ως ένας κομβικός ακρογωνιαίος λίθος στον τομέα της όρασης υπολογιστών, διαπρέποντας στον εντοπισμό και τον χαρακτηρισμό γωνιών σε ψηφιακές εικόνες. Το παρόν κεφάλαιο επιχειρεί να δια φωτίσει τις ιδιαιτερότητες της ροής εργασίας του, να ξετυλίξει το κουβάρι των πολυάριθμων πιθανών εφαρμογών του και να αξιολογήσει αντικειμενικά τα πλεονεκτήματα και τα μειονεκτήματά του.

Ροή εργασίας:

Στον πυρήνα της αποτελεσματικότητας του αλγορίθμου GFTT βρίσκεται μια σχολαστικά κατασκευασμένη ροή εργασίας:

1. Βαθμίδες εικόνας και τανυστής δομής: Ο GFTT ξεκινά το ταξίδι του με τον υπολογισμό των κλίσεων της εικόνας, αποδίδοντας πληροφορίες κατεύθυνσης σχετικά με τις μεταβολές της έντασης των εικονοστοιχείων. Αυτές οι κλίσεις αξιοποιούνται για τη δημιουργία του τανυστή δομής, ενός κρίσιμου συστατικού στην ανίχνευση γωνιών. Ο τανυστής δομής κωδικοποιεί τοπικές δομικές πληροφορίες, ενισχύοντας την ευαισθησία του αλγορίθμου σε μοτίβα που μοιάζουν με γωνίες.

2. Τοπική άθροιση και αποσύνθεση ιδιοτιμών: Εντός προκαθορισμένων γειτονιών, ο GFTT υπολογίζει τα αθροίσματα των προϊόντων κλίσης, επιτρέποντας στον αλγόριθμο να συσσωρεύει πληροφορίες σχετικά με τις δομικές μεταβολές. Στη συνέχεια, χρησιμοποιείται η αποσύνθεση ιδιοτιμών στον τανυστή δομής, αποδίδοντας δύο ιδιοτιμές που αντιπροσωπεύουν τον προσανατολισμό και την ισχύ της τοπικής δομής.

3. Συνάρτηση απόκρισης γωνίας: Ο ακρογωνιαίος λίθος της ικανότητας ανίχνευσης γωνιών του GFTT είναι η συνάρτηση απόκρισης γωνιών, η οποία ενσωματώνει τις υπολογισμένες ιδιοτιμές. Αυτή η συνάρτηση ποσοτικοποιεί την πιθανότητα ένα εικονοστοιχείο να αποτελεί μέρος μιας γωνίας με βάση τα μεγέθη των ιδιοτιμών.

4. Μη μέγιστη καταστολή: Για να εξασφαλιστεί η επιλογή διακριτών γωνιακών σημείων, το GFTT χρησιμοποιεί μη μέγιστη καταστολή. Περιορίζει τον αριθμό των υποψήφιων γωνιών διατηρώντας μόνο εκείνα με απόκριση γωνίας που υπερβαίνει ένα καθορισμένο κατώφλι.

5. Εντοπισμός σημείου κλειδιού: Ο εντοπισμός σημείου κλειδιού βελτιώνει τα ανιχνευμένα γωνιακά σημεία με εκτίμηση ακρίβειας υπό-pixel. Αυτό ενισχύει την ακρίβεια του αλγορίθμου στον εντοπισμό γωνιακών χαρακτηριστικών εντός της εικόνας.

Εφαρμογές

Η ευελιξία του GFTT το καθιστά απαραίτητο σε διάφορες εφαρμογές υπολογιστικής όρασης:

Παρακολούθηση χαρακτηριστικών: σε σενάρια όπως η παρακολούθηση αντικειμένων, η ανάλυση βίντεο και η εκτίμηση κίνησης.

Συρραφή εικόνων: Παίζει καθοριστικό ρόλο στην ευθυγράμμιση και τη συγχώνευση επικαλυπτόμενων περιοχών εικόνων, διευκολύνοντας τη δημιουργία πανοραμικών φωτογραφιών.

Πλεονεκτήματα

Διακριτικότητα: Το GFTT υπερέχει στη σύλληψη διακριτικών γωνιακών χαρακτηριστικών, τα οποία αποτελούν ισχυρά αναγνωριστικά σε διάφορες εργασίες όρασης υπολογιστών.

Υπολογιστική αποδοτικότητα: Ο υπολογιστικός φόρτος του αλγορίθμου είναι διαχειρίσιμος, καθιστώντας τον κατάλληλο για εφαρμογές πραγματικού χρόνου ή περιορισμένων πόρων.

Μειονεκτήματα

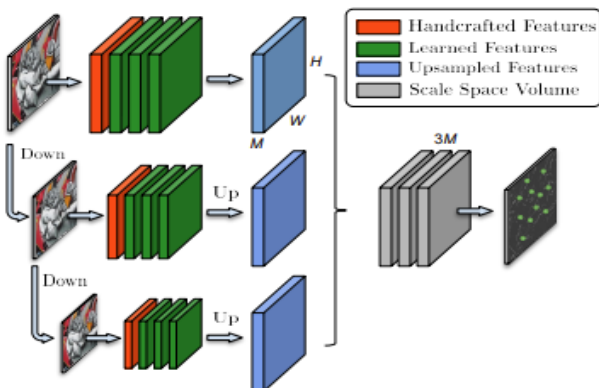
Ευαισθησία στο θόρυβο: Ο GFTT μπορεί να είναι ευαίσθητος στο θόρυβο των εικόνων, οδηγώντας ενδεχομένως σε ψευδώς θετικά αποτελέσματα στην ανίχνευση γωνιών.

Περιορίζεται στις γωνίες: Ως ειδικός ανιχνευτής γωνιών, ο GFTT ενδέχεται να μην υπερέχει στην ανίχνευση άλλων τύπων χαρακτηριστικών.

Ο αλγόριθμος Good Features to Track (GFTT), που προωθείται από την περίπλοκα σχεδιασμένη ροή εργασίας του, αποτελεί ένα απαραίτητο εργαλείο στην όραση υπολογιστών. Η επιδεξιότητά του στην ανίχνευση γωνιών, σε συνδυασμό με την υπολογιστική αποδοτικότητα, τον καθιστά ακρογωνιαίο λίθο στην παρακολούθηση χαρακτηριστικών και τη συρραφή εικόνων. Παρ' όλα αυτά, η ευαισθησία της στο θόρυβο και η εξειδίκευσή της στην ανίχνευση γωνιών δημιουργούν προκλήσεις σε ορισμένα πλαίσια. Συννορίζοντας, η διαρκής σημασία του GFTT υπογραμμίζει τον καίριο ρόλο του στο διαρκώς εξελισσόμενο τοπίο της όρασης υπολογιστών.

KeyNet Algorithm

Ο αλγόριθμος KeyNet έχει σχεδιαστεί για την ανίχνευση keypoints σε εφαρμογές όρασης υπολογιστών. Ο KeyNet συνδυάζει χειροποίητα και μαθημένα χαρακτηριστικά για τη δημιουργία ενός ισχυρού και αποτελεσματικού ανιχνευτή.



Εικόνα 1: Η προτεινόμενη αρχιτεκτονική Key.Net συνδυάζει χειροποίητα και μαθημένα φίλτρα για την εξαγωγή χαρακτηριστικών σε διαφορετικά επίπεδα κλίμακας. Οι χάρτες χαρακτηριστικών αναβαθμίζονται και συνενώνονται. Το τελευταίο φίλτρο που μαθαίνεται συνδυάζει τον όγκο χώρου κλίμακας για να ληφθεί ο τελικός χάρτης απόκρισης.

Το τμήμα "Handcrafted and Learned Filters" του αλγορίθμου KeyNet παίζει καθοριστικό ρόλο στο συνδυασμό των κλασικών τεχνικών όρασης υπολογιστών με τη βαθιά μάθηση. Η χρήση χειροποίητων φίλτρων στο KeyNet αντλεί έμπνευση από καθιερωμένες μεθόδους όπως οι ανιχνευτές Harris και Hessian. Τα χειροποίητα φίλτρα, σε αυτό το πλαίσιο, αναφέρονται σε ένα σύνολο προκαθορισμένων φίλτρων που βασίζονται στις παραγώγους πρώτης και δεύτερης τάξης της εικόνας. Αυτά τα φίλτρα είναι προσεκτικά σχεδιασμένα ώστε να συλλαμβάνουν βασικές πληροφορίες για την εικόνα, όπως οι ακμές και οι

γωνίες. Ένα σημαντικό πλεονέκτημα των χειροποίητων φίλτρων είναι η ερμηνευσιμότητά τους. Δεδομένου ότι αυτά τα φίλτρα είναι προκαθορισμένα, είναι ευκολότερο να κατανοήσουμε πώς συμβάλλουν στην ανίχνευση χαρακτηριστικών. Αυτή η ερμηνευσιμότητα μπορεί να είναι ζωτικής σημασίας για τη λεπτομερή ρύθμιση και την αποσφαλμάτωση του αλγορίθμου.

Από την άλλη πλευρά, το KeyNet ενσωματώνει επίσης μαθημένα φίλτρα μέσω ενός συνελκτικού στρώματος με πολλαπλά φίλτρα, ομαλοποίηση δέσμης και συναρτήσεις ενεργοποίησης ReLU. Αυτά τα στοιχεία αποτελούν ένα μαθημένο μπλοκ μέσα στο δίκτυο. Σε αντίθεση με τα χειροποίητα φίλτρα, τα μαθημένα φίλτρα δεν είναι προκαθορισμένα αλλά βελτιστοποιούνται κατά τη διάρκεια της διαδικασίας εκπαίδευσης. Αυτό επιτρέπει στο δίκτυο να προσαρμόζεται σε συγκεκριμένα σύνολα δεδομένων και να μαθαίνει χαρακτηριστικά που είναι σχετικά με την εργασία.

Ο συνδυασμός χειροποίητων και μαθημένων φίλτρων είναι μια μοναδική πτυχή του KeyNet. Αυτή η υβριδική προσέγγιση αξιοποιεί τα πλεονεκτήματα και των δύο τύπων φίλτρων, παρέχοντας στο δίκτυο μια ισχυρή βάση καλά καθορισμένων χαρακτηριστικών, ενώ του επιτρέπει να προσαρμόζει και να βελτιώνει αυτά τα χαρακτηριστικά μέσω της μάθησης. Αντιμετωπίζει μια κοινή πρόκληση στη βαθιά μάθηση, η οποία είναι η ανάγκη για μεγάλες ποσότητες δεδομένων με ετικέτες. Με την ενσωμάτωση χειροποίητων φίλτρων, το KeyNet μειώνει την εξάρτηση από τεράστια σύνολα δεδομένων για την εκπαίδευση και εξασφαλίζει τη σταθερότητα και τη σύγκλιση της διαδικασίας εκπαίδευσης.

Συνοπτικά, η ενσωμάτωση τόσο των χειροποίητων όσο και των μαθημένων φίλτρων στο KeyNet είναι μια στρατηγική απόφαση που αξιοποιεί τα καλύτερα και από τους δύο κόσμους στην όραση υπολογιστών. Τα χειροποίητα φίλτρα παρέχουν μια σταθερή βάση, ενώ τα φίλτρα που έχουν μάθει προσαρμόζονται στις αποχρώσεις του συνόλου δεδομένων. Αυτή η υβριδική προσέγγιση συμβάλλει στην ευρωστία και την αποδοτικότητα του KeyNet, καθιστώντας το μια συναρπαστική επιλογή για εργασίες ανίχνευσης σημείων κλειδίων.

Η πυραμίδα πολλαπλής κλίμακας. Αυτό το συστατικό έχει σχεδιαστεί για να ενισχύει την ανθεκτικότητα του αλγορίθμου στις μεταβολές της κλίμακας χωρίς την ανάγκη για πολλαπλά περάσματα προς τα εμπρός. Στην όραση υπολογιστών, ο χειρισμός αντικειμένων ή σκηνών σε διαφορετικές κλίμακες αποτελεί θεμελιώδη πρόκληση και το KeyNet αντιμετωπίζει αποτελεσματικά αυτή την πρόκληση.

Η πυραμίδα πολλαπλών κλιμάκων στο KeyNet υλοποιείται συμπεριλαμβάνοντας τρία επίπεδα κλίμακας της εικόνας εισόδου, καθένα από τα οποία είναι ελαφρώς θολωμένο και με υποδειγματοληγία. Αυτή η αναπαράσταση πολλαπλών κλιμάκων επιτρέπει στο KeyNet να ανιχνεύει σημεία κλειδιά σε διάφορες κλίμακες μέσα σε ένα μόνο πέρασμα προς τα εμπρός. Η ιδέα πίσω από αυτό είναι να διασφαλιστεί ότι το KeyNet μπορεί να εντοπίζει σημεία κλειδιά ανεξάρτητα από αλλαγές μικρής κλίμακας, ένα κρίσιμο χαρακτηριστικό για πολλές εφαρμογές υπολογιστικής όρασης.

Στην πράξη, όλοι οι χάρτες χαρακτηριστικών που παράγονται από τα χειροποίητα φίλτρα συνενώνονται για να δημιουργηθούν στοιβες εισόδου για τα φίλτρα που μαθαίνονται σε κάθε επίπεδο κλίμακας. Είναι σημαντικό ότι

και οι τρεις ροές σε διαφορετικές κλίμακες μοιράζονται το ίδιο σύνολο βαρών. Αυτό σημαίνει ότι το KeyNet μπορεί να παράγει παρόμοια σημεία κλειδιών σε διαφορετικές κλίμακες, σχηματίζοντας ένα σύνολο υποψήφιων σημείων κλειδιών για την τελική ανίχνευση σημείων κλειδιών.

Αφού δημιουργηθούν χάρτες χαρακτηριστικών από όλα τα επίπεδα κλίμακας, αναβαθμίζονται, συνενώνονται και τροφοδοτούνται στο τελευταίο φίλτρο συμβολής για να παραχθεί ο τελικός χάρτης απόκρισης. Αυτός ο χάρτης απόκρισης περιέχει πληροφορίες σχετικά με τα σημεία κλειδιά που ανιχνεύθηκαν και λαμβάνει υπόψη την πολυκλιμακωτή φύση της εισόδου, επιτρέποντας στο KeyNet να ανιχνεύει σημεία κλειδιά που μπορεί να διαφέρουν σε μέγεθος λόγω αλλαγών προοπτικής ή άλλων παραγόντων.

Η πυραμίδα πολλαπλών κλιμάκων στο KeyNet είναι μια καινοτόμος προσέγγιση για την ανίχνευση σημείων κλειδιών με αμετάβλητη κλίμακα. Ελαχιστοποιεί την ανάγκη για πολλαπλά περάσματα προς τα εμπρός, τα οποία μπορεί να είναι υπολογιστικά δαπανηρά, ενώ εξακολουθεί να καταγράφει σημεία κλειδιά σε ένα εύρος κλιμάκων. Αυτό καθιστά το KeyNet κατάλληλο για εργασίες όπου οι μεταβολές κλίμακας είναι συχνές, όπως η αναγνώριση αντικειμένων ή η παρακολούθηση στην όραση υπολογιστών. Επιπλέον, ο διαμοιρασμός μιας προσεκτικής προσέγγισης για να διασφαλίσει ότι το μοντέλο μαθαίνει να ανιχνεύει αποτελεσματικά τα σημεία κλειδιά.

Ενότητα Loss Functions του αλγορίθμου KeyNet. Η επιλογή των συναρτήσεων απωλειών είναι μια κρίσιμη πτυχή της εκπαίδευσης οποιουδήποτε νευρωνικού δικτύου και το KeyNet χρησιμοποιεί μια προσεκτική προσέγγιση για να διασφαλίσει ότι το μοντέλο μαθαίνει να ανιχνεύει αποτελεσματικά τα σημεία κλειδιά.

Στην εκπαίδευση με επίβλεψη, η συνάρτηση απωλειών βασίζεται σε μεγάλο βαθμό στις ετικέτες βασικής αλήθειας. Ωστόσο, όταν έχουμε να κάνουμε με σημεία κλειδιά, ο ορισμός της βασικής αλήθειας μπορεί να αποτελέσει πρόκληση, καθώς τα σημεία κλειδιά δεν χρειάζεται να έχουν σταθερές, ακριβείς συντεταγμένες. Αντίθετα, θα πρέπει να ανιχνεύονται με ακρίβεια, ανεξάρτητα από γεωμετρικούς ή φωτομετρικούς μετασχηματισμούς στην εικόνα. Το KeyNet αντιμετωπίζει αυτή την πρόκληση εστιάζοντας στην επαναληψιμότητα των εντοπισμένων σημείων κλειδιών, διασφαλίζοντας ότι μπορούν να εντοπιστούν με συνέπεια σε διαφορετικές εικόνες.

Ορισμένοι ανιχνευτές βασικών σημείων με μάθηση εκπαιδεύουν το δίκτυο να εντοπίζει τα βασικά σημεία χωρίς να περιορίζει τις ακριβείς θέσεις τους. Αντ' αυτού, χρησιμοποιούν το μετασχηματισμό ομοιογραφίας μεταξύ εικόνων ως μέτρο επαναληψιμότητας των σημείων κλειδιών. Άλλοι υιοθετούν μια προσέγγιση βασισμένη σε άγκυρες για την καθοδήγηση της εκπαίδευσής τους, η οποία ενισχύει τη σταθερότητα και οδηγεί σε βελτιωμένη απόδοση. Ωστόσο, οι άγκυρες μπορούν να περιορίσουν την ικανότητα του δικτύου να προτείνει εντελώς νέα σημεία κλειδιά όταν δεν υπάρχει άγκυρα κοντά.

Το KeyNet ακολουθεί μια μοναδική προσέγγιση ενσωματώνοντας χειροποίητα φίλτρα, τα οποία παρέχουν έναν ασθενή περιορισμό που μοιάζει με τις μεθόδους που βασίζονται σε άγκυρες, ενώ επιτρέπουν στον ανιχνευτή να προτείνει νέα, σταθερά σημεία κλειδιά. Αυτή η προσέγγιση

απαιτεί μόνο τον γεωμετρικό μετασχηματισμό μεταξύ των εικόνων για να καθοδηγήσει τον υπολογισμό των απωλειών, καθιστώντας την τόσο σταθερή όσο και ικανή να ανιχνεύει νέα σημεία κλειδιά.

Ένα κρίσιμο στοιχείο στη συνάρτηση απωλειών του KeyNet είναι το στρώμα Index Proposal (IP). Αυτό το στρώμα εξάγει συντεταγμένες για την εκπαίδευση των ανιχνευτών σημείων κλειδιών. Χρησιμοποιεί έναν χωρικό τελεστή softmax για την απόκτηση των συντεταγμένων των σημείων κλειδιών, επιτρέποντας τον προσδιορισμό πολλαπλών παγκόσμιων συντεταγμένων σημείων κλειδιών από μία μόνο εικόνα χωρίς να περιορίζεται ο αριθμός των σημείων κλειδιών στο βάθος του χάρτη χαρακτηριστικών ή στο μέγεθος ενός πλέγματος. Ο τελεστής spatial softmax καθιστά το στρώμα IP πλήρως διαφοροποιήσιμο, εξασφαλίζοντας ότι οι κλίσεις μπορούν να διαδοθούν προς τα πίσω όπου εφαρμόζεται το στρώμα IP.

Το KeyNet επεκτείνει περαιτέρω το στρώμα Index Proposal στην έκδοση πολλαπλών κλιμάκων (M-SIP), η οποία χωρίζει τον χάρτη απόκρισης σε πλέγματα και υπολογίζει υποψήφιες θέσεις σημείων κλειδιών για κάθε παράθυρο πλέγματος. Το M-SIP επιτρέπει στο KeyNet να καταγράφει σημεία κλειδιά σε ένα εύρος κλιμάκων, βελτιώνοντας την ευρωστία του.

Συνοπτικά, η ενότητα Loss Functions του KeyNet αντικατοπτρίζει μια μελετημένη στρατηγική για την καθοδήγηση της εκπαίδευσης του μοντέλου στην ανίχνευση σημείων κλειδιών. Εστιάζοντας στην επαναληψιμότητα, αξιοποιώντας τα χειροποίητα φίλτρα και ενσωματώνοντας το στρώμα Index Proposal, το KeyNet επιτυγχάνει ισορροπία μεταξύ της σταθερότητας και της ικανότητας ανίχνευσης νέων, σχετικών σημείων κλειδιών. Αυτή η προσέγγιση συμβάλλει στην αποτελεσματικότητα του αλγορίθμου σε διάφορες εργασίες υπολογιστικής όρασης.

Πλεονεκτήματα

Ανθεκτικότητα: Το KeyNet συνδυάζει χειροποίητα και μαθημένα χαρακτηριστικά, καθιστώντας το ανθεκτικό σε διάφορους μετασχηματισμούς εικόνας και συνθήκες φωτισμού.

Αποδοτικότητα: Η αρχιτεκτονική μειώνει τον αριθμό των μαθησιακών παραμέτρων, βελτιώνοντας τη σταθερότητα κατά την εκπαίδευση και μειώνοντας τον κίνδυνο υπερπροσαρμογής. Προσφέρει επίσης αποδοτικούς χρόνους εξαγωγής συμπερασμάτων, κατάλληλους για εφαρμογές πραγματικού χρόνου.

Αναλλοίωτο κλίμακας: Η ανάλυση πολλαπλών κλιμάκων του KeyNet του επιτρέπει να ανιχνεύει σημεία κλειδιά σε διαφορετικές κλίμακες, καθιστώντας το ευέλικτο για ένα ευρύ φάσμα σεναρίων.

Απόδοση αντιστοίχισης: Όταν χρησιμοποιείται με έναν περιγραφέα όπως το HardNet, το KeyNet επιτυγχάνει κορυφαίες βαθμολογίες αντιστοίχισης, αποδεικνύοντας την αποτελεσματικότητά του σε εργασίες αντιστοίχισης χαρακτηριστικών.

Μειονεκτήματα

Ευαισθησία στο θόρυβο: Όπως πολλοί αλγόριθμοι υπολογιστικής όρασης, το KeyNet μπορεί να είναι ευαίσθητο στο θόρυβο της εικόνας, γεγονός που μπορεί να οδηγήσει σε ψευδώς θετικά αποτελέσματα στην ανίχνευση σημείων κλειδιών.

Συμβιβασμός πολυπλοκότητας: Ενώ το KeyNet μειώνει την πολυπλοκότητα σε σύγκριση με ορισμένους μαθημένους ανιχνευτές, ενδέχεται να μην έχει τόσο καλές επιδόσεις όσο τα πιο σύνθετα μοντέλα σε ορισμένα σενάρια. Θα πρέπει να λαμβάνεται υπόψη η αντιστάθμιση μεταξύ πολυπλοκότητας και απόδοσης.

Περιορίζεται σε σημεία κλειδιά: Το KeyNet έχει σχεδιαστεί κυρίως για την ανίχνευση σημείων κλειδιών και η απόδοσή του ενδέχεται να μην είναι τόσο ισχυρή στην ανίχνευση άλλων τύπων χαρακτηριστικών.

B. Descriptors

Scale-Invariant Feature Transform (SIFT)

Εισήχθη από τον David G. Lowe το 2004, έχει γίνει ακρογωνιαίος λίθος στον τομέα της όρασης υπολογιστών λόγω της ανθεκτικότητάς του στην ανίχνευση και περιγραφή διακριτών τοπικών χαρακτηριστικών σε εικόνες.

Ο αλγόριθμος SIFT αναπτύχθηκε για να αντιμετωπίσει τις προκλήσεις της ανίχνευσης και της αντιστοίχισης χαρακτηριστικών σε διαφορετικές κλίμακες και προσανατολισμούς σε εικόνες. Υπερέχει σε σενάρια που περιλαμβάνουν σημαντικές μεταβολές στην οπτική γωνία, τον φωτισμό και την κλίμακα, καθιστώντας τον ανεκτίμητο για εργασίες όπως η αναγνώριση αντικειμένων, η συρραφή εικόνων και η τρισδιάστατη ανακατασκευή.

Βασικές αρχές και ροή εργασιών: Ο αλγόριθμος SIFT λειτουργεί μέσω μιας σειράς βασικών βημάτων:

1) Το αρχικό βήμα του SIFT περιλαμβάνει την κατασκευή μιας πυραμίδας κλίμακας-χώρου. Αυτή η πυραμίδα αποτελείται από πολλαπλά επίπεδα, καθένα από τα οποία περιέχει μια έκδοση της εικόνας εισόδου που έχει εξομαλυνθεί με φίλτρα Gauss διαφορετικών κλιμάκων. Ο σκοπός αυτής της πυραμίδας είναι να συλλάβει χαρακτηριστικά σε διαφορετικές κλίμακες, επιτρέποντας στο SIFT να είναι αναλλοίωτο σε αλλαγές στο μέγεθος του αντικειμένου. Αυτό είναι σημαντικό επειδή η ανίχνευση χαρακτηριστικών σε διάφορες κλίμακες εξασφαλίζει ότι το SIFT μπορεί να βρει σημεία-κλειδιά ανεξάρτητα από το αν είναι μικρά ή μεγάλα στην εικόνα. Αυτή η αναλλοίωτη κλίμακα είναι κρίσιμη επειδή τα αντικείμενα μπορεί να εμφανίζονται σε διαφορετικές κλίμακες στις εικόνες λόγω της απόστασής τους από την κάμερα.

2) Εντοπισμός σημείων κλειδιών: Σε κάθε επίπεδο της πυραμίδας του χώρου κλίμακας, το SIFT εντοπίζει πιθανά σημεία κλειδιά αναζητώντας τοπικά ακρότατα στην κανονικοποιημένη κατά κλίμακα Λαπλασιανή της εικόνας. Αυτά τα ακρότατα αντιπροσωπεύουν περιοχές της εικόνας όπου το μέγεθος της κλίσης μεγιστοποιείται τόσο στην κλίμακα όσο και στη χωρική διάσταση. Ωστόσο, αυτά τα πιθανά σημεία κλειδιά πρέπει να εντοπιστούν με ακρίβεια.

Ο ακριβής εντοπισμός εξασφαλίζει ότι τα σημεία κλειδιά τοποθετούνται με ακρίβεια εντός της εικόνας. Αυτός ο εντοπισμός είναι απαραίτητος για την εύρωστη αντιστοίχιση σημείων κλειδιών σε διαφορετικές εικόνες, ακόμη και όταν αυτές υφίστανται μετασχηματισμούς.

3) Ανάθεση προσανατολισμού: Για να καταστούν οι περιγραφείς SIFT αμετάβλητοι στην περιστροφή, σε κάθε σημείο κλειδί ανατίθεται ένας κυρίαρχος προσανατολισμός. Η ανάθεση αυτή βασίζεται στις πληροφορίες τοπικής κλίσης γύρω από το σημείο κλειδί. Το SIFT κατασκευάζει ιστογράμματα των προσανατολισμών κλίσης σε μια κυκλική περιοχή γύρω από κάθε σημείο κλειδί και επιλέγει την κορυφή ως κυρίαρχο προσανατολισμό.

Η ανάθεση προσανατολισμών επιτρέπει στο SIFT να επιτύχει αναλλοίωτο ως προς την περιστροφή. Αυτό σημαίνει ότι ακόμη και αν ένα αντικείμενο εξεταστεί από διαφορετική γωνία, τα σημεία κλειδιά θα εξακολουθούν να αντιστοιχίζονται σωστά.

4) Υπολογισμός περιγραφέα: Το SIFT υπολογίζει περιγραφές χαρακτηριστικών για κάθε σημείο κλειδί λαμβάνοντας υπόψη τις πληροφορίες κλίσης σε μια τοπική γειτονιά γύρω από το σημείο κλειδί. Αυτοί οι περιγραφείς είναι συνήθως ιστογράμματα των προσανατολισμών της κλίσης, δημιουργώντας μια διανυσματική αναπαράσταση υψηλής διάστασης.

Οι περιγραφείς κωδικοποιούν τις τοπικές πληροφορίες υψής και σχήματος γύρω από τα σημεία κλειδιά, καθιστώντας τα διακριτά και κατάλληλα για αντιστοίχιση. Η υψηλή διαστατικότητα αυτών των περιγραφέων επιτρέπει την ισχυρή αντιστοίχιση ακόμη και παρουσία θορύβου και απόκρυψης.

5) Αντιστοίχιση βασικών σημείων: Στο τελικό βήμα, το SIFT αντιστοιχεί σημεία κλειδιά μεταξύ διαφορετικών εικόνων ή καρέ. Αυτό επιτυγχάνεται με τη σύγκριση των περιγραφών των βασικών σημείων και τον εντοπισμό των καλύτερων αντιστοιχιών με βάση μια μετρική απόστασης (π.χ. ευκλείδεια απόσταση).

Γιατί είναι αυτό σημαντικό; Η αντιστοίχιση σημείων κλειδιών επιτρέπει εφαρμογές όπως η ευθυγράμμιση εικόνων, η αναγνώριση αντικειμένων και η τρισδιάστατη ανακατασκευή. Είναι η γέφυρα που επιτρέπει στο SIFT να συσχετίζει αντίστοιχα σημεία σε διαφορετικές προβολές ή καρέ.

Εφαρμογές: Ο αλγόριθμος SIFT βρίσκει εφαρμογές σε διάφορους τομείς:

Αναγνώριση αντικειμένων: Τα χαρακτηριστικά SIFT επιτρέπουν την ισχυρή αναγνώριση αντικειμένων, ακόμη και κάτω από αλλαγές στην κλίμακα και την οπτική γωνία. Συρραφή εικόνων: Ο SIFT είναι θεμελιώδης για τη δημιουργία απρόσκοπτων πανοραμικών εικόνων με την αντιστοίχιση και ευθυγράμμιση επικαλυπτόμενων περιοχών. Τρισδιάστατη ανακατασκευή: Τα σημεία κλειδιά SIFT βοηθούν στην ανακατασκευή τρισδιάστατης σκηνής από πολλαπλές εικόνες.

Πλεονεκτήματα

Αναλλοίωτο κλίμακας και περιστροφή: Τα χαρακτηριστικά SIFT είναι αναλλοίωτα σε αλλαγές της κλίμακας και της περιστροφής, πράγμα κρίσιμο για τον χειρισμό των παραλλαγών του πραγματικού κόσμου. Διακριτικότητα: Οι περιγραφείς SIFT είναι ιδιαίτερα διακριτικοί, ελαχιστοποιώντας τις ψευδείς αντιστοιχίες. Ανθεκτικότητα: Το SIFT αποδίδει καλά σε μεταβολές του φωτισμού και της μερικής απόκρυψης.

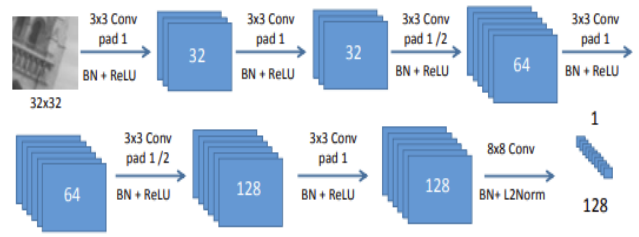
Μειονεκτήματα

Υπολογιστική ένταση: Το SIFT μπορεί να είναι υπολογιστικά απαιτητικό, ιδίως όταν επεξεργάζεται μεγάλες εικόνες ή ροές βίντεο. Απαιτήσεις αποθήκευσης: Ο αλγόριθμος παράγει περιγραφείς υψηλών διαστάσεων, απαιτώντας σημαντική αποθήκευση για μεγάλα σύνολα δεδομένων. Περιορισμοί ευρεσιτεχνίας: Ιστορικά, ο SIFT ήταν κατοχυρωμένος με δίπλωμα ευρεσιτεχνίας, περιορίζοντας τη χρήση του σε ορισμένες εφαρμογές. Ωστόσο, έκτοτε έχει γίνει ευρέως διαθέσιμος.

Ο αλγόριθμος SIFT παραμένει ένα θεμελιώδες εργαλείο στην όραση υπολογιστών, προσφέροντας ισχυρές δυνατότητες ανίχνευσης και περιγραφής χαρακτηριστικών. Η ευελιξία και η αξιοπιστία του τον καθιστούν απαραίτητο για ένα ευρύ φάσμα εφαρμογών. Ενώ οι πρόσφατες εξελίξεις στη βαθιά μάθηση έχουν εισαγάγει εναλλακτικές λύσεις, η διαρκής σημασία του SIFT στην ανάλυση εικόνων βάσει χαρακτηριστικών υπογραμμίζει τη διαρκή αξία του στον τομέα της όρασης υπολογιστών.

HardNet algorithm

Ο αλγόριθμος HardNet είναι μια προσέγγιση που βασίζεται στη βαθιά μάθηση και έχει σχεδιαστεί για τη δημιουργία περιγραφών χαρακτηριστικών υψηλής διακριτικής ικανότητας από τμήματα εικόνας. Χρησιμοποιεί μια αρχιτεκτονική νευρωνικού δικτύου με συνελκτική σύνδεση (CNN) εμπνευσμένη από το VGG, η οποία μαθαίνει αυτόματα περίπλοκα χαρακτηριστικά από τα μπαλώματα εισόδου. Ο πρωταρχικός στόχος του HardNet είναι η μετρική μάθηση, όπου στοχεύει στην αντιστοίχιση παρόμοιων μπαλωμάτων σε κοντινά σημεία ενός χώρου χαρακτηριστικών και στην απομάκρυνση ανόμοιων μπαλωμάτων. Αυτό επιτυγχάνεται μέσω μιας τριπλής συνάρτησης απώλειας περιθωρίου κατά την εκπαίδευση. Ο περιγραφέας που προκύπτει είναι ιδιαίτερα κατάλληλος για εργασίες όπως η στερεοσκοπική όραση, η αντιστοίχιση εικόνων και η τρισδιάστατη ανακατασκευή, όπου ο καθορισμός αντιστοιχιών μεταξύ των χαρακτηριστικών της εικόνας είναι ζωτικής σημασίας. Η αποτελεσματικότητα του HardNet έγκειται στην ικανότητά του να παράγει εύρωστες και κατατοπιστικές αναπαραστάσεις των μπαλωμάτων εικόνας, καθιστώντας το ένα πολύτιμο εργαλείο στον τομέα της όρασης υπολογιστών.



Εικόνα 2: Η αρχιτεκτονική του HardNet

Η ροή του αλγορίθμου αποτελείται από τα παρακάτω στάδια:

Προετοιμασία δεδομένων:

Το πρώτο βήμα στον αλγόριθμο HardNet περιλαμβάνει την προετοιμασία των δεδομένων. Αυτό συνήθως συνίσταται στη συλλογή ενός συνόλου δεδομένων από μπαλώματα εικόνων. Αυτά τα τμήματα είναι συνήθως τετράγωνα και σε κλίμακα του γκρι, τα οποία συχνά εξάγονται από μεγαλύτερες εικόνες. Για παράδειγμα, σε εφαρμογές όπως η στερεοσκοπική όραση ή η αντιστοίχιση εικόνων, αυτά τα patches μπορούν να αντιπροσωπεύουν περιοχές ενδιαφέροντος (ROIs) μέσα στις εικόνες.

Σχεδιασμός αρχιτεκτονικής:

Η βασική αρχιτεκτονική του HardNet βασίζεται σε ένα συνεπτυγμένο νευρωνικό δίκτυο (CNN). Ο σχεδιασμός αυτού του CNN αντλεί εμπνευση από την αρχιτεκτονική VGG, γνωστή για την ικανότητά της να μαθαίνει αυτόματα ιεραρχικά χαρακτηριστικά από δεδομένα. Μέσα στο CNN, υπάρχουν διάφορα επίπεδα, συμπεριλαμβανομένων των στρωμάτων συνελκτικής ανάλυσης, των επιπέδων ομαλοποίησης δέσμης, των συναρτήσεων ενεργοποίησης ReLU (Rectified Linear Unit) και των επιπέδων εγκατάλειψης.

Εξαγωγή χαρακτηριστικών:

Η εξαγωγή χαρακτηριστικών είναι μια κρίσιμη πτυχή του αλγορίθμου HardNet. Τα μπαλώματα εικόνας υποβάλλονται σε επεξεργασία μέσω μιας ακολουθίας στρωμάτων συνελκτικής ανάλυσης. Αυτά τα στρώματα χρησιμοποιούν φίλτρα που έχουν μάθει για να εξάγουν όλο και πιο αφηρημένα και σύνθετα χαρακτηριστικά από τα μπαλώματα εισόδου. Η ομαδοποίηση παρτίδας εφαρμόζεται μετά τα επίπεδα συνελκτικών για την ομαλοποίηση της εξόδου, ενισχύοντας τη σταθερότητα της εκπαίδευσης. Οι συναρτήσεις ενεργοποίησης ReLU εισάγουν μη γραμμικότητα, επιτρέποντας στο δίκτυο να μοντελοποιεί περίπλοκα μοτίβα, και περιλαμβάνονται στρώματα διακοπής για την αποφυγή υπερβολικής προσαρμογής.

Μετρική μάθηση:

Η καρδιά του στόχου του HardNet είναι η μετρική μάθηση. Στοχεύει στη δημιουργία ενός χώρου χαρακτηριστικών όπου τα παρόμοια μπαλώματα βρίσκονται κοντά μεταξύ τους, ενώ τα ανόμοια μπαλώματα τοποθετούνται μακριά μεταξύ τους. Η μετρική μάθηση εκτελείται με τη χρήση της συνάρτησης απώλειας τριπλού περιθωρίου. Αυτό περιλαμβάνει την επιλογή τριπλέτας patches: ένα patch άγκυρας (A), ένα θετικό patch (P) από την ίδια κλάση ή αντίστοιχο patch, και ένα αρνητικό patch (N) από διαφορετική κλάση ή μη αντίστοιχο patch. Η συνάρτηση απωλειών ενθαρρύνει το

δίκτυο να ελαχιστοποιήσει την απόσταση μεταξύ ζευγών άγκυρας-θετικών (d(AP)) και να μεγιστοποιήσει την απόσταση από το πιο σκληρό αρνητικό (d(AN)), εξασφαλίζοντας αποτελεσματική διάκριση μεταξύ παρόμοιων και ανόμοιων patches.

Εκπαίδευση:

Κατά τη διάρκεια της φάσης εκπαίδευσης, γίνεται δειγματοληψία τριπλετών από το σύνολο δεδομένων. Η παραγωγή πληροφοριακών τριπλετών είναι ζωτικής σημασίας για την αποτελεσματικότητα της εκπαίδευσης. Στη συνέχεια χρησιμοποιείται η στοχαστική κάθοδος κλίσης (SGD) ή οι παραλλαγές της για τη βελτιστοποίηση του δικτύου. Ο πρωταρχικός στόχος είναι η ελαχιστοποίηση της απώλειας περιθωρίου τριπλέτας. Οι κλίσεις υπολογίζονται και διαδίδονται μέσω του δικτύου για την επαναληπτική ενημέρωση των βαρών του.

Αξιολόγηση:

Για την αξιολόγηση της απόδοσης του αλγορίθμου HardNet χρησιμοποιούνται διάφορες μετρικές αξιολόγησης. Αυτές περιλαμβάνουν το ποσοστό ψευδώς θετικών αποτελεσμάτων (FPR) σε συγκεκριμένα επίπεδα ανάκλησης, όπως το FPR95, το οποίο μετρά το ποσοστό των ψευδώς θετικών αποτελεσμάτων σε ένα δεδομένο επίπεδο ανάκλησης. Επιπλέον, χρησιμοποιείται η Μέση Μέση Ακρίβεια (mAP), ειδικά σε εργασίες όπως η ανάκτηση εικόνων. Η mAP μετρά την ακρίβεια του περιγραφέα σε πολλαπλά ερωτήματα και υπολογίζεται για διάφορες εργασίες όπως η επαλήθευση, η ανάκτηση και η αντιστοίχιση.

Εφαρμογή:

Αφού εκπαιδευτεί και αξιολογηθεί, ο περιγραφέας HardNet μπορεί να εφαρμοστεί σε ένα ευρύ φάσμα εργασιών υπολογιστικής όρασης. Οι εργασίες αυτές περιλαμβάνουν τη στερεοσκοπική όραση, τη γεωμετρία πολλαπλών όψεων, την αντιστοίχιση εικόνων, την ανάκτηση εικόνων και την τρισδιάστατη ανακατασκευή, καθιστώντας τον ένα ευέλικτο εργαλείο για εργασίες που περιλαμβάνουν την αντιστοίχιση χαρακτηριστικών και την καθιέρωση αντιστοιχίας.

Πλεονεκτήματα

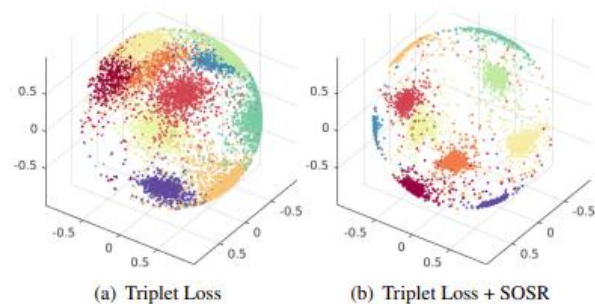
HardNet έχει επιδείξει εντυπωσιακές επιδόσεις σε διάφορα benchmarks και εργασίες όρασης υπολογιστών. Υπερέχει σε εργασίες που απαιτούν αντιστοίχιση χαρακτηριστικών και δημιουργία αντιστοιχίας. Η αρχιτεκτονική του έχει σχεδιαστεί για να μαθαίνει εξαιρετικά διακριτικές ενσωματώσεις, καθιστώντας το ικανό να παράγει περιγραφείς χαρακτηριστικών που διαφοροποιούν αποτελεσματικά μεταξύ αντίστοιχων και μη αντίστοιχων επιφανειών εικόνας. Το HardNet είναι επίσης ανθεκτικό σε διάφορους μετασχηματισμούς, συνθήκες φωτισμού και θόρυβο στις εικόνες, καθιστώντας το κατάλληλο για εφαρμογές του πραγματικού κόσμου όπου τα δεδομένα εισόδου μπορεί να είναι ποικίλα και δύσκολα. Επιπλέον, διατυπώνει την εργασία εκμάθησης περιγραφών ως πρόβλημα μετρικής εκμάθησης, το οποίο είναι κατάλληλο για εργασίες που βασίζονται στην ομοιότητα, όπως η αντιστοίχιση και η ανάκτηση εικόνων.

Μειονεκτήματα

Παρά τα πλεονεκτήματά του, το HardNet έχει ορισμένους περιορισμούς. Τα βαθιά νευρωνικά δίκτυα όπως το HardNet μπορεί να είναι υπολογιστικά δαπανηρά για την εκπαίδευση και την ανάπτυξη, ιδίως σε περιβάλλοντα με περιορισμένους πόρους. Η αρχιτεκτονική του δικτύου μπορεί να έχει μεγάλο αποτύπωμα μνήμης, το οποίο μπορεί να μην είναι κατάλληλο για ανάπτυξη σε συσκευές με περιορισμένη μνήμη. Επιπλέον, η απόδοσή του εξαρτάται σε μεγάλο βαθμό από την ποιότητα και την ποσότητα των δεδομένων εκπαίδευσης και τα ανεπαρκή ή μεροληπτικά δεδομένα εκπαίδευσης μπορεί να οδηγήσουν σε μη βέλτιστα αποτελέσματα. Τέλος, όπως πολλά μοντέλα βαθιάς μάθησης, το HardNet θεωρείται συχνά ένα μοντέλο "μαύρου κουτιού", καθιστώντας δύσκολη την ερμηνεία του τρόπου και του λόγου για τον οποίο κάνει ορισμένες προβλέψεις.

SosNet algorithm

Ο αλγόριθμος SOSNet (Second Order Similarity Network) είναι σχεδιασμένος για την εκμάθηση τοπικών περιγραφών εικόνας. Εισάγει έναν νέο όρο κανονικοποίησης που ονομάζεται Second Order Similarity Regularization (SOSR) για να βελτιώσει την απόδοση αυτών των περιγραφών. Ο SOSNet εκπαιδεύεται σε σύνολα δεδομένων εικόνας μεγάλης κλίμακας και υπερτερεί έναντι των σύγχρονων μεθόδων σε διάφορες εργασίες όρασης υπολογιστών.



Εικόνα 3: Ποιοτικά αποτελέσματα του προτεινόμενου SOSR σε χαρακτηριστικά για τα 10 ψηφία του συνόλου δεδομένων MNIST. Κάθε ψηφίο αναπαριστάται με διαφορετικό χρώμα στη μοναδιαία σφαίρα. Μπορούμε να παρατηρήσουμε ότι με τη χρήση της μεθόδου SOSR που ενθαρρύνει την ομοιότητα δεύτερης τάξης, μαθαίνονται πιο συμπαγείς ατομικές συστάδες σε σύγκριση με την τυπική απώλεια τριπλέτας.

Η ροή του αλγορίθμου περιγράφεται στα παρακάτω βήματα:

Προετοιμασία δεδομένων: Το SOSNet ξεκινά με την προετοιμασία των δεδομένων, η οποία περιλαμβάνει τη δημιουργία ενός ποικίλου συνόλου δεδομένων εκπαίδευσης από μπαλώματα εικόνας. Αυτά τα μπαλώματα εξάγονται από ένα ευρύ φάσμα εικόνων, που καλύπτουν διαφορετικές κατηγορίες ή σκηνές. Αυτή η ποικιλομορφία είναι ζωτικής σημασίας, καθώς διασφαλίζει ότι οι περιγραφείς που μαθαίνονται είναι ανθεκτικοί και μπορούν να γενικευτούν σε διάφορα σενάρια του πραγματικού κόσμου. Η συλλογή ενός εκτεταμένου και αντιπροσωπευτικού συνόλου δεδομένων εκπαίδευσης είναι ζωτικής σημασίας για την απόδοση του αλγορίθμου, καθώς παρέχει στο νευρωνικό δίκτυο τα

απαραίτητα παραδείγματα για την εκμάθηση διακριτικών χαρακτηριστικών.

Δίκτυο περιγραφής: Το SOSNet χρησιμοποιεί μια βαθιά αρχιτεκτονική νευρωνικού δικτύου που έχει σχεδιαστεί για την εκμάθηση περιγραφών εικόνας. Προσαρμόζει μια τροποποιημένη έκδοση της αρχιτεκτονικής L2Net, η οποία είναι γνωστή για την αποτελεσματικότητά της σε αυτό το πλαίσιο. Αυτό το δίκτυο ενεργεί ως το βασικό συστατικό που είναι υπεύθυνο για την εξαγωγή ουσιαστικών αναπαραστάσεων από τα τμήματα εικόνας. Τα βαθιά νευρωνικά δίκτυα έχουν την ικανότητα να συλλαμβάνουν περίπλοκα μοτίβα και χαρακτηριστικά, γεγονός που τα καθιστά κατάλληλα για το έργο της εκμάθησης περιγραφών.

Βελτιστοποίηση απώλειας τριπλέτας: Η καρδιά της διαδικασίας εκπαίδευσης του SOSNet βρίσκεται στη βελτιστοποίηση απώλειας τριπλέτας. Σε κάθε παρτίδα εκπαίδευσης, επιλέγονται τυχαία N ζεύγη επιφανειών. Μέσα σε κάθε ζεύγος, το δίκτυο περιγραφής υπολογίζει τις αποστάσεις L2 μεταξύ των μπαλωμάτων. Στη συνέχεια χρησιμοποιείται η συνάρτηση απώλειας τριπλέτας για την ελαχιστοποίηση της διαφοράς μεταξύ της απόστασης των θετικών ζευγών (patches από την ίδια κλάση) και των αρνητικών ζευγών (patches από διαφορετικές κλάσεις). Αυτή η συνάρτηση απωλειών καθοδηγεί αποτελεσματικά το δίκτυο να μάθει περιγραφείς που διακρίνουν τα patches της ίδιας κλάσης ενώ τα καθιστούν παρόμοια μεταξύ τους, μια θεμελιώδης προϋπόθεση για την επιτυχή αντιστοίχιση εικόνων.

Κανονικοποίηση ομοιότητας δεύτερης τάξης (SOSR): Ο SOSR είναι ένας νέος όρος κανονικοποίησης που εισήγαγε το SOSNet, ο οποίος το διαφοροποιεί από τις παραδοσιακές μεθόδους εκμάθησης περιγραφών. Η SOSR ενθαρρύνει τους περιγραφείς να έχουν παρόμοια στατιστικά στοιχεία δεύτερης τάξης, τα οποία αφορούν τις σχέσεις μεταξύ των patches και των πλησιέστερων γειτόνων τους εντός μιας παρτίδας (K-κοντινότεροι γείτονες). Αυτός ο όρος κανονικοποίησης εισάγει την έννοια της ομοιότητας δεύτερης τάξης, διασφαλίζοντας ότι οι περιγραφείς δεν καταγράφουν μόνο τις διαφορές πρώτης τάξης αλλά και τις σχέσεις ανώτερης τάξης μεταξύ των patches. Αυτό το βήμα είναι καθοριστικό για τη βελτίωση της συνολικής ποιότητας των περιγραφών.

Βελτιστοποίηση της εκπαίδευσης: Αφού οριστούν οι συναρτήσεις απώλειας, συμπεριλαμβανομένης της τριπλής απώλειας και της SOSR, το SOSNet χρησιμοποιεί τεχνικές βελτιστοποίησης όπως η Στοχαστική Κάθοδος Κλίσης (SGD) ή η Adam για την ενημέρωση των παραμέτρων του δικτύου. Είναι ενδιαφέρον ότι το SOSNet παρατηρεί ότι το Adam αποδίδει καλύτερα κατά τη διάρκεια της εκπαίδευσης. Η βελτιστοποίηση είναι η επαναληπτική διαδικασία κατά την οποία το δίκτυο μαθαίνει να ελαχιστοποιεί τις συναρτήσεις απωλειών προσαρμόζοντας τις εσωτερικές του παραμέτρους, βελτιώνοντας σταδιακά τις δυνατότητες εξαγωγής περιγραφών.

Αξιολόγηση επιδόσεων: Μετά την εκπαίδευση, το SOSNet υποβάλλεται σε αυστηρή αξιολόγηση σε μια σειρά εργασιών όρασης υπολογιστών, όπως η αντιστοίχιση μπαλωμάτων, η επαλήθευση, η ανάκτηση και η τρισδιάστατη ανακατασκευή. Αυτή η αξιολόγηση αξιολογεί την πρακτική χρησιμότητα των περιγραφικών που μαθαίνονται σε πραγματικές

εφαρμογές. Η απόδοση του αλγορίθμου μετράται με βάση κριτήρια όπως τα ποσοστά ψευδώς θετικών αποτελεσμάτων, η μέση μέση ακρίβεια και η ποιότητα των τρισδιάστατων ανακατασκευών. Οι αξιολογήσεις αυτές επιβεβαιώνουν την αποτελεσματικότητα του αλγορίθμου και συμβάλλουν στην τεκμηρίωση της υπεροχής του έναντι των υφιστάμενων μεθόδων.

Στην ουσία, η ροή εργασίας SOSNet είναι μια καλά δομημένη διαδικασία που αξιοποιεί την ποικιλομορφία δεδομένων, τα βαθιά νευρωνικά δίκτυα, τις καινοτόμες τεχνικές κανονικοποίησης και την αυστηρή αξιολόγηση για την παραγωγή περιγραφικών εικόνων υψηλής ποιότητας. Κάθε βήμα παίζει καθοριστικό ρόλο στη διαμόρφωση της επιτυχίας του αλγορίθμου, από τη συλλογή δεδομένων έως την εκμάθηση περιγραφών, την κανονικοποίηση και την αξιολόγηση των επιδόσεων. Η μοναδική συμβολή του SOSNet έγκειται στην εισαγωγή του SOSR, το οποίο προάγει τη βαθύτερη κατανόηση της ομοιότητας δεύτερης τάξης, οδηγώντας τελικά σε βελτιωμένες εργασίες αντιστοίχισης εικόνων και όρασης υπολογιστών.

Εφαρμογές

Αντιστοίχιση και ανάκτηση εικόνων: Το SOSNet μπορεί να χρησιμοποιηθεί για την αποτελεσματική αντιστοίχιση και ανάκτηση παρόμοιων εικόνων από μια βάση δεδομένων. Αυτό είναι πολύτιμο σε συστήματα ανάκτησης εικόνων με βάση το περιεχόμενο, όπου οι χρήστες μπορούν να αναζητήσουν εικόνες με βάση την οπτική ομοιότητα και όχι τις ετικέτες κειμένου.

Οπτική SLAM (ταυτόχρονος εντοπισμός και χαρτογράφηση): Στα συστήματα SLAM, το SOSNet βοηθά στην εξαγωγή και την αντιστοίχιση χαρακτηριστικών, η οποία είναι ζωτικής σημασίας για την εκτίμηση της στάσης της κάμερας και τη δημιουργία χαρτών του περιβάλλοντος. Συμβάλλει στην τρισδιάστατη χαρτογράφηση σε πραγματικό χρόνο στη ρομποτική και την επαυξημένη πραγματικότητα.

Ανίχνευση και παρακολούθηση αντικειμένων: Τα χαρακτηριστικά του SOSNet μπορούν να βοηθήσουν στον εντοπισμό και την παρακολούθηση αντικειμένων σε βίντεο και εικόνες. Είναι χρήσιμο σε εφαρμογές όπως η επιτήρηση, τα αυτόνομα οχήματα και η ρομποτική.

Ανακατασκευή 3D: Το SOSNet εφαρμόζεται σε εργασίες τρισδιάστατης ανακατασκευής, όπου βοηθά στην ευθυγράμμιση εικόνων και στη δημιουργία λεπτομερών τρισδιάστατων μοντέλων από εικόνες 2D. Αυτό είναι πολύτιμο σε τομείς όπως η αρχαιολογία, η διατήρηση της πολιτιστικής κληρονομιάς και τα γραφικά υπολογιστών.

Εγγραφή εικόνων: Στην ιατρική απεικόνιση, το SOSNet βοηθά στην καταχώρηση ή την ευθυγράμμιση διαφορετικών ιατρικών εικόνων για τη διάγνωση και το σχεδιασμό της θεραπείας. Ενισχύει την ακρίβεια των διαδικασιών καθοδηγούμενων από εικόνες.

Επαυξημένη πραγματικότητα (AR): Το SOSNet χρησιμοποιείται σε εφαρμογές AR για την παρακολούθηση και αναγνώριση αντικειμένων ή σκηνών στον πραγματικό κόσμο, επικαλύπτοντας απρόσκοπτα ψηφιακές πληροφορίες.

Ανάλυση εγγράφων: Στην OCR (οπτική αναγνώριση χαρακτήρων) και την ανάλυση εγγράφων, το SOSNet συμβάλλει στην ευθυγράμμιση και την αντιστοίχιση

σαρωμένων εγγράφων, βελτιώνοντας την αναγνώριση κειμένου και την κατηγοριοποίηση εγγράφων.

Γεωγραφικά συστήματα πληροφοριών (GIS): Το SOSNet συμβάλλει στην αντιστοίχιση χαρακτηριστικών και την αναγνώριση αντικειμένων σε δορυφορικές και εναέριες εικόνες, βελτιώνοντας εφαρμογές ΓΣΠ όπως η δημιουργία χαρτών και η ταξινόμηση χρήσεων γης.

Ποιοτικός έλεγχος στη μεταποίηση: Το SOSNet χρησιμοποιείται σε βιομηχανικά περιβάλλοντα για την επιθεώρηση και την επαλήθευση της ποιότητας των παραγόμενων προϊόντων με την αντιστοίχιση χαρακτηριστικών σε εικόνες και την ανίχνευση ελαττωμάτων.

Τέχνη και πολιτιστική κληρονομιά: Στην αποκατάσταση έργων τέχνης και στην αρχαιολογία, το SOSNet βοηθά στην ανάλυση και αποκατάσταση έργων τέχνης και αντικειμένων με την ευθυγράμμιση εικόνων και τον εντοπισμό λεπτών λεπτομερειών.

Λιανικό εμπόριο και ηλεκτρονικό εμπόριο: Το SOSNet μπορεί να εφαρμοστεί στην οπτική αναζήτηση και την αναγνώριση προϊόντων για το ηλεκτρονικό εμπόριο, διευκολύνοντας τους χρήστες να βρίσκουν προϊόντα παρόμοια με αυτά που επιθυμούν.

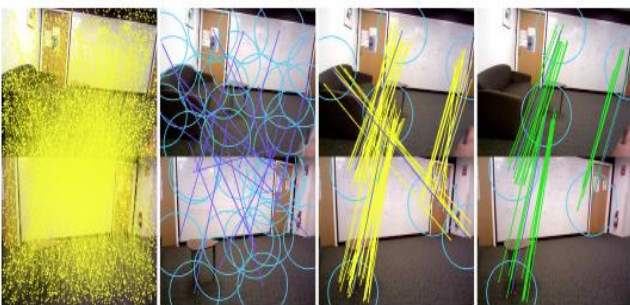
Ψυχαγωγία και παιχνίδια: Στις βιομηχανίες ψυχαγωγίας και παιχνιδιών, το SOSNet χρησιμοποιείται για την εμφύχωση χαρακτήρων, τη σύλληψη κίνησης και τη δημιουργία ρεαλιστικών οπτικών εφέ.

Ρομποτική και αυτοματισμοί: Το SOSNet παίζει ρόλο σε διάφορες εφαρμογές ρομποτικής, συμπεριλαμβανομένης της πλοήγησης, του χειρισμού αντικειμένων και της αλληλεπίδρασης ανθρώπου-ρομπότ.

Ανάλυση ιατρικών εικόνων: Είναι πολύτιμο σε εργασίες ανάλυσης ιατρικών εικόνων, όπως η καταγραφή εικόνων, η ανίχνευση βλαβών και η κατάτμηση οργάνων.

Δορυφορική και τηλεπισκόπηση: Το SOSNet μπορεί να βοηθήσει στην ανάλυση δορυφορικών δεδομένων και δεδομένων τηλεπισκόπησης, συμπεριλαμβανομένης της ταξινόμησης της κάλυψης γης και της ανίχνευσης αλλαγών.

AdaLAM (Adaptive Local Affine Matching)



Εικόνα 4: Κύρια βήματα στο AdaLAM, από αριστερά προς τα δεξιά: 1. λαμβάνουμε ως είσοδο ένα ευρύ σύνολο υποθετικών αντιστοιχιών (με κίτρινο χρώμα), 2. Επιλέγουμε καλά διαδομένες υποθέσεις χονδρικών αντιστοιχιών περιοχών (μπλε κύκλοι), 3. για κάθε περιοχή εξετάζουμε το σύνολο όλων των υποθετικών αντιστοιχιών που είναι συνεπείς με την ίδια υπόθεση της περιοχής, 4. κρατάμε μόνο

τις αντιστοιχίες που είναι συνεπείς με έναν αφινικό μετασχηματισμό με επαρκή υποστήριξη (πράσινο χρώμα).

Το AdaLAM είναι μια πρωτοποριακή εξέλιξη στην όραση υπολογιστών, ειδικά προσαρμοσμένη για ισχυρή και ακριβή αντιστοίχιση χαρακτηριστικών. Στη σφαίρα της όρασης υπολογιστών, η ακριβής αντιστοίχιση οπτικών χαρακτηριστικών σε όλες τις εικόνες είναι θεμελιώδης για εργασίες όπως η τρισδιάστατη ανακατασκευή και ο οπτικός εντοπισμός. Το AdaLAM αντιμετωπίζει την εγγενή πρόκληση της διαφοροποίησης των γνήσιων ταυτίσεων χαρακτηριστικών από τις ψευδείς, ιδίως σε πολύπλοκες σκηνές που χαρακτηρίζονται από διακυμάνσεις στο φωτισμό, την προοπτική ή την ακαταστασία. Στον πυρήνα του, το AdaLAM αξιοποιεί την έννοια της τοπικής συνέπειας, ένα κρίσιμο στοιχείο για το διαχωρισμό των μη-ορθών αντιστοιχιών (σωστών αντιστοιχιών) από τις ακραίες. Το επιτυγχάνει αυτό μέσω μιας νέας προσέγγισης, προσαρμόζοντας δυναμικά το κατώφλι που χρησιμοποιείται για την ταξινόμηση των αντιστοιχίσεων ως ασαφείς ή ασαφείς με βάση τα τοπικά χαρακτηριστικά. Με την εισαγωγή αυτού του προσαρμοστικού μηχανισμού κατωφλίου, το AdaLAM ενισχύει σημαντικά την ευρωστία και την ακρίβεια της αντιστοίχισης χαρακτηριστικών. Η χρήση ενός συγγενικού μοντέλου, πιο ευέλικτου από τις παραδοσιακές ομογραφίες, επεκτείνει περαιτέρω την εφαρμογή του σε διάφορα σενάρια. Οι αξιολογήσεις σε σύνολα δεδομένων εσωτερικών και εξωτερικών χώρων καταδεικνύουν σταθερά την υπεροχή του AdaLAM έναντι των σύγχρονων μεθόδων, τονίζοντας τον καθοριστικό του ρόλο στην ενίσχυση της αξιοπιστίας των εφαρμογών υπολογιστικής όρασης που βασίζονται στην ακριβή αντιστοίχιση χαρακτηριστικών.

Η ροή εργασίας του AdaLAM περιλαμβάνει διάφορα κρίσιμα βήματα, καθένα από τα οποία έχει σχεδιαστεί σχολαστικά για να βελτιώνει την ακρίβεια της αντιστοίχισης χαρακτηριστικών σε εφαρμογές υπολογιστικής όρασης. Κατ' αρχάς, ξεκινά με την εξαγωγή χαρακτηριστικών, όπου εντοπίζονται τα σημεία κλειδιά ή τα διακριτικά σημεία της εικόνας. Αυτό το αρχικό βήμα είναι ζωτικής σημασίας, καθώς η ποιότητα των εξαχθέντων χαρακτηριστικών επηρεάζει άμεσα την επακόλουθη διαδικασία αντιστοίχισης. Το AdaLAM λειτουργεί σε αυτά τα σημεία κλειδιά, οπότε η επιλογή τους είναι κρίσιμη. Μόλις εντοπιστούν τα σημεία κλειδιά, υπολογίζονται οι περιγραφείς τους, οι οποίοι είναι συμπαγείς αριθμητικές αναπαραστάσεις του τοπικού περιβάλλοντος των σημείων κλειδιά. Αυτοί οι περιγραφείς χρησιμεύουν ως βάση για την αντιστοίχιση σημείων κλειδιά σε διαφορετικές εικόνες.

Το δεύτερο βήμα περιλαμβάνει την εύρεση πιθανών αντιστοιχιών μεταξύ σημείων κλειδιά σε διαφορετικές εικόνες. Αυτό επιτυγχάνεται μέσω της αντιστοίχισης του πλησιέστερου γείτονα, όπου ένα σημείο κλειδί σε μια εικόνα αντιστοιχίζεται με το πλησιέστερο αντίστοιχό του σε μια άλλη εικόνα με βάση την ομοιότητα των περιγραφέων τους. Ωστόσο, δεν είναι όλες οι πιθανές αντιστοιχίες γνήσιες-πολλές είναι ακραίες ή αναντίστοιχες που προκαλούνται από αλλαγές στην οπτική γωνία, το φωτισμό ή τις αποκρύψεις. Το τρίτο βήμα του AdaLAM εισάγει μια κρίσιμη καινοτομία: την προσαρμοστική τοπική αφινική μοντελοποίηση. Οι παραδοσιακές μέθοδοι χρησιμοποιούν συνήθως ένα σταθερό κατώφλι για την ταξινόμηση των αντιστοιχιών ως inliers (σωστές αντιστοιχίες) ή outliers

(αναντιστοιχίες). Η AdaLAM, από την άλλη πλευρά, προσαρμόζει δυναμικά αυτό το κατώφλι με βάση τις τοπικές πληροφορίες, εντοπίζοντας αποτελεσματικά αντιστοιχίες που τηρούν τους αναμενόμενους τοπικούς γεωμετρικούς μετασχηματισμούς, ακόμη και παρουσία πολύπλοκων παραλλαγών.

Αυτός ο προσαρμοστικός μηχανισμός κατωφλίου αποτελεί την καρδιά του AdaLAM και είναι απαραίτητος για την ισχυρή αντιστοίχιση. Εξετάζει την τοπική συνέπεια των αντιστοιχίσεων σε μικρές περιοχές γύρω από κάθε ζεύγος σημείων κλειδιών. Αναλύοντας αυτά τα τοπικά μοτίβα, το AdaLAM καθορίζει έξυπνα το κατώφλι για κάθε αντιστοιχία, επιτρέποντάς του έτσι να προσαρμόζεται σε ένα ευρύ φάσμα παραλλαγών στην κλίμακα, τον προσανατολισμό και την προοπτική. Η προσέγγιση αυτή είναι ιδιαίτερα αποτελεσματική σε σενάρια όπου τα παραδοσιακά σταθερά κατώτατα όρια μπορεί να οδηγήσουν είτε σε υπερβολικά αυστηρό φιλτράρισμα, απορρίπτοντας δυνητικά πολύτιμες αντιστοιχίες, είτε σε ανεπαρκές φιλτράρισμα, οδηγώντας σε λανθασμένες αντιστοιχίες.

Το τελικό βήμα της ροής εργασίας του AdaLAM περιλαμβάνει την απόρριψη των ακραίων τιμών με βάση τα προσαρμοστικά καθορισμένα κατώτατα όρια. Οι αντιστοιχίες που δεν πληρούν τα κριτήρια τοπικής συνέπειας κλαδεύονται, αφήνοντας πίσω μόνο τις ισχυρές αντιστοιχίες. Αυτό το βήμα διασφαλίζει ότι μόνο αντιστοιχίες χαρακτηριστικών υψηλής ποιότητας διατηρούνται για επακόλουθες εργασίες υπολογιστικής όρασης, όπως η τρισδιάστατη ανακατασκευή ή ο οπτικός εντοπισμός. Η καινοτόμος ροή εργασίας του AdaLAM, η οποία βασίζεται στην έννοια της προσαρμοστικής κατωφλίωσης και της τοπικής συνέπειας, έχει αποδειχθεί ότι υπερτερεί έναντι των σύγχρονων μεθόδων σε μια σειρά από απαιτητικά σενάρια του πραγματικού κόσμου, γεγονός που την καθιστά σημαντική πρόοδο στον τομέα της όρασης υπολογιστών και της αντιστοίχισης χαρακτηριστικών.

Πλεονεκτήματα

Ανθεκτικότητα: Το προσαρμοστικό μηχανισμό κατωφλίωσης του AdaLAM ενισχύει την ανθεκτικότητα, καθώς προσαρμόζεται σε διάφορους γεωμετρικούς μετασχηματισμούς, καθιστώντας το κατάλληλο για δύσκολα σενάρια του πραγματικού κόσμου.

Υψηλή ακρίβεια αντιστοίχισης: Με τη δυναμική προσαρμογή των κριτηρίων αντιστοίχισης, το AdaLAM υπερέρχει στον ακριβή εντοπισμό έγκυρων αντιστοιχιών χαρακτηριστικών, οδηγώντας σε βελτιωμένη ακρίβεια αντιστοίχισης.

Τοπική συνέπεια: Η εστίαση του AdaLAM στην τοπική συνέπεια του επιτρέπει να χειρίζεται πολύπλοκες παραλλαγές, συμπεριλαμβανομένων των αλλαγών στην κλίμακα, τον προσανατολισμό και την προοπτική, βελτιώνοντας την απόδοσή του σε ακατάστατα και δυναμικά περιβάλλοντα.

Συμβατότητα: Η AdaLAM μπορεί εύκολα να ενσωματωθεί σε υπάρχουσες σωληνώσεις υπολογιστικής όρασης, ενισχύοντας τις δυνατότητες αντιστοίχισης χαρακτηριστικών τους.

Ευελιξία: Η αποτελεσματικότητα του AdaLAM αποδεικνύεται τόσο σε εσωτερικό όσο και σε εξωτερικό

περιβάλλον, αναδεικνύοντας την ευελιξία του σε διάφορες εφαρμογές.

Μειονεκτήματα

Υπολογιστική επιβάρυνση: Η προσαρμοστική κατωφλίωση της AdaLAM απαιτεί πρόσθετους υπολογιστικούς πόρους σε σύγκριση με τις μεθόδους σταθερού κατωφλίου, καθιστώντας την σχετικά πιο αργή σε ορισμένες περιπτώσεις.

Ευαισθησία στις παραμέτρους: Ενώ η τοπική προσαρμογή της AdaLAM αποτελεί πλεονέκτημα, μπορεί επίσης να αποτελέσει μειονέκτημα, καθώς η απόδοση της μεθόδου εξαρτάται από τη σωστή ρύθμιση των παραμέτρων της για συγκεκριμένες εφαρμογές.

Επεκτασιμότητα: Σε σύνολα δεδομένων εξαιρετικά μεγάλης κλίμακας, η AdaLAM μπορεί να αντιμετωπίσει προβλήματα κλιμάκωσης λόγω της προσαρμοστικής της φύσης, η οποία περιλαμβάνει την ανάλυση τοπικών περιοχών γύρω από τα σημεία κλειδιά.

Εξάρτηση από την εξαγωγή χαρακτηριστικών: Η επιτυχία του AdaLAM εξαρτάται σε μεγάλο βαθμό από την ποιότητα της εξαγωγής χαρακτηριστικών, και εάν τα σημεία κλειδιά δεν εξάγονται επαρκώς, αυτό μπορεί να επηρεάσει την απόδοση αντιστοίχισης.

Περιορισμένες παγκόσμιες πληροφορίες: Το AdaLAM επικεντρώνεται κυρίως σε τοπικά χαρακτηριστικά και μπορεί να μην υπερέρχει σε περιπτώσεις όπου το παγκόσμιο πλαίσιο είναι απαραίτητο για την αντιστοίχιση.

Συνοπτικά, η προσαρμοστική προσέγγιση του AdaLAM βελτιώνει σημαντικά την ακρίβεια και την ευρωστία της αντιστοίχισης χαρακτηριστικών, αλλά μπορεί να συνοδεύεται από κάποιο υπολογιστικό κόστος και ευαισθησία στις παραμέτρους, καθιστώντας το ένα ισχυρό εργαλείο για πολλές εφαρμογές υπολογιστικής όρασης, ενώ απαιτεί προσεκτική εξέταση της εφαρμογής και της χρήσης του σε συγκεκριμένα πλαίσια.

Εφαρμογές

Το AdaLAM, με τον προηγμένο μηχανισμό προσαρμοστικής κατωφλίωσης και τη μοντελοποίηση της τοπικής συνέπειας, βρίσκει εφαρμογές σε διάφορους τομείς της όρασης υπολογιστών και όχι μόνο:

SLAM (ταυτόχρονος εντοπισμός και χαρτογράφηση): Οι ισχυρές δυνατότητες αντιστοίχισης χαρακτηριστικών του AdaLAM το καθιστούν πολύτιμο σε συστήματα SLAM για την ακριβή εκτίμηση της θέσης της κάμερας και τη δημιουργία τρισδιάστατων χαρτών του περιβάλλοντος. Εξασφαλίζει αξιόπιστες αντιστοιχίες χαρακτηριστικών, ακόμη και σε δύσκολα σενάρια.

Αναγνώριση και παρακολούθηση αντικειμένων: Το AdaLAM μπορεί να βοηθήσει σε εργασίες αναγνώρισης και εντοπισμού αντικειμένων παρέχοντας ακριβή αντιστοίχιση χαρακτηριστικών. Είναι χρήσιμο σε εφαρμογές ρομποτικής, επιτήρησης και επαυξημένης πραγματικότητας.

Συρραφή εικόνων και δημιουργία πανοράματος: Η ικανότητα του AdaLAM να χειρίζεται διάφορους μετασχηματισμούς, συμπεριλαμβανομένων των περιστροφών και της κλιμάκωσης, το καθιστά ιδανικό για τη

δημιουργία απρόσκοπτων πανοραμάτων και συρραμμένων εικόνων από πολλαπλά καρέ.

Δομή από κίνηση (SfM): Η AdaLAM συμβάλλει στην ανακατασκευή τρισδιάστατων σκηνών από εικόνες 2D με την ακριβή αντιστοίχιση χαρακτηριστικών μεταξύ εικόνων, ένα κρίσιμο βήμα στις σωληνώσεις SfM.

Οπτική Οδομετρία: Χρησιμοποιείται στην εκτίμηση της κίνησης της κάμερας με την παρακολούθηση σημείων κλειδιών σε διαδοχικά καρέ. Αυτό είναι ζωτικής σημασίας για την αυτόνομη πλοήγηση, τα μη επανδρωμένα αεροσκάφη και τη ρομποτική.

Εντοπισμός αντικειμένων σε εναέριες εικόνες: Το AdaLAM μπορεί να βοηθήσει στον εντοπισμό αντικειμένων ή ορόσημων σε εναέριες και δορυφορικές εικόνες, συμβάλλοντας σε εφαρμογές όπως η δημιουργία χαρτών και η παρακολούθηση καταστροφών.

Τρισδιάστατη ανακατασκευή: Στα γραφικά υπολογιστών και την τρισδιάστατη μοντελοποίηση, η AdaLAM βοηθά στη δημιουργία λεπτομερών τρισδιάστατων μοντέλων από εικόνες 2D, βρίσκοντας χρήση στην εικονική πραγματικότητα, την ανάπτυξη παιχνιδιών και τη διατήρηση της πολιτιστικής κληρονομιάς.

Ιατρική απεικόνιση: Το AdaLAM μπορεί να χρησιμοποιηθεί για την ευθυγράμμιση και την αντιστοίχιση ιατρικών εικόνων για εφαρμογές όπως η καταγραφή εικόνων και ο εντοπισμός όγκων στην ακτινολογία και την ακτινοθεραπεία.

Ανάλυση εικόνων εγγράφων: Στο πλαίσιο του OCR (Optical Character Recognition) και της ανάλυσης εγγράφων, η AdaLAM μπορεί να ενισχύσει την ευθυγράμμιση σαρωμένων ή φωτογραφημένων εγγράφων, βελτιώνοντας την αναγνώριση κειμένου.

Γεωγραφικά συστήματα πληροφοριών (GIS): Μπορεί να βοηθήσει στην αντιστοίχιση χαρακτηριστικών για δορυφορικές και εναέριες εικόνες, βοηθώντας στη δημιουργία χαρτών, στην ταξινόμηση χρήσεων γης και στην περιβαλλοντική παρακολούθηση.

Ποιοτικός έλεγχος στη μεταποίηση: Το AdaLAM μπορεί να χρησιμοποιηθεί σε διαδικασίες ποιοτικού ελέγχου για την ανίχνευση ελαττωμάτων και τη διασφάλιση της σωστής συναρμολόγησης προϊόντων με την αντιστοίχιση χαρακτηριστικών σε εικόνες.

Αποκατάσταση έργων τέχνης και αρχαιολογία: Κατά την αποκατάσταση έργων τέχνης ή την ανάλυση αρχαιολογικών αντικειμένων, το AdaLAM βοηθά στην ευθυγράμμιση εικόνων και στη μελέτη λεπτών λεπτομερειών.

Ψυχαγωγία και κινούμενα σχέδια: Μπορεί να εφαρμοστεί στις βιομηχανίες ταινιών και παιχνιδιών για τη δημιουργία ρεαλιστικών οπτικών εφέ, τη σύλληψη κίνησης και την εμπύχωση χαρακτήρων.

Λιανικό εμπόριο και ηλεκτρονικό εμπόριο: Το AdaLAM μπορεί να βοηθήσει στην οπτική αναζήτηση και την αναγνώριση προϊόντων σε εφαρμογές ηλεκτρονικού εμπορίου, βελτιώνοντας την εμπειρία του χρήστη.

Παρακολούθηση της κυκλοφορίας: Η AdaLAM μπορεί να βοηθήσει στην παρακολούθηση οχημάτων και στην παρακολούθηση της ροής της κυκλοφορίας μέσω ανάλυσης βίντεο.

Αναγνώριση χειρονομιών: Η AdaLAM μπορεί να αποτελέσει μέρος συστημάτων που αναγνωρίζουν χειρονομίες του χεριού ή του σώματος σε εφαρμογές όπως η μετάφραση νοηματικής γλώσσας ή τα παιχνίδια.

Αυτές οι εφαρμογές καταδεικνύουν την ευελιξία της AdaLAM για την ενίσχυση της αντιστοίχισης χαρακτηριστικών και τη δημιουργία πιο ακριβών και αξιόπιστων λύσεων υπολογιστικής όρασης σε διάφορους τομείς.

DISK (DIScrete Keypoints) algorithm

Ο αλγόριθμος DISK (DIScrete Keypoints) αντιπροσωπεύει μια πρωτοποριακή προσέγγιση για την εκμάθηση τοπικών χαρακτηριστικών της εικόνας με έναν αξιόπιστο και ολοκληρωμένο τρόπο. Ο DISK αντιμετωπίζει τις εγγενείς δυσκολίες στην εκμάθηση τέτοιων χαρακτηριστικών, που προκύπτουν κυρίως από τη διακριτή φύση της επιλογής και της αντιστοίχισης αραιών σημείων κλειδιών. Αυτή η νέα μέθοδος αξιοποιεί αρχές από την Ενισχυτική Μάθηση (RL) για τη βελτιστοποίηση της δημιουργίας σημείων κλειδιών και περιγραφών, επιδιώκοντας τη μεγιστοποίηση του αριθμού των σωστών αντιστοιχίσεων χαρακτηριστικών. Η δύναμη του DISK έγκειται στο κομψό πιθανοτικό μοντέλο του, το οποίο επιτρέπει τη στενή ευθυγράμμιση μεταξύ των φάσεων εκπαίδευσης και εξαγωγής συμπερασμάτων, μια κρίσιμη πτυχή για την αξιοσημείωτη αποτελεσματικότητά του. Επιπλέον, παρουσιάζει εντυπωσιακές ιδιότητες σύγκλισης, καθιστώντας το ιδιαίτερα κατάλληλο για εκπαίδευση από το μηδέν. Ένα εντυπωσιακό χαρακτηριστικό του DISK είναι η ικανότητά του να παράγει πυκνά καταναμημένα χαρακτηριστικά που παραμένουν διακριτικά. Αυτό αμφισβητεί τις συμβατικές υποθέσεις σχετικά με τα χαρακτηριστικά των καλών σημείων κλειδιών, όπως αποδεικνύεται σε διάφορα σημεία αναφοράς, όπου το DISK επιτυγχάνει σταθερά κορυφαία αποτελέσματα. Ωστόσο, είναι σημαντικό να σημειωθεί ότι η απόδοση του DISK μπορεί να είναι ευαίσθητη στη διαθεσιμότητα πληροφοριών βάθους για την εποπτεία και εξακολουθούν να υπάρχουν προκλήσεις στον χειρισμό περιστροφών και μεγάλων μετασχηματισμών εντός επιπέδου. Παρ' όλα αυτά, το DISK αντιπροσωπεύει μια σημαντική πρόοδο στον τομέα της μάθησης τοπικών χαρακτηριστικών, δίνοντας υποσχέσεις για τη βελτίωση διάφορων εργασιών όρασης υπολογιστών.

Ροή εργασίας

Ο αλγόριθμος DISK (DIScrete Keypoints) ακολουθεί μια μοναδική ροή εργασίας που συνδυάζει αρχές ενισχυτικής μάθησης με πιθανολογική μοντελοποίηση για την εκμάθηση τοπικών χαρακτηριστικών της εικόνας. Τα βήματα είναι τα εξής:

Προετοιμασία δεδομένων: Ο DISK ξεκινά με ένα σύνολο δεδομένων που περιέχει εικόνες και, ιδανικά, αντίστοιχους χάρτες βάθους. Αυτοί οι χάρτες βάθους χρησιμοποιούνται για την παροχή ισχυρής εποπτείας για την εκπαίδευση.

Δίκτυο εξαγωγής χαρακτηριστικών: Το DISK χρησιμοποιεί μια αρχιτεκτονική νευρωνικού δικτύου, η οποία συχνά βασίζεται σε μια παραλλαγή της αρχιτεκτονικής U-Net. Αυτό το δίκτυο χρησιμεύει ως η ραχοκοκαλιά εξαγωγής χαρακτηριστικών. Στην περίπτωση του DISK, το δίκτυο αυτό έχει σχεδιαστεί με ένα μόνο συνελκτικό στρώμα ανά

μπλοκ και χρησιμοποιεί κανονικοποίηση περιπτώσεων και μη γραμμικότητας PReLU.

Παραγωγή σημείων κλειδιών: Η βασική καινοτομία του DISK έγκειται στον μηχανισμό διακριτής παραγωγής σημείων κλειδιών. Διαιρεί την εικόνα εισόδου σε ένα πλέγμα κελιών και μαθαίνει να αποδίδει ένα σκορ πιθανότητας σε κάθε κελί, υποδεικνύοντας την πιθανότητα να δημιουργηθεί εκεί ένα σημείο κλειδί. Τα κελιά του πλέγματος με υψηλή βαθμολογία είναι πιο πιθανό να παράγουν σημεία κλειδιά.

Πιθανολογική δειγματοληψία σημείων κλειδιών: Κατά τη διάρκεια της εκπαίδευσης, τα σημεία κλειδιά επιλέγονται δειγματοληπτικά από το πλέγμα με βάση τις πιθανότητες που τους αποδίδονται. Το δίκτυο ενθαρρύνει την επιλογή σημείων κλειδιών σε περιοχές με ισχυρά χαρακτηριστικά της εικόνας, συνήθως σε γωνίες, ακμές ή περιοχές πλούσιες σε υφή.

Παραγωγή περιγραφής χαρακτηριστικών: Για κάθε σημείο κλειδί που ανιχνεύεται, το DISK παράγει μια περιγραφή χαρακτηριστικών. Αυτοί οι περιγραφείς συλλαμβάνουν τις τοπικές πληροφορίες γύρω από κάθε σημείο κλειδί και είναι κρίσιμοι για τις εργασίες αντιστοίχισης και αναγνώρισης.

Υπολογισμός αντιστοίχισης και ανταμοιβής: Για τη βελτιστοποίηση της επιλογής σημείων κλειδιών, το DISK χρησιμοποιεί ένα πλαίσιο ενισχυτικής μάθησης. Η ανταμοιβή για ένα σύνολο σημείων-κλειδιών υπολογίζεται με βάση την ποιότητα των αντιστοιχίσεων που επιτρέπουν. Εάν ένα σημείο κλειδί οδηγεί σε σωστές αντιστοιχίες, λαμβάνει θετική ανταμοιβή, ενώ εάν οδηγεί σε λανθασμένες αντιστοιχίες, τιμωρείται. Αυτό το σήμα ανταμοιβής καθοδηγεί το δίκτυο να μάθει διακριτικά σημεία κλειδιά.

Εκπαίδευση: Το DISK εκπαιδεύει το δίκτυο εξαγωγής χαρακτηριστικών και την πολιτική επιλογής σημείων κλειδιών χρησιμοποιώντας ενισχυτική μάθηση. Βελτιστοποιεί τις παραμέτρους του δικτύου, όπως τα φίλτρα συνέλιξης, τις παραμέτρους κανονικοποίησης των περιπτώσεων και την πολιτική δημιουργίας σημείων κλειδιών, για να μεγιστοποιήσει την αναμενόμενη ανταμοιβή. Η διαδικασία αποσκοπεί στο να καταστήσει τα keypoints πιο κατατοπιστικά και ανθεκτικά.

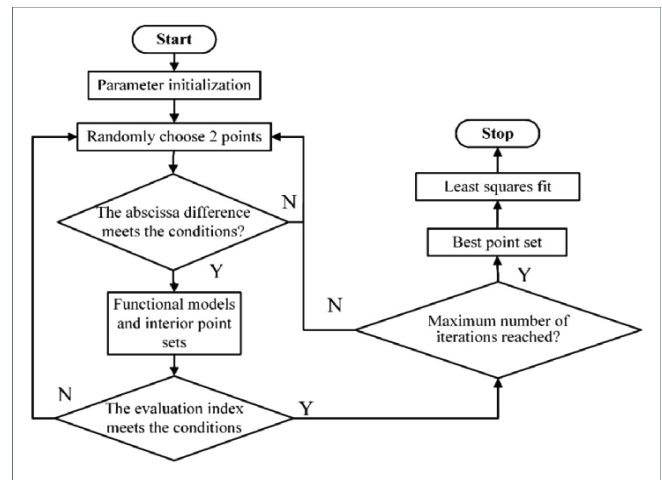
Συμπερασματολογία: Κατά την εξαγωγή συμπερασμάτων, το DISK εφαρμόζει το εκπαιδευμένο μοντέλο σε νέες εικόνες. Χρησιμοποιεί τη μαθημένη πολιτική δημιουργίας σημείων κλειδιών για να επιλέξει σημεία κλειδιών και δημιουργεί περιγραφείς χαρακτηριστικών γύρω από αυτά τα σημεία κλειδιών.

Αντιστοίχιση και μεταγενέστερες εργασίες: Τα παραγόμενα σημεία-κλειδιά και οι περιγραφείς μπορούν να χρησιμοποιηθούν για διάφορες εργασίες υπολογιστικής όρασης, όπως η αντιστοίχιση εικόνων, η στερεοσκοπική όραση ή η ανακατασκευή πολλαπλών όψεων. Η ικανότητα του DISK να παράγει υψηλής ποιότητας σημεία κλειδιά και περιγραφείς είναι ζωτικής σημασίας για την επιτυχία αυτών των εργασιών.

Συνοπτικά, το DISK αξιοποιεί την ενισχυτική μάθηση για την εκμάθηση ενός πιθανοτικού μοντέλου για την παραγωγή σημείων κλειδιών και περιγραφών. Βελτιστοποιώντας τις σωστές αντιστοιχίες χαρακτηριστικών, μπορεί να παράγει διακριτικά σημεία κλειδιά που αμφισβητούν τις παραδοσιακές υποθέσεις για την καλή επιλογή σημείων κλειδιών. Αυτή η ροή εργασίας οδηγεί σε μια ισχυρή και

ευέλικτη μέθοδο για την εκμάθηση τοπικών χαρακτηριστικών στην όραση υπολογιστών.

RANSAC (Random Sample Consensus)



Εικόνα 5: Flowchart του RANSAC

Το RANSAC είναι μια οικογένεια αλγορίθμων που χρησιμοποιούνται για την ισχυρή εκτίμηση των παραμέτρων του μοντέλου από ένα σύνολο δεδομένων που περιέχουν ακραίες τιμές. Αυτοί οι αλγόριθμοι λειτουργούν επιλέγοντας επαναληπτικά τυχαία υποσύνολα των δεδομένων, προσαρμόζοντας ένα μοντέλο σε κάθε υποσύνολο και στη συνέχεια αξιολογώντας πόσο καλά το μοντέλο προσαρμόζεται στο σύνολο των δεδομένων. Ακολουθούν ορισμένοι από τους αλγορίθμους RANSAC και οι παραλλαγές τους:

Standard RANSAC

Αυτός είναι ο βασικός αλγόριθμος RANSAC. Είναι μια εύρωστη και ευρέως χρησιμοποιούμενη μέθοδος για την εκτίμηση μοντέλων από θορυβώδη δεδομένα "μολυσμένα" με ακραίες τιμές. Η ροή εργασίας του ξεκινά με την τυχαία επιλογή ενός ελάχιστου υποσυνόλου σημείων δεδομένων για τον σχηματισμό ενός υποθετικού μοντέλου. Σε επόμενες επαναλήψεις, ο RANSAC αξιολογεί αυτό το μοντέλο εντοπίζοντας σημεία δεδομένων που εντάσσονται εντός ενός συγκεκριμένου ορίου των προβλέψεων του μοντέλου, θεωρώντας τα ως ακραία σημεία. Στη συνέχεια, ο αλγόριθμος συγκεντρώνει αυτά τα ατελή σημεία σε ένα σύνολο συναίνεσης. Μετά από έναν προκαθορισμένο αριθμό επαναλήψεων ή ένα κριτήριο διακοπής, ο RANSAC επιλέγει το μοντέλο με το μεγαλύτερο σύνολο συναίνεσης ως το καλύτερο μοντέλο, θεωρώντας ότι είναι η πιο ισχυρή εκτίμηση. Προαιρετικά, το καλύτερο μοντέλο μπορεί να βελτιωθεί περαιτέρω χρησιμοποιώντας όλες τις ακραίες τιμές, όπως μέσω μιας διαδικασίας βελτιστοποίησης ελαχίστων τετραγώνων. Τα πλεονεκτήματα του RANSAC έγκεινται στην ανθεκτικότητά του στις ακραίες τιμές, στη σχετικά απλή εφαρμογή και στην εφαρμογή του σε διάφορες εργασίες όρασης υπολογιστών. Ωστόσο, έχει μειονεκτήματα, όπως το υπολογιστικό κόστος, η ευαισθησία στις επιλογές των παραμέτρων, ο στοχαστικός χαρακτήρας που απαιτεί πολλαπλές εκτελέσεις και η έλλειψη εγγύησης για την εύρεση του παγκόσμιου βέλτιστου. Παρά τους περιορισμούς αυτούς, παραμένει ένα πολύτιμο εργαλείο για το χειρισμό θορυβωδών δεδομένων και την ισχυρή εκτίμηση μοντέλων.

Οι ερευνητές έχουν αναπτύξει επεκτάσεις και βελτιώσεις για να μετριάσουν ορισμένους από αυτούς τους περιορισμούς.

Locally Optimized RANSAC (LO-RANSAC)

Ο τοπικά βελτιστοποιημένος αλγόριθμος RANSAC (LO-RANSAC) είναι μια βελτίωση του τυπικού αλγορίθμου RANSAC που έχει σχεδιαστεί για να βελτιώσει την αποδοτικότητα και την ακρίβειά του. Η ροή εργασίας του LO-RANSAC ακολουθεί τα θεμελιώδη βήματα του RANSAC, αλλά περιλαμβάνει βήματα τοπικής βελτιστοποίησης για τη βελτίωση των εκτιμήσεων του μοντέλου.

Ο αλγόριθμος ενισχύει την ευρωστία και την ακρίβεια του αλγορίθμου Random Sample Consensus (RANSAC) ενσωματώνοντας βήματα τοπικής βελτιστοποίησης. Ο LO-RANSAC ξεκινά επιλέγοντας τυχαία ένα ελάχιστο υποσύνολο σημείων δεδομένων και στη συνέχεια υπολογίζει ένα αρχικό μοντέλο με βάση αυτά τα δείγματα. Αξιολογεί όλα τα σημεία δεδομένων σε σχέση με αυτό το μοντέλο, ταξινομώντας τα ως ακραία ή μη ακραία σημεία. Η βασική καινοτομία έγκειται στο βήμα της τοπικής βελτιστοποίησης, όπου η LO-RANSAC εντοπίζει συστάδες αθεράπευτων δεδομένων και εφαρμόζει τεχνικές τοπικής βελτιστοποίησης για τη λεπτομερή ρύθμιση των παραμέτρων του μοντέλου εντός αυτών των συστάδων. Το βελτιωμένο μοντέλο συγκρίνεται με το προηγούμενο καλύτερο μοντέλο με βάση ένα σκορ καταλληλότητας, με την υιοθέτηση μοντέλων με καλύτερη προσαρμογή. Η διαδικασία αυτή επαναλαμβάνεται για έναν προκαθορισμένο αριθμό επαναλήψεων ή έως ότου ικανοποιηθεί ένα κριτήριο διακοπής, και το τελικό αποτέλεσμα είναι το καλύτερο μοντέλο που βρέθηκε μαζί με τις σχετικές ακραίες τιμές. Ο συνδυασμός τυχαίας δειγματοληψίας και τοπικής βελτιστοποίησης της LO-RANSAC την καθιστά ιδιαίτερα αποτελεσματική στο χειρισμό θορυβωδών δεδομένων και στην παραγωγή ακριβών μοντέλων, καθιστώντας την ανεκτίμητη σε διάφορες εφαρμογές όρασης υπολογιστών.

Πλεονεκτήματα

Το LO-RANSAC προσφέρει πολλά πλεονεκτήματα:

Βελτιωμένη ακρίβεια: Με την επαναληπτική βελτίωση των εκτιμήσεων του μοντέλου με βάση τις ακραίες τιμές, το LO-RANSAC παράγει ακριβέστερα αποτελέσματα σε σύγκριση με το τυπικό RANSAC, ιδίως παρουσία θορύβου και ακραίων τιμών.

Ανθεκτικότητα: Όπως και το RANSAC, το LO-RANSAC είναι ανθεκτικό σε ακραίες τιμές, καθιστώντας το κατάλληλο για εφαρμογές που ασχολούνται με δεδομένα που περιέχουν σημαντικό θόρυβο ή λανθασμένες μετρήσεις.

Αποδοτικότητα: Το LO-RANSAC συγκλίνει συνήθως ταχύτερα από το τυπικό RANSAC, καθώς αξιοποιεί την τοπική βελτιστοποίηση για την επαναληπτική βελτίωση του μοντέλου. Αυτή η αποδοτικότητα μπορεί να είναι ζωτικής σημασίας για εφαρμογές πραγματικού χρόνου ή εφαρμογές με περιορισμένους πόρους.

Ευελιξία: LO-RANSAC μπορεί να προσαρμοστεί σε διάφορες εργασίες υπολογιστικής όρασης και γεωμετρικής εκτίμησης, από την εκτίμηση θεμελιωδών μητρών στη στερεοσκοπική όραση έως την εκτίμηση ομογραφίας στην εγγραφή εικόνων.

Μειονεκτήματα

Ωστόσο, η LO-RANSAC έχει ορισμένα μειονεκτήματα:

Αυξημένη υπολογιστική πολυπλοκότητα: Το βήμα τοπικής βελτιστοποίησης προσθέτει υπολογιστική επιβάρυνση, καθιστώντας το LO-RANSAC πιο απαιτητικό σε πόρους από το τυπικό RANSAC.

Ευαισθησία στις παραμέτρους: Όπως και η RANSAC, η LO-RANSAC απαιτεί ρύθμιση παραμέτρων και η απόδοσή της μπορεί να είναι ευαίσθητη στην επιλογή αυτών των παραμέτρων. Η εύρεση του βέλτιστου συνόλου παραμέτρων μπορεί να είναι μια μη τετριμμένη εργασία.

Δεν υπάρχει εγγύηση γενικής βελτιστότητας: Η LO-RANSAC, παρόμοια με την RANSAC, δεν εγγυάται την εύρεση του παγκόσμιου βέλτιστου. Μπορεί ακόμη να είναι επιρρεπής στη σύγκλιση σε τοπικά ελάχιστα κατά τη διαδικασία βελτιστοποίησης.

M-estimator Sample and Consensus (MSAC)

Ο αλγόριθμος M-estimator Sample and Consensus (MSAC) είναι μια παραλλαγή του αλγορίθμου Random Sample Consensus (RANSAC) που αντιμετωπίζει τις αδυναμίες της κλασικής προσέγγισης RANSAC, ιδίως όταν υπάρχει μεγάλος αριθμός ακραίων τιμών. Ο MSAC χρησιμοποιεί μια συνάρτηση M-εκτιμητή για να σταθμίζει διαφορετικά την επιρροή των σημείων δεδομένων, δίνοντας μεγαλύτερη βαρύτητα σε εκείνα που είναι πιθανότερο να είναι ακραία και μικρότερη βαρύτητα σε πιθανές ακραίες τιμές. Ακολουθεί μια πιο λεπτομερής ματιά στον τρόπο λειτουργίας της MSAC:

Αρχικοποίηση: Το MSAC ξεκινά επιλέγοντας τυχαία τον ελάχιστο αριθμό σημείων δεδομένων που απαιτούνται για την εκτίμηση του μοντέλου (π.χ. τρία για την προσαρμογή μιας ομοιογραφίας σε έναν προοπτικό μετασχηματισμό).

Εκτίμηση μοντέλου: Χρησιμοποιώντας αυτά τα δειγματοληπτικά σημεία, εκτιμάται ένα αρχικό μοντέλο. Στην περίπτωση της εκτίμησης ομογραφίας, αυτό θα περιλάμβανε την εύρεση του πίνακα μετασχηματισμού που ταιριάζει καλύτερα στα επιλεγμένα σημεία.

Ταξινόμηση εισερχομένων: Στη συνέχεια, η MSAC αξιολογεί όλα τα σημεία δεδομένων με βάση αυτό το αρχικό μοντέλο. Στα σημεία δεδομένων που ταιριάζουν καλά με το μοντέλο (δηλαδή είναι πιθανό να είναι ακραία) δίνεται μεγαλύτερη βαρύτητα με βάση τη συμφωνία τους με το μοντέλο. Για τον υπολογισμό αυτών των βαρών χρησιμοποιείται η συνάρτηση M-estimator.

Επιλογή συναινετικού συνόλου: Αντί να χρησιμοποιείται ένα σταθερό κατώφλι όπως στο RANSAC, το MSAC χρησιμοποιεί ένα δυναμικό κατώφλι με βάση τα εκτιμώμενα βάρη. Αυτό το κατώφλι καθορίζει ποια σημεία δεδομένων θεωρούνται ως ακραία. Τα σημεία δεδομένων με βάρη υψηλότερα από το κατώφλι περιλαμβάνονται στο σύνολο συναίνεσης.

Τοπική βελτίωση: Η MSAC ενσωματώνει συχνά βήματα τοπικής βελτιστοποίησης, παρόμοια με την LO-RANSAC. Εντοπίζει συστάδες ατελειών και εκτελεί τοπική βελτιστοποίηση για τη βελτίωση της ακρίβειας του μοντέλου εντός αυτών των συστάδων.

Αξιολόγηση μοντέλου: Το MSAC επαναλαμβάνει αυτά τα βήματα για έναν προκαθορισμένο αριθμό επαναλήψεων ή μέχρι να ικανοποιηθεί ένα κριτήριο διακοπής. Σε κάθε

επανάληψη, παρακολουθεί το μέγεθος του συνόλου συναίνεσης (αριθμός ατελειών) και την ποιότητα του μοντέλου.

Επιλογή αποτελεσμάτων: Μετά τις επαναλήψεις, η MSAC επιλέγει το μοντέλο που παρήγαγε το μεγαλύτερο σύνολο συναίνεσης (δηλαδή το μοντέλο με τις περισσότερες ατέλειες). Αυτό το μοντέλο θεωρείται η καλύτερη εκτίμηση της υποκείμενης δομής παρουσία ακραίων τιμών.

Πλεονεκτήματα

Η MSAC είναι ισχυρή στο χειρισμό υψηλού ποσοστού ακραίων τιμών.

Παρέχει έναν μηχανισμό για τη διαφορετική στάθμιση των σημείων δεδομένων, δίνοντας μεγαλύτερη σημασία σε εκείνα που είναι πιθανό να είναι ακραία.

Η δυναμική καταστροφή με βάση τα βάρη προσαρμόζεται στην κατανομή των δεδομένων, καθιστώντας την πιο ισχυρή από τα σταθερά κατώφλια.

Μειονεκτήματα

Η MSAC μπορεί να είναι υπολογιστικά δαπανηρή λόγω των πολλαπλών επαναλήψεων και των τοπικών βελτιστοποιήσεων.

Η επιλογή της συνάρτησης M-estimator και των παραμέτρων της μπορεί να επηρεάσει την απόδοση της MSAC και μπορεί να χρειαστεί να ρυθμιστεί για συγκεκριμένες εφαρμογές.

Συνοψίζοντας, η MSAC είναι μια ισχυρή επέκταση της RANSAC που είναι ιδιαίτερα χρήσιμη σε περιπτώσεις όπου υπάρχουν πολλές ακραίες τιμές στα δεδομένα, παρέχοντας μια πιο ισχυρή και ακριβή εκτίμηση των παραμέτρων του μοντέλου.

Progressive Sample Consensus (PROSAC)

Ο αλγόριθμος Progressive Sample Consensus (PROSAC) είναι μια προηγμένη παραλλαγή του αλγορίθμου Random Sample Consensus (RANSAC) που έχει σχεδιαστεί για τη βελτίωση της αποτελεσματικότητας και της ακρίβειας της εκτίμησης των παραμέτρων του μοντέλου, ιδίως παρουσία μεγάλου αριθμού ακραίων τιμών. Ο PROSAC στοχεύει στην εύρεση της καλύτερης προσαρμογής του μοντέλου με λιγότερες επαναλήψεις από τον παραδοσιακό RANSAC. Η ροή εργασίας του έχει ως εξής:

Αρχικοποίηση: Η PROSAC ξεκινά με ένα τυχαία επιλεγμένο ελάχιστο δείγμα, όπως ακριβώς και η RANSAC. Αυτό το δείγμα χρησιμοποιείται για την εκτίμηση ενός αρχικού μοντέλου.

Ταξινόμηση εισερχομένων: Σε αντίθεση με το RANSAC, το PROSAC χρησιμοποιεί μια επαναληπτική στρατηγική για την ταξινόμηση των εισερχομένων και των ακραίων τιμών. Αρχικά θεωρεί όλα τα σημεία δεδομένων ως υπονήφια και αξιολογεί την καταλληλότητά τους για το τρέχον μοντέλο.

Βαθμολόγηση και ταξινόμηση: Σε κάθε σημείο δεδομένων αποδίδεται μια βαθμολογία με βάση την καταλληλότητά του για το μοντέλο. Αυτή η βαθμολογία αντικατοπτρίζει πόσο καλά ένα σημείο συμφωνεί με το τρέχον μοντέλο. Στη

συνέχεια, τα σημεία ταξινομούνται κατά φθίνουσα σειρά της βαθμολογίας τους.

Επιλογή δείγματος: PROSAC εισάγει μια νέα στρατηγική δειγματοληψίας. Αντί για τυχαία επιλογή δειγμάτων όπως κάνει η RANSAC, η PROSAC επιλέγει δείγματα με τρόπο που λαμβάνει υπόψη τις βαθμολογίες των σημείων δεδομένων. Ξεκινά με τα σημεία με τη μεγαλύτερη βαθμολογία και σχηματίζει προοδευτικά μεγαλύτερα δείγματα.

Εκτίμηση μοντέλου: Για κάθε δείγμα, η PROSAC εκτιμά ένα μοντέλο χρησιμοποιώντας τα επιλεγμένα σημεία δεδομένων.

Επέκταση συναινετικού συνόλου: Το σύνολο συναίνεσης περιλαμβάνει αρχικά μόνο τα σημεία δεδομένων από το τρέχον δείγμα που συμφωνούν με το εκτιμώμενο μοντέλο. Στις επόμενες επαναλήψεις, η PROSAC επεκτείνει σταδιακά το σύνολο συναίνεσης περιλαμβάνοντας σημεία δεδομένων που συμφωνούν με το τρέχον μοντέλο.

Αξιολόγηση της ποιότητας του μοντέλου: Σε κάθε επανάληψη, η PROSAC αξιολογεί την ποιότητα του εκτιμώμενου μοντέλου. Εάν το τρέχον μοντέλο είναι καλύτερο από το προηγούμενο καλύτερο μοντέλο (π.χ. έχει μεγαλύτερο σύνολο συναίνεσης), επικαιροποιεί την εκτίμηση του καλύτερου μοντέλου.

Τερματισμός: Η PROSAC τερματίζει όταν ικανοποιείται ένα κριτήριο διακοπής, όπως η επίτευξη ενός προκαθορισμένου αριθμού επαναλήψεων ή όταν η ποιότητα του μοντέλου ξεπεράσει ένα ορισμένο όριο.

Πλεονεκτήματα

Αποδοτικότητα: Η PROSAC προσδιορίζει αποτελεσματικά το καλύτερο μοντέλο, λαμβάνοντας υπόψη τα πιο κατατοπιστικά σημεία δεδομένων νωρίς στη διαδικασία. Αυτό μειώνει τον αριθμό των επαναλήψεων σε σύγκριση με το παραδοσιακό RANSAC.

Ανθεκτικότητα: Με την προοδευτική επέκταση του συνόλου συναίνεσης και την εξέταση σημείων με χαμηλότερη βαθμολογία αργότερα στη διαδικασία, η PROSAC είναι πιο ανθεκτική στο θόρυβο και τις ακραίες τιμές.

Βελτιωμένη ακρίβεια: Η PROSAC συγκλίνει συχνά σε καλύτερη λύση από την RANSAC, καθώς εξερευνά έναν πιο ενημερωμένο χώρο αναζήτησης.

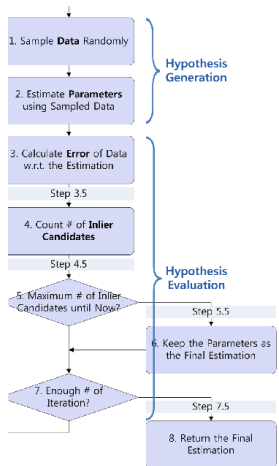
Μειονεκτήματα

Πολυπλοκότητα: Η PROSAC εισάγει πρόσθετη πολυπλοκότητα σε σύγκριση με την RANSAC, ιδίως στην επιλογή δειγμάτων και στην ταξινόμηση με βάση τις βαθμολογίες.

Ευαισθησία παραμέτρων: Η απόδοση της PROSAC μπορεί να είναι ευαίσθητη στην επιλογή παραμέτρων, όπως η συνάρτηση βαθμολόγησης και τα κριτήρια διακοπής.

Συνοπτικά, ο PROSAC είναι μια ισχυρή βελτίωση του αλγορίθμου RANSAC, γνωστού για την αποτελεσματικότητά και την ευρωστία του στην αντιμετώπιση ακραίων τιμών. Βελτιώνει την ποιότητα της εκτίμησης μοντέλων με την αποτελεσματική εξερεύνηση του χώρου των πιθανών λύσεων, καθιστώντας τον κατάλληλο

για διάφορες εργασίες υπολογιστικής όρασης και γεωμετρικής εκτίμησης.



Εικόνα 6: Δομή οικογένειας αλγορίθμων RANSAC

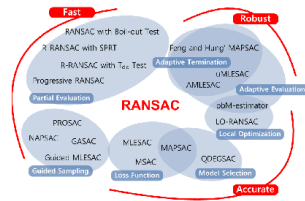
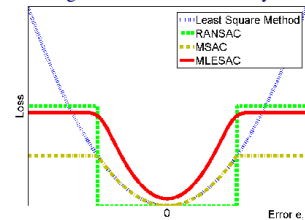


Figure 2: RANSAC Family



Least Median of Squares (LMedS)

Η ελαχίστη διάμεσος των τετραγώνων (LMedS) είναι μια ισχυρή στατιστική μέθοδος που χρησιμοποιείται για την ισχυρή προσαρμογή μοντέλων, ιδίως παρουσία ακραίων τιμών και θορυβωδών δεδομένων. Η LMedS στοχεύει στην εύρεση των παραμέτρων του μοντέλου που ελαχιστοποιούν τη διάμεσο των τετραγωνικών καταλοίπων, καθιστώντας την λιγότερο ευαίσθητη στις ακραίες τιμές σε σύγκριση με τις μεθόδους που ελαχιστοποιούν τη μέση τιμή των τετραγωνικών καταλοίπων, όπως η μέθοδος ελαχίστων τετραγώνων (LS). Ακολουθούν περισσότερες λεπτομέρειες σχετικά με την LMedS:

Ροή εργασίας:

Data Preparation: Η LMedS ξεκινά με ένα σύνολο δεδομένων που περιέχει παρατηρημένα σημεία δεδομένων, τα οποία συνήθως αναπαρίστανται ως ζεύγη (x, y) , όπου x είναι η ανεξάρτητη μεταβλητή και y είναι η εξαρτημένη μεταβλητή.

Υπόθεση μοντέλου: Υποτίθεται μια υπόθεση μοντέλου που συνδέει την ανεξάρτητη μεταβλητή (x) με την εξαρτημένη μεταβλητή (y). Το μοντέλο μπορεί να είναι γραμμικό, πολυωνυμικό ή οποιασδήποτε άλλης κατάλληλης μορφής ανάλογα με το πρόβλημα.

Υπολογισμός υπολοίπων: Για κάθε σημείο δεδομένων, το υπόλειμμα υπολογίζεται ως η κάθετη απόσταση μεταξύ της παρατηρούμενης τιμής (y) και της τιμής που προβλέπει το μοντέλο στο συγκεκριμένο σημείο ($y_{\text{predicted}}$). Μαθηματικά, το υπόλειμμα (r) δίνεται από τη σχέση: $r = y - y_{\text{predicted}}$.

Τετραγωνικά κατάλοιπα: Τα τετραγωνισμένα κατάλοιπα υπολογίζονται για όλα τα σημεία δεδομένων με τον τετραγωνισμό κάθε τιμής υπολοίπου: r^2 .

Ταξινόμηση: Τα τετραγωνικά κατάλοιπα ταξινομούνται σε αύξουσα σειρά.

Υπολογισμός διάμεσου: LMedS επιλέγει τις παραμέτρους του μοντέλου που αντιστοιχούν στα τετραγωνισμένα

κατάλοιπα στη διάμεση θέση. Αυτό σημαίνει ότι τα μισά από τα τετραγωνισμένα κατάλοιπα είναι μικρότερα ή ίσα με τη διάμεση τιμή και τα άλλα μισά είναι μεγαλύτερα.

Εκτίμηση μοντέλου: Χρησιμοποιώντας τα επιλεγμένα τετραγωνικά κατάλοιπα, το LMedS εκτιμά τις παραμέτρους του μοντέλου (π.χ. συντελεστές μιας γραμμικής εξίσωσης) που οδηγούν στη διάμεση τιμή των τετραγωνικών καταλοίπων.

Απόρριξη ακραίων τιμών: Τα ακραία σημεία αναγνωρίζονται ως σημεία δεδομένων των οποίων τα τετραγωνικά κατάλοιπα είναι μεγαλύτερα από τη διάμεση τιμή των τετραγωνικών καταλοίπων. Αυτές οι ακραίες τιμές αποκλείονται από την τελική προσαρμογή του μοντέλου.

Πλεονεκτήματα

Ανθεκτικότητα: Το LMedS είναι ιδιαίτερα ανθεκτικό στην παρουσία ακραίων τιμών και θορύβου. Μπορεί να αντισταθεί αποτελεσματικά στην επιρροή μερικών ακραίων σημείων δεδομένων.

Δεν υπάρχουν υποθέσεις σχετικά με την κατανομή των σφαλμάτων: Σε αντίθεση με μεθόδους όπως η εκτίμηση μέγιστης πιθανοφάνειας (MLE), η LMedS δεν υποθέτει συγκεκριμένη κατανομή για τα σφάλματα, καθιστώντας την εφαρμόσιμη σε διάφορα σενάρια.

Ανίχνευση ακραίων τιμών: Η LMedS αναγνωρίζει εγγενώς τις ακραίες τιμές, καθιστώντας την πολύτιμη για εργασίες καθαρισμού και προεπεξεργασίας δεδομένων.

Εφαρμογή: LMedS μπορεί να εφαρμοστεί σε ένα ευρύ φάσμα προβλημάτων μοντελοποίησης, συμπεριλαμβανομένης της γραμμικής παλινδρόμησης, της ισχυρής εκτίμησης και της προσαρμογής γεωμετρικών μοντέλων.

Μειονεκτήματα

Υπολογιστικό κόστος: LMedS μπορεί να είναι υπολογιστικά δαπανηρή, ιδίως για μεγάλα σύνολα δεδομένων, καθώς περιλαμβάνει την ταξινόμηση όλων των τετραγωνικών καταλοίπων.

Δεν είναι βέλτιστο για δεδομένα που κυριαρχούνται από ακροτάτες: Το LMedS στοχεύει στην ελαχιστοποίηση της διαμέσου των τετραγωνικών καταλοίπων. Εάν το σύνολο δεδομένων περιέχει λίγες ακραίες τιμές σε σύγκριση με τις ακραίες τιμές, άλλες μέθοδοι όπως η RANSAC μπορεί να αποδίδουν καλύτερα.

Γενικά, η LMedS είναι μια στιβαρή και ευρέως χρησιμοποιούμενη μέθοδος για την προσαρμογή μοντέλων όταν πρόκειται για δεδομένα που ενδέχεται να περιέχουν ακραίες τιμές ή σημαντικό θόρυβο. Παρέχει έναν αξιόπιστο τρόπο εκτίμησης των παραμέτρων του μοντέλου, ενώ παράλληλα ανιχνεύει και μετριάξει την επίδραση των ακραίων τιμών, καθιστώντας την πολύτιμη σε τομείς όπως η όραση υπολογιστών, η στατιστική και η μηχανική μάθηση.

Randomized RANSAC (R-RANSAC)

Ο Randomized RANSAC (R-RANSAC) είναι μια παραλλαγή του κλασικού αλγορίθμου RANSAC, ο οποίος χρησιμοποιείται ευρέως για την ισχυρή προσαρμογή μοντέλων παρουσία ακραίων τιμών και θορυβωδών δεδομένων. Ο R-RANSAC εισάγει μια τυχαιοποιημένη στρατηγική δειγματοληψίας για τη βελτίωση της αποδοτικότητας του RANSAC, διατηρώντας παράλληλα την ευρωστία του.

Ροή εργασίας:

Προετοιμασία δεδομένων: Το R-RANSAC ξεκινά με ένα σύνολο δεδομένων που περιέχει παρατηρημένα σημεία δεδομένων, τα οποία συνήθως αναπαρίστανται ως ζεύγη (x, y) , όπου x είναι η ανεξάρτητη μεταβλητή και y η εξαρτημένη μεταβλητή.

Υπόθεση μοντέλου: Υποτίθεται μια υπόθεση μοντέλου που συσχετίζει την ανεξάρτητη μεταβλητή (x) με την εξαρτημένη μεταβλητή (y). Το μοντέλο μπορεί να είναι γραμμικό, πολυωνυμικό ή οποιασδήποτε άλλης κατάλληλης μορφής ανάλογα με το πρόβλημα.

Τυχαία δειγματοληψία: Αντί της τυχαίας επιλογής ενός σταθερού αριθμού σημείων δεδομένων για κάθε επανάληψη RANSAC, η R-RANSAC χρησιμοποιεί μια τυχαία στρατηγική δειγματοληψίας. Επιλέγει τυχαία ένα ελάχιστο υποσύνολο σημείων δεδομένων που απαιτείται για την εκτίμηση του μοντέλου. Ο αριθμός των σημείων δεδομένων σε αυτό το υποσύνολο καθορίζεται από το ελάχιστο που απαιτείται για την εκτίμηση του μοντέλου (π.χ. δύο σημεία για μια γραμμή ή τρία για ένα επίπεδο).

Εκτίμηση μοντέλου: Χρησιμοποιώντας τα τυχαία δειγματοληπτικά σημεία δεδομένων, το R-RANSAC εκτιμά τις παραμέτρους του μοντέλου (π.χ. συντελεστές μιας γραμμικής εξίσωσης) που ταιριάζουν καλύτερα στα δεδομένα του υποσυνόλου. Αυτό το βήμα είναι παρόμοιο με το βήμα εκτίμησης στο τυπικό RANSAC.

Υπολογισμός υπολοίπων: Το R-RANSAC υπολογίζει τα κατάλοιπα για όλα τα σημεία δεδομένων με βάση το εκτιμώμενο μοντέλο. Τα κατάλοιπα αντιπροσωπεύουν την κάθετη απόσταση μεταξύ της παρατηρούμενης τιμής (y) και της τιμής που προβλέπει το μοντέλο στο συγκεκριμένο σημείο ($y_{\text{predicted}}$). Μαθηματικά, το υπόλειμμα (r) δίνεται από τη σχέση: $r = y - y_{\text{predicted}}$.

Επιλογή εισερχομένων: Τα σημεία δεδομένων με κατάλοιπα κάτω από ένα προκαθορισμένο κατώφλι (inliers) θεωρούνται ότι υποστηρίζουν το εκτιμώμενο μοντέλο, ενώ τα σημεία δεδομένων με κατάλοιπα πάνω από το κατώφλι (outliers) θεωρούνται ασυμβίβαστα με το μοντέλο.

Σύνολο συναίνεσης: Το R-RANSAC παρακολουθεί το σύνολο συναίνεσης, το οποίο περιλαμβάνει όλα τα ακραία σημεία που βρέθηκαν σε κάθε επανάληψη.

Επανάληψη και τερματισμός: Το R-RANSAC επαναλαμβάνει τα βήματα τυχαίας δειγματοληψίας, εκτίμησης μοντέλου και επιλογής ακραίων για προκαθορισμένο αριθμό επαναλήψεων ή μέχρι να ικανοποιηθεί μια συνθήκη τερματισμού. Η συνθήκη τερματισμού βασίζεται συχνά στο μέγεθος του συνόλου συναίνεσης ή σε ένα επιθυμητό επίπεδο εμπιστοσύνης.

Βελτίωση μοντέλου: Μετά τον τερματισμό των επαναλήψεων, το R-RANSAC συνήθως βελτιώνει το

μοντέλο χρησιμοποιώντας όλα τα σημεία δεδομένων στο σύνολο συναίνεσης, οδηγώντας σε μια πιο ακριβή εκτίμηση.

Πλεονεκτήματα

Αποδοτικότητα: Η τυχαιοποιημένη στρατηγική δειγματοληψίας της R-RANSAC μπορεί να μειώσει σημαντικά τον αριθμό των σημείων δεδομένων που εξετάζονται σε κάθε επανάληψη, καθιστώντας την υπολογιστικά πιο αποδοτική από την τυπική RANSAC.

Ανθεκτικότητα: Όπως και η RANSAC, η R-RANSAC είναι ανθεκτική σε ακραίες τιμές και θόρυβο στα δεδομένα, καθώς βασίζεται στη συναίνεση των ακραίων τιμών για την εκτίμηση του μοντέλου.

Εφαρμογή: Το R-RANSAC μπορεί να εφαρμοστεί σε διάφορα προβλήματα προσαρμογής μοντέλων, όπως προσαρμογή γραμμής, προσαρμογή επιπέδου και πιο σύνθετα μοντέλα.

Μειονεκτήματα

Τυχαίοτητα: Η τυχαία φύση του R-RANSAC μπορεί να οδηγήσει σε διαφορετικά αποτελέσματα σε κάθε εκτέλεση, αν και συνήθως συγκλίνει σε μια συνεπή λύση σε πολλαπλές εκτελέσεις.

Ρύθμιση παραμέτρων: Όπως η RANSAC, έτσι και η R-RANSAC απαιτεί τη ρύθμιση παραμέτρων, όπως το κατώφλι για την επιλογή των εισερχομένων και τον αριθμό των επαναλήψεων, οι οποίες μπορεί να επηρεάσουν την απόδοσή της.

Συνοπτικά, ο Randomized RANSAC (R-RANSAC) είναι μια παραλλαγή του αλγορίθμου RANSAC που εισάγει τυχαιοποιημένη δειγματοληψία για να βελτιώσει την υπολογιστική απόδοση, διατηρώντας παράλληλα την ευρωστία του RANSAC. Είναι ιδιαίτερα χρήσιμος για την επίλυση προβλημάτων όπου απαιτείται ισχυρή εκτίμηση μοντέλου, όπως στην όραση υπολογιστών, τη γεωμετρική μοντελοποίηση και την παλινδρόμηση με ακραίες τιμές.

Guided RANSAC

Ο καθοδηγούμενος RANSAC είναι μια επέκταση του κλασικού αλγορίθμου RANSAC (Random Sample Consensus) που ενσωματώνει μια στρατηγική καθοδηγούμενης δειγματοληψίας για να βελτιώσει την ευρωστία και την αποτελεσματικότητα της εκτίμησης του μοντέλου, ιδίως σε περιπτώσεις όπου ο παραδοσιακός RANSAC μπορεί να δυσκολευτεί.

Ροή εργασίας

Προετοιμασία δεδομένων: Το Guided RANSAC ξεκινά με ένα σύνολο δεδομένων που περιέχει παρατηρημένα σημεία δεδομένων, τα οποία συνήθως αναπαρίστανται ως ζεύγη (x, y) , όπου x είναι η ανεξάρτητη μεταβλητή και y η εξαρτημένη μεταβλητή. Υποθέτει ότι τα δεδομένα μπορεί να περιέχουν ακραίες τιμές και σφάλματα.

Υπόθεση μοντέλου: Υποτίθεται μια υπόθεση μοντέλου που συνδέει την ανεξάρτητη μεταβλητή (x) με την εξαρτημένη μεταβλητή (y). Το μοντέλο μπορεί να ποικίλλει ανάλογα με

το συγκεκριμένο πρόβλημα, αλλά τα συνήθη παραδείγματα περιλαμβάνουν γραμμικά μοντέλα, πολυωνυμικά μοντέλα ή γεωμετρικά μοντέλα.

Καθοδηγούμενη δειγματοληψία: Σε αντίθεση με το τυπικό RANSAC, το οποίο επιλέγει τυχαία έναν σταθερό αριθμό σημείων δεδομένων σε κάθε επανάληψη, το Guided RANSAC ενσωματώνει μια στρατηγική καθοδηγούμενης δειγματοληψίας. Η στρατηγική αυτή μπορεί να περιλαμβάνει την επιλογή σημείων δεδομένων βάσει ορισμένων κριτηρίων ή γνώσεων σχετικά με το πεδίο του προβλήματος. Για παράδειγμα, προηγούμενες πληροφορίες σχετικά με τις αναμενόμενες παραμέτρους του μοντέλου ή τη χωρική συνοχή των δεδομένων μπορούν να καθοδηγήσουν τη διαδικασία δειγματοληψίας.

Εκτίμηση μοντέλου: Χρησιμοποιώντας τα δειγματοληπτικά σημεία δεδομένων, το Guided RANSAC εκτιμά τις παραμέτρους του μοντέλου που ταιριάζουν καλύτερα στα δεδομένα. Η διαδικασία εκτίμησης είναι παρόμοια με εκείνη του τυπικού RANSAC.

Υπολογισμός υπολοίπων: Το Guided RANSAC υπολογίζει τα κατάλοιπα για όλα τα σημεία δεδομένων με βάση το εκτιμώμενο μοντέλο. Τα κατάλοιπα αντιπροσωπεύουν την κατακόρυφη ή κάθετη απόσταση μεταξύ της παρατηρούμενης τιμής (y) και της τιμής που προβλέπει το μοντέλο στο συγκεκριμένο σημείο ($y_{\text{predicted}}$).

Επιλογή εισερχομένων: Τα σημεία δεδομένων με κατάλοιπα κάτω από ένα προκαθορισμένο κατώφλι θεωρούνται αθεράπευτα, ενώ τα σημεία δεδομένων με κατάλοιπα πάνω από το κατώφλι θεωρούνται ακραία. Οι ακραίες τιμές θεωρείται ότι υποστηρίζουν το εκτιμώμενο μοντέλο.

Σύνολο συναίνεσης: Η καθοδηγούμενη RANSAC διατηρεί ένα σύνολο συναίνεσης, το οποίο περιλαμβάνει όλα τα ατελή σημεία που βρέθηκαν σε κάθε επανάληψη.

Επανάληψη και τερματισμός: Ο αλγόριθμος επαναλαμβάνει τα βήματα της καθοδηγούμενης δειγματοληψίας, της εκτίμησης μοντέλου και της επιλογής ακραίων για προκαθορισμένο αριθμό επαναλήψεων ή μέχρι να ικανοποιηθεί μια συνθήκη τερματισμού. Η συνθήκη τερματισμού βασίζεται συχνά στο μέγεθος του συνόλου συναίνεσης ή σε ένα επιθυμητό επίπεδο εμπιστοσύνης.

Βελτίωση μοντέλου: Μετά τον τερματισμό των επαναλήψεων, το Guided RANSAC συνήθως βελτιώνει το μοντέλο χρησιμοποιώντας όλα τα σημεία δεδομένων στο σύνολο συναίνεσης, οδηγώντας σε μια πιο ακριβή εκτίμηση.

Πλεονεκτήματα

Βελτιωμένη ανθεκτικότητα: Το Guided RANSAC έχει σχεδιαστεί για να χειρίζεται καταστάσεις στις οποίες η προηγούμενη γνώση ή οι πληροφορίες που αφορούν συγκεκριμένο τομέα μπορούν να καθοδηγήσουν τη διαδικασία δειγματοληψίας, καθιστώντας το πιο ανθεκτικό σε σενάρια με δύσκολα δεδομένα και ακραίες τιμές.

Αποδοτικότητα: Η στρατηγική καθοδηγούμενης δειγματοληψίας μπορεί να μειώσει τον αριθμό των επαναλήψεων που απαιτούνται για τη σύγκλιση σε μια λύση, βελτιώνοντας την υπολογιστική αποδοτικότητα σε σύγκριση με την τυπική RANSAC.

Εφαρμογή: Η καθοδηγούμενη RANSAC μπορεί να εφαρμοστεί σε ένα ευρύ φάσμα προβλημάτων προσαρμογής μοντέλων, συμπεριλαμβανομένων εκείνων της όρασης υπολογιστών, της επεξεργασίας εικόνας και της γεωμετρικής μοντελοποίησης.

Μειονεκτήματα

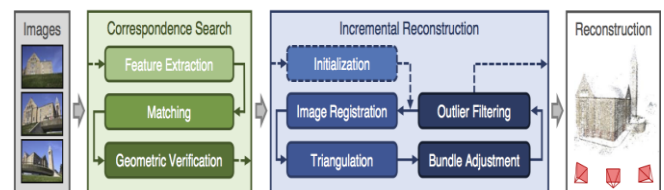
Συντονισμός παραμέτρων: Όπως και το RANSAC, το Guided RANSAC απαιτεί τη ρύθμιση παραμέτρων όπως το κατώφλι εισόδου και ο αριθμός των επαναλήψεων, οι οποίες μπορεί να επηρεάσουν την απόδοσή του και ενδέχεται να απαιτούν συντονισμό.

Απαιτήση καθοδήγησης: Η αποτελεσματική χρήση του Guided RANSAC βασίζεται στην προηγούμενη γνώση ή καθοδήγηση σχετικά με τα δεδομένα και τις αναμενόμενες παραμέτρους του μοντέλου. Σε περιπτώσεις όπου δεν υπάρχει τέτοια καθοδήγηση, η τυπική RANSAC μπορεί να είναι πιο κατάλληλη.

Συνοπτικά, το Guided RANSAC είναι μια επέκταση του RANSAC που ενσωματώνει καθοδηγούμενη δειγματοληψία με βάση την προηγούμενη γνώση ή πληροφορίες για συγκεκριμένο τομέα. Η προσέγγιση αυτή ενισχύει την ευρωστία και την αποτελεσματικότητα του αλγορίθμου σε σενάρια όπου ο παραδοσιακός RANSAC μπορεί να δυσκολεύεται να βρει ακριβή μοντέλα παρουσία ακραίων τιμών και θορυβωδών δεδομένων.

III. BIBΛΙΟΘΗΚΗ COLMAP

Το Colmap είναι μια σουίτα λογισμικού ανοικτού κώδικα σχεδιασμένη για την όραση υπολογιστών και τη φωτογραμμετρία με έμφαση στην ανακατασκευή με δομή από την κίνηση (SfM) και με στερεοφωνικό σύστημα πολλαπλών όψεων (MVS). Είναι ένα ευέλικτο εργαλείο που εφαρμόζεται στην αρχαιολογία, τη γεωχωρική ανάλυση, τη ρομποτική και τα γραφικά υπολογιστών.



Εικόνα 7: Workflow SfM στην COLMAP

Η Colmap περιλαμβάνει μια σειρά λειτουργιών:

- **Feature Extraction and Matching:** Το Colmap εξάγει διακριτικά χαρακτηριστικά από εικόνες και δημιουργεί αντιστοιχίες, ζωτικής σημασίας για την αντιστοίχιση αντίστοιχων σημείων σε πολλαπλές όψεις.
- **Incremental and Global SfM:** Η βιβλιοθήκη χρησιμοποιεί τόσο επαυξητικές όσο και καθολικές προσεγγίσεις βελτιστοποίησης για SfM. Η σταδιακή SfM επεξεργάζεται εικόνες σταδιακά, ενώ η καθολική SfM βελτιώνει ταυτόχρονα τις θέσεις της κάμερας και τα τρισδιάστατα σημεία σε όλες τις εικόνες, οδηγώντας ενδεχομένως σε πιο ακριβείς ανακατασκευές.

- Μέθοδος της Δέσμης: Το Colmap εκτελεί προσαρμογή δέσμης, βελτιώνοντας τις αρχικές εκτιμήσεις των θέσεων της κάμερας και των τρισδιάστατων σημείων. Αυτή η διαδικασία βελτιστοποίησης ευθυγραμμίζει το ανακατασκευασμένο τρισδιάστατο μοντέλο με τις εικόνες 2D, βελτιώνοντας τη συνολική ακρίβεια.
- Υποστήριξη πολλαπλών πλατφορμών: Το Colmap είναι ανεξάρτητο από πλατφόρμες και συμβατό με διάφορα λειτουργικά συστήματα, συμπεριλαμβανομένων των Linux, macOS και Windows. Αξιοποιεί επεξεργαστές πολλαπλών πυρήνων και πολλαπλές GPU για αποδοτική επεξεργασία, καθιστώντας το κατάλληλο για υπολογιστικό περιβάλλον υψηλών επιδόσεων.

Η Colmap διακρίνεται ιδιαίτερα στον τομέα της SfM:

Εκτίμηση της στάσης της κάμερας: Η βιβλιοθήκη εκτιμά τις θέσεις και τους προσανατολισμούς (πόζες) των καμερών σε σχέση με τη σκηνή, κάτι που είναι απαραίτητο για τον τριγωνισμό τρισδιάστατων σημείων από αντιστοιχίες εικόνων 2D.

Ανακατασκευή τρισδιάστατου νέφους σημείων: Το Colmap παράγει ένα αραιό νέφος σημείων που αναπαριστά την τρισδιάστατη δομή της σκηνής, τριγωνοποιώντας τις αντιστοιχίες χαρακτηριστικών σε πολλαπλές εικόνες.

Αρχική ανακατασκευή και σταδιακή βελτίωση: Ξεκινά τη διαδικασία SfM εκτελώντας μια αρχική ανακατασκευή με βάση τις αντιστοιχίες χαρακτηριστικών. Αυτό αποτελεί τη βάση για τα επόμενα βήματα βελτίωσης, τα οποία εκτελούνται σταδιακά, επιτρέποντας τον αποτελεσματικό χειρισμό μεγάλων συνόλων δεδομένων.

Οπτικοποίηση και εξαγωγή: Το Colmap επιτρέπει στους χρήστες να οπτικοποιούν το ανακατασκευασμένο τρισδιάστατο μοντέλο και να το εξάγουν σε διάφορες τυποποιημένες μορφές, όπως PLY ή OBJ.

Οι χρήστες θα πρέπει να γνωρίζουν ότι η επεξεργασία μεγάλων συνόλων δεδομένων μπορεί να απαιτεί σημαντικούς υπολογιστικούς πόρους, συμπεριλαμβανομένης μιας ισχυρής CPU και επαρκούς μνήμης RAM. Επιπλέον, οι χρήστες ενδέχεται να χρειαστεί να εξετάσουν τον συμβιβασμό μεταξύ της ακρίβειας ανακατασκευής και της υπολογιστικής απόδοσης βάσει των επιλεγμένων ρυθμίσεων και αλγορίθμων.

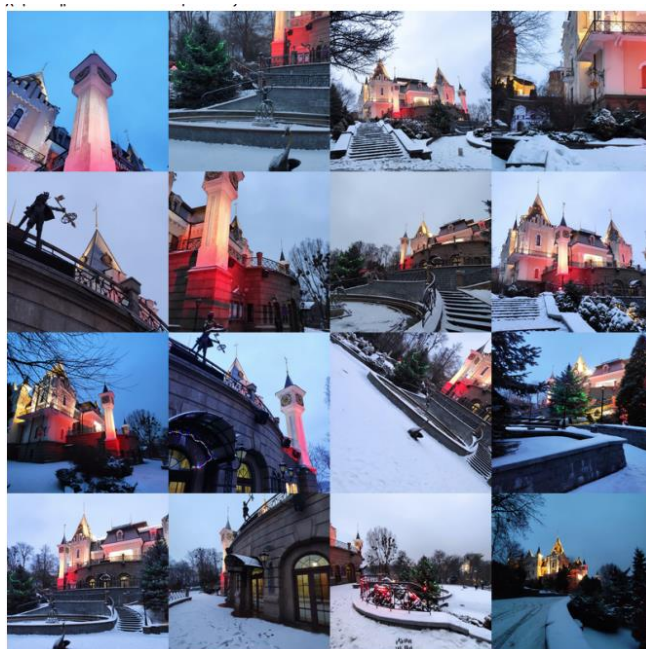
Συνοψίζοντας, το Colmap είναι μια ισχυρή και ευέλικτη σουίτα λογισμικού για τρισδιάστατη ανακατασκευή από εικόνες. Οι προηγμένοι αλγόριθμοί του, η υποστήριξη πολλαπλών πλατφορμών και η ενεργή κοινότητα το καθιστούν πολύτιμο εργαλείο για ένα ευρύ φάσμα εφαρμογών υπολογιστικής όρασης και φωτογραμμετρίας, με ιδιαίτερη υπεροχή στον τομέα της SfM. Για να αξιολογηθούν οι ανιχνευτές και οι περιγραφείς και να κατανοήσουμε καλύτερα ποιοι συνδυασμοί φαίνεται να λειτουργούν καλύτερα, διεξήχθη μια πληθώρα πειραμάτων. Όπως αναφέρθηκε προηγουμένως, τα πειράματα αφορούσαν την εκτίμηση της πόζας για κάθε μοναδική σκηνή του συνόλου δεδομένων, δηλαδή η θέση και ο προσανατολισμός για κάθε εικόνα σε ένα σύνολο εικόνων. Η πόζα περιγράφεται από έναν πίνακα περιστροφής (που αντιπροσωπεύει τον προσανατολισμό) και ένα διάνυμα

μετάθεσης (που αντιπροσωπεύει τη θέση) σε ένα αυθαίρετο σύστημα αναφοράς.

IV. DATASET

Urban

Το σύνολο δεδομένων Urban περιλαμβάνει τρεις διακριτές κατηγορίες: Η καθεμία παρουσιάζει μοναδικές προκλήσεις και χαρακτηριστικά. Urban: Το σύνολο δεδομένων Urban αντικατοπτρίζει το σύνολο δεδομένων IMC2020 PhotoTourism, εστιάζοντας σε φωτογραφίες αστικών κτιρίων. Η πρωταρχική σκηνή, "Kyiv Puppet Theater", περιλαμβάνει 26 εικόνες. Αξίζει να σημειωθεί ότι το σύνολο δεδομένων περιέχει επίσης εικόνες εξαιρετικά συμμετρικών αντικειμένων, όπου οι στάσεις της κάμερας οδηγούν σε μηδενική οπτική επικάλυψη. Η αντιστοίχιση αυτών των εικόνων απαιτεί τη συλλογική τους εξέταση ή τη χρήση πληροφοριών υποβάθρου, κάτι που οι τρέχουσες λύσεις αντιστοίχισης χαρακτηριστικών δεν αντιμετωπίζουν.



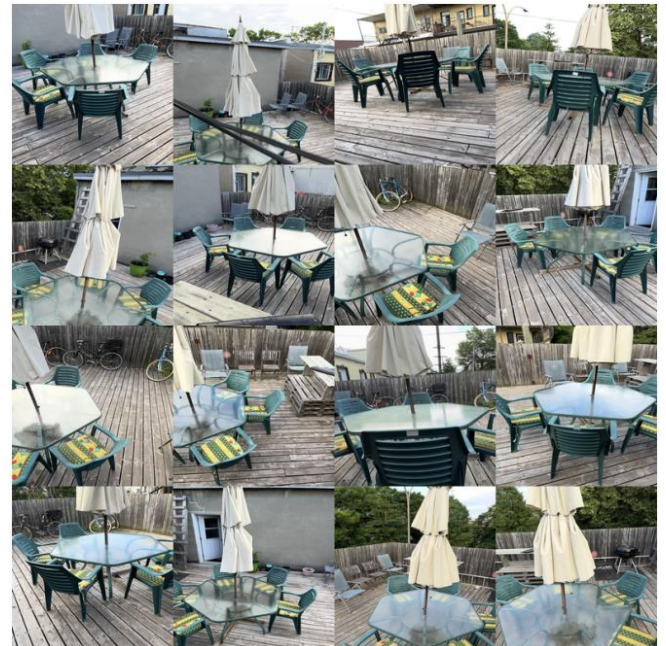
Εικόνα 8: Urban Dataset

Heritage

Το Heritage περιλαμβάνει φωτογραφίες υψηλής ανάλυσης αρχαίων κτιρίων που έχουν ληφθεί τόσο με επίγειες κάμερες DSLR όσο και με UAV. Αυτό το σύνολο δεδομένων αντιμετωπίζει προκλήσεις όπως οι σημαντικές αλλαγές κλίμακας (έως και 20x), η περιστροφή εντός του επιπέδου και το "φαινόμενο μπανάνας", το οποίο παραμορφώνει επίπεδες επιφάνειες λόγω της χαμηλής επικάλυψης μεταξύ των καρέ και των ελαφρών λανθασμένων ευθυγραμμίσεων. Επιπλέον, η επεξεργασία εικόνων υψηλής ανάλυσης θέτει υπολογιστικές προκλήσεις, ιδίως στην προσαρμογή δέσμης. Οι σκηνές Cyprus, Dioscuri και Wall συνεισφέρουν συνολικά 30,174 και 43 εικόνες, αντίστοιχα.



Εικόνα 9: Heritage Dataset



Εικόνα 10: Haipeer Dataset

Haipeer

Το σύνολο δεδομένων Haipeer, που σχεδιάστηκε για εκπαίδευση, υιοθετεί μια παρόμοια προσέγγιση με τα NERFs, χρησιμοποιώντας μια πολυεπίπεδη διάταξη καμερών που περιβάλλουν συγκεκριμένα αντικείμενα. Τα αντικείμενα αυτά διαθέτουν συχνά διακριτικά χαρακτηριστικά, όπως λεπτότητα (π.χ. ποδήλατο) ή έλλειψη υφής (π.χ. άγαλμα). Το δοκιμαστικό τμήμα του συνόλου δεδομένων περιέχει ελάχιστο αριθμό εικόνων, καθιστώντας τις διαφορές οπτικής γωνίας την πρωταρχική πρόκληση. Είναι ενδιαφέρον ότι η χρήση του "Map-free Visual Relocalization" για τις σωληνώσεις εκπαίδευσης θα μπορούσε να είναι επωφελής, αν και αυτή η ευκαιρία παραβλέφθηκε από όλα τα μέρη, συμπεριλαμβανομένων των διοργανωτών. Αυτό το σύνολο δεδομένων περιλαμβάνει τις σκηνές Ποδήλατο, Καρέκλες και Σιντριβάνι, με 15, 16 και 23 εικόνες, αντίστοιχα. Συνοπτικά, κάθε κατηγορία του συνόλου δεδομένων εισάγει μοναδικές προκλήσεις για την επεξεργασία εικόνας και την αντιστοίχιση χαρακτηριστικών. Το σύνολο δεδομένων Urban δίνει έμφαση στις πολύπλοκες στάσεις της κάμερας και στην οπτική επικάλυψη. Η κληρονομιά απαιτεί την αντιμετώπιση ζητημάτων που σχετίζονται με αλλαγές μεγάλης κλίμακας, περιστροφή στο επίπεδο και το "φαινόμενο μπανάνας". Τέλος, το Haipeer εστιάζει σε αντικείμενα με διακριτές ιδιότητες, θέτοντας το δικό του σύνολο προκλήσεων στη μεταβολή της οπτικής γωνίας.

V. ΠΕΙΡΑΜΑΤΑ

A. Μέτρο αξιολόγησης

Μέτρο αξιολόγησης – Mean Average Accuracy (mAA): Η μετρική που χρησιμοποιείται για την αξιολόγηση της ακρίβειας των εκτιμώμενων στάσεων ονομάζεται "Mean Average Accuracy" ή mAA και ακολουθεί τα επόμενα βήματα:

Υπολογισμός του σχετικού σφάλματος: Για κάθε ζεύγος εικόνων στο σύνολο (i και j), υπολογίζεται το σχετικό σφάλμα ως προς την περιστροφή (δr) και τη μεταφορά (δt). Αυτό το σφάλμα είναι η διαφορά μεταξύ της εκτιμώμενης πόζας και της πόζας της βασικής αλήθειας.

Κατώφλι πόζας: Στη συνέχεια, οι εκτιμώμενες πόζες διαχωρίζονται με βάση την ακρίβειά τους όσον αφορά την περιστροφή και τη μεταφορά. Αυτό σημαίνει ότι οι πόζες θεωρούνται ακριβείς εάν εμπίπτουν σε ένα ορισμένο εύρος σφάλματος που ορίζεται από τα κατώφλια. Για τη διαδικασία αυτή χρησιμοποιούνται δέκα ζεύγη κατωφλίων. Για την περιστροφή, τα κατώτατα όρια κυμαίνονται από 1 έως 10 μοίρες και για τη μεταφορά κυμαίνονται από 0,2 έως 5 μέτρα. Αυτά τα κατώφλια επιλέγονται για την αξιολόγηση της ακρίβειας σε διαφορετικά επίπεδα ακρίβειας.

Υπολογισμός του ποσοστού των ακριβών στάσεων: Για κάθε ζεύγος κατωφλίων, υπολογίζεται το ποσοστό των ακριβών ζευγών πόζας. Αυτό σημαίνει το ποσοστό των εκτιμώμενων πόζων που εμπίπτουν στο καθορισμένο εύρος σφάλματος σε σύγκριση με τη πραγματική στάση. Στην συνέχεια, τα αποτελέσματα που προκύπτουν από τα διάφορα επίπεδα κατωφλίου υπολογίζονται κατά μέσο όρο για να προκύψει ένα συνολικό μέτρο ακρίβειας. Αυτό βοηθά στην επιβράβευση των ακριβέστερων εκτιμήσεων. Ολόκληρη αυτή η μετρική υπολογίζεται ξεχωριστά για κάθε σκηνή στο σύνολο δεδομένων. Στη συνέχεια, η mAA για κάθε σκηνή υπολογίζεται κατά μέσο όρο για να ληφθεί μια τιμή mAA

για το συγκεκριμένο σύνολο δεδομένων. Αυτό γίνεται επειδή διαφορετικές σκηνές μπορεί να έχουν διαφορετικά χαρακτηριστικά ή προκλήσεις. Τέλος, οι τιμές mAA που λαμβάνονται για κάθε σύνολο δεδομένων (το οποίο μπορεί να περιέχει πολλές σκηνές) υπολογίζονται κατά μέσο όρο για να προκύψει η τελική μετρική mAA. Αυτή η τελική μετρική δίνει μια συνολική αξιολόγηση της ακρίβειας της εκτίμησης της θέσης της κάμερας για ολόκληρο το σύνολο δεδομένων.

Επομένως για τα πειράματα χρησιμοποιήθηκαν οι παρακάτω ανιχνευτές και περιγραφείς:

Ανιχνευτές

- KeyNet
- Difference-of-Gaussian(DoG)
- CornerGFTT

Περιγραφείς

- HardNet
- SosNet
- SIFT

End-to-end

Επιπρόσθετα προστέθηκε στα πειράματα και ο αλγόριθμος DISK ο οποίος είναι end-to-end για την εύρεση αντιστοιχιών μεταξύ εικόνων. Για όλους τους αλγορίθμους που χρησιμοποιήθηκαν έγιναν πειράματα με κάθε διαθέσιμο συνδυασμό και πειραματιστήκαμε επίσης με τον αριθμό των χαρακτηριστικών που εντοπίστηκαν. Συγκεκριμένα για κάθε συνδυασμό πειραματιστήκαμε με 2000 και 8000 χαρακτηριστικά.

Σύντομη περιγραφή της διαδικασίας

Καθε σκηνή περιέχει έναν αριθμό από διάφορες εικόνες διάσπαρτες. Επειδή αυτός ο αριθμός σε αρκετές περιπτώσεις μπορεί να είναι μεγάλος, με αποτέλεσμα να είναι υπολογιστικά δαπανηρό να αντιστοιχήσουμε τόσες πολλές εικόνες, χρησιμοποιήθηκε ένας μηχανισμός φιλτραρίσματος που βασίζεται στα Συνελκτικά Νευρωνικά Δίκτυα. Πιο συγκεκριμένα, για κάθε εικόνα εξάγαμε υψηλού επιπέδου περιγραφείς από αυτήν με ένα προεκπαιδευμένο στην βάση δεδομένων imagenet δίκτυο, το EfficientNet-B7. Αυτό μας επέτρεψε να φτιάξουμε μια λίστα κατάταξης για καθε εικόνα στην οποία θα τοποθετούνται με σειρά ομοιότητας οι υπόλοιπες εικόνες της σκηνής. Έτσι, αν κάποια σκηνή έχει για παράδειγμα 200 εικόνες και θέλουμε για λόγους ταχύτητας και υπολογισμών και να μην χρειαστεί να τις συγκρίνουμε όλες μεταξύ τους, μπορούμε να επιλέξουμε τις κοντινότερες με την χρήση μιας μετρικής απόστασης. Με αυτόν τον τρόπο αποφεύγουμε την αντιστοιχίση όλων των εικόνων μεταξύ τους και επιλέγουμε τις κοντινότερες με βάση το περιεχόμενο τους. Στην διπλωματική αυτή, επιλέξαμε να χρησιμοποιήσουμε τις 60 κοντινότερες για κάθε εικόνα. Στην ουσία η μόνη σκηνή που επηρεάζεται είναι η dioscuri, καθώς όλες οι άλλες έχουν λιγότερες από 60 εικόνες.

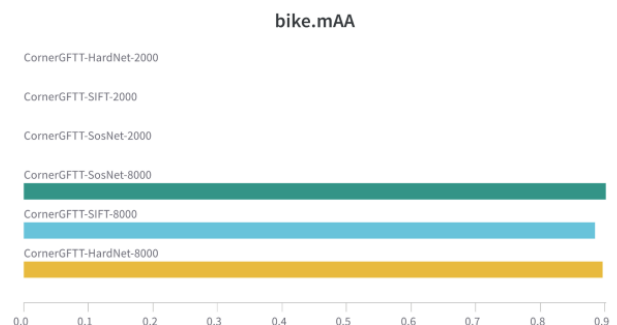
Εφόσον γίνει η διαλογή με την βοήθεια του προεκπαιδευμένου δικτύου, στην συνέχεια πραγματοποιείται η αντιστοίχιση όλων των εικόνων με τις 60 πιο όμοιες τους από την σκηνή. Εδώ αναλαμβάνουν δράση οι ανιχνευτές και οι περιγραφείς για την αντιστοίχιση των εικόνων. Οι

ανιχνευτές βρίσκουν σημεία ενδιαφέροντος για την κάθε εικόνα, ενώ οι περιγραφείς τα περιγράφουν με ένα διάνυσμα διάστασης 128. Το επόμενο βήμα είναι να βρούμε κοινά σημεία ενδιαφέροντος μεταξύ των εικόνων ώστε να μπορεί αργότερα να προσδιοριστεί η σχετική θέση κάθε εικόνας με το σημείο αναφοράς. Αν και συνήθως χρησιμοποιείται κάποιο είδος thresholding για την εύρεση στιβαρών αντιστοιχίσεων, υιοθετήθηκε ο αλγόριθμος adalam ο οποίος είναι αρκετά αποτελεσματικός και διαθέτει γεωμετρική επιβεβαίωση για τις αντιστοιχίες. Αυτό αποτελεί ένα πολύ σημαντικό βήμα για το σύστημα, τόσο γιατί είναι απαραίτητη η εύρεση σημαντικών σημείων ενδιαφέροντος όσο και το φιλτράρισμα των αδύναμων αντιστοιχίσεων ώστε να μην επηρεάσουν την μέθοδο της δέσμης που πραγματοποιείται από το colmap.

Όταν τελειώσει η διαδικασία της αντιστοίχισης και του φιλτραρίσματος για κάθε ζευγάρι εικόνων (από τα 60 κοντινότερα), αποθηκεύονται όλες οι πληροφορίες στην βάση δεδομένων που έχει το colmap. Το colmap έπειτα αναλαμβάνει να ξεκαθαρίσει ακόμα περαιτέρω τις πιθανές αντιστοιχίσεις με γεωμετρική επιβεβαίωση του αλγορίθμο LO-Ransac που αποτελεί μια εκδοχή που παρέχει θεωρητικά μεγαλύτερη ακρίβεια με την προσθήκη ενός επιπλέον βήματος βελτιστοποίησης. Στο επόμενο βήμα γίνεται η χρήση του IncrementalMapper του Colmap, το οποίο αποτελεί μια μέθοδο επαυξητικής καταχώρισης εικόνων και την εκτίμηση της πόζας τους. Συνοπτικά, η μέθοδος ξεκινάει από ένα σημείο αναφοράς, δηλαδή ένα ζευγάρι εικόνων που έχει αντιστοιχιστεί με μεγάλη βεβαιότητα και κτίζει σταδιακά πάνω σε αυτό καταχωρώντας τις υπόλοιπες εικόνες της σκηνής σε σχέση με το σημείο αναφοράς. Μολις ολοκληρωθεί η διαδικασία της καταχώρισης, γίνεται η χρήση της μετρικής αξιολόγησης για κάθε σκηνή ξεχωριστά όπως περιγράφηκε προηγουμένως.

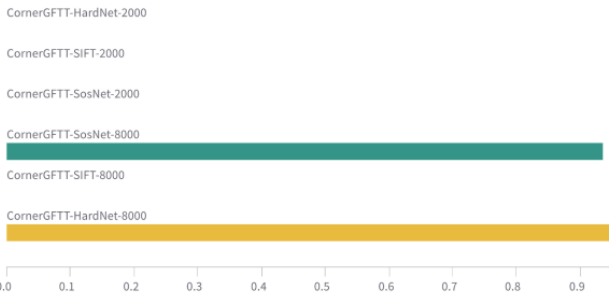
B. Διαγράμματα & αποτελέσματα

CornerGFTT



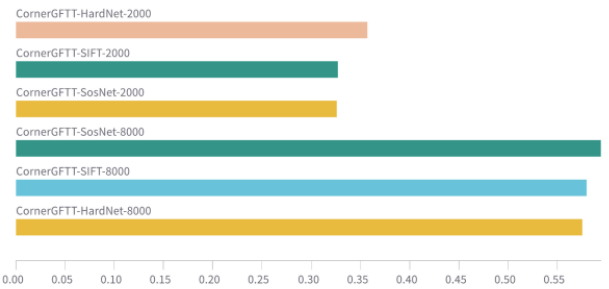
Διάγραμμα 1: Συγκριτική απόδοση CGFTT στη σκηνή bike με όλους τους περιγραφείς για 2000 και 8000 features.

chairs.mAA



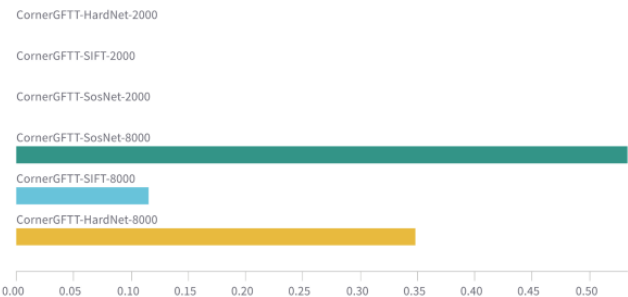
Διάγραμμα 2: Συγκριτική απόδοση CGFTT στη σκηνή chairs με όλους τους περιγραφείς για 2000 και 8000 features.

cyprus.mAA



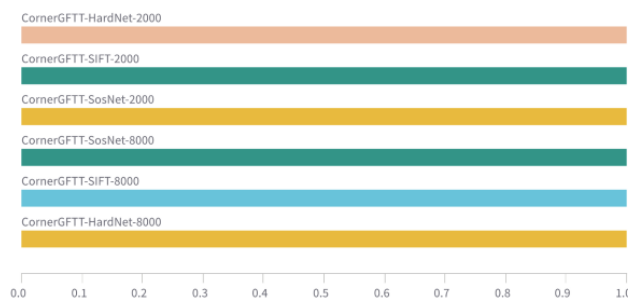
Διάγραμμα 3: Συγκριτική απόδοση CGFTT στη σκηνή cyprus με όλους τους περιγραφείς για 2000 και 8000 features.

dioscuri.mAA



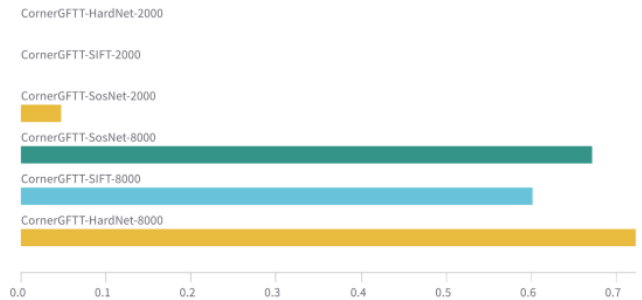
Διάγραμμα 4: Συγκριτική απόδοση CGFTT στη σκηνή dioscuri με όλους τους περιγραφείς για 2000 και 8000 features.

fountain.mAA



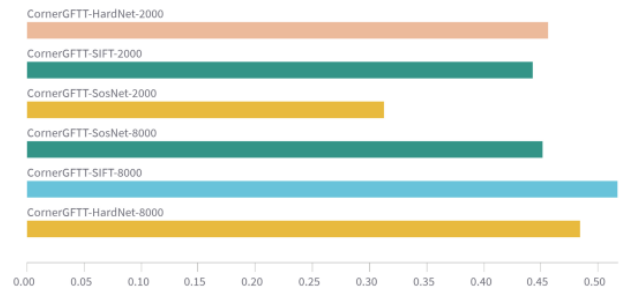
Διάγραμμα 5: Συγκριτική απόδοση CGFTT στη σκηνή fountain με όλους τους περιγραφείς για 2000 και 8000 features.

kyiv-puppet-theater.mAA



Διάγραμμα 6: Συγκριτική απόδοση CGFTT στη σκηνή Kyiv puppet theater με όλους τους περιγραφείς για 2000 και 8000 features.

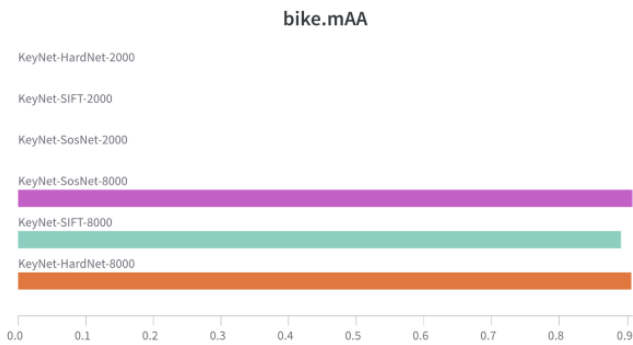
wall.mAA



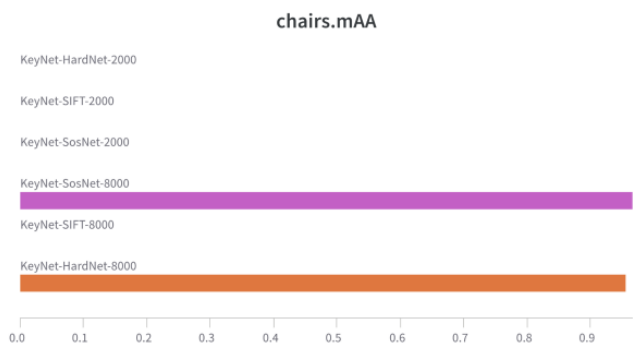
Διάγραμμα 7: Συγκριτική απόδοση CGFTT στη σκηνή wall με όλους τους περιγραφείς για 2000 και 8000 features.

Για αρχή, είναι ξεκάθαρη η υπεροχή των 8000 keypoints έναντι των 2000. Όπως προκύπτει από τα αποτελέσματα, είναι φανερό πως τα 2000 keypoints δεν είναι αρκετά για να χαρακτηρίσουν αποτελεσματικά τις σκηνές ώστε να βρεθούν ισχυρές αντιστοιχίσεις. Ιδιαίτερα σε συγκεκριμένες σκηνές ο αλγόριθμος του colmap αδυνατεί να βρει μέχρι και ένα ζεύγος εικόνων ώστε να εκτιμήσει την στάση τους. Από εκεί και πέρα, με την άυξηση των keypoints παρατηρείται σημαντική βελτίωση σε όλες τις σκηνές, με αξιόλογα αποτελέσματα. Ο συνδυασμός του CornerGFTT με το HardNet και το SosNet πετυχαίνει τα καλύτερα αποτελέσματα. Οσον αφορά τις σκηνές, είναι αναμενόμενα τα αποτελέσματα σε μερικές από αυτές. Για παράδειγμα, η σκηνή dioscuri περιέχει πάρα πολλές εικόνες (174) με αποτέλεσμα να είναι πολύ δύσκολο να καταχωρηθούν όλες οι εικόνες. Η μετρική mAA επηρεάζεται αρνητικά σε μεγάλο βαθμό με τη μη καταχώρηση εικόνων. Άλλη μια σκηνή που δυσκολεύει τους αλγόριθμους είναι η Cyprus. Η σκηνή αυτή παρουσιάζει εικόνες που έχουν μεγάλη ομοιότητα μεταξύ τους αν και μπορεί να βρίσκονται σε διαφορετικά σημεία και επιπλέον, πολλές από αυτές έχουν υποστεί περιστροφή δυσχεραίνοντας περαιτέρω την διαδικασία της αντιστοίχισης.

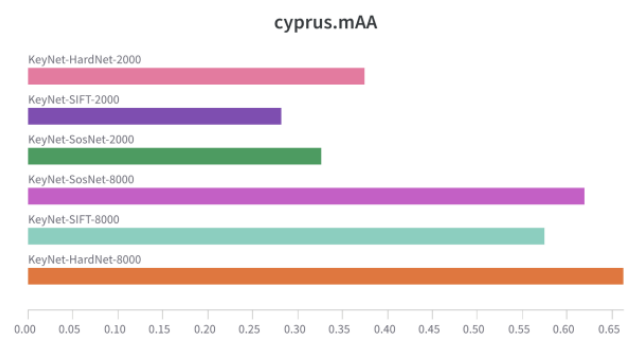
KeyNet



Διάγραμμα 8: Συγκριτική απόδοση KeyNet στη σκηνή bike με όλους τους περιγραφείς για 2000 και 8000 features.

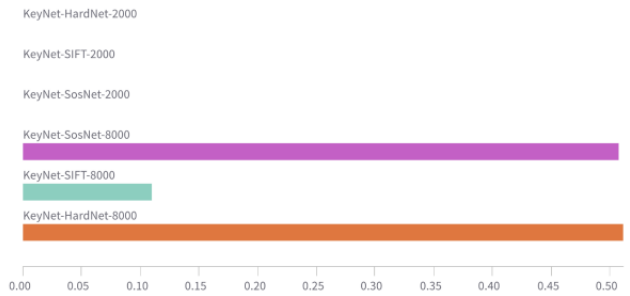


Διάγραμμα 9: Συγκριτική απόδοση KeyNet στη σκηνή chairs με όλους τους περιγραφείς για 2000 και 8000 features.



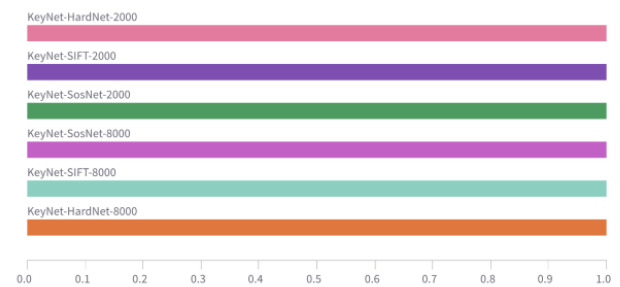
Διάγραμμα 10: Συγκριτική απόδοση KeyNet στη σκηνή cyprus με όλους τους περιγραφείς για 2000 και 8000 features.

dioscuri.mAA



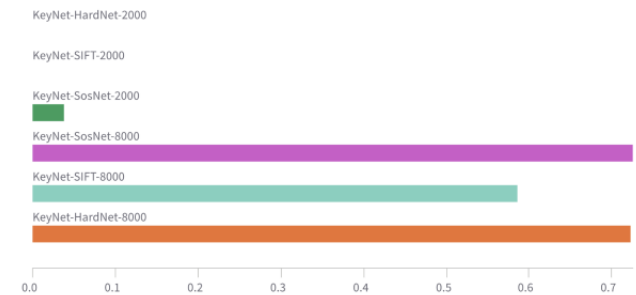
Διάγραμμα 11: Συγκριτική απόδοση KeyNet στη σκηνή dioscuri με όλους τους περιγραφείς για 2000 και 8000 features.

fountain.mAA

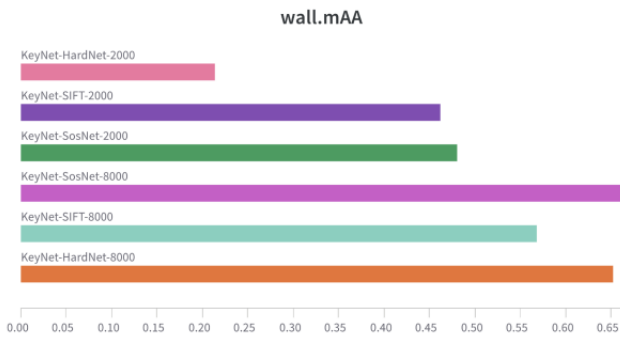


Διάγραμμα 12: Συγκριτική απόδοση KeyNet στη σκηνή fountain με όλους τους περιγραφείς για 2000 και 8000 features.

kyiv-puppet-theater.mAA



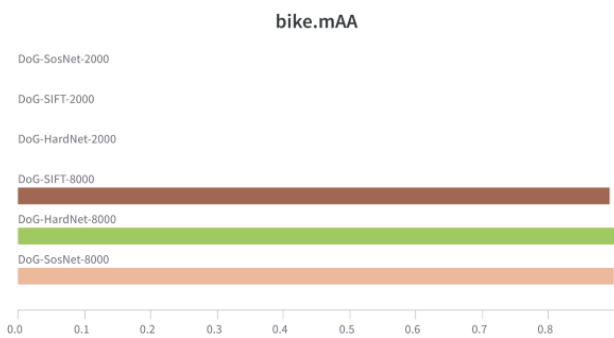
Διάγραμμα 13: Συγκριτική απόδοση KeyNet στη σκηνή Kyiv puppet theater με όλους τους περιγραφείς για 2000 και 8000 features.



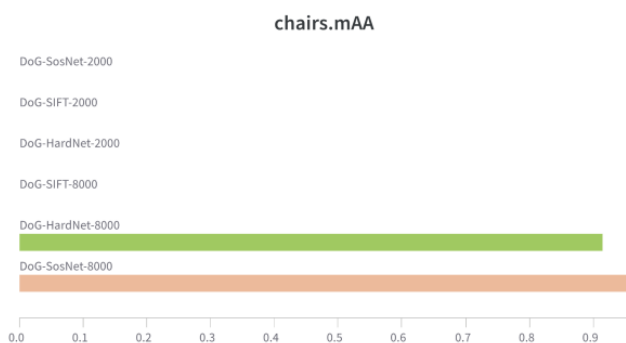
Διάγραμμα 14: Συγκριτική απόδοση KeyNet στη σκηνή wall με όλους τους περιγραφείς για 2000 και 8000 features.

Και στην περίπτωση του KeyNet παρατηρούμε σημαντική αδυναμία των αλγορίθμων χρησιμοποιώντας μόνο 2000 keypoints. Με την αύξηση σε 8000, η μετρική βελτιώνεται σε όλες τις σκηνές. Γενικά η συμπεριφορά του KeyNet είναι παρόμοια με του CornerGFTT ανα σκηνή. Οι σκηνές της κατηγορίας Heritage (Cyprus, Dioscuri και Wall) παρουσιάζουν μεγαλύτερη δυσκολία απο τις υπόλοιπες για τους προαναφερόμενους λόγους και στο KeyNet. Επιπρόσθετα, οι συνδυασμοί KeyNet-SosNet και KeyNet-HardNet δείχνουν να αποδίδουν σταθερά καλύτερα σε σχέση με το KeyNet-SIFT.

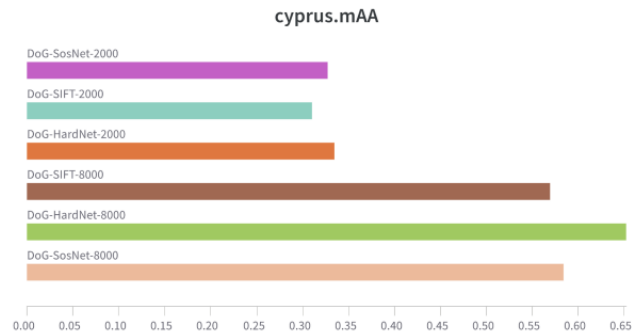
DOG



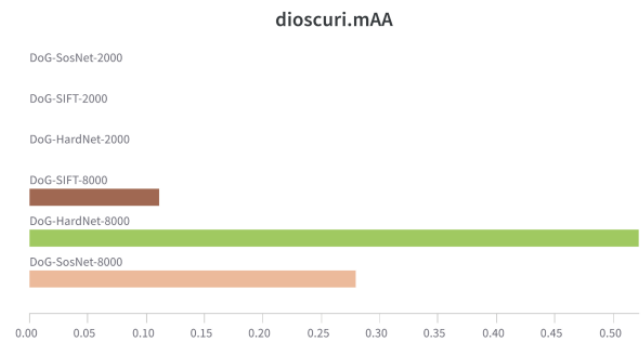
Διάγραμμα 15: Συγκριτική απόδοση DoG στη σκηνή bike με όλους τους περιγραφείς για 2000 και 8000 features.



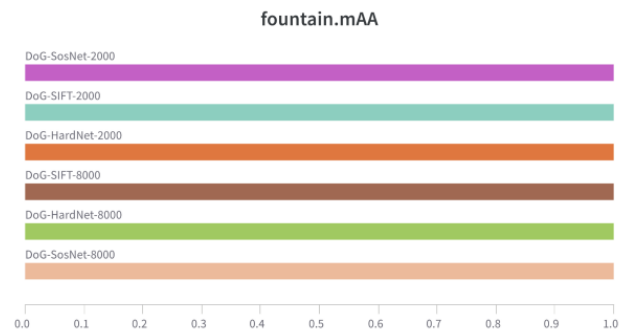
Διάγραμμα 16: Συγκριτική απόδοση DoG στη σκηνή chairs με όλους τους περιγραφείς για 2000 και 8000 features.



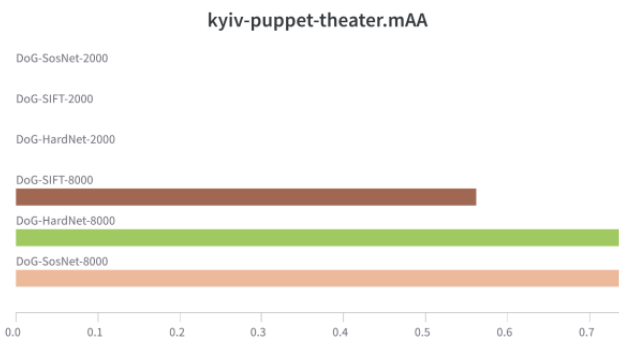
Διάγραμμα 17: Συγκριτική απόδοση DoG στη σκηνή cyprus με όλους τους περιγραφείς για 2000 και 8000 features.



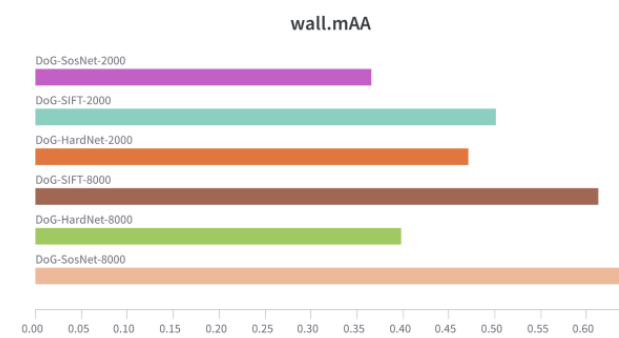
Διάγραμμα 18: Συγκριτική απόδοση DoG στη σκηνή dioscuri με όλους τους περιγραφείς για 2000 και 8000 features.



Διάγραμμα 19: Συγκριτική απόδοση DoG στη σκηνή fountain με όλους τους περιγραφείς για 2000 και 8000 features.



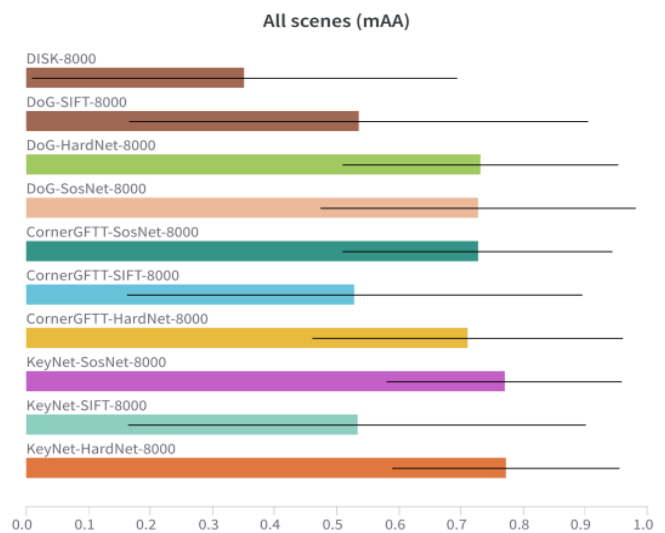
Διάγραμμα 20: Συγκριτική απόδοση DoG στη σκηνή Kyiv puppet theater με όλους τους περιγραφείς για 2000 και 8000 features.



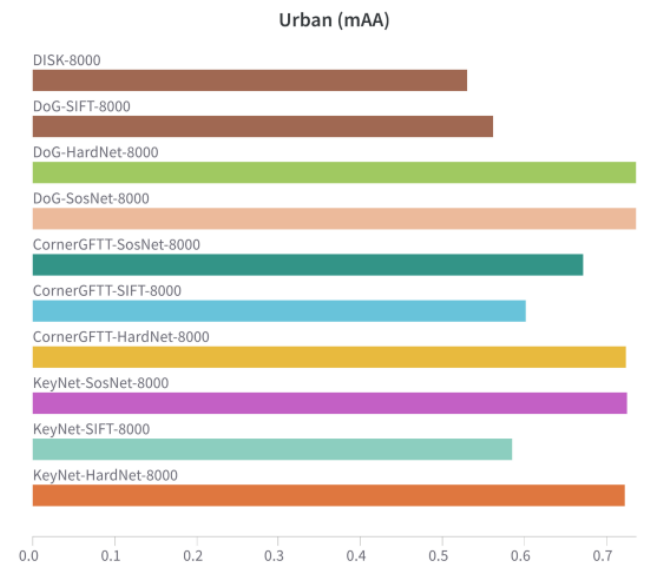
Διάγραμμα 21: Συγκριτική απόδοση DoG στη σκηνή wall με όλους τους περιγραφείς για 2000 και 8000 features.

Τέλος, για άλλη μια φορά τα συμπεράσματα επαληθεύονται και στην περίπτωση του DoG και παρατηρείται παρόμοια ικανότητα στην εξαγωγή αποτελεσματικών σημείων ενδιαφέροντος σε σχέση με τους ανιχνευτές KeyNet και CornerGFTT. Το SIFT ούτε εδώ καταφέρνει να ξεπεράσει τους εξαγωγείς deep learning.

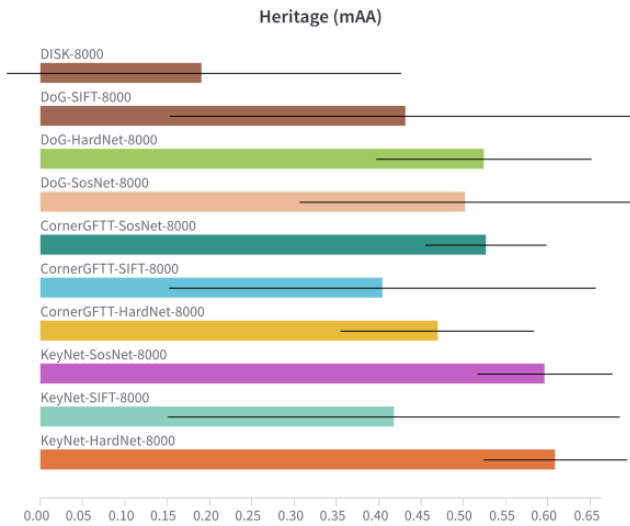
8000 χαρακτηριστικά με όλους τους συνδυασμούς



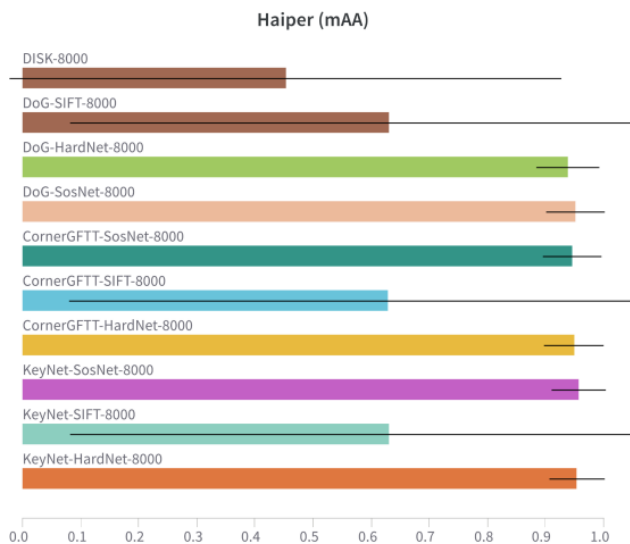
Διάγραμμα 22: Συγκριτική απόδοση όλων των συνδυασμών σε όλες τις σκηνές παρουσιάζοντας το μέσο όρο (χρωματιστές μπάρες) και την τυπική απόκλιση (μαύρη λεπτή γραμμή) με 8000 features.



Διάγραμμα 23: Συγκριτική απόδοση όλων των συνδυασμών στο σύνολο εικόνων Urban με 8000 features.



Διάγραμμα 24: Συγκριτική απόδοση όλων των συνδυασμών στο σύνολο εικόνων Heritage παρουσιάζοντας το μέσο όρο (χρωματιστές μπάρες) και την τυπική απόκλιση (μαύρη λεπτή γραμμή) με 8000 features.



Διάγραμμα 25: Συγκριτική απόδοση όλων των συνδυασμών στο σύνολο εικόνων Haiper παρουσιάζοντας το μέσο όρο (χρωματιστές μπάρες) και την τυπική απόκλιση (μαύρη λεπτή γραμμή) με 8000 features.

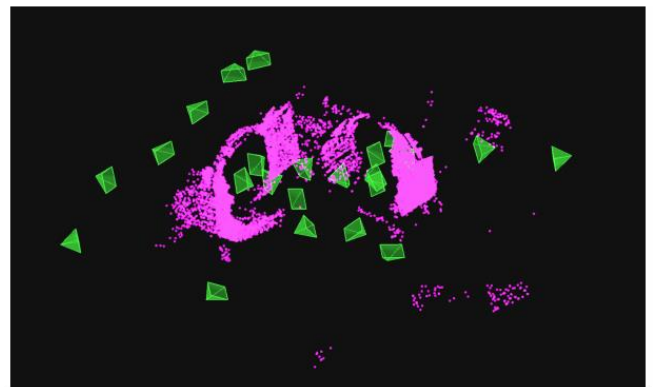
Για να δούμε τα αποτελέσματα πιο αναλυτικά και να είναι εφικτό να εξάγουμε κάποια γενικευμένα συμπεράσματα, επικεντρωθήκαμε στην περίπτωση των 8000 keypoints ώστε κάθε ανιχνευτής να έχει την ευχέρεια να εξαγάγει έναν μεγάλο αριθμό keypoints για να επιδείξει την ικανότητά του. Τα σχήματα περιέχουν την απόδοση των πειραμάτων για κάθε κατηγορία ξεχωριστά αλλά και συνολικά, με τον μέσο όρο και την τυπική απόκλιση να παρουσιάζεται. Όπως συζητήθηκε παραπάνω, όλοι οι συνδυασμοί με τους εξαγωγείς HardNet και SosNet υπερτερούν του SIFT σε όλες τις κατηγορίες (Urban, Heritage, Haiper). Απο εκεί και πέρα, στο Διάγραμμα 22 που

εμφανίζεται ο μέσος όρος των πειραμάτων σε όλες τις σκληρές μπορούμε να εκφέρουμε συμπεράσματα με μεγαλύτερη σιγουριά. Έτσι λοιπόν, ο ανιχνευτής KeyNet αποδίδει καλύτερα από τους υπόλοιπους με τους περιγραφείς HardNet και SosNet με μεγάλη σταθερότητα. Αυτό φαίνεται και από τα τελικά νούμερα στην μετρική mAA αλλά και από την τυπική απόκλιση. Ακολουθεί από κοντά ο ανιχνευτής DoG και πιο πίσω ο CornerGFTT. Όσον αφορά τους περιγραφείς, ο HardNet και ο SosNet είναι αρκετά κοντά με τον HardNet να κερδίζει στις λεπτομέρειες. Τέλος, δοκιμάστηκε ένας state-of-the-art αλγόριθμος στα πειράματα, ο DISK. Παρόλο που πρόκειται για μια αρκετά ισχυρή προσέγγιση, υστερεί σημαντικά σε σχέση με τους υπόλοιπους συνδυασμούς.

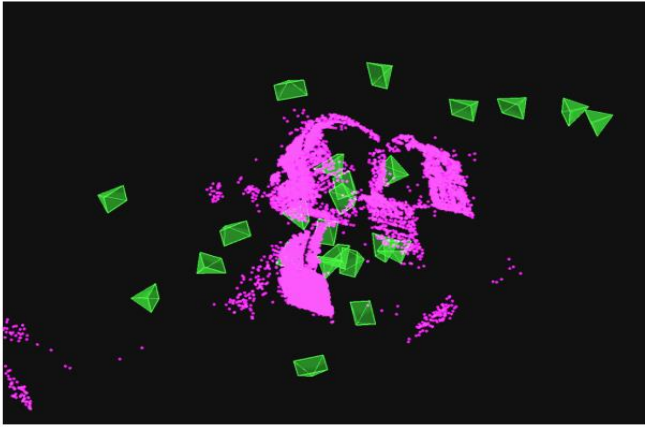
C. Συγκεντρωτικός πίνακας αποτελεσμάτων

	Urban	Heritage			Haiper			ALL
	Kyiv	Cyprus	Dioscuri	Wall	Bike	Chairs	Fountain	
GFTT-Hard	0.723	0.5754	0.348	0.456	0.896	0.95	1	0.711
GFTT-Sos	0.671	0.594	0.533	0.451	0.901	0.935	1	0.726
GFTT-SIFT	0.601	0.580	0.114	0.517	0.883	0.0	1	0.528
KEY-Hard	0.722	0.662	0.51	0.652	0.905	0.955	1	0.772
Key-Sos	0.725	0.619	0.507	0.661	0.907	0.965	1	0.769
Key-SIFT	0.585	0.575	0.109	0.568	0.890	0	1	0.532
DoG-Hard	0.736	0.652	0.521	0.398	0.9	0.914	1	0.731
DoG-Sos	0.736	0.583	0.279	0.642	0.899	0.955	1	0.728
DoG-SIFT	0.561	0.569	0.111	0.398	0.892	0.0	1	0.535
DISK	0.529	0.116	0.0	0.454	0.22	0.138	1	0.351

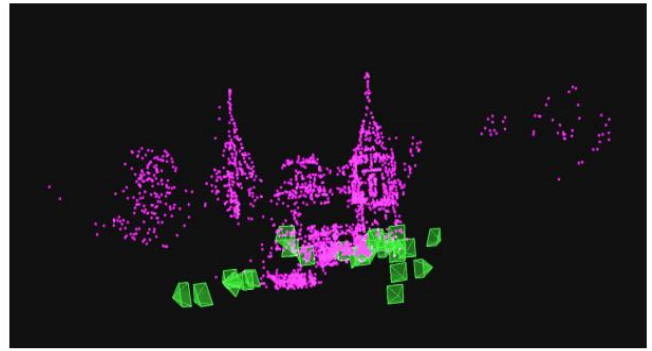
D. Visuals



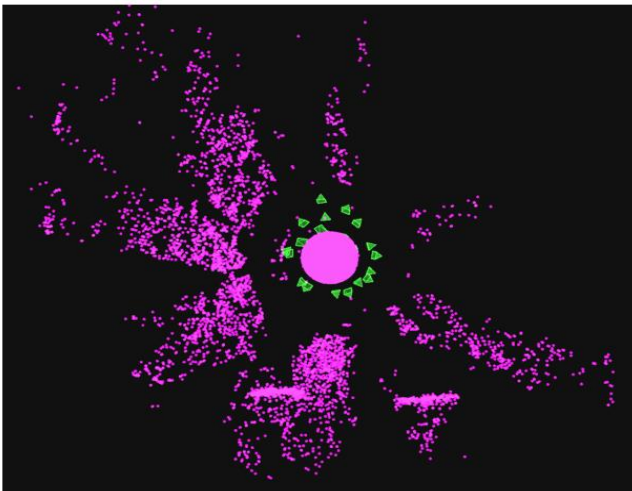
Εικόνα 11: Σκηνή Cyprus mAA 0.662



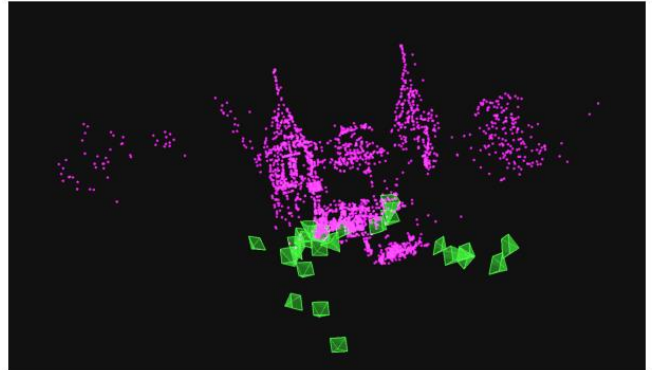
Εικόνα 12: Σκηνή Cyprus mAA 0.662



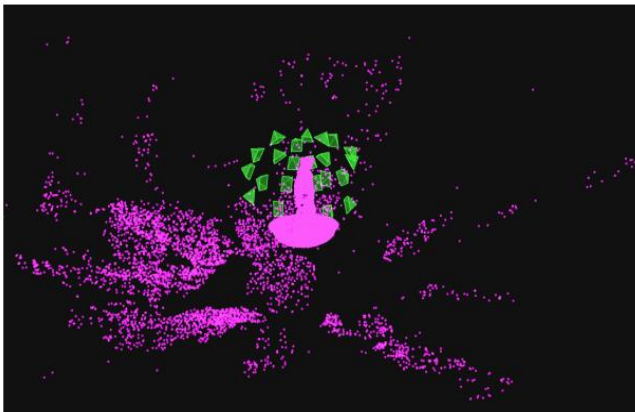
Εικόνα 15: Σκηνή Kyiv puppet theater mAA 0.772



Εικόνα 13: Σκηνή Fountain mAA 1.0



Εικόνα 16: Σκηνή Kyiv puppet theater mAA 0.772



Εικόνα 14: Σκηνή Fountain mAA 1.0

REFERENCES

- [1] Schonberger, Johannes L., and Jan-Michael Frahm. "Structure-from-motion revisited." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.
- [2] Bauer, Zuria, et al. "NVS-MonoDepth: Improving monocular depth prediction with novel view synthesis." *2021 International Conference on 3D Vision (3DV)*. IEEE, 2021.
- [3] Guizilini, Vitor, et al. "Full surround monodepth from multiple cameras." *IEEE Robotics and Automation Letters* 7.2 (2022): 5397-5404.
- [4] Yao, Yao, et al. "Mvsnet: Depth inference for unstructured multi-view stereo." *Proceedings of the European conference on computer vision (ECCV)*. 2018.
- [5] Chang, Angel X., et al. "Shapenet: An information-rich 3d model repository." *arXiv preprint arXiv:1512.03012* (2015).
- [6] Wang, Nanyang, et al. "Pixel2mesh: Generating 3d mesh models from single rgb images." *Proceedings of the European conference on computer vision (ECCV)*. 2018.
- [7] Sitzmann, Vincent, et al. "Deepvoxels: Learning persistent 3d feature

- embeddings." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2019.
- [8] Wang, Zirui, et al. "NeRF--: Neural radiance fields without known camera parameters." *arXiv preprint arXiv:2102.07064* (2021).
- [9] Ma, Zhiliang, and Shilong Liu. "A review of 3D reconstruction techniques in civil engineering and their applications." *Advanced Engineering Informatics* 37 (2018): 163-174.
- [10] Jiang, Jinwei, and Alper Yilmaz. "Good features to track: A view geometric approach." *2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops)*. IEEE, 2011.
- [11] Barroso-Laguna, Axel, et al. "Key. net: Keypoint detection by handcrafted and learned cnn filters." *Proceedings of the IEEE/CVF international conference on computer vision*. 2019.
- [12] Daixian, Zhu. "SIFT algorithm analysis and optimization." *2010 International Conference on Image Analysis and Signal Processing*. IEEE, 2010.
- [13] Pultar, Milan. "Improving the HardNet descriptor." *arXiv preprint arXiv:2007.09699* (2020).
- [14] Tian, Yurun, et al. "Sosnet: Second order similarity regularization for local descriptor learning." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2019.
- [15] Marvaniya, Smit, Raj Gupta, and Anurag Mittal. "Adaptive locally affine-invariant shape matching." *Machine Vision and Applications* 29 (2018): 553-572.
- [16] Kroeger, Till, et al. "Fast optical flow using dense inverse search." *Computer Vision—ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part IV* 14. Springer International Publishing, 2016.
- [17] Raguram, Rahul, Jan-Michael Frahm, and Marc Pollefeys. "A comparative analysis of RANSAC techniques leading to adaptive real-time random sample consensus." *Computer Vision—ECCV 2008: 10th European Conference on Computer Vision, Marseille, France, October 12-18, 2008, Proceedings, Part II* 10. Springer Berlin Heidelberg, 2008.
- [18] Chum, Ondřej, Jiří Matas, and Josef Kittler. "Locally optimized RANSAC." *Pattern Recognition: 25th DAGM Symposium, Magdeburg, Germany, September 10-12, 2003. Proceedings* 25. Springer Berlin Heidelberg, 2003.
- [19] Pleansamai, Korawit, and Krisada Chaiyasarn. "M-estimator sample consensus planar extraction from image-based 3D point cloud for building information modelling." *GEOMATE Journal* 17.63 (2019): 69-76.
- [20] Chum, Ondrej, and Jiri Matas. "Matching with PROSAC-progressive sample consensus." *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)*. Vol. 1. IEEE, 2005.
- [21] Rusiecki, Andrzej. "Robust learning algorithm based on iterative least median of squares." *Neural processing letters* 36 (2012): 145-160.
- [22] Chum, Ondřej, and Jiří Matas. "Optimal randomized RANSAC." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 30.8 (2008): 1472-1482.
- [23] Brachmann, Eric, and Carsten Rother. "Neural-guided RANSAC: Learning where to sample model hypotheses." *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2019.
- [24] Schonberger, Johannes L., and Jan-Michael Frahm. "Structure-from-motion revisited." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.
- [25] Stathopoulou, Elisavet Konstantina, M. Welpone, and Fabio Remondino. "Open-source image-based 3D reconstruction pipelines: Review, comparison and evaluation." *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XLII-2/W17* (2019): 331-338.

