



ΠΑΝΕΠΙΣΤΗΜΙΟ
ΑΘΗΝΩΝ

Σελίδα τίτλου

ΣΧΟΛΗ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΤΗΛΕΠΙΚΟΙΝΩΝΙΩΝ

ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΤΗΛΕΠΙΚΟΙΝΩΝΙΩΝ

ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

**ΑΛΓΟΡΙΘΜΟΙ ΤΡΙΣΔΙΑΣΤΑΤΗΣ ΑΝΑΚΑΤΑΣΚΕΥΗΣ ΜΝΗΜΕΙΩΝ
ΠΟΛΙΤΙΣΤΙΚΗΣ ΚΛΗΡΟΝΟΜΙΑΣ**

Ιωάννης Μπίλλιος

Επιβλέπων: Πέτρος Καρβέλης

Επίκουρος Καθηγητής

Τρίκαλα, Μάρτιος, 2025

3D RECONSTRUCTION ALGORITHMS FOR CULTURAL HERITAGE MONUMENTS

Εγκρίθηκε από τριμελή εξεταστική επιτροπή

Άρτα, 29/07/2025

ΕΠΙΤΡΟΠΗ ΑΞΙΟΛΟΓΗΣΗΣ

1. Επιβλέπων καθηγητής

Πέτρος Καρβέλης,

2. Μέλος επιτροπής

Γιαννακέας Νικόλαος,

3. Μέλος επιτροπής

Τζάλλας Αλέξανδρος,

© Μπίλλιος, Ιωάννης, 2025.

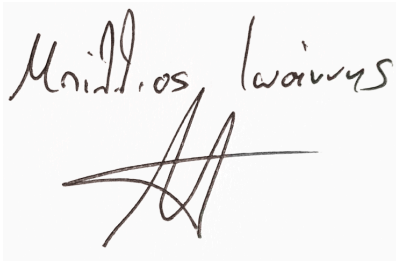
Με επιφύλαξη παντός δικαιώματος. All rights reserved.

Δήλωση μη λογοκλοπής

Δηλώνω υπεύθυνα και γνωρίζοντας τις κυρώσεις του Ν. 2121/1993 περί Πνευματικής Ιδιοκτησίας, ότι η παρούσα μεταπτυχιακή εργασία είναι εξ ολοκλήρου αποτέλεσμα δικής μου ερευνητικής εργασίας, δεν αποτελεί προϊόν αντιγραφής ούτε προέρχεται από ανάθεση σε τρίτους. Όλες οι πηγές που χρησιμοποιήθηκαν (κάθε είδους, μορφής και προέλευσης) για τη συγγραφή της περιλαμβάνονται στη βιβλιογραφία.

Μπίλλιος Ιωάννης

Υπογραφή

The image shows a handwritten signature in dark ink on a light gray background. The signature is written in a cursive style, with the first part appearing to be 'Μπίλλιος' and the second part 'Ιωάννης'. Below the text, there is a stylized, fluid signature mark consisting of several loops and a long horizontal stroke.

ΕΥΧΑΡΙΣΤΙΕΣ

Θα ήθελα να ευχαριστήσω θερμά την οικογένειά μου που με στήριξε όλα αυτά τα χρόνια και για την ατελείωτη υπομονή τους. Επίσης, ευχαριστώ τον επιβλέπων καθηγητή Πέτρο Καρβέλη για το ενδιαφέρον και την καθοδήγηση του. Τέλος, ευχαριστώ την εφετεία αρχαιοτήτων νομού Τρικάλων για την συνεργασία τους όσον αφορά την αδειοδότηση και πρόσβαση που μου παρείχαν στους χώρους και τα μνημεία που βρίσκονται υπό την αρμοδιότητά τους.

ΠΕΡΙΛΗΨΗ

Η παρούσα εργασία έχει ως στόχο την εις βάθος μελέτη και πρακτική αξιολόγηση τεχνικών και αλγορίθμων τρισδιάστατης ανακατασκευής, με έμφαση στη χρήση τους για την τεκμηρίωση και διατήρηση μνημείων πολιτιστικής κληρονομιάς. Η ανάγκη για ψηφιακή διατήρηση πολιτισμικών αγαθών έχει ενταθεί τις τελευταίες δεκαετίες, καθώς η τρισδιάστατη αναπαράσταση αντικειμένων, κτιρίων και χώρων επιτρέπει τη δημιουργία ακριβών ψηφιακών αποτυπώσεων, οι οποίες μπορούν να αξιοποιηθούν σε περιβάλλοντα εικονικής περιήγησης, ψυχαγωγίας, εκπαίδευσης και επιστημονικής μελέτης, καθώς και σε διαδικασίες αποκατάστασης.

Η εργασία ξεκινά με την παρουσίαση του θεωρητικού υπόβαθρου που διέπει την τρισδιάστατη ανακατασκευή, συμπεριλαμβανομένων βασικών εννοιών, ορολογίας, και την κατηγοριοποίηση των μεθόδων που χρησιμοποιούνται. Στη συνέχεια, εξετάζονται συγκεκριμένοι αλγόριθμοι όπως οι Structure from Motion (SfM), Multi View Stereo (MVS), Neural Radiance Fields (NeRF) και Gaussian Splatting, οι οποίοι αντιπροσωπεύουν διαφορετικές προσεγγίσεις και τεχνολογικές εξελίξεις στον χώρο.

Ακολουθεί η εφαρμογή των αλγορίθμων αυτών σε επιλεγμένα σύνολα δεδομένων (datasets) που αφορούν πραγματικά μνημεία, με χρήση υπαρχόντων εργαλείων λογισμικού όπως COLMAP, MicMac, Instant-NGP και άλλα. Στην πορεία, καταγράφονται και αναλύονται τα αποτελέσματα κάθε ανακατασκευής με βάση τόσο ποσοτικά κριτήρια (πυκνότητα, χρόνος εκτέλεσης, ακρίβεια ως προς ground truth όπου είναι διαθέσιμος), όσο και ποιοτικά (οπτική πληρότητα, ανακατασκευή μικρών λεπτομέρειων, κάλυψη).

Σκοπός της εργασίας είναι όχι μόνο η αξιολόγηση των επιδόσεων κάθε τεχνικής σε πρακτικά σενάρια, αλλά και η ανάδειξη της χρησιμότητάς τους ως εργαλεία για την προστασία και μελέτη της πολιτιστικής κληρονομιάς.

Λέξεις-κλειδιά: Τρισδιάστατη ανακατασκευή, αλγόριθμοι, μνημεία πολιτιστικής κληρονομιάς.

ABSTRACT

This work aims at the in-depth study and practical evaluation of three-dimensional reconstruction techniques and algorithms, with emphasis on their use for the archival and maintenance of cultural heritage monuments. The need for digital preservation of cultural points of interest has been amplified in the recent decades, since three-dimensional reconstruction of objects, buildings and spaces allows for the creation of exact digital replicas, which can be utilised in virtual environments for guided tours, entertainment, training and scientific research, as well as preservation efforts.

The essay starts with the presentation of the theoretical background that accompanies three-dimensional reconstruction, including basic concepts, terminology and the classification of the methods that are used. Afterwards, specific algorithms like Structure from Motion (SfM), Multi View Stereo (MVS), Neural Radiance Fields (NeRF) and Gaussian Splatting are examined, which represent different approaches and technological advancements in the field.

Following is the application of these algorithms in specific datasets of real life monuments, using existing software such as COLMAP, MicMac, Instant-NGP and others. The results of each reconstructions are recorded and analysed, focusing on quantitative criteria (density, runtime, accuracy compared to ground truth when applicable) as well as qualitative criteria (visual faithfulness, replication of fine detail, coverage).

The goal of the work is to not only evaluate the performance of each technique on practical scenarios but also the demonstration of their usefulness for the protection and study of cultural heritage.

Keywords: Three-dimensional reconstruction, algorithms, cultural heritage monuments.

ΠΙΝΑΚΑΣ ΠΕΡΙΕΧΟΜΕΝΩΝ

ΕΥΧΑΡΙΣΤΙΕΣ.....	6
ΠΕΡΙΛΗΨΗ.....	7
ABSTRACT.....	8
ΠΙΝΑΚΑΣ ΠΕΡΙΕΧΟΜΕΝΩΝ.....	9
ΚΑΤΑΛΟΓΟΣ ΔΙΑΓΡΑΜΜΑΤΩΝ/ΕΙΚΟΝΩΝ.....	10
ΠΙΝΑΚΑΣ ΣΥΝΤΟΜΟΓΡΑΦΙΩΝ.....	12
ΑΠΟΔΟΣΗ ΟΡΩΝ / ΓΛΩΣΣΑΡΙΟ.....	13
ΕΙΣΑΓΩΓΗ.....	15
1. Θεωρητικό Υπόβαθρο.....	16
1.1 Βασικές αρχές της τρισδιάστατης ανακατασκευής.....	16
1.2 Μέθοδοι τρισδιάστατης ανακατασκευής: θεωρία και κατηγοριοποίηση.....	17
1.2.1 Φωτογραμμετρικές Μέθοδοι (MVS, SfM).....	17
1.2.2 Μέθοδοι με Βάση το Βάθος.....	20
1.2.3 Νευρωνικά μοντέλα NeRF και η μέθοδος 3D Gaussian Splatting.....	25
1.3 Περιπτώσεις χρήσης τρισδιάστατης ανακατασκευής.....	26
2. Μεθοδολογία.....	29
2.1 Επιλογή δεδομένων και μνημείων.....	30
2.2 Χρησιμοποιούμενα εργαλεία και λογισμικό.....	32
2.3 Περιγραφή των αλγορίθμων που θα υλοποιηθούν ή συγκριθούν.....	34
2.4 Κριτήρια αξιολόγησης και μέθοδοι σύγκρισης.....	42
3. Πειραματική διαδικασία και υλοποίηση.....	43
3.1 Προετοιμασία δεδομένων.....	43
3.2 Ανακατασκευή με διαφορετικούς αλγορίθμους.....	43
3.3 Σύγκριση και αξιολόγηση αποτελεσμάτων.....	44
4. Συμπεράσματα και μελλοντικές προεκτάσεις.....	60
4.1 Βασικά συμπεράσματα.....	60
4.2 Προτάσεις για μελλοντική έρευνα.....	62
ΠΑΡΑΡΤΗΜΑ Α.....	64
Ανακατασκευή στο dataset COURT.....	64
Ανακατασκευή στο dataset TZAMI.....	67
Ανακατασκευή στο dataset TRIKKI.....	71
Ανακατασκευή στο dataset PELLINA.....	74
ΒΙΒΛΙΟΓΡΑΦΙΑ.....	76

ΚΑΤΑΛΟΓΟΣ ΔΙΑΓΡΑΜΜΑΤΩΝ/ΕΙΚΟΝΩΝ

Εικόνα 1.1 Σημειακό νέφος του Κολοσσαίου της Ρώμης.....	16
Εικόνα 1.2 Βασική αρχή της στερεοσκοπικής όρασης.....	18
Εικόνα 1.3 Αποτέλεσμα feature matching για SfM.....	20
Εικόνα 1.4 Colidar Mark II.....	21
Εικόνα 1.5 Τεχνική structured light.....	24
Εικόνα 1.6 Microsoft Kinect v1.....	24
Εικόνα 1.7 Βασική αρχή της τεχνικής NERF.....	26
Εικόνα 1.8. Απεικόνιση ενός 3D gaussian.....	26
Εικόνα 1.9 Στολή του Neil Armstrong.....	27
Εικόνα 1.10 Ναός Vang Stave, Κάρπατς, Πολωνία.....	29
Εικόνα 2.1 Ενδεικτικές φωτογραφίες από κάθε dataset.....	32
Εικόνα 2.2 Βηματική απεικόνιση της παραγωγής depth maps από αρχική τυχαία εκτίμηση.....	38
Εικόνα 2.3 Positional encoding.....	40
Διάγραμμα 3.1 COURT - Συνολικός αριθμός points στο point cloud.....	44
Διάγραμμα 3.2 COURT - Μέση πυκνότητα ανά μονάδα επιφάνειας.....	45
Διάγραμμα 3.3 COURT - Χρόνος εκτέλεσης.....	45
Διάγραμμα 3.4 COURT - Μέση απόσταση από το ground truth.....	45
Διάγραμμα 3.5 COURT - Ποσοστό επιτυχημένης χρήσης των φωτογραφιών του dataset.....	45
Εικόνα 3.1 COURT - LIDAR.....	46
Εικόνα 3.2 COURT - Σύγκριση LIDAR και εικόνας εισόδου.....	46
Εικόνα 3.3 COURT - Σύγκριση SfM και MVS.....	47
Εικόνα 3.4 COURT - NERF.....	47
Εικόνα 3.5 COURT - Gaussian Splatting.....	48
Διάγραμμα 3.6 TZAMI - Συνολικός αριθμός points στο point cloud.....	49
Διάγραμμα 3.7 TZAMI - Χρόνος εκτέλεσης.....	49
Διάγραμμα 3.8 TZAMI - Μέση πυκνότητα ανά μονάδα επιφάνειας.....	49
Διάγραμμα 3.9 TZAMI - Ποσοστό επιτυχημένης χρήσης των φωτογραφιών του dataset.....	49
Εικόνα 3.6 TZAMI - Σύγκριση SfM μεταξύ MicMac και COLMAP.....	50
Εικόνα 3.7 TZAMI - Σύγκριση MVS μεταξύ MicMac και COLMAP.....	51
Εικόνα 3.8 TZAMI - NERF.....	52
Εικόνα 3.9 TZAMI - Gaussian Splatting.....	53
Διάγραμμα 3.10 TRIKKI - Συνολικός αριθμός points στο point cloud.....	53
Διάγραμμα 3.11 TRIKKI - Ποσοστό επιτυχημένης χρήσης των φωτογραφιών του dataset.....	53
Διάγραμμα 3.12 TRIKKI - Χρόνος εκτέλεσης.....	54
Διάγραμμα 3.13 TRIKKI - Μέση πυκνότητα ανά μονάδα επιφάνειας.....	54
Εικόνα 3.10 TRIKKI - Structure from Motion.....	54
Εικόνα 3.11 TRIKKI - Multi View Stereo.....	55
Εικόνα 3.12 TRIKKI - Πυκνότητα Multi View Stereo.....	55
Εικόνα 3.13 TRIKKI - NERF με Nvidia Instant NERFs.....	56
Εικόνα 3.14 TRIKKI - Gaussian Splatting.....	57

Διάγραμμα 3.14 PELINNA - Συνολικός αριθμός points στο point cloud.....	57
Διάγραμμα 3.15 PELINNA - Μέση πυκνότητα ανά μονάδα επιφάνειας.....	57
Διάγραμμα 3.16 PELINNA - Χρόνος εκτέλεσης.....	58
Διάγραμμα 3.17 PELINNA - Ποσοστό επιτυχημένης χρήσης των φωτογραφιών του dataset.....	58
Εικόνα 3.15 PELINNA - Έκταση ανακατασκευής.....	58
Εικόνα 3.16 PELINNA - Structure from Motion και Multi View Stereo.....	58
Εικόνα 3.17 PELINNA - Λανθασμένη ανακατασκευή.....	59
Εικόνα 3.18 PELINNA - Σημεία με σωστή ανακατασκευή.....	59
Εικόνα 4.1 Ανατροφοδότηση SfM/MVS.....	62
Εικόνα A1 COURT - LIDAR ground truth.....	64
Εικόνα A2 COURT - Structure from Motion με COLMAP.....	65
Εικόνα A3 COURT - Multi View Stereo με COLMAP.....	66
Εικόνα A4 COURT - Gaussian Splatting με POSTSHOT.....	66
Εικόνα A5 COURT - NERF με Nvidia Instant NERFs.....	67
Εικόνα A6 TZAMI - Structure from Motion με COLMAP.....	68
Εικόνα A7 TZAMI - Multi View Stereo με COLMAP.....	69
Εικόνα A8 TZAMI - Structure From Motion με MicMac (Tapioca, Tapas & AperiCloud).....	69
Εικόνα A9 TZAMI - Multi View Stereo με MicMac (Tapioca & C3DC).....	70
Εικόνα A10 TZAMI - Gaussian Splatting με POSTSHOT.....	70
Εικόνα A11 TZAMI - Nvidia Instant NERFs.....	71
Εικόνα A12 TRIKKI - Structure From Motion με COLMAP.....	72
Εικόνα A13 TRIKKI - Multi View Stereo με COLMAP.....	72
Εικόνα A14 TRIKKI - Structure From Motion με MicMac (Tapioca, Tapas & AperiCloud).....	73
Εικόνα A15 TRIKKI - Multi View Stereo με MicMac (Tapioca & C3DC).....	73
Εικόνα A16 TRIKKI - Gaussian Splatting με POSTSHOT.....	74
Εικόνα A17 PELINNA - Structure from Motion με COLMAP.....	75
Εικόνα A18 PELINNA - Multi View Stereo με COLMAP.....	75

ΠΙΝΑΚΑΣ ΣΥΝΤΟΜΟΓΡΑΦΙΩΝ

TEI-H.....	Τεχνολογικό Εκπαιδευτικό Ίδρυμα Ηπείρου
3D.....	Τρισδιάστατος
SfM.....	Structure from Motion
MVS.....	Multi View Stereo
CAD.....	Computer Aided Design
CAM.....	Computer Aided Manufacturing
LIDAR.....	Light Detection and Ranging
RGB.....	Red, Green, Blue
RGB-D.....	Red, Green, Blue, Depth
H/Y.....	Ηλεκτρονικός Υπολογιστής
NeRF.....	Neural Radiance Fields
MLP.....	Multi Level Perceptron
LTE.....	Long Term Evolution
GPU.....	Graphics Processing Unit
CPU.....	Central Processing Unit
NCC.....	Normalized Cross-Correlation

ΑΠΟΔΟΣΗ ΟΡΩΝ / ΓΛΩΣΣΑΡΙΟ

Vocabulary tree: Ιεραρχική δένδροειδής δομή που χρησιμοποιείται για την κβαντοποίηση τοπικών χαρακτηριστικών εικόνας και την αποδοτική αναπαράστασή τους ως "οπτικές λέξεις", διευκολύνοντας την ανάκτηση και σύγκριση εικόνων μεγάλης κλίμακας.

Perspective-n-Point (PnP): Αλγόριθμος υπολογισμού της θέσης και του προσανατολισμού μιας κάμερας, βάσει αντιστοιχιών μεταξύ 3D σημείων στο χώρο και των 2D προβολών τους στην εικόνα. Χρησιμοποιείται ευρέως σε εφαρμογές υπολογιστικής όρασης όπως SfM, SLAM και επαυξημένη πραγματικότητα.

Camera solving: Διαδικασία υπολογισμού της θέσης και του προσανατολισμού (pose) μιας κάμερας στον 3D χώρο, συνήθως με βάση αντιστοιχίες μεταξύ 2D σημείων εικόνας και 3D σημείων ή άλλων εικόνων. Χρησιμοποιείται σε SfM, φωτογραμμετρία και κινηματογραφικά VFX.

Σκηνή-στόχος: Το αντικείμενο/σύνολο αντικειμένων της τρισδιάστατης ανακατασκευής.

Φωτομετρικός: Με βάση το οπτικό αποτέλεσμα.

Γεωμετρία: Το σχήμα των αντικειμένων της σκηνής-στόχου.

Υφές (textures): Εικόνες ή δισδιάστατες πληροφορίες που αποδίδονται σε επιφάνειες τρισδιάστατων μοντέλων, με σκοπό την αναπαράσταση της οπτικής εμφάνισης (π.χ. χρώμα, μοτίβο, λεπτομέρειες επιφανειών).

Lookup tables: Προϋπολογισμένοι πίνακες τιμών που επιτρέπουν τη γρήγορη αναζήτηση αποτελεσμάτων χωρίς την ανάγκη επανυπολογισμού. Χρησιμοποιούνται για την επιτάχυνση υπολογισμών, όπως σε μετασχηματισμούς εικόνας, χρωματικούς διορθωτές ή γραφικά υπολογιστών.

Shading: Διαδικασία υπολογισμού του χρωματισμού και της φωτεινότητας σε κάθε σημείο μιας επιφάνειας 3D μοντέλου, με βάση τη γεωμετρία, τον φωτισμό και τη θέση της κάμερας. Στοχεύει στη ρεαλιστική απεικόνιση του τρόπου με τον οποίο το φως αλληλεπιδρά με τα αντικείμενα.

Ground Truth: Ένα σύνολο δεδομένων των οποίων η ακρίβεια έχει επιβεβαιωθεί με μετρήσεις και σύγκριση με το πραγματικό αντικείμενο το οποίο περιγράφουν. Στην παρούσα εργασία, το ground truth πρόκειται για ένα μοντέλο σημειακού νέφους του οποίου οι διαστάσεις είναι σωστές για το αντίστοιχο κτίριο που απεικονίζει.

ΕΙΣΑΓΩΓΗ

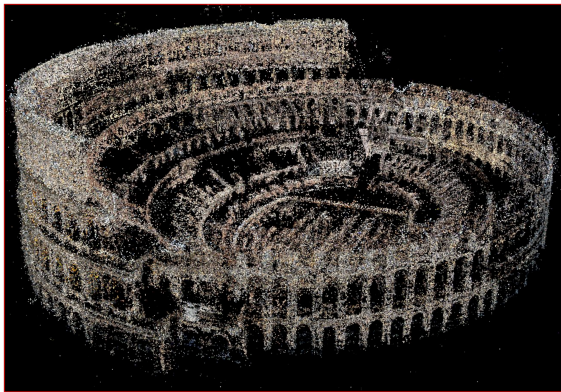
Αν μια εικόνα είναι χίλιες λέξεις, τότε ένα τρισδιάστατο μοντέλο είναι μια ολόκληρη βιβλιοθήκη. Η αποτύπωση ενός αντικειμένου, κτιρίου ή ακόμα και προσώπου ως τρισδιάστατο μοντέλο “κλειδώνει” την μορφή τους στο χρόνο, αποθηκεύοντάς τα αναλλοίωτα και τους επιτρέπει να μελετώνται ατελείωτα στο μέλλον. Κρατά την μορφή τους ασφαλή από το πέρασμα του χρόνου, την φθορά που αυτός επιφέρει σε υλικά και όντα, καθώς και απο εξωγενείς παράγοντες φθοράς όπως η κλιματική αλλαγή, φυσικές καταστροφές και ανθρωπογενείς αιτίες όπως ατυχήματα ή σκόπιμη φθορά μέσω πολεμικών συρράξεων, δολιοφθοράς ή απλά την πρόοδο και αντικατάστασή τους από νεότερες εναλλακτικές.

Η ακρίβεια και αληθοφάνεια της τρισδιάστατης ανακατασκευής την καθιστά αξιόπιστη τεχνολογία για την τεκμηρίωση του σχήματος αντικειμένων σε πολλούς κλάδους του μοντέρνου κόσμου, όπως η αρχαιολογία, η ιατρική, η ψυχαγωγία, η γεωλογία, η αρχιτεκτονική, η βιομηχανία οχημάτων, η ρομποτική και πολλά άλλα. Η εργασία αυτή επικεντρώνεται στην πολιτιστική αξία της τρισδιάστατης ανακατασκευής, συγκεκριμένα στην αποτύπωση και αναπαράσταση έργων τέχνης, ιστορικών κτιρίων, αρχαιολογικών χώρων και άλλων αντικειμένων πολιτιστικής κληρονομιάς. Αναπαραστάσεις σαν και αυτές επιτρέπουν μουσεία να καταγράφουν, να διατηρούν και ακόμα και να εκθέτουν την συλλογή τους με ψηφιακή μορφή. Περαιτέρω, με βάση τα δεδομένα αυτών των τρισδιάστατων αναπαραστάσεων, σε συνδυασμό με άλλες πηγές και κάποιο βαθμό αρτιστικής ελευθερίας, είναι δυνατή η αναπαλαίωση και αποκατάσταση τέτοιων έργων τέχνης και κτιρίων.

1. Θεωρητικό Υπόβαθρο

1.1 Βασικές αρχές της τρισδιάστατης ανακατασκευής

Η βασική ιδέα πίσω από την τρισδιάστατη ανακατασκευή είναι αρκετά απλή στην πρακτική εφαρμογή της. Περιλαμβάνει την συλλογή δεδομένων από το αντικείμενο το οποίο θέλουμε να ανακασκευάσουμε και προαιρετικά, ανάλογα με την μέθοδο, την επεξεργασία αυτών των δεδομένων μέσω κάποιων αλγορίθμων. Μετά την συλλογή και



(Εικόνα 1.1) - Σημειακό Νέφος του Κολοσσαίου της Ρώμης. Πηγή: Condor Framework

επεξεργασία των δεδομένων, δημιουργείται ένα σύνολο στοιχείων στον τρισδιάστατο χώρο, τα οποία αναπαριστούν το σχήμα και πολλές φορές το χρώμα του πραγματικού αντικειμένου του οποίου τα δεδομένα συλλέξαμε. Το σύνολο αυτών των στοιχείων ονομάζεται “point cloud”, ή “σημειακό νέφος”. Το κάθε σημείο στο σημειακό νέφος διαθέτει τις ιδιότητες x,y,z που περιγράφουν την θέση του στον

τρειςδιάστατο χώρο και πολλές φορές και τις ιδιότητες r,g,b που αντιστοιχούν στο χρώμα του αληθινού αντικειμένου της αναπαράστασης, με βάση τα χρώματα Red (κόκκινο), Green (πράσινο) και Blue (μπλε), τα βασικά δηλαδή χρώματα, συνδυασμοί των οποίων μπορούν να αναπαραστήσουν οποιοδήποτε χρώμα. Επίσης συχνά παρατηρείται η εισαγωγή άλλων ιδιοτήτων στα σημεία ενός σημειακού νέφους, όπως πληροφορίες normal για τον υπολογισμό της περιστροφής του σημείου στον τρισδιάστατο χώρο και την υποβοήθηση τυχόν υπολογισμών shading, ανάλογα με την μέθοδο δημιουργίας ή χρήσης του σημειακού νέφους. Υστερα, ανάλογα με τις απαιτήσεις της εφαρμογής, το σημειακό νέφος μπορεί να μετατραπεί σε ένα κλασικό πολυγωνικό μοντέλο μέσω της διαδικασίας “meshing”, το οποίο έχει μια μορφή πιο συμβατή με το υλικό H/Y της αγοράς και μπορεί να χρησιμοποιηθεί σε πιο διαδεδομένες εφαρμογές όπως ψυχαγωγία (ταινίες, βιντεοπαιχνίδια) ή 3D printing ή σε ένα μοντέλο με βάση τις καμπύλες για χρήση σε CAD.

1.2 Μέθοδοι τρισδιάστατης ανακατασκευής: θεωρία και κατηγοριοποίηση

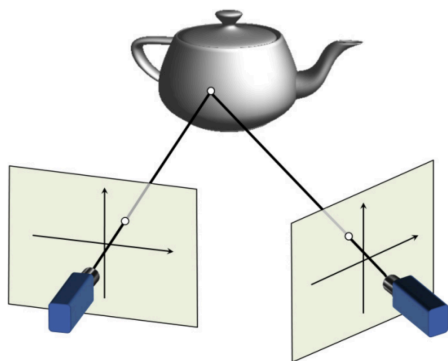
Υπάρχουν διάφορες μέθοδοι τρισδιάστατης ανακατασκευής με διαφορετικές μεθόδους συλλογής δεδομένων, διαφορετικούς τύπους δεδομένων και διαφορετικούς τρόπους επεξεργασίας τους. Γενικά, υπάρχουν τρεις κύριες κατηγορίες αλγορίθμων τρισδιάστατης αναπαράστασης: Φωτογραμμετρικές μέθοδοι, μέθοδοι βασισμένες στο βάθος/απόσταση και ποιο πρόσφατα μέθοδοι βασισμένες σε νευρωνικά μοντέλα. Ο τρόπος απόκτησης της πληροφορίας του βάθους από την σκηνή-στόχο της ανακατασκευής κατηγοριοποιεί τις μεθόδους αυτές σε ενεργές και παθητικές. Οι ενεργές μέθοδοι όπως το LIDAR και οι κάμερες με αισθητήρες βάθους (RGB-D) εκπέμπουν παλμούς φωτός και υπολογίζουν το βάθος με βάση την ανάκλαση του. Οι παθητικές μέθοδοι όπως το Multi View Stereo χρησιμοποιούν αλγόριθμους για να εξάγουν πληροφορίες βάθους από φωτογραφίες. Περεταίρω διαφορές μεταξύ αυτών των μεθόδων έχουν να κάνουν με τις απαιτήσεις υλικού Υ/Η και εξοπλισμού συλλογής δεδομένων και το χρηματικό κόστος που σχετίζεται με αυτά.

1.2.1 Φωτογραμμετρικές Μέθοδοι (MVS, SfM)

Οι φωτογραμμετρικές μέθοδοι είναι παθητικές τεχνικές ανακατασκευής που δέχονται ως δεδομένα ψηφιακές φωτογραφίες του αντικειμένου προς αναπαράσταση. Για αυτόν τον λόγο, η συλλογή δεδομένων για αυτούς τους αλγόριθμους, σε γενικές γραμμές, απαιτεί λιγότερο κόστος και είναι πιο προσβάσιμη. Οποιαδήποτε συσκευή με την ικανότητα λήψης φωτογραφιών ή βίντεο, του οποίου τα καρέ μπορούν να απομονωθούν και να αποθηκευτούν ως διαδοχικές εικόνες, μπορεί να χρησιμοποιηθεί για τη συλλογή δεδομένων. Θα πρέπει όμως να δοθεί προσοχή στον τρόπο λήψης αυτών, ώστε να βελτιστοποιηθεί η αποτελεσματικότητα του αλγορίθμου. Είναι σημαντική η ανάλυση, ο φωτισμός, η γωνία λήψης και η ποσότητα των εικόνων (φωτογραφιών ή καρέ από βίντεο).

Η πρώτη μέθοδος, Structure From Motion (SfM), “Δομή από Κίνηση”, είναι μια μέθοδος που βασίζεται σε τεχνικές μηχανικής όρασης για να μπορέσει να εξάγει δεδομένα βάθους της σκηνής-στόχου με βάση μια σειρά εικόνων εισόδου. Αρχικά, γίνεται εντοπισμός των κοινών σημείων πάνω στα δεδομένα εισόδου (feature matching). Η

εύρεση κοινών σημείων δημιουργεί ζευγάρια εικόνων που απεικονίζουν το ίδιο αντικείμενο ή μέρος της σκηνής-στόχου. Χρησιμοποιώντας τριγωνοποίηση, υπολογίζεται η σχετική θέση και ο προσανατολισμός της κάμερας που αντιστοιχεί σε κάθε εικόνα στον



(Εικόνα 1.2) - Βασική αρχή της στερεοσκοπικής όρασης, η θέση ενός αντικειμένου από δύο ελαφρώς διαφορετικές γωνίες λήψης. (Πηγή: Theo Moons - 3D Reconstruction from Multiple Images)

τρισδιάστατο χώρο. Ύστερα, με βάση τη θέση των κοινών σημείων, καθώς και με βάση τη θέση της κάμερας για κάθε εικόνα εισόδου, μπορεί να υπολογιστεί η θέση κάθε σημείου και κάθε κάμερας στον τρισδιάστατο χώρο [2][3]. Η διαδικασία αυτή ονομάζεται camera solving. Το σύνολο των σημείων που υπολογίστηκαν και τοποθετήθηκαν στον τρισδιάστατο χώρο αποτελεί ένα αραιό

σημειακό νέφος που περιγράφει το βασικό σχήμα της σκηνής-στόχου.

Η δεύτερη μέθοδος, Multi-View Stereo (MVS), αναφέρεται σε μέθοδο μηχανικής όρασης η οποία μας επιτρέπει να εξάγουμε πληροφορίες βάθους μιας σκηνής-στόχου μέσω της σύγκρισης και αντιστοίχισης πολλαπλών εικόνων της ίδιας σκηνής, οι οποίες έχουν ληφθεί από διαφορετικές γωνίες θέασης. Με τον ίδιο ακριβώς τρόπο ο άνθρωπος και άλλα όντα αντιλαμβάνονται την απόσταση αντικειμένων από τα μάτια τους. Συγκεκριμένα, το γεγονός ότι διαθέτουμε δύο μάτια, τα οποία απέχουν μεταξύ τους, μας επιτρέπει να βλέπουμε ταυτόχρονα δύο ελαφρώς διαφορετικές οπτικές γωνίες του ίδιου αντικειμένου. Η διαφορά μεταξύ αυτών των δύο οπτικών γωνιών ονομάζεται “παράλλαξη” [4]. Ο εγκέφαλος δέχεται και επεξεργάζεται αυτές τις δύο οπτικές εισόδους και, με βάση την παράλλαξη, υπολογίζει την απόσταση μεταξύ των ματιών και των αντικειμένων που βλέπει. Στην ίδια λογική βασίζονται οι περισσότερες μέθοδοι απεικόνισης βάθους με ψηφιακά μέσα, όπως τα γυαλιά των 3D κινηματογράφων (ηλεκτρονικά και μη), παιχνίδια τύπου Viewmaster και πιο πρόσφατα, headset εικονικής πραγματικότητας. Όλες αυτές οι συσκευές κατευθύνουν μια ελαφρώς διαφορετική εικόνα στο κάθε μάτι του χρήστη, δημιουργώντας την ψευδαίσθηση του βάθους.

Όσον αφορά το Multi-View Stereo, αντί για δύο μονάχα εικόνες εισόδου, γίνεται επεξεργασία πολλαπλών εικόνων που καλύπτουν τη σκηνή-στόχο. Για την πλήρη τρισδιάστατη ανακατασκευή ενός αντικειμένου, θα χρειαστεί η εκτέλεση Multi-View

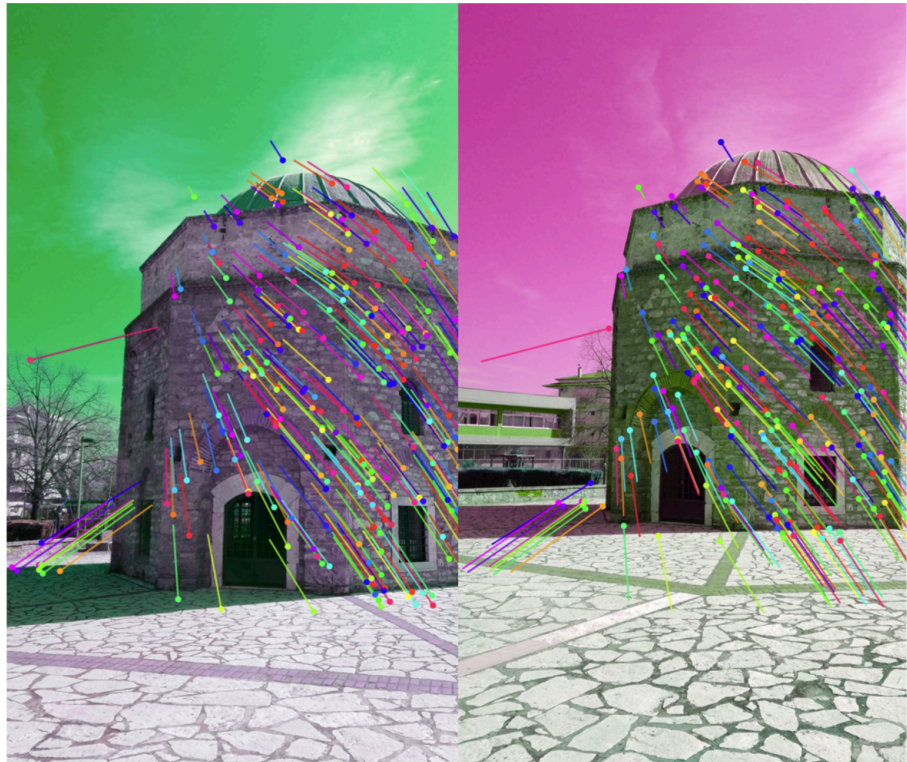
Stereo από όλες τις πλευρές του, ώστε να αποκτήσουμε δεδομένα βάθους για όλα τα σημεία της σκηνής. Σε αυτή την περίπτωση, αρχίζει να εμφανίζεται το μεγαλύτερο μειονέκτημα του MVS: είναι απαραίτητο να γνωρίζουμε την ακριβή θέση κάθε κάμερας στη σκηνή και μεταξύ τους, καθώς και τις παραμέτρους του φακού της κάμερας [6]. Για την επίλυση αυτών των προβλημάτων, το MVS συχνά, αν όχι πάντα, χρησιμοποιείται σε συνδυασμό με την τεχνική SfM.

Για να εξασφαλιστεί ομαλή λειτουργία και καλά αποτελέσματα στη χρήση των SfM και MVS, οι εικόνες εισόδου πρέπει να πληρούν κάποια κριτήρια:

Πρέπει να υπάρχει επικάλυψη (overlap) μεταξύ των εικόνων. Δηλαδή, κάθε σημείο πάνω στο αντικείμενο-στόχο της ανακατασκευής πρέπει να αποτυπώνεται από δύο ή περισσότερες φωτογραφίες. Οι φωτογραφίες πρέπει να ληφθούν από πολλές και διαφορετικές γωνίες λήψης, ώστε να αποτυπώνονται όλες οι πλευρές και λεπτομέρειες του αντικειμένου. Οι διαφορετικές γωνίες λήψης βοηθούν τον αλγόριθμο να “καταλάβει” το σχήμα του αντικειμένου καλύτερα. Από εδώ προέρχεται και η ονομασία του αλγορίθμου, “Δομή από Κίνηση”, με τη λέξη “κίνηση” να αναφέρεται στην αλλαγή της γωνίας λήψης της κάθε φωτογραφίας εισόδου.

Επιπλέον, οι φωτογραφίες πρέπει να έχουν καλή ανάλυση. Καθώς ο αλγόριθμος SfM βασίζεται στην οπτική αναγνώριση κοινών σημείων μεταξύ των φωτογραφιών, χρειάζεται υψηλή ανάλυση ώστε να υπάρχουν περισσότερες λεπτομέρειες και ευκρίνεια για τα σημεία τα οποία μπορεί να αναγνωρίσει. Για τον ίδιο λόγο, το αντικείμενο πρέπει να είναι καλά φωτισμένο. Αν κάποια περιοχή του αντικειμένου είναι υπερβολικά σκοτεινή ή φωτεινή, τότε ο αλγόριθμος SfM θα έχει δυσκολία στο να αναγνωρίσει και να ταιριάζει σημεία σε αυτή. Τέλος, πρέπει να αποφεύγονται αντανάκλασεις και ημιδιάφανες επιφάνειες.

(Εικόνα 1.3) -
Αποτέλεσμα *feature matching* για SfM
μεταξύ δύο
φωτογραφιών μέσω
του λογισμικού
MicMac/Tarioca. Οι
ευθείες απεικονίζουν
την διαφορά στην
θέση των κοινών
σημείων μεταξύ των
φωτογραφιών. Τα
κοινά σημεία που
εντοπίστηκαν είναι
πολλά περισσότερα
αλλά παραλήφθηκαν
για καλύτερη
ορατότητα.



Μετά την εκτέλεση του SfM, το MVS διαθέτει πλέον όλες τις πληροφορίες που χρειάζεται ώστε να ολοκληρώσει την ανακατασκευή της σκηνής από όλες τις οπτικές γωνίες. Ο συνδυασμός των SfM και MVS, για το camera solving και τον υπολογισμό βάθους αντίστοιχα, αποτελεί τον πλέον βασικό τρόπο τρισδιάστατης ανακατασκευής με φωτογραμμετρία, ο οποίος χρησιμοποιείται στις περισσότερες σχετικές εφαρμογές.

Αξίζει να σημειωθεί πως, επειδή τα δεδομένα εισόδου για τους αλγορίθμους SfM/MVS είναι απλές φωτογραφίες, οι αλγόριθμοι αυτοί μπορούν συνήθως να διατηρούν και να αποτυπώνουν το χρώμα των αντικειμένων της αναπαράστασης χωρίς τη χρήση επιπλέον αισθητήρων – σε αντίθεση με άλλες μεθόδους που θα αναλυθούν στη συνέχεια.

1.2.2 Μέθοδοι με Βάση το Βάθος

Σε αυτή την κατηγορία περιλαμβάνονται ενεργές μέθοδοι που χρησιμοποιούν ειδικούς αισθητήρες για την καταγραφή του βάθους μιας σκηνής-στόχου. Κύριοι τύποι τέτοιων εφαρμογών είναι τα συστήματα LIDAR και laser scanning καθώς και ειδικοί τύποι κάμερας με δυνατότητες ανίχνευσης βάθους.

Το LIDAR (πολλές φορές γράφεται ως LiDAR) είναι ακρώνυμο του “light detection and ranging” και παραπλήσιο του RADAR (radio detection and ranging) και πολλές φορές ο όρος ταυτίζεται με τον όρο “laser scanner” ή “3D scanner”. Αν και οι συσκευές laser

scanning χρησιμοποιούν το LIDAR ως το βασικό τους κομμάτι, συχνά περιλαμβάνουν και άλλα όργανα και λειτουργίες που βοηθούν στην δημιουργία ενός νέφους σημείων. Όπως τα συστήματα ραντάρ χρησιμοποιούν ραδιοκύματα για την ανίχνευση αντικειμένων, έτσι και το lidar χρησιμοποιεί κύματα φωτός παραγόμενα από ένα λέιζερ για τον ίδιο σκοπό. Το πρώτο σύστημα μέτρησης απόστασης με βάση το λέιζερ δημιουργήθηκε τον Ιανουάριο του 1961 από την Hughes Aircraft Company, ένα χρόνο μετά την εφεύρεση του λέιζερ. Η εταιρία ήδη πειραματιζόταν επιτυχώς με συστήματα ραντάρ στο φάσμα των



(Εικόνα 1.4) - Ο μηχανικός Rod Smith της εταιρίας Hughes Aircraft Company στοχεύει το Colidar Mark II, το πρώτο εμπορικό μοντέλο lidar. (Πηγή: HRL Laboratories, LLC)

μικροκυμάτων. Υπέθεσαν έτσι ότι αφού η ακρίβεια ενός ραντάρ αυξάνεται όσο μικραίνει το μήκος κύματος της ακτινοβολίας του, τότε είναι δυνατόν να χρησιμοποιηθεί το φως ενός λέιζερ αντί για ραδιοκύματα. Η εφεύρεση ονομάστηκε Colidar (coherent light detection and ranging). Η πρώτη εκδοχή (Mark I) είχε ως βάση ένα ερασιτεχνικό τηλεσκόπιο και έναν τροποποιημένο υπέρυθρο ανιχνευτή αστερών. Η δεύτερη εκδοχή (Mark II) έγινε ποιο

φορητή, με την πηγή λέιζερ και τον αισθητήρα να είναι εφαρμοσμένα πάνω σε κοντάκι πυροβόλου όπλου για εύκολη στόχευση και τα υπόλοιπα ηλεκτρονικά παρελκόμενα μπορούσαν να φορεθούν από τον χρήστη σαν σακίδιο πλάτης (Εικόνα 1.4). Η Hughes Aircraft Company άρχισε να διαθέτει το Mark II στο εμπόριο σε ΗΠΑ και Ευρώπη, με τους πελάτες τους να ενδιαφέρονται κυρίως για στρατιωτικές εφαρμογές, συνήθως για συστήματα στόχευσης σε αρμάτα μάχης [8]. Στις μέρες μας, συστήματα LIDAR χρησιμοποιούνται σε πολλούς τομείς και διατίθενται σε διάφορες μορφές ανάλογα με τον σκοπό χρήσης τους, όπως συστήματα χειρός για τρισδιάστατη ανακατασκευή μικρού ως μεσαίου μεγέθους αντικειμένων, συστήματα που εφαρμόζονται σε οχήματα όπως αυτοκίνητα για την τρισδιάστατη ανακατασκευή δρόμων και αστικών περιβαλλόντων, αεροπλάνα και ελικόπτερα για την τρισδιάστατη ανακατασκευή μεγάλων τμημάτων γης ή πλοία για την καταγραφή του αναγλύφου του βυθού της θάλασσας. Το 2020 η Apple συμπερίλαβε ένα βασικό σύστημα LIDAR στο νέο μοντέλου του κινητού τηλεφώνου τους, δίνοντας στο ευρύ κοινό μια φθηνή λύση σε σύγκριση με τα κατά πολύ ακριβότερα, επαγγελματικά συστήματα LIDAR [10]. Ανάλογα με την εφαρμογή αλλάζει και η

συχνότητα του λέιζερ, συνήθως υπέρυθρο, με συχνότητα 905nm ή 1550nm για υπαίθρια ή εν αέρος χρήση και 515nm για υποβρύχιες εφαρμογές. Η λογική για τον υπολογισμό του βάθους της σκηνής που χρησιμοποιούν τα συστήματα LIDAR ονομάζεται time-of-flight (χρόνος-πτήσης) του φωτός, δηλαδή ο χρόνος τον οποίο χρειάζεται μια δέσμη ή παλμός φωτός για να καλύψει την διαδρομή από την πηγή εκπομπής τού μέχρι κάποιο αντικείμενο-στόχο και ξανά πίσω στον αισθητήρα (ο οποίος συχνότερα βρίσκεται κοντά/στην ίδια συσκευή με την πηγή εκπομπής). Σε συνδυασμό με το γεγονός ότι η ταχύτητα του φωτός είναι σταθερή, η απόσταση μεταξύ της συσκευής και του στόχου μπορεί να υπολογιστεί με την συνάρτηση $d = (c \cdot t) / 2$, όπου d η απόσταση της διαδρομής πηγής-στόχου-αισθητήρα), c η ταχύτητα του φωτός και t ο χρόνος-πτήσης του φωτός. Αλλάζοντας το σημείο πάνω στον στόχο στο οποίο η δέσμη ή ο παλμός φωτός ανακλάται, αποκτούμε πληροφορίες για την απόσταση του κάθε σημείου του στόχου από την συσκευή. Αυτές οι πληροφορίες με την σειρά τους μπορούν να μετατραπούν σε ένα σημειακό νέφος που αντιπροσωπεύει την σκηνή.

Το LIDAR σε γενικές γραμμές θεωρείται πιο ακριβές και αξιόπιστο από τις φωτογραμμετρικές μεθόδους του προηγούμενου κεφαλαίου, καθώς και πιο γρήγορο στην χρήση, διότι η επεξεργασία των δεδομένων της σκηνής-στόχου γίνεται σε πραγματικό χρόνο πάνω στην συσκευή και όχι αργότερα σε κάποιον σταθερό υπολογιστή. Τα πλεονεκτήματα επίσης περιλαμβάνουν την ικανότητα του LIDAR να λειτουργεί ανεξαιρέτως του φωτισμού της σκηνής-στόχου, διότι βασίζεται στο υπέρυθρο φως που εκπέμπει το ίδιο, πράγμα που του επιτρέπει να δουλεύει ακόμη και στο σκοτάδι. Τέλος, ανάλογα με την συχνότητα του laser, το LIDAR μπορεί ακόμα και να διαπεράσει ημιδιάφανες επιφάνειες όπως το φύλλωμα φυτών και δέντρων. Μια χρήσιμη εφαρμογή αυτής της ιδιότητας είναι για παράδειγμα η ανακατασκευή του ανάγλυφου του εδάφους από αέρος σε μια δασική περιοχή, όπου η πυκνή βλάστηση κρύβει το έδαφος και καθιστά φωτογραμμετρικές μεθόδους μη βιώσιμες. Παρόλα αυτά, το LIDAR έχει και αυτό τα μειονεκτήματά του. Κύριο μειονέκτημα είναι η περιορισμένη ακτίνα δράσης του. Η ενέργεια του laser μειώνεται με την απόσταση και επηρεάζεται από τις ατμοσφαιρικές συνθήκες (υγρασία αέρα, ομίχλη, βροχή, χιόνι) και έτσι μπορεί να είναι αδύνατο να ανιχνευθεί η επιστροφή του φωτός στον αισθητήρα και ως εξής η απόκτηση δεδομένων. Επίσης, τα συστήματα LIDAR συνδέονται συνήθως από υψηλό κόστος και έτσι χρησιμοποιούνται συνήθως μόνο σε επαγγελματικές εφαρμογές. Ακόμα, τα συστήματα αυτά είναι αρκετά ενεργοβόρα, πράγμα που περιορίζει την φορητότητα τους όταν η

τροφοδοσία τους βασίζεται σε μπαταρίες. Ένας ακόμα περιοριστικός παράγοντας είναι η μεγάλη πολυπλοκότητα των δεδομένων που παράγουν τα συστήματα. Τα σημειακά νέφη απαιτούν μεγάλο ψηφιακό χώρο αποθήκευσης και μεγάλη υπολογιστική ισχύ για την απεικόνιση και επεξεργασία τους [10]. Πολλά φορητά συστήματα laser scanning συνιστούν την καταγραφή της σκηνής-στόχου σε πολλαπλές, χρονικά σύντομες συνεδρίες, συχνά μόνο λίγων λεπτών, και στην συνέχεια τον συνδυασμό όλων των διαφορετικών σημειακών νεφών που προκύπτουν από αυτές σε έναν σταθερό υπολογιστή με μεγαλύτερη επεξεργαστική δύναμη [11]. Τέλος, αξίζει να σημειωθεί ότι τα συστήματα LIDAR δεν έχουν την ικανότητα καταγραφής του χρώματος της σκηνής-στόχου και έτσι πολλές φορές οι συσκευές laser scanning διαθέτουν συμβατικές έγχρωμες κάμερες που καταγράφουν την σκηνή την ίδια στιγμή και δίνουν χρώμα στο σημειακό νέφος που παράγεται από το LIDAR.

Σε αντίθεση με το LIDAR που αποτελεί καθαρά ενεργή μέθοδο, οι κάμερες με δυνατότητα βάθους μπορούν να είναι και ενεργές αλλά και παθητικές ανάλογα με τον τρόπο που καταγράφουν το βάθος της σκηνής-στόχου. Συχνά, τέτοιες κάμερες αναφέρονται ως κάμερες RGB-D, δηλαδή red, green, blue, δηλαδή τα βασικά χρώματα, και depth (βάθος). Οι κάμερες που καταγράφουν το βάθος παθητικά χρησιμοποιούν ουσιαστικά δύο φακούς και αισθητήρες, δηλαδή στην ουσία είναι δύο κάμερες που λειτουργούν ως μία μονάδα. Χρησιμοποιώντας τις δύο εικόνες ταυτόχρονα, μπορούν να εφαρμόσουν αλγόριθμους τύπου Multi-view Stereo όπως περιγράφεται στο κεφάλαιο 2.2.1. Μεγαλύτερο ενδιαφέρον παρουσιάζουν οι κάμερες βάθους με έναν ή δύο RGB αισθητήρες και μια ενεργή μέθοδο ανίχνευσης βάθους. Οι δύο κύριοι τύποι ενεργών μεθόδων εδώ είναι time-of-flight (χρόνος-πτήσης) και structured light (δομημένο φως) [12].

Η μέθοδος time-of-flight λειτουργεί όπως και περιγράφηκε προηγουμένως στο LIDAR: Μια δέσμη ή ριπή φωτός, πιο συχνά στο υπέρυθρο φάσμα, παράγεται από την συσκευή, ανακλάται από το αντικείμενο-στόχο και επιστρέφει στον αισθητήρα της συσκευής. Ο χρόνος διάρκειας της διαδρομής του φωτός από τον συσκευή ως το αντικείμενο και ξανά πίσω στην συσκευή μετατρέπεται σε μέτρηση βάθους.

Η μέθοδος structured light χρησιμοποιεί μια δέσμη φωτός, ορατό ή σχεδόν υπέρυθρο, που εκπέμπεται από την συσκευή με τρόπο όμοιο με έναν προτζέκτορα κινηματογράφου. Το φως εκπέμπεται έτσι ώστε να δημιουργεί ένα συγκεκριμένο μοτίβο

που συνήθως μοιάζει με παράλληλες ευθείες. Όταν το μοτίβο αυτό εκπέμπεται πάνω στο αντικείμενο-στόχο, το μοτίβο παραμορφώνεται και οι ευθείες δεν είναι πλέον παράλληλες. Η RGB κάμερα τότε μεταφράζει την παραμόρφωση αυτή σε πληροφορίες που περιγράφουν το σχήμα του αντικειμένου-στόχου. Επίσης, η απόσταση μεταξύ των ευθειών καθώς και το πάχος τους αυξομειώνεται με την απόσταση του αντικειμένου-στόχου από τον προτζέκτορα. Η αλλαγές αυτές μεταφράζονται σε πληροφορίες που περιγράφουν την απόσταση του αντικειμένου-στόχου από τον προτζέκτορα.



(Εικόνα 1.5) - Το μοτίβο φωτός ενός 3D scanner με βάση την τεχνική structured light πάνω σε δύο αντικείμενα. (Πηγή: 3DUNDERWORLD-SLS: An Open-Source Structured-Light Scanning System for Rapid Geometry Acquisition)

Ένα αξιοσημείωτο παράδειγμα κάμερας RGB-D με χρήση της λογικής structured light είναι το Microsoft Kinect (v1), μια περιφερειακή συσκευή για την παιχνιδοκονσόλα Xbox 360 της Microsoft. Το Kinect κυκλοφόρησε το 2010 και αποτελείται από έναν υπέρυθρο προτζέκτορα και μία ασπρόμαυρη κάμερα. Ήταν το πρώτο προϊόν που χρησιμοποιούσε κάμερα βάθους που απευθύνεται στο ευρύ κοινό και είχε προσιτή τιμή. Σκοπός του ήταν να επιτρέπει στους χρήστες της κονσόλας να αλληλεπιδρούν με το παιχνίδι χρησιμοποιώντας κινήσεις του σώματός τους, χειρονομίες και εκφράσεις του προσώπου τους. Η προσιτότητα του και η ποιότητα της κάμερας βάθους του επέτρεψαν το kinect να χρησιμοποιείται σε πεδία εκτός των βιντεοπαιχνιδιών, όπως η ρομποτική και το 3D scanning και ακόμα αποτελεί βασικό υλικό για “σπιτικούς” 3D scanners.



(Εικόνα 1.6) - Microsoft Kinect v1 (πηγή: Microsoft)

Σήμερα, οι 3D scanners που χρησιμοποιούν structured light είναι διαδεδομένοι σε τομείς όπως η αυτοκινητοβιομηχανία και τα μηχανουργεία και το χαμηλό κόστος της μεθόδου αυτής την κάνει προσιτή για ερασιτεχνικές ή χαμηλού κόστους συσκευές 3D scanning.

Όπως όλες οι μέθοδοι, το structured light έχει και αυτό τα μειονεκτήματά του. Συγκεκριμένα, αντιμετωπίζει προβλήματα στην ανακατασκευή μαύρων/σκούρων αντικειμένων, διότι το σκούρο χρώμα απορροφά το φως, κάνοντας το μοτίβο του φωτός που εκπέμπει η συσκευή λιγότερο ορατό, μειώνοντας την ακρίβεια του αποτελέσματος. Είναι συχνό φαινόμενο οι μαυρες περιοχές ενός αντικειμένου-στόχου να μην ανακατασκευάζονται και να λείπουν εντελώς από το σημειακό νέφος του αποτελέσματος.

1.2.3 Νευρωνικά μοντέλα NeRF και η μέθοδος 3D Gaussian Splatting

Οι μέθοδοι που αναλύονται σε αυτή την ενότητα αποτελούν πρόσφατες εξελίξεις στον χώρο της τρισδιάστατης ανακατασκευής και ενώ βασίζονται σε κλασικές μεθόδους για την αρχικοποίησή τους (συγκεκριμένα στον αλγόριθμο SfM), διαφοροποιούνται σημαντικά από αυτόν όσον αφορά στην μορφή του αποτελέσματος και την μέθοδο απεικόνισης του, ο οποίος δεν συμβαδίζει με παραδοσιακές τεχνικές τρισδιάστατων γραφικών (νέφη σημείων και πολυγωνικά μοντέλα). Πρόκειται για τις μεθόδους NeRF και Gaussian Splatting, όπου και οι δύο αντιμετωπίζουν το πρόβλημα της τρισδιάστατης ανακατασκευής με νέους και διαφορετικούς τρόπους.

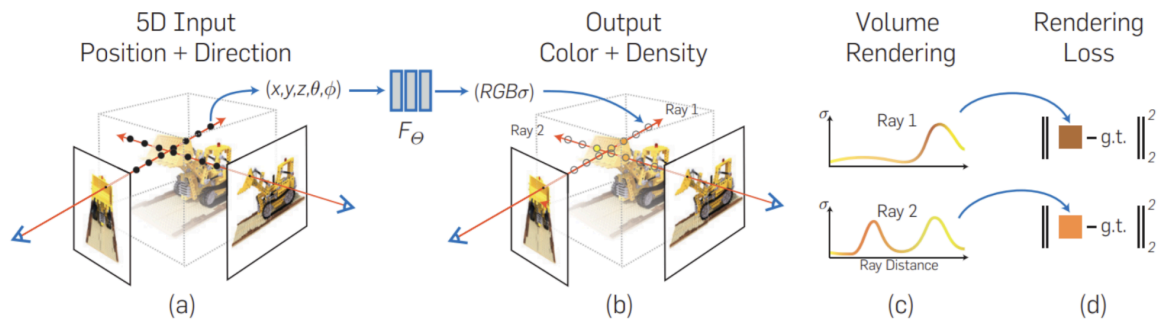
Η μέθοδος NeRF (Neural Radiance Fields) πρόκειται για μια παθητική μέθοδο τρισδιάστατης ανακατασκευής που χρησιμοποιεί ως είσοδο φωτογραφίες (ή καρέ από βίντεο) και πληροφορίες από την μέθοδο SfM, παρόμοια με φωτογραμμετρικές μεθόδους. Αντί όμως να δημιουργεί ένα σημειακό νέφος, οι πληροφορίες αυτές χρησιμοποιούνται για την εκπαίδευση ενός perceptron πολλαπλών στρώσεων το οποίο μπορεί να συνθέσει το οπτικό αποτέλεσμα για όλες τις νέες πιθανές γωνίες θέασης μιας σκηνής-στόχου, συμπεριλαμβανομένου και γωνιών που δεν καλύπτονται άμεσα από τα δεδομένα εισόδου. Δηλαδή, το perceptron χρησιμοποιεί τις φωτογραφίες εισόδου ώστε να “μάθει” την σκηνή και στην συνέχεια “υποθέτει” τι μπορεί να δει μια κάμερα που βλέπει την σκηνή από μία

άλλη,

νέα

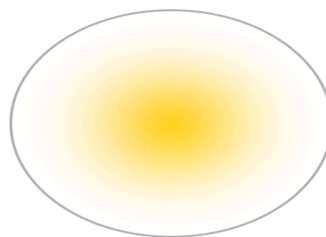
οπτική

γωνία.



(Εικόνα 1.7) - Βασική αρχή της τεχνικής NeRF. (Πηγή: NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis [13])

Το 3D gaussian splatting αποτελεί μια ακόμα νέα μέθοδο απεικόνισης μιας τρισδιάστατης σκηνής αλλά και αυτή έχει τις ρίζες της στην λογική που ακολουθεί η μέθοδος SfM. Συνδυάζει παραδοσιακές τεχνικές τρισδιάστατης ανακατασκευής και τρισδιάστατων γραφικών με μερικά στοιχεία της μεθόδου NeRF ώστε να επιτύχει παρόμοια ποιότητα ανακατασκευής με τα NeRF, αλλά παράλληλα και καλύτερη απόδοση. Η βασική ιδέα πίσω από το gaussian splatting είναι η απεικόνιση ενός κλασικού νέφους σημείων χρησιμοποιώντας σφαίρες που ονομάζονται 3D gaussians, των οποίων η εμφάνιση ακολουθεί μια γκαουσιανή καμπύλη, κάνοντας το κέντρο της σφαίρας αδιαφανές, ενώ η υπόλοιπη σφαίρα γίνεται όλο και πιο διαφανής καθώς πλησιάζουμε πιο κοντά στην επιφάνεια.



(Εικόνα 1.8) - Απεικόνιση ενός 3D gaussian (πηγή: Dylan Ebert - Introduction to 3D Gaussian Splatting)

1.3 Περιπτώσεις χρήσης τρισδιάστατης ανακατασκευής.

Όλες αυτές οι μέθοδοι χρησιμοποιούνται για την ψηφιακή ανακατασκευή μνημείων πολιτιστικής κληρονομιάς και άλλων χώρων για διάφορους σκοπούς.

Συντήρηση και επισκευή

Η τρισδιάστατη ανακατασκευή έχει πολλές εφαρμογές στην αρχαιολογία. Η ακρίβεια των μοντέλων που παράγονται επιτρέπει σε αρχαιολόγους να μελετήσουν ένα αντικείμενο ή μνημείο χωρίς να χρειάζεται απαραίτητα η δια ζώσης επίσκεψή τους σε αυτό. Επίσης του επιτρέπει απεριόριστη πρόσβαση, χωρίς χρονικά περιθώρια και τους επιτρέπει την θέαση του σημείου ενδιαφέροντος από οπτικές γωνίες που θα ήταν δύσκολα προσβάσιμες στην πραγματικότητα, όπως για παράδειγμα στενούς χώρους ή μεγάλα ύψη. Για παράδειγμα, ένα αρχαίο μωσαϊκό, μετά την τρισδιάστατη ανακατασκευή του, θα μπορεί να θεαθεί από ύψος παρόμοιο με drone χωρίς την χρήση τέτοιου εξοπλισμού, ακόμα και όταν στην πραγματικότητα η οπτική γωνία αυτή είναι αδύνατη (π.χ σε μωσαϊκά που βρίσκονται σε εσωτερικό χώρο). Άλλες φορές, η τρισδιάστατη ανακατασκευή χρησιμοποιείται για την αναπαλαίωση και συντήρηση του αντικειμένου/κτιρίου. Ένα παράδειγμα στον Ελληνικό χώρο αποτελεί ο Παρθενώνας της Ακρόπολης της Αθήνας, για την αποκατάσταση του οποίου χρησιμοποιείται τρισδιάστατη ανακατασκευή των υπαρχόντων μαρμάρινων επιφανειών, έτσι ώστε τα νέα μάρμαρα που λαξεύονται για την επισκευή του να εφάπτονται ακριβώς με τα αρχαία.

Τουριστική προώθηση

Η ψηφιοποίηση μνημείων και αντικειμένων επιτρέπει ενδιαφερόμενους να τα “επισκέπτονται” ψηφιακά, όταν δεν είναι δυνατή η δια ζώσης επίσκεψη. Κάποια μουσεία έχουν ψηφιοποιήσει την συλλογή τους για λόγους προσβασιμότητας (Εικόνα 1.9)[15]. Κάποια εκθέματα είναι πολύ ευαίσθητα για να εκτεθούν μπροστά σε φυσικό κοινό και έτσι με την χρήση τρισδιάστατης ανακατασκευής οι επισκέπτες μπορούν να τα θαυμάσουν εικονικά. Σε άλλες περιπτώσεις, το μουσείο χρησιμοποιεί αυτές τις ανακατασκευές για να παράγει ένα πιστό φυσικό αντίγραφο το οποίο εκτίθεται στην θέση του πραγματικού. Η συμβατότητα των αποτελεσμάτων των αλγορίθμων τρισδιάστατης ανακατασκευής με λογισμικά επεξεργασίας τρισδιάστατων μοντέλων (3D editing software) όπως το Blender, Maya, 3DS Max και άλλα, επιτρέπουν την επεξεργασία αυτών των αντικειμένων/κτιρίων από επιστήμονες και 3D artists.



(Εικόνα 1.9) - Τρισδιάστατο μοντέλο της στολής του Neil Armstrong - Μίξη 3D Reconstruction και ανθρωπίνων διορθώσεων. (πηγή: The Smithsonian Museum)

Συνήθως αυτό γίνεται για λόγους αναπαλαίωσης (αφαίρεσης λεκέδων, επισκευή του χρώματος του αντικειμένου, επισκευή ρωγμών, αντικατάσταση κατεστραμμένων κομματιών). Για παράδειγμα, η τρισδιάστατη ανακατασκευή του Παρθενώνα επιτρέπει την ψηφιακή αναπαλαίωσή του η οποία είναι διαθέσιμη στους επισκεπτες ως μια εφαρμογή κινητού τηλεφώνου που τους επιτρέπει να δουν το κτίσμα όπως ήταν στην αρχαιότητα [16].

Ψυχαγωγία

Η χρήση τρισδιάστατης ανακατασκευής μνημείων και αντικειμένων πολιτισμικής κληρονομιάς στον τομέα της ψυχαγωγίας και συγκεκριμένα στους τομείς του κινηματογράφου, της τηλεόρασης και των βιντεοπαιχνιδιών επιτρέπει τους σκηνοθέτες, με την χρήση ψηφιακών εφέ, να βασίσουν το έργο τους σε τόπους και χρόνους που δεν θα ήταν δυνατό υπο κανονικές συνθήκες. Οι σειρές, ταινίες και βιντεοπαιχνίδια με ιστορικό πλαίσιο επωφελούνται ιδιαίτερα από αυτό. Για παράδειγμα, η σειρά “Liberty or Death: Boston Tea Party” που διαδραματίζεται κατά την διάρκεια της Αμερικανικής Επανάστασης χρησιμοποιεί LIDAR για την τρισδιάστατη ανακατασκευή του Mystic Seaport, ενός ναυτικού μουσείου στο Κονέκτικατ των Ηνωμένων Πολιτειών το οποίο περιλαμβάνει ένα χωριό και λιμάνι τα οποία μιμούνται την αρχιτεκτονική και τεχνολογία αντίστοιχων λιμένων του 19ου αιώνα. Το χωριό και τα ιστιοφόρα πλοία, μετά την τρισδιάστατη ανακατασκευή τους, χρησιμοποιήθηκαν ως ψηφιακά σκηνικά στην παραγωγή της σειράς [17,18]. Άλλο ένα παράδειγμα αποτελεί το βιντεοπαιχνίδι The Vanishing of Ethan Carter (The Astronauts, 2014). Γίνεται χρήση τεχνικών φωτογραμμετρίας για διάφορα αντικείμενα του περιβάλλοντος. Σύμφωνα με τους developers, η ρεαλιστικότητα των αντικειμένων εμπλουτίζει την εμπειρία τόσο οπτικά όσο και συναισθηματικά, διότι ο παίκτης μπορεί να συγκεντρωθεί περισσότερο στην ιστορία του παιχνιδιού όταν το περιβάλλον είναι αληθινό και “οικείο” για το υποσυνείδητο του [19].



(Εικόνα 1.10) - Ναός Vang Stave, Κάρπατς, Πολωνία - Το πραγματικό κτίριο αριστερά. Στιγμιότυπο από το παιχνίδι δεξιά.
(πηγή: GOG.com)

Εκπαίδευση

Η ύπαρξη ενός ακριβούς ψηφιακού αντιγράφου μιας τοποθεσίας, κτιρίου ή αντικειμένου επιτρέπει την χρήση του σε ψηφιακά περιβάλλοντα με σκοπό την εκπαίδευση ενός χρήστη για μια δραστηριότητα που παίρνει μέρος στον ίδιο ή παρόμοιο χώρο στον πραγματικό κόσμο. Για παράδειγμα, η εκπαίδευση οδηγών αγώνων ταχύτητας σε κατηγορίες όπως η Φόρμουλα Ένα και άλλες μεγάλες κατηγορίες γίνεται σε προσομειωτές οδήγησης όπου οι εκπαιδευόμενοι οδηγοί οδηγούν ψηφιακά αυτοκίνητα σε ψηφιακά αντίγραφα των διαδρομών και πιστών στις οποίες αγωνίζονται στην πραγματικότητα. Κάθε ανομοιομορφία της ασφάλτου επηρεάζει την οδηγική συμπεριφορά του αυτοκινήτου και έτσι τα αντίγραφα αυτά συνήθως έχουν ακρίβεια χιλιοστού.

2. Μεθοδολογία

Στο κεφάλαιο αυτό θα αναλυθεί το πρακτικό κομμάτι της εργασίας. Στόχος της έρευνας είναι η επιλογή κάποιων μνημείων πολιτιστικής κληρονομιάς, η επίσκεψή τους για συλλογή δεδομένων, η χρήση των δεδομένων αυτών σε κάποιες από τις τεχνικές τρισδιάστατης ανακατασκευής που αναφέρθηκαν στο προηγούμενο κεφάλαιο και η σύγκριση των αποτελεσμάτων. Παράλληλα, οι ίδιοι αλγόριθμοι θα εξεταστούν σε ελεγχόμενο περιβάλλον με δοκιμές που θα στοχεύουν στην αξιολόγηση συγκεκριμένων παραμέτρων του αλγορίθμου κάθε φορά.

2.1 Επιλογή δεδομένων και μνημείων

Όλοι οι αλγόριθμοι που θα χρησιμοποιηθούν είναι παθητικοί και δέχονται φωτογραφίες ως είσοδο. Αυτό καθιστά στην συλλογή δεδομένων για την ανακατασκευή ευκολη όσον αφορά τον εξοπλισμό. Υπάρχουν όμως κάποια ακόμα κριτήρια που πρέπει να ληφθούν υπόψη. Οι φυσικές διαστάσεις του χώρου/αντικειμένου προς ανακατασκευή μπορούν να αποτελέσουν εμπόδιο στην σωστή συλλογή φωτογραφιών και ως αποτέλεσμα την καλή ανακατασκευή μιας σκηνής. Χρειάζεται καλή οπτική πρόσβαση σε ολόκληρο τον χώρο και την επιφάνεια των αντικειμένων. Λόγω της έλλειψης drone ή άλλου εξοπλισμού για την φωτογράφιση αντικειμένων με ύψος, κομμάτια κτιρίων όπως οι σκεπές μπορεί να αποδειχθούν πρόκληση. Ένα ακόμα κριτήριο είναι ο φωτισμός της σκηνής. Χρειάζεται καλός φωτισμός για την αποφυγή απώλειας στην ανακατασκευή. Ιδανικός φωτισμός είναι διάχυτο, λευκό φως που καλύπτει ομοιόμορφα την σκηνή χωρίς να δημιουργεί έντονες σκιές ή υπερέκθεση σε σημεία της σκηνής-στόχου, που μπορούν να κρύψουν λεπτομέρειες στις φωτογραφίες και ως συνέπεια αυτές οι λεπτομέρειες να χαθούν από την ανακατασκευή. Οι καιρικές συνθήκες παίζουν επίσης ρόλο στην συλλογή φωτογραφιών. Δεν συνιστάται η φωτογράφιση για τρισδιάστατη ανακατασκευή κατά την διάρκεια βροχόπτωσης διότι οι βρεγμένες επιφάνειες αλλοιώνουν την ακρίβεια των χρωμάτων και αποτελούν πρόκληση για τους αλγόριθμους ανακατασκευής λόγω των αντανάκλασεων. Ιδανικές συνθήκες αποτελούν μέρες με νεφελώδη ουρανό και χωρίς βροχή. Αυτό αποτρέπει τον ήλιο από το να δημιουργεί σκιες στον χώρο της ανακατασκευής και δημιουργεί έναν πιο ομοιόμορφο φωτισμό που αλληλεπιδρά πιο φυσικά με τυχόν τεχνητό φωτισμό που μπορεί να χρησιμοποιηθεί στο μοντέλο-εξόδο της ανακατασκευής.

Πέρα των τεχνικών προκλήσεων, υπάρχουν και νομικά θέματα στα οποία πρέπει να υπακούει η έρευνα. Όλες οι επισκεψεις και τρισδιάστατες ανακατασκευές γίνονται με την συναίνεση των αρμόδιων αρχών για τις διάφορες σκηνές-στόχους της έρευνας. Η συνεργασία των αρμόδιων αρχών είναι απαραίτητη για την σωστή φωτογράφιση καθώς και εκπόνηση έρευνας σχετικά με αυτές. Αν και δεν υπάρχει σαφές πρωτόκολλο για την τρισδιάστατη ανακατασκευή, τουλάχιστον στις σκηνές-στόχους της παρούσας εργασίας, η αδειοδότηση έγινε μετά από συνεννόηση με τις αρμόδιες αρχές στα πλαίσια μιας απλής φωτογράφισης. Τελικό κριτήριο είναι η προσβασιμότητα σε σχέση με τον τόπο κατοικίας του ερευνητή. Καθώς η συλλογή δεδομένων απαιτεί την δια ζώσης επίσκεψη του χώρου, όλες οι τοποθεσίες βρίσκονται σε κοντινή απόσταση από τα Τρίκαλα Θεσσαλίας.

Επιπλέον παρατηρήσεις:

Για την εξασφάλιση των αποτελεσμάτων, θα χρησιμοποιηθούν και δεδομένα που συλλέχθηκαν από άλλες ομάδες και η αποτελεσματικότητά τους σε χρήση αλγορίθμων τρισδιάστατης ανακατασκευής είναι επιβεβαιωμένη. Επίσης, τα δεδομένα αυτά παρέχουν ένα ελεγμένο τρισδιάστατο μοντέλο της αντίστοιχης σκηνής το οποίο θα χρησιμοποιηθεί ως βάση σύγκρισης (ground truth) για τους αλγορίθμους που αξιολογούνται στην εργασία αυτή.

Με βάση τα παραπάνω, επιλέχθηκαν τα εξής μνημεία για επισκεψη για λόγους της εργασίας:

- Μουσείο Οσμάν Σαχ στην πόλη των Τρικάλων. Πρόκειται για οκτάγωνο ταφικό κτίσμα που βρίσκεται δίπλα από το τζαμί Οσμάν Σαχ. Ο ίδιος απεβίωσε το 1567 ή 1568 και είναι ενταφιασμένος στο κτίσμα [20]. Η τοποθεσία επιλέχθηκε λόγω της εύκολης πρόσβασης, της μικρής και γεωμετρικής (οκταγωνης) περιμέτρου του κτιρίου και της οροφης που έχει σχήμα τρούλου, πράγμα που την κάνει πιο ορατή από το έδαφος σε σύγκριση με παραδοσιακές επίπεδες ή λοξές σκεπές.
- Ανασκαφή κτήριο Α', Αρχαιολογικός χώρος Αρχαίας Πέλιννας, Πετρόπορο, Τρίκαλα. Πρόκειται για ελληνιστική και ύστερα βυζαντινή πόλη με οχύρωση και κτίσματα (3ος - 2ος αι. π.Χ.). Οι ανασκαφές είναι σε εξέλιξη [21]. Η τοποθεσία επιλέχθηκε λόγω της εύκολης πρόσβασης, της καλής οπτικής σε όλη την ανασκαφή, της πολυπλοκότητας και το μέγεθος του χώρου.
- Γλυπτό "ΤΡΙΚΚΗ", πεζόδρομος Ασκληπιού, Τρίκαλα. Προκειται για μαρμάρινο γλυπτό που φιλοτεχνήθηκε από τον γλύπτη Αναστάσιο Κρατίδη το 2004 στα πλαίσια του 1ου Συμποσίου Γλυπτικής στα Τρίκαλα, με την επιμέλεια του ΠΟΔΤ [22].

Επίσης, επιλέχθηκαν τα εξής μνημεία με δεδομένα από άλλες ομάδες:

- Μουσείο Ιστορίας της κομητείας San Mateo στο Ρέντγουντ Σίτι, Καλιφόρνια, Ηνωμένες Πολιτείες. Πρώην δικαστικό μέγαρο χτισμένο το 1910.



(Εικόνα 2.1) - Ενδεικτικές φωτογραφίες από κάθε dataset.

2.2 Χρησιμοποιούμενα εργαλεία και λογισμικό

Λογισμικό

Στο πλαίσιο της παρούσας εργασίας θα χρησιμοποιηθεί το ακόλουθο λογισμικό για την εκτέλεση της διαδικασίας τρισδιάστατης ανακατασκευής με τις μεθόδους στις οποίες εστιάζει η έρευνα.

COLMAP για Structure from Motion και Multi View Stereo [23][24]. Open source λογισμικό με δυνατότητα εκτέλεσης Structure from Motion and Multi View Stereo. Το COLMAP αρχικά γράφτηκε απο τον Johannes Schönberger με χρηματική υποστήριξη απο τους επιβλέπων καθηγητές για το PhD του, τους Jan-Michael Frahm και Marc Pollefeys. Η ανάπτυξή του συνεχίζεται ενεργά μέσω της συνεισφοράς της ευρύτερης κοινότητας, με βασικούς συντελεστές τους Schönberger, Paul-Edouard Sarlin και Shaohui Liu. Το

COLMAP διακρίνεται για την ακρίβεια των αποτελεσμάτων του και την ευελιξία που προσφέρει κατά την επιλογή παραμέτρων και ακολουθιών εκτέλεσης. Διαθέτει δυνατότητες εκτέλεσης μέσω γραμμής εντολών καθώς και γραφική διεπαφή.

MicMac για Structure from Motion και Multi View Stereo [25]. Open source λογισμικό με δυνατότητα εκτέλεσης Structure from Motion and Multi View Stereo. Το MicMac δημιουργήθηκε από το Εθνικό Ινστιτούτο Γεωγραφικών και Δασικών Πληροφοριών, την Γαλλική Υπηρεσία Χαρτογράφησης - (IGN) και την Εθνική Σχολή Γεωγραφικών Επιστημών (ENSG) στο εργαστήριο LASTIG.

NVIDIA Instant NeRF για NeRF [26]. Το λογισμικό Instant-NGP επιτρέπει την εκπαίδευση και οπτικοποίηση νευρωνικών αναπαραστάσεων φωτεινότητας (Neural Radiance Fields – NeRF) με εξαιρετικά υψηλή απόδοση, αξιοποιώντας βελτιστοποιήσεις επιπέδου GPU [26]. Αναπτύχθηκε από την NVIDIA και διατίθεται υπό την άδεια Nvidia Source Code License-NC. Το συγκεκριμένο εργαλείο χρησιμοποιείται για τη σύγκριση των σύγχρονων νευρωνικών μεθόδων ανακατασκευής σε σχέση με τις παραδοσιακές γεωμετρικές.

POSTSHOT για Gaussian Splatting [27]. Το PostShot είναι δωρεάν λογισμικό το οποίο χρησιμοποιείται για την υλοποίηση και απεικόνιση της τεχνικής Gaussian Splatting [27]. Αναπτύχθηκε από την εταιρεία Jawset Visual Computing. Η συγκεκριμένη τεχνική βασίζεται στη χρήση τρισδιάστατων γκαουσιανών με ιδιότητες opacity, scale και orientation για την απόδοση της σκηνής, προσφέροντας ταχύτατη απεικόνιση και ρεαλιστικά αποτελέσματα, ιδιαίτερα σε ότι αφορά αντανάκλασεις και ημιδιαφανείς επιφάνειες.

CloudCompare για την σύγκριση των αποτελεσμάτων μεταξύ των διαφορετικών τεχνικών [28]. Είναι επίσης open-source εργαλείο και βασίζεται στο CCLib για την επεξεργασία σημειακών νεφών (point clouds). Διαθέτει δυνατότητες για σύγκριση μοντέλων, υπολογισμό αποστάσεων, μετρικών πυκνότητας και οπτική απεικόνιση σημειακών δεδομένων και πολυγωνικών μοντέλων. Αποτελεί βασικό εργαλείο για την ποσοτική αξιολόγηση της ποιότητας ανακατασκευής και προσφέρει μια κοινή πλατφόρμα απεικόνισης για την ποιοτική αξιολόγηση του μεγαλύτερου συνόλου των αποτελεσμάτων των πειραμάτων.

Πλατφόρμα Εκτέλεσης Πειραμάτων

Τα πειράματα εκτελέστηκαν σε σύστημα με τα εξής χαρακτηριστικά:

- **Λειτουργικό σύστημα:** Windows 11
- **Κεντρική μονάδα επεξεργασίας (CPU):** Intel Core i5 12600KF
- **Κάρτα γραφικών (GPU):** NVIDIA GeForce RTX 4070 (12GB VRAM)
- **Μνήμη RAM:** 32GB DDR4 @ 3600 MHz
- **Μονάδα αποθήκευσης:** SSD Crucial P5 Plus (NVMe)

Όλα τα δεδομένα εισόδου και εξόδου διαβάζονται και αποθηκεύονται τοπικά στον SSD, ώστε να εξασφαλίζεται η μέγιστη δυνατή απόδοση κατά την εκτέλεση των αλγορίθμων και να μειωθούν όσο το δυνατόν περισσότερο τυχόν καθυστερήσεις που οφείλονται σε διαδικασίες ανάγνωσης και εγγραφής στο δίσκο.

2.3 Περιγραφή των αλγορίθμων που θα υλοποιηθούν ή συγκριθούν

Οι αλγόριθμοι που θα συγκριθούν είναι οι εξής:

Structure From Motion (SfM)

Στην εργασία αυτή, η τεχνική SfM εκτελείται από δύο διαφορετικές εφαρμογές, το COLMAP και το MicMac. Παρόλο που πρόκειται για την ίδια τεχνική, ο αλγόριθμος έχει διαφοροποιηθεί σε κάθε εφαρμογή. Μία κλασική εφαρμογή του SfM λειτουργεί με τον ακόλουθο τρόπο: Το πρώτο στάδιο (Feature Extraction) εκτελεί τον αλγόριθμο SIFT στις φωτογραφίες του dataset. Ο SIFT είναι αλγόριθμος μηχανικής όρασης που εντοπίζει σημεία πάνω σε φωτογραφίες με βάση το αποτέλεσμα της αφαίρεσης ενός αντιγράφου της φωτογραφίας με γκαουσιανό φίλτρο θολώματος από τον εαυτό της, χωρίς φίλτρο. Το αποτέλεσμα αυτής της διαδικασίας αποκαλύπτει σημεία πάνω στην φωτογραφία με υψηλή αντίθεση, τα οποία αποθηκεύονται σε μία βάση δεδομένων. Αυτά είναι τα features που θα χρησιμοποιηθούν για τα επόμενα βήματα του SfM. Το αμέσως επόμενο βήμα είναι το Feature Matching, δηλαδή η αντιστοίχιση κοινών features μεταξύ φωτογραφιών. Σε αυτό το βήμα γίνεται εξαντλητική σύγκριση των features όλων των εικόνων και τα κοινά features αντιστοιχίζονται και αποθηκεύονται στην βάση δεδομένων. Επίσης δημιουργούνται και αποθηκεύονται “multi-view tracks”, ευθύγραμμα τμήματα που περιγράφουν την οπτική διαφορά (παράλλαξη) κοινών features μεταξύ εικόνων. Το ζευγάρι εικόνων με τον υψηλότερο αριθμό κοινών features επιλέγεται ως η βάση της

τρισδιάστατης ανακατασκευής και δημιουργείται το σημειακό νέφος με τα points αυτών των δύο εικόνων καθώς και η αρχικοποίηση της θέσης των αντίστοιχων καμερών στον τρισδιάστατο χώρο. Το επόμενο βήμα είναι το Incremental Reconstruction, η σταδιακή επεξεργασία δηλαδή των υπόλοιπων φωτογραφιών του dataset. Αυτό γίνεται συγκρίνοντας την υπάρχουσα βάση με τις υπόλοιπες φωτογραφίες και διαλέγοντας την φωτογραφία με τα περισσότερα κοινά features με τις φωτογραφίες της βάσης ή την φωτογραφία με τα περισσότερα features που τέμνονται από τα multi-view tracks των εικόνων της βάσης. Μετά την επιλογή της νέας φωτογραφίας, εκτελείται το βήμα Triangulation και δημιουργούνται και αποθηκεύονται τα νέα multi-view tracks μεταξύ αυτής και της προηγούμενης, καθώς και τα νέα points που προστίθενται στο σημειακό νέφος και οι αντίστοιχη θέση της κάμερας. Ακολουθεί το βήμα Bundle Adjustment, το οποίο διορθώνει σφάλματα στις θέσεις των καμερών και των points στο χώρο. Η διαδικασία βασίζεται στη σύγκριση της προβολής κάθε point σε κάθε φωτογραφία (μέσω του μοντέλου της αντίστοιχης κάμερας) με τη θέση του αντίστοιχου feature στην φωτογραφία. Σκοπός είναι η ελαχιστοποίηση της διαφοράς μεταξύ αυτών των δύο προβολών, ώστε να επιτευχθεί η βέλτιστη γεωμετρική ευθυγράμμιση του μοντέλου. Τα ζευγάρια προβολών με πολύ μεγάλη απόκλιση αφαιρούνται. Αυτό το βήμα εκτελείται για όλες τις κάμερες και γι αυτό είναι υπολογιστικά ακριβό. Μετά το τέλος του Bundle Adjustment, ο αλγόριθμος επιστρέφει στο βήμα “Incremental Reconstruction” και προσθέτει την επόμενη φωτογραφία. Αν δεν υπάρχει άλλη φωτογραφία τότε ο αλγόριθμος τερματίζει.

Σύμφωνα με την διατριβή του συντάκτη του COLMAP, Johannes Schönberger, το SfM στο COLMAP αποτελεί μια νέα προσέγγιση που σκοπεύει να μπορεί να επεξεργαστεί μεγάλο αριθμό φωτογραφιών εισόδου. Η προσέγγιση αυτή βασίζεται στην αντικατάσταση κάποιων υπολογιστικά ακριβών αλγορίθμων που χρησιμοποιεί το κλασικό SfM με πιο αποδοτικές εναλλακτικές. Αρχικά, στο βήμα Feature Extraction, ο αλγόριθμος SIFT παραμένει αλλά μπορεί πλέον να τρέχει στο GPU αντί του CPU και έτσι είναι ταχύτερος. Στο βήμα Feature Matching, το COLMAP δίνει εναλλακτικές μεθόδους, πέρα της εξαντλητικής αναζήτησης. Δίνεται η επιλογή του διαδοχικού (sequential) matching, το οποίο ψάχνει κοινά features μόνο μεταξύ εικόνων που εισάγονται διαδοχικά στον αλγόριθμο, το οποίο είναι ιδανικό για περιπτώσεις όπου οι φωτογραφίες-είσοδοι είναι διαδοχικά καρέ ενός βίντεο καθώς και η αναζήτηση μεταξύ εικόνων των οποίων οι συντεταγμένες GPS βρίσκονται αρκετά κοντά μεταξύ τους (τα κινητά τηλέφωνα καθώς και πολλές σύγχρονες φωτογραφικές μηχανές μπορούν να ενσωματώσουν δεδομένα GPS στις

φωτογραφίες τους). Επίσης, δίνεται η επιλογή αναζήτησης με βάση ένα vocabulary tree. Μετά την εκτέλεση του SIFT στο προηγούμενο βήμα, κάθε feature αποκτά μια “λεξή”, μια μοναδική τιμή αποτελούμενη από ένα vector 128 διαστάσεων που το χαρακτηρίζει. Η παραγωγή αυτών των τιμών γίνεται με βάση ένα vocabulary tree εκπαιδευμένο σε χιλιάδες φωτογραφίες το οποίο παρέχεται μαζί με το λογισμικό COLMAP ή εναλλακτικά αφήνει τον χρήστη να δημιουργήσει το δικό του. Η αντιστοίχιση εικόνων με βάση τις κοινές λέξεις που περιέχουν είναι πολύ γρηγορότερη σε σχέση με την εξαντλητική αναζήτηση του κλασικού SfM, μπορεί όμως να έχει λιγότερες αντιστοιχίσεις. Στο βήμα Incremental Reconstruction, το COLMAP αποφεύγει την εξαντλητική σύγκριση του αριθμού των features ώστε να επιλέξει την επόμενη εικόνα. Αυτό επιτυγχάνεται με την εισαγωγή μιας νέας λογικής βαθμολόγησης των εικόνων στο βήμα Feature Matching. Όλες οι φωτογραφίες αποκτούν έναν βαθμό ο οποίος πηγάζει από την αξιολόγηση του αριθμού features και πως αυτά είναι κατανομημένα στην φωτογραφία. Η λίστα με τις βαθμολογίες αναδιατάσσεται με τις φωτογραφίες με την μεγαλύτερη βαθμολογία να είναι στην κορυφή της λίστας. Οι φωτογραφίες με βαθμολογία μικρότερη κάποιας βάσης απορρίπτονται. Η εικόνα με την μεγαλύτερη βαθμολογία επιλέγεται ως η επόμενη για την ανακατασκευή και μετά την εισαγωγή και επεξεργασία της, η λίστα βαθμολογιών ανανεώνεται ώστε η νέα γεωμετρία που δημιουργήθηκε να επηρεάζει την επιλογή της αμέσως επόμενης φωτογραφίας. Το βήμα Bundle Adjustment είναι επίσης βελτιστοποιημένο διότι εκτελείται μόνο στις πρόσφατες φωτογραφίες με τα περισσότερα matches αντί για το σύνολο των εικόνων που έχουν εισαχθεί ως τώρα. Επίσης, σε αντίθεση με το κλασικό SfM, το Bundle Adjustment του COLMAP διαιρεί την σκηνή σε μικρότερα κομμάτια, μέσα στα οποία τυχόν κάμερες που βρίσκονται κοντά μεταξύ τους αφομοιώνονται και συμπεριφέρονται σαν μία κάμερα, πράγμα το οποίο μειώνει τον συνολικό αριθμό καμερών και έτσι εκτελείται πιο γρήγορα [31].

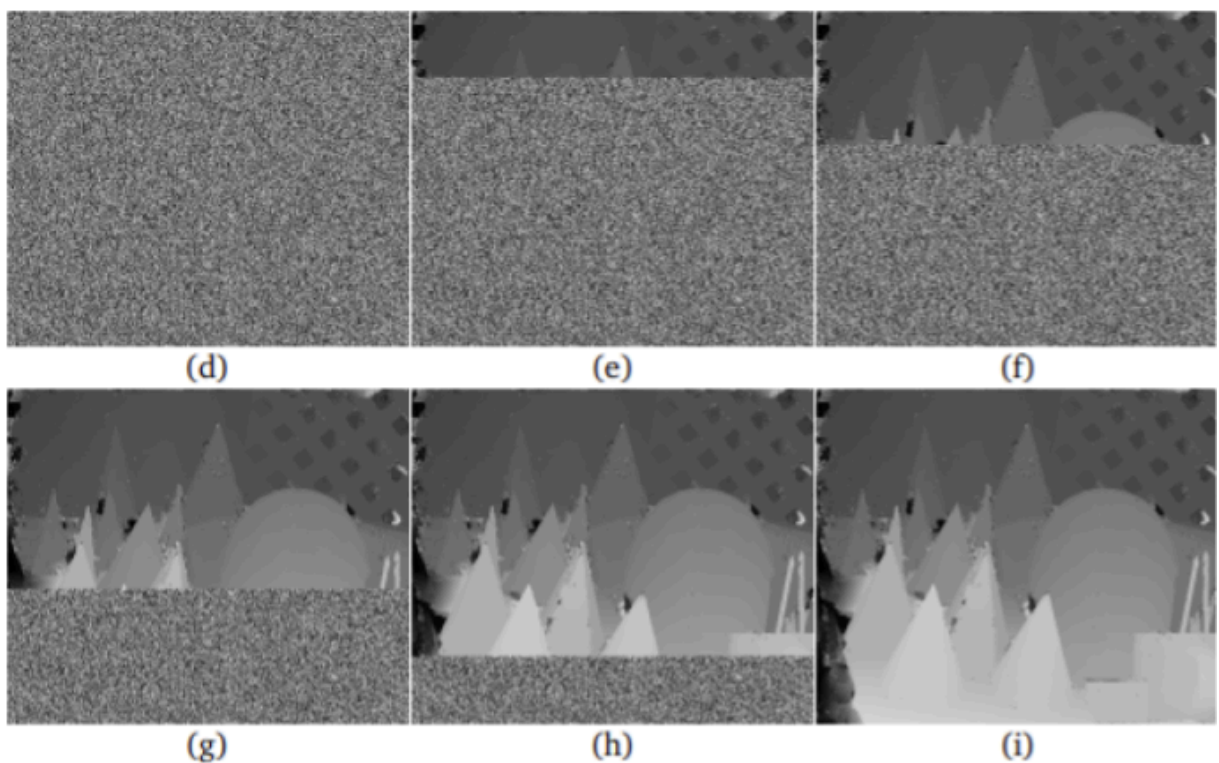
Multi View Stereo (MVS)

Η κλασική μορφή του MVS αρχίζει με την εισαγωγή δεδομένων Feature Extraction και Feature Matching, με παρόμοιο δηλαδή τρόπο με τον αλγόριθμο SfM. Για τον λόγο αυτό, συχνά όταν οι δυο αλγόριθμοι συναντιούνται στο ίδιο λογισμικό, ο MVS δέχεται δεδομένα από τον SfM ως είσοδο. Το πρώτο βήμα του MVS ονομάζεται Seeding και έχει σκοπό την δημιουργία μικρών, ορθογώνιων περιοχών πυκνού σημειακού νέφους τα οποία ονομάζονται patches. Οι περιοχές αυτές επιλέγονται και δημιουργούνται με βάση κάποιο ελάχιστο όριο βεβαιότητας, πρέπει δηλαδή να είναι ορατά από τουλάχιστον N

φωτογραφίες ώστε η γεωμετρική τους θέση στον χώρο να είναι βέβαιη. Η βεβαιότητα της γεωμετρικής θέσης αξιολογείται με τον αλγόριθμο NCC. Το επόμενο βήμα είναι το Expansion, η εξάπλωση του πυκνού σημειακού νέφους από το κέντρο των patches προς τα έξω. Αυτό γίνεται με την εύρεση γειτονικών patch και την εισαγωγή νέων points. Η διαδικασία που αποφασίζει αν τα patch είναι γειτονικά χωρίζει τις φωτογραφίες σε ένα πλέγμα και ύστερα ελέγχει αν τα patches ανήκουν σε γειτονικά κελιά στο πλέγμα και αν η εφαπτομένη των δυο patch είναι παρόμοια. Ο έλεγχος της εφαπτομένης των patches λειτουργεί σαν έλεγχος για το αν τα δυο patches αντιπροσωπεύουν την ίδια επιφάνεια της σκηνής-στόχου. Αν αυτά τα κριτήρια εκπληρώνονται, τότε γίνεται ένωση των δυο patches με την τοποθέτηση νέων patches στο μεσοδιάστημα. Το επόμενο βήμα είναι το Filtering το οποίο έχει σκοπό την περαιτέρω βελτίωση της ανακατασκευής και την αφαίρεση των patches που δεν ανήκουν σε κάποια πραγματική επιφάνεια. Ο πρώτος σκοπός επιτυγχάνεται με την αξιολόγηση της θέσης κάθε patch στον χώρο σε σύγκριση με τα υπόλοιπα γειτονικά του. Αν κάποιο patch αποκρύπτει πολλαπλά άλλα patches (βρίσκεται μεταξύ της πλειοψηφίας των γειτόνων του και της κάμερας) τότε το patch αυτό αφαιρείται. Αντίστοιχα, αν κάποιο patch αποκρύπτεται από την πλειοψηφία των γειτονικών του patches τότε επίσης αφαιρείται. Ο δεύτερος στόχος επιτυγχάνεται με την σύγκριση της ορατότητας κάθε patch από την αντίστοιχη περιοχή από τις θέσεις των φωτογραφιών που το καλύπτουν και την απόρριψη του patch σε περίπτωση που η ορατότητα είναι χαμηλή ή παραμορφωμένη. Η διαδικασία αυτή επαναλαμβάνεται για όλα τα patches μέχρι η σκηνή-στόχος να έχει αποτυπωθεί ικανοποιητικά στο νέφος. [32]

Και εδώ το COLMAP έχει κάποιες διαφορές στην εφαρμογή του αλγορίθμου MVS, κυρίως με στόχο την βελτίωση της ταχύτητας και της λειτουργίας του με dataset μεγάλου αριθμού φωτογραφιών, αλλά τα βήματα παραμένουν ίδια με το “κλασσικό” MVS, δηλαδή Seeding, Expansion και Filtering. Το COLMAP αποφεύγει την αξιολόγηση βάθους όλης της γεωμετρίας της σκηνής και αντί για αυτό εκτελεί αξιολόγηση βάθους για τα εικονοστοιχεία (pixels) της κάθε εικόνας. Η αξιολόγηση γίνεται φωτομετρικά, δηλαδή με τον έλεγχο της ορθότητας του βάθους σε σχέση με την παράλλαξη (με βάση τις άλλες φωτογραφίες που περιγράφουν το σημείο). Συγκεκριμένα, αντί για τις υπολογιστικά ακριβείς μεθόδους του κλασσικού MVS, το COLMAP χρησιμοποιεί PatchMatch Stereo, έναν αλγόριθμο που εκτελεί την αξιολόγηση βάθους της σκηνής με βάση depth maps, ειδικές εικόνες οι οποίες εκφράζουν την απόσταση κάθε εικονοστοιχείου για κάθε φωτογραφία από την κάμερα που τις αντιστοιχεί. Η δημιουργία αυτών των maps αρχίζει

με την θέσπιση ενός διαστήματος τιμών που αντιπροσωπεύουν την ελάχιστη και μέγιστη τιμή βάθους του map. Αυτό μπορεί να υπολογιστεί με βάση το αποτέλεσμα του SfM καθώς και με τα δεδομένα που διαθέτουμε για την κάμερα στην οποία αντιστοιχεί η φωτογραφία (ανάλυση, εστιακό βάθος, κ.α). Αρχικά, κάθε εικονοστοιχείο αποκτά μια τυχαία εκτίμηση για το βάθος και τη διεύθυνση της εφαπτομένης (normal), με τιμές εντός του επιτρεπτού βάθους. Αυτές λειτουργούν ως αρχικές υποθέσεις για τη γεωμετρία της σκηνής. Στη συνέχεια, το βάθος και η εφαπτομένη κάθε εικονοστοιχείου βελτιώνονται σταδιακά μέσω μιας διαδικασίας διάδοσης και τυχαίας δοκιμής (propagation and random refinement), επιλέγοντας πάντα την εκτίμηση που εμφανίζει τη μέγιστη φωτομετρική συνέπεια με τις υπόλοιπες φωτογραφίες. Η διαδικασία αυτή μπορεί να εκτελεστεί πάνω από μία φορά για την διόρθωση τυχόν λαθών. Όταν εντοπιστεί η βέλτιστη εκτίμηση για το βάθος και την εφαπτομένη (normal) ενός εικονοστοιχείου, η τιμή του βάθους καταγράφεται στον αντίστοιχο depth map της εικόνας, ενώ το διάνυσμα της εφαπτομένης αποθηκεύεται στον αντίστοιχο normal map [33], [34].



(Εικόνα 2.2) - Βηματική απεικόνιση της παραγωγής depth maps από αρχική τυχαία εκτίμηση
(Πηγή: PatchMatch Algorithms for Motion Estimation and Stereo Reconstruction [35])

Στο βήμα Expansion, το COLMAP συγκρίνει depth και normal maps μεταξύ γειτονικών pixels που αντιστοιχούν στην ίδια επιφάνεια και προσαρμόζει τις τιμές αντίστοιχα για καλύτερη συνέχεια μεταξύ τους. Αυτό βελτιώνει την ομοιομορφία των επιφανειών της

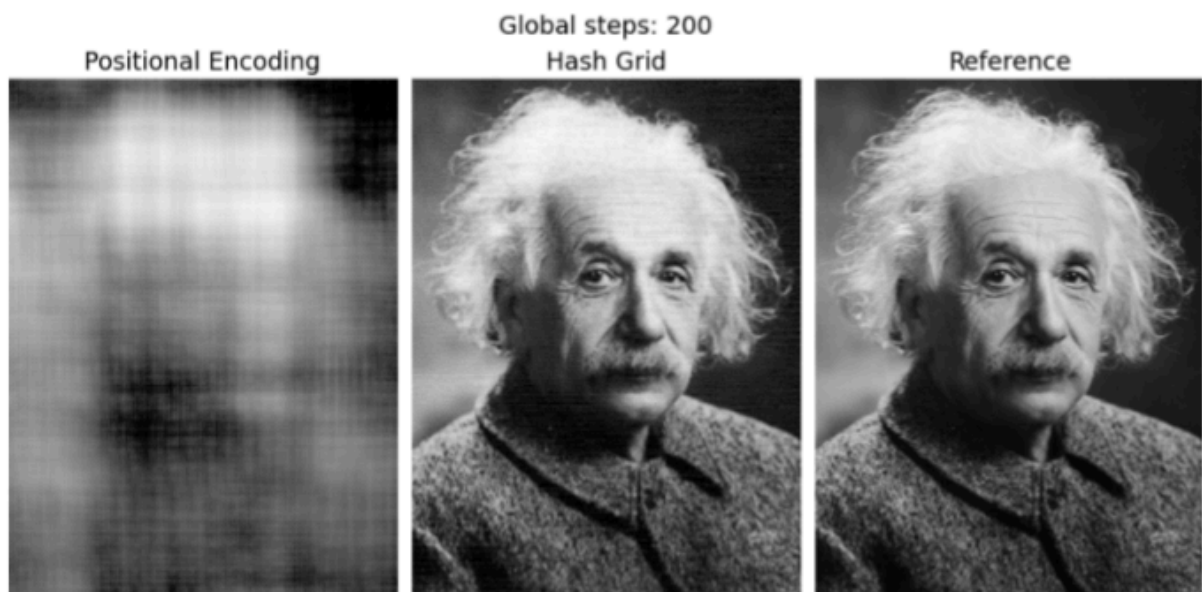
ανακατασκευής, αν και μπορεί να προκαλέσει απώλεια μικρών γεωμετρικών λεπτομερειών. Η τελική ανακατασκευή αποτελείται από το σύνολο των points τοποθετημένα στον χώρο με οδηγό τις πληροφορίες που περιέχονται στα depth και τα normal maps.

NeRF

Ο αλγόριθμος NeRF περιγράφεται ως εξής από τους δημιουργούς του: “Ο αλγόριθμός μας αναπαριστά μια σκηνή χρησιμοποιώντας ένα πλήρως συνδεδεμένο (μη συνελικτικό) βαθύ νευρωνικό δίκτυο, MLP, του οποίου η είσοδος είναι μια συνεχής πενταδιάστατη συντεταγμένη (χωρική θέση (x, y, z) και κατεύθυνση θέασης (θ, ϕ)) και η έξοδος είναι η πυκνότητα όγκου και η εξαρτώμενη από την οπτική γωνία εκπεμπόμενη ακτινοβολία σε εκείνη τη χωρική θέση (radiance). Συνθέτουμε όψεις ερωτώντας πενταδιάστατες συντεταγμένες κατά μήκος των ακτίνων της κάμερας και χρησιμοποιούμε κλασικές τεχνικές απεικόνισης όγκου για να προβάλλουμε τα χρώματα και τις πυκνότητες εξόδου σε μια εικόνα [36].” Πρόκειται λοιπόν για μια εντελώς διαφορετική προσέγγιση στον τρόπο ανακατασκευής και απεικόνισης μιας τρισδιάστατης σκηνής που δεν χρησιμοποιεί σημειακό νέφος ή πολυγωνα για την αναπαράσταση της γεωμετρίας. Το αποτέλεσμα είναι η ικανότητα ρεαλιστικής αναπαράστασης όχι μόνο από τις οπτικές γωνίες που καλύπτονται από το dataset της εισόδου αλλά και νέες γωνίες (εντός ορίων), παρόμοια με την ελευθερία της επιλογής γωνίας θέασης που παρέχει ένα κλασικό τρισδιάστατο μοντέλο. Για την υλοποίηση ενός NeRF απαιτείται η χρήση ενός dataset από φωτογραφίες με γνωστές τις θέσεις και τα χαρακτηριστικά της κάμερας. Για τον λόγο αυτό χρησιμοποιείται ο αλγόριθμος SfM που περιγράφεται νωρίτερα. Με βάση τις φωτογραφίες εισόδου και τις γνωστές πλέον θέσεις από τις οποίες λήφθηκαν, το MLP εκπαιδεύεται έτσι ώστε το οπτικό αποτέλεσμα της αναπαράστασης να είναι όσο το δυνατόν πιο όμοιο στην φωτογραφία-είσοδο, όταν η σκηνή παρατηρείται από την ίδια οπτική γωνία. Άλλες έρευνες πάνω στα MLP έχουν αποδείξει ότι τα MLP λειτουργούν καλύτερα (bias) στην μάθηση λεπτομέριας χαμηλής συχνότητας και όχι υψηλής συχνότητας. Αυτό αποτελεί μειονέκτημα για την χρήση τους στην αναπαράσταση τρισδιάστατων σκηνών διότι ραγδαίες αλλαγές στην γεωμετρία της σκηνής αποτελούν εκφράσεις υψηλής συχνότητας οι οποίες, λόγω του bias, θα χαθούν. Για να αντιμετωπίσει αυτό το πρόβλημα, το NeRF εφαρμόζει χωρική κωδικοποίηση (positional encoding) στις εισόδους του, μια μέθοδο που μετασχηματίζει τις αρχικές συντεταγμένες χρησιμοποιώντας ημιτονοειδείς και συνημιτονοειδείς συναρτήσεις σε πολλαπλές συχνότητες. Αυτό επεκτείνει τον χώρο των

εισόδων και βοηθά το MLP να μοντελοποιεί ευκολότερα λεπτές γεωμετρικές και φωτομετρικές λεπτομέρειες, οι οποίες είναι κρίσιμες για τη ρεαλιστική απόδοση μιας σκηνής, με καθαρότερες υφές και αλλαγές στην γεωμετρία. Μετά την εκπαίδευση, η απόδοση μιας εικόνας (rendering) γίνεται μέσω εκπομπής ακτίνων από το εικονικό κέντρο της κάμερας, μία ανά pixel. Κάθε ακτίνα διασχίζει τη σκηνή και δειγματοληπτεί πολλαπλά σημεία στο χώρο, στα οποία το MLP προβλέπει χρώμα και πυκνότητα. Στη συνέχεια, με βάση ένα μοντέλο ογκομετρικής σύνθεσης (volume rendering), τα δεδομένα αυτά συνδυάζονται ώστε να παραχθεί η τελική τιμή χρώματος για κάθε pixel που συνολικά συνθέτουν την εικόνα.

Στην παρούσα εργασία, η εφαρμογή των NeRF γίνεται μέσω του αλγορίθμου Instant NERF της Nvidia, ο οποίος εφαρμόζει αλλαγές σε σχέση με το κλασσικό NeRF κυρίως προς όφελος της ταχύτερης εκπαίδευσης και απεικόνισης.



(Εικόνα 2.3) - Σύγκριση positional encoding (ημιτονοειδής και συνημιτονοειδής είσοδος), hash grid και εικόνα αναφοράς. Το οπτικό αποτέλεσμα στα 200 βήματα είναι σχεδόν πανομοιότυπο με την εικόνα αναφοράς στο hash grid, ενώ το positional encoding έχει μόλις αρχίσει να δημιουργεί μια αναγνωρίσιμη φιγούρα.
(Πηγή: Hash Grid Encoding - Github [38])

Συγκεκριμένα, η μέθοδος Instant NERF αντικαθιστά την ημιτονοειδή και συνημιτονοειδή είσοδο της συνάρτησης με ένα hash-based grid encoding πολλαπλής ανάλυσης με lookup tables, μέσα στο οποίο η ανάκτηση χαρακτηριστικών είναι πολύ γρηγορότερη, ιδιαίτερα όταν συνδυάζεται με την επεξεργαστική δύναμη σύγχρονων GPU [37], [38]. Λόγο του ότι το MLP έχει να διαχειριστεί μικρότερο όγκο πληροφοριών και αυτές οι πληροφορίες είναι πιο ευκολα και γρήγορα προσβάσιμες, η πολυπλοκότητα του MLP είναι πολύ μικρότερη στην εφαρμογή Instant NERF σε σχέση με το παραδοσιακό NERF. Ως αποτέλεσμα, η

εκπαίδευσή του για ένα νέο dataset είναι πολύ γρηγορότερη. Η Nvidia ισχυρίζεται ότι το Instant NERF είναι πάνω από 1000 φορές γρηγορότερο από το παραδοσιακό NERF [39].

Gaussian Splatting

Ο αλγόριθμος Gaussian Splatting αποτελεί μια διαφορετική προσέγγιση στην απεικόνιση μιας τρισδιάστατης σκηνής με πυκνότητα όγκου και εξαρτώμενη από την οπτική γωνία εκπεμπόμενη ακτινοβολία σε εκείνη τη χωρική θέση (radiance). Δηλαδή, επιδιώκει να έχει παρόμοια αποτελέσματα με τον αλγόριθμο NeRF αλλά με σημαντικές βελτιώσεις στην ταχύτητα εκτέλεσης και την απαίτηση υλικού. Η κύρια διαφορά είναι η χρήση τρισδιάστατων γκαουσιανών στοιχείων (3D Gaussians) για την έκφραση της σκηνής.

Ο αλγόριθμος αρχικοποιείται με πληροφορίες που προέρχονται από τον αλγόριθμο Structure-from-Motion (SfM), όπως οι θέσεις και οι παράμετροι των καμερών, καθώς και ένα αραιό νέφος (sparse point cloud). Σε αντίθεση με το NeRF που λειτουργεί απευθείας σε συνεχείς συντεταγμένες, εδώ το νέφος εκφράζεται ως σύνολο 3D Gaussians, τα οποία τοποθετούνται στο χώρο με τέτοιο τρόπο ώστε να καλύπτουν το μεσοδιάστημα μεταξύ των αρχικών σημείων του νέφους. Η επόμενη φάση είναι η εκπαίδευση ή βελτιστοποίηση της σκηνής, κατά την οποία τα 3D Gaussians παραμορφώνονται, μεταβάλλοντας τη θέση, το μέγεθος, το σχήμα, το χρώμα και την αδιαφάνειά τους, ώστε να γεμίσουν τα κενά του χώρου και να αναπαραστήσουν ρεαλιστικά την όψη της σκηνής. Η διαδικασία βελτιστοποίησης βασίζεται στη σύγκριση του οπτικού αποτελέσματος που παράγεται από το σύνολο των Gaussians με τις φωτογραφίες εισόδου, προσαρμόζοντας τις παραμέτρους ώστε να ελαχιστοποιηθεί το σφάλμα απόδοσης. Ακολουθώντας αυτή την διαδικασία, το Gaussian Splatting δεν απαιτεί χρήση MLP και της χρονοβόρας εκπαίδευσής του.

Η διαδικασία αυτή δίνει σε κάθε 3D gaussian τις εξής παραμέτρους:

- Θέση: Που βρίσκεται στον χώρο της σκηνής (x, y, z)
- Συνδιακύμανση: πώς παραμορφώθηκε η σφαίρα (3×3 matrix)
- Χρώμα: Τι χρώμα θα έχει αυτό το 3D gaussian (RGB)
- Αδιαφάνεια (Alpha): Πόσο διάφανο είναι το 3D gaussian (α)

Η λογική της βελτιστοποίησης δουλεύει ως εξής:

- Αν το 3D gaussian είναι πολύ μεγάλο, να γίνει διαίρεση του σε δύο μικρότερα.

- Αν το 3D gaussian είναι πολύ μικρό, να γίνει αντιγραφή του ώστε να γίνει καλύτερη αναπαράσταση της γεωμετρίας.
- Αν το Alpha (α) ενός 3D gaussian είναι πολύ μικρό, να γίνει διαγραφή του ώστε να μην επιβαρύνει το μοντέλο.

Η απόδοση εικόνας εξόδου (rendering) αποτελεί άλλο βήμα στο οποίο παρατηρείται μεγάλο κέρδος στην αποδοτικότητα του Gaussian Splatting σε σύγκριση με το NERF. Αυτό πραγματοποιείται μέσω rasterization, μια κλασική μέθοδο απόδοσης τρισδιάστατων γραφικών σε δισδιάστατη εικόνα. Τα Gaussians προβάλλονται στο δισδιάστατο επίπεδο της εικόνας από το σημείο θέασης της εικονικής κάμερας. Αυτή η προβολή μετασχηματίζει κάθε τρισδιάστατο ελλειψοειδές Gaussian σε μια δισδιάστατη έλλειψη στην οθόνη. Η διαδικασία αυτή είναι πολύ λιγότερο υπολογιστικά ακριβή σε σύγκριση με το volume rendering που συναντάνται στα NERF, επιτρέποντας την άμεση και αποδοτική απεικόνιση της σκηνής [40].

2.4 Κριτήρια αξιολόγησης και μέθοδοι σύγκρισης

Ποσοτική Σύγκριση:

Η ποσοτική σύγκριση έχει σκοπό την συλλογή και την αξιολόγηση αριθμητικών δεδομένων για τους αλγόριθμους και το αποτέλεσμά τους, όπως ο χρόνος εκτέλεσης, ο αριθμός points στο σημειακό νέφος του αποτελέσματος, την πυκνότητα του νέφους και άλλα.

Ποιοτική Σύγκριση:

Η ποιοτική σύγκριση έχει σκοπό να αξιολογηθεί οπτικά η ομοιότητα της ανακατασκευής με την σκηνή-στόχο, συγκρίνοντας το οπτικό αποτέλεσμα κάθε αλγορίθμου με τις φωτογραφίες-εισόδους του dataset. Θα αξιολογηθεί η γεωμετρική λεπτομέρεια και η απόδοση της υφής των αντικειμένων στην σκηνή-στόχο. Επίσης, θα σημειωθούν τυχόν λάθη, ελλείψεις και κενά στην ανακατασκευή.

3. Πειραματική διαδικασία και υλοποίηση

3.1 Προετοιμασία δεδομένων

Η πειραματική διαδικασία αρχίζει με την συλλογή των φωτογραφιών που θα αποτελέσουν τα dataset που θα εκτελεστούν. Έγινε επίσκεψη των χώρων και φωτογράφιση των σκηνών-στόχων της ανακατασκευής ή λήψη φωτογραφιών από άλλη πηγή, με την τήρηση, όσο το δυνατόν καλύτερα, των κριτηρίων που περιγράφονται στην ενότητα 2.1. Σε περίπτωση που υπάρχουν φωτογραφίες διαφορετικών πηγών ή ρυθμίσεων σε κάποιο dataset, θα υπάρξει σχετική αναφορά στην αντίστοιχη υποενότητα. Κάθε dataset θα ονομαστεί με ένα όνομα που περιγράφει την σκηνή-στόχο που αφορά.

ΤΙΤΛΟΣ	ΠΕΡΙΓΡΑΦΗ	ΠΛΗΘΟΣ ΦΩΤΟΓΡΑΦΙΩΝ
TZAMI	Μαυσωλείο Οσμάν Σαχ	99 Φωτογραφίες
PELINNA	Αρχαία Πέλινα	1539 Φωτογραφίες
TRIKKI	Γλυπτό ΤΡΙΚΚΗ	22 Φωτογραφίες
COURT	Μουσείο Ιστορίας της κομητείας San Mateo	1106 Φωτογραφίες + LIDAR

Πίνακας 3.1 - Στοιχεία των dataset

3.2 Ανακατασκευή με διαφορετικούς αλγορίθμους

Έγινε επεξεργασία και των τεσσάρων dataset όπου αυτό ήταν δυνατό από τα λογισμικά και τους αντίστοιχους αλγορίθμους. Ξεκινώντας με το dataset “COURT”, το οποίο περιλαμβάνει ground truth με την μέθοδο LIDAR, έγινε εκτέλεση των αλγορίθμων Structure From Motion, Multi View Stereo, Gaussian Splatting και NERF. Η εκτέλεση SfM και MVS έγινε μόνο μέσω COLMAP διότι το MicMac απέτυχε. Το dataset “TZAMI” υποβλήθηκε σε επεξεργασία μέσω όλων των λογισμικών και αλγορίθμων, συμπεριλαμβανομένου και SfM και MVS μέσω MicMac. Αυτό επιτρέπει την σύγκριση της αποτελεσματικότητας των ίδιων αλγορίθμων από δύο διαφορετικές εφαρμογές. Αντίστοιχες εκτελέσεις έγιναν και στο dataset TRIKKI, με όλους του αλγόριθμους να το επεξεργάζονται επιτυχώς μέσω όλων των εφαρμογών. Τέλος, το dataset PELINNA το οποίο ήταν και το μεγαλύτερο εκτελέστηκε σε SfM και MVS μέσω COLMAP, ενώ το

MicMac και NERF απέτυχαν να ολοκληρώσουν. Οι εικόνες των αποτελεσμάτων βρίσκονται στο ΠΑΡΑΡΤΗΜΑ Α και στην ανάλυση των αποτελεσμάτων στο κεφάλαιο 4.

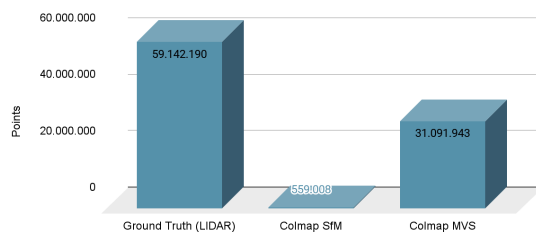
3.3 Σύγκριση και αξιολόγηση αποτελεσμάτων

Μετά την εκτέλεση των αλγορίθμων, τα αντίστοιχα σημειακά νέφη ευθυγραμμίστηκαν στο λογισμικό CloudCompare για ποσοτική αξιολόγηση. Για την ποιοτική αξιολόγηση, τα αποτελέσματα οπτικοποιούνται στο CloudCompare για όλα τα σημειακά νέφη και στο αντίστοιχο viewer για τα gaussian splats και NERF.

COURT - Ποσοτική αξιολόγηση με ground truth

Διάγραμμα 3.1 - COURT - Συνολικός αριθμός points στο point cloud

Υψηλότερες τιμές συνήθως υποδεικνύουν μεγαλύτερη ακρίβεια

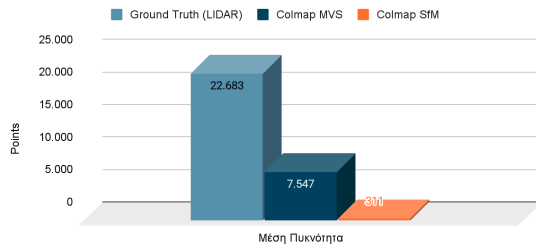


Η ύπαρξη ground truth για το συγκεκριμένο dataset μας δίνει την δυνατότητα σύγκρισης της ακρίβειας των παθητικών αλγορίθμων τρισδιάστατης ανακατασκευής (SfM και MVS) με την ενεργή μέθοδο LIDAR. Το αποτέλεσμα του

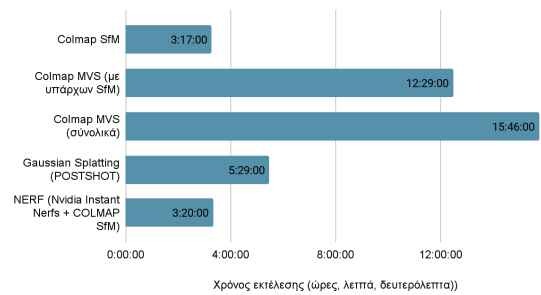
LIDAR θα θεωρηθεί ως η αξιόπιστη βάση πάνω στην οποία θα γίνουν συγκρίσεις. Η αξιοπιστία του LIDAR είναι άμεσα προφανής όταν συγκρίνουμε τον αριθμό των points από τα οποία αποτελείται κάθε σημειακό νέφος που παρήγαγε κάθε αλγόριθμος. Το LIDAR αποτελείται από 59.142.190 points, με μέση πυκνότητα 22.683 σημεία ανα μονάδα επιφάνειας. Αντίστοιχα, το σημειακό νέφος από το SfM αποτελείται από 559.008 σημεία, με πυκνότητα 310 σημεία ανά μονάδα επιφάνειας, ενώ το MVS παρήγαγε σημειακό νέφος που αποτελείται από 31.091.943 points και έχει πυκνότητα 7.547 σημεία ανά μονάδα επιφάνειας. Όσον αφορά τον χρόνο εκτέλεσης των αλγορίθμων, το SfM εκτελέστηκε σε 3 ώρες και 17 λεπτά. Το MVS εκτελέστηκε σε 12 ώρες και 29 λεπτά με χρήση του camera solving από το SfM. Για την πλήρη εκτέλεση του MVS χρειάστηκαν δηλαδή 15 ώρες και 46 λεπτά. Το gaussian splatting και NERF χρειάστηκαν 5 ώρες και 29 λεπτά και 3 ώρες και 20 λεπτά αντίστοιχα για την εκτέλεσή τους, με τον χρόνο εκτέλεσης του NERF να περιλαμβάνει τον χρόνο εκτέλεσης SfM από το COLMAP του οποίου τα αποτελέσματα χρησιμοποιεί το NERF ως είσοδο.

Διάγραμμα 3.2 - COURT - Μέση πυκνότητα ανά μονάδα επιφάνειας

Υψηλότερες τιμές συνήθως υποδεικνύουν μεγαλύτερη ακρίβεια



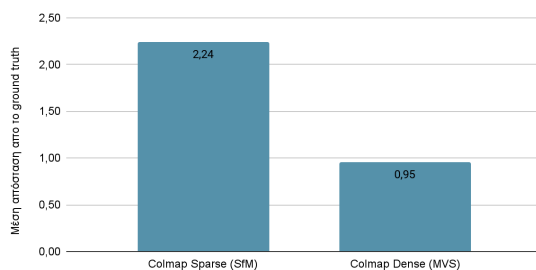
Διάγραμμα 3.3 - COURT - Χρόνος εκτέλεσης



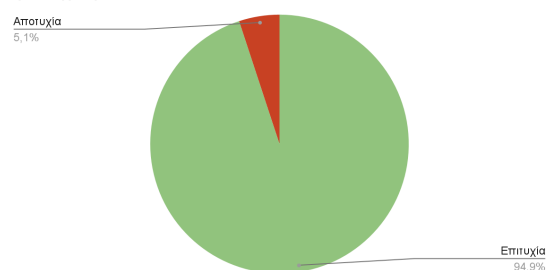
Όσον αφορά την ακρίβεια της ανακατασκευής, πέραν την απλής καταμέτρησης των σημείων, υπολογίζουμε πόσο πιστά είναι τα σημειακά νέφη που προκύπτουν από τις μεθόδους SfM και MVS σε σχέση με το ground truth του LIDAR, μετρώντας την απόσταση μεταξύ του ground truth και του σημειακού νέφους μετά την ευθυγράμμιση και την επικάλυψη των δύο. Ένα μικρό περιθώριο σφάλματος πρέπει να ληφθεί υπόψη σε αυτά τα δεδομένα, καθώς η ευθυγράμμιση δεν μπορεί να είναι απόλυτη και διότι διάφορα αντικείμενα στην σκηνή τα οποία μετακινήθηκαν καθώς και αντικείμενα τα οποία δεν περιλαμβάνονται στην σκηνή του ground truth μετά την επεξεργασία του (για παράδειγμα αφαιρέθηκε η βλάστηση). Τέλος, γίνεται μέτρηση του ποσοστού χρήσης των φωτογραφιών του dataset. Ορίζεται ως ο λόγος του αριθμού των εικόνων που ενσωματώθηκαν επιτυχώς στο τελικό μοντέλο, προς το συνολικό πλήθος των εικόνων του dataset. Το ποσοστό αυτό αποτελεί ένδειξη της ικανότητας του αλγορίθμου να αξιοποιεί πλήρως την διαθέσιμη πληροφορία, κάτι που επηρεάζει άμεσα την πληρότητα της ανακατασκευής. Η μέτρηση αυτή αφορά μόνο το SfM. Απο τις 1106 εικόνες, ο αλγόριθμος κατάφερε να ενσωματώσει επιτυχώς 1050, δηλαδή ποσοστό 94.9%.

Διάγραμμα 3.4 - COURT - Μέση απόσταση από το ground truth

Χαμηλότερες τιμές συνήθως υποδεικνύουν μεγαλύτερη ακρίβεια



Διάγραμμα 3.5 - COURT - Ποσοστό επιτυχημένης χρήσης των φωτογραφιών του dataset



COURT - Ποιοτική αξιολόγηση



(Εικόνα 3.1) - COURT - LIDAR

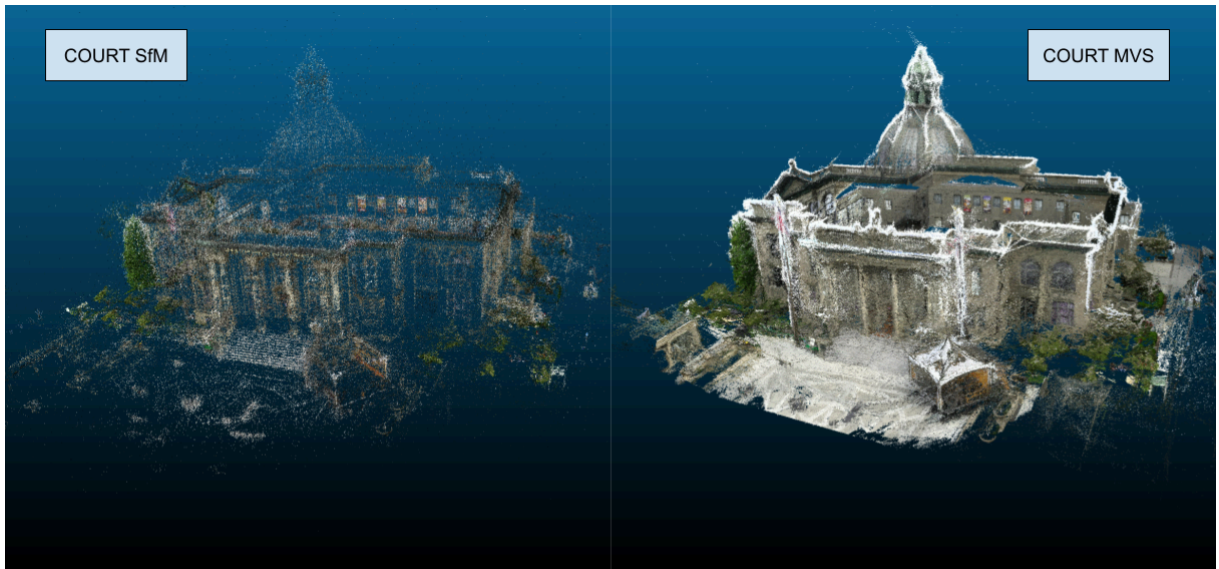
Στην οπτική σύγκριση, το LIDAR συνεχίζει να υπερσχύει των παθητικών μεθόδων τρισδιάστατης ανακατασκευής (εικόνα 3.1). Το σημειακό νέφος είναι πυκνό ακόμα και σε μεγάλη απόσταση από το σημείο που τοποθετήθηκε η συσκευή 3D scanner. Αντίστοιχα υψηλή είναι και η γεωμετρική λεπτομέρεια. Μικρά αντικείμενα όπως οι βάσεις του φωτισμού στην είσοδο του κτιρίου έχουν ανακατασκευαστεί με ακρίβεια. Η απόχρωση

του κτιρίου δείχνει να είναι διαφορετική από τις φωτογραφίες του dataset αλλά αυτό πιθανόν οφείλεται στις διαφορετικές καιρικές συνθήκες και την ώρα της ημέρας που έγιναν οι λήψεις (εικόνα 3.2). Ο τρούλος και η οροφή του κτιρίου είναι ημιτελή, όπως είναι αναμενόμενο, λόγω της έλλειψης οπτικής επαφής από το έδαφος.



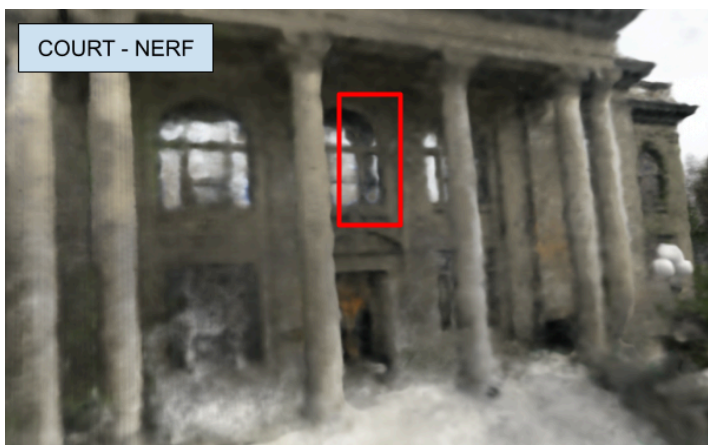
(Εικόνα 3.2) - COURT - Σύγκριση LIDAR και εικόνας εισόδου

Στους παθητικούς αλγορίθμους SfM και MVS, το SfM είχε ως αποτέλεσμα σημειακό νέφος με καλή κάλυψη όλων των περιοχών που καλύπτονται από το dataset φωτογραφιών εισόδου. Η πυκνότητα είναι χαμηλή, όπως είναι αναμενόμενο από την μέθοδο SfM. Πρέπει να σημειωθεί η ύπαρξη points που δεν φαίνεται αν ανήκουν σε κάποια επιφάνεια της σκηνής στόχου.



(Εικόνα 3.3) - COURT - Σύγκριση SfM και MVS

Αυτό ίσως οφείλεται στην ύπαρξη θορύβου (noise) στις φωτογραφίες του dataset, πιθανών από το γεγονός ότι η προέλευσή τους είναι η απομόνωση καρέ από βίντεο. Ο θόρυβος αυτός φαίνεται να μεταφράζεται και στο βήμα MVS (εικόνα 3.3 - τα λευκά points γύρω από το κτίριο). Το σημειακό νέφος του MVS αποτυπώνει την σκηνή σε πολύ καλό βαθμό και καταφέρνει να ανακατασκευάσει ακόμα και μικρές λεπτομέρειες όπως οι βάσεις του φωτισμού της εισόδου του κτιρίου, όπως δηλαδή και το LIDAR. Και οι τρεις μέθοδοι (Lidar, SfM, MVS) αντιμετωπίζουν δυσκολία με τις αντανάκλασεις στα παράθυρα του κτιρίου. Ο αλγόριθμος NERF κατάφερε να ανακατασκευάσει την σκηνή σε πολύ βασικό επίπεδο. Το σχήμα και γενικά η διάταξη της σκηνής ήταν επιτυχής αλλά υπάρχει σοβαρή έλλειψη μικρών λεπτομερειών και ρεαλιστικών υφών (texture). Τα αντικείμενα είναι θολά και απλοποιημένα. Αξιοσημείωτη είναι όμως η παρουσία αντανάκλασεων στα παράθυρα του κτιρίου (εικόνα 3.4).



(Εικόνα 3.4) - COURT - NERF με Nvidia Instant NERFs

Το Gaussian Splatting έδωσε πολύ καλύτερα αποτελέσματα, με καθαρές υφές και γεωμετρία. Υπάρχουν όμως κάποιες λεπτομέρειες που λείπουν και άλλες που είναι παραμορφωμένες λόγω της φύσης των gaussians που αποτελούν την σκηνή. Έντονη είναι η ύπαρξη gaussians που φαίνεται να μην ανήκουν σε κάποια επιφάνεια, αν και η θέση τους είναι περιορισμένη στα σημεία της σκηνής μακριά από το κτίριο και έτσι, πέρα από κάποια προβλήματα ορατότητας, δεν αποτελούν πρόβλημα. Για άλλη μια φορά σημειώνεται η ύπαρξη αντανakλάσεων, αλλά όπως και με το NERF, η γεωμετρική επιφάνεια του παραθύρου δεν είναι ακριβής (εικόνα 3.5).



(Εικόνα 3.5) - COURT - Gaussian Splatting με POSTSHOT

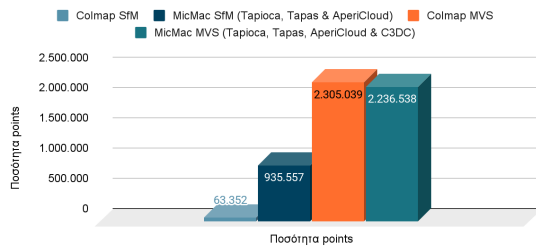
TZAMI - Ποσοτική αξιολόγηση

Σε αυτό το dataset έχουμε εκτέλεση των SfM και MVS μέσω δυο διαφορετικών εφαρμογών (COLMAP και MicMac). Ενδιαφέρον αποτελεί η σύγκριση των αντίστοιχων σημειακών νεφών του κάθε βήματος. Το COLMAP SfM παρήγαγε σημειακό νέφος με 63.352 points με χρόνο εκτέλεσης 2 λεπτά και μέση πυκνότητα ανά μονάδα επιφάνειας 174 σημεία ενώ το MicMac SfM (Tapioca, Tapas & AperiCloud) παρήγαγε ένα σημειακό νέφος με 935.557 points με χρόνο εκτέλεσης 1 ώρα και 7 λεπτά και μέση πυκνότητα ανά μονάδα επιφάνειας 969 σημεία. Η μεγάλη διαφορά στο χρόνο εκτέλεσης φαίνεται να πηγάζει από την προτίμηση του MicMac να δημιουργεί ένα σχετικά πυκνό σημειακό νέφος ακόμα και στο βήμα SfM. Στο βήμα MVS, το COLMAP παρήγαγε σημειακό νέφος με 2.305.039 points με χρόνο εκτέλεσης 47 λεπτά και μέση πυκνότητα ανά μονάδα επιφάνειας 2.902 σημεία, ενώ το MicMac (Tapioca, Tapas, AperiCloud & C3DC)

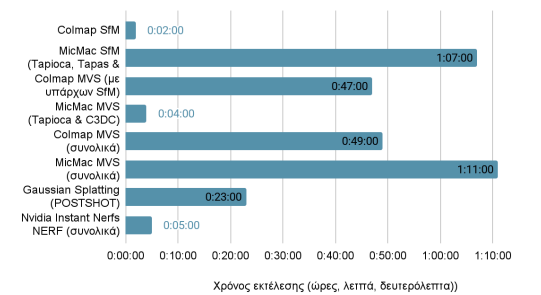
παρήγαγε σημειακό νέφος με 2.236.538 points με χρόνο εκτέλεσης 1 ώρα και 11 λεπτά και μέση πυκνότητα ανά μονάδα επιφάνειας 3.796 σημεία. Το gaussian splatting εκτελέστηκε σε 23 λεπτά και το NERF σε 5 λεπτά, 2 απο τα οποία είναι η εκτέλεση του SfM μέσω COLMAP.

Διάγραμμα 3.6 - TZAMI - Συνολικός αριθμός points στο point cloud

Υψηλότερες τιμές συνήθως υποδεικνύουν μεγαλύτερη ακρίβεια



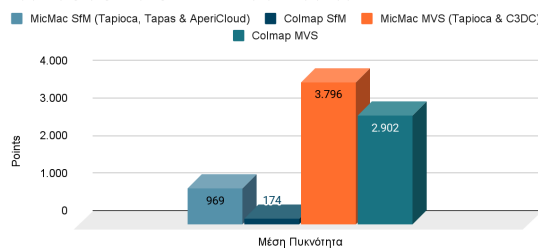
Διάγραμμα 3.7 - TZAMI - Χρόνος εκτέλεσης



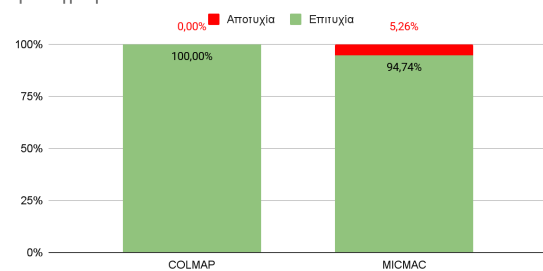
Απο τις 95 φωτογραφίες του dataset, το COLMAP ενσωμάτωσε και τις 95 (100%) ενώ το MicMac ενσωμάτωσε 90 απο τις 95 (94,74%).

Διάγραμμα 3.8 - TZAMI - Μέση πυκνότητα ανά μονάδα επιφάνειας

Υψηλότερες τιμές συνήθως υποδεικνύουν μεγαλύτερη ακρίβεια

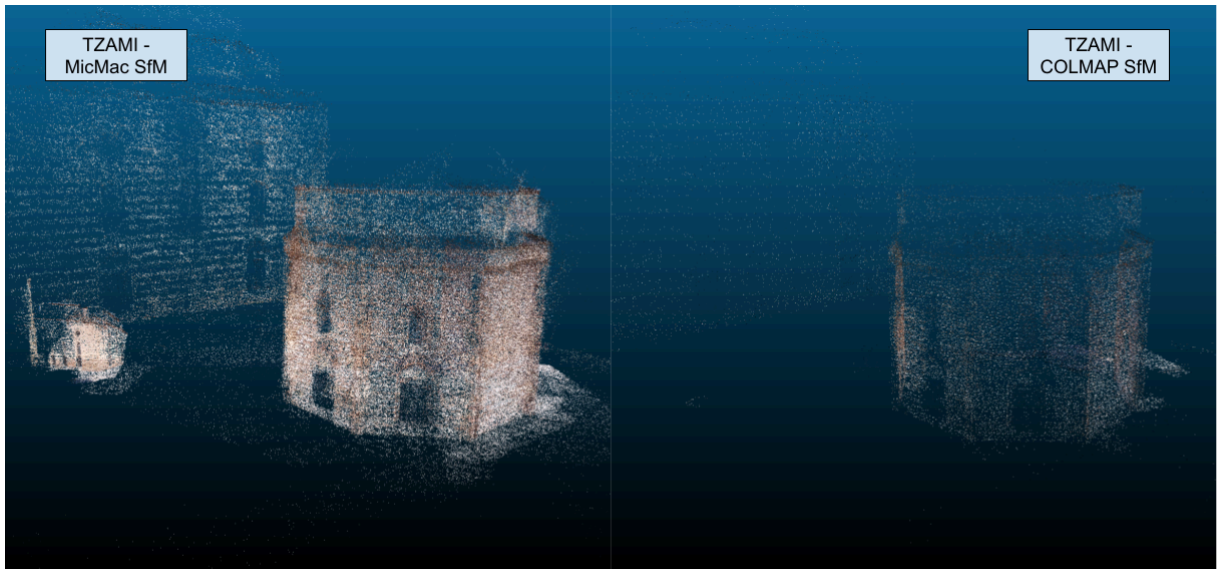


Διάγραμμα 3.9 - TZAMI - Ποσοστό επιτυχημένης χρήσης των φωτογραφιών του dataset



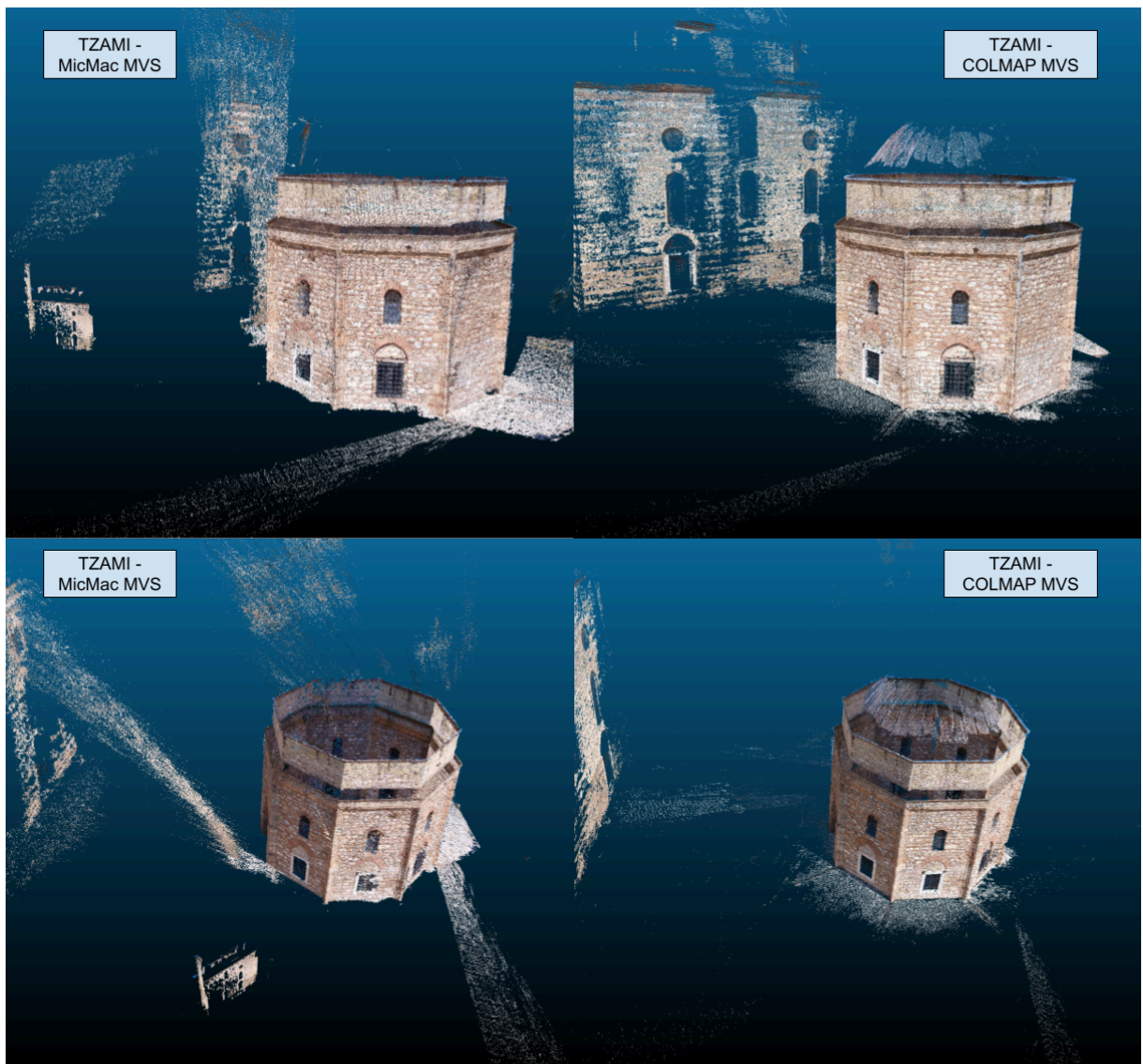
TZAMI - Ποιοτική αξιολόγηση

Στην οπτική σύγκριση, το MicMac παρήγαγε σημειακό νέφος με μεγάλη πυκνότητα ακόμα και στο βήμα SfM, ενώ το COLMAP έχει πυκνότητα πιο χαρακτηριστική για ένα σημειακό νέφος της μεθόδου SfM. Παράλληλα, το σημειακό νέφος του COLMAP είναι καθαρό σε σχέση με αυτό του MicMac, με τα points να ακολουθούν πιο αυστηρά την γεωμετρία της σκηνής. Στο νέφος του MicMac παρατηρείται ένα παραμορφωμένο αντίγραφο της ανακατασκευής του κτιρίου δίπλα από το κτίριο-στόχο, ενώ το COLMAP έχει μόνο την σωστή γεωμετρία (εικόνα 3.6). Στην σύγκριση των αποτελεσμάτων του βήματος MVS, το MicMac καθάρισε αρκετά την θορυβώδη όψη της σκηνής, παρολα αυτα διατήρησε το παραμορφωμένο αντίγραφο του δεύτερου κτιρίου, ενώ η γεωμετρία του μη παραμορφωμένου αντιγράφου έχει αφαιρεθεί.



(Εικόνα 3.6) - TZAMI - Σύγκριση SfM μεταξύ MicMac και COLMAP

Η πυκνότητα της ανακατασκευής είναι παρόμοια μεταξύ των δύο στο μεγαλύτερο κομμάτι της σκηνής, με το MicMac να παράγει εξαιρετικά πυκνότερο νέφος σε ορισμένα σημεία. Αντίθετα, το MicMac απέτυχε να κατασκευάσει τον τρούλο του κτιρίου. Το COLMAP είχε καλύτερα αποτελέσματα σε αυτό το σημείο.



(Εικόνα 3.7) - TZAMI - Σύγκριση MVS μεταξύ MicMac και Colmap

Γενικά, το MicMac φαίνεται να είναι πιο λεπτομερές αλλά και αυστηρότερο όσον αφορά την διατήρηση/αφαίρεση points σε σημεία με έλλειψη κάλυψης από το dataset, πράγμα που μπορεί να δημιουργήσει κενά στο νέφος. Αντίθετα, το COLMAP παρέχει πιο ομοιόμορφη κάλυψη εις βάρος της ολικής πυκνότητας. Και στις δύο ανακατασκευές, το χρώμα είναι ακριβές και πιστό στις εικόνες εισόδου.

Όσον αφορά τον αλγόριθμο NERF, το αποτέλεσμα ήταν ικανοποιητικό για οπτικοποίηση του χώρου, με καλή σχετικά υφή και γεωμετρία. Υπάρχει καλή κάλυψη όλης της σκηνής, συμπεριλαμβανομένου και του τρούλου του κτιρίου, κάτι που οι προηγούμενες τεχνικές δεν μπόρεσαν να απεικονίσουν. Η ποιότητα της ανακατασκευής είναι χαμηλότερη από την μεθόδο MVS, με μικρές λεπτομέρειες όπως η ανάγλυφη υφή του κτιρίου να είναι μεταβίας αντιληπτή. Το NERF φαίνεται να έχει ανακατασκευάσει και τον περίγυρο του κτιρίου, με όσο το δυνατόν περισσότερη ακρίβεια από την περιορισμένη κάλυψη του απο το dataset, καθώς και έναν υποτυπώδη “ουρανό” για την σκηνή, πράγμα που κάνει την απεικόνιση ιδιαίτερα ρεαλιστική και οπτικά παρόμοια με τις φωτογραφίες του dataset (εικόνα 3.8).



(Εικόνα 3.8) - TZAMI - NERF με Nvidia Instant NERFs

Το Gaussian Splatting είχε παρόμοια αποτελέσματα με το NERF αλλά με καλύτερη ανάλυση της υφής των αντικειμένων. Διατηρούνται λεπτομέρειες από τις φωτογραφίες-εισόδους που δεν επιβιώνουν στο NERF, όπως η λεπτομέρεια στους αρμούς ανάμεσα στις πέτρες του κτιρίου. Όπως και με το NERF, υπάρχει ουρανός και ικανοποιητική ανακατασκευή του χώρου γύρω από το κτίριο. Αντίθετα με το NERF όμως, υπάρχουν κενά όπως στους τρούλους των δυο κτιρίων καθώς και “ορφανά” gaussians που αιωρούνται σε διάφορα σημεία (εικόνα 3.9).



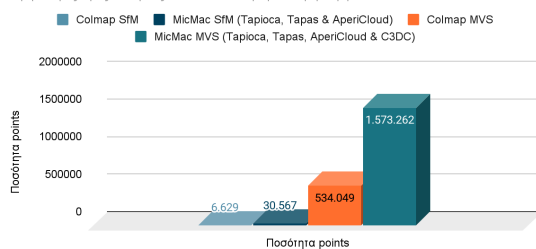
(Εικόνα 3.9) - TZAMI - Gaussian Splatting με POSTSHOT

ΤΡΙΚΚΙ - Ποσοτική αξιολόγηση

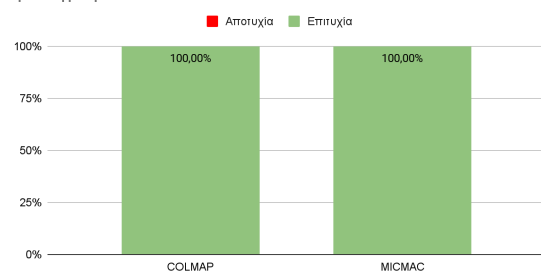
Αυτό το dataset αποτελείται από περιορισμένο αριθμό εικόνων εισόδου. Είναι αναμενόμενο η ανακατασκευή να έχει κενά. Το COLMAP SfM παρήγαγε σημειακό νέφος με 6.629 points με χρόνο εκτέλεσης 11 δευτερόλεπτα ενώ το MicMac SfM (Tapioca, Tapas & AperiCloud) παρήγαγε σημειακό νέφος με 30.567 points με χρόνο εκτέλεσης 2 λεπτά. Στο βήμα MVS, το COLMAP παρήγαγε σημειακό νέφος με 534.049 points με χρόνο εκτέλεσης 9 λεπτά, ενώ το MicMac (Tapioca, Tapas, AperiCloud & C3DC) παρήγαγε σημειακό νέφος με 1.573.262 points με χρόνο εκτέλεσης 2 λεπτά.

Διάγραμμα 3.10 - ΤΡΙΚΚΙ - Συνολικός αριθμός points στο point cloud

Υψηλότερες τιμές συνήθως υποδεικνύουν μεγαλύτερη ακρίβεια



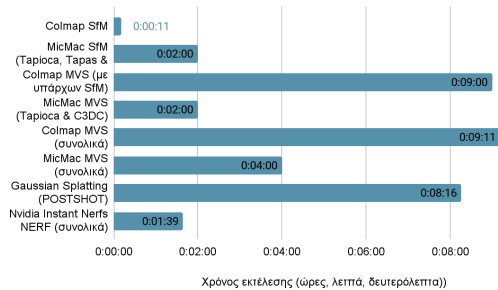
Διάγραμμα 3.11 - ΤΡΙΚΚΙ - Ποσοστό επιτυχημένης χρήσης των φωτογραφιών του dataset



Όλες οι φωτογραφίες ενσωματώθηκαν χωρίς σφάλματα. Το SfM μέσω COLMAP παρήγαγε σημειακό νέφος μέσης πυκνότητας 505 points/μονάδα επιφάνειας ενώ το αντίστοιχο SfM μέσω MicMac παρήγαγε σημειακό νέφος μέσης πυκνότητας 1442 points/μονάδα επιφάνειας. Το βήμα MVS παρήγαγε σημειακό νέφος μέσης πυκνότητας 26.667 points/μονάδα επιφάνειας και 112.096 points/μονάδα επιφάνειας για το COLMAP

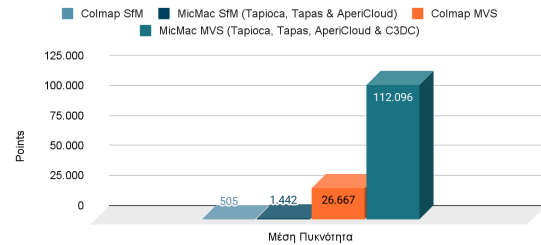
και MicMac αντίστοιχα. Το gaussian splatting εκτελέστηκε σε 8 λεπτά και 16 δευτερόλεπτα και το NERF σε 1 λεπτό και 39 δευτερόλεπτα, 11 απο τα οποία είναι η εκτέλεση του SfM μέσω COLMAP.

Διάγραμμα 3.12 - TRIKKI - Χρόνος εκτέλεσης



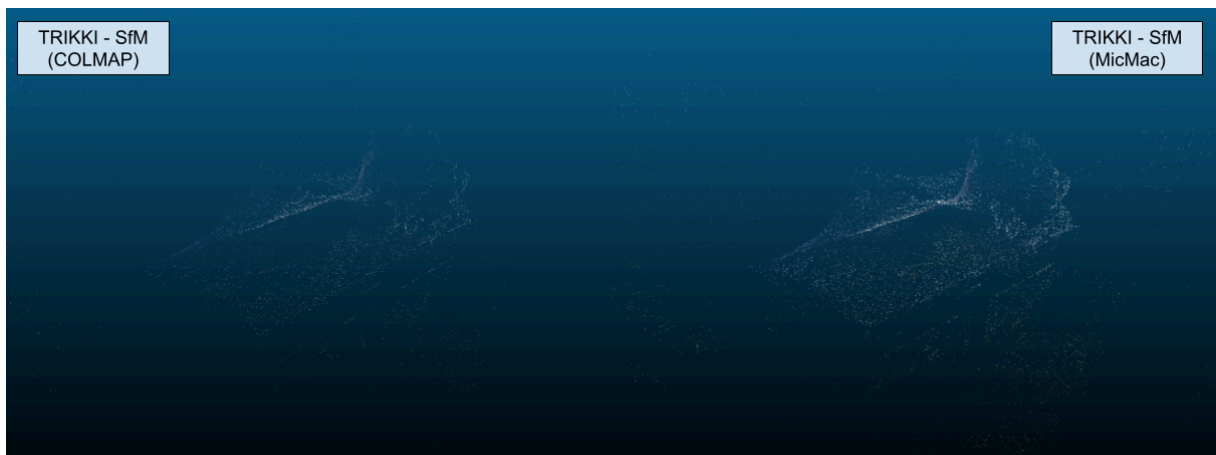
Διάγραμμα 3.13 - TRIKKI - Μέση πυκνότητα ανά μονάδα επιφάνειας

Υψηλότερες τιμές συνήθως υποδεικνύουν μεγαλύτερη ακρίβεια



ΤΡΙΚΚΙ - Ποιοτική αξιολόγηση

Τα αραιά σημειακά νέφη που προκύπτουν απο το SfM δεν έχουν ιδιαίτερη οπτική αξία και εξυπηρετούν μόνο ως εισοδοι στο MVS. Το MicMac φαίνεται να έχει αποτυπώσει περισσότερες λεπτομέρειες της σκηνής-στόχου, με το έδαφος γύρω από το άγαλμα καθώς και την ίδια την βάση του να είναι περισσότερο ορατά σε σύγκριση με το σημειακό νέφος του COLMAP.



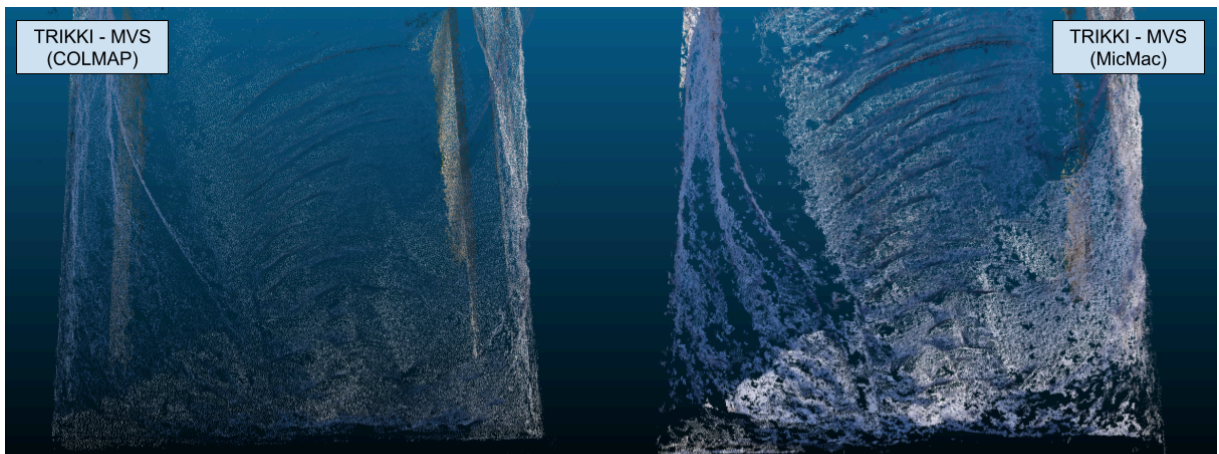
(Εικόνα 3.10) - TRIKKI - Σύγκριση Structure From Motion

Στο βήμα MVS, η κάλυψη των λεπτομερειών φαίνεται να αντιστρέφεται. Το MicMac δείχνει να αφαιρεί κάποια μέρη της σκηνής όπως η βάση του αγάλματος. Το έδαφος δεν έχει ανακατασκευαστεί καθόλου στο σημειακό νέφος του MicMac, ενώ κάποια σημεία του υπάρχουν στο σημειακό νέφος του COLMAP. Η πυκνότητα στο σημειακό νέφος του COLMAP είναι χαμηλότερη από αυτή του MicMac



(Εικόνα 3.11) - TRIKKI - Σύγκριση Multi View Stereo

αλλά η ανακατασκευή είναι πιο πλήρης και δημιουργεί ένα καλύτερο οπτικό αποτέλεσμα εις βάρος λεπτομέρειας στην γεωμετρία. Αντίθετα, το MicMac φαίνεται να επικεντρώνεται στην αποτύπωση της λεπτομέρειας όπου είναι δυνατόν και την απόρριψη points με χαμηλή “βεβαιότητα” θέσης.



(Εικόνα 3.12) - TRIKKI - Πυκνότητα Multi View Stereo

Το NERF του dataset TRIKKI απέδωσε την σκηνή ρεαλιστικά, με όλες τις επιφάνειες να καλύπτονται, συμπεριλαμβανομένου και της πίσω όψης του αγάλματος, η οποία δεν αποτυπώνεται επαρκώς στο dataset και δεν ανακατασκευάστηκε στους αλγορίθμους SfM και MVS. Για ακόμη μια φορά το NERF δείχνει να είναι ικανό να “γεμίσει” τα κενά του dataset με ρεαλιστικό τρόπο. Η υφές της σκηνής δεν αποδίδονται τόσο καθαρά όσο στις υπόλοιπες μεθόδους και αρκετή από την γεωμετρική λεπτομέρεια στην επιφάνεια του αγάλματος έχει χαθεί.



(Εικόνα 3.13) - TRIKKI - NERF με Nvidia Instant Nerfs

Η μέθοδος gaussian splatting είχε καλύτερο αποτέλεσμα στην απόδοση της υφής της σκηνής. Η γεωμετρία είναι πυκνή αλλά παρουσιάζει κενά στις επιφάνειες, παρομοίως με την τεχνική MVS του MicMac. Παρόμοια με το NERF, η πίσω όψη του αγάλματος αποδίδεται επαρκώς.



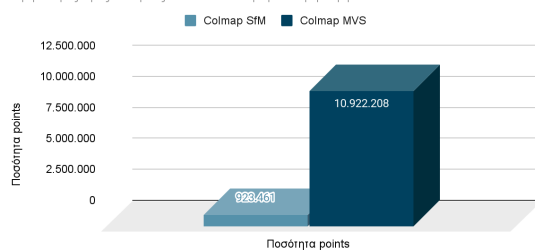
(Εικόνα 3.14) - TRIKKI - Gaussian Splatting

PELINNA - Ποσοτική αξιολόγηση

Το database PELINNA εκτελέστηκε μόνο σε COLMAP λόγο πολυπλοκότητας. Τα αποτελέσματα είναι γενικά χαμηλότερης ποιότητας σε σύγκριση με τα προηγούμενα dataset που αποτελούνται από μικρότερες σκηνές. Το βήμα SfM παρήγαγε σημειακό νέφος με 923.461 σημεία με χρόνο εκτέλεσης 13 ώρες και 33 λεπτά και μέση πυκνότητα 53 σημεία/μονάδα επιφάνειας. Το βήμα MVS παρήγαγε σημειακό νέφος με 10.922.208 σημεία με χρόνο εκτέλεσης 10 ώρες και 50 λεπτά και μέση πυκνότητα 33 σημεία/μονάδα επιφάνειας.

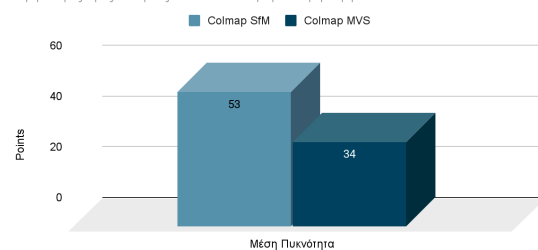
Διάγραμμα 3.14 - PELINNA - Συνολικός αριθμός points στο point cloud

Υψηλότερες τιμές συνήθως υποδεικνύουν μεγαλύτερη ακρίβεια

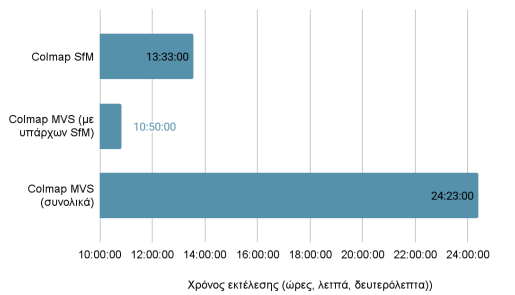


Διάγραμμα 3.15 - PELINNA - Μέση πυκνότητα ανά μονάδα επιφάνειας

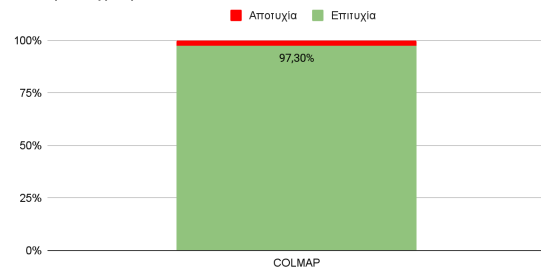
Υψηλότερες τιμές συνήθως υποδεικνύουν μεγαλύτερη ακρίβεια



Διάγραμμα 3.16 - PELINNA - Χρόνος εκτέλεσης

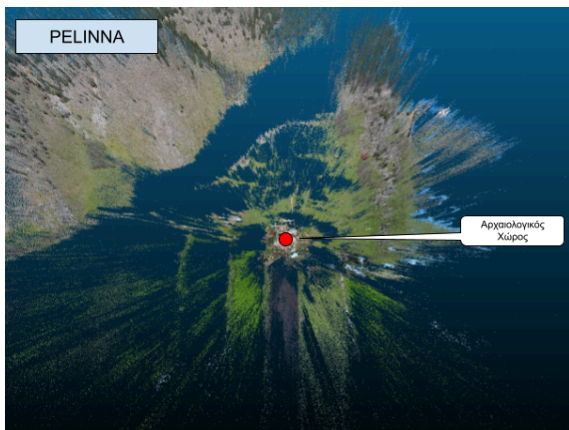


Διάγραμμα 3.17 - PELINNA - Ποσοστό επιτυχημένης χρήσης των φωτογραφιών του dataset



Απο το σύνολο των 1539 φωτογραφιών, ενσωματώθηκαν σωστά οι 1497, δηλαδή ποσοστό επιτυχίας 97,30%.

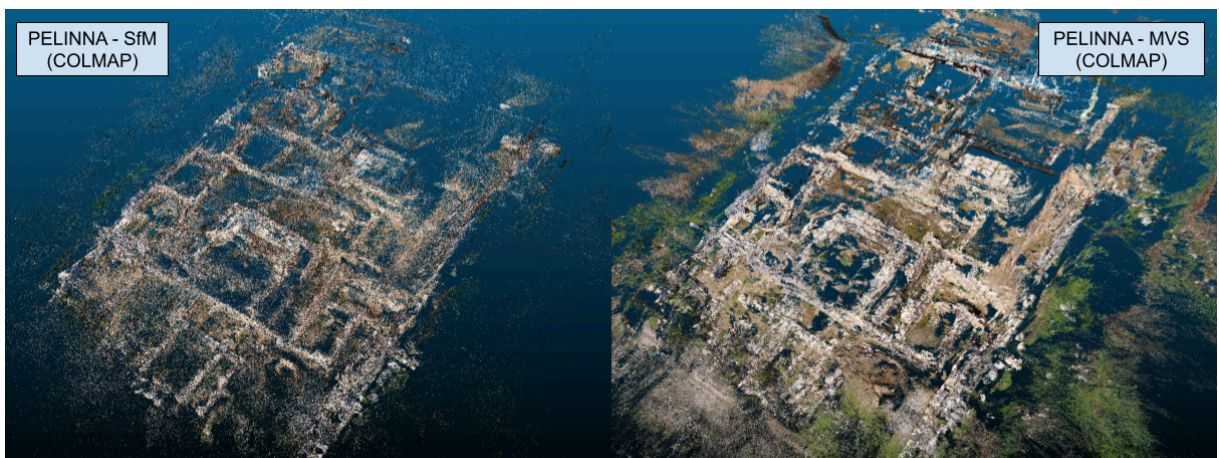
PELINNA - Ποιοτική αξιολόγηση



(Εικόνα 3.15) - PELINNA - Έκταση ανακατασκευής

Η πολύ χαμηλή πυκνότητα των νεφών πιθανώς προέρχεται από την μεγάλη έκταση της ανακατασκευής που περιλαμβάνει πυκνές και αραιές συγκεντρώσεις points (Εικόνα 3.15). Αν και η κάλυψη της σκηνής είναι σχετικά πλήρης, η δαιδαλώδης φύση του αρχαιολογικού χώρου προκαλεί πολλά “τυφλά” σημεία που δεν ανακατασκευάζονται σε ικανοποιητικό επίπεδο και ίσως η χρήση εναέριων λήψεων

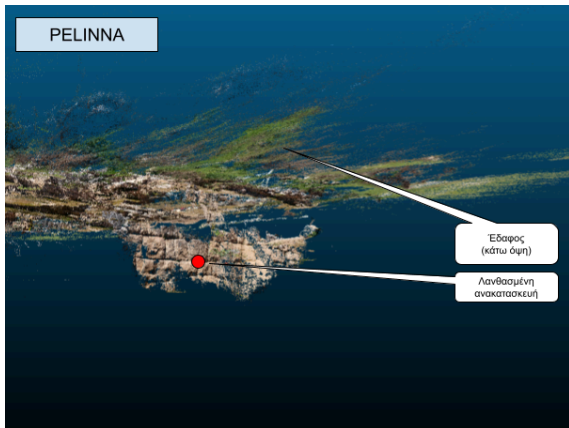
(π.χ απο drone) να επιφέρουν καλύτερα αποτελέσματα.



(Εικόνα 3.16) - PELINNA - Structure from Motion και Multi View Stereo

Παρομοίως, και οι δύο αλγόριθμοι αποτυγχάνουν να ανακατασκευάσουν το έδαφος/δάπεδο ανάμεσα στους αρχαίους τοίχους, παρά την εκτεταμένη τους κάλυψη απο

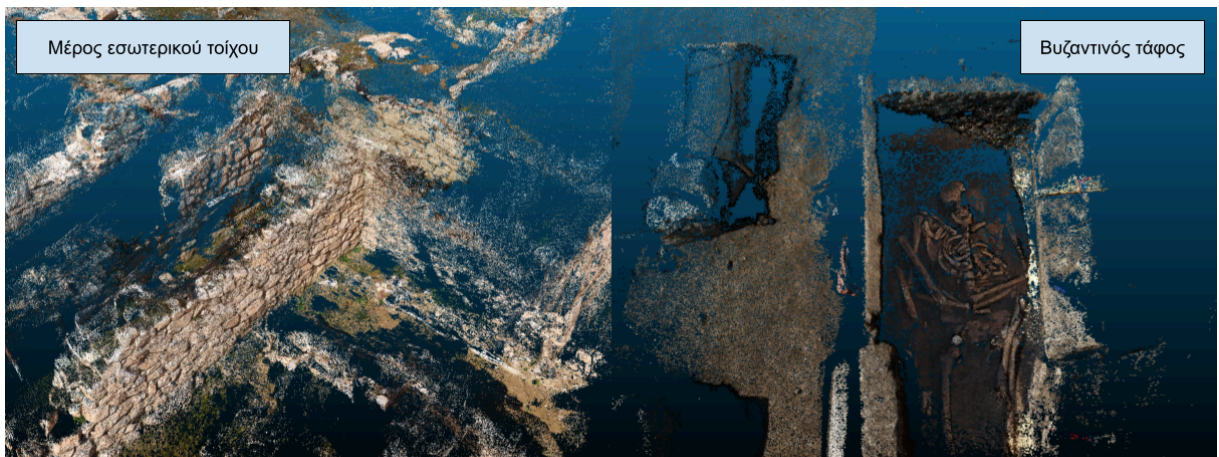
το dataset εισόδου και των πολλών οπτικών σημείων αναφοράς που κανονικά θα ευνοούσαν την ανακατασκευή. Και τα δυο σημειακά νέφη έχουν μεγάλο αριθμό ορφανών σημείων. Επίσης παρατηρείται λανθασμένη ανακατασκευή μέρους του αρχαίου τοίχου, το οποίο έχει τοποθετηθεί σε λάθος σημείο και με λάθος κλίμακα σε σχέση με το υπόλοιπο νέφος (Εικόνα 3.17).



(Εικόνα 3.17) - PELINNA - Λανθασμένη ανακατασκευή

Στα σημεία που η ανακατασκευή δούλεψε καλύτερα υπάρχει υψηλή συγκέντρωση points και η γεωμετρία καθώς και η υφή της σκηνής είναι ακριβής. Παρατηρούνται μικρές λεπτομέρειες όπως οι πέτρινοι τοίχοι και τα ανθρώπινα οστά που βρέθηκαν στον χώρο. Γενικά, η ποιότητα της ανακατασκευής είναι καλή αλλά σε αυτή την περίπτωση χρήζει περεταίρω φωτογραφίες εισόδου ή ακόμα και μια

διαφορετική προσέγγιση στην μέθοδο λήψης.



(Εικόνα 3.18) - PELINNA - Σημεία με σωστή ανακατασκευή

4. Συμπεράσματα και μελλοντικές προεκτάσεις

4.1 Βασικά συμπεράσματα

Με βάση τα παραπάνω δεδομένα, συμπεραίνουμε ότι η τεχνική LIDAR αποτελεί την καλύτερη επιλογή για ακριβή αναπαράσταση της σκηνής-στόχου όσον αφορά την πιστότητα της γεωμετρίας και ανάλογα με την συσκευή (3D scanner), την πιστότητα της υφής των αντικειμένων. Δημιουργεί καθαρά και πυκνά σημειακά νέφη, σε σημείο που είναι δύσκολο να διακριθεί αν το μοντέλο αποτελείται από points ή παραδοσιακά πολύγωνα. Υστερεί όμως στην προσβασιμότητα λόγω του ακριβούς υλικού που χρειάζεται. Υστέρημα μπορεί να θεωρηθεί και το γεγονός ότι η ανακατασκευή εκτελείται αυστηρά και μόνο σε επιφάνεια οι οποίες ήταν σε οπτική επαφή με την συσκευή, χωρίς την δυνατότητα “γεμίσματος” κάποιου κρυφού σημείου, πράγμα που δημιουργεί κενά στην ανακατασκευή. Υστερεί στην ανακατασκευή επιφανειών με ανακλαστικές και διαθλαστικές ιδιότητες.

Ο αλγόριθμος SfM (Structure from Motion) είναι πολύ πιο προσβάσιμος, αφού οι απαιτήσεις υλικού για την συλλογή δεδομένων εισόδου είναι μόνο η δυνατότητα λήψης φωτογραφιών ή βίντεο, πράγμα που αποτελεί δυνατότητα όλων των σύγχρονων κινητών τηλεφώνων. Οι απαιτήσεις υλικού για την εκτέλεση του αλγορίθμου και επεξεργασίας των δεδομένων είναι προσωπικός υπολογιστής με σχετικά χαμηλές επιδόσεις, με παράγοντες όπως η μνήμη συστήματος (RAM) να είναι μεταβλητές ανάλογα με το μέγεθος του dataset. Ο αλγόριθμος SfM όμως, υστερεί στο ότι ανακατασκευάζει μόνο μια απλοποιημένη εκδοχή της σκηνής-στόχου και έτσι δεν αποτελεί βιώσιμη λύση για εφαρμογές που απαιτούν ακρίβεια στην γεωμετρία ή ρεαλιστικές υφές. Ως αποτέλεσμα, συχνά χρησιμοποιείται μονάχα για την ικανότητά του να υπολογίζει την θέση της κάμερας στον τρισδιάστατο χώρο και στην συνέχεια να παρέχει το αποτέλεσμα αυτό ως είσοδο στον αλγόριθμο MVS (Multi View Stereo). Και ο SfM υστερεί στην ανακατασκευή επιφανειών με ανακλαστικές και διαθλαστικές ιδιότητες.

Ο αλγόριθμος MVS είναι εξίσου προσβάσιμος με τον SfM και συνήθως συναντώνται και οι δύο μαζί σε εφαρμογές τρισδιάστατης ανακατασκευής, διότι ο MVS χρειάζεται τα αποτελέσματα του SfM για να αρχικοποιηθεί. Παράγει σημειακό νέφος με

υψηλή πυκνότητα και ρεαλιστική απόδοση γεωμετρίας και υφής. Είναι κατάλληλος για τις περισσότερες εφαρμογές που επικεντρώνονται στην ρεαλιστική απόδοση μιας σκηνής με επίκεντρο κάποιο συγκεκριμένο υποκείμενο, όπως ένα αντικείμενο. Όπως και οι υπόλοιποι αλγόριθμοι, υστερεί στην ανακατασκευή επιφανειών με ανακλαστικές και διαθλαστικές ιδιότητες.

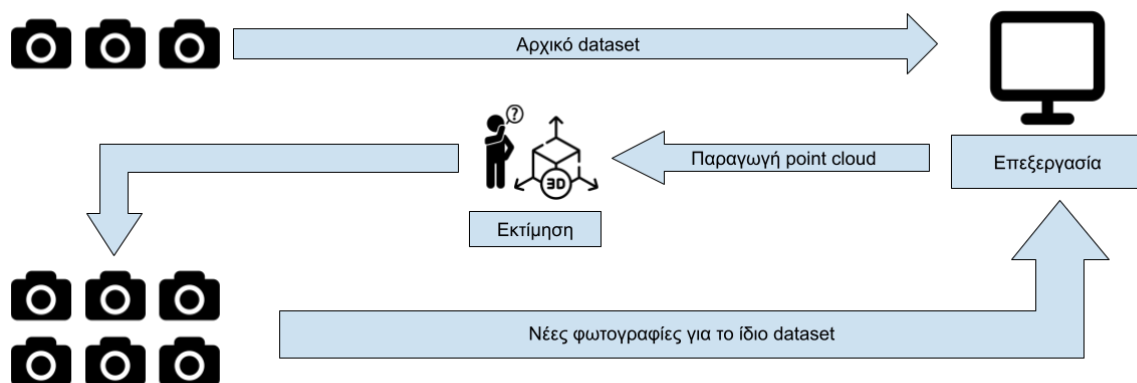
Ο αλγόριθμος gaussian splatting βασίζεται και αυτός στο αποτέλεσμα του SfM και είναι εξίσου προσβάσιμος, αλλά χρησιμοποιεί διαφορετική προσέγγιση στην παρουσίαση (rendering) του αποτελέσματος. Η προσέγγιση αυτή του επιτρέπει να κατασκευάζει μια πιο ολοκληρωμένη εικόνα της σκηνής-στόχου, με την ικανότητα να δημιουργεί την ψευδαίσθηση αντανάκλασεων και διάθλασης στις αντίστοιχες επιφάνειες της σκηνής. Διατηρεί γενικά υψηλή ακρίβεια γεωμετρίας, μέχρι κάποιο σημείο, μετά το οποίο η φύση του μοντέλου που αποτελείται από gaussians αρχίζει να γίνεται ορατή. Η ικανότητά του αυτή να παρουσιάζει μια πιο ρεαλιστική (οπτικά) αναπαράσταση της σκηνής κάνει το gaussian splatting κατάλληλο για εφαρμογές όπου δεν απαιτείται απόλυτη ακρίβεια στο μοντέλο και όπου η ρεαλιστική απεικόνιση είναι πιο σημαντική (ψυχαγωγία, ψηφιακά μουσεία, κα).

Ο αλγόριθμος NERF είναι και αυτός προσβάσιμος διότι βασίζονται στην ανακατασκευή από εικόνες εισόδου και την έξοδο του αλγορίθμου SfM αλλά λόγω των πιο πολύπλοκων τεχνικών rendering του αποτελέσματος, η απαιτήσεως υλικού (κυρίως μνήμης κάρτας γραφικών VRAM) αυξάνονται ραγδαία με το μέγεθος του dataset. Το NERF εκτελεί volume rendering με βάση το ray tracing, μια πολύ υπολογιστικά απαιτητική τεχνική που απαιτεί δυνατή κάρτα γραφικών στο σύστημα που εκτελείται ο αλγόριθμος. Το αποτέλεσμα του έχει τα ίδια πλεονεκτήματα με το gaussian splatting, με εξαίρεση την ακρίβεια της γεωμετρίας και υφών της σκηνής. Το γεγονός ότι το αποτέλεσμα του NERF δεν είναι ένα παραδοσιακό μοντέλο στον τρισδιάστατο χώρο αλλά είναι η “αντίληψη” που έχει ένα νευρωνικό δίκτυο για αυτό το μοντέλο δημιουργεί μια πιο αφηρημένη αποτύπωση της σκηνής, όπου μικρές λεπτομέρειες και πολύπλοκες υφές χάνονται. Η ίδια αφηρημένη αποτύπωση όμως μπορεί να βοηθήσει στην γενική πληρότητα της ανακατασκευής, γεμίζοντας κενά που μπορεί να υπάρχουν στο dataset. Η υψηλές υπολογιστικές απαιτήσεις, όχι μόνο για την εκτέλεση του αλγορίθμου αλλά και για την απεικόνιση του αποτελέσματος, κάνουν το NERF μη βιώσιμο για μεγάλα dataset. Ταυτόχρονα όμως, το ίδιο μειονέκτημα μπορεί να αποτελέσει πλεονέκτημα σε μικρά datasets, τα οποία μπορούν

να απεικονιστούν απο λιγότερο ισχυρά συστήματα. Επειδή το NERF είναι ένα σύνολο εντολών για νευρωνικά δίκτυα, απαιτεί λιγότερο αποθηκευτικό χώρο σε σχέση με παραδοσιακά μοντέλα πολυγώνων ή σημειακών νεφών.

4.2 Προτάσεις για μελλοντική έρευνα

Σίγουρο παράγοντα για την αποτελεσματικότητα και ποιότητα του αποτελέσματος για όλους τους παραπάνω αλγορίθμους αποτελεί η ποιότητα του dataset. Υπάρχει χώρος για βελτίωση σε όλα τα dataset της έρευνας καθώς και ένα θεωρητικό βέλτιστο. Παρόλα αυτά, η έρευνα είχε σκοπό την αξιολόγηση των αλγορίθμων σε πραγματικές εφαρμογές και όχι σε ιδανικές συνθήκες. Ενδιαφέρουσα θα ήταν η σύγκριση των αλγορίθμων σε ιδανικές συνθήκες. Επίσης άξιες αναφοράς είναι και διάφορες εναλλακτικές ή πειραματικές μέθοδοι εκτέλεσης και εφαρμογής των αλγορίθμων που μπορούν να βελτιώσουν τα αποτελέσματα ή να ανοίξουν τον δρόμο για νέες περιπτώσεις χρήσης της τρισδιάστατης ανακατασκευής. Κάποιες εφαρμογές SfM/MVS παρέχουν την δυνατότητα “live reconstruction”, μιας ικανότητας που αφήνει τον χρήστη να προσθέσει φωτογραφίες στο dataset αργότερα και να έχει άμεση ανατροφοδότηση για το ποια σημεία της σκηνης-στόχου δεν έχουν επαρκή κάλυψη (εικόνα 4.1).



(Εικόνα 4.1) - Ανατροφοδότηση SfM/MVS

Αυτό μειώνει δραματικά τα ποσοστά λανθασμένης ανακατασκευής και είναι ιδιαίτερα χρήσιμο σε εφαρμογές τρισδιάστατης ανακατασκευής μνημείων πολιτιστικής κληρονομιάς διότι εξασφαλίζει το καλύτερο δυνατό αποτέλεσμα με μία μόνο επίσκεψη στον χώρο. Οι επισκέψεις σε αρχαιολογικούς χώρους ή πρόσβαση σε αρχαιολογικά αντικείμενα για σκοπούς τρισδιάστατης ανακατασκευής μπορεί να είναι μια χρονοβόρα και κοστοβόρα διαδικασία από την πλευρά των τεχνικών που εκτελούν την ανακατασκευή. Παράλληλα, οι

λιγότερες επισκέψεις μειώνουν την φθορά στα μνημεία-στόχους και προκαλούν λιγότερες διακοπές σε τυχόν καθημερινό πρόγραμμα ενός αρχαιολογικού χώρου/μουσείου/κ.α. Σε συνδυασμό με τα παραπάνω, οι εξελίξεις στον τομέα της ασύρματης επικοινωνίας επιτρέπουν την επεξεργασία dataset τρισδιάστατης ανακατασκευής απο απομακρυσμένες τοποθεσίες, χωρίς την ανάγκη παρουσίας ενός ηλεκτρονικού υπολογιστή ή ακόμα και παροχή ηλεκτρισμού. Είναι πλέον δυνατόν οι φωτογραφίες να μεταφέρονται από την κάμερα/κινητό στον υπολογιστή που θα εκτελέσει τον αλγόριθμο μέσω του ασύρματου δικτύου ίντερνετ (4G LTE/5G) που παρέχουν πλέον οι περισσότερες εταιρείες κινητής τηλεφωνίας. Η επεξεργασία γίνεται απομακρυσμένα και τα αποτελέσματα επιστρέφονται μέσω της ίδιας μεθόδου στους ερευνητές στον χώρο.

ΠΑΡΑΡΤΗΜΑ Α

Σε αυτή αυτό το παράρτημα συμπεριλαμβάνεται το υλικό που προέκυψε από την εκτέλεση των αλγορίθμων τρισδιάστατης ανακατασκευής για κάθε dataset. Το αποτέλεσμα κάθε αλγορίθμου θα συνοδεύεται από παρατηρήσεις και μια εικόνα, καθώς και σχόλια σε περίπτωση αδυναμίας εκτέλεσης κάποιου από αυτούς. Τα δεδομένα εξόδου και οι συγκρίσεις που γίνονται σε αυτά έχουν βάση τον στόχο κάθε αλγορίθμου όπως περιγράφεται στην ενότητα 2.2.

Ανακατασκευή στο dataset COURT

Το dataset COURT θα εκτελεστεί πρώτο διότι περιλαμβάνει ground truth και είναι συμβατό με όλες τις δοκιμές της που θέλουμε να διεξάγουμε. Όλες οι φωτογραφίες του dataset προέρχονται από βίντεο ανάλυσης 4K (3840 x 2160 pixel) στα 29.97 FPS.

COURT - Ground truth (LIDAR)

Το ground truth θα αποτελεί το μέτρο σύγκρισης για τα αποτελέσματα των αλγορίθμων που θα εκτελεστούν σε αυτό το dataset.



(Εικόνα Α1) - COURT - LIDAR ground truth.

COURT - Structure From Motion με COLMAP



(Εικόνα Α2) - COURT - Structure from Motion με COLMAP

Παρατηρήσεις:

Η εκτέλεση διήρκησε 3 ώρες και 17 λεπτά. Απο τις 1106 φωτογραφίες του dataset, υλοποιήθηκαν σωστά οι 1050. Το αποτέλεσμα είναι σημειακό νέφος με 559.008 points. Το σημειακό νέφος που προκύπτει έχει τυχαία περιστροφή, θέση στον τρισδιάστατο χώρο και κλίμακα. Για την σωστή σύγκριση χρειάζεται η ευθυγράμμιση του με το σημειακό νέφος του ground truth.

COURT - Structure From Motion με MICMAC

Η εκτέλεση του MicMac απέτυχε λόγω του μεγάλου όγκου των φωτογραφιών (1116 εικόνες), με αποτέλεσμα να μην ολοκληρώνεται η διαδικασία πυκνής ανακατασκευής ακόμη και μετά από πολλαπλές προσπάθειες. Αυτό επιβεβαιώνει την ανάγκη περιορισμού των εικόνων ή χρήσης εξειδικευμένων παραμέτρων σε τέτοιου είδους datasets.

COURT - Multi View Stereo με COLMAP

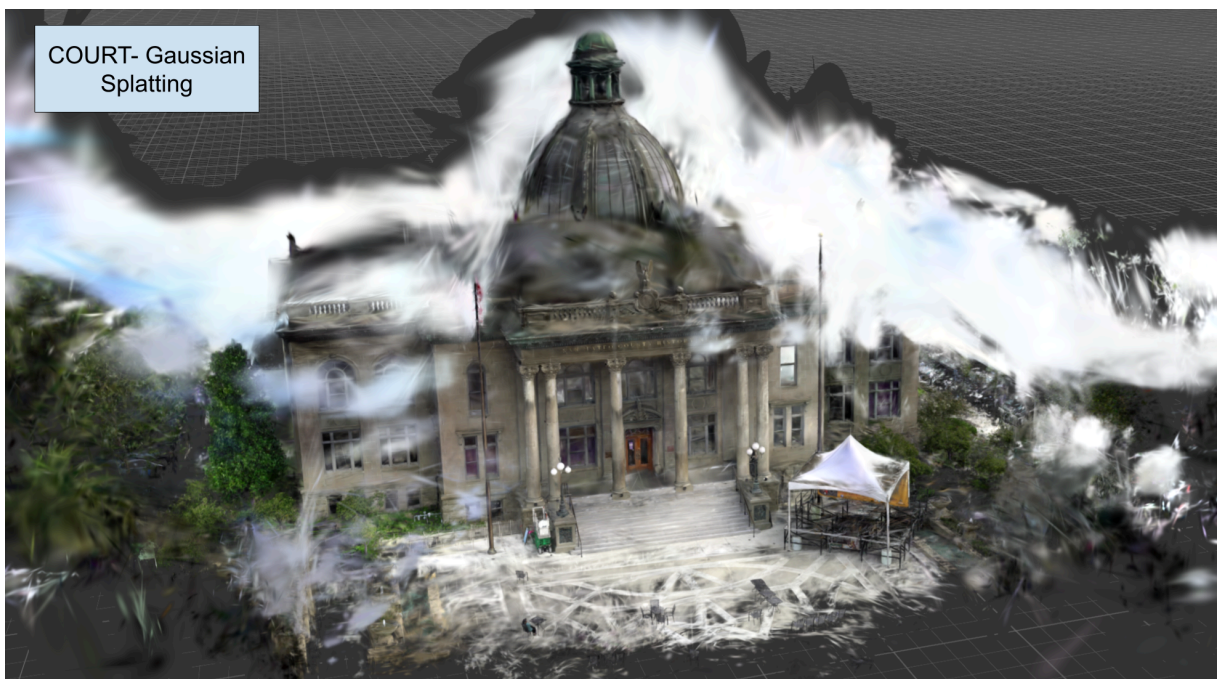


(Εικόνα A3) - COURT - Multi View Stereo με COLMAP

Παρατηρήσεις:

Χρησιμοποιήθηκαν τα δεδομένα camera solving απο την προηγούμενη εκτέλεση του structure from motion. Η εκτέλεση διήρκησε 9 ώρες και 12 λεπτά. Το αποτέλεσμα είναι ένα σημειακό νέφος με 31.091.943 points. Το σημειακό νέφος που προκύπτει έχει τυχαία περιστροφή, θέση στον τρισδιάστατο χώρο και κλίμακα. Για την σωστή σύγκριση χρειάζεται η ευθυγράμμιση του με το σημειακό νέφος του ground truth.

COURT - Gaussian Splatting με POSTSHOT



(Εικόνα A4) - COURT - Gaussian Splatting με POSTSHOT

Παρατηρήσεις:

Η επεξεργασία εκτελέστηκε με τις εξής ρυθμίσεις:

- Downsampling των φωτογραφιών εισόδου στα 1000px
- 3000 max splats.
- 50.000 training steps.

Μετά την καταγραφή των ποσοτικών αποτελεσμάτων, έγινε χειροκίνητη αφαίρεση κάποιων gaussians για καλύτερη ορατότητα του μοντέλου. Η εκτέλεση διήρκησε 5 ώρες και 29 λεπτά.

COURT - NERF με Nvidia Instant Nerfs



(Εικόνα A5) - COURT - NERF με Nvidia Instant Nerfs

Παρατηρήσεις: Η επεξεργασία εκτελέστηκε με είσοδο το σημειακό νέφος απο το COLMAP. Η εκπαίδευση του NERF διήρκησε περίπου 2 λεπτά.

Ανακατασκευή στο dataset TZAMI

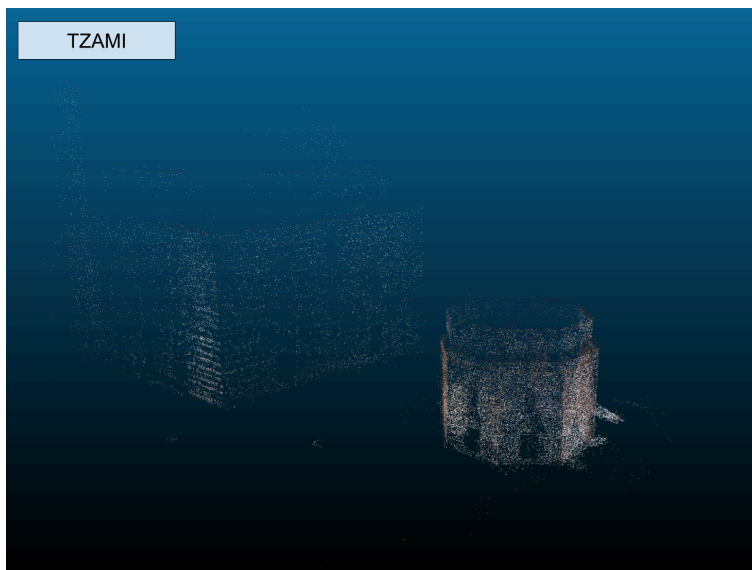
Το dataset TZAMI έχει ως σκοπό την εφαρμογή των αλγορίθμων σε dataset με σκηνή-στόχο ένα κτίριο με χαμηλή πολυπλοκότητα. Το οκτάγωνο σχήμα παρέχει εύκολη πρόσβαση για φωτογράφιση και δεν έχει κρυφά σημεία.

Η φωτογράφιση έγινε με κινητό τηλέφωνο Samsung Galaxy S21 με τις εξής ρυθμίσεις:

- Ανάλυση: 3024 x 4032 pixels (portrait mode).

- Βάθος χρώματος: 24-bit.
- Διάφραγμα (F-stop): f/1.8.
- Χρόνος έκθεσης: 1/6000 του δευτερολέπτου.
- ISO: 160.
- Focal length: 5mm.

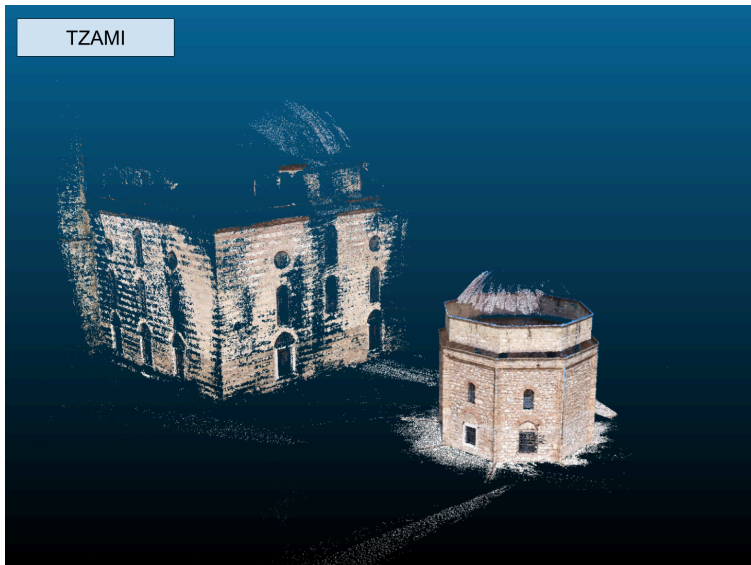
TZAMI - Structure From Motion με COLMAP



(Εικόνα Α6) - TZAMI - Structure From Motion με COLMAP.

Παρατηρήσεις: Η εκτέλεση διήρκησε 2 λεπτά. Το αποτέλεσμα είναι ένα σημειακό νέφος με 63.352 points.

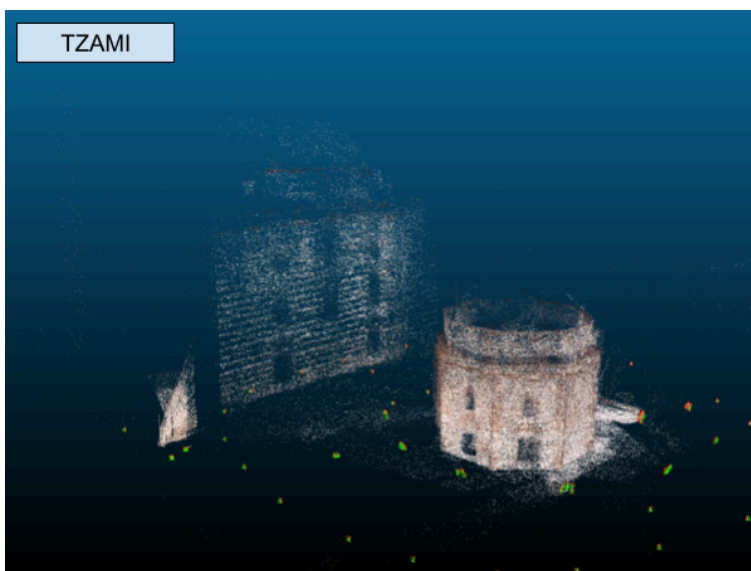
TZAMI - Multi View Stereo με COLMAP



(Εικόνα A7) - TZAMI - Multi View Stereo με COLMAP

Παρατηρήσεις: Χρησιμοποιήθηκαν τα δεδομένα camera solving απο την προηγούμενη εκτέλεση του structure from motion. Η εκτέλεση διήρκησε 47 λεπτά. Το αποτέλεσμα είναι ένα σημειακό νέφος με 2.305.069 points.

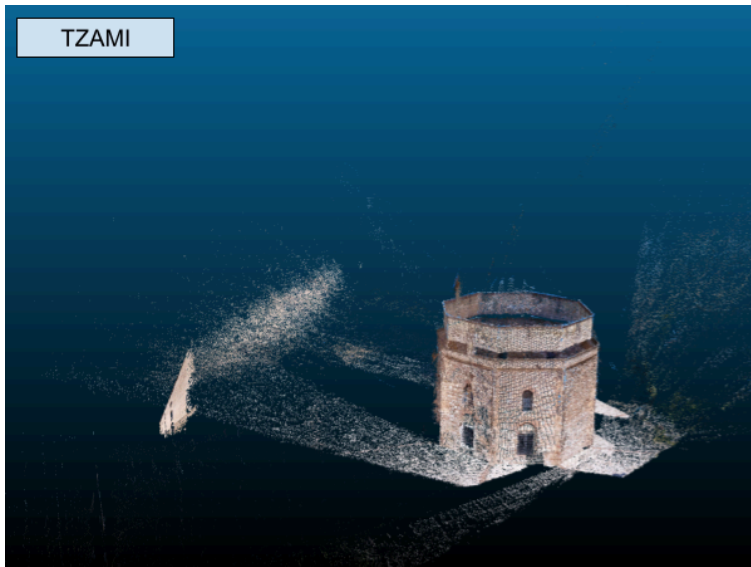
TZAMI - Structure From Motion με MicMac (Tapioca, Tapas & AperiCloud)



(Εικόνα A8) - TZAMI - Structure From Motion με MicMac (Tapioca & AperiCloud).

Παρατηρήσεις: Η εκτέλεση διήρκησε 1 ώρα και 7 λεπτα. Το αποτέλεσμα είναι ένα σημειακό νέφος με 935.557 points.

TZAMI - Multi View Stereo με MicMac (Tapioca & C3DC)



(Εικόνα A9) - TZAMI - Multi View Stereo με MicMac (Tapioca & C3DC)

Παρατηρήσεις: Η εκτέλεση διήρκησε 86 λεπτα. Χρησιμοποιήθηκαν τα δεδομένα camera solving απο το Tapioca του προηγούμενου βήματος. Το αποτέλεσμα είναι ένα σημειακό νέφος με 2.236.538 points.

TZAMI - Gaussian Splatting με POSTSHOT



(Εικόνα A10) - TZAMI - Gaussian Splatting με POSTSHOT

Παρατηρήσεις: Η εκτέλεση διήρκεσε 23 λεπτα.

TZAMI - NERF με Nvidia Instant Nerfs



(Εικόνα A11) - TZAMI - NERF με Nvidia Instant Nerfs

Παρατηρήσεις: Η επεξεργασία εκτελέστηκε με είσοδο το αραιό σημειακό νέφος απο το COLMAP. Η εκπαίδευση του NERF διήρκεσε περίπου 3 λεπτά.

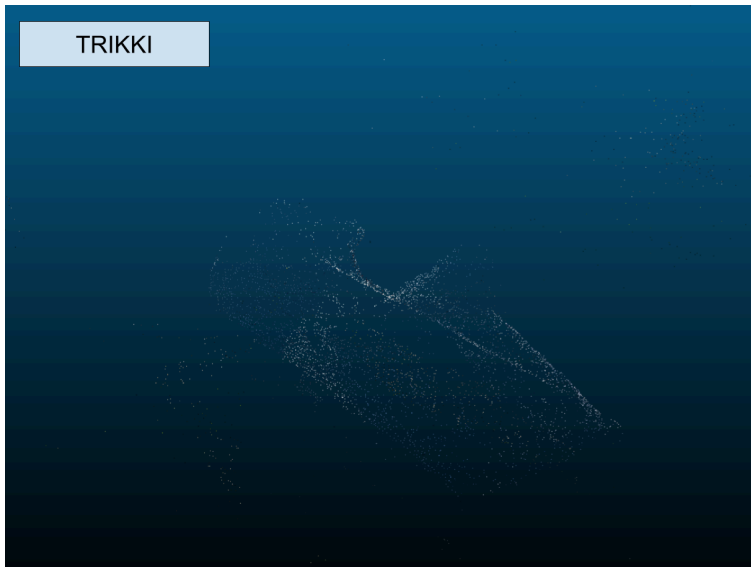
Ανακατασκευή στο dataset TRIKKI

Το dataset TRIKKI έχει ως σκοπό την εφαρμογή των αλγορίθμων σε dataset με μικρό αριθμό εικόνων εισόδου. Η σκηνή-στόχος είναι ένα άγαλμα με τυφλά σημεία στο σχήμα του και μη πλήρη κάλυψη από όλες τις οπτικές γωνίες κατά την διάρκεια της φωτογράφισης. Δοκιμάζεται η ικανότητα των αλγορίθμων να χειριστούν περιπτώσεις με ελλιπής πληροφορίες.

Η φωτογράφιση έγινε με κινητό τηλέφωνο Samsung Galaxy S21 με τις εξής ρυθμίσεις:

- Ανάλυση: 4032 x 3024 pixels.
- Βάθος χρώματος: 24-bit.
- Διάφραγμα (F-stop): f/1.8.
- Χρόνος έκθεσης: 1/6000 του δευτερολέπτου.
- ISO: 250.
- Focal length: 5mm.

TRIKKI - Structure From Motion με COLMAP

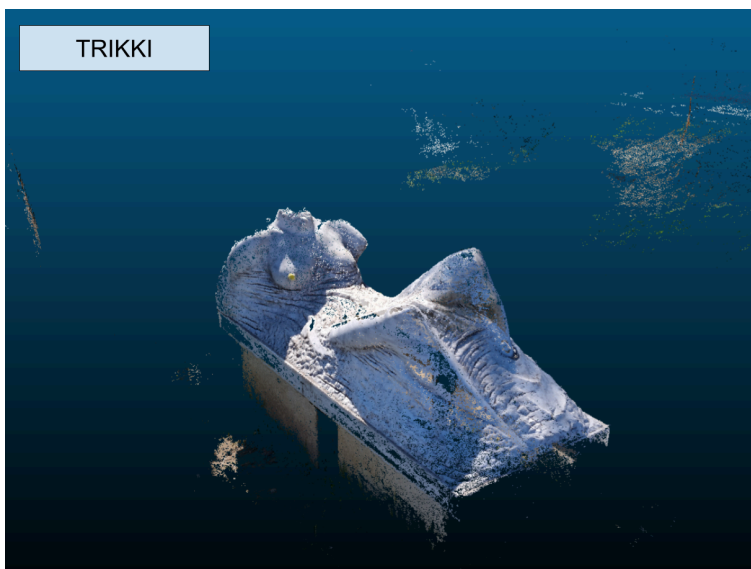


(Εικόνα A12) - TRIKKI - Structure From Motion με COLMAP

Παρατηρήσεις:

- Η εκτέλεση διήρκεσε 11 δευτερόλεπτα.
- Το αποτέλεσμα είναι ένα σημειακό νέφος με 6.629 points.

TRIKKI - Multi View Stereo με COLMAP



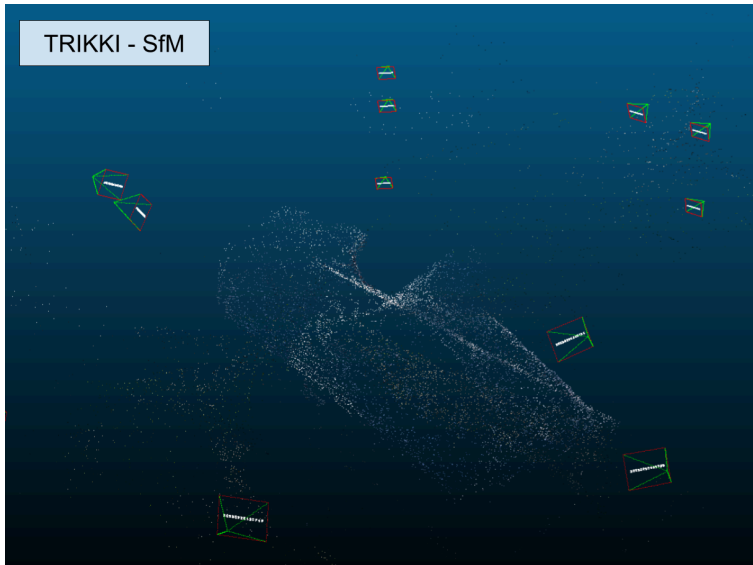
(Εικόνα A13) - TRIKKI - Multi View Stereo με COLMAP

Παρατηρήσεις:

- Χρησιμοποιήθηκαν τα δεδομένα camera solving απο την προηγούμενη εκτέλεση του structure from motion.

- Η εκτέλεση διήρκεσε 9 λεπτά.
- Το αποτέλεσμα είναι ένα σημειακό νέφος με 534.049 points.

TRIKKI - Structure From Motion με MicMac (Tapioca, Tapas & AperiCloud)



(Εικόνα A14) - TRIKKI - Structure From Motion με MicMac
(Tapioca, Tapas & AperiCloud)

Παρατηρήσεις: Η εκτέλεση διήρκεσε 2 λεπτά. Το αποτέλεσμα είναι ένα σημειακό νέφος με 30.567 points.

TRIKKI - Multi View Stereo με MicMac (Tapioca & C3DC)



(Εικόνα A15) - TRIKKI - Multi View Stereo με MicMac
(Tapioca & C3DC)

Παρατηρήσεις: Χρησιμοποιήθηκαν τα δεδομένα camera solving απο την προηγούμενη εκτέλεση του structure from motion. Η εκτέλεση διήρκεσε 2 λεπτά. Το αποτέλεσμα είναι ένα σημειακό νέφος με 1.573.262 points.

TRIKKI - Gaussian Splatting με POSTSHOT



(Εικόνα A16) - *TRIKKI - Gaussian Splatting με POSTSHOT*

Παρατηρήσεις: Η εκτέλεση διήρκεσε 8 λεπτά.

TRIKKI - NERF με Nvidia Instant Nerfs

Παρατηρήσεις: Η εκτέλεση διήρκεσε 3 λεπτά.

Ανακατασκευή στο dataset PELLINA

Το dataset PELLINA έχει ως σκοπό την πρακτική εφαρμογή των αλγορίθμων σε ένα πραγματικό αρχαιολογικό σενάριο. Η σκηνή-στόχος είναι μεγάλης έκτασης και υψηλής πολυπλοκότητας. Εκτελέστηκε μόνο σε COLMAP λόγω πολυπλοκότητας.

Η φωτογράφιση έγινε με κινητό τηλέφωνο Samsung Galaxy S21 με τις εξής ρυθμίσεις:

- Ανάλυση: 4032 x 3024 pixels (portrait mode).
- Βάθος χρώματος: 24-bit.
- Διάφραγμα (F-stop): f/1.8 και f/2.2.
- Χρόνος έκθεσης: 1/6000 και 1/1752 του δευτερολέπτου.
- ISO: 50 και 250.

- Focal length: 2 mm και 5 mm.

Οι διαφορετικές ρυθμίσεις προκύπτουν διότι έγιναν δύο επισκέψεις στον χώρο. Έγινε χρήση και του ευρυγώνιου φακού του τηλεφώνου.

PELLINA - Structure From Motion με COLMAP



(Εικόνα A17) - PELINNA - Structure from Motion με COLMAP

Παρατηρήσεις: Η εκτέλεση διήρκεσε 13 ώρες και 33 λεπτά.

PELLINA - Multi View Stereo με COLMAP



(Εικόνα A18) - PELINNA - Multi View Stereo με COLMAP

Παρατηρήσεις: Χρησιμοποιήθηκαν τα δεδομένα camera solving από την προηγούμενη εκτέλεση του structure from motion. Η εκτέλεση διήρκεσε 10 ώρες και 50 λεπτά.

PELLINA - Gaussian Splatting με POSTSHOT

Παρατηρήσεις: Η εκτέλεση απέτυχε λόγω υπέρβασης της μνήμης της κάρτας γραφικών του υπολογιστή. Αυτό σημαίνει ότι το Gaussian Splat χρειάστηκε πάνω από 12 GB μνήμης για να απεικονισθεί.

PELLINA - NERF με Nvidia Instant Nerfs

Παρατηρήσεις: Η εκτέλεση απέτυχε λόγω υπέρβασης της μνήμης της κάρτας γραφικών του υπολογιστή. Αυτό σημαίνει ότι το NERF χρειάστηκε πάνω από 12 GB μνήμης για να απεικονισθεί.

ΒΙΒΛΙΟΓΡΑΦΙΑ

- [1] R. Fergus Prof, “Lecture 6: Multi-view Stereo & Structure from Motion”.
- [2] Computerphile, “Structure from Motion Octocopter.” Feb. 05, 2016. Accessed: Mar. 04, 2025. [Online]. Available: https://www.youtube.com/watch?v=7DNmN_qWUgM
- [3] T. Moons, *3D Reconstruction from Multiple Images Part 1: Principles*. now Publishers Inc, 2008. Accessed: Mar. 05, 2025. [Online]. Available: <https://doi.org/10.1561/9781601982858>
- [4] Ελληνικό Ινστιτούτο Ολογραφίας, “Περί Στερεοσκοπίας.” Accessed: Mar. 05, 2025. [Online]. Available: <http://www.hih.org.gr/el/holography/stereoscopy.html>
- [5] I. RealSense, “The basics of stereo depth vision – Intel® RealSense™ Depth and Tracking Cameras,” Intel® RealSense™ Depth and Tracking Cameras. Accessed: Mar. 05, 2025. [Online]. Available: <https://www.intelrealsense.com/stereo-depth-vision-basics/>
- [6] J. Cohen, “Photogrammetry Explained: From Multi-View Stereo to Structure from Motion,” PyImageSearch. Accessed: Mar. 05, 2025. [Online]. Available: <https://pyimagesearch.com/2024/10/14/photogrammetry-explained-from-multi-view-stereo-to-structure-from-motion/>
- [7] M. Mendez, “What is LiDAR? Transforming Industries with Precision Mapping,” Inertial Labs. Accessed: Mar. 21, 2025. [Online]. Available: <https://inertiallabs.com/what-is-lidar-transforming-industries-with-precision-mapping/>
- [8] T. Neff, *The Laser That’s Changing the World: The Amazing Stories behind Lidar, from 3D Mapping to Self-Driving Cars*. Prometheus Books, 2018.
- [9] E. Hetherington, “LiDAR on a phone - a new way to capture 3D data - Resource Centre,” Resource Centre | Esri UK. Accessed: Mar. 23, 2025. [Online]. Available: <https://resource.esriuk.com/blog/lidar-on-a-phone-a-new-way-to-capture-3d-data/>

- [10]M. Mendez, “What is LiDAR? Transforming Industries with Precision Mapping,” Inertial Labs. Accessed: Mar. 21, 2025. [Online]. Available: <https://inertiallabs.com/what-is-lidar-transforming-industries-with-precision-mapping/>
- [11]M. Schimanko, “3DMakerPro EAGLE LIDAR Scanner Review.” [Youtube]. Available: <https://youtu.be/PlxEBllHLfc>
- [12]P. Kumar, “What are RGBD cameras? Why RGBD cameras are preferred in some embedded vision applications?,” e-con Systems. Accessed: Mar. 23, 2025. [Online]. Available: <https://www.e-consystems.com/blog/camera/technology/what-are-rgb-d-cameras-why-rgb-d-cameras-are-preferred-in-some-embedded-vision-applications/>
- [13]B. Mildenhall, P. P. Srinivasan, M. Tancik, J. T. Barron, R. Ramamoorthi, and R. Ng, “NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis,” in *Lecture Notes in Computer Science*, Cham: Springer International Publishing, 2020, pp. 405–421. Accessed: Mar. 25, 2025. [Online]. Available: https://doi.org/10.1007/978-3-030-58452-8_24
- [14]D. Ebert PhD, “Introduction to 3D Gaussian Splatting,” Hugging Face. Accessed: Mar. 29, 2025. [Online]. Available: <https://huggingface.co/blog/gaussian-splatting>
- [15]“Neil Armstrong Spacesuit,” 3D Digitization. Accessed: Apr. 07, 2025. [Online]. Available: <https://3d.si.edu/object/3d/neil-armstrong-spacesuit:d8c63ba6-4ebc-11ea-b77f-2e728ce88125>
- [16] Accessed: Apr. 07, 2025. [Online]. Available: <https://www.cosmote.gr/cs/cosmote/gr/CHRONOS.html>
- [17]“Liberty or Death: Boston Tea Party,” *IMDb*. Nov. 09, 2023. Accessed: Apr. 07, 2025. [Video]. Available: <https://www.imdb.com/title/tt29920137/>
- [18]smpteconnect, “Emerging Tech Stage How to Use 3D Laser Scanning in Film Production 10/22/2024,” *Smppteconnect*. Dec. 02, 2024. Accessed: Apr. 07, 2025. [Online]. Available: <https://www.youtube.com/watch?v=HhNpTjvgaOY>

- [19]The Astronauts. [Online]. Available: <https://www.theastronauts.com/2014/03/visual-revolution-vanishing-ethan-carter/>
- [20]“Τέμενος του Οσμάν Σαχ,” Δήμος Τρικκαίων. Accessed: Mar. 29, 2025. [Online]. Available: <https://trikalacity.gr/building/temenos-toy-osman-sach/>
- [21]“Αρχαία Πέλινα, Πετρόπορος,” ΟΔΥΣΣΕΥΣ. Accessed: Mar. 29, 2025. [Online]. Available: http://odysseus.culture.gr/h/3/gh351.jsp?obj_id=6765
- [22]N. Σαλέπης, “trikalaola.gr,” Trikalaola.gr. Accessed: Apr. 08, 2025. [Online]. Available: <https://www.trikalaola.gr/%CE%BC%CE%BD%CE%B7%CE%BC%CE%B5%CE%AF%CE%B1-%CE%B1%CE%B3%CE%AC%CE%BB%CE%BC%CE%B1%CF%84%CE%B1-%CF%80%CF%81%CE%BF%CF%84%CE%BF%CE%BC%CE%AD%CF%82-%CF%84%CF%89%CE%BD-%CF%84%CF%81%CE%B9%CE%BA%CE%AC%CE%BB/>
- [23]J. L. Schonberger and J.-M. Frahm, “Structure-from-Motion Revisited,” in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, Jun. 2016. Accessed: Apr. 08, 2025. [Online]. Available: <https://doi.org/10.1109/cvpr.2016.445>
- [24]J. L. Schönberger, T. Price, T. Sattler, J.-M. Frahm, and M. Pollefeys, “A Vote-and-Verify Strategy for Fast Spatial Verification in Image Retrieval,” in *Lecture Notes in Computer Science*, Cham: Springer International Publishing, 2017, pp. 321–337. Accessed: Apr. 08, 2025. [Online]. Available: https://doi.org/10.1007/978-3-319-54181-5_21
- [25]E. Rupnik, M. Daakir, and M. Pierrot Deseilligny, “MicMac – a free, open-source solution for photogrammetry,” *Open Geospatial Data, Software and Standards*, vol. 2, no. 1, Jun. 2017, doi: 10.1186/s40965-017-0027-2.
- [26]T. Müller *et al.*, “Instant Neural Radiance Fields,” in *Special Interest Group on Computer Graphics and Interactive Techniques Conference Real-Time Live!*, New York, NY, USA: ACM, Aug. 2022, pp. 1–2. Accessed: Apr. 08, 2025. [Online]. Available: <https://doi.org/10.1145/3532833.3538678>

- [27]“Jawset Postshot.” Accessed: May 09, 2025. [Online]. Available: <https://www.jawset.com/>
- [28]D. Girardeau-Montaut, “CloudCompare: 3D point cloud and mesh processing software,” 2015.
- [29]B. Kerbl, G. Kopanas, T. Leimkuehler, and G. Drettakis, “3D Gaussian Splatting for Real-Time Radiance Field Rendering,” *ACM Transactions on Graphics*, vol. 42, no. 4, pp. 1–14, Jul. 2023, doi: 10.1145/3592433.
- [30] [Online]. Available: <https://www.flaticon.com/>
- [31]J. L. Schonberger and J.-M. Frahm, “Structure-from-Motion Revisited,” in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, Jun. 2016. Accessed: Jun. 06, 2025. [Online]. Available: <https://doi.org/10.1109/cvpr.2016.445>
- [32]Y. Furukawa and J. Ponce, “Accurate, Dense, and Robust Multi-View Stereopsis,” in *2007 IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, Jun. 2007, pp. 1–8. Accessed: Jun. 06, 2025. [Online]. Available: <https://doi.org/10.1109/cvpr.2007.383246>
- [33]T. Gee, “PatchMatch Stereo matching.” Accessed: Jun. 06, 2025. [Online]. Available: <https://wildboar-dev.github.io/PatchMatch/>
- [34]M. Bleyer, C. Rhemann, and C. Rother, “PatchMatch Stereo - Stereo Matching with Slanted Support Windows,” in *Proceedings of the British Machine Vision Conference 2011*, British Machine Vision Association, 2011, p. 14.1-14.11. Accessed: Jun. 06, 2025. [Online]. Available: <https://doi.org/10.5244/c.25.14>
- [35]F. Kuang, “PatchMatch Algorithms for Motion Estimation and Stereo Reconstruction,” 2017. Accessed: Jun. 06, 2025. [Online]. Available: <https://objects.scraper.bibcitation.com/user-pdfs/2025-06-06/e8754b72-7fbe-4292-9118-c7093a5210ad.pdf>

[36]B. Mildenhall, P. P. Srinivasan, M. Tancik, J. T. Barron, R. Ramamoorthi, and R. Ng, “NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis,” in *Lecture Notes in Computer Science*, Cham: Springer International Publishing, 2020, pp. 405–421. Accessed: Jun. 06, 2025. [Online]. Available: https://doi.org/10.1007/978-3-030-58452-8_24

[37]T. Müller, A. Evans, C. Schied, and A. Keller, “Instant neural graphics primitives with a multiresolution hash encoding,” *ACM Transactions on Graphics*, vol. 41, no. 4, pp. 1–15, Jul. 2022, doi: 10.1145/3528223.3530127.

[38]E. Hsiao, GitHub. Accessed: Jun. 06, 2025. [Online]. Available: <https://github.com/Ending2015a/hash-grid-encoding?tab=readme-ov-file>

[Οπισθόφυλλο. Κενή σελίδα]