



Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης
Πολυτεχνική Σχολή
Τμήμα Ηλεκτρολόγων Μηχανικών &
Μηχανικών Υπολογιστών
Τομέας Ηλεκτρονικής και Υπολογιστών

Διπλωματική Εργασία

3D ανακατασκευή αντικειμένων μέσω έμμεσων
αναπαραστάσεων με έμφαση στην κωδικοποίηση
υψηλοσυχνοτικού περιεχομένου

Εκπόνηση:
Χάρης Φίλης
ΑΕΜ: 9449

Επίβλεψη:
Καθ. Αναστάσιος Ντελόπουλος
Υπ. Δρ. Αντώνιος Καρακώττας

Δήλωση Συγγραφικής Ιδιότητας

Εγώ ο Φίλης Χάρης, δηλώνω ότι η πτυχιακή εργασία με τίτλο, 3D ανακατασκευή επιφανειών μέσω έμμεσων αναπαραστάσεων με έμφαση στην κωδικοποίηση υψηλοσυχνοτικού περιεχομένου, καθώς και η τεχνική δουλειά που παρουσιάζεται είναι δικά μου. Επιβεβαιώνω ότι:

- Αυτή η δουλειά πραγματοποιήθηκε ολοκληρωτικά ή κυρίως κατά την υποψηφιότητά μου για τίτλο προπτυχιακών σπουδών σε αυτό το πανεπιστήμιο.
- Ὁπου οποιοδήποτε μέρος αυτής της πτυχιακής εργασίας έχει προηγουμένως κατατεθεί για την απόκτηση πτυχίου ή άλλου τίτλου σε αυτό ή άλλο πανεπιστήμιο, αυτό διατυπώνεται ξεκάθαρα.
- Ὁπου έχω συμβουλευτεί την δημοσιευμένη δουλειά τρίτων, αυτό αποδίδεται ορθώς
- Ὁπου έχω παραθέσει από δουλειά τρίτων, η πηγή δίνεται πάντα. Με εξαίρεση αυτές τις παραθέσεις, αυτή η πτυχιακή εργασία είναι εξ ολοκλήρου προσωπική μου δουλειά.
- Έχω παραθέσει όλες τις κύριες πηγές βοήθειας.
- Ὁπού αυτή η πτυχιακή εργασία είναι βασισμένη σε συνεργατική δουλειά δική μου και τρίτων, έχω καταστήσει ξεκάθαρο ποια κομμάτια έχουν πραγματοποιηθεί από άλλους και πώς συνέβαλα εγώ.

Life is like riding a bicycle, to keep your balance you must keep moving
—Albert Einstein

"The task is ... not so much to see what no one has yet seen; but to think what nobody has yet thought, about that which everybody sees."
— Erwin Schrödinger

I am enough of an artist to draw freely upon my imagination. Imagination is more important than knowledge. Knowledge is limited. Imagination encircles the world.
— Albert Einstein

Ευχαριστίες

Καταρχάς, θα ήθελα να ευχαριστήσω τον καθηγητή μου Α.Ντελόπουλο που μου έδειξε στην πράξη πόσο ενδιαφέρων είναι ο κόσμος των πολυμέσων και των αλγορίθμων γραφικής και μου ενέπνευσε να βρω και εγώ το μονοπάτι μου, με διάφορες συμβουλές και καθοδηγήσεις. Παράλληλα, μεγάλο στήριγμά μου σε όλη αυτή την διάρκεια εκπόνησης της διπλωματικής, και βοηθό σε διάφορα τέλματα που προέκυψαν υπήρξε ο υποψήφιος διδάκτωρ Αντώνης Καρακώττας με τον οποίο είχα άμεση και συνεχή επικοινωνία και οι συμβουλές του ήταν καίριες.

Αναγκαίο και απαραίτητο κομμάτι στην εκπόνηση των σπουδών μου ήταν η στήριξη της οικογένειάς μου σε όλη αυτήν την εξαιρετικά δύσκολη δοκιμασία των σπουδών και ειδικά της μητέρας μου που μου στάθηκε ως βράχος στο να μην τα παρατήσω. Οπότε χωρίς αυτή την στήριξη πιθανότατα να μην βρισκόμουν στην ολοκλήρωση των σπουδών μου.

Θα ήθελα σε αυτό το σημείο να ευχαριστήσω και όλους του φίλους/φίλες που στάθηκαν στο πλευρό μου, όλο αυτό το πολυετές διάστημα φοίτησης στο τμήμα και με στήριξαν ψυχολογικά σε περιόδους που δεν φαινόταν στον ορίζοντα η αποφοίτηση αλλά και τώρα στο τέλος σε αυτό το ερευνητικό εγχείρημα για την απόκτηση διπλώματος.

Σύνοψη

Ανακατασκευή 3D Επιφανειών μέσω έμμεσων αναπαραστάσεων με χρήση Δικτύων Κωδικοποίησης

Το τελευταίο διάστημα υπάρχει άνθιση των επιτυχιών της βαθιάς μάθησης στην επεξεργασία εικόνων, καθώς και ανάγκη διαχείρισης 3D πληροφορίας σε εφαρμογές υπολογιστικής όρασης αλλά και σε μηχανές γραφικών. Μέχρι σήμερα επιστήμονες υγείας που ασχολούνται με τρισδιάστατα μοντέλα οστών/κοιλοτήτων, κατασκευαστές μηχανολογικών εξαρτημάτων, καλλιτέχνες, σχεδιαστές παιχνιδιών ή τεχνικοί οπτικών συστημάτων εφέ ταινιών που θέλουν να αποτυπώσουν εύκολα πραγματικά αντικείμενα στο υπολογιστή στις περισσότερες περιπτώσεις χρησιμοποιούν κλασσικούς τρόπους φωτογραφιμετρίας για να δημιουργήσουν 3D σαρώσεις (*3D scans*) αντικειμένων, μια διαδικασία που εκτός από χρονοβόρα είναι και άπλοτη ως προς το πώς αναπαριστάται στη μνήμη το τελικό αποτέλεσμα. Σήμερα τα υψηλής ακρίβειας τρισδιάστατα μοντέλα είναι απαραίτητο να αποθηκεύονται στον υπολογιστή με λιγότερους δυνατούς χωρικούς συμβιβασμούς αλλά και η διαδικασία παραγωγής τους να είναι πιο διαισθητική και απλή παίρνοντας λίγες φωτογραφίες τους.

Στην παρούσα διπλωματική εργασία γίνεται μια προσπάθεια βελτίωσης των μεθόδων 3D ανακατασκευής μέσω αντίστροφης αποτύπωσης με χρήση νευρωνικών δικτύων. Αυτά τα βαθιά νευρωνικά δίκτυα, προσπαθούν να προσεγγίζουν ένα έμμεσο πεδίο γεωμετρίας και εμφάνισης μιας συμπαγούς επιφάνειας μέσω συνάρτησης προσημασμένης απόστασης, *SDF* (*signed distance function*). Τα δίκτυα αυτά δεν είναι σε θέση να συλλάβουν το περιεχόμενο χωρικής συχνότητας της επιφάνειας που αποτυπώνεται σε πολλαπλές προβολές (εικόνες που χρησιμοποιούνται για επίβλεψη). Συνεπώς, προτείνονται μέθοδοι που μετασχηματίζουν τον τρόπο λειτουργίας αυτών των δικτύων επιτρέποντας την αποτύπωση λεπτομερειών αλλά και τη σύλληψη υψηλοσυχνοτικών μεταβολών σε επύπεδο γεωμετρίας και εμφάνισης των επιφανειών.

Προτείνεται μια αρχιτεκτονική βαθιάς μάθησης, που βασίζεται σε προηγμένες τεχνικές επεξεργασίας σήματος. Για παράδειγμα, χρησιμοποιείται η ιδέα του μετασχηματισμού κυματιδίων (*Wavelet Decomposition*) μέσω Νευρωνικών Δικτύων αποθήκευσης συχνοτικού περιεχομένου (*Neural Filter Banks* [1]) ώστε να αποδοθεί η καλύτερη δυνατή είσοδος για το νευρωνικό δίκτυο. Σε αυτή την προσπάθεια νιοθετούνται καινοτόμες ιδέες σχετικά με την κωδικοποίηση όπως αυτή που προτείνεται το *Neural Graphics Primitives*[2], καταδεικνύοντας το πώς μπορεί μια κωδικοποίηση κατακερματισμού του χώρου σε πολλά επίπεδα ανάλυσης (*Multi-Resolution hash encoding*) να αποτελέσει χωρική κωδικοποίησή του. Έτσι αυτό επιτρέπει στο δίκτυο να συλλάβει μεταβολές της γεωμετρίας και της εμφάνισης κατά τη σύνθεση μεγάλων σκηνών. Τέλος, γίνεται περαιτέρω επέκταση δίνοντας δυνατότητα προσοχής στο ποιοι χάρτες συχνοτικών χαρακτηριστικών των όψεων είναι πιο αποτελεσματικοί στην καλύτερη ανακατασκευή. Αυτό πραγματοποιείται κάνοντας διαμόρφωση και αποδιαμόρφωση των χαρτών χαρακτηριστικών της εισόδου των δικτύων έμμεσης αναπαράστασης (άτλαντας χαρακτηριστικών απεικόνισης).

Abstract

Implicit Neural Surface Reconstruction with 3D Neural High Frequency Embeddings

Recently, there has been an overflow of deep learning success in image processing, as well as a need to manage *3D* information in computer vision applications and graphics machines. To date, health scientists working with *3D* bone/cavity models, mechanical component manufacturers, artists, game designers or film visual effects technicians who want to easily capture real objects on the computer in most cases use classical photogrammetry methods to create *3D* scans of objects, a process that is both time-consuming and greedy in terms of how the result is represented in memory. Nowadays, accurate *3D* models need to be stored in the computer with the least possible spatial compromises, and the process of producing them should be intuitive and simple, done just taking a few pictures of them.

In the present diploma thesis, an attempt is made for the improvement of the input encoding using deep neural networks on the three-dimensional implicit surface reconstruction models, i.e., networks representing *3D* space through implicit representations. These deep neural networks (*Mostly MLPs*) attempt to approximate the geometry field and the appearance of a watertight surface through signed distance functions (*SDFs*). This process is challenging (*requiring a lot of training time*), especially when these networks are not able to capture the high frequency content of the surface that is being reconstructed, using multi-view synthesis (*real images are used for supervision of the network's weights*). Therefore, methods are proposed that transform the way the data is processed, allowing both the capture of high frequency details, spatial features in terms of the surface geometry and appearance for in general high fidelity novel view synthesis and fast convergence of the model.

Thus, a deep learning architecture is proposed, based on advanced signal processing concepts such as wavelet decomposition via Neural Fourier filter Banks [1] to construct the best possible input encoding for the neural surface reconstruction network. In this effort, innovative ideas about encoding the input are proposed, as in Neural Graphics Primitives [2], demonstrating the effect that multi-resolution hash grid encoding has on spatial embedding of the sparse high frequency content. This allows capturing faster the changes in geometry and appearance instantly during surface composition. Finally, further development is made by giving the network the ability to pay attention on which frequency/spatial feature maps of the input are most effective for better reconstruction using modulation and demodulation of the existing feature maps in every grid resolution.

Keywords: Photogrammetry, 3D Neural Surface Reconstruction, Signed Distance Functions, Multi-Resolution Hash Grid Encoding, Neural Fourier Filter Banks, Feature Modulation/Demodulation

Περιεχόμενα

Ευχαριστίες	iv
Περίληψη	v
Abstract	vi
Ακρωνύμια/Ορολογίες	xiv
1 Εισαγωγή	1
1.1 Διατύπωση του Προβλήματος 3D Ανακατασκευής	1
1.2 Διατύπωση προβλήματος 3D ανακατασκευής επιφανειών με έμμεση μορφή ως πρό- βλημα βελτιστοποίησης με νευρωνικά δίκτυα	2
1.3 Σκοπός της διπλωματικής εργασίας	2
1.4 Δομή Διπλωματικής Εργασίας	3
2 Θεωρητικό Υπόβαθρο	5
2.1 Αναπαραστάσεις Γραφικών	5
2.1.1 Άμεσες/Ρητές Αναπαραστάσεις 3D Γραφικών	5
2.1.2 Παραμετρικές Αναπαραστάσεις Γραφικών	6
2.1.3 Έμμεσες αναπαραστάσεις Γραφικών	7
2.2 Μέθοδοι Αποτύπωσης 3D γραφικών σε εικόνα	9
2.2.1 Αλγόριθμος Ιχνηλάτισης Ακτίνων Ray Tracing	10
2.2.2 Απόδοση Όγκου Αλγόριθμος Βηματισμού Πάνω στην Ακτίνα Ray Marching .	11
2.2.3 Διαφορίσιμη αποτύπωση έμμεσων πεδίων <i>Implicit Differentiable Rendering</i>	13
2.3 Εκπαιδευόμενο Πεδίο Προσημασμένης Απόστασης - Neural SDF	14
2.3.1 Επιφάνειες που ορίζονται από πεδία Απόστασης	14
2.3.2 Διαφορίσιμες πολλαπλότητες και Γενικευμένο Θεώρημα Αναπαράστασης .	15
2.3.3 Θεωρία πίσω από τον νευρωνικό έλεγχο διαφορίσιμων πολλαπλοτήτων(<i>neural level sets</i>)[17], [14]	16
2.3.4 Απόδοση Γραφικών σε Έμμεσες Διαφορίσιμες Επιφάνειες	17
2.4 Βαθιά μάθηση	19
2.4.1 Αρχιτεκτονική	19
2.4.2 Βασικό Πρόβλημα Ισοζυγίου Bias-Variance	21
2.4.3 Αρχικοποίηση Βαρών - Θεωρία Ομαλοποίησης	22
2.4.4 Αντικειμενική Συνάρτηση Βελτιστοποίησης - Loss Function	22
2.5 Δίκτυα & Μηχανισμοί Κωδικοποίησης 3D Εισόδου DeepSDF Frequency Feature Embeddings	23
2.5.1 Κωδικοποίηση μέσω συναρτήσεων εφαπτόμενου νευρωνικού πυρήνα Fourier - Fourier Feature Neural Tangent Kernel	23

2.5.2	Χρήση χωρικής πληροφορίας στον εφαπτόμενο νευρωνικό πυρήνα συναρτήσεων - Positional Encoding	24
2.5.3	Κωδικοποίηση με χρήση Σφαιρικών Αρμονικών [25]	25
2.5.4	Τρισδιάστατος χωρικός κατακερματισμός- Multi-resolution Hash 3D Encoding	26
2.5.5	Θυρίδες Κατακερματισμένων φίλτρων Fourier NFFB [1]	27
2.6	Διαμόρφωση Δικτύων Κωδικοποίησης Υψηλοσυχνοτικού Περιεχομένου	29
3	Επισκόπηση της Ερευνητικής Περιοχής	31
3.1	Τεχνικές Βαθιάς Μάθησης, Τρισδιάστατης Ανακατασκευής και Κωδικοποίησης	31
3.2	MVSR-IDR	33
3.3	Let Neural Nets learn high frequency content in low dimensional domain	35
3.4	Instant Neural Graphics Primitives Multi-Resolution Hash Encoding	36
3.5	Neural Fourier Filter Banks	37
3.6	StyleGAN Modulation / Demodulation, Attention is all you need	38
4	Μεθοδολογία	41
4.1	Δεδομένα Ελέγχου - Προ επεξεργασία	41
4.1.1	Σύνολο Δεδομένων DTU, <i>DTU Robot Image Datasets</i> [81]	41
4.1.2	Εξαγωγή Δεδομένων Τρισδιάστατης Στερεομετρίας σε ένα αντικείμενο Scene Dataset	42
4.2	Δίκτυο Γεωμετρικής Ανακατασκευής και Χρωματικής Αποτύπωσης	42
4.2.1	Συμβολισμοί	42
4.2.2	Διαφορίσιμη νευρωνική αποτύπωση σε βήματα	43
4.2.3	Διαφορική μορφή διασταύρωσης ακτίνας με σημείο της γεωμετρίας	45
4.2.4	Υπολογισμός Πεδίου Ακτινοβολίας - Rendering Network	47
4.2.5	Διαδικασία Εκπαίδευσης	48
4.3	Συχνοτική Κωδικοποίηση Εφαπτόμενου Πυρήνα Συναρτήσεων Fourier	50
4.4	Χωρικός Κατακερματισμός Πολλαπλής Ανάλυσης	52
4.5	Θυρίδες Βαθιάς Συχνοτικής/Χωρικής Αποσυχέτισης και Κωδικοποίησης	54
4.6	Διαμόρφωση/Αποδιαμόρφωση Χαρακτηριστικών Σήματος	56
4.7	Υλοποιημένα Δίκτυα και Παράμετροι	56
5	Πειράματα - Αποτελέσματα	59
5.1	Πειραματική Φάση	59
5.1.1	Λεπτομέρειες DTU Συνόλου δεδομένων	59
5.1.2	Πειράματα	59
5.1.3	Καμπύλες Εκπαίδευσης Αλγορίθμων	61
5.1.4	Σχόλια για την ταχύτητα των αλγορίθμων	61
5.1.5	Σχόλια για την αξιολόγηση	62
5.2	Παρουσίαση Φωτογραφικών Αποτελεσμάτων	62
5.2.1	Φωτογραφίες Ανακατασκευής ανά Εποχή Εκπαίδευσης και στην Αξιολόγηση	64
5.2.2	Αποτύπωσης Προσέγγισης Δυαδικής μάσκας - Μάσκα «βάθους» ανά εποχή	68
5.3	Αξιολόγηση 3D ανακατασκευασμένων μοντέλων	70
5.4	Μετρικές Αξιολόγησης Αλγορίθμων	70
5.4.1	PSNR - Λόγος σήματος κορυφής προς θόρυβο	70
5.4.2	SSIM(Structure Similarity Index Measure) - Μετρική Δείκτη Δομικής Ομοιότητας Εικόνων	71

5.4.3	LPIPS(Learned Perceptual Image Patch Similarity)- Αντιληπτική Ομοιότητα Μπαλωμάτων Εικόνας σε μοντέλα αναγνώρισης εικόνας	72
5.5	Μετρικές Αξιολόγησης Γεωμετρικής Ομοιότητας	72
5.5.1	Παρουσίαση Γεωμετρίας Μοντέλων	73
5.5.2	Πίνακες Απόστασης Chamfer	75
6	Συμπεράσματα & Μελλοντικές επεκτάσεις	79
6.1	Συμβολή της Εργασίας - Συμπεράσματα	79
6.2	Μελλοντικές Βελτιώσεις/Επεκτάσεις	79
7	Παράρτημα	83
	Βιβλιογραφία	96

Σχήματα

2.1	Διαφορετικές Μορφές αναπαραστάσεων Γραφικών	5
2.2	3D πλέγμα τριγώνων που αναπαριστά δελφίνι	6
2.3	Πολλαπλότητες ορίζουν έμμεσα ισομετρικές επιφάνειες χαμηλότερης διάστασης σε όριο απόφασης	7
2.4	Οκταδικό Δέντρο	8
2.5	Πλεονεκτήματα και μειονεκτήματα Γραφικών Αναπαραστάσεων (Πηγή [12])	9
2.6	Σύστημα Κάμερας σε μηχανές Γραφικών που εκτελούν Ιχνηλάτιση Ακτίνας	10
2.7	Πολλαπλά Ιχνη ανάκλασης και μετάδοσης ακτίνας που προέρχεται από το σημείο αναφοράς προβολής, περνάει μέσα από την θέση του Πηγή:[11]	10
2.8	Το διάγραμμα ροής του <i>Ray Tracing</i> Πηγή: Ray Marching Video	11
2.9	Χρήση Βαδίσματος Ακτίνας για την εύρεση Σημείων τομής με δομές του χώρου με χρήση σφαιρικού έλεγχου προσημασμένης απόστασης(Sphere Tracing) - 2D	12
2.10	Ομοιομορφισμός	14
2.11	Ορισμός <i>SDF</i> πεδίου. Πηγή [6]	15
2.12	Διάγραμμα που δείχνει τα διανύσματα που ορίζουν την BRDF Πηγή [19]	18
2.13	Νευρωνικό δίκτυο Perceptron τεσσάρων κρυφών στρωμάτων	19
2.14	Τρεις συνήθεις συναρτήσεις ενεργοποίησης	20
2.15	Οπτική Παρουσίαση Εφαρμογής Fourier Feature Mapping σε Δισδιάστατες Εικόνες, Πηγή:[(24)]	23
2.16	Ανακατασκευή εικόνας με Κωδικοποίηση Ευρύ Εφαπτόμενου Πυρήνα Συναρτήσεων Fourier (μεγάλο σ κατανομής δειγματοληψίας συχνοτήτων),Πηγή [24]	24
2.17	Ανακατασκευή εικόνας με Κωδικοποίηση Στενού Εφαπτόμενου Πυρήνα Συναρτήσεων Fourier (μικρό σ), Πηγή [24]	24
2.18	Ανακατασκευή Σήματος με χρήση σφαιρικών αρμονικών, πηγή [25](Robin Green)]	26
2.19	Αναπαράσταση Wavelet κατακερματισμένης κωδικοποίησης υψηλών και χαμηλών συχνοτήτων σε μια και δύο διαστάσεις με χρήση φίλτρων, πηγή [1]	28
2.20	Αυτό-Διαμόρφωση MLP δικτύου, πηγή [29]	29
3.1	Συνολικό Δίκτυο IDR, Πηγή [6]	34
3.2	Βελτίωση αποτελεσμάτων σε Coordinate Based MLPs για μια ποικιλία υψηλοσυχνοτικού περιεχομένου εφαρμοσμένο σε χαμηλοσυχνοτική εφαρμογή της ανακατασκευής εικόνων και 3D σχημάτων, Πηγή [24]	35
3.3	Προτεινόμενη αρχιτεκτονική NFFB, Πηγή [1]	37
3.4	Διαμόρφωση και Αποδιαμόρφωση βαρών στο αναθεωρημένο StyleBlock-StyleGan2, Πηγή [28]	39
4.1	DTU MVS σύνολο δεδομένων	41
4.2	Νευρωνική Αποτύπωση - Διαδικασία	43

4.3	2D ανάλογο end-to-end αρχιτεκτονικής δικτύου με χρήση Multi Resolution Hash Grid Encoding	53
4.4	3D Hashgrid MLP IDR Proxy	53
4.5	Αρχιτεκτονική δικτύου 3D Κωδικοποίησης NFFB	55
5.1	Σύγκριση Σύγκλισης Αλγορίθμων στο Σφάλμα Εκπαίδευσης	61
5.2	Καμπύλες Σφαλμάτων εκπαίδευσης NFFB, σε όλη την διάρκεια της εκπαίδευσης	62
5.3	Σηματοθορυβική σύγκριση αλγορίθμων κωδικοποίησης στην σκηνή της κουκουβάγιας	71
5.4	Οπτική αποτύπωση Chamfer απόστασης	77
1	Γενικευμένα Πολυώνυμα Bernstein 5^{ou} βαθμού B_k^5 στο $[0,1]$. Πηγή [10]	84
2	Βασική Γραμμή Παραγωγής Γραφικών	84
3	Σύστημα προοπτικής Προβολής [86]	85
4	εξωτερικές και εξωτερικές παράμετροι κάμερας σε μαθηματική μορφή συστήματος προβολής	85
5	Διανύσματα, Διάχυτη Ανάκλαση, Κατοπτρικό Ανάκλαση, Μοντέλο Ανάκλασης Phong, Πηγή [89]	86
6	Απεικόνιση Lookup Table ανάλυσης 8, αλγορίθμου Marching Cubes	91

Πινάκες Αποτελεσμάτων

3.1 Πίνακας Ερευνών πάνω στις 3D Αναπαραστάσεις [7]	32
5.1 Αποτελέσματα 3D Ανακατασκευής Σκηνής 65	64
5.2 Αποτελέσματα 3D Ανακατασκευής Σκηνής 122	65
5.3 Αποτελέσματα 3D Ανακατασκευής Σκηνής 110	66
5.4 Αποτελέσματα 3D Ανακατασκευής Σκηνής 114	67
5.5 Αποτελέσματα 3D Ανακατασκευής Σκηνής 65 - Μάσκα Βάθους	68
5.6 Αποτελέσματα 3D Ανακατασκευής Σκηνής 122 - Μάσκα Βάθους	68
5.7 Αποτελέσματα 3D Ανακατασκευής Σκηνής 110 - Μάσκα Βάθους	69
5.8 Αποτελέσματα 3D Ανακατασκευής Σκηνής 114 - Μάσκα Βάθους	69
5.9 Αποτελέσματα 3D Ανακατασκευής Σκηνής 122 - SHencoder	70
5.10 Αποτελέσματα αξιολόγησης με τη μετρική PSNR	71
5.11 Αποτελέσματα αξιολόγησης με τη μετρική SSIM	72
5.12 Αποτελέσματα αξιολόγησης με τη μετρική LPIPS	72
5.13 Πίνακας Γεωμετρίας 3D μοντέλων	74
5.14 Πίνακας Γεωμετρίας Κρυφής Όψης 3D μοντέλων - Έλεγχος Γενίκευσης	75
5.15 Πίνακας Μετρικών Μέσης Απόστασης Chamfer	75

Αλγόριθμοι

1	Απλοική μορφή Ray Marching	12
2	Sphere Tracing Algorithm	13
3	HashGrid3D-Neural Coordinate Encoding	89
4	Ray Marching Algorithm using LSTM step length Prediction	90
5	3D Marching Cubes Algorithm	91

Ακρωνύμια/Ορολογίες Εγγράφου

Παρακάτω παρατίθενται ορισμένα από τα πιο συχνά χρησιμοποιούμενα ακρωνύμια και ορολογίες της παρούσας διπλωματικής εργασίας¹:

3D/3Δ	→ 3 Dimensions 3 Διαστάσεις
2D/2Δ	→ 2 Dimensions 2 Διαστάσεις
Voxel	→ Volumetric Pixel Ογκομετρικό Εικονοστοιχείο
HF	→ High Frequency High Fidelity
CAGD	→ Computer Aided Geometric Design
Ray Tracing	→ Ιχνηλάτιση Ακτινών
Ray Marching	→ Αλγόριθμος Βάδισης πάνω στην ακτίνα
Sphere Tracing	→ Σφαιρικός έλεγχος αλγορίθμου βάδισης
BRDF	→ Bidirectional Reflectance Distribution Function
NN	→ Neural Network
ANN	→ Artificial Neural Network
DL	→ Deep Learning
DNN	→ Deep Neural Network
MLP	→ Multilayer Perceptron
CNN	→ Convolutional Neural Network
RNN	→ Recurrent Neural Network
ResNet	→ Residual Neural Network
GAN	→ Generative Adversarial Network
LSTM	→ Long Short-Term Memory Network
CPU	→ Central Processing Unit
GPU	→ Graphics Processing Unit
SDF	→ Signed Distance Field
IDR	→ Implicit Differentiable Rendering
NEUS	→ Neural Surface Reconstruction
NeRF	→ Neural Radiance Fields
NFFB	→ Neural Fourier Filter Banks
XOR	→ Exclusive OR
MOD	→ Modulation
StyleMod	→ Style Modulation
AdaIN	→ Adaptive Instance Normalization

¹Οι αγγλικοί όροι μεταφράζονται στα σημεία που χρησιμοποιούνται

1

Εισαγωγή

1.1 Διατύπωση του Προβλήματος 3D Ανακατασκευής

Η ανακατασκευή τρισδιάστατων αναπαραστάσεων αναφέρεται σε όλες τις μεθόδους που έχουν ως στόχο να ανακτήσουν την πληροφορία της γεωμετρίας και της εμφάνισης ενός 3D πραγματικού αντικειμένου.

Το πρόβλημα της 3D ανακατασκευής λόγω των πολλών εφαρμογών της, έχει ερευνηθεί εκτενώς με διάφορες προσεγγίσεις. Σε αυτές τις μεθόδους συμμετέχουν οι ενεργές μέθοδοι που αναφέρονται στην αριθμητική αντιμετώπιση του προβλήματος. Αυτό επιχειρείται, μέσα από έναν χάρτη βάθους (*depth map*) ο οποίος αποκτάται χρησιμοποιώντας μεθόδους όπως το «δομημένο φως» (*structured light*). Ενδεικτικές εφαρμογές, οι χρονοβόρες τρισδιάστατες σαρώσεις 3D *scans*.

Από την άλλη πλευρά, υπάρχουν οι παθητικές μέθοδοι που δεν παρεμβαίνουν άμεσα στο ανακατασκευασμένο αντικείμενο. Αντίθετα, χρησιμοποιούν πληροφορίες μόνο από την ακτινοβολία που ανακλάται από τα αντικείμενα και αποτυπώνεται στον αισθητήρα της κάμερας. Συνεπώς, τα δεδομένα που χρησιμοποιούν είναι μόνο οι πληροφορίες που παρέχει μια φωτογραφία. Σε αυτόν το τομέα, υπάρχουν κλασσικές μέθοδοι που εξετάζουν την απόκτηση 3D σχήματος από την σκίαση που εφαρμόζεται *Shape-from-Shading*, άλλες από την υφή που παραμορφώνεται κατά την προβολή *Shape-from-texture*, και μία άλλη πιο εκλεπτυσμένη προσέγγιση που αναφέρεται στην φωτομετρική στερεογραφία ή φωτογραμμετρία. Η τελευταία ανακτά την πληροφορία βάθους από φωτογραφίες του αντικειμένου σε διαφορετικές συνθήκες φωτισμού. Σε σύγκριση με τις ενεργές μεθόδους αυτές οι μέθοδοι έχουν πολύ περισσότερα πεδία εφαρμογής.

Όλες οι προηγούμενες μέθοδοι αντιστοιχούν σε πρακτικές που χρησιμοποιούνται σήμερα στους τομείς της υγείας και της βιομηχανίας, αλλά είναι εξαιρετικά αργές και δεν μπορούν να χρησιμοποιηθούν σε πραγματικό χρόνο. Για αυτόν τον λόγο, η στερεοπτική με υπολογιστές (*Computer Stereo Vision*) έχει αναπτύξει αλγορίθμους που χρησιμοποιούν ταυτόχρονα εικόνες από διάφορες θέσεις καμερών αντιστρέφοντας το πρόβλημα. Συγκεκριμένα προσπαθούν να υπολογίσουν τις τρισδιάστατες συντεταγμένες του αντικειμένου κάνοντας τεχνικές αντίστροφες από την προβολή που κάνει η κάμερα. Για να γίνει αυτό, χρησιμοποιούνται μέθοδοι αποτύπωσης γραφικών σε εικόνα που λαμβάνουν

υπόψιν τον ολικό φωτισμό της σκηνής με μεθόδους όπως η ιχνηλάτιση ακτίνων των οποίων η ρίψη ξεκινά από την κάμερα και αλγόριθμοι που εκτιμούν την τομή των ακτίνων που περνούν από την εικόνα πάνω στο 3D μοντέλο. Αυτές οι μέθοδοι, απαιτούν τον υπολογισμό των σημείων τομής ακτίνας επιφάνειας, τον υπολογισμό του ανακλώμενου χρώματος και ταυτόχρονα η αποτυπωμένη εικόνα πρέπει να μοιάζει με την πραγματική φωτογραφία.

1.2 Διατύπωση προβλήματος 3D ανακατασκευής επιφανειών με έμμεση μορφή ως πρόβλημα βελτιστοποίησης με νευρωνικά δίκτυα

Στο πεδίο της Υπολογιστικής Όρασης (*Computer Vision*) και των 3D Γραφικών οι επιφάνειες συνήθως αναπαριστώνται με άμεσους τρόπους. Τέτοιες αναπαραστάσεις είναι είτε πολυγωνικές (συνήθως τρίγωνα με σημεία στον 3D χώρο για κορυφές), είτε ογκομετρικές αναπαραστάσεις (νέφη σημείων *Point Clouds*).

Μετά από διάφορα εγχειρήματα βελτίωσης του προβλήματος της ανακατασκευής, η επιστημονική κοινότητα κατέληξε πως οι άμεσες αναπαραστάσεις όπως ένα ακατέργαστο νέφος σημείων αυξάνουν την πολυπλοκότητα των αλγορίθμων.

Σε αυτό το πλαίσιο προτάθηκαν οι έμμεσες αναπαραστάσεις γραφικών οι οποίες αλλάζουν το πρόβλημα από την ανακατασκευή μια τρισδιάστατης επιφάνειας, στον υπολογισμό ενός πεδίου προσημασμένης απόστασης *SDF*. Το πεδίο αυτό, αντιστοιχεί κάθε σημείο τους χώρου σε μια προσημασμένη απόσταση από την επιθυμητή προς ανακατασκευή επιφάνεια. Στην συνέχεια με έναν αλγόριθμο ισοδυναμικής επιφάνειας εξάγεται η άμεση αναπαράσταση από το πεδίο (βλ. K7 *Marching Cubes*)

Πώς συνδέεται η τεχνητή νοημοσύνη;

Με την ανάπτυξη των τεχνικών εκπαίδευσης των τεχνητών νευρωνικών δικτύων (*ANNs*) και την απαγκίστρωση μας από το γενικό θεώρημα προσέγγισης (*General Approximation Theorem*), είμαστε πλέον στην άνθιση των βαθιών νευρωνικών δικτύων (*DNNs*). Συνεπώς το πρόβλημα εκτίμησης των έμμεσων αναπαραστάσεων γραφικών μπορεί να λυθεί με νευρωνικά δίκτυα.

Οι μέθοδοι που προτάθηκαν δίνουν εκλεπτυσμένες μεθόδους για την αποτύπωση των γραφικών σε εικόνα με διαφορίσιμη μάλιστα μορφή, μετατρέποντάς το σε πρόβλημα βελτιστοποίησης που μπορούν να το χειριστούν τα βαθιά νευρωνικά δίκτυα. Παρ' όλα αυτά τα δίκτυα συντεταγμένων, που αναπαριστούν *SDF*, μέχρι σήμερα δεν είναι δίκτυα που εκπαιδεύονται γρήγορα και παρουσιάζουν προβλήματα σε περιοχές με υψηλοσυχνοτικές μεταβολές της γεωμετρίας.

Αυτό το πρόβλημα αποσκοπεί να ερευνήσει η παρούσα εργασία προτείνοντας μεθόδους που βασίζονται στην υψηλοσυχνοτική κωδικοποίηση εισόδου.

1.3 Σκοπός της διπλωματικής εργασίας

Η έρευνα που πραγματοποιείται, είναι μια από τις πρώτες εκτενείς προσπάθειες επίλυσης του προβλήματος της αξιόπιστης ανακατασκευής 3D επιφανειών, με την έννοια του φωτορεαλισμού. Το πρόβλημα αντιμετωπίζεται, μέσα από εκπαίδευση έμμεσων αναπαραστάσεων και την χρήση λίγων εικόνων, διαφόρων όψεων, ως μέσο επίβλεψης στην διαδικασία αντίστροφης νευρωνικής αποτύπωσης.

Τα νευρωνικά δίκτυα εκ φύσεως αδυνατούν να συλλάβουν το υψηλοσυγχνοτικό περιεχόμενο της ακατέργαστης πληροφορίας, κάτι που πηγάζει από τον τρόπο εκπαίδευσης και λειτουργίας τους. Επομένως, είναι από τις λίγες εργασίες που δίνει κυρίαρχη έμφαση στην κωδικοποίηση των έμμεσων αναπαραστάσεων της εισόδου που δίνονται στο δίκτυο. Κύριος στόχος αυτού, η ανακατασκευή να έχει αξιόπιστα αποτελέσματα σε υψηλοσυχνοτικές περιοχές αλλά και ταχύτερη σύγκλιση.

Στην επιστημονική κοινότητα δεν έχει ακόμα διακριθεί ποια είναι η ιδανική κωδικοποίηση περιεχομένου για την συγκεκριμένη εφαρμογή. Οι περισσότερες εργασίες αντιμετωπίζουν το πρόβλημα ως πρόβλημα υπερεκπαίδευσης πεδίων απόστασης. Σύγχρονες μελέτες δείχνουν όμως, πως υπάρχει χώρος επίλυσης, με συνδυασμό μεθόδων κωδικοποίησης.

Συνεπώς, στην παρούσα διπλωματική εργασία εξετάζεται η χρήση κωδικοποίησεων και η σύλληψη υψηλοσυχνοτικού περιεχομένου για την εκπαίδευση νευρωνικών δικτύων που χρησιμεύουν ως εκτιμητές συνεχών συναρτήσεων(*function approximators*) πάνω στο πρόβλημα 3D ανακατασκευής μοντέλων, φωτογραφημένων αντικειμένων. Αυτή η εκτενής χρήση δικτύων κωδικοποίησης μετασχηματίζει την είσοδο και οδηγεί το δίκτυο να συγκλίνει πιο γρήγορα σε ακριβείς εκτιμήσεις ανεξαρτήτως γεωμετρικής φύσης του 3D μοντέλου που ανακατασκευάζεται. Ταυτόχρονα φαίνεται να γενικεύει καλύτερα το δίκτυο ενώ αναπαριστά καλύτερα μεγάλες σκηνές.

1.4 Διπλωματικής Εργασίας

Η διπλωματική εργασία ασχολείται με δύο αντικείμενα. Αρχικά το κομμάτι του δικτύου που είναι υπεύθυνο για την εκτίμηση της γεωμετρίας της επιφάνειας στον τρισδιάστατο χώρο μέσω πεδίων προσημασμένης απόστασης (*SDF*). Αυτό, πραγματοποιείται μέσα από την έμμεση μορφή αντίστροφης αποτύπωσης διαφορίσιμων ισομετρικών επιφανειών (*Implicit Differentiable Rendering*), με ξεχωριστό τον υπολογισμό του πεδίου ακτινοβολίας (*Rendering Network*).

Δεύτερο και πεδίο έρευνας της παρούσας εργασίας, είναι οι αλγόριθμοι και τα δίκτυα βαθιάς κωδικοποίησης των έμμεσων πεδίων αναπαράστασης γεωμετρίας και φωτισμού.

Επομένως, πρώτα γίνεται μια ανάλυση του θεωρητικού υπόβαθρου πάνω στις γραφικές αναπαραστάσεις στους υπολογιστές και στις μεθόδους που χρησιμοποιούνται για την την αποτύπωση και κωδικοποίηση αυτών. Στην συνέχεια, παρουσιάζονται οι πιο πρόσφατες προσπάθειες από την επιστημονική κοινότητα, στην προσέγγιση του προβλήματος με αντίστοιχες ερευνητικές εργασίες. Εργασίες, σε οποίες βασίζεται και η παρούσα εργασία, χρησιμοποιώντας παράλληλα και μοντέλα για την απόδοση και του χρώματος των σκηνών ξεχωριστά από την γεωμετρία. Στην συνέχεια αναλύεται το θεωρητικό υπόβαθρο που απαιτείται, δίνεται η υλοποίηση που έγινε πάνω στο δίκτυο κωδικοποίησης της εισόδου των δικτύων ανακατασκευής επιφάνειας (το οποίο μπορεί να εφαρμοστεί και σε δίκτυα απόδοσης όγκου [3], [4]), αλλά και οι μετατροπές στα ίδια τα δίκτυα εκτίμησης της ισομετρικής επιφάνειας του 3D μοντέλου. Συνεπώς η εργασία διαρθρώνεται ως εξής:

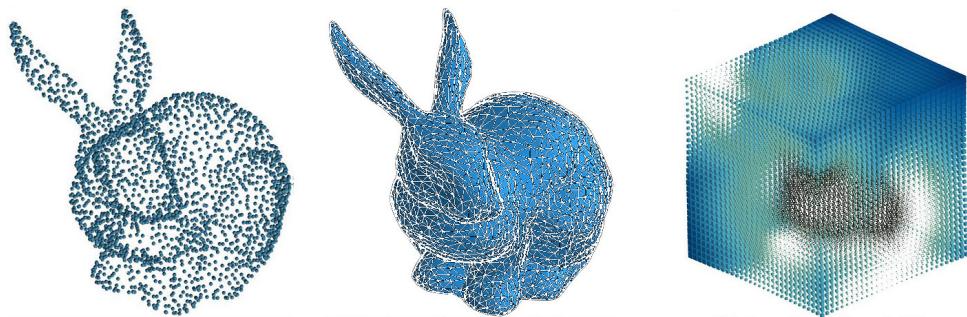
- **Κεφάλαιο 2:** Περιγράφονται θεωρητικά στοιχεία στα οποία βασίστηκαν οι υλοποιήσεις. Γίνεται εισαγωγή στην επιστήμη της γραφικής υπολογιστών, της υπολογιστικής όρασης καθώς και της βαθιάς μάθησης και στην συνέχεια δίνεται βάση, στο πώς αυτές οι επιστήμες συνδυάζονται στο κομμάτι της ανακατασκευής μοντέλων. Επιπλέον, παρουσιάζεται το θεωρητικό υπόβαθρο αλγορίθμων αποτύπωσης 3D σκηνής σε εικόνα από έμμεσες επιφάνειες(*implicit differentiable rendering* [5–7]), ενσωμάτωσης πληροφορίας και εκτίμησης σημείων μέσω πολυ-στρωματικών νευρωνικών δικτύων.
- **Κεφάλαιο 3** Γίνεται ανασκόπηση της ερευνητικής περιοχής που αφορά την ανακατασκευή 3D σκηνής, αλλά και των τεχνικών με τις οποίες δίνεται δυνατότητα στα δίκτυα εκτίμησης 3D σκηνής να μάθουν το υψηλοσυχνοτικό περιεχόμενο των δεδομένων.

- **Κεφάλαιο 4:** Πλήρης περιγραφή των υλοποιήσεων διαφόρων μοντέλων (*Implicit Differential Network* [6], *Rendering Network* [3], *Multi-Resolution Hash Embedding* [2], *NFFB*[1], *Style Modulation* [8, 9] με βάση τις τεχνικές που ακολουθήθηκαν για την βελτιστοποίηση νευρωνικών ανακατασκευής 3D σκηνών.
- **Κεφάλαιο 5:** Παρουσιάζεται η μεθοδολογία των πειραμάτων και τα αποτελέσματα.
- **Κεφάλαιο 6:** Παρουσιάζονται τα τελικά συμπεράσματα, ενώ ταυτόχρονα αναφέρονται τα προβλήματα που προέκυψαν και προτείνονται θέματα για μελλοντική μελέτη, αλλαγές και επεκτάσεις.
- **Παράρτημα κ.7:** Αφήνονται για τον αναγνώστη πολλές σημαντικές πληροφορίες που δεν μπορούσαν να ενταχθούν άμεσα στο βασικό κείμενο της εργασίας

2

Θεωρητικό Υπόβαθρο

2.1 Αναπαραστάσεις Γραφικών

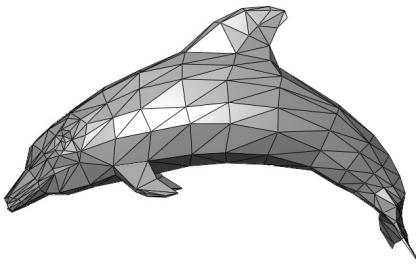


Σχήμα 2.1: Διαφορετικές Μορφές αναπαραστάσεων Γραφικών

2.1.1 Άμεσες/Ρητές Αναπαραστάσεις 3D Γραφικών

3D - Αναπαραστάσεις ως πλέγματα τριγώνων

Στον χώρο της γραφικής υπολογιστών στις περισσότερες περιπτώσεις οποιοδήποτε γραφικό στοιχείο έχει ως δομική μονάδα αναπαράστασης τα κυρτά πολύγωνα. Αυτό, ονομάζεται ρητή αναπαράσταση γραφικών(*explicit representation*). Η συγκεκριμένη τακτική ακολουθείται, καθώς η κυρτότητα του σχήματος δίνει την δυνατότητα διάκρισης του χώρου του οποίο καταλαμβάνει και κατά συνέπεια αυτό βοηθάει σε αλγορίθμους σκίασης. Οι αλγόριθμοι σκίασης, απαιτούν την επίγνωση των ορίων κάθε στοιχειώδους επιφάνειας για τον χρωματισμό της. Κατά σύμβαση η συνηθέστερη επιλογή κυρτού πολυγώνου είναι το τρίγωνο χάρη στα πλεονεκτήματα της απλότητας ως προς αναπαράσταση.



Σχήμα 2.2: 3D πλέγμα τριγώνων που αναπαριστά δελφίνι.

Πλειάδες αυτών των κυρτών πολυγώνων (*Triangular Meshes*), αναπαριστούν 3D μοντέλα και σκηνές του χώρου και κατά συνέπεια κάθε κυρτό πολύγωνο έχει μια τιμή χρώματος που του έχει αποδοθεί από τον αλγόριθμο σκίασης. Για να γίνει αυτό στον 3D χώρο, αυτός διαμερίζεται σε ογκομετρικό πλέγμα (*Voxel Grid*).

Παρόμοιες αναπαραστάσεις στον 3D χώρο είναι και τα νέφη σημείων *Point Clouds*, τα οποία ωστόσο δεν είναι ιδανικές αναπαραστάσεις στερεών αντικειμένων ή ρευστών (είναι μη συνεχείς και μη πραγματικές συναρτήσεις). Συνήθως χρησιμοποιούνται σε ογκομετρικά δεδομένα και πεδία πυκνότητας (γραφικά σύννεφων) όπου δεν απαιτούνται αναπαραστάσεις συμπαγών επιφανειών.

Για τον υπολογισμό του χρώματος σε τέτοιες δομές, απαιτείται τόσο από την υφή αλλά και ο φωτισμός του αντικειμένου. Κατά συνέπεια όλες οι ρητές αναπαραστάσεις γραφικών πριν την προβολή τους, είναι πίνακες που αναπαριστούν τρισδιάστατα ογκομετρικά πλέγματα. Κατά την διαδικασία της απόδοσης που θα αναλυθεί στην συνέχεια είτε με την μέθοδο της ραστεροποίησης(*rasterization*) με χρήση κλασικών μοντέλων φωτισμού ή χρησιμοποιώντας και μεθόδους ολικού φωτισμού με την μέθοδο ρίψης ακτίνων (*ray tracing, ray casting*) αποτυπώνονται στην οθόνη με δισδιάστατη μορφή. Στα σύγχρονα συστήματα απόδοσης γραφικών (πχ. *Blender, Meshlab* κτλ.) ο χρήστης έχει τη δυνατότητα να δει προοπτικές προβολές του 3D μοντέλου του σε ένα συνεχές χώρο διάφορων οπτικών γωνιών μέσα από το *viewport*(επεξεργάσιμος καμβάς γραφικών που χρησιμοποιείται από τον σχεδιαστή), μετακινώντας την θέση της εικονικής πλέον κάμερας.

2.1.2 Παραμετρικές Αναπαραστάσεις Γραφικών

Πέραν των άμεσων αναπαραστάσεων γραφικών που είναι ο κατ' εξοχήν τρόπος με τον οποίο αναπαριστώνται τα στατικά γραφικά στην συντριπτική τους πλειοψηφία, σε διάφορες εφαρμογές, κυρίως στον τομέα της σχεδίασης CAD, χρησιμοποιούνται μαθηματικές αναπαραστάσεις που υπάγονται στην κατηγορία των παραμετρικών αναπαραστάσεων γραφικών. Αυτές οι μορφές γραφικών αναπαραστάσεων μπορεί να γίνουν αρκετά σύνθετες και βασίζονται σε γνωστές συναρτήσεις.

Πιο συγκεκριμένα, η τεχνική που ακολουθείται είναι η τμηματική προσέγγιση καμπυλών ή επιφανειών, μέσω συναρτήσεων που ονομάζονται πολυώνυμα βάσης [10]. Ανάλογα τον βαθμό του πολυωνύμου περιγράφονται αντίστοιχης διάστασης γεωμετρικοί τόποι.

Σε αυτή την κατηγορία μεθόδων η θέση τυχαίου σημείου $P \in \mathbb{R}^n$ προσέγγιζεται από μία διανυσματική συνάρτηση $\mathbf{P}(\tau)$ παραμετροποιημένη ως προς την μεταβλητή τ . Η εν λόγω συνάρτηση, παριστάνεται ως ανάπτυγμα σε ένα πεπερασμένο σύνολο συναρτήσεων βάσης με συντελεστές του αναπτύγματος να αποτελούν μια πλειάδα σημείων του χώρου (τρισδιάστατος χώρος ή δισδιάστατο επίπεδο), που αποκαλούνται σημεία ελέγχου κάτι το οποίο εισάγει και το πρόβλημα των βαθμών ελευθερίας.

Η βασική μορφή της συνάρτησης $\mathbf{P}(\tau)$ για τα αναπτύγματα σε συναρτήσεις βάσης είναι η ακόλουθη:

$$\mathbf{P}(\tau) = \sum_{n=0}^N p_i * \phi_i(\tau), \quad (2.1)$$

δηλαδή το άθροισμα πεπερασμένου πλήθους συναρτήσεων βάσης ϕ σε $n + 1$ σημεία ελέγχου. Προφανώς η έκφραση 2.1 μπορεί να χρησιμοποιηθεί με ποικίλα πολυώνυμα, ωστόσο υπάρχουν ορισμένες οικογένειες

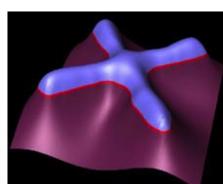
πολυωνύμων, οι οποίες έχουν επικρατήσει λόγω χρήσιμων ιδιοτήτων τους στην περιγραφή μεγάλου συνόλου γραφικών αναπαραστάσεων. Περισσότερα για τα πολυώνυμα βάσης στο Παράρτημα K.7.

Παρά την αναλυτικότητα αυτών των μορφών αναπαράστασης υπάρχουν σημαντικά μειονεκτήματα. Παράδειγμα, η διακριτή και αναγκαία παραμετροποίηση τους αλλά και η αδυναμία περιγραφής τους από συναρτήσεις που μπορούν να αναπαρασταθούν από νευρωνικά δίκτυα. Συνεπώς, δεν αποτελούν κατάλληλη επιλογή ως μέσο αποτύπωσης επιφανειών που βελτιώνει τις παραμέτρους του ώστε να καταφέρει να μάθει την γεωμετρία άγνωστης επιφάνειας (εφόσον δεν είναι εύκολα εκπαδεύσιμες μορφές), μιας και οι παράγωγοι τους σε ορισμένες περιπτώσεις είναι ήδη αρκετά μεγάλου βαθμού και υπάρχει ήδη πεπερασμένος βαθμός ελευθερίας λόγω των σημείων ελέγχου [10].

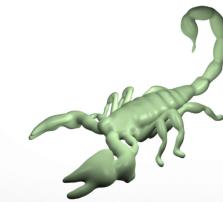
2.1.3 Έμμεσες αναπαραστάσεις Γραφικών

Οι άμεσες αναπαραστάσεις αυξάνουν το υπολογιστικό κόστος των αλγορίθμων καθώς είναι ακατέργαστες μορφές ενώ οι παραμετρικές είναι δύσκολες στην βελτιστοποίηση. Σε αυτό το πρόβλημα, δίνουν λύση οι έμμεσες αναπαραστάσεις δηλαδή η περιγραφή των σημείων του χώρου για παράδειγμα μέσα από την απόστασή του από την επιθυμητή επιφάνεια. Έτσι, για παράδειγμα, η επιφάνεια είναι το σύνολο των σημείων που δίνουν μηδέν για αυτήν την συνάρτηση απόστασης.

Αυτές οι αναπαραστάσεις, έχουν πλήθος πλεονεκτημάτων με βασικότερο όλων ότι μπορούν να περιγραφούν μέσω νευρωνικών δικτύων, μιας και ένα νευρωνικό δίκτυο ουσιαστικά περιγράφει σύνολα επιπέδων(πολλαπλότητες 2.3.1 , απόδειξη 7). Έτσι, με κατάλληλες παραμέτρους και έλεγχο της επιθυμητής εξόδου ένα δίκτυο μπορεί να προσεγγίσει έμμεσα μια τρισδιάστατη επιφάνεια προσεγγίζοντας την συνάρτησή προσημένης απόστασης. Έτσι προσεγγίζει μια πολύ πιο ευέλικτη δομή αυτή των οριακών πολλαπλοτήτων ή πιο συγκεκριμένα της συνάρτησης που αναπαριστά ισομετρική επιφάνεια για συγκεκριμένη τιμή.



Πολλαπλότητες δύο διαστάσεων ορίζουν καμπύλη μιας διάστασης



Πολλαπλότητες τριών διαστάσεων ορίζουν έμμεσα επιφάνεια σκορπιού

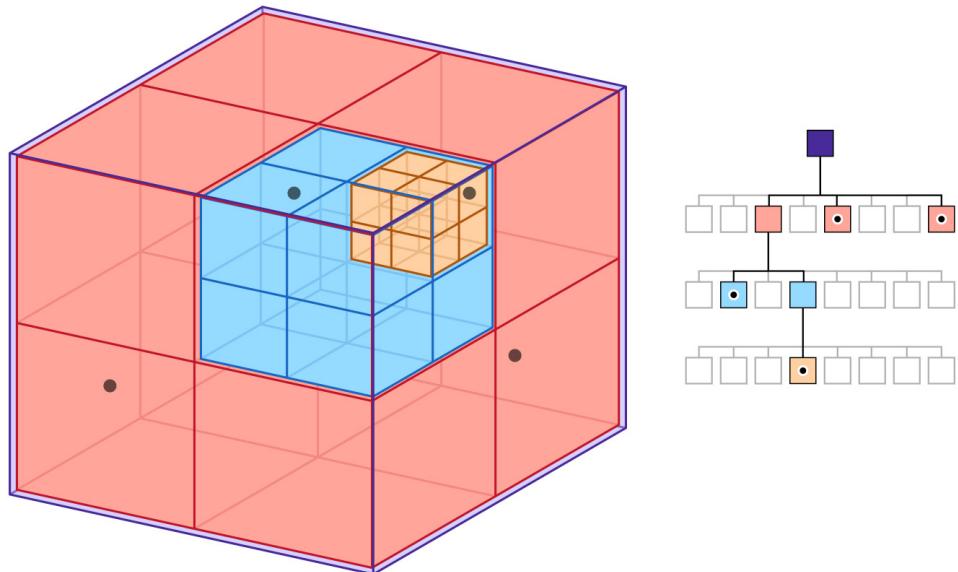
Σχήμα 2.3: Πολλαπλότητες ορίζουν έμμεσα ισομετρικές επιφάνειες χαμηλότερης διάστασης σε όριο απόφασης

Μια γενική μορφή έμμεσης συνάρτησης ορίζεται ως εξής:

- Εσωτερικά Σύνολα Επιπέδων: $F(x, y, z) < 0$
- Εξωτερικά Σύνολα Επιπέδων: $F(x, y, z) > 0$
- Οριακή Επιφάνεια: $F(x, y, z) = 0$

Σκοπός είναι η επιτυχής περιγραφή της οριακής επιφάνειας. Σε κάθε περίπτωση, η κλίση ∇F είναι ορθογώνια στα σύνολα επιπέδων (εφόσον ταυτίζεται με το κάθετο διάνυσμα στο εφαπτόμενο επίπεδο) και ειδικότερα στην περίπτωση που η έμμεση επιφάνεια περιγράφεται από συνάρτηση προσημασμένης απόστασης ($sign \in \{-1, 1\}$), το $|\nabla F| = 1$. Εφόσον αυτού του είδους οι έμμεσες συναρτήσεις ή πεδία χρησιμοποιούνται για να περιγράψουν την γεωμετρία των εκπαιδευόμενων επιφανειών, γίνεται περισσότερη ανάλυση στο κεφάλαιο 2.3.

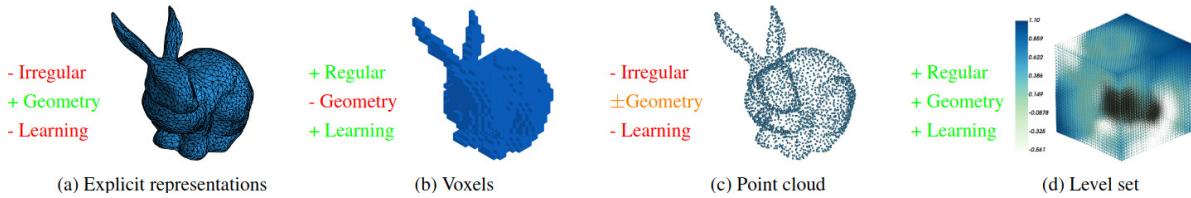
Ταυτόχρονα, υπάρχουν και άλλες μορφές έμμεσων αναπαραστάσεων γραφικών όπως είναι τα οκταδικά δέντρα [11] ή καλύτερα octrees τα οποία θεωρούνται δομές που αναπαριστούν την κάλυψη ενός ογκομετρικού στοιχείου (voxel) από σημεία. Έτσι, τα οκταδικά δέντρα χρησιμοποιούνται συχνότερα για τη διαίρεση ενός τρισδιάστατου χώρου διαιρώντας τον αναδρομικά σε n οκτάδες¹ και περιγράφουν σε ποιες περιοχές υπάρχουν σημεία. Αυτή η δομή, χρησιμοποιεί την a-priori ιδιότητα της συνοχής του χώρου για να ελαττώσει τις αποθηκευτικές απαιτήσεις των τρισδιάστατων αντικειμένων και επίσης παρέχει μια βολική περιγραφή για την αποθήκευση πληροφορίας σχετικά με το εσωτερικό των αντικειμένων. Ωστόσο, δεν αποτελεί κατ' εξοχήν συνεχή μορφή αναπαράστασης που σημαίνει ότι χωλαίνει στην αναπαράσταση συμπαγών επιφανειών υψηλής λεπτομέρειας. Παρακάτω φαίνεται ένα παράδειγμα οκταδικού δέντρου και πως δίνει πληροφορία για τρισδιάστατα δεδομένα.



Σχήμα 2.4: Οκταδικό Δέντρο

¹(όπου n αφορά την ανάλυση)

Συνοψίζοντας, γίνεται επιλογή έμμεσων επιφανειών για αναπαράσταση 3D δομών λόγω των πλεονεκτημάτων που αναφέρθηκαν και φαίνονται συνοπτικά στην παρακάτω εικόνα :



Σχήμα 2.5: Πλεονεκτήματα και μειονεκτήματα Γραφικών Αναπαραστάσεων (Πηγή [12])

Ο τρόπος εξαγωγής πλεγμάτων τριγώνων (τυπική μορφή αποθήκευσης τρισδιάστατων δεδομένων) από έμμεση αναπαράσταση επιφάνειας είναι αρκετά ενδιαφέρων με έναν από τους κλασσικούς πλέον αλγορίθμους την γραφική υπολογιστών να αποτελεί τον αλγόριθμο *contouring* (περιγράμματος) Marching Cubes, ο οποίος αφήνεται στο παράρτημα Κεφ.7 για τον αναγνώστη.

2.2 Μέθοδοι Αποτύπωσης 3D γραφικών σε εικόνα

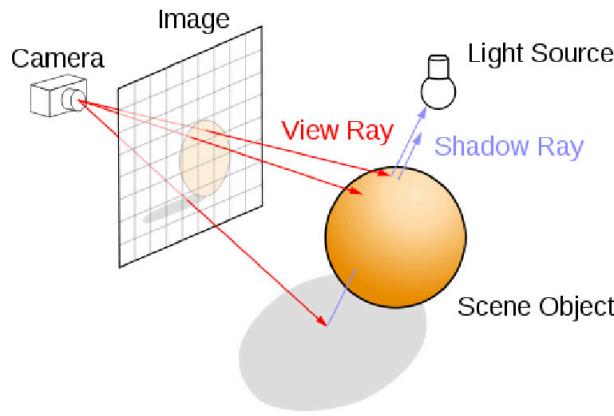
Όταν γίνεται λόγος για μεθόδους αποτύπωσης γραφικών αναφερόμαστε στο σύστημα που συνολικά συνθέτει την τρισδιάστατη σκηνή έχοντας ως δεδομένα τις μαθηματικές αναπαραστάσεις των γεωμετρικών οντοτήτων (*objects*) στο χώρο. Μετέπειτα η μηχανή γραφικών αναλαμβάνει την προβολή της σκηνής στο δισδιάστατο επίπεδο ανάλογα με το μοντέλο προβολής που έχει επιλεχθεί.

Βέβαια, αυτή η ερμηνεία του συστήματος απόδοσης δεν είναι δόκιμη. Η αποτύπωση γραφικών σε εικόνα αποτελεί, ουσιαστικά, τον τρόπο υπολογισμού του χρώματος των επιφανειών των στοιχειωδών αναπαραστάσεων, που δομούν τις επιφάνειες, με βάση το μοντέλο εικονικού φωτισμού και το μοντέλο σκίασης που ακολουθείται. Κυρίαρχο ρόλο σε αυτό παίζει η τεχνική που ακολουθείται στο κομμάτι υπολογισμού το χρώματος των εικονοστοιχείων με βάση το μοντέλο φωτισμού που ακολουθείται (είδος υλικού, είδη φωτισμού). Πάνω σε αυτό το κομμάτι υπάρχουν και φωτορεαλιστικές προσεγγίσεις του προβλήματος (πχ. ολικός φωτισμός με χρήση ιχνηλάτισης ακτίνων).

Αναφορικά στις κατηγορίες των μεθόδων σκίασης ανήκουν οι αλγόριθμοι *Lambert* ή *Flat Shading* και *Gouraud* ή *Smooth Shading*. Στο κομμάτι της προσομοίωσης του φωτισμού χωρίς φωτορεαλιστικά αποτελέσματα ανήκουν μέθοδοι όπως ο αυτοφωτισμός, ο φωτισμός διάχυσης, ο κατοπτρικός φωτισμός και το μοντέλο *Phong* κατοπτρικής ανάκλασης, κλασσικά τεχνάσματα φωτισμού που αποδίδουν και διαφορετικές υφές υλικών (*Βασικά Μοντέλα Φωτισμού Κεφ.17 στο [11]*). Όλα αυτά τα μοντέλα φωτισμού έχουν πρόβλημα στον υπολογισμό των σκιών οι οποίες δεν υπολογίζονται αυτόματα από το σύστημα αλλά επίσης θα πρέπει τα γραφικά να είναι άμεσα ορισμένα που σημαίνει στοιχεία που προκύπτουν από αλληλεπιδράσεις του φωτός με τα αντικείμενα θα πρέπει να είναι τεχνητά φτιαγμένα. Συνεπώς, τέτοιες πρακτικές δεν αρκούν στην ανακατασκευή γραφικών που όταν αποτυπώνονται σε εικόνες προσεγγίζουν πραγματικές φωτογραφίες.

Εκτός από το κομμάτι του χρώματος, για να γίνει και γεωμετρική βελτίωση του μοντέλου χρειάζεται αντίστροφη αποτύπωση, δηλαδή μετάβαση από την εικόνα στον τρισδιάστατο χώρο και εύρεση των ακτίνων φωτός που πέφτουν πάνω στην επιφάνεια. Τέλος, οι οντότητες μια σκηνής προφανώς θα πρέπει να μπορούν να περιγραφούν και με έμμεσο τρόπο και όχι μόνο με άμεσο τρόπο (πλέγματα πολυγώνων [2]).

2.2.1 Αλγόριθμος Ιχνηλάτισης Ακτίνων | Ray Tracing

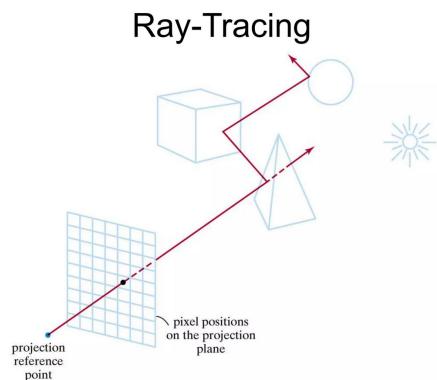


Σχήμα 2.6: Σύστημα Κάμερας σε μηχανές Γραφικών που εκτελούν Ιχνηλάτιση Ακτίνας

Στην γραφική υπολογιστών το κομμάτι του φωτισμού, είναι μια απαιτητική και αναγκαία διαδικασία για την απόδοση χαρακτηριστικών όπως το χρώμα και η υφή του υλικού το οποίο απεικονίζεται στην σκηνή. Η πορεία του φωτός παίζει εξίσου σημαντικό ρόλο εδώ, καθώς πέραν του ότι καθορίζει την θέση των σκιών δίνει πληροφορίες για ανακλάσεις που δημιουργούνται από αλλεπάλληλες ανακλάσεις του φωτός σε αντικείμενα.

Η έννοια της ρίψης ακτίνων (*ray casting*) χρησιμοποιείται στην δημιουργική στερεομετρία για τον εντοπισμό των τομών πάνω σε επιφάνειες, με ακτίνες που προκύπτουν από τις θέσεις των εικονοστοιχείων(UV επίπεδο βλ.Παράρτημα Κεφ.[7]).

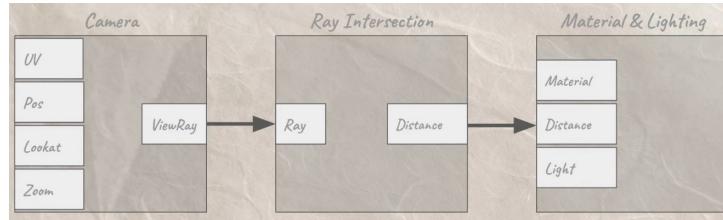
Ο αλγόριθμος ιχνηλάτισης ακτίνων είναι μια γενικευμένη μέθοδος ρίψης ακτίνων ως ένα μέσο εντοπισμού των ορατών επιφανειών μιας σκηνής και προτάθηκε από την επιστημονική κοινότητα όταν οι απλοί μέθοδοι φωτισμού δεν παρήγαγαν το επιθυμητό οπτικό αποτέλεσμα υπό το πρίσμα της φωτογραφικά ρεαλιστικής παραγόμενης σκηνής.



Σχήμα 2.7: Πολλαπλά Ίχνη ανάκλασης και μετάδοσης ακτίνας που προέρχεται από το σημείο αναφοράς προβολής, περνάει μέσα από την θέση του Πηγή:[11]

Το πρόβλημα είναι πως οι πηγές δεν είναι σημειακές αλλά και επίσης τα σώματα δεν επηρεάζονται αποκλειστικά και μόνο από το φως που προκύπτει από την πηγή φωτός αλλά και από τις αλληλεπιδράσεις και τις ανακλάσεις που κάνει το φως με άλλα σώματα στην σκηνή. Έτσι, αντί απλά να ψάχνεται μια θέση ενός εικονοστοιχείου για την ορατή επιφάνεια, συνεχίζουμε την πορεία της ακτίνας που περνά από ένα εικονοστοιχείο για να γίνει συλλογή εντάσεων ακτινοβολίας από διάφορες ανακλάσεις στις επιφάνειες που περιλαμβάνει η τρισδιάστατη σκηνή.

Συνεπώς, αν θα μπορούσαμε να οριοθετήσουμε τον αλγόριθμο *Ray Tracing* σε μια διαδικασία που απεικονίζεται με διάγραμμα ροής και βασίζεται στις παραπάνω εικόνες θα ήταν η εξής:



Σχήμα 2.8: Το διάγραμμα ροής του *Ray Tracing* Πηγή: Ray Marching Video

Το πιο απαιτητικό κομμάτι στο διάγραμμα που καθορίζει και την γεωμετρική λεπτομέρεια που αποδίδεται στην εικόνα (και είναι αλληλένδετο με την γεωμετρική λεπτομέρεια που υπάρχει στην 3D πληροφορία), είναι το *Ray Intersection*(υπολογισμός των σημείων τομής της ακτίνας με την επιφάνεια). Σε αυτό το τμήμα, έχουν βοηθήσει αρκετά νέοι αλγόριθμοι ιχνηλάτισης ακτίνας, ιδιαιτέρως σε έμμεσες αναπαραστάσεις επιφανειών με χρήση πεδίων απόστασης.

2.2.2 Απόδοση Όγκου

Αλγόριθμος Βηματισμού Πάνω στην Ακτίνα | Ray Marching

Ο αλγόριθμος ιχνηλάτισης ακτίνας, υπολογίζει επακριβώς τα σημεία τομής της επιφάνειας με την ακτίνα που πάνω στην κατεύθυνση προβολής. Αυτό μπορεί να αποβεί μοιραίο υπολογιστικά, ιδιαιτέρως όταν η γεωμετρία που αναπαριστάται είναι αρκετά σύνθετη και δεν περιγράφεται άμεσα. Για αυτόν τον λόγο οι κλασσικές μέθοδοι ιχνηλάτισης ακτίνας δεν χρησιμοποιούνται στην απόδοση ογκομετρικών γραφικών.

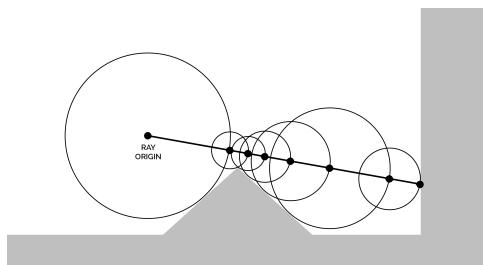
Μια πιο συνετή επιλογή είναι ο Αλγόριθμος Βαδίσματος πάνω στην ακτίνα, *Ray Marching*, που σκοπό δεν έχει να βρει το ακριβές σημείο πάνω στην επιφάνεια αλλά να κάνει βήματα από την αρχή της ακτίνας προς την επιφάνεια και να φτάσει όσο τον δυνατόν πιο κοντά στην επιφάνεια. Αποτέλεσμα αυτού, πρώτον ότι δεν είναι απαραίτητοι όλοι οι υπολογισμοί που κάνει το *Ray Tracing* και δεύτερον η αξιοπιστία των αποτελεσμάτων παραμένει υψηλή. Τέλος μετασχηματισμοί και παραμορφώσεις σε γραφικές αναπαραστάσεις δεν επιβαρύνουν την λειτουργία του αλγορίθμου (ο οποίος ψάχνει σημεία τομής), ενώ αντιθέτως στο *Ray Tracing* αυτό αποτελεί τροχοπέδη, για αυτό συνήθως ο τελευταίος χρησιμοποιείται για στατικά γραφικά απλής γεωμετρίας που συνήθως ορίζονται άμεσα.

Επί του πρακτέου, χρησιμοποιείται ο αλγόριθμος βαδίσματος πάνω στην ακτίνα ο οποίος αποτελεί εκ φύσεως, έναν άπληστο αλγόριθμο αφού ψάχνει τον χώρο για τομή με επιφάνειες. Εποι, για την εφαρμογή του γίνεται μια σειρά πεπερασμένων βημάτων κατά την διεύθυνση της ακτίνας μέχρι να έρθει σε επαφή η ακτίνα πάνω σε κάποιο αντικείμενο της σκηνής ή να ξεπεραστεί ο αριθμός των πεπερασμένων βημάτων που επιλέγονται ως συνθήκη τερματισμού του αλγορίθμου². Συνεπώς μια συνοπτική αλγορίθμιμη παρουσίαση του αλγορίθμου περιλαμβάνει τα εξής βήματα:

²για να μη επεκτείνει την ακτίνα προς το άπειρο

Algorithm 1 Απλοική μορφή Ray Marching

- 1: Ρίψη ακτίνας προς την 3D σκηνή μέσα από το UV επίπεδο της φωτογραφίας
- 2: Αρχικοποίηση μεταβλητών
- 3: Ορισμός μεγίστου αριθμού βημάτων MAX_STEPS
- 4: **for** κάθε pixel στο UV επίπεδο της φωτογραφίας **do**
- 5: $p \leftarrow$ αρχική θέση της ακτίνας
- 6: $d \leftarrow$ αρχική απόσταση
- 7: $i \leftarrow 0$
- 8: **while** $i < \text{MAX_STEPS}$ και $d > \text{MIN_DISTANCE}$ **do**
- 9: Υπολογισμός πεδίου απόστασης $d(p)$
- 10: Προώθηση της ακτίνας προς την κατεύθυνση του κοντινότερου object
- 11: $i \leftarrow i + 1$
- 12: **end while**
- 13: Εφαρμογή επεξεργασίας για την απόδοση του χρώματος στο pixel
- 14: **end for**

Χρήση Πεδίων Προσημασμένης Απόστασης σε συνδυασμό με τον αλγόριθμο Ray Marching - Sphere Tracing

Σχήμα 2.9: Χρήση Βαδίσματος Ακτίνας για την εύρεση Σημείων τομής με δομές του χώρου με χρήση σφαιρικού έλεγχου προσημασμένης απόστασης(Sphere Tracing) - 2D

Ο John C. Hart στην αντίστοιχη εργασία του [13] έδωσε την ιδέα υλοποίησης του αλγορίθμου, για απόδοση τρισδιάστατων σκηνών από έμμεσες επιφάνειες πολλών μορφών μεταξύ και πεδίων προσημασμένης απόστασης, όπου η επιφάνεια ορίζεται ως επιφάνεια απόστασης $d=0$ στην συνάρτηση προσημασμένης απόστασης και εντός της η απόσταση είναι αρνητική.

Algorithm 2 Sphere Tracing Algorithm

Initialize the ray with an initial position and direction.
Set the step size t and maximum distance T .
while Termination condition is not met **do**
 Calculate the current position along the ray: $p = \text{initial position} + t \cdot \text{direction}$.
 Evaluate the implicit function at the current position: $f(p)$.
 if $f(p)$ is less than a predefined threshold (indicating proximity to the surface), terminate the loop. **then**
 end if
 Update t using a specific step size function.
end while
Calculate the surface normal at the intersection point.
Calculate the final color and shading at the intersection point for rendering.

Αυτός ο αλγόριθμος είναι μια απλοϊκή μορφή του *Sphere Tracing* που χρησιμοποιείται και στην παρούσα εργασία.

2.2.3 Διαφορίσιμη αποτύπωση έμμεσων πεδίων *Implicit Differentiable Rendering*

Στην παρούσα εργασία χρησιμοποιείται ο αλγορίθμος βαδίσματος πάνω στην ακτίνα *Ray Marching/Sphere Tracing* μέσω του οποίου γίνεται δυνατή η εκπαίδευση της γεωμετρίας του χώρου που αναπαρίσταται. Η διαφορά έγκειται, στο ότι έχουμε ένα εκπαίδευσιμο πεδίο προσημασμένης απόστασης το οποίο μεταβάλλεται και χρησιμοποιείται ταυτόχρονα η διαδικασία του *Sphere Tracing* κατά την εκπαίδευση. Περισσότερα στο κεφάλαιο της μεθοδολογίας με την διαφορική μορφή εξίσωσης απόδοσης 4.3.

2.3 Εκπαιδευόμενο Πεδίο Προσημασμένης Απόστασης - Neural SDF

Στο σημείο αυτό ορίζονται κάποιες βασικές έννοιες γύρω από τοπολογικούς χώρους. Ορίζονται, επίσης, οι βασικές δομές των συναρτήσεων οι οποίες μετρούν τα όρια των γεωμετρικών αποστάσεων με τις έμμεσες επιφάνειες.

Ορισμός 2.3.1 (Συνάρτηση Προσημασμένης Απόστασης). Αν Ω είναι ένα υποσύνολο ενός μετρικού χώρου X με μετρική d , τότε η συνάρτηση προσημασμένης απόστασης f ορίζεται ως εξής

$$f(x) = \begin{cases} d(x, \partial\Omega) & \text{if } x \in \Omega \\ -d(x, \partial\Omega) & \text{if } x \in \Omega^c \end{cases}$$

όπου $\partial\Omega$ δηλώνει το όριο του Ω . Για κάθε $x \in X$,

$$d(x, \partial\Omega) := \inf_{y \in \partial\Omega} d(x, y)$$

όπου \inf δηλώνει το μέγιστο κάτω φράγμα ((infimum)). Βασική συνθήκη στην οποία υπακούν συγκεκριμένα τα πεδία απόστασης με πρόσημο είναι η εικονική εξίσωση (eikonal equation). Στην περίπτωση των πεδίων με των οποίων ασχολούμαστε αντιστοιχεί στο ότι ορίζεται για τα σημεία του χώρου που ισχύει

$$|\nabla F| - 1 = 0$$

, κάπι το οποίο αποτελεί πόρισμα της εξίσωσης στον ορισμό 2 που δίνεται παρακάτω 2.2

Ορισμός 2.3.2 (Ομοιομορφισμοί). Οι Ομοιομορφισμοί είναι απεικονίσεις που διατηρούν τις αποστάσεις. Αν το \mathbf{I} είναι ισομετρία (πολλαπλότητα), η απόσταση που επιστρέφεται από τον f δεν χρειάζεται προσαρμογή και είναι πλήρως αντιστρέψιμος.

$$d(\mathbf{x}, \mathbf{I} \circ f^{-1}(0)) = d(\mathbf{I}^{-1}(\mathbf{x}), f^{-1}(0)).$$

Οι ομοιομορφισμοί περιλαμβάνουν περιστροφές, μετατοπίσεις και ανακλάσεις.

Πόρισμα 2.3.1. Μπορούμε να θεωρήσουμε ένα ομοιομορφισμό, ως συνεχή απεικόνιση σε έκταση και κάμψη μιας τοπολογίας με τρόπο ώστε να μετα(παρα)μορφώνεται σ' ένα νέο σχήμα μέσω αυτού. Παράδειγμα μια κούπα καφέ και ένα ντόνατ είναι ομοιομορφικά το ένα σε σχέση με το άλλο, επειδή η κούπα μπορεί, με συνεχή παραμόρφωση, να μεταμορφωθεί σε ντόνατ (τόρο) και το αντίστροφο.



Σχήμα 2.10: Ομοιομορφισμός

2.3.1 Επιφάνειες που ορίζονται από πεδία Απόστασης

Υπό μαθηματικούς όρους η συνάρτηση προσημασμένης απόστασης είναι ουσιαστικά μια συνάρτηση που περιγράφει την ορθογώνια απόσταση ενός δεδομένου στοιχείου χ από το όριο ενός συνόλου Ω σε έναν μετρικό χώρο. Αυτή η συνάρτηση, ή πεδίο, εφόσον η είσοδος και η έξοδος είναι ίσης διάστασης, μπορεί να περιγράψει κάλλιστα μια επιφάνεια Ω η οποία αναπαριστά ένα μοντέλο του φυσικού χώρου. Συνεπώς όταν αυτός ο χώρος είναι άγνωστος, εφόσον μπορούμε να ελέγξουμε τις παραμέτρους της συνάρτησης μπορούμε να ορίσουμε ποιον

χώρο ακριβώς περιγράφει και η διαδικασία αυτή μπορεί να γίνει προσεγγιστικά σε νευρωνικά επίπεδα που καταγράφουν την γεωμετρία και την εμφάνιση της επιφάνειας αυτής στις παραμέτρους τους. Αυτά οδηγούνται σε μηδενική τιμή όταν αυτή η προσημασμένη απόσταση μηδενίζεται και τότε αναπαριστούν την επιθυμητή επιφάνεια. Πιο συγκεκριμένα, ο έλεγχος ενός συνόλου νευρωνικών επιπέδων που αποτυπώνει μια γεωμετρία είναι το πεδίο που εν τέλει οδηγεί στην ανακατασκευή άγνωστων επιφανειών.

Συνεπώς πιο δόκιμοι ορισμοί σύμφωνα με [13]:

Ορισμός 2.3.3 (Manifold). Έστω συνάρτηση \mathcal{F} , με $\mathcal{F} : \mathbb{R}^n \rightarrow \mathbb{R}$ αυτή περιγράφει πολλαπλότητες, ένα σύνολο δεδομένων που αντιστοιχεί σε ισοδυναμικές επιφάνειες και μοιάζει με ευκλείδειο χώρο $\mathcal{A} \subset \mathbb{R}^n$ ως τον γεωμετρικό τόπο των σημείων όπου:

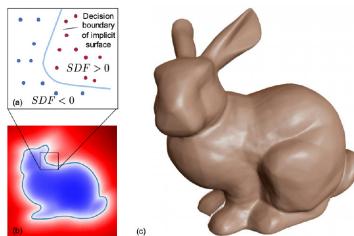
$$\mathcal{A} = \{x : \mathcal{F}(x) \leq 0\}$$

Ορισμός 2.3.4 (Manifold as a Signed Distance Field). Μια συνάρτηση $f : \mathbb{R}^3 \rightarrow \mathbb{R}$ είναι ένα όριο απόστασης της έμμεσης επιφάνειάς της $f^{-1}(0)$ αν και μόνο αν

$$|f(\mathbf{x})| \leq d(\mathbf{x}, f^{-1}(0)), \quad (2.2)$$

Εάν ισχύει η ισότητα για την 2.2, τότε η f είναι μια συνάρτηση απόστασης με πρόσημο. [13]

Ορισμένα πρωταρχικά στοιχεία γραφικών, όπως η σφαίρα, ορίζονται εύκολα με προσημασμένες συναρτήσεις απόστασης. Η εύρεση της απόστασης σε άλλα σχήματα μπορεί να είναι αρκετά δύσκολη. Εκεί υπεισέρχονται τα τεχνητά νευρωνικά δίκτυα προσεγγίζοντας την συνάρτηση προσημασμένης απόστασης με παλινδρομική διαδικασία. Το γεγονός ότι μπορούμε να βαδίσουμε πάνω σε μια ακτίνα με αποτέλεσμα να φτάσουμε είτε πάνω στην επιφάνεια, είτε εντός, είτε εκτός, σε συνδυασμό με το πεδίο προσημασμένης απόστασης να δίνει μηδέν στην έξοδο είτε εκτός με θετική απόσταση είτε εντός με αρνητική απόσταση ανάγει το πρόβλημα σε ένα πρόβλημα παλινδρόμησης για την εκπαίδευση παραμέτρων που αναπαριστούν άλλες γεωμετρίες.



Σχήμα 2.11: Ορισμός SDF πεδίου. Πηγή [6]

2.3.2 Διαφορίσιμες πολλαπλότητες και Γενικευμένο Θεώρημα Αναπαράστασης

Αναπαράσταση 3D επιφάνειας μέσω νευρωνικού δικτύου

Θεώρημα 2.3.1 (θεώρημα Bocher ή θεώρημα Representer). Κάθε συμπαγής, μη υποχρεωτικά φραγμένη, τμηματικά γραμμική υπερεπιφάνεια $\mathcal{M} \subset \mathbb{R}^d$ μπορεί να αναπαρασταθεί ως ένα σύνολο νευρωνικών επιπέδων (neural level set) \mathcal{S} ενός πολυστρωματικού δικτύου perceptron με ReLU (Rectified Linear Unit) συναρτήσεις ενεργοποίησης σε κάθε στρώμα, $F : \mathbb{R}^d \rightarrow \mathbb{R}$ [14]³.

³ Απόδειξη στο Παράρτημα B K.7

Σε διάφορες εργασίες δίνεται η δυνατότητα ελέγχου των δικτύων που αναπαριστούν SDF συναρτήσεις, όπως για παράδειγμα στην εκμάθηση γεωμετρίας σχηματικών αναπαραστάσεων είτε χωρίς είτε με την χρήση της πληροφορίας παραγώγων (*Sign Agnostic Learning Shape*) [15, 16] που αποδίδουν έμμεσα τις ισομετρικές επιφάνειες. Έτσι, ενδιαφέρον είναι να δούμε πώς γίνεται η εκπαίδευση πεδίων SDF που αναπαριστούν διάφορες γεωμετρίες.

2.3.3 Θεωρία πίσω από τον νευρωνικό έλεγχο διαφορίσιμων πολλαπλοτήτων(*neural level sets*)[17], [14]

Η χρήση δικτύων που αναπαριστούν ισομετρικές επιφάνειες και ο έλεγχος τους ώστε να αναπαριστούν τις επιθυμητές ισομετρικές επιφάνειες βασίζεται στις διαφορίσιμες πολλαπλότητες και την αμφιδιαφορισμότητα.

Ορισμός 2.3.5 (Αμφιδιαφόριση - Διαφορομοιομορφισμός). Αν \mathcal{X}, \mathcal{Y} ανοιχτά σύνολα στο \mathbb{R}^n , για κάποιο $n \in \mathbb{N}$. Τότε, λέμε ότι η απεικόνιση $f : \mathcal{X} \rightarrow \mathcal{Y}$ είναι αμφιδιαφορίσιμη (διαφορο-ομοιομορφική *diffeomorphism*), εάν ισχύουν οι ακόλουθες συνθήκες:

- η f είναι ομοιομορφισμός
- αμφότερες f, f^{-1} είναι συνεχώς διαφορίσιμες απεικονίσεις.

Σε αυτή την περίπτωση, τα \mathcal{X}, \mathcal{Y} λέγονται διαφορομορφικά σύνολα ή διαφορίσιμοι ισομετρικοί χώροι το ένα προς το άλλο. Εάν οι f, f^{-1} είναι αμφότερες k -φορές συνεχώς διαφορίσιμες, τότε η f απεικόνιση αποκαλείται αμφιδιαφόριση C^k .

Ορισμός 2.3.6 (Διαφορίσιμη Ισομετρική Επιφάνεια [14]). Μια C^k διαφορίσιμη πολλαπλότητα \mathcal{M} διάστασης n με σύνολο χαρτών επικάλυψης ή άτλαντα ομοιομορφισμών $(X_i, Y_i, f_i), i = 1, \dots, L$ είναι ένα τοπολογικό σύνολο (γενικότερη μορφή από ισομετρική επιφάνεια), όπου κάθε Y_i είναι ανοικτό σύνολο στο \mathbb{R}^n τέτοιο ώστε οι αντιστοιχίσεις επικάλυψης f_j ; να είναι όλες αμφιδιαφορίσιες C^k

Πόρισμα 2.3.2 (Open Mapping Theorem). Πόρισμα αυτού του ορισμού είναι ότι για κάθε n -διάστατο σημείο που ανήκει σε διαφορίσιμη πολλαπλότητα, $x \in \mathcal{M}$, υπάρχει ένας επιτρεπτός χάρτης απεικονίσεων $\mathcal{X}, \mathcal{Y}, f$, όπου για $x \in \mathcal{X}$ η f το αντιστοιχίζει στο \mathcal{Y} που αναπαριστά άλλη τοπολογία.

Τα σύνολα επιπέδων νευρωνικών δικτύων (*Neural Level Sets*), αναπαριστούν θεμελιώδεις ιδιότητες όπως ορια αποφάσεων κατηγοριοποιητών (*classifier decision boundaries*), αλλά χρησιμοποιούνται και στην μοντελοποίηση μη γραμμικών πολλαπλοτήτων όπως καμπύλες και ισομετρικές επιφάνειες. Συνεπώς, ο έλεγχος αυτών των νευρωνικών επιπέδων μπορεί να οδηγήσει σε διάφορες εφαρμογές τεχνητής νοημοσύνης όπως στην προκείμενη περίπτωση την ανακατασκευή 3D επιφανειών μέσω ελεγχόμενων ισομετρικών επιφανειών από τις παραμέτρους νευρωνικού δικτύου SDF.

Δεδομένου ενός νευρωνικού δικτύου $F(x; \theta) : \mathbb{R}^d \times \mathbb{R}^m \rightarrow \mathbb{R}^l$ το $0 \in \mathbb{R}^l$ σύνολο επίπεδου του ή (ρύθμιση του στο μηδέν στο επίπεδο εξόδου του), ορίζεται από την επιφάνεια:

$$S(\theta) = \{x | F(x; \theta) = 0\}$$

Εστω πως συμβολίζεται με $D_x F(p; \theta) \in \mathbb{R}^{l \times d}$ ο πίνακας των μερικών παραγώγων της F ως προς x . Εστω επίσης πως οι παράμετροι γεωμετρίας θ είναι σταθερές, με $F(p; \theta) = 0$ και ότι ο πίνακας των μερικών παραγώγων D είναι ένας πίνακας πλήρους βαθμού (κάθε γραμμή και στήλη είναι γραμμικά ανεξάρτητες), πόρισμα του γενικού θεωρήματος ασφάρων συναρτήσεων υπονοεί πως η S_θ ισομετρική επιφάνεια είναι μια πολλαπλότητα $d - l$ διάστασης στην περιοχή του $p \in S_\theta$.

Στόχος είναι, να ενσωματωθεί σε αυτή την διαφορίσιμη πολλαπλότητα που αναπαρίσταται από νευρωνικό δίκτυο ένα διαφορικό σφάλμα. Αυτό επιτυγχάνεται εκτελώντας την παρακάτω διαδικασία σε κάθε εποχή εκπαίδευσης.

- i) Δειγματοληψία n σημείων πάνω στην ισομετρική επιφάνεια τέτοια ώστε: $p_i \in S\theta, i \in [n]$
- ii) Δημιουργία του δικτύου δειγματοληψίας $p_i(\theta), i \in [n]$ κάνοντας χρήση ενός γραμμικού νευρωνικού στρώματος στο δίκτυο της έμμεσης αναπαράστασής της γεωμετρίας της ισομετρικής επιφάνειας $F(x; \theta)$ και
- iii) ενσωμάτωση μιας συνάρτησης σφάλματος στο δίκτυο δειγματοληψίας το οποίο χρησιμοποιείται ως διαμεσολαβητής στην ισομετρική επιφάνεια που αναπαρίσταται έμμεσα από την F .

2.3.4 Απόδοση Γραφικών σε Έμμεσες Διαφορίσιμες Επιφάνειες

Παρακάτω δίνεται μια συνεκτική περιγραφή του αλγορίθμου αποτύπωσης εκπαίδευσης αναπαράστασης. Ένας αλγόριθμος αποτύπωσης [5], μιας δεδομένης αναπαραστώμενης $3D$ σκηνής Φ . Ο Αλγόριθμος που συμβολίζεται με Θ , ο οποίος αντιστοιχίζει μια αναπαράσταση σκηνής Φ καθώς και τις εξωτερικές \mathbf{K} και εσωτερικές \mathbf{E} παραμέτρους της κάμερας (βλ. για το σύστημα κάμερας στο Κεφ.7) σε μια εικόνα \mathcal{I} περιγράφεται ως εξής:

$$\Theta : \mathcal{X} \times \mathbb{R}^{3 \times 4} \times \mathbb{R}^{3 \times 3} \rightarrow \mathbb{R}^{H \times W \times 3}, \quad (\Phi, \mathbf{E}, \mathbf{K}) \mapsto \Theta(\Phi, \mathbf{E}, \mathbf{K}) = \mathcal{I},$$

όπου \mathcal{X} είναι ο χώρος όλων των συναρτήσεων Φ . Η βασική επιπλοκή στην απόδοση μιας σκηνής Φ , είναι ότι η γεωμετρία αναπαρίσταται έμμεσα. Η επιφάνεια ενός ξύλινου τραπεζιού, για παράδειγμα, ορίζεται από τον υποχώρο του \mathbb{R}^3 όπου το Φ υφίσταται μια αλλαγή από ένα διάνυσμα χαρακτηριστικών που αναπαριστά τον ελεύθερο χώρο σε ένα που αναπαριστά το ξύλο. Αυτό σκοπεύει και να βελτιώσει και η παρούσα εργασία.

Για να αποτυπώσουμε ένα μεμονωμένο εικονοστοιχείο στην εικόνα που παρατηρείται από μια εικονική κάμερα, πρέπει να λύσουμε επομένως δύο υπό προβλήματα: (i) την εύρεση των παγκόσμιων συντεταγμένων των τομών των αντίστοιχων ακτίνων της κάμερας με τη γεωμετρία της σκηνής και (ii) την αντιστοίχηση του διανύσματος χαρακτηριστικών \mathbf{v} σε αυτή τη χωρική συντεταγμένη σε ένα χρώμα και την προβολή στην εικόνα.

Αλγόριθμος ανίχνευσης των Τομών Ακτίνων

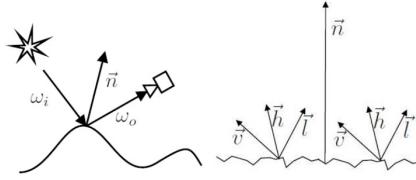
Για το κομμάτι εύρεσης των τομών των ακτίνων έγινε παραπάνω μια νύξη στον αλγόριθμο *Sphere Tracing* δεδομένης γεωμετρίας της σκηνής. Θα μπορούσε να προταθεί ένα πιο εξελιγμένο νευρωνικό δίκτυο που θα εφάρμοζε ένα νευρωνικό αλγόριθμο βάσισης ακτίνων με εκπαίδευσμα, προσαρμοστικό μέγεθος βήματος για την εύρεση των τομών των ακτίνων με τη γεωμετρία της σκηνής. Περισσότερα στο Παράρτημα A [Κεφάλαιο 7].

Δίκτυο Χρωματισμού Εικονοστοιχείων

Δίκτυο που κάνει χρήση BRDF [18]

Το δίκτυο που αφορά τον χρωματισμό εικονοστοιχείων, αποτελεί αντιστοίχηση του διανύσματος χαρακτηριστικών σε αυτή την χωρική συντεταγμένη σε ένα χρώμα και βασικό στέλεχος του δικτύου απόδοσης *Rendering Network*. Αυτό το δίκτυο στην κλασσική περίπτωση μπορεί να υλοποιεί μια συνάρτηση κατανομής της αμφίδρομης ανάκλασης *BRDF(Bidirectional Reflectance Distribution Function)*.

Η συνάρτηση BRDF $\mathcal{B}(\hat{x}, \hat{n}, \omega^i, \omega^o)$ περιγράφει το μέρος της ανακλώμενης ακτινοβολίας (δηλαδή της ηλεκτρομαγνητικής ροής του φωτός) σε κάποιο μήκος κύματος (δηλαδή το χρώμα) η οποία φεύγει από το σημείο x με κανονικό διάνυσμα \hat{n} κατά κατεύθυνση ω^o σε σχέση με την προστύπτουσα ακτινοβολία από την κατεύθυνση ω^i . Η BRDF επίσης βασίζεται επίσης στο κανονικό διάνυσμα \hat{n} στο σημείο της επιφάνειας (βλ.7). Οι πηγές φωτός περιγράφονται από την συνάρτηση $L^e(\hat{x}, \omega^o)$ και δίνουν το μέτρο της ακτινοβολίας που πηγάζει από την φωτεινή πηγή σε κάποιο μήκος κύματος στο σημείο x κατά την κατεύθυνση του ω^o . Η ολοκλήρωση της BRDF σε καθορισμένες στερεές γωνίες πρόσπτωσης και ανάκλασης ορίζει την ανακλαστικότητα, η οποία



Σχήμα 2.12: Διάγραμμα που δείχνει τα διανύσματα που ορίζουν την BRDF Πηγή [19]

μπορεί εύκολα να συσχετιστεί με την απορροφητικότητα (ή την ικανότητα εκπομπής) ενός υλικού. Συνεπώς η συνολική εξίσωση αποτύπωσης δίνεται από τον εξής τύπο:

$$L(\hat{\mathbf{x}}, \omega^o) = L^e(\hat{\mathbf{x}}, \omega^o) + \int_0 B(\hat{\mathbf{x}}, \hat{\mathbf{n}}, \omega^i, \omega^o) L^i(\hat{\mathbf{x}}, \omega^i) (\hat{\mathbf{n}} \cdot \omega^i) d\omega^i = M_0(\hat{\mathbf{x}}, \hat{\mathbf{n}}, \mathbf{v}) \quad (2.3)$$

Αυτό είναι και το μοντέλο που χρησιμοποιεί και το [IDR [6]] το οποίο υιοθετείται για την απόδοση σκηνής.

Χρήση Ογκομετρικού Μοντέλου Ακτινοβολίας NeRF [3]

Πολλά μοντέλα διαχειρίζονται το πρόβλημα της ακτινοβολίας χρησιμοποιώντας ως είσοδο 5D (πέντε διαστάσεων). Αυτά τα μοντέλα υπολογίζουν την εκ λαμβανόμενη ακτινοβολία για κάθε ακτίνα που περνά από τον καμβά της εικόνας σε όλο το μήκος της μέσω της κλασσικής εξίσωσης ογκομετρικής απόδοσης 2.4.

Ορισμός 2.3.7. *Neural Volume Rendering* Η νευρωνική απόδοση όγκου αναφέρεται σε μεθόδους που παράγουν εικόνες με την ιχνηλάτηση μιας ακτίνας στη σκηνή και την ολοκλήρωση κατά το μήκος της ακτίνας. Συνήθως ένα νευρωνικό δίκτυο όπως ένα πολυεπίπεδο perceptron κωδικοποιεί μια συνάρτηση από τις τρισδιάστατες συντεταγμένες της ακτίνας σε ποσότητες όπως η πυκνότητα και το χρώμα, οι οποίες ενσωματώνονται για να προκύψει μια εικόνα. Χρησιμοποιείται κατ' εξοχήν στα μοντέλα NeRF

Πρακτικά σύμφωνα με το NeRF [3], δεδομένης συνάρτησης πυκνότητας όγκου, το αναμενόμενο χρώμα που γυρνά στην κάμερα $C(\mathbf{r})$ κατά την διεύθυνση της ακτίνας $\mathbf{r}(t) = \mathbf{o} + t\mathbf{d}$ με κοντινά και μακριά όρια t_n, t_f δίνεται από την παρακάτω ολοκλήρωση:

$$C(\mathbf{r}) = \int_{t_n}^{t_f} T(t) \sigma(\mathbf{r}(t)) \mathbf{c}(\mathbf{r}(t), \mathbf{d}) dt, \quad (2.4)$$

οπού $T(t) = \exp\left(-\int_{t_n}^t \sigma(\mathbf{r}(s)) ds\right)$ αναφέρεται στην συνολική διαφάνεια ή διαπερατότητα φωτός κατά μήκος της ακτίνας⁴.

Το σημαντικό με τα ογκομετρικά μοντέλα είναι πως δεν χρειάζονται απαραίτητα επίβλεψη μάσκας για την θέση του αντικειμένου στην σκηνή. Μπορούν να ξεκινήσουν από ακατέργαστα σημεία του τρισδιάστατου χώρου, με τυχαία αρχικοποίηση παραμέτρων κάμερας και να εκτιμήσουν το πεδίο ακτινοβολίας στην πορεία της ακτίνας, χωρίς να αποκλίνουν.

Η παρούσα εργασία ασχολείται πάνω στο κομμάτι της διαφορίσιμης αποτύπωσης που δεν αφορά όμως τέτοιου είδους δίκτυα. Χρειάζεται ο έλεγχος στο επίπεδο της φωτογραφίας με χρήση δυαδικής μάσκας (*object masked rendering*) για καλύτερη σύγκλιση. Παράλληλα, οι εξωτερικές παράμετροι της κάμερας (7), πρέπει να μη έχουν αρχικοποιηθεί με τυχαίο τρόπο αλλά μπορούν εκπαιδευτούν μαζί με το δίκτυο. Μια αναλυτική περιγραφή της μεθόδου που ακολουθείται περιγράφεται στην μεθοδολογία (Κεφ. 4).

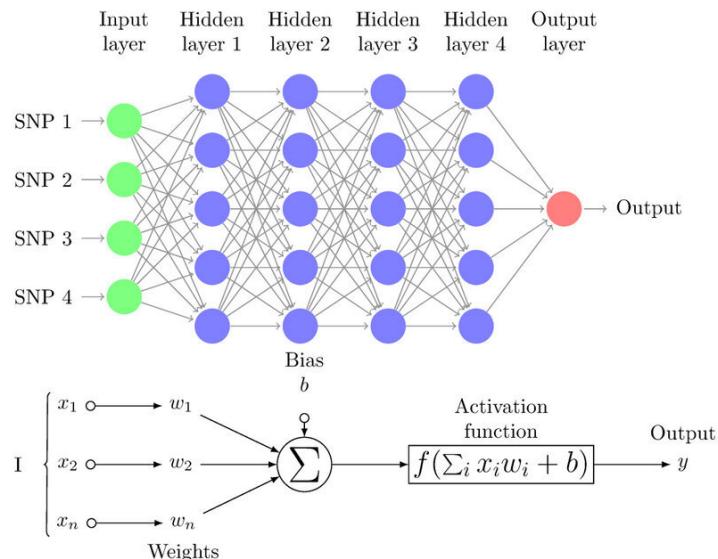
⁴αυτό είναι έμμεσο πεδίο διαφάνειας που υπολογίζεται σε μοντέλα ογκομετρικής απόδοσης

2.4 Βαθιά μάθηση

Εφόσον στην πλειοψηφία του, το λειτουργικό κομμάτι της παρούσας διπλωματικής βασίζεται στην μάθηση επιφανειών από έμμεσες αναπαραστάσεις μέσω βαθιών νευρωνικών δίκτυων (δίκτυα συντεταγμένων), γίνεται μια μικρή νύξη στην στα βαθιά νευρωνικά δίκτυα (DNN (*Deep Neural Networks*) και στον αλγόριθμο οπισθοδιάδοσης σφάλματος που είναι η κύρια μέθοδος βελτιστοποίησης των νευρωνικών δίκτυων.

2.4.1 Αρχιτεκτονική

Ένα τεχνητό νευρωνικό δίκτυο (*ANN*) αναπαριστάται στην παρακάτω εικόνα 2.13 και η βασική δομική του ονομάζεται νευρώνας *perceptron*. Πλειάδα αυτών των δομών σχηματίζει τον υπολογιστικό γράφο που ονομάζουμε πολυστρωματικό δίκτυο (*MLP*).



Σχήμα 2.13: Νευρωνικό δίκτυο Perceptron τεσσάρων κρυφών στρωμάτων

Ανάγκη μη γραμμικότητας

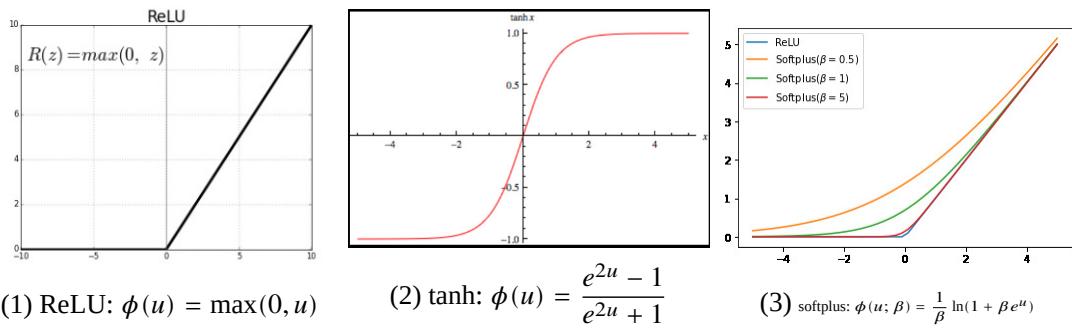
Η μη γραμμικότητα στο δίκτυο εισάγεται μέσω την συνάρτησης ενεργοποίησης και είναι απαραίτητη για να δουλέψει η διαδικασία της εκπαίδευσης με τον αλγόριθμο οπισθοδιάδοσης των βαρών. Φαίνεται πως, ένα δίκτυο με γραμμικές συναρτήσεις και n στρώματα με m κρυφές μονάδες είναι ισοδύναμο με ένα γραμμικό νευρωνικό δίκτυο χωρίς εσωτερικά στρώματα και αυτό φαίνεται κάνοντας τις παρακάτω πράξεις:

$$\begin{aligned} y = h(\mathbf{x}) &= \mathbf{b}_n + W_n (\mathbf{b}_{n-1} + W_{n-1} (\dots (\mathbf{b}_1 + W_1 \mathbf{x}) \dots)) \\ &\quad + W_n W_{n-1} \mathbf{b}_{n-2} + \dots + W_n W_{n-1} \dots W_1 \mathbf{x} = \mathbf{b}' \end{aligned}$$

Συνεπώς το δίκτυο δεν βελτιώνει την ικανότητα προσέγγισης του βάζοντας περισσότερα στρώματα, εφόσον οι παράγωγοι των εξόδων των νευρώνων δεν είναι συνεχείς διαφορίσιμες συναρτήσεις, ενώ ταυτόχρονα

δεν μπορεί να αντιληφθεί τις μη γραμμικές εξαρτήσεις μεταξύ των δεδομένων. Το καθολικό θεώρημα προσέγγισης ταυτόχρονα απαιτεί αυτήν την μη γραμμικότητα να ισχύει. Σύμφωνα με αυτό το θεώρημα κάτω από συγκεκριμένες συνθήκες κάθε συνάρτηση $f : [0, 1]^d \rightarrow \mathbb{R}$, είναι σε θέση να αναπαραστήσει οποιαδήποτε μορφής συνάρτησης και μάλιστα να εκτιμήσει συναρτήσεις ρυθμίζοντας κατάλληλα τα συναπτικά βάρη $\omega_1, \dots, \omega_n$ χωρίς την ανάγκη να υπάρχουν περισσότερα από ένα επίπεδα τεχνητών νευρώνων.

Συνήθεις συναρτήσεις ενεργοποίησης που εισάγουν μη γραμμικότητα στα δίκτυα και ταυτόχρονα δεν δημιουργούν προβλήματα όπως αυτά του εξ αφανισμού του βαρών ή της τεράστιας αύξησης τους (κάτι που σχετίζεται με την αρχικοποίηση και την μετέπειτα κανονικοποίηση των βαρών αλλά και της εισόδου [20]) είναι οι ακόλουθες:



Σχήμα 2.14: Τρεις συνήθεις συναρτήσεις ενεργοποίησης

Διαδικασία Εκπαίδευσης

Τα βαθιά νευρωνικά δίκτυα χωρίζονται σε κατηγορίες ανάλογα με τον τρόπο που εκπαιδεύονται επίσης. Συγκεκριμένα υπάρχουν δύο μεγάλες κατηγορίες μάθησης οι οποίες βασίζονται στον αν παρέχεται η όχι επίβλεψη στην παραγόμενη έξοδο από το νευρωνικό δίκτυο. Συνήθως όταν παρέχονται ετικέτες επίβλεψης στα δεδομένα δηλαδή γενικώς αποδεκτές τιμές που θα πρέπει να βγάλει ως έξοδο το δίκτυο με βάση την είσοδο του τότε η μάθηση είναι επιβλεπόμενη. Αντιθέτως, όταν τα δεδομένα δεν έχουν ετικέτες επίβλεψης τα δίκτυα προσπαθούν να μάθουν ιδιότητες δεδομένων, ή και τα ίδια τα δεδομένα και η μάθηση είναι μη επιβλεπόμενη. Στα πλαίσια της παρούσας εργασίας η εκπαίδευση είναι επιβλεπόμενη καθώς υπάρχουν εικόνες των δεδομένων εκπαίδευσης οι οποίες αντιταραβάλλονται με φωτογραφίες από το ανακατασκευασμένο 3D μοντέλο και υπολογίζεται η διαφορά τους εικονοστοιχείο προς εικονοστοιχείο. Προβλήματα μη επιβλεπόμενης μάθησης συνήθως δεν έχουν τόσο σαφείς στόχους όπως η παρούσα εργασία και συνήθως μαθαίνουν τα δεδομένα υπό τις έννοιες της ομαδοποίησης, της συσχέτισης ή και μείωσης διάστασης ώστε να μπορέσουν είτε να παράγουν δεδομένα όμοια με αυτά είτε να βοηθήσουν σε κάποια λειτουργία που βασίζεται στις προαναφερθείσες διαδικασίες.

Ο αλγόριθμος Οπισθοδιάδοσης Βαρών Back Propagation [14, 21]

Στην βαθιά μάθηση υπάρχει μια αντικειμενική συνάρτηση στόχου η οποία πολλές φορές ονομάζεται και συνάρτηση κόστους την οποία θέλουν να βελτιώσουν μειώνοντας το σφάλμα που παρατηρείται στην έξοδο του δικτύου. Αρχικά γίνεται η ανάκληση του δικτύου που παρουσιάζει το σφάλμα στο επίπεδο εξόδου. Το σφάλμα ακολουθεί μια διαδικασία οπισθοδιάδοσης στους στα προηγούμενα στρώματα του δικτύου, όπου υπολογίζεται το επιμέρους σφάλμα που παρουσιάζει η έξοδος του εκάστοτε νευρώνα. Έπειτα, υπολογίζεται υπολογίζεται η κλίση της συνάρτησης κόστους ως προς τα βάρη εκπαίδευσης. Ανάλογα της μορφής εκπαίδευσης αναβαθμίζονται κατάλληλα τα συναπτικά βάρη w_1, w_2, \dots, w_n με τα οποία πολλαπλασιάζεται ο νευρώνας. Έτσι επιγραμματικά ο αλγόριθμος οπισθοδιάδοσης εκτελεί τα εξής βήματα:

1. Αρχικοποίηση βαρών (περισσότερα στο 2.4.3).
2. Παρουσιάσεις δειγμάτων εκπαίδευσης (οργανωμένο σε πακέτα (*batches*) ανά εποχή εκπαίδευσης (εξαρτάται από τον αλγόριθμο βελτιστοποίησης).
3. Εμπρόσθια διάδοση

$$u^{(l)}(n) = \sum_i \omega_{ji}^{(l)}(n) y_i^{(l-1)}(n),$$

με $y_j^{(l)} = \phi_j(u_j(n))$ και φ συναρτήσεις ενεργοποίησης, και υπολογισμός του σφάλματος εξόδου

$$e_j(n) = d_j - o_j(n)$$

4. Υπολογισμός των δυναμικών ή κλίσεων δ με βάση τον κανόνα βελτιστοποίησης οποίος θα μπορούσε να βασίζεται σε κάποιο σύγχρονο αλγόριθμο κατάβασης δυναμικού όπως ο *SGD* ή πιο σύγχρονους προσαρμοστικούς αλγορίθμους κατάβασης δυναμικού όπως ο *Adam* (παρουσιάζεται ο κλασσικός Δέλτα από Πηγή: [14]) Επομένω έχουμε:

- Υπολογισμός τοπικών κλίσεων:

$$\delta_j^{(l)}(n) = \begin{cases} e_j^{(L)}(n) * \dot{\phi}_j(u_j^{(L)}(n)), & \text{για τον νευρώνα } j \text{ στο επίπεδο εξόδου } L \\ \dot{\phi}_j(u_j^l(n)) \sum_k \delta_k^{(l+1)}(n) \omega_{kj}^{l+1}(n), & \text{για τον νευρώνα } j \text{ στο κρυφό επίπεδο } l \end{cases}$$

όπου το σύμβολο της πρώτης παραγώγου στην $\dot{\phi}_j(\cdot)$ συμβολίζει την διαφόριση αναφορικά ως προς το όρισμα και ο υπολογιστικός γράφος αντικατοπτρίζει μια διαδικασία αλυσιδωτής παραγώγισης που είναι σε θέση να υπολογίσει παραγώγους υψηλής τάξης.

- Ταυτόχρονα επιτελείται και η διαδικασία ενημέρωσης των βαρών στο επίπεδο l που σύμφωνα με τον γενικευμένο κανόνα Δέλτα αντιστοιχεί:

$$\omega_{ji}^{(l)}(n+1) = \omega_{ji}^{(l)} + \alpha[w_{ji}^{(l)}(n-1)] + \eta \delta_j^{(l)}(n) y_i^{(l-1)}(n),$$

όπου η είναι το βήμα εκπαίδευσης (*learning rate*), και α είναι η σταθερά ορμής.

2.4.2 Βασικό Πρόβλημα Ισοζυγίου Bias-Variance

Τα νευρωνικά δίκτυα αντιμετωπίζουν ένα βασικό πρόβλημα που έχει τις ρίζες στον τρόπο εκπαίδευσης τους. Από αυτό το πρόβλημα εκπορεύονται δύο φαινόμενα που αποτελούν συγκοινωνούντα δοχεία.

1. Το σφάλμα μεροληψίας (*Bias Error*), το οποίο είναι το σφάλμα που προκύπτει από υψηλό σφάλμα κατά την εκπαίδευση. Υψηλό σφάλμα μεροληψίας οδηγεί σε χάσιμο των σχέσεων των δεδομένων και των χαρακτηριστικών τους από τον στόχο και γενικά μέτρια αποτελέσματα (*under fit*), τα οποία όμως εμφανίζονται σε μεγάλο εύρος δεδομένων.
2. Το σφάλμα διακύμανσης (*Variance Error*), αναφέρεται στην ανοχή του δικτύου σε διαφορετικά δεδομένα. Υψηλό σφάλμα διακύμανσης εμφανίζεται όταν το δίκτυο έχει υπερεκπαίδευτεί (*over fit*) σε συγκεκριμένα δεδομένα και καταλήγει να τα παπαγαλίζει με συνέπεια να εμφανίζει μεγάλη διακύμανση ακρίβειας σε διαφορετικά σύνολα δεδομένων. Επί της ουσίας ο ορός διακύμανση αναφέρεται στην διαφορά στην ακρίβειας σε διαφορετικά δεδομένα.

Υψηλό σφάλμα διακύμανσης ή διακύμανση ακρίβειας του μοντέλου σε διαφορετικά δεδομένα εισόδου, συνοδεύεται με χαμηλό σφάλμα μεροληψίας και το ανάποδο. Στην μηχανική μάθησή στόχος είναι η να βρεθεί η χρυσή τομή κάτι που επιτυγχάνεται με την θεωρία ομαλοποίησης.

2.4.3 Αρχικοποίηση Βαρών - Θεωρία Ομαλοποίησης

Σύμφωνα με [22] μια διαδικασία προ-εκπαίδευσης των δικτύων θα μπορούσε να δώσει την αρχική τιμή των βαρών. Ωστόσο, αυτό πολλές φορές δεν είναι επιθυμητό, καθώς το δίκτυο μεροληπτεί πλέον προς αυτά τα δεδομένα (σε περίπτωση που δεν μαθαίνει μόνο κανόνες εκπαίδευσης). Μια πιο συνετή επιλογή που είναι εφαρμόσιμη στην δική μας περίπτωση για την επίλυση «κακώς τοποθετημένων» προβλημάτων όπως το πρόβλημα που πραγματεύεται η παρούσα εργασία, βασίζεται στην θεωρία ομαλοποίησης [Regularization Networks [14] Κεφ.7 Tikhonov]. Η αρχική ιδέα είναι να κάνει ευσταθή την έξοδο ενός μη-αρνητικού μηχανισμού έμμεσης συνάρτησης υπό την έννοια στην περίπτωση της ανακατασκευής: Για να είναι ομαλή μια αντιστοίχηση εισόδου-εξόδου, όμοιες είσοδοι πρέπει να παράγουν όμοιες εξόδους. Επομένως τα γραφικά που αναπαρίστανται μέσω έμμεσης συνάρτησης πρέπει να παίρνουν ως είσοδο όμοια σημεία, δηλαδή σημεία ισομετρικού χώρου ή γενικότερα πολλαπλότητας.

Για τον σκοπό αυτό γίνεται έμμεση γεωμετρική ομαλοποίηση (IGR Implicit Geometric Regularization [23]), του δικτύου που αναλαμβάνει την ανακατασκευή της γεωμετρίας έτσι ώστε αρχικά το δίκτυο να αποδίδει γεωμετρικά μια μοναδιαία ισομετρική σφαίρα η οποία είναι όμοια ως προς τις ιδιότητες με το προς ανακατασκευή μοντέλο. Περισσότερα στην επισκόπηση στο κεφάλαιο 3.1.

2.4.4 Αντικειμενική Συνάρτηση Βελτιστοποίησης - Loss Function

Στην διαδικασία της επιβλεπόμενης εκπαίδευσης ενός νευρωνικού $f(x; \theta, \gamma)$ ορίζεται μια συνάρτηση $L(\hat{y}, y)$, η οποία ονομάζεται αντικειμενική συνάρτηση κόστους η οποία αποτελεί μέτρο σύγκρισης των πραγματικά αποδεκτών εξόδων του δικτύου y (*Ground Truth*), με τις εκτιμήσεις που δίνει το δίκτυο \hat{y} . Η αναμενόμενη τιμή της συνάρτησης ως ρίσκο του δικτύου ή κόστος.

$$J(\theta, \gamma) = E_{(x,y)}[L(\hat{y}, y)],$$

το οποίο προσπαθεί να ελαχιστοποιήσει το δίκτυο εκτιμώντας κατάλληλες παραμέτρους θ, γ . Ετσι κάνει εκτίμηση των παραμέτρων με την παρακάτω διαδικασία βελτιστοποίησης σε ένα πλήθος εξόδων:

$$\hat{\theta}, \hat{\gamma} = \operatorname{argmin}_{\theta, \gamma} \left\{ \frac{1}{n} \sum_{i=1}^n L(\hat{y}_i, y_i) \right\}$$

Η συνάρτηση $e_j(n)$ που αναφέρεται στην διαδικασία εκπαίδευσης μέσω του αλγορίθμου οπισθοδιάδοσης βαρών αποτελεί την αντικειμενική συνάρτηση βελτιστοποίησης του δικτύου. Εφόσον το πρόβλημα είναι ξεκάθαρα επιβλεπόμενης μάθησης έχουμε δεδομένα που αποτελούν στόχους του δικτύου και μπορεί να κριθεί η τιμή της συνάρτησης από την απόσταση αυτών των δεδομένων με τα δεδομένα που παράγει το δίκτυο. Για την επιλογή της συνάρτησης βελτιστοποίησης πρέπει να γίνει πρώτα ένα ξεκαθάρισμα στο ποια είναι η διαδικασία που επιτελεί το δίκτυο.

Σε περιπτώσεις που το δίκτυο κάνει εκτίμηση για το αν ένα δεδομένο ανήκει σε μια συγκεκριμένη κλάση στόχο και σε καμία άλλη αυτή η συνάρτηση μπορεί να είναι το Ετερο-εντροπικό Σφάλμα (*Cross Entropy Loss*) των νευρώνων της εξόδου που αναπαριστούν πιθανότητες.

Σε περιπτώσεις που το δίκτυο είναι δίκτυο παλινδρόμησης η έννοια της αντικειμενικής συνάρτησης κόστους μπορεί να πάρει πολλές μορφές ανάλογα και με την μορφή των δεδομένων εισόδου (εικόνα, τρισδιάστατη πληροφορία, βίντεο, μονοδιάστατο σήμα) και το πόσο αναγκαία συνθήκη αποτελεί ο μηδενισμός του 2.4.4. Ετσι υπάρχουν οι L1, L2 νόρμες που ορίζουν την απόσταση και την μέση τετραγωνική απόσταση αλλά και το μέσο απόλυτο ποσοστιαίο σφάλμα *MAPE*(Mean absolute percentage error) που δίνει μεγαλύτερη βαρύτητα όχι σε αριθμητικά σφάλματα αλλά σε ποσοστιαία.

Στο κεφάλαιο 4 της μεθοδολογίας γίνονται συγκεκριμένες περιγραφές των συναρτήσεων που χρησιμοποιήθηκαν για τον έλεγχο του πεδίου SDF αλλά και την εκτίμηση του φωτορεαλιστικού κόστους αλλά και τον έλεγχο του δικτύου με χρήση μάσκας εικονοστοιχείων κατά την απόδοση σκηνών για την αποφυγή αστάθειας σε μεγάλη κλίμακα επιβλεπόμενης μάθησης.

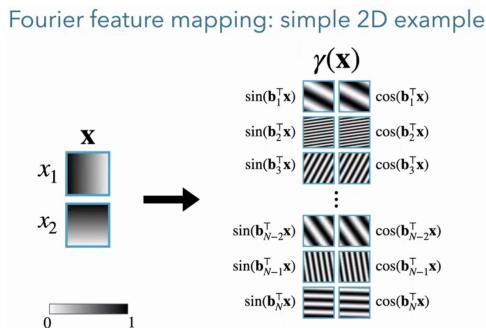
2.5 Δίκτυα & Μηχανισμοί Κωδικοποίησης 3D Εισόδου

DeepSDF Frequency Feature Embeddings

Στο πλαίσιο της καλύτερης δυνατής ανακατασκευής 3D γραφικών όσον αφορά την ακρίβεια του σήματος που ισοδυναμεί με την αποτύπωση υψηλοσυχνοτικών περιοχών εισάγεται η έννοια των δικτύων κωδικοποίησης. Αυτός είναι η βασική μεθοδολογία με την οποία δίνεται η δυνατότητα η εκπαίδευση των νευρωνικά ελεγχόμενων διαφορίσμων ισομετρικών επιφανειών (*Neural Level Sets*) να μπορέσει να συλλάβει υψηλοσυχνοτικό περιεχόμενο. Επομένως, η βάση της διπλωματικής είναι οι τρισδιάστατες κωδικοποίησεις 3D πληροφορίας με χρήση βαθιών νευρωνικών δικτύων που αναλύονται παρακάτω.

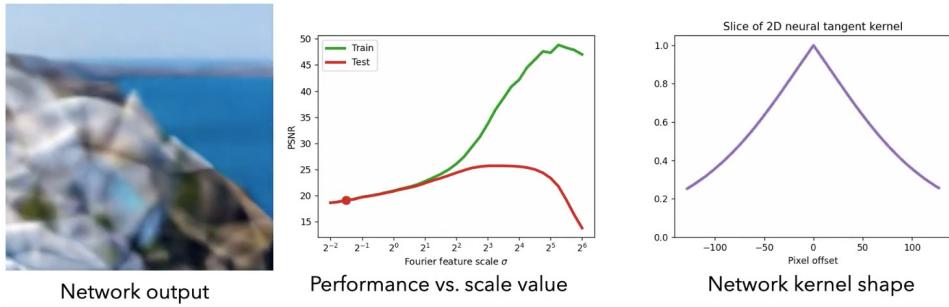
2.5.1 Κωδικοποίηση μέσω συναρτήσεων εφαπτόμενου νευρωνικού πυρήνα Fourier - Fourier Feature Neural Tangent Kernel

Παρά του θεωρήματος γενικευμένης δυνατότητας αναπαράστασης συναρτήσεων από πλήρως συνδεδεμένα γραμμικά δίκτυα MLP (ακόμα και μονο-στρωματικά), τα τεχνητά νευρωνικά δίκτυα στην πραγματικότητα δεν μπορούν να ανταποκριθούν πλήρως σε αυτήν την ιδιότητα. Θεωρητικά πρέπει να ισχύουν, οριακές συνθήκες που προβλέπουν να υπάρχει άπειρο πλήθος νευρώνων στο στρώμα, ενώ το βήμα εκπαίδευσης στον αλγόριθμο βελτιστοποίησης για παράδειγμα SGD να (*Stochastic Gradient Descent*) τείνει στο 0. Αυτό το πρόβλημα καλούνται να λύσουν εξειδικευμένες συναρτήσεις βελτιστοποίησης, τεχνικές αποκοπής νευρώνων αλλά ακόμα και ειδική προ επεξεργασία δεδομένων που ωστόσο αποτυγχάνουν μιας και δεν καταφέρνουν να συλλάβουν τις σχέσεις δεδομένων σε ανώτερες διαστάσεις. Σε αυτό το πλαίσιο επιδρούν πυρήνες συναρτήσεων ή δίκτυα κωδικοποίησης που μετασχηματίζουν την είσοδο εισάγοντας ποιοτικά χαρακτηριστικά που σχετίζονται με την έξοδο προς αναπαράσταση κάνοντας αντιστοίχηση των εισόδων των δικτύων σε χώρους μεγαλύτερης διάστασης ή με ιδιαίτερα χαρακτηριστικά.

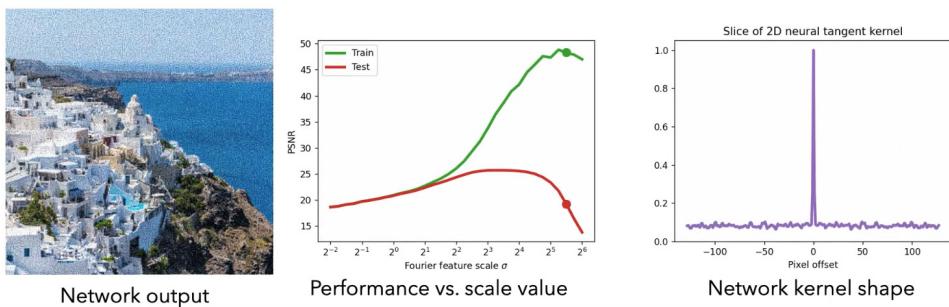


Σχήμα 2.15: Οπτική Παρουσίαση Εφαρμογής Fourier Feature Mapping σε Δισδιάστατες Εικόνες. Πηγή:([24])

Παρατηρώντας το Σχήμα 2.15, γίνεται αντιληπτό το πως λειτουργεί ένας εφαπτόμενος πυρήνας Fourier συχνοτικής κωδικοποίησης εισόδου. Στα πλαίσια των νευρωνικών δικτύων αυτό που πρακτικά εφαρμόζεται στην συνέχεια είναι μια παλινδρόμηση πυρήνα ώστε να γίνει εκπαίδευση των ενσωματωμένων συχνοτήτων που διαθέτει πλέον η κωδικοποιημένη είσοδος. Σημαντικό ρόλο σε αυτό το κομμάτι παίζει το εύρος των συναρτήσεων πυρήνα. Σε περίπτωση που το εύρος είναι πολύ μεγάλο παρατηρείται αξιόπιστη, αλλά πολύ ομαλοποιημένη απόδοση χαρακτηριστικών ενώ σε μικρό εύρος συναρτήσεων πυρήνα παρατηρούνται φαινόμενα μη σωστής παλινδρόμησης και συνεπώς μη συνεπής ανακατασκευή που προέρχεται από φαινόμενα υπερεκπαίδευσης σε συγκεκριμένα χαρακτηριστικά.



Σχήμα 2.16: Ανακατασκευή εικόνας με Κωδικοποίηση Ευρύ Εφαπτόμενου Πυρήνα Συναρτήσεων Fourier (μεγάλο σ κατανομής δειγματοληψίας συχνοτήτων), Πηγή [24]



Σχήμα 2.17: Ανακατασκευή εικόνας με Κωδικοποίηση Στενού Εφαπτόμενου Πυρήνα Συναρτήσεων Fourier (μικρό σ), Πηγή [24]

Αυτή η τεχνική μπορεί να χρησιμοποιηθεί και σε τρισδιάστατη κωδικοποίηση πληροφορίας (τρισδιάστατα σημεία εισόδου, τρισδιάστατα διανύσματα κατεύθυνσης όψης) και επιφέρει ανάλογα αποτελέσματα στις έμμεσες αναπαραστάσεις που δημιουργούνται τόσο στο κομμάτι της απόδοσης χρώματος όσο και στην γεωμετρική απόδοση.

2.5.2 Χρήση χωρικής πληροφορίας στον εφαπτόμενο νευρωνικό πυρήνα συναρτήσεων - Positional Encoding

Μια ειδική κατηγορία της παραπάνω μεθόδου κωδικοποίησης είναι το *Positional Encoding*. Η πρώτη εφαρμογή αυτού γίνεται στα πλαίσια της κωδικοποίησης της εισόδου σε μηχανισμού προσοχής σε δίκτυα που αφορούν την ερμηνεία φυσικής γλώσσας (*Natural Language Processing*). Στον τομέα της εργασίας πρώτη φορά αυτού του είδους η κωδικοποίηση της εισόδου εφαρμόστηκε σε δίκτυα που μαθαίνουν την ακτινοβολία μιας σκηνής *Neural Radiance Fields* [24].

Η παλινδρόμηση πυρήνα ή χρήση συναρτήσεων πυρήνα σε νευρωνικά δίκτυα παλινδρόμησης είναι μια συνήθης τακτική. Ωστόσο, αυτή η μέθοδος έχει πολλά να προσφέρει και ως ένα τρόπο διαμεσολάβησης σε γραμμικά δίκτυα MLP που δεν είναι πολλές φορές σε θέση να καταγράψουν συχνοτικό περιεχόμενο, ειδικά όταν αυτά τα δίκτυα είναι δίκτυα συντεταγμένων (*coordinate based networks*). Όπως είναι και τα βαθιά δίκτυα SDF.

Εστω f να αναπαριστά ένα πλήρως συνδεδεμένο νευρωνικό δίκτυο με βάρη θ που είναι αρχικοποιημένα σε με βάση μια Γκαουσιανή κατανομή N .

Το εύρος των στρωμάτων, δηλαδή το πλήθος νευρώνων τείνει στο άπειρο και το βήμα εκμάθησης στον αλγόριθμο back propagation που βελτιστοποιεί τα βάρη να τείνει στο 0.

Όταν συμβαίνει αυτό, οι συναρτήσεις τύπου $f(\chi; \theta)$ συγκλίνουν κατά την διάρκεια της εκπαίδευσης κάτι

το οποίο προσομοιώνει, παλινδρόμηση πυρήνα ή καλύτερα *kernel regression* που αναλογεί στην χρήση ενός εφαπτόμενου νευρωνικού δικτύου με συναρτήσεις πυρήνα, (*NTK Neural (Stationary) Tangent Kernel*):

$$K_{ntk}(x_i; x_j) = E_{\theta \sim N} < \frac{\partial f(x_i; \theta)}{\partial \theta}, \frac{\partial f(x_j; \theta)}{\partial \theta} >$$

Όταν οι είσοδοι είναι περιορισμένες σε μια υπερσφαίρα, ο εφαπτόμενος νευρωνικός πυρήνας για ένα MLP δίκτυο μπορεί να γραφτεί ως το εσωτερικό γινόμενο πυρήνα ($h_{NTK}(x_i^T x_j)$, όπου το $h_{NTK} : \mathbb{R} \rightarrow \mathbb{R}$ είναι βαθμωτό μέγεθος)

Χρήση των Χαρακτηριστικών Fourier για τον υπολογισμό εφαπτόμενου πυρήνα κωδικοποίησης

Η εκτίμηση μιας ακατέργαστης στάσιμης συνάρτησης πυρήνα με το θεώρημα Bocher's (Θεώρημα 2.3.1) είναι εφικτή με υψηλότερη ακρίβεια. Για την διαδικασία αυτή χρησιμοποιείται αντιστοίχηση της εισόδου σε χαρακτηριστικά Fourier γ για να δοθεί στις συντεταγμένες εισόδου, συχνοτικό χαρακτηριστικό εμφάνισης, πριν να οδηγηθούν στο βασικό δίκτυο MLP που αναλαμβάνει την την αντιστοίχηση τους πάνω σε ισομετρίες.

Συγκεκριμένα η συνάρτηση γ αντιστοιχεί τα δεδομένα εισόδου, που είναι ένα διάνυσμα $\vec{v} \in [0, 1]^d$, σε ισομετρική επιφάνεια μια υψηλότερου βαθμού και στην περίπτωση της εργασίας μια υπερσφαίρα (λόγω και έμμεσης γεωμετρικής ομαλοποίησης δικτύου SDF [23]) με ένα σύνολο ημιτονοειδών συναρτήσεων που επηρεάζονται από ένα παράγοντα κλίμακας ή βάρους α .

$$\vec{\gamma}(v) = [\alpha_1 \cdot \cos(2\pi b_1^T v), \alpha_2 \cdot \sin(2\pi b_2^T v)], \dots, \alpha_m \cdot \cos(2\pi b_m^T v)]^T$$

Επειδή από την βασική τριγωνομετρία, $\cos(\alpha - \beta) = \cos a \cdot \cos b - \sin a \cdot \sin b$ αυτή η συνάρτηση πυρήνα και η κωδικοποίηση που συνεπάγεται εισάγει στις συντεταγμένες τον εξής μετασχηματισμό:

$$K_\gamma = \gamma(v_1)^T \cdot \gamma(v_2) = \sum_{j=1}^m \alpha_j^2 \cos(2\pi b_j^T (v_1 - v_2)) = h_\gamma(v_1 - v_2)$$

που αποτελεί μια στάσιμη εφαρμογή πυρήνα κωδικοποίησης Fourier, δηλαδή μια συνάρτηση που διατηρεί την ισομετρική ιδιότητα των δεδομένων και είναι αμετάβλητη σε περιστροφές ή μετατοπίσεις μιας και είναι κωδικοποίηση που αφορά μόνο την απόσταση των σημείων για αυτό αποκαλείται και κωδικοποίηση τοπολογική ή *positional encoding*.

Με την χρήση αυτής την κωδικοποίησης σαν πυρήνας διαμεσολάβησης στο δίκτυο γεωμετρικής αποτύπωσης και απόδοσης, έστω f , καταλήγουμε σε μια νέα μορφή αναπαράστασης που έχει ενσωματωμένη την συχνοτική πληροφορία του χώρου με παραμέτρους προς εκπαίδευση, τις α_i, b_j που αποτελούν το χωρικό μέτρο και συχνότητα των συναρτήσεων βάσης.

$$\hat{f} = (h_{NTK} \circ h_\gamma) \cdot \sum_{i=1}^n \theta_i \cdot \delta_{v_i}$$

2.5.3 Κωδικοποίηση με χρήση Σφαιρικών Αρμονικών [25]

Οι σφαιρικές συντεταγμένες είναι μια μεθοδολογία που αξιοποιείται σε σύγχρονες μεθόδους ανακατασκευής σκηνών [26], για την πρόβλεψη του πεδίου ακτινοβολίας καθώς διαφέρει από όψη σε όψη και οι εξωτερικές παράμετροι της κάμερας στο σύνολο δεδομένων περιγράφουν θέσεις σε τροχιά γύρω από το αντικείμενο προβολής.

Η γενική μορφή των σφαιρικών αρμονικών είναι η εξής:

$$Y_l^m(\theta, \phi) = K_l^m P_l^m(\cos(\theta)) e^{im\phi}$$

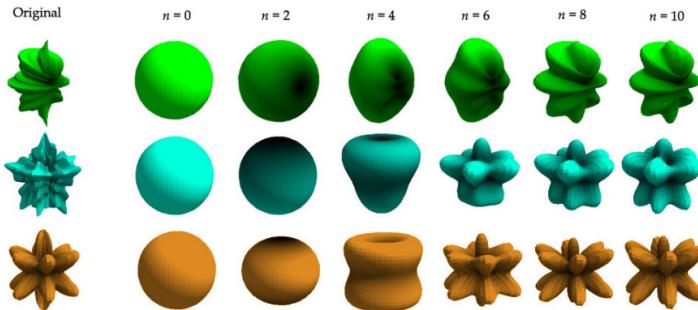
Στην πραγματική της μορφή η παραπάνω εξίσωση δίνει τον εξής τύπο:

$$Y_l^m(\theta, \phi) = \begin{cases} \sqrt{2} K_l^{-m} P_l^{-m}(\cos(\theta)) \sin(-m\phi) & \text{αν } m < 0 \\ K_l^0 P_l^0(\cos(\theta)) & \text{αν } m = 0 \\ \sqrt{2} K_l^m P_l^m(\cos(\theta)) \cos(m\phi) & \text{αν } m > 0 \end{cases}, \quad (2.5)$$

όπου $K_l^m = \sqrt{\frac{(2l+1)}{4\pi} \frac{(l-m)!}{(l+m)!}}$ ο οποίος αποτελεί παράγοντα κανονικοποίησης που βασίζεται στα πολυώνυμα Legendre[27].

Για την ανακατασκευή ενός σήματος που έχει κωδικοποιηθεί με σφαιρικές αρμονικές στους συντελεστές C_l η αποκωδικοποίηση τους ακολουθεί τον εξής τύπο:

$$f(\theta, \phi) \simeq \sum_{l=0}^N \sum_{m=-l}^l C_l^m Y_l^m(\theta, \phi), \quad \text{όπου } N \text{ είναι ο βαθμός των σφαιρικών αρμονικών}$$



Σχήμα 2.18: Ανακατασκευή Σήματος με χρήση σφαιρικών αρμονικών, πηγή [25](Robin Green)]

Έτσι προκύπτει μια αποδοτική κωδικοποίηση των διανυσμάτων προβολής στα δίκτυα ακτινοβολίας.

2.5.4 Τρισδιάστατος χωρικός κατακερματισμός- Multi-resolution Hash 3D Encoding

Νέες μέθοδοι αναφέρονται στην κωδικοποίηση έμμεσων αναπαραστάσεων που γενικότερα αναφέρονται και στην βιβλιογραφία ως πρότυπα γραφικών (*Neural Graphics Primitives* [2]). Αυτές, καταδεικνύουν πως γεωμετρίες έμμεσων πεδίων προσημασμένης απόστασης, ογκομετρικών πεδίων, πεδίων ακτινοβολίας, ακόμα και εικόνων όταν κωδικοποιούνται με πυρήνες συναρτήσεων που τα αντιστοιχούν σε χώρους μεγαλύτερων διαστάσεων, για να αποτυπώσουν ανώτερες συσχετίσεις των δεδομένων, πρακτικά δημιουργούν δομές που εκπαιδεύονται δύσκολα. Έτσι βασίζονται σε ευριστικά ή δομικά τεχνάσματα που εφαρμόζονται στα δίκτυα κάποια από τα οποία εμφανίζονται στα πλαίσια της θεωρίας ομαλοποίησης και άλλα βασίζονται στην εκμάθηση υπολεπτομένων αναπαραστάσεων σε *ResNet*. Αυτό γίνεται μιας και η απόδοση της εκπαίδευσης των δικτύων μειώνεται δραματικά και ο έλεγχος σε δεδομένα υψηλής διάστασης, όπου δεικτοδότηση είναι ακριβή υπολογιστικά, είναι χρονοβόρος και απαιτεί μεγάλο αριθμό πόρων.

Αυτό καλούνται να λύσουν οι μέθοδοι 3D κωδικοποίησης που κάνουν χωρικό κατακερματισμό σε πολλαπλά επύπεδα ανάλυσης δίνοντας αυτόματα προτεραιότητα σε αραιά χαρακτηριστικά περιοχών και σταδιακά αυξάνοντας την ανάλυση κωδικοποίησης οδηγούν στην εκμάθηση και λεπτομερειών υψηλής ακρίβειας.

Μια τέτοια μέθοδος είναι και η κωδικοποίηση κατακερματισμού *Hash Encoding* που εφαρμόζεται σε δίκτυα συντεταγμένων και δημιουργεί χάρτες χαρακτηριστικών σε αυξανόμενες αναλύσεις επιμεριζοντας το πρόβλημα σε απλούστερα υπό προβλήματα τα οποία είναι σε θέση να είναι αποσυσχετισμένα μεταξύ τους μιας

και επιλέγεται διαδικασία κατακερματισμού που επιτρέπει με χρήση της πράξης XOR(*Exclusive OR*) στις συντεταγμένες και έχει τη δυνατότητα να αποφύγει από μόνη της συμπίπτουσες κωδικοποιημένες συντεταγμένες.

$$h(\mathbf{x}) = \bigoplus_{i=1}^d x_i \pi_i,$$

όπου d η διάσταση των δεδομένων στον πίνακα κατακερματισμού και π_i διάνυσμα ακεραίων αριθμών, με τις οποίες εφαρμόζεται αποκλειστική διάζευξη σε επίπεδο bit (*Bitwise XOR*).

Οι χάρτες κατακερματισμένων χαρακτηριστικών (άτλαντες διαφορίσιμων πολλαπλοτήτων) σε διάφορες αναλύσεις του χώρου έχουν συγκεκριμένο μέγεθος και αντιστοιχίζονται με παρεμβολή (χρήση interpolation στα χαρακτηριστικά κελιών) σε βάρη, όπως εκτελούν και στρώματα προσοχής (*Attention Layers* [8]).

Οι χωρικές ιεραρχίες (*spatial hierarchical encoding*) που εισάγονται με αυτού του τύπου των κωδικοποιήσεων, επιτρέπουν με ίδιο αριθμό παραμέτρων πολύ πιο γρήγορη κωδικοποίηση και εκπαίδευση δίνοντας ανάλογα αποτελέσματα χωρίς να κωδικοποιούνται δύσκολες συσχετίσεις δεδομένων σε υψηλές διαστάσεις. Περισσότερα στην μεθοδολογία Κεφ.4.

2.5.5 Θυρίδες Κατακερματισμένων φίλτρων Fourier NFFB [1]

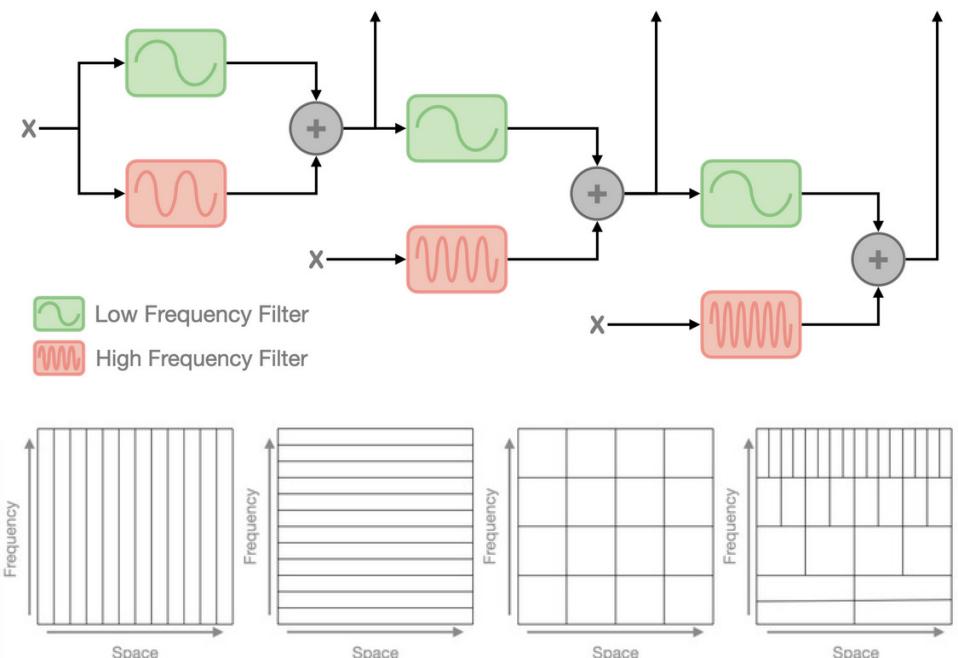
Παρά την επιτυχία της προηγούμενης μεθόδου, παρατηρούνται προβλήματα όταν γίνεται αποκλειστική χρήση χαρτών κατακερματισμού στην εκπαίδευση πεδίων προσημασμένης απόστασης που έχουν να κάνουν με την ποιότητα του αποτελέσματος εμφανίζονται *artifacts*. Επομένως μια καλή επιλογή είναι ο συγκερασμός των δύο μεθόδων με το σκεπτικό του μετασχηματισμού κυματιδίων *wavelets*. Αυτό, συνδυάζει τις δύο παραπάνω μορφές κωδικοποίησης ταυτόχρονα και αρμονικά μεταξύ τους και ονομάζεται Νευρωνικές Θυρίδες κατακερματισμένων Φίλτρων Fourier ή *Neural Fourier Filter Banks* [1] τα οποία πλέον αποτελούν τον άτλαντα χαρακτηριστικών των αμφισδιαφορίσιμων πεδίων γεωμετρίας και φωτισμού.

Μετασχηματισμός Κυματιδίων - Wavelet Transform

Η χρήση του μετασχηματισμού κυματιδίων, έχει ερευνηθεί αρκετά στην βιβλιογραφία της βαθιάς μάθησης. Για παράδειγμα, χρησιμοποιούνται για δειγματοληψία χαρακτηριστικών βασισμένη στα κυματίδια (*wavelet-based feature pooling*). Τέτοια διαδικασία, εισάγει βελτιώσεις στο κομμάτι της μεταφοράς χαρακτηριστικών ή στην αποθορυβοποίηση εικόνων για ιατρική ανάλυση. Στο κομμάτι των παραγωγικών-αντιπαραθετικών δικτύων (*GANs*) που γεννούν εικόνες κάποιες υλοποιήσεις βασίζονται εξολοκλήρου στην αποσύνθεση σημάτων με χρήση μετασχηματισμού κυματιδίων. Άλλες, προτείνουν επαυξημένο διαχωριστή (*discriminator*, δομή των *GAN* δικτύων που προσπαθεί να διαχωρίσει τα παραγόμενα από τα πραγματικά δεδομένα), πάνω στις υποζώνες(*sub bands*) συχνοτήτων που προκύπτουν από έναν μετασχηματισμό κυματιδίων. Η χρήση του λοιπόν είναι σημαντική και μπορεί να εφαρμοστεί στην κωδικοποίηση τρισδιάστατης πληροφορίας με εφαρμογές στην 3D όραση.

Η ουσία των νευρωνικών θυρίδων Fourier είναι πως λειτουργούν παρόμοια με έναν μετασχηματισμό κυματιδίων. Συγκεκριμένα, γίνεται αναπαράσταση χαρακτηριστικών σε μια κλίμακα κατακερματισμού που επιτρέπει την απόκτηση των πρωτότυπων βαθιών χαρακτηριστικών από μητρικά χαρακτηριστικά μέσω αλλαγής της κλίμακας του χώρου δηλαδή της ανάλυσης του κατακερματισμού. Έτσι αποκτάμε θυρίδες που έχουν χαρακτηριστικά τα οποία μπορούν να αντιστοιχούν σε δεδομένα σε υψηλές διαστάσεις αλλά δεν χρησιμοποιούνται απευθείας στην διαδικασία εκπαίδευσης όπως είναι, με αποτέλεσμα να οδηγούμαστε με μικρές, αλλά ουσιώδης μορφές κωδικοποιήσεων με υψηλή ακρίβεια αναπαραστάσεων χωρίς λανθάνων τεχνουργήματα (*artifacts*).

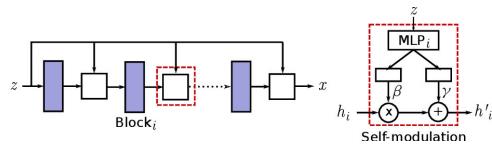
Αυτή η διαδικασία αναπαριστάται συμβολικά στις παρακάτω εικόνες που δίνεται η διαφορά στους τρόπους χωρικής/συχνοτικής κωδικοποίησης και πως γίνεται ο συνδυασμός τους.



Σχήμα 2.19: Αναπαράσταση Wavelet κατακερματισμένης κωδικοποίησης υψηλών και χαμηλών συχνοτήτων σε μια και δύο διαστάσεις με χρήση φίλτρων, πηγή [1]

2.6 Διαμόρφωση Δικτύων Κωδικοποίησης Υψηλοσυχνοτικού Περιεχομένου

Στα πλαίσια της διαμόρφωσης των δικτύων κωδικοποίησης, εντάσσονται όλες οι τεχνικές που ασχολούνται με την διαδικασία κανονικοποίησης των εξαγόμενων χαρακτηριστικών και έχουν να κάνουν συνήθως με το ποιά χαρακτηριστικά ευνοεί και ποια τα χαρακτηριστικά το ρίχνει το δίκτυο. Μια ικανοποιητική ανακατασκευή απαιτεί κατά περίπτωση έκφραση όλων των χαρακτηριστικών που είναι απαραίτητα. Επομένως ακολουθούνται τεχνικές προσαρμοστικής κανονικοποίησης σε κάθε χαρακτηριστικό κατά την διάρκεια της εκπαίδευσης τα οποία περιγράφονται τόσο ως έμπνευση από εργασίες που ήδη χρησιμοποιούν αυτές τις τεχνικές με επιτυχία όπως το *StyleGAN2* [28]. Στο κομμάτι της μεθοδολογίας που αναφέρεται πώς ακριβώς εφαρμόστηκε ο μηχανισμός αυτός που βασίστηκε περισσότερο στην έννοια της προσοχής. [8].



Σχήμα 2.20: Αυτό-Διαμόρφωση MLP δικτύου, πηγή [29]

3

Επισκόπηση της Ερευνητικής Περιοχής

3.1 Τεχνικές Βαθιάς Μάθησης, Τρισδιάστατης Ανακατασκευής και Κωδικοποίησης

Ο τομέας της υπολογιστικής όρασης(*Computer Vision*) καθώς και της αποτύπωσης 3D δεδομένων του χώρου σε εικονικά περιβάλλοντα (*Photogrammetry, 3D Scanning*) έχει γίνει επιτακτικός τομέας απασχόλησης του επιστημονικού δυναμικού με αποκορύφωμα να υπάρχουν εργασίες βιβλιογραφικής επισκόπησης επί του συνόλου των σύγχρονων μεθόδων ανακατασκευής [7], λόγω εφαρμογών που απαιτούν είτε εικονικές συνθήκες αξιολόγησης (κίνηση ρομπότ στον χώρο), είτε αποτύπωση πραγματικών αντικειμένων σε εικονικά περιβάλλοντα για χάρη του ρεαλισμού (*Ρεαλιστικά περιβάλλοντα γραφικών*), είτε ακόμα και παραγωγή 3D δεδομένων που δεν έχουμε (πχ. ανακατασκευή *Παρθενώνα* όπως ήταν επί χρυσού αιώνα), χωρίς την μεγάλη προσπάθεια των κλασσικών μεθόδων που δεν έχουν μεθόδους προσέγγισης και ανακατασκευής δεδομένων. Στο μεταξύ, η αναλυτικότητα που προσφέρουν τα δίκτυα βαθιάς μάθησης και η δυνατότητα τους να μαθαίνουν πεδία του χώρου τα οποία περιγράφονται δύσκολα χωρίς όλα τα απαραίτητα δεδομένα και συγκεκριμένα το να μαθαίνεται ένα 3D σχήμα παρέχοντας μόνο εικόνες διευκολύνει το εγχείρημα ανακατασκευής χωρικών δεδομένων και μοντέλων.

Οι εργασίες που ασχολούνται με την παραμόρφωση και γενικότερα την επεξεργασία άμεσων αναπαραστάσεων είναι πολυπληθείς και έχουν χρησιμοποιηθεί κατά καιρούς σε μοντέλα εκτίμησης ανθρώπινου σώματος [30], [31], [32] και προσώπου [33]. Βέβαια, αυτές οι μέθοδοι εφόσον έχουν διαθέσιμη την άμεση 3D πληροφορία δεν αφορούν άμεσα την ανακατασκευή που πραγματοποιείται με μοντέλα που χρησιμοποιούν απλά εικόνες για επίβλεψη.

Οι έρευνες που πρώτες επέκτειναν αυτή την λογική παραμόρφωσης βασιζόμενες στην παραμόρφωση πρωτότυπων γραφικών (*Deform one primitive*) ανακατασκευάζοντας 3D σχήματα με χρήση κωδικοποίησης πληροφορίας αλλά επέμειναν στην τρισδιάστατη επίβλεψη νεφών σημείων (*Point Clouds*) οδήγησαν σε εργασίες όπως: *Pixel2Mesh*[34] και *Pixel2Mesh++*[35] που αντιμετωπίζει το πρόβλημα σε κάτω από πολλές προβολές.

Ένας συγκεντρωτικός πίνακας που δείχνει όλες τις εργασίες με τον αντίστοιχο τρόπο με τον οποίο χειρίζονται τις τρισδιάστατες αναπαραστάσεις με πλήθος διαφορετικών αρχιτεκτονικών δικτύων (αναφορικά *CNN*, *RNN*, *MLP*), χρήση και επιβλεπόμενης, μη-επιβλεπόμενης, ημι-επιβλεπόμενης αλλά και αυτό-ενισχυτικής εκπαίδευσης (*reinforcement learning*), φαίνεται παρακάτω.

Διαδικασία	Οικογένεια Αναπαραστάσεων	Ερευνα
Deform one template	Mesh	”Human from single image” [36]
Retrieve and deform template	Mesh	ShapeFlow [37]
Deform one primitive	Mesh	Pixel2Mesh [34]
Deform multiple primitives	Mesh	AtlasNet [38]
Set of primitives	Mesh	BSP-Net [39]
Primitive detection	Parametric surface	ParSeNet [40]
Grid mesh	Mesh	Neural Marching Cubes [41]
Grid polygon soup	Mesh	Adaptive O-CNN [42]
Grid voxels	Implicit	3D-R2N2 [43]
Neural implicit	Implicit	IM-Net [44], DeepSDF [45]
Primitive CSG	CSG tree	CAPRI-Net [46]
Sketch and extrude	CSG tree	ExtrudeNet [47]
Connect given vertices	Mesh	PointTriNet [48]
Generate and connect vertices	Mesh	PolyGen [49]
Sequence of edits	Mesh	Modeling 3D Shapes by RL [50]

Πίνακας 3.1: Πίνακας Ερευνών πάνω στις 3D Αναπαραστάσεις [7]

Πρόσφατα αρκετές ερευνητικές εργασίες έχουν ασχοληθεί πάνω στην έμμεση αναπαράσταση συμπαγών τρισδιάστατων αντικείμενων, βελτιστοποιώντας απευθείας νευρωνικά δίκτυα που αντιστοιχούν x, ψ, z τρισδιάστατες συντεταγμένες με εκπαίδευση όμενη γεωμετρία σε πεδία προσημασμένης απόστασης (*SDF*) που είδαμε προηγουμένως ή δίκτυα κάλυψης (*voxel occupancy fields*), τα οποία έχουν να κάνουν με άλλες έμμεσες αναπαραστάσεις (*implicit representations*) όπως τα *octrees* (βλ. 2.4). Έχοντας ως απαρχή το *CVPR(Computer Vision and Pattern Recognition Conference)* του 2019 με τρεις ταυτόχρονες εργασίες πάνω στην εκπαίδευση τέτοιων αναπαραστάσεων σε δίκτυα όπως το IM-Net[44], OccNet[51], DeepSDF[45], τα οποία είναι μοντέλα **voxel-based**, δηλαδή μοντέλα που περιγράφουν μη παραμετρικά όγκους με τρισδιάστατα πλέγματα τιμών και είναι ίσως η πιο φυσική επέκταση στον τρισδιάστατο τομέα μάθησης.

Επί της ουσίας, οι έμμεσες νευρωνικές αναπαραστάσεις (όπως είδαμε 2.3), είναι ένα MLP (Multi-Layer Perceptron) το οποίο παίρνει μια συντεταγμένη ενός σημείου ως είσοδο και βγάζει ως έξοδο το εσωτερικό-εξωτερικό πρόσημο [44], [52], ή την απόσταση με πρόσημο αυτού του σημείου από την επιφάνεια [45]. Το ίδιο το MLP έτσι αντιπροσωπεύει ένα τρισδιάστατο σχήμα υπό την μορφή έμμεσης συνάρτησης. Ο τρόπος που τέτοια δίκτυα γεννούν από μόνα τους τρισδιάστατες γεωμετρίες γραφικών πρωτοτύπων έχει αναλυθεί στο θεωρητικό υπόβαθρο (Κεφ.2 2.3) με την βοήθεια ειδικών μορφών γεωμετρικής κανονικοποίησης των βαρών (*Implicit Geometric Regularization* [23]). Βέβαια, για να γεννηθεί κάποια διαφορετική έξοδος με βάση την είσοδο, το MLP θα μπορούσε να ελεγχθεί συνενώνοντας ένα σχήμα σε μορφή *latent code*¹ με τις συντεταγμένες εισόδου του πριν την κλήση της εμπρόσθιας διάδοσης[53], [54].

¹κωδικοποίηση που προκύπτει από συνελικτικό συνήθως δίκτυο CNN

Η εργασία παρά όλη αυτήν την πλούσια και εξειδικευμένη βιβλιογραφία, δεν βασίστηκε σε όλο το φάσμα της, καθώς πολλές από αυτές τις εργασίες εξειδικεύουν το ενδιαφέρον τους σε άλλα πράγματα και όχι στην υψηλοσυχνοτική κωδικοποίηση των επιφανειών και του φωτισμού. Συνεπώς, παρουσιάζονται οι απλές μορφές εργασιών στις έμμεσες μορφές αναπαραστάσεων από πολλές όψεις που απαιτούν ακόμα «κακώς τοποθετημένα» μαθηματικά προβλήματα, όπως στο *MVSR-IDR* [6] (*Multi-view Neural Surface Reconstruction - Implicit Differential Rendering*) και όχι το *NeuS* [4] (*Neural Surface Reconstruction - Volumetric Rendering*) στην ομπρέλα της νευρωνικής απόδοσης σκηνών.

Στο κομμάτι της κωδικοποίησης αντλείται ενδιαφέρον από την χρήση κωδικοποίησης κατακερματισμού (*hash*) στο *Instant-NGP(Neural Graphics Primitives)* [2], η κωδικοποίηση υψηλοσυχνοτικού περιεχομένου φωτισμού όπως το (*NeRF*, ie. *Neural Radiance Field*) [3], [24], αλλά και οι βαθιές κωδικοποίησεις που βασίζονται στην λειτουργία κυματιδίων *NFFB* [1], καθώς και η διαμόρφωση χαρακτηριστικών που εξάγουν τα νευρωνικά δίκτυα όπως αυτή παρουσιάστηκε στο *StyleGan2* [28].

3.2 MVSR-IDR

Ανακατασκευή 3D επιφανειών με χρήση Νευρωνικών Δικτύων μέσω πολλαπλών όψεων διαχωρίζοντας τη γεωμετρία και την εμφάνιση

Στο πλαίσιο της 3D ανακατασκευής με την οποία ασχολείται η παρούσα εργασία σημαντική έρευνα είναι αυτή της *Lior Yariv* με τίτλο: **Ανακατασκευή επιφανειών με χρήση νευρωνικών δικτύων, μέσω έμμεσων αναπαραστάσεων που προκύπτουν από εικόνες πολλαπλών όψεων**, [6, Multiview neural surface reconstruction by disentangling geometry and appearance].

Τα διαφορίσιμα συστήματα απόδοσης γραφικών, που στοχεύουν την εκμάθηση τρισδιάστατης γεωμετρίας, κυκλοφορούν σε δύο εκδοχές: *differentiable rasterization* και *differentiable ray casting*. Στην παρούσα εργασία, εφόσον μιλάμε για έμμεσες επιφάνειες που αποδίδονται ογκομετρικά και σκοπεύεται ο φωτορεαλισμός επιλέγεται η δεύτερη μέθοδος. Γίνεται περιγραφή συναφών εργασιών για την ανακατασκευή επιφανειών πολλαπλών όψεων και την σύνθεση όψεων με χρήση νευρωνικών δικτύων.

Σε εργασίες όπως [55] προσεγγίζεται η τιμή της συνάρτησης πολλών μεταβλητών SDF και των κανονικών της διανυσμάτων σε κάθε ογκομετρικό κελί (voxel). Στην εργασία [56] χρησιμοποιείται ο αλγόριθμος *sphere tracing* (που αναλύθηκε στο 2.2), σε ένα προ εκπαιδευμένο *DeepSDF* μοντέλο [45]. Εκεί, προσεγγίζονται οι παράγωγοι των βαθών ως προς το διάνυσμα εξόδου του *DeepSDF* νευρωνικού δικτύου δηλαδή ως προς το *latent codet* διαφορίζοντας το κάθε ξεχωριστό βήμα του αλγορίθμου του *Differential Sphere Tracing*. Σε αυτή την εργασία [57] χρησιμοποιείται ανιχνευτικό πεδίο για την διευκόλυνση της λειτουργίας του αλγορίθμου ρίψης ακτίνων ώστε να γίνεται εκπαίδευση χωρίς επίβλεψη. Σε αντίθεση με αυτές τις εργασίες, το *IDR* κάνει χρήση διαφορίσιμου σημείου επαφής ακτίνας με επιφάνεια το οποίο έχει ακριβή σε αριθμητική τιμή πρώτη παράγωγο. Επίσης χρησιμοποιείται και το κάθετο διάνυσμα πάνω στο σημείο αυτό με σκοπό την πιο γενικεύσιμη μορφή του μοντέλου εμφάνισης. Τέλος χειρίζεται και κάμερες με θόρυβο.

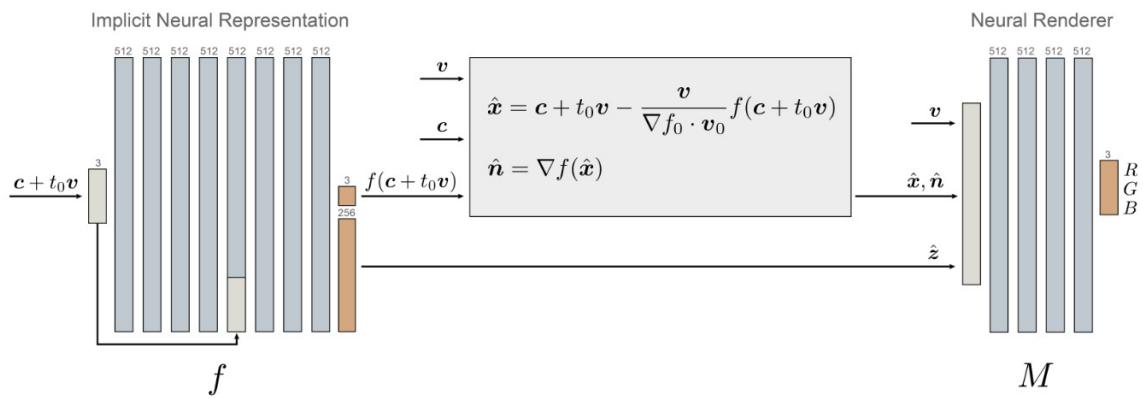
Ανακατασκευή από Πολλαπλές Όψεις - Multi-View Synthesis[MVS]

Κατά την διάρκεια της φωτογράφησης μια εικόνας, η πληροφορία του βάθους χάνεται. Δεδομένης της θέσης και του προσανατολισμού της κάμερας, η κλασσική θεωρία της στερεογραφίας πολλαπλών όψεων (Multi-View Stereo - MVS) έχει μεθόδους [58], [59], [60], [61] που προσπαθούν να αναπαράγουν την πληροφορία του βάθους μέσα από την επικάλυψη ή το ταίριασμα χαρακτηριστικών των εικόνων από διάφορες όψεις. Παρ' όλα αυτά, μια μετά-επεξεργασία βάθους (*depth fusion post processing*) [62], [63] ακολουθούμενη από τον αλγόριθμο ανακατασκευής επιφανειών *Poisson* [64] είναι απαιτούμενα για την παραγωγή συμπαγούς ανακατασκευής της επιφάνειας. Πρόσφατες μέθοδοι χρησιμοποιούν μια συλλογή σκηνών για την εκπαίδευση βαθιών νευρωνικών μοντέλων, σε επιμέρους εργασίες όπως το *pipeline MVS* (μέσω αντιστοίχησης χαρακτηριστικών

όπως γίνεται στην εργασία [65]), ή με μέθοδο σύντηξης βάθους (*depth fusion*) στις αντίστοιχες εργασίες [66], [67]. Κάποιες άλλες εργασίες συμβάλλουν σε μια ολοκληρωμένη γραμμή παραγωγής στερεομετρίας από πολλαπλές όψεις *MVS*[68], [69], [70] και παράγουν αξιόπιστες σαρώσεις *MVS 3D scans*. Πολλές φορές οι παράμετροι της κάμερας βέβαια δεν είναι διαθέσιμες. Σε τέτοιες περιπτώσεις και δεδομένου συνόλου εικόνων από μια συγκεκριμένη σκηνή, υπάρχουν μέθοδοι που τις εξάγουν με *Structure Form Motion* (SFM) με αντίστοιχες μεθοδολογίες [71], [72], [73], [74]. Συγκεκριμένα, εφαρμόζονται για να αναπαράγουν τις κάμερες σε μια αραιή τρισδιάστατη ανακατασκευή εκτιμώντας γραμμικά και τις παραμέτρους των καμερών και μάλιστα η τελευταία χρησιμοποιείται και στην παρούσα εργασία. Τέλος στην εργασία του, ο *Tang Tan* [75], χρησιμοποιεί μια αρχιτεκτονική βαθιού νευρωνικού δικτύου με έναν ολοκληρωμένο διαφορίσματος νευρωνικό επίπεδο *Bundle Adjustment* [76] (Δέσμης Προσαρμογής) για να εξάγει μια γραμμική βάση για το βάθος ενός πλαισίου αναφοράς στον χώρο, και τα χαρακτηριστικά από τις γειτονικές εικόνες, ώστε να βελτιστοποιήσει το βάθος και τις παραμέτρους της κάμερας σε κάθε εμπρόσθιο πέρασμα του δικτύου.

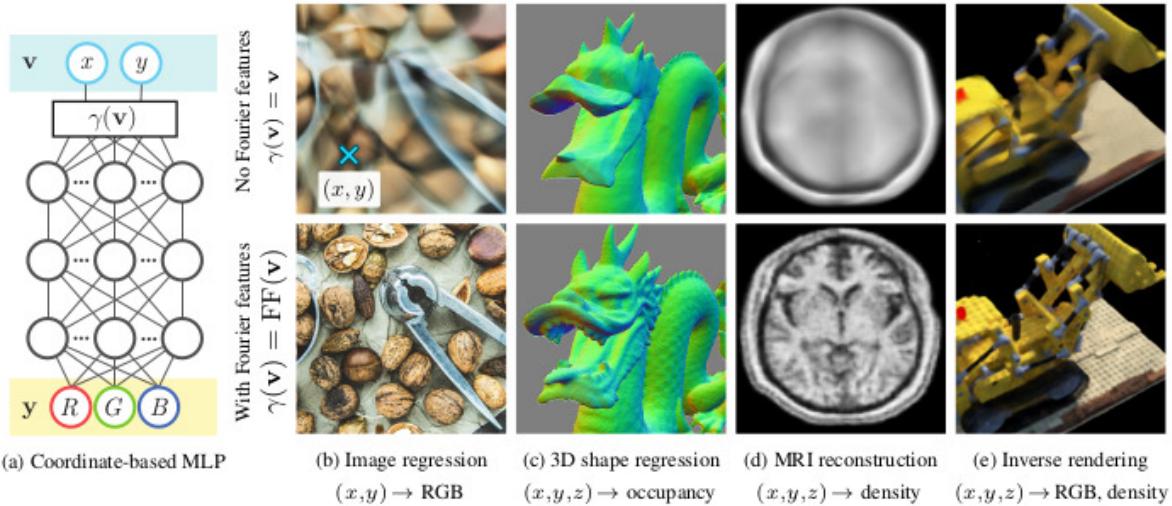
Σε αντίθεση με αυτές τις εργασίες, το *IDR* εκπαιδεύεται με εικόνες από μια μόνο σκηνή-στόχο, παράγοντας μια ακριβή στεγανή τρισδιάστατη ανακατασκευή επιφάνειας.

Όπως ειπώθηκε ήδη, στόχος της συγκεκριμένης εργασίας είναι να επινοήσει ένα ολοκληρωμένο σύστημα όσον αφορά αρχιτεκτονικής βαθιών νευρωνικών δικτύων, το οποίο μπορεί να μάθει τρισδιάστατες γεωμετρίες από *masked* (μάσκα που δείχνει αν καλύπτεται το εξής pixel από το αντικείμενο ενδιαφέροντος) εικόνες 2D και εκτιμήσεις καμερών, ενώ δεν απαιτεί περαιτέρω επίβλεψη. Στο Σχήμα:3.1, απεικονίζεται η συνολική αρχιτεκτονική του δικτύου, η διαφορίσματος συνάρτηση απόδοσης (*differentiable rendering function* με έμφαση στο διαφορίσματος σημείο τομής της ακτίνας με την επιφάνεια και την επιστροφή των κλισεων των κάθετων διανυσμάτων, το οποίο αναλύεται και στην μεθοδολογία της διαφορίσματος εξίσωσης απόδοσης γραφικών βλ.4.3)



Σχήμα 3.1: Συνολικό Δίκτυο IDR, Πηγή [6]

3.3 Let Neural Nets learn high frequency content in low dimensional domain



Σχήμα 3.2: Βελτίωση αποτελεσμάτων σε Coordinate Based MLPs για μια ποικιλία υψηλοσυχνοτικού περιεχομένου εφαρμοσμένο σε χαμηλοσυχνοτική εφαρμογή της ανακατασκευής εικόνων και 3D σχημάτων, Πηγή [24]

Τα δίκτυα που ασχολούνται με την νευρωνική απόδοση όγκου καθώς και δίκτυα ανακατασκευής σημάτων ασχολήθηκαν σε βάθος με την κωδικοποίηση της εισόδου για την καταγραφή του συχνοτικού περιεχομένου που έχει η επιφάνεια και πρέπει να φανεί τόσο στην γεωμετρία της όσο και στις παραμέτρους που αναπαριστούν το χρώμα της κατά την απόδοση της σκηνής.

Αυτή η ιδέα παρουσιάζεται στο NeRF [3], με την μορφή του *Positional Encoding* το οποίο αναλύθηκε και στο 2.5. Ωστόσο, το βασικό θεωρητικό υπόβαθρο αυτής της τεχνικής, η οποία είναι σε θέση να βρει υψηλοσυχνοτικές σχέσεις δεδομένων αντιστοιχώντας τα με ειδικές συναρτήσεις πυρήνα Fourier σε υψηλότερες διοιστάσεις όπου αναπαριστώνται καλύτερα χαρακτηριστικά (*Hilbert Space*), διατυπώθηκε ξεκάθαρα στα πλαίσια της εργασίας *Fourier Feature Networks / Let Neural Networks Learn High frequency content from low dimensional input* [24].

Συγκεκριμένα τα νευρωνικά δίκτυα προσπαθούν να προσεγγίσουν συναρτήσεις ρυθμίζοντας τις παραμέτρους των βαρών τους μέσω αλγορίθμων που βασίζονται στην πτωτική κλίση σφάλματος (*Stochastic Gradient Descent* [22]). Όταν ο στόχος είναι το μοντέλο να γενικεύει καλά στην διαδικασία προσέγγισης, τα δίκτυα είναι καταδικασμένα να ενσωματώσουν πολλές περισσότερες παραμέτρους για να μοντελοποιήσουν συναρτήσεις που δίνουν ακριβείς τιμές. Αυτό σε έναν βαθμό βελτιώνεται όταν προσπαθούμε να προσαρμόσουμε τα βάρη στα δεδομένα μέσω θεωρίας ομαλοποίησης βαρών. Έτσι θεωρείται ότι δίνονται βαθμοί ελευθερίας ώστε να μπορέσει το δίκτυο να δράσει καλύτερα εκτός του συνόλου εκπαίδευσης. Ωστόσο, μη επαρκής ομαλοποίηση οδηγεί το δίκτυο σε προβλήματα υπερεκπαίδευσης το οποίο θεωρείται ότι συνήθως μπορεί να ξεπεραστεί στα πλαίσια της δημιουργίας συνόλου δεδομένων υψηλής εντροπίας, και τεχνικών εκπαίδευσης όπως στρώματα *Dropout* που τυχαία κόβουν υπολογιστικούς κόμβους, ή τεχνικές κανονικοποίησης / ομαλοποίησης που εμφανίζονται στα πλαίσια εκπαίδευσης (πχ. *Batch Normalization*, δηλαδή κανονικοποίηση κατά ομάδες δεδομένων εκπαίδευσης που συνδυάζεται με μεθόδους εκπαίδευσης *mini Batch* όχι *online*).

Παρ' όλα αυτά, στα πλαίσια της εργασίας τα δίκτυα είναι πλήρως συνδεδεμένα MLPs συνταγμένων. Δηλαδή η διάσταση είναι αρκετά μικρή και η τυχαιότητα των δεδομένων δεν είναι προφανής. Συνεπώς στα πλαίσια

των εργασιών που προαναφέρθηκαν γίνεται μια προσπάθεια, αντί να προσεγγιστεί το πεδίο $f(x, y, z; \theta)$ για παράδειγμα, που είναι το έμμεσο πεδίο απόστασης από την επιθυμητή επιφάνεια, να προσεγγιστεί το πεδίο $\gamma(v; \theta)$. Αυτό το πεδίο, έχει κωδικοποιήσει την είσοδο μέσα από έναν εφαπτόμενο πυρήνα ημιτονειδών συναρτήσεων δηλαδή συναρτήσεων που ενσωματώνουν συχνοτικό περιεχόμενο. Το είδος των συναρτήσεων εξασφαλίζει ότι δεν χάνεται το συχνοτικό περιεχόμενο των δεδομένων όταν περνούν από συστήματα που μπορεί να μην μπορούν να το αντιληφθούν επειδή είναι υψηλότερης διάστασης από τα ίδια τα δεδομένα. Τέτοιου είδους πρόβλημα αντιμετωπίζουν από την φύση της εκπαίδευσής τους οι υπολογιστικοί γράφοι των *DNNs*(βαθιών νευρωνικών δικτύων).

NeRF Positional Encoding

Έτσι το NeRF [3], προσπαθεί να καταγράψει υψηλοσυχνοτικές σχέσεις στην απόδοση χρώματος στην ογκομετρική απόδοση χώρου εφαρμόζοντας το *Positional Encoding* στα δεδομένα μια πρακτική που δίνει προ-οπτική στα αποτελέσματα ανακατασκευής του πεδίου ακτινοβολίας αλλά και στην ταχύτητα σύγκλισης σε καλά επίπεδα σφάλματος.

Σφαιρικές Αρμονικές | Spherical Harmonic Encoder

Τέλος η χρήση σφαιρικών αρμονικών είναι μια τεχνική που χρησιμοποιείται αποκλειστικά για την κωδικοποίηση των διανυσμάτων προσανατολισμού της κάμερας, καθώς μοντελοποιούν ανακλαστικές ιδιότητες που έχουν να κάνουν με πεδία ακτινοβολίας ωστόσο αντιστοιχούν σε παρόμοιες μορφές κωδικοποίησης και όλη η προήλθε από την παρακάτω πηγή [25] μιας και οι σφαιρικές αρμονικές είναι κλασσική μέθοδος ανακατασκευής σήματος και προήλθε από έρευνα κβαντομηχανικής πάνω στο σωματίδιο του υδρογόνου H. Έτσι ερευνήθηκε και το μοντέλο κωδικοποίησης σφαιρικών αρμονικών που βασίζεται στον υπολογισμό των παραμέτρων C στην εξίσωση πραγματικών σφαιρικών αρμονικών 2.5.

Σε κάθε περίπτωση η συνένωση HF(υψηλοσυχνοτικών) κωδικοποιημένων εισόδων μαζί με πραγματικών συντεταγμένων, επιτρέπει στα δίκτυα να προσεγγίσουν με πολύ μεγαλύτερη αξιοπιστία συναρτήσεις στο πρόβλημα της παλινδρόμησης συντεταγμένων, με αποτέλεσμα να αποτυπώσουν υψηλοσυχνοτικό περιεχόμενο στις αναπαραστάσεις σε λίγες εποχές εκπαίδευσης με λιγότερες παραμέτρους εκπαίδευσης.

3.4 Instant Neural Graphics Primitives | Multi-Resolution Hash Encoding

Μέσα στο 2022 εμφανίστηκε μια έρευνα που ασχολείται με την κωδικοποίηση δικτύων. Συγκεκριμένα, το *Instant-NGP*(Άμεσα νευρωνικά πρωτότυπα γραφικών), από τους επιστημονικούς συνεργάτες της *Nvidia*, *T.Muller et al.*, που επιτάχυνε την όλη ανακατασκευή έμμεσων επιφανειών δηλαδή την εκπαίδευση των πεδίων αυτών στημαντικά.

Αυτή εργασία βασίζεται πάλι στην κωδικοποίηση της εισόδου, αλλά με τρόπο που δεν αντιστοιχεί τις συντεταγμένες σε χώρο υψηλότερης διάστασης αλλά σε αντίθεση με προηγούμενες εργασίες που παρουσιάστηκαν, θεωρεί πως δεν απαιτείται δομικός μετασχηματισμός των δεδομένων σε κανένα σημείο κατά την διάρκεια της εκπαίδευσης. Συγκεκριμένα δεν πρέπει να γίνεται στόχος η μετατροπή της πληροφορίας με σκοπό το δίκτυο να μάθει πιο εύκολα τις κωδικοποίησεις των δεδομένων. Αντίθετα, να γίνει μια προσαρμόσιμη μορφή του δικτύου ανάλογα την περιοχή που παράγει την κατάλληλη μορφή κωδικοποίησης και να υπάρχει τρόπος διακριτοποίησης των κωδικοποιήσεων από περιοχή σε περιοχή του χώρου γνωστό και ως *Spatial Coordinate Encoding*. Η συγκεκριμένη εργασία καταφέρνει να κωδικοποιήσει τις συντεταγμένες με τέτοιο τρόπο που αποφεύγονται ίδια χαρακτηριστικά *hash* που παράγει με συνάρτηση κατακερματισμού, το οποίο ονομάζεται *hash collision*. Ταυτόχρονα επιτυγχάνει, σε λίγες εποχές να ανακατασκευάσει σύνθετες τρισδιάστατες αναπαραστάσεις με έμμεσα

δίκτυα όπως *Occupancy Fields* ή *Truncated SDF fields* ή ακόμα και πεδία φωτισμού με μονοσκοπική πληροφορία από βίντεο (*intristic-NGP NeRF*) [77]). Η συγκεκριμένη τεχνική χρησιμοποιήθηκε και στην παρούσα εργασία για την χωρική κωδικοποίηση.

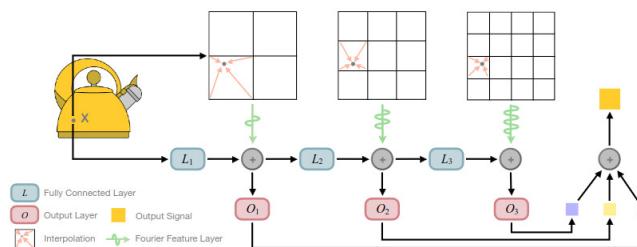
3.5 Neural Fourier Filter Banks

Στο πεδίο της 3D όρασης, (De Queiroz κ.α. [78]) πρότειναν την χρήση προσαρμοστικής περιοχής του ιεραρχικού χωρικού μετασχηματισμού κυματιδίων (*Adaptive region Haar Wavelets*) για την διευκόλυνση συμπίεσης νέφους σημείων 3D. Συγχρόνως οι Isik, κ.α. [79] προτείνουν την εκμάθηση των ογκομετρικών χαρακτηριστικών που προκύπτουν από την προηγούμενη εργασία από δίκτυα συντεταγμένων. Τέλος δίνεται μια προοπτική στην ανακατασκευή αναπαραστάσεων με φίλτρα κυματιδίων(*wavelet filtering*), για συμπαγή πεδία νευρικής ακτινοβολίας στην εργασία [80].

Πριν λίγους μήνες από το εργαστήριο *UBC-Vision*(*University of British Columbia Computer Vision Group*) δημοσιεύτηκε ένα κείμενο που προτείνει έναν συνδυασμό των παραπάνω μεθόδων κωδικοποίησης που μπορεί να εφαρμοστεί και σε δεδομένα εικόνας αλλά σημαντικό είναι πως μπορεί να εφαρμοστεί και σε δεδομένα που αναπαριστούν έμμεσα πεδία γεωμετρίας.

Ο λόγος γίνεται για την έρευνα *Neural Fourier Filter Banks* [1], η οποία βασιζόμενη στην ιδέα συχνοτικής αποσύνθεσης του σήματος με χρήση κυματιδίων (*Wavelet Decomposition*), προτείνει την επικουρική χρήση και των δικτύων που εφαρμόζουν κωδικοποίηση της εισόδου με χρήση εφαπτόμενου πυρήνα Fourier, και των δικτύων που εφαρμόζουν χωρική κωδικοποίηση συντεταγμένων με έναν τρόπο που διατηρούνται τα χαρακτηριστικά και των δύο μορφών κωδικοποίησης.

Η αρχιτεκτονική του δικτύου που προτείνεται φαίνεται στην παρακάτω εικόνα και στην παρούσα εργασία με βάση την παρακάτω εικόνα προσπαθεί να επινοηθεί μια μέθοδος που εφαρμόζει αυτή την διαδικασία στα δεδομένα στοχεύοντας αποτελέσματα υψηλής αξιοπιστίας ανακατασκευής σε μικρότερο αριθμό εποχών εκπαίδευσης του δικτύου απόδοσης IDR [6].



Σχήμα 3.3: Προτεινόμενη αρχιτεκτονική NFFB, Πηγή [1]

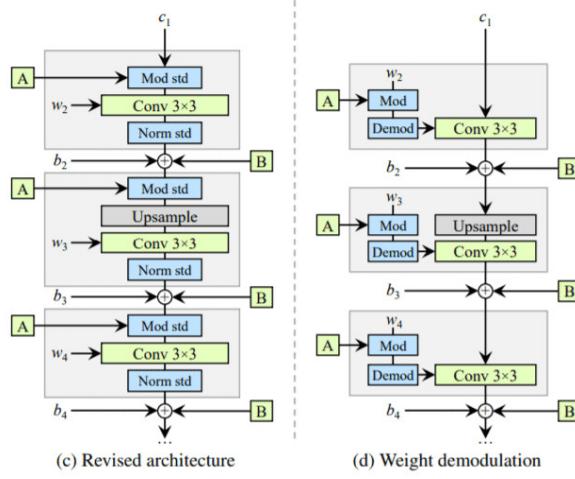
3.6 StyleGAN Modulation / Demodulation, Attention is all you need

Πρόσφατες μέθοδοι εφαρμόζουν μηχανισμό σύγχρονης/ *online* ενσωμάτωσης των βαθιών χαρακτηριστικών που παράγουν τα δίκτυα κωδικοποίησης² στα βάρη του δικτύου. Με αυτό το σκεπτικό αντλήθηκαν πληροφορίες από δίκτυα που εφαρμόζουν διαμόρφωση και αποδιαμόρφωση του χαρακτηριστικών σκοπεύοντας την ομαλοποίηση των δικτύων και την προσθήκη στο επόπεδο γνώσης του *style*/στιλ των δεδομένων. Αυτή η έννοια δεν αναφέρεται κυριολεκτικά αλλά αναφέρεται σε στατιστικές που έχουν τα δεδομένα και στην απόκτηση γνώσης μέσα από αυτές.

Αρχικά παρουσιάστηκε μια έρευνα πάνω στην προσαρμοστική κανονικοποίηση του δικτύου. Αυτή (*Adaptive Instance Normalization* [9]) στοχεύει το ίδιο το δεδομένο εισόδου και όχι με βάση τις ομάδες δεδομένων (*batches*) εκπαίδευσης που περνούν από το δίκτυο μέχρι την επόμενη αναβάθμιση των βαρών. Το τελευταίο είναι κλασσική τακτική, που στην ορολογία των νευρωνικών δικτύων αναφέρεται ως *Batch Normalization*. Αυτού του είδους η προσαρμοστική κανονικοποίηση δείγματος, θεωρεί μη αποδοτική την χρήση *Batch Normalization* καθώς κανονικοποιεί τα δεδομένα με βάση την στατιστική ενός πλήθους δειγμάτων, διαισθητικά ομαλοποιώντας όλο το *batch* γύρω από ένα κεντρικό στυλ που φαίνεται από τις στατιστικές του χαρακτηριστικού. Στην περίπτωση των δικτύων ανακατασκευής που δέχονται ως είσοδο συντεταγμένες, και επομένως σχηματίζουν χαρακτηριστικά διανύσματα συντεταγμένων, κάθε δείγμα δεδομένων είναι σημαντικό ως προς τις στατιστικές του. Συνεπώς ο *AdaIN* αλγόριθμος προσπαθεί να εφαρμόσει το στυλ της εισόδου στα δεδομένα που περνούν από το δίκτυο κάνοντας χρήση ανακλιμάκωσης του πίνακα χαρακτηριστικών (*rescaling feature map*) με βάση ένα διάνυσμα στατιστικών της εισόδου υπολογίζεται (*Style Vector*).

Πάνω σε αυτή την μέθοδο, η ήδη υπάρχουσα έρευνα *StyleGAN* που προτείνει ένα παραγωγικό - αντιπαραθετικό δίκτυο δημιουργίας εικόνων ανθρώπων, έκανε επεκτάσεις (*StyleGAN2* [28]) εισάγοντας τες, στην έννοια του *StyleBlock*. Συγκεκριμένα, έκρινε πως αυτή η διαδικασία που ακολουθεί ο *AdaIN* μπορεί να εφαρμοστεί κάνοντας διαδοχικές αποδιαμορφώσεις και διαμορφώσεις των βαρών των συνελίξεων που παίρνουν μέρος στο *StyleBlock*. Όλο αυτό βασίζεται στην λογική ότι οι πίνακες χαρακτηριστικών αλλάζουν χρησιμοποιώντας συνελίξεις και η διαμόρφωση και αποδιαμόρφωση που εφαρμόζεται με την μορφή κλιμάκωσης και αποκλιμάκωσης βαρών, βοηθά τα δίκτυα να αποδώσουν ορισμένα χαρακτηριστικά των δεδομένων περισσότερο από άλλα με βάση το ποια χαρακτηριστικά ακολουθούν τα δεδομένα. Έτσι, στην αναθεωρημένη δεύτερη εκδοχή το μοντέλο *Stylegan2* [28], μεταφέρει τον μη ακολουθιακό θόρυβο που κατά παράδοση χρειάζονται τα παραγωγικά δίκτυα εκτός του *StyleBlock* όπως φαίνεται και στην παρακάτω εικόνα.

²όπως και τα *DeepSDF embedding* της εργασίας



Σχήμα 3.4: Διαμόρφωση και Αποδιαμόρφωση βαρών στο αναθεωρημένο StyleBlock-StyleGAN2, Πηγή [28]

Τα δίκτυα που χρησιμοποιούνται σε αυτή την εργασία είναι μη συνελικτικά πλήρως συνδεδεμένα *MLPs* μιας και η λογική της ανακατασκευής απαιτεί να μην γίνει μετασχηματισμός του χώρου του συνταγμένων με τρόπο που δεν λαμβάνει υπόψιν την ισομετρική ιδιότητα του χώρου. Συνεπώς εργασίες όπως [8] (*Attention is all you need*) δίνουν μια πιο γενική μορφή στο πως να βοηθήσουμε την έμφαση και την προσοχή που δίνει το δικτύου σε ορισμένα μόνο χαρακτηριστικά, ανεξαρτήτως αρχιτεκτονικής. Σε αυτή την λογική, βασίζονται πλέον όλα τα σύγχρονα μοντέλα που θέλουν να καταγράψουν το στυλ των δεδομένων.

4

Μεθοδολογία

4.1 Δεδομένα Ελέγχου - Προ επεξεργασία

4.1.1 Σύνολο Δεδομένων DTU, *DTU Robot Image Datasets* [81]

Το σύνολο δεδομένων DTU MVS 2014 [81], αποτελεί μια συλλογή εικόνων τρισδιάστατων μοντέλων πολλαπλών όψεων. Το συγκεκριμένο σύνολο δεδομένων περιέχει 80 σκηνές μεγάλης φωτομετρικής ποικιλίας, δηλαδή 80 διαφορετικά μοντέλα που φωτογραφήθηκαν με χρήση ρομποτικού βραχίονα κάτω από ίδιες συνθήκες φωτισμού και τοποθέτησης της κάμερας. Συγκεκριμένα κάθε σκηνή περιέχει είτε 49 είτε 64 διαφορετικές ακριβείς θέσεις καμερών και σαρώσεις δομημένου φωτός, όλες αποκτημένες από ρομποτικό βραχίονα 6 βαθμών ελευθερίας (6-DOF).



Σχήμα 4.1: DTU MVS σύνολο δεδομένων

4.1.2 Εξαγωγή Δεδομένων Τρισδιάστατης Στερεομετρίας σε ένα αντικείμενο Scene Dataset

Για να γίνει εκπαίδευση του δικτύου και να γίνει επίβλεψη του με το σύνολο δεδομένων εικόνων που έχει το DTU γίνεται χρήση μια κλάσης που αποτελεί την δομή δεδομένων που αποθηκεύει σύνολο εικόνων μια σκηνής. Συγκεκριμένα η κλάση *SceneDataset* δέχεται τις εικόνες μια σκηνής (που την δεδομένη στιγμή αναπαριστά το δίκτυο απόδοσης), τις μάσκες των ενδιαφέρον ή ενεργών pixel για τα αντικείμενα της σκηνής. Κάνοντας χρήση του ευρετηρίου της κάμερας σε συνδυασμό έναν πίνακα κλίμακας *scale matrix* μαζί με τις παραμέτρους των καμερών (βλ. Κεφ.7) που στοχεύουν το αντικείμενο δημιουργεί το αντικείμενο της κάμερας όπως περιγράφεται στο παράρτημα. Σε περίπτωση που δεν υπάρχουν δεδομένες θέσεις καμερών, χρειάζεται να οριστούν οι αρχικές θέσεις των καμερών με μια αιτιοκρατική μέθοδο που υπάγεται στην προ επεξεργασία καμερών. Συγκεκριμένα γίνεται χρήση της μεθόδου [74], που είναι σε θέση να κάνει μια καθολική γραμμική αρχικοποίηση αυτών των παραμέτρων αν δεν παρέχονται από το σύνολο δεδομένων. Στην συνέχεια αυτές οι παράμετροι εκπαιδεύονται μιας και αναπαριστούν κάμερες με θόρυβο(για περισσότερα βλ. Κεφ.7). Έτσι το αντικείμενο σε κάποιο δείκτη μπορεί να δώσει κανονικοποιημένες εγγενείς παραμέτρους κάμερας, pixel των πολλαπλών εικόνων της σκηνής που εκτιμάται κατά την φωτογράφιση καθώς και της αντίστοιχη μάσκα. Ταυτόχρονα σε περίπτωση εκπαιδεύσιμων παραμέτρων κάμερας δίνει και την πόζα¹ της κάμερας που εκτιμάται ότι έχει. Παράλληλα το αντικείμενο επιστρέφει και τα pixel της πραγματικής εικόνας, δηλαδή όχι την φωτογραφία της παρούσας σκηνής που αναπαριστά το έμμεσο πεδίο.

Παράλληλα παρέχεται συνάρτηση που συνενώνει όλα αυτά τα δεδομένα σε τανυστές και η συνάρτηση που εκτιμά την αρχική πόζα της κάμερας.

4.2 Δίκτυο Γεωμετρικής Ανακατασκευής και Χρωματικής Απότυπωσης

4.2.1 Συμβολισμοί

Οι συμβολισμοί που συνοδεύουν τις παρακάτω αναλύσεις είναι οι εξής:

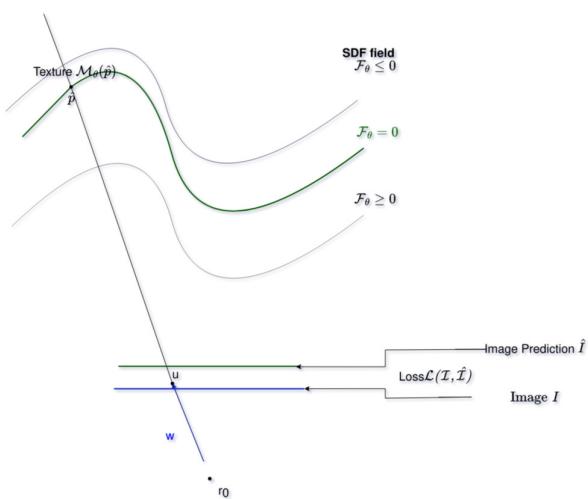
- u - Θέση pixel στην εικόνα
- c - Θέση κάμερας
- \hat{p} - Εκτιμώμενο σημείο στην επιφάνεια
- f_θ - Νευρωνικό πεδίο προσημασμένης απόστασης SDF
- \mathcal{M}_θ - Νευρωνικό πεδίο αποτύπωσης χρώματος στην εικόνα
- $\mathcal{L}_u(\theta, \gamma, \tau)$ - Πεδίο, έμμεσης διαφορίσιμης αποτύπωσης
- I, \hat{I} - Εικόνα και εκτίμηση εικόνας
- S_θ - Αναπαραστώμενη επιφάνεια
- w - Μοναδιαίο διάνυσμα πάνω στην ακτίνα με κατεύθυνση της κατεύθυνση όψης
- $R_u(t)$ - Παραμετρική εξίσωση ακτίνας που διέρχεται από pixel u και έχει παράμετρο t
- θ, γ, τ - Παράμετροι βελτιστοποίησης

¹quaternion πίνακας με το πίνακα στροφής της κάμερας

4.2.2 Διαφορίσιμη νευρωνική αποτύπωση σε βήματα

Προς το παρόν είδαμε, υπό γενική μορφή, πως αναπαρίστανται πολλαπλότητες μέσω νευρωνικών δικτύων αλλά και πως ελέγχονται οι διαφορίσιμες πολλαπλότητες. Στο παρόν κομμάτι παρουσιάζεται αναλυτικά πως αυτά εφαρμόζονται στην γενική περίπτωση ανακατασκευής επιφανειών με νευρωνικά δίκτυα, κάνοντας χρήση εικόνων ως κύριο μέσο επίβλεψης. Κατά το εμπρόσθιο πέρασμα του δικτύου αυτά που εκτελούνται είναι τα εξής:

- Για κάθε ενδιαφέρον pixel u στην μάσκα της εικόνας I που καλύπτει την περιοχή του αντικείμενου
- Βρες το σημείο \hat{p} πάνω στην ακτίνα w εκτελώντας Sphere Tracing με αφετηρία r_0
- Εντός της επιφάνειας $f_\theta \leq 0$, Εκτός της επιφάνειας $f_\theta \geq 0$
- Για δεδομένο σημείο \hat{p} με εκτιμώμενη προσημασμένη απόσταση $f(\hat{p})$
- Υπολόγισε το χρώμα (Texture Field) $\mathcal{M}_\theta(\hat{p})$ στο σημείο \hat{p}
- Τοποθέτησε το χρώμα $\mathcal{M}_\theta(\hat{p})$ στο pixel u



Σχήμα 4.2: Νευρωνική Αποτύπωση - Διαδικασία

Το σημαντικό είναι ότι κατά τον αλγόριθμο *Back Propagation* εκπαιδεύεται και το SDF πεδίο. Αυτό απαιτεί, για δεδομένο φωτομετρικό σφάλμα έστω η νόρμα L_1 και για όλα τα ενδιαφέροντα pixel της εικόνας, ο αλγόριθμος βελτιστοποίησης να ανανεώνει **και τις παραμέτρους θ** (υπολογίζοντας τις κλίσεις του σφάλματος και αναβαθμίζοντας τα βάρη στο δίκτυο f). Έτσι:

- Για κάθε pixel παρατήρησης u . που καλύπτεται από το αντικείμενο στην μάσκα της εικόνας I και εκτιμώμενο χρώμα αυτής \hat{I}
- Ορίζεται το φωτομετρικό σφάλμα ως το L_1 Loss $\mathcal{L}(I, \hat{I}) = \sum_u |\hat{I}_u - I|$
- Παραγώγιση της συνάρτησης κόστους επιστρέφοντας κλίσεις στα δίκτυα f , \mathcal{M}

- Με χρήση κανόνα αλυσίδας

$$\frac{\partial \mathcal{L}}{\partial \theta} = \sum_{\mathbf{u}} \frac{\partial \mathcal{L}}{\partial \hat{\mathbf{I}}_{\mathbf{u}}} \cdot \frac{\partial \hat{\mathbf{I}}_{\mathbf{u}}}{\partial \theta}$$

$$\frac{\partial \hat{\mathbf{I}}_{\mathbf{u}}}{\partial \theta} = \frac{\partial \mathbf{M}_{\theta}(\hat{\mathbf{p}})}{\partial \theta} + \frac{\partial \mathbf{M}_{\theta}(\hat{\mathbf{p}})}{\partial \hat{\mathbf{p}}} \cdot \frac{\partial \hat{\mathbf{p}}}{\partial \theta}$$

- Έμμεση διαφόριση ως προς τις παραγώγους γεωμετρίας βελτιώνοντας το SDF πεδίο. Έτσι η διαφόριση της $f_{\theta}(\hat{\mathbf{p}}) = 0$ επιστρέφει:

$$\frac{\partial \hat{\mathbf{p}}}{\partial \theta} = -\mathbf{w} \left(\frac{\partial f_{\theta}(\hat{\mathbf{p}})}{\partial \hat{\mathbf{p}}} \cdot \mathbf{w} \right)^{-1} \frac{\partial f_{\theta}(\hat{\mathbf{p}})}{\partial \theta}$$

Μεθοδολογία IDR - Επέκταση Νευρωνικής Αποτύπωσης [6]

Στόχος του δικτύου της έμμεσης διαφορίσιμης αποτύπωσης και στο IDR είναι να ανακατασκευάσει την γεωμετρία ενός αντικειμένου από μάσκες 2D εικόνων (των pixel που καλύπτουν το ενδιαφέρον αντικείμενο) οι οποίες πιθανόν μπορεί να έχουν θόρυβο είτε από μόνες τους ή λόγω θορύβου στις πληροφορίες της κάμερας. Ταυτόχρονα κάνει χρήση και του μοναδιαίου διανύσματος όψης \mathbf{w} στην εξίσωση αποτύπωσης.

Πιο συγκεκριμένα οι άγνωστες παράμετροι που μαθαίνονται στα δίκτυα είναι οι εξής:

- γεωμετρία ισομετρικής επιφάνειας, η οποία αναπαριστάται ως $\theta \in \mathbb{R}^m$
- εμφάνιση ισομετρικής επιφάνειας, η οποία αναπαριστάται ως $\gamma \in \mathbb{R}^n$
- παράμετροι κάμερας, οι οποίες αναπαρίστανται ως $\tau \in \mathbb{R}^k$

Όπως αναλύθηκε και στο κομμάτι του νευρωνικού ελέγχου διαφορίσιμων πολλαπλοτήτων η γεωμετρία αναπαριστάται ως ένα σύνολο μηδενικού επιπέδου ενός MLP νευρωνικού f ,

$$S_{\theta} = \{\mathbf{p} \in \mathbb{R}^3 | f(\mathbf{p}; \theta) = 0\} \quad (4.1)$$

Το αξιοσημείωτο με το παρόν μοντέλο είναι ότι διαχωρίζει το δίκτυο απόδοσης χρωμάτων M της σκηνής από την γεωμετρία f , υπό την έννοια ότι αποτελούν διαφορετικά επιμέρους δίκτυα ενός συνολικού μοντέλου ανακατασκευής σκηνής. Αυτό σε συνδυασμό με την απλότητα του μοντέλου το καθιστούν πολύ καλή επιλογή ως δίκτυο βάσης μιας και δεν καταφέρνει το ίδιο να αποδώσει υψηλοσυχνοτικό περιεχόμενο.

Εμπρόσθιο Δίκτυο IDR - IDR Forward Resnet

Το IDR αποτελεί σημαντική επέκταση της απλής νευρωνικής απόδοσης στοχεύοντας να εκπαιδεύσει περισσότερες παραμέτρους κατά την διάρκεια της εκπαίδευσης, εκτελώντας λίγο διαφορετικό τρόπο υπολογισμού των σημείων και εισάγοντας διαφορίσιμα σημεία επαφής με τις επιφάνειες. Οι παρακάτω αναλύσεις αποτελούν ερμηνεία της ανάλυσης της λειτουργίας του IDR [6].

Έστω εικονοστοιχείο (pixel), με δείκτη \mathbf{u} , το οποίο συνδέεται με κάποια εικόνα εισόδου I_u . Έστω τώρα, $R_u(t) = \{c_u + t w_u \mid t \geq 0\}$, το οποίο συμβολίζει την διερχόμενη ακτίνα από το εικονοστοιχείο της εικόνας όψεως \mathbf{u} , όπου $c_u = c_u(\tau)$ (αντίστοιχο του r_0 στο σχήμα 4.2) συμβολίζει το (μη απαραίτητα γνωστό) κέντρο της αντίστοιχης κάμερας κάμερας (της εικόνας προβολής που έχει το εικονοστοιχείο \mathbf{u}) με $\mathbf{w}_u = \mathbf{w}_u(\tau)$ να συμβολίζει της κατεύθυνση της ακτίνας (δηλαδή, το διάνυσμα που στοχεύει από την θέση της κάμερα c_u προς το εικονοστοιχείο \mathbf{u}).

Ας ορίσουμε, $\hat{p}_u = \hat{p}_u(\theta, \tau)$ την πρώτη διασταύρωση της ακτίνας R_u και της επιφάνειας S_{θ} . Η ακτινοβολία κατά μήκος της ακτίνας R_u , η οποία καθορίζει το προς αποτύπωση χρώμα (texture) στο εικονοστοιχείο \mathbf{u} , $\mathcal{M}_u = L_u(\theta, \gamma, \tau)$, να είναι μία συνάρτηση των ιδιοτήτων της επιφάνειας στο \hat{p}_u , της επερχόμενης ακτινοβολίας στο \hat{p}_u , και της κατεύθυνσης θέασης w_u . Στο IDR γίνεται η υπόθεση ότι η ιδιαιτερότητά/ιδιότητα ή υλικό της

επιφάνειας και η ερχόμενη ακτινοβολία είναι συναρτήσεις του σημείου \hat{x}_p και του αντίστοιχου στο σημείο κανονικού ανύσματος $\hat{n}_u = \hat{n}_u(\theta)$, της κατεύθυνσης όψης \mathbf{w}_u και **ενός διανύσματος γεωμετρίας** $\hat{z}_u = \hat{z}_u(\hat{p}_u; \theta)$. Το IDR πρόσθιο μοντέλο MLP λοιπόν αναπαριστάται μαθηματικά με την παρακάτω μορφή:

$$L_p(\theta, \gamma, \tau) = M(\hat{p}_u, \hat{n}_u, \hat{z}_u, \mathbf{w}_u; \gamma), \quad (4.2)$$

όπου M , είναι το δεύτερο νευρωνικό δίκτυο MLP εκτίμησης χρώματος των σημείων της επιφάνειας. Γίνεται χρήση της συνάρτησης απεικόνισης χρώματος L_u στο εικονοστοιχείο p για τον υπολογισμό του σφάλματος στο χρώμα με την τιμή του χρώματος στο εικονοστοιχείο της εικόνας εισόδου I_u έτσι ώστε απευθείας να εκπαιδεύσουμε το δίκτυο πάνω στις παραμέτρους θ, γ, τ , οι οποίες είναι ουσιαστικά τα βάρη των πολυστρωματικών δικτύων.

4.2.3 Διαφορική μορφή διασταύρωσης ακτίνας με σημείο της γεωμετρίας

Έστω λοιπόν, ένα σταθερό εικονοστοιχείο u , και αφαιρείται ο δείκτης u από τους συμβολισμούς, χωρίς να αφαιρείται η γενικότητα των παραδοχών.

Πρώτο βήμα είναι η αναπαράσταση του διαφορίσμου σημείου τομής $\hat{p}(\theta, \tau)$, ως ένα νευρωνικό δίκτυο με παραμέτρους θ, τ . Αυτό μπορεί να γίνει με μικρή παραμετροποίηση του δικτύου γεωμετρίας Implicit Differentiable Network δηλαδή της συνάρτησης προσημασμένης απόστασης f .

Έστω $\hat{p}(\theta, \tau) = c + \tau(\theta, c, \mathbf{w}) \cdot \mathbf{w}$, ορίζεται το σημείο τομής της ακτίνας με την επιφάνεια. Εφόσον επιδιώκεται να γίνει χρήση του \hat{p} σε έναν αλγόριθμο που μοιάζει με gradient descent όπως αυτός της οπισθοδιάδοσης της παραγώγου του σφάλματος 2.4, αυτό που απαιτείται είναι να εξασφαλιστεί ότι οι παράγωγοι είναι σωστές σε αριθμητική τιμή και έχουν πρώτες παραγώγους στις συγκεκριμένες παραμέτρους, οι οποίες συμβολίζονται ως θ_0, τ_0 . Αντίστοιχα συμβολίζουμε ως $c_0 = c(\tau_0)$ και $\mathbf{w}_0 = \mathbf{w}(\tau_0)$, $\tau_0 = \tau(\theta_0, c_0, \mathbf{w}_0)$, την αρχική θέση κάμερας, την κατεύθυνση κάμερα και αρχική τιμή χρόνου με σημείο τομής (πρώτης παραγώγου) το $p_0 = \hat{p}(\theta_0, \tau_0) = c_0 + \tau_0 \mathbf{w}_0$

Λήμμα 4.2.1. Έστω S_θ ορίζεται ως η εξίσωση της επιφάνειας 4.1 (δηλαδή το σύνολο μηδενικών την έξοδο της f). Το διαφορίσμο σημείο τομής της ακτίνας $R(\tau)$ και της επιφάνειας S_θ μπορεί να αναπαρασταθεί με την παρακάτω εξίσωση:

$$\hat{p}(\theta, \tau) = \vec{c} + \tau_0 * \vec{\mathbf{w}} - \frac{\vec{\mathbf{w}}}{\nabla p f(p_0; \theta_0) \cdot \mathbf{w}_0} \cdot f(\vec{c} + \tau_0 \vec{\mathbf{w}}; \theta) \quad (4.3)$$

και αυτό το σημείο έχει ακριβή αριθμητική τιμή και πρώτη παραγώγος ως προς το θ και το τ στο $\theta = \theta_0, \tau = \tau_0$.

Για να αποδείξουμε αυτή την συναρτησιακή εξάρτηση του \hat{x} στις παραμέτρους, γίνεται χρήση έμμεσης διαφορίστης με το δίκτυο σε συνδυασμό με το δίκτυο δειγματοληψίας που παρουσιάζεται παρακάτω 4.2.3, έτσι, διαφορίζοντας την εξίσωση $\mathbf{f}(\hat{x}; \theta) \equiv 0$ ως προς $\mathbf{u}, \mathbf{c}, \theta$ γίνεται η επίλυση της για της παραγώγους ως προς \mathbf{t} . Στην συνέχεια, μπορεί να ελεγχθεί πως η εξίσωση παραπάνω έχει της σωστές παραγώγους.

Στρώμα Δειγματοληψίας Διαφορίσμων Σημείων - Sample Network

Για δεδομένη ισομετρική επιφάνεια που αναπαριστάται από σύνολο μηδενικού επιπέδου του MLP έμμεσης γεωμετρίας, 4.1, γίνεται χρήση του δικτύου δειγματοληψίας ως μέσο ελέγχου. Η διαδικασία της δειγματοληψίας συνοψίζεται στις εξής διαδικασίες:

- i) Δειγματοληψία n σημείων πάνω στην ισομετρική επιφάνεια τέτοια ώστε: $p_i \in S_\theta, i \in [n]$,
- ii) Δημιουργία του δικτύου δειγματοληψίας $p_i(\theta), i \in [n]$ κάνοντας χρήση ενός γραμμικού νευρωνικού στρώματος στο δίκτυο της έμμεσης αναπαράστασης της γεωμετρίας της ισομετρικής επιφάνειας $f(x; \theta)$ και

- iii) ενσωμάτωση μιας συνάρτησης σφάλματος στο δίκτυο δειγματοληψίας το οποίο χρησιμοποιείται ως διαμεσολαβητής (στρώμα ελέγχου πριν και μετά) στην ισομετρική επιφάνεια που αναπαρίσταται έμμεσα από την f .

Δειγματοληψία Σημείων Διαφορίσιμης Πολλαπλότητας - Level Set Sampling[17]

Το δίκτυο δειγματοληψίας αναπαρίσταται συνολικά από την εξίσωση διαφορίσιμων σημείων τομής και διανυσμάτων προβολής στην εξ. 4.3. Πώς φτάνουμε σε αυτή την εξίσωση όμως;

Συμβολίζεται $D_x f(p; \theta) \in \mathbb{R}^{l \times d}$ ο πίνακας μερικών παραγώγων του f ως προς p . Υποτίθεται συγκεκριμένες παράμετροι γεωμετρίας θ (έστω οι παράμετροι αρχικοποίησης στην σφαίρα).

Ο γενικευμένος αλγόριθμος Newton [82] για εύρεση λύσης της 4.1 μπορεί να αναπαρασταθεί ως εξής:

$$p_{\text{next}} = p - D_x f(p; \theta_0)^{\dagger} f(p; \theta_0), \quad (4.4)$$

όπου $D_x f(p; \theta_0)^{\dagger} f$ είναι ο ψευδό αντίστροφος.

Στην περίπτωσή μας τα δειγματοληπτημένα σημεία ανήκουν στο δίκτυο $f(p; \theta_0)$. Η εξάρτηση του δείγματος p από τις παραμέτρους γεωμετρίας θ ορίζεται με τον συμβολισμό $p(\theta_0)$, ο οποίος εισάγεται και στο πεδίο $f(p(\theta); \theta) = c$. Αυτό εξασφαλίζει ότι αναπαριστώνται σημεία που ανήκουν σε γειτονική πολλαπλότητα για κάποιο κάτω φράγμα c το οποίο θεωρείται όριο γειτονίας όσο οι παράμετροι θ αλλάζουν. Στόχος είναι λοιπόν η εκπαίδευση των παραμέτρων θ ώστε να προσεγγίζεται η επιθυμητή γεωμετρία με αυτά τα σημεία. Εκεί βασίζεται και η έμμεση διαφόριση για την εκπαίδευση γεωμετρικών παραμέτρων και έχει ουσία όταν υπάρχει αριθμητική τιμή των σημείων δειγματοληψίας και της πρώτης παραγώγου του πεδίου ως προς θ . Δηλαδή:

$$p(\theta_0) = p \quad ; \quad \left. \frac{\partial}{\partial \theta} \right|_{\theta=\theta_0} f(p(\theta); \theta) = 0 \quad (4.5)$$

Χρησιμοποιώντας και τον κανόνα αλυσίδας στον πίνακα μερικών παραγώγων του πεδίου προσημασμένης απόστασης:

$$D_x f(p, \theta_0) D_\theta p(\theta_0) + D_\theta f(p, \theta_0) = 0. \quad (4.6)$$

Λύνοντας το σύστημα με των γραμμικών εξισώσεων η λύση αναπαρίσταται ως εξής με τον ψευδο-αντίστροφο της γενικευμένης μορφής Newton:

$$D_\theta p(\theta_0) = -D_x f(p, \theta_0)^{\dagger} D_\theta f(p, \theta_0) \quad (4.7)$$

Τελικά το δίκτυο δειγματοληψίας παράγει μια λύση σημείων πιο ευέλικτη υπολογιστικά:

$$p(\theta) = p - D_x f(p; \theta_0)^{\dagger} [f(p; \theta) - c], \quad (4.8)$$

Το οποίο αντιστοιχεί σε ένα δίκτυο δειγματοληψίας διαφορίσιμων σημείων

$$G\left(p, D_x F(p; \theta_0)^{\dagger}; \theta\right) := p(\theta) \quad (4.9)$$

οπού συλλογή σημείων αποτελεί την είσοδο του δικτύου. Αυτή η αναπαράσταση επιτρέπει την επιστροφή κλίσεων του πεδίου f όπως αναπαρίσταται στην διαφορίσιμη εξίσωση σημείων και 4.3.

Τελευταίο αλλά εξίσου σημαντικό, είναι ότι αυτά τα σημεία δεν γνωρίζουμε αν ανήκουν μαθηματικά στο πεδίο προσημασμένης απόστασης. Επομένως θα πρέπει να ισχύει η εικονική εξίσωση για της οποίας η παραγώγος εισάγεται στο δίκτυο και ελέγχεται η δεύτερη νόρμα L_2 :

$$((\|\nabla_{\mathbf{x}} f(\mathbf{x}; \theta)\| - 1)^2 = 0$$

Το δίκτυο γεωμετρίας f το οποίο αναπαριστά διαφορίσιμη πολλαπλότητα μέσω έμμεσης συνάρτησης προσημασμένης απόστασης, είναι αρχικοποιημένο σε σφαίρα λόγω έμμεσης γεωμετρικής ομαλοποίησης των βαρών του (*Implicit Geometric Regularization* [23]). Η ομαλοποίηση αυτή των βαρών, είναι απαραίτητη στην έμμεση διαφόριση της γεωμετρίας καθώς αυτή εξυπηρετείται από τα στρώματα δικτύων δειγματοληψίας που χρησιμοποιούνται ως μέθοδος ελέγχου της ισομετρικής απεικόνισης και παρουσιάζονται στο 4.2.3. Η διαδικασία αυτή είναι κομμάτι της έρευνας [17].

4.2.4 Υπολογισμός Πεδίου Ακτινοβολίας - Rendering Network

Το IDR [6] δίνει ιδιαίτερη προσοχή στην εκτίμηση του πεδίου ακτινοβολίας μια και εκπαιδεύεται σε ένα σύνολο δεδομένων που έχει πολύ σύνθετα στοιχεία υλικών και ο φωτισμός είναι σύνθετος. Συνεπώς πρέπει τα πεδία να καταφέρουν να συλλάβουν πολύ καλά όλους τους τύπους φωτισμού και ειδικά φωτισμούς που παράγει το μοντέλο ανάκλασης *Phong* (βλ. 4). Η συνάρτηση που το εκτελεί αυτό είναι μια *BRDF* η οποία καταφέρνει σε ένα περιορισμένο μοντέλο από τα κανονικά διανύσματα της επιφάνειας ($\mathcal{M}(\hat{p}_u, \hat{n}_u, \hat{z}_u, \mathbf{w}_u; \gamma)$) να υπολογίσει με τρόπο που γενικεύει καλά αντιπροσωπεύοντας ένα συνολικό μοντέλο φωτισμού.

Αναγκαιότητα Υπολογισμού των κανονικών διανυσμάτων - Material Training (BRDF)

Ο υπολογισμός του κάθετου διανύσματος πάνω στην επιφάνεια που δημιουργείται με βάση και την παραπάνω εξίσωση διαφορίσμων σημείων στην επιφάνεια 4.3, δίνεται ως προς τις παραμέτρους της κάμερας και της γεωμετρίας με τον τύπο:

$$\hat{\mathbf{n}}(\theta, \tau) = \frac{\nabla_p f(\hat{p}(\theta, \tau), \theta)}{\|\nabla_p f(\hat{p}(\theta, \tau), \theta)\|_2} \quad (4.10)$$

το οποίο προκύπτει από την θεωρία του λογισμού πάνω σε πεδία συναρτήσεων και αναφέρεται για συγκεκριμένο σημείο p και για διαφορίσμα σημείο \hat{p} . Οι κλίσεις των κανονικών διανυσμάτων επιτρέπουν επιτρέπουν την προσαρμογή των παραμέτρων κάμερας και εμφάνισης.

Το συνολικό μοντέλο ανακλώμενης ακτινοβολίας που δίνει η BRDF μαζί με την ακτινοβολία από τις πηγές σε διαφορίσμα σημείο επιφάνειας \hat{p} (αναλύθηκε στο θεωρητικό υπόβαθρο) προκύπτει από τον παρακάτω τύπο και αντιστοιχείται στο νευρωνικό δίκτυο αποτύπωσης χρώματος *Rendering Network* για δεδομένο διάνυσμα ανακλώμενης ακτινοβολίας \mathbf{w}^0 :

$$L(\hat{p}, \mathbf{w}^0) = L^e(\hat{p}, \mathbf{w}^0) + \int_{\Omega} B(\hat{p}, \hat{n}, \mathbf{w}^i, \mathbf{w}^o) L^i(\hat{p}, \mathbf{w}^i) (\hat{n} \cdot \mathbf{w}^i) d\mathbf{w}^i = \mathcal{M}_0(\hat{p}, \hat{n}, \hat{w}) \quad (4.11)$$

Η συνάρτηση \mathcal{M}_0 στο IDR εκπαιδεύεται ως προς παραμέτρους γ με τέτοιο τρόπο ώστε να παριστάνει ένα συνεχές πεδίο φωτισμού αντικαθιστώντας την μορφή της με την παρακάτω:

$$L(\theta, \gamma, \tau) = \mathcal{M}(\hat{p}, \hat{n}, \mathbf{w}; \gamma) \quad (4.12)$$

Με αυτόν τον τρόπο δουλεύει η έμμεση διαφόριση και η αντίστροφη αποτύπωση γυρνώντας πίσω κλίσεις (*back propagate gradients*) που εκπαιδεύουν και το δίκτυο γεωμετρίας $f(p; \theta)$.

Γιατί είναι απαραίτητα τα κανονικά διανύσματα στην επιφάνεια; Εφόσον το δίκτυο που υπολογίζει το πεδίο φωτισμού είναι πλήρες κάτι το οποίο στην βιβλιογραφία δίνεται ως *P-universal*, πρέπει να αποτυπώνει όλα τα μοντέλα ανακλάσεων στο χρώμα. Κάποια μοντέλα όπως το μοντέλο κατοπτρικής ανάκλασης επηρεάζονται κυρίως από την κλίση του διανύσματος που επιστρέφει στην κάμερα. Μόνο με τις παραμέτρους γεωμετρίας και την παράμετρο του διανύσματος όψης να εκπαιδεύονται δεν μπορεί να βρεθεί το διάνυσμα ανάκλασης από τον νόμο ανάκλασης επομένως οι παράμετροι εμφάνισης εκπαιδεύονται μόνο ως προς μοντέλα φωτισμού που δεν λαμβάνουν υπόψη το πόσο τραχιά είναι επιφάνεια δηλαδή η υφή. Συνεπώς, για να μπορεί να αποδοθεί στην πληροφορία του χρώματος η πληροφορία του υλικού πρέπει η συνάρτηση απόδοσης να πάρει τα διανύσματα προβολής και το κανονικό διάνυσμα της επιφάνειας σαν είσοδο. Ταυτόχρονα για μεγαλύτερη αξιοπιστία πρέπει να κωδικοποιηθούν κατάλληλα έστω ένα από αυτά τα διανύσματα. Ξεχωρίζοντας τα δύο δίκτυα γεωμετρίας και φωτισμού ως προς τις παραμέτρους και δημιουργώντας ένα *P-Universal* πεδίο φωτισμού, εφικτό να μεταφερθεί το χρώμα μιας σκηνής στην γεωμετρία της άλλης.

Μοντέλο Ιχνηλάτισης Ακτίνας - Ray Tracer

Όσον αφορά την διαδικασία ιχνηλάτισης της ακτίνας μέσα στο πεδίο *SDF*, ακολουθείται όπως αναφέρθηκε και στο θεωρητικό υπόβαθρο η τεχνική του *Sphere Tracing*.

Σύμφωνα με αυτή την τεχνική, δεδομένης θέσης κάμερας c και διανύσματος κατεύθυνσης $\mathbf{w} \in \mathcal{S}$ που μαζί χαρακτηρίζουν μια θέση pixel με βάση το πρότυπο της κάμερας 2.7, 2, 3 εφαρμόζεται ο παραμετροποιημένος αλγόριθμος *Ray Marching* που ονομάζεται *Sphere Tracing*, πάνω στην ακτίνα $\mathbf{r} = \{\mathbf{c} + t\mathbf{w}\}$. Ο αλγόριθμος είναι αρχικοποιημένος στο p_0 το οποίο ταυτίζεται με το κέντρο της κάμερας, δηλαδή $p_0 = c$. Για την επιτάχυνση του αλγορίθμου, θεωρείται το p_0 ως το πρώτο σημείο τομής της εξίσωσης της ακτίνας με την περιβάλλουσα σφαίρα. Εφόσον η S_θ , δίνεται μέσω έμμεσου πεδίου προσημασμένης απόστασης SDF, f , ο αλγόριθμος προσεγγίζει την επιφάνεια σε βήματα κάνοντας χρήση μιας παραμέτρου σύγκλισης $\epsilon > 0$ (αρχική τιμή $\epsilon = 1e - 5$) που είναι το κάτω όριο σύγκλισης, και $r > 0$ (αρχική τιμή $r = 1$) που είναι η ακτίνα της περιβάλλουσας σφαίρας την οποία δίνει το δίκτυο μη εκπαιδευμένο λόγω γεωμετρικής κανονικοποίησης των βαρών [23]). Επιπλέον τίθεται και ο παράγοντας α ο οποίος χρησιμοποιείται στην εκπαίδευση και αποτελεί ένα όριο για συντηρητικές εκτιμήσεις πάνω στην μάσκα ενεργών pixel της εικόνας (*masked rendering*).

Χρήση Μάσκας Καθορισμού Ενεργών Pixel - Μονοχρωματική Μάσκα

Μια χρήση μορφή δισδιάστατης επίβλεψης που χρησιμοποιείται από το δίκτυο για την ανακατασκευή γεωμετριας 3D σκηνής είναι οι μονοχρωματικές μάσκες. Οι μάσκες αποτυπώνουν σε ένα κανάλι με δυαδική μορφή τα σημεία της εικόνας των οποίων τα pixel u πληρούνται από το αντικείμενο αναπαράστασης. Οι μάσκες μπορούν να αποτελούν μέρος του συνόλου δεδομένων (*dataset*), η να εκτιμηθούν με κλασικού αλγορίθμους εκτίμησης περιοχών κάλυψης ή τμηματοποίησης εικόνων. Μια συνάρτηση που περιγράφει αν ένα pixel καταλαμβάνεται από το αντικείμενο που αποτυπώνεται για συγκεκριμένη ακτίνα που διέρχεται από το pixel είναι η εξής:

$$S(\theta, \tau) = \begin{cases} 1 & R(\tau) \cap S_\theta \neq \emptyset \\ 0 & \text{αλλιώς} \end{cases}$$

Επειδή δεν είναι διαφορίσιμη μορφή αυτή, η απόδοση του βάθους γίνεται με διαφορίσιμη προσέγγιση της μάσκας με την εξίσωση:

$$S_\alpha(\theta, \tau) = \text{sigmoid}\left(-\alpha \min_{t \geq 0} f(c + t\mathbf{w}; \theta)\right), \quad (4.13)$$

η παράμετρος $\alpha > 0$ είναι θετική. Επειδή, λόγω σύμβασης $f < 0$ εντός της επιφάνειας και $f > 0$ εκτός, επιβεβαιώνεται πως όσο το a τείνει στο άπειρο η μορφή μάσκας τείνει στην δυαδική μάσκα. Η διαφόριση ως προς c, \mathbf{w} αυτής της μορφής μάσκας βασίζεται στο θεώρημα της περιβάλλουσας συνάρτησης (*Envelope Theorem*) εφόσον οι παράμετροι της κάμερας και του διανύσματος προβολής είναι συγκεκριμένες τιμές και υπάρχουν η πρώτοι παράγωγοι τους. Αναφορικά $\frac{\partial c}{\partial t} \min_{t \geq 0} \{f(c + t\mathbf{w}; \theta)\} = \frac{\partial c}{\partial t^*} f(c + t^*\mathbf{w}; \theta)$.

4.2.5 Διαδικασία Εκπαίδευσης

Συνολική συνάρτηση κόστους - Loss Function

Στην περίπτωσή μας, η επίβλεψη των δικτύων γίνεται μέσω 2D εικόνων. Εστω πολυκάναλη εικόνα $I_u \in [0, 1]^3$, και με δυαδική μάσκα ενεργών pixel της εικόνας $\mathcal{O}_u \in [0, 1]$ που αντιστοιχεί για δεδομένο εικονοστοιχείο u σε εικόνα που λήφθηκε από κάμερα $c_u(\tau)$ και κατεύθυνση ακτίνας $\mathbf{w}_u(\tau)$ όπου τα $u \in P := \{UV/\text{καμβάς συνόλου εικόνων}\}$ και $\tau \in \mathbb{R}^k$ να αναπαριστά τις παραμέτρους της κάμερας στη σκηνή. Το συνολικό σφάλμα κατηγοριοποιείται σε τρεις κατηγορίες, μιας και το εγχείρημα απαιτεί και φωτομετρική και γεωμετρική ακρίβεια των πεδίων. Συγκεκριμένα απαρτίζεται στο φωτομετρικό σφάλμα $Loss_{RGB}$ μεταξύ των εικόνων, στο σφάλμα μονοχρωματικής μάσκας αποτύπωσης βάθους $Loss_{Mask}$, και στο σφάλμα της εικονικής εξίσωσης *eikonal equation* που χρησιμοποιείται ώστε να διασφαλίσει ότι το δίκτυο δειγματοληψίας επιλέγει σημεία που ανήκουν σε δίκτυο SDF. Συνολικά, η συνάρτηση έχει την παρακάτω μορφή:

$$\text{loss}(\theta, \gamma, \tau) = \text{loss}_{\text{RGB}}(\theta, \gamma, \tau) + \rho \text{loss}_{\text{MASK}}(\theta, \tau) + \lambda \text{loss}_{\text{E}}(\theta), \quad (4.14)$$

όπου ρ, λ βάρη που εφαρμόζονται στα σφάλματα. Η εκπαίδευση γίνεται σε ομάδες των pixel P (πχ.10000) πολλών εικόνων (*mini-batches*) για αυτό στην παρακάτω ανάλυση χρησιμοποιείται ο συμβολισμός P για όλο το πλήθος των pixel της διερχόμενης από το δίκτυο ομάδας pixel.

Φωτομετρικό σφάλμα $Loss_{RGB}$

Για κάθε pixel $u \in P$ που αποτελεί είσοδο του δικτύου, χρησιμοποιώντας τον αλγόριθμο *sphere-tracing* αποκτάται το σημείο τομής της επιφάνειας S_θ με την ακτίνα. Έστω ότι το $P^{in} \subset P$, υποσύνολο pixels u που βρίσκουν την επιφάνεια και είναι ενεργά pixel $\mathcal{O}_u = 1$. Το χρώμα που ακτινοβολεί η επιφάνεια υπολογίζεται από το δίκτυο *Rendering*, έστω $L_u(\theta, \gamma, \tau) = \mathcal{M}(\hat{p}_u, \hat{n}_u, \hat{z}_u, \mathbf{w}_u; \gamma)$ όπου τα \hat{p}_u, \hat{n}_u ορίζονται από το πεδίο γεωμετρίας 4.3, και πεδίο φωτισμού 4.12 και το \hat{z}_u είναι οι παράμετροι χαρακτηριστικών. Το σφάλμα του δικτύου ως προς το χρώμα υπολογίζεται σε συνάρτηση όλων αυτών με την L_1 νόρμα της απόλυτης τιμής διαφοράς στις τιμές του χρώματος και εκτίμησης και πραγματικής εικόνας. Δηλαδή:

$$loss_{RGB}(\theta, \gamma, \tau) = \frac{1}{|P|} \sum_{u \in P^{in}} |I_u - L_u(\theta, \gamma, \tau)| \quad (4.15)$$

Σφάλμα Μάσκας

Ταυτόχρονα υπολογίζεται το σφάλμα μάσκας η οποία εφόσον προσεγγίζεται δίνει πόσα σημεία της εισόδου από την ομάδα pixel δεν απέχουν ως προς την γεωμετρία του πεδίου που δημιουργείται από την επιφάνεια και τις προσπίπτουσες ακτίνες. Έστω $P^{out} = \frac{P}{P^{in}}$, ορίζει του δείκτες των pixel από τα οποία δεν περνά ακτίνα που να πέφτει στην επιφάνεια $\mathcal{O}_u = 0$. Το σφάλμα μάσκας δίνεται από την παρακάτω συνάρτηση κόστους:

$$loss_{MASK}(\theta, \tau) = \frac{1}{\alpha |P|} \sum_{u \in P^{out}} CE(\mathcal{O}_u, \mathcal{S}_{u,\alpha}(\theta, \tau)), \quad (4.16)$$

όπου το CE υποδηλώνει το σφάλμα cross-εντροπία (Cross Entropy Loss) αφού η μάσκα είναι δυαδική και μοιάζει με πιθανότητα. Δίνεται από τον τύπο:

$$CE := - \sum_{i=1}^N [y_i \log(p_i) + (1 - y_i) \log(1 - p_i)],$$

όπου y_i είναι η πραγματική τιμή της μάσκας της εικόνας και p_i είναι η εκτίμηση με βάση την εκτιμήτρια συνάρτηση S_α 4.13. Με αυτό το σφάλμα εκπαιδεύονται σημαντικά οι θέσεις των καμερών και λίγο και η γεωμετρία παρά η εμφάνιση.

Σφάλμα απεικόνισης eikonal loss

Επειδή το πεδίο f πρέπει να προσεγγίζει την γεωμετρία της 3D σκηνής με βάση το θεώρημα της περιβάλλοντας συνάρτησης και δηλαδή να είναι, κατά προσέγγιση πεδίο προσημασμένης απόστασης εισάγεται ο παράγοντας κανονικοποίησης της γεωμετρίας ή σφάλμα απεικόνισης το οποίο δείχνει πόσο κοντά είναι το πεδίο σε ένα πεδίο SDF. Δηλαδή πόσο ισχύει η συνθήκη ομοιόμορφα για ένα φραγμένο χώρο για τα σημεία ανήκουν στο πεδίο.

$$Loss_E(\theta) = \mathbb{E}_x (\|\nabla_x f(x; \theta)\| - 1)^2 \quad (4.17)$$

Εκπαίδευση IDR

Η εκπαίδευση ξεκινά με την φόρτωση των δεδομένων σε αντικείμενα της PyTorch *data loaders*. Εκτελείται απευθείας *Sphere Tracing* της σκηνής, αποτυπώνοντας το αρχικό πεδίο γεωμετρίας στο οποίο είναι αρχικοποιημένο το δίκτυο (περιβάλλουσα σφαίρα, *bounding sphere*, *Implicit Geometric Regularization* [23]). Στην συνέχεια ξεκινά το δίκτυο δειγματοληψίας να ελέγχει το έμμεσο πεδίο γεωμετρίας και οι παράμετροι μεταβάλλονται

με βάση το συνολικό σφάλμα της συνάρτησης κόστους, που υπολογίζεται στην έξοδο. Τα δίκτυα υψηλοσυχνοτικής κωδικοποίησης εφαρμόζονται στην αρχή των δικτύων γεωμετρίας και αποτύπωσης και τα σφάλματα γυρνάνε πίσω σε αυτά μέσω της διαδικασίας της *PyTorch*, *torch.autograd* η οποία υπολογίζει τις κλίσεις των διανυσμάτων σφάλματος σε όλο τον υπολογιστικό γράφο. Κάθε 100 εποχές γίνεται κλήση του *Marching Cubes* αλγορίθμου για την εξαγωγή αρχείου πλεγμάτων τριγώνων της 3D σκηνής από την έμμεση αναπαράσταση της *SDF*. Ταυτόχρονα το μοντέλο φωτογραφίζεται για την δεδομένη πόζα δηλαδή εγγενείς παραμέτρους κάμερας και αποθηκεύεται η εικόνα αποτύπωσης σε σχέση με την πραγματική εικόνα της φωτογραφίας. Σε αυτή την φάση έχουμε διαδικασία ανάκλησης του μοντέλου (*inference*), η οποία διακόπτει την εκπαίδευση προσωρινά. Η εκπαίδευση διαρκεί 2000 εποχές και ο αλγόριθμος βελτιστοποίησης αυξάνει σταδιακά τον παράγοντα που προσεγγίζει την δυαδική μάσκα α ², του βήματος μάθησης (*learning rate*) και άλλων παραμέτρων όπως αναφέρεται παρακάτω.

Αλγόριθμος Βελτιστοποίησης

Στα πλαίσια της εκπαίδευσης και συγκεκριμένα της ανανέωσης των βαρών το σύνολο του αλγορίθμου εκπαίδευσης, χρησιμοποιεί τον αλγόριθμο βελτιστοποίησης **Adam** με δρομολόγηση πτώσης βήματος εκπαίδευσης σε συγκεκριμένες εποχές (*Scheduled Learning Rate Decay*). Αυτό περιλαμβάνει σταδιακή αύξηση για την παράμετρο της επίδρασης της μάσκας α ώστε η συνάρτηση f να γίνεται σταδιακά η προσέγγιση της μάσκας S_α με βάση το *Envelope Theorem*. Παράλληλα γίνεται πτώση του βήματος εκπαίδευσης στο μισό στις εποχές [250, 500, 750, 1000, 1250]. Όσον αφορά τα βάρη για το σφάλμα μάσκας και το *eikonal loss* αυτά παραμένουν σταθερά.

Όλες αυτές οι ρυθμίσεις αποσκοπούν στην ευσταθή μορφή εκπαίδευσης ώστε να διατηρείται η συνέχεια και η τμηματική διαφορισμότητα της συνάρτησης αναπαράστασης γεωμετρίας αλλά και του πεδίου φωτισμού.

4.3 Συχνοτική Κωδικοποίηση Εφαπτόμενου Πυρήνα Συναρτήσεων Fourier

Η εργασία βασιζόμενη στην ανάλυση που έγινε στο θεωρητικό υπόβαθρο (βλ. 2.5), προτείνει την βελτιστοποίηση της σύγκλισης του σφάλματος κόστους των νευρωνικών δικτύων μέσω παραμετροποίησμων στάσιμων εφαπτόμενων κωδικοποίησεων στην έσοδο. Καταλήγει, έτσι σε τρεις τύπους εφαπτόμενων πυρήνων που είναι σε θέση να αποτυπώσουν υψηλοσυχνοτικές σχέσεις των δεδομένων, οι οποίες είναι καλό να εφαρμόζονται σε δίκτυα συντεταγμένων (κρίσιμο όταν πρόκειται για δεδομένα εικόνας από πολλές όψεις). Αυτοί οι εφαπτόμενοι πυρήνες *Neural Tangent Kernels* διαφέρουν ως προς την μορφή των συναρτήσεων πυρήνα που χρησιμοποιούνται και είναι οι εξής:

- Βασική κωδικοποίηση απεικονίσεων όπου απλά γίνεται μετασχηματισμός των εισόδων γύρω από κύκλο ή σφαίρα με χρήση ημιτονοειδών συναρτήσεων με NTK:

$$\gamma(\mathbf{v}) = [\cos(2\pi \mathbf{v}), \sin(2\pi \mathbf{v})]^T$$

- Κωδικοποίηση με βάση την θέση των συντεταγμένων - *Positional Encoding*. Γίνεται δειγματοληψία συχνοτήτων πάνω σε λογαριθμική κλίμακα σε κάθε διάσταση των συντεταγμένων που κωδικοποιούνται και το εύρος σ επηρεάζει την δυνατότητα του μοντέλου να γενικεύει ή και την ακρίβειά του (είναι παράμετρος που εκπαιδεύεται, και τα αποτελέσματά της είναι εμφανή στα σχήματα 2.16 2.17)). Το εφαπτόμενο πεδίο συναρτήσεων πυρήνα περιγράφεται από την παρακάτω μαθηματική μορφή:

$$\gamma(\mathbf{v}) = [\dots, \cos(2\pi \sigma^{j/m} \mathbf{v}), \sin(2\pi \sigma^{j/m} \mathbf{v}), \dots]^T \text{ για } j = 0, \dots, m-1$$

²To α στο S_α είναι η παράμετρος όταν τείνει στο άπειρο η προσέγγιση της μάσκας τείνει στην δυαδική μάσκα

- Κωδικοποίηση με τυχαία Γκαουσιανή κατανομή. Κρίθηκε πως η κατανομή δειγματοληψίας των συχνοτήτων και οι ίδιες οι συχνότητες \mathbf{B} , είναι και αυτές παράμετροι προς εκπαίδευση και μπορούν να δειγματοληπτηθούν αρχικά από κάποια κανονική κατανομή $\mathcal{N}(0, \sigma^2)$

$$\gamma(\mathbf{v}) = [\cos(2\pi\mathbf{B}\mathbf{v}), \sin(2\pi\mathbf{B}\mathbf{v})]^T$$

Έτσι, γίνεται χρήση ενός πυρήνα ημιτονοειδών συναρτήσεων κωδικοποίησης στο IDR ή μετασχηματιστή (*transformer*), θα μπορούσαμε να πούμε, $\gamma(x; \theta, \gamma, \tau)$ το οποίο στην μία εκδοχή του χαρακτηρίζεται Fourier Features όταν χρησιμοποιεί ημιτονοειδή συναρτήσεις πυρήνα σε συχνότητες από Gaussian κατανομή.

Σε άλλη εκδοχή του πυρήνα συναρτήσεων κωδικοποίησης, έχουμε *positional embedding* μετασχηματισμό εισόδου, που η δειγματοληψία σχετίζεται με την θέση. Ωστόσο και τα δύο δείχνουν ότι η εκπαίδευση τους δίνει παρόμοια αποτελέσματα χωρίς ιδιαίτερη επιτάχυνση της σύγκλισης στην εκπαίδευση του *IDR*.

4.4 Χωρικός Κατακερματισμός Πολλαπλής Ανάλυσης

Η διαδικασία *multi-resolution hash encoding* όπως είπαμε αφορά την κωδικοποίηση του δικτύου με βάση την διαμέριση του χώρου.

Ο τρόπος που εφαρμόζεται κάτι τέτοιο είναι με το σπάσιμο του χώρου σε μερικά πλέγματα (*Grids*) τετραγώνων εφόσον το πεδίο που εφαρμόζεται ας πούμε είναι εικόνες τα οποία θα αυξάνουν σε ανάλυση. Αυτό θα εφαρμόζεται iεραρχικά και μπορεί να οριστεί όμως σε οποιοδήποτε χώρο δεδομένων (παραδείγματος χάρη μπορεί να εφαρμοστεί στο NeRF που χρησιμοποιεί μαζί το σημείο της εικόνας με το τρισδιάστατο διάνυσμα της ακτίνας που περνά από το pixel).

Έστω πως \mathbf{d} είναι οι διαστάσεις της εισόδου. Η ανάλυση του κάθε πλέγματος διαμέρισης είναι σταθερή και διαφορετική από το προηγούμενο επίπεδο διαμέρισης αλλά δεν χρειάζεται ο τρόπος με τον οποίο ανεβάζουμε την ανάλυση διαμέρισης να είναι σταθερός. Έτσι διακρίνονται οι παρακάτω υπερπαράμετροι αυτής της μεθοδολογίας:

- $L :=$ Πλήθος Επιπέδων Ανάλυσης
- $N_{min} :=$ Ελάχιστη Ανάλυση πλέγματος
- $N_{max} :=$ Μέγιστη Ανάλυση πλέγματος

και ο τρόπος υπολογισμού του παράγοντα αύξησης ανάλυσης να δίνεται από τον τύπο (γεωμετρική πρόοδος μεταξύ χαμηλής και υψηλής ανάλυσης ογκομετρικού πλέγματος *grid*)

$$b := \exp \frac{\ln N_{max} - \ln N_{min}}{L - 1}$$

Συνεπώς η ανάλυση του πλέγματος σε ένα επίπεδο ℓ είναι $N_\ell := \lfloor N_{min} \cdot b^\ell \rfloor$

Ο μετασχηματισμός της εισόδου σε κάποια επιλεγμένη ανάλυση γίνεται πρώτα κλιμακώνοντας το ζητούμενο σημείο x κατά την ανάλυση του *grid*, N_ℓ και εφαρμόζοντας στρογγυλοποίηση προς τα πάνω ή κάτω για να βρούμε το κελί που έχει το x . Αυτό το κελί αποτελεί τη χωρική θέση του σημείου στο *grid*.

Αφού βρεθεί η θέση για να αποδοθεί το χαρακτηριστικό των συντεταγμένων εφαρμόζεται μετασχηματισμός κατακερματισμού *Hash Encoding*. Όπου εισάγονται άλλες δύο υπερπαράμετροι που καθορίζουν και την αναλυτικότητα κωδικοποίησης:

- $T :=$ Μέγεθος Πίνακα Κατακερματισμού,
- $F :=$ Μέγιστος αριθμός χαρακτηριστικών συντεταγμένων ανά κελί,

με την συνάρτηση κατακερματισμού να είναι μια συνάρτηση που εφαρμόζει *bitwise XOR*, όπως αναφέρθηκε και στο θεωρητικό υπόβαθρο και η εφαρμογή του *hash* χαρακτηριστικού στον πίνακα να γίνεται απλά εφαρμόζοντας την πράξη της ακέραιας διαίρεσης του *hash* με τον μέγεθος του πίνακα T

$$h(\mathbf{x}) = \bigoplus_{i=1}^d x_i \pi_i \mod T \quad (4.18)$$

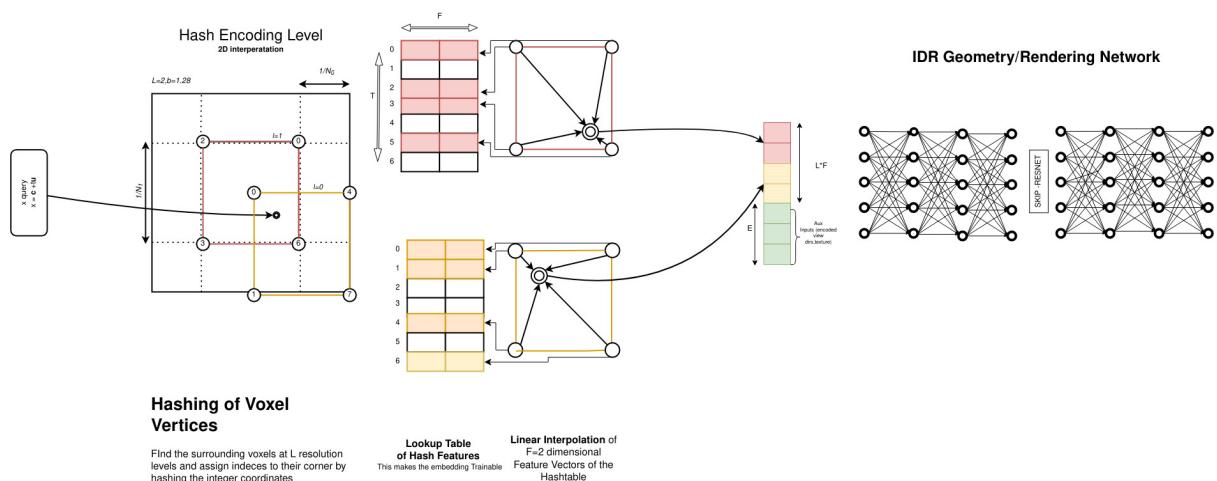
Κάθε κελί έχει F παραμέτρους εκπαίδευσης. Κατά την διάρκεια της εκπαίδευσης γίνεται οπισθοδιάδοση των κλίσεων μέχρι και τους πίνακες *hash* ώστε να δυναμικά να εκπαιδευτεί η κωδικοποίηση και αυτή η χωρική διαμέριση. Όπως αναφέρθηκε δεν γίνεται χειρισμός των *hash* συμπίπτουν (λόγω φύσης της συνάρτησης κατακερματισμού), αλλά κατά την διάρκεια της εκπαίδευσης γίνονται τα χαρακτηριστικά *hash* ανθεκτικά σε αυτό το φαινόμενο, που υπό βοηθείας και από τους παραμέτρους π_i που βελτιώνουν την συνοχή προσωρινής μνήμης αλλά σε μικρή κλίμακα δεν λύνει το πρόβλημα.

Τρίτο βήμα σε αυτή την διαδικασία αποτελεί η παλινδρόμηση η οποία αφορά την γραμμική προσέγγιση των 2^d πινάκων *hash* που αντιστοιχούν στο παρόν κελί του πλέγματος. Αυτή η διαδικασία πραγματοποιείται

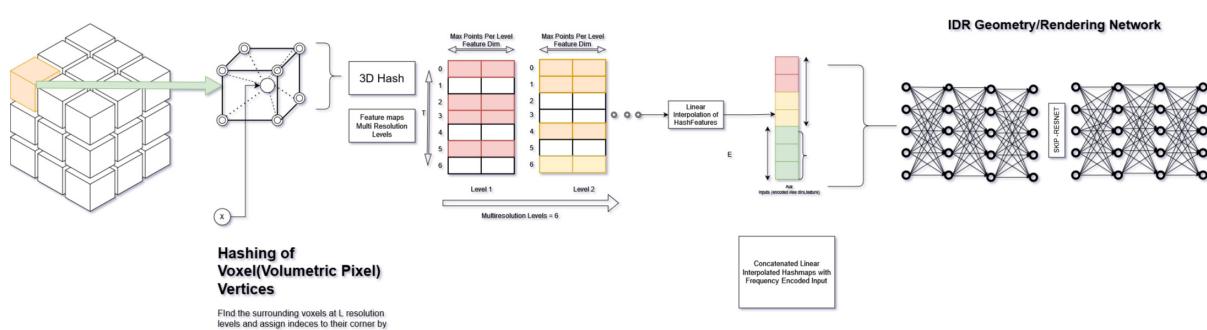
είτε με δι-γραμμική παρεμβολή σε δύο διαστάσεις είτε με τριγραμμική. Δηλαδή στις δύο διαστάσεις με δυο οριζόντιες παρεμβολές στις τετμημένες πρώτα και μετά στις τεταγμένες. Σε αυτό το σημείο εφαρμόζεται και το νευρωνικό στρώμα προσοχής το οποίο δεν διακριτοποιεί την συνάρτηση παλινδρόμησης άλλα αντιστοιχεί τις διακριτές τιμές των πινάκων σε βάρη τα οποία βελτιώνονται κατά την διαδικασία της εκπαίδευσης με τον αλγόριθμο βελτιστοποίησης (π.χ. Adam). Συνεπώς επί του συνόλου το *hash encoding* αποτελεί συνεχή μορφή κωδικοποίησης παρ' όλη την διακριτή μορφή ορισμένων διαδικασιών που χρησιμοποιεί.

Τελευταίο βήμα αυτής της διαδικασίας είναι η συνένωση των χαρακτηριστικών διανυσμάτων που αντιστοιχήθηκε η είσοδος δημιουργώντας ένα συνολικό διάνυσμα διάστασης $L * F$ εφόσον έχουμε L πλήθος αναλύσεις F χαρακτηριστικών σημείων ανά ανάλυση. Τα οποία συνενώνονται με την ίδια την είσοδο δίνοντας μια ικανοποιητική μορφή πεδίου για το *IDR*.

Παρακάτω φαίνεται η αναπαράσταση του δικτύου μετασχηματισμού με είσοδο δύο διαστάσεων όπως περιγράφεται και τριών διαστάσεων, όπως αυτή που υλοποιήθηκε υλοποιήθηκε:



Σχήμα 4.3: 2D ανάλογο end-to-end αρχιτεκτονικής δικτύου με χρήση Multi Resolution Hash Grid Encoding



Σχήμα 4.4: 3D Hashgrid MLP IDR Proxy

Στο Παράρτημα Γ 7, υπάρχει μια πιο αναλυτική μορφή του αλγορίθμου που υλοποιήθηκε σε *python* σε μορφή ψευδογλώσσας.

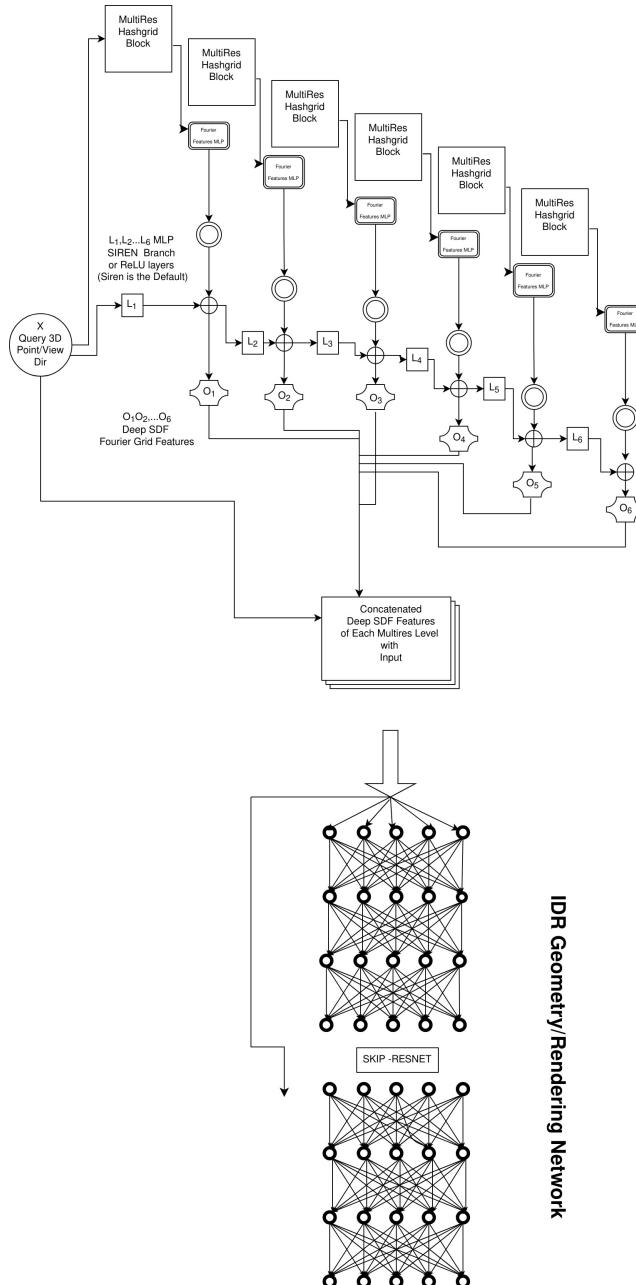
4.5 Θυρίδες Βαθιάς Συχνοτικής/Χωρικής Αποσυσχέτισης και Κωδικοποίησης

Το δίκτυο της $3D$ κωδικοποίησης μέσω νευρωνικών θυρίδων Fourier. Συνδυάζει τις παραπάνω μεθόδους κωδικοποίησης σε ένα ολοκληρωμένο δίκτυο το οποίο αποσυσχετίζει συχνοτικά τον τρισδιάστατο χώρο σε υψηλής και χαμηλής συχνότητας χαρακτηριστικά με το σκεπτικό του μετασχηματισμού κυματιδίων (*Wavelet Transform*).

Η ιδέα προήλθε από το κείμενο της έρευνας *Neural Fourier Filter Banks* [1]. Κατεύθυνση της ιδέας είναι να αξιοποιήσει την χωρική διαμέριση που προσφέρει η νευρωνική κωδικοποίηση *Multi-Resolution Hash Encoding* και να κωδικοποιήσει τα υψηλής διάστασης χαρακτηριστικά ζεχωριστά σε κάθε διαμέριση με συχνοτικούς μετασχηματισμούς ανάλογους του *Positional Encoding*. Έτσι το νευρωνικό δίκτυο αναπαριστά φίλτρα χαμηλών και υψηλών συχνοτήτων τα οποία αντιστοιχίζουν με αυτό τον τρόπο τις συντεταγμένες σε ειδικές δομές που λέγονται *Fourier Grid Features* και με αυτόν τον τρόπο επιτρέπεται καλύτερος εντοπισμός στην κλίμακα αλλά και στην συχνότητα βαθιών *SDF* χαρακτηριστικών.

Η διαδικασία περιλαμβάνει την μετατροπή των *Grid Features* (κεφ.4.4) σε *Fourier Features* εφαρμόζοντας τον μετασχηματισμό Fourier μέσω συναρτήσεων πυρήνα που περιγράφηκε στο κεφάλαιο 4.3. Ταυτόχρονα για την διατήρηση της υψηλοσυχνοτικής πληροφορίας χρησιμοποιείται συμπληρωματικό πλήρως συνδεδεμένο δίκτυο με ημιτονοειδή συναρτήσεις που οδηγεί τα χαμηλοσυχνοτικά και υψηλοσυχνοτικά χαρακτηριστικά (ανάλογο του *SIREN* δικτύου).

Η συνολική μορφή που παίρνει το δίκτυο IDR μπορεί να αποδοθεί με το παρακάτω διάγραμμα.



Σχήμα 4.5: Αρχιτεκτονική δικτύου 3D Κωδικοποίησης NFFB

4.6 Διαμόρφωση/Αποδιαμόρφωση Χαρακτηριστικών Σήματος

Η τεχνική που ακολουθείται στην διαμόρφωση και αποδιαμόρφωσης βαρών βασίζεται στον αλγόριθμο *AdaIN*(*Adaptive Instance Normalization*, ο οποίος δεν λειτουργεί σε ομάδα δεδομένων αλλά σε κάθε δεδομένο χαρακτηριστικό ξεχωριστά.

Η τεχνική αυτή αποδίδεται με χρήση ενός *Style Modulation Block* το οποίο εισάγεται στα γραμμικά στρώματα της νευρωνικής θυρίδας φύλτρων Fourier (*NFFB*) και χρησιμοποιεί την είσοδο ως δεδομένο και την κωδικοποίηση Fourier σε δεδομένη ανάλυση του πλέγματος κατακερματισμού ως διάνυσμα χαρακτηριστικών. Αντί να εφαρμόσει τον κλασικό αλγόριθμο προσαρμοστικής κανονικοποίησης πάνω στις στατιστικές του διανύσματος χαρακτηριστικών που περιγράφεται με την εξίσωση:

$$\text{AdaIN}(x, y) = \mu(y) + \sigma(y) \cdot \left(\frac{x - \mu(x)}{\sigma(x)} \right), \quad (4.19)$$

εφαρμόζεται ένα επίπεδο προσοχής το οποίο υπολογίζει τα βάρη προσοχής ως το εσωτερικό γινόμενο πραγματικής εισόδου με διάνυσμα χαρακτηριστικών. Αυτά τα βάρη κανονικοποιούνται με συνάρτηση *Softmax* στο διάστημα $(0, 1)$ και χρησιμοποιούνται για τον υπολογισμό του αποδιαμόρφωμένου διανύσματος χαρακτηριστικών. Έτσι εφαρμόζεται η τεχνική διαμόρφωσης και αποδιαμόρφωσης των βαρών του δικτύου χωρίς να βασιζόμαστε σε στατιστικές των δεδομένων ενός συνόλου εισόδων.

4.7 Υλοποιημένα Δίκτυα και Παράμετροι

Ένα μεγάλο μέρος των δικτύων έμμεσης διαφορίσιμης αποτύπωσης, που έχουν χρησιμοποιηθεί, βασίζονται σε *PyTorch* και έχουν δανειστεί από το αποθετήριο του IDR³ <https://github.com/lioryariv/idr>. Ευχαριστούμε την παραπάνω ερευνητική εργασία που επέτρεψε να δοθεί έμφαση στο κομμάτι των βαθιών δικτύων κωδικοποίησης των πεδίων που αποτυπώνει.

Βέβαια έχουν γίνει αρκετές παραμετροποίησεις στον κώδικα του δικτύου στο σημείο που να μην λειτουργούν με τον ίδιο τρόπο επομένως έχει δημιουργηθεί και ειδικό αποθετήριο για την παρούσα εργασία που δίνεται στο παράρτημα (βλ. 7 Παράρτημα Γ Δείγματα Αλγορίθμων).

Το δίκτυο γεωμετρικής εκτίμησης της επιφάνειας, $F(p; \theta) = (f(x; \theta), z(x; \theta)) \in \mathbb{R} \times \mathbb{R}^{256}$, είναι ένα δίκτυο MLP 8 επιπέδων πλάτους 512 νευρώνων και αποτελεί ResNet (*Residual Net*) γιατί πραγματοποιεί διακοπή της σύνδεση στο 4o επίπεδο όπου εισάγονται ξανά οι είσοδοι ώστε να διατηρηθεί η αριθμητική τιμή των κλίσεων των βαρών (και να μην χαθεί η πληροφορία των δεδομένων). Οι παράμετροι των γραμμικών του επιπέδων, αρχικοποιούνται ώστε να περιγράφει μια περιβάλλουσα μοναδιαία σφαίρα ενώ οι παράμετροι z αποτελούν τα DeepSDF χαρακτηριστικά τα οποία βελτιώνονται μετασχηματίζοντας την είσοδο με την βαθιά δίκτυα κωδικοποίησης. Σημαντική αλλαγή είναι πως χρησιμοποιείται συνάρτηση πυκνότητας Laplace στο μέρος της εξόδου που αφορά το σημείο της επιφάνειας μαζί με συνάρτηση υπερβολικής εφαπτομένης ώστε να προσεγγίζουμε ομαλά με προτίμηση από έξω δηλαδή από τιμές SDF θετικές και να κόβονται ακραίες τιμές που υπερβαίνουν την μοναδιαία σφαίρα. Αυτή η τακτική *Truncated SDF* χρησιμοποιείται συχνά ως μέτρο ευστάθειας, ειδικά όταν υπάρχουν έξοδοι που δίνουν πολύ μεγάλα βάρη στην κωδικοποίηση της εισόδου.

Το δίκτυο αποτύπωσης $\mathcal{M}(\hat{p}, \hat{n}, \hat{z}, \mathbf{w}; \gamma) \in \mathbb{R}^3$, δίνει την εκτίμηση του χρώματος στο πρότυπο [R, G, B] του pixel της εικόνας από το οποίο διέρχεται η ακτίνα με μοναδιαίο διάνυσμα κατεύθυνσης \mathbf{w} . Είναι ένα MLP 4 επιπέδων με ίδιο πλήθος νευρώνων με το δίκτυο γεωμετρικής εκτίμησης δηλαδή 512. Η είσοδος του δικτύου αποτελεί συνενωμένο το διάνυσμα διαφορίσιμων σημείων, κάθετων διανυσμάτων, διανυσμάτων όψης και βαθιών χαρακτηριστικών SDF ώστε να μπορέσει να υλοποιηθεί η έμμεση εκπαίδευση του δικτύου γεωμετρίας από τις εικόνες με μια συνεχή συνάρτηση που αποτελεί **P-Universal** σύστημα αποτύπωσης (βλ. 7 Phong Reflection).

³διατίθεται με MIT άδεια χρήσης, δηλαδή μπορεί να χρησιμοποιηθεί και να επεκταθεί για ερευνητικούς σκοπούς όπως γίνεται στην προκειμένη περίπτωση

Τα προαναφερθέντα δίκτυα, είναι διαφοροποιημένα στα μέτρα της εργασίας και είναι απλά MLP. Η παρούσα έρευνα αποτελεί κυρίως την υλοποίηση των παρακάτω δικτύων και την συνολική ρύθμιση της συνολικής γραμμής του αλγορίθμου ώστε να δουλεύει σωστά με την επέκταση («*plug-in*») των δίκτυα κωδικοποίησης.

Οι πυρήνες συναρτήσεων μετασχηματισμού και τα δίκτυα υψηλοσυχνοτικής κωδικοποίησης εφαρμόζονται κατά το εμπρόσθιο πέρασμα των παραπάνω δικτύων στις εισόδους τους και επιστρέφουν κωδικοποιημένα σημεία διάστασης 259 (όση δηλαδή το *feature vector* z μαζί με την διάσταση της εισόδου). Τα δίκτυα γεωμετρίας και αποτύπωσης, εκτελούν τον υπολογισμό προς τα πίσω των κλίσεων των σφαλμάτων και στα δίκτυα κωδικοποίησης.

Οι αλγόριθμοι και τα δίκτυα κωδικοποίησης, επίσης, μπορούν να αναπαριστούν λόγω της ιδιαίτερης δομής τους με MLP στρώματα και από μόνα τους σημεία στον χώρο (αναπαριστούν συνεχείς συναρτήσεις). Χρησιμοποιούνται κυρίως όμως, ως δίκτυα μετασχηματισμού της εισόδου στην παρούσα εργασία καθώς η είσοδος και η έξοδός τους δεν χειρίζεται με τρόπο ώστε να αναπαριστούν πεδία πληρότητας σημείων (*occupancy fields*). Οι παράμετροι τους δίνονται στα πειράματα.

Πλήρως Συγχωνευμένα Δίκτυα Βαθιάς Κωδικοποίησης TinyCudaNN

Η Nvidia παρέχει μια βιβλιοθήκη σύνθεσης πλήρως συγχωνευμένων δικτύων *MLP* που υπολογίζουν το *Multi-Resolution HashGrid Encoding* της οποίας η χρήση γίνεται στα δίκτυα κωδικοποίησης που αναφέρθηκαν προηγουμένως (ως εναλλακτική στα δίκτυα που υλοποιήθηκαν αρχικά). Ο λόγος γίνεται για την *TinyCudaNN*[83], μια πλήρως παραλληλοποιημένη επέκταση της *PyTorch* που δίνει την δυνατότητα χρήσης και πράξεων με αριθμούς κινητής υποδιαστολής μισής ακρίβειας (*Half operations*), για όσους διαθέτουν νέες μορφές καρτών γραφικών (διαθέτουν ειδική μονάδα επεξεργασίας για γραφικά που βασίζονται στην ιχνηλάτιση ακτίνας - *RT cores*). Τα δίκτυα αυτά είναι σε θέση βάζοντας περισσότερους κόμβους σε κάθε επίπεδο αλλά διατηρώντας μικρό πλάτος δικτύων (εξ ου και το *tiny* στην ονομασία τους), να επιταχύνουν την διαδικασία εκπαίδευσης των χωρικών πλεγμάτων των MLPs που αναπαριστούν την κωδικοποίηση συνταγμένων *Hash*. Επειδή η ταχύτητα εκπαίδευσης είναι κρίσιμη παράμετρος ερευνήθηκαν και μέθοδοι *CUDA* δικτύων ως *C++ binding* επεκτάσεις στην *PyTorch* του *HashGrid Encoding* και δημιουργήθηκαν κάποια από τα προαναφερθέν δίκτυα με την χρήση της βιβλιοθήκης αλλά και με χρήση *CUDA* υπολογισμών⁴.

⁴Επειδή δεν υπήρχε διαθέσιμη σύγχρονη μορφή κάρτας δεν μπορούσαν να χρησιμοποιηθούν *half operations* ούτε *mixed precision* εκπαίδευση αλλά η βιβλιοθήκη παρέχει δίκτυο συμβατό που το ονομάζει *CutlassMLP*

5

Πειράματα - Αποτελέσματα

5.1 Πειραματική Φάση

5.1.1 Λεπτομέρειες DTU Συνόλου δεδομένων

Το σύνολο στο οποίο εφαρμόζονται τα μοντέλα υψηλοσυχνοτικής 3D ανακατασκευής είναι πραγματικές 2D εικόνες από το DTU MVS αποθετήριο[81]. Τα πειράματα τρέχουν σε 4 απαιτητικές σκηνές εκ των οποίο η πρώτη περιέχει 49 και οι υπόλοιπες 64 υψηλής ποιότητας εικόνες με γεωμετρία και φωτισμό και υφή που διαφέρει. Το σύνολο δεδομένων διαθέτει επίσης 3D γεωμετρικές αναπαραστάσεις και συγκεκριμένες παραμέτρους καμερών οι οποίες υπολογίστηκαν αφού το σύνολο φωτογραφήθηκε με ρομποτικό βραχίονα. Οι δυαδικές μάσκες εκτός από την σκηνή 65, που παρέχεται, υπολογίζονται.

5.1.2 Πειράματα

Στην πειραματική διαδικασία εκτελέστηκε πλήθος πειραμάτων πάνω στο *IDR* με διαφορετικούς αλγορίθμους και δίκτυα κωδικοποίησης της εισόδου. Αρχικά έγινε μια μελέτη του γνήσιου *IDR* και πώς θα μπορούσε να ενσωματώσει αλγορίθμους και δίκτυα υψηλοσυχνοτικής κωδικοποίησης του διανύσματος εισόδου. Στην συνέχεια χρησιμοποιήθηκαν βασικοί μέθοδοι πάνω στον εφαπτόμενο πυρήνα *Fourier*. Ένα πείραμα αφορούσε την δειγματοληψία συχνοτήτων από *Gaussian* κατανομή το οποίο αναφέρεται ως αλγόριθμος *Fourier Features*, ενώ το άλλο αφορούσε την λογαριθμική δειγματοληψία που λαμβάνει υπόψιν και την θέση των δεδομένων δηλαδή *Positional Encoding*. Αυτοί οι αλγόριθμοι κατάφεραν να δώσουν αξιόπιστα αποτελέσματα και κατά την διάρκεια εκπαίδευσης επιλέχτηκε αντιπροσωπευτικό εύρος συχνοτήτων ώστε να μην έχουμε πρόβλημα υπερεκπαίδευσης.

Τα βασικά πειράματα ξεκίνησαν με την εισαγωγή της μεθόδου νευρωνικής κωδικοποίησης κατακερματισμού πολλαπλών αναλύσεων (*Multi-Resolution Hash Encoding*). Αυτό αφορά την εκπαίδευση κωδικοποίησης hash μέσω του στρώματος ενσωμάτωσης των βαρών παρεμβολής μεταξύ των κωδικοποιημένων γειτόνων. Στα πλαίσια αυτά έγινε μια υλοποίηση του μοντέλου σε *PyTorch* και ακολουθήθηκαν αρχικά οι παράμετροι που πρότεινε η έρευνα *Instant-NGP* [2].¹. Αυτό ωστόσο, λόγω του μεγάλου πλήθους επιπέδων ανάλυσης που αντι-

¹ χρησιμοποιήθηκε και το ίδιο *Instant-NGP* για να δούμε πως λειτουργεί

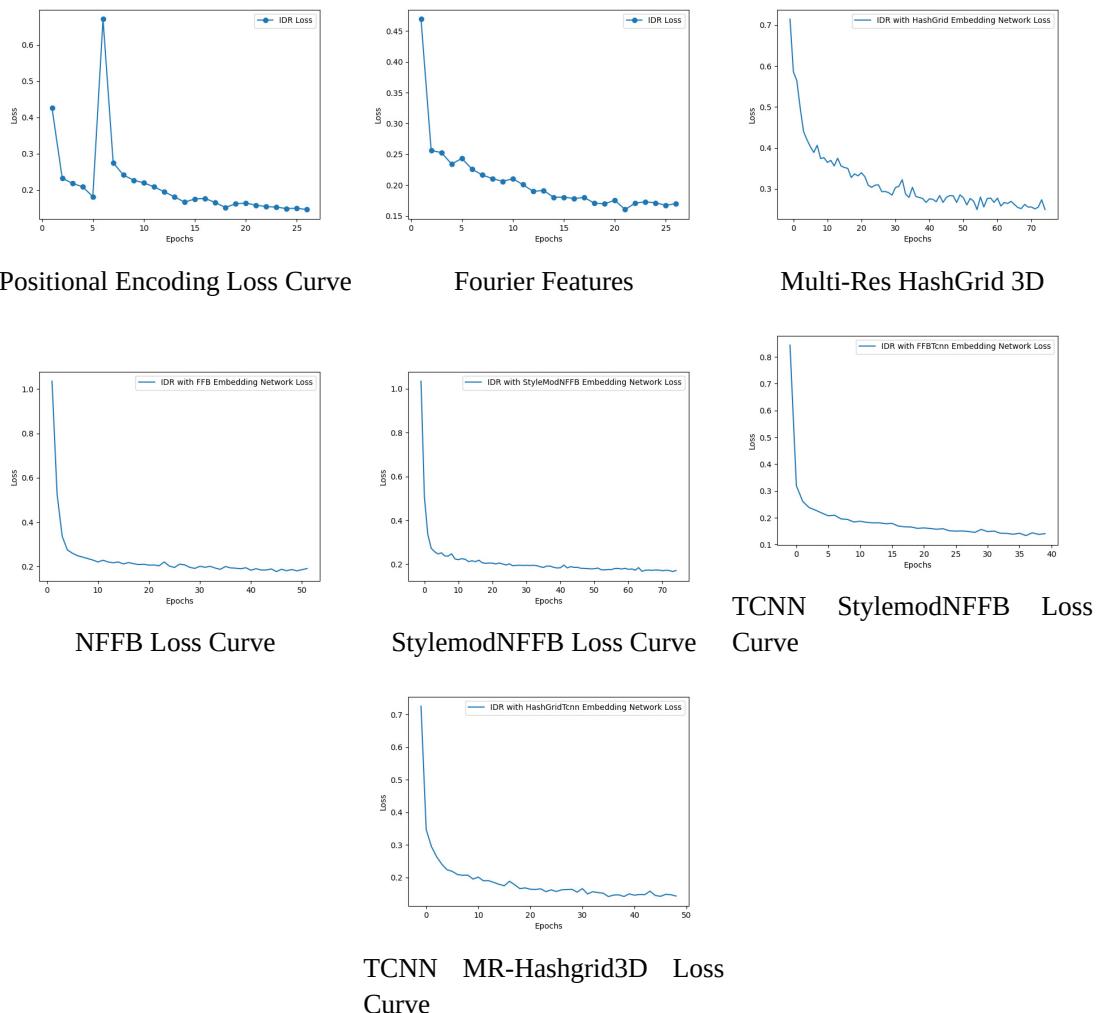
στοιχούν σε εύρος χωρικών διαμερισμών, οδήγησε γρήγορα σε προβλήματα πόρων επομένως επιλέχθηκαν οι κλασσικοί παράμετροι που προτείνει το *IDR* με συμβιβασμό σε λιγότερα επίπεδα αναλύσεων. Παράλληλα εξετάστηκε και η βιβλιοθήκη των πλήρως συγχωνευμένων δικτύων κωδικοποίησης *Hash Grid*, *TinyCudaNN*[83]. Αυτή διαθέτει ήδη υλοποιημένο το δίκτυο κωδικοποίησης *Hash Grid* και απλά επιτρέπει το ορισμό των υπερπαραμέτρων του. Έτσι δημιουργήθηκαν και μοντέλα που χρησιμοποιούν αυτές τις υλοποιήσεις και αναφέρονται συνήθως με την επέκταση *TCNN*. Τέλος το δίκτυο *Hash Grid 3D* υπάρχει και σε υλοποίηση *CUDA* που ωστόσο δεν χρησιμοποιήθηκε εκτενώς στην πειραματική φάση. Έτσι επίσημα τουλάχιστον, έγιναν 2 σειρές πειραμάτων στο *MR-HashGrid-3D* που είναι η *PyTorch* υλοποίηση του δικτύου και στο *HashGrid_TCNN* που βασίζεται στην βιβλιοθήκη της *Nvidia*².

Τα πειράματα ολοκληρώνονται με τα δίκτυα νευρωνικών θυρίδων *Fourier* και της παραλλαγής του που εφαρμόζει την μέθοδο *style modulation*. Τα συγκεκριμένα δίκτυα εισάγουν και τους αλγορίθμους συχνοτικής κωδικοποίησης με τον εφαπτόμενο πυρήνα *Fourier* και τα δίκτυα κατακερματισμού πολλαπλής ανάλυσης *Hash* μαζί με γραμμικά πλήρως συνδεδεμένα δίκτυα τα οποία μοιάζουν με το δίκτυο *SIREN* [84]. Τα συγκεκριμένα πειράματα είχαν αρκετή δυσκολία στην εύρεση κατάλληλων υπερπαραμέτρων ώστε να δουλεύουν σωστά. Ταυτόχρονα επιχειρήθηκε να χρησιμοποιηθούν απευθείας ως έμμεσο δίκτυο γεωμετρίας αλλά χωρίς καλά αποτελέσματα. Έτσι δημιουργήθηκαν τρεις σειρές τελικών πειραμάτων για τις 4 σκηνές *NFFB*, *StylemodNFFB*, *StylemodNFFB_TCNN*. Σύνολο 7 πειράματα για όλες τις σκηνές δηλαδή 28 αυτοτελή πειράματα.

²στο *HashGrid_TCNN* εκπαιδεύτηκαν μόνο 2 από τις 4 σκηνές λόγω περιορισμένου χρόνου πειραμάτων

5.1.3 Καμπύλες Εκπαίδευσης Αλγορίθμων

Καμπύλες Σφάλματος 50-100 Εποχών



Σχήμα 5.1: Σύγκριση Σύγκλισης Αλγορίθμων στο Σφάλμα Εκπαίδευσης

Η ταχύτητα σύγκλισης των αλγορίθμων θα μπορούσε να συγκριθεί με την κλίση των καμπυλών εκπαίδευσης και τα τελικά ορόσημα σφαλμάτων που επιτυγχάνονται στις 2000 εποχές εκπαίδευσης. Η εκπαίδευση που εκτελέστηκε ήταν διακοπτόμενη και με επανεκκινήσεις για λόγους ευστάθειας του εξοπλισμού αφού σε ορισμένες περιπτώσεις συσσωρεύονται ταυτότετες κλίσεων σφαλμάτων που υπερχειλίζουν την μνήμη καθώς οι υπολογισμοί γίνονται σε μεγάλο βάθος δικτύων.

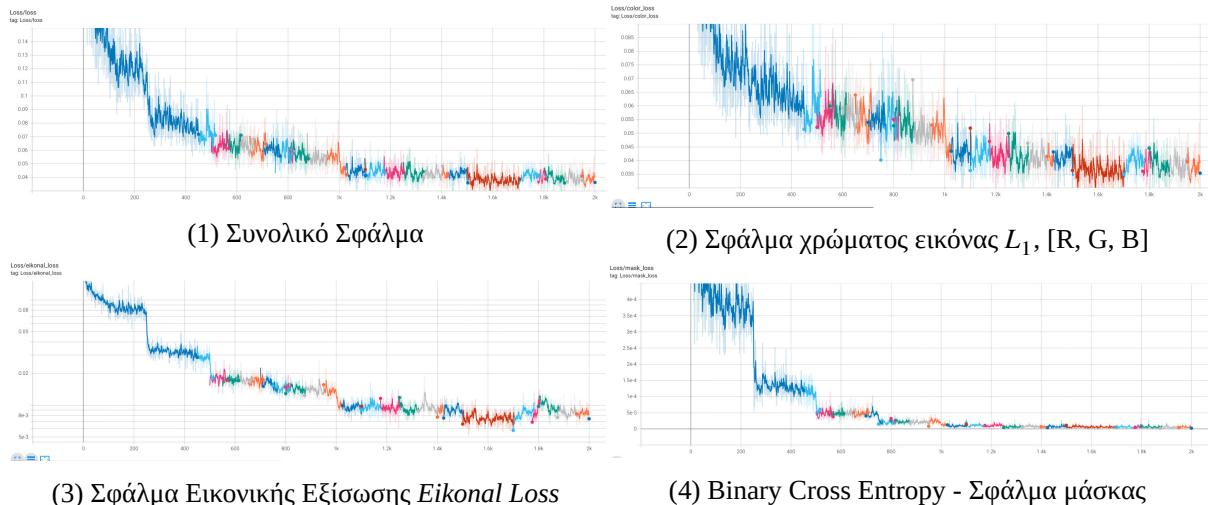
5.1.4 Σχόλια για την ταχύτητα των αλγορίθμων

Οι αλγόριθμοι διαφέρουν ως προς τον χρόνο που ολοκληρώνουν την εκπαίδευση ξεκινώντας από του αλγορίθμους κωδικοποίησης με πυρήνες Fourier να απαιτούν 24 ώρες για την ολοκλήρωση των 2000 εποχών ενώ το δίκτυο κατακερματισμού πολλαπλών αναλύσεων *Multi-Resolution HashGrid3d* να χρειάζεται κάτι λιγότερο από μια μέρα για μία σκηνή του συνόλου δεδομένων. Σημαντική επιτάχυνση παρατηρείται στα πλήρως

συγχωνευμένα δίκτυα κωδικοποίησης *hash*, μέσω της βιβλιοθήκης *TinyCUDANN*. Αυτά εκπαιδεύονται πλήρως σε λιγότερο από 14 ώρες. Βέβαια αυτή η διαδικασία εκπαίδευσης είναι μια διαδικασία υπερεκπαίδευσης. Πρακτικά απαιτούνται 500 εποχές για να έχουμε ικανοποιητική αναπαράσταση.

Καμπύλες 2000 εποχών εκπαίδευσης

Ακολουθούν γραφήματα που παρουσιάζουν την συνολική πορεία όλων των επιμέρους σφαλμάτων ενδεικτικά για το δίκτυο NFFB. Παράγονται κατά την διάρκεια της εκπαίδευσης με χρήση του εργαλείου *Tensorboard*.



Σχήμα 5.2: Καμπύλες Σφαλμάτων εκπαίδευσης NFFB, σε όλη την διάρκεια της εκπαίδευσης

5.1.5 Σχόλια για την αξιολόγηση

Η εκπαίδευση δεν κάνει διαχωρισμό των ομάδων pixel σε pixel εκπαίδευσης και pixel αξιολόγησης μιας και αυτό δεν είναι το ζητούμενο σε ένα πρόβλημα ανακατασκευής. Δηλαδή δεν είναι στόχος να κρύψουμε δεδομένα από το δίκτυο αποτύπωσης αλλά να καταφέρουμε να κάνουμε το δίκτυο να ελαχιστοποιήσει όσο το δυνατόν περισσότερο την αντικειμενική συνάρτηση βελτιστοποίησης με βελτιώνοντας όλα τα δίκτυα. Ετοι πρακτικά μπορούμε να πούμε πως όλες οι καμπύλες που παρουσιάζονται είναι καμπύλες που δείχνουν την σύγκλιση των συνεχών συναρτήσεων και το μεγαλύτερο μέρος αυτής της διαδικασίας φαίνεται στις πρώτες 50-100 εποχές.

Η διαδικασία αξιολόγησης παράγει κυρίως τα αποτελέσματα μετρικών για την ανακατασκευή των μοντέλων χρησιμοποιώντας ως είσοδο τις ίδιες εικόνες επίβλεψης που χρησιμοποιούνται και στην εκπαίδευση. Είναι εφικτό να κρύψουμε ορισμένες εικόνες από το δίκτυο και να σπάσουμε το πρόβλημα σε δεδομένα εκπαίδευσης και δεδομένα αξιολόγησης ωστόσο αυτό δεν έχει ουσιώδες νόημα στα πλαίσια της ανακατασκευής, μιας και αντικειμενικός στόχος βελτιστοποίησης είναι να δημιουργηθεί μια υψηλής ακρίβεια ανακατασκευή με ήδη λίγα δεδομένα εκπαίδευσης.

5.2 Παρουσίαση Φωτογραφικών Αποτελεσμάτων

Δεδομένου ότι ασχολούμαστε με την 3D ανακατασκευή υψηλοσυχνοτικού περιεχομένου σκηνών με διάφορα δίκτυα και μετασχηματισμούς υψηλοσυχνοτικής κωδικοποίησης, οι πίνακες που δίνονται είναι συναρτήσει των διαφορετικών δικτύων και αλγορίθμων κωδικοποίησης. Οι σκηνές που εκπαιδεύτηκαν με όλα τα

μοντέλα κωδικοποίησης από το DTU είναι 4. Η κουκουβάγια με δείκτη στο σύνολο δεδομένων 122, η νεκροκεφαλή με δείκτη 65, το χρυσό κουνέλι με δείκτη 110 και ο πορσελάνινος Βούδας με δείκτη 114. Όλα τα μοντέλα αυτά, όπως και η πολυψήφια των δεδομένων του DTU, περιέχουν υψηλοσυχνοτικό περιεχόμενο στην γεωμετρία σε συνδυασμό με σύνθετο φωτισμό. Στα πλαίσια της εργασίας χρησιμοποιήθηκαν δεδομένες θέσεις καμερών από το dataset αν και θα μπορούσαν να μην δοθούν και να γίνει αρχικοποίηση [74] και εκπαίδευσή τους. Αυτό έγινε καθώς το ζητούμενο δεν είναι η καλύτερη εκτίμηση παραμέτρων κάμερας.

5.2.1 Φωτογραφίες Ανακατασκευής ανά Εποχή Εκπαίδευσης και στην Αξιολόγηση

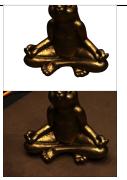
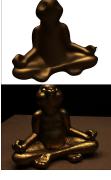
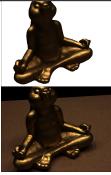
Παρακάτω παρουσιάζονται τα αποτελέσματα ως προς την αποτύπωση εικόνων των σκηνών. Αυτό που ενδιαφέρει είναι η αξιοπιστία δηλαδή το πόσο κοντά στην πραγματική εικόνα που συνοδεύει, είναι η εικόνα αποτύπωσης του υπό κατασκευή μοντέλου.

Αλγόριθμος	Εποχή 100	Εποχή 500	Εποχή 1000	Εποχή 2000	Eval Πόζα 35
Positional Encoding					
FourierNTK					
MR-HashGrid3D					
NFFB					
StyleMod NFFB					
StyleModNFFB (TinyCudaNN)					

Πίνακας 5.1: Αποτελέσματα 3D Ανακατασκευής Σκηνής 65

Αλγόριθμος	Epoch 100	Epoch 500	Epoch 1000	Epoch 2000	Eval Πόζα 55
Positional Encoding					
FourierNTK					
MR-HashGrid3D					
NFFB					
StyleMod NFFB					
StyleModNFFB (TinyCudaNN)					
HashGrid3D (TinyCudaNN)					

Πίνακας 5.2: Αποτελέσματα 3D Ανακατασκευής Σκηνής 122

Αλγόριθμος	Epoch 100	Epoch 500	Epoch 1000	Epoch 2000	Eval Πόζα 55
Positional Encoding					
FourierNTK					
MR-HashGrid3D					
NFFB					
StyleMod NFFB					
StyleModNFFB (TinyCudaNN)					
HashGrid3D (TinyCudaNN)					

Πίνακας 5.3: Αποτελέσματα 3D Ανακατασκευής Σκηνής 110

Αλγόριθμος	Epoch 100	Epoch 500	Epoch 1000	Epoch 2000	Eval Πόζα 55
Positional Encoding					
FourierNTK					
MR-HashGrid3D					
NFFB					
StyleMod NFFB					
StyleModNFFB (TinyCudaNN)					

Πίνακας 5.4: Αποτελέσματα 3D Ανακατασκευής Σκηνής 114

Παρατηρώντας τα αποτελέσματα, βλέπουμε ότι τα μοντέλα *SylemodNFFB*, *NFFB*, και οι αντίστοιχες συγχωνευμένες μορφές τους όσον αφορά την κωδικοποίηση *HashGrid TCNN* συγκλίνουν πολύ πιο γρήγορα στις επιθυμητές παραμέτρους εμφάνισης (χρειάζονται μόνο 500 εποχές) και μάλιστα κάποια από αυτά, είναι σε θέση να αποδώσουν στοιχεία φωτισμού που αντιστοιχούν στο υλικό/υφή (*Phong Material*). Το κλασσικό *IDR*, προσπαθεί να πετύχει αποτελέσματα αυτής της μορφής, αλλά δεν τα καταφέρνει για όλα τα μοντέλα αφού δεν εκπαιδεύεται για τέτοια χαρακτηριστικά του πεδίου φωτισμού. Παρ' όλα αυτά, σε κάποιες εικόνες μπορεί να δούμε λανθασμένα τεχνουργήματα (*artifacts*) του δικτύου. Πόρισμα είναι ότι έχουν να κάνουν με την σφάλμα μάσκας το οποίο δεν διαδίδεται με σωστές αριθμητικές τιμές στο δίκτυο γεωμετρίας.

Οι εικόνες που λείπουν είναι σε εποχές που διακόπηκε η εκπαίδευση ή δεν εκτύπωσαν το αποτέλεσμα(πειράματα υπολογιστικής συστοιχίας) λόγω επανεκκίνησης της εκπαίδευσης.³

³Η διακοπή της εκπαίδευσης, μπορεί να συμβεί λόγω υπέρβασης των ορίων μνήμης ή κάποιας άλλης διακοπής. Ωστόσο, αποθηκεύεται μοντέλο σε τακτά χρονικά διαστήματα .

5.2.2 Αποτύπωσης Προσέγγισης Δυαδικής μάσκας - Μάσκα «βάθους» ανά εποχή

Αλγόριθμος	Εποχή 100	Εποχή 500	Εποχή 1000	Εποχή 2000
Positional Encoding				
FourierNTK				
MR-HashGrid3D				
NFFB				
StyleMod NFFB				
StyleModNFFB (TinyCudaNN)				

Πίνακας 5.5: Αποτελέσματα 3D Ανακατασκευής Σκηνής 65 - Μάσκα Βάθους

Αλγόριθμος	Εποχή 100	Εποχή 500	Εποχή 1000	Εποχή 2000
Positional Encoding				
FourierNTK				
MR-HashGrid3D				
NFFB				
StyleMod NFFB				
StyleModNFFB (TinyCudaNN)				

Πίνακας 5.6: Αποτελέσματα 3D Ανακατασκευής Σκηνής 122 - Μάσκα Βάθους

Αλγόριθμος	Εποχή 100	Εποχή 500	Εποχή 1000	Εποχή 2000
Positional Encoding				
FourierNTK				
MR-HashGrid3D				
NFFB				
StyleMod NFFB				
StyleModNFFB (TinyCudaNN)				

Πίνακας 5.7: Αποτελέσματα 3D Ανακατασκευής Σκηνής 110 - Μάσκα Βάθους

Αλγόριθμος	Εποχή 100	Εποχή 500	Εποχή 1000	Εποχή 2000
Positional Encoding				
FourierNTK				
NFFB				
StyleMod NFFB				
StyleModNFFB (TinyCudaNN)				

Πίνακας 5.8: Αποτελέσματα 3D Ανακατασκευής Σκηνής 114 - Μάσκα Βάθους

Παρατηρείται πως ορισμένες μάσκες βάθους δεν έχουν σωστή προσέγγιση. Αυτό συμβαίνει επειδή υπήρξε το πρόβλημα συσσώρευσης βαρών στο δίκτυο που προκαλούσε διαρροή μνήμης και αναγκαστική επανεκκίνηση με διαφορετικό βάρος μάσκας και παράγοντα προσέγγισης δυαδικής μάσκας α ⁴. Ταυτόχρονα παρατηρείται η αντιληπτική ικανότητα του δικτύου στην εύρεση περιοχών που υπάρχει υψηλοσυχνοτική εναλλαγή βάθους και μεταβολές σε pixel που περιέχουν ή όχι το σημεία του μοντέλου.

⁴ Αυτά τίθενται από τον learning scheduler ο οποίος θέτει σχετικά κοντινή τιμή στην προηγούμενή τους τιμή με βάση τις εποχές εκπαίδευσης

SHencoder

Ως έξτρα πείραμα, πραγματοποιήθηκε εκπαίδευση του μοντέλου *Multi-Resolution HashGrid3D* με χρήση μοντέλου κωδικοποίησης σφαιρικών αρμονικών *SHEncoder* στα διανύσματα κατευθύνσεων των ακτίνων. Δηλαδή διαφορετική κωδικοποίηση σε σημεία και διανύσματα κατευθύνσεων. Το αποτέλεσμα ότι ο αλγόριθμος έτρεχε πιο γρήγορα αλλά η αποτύπωση ήταν πολύ πιο ομαλοποιημένη. Ενδεικτικά παρουσιάζεται η μάσκα βάθος και η αποτύπωση φωτογραφίας στις 500 εποχές.

Αλγόριθμος	Εποχή 500	Εποχή 500
SHEncoder		

Πίνακας 5.9: Αποτελέσματα 3D Ανακατασκευής Σκηνής 122 - SHencoder

5.3 Αξιολόγηση 3D ανακατασκευασμένων μοντέλων

Στα πλαίσια της αξιολόγησης του πόσο καλά λειτουργεί το δίκτυο ανακατασκευής 3D επιφανειών με χρήση των βαθιών δικτύων κωδικοποίησης γίνεται λόγος για το πόσο μοιάζει γεωμετρικά το μοντέλο που παράγεται από το δίκτυο σε σχέση με το αντίστοιχο 3D scan του DTU αλλά και η ομοιότητα της σκηνής με την αντίστοιχη φωτογραφία επίβλεψης όταν γίνεται αξιολόγηση και του πεδίου φωτισμού στα πλαίσια της εμφάνισης. Άλλα, ας ξεκινήσουμε με τις μετρικές αξιολόγησης σε επόπεδο φωτογραφίας.

5.4 Μετρικές Αξιολόγησης Αλγορίθμων

5.4.1 PSNR - Λόγος σήματος κορυφής προς θόρυβο

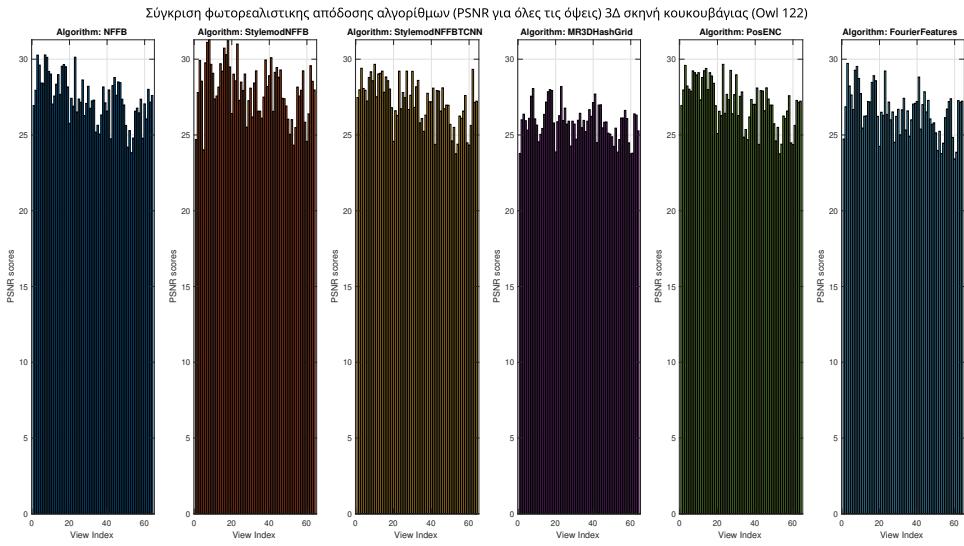
Για δεδομένη 3D σκηνή, αυτή της κουκουβάγιας, φαίνεται πως στο παρακάτω διάγραμμα όπου συγκρίνεται η φωτορεαλιστική απόδοση των διαφορετικών αλγορίθμων για πλήθος θέσεων κάμερας δηλαδή όψεων. Το PSNR ορίζεται πιο εύκολα μέσω του μέσου τετραγωνικού σφάλματος (MSE). Δεδομένης μιας μονοχρωματικής εικόνας I με μέγεθος $m \times n$ και της θορυβώδους προσέγγισής της K , το MSE ορίζεται ως

$$MSE = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} [I(i,j) - K(i,j)]^2.$$

To PSNR (σε dB) ορίζεται ως

$$PSNR = 10 \cdot \log_{10} \left(\frac{MAXI^2}{MSE} \right) = 20 \cdot \log_{10} \left(\frac{MAXI}{\sqrt{MSE}} \right) = 20 \cdot \log_{10}(MAXI) - 10 \cdot \log_{10}(MSE).$$

Εδώ, το MAXI είναι η μέγιστη δυνατή τιμή εικονοστοιχείου της εικόνας. Όταν τα εικονοστοιχεία αναπαρίστανται χρησιμοποιώντας 8 bits ανά δείγμα, αυτή είναι η τιμή 255. Πιο γενικά, όταν τα δείγματα αναπαρίστανται χρησιμοποιώντας γραμμικό PCM με B bits ανά δείγμα, το MAXI είναι $2^B - 1$.



Σχήμα 5.3: Σηματοθορυβική σύγκριση αλγορίθμων κωδικοποίησης στην σκηνή της κουκουβάγιας

Η μετρική PSNR δείχνει πόσο υψηλής ποιότητας είναι η 3D ανακατασκευή στο επίπεδο της φωτογραφίας του μοντέλου. Μικρές μεταβολές μπορεί να αντιστοιχούν σε πολλαπλάσιο ποσό θορύβου φωτογραφίας μιας και η κλίμακα είναι λογαριθμική. Συγκεκριμένα για 1 μονάδα αύξηση έχουμε 10 φορές λιγότερο θόρυβο. Έτσι αυτό που παρατηρείται στην ανακατασκευή της σκηνής της κουκουβάγιας είναι ότι τα μοντέλα φίλτρων πλεγμάτων Fourier (*Fourier Filter Banks*) είναι σε θέση να μειώσουν το θόρυβο σε όλες τις όψεις αρκετά. Ταυτόχρονα το hashgrid δεν μπορεί να συλλάβει υψηλοσυχνοτικό περιεχόμενο από μόνο του γιατί το πλήθος των επιπέδων ανάλυσης που χρησιμοποιείται είναι μικρό. Ταυτόχρονα η συχνοτική κωδικοποίηση με μπορεί να κάνει καλύτερη δουλειά για λίγα επίπεδα συχνοτήτων αλλά παρουσιάζει μεγάλη διακύμανση στις τιμές και απαιτεί το *Hash Grid Encoding* για εντοπισμό των χαρακτηριστικών περιοχών. Τέλος, τα συγχωνευμένο δίκτυο θυρίδων (*StylemodNFFBTNN*) φαίνεται να αποτελεί μια ιδανική λύση δεδομένου μικρότερου χρόνου εκπαίδευσης και χρησιμοποιούν την συνδυασμένη μέθοδο των φίλτρων πλεγμάτων Fourier.

PSNR Metric	DTU Scenes			
Embedding Algorithm for IDR Input	Owl[122]	Rabbit[110]	Skull[65]	Buda[114]
Positional Encoding	27.15	22.7	24.58	26.09
Fourier Features	26.51	22.21	22.07	23.23
Multi-Resolution Hash Grid	25.85	22.73	22.74	21.81
Multi-Resolution Hash Grid (<i>TinyCudaNN</i>)	27.26	22.33		
NFFB	27.4	21.43	21.94	23.35
StyleMod NFFB	28.04	21.97	22.96	23.96
StyleModNFFB (<i>TinyCudaNN</i>)	27.57	20.57	23.48	24.55

Πίνακας 5.10: Αποτελέσματα αξιολόγησης με τη μετρική PSNR

5.4.2 SSIM(Structure Similarity Index Measure) - Μετρική Δείκτη Δομικής Ομοιότητας Εικόνων

Η μετρική δείκτη δομικής ομοιότητας (SSIM) είναι μια μέθοδος πρόβλεψης της αντιλαμβανόμενης ποιότητας ψηφιακών τηλεοπτικών και κινηματογραφικών εικόνων, καθώς και άλλων ειδών ψηφιακών εικόνων και

βίντεο. Το SSIM είναι ένα μοντέλο βασισμένο στην αντίληψη που θεωρεί την υποβάθμιση της εικόνας ως αντιληπτή μεταβολή της δομικής πληροφορίας, ενώ ενσωματώνει επίσης σημαντικά αντιληπτικά φαινόμενα, συμπεριλαμβανομένων τόσο των όρων κάλυψης φωτεινότητας όσο και των όρων κάλυψης αντίθεσης. Η διαφορά με άλλες τεχνικές όπως το MSE ή το PSNR είναι ότι οι προσεγγίσεις αυτές εκτιμούν τα απόλυτα σφάλματα. Η δομική πληροφορία είναι η ιδέα ότι τα εικονοστοιχεία έχουν ισχυρές αλληλεξαρτήσεις, ιδίως όταν βρίσκονται σε χωρική εγγύτητα.

SSIM Metric	DTU Scenes			
	Owl[122]	Rabbit[110]	Skull[65]	Buda[114]
Embedding Algorithm for IDR Input	0.96	0.93	0.95	0.91
Positional Encoding	0.95	0.93	0.94	0.88
Fourier Features	0.95	0.93	0.95	0.86
Multi-Resolution Hash Grid	0.95	0.93	0.95	0.88
Multi-Resolution Hash Grid (<i>TinyCudaNN</i>)	0.96	0.94		
NFFB	0.96	0.93	0.94	0.88
StyleMod NFFB	0.96	0.93	0.94	0.89
StyleModNFFB (<i>TinyCudaNN</i>)	0.96	0.92	0.96	0.88

Πίνακας 5.11: Αποτελέσματα αξιολόγησης με τη μετρική SSIM

5.4.3 LPIPS(Learned Perceptual Image Patch Similarity)- Αντιληπτική Ομοιότητα Μπαλωμάτων Εικόνας σε μοντέλα αναγνώρισης εικόνας

Η συγκεκριμένη μετρική είναι μια μέθοδος αξιολόγησης εικόνων που πλέον γίνεται από νευρωνικά δίκτυα. Στην συγκεκριμένη περίπτωση χρησιμοποιείται το Alexnet (προ εκπαίδευμένο μοντέλο αναγνώρισης εικόνας) για να δείξει πόσο διαφέρουν οι εικόνες μεταξύ τους. Όσο μικρότερη είναι η τιμή το LPIPS σημαίνει τόσο κοντά είναι όχι μόνο σε επίπεδο χρώματος(τιμή [r, g, b]) αλλά και σε αυτό που αναπαρίσταται στις δύο εικόνες.

LPIPS Metric	DTU Scenes			
	Owl[122]	Rabbit[110]	Skull[65]	Buda[114]
Embedding Algorithm for IDR Input	0.08	0.15	0.09	0.19
Positional Encoding	0.25	0.16	0.11	0.10
Fourier Features	0.08	0.14	0.1	0.19
Multi-Resolution Hash Grid	0.06	0.11		
Multi-Resolution Hash Grid (<i>TinyCudaNN</i>)	0.09	0.03	0.1	0.22
NFFB	0.08	0.16	0.1	0.2
StyleMod NFFB	0.07	0.15	0.08	0.16
StyleModNFFB (<i>TinyCudaNN</i>)				

Πίνακας 5.12: Αποτελέσματα αξιολόγησης με τη μετρική LPIPS

5.5 Μετρικές Αξιολόγησης Γεωμετρικής Ομοιότητας

Για την μέθοδο αξιολόγησης της γεωμετρίας των ανακατασκευασμένων μοντέλων, χρησιμοποιήθηκε η μετρική *chamfer distance* η οποία μετράει με την μέθοδο των πλησιέστερων γειτόνων πόσο κοντά είναι μεταξύ τους δύο νέφη σημείων δύο τρισδιάστατων αναπαραστάσεων.

Ορισμός 5.5.1. Απόσταση Chamfer Απόσταση Chamfer ορίζεται ως το άθροισμα των θετικών αποστάσεων των σημείων της 3D ανακατασκευασμένης επιφάνειας με την πραγματική επιφάνεια. Ο υπολογισμός της γίνεται επιλέ-

γοντας το πλησιέστερο γειτονικό σημείο της επιφάνειας με το αντίστοιχο σημείο της άλλης αποτελεί το άθροισμα των αποστάσεων όλων των σημείων των 3D μοντέλων όπως δίνεται από τον παρακάτω τύπο:

$$d_{CD}(S_1, S_2) = \sum_{x \in S_1} \min_{y \in S_2} \|x - y\|_2^2 + \sum_{y \in S_2} \min_{x \in S_1} \|x - y\|_2^2 \quad (5.1)$$

5.5.1 Παρουσίαση Γεωμετρίας Μοντέλων

Στα πλαίσια της οπτικής αξιολόγησης της γεωμετρίας των σχημάτων παρουσιάζονται τα αρχεία των πλεγμάτων πολυγώνων που τυπώνει ο αλγόριθμος *Marching Cubes*. Η ανάλυση με την οποία έτρεξε ο αλγόριθμος στην πληθώρα των περιπτώσεων είναι 400. Ωστόσο στο *FourierNTK* δίκτυο, ο αλγόριθμος έτρεξε σε 512 ανάλυση που σημαίνει ότι ενδέχεται να παρουσιάζονται καλύτερα αποτελέσματα.⁵

⁵ Δεν επετεύχθη να τρέξει ο αλγόριθμος για όλα τα μοντέλα σε ανάλυση 512 λόγω απαιτήσεων μνήμης GPU

Algorithm Name	Scan 65	Scan 110	Scan 114	Scan 122
Positional Encoding				
FourierNTK				
MR HashGrid3D				
NFFB				
Stylemod NFFB				
Stylemod NFFB TCNN				
HashGrid3D TCNN				

Πίνακας 5.13: Πίνακας Γεωμετρίας 3D μοντέλων

Ενδιαφέρον παρουσιάζει και η κρυφή όψη:

Algorithm Name	Scan 65	Scan 110	Scan 114	Scan 122
Positional Encoding				
Stylemod NFFB				
StylemodNFFB TCNN				

Πίνακας 5.14: Πίνακας Γεωμετρίας Κρυφής Όψης 3D μοντέλων - Έλεγχος Γενίκευσης

Τα παραπάνω αποτελούν τις γεωμετρίες που εξάγονται από το πεδίο *SDF* με χρήση του αλγορίθμου *Marching Cubes*. Παρατηρείται, ότι τα δίκτυα υψηλοσυχνοτικής κωδικοποίησης πετυχαίνουν μεγαλύτερη λεπτομέρεια και γενικεύουν καλύτερα στην πίσω όψη που δεν υπάρχουν δεδομένα. Επίσης καταφέρνουν να αποτυπώσουν περιοχές που παρουσιάζουν μεγάλη μεταβολή της παραγώγου δηλαδή τρύπες στα μοντέλα κάτι το οποίο το IDR χωρίς το δίκτυο κωδικοποίησης δεν μπορεί να μάθει.

5.5.2 Πίνακες Απόστασης Chamfer

Η αξιολόγηση ολοκληρώνεται παρουσιάζοντας την μετρική αξιολόγησης της γεωμετρίας των 3D μοντέλων. Δίνονται, οι συνολικές αποστάσεις *chamfer* μεταξύ των ανακατασκευασμένων 3D μοντέλων με τα δεδομένα γεωμετρίας που δίνει το DTU (STL) καθώς και με 3D μοντέλα που εξάγονται από άλλους αλγορίθμους Multi-Vision Stereo οποίοι θεωρούνται ότι παράγουν αξιόπιστα 3D συμπαγή σχήματα. Ο αλγόριθμος που χρησιμοποιήθηκε βασίζεται στον κώδικα που δίνει το *DTU* για την αξιολόγηση [85]. Παρόλα αυτά δεν είναι ο κώδικα σε Matlab που παρέχεται στο MVS DTU αποθετήριο, αλλά είναι δημοσιευμένος (<https://github.com/jzhangbs/DTUEval-Python>) κώδικας αξιολόγησης γραμμένος σε python και χρησιμοποιεί την *sklearn* για τα μοντέλα πλησιέστερου γείτονα (s2d, d2s eval data to scan, scan to data). Δημιουργεί το νέφος σημείων εφαρμόζοντας υποδειγματοληψία των συμπαγών δομών και υπολογίζει την απόσταση των γειτόνων μέσω αλγορίθμου δένδρου πλησιέστερων γειτόνων διαμερισμού του χώρου (*kd-tree*).

Μέθοδος Εξαγωγής GT Mesh	Chamfer Distance(mm) ανά scan [65, 110, 114, 122]															
	STL DTU scans				Camp MVS algorithm				Furu MVS algorithm				Tola MVS algorithm			
	0.971	1.117	0.5	0.58	1.27	0.84	0.49	0.55	1.70	1.20	0.74	0.83	2.36	0.96	0.59	1.02
Positional Encoding	1.13	1.32	0.63	0.601	1.22	0.99	0.491	0.52	1.72	1.31	0.80	0.83	2.13	1.11	0.67	1.03
NFFB	1.00	1.36	0.43	0.58	1.29	1.05	0.413	0.523	1.75	1.34	0.64	0.81	2.24	1.15	0.46	1.00
StylemodNFFB	0.957	1.50	0.45	0.543	1.2	1.03	0.42	0.52	1.65	1.44	0.65	0.79	2.14	1.29	0.47	0.99
StylemodNFFB (TinyCudaNN)	1.11	2.44	1.66	0.59	1.48	2.02	1.519	0.54	1.99	2.39	1.70	0.81	2.44	2.46	1.81	0.98

Πίνακας 5.15: Πίνακας Μετρικών Μέσης Απόστασης Chamfer

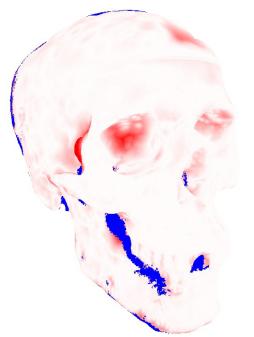
Παρατηρώντας τον παραπάνω πίνακα, βλέπουμε πως διαφορετικοί αλγόριθμοι είναι πιο καλοί ανά κωδικοποίηση. Κατά πλειοψηφία βέβαια οι καλύτερες ανακατασκευές έχουν γίνει με το *StylemodNFFB*. Η ασυνέπεια που παρατηρείται οφείλεται στο δύσκολο περιεχόμενο φωτισμού των σκηνών 110 και 114 που μπερδεύουν το πεδίο γεωμετρίας. Ταυτόχρονα η εξαγωγή των πλεγμάτων από τις έμμεσες αναπαραστάσεις στο *Positional Encoding* λόγω έγινε με λίγο μεγαλύτερη ανάλυση επομένως έχει μεγαλύτερη λεπτομέρεια χωρίς να είναι απαραίτητα καλύτερη ανακατασκευή.

Πάνω σε αυτό σημαντική σημείωση για τον αναγνώστη είναι πως λόγω περιορισμών μνήμης στη *GPU(8Gb Vram)* που χρησιμοποιήθηκε για τα πειράματα, η εξαγωγή των πλεγμάτων από τις έμμεσες αναπαραστάσεις έχει γίνει σε χαμηλή ανάλυση(ανάλυση αλγόριθμου *Marching Cubes* βλ.7). Αυτό έχει άμεσο αντίκτυπο στα αποτελέσματα αφού παρέχεται μικρότερο επίπεδο ανάλυσης από ότι δύναται να δοθεί και συνεπώς τα αποτελέσματα της μετρικής *Chamfer* είναι χειρότερα από όσο θα μπορούσαν να είναι.⁶

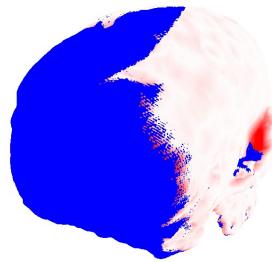
⁶η κωδικοποίηση FourierFeatures παραλείπεται γιατί η εξαγωγή των πλεγμάτων έχει γίνει με σημαντικά μεγαλύτερη ανάλυση στην υπολογιστική συστοιχία του ΑΠΘ και τα αποτελέσματα είναι πολύ καλύτερα αλλά δεν μπορούν να συγκριθούν με τα υπόλοιπα.

Οπτική Παρουσίαση Chamfer Απόστασης

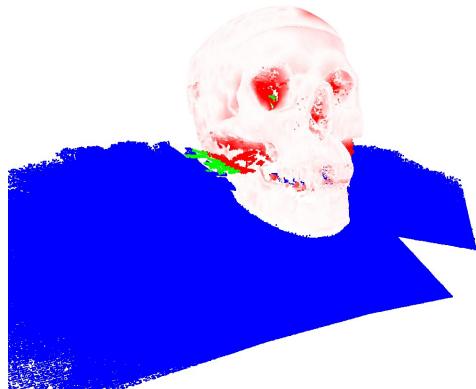
Ενδεικτικά δίνεται το οπτικό αποτέλεσμα που παράγει ο αλγόριθμος αξιολόγησης πλεγμάτων για το δίκτυο StylemodNFFB. Το οπτικό αποτέλεσμα που παράγεται είναι και αυτό νέφος σημείων και των δύο σκηνών έτσι δίνεται φωτογραφία. Οι εικόνες είναι 2 μια που δείχνει την απόσταση των παραγόμενων δεδομένων από το τρισδιάστατο scan(d2s) και μια το ανάποδο(s2d).



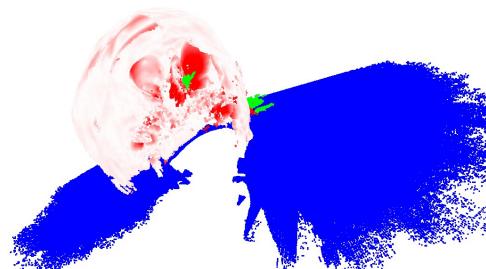
(1) d2s front



(2) d2s back



(3) s2d front



(4) s2d back

Σχήμα 5.4: Οπτική αποτύπωση Chamfer απόστασης

Τα κόκκινα σημεία είναι μικρές αποστάσεις στα όρια του στατιστικού λάθους. Τα μπλε σημεία είναι σημεία που αποτυπώνει η ανακατασκευή, ενώ δεν έχει το point cloud της σάρωσης στην d2s περίπτωση. Αντίθετα στην άλλη περίπτωση s2d είναι σημεία που σαρώνονται του αντικειμένου από τις κλασσικές μεθόδους σάρωσης και δεν μας ενδιαφέρουν. Τέλος τα πράσινα σημεία είναι σημεία που απέχουν μεταξύ ανακατασκευής και σάρωσης τα οποία λανθασμένα βγάζουν οι κλασσικές μέθοδοι σάρωσης MVS.

6

Συμπεράσματα & Μελλοντικές επεκτάσεις

6.1 Συμβολή της Εργασίας - Συμπεράσματα

Η παρούσα εργασία καταφέρνει να εξετάσει μια πληθώρα μεθόδων κωδικοποίησης σημείων και διανυσμάτων όψης, με σκοπό την αξιόπιστη τρισδιάστατη ανακατασκευή. Έτσι τα δίκτυα αυτά μπορούν να αποτελέσουν μια λύση όσον αφορά το πεδίο των *Vision Transformers* και συγκεκριμένα των *3D Vision coordinate based transformers*.

Φυσικά η έρευνα χρειάζεται επέκταση και μετατροπή σε κώδικα χαμηλότερου επιπέδου με σκοπό την αποφυγή προβλημάτων διαρροής μνήμης που υπάρχουν. Τα αποτελέσματα ωστόσο εξακολουθούν να ξεπερνούν σε επίπεδο φωτογραφίας τα αποτελέσματα του IDR ειδικά σε λεπτομέρειες όπως για παράδειγμα στην απούπωσης της υφής του χρυσού κούνελου(σκηνή 110), η οποία αποτελεί από τις πιο δύσκολες σκηνές όσον αφορά την *3D* ανακατασκευή.

Σαφώς δεν έχουν εξεταστεί όλες οι περιπτώσεις δεδομένων και ίσως θα έπρεπε το δίκτυο να μην εκπαιδεύεται σε τόσο μεγάλη κλίμακα αφού βλέπουμε στα τελευταία μοντέλα *StylemodNFFB*, *StylemodNFFB_TCNN* πως ήδη από τις πρώτες 100 εποχές έχει αποτυπωθεί λεπτομέρεια.

Τέλος παρατηρούνται *artifacts* σε περιπτώσεις που κωδικοποιείται σύνθετο περιεχόμενο φωτισμού με απλό περιεχόμενο φωτισμού στα πλήρως συγχωνευμένα δίκτυα (*TCNN*) του *StylemodNFFB* για μεγάλη παράμετρο μεγέθους πίνακα κατακερματισμού και επιπέδων κατακερματισμού. Ίσως θα μπορούσε το ίδιο το δίκτυο *NFFB* να αποτελέσει το έμμεσο πεδίο διαφορίσμης γεωμετρίας ώστε να μην κάνουν μεγάλη οπισθοδάδοση τα σφάλματα και υπάρχει θέμα τις κλίσεις σφαλμάτων όπως το σφάλμα μάσκας που είναι ήδη μικρό.¹

6.2 Μελλοντικές Βελτιώσεις/Επεκτάσεις

Η προσπάθεια κωδικοποίησης υψηλοσυχνοτικού περιεχομένου με χρήση τρισδιάστατων δικτύων κωδικοποίησης για το μέρος της έρευνας που αφορά την ανακατασκευή τρισδιάστατων επιφανειών μέσω έμμεσων

¹Σε αυτές τις περιπτώσεις παρατηρείται να έχει μπερδευτεί το δίκτυο γεωμετρίας με το δίκτυο εμφάνισης και ενώ το χρώμα είναι καλό και το σφάλμα μάσκας μικρό, να έχουμε μέρη γεωμετρίας που δεν είναι πάνω στο μοντέλο

νευρωνικών αναπαραστάσεων, παρείχε σαφή βελτίωση τόσο στην ταχύτητα εκπαίδευσης των πεδίων προσημασμένης απόστασης αλλά και στην αξιοπιστία της ανακατασκευής υψηλούσυχνοτικών περιοχών.

Βέβαια η πραγματική δύναμη του *Multi-Resolution HashGrid Encoder* και κατά συνέπεια και του *NFFB*, *StylemodNFFB* θα μπορούσε να φανεί στην ανακατασκευή επιφανειών μεγάλων εκτάσεων όπου η λεπτομέρειες είναι μικρής κλίμακας και πολύ περισσότερες. Δυστυχώς, δεν υπήρχαν επαρκή δεδομένα για να γίνει τέτοια έρευνα μια και το δίκτυο απαιτεί ως είσοδο έστω παραμέτρους κάμερας έστω και εκτιμήσεις αυτών.

Δεδομένου του παραπάνω μια μελλοντική επέκταση θα ήταν η εφαρμογή του τρισδιάστατου κωδικοποιητή συνταγμένων σε ήδη υπάρχοντα δίκτυα ογκομετρικής απόδοσης όπως το *NeRF*[3] και το *NeuS* [4] και παράλληλα αφαίρεση της επίβλεψης μέσω μασκών εικόνων κατά την ανακατασκευή που το παρόν δίκτυο χρειάζεται για να συγκλίνει ομαλά. Η διαδικασία επεκτείνεται με εφαρμογή των κωδικοποιήσεων σε εργασίες που κάνουν χρήση μονοσκοπικού βίντεο για συλλογή δεδομένων στα οποία εκπαιδεύεται το δίκτυο σε λίγα λεπτά.

Τέλος μιας και η τάση της εποχής είναι τα παραμετρικά νευρωνικά πεδία (*Conditional Neural Fields*) θα μπορούσε να γίνει εφαρμογή των μεθόδων που αναφέρθηκαν σε δίκτυα τα οποία χρησιμοποιούν έμμεσες αναπαραστάσεις ως *latent code* (διάνυσμα εξόδου) σε παραγωγικά αντιπαραθετικά δίκτυα (*GANs*) τα οποία θα μπορούν να παράγουν τυχαία μορφή *3D* μοντέλων με τυχαίο χρώμα κάτι το οποίο θα οδηγούσε πιθανόν και σε βελτιωμένες μορφές απόδοσης παραμορφώσεων σε μηχανές γραφικών.

7

Παράρτημα

Παράρτημα Α: Επεκτάσεις Θεωρίας

Πολυώνυμα Bernstein

Για κάθε φυσικό αριθμό $n \in \mathbb{N}$, η πολυωνιμική συνάρτηση $\phi(\tau)$ ή χρησιμοποιώντας το σύμβολο $B(\tau)$ για τα πολυώνυμα Bernstein, δίνεται από την μορφή:

$$B_k^n(\tau) := \binom{n}{k} \tau^k (1 - \tau)^{n-k}, \forall k \in [0, n] \quad (1)$$

Προσέγγιση Bézier - Bernstein

Τα πολυώνυμα Bernstein $B_k^n(\tau)(\tau; \vec{\alpha})$, όπου το διάνυσμα $\vec{\alpha} = (\alpha_1, \alpha_2, \dots, \alpha_n)$ είναι μια ακολουθία παραμέτρων, εισάγουν την έννοια των βαθμών ελευθερίας μπορούν να προσεγγίσουν με την μέθοδο Bézier καμπύλες και επιφάνειες ως ανάπτυγμα αυτών των παραμετροποιημένων όρων με την παρακάτω μορφή με βάση τις 2.1, 1 :

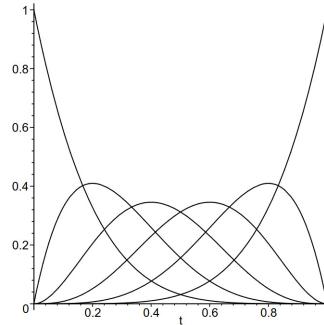
$$\mathbf{P}(t) = \sum_{i=0}^n \mathbf{p}_i B_n^i(t), \quad 0 \leq t \leq 1$$

Συνήθως αυτή με αυτή την μεθοδολογία προσεγγίζονται καμπύλες τμηματικά, δηλαδή σε περιοχές που ελέγχονται από 4 σημεία ελέγχου (n=3).

Πολυώνυμα Spline

Μια άλλη οικογένεια πολυωνιμικών συναρτήσεων βάσης είναι τα πολυώνυμα Bsplines των οποίων ο τύπος για τη βαθμού είναι ο εξής:

$$\phi_{k,0}^{BS}(t) = \begin{cases} 1 & t_k \leq t < t_{k+1} \\ 0 & \text{αλλού} \end{cases}$$



Σχήμα 1: Γενικευμένα Πολυώνυμα Bernstein $5^{ου}$ βαθμού B_k^5 στο $[0,1]$. Πηγή [10]

και απαιτείται η επιλογή $n + m + 2$ διακριτών τιμών της τιμής t έτσι:

$$\phi_{k,m}^{BS}(t) = \frac{t - t_k}{t_{k+m} - t_k} \phi_{k,m-1}^{BS}(t) + \frac{t_{k+m+1} - t}{t_{k+m+1} - t_{k+1}} \phi_{k+1,m-1}^{BS}(t), \quad k = 0, \dots, r$$

επομένως ισχύει η ακόλουθη έκφραση:

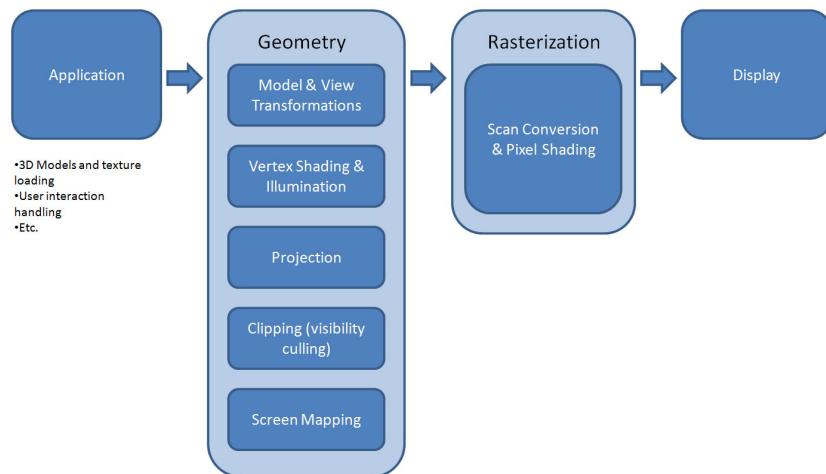
$$\phi_{k,m}^{BS}(t_k) = \phi_{k,m}^{BS}(t_{k+m+1}) = 0, \quad n \geq 1.$$

για αυτό οι τιμές t_k είναι αποκαλούνται και δεσμοί των καμπυλών. Με την προσέγγιση Bezier παίρνουμε ανάλογα αναπτύγματα των αναπτυγμάτων Bezier Bernstein που είναι σε θέση να γενικεύσουν την ικανότητα αναπαράστασης οδηγώντας σε γενικευμένες καμπύλες Bspline.

Μοντέλα Προβολής | Κάμερα και Θέαση

Ένα σχετικά αφαιρετικό αλλά περιεκτικό διάγραμμα που παρουσιάζει την πλήρη γραμμή παραγωγής γραφικών φαίνεται παρακάτω 2:

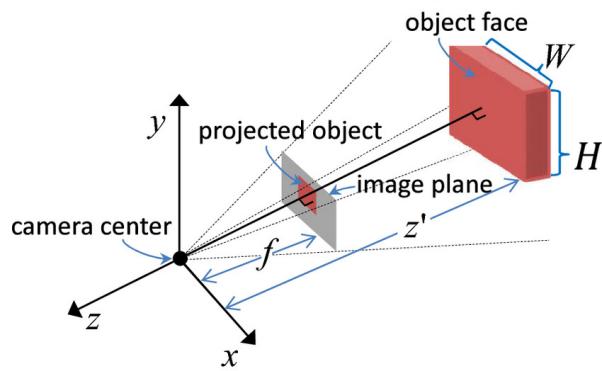
Real-Time Graphics Pipeline



Σχήμα 2: Βασική Γραμμή Παραγωγής Γραφικών

Το κομμάτι των μετασχηματισμών αναφέρεται σε όλους του μετασχηματισμούς περιστροφής, και τους μετασχηματισμούς μετατόπισης αλλά και κλίμακας που εφαρμόζονται στα γραφικά πριν την προβολή τους.

Συγκεκριμένα όσον αφορά την εργασία γίνεται εκτενής χρήση μεθόδων απόδοσης που βασίζονται στην ιχνηλάτιση ακτίνας καθώς και χρήση εικόνων για την εκτίμηση σφάλματος του δικτύου οπότε είναι επιτακτικό να γνωρίζουμε την βασική γραμμή παραγωγής σκηνών μέσω μηχανών γραφικών Στα πλαίσια της γραφικής υπολογιστών όταν γίνεται λόγος για απόδοση γραφικών αυτό αφορά και την προβολή τους σε κάποιο μέσο αναπαραγωγής. Τα ανθρώπινα μάτια δεν είναι σε θέση να δουν τρισδιάστατη πληροφορία κάτι που πηγάζει από τον τρόπο λειτουργίας τους. Συγκεκριμένα αυτό που αποτυπώνεται σε κάθε μάτι είναι ένα ανεστραμμένο είδωλο της 3D πληροφορίας ως μια εικόνα που ερεθίζει το οπτικό μας νεύρο. Η αντίληψη της τρισδιάστατης υπόστασης του χώρου είναι μια εμπειρική ικανότητα που οφείλεται στην απόσταση των ματιών μας (τεχνική που ακολουθείται και από κάμερες βάθους) αλλά κυρίως από άλλες αισθήσεις μας. Η διαδικασία παραγωγής



Σχήμα 3: Σύστημα προοπτικής Προβολής [86]

μιας απεικόνισης γραφικών μιας τρισδιάστατης σκηνής είναι κάπως ανάλογη με τις διαδικασίες που σχετίζονται με την λήψη φωτογραφίας. Επί της ουσίας εισάγουμε ένα μοντέλο εικονικής κάμερας που παίζει τον ρόλο των ματιών του παρατηρητή και έχουμε δύο είδη προβολών. Την ορθογραφική η οποία δεν κάνει χρήση του πόσο κοντά είναι το αντικείμενο στην κάμερα και την προοπτική που δίνει την αίσθηση της μείωσης βάθους καθώς το αντικείμενο απομακρύνεται από την θέση της κάμερας. Το συνολικό μοντέλο προβολής πάνω στο πέτασμα της κάμερας όπου φαίνονται και σημαντικές παράμετροι της κάμερας (εσωτερικές, εξωτερικές ή όπως αναφέρεται στην διεθνή βιβλιογραφία *intristic, extrinsic*.

$$s \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

2D Image Coordinates Intrinsic properties (Optical Centre, scaling) Extrinsic properties (Camera Rotation and translation) 3D World Coordinates

Σχήμα 4: εξωτερικές και εξωτερικές παράμετροι κάμερας σε μαθηματική μορφή συστήματος προβολής

Ο χώρος συντεταγμένων UV είναι ο χώρος των εικονοστοιχείων της εικόνας. Οι εξωτερικές παράμετροι αφορούν τον προσανατολισμό και την θέση της κάμερας ενώ οι εξωτερικές παράμετροι αφορούν το μήκος του φακού δηλαδή το ποσό της προοπτικής κλιμάκωσης καθώς και το οπτικό κέντρο. Όταν γίνεται λόγος για φωτογραφική ομοιότητα μιλάμε για το πόσο ίδιες είναι οι σκηνές υπό το πρίσμα της απόστασης σε τιμή φωτεινότητας (σε όλα τα κανάλια πχ. [R, G, B]) των εικονοστοιχείων στο UV επίπεδο.

Τρόπος αναπαράστασης και αρχικοποίησης καμερών στην εργασία

Μια κάμερα αναπαριστάται με ένα σύνολο διανυσμάτων $\mathcal{C} = (\mathbf{q}, c, \mathbf{K})$ όπου το $\mathbf{q} \in \mathbb{R}^4$ είναι το *quaternion* διάνυσμα¹, το οποίο αναπαριστά την πόζα ή στροφή της κάμερας, $\mathbf{c} \in \mathbb{R}^3$ αναπαριστά την θέση της κάμερας, και $\mathbf{K} \in \mathbb{R}^{3 \times 3}$ αναπαριστά τον πίνακα εσωτερικών παραμέτρων της κάμερας.

Οι παράμετροι της κάμερας είναι οι $\tau = (\mathcal{C}_1, \dots, \mathcal{C}_N)$, όπου N ο αριθμός των καμερών (και ταυτοχρόνως και εικόνων όψης του αντικειμένου). Έστω $\mathbf{Q}(\mathbf{q}) \in \mathbb{R}^{3 \times 3}$ συμβολίζει τον πίνακα στροφής που αντιστοιχεί στο *quaternion* \mathbf{q} . Τότε, για την κάμερα $\mathcal{C}_i, i \in [N]$, και για pixel u έχουμε:

$$c_u(\tau) = c_i, \quad (2)$$

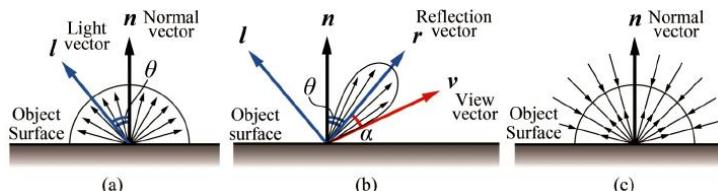
$$\mathbf{w}_u(\tau) = \frac{1}{\|\mathbf{K}_i^{-1}u\|_2} \mathbf{Q}(\mathbf{q}_i) \mathbf{K}_i^{-1}u, \quad (3)$$

όπου το $\mathbf{u} = (u_x, u_y, 1)^T$ είναι το pixel u σε ομογενής συντεταγμένες.

Στα πειράματα που πραγματοποιήθηκαν, σε περιπτώσεις που οι παράμετροι κάμερας για μια σκηνή που αντιστοιχεί σε ένα 3D αντικείμενο δεν υπάρχουν, δημιουργούνται σχετικές μετακινήσεις μεταξύ ζευγαριών κάμερα χρησιμοποιώντας τον αλγόριθμο *SIFT feature matching* [87] και ένας ανθεκτικός στατιστικά πίνακας εκτίμησης (RANSAC), ο οποίος ακολουθείται από την αποσύνθεση των σχετικών περιστροφών και μεταπόσεων μεταξύ των καμερών [88]. Οι σχετικές μετακινήσεις ως είσοδοι στην γραμμική μέθοδο κανονικοποίησης που χρησιμοποιείται [74] παράγει κάμερες με θόρυβο. Θεωρούνται γνωστοί οι παράμετροι της κάμερας όπως συχνά θεωρείται και το SFM [72].

Μοντέλο Phong Φωτισμού - P-Universal Renderer

Στην θεωρία γραφικής υπολογιστών ο φωτισμός είναι ένα μεγάλο κεφάλαιο. Η μεθοδολογία που ακολουθείται είναι η συνολική εκτίμηση όλων των μορφών του φωτισμού και ο υπολογισμός της επίδρασης που έχουν στο χρώμα. Οι μορφές φωτισμού που υπάρχουν επηρεάζονται ιδιαιτέρως από την υφή της επιφάνειας δηλαδή την κατανομή των κανονικών διανυσμάτων στα πιο σύνθετα μοντέλα φωτισμού όπως η κατοπτρική ανάκλαση και ο περιορισμός του πεδίου φωτός που επιβάλουν οι συντελεστές Phong.



Σχήμα 5: Διανύσματα, Διάχυτη Ανάκλαση, Κατοπτρικό Ανάκλαση, Μοντέλο Ανάκλασης Phong, Πηγή [89]

Όπως φαίνεται και στο σχήμα 5, ανάλογα το είδος του φωτισμού που μελετάμε μας ενδιαφέρουν διαφορετικά διανύσματα είτε πρόσπτωσης(φως διάχυσης) είτε ανάκλασης(κατοπτρικής ανάκλαση) στην επιφάνεια. Σε όλα τα μοντέλα όμως είναι απαραίτητη η επίγνωση του κανονικού διανύσματος στην επιφάνεια. Το μοντέλο Phong συνδυάζει όλα τα μοντέλα φωτισμού περιλαμβάνοντας και παραμέτρους για την υφή των υλικών η οποία. Συνολικά η περιγραφή του περιορισμένου μοντέλου φωτισμού για δοθέν σημείο πρόσπτωσης \hat{x} και διάνυσμα ανάκλασης \mathbf{w}^o δίνεται από την εξίσωση:

$$L(\hat{x}, \mathbf{w}^o) = k_d O_d I_\alpha + k_d O_d I_d (\hat{n} \cdot \frac{\mathbf{I} - \hat{x}}{\|\mathbf{I} - \hat{x}\|}) + k_s O_s I_d (\hat{r} \cdot \frac{\mathbf{I} - \hat{x}}{\|\mathbf{I} - \hat{x}\|})^{n_{\text{phong}}}, \quad (4)$$

¹τα *quaternions* ή τετραδρόνια είναι αναπαραστάσεις διανυσμάτων τεσσάρων διαστάσεων που επεκτείνουν το σύστημα των μιγαδικών αριθμών και χρησιμοποιούνται συχνά σε εφαρμογές ρομποτικής και γραφικής σε ομογενείς μετασχηματισμούς (στροφές) όπου δεν ισχύουν αντιμεταθετικές ιδιότητες

με α να είναι το μήκος κύματος διάχυτης ακτινοβολίας, k_d, k_s συντελεστές διάχυτης και κατοπτρικής ανάκλασης, I_a, I_d , το διάχυτο φως το φως σημειακής πηγής(χρώμα πηγής) και O_d, O_s , είναι τα χρώματα της επιφάνειας. Η θέση της σημειακής πηγής καθορίζεται από το $\mathbf{I} \in \mathbb{R}^3$ και $\mathbf{w}^0 = -\mathbf{w}$ δηλαδή το διάνυσμα ανάποδο από το διάνυσμα όψης ή ν όπως φαίνεται στο σχήμα 5. Το διάνυσμα ανάκλασης \hat{r} δίνεται από τον νόμο ανάκλασης $\hat{r} = -(I - 2\hat{n}\hat{n}^T)\mathbf{w}^0$ σε συνάρτηση του κανονικού διανύσματος. Ο συντελεστής n_{phong} είναι ο συντελεστής του κατοπτρικού μοντέλου Phong που αλλάζει τα κατοπτρικά φαινόμενα στο χρώμα.

Η οικογένεια αυτή δεν περιλαμβάνει όλα τα στοιχεία φωτισμού, παραδείγματος χάρη δεν υπολογίζει σκιές από το ίδιο το σώμα που φωτίζεται, αφού για δεύτερη ή υψηλότερη παράγωγο του L η γεωμετρία δεν παίζει ρόλο στον υπολογισμό του χρώματος.

Η έρευνα χρησιμοποιεί μια συνάρτηση για να περιγράψει όλα αυτά τα μοντέλα φωτισμού την $\mathcal{M}(x, n, \mathbf{w}; \gamma)$ όπου M είναι ένα MLP δίκτυο. Η συνεχή αυτή συνάρτηση που προσεγγίζει το δίκτυο μπορεί να περιγράψει πληροφορία για πλήθος φωτισμών μεγιστοποιώντας το διάνυσμα εξόδου σε συνάρτηση με τις παραμέτρους γ της εμφάνισης. Αυτό βασίζεται στην παρακάτω θεωρία του πως ένα MLP δίκτυο που περιγράφει Level Set πολλαπλότητες μπορεί να αναπαραστήσει οποιαδήποτε τμηματικά συνεχή συνάρτηση μεγιστοποιώντας ένα διάνυσμα εισόδου. Στην συγκεκριμένη περίπτωση το διάνυσμα εισόδου περιλαμβάνει τα κάθετα διανύσματα οπότε αποτυπώνονται κατοπτρικά φαινόμενα. Πρακτικά περιγράφεται η επιφανειακή κατανομή χρώματος ή Texture. Τότε το δίκτυο ονομάζεται P-Universal Renderer.

Παράρτημα B: Αποδείξεις

Θεώρημα [0].0.1. Θεώρημα αναπαράστασης συμπαγών επιφανειών μέσω ελεγχόμενων νευρωνικών επιπέδων Κάθε συμπαγής, όχι αναγκαστικά φραγμένη, τμηματική γραμμική υπερεπιφάνεια $M \subset \mathbb{R}^d$ μπορεί να αναπαρασταθεί με ακρίβεια, ως ένα νευρωνικά διαφορίσμη πολλαπλότητα (τοπολογία, ισομετρική επιφάνεια) S ενός πολυστρωματικού perceptron δικτύου με ReLU συναρτήσεις ενεργοποίησης, $F : \mathbb{R}^d \rightarrow \mathbb{R}$.

Απόδειξη. Έστω $h_i(x) = a_i^T x + b_i = 0 \ i \in [k]$ συμβολίζει την επιφάνεια η οποία υποστηρίζεται από τις έδρες του M όπου τα a_i είναι επιλεγμένα ώστε να είναι τα προς τα έξω κανονικά διανύσματα της επιφάνειας σε αυτές τις έδρες. Εφόσον η M είναι συμπαγής δηλαδή αποτελεί Convex Hull ή κυρτή τοπολογία του χώρου, αποτελεί ένα φράγμα d-διάστατου χώρου P .

Για κάθε $\lambda \in -1, 0, 1^k$, έστω $P_\lambda = \cap_{i \in [k]} \{x | \lambda_i h_i(x) \geq 0\}$. Πιο απλά, P_λ είναι ένας χώρος όπου κάθε τομή κλειστών ημι-χώρων ορίζεται από τα υπερεπίπεδα h_i . Από όλα τα P_λ , μας ενδιαφέρει αυτά που περιέχονται στο P , έτσι ορίζεται $\Lambda = \{\lambda | P_\lambda \subseteq P\}$. Έτσι τώρα, αρκεί να δείξουμε ότι κάθε σημείο στο εσωτερικό μια μεγάλης τοπολογίας βρίσκεται και αυτό αναγκαστικά εσωτερικά κάποιων μικρών τοπολογιών στην συλλογή, δηλαδή $\cup_{\lambda \in \Lambda} P_\lambda^\circ = \mathring{P}$. Έτσι για δεδομένο $x \in \mathring{P}$ υπάρχουν δύο περιπτώσεις στο που μπορεί να βρίσκεται:

Περίπτωση 1: $h_i \neq 0 \forall i \in [k]$. Δεδομένου αυτού το, x δεν βρίσκεται ακριβώς σε κάποιο υπερεπίπεδο. Μπορούμε να πάρουμε τις υπόλοιπες πολυδιάστατες τοπολογίες P_λ οι οποίες περιέχουν εσωτερικά το x θέτοντας: $\lambda_i = sign(h_i(x))$. Σημειώνεται ότι, $\lambda \in \{-1, 1\}^k$, και για αυτές τις τιμές η πολυδιάστατη τοπολογία που αναπαρίσταται λέγεται ελάχιστη (*minimal*). Διατυπώνεται ο ισχυρισμός ότι η ελάχιστη τοπολογία είναι στο εσωτερικό ή εντελώς εξωτερικά στο σύνολο P . Αυτό επιβεβαιώνεται καθώς σε άλλη περίπτωση, η ελάχιστη τοπολογία θα περιείχε δύο σημεία τα οποία βρίσκονται σε διαφορετικές πλευρές κάποιου υπερεπιπέδου, το οποίο δεν συμβαδίζει για τις τιμές του $\lambda \in \{-1, 1\}^k$. Συνεπώς στην περίπτωση μας, ξέρουμε ότι P_λ και P περιέχουν το x ως εσωτερικό σημείο, και αυτό σημαίνει ότι απαραίτητα $P_\lambda \subseteq P$ άρα και $\lambda \in \Lambda$.

Περίπτωση 2: $\exists \{i_1, \dots, i_l\} \subseteq [k]$ τέτοιο ώστε $h_i(x) = 0 \forall i \in \{i_1, \dots, i_l\}$, και $h_i \neq 0 \forall i \in [k] \setminus \{i_1, \dots, i_l\}$. Σε αυτή την περίπτωση, δεν υπάρχει ελάχιστη τοπολογία η οποία να περιέχει το x εσωτερικά της. Έτσι ας υποθέσουμε ότι τις όλες οι ελάχιστες τοπολογίες οι οποίες περιέχουν το x στο όριο τους. Έστω P_u μια τέτοια ελάχιστη τοπολογία πολλών διαστάσεων. Όπως προαναφέρθηκε, το εσωτερικό P_u της είναι είτε εντελώς εντός του συνόλου τροπολογιών P ή εντελώς εκτός, αλλά αφού το x είναι ταυτόχρονα στα όρια της P_u και εντός της P , αναγκαστικά P_u είναι εσωτερικά της P , δηλαδή $P_u \subset P$. Μας ενδιαφέρει το σύνολο όλων αυτών των ελάχιστων τοπο-

λογιών. Σημειώνεται πως για τέτοια ελάχιστη τοπολογία, P_u , αναγκαστικά $u_i = \text{sign}(h_i(x)) \forall i \in [k] \{i_1, \dots, i_l\}$. Για $i \in \{i_1, \dots, i_l\}$, το u_i μπορεί να πάρει τιμές στο $\{1, -1\}$. Έτσι, η ένωση όλων αυτών των ελάχιστων τοπολογιών είναι οι P_λ όπου:

$$\lambda_i = \begin{cases} 0, & \text{if } i \in \{i_1, \dots, i_l\} \\ \text{sign}(h_i(x)), & \text{αλλού} \end{cases}$$

οι οποίες προφανώς περιέχει το σημείο x στο εσωτερικό τους και οι ίδιες περιέχονται στο P (ως ένωση τοπολογιών που περιέχονται στο P), δηλαδή $\lambda \in \Lambda$.

Έχοντας δείξει τα παραπάνω μπορούμε να ορίσουμε μια συνάρτηση η οποία δέχεται θετικές τιμές στο εσωτερικό του συνόλου P , αρνητικές τιμές στο εξωτερικό του P (ή μπορεί και ανάποδα ανάλογα την σύμβαση). Έτσι ορίζεται συνάρτηση πολλαπλότητας:

$$f(x) = \max_{\lambda \in \Lambda} \min_{i \in [k]} \lambda_i h_i(x)$$

Η f , είναι τιμηματικά συνεχής συνάρτηση, συνεπώς, με βάση το θεώρημα μπορεί να κωδικοποιηθεί ως δίκτυο MLP με συναρτήσεις ενεργοποίησης ReLU. Η συνολική ιδέα βασίζεται στην δημιουργία τελεστών \max χρησιμοποιώντας γραμμικά στρώματά και ReLU δηλαδή $\max\{a, b\} = \sigma(a - b)^2 + \sigma(b - a)^2 + (a + b)^2$, όπου $\sigma(x) = \max(0, x)$ η συνάρτηση ReLU. Χρησιμοποιώντας αυτούς τους δυαδικούς τελεστές μπορεί κανείς να δημιουργήσει τον μέγιστο ενός διανύσματος και το ελάχιστο με τον ίδιο τρόπο προσεγγίζοντας οποιαδήποτε τοπολογία.

□

Παράδειγμα Γ: Ενδεικτικοί Αλγόριθμοι που Υλοποιήθηκαν ή Αναφέρονται

Αλγόριθμος Νευρωνικής Κωδικοποίησης Κατακερματισμού Hash - Multi-Resolution HashGrid3D Encoding

Algorithm 3 HashGrid3D-Neural Coordinate Encoding

Ορισμός σταθερών: π_i .

Ορισμός ενός πίνακα πρώτων αριθμών HASH_PRIMES.

procedure HashGrid3D(**d**, **F**, $\log(T)$, N_ℓ)

 Αρχικοποίηση του HashGrid3D με τις διαστάσεις εισόδου, μέγεθος τιμών κατακερματισμού, μέγεθος χάρτη και ανάλυση.

 Δημιουργία πίνακα αναζήτησης για τιμές κατακερματισμού T .

 Αρχικοποίηση μάσκας διασταύρωσης *Binary Mask*.

 Αρχικοποίηση βαρών EmbeddingLayer με ομοιόμορφη κατανομή μικρής τυπικής απόκλισης.

end procedure

procedure HashGrid3Dforward(**x**, υπολογισμός_κλίσης)

$x_i = \lfloor x \rfloor$.

$x_f = x_i - x$.float() # Κλιμάκωση εισόδου στην δοθείσα ανάλυση.

neighbors = bin_mask($x_i, x_i + 1$)

w_ids = bin_mask($1 - x_f, x_f$) # Υπολογισμός γειτόνων, παρεμβολή των δεδομένων των γειτόνων στην δοθείσα ανάλυση.

 Εφαρμογή συνάρτησης hash στα ids των γειτόνων και αντιστοίχηση τους σε βάρη του δικτύου προσοχής.

neig_hash_data = EmbeddingLayer(hash_func(**w_ids**, HASH_PRIMES, T))

 Επέστρεψε $\sum neig_{data} * w_{ids} * w_{ids}^T$ # Παρεμβολή των δεδομένων των γειτόνων.

end procedure

procedure MultiResHashGrid3D(**includeInput**, **d**, **L**, **F**, $\log(T)$, N_{min} , N_{max})

 Αρχικοποίηση MultiResHashGrid3D με παραμέτρους.

 Υπολογισμός $N_\ell := \lfloor N_{min} \cdot b^\ell \rfloor$

 Δημιουργία Πολλαπλής Ανάλυσης Πλεγμάτων με κλήση HashGrid3D.

if **includeInput** **then**

 Αρχικοποίηση FourierFeatureNTK για κωδικοποίηση Βοηθητικής εισόδου.

end if

end procedure

procedure MultiResHashGrid3Dforward(**x**, υπολογισμός_κλίσης)

 Συγχώνευση των πλεγμάτων κωδικοποίησης από κάθε επίπεδο ανάλυσης.

if **includeInput** **then**

 Συγχώνευση FFNTK(**x**), MultiResHashGrid3D(**x**).

end if

 Επιστροφή συγχωνευμένου διανύσματος χαρακτηριστικών

end procedure

Στο *backward pass* του δικτύου επιστρέφονται σφάλματα μέχρι τα βάρη των γειτόνων και ρυθμίζονται οι παράμετροι του *Embedding* στρώματος νευρώνων. Αυτά αντιστοιχούν με *Lookup Table* στα διακριτά χαρακτηριστικά των γειτονικών *hash*, μετατρέποντας την διαδικασία από διακριτό μετασχηματισμό σε γραμμικό. Με αυτό τον τρόπο επιτρέπεται η εκπαίδευση των κωδικοποιήσεων των συντεταγμένων/διανυσμάτων όψης σε

συνάρτηση με την αξιοπιστία 3D ανακατασκευής.

Χρήση Δικτύου LSTM για τον υπολογισμό των μήκους των πεπερασμένων βημάτων

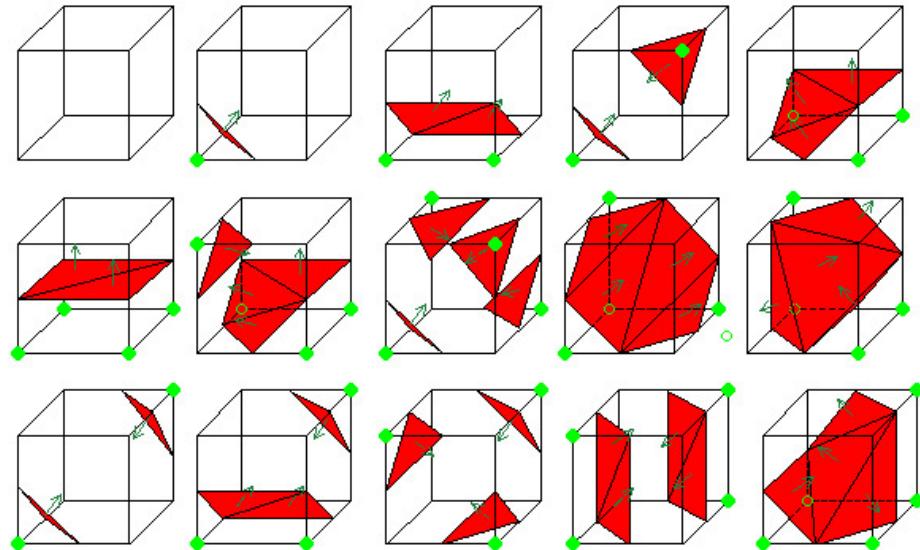
(Από το κεφάλαιο Μέθοδοι Απόδοσης - Ray Marching 2.2) Πολλές φορές είναι θεμιτό να βελτιστοποιηθούν όσο τον δυνατόν περισσότερο χρονοβόροι αλγόριθμοι που εκτελούν απόδοση γραφικών. Στην περίπτωση του Ray Marching αυτό γίνεται εφικτό με έναν περιορισμό στο πλήθος των βημάτων που γίνονται προς μια κατεύθυνση. Ωστόσο, δεν είναι γνωστές οι διαστάσεις της σκηνής, ο χώρος ο οποίο αναπαρίσταται δηλαδή, η επιλογή του βήματος βάσισης δηλαδή του μήκους το οποίο θα διανύσει μια ακτίνα για να ελέγξει ξανά την απόστασή της από την έμμεση επιφάνεια, μπορεί να μην είναι ιδανική. Η επιστημονική κοινότητα εξελίσσοντας τα δίκτυα LSTM δηλαδή Long Short Term Memory το οποίο αναπαριστώντας διάφορες σκηνές από διάφορες όψεις ενός χώρου είναι σε θέση να εκτιμήσει το ιδανικό μήκος βήματος δ και να το εισάγει στον αλγόριθμο δίνοντας καλύτερα αποτελέσματα σε λιγότερο χρόνο. Στο κομμάτι της εργασίας δεν υιοθετείται LSTM εκδοχή του δικτύου ray tracing, ωστόσο είναι μια καλή βελτιστοποίηση που θα μείωνε τον χρόνο απόδοσης της σκηνής σε εικόνα.

Algorithm 4 Ray Marching Algorithm using LSTM step length Prediction

```
function RayMarch( $\Phi, K, E, (u, v)$ )
     $d_0 \leftarrow 0.05$                                      # Near plane
     $(h_0, c_0) \leftarrow (0, 0)$                          # Initial state of LSTM
    for  $i \leftarrow 0$  to  $max\_iter$  do
         $x_i \leftarrow r_{u,v}(d_i)$                            # Calculate world coordinates
         $v_i \leftarrow \Phi(x_i)$                              # Extract feature vector
         $(\delta, h_{i+1}, c_{i+1}) \leftarrow LSTM(v_i, h_i, c_i)$  # Predict steplength using ray marching LSTM
         $d_{i+1} \leftarrow d_i + \delta$                          # Update d
    end for
    return  $r_{u,v}(d_{max\_iter})$ 
end function
```

Αλγόριθμος εξαγωγής Πλέγματος Τριγώνων από Έμμεση αναπαράσταση επιφάνειας | Marching Cubes

Πλήθος μεθόδων από την δεκαετία του 80 ακόμα προσπάθησαν να μετατρέψουν τις έμμεσες αναπαράστασης γραφικών σε πλέγματα πολυγώνων με κάποια ανοχή στα σφάλματα και χωρίς βλάβη της γενικότητας, επιτρέποντας την χρήση των υπαρχόντων πολυγώνων. Το πρόβλημα αυτό ήρθε να λύση οριστικά δίνοντας την χρυσή τομή στο ισοζύγιο υπολογιστικής πολυπλοκότητας και ανάλυσης ο αλγόριθμος *Marching Cubes*.



Σχήμα 6: Απεικόνιση Lookup Table ανάλυσης 8, αλγορίθμου Marching Cubes

Ο αλγόριθμος Marching Cubes[90], στην τρισδιάστατη μορφή του κάνει χρήση της δομής του ογκομετρικού διαμερισμού του χώρου σε voxels. Τα βήματα που ακολουθεί είναι τα εξής:

Algorithm 5 3D Marching Cubes Algorithm

Classify grid nodes as inside/outside:

for each grid node $(x_{i,j,k})$ **do**

 Is $F(x_{i,j,k}) > 0$ or $F(x_{i,j,k}) < 0$?

end for

Classify cell: $2^{resolution}$ configurations:

for each cell **do**

 In/out for each corner

end for

Compute intersection points:

for each edge **do**

 Linear interpolation along edges

end for

Connect them by edges:

Look-up table for path configuration

Disambiguation by modified table [Montani '94]

Παράμετροι δικτύων

Χρήση Δικτύου IDR

Το δίκτυο IDR[6] έχει παραμετροποιηθεί ώστε να χρησιμοποιεί το κάθε μοντέλο κωδικοποίησης με παραμέτρους που ορίζονται σε ένα αρχείο HOCON (JSON). Παράλληλα εκεί ορίζονται όλοι οι υπερπαράμετροι του δικτύου καθώς και αντίστοιχο αρχείο χρησιμοποιείται για την ρύθμιση του δικτύου σε άγνωστη θέση καμερών η οποία εκτιμάται και εκπαιδεύεται. Στα πλαίσια της εργασίας δεν κρίθηκε απαραίτητο να εκπαιδευτούν οι θέσεις των καμερών.

Οι παράμετροι του IDR με τις οποίες εκπαιδεύτηκε είναι οι εξής:

- πλήθος pixel ανά batch, training_pixel_batch_size = 2048
- ορόσημα παραμέτρου σύγκλισης μάσκας, α -milestones = [250, 500, 750, 1000, 1250]
- αρχικό βήμα εκπαίδευσης, learning_rate = 0.0001
- ποσοστό μείωσης του learning_rate, sched_factor = 0.5
- εποχές μείωσης του βήματος εκπαίδευσης, sched_milestones = [1000, 1500]
- βάρος απόκλισης από την εικονική εξίσωση, eikonal_weight = 0.1
- βάρος σφάλματος μάσκας, mask_weight = 100 ή 200 στο NFFB

Παράμετροι δικτύων, Implicit Network

- διάσταση εισόδου, din = 3
- διάσταση εξόδου, dout = 1
- διαστάσεις κρυφών επιπέδων, dims = [512, 512, 512, 512, 512, 512, 512, 512]
- γεωμετρική κανονικοποίηση, geometric_init = True
- bias βαρών, bias = 0.6
- διακοπή δικτύου στο επύπεδο 4, skip_in = [4]
- κανονικοποίηση βαρών, weight_norm = True
- Πλήθος Συχνοτήτων Κωδικοποίησης/Επιπέδων Βαθιάς Κωδικοποίησης = 6

Rendering Network

- διάσταση εισόδου, din = 9
- διάσταση εξόδου, dout=3
- Είδος κωδικοποίησης διανυσμάτων όψης, [Δικτύο κωδικοποίησης]
- διαστάσεις κρυφών επιπέδων, dims = [512, 512, 512, 512]
- κανονικοποίηση βαρών,
- Πλήθος Συχνοτήτων Κωδικοποίησης/Επιπέδων Βαθιάς Κωδικοποίησης = 4

Δίκτυο Βαθιάς Κωδικοποίησης NFFB

- Μέγεθος πίνακα Hash, $T = 10^{19}$
- Πλήθος χαρακτηριστικών ανά κελί $F = 2$
- Αρχική Ανάλυση (low res) $N_0 = 16$
- Τελική Ανάλυση (coarse res) $N_{max} = 512$
- Συντελεστής κανονικοποίησης εισόδου, bound = 0.42

Κωδικοποίηση HashGrid3D

- Μέγεθος πίνακα Hash, $T = 10^5$
- Πλήθος χαρακτηριστικών ανά κελί $F = 2$
- Αρχική Ανάλυση (low res) $N_0 = 8$
- Τελική Ανάλυση (coarse res) $N_{max} = 512$
- Συντελεστής κανονικοποίησης εισόδου, bound = 1.0

Κωδικοποίηση HashGrid3D_TCNN

- Μέγεθος πίνακα Hash, $T = 10^{15}$
- Πλήθος χαρακτηριστικών ανά κελί $F = 2$
- Αρχική Ανάλυση (low res) $N_0 = 16$
- Τελική Ανάλυση (coarse res) $N_{max} = 512$
- Συντελεστής κανονικοποίησης εισόδου, bound = 1.0

Δίκτυο Βαθιάς Κωδικοποίησης StylemodNFFB

- Μέγεθος πίνακα Hash, $T = 10^5$
- Πλήθος χαρακτηριστικών ανά κελί $F = 2$
- Αρχική Ανάλυση (low res) $N_0 = 16$
- Τελική Ανάλυση (coarse res) $N_{max} = 512$
- Συντελεστής κανονικοποίησης εισόδου, bound = 0.45

Δίκτυο Βαθιάς Κωδικοποίησης StylemodNFFB_TCNN

- Μέγεθος πίνακα Hash, $T = 10^{19}$
- Πλήθος χαρακτηριστικών ανά κελί $F = 2$

- Αρχική Ανάλυση (low res) $N_0 = 16$
- Τελική Ανάλυση (coarse res) $N_{max} = 512$
- Συντελεστής κανονικοποίησης εισόδου, bound = 0.45

Η επιλογή των παραμέτρων έγινε με βάση πειραματικές διαδικασίες, ανάλυση του αν δουλεύουν οι προτεινόμενες από τις δημοσιευμένες έρευνες και εύρεσης κατάλληλων παραμέτρων με βάση μετρικών εκπαίδευσης όπως το σφάλμα.

Παράρτημα Δ: Αποθετήριο Έρευνας

Η συντριπτική πλειοψηφία των αλγορίθμων που παρουσιάζονται στην παρούσα εργασία έχουν υλοποιηθεί σε *Python* με χρήση διαφόρων βιβλιοθηκών και ακόμα και χρήση συναρτήσεων πάνω σε *C++* με κύρια περιβάλλον τεχνητής νοημοσύνης το σύνολο των βιβλιοθηκών (*framework*) της *PyTorch*. Έχει δημιουργηθεί ένα αποθετήριο που περιέχει το σύνολο του τεχνικού μέρους της εργασίας στον σύνδεσμο: *HashModNFFBanks.git*. Σε κάθε περίπτωση παρουσίαση των ουσιώδη αλγορίθμων με χρήση ψευδογλώσσας έχει γίνει μέσα στο βασικό σκέλος της εργασίας και δίνονται οι απαραίτητες περαιτέρω πληροφορίες και για την διαδικασία χρήσης του κώδικα στο παραπάνω αποθετήριο.

Βιβλιογραφία

- [1] Z. Wu, Y. Jin, and K. M. Yi, “Neural fourier filter bank,” 2023.
- [2] T. Müller, A. Evans, C. Schied, and A. Keller, “Instant neural graphics primitives with a multiresolution hash encoding,” *ACM Trans. Graph.*, vol. 41, pp. 102:1–102:15, July 2022.
- [3] 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,” 2020.
- [4] P. Wang, L. Liu, Y. Liu, C. Theobalt, T. Komura, and W. Wang, “Neus: Learning neural implicit surfaces by volume rendering for multi-view reconstruction,” *CoRR*, vol. abs/2106.10689, 2021.
- [5] V. Sitzmann, M. Zollhöfer, and G. Wetzstein, “Scene representation networks: Continuous 3d-structure-aware neural scene representations,” 2020.
- [6] L. Yariv, Y. Kasten, D. Moran, M. Galun, M. Atzmon, B. Ronen, and Y. Lipman, “Multiview neural surface reconstruction by disentangling geometry and appearance,” *Advances in Neural Information Processing Systems*, vol. 33, pp. 2492–2502, 2020.
- [7] Z. Chen, “A review of deep learning-powered mesh reconstruction methods,” 2023.
- [8] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, “Attention is all you need,” 2023.
- [9] X. Huang and S. Belongie, “Arbitrary style transfer in real-time with adaptive instance normalization,” 2017.
- [10] R. Winkel, “Generalized bernstein polynomials and bézier curves: An application of umbral calculus to computer aided geometric design,” *Advances in Applied Mathematics*, vol. 27, no. 1, pp. 51–81, 2001.
- [11] D. Hearn, M. Baker, and W. Carithers, *Computer Graphics with OpenGL*. Prentice Hall, 2011.
- [12] M. Michalkiewicz, J. K. Pontes, D. Jack, M. Baktashmotagh, and A. Eriksson, “Implicit surface representations as layers in neural networks,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 4743–4752, 2019.
- [13] J. Hart, “Sphere tracing: A geometric method for the antialiased ray tracing of implicit surfaces,” *The Visual Computer*, vol. 12, 06 1995.
- [14] S. Haykin, *Neural networks and learning machines*, 3/E. Pearson Education India, 2009.

- [15] M. Atzmon and Y. Lipman, “Sal: Sign agnostic learning of shapes from raw data,” 2020.
- [16] M. Atzmon and Y. Lipman, “Sald: Sign agnostic learning with derivatives,” 2020.
- [17] M. Atzmon, N. Haim, L. Yariv, O. Israelov, H. Maron, and Y. Lipman, “Controlling neural level sets,” *CoRR*, vol. abs/1905.11911, 2019.
- [18] W. Matusik, *A data-driven reflectance model*. PhD thesis, Massachusetts Institute of Technology, 2003.
- [19] K. Berger, A. Weidlich, A. Wilkie, and M. Magnor, “Modeling and verifying the polarizing reflectance of real-world metallic surfaces,” *IEEE Computer Graphics and Applications*, vol. 32, no. 2, pp. 24–33, 2012.
- [20] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, “Gradient-based learning applied to document recognition,” *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.
- [21] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, “Learning representations by back-propagating errors,” *Cognitive modeling*, vol. 5, no. 3, p. 1, 1988.
- [22] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
- [23] A. Gropp, L. Yariv, N. Haim, M. Atzmon, and Y. Lipman, “Implicit geometric regularization for learning shapes,” 2020.
- [24] M. Tancik, P. P. Srinivasan, B. Mildenhall, S. Fridovich-Keil, N. Raghavan, U. Singhal, R. Ramamoorthi, J. T. Barron, and R. Ng, “Fourier features let networks learn high frequency functions in low dimensional domains,” 2020.
- [25] “SphericalHarmonicsPortal — Wakapon — wiki.nuaj.net.” <https://wiki.nuaj.net/index.php?title=SphericalHarmonicsPortal>. [Accessed 26-10-2023].
- [26] J. Yao, C. Wang, T. Wu, and C. Li, “Geometry-guided ray augmentation for neural surface reconstruction with sparse views,” 2023.
- [27] “Legendre Polynomials — hyperphysics.phy-astr.gsu.edu.” <http://hyperphysics.phy-astr.gsu.edu/hbase/Math/legend.html>. [Accessed 26-10-2023].
- [28] T. Karras, S. Laine, M. Aittala, J. Hellsten, J. Lehtinen, and T. Aila, “Analyzing and improving the image quality of stylegan,” 2020.
- [29] T. Chen, M. Lucic, N. Houlsby, and S. Gelly, “On self modulation for generative adversarial networks,” *ArXiv*, vol. abs/1810.01365, 2018.
- [30] G. Pavlakos, L. Zhu, X. Zhou, and K. Daniilidis, “Learning to estimate 3d human pose and shape from a single color image,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 459–468, 2018.
- [31] F. Bogo, A. Kanazawa, C. Lassner, P. Gehler, J. Romero, and M. J. Black, “Keep it smpl: Automatic estimation of 3d human pose and shape from a single image,” in *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part V* 14, pp. 561–578, Springer, 2016.
- [32] N. Kolotouros, G. Pavlakos, and K. Daniilidis, “Convolutional mesh regression for single-image human shape reconstruction,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 4501–4510, 2019.

- [33] Y. Deng, J. Yang, S. Xu, D. Chen, Y. Jia, and X. Tong, “Accurate 3d face reconstruction with weakly-supervised learning: From single image to image set,” 2020.
- [34] N. Wang, Y. Zhang, Z. Li, Y. Fu, W. Liu, and Y.-G. Jiang, “Pixel2mesh: Generating 3d mesh models from single rgb images,” in *Proceedings of the European conference on computer vision (ECCV)*, pp. 52–67, 2018.
- [35] C. Wen, Y. Zhang, Z. Li, and Y. Fu, “Pixel2mesh++: Multi-view 3d mesh generation via deformation,” in *ICCV*, 2019.
- [36] Z. Zheng, T. Yu, Y. Wei, Q. Dai, and Y. Liu, “Deephuman: 3d human reconstruction from a single image,” in *The IEEE International Conference on Computer Vision (ICCV)*, October 2019.
- [37] C. M. Jiang, J. Huang, A. Tagliasacchi, and L. Guibas, “Shapeflow: Learnable deformations among 3d shapes,” 2021.
- [38] T. Groueix, M. Fisher, V. G. Kim, B. C. Russell, and M. Aubry, “Atlasnet: A papier-mâché approach to learning 3d surface generation,” 2018.
- [39] Z. Chen, A. Tagliasacchi, and H. Zhang, “Bsp-net: Generating compact meshes via binary space partitioning,” 2020.
- [40] W. Liu, A. Rabinovich, and A. C. Berg, “Parsenet: Looking wider to see better,” 2015.
- [41] Z. Chen and H. Zhang, “Neural marching cubes,” *ACM Transactions on Graphics*, vol. 40, pp. 1–15, dec 2021.
- [42] P.-S. Wang, C.-Y. Sun, Y. Liu, and X. Tong, “Adaptive o-CNN,” *ACM Transactions on Graphics*, vol. 37, pp. 1–11, dec 2018.
- [43] C. B. Choy, D. Xu, J. Gwak, K. Chen, and S. Savarese, “3d-r2n2: A unified approach for single and multi-view 3d object reconstruction,” 2016.
- [44] Z. Chen and H. Zhang, “Learning implicit fields for generative shape modeling,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019.
- [45] J. J. Park, P. Florence, J. Straub, R. Newcombe, and S. Lovegrove, “Deepsdf: Learning continuous signed distance functions for shape representation,” 2019.
- [46] F. Yu, Z. Chen, M. Li, A. Sanghi, H. Shayani, A. Mahdavi-Amiri, and H. Zhang, “Capri-net: Learning compact cad shapes with adaptive primitive assembly,” 2021.
- [47] D. Ren, J. Zheng, J. Cai, J. Li, and J. Zhang, “Extrudenet: Unsupervised inverse sketch-and-extrude for shape parsing,” 2022.
- [48] N. Sharp and M. Ovsjanikov, “Pointtrinet: Learned triangulation of 3d point sets,” *ArXiv*, vol. abs/2005.02138, 2020.
- [49] C. Nash, Y. Ganin, S. M. A. Eslami, and P. W. Battaglia, “Polygen: An autoregressive generative model of 3d meshes,” 2020.
- [50] C. Lin, T. Fan, W. Wang, and M. Nießner, “Modeling 3d shapes by reinforcement learning,” 2020.
- [51] C. Sima, W. Tong, T. Wang, L. Chen, S. Wu, H. Deng, Y. Gu, L. Lu, P. Luo, D. Lin, and H. Li, “Scene as occupancy,” 2023.

- [52] L. Mescheder, M. Oechsle, M. Niemeyer, S. Nowozin, and A. Geiger, “Occupancy networks: Learning 3d reconstruction in function space,” in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 4460–4470, 2019.
- [53] S. Peng, M. Niemeyer, L. Mescheder, M. Pollefeys, and A. Geiger, “Convolutional occupancy networks,” in *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part III 16*, pp. 523–540, Springer, 2020.
- [54] J. Chibane, T. Alldieck, and G. Pons-Moll, “Implicit functions in feature space for 3d shape reconstruction and completion,” in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 6970–6981, 2020.
- [55] Y. Jiang, D. Ji, Z. Han, and M. Zwicker, “Sdfdiff: Differentiable rendering of signed distance fields for 3d shape optimization,” 2022.
- [56] S. Liu, Y. Zhang, S. Peng, B. Shi, M. Pollefeys, and Z. Cui, “Dist: Rendering deep implicit signed distance function with differentiable sphere tracing,” 2020.
- [57] S. Liu, S. Saito, W. Chen, and H. Li, “Learning to infer implicit surfaces without 3d supervision,” 2019.
- [58] Y. Furukawa and J. Ponce, “Accurate, dense, and robust multiview stereopsis,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 8, pp. 1362–1376, 2010.
- [59] J. L. Schönberger, E. Zheng, J.-M. Frahm, and M. Pollefeys, “Pixelwise view selection for unstructured multi-view stereo,” in *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part III 14*, pp. 501–518, Springer, 2016.
- [60] N. D. Campbell, G. Vogiatzis, C. Hernández, and R. Cipolla, “Using multiple hypotheses to improve depth-maps for multi-view stereo,” in *Computer Vision–ECCV 2008: 10th European Conference on Computer Vision, Marseille, France, October 12–18, 2008, Proceedings, Part I 10*, pp. 766–779, Springer, 2008.
- [61] E. Tola, C. Strecha, and P. Fua, “Efficient large-scale multi-view stereo for ultra high-resolution image sets,” *Machine Vision and Applications*, vol. 23, pp. 903–920, 2012.
- [62] B. Curless and M. Levoy, “A volumetric method for building complex models from range images,” in *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*, pp. 303–312, 1996.
- [63] P. Merrell, A. Akbarzadeh, L. Wang, P. Mordohai, J.-M. Frahm, R. Yang, D. Nistér, and M. Pollefeys, “Real-time visibility-based fusion of depth maps,” in *2007 IEEE 11th International Conference on Computer Vision*, pp. 1–8, Ieee, 2007.
- [64] M. Kazhdan, M. Bolitho, and H. Hoppe, “Poisson surface reconstruction,” in *Proceedings of the fourth Eurographics symposium on Geometry processing*, vol. 7, p. 0, 2006.
- [65] V. Leroy, J.-S. Franco, and E. Boyer, “Shape reconstruction using volume sweeping and learned photoconsistency,” in *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 781–796, 2018.
- [66] S. Donne and A. Geiger, “Learning non-volumetric depth fusion using successive reprojections,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 7634–7643, 2019.

- [67] G. Riegler, A. O. Ulusoy, H. Bischof, and A. Geiger, “Octnetfusion: Learning depth fusion from data,” in *2017 International Conference on 3D Vision (3DV)*, pp. 57–66, IEEE, 2017.
- [68] P.-H. Huang, K. Matzen, J. Kopf, N. Ahuja, and J.-B. Huang, “Deepmvs: Learning multi-view stereopsis,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2821–2830, 2018.
- [69] Y. Yao, Z. Luo, S. Li, T. Fang, and L. Quan, “Mvsnet: Depth inference for unstructured multi-view stereo,” 2018.
- [70] Y. Yao, Z. Luo, S. Li, T. Shen, T. Fang, and L. Quan, “Recurrent mvsnet for high-resolution multi-view stereo depth inference,” in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 5525–5534, 2019.
- [71] N. Snavely, S. M. Seitz, and R. Szeliski, “Photo tourism: exploring photo collections in 3d,” in *ACM siggraph 2006 papers*, pp. 835–846, 2006.
- [72] J. L. Schonberger and J.-M. Frahm, “Structure-from-motion revisited,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4104–4113, 2016.
- [73] Y. Kasten, A. Geifman, M. Galun, and R. Basri, “Algebraic characterization of essential matrices and their averaging in multiview settings,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 5895–5903, 2019.
- [74] N. Jiang, Z. Cui, and P. Tan, “A global linear method for camera pose registration,” in *Proceedings of the IEEE international conference on computer vision*, pp. 481–488, 2013.
- [75] C. Tang and P. Tan, “Ba-net: Dense bundle adjustment network,” *arXiv preprint arXiv:1806.04807*, 2018.
- [76] B. Triggs, P. F. McLauchlan, R. I. Hartley, and A. W. Fitzgibbon, “Bundle adjustment—a modern synthesis,” in *Vision Algorithms: Theory and Practice: International Workshop on Vision Algorithms Corfu, Greece, September 21–22, 1999 Proceedings*, pp. 298–372, Springer, 2000.
- [77] B. Peng, J. Hu, J. Zhou, X. Gao, and J. Zhang, “Intrinsiccngp: Intrinsic coordinate based hash encoding for human nerf,” 2023.
- [78] R. L. De Queiroz and P. A. Chou, “Compression of 3d point clouds using a region-adaptive hierarchical transform,” *IEEE Transactions on Image Processing*, vol. 25, no. 8, pp. 3947–3956, 2016.
- [79] B. Isik, P. A. Chou, S. J. Hwang, N. Johnston, and G. Toderici, “Lvac: Learned volumetric attribute compression for point clouds using coordinate based networks,” *Frontiers in Signal Processing*, vol. 2, p. 1008812, 2022.
- [80] D. Rho, B. Lee, S. Nam, J. C. Lee, J. H. Ko, and E. Park, “Masked wavelet representation for compact neural radiance fields,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 20680–20690, June 2023.
- [81] H. Aanæs, R. R. Jensen, G. Vogiatzis, E. Tola, and A. B. Dahl, “Large-scale data for multiple-view stereopsis,” *International Journal of Computer Vision*, pp. 1–16, 2016.
- [82] A. Ben-Israel, “A newton-raphson method for the solution of systems of equations,” *Journal of Mathematical analysis and applications*, vol. 15, no. 2, pp. 243–252, 1966.

- [83] T. Müller, “tiny-cuda-nn - pytorch extension library based on fully fused mlps.” <https://github.com/NVlabs/tiny-cuda-nn>, 4 2021. Version 1.7.
- [84] V. Sitzmann, J. N. Martel, A. W. Bergman, D. B. Lindell, and G. Wetzstein, “Implicit neural representations with periodic activation functions,” in *Proc. NeurIPS*, 2020.
- [85] R. Jensen, A. Dahl, G. Vogiatzis, E. Tola, and A. s H., “Large scale multi-view stereopsis evaluation,” in *CVPR’14*, 2014.
- [86] H. Hachiya, Y. Saito, K. Iteya, M. Nomura, and T. Nakamura, “Distance estimation with 2.5d anchors and its application to robot navigation,” *ROBOMECH Journal*, vol. 5, 09 2018.
- [87] D. G. Lowe, “Distinctive image features from scale-invariant keypoints,” *International journal of computer vision*, vol. 60, pp. 91–110, 2004.
- [88] R. Hartley and A. Zisserman, *Multiple view geometry in computer vision*. Cambridge university press, 2003.
- [89] T. Kurihara and Y. Takaki, “Shading of a computer-generated hologram by zone plate modulation,” *Optics express*, vol. 20, pp. 3529–40, 02 2012.
- [90] W. Lorensen and H. Cline, “Marching cubes: A high resolution 3d surface construction algorithm,” *ACM SIGGRAPH Computer Graphics*, vol. 21, pp. 163–, 08 1987.