Paper: Onderzoeksvoorstel

Het automatisch genereren van een dynamische 3D ruimte met oog op een didactische meerwaarde voor de opleiding ergotherapie .

Bachelor Proef, 2023-2024

Wolf Degol

E-mail: wolf.degol@student.hogent.be
Co-promotor: Jana Van Damme (HOGENT, jana.vandamme@hogent.be)

Samenvatting

Het aanmaken van 3D ruimtes vergt tijd en expertise over game-engines zoals Unity of Unreal Engine en 3D-editors zoals Maya of Blender. Als reactie op dit probleem wordt hier voorgesteld om te onderzoeken of het proces van 3D-ruimtecreatie (deels) kan geautomatiseerd worden en uitgevoerd door lector of studenten ergotherapie. Er is namelijk nood aan een gebruiksvriendelijke 3D editor om in de les te gebruiken en combineren met een VR-bril zoals de Meta Quest 2. Recente ontwikkelingen in computer visie en generatieve artificiële intelligentie bieden veel nieuwe mogelijkheden tot creatie van beeld materiaal in zo wel 2 als 3 dimensies. In dit onderzoek tracht er achterhaald te worden welke algoritmes en tools het meest efficiënt zijn in dit geval.

Keuzerichting: Mobile & Enterprise development
Sleutelwoorden: Applicatieontwikkeling; 3D-Asset Generation; Generative AI; 360-video; VR; NeRF

Inhoudsopgave

Opmerking

1. Inleiding

Warre Neufkens trachtte in 2023 een dynamische 3D ziekenhuiskamer te creëren voor het ondersteunen van de studenten ergonomie te Hogent in zijn bachelor proef voor de opleiding toegepaste informatica. Concreet maakte hij een virtuele ziekenhuiskamer met een bed en kasten. Je kon per object een tekst laten verschijnen met een uitleg over het aangeklikte object.

De applicatie slaat de bal mis bij de requirement van het dynamisch kunnen toevoegen en verwijderen van objecten in de ruimte. Dit wijt hij aan het hebben van weinig knowhow over de game engine zelf. Deze tekortkoming zorgt ervoor dat de applicatie aan zich niet bruikbaar is in een reële situatie. Het veranderen van de positie, kleur en vorm van meubels, ramen, deuren, ... is onder andere waar de doelgroep naar op zoek is.

Recente ontwikkelingen in computer visie, object-detectie, object-classificering en generatieve kunstmatige intelligentie doen dan de vraag rijzen of het misschien mogelijk is om dat maak-proces van zo'n 3D ruimte (gedeeltelijk) te kunnen automatiseren. Zo'n applicatie zou het doelpubliek van Warre Neufkens zijn originele POC in staat stellen om zelf een 3D omgeving aan te maken. Beter nog zou er dan geen voorkennis van programmeren en/of het gebruik van 3D ontwikkelingssoftware vereist zijn. Dit brengt met zich mee dat het een pak goedkoper is om zo'n ruimte op te stellen. Het creëer proces kan op die manier ook als een didactisch proces worden gezien.

Welke technologieën zijn het meest geschikt voor het automatisch genereren van een 3D omgeving in functie van een gebruiksvriendelijke applicatie voor de lector/student ergotherapie?Meer specifiek moet er onderzocht worden in hoeverre we de reële ruimte van een ziekenhuiskamer gebruiken bij het genereren van de 3D omgeving. Op welke manier zullen 3D assets worden gegenereerd? Welk algoritme is het meest geschikt om 3D assets te genereren op basis van de reële omgeving? Welke hardware is het meest geschikt om de ruimte te scannen?

Ten slotte is de gebruiksvriendelijkheid ook van groot belang. Het doel van dit onderzoek is het bekomen van een gebruiksvriendelijke applicatie om in te zetten voor de opleiding ergotherapie. Wat is het juist dat een lector van de applicatie verlangt? Waar zouden de studenten de applicatie voor willen gebruiken? Hoeveel zal de gebruiker willen aanpassen aan deze gegenereerde ruimte? Wat is de minimale kwaliteit van de gewenste 3D ruimte? Er zal een requirement analyse moeten gebeuren om te antwoorden op deze vragen.

2. Literatuurstudie

2.1. Manieren om 3D-Omgevingen te bekijken

2.1.1. Stereoscopische foto's/video's

Stereoscopie betekend het bekijken van 2 beelden. Doormiddel van twee beelden te laten zien, denken onze hersenen dat een beeld in 3 dimensies bekijken. De camera die op de HOGENT beschikbaar is, is de Ricoh Theta Z1. Die camera maakt doormiddel van twee lenzen een beeld van een ruimte in alle richtingen. Softwarematig wordt dit beeld dan samengevoegd tot een bolvormige foto. Als kijker word je dan als het ware in het middelpunt van de bol gezet en kan je vervolgens rondkijken in de foto alsof het een echte 3-dimensionale foto was.

2.1.2. Virtuele 3D omgeving

Een virtuele 3D omgeving is een ruimte gecreëerd door middel van software. Structuren en objecten hebben een volume, een textuur en belichting waardoor er een idee van ruimte ontstaat. Toepassingen zijn te vinden in entertainment zoals films en games, maar ook simulaties en in dit geval didactisch. Het grote verschil met stereoscopische foto's en video's is dat in een 3D omgeving objecten uit verschillende hoeken kunnen worden bekeken en er eventueel zelfs invloed kan worden uitgeoefend op de omgeving. Software als Maya of Blender worden gebruikt om 3D objecten te maken, terwijl game engines als Unity of Unreal Engine gebuikt worden om de interactiviteit te verzorgen.

2.2. Manieren om 3D-Objecten virtueel weer te geven

Er zijn verschillende manieren om zo'n virtuele omgeving weer te geven, en elke manier heeft zo zijn eigen toepassingen.

2.2.1. Mesh

Mesh bestaat uit twee onderdelen: een vorm bepaald door de hoekpunten en geconnecteerde meshes bepaald door de randen of vlakken die de hoekpunten verbinden. Mesh bestaan meestal enkel uit driehoeken. (Luebke e.a., 2002)

2.2.2. Impliciete functies

Vormen in 3 dimensies kunnen weergegeven worden als wiskundige functies. Het is zelfs mogelijk complexere vormen te maken door het combineren van verschillende functies. Een nadeel is hier dat wij als mens slechts beperkt zijn in het combineren van genoeg functies om een juiste interpretatie trachten te maken van bijvoorbeeld een doornstruik. Hiervoor zijn nieuwe methodes nodig zoals bv. machine learning of gaussian splatting zoals we hieronder bespreken. (Tancik, 20)

2.2.3. Point Cloud

Een point cloud is in feite een groep van samenhangende punten. Ze stellen een reeks punten in de ruimte voor waar er een deel van een object te vinden is. Deze manier van voorstelling wordt vaak gebruikt bij laser scanners of NeRF modellen. De granulariteit en kleur voorstelling van de punten kunnen worden aangepast aan de noden van de toepassing. Point cloud data kan omgezet worden in een mesh door middel van algoritmes als Ball Pivoting Algorithm of Marching Cubes. (Fisher, 2014)

2.3. Manieren om om te zetten naar 3D

2.3.1. Fotogrammetrie

Het nemen van een foto is eigenlijk een projectie van een 3D-scène maken op een 2D-vlak, waarbij er informatie over de diepte verloren gaat. Het doel van fotogrammetrie is eigenlijk het tegenovergestelde: men wil aan de hand van overlappende foto's een zo gedetailleerd mogelijke 3D omgeving genereren. (FormLabs, g.d.) Men gaat een techniek toepassen genaamd Feature Extraction waarbij gemeenschappelijke kenmerken in foto's efficiënt worden opgeslagen en vergeleken met andere foto's. Op die manier kan men de positie van de camera-poses van de originele foto's berekenen. Door die poses en de gemeenschappelijke kenmerken te vergelijken, kan men een 3D schatting maken van de positie van deze kenmerken, dat heet Structure from motion (SfM) (Schonberger & Frahm, 2016). Daarna zal er een mesh worden gegenereerd op basis van deze structuur

en wordt per camera vergeleken welke kleur overeenkomt met een kenmerk, om zo een texture te bekomen. Die zal dan door middel van een UV map op de 3D structuur worden geplakt. Een veelgenoemd software pakket voor SfM te berekenen is COLMAP. Voor fotogrammetrie heb je dan weer 3DF Zephyr of AliceVision.

2.3.2. NeRF

NeRF of Neural Radiance Fields is een vernieuwende manier om overlappende foto's om te zetten in een point cloud. Anders dan bij fotogrammetrie wordt er bij deze techniek gebruik gemaakt van een neuraal netwerk. Een neuraal netwerk is onderdeel van machine learning, een tak van de computerwetenschappen en meer bepaald artificiële intelligentie waarbij statistische algoritmes trachten (zelfstandig) te leren en generaliseren van bepaalde data. Neurale netwerken worden gebruikt om modellen te bouwen die op hun beurt proberen data te voorspellen op basis van input.

Het neural network in het geval van NeRF krijgt foto's toegestopt die hij op zijn beurt zal analyseren. In het kort worden per foto willekeurige 'lichtstralen' uitgestuurd. Waar die dan snijden met een bepaald kenmerk, daar krijg je dan een punt. Als je dit blijft herhalen voor meerdere foto's voor een bepaalde periode, dan krijg je een verzameling van allemaal punten, of een point cloud.

Door alle foto's te laten analyseren door het neuraal netwerk krijg je dan een model, dat heet dan een Neural Radiance Field, of NeRF. Informatie over diepte en kleur zit dan vervat in dat model en dan kan je aan de hand van een punt in de ruimte (x,y,z) en de kijkrichting (□, □) vragen om een output. Dan krijg je dus de dichtheid van het volume en de kijkrichting-afhankelijke straling of 'radiance' die de kleur bepaald. Het model zal dus op elke plaats in de ruimte en elke kijkhoek proberen 'voorspellen' hoe het er uit zou moeten zien aan de hand van de foto's waarmee hij getraind is. (Mildenhall e.a., 2020)

Sinds de paper van NeRF uitkwam is er een hele reeks aan vervolg papers gekomen die het proces trachten te optimaliseren. Hoewel NeRF erg traag is in computatie, zijn er methodes ontwikkeld voor het optimaliseren van bijna elk stuk in de pijplijn. Hieronder zijn er een paar opgesomd.

Nerf Studio is een API geschreven door de originele bedenkers van NeRF. Je kan er op een makkelijke manier NeRFs creëren, trainen en testen. (Tancik e.a., 2023)

InpaintNeRF360 is een framework dat taal gebruikt om een Neural Radiance Field te bewerken. Dit gebeurd door middel van de training foto's te bewerken met een generatief model dat gegeven foto's kan bewerken door middel van een prompt. (Wang e.a., 2023)

Nvidia Instant Nerf is een framework gebouwd door onderzoek ondersteund door Nvidia. Je kan er op enkele minuten tijd een NeRF model mee trainen. In een 10-tal ms kan je neurale graphics renderen. Het model zelf trainen duurt volgens hen slechts 5 minuten. (Müller e.a., 2022)

NeRF Shop is een tool om een point cloud model mee te bewerken. Door het aanduiden van een deel van het model door middel van wat krabbels, kan de software volume detecteren. Na vervolgens het kiezen hoeveel je van het volume wil selecteren, kan je beginnen aanpassen. Je kan objecten verplaatsen, verbuigen of verwijderen.(Jambon e.a., 2023)

2.3.3. Gaussian Splatting

Geïnspireerd door NeRF gaat Gaussian splatting ook Radiance Fields bij houden. Gaussian splatting zal deze keer zonder een neuraal netwerk tewerk gaan. Graphics worden benaderd doormiddel van 3D gaussians, een reeks 3D gauss curves waarvan vorm en doorzichtigheid wordt gebruikt om het originele beeld te benaderen. Die nieuwe beelden worden dan afgetoetst met de originele beelden, om zo tot een accuraat model te komen. GS biedt een enorme vooruitgang in kwaliteit en snelheid waaraan beelden kunnen worden geladen. Er is sprake van een real-time rendering. (Kerbl e.a., 2023)

2.3.4. Lidar

LIDAR of Laser Imaging Detection And Ranging is een techniek die, zoals het zelf zegt, de afstand tot een object bepaald tot een object. Het kan heel precies de vorm van een omgeving scannen en wordt gebruikt in velden als agricultuur, archeologie en zelfrijdende auto's. Je hebt er gespecialiseerde sensors voor nodig om zo'n metingen te kunnen uitvoeren.

2.3.5. Object herkenning en classificering

Het doel van object herkenning is het herkennen van bepaalde kenmerken en die linken aan een bepaald woord, of classificering. Hierboven werd kort gesproken over Feature Extraction. Net zoals daar wordt gebruik gemaakt van een algoritme zoals bijvoorbeeld SIFT van Schonberger en Frahm (2016) om het in kaart te brengen van prominente kenmerken waarmee we aan de slag kunnen. Een andere manier van werken zou kunnen zijn om de omgeving te scannen op objecten door middel van object herkenning en classificering. Op basis van die data zou je voorgeprogrammeerde objecten kunnen samenvoegen in een ruimte. Enkele voorbeelden van platforms waar je modellen zou kunnen trainen zijn Tensorflow, Pytorch, Azure Vision Studio of customvision.ai.

2.3.6. Oriëntatie van objecten

Het bepalen van de oriëntatie van een gedetecteerd object is nodig wanneer je de ruimte

wil nabootsen. Een bed staat meestal evenwijdig met de muur bijvoorbeeld. Saxena e.a. (2009) stelt een manier voor bijvoorbeeld om de oriëntatie van een object te proberen achterhalen.

2.3.7. Asset generatie

Door middel van object detectie kunnen we ofwel een voorgemaakt model tonen op basis van de semantiek van het gedetecteerd object. Een andere mogelijkheid is om het gevonden object in een generatief model in te geven. Er zijn een aantal projecten lopende die prompts omzetten in 3D modellen: Raj e.a. (2023) stelt dreambooth3d voor. Een algoritme om met enkele foto's en een tekst-prompt een 3D-model te genereren. Xu e.a. (2023) stelt dan weer Neuralift – 360 voor. Een algoritme om een 3D model te genereren vanuit één enkele foto.

3. Methodologie

3.1. Requirement Analyse

3.1.1. Functionele requirements

Meta Quest 2 draait op Android Open Source Project (AOSP) Er zal ook duidelijk moeten worden welk algoritme en frameworks er zullen worden gebruikt, welke specificaties de pc's hebben om de modellen te trainen, ...

3.1.2. Niet-Functionele requirements

Daarnaast moet het ook duidelijk worden wat de lector specifiek verwacht van de applicatie. Tot nu toe weten we dat er een 3D ruimte nodig is met aanpasbare kenmerken. Ook weten we dat de oplossing die we trachten te maken gratis moet zijn aangezien er geen extra budget is.

3.1.3. MOSCOW-analyse

Uit deze requirements maken we onderscheid tussen must-haves, should-haves, could-haves en nice-to-haves. Uiteindelijk willen we een werkende proof-of-concept die op tijd klaar is. Dan is een volgorde zeker aangewezen.

3.2. Long List

De lijst van alle verschillende mogelijkheden van frameworks, tools en algoritmes. Die wordt dan in een tabel geschreven naast alle requirements met alle must-haves en zoveel mogelijk nice-to-haves.

3.3. Short List en Proof-Of-Concept

De proof of concept zal dienen als uiteindelijke test voor de frameworks tools en algoritmes die de short list hebben gehaald. Een functionele, gebruiksvriendelijke applicatie is hopelijk het resultaat.

4. Verwachte resultaten

Er zijn heel wat mogelijkheden om een 3D ruimte automatisch te laten genereren. Sommige zijn niet gratis en sommige zijn enkel nog maar in papers gebruikt. Het resultaat zal afhangen van de requirements die uiteindelijk uit de bus komen. De mogelijkheden voor de ontwikkeling zijn talrijk, dus ik heb er vertrouwen in dat een proof-of-concept die min of meer werkt wel moet lukken.

Referenties

Fisher, M. (2014). *Marching Cubes*. https://graphics.stanford.edu/~mdfisher/MarchingCubes.html#:~:text=Marching%20cubes%20is%20a%20simple,a%20region%20of%20the%20function.

FormLabs. (g.d.). *Photogrammetry: Step-by-Step Guide and Software Comparison*. https://formlabs.com/asia/blog/photogrammetry-guide-and-software-comparison/

Jambon, C., Kerbl, B., Kopanas, G., Diolatzis, S., Leimkühler, T., & Drettakis, G. (2023). NeRFshop: Interactive Editing of Neural Radiance Fields". *Proceedings of the ACM on Computer Graphics and Interactive Techniques*, *6*(1). https://repo-sam.inria.fr/fungraph/nerfshop/

Kerbl, B., Kopanas, G., Leimkühler, T., & Drettakis, G. (2023). 3D Gaussian Splatting for Real-Time Radiance Field Rendering. *ACM Transactions on Graphics*, *42*(4). https://repo-sam.inria.fr/fungraph/3d-gaussian-splatting/

Luebke, D., Reddy, M., Cohen, J., Varshney, A., Watson, B., & Huebner, R. (2002). *Level of detail for 3D graphics* (1st edition). Morgan Kaufmann.

Mildenhall, B., Srinivasan, P. P., Tancik, M., Barron, J. T., Ramamoorthi, R., & Ng, R. (2020). NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis. *CoRR*, *abs/2003.089*

Müller, T., Evans, A., Schied, C., & Keller, A. (2022). Instant Neural Graphics Primitives with a Multiresolution Hash Encoding. *ACM Trans. Graph.*, *41*(4), 102:1–102:15. https://doi.org/10.1145/3528223.3530127

Raj, A., Kaza, S., Poole, B., Niemeyer, M., Ruiz, N., Mildenhall, B., Zada, S., Aberman, K., Rubinstein, M., Barron, J., Li, Y., & Jampani, V. (2023). DreamBooth3D: Subject-Driven Text-to-3D Generation.

Saxena, A., Driemeyer, J., & Ng, A. Y. (2009). Learning 3-D object orientation from images. *2009 IEEE International Conference on Robotics and Automation*, 794–800. https://doi.org/10.1109/ROBOT.2009.5152855

Schonberger, J. L., & Frahm, J.-M. (2016). Structure-From-Motion Revisited. *Proceedings of the*

IEEE Conference on Computer Vision and Pattern Recognition (CVPR).

Tancik, M. The Metaverse: What, How, Why and When. In: 20. https://www.youtube.com/watch?v=isKbsNKArJU

Tancik, M., Weber, E., Ng, E., Li, R., Yi, B., Kerr, J., Wang, T., Kristoffersen, A., Austin, J., Salahi, K., Ahuja, A., McAllister, D., & Kanazawa, A. (2023). Nerfstudio: A Modular Framework for Neural Radiance Field Development. *ACM SIGGRAPH 2023 Conference Proceedings*.

Wang, D., Zhang, T., Abboud, A., & Süsstrunk, S. (2023). InpaintNeRF360: Text-Guided 3D Inpainting on Unbounded Neural Radiance Fields. *arXiv.org*.

Xu, D., Jiang, Y., Wang, P., Fan, Z., Wang, Y., & Wang, Z. (2023). NeuralLift-360: Lifting an In-the-Wild 2D Photo to a 3D Object With 360deg Views. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 4479–4489. https://arxiv.org/abs/2211.16431