Converting 2D Images to 3D Anaglyphs using Deep Learning Models

Varich Boonsanong, Ivy Ding CSE 493V VR Systems

PROBLEM & MOTIVATION

Technological innovation in the realm of virtual reality (VR) has progressed significantly with devices as ubiquitous as smartphones capable of showing 3D content. Since the advent of camera technology, we have over a century of 2D media material at our disposal, which we believe has untapped potential for future 3D-generated content. Our project focuses on leveraging the advancement in Depth Estimation Deep Learning Models to approximate the depth of an object(s) in a 2D image and convert it into a 3D anaglyph.

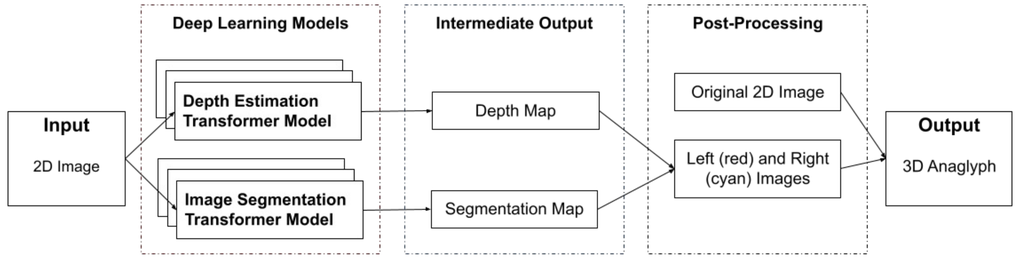
RELATED WORK

- Nasir Khalid's 2D to 3D Sneakers Project
- Mark Alcaraz's Realtime 2D to 3D Faces
- Michael Callahan's Stable Diffusion 2D to 3D Video Synthesis
- Eric Xie's 2D to 3D Video Conversion with CNNs
- Google Starline Project

SOLUTION

We use deep learning models to estimate the depth of each pixel in an image as well as determine the pixels associated with different objects. That info is then processed into a 3D anaglyph or image that can be displayed on a 3D screen, such as in a VR headset.

PIPELINE

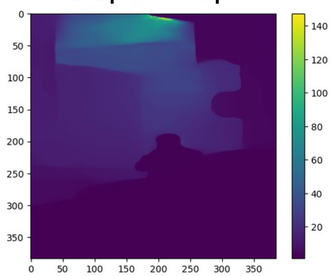


Our process uses two different pre-trained Deep Learning models. Given a 2D image we want to convert, we generate a depth map and segmentation map. In the post-processing step, using the segmentation map, we determine the object(s) of focus and decrease the depth of the pixels associated with it to give the illusion of that object(s) being closer. This helps us determine how much to shift the red and blue channels of the object(s) to give a 3D effect. Combining with the original 2D image, we supersede the red and blue channels to produce the final 3D anaglyph.

RESULTS

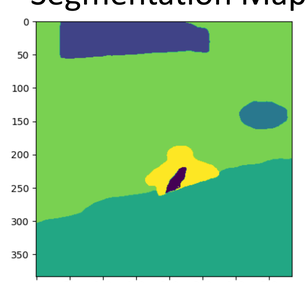
Original 2D Image

Depth Map



Full Image 3D Anaglyph

Segmentation Map



Selective Anaglyph

*view with 3D glasses to experience the effect

3D REAL-TIME VIDEO

We can extrapolate our process to work on videos. However, current deep learning models require immense computational power, thus, we struggle with low frame rates and deteriorating video call quality

Video Example

*similar to a lite version of Starline

REFERENCES

[Alc18] Marc Alcaraz. MARCALCARAZF/realtime-2d-to-3d-faces: Reconstructing real-time 3D faces from 2D images using deep learning. 2018. https://github.com/marcalcarazf/realtime-2D-to-3D-faces.

[Cal23] Michael Callahan. Mcallahan/stable-diffusion-2d-to-3d-video-synthesis: Automatic conversion of 2D video into 3D video by using stable diffusion depth map generation to create the secondary images. 2023. https://github.com/mcallahan/stable-diffusion-2D-to-3D-video-synthesis.

[Che+21] Bowen Cheng et al. "Masked-attention Mask Transformer for Universal Image Segmentation". In: CoRR abs/2112.01527 (2021). arXiv: 2112.01527. https://arxiv.org/abs/2112.01527.

[Kha20] Nasir Khalid. Nasirkhalid24/2Dto3D-shoes: Single View Shoe to 3D model using neural networks + Nvidia kaolin. 2020. https://github.com/NasirKhalid24/2Dto3D-Shoes.

[Law+21] Jason Lawrence et al. "Project Starline: A High-Fidelity Telepresence System". In: ACM Trans. Graph. 40.6 (Dec. 2021). issn: 0730-0301. doi: 10.1145/3478513.3480490. https://doi.org/10.1145/3478513.3480490.

[RBK21] René Ranftl, Alexey Bochkovskiy, and Vladlen Koltun. "Vision Transformers for Dense Prediction". In: CoRR abs/2103.13413 (2021). arXiv: 2103.13413. https://arxiv.org/abs/2103.13413.

[Xie16] Eric Xie. Piiswrong/deep3d: Automatic 2D-to-3d video conversion with cnns. 2016. https://github.com/piiswrong/deep3d.