The Application and Performance Comparison of Image Segmentation in VR Environments

Chih-Chuan Huang Advisor: Min-Chun Hu

INTRODUCTION

Generating innovative ideas is often the most challenging aspect of the creative process. Creators typically base their work on real-life people, events, and objects, blending them with their imagination. According to V. Rieuf et al. [1], virtual reality (VR) offers greater immersion than flat images, providing lifelike experiences. Thus, VR serves as a limitless creative space, enabling users to explore and realize ideas without real-world constraints. Additionally, artists often create moodboards before beginning their work, collecting diverse visual elements to spark creativity and deepen their concepts. Moodboards, which include collections of images, colors, materials and textures play a crucial role in the creative process.—aid in inspiring ideas, conveying emotions, and fostering consensus. They capture inspiration during the creative process and integrate these elements into artistic concepts, enhancing the work's depth and expressiveness.

Method

This study explores generating VR scenes aligned with creative themes and assisting artists in swiftly capturing intriguing visual elements to support their artistic creation. It introduces three input formats corresponding to different models and methods:

1. Segment Anything 2, which uses 2D coordinates as input.
2. CLIPSeg, which uses voice-to-text or text input.
3. A simple cropping tool (Crop) to retain the image within the object's boundaries.

The performance, speed, and application of these approaches will be compared to assess their potential and effectiveness in various creative scenarios.

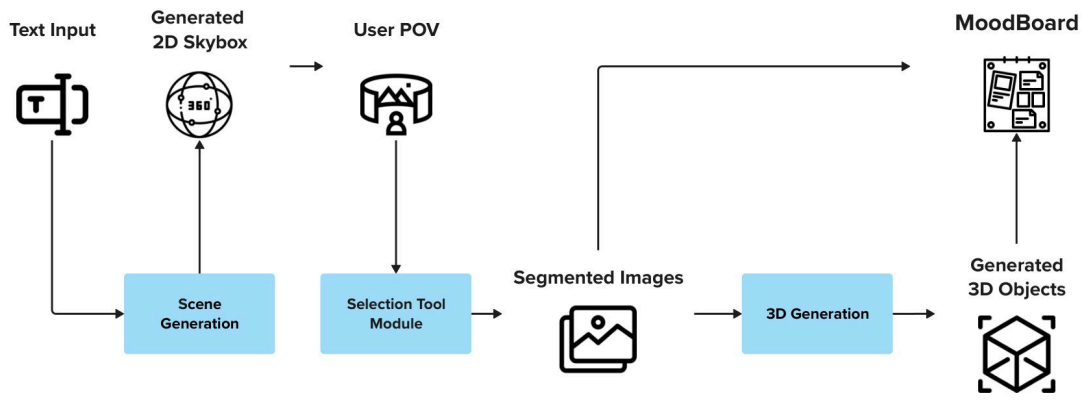

Figure 1. System overview

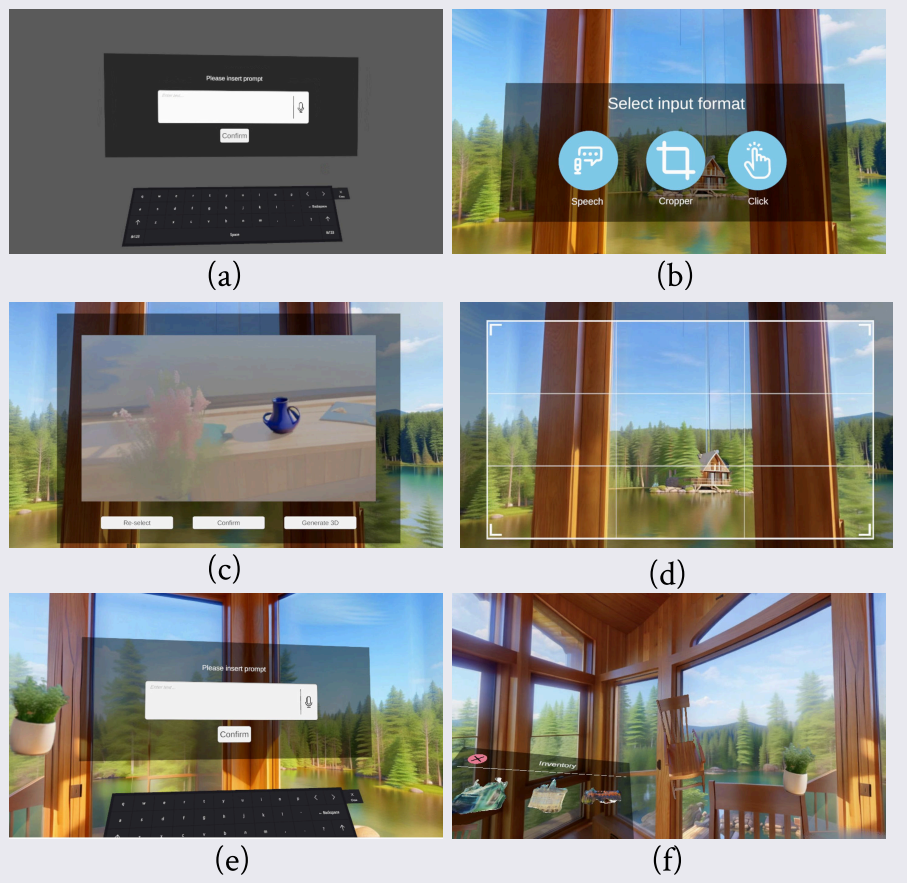

Figure 2. Userflow (a) Submit prompts to generate the environment (b) Select segmentation tools to extract items from the skybox (c) Evaluate segmentation results (d) Use the crop tool to save selected items with their surroundings (e) Submit prompts for CLIPSeg through the interface (f) View and integrate 2D and 3D objects in the Moodboard to enhance creativity

Evaluation

Table 1. Operational Efficiency

Model	Time (s)
SAM2	0.9059
Clipseg (1 output)	2.95
Clipseg (4 output)	3.08
Crop	Instance

1. The input format of SAM2 is a single point

Table 2. Segmentation Accuracy

Model	Metrix	Average	Min	Max
SAM2	F1 score	0.9059	0.3490	0.9868
Clipseg	F1 score	0.6530	0.0633	0.9736
SAM2	IoU	0.8517	0.2114	0.9740
Clipseg	IoU	0.7486	0.0327	0.9486

1. The Accuracy was evaluated on COCO 2017 Val dataset
2. The input format of SAM2 is a single point
3. The input prompts of CLIPSeg are relatively general, for example: "the bear", "the clothes," or "the trees."

Figure 3. Result Comparison.This comparison illustrates that CLIPSeg is particularly effective at segmenting multiple similar objects, the Crop tool effectively retains the relative background, and Segment Anything 2 achieves highly precise segmentation.

Conclusion

Each method has its strengths and is best suited for specific environments, as shown in Figure 3 : Segment Anything 2 excels in precise and efficient object segmentation, significantly reducing background noise and providing ideal material for further process such as 3D generation or integrating into various scenes. CLIPSeg, while less precise, effectively segments objects when textual descriptions are clear and can simultaneously segment multiple similar objects, despite some background noise. Besides, Its another advantage is text-based input, allowing users to freeing their hands for drawing. This makes it suitable for both VR and adaptable to XR environments. The Crop method is particularly effective when the background closely relates to the object, offering different perspectives and reference points during creation.

Reference

[1]V. Rieuf, C. Bouchard, V. Meyrueis, and J.-F. Omhover, "Emotional activity in early immersive design: Sketches and moodboards in virtual reality," DESIGN STUDIES, vol. 48. ELSEVIER SCI LTD, THE BOULEVARD, LANGFORD LANE, KIDLINGTON, OXFORD OX5 1GB, OXON, ENGLAND, pp. 43–75, Jan. 2017. doi: 10.1016/j.destud.2016.11.001