3D Person Segmentation and Anaglyph Generation

Alexander Hortua
Hugging Face Spaces: Object Segmentation App

Abstract

This project presents a 3D image processing web application that integrates state of the art AI based person segmentation with traditional stereoscopic image manipulation techniques. By utilizing the SegFormer deep learning model and classical computer vision tools, the system transforms 2D images into red cyan anaglyphs suitable for stereoscopic viewing. The application features an intelligent segmentation mask processor, depth simulation via stereo pair generation, and a fully interactive Gradio based interface that allows real time image customization and preview.

I. Introduction

With the rise of immersive content and growing interest in 3D visual experiences, this project introduces a novel approach for generating stereoscopic imagery from conventional 2D photographs. Leveraging transformer based segmentation models and image alignment algorithms, the developed system enables users to isolate people from images and produce realistic 3D effects, including red cyan anaglyphs.

This application supports:

- Semantic person segmentation using the SegFormer model
- Real time stereo effect rendering via horizontal parallax simulation
- Custom background integration
- Interactive control over object scaling and stereo depth
- Intuitive web interface powered by Gradio

The project highlights a practical integration of modern AI with classical stereoscopic vision.

II. Methodology

A. System Overview

The solution consists of four main components:

1. **Person Segmentation** – Identifies and isolates people in the image
2. **Mask Processing** – Resizes and centers the segmented object with proper transparency
3. **Stereoscopic Rendering** – Generates stereo image pairs by simulating horizontal disparity
4. **Anaglyph Generation** – Produces red cyan 3D images for stereoscopic viewing

B. Tools and Libraries

- **SegFormer (nvidia/segformer b0)** – Transformer based semantic segmentation
- **PyTorch** – Model inference and AI operations
- **OpenCV** – Image manipulation (resizing, shifting, merging)
- **Gradio** – Front end interface for real time interactions
- **NumPy & PIL** – Numerical operations and image conversions

C. Implementation Details

1) Person Segmentation

- Model: SegFormer, fine tuned on the ADE20K dataset
- Preprocessing: Input normalization and resizing
- Postprocessing: Morphological filtering (erosion), Gaussian blur, alpha channel handling

2) Mask Handling

- Mask resizing based on image and object proportions
- Transparent padding to retain original image aspect ratio
- Centering logic for consistent placement across various resolutions

3) Stereoscopic Simulation

- Generation of stereo pairs using horizontal pixel shifts (parallax)
- Configurable interaxial distance (0–10 px)
- Parallel stereo view generation for both left and right eye simulation

4) Anaglyph Composition

- Color channel manipulation (Red for left image, Cyan for right image)
- Depth aware alignment for improved 3D realism
- Foreground and background blending with minimal artifacts

III. Results

The final system generates three key outputs:

1. Segmentation Mask

- Accurate isolation of human subjects
- Transparent background support
- Smooth edges and adaptive alpha blending

2. Stereo Image Pair

- Side by side stereo visualization
- User controlled depth effect
- High fidelity in object preservation and placement

3. Anaglyph Output

- Red cyan stereoscopic rendering
- Adjustable 3D strength
- Low ghosting artifacts and crisp visual output

Feature Highlights:

- Person scaling (10–200%)
- Interaxial distance control
- Background customization
- Smart mask alignment and resizing

IV. Discussion

A. Technical Challenges

- **Mask Alignment**: Managing masks across varying image sizes while preserving alignment and proportions
- **Stereo Rendering**: Avoiding artifacts and ensuring visual comfort with variable interaxial distances
- **Performance Optimization**: Achieving real time responsiveness and efficient memory usage
- **Transparency Handling**: Ensuring edge smoothness and clean alpha blending during overlays
- **Difference between images types:** The biggest issue with this project is the fact of working with different image types.

B. Learning Outcomes

- Practical integration of AI segmentation models into user facing applications, and who to use transformers in this context
- Understanding of stereoscopic rendering principles.
- Advanced image processing pipeline development, Understanding how the mask, and different information can be manipulated to accomplish the image that we are looking for.
- Mask difference, how the mask generates a 1, 0 that can be used as filter on other images to remove the background.

V. Conclusion

This project successfully demonstrates how AI driven segmentation (with transformers) can enhance traditional 3D imaging workflows. By merging cutting edge machine learning models with classical vision techniques, the system provides an intuitive and effective method for creating stereoscopic and anaglyph images from standard 2D photos.

VI. Future Work

To build upon the current system and further enhance the user experience, several potential improvements are proposed:

1. **User-Controlled Object Placement**
   Enabling users to manually select where the segmented person is positioned within the final image would greatly enhance creative flexibility. This feature would allow dynamic placement through drag-and-drop or coordinate selection, making the output more customizable and visually balanced.

2. **Depth-Aware Rendering**
   Integration of depth estimation techniques to simulate more realistic 3D effects based on actual scene geometry rather than fixed horizontal shifts.