Medial Axis Techniques for Stereoscopic Extraction

Evan Goldberg

Walt Disney Animation Studios

1   Introduction

To perform the stereoscopic conversion of Disney's *Beauty and the Beast*, we developed novel extensions to standard medial axis techniques. Distance transforms, directional influences, and segmentation variations were used with medial axis skeletonization to automatically generate depth maps from the hand-drawn images. Overall, our methods significantly reduced the amount of manual rotoscoping and modeling traditionally required for a stereoscopic conversion process.

We define a depth map where white pixels are closer to the viewer and black pixels are farther away. Given the distance between two cameras and their convergence angle, we can derive an invertible set of functions between depth maps, displacement maps, physical geometry, inter-ocular disparity maps, and our final stereoscopic images.
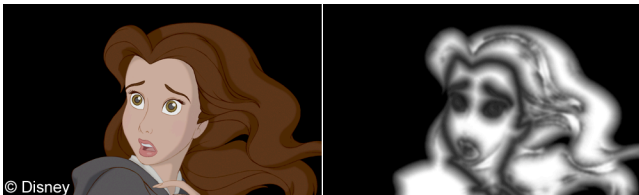
2   Medial Axis

Figure 1:   (a) Input   (b) Depth map

In computer graphics, the medial axis is a common technique to procedurally calculate skeletal structure from a volumetric region. We use the Pfaltz and Rosenfeld [1967] definition for the medial axis MA of a shape S described as the set of points that lie at the center of all maximally inscribed circles in S (i.e. all points equidistant to at least two closest points along the perimeter of S). A medial axis computation of color regions allows us to generate depth maps and ultimately a stereoscopic image (Figure 1b).

The first step of our process is to partition the original image into segmented regions S_i. This can be done manually via rotoscope techniques or programmatically by defining a threshold over red, green, blue, alpha, or luma color channels. Once appropriate regions have been defined, a medial axis MA_i is constructed through each S_i.

3   Distance Transform

Next, we define a distance transform for the medial axis skeleton. The borders of S_i can be likened to the seam of a Mylar balloon, where minimum inflation exists along the edges and maximum inflation at MA_i. The depth value of all intermediate pixels can be defined by an interpolation function that achieves the desired contour from edge to center. The profile of this curve can be linear (for regions with sharp peaks), convex (for areas that bow outward), or concave (for sections that bend inward). Depth gradation is heavily based on our choice of curvature, so it is important to define this interpolation function judiciously.

evan.goldberg@disney.com
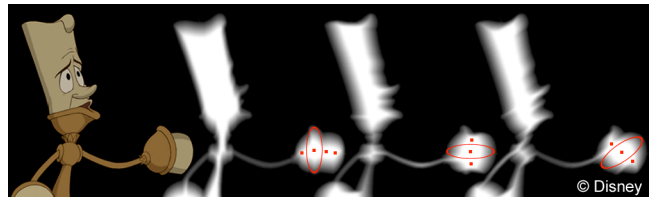
4   Directional Influence

Figure 2: (a) Input (b) Vertical (c) Horizontal (d) Rotated

A traditional medial axis is based on a symmetrical inward traversal from the shape periphery -- we extend this definition to accommodate cases where it is beneficial to influence this computation in a particular direction. A character's face, for example, often requires a gradual rounding around the cheeks (sides) and an abrupt falloff near the forehead and chin (top and bottom). In this case, we modify our medial axis definition to inscribe ellipses in S_i with a short vertical axis and a long horizontal axis (Figure 2c). This produces the ideal skeletal structure. Other scenarios benefit from ellipses with a short horizontal and a long vertical axis (Figure 2b). Thin, pointed areas, such as Lumiere's nose, are best fitted with an ellipse that is rotated about its center to capture the region's narrow shape as well as its upwards tilt (Figure 2d).

As the character animates over time, the best directional influence for S_i may be different on every frame. Thus, we animate the inscribed shape to capture the most salient geometric features of each pose.

5   Segmentation Variation

Figure 3: (a) Input   (b) Detailed   (c) General

The key to a successful depth map is the combination of large segmentation regions and small, localized details. A large area, such as a character's alpha mask, produces a good general rounding of the entire figure when inflated to its medial axis (Figure 3c). To extract interior details, the image is analyzed by its color channels, allowing the ink lines to form segmentation boundaries (Figure 3b). Individually, the global extraction is too general and the local extraction too extreme. However, using the local as a detailed bump map on top of the global produces the desired result.

References

PFALTZ, J. L. and ROSENFELD, A. 1967. Computer Representation of Planar Regions by Their Skeletons, *Communications of the ACM*, 10, 2, 119-125.