MVA2005 IAPR Conference on Machine VIsion Applications, May 16-18, 2005 Tsukuba Science City, Japan

3-6

Silhouette Extraction for Visual Hull Reconstruction

Youngbae Hwang, Jun-Sik Kim, Inso Kweon
Dept. of Electrical Engineering and Computer Science
Korea Advanced Institute of Science and Technology (KAIST)
Daejeon 305-701, Korea
unicorn@rcv.kaist.ac.kr, jskim@rcv.kaist.ac.kr, iskweon@kaist.ac.kr

Abstract

Most of volumetric 3D reconstruction methods need accurate silhouette information of foreground objects for the reliable 3D scene reconstruction of the objects. We present an approach to silhouette extraction based on our change detection method [1] using a statistical model and graph cuts based optimization. We describe some requirements for visual hull reconstruction. We show that our change detection method is appropriate to satisfy these requirements. We analyze our methodology through experiments and have shown the feasibility of reliable extraction of silhouettes. Consequently, we can reconstruct the 3D shape more accurately than by other change detection methods.

1 Introduction

In computer vision, volumetric 3D reconstruction from multiple cameras has long been an important and active research topic. Many volumetric 3D reconstruction methods such as visual hull reconstruction [2], voxel coloring [3], space carving [4] and probabilistic optimization-based approach [5] use silhouette information of foreground objects because the visual hull reconstruction using silhouette information provides an upper bound on the shape of the object [3-5]. When the visual hull is reconstructed, these methods carve voxels to be close to true shape using color information and various approaches. The silhouette information is obtained manually or by background subtraction or change detection. Since the precision of visual hull is very important to the results of 3D reconstruction, the accurate detection of silhouettes is critical.

Most silhouette extraction for 3D reconstruction is based on background subtraction in gray images. We can detect silhouette pixels of objects by simple intensity differences of corresponding pixels. In background subtraction, determining an appropriate threshold is important. A higher threshold suppresses significant changes and a lower one, however, swamps the difference map with spurious changes. Although the correct threshold is determined, the extracted silhouette may not show appropriate results to apply to 3D reconstruction because of limited quantization levels and noises in gray images.

In this paper, we show that the accuracy of silhouettes determines the accuracy of reconstructed 3D shape by visual hull reconstruction as well as other volumetric reconstruction methods [3-5]. We present three requirements of silhouette extraction for accurate visual hull reconstruction. And we show that our change detection method [1] fulfills these requirements suitably.

Our change detection method consists of two parts. One is to obtain an optimal threshold to classify object regions from background regions. The other is to refine previous thresholding result based on the threshold and relationships among neighborhood pixels. By means of these parts, we can extract accurate silhouette pixels and satisfy the requirements of silhouette extraction.

This paper is organized as follows. In Section 2, we present three requirements of silhouette extraction to obtain accurate 3D reconstruction. We show why the requirements are important by experiments. Then, we briefly explain our change detection algorithm [1] in Section 3. We show how our algorithm was designed to satisfy requirements. Section 4 shows silhouette extraction results and we validate our method to satisfy three requirements. Reconstruction results are given to illustrate the performance of the proposed algorithm. Finally, we conclude the paper in Section 5.

2 Requirements of silhouette extraction

The requirements of silhouette extraction for accurate volumetric 3D reconstruction are shown below.

- No missing pixel: Missing silhouette pixels in object regions are so critical that they make some holes in the reconstruction results. Although we go through further reconstruction process such as voxel coloring or space carving, we cannot recover these errors. Therefore, we have to minimize the number of missing pixels in object regions.

- An accurate boundary in the visual hull: The visual hull is built by an intersection of the reprojected silhouette images. So, boundaries of silhouettes determine the accuracy of reconstructed 3D shape. But it is hard to measure precision of an accurate boundary in the image space because the determination of a criterion is not clear. Therefore, we compare visual hull using silhouettes by our method with visual hull using silhouettes obtained manually.

- Automatic extraction: If we manually indicate the regions corresponding to a target object and a background by some lines, we can extract silhouettes of an interesting object in a single image [7][8]. Our approach uses two images corresponding to background and object images. Therefore, if our method can be comparable with the methods using a single image, we have to extract silhouettes automatically. For automatic extraction, we keep the minimal number of user-defined parameters and make the system that is not sensitive to these parameters.

Figure 1. Test silhouette images that have many extra pixels (There are some missing pixels in the third image)

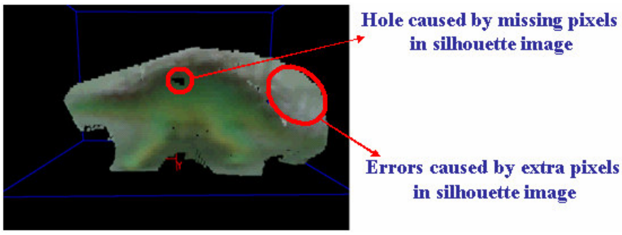


Figure 2. 3D reconstruction result using silhouette images in figure 1 by voxel coloring algorithm

We manually over-segment out objects to have inaccurate boundaries with some missing pixels in the third image. We reconstruct 3D shape using voxel coloring as shown in figure 2. There is a hole in the center caused by missing pixels. Some extra voxels around a target object are reconstructed because all silhouette images have many extra pixels around the object. Although voxel coloring can carve voxels corresponding to the background using color similarities, the reconstruction result in figure 2 shows many extra voxels because the background has homogeneous colors.

Through this experiment, we know that for accurate volumetric 3D reconstruction, we have to reduce missing pixels and extract accurate boundaries of a target object. Since the visual hull is reconstructed by the intersection of the reprojected silhouette images, all of silhouette boundaries need not to be accurate. This fact makes us decide a measure not in the image but in the reconstructed visual hull.

3 Change detection method using a statistical model and the graph cuts based optimization

Our change detection method [1] is designed to satisfy the requirements described in the previous section. To reduce missing pixels, our method applies graph cuts based optimization to an initial result obtained by simple thresholding. All nodes have energies corresponding to color difference between background and object images. Since each node is linked with neighborhood nodes, some nodes that can be missed by an image noise in simple thresholding step are recovered by cutting process.

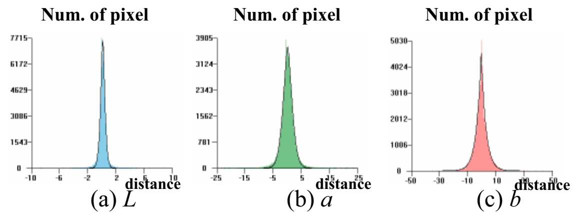


Figure 3. The estimation result of each channel in the CIELab color space

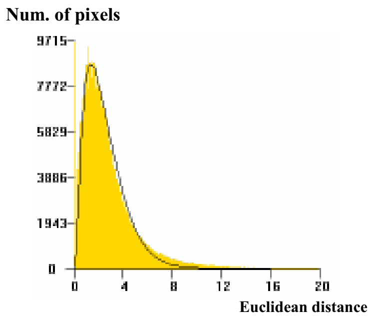


Figure 4. The estimation result of the Euclidean distance in the CIELab color space

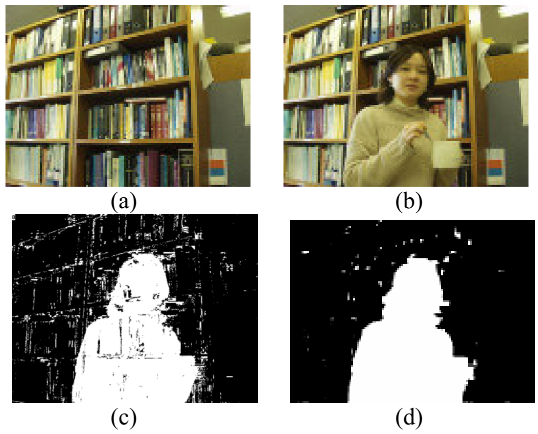


Figure 5. Change detection result: (a) background (b) foreground (c) simple thresholding using a statistical modeling (d) region-based optimization based on graph cut

To obtain accurate boundaries in the visual hull, we propose an accurate thresholding strategy based on a statistical modeling of the image noise. Since edge pixels produce remarkable differences compared with other pixels, we can detect accurate boundaries simply by thresholding the differences based on an estimated distribution of the image noise. Our measure to separate an image into background and object regions is the Euclidean distance of three channels in color space. This Euclidean distance corresponding to the image noise combines distributions corresponding to the each channel noise. We propose a generalized exponential distribution, called the Generalized Exponential Model (GEM) as the noise distribution for each color channel. Figure 3 shows the channels' histogram in the CIELab color space. Based on the estimated noise characteristic of each channel, we finally estimated the Euclidean distance of each channel noise as shown in figure 4.

In figure 5, we show the performance of our change detection algorithm. In figure 5(c), we can confirm that there are some missing and extra pixels, but a boundary of a person is preserved accurately by simple thresholding using our statistical modeling. Then, we reduce all of missing pixels in object regions by graph cuts based optimization. As shown in figure 5, our change detection can reject missing pixels by graph cuts and detect accurate boundaries by proposed statistical modeling.

The number of parameters in our method has been kept minimal to enable automatic silhouette extraction. For example, thresholds are automatically determined from the fitted histograms of the image noise. Some parameters used in graph cuts are not sensitive to the final extraction results. Therefore, we fixed them in the experiments.

4 Experimental results

In this section, we apply our method to the extraction of silhouettes for a set of background and foreground images. We compare our result with a well-known change detection algorithm, Integrating Intensity and Texture Differences (IITD) [6]. Then, we compare the 3D reconstruction results using silhouettes by the proposed algorithm with the results obtained manually and by IITD. We evaluate our result in terms of the requirements such as no missing pixel, an accurate boundary in the visual hull. The test images for our experiments have 640x480-pixels with 24 bit RGB color bands as shown in figure 6. For 3D reconstruction, voxel size is 2mm and voxel space is set to 200x200x200.

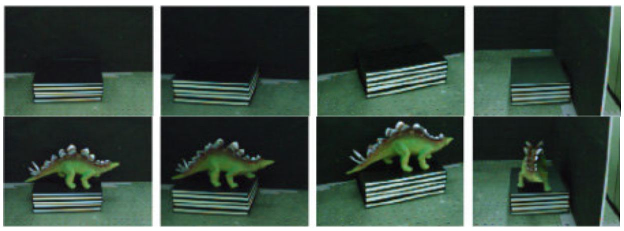


Figure 6. Test images for silhouette extraction

4.1 Silhouette extraction results for multiple cameras

Figure 6 shows the extracted silhouette of the test images. The result of IITD has many missing pixels so that it leads to many holes in the reconstruction result. Our final results shown in figure 7(d) are close to the ground truth shown in figure 7(a). Graph cuts based optimization reduces error pixels caused by the image noise and self-shadows. The shadow region is extracted as silhouette pixels in the fourth image of figure 7(d). However, since this shadow isn't extracted in other images, it will not contribute to the reconstruction.

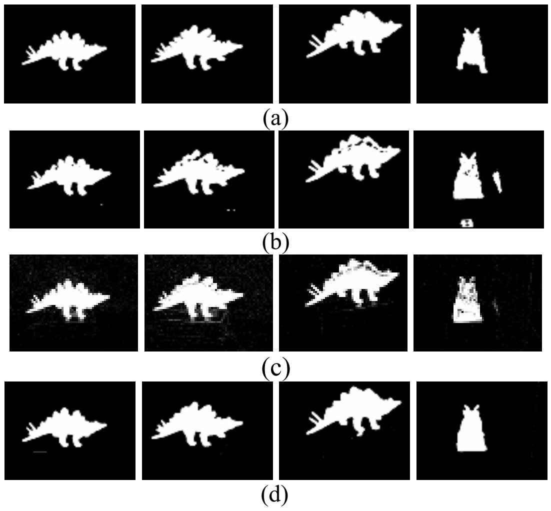
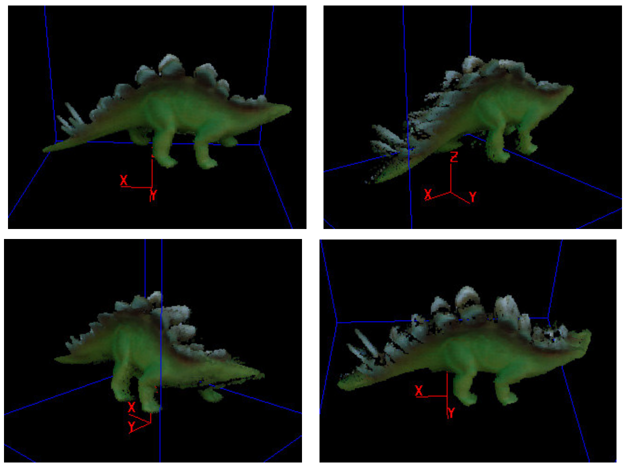


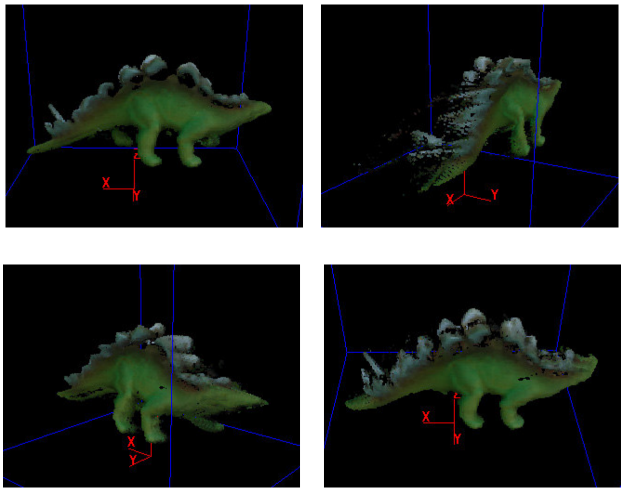
Figure 7. The result of change detection applied to the image set in figure 6: (a) ground truth (b) IITD (c) our method (simple thresholding) (d) our method (region-based optimization)

4.2 3D reconstruction results

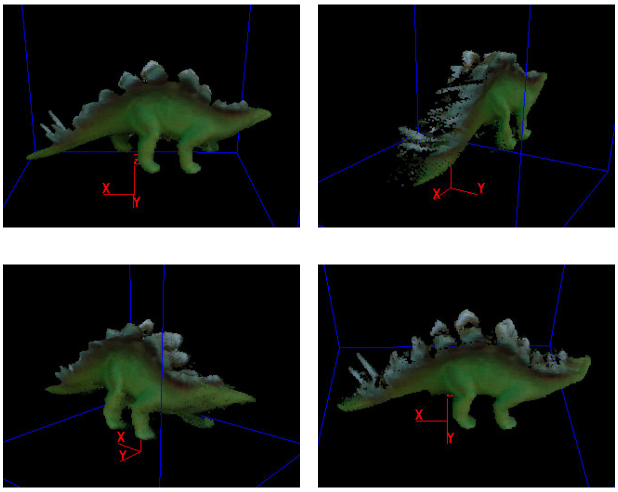
The silhouettes in the previous section can reconstruct visual hull. We compare the 3D reconstruction results using silhouettes by our algorithm with the results obtained manually and by IITD in figure 8. The result by IITD has many holes by missing pixels and some parts not are reconstructed properly by inaccurate boundaries of silhouettes. The result by our method shows good result which is close to the result by the ground truth silhouette. Our result has no hole in the object region because our silhouettes have no missing pixel.



(a)



(b)



(c)

Figure 8. The visual hull reconstruction results using obtained silhouette: (a) ground truth (b) IITD (c) our method (region-based optimization)

4.3 Evaluation of the requirements

We define a pixel as an object pixel when all of five by five neighborhood pixels are silhouette pixels. If we miss an object pixel, we regard it as a missing pixel. As shown in table 1, our proposed method has a smaller number of missing pixels compared with IITD.

Table 1. Missing pixels in silhouette images

	IITD	Proposed
Camera 1	164	0
Camera 2	983	3
Camera 3	1810	13
Camera 4	1876	15
Camera 5	575	6

We compare the reconstructed visual hulls with the visual hull by the true silhouettes for the accuracy evaluation of a boundary in the visual hull. We have much less extra and missing voxels than IITD as shown in table 2. As a matter of fact, manual silhouette extraction cannot guarantee the absolute true value because the boundary pixel between object and background is ambiguous. Therefore, we only use silhouettes obtained manually as a reference. Since our reconstruction result has no hole and detects accurate silhouette boundaries, the 3D shape by our method is not very different from the result by true silhouettes. It means that our result may be even closer to the true shape than the shape by manual extraction.

Table 2. Reconstructed visual hulls comparison

	True silhouette	By IITD	By our method
Total voxels	179065	182976	182000
Missing voxels	0	40002	7154
Extra voxels	0	43913	10089

5 Conclusions

In this paper, we present a silhouette extraction method based on our change detection methodology. We define three requirements to extract silhouette pixels: no missing pixel, an accurate boundary in the visual hull and automatic extraction. A statistical modeling of image noise based on our proposed noise model, GEM, detects the accurate boundaries of silhouettes. To reduce missing pixels in object regions, we adopt graph cuts so that we can consider the effect of neighborhood pixels as well as the color difference. We can extract silhouette pixels that satisfy the requirements automatically. We have analyzed our proposed methodology experimentally. By 3D reconstruction results, we show that our method can extract silhouette pixels for the accurate 3D reconstruction.

References

[1] Y. Hwang, J. Kim, I. Kweon, “Change detection using a statistical model of the noise in color images”, IEEE Conference on Intelligent Robots and System, vol. 3, Sep. 2004, pp. 2713-2718,

[2] A. Laurentini, “The visual hull concept for silhouette-based image understanding”, IEEE Transaction on Pattern Analysis and Machine Intelligence, 16(2), pp. 150-162, Feb, 1994

[3] S. Seitz, C. Dyer, “Photorealistic scene reconstruction by voxel coloring”, IEEE Conference on Computer Vision and Pettern Recognition, pp. 1067-1073, June, 1997

[4] K. N. Kutulakos, S. Seitz, “A theory of shape by space carving”, IEEE Conference on Computer Vision, pp. 307-314, 1999

[5] H. Kim and I. Kweon, "Probabilistic optimization-based approach for the volumetric full-view scene recovery", To be published in International Journal of Computer Vision

[6] L. Li, M. K. H. Leung, “Integrating intensity and texture differences for robust change detection”, IEEE Transaction on Image Processing, Feb 2002, vol. 11, No. 2, pp. 105-112

[7] J. Sun, J. Jia, C. Tang, H. Shum, “Possion matting”, ACM SIGGRAPH 2004, August 2004, vol. 23, pp. 315-321

[8] C. Rother, V. Kolmogorov, A. Blake, “GrabCut : interactive foreground extraction using iteratived graph cuts”, ACM SIGGRAPH 2004, August 2004, vol. 23, pp. 309-314