VIEW SYNTHESIS OF ARTICULATING HUMANS USING VISUAL HULL

Zhanfeng Yue, Liang Zhao and Rama Chellappa

Center for Automation Research
University of Maryland, College Park, MD 20742
{zyue, rama}@cfar.umd.edu, lzhao@umiacs.umd.edu

ABSTRACT

In this paper, we present a method which combines image-based visual hull and human body part segmentation for overcoming the inability of the visual hull method to reconstruct concave regions. The virtual silhouette image corresponding to the given viewing direction is first produced with image-based visual hull. Human body part localization technique is used to segment the input images and the rendered virtual silhouette image into convex body parts. The body parts in the virtual view are generated separately from the corresponding body parts in the input views and then assembled together. The previously rendered silhouette image is used to locate the corresponding body parts in input views and avoid the unconnected or squeezed regions in the assembled final view. Experiments show that this method can improve the reconstruction of concave regions for human postures and texture mapping.

1. INTRODUCTION

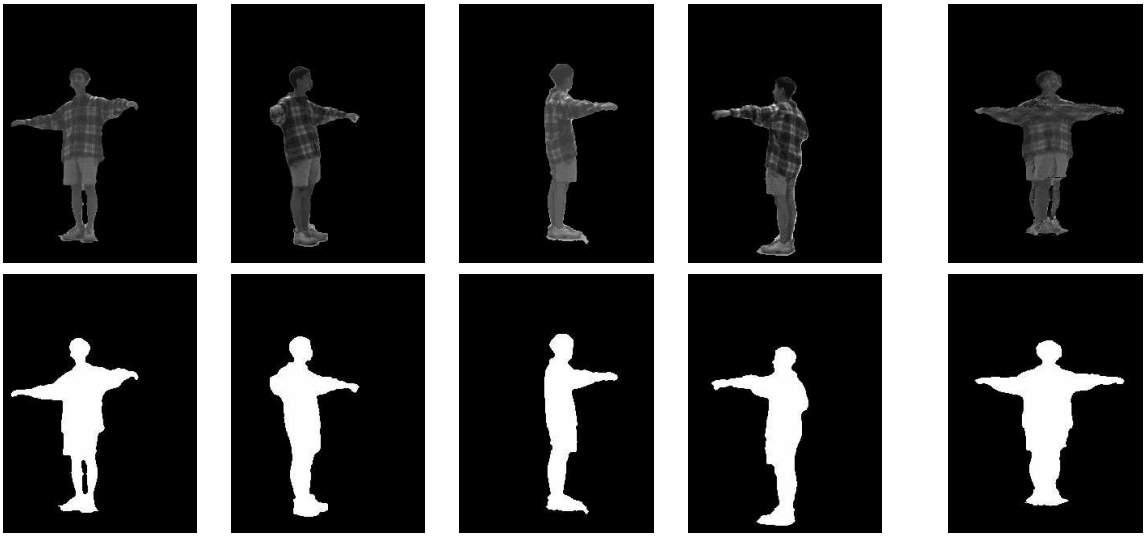
View synthesis is the technique of visualizing and manipulating the appearance of the object for a given viewing direction from several existing viewpoints. An effective and fast volumetric scene reconstruction method for view synthesis is shape from silhouettes in which the intersection of the generalized cones associated with a set of cameras defines a volume of scene space containing the object. The silhouette-based reconstruction encloses the true volume and only approximates the true 3D shape, depending on the number of views, the positions of the viewpoints, and the complexity of the object. In particular, the concave patches are not observable in any silhouette. In this paper, we present a method to overcome the inability of the shape from silhouettes method to reconstruct concave regions for human postures. We resort to contour based human body part segmentation method to disassemble each input silhouette image into convex parts and use image based visual hull to generate the novel view for each body part. These rendered parts are assembled to give the final result. To avoid the presence of possibly unconnected or squeezed regions in the final view, the virtual silhouette image is first generated without using body part segmentation method.

A visual hull of an object is the intersection of all the extruded cone-like shapes that result from lifting the silhouettes in all views [1]. Hence, visual hull can be obtained by volume carving. It is possible to reduce the computation of visual hull to 2D operations since it contains only points that project onto the silhouettes. Image based visual hull [2] is an efficient geometrically-valid pixel reprojection method to compute the visual hull. For each pixel in the desired view, the epipolar line in each input view is intersected with the contour approximation, then the intersected 2D line segment is projected back to 3D space to form the visual hull. The algorithm is able to render a desired view of n^2 pixels in $O(kn^2)$ where k is the number of input images. After the visual hull is constructed, its surface is texture mapped using the weighted sum of intensity values in the input images [3]. Considering the visibility during the texture mapping process, an occlusion-compatible warping ordering scheme [4] was used to solve the object occlusion and dis-occlusion problem. An advantage of the image based visual hull technique is its tradeoff between accuracy and efficiency. With the widely-positioned views as inputs, image based visual hull allows us to produce the virtual view without finding the wide baseline correspondence. It also provides information about the object's 3D shape and location. Besides, since the visual hull is formed by volume carving, the noise from input images is greatly reduced in the intersecting process.

Fig 1 shows an example of view synthesis with image based visual hull. We can observe from Fig 1 that the person stands with a 3D concave posture which is formed by the stretching of arms and torso. Although the rendered silhouette image shown as the bottom image in Fig 1 (b) is correct (because the eye can be fooled into perceiving convex and concave regions with only silhouette images), the error coming from the concave regions can be easily observed on the texture-mapped chest part in the top image in Fig 1 (b). [1] stated that the visual hull of an object depends not only on the object itself but also on the region allowed to the viewpoint. The *external visual hull* is related to the convex hull, and the *internal visual hull* can not be observed from any viewpoint outside the convex hull.

Observing that in many cases the concave human posture is formed due to the position of arms, we are inspired to explore the possibility of body part based view synthesis with visual hull. Several methods have been proposed for human body part segmentation from silhouette (contour). The work in [5] gives a silhouette-based human body labelling template by using topological order-constraints of body parts for different postures. A contour-based body part localization method was presented in [6] with a probabilistic similarity measure which combines the local shape and global relationship constraints to guide body part identification. More recently, a hierarchical model fitting method to estimate 3D shape with density fields was proposed in [7]. The body parts of the human can be described accurately with the estimated parameters. In this paper we use the work in [6] for body part segmentation because of its simplicity and robustness. The silhouette image in each input view is partitioned into arms and torso (with legs) so that each human part is a convex object. All the parts are separately processed with image based visual hull, and assembled together to

Partially supported by an ONR Grant N00014-01-1-0265



(a) input (b) output

Fig. 1. An example of image based visual hull: (a) the images observed from the 4 static cameras (top) and corresponding silhouette images (bottom). (b) The rendered image corresponding to a novel view with texture (top) and without texture (bottom) obtained with IBVH.

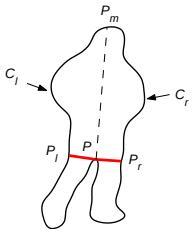


Fig. 2. Computing the cuts passing through point P

get the final result. It is possible that the final view has some unconnected or squeezed regions since it is obtained by "stitching" the separately processed body parts. To prevent this problem, a silhouette image for the desired viewing direction is first generated without segmenting the body parts.

The paper is organized as follows: In section 2 we introduce the body part segmentation method which utilizes the silhouette image and considers three factors affecting the saliency of a part. Section 3 presents the body part based visual hull formulation process, in which the desired silhouette image is first generated using the image based visual hull and then helps to find the corresponding body parts in input views and prevents the unconnected or squeezed region for the assembled final view. Conclusions are given in section 4.

2. BODY PART SEGMENTATION

In order to remove the errors of texture mapping from a non-convex shape, we break the silhouette of a person into convex parts. We segment a human body into parts at *negative minima of curvature* so that the decomposed parts are convex regions. Singh *et al.* noted that when boundary points can be joined in more than one way to decompose a silhouette, human vision prefers the partitioning scheme which uses the shortest cuts (A cut is the boundary between a part and the rest of the silhouette). They further restrict a cut to cross a symmetry axis in order to avoid short but undesirable cuts. However, most symmetry axes are very sensitive to noise and are expensive to compute. In contrast, we use the constraint on the saliency of a part to avoid short but undesirable cuts. According to Hoffman and Singh's [8] study there are three factors that affect the saliency of a part: the size of the part relative to the whole object, the degree to which the part protrudes, and the strength of its boundaries. Among these three factors, the computation of a part's protrusion (the ratio of the perimeter of the part (excluding the cut) to the length of the cut) is more efficient and robust to noise and partial occlusion of the object. Thus, we employ the protrusion of a part to evaluate its salience; the salience of a part increases as its protrusion increases.

In summary, we combine the short-cut rule and the salience requirement to constrain the other end of a cut. For example in Fig 1, let S be a silhouette, C be the boundary of S, P be a point on C with negative minima of curvature, and P_m be a point on C so that P and P_m divide the boundary C into two curves C_l, C_r of equal arc length. Then two cuts are formed passing through point P: $\overline{PP_l}$, $\overline{PP_r}$ such that points P_l and P_r lies on C_l and C_r, respectively. The ends P_l and P_r of the two cuts are located as follows:

$$P_l = arg\min_{P'} \|\overline{PP'}\| \quad \text{s.t.} \quad \frac{\|\overset{\frown}{PP'}\|}{\|\overline{PP'}\|} > T_p, P' \in C_l, \overline{PP'} \in S \tag{1}$$

$$P_r = arg\min_{P'} \|\overline{PP'}\| \quad \text{s.t.} \quad \frac{\|\overset{\frown}{PP'}\|}{\|\overline{PP'}\|} > T_p, P' \in C_r, \overline{PP'} \in S \tag{2}$$

where $\overset{\frown}{PP'}$ is the smaller part of boundary C between P and P', $\|\overset{\frown}{PP'}\|$ is the arc length of $\overset{\frown}{PP'}$, and $\frac{\|\overset{\frown}{PP'}\|}{\|\overline{PP'}\|}$ is the salience of the part bounded by curve $\overset{\frown}{PP_l}$ and cut $\overline{PP_l}$.

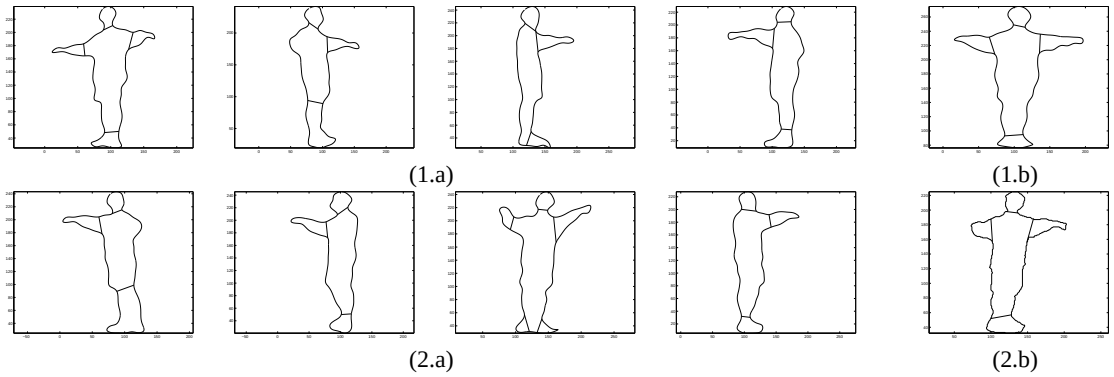


Fig. 3. Two examples of human body part segmentation results: (1.a) and (2.a) are the body part segmentation results for input views. (1.b) and (2.b) are the body part segmentation result for rendered silhouette images.

Eq. (1) means that point P_l is located so that the cut $\overline{PP_l}$ is the shortest one among all cuts sharing the same end P, lying within the silhouette with the other end lying on contour C_l, and resulting in a significant part whose salience is above a threshold T_p. The other point P_r is located in the same way using Eq. (2).

Since negative minima of curvature are obtained by local computation, their computation is not robust in real digital images. We take several computationally efficient strategies to reduce the effects of noise. First, a B-spline approximation is used to moderately smooth the boundary of a silhouette, since B-spline representation is stable and easy to manipulate locally without affecting the remaining part of the silhouette. Second, the negative minima of curvatures with small magnitudes are removed to avoid parts due to noise or small local deformations. However, the curvature is not scale invariant (e.g. its value doubles if the silhouette shrinks by half). One way to transform curvature into a scale-invariant quantity is to first find the chord joining the two closest inflections which bound the point, then multiply the curvature at the point by the length of this chord. The resulting normalized curvature does not change with scale — if the silhouette shrinks to half size, the curvature doubles but the chord halves, so the product is constant.

3. VIEW SYNTHESIS OF ARTICULATING HUMANS USING VISUAL HULL

Having segmented each input image into convex body parts, we need to render the image for each body part in the given viewing direction and assemble them together. In order to generate each body part separately for the desired view, we have to use the corresponding body part in each input image. Since the body part localization method in previous section does not give such corresponding relationship between views, we can not tell which body part is left arm and which one is right arm from the input silhouette images. In the assembling process, it is possible that the "stitched" final view has unconnected or squeezed regions because the separately-generated virtual parts are not guaranteed to match each other.

To solve these two problems, a virtual silhouette image corresponding to the given viewing direction is first generated using the image based visual hull computed from the input silhouette images. In this process, we only need to decide whether each pixel in the virtual view belongs to the foreground or the background. If a pixel's corresponding 3D ray intersection in the visual hull formulation process is not null, the pixel is marked as a foreground pixel and the intersection coordinates are stored in a table for later use. Each input image is segmented into left arm, right arm and torso (with legs). The rendered silhouette image can also be segmented into body parts in the same way. Since the visual hull of the person has been built, the 3D centroid for each body part can be roughly approximated with the center of gravity of the body part's visual hull. By projecting the 3D centroid to each input image, we are able to locate the corresponding body part in each input image for the rendered body part in the synthetic image.

To map the texture for the foreground pixels in the desired view, a nearest neighboring scheme is used [2]. For each foreground pixel, the 3D closet frontal point is retrieved from the stored table and projected onto each input view. The intensity value P for the desired view pixel is a weighted sum of intensity values P_i of the corresponding pixels in the input views, $P = \sum P_i \cos\theta_i$, where θ_i is the angle between the 3D ray from the desired view foreground pixel and the 3D ray from the corresponding pixel in input view i if the closet frontal point is visible in this view. If the concave regions are not considered in the formulation of the visual hull, the pixels in the desired view projected by the points inside the concavities will have erroneous 3D closet frontal points and their intensity values will be wrong. In order to obtain correct visual hull and texture mapping result, the human body part segmentation method is used in the reconstruction process. For the desired view, each foreground pixel in a segmented body part will have its epipolar line intersected with the corresponding body part contour in each input view. These 2D line intersections are projected back into 3D space and intersect with the retrieved 3D ray starting from the pixel in the desired view. If the pixel is the projection of a 3D point which lies on the concave region, the new 3D ray intersection will be shorter compared to the previously-stored intersection because the epipolar line only intersects with the corresponding body part instead of the whole body contour. Hence, the 3D closet frontal points for these pixels are closer to their correct positions so that their intensity values can be decided with the correct corresponding pixels in the input views. For the pixels corresponding to the 3D points which do not lie on concave regions, the 3D ray intersections are same as the stored ones. In this way, even if the epipolar line of a pixel in a desired view body part has no intersection with the corresponding body part contour in the input views, this pixel is still marked as a foreground pixel and has its intensity value decided using the nearest neighboring

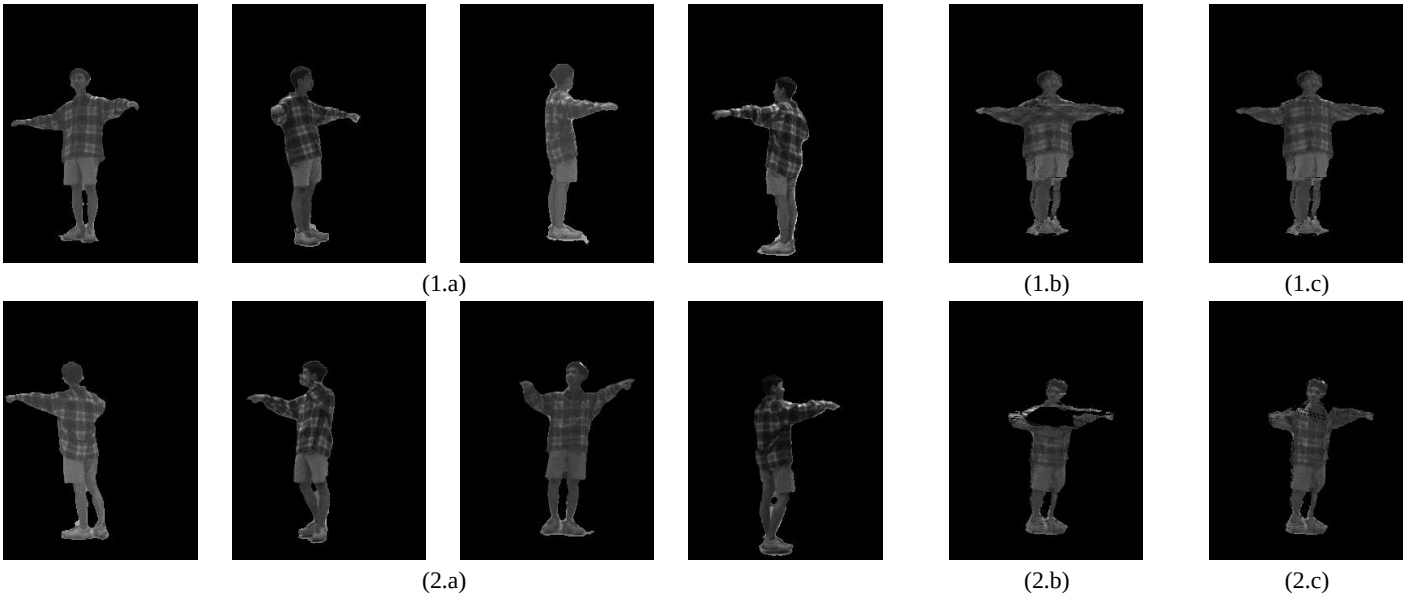


Fig. 4. Two examples of view synthesis of articulating humans with visual hull: (1.a) and (2.a) are the input views. (1.b) and (2.b) are the texture mapping results without using body part segmentation method. (1.c) and (2.c) are the texture mapping results using body part segmentation segmentation method.

scheme. Therefore, no unconnected region will be observed in the assembled view. Since the separately processed body parts are segmented from the previously generated silhouette image, no region will be squeezed together in the assembled view.

The body body segmentation results for four input views and the rendered silhouette image are shown in Fig 3. The texture mapping results obtained with and without using the body part segmentation method are shown in Fig 4. The hole on the chest part of Fig 4 (2.b) is because the concave region formed by the arms and the torso is treated as a convex region. Since the desired viewing direction is from above the concave region while the input viewing directions are either from below the concave region or make the concave region occluded, so the front-most points corresponding to these pixels are not visible in any of the input views and marked as invisible. From Fig 4 (1.c) and Fig 4 (2.c) we can observe that the texture mapping results are greatly improved with body part based method being used. It should be mentioned that if the desired viewing direction makes the rendered image have self occlusion between the limbs and the torso, the rendered image has no obvious improvement compared to the result obtained without using the body part based method.

4. CONCLUSION

We have presented a method which combines image-based visual hull and human body part segmentation for overcoming the inability of the visual hull method to reconstruct concave regions for human postures. A contour-based human body part segmentation method is introduced and used to segment the input images and the previously rendered silhouette image into convex body parts. The body parts in the desired view are generated separately from the corresponding body parts in the input views and are assembled together to give the final view. Experiments show that this method can improve the reconstruction of concave regions for human postures and the texture mapping result.

5. REFERENCES

[1] A. Laurentini, "The visual hull concept for silhouette-based image understanding," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 16, no. 2, pp. 150–162, 1994.

[2] W. Matusik, C. Buehler, R. Raskar, S. Gortler, and L. McMillan, "Image-based visual hulls," *Proc. SIGGRAPH 2000*, pp. 369–374, 2000.

[3] W. Matusik, C. Buehler, and L. McMillan, "Polyhedral visual hulls for real-time rendering," *Proc. Eurographics Workshop on Rendering '01*, 2001.

[4] L. McMillan, *An Image-Based Approach to Three Dimensional Computer Graphics*, Ph.D. thesis, University of North Carolina, Chapel Hill, NC, 1997.

[5] I. Haritaoglu, D. Harwood, and L. Davis, "Ghost: A human body part labeling system using silhouettes," *Proc. Int. Conf. on Pattern Recognition*, pp. 77–82, 1998.

[6] L. Zhao, *Dressed Human Modeling, Detection, and Parts Localization*, Ph.D. thesis, Carnegie Mellon University, Pittsburg, PA, 2001.

[7] E. Borovikov and L. Davis, "3d shape estimation based on density driven model fitting," *Proc. The 1st International Symposium on 3D Data Processing, Visualization and Transmission (3DPVT)*, 2002.

[8] D. Hoffman and W. Richards, "Salience of visual parts," *Cognition*, vol. 63, pp. 29–78, 1997.