3D Shape Reconstruction from Incomplete Silhouettes in Multiple Frames

Masahiro Toyoura Masaaki Iiyama Takuya Funatomi Koh Kakusho Michihiko Minoh
Academic Center for Computing and Media Studies
Kyoto University

Abstract

3D shapes are reconstructed from silhouettes obtained by multiple cameras with the volume intersection method. In recent work, methods of integrating silhouettes in time sequences have been proposed. The number of silhouettes can be increased by integrating silhouettes in multiple frames. The silhouettes of a rigid object in multiple frames are integrated with its rigid motion. This motion is often estimated with 3D feature points extracted from silhouettes. When the estimated motion has large error, shapes are reconstructed with missing parts. This error is given by the incomplete extraction of 3D feature points, which is caused by additional and missing regions of extracted silhouettes. We cannot prevent silhouettes from being extracted with the additional and missing regions in real environments.

Here, we propose an intelligent method of integrating incomplete silhouettes where outcrop points, which are 3D feature points for estimating motion, play an important role. The reconstructed shape can be evaluated referring to how many outcrop points have been included in the reconstructed shape of another frame. Although the evaluation does not represent the accuracy of estimated motion directly, it does guarantee that outstanding parts will be preserved in the reconstructed shape. Silhouettes in multiple frames can be integrated with fewer missing and additional parts based on this evaluation.

1 Introduction

Shapes of objects are reconstructed from silhouettes with the volume intersection method [5]. The silhouettes are extracted from images obtained by multiple cameras. The reconstructed shapes are called *Visual Hulls*, or *VH*. The volume intersection method is not affected by colors or surface characteristics of objects because the method reconstructs the shapes of the objects from their silhouettes only.

The visual hulls include many additional regions when few cameras are used in the volume intersection method. These are regions that are not included in the object region. The additional regions are decreased by increasing the number of cameras. Decreasing the additional regions means that the reconstructed shapes become more accurate. However, it is not realistic to install too many cameras around the object. The number of cameras is limited by their physical size, the space for setting them up and the cost. A method of reconstructing shapes that integrates silhouettes in multiple frames has been proposed to overcome these limitations [1, 6].

Let us assume that the object is moving rigidly and cameras change their relative positions to it in every frame. If the rigid motion of the object can be estimated, the images captured by the cameras in different frames can be treated virtually as those at different positions. More accurate shapes can be reconstructed with these virtual images, without physically increasing the number of cameras.

A problem with the volume intersection method in multiple frames is to reconstruct shapes with large missing parts, when the motion of the object is estimated with large errors. The reconstructed shape in multiple frames is calculated as the intersection of visual hulls of all frames. Even if only a few frames have large errors, these are accumulated into the reconstructed shape in multiple frames. These are caused by the failure to extract 3D feature points due to missing or additional regions in extracting silhouettes. The CSPs [1] and outcrop points [6], which are 3D feature points, are extracted assuming that the silhouettes are completely extracted.

We propose a method of suppressing missing reconstructed shapes in multiple frames. Although the method does not solve incompleteness of silhouettes or error of estimating motion directly, it does guarantee that outstanding parts will be preserved in the reconstructed shape. These outstanding parts characterize what the object is [4].

We focused on the fact that the outcrop points tended to be extracted from outstanding parts of the object. Using this fact, we could select frames that retained outstanding parts and process those into the reconstructed shape. We defined a function for measuring how the outstanding points were

retained in the reconstructed shape by integrating the visual hulls of the frames. By only integrating the visual hulls of frames with high scores, the shape in multiple frames was guaranteed to be reconstructed by including the outstanding parts.

2 Reconstruction of Shape from Silhouettes in Multiple Frames

Let us denote the cameras that are placed around target object O to capture it by $C_j(j = 1, \cdots, N)$, where N denotes the number of cameras ($N > 1$). All the cameras observe the object synchronously. Time i can be replaced as the i-frame. The 2D region that corresponds to object O is extracted from the images of C_j. The projection matrix of C_j is represented as P_j. This region is called the *silhouette* and denoted by S_{ij}. When the object is in motion, observed silhouettes differ among frames. O is guaranteed to be included in a cone with the apex at the optical center of C_j and the base at S_{ij} is then calculated. This cone is called the *visual cone* of camera C_j, and denoted by $V_{ij} = \{v \mid P_j v \in S_{ij}\}$, where v represents the occupation of a small 3D region, or a *voxel*. *Visual hull* V_i of the i-frame is defined as the intersection of visual cones $V_{i1}, \cdots, V_{iN}$.

Let the object move rigidly and its motion be known. An intersection of visual hulls $V_i(i = 1, \cdots, M)$ means the shape derived from silhouettes of all frames. When the k-frame is selected as a base frame, intersection V^k is calculated by the estimated motion, D_{ik}, between the k-frame and i-frame $(i = 1, \cdots, M)$. Intersection V^k is called an integrated shape.

$$V^k = \{v \mid \forall i, D_{ik} v \in V_i\}. \tag{1}$$

To estimate the motion, some kinds of 3D feature points are required to be extracted from obtained images. The *outcrop point (OP)* [6] is a 3D feature point that is extracted from silhouettes of multiple cameras. Although the frontier point [3] is also a 3D feature point, it is not guaranteed to be included in the object region of the visual hull completely.

When voxel v in the visual hull satisfies the conditions we call v an *outcrop point*:

1. When v is projected onto the image plane of each camera, the projected pixel of v is in the contour of the silhouette for at least one camera.

2. For each camera satisfying the condition above, no other voxel of the visual hull is projected to the pixel.

Theoretically, the outcrop point is guaranteed to be included in the object region. If outcrop point v satisfies condition 1 despite the fact that v is not actually included in the object region, any other voxels are required to be projected to the pixel to which v is projected. Due to condition 2, no other voxels are projected to the pixel. A pixel to which no voxels are projected is not a element of the silhouette.

The outcrop points are often extracted from the outstanding parts of the object. The corresponding outcrop points are robustly extracted even though the relative position between the object and cameras is changed.

3 Evaluation Function Based on Preserving Outcrop Points

When the rigid object motion, D_{ik}, between the i-frame and k-frame is correctly estimated, the outcrop points, OP_i, which are extracted in the i-frame, are included in V_k by translating D_{ki}, because V_k should include the whole object region and OP_i should be included in the object region. Similarly, the outcrop points, OP_k, which are extracted in the k-frame, are included in V_i by translating D_{ik}.

$$p_i \in OP_i, D_{ik} p_i \in V_k \quad and \tag{2}$$
$$p_k \in OP_k, D_{ki} p_k \in V_i. \tag{3}$$

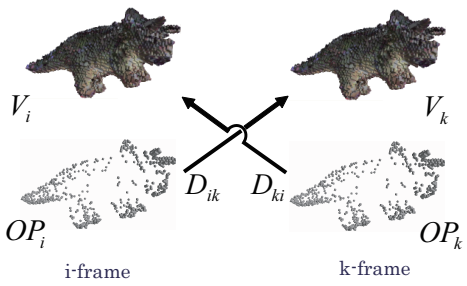


Figure 1. Evaluation of frames from visual hulls and outcrop points.

If Eqs. (2) and (3) are completely satisfied, all outstanding parts of the object shape will be completely included in the reconstructed shape in multiple frames. They are not completely satisfied in real environments, because there are additional and missing parts of visual hulls or errors in the estimated motion. To evaluate how many outstanding parts are included in the reconstructed shape, we can utilize the rate of outcrop points that satisfies Eqs. (2) and (3). The rate, $E_m(i, k)$, is defined by

$$E_m(i,k) = \frac{n_{ik} + n_{ki}}{2}, \tag{4}$$
$$n_{ik} = \frac{n\{p_i | p_i \in OP_i, D_{ik} p_i \in V_k\}}{n\{p_i | p_i \in OP_i\}},$$

$$n_{ki} = \frac{n\{p_k | p_k \in OP_k, D_{ki} p_k \in V_i\}}{n\{p_k | p_k \in OP_k\}}.$$

Here, $n\{\cdot\}$ is the number of voxels included in a set. OP_i and OP_k are sets of outcrop points in the i-frame and k-frame. V_i and V_k are sets of voxels included in the visual hulls of the i-frame and k-frame.

The evaluation function, $E_m(i,k)$, ranges from 0 to 1. When $E_m(i,k)$ indicates a large value, the integrated shape preserves many outcrop points within it. Conversely, a small value for $E_m(i,k)$ means that the integrated shape has lost most of the outcrop points. By only using frames where $E_m(i,k)$ has large values, the integrated shape can preserve outstanding parts. If threshold E_m^{th} is given, appropriate frames can be selected by $E_m(i,k) > E_m^{th}$. When the 0-th frame is chosen as the base frame, the frames that satisfy $E_m(0,i) > E_m^{th}$ are selected. Relabeling the frames as $i'(i' = 1, \cdots, M')$, the integrated shape is calculated as an intersection of the visual hulls of $V_{i'}$.

4 Experimental Results

The experimental results for simulated and real objects were used to evaluate how valid our proposed function, $E_m(i,k)$, was. We examined whether the outstanding parts of the object were preserved.

We obtained silhouettes from the simulation data of a triceratops toy. We arranged 12 cameras to observe the toy. We adopted our proposed method of volume integration for the silhouettes. The experimental results are presented in Figure 2. The silhouettes have regions randomly added and missed in specified percentages. E_m^{th} was set to 0.97. The visual hulls in one frame (II) include many additional regions on their surfaces, which angulate visual hulls. Some additional regions are floating away from the visual hulls. The shapes integrated without our method (III) have many missing parts. This is caused by frames that have large errors in estimating motion. The integrated shapes (IV) have accurate shapes with our method. The integrated shape includes the original regions. To conclude, our proposed evaluation function, $E_m(i,k)$, accurately preserves the outstanding parts of the object when the percentages of the missing and additional parts are small enough.

We captured a triceratops toy and a mammoth toy with multiple cameras in real world. Their shapes were reconstructed from the silhouettes in multiple frames. The integrated shapes with and without our proposal method are shown in Figures 3(a) and 3(b). Averages of 8.54% and 3.22% of the silhouettes were missing. Additional silhouettes were 5.27% and 4.29%. The silhouettes were refined with a silhouette refining method [7]. The threshold E_m^{th} was set to 0.95. There were some small missing parts in the integrated shapes obtained with the conventional method of integration (III). The feet of the triceratops and the head of the mammoth have gaps. The integrated shapes with our method (IV) did not have such missing parts. Outstanding parts of the objects were preserved. Compared with the shapes reconstructed from silhouettes in one frame (II), the areas of additional regions on the surface of the shapes are decreased.

5 Conclusions and Future work

We proposed an intelligent method of integrating silhouettes in multiple frames, which enabled us to reconstruct accurate shapes even if there were missing and additional parts in the silhouettes. We designed an evaluation function, $E_m(i,k)$, which indicates how many outcrop points are preserved in the integrated shape. Some integrated shapes with $E_m(i,k)$ were presented as experimental results. These shapes included outstanding parts of the object. We solved the problem where integrated shapes were missing when motion was incompletely estimated. The shapes were also more accurate than when the visual hull was calculated in one frame.

In future work, we intend to set threshold E_m^{th} automatically, which will be set by the percentages of missing and additional parts of silhouettes, or the numbers of available frames.

References

[1] G. K. M. Cheung, S. Baker, and T. Kanade. Shape-from-silhouette across time part i: Theory and algorithms. *International Journal of Computer Vision*, 62(3):221–247, May 2005.

[2] G. K. M. Cheung, T. Kanade, J.-Y. Bouguet, and M. Holler. A real time system for robust 3d voxel reconstruction of human motions. In *IEEE Conference on Computer Vision and Pattern Recognition*, volume 2, pages 714–720, June 2000.

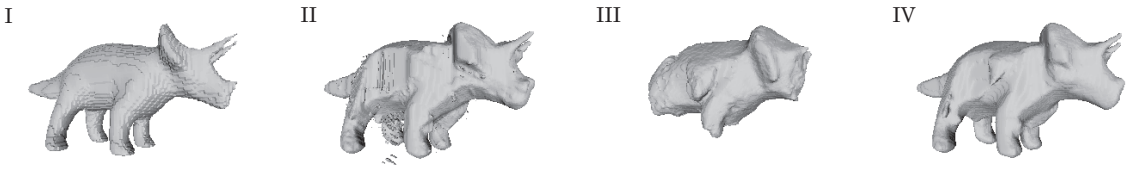
[3] R. Cipolla, K. Astrom, and P. Giblin. Motion from the frontier of curved surfaces. In *IEEE International Conference on Computer Vision (ICCV)*, pages 269–275, 1995.

[4] C. Dorai and A. K. Jain. Shape spectrum based view grouping and matching of 3d free-form objects. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(10):1139–1146, 1997.

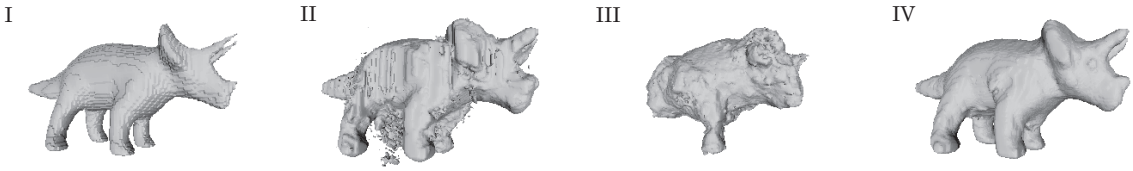
[5] W. Martin and J. Aggarwal. Volumetric description of objects from multiple views. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 5(2):150–158, 1983.

[6] M. Toyoura, M. Iiyama, K. Kakusho, and M. Minoh. Extraction of outcrop points from visual hulls for motion estimation. In *IEEE International Conference on Multimedia & Expo (ICME)*, pages 217–220, July 2006.

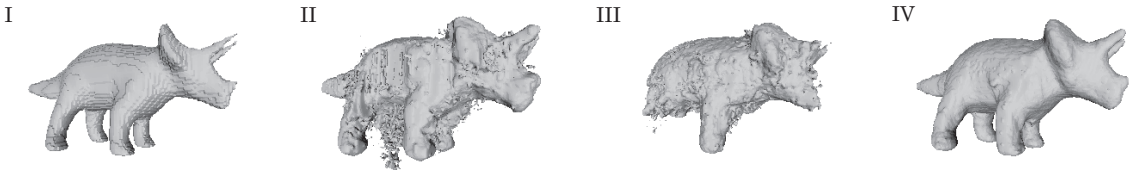
[7] M. Toyoura, M. Iiyama, K. Kakusho, and M. Minoh. Silhouette extraction with random pattern backgrounds for the volume intersection method. In *The 6th International Conference on 3-D Digital Imaging and Modeling (3DIM 2007)*, pages 225–232, August 2007.



(a) (Percent missing, Percent added)=(1%, 1%)



(b) (Percent missing, Percent added)=(2%, 2%)



(c) (Percent missing, Percent added)=(5%, 5%)

Figure 2. Reconstructed shapes of triceratops. Original shapes (leftmost, I) and visual hulls reconstructed from silhouettes of one frame with SPOT[2] (left, II). Visual hulls reconstructed from silhouettes of 50 frames with SPOT (right, III) and visual hulls reconstructed from silhouettes of 50 frames with SPOT and our proposed method (rightmost, IV).

(a) Reconstructed shape for triceratops toy.

(b) Reconstructed shape for mammoth toy.

Figure 3. Reconstructed shapes for real toys. Examples of obtained images (leftmost, I) and visual hulls reconstructed from silhouettes of one frame with SPOT[2] (left, II). visual hulls reconstructed from silhouettes of 50 frames with SPOT (center, III), visual hulls reconstructed from silhouettes of 50 frames with SPOT and our proposed method (right, IV), and colored ones with obtained images (rightmost, V).