EXTRACTION OF OUTCROP POINTS FROM VISUAL HULLS FOR MOTION ESTIMATION

Masahiro Toyoura†, Masaaki Iiyama‡, Koh Kakusho‡ and Michihiko Minoh‡

† Graduate School of Informatics, Kyoto University
‡ Academic Center for Computing and Media Studies, Kyoto University
†toyo@mm.media.kyoto-u.ac.jp, ‡{iiyama,kakusho,minoh}@media.kyoto-u.ac.jp

ABSTRACT

In this article, we discuss 3D shape reconstruction of an object in a rigid motion with the volume intersection method. When the object moves rigidly, the cameras change their relative positions to the object at every moment. To estimate the motion correctly, we propose new feature points called *outcrop points* on the reconstructed 3D shape. These points are guaranteed to be located on the real surface of the object.

If the rigid motion of the object can be correctly estimated, cameras at different moments serve as the cameras in different positions virtually. With these cameras in time sequences, we can increase accuracy of the reconstructed 3D shape without increasing the number of cameras. Based on this idea, we reconstruct an accurate shape of the object in motion from images obtained by limited number of cameras. As the result, we can acquire an accurate shape from images in time sequences.

1. INTRODUCTION

With the volume intersection method [5], a shape of a 3D physical object is reconstructed from silhouettes of obtained images. The reconstructed shapes are called *visual hulls*, or *VHs*. The silhouettes show regions which the object is projected. In the volume intersection method, textures of the objects are not used for the shape reconstruction. It results that the method can be applied even for textureless objects.

With the volume intersection method, additional regions, which are not included in the object's region, decrease with increasing the number of cameras. Decreasing the additional regions means that the reconstructed shapes become more accurate. However, it is not realistic to install so many cameras around the object due to constraints of mounting them.

Let us suppose that the object is in a rigid motion. When the object moves rigidly, cameras change their relative positions to the object at every moment. If the rigid motion of the object can be correctly estimated, images obtained by the cameras at different moments are treated as the images in different positions virtually. With these virtual images, an accurate VH is reconstructed without increasing the number of cameras. Based on the idea, Cheung et al. [1] have proposed a method for reconstructing an accurate shape from silhouettes in time sequences.

Frontier points [2] [3] [4] [6] [7] are generally used as the feature points for estimating the motion of the object. Cheung et al. also employ the frontier points. They are extracted from the VH. There is a tendency to be extracted a large number of the frontier points when the many cameras are used for shape reconstruction. The problem of the motion estimation with the frontier points is that it costs time to estimate the motion with the large number of the feature points. To solve the problem, we propose a new kind of feature points called *outcrop points*. The outcrop points form a subset of the frontier points. Imposing a new condition for the frontier points provides the outcrop points.

To estimate the motion, the feature points at every moment must be the same. When the feature points at a moment do not correspond to the feature points at the next moment, it is difficult to estimate the motion accurately. Since the VH includes the additional regions, the frontier points may be included in the additional regions. When they are included in the additional regions, they may not be extracted from the VH at another moment. Compared to the frontier points, the outcrop points are guaranteed to be included in the surface of the object. With the outcrop points, the outcrop points provide the motion estimation of the object with small number of points.

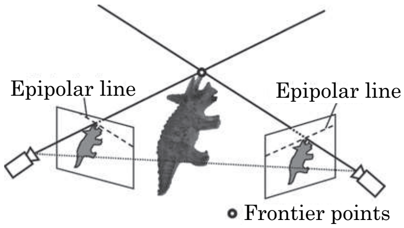
In Sections 2, we propose to extract new feature points, i.e., the outcrop points. Experimental results are given in Section 3, and future work is discussed in Section 4.

2. EXTRACTING FEATURE POINTS

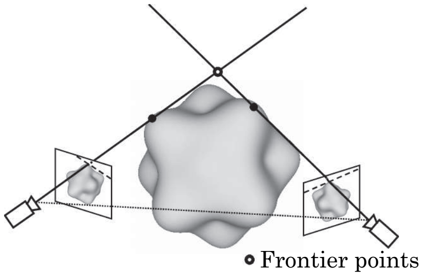
2.1. Conditions Required for Feature Points

In order to estimate the motion of the object, we have to extract some feature points from the VH at each moment.

Whereas the object region is included in the VH for all the time during the object moves, the additional regions around the object region change in time sequences due to change relative positions between the object and the cameras. The feature points that are not included in the object region may not be included in another VH obtained from another pair of silhouettes. The feature points have to be included in the object



(a) Extraction of frontier points



(b) Fault of extraction of frontier points

Fig. 1. Extraction of frontier points.

regions of the VH, and thus they must be included in the VH at any moment.

However, it is not easy to extract the points that are guaranteed to be in the object regions from the VH itself, because the additional regions around the object region change at every moment. Any point of the object regions may be occluded from the additional regions.

2.2. Frontier Points

In the previous work, there are some proposals for extracting the feature points from silhouettes. These feature points are called *frontier points* or *epipolar tangencies* [2] [3] [4] [6] [7]. In this subsection, we discuss the problem of these feature points.

When the object is observed by a pair of cameras, a point on the surface of the object and the optical centers of the two cameras form a plane called an epipolar plane. Based on this epipolar geometry, if the point constitutes the same epipolar plane together with a point on the contour of the silhouette as illustrated in Figure 1(a), the point can be considered to be in the object region. The point is called the frontier point.

The frontier point seems to be useful as a feature point that satisfies the condition of the previous subsection, which they must be included in the VH at any moment. However, the frontier point is not actually guaranteed to be included in the object region. Let us suppose the case in which an epipolar plain has more than one tangent point with the object region as shown in Figure 1(b). In this case, the point represented by an open circle in the figure is included in the VH constituted with the silhouettes of the two cameras. That point is regarded as the frontier point in spite that it is not actually included in the object region.

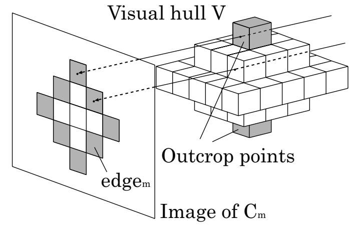


Fig. 2. Extraction of outcrop points.

The frontier points are extracted based on the epipolar geometry with each pair of cameras. When many cameras are used, many fake feature points are extracted. Let us denote the number of the feature points is represented by N, it costs time of $O(N^2)$ to estimate the motion of the object to make correspondence. To estimate the motion effectively, only the feature points that are included in the object regions must be extracted.

2.3. Outcrop Points

When a voxel v in the VH satisfies the following conditions, we define v as *outcrop point* as illustrated in Figure 2:

1. The projected pixel of v is included in the contour pixels of the silhouette of at least one image.

2. For each image satisfying 1, the projected pixel of v does not coincide with the projected pixel of any other voxel of the VH.

The condition 1. corresponds that of extraction of the frontier points. In Figure 2, $edge_m$ is a set of pixels on the contour of the silhouette, and V is the VH composed with the silhouettes of all the cameras.

In principle, the outcrop points are guaranteed to be included in the object region. If the voxel v satisfies 1 for an image, at least one voxel of the object region can be projected to the pixels to which v is projected. When v also satisfies the condition 2, v is only the voxel that is projected to the pixel. It means that v is guaranteed to be included in the object region.

The outcrop points are extracted based on the silhouettes and the VH. Even if the many cameras are used, no fake feature points are extracted as the outcrop points. Due to the difference between the principles of extraction, the number of the outcrop points is smaller than that of the frontier points.

3. EXPERIMENTAL RESULTS

In order to compare our proposed outcrop points with the frontier points, we show experimental results for some simulated shapes. Since the frontier points and the outcrop points are extracted from the VH, they are difficult to be extracted for smooth shapes, which do not represent their features on the obtained silhouettes. We employ the smooth spherical shapes as follows for our experiments.

$$\begin{cases} x' = cos\theta cos\phi \\ y' = sin\theta cos\phi \\ z' = sin\phi \end{cases}$$
$$(0 \leq \theta \leq 2\pi, -\pi/2 \leq \phi \leq \pi/2)$$
$$r' = A_l cos F_l \pi x' \cdot cos F_l \pi y' \cdot cos F_l \pi z'$$
$$\begin{cases} x = (r + r')cos\theta cos\phi \\ y = (r + r')sin\theta cos\phi \\ z = (r + r')sin\phi \end{cases}$$

where A_l denotes the amplitude of the fluctuation on surface, and F_l denotes its frequency. The radius of the sphere r is set to 50. Since the outstanding points are created by the fluctuation is controlled by A_ls and F_ls, the feature points are easy to be expected with larger A_l and larger F_l. Appropriate feature points give accurate reconstructed shapes by integrating images in time sequences.

The results for $F_l = 4, 6$ and $A_l = 2, 4$ are shown in Figure 3. In these experiments, the objects are translated by 1 voxel along X,Y,Z-axes and rotated by 2 degrees around Z-axis for each frame. 20 cameras are set on the vertices of a dodecahedron surrounding the object. We employ the VH by using 40000 cameras in order to evaluate the error of each experimental result. We call this shape a *correct shape* of the object in the volume intersection method, because concave surfaces of the object can not be reconstructed with volume intersection method.

Each experimental result is compared with the correct shape based on the three types of voxels: missing voxels, additional voxels and error voxels. The missing voxels are those included in the correct shape yet not in the result, whereas the additional voxels are those included in the result yet not in the correct shape. The error voxels are the summations of the number of the missing voxels and the additional voxels. Figure 3(c), (f), (i) and (l) illustrate each type of voxels.

If the outcrop points or the frontier points are used for the motion estimation, except for the object with $F_l = 4, A_l = 2$, the error voxels monotonically decrease. The error voxels for the object with $F_l = 4, A_l = 2$ increase after the number of the image frames used for shape reconstruction exceeds nine, because the surface of the object does not have sufficient number and length of salient points to keep extracted as the outcrop points.

For the VH of each object through all frames, the number of the frontier points is 27142 on average. Compared to the number of the frontier points, only 1182 points are extracted as the outcrop points on average. As the result, the outcrop points are superior to the frontier points, because both kinds of the points give almost same motion estimation and the number of the outcrop points is much smaller than that of the feature points.

From this experimental results, F_l should be larger than 6 or A_l should be larger than 4 in order to increase accuracy of the 3D shape with the outcrop points as well as with the frontier points. Since we set r to be 50, $A_l = 4$ corresponds to 8% of the length of the whole object, and 6.24 pixels in the silhouettes. In the shape with $F_l = 6$, salient points exist every 30 degrees on the surface. These are the conditions required for the object shapes for increasing accuracy with the outcrop points.

4. CONCLUSIONS AND DISCUSSION

In this paper, we introduce a method for increasing accuracy of the reconstructed shape by using the rigid motion of the object. In order to estimate the motion of the object from the VH at each moment correctly, we proposed new feature points called the outcrop points. In the experiments with the simulated objects, it is shown that the outcrop points are effective for the motion estimating of the objects. By integrating images in time sequences based on the estimated motion, the reconstructed shape is more accurate than that from images of a single frame. Compared from the frontier points, the out crop points represents the feature of the object with the small number of points.

Similar to extraction of the frontier points, extraction of the outcrop points is sensitive to the error of image processing for extracting the silhouettes. It is necessary to improve image processing for extracting the correct silhouette in real environment.

5. REFERENCES

[1] German K.M. Cheung, Simon Baker and Takeo Kanade, “Visual Hull Alignment and Refinement Across Time : A 3D Reconstruction Algorithm Combining Shape-From-Silhouette with Stereo, ” *Proc. CVPR*, pp.77-84, 2003.

[2] R.Cipolla, K.E.Astrom and P.J.Giblin, “Motion from the frontier of curved surfaces, ” *Proc. ICCV*, pp.269-275, 1995.

[3] Y. Furukawa, A. Sethi, J. Ponce and D. Kriegman, “Structure and Motion from Images of Smooth Textureless Objects, ” *Proc. ECCV*, vol. 2, pp. 287-298, 2004.

[4] C. Hernández Esteban and F. Schmitt, “Using Silhouette Coherence for 3D Image-based Object Modeling under Circular Motion, ” *Technical report 2003D011, Ecole Nationale Supérieure des Télécommunications*, September, 2003.

[5] W.N.Martin and J.K.Aggarwal, “Volumetric Description of Objects from Multiple Views, ” TPAMI, vol.5, no.2, pp.150–158, 1983.

[6] Dan Snow, Paul Viola and Ramin Zabih, “Exact Voxel Occupancy with Graph Cuts, ” *Proc. CVPR*, 2000.

[7] Sundar Vedula, Simon Baker, Steven Seitz and Takeo Kanade, “Shape and Motion Carving in 6D, ” *Proc. CVPR*, 2000.

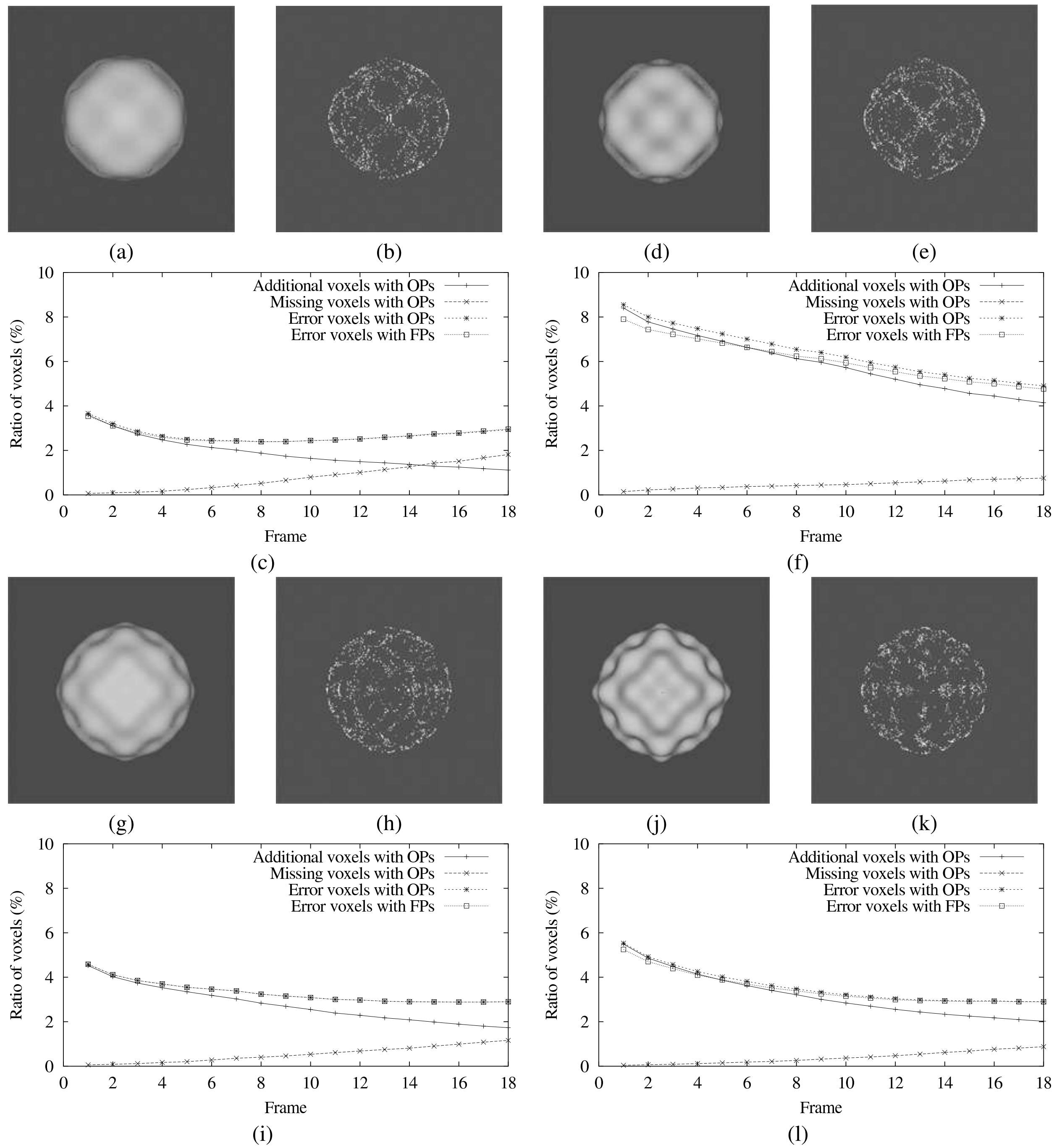


Fig. 3. Integratation for images of 18 frames : (a) An object shape of F_l=4, A_l=2. (d) F_l=4, A_l=4. (g) F_l=6, A_l=2. (j) F_l=6, A_l=4. (b), (e), (h) and (k) are outcrop points (OPs) for each shape. (c), (f), (i) and (l) show ratios of error voxels of integrated VHs. In (c), summations of additional voxels and missing voxels are increasing after 9th frame, even if the frontier points (FPs) are used. In (f), (i) and (l), the summations are decreasing until 18th frame with both kinds of feature points.