MVA2000 IAPR Workshop on Machine Vision Applications, Nov. 28-30, 2000, The University of Tokyo, Japan

13—18

A Unique Recovery of 3D Shape and Size from Motion

Akira Amano * Tsuyoshi Migita * Naoki Asada *
Department of Intelligent Systems
Hiroshima City University

Abstract

Factorization is one of the most practical method to recover 3D shape and motion simultaneously from 2D images with stable and fast computation. However, there still remain two crucial problems to be solved in real situations; one is to determine the true shape from a pair of visually equivalent candidates and the other is to measure the actual size of the object. This paper presents a method to solve the shape and size problems by the factorization with action; that is, 3D recovery is performed from an image sequence with a known trajectory of a single feature point given by the computer-controlled robot hand, and we determine the shape and size by evaluating the consistency between computed shape and the given trajectory. Experiments results performed in simulation study and in real world have shown the effectiveness of our method.

1 Introduction

A major concern in computer vision is to recover 3D shape from 2D images. Several methods have been developed so far to estimate depth information, from a set of images taken by varying a lens parameter in depth from focus[1], from a pair of images in shape from stereo[2], from a set of images taken under a structured illumination in photometric stereo[3], from a image sequence taken by a moving camera in factorization[4] or in active vision[5], and so on.

In all vision studies, problems are solved under an intrinsic condition; namely, nothing but the visual information is available. In other words, the observer is not permitted to alter the scene configuration but allowed to watch the scene.

In practical situations, however, the output of vision task such as 3D shape and position of an object is expected to be an input of action task such that the object is pushed, picked up or moved. This concatenation of successive two tasks suggests a new paradigm *Active Recognition* in which vision problems are solved with physical information obtained actively by touching or pushing the objects. In this paper, we present an active recognition approach to overcome the limitation of the conventional vision paradigm.

The factorization is one of the most practical method to recover 3D shape and motion simultaneously from 2D images with stable and fast computation. However, there still remain two crucial problems to be solved for the succeeding action task; that is,

- Shape uncertainty:

  While the factorization yields a unique solution mathematically based on the linear algebra, there are two candidates of 3D shape physically. This uncertainty comes from that the axis of depth is reversible in the orthographic projection. This is known as a Necker's reversal or depth reversal problem.

- Size uncertainty:

  The factorization tells no information about the actual size of the object. This is why the arbitrary scale factor is involved in the projective transformation.

This paper proposes a method to solve these shape and size problems by the factorization with action; that is, 3D recovery is performed from an image sequence with a known trajectory of a single feature point given by the computer-controlled robot hand, and we determine the shape and size by evaluating the consistency between computed shape and the given trajectory. In a situation where a remote robot having camera and manipulator that is controlled by a human operator, such a trajectory is once obtained by pushing an object through a human-aided operation, the robot can compute the precise shape and size of the object so that an action task such that the object is picked up would be performed.

In the following section, we first discuss the shape and size problems in the formulation of the factorization, then present the method to reach the unique

*Address: 3-4-1 Ozuka-Higashi, Asa-Minami, Hiroshima 731-3194, Japan. E-mail: {a-amano,migita,asada}@cv.its.hiroshima-cu.ac.jp

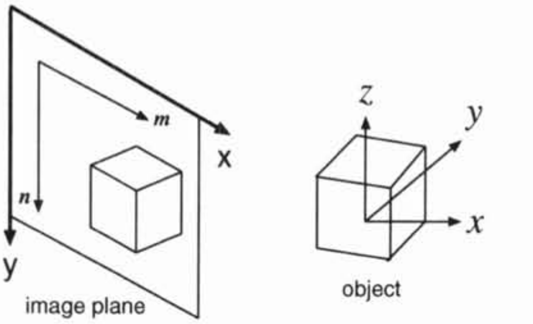


Figure 1: Image and object-centered coordinate systems.

solution by using a known trajectory. Finally, experimental results carried out in simulation as well as in real world will be shown to demonstrate the effectiveness of our method.

2 Solution of Shape and Size Problems

2.1 Shape and size uncertainties

Suppose a situation where an image sequence of a moving object is taken by a static camera and all feature points on the object are tracked throughout the image sequence.

Using the following notations, (Fig. 1)

- $\boldsymbol{s}_p = (x_p, y_p, z_p)^T$: 3D object-centered coordinates of the pth feature point, where $1 \le p \le P$.
- $\boldsymbol{m}_f$ and $\boldsymbol{n}_f$: 3D vectors representing the x and y axis of the image plane, respectively, where $1 \le f \le F$.
- (x_{fp}, y_{fp}): 2D image coordinates of the pth feature point in the fth frame.
- $[cx_f \ cy_f]^T$: 2D translation vector of the object center in the image coordinates.

we have the relation between 3D and 2D coordinates, as follows,

$$\begin{bmatrix} x_{fp} \\ y_{fp} \end{bmatrix} = \begin{bmatrix} \boldsymbol{m}_f^T \\ \boldsymbol{n}_f^T \end{bmatrix} \boldsymbol{s}_p + \begin{bmatrix} cx_f \\ cy_f \end{bmatrix}. \quad (1)$$

For the physical interpretation, it is necessary to introduce a matrix $Q = diag(1, 1, q)$, where $q \in \{-1, 1\}$, and a scale parameter $z \in \mathcal{R}$, thus we obtain,

$$\begin{bmatrix} x_{fp} \\ y_{fp} \end{bmatrix} = \left(\begin{bmatrix} \boldsymbol{m}_f^T \\ \boldsymbol{n}_f^T \end{bmatrix} Qz \right) (z^{-1} Q^{-1} \boldsymbol{s}_p) + \begin{bmatrix} cx_f \\ cy_f \end{bmatrix}. \quad (2)$$

The matrix Q represents a pair of visually equivalent candidates of 3D shape which corresponds to the

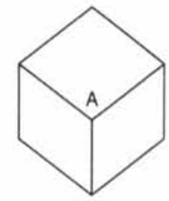


Figure 2: Depth reversal problem. Is vertex A convex or concave?

depth reversal problem; that is, the object is convex or concave around the point A in Fig.2. The scale factor z is necessary to represent the actual size of the object, and there is no way to determine z in the projective transformation.

2.2 Factorization with a known trajectory

To obtain a unique solution of 3D shape and size, we solve Eq.(2) by the factorization with a known trajectory of the object. Such trajectory is given by a computer-controlled robot hand whose coordinates is calibrated with respect to the camera coordinate system. The object is pushed by the end of the robot hand, and the object motion is recorded in a image sequence together with the accurate 3D position of the robot hand. Thus, we have many 2D trajectories of the feature points detected from the images and one 2D trajectory as a projection of the known 3D trajectory. We define the $(P+1)$th feature point by the given 3D trajectory which is represented as a vector $\boldsymbol{v}$.

$$\begin{aligned} \boldsymbol{v} &= [X_{1P+1}, Y_{1P+1}, Z_{1P+1}, \cdots, \\ &\quad X_{FP+1}, Y_{FP+1}, Z_{FP+1}]^T \end{aligned} \quad (3)$$

For each $q \in \{1, -1\}$ in matrix Q, 3D coordinates $(X_{fpq}, Y_{fpq}, Z_{fpq})$ of the pth feature point in the fth frame are calculated from Eq.(4),

$$\begin{bmatrix} X_{fpq} \\ Y_{fpq} \\ Z_{fpq} \end{bmatrix} = \frac{z_q}{||\boldsymbol{m}_f||} \left[\begin{bmatrix} \boldsymbol{m}_f^T \\ \boldsymbol{n}_f^T \\ q\frac{(\boldsymbol{m}_f \times \boldsymbol{n}_f)^T}{||\boldsymbol{m}_f||} \end{bmatrix} \boldsymbol{s}_p + \begin{bmatrix} cx_f \\ cy_f \\ l \end{bmatrix} \right], \quad (4)$$

where l denotes the focal length of the camera, and z_q is a scale factor for each $q \in \{1, -1\}$.

Using all 2D trajectories including the 2D projection of $(P+1)$th feature point, 3D trajectories are calculated by the factorization with Eq.(4), and we have the computed 3D trajectory of the $(P+1)$th feature point which is denoted by a vector $\boldsymbol{u}_q$,

$$\begin{aligned} \boldsymbol{u}_q &= \tfrac{1}{z_q} \ [X_{1P+1q}, Y_{1P+1q}, Z_{1P+1q}, \cdots, \\ &\quad X_{FP+1q}, Y_{FP+1q}, Z_{FP+1q}]^T. \end{aligned} \quad (5)$$

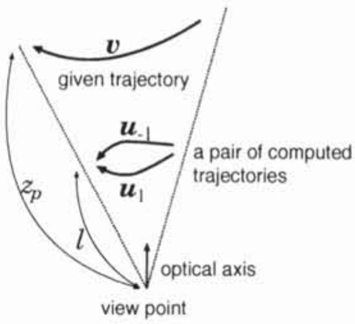


Figure 3: Relation between given trajectory $\boldsymbol{v}$ and computed trajectories $\boldsymbol{u}_{\pm 1}$.

From the relation $z_q \boldsymbol{u}_q = \boldsymbol{v}$, the scale factor z_q for each q are determined by the least squares method,

$$z_q = (\boldsymbol{u}_q^T \boldsymbol{v})/(\boldsymbol{u}_q^T \boldsymbol{u}_q). \quad (6)$$

Figure 3 illustrates the relation between $\boldsymbol{v}$ and $\boldsymbol{u}_q$. Now, we can determine that the true shape either of $q = 1$ or $q = -1$ by evaluating the remainder of

$$||z_q \boldsymbol{u}_q - \boldsymbol{v}||^2. \quad (7)$$

3 Experiments

3.1 Simulation study

We used a cube($100 \times 100 \times 100$mm) for the simulation study. The motion was given in 30 degrees rotation and 200mm horizontal translation from right to left. The object is viewed from a camera placed at 950mm distant and 345mm above from the object.

From sixty one 2D trajectories of the feature points in fifty frames and one known 3D trajectory of the top right vertex of the cube, the shape and size were recovered by the proposed method. We obtained two candidates of the scale factor, i.e. z_1 and z_{-1}, by using Eq.(6). The remainders of Eq.(7) corresponding to $q = 1$ and $q = -1$ were 1.40×10^2 and 4.12×10^4, respectively. Thus, we can determine that the shape and size of $q = 1$ is true.

Using the scale factor z_1, the estimated size of the object was $103.6 \times 104.4 \times 104.4$mm. While we have 13% volumetric error, it is important that we can estimate the real size of the object which is not obtained in the conventional factorization. Note that the error of the shape and size mainly comes from the linear approximation of the perspective projection.

3.2 Experiment in real environment

Figure 4 illustrates our experimental environment in remote operation environment. The remote operator first moves the object manually by the robot hand, then the unique shape and size of the object is recovered by the proposed method.

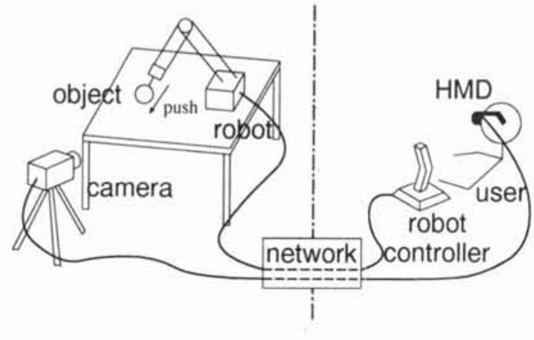
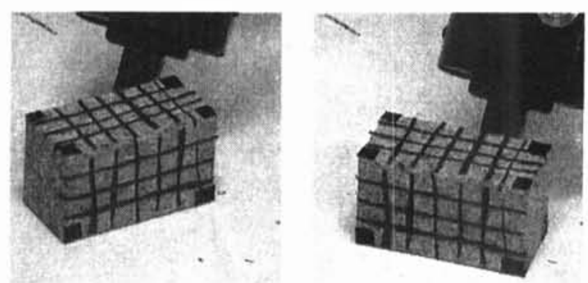


Figure 4: Experimental environment.



(a) The first frame (b) The last frame

Figure 5: An image sequence of a block pushed by robot hand.

We used a cork block whose size was $43 \times 49 \times 86$mm, and the object was pushed by a robot hand (MITSUBISHI RV-E3-ST) which is controlled by the remote operator through a computer network. The object is viewed by a camera (SONY EVI-G20) at 1100mm distant and 300mm above. The hand-eye calibration were performed in this environment.

Fifty seven feature points detected on the object are tracked by the CMU method[6] throughout sixty five images while the object travels along 60mm translation pushed by the robot hand. The first and last frames are shown in Fig. 5. Figure 6 shows the feature points in the first image.

The remainders of Eq.7 for $q = 1$ and $q = -1$ were 4.8×10^2 and 2.3×10^3, respectively. Thus we have found that $q = 1$ gives the true shape and size of the object. Figure 7 shows the trajectories of $\boldsymbol{u}_1$ and $\boldsymbol{u}_{-1}$ with the known $\boldsymbol{v}$ trajectory drawn in the bird-eye view. It is clear that $\boldsymbol{u}_1$ shows good consistency with $\boldsymbol{v}$ and $\boldsymbol{u}_{-1}$ does not.

Using the scale factor z_1 obtained by Eq.(6), the estimated size of the cork block was $41 \times 49 \times 83$mm.

Figure 6: Detected feature points in the first frame.

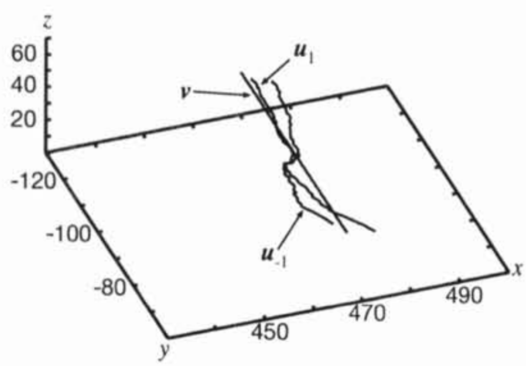


Figure 7: Consistency evaluation between $\boldsymbol{v}$ and $\boldsymbol{u}_{\pm 1}$.

The volumetric error was at most 8.0%, and such a good result shows that our method works successfully in real environment.

The side views of the recovered object corresponding to $q = 1$ and $q = -1$ are shown in Fig. 8. The horizontal lines show the surface of the table. Since $q = 1$ gives correct shape in this case, the recovered pose is stable as shown in Fig. 8 (a). Note that $q = -1$ gives a skewed shape with unstable pose. Figure 9 is a snapshot where the object was grabbed by the robot hand successfully as a result of 3D shape and size recovery.

4 Conclusion

This paper proposes an active recognition approach to solve the shape and size problems in 3D recovery. We first discussed that two parameters representing shape candidates and scale factor are necessarily involved in the formulation of the projective transformation, then described the procedure to obtain the unique solution by the factorization with a known 3D trajectory.

Experiments were performed in simulation study as well as in real environment, and both results have shown the potential usefulness of our method. In future, we will examine the robustness and stability of the proposed method in more realistic situations.

(a) $q = 1$ (b) $q = -1$

Figure 8: Side view of the recovered object.

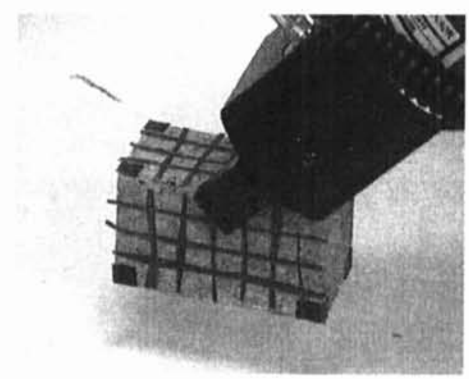


Figure 9: A snapshot of the object grabbed by robot hand.

References

[1] N. Asada, H. Fujiwara, T. Matsuyama: "Edge and depth from focus," IJCV, Vol.26, No.2, pp.153–163, 1998.

[2] S.T. Barnard, M.A. Fischler: "Computational Stereo," Computing Surveys, Vol. 14, No. 4, pp.553–572, 1982.

[3] R.J. Woodham: "Analyzing images of curved surfaces," Artificial Intelligence, vol. 17(1-3), pp. 117–140, 1981.

[4] C. Tomasi, T. Kanade: "Shape and motion from image streams under orthography: A factorization method," IJCV, 9-2, pp. 137–154, 1992.

[5] Y. Aloimonos, A. Bandopadhay, I. Weiss: "Active Vision," Proc. 1st ICCV, pp.35–54, 1987.

[6] C. Tomasi, T. Kanade: "Detection and Tracking of Point Features", CMU-CS-91-132, 1991.