High accuracy online 3D-reconstruction by multiple cameras

Yoshikazu OOTA¹, Yaodong PAN², Katuhisa FURUTA³

1,2,3 Department of Computer and Systems Engineering, Tokyo Denki University(TDU), Hatoyama-cho Hiki-gun Saitama 350-0394 JAPAN
1 oota@furutalab.k.dendai.ac.jp
2 pan@ham.coe.dendai.ac.jp
3 furuta@furutalab.k.dendai.ac.jp

Abstract: For online high accurate reconstruction of an object from an visual information, a linear reconstruction method for multiple images is popular. Basically this method needs many cameras or many different screen shots from different view points. This method, however, has the benefit of less calculation and is adequate for a real time application by comparing other popular method. In this paper, online reconstruction system using more than three cameras is treated. An evaluation method of cameras’ position, and of the number is derived for the linear reconstruction method. To decrease errors that are caused from skew of lens, positional error between corresponding points is taken into consideration on the evaluation. The proposed evaluation method enables estimation of the adequate number of cameras and then of feasible view locations. Additionally, repeating search of epipolar lines enables estimation of the hidden point. Comparing with result of an average error analysis, it was confirmed that the proposed methods works effectively.

Keywords: 3D Reconstruction, Image Processing

1. Introduction

To reconstruct a 3D-position accurately, cameras in the motion capture system should be placed properly [1]. And, it is necessary to consider the occlusion which decreases the accuracy of the reconstruction [2].

This paper will present an estimation method to evaluate the reconstruction accuracy depending on the number of cameras and cameras’ locations, and a method of estimating corresponding points which are hidden by occlusion. The former method can evaluate the accuracy by using three, four or more cameras. The latter method makes it possible to keep tracking invisible marker. Thus, the latter method can avoid reducing reconstruction accuracy by losing marker.

In this research, these methods are developed to the application of an optical motion capture system. After color filtering, a colored marker is detected as a corresponding point in a window that includes the colored marker. The procedure of reconstruction is shown below.

1. Determining cameras’ locations based on the proposed evaluation method
2. Keeping the tracking window to have the colored marker in the center of window
3. Detecting the center of the colored marker as the corresponding point
4. Estimating the corresponding point by using the proposed estimation method
5. Calculating the position of each tracking window that has the marker on the center of window
6. Reconstructing the 3D-position from corresponding points given by step 3
7. Return to step 2

In this paper, section 2 describes a liner reconstruction method. Section 3 proposes an evaluation method of reconstruction accuracy. Section 4 discusses the estimation method of an occluded corresponding point. Section 5 validates the usefulness of proposed methods by numerical simulations. Section 6 shows experimental results.

2. Linear Reconstruction by Multiple Cameras

We define the origin O₁ of a camera coordinate system be the focal point, and the optical axis is taken as the Z-axis as shown in Fig.3 (t₁₁ = [0, 0, 0]ᵀ, R₁₁ = I). Assume that rotating matrix and the translational vector from the i-th camera to the 1st one are Rᵢ₁ and tᵢ₁, respectively, which are said the motion parameters. Moreover, we assume the i-th camera’s corresponding point xᵢ = [xᵢ, yᵢ, 1]ᵀ. Here, ( xᵢ, yᵢ ) (i = 1, 2, · · · , n, n is the number of cameras ) are
location of corresponding points on a normalized image with the focal length $f = 1$. Ideally, Eqn.(1) holds.

$$X = s_1x_1 = s_2R_{12}x_2 + t_{12} = \cdots = s_nR_{1n}x_n + t_{1n} \tag{1}$$

Here, $s_i$ is the length from $O_i$ to $X$. However, the measured $x_i$ is corrupted by some noise. Thus, we should find $X$ and $s_i$ that minimize the quadratic criterion function

$$V_n = \sum_{i=1}^{n} ||X - s_iR_{1i}x_i - t_{1i}||^2. \tag{2}$$

Setting the partial derivative of $V_n$ by $s_i$ to 0, $s_i$ can be obtained using $R_{1j}^TR_{ij} = I$ as

$$s_i = \frac{x_i^TR_{1i}(X - t_{1i})}{x_i^Tx_i} \tag{3}$$

For the partial derivative of $V_n$ by $X$, we have

$$\frac{\partial V_n}{\partial X} = 2\sum_{i=1}^{n}\{X - s_iR_{1i}x_i - t_{1i}\} = 0 \tag{4}$$

The substitution of Eqn.(4) into Eqn.(3) gives $X$ minimizing the criterion function Eqn.(2) as

$$X = B_n^{-1}b_n \tag{5}$$

where

$$B_n = \sum_{i=1}^{n}\left\{I - \frac{R_{1i}x_i^TR_{1i}^T}{x_i^Tx_i}\right\} \tag{6}$$

$$b_n = \sum_{i=1}^{n}\left\{I - \frac{R_{1i}x_i^TR_{1i}^T}{x_i^Tx_i}\right\}t_{1i}. \tag{7}$$

![Fig. 3. Geometric Relation between Cameras](image)

**3. Evaluation of Reconstruction Accuracy**

Eqn.(5) can be rewritten as

$$X = B_n^{-1}A_n\tau_n. \tag{8}$$

Here, $A_n$ and $\tau$ are respectively defined as

$$A_n = \left[I - \frac{R_{11}x_1^TR_{11}^T}{x_1^Tx_1}, \cdots, I - \frac{R_{1n}x_n^TR_{1n}^T}{x_n^Tx_n}\right] \tag{9}$$

$$\tau_n = \left[t_{11}^T, \cdots, t_{1n}^T\right]^T. \tag{10}$$

![Fig. 4. Estimating Corresponding Point from Epipolar Lines](image)

**4. Estimation of Occluded Corresponding Points**

We can estimate an occluded corresponding point by making two epipolar lines on the visual image. Vectors $x_i, R_{ij}x_j$ and $t_{ij}$ should be coplanar. Then, the following epipolar equation should be satisfied.

$$x_i^T(R_{ij}x_j) = 0 \tag{13}$$

$$x_i^TE_{ij}x_j = 0 \tag{14}$$

The matrix $E_{ij}$ is called the essential matrix. There are image planes for cameras $i,j$ and $k$, respectively. Giving $x_i, x_j$ and motion parameters $(R_{ik}, t_{ik})$ and $(R_{jk}, t_{jk})$, we can draw two epipolar lines on the image plane $k$ as

$$a_{ik}y_k + b_{ik}x_k + c_{ik} = 0, \quad x_j^TE_{ik} = [b_{ik}, a_{ik}, c_{ik}] \tag{15}$$

$$a_{jk}y_k + b_{jk}x_k + c_{jk} = 0, \quad x_j^TE_{jk} = [b_{jk}, a_{jk}, c_{jk}] \tag{16}$$

The estimated corresponding point is located at the intersecting point $(x_k, y_k)$ determined by

$$y_k = \frac{a_{jk}c_{ik} - a_{ik}c_{jk}}{a_{jk}b_{ik} - a_{ik}b_{jk}} = x_k \tag{17}$$

However, by this method, the point $(x_k, y_k)$ can not be calculated from parallel epipolar lines. Orthogonal epipolar lines are desirable for reducing the influence of noise.

**5. Simulation Result**

**5.1. Camera-formation Survey**

Using three cameras, we evaluate the proposed method. We set the origin as the focal point of camera 1, and
define translation vectors as \( t_{12} = [0, 5, 0]^T \) and \( t_{13} = [5 \sin(\theta), 5 \cos(\theta), 0]^T \), respectively. \( \theta \) is moved by 20[degree]. Rotating matrices are defined as \( R_{1i} = I(i = 2, 3) \). The noise is assumed Gaussian in the measurement of each corresponding point \( x_i \). The average error and \( J \) values are shown in Fig.6, which indicates the close relation between \( J \) and \( E \). The error of \( E \) is defined as

\[
E = \sum_{p=1}^{m} ||(\bar{X}_p - X_p)||_2. 
\]

(18)

Here, \( \bar{X}_p \) is the true position of the \( p \)-th target point, and \( X_p \) is the reconstructed position obtained by Eqn.(5).

### 5.2. Accuracy Depending on the Number of Cameras

We show the average errors and \( J \) values by using two, three, four and five cameras. Locations of five cameras are shown in Fig.7. Noise is assumed to be added to the corresponding point in the same way as the above simulation. The error \( E \) is defined Eqn.(18). Fig.8 shows average errors and evaluation values. Using proposed evaluation method, we can estimate the reconstruction accuracy that depends on the camera number and motion parameters.

### 5.3. Estimation of Corresponding Points

We show the estimation of corresponding points. Three cameras are located as shown in Fig.9. Translation vectors are \( t_{12} = [4, 0, 0] \) and \( t_{13} = [2, 1.5, -2] \), and rotating matrices are \( R_{1i} = I(i = 2 \text{ and } 3) \). Noise is assumed to be added to the corresponding point in the same way as the above simulation. Fig.10 shows the occluded corresponding points by using the proposed method.
6. Real Image Application
Cameras are moved using a robot arm to set motion parameters. We took five images from one camera. Assume that these images are taken from single moving camera. Translation vectors are \( t_{12} = [0, 15, 0] \), \( t_{13} = [15, 0, 0] \), \( t_{14} = [0, -15, 0] \) and \( t_{15} = [-15, 0, 0] \), and rotating matrixes are \( R_{1i} = I \) (\( i = 2, 3, 4, 5 \)). Fig. 15 shows reconstruction results. Measuring the exact target position is difficult, we evaluate the variances of edge length between target points. Fig. 15 shows that the decrease of \( J \) leads to that of variance. Moreover, we estimate corresponding points from other images with the proposed estimation method. Fig. 16 shows estimation results. The estimated occluded corresponding points have enough accuracy by using real images.

7. Conclusion
We proposed two methods to realize a high accuracy reconstruction. From simulation results, the proposed methods could evaluate the reconstruction accuracy and estimate occluded corresponding points. These methods need less calculation efforts. Thus it is expected that the proposed methods can be applied easily for online applications.

References