3D Motion Parameters Determination Based on Binocular Sequence Images

ZHANG Jianqing  ZHANG Chunsen  HE Shaojun

ABSTRACT  Exactly capturing three dimensional (3D) motion information of an object is an essential and important task in computer vision, and is also one of the most difficult problems. In this paper, a binocular vision system and a method for determining 3D motion parameters of an object from binocular sequence images are introduced. The main steps include camera calibration, the matching of motion and stereo images, 3D feature point correspondences and resolving the motion parameters. Finally, the experimental results of acquiring the motion parameters of the objects with uniform velocity and acceleration in the straight line based on the real binocular sequence images by the mentioned method are presented.

KEY WORDS  binocular sequence image; camera calibration; image matching; motion parameter

CLC NUMBER  TP751

Introduction

Among existing vision monitoring and the estimation of 3D motion, nearly all investigations are monitoring and tracing the motion object based on single sequence images. The motion information by analyzing the single sequence images is relative, which includes a scale of factor related to the depth. In order to gain the absolute translation, an additional condition is necessary. On the basis of binocular sequence images, the 3D space coordinates of the feature points of moving object in different time can be obtained. Then, the moving parameters (rotation matrix R and translation vector T) are computed according with the 3D position information. Compared with the analysis method with single sequence images, not only the computation is simple, but also the absolute translation in space can be acquired. The key steps of the method include the acquisition of 3D coordinates in space, the extraction of feature point position on moving object, effective stereo and movement matching between feature points, and calculation of movement parameters. In this paper, an experimental binocular vision system and the calibration of its vidicons are introduced firstly. Then, the stereo and movement matching are performed for the feature points of movement object in the binocular stereo sequence images from the system, as well as to guarantee the correspondence of the feature points on the movement object in different time. The calibration result is used to obtain the coordinates of the feature points in object space, and the movement parameters of the object are estimated.

1 Binocular vision system and the calibration of its vidicons

Fig. 1 shows the binocular vision system installed in the School of Remote Sensing and Information Engineering of Wuhan University. The system consists of double video cameras, servo platforms with two freedoms of motion, image transmission and system control parts. The vidicon is video Color CCD Hitachi, lens is
128×Zoom, auto focus \( F = 3.9 - 63.0 \) mm. The image size is 240 pixels × 320 pixels. The base between two vidicons can be adjusted, which is about 500 mm usually.

The calibration of the camera is an essential procedure in the 3D position based on binocular sequence image. The precision and the reliability of camera calibration directly influence spatial position precision of the stereovision system. Usually the camera should be calibrated before applying. The principle point of the camera should be invariable nearly, and the focuses should be controllable. Because the focuses of the CCD cameras in the experimental system are auto-variable, the real-time in-suit calibration of CCD camera is adopted in the study case, that is on-line calibration based on a reference object. Two grids perpendicular each other are setting on the background. The parameters of each image are computed real-timely by use of the grid points. Fig. 2 shows the scene of in-suit calibration. The calibration model is direct linear transformation (DLT) with distortion correction.

\[
\begin{align*}
    i + v_i + \delta_i + \frac{L_1 X + L_2 Y + L_3 Z + L_4}{L_5 X + L_6 Y + L_7 Z + 1} &= 0 \\
    j + v_j + \delta_j + \frac{L_8 X + L_9 Y + L_{10} Z + L_{11}}{L_{12} X + L_{13} Y + L_{14} Z + 1} &= 0
\end{align*}
\]

where \( (L_1, L_2, \cdots, L_{11}) \) are 11 unknown DLT parameters; \( (X, Y, Z) \) is the space coordinate of grid points in object space; \( (i, j) \) is the pixel coordinate of grid point; \( (\delta_i, \delta_j) \) is the pixel distortion value; \( (v_i, v_j) \) is the correction value of \( (i, j) \) when redundancy exists. The transformation parameters \( L_1, L_2, \cdots, L_{11} \) can be solved from Eq. (1) and the image and space coordinates of grid points.

2 Image matching and solving special points

2.1 Extraction of feature point on the object

Acquiring the feature points from the images is the first step of matching. Harris operator\(^{[25]}\), which is simple, stable, and insensitive to the noise and illumination is applied in the feature extraction, which is commonly used in the computer vision community. It can quantitively extract feature points, and the distribution of obtained feature point is reasonable. The expression of Harris operator is as below.

\[
M = G(s) \otimes \begin{bmatrix} g_x & g_y \\ g_y & g_y \end{bmatrix} \]

\[
I = \det(M) - k \mathrm{tr}^2(M), k = 0.04
\]

where \( g_x \) is the gradient in \( x \) direction, and \( g_y \) is the gradient in \( y \) direction; \( G(s) \) is Gauss template; \( \otimes \) is convolution operation; \( I \) is interest value of each point; \( \det \) is matrix determinant; \( \mathrm{tr} \) is matrix trace; and \( k \) is a constant.

2.2 Initial matching

The goal of initial matching is to determine a candidate matching set \( T \). The grey correla-
tion is used here. For each feature point $m_i$ $(u_i, v_i)$ of the image 1, if the coordinate difference between $m_1$ and $m_2$ $(u_2, v_2)$ of the image 2 is less than a certain threshold, the gray correlation coefficient $\rho$ of $(2m+1) \times (2n+1)$ windows centered in $m_1$ and $m_2$ is calculated.

$$\rho(m_1, m_2) = \frac{\sum_{i=-m}^{m} \sum_{j=-n}^{n} [I_1(u_i + i, v_i + j) - E(I_1)] \times [I_2(u_2 + i, v_2 + j) - E(I_2)]}{(2n+1)(2m+1) \sqrt{\sigma^2(I_1) \times \sigma^2(I_2)}}$$

where $I_1$, $I_2$ are the gray values of two windows, $E(I_1)$, $E(I_2)$ are the gray averages of two windows, and $\sigma_1$, $\sigma_2$ are the gray standard deviations of two windows. For a given couple of points to be considered as a candidate matching, the correlation coefficient must be higher than a given threshold (threshold is 0.8 in the experiment). The size of search window is usually determined by priori knowledge (the size of correlated window is $11 \times 11$ in the experiment). Therefore, the candidate matching relations between a certain feature point in image 1 and some feature points in image 2 are established. This point is then joined to the candidate match set $T$.

### 2.3 Relaxation method based matching for motion and stereo images

The relaxation method is to let the candidate matching pair in $T$ satisfy the continuity and uniqueness furthestly through the iteration. The continuity refers to that the massive other correct matching pair usually exists in the neighborhood of a correct matching pair. Uniqueness refers to that one feature point exists in only one matching pair, or it can be expressed as: if candidate matching pair is right, there must be many consistent candidate matching pairs around it; while if candidate matching pair is wrong, there are less consistent candidate matching pairs around it. After the initial matching probability of each candidate matching pair was defined, the matching probability is modified according to the consistency. The processing is iterated till the matching probabilities of all point pairs are less than known threshold, except the most possible matching pair.

The efficient correspondence of three dimensional feature points at different time is a difficulty in the computation of motion parameters. Using the image matching method discussed above, it can be effectively solved through the reasonable combination about the motion and stereo matches. At first, the feature points in the moving target are extracted and the relaxation matching is performed on images taken at different times in the same sequence. The correspondences of feature points between images in left (right) sequence are acquired, which are called the single-sequence correspondences or motion correspondence. Then search the matching points for the single-sequence points extracted ahead on the images, which are taken at the same times but in right (left) sequence, in order to establish the correspondence called stereo correspondence for stereo images got at the same time, then acquire the stereo correspondence points at different time during motion.

### 2.4 Computation of spatial points

After the coordinates at left and right images of each feature point on moving target are acquired according to feature points extraction and images matching, the spatial coordinates $(X,Y,Z)$ of feature points in object space can be calculated. Suppose the parameters of two images are $(L_1, L_2, \ldots, L_{11})$ and $(L'_1, L'_2, \ldots, L'_{11})$ respectively, and the left and right image coordinates of a feature point are $(x, y)$ and $(x', y')$. By the set of linear equations from Eq. (1), the object spatial coordinates $(X,Y,Z)$ for feature points can be computed.

$$\begin{bmatrix}
L_1 + xL_5 + L_9 + xL_{10} & L_2 + xL_{11} \\
L_5 + yL_9 & L_1 + yL_{10} \\
L'_1 + x'L'_9 & L'_2 + x'L'_{10} \\
L'_5 + y'L'_9 & L'_1 + y'L'_{10}
\end{bmatrix} = \begin{bmatrix}
L_1 + xL_5 + L_9 + xL_{10} & L_2 + xL_{11} \\
L_5 + yL_9 & L_1 + yL_{10} \\
L'_1 + x'L'_9 & L'_2 + x'L'_{10} \\
L'_5 + y'L'_9 & L'_1 + y'L'_{10}
\end{bmatrix}.$$
\[
\begin{bmatrix}
X \\
Y \\
Z
\end{bmatrix} + \begin{bmatrix}
L_x + x \\
L_y + y \\
L_z + x'
\end{bmatrix} = 0
\] (4)

3 Determination of three-dimensional motion parameter

3D spatial point sequences have been acquired according to the matching of stereo motion images. As a result the motion parameters can also be estimated on the basis of 3D spatial point sequences.

3.1 Basic estimation method of three-dimensional motion parameter in binocular sequential images

According to the theory of the dynamics and the space analytic geometry, rigid body movement in the three dimensional space can be decomposed into rotation and translation. If the three dimensional coordinates of any feature point \( P \) at time \( t \) on moving rigid body is \((X,Y,Z)^T\), after time \( \Delta t \) it moves to feature point \( P'(X',Y',Z')^T(i=1,2,\cdots,n) \), then, the corresponding relationship of \( P \) and \( P' \) is as the following rigid body moving equation.

\[
P'_i = RP_i + T (i = 1,2,\cdots,n)
\] (5)

where \( R \) is a \( 3 \times 3 \) rotational matrix, \( T = (AX,AY,AZ)^T \) is the translation vector. Thus determining the 3D motion parameters from sequential images is computing the rotational matrix \( R \) and the translation vector \( T \). If there are three corresponding feature points on the object before and after the motion, then according to the Eq. (5),

\[
R \begin{bmatrix}
P_1 - P_2 \\
P_2 - P_3 \\
P_3 - P_1
\end{bmatrix} = \begin{bmatrix}
P'_1 - P'_2 \\
P'_2 - P'_3 \\
P'_3 - P'_1
\end{bmatrix}
\] (6)

The Eq. (6) is linear about the \( R \). The condition with the unique solution is that the rank of the coefficient matrix is not less than 2. In another word, \( P_1 - P_2 \) and \( P_2 - P_3 \) must not be colinear. It means that the three-dimensional motion parameter can be uniquely determined so long as more than three 3D feature points with un-colinear on the object are obtained. In order to guarantee the precision and the computation speed, in the actual computation the centered method for coordinates of the 3D feature points is adopted. If the gravity centers of the 3D feature point sets \{\( P_i \)\} and \{\( P'_i \)\} are:

\[
\bar{P} = \frac{1}{n} \sum_{i=1}^{n} P_i, \quad \bar{P'} = \frac{1}{n} \sum_{i=1}^{n} P'_i
\]

thus

\[
\bar{P'} = R\bar{P} + T
\]

then

\[
(P'_i - \bar{P'})^T R = (P_i - \bar{P})^T (i = 1,2,\cdots,n)
\] (8)

After calculating rotation matrix \( R \), the translation vector can be computed with following equation.

\[
T = \bar{P'} - R\bar{P}
\]

3.2 Solution of rotational matrix \( R \)

Above-mentioned \( R \) is a \( 3 \times 3 \) orthogonal matrix. The algorithm for calculating 9 components of \( R \) directly, is complex, sometimes the solution is not orthogonal, and the reliability is poor. Because there are only 3 independent variables in the rotational matrix, there is no doubt that it will reduce computation workload and enhance the reliability of algorithm by selecting these 3 independent variables for computation. By a algorithm based on a decomposition of a skew-symmetry matrix to determine the rotational parameters of three-dimensional motion, not only the computation workload is reduced, but also linear computation is realized.

As known from Cayley theorem, if a 3D orthogonal matrix \( R \) satisfies that \( I + R \) is full rank, it can be decomposed into a skew-symmetry matrix \( S \) and a unit matrix \( I \) uniquely, that is

\[
R = (I + S) (I - S)^{-1}
\]

where

\[
S = \begin{bmatrix}
0 & -c & b \\
c & 0 & -a \\
-b & a & 0
\end{bmatrix}, \quad I = \begin{bmatrix}
1 & 0 & 0 \\
0 & 1 & 0 \\
0 & 0 & 1
\end{bmatrix}
\]

\[
R = \frac{1}{1 + a^2 + b^2 + c^2} \begin{bmatrix}
1 + a^2 - b^2 - c^2 & 2ab - 2c & 2ac + 2b \\
2ab + 2c & 1 - a^2 + b^2 - c^2 & 2bc - 2a \\
2ac + 2b & 2bc + 2a & 1 - a^2 + b^2 + c^2
\end{bmatrix}
\] (11)
substituting (10) into (8), then
\[(P' - P)(I + S) = (P - P)(I - S)\]  
(12)

Independent variable \((a, b, c)\) can be obtained by solving Eq. (12), thereby the rotational matrix \(R\) is acquired.

4 Experimental results

4.1 Motion with uniform velocity along straight line

Using the binocular vision system a pair of binocular image sequences about a toy car in constant interval at 8 different times is taken (with 7 time intervals, in each time interval the car moves \(\Delta X = 17.6\, \text{mm}, \Delta Y = -17.6\, \text{mm}\), which moves with uniform velocity along straight line on an \(XY\) plane. The three-dimensional motion parameters of the moving object from images sequence are calculated. The results of rotational matrixes and translation vectors in 7 time intervals are shown in Table 1. The errors between calculated values and true values in \(X\) and \(Y\) directions in 7 time intervals are shown in Table 2. The mean errors of in \(X\) and \(Y\) directions are 2.7 mm and 2.5 mm respectively. Table 3 shows the root mean square of errors between spatial positions \(P'_i(X', Y', Z'_i)^T\) and \(P'_i(X''_i, Y''_i, Z''_i)^T\) of feature points, where \(P'_i(X', Y', Z'_i)^T\) (the calculated position of the points at time \(t + \Delta t\) are calculated from \(P'_i(X_i, Y_i, Z_i)^T\) (the real position of the points at time \(t\)) according to Eq. (5) when rotation matrix \(R\) and translation vector \(T\) are acquired already, and \(P'_i(X''_i, Y''_i, Z''_i)^T\) is the reconstructed position of the points corresponsive to \(P'_i(X_i, Y_i, Z_i)^T\) at time \(t + \Delta t\), where \(i = 1, 2, \ldots, 16\).

| \(\phi (^\circ)\) | \(\omega (^\circ)\) | \(\kappa (^\circ)\) | \(\Delta X/\text{cm}\) | \(\Delta Y/\text{cm}\) | \(\Delta Z/\text{cm}\) |
|------------------|-----------------|-----------------|----------------|----------------|----------------|
| 1                | 1,439 006       | 1,104 751       | 0.737 537      | 1.74           | -1.70          | -0.49          |
| 2                | -4,010 399      | -3,376 598      | -0.891 974     | 1.78           | -1.36          | 1.27           |
| 3                | 0.816 508       | 1.320 509       | -2,736 706     | 1.35           | -1.40          | -0.13          |
| 4                | -2,084 656      | -1,176 902      | -1,107 007     | 1.95           | -1.23          | 0.92           |
| 5                | 0.428 672       | 1.241 595       | -1,233 721     | 1.23           | -1.71          | -0.55          |
| 6                | 2,688 088       | 2,701 891       | -0.868 271     | 1.18           | -1.63          | -1.16          |
| 7                | -1,932 762      | -1,438 182      | -0.946 721     | 1.93           | -1.55          | 0.81           |

| \(d_{XX}\) | \(d_{XY}\) |
|----------|----------|
|         |          |
| 1       | -0.00    | -0.00    |
| 2       | -0.04    | -0.04    |
| 3       | 0.18     | 0.06     |
| 4       | -0.54    | -0.06    |
| 5       | -0.59    | -0.14    |
| 6       | 0.16     | -0.22    |

| \(m_x\) | \(m_y\) | \(m_z\) |
|----------|----------|----------|
|          |          |          |
| 1       | 0.27     | 0.29     |
| 2       | 0.15     | 0.13     |
| 3       | 0.22     | 0.23     |
| 4       | 0.37     | 0.33     |
| 5       | 0.08     | 0.10     |
| 6       | 0.15     | 0.18     |
| 7       | 0.34     | 0.29     |

4.2 Motion with uniform acceleration along straight line

A target is moving with a uniform acceleration along straight line, where acceleration \(a = 2 \, \text{cm}/\, \text{s}^2\) and initial velocity is 0. Binocular stereo cameras took images at 8 different times, \(t_1, t_2, \ldots, t_8\), with the same time interval. The predictive values of moving target’s positions at \(t_1, t_2, t_8\) are calculated on the basis of estimation method by square extrapolating. In order to ensure that the target is in the center of view scope, the horizontal and vertical rotation of the servo platforms are controlled by the computer in the binocular vision system, so that the requirement to trace the moving target is achieved. The tracking results for the motion with uniform acceleration along straight line are shown in Table 4.
where $X$, $Y$, $Z$ are estimated values of positions at times $t_4, t_6, t_8$ corresponding to time $t_1$, $S_r$ is the estimated translation quantities calculated from $X$, $Y$, $Z$ at $t_4, t_6, t_8$ corresponding to the position at $t_1$. $S_r$ is the real distance between the position at $t_1$ and the positions at times $t_4, t_6, t_8$, and $\Delta S$ is $S_r - S_S$.

Table 4 Tracing results of the linear motion with uniform acceleration/cm

|   | $X$  | $Y$  | $Z$  | $S_r$ | $S_S$ | $\Delta S$ |
|---|------|------|------|-------|-------|------------|
| $t_4$ | -0.75 | 0.57 | 0.02 | 0.94  | 0.90  | 0.04       |
| $t_6$ | -2.03 | 1.61 | 0.01 | 2.59  | 2.50  | 0.09       |
| $t_8$ | -4.05 | 3.07 | -0.00| 5.09  | 4.90  | 0.19       |

5 Conclusions

The research introduced in this paper combined the photogrammetry and computer vision effectively and an entire method to determine the parameters of moving target from binocular sequence images is concluded, in which the problem of feature points correspondence is solved effectively. As a result, the real tracing and positioning the target moving with a uniform velocity and a uniform acceleration is realized. The further research is to trace and position the target in arbitrary motion based on current work.

REFERENCES

[1] Ma Songde, Zhang Zhengyou (1998) Computer vision—the theory of computation and the basic algorithm[M]. Beijing: Science Press (in Chinese)

[2] Feng Wenhao (2002) Close range photogrammetry—photography way to determine the targets’ shape and moving condition[M]. Wuhan: Wuhan University Press (in Chinese)

[3] Harris C, Stephens M (1988) A combined corner and edge detector[C]. The 4th Alvey Vision Conference, Manchester, UK

[4] Zhang Jianqing, Wang Shugen, Pan Li (2003) Photogrammetry[M]. Wuhan: Wuhan University Press (in Chinese)

[5] Zhang Zuxun, Zhang Jianqing, Wu Xiaoliang, et al. (1992) Global image matching with relaxation method[C]. The International Colloquium on Photogrammetry, Remote Sensing and Geographic Information Systems, Wuhan, China

[6] Zhang Zhengyou, Derche R, Faugeras O, et al. (1995) A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry[J]. Artificial Intelligence, 78(1/2): 87-119

[7] Aggarwal J K, Chandakumar N (1988) On the computation of motion from sequences of images: a review[J]. IEEE, 76(8): 917-935

[8] Xu Wenli, Liu Wenhuang (1992) Estimation motion parameters from position vectors of feature points[J]. Acta Automatica Sinica, 18(4):440-447