A Novel Computer Vision Technique Used on Sport Video

Qiu Xian-jie
Institute of Computing Technology, CAS
Beijing 100080, China
qxj@ict.ac.cn

Wang Zhao-qi
Institute of Computing Technology, CAS
Beijing 100080, China
zqwang@ict.ac.cn

Xia Shi-hong
Institute of Computing Technology, CAS
Beijing 100080, China
xsh@ict.ac.cn

ABSTRACT

A method based on computer vision technologies is presented to achieve the function that the simulated motion in sport simulation system and the motion in sport video are presented on the same screen and at the same view point. The proposed method first applies the camera self-calibration theory to obtaining camera intrinsic parameters in sport video according to the 2D video features correspondence. Next it makes use of the feature 3D reconstruction to get a feasible estimation of extrinsic parameters. The extracted camera parameters information is applied to sport simulation and training system to achieve the function that the simulated normal motion of 3D virtual athlete in sport training system and the athlete motion in sport video are presented on the same screen and at the same view point. So we can quickly and accurately find the difference between the athlete motion in sport video and the simulated motion. It is very helpful to coaches and the training of athletes.

Keywords

sport video, camera calibration, camera reference frame, world reference frame, intrinsic parameters, extrinsic parameters

1. INTRODUCTION

The simulation and analysis of human motion is a popular issue in simulation area. It is highly valued in sport motion training and analysis. Video is the main record form of athlete train. It is significant if we put the athlete motion of video and the normal motion of 3D virtual athlete of simulation system into the same screen because we can easily and accurately compare the difference in them. And it will greatly benefit the athletes so that they can compare their motion with the normal motion of 3D virtual athlete and improve their competition level. While the motion of 3D virtual athlete can be observed from different view point and the appearance will be different, the comparison is rather difficult. In order to ensure the view point from which we watch the motion of virtual athlete in the simulation system and the view point of the sport video are the same, we have to adjust the view point of the simulation system by hand. It will usually become inconvenient and inaccurate. This shortcoming prevents the further use of human motion simulation and training system in athlete training.

We have presented a novel computer vision technique. With this technique, we can get the camera parameters from sport video and use the camera information to adjust the view point of the virtual camera of 3D simulation system automatically. We have implemented this technique in the trampoline sport simulation and training system and sound results have been obtained.

Our main contribution is to synthetically use computer vision algorithms such as camera self-calibration and feature tracking in the course of realizing the technique presented here in the sport video. In this way we have greatly extended the function of sport training system. To our knowledge, it is a novel try. On the other hand, the estimation of extrinsic parameters by using feature 3D reconstruction appears to be new.

Journal of WSCG, Vol.12, No.1-3, ISSN 1213-6972
WSCG'2004, February 2-6, 2003, Plzen, Czech Republic.

We design in the following paragraphs all the algorithmic steps required for the computer vision technique. At first, we present a global frame of the visual reality technique in section 2. Section 3 presents the method of camera intrinsic calibration. The method for computing the camera motion between the successive frames and 3D reconstruction is explained in section 4. Section 5 illustrates the method of extrinsic parameters estimation and the experimental results obtained by using the computer vision technique are illustrated in section 6. In section 7,we conclude our work.

2. TECHNIQUE OUTLINE

The computer vision technique we present here can be widely applied not only in the trampoline sport as we have done but also in diving and gymnastics, etc, if only there are camera motions included in the sport video and the normal sport field can be viewed in it.

The computer vision technique will be implemented in two steps:

1) 、camera intrinsic parameter calibration of sport video. This step includes feature selection, tracking, and camera self-calibration.

2) 、camera extrinsic parameter is extracted from sport video with feature 3D reconstruction and the view point of virtual camera in the simulation system is adjusted automatically with this parameter into the same view point as in the trampoline sport video.

3. CAMERA SELF-CALIBRATION

Camera calibration is traditionally determined by observing a known calibration object. However, there are several applications for which a calibration object is not available, or its use is too cumbersome. For example, in sports training, the data of athlete training is usually stored in video. It is impossible to always put a known calibration object into a gym. For this applications a self-calibration techniques is useful and the first algorithm that solves the self-calibration was proposed by Faugeras and Luong [Fau92] [Luo97]. Compared with traditional calibration techniques, self-calibration does not require a calibration object with known 3D geometry, but only needs point correspondences from images to solve for the intrinsic parameters.

The calibration method we propose is based on the calibration method that Mendonca and Cipolla proposed. The main steps of the method include:

1)、extraction and matching of points through the trampoline sport image sequence.

2)、estimation the Fundamental matrix between two successive frames.

3)、based on the two steps above and the essential matrix properties, a cost function is used which takes the intrinsic parameters as arguments and the fundamental matrix as parameters.

3.1、Extraction and matching of points

Most of classical stereoscopic methods distinguish between extraction and matching of features.

In this respect, Deriche [Der90] uses contour points to identify the features; Harris [Har88] considers interest points and Sistiaga [Sis00] local invariants. The features to be matched are extracted from both images and matched using correlation techniques or a measure of distance between differential attribute vectors.

In normal sport video, image sequences have little displacement between two successive images. A tracking algorithm based on the principles developed by Kanade-Lucas-Tomasi (KLT) [Shi94] is found to be more appropriate in our application. Opposed to the methods mentioned above, the KLT algorithm has the particularity to extract characteristics only in a first image and to track them through a short image sequence.

3.1.1、Extraction of points

The selected primitive of KLT is a textured patch with a large variation of intensity in the two directions x and y. Given the intensity function $I(x, y)$, the matrix of variation of local intensity Z is:

$$Z = \int_W g g^T dW = \int_W \begin{bmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{bmatrix} dW \qquad (1)$$

where:

-g is the local gradient defined as $g = \left(\frac{\partial I}{\partial x}, \frac{\partial I}{\partial y} \right)$.

- I_x and I_y are the first derivatives of the function I along directions x and y.

A patch defined by a squared window W (for example 15 pixels) is accepted as an interesting primitive if the two eigenvalues (λ_1, λ_2) of Z are superior to a threshold value λ:

$$\min(\lambda_1, \lambda_2) > \lambda \qquad (2)$$

The matrix Z can be computed in function of x and y by displacing the window W by a number of pixels smaller than the window size.

Eigenvalues of the matrix Z are computed for each displacement of the window. Thresholding according to equation (2) gives a set of feature points.

3.1.2、Feature tracking

If the residual displacement between two successive images I and J are smaller than the observation window, features can be tracked using a measure of similarity. Because two images are extracted from a continuous sequence, the displacement is assumed to be small and can be approximated by a translation. The image can then be defined as a function of three variables (x, y, t):

$$I(x, y, t) = I(x + \zeta, y + \eta, t + \tau) \qquad (3)$$

where $d = (\xi, \eta)$ is the displacement between times t and $t+\tau$ at point $x=(x, y)$.

We can write:

$$I(x + d) = I(x + \xi, y + \eta, t + \tau) \qquad (4)$$

A second image can then be written as:

$$J(x) = I(x + d) + n(x) \qquad (5)$$

where n is a function representing the image noise.

What we shall do to solve the problem of tracking is to determine d by minimizing the dissimilarity between the two windows W in I and J.

$$\varepsilon = \int_W [I(x + d) - J(x)]^2 \omega(x) dx \qquad (6)$$

Because of the small value of the displacement vector d, the intensity function can be approximated by a Taylor series truncated to the linear term:

$$I(x + d) = I(x) + g \bullet d \qquad (7)$$

where:

-g corresponds to the image gradient.

According to [Shi94], we have:

$$Zd = e \qquad (8)$$

where :

- Z can be expressed by the truncated Taylor series:

$$Z = \int_W g(x) g^T(x) \omega(x) dx \qquad (9)$$

- e is a column vector representing the difference between the two images:

$$e = \int_W [I(x) - J(x)] g(x) \omega(x) dx \qquad (10)$$

The d is therefore computed by solving the system (8).

3.2、Epipolar geometry

The projective geometry corresponding to two images of a same scene taken in two different view points is called epipolar geometry: the first image of any point must lie in the plane formed by its second image and the optical centers of the two cameras. The epipolar constraint can be represented algebraically by a 3×3 matrix, called the fundamental matrix in the following way:

$$q_i'^T F q_i = 0 \quad \forall i \in [1, n] \qquad (11)$$

Several methods exist to identify the fundamental matrix. One of these methods [Fau92] consists in minimizing the sum of distance squares from a point to the epipolar line in the two images, which is supposed to pass by this point. A further approach is based on the minimization of (11) by weighing the gradient of its variance. On the contrary of these two methods, which are non-linear, Hartley [Har95] proposes to us a linear method.

3.2.1 Fundamental matrix estimation

In our approach, we estimate the fundamental matrix by using Hartley's normalized 8-point algorithm [Har95]. The problem is solved by first normalizing the matched point coordinates. During this stage, each image reference frame is first translated to the centroid of the set of all points. Then, an isotropic scaling of points allows us to reduce the distance at the origin so that the average value of the distances is equal to $\sqrt{2}$. Given matched point $q_i' \leftrightarrow q_i$, In particular, writing $q = (u, v, 1)^T$ and $q' = (u', v', 1)^T$ one point match gives rise to one linear equation in the unknown entries of F':

$$uu' f_{11} + uv' f_{21} + uf_{31} + vu' f_{12} + vv' f_{22} + vf_{32} + u' f_{13} + v' f_{23} + f_{33} = 0 \qquad (12)$$

The normalized 8-points algorithm allows us to compute the fundamental matrix from a set of at least eight matched point.

This equation (12) can be written in the form:

$$U^T f = 0 \qquad (13)$$

where:

$$U = [uu', uv', u, vu', vv', v, u', v', 1]$$

$$f = [f_{11}, f_{21}, f_{31}, f_{12}, f_{22}, f_{32}, f_{13}, f_{23}, f_{33}]^T$$

If we have n correspondences, we obtain the following matrix equation:

$$Af = \begin{bmatrix} u_1u_1' & u_1v_1' & u_1 & v_1u_1' & v_1v_1' & v_1 & u_1' & v_1' & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ u_nu_n' & u_nv_n' & u_n & v_nu_n' & v_nv_n' & v_n & u_n' & v_n' & 1 \end{bmatrix} f = 0 \quad (14)$$

The fundamental matrix F is defined only up to an unknown scale factor. To avoid a trivial solution of f, a constraint was added: $\| f \| = 1$.

If the data are noisy and inaccurate, the estimation of F is obtained by using linear least squares. So, we seek to estimate the f vector that minimizes $\| Af \|$. The solution of f is the eigenvector corresponding to the smallest eigenvalue of AA^T.

The fundamental matrix is associated with a constraint that has to be of rank 2. Therefore the fundamental matrix F is replaced by the matrix F' that minimizes the Frobenius norm $\| F - F' \|$ subject to the condition: $\det(F') = 0$.

The complete normalized 8-point algorithm is summarized as follows:

Algorithm : The normalized 8-point algorithm

(i) Normalization: Transform the image coordinates according to $\widehat{x_i} = Tx_i$ and $\widehat{x_i}' = T'x_i'$, where T and T' are normalizing transformations consisting of a translation and scaling.

(ii) Find the fundamental matrix $\widehat{F}'$ corresponding to the matches $\widehat{x_i} \leftrightarrow \widehat{x_i}'$ by

(a) Linear solution: Determine $\widehat{F}$ from the singular vector corresponding to the smallest singular value of $\widehat{A}$, where $\widehat{A}$ is composed from the matches $\widehat{x_i} \leftrightarrow \widehat{x_i}'$ as defined in (14).

(b) Constraint enforcement: Replace $\widehat{F}$ by $\widehat{F}'$ such that det $\widehat{F}'$=0 using the SVD.

(iii) Denormalization: Set $F = T'\widehat{F}'T$. Matrix F is the fundamental matrix corresponding to the original data $\widehat{x_i} \leftrightarrow \widehat{x_i}'$.

3.2.2 RANSAC algorithm

So far, we have assumed that the matches in two frames are correct. But in fact, there are false matches existing unavoidably in the results of KLT algorithm. If we use all the matches including the good and false in the estimation of F matrix and camera intrinsic parameters, the result is wrong and useless. Therefore the elimination of bad point matches is essential. In order to do this, we suggest the use of the RANSAC (RANdom Sample Concensus) algorithm as developed by Fischler and Bolles [Fis81]. The integration of the RANSAC algorithm in the process of self-calibration is able to significantly reduce the errors.

The working principle of this algorithm is to define a model and to determine it with N random samples of n points. From these N determinations of the model, a classification of matches (as valid or not) is possible by using a criterium of points validity.

The number of samples N is chosen sufficiently high to ensure with a probability p, that at least one of the random samples of n points consists solely of valid matches. N is given by:

$$N = \log(1-p) / \log(1-(1-\varepsilon)^n) \quad (15)$$

with p=99.0 and ε the probability that a selected point is an "invalid point".

The RANSAC algorithm is coupled with the computation of the fundamental matrix between two successive images in the course of camera intrinsic parameters estimation. The model, which represents the mathematical support allowing to verify the point matches is also expressed by the computation of the fundamental matrix.

The complete RANSAC algorithm is summarized as follows:

Algorithm : RANSAC: Fitting a Fundamental Matrix Using Random Sample Consensus

Determine:

The smallest number of corresponding pairs of points required is seven

k-----the number of iterations required

t-----the threshold used to identify a point that fits well

d----the number of nearby points required to assert a model fits well

Until *k* iterations have occurred

Draw a sample of seven correspondences from the data uniformly and at random

Use the seven points algorithm to obtain an estimate of the fundamental matrix F_0.

For each putative correspondence outside the sample

Test the distance from the closest points consistent with F_0 to the measured points against *t;*

if the distance is less than *t*, the correspondence is consistent

end

If there are d or more consistent correspondences then there is a good fit. Refit the estimate of the fundamental matrix using all these points.

end

Use the best fit from this collection, using the fitting error as a criterion

3.3 Intrinsic parameters estimation

The camera's intrinsic parameters are calculated by using the Mendonça and Cipolla method [Men99] which is an extension of Hartley's self-calibration technique based on the properties of essential matrix, this method is applied to a set of five images taken from video sequence.

The Mendonça and Cipolla method allow for the stable computation of varying focal lengths and principal point. The three singular values of the essential matrix have to satisfy two conditions: one of them must be zero and the two others must be equal. For that, Mendonça and Cipolla use a cost function which takes the intrinsic parameters as arguments and the fundamental matrix as parameters. This function minimizes a positive value proportional to the difference between the two non-zero singular values of the essential matrix.

The essential matrix is given by:

$$E = [t]_x R \tag{16}$$

where:

- $[t]_x$: antisymmetric matrix associated to the translation vector *t*,
- R : rotation matrix.

The expression of the fundamental matrix is (with *K* the matrix of intrinsic parameters):

$$F = K^{-T}[t]_x R K^{-1} \tag{17}$$

$$F = K^{-T} E K^{-1} \tag{18}$$

then:

$$K^T F K = E \tag{19}$$

Let F_{ij} be the fundamental matrix associated to consecutive images *i* and *j*, and ${}^1\sigma_{ij} > {}^2\sigma_{ij}$ be the non-zero singular values of E_{ij} obtained by making a singular value decomposition (SVD) of E_{ij}. The cost function is:

$$C(K) = \sum_{i=1}^{n}\sum_{j>i}^{n} \omega_{ij} \frac{{}^1\sigma_{ij} - {}^2\sigma_{ij}}{{}^2\sigma_{ij}} \tag{20}$$

where:

- $K = f(\alpha_u, \alpha_v, u_0, v_0)$ where α_u, α_v, u_0, v_0 correspond respectively to products of the scale factors according to the axis *u* and *v* by the focal and to the coordinates of the intersection of the optical axis with the image plane,
- w_{ij} is the degree of confidence of the fundamental matrix F_{ij} estimation.

4. CAMERA MOTION AND 3D RECONSTRUCTION

4.1、camera motion

Once we have estimated the essential marix *E*, we can recover the motion（*R*，*t*） between the successive frames. As $E^T t = 0$, the relative location t is the solution of the following problem:

$$\min_t \| E^T t \|^2, \text{ subject to } \| t \| = 1 \tag{21}$$

Consequently, *t* is the unit eigenvector of EE^T corresponding to the smallest eigenvalue. If the sign of *E* is correct, the ambiguity of the sign of t can be resolved as follows. Indeed, for j^{th} correspondence (p_{1j}, p_{2j}), if

$$(t \times \widetilde{p_{2j}}) \bullet (E\widetilde{p_{1j}}) > 0 \tag{22}$$

then the sign of *t* is compatible with the sign of *E*; otherwise, reverse the sign of ***t***.

We now turn to the problem of estimating the rotation matrix R. As $E = [t]_X R$ by definition, we find R by solving

$$\min_R \sum \| E - [t]_X R \|^2,$$

$$\text{subject to } R^T R = I \text{ and } \det(R) = 1 \tag{23}$$

since

$$E - [t]_X R = (ER^T - [t]_X)R,$$

$$\| E - [t]_X R \|^2 = \| ER^T - [t]_X \|^2.$$

The above problem become

$$\min_R \sum_{j=1}^{3} \| \mathrm{Re}_j - \tau_j \|^2,$$

subject to $RR^T = I$ and $\det(R) = 1$ (24)

where e_j and τ_j are the j^{th} row vectors of matrices E and $[t]_X$, This can be easily solved using the quaternion representation of 3-D rotations.

4.2、3D reconstruction

Once we know the motion (*R, t*), given a match (m_1, m_2), we can estimate the 3D coordinates *M* . The method of 3D reconstruction is described below.

Under the pinhole model, we have the following two equations:

$$s_1 m_1 = A_1 [I \quad 0] M \qquad (25)$$

$$s_2 m_2 = A_2 [R \quad t] M \qquad (26)$$

where A_1 and A_2 are the intrinsic matrices of the first and second cameras,

It is a straightforward matter to formulate a linear least-squares to estimate the 3D coordinates M by eliminating s_1 and s_2 from (25) and (26). Let $m_1 = [u_1, v_1]^T, m_2 = [u_2, v_2]^T$ and $B_2 = A_2 R$, then we have

$$\begin{bmatrix} a_1^T - u_1 a_3^T \\ a_2^T - v_1 a_3^T \\ b_1^T - u_2 b_3^T \\ b_2^T - v_2 b_3^T \end{bmatrix} M = \begin{bmatrix} 0 \\ 0 \\ (u_2 c_3 - c_1)^T t \\ (v_2 c_3 - c_2)^T t \end{bmatrix} \qquad (27)$$

or $$ZM = z \qquad (28)$$

where a_i^T, b_i^T, c_i^T are respectively the i^{th} row of matrices A_1, B_2, A_2 . The solution is given by $\widehat{M} = (Z^T Z)^{-1} Z^T z$. A more elaborate technique consists in minimizing the distance between the back-projection of the 3D reconstruction and the observed image points, that is

$$M = \arg\lim_{M} (\| m_1 - h_1(a, M) \|^2 + \| m_2 - h_2(a, M) \|^2) \qquad (29)$$

where $h_1(a, M), h_2(a, M)$ are the camera projection functions corresponding to (25) and (26) respectively.

The reconstructed 3-D points are referred to the left camera reference frame.

5. EXTRINSIC PARAMERTERS ESTIMATION

The extrinsic parameters are defined as parameters (rotate matrix R and translation vector T) that identify uniquely the transformation between the camera reference frame and the world reference frame.

We realized the computer vision technique in the trampoline sport simulation and training system to compare the athlete motion of trampoline sport video to the normal motion of 3D virtual athlete in simulation system. In order to make the view point of 3D model motion of simulation system appear the same as that in the sport video, we regard the world reference frame in the simulation system and the 3D space of feature reconstruction as the same. The trampoline plane is regarded as XZ plane of the world reference frame. That is, we regard one of the long sides of the trampoline as the X-axis of world reference frame and one of the short sides that intersect with the X-axis side of the trampoline as the Z-axis. The intersection point for the X-axis and Z-axis sides of the trampoline is regarded as the origin of world reference frame. So the Y-axis of world reference frames is the axis that is orthogonal with the X and Z-axis according to the right hand rule. The selection of world reference frame is illustrated in figure1.

If we have got the 3D coordinate of three features corresponding with the intersection point of the trampoline sides as A、B and O that are illustrated in figure4 , we will get the camera extrinsic parameters and then the view point of virtual camera can be automatically adjusted in simulation system.

The particular steps are listed below:

1）、Because the intersection points of the trampoline sides are sure to be textured patches with a large variation of intensity in the two directions *x* and *y,* some features located at the point intersection

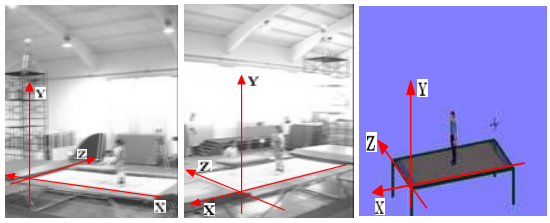


Figure 1. The selection of world reference frames.

are supposed to be automatically extracted in the first frame and well tracked in the successive frame. Select three points that lie at the intersection of the trampoline sides among the successfully tracked feature points by hand. Assume that *A* is the extracted feature point lying at the intersection of the

long and short sides of the trampoline. Make *B* and *O* extracted in the same way. The selection of A、B and O is illustrated in figure 4. We regard the camera coordinate frames of first image as camera reference frame. And the 3-D reconstruction feature coordinate is referred to the camera reference frame. After the 3-D reconstruction, the 3-D coordinate of *A* 、 *B* and *O* are (a_1, a_2, a_3) , (b_1, b_2, b_3) and (o_1, o_2, o_3) respectively. We regard *O* as the origin of the world reference frame.

2)、 Vector $\overrightarrow{AO} = (o_1 - a_1, o_2 - a_2, o_3 - a_3)$ and $\overrightarrow{BO} = (o_1 - b_1, o_2 - b_2, o_3 - b_3)$ are coherent with the X axis and Z-axis of world reference frames. Since there are noise in the image, the vectors $\overrightarrow{AO}$, $\overrightarrow{BO}$ isn't orthogonal, we should make them orthonormal. Then the vector $\overrightarrow{AO} \times \overrightarrow{BO}$ is coherent with the Y-axis of world reference frame.

3）、From the vector $\overrightarrow{AO}$ 、 $\overrightarrow{BO}$ and $\overrightarrow{AO} \times \overrightarrow{BO}$, we can get the matrix of direction consine matrices easily and then transfer it to the Euler angular coordinate[Jia99]. Therefore we get the rotation matrix described as Euler angle.

4）、The 3-D coordinate (o_x, o_y, o_z) of the origin of world reference frame referred to the camera reference frame is regarded as the translate vector $T = (T_x, T_y, T_z)$.

5）、With the extrinsic parameters, we can adjust the virtual camera of simulation system: rotate the virtual camera axis according to the Euler angle of rotate matrix *R* and translate the origin according to the translate vector *T*.

6. EXPERIMENTAL RESULTS

We shoot the trampoline sport video from two different view point with digital handle camera. Since we have to extract the intrinsic parameters of the camera, the camera is control to make a slight shake at the beginning of the sport video (because some sport video, such as the videos of diving and gymnastics, has included camera motion in it, the shake will be unnecessary).

We can get the camera parameters from the segment of video that includes camera motion and achieve the view point adjustment of the trampoline sport simulation and training system.

Results presented below are organized into three parts. First we present results of the tracking method applied to the two sequences of the trampoline sport images. Secondly we present the quality of the reconstruction of points A、B and O lying in the intersection of the trampoline sides. At last, we present final results of the computer vision technique applied in the trampoline sport simulation and train system.

6.1 、 The feature tracking and outlier elimination in the trampoline sport video sequences

Given two sequences of trampoline sport images, characteristic points are extracted in the first image and tracked through the images of the sequence. Below we show this process (figure 2 & 3) with the tracked points drawn in the first and last frame of a five- image sequence.

In figure 2a and 3a, 150 feature points are exacted from the first trampoline sport image, and despite the complexity of the image, 112 and 83 features are positively tracked through the sequence (figure 2b and 3b) respectively. The points are well distributed in the image. And with the outlier elimination using the RANSAC algorithm, we ensure the validity of feature matching in the two image sequences.

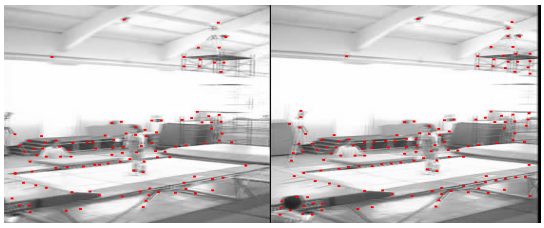
The experiments carried out on the two image sequences confirm the efficiency and the robustness of the KLT algorithm while processing sport video sequences.

6.2、3-D reconstruction

Based on the tracking results of the trampoline sport sequences, the next step we should take is feature reconstruction. We do not need to reconstruct all successfully tracked features. We just have to reconstruct the 3D coordinate of the three features *A, B* and *O* lying in the intersection of the trampoline sides. So we have to select the intersection point *A, B* and *O* of the trampoline sides by hand from the successfully tracked features in the fifth frame of the trampoline sport sequences as illustrate in figure 4. The 3D coordinates are referred to the camera reference frame. The nonlinear minimization in (29) is done with the Levenberg-Marquardt algorithm implemented in the Minpack library.

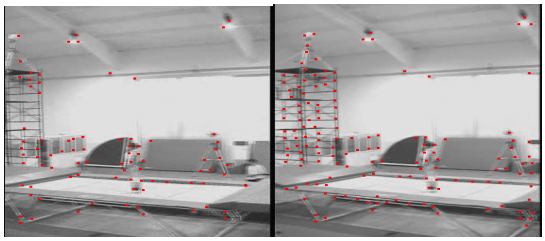
The reconstruction results of the selected features are displayed in figure 5. We can see that the relationship of the reconstructed features A、B and O are consistent with the actual pattern in figure 4.

The reconstruction is up to a scale factor. In order to get the real distance between the origins of the camera reference frame and the world reference frame, some related knowledge is used. The length of long and short sides of trampoline is normal. So we can get the scale factor easily and get the real translate vector *T*.



2a: image1 *2b: image5*

Figure 2. Sequence 1 of the trampoline sport video.



3a: image1 *3b: image5*

Figure 3. Sequence 2 of the trampoline sport video.

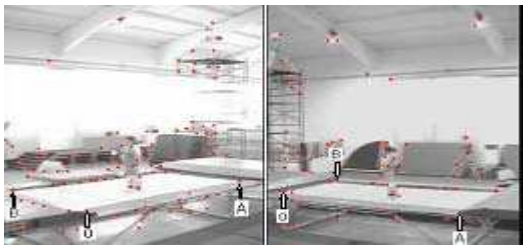


Figure 4. The 3 points selected by hand from the successfully tracked features of two sequences in the trampoline sport video. The vector $\overrightarrow{AO}$、$\overrightarrow{BO}$ and $\overrightarrow{AO}\times\overrightarrow{BO}$ are corresponding to the X-axis、Z-axis and Y-axis of world reference frame respectively.

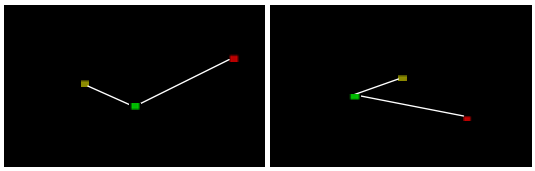


Figure 5. The 3 reconstructed points A、B and O seen from the origin of the virtual camera in OpenGL. The green, red and yellow points are corresponding to the O、A and B respectively.

6.3、The use of extrinsic parameters in the trampoline sport training system

From the first few image frames in the trampoline sport video sequences, we get the estimation of extrinsic parameters, which are used in the trampoline sport simulation and training system. Below we show this process with the use of extrinsic parameters (figure 6).

The trampoline sport training system and the relevant sport video are displayed on the same screen. At first, the view point of virtual camera in the trampoline sports training system is adjusted according to the extrinsic parameters in the first frame (figure 61a & figure 62a). And then, the successive frames are displayed at the same frame rate as in the trampoline sports video. Figure 61b-61d and figure 62b-62d are the frames extracted from the successive image sequences. We can see that the view point in simulation system and that in the video are just similar and the motions are just synchronous. So we can quickly and accurately find the difference between the human motion in video and the simulation normal virtual athlete motion in training system.

7. CONCLUSION

We present a novel computer vision technique based on sport video and have made it work in the trampoline sport simulation and training system. From the experimental results, we find that it works well and has extended the function of sport training system greatly. The technique can extract the intrinsic and extrinsic parameters steadily from the sport video that includes camera motion. It can be used in diving or gymnastics sport training system as well.

In the trampoline sports video, we can see that the camera keeps still during the actions of the athletes. While in other sport such as diving, the camera is moving. A future study will aim at the automatic extraction of global motion from sport video real-time and online. And at the same time, we control the virtual camera of sports simulation system so that it can perform the same motion as sport video does. As a matter of fact, that is just the project we are working on and will be finished in not a long time.

8. ACKNOWLEDGEMENTS

This work is supported by National Natural Science Foundation of China(NSFC), Grant 60103007; 863 Plan of China, Grant 2001AA115131; National Special Item for Olympics, Grant 2001BA904B08; National 973 Project, Grant 2002CB312104 and Knowledge Innovation Item, Grant 20036070. We thank for the contributions to this work from all our colleagues in the Digital laboratory. Special thanks go to Wu Yong-dong, Liu Li, Li Yan and Wei Yi.

9. REFERENCES

［Fau92］、 O. Faugeras, D. Luong, and Q. Maybank. Camera self-calibration: Theory and experiments. ECCV'92, Lecture notes in Computer Science, 588:321-334, 1992.

［Luo97］、 Q. Luong and O. Faugeras. Self-calibration of a moving camera from point correspondences and fundamental matrices. IJCV, 22(3):261{89, 1997.

［Der90］、Deriche, R. and Faugeras, O., 2D Curve matching using high curvature points : Application to stereo vision, In Proceedings International Conference on Pattern Recognition, pp. 240-242, 1990.

［Har88］、Harris, C. and Stephens, M., A combined corner and edge detector, In Proceedings fourth Alvey Vision Conference, pp 174-151, Manchester, England, August 1988.

［Sis00］、Sistiaga, M., Navigation référencée images de terrain pour engins sous-marins, Ph.D Thesis, University of Montpellier II, September 2000.

［Shi94］、Shi, J. and Tomasi, C., Good Features to Track, In IEEE Conference on Computer Vision and Pattern Recognition, Seattle, June 1994.

［Har95］、Hartley，R.,In Defence of the 8-points Algorithm, In Proceedings Fifth International Conference on Computer Vision ,Cambridge, Mass.,June 1995.

［Fis81］、Fischler, M.A. and Bolles, R.C., Random Sample Concensus: a paradigm for model fitting with application to image analysis and automated cartography, Communication Association and Computing Machine, 24(6),pp. 381-395,1981.

［Har97］、Hartley, R. and Sturm, P., Triangulation, In Computer Vision and Image Understanding, 68(2), pp. 146-157, 1997.

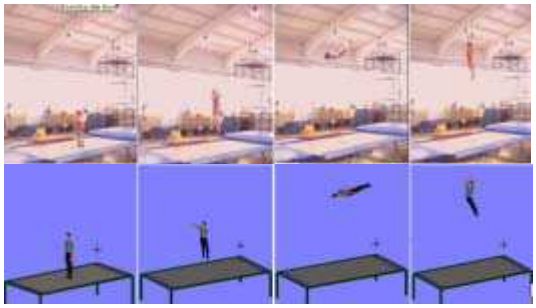
［Men99］、Mendonca,P. and Cipolla, R., A Simple Technique for Self-Calibration, In Conference on Computer Vision and Pattern Recognition, 1999.

［Har00］、Hartley R, Zisserman A. Multiple View Geometry in Computer Vision. Cambridge: Cambridge University Press, 2000.

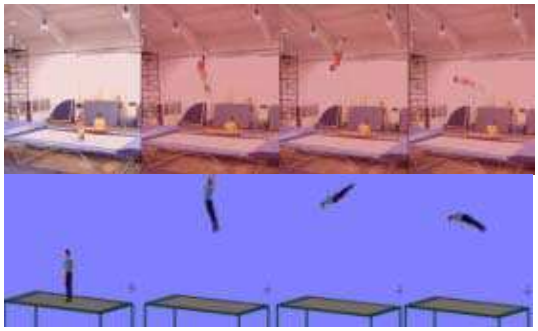
［Dav03］、David A. Forsyth and Jean Ponce, Computer Vision：A Modern Approach, Prentice Hall press. 2003.

［Jia99］、Jiazhen Hong, Computational Dynamics of Multibody Systems, Higher Education press China. 1999.

［Zhang96］、Zhang, Z, A new multistage approach to motion and structure estimation: From essential parameters to euclidean motion via fundamental matrix, Research Report 2910, INRIA Sophia-Antipolis, France, 1996.



61a *61b* *61c* *61d*



62a *62b* *62c* *62d*

Figure 6. The simulated normal motion of virtual athlete in the trampoline sport training system and the motion in the video are displayed on the same screen: The view point of virtual camera is automatically adjusted in the first frame according to the extrinsic parameters extracted from the video and the successive frames are displayed at the same rate as in the trampoline sport video.