Computing and Informatics, Vol. 21, 2002, 591–606

ROBUST MULTIPLE-VIEW GEOMETRY ESTIMATION BASED ON GMM*

Mingxing Hu, Qiang Xing, Baozong Yuan, Xiaofang Tang

Institute of Information Science
Northern Jiaotong University
Beijing, China 100044
e-mail: gary_hu@263.net

Manuscript received 10 June 2002; revised 2 December 2002
Communicated by Horst Bischof

Abstract. Given three partially overlapping views of the scene from which a set of point or line correspondences have been extracted, 3D structure and camera motion parameters can be represented by the trifocal tensor, which is the key to many problems of computer vision on three views. Unlike in conventional typical methods, the residual value is the only rule to eliminate outliers with large value, we build a Gaussian mixture model assuming that the residuals corresponding to the inliers come from Gaussian distributions different from that of the residuals of outliers. Then Bayesian rule of minimal risk is employed to classify all the correspondences using the parameters computed from GMM. Experiments with both synthetic data and real images show that our method is more robust and precise than other typical methods because it can efficiently detect and delete the bad corresponding points, which include both bad locations and false matches.

Keywords: Multiple-view geometry, Gaussian mixture model, trifocal tensor

1 INTRODUCTION

Given three partially overlapping views of the scene from which a set of point or line correspondences have been extracted, 3D structure and camera motion parameters

* The research is supported by: National Natural Science Foundations (No. 69789301) and Doctoral Foundations of China (No. 97000409).

can be represented by a $3 \times 3 \times 3$ tensor, named trifocal tensor. And the associated trilinear constraints are now considered as the fundamental equations for several problems related to three views, such as motion analysis [1], self-calibration [2], and view synthesis [3].

Therefore, the geometry of perspective projection for images and the trilinear constraints have provoked much interest. First Spetsakis and Aloimonos[4], then Weng el al. [5] pointed out that there is a constraint on the projected positions of lines over three images, demonstrating how this might be used for reconstruction in the calibrated case. Then Shashua [6] showed that the coordinates of three corresponding points satisfied a set of algebraic relations of degree 3, called trilinear constraints. It was later on pointed out by Hartley[7] that those trilinear constraints were in fact arising from a tensor that governed the correspondences of lines among three views, which he called the trifocal tensor. Different from the above methods, Faugeras[8] et al. derived the trifocal tensor based on Grassmann-Cayley algebra; they gave a set of algebraic constraints, which are satisfied by the 27 coefficients of the trifocal tensor and allowed to parameterize it minimally with 19 coefficients. Recently, Torr and Zisserman [9] have presented a robust algorithm for computing a maximum likelihood estimation (MLE) of the trifocal tensor. And Gideon [10] also investigates the linear degenerations of projective structure estimation from line features across three views, when the scene is a Linear Line Complex (a set of lines in space intersecting at a common line).

From all that has been presented above, we can see that the methods either focus on the new geometrical interpretation of trilinear constraints or estimate the trifocal tensor by starting with a line solution and improving it further through employing a numerical Gauss-Newton style iterative procedure. However, in many applications, image data not only are noisy, but also contain outlier data that are in gross disagreement with a specific postulated model. Outliers, which are inevitably included in an initial fit, can so distort a fitting process that the fitted parameters become arbitrary. Nevertheless, the above methods either pay less or no attention to the situation, or only use the residual value as the only rule to classify correspondences into inliers and outliers, though it is possible that even a false match can have a small residual value when robust algorithms such as M-estimators [12] and RANSAC [13] (random sample consensus paradigm) are used.

In this paper, we present an approach to multiple-view geometry estimation based on Gaussian mixture model (GMM) [15, 16]. GMM represents a statistical pattern recognition approach to machine monitoring that enables optimal processing of data both for training the classifier (EM algorithm) and for performing on-line classification. In addition, while GMM possesses many of discriminant surface modeling capabilities of more complex nonparametric classifiers, the GMM is parametric, making it more robust to the effects of a limited amount of training data. In our work, Gaussian mixture model is built assuming that the residuals corresponding to inliers comes from Gaussian distributions different from that of the residuals of outliers. Then Bayesian rule of minimal risk is employed to classify all the correspondences. Experiments with both synthetic data and real images show that our

method is more robust than other typical algorithms and relatively unaffected by outliers.

The rest of the paper is organized as follows. Section 2 provides a brief overview of the trifocal tensor and several robust methods for estimation. In Section 3, a new approach to multiple-view geometry estimation using GMM is presented in detail, including problem formulation, model building and decision rule. Experimental results with synthetic data and real images are described in Section 4. Finally, conclusion is given.

Notations: we use the covariant-contravariant summation convention: a point is an object whose coordinates are specified with superscripts, i.e., $\mathbf{p}^i = (p^1, p^2, \Lambda)$. These are called contravariant vectors. The element in the dual space (representing hyper-planes — lines in P^2) is called a covariant vector and is represented by subscripts, i.e., $\mathbf{s}_j = (s^1, s^2, \Lambda)$. Indices repeated in covariant and contravariant forms are summed over, i.e., $\mathbf{p}^i\mathbf{s}_i = (p^1 s_1 + p^2 s_2 + \Lambda + p^n s_n)$. This is known as contraction. An outer-product of two 1-valence tensors (vectors), $\mathbf{a}_i\mathbf{b}^j$, is a 2-valence tensor (matrix) $\mathbf{c}_i^j$ whose i, j, entries are $\mathbf{a}_i\mathbf{b}^j$ —note that in matrix form $\mathbf{C} = \mathbf{b}\mathbf{a}^T$.

2 TRIFOCAL TENSOR AND ROBUST METHODS FOR ESTIMATION

Consider a single point $\mathbf{X}$ in space projected onto 3 views with camera matrices $\mathbf{P}$, $\mathbf{P}'$, $\mathbf{P}''$, with image points $\mathbf{p}$, $\mathbf{p}'$, $\mathbf{p}''$, respectively. Note that $\mathbf{X} = (x, y, 1, \Lambda)$ for some scalar λ. Consider $\mathbf{P} = [\mathbf{1}|\mathbf{0}]$ and $\mathbf{P}' = [\mathbf{A}|\mathbf{v}']$ where $\mathbf{A}$ is the 3×3 principle minor of $\mathbf{P}'$ and $\mathbf{v}'$ is the fourth column of $\mathbf{P}'$. Consider $\mathbf{p}' \cong \mathbf{P}'\mathbf{X}$ and eliminate the scale factor:

$$x' = \frac{\mathbf{a}_i^T x}{\mathbf{a}_3^T x} = \frac{\mathbf{a}_1^T \mathbf{p} + \lambda \mathbf{v}'_1}{\mathbf{a}_3^T x + \lambda \mathbf{v}'_3} \tag{1}$$

$$y' = \frac{\mathbf{a}_2^T x}{\mathbf{a}_3^T x} = \frac{\mathbf{a}_2^T \mathbf{p} + \lambda \mathbf{v}'_2}{\mathbf{a}_3^T x + \lambda \mathbf{v}'_3} \tag{2}$$

where $\mathbf{a}_i$ is the i-th row of $\mathbf{A}$. These two equations can be written more compactly as follows:

$$\lambda \mathbf{s}'^T \mathbf{v}' + \mathbf{s}'^T \mathbf{A}\mathbf{p} = 0 \tag{3}$$

$$\lambda \mathbf{s}''^T \mathbf{v}' + \mathbf{s}''^T \mathbf{A}\mathbf{p} = 0 \tag{4}$$

where $\mathbf{s}' = (0, -1, y')$ and $\mathbf{s}'' = (0, -1, y')$. Yet in a more compact form consider $\mathbf{s}'$, $\mathbf{s}''$ as row vectors of the matrix

$$\mathbf{s}_j^\mu = \begin{bmatrix} -1 & 0 & x' \\ 0 & -1 & y' \end{bmatrix}, \tag{5}$$

where $j = 1, 2, 3$ and $\mu = 1, 2$. Therefore, the compact form we obtain is described below:

$$\lambda \mathbf{s}_j^\mu \mathbf{v}'^j + \mathbf{p}^i \mathbf{s}_j^\mu \mathbf{a}_i^j = 0 \tag{6}$$

where μ is a free index (i.e., we obtain one equation per range of μ). Similarly, let $\mathbf{P}'' = [\mathbf{B}|\mathbf{v}'']$ for the third view $\mathbf{p}'' \cong \mathbf{P}''\mathbf{X}$ and let $\mathbf{r}_k^\rho$ be the matrix,

$$\mathbf{r}_k^\rho = \begin{bmatrix} -1 & 0 & x'' \\ 0 & -1 & y'' \end{bmatrix} \tag{7}$$

and likewise,

$$\lambda \mathbf{r}_k^\rho \mathbf{v}''^k + \pi^i \mathbf{r}_k^\rho \mathbf{b}_i^k = 0, \tag{8}$$

where $\rho = 1, 2$ is a free index. We can eliminate λ from (6) and (8) and obtain a new equation:

$$p^i \mathbf{s}_j^\mu \mathbf{r}_k^\rho \left(\mathbf{v}'^j \mathbf{b}_i^k - \mathbf{v}''^k \mathbf{a}_i^j \right) = 0 \tag{9}$$

and the term in parenthesis is a trivalent tensor we call the trilinear tensor:

$$\mathbf{T}_i^{jk} = \mathbf{v}'^j \mathbf{b}_i^k - \mathbf{v}''^k \mathbf{a}_i^j. \tag{10}$$

Hence, we have four trilinear equations (note that $\mu, \rho = 1, 2$). In more explicit form, these trilinearities look like:

$$\begin{cases} x''\mathbf{T}_i^{13}p^i - x''x'\mathbf{T}_i^{33}p^i + x'\mathbf{T}_i^{31}p^i - \mathbf{T}_i^{11}p^i = 0 \\ y''\mathbf{T}_i^{13}p^i - y''x'\mathbf{T}_i^{33}p^i + x'\mathbf{T}_i^{32}p^i - \mathbf{T}_i^{12}p^i = 0 \\ x''\mathbf{T}_i^{23}p^i - x''y'\mathbf{T}_i^{33}p^i + y'\mathbf{T}_i^{31}p^i - \mathbf{T}_i^{21}p^i = 0 \\ y''\mathbf{T}_i^{23}p^i - y''y'\mathbf{T}_i^{33}p^i + y'\mathbf{T}_i^{33}p^i - \mathbf{T}_i^{22}p^i = 0 \end{cases} \tag{11}$$

Equation (11) was first introduced by Shashua in [6], from where we can see that the trifocal tensor has 27 elements, but only their ratios are significant, leaving 26 coefficients to be specified. Each triplet of point correspondences can provide four independent linear equations for the elements of the tensor. Therefore the tensor can be computed from a minimum of 7 points using a linear algorithm (LA). But it is too sensitive to noise and outliers.

However, the tensor has only 18 independent degrees of freedom, which can be seen by considering three 3×4 projection matrices, less 15 projective degrees of freedom, that is, $3 \times 11 - 15 = 18$. Then six points is enough to estimate the trifocal tensor with computing an invariant of six points from three views [13, 11]. The method involves the solution of a cubic, and correspondingly provides one or three real solutions for the trifocal tensor. Consequently the best of them is selected as the final solution, when measuring support for each from the full set of correspondences.

An alternative to the LA is M-estimator [12], whose aim is to follow maximum-likelihood formulations by deriving optimal weighting for the data. The estimators minimize the sum of a symmetric, positive-definite function $\rho(d_i)$ of the d_i. That is, the parameters are sought that minimize

$$\sum_i^n \rho(d_i) = \sum_i^n (\gamma_i d_i)^2. \tag{12}$$

The form of ρ is derived from a particular chosen density function so that ρ is some weighting, $\rho(d_i) = (\gamma_i d_i)^2$.

$$d_i^2 = (\hat{\mathbf{p}}_i - \mathbf{p}_i)^2 + (\hat{\mathbf{p}}'_i - \mathbf{p}'_i)^2 + (\hat{\mathbf{p}}''_i - \mathbf{p}''_i)^2, \tag{13}$$

where $\hat{\mathbf{p}}_i$, $\hat{\mathbf{p}}'_i$, $\hat{\mathbf{p}}''_i$ are estimated point correspondences, just like the measured point correspondences $\mathbf{p}_i \leftrightarrow \mathbf{p}'_i \leftrightarrow \mathbf{p}''_i$. A typical weighting scheme in the statistics literature is

$$\gamma_i = \begin{cases} 1 & d_i < \delta \\ \delta/|d_i| & \delta < d_i < 3\delta \\ 0 & d_i > 3\delta \end{cases} \tag{14}$$

where δ is the standard deviation of the error, estimated from the median $\delta = \frac{med_i d_i}{0.6745}$.

The experiments show that M-estimator is robust to those outliers, which are produced by bad location. It is, however, not robust to false matches, because it is highly vulnerable to poor starting conditions, which makes the algorithm converge to a local minimum.

Torr et al. [9] gave another highly robust fitting algorithm — the random sample consensus paradigm (RANSAC). Rather than using as many data as possible to obtain an initial solution and then attempting to identify outliers, as small a subset of data as feasible to estimate the parameters is used (e.g., seven triplet of correspondences for a trifocal tensor), and this process is repeated enough times m on different subset to ensure that there is a 95% chance that one of the subsets will contain only good data points. As pointed out by Fischler and Bolles [13], the number m of samples is chosen by making the probability

$$\tau = 1 - (1 - (1 - \varepsilon)^q)^m, \tag{15}$$

where ε is the fraction of contaminated data, and q the number of features in each sample. Outliers are typically discriminated from inliers by using

$$i \in \begin{cases} \textit{set of inlier} & \text{if } d_i \leq 1.96\delta \\ \textit{set of outlier} & \text{otherwise} \end{cases}, \tag{16}$$

where $\delta = 1.4828\left(1 + \frac{5}{n-f}\right)\sqrt{med_i|d_i|}$, n is the number of data, and f the dimensionality of the parameter. Then the best solution is that which maximizes the number of points whose residual is below a threshold. Experiments show that the convergence of RANSAC is superior to that of M-estimator, and the solution is typically more accurate. But RANSAC only uses random sample to search for the optimal solution; when a large number of outliers are involved, the computation efficiency will decrease significantly.

In [14], Torr and Zisseramn also addressed another robust estimator MLESAC for epipolar geometry estimation, which is taken as generalization of the RANSAC estimator. It adopts the same sampling strategy as RANSAC to generate putative

solutions, but chooses the solution that maximizes the likelihood rather than just the number of inliers.

First, the noise in the two images is assumed to be Gaussian with zero mean and uniform standard deviation σ, and the outlier distribution is uniform with $\left[-\frac{v}{2}, \frac{v}{2}\right]$ being the pixel range within which outliers are expected to fall. Thus the error is modeled as a mixture model

$$Pr = \left(\gamma \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{e^2}{2\sigma^2}\right) + (1-\gamma)\frac{1}{v}\right), \tag{17}$$

where γ is the mixing parameter. It can be estimated by using expectation maximization (EM) algorithm.

Thus the error minimized is the negative log likelihood

$$-L = \sum_i \log\left(\gamma\left(\frac{1}{\sqrt{2\pi\sigma}}\right)'' \exp\left(-\frac{\sum\left(\underline{x}_i^j - x_i^j\right)^2 + \left(\underline{y}_i^j - y_i^j\right)^2}{2\sigma^2}\right) + (1-\gamma)\frac{1}{v}\right). \tag{18}$$

So the output of MLESAC is an initial estimate, together with a likelihood that each correspondence is consistent with the epipolar constraint. Then it is improved using a gradient descent method. In this paper, we also employ MLESAC for trifocal tensor estimation, and compare its results with those of our method.

3 ESTIMATION BASED ON GAUSSIAN MIXTURE MODEL

From the methods presented above, M-estimator and RANSAC both take outliers into account, but they do not work well for some reasons. The M-estimator tries to blindly compensate the effect of outliers by replacing the Gaussian distribution assumption by a long tailed distribution. The performance of the M-estimator therefore depends on how well the new distribution corresponds to the actual residual which is, however, unknown a priori. On the other hand, the RANSAC algorithm does not perform well if a great number of outliers are involved in the estimation, because it only employs a simple threshold to classify correspondences (inliers and outliers), which are perturbed by different reasons (Gaussian image noise, or bad locations and false matches).

Our solution to the multiple-view geometry estimation problem is based on the fact that the residuals corresponding to inliers come from distributions different from that of the residuals of outliers.

3.1 Problem Formulation

In the following, it is assumed for simplicity, but without loss of generality, that the residuals of inliers are considered normally distributed, and those of outliers are considered to follow other Gaussian distributions on each image coordinate in

all three views. Thus the density model of residuals can be viewed as a type of mixture model, which comprises a couple of component Gaussian functions, together providing a multi-model density.

Let $\mathbf{u}_i$ be a vector of the measured x, y coordinates in each image of a correspondence i over three views, namely, $\mathbf{u}_i = (x_i, y_i, x_i', y_i', x_i'', y_i'')$, where $i = 1, \Lambda, n$ labels points. Thus given a true correspondence vector (perfect or noise free quantity), $\bar{\mathbf{u}}_i = (\bar{x}_i, \bar{y}_i, \bar{x}_i', \bar{y}_i', \bar{x}_i'', \bar{y}_i'',)$, the residual of the correspondence $\mathbf{v}_i = \mathbf{u}_i - \bar{\mathbf{u}}_i$ is a six-dimensional feature vector and $\mathbf{V} = \{\mathbf{v}_1, \mathbf{v}_2, K, \mathbf{v}_n\}$ is the feature vectors set. We use the following probability density function to describe the distribution of $\mathbf{v}_i$

$$P(v_i|\Theta) = \sum_{m=1}^{K} \alpha_m P_m(\mathbf{v}_i|\theta_m) \tag{19}$$

where $K = 2$ is the model order, or the number of components to be incorporated into the mixture model. α_m is the weight of each single Gaussian component. It corresponds to the prior probability that feature vector $\mathbf{v}_i$ is generated by component θ_m and

$$\sum_{m=1}^{K} \alpha_m = 1, \quad \alpha_m \geq 0. \tag{20}$$

Each model density $P_m(\mathbf{v}_i|\theta_m)$ is a K — variate Gaussian function of the form

$$P_m(\mathbf{v}_i|\theta_m) = \frac{1}{(2\pi)^{\frac{K}{2}}|\Sigma_m|^{\frac{1}{2}}} \exp -\frac{1}{2}(\mathbf{v}_i - \mu_m)^T \Sigma_m^{-1}(\mathbf{v}_i - \mu_m) \tag{21}$$

with mean vector μ_m and covariance matrix Σ_m.

3.2 GMM Building and Decision Rule

In practice, the true correspondences cannot be obtained directly from the image data, therefore we use the estimated correspondence vector $\hat{\mathbf{u}}_\mathbf{i} = (\hat{x}_i, \hat{y}_i, \hat{x}_i', \hat{y}_i', \hat{x}_i'', \hat{y}_i'')$ instead. Then the residual of correspondence $\mathbf{v}_i$ is rewritten as

$$\mathbf{v}_i = \mathbf{u}_i - \hat{\mathbf{v}}_i = (x_i - \hat{x}_i, y_i - \hat{y}_i, x_i' - \hat{x}_i', y_i' - \hat{y}_i', x_i'' - \hat{x}_i'', y_i'' - \hat{y}_i'') \tag{22}$$

From the experiment, we find that the range of $\mathbf{v}_i$ is too wide to be employed directly for mixture model estimation. Thus we normalize it with the following equation

$$\mathbf{v}_i = \left(\frac{x_i - \hat{x}_i}{\|\mathbf{x} - \hat{\mathbf{x}}\|}, \frac{y_i - \hat{y}_i}{\|\mathbf{y} - \hat{\mathbf{y}}\|}, \frac{x_i' - \hat{x}_i'}{\|\mathbf{x}' - \hat{\mathbf{x}}'\|}, \frac{y_i' - \hat{y}_i'}{\|\mathbf{y}' - \hat{\mathbf{y}}'\|}, \frac{x_i'' - \hat{x}_i''}{\|\mathbf{x}'' - \hat{\mathbf{x}}''\|}, \frac{y_i'' - \hat{y}_i''}{\|\mathbf{y}'' - \hat{\mathbf{y}}''\|} \right), \tag{23}$$

where $\|w\| = (\sum_{i=1}^{n} w_i^2)^{\frac{1}{2}}$.

Because the feature vectors are assumed to be independent and each element belongs to some Gaussian distribution, it is easy to derive that the normalizing process will not destroy the Gaussian mixture model at all.

Given $\mathbf{V}$, our goal in the stage of building the mixture model is to estimate the GMM parameters, including mean vectors, covariance matrices, and mixture weights of each model. This way we get the maximum value of $P(\mathbf{V}|\mathbf{\Theta})$

$$P(\mathbf{V}|\mathbf{\Theta}) = \prod_{i=1}^{n} P(\mathbf{v}_i|\mathbf{\Theta}). \tag{24}$$

To estimate Θ we apply log-function which is monotonically increasing to simplify the problem. Thus the following equation is obtained:

$$f = \nabla_{\Theta}(\ln P(\mathbf{V}|\mathbf{\Theta})) = 0. \tag{25}$$

Then parameter estimates can be obtained iteratively using an expectation and maximization (EA) algorithm. EM is a well established maximum likelihood algorithm for fitting a mixture model to a set of training data. On each EM iteration, the re-estimation formulas are used which guarantee a monotonic increase in the likelihood value of the model:

$$\mu_m = \frac{\sum_{i=1}^{n}(P_{m,i}\mathbf{v}_i)}{\sum_{i=1}^{n} P_{m,i}} \tag{26}$$

$$\sum_m = \frac{\sum_{i=1}^{n} P_{m,i}(\mathbf{v}_i - \mu_m)(\mathbf{v}_i - \mu_m)^T}{\sum_{i=1}^{n} P_{m,i}} \tag{27}$$

$$\alpha_m = \frac{\sum_{i=1}^{n} P_{m,i}}{n} \tag{28}$$

where the posterior probability for class m in the j-th partitioned region is given by

$$P_m^i = \frac{\alpha_m P(\mathbf{v}_i|\theta_m)}{P(v_i|\Theta_m)}. \tag{29}$$

After the model building, to decide to which class a new-coming sample $\mathbf{v}$ belongs, we use the Bayesian rule of minimal risk which assigns $\mathbf{v}$ to the class that maximizes the class posterior probability

$$\Pr\{class = m\} = \frac{\alpha_m P(\mathbf{v}_i|\theta_m)}{\sum_{i=1}^{K} \alpha_i P(\mathbf{v}_i|\theta_i)} \tag{30}$$

where the prior of a particular class is estimated by the proportion of samples of that class used for building.

3.3 Stages of Estimation Based on GMM

From the description presented above, the corresponding residuals are the only information we can use for mixture model building. Thus how well the trifocal tensor is computed for initial is of great importance to building a more accurate mixture

model. In our experiment, RANSAC is employed first to estimate the multiple-view geometry, through which the residual of correspondence has perhaps more implicit information for classification of the inliers and outliers. Then our proposed method for the trifocal tensor estimation can be summarized as follows.

1. Compute an initial estimate for the multiple-view geometry with the RANSAC algorithm. One thing we would like to emphasize is that we need not iterate the search steps as much as the pure RANSAC method does. Because with the following steps, we will get more statistical information of inliers and outliers, which helps us classify all the correspondences, and the application of RANSAC is only for initialization. Only about half of subsets (six correspondences for one sample) that pure RANSAC has used are employed in the experiment and the best solution that maximizes the number of inliers is selected.
2. Normalize the corresponding residuals $\mathbf{v}_i$ with equation (23), in order to decrease the range of $\mathbf{v}_i$.
3. Build the Gaussian mixture model using the normalized residuals. One advantage of our method is its fast convergence: in our experiments the convergence is usually obtained after not more than 20 iterations.
4. Use Bayesian rule to classify the correspondences into inliers and outlier with GMM.
5. Re-estimate the trifocal tensor using the inliers provided by the last step.

4 EXPERIMENTAL RESULTS

In this section we will discuss the result of multiple-view geometry estimation based on GMM, using both synthetic data and real images. In order to compare how well these robust methods (M-estimator, MLESAC, GMM) will classify the correspondences, they are employed first to detect outliers, then the inliers computed from the above step are used to re-estimate the trifocal tensor with linear algorithm.

4.1 Experiments with Synthetic Data

In our experiments, the correspondences are randomly generated by space points in the region of R^3 visible to three positions of a synthetic camera: $\mathbf{P} = \mathbf{C}[\mathbf{1}|\mathbf{0}]$ ($\mathbf{C}$ stands for camera intrinsic matrix), $\mathbf{P}' = \mathbf{C}[\mathbf{R}'|\mathbf{t}']$ and $\mathbf{P}'' = \mathbf{C}[\mathbf{R}''|\mathbf{t}'']$, where the camera makes rotations $\mathbf{R}'$, $\mathbf{R}''$ and translations $\mathbf{t}'$, $\mathbf{t}''$. Here the total number of space points is 300. As shown in Table 1, we select the first ten space points and their projective correspondences.

The experiments can be divided into two parts:

<1> Six different groups of Gaussian noise are added to the projective correspondences, whose means are 0 and variances vary from 0.5 to 3.0 (at 0.5 step).

No.	$\mathbf{X}$	$\mathbf{p}$	$\mathbf{p}'$	$\mathbf{p}''$
1	(5.47, 1.78, 4.80)	(239.4, 40.8)	(509.7, 39.7)	(316.4, 240.3)
2	(1.72, 7.82, 9.08)	(39.8, 94.7)	(150.9, 18.0)	(54.8, 74.1)
3	(6.01, 4.33, 6.43)	(196.2, 74.1)	(355.6, 43.0)	(173.1, 175.4)
4	(0.25, 7.40, 9.82)	(5.4, 82.9)	(128.0, 6.2)	(55.9, 56.2)
5	(9.30, 8.23, 4.02)	(486.2, 225.3)	(418.4, 140.1)	(75.1, 297.8)
6	(4.93, 6.28, 9.43)	(109.8, 73.2)	(243.8, 19.6)	(114.8, 110.7)
7	(4.60, 8.34, 6.79)	(142.5, 135.3)	(210.3, 52.5)	(48.4, 126.1)
8	(3.63, 7.12, 6.44)	(118.5, 121.7)	(201.2, 43.9)	(58.2, 116.0)
9	(6.31, 7.90, 4.43)	(299.0, 196.0)	(282.9, 93.9)	(45.1, 196.3)
10	(2.41, 3.30, 5.82)	(86.9, 62.3)	(228.8, 13.5)	(135.7, 104.5)

Table 1. The first ten space points and their projective correspondences

<2> The means and variances of Gaussian noise are fixed to 0, 1, respectively; the percentage of outliers disturbed by the bad locations and false matches varies from 10 % to 60 % (at 10% step).

In order to compare the quality of the re-estimated trifocal tensor, the following formula is used to compute the residuals of inliers only

$$E = \frac{1}{N_{in}} \sum_{i=1}^{N_{in}} \left((x-\hat{x})^2 + (y-\hat{y})^2 + (x'-\hat{x}')^2 + (y'-\hat{y}')^2 + (x''-\hat{x}'')^2 + (y''-\hat{y}'')^2 \right)^{\frac{1}{2}} \tag{31}$$

where N_{in} is the number of all the inliers.

Tables 2 and 3 show the residuals under various Gaussian noises and percentages of outliers, respectively. They are illustrated in Figures 1 and 2 as well, the curves at the bottom being the results of GMM. It can be seen in Table 2 that our method can gain better results compared with other three algorithms, especially when the variances of noise are more than 2.0. In Table 3, as the percentage of outliers increases, residual of GMM keeps increasing smoothly while at the same time it remains the smallest of those derived from the four methods. In Experiment <1>, when the variance is 0.5, the residuals of LA, M-estimator and MLESAC are 1.366, 1.102, 0.933 times, respectively, as much as that of GMM; when the variance increases to 3.0, they are 2.310, 1.409, 1.099 times, respectively. In Experiment <2>, when the percentage of outliers equals to 10 %, the residuals of LA, M-estimator and MLESAC are 1.780, 1.195, 0.967 times, respectively, as much as that of GMM; when the variance increases to 60 %, they are 3.238, 1.832, 1.063 times, respectively.

4.2 Experiments with Real Images

Three different images of the same scene are employed to compare the four methods. First we discard the correspondences, which are destroyed by false matches or bad locations. Then the percentage of outliers increases from 10% to 60% (at 10 % step) by disturbing correspondences using false matches and bad locations.

variance	0.5	1.0	1.5	2.0	2.5	3
LA	1.512	3.963	4.549	6.978	8.016	10.374
M-Estimator	1.307	2.044	2.729	3.576	5.436	6.328
MLESAC	1.117	1.434	1.954	2.770	4.511	4.934
GMM	1.186	1.423	1.847	2.531	4.168	4.490

Table 2. Residuals under various Gaussian noises

outlier percentage	10%	20%	30%	40%	50%	60%
LA	6.103	10.796	12.107	14.225	16.784	24.360
M-Estimator	4.099	7.119	9.260	10.928	12.464	13.785
MLESAC	3.317	5.164	5.737	5.992	7.047	7.996
GMM	3.429	5.015	5.458	5.516	6.870	7.523

Table 3. Residuals under various percentages of outliers disturbed by noise and false matches

Table 4 shows the results of the experiment. When the percentage of outliers is 10%, the residuals of LA, M-estimator and MLESAC are 2.519, 1.764, 0.984 times, respectively, as much as that of GMM; when the variance increases to 60%, they are 2.976, 2.494, 1.025 times, respectively. Figure 6 also shows the results of the four algorithms under 30% percentage of outliers (the white crosses stand for the measured correspondences; the white blocks stand for the estimated correspondences; the white line segments stand for the residuals between measured and estimated correspondences). From the tables and images, we can see that our method is robust

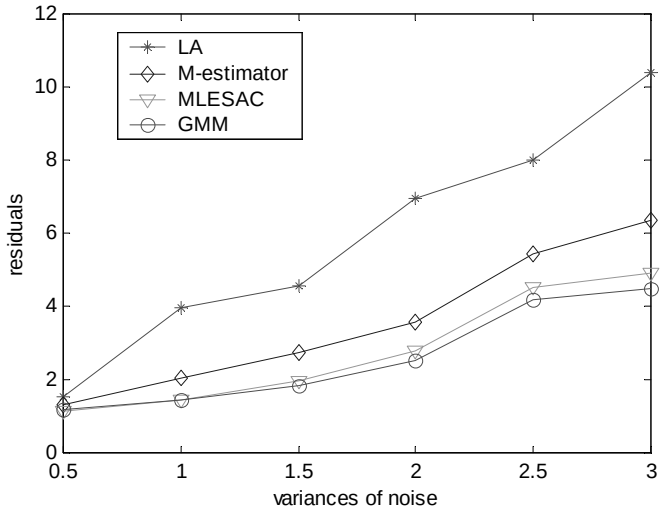

Fig. 1. The results of adding Gaussian noise to the correspondences, whose means are 0 and variances vary from 0.5 to 3.0 (at 0.5 step)

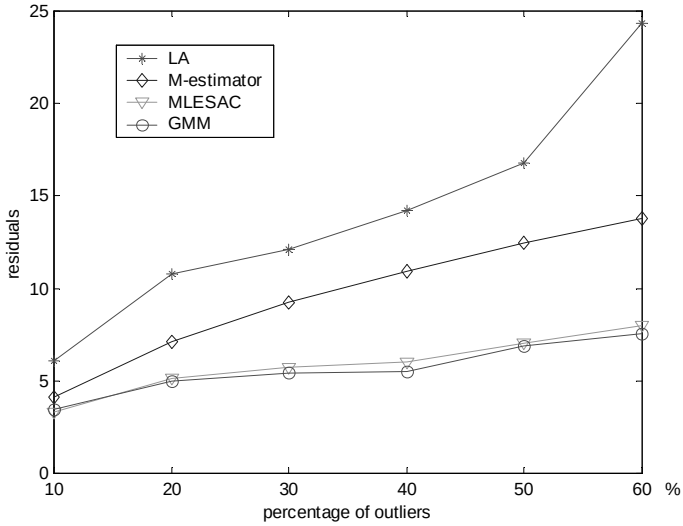

Fig. 2. The results under various percentages of outliers, varying from 10% to 70% (at 10 % step)

to noise and outliers; even when the percentage of outlier is great, the residuals of GMM are less than those of others all the time.

From the above experiments, we can notice that the M-estimator depend too much on the value of residuals and make too little use of the underlying information it contains, i.e. multivariate Gaussian distribution, while that is just what we exploit to construct Gaussian mixture model as the principle of classification.

outlier percentage	10%	20%	30%	40%	50%	60%
LA	11.816	14.307	16.844	23.927	26.787	32.405
M-Estimator	8.275	10.261	11.138	19.904	22.523	27.153
MLESAC	4.614	5.176	5.649	7.635	9.995	11.163
GMM	4.691	4.982	5.435	7.398	9.512	10.889

Table 4. Residuals under various percentages of outliers disturbed by noise and false matches

5 CONCLUSION

In this paper, we propose a new robust estimation method for the multiple-view geometry employing Gaussian mixture model. Unlike in other robust methods, outliers are found only according to the great residual value obtained from the trifocal tensor computation. We derive parameters from the Gaussian mixture model constructed by the residuals. Then according to the parameters we get the posterior

Fig. 3. The results of the four algorithms under 30% percentage of outliers; (a) the image of the first view; (b) the image of the second view; (c) the result of LA in the third view; (d) the result of M-estimator in the third view; (e) the result of MLESAC in the third view; (f) the result of GMM in the third view

probability, based on which classification is made. Experiments with both synthetic data and real images show that our method is more robust and precise than other typical methods (LA, M-estimator) because it can efficiently detect and delete bad corresponding points, which include both bad locations and false matches.

Acknowledgments

Thanks to the reviewers for helpful suggestions and to Robotics Research Group of Oxford University for supplying real image data.

REFERENCES

[1] Torr, P. H. S.: Outlier Detection and Motion Segmentation. PhD thesis, University of Oxford, Engineering. Dept., 1995.

[2] Armstrong, M.—Zisserman, A.—Hartley, R.: Self-Calibration from Image Triplets. In Bernard Buxtion, ed., Proceedings of the 4th European Conference on Computer Vision, Cambridge, UK, April 1996.

[3] Avidan, S.—Shashua, A.: Novel View Synthesis by Cascading Trilinear Tensors. IEEE Transactions on Visualization and Computer Graphics (TVCG), Vol. 4, 1998, No. 4.

[4] Spetsakis, M.—Alomonos, J.: Structure from Motion Using Line Correspondences. International Journal of Computer Vision, Vol. 4, 1990, No. 3, pp. 171–183.

[5] Weng, J.—Ahuja, N.—Huang, T.: Optimal Motion and Structure Estimation. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 15, 1993, No. 9, pp. 864–884.

[6] Shashua, A.: Algebraic Functions for Recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 17, 1995, No. 8, pp. 779–789.

[7] Hartley, R. I.: Projective Reconstruction from Line Correspondences. In Proc. IEEE Conference on Computer Vision and Pattern Recognition, 1994.

[8] Faugeras, O. D.—Papadopoulo, T.: A Nonlinear Method for Estimating the Projective Geometry of Three Views. In Proceedings of the 6th International Conference on Computer Vision, Bombay, India, pp. 477–484, IEEE Computer Society Press, 1998.

[9] Torr, P. H. S.—Zisserman, A.: Robust Parameterization and Computation of the Trifocal Tensor. Image and Vision Computing, Vol. 15, 1997, pp. 591–605.

[10] Stein, G. P.—Shashua, A.: On Degeneracy of Linear Reconstruction from Three Views: Linear Line Complex and Applications. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 21, 1999, No. 3, pp. 244–251.

[11] Quan, L.: Invariants of 6Points from 3 Uncalibrated Images. In J. O. Eckland, editor, Proc. 3rd European Conference on Computer Vision, LNCS 800/801, Stockholm, pp. 459–469, 1994.

[12] Maronna, R. A.: Robust M-Estimator of Multivariate Location and Scatter. Annals of Statistics 4, pp. 51–67, 1976.

[13] Fischler, M. A.—Bolles, R. C.: Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. Comm. Assoc. Comp. Mach., Vol. 24, 1981, No. 6, pp. 381–395.

[14] Torr, P. H. S.—Zisserman, Z.: MLESAC: A New Robust Estimator with Application to Estimating Image Geometry. Computer Vision and Image Understanding, 78, pp. 138–156, 2000.

[15] Jain, A. K.—Duin, R. P. W.—Mao, J. C.: Statistical Pattern Recognition: A Review. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 22, 2000, No. 1, pp. 4–37.

[16] Reynolds, D. A.—Rose, R. C.: Robust Text-Independent Speaker Identification Using Gaussian Mixture Speaker Models. IEEE Transactions on Speech and Audio Processing, Vol. 3, 1995, No. 1.

Mingxing Hu was born in 1975. He received the B.S. degree in Jilin University, in 1998. He is currently pursuing the Ph.D. degree in signal and information processing at Institute of Information Science, Northern Jiaotong University. His current research interests include 3D reconstruction, computer vision, and virtual reality.

Qiang Xing was born in 1975. He received the B.S. degree in Petroleum University, in 1997. He is currently pursuing the Ph.D. degree in signal and information processing at Institute of Information Science, Northern Jiaotong University. His current research interests include image processing and image retrieval.

Baozong Yuan was born in 1932. He received the Ph.D. degree in electrical engineering from Leningrad Institute of Railway Engineering, USSR, in 1960. He has joined the Northern Jiaotong University in 1953. He was a visiting professor at the University of Pittsburgh, USA and the University of Wales, UK in 1982, 1983, and 1988, respectively. Dr. Yuan is Chairman of Computer Chapter of IEEE Beijing Section, Fellow of British Royal Society, IEE Fellow, Vice Chairman of IEE Beijing Center Development. His research interests include computer vision, virtual reality, image processing, computer graphics, speech signal processing, and multimedia information processing and data communication.

Xiaofang Tang is an engineer of Northern Jiaotong University. She is a member of IEEE. Her research interests include computer vision, virtual reality, multimedia information processing and data communication, and speech signal processing.