第 29 卷 第 2 期
2003 年 3 月

自 动 化 学 报
ACTA AUTOMATICA SINICA

Vol. 29, No. 2
Mar., 2003

5-point and 4-point Algorithm to Determine of the Fundamental Matrix[1)]

WU Fu-Chao HU Zhan-Yi
(*National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing* 100080)
(E-mail: {fcwu, huzy}@nlpr.ia.ac.cn)

Abstract The fundamental matrix encapsulates all the information between two images, and plays a very important role in camera calibration and 3D reconstruction. In this paper, the following conclusions have been rigorously proved: If the camera motion is of a pure translation, then given 5 point correspondences across two images, the fundamental matrix can be linearly determined if four correspondences of the 5 ones are from coplanar space points (called coplanar correspondences). In addition, we show that if the distortion factor in the pinhole camera model is null, then the fundamental matrix can be linearly determined by only these 4 coplanar correspondences. To our knowledge, such results have not been reported yet in the literature.

Key words Fundamental matrix, homography, camera intrinsic parameters

1 Introduction

The epipolar geometry is the fundamental constraint between two images. It is independent of scene structure, and depends only on the camera internal parameters and their relative pose. The fundamental matrix encapsulates this epipolar geometry, and plays a very important role in 3D computer vision[1,2,3], such as camera calibration, 3D reconstruction. The fundamental matrix is a 3×3 matrix of rank 2, and has 7 degrees of freedom[4]. When the scene structure and camera's motion are unknown, the fundamental matrix can be non-linearly determined using at least 7 point correspondences[4]. In general at least 8 point correspondences are required to linearly determine the fundamental matrix[5]. When there exist 4 correspondences induced from coplanar 3D points (hereinafter, such a correspondence is called a coplanar one), only 6 point correspondences are required to linearly determine the fundamental matrix[6]. If there exist 4 coplanar correspondences, and if additional knowledge on camera motion is available, then can the number of point correspondences be further reduced? Our answer is affirmative. In this paper, the following conclusions will be proved: If the camera motion is of a pure translation, then given 5 point correspondences across two images, the fundamental matrix can be linearly determined if 4 of them are coplanar. In addition, we will show that if the distortion factor in the pinhole camera model is null, then the fundamental matrix can be linearly determined by only these 4 coplanar correspondences.

The paper is organized as follows. In Section 2, some preliminaries on the fundamental matrix and homography are briefed. In Sections 3 and 4, 5-point and 4-piont algorithm for the fundamental matrix computation are elaborated. Some conclusions are given in Section 5.

2 The fundamental matrix and homography

The fundamental matrix F

Let (R,t) be the camera's motion, that is, the transformation from the first camera

1) Supported by the National Natural Science Foundation of P. R. China(60075004, 60033010) and Multidisciplinary Research Program of CAS(KJCX1-07)
Received November 26, 2001; in revised form June 20, 2002
收稿日期 2001-11-26; 收修改稿日期 2002-06-20

frame (before motion) to the second (after motion) is $\boldsymbol{x}^{(2)} = R\boldsymbol{x}^{(1)} + \boldsymbol{t}$, where R is a rotation matrix and $\boldsymbol{t}$ a translation vector. Suppose that the internal parameters of the two cameras are respectively:

$$K^{(j)} = \begin{pmatrix} f_u^{(j)} & s^{(j)} & u_0^{(j)} \\ 0 & f_v^{(j)} & v_0^{(j)} \\ 0 & 0 & 1 \end{pmatrix}, \quad j = 1,2 \tag{1}$$

The fundamental matrix F corresponding to the two images $(\boldsymbol{I}^{(1)}, \boldsymbol{I}^{(2)})$ can be expressed as[6]:

$$F = (K^{(2)})^{-\mathrm{T}}[\boldsymbol{t}]_\times R(K^{(1)})^{-1} \tag{2}$$

where $[\boldsymbol{t}]_\times = \begin{pmatrix} 0 & -t_3 & t_2 \\ t_3 & 0 & -t_1 \\ -t_2 & t_1 & 0 \end{pmatrix}$ is the skew-symmetric matrix induced by the translation vector $\boldsymbol{t} = (t_1, t_2, t_3)^\mathrm{T}$.

If $\{\tilde{\boldsymbol{u}}_j^{(1)} = (u_j^{(1)}, v_j^{(1)}, 1)^\mathrm{T} \leftrightarrow \tilde{\boldsymbol{u}}_j^{(2)} = (u_j^{(2)}, v_j^{(2)}, 1)^\mathrm{T}\}$ is a set of point correspondences, then the following matrix equality holds:

$$F\tilde{\boldsymbol{u}}_j^{(1)} = \lambda \boldsymbol{e} \times \tilde{\boldsymbol{u}}_j^{(2)} \text{ or equivalently, } (\tilde{\boldsymbol{u}}_j^{(2)})^\mathrm{T} F\tilde{\boldsymbol{u}}_j^{(1)} = 0 \tag{3}$$

where λ is a non-zero scale factor, and $\boldsymbol{e}$ is the epipole in the second image.

The fundamental matrix is unique up to a non-zero scale factor. In general, given at least 8 point correspondences, it is possible to determine the fundamental matrix by solving a set of linear equations as shown in (3).

The homography H

The homography H of the two images $(\boldsymbol{I}^{(1)}, \boldsymbol{I}^{(2)})$ from a same space plane π can be expressed as[6]:

$$H \approx K^{(2)} R (K^{(1)})^{-1} + \frac{1}{d} K^{(2)} \boldsymbol{t}\boldsymbol{n}^\mathrm{T} (K^{(1)})^{-1} \tag{4}$$

where $\boldsymbol{n}$ is the normal of plane π under the first camera frame, and d is the distance of the plane π to the origin.

The homography H is uniquely defined up to a non-zero scale factor. If $\{\tilde{\boldsymbol{u}}_j^{(1)} \leftrightarrow \tilde{\boldsymbol{u}}_j^{(2)}\}$ is a set of coplanar correspondences, then the following equality holds:

$$\lambda \tilde{\boldsymbol{u}}_j^{(2)} = H\tilde{\boldsymbol{u}}_j^{(1)} \tag{5}$$

Thus, given a point correspondence, we can obtain two linear constraints on H. Hence, in general, at least 4 point correspondences are required to linearly compute H.

The relation between F and H

Let F be the fundamental matrix corresponding to the two images $(I^{(1)}, I^{(2)})$, and $\boldsymbol{e}$ be the epipole in the second image. Suppose that H is the homography of a plane π. Then the following relation exists among F, H and $\boldsymbol{e}$ [6]:

$$F \approx [\boldsymbol{e}]_\times H \tag{6}$$

Given matrix H, epipole $\boldsymbol{e}$ can be linearly determined using at least two point correspondences (corresponding 3D points do not belong to the plane π) by (3) and (6). Hence, at least 6 point correspondences are required to linearly determine F when there exists a plane in the 3D scene.

In this paper, suppose the camera motion is of a pure translation. Our objectives are: (1) how to linearly determine the fundamental matrix from 5 known point correspondences, of which 4 are coplanar; (2) how to linearly determine the fundamental matrix from only these 4 coplanar correspondences if the distortion factor in the pinhole camera model equals zero.

3 Fundamental matrix determination with non-zero distortion factor

In this section, the distortion factor in the pinhole camera model is assumed to be nonzero. Let $t = (t_1, t_2, t_3)^T$ be the translation vector of camera. Given 5 point correspondences, denoted by $\tilde{u}_j^{(1)} = (u_j^{(1)}, v_j^{(1)}, 1)^T \leftrightarrow \tilde{u}_j^{(2)} = (u_j^{(2)}, v_j^{(2)}, 1)^T (j = 1, 2, \cdots, 5)$, we suppose the point correspondences $\tilde{u}_j^{(1)} \leftrightarrow \tilde{u}_j^{(2)} (j = 1, 2, 3, 4)$ are from plane π, and that n is the normal vector of plane π under the first camera frame, d the distance to the origin. If the homography H of plane π has been determined by these 4 coplanar correspondences, then in order to compute the fundamental matrix F by (6), we need only to compute the epipole e. In the following, we will show how to determine e under two different circumstances.

(I) The translation vector is not parallel to the image plane

Proposition 1. Suppose plane π is not parallel to the image plane; then the epipole e can be linearly determined if and only if $v_5^{(2)} \neq \hat{v}_5^{(2)}$, where $(\hat{u}_5^{(2)}, \hat{v}_5^{(2)}, 1)^T \approx H\tilde{u}_5^{(1)}$.

Proof. Since the translation vector t is not parallel to the image plane, $t_3 \neq 0$. By denoting $e = (e_1, e_2, 1)^T = \dfrac{K^{(2)} t}{t_3}$, $a^T = (a_1, a_2, a_3) = \dfrac{t_3}{sd} n^T (K^{(1)})^{-1}$, and

$$sK^{(2)} (K^{(1)})^{-1} = \begin{bmatrix} k_1 & k_2 & k_3 \\ 0 & k_4 & k_5 \\ 0 & 0 & k_6 \end{bmatrix} \tag{7}$$

the matrix H can be written in the form

$$H = sK^{(2)} (K^{(1)})^{-1} + ea^T = \begin{bmatrix} k_1 + e_1 a_1 & k_2 + e_1 a_2 & k_3 + e_1 a_3 \\ e_2 a_1 & k_4 + e_2 a_2 & k_5 + e_2 a_3 \\ a_1 & a_2 & k_6 + a_3 \end{bmatrix} \tag{8}$$

Now note that $n_1 \neq 0$ due to the fact that the plane is not parallel to the x axis of the camera frame, and

$$(K^{(1)})^{-1} = \begin{bmatrix} 1/f_u^{(1)} & -s^{(1)}/(f_u^{(1)} f_v^{(1)}) & (-u_0^{(1)} f_v^{(1)} + s^{(1)} v_0^{(1)})/(f_u^{(1)} f_v^{(1)}) \\ 0 & 1/f_v^{(1)} & -v_0^{(1)}/f_v^{(1)} \\ 0 & 0 & 1 \end{bmatrix}$$

So we have $a_1 = \dfrac{n_1}{sdf_u^{(1)}} \neq 0$. Therefore equation (8) implies

$$e_2 = \frac{H_{21}}{H_{31}} \tag{9}$$

where H_{ij} is the $(i, j)^{th}$ element of matrix H.

Since $F = [e]_\times H$ is the fundamental matrix and $\{\tilde{u}_5^{(1)} \leftrightarrow \tilde{u}_5^{(2)}\}$ is a point correspondence, we have

$$(\tilde{u}_5^{(2)})^T [e]_\times H\tilde{u}_5^{(1)} = 0 \tag{10}$$

Substituting $(\hat{u}_5^{(2)}, \hat{v}_5^{(2)}, 1)^T \approx H\tilde{u}_5^{(1)}$ into (10) leads to

$$(\hat{v}_5^{(2)} - v_5^{(2)}) e_1 = (u_5^{(2)} \hat{v}_5^{(2)} - v_5^{(2)} \hat{u}_5^{(2)}) + e_2 (u_5^{(2)} - \hat{u}_5^{(2)}) \tag{11}$$

Equation (11) has a unique solution on e_1 if and only if $v_5^{(2)} \neq \hat{v}_5^{(2)}$, i. e.,

$$e_1 = \frac{(u_5^{(2)} \hat{v}_5^{(2)} - v_5^{(2)} \hat{u}_5^{(2)}) + e_2 (u_5^{(2)} - \hat{u}_5^{(2)})}{(\hat{v}_5^{(2)} - v_5^{(2)})} = \frac{u_5^{(2)} \hat{v}_5^{(2)} - v_5^{(2)} \hat{u}_5^{(2)} - \dfrac{H_{21}}{H_{31}} (\hat{u}_5^{(2)} - u_5^{(2)})}{\hat{v}_5^{(2)} - v_5^{(2)}} \tag{12}$$

Thus, by (9) and (12) we have

$$e = \left(\frac{u_5^{(2)} \hat{v}_5^{(2)} - v_5^{(2)} \hat{u}_5^{(2)} - \dfrac{H_{21}}{H_{31}} (\hat{u}_5^{(2)} - u_5^{(2)})}{\hat{v}_5^{(2)} - v_5^{(2)}}, \frac{H_{21}}{H_{31}}, 1 \right)^T \tag{13}$$

□

Remark 1. When the plane π is parallel to the x axis of the camera frame, we have $a_1 = \dfrac{n_1}{sdf_u^{(1)}} = 0$. Hence, it is impossible to obtain any linear constrains on the epipole e

from the matrix H. As a result, the fundamental matrix F can not be determined using 5 point correspondences in this case.

Proposition 2. If the plane π is not parallel to the x axis of the camera frame and $v_5^{(2)} \neq 0$, then the epipole $\boldsymbol{e}$ can be linearly determined.

Proof. We need only to prove that equation (11) has a unique solution. Otherwise we have

$$\hat{v}_5^{(2)} - v_5^{(2)} = 0 \tag{14}$$

and

$$(u_5^{(2)}\hat{v}_5^{(2)} - v_5^{(2)}\hat{u}_5^{(2)}) + e_2(u_5^{(2)} - \hat{u}_5^{(2)}) = 0 \tag{15}$$

Let $\boldsymbol{l}$ be the line through the first camera center and the 3D point $\tilde{\boldsymbol{x}}_5$ corresponding to images $(\tilde{\boldsymbol{u}}_5^{(1)} \leftrightarrow \tilde{\boldsymbol{u}}_5^{(2)})$, and $\hat{\boldsymbol{x}}_5$ be the intersection point of the line $\boldsymbol{l}$ with plane π. Then from Fig. 1, we know that the projections of point $\hat{\boldsymbol{x}}_5$ on the two image planes are $\tilde{\boldsymbol{u}}_5^{(1)}$ and $\hat{\boldsymbol{u}}_5^{(2)}$, respectively. It follows that $\boldsymbol{l}_{\tilde{\boldsymbol{u}}_5^{(1)}}^{(2)} = \hat{\boldsymbol{u}}_5^{(2)} \times \tilde{\boldsymbol{u}}_5^{(2)}$ is the epipolar line in the second image corresponding to $\tilde{\boldsymbol{u}}_5^{(1)}$.

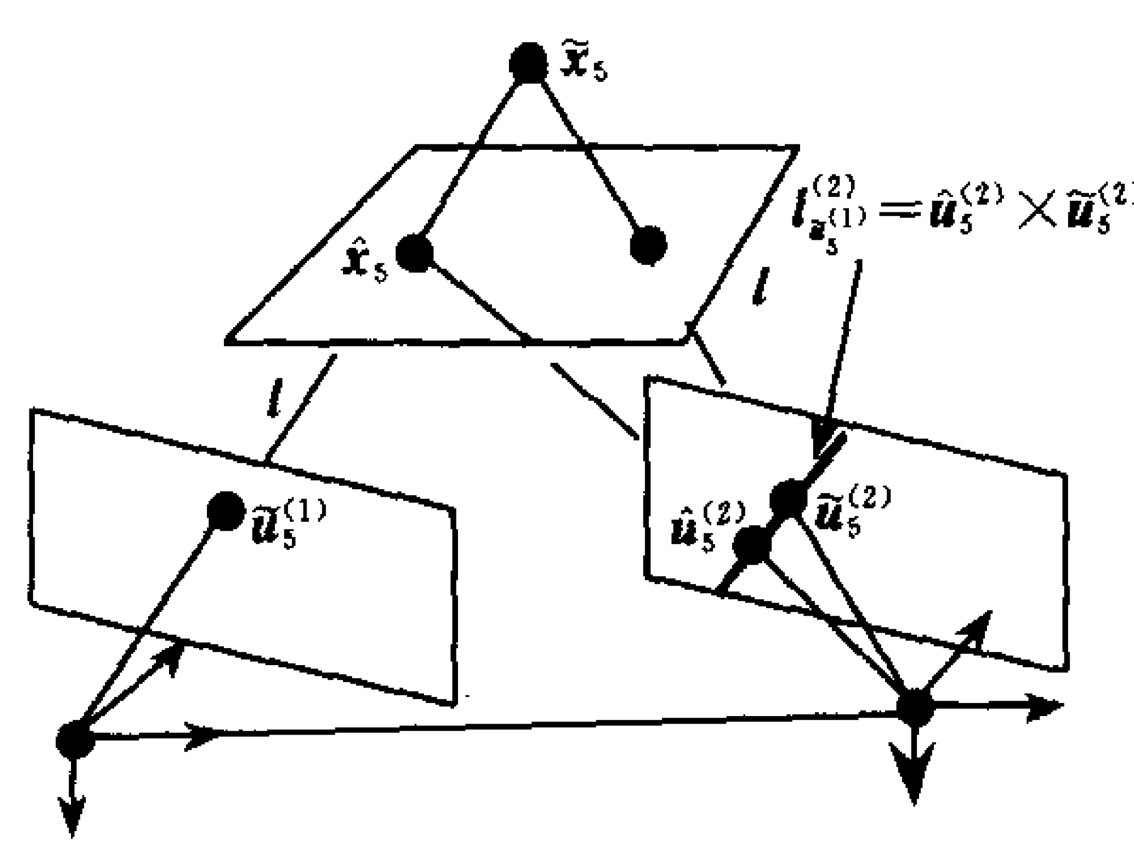

Fig. 1 $\boldsymbol{l}_{\tilde{\boldsymbol{u}}^{(1)}}^{(2)} = \hat{\boldsymbol{u}}_5 \times \tilde{\boldsymbol{u}}_5$ is the epipolar line in the second image corresponding to $\tilde{\boldsymbol{u}}_5^{(1)}$

According to (14), we have $\boldsymbol{l}_{\tilde{\boldsymbol{u}}_5^{(1)}}^{(2)} = (0,1,-v_5^{(2)})(=(0,1,-\hat{v}_5^{(2)}))$. This implies $e_2 = \hat{v}_5^{(2)} = v_5^{(2)} \neq 0$. Thus, equation (15) can be simplified into the following form:

$$v_5^2(\hat{u}_5^{(2)} - u_5^{(2)}) = 0 \tag{16}$$

So, $\hat{u}_5^{(2)} - u_5^{(2)} = 0$, and $\hat{\boldsymbol{u}}_5^{(2)} = \tilde{\boldsymbol{u}}_5^{(2)} \approx H\tilde{\boldsymbol{u}}_5^{(1)}$. From (14), it implies that $\tilde{\boldsymbol{x}}_5$ belongs to the plane π. That is contradictory to our basic assumption. □

(II) The translation vector is parallel to the image plane

Proposition 3. From the homography H and point correspondences $(\tilde{\boldsymbol{u}}_5^{(1)} \leftrightarrow \tilde{\boldsymbol{u}}_5^{(2)})$, we can linearly determine the epipole $\boldsymbol{e}$.

Proof. Since the translation vector $\boldsymbol{t}$ is parallel to the image plane, $t_3 = 0$ and $\boldsymbol{e}$ is at the infinity. Let $\boldsymbol{e} = (e_1, e_2, 0)^{\mathrm{T}}$. Then the equation

$$(\tilde{\boldsymbol{u}}_5^{(2)})^{\mathrm{T}}[\boldsymbol{e}]_\times H\tilde{\boldsymbol{u}}_5^{(1)} = 0 \tag{17}$$

becomes

$$(\hat{v}_5^{(2)} - v_5^{(2)})e_1 = e_2(u_5^{(2)} - \hat{u}_5^{(2)}) \tag{18}$$

Note that the two entities $\hat{v}_5^{(2)} - v_5^{(2)}$, $u_5^{(2)} - \hat{u}_5^{(2)}$ are not all zero since 3D point $\tilde{\boldsymbol{x}}_5$ corresponding to images $(\tilde{\boldsymbol{u}}_5^{(1)} \leftrightarrow \tilde{\boldsymbol{u}}_5^{(2)})$ does not belong to the plane π. Hence, we have

$$\boldsymbol{e} = e_1\left(1, \frac{u_5^{(2)} - \hat{u}_5^{(2)}}{\hat{v}_5^{(2)} - v_5^{(2)}}, 0\right)^{\mathrm{T}} \approx \left(1, \frac{u_5^{(2)} - \hat{u}_5^{(2)}}{\hat{v}_5^{(2)} - v_5^{(2)}}, 0\right)^{\mathrm{T}} \text{ when } \hat{v}_5^{(2)} - v_5^{(2)} \neq 0;$$

$$\boldsymbol{e} = e_2\left(\frac{v_5^{(2)} - \hat{v}_5^{(2)}}{u_5^{(2)} - \hat{u}_5^{(2)}}, 1, 0\right)^{\mathrm{T}} \approx \left(\frac{v_5^{(2)} - \hat{v}_5^{(2)}}{u_5^{(2)} - \hat{u}_5^{(2)}}, 1, 0\right)^{\mathrm{T}} \text{ when } u_5^{(2)} - \hat{u}_5^{(2)} \neq 0. \quad □$$

4 Fundamental matrix determination with zero distortion factor

Suppose that the distortion factor in the pinhole model equals zero, and the translation vector $\boldsymbol{t}$ is not parallel to the image plane. Then we have the following propostion.

Proposition 4. If neither the x-axis nor the y-axis of the camera frame is parallel to the plane π, then the epipole $\boldsymbol{e}$ can be linearly determined from the homography H of plane π.

Proof.

$$K^{(2)} = \begin{bmatrix} f_u^{(2)} & 0 & u_0^{(2)} \\ 0 & f_v^{(2)} & v_0^{(2)} \\ 0 & 0 & 1 \end{bmatrix}, \quad (K^{(1)})^{-1} = \begin{bmatrix} 1/f_u^{(1)} & 0 & -u_0^{(1)} f_v^{(1)}/(f_u^{(1)} f_v^{(1)}) \\ 0 & 1/f_v^{(1)} & -v_0^{(1)}/f_v^{(1)} \\ 0 & 0 & 1 \end{bmatrix}$$

Since $s^{(1)} = s^{(2)} = 0$, the matrix H may be written in the form

$$H = sK^{(2)}(K^{(1)})^{-1} + \boldsymbol{e}\boldsymbol{a}^{\mathrm{T}} = \begin{bmatrix} k_1 + e_1 a_1 & e_1 a_2 & k_3 + e_1 a_3 \\ e_2 a_1 & k_4 + e_2 a_2 & k_5 + e_2 a_3 \\ a_1 & a_2 & k_6 + a_3 \end{bmatrix} \tag{19}$$

In addition, since neither the x-axis nor the y-axis of the camera frame is parallel to the plane π, we have $n_1 \neq 0, n_2 \neq 0$, and

$$a_1 = \frac{n_1}{sdf_u^{(1)}} \neq 0, a_1 = \frac{n_2}{sdf_2^{(1)}} \neq 0 \tag{20}$$

Then according to (19) we have $e_1 = \dfrac{H_{12}}{H_{32}}, e_2 = \dfrac{H_{21}}{H_{31}}$, and $\boldsymbol{e} = \left(\dfrac{H_{12}}{H_{32}}, \dfrac{H_{21}}{H_{31}}, 1\right)^{\mathrm{T}}$. □

Remark 2. If the plane π is parallel to only one of the two axes of the camera frame, one additional point correspondence (corresponding 3D point does not belong to the plane π) is needed to linearly determine epipole $\boldsymbol{e}$.

Remark 3. If the plane π is parallel to the image plane, then two additional point correspondences (both two corresponding 3D points do not belong to the plane π) are needed to linearly determine the epipole $\boldsymbol{e}$.

5 Conclusions

In this paper, we show that if some information about camera motion is available, the minimum number of point correspondences can be reduced for determination of the fundamental matrix. In particular, the following conclusions have been rigorously proved: If the camera motion is of a pure translation, then given 5 point correspondences across two images, the fundamental matrix can be linearly determined if 4 point correspondences among the 5 ones are coplanar. In addition, we show that if the distortion factor in the pinhole camera model is null, then the fundamental matrix can be linearly determined by only these 4 coplanar correspondences. To our knowledge, our algorithms are the most economical ones which use only 4 or 5 point correspondences rather than 6, 7, or 8 as reported in the literature.

References

1 Koch R, Pollefeys M, Van Gool L. Multi viewpoint stereo from uncalibrated video sequences. In: Proceedings European Conference on Computer Vision, Freiburg, Germany, 1998, I: 55~71

2 Faugeras O, Rorbert L, Laveaw S. 3-D reconstruction of urban scenes from image sequences. *Computer Vision and Image Understanding*, 1998, **69**(3): 392~309

3 Hartley R I. Kruppa's equations derived from the fundamental matrix. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1997, **19**(2): 100—102

4 Luong Q T, Faugeras O D. The fundamental matrix: Theory, algorithms, and stability analysis. *International Journal of Computer Vision*, 1996, **17**(2): 43~75

5 Hartley R I. In defense of the 8-point algorithm. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1997, **19**(6): 580~593

6 Hartley R I. Multi Viewpoint Geometry in Computer Vision. Cambridge, UK: Cambridge University Press, 2000

WU Fu-Chao Received his bachelor degree in Mathematics from Anqing Normal College in 1982. From 1984 to 1994, he acted first as a lecturer then an associate professor in Anqing Normal College. From 1995 to 1999, he acted as a professor in Anhui University. Since 2000, he has been with the Institute of Automation, Chinese Academy of Sciences, where he is now a professor. His research interests are now in Computer Vision, which include 3D Reconstruction, Active Vision, and Image Based Modeling and Rendering. He has received the Second Best Award of Natural Science of Anhui province in 1999.

HU Zhan-Yi Received his bachelor degree in Automation from the North China University of Technology in 1985, the Ph. D. Degree (Docteur d'Etat) in Computer Vision from the University of Liege, Belgium, in Jan. 1993. Since 1993, he has been with the Institute of Automation, Chinese Academy of Sciences, where he is now a professor. From May 1997 to May 1998, he also acted as a visiting scholar of Chinese University of Hong Kong on invitation. His research interests are in Robot Vision, which include Camera Calibration and 3D Reconstruction, Active Vision, Geometric Primitive Extraction, Vision Guided Robot Navigation, Image Based Modeling and Rendering. He has received the Second Best Award of Natural Science of CAS in 1996.

基本矩阵的 5 点和 4 点算法

吴福朝　　胡占义

（中国科学院自动化研究所模式识别国家重点实验室　北京　100080）

（E-mail：{fcwu，Huzy}@nlpr. ia. ac. cn）

摘　要　基本矩阵(Fundamental Matrix)是两幅图像之间的基本约束，在摄像机标定和三维重建中起着至关重要的作用. 本文证明，当摄像机在两幅图像之间的运动为纯平移运动时，给定 5 对图像对应点，如果其中的 4 对对应点为共面空间点的投影(称为共面对应点)，则可以线性确定基本矩阵. 另外，如果摄像机不是 5 参数模型(完全针孔模型)，而是 4 参数模型(畸变因子为零)，则此时仅使用该 4 对共面对应点即可线性确定基本矩阵. 据我们所知，这些结果在文献中还没有类似的报导.

关键词　基本矩阵，单应矩阵，摄像机，摄像机内参数

中图分类号　TP18