On the Epipolar Geometry in Stereo Vision

Ebroul Izquierdo M.
Heinrich-Hertz-Institute for Communication Technology (HHI)
Image Processing Department
Einsteinufer 37
D-10587 Berlin, Germany
e-mail: ebroul@hhi.de
Tel.: (30)(030)31002619

Abstract

In this paper we are interested in the calculation of the equation governing pairs of epipolar lines for unregistered images, i.e. when the image acquisition parameters are unknown or no calibration object is available. First we do a thorough analysis of the mathematics required to find the epipolar line equations in the classical sense. Thereafter we show that for a broad class of stereo images the epipolar line equation can be obtained directly from the images without restrictive assumptions on the stereo geometry and without anyone 3D reference point. Therefore, the estimation of the epipolar line equations reduce to solve a linear optimization problem. Finally, we give an efficient algorithm to find the epipolar line equations directly from the image data, without knowledge of the image acquisition parameters.

Keywords and phrases: Stereo Imaging, unregitered images, Epipolar line, correspondence problem.

AMS(MOS) subject classifications: 65M55, 68T99, 35A40

1 Introduction

Real world scenes taken from two different perspectives can provide structural information of the world. The process of analyzing different images of the same scene is called stereo vision and it works in a very large variety of applications in the computer vision. For example in the robotics two or more cameras mounted in a Robot take stereo picture pairs, each of these stereo pair is processed to obtain the needed three-dimensional-information and so the Robot can move through an environment with different obstacles and navigate to a desired location.

The main problem in stereo vision is to find corresponding points in images taken from different viewpoints, i.e. to detect the differences occur between images. These differences are called disparity and can be seen as a vector field mapping one stereo image into the other. The determination of this vector field has been called the correspondence problem and its solution is necessary to obtain depth information of a scene [6], [7]. Once correspondences are determined, the three-dimensional position of an object is easily computed by the so-called triangulation technique.

In order to reduce the amount of potential diplacements by the matching process, most matching methods constraint the search along the epipolar lines [1], [5]. This constraint demand that epipolar lines in one stereo image match the corresponding epipolar lines in the other one. The corresponding epipolar lines are defined as intersection of a plane, containing both optic centers and a given point in the real world scene, with the image planes. The simplest case is given when the camera configuration is parallel, i.e., when the image planes are coplanar, the vertical camera coordinate axes are parallel and the distortion coefficient is equal zero. In this case the vertical component of the disparity vectors is zero and the epipolar lines coincide with the scan lines. Several algorithms assume this situation [1], [5], in practice it is very difficult to realize this camera geometry, so that a method to calculate the epipolar line equations will be required. There are several methods to calculate the equations of the epipolar lines, but most require a stereoscopic camera calibration step [2], [4], [6]. Stereoscopic camera calibration is the process of determining the internal camera parameters and optical characteristics, intrinsic parameters, and the position and orientation of both cameras respect a fix defined world coordinate system, extrinsic parameters. After estimation of the camera parameters it is a trivial task to find the relation governing the epipolar lines. This method is suitable for studio applications or in industrial environments that not require real time processing and where the cameras are not moved after calibration.

Often we have to solve the correspondence problem for unregistered stereo sequences [8] i.e., sequences for which the respective camera parameters are unknown, or in case of changes in the camera parameters during the sequence shooting process. In such situation we have to estimate the epipolar lines directly from the stereo images.

In this paper we present a method to estimate the equations of the epipolar lines using only image information and without knowledge of camera parameters. Firstly, we select a few number of points in one stereo image and estimate very accurately the disparities for these image positions. The so obtained information is used in order to find the relation governing the epipolar lines in the whole image.

The paper is organized as follows. In section 2 we introduce the considered stereo camera model. Section 3 gives a formal definition of the epipoles and the epipolar lines. In section 4 we describe the basic concept used to estimate the quations of the epipolar lines without camera calibration as well as a suited algorithm to calculate them.

2 Stereo Image Model

Intrinsic Parameters We assume that the considered cameras can be modeled by pin-hole cameras. The basic camera model is shown in figure 1. In the following we describe in detail the used notation.

(X_W, Y_W, Z_W)	the 3D coordinate of the object point P in the 3D world coordinate system
(X, Y, Z)	the 3D coordinate of the object point P in 3D camera coordinate system, whereby the Z axis coincides with the optical axis
f	distance between image plane and the optical center O
(x, y)	the image coordinate system with the origin at the point (O, O, f) with respect to the 3D camera coordinate system and the image plane is the same as the plane $Z = f$
(u, v)	the coordinate system used in the computer or in the respective digitized image

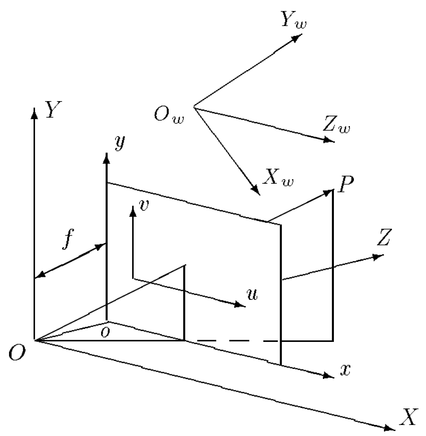

Figure 1. *The Camera Model*

The relationship between the 3D camera coordinate system and the image coordinate system is

$$\frac{f}{Z} = \frac{x}{X} = \frac{y}{Y} \tag{1}$$

Let k_n and k_v be the image scale factors in the direction u, respectively v. The following both equations give the scanning and sampling relationship.

$$x = \frac{u}{k_u}, \qquad y = \frac{v}{k_v}.$$

Using a coordinate translation we can transform the scaled image coordinate (x, y) in the respective computer coordinate

$$\begin{aligned} u &\Rightarrow u - u_0, \\ v &\Rightarrow v - v_0. \end{aligned}$$

Hereby the point (u_0, v_0) coincides with the image center O.

Hence, from the last relations we obtain the equations

$$x = \frac{u - u_0}{k_u}, \qquad y = \frac{v - v_0}{k_v}$$

Inserting the last equations into (1) we get

$$u = \frac{\alpha_u X}{Z} + u_0, \qquad v = \frac{\alpha_v Y}{Z} + v_0, \tag{2}$$

with $\alpha_u = fk_u$ and $\alpha_v = fk_v$.

Note, that in these relations lens distortion has not been considered. In this case the five parameters u_0, v_0, k_u, k_v and f represent the intrinsic camera parameters. For an analysis in which distortion coefficients are considered we refer the reader to [9].

Now we can write (2) as follow

$$ZU = \mathcal{M}\tilde{X} \tag{3}$$

with $U = \begin{pmatrix} u \\ v \\ 1 \end{pmatrix}$, $\mathcal{M} = \begin{pmatrix} \alpha_u & 0 & u_0 \\ 0 & \alpha_v & v_0 \\ 0 & 0 & 1 \end{pmatrix}$ and $\hat{X} = \begin{pmatrix} X \\ Y \\ Z \end{pmatrix}$.

The matrix $\mathcal{M}$ is called perspective transformation matrix [2], [3], [9].

Extrinsic Parameters The parameters for the transformation from 3D object world coordinate system to the camera 3D coordinate system with the optical center as origin are called extrinsic parameters. This transformation is a mapping between the world and the camera coordinate system. In order to define this mapping in a unique way we rotate the camera coordinate system first and then we translate it to the origin of the world coordinate system. This process yields the following mapping

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = R \begin{pmatrix} X_W \\ Y_W \\ Z_W \end{pmatrix} + T. \tag{4}$$

Hereby are R the rotation matrix and T the translation vector.

The orthonormal matrix R is given by the Euler angles yaw, pitch and tilt. The six extrinsic camera parameters consist of these three angles together with the three components of the translation vector T.

Stereo Camera Model In the first part of this section monocular image geometry has been described. Let us now consider a stereo camera arrangement.

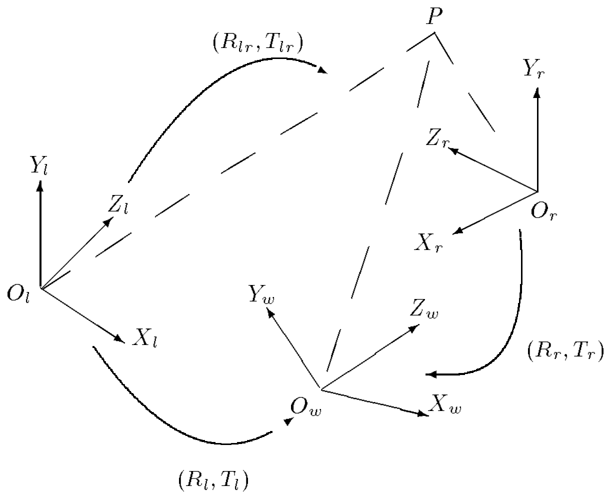

Figure 2. *The Coordinate Systems*

Let $\hat{X}_l = (X_l, Y_l, Z_l)$ and $\hat{X}_r = (X_r, Y_r, Z_r)$ be the left and the right 3D world coordinates of one object point P in the left respectively right 3D camera coordinate systems (see fig. 2).

Applying (4) to $\hat{X}_l$ we obtain

$$\hat{X}_l = R_l \hat{X}_W + T_l, \tag{5}$$

when $\hat{X}_W$ defines the coordinates of P in the 3D world coordinate system and R_l, T_l are the rotation matrix and the translation vector from left camera coordinate system to the world coordinate system.

Analogously, the relation

$$\hat{X}_r = R_r \hat{X}_W + T_r, \tag{6}$$

is valid for the right camera coordinate system.

Inserting (6) into (5) we get

$$\hat{X}_l = R_{lr} \hat{X}_r + T_{lr}. \tag{7}$$

Here $R_{lr} = R_l R_r^T$ is the rotation matrix from left to right camera coordinate systems and $T_{lr} = T_l - R_{lr} T_r$ is the respective translation vector.

Suppose that a point P is observed by two cameras at $z_l = (u_l, v_l)$ and at $z_r = (u_r, v_r)$ on the image planes, respectively. Their relationship is given as follows

$$Z_l U_l = Z_r M U_r + N, \tag{8}$$

where $M = \mathcal{M}_l R_{lr} \mathcal{M}_r^{-1}$ and $N = \mathcal{M}_l T_{lr}$. Note that in (8) M and N depend on the intrinsic and extrinsic camera parameter exclusively.

3 Epipoles and Epipolar Lines

The point where the interocular axis intersects the left (respectively the right) image plane is called left (respectively right) epipole (see fig. 3).

Let $z_{el} = (u_{el}, v_{el})$ and $z_{er} = (u_{er}, v_{er})$ be the left respectively the right epipole. We can determine directly from the equation (8) the image coordinates of z_{el} and z_{er} in dependence on the camera parameters. In order to do it we define $N = (n_1, n_2, n_3)^T$ and $-M^{-1}N = (\tilde{n}_1, \tilde{n}_2, \tilde{n}_3)^T$. Therewith the following relations are valid for the epipoles

$$z_{el} = \left(\frac{n_1}{n_3}, \frac{n_2}{n_3}\right) \tag{9a}$$

if $n_3 \neq 0$ and

$$z_{er} = \left(\frac{\tilde{n}_1}{\tilde{n}_3}, \frac{\tilde{n}_2}{\tilde{n}_3}\right) \tag{9b}$$

if $\tilde{n}_3 \neq 0$.

Note that $z_{el} \longrightarrow \infty$, if $n_3 \longrightarrow 0$ and $z_{er} \longrightarrow \infty$, if $\tilde{n}_3 \longrightarrow 0$. In this case the left and right image planes are coplanar and the epipolar lines coincide with the scan lines.

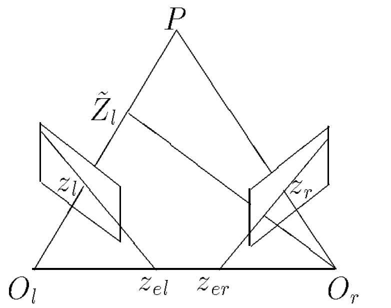

Figure 3. *The Epipoles and the Epipolar Lines*

Let E_i be the family of all planes in the 3D space passing through the interocular axis. Each such plane intercepts the left and the right image plane in a straight line. The so arisen two lines are called the corresponding epipolar lines, i.e., the left (respectively right) epipolar lines are all the lines lying on the left (respectively right) image plane and passing through the respective epipole (see fig. 3).

Let

$$v_l - v_{el} = m_l(u_l - u_{el}) \tag{10a}$$

be the equation any left epipolar line with slope m_l. The corresponding right epipolar line is

$$v_r - v_{er} = m_r(u_r - u_{er}). \tag{10b}$$

Assuming that all camera parameters are known, we can use (8), (9) and (10) in order to find the corresponding epipolar line for any given image point $\tilde{z}_l = (\tilde{u}_l, \tilde{v}_l)$ in the left stereo image. Inserting $\tilde{z}_l$ and (9) into (10) we can write the equation of the left epipolar line

ρ_l as follows:

$$\rho_l : v = \frac{n_3\tilde{v}_l - n_2}{n_3\tilde{u}_l - n_1}u + \frac{n_2\tilde{u}_l - n_1\tilde{v}_l}{n_3\tilde{u}_l - n_1}. \tag{11}$$

In order to determine the equation of the corresponding right epipolar line ρ_r we need an additional point $\tilde{z}_r$ lieying on ρ_r. Using the relation (8) and fixing any value for the depth Z_l we can find such point $\tilde{z}_r$. Let $\tilde{Z}_l$ be the fixed value for the depth (see fig. 3). Inserting $\tilde{Z}_l$ and $\tilde{z}_l$ into (8) we get $\tilde{z}_r$. The remaining unknown value Z_r can be eliminated inserting the last equation of the system

$$Z_r U_r = M^{-1}(\tilde{Z}_l \tilde{U}_l - N)$$

into the other two.

Note that the so deduced image point $\tilde{z}_r = (\tilde{u}_r, \tilde{v}_r)$ lies on the right epipolar line, but it not necessarily coincide with the corresponding image point of $\tilde{z}_l$.

Using both points z_{er} and $\tilde{z}_r$ is now a triviality to write the analytical equation of the right epipolar line.

4 A Method for Epipolar Line Estimation without Camera Calibration

Theoretical Background In the preceeding section the equations governing corresponding epipolar lines have been deduced, under the assumption that the intrinsic and extrinsic camera parameters are known in advance. The given procedure can be used in order to solve the correspondence problem, if a calibration step had been carried out. Such step is not feasible in a lot of situations. Moreover several times the correspondence problem must be solved for unregistered stereo sequences. In all these cases it is not possible to find the epipolar lines using the procedure described in the section 3. Alternatively the epipolar line equations can be estimated using exclusively the image information. In this section we give a method for epipolar line estimation without knowledge of the stereo geometry. The equation of corresponding epipolar lines is estimated directly by analyzing the stereo images. The basic assertion used in the method is given by the following theorem.

Theorem 1 *Let $z_l = (u_l, v_l)$ and $z_r = (u_r, v_r)$ be two corresponding image points. Then constants K_i, $i = 1, 2, \ldots, 9$, exist satisfying the equation*

$$\begin{aligned} &K_1 u_l u_r + K_2 u_l v_r + K_3 v_l u_r + K_4 v_l v_r + \\ &K_5 u_l + K_6 v_l + K_7 u_r + K_8 v_r + K_9 = 0. \end{aligned} \tag{12}$$

The constants K_i depend on the camera parameters exclusively and not on the image points z_l and z_r.

An immediately consequence of this theorem is that if a set of corresponding points are known, (12) can be used in order to compute the constants K_i, i.e., assuming that a sufficient number of true correspondences has been established, we may then estimate the nine constants K_i, solving a system of linear equations. After these constants have been estimated we can deduce for each image point $\tilde{z}_l = (\tilde{u}_l, \tilde{v}_l)$ on the left stereo image the corresponding right epipolar line as follows:

Inserting $\tilde{z}_l$ and K_i, $i = 1, 2, \ldots, 9$, into (12) we obtain

$$v_r = m_r u_r + a \tag{13}$$

with $m_r = -\dfrac{K_1\tilde{u}_l + K_3\tilde{v}_l + K_7}{K_2\tilde{u}_l + K_4\tilde{v}_l + K_8}$ and $a = -\dfrac{K_5\tilde{u}_l + K_6\tilde{v}_l + K_9}{K_2\tilde{u}_l + K_4\tilde{v}_l + K_8}$.

For any given point $\tilde{z}_l$ on the left image (13) gives the equation of the corresponding right epipolar line.

Description of the Method In order to find a set of correspondences a matching method is used. The established corresponding points, usually contain perturbations e.g. due to sampling and numerical errors or due to the ill-posed nature of the correspondence problem. For this reason the number of established correspondences has to be much larger then nine, so that solving the overdetermined system a very accurate approximation for K_i, $i = 1, 2, \ldots, 9$, can be found. A least squares approach is used in order to solve this overdetermined system. The following is the proposed algorithm for the estimation of epipolar lines in stereo images. The method an be used in order to process stereo image sequences in real time.

Algorithm

1. Select $n > 9$ interesting points on the left stereo image.
2. Find the corresponding points on the right stereo image.
3. Use (12) and the found correspondences in order to estimate the constants K_i, $i = 1, 2, \ldots, 9$.
4. For each image point in the left stereo image use (13) in order to compute the corresponding right epipolar line.

In the first step an interest operator may be used. Variance based operators are very robust and can be implemented easily. The so-called scale-space and the convolution of the image with the Laplacian of Gaussian take advantage of different image resolution levels. This procedure is more complicated compared with simple variance based interest operators but it gives feature points very accurately.

The establishment of corresponding point is done using a hierarchical block matching approach. Firstly rough disparity vectors to a pixel accuracy are estimated. These global vectors are selected from a large searching area and a low-pass filtered enlarged measurement window. Then a local step is performed with subpixel accuracy starting from the vectors estimated in the global step. A reliability check is followed and suspected vectors are eliminated from the set of correspondences. As a result of the correspondence estimation and the relation (12) we get a overdetermined system of linear equations. This system is solved using a least square method.

Proof of theorem 1 Equations (2) are equivalent to

$$\begin{aligned} \alpha_u X + (u_0 - u)Z = 0, \\ \alpha_v Y + (v_0 - v)Z = 0, \end{aligned} \tag{14}$$

(14) describe the equations of two planes E_1 and E_2 in the 3D camera coordinate system (X, Y, Z). The intersection of these two planes generates the line $\overline{OP}$ (see fig. 1). Let η_1 and η_2 be the normal vectors to E_1 and E_2, i.e.,

$$\eta_1 = (\alpha_u, 0, u_0 - u)^T,$$

and

$$\eta_2 = (0, \alpha_v, v_0 - v)^T.$$

A unit vector W on the line $\overline{OP}$ is given by the cross product of η_1 and η_2.

$$\begin{aligned} W = \eta_1 \times \eta_2 = \begin{vmatrix} i & j & k \\ \alpha_u & 0 & u_0 - u \\ 0 & \alpha_v & v_0 - v \end{vmatrix} = \\ (-\alpha_v(u_0 - u), -\alpha_u(v_0 - v), \alpha_u \alpha_v)^T. \end{aligned} \tag{15}$$

Now let us consider the stereo model shown in fig. 2. The line $\overline{OP}$ in the left camera coordinate system is given by the following equation

$$\hat{X}_l = \lambda W_l \tag{16}$$

with $W_l = (-\alpha(u_{0_l} - u), -\alpha_{ul}(v_{0l} - v_l), \alpha_{ul}\alpha_{vl})^T$ and $\lambda \in I\!R$. The corresponding right epipolar line of a given image point $z_l = (u_l, v_l)$ is the projection of line (16) onto the right image plane, i.e., the right epipolar line consists of the points (u_r, v_r), which satisfy the following relationship

$$Z_r U_r = M_r \tilde{X}_l + N_r = M_r \lambda W_l + N_r. \tag{17}$$

Here are $M_r = \mathcal{M}_r R_{rl}$ and $N_r = \mathcal{M}_r T_{rl}$.

It should be noted that (17) corresponds to the inverse transformation of (8), i.e., $\mathcal{M}_r$ is the perspective transformation matrix of the right camera, R_{rl} is the rotation matrix from right to left camera coordinate system and T_{rl} is the respective translation.

Let:

$$M_r = \begin{pmatrix} m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ m_{31} & m_{32} & m_{33} \end{pmatrix}$$

and

$$N_r = (n_{r1}, n_{r2}, n_{r3})^T.$$

Inserting (16) into (17) we get the following equation system

$$\begin{aligned} Z_r u_r &= -m_{11}\lambda\alpha_{vl}(u_{0l} - u_l) - m_{12}\lambda\alpha_{ul}(v_{0l} - v_l) + \\ & \quad m_{13}\lambda\alpha_{ul}\alpha_{vl} + n_{rl}, \qquad\qquad (18a) \\ Z_r v_r &= -m_{21}\lambda\alpha_{vl}(u_{0l} - u_l) - m_{22}\lambda\alpha_{ul}(v_{0l} - v_l) + \end{aligned}$$

$$
\begin{aligned}
& m_{23}\lambda\alpha_{ul}\alpha_{vl} + n_{r2}, && (18b)\\
Z_r \;=\;& -m_{31}\lambda\alpha_{vl}(u_{0l} - u_l) - m_{32}\lambda\alpha_{ul}(v_{0l} - v_l) + \\
& m_{33}\lambda\alpha_{ul}\alpha_{vl} + n_{r3}, && (18c)
\end{aligned}
$$

Multiplying (18c) by u_r and v_r respectively and inserting the so obtained equations into (18a) and (18b) respectively, we arrive at

$$
\begin{aligned}
\lambda(a_3u_r - a_1)u_l + \lambda(b_3u_r - b_1)v_l + \lambda(c_3u_r - c_1) &+ \\
\lambda(d_3u_r - d_1) + \lambda(e_3u_r - e_1) + f_3u_r - f_1 &= 0
\end{aligned}
$$

$$
\begin{aligned}
\lambda(a_3v_r - a_2)u_l + \lambda(b_3v_r - b_2)v_l + \lambda(c_3v_r - c_2) &+ \\
\lambda(d_3v_r - d_2) + \lambda(e_3v_r - e_2) + f_3v_r - f_2 &= 0
\end{aligned}
$$

where $a_i = m_{il}\alpha_{vl}$, $b_i = m_{i2}\alpha_{ul}$, $c_i = -m_{i2}\alpha_{ul}v_{0l}$, $d_i = -m_{i1}\alpha_{vl}u_{0l}$, $e_i = m_{i3}\alpha_{ul}\alpha_{vl}$, $f_i = n_{ri}$ for $i = 1, 2, 3$.

The parameter λ can be eliminated from the last two equations, so we get

$$
\begin{aligned}
(f_1 - f_3u_r)(a_3u_lv_r + b_3v_lv_r - a_2u_l - b_2v_l + g_3v_r - g_2) &= \\
(f_2 - f_3v_r)(a_3u_lu_r + b_3u_rv_l - a_1u_l - b_1v_l + g_3u_r - g_1)&
\end{aligned}
$$

Hereby are $g_i = c_i + d_i + e_i$ for $i = 1, 2, 3$. Finally (12) follows directly from this equation with

$$
\begin{aligned}
K_1 &= f_3a_2 - f_2a_3 & K_6 &= f_2b_1 - f_1b_2\\
K_2 &= f_1a_3 - f_3a_1 & K_7 &= f_3g_2 - f_2g_3\\
K_3 &= f_3b_2 - f_2b_3 & K_8 &= f_1g_3 - f_3g_1\\
K_4 &= f_1b_3 - f_3a_1 & K_9 &= f_2g_1 - f_1g_2\\
K_5 &= f_2a_1 - f_1a_2 & &
\end{aligned}
$$

References

[1] M. Ernst, H. D. Höhne, E. Izquierdo: *Signalverarbeitung für Vision-Like-Tele-Vision (VLTV) Systeme.* Annual Report -FKZ: 01BK304-, Heinrich Hertz Institute, Berlin, 1995.

[2] O. D. Faugeras and G. Toscani: *The calibration problem for stereo.* In Proc. of CVPR86, pp. 15-20, 1986.

[3] O. D. Faugeras, Q. T. Luong and S. J. Maybank: *Camera self-calibration: theory and experiments.* In Proc. 2nd ECCV, 1992.

[4] D. B. Gennery: *Stereo-Camera calibration.*, Proc. Image Understanding Workshop, pp 101-108, 1979.

[5] E. Izquierdo and M. Ernst: *Motion/Disparity Analysis and Image Synthesis for 3DTV.* Proc. on Int. Workshop on HDTV94, Torino, Italy, 1994.

[6] E. Izquierdo and M. Ernst: *Motion/Disparity Analysis for 3D-Video-Conference Applications.* Proc. on Int. Workshop on Stereoscopic and Three Dimensional Imaging, Santorini, Greece, 1995.

[7] H. C. Longuet-Higgins: *The reconstruction of a scene from two projections - configurations that defeat the 8-point algorithm.* Proc., 1st Int. Conf. Appl. AI, Denver, CO, USA, pp.395-397, 1984.

[8] H. P Travedi: *On the reconstruction of a scene from two unregistred images.* Proc. 5th National Conf. AAAI-86, Philadelphia, PA, USA, pp 652-656, 1986.

[9] R. Y. Tsai: *An efficient and accurate camera calibration technique for 3D machine vision.* Proc. IEEE Conference on Computer Vision and Pattern Recognition, pp 364-374, 1986.