Robotics Tools for Constructing Stereo Systems
Using off-the-Shelf Cameras

VITALIY RYBAK, ISMAEL DE J. PÉREZ V.
División de estudios de posgrado
Universidad Tecnológica de la Mixteca
Carretera a Acatlima Km. 2.5 Huajuapan de León, Oaxaca.
MÉXICO
rybak@mixteco.utm.mx, ps2008080007@ndikandi.utm.mx

Abstract: The presented work is a part of the development of the sensory system of a service robot. Tools developed to construct the low cost canonic and convergent stereosystems using off-the-shelf cameras are presented. The method of constructing proposed is based on the simultaneous individual calibration of every camera using the same mask. To adjust a camera with regard to other one a manipulator of 6 degrees of freedom is used. The developed software allows us to realize the change of the position and orientation of the camera in an automatic way. Using the calibration system the developed tools are applicable for examination and equalization of the focal lengths of two cameras of the stereo system under constructing and also of the cameras of ready stereosystems. As the mechanism to construct the stereosystems the manipulator PowerCube is used.

Keywords: Camera calibration, Off-the-shelf camera, Epipolar geometry, Stereoscopic system, Adjustment tool, Image distortion.

1 Introduction

Computational stereo vision is an important cue for visually guided robotics. It can be used for objects shapes and pose recognition in the task of vision-guided navigation and manipulation. Besides this area of application a large number of important applications such as surveying and mapping, engineering, architecture, involve quantitative measurements of coordinates of 3D points from a stereo pair or multiple images of 3D. Stereo vision is used also in geology, forensics, biology (e.g. growth monitoring), biometrics (e.g. 3D facial models), bioengineering and many other application areas.

The stereoscopic vision is applied for the measurement of the values of 3D coordinates of the points of a three-dimensional space by processing the images of two cameras that see a scene from two different positions. There are two types of stereoscopic system – the parallel, which has parallel optical cameras axes, and the convergent one. Simpler in the sense of the number of operations for the calculations of the three-dimensional coordinates is the parallel system. When the planes of images of two cameras coincide, the cameras have the same focal length, and the corresponding scanning lines coincide, such system is a canonic stereoscopic system. In this case for each point in one image, its corresponding point in the other image can be found by looking only along a corresponding horizontal line. The simplification of the calculations is an obligatory condition to construct the real-time systems of stereoscopic vision. The 3D measurements are based on stereo matching that is used to find points in stereo images depicting the same observed scene point. In spite of the fact that the problem of corresponding image points searching under admissible geometric and photometric distortions of stereo images is the central problem of computational stereo vision, consideration of this problem is beyond the present work.

The cost of a stereoscopic system has an influence on its application. The robotic tools for constructing the low cost canonic stereosystems using off-the-shelf cameras were presented in [1]. In the current work we present the means to construct both the canonic and convergent stereosystems using the canonic as intermediate phase. The main attention is paid to solve the problem of correction of the intrinsic characteristics of the cameras to reach acceptable exactitude of stereo measuring.

It is necessary to mention that in the field of measurement of the three-dimensional coordinates with the cameras which are placed in different positions with respect to the object of interest, more attention puts itself to the development of the calibrated and not calibrated systems based on the restrictions of the epipolar geometry [2], [3], and [4]. When two cameras watch a 3D scene from two different positions, there are several geometric relations between the 3D points and their projections in the 2D images that lead to restrictions between the
image points. These restrictions are generalized by fundamental matrix, when it is necessary to find the extrinsic and intrinsic parameters of the cameras, and essential matrix, when the intrinsic parameters are well-known. The existence of a big number of publications in this area demonstrates so much the big interest to this problem as the complexity of its solution. We use the calibration technique to get the fundamental matrix directly.

Reconstruction of 3D scenes from stereo pairs is based on matching of corresponding points in the left and right images. An analysis of different methods of stereo matching is presented in many publications. Corresponding points represent one and the same surface elements on each of the stereo pair images. The distance to the scene points visible on the left and right stereo-images are unambiguously determined by the relative displacement (disparity, parallax) of identical points on the images. The disparities (parallaxes) of corresponding points are defined as the differences of these points coordinates in a conventional coordinate grid attached to both images. In particular, the epipolar constraint reduces a 2D search space to 1D search along epipolar lines. For each point observed in one image, the same point must be observed in the other image on the associated epipolar line. The epipolar lines corresponds to the lines of intersection of each camera’s image plane by the epipolar plane passed over the analyzed scene point \( P \) and the focal points of the left and of the right cameras. The epipolar geometry is simplified if the image planes of two cameras coincide. In this case, the epipolar lines also coincide. Furthermore, the epipolar lines are parallel to the line between the focal points and can in practice be aligned with the horizontal axes of the two images (canonic stereo system). It means that for each point in one image, its corresponding point in the other image can be found by looking only along the same horizontal line. To determine distance to an object in real-time mode, stereo vision systems use triangulation based on epipolar geometry with horizontal epipolar lines. In case of deviation between the geometry of the physical stereoscopic system and the canonic system a rectification of the images can be applied forming the virtual canonic system [5]. This rectification needs to do the multiplication of a matrix 3x3 by a three-dimensional vector for every point of the image of every camera. The corresponding matrices are received during the calibration of the camera. It is possible to reduce the number of operations of rectification up to a multiplication of the homogeneous vector 1x4 by the matrix 4x4 received during the calibration, if we receive like calibration result the fundamental matrix \( C^c_iT_{c_i} \) of transformation of the coordinate system of one camera to the coordinate system of another one. The proposed method of adjustment of the physical stereoscopic system to the physical canonic system also is based on the use of the matrix \( C^c_iT_{c_i} \).

2 Constructing Canonic Stereosystems

Figure 1 shows the geometry of the canonic stereoscopic system. Here \( B \) is the base of the stereoscopic system, 3D Z-axis coincides with optical axis of the left camera, 3D X-axis and Y-axis are parallel to the x-axis and y-axis of the images coordinate systems. The relations between the coordinates of spatial point \( P(X,Y,Z) \) and the coordinates of its projections in both cameras are defined as following:

\[
X = \frac{B}{d}x_L; \quad Y = \frac{B}{d}y_L; \quad Z = \frac{B}{d}f
\]

where \( f \) is a focal length of the cameras,

\[
d = x_L - x_R\]

is the disparity or parallax of the system,

\( x_L \) and \( x_R \) are abscissas of the projection points of the spatial point \( P \) in the image planes of the left and right camera respectively.

![Geometry of the canonic stereoscopic system](image)

Fig. 1. Geometry of the canonic stereoscopic system.

To construct the canonic stereoscopic system it is necessary to have the cameras with the same value of focus and to have the physical means to provide the satisfaction of other conditions such as belonging of the image planes to the same plane, coincidence of the directions of the axes, parallelism of the straight line that passes across the projection centers with the scanning lines, and coincidence of the distance between projection centers of the cameras with the base of the system.

In case of using the commercial cameras it is necessary to have the means to equalize the focal distance of the pinhole model of the cameras and to
provide the satisfaction of other mentioned conditions.

In the current work we realize the adjustment mentioned previously, using the common mask for both cameras calibration. We use the well known Tsai method [6] that allows us to find the extrinsic and intrinsic parameters of each of the cameras by processing the images of the flat mask (coplanar calibration) or the images of a sequence of parallel flat masks located on the different known distances from an initial plane (non coplanar calibration).

The extrinsic parameters for one camera define the geometric relations between the camera coordinate system and the mask coordinate system. The transformation from the camera coordinate system to the mask coordinate system is presented by the homogeneous matrix:

\[ M_T C = \begin{bmatrix} r_{11} & r_{12} & r_{13} & mX_C \\ r_{21} & r_{22} & r_{23} & mY_C \\ r_{31} & r_{32} & r_{33} & mZ_C \\ 0 & 0 & 0 & 1 \end{bmatrix} \]

where the upper left 3x3 submatrix represents rotation matrix and the upper right 3x1 submatrix represents the position vector of the origin of the camera coordinate system with respect to the mask coordinate system. The vectors \( (r_{11}, r_{21}, r_{31})^T, (r_{12}, r_{22}, r_{32})^T, (r_{13}, r_{23}, r_{33})^T \) are the representations of the camera axis unitary vectors expressed in terms of the axis unit vectors of the mask coordinate system.

The intrinsic parameters include: 
- \( f \) - Effective focal length, or image plane to projective center distance, 
- \( k_i \) - Lens distortion coefficient, 
- \( s_x \) - Uncertainty scale factor for \( x \), due to TV camera scanning and acquisition timing error, 
- \( (C_x, C_y) \) - Computer image coordinates for the origin in the image plane.

In case of the difference between the focal lengths of the camera models it is possible to equalize them with the iterative calibration after every focal lengths correction.

In the result of calibration of two cameras with the same mask the matrices of transformation between the mask coordinate system and coordinate systems of the left (L) and right (R) cameras \( MTC_L \) and \( MTC_R \) are known. So, it is possible to find the space relations between two cameras:

\[ C_TC_{6e} = C_M MTC_{6e} \]  

To define if the requirements mentioned previously are satisfied the expression of the matrix \( C_TC_{6e} \) for the canonic stereoscopic system is used:

\[ C_TC_{6e} = \begin{bmatrix} 1 & 0 & 0 & B \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \]  

The problem to fit the two cameras to receive the canonic stereoscopic system is realized by the movement of a camera \( C_R \) with respect to the immovable camera \( C_L \) that provides the correspondence of the matrix \( C_TC_{6e} \) with the Equation (2). If the camera \( C_R \) is fixed in a tool of the camera movement, the adjustment problem is formulated like the problem to find the space state of the tool, which corresponds with the position and orientation of the fitted camera \( C_R \) with respect to the coordinate system of the camera \( C_L \). In the current work a manipulator of 6 degrees of freedom PowerCube is used as the adjustment tool [7].

The movable camera \( C_R \) is rigidly fixed by the gripper of the manipulator. The relation between the coordinate system of the camera and the coordinate system of the gripper is received by the calibration of the camera-gripper system (matrix \( MTC_{6e} \), where superscript 6 identifies the gripper coordinate system). Before realizing this calibration the relations between the mask coordinate system and the manipulator basic coordinate system are defined. An extension (needle) is rigidly fixed in the gripper along the axis \( Z_e \) with the distance \( e \) between extension end point and the origin \( O_e \) of the gripper coordinate system. The vector \( (0,0,e)^T \) defines the end point position with respect to the gripper coordinate system. For each point \( M P_i \) \( (i = 0,1,2) \) by the solution of the direct kinematic problem the position of the origin is defined.

The end point is combined with three points of the mask \( M P_0, M P_1, M P_2, \) located correspondingly on the origin and on the axes of the mask coordinate system. For each of these values the positions of the manipulator joints angles are stored and by the solution of the direct kinematic problem the matrices of transformation between the gripper coordinate system and the manipulator basic coordinate system \( T_{e(i)} \), \( i = 0,1,2 \) are calculated. The position of these
three points with respect to the manipulator basic coordinate system is defined as:

\[ \mathcal{T}_M^0 = T_k^{0,i}(0 \ 0 \ e)^T, \quad i = 0, 1, 2 \] (3)

The knowledge of the points allows us to define the origin and the representations of the mask axis unitary vectors with respect to the manipulator basic coordinate system. The matrix of transformation from the mask coordinate system to the manipulator basic coordinate system is:

\[ 0_T^M = \begin{pmatrix} \bar{x}_M & \bar{y}_M & \bar{z}_M & \bar{P}_0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \]

Where

\[ \bar{x}_M = (\bar{P}_1 - \bar{P}_0) / \| \bar{P}_1 - \bar{P}_0 \| \]
\[ \bar{y}_M = (\bar{P}_2 - \bar{P}_0) / \| \bar{P}_2 - \bar{P}_0 \| \]
\[ \bar{z}_M = \bar{x}_M \times \bar{y}_M \]

With the matrix \( M_T_0 = 0_T^M \) the matrix of spatial relations between the camera \( C_r \) and gripper is:

\[ c_r T_6 = c_r T_M^M T_0 0_T^T \] (4)

where the matrix \( 0_T^T \) is defined by solution of the direct kinematics problem for the manipulator configuration (joint angles) that corresponds with the camera \( C_r \) spatial position during the camera calibration.

The matrix \( 0_{T_6}^T \) that defines the spatial state of the gripper with camera \( C_r \) fitted to canonic stereo system is calculated as:

\[ 0_{T_6}^T = 0_T^M M_T c_r^T c_r T_6 \]

For the matrix \( 0_{T_6}^T \) the problem of inverse kinematics is solved to find the configuration of the manipulator that allows reach the position and direction necessary to form the canonic stereoscopic system. In order to verify that the cameras form a canonic stereoscopic system the space relations among them by means of the repetition of the calibration process and the application of the equation (1) are calculated again. In case of having a significant difference in the present position and the equation (2) it is possible to realize a correction. In order to realize this correction the matrix \( 6_{T_6}^T \) calculates by means of the equation (4) and:

\[ 6_{T_6}^T c_r = \begin{pmatrix} 6_{x C_r} & 6_{y C_r} & 6_{z C_r} & 6_{P C_r} \\ 0 & 0 & 0 & 1 \end{pmatrix} = c_r T_6^{-1} \]

The new matrix \( 0_{T_6}^{cor} \) corrected one calculates like:

\[ 0_{T_6}^{cor} = \begin{pmatrix} R & 0 \\ 0 & 1 \end{pmatrix} \]

where \( R \) is the upper left 3x3 submatrix of matrix \( 0_{T_6}^T \). Solving the problem of inverse kinematics for the matrix \( 0_{T_6}^{cor} \) we calculate the configuration of the manipulator that allows form a canonic stereoscopic system.

After fixation of the fitted cameras rigidly each to another and release movable camera from the gripper and immovable one from its supporting device, canonic stereoscopic system is constructed and again its resultant matrix \( c_r T_6^{cor} \) is calculated.

Because the formulas of the spatial coordinate’s calculation corresponds to the ideal cameras it is necessary to provide the correction of distortion of the cameras lenses using the value of lens distortion coefficient \( k_1 \).

To carry out it, we use the method of bilinear interpolation with taking into consideration the possibility of real-time realization by using the FPGA technology [8].

The compensation of image distortion transforms the stored images to canonic presentation: the centers of images coincide with optical centers and the scanning lines of the same ordinate coincide with corresponding epipolar line.

3 Constructing Convergent Stereosystems

To transform the stereo system from an canonic mode to a convergent one it is sufficient to rotate the vector \( c_r \hat{B} = (B, 0, 0)^T \), that represents the stereo system base, together with the camera \( C_r \) about the \( Y_{C_r} \) axis of the camera \( C_r \) with \( \theta \) convergence angle. The
The matrix $T_{6}^{C_{R}}$ that defines the spatial state of the gripper with camera $C_{R}$ fitted to convergent stereo system is calculated as:

$$T_{6}^{C_{R}} = T_{M}^{S} T_{C_{L}}^{-1} T_{C_{R}}^{C_{L}}$$

The convergent stereo system received has the same base distance $B$ as canonc one. It is necessary to notice that we used the canonic stereo system presentation as a intermediate phase of constructing the convergent stereo system only to get the explicit form of the matrix $T_{6}^{C_{R}}$. Resulting formula for $T_{6}^{C_{R}}$ does not require the calculation of $T_{6}^{C_{L}}$.

The calculation of the spatial coordinates of point $P(X,Y,Z)$ is based on the search of the point of intersection of two projective beams passing through the corresponding points defined with respect to the cameras coordinate systems by the vectors $\tilde{P}_i = (c_X, c_Y, f)^T$, $i = L, R$, [9].

The coordinate system $O_SX_S Y_S Z_S$ of the convergent stereo system is presented in the Fig. 2.

![Fig. 2. Geometry of the convergent stereoscopic system.](image)

The coordinates of the projection centers of the left and right cameras with respect to this coordinate system are defined by the vectors:

$${}^S\tilde{O}_L = \left(-\frac{B}{2},0,0\right)^T$$ and $${}^S\tilde{O}_R = \left(\frac{B}{2},0,0\right)^T$$

The matrix of transformation from the cameras coordinate systems to the stereo system coordinate system are:

$$sT_{C_{L}} = \begin{bmatrix}
\cos\theta & 0 & \sin\theta & -\frac{B}{2} \\
0 & 1 & 0 & 0 \\
-\sin\theta & 0 & \cos\theta & \frac{B}{2} \\
0 & 0 & 0 & 1
\end{bmatrix}$$

$$sT_{C_{R}} = \begin{bmatrix}
\cos\theta & 0 & \sin\theta & \frac{B}{2} \\
0 & 1 & 0 & 0 \\
-\sin\theta & 0 & \cos\theta & -\frac{B}{2} \\
0 & 0 & 0 & 1
\end{bmatrix}$$

With respect to the coordinate system $O_SX_S Y_S Z_S$ the corresponding points are defined by the vectors:

$$s\tilde{P}_i = sT_{C_{L}}^{-1}\tilde{P}_i, \quad i = L, R$$

The unitary vectors that correspond to the projective beams are calculated as:

$$\hat{u}_i = \frac{s\tilde{P}_i - s\tilde{O}_i}{\sqrt{||s\tilde{P}_i - s\tilde{O}_i||^2}}, \quad i = L, R$$

where $s\tilde{P} = s\tilde{O}_L + a\hat{u}_L$. In an ideal case does exist two numbers $a$ and $b$ such that $a\hat{u}_L = B + b\hat{u}_R$. In this case $s\tilde{P} = s\tilde{O}_L + a\hat{u}_L$. Because the possible errors, in practice two projective beams do not cross each another and the problem is formulated as a problem of searching the point that lie on the same distance from the projective lines in the place of their maximum closing.

$$\tilde{P} = \frac{a\hat{u}_L + (s\tilde{B} + b\hat{u}_R)}{2}$$

where

$$a_0 = \frac{(\hat{u}_L \cdot s\tilde{B}) - (\hat{u}_L \cdot \hat{u}_R)(\hat{u}_R \cdot s\tilde{B})}{1-(\hat{u}_L \cdot \hat{u}_R)^2}$$

$$b_0 = \frac{(\hat{u}_L \cdot \hat{u}_R)(\hat{u}_L \cdot s\tilde{B}) - (\hat{u}_R \cdot s\tilde{B})}{1-(\hat{u}_L \cdot \hat{u}_R)^2}$$

correspond to the meaning of $a$ and $b$ that minimize the value of $J(a,b) = (a\hat{u}_L - (B + b\hat{u}_R))^2$.

4 Results

Hardware to construct stereoscopic systems using commercial cameras is presented in Fig. 3.
In figure 4 show a canonic stereoscopic system after fixation of the fitted cameras rigidly each to another and release movable camera from the gripper and immovable one from its supporting device.

An example of matrix $T_{c_r}^{c_s}$ for canonic stereo system with cameras DCAM with focus 2.9 mm. is presented below:

$$T_{c_r}^{c_s} = \begin{bmatrix} 0.999997 & -0.002282 & 0.000316 & 124.517 \\ 0.002282 & 0.999998 & -0.000342 & -0.015927 \\ -0.000315 & 0.000343 & 1.0 & 0.062144 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

The upper left 3x3 submatrix represents the rotation of the right camera coordinate system and the upper right 3x1 submatrix represents the position vector of the origin of the right camera coordinate system with respect to the left camera coordinate system.

It is visible that the matrix $T_{c_r}^{c_s}$ differs from the matrix $T_{c_r}^{c_s}$. The difference in orientation is depreciatingly insignificant. The base value is 124.517 mm. The right camera origin is pulled down along the ordinate axis at -0.016 mm. that corresponds to 2.88 pxs. The displacement along the optical axis is 0.062 mm. that corresponds to about 2% of the focal length that is acceptable for such kind of stereo systems. It is necessary to take into account the displacement of the right image along the ordinate axis by bringing the left camera scan line in correspondence with displaced at 3 pxs the right camera scan line.

5 Conclusion

The developed tools allow construct canonic and convergent stereoscopic systems using off-the-shelf cameras, and they offer the possibility of being able to experiment with systems of vision of diverse characteristics and low cost. The tools are universal and applicable to form stereoscopic systems independent of the type of used cameras. The adjustment tools are accessible without dependency of the type of manipulating robot of six degrees of freedom under the condition of which the solution of the problem of inverse kinematics is well-known.

References: