# A simple Method of Multiple Camera Calibration for the Joint Top View Projection

Mikhail Mozerov[1], Ariel Amato[1], Murad Al Haj[1], and Jordi Gonzàlez[2]

[1] Computer Vision Center and Department d'Informàtica. Universitat Autònoma de Barcelona, 08193 Bellaterra, Spain mozerov@cvc.uab.es

[2] Institut de Robòtica i Informàtica Industrial(UPC-CSIC),Edifici U Parc Tecnològic de Barcelona. 08028, Spain.

**Summary.** A simple method for multiple camera calibration based on a novel geometric derivation is presented. The main advantage of this method is that it uses only three points in the world coordinate system to achieve the calibration. Rotation matrix and translation vector for each camera coordinate system are obtained via the given distance between the vertices of the marker triangle formed by the three points. Therefore, the different views from the different cameras can be converted into one top view in the world coordinate system. Eventually, the different trajectories traced by certain tracked agents on the floor plane can be obtained from different viewpoints and can be matched in a joint scene plane.

## 1 Introduction

Camera calibration techniques play an important role in the different applications of computer vision. The objective of the calibration process is to obtain both the intrinsic and the extrinsic camera parameters. The intrinsic parameters are decided by the inner geometry and the optical characteristics of the camera; these include focal ratio and radial distortion factor. The extrinsic parameters reflect the relationship between the image plane and the world plane; these include rotation matrix **R** and translation vector **T**. However, the importance of each of these parameters, intrinsic or extrinsic, depends on the problem to be solved and on the camera model. For example, a new generation of digital cameras provides rectified image sequences; therefore, the obtained images are not affected by the radial distortion of the camera. Most traditional camera calibration techniques require specific knowledge about the geometric characteristics of the referenced object, such as Direct Linear Transformation (DLT)[1], which solves the perspective matrix linearly; similar methods include Tsais[2],and Zhangs [3].Plane based methods[4]-[5],use the same DLT paradigm, but show more flexibility. These methods use multiple view approach for each camera and are not very useful for our problem. The main goal of this work is to use calibration in order to convert the image planes obtained from multiple camera views into one top view of an inspected scene, as shown in Fig. 2. This allows us to interpret certain actions, after extracting them from the scene, in a chosen world coordinate system. The similar approach is used in the paper of Lee et al.[6]. This can be done by performing trajectory matching of the tracked foot points of the moving agent. Many works use these approach to handle uncertainty of one point view[7]-[8]. In such a scenario, the most important aspect is to achieve sufficient accuracy in the ground plan which contains the marker triangle (the floor plane on which the agent is walking). Our experiments show that in this situation, the focal ratio does not play an important role and the lens focus scope ($\pm 5\%$)produces minimal distortion in the ground plan measurements. Furthermore, in the extreme case where the image plane and the world plane coplanar, the accuracy of the ground plane measurements do not depend on the focal ratio. Of course, the focal ratio is essential in determining the distance between the camera and the scene or in measuring the heights of objects in the scene; however, in these case, the focal length can be obtained a priori, either through a simple geometric approximation or from the datasheet of the camera and lens. The DLT methods require more than three points in the world coordinate system, for example, the Tsais calibration technique requires five points. Three point problem usually is considered as theoretical problem without implementing in real experiments[9]-[10]. Our main contribution is to propose calibration method that uses only three world coordinate points, eventually calibration process can be considered as two independent and intuitively clear parts with simple geometrical interpretation; the first part is obtaining values of marker triangle points in a camera coordinate system while the second is deriving $\mathbf{R}$ and $\mathbf{T}$. This paper is organized as follows. Section 2 presents our camera model while section 3 discusses our calibration process. Section 4 describes inverse perspective mapping. Section 5 shows our experimental results. Concluding remarks are made in section 6.

## 2 Camera Model

As shown in Fig. 2 pinhole camera model is used. We suppose that the focal ratio parameter is known or predetermined. The right-hand camera coordinate system $\mathbf{p}_0$, $\mathbf{X}_c, \mathbf{Y}_c, \mathbf{Z}_c$ is defined as origin $\mathbf{p}_0$,, which coincide with optic center of camera,$\mathbf{Z}_c$ axis coincident with optical axis and has invers direction, which means that all visible point including projection points have depth values strongly less than zero. A scene point $\mathbf{P}$ can be represented in the camera coordinate system as $\mathbf{P}_n = [X_n, Y_n, Z_n]^t$ and in the world coordinate system $(\mathbf{O}w, \mathbf{X}_w, \mathbf{Y}_w, \mathbf{Z}_w)$ as $\hat{\mathbf{P}}_n = \left[\hat{X}_n, \hat{Y}_n, \hat{Z}_n\right]^t$. The projection of any world point onto image plane, which is parallel to the camera plane, is denoted by lowercase as $\mathbf{p}_n = [x_n, y_n, -f]^t$, where $f$ is the camera focal length. The relationship between these two coordinate values is as follow.

$$\mathbf{P}_n = \frac{Z_n}{f}\mathbf{p}_n \tag{1}$$

As we work with the raster images, it is useful also to consider the raster representation of the image plane points $\mathbf{m}_n = [i_n, j_n, -f_{pix}]^t$ , $i \in [-I, I]$ , $j \in [-J, J]$ , where $f_{pix}$ is the focal length expressed in the raster pixels, I and J are the half size of the image matrix in $X$ and $Y$ directions respectively. If to denote the distance between two neighbor raster pixels as $\mu = CCD_x/2I$ , where CCD$_x$ is the physical size of the camera CCD matrix in $X$ direction, then, the focal length in pixels is $f_{pix} = f/\mu$.Let us denote the inverse of the intrinsic constant $f_{pix}^{-1}$ as $\varphi$, and now we are prepared to express an image plane point via raster representation:

$$\mathbf{p}_n = \mu\mathbf{m}_n = f\mathbf{u}_n \tag{2}$$

where $\mathbf{u}_n = \varphi\mathbf{m}_n = [\varphi i_n, \varphi j_n, -1]^t$ is a convenient raster representation of a point in the image plane.

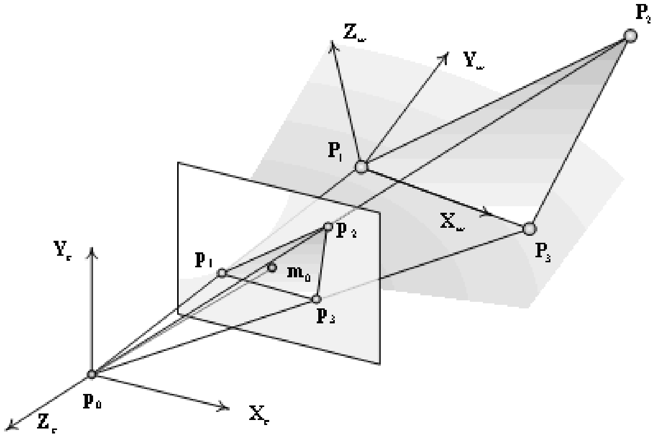

Fig. 1: Pinhole camera model for calibration with three world coordinate points.

To obtain focal ratio of our cameras we use a simple geometrical approach with just two points: it is assumed that two world coordinate points $\mathbf{P}_1$, $\mathbf{P}_2$ have two given projection on the image plane $\mathbf{p}_1$, $\mathbf{p}_2$ as it is shown in Fig. 1. We also suppose that the distance between the two world points $d12 = |\mathbf{P}_2 - \mathbf{P}_1|$ is known and distances between optical center $\mathbf{p}_0$ and given points $\mathbf{P}_1$, $\mathbf{P}_2$ are measured (for example using laser distancemeter as it is in our experiments). Now, the angle between unknown vectors $\mathbf{p}_1$, $\mathbf{p}_2$ can be calculated using cosine theorem, and finally cos( $\mathbf{P}_1$,$\mathbf{P}_2$)= cos($\mathbf{p}_1$,$\mathbf{p}_2$). Then the focal ratio is the a simple solution of this equality.

## 3 Calibration Algorithm

Now, we assume that the intrinsic constant $\varphi$ is given or predetermined. Note, that in this case only the depth parameter $Z_n$ is needed to completely describe all the visible points in the world coordinate system of a camera (see Eq.( 1-2). On the other hand, three points of the world coordinate system (if they do not belong to the same line) are enough to determine extrinsic coordinate system. So, the calibration process can be divided into two independent parts. First, we calculate values of marker triangle points in a camera coordinate system and then obtain extrinsic parameters $\mathbf{R}$ and $\mathbf{T}$. Let us denote the distances between three point in the marker triangle $\mathbf{P}_1$,$\mathbf{P}_2$,$\mathbf{P}_3$ and scalar product matrix of its projection points as $d_{kl} = |\mathbf{P}_k - \mathbf{P}_l|$and $\sigma_{kl} = \mathbf{u}_k \mathbf{u}_l$ with $k$,$l$=1,2,3. Then the three perspective points problem can be defined as follow: find three desired values $Z_1$, $Z_2$, $Z_3$ that satisfy the system of three equations

$$Z_k^2 - 2Z_k Z_l a_{kl} + Z_l^2 b_{kl} - c_{kl} = 0 \tag{3}$$

where $a_{kl} = \sigma_{kl}/\ \sigma_{kk}$, $b_{kl} = \sigma_{ll}/\ \sigma_{kk} c_{kl} = d_{kl}/\ \sigma_{kk}$.

The system of these nonlinear equations can be solved in two ways. One way leads to a biquadratic closed form, which is a big problem itself. We propose another way with a fast iterative algorithm that provides computer level accuracy. Let us suppose that $0 > Z_3 \geq Z_1 \geq Z_2$. In this case the system has just one solution and the values of $Z_3$, $Z_1$ can be calculated via value $Z_2$

$$Z_k = Z_2 a_{k2} + \sqrt{Z_2^2\left(a_{k2}^2 - b_{k2}\right) - c_{k2}} = 0,\ k = 1,3 \tag{4}$$

We remember that depth values of the marker triangle points satisfy $0 > Z_3 \geq Z_1 \geq Z_2$ and the expression under square root in Eq.( 4) must not be negative. These constraints provide us with the limits of $Z_2$ variable domain. Then substituting Eq.( 4) in ( 3) we have one equation to solve

$$\left(Z_1(Z_2)\right)^2 - 2Z_1(Z_2)Z_3(Z_2)a_{13} + Z_l^2 b_{13} - c_{13} = 0 \tag{5}$$

This equation with given order constrain $0 > Z_3 \geq Z_1 \geq Z_2$ has only one solution within the calculated segment [min($Z_2$),max($Z_2$)]. In such a case, the solution can be obtained iteratively using dichotomic division algorithm of the desired variable domain. The double float precision computer accuracy can be achieved with less than 50 steps. Now we are able to derive rotation matrix and translation vector that transform the new world coordinates of a point into the camera world coordinates

$$\mathbf{P}_n = \mathbf{R}\hat{\mathbf{P}}_n + \mathbf{T} \tag{6}$$

To obtain the transformation parameter, first, it is necessary to define new world coordinate system. Let us put the origin of our world coordinate system to one of the marker triangle vertices $\mathbf{O}_w = \mathbf{P}_1$, and take one of the triangle leg as the **X** coordinate axis. It is reasonable to put our marker triangle into the new coordinate system plane with Z=0. Then, the new world coordinate system can be defined by its basis

$$\mathbf{X}_w = \frac{(\mathbf{P}_3 - \mathbf{P}_1)}{|\mathbf{P}_3 - \mathbf{P}_1|}; \, \mathbf{Z}_w = \frac{\mathbf{X}_w \times (\mathbf{P}_2 - \mathbf{P}_1)}{|\mathbf{X}_w \times (\mathbf{P}_2 - P_1)|}; \, \mathbf{Y}_w = \mathbf{Z}_w \times \mathbf{X}_w; \tag{7}$$

Now, rotation matrix can be represented as a simple combination of the basis vectors and translation vector is equal to origin vector

$$\mathbf{R} = [\mathbf{X}_w \mathbf{Y}_w \mathbf{Z}_w] \, ; \; \mathbf{T} = \mathbf{P}_1; \tag{8}$$

## 4 Inverse Perspective Mapping

One of the implementations of the multiple camera calibration is inverse perspective mapping. Once your tracking algorithm localizes the pixel of interest (i,j) in the image of one camera (in our case it is a foot point of an agent) you have to project this pixel onto joint plane in the world coordinate system for the matching with another point obtained from the image of the second camera. This projection process is referred to as inverse perspective mapping. In other words, we need to derive a function $\hat{\mathbf{P}}(i,j)$ that project arbitrary pixel with indexes (i,j) into the real world plane. To derive the desired function, first, we need to obtain the value of the optical center point in the joint world coordinate system

$$\hat{\mathbf{p}}_0 = \mathbf{R}^t \left( \mathbf{p}_0 - \mathbf{T} \right) = -\mathbf{R}^t \mathbf{T} \tag{9}$$

Let us denote the pure rotation of the image plane vectors as

$$\breve{\mathbf{u}}_{ij} = \mathbf{R}^t \mathbf{u}_{ij} = \mathbf{R}^t \left[ \varphi i, \varphi j, -1 \right]^t \tag{10}$$

then, the desired inverse perspective mapping function is

$$\hat{\mathbf{P}}_{ij} = \hat{\mathbf{p}}_0 - \breve{\mathbf{u}}_{ij} \frac{\hat{z}_0}{\breve{z}_{ij}} \tag{11}$$

If the localization algorithm is working properly and pixels of interest (i,j) in the image of one camera and (i',j') in the image of other camera belong to the same agent the distance between these two points $\left| \hat{\mathbf{P}}(ij) - \hat{\mathbf{P}}(i'j') \right|$ in the joint world coordinate plane might be zero. Of course, the excellent accuracy of the method can be flawed by the real world ground curvature, and to check it we performed experiments with real image sequences.

## 5 Computer Experiments

In our experiments we use the real data sequences of two view point. In general the goal is to compare trajectories form different points of view in one joint plane. The marker triangle (Egyptian triangle with 3x4x5 meter legs) is formed using the corner points of colored square markers Fig. 2. To check the accuracy of the method we map two images in one joint top view. The green border of the in Fig. 2 shows the overlapping effect of the two inverse mapping. We can see that visual features (zebra crossing lines e. g.) coincide on the arbitrary drown fringe. The inverse perspective mapping also performed for two trajectories taken from different cameras and belongs to the same tracked agent (the green colored line from the first camera and red colored line form the second).

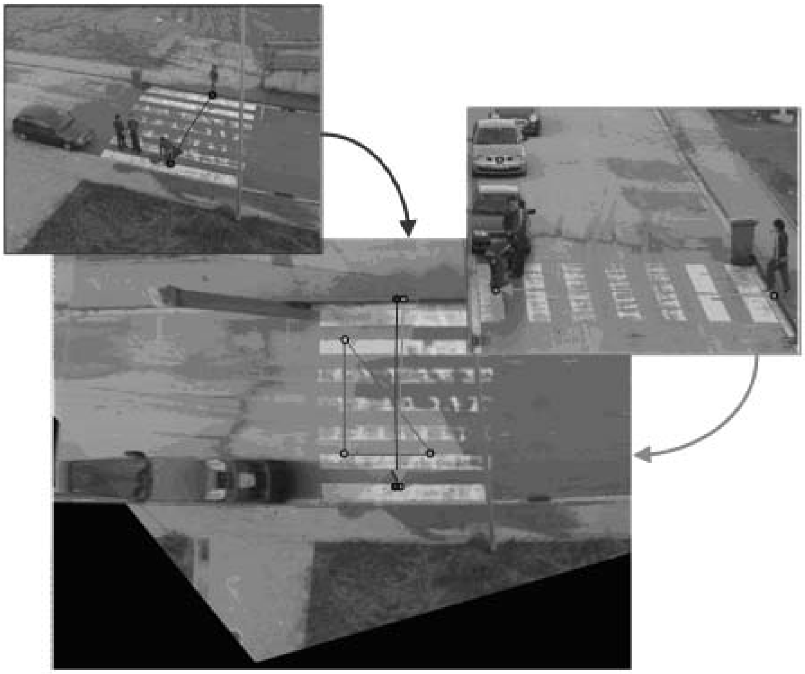

Fig. 2: Projection of two multiple camera views onto the top view of the inspected scene.

We estimate the distances in the joint ground plan and maximum error value is 45 cm. Note, that this error is not a result of inaccurate calibration or inverse mapping, but mostly due to inaccuracy of the foot point localization in the image. The no planarity of the scene affects on the curvature of zebra crossing relative to strait black line in Fig. 3, but this effect produces error measurements less than 10 cm in the field of interest. So, we can conclude that the proposed calibration method has a sufficient accuracy for handling multiple cameras views.

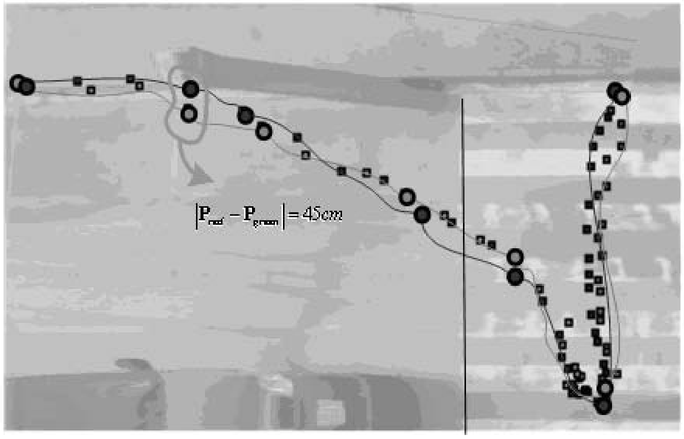

Fig. 3: The result of two trajectory matching.

## 6 Conclusion

We propose simple calibration method, which is sufficient for multiple camera handling. Trajectories of the foot points of the tracked agents are usually not very accurate due to intrinsic property of segmentation and tracking algorithms. The experimental results show that the accuracy of the trajectories matching with proposed calibration and geometrical mapping technique highly overpass tracking approximation.

## Acknowledgements

This work is supported by EC grants IST-027110 for the HERMES project and IST-045547 for the VIDI-video project, and by the Spanish MEC under projects TIN2006-14606 and DPI-2004-5414. M. Mozerov and Jordi Gonzàlez also acknowledge the support of a Ramon y Cajal and Juan de la Cierva Postdoctoral fellowships from the Spanish MEC, respectively.

## References

1. Y. I. Abdel-Aziz, and H.M. Karara, “Direct linear transformation from comparator coordinates into object space coordinates in close-range photogrammetry”, *Proc., of the ASP Symposium on Close-Range Potogrammetry,* pp. 1-18, 1971.
2. Tsai, R., “A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses, *IEEE J. Robot. Autom.,* Vol. 3 pp. 323-344, 1987.
3. Zhang, Z, “A flexible new technique for camera calibration, *IEEE Trans. Pattern Analysis and Machine Intelligence,* Vol. 22 pp. 1330-1334, 2000.
4. Sturm, P., and Maybank, S., “On plane-based camera calibration: A general algorithm, singularities, applications, *IEEE CVPR,* pp. 432-437, 1999.
5. Lucchese, L., “Geometric calibration of digital cameras through multiview rectification, *Image Vis. Comput,* Vol. 23 pp. 517-539, 2005.
6. Lee, L., Romano, R., Wang, L., and Stein, G., “Monitoring activities from multiple video streams: Establishing a common coordinate frame, *IEEE Trans. Pattern Analysis and Machine Intelligence,* Vol. 22 pp. 758–767, 2000.
7. Cai, Q., and Aggarwal, J. K., “Human motion in structured environments using a distributed camera system, *IEEE Trans. Pattern Analysis and Machine Intelligence,,* Vol. 21 pp. 1241–1247, 1999.
8. Chang, T.-H., and Gong, S., “Tracking multiple people with a multi-camera system, *IEEE Workshop Multi-Object Tracking,* pp. 19–26, 2001.
9. Haralick, R., Lee, C., Ottenberg, K., Nolle, M., “Review and Analysis of Solutions of the Three Point Perspective Pose Estimation Problem, *International Journal of Computer Vision,* Vol. 13 pp. 331–356 1994.
10. Nister, D., “A minimal solution to the generalized 3-point pose problem. On plane-based camera calibration A general algorithm, singularities, applications, *IEEE CVPR* Volume 1 pp. I560–I567 2004.