Calibration Method for an Augmented Reality System

S. Malek, N. Zenati-Henda, M. Belhocine and S. Benbelkacem

Abstract—In geometrical camera calibration, the objective is to determine a set of camera parameters that describe the mapping between 3D references coordinates and 2D image coordinates. In this paper, a technique of calibration and tracking based on both a least squares method is presented and a correlation technique developed as part of an augmented reality system. This approach is fast and it can be used for a real time system

Keywords—Camera calibration, pinhole model, least squares method, augmented reality, strong calibration.

I. INTRODUCTION

THE term Augmented Reality (AR) is used to describe systems that blend computer generated virtual objects with real environments [1]–[2]. AR is defined as a technology in which a user's view of the real world is enhanced or augmented with additional information generated by a computer [3]. This augmentation may include labels (text), 3D rendered models, or shading and illumination changes. AR allows a user to work with and examine the physical world [4], while receiving additional information about the objects in it.

In order for AR to be effective, the real and computer-generated objects must be accurately positioned relative to each other and properties of certain devices must be accurately specified. This implies that certain measurements or calibrations need to be made at the start of the system [5].

Calibration is the first step in an AR system. Camera calibration in the context of three-dimensional computer vision is the process of determining the internal camera geometric and optical characteristics (intrinsic parameters) and the 3D position and orientation of the camera frame relative to a certain world coordinate system (extrinsic parameters) [6]. In many cases, the overall performance of the computer vision system strongly depends on the accuracy of the camera calibration [7].

Salim Malek is an attached researcher at the vision team, Robotics division, Advanced Technologies Development Center, CDTA (corresponding author: e-mail: s_malek@cdta.dz).

Nadia Zenati-Henda is a Master researcher at the vision team, Robotics division, Advanced Technologies Development Center, CDTA (e-mail: nzenati@cdta.dz).

Mahmoud Belhocine is a Master researcher at the vision team, Robotics division, Advanced Technologies Development Center, CDTA (e-mail: mbelhocine@cdta.dz).

Samir Benbelkacem is an engineer at the vision team, Robotics division, Advanced Technologies Development Center, CDTA (e-mail: sbenbelkacem@cdta.dz).

There are different methods used to estimate the parameters of the camera model. They are classified in three groups:

Non linear optimization techniques: the camera parameters are obtained through iteration with the constraint of minimizing a determined function. This technique is used in many works [8]–[9]–[10].

Linear techniques which compute the transformation matrix: due to slowness and computational burden of the first technique, closed-form solutions have been also suggested. These techniques use the least squares method to obtain a transformation matrix which relates 3D points with their projections. This technique is fast and can be used in a real time application, but it ignores the nonlinear radial and tangential distortion components. Also, it was revised in several works [11]–[12].

Two-steps techniques. These approaches [13]-[14] consider a linear estimation of some parameters while the others are estimated iteratively.

The technique described in this paper has been developed as part of an AR system. Furthermore, it can be used in other applications. The least squares method to calibrate the camera is used in this case and a correlation technique to track the virtual object in the images sequence.

This paper is organized as follows: In Section II, a brief survey of a camera model is given. In Section III, the details of the camera calibration approach is presented, followed by a description of the technique of tracking in section IV. A description and discussion of experimental results are presented in Section V. Finally, in Section VI, conclusions are given.

II. CAMERA MODEL

The model is a mathematical formulation which approximates the behavior of any physical device, i.e. a camera. In such a case, the internal geometry and the position and orientation of the camera in the scene are modeled.

In an AR system, there are both real entities in the user's environment and virtual entities. Calibration is the process of estimating the parameters of camera in order to match the virtual objects with their physical counterparts. These parameters may be the optical characteristics of a physical camera as well as position and orientation information of various entities such as the camera and the various objects.

World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:2, No:9, 2008

In the following, the pinhole model used in the application will be described.

A. A Pinhole Model

There are several camera models depending on the desired accuracy. The simplest model is the pinhole Model proposed by Hall [15].

In an AR system, it's necessary to know the relationship between the 3D object coordinates and the image coordinates. This transformation is determined in geometric camera calibration by solving the unknown parameters of the camera model.
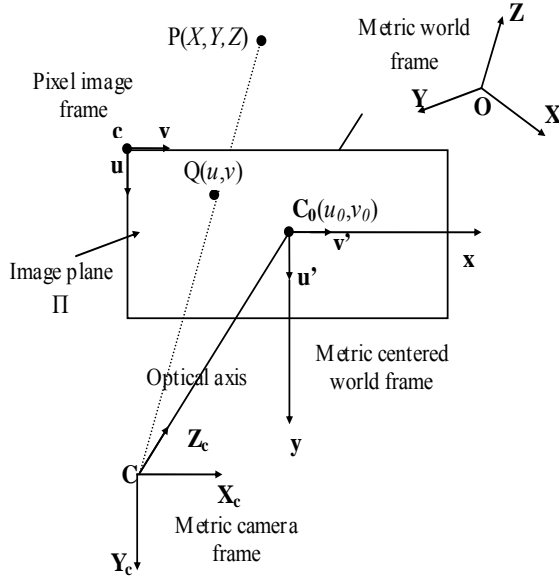

Fig. 1 Block diagram of the model pinhole

A simple pinhole model (Fig. 1) is used for the camera, which defines the basic projective imaging geometry with which the 3D objects are projected onto the 2D image plane. This is an ideal model commonly used in computer graphics and computer vision to capture the imaging geometry [16]. It does not account for certain optical effects (such as non-linear distortions) that are often properties of real cameras but can be ignored in most cases.

The camera can be modeled by a set of intrinsic and extrinsic parameters. The intrinsic parameters are those that define the optical properties of the camera such as the focal length, the aspect ratio of the pixels, and the location of the image centre where the optical axis intersects the image plane. The extrinsic parameters define the position and orientation (pose) of the camera with respect to some external world coordinate system. The transformation that maps the 3D world points into the 2D image coordinates can be characterized by writing the transformation matrices for:

- The rigid transformation matrix defining the camera pose,
- The projection matrix defining the image formation process.

B. Camera Parameters

Fig. 1 shows the different elements of the pinhole model. The transformation made by the camera as a parfait perspective transformation of the camera optical centre C (also the centre of the Metric Camera frame) is considered. The image plane Π (C_0xy or cuv) is parallel to the CX_cY_c plane. The optical axis $C\,Z_c$ (see Fig. 1) pierces the image plane at the principal point $C_0(u_0,v_0)$ called also the image centre. The distance CC_0 is the focal length "f" of the camera.

Let P be an arbitrary 3D point located on the positive side of the Z_c axis and Q its projection on Π. The coordinates of P in the camera frame (C,X_c,Y_c,Z_c) are $[X_c,Y_c,Z_c]^T$, and in the world frame (O,X,Y,Z) the coordinates are $[X, Y, Z]^T$. The coordinates of Q in the image frame are $[u, v]^T$ and $[x, y]^T$ in the metric image frame. The relations between P and Q are:

$$\frac{f}{Z_c} = \frac{x}{X_c} = \frac{y}{Y_c} \qquad (1)$$

If the measure units of x and y axes in the image plane are changed like (scanning):

$$x \leftarrow \frac{u'}{k_u}$$

$$y \leftarrow \frac{v'}{k_v}$$

and the centre (acquisition) is translated:

$$u' \leftarrow u - u_0$$
$$v' \leftarrow v - v_0$$

The following relations are found:

$$x = \frac{u - u_0}{k_u} \qquad (2)$$

$$y = \frac{v - v_0}{k_v} \qquad (3)$$

If $\alpha_u = k_u.f$ and $\alpha_v = k_v.f$ are put, equations (1), (2) and (3) can be written with $Z_c \neq 0$ as linear relation in an homogeneous coordinate [17] :

$$\begin{pmatrix} su \\ sv \\ s \end{pmatrix} = \begin{pmatrix} \alpha_u & 0 & u_0 & 1 \\ 0 & \alpha_v & v_0 & 1 \\ 0 & 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{pmatrix} \qquad (4)$$

The physical signification of the parameters k_u, k_v, α_u, α_v, u_0 and v_0 are:

α_u : horizontal focal length in pixel.

$\dfrac{1}{k_u}$: horizontal dimension of pixel (in meter)

α_v : vertical focal length in pixel.

$\dfrac{1}{k_v}$: vertical dimension of pixel (in meter)

(u_0,v_0) : the pixel coordinate of the image centre.

World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:2, No:9, 2008

Equation (4) can be also written in the following form:

$$U = M.X \qquad (5)$$

The matrix M is cited in literature [17] as a perspective transformation matrix.

The system (5) can be considered in different aspect. If X and M are known, (5) allow us to find the 2D coordinate (u,v) of P. If U and M are known, (5) permit a three-dimensional reconstruction.

The case U and X are known allows calculation of the perspective transformation matrix M.

According the metric world frame, following equations can be written:

$$\begin{pmatrix} X_c \\ Y_c \\ Z_c \end{pmatrix} = R \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} + t \qquad (6)$$

$$\begin{pmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{pmatrix} = \begin{pmatrix} R & t \\ 0 & 1 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \qquad (7)$$

$$\begin{pmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{pmatrix} = \begin{pmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \qquad (8)$$

Where $t=[t_x, t_y, t_z]^T$ describes the translation between the two frames (camera frame and world frame), and R is a 3 by 3 orthonormal rotation matrix which can be defined by the three Euler angles.

The matrix M becomes:

$$\mathbf{M} = \begin{pmatrix} \alpha_u & 0 & u_0 \\ 0 & \alpha_v & v_0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \end{pmatrix} \qquad (9)$$

$$\mathbf{M} = \begin{pmatrix} m_{11} & m_{12} & m_{13} & m_{14} \\ m_{21} & m_{22} & m_{23} & m_{24} \\ m_{31} & m_{32} & m_{33} & m_{34} \end{pmatrix} \qquad (10)$$

III. CALIBRATION

While working in an AR system, it is important to have a reference coordinate system where the locations of the real and virtual objects can be specified. In practice, this coordinate system is set in a location which stays fixed during runtime [18].

The principle of the calibration is to use a calibration grid, or any other calibration object, in which the positions of points marked on it, called checkpoints, are known (Fig. 2).

These points may be wedges [17], points [19], and intersections of lines [20] or any other primitives which can be easily extracted from digital images.
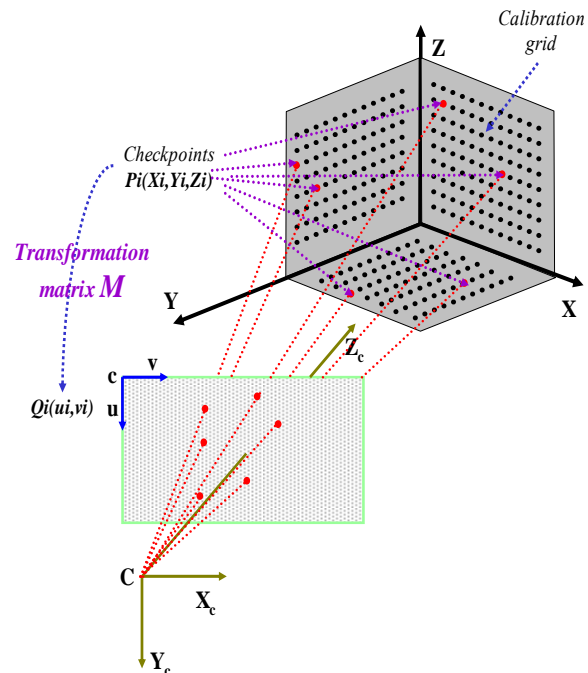

Fig. 2 Diagram of the calibration method

The problem of calibration can be formulated in the following way: given a set of checkpoints P_i which their 3D coordinates (X_i, Y_i, Z_i) are known, determine the parameters of the camera projection function so that their projections are at best the same with the points extracted from images $Q_i (u_i, v_i)$.

The projection (u,v) of each 3D point P(X, Y, Z, 1) on the image is given by:

$$\begin{pmatrix} su \\ sv \\ s \end{pmatrix} = \mathbf{M} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \qquad (11)$$

From (10) and (11) we get:

$$\begin{cases} u = \dfrac{m_{11}X + m_{12}Y + m_{13}Z + m_{14}}{m_{31}X + m_{32}Y + m_{33}Z + m_{34}} \\ v = \dfrac{m_{21}X + m_{22}Y + m_{23}Z + m_{24}}{m_{31}X + m_{32}Y + m_{33}Z + m_{34}} \end{cases} \qquad (12)$$

Each 3D point gives two equations. So, six points are then sufficient to estimate the twelve coefficients of the matrix M. But more than six points can be used if best precision is needed. In this case, the constraint $m_{34} = 1$ is used.

To solve the system (12), it is first transformed it in a linear system as described by (13):

World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:2, No:9, 2008

$$\begin{cases} u_1 = m_{11}X_1 + m_{12}Y_1 + m_{13}Z_1 + m_{14} - m_{31}X_1u_1 - m_{32}Y_1u_1 - m_{33}Z_1u_1 \\ v_1 = m_{21}X_1 + m_{22}Y_1 + m_{23}Z_1 + m_{24} - m_{31}X_1v_1 - m_{32}Y_1v_1 - m_{33}Z_1v_1 \\ . \\ . \\ . \\ u_N = m_{11}X_N + m_{12}Y_N + m_{13}Z_N + m_{14} - m_{31}X_Nu_N - m_{32}Y_Nu_N - m_{33}Z_Nu_N \\ v_N = m_{21}X_N + m_{22}Y_N + m_{23}Z_N + m_{24} - m_{31}X_Nv_N - m_{32}Y_Nv_N - m_{33}Z_Nv_N \end{cases} \quad (13)$$

Then, this system is transformed in a matrix form:

$U = P.V_m$ (eq 14)

$$\begin{pmatrix} u_1 \\ v_1 \\ . \\ . \\ . \\ u_N \\ v_N \end{pmatrix} = \begin{pmatrix} X_1 & Y_1 & Z_1 & 1 & 0 & 0 & 0 & 0 & X_1 & Y_1 & Z_1 \\ 0 & 0 & 0 & 0 & X_1 & Y_1 & Z_1 & 1 & X_1 & Y_1 & Z_1 \\ . & . & . & . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . & . & . & . \\ X_6 & Y_6 & Z_6 & 1 & 0 & 0 & 0 & 0 & X_N & Y_N & Z_N \\ 0 & 0 & 0 & 0 & X_N & Y_N & Z_N & 1 & X_N & Y_N & Z_N \end{pmatrix} \begin{pmatrix} m_{11} \\ m_{12} \\ m_{13} \\ m_{14} \\ m_{21} \\ m_{22} \\ m_{23} \\ m_{24} \\ m_{31} \\ m_{32} \\ m_{33} \end{pmatrix} \quad (14)$$

To find the m_{ij} parameters, the least squares method is used. The following relation is obtained:

$$V_m = (P^T.P)^{-1}. P^T.U \quad (15)$$

The system of equations obtained can be then solved by using a numerical technique as Gauss-Jacobi technique.

IV. TRACKING STRATEGY

Tracking is very important in an AR system. In this work, visual information to locate the position of virtual objects in the real scene (image) is used.

In this case color images are used, then, in the calculation, the three color information (green, red, blue) are also used.

The principle of this method is described in the following algorithm:

Consider a set of "K" images.

1. Calibrate the camera using the first image (Ima$_1$),
   a. Select N (N>=6) checkpoints $Q_i^1(u_i^1, v_i^1)$ on the first image (Ima$_1$) where their 3D coordinate on the world coordinate system are known.
   b. Calculate the transformation matrix of the first image (M$_1$).
   c. Insert virtual objects on Ima$_1$ using M$_1$.
2. for j:=2 to K do,
   a. Find the corresponding checkpoints of $Q_i^{(j-1)}(u_i, v_i)$ from the last image (Ima$_{j-1}$) on the current image Ima$_j$ ($Q_i^j(u_i, v_i)$).
   b. Calibrate the camera using the current image (Ima$_j$)
   c. Calculate the transformation matrix of Ima$_j$ (M$_j$).
   d. Insert virtual objects on Ima$_j$ using M$_j$.

V. EXPERIMENTATION AND RESULTS

To calibrate the camera, a calibration grid is used and a couple of image is taken. First, N (N=6) points are selected on the first image where the 3D coordinates in the world frame are known, and then their corresponding points in the second image are searched. The "Table I" shows an example of 3D selected points, their projection on the first image and their corresponding points in the second one. These points are represented in Fig. 3 and Fig. 4 by red crosses.

TABLE I
CONTROL POINTS SELECTED FOR THE CALIBRATION

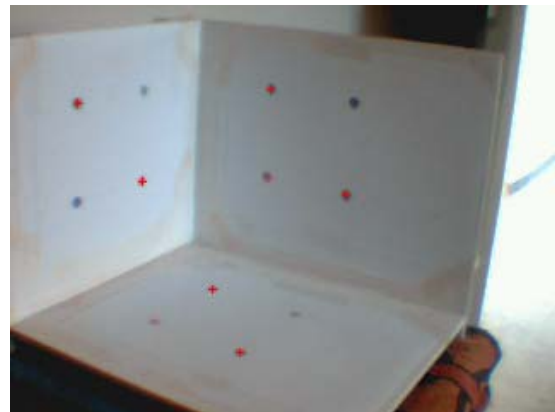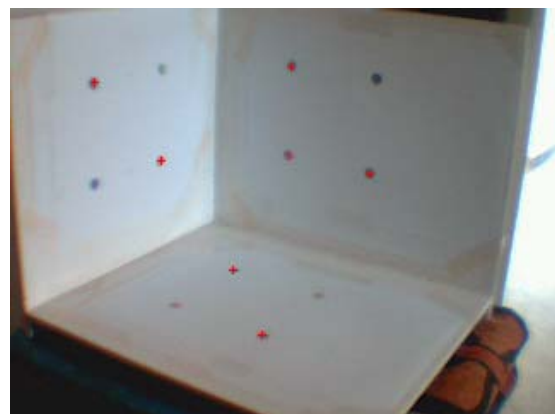| Check-points	3D coordinates (cm)			2D coordinate (Pixel) Ima$_1$ | Ima$_2$	
	X_1	Y_1	Z_1	u_1 / v_1	u_2 / v_2
P1	10	20	0	197/111	211/97
P2	20	10	0	152/49	164/34
P3	0	10	10	77/103	88/89
P4	0	20	20	39/57	49/43
P5	10	0	10	118/166	130/153
P6	20	0	20	134/203	147/191

Fig. 3 Selection of the checkpoints (Ima$_1$)

Fig. 4 Corresponding points found on the second image

After the selection, the transformation matrices are calculated. For the first image:

World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:2, No:9, 2008

$$M_1 = \begin{pmatrix} 3,1013 & -0,2073 & -3,6040 & 106,4857 \\ 0,2278 & -4,8991 & 0,1760 & 138,6932 \\ -0,0059 & -0,0032 & -0,0081 & 1 \end{pmatrix}$$

and for the second one:

$$M_2 = \begin{pmatrix} 2,9928 & -0,2760 & -3,7390 & 118,0704 \\ 0,2605 & -4,8770 & 0,3110 & 124,4680 \\ -0,0067 & -0,0032 & -0,0082 & 1 \end{pmatrix}$$

Using these matrices, an augmentation is applied. It consists of:

- Drawing lines to show the orthogonal frame on the calibration grid.
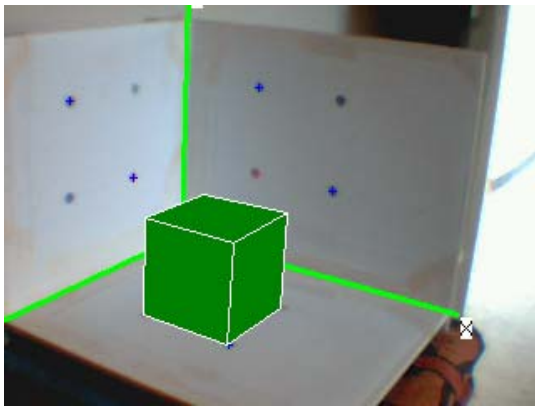- Labeling the axis (text augmentation).
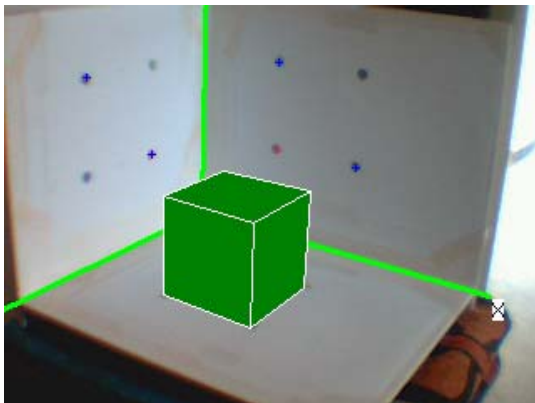- Drawing a 3D object (Cube)

Fig. 5 Image augmentation (Ima$_1$)

Fig. 6 Image augmentation (Ima$_2$)

The obtained results are shown in the Fig. 5 for the first example and Fig. 6 for the second one.

Lines and texts augmentation permit locating and showing the world frame axis and their indexation (Fig. 5 and Fig. 6). Concerning the 3D object augmentation, the virtual object is placed in the desirable position when a small error can be noticed.

For the second example, other virtual object more complicate than the first one and a sequence of eight images are used.

In this example, nine points (checkpoints) are selected. After calculating the transformation matrix (M_1), an augmentation for the first image by the virtual object is done. For the other images, the corresponding checkpoints are searched to calculate the transformation matrix (M_i) for each image. Finally, the virtual object into the sequence using the corresponding matrix for each image "M_i" is inserted. Following figures (Fig. 7) shows the obtained results.
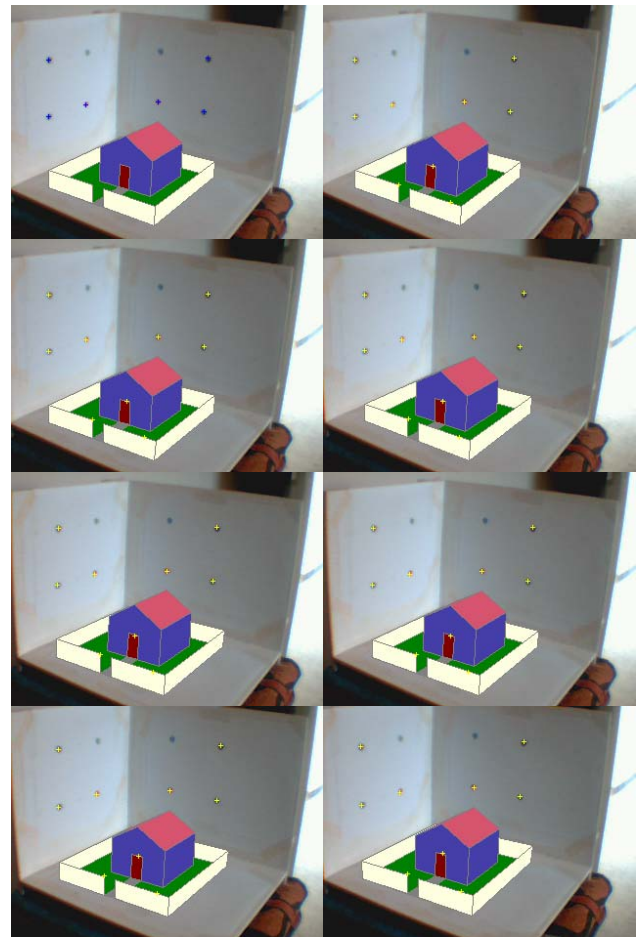

Fig. 7 Sequence of image augmentation

The checkpoints selected are represented by blue crosses in the first image of the Fig. 7. The checkpoints finding by correlation are represented by yellow crosses on the other images.

VI. CONCLUSION

In this paper, a camera calibration procedure and tracking for augmented reality systems are presented. A method based on the following steps is proposed:

- Selection of N (N>=6) points (checkpoints) on the first image which their 3D coordinate on the world frame are known.
- Search the checkpoints on the other images using the correlation technique.

World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:2, No:9, 2008

- Calculate the transformation matrix for each image.
- Insertion of the virtual object.

The obtained results using this method are acceptable and the technique used is fast. In the future work, this method will be applied on a real time system using a camera.

REFERENCES

[1] R. Azuma, "A Survey of Augmented Reality". Presence: Teleoperators and Virtual Environment, 6(4): pp. 355-385, 1997.
[2] R. Azuma, Y. Baillot,R. Behringer, S. Feiner, and S. Julier, "Recent advances in augmented reality", IEEE Computer Graphics and Applications, 21(6), pp. 34-47, 2001.
[3] D. Koller, G. Klinker, E. Rose, D. Breen, Ross Whitaker, and M. Tuceryan, "Real-time Vision-Based Camera Tracking for Augmented Reality Applications", VRST, Lausanne, Switzerland, pp. 87–94, September, 1997.
[4] Y. Genc M. Tuceryan, and N. Navab, "Practical Solutions for Calibration of Optical See-Through Devices", International Symposium on Mixed and Augmented Reality, pp. 169-175, 2002
[5] E. McGarrity, and M. Tuceryan, "A Method for Calibrating See-through Head-mounted Displays for AR", 2nd International Workshop on Augmented Reality, San Francisco, CA, Oct 20-21, 1999.
[6] R. Y. Tsai, "A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off the- shelf TV cameras and lenses", IEEE Journal of Robotics and Automation RA, 3(4) pp. 323-344, 1987.
[7] J. Heikkilä, and O. Silvén, "A Four-step Camera Calibration Procedure with Implicit Image Correction", IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 1106-1112, 1997.
[8] A. Janin, D. Mizell, and T. Caudell, "Calibration of headmounted displays for augmented reality applications", In Proceedings of the Virtual Reality Annual International Symposium, pp. 246–255, September 1993.
[9] R. Tsai, "An efficient and accurate camera calibration technique for 3D machine vision", IEEE Proceedings CVPR, pp. 364-374, June 1986.
[10] M. Tuceryan, M., Greer, D., Whitaker, R., Breen, D., Crampton, C., Rose, E., and Ahlers, K., "Calibration Requirements and Procedures for Augmented Reality", IEEE Trans. On Visualization and Computer Graphics, pp. 255-273, September 1995.
[11] O.D. Faugeras and G. Toscani. "The calibration problem for stereo", In Proceedings of CVPR, pp. 15-20, 1986.
[12] Z. Zhang, "Camera calibration with one-dimensional objects", In Proc. 7th European Conference on Computer Vision, (IV), pp. 161–174, 2002.
[13] J. Weng, P. Cohen, and M. Herniou. "Camera calibration with distortion models and accuracy evaluation" IEEE Transactions on Pattern Analysis and Machine Intelligence, 14(10), pp. 965-980, October 1992.
[14] R. Y. Tsay, "A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses", IEEE Journal of robotics and automation, 3(4), pp. 323-342, August 2002.
[15] E. L. Hall, J. B. K. Tio, C. A. McPherson, and F. A. Sadjadi, "Measuring curved surfaces for robot vision", Computer Journal, pp. 42–54, December 1982.
[16] O.D. Faugeras, Three-Dimensional Computer Vision, MIT Press, 1993.
[17] G. Toscani, Système de calibration et perception du mouvement en vision artificielle, Doctorate thesis. PARIS-SUD University, France, 1987.
[18] E. McGarrity, Y. Genc, N. Navab, M. Tuceryan, and C. Owen, "Evaluation of Calibration for Optical See-Through Augmented Reality Systems", In Proceedings of Euroimage International Conference on Augmented, Virtual Environments and Three-Dimensional Imaging, Mykonos, Greece, pp. 18–21, May 30–June 1, 2001.
[19] R. G.Willson, Modeling and Calibration of Automated Zoom Lenses, PhD thesis, Department of Electrical and Computer Engineering, Carnegie Mellon University, 1994.
[20] P. Brand, R. Mohr and P. Bobet, "Distorsions optiques : correction dans un modèle projectif", technical rapport N° 1933, LIFIA–INRIA Rhône-Alpes, 1993.