Reconstructing while registering: a novel approach for markerless augmented reality

Gilles Simon, Marie-Odile Berger

▶ To cite this version:

HAL Id: inria-00099442
https://hal.inria.fr/inria-00099442

Submitted on 26 Sep 2006

Reconstructing while registering: a novel approach for markerless augmented reality

Gilles Simon and Marie-Odile Berger

LORIA - UHP Nancy I - INRIA Lorraine, BP 239, 54506 Vandœuvre-les-Nancy Cedex, France
{gsimon,berger}@loria.fr

Abstract

This paper addresses the registration problem for un-prepared multi-planar scenes. An interactive process is proposed to get accurate results using nothing else than the texture information of the planes. In particular, the classical preparation steps (camera calibration, scene acquisition) are greatly simplified, since included in the on-line registration process. Some results are shown on indoor and outdoor scenes. Videos available at url `http://www.loria.fr/~gsimon/Ismar`.

1 Introduction

The objective of Augmented Reality (AR) is to add virtual objects to real video sequences, allowing computer-generated objects to be overlaid on the video in such a manner as to appear part of the viewed 3D scene. Applications of this concept concern interior design, architectural design, computer-aided repair and learning systems, medicine, and special effects for broadcast industry [1].

While there are several problems in building AR systems, one of the most basic challenge to overcome is the registration problem: the objects in the real and the virtual world must be properly aligned with respect to each other or the illusion that the two worlds coexist will be compromised. It is therefore essential to determine accurately the position of the cameras.

In this paper, we address the registration problem for interactive AR applications. Though this problem has received a lot of attention in the AR community, only little works have been devoted to real-time algorithms for unprepared environments. In the present paper we propose a vision-based registration method designed for multi-planar environments (like indoor or urban environments). The novelty of our method resides in that it does not require any marker in the scene and is therefore also well-suited for outdoor environments. Moreover, the traditional preparation steps (camera calibration, scene acquisition) are largely simplified and performed while the camera is moving.

2 Background

The registration may be achieved using different kinds of sensors: mechanic or magnetic sensors, GPS, compass and so on. However, the use of sensors has proven to be constraining in practice (extensive calibration, restricted user displacements, perturbations from the environment, poor accuracy ...). By contrast, vision-based registration does not require any instrument except the acquiring camera. Moreover, the augmentation results are generally more accurate than the results obtained from sensors (except for abrupt motions), as they are directly computed from features extracted from the images to be augmented.

Vision-based techniques rely on the identification in the images of features of an object model. Examples include point features [10, 11, 14], edges [6] or curves [12]. Artificial markers can help generate such features in the images [10, 14]. Other works extract them from the natural structure of the scene [6, 11, 16, 12]. Pose estimation techniques can then be used to estimate the camera position [2, 12] on each single image. However, it is commonly true that few features are available for registration. As a result, the tracking suffers from high-frequency jitter. More importantly, such methods require significant manual intervention to construct the model (especially outdoors), and the camera has to be calibrated off-line.

By contrast, using planar structures of the scene allows to track the camera with minimal intervention [15, 8, 13]. In this paper, we present a novel approach for registering unprepared multi-planar environments. When one or several planes are visible in the scene, it is very simple to indicate these planes and calibrate the camera in an on-line process. As soon as one plane is designed, it can be registered using its texture information, and a virtual object can be added into the scene. Other planes may then been added as the camera is moving, making the registration process more accurate.

The paper is organized as follows: the multi-planar registration method is presented in section 3 ; section 4 describes the interactive scene-and-camera reconstruction process ; finally, experimental results are shown in section 5.

3 The registration method

The registration method is an extension of [13] to the case where more than one plane are visible in the scene. The projection matrices are computed from sets of key-points belonging to the planes, yielding a fast, sequential, robust and accurate registration process.

3.1 Single-plane registration

A single-plane temporal registration system was described in [13]. In this section, we remind how to compute a 3×4 projection matrix P^i in image i, from P^{i-1} in image $i-1$ and a planar homography H^i between these images.

We consider the pinhole camera model, which associates a point $\mathbf{x}^i$ in image i to a point $\mathbf{X}$ in the scene:

$$\mathbf{x}^i = \mathrm{P}^i \mathbf{X} = \mathrm{K} \left(\ \mathrm{R}^i | \mathbf{t}^i \ \right) \mathbf{X} \qquad (1)$$

The matrix K represents the internal calibration parameters of the camera. In this section, these parameters are supposed to be known. $\left(\ \mathrm{R}^i | \mathbf{t}^i \ \right)$ is the viewpoint matrix to be estimated.

Let us now restrict $\mathbf{X}$ to lie on a plane Π and suppose that we know the associated planar homography H^i between image i and image $i-1$. The following relation is valid for all points on plane Π (Fig. 1):

$$\mathbf{x}^i \propto \mathrm{H}^i \mathbf{x}^{i-1}, \qquad (2)$$

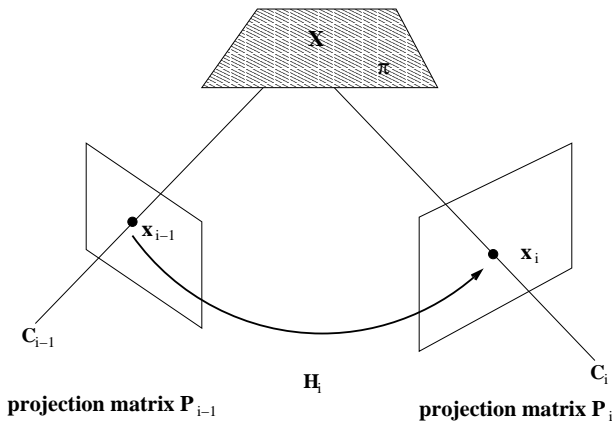where $\propto$ denotes an equality up to a multiplicative factor.

Figure 1. Homography induced by a plane.

Let M be a transformation matrix between $\mathcal{R}_0$, a coordinate system where the equation of Π is $z = 0$, and the world-coordinate system. For all points $\mathbf{X}$ on plane Π, we have:

$$\mathbf{x}^i = \mathrm{P}^i \mathbf{X} = \mathrm{P}^i \mathrm{M} \begin{pmatrix} \mathrm{X} \\ \mathrm{Y} \\ 0 \\ 1 \end{pmatrix} = \langle \mathrm{P}^i \mathrm{M} \rangle \begin{pmatrix} \mathrm{X} \\ \mathrm{Y} \\ 1 \end{pmatrix} \qquad (3)$$

where $\langle \mathrm{A} \rangle$ denotes the matrix A deprived of its third column. $\langle \mathrm{P}^i \mathrm{M} \rangle$ is invertible unless plane Π goes through the origin of the camera.

As a result, combining equations (2) and (3) yields:

$$\langle \mathrm{P}^i \mathrm{M} \rangle \propto \mathrm{H}^i \langle \mathrm{P}^{i-1} \mathrm{M} \rangle \qquad (4)$$

Depriving $\mathrm{P}^i \mathrm{M}$ of its third column does not prevent from recovering the full viewpoint parameters. Indeed, knowing $\langle \mathrm{P}^i \mathrm{M} \rangle$ from equation (4), as well as the internal parameters K leads to:

$$\mathrm{K}^{-1} \langle \mathrm{P}^i \mathrm{M} \rangle \propto \left(\ \mathbf{r}_1 \mathbf{r}_2 \mathbf{t} \ \right) \qquad (5)$$

where $\mathbf{r}_1$ and $\mathbf{r}_2$ are orthonormal vectors. The third column for the rotation matrix of the viewpoint is merely given by $\mathbf{r}_3 = \mathbf{r}_1 \times \mathbf{r}_2$. In practice, the orthonormality conditions are never perfectly met, and renormalization must be applied ($\mathbf{r}_2 = \mathbf{r}_2 / \|\mathbf{r}_2\|, \mathbf{r}_3 = \mathbf{r}_1 \times \mathbf{r}_2 / \|\mathbf{r}_1 \times \mathbf{r}_2\|, \mathbf{r}_1 = \mathbf{r}_2 \times \mathbf{r}_3$).

3.2 Multi-plane registration

In this section, we suppose that $n > 1$ homographies H_p^i are known for n planes Π_p. From equation (4), we get for each plane:

$$\mathrm{H}_p^i \propto \langle \mathrm{P}^i \mathrm{M}_p \rangle \langle \mathrm{P}^{i-1} \mathrm{M}_p \rangle^{-1}. \qquad (6)$$

For each correspondence $\mathbf{x}_j \leftrightarrow \mathbf{x}'_j$ between image $i-1$ and image i, we have

$$\mathbf{x}'_j \propto \mathrm{H}_p^i \mathbf{x}_j, \qquad (7)$$

where p is the subscript of the plane containing the related 3D point $\mathbf{X}_j$.

Combining equations (6) and (7), and considering that $\langle \mathrm{P}^i \mathrm{M}_p \rangle = \mathrm{P}^i \langle \mathrm{M}_p \rangle$, leads to:

$$\mathbf{x}'_j \propto \mathrm{P}^i \langle \mathrm{M}_p \rangle \langle \mathrm{P}^{i-1} \mathrm{M}_p \rangle^{-1} \mathbf{x}_j,$$

$$\Leftrightarrow \mathbf{x}'_j \propto \begin{pmatrix} \mathbf{p}^{1\top} \mathrm{Q}_p \mathbf{x}_j \\ \mathbf{p}^{2\top} \mathrm{Q}_p \mathbf{x}_j \\ \mathbf{p}^{3\top} \mathrm{Q}_p \mathbf{x}_j \end{pmatrix}, \qquad (8)$$

where $\mathbf{p}^{k\top}$ is the kth row of matrix P^i, and $\mathrm{Q}_p = \langle \mathrm{M}_p \rangle \langle \mathrm{P}^{i-1} \mathrm{M}_p \rangle^{-1}$. Writing equations (8) in terms of cross

products gives

$$\begin{pmatrix} y'_j \mathbf{p}^{3\top} \mathbf{Q}_p \mathbf{x}_j - w'_j \mathbf{p}^{2\top} \mathbf{Q}_p \mathbf{x}_j \\ w'_j \mathbf{p}^{1\top} \mathbf{Q}_p \mathbf{x}_j - x'_j \mathbf{p}^{3\top} \mathbf{Q}_p \mathbf{x}_j \\ x'_j \mathbf{p}^{2\top} \mathbf{Q}_p \mathbf{x}_j - y'_j \mathbf{p}^{1\top} \mathbf{Q}_p \mathbf{x}_j \end{pmatrix} = \mathbf{0},$$

where $\mathbf{x}'_j = \begin{pmatrix} x'_j & y'_j & w'_j \end{pmatrix}^\top$. Finally, as $\mathbf{p}^{k\top} \mathbf{Q}_p \mathbf{x}_j = \mathbf{x}_j^\top \mathbf{Q}_p^\top \mathbf{p}^k$, we get the linear system of equations:

$$\begin{pmatrix} \mathbf{0}^\top & -w'_j \mathbf{x}_j^\top \mathbf{Q}_p^\top & y'_j \mathbf{x}_j^\top \mathbf{Q}_p^\top \\ w'_j \mathbf{x}_j^\top \mathbf{Q}_p^\top & \mathbf{0}^\top & -x'_j \mathbf{x}_j^\top \mathbf{Q}_p^\top \\ -y'_j \mathbf{x}_j^\top \mathbf{Q}_p & x'_j \mathbf{x}_j^\top \mathbf{Q}_p & \mathbf{0}^\top \end{pmatrix} \begin{pmatrix} \mathbf{p}^1 \\ \mathbf{p}^2 \\ \mathbf{p}^3 \end{pmatrix} = \mathbf{0}. \quad (9)$$

Although there are three equations, only two of them are linearly independent (we omit the third equation). Therefore, each correspondence that belongs to one plane Π_p gives two equations in the entries of $\mathbf{P}^i$. Finally, we obtain a linear system of equations having the form $\mathbf{Ap} = \mathbf{0}$, where $\mathbf{A}$ is a $2m \times 12$ matrix, m being the number of point correspondences.

This system may be solved very quickly using its singular value decomposition. However, no profit is taken from the a priori knowledge of matrix $\mathbf{K}$ and that $\mathbf{R}$ is a rotation matrix. To overcome this problem, we suppose that the camera rotation between two images $\Delta \mathbf{R}^i = \mathbf{R}^i (\mathbf{R}^{i-1})^t$ is small. A first order approximation of this rotation can thus be performed. We obtain a linear expression of the entries of $\mathbf{P}^i$ in the coefficients $t_x, t_y, t_z, \Delta\alpha, \Delta\beta, \Delta\gamma$:

$$\mathbf{P}^i \approx \mathbf{K} \left[\begin{pmatrix} 1 & -\Delta\alpha & \Delta\beta \\ \Delta\alpha & 1 & -\Delta\gamma \\ -\Delta\beta & \Delta\gamma & 1 \end{pmatrix} \mathbf{R}^{i-1} \quad \begin{matrix} t_x \\ t_y \\ t_z \end{matrix} \right]. \quad (10)$$

Combining equations (9) and (10) provides a linear system in the viewpoint parameters.

3.3   Implementation

The implementation of the different steps of the algorithm is detailed below. Its leads to an approximate processing rate of 16 frames per second on a Pentium III, 900 MHz processor.

Extraction and matching of key-points

The Harris detector [7] is used to detect key-points and the matching process is performed in a classical way. The normalized cross correlation score [17] is computed between each key-point in image i and all key-points lying in its neighbourhood in image $i - 1$. We only retain the set of candidate matches presenting with the maximum score, provided that this score is greater than a predefined threshold (typically, $s = 0.8$).

Robust estimation of the homography

The scene is modelled as a set of 3D planar polygons $\mathcal{L}_p (1 \le p \le n)$. Let $\mathcal{C}_p^i$ be the set of correspondences $\mathbf{x}_j \leftrightarrow \mathbf{x}'_j$ such that $\mathbf{x}_j$ belongs to the projection of $\mathcal{L}_p$ in image $i - 1$ using $\mathbf{P}^{i-1}$. Robust estimation of the homography $\mathbf{H}_p^i$ can be achieved using the RANSAC paradigm [4]: random samples of four pairs are selected in $\mathcal{C}_p^i$ and the corresponding homography is computed. The homography is tested against all the correspondences: the set of inliers is the set of pairs $\mathbf{x}_j \leftrightarrow \mathbf{x}'_j$ for which the distance between $\mathbf{x}'_j$ and $\mathbf{H}_p^i \mathbf{x}_j$ is below a predefined threshold $r = 1.25$. Finally, the homography with the largest consensus set is chosen.

In the following, the set of inliers computed for each planar polygon $\mathcal{L}_p$ is denoted $\mathcal{I}_p^i$.

Computation of $\mathbf{P}^i$

The projective matrix $\mathbf{P}^i$ is computed using the full set of inliers $\bigcup_p \mathcal{I}_p^i$ stemming from the planar structures. Once this computation is achieved, the homographies are updated using equation (6). Then, new inliers are detected and used to update $\mathbf{P}^i$. This loop is continued until the number of inliers does not change. This refinement stage generally converges in 2 or 3 iterations.

3.4   Evaluation

Our registration method has been evaluated on a 98-frames sequence of a calibration target. The camera is moving around the target, roughly pointing the intersection of the three planes (see Fig. 2). The initial matrix $\mathbf{P}^0$ is obtained using the method proposed in [3], from 3D/2D correspondences of points belonging to the target.
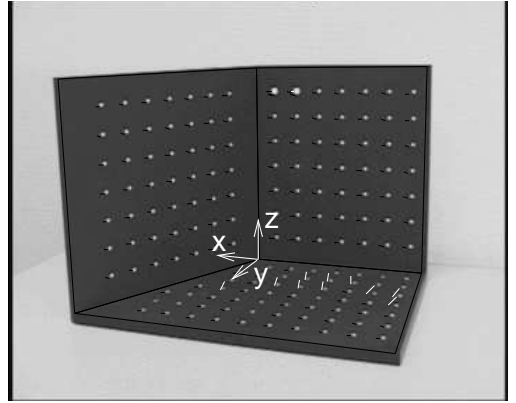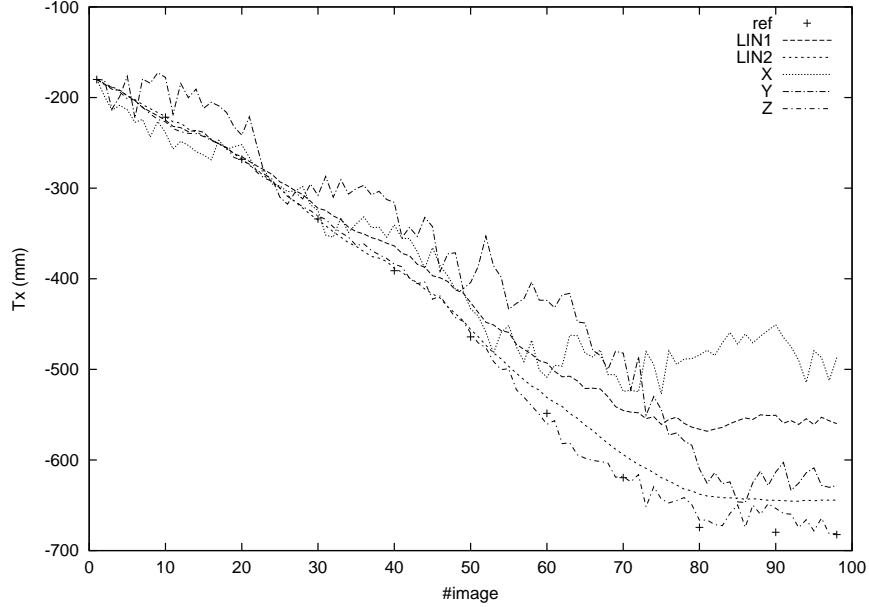

Figure 2. One image of the target sequence.

Fig. 3 shows the temporal evolution of one viewpoint parameter (x translation), for the single-plane and the multiplane registration methods. Actual values (represented by

Figure 3. Temporal evolution of t_x for the single-plane and the multi-plane registration methods.

crosses in Fig. 3), are computed every ten images, using classical calibration [3] from points on the target. The curve called LIN1 shows the results obtained by solving equations (9), using the three planes of the target. These results are not accurate, due to the fact that no constraint is set on the rotation or the intrinsic parameters. More accurate results were obtained by solving equations (10) (curve LIN2). These results are relatively close to the actual values, but they slowly diverge from them: this expresses an accumulation error due to the successive approximations of the rotation matrix. However, the pose estimation error obtained at the end of the sequence is only 6.3 cm for a distance target - camera equal to 127 cm. This error is almost not perceptible with regard to the projection of the target.

Fig. 3 also shows the results obtained by using only one plane of the target, and the method described in section 3.1: the vertical left plane (curve X), the vertical right plane (curve Y) or the horizontal plane (curve Z). These results are much more irregular than the results obtained from three planes, which illustrate our contribution with regard to [13]. It is interesting to note that the most accurate results were obtained when the horizontal plane was used. This is due to the fact that the depth information is better represented by that plane, particularly in the second half of the sequence.

4 Reconstructing while registering

Section 3 supposed that the internal parameters of the camera and the planar structures of the scene are known.

This section shows how these parameters can be recovered interactively while the camera is moving.

The algorithm of this process is summarized in Fig. 4, and described in more details below. A typical session consists on defining a reference plane and clicking four points on that plane as the camera is moving. This is enough to augment the scene, but the user can also add other planes (perpendicular to the reference plane) in order to increase the accuracy of the registration, and also his displacement possibilities.

4.1 Initialization

If the intrinsic parameters of the camera are unknown, they can be approximated by using two sets of parallel lines [9]. Assuming skew is 0, aspect ratio is 1 and principal point is the center of the image, the focal length f can be computed using equation $x_v x_w + y_v y_w + f^2 = 0$, where $\mathbf{v} = (x_v, y_v, 1)$ and $\mathbf{w} = (x_w, y_w, 1)$ are the vanishing points of the sets of lines.

Moreover, it is shown in [13] how the position and orientation of the camera can be retrieved from known intrinsic parameters and the pixel coordinates of four points on a world rectangle, whose aspect ratio is unknown. As a rectangle also provides two sets of parallel lines, it is therefore sufficient to click four points on a rectangle to compute the projection matrix $\mathbf{P}^0$, assuming the rectangle is on the plane $z = 0$. These four points are called *calibration points* in Fig. 4.

Once $\mathbf{P}^0$ is known, the camera can be tracked if at least
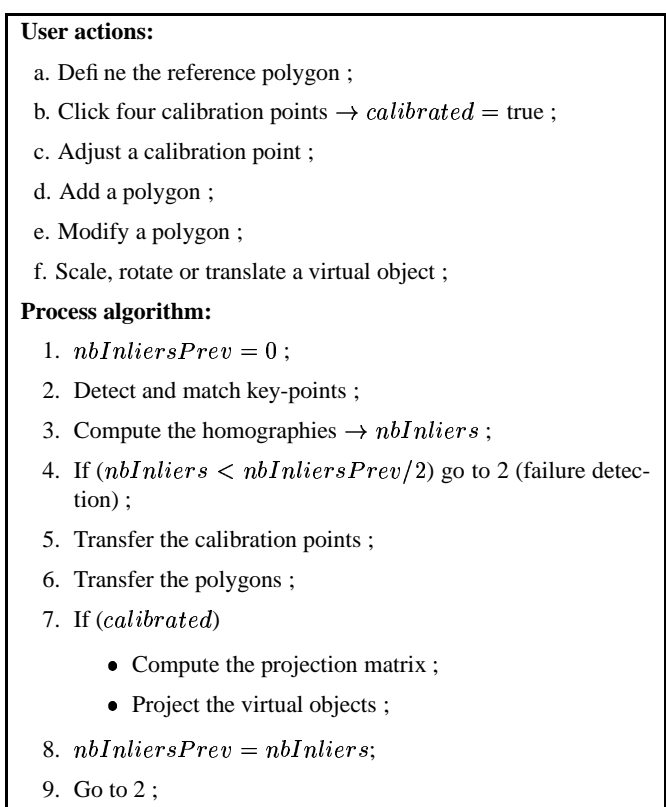
Figure 4. Reconstructing while registering: overview of the algorithm.

one plane is specified. Predefined virtual objects can also be projected onto the reference plane and the user can scale and rotate these objects, or shift them on the reference plane.

In practice, the user first outlines a planar region in the image (called *reference polygon* in Fig. 4), and then specifies four points on the plane defined by that region (*reference plane*). This enables the user to click the calibration points while the camera is moving. Indeed, any point belonging to the reference plane can be transferred along the sequence, using the related homography (Fig.5.a-b). In the same way, the calibration points can be moved once they have been defined in order to visually refine the calibration result (see Fig. 5.c).

This last possibility may also be very useful in the case where no rectangle is visible is the scene: the user can try to "imagine" the rectangle by approximately clicking four points on the plane and adjusting these points until the perspective of the virtual object fits the perspective of the real scene.

4.2 Definition of the planar surfaces

Once the reference plane is defined, it is very simple to add other planes, assuming they are perpendicular to the reference plane. This is commonly true in indoor or urban environments, when the reference plane is the ground and the other planes are vertical walls.

The position and orientation of a plane perpendicular to the reference plane are obtained by back-projecting a 2D line onto the plane $z = 0$. The new plane is the vertical plane passing through the back-projected line.

In practice, the user specifies $v \geq 3$ vertices of a polygon. The first two vertices are used to compute the equation of the plane: they must be common to the new plane and the reference plane. The other vertices are used to outline the region to be tracked (Fig. 5.d-e).

As the user travels through the scene, existing planes may be extended by moving their vertices (Fig. 5.f), or new planes may be added, making the registration process more accurate, and increasing the user displacement possibilities.

4.3 Failure detection

If an abrupt motion occurs during the session, the matching process may fail because of a too large difference between consecutive frames. It is possible to detect these failures by counting the total number of inliers with regard to the computed homographies. If this number decreases rapidly (typically is divided by 2 or more), a failure is detected.

In that case, the projection matrix is not updated until the number of inliers becomes large enough again. This means that the user has to bring the camera back to a position close to its position before failure. Projecting the structure of the scene with the last computed projection matrix may help the user perform this task (Fig. 6).

Note that this method may also be used to start a session from a prepared scene. An experimented user may calibrate the camera, reconstruct the scene, add virtual objects and stop the process. Another person would then be able to immediately see the augmented scene by just starting to look approximately in the same direction as the one which was used at the end of the preparing process.

4.4 Known scene or camera

If the structure of the scene or the camera intrinsic parameters are known *a priori*, this knowledge can easily be included in our system:

- if the intrinsic parameters are known but not the scene, the four calibration points are used to compute the pose only ;
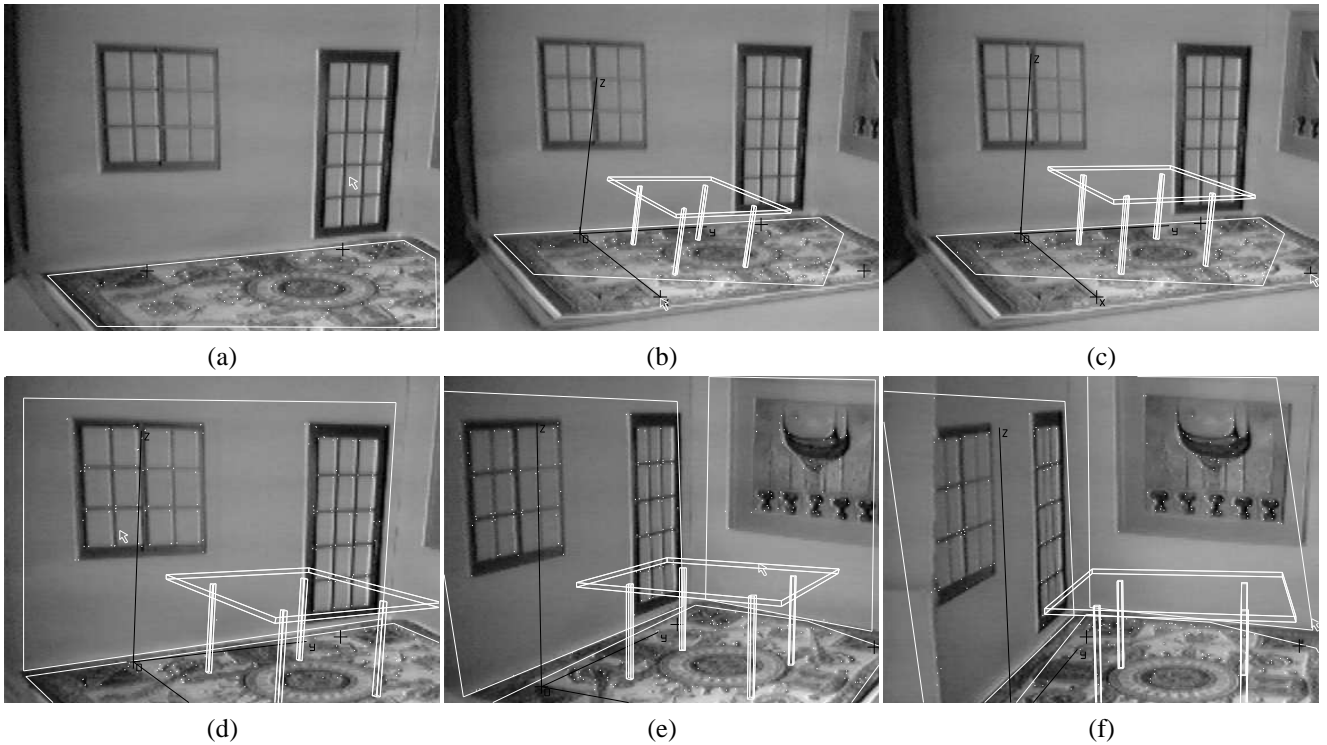
Figure 5. Reconstructing while registering. White segments connect the key-points in the current frame to their homologous in the previous frame. (a)-(b) The points are clicked while the camera is moving. (c) Calibration points can be adjusted to refine the calibration result. (d)-(f) New polygons can be added or existing polygons can be extended as the camera goes over the room.

- if the scene is known but not the intrinsic parameters, matching six points between the model and the image provides the first projection matrix ;

- if both the scene and the intrinsic parameters are known, two methods may be used to initialize the process: matching four points between the model and the image, or using the "prepared scene" principle exposed in 4.3.

This is of particular interest when the application requires a high level of accuracy, or concerns tasks which are repeated regularly in the same conditions. This also allows to handle planes which are not necessary perpendicular to the reference plane.

5 Experimental results

In this section, the on-line process is demonstrated on a miniature sequence. For practical reasons, the other two sequences were treated off-line. The second sequence proves that little texture information is enough to

get accurate results. The third sequence shows that our method is also well-suited for outdoor environments. All the videos presented below are available on our web site: http://www.loria.fr/~gsimon/Ismar.

5.1 A miniature scene

A miniature scene was built to illustrate the on-line process: three sheets of paper showing drawings of indoor elements were glued onto the faces of an uncalibrated box. A simple web-cam connected to a Pentium III, 900 MHz computer was used to shoot the scene.

Snapshots of the sequence are shown in Fig. 5 and 6. The user first designates the horizontal plane and a rectangle on it. As soon as the fourth calibration point is specified, a wire-frame table is added on the plane. This augmentation is refined when the user moves one calibration point. Finally, two other planes are defined interactively, and an abrupt motion is simulated in order to assess a failure detection. Note that a jittering effect is observed until a second plane is defined, which is consistent with Fig. 3.

This sequence proves that our method is relevant and

6
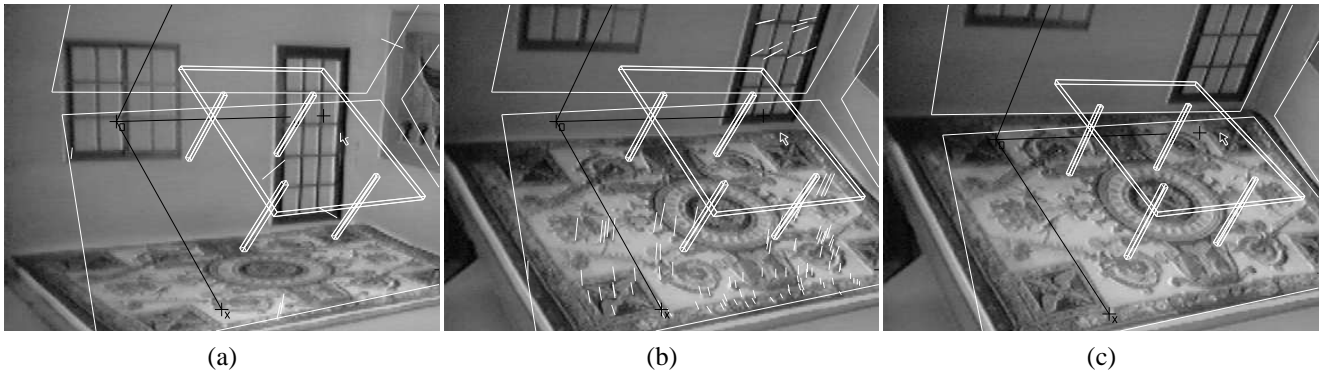
Figure 6. Failure detection. (a) Due to an abrupt motion, the registration failed. (b) The user brings the camera back to a position close to its position before failure. (c) The scene is registered again.

leads very quickly to augmentation results, without any *a priori* knowledge about the camera or the scene (except that the walls are vertical).

5.2 An indoor scene

Results were also obtained on a 200-frames indoor sequence, shot in the basement of our laboratory. These experiments were conducted within a European project for e-commerce applications (ARIS project). One of the goal is to allow the customer to visualize the furniture proposed for selling in the future environment. An ordinary poster was put on the ground to help the calibration process.

This sequence is particularly difficult to treat because the scene is very poorly textured (hardly a few stains on the ground and walls). Moreover, the camera motion is relatively fast in the second half of the sequence (up to 20 pixels of disparity between two images), and some images are blurred.

Despite these difficulties, the system succeeded in registering the two or three planes that were visible during the shot. Fig. 7 shows the matching result and the projection of a cube after registration, in four images of the sequence. A final composition is shown in Fig. 8.
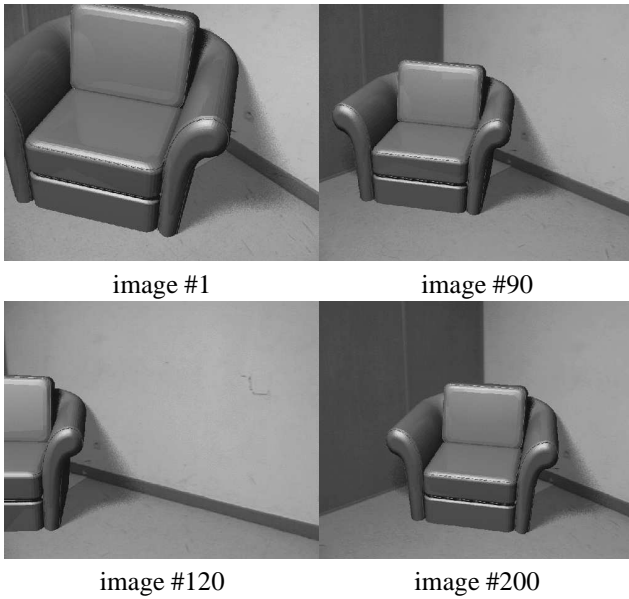
5.3 An outdoor scene

The effectiveness of our approach was also demonstrated on an 600-frames outdoor sequence. The campus of our University was shot by a pedestrian walking with a hand-held camera. Our aim was to add annotations in order to help the visitor find some departments in our university: the mathematic research center and the biological research center (BIO RC and MATH RC). A Maya statue was also added to bring an exotic impression to our campus (Fig. 10).

image #1   image #90

image #145   image #200
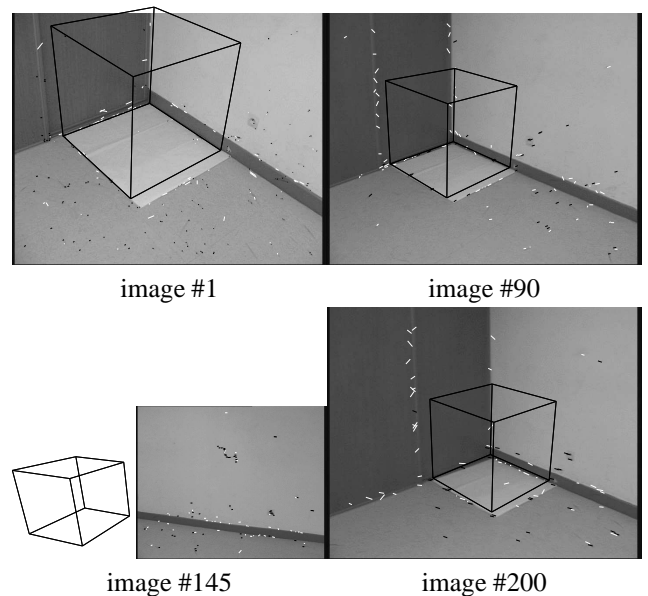
Figure 7. The registration process for the indoor sequence. Correct matches are drawn in black, outliers in white.

In this application the ground plane and two façades were used to compute the viewpoint. Fig. 9 exhibits the points which were tracked in each planar structure to recover the homography, and the projection of a wire-frame model of a part of the scene. Note that this model was build using the same principle as what was described in section 4.2: the orientation and position of the wireframe boxes were defined by clicking an edge on the horizontal plane, and scaled and translated along the vertical axis in order to fit their image counterpart.

7

image #1   image #90

image #120   image #200

Figure 8. Augmented sequence: a sofa has been added.

image #1   image #116

image #372   image #455

Figure 10. Some views of the augmented campus. Annotations have been added on the mathematical research center and the biological research center as well as a maya statue.
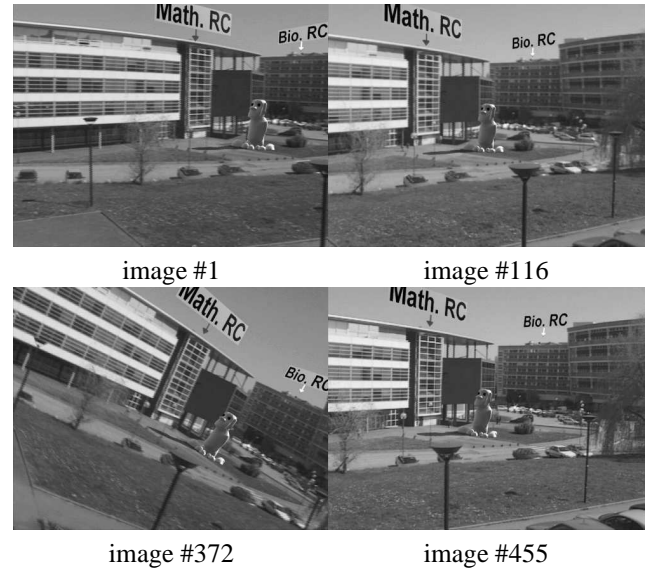
Snapshots of the augmented scene are shown in Fig. 10. The visual impression is good and the annotations seem to be part of the scene. However, if we look carefully at the full video available at out web site, we notice that the Maya statue seems to slide slightly along the ground at instants.

6 Discussion

A novel approach has been proposed for registering multi-planar surfaces. This approach leads very quickly to augmentation results, without any *a priori* knowledge about the scene or the camera. We proved that this method is applicable to a wide range of environments both for complex indoor and outdoor urban scenes. The accuracy obtained on the viewpoint is sufficient for numerous applications: it especially allows annotations to be displayed, providing the relevant, critical information for a user's context.

We now plan to further investigate how this framework can be improved to be able to handle longer sequences. Indeed, the method may progressively diverge because of successive approximations. This problem may be shaped by considering homography with more distant images, or by performing a bundle adjustment on a small number of images (the last five images for example, in the spirit of [5]). Of course, a hybrid system could also increase robustness and avoid drift by taking advantage of a partial 3D knowledge on the scene.

Finally, the possibility of automatically recovering the

multi-planar structure of the scene is currently under study. This will be of particular interest when the structure of the observed scene is not obvious, especially when the observed planes are not perpendicular.

Acknowledgements

References

[1] R. T. Azuma, Y. Baillot, R. Behringer, S. Feiner, S. Julier, and B. MacIntyre. Recent Advances in Augmented Reality. *IEEE Computer Graphics and Applications*, pages 34–47, Dec. 2001.

[2] D. Dementhon and L. Davis. Model Based Object Pose in 25 Lines of Code. *International Journal of Computer Vision*, 15:123–141, 1995.

[3] O. D. Faugeras and G. Toscani. The Calibration Problem for Stereo. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL (USA)*, pages 15–20, 1986.

[4] M. A. Fischler and R. C. Bolles. Random Sample Consensus: A Paradigm for Model Fitting with Ap-

plications to Image Analysis and Automated Cartography. *Communications of the ACM*, 24(6):381–395, 1981.

[5] A. Fitzgibbon and A. Zisserman. Automatic Camera Recovery for Closed or Open Images Sequences. In *Proceedings of 5th European Conference on Computer Vision, University of Freiburg (Germany)*, pages 311–326, June 1998.

[6] D. Gennery. Visual Tracking of Known Three Dimensional Objects. *International Journal of Computer Vision*, 7(3):243–270, 1992.

[7] C. Harris and M. Stephens. A Combined Corner and Edge Detector. In *Proceedings of 4th Alvey Conference*, Cambridge, Aug. 1988.

[8] R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, ISBN: 0521623049, 2000.

[9] D. Liebowitz, A. Criminisi, and A. Zisserman. Creating Architectural Models from Images. In *EUROGRAPHICS'99, Milano, Italy*, 1999.

[10] B. Peuchot. Virtual Reality As An Operative Tool During Scoliosis Surgery. In *Proceedings of Imagina, Monte-Carlo, France*, page 262, Feb. 1995.

[11] S. Ravela, B. Draper, J. Lim, and R. Weiss. Tracking Object Motion Across Aspect Changes for Augmented Reality. In *ARPA Image Undertanding Worshop, Palm Spring (USA)*, Aug. 1996.

[12] G. Simon and M.-O. Berger. A Two-stage Robust Statistical Method for Temporal Registration from Features of Various Type. In *Proceedings of 6th International Conference on Computer Vision, Bombay (India)*, pages 261–266, Jan. 1998.

[13] G. Simon, A. Fitzgibbon, and A. Zisserman. Markerless tracking using planar structures in the scene. In *Proc. International Symposium on Augmented Reality*, pages 120–128, October 2000.

[14] A. State, G. Hirota, D. Chen, W. garett, and M. Livingston. Superior Augmented Reality Registration by Integrating Landmark Tracking and Magnetic Tracking. In *Computer Graphics (Proceedings Siggraph New Orleans)*, pages 429–438, 1996.

[15] B. Triggs. Autocalibration from planar scenes. In *European Conference on Computer Vision*, Freiburg, June 1998.

[16] M. Uenohara and T. Kanade. Vision based object registration for real time image overlay. *Journal of Computers in Biology and Medecine*, 1996.

[17] Z. Zhang, R. Deriche, O. Faugeras, and Q. Luong. A Robust Technique for Matching Two Uncalibrated Images Through the Recovery of the Unknown Epipolar Geometry. *Artificial Intelligence*, 78:87–119, Oct. 1995.

image #1

image #116

image #372

image #455

Figure 9. The registration process for two views of the campus sequence.