VISION-INERTIAL SYSTEM CALIBRATION FOR TRACKING IN AUGMENTED REALITY

Madjid Maidi, Fakhr-Eddine Ababsa and Malik Mallem
Complexes Systems Laboratory
University of Evry Val d'Essonne
40 Rue du Pelvoux, CE 1455 Courcouronnes, 91020 Evry Cedex, France

Keywords: Hybrid sensor, tracking system, calibration, computer vision, augmented reality.

Abstract: High accuracy registration between real and virtual environments is crucial in Augmented Reality (AR) systems. However, when a vision/inertial hybrid tracker is used,such accuracy depends mostly on the calibration procedure to determine transformations between the sensors frames. This calibration allows to project all data in a single reference frame. In this paper, we describe a new calibration method for a hybrid tracking system. It consists on rigidly assembling the hybrid tracker to a 6DOF robot in order to simulate the users head motion while tracking targets in AR environment. Our approach exploits the robot positioning to obtain a high accuracy for the tracker calibration. Experimental results and accuracy analyses are presented and demonstrate our approach effectiveness.

1 INTRODUCTION

Augmented Reality (AR) is the term used to describe systems in which the user's view of the real environment is enhanced by inserting computer graphics. These graphics must be generated in such a way that the user believes that the synthetic objects exist in the real environment (Jacobs et al., 1997). However, if there is a misregistration between virtual and real objects, the augmentation fails. In order to overcome this problem, over the past years, a new technology has focused on the use of hybrid tracking devices in AR systems. Fusing the multiple data sources provided by several sensors, gives an accurate information for virtual objects registration (Azuma, 1997), (Azuma and Bishop, 1995), (Bajura and Neumann, 1995).

Research works in this domain employ different sensing technologies for the motion tracking systems to compensate for the shortcomings of the used sensors and produce robust results (You et al., 1999). However, each technology has its strengths and weakness and uses a calibration method which depends on the employed system and the required accuracy of the application.

Azuma and Bishop (Azuma and Bishop, 1995) developed an optoelectronic tracking system to improve dynamic registration. For the calibration, the authors used directly the viewing measures parameters relying on geometric constraints. You, Neumann and Azuma (You et al., 1999) developed a tracking system that integrates inertial and vision-based technology to compensate for the limitations in each system component. The system was calibrated using a motion-based calibration (You et al., 1999), (You and Neumann, 2001). Foxlin (Foxlin, 2003) used a self tracker system composed of inertial and vision sensors. Auto-calibration algorithms were used to get high accuracy measurement without expensive calibration equipment. Chai, Hoff and Vincent (Chai et al., 2002) used inertial sensors with two cameras for the tracking process. To simplify the kinematics model of the system, the authors coincided the different sensors frames.

In the present work we develop a new technique for calibrating a camera with an inertial sensor using a 6DOF robot. Having both camera and robot position data while observing some features points of the environment, the transformation between the camera and the inertial sensor frames is determined.

The remainder of the paper is organized as follows. Section 2 describes the system components and represents the different sensors frames. The calibration procedure using a robot is presented in section 3, the models of sensors and the frames transformations are reported. Section 4 shows the experimental setup and

Maidi M., Ababsa F. and Mallem M. (2005).
VISION-INERTIAL SYSTEM CALIBRATION FOR TRACKING IN AUGMENTED REALITY.
In *Proceedings of the Second International Conference on Informatics in Control, Automation and Robotics - Robotics and Automation*, pages 156-162
DOI: 10.5220/0001183901560162

the obtained results. We conclude by section 5 where we present conclusions and future work.

2 HYBRID TRACKING SYSTEM

Our hybrid system is composed of an inertial sensor and a CCD camera rigidly mounted onto a robot as illustrated in Figure 1.

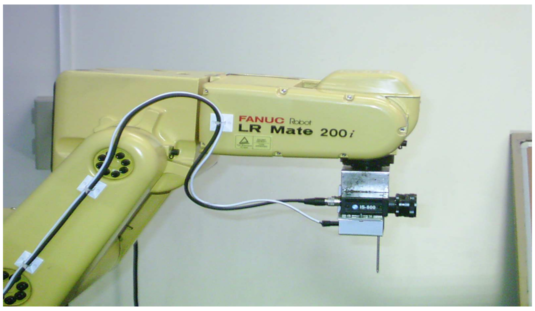

Figure 1: Hybrid system mounted onto the robot.

2.1 Inertial Sensor

The inertial sensor (MT9-B from Xsens) measures accelerations, rate of turn and earth magnetic field. All these data are in the right handed cartesian coordinate system, $\{I\}$, as defined in Figure 2. This coordinate system is the body-fixed to the Inertial Measurement Unit (IMU) and it is substantially aligned to the external housing of the IMU. The IMU software computes the rotation of the IMU frame, $\{I\}$, with respect to a global coordinate system, $\{G\}$, defined as a right handed cartesian coordinate system (Figure 2) with

- X positive when pointing to the local magnetic north.
- Y according to right handed coordinates (West).
- Z positive when pointing up.

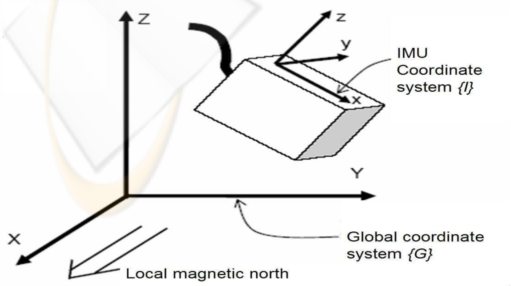

Figure 2: IMU related frames.

2.2 Camera Model

Our vision sensor is a CCD camera (IS-800 from i2S with $8mm$ focal length). Figure 3 illustrates the different frames used for the camera calibration. The calibration procedure simulates the camera by a theoretical model which describes the transformation of the scene (3D objects) toward the image.

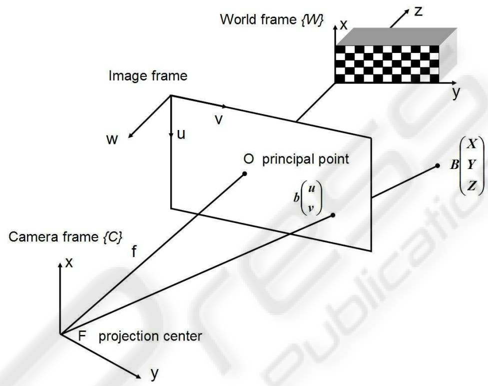

Figure 3: Coordinate systems used in camera calibration.

The camera calibration determines the geometrical model of an object and the corresponding image formation system which is described by the following equation

$$\begin{pmatrix} su \\ sv \\ s \end{pmatrix} = A\,[R \;\; T] \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \tag{1}$$

where s is an arbitrary scale factor, (R, T) called the extrinsic parameters, is the rotation and translation which relate the world coordinate system, $\{W\}$, to the camera coordinate system, $\{C\}$, and A called the camera intrinsic matrix given by

$$A = \begin{pmatrix} \alpha_u & 0 & u_0 \\ 0 & \alpha_v & v_0 \\ 0 & 0 & 1 \end{pmatrix} \tag{2}$$

with (u_0, v_0) the coordinates of the principal point and α_u and α_v the scale factors according to u and v image axes.

3 CALIBRATION PROCEDURE

We have rigidly mounted our hybrid tracker onto a 6DOF robot (LR Mate 200i from FANUC Robotics) in order to exploit the accuracy of the robot positioning in a calibration process. The coordinate frames used in the calibration procedure are illustrated in

Figure 4. The transformation between two frames is represented by the rotation matrix and the translation vector. (R_{CI}, T_{CI}) is the transformation of the IMU frame with respect to the camera frame, (R_{IT}, T_{IT}) is the transformation of the tool frame with respect to the IMU frame, (R_{WT}, T_{WT}) is the transformation of the robot tool frame with respect to the world frame and (R_{CW}, T_{CW}) is the transformation of the world frame with respect to the camera frame.

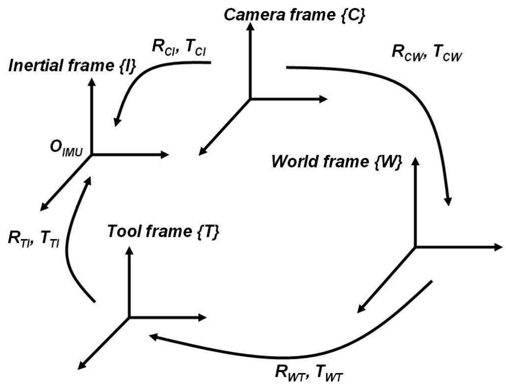

Figure 4: Coordinate frames related to the hybrid system.

The used robot is a manipulator arm (Figure 5), it is composed of six rotation axes. The robot is completely articulated with its six axes, its tool (axis 6) is the reference for motion and all settings applications.

Figure 5: Robot and rotation axes.

The robot provides the coordinates of its tool frame with respect to a user reference frame. Consequently, to determine the position of a new tool frame in a new user frame, it is necessary to make a learning of both frames: tool and user.

3.1 Robot Frames

By default, the user frame is related to the basis of the robot (axis 1). The robot data are expressed in the tool frame with respect to the robot user frame. Nevertheless, the transformation between the robot user frame and the world frame of the camera is unknown. For this purpose, we define a new user frame, $\{U\}$, to have the same orientation as the camera world frame, $\{W\}$. We learn also a new tool frame, $\{T\}$, so that $\{T\}$ and $\{I\}$ are coincident. The aim of these frames definitions is to derive directly the pose of the IMU frame from the robot computed coordinates.

3.1.1 Definition of the Tool Frame

This frame is related to the last axis of the robot. As we want to align the tool frame and the IMU frame, we define the axes of $\{T\}$ according to $\{I\}$ axes using six points method for learning.

3.1.2 Definition of the User Frame

The three points method is used to learn the user frame. It consists on moving the tool to the beginning and according to two other points on the X and Y axes of the user frame which is in this case the world camera frame.

3.2 IMU Calibration

3.2.1 IMU Orientation

Method 1 We can determine the IMU orientation with respect to the global frame which is by definition related to the earth magnetic field. We can also compute the IMU orientation in an earth fixed coordinate frame that is different from the global coordinate frame. In this work, we defined a new global coordinate frame, $\{G\}$, the IMU has to be orientated in such a way that the sensor axes all point onto exactly the same direction as the axes of the global coordinates frame (Figure 6). Afterwards, the orientation output will be with respect to the newly defined global axes. The IMU is used to record orientation of 3D object in real time. However, when the IMU is mounted onto an object which contains ferromagnetic materials (for example a camera) the measured magnetic field is distorted and this will cause errors on the orientation measurement.

We can also obtain the position of the IMU by integrating the acceleration data. It's theoretically possible to double integrate accelerometer data, after coordinate transformations and subtraction of the acceleration due to gravity, we obtain 3D position data.

To implement this, some practical issues will be encountered:

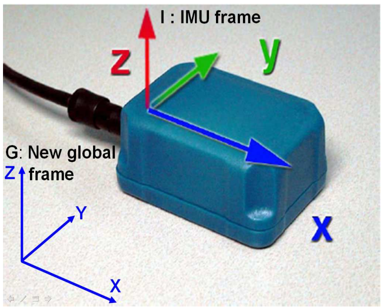

Figure 6: New global frame of the IMU.

- We need a "starting point", a reference 3D position, from which we can start to integrate the 3D acceleration data.
- Noise on the acceleration data and small offset errors and/or incorrectly subtracted acceleration due to gravity, will be integrated and over time will cause huge (drift) errors in the position estimate if used longer than a few seconds without any external update of true position.

The conclusion is that the orientation and also the position determined by this method depends very much on the type of motion and the environment in which we are operating. For the position estimation typically, short duration motions, preferably cyclical, with known reference positions will work well. We must also take into account the magnetic perturbations, actually, the orientation measured by the IMU is affected by the disturbances caused by the ferromagnetic objects present in the environment. These constraints and problems obliged us to choose another method more appropriated for our application which needs accurate orientation and position measurements.

Method 2 As we already evoked, the robot gives the position of the tool located at the end of its last axis with respect to a defined user frame. Knowing the positioning of the IMU with respect to the robot tool, we deduce the transformation between the IMU related frame, $\{I\}$, and the user frame, $\{U\}$, where $\{U\}$ represents the camera world frame.

The IMU rotation with respect to the camera frame is

$$R_{CI} = R_{CW}.R_{WT}.R_{TI} \tag{3}$$

where R_{CI} and R_{CW} are respectively the rotation of the IMU frame and the world frame with respect to the camera frame, R_{WT} is the rotation of the tool frame with respect to the world frame and R_{TI} is the rotation of the IMU frame with respect to the tool frame.

3.2.2 IMU Translation

For this part, we use also the data provided by the robot, which are the coordinates of its tool frame, $\{T\}$, with respect to its user frame, $\{U\}$. A simple reading on the robot control tool, allows to know the three translation components of the tool with respect to the user frame.

Nevertheless, we need to determine the translation of the IMU frame, $\{I\}$, with respect to the camera frame, $\{C\}$. Then, it is important to know exactly the position of the IMU frame with respect to the robot tool frame, $\{T\}$.

Indeed, the coordinates of the IMU frame origin, O_{IMU}, according to the tool frame, represent the translation of the IMU frame with respect to the tool frame, we denote it T_{TI}. This translation is computed from reported measurements and manufacturer data. T_{TI} is known, we compute T_{WI}, the translation of the IMU frame with respect to the world frame. We apply the following formula of coordinate transformation to determine T_{WI}

$$T_{WI} = R_{WT} \cdot T_{TI} + T_{WT} \tag{4}$$

Finally, the translation to T_{CI} is given by

$$T_{CI} = R_{CW} \cdot T_{WI} + T_{CW} \tag{5}$$

3.3 Camera Calibration

In this work, we have used a calibration method which is based on Zhang technique (Zhang, 1998). The camera observes a planar pattern from a few (at least two) different orientations. We can move either the camera or the planar pattern, the motion does not need to be known. The camera intrinsic and extrinsic parameters are solved using an analytical solution, followed by a nonlinear optimization technique based on the maximum likelihood criterion (Zhang, 1998). Radial and tangential lens distortions are also modeled and very good results have been obtained compared with classical techniques which use two or more orthogonal planes.

From (1) the rotation matrix and the translation vector are computed during the determination of the camera parameters in the calibration procedure. This transformation expresses the orientation and the translation of the camera frame, $\{C\}$, with respect to the camera world frame, $\{W\}$.

4 EXPERIMENTS

4.1 Experimental Setup

The hybrid tracker calibration procedure described in the previous section was experimented. We have determined the rigid transformation between the camera and the IMU frame using the calibration bench illustrated in Figure 7. The hybrid system is mounted onto the robot. First, we fix a metal point to the IMU housing (Figure 1 and Figure 7). The end of this point defines our tool frame which represents the motion reference and used for all frame learning operations. We set a test card opposite to the system to calibrate the camera and to learn the user frame of the robot. Then, we report the robot tool coordinates for each recorded position.

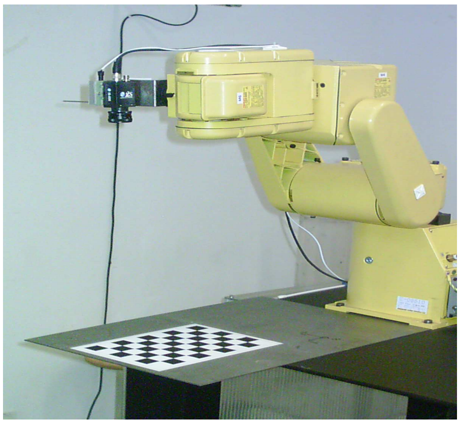

Figure 7: Calibration bench.

4.2 Orientation between IMU Frame and Camera Frame

Since the first method which uses the orientation computed by the IMU software does not give good experimental results, we use the robot data which are the orientation and the position of the tool frame with respect to the user frame.

As we already said, for the kinematics simplicity, we define a tool frame, $\{T\}$, which has the same orientation as the IMU frame, $\{I\}$.
Hence, the rotation between these two frames is an identity matrix

$$R_{TI} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{6}$$

The user frame, $\{U\}$, is aligned to the camera world frame, $\{W\}$. Then, the IMU orientation is directly given by the orientation of $\{T\}$ with respect to $\{W\}$.

The camera orientation is computed by calibration. Several images taken from different viewpoints were used for this procedure (Figure 8).

We used a single camera with $8mm$ lenses and 640×480 8-bit grayscale images. For the experiment,

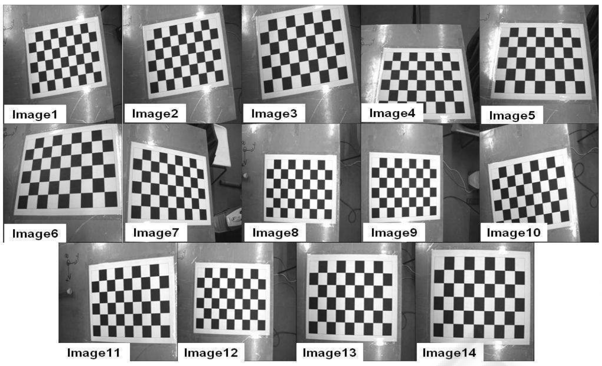

Figure 8: Camera images used for the calibration.

we have tried various numbers of images. The used formulation needs at least 2 images in different orientations for the pose estimation. On the other hand, we found that using more than 14 images did not increase the accuracy any more. In the experiment, there are 40 corner points in each image. After calibration, the obtained results for the camera internal parameters are: the scale factors $(\alpha_u, \alpha_v) = (989, 986)\ pixel$, the image center $(u_0, v_0) = (380, 283)\ pixel$, the radial distortions $(k_1, k_2) = (-0.2395, 0.3938)$ and the tangential distortions $(t_1, t_2) = (-0.0004, -0.0018)$. The extrinsic parameters are represented by the rotation matrix and the translation vector of patterns position in the image.

The RMS (Root Mean Square) error between the original and the reconstituted image points is equal to $2.6525\ pixels^2$. Of course we introduced the radial and tangential distortions into the perspective projection matrix to correct geometric errors of the camera.

The rotation of the IMU frame, $\{I\}$, with respect to the camera frame, $\{C\}$, is determined by

$$R_{CI} = R_{CW}.R_{WI} \tag{7}$$

The rotation angles of the IMU with respect to the camera are finally computed from the rotation matrix R_{CI}.

We used 14 positions of our hybrid sensors system and we notice on the whole that the obtained rotation angles are practically the same. However, to evaluate efficiently the performance of this method, the rotation measurement errors are computed (Figure 9).

Evaluation of the Rotation Errors The mean value of each angle is computed and the static error corresponding to each angle measurements is determined.

The mean rotation errors (MRE) of the angles are

$$\begin{aligned} MRE_\psi &= 0.32° \\ MRE_\theta &= 0.30° \\ MRE_\phi &= 0.18° \end{aligned} \tag{8}$$

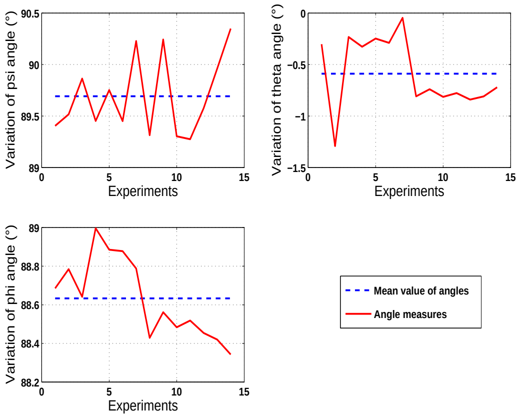

Figure 9: Variation of the IMU rotation angles.

Finally, the rotation angles of the IMU frame with respect to the camera frame are given by the following mean values

$$\begin{aligned} \psi &= 89.69^\circ \\ \theta &= \text{-}0.59^\circ \\ \phi &= 88.63^\circ \end{aligned} \quad (9)$$

4.3 Translation between IMU Frame and Camera Frame

To compute the translation of the IMU origin, O_{IMU}, in the camera frame, it is necessary to know the position of O_{IMU} in the world frame, T_{WI}, and then project into the camera frame. First, we determine the translation of the IMU frame with respect to the tool frame which is computed using the projection of O_{IMU} coordinates into the robot tool frame (see (4) and (5)). The coordinates of O_{IMU} with respect to the tool frame are

$$O_{IMU}\,(mm) = \begin{pmatrix} \text{-}6.0 \\ 7.8 \\ 75.0 \end{pmatrix}_{\{T\}} \quad (10)$$

For this experiment, we use the same positions and orientations of the robot which were used to compute the IMU rotation in the camera frame.

We replace the values of T_{WT} and T_{CW} in (4) and (5) where T_{TI} is O_{IMU} given in (10).

Evaluation of the Translation Errors We compute the mean value of each component of the T_{CI} coordinates for all robot positions used for this calibration (Figure 10). The mean translation errors (MTE) components are

$$\begin{aligned} MTE_X &= 1.5\ mm \\ MTE_Y &= 1.5\ mm \\ MTE_Z &= 1.2\ mm \end{aligned} \quad (11)$$

Finally, the translation of the IMU frame with respect to the camera frame is given by the following vector which expresses the coordinates of the IMU with respect to the camera frame

$$T_{CI}\,(mm) = \begin{pmatrix} 7.2 \\ 40.8 \\ \text{-}41.6 \end{pmatrix} \quad (12)$$

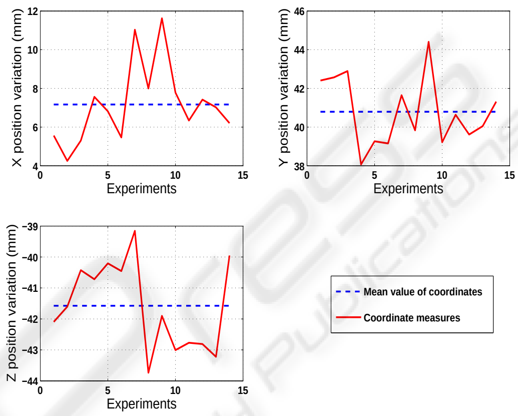

Figure 10: Variation of the IMU translation coordinates.

5 CONCLUSIONS

In this work, we presented a new approach to calibrate a hybrid tracking device for an augmented reality system. The system consists on a camera and an inertial measurement unit rigidly attached and mounted onto a robot tool axis. This robot allows the displacement of its tool in a workspace and computes the position and the orientation of the tool frame in a defined reference frame. The calibration of the camera and the coordinates provided by the robot determine the transformation between the inertial measurement unit and the camera with high accuracy.

The obtained calibration accuracy is sufficient for the tracking application for which this hybrid system was concerned. The evaluation of the numerical results showed the validity and the effectiveness of the proposed approach.

Our future work is to test prediction filters with real data provided by this hybrid system. We will integrate the robot information data to correct and evaluate the tracking methods before implementing prediction algorithms on a portable AR system.

REFERENCES

Azuma, R. (1997). A survey of augmented reality. In *Presence:Teleoperators and Virtual Environments, Vol.6, No.4, pp. 355-385*.

Azuma, R. and Bishop, G. (1995). Improving static and dynamic registration in an optical see-through hmd. In *Proceedings of SIGGRAPH 95*.

Bajura, M. and Neumann, U. (1995). Dynamic registration correction in augmented reality systems. In *Proceedings of IEEE Virtual Reality Annual International Symposium, pp. 189-196*.

Chai, L., Hoff, W., and Vincent, T. (2002). Three-dimensional motion and structure estimation using inertial sensors and computer vision for augmented reality. In *Presence: Teleoperators and Virtual Environments, Vol.11, pp. 474-492*.

Foxlin, E. (2003). Vis-tracker: A wearable vision-inertial self-tracker. In *Proceedings of the IEEE VR2003*.

Jacobs, M., Livingston, M., and State, A. (1997). Managing latency in complex augmented reality systems. In *Proceedings of the 1997 Symposium on Interactive 3D graphics*.

You, S. and Neumann, U. (2001). Fusion of vision and gyro tracking for robust augmented reality applications. In *Proceedings of IEEE VR2001, pp. 71-78*.

You, S., Neumann, U., and Azuma, R. (1999). Hybrid inertial and vision tracking for augmented reality registration. In *Proceedings of IEEE VR'99, pp. 260-267*.

Zhang, Z. (1998). A flexible new technique for camera calibration. In *Technical Report MSR-TR-98-71*.