Analysis of Video-Based 3D Tracking Accuracy by Using Electromagnetic Tracker as a Reference

Matjaž Divjak, Damjan Zazula
University of Maribor
Slovenia

1. Introduction

In recent years video-based tracking systems have been gaining widespread attention in several application fields. They are often used in military or surveillance applications (Ellis & Black, 2003; Collins et al, 2000; Cupillard et al., 2003; Fischer et al.,2004; Safeguards, 2007), in medicine (Grimson et al., 1998; Bornik et al, 2003; Pandya & Siadat, 2001; Tang et al., 1998; Bernd & Seibert, 2004), entertainment industry (Stapleton et al., 2002; Wren et al., 1997; Huang & Yan, 2002; Collomosse et al., 2003; Fua & Plankers, 2003) or sport (Qiu et al., 2004; Gueziec, 2002; Kristan et al., 2006), for research on human-computer interaction (Sato et al., 2004; Bradley & Roth, 2005; Polat et al., 2003), intelligent environments (Krumm et al., 2000) and similar. Continuous technological development and increasing competition among vendors have led to a great selection of tracking systems that are available on the market today with a variety of capabilities.

To compare them, several factors have to be considered. While price, speed or technical limitations may be very important for initial selection, the tracking accuracy is usually the most important property. To assess a tracking system and its precision, we need a reliable measure which allows for comparison of tracking system performance, provides estimates of tracking errors and indicates how to optimize the tracking system parameters.

The natural way to analyze the accuracy of any tracking system is to compare it to some reliable reference data. While a selection of comparison methods is readily available to the research community (Needham & Boyle, 2003), a reliable reference data (ground truth) can be hard to obtain, especially if greater accuracy is desired. Publicly available collections of video recordings with registered 3D ground truth information can be helpful, but are very scarce and with limited selection (Scharstein & Szeliski, 2003; CVTI, 2007). Such collections can be very useful in the development and testing of tracking algorithms, but are not enough for evaluation of a complex video tracking system in its actual operating environment.

Instead, one of the most popular approaches to obtain the reference data is to resort to an electromagnetic tracking device. These devices offer fast and accurate measurements and are insensitive to the line-of-sight requirements of optical motion trackers, which makes them ideally suited for tracking free-moving objects.

Source: Scene Reconstruction, Pose Estimation and Tracking, Book edited by: Rustam Stolkin,
ISBN 978-3-902613-06-6, pp.530, I-Tech, Vienna, Austria, June 2007

In this chapter we describe a general framework for assessing the 3D accuracy of video-based tracker by comparing it to an electromagnetic tracking device. Since both devices record data within their local coordinate systems, the data needs to be aligned accordingly before any comparison. This transformation between the coordinate systems of a video camera and the reference tracker is crucial for reliable and unbiased analysis of the optical tracking algorithm performance.

We analyze three possible models for the coordinate system alignment, based on measuring the position and orientation of video camera inside the reference coordinate frame. We also derive methods and metrics for comparing the models and their sensitivity. The transformation error is analytically and statistically separated from the tracking error of the algorithm, making it possible to compare 3D tracking accuracy of different algorithms in the same experimental setting.

The last part of the chapter demonstrates the applied value of the introduced models by a real-world experiment. The accuracy of a stereo camera-based face and hand tracker is analyzed by comparing the simultaneous measurements from the Polhemus 3Space Fastrak electromagnetic tracker (Polhemus, 1998). Three various transformation models are tested and compared using the derived metrics. Finally, the algorithm's tracking error is estimated by statistically separating it from the transformation-induced error.

2. Survey of the performance characterization of optical 3D tracking systems

Performance characterization of 2D tracking systems is a well developed field. Its maturity is confirmed by the growing success of conferences such as IEEE *Performance Evaluation of Tracking and Surveillance – PETS* (PETS, 2005), along with other workshops and specialised conference sections. The European project *Performance Characterization in Computer Vision – PCCV* (PCCV, 2007) also boosted the growing awareness and interest in the scientific community. A comprehensive review of the field can be found in (Christensen & Förstner, 1997; Gavrila, 1999; Black et al., 2003; Bashir & Porikli, 2006; Georis et al., 2003). In (Needham & Boyle, 2003), several metrics are presented for comparing the tracked trajectories, but they are still limited to 2D. The paper also describes an example of how to generate ground truth data by manually marking the video sequence. This approach is often used, despite the fact that it is very labour intensive, time demanding and unreliable. To make the process easier, several authors developed semi-automatic procedures that use existing collections of ground truth data to generate new reference data (Jaynes et al., 2002; Doermann & Mihalcik, 2000; Black et al., 2003; Georis et al., 2004).

Performance characterization of 3D optical trackers is faced with a serious obstacle, since reliable ground truth data is much harder to obtain than for 2D trackers. Manual and semi-automatic annotation of video streams with 3D reference information still have all the drawbacks of 2D approaches, and are even less precise due to difficulties in estimating the depth, which makes it generally unsuitable for such tasks. The best approach is to measure ground truth using a second 3D tracking or measuring device with significantly better accuracy than the tested device. Electromagnetic tracking devices, marker-based optical systems and laser scanners are all frequently used for this purpose.

Electromagnetic trackers such as (Polhemus, 1998) and (Ascension, 2007) are examples of the most popular solutions, and have been in use for more than 30 years. The latest models can produce measurements of a sensor's position and orientation (6 DOF) with sample rates up to 240 Hz and static accuracy of 0.8 mm RMS for position and 0.15° RMS for orientation.

They are insensitive to occlusions, which makes them very suitable for tracking free-moving targets, such as humans and their body parts in movement. The majority of products use wired sensors which can be cumbersome to wear and may interfere with the free movement of the object. However, newer devices solve this problem by using wireless, battery-powered sensors. A bigger concern is electromagnetic interference which greatly affects the actual device's precision and is very hard to avoid in any urban environment. Precision also decreases rapidly once the distance from the transmitter crosses a certain limit. Therefore, appropriate means should be taken to reduce the effect of environment prior to performing any experiments (Kindarenko, 2000; La Cascia et al., 2000). Recent reports on the usage of a magnetic tracker for tracking the position of head movements were published in (Xiao et al., 2003; La Cascia & Sclaroff, 1999), while (Rehg & Kanade, 1994) reports using it for tracking the hand movements. In (Bernd & Seibert, 2004) a specially designed magnetic sensor was implemented to guide an augmented reality system during minimally invasive surgery.

Marker-based optical systems mean another attractive solution. To ensure the accuracy which is required for a reliable ground truth, the reference optical trackers usually depend on active or passive markers that are attached to the target. The NDI Optotrak Certus system (NDI, 2007) uses up to 512 markers at distances up to 2.25 m. Markers are scanned at 1500 Hz with accuracy of 0.15 mm RMS. NaturalPoint (NP, 2007) and ARTracking (ART, 2007) also supply various marker-based trackers. Besides their speed and reasonably good accuracy, optical trackers have another advantage. To calibrate them, a specially designed target is usually shown to the camera (Bornik et al, 2003). This same target can also be used to calibrate the optical tracker whose accuracy is being measured, so the same coordinate system is used, which greatly simplifies the data comparison. However, the main obstacle remains their sensitivity to occlusions, which is undesired when tracking the complex movements. It also hinders a reliable performance evaluation of the video-based tracker. Recent examples of application include tracking the position of the head (Vogt et al., 2006), the body (Herda et al., 2001), person tracking (Balan et al., 2005), in medicine (Keemink et al., 1991; Bornik et al, 2003), etc.

While the electromagnetic devices and marker-based optical trackers can only provide measurements for a limited number of 3D points, laser scanners can scan the whole scene and obtain dense range measurements with great accuracy. For example, the systems (Optix, 2007) and (VIVID, 2007) achieve the resolution of 0.05 mm at 100 mm distance and 0.5 mm at 900 mm distance. Dense range information is very useful for a number of applications, but comes at a price: the scene is usually scanned through a lens by a single laser and this operation typically takes a couple of seconds on modern devices. This currently makes laser trackers inappropriate for tracking any reasonably fast movement, but they can provide an excellent reference for static scenes. A combination of laser and optical tracking system for neuro-surgery application is described in (Grimson et al., 1998).

Unfortunately, a surprisingly low number of papers can be found on general evaluation of 3D tracking accuracy. Some authors (Yao & Li, 2004; La Cascia et al., 2000; Kindarenko, 2000) inspect this issue in more detail, but they ignore the relationship between the two coordinate systems, i.e. of the verified system and of the reference, and usually align the two sets of measurements by only looking for an optimal fit (Needham & Boyle, 2003). Such performance analysis is insufficient, as it masks possible tracker alignment errors and doesn't give real accuracy information. To clarify this issue the next chapter focuses on electromagnetic tracking device as an example of a ground truth for video-based tracking evaluation. We also explain the necessary coordinate system transformation and evaluate the factors involved in it.

3. Electromagnetic tracker as a reference for video-based tracking

In this section we describe the general framework for assessing the 3D accuracy of video-based tracker by comparing it to an electromagnetic tracking device. Fig. 1 depicts the usual approach. In order to compare the tracking performance, the target's position must be measured by both systems simultaneously. The magnetic sensor is firmly attached to the target object. Each time a frame of the scene is captured by the camera, the sensor's position is read and stored into a file, thus forming a motion trajectory of the target as detected by the magnetic tracker (a reference trajectory). Afterwards, the video is processed by a tracking algorithm to reconstruct the vision-based trajectory. Each trajectory is expressed in its own coordinate system (CS). In order to compare them, they need to be transformed into a common CS. Without loss of generality we select the coordinate system of magnetic tracker as the common CS in this discussion.

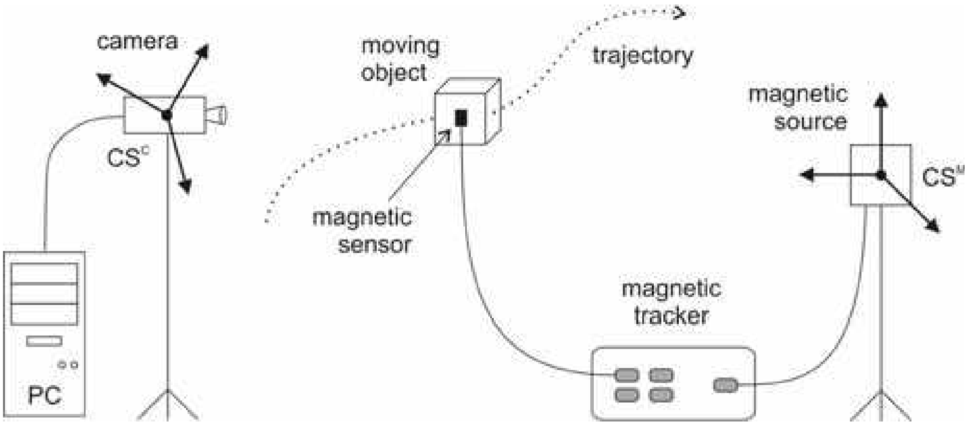


Figure 1. A general approach to analyzing the accuracy of video-based tracking with an electromagnetic tracking device. Suitable alignment of coordinate systems is necessary for comparison of detected motion trajectories

This transformation between the coordinate systems is crucial for a reliable and unbiased analysis of the optical tracking algorithm performance. Once the tracking data is properly aligned, it can be compared using any standard metrics, such as Root Mean Square (RMS) for example. The most frequently used method for aligning two trajectories uses optimization that minimizes the distances between them. Such a solution completely ignores possible bias errors and gives little information on how well the tracking algorithm follows the actual movement of the object. For example, if an algorithm consistently provides overestimated depths, the aligned trajectories can still show a close match. Another solution to this problem is aligning of the two coordinate systems physically by carefully positioning the camera and the magnetic tracker. Although this might seem a fast and simple procedure, such alignment is never perfect and results in considerable transformation errors. A quick calculation shows that an orientation error of 1° results in the position error of 3.5 cm at a distance of 2 meters from the camera.

A better approach to align the coordinate systems is by measuring the position and orientation of video camera using the magnetic tracker's sensors. This gives us enough information to derive a mathematical transformation between the CS of video camera (CS^C) and magnetic tracker (CS^M). Such alignment enables more thorough study of the transformation and its parameters, as well as comparison between the errors caused by the transformation and by the tested tracking algorithm. Although the idea seems straightforward, its implementation must be carefully considered, as will be explained in the next subsections.

3.1 Transformation models for coordinate systems

Assume we have a point in 3D space that needs to be expressed in two coordinate systems simultaneously. In CS^C we denote it by $\mathbf{p}^{\mathrm{C}} = (p_1^{\mathrm{C}}, p_2^{\mathrm{C}}, p_3^{\mathrm{C}}, 1)^{\mathrm{T}}$ and in CS^M by $\mathbf{p}^{\mathrm{M}} = (p_1^{\mathrm{M}}, p_2^{\mathrm{M}}, p_3^{\mathrm{M}}, 1)^{\mathrm{T}}$, respectively (using homogenous coordinates and denoting the transposition of vectors by $^{\mathrm{T}}$). Since both vectors $\mathbf{p}^{\mathrm{M}}$ and $\mathbf{p}^{\mathrm{C}}$ represent the same point in space, the following equation holds:

$$\mathbf{p}^{\mathrm{M}} = \mathbf{A}\mathbf{p}^{\mathrm{C}} . \tag{1}$$

Transformation matrix $\mathbf{A}$ contains the information about translation and rotation of CS^C with regards to CS^M. The position of camera's origin can be described by point $\mathbf{o}^{\mathrm{C}} = (o_1, o_2, o_3)^{\mathrm{T}}$, while base vectors $\mathbf{i}^{\mathrm{C}} = (i_1, i_2, i_3)^{\mathrm{T}}$, $\mathbf{j}^{\mathrm{C}} = (j_1, j_2, j_3)^{\mathrm{T}}$ and $\mathbf{k}^{\mathrm{C}} = (k_1, k_2, k_3)^{\mathrm{T}}$ describe its orientation. If homogenous coordinates are used, matrix $\mathbf{A}$ has the following structure:

$$\mathbf{A} = \begin{bmatrix} i_1 & j_1 & k_1 & o_1 \\ i_2 & j_2 & k_2 & o_2 \\ i_3 & j_3 & k_3 & o_3 \\ 0 & 0 & 0 & 1 \end{bmatrix} . \tag{2}$$

Vectors $\mathbf{i}^{\mathrm{C}}$, $\mathbf{j}^{\mathrm{C}}$, $\mathbf{k}^{\mathrm{C}}$ and $\mathbf{o}^{\mathrm{C}}$ that define $\mathbf{A}$ depend on a set of parameters $\boldsymbol{\Theta}$, $\boldsymbol{\Theta} = \{\Theta_l\}$, $l = 1, \ldots, N$. The exact number of parameters, N, depends on the procedure selected for building the transformation model. One of the most important parameters is the exact camera position. Of course, this information is usually not readily available, but it can be measured by placing one of the magnetic sensors on the camera and reading its position and orientation data. This simple approach has several shortcomings:

- The origin of CS^C is usually located inside the camera body and is impossible to be measured directly.
- The camera housing is usually metallic and therefore distorts the sensor's electromagnetic field.
- While inaccurate measurements of camera position have a relatively small effect on the overall accuracy, the erroneous camera orientation can cause significant deviations in results.

To address the abovementioned problems we present three different models for transformation of CS^C into CS^M. In all three models, the magnetic tracker is used to measure only the position of a number of control points around the camera that are used to calculate its position and orientation. Positional information shows significantly lower level of signal distortion than the information about orientation, as we have indicated. For a unique solution at least three control points in space are needed. They can be selected in a number of ways, but due to physical limitations of the used equipment (presence of ferromagnetic materials, camera range) this selection can affect the quality of transformation. The following options will be examined:

- All three control points are measured away from the camera (model A).
- Two points are measured on the camera and one away from it (model B).
- One point is measured on the camera and the other two away from it (model C).

Model A

To ensure that the camera body does not interfere with measuring magnetic sensor, all three control points are measured at a certain distance from it. The camera housing is fixed to a flat wooden board and accurately aligned with the board's sides (Fig. 2). Three corners of the board are selected and their coordinates are measured by magnetic sensor to obtain three control points $\mathbf{T}_1$, $\mathbf{T}_2$ and $\mathbf{T}_3$. Since it is assumed that camera's coordinate axes are completely aligned with the board, the base vector $\mathbf{i}^C$ can be expressed by $\overline{\mathbf{T}_3\mathbf{T}_1}$, the base vector $\mathbf{k}^C$ by $\overline{\mathbf{T}_2\mathbf{T}_1}$ and the base vector $\mathbf{j}^C$ is determined by the cross product (Fig. 2):

$$\mathbf{i}^C = \frac{\overline{\mathbf{T}_3\mathbf{T}_1}}{\left\|\overline{\mathbf{T}_3\mathbf{T}_1}\right\|},\ \mathbf{k}^C = \frac{\overline{\mathbf{T}_2\mathbf{T}_1}}{\left\|\overline{\mathbf{T}_2\mathbf{T}_1}\right\|},\ \mathbf{j}^C = \mathbf{k}^C \times \mathbf{i}^C \ . \tag{3}$$

The position of the camera's CS origin, $\mathbf{o}^C$, is expressed in CS^M by manually measuring the relative distances d_1 and d_2 between control point $\mathbf{T}_1$ and $\mathbf{o}^C$ (Fig. 2):

$$\mathbf{o}^C = \mathbf{T}_1 - d_1\mathbf{i}^C - d_2\mathbf{j}^C \ . \tag{4}$$

This way the transformation model A can be completely described by 11 parameters $\mathbf{\Theta}_A = \{x_1, y_1, z_1, x_2, y_2, z_2, x_3, y_3, z_3, d_1, d_2\}$, where $\mathbf{T}_1 = (x_1, y_1, z_1)$, $\mathbf{T}_2 = (x_2, y_2, z_2)$ and $\mathbf{T}_3 = (x_3, y_3, z_3)$.

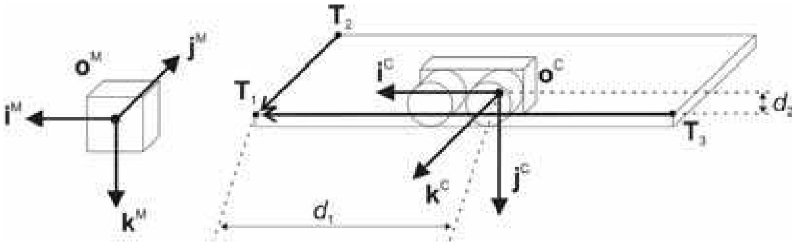


Figure 2. The setup of the magnetic tracker (left) and the camera (right) for model A. Vectors $\mathbf{i}^M$, $\mathbf{j}^M$, $\mathbf{k}^M$ denote base vectors of CS^M, while $\mathbf{i}^C$, $\mathbf{j}^C$, $\mathbf{k}^C$ denote base vectors of CS^C. $\mathbf{T}_1$, $\mathbf{T}_2$ and $\mathbf{T}_3$ mark the control point positions, while d_1 and d_2 mark manual measurements

Model B

The second transformation model neglects the fact that the camera housing disturbs the measurements, but it can be implemented only if those disturbances are proven very small compared to the errors caused by false orientation data. If the magnetic sensor is placed on the camera lens' face (Fig. 3), the disturbances caused by the camera housing are reduced to a minimum. This model also considers a stereoscopic (Jain et al., 1995) camera setup with two lenses that are used as two acceptable position measuring spots. The proposed model could be extended to a single camera, but the second measured point on the camera, $\mathbf{T}_3$, would be more problematic. This is why we are developing only the stereoscopic setup here.

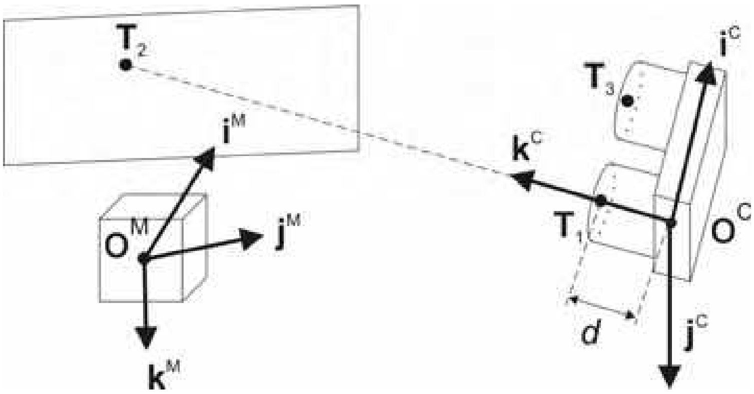


Figure 3. The setup of the magnetic tracker and the camera for model B. Vectors $\mathbf{i}^M$, $\mathbf{j}^M$, $\mathbf{k}^M$ denote base vectors of CS^M, while $\mathbf{i}^C$, $\mathbf{j}^C$, $\mathbf{k}^C$ denote base vectors of CS^C. $\mathbf{T}_1$, $\mathbf{T}_2$ and $\mathbf{T}_3$ mark the control point positions, while d marks the lens' focal length

First, the position of control point $\mathbf{T}_1$ on the face of the left camera lens (Fig. 3) is measured. Next, the magnetic sensor is attached to an arbitrary flat screen in front of the camera (control point $\mathbf{T}_2$ in Fig. 3) and the camera is aligned in such a way that the sensor is visible exactly in the centre of the left image. This step insures that the point $\mathbf{T}_2$ lies on the camera's (i.e. the lens') left optical axis. Since base vector $\mathbf{k}^C$ has the same direction as this optical axis, we calculate it from $\mathbf{T}_1$ and $\mathbf{T}_2$. The base vector $\mathbf{i}^C$ is obtained by measuring the coordinates of the third control point $\mathbf{T}_3$ on the face of the right camera lens (Fig. 3):

$$\mathbf{k}^C = \frac{\overrightarrow{\mathbf{T}_1\mathbf{T}_2}}{\left\|\overrightarrow{\mathbf{T}_1\mathbf{T}_2}\right\|},\ \mathbf{i}^C = \frac{\overrightarrow{\mathbf{T}_1\mathbf{T}_3}}{\left\|\overrightarrow{\mathbf{T}_1\mathbf{T}_3}\right\|}. \tag{5}$$

Base vector $\mathbf{j}^C$ is calculated from Eq. (3). The origin of CS^C lies on the camera's left optical axis and is determined by displacing the point $\mathbf{T}_1$ by d, i.e. the camera's focal length (Fig. 3):

$$\mathbf{o}^C = \mathbf{T}_1 - d\,\mathbf{k}^C. \tag{6}$$

The transformation model B is therefore described by 10 parameters:

$$\mathbf{\Theta}_B = \{x_1, y_1, z_1, x_2, y_2, z_2, x_3, y_3, z_3, d\},$$

where the initial 9 parameters have the same meaning as with $\mathbf{\Theta}_A$, and d is the focal length. When aligning $\mathbf{T}_2$ with the centre of left image, a quantization error of at least ½ pixel is unavoidable. This error causes a displacement of $\mathbf{T}_2$ from the optical axis by r_H in horizontal direction and r_V in vertical direction (relative to camera's left optical axis). At distance h from the screen, camera's field of view measures equal $u_H \times u_V$ (Fig. 4). At the same time, this area is represented in the image by $v_H \times v_V$ pixels. Therefore, an error of 1 pixel (± 0.5 pixel) causes a displacement of point $\mathbf{T}_2$ by

$$r_H = \frac{u_H}{v_H} = \frac{2h\,\mathrm{tg}\left(\frac{\varphi_H}{2}\right)}{v_H}, r_V = \frac{u_V}{v_V} = \frac{2h\,\mathrm{tg}\left(\frac{\varphi_V}{2}\right)}{v_V}, \tag{7}$$

where φ_H and φ_V mark the camera's horizontal and vertical view angles (Jain et al, 1995). Other parameters of model B are not affected by the image quantization error.

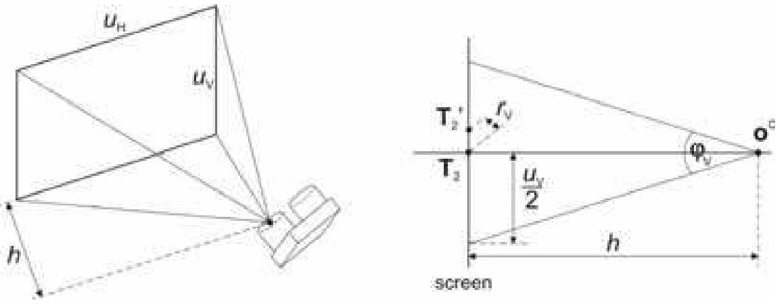


Figure 4. Left: stereo camera's field of view. Right: vertical displacement r_V of point $\mathbf{T}_2$ due to 1 pixel of error when aligning it with the centre of the left image

Model C

This model is similar to model B, except that control point $\mathbf{T}_3$ is also measured on the screen in front of the camera (Fig. 5). The camera should be aligned so that its left image displays $\mathbf{T}_2$ in the centre, while $\mathbf{T}_3$ is displaced from the image centre by m_H pixels horizontally and m_V pixels vertically. Therefore, base vectors $\mathbf{k}^C$ and $\mathbf{j}^C$ are obtained by using the same procedure as with model B.

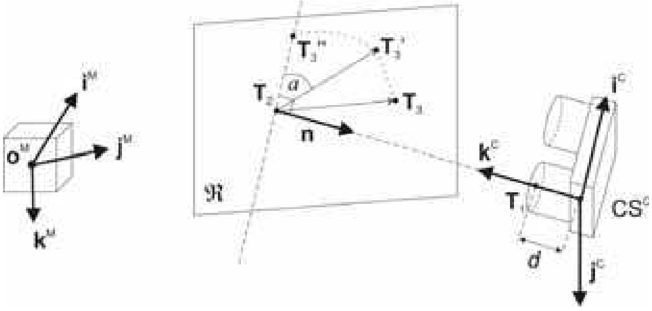


Figure 5. The setup of the magnetic tracker and the camera for model C. Vectors $\mathbf{i}^M$, $\mathbf{j}^M$, $\mathbf{k}^M$ denote base vectors of CS^M, while $\mathbf{i}^C$, $\mathbf{j}^C$, $\mathbf{k}^C$ denote base vectors of CS^C. $\mathbf{T}_1$, $\mathbf{T}_2$ and $\mathbf{T}_3$ mark the original control point positions, $\mathbf{T}_3'$ and $\mathbf{T}_3''$ mark recalculated position of $\mathbf{T}_3$, while d denotes the lens' focal length. Plane $\Re$ is perpendicular to the left optical axis

Base vector $\mathbf{i}^C$ could be determined by $\mathbf{T}_2$ and $\mathbf{T}_3$, but since the camera's optical axis is, in general, not perpendicular to the screen, the point $\mathbf{T}_3$ position must be recalculated accordingly. Imagine a plane $\Re$ which is perpendicular to the optical axis and intersects it in

$\mathbf{T}_2$. Point $\mathbf{T}_3' \in \Re$ can be determined by the intersection of $\Re$ and a line that is passing through $\mathbf{T}_3$ and is perpendicular to $\Re$ (Fig. 5). The new vector $\overline{\mathbf{T}_2\mathbf{T}_3'}$ is coplanar with $\mathbf{i}^\mathrm{C}$, but needs to be rotated around the camera's left optical axis to get parallel with $\mathbf{i}^\mathrm{C}$. The required angle of rotation α is determined by m_H and m_V:

$$\alpha = \mathrm{arctg}\frac{m_\mathrm{V}}{m_\mathrm{H}}\,. \tag{8}$$

The new point $\mathbf{T}_3''$, obtained after the rotation, is finally used to calculate the base vector $\mathbf{i}^\mathrm{C}$:

$$\mathbf{i}^\mathrm{C} = \frac{\overline{\mathbf{T}_2\mathbf{T}_3''}}{\left\|\overline{\mathbf{T}_2\mathbf{T}_3''}\right\|}\,. \tag{9}$$

The origin of CS^C is determined by focal length d, as in Eq. (6). The third transformation model is therefore described by 12 parameters:

$$\mathbf{\Theta}_\mathrm{C} = \{x_1, y_1, z_1, x_2, y_2, z_2, x_3, y_3, z_3, d, m_\mathrm{H}, m_\mathrm{V}\}.$$

3.2 Sensitivity of transformation models and their parameters

Inaccuracies in matrix $\mathbf{A}$ (Eq. (2)) cause erroneous transformations of coordinate systems, that depend also on the measurement model applied. To assess the appropriateness of a particular model, a measure for comparing the transformations and their parameters is needed. When parameter values are measured, the inherent measurement error can be statistically estimated using standard techniques (Stoodley, 1984). However, the effect of each parameter on the final transformation error depends also on its sensitivity. To determine the sensitivity of transformation matrix $\mathbf{A}$ to the parameter set $\mathbf{\Theta}$, each element $a_{u,v} \in \mathbf{A},\ \forall u,v \in [1,2,3,4]$, will be described as a function of parameters $\Theta_l \in \mathbf{\Theta}$:

$$a_{u,v} = f_{u,v}(\Theta_1, \Theta_2, ..., \Theta_N)\,. \tag{10}$$

Sensitivity of transformation matrix $\mathbf{A}$ can be expressed by derivatives:

$$\frac{\partial \mathbf{A}}{\partial \Theta_l} = \frac{\partial a_{u,v}}{\partial \Theta_l} = \frac{\partial f_{u,v}(\Theta_1, \Theta_2, ..., \Theta_N)}{\partial \Theta_l},\ \text{for } \forall u,v \in [1,2,3,4], \forall l \in [1,...,N]\,. \tag{11}$$

Unfortunately, the resulting mathematical expressions in Eq. (11) are too complex for direct comparison. Instead, the derivatives can be compared numerically by using real parameter values obtained from experiments (Section 4). This procedure gives an estimate on the largest contributor to the transformation error.

The effect of a mutual interaction of parameter errors on the sensitivity of matrix $\mathbf{A}$ is generally too complex to determine, but the overall upper error bound of each transformation model can still be estimated. The magnitude of error amplification for a certain parameter can be expressed if Eq. (1) is differentiated:

$$\left\|\frac{\partial \mathbf{p}^\mathrm{M}}{\partial \Theta_l}\right\| = \left\|\frac{\partial \mathbf{A}}{\partial \Theta_l}\mathbf{p}^\mathrm{C}\right\| \le \left\|\frac{\partial \mathbf{A}}{\partial \Theta_l}\right\| \cdot \left\|\mathbf{p}^\mathrm{C}\right\|, \tag{12}$$

$$S_l = \frac{\left\| \frac{\partial \mathbf{p}^{\mathrm{M}}}{\partial \Theta_l} \right\|}{\left\| \mathbf{p}^{\mathrm{M}} \right\|} \leq \frac{\left\| \frac{\partial \mathbf{A}}{\partial \Theta_l} \right\| \cdot \left\| \mathbf{p}^{\mathrm{C}} \right\|}{\left\| \mathbf{p}^{\mathrm{M}} \right\|}, \text{for } \forall l = 1, \ldots, N \,. \tag{13}$$

Expression (13) describes the relative sensitivity S_l of point $\mathbf{p}^{\mathrm{M}}$ with regards to parameter Θ_l . For matrix norm calculation, a spectral norm $\left\| \mathbf{A} \right\|_2$ is suggested (Meyer, 2001). Replacing $\mathbf{p}^{\mathrm{C}}$ in (13) by relationship from Eq. (1), an expression for calculating the relative sensitivity for individual parameters yields:

$$S_l = \frac{\left\| \frac{\partial \mathbf{p}^{\mathrm{M}}}{\partial \Theta_l} \right\|}{\left\| \mathbf{p}^{\mathrm{M}} \right\|} \leq \frac{\left\| \frac{\partial \mathbf{A}}{\partial \Theta_l} \right\| \cdot \left\| \mathbf{A}^{-1} \mathbf{p}^{\mathrm{M}} \right\|}{\left\| \mathbf{p}^{\mathrm{M}} \right\|} \leq \frac{\left\| \frac{\partial \mathbf{A}}{\partial \Theta_l} \right\| \cdot \left\| \mathbf{A}^{-1} \right\| \cdot \left\| \mathbf{p}^{\mathrm{M}} \right\|}{\left\| \mathbf{p}^{\mathrm{M}} \right\|} = \left\| \frac{\partial \mathbf{A}}{\partial \Theta_l} \right\| \cdot \left\| \mathbf{A}^{-1} \right\|, \text{for } \forall l = 1, \ldots, N \,. \tag{14}$$

Finally, the upper relative sensitivity limit of the whole model (S^{MAX}) equals the sum of individual sensitivities:

$$S^{\mathrm{MAX}} = \frac{\left\| \frac{\partial \mathbf{p}^{\mathrm{M}}}{\partial \boldsymbol{\Theta}} \right\|}{\left\| \mathbf{p}^{\mathrm{M}} \right\|} \leq \left(\left\| \frac{\partial \mathbf{A}}{\partial \Theta_1} \right\| + \left\| \frac{\partial \mathbf{A}}{\partial \Theta_2} \right\| + \ldots + \left\| \frac{\partial \mathbf{A}}{\partial \Theta_N} \right\| \right) \cdot \left\| \mathbf{A}^{-1} \right\| \,. \tag{15}$$

Sensitivity of a certain model, S^{MAX}, shows how much the inaccuracies of measured model parameters destroy the correct coordinate system's alignment. It can serve as a model robustness measure. On the other hand, the transformation sensitivities related to the individual parameters, S_l, indicate how much uncertain parameter measurements can ruin a good alignment. Thus, they rank the parameters according to their devastating influence on the correct alignment and point out those whose measurements must be done most accurately.

3.3 Decomposition of vision-based tracking error

With a suitable reference, such as a magnetic tracker, the tracking error of a vision-based system can always be assessed. However, as we showed in previous sections, this error consists of two contributions: the error which emerges from the tracking algorithm and the error caused by inaccurate coordinate system transformation. The latter depends on the combination of parameter values, and can be made in favour for any of the models A, B or C from Subsection 3.1 just with adequate choice of parameter values. It is therefore important that any comparison of the models respect the same specific set of parameters, related to one specific setup of the camera and magnetic tracker. All comparison results and conclusions are thus valid for this selected setup only.

Performance of transformation models can be most realistically evaluated by comparing the actual motion trajectories obtained by the magnetic tracker with the aligned trajectories from the tracking algorithm. The RMS difference of matching coordinate pairs reveals the average deviation of the algorithm results from the magnetic tracker's reference. However, this error does not reveal the true accuracy of the tracking algorithm, because the transformation errors caused by inaccurate parameter values also contribute to the difference between trajectories. For detailed analysis, the transformation error needs to be separated from the tracking error of the algorithm. In the sequel, we describe two possible approaches.

Analytical approach

Eq. (1) explains the transformation of a 3D point from CS^C into CS^M under ideal circumstances. In reality, the measurements of control point positions $\mathbf{T}_1$, $\mathbf{T}_2$ and $\mathbf{T}_3$ contain inaccuracies. As a result, the transformation matrix $\mathbf{A}$ is determined corrupted. Denote it by $\mathbf{A}_e$:

$$\mathbf{A}_\mathrm{e} = \mathbf{A} \cdot \Delta\mathbf{A}\,, \tag{16}$$

where $\Delta\mathbf{A}$ is a 4×4 matrix representing the transformation errors. Any vision-based tracking algorithm is also incapable of estimating the exact location of a target $\mathbf{p}^C$, but instead reports corrupted coordinate position $\mathbf{p}_\mathrm{e}^\mathrm{C}$:

$$\mathbf{p}_\mathrm{e}^\mathrm{C} = \Delta\mathbf{P}^\mathrm{C} \cdot \mathbf{p}^\mathrm{C}\,. \tag{17}$$

Error matrix $\Delta\mathbf{P}^C$ contains unknown coordinate errors dp_1, dp_2 and dp_3 that are added to $\mathbf{p}^C$:

$$\Delta\mathbf{P}^\mathrm{C} = \begin{bmatrix} 1 & 0 & 0 & dp_1 \\ 0 & 1 & 0 & dp_2 \\ 0 & 0 & 1 & dp_3 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \tag{18}$$

The errors of magnetic tracker are considered significantly smaller than algorithm-based errors, so points $\mathbf{p}^M$ are considered exact in this derivation. Finally, the transformation from CS^C into CS^M can, under realistic circumstances, be expressed by

$$\mathbf{p}^\mathrm{M} = \mathbf{A}_\mathrm{e} \cdot \mathbf{C} \cdot \mathbf{p}_\mathrm{e}^\mathrm{C}\,, \tag{19}$$

where matrix $\mathbf{C}$ compensates the errors of both the algorithm and the transformation. Eq. (19) is valid for any pair of points $\mathbf{p}^\mathrm{M}$ and $\mathbf{p}^\mathrm{C}$. So, we can observe more of them together. If four points are selected, they together can be described by a matrix $\mathbf{P}^\mathrm{C} = \left[\mathbf{p}_1^\mathrm{C} \quad \mathbf{p}_2^\mathrm{C} \quad \mathbf{p}_3^\mathrm{C} \quad \mathbf{p}_4^\mathrm{C}\right]$. This matrix contains ideal homogenous coordinates of four arbitrary points. Analogously, the corresponding error-corrupted points can be joint in matrix $\mathbf{P}_\mathrm{e}^\mathrm{C}$ and related magnetic measurements in matrix $\mathbf{P}^M$. If we can find four points whose coordinate errors dp_1, dp_2 and dp_3 are the same, Eq. (17) can be extended to all four points together. Although this condition is hard to verify in practice, four measurements with the most similar error can still be found by searching through all the combinations of the observed points. A criterion for the error similarity will be presented at the end of this section. At this point, we suppose that $\Delta\mathbf{P}^C$ contains the identical error of four selected points.

By substitution of Eq. (1) in Eq. (17), the following relationship is obtained:

$$\mathbf{P}^\mathrm{M} \cdot \left(\mathbf{P}_\mathrm{e}^\mathrm{C}\right)^{-1} = \mathbf{A} \cdot \mathbf{P}^\mathrm{C} \cdot \left(\mathbf{P}^\mathrm{C}\right)^{-1} \cdot \left(\Delta\mathbf{P}^\mathrm{C}\right)^{-1} = \mathbf{A} \cdot \left(\Delta\mathbf{P}^\mathrm{C}\right)^{-1}. \tag{20}$$

Since the left-hand side of (20) is known and $\Delta\mathbf{P}^C$ has a specific structure, the contents of the ideal matrix $\mathbf{A}$ can be reconstructed as:

$$\mathbf{A}\cdot\left(\Delta\mathbf{P}^{\mathrm{C}}\right)^{-1}=\begin{bmatrix} i_1 & j_1 & k_1 & o_1 \\ i_2 & j_2 & k_2 & o_2 \\ i_3 & j_3 & k_3 & o_3 \\ 0 & 0 & 0 & 1 \end{bmatrix}\cdot\begin{bmatrix} 1 & 0 & 0 & -dp_1 \\ 0 & 1 & 0 & -dp_2 \\ 0 & 0 & 1 & -dp_3 \\ 0 & 0 & 0 & 1 \end{bmatrix}=\begin{bmatrix} i_1 & j_1 & k_1 & -i_1dp_1-j_1dp_2-k_1dp_3+o_1 \\ i_2 & j_2 & k_2 & -i_2dp_1-j_2dp_2-k_2dp_3+o_2 \\ i_3 & j_3 & k_3 & -i_3dp_1-j_3dp_2-k_3dp_3+o_3 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \quad (21)$$

First three columns of the resulting matrix in (21) represent the rotational part of ideal transformation matrix **A**, only the translation part (the fourth column) cannot be directly determined. However, new base vectors **i**, **j**, **k**, that are more reliable, can be computed and used to construct new transformation matrix $\hat{\mathbf{A}}$. This matrix represents the best estimate of ideal matrix **A**. If $\hat{\mathbf{A}}$ is used in Eq. (16) instead of **A**, the estimated transformation error Δ**A** can be obtained:

$$\Delta\mathbf{A}=\left(\hat{\mathbf{A}}\right)^{-1}\cdot\mathbf{A}_{\mathrm{e}}\,. \quad (22)$$

Using Eqs. (1) and (17), Eq. (19) can be rearranged into

$$\mathbf{p}^{\mathrm{C}}=\Delta\mathbf{A}\cdot\mathbf{C}\cdot\Delta\mathbf{P}^{\mathrm{C}}\cdot\mathbf{p}^{\mathrm{C}}\,, \quad (23)$$

which proves that

$$\Delta\mathbf{A}\cdot\mathbf{C}\cdot\Delta\mathbf{P}^{\mathrm{C}}=\mathbf{I}\,, \quad (24)$$

where **I** stands for identity matrix.

Since matrix **C** can be calculated from Eq. (19), the error matrix Δ$\mathbf{P}^{\mathrm{C}}$ can also be determined, giving also the matrix $\mathbf{P}^{\mathrm{C}}$ afterwards:

$$\begin{aligned}\Delta\mathbf{P}^{\mathrm{C}}&=\left(\mathbf{C}\right)^{-1}\cdot\left(\Delta\mathbf{A}\right)^{-1},\\ \mathbf{P}^{\mathrm{C}}&=\left(\Delta\mathbf{P}^{\mathrm{C}}\right)^{-1}\cdot\mathbf{P}_{\mathrm{e}}^{\mathrm{C}}.\end{aligned} \quad (25)$$

Finally, the exact vision-based points $\mathbf{P}^{\mathrm{C}}$ can be compared to magnetic tracker reference data, which results in a reliable tracking accuracy analysis.

Of course, this conclusion is based on the assumptions that four point vectors joint in matrix $\mathbf{P}_{\mathrm{e}}^{\mathrm{C}}$ contain the same error Δ$\mathbf{P}^{\mathrm{C}}$ and that the magnetic tracker can be considered error-free. If such a set of four points can be found that the reconstructed rotation vectors (denoted by $\hat{\mathbf{A}}_{\mathrm{ROT}}$) are orthonormal to each other, then both assumptions are satisfied. This property can be used as a criterion function when searching for a suitable set of points:

$$f_{\mathrm{ORTO}}\left(\hat{\mathbf{A}}_{\mathrm{ROT}}\right)=\sum_{\forall a\in\hat{\mathbf{A}}_{\mathrm{ROT}}}\left|\left(\hat{\mathbf{A}}_{\mathrm{ROT}}\right)^{\mathrm{T}}\cdot\hat{\mathbf{A}}_{\mathrm{ROT}}-\mathbf{I}\right|. \quad (26)$$

Four points obtained by tracking algorithm that generate the matrix $\hat{\mathbf{A}}_{\mathrm{ROT}}$ (using Eq. (20)) with the lowest f_{ORTO} value are the ones that best satisfy the assumptions.

Statistical approach

Another approach to separate the coordinate-system transformation error from the tracking-algorithm error is based on statistics. It can always be implemented, which makes it preferred to the analytical approach with four selected trajectory points whose proper choice is not always guaranteed in practice. Eq. (1) can again serve as a starting point, but instead of grouping four camera-based trajectory points with similar errors, we express the transformation of each point separately. Same as before, the ideal values of $\mathbf{A}$ and $\mathbf{p}^{\mathrm{C}}$ are unknown, only error-contaminated $\mathbf{A}_{\mathrm{e}}$ and $\mathbf{p}_{\mathrm{e}}^{\mathrm{C}}$ are available:

$$\begin{aligned}\mathbf{A}_{\mathrm{e}} = \mathbf{A} + \Delta\mathbf{A}, \mathbf{p}_{\mathrm{e}}^{\mathrm{C}} = \mathbf{p}^{\mathrm{C}} + \Delta\mathbf{p}^{\mathrm{C}},\\ \mathbf{p}^{\mathrm{M}} + \mathbf{e} = \mathbf{A}_{\mathrm{e}}\mathbf{p}_{\mathrm{e}}^{\mathrm{C}}\end{aligned} \tag{27}$$

Note that the transformation error $\Delta\mathbf{A}$ and algorithm-based error $\Delta\mathbf{p}^{\mathrm{C}}$ are handled differently than in previous approach, although the same notation is adopted. Instead of multiplicative error model, an additive error model is used here. Vector $\mathbf{e}$ denotes the total deviation of each transformed point $\mathbf{A}_{\mathrm{e}}\mathbf{p}_{\mathrm{e}}^{\mathrm{C}}$ from its magnetic reference position $\mathbf{p}^{\mathrm{M}}$. The error of magnetic tracker is considered to be insignificant compared to other errors, so the tracker's measurements are considered exact.

Using Eq. (27), the total tracking error can be expressed by

$$\begin{aligned}\mathbf{e} = -\mathbf{p}^{\mathrm{M}} + (\mathbf{A} + \Delta\mathbf{A})(\mathbf{p}^{\mathrm{C}} + \Delta\mathbf{p}^{\mathrm{C}}),\\ \mathbf{e} = \mathbf{A}\Delta\mathbf{p}^{\mathrm{C}} + \Delta\mathbf{A}\mathbf{p}_{\mathrm{e}}^{\mathrm{C}}.\end{aligned} \tag{28}$$

Since the actual value of $\mathbf{e}$ can be calculated for each point, only $\mathbf{A}$, $\Delta\mathbf{A}$, and $\Delta\mathbf{p}^{\mathrm{C}}$ remain unknown. Matrices $\mathbf{A}$ and $\Delta\mathbf{A}$ remain constant throughout the analysis. Consequently, some estimates about their value can be made using statistical methods. First, the maximum expected error of each transformation parameter Θ_l must be realistically estimated. Then, a set of random, normally distributed errors is generated and added to the measured parameter values of a selected coordinate-system transformation model, resulting in a transformation matrix $\mathbf{A}_{\mathrm{SIM}}$ whose coefficients are influenced by additional errors introduced artificially and, thus, exactly known. This matrix is used to transform the points $\mathbf{p}_{\mathrm{e}}^{\mathrm{C}}$, recognized by vision-based algorithm. A new trajectory is obtained which is a variation of the proper camera-tracked trajectory in CS$^{\mathrm{M}}$. By repeating this process and generating a large set of possible transformation matrices, their mean transformation error $m_{\mathrm{RMS}}(\mathbf{A}_{\mathrm{SIM}})$ can be calculated. Due to the averaging properties of the RMS metrics, this error is expected to approximate the actual mean transformation error $m_{\mathrm{RMS}}(\mathbf{A}_{\mathrm{e}})$. Experiments confirm this, provided that $\mathbf{A}_{\mathrm{e}}$ is sufficiently close to ideal $\mathbf{A}$.

If a large enough set of errors is simulated, one or more of the resulting trajectories may closely resemble the ideal transformation. Unfortunately, it cannot be specifically identified since the initial measurement error of $\mathbf{A}_{\mathrm{e}}$ remains unknown. The best we can do is to find the simulated trajectories that minimally or maximally deviate from the $m_{\mathrm{RMS}}(\mathbf{A}_{\mathrm{SIM}})$ and use them as estimates of minimal and maximal expected transformation error, $\Delta\mathbf{A}_{\mathrm{SIM}}^{\mathrm{MIN}}$ and $\Delta\mathbf{A}_{\mathrm{SIM}}^{\mathrm{MAX}}$. When those values are entered into Eq. (28) together with matching $\mathbf{A}_{\mathrm{SIM}}$ and $\mathbf{e}$ values, the estimated $\Delta\mathbf{p}_{\mathrm{SIM}}^{\mathrm{C}}$ can be calculated, and consequently the estimated $\mathbf{p}_{\mathrm{SIM}}^{\mathrm{C}}$ as well (Eq. (27)). Those simulation-based estimates can be used to statistically compare individual factors involved in the presented tracking accuracy analysis.

4. A practical example: stereo video tracking compared to the Polhemus Fastrak magnetic tracker

To illustrate the presented ideas on a practical example, we describe an experiment in which the Polhemus 3Space Fastrak magnetic tracker (Polhemus, 1998) is used as a reference for analysis of 3D tracking algorithm based on images from the Videre Design's STH-MD1-C stereo head (Videre, 2001). The obtained motion trajectories are aligned using all three presented transformation models and analyzed according to procedures explained in Section 3.

The Fastrak tracker uses four wired sensors and produces measurements with static resolution of 0.8 mm RMS for position and 0.15° RMS for orientation. This accuracy is only achieved when the sensor is less than 75 cm away from magnetic transmitter. We adapted the experiment to this requirement and took several measures to ensure that electromagnetic interference was minimal.

The digital STH-MD1-C stereo head uses two synchronized CMOS sensors with 9 cm of baseline distance and was in our case equipped with *f* = 48 mm lenses. At maximum resolution of 1288 × 1032 pixels the camera captures only 7.5 frames per second (fps), but if the frame size is reduced to 320 × 240 pixels, the frame rate increases to 110 fps. During all our experiments the camera was positioned approximately 1 meter away from the test subject. Detailed schematics of camera, Fastrak and test object are depicted in Fig. 6.

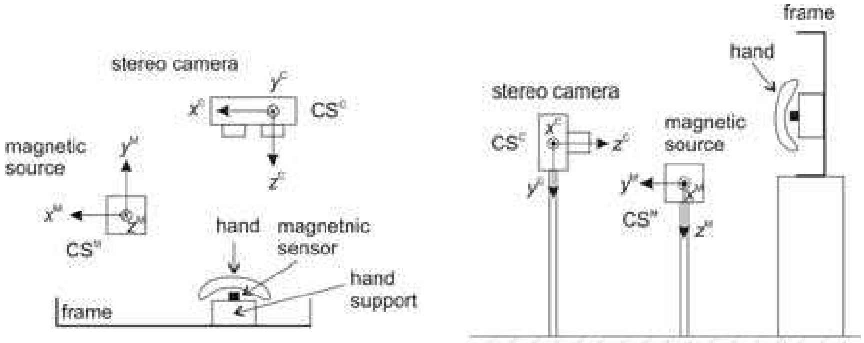


Figure 6. Schematics of the experiment setup including a stereo camera (CS^C), a source of magnetic pulses (CS^M), magnetic sensor and a frame for limiting the movement of the user's hand. Left side shows top view of the setup, right side shows side view

Video data was processed by our algorithm for detection of human hands and faces (Divjak, 2005). The algorithm uses bimodal colour and range information to detect consistent skin coloured regions. 3D centroids of those regions are tracked temporally by a Kalman filter-based prediction algorithm, resulting in smooth 3D motion trajectories of the tracked objects (Fig. 7).

The positions of all control points and other transformation model parameters were measured before conducting the experiment (Tables 1 and 2). Then, one of Fastrak's sensors was attached to the back of the test subject's right hand. The test subject moved his hand along a predefined, physically limited path so that the movement remained practically the same during all the experiments. Every time the stereo camera captured a pair of images, the position of magnetic sensor was read and stored. Three different video sequences were captured, each consisting of 120 – 200 colour image pairs with 320 × 240 pixels, and, in parallel, also the magnetic tracker reference data.

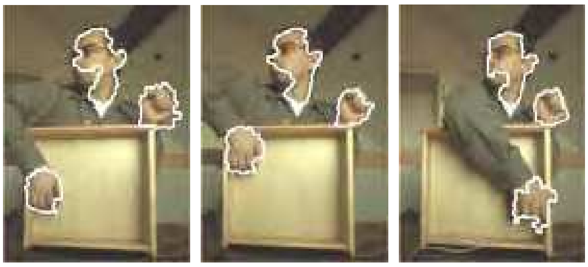


Figure 7. A few frames of the captured video overlaid with the object region borders (in white), as detected by the stereo tracking algorithm

Parameter	d_1	d_2	d	m_H	m_V
Value	490 mm	14 mm	48 mm	85 pixels	9 pixels

Table 1. Manually measured transformation parameter values for models A, B and C

Parameter	x_1	y_1	z_1	x_2	y_2	z_2	x_3	y_3	z_3
Model A (mm)	201.4	241.9	91.3	211.2	522.2	100.2	-68.1	266.0	124.5
Model B (mm)	-83.7	204.8	88.9	-62.4	-99.4	-41.7	-97.4	198.8	84.8
Model C (mm)	-83.7	204.8	88.9	-62.4	-99.4	-41.7	107.4	-17.7	-10.6

Table 2. Coordinates of control point T_1, T_2, T_3 for models A, B and C as determined by the Polhemus magnetic tracker

4.1 Comparison of transformation models

Using parameter values from Table 1 and Table 2 the base vectors $\mathbf{i}^C$, $\mathbf{j}^C$, $\mathbf{k}^C$ and coordinate system origin $\mathbf{o}^C$ were calculated for each model (Table 3). Those vectors can be used to construct transformation matrix **A** by Eq. (2). Relative sensitivity of model parameters is presented in Table 4. Finally, the upper sensitivity limit S^{MAX} of each transformation model is compared in Table 5. With our selection of parameter values, the model A turned out to be the most sensitive.

Model	Calculated CS^C base vector values
A	$\mathbf{i}^C = \begin{bmatrix} 0.999 \\ -0.028 \\ -0.038 \end{bmatrix}, \mathbf{j}^C = \begin{bmatrix} 0.037 \\ -0.033 \\ 0.999 \end{bmatrix}, \mathbf{k}^C = \begin{bmatrix} -0.035 \\ -0.999 \\ -0.032 \end{bmatrix}, \mathbf{o}^C = \begin{bmatrix} -288.6 \\ 255.9 \\ 96.0 \end{bmatrix}$
B	$\mathbf{i}^C = \begin{bmatrix} 0.999 \\ -0.028 \\ -0.038 \end{bmatrix}, \mathbf{j}^C = \begin{bmatrix} 0.037 \\ -0.038 \\ 0.999 \end{bmatrix}, \mathbf{k}^C = \begin{bmatrix} -0.027 \\ -0.999 \\ -0.037 \end{bmatrix}, \mathbf{o}^C = \begin{bmatrix} -289.2 \\ 250.0 \\ 96.1 \end{bmatrix}$
C	$\mathbf{i}^C = \begin{bmatrix} 0.998 \\ -0.045 \\ -0.040 \end{bmatrix}, \mathbf{j}^C = \begin{bmatrix} 0.038 \\ -0.038 \\ 0.998 \end{bmatrix}, \mathbf{k}^C = \begin{bmatrix} -0.027 \\ -0.999 \\ -0.037 \end{bmatrix}, \mathbf{o}^C = \begin{bmatrix} -289.2 \\ 250.0 \\ 96.1 \end{bmatrix}$

Table 3. Base vectors of CS^C and its origin (expressed in CS^M), as defined by measured parameter values

Model A		Model B		Model C	
Parameter	S_l value	Parameter	S_l value	Parameter	S_l value
x_1	398.5	x_1	410.4	x_1	408.8
y_1	175.2	y_1	390.7	y_1	389.1
z_1	174.8	z_1	410.2	z_1	408.5
x_2	1.4	x_2	20.4	x_2	20.3
y_2	0.7	y_2	3.4	y_2	3.4
z_2	19.9	z_2	20.3	z_2	20.2
x_3	10.6	x_3	0.4	x_3	0.1
y_3	224.3	y_3	6.9	y_3	0.1
z_3	224.3	z_3	4.5	z_3	0.8
d_1	398.6	d	390.1	d	388.5
d_2	398.6			m_H	0.5
				m_V	4.6

Table 4. Numerical relative sensitivity values S_l for all parameters of models A, B and C

Model	A	B	C
S^{MAX}	2026.9	1661.1	1651.2

Table 5. The upper sensitivity limit (S^{MAX}) of models A, B and C for our experimental setup

4.2 Trajectory comparison

Fig. 7 shows an example of how the algorithm detected the image regions that represent the tracked objects. However, it doesn't give us any clue about how accurate is matching between the reconstructed and the reference 3D position. To obtain this information, all captured trajectories were transformed from CS^C into CS^M (using constructed matrices **A**) and their deviation from the magnetic reference was estimated. Table 6 reports RMS differences for all three transformation models. An example of the aligned trajectories is depicted in Fig. 8. We experimented with 3 trajectories consisting of 500 3D points all together.

Model	X RMS difference (mm)	Y RMS difference (mm)	Z RMS difference (mm)	Total RMS difference (mm)
A	16.1 ± 6,1	5.8 ± 1.2	10.7 ± 1.7	20.5 ± 4.9
B	14.8 ± 2.9	12.7 ± 0.7	72.5 ± 6.4	75.1 ± 6.8
C	34.7 ± 3.2	12.7 ± 0.4	55.1 ± 6.7	66.4 ± 6.9

Table 6. The RMS difference between the coordinates of video-based and magnetic-based trajectories, as aligned by each transformation model. Mean values plus standard deviation for all captured trajectories are shown

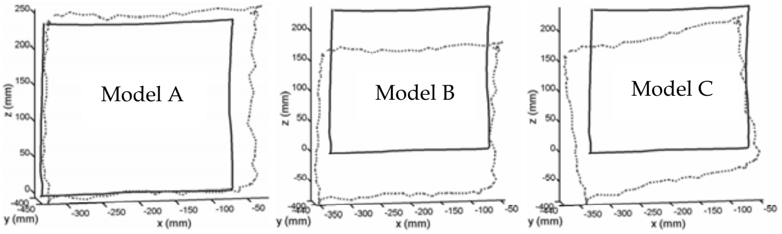


Figure 8. An example of transformation of vision-based trajectories from CS^C to CS^M, for each transformation model. The magnetic tracker data is depicted by solid lines, the video-based tracking algorithm data is depicted by dotted lines

4.3 Error analysis

The total tracking error (Table 6) origins in the transformation error and algorithm error, as described in Section 3.3. We tried to decompose the total error into its constituent components by the proposed statistical approach. First, we empirically estimated maximal expected errors of all model parameters. For Fastrak measurements, its factory specified accuracy of ± 0.8 mm was used, while for manual measurements with a tape measure we estimated accuracy of ± 0.5 mm. For measurements in pixels we estimated an error of ± 0.5 pixel, which according to our setup (the camera was 1 m away from the object) is equivalent to ± 4 mm. Using those values we generated a set of random, normally distributed measurement errors (zero mean, 1000 Monte-Carlo runs) that were added to actual measured parameter values, simulating the effect of error matrix $\Delta\mathbf{A}$. Mean values of vectors of the simulated matrix $\mathbf{A}_{\mathrm{SIM}}$ are shown in Table 7.

Model	Simulated CS^C base vector values
A	$\mathbf{i}^C = \begin{bmatrix} 0{,}999 \pm 0{,}000 \\ -0{,}028 \pm 0{,}001 \\ -0{,}038 \pm 0{,}001 \end{bmatrix}, \mathbf{j}^C = \begin{bmatrix} 0{,}037 \pm 0{,}001 \\ -0{,}033 \pm 0{,}004 \\ 0{,}999 \pm 0{,}000 \end{bmatrix}, \mathbf{k}^C = \begin{bmatrix} -0{,}035 \pm 0{,}004 \\ -0{,}999 \pm 0{,}000 \\ -0{,}032 \pm 0{,}004 \end{bmatrix}, \mathbf{o}^C = \begin{bmatrix} -288{,}7 \pm 0{,}8 \\ 249{,}9 \pm 0{,}6 \\ 95{,}8 \pm 0{,}7 \end{bmatrix}$
B	$\mathbf{i}^C = \begin{bmatrix} 0{,}999 \pm 0{,}001 \\ -0{,}028 \pm 0{,}011 \\ -0{,}038 \pm 0{,}011 \end{bmatrix}, \mathbf{j}^C = \begin{bmatrix} 0{,}037 \pm 0{,}011 \\ -0{,}038 \pm 0{,}003 \\ 0{,}999 \pm 0{,}000 \end{bmatrix}, \mathbf{k}^C = \begin{bmatrix} -0{,}027 \pm 0{,}003 \\ -0{,}999 \pm 0{,}000 \\ -0{,}037 \pm 0{,}003 \end{bmatrix}, \mathbf{o}^C = \begin{bmatrix} -289{,}2 \pm 0{,}7 \\ 249{,}9 \pm 0{,}8 \\ 96{,}1 \pm 0{,}7 \end{bmatrix}$
C	$\mathbf{i}^C = \begin{bmatrix} 0{,}999 \pm 0{,}000 \\ -0{,}025 \pm 0{,}003 \\ -0{,}038 \pm 0{,}011 \end{bmatrix}, \mathbf{j}^C = \begin{bmatrix} 0{,}038 \pm 0{,}007 \\ -0{,}038 \pm 0{,}003 \\ 0{,}999 \pm 0{,}000 \end{bmatrix}, \mathbf{k}^C = \begin{bmatrix} -0{,}027 \pm 0{,}003 \\ -0{,}999 \pm 0{,}000 \\ -0{,}037 \pm 0{,}003 \end{bmatrix}, \mathbf{o}^C = \begin{bmatrix} -289{,}2 \pm 0{,}8 \\ 249{,}9 \pm 0{,}8 \\ 96{,}1 \pm 0{,}8 \end{bmatrix}$

Table 7. Base vectors of CS^C and its origin (expressed in CS^M), obtained by a simulated matrix $\mathbf{A}_{\mathrm{SIM}}$. Mean values and standard deviations were estimated by 1000 iterations

The effect of transformation errors was evaluated on all available trajectories, detected by the tracking algorithm during our experiments. Each trajectory was transformed into CS^M using matrix $\mathbf{A}_{\mathrm{SIM}}$ and compared to the magnetic reference. Trajectories with minimal and maximal deviation from the mean simulated value were identified and the resulting transformation errors $\Delta\mathbf{A}_{\mathrm{SIM}}^{\mathrm{MIN}}$ and $\Delta\mathbf{A}_{\mathrm{SIM}}^{\mathrm{MAX}}$ were used to calculate the lower and upper bounds of estimated tracking errors, separating the transformation error from the error of the tracking video-based algorithm (Table 8).

Model	Min. transformation error (mm RMS)	Max. transformation error (mm RMS)	Min. algorithm error (mm RMS)	Max. algorithm error (mm RMS)
A	2.5 ± 0.002	11.6 ± 0.02	17.1	20.5
B	2.3 ± 0.02	9.5 ± 0.23	20.9	22.8
C	1.7 ± 0.01	6.3 ± 0.08	21.4	22.2

Table 8. Separation of total tracking error into transformation-induced error and algorithm-induced error. Values shown are mean estimates based on $\Delta\mathbf{A}_{\text{SIM}}^{\text{MIN}}$ and $\Delta\mathbf{A}_{\text{SIM}}^{\text{MAX}}$, for all captured trajectories

4.4 Discussion

Our experiment demonstrates the importance of trajectory transformation models and their effects on the estimated tracking error. The main difference between the presented models was the placement and the way of measuring the control point positions. In model A, control points $\mathbf{T}_1$, $\mathbf{T}_2$, $\mathbf{T}_3$ are measured at a safe distance from the stereo camera and its coordinate system centre (approximately 50 cm). When a measurement error is made, its effect on the calculation of camera orientation is much smaller than if the same measurement error is made at close distance to the coordinate origin. In models B and C the metal body of the camera distorted the measurements slightly, but close to the CS^{C} origin the prevailing errors appear again.

It is important to notice that an imprecise physical alignment of the stereo camera's two image sensors significantly corrupts the trajectory comparison. This is particularly obvious for models B and C, because they require manual alignment of control point $\mathbf{T}_2$ with the left optical axis. Any displacement of the optical axis can be verified during the calibration of the camera and should be corrected accordingly.

Relative sensitivity values in Table 4 show which parameters amplify the measurement errors the most, possibly causing a significant transformation error. In model A, such parameters are $\mathbf{T}_1$, $\mathbf{T}_3$, d_1 and d_2. In models B and C, parameters $\mathbf{T}_1$ and d are the most sensitive. Comparison of the upper relative sensitivity limits S^{MAX} (Table 5) also confirms that model A is the most sensitive, while model C is the least. But, we need to consider the fact that a highly sensitive parameter with low actual numerical value can have less impact on the transformation than a parameter with low sensitivity and large numerical value. For example, if a measurement error of a few millimetres is made at close distance to the camera (like control point $\mathbf{T}_3$ in model B), it would greatly distort the trajectory comparison, even if the parameter in question has very low sensitivity.

Consequently, the trajectory comparison results cannot be matched directly with the estimated S^{MAX}. In our experiments, the real parameter values generated such coordinate system transformations that model A produced the lowest tracking error, while model B performed the worst (Table 6, Fig. 8). On the other hand, the simulation of the transformation errors $\Delta\mathbf{A}$ determined the trajectories with the minimum and maximum errors (Table 8). In this context, model C corresponds to the lowest expected error, while model A to the biggest, which is consistent with the S^{MAX} predictions. Thus, a conclusion based on the model sensitivity about the accuracy of the proposed coordinate-system transformation is that model C is theoretically the least error-prone, while model A is the most.

Since the identical tracking algorithm data was used in all comparisons, the error of the tracking algorithm should appear the same for all three transformation models. Our best estimate of the total tracking error is 20.5 mm RMS, as obtained by model A (Table 6). The minimal and the maximal transformation errors of individual models can be used to decompose the total tracking error, resulting in estimates of the lower and upper bound of the algorithm error. For model A, the transformation error is estimated between 2.5 and 11.6 mm RMS, while the algorithm video-based error is between 17.1 and 20.5 mm RMS (Table 8).

Reference measurements of the Fastrak tracker also contain certain inaccuracies, but since their magnitude is significantly lower than transformation error or algorithm error, they are ignored and Fastrak measurements are considered error-free. However, in experiments where the magnetic sensor is more than 75 cm away from the magnetic transmitter, those errors should be considered and compensated accordingly.

5. Conclusion

When a magnetic tracker is used as a reference for vision-based tracking, a reliable transformation of their coordinate systems is crucial for proper tracking accuracy estimation. To address this issue in more detail, three different models of coordinate system alignment were developed. By analyzing the worst-case sensitivity of transformation models a limited comparison of those models is possible. The most influential model parameters are easy to identify and should be measured with special care. However, the actual parameter values used also have a significant effect on the final transformation error. With appropriate selection of parameter values, any model can be manipulated to produce the most accurate transformation. Therefore, such comparisons are only reasonable if the parameters are fixed to a certain setup of a camera and a magnetic tracker.

In our experiment the transformation model A resulted in the lowest total trajectory difference, despite being the most sensitive. Statistical separation of this error into estimates of the tracking algorithm’s error and the transformation-induced error provides more detailed discrepancy analysis.

The presented approach is applicable to any setup where the performance of video-based tracking is to be estimated by a reference device with its separate coordinate system. Experimentally determined parameter values and conclusions are valid only for the specific setup, but the proposed methodology can be applied to any similar problem.

6. References

Ascension (2007). *Ascension Technology Corporation*, Burlington, USA, http://www.ascension-tech.com/.

ART (2007). *Advanced Realtime Tracking GmbH*, Munich, Germany, http://www.ar-tracking.de/.

Balan, A.O.; Sigal, L.; Black, M.J. (2005). A Quantitative Evaluation of Video-based 3D Person Tracking. *The Second Joint IEEE International Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance*, Beijing, China, October 15-16, pp. 349-356.

Bashir, F. & Porikli, F. (2006). Performance Evaluation of Object Detection and Tracking Systems. *IEEE International Workshop on Performance Evaluation of Tracking and Surveillance*, June.

Black, J.; Ellis, T.; Rosin, P. (2003). A Novel Method for Video Tracking Performance Evaluation. *The Joint IEEE International Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance*, Nice, France, pp. 125-132.

Bornik, A.; Beichel, R.; Reitinger, B.; Gotschuli, G.; Sorantin, E.; Leberl, F.; Sonka, M. (2003). Computer Aided Liver Surgery Planning Based on Augmented Reality Techniques. *Workshop Bildverarbeitung für die Medizin*, Springer Verlag, pp. 249-253.

Bradley, D. & Roth, G. (2005). *Proceedings of the ACM SIGCHI International Conference on Advances in Computer Entertainment Technology*. Valencia, Spain, pp. 19 – 26.

Christensen, H.I. & Förstner, W. (1997). Performance characteristics of vision algorithms. *Machine Vision and Applications*, Springer-Verlag, Vol. 9, No. 5-6, pp. 215-218.

Collins, R.T.; Lipton, A.J.; Kanade, T. (2000). Introduction to the Special Section on Video Surveillance. *IEEE Transactions on PAMI*, Vol. 22, No. 8, pp. 745 - 746.

Collomosse, J.P.; Rowntree D.; Hall, P.M. (2003). Video Analysis for Cartoon-like Special Effects. *14th British Machine Vision Conference*, UK.

Cupillard, F.; Bremond, F.; Thonnat, M. (2003). Behaviour recognition for individuals, groups of people and crowd. *IEE Proceedings of the IDSS Symposium Intelligent Distributed Surveillance Systems*, February, London.

CVTI (2007). *Computer Vision Test Images (on-line collection)*. Calibrated Imaging Lab, Carnegie Mellon University, USA, http://www.cs.cmu.edu/~cil/v-images.html.

Divjak, M. (2005). *Evaluation of models and procedures for 3D tracking of human body using a stereocamera*. PhD Dissertation, Faculty of Electrical Engineering and Computer Science, University of Maribor, Slovenia.

Doermann, D. & Mihalcik, D. (2000). Tools and Techniques for Video Performance Evaluation. *Proceedings of the International Conference on Pattern Recognition*, Barcelona, Spain, September, pp. 4167–4170.

Ellis, T.J.; Black, J. (2003). A Multi-view surveillance system. *IEE Intelligent Distributed Surveillance Systems*, London, UK, February.

Fischer, J.; Neff, M.; Freudenstein, D.; Bartz. D. (2004). Medical Augmented Reality based on Commercial Image Guided Surgery. *Eurographics Symposium on Virtual Environments*, The Eurographics Association, Germany.

Fua, P. & Plankers, R. (2003). Articulated Soft Objects for Multiview Shape and Motion Capture. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 25, No. 9, pp. 1182 – 1187.

Gavrila, D.M. (1999). The visual analysis of human movement: A survey. *Computer Vision and Image Understanding*, Vol. 73, No. 1, pp. 82-98, Academic Press.

Georis, B.; Brémond, F.; Thonnat, M.; Macq, B. (2003). Use of an Evaluation and Diagnosis Method to Improve Tracking Performances. *The 3rd IASTED International Conference on Visualization, Imaging and Image Proceeding*, September 8-10, Benalmadena, Spain.

Grimson, E.; Leventon, M.; Ettinger, G.; Chabrerie, A.; Ozlen, F.; Nakajima, S.; Atsumi, H.; Kikinis, R.; Black, P. (1998). Clinical Experience with a High Precision Image-Guided Neurosurgery System. *Lecture Notes In Computer Science*, Springer-Verlag, London, UK, Vol. 1496, pp. 63 – 73.

Gueziec, A. (2002). Tracking pitches for broadcast television, *Computer*, IEEE, March, Vol. 35, No. 3, pp. 38-43.

Herda, L.; Fua, P.; Plankers, R.; Boulic, D.R.; Thalmann D. (2001). Using skeleton-based tracking to increase the reliability of optical motion capture. *Human Movement Science*, No. 20, pp. 313—341.

Huang, D. & Yan, H. (2002). Modeling and animation of human expressions using NURBS curves based on facial anatomy. *Proceedings of the 6th International Computer Science Conference*, Hong Kong, China, December 18-20, No. 17, pp. 457-465.

Jain, R.; Kasturi, R.; Schunck, B.G. (1995). *Machine Vision*, International Edition, McGraw-Hill Inc., 1995.

Jaynes, C.; Webb, S.; Steele, R.; Xiong, Q. (2002). An Open Development Environment for Evaluation of Video Surveillance Systems. *Proceedings of the Third International Workshop on Performance Evaluation of Tracking and Surveillance*, Copenhagen, June.

Keemink, C.J.; Hoek van Dijkel, G.A.; Snijders, C.J. (1991). Upgrading of efficiency in the tracking of body markers with video techniques. *Medical and Biological Engineering and Computing*, Vol. 29, No. 1, January.

Kindarenko, V. (2000). A Survey of Electromagnetic Position Tracker Calibration Techniques. *Virtual Reality*, Vol. 5, No. 3, September, pp. 169–182.

Kristan, M.; Perš, J.; Perše, M.; Kovačič, S. (2006). Towards fast and efficient methods for tracking players in sports. *Proceedings of the ECCV Workshop on Computer Vision Based Analysis in Sport Environments*, May, pp. 14-25.

Krumm, J.; Harris, S.; Meyers, B.; Brumitt, B.; Hale, M.; Shafer S. (2000). Multi-Camera Multi-Person Tracking for EasyLiving. *Third IEEE International Workshop on Visual Surveillance*, IEEE Computer Society, Washington, DC, USA, pp. 3.

La Cascia, M.; Sclaroff, S.; Athitsos, V. (2000). Fast, Reliable Head Tracking under Varying Illumination: An Approach Based on Registration of Texture-Mapped 3D Models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 22, No. 4, pp. 322 - 336.

La Cascia, M. & Sclaroff, S. (1999). Fast, reliable head tracking under varying illumination. *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 604-610.

Meyer, C. D. (2001). *Matrix Analysis and Applied Linear Algebra*. Society for Industrial and Applied Mathematics, Philadelphia, USA.

Needham, C.J.; Boyle, R.D. (2003). Performance Evaluation Metrics and Statistics for Positional Tracker Evaluation. *International Conference on Computer Vision Systems*, Graz, Austria, April, pp. 278-289.

NDI (2007). *NDI Optotrak Certus System*, NDI International, Waterloo, Canada, http://www.ndigital.com/.

NP (2007). *NaturalPoint Inc.*, Corvallis, Oregon, USA, http://www.naturalpoint.com/.

Optix (2007). *Optix 400 Series Laser Scanner*, 3D Digital Corp., USA, http://www.3ddigitalcorp.com/products.htm.

Pandya, A. & Siadat, M. (2001). Tracking Methods for Medical Augmented Reality. *Proceedings of Medical Image Computing and Computer-Assisted Intervention*, Utrecht, The Netherlands, October 14-17 Springer Berlin, pp. 1404-1405.

PCCV (2007). *Performance Characterization in Computer Vision*. PEIPA - Pilot European Image Processing Archive, http://peipa.essex.ac.uk/index.html.

PETS (2005). *IEEE International Workshop on Performance Evaluation of Tracking and Surveillance*, IEEE Winter Vision Multi-Meeting, Breckenridge, Colorado, USA, http://pets2005.visualsurveillance.org/.

Polat, E.; Yeasin, M.; Sharma, R. (2003). Robust tracking of human body parts for collaborative human computer interaction. *Computer Vision and Image Understanding*, Vol. 89, No. 1, pp. 44-69.

Polhemus (1998). *3Space Fastrak User's Manual*. Polhemus Inc., USA, http://www.polhemus.com.

Qiu, X.; Wang, Z.; Xia S. (2004). A Novel Computer Vision Technique Used on Sport Video, *Proceedings of International Conference on Signal Processing*, Vol. 2, No. 31 pp. 1296 – 1300.

Rehg, J.M. & Kanade, T. (1994). Visual Tracking of High DOF Articulated Structures: an Application to Human Hand Tracking. *3rd European Conference on Computer Vision*, Stockholm, Sweden, pp. 35-46.

Safeguards (2007). *Safeguards Technology Inc.*, Hackensack, USA, http://www.safeguards.com/.

Sato, Y; Oka, K.; Koike, H.; Nakanishi, Y. (2004). Video-based tracking of user's motion for augmented desk interface. *Proceedings of Sixth IEEE International Conference on Automatic Face and Gesture Recognition*, May 17-19, pp. 805- 809.

Scharstein, D. & Szeliski, R. (2003). High-accuracy stereo depth maps using structured light. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Vol. 1, pp. 195-202, Madison, WI, USA.

Stapleton, C.; Hughes, C.; Moshell, M.; Micikevicius, P.; Altman, M. (2002). Applying mixed reality to entertainment. *Computer*, IEEE, Vol. 35, No. 12, pp. 122 - 124.

Stoodley, K.D.C. (1984). *Applied and Computational Statistics: A First Course*. Ellis Horwood Ltd.

Tang, S.L.; Kwoh, C.K.; Teo, M.Y.; Sing, N.W.; Ling, K.V. (1998). Augmented reality systems for medical applications. *Engineering in Medicine and Biology Magazine*, IEEE, May/June, Vol. 17, No. 3, pp. 49-58.

Videre (2001). *STH-MD1/-C Stereo Head*. User's Manual, Videre Design Inc., USA, http://www.videredesign.com.

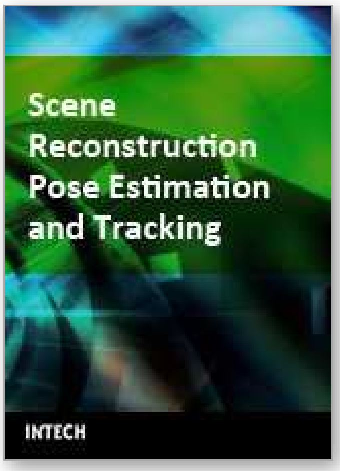
VIVID (2007). *VIVID 910 Series 3D Digitizer*, Konica Minolta Holdings, http://konicaminolta.com/.

Vogt, S.; Khamene, A.; Sauer, F. (2006). Reality Augmentation for Medical Procedures: System Architecture, Single Camera Marker Tracking, and System Evaluation. *International Journal of Computer Vision*, Springer Netherlands, Vol. 70, No. 2, November, pp. 179-190.

Wren, C.R. et al. (1997). Perceptive Spaces for Performance and Entertainment: Untethered Interaction using Computer Vision and Audition. *Applied Artificial Intelligence*, Taylor & Francis, Vol. 11, No. 4, pp. 267 – 284.

Xiao, J.; Moriyama, T.; Kanade, T.; Cohn, J. F. (2003). Robust full-motion recovery of head by dynamic templates and re-registration techniques. *International Journal of Imaging Systems and Technology*, No. 13, pp. 85-94.

Yao, Z. & Li, H. (2004). Is A Magnetic Sensor Capable of Evaluating A Vision-Based Face Tracking System? *Conference on Computer Vision and Pattern Recognition Workshop*, Vol. 5, Washington, USA.



Scene Reconstruction Pose Estimation and Tracking
Edited by Rustam Stolkin

ISBN 978-3-902613-06-6
Hard cover, 530 pages
Publisher I-Tech Education and Publishing
Published online 01, June, 2007
Published in print edition June, 2007

InTech Europe
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821