Identification of most influential factors in a Virtual Reality tracking system using hybrid method
Fabien Ezedine, Jean-Marc Linares, Wan Mansor Wan Muhamad, Jean-Michel Spraul

▶ To cite this version:

HAL Id: hal-01409793
https://hal-amu.archives-ouvertes.fr/hal-01409793
Submitted on 6 Dec 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Identification of most influential factors in a Virtual Reality tracking system using hybrid method

Fabien Ezedine¹, Jean-Marc Linares², Wan Mansor Wan Muhamad¹, Jean-Michel Spraul²

¹ Universiti Kuala Lumpur UNIKL - MFI, Malaysia, Seksen 14, Jalan Teras Jernang, 43650 Bandar Baru Bangi, Malaysia
² Aix-Marseille Université, CNRS, ISM UMR 7287, 13288 Marseille cedex 09, France

ABSTRACT
This paper studies the factors that influence the accuracy of Virtual Reality (VR) systems, in particular for applications in a Cave Automatic Virtual Environment (CAVE). The CAVE can be used to train surgeon students. For this purpose, an application for a total knee arthroplasty surgery is investigated. To meet the requirements for a high quality training, the accuracy of the tracking system in the CAVE has to be improved. First, a complete model of the tracking system is created based on the extrinsic and intrinsic parameters of the eight-camera system. With this model, the uncertainty of the tracking system is determined for one location in the CAVE. Next, a hybrid method, comprising the Monte Carlo Method and Design Of Experiment, is used to find the most important factors influencing the tracking accuracy in the CAVE.

1. INTRODUCTION
Carolina Cruz-Neira et al. [1] defined VR in 1992 as a "system which provides real-time viewer-centered head-tracking perspective with a large angle of view, interactive control, and binocular display".

The concept and the relation between the VR system and the user can be summarized as follows [2]:
- capture of user's actions;
- computation of these data and creation of a tailored response;
- transmission of the response toward the user.

This concept was greatly improved during the 80's thanks to the parallel evolution of computer science, and the development of new external devices in haptic interaction and visualisation. Since then, more and more applications are developed in manufacturing, research, teaching or therapeutics fields, through multiple devices such as the Head-Mounted Display (HMD), the BOOM or the Cave Automatic Virtual Environment (CAVE). This research work deals with the CAVE.

Figure 1. Model of location and orientations of the eight-cameras system in the CAVE.
Our CAVE consists of a four projection-sides environment (three walls and a ground), a set of cameras (eight cameras in the studied system) and four stereoscopic projectors. Figure 1 models approximate locations and orientations of cameras in the CAVE structure. Interactions between a numerical model, a system of projection, a system of tracking and a set of tracked spheres hung to user are created. A user which is immersed in this virtual world wears shutter glasses, in order to get a depth perception of the numerical model projected by projectors, via mirrors. Cameras are able to locate glasses, thus defining the field of vision of the user, through spherical markers tracked by these cameras shown in Figure 2. Then, this information is computed and sent back to the user. The technology used by the system in place strongly influences the quality of the immersion experienced and sensed by the user. These devices provide information in order to get the location of the user in the CAVE, using the tracking of optical markers [3].

Applications require a more or less good accuracy: for example, a visitor in a museum travelling inside an architecture, or a medical student training on a knee surgery, are differentiated by the precision needed to really feel the immersion inside this virtual world. Therefore, before loading an application inside a VR system, checking the compatibility between the precision required by the application and the real accuracy delivered is a compulsory step. If this process does not bring any convincing results, then two options are possible: either the application cannot be loaded in the CAVE with its proper features, or a method has to be found to improve the accuracy of the tracking system, which is the purpose of this paper.

The method, first, to model the tracking system in order to define the uncertainties of the captured coordinates: these uncertainties will be derived from a covariance matrix which is found using Monte Carlo simulations; and second, to identify the most influential factors using a hybrid method involving Monte Carlo simulations, Design of Experiment, Hadamard matrix and Bayesian analysis.

2. THE PIN-HOLE MODEL

In the tracking system, ARTtrack cameras are modelled using the principle of the pin-hole camera and projective geometry theory, which is shown in Figure 3. Linear algebra is used to provide the Cartesian coordinates \( M(u, v) \) in the CCD frame of the image of a marker tracked by each camera \( i \). These coordinates are derived from the position \( S(x, y, z) \) of the marker in the CAVE global coordinate system. These coordinates are given as a function of the extrinsic and intrinsic factors of the camera [4].

Extrinsic factors define the position and the orientation of each camera. The position of each camera is characterized by the coordinates of the lens centre \( O_i \). The orientation of each camera is defined by azimuth and inclination angles.

Intrinsic factors define the proper characteristics of the camera, such as the focus distance \( f \), the skew coefficient \( \alpha \), defining the angular error between the horizontal and vertical directions of pixels and the coordinates \( (u_0, v_0) \) of the intercept \( C_i \) (Figure 3) between the focus axis and the image plane of the camera \( i \).

The reference frames used in the model are \((i\) refers to the index of a camera, \(1 < i < 8\)):

- \( R_w(O, X, Y, Z) \) : bound to the CAVE, the global coordinates system;
- \( R_{cai}(O, X_{cai}, Y_{cai}, Z_{cai}) \) : bound to the camera \( i \);
- \( R_{im}(A_i, u_i, v_i) \) : bound to the image plane of the camera \( i \).

The characteristic points are defined as following:

- \( S(X, Y, Z) \) : coordinates of \( S \) into \( R_w \);
- \( S(x, y, z) \) : coordinates of \( S \) into \( R_{cai} \);
- \( M_i(u, v) \) : coordinates of \( M_i \) into \( R_{im} \);
- \( C_i(u_0, v_0) \) : coordinates of \( C_i \) into \( R_{im} \);

The pin-hole model is used to provide a realistic description of a camera. In this approach, the real image plane is shifted to the location \( z_{ci} = f = O_i C_i \) and reversed, making the analysis easier [5]. The pin-hole model allows defining the image \( M_i \) of
any tracked point \( S \) in the local reference frame of the camera. The coordinates \((u_i, v_i)\) of \( M_i \) are given as a function of the focus distance \( f \), the skew coefficient \( a \) and the scale factor \( b \) (Equation 1):

\[
\begin{bmatrix}
    x_i \\
    y_i \\
    z_i \\
    1
\end{bmatrix} =
\begin{bmatrix}
    f & a & n_{0i} & 0 \\
    0 & f & r_{0i} & 0 \\
    0 & 0 & 1 & 0
\end{bmatrix}
\begin{bmatrix}
    x_i \\
    y_i \\
    z_i \\
    1
\end{bmatrix}
\]

(Classical matrix transformations are used to derive the local coordinates of \( S \) from the fixed coordinates in the reference frame \( R_{ef} \) of the CAVE:

\[
\begin{bmatrix}
    x_i \\
    y_i \\
    z_i \\
    1
\end{bmatrix} = [T_m][R_m][X]
\]

(2)

\[
\begin{bmatrix}
    X \\
    Y \\
    Z \\
    1
\end{bmatrix} =
\begin{bmatrix}
    1 & 0 & 0 & -t_{xd} \\
    0 & 1 & 0 & -t_{yd} \\
    0 & 0 & 1 & -t_{zd}
\end{bmatrix}
\begin{bmatrix}
    x_i \\
    y_i \\
    z_i \\
    1
\end{bmatrix}
\]

and \([R_m] = [R_{yi}][R_{Xy}]\)

Considering, in Cartesian coordinates:

- \([T_m] \) : relative to the translation vector \( \vec{O}'\vec{O} \)
- \([R_{Xy}] \) : relative to the orientation of the camera around X-axis:

\[
[R_{Xy}] = \begin{bmatrix}
1 & 0 & 0 \\
0 & \cos(r_{xy}) & \sin(r_{xy}) \\
0 & -\sin(r_{xy}) & \cos(r_{xy})
\end{bmatrix}
\]

(3)

- \([R_{yi}] \) : relative to the orientation of the camera around Y-axis, using axial rotation (quaternion theory). Considering \( n_i(n_{xi}, n_{yi}, n_{zi}) \), the transformation matrix is:

\[
[R_{yi}] = \cos(r_{yi}) \cdot \begin{bmatrix}
1 & 0 & 0 \\
0 & 1 & 0 \\
0 & 0 & 1
\end{bmatrix}
+ (1 - \cos(r_{yi})) \cdot \begin{bmatrix}
0 & -n_{yi} & n_{zi} \\
-n_{zi} & 0 & n_{yi} \\
-n_{yi} & n_{zi} & 0
\end{bmatrix}
+ \sin(r_{yi}) \cdot \begin{bmatrix}
0 & -n_{zi} & n_{yi} \\
-n_{yi} & 0 & -n_{zi} \\
-n_{zi} & n_{yi} & 0
\end{bmatrix}
\]

Some distortions caused by the lens have to be considered in order to extend the pin-hole model. The vector distortion is \((A_{x}, A_{y})\). These distortions can be radial, image decentering, and prismatic. Equation (5) describes the detailed model R3D1P1, a non-linear polynomial distortion model for cameras [6]:

\[
\begin{align*}
\Delta_x &= x_{M_i} + \sum_{i=1}^{3} r_{xi} \rho_{x_i}^2 + d_{x,i}(2x_{M_i}^2 + \rho_{x_i}^2) + 2d_{x,i}x_{M_i}y_{M_i} + p_{x,i}\rho_{x_i}^2 \\
\Delta_y &= y_{M_i} + \sum_{i=1}^{3} r_{yi} \rho_{y_i}^2 + d_{y,i}(2y_{M_i}^2 + \rho_{y_i}^2) + 2d_{y,i}y_{M_i}x_{M_i} + p_{y,i}\rho_{y_i}^2 \\
\end{align*}
\]

(5)

where:

\[
\rho_{x_i}^2 = (x_{M_i}^2 + y_{M_i}^2) \]

Then, Equation (6) describes the full model providing the coordinates of the image of a tracked marker, inside the CCD frame of the camera \( i \) as a function of \((X,Y,Z)\):

\[
\begin{align*}
u_i &= X_{r_{xi}1} + Y_{r_{xi}2} + Z_{r_{xi}3} + t_{xi} + x_{r_{xi}1} + Y_{r_{xi}2} + Z_{r_{xi}3} + t_{xi} + X_{r_{yi}1} + Y_{r_{yi}2} + Z_{r_{yi}3} + t_{yi} + \Delta_{x_i} + n_{ui} \\
v_i &= X_{r_{yi}1} + Y_{r_{yi}2} + Z_{r_{yi}3} + t_{yi} + X_{r_{yi}1} + Y_{r_{yi}2} + Z_{r_{yi}3} + t_{yi} + \Delta_{y_i} + p_{yi}
\end{align*}
\]

(6)

3. TRACKING UNCERTAINTY COMPUTATION

In the pin-hole model, the intrinsic and extrinsic parameters are assumed to be perfectly defined but it is not really the case. In fact, these factors are calibrated by preliminary experiments and, therefore, are only ranged in a given uncertainty interval. For this reason, the real images of any marker captured by the camera do not fit the position calculated from the nominal parameters but are subjected to some deviations. Therefore, the approximated 3D coordinates of the tracked point computed by the multiple camera system do not correspond to the real ones. Our aim is consequently to appraise the uncertainties of the 3D coordinates evaluated by the tracking system. For that purpose, the intrinsic and extrinsic parameters are randomly perturbed assuming a uniform distribution in their uncertainty intervals, in order to be placed in the most critical case [7]. Each simulation is assumed to correspond to a real configuration of the tracking system. For this end, the coordinates of the images \( M_i \) as they would be captured by the eight cameras are
computed. The theoretical positions of \( M_i \) are also derived from the nominal mean parameters. The difference between both points is called projection error \( (e_i) \). As in a real experimental configuration, the coordinates of the tracked point \( S \) are supposed unknown. After computing the projection error for each camera using the full model, the least squares method is, therefore, used to evaluate the location of \( S \) in \( R_W \). The deviation \( \Delta M = (dX, dY, dZ) \) between the real and evaluated locations of \( S \) is also computed.

This calculation procedure is repeated 30000 times using a Monte Carlo simulation approach \cite{8}. This method provides the variance covariance matrix \( V_M \) of the deviations \( \Delta M \). The error zone, an ellipsoidal shape, is also created. If the eigenvalues and eigenvectors of the variance covariance matrix are calculated, the dimension and orientation of the error zone can be computed. A CAVE with eight tracking cameras counts until one hundred and twenty eight significant factors, sixteen per camera \cite{5}. In order to improve the accuracy of the tracking system and consequently the quality of the user immersion in the CAVE, it is important to target adjustable factors which have a significant influence.

### 3.1. The hybrid method

As shown in Figure 4, the flowchart explains the first step of this hybrid method which is to detect adjustable factors. It is assumed that intrinsic factors cannot be modified and/or adjusted by the user. Indeed, the cameras used are off-the-shelf ones from ARTtrack and the only factors we can work on are the extrinsic ones. Moreover, preliminary calculations have shown that the distortion factors of the camera do not have significant effect on the components of the variance covariance matrix \( V_M \). Therefore, only effects of the position and the orientations of the cameras are studied.

The cameras are fixed on the framework of the CAVE. The position referring to the camera \( I \) is defined as the location of lens optic centre using Cartesian coordinates in \( R_W \) (CamiTX, CamiTY, CamiTZ). The orientation of the optic axis is defined by CamiRX and CamiRY. Depending on the location of cameras inside the framework, all parameters cannot be adjusted. Then, only twenty six extrinsic factors are adjustable (Cam1TY, Cam2TZ, Cam3TX, Cam3TZ, Cam4TX, Cam5TX, Cam6TX, Cam6TZ, Cam7TZ, Cam8TY) and CamiRX, CamiRY for \( 1 \leq i \leq 8 \). These factors of position and orientation of cameras are detailed in Figure 5, where \( C_i \) refers to the location of the camera \( i \).

The processing for adjustable and non-adjustable factors is different. In the real configuration of the CAVE, the non-adjustable factors \( X_i \) are fixed and calibrated during manufacturing. However in order to compare the effects of the adjustable factors and the unknown influences of the internal parameters of the cameras, the later values were generated randomly in the Monte Carlo simulation, leading to random perturbations of the calculated results. This approach permitted the application of statistical tests to discriminate the significant adjustable factors.

The effects of the adjustable factors \( X_i \) are studied by a DOE methodology \cite{10}. For this purpose, only the lower and higher values of the factors were considered and normalized to \((-1,1)\). In a classical approach, it would lead to \( 2^{26} \) simulations. But, in order to decrease the number of random generations, an optimized strategy is used, based on Hadamard design matrix of dimension \( 28 \times 26 \) \cite{11}. The DOE analysis leads to twenty eight Monte Carlo simulations.

Two responses \( (8) \), function of factors \( X_i \), are provided by the covariance matrix after Monte Carlo simulations: the average dimension of the error zone, \( Y_1 \); the distortion of the error zone, \( Y_2 \), such as:

![Figure 4. Flowchart of Hybrid Method (MCM, DOE).](image)

![Figure 5. Translation and rotation adjustments available for each camera.](image)
\[
Y_1 = m = \frac{1}{3} J_r(V_M) \\
Y_2 = \frac{3}{\sqrt{2}} \sum D^i_y
\]

where \(D = V_M - mI\) (8)

A screening process is used to detect the influential factors. The screening models are linear systems which link \(Y_1, Y_2\) the response vectors to \(X\), the experimental matrix. Vectors \(B_1, B_2\) of components \(b_{1j}, b_{2j}\) defined the coefficients of the model and \(E_1, E_2\) the error vectors:

\[
\begin{align*}
Y_1 &= X.B_1 + E_1 \\
Y_2 &= X.B_2 + E_2
\end{align*}
\]

The error vectors \(E_1, E_2\) describe two kinds of deviations: the best fit approximation error and the random perturbations introduced by the non-adjustable factors. Due to the large number of intrinsic and extrinsic parameters of the Monte Carlo simulation, the distribution of these errors becomes practically Gaussian, even if the random generations used a uniform repartition.

These systems get a number of equations greater than the number of unknowns. The least squares method is therefore used to solve the system. The estimates \(\hat{B}_1, \hat{B}_2\) of vectors \(B_1, B_2\) are obtained:

\[
\begin{align*}
\hat{B}_1 &= (X^T.X)^{-1}.X^T.Y_1 \\
\hat{B}_2 &= (X^T.X)^{-1}.X^T.Y_2
\end{align*}
\]

Then, statistical tools are used to create a statement of factors, concerning their influence on responses.

Figure 5 summarizes the whole hybrid method used to highlight the most influential factors of the tracking system in the CAVE.

### 3.2. Statistical tools

Two tools are used to analyze and state the influence of the factors: the graph of effects and the Bayesian analysis.

Regarding the graph of effects, it is assumed that \(b_{1j}, b_{2j}\) coefficients are propagated using a Student law. The degree of freedom is one, because twenty seven factors \(b_i\) and twenty eight different equations given by the Hadamard matrix are considered. With 5% risk, the Student table provides:

\[t_{0.025} = 12.7\]

95% of \(b_{ij}\) values belong to this range of confidence:

\[
[-t_{0.025} \cdot \text{var}(\hat{b}_{ij}) + t_{0.025} \cdot \text{var}(\hat{b}_{ij})]
\]

Then, in order to look after the range of confidence of \(b_{1j}, b_{2j}\) coefficients, the error matrix of two answers \(Y_1, Y_2\) are computed as follow:

\[
\begin{align*}
V &\cdot \text{AR}(\hat{B}_1), V &\cdot \text{AR}(\hat{B}_2)
\end{align*}
\]

Then, \(V\cdot \text{AR}(\hat{B}_1), V\cdot \text{AR}(\hat{B}_2)\) are computed using these equations:

\[
\begin{align*}
V &\cdot \text{AR}(\hat{B}_1) = (X^T.X)^{-1} \cdot \sigma^2_{\hat{b}_{1j}} \\
V &\cdot \text{AR}(\hat{B}_2) = (X^T.X)^{-1} \cdot \sigma^2_{\hat{b}_{2j}}
\end{align*}
\]

Regarding the Bayesian analysis, the principle of the test is to consider a posteriori that every factor is active \[12][13\]. Two parameters are considered:

- the probability that a factor is active a priori (\(p\));
- the ratio between variances of active factors and variances of non-active factors (\(q\)).

Then, the aim is to calculate probabilities a posteriori for any combination (\(p, q\)) ranging : \(0.1 \leq p \leq 0.4\) for \(p\) and \(5 \leq q \leq 20\) for \(q\).

<table>
<thead>
<tr>
<th>Factor</th>
<th>Domain</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cam1TY</td>
<td>[-4;4]</td>
</tr>
<tr>
<td>Cam2TZ</td>
<td>[-8;0]</td>
</tr>
<tr>
<td>Cam3TY</td>
<td>[0;8]</td>
</tr>
<tr>
<td>Cam4TX</td>
<td>[0;8]</td>
</tr>
<tr>
<td>Cam4TX</td>
<td>[-8;4]</td>
</tr>
<tr>
<td>Cam5TX</td>
<td>[-8;4]</td>
</tr>
<tr>
<td>Cam6TX</td>
<td>[-8;0]</td>
</tr>
<tr>
<td>Cam7TZ</td>
<td>[0;8]</td>
</tr>
<tr>
<td>Cam8TY</td>
<td>[-4;4]</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Factor</th>
<th>Domain</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cam1RY</td>
<td>[-10;10]</td>
</tr>
<tr>
<td>Cam2RY</td>
<td>[-10;10]</td>
</tr>
<tr>
<td>Cam3RY</td>
<td>[-10;10]</td>
</tr>
<tr>
<td>Cam4RY</td>
<td>[-10;10]</td>
</tr>
<tr>
<td>Cam5RY</td>
<td>[-10;10]</td>
</tr>
<tr>
<td>Cam6RY</td>
<td>[-10;10]</td>
</tr>
<tr>
<td>Cam7RY</td>
<td>[-10;10]</td>
</tr>
<tr>
<td>Cam8RY</td>
<td>[-10;10]</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Factor</th>
<th>Domain</th>
</tr>
</thead>
<tbody>
<tr>
<td>CamRX</td>
<td>[-2;2]</td>
</tr>
<tr>
<td>CamRY</td>
<td>[-2;2]</td>
</tr>
</tbody>
</table>

Figure 6. Bayesian analysis of Y1.
3.3. Adjustable factor settings

The location of the marker studied in Monte Carlo simulations is (0, 1200, 0), all coordinates are given in millimetres.

The model was programmed in Visual Basic. Depending on the characteristics of the computers used, a simulation may take four to six hours.

The domain of every factor has to be chosen. As said previously and concerning the positions of cameras, their location influences this domain, whereas the domain of factors about orientation remains the same. The domains per adjustable factor are shown in Table 1 and Table 2. The unit used for the position is the millimetre and the one used for orientation is the degree. These variations are given as a function of the theoretical location and orientation of cameras inside the reference frame of the CAVE as shown in Figure 5.

4. RESULT ANALYSIS

4.1. Analysis for Y1

These two statistical analyses are strongly similar and lead to the same findings:
- The positions of camera inside the CAVE do not influence the dimension of the error zone;
- Only the orientations of cameras are adjustable factors influencing the dimension of the error zone.
- The symmetry of the CAVE is respected;
- The influential factors over Y1 are the rotations: Cam1RY, Cam2RX, Cam3RY, Cam4RX, Cam5RX, Cam6RY, Cam7RX and Cam8RY.

In the case of Cam4RY and Cam5RY, a reflection has to be done as the geometrical symmetry of the CAVE is not respected.

As per Figure 6 and Figure 7, the statistical analysis shows that the orientations of each camera might have an influence on the dimension of the error zone.
4.2. Analysis for Y2

As the response mentioned above, the two analyses lead to the same findings:

- The positions of camera inside the CAVE do not influence the dimension of the error zone;
- Only the orientations of cameras are adjustable factors influencing the dimension of the error zone.
- The symmetry of the CAVE is respected;
- The influential factors over $Y_2$ are the rotations: Cam1RX and Cam8RX.

As per Figure 8 and Figure 9, the statistical analysis shows that the orientations of cameras 2, 5, 6 and 7 might have a low influence over the distortion of the error zone.

Cameras 1 and 8 strongly influence the distortion of the error zone in comparison to others.

5. CONCLUSION

This study focused on the tracking system of a Cave Automatic Virtual Environment. In order to improve the accuracy of any position captured by the system a complete modelling of the cameras has been developed. This approach permitted the simulation of the tracking device of the CAVE. Using a hybrid method based on a DOE and Monte Carlo simulations, the effect of the positions and the orientations of the cameras over distortion and average dimension of the error zone were studied. The final result revealed that the only adjustable factors which has to be considered for a strong influence on the dimension and the distortion of the error zone are the orientations of cameras.

The next step of this research will be to analyse the possible interactions between the key factors and to create the response surfaces.

ACKNOWLEDGEMENT

This work was realized in collaboration with the Centre de Réalité Virtuelle de Marseille (CRVM) in France. We thank all members of CRVM team, especially its Director Prof. Daniel Mestre and Engineers Pierre Mallet, Vincent Perrot and Jean-Marie Pergandi, for their help and great support.

This research work was also supported by an STRG grant from the Universiti Kuala Lumpur, UNIKL.

REFERENCES


