Automatic 3D scan view registration using 6D scene differentials

Joris Vergeest
Delft University of Technology
Delft, The Netherlands
j.s.m.vergeest@tudelft.nl

Wolf Song
Delft University of Technology1st
Delft, The Netherlands
y.song@tudelft.nl

ABSTRACT

Commonly the determination of one coordinate system relative to another is based on the equivalence of scenes measured in the two frames. If 3D scene data is sufficiently accurately available in both frames, the calibration in 6D configuration space can even be carried out fully automatically. However, in some applications, no scene data is explicitly available, but merely differentials in 6D space are known. One important application is automatic registration of 3D scan views, using data from an external placement measurement device, but without using markers. In this paper we show that from two linearly independent scene differentials, each measured relative to two frames, a calibration of the frames can be achieved. The problem can be reduced to minimization of a bi-variate function of two real parameters. We have investigated the conditions in which the method is sufficiently accurate. We report on this investigation and describe the application of the technique to quick 3D scanning of hand-held objects.

Keywords

Alignment of 3D scenes, scan view registration, automatic calibration, practical application

1. INTRODUCTION

Finding the relative placements of coordinate systems, or frames, in 3D space is a common issue in applications such as robotics, vision, augmented realty and 3D scanning. In most applications a calibration of two devices is based on the equivalence of geometric data recorded relative to each frame. If the data is sufficiently accurate, the transformation matrix for the two frames can be directly calculated or numerically approximated [Grasset 2001], [Wheeler 1998]. We apply this principle to scan view registration, which involves finding the transformation matrix to align one set of point data to another, as to merge the two point sets (or surface meshes) into a single set. Additional point sets could be merged subsequently as to obtain a representation of the entire outer surface of the scanned object. However, in general the alignment transformation cannot be directly calculated from two point data sets unless the correspondence is known of at least three points from one set to three points from the other set. There are several ways to supply this information to the registration software. In common practice the system prompts the user to designate correspondences interactively on the computer's screen. If large objects are being scanned, the object is usually left untouched and the scanning device is moved around the object to take different scan views. If the scanning device's position and orientation are tracked then the alignment matrix can be derived from the tracking data. When the scanner is mounted on a mechanical arm, the tracking data is very accurate and all scan views can be merged immediately. However, for scanning relatively small objects, usually the scanner is at rest and the object is manually repositioned for every scan view taken. Or, as for example with the Handyscan, [Han 2007] both the object and the scanner are moved between takes. In each of the latter cases either the user should supply correspondence information, as mentioned, or artificial features must be attached to the object. These features could be visual marks detectable by the scanner software.

We have developed an automatic registration method to support the scanning of handheld objects, where the scanning device, a Minolta Vivid 700 [Min 2006] is fixated to the ground. A 6 degrees-of-freedom (DoF) sensor, called Flock of Birds (FOB) from Ascension Tecnology [Asc 2006] has been attached to the object as to supply position and orientation data of the object to the system. The sensor can be fixed to any part onto or inside, the object [Vergeest 2007]. A schematic picture of the setup is presented in Figure 1. The sensor supplies 6D placement information associated to each scan view taken. However, since that placement data are relative to a frame differing from the frame in which the scanned points are measured, a calibration is needed to determine the transformation from scanner frame to sensor frame.

In this paper we present a mathematical formulation of the calibration. We describe a numerical method to determine the transformation matrix, we report on an analysis of the accuracy and present results of the application to practical 3D scanning.

2. CALIBRATION METHOD

A scanning device captures points from the surface of a physical object. Let V_i represent the points (or facets) captured during take *i* and V_{i+1} the points captured during the next take, *i*+1. To combine these two sets into a single set requires finding the transformation *A* such that V_i and AV_{i+1} get aligned, *i.e.* the two sets regain their "true" relative placement. The process of aligning and then merging of scan views is called scan view registration. Matrix *A* can be regarded as a discrete scene differential of the scenes V_i and V_{i+1}.

It is assumed that the scanning device, having reference frame *S*, is stationary relative to the ground. The object to be scanned can be hold in a different placement (position and orientation) for scan view *i*+1 compared to its placement for scan view *i*. If there happens to be no difference between the two placements then *A* would equal the identity transformation. If the difference between the two placements is not too large then *A* can be approximated by any shape matching algorithm commonly used to process scanner data. In general, a shape matching algorithm would find a local optimum of scan view alignment, which is not the optimal alignment. In that case, as mentioned, additional input is required, *e.g.* from the user.

In our setup, a 6 DoF sensor has been attached to the scanned object. The sensor continuously transmits data representing its position and orientation relative to the sensor's frame of reference *T*, where we assume that *T* is stationary relative to the ground.

We denote the sensor data corresponding to scan views *i* and *i*+1 by F_i and F_{i+1}, respectively. F_i defines the placement of the sensor relative to *T* at the time scan view *i* is taken. There exist a unique transformation *B* such that $F_i = BF_{i+1}$.

Both *A* and *B* represent the change of placement (translation and rotation) that would be needed to return the object from its placement during scan *i*+1 back to its placement during scan *i*. However, *A* and *B* are specified in different frames of reference (namely *S* and *T*) and therefore are in general not equal. Since *A* and *B* represent the same transformation they are called similar transformations. For any pair of similar transformations *A* and *B* there exist transformations *X* such that

$$X\,BX^{-1} = A. \qquad (1)$$

If we denote transformations in 3D space by the usual 4×4 matrices, then the matrices *A* and *B* have the same eigenvalues, and, as Henri Poincarré observed, the total amount of rotation defined by the two matrices are equal [Vergeest 2007]. It was also found by Poincarré that the rotation axis direction can be calculated from the transformation matrix, and we denote the direction vectors implied by *A* and *B* by ${}^{S}a$ and ${}^{T}a$, respectively. The upper index denotes the frame of reference. Since *A* and *B*, being 4×4 matrices contain translation information, the actual rotation axes can be computed as well, and we denote these lines by ${}^{S}l$ and ${}^{T}l$, respectively.

As mentioned, *S* and *T* are both stationary relative to the ground, but as yet still unknown. In classical calibration problems we would have data ${}^{S}z_i$ and ${}^{T}z_i$ representing one particular object in *S* and *T*, respectively. From that data the calibration matrix ${}^{S}T$ would be derived by solving an equation of the form ${}^{S}z_i = {}^{S}T\ {}^{T}z_i$.

The matrix ${}^{S}T$ represents the placement of the sensor's frame *T* relative to the scanner's frame *S*. If we would find ${}^{S}T$ we can derive the change of placement relative to *S* from the change of placement relative to *T* and hence predict the registration of scan views based on data from the sensor.

In our problem at hand, we do have explicit data ${}^{S}z_i$, in the form of scanned data points relative to *S*.. However, we do not have corresponding data points relative to *T*.

Instead we can obtain the transformations *A* and *B*. For the purpose of calibration, we chose views V_i and V_{i+1} such that the scanner's surface alignment software can compute *A*. *B* can be simply computed as $B = F_{i+1}\,(F_i)^{-1}$. By realizing that

$A = {}^{S}F_i\ {}^{Fi+1}S$ and

$B = {}^{T}F_i\ {}^{Fi+1}T,$

it can be seen that if $X = {}^{S}T$ then equation (1) holds. However, this solution is not unique. In other words, from the two transformations *A* and *B* alone we cannot derive the calibration matrix ${}^{S}T$. It can be

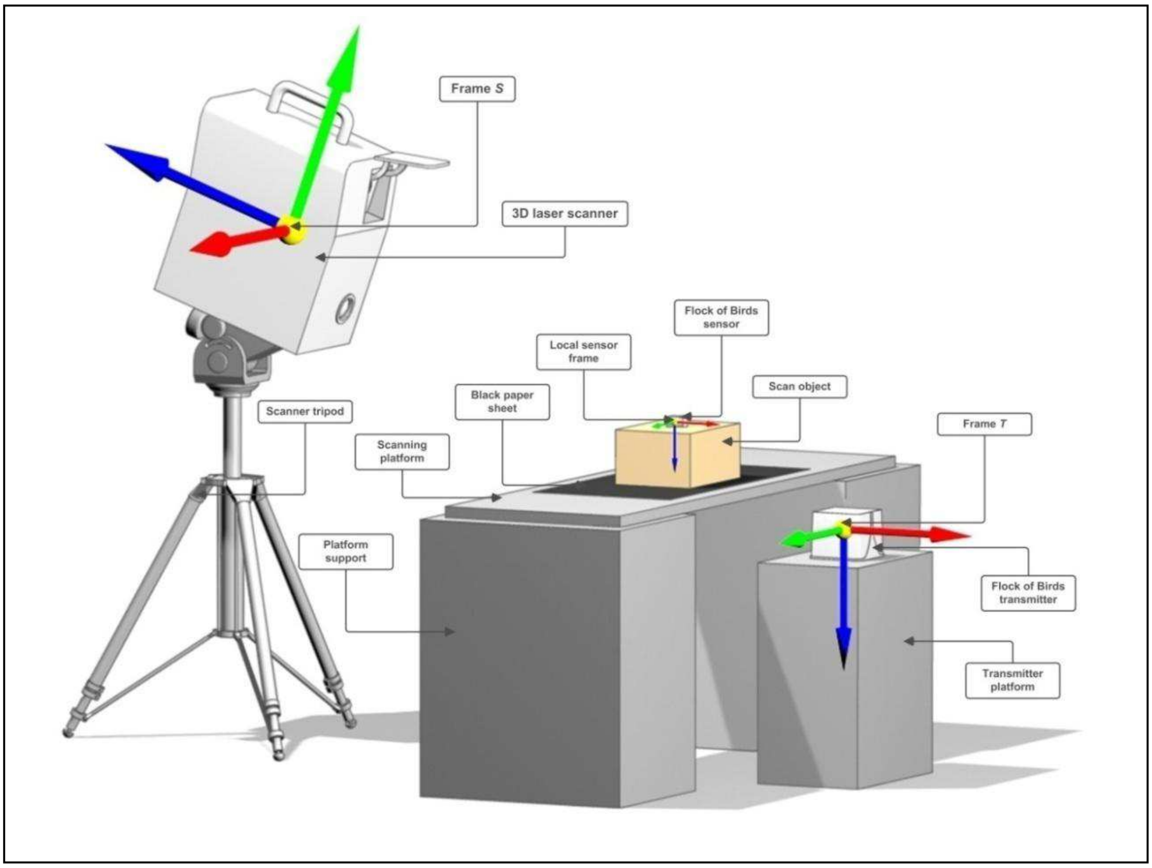

Figure 1. Setup of the accuracy measurements. The FOB's sensor is attached to the scan object. Scene differences are measured relative to frame *S* (of scanner) and to frame *T* (of FOB transmitter).

shown that it should hold for X that it 1) transforms ${}^{T}a$ to ${}^{S}a$ and 2) transforms any point in line ${}^{T}l$ to a point somewhere in line ${}^{S}l$. And this should hold for *all* possible pairs (A, B).

The key of our method is that, based on this knowledge, we can construct coordinate frames ${}^{S}L$ and ${}^{T}L$ and apply the "classical" calibration process using the frames as scene data measured relative to S and T. Based on A and B, however, the frames ${}^{S}L$ and ${}^{T}L$ can only be partially defined. Therefore, at least two pairs (A, B) must be measured in order to resolve for X.

It holds that

$$X = Y(\delta', \gamma'),$$

for some (still to be determined) values δ' and γ', where

$Y(\delta, \gamma) = {}^{S}L\ ({}^{T}L\ (\delta, \gamma))^{-1}$, where

$${}^{T}L\ (\delta, \gamma) = {}^{T}L \begin{pmatrix} \cos\gamma & -\sin\gamma & 0 & 0 \\ \sin\gamma & \cos\gamma & 0 & 0 \\ 0 & 0 & 1 & \delta \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

and ${}^{S}L$ and ${}^{T}L$ represent frames constructed as having their origin on any point in the axes ${}^{S}a$ and ${}^{T}a$ of rotation defined by A and B, respectively and with z-axis pointing into the directions of these rotation axis. For a full description we refer to [Vergeest 2007]. The angle γ specifies the orientation of ${}^{T}L\ (\delta, \gamma)$ about its z-axis and δ the location of its origin in line ${}^{S}l$, relative to frame ${}^{T}L$.

Numerically, δ' and γ' are determined by means of (A', B'), a *second pair* of measured transformations, yielding axes of rotation ${}^{S}l'$ and ${}^{T}l'$ and directions ${}^{S}a'$ and ${}^{T}a'$. Then

$$(\delta', \gamma') = \arg\min |f(\delta, \gamma)|^2 + |g(\delta, \gamma)|^2 , \quad (2)$$

where

$f(\delta, \gamma) = |Y(\delta, \gamma)\ {}^{T}a' - {}^{S}a'|$

$g(\delta, \gamma) = (({}^{S}a' \times (h' \times {}^{S}a')).h')^{1/2}$

$h' = Y(\delta, \gamma)\ {}^{T}d' - {}^{S}d'$,

where ${}^{S}d'$ and ${}^{T}d'$ are the origins chosen for the construction of ${}^{S}L'$ and ${}^{T}L'$, respectively. $f(\delta, \gamma)$ is the size of the difference between the two unit direction vectors of rotation. $g(\delta, \gamma)$ is the distance of point

$Y(\delta, \gamma)^T d$ to line $^S a'$. Note that the function $Y(\delta, \gamma)$ is the one determined from the first (not the second) measurement.

Equation (2) can be solved numerically using any 2-parameter nonlinear minimization method. We have used the Levenberg-Marquardt algorithm [Lourakis 2004] where we set the number of parameters to two and the number of output functions to two as well, with output functions $|f(\delta, \gamma)|$ and $|g(\delta, \gamma)|$.

The method can be summarized by the following 9 steps:

1. Take two scan views
2. Determine A and B
3. Compute $^S L$, $^T L$, $Y(\delta, \gamma)$
4. Take two new scan views
5. Determine A' and B'
6. Compute $^S a'$, $^T a'$, $^S d'$ and $^T d'$
7. Minimize both $| Y(\delta, \gamma)\, ^T l' - {}^S l' |$ and $| Y(\delta, \gamma)\, ^T d' - {}^S d' |$, thus finding δ and γ, using equation (2)
8. From δ and γ calculate $^S T = X$
9. Optionally repeat steps 4-9 for renewed computation of X.

3. ACCURACY OF THE INDIVIDUAL DEVICES

To evaluate the accuracy of the calibration method we first analyzed the matrices A and B themselves under two conditions, 1) no change in placement between V_i and V_{i+1} and 2) a small displacement involving a rotation of approximately 5 degrees. The object for scanning was a simple easy-to-scan part, to which the FOB's sensor was attached. Condition 1) was implemented by taking 11 scan views of the same object without moving the object at all. The 11 placements from the FOB were recorded as well. We determined the displacement of scan views 2...11 relative to the very first scan view. As expected the 10 measurements of A and B resulted in matrices close to identity, however, the FOB data appeared less accurate than 3D scan registration. The total intrinsic rotations of the A matrices (which should be zero, theoretically) were found between 0.00 and 0.04 degrees, whereas the B matrices show variations between 0 and 0.98 degrees, approximately 25 times as wide, see Figure 2.

Condition 2 was created as follows. Scan views V_i and V_{i+1} were taken, where V_{i+1} was slightly displaced relative to V_i. From the two scenes the matrices A and B were derived. Without moving the object anymore, 19 more scan views were taken and the FOB placements were recorded as well. Using these data, 19 more matrices A and B were derived as displacements relative to V_i. The variations among the matrices A and among the matrices B provide an indication of their accuracy. The histograms of intrinsic rotation of the A and B matrices are shown in Figure 3.

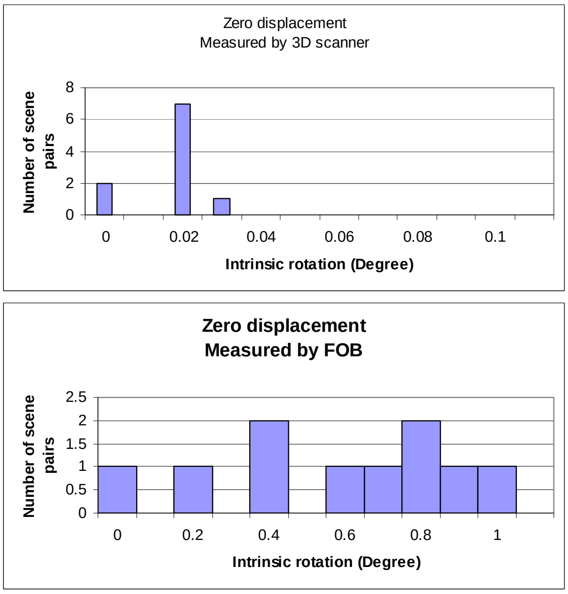

Figure 2. Intrinsic rotation of matrices *A* (top) and *B* (bottom), for fixed scenes. Please note the difference in scale of the horizontal axis scales.

The spread of intrinsic rotation appears to be 0.03 degrees for the A matrices and 0.50 degrees for the B matrices. The relatively large spread in the FOB data was observed earlier [Kroes 2007] and is partly attributed to metal objects near the FOB's sensor. The experiment also confirmed that 3D scan view registration is robust for scenes differing by rotations as small as 5 degrees.

The axis of rotation cannot be accurately derived in condition 1), since without rotation the axis would be undefined. In condition 2, we could determine the spread of the 11 rotation directions in terms of solid angular deviation from the average direction vector. The results are shown in Figure 4. The rotational component of the registration of 3D scan views appears about twice more accurate than those of the FOB's displacement measurement.

To further analyze this difference in accuracy we determined the 11 directions of the rotation axes relative to a reference frame having its average rotation direction as its z-direction and its x-direction defined as the cross product of the z-direction and vector $(0, 0, 1, 0)^T$. The projections of the unit rotation axis direction vectors onto the xy-plane are presented in Figure 5, for both the A and the B matrices.

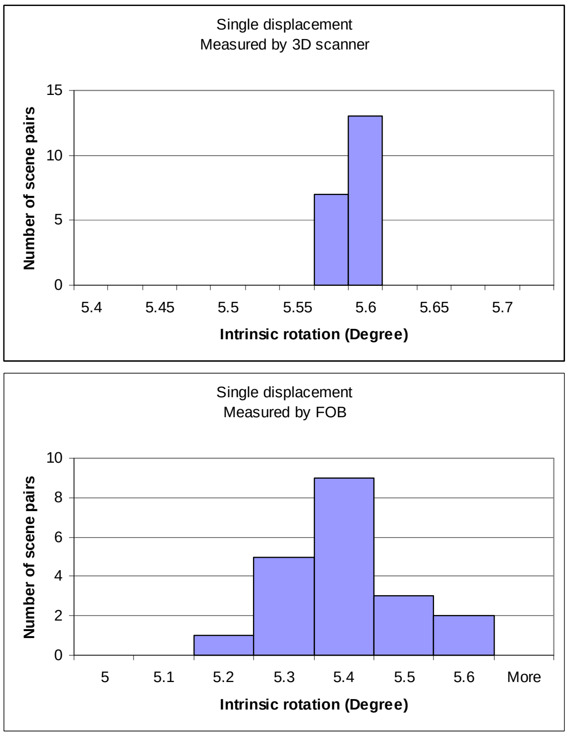

Figure 3. Intrinsic rotation of matrices *A* (top) and *B* (bottom), for slightly displaced scenes.

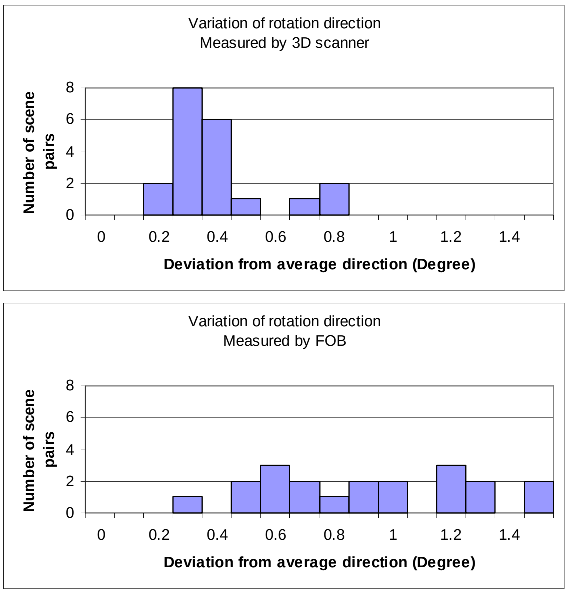

Figure 4. Variation of the direction of the axis of intrinsic rotation for *A* matrices (top) and *B* matrices (bottom).

The spread of the measured rotation axis can be almost fully attributed to the variation of the *x*-component, which is approximately horizontal, from left to right relative to the scanner's reference frame (see Figure 1). This asymmetry of precision can be partly explained by the shape of the geometry of the measured object, which indeed extended mostly in the *y*-direction, and therefore provides better angular registration precision in that direction. In contrast, the FOB deviations are more evenly spread over the *xy*-plane.

Finally we determined the spatial distance among the rotation axes. Here the 4th column of the *A* and *B* matrices come into play [Vergeest 2007]. The axes of rotation form a bundle of lines, nearly parallel, and nearly coinciding. The degree to which the lines are parallel is already depicted in Figures 4 and 5. The distance between the lines appears to be less than 1mm when derived from the scanner's registration matrices *A*, see top of Figure 6. However, the FOB's sensor produces lines which are up to 20mm apart from each other.

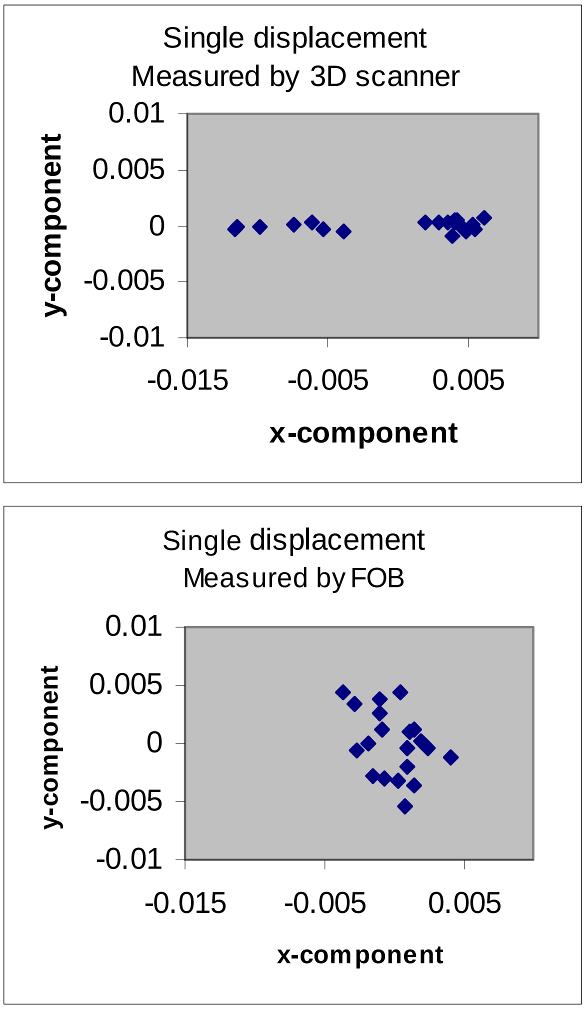

Figure 5. *xy*-variation of the unit rotation axis direction vector for *A* matrices (top) and *B* matrices (bottom).

From this brief accuracy analysis we conclude that a displacement matrix is more precisely determined by registration of 3D scan views than by computing the relative placement of the 6D sensor, for the particular types of devices we used. The spread in rotation directions for the FOB system is about 2 times as large compared to 3D scan registration, whereas the spread in intrinsic rotation angle, and also in distance among rotation axes is about 20 times larger for the FOB compared to 3D scan registration.

Another question is to which extent this inaccuracy affects the precision of the calibration, which is addressed in the next section.

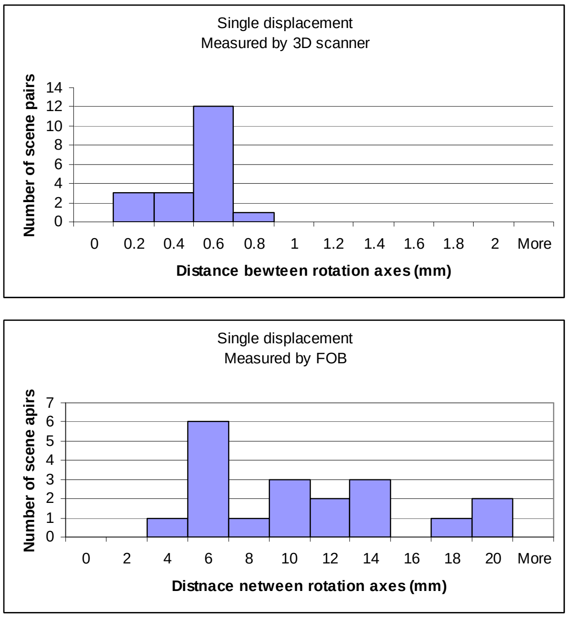

Figure 6. Distance between rotation axes (mm) for *A* matrices (top) and *B* matrices (bottom).

4. ACCURACY OF THE CALIBRATION

In the previous section we used one single displacement of the object, and derived the object's displacement matrix (or scene differential) A and B as observed in frames S and T, respectively. From a pair of matrices (A, B) we can obtain the representations of a particular line, namely ${}^{S}l$ the axis of the object's rotation relative to S and ${}^{T}l$, the same line, relative to T. From this information only we can determine the calibration $X = {}^{S}T$ (that is frame T as measured relative to frame S) up to a translation of T along ${}^{S}l$ and a rotation of T about ${}^{S}l$. The amount of translation and rotation can be resolved by measuring another scene differential Due to the inaccuracy in ${}^{S}l$ and ${}^{T}l$, the determination of X will be of finite accuracy..

To give an impression of the variance of the results, we determined X 10 times under two conditions. In the first the scan view pairs differed by motions and rotations about arbitrary axes, whereas in the second condition, the scan views were in relative orientations and positions repeated in a mechanically identical way. Calibration matrices measured under two mechanically identical conditions are in general still slightly different for two reasons. First, the 3D scan view registrations may not be exactly the same because the scan views themselves may be slightly different, even if they are captured subsequently without moving anything or changing anything to the settings of the scanner. Second, when the FOB's sensor is fixed it still appears not to generate exactly constant data values.

From each measured X we derived its intrinsic amount of rotation θ_X as a quantity to make a comparison. It turned out that the spread under the second condition was about 0.05 degrees, significantly smaller than the spread (about 015 degrees) achieved for the general calibration procedure.. The data from the measurements are shown in Figure 7.

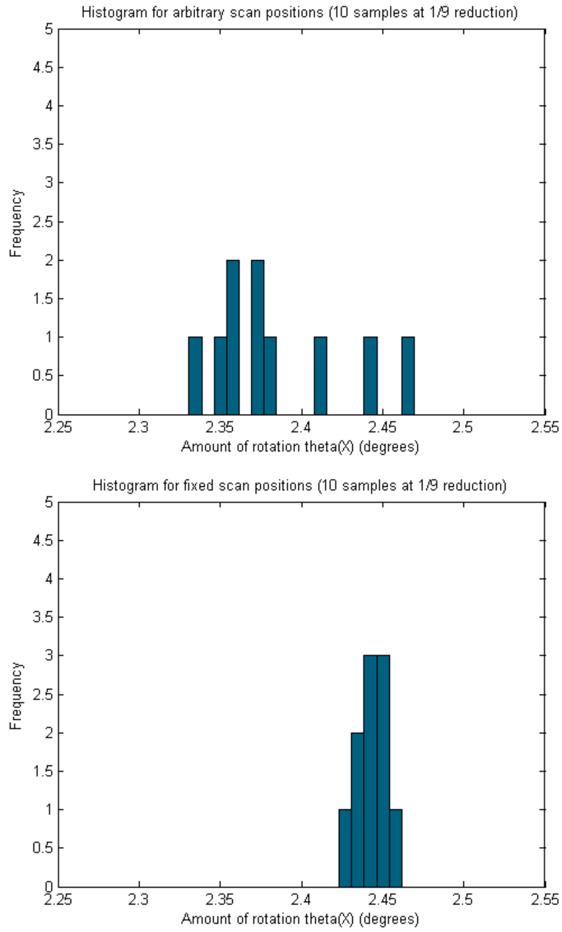

Figure 7. Spread of the intrinsic rotation θ_X for arbitrary measurements (top) and mechanically identical measurements (bottom).

5. APPLICATION TO 3D SCANNING

The development of the calibration technique was motivated by the endeavour to simplify the 3D scanning procedure. The type of scanner we use requires from the user a tedious process on the computer to designate corresponding points in scan views, as to provide a start condition to the registration software. This effort is avoided when the

FOB's sensor remains attached to the scanned object, so that the object's placement is known at any time a scan view is taken. The starting condition for registration can then be derived from the FOB data, provided that the calibration matrix *X* is known.

We have tested our method with objects of different sizes and shapes [Kroes 2007]. The accuracy of the calibration matrix *X* appeared sufficient to support the scanning process in most cases. Occasionally the FOB data was too noisy causing automatic registration to fail. One test case involved a 1:5 scale car body model made of clay. To obtain a 3D scan from this object eight scan views were taken, which could be pair-wise registered and merged into one facetted model fully automatically, see Figure 8.

Figure 8 A 1:5 automotive model (top) was fully automatically registered and merged into a surface model (bottom).

On average the time needed for a 3D scanning process was reduced from 50 to 8 minutes or less, an improvement of efficiency by a factor 6 to 10. The difference in performance is shown in Figure 9. Moreover, the users commented that 3D scanning had become more attractive since the tedious registration process disappeared.

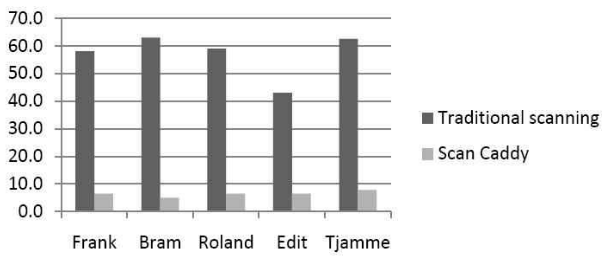

Figure 9 Time (in minutes) it took test persons to carry out a 3D scanning process with and without supplying the FOB data. The FOB facility was called Scan Caddy in the particular experiment.

6. CONCLUSIONS

We have developed a method to automatically calibrate a 3D scanning device using an external 6 DoF sensor. The calibration is achieved by taking a few scan views of an object, where the object is only slightly displaced between the takes. The calibration matrix *X* is then automatically computed from registration matrices and the associated placements measured by the 6 DoF measuring device. The scanning system is then ready for practical operation. The user has to focus on taking the scan views only; the registration and merging process remain unnoticed for the user, which makes the whole facility much more attractive to non-expert users.

Two ways will be investigated to further enhance the calibration process. First, we will replace the FOB device with a wireless 6 DoF sensor, which is hopefully less sensitive to noise and metal objects. Also the cable from sensor to the transmitter would then be absent, which makes the facility more convenient. Secondly, the accuracy of *X* could be improved during normal operation of the scanner. Registration matrices and corresponding placements as measured by the 6 DoF device are continuously available, and can be used to incrementally improve the accuracy of *X*.

7. REFERENCES

[Asc2006] http://www.ascension-tech.com

[Grasset2001] R. Grasset, X. Decoret, J-D Gascuel, "Augmented Reality Collaborative Environment Calibration and Interactive Scene Editing", VRIC, Virtual Reality International Conference,.

[Han2007] http://www.handyscan3d.com

[Kato1999] H. Kato, M. Billinghurst."Marker Tracking and HMD Calibration for a Video-Based Augmented Reality Conferencing", 2nd IEEE and ACM International Workshop on Augmented Reality, pp 85

[Kroes2007] T. Kroes, Easing 3D scanning by pre-aligning scan data with a motion tracking device. MSc Theses, Technical Report, Delft University of Technology, 15 March 2007.

[Lourakis2004] M.I.A. Lourakis, A.A. Argyros. The design and implementation of a generic sparse bundle adjustment software package based on the Levenberg-Marquardt algorithm. Report FORTH-ICS/TR-340, http://www.ics.forth.gr/~lourakis/sba..

[Min2006] http://www.konicaminolta.com

[Smyth2000] S.N. Smyth and D.R. Wallace, Towards the synthesis of aesthetic product form. Proc. DETC2000/DTM-14554, ASME, New York.

[Song2005] Song, Y, Vergeest, JSM, & Bronsvoort, WF. Fitting and manipulating freeform shapes using templates Journal of Computing and Information Science in Engineering, 5(2), 86-94.

[Vergeest2007] Vergeest, JSM, Kroes, T, Song Y. Associating 6 DoF sensor data to 3D scan view registration. In: V. Skala, Proc. 15th Int. Conf. on Computer Graphics, Visualization and Computer Vision 2007, pp 233-240. Plzen, Union Agency.

[Wheeler1998] Wheeler A. Pretlove J. Parker G. "Augmented reality : calibration of the real and virtual worlds", Telemanipulator and telepresence technologies IV, vol. 3206, pp. 30-40.