



Technical section

# Dynamic superimposition of synthetic objects on rigid and simple-deformable real objects

Yann Argotti<sup>a</sup>, Larry Davis<sup>a</sup>, Valerie Outters<sup>b</sup>, Jannick P. Rolland<sup>a,b,\*</sup><sup>a</sup>*School of Electrical Engineering and Computer Science, University of Central Florida, 4000 Central Florida Blvd, Orlando, FL 32816-2700 USA*<sup>b</sup>*School of Optics, CREOL, University of Central Florida, 4000 Central Florida Blvd, Orlando, FL 32816-2700, USA*

---

## Abstract

A current challenge in augmented reality applications is the accurate superimposition of synthetic objects on real objects within the environment. This challenge is heightened when the real objects are in motion and/or are non-rigid. In this article, we present a robust method for real-time, optical superimposition of synthetic objects on dynamic, rigid and simple-deformable real objects. Moreover, we illustrate this general method with the VRDA Tool, a medical education application enabling the visualization of internal human knee joint anatomy on a real human knee.

© 2002 Elsevier Science Ltd. All rights reserved.

*Keywords:* Augmented reality; Dynamic optical superimposition; Tracking; Simple-deformable objects; VRDA tool

---

## 1. Introduction

In a large range of fields, the ability to enhance reality with synthetic information is an exciting alternative to traditional methods of presenting information. Applications where computer-generated objects are employed to augment the perception of the real environment are referred to as augmented reality (AR) applications. AR technology typically includes input devices that enable users to interact with the synthetic world and output devices to allow them to visualize the augmented environment. Input devices can be as simple as keyboards or as complex as tracking systems that are used to determine the location and orientation of real objects in the environment. For example, the point of view of a user in the synthetic world, which is typically considered correlated to the user's head orientation in the real world, is obtained by tracking the head of the user. The

output device corresponds to the visualization of computer-generated synthetic objects. A common output device is the head-mounted display (HMD). Variations of HMDs that are used in AR applications include optical see-through, video see-through HMD [1], and projective see-through [2].

A significant challenge in AR applications is the correct superimposition of synthetic objects on real objects within the environment. Real and synthetic objects must be placed *into register*, that is, spatial coincidence, from a common reference. The superimposition becomes more challenging when the real objects are moving. In general, real objects are considered rigid with respect to tracking, which represents only a subset of the possibilities given that the real objects in the environment may be non-rigid. The contribution of this article is to present a method for dynamic superimposition that is robust, accurate, interactive-speed, and applicable to rigid and simple-deformable real objects. The method presented is applied to the Virtual Reality Dynamic Anatomy (VRDA) Tool, a visualization system developed for the study of complex joint motions [3]. Furthermore, the method presented here is applicable to other areas of

---

\*Corresponding author. School of Optics, CREOL, University of Central Florida, 4000 Central Florida Blvd, Orlando, FL 32816-2700 USA. Tel.: +1-407-823-6870; fax: 407-823-6880.

*E-mail address:* jannick@odalab.ucf.edu (J.P. Rolland).

AR, including surgical assistance, engineering applications, military simulation, and entertainment.

In the following sections, we first discuss related work in augmented reality, tracking, and anatomical motion tracking. We then summarize the calibration technique associated with the dynamic superimposition detailed in Argotti et al. [4,5], and focus the contribution of the paper on detailing the real-time, dynamic superimposition method, which was first summarized in Argotti et al. (2001) [4,5]. Finally, we demonstrate the method as it applies to the VRDA Tool, present visual results of the superimposition, and discuss the future direction of the research.

## 2. Previous work

For correct superimposition within AR applications, real objects must be located or tracked and synthetic objects must be displayed accordingly. Primarily, the focus in AR research is upon tracking the motion of objects within the environment that are considered rigid. For instance, Sutherland [6], Caudell and Mizell [7], and Feiner et al. [8] implemented rigid head tracking. Developments for tracking rigid objects within AR applications also included quantification of sources of registration error [9], the use of predictive filtering to improve tracking performance [10,11], and inside-out optical tracking [12].

In superimposing synthetic objects on real objects, most AR applications treat the objects rigidly and thus apply static registration methods. Among the research in static registration methods for AR, Mellor introduced a static marker-based tracking method that was able to recover depth information using a single video source [13]. Grimson et al. created an AR system that featured interactive updates of a patient's brain using MRI data [14]. Fuchs et al. developed a system to aid in laparoscopic surgery that used structured light patterns for tracking [15].

Furthermore, dynamic superimposition (i.e. the objects as well as the user position are dynamic) within AR applications has been typically limited to rigid objects. Bajura and Neumann used a closed loop registration correction method to enhance dynamic superimposition [16]. Uenohara and Kanade implemented a method to dynamically track rigid objects using the video outputs of cameras in a video see-through HMD [17]. State et al. implemented an occlusion-resistant, hybrid tracking scheme to achieve dynamic superimposition [18]. In the realm of wearable computing, Starner et al. created a system to aid in everyday tasks [19], while Billingham and Kato developed a system that used a dynamic collaborative writing surface [20].

In recent years, however, there has been increased interest in tracking the motion of non-rigid objects.

Halvey and Weinshall implemented a method for tracking non-rigid objects in video sequences based upon optical flow methods [21]. Comaniciu et al. implemented a method for tracking non-rigid objects based upon statistical properties [22]. Within the context of tracking non-rigid, dynamic objects, one of the most challenging tracking tasks is tracking the motion of human anatomy. Spoor and Veldpaus published a method for calculating rigid body motion from the spatial coordinates of markers that has been adapted to tracking skeletal motion [23]. In addition, techniques have been devised that address the problems associated with accurately tracking anatomical motion [24,25].

In fact, much is known about the motion of anatomical structures, but they still pose a significant challenge to inclusion within AR systems. Aside from the fact that markers cannot be directly positioned on bones in daily settings, anatomical structures can be also non-rigid. This characteristic increases the difficulty of registering real anatomical structures with synthetic structures in three dimensions. Moreover, attempting to track anatomical structures at interactive speed while maintaining registration of synthetic objects is especially challenging. In this paper, we examine the problem of tracking simple-deformable bodies within an augmented reality system and present a general method for dynamically superimposing synthetic objects on these real objects at interactive speeds.

## 3. Method overview

The overall method includes calibration and dynamic superimposition procedures. Both procedures assume the use of a marker-based tracking system capable of providing the 3D location of the markers. A cluster of markers placed on its surface defines each real object in the system. The procedures also assume the use of a stereoscopic display device with markers attached to it for determining the location and orientation of the head of the user.

In the method presented, we call the tracker coordinate system the global coordinate system or global frame. Moreover, for each object in the environment (real or synthetic), we associate a local coordinate system or local frame, as shown in Fig. 1. We refer to the transformation matrices between coordinate systems within the environment as links. Objects that have an expressed transformational relationship between one another are referred to as linked objects. Simple-deformable objects are defined as objects that are slightly changing in shape compared to an equivalent rigid object. The change in shape can be quantified by the change in the eigenvalues of the dispersion matrix associated with a cluster of markers placed on the object [4].

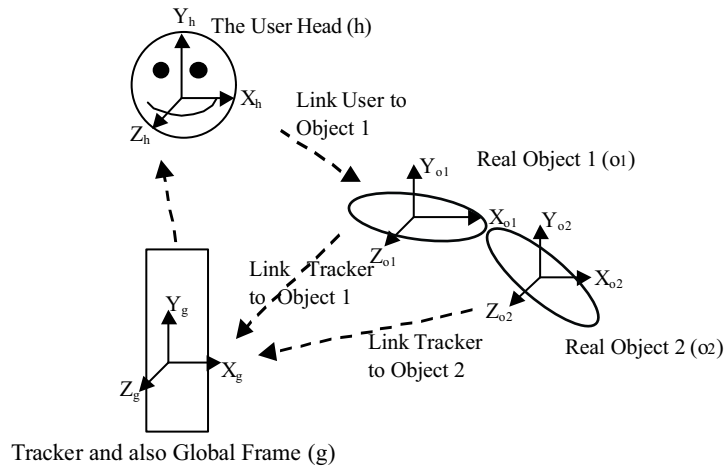


Fig. 1. Representation of the coordinate systems of the real objects in the overall system: The tracker whose coordinate system is the global coordinate system; The head of the user; And two real objects perceived by the user. The synthetic objects that will get overlaid on the real objects also have their own separate associated coordinate systems.

The general steps of the basic calibration procedure detailed in [4] can be summarized as follows: First, the local coordinate frames for the real objects in the system are defined and computed based upon the cluster of markers placed on them. The computation is based on computing a dispersion matrix of the markers around their centroid, as well as the eigenvectors and eigenvalues of this matrix. The eigenvectors of the dispersion matrix are chosen to define the local frames. An optimization method for accounting for all markers around the real objects in the computation of the eigenvectors was detailed in [4]. The local coordinates of all markers are now known. Next, the correspondence between real and synthetic objects is determined in the environment based on the registration of landmarks present on both objects. The correspondence is established by expressing the coordinates of each landmark in both coordinate frames, and computing the scaling, translation and rotation matrices that bring these landmarks in best correspondence according to a root mean-square error. The scaling is in fact done first by computing a mean scaling as well as its standard deviation for all the landmarks and establishing whether uniform scaling is appropriate or not for that object [4]. Finally, the optical properties of the system are quantified. The calibration steps, expressed in a flowchart, are shown in Fig. 2.

The work presented hereafter focuses on the dynamic superimposition process now summarized and detailed in Section 4. The first step in the dynamic superimposition procedure is to measure the global locations of at least three markers on each real object. Given also the location of these markers in their local coordinate

frame, which were obtained during the calibration process, an optimization method described in Section 4.1 is applied to estimate the rotation and translation matrices that yield the measured global coordinates when applied to the local coordinates. The local motion of markers on a simple-deformable object is managed during this step. Next, because there may be a need for collision detection and/or motion constraints between linked objects within the environment, the transformation matrix which links the real objects is used as an input to a kinematic model of motion [26]. The last step is the stereoscopic rendering process that combines all the required transformation matrices and allows defining the connection between the real and synthetic world. Included in the final step is the correction of optical distortion that may be introduced by the display device. The steps of the dynamic superimposition, expressed in a flowchart, are shown in Fig. 3.

#### 4. Dynamic superimposition process

The dynamic superimposition process requires continually locating all real objects in the environment from the visible markers and accounting for occluded markers. Furthermore, the process can be optimized by pre-assigning a level of fidelity to each marker, which will also be described in Section 4.1. The process further requires the computation of the relative position and orientation of the real objects. Finally, it requires rendering the synthetic objects. The superimposition method is robust, taking into account noise in the tracking data.

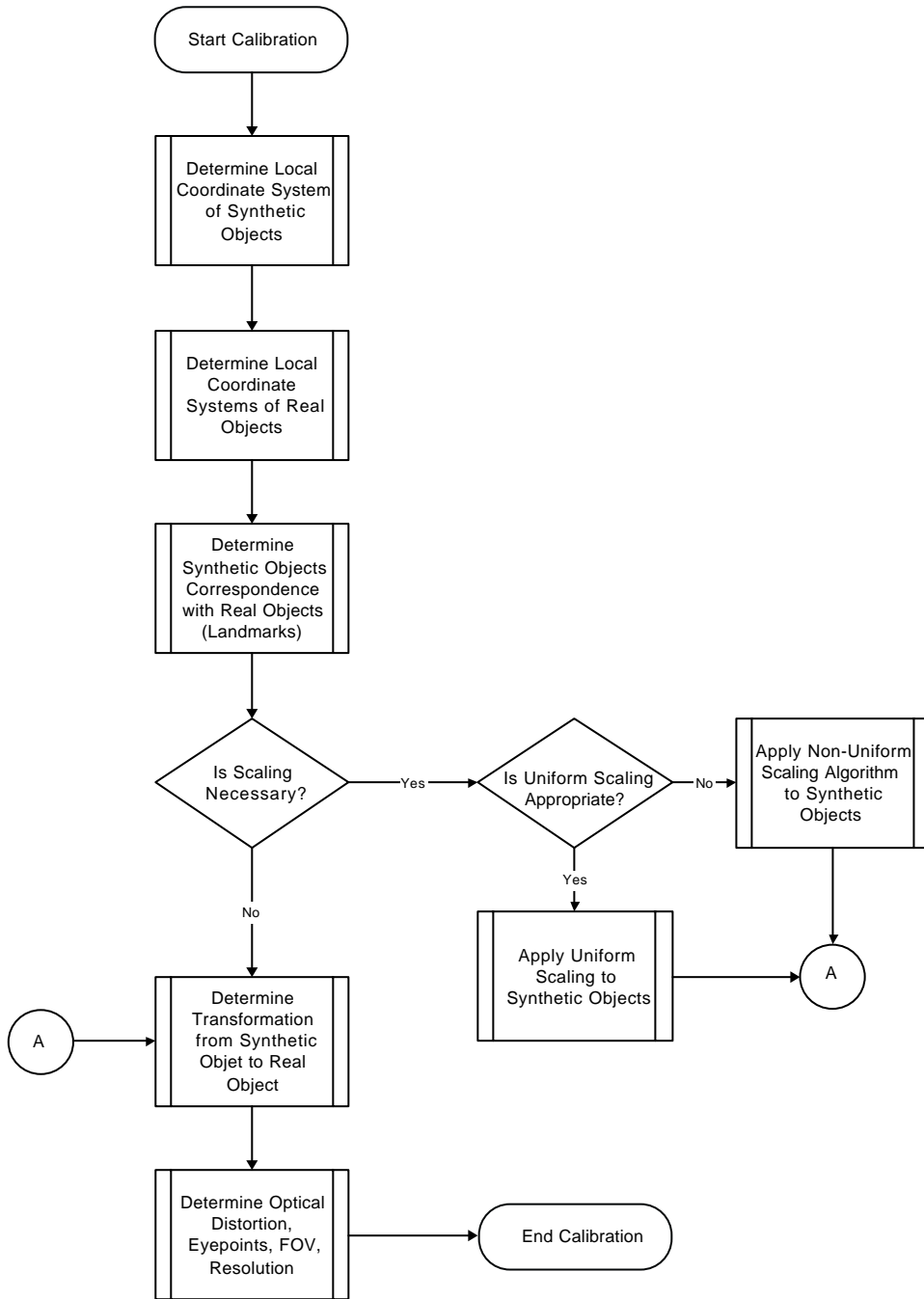


Fig. 2. The calibration process.

4.1. Locating real objects with all markers visible

The result of the calibration process is the transformation matrix from the synthetic object frames to the real object frames. To properly display synthetic objects from the eye viewpoint, the matrix describing the

transformation from the real object local frame to the global frame is needed.

Given  $n$  markers on the surface of the real object, if  $x_i$  is the  $i$ th real object marker coordinate expressed in its local frame, and  $y_i$  is the  $i$ th real object marker coordinate expressed in the global frame, then the

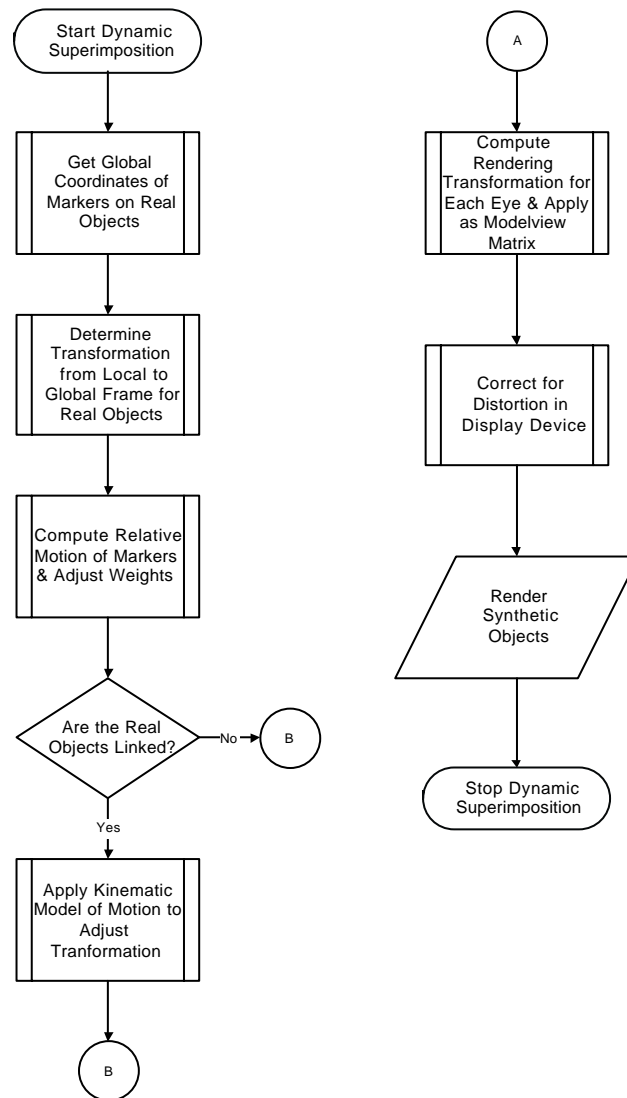


Fig. 3. The dynamic superimposition process.

desired transformation transforms  $x_i$  into  $y_i$ . The transformation can be decomposed into a rotation matrix,  $R$ , and a translation vector,  $T$ . Scaling is unnecessary because the transformation is between normalized frames. Thus, the relationship between  $x_i$  and  $y_i$  is

$$y_i = Rx_i + T \quad \text{with } i \in [1; n]. \quad (1)$$

Furthermore, the data from the tracking system are intrinsically noisy and, in the case of a simple-deformable object, the markers may move with respect to each other. To take into account both the noise of the tracking data and the possible relative motion of the markers, a weight is applied to each marker to represent its fidelity. The more a marker moves or is subject to

noise, the smaller will be its weight in order to assign more importance to the higher fidelity markers.

The weight for each marker is determined as an iterative process. From the extended position of the leg during the calibration procedure, a local frame was defined based on the eigenvectors of the dispersion matrix computed from the distribution of markers around their centroid. The impact of the leg being a simple deformable model is the change in the distribution of markers around their centroid, and therefore the change of the local frame. Therefore, to assess the relative motion of each marker around their centroid, we conduct repeated measures (e.g. 20) of the local frame, as defined during the calibration procedure, but now after the leg goes through a full cycle of flexion–extension. The relative displacement  $\Delta x$  of each marker

with respect to the centroid between two consecutive local frames is computed using

$$\Delta x_i = x'_i - x_i. \quad (2)$$

The average variation  $\Delta x$  measured over at least 20 repeated measures is computed. The error or sensitivity of each marker location is then determined by comparing the tracking system precision to the current marker relative motion and taking the largest value. Thus, the weight of the  $i$ th marker,  $w_i$  is then given by

$$w_i = 1 - \frac{Error_i}{\sum_{i=0}^n Error_i}, \quad (3)$$

where  $n$  is the total number of markers. In Section 4.2, we shall express how these weights should be modified if all markers are not visible.

Given a set of weights, for each position of the leg during motion, the matrix  $R$  and the vector  $T$  that transform the local coordinate of a marker into its global coordinate as expressed by Eq. (1) are then estimated using least-squares minimization. The error,  $e$ , to minimize is given by

$$e(R, T) = \sum_{i=1}^n w_i \|y_i - Rx_i - T\|^2, \quad (4)$$

which can be solved using SVD [27]. Once estimated,  $R$  and  $T$  are arranged in a  $4 \times 4$  matrix format. Expressions of  $R$  and  $T$  are derived in the Appendix. We chose to implement a method based on SVD [32] because this classical optimization method is robust [28], gives the best possible solution at all times, is computationally efficient [29], and converges quickly to a solution.

#### 4.2. Locating real objects with occluded markers

During the tracking process, some markers may not be detected by the tracking system, while other markers are detected. Since we are weighting the marker coordinates to give more importance to the markers that are less perturbed by measurement noise and by relative motion on the real object, we now limit the computation to take into account only markers that have been detected. In this case, we have to compute the new weight,  $w'_i$  associated to the  $i$ th marker whose current weight is  $w_i$ .

$$w'_i = \frac{w_i}{\sum_{j=1}^n w_j \cdot Visible(j)}, \quad (5)$$

where  $n$  is the total number of markers of the current real object, and  $Visible(j)$  is a function that gives 1 if the  $j$ th marker is detected and 0 otherwise. Finally, we can determine the optimal solution for  $R$  and  $T$  using SVD, resulting in the transformation from the real object local frame to the global frame, as detailed in Section 4.1. This determination is robust because SVD is an optimization technique.

#### 4.3. Computation of relative position and orientation of real objects

We have determined the transformation matrices,  $M_{g\_ri}$ , that relate the global coordinate system to the local coordinate system of the  $i$ th real object. We shall now determine the transformation from one real object local frame to another by multiplying matrices.  $M_{o1\_o2}$ , the transformation from the local frame of object 2 to object 1, is then given by

$$M_{o1\_o2} = M_{g\_o1}^{-1} \cdot M_{g\_o2}. \quad (6)$$

We compute a matrix inverse to determine  $M_{o1\_o2}$  instead of resorting to SVD, because the computation of the inverse transformation matrix is not computationally intensive. Also, in the SVD method, the scaling, rotation and translation matrices are computed individually, as opposed to a four by four matrix, given by direct inversion.

After computing the link between objects, the method may include a control process to account for the motion of objects with respect to each other. In this case, the transformation matrices describing the global position and orientation for each real object are given as entries in lookup tables. The lookup tables describe the motion between real objects and give the accurate location and orientation of their synthetic counterparts to avoid collision. The location and orientation information can be used to modify the synthetic object transformation matrices. A kinematic model applied in a pre-computed lookup table allows optimization of the speed of rendering [10].

#### 4.4. Rendering synthetic objects

The last part of the dynamic superposition process is the rendering of synthetic objects for each eye view and the correction of optical deformations introduced by the HMD optics. Thus, the HMD specifications and user parameters are required for proper placement of the synthetic objects within the application [30]. Moreover, it may be necessary to have a means for correcting residual optical distortions in the HMD for proper visualization of the synthetic objects [28].

User head tracking or equivalently display device tracking is utilized in order to extract user head location and orientation. As detailed in Section 4.1, SVD is applied to find the transformation matrix,  $M_{g\_h}$ , from the display device or equivalently head local frame,  $h$ , to the global frame,  $g$ . The transformation matrices  $M_{sn\_re}$  and  $M_{sn\_le}$  from the user's right eye (re) and left eye (le), respectively, to the synthetic object  $n$  (i.e. sn), are computed as

$$M_{re\_sn} = M_{re\_h} \cdot M_{h\_g} \cdot M_{g\_m} \cdot M_{m\_sn}, \quad (7)$$

$$M_{le\_sn} = M_{le\_h} \cdot M_{h\_g} \cdot M_{g\_rn} \cdot M_{rn\_sn}, \quad (8)$$

where  $M_{h\_re}$  is the transformation matrix from the right eye frame to the user's head frame,  $M_{h\_le}$  is the transformation matrix from left eye frame to the head frame,  $M_{rn\_g}$  is the transformation matrix from the global frame to the real object  $n$  frame, and  $M_{rn\_sn}$  is the transformation matrix from the synthetic object  $n$  frame to the real object  $n$  frame. In OpenGL, we set the modelview matrix to  $M_{re\_sn}$  and  $M_{le\_sn}$  before rendering the synthetic object  $n$ .

To correct the residual optical distortion present in the display device, we apply the rendered image on a deformed polygon mesh by texture mapping [31]. We define the optical distortion for each eye and compute a polygon mesh for each eye that combines the optical distortion and corrects it. The textures applied on these meshes are images of the synthetic scene rendered off-screen, via the pixel buffer in OpenGL.

### 5. An application: dynamic superimposition of a knee joint on a patient's leg

The method described for dynamic superimposition in augmented reality systems is well suited for implementation in complex AR systems. The Virtual Reality Dynamic Anatomy (VRDA) Tool is a system that allows medical practitioners to visualize anatomical structures superimposed on their real counterparts. To realize this effect, the medical practitioners wear a HMD to view a computer graphics model of the knee superimposed on the real leg of a model patient. In the following sections, we demonstrate how the method is integrated within the VRDA Tool (Fig. 4).



Fig. 4. The VRDA Tool: from markers on a real leg to a see-through head-tracked bench-mounted display.

#### 5.1. System setup

The knee is one of the most complex anatomical human joints with regards to its structure and its motion. Fortunately, the complexity of motion is not a limiting factor in the proposed method. In order to demonstrate the methods, we treat the leg as two separate real objects and the virtual model as their synthetic counterparts; the first object is associated with the femur and the second object is associated with the tibia. They are naturally linked together. The 3D models we are using to represent the complete knee joint anatomy are high-resolution models from Viewpoint Corporation, acquired by digitizing the anatomy of a cadaver. The tracking system we employ is an optical tracker, OPTOTRAK™, which uses active, infrared LEDs as markers. The choice of this system is based upon its resolution, robustness against common perturbations, and speed. The display device is a prototype see-through bench mounted display. We are currently using a Silicon Graphics Deskside Onyx2 with an Infinite Reality2 graphics pipeline to run the application. We perform both computations and stereoscopic rendering on this computer. However, we can consider the leg segments as rigid objects only as a first approximation. The muscles and the skin create many perturbations in the 3D marker locations with respect to the bones that must be taken into consideration for optimal performance [36]. Therefore the methods presented are necessary for full optimization of the optical superimposition.

#### 5.2. Application of the method

Each part of the real leg is tracked independently. To find the best location of the markers, we considered the shape of the leg and chose the marker locations where they would probably move the least [20]. Also, the correspondence between the real object and the synthetic object is realized by defining some common landmarks. We defined the landmarks in places where there is less flesh, allowing the landmarks to be closer to the bones to reduce scaling or location errors. The synthetic model is scaled based on some landmarks location measure (Fig. 5).

To estimate the relative motion of the markers on the leg, we made 1000 measurements of the global 3D location of the markers over a 10 s interval of standard motion for the leg. We found that the maximum standard deviation of the motion of markers is less than 15 mm. In a first implementation, weights were assigned accordingly. While it was beyond the scope of the first implementation, optimal weights can be established by redefining the local frames attached to each object over repeated measures as described in the overall mathematical framework, and iterating the process until the error

function given by Eq. (4) reaches a minimum. For the eyepoints location from which the projection of the scene is rendered, we chose the centers of rotation of the eyes [33]. The field of view of the HMD is  $26.11^\circ$  and the display resolution is  $640 \times 480$  pixels. We also applied a coating to the LCD displays to minimize the pixelization of our synthetic objects [34]. Finally, by applying a kinematic model of motion to the synthetic objects, collisions between the synthetic objects can be avoided [10,35]. We adopted an approach where the entry is the transformation between the two synthetic objects and the table returns the real-time location of the synthetic objects. The technique allows a smooth, realistic motion without collisions.

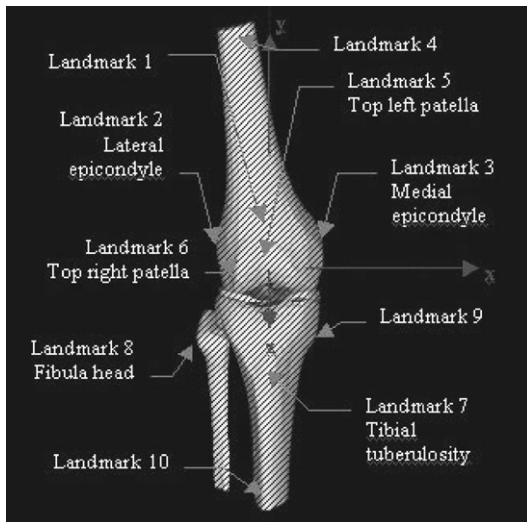


Fig. 5. Ten selected landmarks to associate the real knee with the virtual knee.



Fig. 6. A flexion (a) and partial extension (b) of the leg of a model patient demonstrating the optical superimposition of knee joint anatomy on a real leg.

### 5.3. Results and discussion

The implementation of the methods presented allows superimposition at interactive-speed. We are currently able to achieve frame rates of up to 26.6 Hz. Furthermore, because of the choice of the SVD method and the enhancement of noise attenuation, the superimposition process is robust. Two views of the dynamic superimposition are shown in Fig. 6.

In the first implementation of the methods, we currently use the minimum number of markers that insures that the tracker sees at least three from each viewpoint and camera, a requirement of the tracker. Furthermore, the markers are currently mostly located at key anatomical landmark points as described in the paper to minimize the motion of the markers with respect to the skin during leg flexion–extension. Future work will investigate the potential enhancement of the methods by uniformly distributing markers on the leg, combined with control points at key anatomical landmarks. Based on the optical tracker maximum marker rate (i.e. for one marker) of 3500 Hz, increasing the number of markers beyond about 100 will affect the update rate of the superimposition if all markers are sequentially activated. However in the case of the leg, further optimization of the tracking algorithm may be established by flashing only markers visible from the user viewpoint. According to the current methods and the current implementation with 16 markers, the number of markers does not limit the frame rate.

Regarding the quantification of the methods, it is not currently possible to perform such quantification meaningfully given that we do currently use a generic anatomical joint model rather than a real patient specific model. The visual superimposition indicates however





Fig. 7. A superimposition with full internal anatomy.

that accuracy in registration is around 5–10 mm r.m.s. error. Performance measures following the optimization of the number of markers as well as their individual weight, combined with MRI patient specific models is under investigation and will be reported in future work. Finally, long term research will also demonstrate deformable structures such as ligaments and muscles with respect to the bones as well. A static superimposition in extension is demonstrated in Fig. 7. Applications of augmented reality methods presented here will be further extended to perform full body motion capture.

## 6. Conclusion

In this paper we have presented a robust method that allows real-time, optical superimposition of synthetic objects on dynamic rigid and simple-deformable real objects. We illustrated these methods with the VRDA Tool, a medical education application for the visualization of internal anatomy on real human anatomy. In the demonstration, while the deformable structures of the knee joint are accounted for in part in the kinematic model we established in previous work, we currently represent solely the internal motion of the bones on a subject. Furthermore we currently employ a generic knee joint model. Therefore, while the visual rendering indicates accuracy in registration between 5 and 10 mm r.m.s. error, it is non-meaningful to further quantify the registration until we complete the development and implementation of the methods for patient specific models. Generic models can ultimately be employed after advanced methods of anatomical scaling of generic

models to patient specific models have been developed, implemented, and validated.

## Acknowledgements

The authors thank Benjamin J. Fregly from the University of Florida and Scott Banks from the Research Orthopaedic Laboratory in West Palm Beach Florida for stimulating discussion about this work and their current joint partnership with us to apply the technique to Orthopaedic Research using real patient models. We further thank Ami Sun for providing comprehensive video footage of the VRDA Tool. Videotapes are available upon request <http://odalab.ucf.edu>. This work was supported by a First Award from the National Library of Medicine of the National Institute of Health 1-R29-LM06322-01A1 to Jannick Rolland and a Fellowship from the Florida Education Fund to Larry Davis. Furthermore, the overall virtual environment research that supports many components of the research presented in this paper is supported by instrumentation grants from the National Science Foundation EIA-99-86051 and the Lockheed-Martin Corporation.

## Appendix

We chose to use a least-squares method to find  $R$  and  $T$ , minimizing the weighted error,  $e$ , defined as

$$e(R, T) = \sum_{i=1}^n w_i \|y_i - Rx_i - T\|^2,$$

where  $y_i$  is the  $i$ th marker coordinate defined in the real object global frame and  $x_i$  is the  $i$ th marker coordinate defined in the real object local frame,  $w_i$  is the weight of the  $i$ th marker that quantifies the robustness of the marker data against noise measurements and its relative motion on the real object.

We define and compute new coordinates with respect to the weighted centroids,  $x$  and  $y$ , of the markers in each frame given by

$$x'_i = x_i - x \quad \text{and} \quad y'_i = y_i - y, \quad \forall i, 1 \leq i \leq n,$$

where

$$x = \frac{\sum_{i=1}^n w_i x_i}{\sum_{i=1}^n w_i} \quad \text{and} \quad y = \frac{\sum_{i=1}^n w_i y_i}{\sum_{i=1}^n w_i}.$$

The error function can be rewritten as

$$e(R, T) = \sum_{i=1}^n w_i \|y'_i - Rx'_i - T'\|^2$$

with

$$T' = T - y + Rx.$$

Moreover, we can perform a Taylor Series expansion of  $e$ , and taking into account the fact that

$$\sum_{i=1}^n w_i x'_i = \mathbf{0} \quad \text{and} \quad \sum_{i=1}^n w_i y'_i = \mathbf{0},$$

we obtain

$$\begin{aligned} e(R, T) &= \sum_{i=1}^n w_i \|y'_i - Rx'_i\|^2 \\ &\quad - 2T' \sum_{i=1}^n w_i (y'_i Rx'_i) + \left( \sum_{i=1}^n w_i \right) \|T'\|^2 \\ &= \sum_{i=1}^n w_i \|y'_i - Rx'_i\|^2 + \left( \sum_{i=1}^n w_i \right) \|T'\|^2. \end{aligned}$$

Minimizing with respect to  $T'$ , we must have

$$\|T'\|^2 = 0 \Rightarrow T = y - Rx.$$

We can then write  $e$  as a function of  $R$  as

$$e(R) = \sum_{i=1}^n w_i \|y'_i - Rx'_i\|^2.$$

Expanding  $e$ , it can be written as

$$\begin{aligned} e(R) &= \sum_{i=1}^n w_i (y'_i - Rx'_i)^t (y'_i - Rx'_i) \quad (\text{scalars}) \\ &= \sum_{i=1}^n w_i (y_i^t y'_i + x_i^t R^t R x'_i - y_i^t R x'_i - x_i^t R^t y'_i). \end{aligned}$$

By applying  $R^t R = I$ ,  $e$  reduces to

$$e(R) = \sum_{i=1}^n w_i (y_i^t y'_i + x_i^t x'_i - 2y_i^t R x'_i).$$

Thus, because the products  $(y_i^t y'_i)$  and  $(x_i^t x'_i)$  are scalars, minimizing  $e(R)$  is equivalent to maximizing a function of  $R$  defined as

$$f(R) = \sum_{i=1}^n w_i y_i^t R x'_i.$$

At this point, the idea is to solve the maximization of  $f(R)$ , then use SVD to solve for  $R$ . Furthermore, when  $R$  is found,  $T$  is obtained by

$$T = y - Rx.$$

Finally  $f(R)$  can be rewritten as

$$\begin{aligned} f(R) &= \sum_{i=1}^n w_i y_i^t R x'_i = \text{Trace} \left( \sum_{i=1}^n w_i R x'_i y_i^t \right) \\ &= \text{Trace}(RH), \end{aligned}$$

where (with some computation),

$$H = \sum_{i=1}^n w_i x'_i y_i^t.$$

From this equality, we can apply the following lemma:

**Lemma.** For any positive definite matrix  $AA^t$ , and any orthogonal matrix  $B$ , we have

$$\text{Trace}(AA^t) \geq \text{Trace}(BAA^t)$$

**Proof.** Let  $a_i$  be the  $i$ th column of  $A$ . Then, (with some computation)

$$\begin{aligned} \text{Trace}(BAA^t) &= \text{Trace}(A^t B A) \\ &= \sum_{i=1}^3 a_i^t (B a_i). \end{aligned}$$

But, by the Schwarz inequality [18],

$$a_i^t (B a_i) \leq \sqrt{(a_i^t a_i)(a_i^t B^t B a_i)} = a_i^t a_i.$$

Hence

$$\text{Trace}(BAA^t) \leq \sum_{i=1}^3 a_i^t a_i = \text{Trace}(AA^t).$$

As an intermediate result, we conclude that if we can decompose  $RH$  to a form  $AA^t$ , which is positive definitive, then  $\text{Trace}(RH)$  will maximize  $f(R)$ , by the previously stated lemma. Also, by maximizing  $f(R)$ ,  $R$  will be determined.

SVD helps to solve this problem. We take the SVD of  $H$  to be

$$H = U \Gamma V^t,$$

where  $U$  and  $V$  are orthonormal matrices, and  $\Gamma$  (the matrix of singular values) is a diagonal matrix with non-negative elements expressed as

$$\begin{aligned} \Gamma &= \begin{pmatrix} \gamma_1 & 0 & 0 \\ 0 & \gamma_2 & 0 \\ 0 & 0 & \gamma_3 \end{pmatrix} \\ &= \begin{pmatrix} \sqrt{\gamma_1} & 0 & 0 \\ 0 & \sqrt{\gamma_2} & 0 \\ 0 & 0 & \sqrt{\gamma_3} \end{pmatrix} \begin{pmatrix} \sqrt{\gamma_1} & 0 & 0 \\ 0 & \sqrt{\gamma_2} & 0 \\ 0 & 0 & \sqrt{\gamma_3} \end{pmatrix} \\ &= CC^t. \end{aligned}$$

Let  $X = VU^t$ , which is orthonormal, then

$$XH = V U^t U \Gamma V^t = V C C^t V^t.$$

Let  $A = V C$ , then

$$X H = A A^t.$$

$XH$  is symmetrical and positive definite. Thus, among all orthonormal matrices,  $X$  maximizes the function

$$f(R) = \text{Trace}(RH).$$

Finally, the minimum of the error function  $e(R, T)$  occurs when

$$X = VU^t.$$

We have two possible results because  $VU^t$  is orthonormal, meaning the determinant of  $X$  is 1 or  $-1$ .

If  $\text{Det}(VU^t) = 1$ , the orthonormal matrix  $X$  is a rotation, and

$$R = VU^t.$$

If  $\text{Det}(VU^t) = -1$ , the orthonormal matrix  $X$  is a reflection. We can find the corresponding rotation by

$$R = V \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix} U^t.$$

$T$  is then found by

$$T = y - Rx \quad \blacksquare$$

## References

- [1] Rolland JP, Fuchs H. Optical versus video see-through head-mounted displays in medical visualization. *Presence: Teleoperators and Virtual Environments* 2000;9(3): 287–309.
- [2] Hua H, Girardot A, Gao CY, Rolland JP. Engineering of head-mounted projective displays. *Applied Optics* 2000;39(22):3814–24.
- [3] Wright DL, Renner JB, Jordan JM, Rolland JP. Predicting bone diameter from tissue diameter. Technical Report TR02-003 University of Central Florida, 2002.
- [4] Argotti Y, Outters V, Davis L, Sun A, Rolland JP. Technologies for augmented reality: calibration for real-time superimposition on rigid and simple-deformable real objects. *Proceedings of the MICCAI'01, Utrecht, Netherlands, October 2001*.
- [5] Argotti Y, Davis L, Outters V, Rolland JP. Dynamic superimposition of synthetic objects on rigid and simple-deformable real objects. *Proceedings of the ISAR'01, New York, NY: IEEE-Computer Society, October 2001*. p. 5–10.
- [6] Uenohara M, Kanade T. Vision-based object registration for real-time image overlay. *International Journal of Computers in Biology and Medicine* 1995;25(2):249–60.
- [7] Caudell TP, Mizell DW. Augmented reality: an application of heads-up display technology to manual manufacturing processes. *Proceedings of the Hawaii International Conference on System Sciences*, 1992. p. 659–69.
- [8] Feiner S, MacIntyre B, Seligmann D. Knowledge-based augmented reality. *Communications of the ACM* 1993;36(7):52–62.
- [9] Holloway R. An analysis of registration errors in a see-through head-mounted display system for craniofacial surgery planning. Ph.D. Dissertation, University of North Carolina, Chapel Hill, 1995.
- [10] Azuma R, Bishop G. Improving Static and Dynamic Registration in an Optical See-through HMD. In: *Proceedings of the SIGGRAPH '94, Orlando, FL, July 24–29, 1994*. Computer Graphics Proceedings, Annual Conference Series. ACM SIGGRAPH, 1994. p. 197–204.
- [11] Foxlin E. Inertial head-tracker sensor fusion by a complementary separate-bias Kalman filter. *Virtual Reality Annual International Symposium* 1996;267:185–94.
- [12] Wright DL, Rolland JP, Kancherla AR. Using virtual reality to teach radiographic positioning. *Radiologic Technology* 1995;66(4):167–72.
- [13] Mellor JP. Realtime camera calibration for enhanced reality visualization. *Proceedings of the First International Conference on Computer Vision, Virtual Reality and Robotics in Medicine, Nice, France, 3–5 April 1995*.
- [14] Grimson WEL, Kikinis R, Jolesz FA, Black PMcL. Image guided surgery. *Scientific American* 1999;282(1):63–9.
- [15] Fuchs H, Livingston MA, Raskar R, Colucci D, Keller K, State A, Crawford JR, Rademacher P, Drake SH, Meyer AA. Augmented reality visualization for laparoscopic surgery. *MICCAI '98* 1998. p. 934–43.
- [16] Bajura M, Neumann U. Dynamic registration correction in video-based augmented reality systems. *IEEE Computer Graphics and Applications* 1995;15(9):52–60.
- [17] Van Sint Jan SL, Clapworthy GJ, Rooze M. Visualization of combined motions in human joints. *IEEE Computer Graphics and Applications* 1998;18(6):10–4.
- [18] Stewart GW. *Introduction to matrix computations*. New York: Academic Press Inc., 1973.
- [19] State A, Hirota G, Chen DT, Garrett WF, Livingston MA. Superior augmented-reality registration by integrating landmark tracking and magnetic tracking. In: *Proceedings of the SIGGRAPH 96, New Orleans, LA, August 4–9, 1996*. Computer Graphics Proceedings, Annual Conference Series. ACM SIGGRAPH, 1996. p. 429–38.
- [20] Billinghurst M, Kato H. Collaborative mixed reality. In: *Proceedings of the ISMR '99, Berlin, 1999*. *Proceedings of the First International Symposium on Mixed Reality, 1999*, Springer, Berlin, p. 261–84.
- [21] Halvey G, Weinsall D. Motion of disturbances: detection and tracking of multi-body, non-rigid motion. *Machine Vision and Applications* 1999;11:122–37.
- [22] Comaniciu D, Ramesh V, Meer P. Real-time tracking of non-rigid objects using mean shift. *Proceeding of the IEEE Computer Vision and Pattern Recognition* 2000;2:142–9.
- [23] Starner T, Mann S, Rhodes B, Levine J, Healey J, Kirsch D, Picard R, Pentland A. Augmented reality through wearable computing. *Presence: Teleoperators and Virtual Environments* 1997;6(4):386–98.
- [24] Andriachhi TP. Clinical applications of the SELSPOT System. *Proceedings of the Biomechanics Symposium, ASME, 1987;84:339–42*.
- [25] Cappozzo A, Catani F, Leardini A. Skin movement artifacts in human movement photogrammetry. *Proceedings of the XIVth Congress of the International Society of Biomechanics* 1993;1:238–9.
- [26] Baillot Y, Rolland JP, Lin K, Wright DL. Automatic modeling of knee-joint motion for the virtual reality dynamic anatomy (VRDA) tool. *Presence: Teleoperators and Virtual Environments* 2000;9(3):223–35.
- [27] Sutherland IE. A head-mounted three-dimensional display. *AFIPS Conference Proceedings* 1968;33(Part I): 757–64.

- [28] Robinett W, Rolland JP. A computational model for stereoscopic optics of a head-mounted display. Presence: Teleoperators and Virtual Environments 1992;1(1):45–62.
- [29] Arun KS, Huang TS, Blostein SD. Least-squares fitting of two 3-D point sets. IEEE Transactions on Pattern Analysis and Machine Intelligence 1987;PAMI-9(5):698–700.
- [30] Spoor CW, Veldpaus FE. Technical Note: rigid body motion calculated from spatial coordinates of markers. Journal of Biomechanics 1980;13:391–3.
- [31] Adelson SJ, Hodges LF. Generating exact ray-traced animation frames by reproduction. IEEE Computer Graphics and Applications 1995;15(3):43–52.
- [32] Press W, Teukolsky S, Vetterling W, Flannery B. Numerical Recipes in C: the Art of Scientific Computing, 2nd ed. Cambridge: Cambridge University Press, 1992.
- [33] Rolland JP, Gibson W, Ariely D. Towards quantifying depth and size perception in virtual environments. Presence: Teleoperators and Virtual Environments 1995;4(1):24–49.
- [34] Rolland JP, Meyer C, Arthur K, Rinalducci E. Methods of adjustments versus method of constant stimuli in the quantification of accuracy and precision of rendered depth in head-mounted displays. Presence: Teleoperators and Virtual Environments, 2002;11(6), in press.
- [35] Welch G, Bishop G. 1997. SCAAT: Incremental Tracking with Incomplete Information. In: Proceedings of the SIGGRAPH 97, Los Angeles, CA, August, 1997. Computer Graphics Proceedings of the Annual Conference Series 1997, ACM SIGGRAPH.
- [36] Cappozzo A, Cappello A, Della Croce U, Pensalfini F. Surface-maker cluster design criteria for 3-D bone movement reconstruction. IEEE Transactions Biomed Eng 1997;44:1165–74.