# A Markerless Augmented Reality Environment for Medical Data Visualization

**Márcio C. F. Macedo[1], Caio S. de B. Almeida[1], Antonio C. S. Souza[3,1], Josildo P. Silva[3,1], Antonio L. Apolinário Jr.[1], Gilson A. Giraldi[2]**

[1]Federal University of Bahia – Salvador,BA – Brazil

[2]National Laboratory for Scientific Computing – Petropolis,RJ – Brazil

[3]Federal Institute of Bahia – Salvador,BA – Brazil

{marciocfm,caiosba,antoniocarlos,josildo,apolinario}@dcc.ufba.br

gilson@lncc.br

***Abstract.*** *Augmented Reality (AR) techniques can be applied in medicine to help physicians in diagnosis, treatment planning, surgery simulation, among others. In craniofacial treatments, it can also be used to support prediction in patient's own body. In this context, AR applications should fill more requirements than usual, like markerless support, tracking deformable objects and volume rendering. This paper presents a markerless AR environment, with support to deformable models, for volumetric medical data visualization based on a simple off-the-shelf hardware.*

## 1. Introduction

Images have been used in medicine since the nineteenth century, giving physicians a tool to improve diagnosis. Magnetic Resonance Imaging (MRI) and Computed Tomography (CT) take one step further, with the possibility to compound a stack of images in a volumetric dataset. Nowadays, with sophisticated computational tools, those images can be visualized in several different ways, improving even more the physician's repertoire of tools to interpret this kind of dataset. Direct volume rendering plays an important role in this process, as it can provide a realistic visualization of internal structures of a patient in 3D. So, the next step to enhance diagnosis could be in-situ visualization. That means, the medical data could be viewed in a way that mimics an exploratory surgery, allowing the physician to see internal structures of the patient in its real environment.

Augmented Reality (AR) is the key technology to achieve in-situ visualization. It makes possible to combine synthetic images in a stream of real scene images, in such a way that real and virtual worlds look like perfectly mixed. The main issues that an AR application must address are: (i) localizing and tracking of objects of interest, (ii) alignment of synthetic objects in the real scene, and (iii) realistic mixing of real and synthetic images [Azuma et al. 2001].

Traditionally, localizing and tracking are made using fiducial markers, which simplify the task of searching for referential objects. Although simple, this strategy has some problems, like occlusion of the fiducial markers, for example. Furthermore, for some application, the inclusion of fiducial markers in the scene is not feasible, like in medicine. Markerless AR is based on two main approaches. The first is based on image analysis. As

previously in fiducial marker based methods, occlusion and lighting conditions still be a challenge. The second approach is based on geometric information. Using depth sensors, a set of 3D points can be acquired from the scene and used to locate and track objects of interest. Although more computationally expensive, this approach is usually independent of lighting conditions. Mixed systems can use both techniques to overcome each fragility.

The focus of this work is to construct a computational markerless AR environment to be used by craniofacial specialists, to simulate and predict surgery results. Craniofacial treatments involve not only functional correction of malformations, but also aesthetic rehabilitation. So, this kind of application is even more complex than ordinary AR applications, because it must deal with deformable objects. The deformation will have a great influence on the whole AR process.

In this scenario, we present an integrated solution to provide physicians a markerless AR environment for medical data visualization focused on volumetric data. Using a simple computational infrastructure, based on AR glasses, depth sensors and personal computers, the physician will be able to visualize 3D volumetric data in-situ, on patient's body. Both patient and physician can move independently because the system will track their movements and provide the physician the right visualization of the patient data. This is achieved by using two Kinect sensors which allow us to reconstruct 3D models of both patient and physician and track their movements using the KinectFusion Algorithm. The visualization is done using volume rendering techniques, allowing a more sophisticated evaluation of medical data, for example, by applying cutting planes or restricting the projected area to some region of a patient's body. This solution also must support deformation of the target object. From an evaluation of the state-of-the-art works in this field, we can conclude that a real-time markerless AR solution with support to volume rendering, non-rigid registration and based on low-cost hardware components has never been proposed before.

The paper is organized as follows. The next section briefly presents some related works. Section 3 gives an overview of the solution proposed in this article. Next, section 4 shows preliminary results and discussions. Finally, the conclusions and future works will be presented on section 5.

## 2. Related work

Medical AR applications can be divided into two basic groups: marker-based and markerless. Over the past decades, many relevant approaches have been proposed in the field of medical AR based on markers. The work most closely related to our approach was proposed in 2008 by Kutter et al. [Kutter et al. 2008]. They proposed a marker-based method for real-time high quality on-patient visualization of volumetric medical data using a Head Mounted Display (HMD). That work focuses on efficient implementations for high quality volume rendering in an AR environment. They also provide occlusion handling for physician's hands. An improvement of this work was proposed by Wieczorek et al. [Wieczorek et al. 2010] to handle occlusions due to medical instruments as well. Also, they included additional effects in the application, such as virtual mirror and multi-planar reformations (MPR).

More recently, a few works have been proposed to deal with markerless medical AR. Maier-Hein et al. [Maier-Hein et al. 2011], for example, proposed a method for

markerless mobile AR for on-patient visualization of medical volumetric data. They use a system in which a 3D depth sensor is mounted on a mobile device and the physician might move this device along the body of the patient to see his anatomical information. This method does not achieve real-time frame rate, but only 10 frames per second (FPS).

The use of fiducial markers provides fast and accurate tracking. However, they are still intrusive, because they are not part of the original scene. Moreover, the hardware of the optical tracking system in some applications is expensive. The computational cost of the markerless tracking in conjunction with the volume rendering techniques prevent to obtain real-time performance in AR applications. However, the approach proposed in this paper is based on markerless tracking and is expected to run entirely in real-time with low-cost hardware components and volume visualization facilities.

## 3. Technique overview

An overview of the whole process proposed in this work can be seen in Figure 1. From an RGB-D (i.e. color + depth) live stream (Figure 1-A), a face detector is used to locate and segment the face from the rest of the scene (Figure 1-B). A 3D reference model is reconstructed with KinectFusion algorithm (Figure 1-C). The 3D reconstruction is stopped and the medical volume is rendered (Figure 1-D). From the localization of the 3D reference model, the medical volume is positioned into the scene and the user is responsible for the semi-automatic fine adjustment of the medical data in terms of 3D scale, orientation and positioning. Markerless live tracking is done based on the registration between the 3D reference model previously reconstructed and the 3D model captured by the sensor in the current frame (Figure 1-E). The region of interest in the medical data is extracted with the context-preserving volume rendering technique (Figure 1-F). Finally, the volumetric data can be rendered based on the focus + context visualization paradigm [Card et al. 1999] (Figure 1-G), in which the patient's anatomy is shown as a focus region inside the context of the patient's body. Moreover, the process supports occlusion (Figure 1-H). Throughout this section the main stages of this process will be discussed, but first a brief description of the computational setup will be presented.

### 3.1. Computational Infrastructure Environment

The computational infrastructure proposed by this work aims to contribute to augmented reality applications on which virtual images should be merged with real images in real time, without using fiducial markers, and considering the angle of view of the observer and the position of the object of interest. The scope of the current version of this markerless augmented reality environment was defined, and it is pictured in Figure 2.

The observer wears augmented reality glasses and is positioned in front of the target (a person, for example). Typically, the observer would be the physician, who will analyse the combination of the real image (part of the patient body) with the virtual image (volumetric rendering result). The observer wears augmented reality glasses, which have two cameras, whose images captured from the observed object will be merged with the virtual image and displayed on two screens. The movement of the observer can be determined by the sensors of the glasses, which returns values that define movements along longitudinal, transversal and vertical axes. The virtual image must be re-rendered in real time according to the movements.

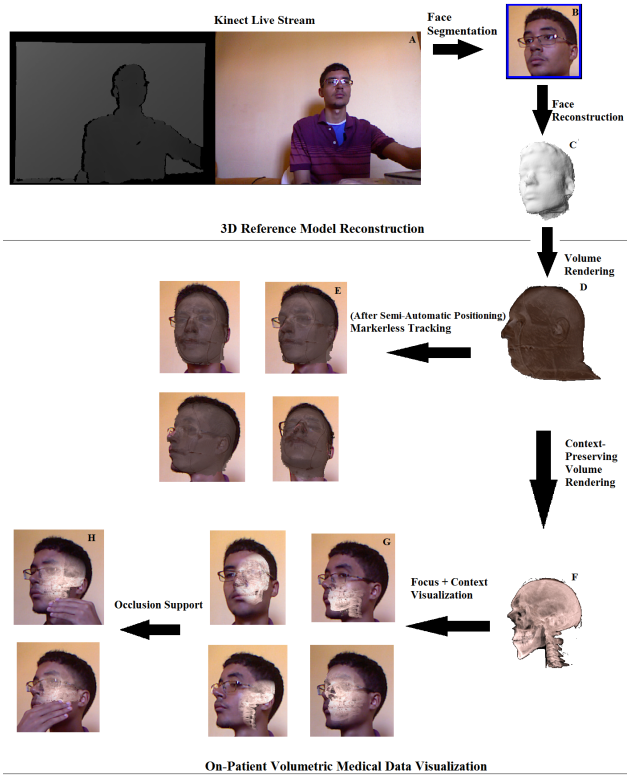


**Figure 1. Schematic overview of the markerless medical AR process proposed in this paper. A) RGB-D live stream. B) A face detector is used to locate the patient's face in the color image. C) The 3D reference model is reconstructed by using the KinectFusion algorithm. D) The medical volume is rendered. E) Markerless AR with the medical volume and the patient's face. F) The region of interest in the medical volume is extracted with the context-preserving volume rendering technique. G) Focus+context visualization is applied to improve the visual perception of the scene. H) The proposed approach supports occlusion.**

The target, typically the patient, is positioned in front of the observer and does not use any fiducial marker. The virtual image should be rendered over the target. To identify the pose of the observed object, two sensors will be used (Figure 2), one to capture the observer and the other to capture the observed object. The sensor that gets information from the observed object contains its model, which will be used to render the virtual image.

The environment described above uses simple hardware. The glasses are Vuzix Wrap 920AR[1], which has a movement tracker. The sensors are Kinect devices[2]. Data from the glasses are parsed and converted into values of yaw, pitch and roll, which define three degrees of freedom. Raw data from the glasses are formatted as 42 bytes blocks, related to x, y and z coordinates of accelerometer, magnetometer and high and low density gyros. The tracker of the glasses is calibrated, assigning minimum and maximum values for each of the four sensors, particularly raw returned values are converted to a range from -180°to 180°(yaw and roll) or a range from -90°to 90°(pitch). Data from the tracker

[1]http://tinyurl.com/wrap920ar
[2]http://www.microsoft.com/en-us/kinectforwindows

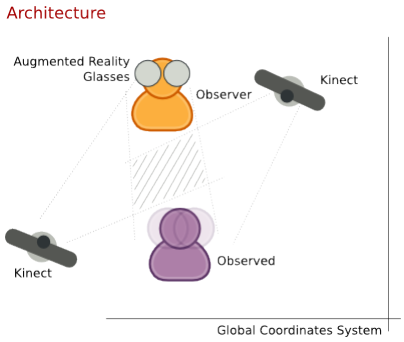


**Figure 2. Global view of the computational environment**

is used to determine the movement of the observer's head. Other positions are determined by Kinect devices.

Two Kinect devices are used to avoid the usage of fiducial markers. Each Kinect provides an RGB-D data stream. Depth map and RGB map are calibrated (*depth registration*) using configuration data stored in Kinect's firmware. After that, the position of the Kinect in relation to a common reference point must be determined. Intrinsic parameters are obtained for each Kinect using traditional calibration methods. The two Kinects are not calibrated simultaneously, this way, interference can be avoided at this point.

The translation *t* (3x1) and rotation *R* (3x3) of an object relative to the camera, equals to the transformation of the object to the camera space, given by $v' = R \cdot v + t$. The reversal of the rotation matrix is simply its transpose, so it's possible to get the transformation of the camera on the global space $R^{-1} = R^T \therefore v = R^T \cdot v' - R^T \cdot t$. Finally, the homogeneous transformation matrix is build based on $R$ and $t$.

Since the reference point is the same for both Kinects, the position of a Kinect in relation to another can be estimated. Once the Kinects were calibrated, the objects that each one sees can be represented in a common coordinate system. So, the next step is define a way to identify and track objects of interest. The strategy to achieve this goal will be explained hereafter.

### 3.2. Markerless Tracking

To track the volumetric medical data in the AR environment without markers, a 3D reference model of the region of interest in the patient is generated. In this work, the region of interest consists in the patient's face. To segment the face from the scene, a Viola-Jones face detector [Viola and Jones 2004] implemented in the Graphics Processing Unit (GPU) is applied in the color image provided by the Kinect sensor. As long as the color and depth sensors of the Kinect are calibrated, the segmention result can be transposed to the depth image. The depth map is denoised using a bilateral filter [Tomasi and Manduchi 1998] and converted into a vertex and a normal map. Then, the KinectFusion algorithm [Izadi et al. 2011] is used to reconstruct the 3D reference model in real-time. As evaluated in [Meister et al. 2012], the KinectFusion algorithm has accuracy of $10mm$. Therefore, it is assumed that its reconstructed models are adequate to be used as reference for tracking in augmented reality applications.

The 3D reference model reconstruction is done only once and it is the basis

for the markerless augmented reality live tracking. The reconstruction is stopped semi-automatically and the user can position the medical data into the scene. Afterwards, the markerless tracking can be started.

The live tracking is done in two steps: during the reconstruction of the 3D reference model, to integrate the different viewpoints into a single reference model, and during the markerless AR with the patient and the medical data. The Iterative Closest Point (ICP) [Rusinkiewicz and Levoy 2001] algorithm is used to estimate the transformation that aligns the current depth frame captured by the Kinect sensor with the previous one represented by the 3D reference model. However, in presence of fast rigid motion, the ICP may fail. To minimize this problem, a real-time head pose estimation is used to give a new initial guess to the ICP to compute correctly the current transformation [Macedo et al. 2013].

Although the estimation of the rigid transformation gives a good approximation of position and orientation of the patient, if its face deforms, ICP can produce a wrong alignment. To overcome this limitation, a deformable model should be introduced, as follows.

### 3.3. Deformation Model

One of the limitations of the markerless tracking described in the previous section is that it does not support non-rigid motion. Therefore, the patient must move his face in front of the sensor with a neutral expression as-rigid-as-possible.

One way to solve this issue is building a space deformation represented by a collection of affine transformations associated with each node in a graph structure. Each node influencing the deformation to the nearby space [Sumner et al. 2007]. To achieve a fast and high quality non-rigid registration, this algorithm must be extended by using GPU implementation and the graph structure must be adapted according to the facial expression, increasing nodes where the deformation is high and decreasing the nodes otherwise. From the graph structure, the affine transformations estimated can be applied on the 3D reference model to align it with the current facial expression captured by the Kinect sensor. Once this solution is fast but does not run in real-time, it is currently being applied in a multi-frame way, for each 10/20 frames, depending on the object's deformation.

### 3.4. Soft Tissue Deformation

The ability to predict the behaviour of the soft tissue under some treatment procedure is an interesting feature that an AR environment could provide to physicians. In the last decades, a wide variety of physically based models has been developed to address the challenge of simulating deformable materials [Nealen et al. 2006]. Those models can be divided into continuum-based, like Finite Element Models (FEM) or discrete approaches as Mass Spring Models (MSM) [Xu et al. 2010]. FEM can describe accurately the behaviour of a wide range of materials, but has a high complexity and computational cost. MSM systems, on the contrary, are widely used in interactive applications in computer graphics, because of its simple mathematical formulation and great versatility for topological changes [Volino and Thalmann 2000]. However, it is difficult to configure its parameters to precisely represent the behaviour of a real deformable object.

We choose the MSM model represent the patient's face deformation in order to obtain realistic behavior and real-time performance. Currently a FEM-based methodology for deriving MSM parameters is under development, based on the method shown in [Baudet et al. 2009] to compute springs constants applying optimization algorithms. Although not incorporated yet in our computational system, the first simulations show encouraging results.

### 3.5. Medical Volume Rendering

Volume rendering is concerned with techniques for generating images from volume data [Hadwiger et al. 2009]. These images can be generated by solving the volume rendering integral based on a emission-absorption optical model, as shown in Equation 1.

$$I(D) = I_0 e^{-\int_{s0}^{D} k(t)dt} + \int_{s0}^{D} q(s) e^{-\int_{s}^{D} k(t)dt} ds. \quad (1)$$

The radiance energy *I(D)* is the result of integrating from entry point into the volume ($s = s_0$) to the exit point toward the camera ($s = D$). The absorbed energy and emission components are represented by the absorption and emission coefficients $k$ and $q$ respectively. The term $I_0$ is the radiance in the entry point $s_0$.

The volume rendering integral cannot be evaluated analytically. Therefore, the numerical computation of the Equation 1 is typically performed according to a compositing scheme known as front-to-back Direct Volume Rendering (DVR). Given the voxel $i$ being traversed, the front-to-back DVR is defined by $C_{dst} = C_{dst} + (1 - \sigma_{dst})C_{src}$ and $\sigma_{dst} = \sigma_{dst} + (1 - \sigma_{dst})\sigma_{src}$ where $C_{dst} = c_{i+1}$, $C_{src} = c_i$, $\sigma_{dst} = 1 - T_{i+1}$, $\sigma_{src} = \sigma_i$ where $C$ represents the color contribution and $\sigma$ the opacity of the voxel.

To render the medical data based on DVR, a single rendering pass ray casting is applied based on the bounding box of the volume [Hadwiger et al. 2009]. To render high quality images from the medical data in real time, several techniques are employed: stochastic jittering (i.e. random ray-start off-setting) to reduce sampling artifacts, empty-space leaping to skip non-visible voxels [Li et al. 2003], early ray termination if the opacity accumulated is sufficiently high, pre-integrated transfer functions [Engel et al. 2001] to capture the high frequencies introduced in the transfer functions with low sampling rates and Blinn-Phong illumination with on-the-fly gradient computation to add realism in the final rendering. To dynamically define the region of interest to be rendered in the medical volume, the context-preserving volume rendering [Bruckner et al. 2005] is applied.

After the medical volume rendering, the color frame buffer of the volume is loaded and sent to a shader to blend it with the RGB data coming from the Kinect sensor. The blending is done by the linear interpolation $I_{final} = \beta * I_{real} + (1 - \beta) * I_{medical}$ where $I_{real}$ is the image captured by the sensor, $I_{medical}$ is the image corresponding to the medical volume, and $I_{final}$ is the resulting image. The contribution of each image (the value $\beta$) is dynamically defined by using the focus + context visualization proposed in [Bichlmeier et al. 2007].

Incorrect occlusion of virtual and real objects in an augmented scene is one of the fundamental problems in AR. To solve it, the depth images of the 3D reference object pre-

viously reconstructed (reference) and the 3D object coming from the sensor's live stream are used. If the live object is in front of the reference object, the volume is the ocludee, otherwise, it is the occluder.

## 4. Results and Discussion

In this section we analyse the performance and the visual quality of the proposed approach for each one of the techniques described in the previous section.

The evaluation of the proposed approach is conducted in a scenario where the patient's head is augmented with a generic CT volumetric dataset of a head. The use of a generic volume does not affect our evaluation, as our main interest is evaluate the performance and visual quality of the proposed approach. Meanwhile, the deformation model presented is evaluated in terms of performance and accuracy in a real situation.

The medical dataset used is the CT volumetric data of the Visible Male's head (from Visible Human Project) [3] of resolution $128 \times 256 \times 256$.

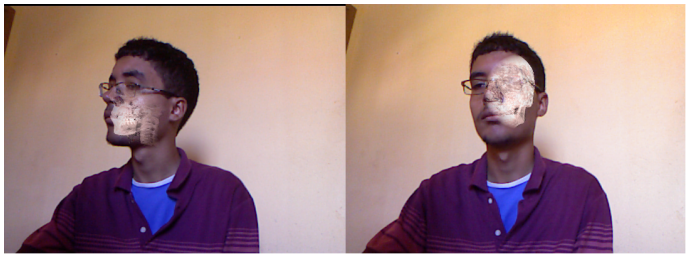

**Figure 3. On-patient medical data focus+context visualization. (Left) focus region in the context of the patient's face. (Right) real structures with high curvature (e.g. nose) are still visible even if they are inside the target region.**

The 3D reference model reconstruction runs in 30 FPS. The on-patient medical data focus + context visualization based on markerless tracking runs in 20 FPS, being the context-preserving volume rendering the most expensive method in this solution. As can be seen in Figure 3, the visualization method enhances the comprehension of the scene by enabling the visualization of a focus region on the medical data in the context of the patient's face. Also, this method preserves the visualization of regions with high curvature even if it is inside the focus region.

The deformation model was tested on a typical dataset captured by the Kinect sensor. To illustrate a real situation we asked the patient to deform his face by inflating his cheeks. As can be seen in Figure 4, the deformation algorithm captured the main deformation presented on the cheeks. The accuracy obtained with the current implementation is $70\%$ on average, for an average time of $400$ms. The accuracy can be increased at the expense of runtime.

## 5. Conclusion and Future Works

In this paper it was presented a multiview, marker-free augmented reality approach for on-patient volumetric medical data visualization. The user wears augmented reality glasses in order to interact with this environment. The markerless live tracking is achieved by

[3] http://www.nlm.nih.gov/research/visible/

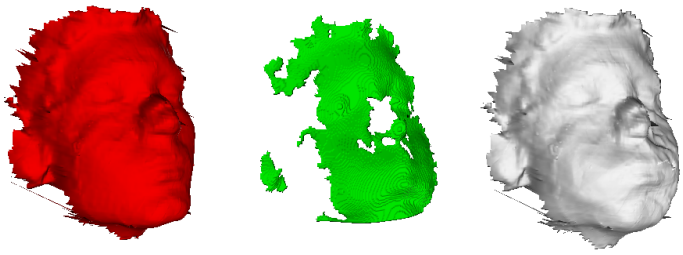

**Figure 4. (Left) Source object. (Middle) Target object. (Right) Source object post non-rigid registration.**

using a 3D reference model of the patient's head and the ICP algorithm. An algorithm for fast non-rigid registration of surfaces using GPU was presented, which is used in a multi-frame scheme to adjust the rigid registration. Moreover, volume rendering techniques were applied to enable the focus + context visualization of the medical structures in the real scene. This approach runs in real-time with a generic volume of typical size. Also, it improves the human visual perception of the scene.

To further improve the visual quality of the scene, fast global illumination or the recovery of the real illumination of the scene could be applied to improve the realism of the volume rendering and the integration with the real scene. The non-rigid registration balance between performance and precision is statically and manually set. An algorithm that automatically controls the precision and time, and obviously the multi-frame rate, is currently under investigation. Moreover, an evaluation in a real surgical environment must be performed in order to validate the accuracy of the full solution proposed in this paper.

## Acknowledgments

The authors would like to acknowledge the support of FAPESB, CAPES and CNPq for this work. Also, they are grateful to the PCL project for providing the open-source implementation of the KinectFusion algorithm.

## References

Azuma, R., Baillot, Y., Behringer, R., Feiner, S., Julier, S., and MacIntyre, B. (2001). Recent advances in augmented reality. *IEEE Comput. Graph. Appl.*, 21(6):34–47.

Baudet, V., Beuve, M., Jaillet, F., Shariat, B., and Zara, F. (2009). Integrating tensile parameters in mass-spring system for deformable object simulation. Technical Report RR-LIRIS-2009-034, LIRIS UMR 5205 CNRS/INSA de Lyon/Université.

Bichlmeier, C., Wimmer, F., Heining, S. M., and Navab, N. (2007). Contextual anatomic mimesis hybrid in-situ visualization method for improving multi-sensory depth perception in medical augmented reality. In *ISMAR*, pages 1–10. IEEE Computer Society.

Bruckner, S., Grimm, S., Kanitsar, A., and Gröller, M. E. (2005). Illustrative context-preserving volume rendering. In *EUROVIS*, pages 69–76. Eurographics Association.

Card, S. K., Mackinlay, J. D., and Shneiderman, B., editors (1999). *Readings in Information Visualization: Using Vision to Think*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.

Engel, K., Kraus, M., and Ertl, T. (2001). High-quality pre-integrated volume rendering using hardware-accelerated pixel shading. In *Proceedings of the ACM SIGGRAPH/EUROGRAPHICS workshop on Graphics hardware*, pages 9–16. ACM.

Hadwiger, M., Ljung, P., Salama, C. R., and Ropinski, T. (2009). Advanced illumination techniques for gpu-based volume raycasting. In *ACM SIGGRAPH 2009 Courses*, pages 2:1–2:166. ACM.

Izadi, S., Kim, D., Hilliges, O., Molyneaux, D., Newcombe, R., Kohli, P., Shotton, J., Hodges, S., Freeman, D., Davison, A., and Fitzgibbon, A. (2011). Kinectfusion: real-time 3d reconstruction and interaction using a moving depth camera. In *UIST*, pages 559–568. ACM.

Kutter, O., Aichert, A., Bichlmeier, C., Traub, J., Heining, S. M., Ockert, B., Euler, E., and Navab, N. (2008). Real-time Volume Rendering for High Quality Visualization in Augmented Reality. In *AMI-ARCS*.

Li, W., Mueller, K., and Kaufman, A. (2003). Empty space skipping and occlusion clipping for texture-based volume rendering. In *VIS*, pages 317–324. IEEE.

Macedo, M., Apolinario, A., and Souza, A. C. (2013). A Robust Real-Time Face Tracking using Head Pose Estimation for a Markerless AR System. In *SVR*.

Maier-Hein, L., Franz, A. M., Fangerau, M., Schmidt, M., Seitel, A., Mersmann, S., Kilgus, T., Groch, A., Yung, K., dos Santos, T. R., and Meinzer, H.-P. (2011). Towards mobile augmented reality for on-patient visualization of medical images. In *Bildverarbeitung für die Medizin*, pages 389–393. Springer.

Meister, S., Izadi, S., Kohli, P., Hämmerle, M., Rother, C., and Kondermann, D. (2012). When can we use kinectfusion for ground truth acquisition? In *IROS*. IEEE.

Nealen, A., Müller, M., Keiser, R., Boxerman, E., and Carlson, M. (2006). Physically based deformable models in computer graphics. *Computer Graphics Forum*, 25(4):809–836.

Rusinkiewicz, S. and Levoy, M. (2001). Efficient variants of the ICP algorithm. In *3DIM*.

Sumner, R. W., Schmid, J., and Pauly, M. (2007). Embedded deformation for shape manipulation. *ACM Trans. Graph.*, 26(3).

Tomasi, C. and Manduchi, R. (1998). Bilateral filtering for gray and color images. In *Computer Vision, 1998. Sixth International Conference on*, pages 839 –846.

Viola, P. and Jones, M. J. (2004). Robust real-time face detection. *Int. J. Comput. Vision*, 57(2):137–154.

Volino, P. and Thalmann, N. (2000). *Virtual Clothing.: Theory and Practice.* Springer-Verlag GmbH.

Wieczorek, M., Aichert, A., Kutter, O., Bichlmeier, C., Landes, J., Heining, S. M., Euler, E., and Navab, N. (2010). GPU-accelerated Rendering for Medical Augmented Reality in Minimally-Invasive Procedures. In *Proceedings of BVM 2010*. Springer.

Xu, S., Liu, X., Zhang, H., and Hu, L. (2010). An improved realistic mass-spring model for surgery simulation. In *HAVE*, page 1–6. IEEE.