3D Annotation and Manipulation of Medical Anatomical Structures

Dime Vitanovski, Christian Schaller, Dieter Hahn, Volker Daum, Joachim Hornegger
Chair of Pattern Recognition, Martensstr. 3, 91058 Erlangen, Germany

ABSTRACT

Although the medical scanners are rapidly moving towards a three-dimensional paradigm, the manipulation and annotation/labeling of the acquired data is still performed in a standard 2D environment. Editing and annotation of three-dimensional medical structures is currently a complex task and rather time-consuming, as it is carried out in 2D projections of the original object. A major problem in 2D annotation is the depth ambiguity, which requires 3D landmarks to be identified and localized in at least two of the cutting planes. Operating directly in a three-dimensional space enables the implicit consideration of the full 3D local context, which significantly increases accuracy and speed. A three-dimensional environment is as well more natural optimizing the user's comfort and acceptance. The 3D annotation environment requires the three-dimensional manipulation device and display. By means of two novel and advanced technologies, Wii Nintendo Controller and Philips 3D WoWvx display, we define an appropriate 3D annotation tool and a suitable 3D visualization monitor. We define non-coplanar setting of four Infrared LEDs with a known and exact position, which are tracked by the Wii and from which we compute the pose of the device by applying a standard pose estimation algorithm. The novel 3D renderer developed by Philips uses either the Z-value of a 3D volume, or it computes the depth information out of a 2D image, to provide a real 3D experience without having some special glasses. Within this paper we present a new framework for manipulation and annotation of medical landmarks directly in three-dimensional volume.

Keywords: Image Display, Visualization

1. DESCRIPTION OF PURPOSE

The idea of annotating/labeling medical structures is to create correspondences between 2D slices. Once structures or medical landmarks are annotated, it can be constructed a segmented mesh of the anatomical structure. Even though the medical scanners are rapidly moving towards a three-dimensional standard, the manipulation and annotation/labeling of the acquired data is still performed in a regular 2D environment, where the procedure is repeated slice by slice in order to annotate the whole three-dimensional object. Since 3D medical structures can be very complex, annotating the surface of anatomical shape might become a quite time-consuming task, involving not only the labeling step, but also rotating, translation and zooming of the view is required, in order to find the portions of the surface to draw on. A major problem in 2D annotation is the depth ambiguity, which requires 3D landmarks to be identified and localized in at least two of the cutting planes. The standard planes in clinical settings are sagittal, coronal and axial (see Figure 1).

Since annotation has a major function in different fields (e.g. machine learning algorithms use the annotated structures as prior knowledge for training classifiers, extracting clinical features out of an annotated mesh or pre-annotated 3D anatomical atlases for students education purposes) different research groups has developed diverse 3D annotations tools. The Scientific Computing and Imaging (SCI) at the University of Utah has developed Annot3D, an 3D annotation system that allows users to add various 3D annotations to 3D volume data from CT scans (DICOM format).[1] 3D pen based annotation approach has been presented by the research group of the technical university of Cluj-Napoca.[2] The lack of current labelling environments is still the depth information. In order to avoid the problems by annotating in 2D, the annotation has to be performed direct in the three dimensional object, where information about the object and its features is complete and the level of annotation can be therefore more accurate. In this paper we introduce a new framework for annotation directly in 3D volume by using the 3D Philips WoWvx display.

Further author information: (Send correspondence to Dime Vitanovski)
Dime Vitanovski: E-mail: vitanovski@gmail.com, Telephone: +49 179 4869157

Medical Imaging 2009: Visualization, Image-Guided Procedures, and Modeling,
edited by Michael I. Miga, Kenneth H. Wong, Proc. of SPIE Vol. 7261, 72611H
 · doi: 10.1117/12.812453

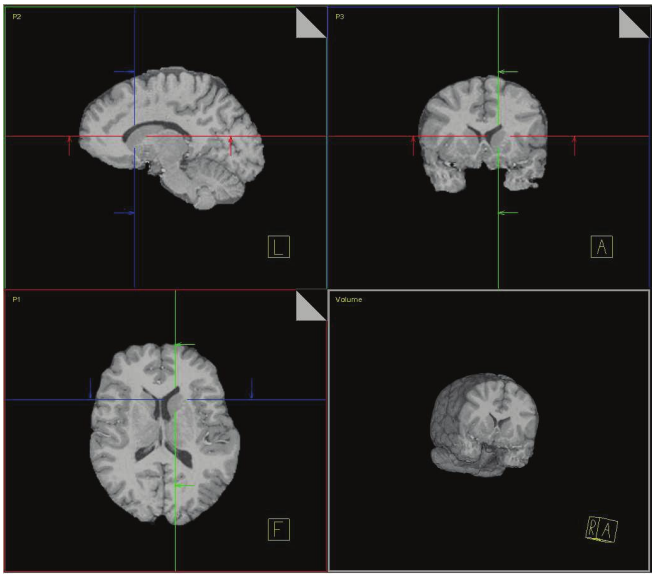

Figure 1. Standard Clinical View Planes(Axial, Sagittal, Coronal)

2. METHODS

The 3D annotation environment defines the three-dimensional manipulation device and display. The user interaction within our system is intuitive and accurate, while the conversion into the 3D environment is based on standard technology. The Nintendo Wii Controller,[3] also called Wiimote, (see Figure 2), developed by the company Nintendo, can measure the earth acceleration and can track up to 4 Infrared LEDs (IR) but does not detect its exact pose. We defined non-coplanar setting of four Infrared LEDs with a known and exact position,

Figure 2. Wii Nintendo Controller[4]

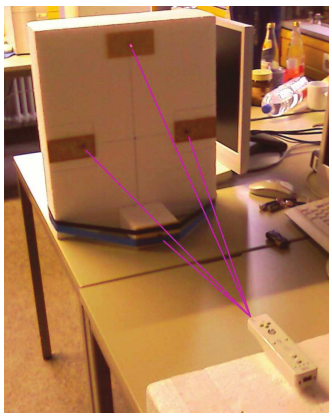

Figure 3. Infrared beacon

which are tracked by the Wii Controller and from which we compute the pose of the device. The translation and orientation is determined using the Direct Linear Transformation (DLT), which performs a non-linear optimization on the four point correspondence. The energy function for the camera pose estimation can be defined as the quadratic difference between the ground truth value and the projected values. The Wiimote pose estimation algorithm is more detailed introduced in the following sections.

2.1 Camera models

The camera pose estimation requires a definition of a corresponding camera model. In the case of the Wiimote we use the pinhole camera model. In[5] two camera models are presented

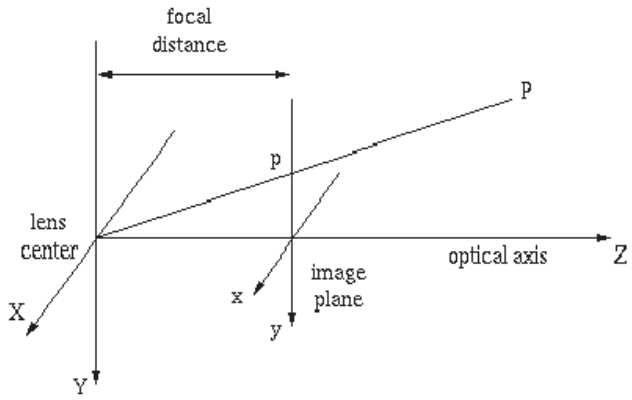

Figure 4. Pinhole Camera Model

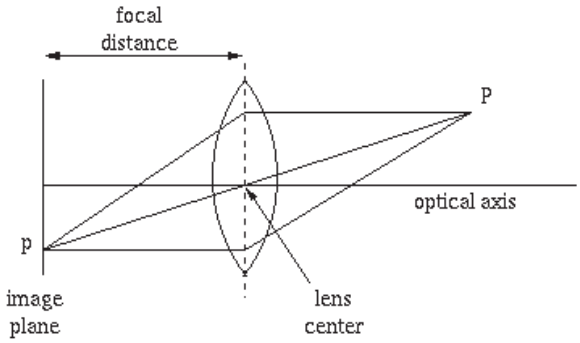

Figure 5. Thin Lense Camera Model

- Pinhole Camera and
- Thin Lens Camera

The pinhole camera does not have a conventional glass lens. It has an small hole in a thin material that can focus light by confining all rays from a scene through a single point (Figure 4).

With the thin lens camera model, the mapping from a 3D object point to the 2D image plane is performed by a convex lens (Figure 5). The object is sharply mapped onto the 2D image plane only when all projection rays pass through the same point on the image plane. If the variances of the projection rays is too small, so that the human eye cannot detect it, the object mapping remains sharp. The perspective projection corresponds to the pinhole camera model. From (Figure 4) one can derivate the mathematical formula for the projection of a 3D point in camera coordinates to the 2D image coordinates:

$$\frac{x_A}{x_{A'}} = \frac{z_A}{f} \tag{1}$$

2.2 Intrinsic Camera Parameters

The computation of the intrinsic camera parameters is one of the most studied problems in computer vision. The traditional way to compute the intrinsic parameters is to use a known calibration object and as an input only information for 3D/2D point correspondences is required. Since the parameters are not depend on the position and orientation of the camera in space, they can be computed once by doing offline camera calibration, or if a sufficient number of point correspondences is available, they can be incorporated in the camera pose estimation. In the case of the Wiimote, an ofline camera calibration is required.
The intrinsic camera matrix is defined with five parameters: f_x, f_y, s, H_x, H_y. The real pixel coordinates are

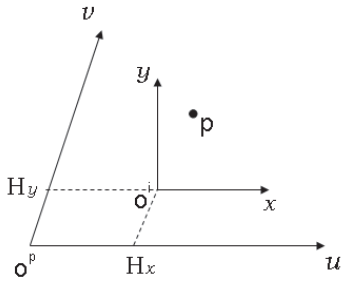

Figure 6. Intrinsic Camera Parameters(o^i : origin of the ideal image coordinate system, o^p: origin of the pixel coordinate system)

computer by multiplication with the camera matrix, where f_x, f_y is the focal length of the camera, s(skew) is the angle between the sensors and H_x, H_y is the principal point of the real coordinate system u/v (Figure 6)

$$K = \begin{bmatrix} f_x & s & H_x \\ 0 & f_y & H_y \\ 0 & 0 & 1 \end{bmatrix} \tag{2}$$

The cartesian coordinate system x/y is the ideal image coordinate system with origin o^i. The coordinate transformation for the ideal coordinate system is performed by multiplying with the perspective projection.

$$P = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \tag{3}$$

2.3 Pose Estimation with DLT Method from known 4 2D-3D Point Correspondences

The Direct Linear Transformation (DLT) is one of the most commonly used approaches to camera calibration. With 6 point correspondences it can be rewritten in a closed form solution, which can be solved using SVD. Since the Wiimote IR camera can track up to 4 IR points, a non-linear optimization technique is required, in order to solve the problem of pose estimation. Hence, an ofline camera calibration is necessary.

Equation 4 defines the formula for the 2D/3D correspondences:

$$\begin{aligned} \begin{pmatrix} x^p \\ y^p \end{pmatrix} &= \begin{pmatrix} f_x & s & H_x \\ 0 & f_y & H_y \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} R^T & -R^T T \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x^w \\ y^w \\ z^w \\ 1 \end{pmatrix} \\ &= \begin{pmatrix} p_{11} & p_{12} & r_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \end{pmatrix} \begin{pmatrix} x^w \\ y^w \\ z^w \\ 1 \end{pmatrix} \end{aligned} \tag{4}$$

from which the ground truth values x_p, y_p can be computed:

$$x^p = \frac{x^w p_{11} + y^w p_{12} + z^w p_{13} + t_x}{x^w p_{31} + y^w p_{32} + z^w p_{33} + t_z} \tag{5}$$

$$y^p = \frac{x^w p_{21} + y^w p_{22} + z^w p_{23} + t_y}{x^w p_{31} + y^w p_{32} + z^w p_{33} + t_z} \tag{6}$$

The equations can be linearized by multiplying with the denominator:

$$\begin{aligned} x^p(x^w p_{31} + y^w p_{32} + z^w p_{33} + t_z) &= x^w p_{11} + y^w p_{12} + z^w p_{13} + t_x \\ y^p(x^w p_{31} + y^w p_{32} + z^w p_{33} + t_z) &= x^w p_{21} + y^w p_{22} + z^w p_{23} + t_y \end{aligned} \tag{7}$$

The energy function for the camera pose estimation can be defined as the quadratic difference between the ground truth value and the projected values:

$$f(R,T) = \underset{\hat{R},\hat{T}}{\operatorname{argmin}} \sum_{i=1}^{4} \begin{pmatrix} x_i^p(x_i^w p_{31} + y_i^w p_{32} + z_i^w p_{33} + t_z) - x_i^w p_{11} + y_i^w p_{12} + z_i^w p_{13} + t_x \\ y_i^p(x_i^w p_{31} + y_i^w p_{32} + z_i^w p_{33} + t_z) - x_i^w p_{21} + y_i^w p_{22} + z_i^w p_{23} + t_y \end{pmatrix}^2$$

The camera pose is computed by optimizing the energy function with respect to the rotation and the translation. In section 2.4 one optimization technique is introduced.

2.4 Non-Linear Programming with IPOpt

In this section one advanced technique for non-linear programming is presented, which is used for optimization of the energy functions defined in the previous sections.
A Nonlinear Program (NLP) is a problem that can be defined in the following form:

$$\begin{array}{lllll} \text{minimize} & F(\boldsymbol{x}) & & & \\ \text{subject to} & g_i(\boldsymbol{x}) = 0 & \text{for } i = 1, ..., m_1 & \text{where } m_1 >= 0 \\ & h_j(\boldsymbol{x}) >= 0 & \text{for } j = m_{1+1}, ..., m & \text{where } m >= m_1 \end{array}$$

$F(\boldsymbol{x})$ is the objective function, where $\boldsymbol{x}$ is a vector of several variables, and $g_i(x), h_i(x)$ are constraints. The goal of the NLP is to minimize the objective function fulfilling the constraints limitations.
One of the greatest challenges in NLP is that some problems exhibit local optima that is, spurious solutions that merely satisfy the requirements on the derivatives of the functions. Algorithms that propose to overcome this difficulty are termed Global Optimization.
Andreas Waechter and Lorenz T. Biegler introduce in[6] a new approach to NLP, the so called IPOPT (**I**nterior **P**oint **OPT**imizer) algorithm, used for large-scale nonlinear programming, which implements a primal-dual interior point method, and uses line search based on filter methods. In particular, these methods provide an attractive alternative to active set strategies in handling problems with large numbers of inequality constraints. A detailed mathematical explanation for IPOPT can be found in.[6]
In order to optimize the energy functions, described in section 2.3, analytical or numerical computation of the first and second order derivates is required and the sparse structure for the Jacobian and Hessian matrix. Because of its fact convergence, IPOPT can provide a real-time optimization. Using the mathematical framework presented in the previous sections we could estimate the position and the orientation of the Wiimote in real-time. Using the mathematical framework presented in the previous sections we could estimate the position and the orientation of the Wiimote in real-time. In our framework one can use the Wiimote as a 3D mouse for 3D volume manipulation. In our In the next section the state-of-the-art in visualization techniques is presented.

2.5 3D Displays

Although most of the applications use 3D dataset, at the end they deliver 2D image. Standard monitors (displays) can provide only 2D information, or they can visualize a 3D object, where the depth information about the object is missing, and therefore, they are not suitable for reliable, fast and accurate direct 3D annotation. Recent developed monitors can merge 2D image along with its depth information to enhance the 3D object visualization (Philips 3D WOWvx).[7] This new technology enables a realistic 3D volume rendering without special glasses(see Figure 7 and Figure 8).

The 3D display uses the Z-information of a 3D volume, or it estimates the depth value out of a 2D image, to provide real 3D visualization. The interface to a Philips 3D Display is based on 2D and Depth images, where the

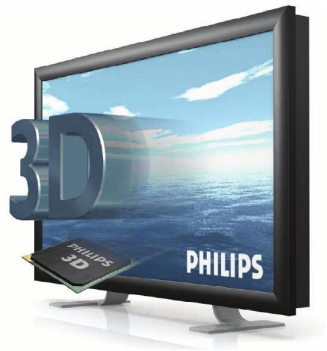

Figure 7. Philips 3D Display[8]

Figure 8. Philips 3D Display[9]

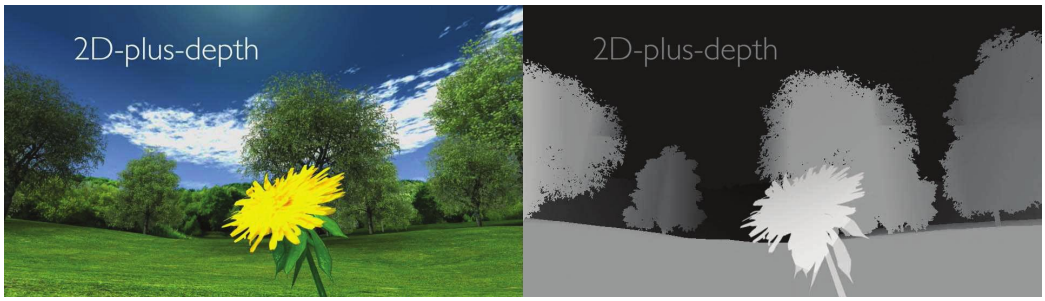

Figure 9. 2D image with its depth map[10]

depth map describes the disparity (see Figure 9). Pixels of the depth map correspond to pixels of the 2D image and indicate the distance of the corresponding 2D pixel to the observer. From the formula for the disparity computation (see Formula 8), one can estimate the Z-value for a particular voxel (see Formula 9).[11]

$$D(Z) = M * (1 - \frac{vz}{Z - Zd + vz}) + C \tag{8}$$

$$Z = \frac{M * vz}{M + C - D} + Zd - vz \tag{9}$$

where D is the disparity, Z the depth and Zd,vz,C and M are constants. Using the APIs provided by Philips we have implemented a bridge between the Visualization Toolkit (VTK) and the Philips display. With the VTK platform we have constructed a 3D volume out of a CT DICOM series and passed it to the Philips 3D renderer. The manipulation control together with the enhanced 3D visualization defines the base line for an advance three-dimensional environment.

3. RESULTS

The accuracy of the manipulation system was evaluated with ideal data from a robot arm. In the following table the accuracy of the controller is introduced:

	R. around X axis	R. around Y axis	T. in X direction	T. in Y direction	T. in Z direction
Mean error	0.16212	0.28624	0.38446cm	0.41119cm	4.94116cm
Std.Deviation	0.40264	0.53502	0.62005cm	0.64124cm	2.22287cm

Table 1. Wiimote Accuracy Measurements (R = rotation, T = translation)

Since the robot arm (Scorbot -ER VII, see Figure 10) can not rotate around the Y-axis, it was not possible for us the measure the accuracy of the Wiimote for this axis. We belief, that the precision of the controller lie in the same range as for rotation error for the X-axis and the Z-axis.

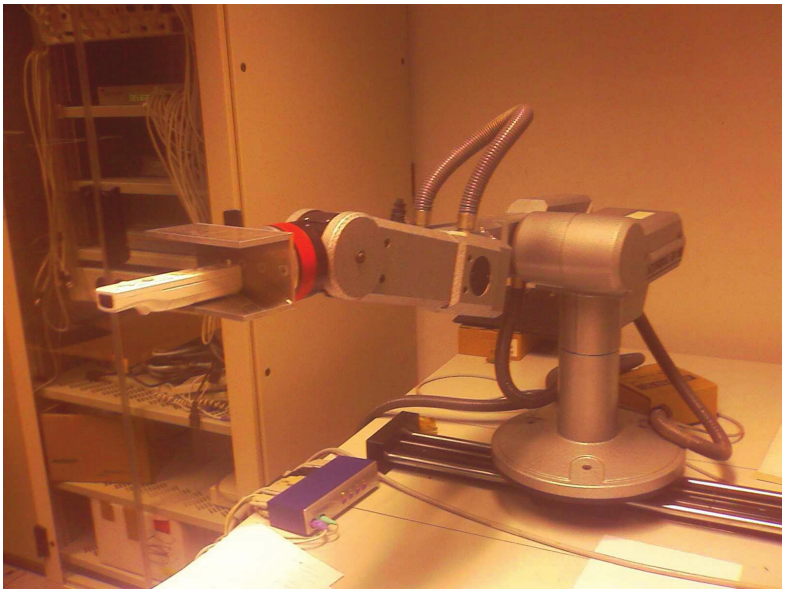

Figure 10. Robot arm (Scorbot ER VII)

4. CONCLUSION

Even though the standard medical scanners provide the clinicians with three-dimensional information, the manipulation and annotation of the acquired data is still performed in 2D environment. Although many annotation frameworks claim a 3D annotation environment, at the end they provide a 2D image. The lack of the currently annotation frameworks is the missing depth information. 3D annotation environment requires a three-dimensional manipulation device and display. Using two standard technologies, Nintendo Wiimote Controller and Philips WoWvx, we introduced a new framework for annotating anatomical structures direct in the three-dimensional volume. By default the Wiimote controller only measure rough acceleration and orientation. Therefore, a pose estimation algorithm was developed for computing accurate position and orientation in 3D space. Within our framework the exact rotation and translation of the Wiimote is determined by optimizing an energy function by means of IPOPT. In order to overcome the measurements noise and estimation error, pose estimation is filtered using the standard Kalman model. The presented mathematical framework enables an accurate pose estimation computed in real-time which satisfies catheter navigation requirements. Operating directly in a three-dimensional space enables the implicit consideration of the full 3D local context, which significantly increases accuracy and speed. A three-dimensional environment is as well more natural optimizing the user's comfort and acceptance.

REFERENCES

[1] Simpson, J. and Balling, J., "Annot3d and packaging of 3d visualizations for educational purposes," in [*In Proceedings of The 12th Annual Medicine Meets Virtual Reality Conference*], pp. in (press) (2004).

[2] Gorgan, D., Stefanut, T., and Gavrea, B., "Pen based graphical annotation in medical education," *Twentieth IEEE International Symposium on Computer-Based Medical Systems* (2007).

[3] Wiimote, "http://wii.com/," (2006).

[4] Source: http://www.newscientist.com.

[5] Wenhardt, J., Schmidt, H., and Niemann, H., [*Rechnersehen mit Anwendung in der Augmented Reality sowie beim bildbasierten Rendering*], ch. Projektionsmodelle, Kameraparameter, 3-D-Rotationen, 2–13, LME (2006).

[6] Waechter, A. and Biegler, L. T., "On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming," 25–57 (2006).

[7] Philips 3D WoWvx, Source: http://www.wowvx.com/.

[8] Source: www.technologydigger.com.

[9] Source: www.tvsreview.com.
[10] Source: http://images.dailytech.com.
[11] Source: http://www.business-sites.philips.com.