Augmented virtuality based on stereoscopic reconstruction in multimodal image-guided neurosurgery: methods and performance evaluation.

Perrine Paul, Oliver J. Fleig, Pierre Jannin

HAL Id: inserm-00136469

https://inserm.hal.science/inserm-00136469

Submitted on 16 May 2007

Augmented Virtuality based on Stereoscopic Reconstruction in Multimodal Image-Guided Neurosurgery: Methods and Performance Evaluation

Perrine Paul, Oliver Fleig and Pierre Jannin, *Member IEEE*

Abstract— This paper presents a new method for displaying in the same 3D scene multimodal preoperative images of a patient and images of the operative field viewed through surgical microscope binoculars for image-guided neurosurgery.

Matching real world information, i.e., the operative field, with virtual world information, i.e., preoperative images of the patient, is an important issue in image-guided neurosurgery. This can be achieved by superimposing preoperative images onto a surgical microscope ocular or a head-mounted display. Such an approach is usually called augmented reality (AR). When surgery is performed in functional areas, such as the eloquent cortex, multimodal images are required. Therefore, preoperative images consist of a complex 3D multimodal scene which can hamper the vision of the real world when displayed in the neurosurgeons view of the operative field.

The approach, introduced in this paper, is called augmented virtuality (AV) and involves displaying the operative field view in the virtual world, i.e., in the 3D multimodal scene which includes preoperative images of the patient. Information from the operative field consists of a 3D surface reconstructed from two stereoscopic images from surgical microscope binoculars using stereovision methods. As the microscope is part of a neuronavigation system and is tracked by an optical 3D localizer, the 3D reconstructed surface is directly expressed in the physical space coordinate system. Using the image-to-physical space transformation computed by the neuronavigation system, this 3D surface can also be directly expressed in the image coordinate system.

In this paper, we present the method for reconstructing 3D surfaces of the operative field from stereoscopic views and matching the reconstructed surface with the preoperative images. Performance evaluation of this method was performed using a physical skull phantom. 300 image pairs of this phantom were acquired. The distance between the reconstructed surfaces and the skull surface segmented from a CT data set of this phantom was used as a system accuracy measurement. Our method was used for 6 clinical cases with lesions in eloquent areas. For the minimum microscope focus value, 3D reconstruction accuracy alone was shown to be within 1mm (median: 0.76mm ± 0.27), whereas virtual and real image matching accuracy was shown to be within 3mm (median: 2.29mm ± 0.59), including the image-to-physical space registration error.

Clinical use of this system has proved the relevance of our approach. In addition to seeing beyond the surface, augmented virtuality can be used to see around the surgical area. With this system, neurosurgeons and clinical staff in the OR were able to interact with the resulting 3D scene by rotating and modifying transparency features. This AV system facilitates understanding of the spatial relationship between the operative field and the complex 3D multimodal scene, which includes preoperative images of the patient.

Index Terms— Preoperative and Intraoperative Multimodal Images, Stereopsis, Performance Evaluation, Neuronavigation, Image-Guided Neurosurgery.

I. INTRODUCTION

MULTIMODAL Image-Guided Neurosurgery (MIGN) involves assisting the neurosurgeon with multimodal preoperative images of the patient during both the planning and operative stages [1], [2], [3], [4]. During the planning stage, relevant information is selected from the functional and anatomical preoperative images: the target (e.g., lesion), areas to be avoided (e.g., language, motor or somesthetic areas from functional MRI (fMRI) or MagnetoEncephalography (MEG)) and reference anatomical structures (e.g., cortical sulci). During the operative stage, image-to-physical space geometrical transformation is computed using a 3D localizer and registration methods, parts of neuronavigation systems. Image guidance mainly consists of visually matching preoperative images with the neurosurgeons vision of the operative field. This visual matching still remains an important issue in MIGN. The usual approach is to enhance reality with virtual information. This approach is called *Augmented Reality* (AR). Preoperative images are displayed as 2D monochromatic contours inside microscope oculars, aligned with the patient [5] whereas these images are multimodal and 3D. To account for this limitation, AR with 3D graphics has been proposed by [6], but only to display the 3D surface of the target lesion. In [7] a head-mounted display with stereo is used to visualize brain structures. The 3D information can also be overlayed onto a 2D video of the patient [8], [9], [10] displayed on an external screen, by registering the images to the patient and computing camera pose. Extending this approach to multimodal images raises the issue of how to display a large amount of information without obstructing the normal vision of the operative field of view (FOV). A complementary approach is to superimpose real information from the neurosurgeon's FOV onto the virtual information, instead of superimposing virtual information onto the real information. An external media, such as the neuronavigation workstation screen, is used for display. This approach is called *Augmented Virtuality* (AV) [11]. Visual matching between real and virtual images can be partially characterized by the dimension of the information shared between the real and virtual worlds [12]. In neuronavigation systems, the information sent to the virtual world is 0D or 1D, i.e., the pointer location or trajectory displayed in the preoperative images. More recently, the use of an endoscopic

camera was proposed to either texture the 3D preoperative surfaces with camera views [13], or create transparent virtual endoscopic views beyond which 3D preoperative structures can be seen [14]. In these systems, the information sent to the virtual world is 2D. In our AV approach, this information consists of 3D surfaces. The higher the dimension, the easier it is to understand the spatial relationship between the virtual and real world as more information is available to the surgeon. We suggest an AV approach which is totally non-invasive and which does not require any additional expensive hardware. We previously proposed mapping 2D direct light images of the intraoperative FOV to preoperative images [11]. We then applied stereoscopic reconstruction techniques to this approach to acquire 3D surface intraoperative data sets [15]. Similar proposals, using images acquired by cameras fixed to the surgical microscope, have been developed by [16] and [17]. In [16], Skrinjar made a global lambertian assumption and reconstructed surfaces from image gradient computation. In [17], Sun applied his system,which is quite similar to ours, for deformation quantification, not for visualization. In [18], a laser grid was projected onto the brain surface using a commercial 3D laser range sensor again for surface deformation tracking purposes. In [19], the same kind of system was used to register preoperative data with the patient in the operating room.

This paper presents a method for creating AV 3D scenes and focuses on visualization features using surface mesh reconstructions of the operative FOV. We also describe the performance evaluation of our method using a physical phantom. Accuracy and influence of certain input parameters are evaluated on 300 reconstructed rough surface meshes and on the same number of smooth surface meshes. The method was tested on 6 clinical cases, one of which is detailed in this paper. The clinical advantages of this new visualization approach are emphasized in the context of brain surgery, mainly surgery of cortical lesions located in eloquent areas where multimodal preoperative images are needed.

II. Materials and Methods

A. Multimodal acquisition and planning

Before surgery, the patient underwent anatomical and functional examinations. Volumes of interest were segmented from these examinations and registered with the anatomical examination (e.g., 3D T1 MRI). The anatomical examination coordinate system is called the "image coordinate system". These volumes of interest were displayed as 3D surfaces and included the cortical surface, target volume, functional areas to be avoided, cortical sulci used as references or surgical trajectories [1]. The neurosurgeon planned his or her operative gesture by selecting relevant information that he or she would like to have available during surgery. This information was saved as 3D surfaces in Visualization Toolkit (VTK) format [20].

B. The neuronavigation system

The operative room was equipped with a standard neuronavigation system (SNN-SNS) and a surgical microscope (Carl Zeiss NC4) featuring light emitting diodes (LEDs). LEDs were also fixed on a pointer and a Dynamic Reference Frame (DRF) attached to the patient's head. The coordinate system defined by the DRF will be called "the physical coordinate system". The pointer and microscope 3D locations were tracked by a 3D optical localizer and defined in the physical coordinate system. Computation of geometrical transformation between the physical and image coordinate systems $T_{image->phys}$ (see Fig. 1), was based on anatomical landmarks, attached fiducial markers or point clouds scanning the skin's surface.

C. Surface mesh reconstruction

1) Acquisition: A 3D compact video camera (Carl Zeiss 3D Compact Video Camera S2) was installed between the NC4 surgical microscope and the binocular tube. It consisted of two CCD compact cameras. Its control unit contained the two camera controls for processing the video signals from the left and right videos cameras, and it had a video out socket connected to an external computer. Acquisition consisted of sending, via an Ethernet connexion to a computer, a pair of left and right images, the corresponding microscope location in the physical coordinate system and the magnification and focus settings. Only minimum magnification with minimal and maximum focus levels was used.

2) Calibration: The first stage of the reconstruction process was to calibrate the camera system once, off-line. We used a pinhole camera model for both cameras. A flat calibration grid, printed on a high precision laser printer, was mounted on a flat surface. White squares on black background were chosen as a pattern. About 20 acquisitions (images il and ir, and microscope locations $T_{mic->phys}^{il}$ and $T_{mic->phys}^{ir}$) were made per focus and magnification setting pair. For each pair, points of interest, which matched in the left image and right image, were automatically extracted using curvature operators [21]. A dozen of these points were manually matched with calibration grid points. The homography of one image to the pattern image $T_{i2D->p2D}$, and the transformation $T_{p2D->phys}$ between the pattern image and the pattern in the physical space were then computed. The locations of these points in the physical coordinate system were known thanks to the localizer pointer. As the microscope location $T_{mic->phys}^{i0}$ was known for each image $i0$, the geometrical transformation between the calibration grid and the camera in the microscope ($T_{p2D->mic}$) could be computed. The calibration process was iterative and was initialized by a Tsai calibration procedure [22]. The extrinsic and intrinsic parameters of each camera were computed by maximizing the likelihood of the pinhole camera model with the 20 acquisitions for each magnification and focus setting pair.

The fundamental matrix is the mapping between the pixel coordinates of each image and the epipolar constraint of the stereo system. It was computed with the cameras parameters, once, off-line. Indeed, in a stereo system, the matching points lie in conjugated epipolar lines. The rectification matrix, R_l for the left images and R_r for the right images, is the image transformation that makes conjugated epipolar lines collinear and parallel to the horizontal image axis. Both were thus

computed once (and off-line) from the fundamental matrix and from the points matched in the left and right images [21].

3) Reconstruction: During surgery, each new image pair (left image I_l^0 and right I_r^0) was rectified.

$$
\begin{aligned}
I_l^R &= R_l \times I_l^0 \\
I_r^R &= R_r \times I_r^0
\end{aligned}
$$

The matching point of a pixel of I_l^R was searched on the line with the same ordinate in I_r^R. The difference of abscissae in pixel between two matched points is called the disparity d. Points were matched by maximizing the similarity between two points, measured by ZNSSD (*Zero-mean Normalized Sum of Squared Differences*) [21]:

$$
\mathrm{ZNSSD}_{x,y}(d, I_l^R) = -\frac{\sum_{i,j}(I_l'(x+i,y+j)-I_r'(x+i+d,y+j))^2}{\sqrt{\sum_{i,j}\overline{I_r^R}(x+i+d,y+j)^2}}
$$

where d denotes the disparity and $I' = I^R(x,y) - \overline{I}(x,y)$, with $I^R(x,y)$ for the intensity of luminance at rectified image point (x,y) and $\overline{I}(x,y)$ for the mean intensity over the correlation window of size 11×11. Indexes l and r denote the left and right images respectively. For the sums, we have $-10 \le i \le 10$ and $-10 \le j \le 10$ with $i,j \in N$. Additional robustness for point pairing was achieved by respecting the ordering constraint, which says that the ordering of matching points remains unchanged on conjugated epipolar lines. The search interval for the disparity was determined by the optical geometry of the microscope and cameras, and the location of the surface in relation to the microscope. Matching points were only validated if the same value of d was found for $\mathrm{ZNSSD}_{x,y}(d, I_l^R)$ and $\mathrm{ZNSSD}_{x,y}(d, I_r^R)$. The disparity map consisted of the disparity value for each pair of validated matched points (see Fig. 2(c)). The disparity map was smoothed by a median filter and mathematical morphology, then by computation of the tangent plane in each pixel [23] (see Fig. 2(d)). Thanks to the parameters of each camera, the rectification matrices R_l and R_r, and $T_{mic->phys}^{I_l^0}$ and $T_{mic->phys}^{I_r^0}$, we computed the stereoscopic system parameters and obtained the depth of the point from the disparity map, and consequently its 3D coordinates in the physical coordinate system. The computed reconstruction matrix is called $T_{d->phys}$. Using VTK [20], a surface mesh was computed, composed of triangles for which each vertex had been computed from the disparity map and $T_{d->phys}$. Since each vertex has known pixel coordinates (x,y) in the disparity map, the left image was mapped to the surface mesh using the pixel coordinates (x_l, y_l) where $(x_l, y_l) = R_l^{-1} \times (x,y)$. The surface meshes obtained at the end of this stage are called "rough surface meshes" (see Fig. 2(e)).

The rough surface meshes had some holes, due to the points where disparity was not validated, and some outliers. Surface meshes were then decimated and smoothed using Laplacian smoothing. These surface meshes are called "smooth surface meshes" (see Fig. 2(f)).

D. Visualization

The geometrical transformation from the image coordinate system to the physical coordinate system was applied to the 3D surfaces extracted from the preoperative images. The textured smooth surface mesh was superimposed onto the preoperative 3D surfaces in the physical coordinate system, stored in a VTK file and displayed on the computer screen. Merging reality and virtuality did not require any additional registration, since we used the image-to-physical geometrical transformation.

For each vertex of the 3D surfaces from the preoperative images, with coordinates $(P_x, P_y, P_z)^T$ in the image coordinate space, let $(P_x^p, P_y^p, P_z^p)^T$ be its coordinates expressed in the physical coordinate system.

$$
\begin{pmatrix} P_x^p \\ P_y^p \\ P_z^p \\ 1 \end{pmatrix} = T_{image->phys} \times \begin{pmatrix} P_x \\ P_y \\ P_z \\ 1 \end{pmatrix}
$$

For each pixel $(x,y)^T$ of the disparity map, with a disparity value of $d(x,y)$, let $(X, Y, Z)^T$ be the coordinates of the reconstructed vertex expressed in the physical coordinate system.

$$
\begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} = T_{d->phys} \times \begin{pmatrix} x \\ y \\ d(x,y) \\ 1 \end{pmatrix}
$$

where $T_{d->phys}$ is the reconstruction matrix and $T_{image->phys}$ the image-to-physical geometrical transformation.

Since the result was a 3D virtual scene, all 3D surface rendering possibilities were available. For instance, the neurosurgeon was able to toggle on/off each component in the scene, and change its opacity or color. He or she was also able to change the pose of the global scene.

III. PERFORMANCE EVALUATION ON PHANTOM

The objective of performance evaluation was to study the accuracy of the AV approach in clinical settings, i.e., intraoperative visual matching of reconstructed smooth surfaces along with preoperative images of the patient. Evaluating accuracy on rough surfaces meshes was mainly motivated by future quantification applications. Performance evaluation requires a reference. To compare surface meshes with a reference surface, we used a physical skull phantom. Preliminary results were presented in [24] for the rough surface meshes only, and for two reference surfaces. In this paper, we extend the evaluation study with a third reference surface. In addition, the three reference surfaces were compared to both rough and smooth surface meshes. The first reference surface (see Fig. 3(a)), used by the FIDSCAN procedure, was the 3D surface of a physical skull phantom segmented from a CT scan and registered to the physical coordinate system by means of attached fiducial markers. The second reference surface, used by the LOCSCAN procedure, was obtained by simply changing the image-to-physical registration method. The CT scan volume was registered using both fiducials and a surface described by a cloud of 2,000 points acquired with the neuronavigation system pointer. The third reference surface (see Fig. 3(b)), used by the LOCCLOUD procedure, was this point cloud. Reconstructed surface meshes and reference

surfaces were expressed in the same physical coordinate system. The FIDSCAN procedure was used to study the overall performance of the AV method including image-to-physical registration error. The LOCSCAN procedure was used to check the influence of the image-to-physical registration error. Finally, the LOCCLOUD procedure was used to estimate the error due to the calibration and reconstruction methods alone. A standardized framework, as suggested in [25], was applied to describe these three procedures (see Table I). This evaluation data (images and reference surfaces) can be downloaded from *http://idm.univ-rennes1.fr/theme1/AV_validation*.

A. Evaluation data sets

For the three performance evaluation procedures, the same physical bone skull phantom (referred to as D_I in Table I) was used.

B. Input parameters

The influence of two input parameters on method performance was studied, i.e., the FOV type and microscope focus (referred to as P_I in Table I). The studied FOV types were mat or bright. The bright FOV was obtained by covering the skull phantom with gel. Two values for microscope focus were used: minimum focus (f_{min}) and maximum focus (f_{max}). These two focus values corresponded to the two extrema values of possible microscope focus settings.

C. Reference 1

For the first procedure, FIDSCAN, the skull phantom had a high resolution CT scan ($512 \times 512 \times 425$ with an isotropic voxel size of 0.48 millimeters), with 8 attached fiducial markers. The locations of these markers were identified in the CT scan images and localised in the physical coordinate system with the pointer. The rigid geometrical transformation computed between image and physical space was subsequently applied to the CT scan volume. To obtain an estimate of registration quality, the *Target Registration Error* (TRE) [26] was computed over the surface of the skull phantom. The surface of the registered CT scan volume was then extracted using a *Marching Cubes* algorithm. Since the skull had thickness, the extracted surface was made up of an inner and an outer surface. To be comparable with the reconstructed surface meshes, Reference 1 (referred to as $\hat{S}_{SNR}$ in Table I) was defined as the o uter surface of the skull phantom.

D. Reference 2

The second procedure, LOCSCAN, was similar to the first one. Only the image-to-physical registration was different. For this procedure, we acquired a cloud of 2,000 points by scanning the physical skull phantom surface with the pointer. The CT skull surface was then registered using both fiducial markers and this point cloud, using the *Iterative Closest Point* (ICP) method.

E. Reference 3

Reference 3 was the cloud of 2,000 points acquired on the physical skull phantom and expressed in the physical coordinate system. This cloud was triangulated by Delaunay's method.

F. Surface meshes used for evaluation

Each evaluation procedure was performed twice, for the rough surface meshes and smooth surface meshes. A full 4 (2 FOV types $\times$ 2 focus levels) factorial design was used, 75 image pairs were acquired, i.e., 75 surface meshes intended for reconstruction, with a size of 5×3cm (obtained for f_{min}) and 8×6cm (obtained for f_{max}). These 300 image pairs were acquired from different viewpoints, moving the microscope to cover the whole facial part of the skull. For the FIDSCAN and LOCSCAN procedures, the 300 reconstructed surface meshes (referred to as S_M in Table I) were studied, while in LOCCLOUD only 178 meshes which had a covering rate of 100% with $\hat{S}_{NR}$ (see Table I) were manually selected. All acquisitions were performed with a minimum magnification level.

G. Evaluation metric

For the three FIDSCAN, LOCSCAN and LOCCLOUD procedures, the evaluation metric was the distance in millimetres from each vertex of a mesh (more than 30,000 vertices per mesh) to the nearest $\hat{S}_{NR}$ surface point (see Table I). This was achieved using the ICP algorithm.

H. Discrepancy

For each reconstructed surface mesh, ICP was performed with Reference 1, 2 or 3 as the target. As ICP results for one mesh, we kept the median (referred to as i.median), standard deviation (referred to as i.stddev) and maximum value (written as i.max). We defined the success ratio S_R as the percentage of points reconstructed at a correct location compared with the total number of points, expressed by $S_R = \frac{R_R * Q_R}{100}$ where: R_R stands for the reconstruction ratio (as a percentage), i.e., the number of points which have been reconstructed compared with the total number of points (pixels) studied; and Q_R stands for the quality ratio, i.e., the number of points of the reconstructed surface mesh being reconstructed under a threshold compared with the total number of reconstruction points. The threshold was determined as the worst i.median.

I. Quality indices

Quality indices were defined as follows for the three procedures:
- Precision was given by standard deviation and the maximum i.median, i.max and i.stddev values.
- Accuracy was given by the mean of the different i.median values.
- Robustness was given by the results obtained for the worst parameter combination and the success ratio S_R.
- Input parameter influence was checked by analysis of variance of i.mean and R_R.

TABLE I

DESCRIPTION OF THE THREE PERFORMANCE EVALUATION PROCEDURES

	FIDSCAN	LOCSCAN	LOCCLOUD
D_I: Evaluation data sets	Physical skull	Physical skull	Physical skull
P_I: Input parameters	FOV type, focus	FOV type, focus	FOV type, focus
F_M: Method to be evaluated	AV with surface meshes	AV with surface meshes	AV with surface meshes
$\hat{S}_M$: Surfaces computed by F_M	292 rough meshes 288 smooth meshes	292 rough meshes 288 smooth meshes	178 rough meshes 152 smooth meshes
F_R: Function which computes reference from D_I & P_I	CT scan with fiducial based registration	CT scan with fiducial and point cloud based registration	Point cloud acquired with pointer
$\hat{S}_R$: Output data from F_R (i.e., reference)	CT scan volume in physical coordinate system	CT scan volume in physical coordinate system	2,000 points in physical coordinate system
$\hat{E}_R$: Estimated error in relation to the computation of $\hat{S}_R$ by F_R	TRE	$<$TRE	Pointer precision
F_{NSR}: Function which transforms $\hat{S}_R$ for comparison	Marching Cube + inner surface removal	Marching Cube + inner surface removal	Delaunay triangulation
$\hat{S}_{NR}$: Normalized results for comparison	Outer surface of F_R	Outer surface of F_R	Front surface of D_I
F_C: Evaluation metric to compare $\hat{S}_M$ and $\hat{S}_{NR}$	ICP point-to-surface distance	ICP point-to-surface distance	ICP point-to-surface distance
O_C: Discrepancy between $\hat{S}_M$ and $\hat{S}_{NR}$ computed by F_C	i.median[a][b], i.max[a][b] i.stddev[a][b], S_R	i.median[a][b], i.max[a][b] i.stddev[a][b], S_R	i.median[a][b], i.max[a][b] i.stddev[a][b], S_R
O_{QI}: Quality index from O_C	Mean[a], stddev[a], max[a],min[a]	Mean[a], stddev[a], max[a],min[a]	Mean[a], stddev[a], max[a],min[a]
F_H: Statistical tests	Wilcoxon test, ANOVA	Wilcoxon test, ANOVA	Wilcoxon test, ANOVA

[a] in millimetres.
[b] i.* stands for results of ICP.

IV. RESULTS

A. Performance evaluation results

1) Input parameter influence: To study input parameter influence, rough reconstructed surface meshes were grouped by combining parameters into 4 samples. The distribution of i.median (median values) is shown in Fig. 4 for the three procedures (FIDSCAN, LOCSCAN and LOCCLOUD). Parameter influence was not related to the procedure. Analysis of variance (ANOVA) of the i.mean (mean distance) was performed on the samples grouped by focus. Residuals were Gaussian (p-value $< 10^{-16}$). Consequently, the high impact of focus on accuracy was assumed. These results were similar when studying the smooth surface meshes. For 4 rough surface meshes with a maximum level focus and a very low reconstruction ratio, the smooth surfaces were considered as failures. The same ANOVA was performed on the f_{min} samples to test the impact of the FOV type. These parameters had no significant impact on accuracy (p-value > 0.5). The reconstruction ratio R_R was also studied: only the FOV type had an impact on R_R.

2) Results of the FIDSCAN procedure: We have studied the 300 rough surface meshes and the 296 smooth surface meshes. The i.median values, computed for over 30,000 vertices for each of the 150 reconstructed rough surface meshes with f_{min}, had a mean of 2.29 ± 0.59mm with a Q75 of 2.65mm and a maximum value of 4.32mm. For f_{min} and the 146 smooth surface meshes, the i.median values were a little higher, with a mean of 2.56 ± 0.57 and a maximum value of 4.61mm. When both focus levels were considered, results for the mean of i.median were 3.47 ± 1.36mm and 3.53 ± 1.17 for rough and smooth data respectively. The mean TRE was 3mm on the front part of the skull, where stereo images were acquired. A Wilcoxon non-parametric test showed that the Q95 of i.RMS, for all focus values, can be considered as below the TRE for both rough and smooth surface meshes. For all focus values and rough surface meshes, the success ratio S_R had a mean of 62.15% with a threshold of 5mm. The minimum S_R was 24.38%, however 219 out of 300 reconstructed surface meshes had S_R higher than 55%. The best result of 84.15% was obtained for a bright FOV type and a minimum focus value.

3) Results of the LOCSCAN procedure: For each of the 150 reconstructed rough surface meshes with f_{min}, the i.median values, computed for over 30,000 vertices, had a mean value of 1.09 ± 0.31mm with a Q75 of 1.26mm and a maximum value of 2.17mm. For the 146 smooth surface meshes with f_{min}, the values were a little higher with 1.8 ± 0.45mm. For all focus values, the maximum distance was 6.73mm for the rough meshes and 5.65mm for the smooth meshes.

4) Results of the LOCCLOUD procedure: For all 111 reconstructed surface meshes with f_{min}, the mean i.median value was 0.76 ± 0.27mm for the rough surface meshes and 1.07 ± 0.47mm for the smooth surface meshes. For minimum focus values, a Wilcoxon test showed that the median can be expected to be within 1mm (p-value $< 10^{-11}$) for rough surface meshes and within 1.5mm (p-value $< 10^{-16}$) for smooth surface meshes.

B. Clinical cases

The method has been performed on 6 patients who presented 3 brain cavernomas and 3 low-grade brain tumors. The cavernomas were located in the right frontal lobe, the left medial temporal lobe and the right central region. The

low-grade tumors were located, for two patients, in the left frontal lobe and, for one patient, in the right frontal lobe. Here, we describe the use of MIGN with AV for the resection of a left frontal low-grade glioma on a 31-year-old male patient. The patient had preoperative three-dimensional magnetic resonance imaging (MRI) and fMRI of language and motor functions. Three-dimensional surfaces in VTK file format were automatically or semi-automatically extracted from the MR examination: cortex, skin, target lesion and cortical sulci. The neurosurgeon selected superior and inferior frontal sulci, and the mean and superior part of the pre-central sulcus. fMRI was registered to 3D MRI, and showed that the linguistic activations were mostly located in the posterior and inferior part of the lesion, and even inside the lesion. Motor activations were located behind the precentral sulcus, at about 10 millimeters from the posterior part of the lesion. This functional information was also extracted and stored in VTK file format. A whole 3D scene was created with transparency and colour being chosen for each image entity (see Fig. 5). During the planning stage, the neurosurgeon was able to interact with the 3D scene and the original images. During surgery, the image-to-physical geometrical transformation ($T_{image->phys}$) was computed using anatomical landmarks and a cloud of frontal points, and applied to the preoperative 3D scene. Surgery was performed with the patient in the dorsal decubitus and semisitting position for the procedure. A reconstruction process from stereoscopic images took about one minute. An initial surface mesh of the patient's skin was reconstructed. This reconstruction enabled the neurosurgeon to check the results of the image-to-physical registration. After the scalp incision and elevation, a second reconstruction was requested by the surgeon and was used to understand the spatial relationship between the operative FOV and the preoperative scene. After opening the bone, stereoscopic image acquisition of the dura-mater was performed. During dura-mater opening, care was taken to preserve the arachnoid plane to prevent brainshift. A surface mesh of the arachnoid was reconstructed (see Fig. 6). The neurosurgeon located white tissues surrounding the cerebral convolutions. He checked the anatomical and functional information by navigating in the AV scene. When the arachnoid was opened, the cortical surface was deformed. Another surface mesh of the cortical surface was reconstructed and highlighted the brainshift. The AV was used to qualitatively evaluate the anatomical deformations. At the end of the surgery, a new surface mesh was reconstructed. Surgical tumor removal, made behind the precentral sulcus, was tangent to the inferior and superior frontal sulci. The AV scene again confirmed the importance of the anatomical deformations.

V. DISCUSSION AND CONCLUSION

We developed an Augmented Virtuality approach which involves merging textured surface meshes of the operative FOV with a 3D preoperative virtual scene. Performance evaluation of this approach and the first clinical cases have proved its feasibility and relevance.

A. Performance evaluation

This AV application includes many software components, each one being a potential source of errors (e.g., 3D localisation, image-to-physical registration, fiducial designation, reconstruction, calibration). Three evaluation procedures were used to differentiate between global system error and intrinsic error related to the surface mesh reconstruction component, including calibration. Global process accuracy was given by the FIDSCAN and LOCSCAN procedures. For both procedures, this accuracy was within the image-to-physical registration error, with the best results for LOCSCAN, where this registration (using a combination of fiducials and surface) has been shown in [27] to have an error lower than the fiducial-based registration method used in FIDSCAN. We have thus shown that the global error mainly depended on image-to-physical registration accuracy. The accuracy of a subpart of the system, including reconstruction and calibration, was computed by LOCCLOUD and was shown to be within 1.5mm for the smooth surface meshes. For results with only the minimum focus value, precision, given by the standard deviation, was below 0.6mm, for all evaluation procedures and for both rough and smooth surface meshes. The robustness of our method has been demonstrated. Correct results were obtained for all 300 rough surface meshes and for 98.6% of the smooth surface meshes. The failures observed on the smooth surface meshes can be explained by the low reconstruction ratio of the corresponding rough surface meshes. We have shown the impact of focus level on the method. It can be explained by the fact that, for a same magnification value, a maximum focus requires the microscope to be a long way from the operative FOV, while a minimum focus level enables higher image spatial resolution.

However, some limitations in the performance evaluation require emphasis. Firstly, with our former neuronavigation system, we did not have access to microscope setting numerical values. Consequently, calibration was only processed for the extreme values. Ideally, the calibration must be processed for all focus and magnification settings. Interpolation can also be used for focus and magnification values which were not calibrated [6].

Secondly, our AV method was tested on 6 clinical cases, but only the reconstruction ratio R_R was computed. Results of R_R with skin, bone, or cortex images were similar to those computed on the phantom. For one clinical case only, described in section IV-B, the reconstructed surface mesh of the skin with a maximum focus value was compared with the skin surface segmented from preoperative MRI and rigidly registered to the physical space. The results obtained were promising: 3.05 ± 1.54mm with a maximum distance of 6mm. Finally, the performance evaluation is limited by the fact that the ICP evaluation metric might over-evaluate performance as ICP does not give a symmetrical distance. However, using a statistical estimator like the median for all the point-to-point distances of the same surface mesh moderates this underestimation of the real distance.

B. Comparison with other works

In our method, local deformations are respected by our rough and smooth surface meshes, whereas in [17] the reconstructed 3D surfaces were fitted to a spherical model. Moreover, local zones which are too dark or which have too high a specularity do not interfere with the rest of the surface reconstruction, unlike [16]. Besides, while in [17] accuracy has been studied with 5 points of one reconstructed surface mesh, our performance evaluation related to more than 30,000 points for each of the 178 to 300 rough surface meshes and to more than 5,000 points for each of the 152 to 296 smooth surface meshes.

Our work involves combining the real-world scene with preoperative virtual images. Two approaches are available: AR and AV. One of the selling points of AR is that direct over-laying of preoperative images in the microscope ocular means that the neurosurgeon need not look up at the neuronavigation workstation, and can consequently focus his or her attention on the surgical FOV. However, during surgery, the neurosurgeon needs to raise his or her head from the microscope, to think about his or her surgical gestures or rest at critical moments. At these moments, the AV approach seems to us to be a great advantage. Furthermore, despite AR display constraints, AV allows the surgeon to display the 3D scene according to any viewing direction. Concerning related AV approaches, most of them only use 2D intraoperative information, as with endoscopic cameras [13], [14]. Here we propose more than 2D information, but less than a volume, by providing the neurosurgeon with the 3D surface of the intraoperative FOV.

C. Advantages and limitations of our AV method

Our system is independent of the neuronavigation software. It was successively used on an SMN system from ZEISS, STN_{SNN} from SNS and it is currently being implemented on the StealthStation from Medtronic SNT. Our method does not clutter up the operative room with additional and expensive hardware: the VR scene can be computed and displayed on the workstation of any neuronavigation system and the 3D compact video cameras can simply be plugged into the microscope. Since our method worked with a neuronavigation system, another characteristic of our approach is that the intraoperative surface meshes showing the operative field of view were directly expressed in the physical coordinate system. Therefore they can be directly merged with the preoperative scene registered to the physical space.
After opening the arachnoid, deformations of the cortical surface are significant [28], and the preoperative information no longer corresponds to the anatomical reality of the patient. Consequently, the 3D surfaces from the preoperative images, only registered with image-to-physical transformation $T_{image->phys}$, are shifted compared with the surface mesh which respected the actual anatomy of the patient. Our method is therefore less advantageous when surgical resection actually begins. However, the rough surface meshes seem to be accurate enough to be used for quantification of cortical surface deformations. This quantification might be used to update preoperative information, located near the cortical surface only. Indeed, further deep structures are subject to deformations which are difficult to predict from surface deformations alone [29].

To conclude, this AV system facilitates the intraoperative understanding of the spatial relationship between the operative field and the complex 3D multimodal scene, which includes preoperative images of the patient. Other information can be added to this 3D scene (e.g., new image modalities, intraoperative images or information from anatomical digital atlases registered with the patient). With this system, neurosurgeons and clinical staff in the OR are able to freely interact with the resulting 3D scene in the OR. In addition, this display approach presents a real clinical advantage as it reduces surgeon stress, and a pedagogical advantage for clinical staff present during surgery.

REFERENCES

[1] P. Jannin, X. Morandi, O. Fleig, E. Le Rumeur, P. Toulouse, B. Gibaud, and J.-M. Scarabin, "Integration of sulcal and functional information for multimodal neuronavigation," *J Neurosurg*, no. 96, pp. 713–723, 2002.

[2] P. Jannin, O. Fleig, E. Seigneuret, C. Grova, X. Morandi, and J.-M. Scarabin, "A data fusion environment for multimodal and multi-informational neuro-navigation," *Computer Aided Surgery*, vol. 5, no. 1, pp. 1–10, 2000.

[3] O. Ganslandt, R. Fahlbusch, C. Nimsky, H. Kober, M. Möller, R. Stein-meier, J. Romstöck, and J. Vieth, "Functional neuronavigation with magnetoencephalography: outcome in 50 patients with lesions around the motor cortex," *J Neurosurg*, no. 91, pp. 73–79, 1999.

[4] M. Schulder, J. Maldjian, W. Liu, A. Holodny, A. Kalnin, I. Mun, and P. Carmel, "Functional image-guided surgery of intracranial tumors in or near the sensorimotor cortex," *J Neurosurg*, no. 89, pp. 412–418, 1998.

[5] J. Luber and A. Mackevics, "Multiple co-ordinate manipulator (mkm): A computer-assisted microscope," in *Computed Assisted Radiology*, H. L. ed., Ed. Berlin: Springer-Verlag, 1995, pp. 1121–1125.

[6] P. Edwards, A. King, C. J. Maurer, D. de Cunha, D. Hawkes, D. Hill, R. Gaston, M. Fenlon, A. Jusczyzck, A. Strong, C. Chandler, and M. Gleeson, "Design and evaluation of a system for microscope-assisted guided interventions (MAGI)," *IEEE Trans Med Imaging*, vol. 19, no. 11, pp. 1082–1093, 2000.

[7] C. Maurer, F. Sauer, B. Bascle, B. Geiger, F. Wenzel, F. Reechi, T. Rohlfing, C. Brown, R. Bakos, R. Maciunas, and A. Bani-Hashemi, "Augmented reality visualization of brain structures with stereo and kinetic depth cues: System description and initial evaluation with head phantom," *Medical Imaging 2001: Visualization, Display, and Image-Guided Procedures*, vol. Proc SPIE 4319, pp. 445–456, 2001.

[8] W. Grimson, G. Ettinger, S. White, T. Lozano-Pérez, W. Wells, and R. Kikinis, "An automatic registration method for frameless stereotaxy, image guided surgery, and enhanced reality visualization," *IEEE Trans Med Imaging*, vol. 15, no. 2, pp. 129–140, 1996.

[9] A. Colchester, J. Zhao, K. Holton-Tainter, C. Henri, N. Maitland, P. Roberts, C. Harris, and R. Evans, "Development and preliminary evaluation of vislan, a surgical planning and guidance system using intra-operative video imaging," *Med Image Anal.*, vol. 1, no. 1, pp. 73–90, 1996.

[10] M. Uenohara and T. Kanade, "Vision-based object registration for real-time image overlay," *Comput Biol Med.*, vol. 25, no. 2, pp. 249–260, 1995.

[11] P. Jannin, A. Bouliou, E. Journet, and J.-M. Scarabin, "A ray-traced texture mapping for enhanced virtuality in image-guided neurosurgery," *Stud Health Technol Inform*, vol. 29, pp. 553–563, 1996.

[12] P. Jannin, A. Bouliou, J.-M. Scarabin, C. Barillot, and J. Luber, "Visual matching between real and virtual images in image guided neurosurgery," in *SPIE Medical Imaging*, vol. 3031, 1997, pp. 518–526.

[13] D. Dey, D. G. Gobbi, P. J. Slomka, K. J. Surry, and T. M. Peters, "Automatic fusion of freehand endoscopic brain images to three-dimensional surfaces: creating stereoscopic panoramas," *IEEE Trans Med Imaging.*, vol. 21, no. 1, pp. 23–30, 2002.

[14] R. Shahidi, M. Bax, C. J. Maurer, J. Johnson, E. Wilkinson, B. Wang, J. West, M. Citardi, K. Manwaring, and R. Khadem, "Implementation, calibration and accuracy testing of an image-enhanced endoscopy system," *IEEE Trans Med Imaging.*, vol. 21, no. 12, pp. 1424–1435, 2002.

[15] O. Fleig, F. Devernay, J.-M. Scarabin, and P. Jannin, "Surface reconstruction of the surgical field from stereoscopic microscope views in neurosurgery," in *Proc. CARS'2001*, 2001, pp. 259–264.

[16] O. Skrinjar, H. Tagare, and J. Duncan, "Surface growing from stereo images," in *IEEE CVPR'00*, vol. II, 2000, pp. 571–576.

[17] H. Sun, H. Farid, K. Rick, A. Hartov, D. Roberts, and K. Paulsen, "Estimating cortical surface motion using stereopsis for brain deformation models," in *Proc. Medical Image Computing and Computer-Assisted Intervention MICCAI'03*, 2003, pp. 794–801, LNCS 2878.

[18] M. Audette, K. Siddiqi, and T. Peters, "Level-set surface segmentation and fast cortical range image tracking for computing intrasurgical deformations," in *Proc. Medical Image Computing and Computer-Assisted Intervention MICCAI'99*, 1999, pp. 788–797, LNCS 1679.

[19] M. Miga, T. Sinha, D. Cash, R. Galloway, and R. Weil, "Cortical surface registration for image-guided neurosurgery using laser-range scanning," *IEEE Trans Med Imaging*, no. 22, pp. 973–985, 2003.

[20] W. Schroeder, K. Martin, and B. Lorensen, *The Visualization Toolkit An Object-Oriented Approach To 3D Graphics, 3rd Edition.* Kitware, Inc. publishers, 1993, http://www.vtk.org.

[21] F. Devernay, "Vision stéréoscopique et propriétés différentielles des surfaces," Ph.D. dissertation, École Polytechnique, 1997.

[22] R. Y. Tsai, "A versatile camera calibration technique for high-accuracy 3d machine vision metrology using off-the-shell tv cameras and lenses," *IEEE Journal of Robotics and Automation*, vol. 3, no. 4, pp. 323–344, 1987.

[23] F. Devernay and O. Faugeras, "Computing differential properties of 3-d shapes from stereoscopic images without 3-d models," in *Proceedings of the International Conference on Computer Vision and Pattern Recognition*, 1994, pp. 208–213.

[24] P. Paul, O. Fleig, S. Tranchant, and P. Jannin, "Performance evaluation of a stereoscopic based 3d surface localiser for image-guided neurosurgery," in *MICCAI 04*, 2004, in press.

[25] P. Jannin, M. Fitzpatrick, D. Hawkes, X. Pennec, R. Shahidi, and M. Vannier, "Validation of medical image processing in image-guided therapy," *IEEE Trans Med Imaging*, vol. 21, no. 11, pp. 1445–1449, 2002.

[26] J. Fitzpatrick, J. West, and C. Maurer, "Predicting error in rigid-body, point-based registration," *IEEE Trans Med Imaging*, vol. 17, no. 5, pp. 694–702, 1998.

[27] C. Maurer, R. Maciunas, and J. Fitzpatrick, "Registration of head CT images to physical space using a weighted combination of points and surfaces," *IEEE Trans Med Imaging*, vol. 17, no. 5, pp. 753–761, 1998.

[28] C. Maurer, D. Hill, A. Martin, H. Liu, M. McCue, D. Rueckert, D. Lloret, W. Hall, R. Maxwell, D. Hawkes, and C. Truwit, "Investigation of intraoperative brain deformation using a 1.5-t interventional mr system: Preliminary results," *IEEE Trans Med Imaging*, vol. 17, no. 5, pp. 817–825, 1998.

[29] T. Hartkens, D. Hill, A. Castellano-Smith, D. Hawkes, C. Maurer, A. Martin, W. Hall, H. Liu, and C. Truwit, "Measurement and analysis of brain deformation during neurosurgery," *IEEE Trans Med Imaging*, vol. 22, no. 1, pp. 88–92, 2003.
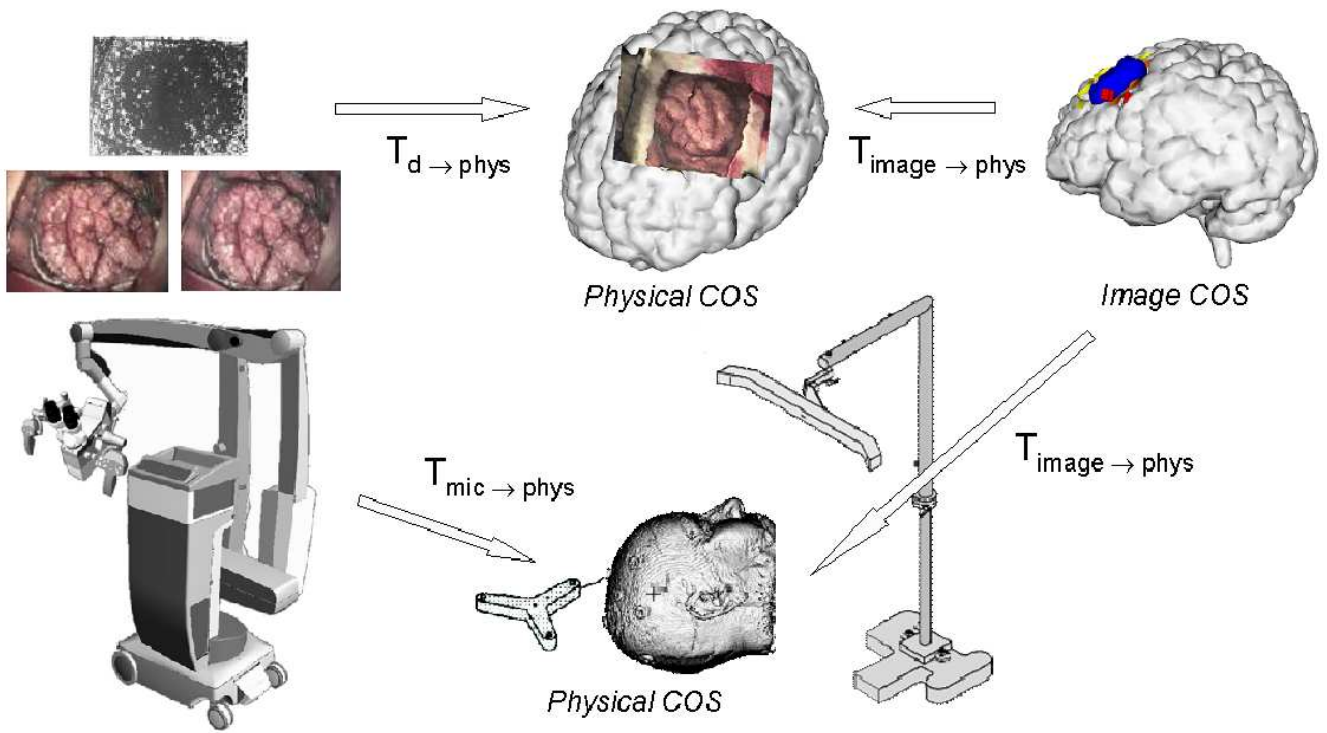
Fig. 1.   Principle of our AV method, with the principal geometrical transformations, COS stands for CoOrdinate System.

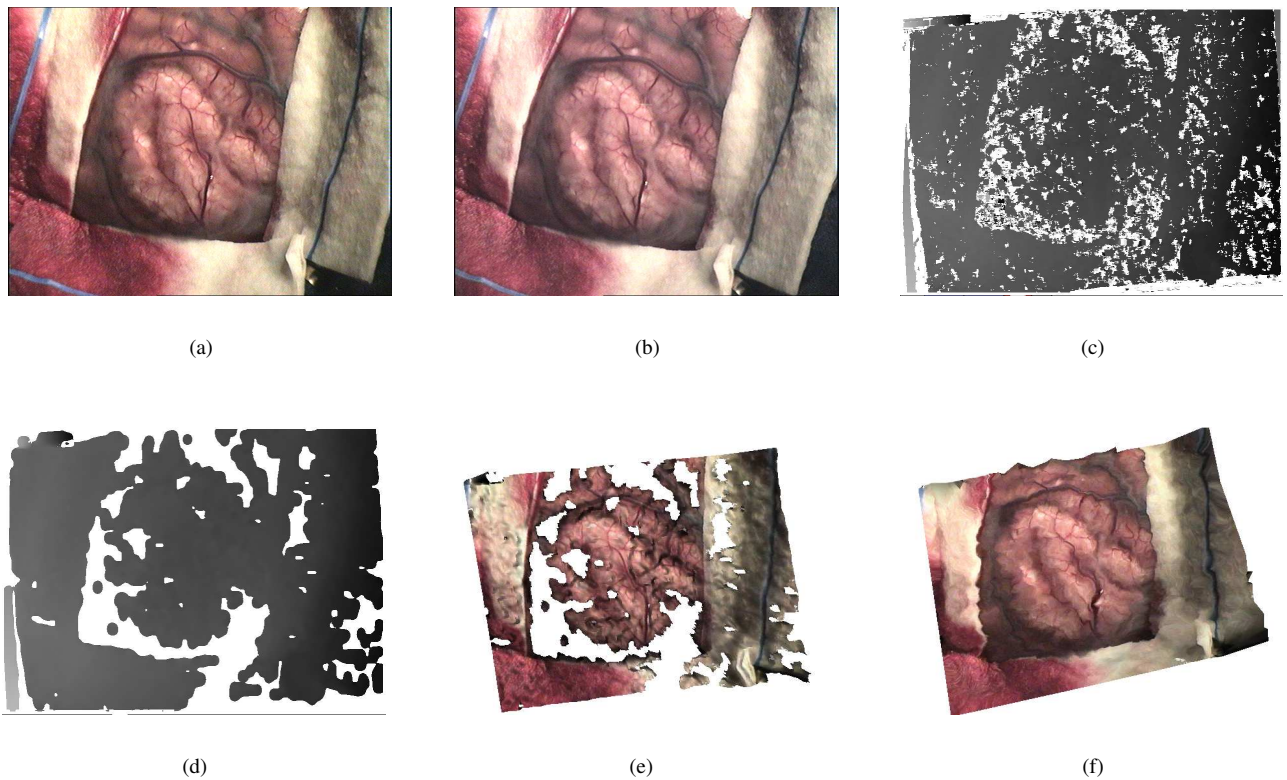
(a)

(b)

(c)

(d)

(e)

(f)

Fig. 2.   3D surface reconstruction of the operative FOV. (a) Left image. (b) Right image. (c) Corresponding disparity map before smoothing. (d) Corresponding smoothed disparity map. (e) Corresponding rough surface mesh. (f) Corresponding smooth surface mesh.

(a)

(b)

Fig. 3.   Surface references for performance evaluation. (a) References 1 and 2, i.e., 3D surface segmented from the CT scan. (b) Reference 3, i.e., 3D surface triangulated from the point cloud
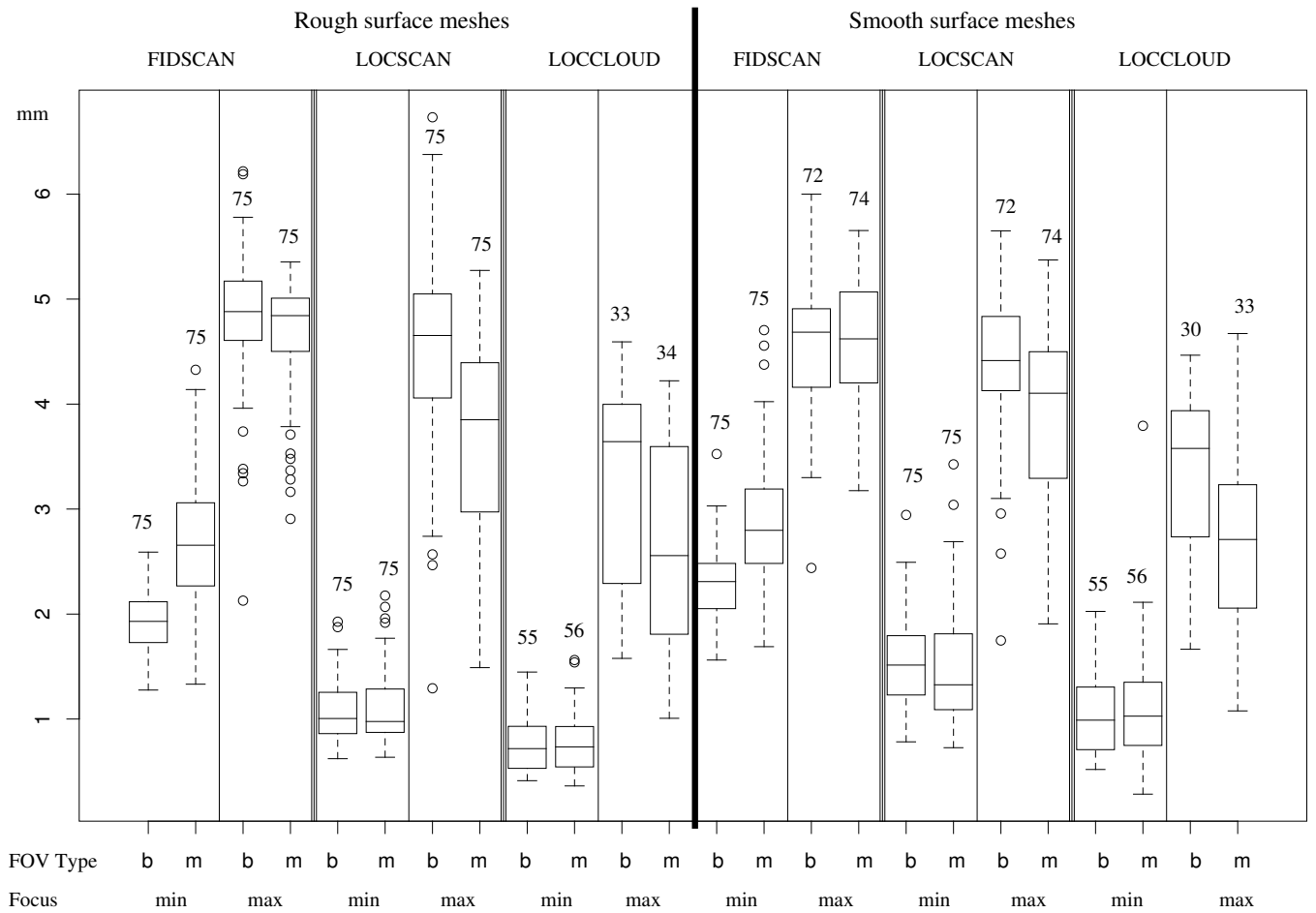

Fig. 4.   Distribution of the median of the distance between reconstructed surface meshes and the reference surface for each sample corresponding to a parameters combination, and for the three evaluation procedures. The number of data in each sample is indicated by the number above its box-plot, b and m stand for bright and mat respectively
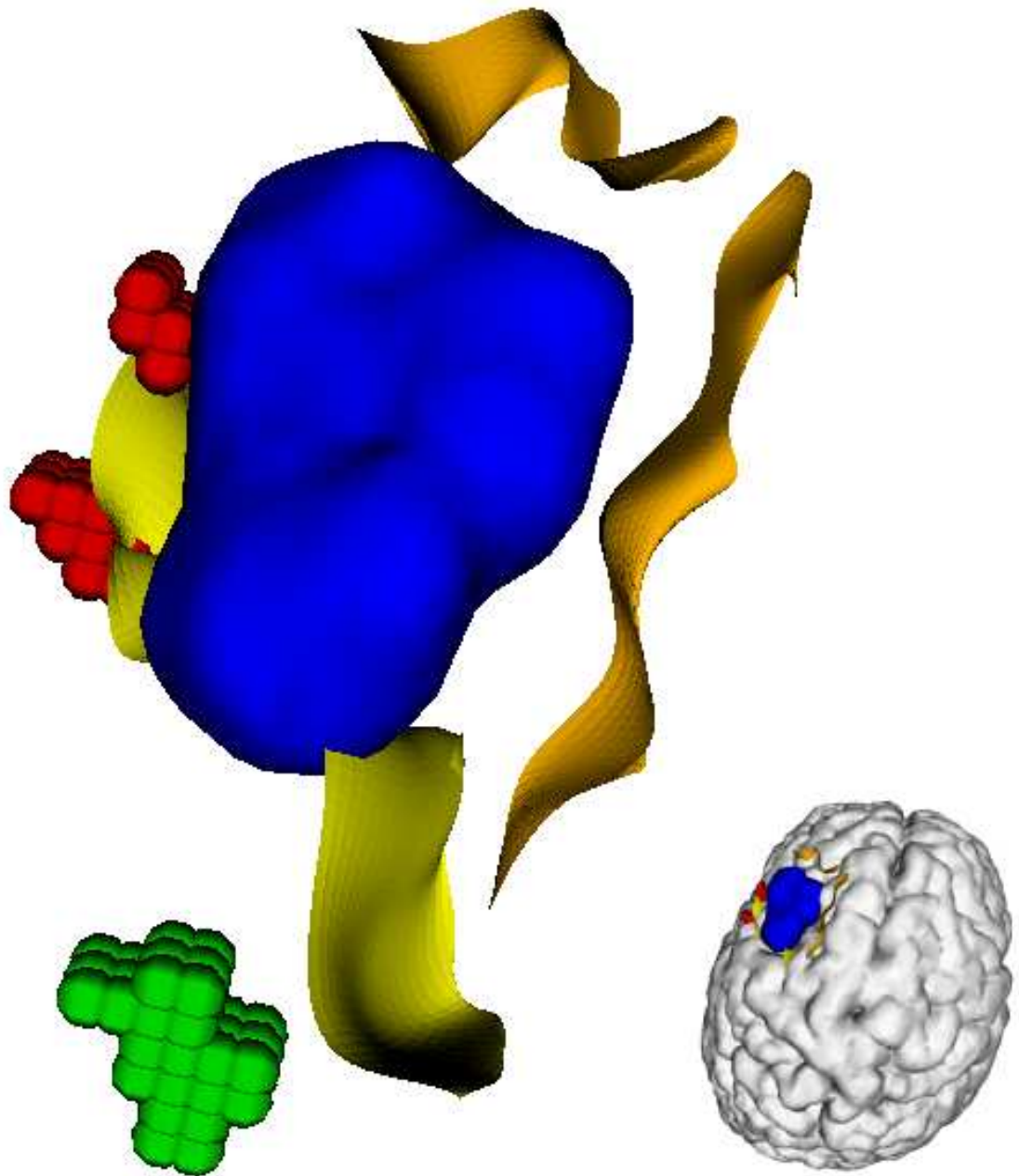
Fig. 5.   Multimodal 3D scene from preoperative images, including the glioma, precentral and frontal sulci, linguistic and motor areas, with global orientation of the scene in the lower right corner of the image.

(a)

(b)

Fig. 6. Two viewpoints of the AV scene including the surface mesh of the arachnoid merged with the multimodal 3D scene.