

Automatic Fusion of Freehand Endoscopic Brain Images to Three-Dimensional Surfaces: Creating Stereoscopic Panoramas

Damini Dey, David G. Gobbi, Piotr J. Slomka, *Member, IEEE*, Kathleen J. M. Surry, and Terence M. Peters*, *Senior Member, IEEE*

Abstract—A major limitation of the use of endoscopes in minimally invasive surgery is the lack of relative context between the endoscope and its surroundings. The purpose of this work was to fuse images obtained from a tracked endoscope to surfaces derived from three-dimensional (3-D) preoperative magnetic resonance or computed tomography (CT) data, for assistance in surgical planning, training and guidance. We extracted polygonal surfaces from preoperative CT images of a standard brain phantom and digitized endoscopic video images from a tracked neuro-endoscope. The optical properties of the endoscope were characterized using a simple calibration procedure. Registration of the phantom (physical space) and CT images (preoperative image space) was accomplished using fiducial markers that could be identified both on the phantom and within the images. The endoscopic images were corrected for radial lens distortion and then mapped onto the extracted surfaces via a two-dimensional 2-D to 3-D mapping algorithm. The optical tracker has an accuracy of about 0.3 mm at its centroid, which allows the endoscope tip to be localized to within 1.0 mm. The mapping operation allows multiple endoscopic images to be “painted” onto the 3-D brain surfaces, as they are acquired, in the correct anatomical position. This allows panoramic and stereoscopic visualization, as well as navigation of the 3-D surface, painted with multiple endoscopic views, from arbitrary perspectives.

Index Terms—endoscopy, image-guided neurosurgery, stereoscopic panorama, texture mapping, 2-D to 3-D registration, virtual endoscopy.

I. INTRODUCTION

IN minimally invasive surgery, endoscopes are introduced through a small entry port into a body cavity from where they provide high-resolution two-dimensional (2-D) real-time video images of a limited view of the surgical site. They are regularly used in arthroscopic, abdominal, and minimally invasive cardiac

surgery, as well as in the ventricles of the brain for neurosurgical applications [1], [2]. Since the field of view (FOV) of the endoscope is small, a major limitation is the lack of relative context between the endoscope and its surroundings [3]. Relating endoscopic images to their surroundings and to preoperative image data, is crucial for surgical procedures; however, because of the limited FOV, it is often a challenging task for the surgeon.

In recent years, virtual endoscopy has emerged as a means of “close-up” examination of complex three-dimensional (3-D) datasets [4]–[8]. A virtual camera is placed within the 3-D volume and images that correspond to “what the endoscope would see” are generated. Virtual endoscopic navigation, combined with a standard display of a set of orthogonal planes, has been utilized for enhancing clinical diagnosis and for planning surgery and treatment [4]–[8].

The integration of video with virtual scenes has been demonstrated in the tele-robotics field [9] and several investigators have attempted to integrate intraoperative endoscopic video with anatomical images. The endoscopic video images are typically displayed alongside the corresponding virtual camera views generated from the preoperative images [4], [10], [11]. Konen *et al.* mapped anatomical landmarks derived from preoperative magnetic resonance (MR) images to live video sequences acquired with a tracked endoscope [12]. Jannin *et al.* merged single microscope images onto surfaces derived from anatomical images for enhanced visualization [13]. Clarkson *et al.* updated the 3-D position of stereo video cameras by registering a virtual view of the surface to a real video image [14]. In a more recent work, Clarkson *et al.* [15] demonstrated improved tracking of stereo video cameras by registering a virtual view of the surface to which the real view has been texture-mapped, to the real video image. However, to our knowledge, for applications in medical imaging, creating stereoscopic panoramas by fusing *multiple* 2-D images of an object acquired by a freehand camera to its 3-D geometrical counterpart, has not been reported before.

We have developed computer graphics-based methods for near real-time integration of video endoscopic images with preoperative 3-D magnetic resonance imaging (MRI) or CT volumes, for assistance in the planning and guidance of surgical procedures. In addition to generating virtual endoscopic views, we demonstrate methods to fuse multiple intraoperative endoscopic views onto irregular polygonal surfaces generated from the preoperative images. This approach enables the acquired endoscopic images to remain permanently mapped onto the surface, even when the endoscope has moved to a

Manuscript received May 18, 2000; revised November 6, 2001. This work was supported by in part the Canadian Institutes for Health Research (CIHR) under Grant MT11540, in part by the Institute for Robotics and Intelligent Systems (IRIS), and in part by the Canadian Foundation for Innovation (CFI). The Associate Editor responsible for coordinating the review of this paper and recommending its publication was D. Hawkes. *Asterisk indicates corresponding author.*

D. Dey, D. G. Gobbi and K. J. M. Surry are with the Imaging Research Laboratories, John P. Robarts Research Institute, London, ON N6A 5K8, Canada.

P. J. Slomka is with the Imaging Research Laboratories, John P. Robarts Research Institute, London, Canada, and also with the Department of Diagnostic Radiology and Nuclear Medicine, London Health Sciences Center, London, ON N6A 5K8, Canada.

*T. M. Peters is with the Imaging Research Laboratories, John P. Robarts Research Institute, University of Western Ontario, 100 Perth Drive, London, ON N6A 5K8, Canada (e-mail: tpeters@irus.ri.ca).

Publisher Item Identifier S 0278-0062(02)01041-8.

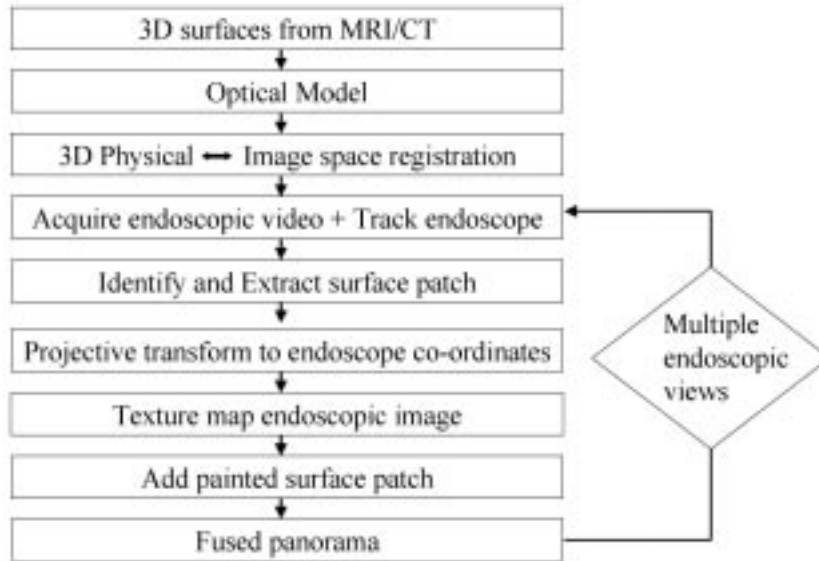


Fig. 1. Flowchart demonstrating our system for automatic fusion of intraoperative endoscopic images to 3-D surfaces.

different location. In addition to enhancing visualization, this provides the necessary 3-D relative context in a complete, visual manner. By simply mapping the 2-D endoscopic images, which are acquired monoscopically, to 3-D surfaces, we can impart stereoscopic depth cues to them, allowing stereoscopic visualization and navigation of the painted surface.

Fusing a real 2-D view to three dimensions is essentially the “inverse” mathematical problem to virtual endoscopy, requiring accurate 2-D to 3-D registration. In addition, creating a fused panorama of composite endoscopic images requires a fast automatic surface-painting algorithm. In our work, the fused panorama is the end result of several key steps: deriving the optical characteristics of the endoscope, tracking the endoscope and, finally, applying the optical characteristics of the tracked endoscope and computer graphics techniques to fuse multiple 2-D endoscopic images to the 3-D surface. We use the characteristics of the virtual camera, which is based on our optical model of the endoscope, to generate a virtual endoscopic view and fuse the intraoperative endoscopic image onto the corresponding part of the surface. Several researchers have worked on these individual components, such as optical calibration of an endoscope [12], or generating virtual endoscopic views from the viewpoint of the endoscope [4]–[8], [12]. However, we believe that this is the first study to integrate all these various aspects, fuse multiple real endoscopic views to irregular 3-D surfaces, and to yield stereoscopic fused panoramas of the surface.

We have previously described an automatic surface-painting algorithm [16], [17] based on ray-traced texture mapping, in preliminary conference proceedings. Our new approach, described here, is simpler, faster, and more robust. It is based on a projective transformation to endoscope coordinates, followed by texture mapping.

II. METHODS

A. Geometrical Framework

In general, a video image is a 2-D projection of the 3-D scene. This process can be represented using the pinhole camera model

[18]. A point \mathbf{x} in three dimensions, defined by the homogeneous coordinates $\mathbf{x} = (x, y, z, 1)$, is related through the 3×4 transformation matrix, \mathbf{T} , to a point $\mathbf{u} = (u, v, 1)$ on the 2-D image, as follows:

$$w\mathbf{u}^T = \mathbf{T}\mathbf{x}^T = \mathbf{P}\mathbf{C}\mathbf{x}^T \quad (1)$$

where the matrix \mathbf{T} can be further decomposed into: a rigid-body transformation, given by the 4×4 matrix \mathbf{C} , from 3-D scene (or world) coordinates to 3-D camera coordinates and a projective transformation onto the 2-D view plane (or imaging plane), given by the 3×4 matrix \mathbf{P} . The variable w represents a scaling factor for homogeneous coordinates. If the camera characteristics that define the projective transformation are known, a virtual camera view can be generated by placing a virtual camera at any position in the 3-D scene. Since the 3-D-to-2-D mapping described in (1) is generally not invertible, the inverse of this mapping is generally not unique [18]. In this paper, we have developed a computer graphics-based approach to solve this problem and merge multiple 2-D images back to 3-D.

B. Overview

In our approach, there are seven principal steps required to map endoscopic images to object surfaces, the first two of which must be performed prior to the intraoperative procedure. These steps (outlined in Fig. 1) are as follows.

- 1) Acquire the anatomical data (CT, MRI) and process it to extract the relevant surfaces.
- 2) Characterize the optics of the endoscope.
- 3) Register the patient (physical space) to the image volume.
- 4) For each acquired frame of the endoscopic video, obtain the position and orientation of the endoscope. This is accomplished by tracking the endoscope with an optical tracking system.
- 5) Using the optical model, identify and extract the surface-patch that is visible to the endoscope.

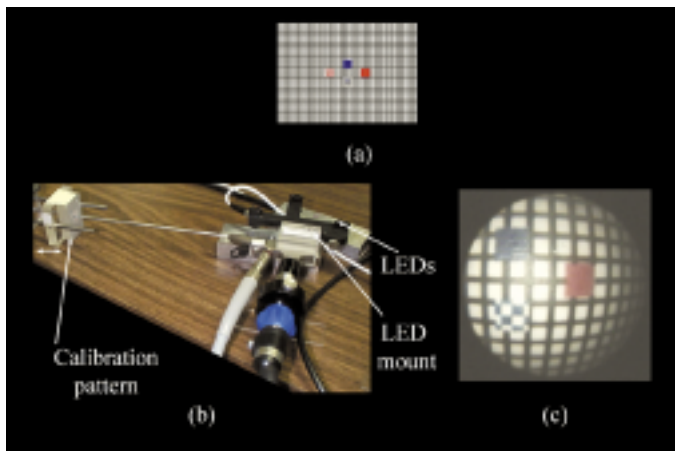


Fig. 2. Optical calibration of endoscope. (a) Our calibration pattern. (b) Calibration setup showing tracked endoscope and pattern. (c) Endoscopic view of calibration pattern.

- 6) Apply a 3-D to 2-D projective transform to endoscope coordinates and generate a virtual endoscopic view.
- 7) Texture map the endoscopic image onto the visible surface patch.
- 8) Add the surface patch and repeat steps 4–8 for the next endoscopic view.

C. Preoperative Imaging and Surface Extraction

To test our methods, we acquired 3-D CT images of a standard brain phantom (Kilgore International Inc., Coldwater, MI). We acquired a 3-D $512 \times 512 \times 250$ (0.49 mm \times 0.49 mm \times 1 mm) volume data set of the phantom with a GE HiSpeed CT scanner (General Electric, Milwaukee, WI). The images were rebinned to 256×256 matrix size. The surface of the phantom was extracted from the CT images using the Marching Cubes algorithm [19] and then decimated using polygon reduction algorithms in the Visualization Toolkit (VTK) [20]. For the surface shown in this paper, the total number of vertices is approximately 33 000.

D. Optical Model

We derived an optical model of the endoscope, similar to the well-known method of Tsai [21]. Our approach consists of imaging a simple calibration pattern [shown in Fig. 2(a)] at several known distances from the endoscope tip [Fig. 2(b)] and determining a model of the endoscopic viewing cone from the resulting images.

We adopted the method described by Haneishi *et al.* [22] to extract a model of the radial (or barrel) distortion of the endoscope lens from the same set of images of the calibration pattern. The intersections of the lines of the grid pattern are identified and the radial distance of each point from the center of distortion are calculated [Fig. 2(c)]. The corrected radial distance is expressed as a fourth-order polynomial function. The coefficients that characterize the distortion can be determined by iteratively optimizing the polynomial coefficients and estimating the distortion center [22].

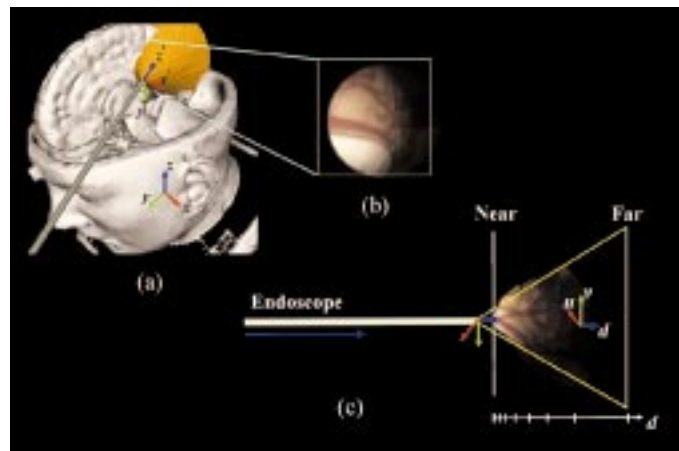


Fig. 3. Three-dimensional geometry. (a) Surface with tracked endoscope. The modeled viewing cone is shown in yellow, (x, y, z) correspond to world coordinates and (x', y', z') correspond to endoscope view coordinates. (b) Acquired endoscopic image. (c) Projective geometry, looking along the endoscope in the direction of projection. The virtual camera is placed at the tip of the endoscope viewing cone. (u, v) correspond to 2-D projected coordinates and d corresponds to the distance from the view plane along the direction of projection.

E. Physical to Image Space Registration

To register the phantom to the preoperative image space, we used five small plexiglass blocks with divots in the center serving as fiducial landmarks. These were identified on the phantom and were also localized on a stereoscopic triplanar (transverse, coronal, and sagittal) view of the CT image volume, with the POLARIS pointer. Physical to image space registration was accomplished by finding the rigid-body transformation that minimized the least-squares residual between the two sets of homologous points [23].

F. Tracking the Endoscope

Endoscopic images were acquired with a 2.7-mm straight AESCULAP neuro-endoscope, (AESCULAP, Tuttlingen, Germany) which was tracked by the POLARIS Optical Tracking System (Northern Digital Inc., Waterloo, ON, Canada). The tracking system consists of a T-shaped assembly of infrared light-emitting diodes (LEDs) and a position-sensor consisting of two charge-coupled device cameras spaced 1 m apart. The LED assembly was attached to the shaft of the endoscope via a plexiglass mount [Fig. 2(b)]. The LEDs are continuously tracked by the POLARIS. We have developed a fast tip localization method, which calculates the position and the six degrees of freedom of the tip from the measured LED position. The endoscopic video images were digitized using a Matrox Corona framegrabber board (Matrox Inc., St Laurent, QC, Canada) in a Pentium II 450-MHz PC. Each endoscopic video frame consisted of a $640 \times 480 \times 24$ -bit color image.

G. Identifying and Extraction of the Surface Patch Matching the Endoscopic View

Fig. 3 shows the 3-D to 2-D geometry. From our tip localization procedure, for each endoscopic view, we know the position and orientation of the endoscope tip in three dimensions.

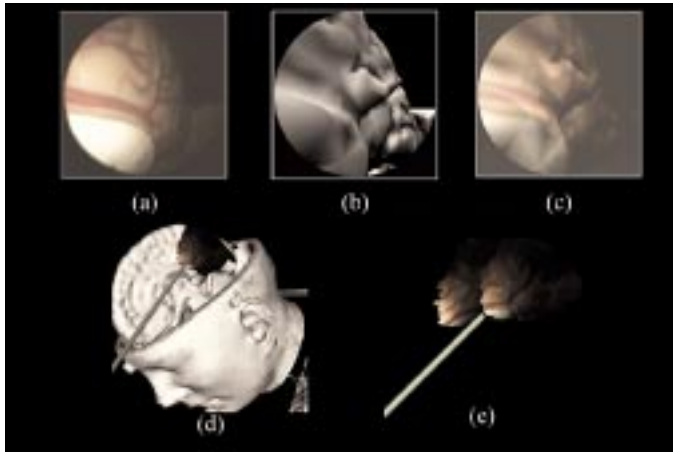


Fig. 4. (a) Acquired endoscopic image for the pose in Fig. 3(a), (b) Virtual endoscopic image corresponding to (a). (c) Endoscopic image shown in (a) mapped to the surface, viewed through the virtual camera. (d) Texture-mapped surface viewed from another perspective. (e) Implementation of multiple endoscopic views with surface patches.

We position the modeled endoscopic viewing cone at the endoscope tip. The surface is clipped with a 3-D mathematical function describing this viewing cone [Fig. 3(a)]. The surface patch corresponding to all the triangles and pieces of triangles inside the cone is extracted. This is the surface patch that corresponds to the endoscopic view.

H. Texture Mapping Algorithm

Texture mapping is a computer graphics technique commonly used to add realism to virtual 3-D scenes [24]. It is typically used to map 2-D images (textures) to surfaces. In recent years, mapping of textures and navigation of surfaces with textures has been accelerated by inexpensive, off-the-shelf graphics hardware developed to support the PC gaming industry [24]. In order to map a texture to a surface, the correct texture coordinates for the vertex of each surface polygon must be computed. While this is quite simple for mapping a texture to a regular surface such as a plane, cylinder, or a sphere, the projection of a texture to an irregular 3-D surface, from an arbitrary 3-D position is not straightforward.

Texture mapping of multiple freehand endoscopic views was implemented via a series of computer graphics operations, as illustrated in Fig. 4.

Texture Coordinate Calculation: In this algorithm, we introduced two special coordinate transformations that simplify all the calculations. We use these transformations to accomplish three tasks: to generate a virtual endoscopic view of the scene, to determine if the polygon is visible to the virtual endoscope, and to determine the texture coordinates for each polygon vertex. We utilize the characteristics of the virtual camera, which is based on our optical model, to perform the necessary calculations.

First, we transform the 3-D world coordinates of each vertex in the surface patch to endoscope view coordinates at the tip of the endoscope viewing cone. For example, if $\mathbf{V} = (x, y, z)$ represents the world coordinates of a vertex of a triangle in the surface patch, then we can define homogeneous coordinates for $\mathbf{V} = (x_v, y_v, z_v, w_v)$ where $x = x_v/w_v, y = y_v/w_v, z = z_v/w_v, w_v = 1$.

Then in the endoscope view coordinate system $\mathbf{V}' = (x'_v, y'_v, z'_v, w'_v)$ can be given by

$$\begin{bmatrix} x'_v \\ y'_v \\ z'_v \\ w'_v \end{bmatrix} = \mathbf{C} \begin{bmatrix} x_v \\ y_v \\ z_v \\ w_v \end{bmatrix} \quad (2)$$

where \mathbf{C} is a 4×4 transformation matrix that corresponds to a transformation from world to view coordinates. In our system, \mathbf{C} is known from the optical model and the physical-image space registration. The value $d = z'_v/w'_v$ corresponds to the distance of the vertex from the endoscope tip along the direction of projection [see Fig. 3(c)]. The points with the subscript v must first be precorrected using Tsai's method [21].

Second, we transform the 3-D view endoscope coordinates given by \mathbf{V}' to the 2-D view plane [Fig. 3(b) and (c)]. This 3-D-to-2-D transformation is described by

$$\begin{bmatrix} wu_v \\ wv_v \\ w \end{bmatrix} = \mathbf{P} \begin{bmatrix} x'_v \\ y'_v \\ z'_v \\ w'_v \end{bmatrix} \quad (3)$$

where \mathbf{P} is a 3×4 transformation matrix defined by the camera characteristics (which are known from the optical model) and (wu_v, wv_v, w) are the projected 2-D coordinates for V . This is equivalent to placing a virtual camera at the tip of the viewing cone of the endoscope. This transform gives us a virtual endoscopic view of the scene [Fig. 4(b)].

While projecting textures to a complex surface, it is necessary to eliminate the polygons that are not visible to the camera. To extract visible polygons, we utilize the depth buffer (or the z buffer) associated with the virtual endoscopic view [25]. For a screen pixel with 2-D coordinates given by (u, v) , the z buffer value corresponds to the distance from the camera view plane to the closest 3-D point along the direction of projection. In our algorithm, the depth value d of each vertex is compared with the corresponding depth buffer value obtained from the display device. If d is greater than the depth buffer value plus a predefined tolerance, then the polygon that contains that vertex is regarded as invisible and is discarded.

Finally, for all the visible polygons, the texture coordinates are generated by scaling the virtual view to the range of the texture [Fig. 4(c)]. The texture coordinates are used to effectively "pin" the texture to the surface. Once the texture coordinates are calculated, the texture is automatically mapped to the surface.

I. Painting Multiple Endoscopic Views

As the endoscope sweeps across the object surface, a surface "patch" visible to the endoscope is extracted, the texture coordinates are computed, and the endoscopic image is texture-mapped to the patch. For each pose of the endoscope, a new texture-mapped surface patch is added to the rendered scene [Fig. 4(e)]. The patch corresponding to the most recent endoscopic pose is added last. For the endoscopic views shown in this paper, the number of vertices for the surface patches varies from 4496 to 7703.

This software is written primarily in C++ and in the Python Programming Language and is interfaced with the graphics

toolkit *The Visualization Toolkit (VTK)* [26]. This toolkit is layered on top of OpenGL graphics libraries and can take advantage of any available graphics hardware-based acceleration.

J. Evaluation of Error

The global accuracy with which the integration of endoscopic images to preoperative 3-D images can be achieved is governed by the accuracy of several of the subcomponents of the system. To ensure that no new errors are introduced with our surface-painting algorithm, we performed the following tests.

- a) From the optical endoscopic calibration, we modeled a virtual camera.
- b) The virtual camera was placed in the 3-D scene, with the surface rendered in a different color.
- c) Several virtual endoscopic views were captured.
- d) The virtual endoscopic views were then mapped back to the 3-D surface.

From these experiments, we evaluated the error in each sub-component of our system, as follows.

1) *Tip Localization Error*: The tracked endoscope was placed at fixed distances from the plane defined by the calibration pattern for seven trials (Fig. 2). The distance from the endoscope tip to the center of the calibration pattern was determined by our tip localization method and was also measured experimentally. For each trial, the tip localization error was defined as the difference between the measured distance and the distance calculated by our tip localization method. From all the trials, the maximum, mean, and standard deviation error values were calculated.

2) *Texture Mapping Error*: The 3-D positions of seven anatomical landmarks were first measured on the preoperative images by localizing each landmark on a stereoscopic triplanar (transverse, coronal, and sagittal) view of the CT image volume, with the POLARIS pointer. Next, a wireframe sphere was manually placed on the texture-mapped surface at each landmark. The wireframe sphere is moved manually with the mouse of the computer, but it is forced to move so that it is always centered along the surface.

For each landmark, the texture-mapping error, e , was defined as the 3-D distance (in millimeters) between the measured and the texture-mapped coordinate positions

$$e = \sqrt{(x_m - x_s)^2 + (y_m - y_s)^2 + (z_m - z_s)^2} \quad (4)$$

where (x_m, y_m, z_m) are the 3-D coordinates of the landmarks, measured on the preoperative images with the POLARIS pointer and (x_s, y_s, z_s) are the 3-D coordinates where the landmark is mapped, determined by positioning the sphere. Since the surface loses some detail due to segmentation and surface decimation, for better accuracy, we used the POLARIS pointer to measure the 3-D coordinates (x_m, y_m, z_m) on the preoperative images instead of moving a wireframe sphere on the surface. If our surface extraction, physical to image space registration, tip localization, and optical modeling were absolutely accurate, the texture-mapped coordinates would correspond to the actual coordinates and e would be zero. The maximum, mean, and standard deviation error values were calculated from values of e for all the landmarks.

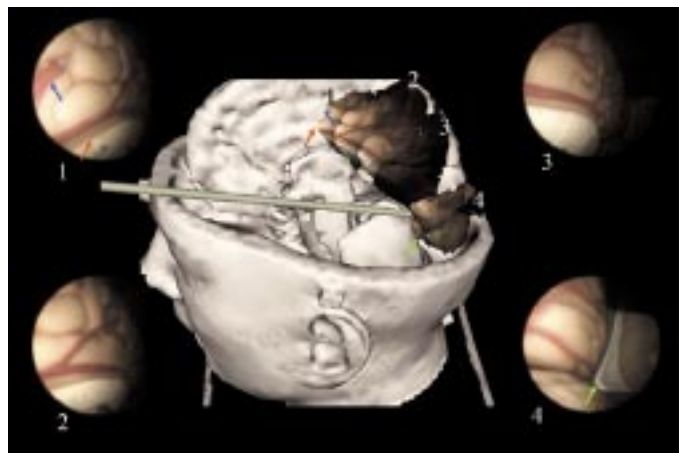


Fig. 5. Panoramic view showing the endoscopic views marked 1–4 texture mapped to the surface. Examples of texture-mapping errors are indicated with pairs of colored arrows. The blue arrow indicates a landmark on view #1 that was not painted onto the surface due to incorrect clipping. The orange arrow indicates the top of the corpus callosum on view #1 and the estimated corresponding position on the surface. The green arrow indicates the tip of the skull at the confluence of the occipital lobe and the cerebellum on view #4 and the estimated corresponding position on the surface.

III. RESULTS

A. Visual Assessment

Fig. 4(a) shows an original endoscopic video image, Fig. 4(b) the corresponding virtual endoscopic view, while Fig. 4(c) presents a view of this image, texture-mapped to the surface, through the virtual endoscope. Fig. 4(d) shows the texture-mapped surface from another perspective. This particular endoscopic view is an oblique projection of the 3-D scene, as shown clearly by the lighting in the endoscopic image itself [Fig. 4(a)] and from the virtual endoscopic view [Fig. 4(b)] as well as the obliquity of the modeled viewing cone.

Transparency Effects: Fig. 5 shows a panorama of multiple endoscopic views mapped to the surface, along with the individual endoscopic images used to create the panorama. The light intensity of the endoscope decreases as the distance between the surface and the endoscope tip increases. Therefore, for oblique views, such as the one in Fig. 4(d), large parts of the texture are dark and contain limited information because no light reaches the surface. In our system, we provide the following option: we can decrease the transparency of the darker pixels in the endoscopic image so that the underlying surface, which contains more information, can show through. In Fig. 6, the endoscopic views shown in Fig. 5 are texture mapped to the surface with the transparency option enabled. The transparency of the endoscopic image is set to decrease linearly between two predefined upper and lower luminance threshold values. These thresholds can be predetermined from the lighting model of the endoscope and can be adjusted interactively.

For comparison, Fig. 7 shows a photograph of the physical phantom from approximately the same perspective as Figs. 5 and 6. Fig. 8 shows a stereoscopic rendering of the texture-mapped surface, indicating the manner in which the underlying 3-D character of the phantom can be imparted to the monoscopically acquired endoscopic images.

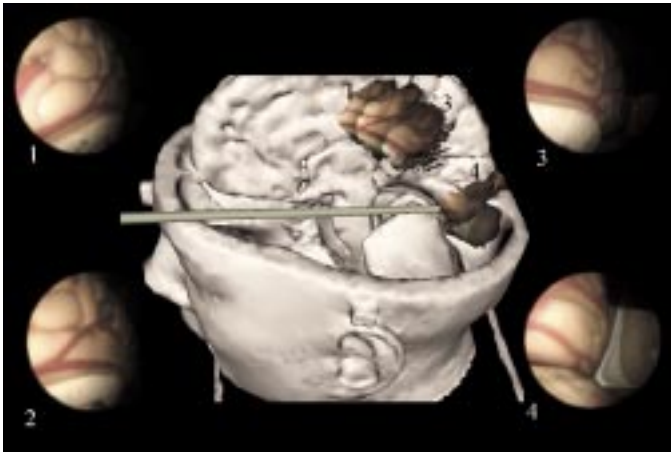


Fig. 6. Panoramic view showing the same four endoscopic views as in Fig. 5 texture mapped to the surface, with the transparency option enabled.

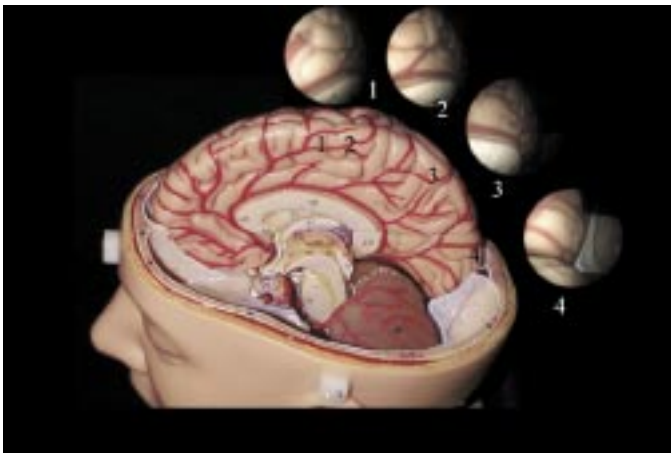


Fig. 7. Photograph of the phantom from approximately the same perspective as in Figs. 5 and 6.

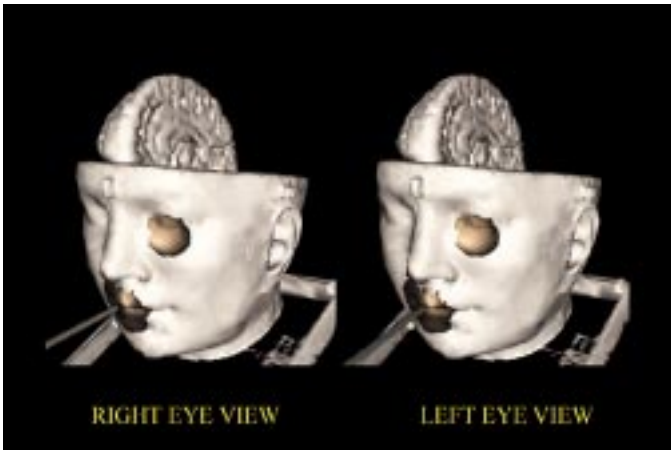


Fig. 8. Stereoscopic view of texture-mapped surface. To view this in stereo, the left and right eyes should be crossed to allow the two views to fuse together.

B. Errors

From our preliminary error assessment, the measured errors in tip localization and texture mapping are shown in Table I. The residual error for registering physical space and image space was found to be 0.8 mm.

In Table I, the tip localization error is governed by the tracking accuracy of the optical tracking system, as well as errors in 3-D

TABLE I
TIP LOCALIZATION ERROR AND TEXTURE MAPPING ERROR

	Mean	Standard Deviation	Maximum
	(mm)	(mm)	(mm)
Tip Localization	1.0	0.3	1.3
Error (n=7)			
Texture mapping	2.4	0.7	3.1
Error (n=7)			

physical-to-image space registration. The texture-mapping error is the overall error of our methods. Since the endoscopic image is projected to the surface, both the tip localization error and errors in our optical model affect the overall error. The average texture-mapping error value is, therefore, greater than the tip localization error.

By comparing Figs. 5–7, it can be seen that the positioning and scaling of the endoscopic views after texture mapping, generally correspond with the actual phantom. However, errors do exist and a few can be seen on Fig. 5. There are errors in the position and orientation of the fused image. One example is the top of the corpus callosum, indicated with orange arrows on the endoscopic view (view #1) and the estimated corresponding position on the surface (Fig. 5). Another example is the tip of the skull, at the confluence of the occipital lobe and the cerebellum, indicated with green arrows on the endoscopic view (view #4) and the estimated corresponding position on the surface (Fig. 5). A few segments of the endoscopic image can escape being mapped altogether. One example of this type of error is indicated with a blue arrow on the endoscopic view (view #1) and the estimated corresponding position on the surface (Fig. 5).

IV. DISCUSSION

In this paper, we have demonstrated real-time merging of multiple tracked 2-D endoscopic views to the corresponding irregular 3-D surfaces. This approach allows panoramic and stereoscopic visualization of the “painted” surface from arbitrary perspectives, as well as navigation within the volume containing these surfaces, even after the endoscope has been removed. Navigation within the texture-mapped surface subsequent to the surgical procedure can also be a valuable tool for training new surgeons for endoscopy-assisted surgical procedures.

The computer graphics approaches we have adopted to generate the stereoscopic panoramas are novel and differ from other methods described in the literature. In computer graphics, standard texture mapping consists of mapping a single picture or a scene to a regular geometrical object such as a plane, box, or sphere [24]. In some commercial image visualization applications such as *Quicktime VR* [26] and *Surround Video* [27] multiple photographic or video images are “stitched” together to form visual panoramas. In these methods, the camera acquiring the images is required to be under some kind of constraint, for example, to be made to rotate about a fixed central axis or to follow a predefined path [26], [27]. The acquired im-

ages, whether acquired by a constrained [26], [27] or a freehand [28] camera, are combined and mapped back to regular geometrical objects such as planar, spherical, or cylindrical surfaces to form visual panoramas. However, this approach is not suited for freehand endoscopic images since the 3-D surfaces are not regular and the endoscope is physically close to the surface.

In computer graphics research, for art and architecture-related applications, some investigators have mapped multiple real views of physical objects to their geometrical counterparts, to generate realistic software models of pieces of art and architecture. Debevec *et al.* [29] have mapped real photographs of architectural objects to their geometrical counterparts via view-dependent texture mapping. Wood *et al.* [30] used surface light fields to map multiple 2-D snapshots taken from multiple cameras spherically distributed around objects of art, back onto their surface representation.

In medical imaging applications, Jannin *et al.* [13] demonstrated merging of single microscope images to the corresponding 3-D surfaces. Clarkson *et al.* [15] texture-mapped a single stereoscopic 2-D video image of an object to its 3-D geometrical counterpart and utilized this mapping to track stereo video cameras with respect to the surface. Our method differs from all these approaches. Here, it is necessary to adopt computer graphics techniques, specially adapted for endoscope optics, that would allow us to fuse *multiple* endoscopic images and would perform well with a large number of endoscopic views.

Another surface-painting method common in computer graphics is that of vertex coloring. In this method polygon vertices would be colored via a color lookup table based on the corresponding pixel values in the endoscopic image. Since the spatial resolution of the endoscope is far superior to that of preoperative MR and CT (up to 60 times from our measurements), surfaces painted by vertex coloring appear less detailed and incomplete when compared with the texture-mapped surfaces. In addition, texture mapping can be accelerated by standard graphics hardware. Surface light fields, as proposed by Wood *et al.* [30], are also generally not suitable since the spatial resolution of the endoscope far exceeds that of the preoperative images and some physical details may be lost in the extracted surface. In their paper, Wood *et al.* acquired a very accurate surface with laser range scanning [30]. We could further refine our texture-mapping method by blending overlapping textures by assigning weights based on the viewpoint of the observer, as described by Debevec *et al.* [29].

Our graphics libraries are layered on top of OpenGL. Until recently, only one texture could be assigned to a surface at a time; that is, if a second endoscopic view were added, it would replace the previous endoscopic view. Therefore, a surface patch corresponding to each endoscopic view was necessary. However, the newly introduced OpenGL 1.2 standard has multitexturing options [25] which would allow us to add and blend multiple endoscopic views together, while eliminating the time-intensive clipping step. However, if multitexturing is used for surface painting, for each endoscopic view, texture coordinates must be computed for all the vertices in the surface, instead of just the vertices in a small patch. Our current algorithm allows us to run our application with both OpenGL 1.2 and its earlier versions.

When two endoscopic views overlap, several triangles from the two surface patches become superimposed. Since the depth of overlapping triangles can be the same, this can confuse the rendering engine and lead to rendering artifacts. Typically, this manifests in the form of jagged unpainted triangles in the last mapped view. This problem would also be resolved if we could assign multiple textures to just one surface.

Our hidden-surface-removal algorithm has limitations, since removing the whole polygon, if it is partly hidden by another polygon, leads to unnecessary gaps in the panorama. We are currently implementing an improved hidden-surface-removal algorithm, which splits a partially hidden polygon into visible and hidden subcells. This deficiency in our current hidden-surface-removal method, however, does not affect the panoramas shown in this paper, since for the surface patches corresponding to the endoscopic images shown here, the depth buffer values were within the specified tolerance and, hence, no polygons were removed.

The work described here has been based on phantom validation experiments, where the preoperative image data accurately reflects the intraoperative state. In an actual endoscopy assisted procedure, however, the physical-to-image-space registration will no longer be accurate because of soft-tissue motion. We intend to use the endoscopic image to update the registration between the surface patch and the endoscopic image, similar to the methods proposed by Clarkson *et al.* [14], [15], who utilized the real video image to update the physical-to-image-space registration.

Accurate tracking, tip localization and an accurate optical model are key components to the success of our 2-D-to-3-D registration. From our experience, before every endoscopic procedure, the endoscope must be optically calibrated, a procedure that needs to be made faster and more reproducible and should be extended to accommodate angled endoscopes. In this paper, our algorithm was applied to individual digitized video frames acquired with a tracked endoscope, but we are adapting the algorithm to digitized streaming video.

Our methods are for rigid endoscopes, with their tip position extrapolated from the position of the tracking LEDs external to the body. However, tracking technologies exist that employ optical fiber properties [ShapeTape (Measurand Inc., Fredericton, Canada)] and electromagnetic localization principles (Biosense Inc, Seatauket, NY and Northern Digital, Waterloo, Canada) that could detect the position and orientation of the tip itself. This could enable us to employ the techniques described in this paper with flexible endoscopes.

Our surface-painting algorithm is not limited to endoscopic images only. It could be applied to map video or still images of any object, acquired by any calibrated camera, back to the corresponding surfaces.

V. CONCLUSION

We have demonstrated the first steps toward multimodal integration of endoscopic images to 3-D preoperative images. Our method allows panoramic and stereoscopic visualization of the merged 3-D dataset from arbitrary perspectives and navigation

of the painted dataset after the procedure. While originally conceived as a means of augmenting neuro-endoscopic procedures, it is clear that this approach will have application in other endoscopy-assisted surgery (e.g., cardiac or orthopedic) and as a means of enhancing the realism of computer-modeled organs in surgical simulators.

ACKNOWLEDGMENT

The authors would like to thank their colleagues Dr. Y. Starreveld and Dr. A. Parrent for many useful discussions. They would also like to thank Trudell Medical Inc. and Smith and Nephew for the loan of endoscopic equipment and Nuclear Diagnostics, Sweden, for the use of Multimodality software for image processing and segmentation. Finally, they would like to thank Dr. R. Grzeszczuk (Intel Corporation Research Lab) for his insightful comments.

REFERENCES

- [1] A. Perneczky and G. Fries, "Endoscope assisted brain surgery: Part 1—Evolution, basic concept and current technique," *Neurosurgery*, vol. 42, pp. 219–224, 1998.
- [2] G. Fries and A. Perneczky, "Endoscope assisted brain surgery: Part 2—Analysis of 380 procedures," *Neurosurgery*, vol. 42, pp. 226–231, 1998.
- [3] J. D. Stefansic, A. J. Herline, W. C. Chapman, and R. L. Galloway, "Endoscopic tracking for use in interactive, image-guided surgery," in *Proc. SPIE Conf. Image Display Proceedings*, vol. 3335, 1998, pp. 208–129.
- [4] L. M. Auer and D. P. Auer, "Virtual endoscopy for planning and simulation of minimally invasive neurosurgery," *Neurosurgery*, vol. 43, pp. 529–548, 1998.
- [5] S. L. Aquino and D. J. Vining, "Virtual bronchoscopy," *Clin. Chest Med.*, vol. 20, pp. 725–730, 1999.
- [6] F. Liewald, G. Lang, T. Fleiter, R. Sokiranski, G. Halter, and K. H. Orend, "Comparison of virtual and fiberoptic bronchoscopy," *Thorac. Cardiovasc. Surg.*, vol. 46, pp. 361–364, 1998.
- [7] O. Kronborg, "Colons polyps and cancer," *Endoscopy*, vol. 32, pp. 124–130, 2000.
- [8] D. Paik, C. F. Beaulieu, R. B. Jeffrey Jr., C. A. Karadi, and S. Napel, "Visualization modes for CT colonography using cylindrical and planar map projections," *J. Comput. Assist. Tomogr.*, vol. 24, pp. 179–188, 2000.
- [9] D. Drascic and P. Milgram, "Positioning accuracy of a virtual stereographic pointer in a real stereoscopic video world," in *Proc. SPIE Conf. Stereoscopic Display and Applications II*, vol. 1457, 1991, pp. 58–69.
- [10] D. L. G. Hill, L. A. Langsaeter, P. N. Poynter-Smith, C. L. Emery, P. E. Summers, S. F. Keevil, J. P. M. Pracy, R. Walsh, D. J. Hawkes, and M. J. Gleeson, "Feasibility study of magnetic resonance imaging-guided intranasal flexible microendoscopy," *Comput. Aided Surg.*, vol. 2, pp. 264–275, 1997.
- [11] A. J. Sherbondy, A. P. Kiraly, A. L. Austin, J. P. Herferty, S. Y. Wan, J. Z. Turlington, T. Yang, C. Zhang, E. A. Hoffman, G. McLennan, and W. E. Higgins, "Virtual bronchoscopic approach for combining 3D CT and endoscopic video," in *Proc. SPIE Conf. Image Display*, vol. 3978, 2000, pp. 104–116.
- [12] W. Konen, M. Scholz, and S. Tombrock, "The VN project: Endoscopic image processing for neurosurgery," *Comput. Aided Surg.*, vol. 3, pp. 144–148, 1998.
- [13] P. Jannin, A. Bouliou, J. M. Scarabin, C. Barillot, and J. Luber, "Visual matching between real and virtual images in image guided neurosurgery," in *Proc. SPIE Conf. Image Display*, vol. 3031, 1998, pp. 518–526.
- [14] M. J. Clarkson, D. Rueckert, D. L. G. Hill, and D. J. Hawkes, "Registration of multiple video images to pre-operative CT for image guided surgery," in *Proc. SPIE Conf. Image Processing*, vol. 3661, 1999, pp. 14–23.
- [15] M. J. Clarkson, D. Rueckert, A. P. King, P. J. Edwards, D. L. G. Hill, and D. J. Hawkes, "Registration of video images to tomographic images by optimising mutual information using texture mapping," in *Lecture Notes in Computer Science*. Berlin, Germany: Springer-Verlag, 1999, vol. 1679, Proceedings MICCAI, pp. 290–300.
- [16] D. Dey, D. G. Gobbi, K. J. M. Surry, P. J. Slomka, and T. M. Peters, "Mapping of endoscopic images to object surfaces via ray-traced texture mapping for image guidance in neurosurgery," in *Proc. SPIE Conf. Image Display*, vol. 3976, 2000, pp. 290–300.
- [17] D. Dey, P. J. Slomka, D. G. Gobbi, and T. M. Peters, "Mixed reality merging of endoscopic images and 3-D surfaces," in *Lecture Notes in Computer Science*. Berlin, Germany: Springer-Verlag, 2000, vol. 1935, Proceedings MICCAI, pp. 796–803.
- [18] J. Foley, A. van Dam, S. Feiner, and J. Hughs, *Computer Graphics*, 2nd ed: Addison Wesley, 1990.
- [19] W. E. Lorensen and H. E. Cline, "Marching cubes: A high resolution 3D surface construction algorithm," *Comput. Graph.*, vol. 21, pp. 163–169, 1987.
- [20] W. Schroeder, K. Martin, and B. Lorensen, *The Visualization Toolkit: An Object-Oriented Approach to 3D Graphics*, 2nd ed. Englewood Cliffs, NJ: Prentice-Hall, 1997.
- [21] R. Y. Tsai, "A versatile camera calibration technique for high-accuracy 3-D machine vision metrology using off-the-shelf TV cameras and lenses," *IEEE J. Robot. Automat.*, vol. RA-3, pp. 323–345, 1987.
- [22] H. Haneishi, Y. Yagihashi, and Y. Miyake, "A new method for distortion correction of electronic endoscope images," *IEEE Trans. Med. Imag.*, vol. 14, pp. 548–555, Sept. 1995.
- [23] B. K. P. Horn, "Closed-form solution of absolute orientation using unit quaternions," *J. Opt. Soc. Amer. A*, vol. 4, pp. 629–642, 1987.
- [24] P. Heckbert, "Survey of texture mapping," *IEEE Comput. Graph. Applicat.*, pp. 56–67, 1986.
- [25] M. Woo, J. Neider, T. Davis, and D. Shreiner, *OpenGL Programming Guide*, 3rd ed. Reading, MA: Addison-Wesley, 1999.
- [26] QuickTime VR (2001, July). [Online]. Available: <http://www.apple.com/quicktime/>
- [27] Surround Video (2001, July). [Online]. Available: <http://www.immersivemedia.com>
- [28] R. Szeliski and H. Y. Shum, "Creating full view panoramic image mosaics and environment maps," in *Proc. SIGGRAPH*, Computer Graphics Proceedings, 1997, pp. 251–258.
- [29] P. E. Debevec, C. J. Taylor, and J. Malik, "Modeling and rendering architecture from photographs: A hybrid geometry- and image-based approach," in *Proc. SIGGRAPH*, Computer Graphics Proceedings, 1996, pp. 11–20.
- [30] D. N. Wood, D. I. Azuma, K. Aldinger, B. Curless, T. Duchamp, D. H. Salesin, and W. Stuetzle, "Surface light fields for 3D photograph," in *Proc. SIGGRAPH*, Computer Graphics Proceedings, 2000, pp. 287–206.