

STEPS - an Application for Simulation of Transsphenoidal Endonasal Pituitary Surgery

André Neubauer*
Lukas Mroz‡

Stefan Wolfsberger†
Rainer Wegenkittl*

Marie-Thérèse Forster†
Katja Bühler*

*VRVis Research Center
Vienna, Austria

† Department of Neurosurgery
Medical University Vienna, Austria

‡ Tiani Medgraph AG
Brunn/Gebirge, Austria

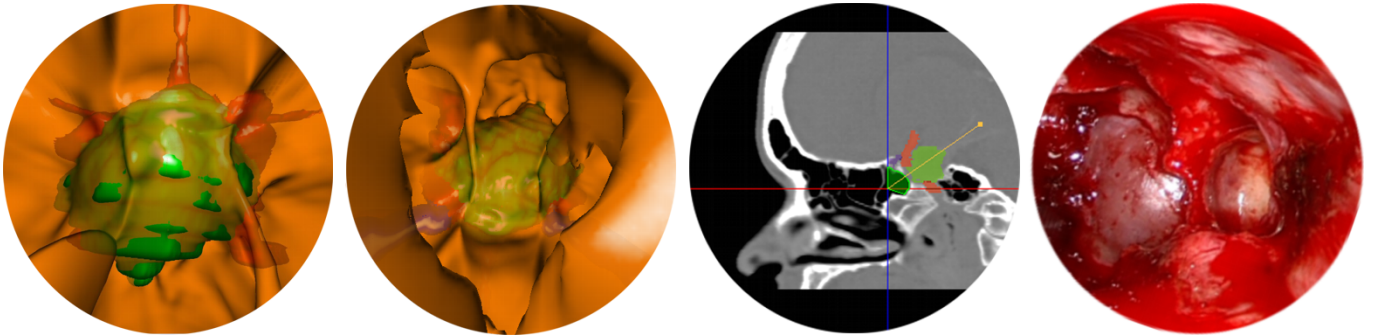


Figure 1: Left: A tumor (green) partly penetrating the sellar floor as also observed intraoperatively, internal carotid artery (red); Center left: The sphenoid sinus with a tumor (green), vessels (red) and the optic nerve (purple) in the background; Center right: A cross-section of the raw CT data - the tumor (light green), vessels (red) and the optic nerve (purple) are indicated as well as the current eye-point (intersection of red and blue line, inside the sphenoid sinus), the viewing vector (yellow) and the intersection of the viewing frustum with the cross-section (dark green); Right: Real endoscopic view of the sellar floor

ABSTRACT

Endonasal transsphenoidal pituitary surgery is a minimally invasive endoscopic procedure, applied to remove various kinds of pituitary tumors. To reduce the risk associated with this treatment, the surgeon must be skilled and well-prepared. Virtual endoscopy can be beneficial as a tool for training, preoperative planning and intraoperative support. This paper introduces STEPS, a virtual endoscopy system designed to aid surgeons in getting acquainted with the endoscopic view, the handling of instruments, the transsphenoidal approach and challenges associated with the procedure. STEPS also assists experienced surgeons in planning a real endoscopic intervention by getting familiar with the individual patient anatomy, identifying landmarks, planning the approach and deciding upon the ideal target position of the actual surgical activity.

Besides interactive visualization using two different first-hit ray casting techniques, the application provides navigation and perception aids and the possibility to simulate the procedure, including haptic feedback and simulation of surgical instruments.

CR Categories: K.6.1 [Computer Graphics]: Picture/Image Generation—Display Algorithms;

Keywords: virtual endoscopy, ray casting, iso-surfacing, pituitary surgery

*{neubauer,wegenkittl,buehler}@vrvis.at

†stefan.wolfsberger@univie.ac.at, marie.johanna@chello.at

‡lukas.mroz@tiani.com

1 INTRODUCTION

Minimally invasive endoscopic procedures are constantly gaining importance in modern medicine [2]. One example is endonasal transsphenoidal pituitary surgery, used for the removal of various kinds of pituitary tumors: A rigid endoscope and surgical instruments, including a bone punch (*rongeur*), are inserted into the patient's nose and advanced through the natural nasal airway into the *sphenoid sinus*. This cavity of the skull is separated from the pituitary gland only by a thin bony structure, the *sellar floor*. The sellar floor is opened using the *rongeur*, the tumor is cut off surrounding tissue and removed through the patient's nose.

The procedure is associated with some risk for the patient: The *internal carotid artery* (ICA) which supplies large parts of the brain passes the pituitary gland behind the sellar floor and is therefore invisible to the surgeon. Damage to this large vessel may be life-threatening to the patient. Although this happens very infrequently (in less than 0.2 percent of cases), still all possible measures have to be applied to further reduce the risk.

The optic nerve and, of course, the pituitary gland itself are also obscured by the sellar floor and endangered to be damaged by the *rongeur*. In order to keep risk at a low level, it is obviously of utmost importance for the surgeon to be skilled and experienced as well as to get familiar with the patient's anatomy prior to each surgical procedure. Acquisition and analysis of both MRI (Magnetic Resonance Imaging) and CT (Computer Tomography) scans of the patient's head are thus usually part of the preoperative routine diagnostic workup. Besides standard 2D viewing techniques, the surgeon can therefore, without any additional radiological examinations, use virtual endoscopy which can be of benefit in three different ways:

Preoperative Planning: While careful slice-by-slice-analysis of the raw images obtained from the modalities can give a good overall picture of the situation to a skilled surgeon, this picture can still be greatly enhanced by additionally using simulated 3D views provided by virtual endoscopy. This helps the surgeon to get familiar with the anatomy of the patient, decide on the route to take and plan where exactly to open the sellar floor.

One of the most important capabilities of virtual endoscopy is that the boundaries of the investigated cavity can be displayed semi-transparently with objects of interest in the background. Rendering the tumor, the pituitary gland, the ICA and the optic nerve behind the semi-transparent sellar floor can give the surgeon a profound impression of what to expect during the real surgical procedure and of the risks associated with it.

Intraoperative Support: Virtual endoscopy can be used during the real surgical procedure both as a navigation aid and to decide on the exact position of application of the rongeur. If the position of the real endoscope is tracked, the viewing parameters can be retrieved and transferred to the virtual endoscopy system.

Training: Simulation of the real endoscopic operation by not only providing the visual impression but also modeling haptic parameters and constraints of movement can serve as a tool to exercise the complete procedure from the insertion of the endoscope into the patient's nostril to the opening of the sellar floor. Compared to training on cadavers, virtual endoscopy offers advantages like lower cost, unbounded repeatability and the possibility to use navigation aids to speed up the learning process.

Virtual endoscopy has been reported to be usable for visualization of the nasal cavity and the paranasal sinuses [14]. Talala et al. assessed the benefits of a combination of virtual endoscopy with non-perspective semi-transparent imaging of surface landmarks for transphenoidal pituitary surgery [17].

This paper describes a virtual endoscopy system designed to aid neurosurgeons as a training and planning device for endonasal pituitary surgery. Section 2 describes the clinical workflow and points out the requirements arising from it. Sections 3 and 4 introduce the implemented application. Section 5 discusses benefits and problems experienced by neurosurgeons working with the application.

2 WORKFLOW AND REQUIREMENTS

The first step in clinical workflow is data acquisition. Radiological images of the patient's head are generated using both CT and MRI. CT data give information about skull anatomy. The pituitary gland, the tumor and the optic nerve are usually well captured by MRI. The ICA is emphasized in both MRI and CT by the application of contrast media. Data set sizes usually amount to approximately $512 \times 512 \times 170$ voxels (CT) and $512 \times 512 \times 100$ voxels (MR).

In order to be able to visualize distinct objects, these objects must be extracted (segmented) from the acquired data volumes prior to the virtual fly-through. Apart from the artery, all objects must be segmented in the MR image, since they can hardly be distinguished from surrounding tissue in the CT image. The visualization of the boundaries of the investigated cavities during the virtual endoscopy procedure is, however, based on the CT-data. Therefore, the two data sets must be geometrically aligned (registered) to ensure exact correlation of object locations. After registration, objects of interest can be segmented in the MR image and displayed at correct positions during the CT-based virtual fly-through. The system must therefore provide accurate and easy-to-use modules for image segmentation and registration.

After the extraction of objects of interest, virtual endoscopy can be started. For this, fast and flexible perspective visualization is required. The visualization system must support semi-transparent viewing, displaying the user-defined objects in the background.

The parameters of the system depend on whether training, preoperative planning or intraoperative support is required. In the case of training, the virtual endoscopy system should simulate conditions as encountered during a real endoscopic procedure. Thus the system should be capable of mimicking the challenges and the feeling associated with endonasal pituitary surgery. This includes modeling the correct parameters and constraints of movement, as well as soft tissue deformation and haptic feedback. In the case of preoperative planning and intraoperative support, the emphasis is laid on studying the individual patient anatomy - exact simulation is thus of minor importance.

In order to keep the workflow efficient, all functions should be embedded into a single medical workstation, possibly even connected to a PACS (Picture Archiving and Communication System).

3 THE APPLICATION

STEPS was realized as a collection of three modules, embedded into the Java-based PACS J-Vision (Tiani Medgraph AG, Brunn/Gebirge, Austria, <http://www.tiani.com>). Registration, segmentation and virtual endoscopy were each encapsulated into one module. The following paragraphs describe the registration and segmentation modules. The virtual endoscopy module is covered in detail in section 4.

3.1 Registration

The registration module aligns two three-dimensional data sets of the head by using rigid body transformations with six degrees of freedom (three translations and three rotations). The alignment is achieved by optimizing a similarity criterion (based on mutual information) using adaptive simulated annealing [3]. Usually a result is obtained within approximately 10 seconds. In case the global optimum is not found automatically, the registration process can be supported manually.

3.2 Segmentation

The segmentation module is used to extract objects of interest. Two different segmentation modes are provided:

Manual Segmentation: Voxels are added to or removed from the object manually in a paint/erase fashion.

Watershed From Markers: Different objects are defined by the user by placing a set of markers on the image, each marker belonging to one object. For each voxel the cheapest path to any of the markers is the automatically identified. The voxel is added to the object to which this marker belongs [4]. Usually manual refinement or additional executions of the algorithm are needed to obtain a result of the desired quality.

User interaction in the segmentation module takes place via cross-sections of the volume, which can be arbitrarily moved and rotated. The segmentation result can be previewed in 3D at any time.

Other semi-automatic segmentation methods are currently being developed in order to reduce the time needed for segmentation (currently about 30 minutes per patient).

The result of the segmentation process is a binary volume stating for each voxel whether it is part of the segmented object. This bit-mask can be saved to disk and later loaded into the virtual endoscopy module to define a new background object.

4 THE VIRTUAL ENDOSCOPY MODULE

The virtual endoscopy module is the main part of the application. Sections 4.1 and 4.2 deal with visualization. Section 4.3 describes the modes of operation that were implemented. Section 4.4 introduces additional features, common to all modes of operation.

4.1 Visualization - Related Work

To be applicable, a virtual endoscopy system must provide instant reaction to user input. This results in the need for high frame rates and thus fast visualization. Most existing systems therefore depict iso-surfaces instead of using transfer-function-based direct volume rendering [2]. An iso-surface is defined by the *threshold* t and separates voxels whose intensities are smaller than t from those whose intensities are greater than t . Roughly two groups of iso-surfacing algorithms can be identified: those which use hardware-accelerated polygonal surface rendering and those using CPU-based first-hit ray casting.

Polygonal surfaces are often extracted using the Marching Cubes algorithm [11] or similar techniques. These algorithms usually create a large number of polygons. Therefore, to enable interactive rendering, usually mesh simplification techniques [7] and occlusion culling [5, 6] have to be applied.

In first-hit ray casting [9], for each pixel a viewing ray is tracked through the volume and intersected with the iso-surface. As soon as the first intersection has been found, ray tracking is stopped and the pixel is shaded, the pixel intensity depending on the orientation of the iso-surface at the intersection point. Plain first-hit ray casting is usually too CPU-intensive to be performed at interactive frame rates. Therefore optimization strategies, which either reduce ray traversal distances or increase ray tracking efficiency, have to be applied. Distance fields [21] are a well-known method of accelerating ray-traversal through empty spaces: Each voxel is assigned its distance to the nearest point of the iso-surface. This distance can be skipped by a ray passing the voxel. The distance field, however, must be reprocessed whenever the threshold is changed, which reduces the frame rate during threshold adjustment. Also, distance fields require a considerable amount of memory. A more flexible approach is the so-called Lipschitz method [16]: The data value at the current ray position, the maximum data gradient inside the currently traversed sub-volume and the threshold are used to calculate a minimum distance to the iso-surface on-the-fly. Changes of the threshold therefore do not induce any performance loss. The distances that can be skipped, however, tend to be rather small, limiting the gain of computation speed.

There are a number of approaches which pay for increased performance by reducing image quality [8, 15, 18] or flexibility [19].

4.2 Visualization - The Implemented Solution

For the described application, optimized first-hit ray casting techniques are applied for the following two reasons:

Flexibility: Polygonal surface rendering bears the disadvantage of reduced flexibility: Due to the costly surface extraction process, it is generally not possible to react to changes of voxel values or the iso-threshold at interactive frame rates. In STEPS, however, flexibility is of major importance, as pointed out in sections 4.4 (volume manipulation for simulation of the rongeur) and 5.1 (interactive threshold adjustment).

Company Policy: The application was embedded into the J-Vision PACS by Tiani Medgraph AG, which for reasons of hardware-independence completely abandoned the use of polygon rendering.

For STEPS, two different first-hit ray casting techniques were implemented, one depicting the foreground (the inner walls of the investigated cavities, an iso-surface in the original CT data) and one visualizing pre-segmented background objects. Since the foreground always covers the complete image and thus each pixel has to be processed, an optimized image-order (pixel by pixel) ray casting technique is applied. Background objects cover a varying number of pixels, depending on their shapes and their positions relative to the viewing frustum. When background objects cover only few pixels, application of image-order ray casting results in a large number of rays cast in vain. Therefore the background is generated by applying an optimized object-order (based on object projections) first-hit ray casting algorithm. In each frame, the two resulting images (foreground and background) are merged using a z-buffer. The following paragraphs summarize the two techniques:

4.2.1 Foreground Visualization

The foreground is rendered using image-order first-hit ray casting. To achieve interactive frame rates, several acceleration techniques are applied:

Exploitation of Pixel-Space Coherence: The image is subdivided into tiles of 8×8 pixels and a ray is cast for each corner of a tile. If the colors obtained from the cast rays are sufficiently similar, the remaining pixels of the tile can be computed from the corner pixels using bilinear interpolation. Otherwise the tile is recursively subdivided and further rays have to be cast for the sub-tiles [10]. Using this approach, rays have to be cast for only 10-25 percent of all pixels in a typical endonasal view. This strategy occasionally leads to artifacts, which are, however, hardly noticed during interaction. This optimization is not used when rendering the final quality image when the viewing frustum remains still.

Acceleration of Memory Access: Retrieving volume data along an arbitrarily oriented ray usually exhibits low utilization of the CPU cache. To reduce the number of accesses to the data volume, two binary acceleration volumes (1 bit per voxel) are used. A *void volume* marks volume cells which do not contain the object's surface. During ray traversal marked cells can be simply skipped.

Space Leaping: A *leap volume* marks cells whose distances from the iso-surface exceed a value v . Whenever a marked cell is encountered during ray casting, the ray can be advanced by a distance of v without missing the surface. The leap volume approximates the benefits of distance fields at much lower memory cost.

The acceleration volumes have to be recomputed after a change of the iso-threshold. The computation, however, is fast: the void volume is generated trivially by checking the voxel values for each cell and the leap volume is calculated by eroding the void volume. During threshold adjustment the acceleration volumes are not used resulting in a slight drop of the frame rate (by approximately 10 percent). The time that it takes for the user to switch between threshold adjustment and any other interaction is usually sufficient to recalculate the acceleration volumes.

In the vicinity of the iso-surface rays are tracked efficiently from cell-boundary to cell-boundary [1] in order to avoid missing thin structures.

4.2.2 Background Visualization

Background visualization is performed on a synthetic background volume which is generated by combining a set of binary segmen-

tation masks as described in section 4.2.3. For each object an iso-surface is displayed, the threshold can be adapted individually for each object. The object-order *cell-based first-hit ray casting* algorithm [12, 13] is applied: The volume extracted from the CT scan of the patient's head is divided into cubic bricks, referred to as *macro-cells*. These bricks form the leaves of a min-max-octree, implemented as an integer-array [20]. This octree can be efficiently traversed to find all macro-cells containing a part of an iso-surface in order of increasing distance from the eye-point. Each of those macro-cells is then trimmed by removing redundant parts and projected to the image plane. The projection is rasterized to the screen, scan line by scan line. For each pixel which is covered by the projection and has not yet been assigned a color, a local ray segment is launched and tracked inside the macro-cell from cell boundary to cell boundary [1]. When a cell containing a part of an iso-surface is encountered, a ray-surface intersection test is performed. This basic algorithm is accelerated using, among others, the following concepts:

Early Scan Line Termination: Early scan line termination is a heuristic measure yielding a reduction of the number of ray segments tracked. Based on an analysis of the iso-surface geometry inside the macro-cell, a decision is made, whether rasterization of a scan line can be stopped as soon as a ray segment did not intersect the iso-surface.

Ray Segment Concatenation: Instead of confining tracking of a ray segment to the current macro-cell, the ray segment may be expanded across a macro-cell boundary, if the neighboring macro-cell also contains a part of an object. This measure reduces the number of ray segments and therefore the administrative cost associated with launching them.

Exploitation of Pixel-Space Coherence: To avoid tracking lots of ray segments into an empty part of the volume, the screen is sparsely sampled for parts of the iso-surface before switching to the final resolution. Also, pixels whose neighbor pixels are shaded similarly, may acquire their shading through interpolation between neighbor pixels, rather than through ray casting.

Adaptive Rendering: The way a macro-cell is processed can be adapted to the size of the projection of the macro-cell. This includes reduction of sampling frequency in far macro-cells as well as optimizations being applied only when they are expected to be effective.

4.2.3 Background Volume Generation

For visualization of the selected background objects, a synthetic volume is generated from all segmentation masks prior to the virtual endoscopic procedure. In this background volume, each background object is represented by an iso-surface. Thresholds are either fixed or can be adapted by the user individually for each object. This is because there are two kinds of objects: Those which can be represented as an iso-surface in a part of the original volume and those that cannot.

An example for objects of the first group is the ICA. Since a contrast agent is used during data acquisition, the artery is imaged at higher intensity than most of its surroundings. Still, an iso-surface in the original CT volume with the correct threshold, representing the vessel boundary, would also include parts of bones and, of course, other large vessels of the head. In the contrasted MRI volume, at least other vessels would be included. The best way to get a clean high-quality image of the artery is to manually create a segmentation mask containing the vessel and its low-intensity surroundings and to copy the values of the voxels contained by the

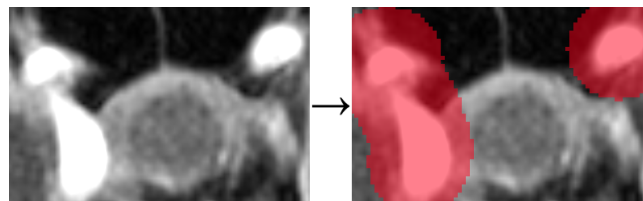


Figure 2: Manual segmentation of the vessel: The vessel is only roughly segmented (red) in the MR-Angio data set. The underlying intensity values are transferred to the background volume. Segmentation is completed using interactive threshold adjustment during virtual endoscopy.

mask from the original contrasted MRI volume to the background volume (see Figure 2). Since cell-based first-hit ray casting allows interactive threshold adjustment, the user can then, during the virtual investigation, complete the segmentation of the artery by selecting the correct threshold. The default value of the threshold is an empirically determined average. With this method, the vessel does not have to be reconstructed from a binary mask, the appearance of the vessel boundary as recorded by the CT is maintained.

The second group consists of those segmented objects whose boundaries cannot be represented by iso-surfaces in the original data volume. Examples are the tumor, the pituitary gland and the optic nerve. For each of those objects, an artificial iso-surface must be constructed in the background volume. This can be achieved by selecting a fixed threshold t for the object and assigning a value smaller than t to each voxel surrounding the object and a value larger than t to each voxel that is part of the object. In order to keep the boundary smooth, the following strategy is applied [12]: A *reference mask* is created by eroding the binary segmentation mask using a spherical kernel of radius n . The voxel values are then selected such that the resulting iso-surface surrounds the reference mask at a constant distance d with $d = n$. Resulting filtering-artifacts are corrected in a post-processing step. All this is done automatically and takes a few seconds per object.

In order to prevent interferences among overlapping objects, the calculated values are written to the background volume using a maximum-operator: If a voxel obtains more than one value greater than zero (for two or more objects), the largest of these values is finally assigned. Each voxel belongs to one object at most - to the one which determines the voxel value. Rendering neighboring objects with different thresholds can, however, still result in artifacts. To keep these artifacts at an insignificant level, the fixed threshold for reconstructed objects is set within the interval of reasonable thresholds for the vessel. Also, for object reconstruction, a similar data range as encountered in the MR-Angio data is used. Therefore, in relation to the data range, thresholds usually differ only slightly.

4.2.4 Image Fusion

In each frame during a virtual endoscopy investigation, two images are generated, one depicting the foreground and one showing background objects of interest. The two images are combined using a weighted average operator, yielding the impression of the foreground being semi-transparent. A z-buffer is used to determine those regions of the image where the background is not covered by the foreground. This allows previewing the effect of opening the sellar floor (see section 4.4). The z-buffer is also used to enhance depth perception: The distance between foreground and background at each pixel determines foreground opacity (see Figure 3). This makes the association between foreground and background more intuitive and navigation easier.

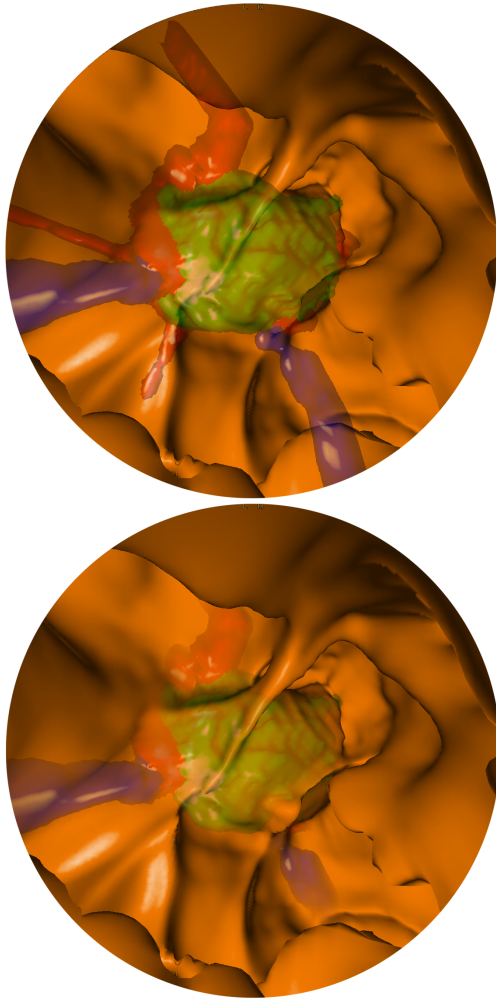


Figure 3: Top: simple blending, constant foreground opacity; Bottom: improved depth perception, foreground opacity z-buffer-dependent

4.3 Modes of Operation

As pointed out in the introduction, STEPS can be used for various purposes - as a training device, for preoperative planning and for intraoperative support. Therefore the application provides different modes of operation. The modes differ in the way user interaction is implemented:

4.3.1 Free Flight

In the free flight mode, the user is allowed to move freely, with the single restriction that the eye point is not allowed to cross an iso-surface. This mode is used for preoperative planning and intraoperatively - to get familiar with a patient's anatomy, view landmarks from different viewing angles, and plan where to open the sellar floor.

4.3.2 Surgery Simulation

In this mode of operation, the process of advancing the endoscope through the narrow nasal cavities and finding the correct way to the sphenoid sinus can be trained as well as the most crucial part of the actual surgical operation, the opening of the sellar floor. For training purposes, it is important to model the degrees of freedom and

constraints experienced when working with a rigid endoscope as accurately as possible. The trainee should be confronted with similar challenges as in the real procedure and guided by haptic feedback. Therefore, in this mode of operation, the most important parts of the user-application interface are provided by a force-feedback joystick (WingMan Force 3D, Logitech, <http://www.logitech.com>, see Figure 4). The following paragraphs describe this interface.

Control: The simulation starts with the user selecting a *start point* in the CT volume. This point defines the position where the endoscope is entered into the patient and is usually located inside one of the patient's nostrils. After selection of the start point, the eye point is automatically moved to this position. The user can move forwards (into the viewing direction) and backwards using two joystick buttons and rotate the virtual endoscope by moving the joystick handle.

The endoscope is modeled as the line segment connecting the start point and the eye point. It can only be rotated around the start point. The start point remains constant most of the time. Only if the user requests a rotation which cannot be performed due to collisions (see the following paragraph), the start point may be moved within a bounded region surrounding the specified initial start point. This simulates the fact that the surgeon is able to deform the patient's nose to some extent, which increases the mobility of the endoscope and the accessibility of certain structures and pathways.

Collision Detection: The nasal airway consists of a number of cavities in the skull. These cavities are separated from each other by thin bony walls, so-called *septa*. Cavities can additionally be divided into chambers by *incomplete septa*. Bones form the barriers for the movement of the endoscope. They are, however, usually covered by soft tissue (*mucosa*), which also poses resistance to movement, but can be deformed to some extent.

The threshold used for foreground visualization is usually selected such that soft tissue is depicted. In the simulation, actions which would cause the eye point to move across the visualized foreground-iso-surface are disallowed. To prevent the eye-point from jumping past thin walls, after each movement, the path from the previous to the current eye point is sampled. If a surface intersection is found, the eye-point is restored to its previous position. Sampling is performed by casting a viewing ray, such that any visible structures are encountered with a very high probability. Apart from the eye-point, the virtual endoscope is allowed to be pushed into soft tissue. This is accomplished by using two thresholds, a soft threshold t_{soft} and a firm threshold t_{firm} . Threshold t_{soft} represents the surface of the mucosa (≈ -100 Hounsfield Units) and is equal to the iso-value used for foreground visualization. Threshold t_{firm} represents the boundary between mucosa and bony structures (≈ 600 HU). No

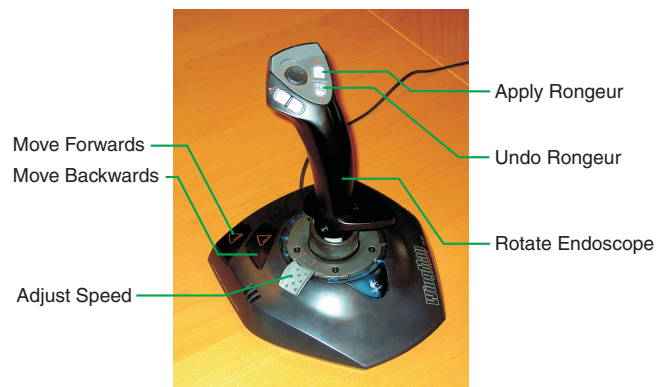


Figure 4: The joystick and the mapping of operations

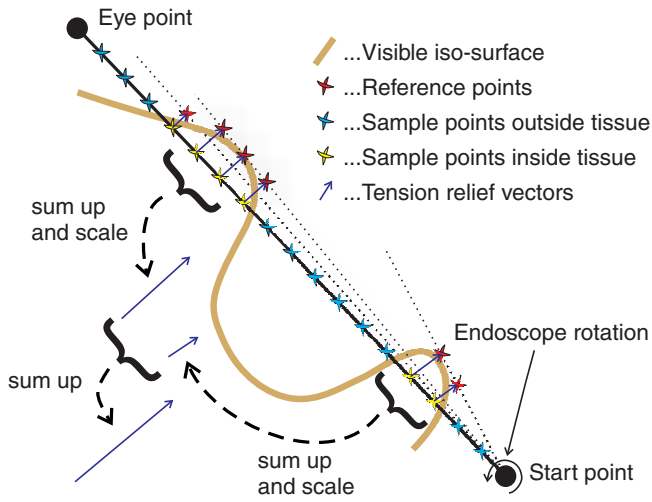


Figure 5: Calculation of the tension relief vector after endoscope rotation

part of the virtual endoscope is allowed to be moved to an intensity value greater than t_{firm} .

Application of Force-Feedback: If a part of the virtual endoscope is inside soft tissue, force-feedback is exerted with the goal to press the endoscope back into regions of lower image intensity. The direction and intensity of force-feedback is calculated using the following algorithm, illustrated in Figure 5: The sorted set of intersection points of the virtual endoscope with the visible iso-surface is a sequence of alternately entry points (of the endoscope into soft tissue) and exit points (out of soft tissue). Each *intersection segment* (a segment of the endoscope confined by an entry point and the subsequent exit point) is associated with a source of tension, because the mucosa that has been pushed aside exerts some pressure on the endoscope. The direction into which the endoscope should be moved in order to reduce the tension caused by the n th intersection segment most effectively, is defined by the *tension relief vector* $v_{tension}^n$.

Tension relief vectors are approximated as follows: In each frame, each $v_{tension}^i$ is, first of all, set to zero. Then the CT data is sampled at evenly-spaced points along the virtual endoscope, from the start point to the eye point. The inter-sample distance d remains constant throughout the simulation. The λ^{th} sample point is therefore always located at distance $\lambda \cdot d$ from the start point. For each λ a reference point is stored, indicating the last position of a λ^{th} sample point whose intensity value was lower than t_{soft} . If the intensity value at a sample point is smaller than t_{soft} , the corresponding reference point is updated. Otherwise, the vector from the sample position to the reference point is added to $v_{tension}^i$ with i being the number of the intersection segment that spans the sample point. Each resulting vector $v_{tension}^i$ is normalized and scaled with a coefficient c^i which is directly proportional to the largest value in the (slightly low-pass filtered, to reduce the impact of noise) sequence of intensity samples between the i^{th} entry and the i^{th} exit point and directly proportional to the distance between the i^{th} entry point and the start point (to simulate torque). Finally the sum of all tension relief vectors is projected to the screen to acquire a summed tension relief vector in screen coordinates. This vector can be directly passed on to the force-feedback-joystick. The intensity of force-feedback is determined by the Euclidean norm of the vector.

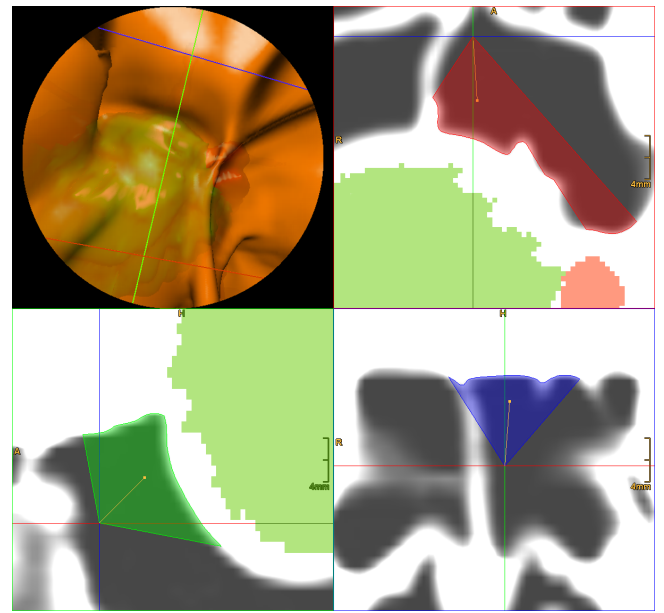


Figure 6: The 3D view (top left) and the three section views are linked to each other. Note the septum on the left hand side of the 3D view, rendered incompletely due to a too high threshold, as confirmed by the blue section view (bottom right).

4.3.3 Guided Navigation

This mode of operation is also used for training purposes - here the trainee is confronted with the same constraints as in the simulation mode - the virtual endoscope is constrained to movements which would be possible also with a real rigid endoscope. In this mode, haptic feedback is not aimed at simulating the feeling encountered in reality, but used instead to guide the trainee towards the target, the tumor. The user must select a start point and a *target point* (usually inside or near the tumor). Haptic feedback is given only when the user diverges too significantly from the path defined by these two points.

4.4 Additional Features

The following features are available in each mode of operation:

Linked Views: The screen is separated into four parts: the 3D-view displaying the simulated endoscopic image and a multi-planar reconstruction (MPR) consisting of three *section views* displaying raw CT data and selected background objects on three mutually orthogonal cross-sections of the volume. The current eye point is the intersection point of the three cross-sections. The cutting planes defining the cross-sections can be arbitrarily rotated by the user. The section views are linked to each other - the intersection between any pair of cross-sections is displayed in both section views. The user can also link any of the section views to the 3D view. If a section view is linked, the intersection of its cutting plane with the viewing frustum is displayed in both the 3D view and the section view (see Figure 6). This tremendously improves navigation, since structures seen in the 3D view are mapped to their locations in the section view in a very intuitive way. This technique also serves as an instrument to find the ideal threshold for the foreground-iso-surface.

Simulation of opening the sellar floor: The user can select the radius and the depth of the rongeur to use. The region of im-

part of the currently selected rongeur is displayed on the iso-surface (see Figure 7, left image). Those parts of the background which are within the range of the rongeur can be highlighted (see Figure 7, center image). Also the actual opening can be simulated - the intensity values within a cylindrical region, the region of impact of the rongeur, are adapted such that a hole is virtually cut into foreground tissue (see Figure 7, right image). Information about which objects have been hit by the rongeur can be retrieved. An undo-function is provided, thus allowing several attempts.

Angled Endoscopes: Angled endoscopes have the special property that the principal viewing vector is not parallel to the axis of the endoscope, but slightly rotated (either 30, 45 or 70 degrees). This enables the surgeon to obtain a significantly enhanced field of view. STEPS supports simulation of angled endoscopes.

Lens Distortion: Endoscope lenses usually exhibit slight barrel distortion resulting in non-linear projection. This can be simulated by STEPS: The foreground image is distorted by altering ray directions. The background image must be distorted in a post-processing step, since cell-based first-hit ray casting relies on the linearity of projections.

Guidance: If a target point is specified, the direction to the target is, as a navigation aid, indicated during the virtual endoscopic procedure.

5 DISCUSSION

The feasibility of STEPS has been tested within the scope of a medical research project. The value of the application in the field of endonasal transsphenoidal pituitary surgery was assessed.

5.1 Visualization

Feedback from the medical doctors concerning the visualization was predominantly positive. The surgeons observed a good correlation of the intraoperatively observed and the virtually reconstructed anatomy. Rendering speed (between 6 and 16 fps, depending on the sizes and shapes of background object projections, and the roughness of the foreground surface) was considered sufficient. The ability to interactively adjust the thresholds of the foreground and each background object was accepted well: Interactive adjustment of the foreground threshold even added some unexpected features to the application, as, for example, the virtual removal of septa.

A well-known problem of iso-surfacing based on intensity values as obtained from a CT is that iso-surfaces generally do not entirely correspond to boundaries as observed in real endoscopy. The reason is that optimal threshold values differ between structures, even between different structures seen in the same image. Due to the fact that CT data appear slightly low-pass filtered, a small variation of the threshold can have a quite drastic effect on image appearance. So a slightly wrong threshold can result in a certain structure appearing quite differently from its real shape as seen during the real endoscopic procedure. However, adding an additional dimension to user interaction by providing interactive threshold adjustment significantly reduces the impact of this problem.

Interactive threshold adjustment is not only feasible for the foreground, but also for background objects, since it enables the user to provide only a rough manual segmentation of a certain structure (e.g., a well-contrasted blood vessel) and finish the segmentation process interactively during the 3D virtual investigation (see section 4.2.3).

5.2 Application as a Training Device

Endoscopic procedures are relatively new in the field of neurosurgery. Neurosurgeons are used to the handling and the three-dimensional view provided by operating microscopes and often unfamiliar with the magnified, laterally distorted and sometimes angled field of view of an endoscope. Before performing endoscopic pituitary surgery on a real patient for the first time, a surgeon must therefore go through a training process. STEPS was recognized to be beneficial as a training tool. The only alternative to virtual endoscopy is cadaver training which is associated with much cost and effort. Advantages of virtual endoscopy over cadaver training include increased reproducibility (a patient can be examined a virtually unbounded number of times and effects of applying the rongeur can be reverted), and an improved learning curve by utilizing navigation aids.

Still, training as offered by the current application suffers from some drawbacks: A force-feedback-joystick is a cheap but not ideal solution to the problem of modeling haptic feedback as experienced in real life. Although the force-feedback algorithm that is currently used (see section 4.3.2) yields a good feeling of soft-tissue-deformation during endoscope rotation, a realistic simulation cannot be provided yet: The eye point is prevented from crossing the visible iso-surface. This reduces the maneuverability of the virtual endoscope. It, for instance, sometimes prevents the virtual endoscope from being retreated. The reason is that if the endoscope intersects the visible iso-surface (see Figure 5), pulling it in direction of the inverted viewing vector, towards the start point, would at some point cause the eye point to cross the iso-surface, which is forbidden. In real endoscopy, this restriction does not exist. Geometrical simulation of soft tissue deformation is needed to overcome this problem. A suitable model is currently being developed. Another drawback identified during the tests is that there is no haptic feedback caused by tissue contact during advancement of the endoscope. This is because force-feedback evidently can only be exerted through the joystick handle. The endoscope, however, is advanced and pulled back using joystick buttons, rather than the handle which is used only for endoscope rotations. This problem can only be overcome by using some highly sophisticated haptic feedback device which comes at very high cost.

Because of these downsides, STEPS does not yet provide the 'real' feeling experienced by the surgeon during the actual endoscopic procedure. It is therefore not yet capable of fully replacing cadaver training. Still, the advantages outlined in this section make it a valuable addition.

5.3 Application for Preoperative Planning

For neurosurgeons who are experienced in endoscopic pituitary surgery, STEPS is beneficial as a tool for preoperative planning because it provides prospective information about the anatomic variations in the individual patients. The value of STEPS in preoperative planning was assessed in a series of 22 patients: In 12 of 22 patients, visualization of the position of the pituitary gland was useful for the planning of an individually tailored approach to the tumor minimizing possible damage to the gland. In 10 of 22 patients STEPS showed that the tumor partially encased the ICA. In these cases, STEPS facilitated planning of the operative removal of the tumor, thus minimizing the risk of damage to the ICA. Of the 22 patients studied, 15 had more than one complete or incomplete sphenoid septum. In all of these patients, STEPS successfully identified the individual anatomy of the sphenoid sinus chambers, facilitating selection of the opening site of the sellar floor.

Apart from this, STEPS improved preoperative planning in the following ways:

One decision aided by STEPS is about the side of the approach. The surgeon has the choice, which nostril to use as entry point for

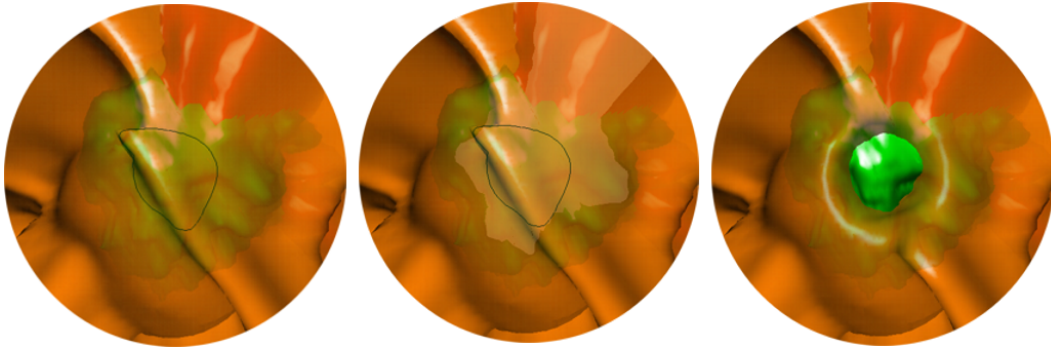


Figure 7: Simulation of opening the sellar floor: Left: The region of impact of the rongeur is indicated; Center: Parts of the background which are within range of the rongeur are highlighted; Right: The sellar floor has been opened.

the endoscope. STEPS can quickly give information about the accessibility of the sphenoid sinus through each nostril. Septa might have to be opened for sufficient access. The simulation tells the surgeon, which septa will have to be removed and whether removal of a septum might cause damage to the ICA. In order to plan the route to take during the real procedure, the surgeon can preoperatively determine landmarks (e.g., septa). Septa overlying the carotid artery can also serve as landmarks in the decision as to where to open the sellar floor. The semi-transparent display provided by STEPS allows intuitive spatial association of the segmented objects with foreground landmarks. This effect is enhanced by the improved depth cueing as described in section 4.2.4 and by highlighting those parts of background objects which are within reach of the rongeur (see section 4.4).

6 FUTURE WORK

Currently a model for simulation of tissue deformation is being developed. This is expected to greatly enhance the realism of the simulation and therefore further increase the applicability of the system as a training device. Another planned extension is to connect the application to a tracking system to optimize intraoperative support.

ACKNOWLEDGEMENTS

This work has been carried out as part of the research project 'Virtual Reality for Scientific Applications' at the VRVis research center (<http://www.vrvis.at>) in Vienna, Austria, which is funded by the Austrian governmental research project Kplus. Cooperating partners were Tiani Medgraph AG (<http://www.tiani.com>) and the Department of Neurosurgery, Medical University Vienna.

REFERENCES

- [1] J. Amanatides and A. Woo. A fast voxel traversal algorithm for ray tracing. In *Proc. of Eurographics '87*, pages 3–10, 1987.
- [2] D. Bartz. Virtual endoscopy in research and clinical practice. In *Eurographics State-of-the-Art-Reports 2003, (S4)*, 2003.
- [3] M. Capek and R. Wegenkittl. Robust and fast medical registration of 3D-multi-modality data sets. In *Proc. of Medicon '01*, pages 515–518, 2001.
- [4] P. Felkel, R. Wegenkittl, and M. Bruckschwaiger. Implementation and complexity of the watershed-from-markers algorithm computed as a minimal cost forest. In *Proc. of Eurographics '01*, pages 26–35, 2001.
- [5] R. Hietala and J. Oikarinen. A visibility determination algorithm for interactive virtual endoscopy. In *Proc. of IEEE Visualization '00*, pages 29–36, 2000.
- [6] L. Hong, S. Muraki, A. Kaufman, D. Bartz, and T. He. Virtual voyage: interactive navigation in the human colon. In *Proc. of International Conference on Computer Graphics and Interactive Techniques (SIGGRAPH 97)*, pages 27–34, 1997.
- [7] Hugues Hoppe. Progressive meshes. *Computer Graphics*, 30(Annual Conference Series):99–108, 1996.
- [8] K.A. Kreeger, I. Bitter, F. Dacheille, B. Chen, and A. Kaufman. Adaptive perspective ray casting. In *Proc. of Symposium on Volume Visualization*, pages 55–62, 1998.
- [9] M. Levoy. Display of surfaces from volume data. *IEEE Computer Graphics and Applications*, 8(5):29–37, 1988.
- [10] M. Levoy. Volume rendering by adaptive refinement. *The Visual Computer*, 6(1):2–7, 1990.
- [11] W. E. Lorensen and H. E. Cline. Marching cubes: a high resolution 3D surface construction algorithm. In *Proc. of SIGGRAPH '87*, pages 163–169, 1987.
- [12] A. Neubauer, M. Forster, R. Wegenkittl, L. Mroz, and K. Bühler. Efficient display of background objects for virtual endoscopy using flexible first-hit ray casting. In *Proc. Eurographics/IEEE TCVG Symposium on Visualization (VisSym '04)*, pages 301–310, 2004.
- [13] A. Neubauer, L. Mroz, H. Hauser, and R. Wegenkittl. Cell-based first-hit ray casting. In *Proc. Eurographics/IEEE TCVG Symposium on Visualization (VisSym '02)*, pages 77–86, 2002.
- [14] M. De Nicola, L. Salvolini, and U. Salvolini. Virtual endoscopy of nasal cavity and paranasal sinuses. *Eur J Radiol*, 24:175–180, 1997.
- [15] H. Qu, M. Wan, J. Qin, and A. Kaufman. Image based rendering with stable frame rates. In *Proc. of IEEE Visualization 2000*, pages 251–258, 2000.
- [16] B. T. Stander and J. C. Hart. A Lipschitz method for accelerated volume rendering. In *Proc. of IEEE Visualization '95*, pages 107–114, 1995.
- [17] T. Talala, T. Pirila, V. Karhula, E. Ilkko, and I. Suramo. Preoperative virtual endoscopy and three-dimensional imaging of the surface landmarks of the internal carotid arteries in trans-sphenoidal pituitary surgery. *Acta Otolaryngol*, 120:783–787, 2000.
- [18] A. Vilanova Bartroli, R. Wegenkittl, A. König, and M. E. Gröller. Perspective projection through parallelly projected slabs for virtual endoscopy. In *Proc. of SCCG '01- Spring Conference on Computer Graphics*, pages 287–295, April 2001.
- [19] R. Wegenkittl, A. Vilanova, B. Hegedüs, D. Wagner, M. C. Freund, and M. E. Gröller. Mastering interactive virtual bronchoscopy on a low-end PC. In *Proc. of IEEE Visualization 2000*, pages 461–464, 2000.
- [20] J. Wilhelms and A. van Gelder. Octrees for faster isosurface generation. *ACM Transactions on Graphics*, 11(3):201–227, July 1992.
- [21] K. J. Zuiderveld, A. H. J. Koning, and M. A. Viergever. Acceleration of ray-casting using 3D distance transforms. In *Visualization in Biomedical Computing II, Proc. SPIE 1808*, pages 324–335, 1992.