Human-machine interface for a VR-based medical imaging environment

Christian Krapichler[a], Michael Haubner[a], Andreas Lösch[a],
Manfred K. Lang[b], Karl-Hans Englmeier[a]

[a]GSF-Institute of Medical Informatics and Health Services Research,
85764 Neuherberg, Germany

[b]Institute for Human-Machine Communication, Technical University of Munich,
80290 Munich, Germany

ABSTRACT

Modern three-dimensional scanning techniques like magnetic resonance imaging (MRI) or computed tomo graphy (CT) produce high-quality images of the human anatomy. Virtual environments open new ways to display and to analyze those tomograms. Compared with today‘s inspection of two-dimensional image sequences, physicians are empowered to recognize spatial coherencies and examine pathological regions more facile, diagnosis and therapy planning can be accelerated.

For that purpose a powerful human-machine interface is required, which offers a variety of tools and features to enable both exploration and manipulation of the 3D data. Man-machine communication has to be intuitive and efficacious to avoid long accustoming times and to enhance familiarity with and acceptance of the interface. Hence, interaction capabilities in virtual worlds should be comparable to those in the real world to allow utilization of our natural experiences. In this paper the integration of hand gestures and visual focus, two important aspects in modern human-computer interaction, into a medical imaging environment is shown.

With the presented human-machine interface, including virtual reality (VR) displaying and interaction techniques, radiologists can be supported in their work. Further, virtual environments can even alleviate communication between specialists from different fields or in educational and training applications.

Keywords: virtual reality, human-computer interaction, gesture recognition, eye tracking, biosignals, medical image analysis, visualization, image segmentation

1. INTRODUCTION

Progress in the field of medical engineering introduced innovative scanning technologies which empower radiologists to obtain three-dimensional images of the human anatomy. Techniques like computed tomography (CT), magnetic resonance imaging (MRI), positron emission tomography (PET) or 3D ultrasound imaging (US) can produce high-quality recordings, represented as a stack of two-dimensional greyscale images.

However, analysis of those image sequences for diagnosis and therapy has been an arduous and time-consuming effort so far. There is remarkable success in accelerating parts of this process by automation,[1] but nevertheless human interaction will be inevitable in foreseeable future.[2] Three-dimensional image segmentation, computer-aided construction of bone implants or planning of radiotherapy are typical tasks requiring

---

Further author information –
C.K. (correspondence): Email: krapichler@gsf.de; WWW: http://ilona.gsf.de/~krapichl; Telephone: +49-89-3187-4455; Fax: +49-89-3187-4243
M.H.: Email: haubner@gsf.de; Telephone: +49-89-3187-4454
A.L.: Email: loesch@gsf.de; Telephone: +49-89-3187-4458
M.K.L.: Email: lg@mmk.e-technik.tu-muenchen.de; WWW: http://www.mmk.e-technik.tu-muenchen.de/persons/lg.html; Telephone: +49-89-289-28541; Fax: +49-89-289-28535
K.-H.E.: Email: englmeier@gsf.de; Telephone: +49-89-3187-4148

numerous user actions. Hence, our approach aims at the support of specialists in their work by modern means of man-machine communication. New techniques have to be applied which circumvent today‘s editing and manipulation of all the single tomogram slices.

Virtual reality applications[3] offer the possibility to visualize three-dimensional images, giving users the impression of real spatial perception. Thus, specialists can recognize topological coherencies in a much faster and more natural way. Immersion into a virtual environment facilitates the exploration of a scene of three-dimensional medical objects or tomograms and the examination of pathological regions. The benefits of VR technologies in industrial fields like architecture or flight simulation led to intensified research activities and an increasing number of virtual reality applications in medicine,[4] e.g. training and education programs, therapy planning or surgery assistance. But most of the available industrial software cannot fulfil the requirements in medical environments: In opposite to industrial applications, where virtual scenes commonly consist of some thousands of (usually textured) polygons, medical scenes are much more complex (up to millions of polygons). Even maximum performance computer hardware cannot achieve sufficient frame rates for larger virtual scenes, which would be necessary for immersion. But as no information must be distorted or lost, reduction of scene complexity is quite a delicate problem. And on the way from radiological image sequences to a virtual scene of geometrically described object surfaces, a lot of time-consuming steps are necessary (e.g. filtering, segmentation, triangulation), whereas data from e.g. CAD systems can easily be used for VR scenes.

Hence, most of medical applications, especially in education and training, are based upon models of the human anatomy, where the laborious model generation is only done once. Merely a few surgery systems can be found which evaluate and work with tomograms from individual patients. To empower specialists to explore and diagnose the radiological data in every particular case and use them for further treatment, we improved not only the visualization and simulation tasks, but also accelerated the processing of the 3D image sets. Therefore, refinements in visualization techniques and in human-computer interaction had to be achieved, honing existing applications and developing new methods in those areas.

In addition to spatial visualization, the human-machine interface has to provide efficient instruments for analysis of and interaction with the 3D data. Such an interface should support control of image processing and segmentation algorithms as well as offering numerous tools for analyzing, manipulating and representing both volume- and surface-oriented structures. And in spite of the desired variety of features, usage has to be intuitive and convenient. One characteristic feature of virtual reality is the adaptation of interaction devices to human senses. In an ideal case users could interact in virtual worlds in the same way as they would do in reality. In opposite to standard input devices like a keyboard or a 2D-mouse, operations in three dimensions with six degrees of freedom (6-DOF) have to be offered. Therefore, interaction devices should recognize and interpret human actions and gestures and transform them into corresponding manipulations of the virtual scene.

2. METHODS

An important aspect of virtual environments is the degree of immersion. Sufficiently high frame rates and minimal time delays between human actions and corresponding machine reactions are important factors. Hence, employment of high performance hardware still is an indispensable condition. But just as important for an increased system performance is the use and optimization of fast and efficient algorithms and software applications.

To achieve a broad acceptance for the new VR working environment, handling of the entire system has to be user-adequate. Physicians should be able to concentrate on the medical problems, not on the utilization of the equipment, and benefit from the new technologies without long periods of learning and adaptation.[5] Working with the application has to be as intuitive as possible, even with no technical background, while experience in the field of medicine could be taken into account. As one characteristic of virtual reality systems is the adaptation of input/output devices to the sensory, motoric and cognitive capabilities of humans, it seems to be predestined to fulfill those requirements.

2.1 Hardware environment

The underlying computer hardware for all VR applications is an Onyx InfiniteReality system from Silicon Graphics (SGI, Inc.), with a dual pipe configuration, twelve R4400 processors (200MHz), four raster managers, and 2GB main memory. But with some performance restrictions, the applications should be running on desktop workstations, too. The human-machine interface is based on the following components:

2.1.1. Display systems

- For a larger group of spectators or protagonists, a stereo projection system (TAN GmbH) with two video beamers is used, equipped with pairwise perpendicularly oriented polarizing filters. Each viewer is wearing spectacles with adequately oriented polarizing foils in order to perceive a stereoscopic image.
- The BOOM 3C (Binocular Omni-Orientation Monitor, Fakespace, Inc.) is a head-coupled stereoscopic color display and contains two cathode ray tubes (CRTs). It is able to track the head position of the user, evaluating the angles of all joints of the counterbalanced boom construction. Corresponding to this information, the current viewing position is determined and the virtual scene is calculated for each eye in real-time. A rocker switch with three positions (center/left/right) and a joystick-like button (center/up-down/left-right) allow further user interaction.
- Another stereo viewing system is a head-mounted display (HMD), the Datavisor 10X (n-Vision, Inc.) with see-through option. Equipped with two CRTs and semi-translucent mirrors, this high-resolution color stereo display offers simultaneous sight on the virtual and the real environment (augmented reality).

2.1.2. Input devices

- A desktop-based input device is the Space Mouse (SpaceControl GmbH). It offers an opto-electronically tracked control ball which allows full 6-DOF actions. Moving the ball in all spatial directions controls translation coordinates, toppling and twisting it enables rotafromtion. Additional interaction is possible via nine user-definable buttons.
- For user input via natural gestures, the optical fiber based 5th Glove (Fifth Dimension Technologies, Inc., 5DT) is used. With optical fibres and five sensors the flexion of each finger can be determined, and a tilt sensor measures roll and pitch values of the hand. The degrees of finger flexions are measured with eight bits resolution. Bending of finger joints is not determined separately. There is provided one measured value per finger, representing current finger flexion as a whole.
- The HMD (head position), a three-button 3D mouse and the glove (hand position) are tracked simultaneously with the pulsed DC magnetic field of the modular Flock of Birds system (Ascension Technology Corp.). The extended range transmitter allows motions within a radius of up to ten feet. Accuracy of magnetic tracking is about 2.5mm or 0.5° RMS, respectively, with corresponding resolutions of 0.8mm/0.1°. For each device a separate receiver is needed which provides the application software with information about the absolute position and orientation.
- Eye positions are determined with an electro-oculography (EOG) system (BioMuse[6,7], BioControl Systems, Inc.). Because of different potentials of cornea and retina, the eyeball behaves like a rotating dipole. The resulting biosignals are measured with silver/silver chloride electrodes, which are attached to a headband and placed around each eye. A digital signal processing unit amplifies, filters, digitizes and converts those raw analog signals in four digital data streams, containing information about horizontal and vertical position of each eye.

2.2 Software applications

Hands and eyes play a preeminent role in human‘s natural interaction. Recognizing and interpreting hand gestures and actions improves intuitiveness and efficiency of human-machine interfaces.[8] For this reason, a multiplicity of tools and actions is offered for image analysis and visualization, controlled via hand gestures.

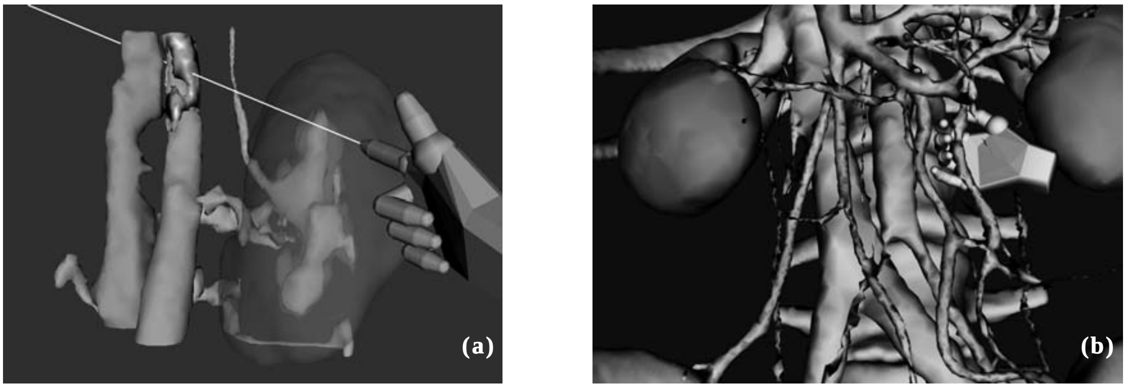

Fig. 1: Two examples of interaction in a virtual scene (abdominal area) via hand gestures: (a) selection of a calculus with the *pointing* gesture; (b) *grasping* a vessel (right).

And knowledge about user's current visual focus and its motion can be applied to control actions or tools, to adequately display VR scenes, or to log the eye path for later evaluation.

2.2.1. Hand gestures

All positions, motions and finger states of the fiber optic glove are reflected by a hand model in the virtual scene, showing location and orientation of the whole hand and current finger flexions. In the see-through mode of the HMD the virtual hand is congruent with the real hand, even the depth perception of both is matching. Therefore, two data streams have to be evaluated: the tracker information for hand position and orientation in space, and the glove information about current bending states of the five fingers. As HMD and glove are tracked with the same tracking device, their relative position is correct. If others than the HMD are used as display system, the position of the glove has to be calibrated every time the initial position of the display in relation to the magnetic tracking transmitter changes.

With different gestures out of a calibrated reference set (e.g. pointing, grasping, navigation, release) various actions are possible. For the definition of reference gestures natural human hand gestures were imitated for different actions. For example, stretching of the index finger, while all other fingers are bent, is the natural way of pointing to objects. And for grasping things, all fingers are bent like clenching a fist (Fig. 1). Calibrated gestures and current flexion states are represented as five-dimensional vectors. To determine the current user gesture, angle and size of the measured glove vectors are compared versus the vectors of all reference gestures. That gesture is chosen which shows least deviation in angle and size and stays within a predefined tolerance range.

2.2.2. Eye tracking

To determine screen coordinates of the visual focus, the eye tracking system has to be calibrated to get the eye position in relation to the display and the range of eye movements. After signal processing (digital filtering, correlation analysis) of the four data streams from the biosignal controller, horizontal and vertical positions of each eye are calculated, taking the calibration values into account (Fig. 2). As the HMD position is constant in relation to the viewer's eyes, it is not necessary to evaluate head tracking data for focus calculations. This is required for applications which allow users to move in front of a screen, e.g. the stereo projection system.

From a user's point of view there are two ways to apply eye movements to virtual environments: *active* control, for example moving an object or selecting a tool, or *passive* control, i.e. eye movements are evaluated in the background, the user doesn't notice any effects or reactions.

Because an active control of tools or objects with eye movements is not possible in the real world, passive evaluation seems to be more reasonable. As a first step, eye movements influence the image representation:

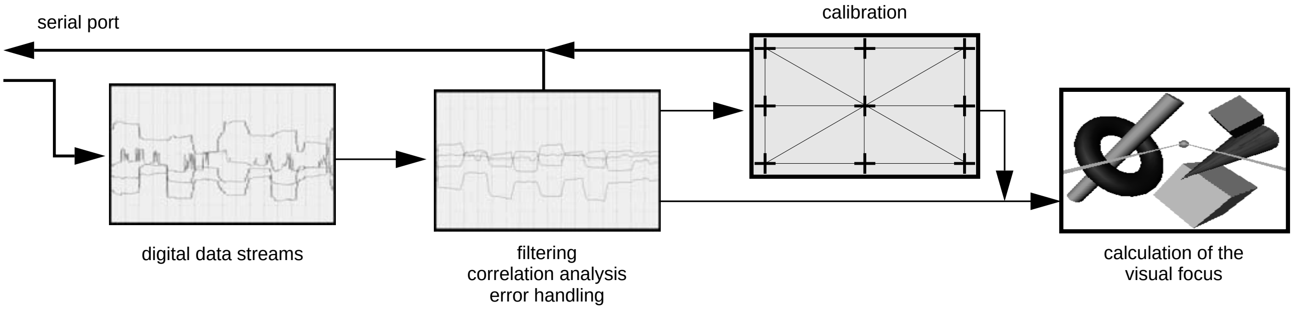

Fig. 2: Simplified flow chart of the eye tracking application to determine the current visual focus.

The visual focus is used as a parameter for a level-of-detail (LOD) algorithm. It allows different degrees of surface mesh resolution even within the same object. Only the region of interest is displayed accurately, peripheral regions show less details.

For an active control of virtual objects, some additional gestures would be required. But besides of interpreting eye blinking e.g. as different button clicks (left eye, right eye, both eyes) or connecting some extreme eye positions with particular actions, which would both be quite annoying in practice, eyes are not able to perform gestures. This means that active control by eye tracking would require a combination with other input sources (e.g. speech recognition).

3. RESULTS AND DISCUSSION

An important and difficult step in computer-aided analysis of tomography data is the segmentation of the medical images. As a lot of approaches still require human interaction, this work is supported and facilitated. One problem of segmentation algorithms emerges when two areas with similar greyscale values, but belonging to different objects, overlap in one or more layers of the tomogram. In such cases most volume growing or edge detection algorithms recognize those two objects as one. To correct such misinterpretations, it is necessary to work on the affected regions manually. Avoiding time-consuming layer-by-layer editing, selection of a three-dimensional barrier (plane, hemisphere, cube) is supported to place it around the location of interest and adapt it by scaling, rotation and translation. Because segmentation algorithms are prohibited to exceed these boundaries, the desired object can be detected automatically.

As the surfaces of the medical objects are highly complex, triangulation of larger scenes produces meshes with hundreds of thousands of polygons. Even today‘s high-end graphics hardware cannot provide sufficient frame rates. For that reason, scene complexity has to be reduced. But as visualization in medicine requires high accuracy, no information must be lost. The implemented LOD-algorithm reduces the number of polygons significantly. Depending on the objects of a scene, the number of triangles can be restricted to 3-25% of the original number. In combination with the eye tracking parameters, users do not notice this reduction. The region they are inspecting is displayed with a high resolution of the geometry mesh, while the number of polygons in peripheral regions can be reduced.

Advantages of the applied EOG method are the limited computing performance required for data processing, allowing real-time applications,[9] and the ability to wear the equipment in combination with VR displays. A major disadvantage of EOG is the sensitivity to several sources of interference.[10] For example, varying skin temperature or eye lid movements[11] can lead to drifting signals. Besides of standard digital signal processing, larger errors are reduced by a combined evaluation of eye tracking and hand tracking: selecting and moving objects requires to pay attention to them, they are focused for a longer period. Hence, object coordinates can be used to recalibrate the drifting EOG signals permanently.

Moving in a virtual scene is usually done by simply walking around, because head tracking data specify the viewer’s location and orientation, and the corresponding perspective of the scene representation is calcu-

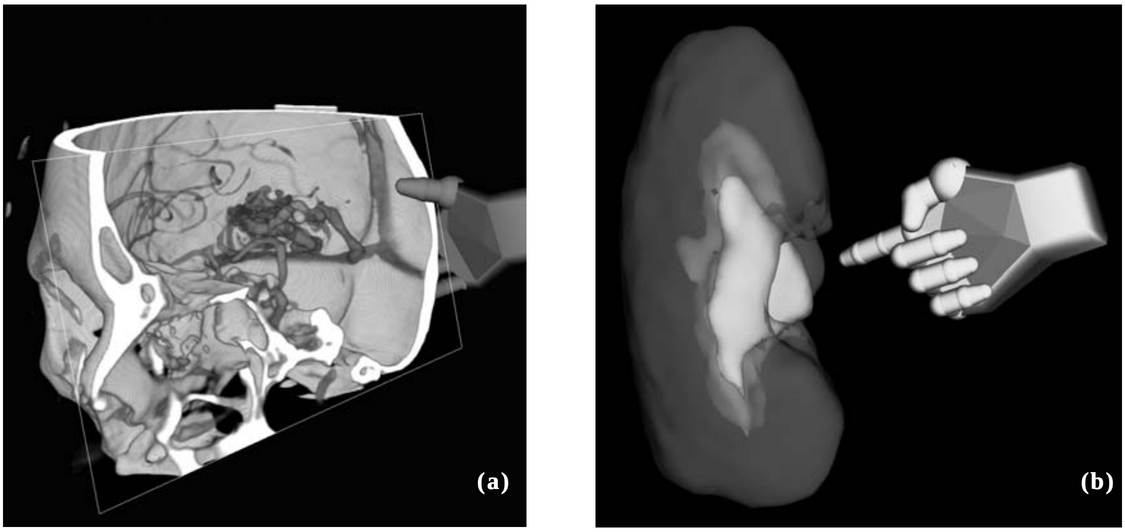

Fig. 3: Two possibilities to view internal structures of objects: (a) moving a clipping plane through a volume-rendered skull to examine a tumour; (b) transparent representation of the surface-rendered kidney to inspect a renal calculus.

lated for every frame. Additionally, navigation is possible with input devices like the Space Mouse or with a hand gesture. While performing such a *navigation* gesture, orientation of the hand is transformed to a change of viewing position and orientation, evoking an impression of flying through the scene. This empowers specialists to examine all relevant areas of the scene from different angles and positions, for example in virtual bronchoscopy. Even during fast user motions frame rates have to be sufficiently high to display images without noticeable delay. This is achieved with the above mentioned LOD algorithm. Thus an improved feeling of immersion is provided and simulator sickness can be avoided.

By selecting, moving and rotating the medical objects of interest, they can be explored without the need to walk around or fly through them. It is an intuitive way to look at things: take them in the hand and turn them to examine them from every side and every angle. Thus, diagnosis and therapy planning of complicated injuries or diseases, e.g. tumors or compound bone fractures, can be accelerated. To touch e.g. menu buttons and to grasp objects, the collision of the hand model with the objects of the virtual scene has to be detected. Therefore, a quite fast and accurate algorithm was implemented which uses oriented bounding box hierarchies (OBB-Trees[12]) for intersection tests.

Surfaces which hide other objects inside could simply be grasped and removed from the virtual scene. But to conserve information about location and orientation of the enclosed objects with respect to the hiding object and to keep the orientation inside the scene, there are two possible solutions: The first is to place a clipping plane through the surrounding object with a hand gesture. The outer object is opened and permits a view inside. Fig. 3a shows a volume-rendered skull with a tumour inside. By moving the clipping plane through the bone structure, the tumour becomes visible. As another example, it is necessary to examine the structure inside a renal calculus to predict the result of an ultrasound treatment. The calculus can be cut with the clipping plane and the original tomographical data inside the object can be analyzed. The second possibility is illustrated in Fig. 3b: if the internal object (the renal calculus) would also be cut by the clipping plane, though it should be shown as a whole, the enclosing object (the kidney) can be switched to a transparent state with a single gesture.

A combined representation of both original volumetric tomography data and segmented objects is a helpful support of diagnosis.[13] Fig. 4a shows an example of virtual angioscopy, a view inside the segmented aorta. The nebular looking bright areas are volume rendered original data. They represent the flow of the blood which is enhanced with a contrast medium. At the free lumen inside the aorta there is a tumour which restricts blood flow. Figs. 4b-d show further examples of this hybrid visualization.

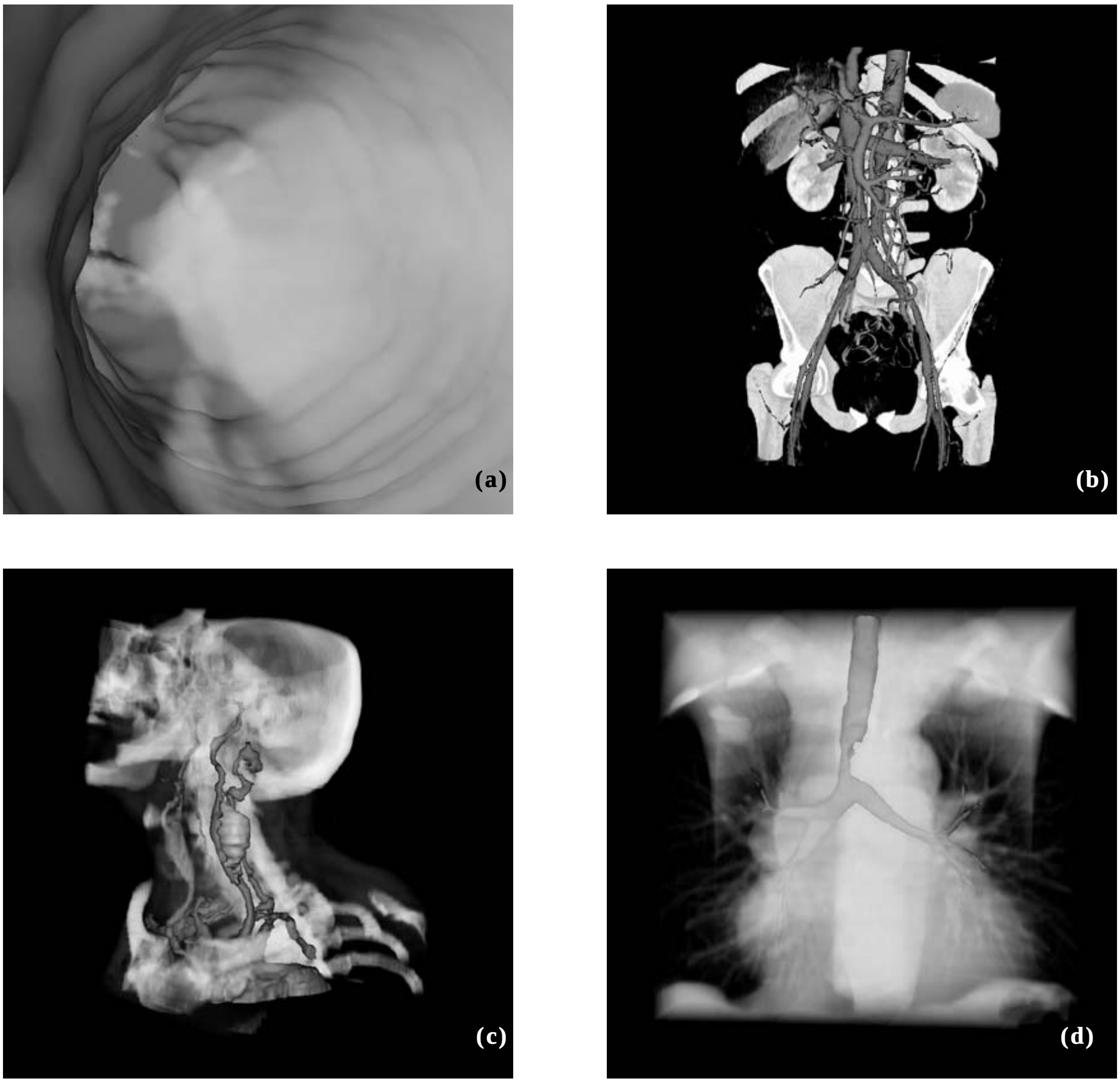

Fig. 4: Hybrid visualization: (a) surface representation of the aorta, volume-rendered blood flow (virtual angioscopy); (b) abdominal scene with a Leriche syndrome, volume-rendered skeleton and kidneys, surface-rendered vessels; (c) volume-rendered bony structures of the skull and neck, surface representation of the vessels and a thrombus; (d) surface representation of the trachea with a tumour, volume representation of the lung structure with a metastasis in the right lung.

4. CONCLUSION

The presented environment offers physicians an improved access to radiological data. Not only anatomical models can be visualized in three dimensions, but also image sets of individual patients. With a powerful man-machine interface, they can be analyzed and manipulated with the necessary precision. Compared with the conventional procedure of layer-by-layer editing, that work benefits from this VR technology. It is predetermined to support diagnosis, therapy, surgical planning, simulation or training and education in the field of medicine. For that purpose, the human-computer communication should offer powerful tools and features and has to be as intuitive and convenient as possible, though a lot of technical equipment is used. Then acceptance of VR will increase and it is likely to become a standard tool for medicine of the future.[14,15] Another important factor for its acceptance are the costs of hardware. But with progress in technology high performance graphics equipment will become cheaper, and even today it can be used with desktop workstations if some limitations are accepted.

Multimodality, one aspect of virtual reality, is an important factor in the realization of user-adequate interfaces, since human perception and interaction is based largely upon it. Several parallel channels should transmit information and provide communication, allowing transition between the different modi. Hence, speech recognition and voice output could be reasonable supplementations of the environment. And as virtual reality has made computer interfaces more intuitive but not more intelligent, knowledge-based interaction could enhance the flexibility and friendliness of the interfaces, thus broadening the application spectrum of human-computer interaction.

REFERENCES

1. N. Ayache, J.D. Boissonnat, L. Cohen, B. Geiger, J. Levy-Vehel, O. Monga, and P. Sander, ”Steps Toward the Automatic Interpretation of 3D Images“, *3D Imaging in Medicine*, *NATO ASI Series* **F60**, pp. 107-120, Springer, 1990.
2. M.W. Krueger, ”Automating Virtual Reality“, *IEEE Comp. Graph. & Appl.* **15(1)**, pp. 9-11, 1995.
3. S.R. Ellis, ”What are Virtual Environments?“, *IEEE Comp. Graph. & Appl.* **14(1)**, pp. 17-22, 1994.
4. K.-F. Kaltenborn, ”Virtual Reality in medicine“, *Meth. Inform. Med.* **32**, pp. 407-417, 1993.
5. M.K. Lang, ”Kommunikation zwischen Mensch und Maschine“, *mikroelektronik* **5(6)**, pp. 212-219, vde, 1991.
6. H.S. Lusted, R.B. Knapp, A.M. Lloyd, ”Biosignal processing and biocontrollers“, *Virtual Reality Systems* **1(1)**, pp. 38-39, 1993.
7. H.S. Lusted, R.B. Knapp, ”Medical applications for biocontroller technology“, *Proc. Medicine meets Virtual Reality II: Interaction Technologies & Healthcare*, pp. 45-47, Aligned Management Assoc., San Diego, 1994.
8. D.J. Sturman, D. Zeltzer, ”A survey of glove-based input“, *IEEE Comp. Graph. & Appl.* **14(1)**, pp. 30-39, 1994.
9. R.B. Knapp, H.S. Lusted, ”Real-time digital signal processing system for bioelectric control of music“, *Proc. of the IEEE ICASSP ‘88*, pp. 2556-2557, 1988.
10. L.R. Young, D. Sheena, ”Survey of eye movement recording methods“, *Behav. Res. Methods Intrum.*, pp. 397-429, 1975.
11. W. Barry, G. Melvill Jones, ”Influence of eye lid movement upon electro-oculographic recording of vertical eye movements“, *Aerospace Medicine* **36**, pp. 855-858, 1965.
12. S. Gottschalk, M.C. Lin, D. Manocha, ”OBB-Tree: A Hierarchical Structure for Rapid Interference Detection“, *Proc. of the ACM Siggraph ‘96*, pp. 171-180, 1996.
13. K.-H. Englmeier, M. Haubner, H.M. Förterer, W. Perzl, B.K. Fink, U. Fink, ”A Method for Combining Fast Surface and Exact Volume Visualization Techniques in Medicine“, *Proc. of the SPIE Medical Imaging* **2164**, pp. 22-33, 1994.
14. S. Weghorst, ”Virtual Reality in medicine“, *AI in medicine* **6**, pp. 277-279, 1994.
15. R.M. Satava, ”The role of Virtual Reality in medicine of the 21st century“, *Virtual Reality Systems* **1(2)**, pp. 43-52, 1993.