OP 2000 and MedSeC

A Concept to Extend the DICOM Standard to Include Digital Stereoscopic Video Sequences

G. Bellaire[1], D. Steines[1], G. Graschew[1], A. Thiel[2], J. Bernarding[2], T. Tolxdorff[2], P. M. Schlag[1]

[1] Surgical Research Unit OP 2000, Robert Rössle Klinik am Max-Delbrück-Centrum für Molekulare Medizin, Universitätsklinikum Charité, Medizinische Fakultät der Humboldt Universität zu Berlin,
[2] Department of Medical Informatics, University Hospital Benjamin Franklin, Free University of Berlin
Email: bellaire@rrk-berlin.de

Abstract. We present a concept for a medical video server extended by a DICOM interface. DICOM conform security mechanisms will be integrated to enable secure Internet access. Especially *d*igital *s*tereoscopic *v*ideo *s*equences (DSVS) that are relevant for surgery should be examined regarding clip length necessary for diagnosis and documentation and clip size manageable with today's hardware. Methods for DSVS compression are described, implemented and tested. DSVS are provided by a stereoscopic surgery microscope and a stereo video camera.

Keywords: Video Server, DICOM, Compression of Stereoscopic Video, MPEG-2

1 Introduction

The Surgical Research Unit OP 2000 at the Robert-Rössle-Klinik, Humboldt University Berlin, and the Department of Medical Informatics of the University Hospital Benjamin Franklin are developing a video server that integrates security mechanisms, digital monoscopic and *d*igital *s*tereoscopic *v*ideo *s*equences (DSVS) into the DICOM standard. The use of public networks such as the internet for transmission of personal patient data must be in accordance with local legislation concerning data security and integrity. The tremendous amount of data with approximately 30 MB/s for each channel can only be handled with video compression and high speed hard disk systems. The MPEG-2 encoding method should be appropriate for video compression. However real-time coding requires specialized hardware, leading to a reduction of the data stream to 6 MBit/s. To significantly reduce the data and to facilitate image management we take advantage of redundant information especially in stereoscopic video sequences.

2 MedSEC and Mallinckrodt CTN-DICOM Server

The typical heterogeneous hardware and software infrastructure within a hospital requires a standardized data exchange format and a protocol as provided by the

DICOM standard in Medicine. Unfortunately, DICOM does not yet include inherent data security concepts. This aggravates the security problem, since the use of standardized protocols and formats weakens the data security: once an intruder has gained access to the data, no special knowledge or software is required to decipher image and patient data. Due to differing national data security regulations, the connection of hospitals to public networks may vary between full connection, restricted access via firewalls, and total isolation. For a university hospital offering special medical services and international research activities, total isolation would be counterproductive and is therefore usually not supported.

The experience in transfer of medical data exposes an additional problem: the compliance of medical users requires high-speed transfer methods, especially for the exchange of image data. This restrictive condition renders time-consuming encryption methods impractical for daily use.

Goal of the MedSeC project is therefore the evaluation of a secure high speed transfer of huge amounts of data within the local high-speed network that links project partners. To investigate different speed-optimized data security concepts, we constructed a scenario with users in- and outside a hospital. To account for the heterogeneous hard- and software infrastructure, the distributed medical services had to use the DICOM 3.0 standard. For the data management (encryption, database management, data transfer activities etc.) a platform-independent user interface had to be developed.

In order to develop a server security concept that permits the storage of video data, mechanisms for the coding of video-specific data into the DICOM object are needed. This concept must also integrate a security mechanism. The structure of DICOM allows the definition of additional data elements in new data groups. These so-called "private" groups have the sole restriction that they must use odd group number. We have developed a toolkit that easily allows to modify the publicly accessible DICOM software by the Mallinckrodt Institute of Radiology [2].

3 Video Server Concept

The video server concept consists of two main components: a specialized *v*ideo *f*ile *s*erver (VFS) on the one hand, and an object-relational *d*ata*b*ase *m*anagement *s*ystem (DBMS) on the other. The DBMS is responsible for the registration and administration of video sequences stored by the VFS. This allows the integration of broadcast-quality (stereoscopic) video into a multimedia patient record in such a way that the access to information as well as specific attributes of the video data are handled by the DBMS together with other patient data. Specific attributes can be time and date of video capture, keywords describing the examination and content of the sequence, or selected frames for content based retrieval of sequences.

The VFS captures and stores video data. It communicates with the DBMS via a special interface that is realized as an extension to the latter [3]. The DBMS receives the information necessary to register newly recorded sequences from the VFS.

Client access to video data uses database requests that are forwarded to the VFS, which returns a handle to the client via the database server. Using this handle, the client now has direct access to the sequence from the VFS.

4 Compression of Stereoscopic Video

To optimize patient documentation and diagnosis the question of the adequate clip size (storage space, access time) has to be solved. The parameters determining the clip size are clip length, image resolution, frame rate and compression scheme. In the following we describe theory and first results of the compression of stereoscopic video sequences.

4.1 Theory of Video Compression using the Epipolar Constraint

Compression schemes for digital video sequences that exploit the correlation between temporally adjacent frames (e.g. MPEG [5]), suggest the exploitation of spatially adjacent still frames (i.e. left-right stereoscopic image pairs).

Three kinds of redundancies can be exploited to compress DSVS [8]. The first two have no specific relation to stereoscopy. The third takes advantage of the camera geometry, which is used to capture real live DSVS, or the geometrical constraints, which are implemented to generate stereoscopic visualizations of 3-D computer graphics (e.g. for medical video conferences [1,7]).

- Spatial correlation: Still-image compression is often based on internal predictability, meaning neighboring pixels are most likely identical or nearly identical.
- Temporal correlation: Frames in a sequence imply a frame-to-frame predictability.
- Epipolar constraint: The left/right frames of a stereoscopic image pair contain similar image areas. The known camera geometry (epipolar lines) permits efficient coding using disparity predictability.

The disparity denotes the relative offset between corresponding points in a stereoscopic image pair. [8,10,11] use variations of the basic disparity-based approach for the coding of DSVS. In a *s*tandard *c*amera *c*onfiguration (SCC, parallel epipolar lines and complanar image planes), the disparity search can be performed in a single horizontal image line [4]. An image pair can be synthesized given one image and a low resolution map of the disparity of the image pairs. *D*isparity *m*ap *e*ntries (DMEs) can be represented by few bits.

- Suppose an image (true color) represented in YCbCr-4:2:0 color space (as we use in our experiments) is coded with 12 Bits/Pixel,
- suppose the disparity does not vary significantly over eight pixels and
- suppose the horizontal disparity can be coded with 5 Bits/DME (vertical disparity with 3 Bits/DME). Depending on the underlying geometry a disparity vector consists of one (SCC) or two components.

That implies net compression factors of 58% (SCC) / 30% (one component / two components). The main disadvantage of the mentioned approach is the missing representation of occlusions.

4.2 MPEG-2 Using the Epipolar Constraint

The methods of the MPEG-2 video compression standard [5] can be adapted for stereoscopic video in the following way: The same algorithms applied in the computation of motion prediction may be used for perspective redundancies between corresponding frames in the two channels. However, in contrast to motion prediction the search space for this disparity prediction is very limited, which will reduce encoding time. The reason for this is the missing vertical offset (using SCC for imaging) and the fact that the horizontal offset can be estimated beforehand from the distance between the two lenses and the focal distance. Taking channel 1 as the main channel and encoding it in the usual manner, the frames of channel 2 can be compressed as p-frames referencing the nearest preceding i- or p-frame of channel 1. Depending on the predetermined order of p- and b-frames in the first channel, this results in the sequences shown in Figure 1.

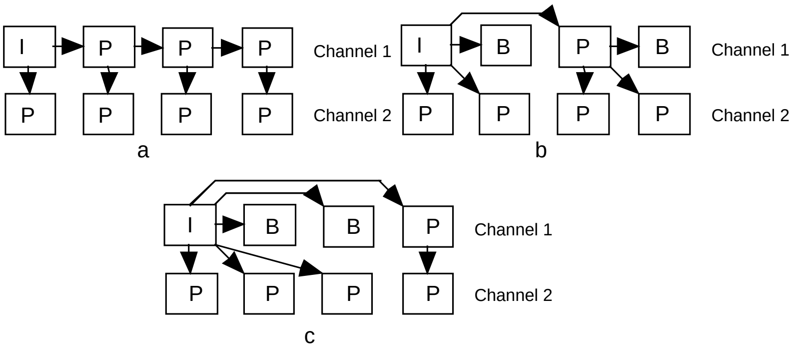

Figure 1: Coding sequence depending on channel 1 frame order;
a) IPPPP... b) IBPBPBPBPB c) IBBPBBPBBP

4.3 Experiments and Results

Some basic tests were carried out in order to determine the most appropriate method for compression of stereoscopic video. The reference MPEG-2 encoder from the University of California (Berkeley) [6] was customized to handle stereoscopic video as described in the preceding subsection. The horizontal search width for disparity prediction was used as a parameter for the experiments, whereas the vertical search width was set to zero pixels.

Seq Ch1	HSW 10 pix			HSW 30 pix			Separate compression		
	Time	Size	Ratio	Time	Size	Ratio	Time	Size	Ratio
IPPPP...	560	4710	61%	590	4812	60%	954	586	95%
IBPBP...	490	831	93%	522	836	93%	842	586	95%
IBBPBB...	473	481	96%	503	481	96%	800	586	95%

Table 1: Compression time [sec], size of compressed stereoscopic video [KB], and compression ratio depending on channel 1 frame order; HSW: Horizontal search width

A short sequence of 10 frames per channel of a phantom object (high contrast in shape and color) was captured from a Leica WILD M680 stereoscopic surgical microscope (2x 3 CCD ½’’ Camera KC233, Ikegami) on a Silicon Graphics Onyx RE II workstation equipped with two Sirius video boards. The uncompressed sequence with 720x576 pixels resolution and YCbCr-4:2:0 color format was approx. 11.9 MB in size.

Encoding time and size of compressed stereoscopic video data were measured on the Onyx (using 1 R4400 CPU) and compared to the separate compression of each channel in the conventional way. The results for the microscope sequence are shown in Table 1. Captures from another stereoscopic device (room camera Ikegami LH33, 6 CCD ¾'') showed similar results.

5 Conclusion and Outlook

The results show that making use of redundancy in stereo image pairs can reduce encoding time up to 60%. This method, however, does not improve compression ratio (up to 96%). Apparently redundancy between consecutive frames outweighs that between parallel frames.

The presented concept for a video server using a commercial database with the Mallinckrodt DICOM Server opens new perspectives for the medical documentation. The *offline* second opinion procedure for diagnosis can be simplified by secure internet access. The mentioned DSVS compression approach as well as methods using highly lossy compression for the second channel [8,9] will be medically verified (by image sources like stereoscopic laparoscope, stereoscopic surgical microscope, stereoscopic camera and artificial stereoscopic sequences) to define standards for DVSV in surgical applications.

6 References

1. Bellaire G, Graschew G, Engel-Murke F, Krauss M, Neumann P, Schlag PM: Interactive telemedicine in surgery: Fast 3-D visualization of medical volume data, Minimal Invasive Medizin:. Vol.8, no.1-2, 1997
2. CTN, DICOM Software Documentation, ftp://ftp.erl.wust1.edu/pub/dicom/software/ctn/doc
3. Informix Video Foundation DataBlade Module User's Guide, Version 1.1
4. Koschan A: A Framework for Area-Based and Feature-Based Stereo Vision. Machine GRAPHICS & VISION, Vol. 2, No. 4, 1993, pp. 285-308.
5. MPEG Draft International Standard. ISO/IEC 13818-2: Generic Coding of Moving Pictures and Associated Audio: Video, 1995
6. MPEG Software Simulation Group: mpeg2encodeVersion 1.2, ftp://mm-ftp.cs.berkeley.edu/pub/multimedia/mpeg2/software
7. Schlag PM., Rau B, Quack A, Below C, Graschew G, Meyer zur Heyde M, Rakowsky S, Engel-Murke F, Papst M, Balanos E, Breide S, Gaus H: New Technologies in Surgical Oncology. 3-D Video conference via Satellite from Berlin to Paris, CAR 96. Minimal Invasive Medizin 7: 184-186, 1996.
8. Siegel MW, Gunatilake P, Sethuraman S, Jordan AG: Compression of Stereo Image Pairs and Streams. Proc. SPIE Vol 2177, p. 258-268, Stereoscopic Displays and Virtual Reality Systems, 1994.
9. Tseng BL, Anastassiou D: Multi-Viewpoint Video Coding with MPEG-2 Compatibility. IEEE Transactions on Circuits and Systems for Video Technology, Vol.6, No.4, 1996
10. Tzovaras D, Grammalidis N, Strintzis MG: Object-Based Coding of Stereo Image Sequences using Joint 3-D Motion/Disparity Compensation, IEEE Trans. Video Technology 7(2), 1997
11. Woo O, Ortega A: Dependent Quantization for Stereo Image Coding. Proc. of PW-EI-VCIP, San Jose, USA, Jan., 1998