2011 IEEE 13th International Conference on e-Health Networking, Applications and Services

Real-Time Stereoscopic Streaming of Robotic Surgeries

Jiri Navratil, Milan Sarek, Sven Ubik
CESNET
Zikova 4, Prague 6
Czech Republic

Jiri Halak, Petr Zejdl
CESNET / CTU
Zikova 4, Prague 6 / Kolejni 550, Prague 6
Czech Republic

Pavel Peciva, Jan Schraml
Masaryk Hospital
Ústí nad Labem
Czech Republic

***Abstract*—Modern robotic surgery using stereoscopic video views of surgical elements can also enable novel E-health applications, such as remote medical students training or presentations of surgical procedures on symposia. The requirement is that such video signals are transmitted remotely in high quality and low latency. In this paper we describe our experience with real-time long distance stereoscopic transmissions of robotic surgeries using a system for low latency streaming over packet networks that we have developed.**

I. INTRODUCTION

Robotic surgery brings several advantages to modern surgery techniques - precision, smaller incisions, decreased blood loss and consequently quicker healing time.

A stereoscopic camera is used to provide the surgeon with a view of the surgical elements. The signal from this camera can also be used for E-health applications, such as remote medical students training or presentations of surgical procedures on symposia.

For such E-health applications, we need to transfer the stereoscopic high-resolution signal from the surgery robot to the audience over potentially large distances at high quality and with short latency. The later is particularly important to provide interactive experience.

In this paper we describe our system for long-distance stereoscopic transmissions of high-definition video and our experience with using this system for transmissions of surgical procedures for E-learning in medicine.

The paper is organized as follows. The technical aspects of our system are described in Section II. It includes statement of our requirements, description of interfacing to the video signal and to the network, as well as options for stereoscopic projection. The next Section III describes our practical experience with using the system for transmissions of robotic surgeries. The related work is referenced in Section IV and we draw conclusions of main contributions and future directions in Section V.

II. SYSTEM DESIGN

A. Requirements

After speaking with prospective users of the targeted application in medical field and based on our experience with real-time video streaming we set forth the following goals to be achieved by the projected system:

- Real-time transmission of high-definition stereoscopic video signals from the da Vinci Surgical System to a remote audience over unlimited distance
- Operation over a regular packet-based network infrastructure with routers and switches (not a dedicated fibre)
- Uncompressed transmission in high quality and with low latency
- Distribution to multiple audiences simultaneously
- Lightweight, easy to deploy

Majority of robotic operations are now performed using the da Vinci Surgical System [1], developed by the Intuitive Surgical company. This system consists of three main parts - a console with hand grips and 3D vision for the surgeon, the robotic arms with 3D camera and an electronics box.

Uncompressed transmissions eliminate processing delay that would be otherwise caused by compression and decompression algorithms and buffering. The total delay added by uncompressed sender and receiver devices in our system is less than 1 ms, enabling truly real-time experience.

There are three classes of devices for high-definition video transfer. The first type of devices are electro-optical convertors or extenders. They transfer the video signals over a dedicated fibre or wavelength without packetization. Therefore, the distance is limited to the nearest switch or router in the network.

Devices of the second type can transmit the video signals over a synchronous network, such as SONET/SDH, with constant bit-rate bandwidth. These devices are relatively simple, because they do not need to care about network delay variation and the receiver can derive its clock from the network payload rate. They can transmit over long distances, but not over packet-based networks, such as Ethernet, which is increasingly replacing expensive SONET/SDH network and is often the only choice in access networks.

Finally, the devices of the third class can transfer video signals to unlimited distances over the packet-based Internet. These devices need to compensate for the network delay variation and the receiver must adapt its clock rate to the incoming data. We aimed at this category when developing our system.

B. Video interface

The da Vinci surgical system provides the stereoscopic video signal on two HD-SDI [2] interfaces. HD-SDI is a commonly used interface in professional video equipment. It is a synchronous interface carrying a continuous bit rate of 1.485 Gbit/s or 2.97 Gb/s for 3G-SDI [3]. Mapping of various digital video signals into these interfaces is specified in SMPTE standards. Mapping of the HD (1920x1080) format in particular is specified in [4], [5].

A structure of a HD frame when mapped into HD-SDI is shown in Fig. 1. The blank period includes embedded audio and various synchronization and data fields. Transmission can include the complete HD-SDI data or just the active period (if audio is not required). Since we designed our system also for ultra-high-definition video transfers, such 4K (4096x2160), which uses 4 or 8 HD-SDI interfaces depending on color subsampling, we decided to transfer the active period only. One of the HD-SDI interfaces can be configured to be transferred including the blank period to carry embedded audio.

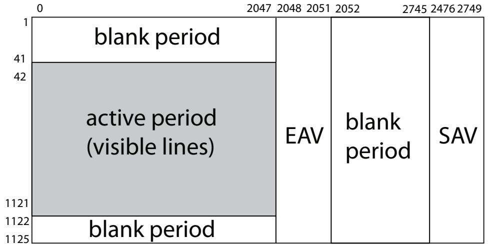

Fig. 1. HD-SDI frame format

C. Network interface and data volume

One HD channel at 25 frames per second with 10-bit color depth and 4:2:2 color subsampling requires a bit rate of 1980 x 1080 x 25 x 10 x 2 = 1.07 Gb/s. With the header overhead added (Ethernet, IP, UDP and service headers), the total bit rate is approx 1.25 Gb/s for one HD channel and 2.5 Gb/s for 3D HD transmission. When using Ethernet technology, a 10 Gb/s Ethernet network interface is required.

There are two principal options for long-distance network connectivity. The regular shared Internet and a dedicated bandwidth connection. Both categories have their advantages and limitations. The shared Internet allows to quickly make a connection between two remote users without any permissions or prior intervention of network providers. The data transmission can however be affected by increased jitter or even packet loss due to traffic concurency.

Setting up a dedicated network path currently needs a lot of coordination by all providers along a network path, who need to configure there networks. The data transmission is then guaranteed to be free of jitter and packet loss.

Backbone links in academic networks such as our network CESNET, European network GÉANT or US network Internet2 are built on DWDM (Dense Wavelength Division Multiplexing) principles, with multiple parallel wavelengths on a single optical fiber. This facilitates creation of dedicated network paths. Additionally, GLIF (Global Lambda Interconnection Facility) [6], [7] network is designed to support dedicated network paths. GLIF is a virtual organization where member institutions provide optical lambdas internationally as a facility to support data-intensive research in global dimensions.

We tested our device in both environments. Patient privacy is particularly sensitive in the case of a shared network and needs to be cared about by signing a proper agreement and using an encrypted or proprietary data format. We currently do not use encryption, but the data is in proprietary format. Therefore we use dedicate bandwidth where possible.

CESNET uses dedicated wavelengths in a DWDM system to inter-connect several hospitals in Czech Republic in a facility called POSN (Private Optical Network for Hospitals). Most other large hospitals in the country are connected over a regular IP network (Fig. 2). Universities with medical faculties are also connected by high-speed networks. Therefore, interconnection between hospitals and medical faculties is possible.

Distribution to multiple audiences simultaneously can be done either by transmission to a multicast destination IP address, if the network supports IP multicast, or by optical multicast, where the signal is split into multiple fibers along the path.

The HD-SDI signals can also be compressed, stored on a disk drive and replayed later with transmission to the audience from the record. We use a Blackmagic Design DeckLink HD Extreme 3D adapter and Media Express software for Motion JPEG 2000 compression.

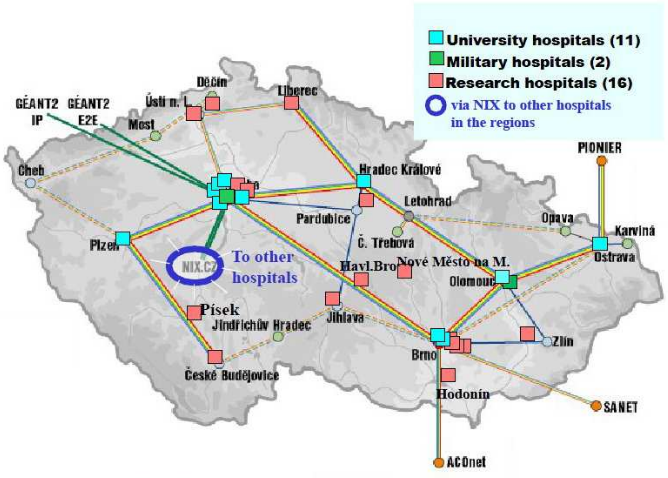

Fig. 2. Connecting hospitals to the network

D. System architecture

Real-time processing of multi-gigabit data rates is difficult on PC-based platforms with standard operating systems not designed for real-time operation. Highly parallel and pipelined real-time data processing is necessary. Therefore, hardware acceleration of data processing is needed. DSP (Digital Signal Processor) and FPGA (Field Programmable Gate Arrays) are the standard technologies in this area. We selected FPGA, due to its high data bandwidth and our design having no requirements with regard to floating-point operations.

The system architecture is shown in Fig. 3. The HD-SDI board converts electrical levels and timing between input and output HD-SDI signals on one side and the RocketIO channels of the Xilinx FPGA on the other side. The FPGA board processes the video signal and is connected to an optical transceiver, which converts electrical and optical signals for network transmission.

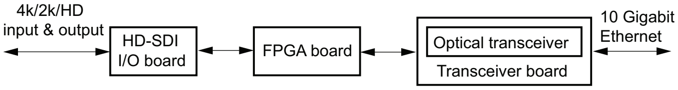

Fig. 3. System architecture

E. Receiver control

Real-time video streaming requires that the speed of rendering on the receiver side matches the rate of video source on the sender side. When the sender and receiver are connected over an asynchronous network, such as Ethernet, the receiver cannot directly synchronize its clock with the sender.

Therefore, the receiver needs to implement a technique that adjusts the speed of data arrival and maintains a continuous stream of video and audio data to the rendering device, preferably with minimal added latency for remote interactive applications. We use tunable oscillators driven by a PLL controller for this purpose [8]. Empirical experience has shown that the maximum acceptable one-way latency for remote interactive work not perceived by users as limiting factor is around 150 ms [13]. Such latency can easily be caused just by network propagation delay. Therefore, the video transfer system should add minimal further latency. Buffering of one frame at 25 frames per second adds 40 milliseconds. Tunable oscillators enable us to use subframe buffering. This together with uncompressed transmission (the compression would require pipeline processing) results in processing latency less than 1 ms added by the system to the unavoidable network propagation delay. An overall system appearance is illustrated in Fig. 4. The MVTP stands for Modular Video Transfer Platform. The device is a size of a VCR player and weighs about 5 lbs. It is easily transportable.

Fig. 4. A prototype of the MVTP transmission device

F. Stereoscopic projection

Stereoscopic projection without glasses is still not mature enough to provide sufficient quality for medical E-learning as of the end of 2010.

Stereoscopic projection with glasses can be divided into two categories. The first category are system that use two projectors, one for each eye. Each projector has a fixed polarizing filter mounted on its lens, with the polarization plane perpendicular to each other. The passive glasses use a corresponding polarization for the left and right eye, so that each eye sees the signal from one projector only. These systems usually require special silver screens, but the glasses are inexpensive, making the system suitable for large audiences. The projection system itself is however expensive, because it requires two projectors and an alignment construction. These systems can be easily connected to the receiving device by two HD-SDI to DVI converters, one for each eye. We used such a system with two JVC D-ILA projectors during the CineGrid 2010 demo described below.

The other category are systems that use one projector, which interlaces frames for the left and right eye. Active glasses are needed, changing polarization in phase with frame interlacing on the projector. Multiple communication protocols exist between glasses and projectors, the proper glasses must be assured in the projection venue. Glasses are more expensive, making this option suitable for smaller audiences. A large consumer 3D TV set can also be used. We tested our projection with Benq MP780 projector and Panasonic TX-P42GT20E TV. A conversion box is needed with two HD-SDI inputs and one HDMI output, such as BlackMagicDesign HDLinkPro 3D. Very convenient devices in this category are projectors with single lens, but two separate DVI inputs, such as Projection Design F10 AS3D. It is however rather expensive at approx. $25000.

III. PRACTICAL EXPERIENCE

The work described in this paper was implemented as part of long-term research plan of CESNET, which is NREN (National Research and Educational Network) in Czech Republic, in cooperation with the Masaryk hospital in Ústí nad Labem and KEK research institute in Tsukuba, Japan.

As our other targeted applications is remote collaboration on streaming media, we first demonstrated our system at the CineGrid 2009 workshop [1]. At this event, a stream of uncompressed 4K video was transferred in real time from Prague to the venue in San Diego over a distance of more than 10000 km. The aim was to demonstrate that such technology can enable real-time remote color grading in a distributed team and thus increase productivity. The second demonstration was during the CineGrid 2010 demo when we used the same system for real-time long-distance collaboration on processing of 3D streams.

The first demonstration of the system suitability for medical E-learning applications was a real-time tranmission of two surgical operations from the Masaryk hospital in Ústí nad Labem to the 5th International Congress of Miniinvasive and Robotic Surgery (KMRCH) [2] in Brno in October 2010. The surgery was performed by MUDr. Jan Schraml, the head of the Department of Urology (Fig. 6). The transmission setup

---

[1] http://www.cinegrid.org

[2] http://www.kmrch.cz

is illustrated in Fig. 5. The distance between the operation room and the audience was approximately 350 kilometers. The network was formed by links of the CESNET network infrastructure with the 10 Gb/s capacity. The bandwidth used was approximately 2.5 Gb/s.

The original stereoscopic HD signal in 1080i format (1920x1080 pixels, interlaced rendering) and 59.94 frames per second was transferred uncompressed. Therefore with ultimate quality and very low latency of approximately 2 ms including the network propagation delay. The stereoscopic view of the surgery is illustrated in Fig. 7. The situation in the audience room is shown in Fig. 8. The whole operation was commented by Dr. Schraml for the audience. A parallel HD videoconference was used by the audience to ask questions, which were responded by Dr. Schraml.

We used a Projection Design F10 AS3D projector. The native resolution of this projector is SXGA+ (1400x1050), therefore the picture was resampled in the projector to a slightly lower resolution.

The picture quality was subjectively approved by invited medical experts as well suitable for highly illustrative student training and presentations of surgical procedures on symposia. The commentary of the operation and the possibility to ask questions was highly appreciated.

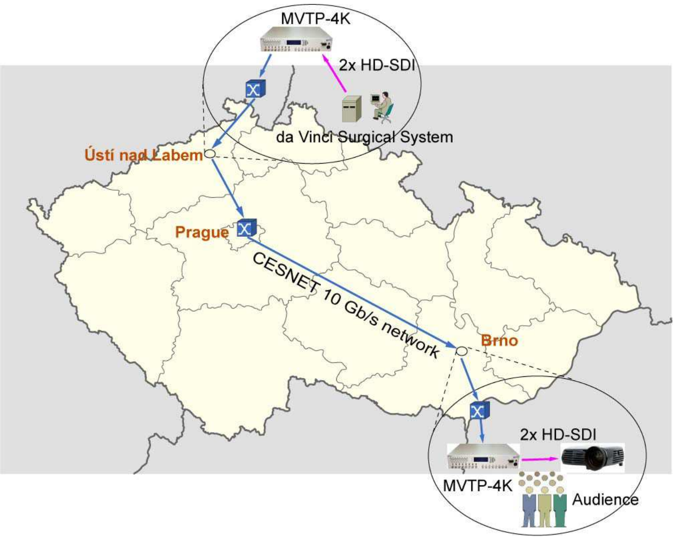

Fig. 5. Schematic diagram of the transmission to the KMRCH congress

Fig. 6. Robotic surgery in the Masaryk Hospital

Fig. 7. Remote stereoscopic view of the surgery

Fig. 8. Remote audience viewing the surgery at the KMRCH congress

Our second demonstration of the use for remote medical E-learning was a long-distance transmission from the Masaryk hospital in Ústí nad Labem in Czech Republic to the KEK research center in Tsukuba in Japan. The surgery was again performed by Dr. Schraml. The transmission setup is shown in Fig. 9. The distance between the operation room and the audience was approximately 17600 kilometers.

The network connection was formed by the CESNET network (from Ústí nad Labem to Prague), the Pan-European Géant network (from Prague to Frankfurt), the link to the MAN LAN interconnection center in New York, the link to Japan and finally the SINET3 network in Japan. The installed capacity of all links was 10 Gb/s. However, some links were already heavily loaded by 3-5 Gb/s of traffic. Therefore, we requested a Premium IP service (packet priority) in the Géant network from Dante who operates the network. The maximum Premium IP bandwidth available was 2 Gb/s. Therefore we used two synchronized scaling boxes to change the original signal 1080i/59.97 to a lower resolution 720p/50.

The audience room is shown in Fig. 10. The operation wasagain commented by Dr. Schraml for the audience and a parallel HD videoconference was used by the audience to ask questions. The networking delay was approximately 150 ms. The processing delay in the sender and receiver devices was less than 1 ms in total, therefore negligible. The audience evaluated the picture quality as excellent for remote E-learning applications.

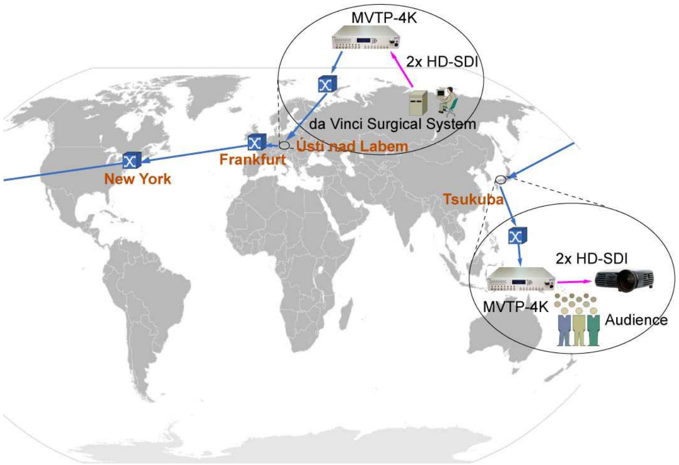

Fig. 9. Schematic diagram of the transmission from Europe to Japan

Fig. 10. Remote audience in Tsukuba viewing the surgery in real-time

IV. RELATED WORK

Net Insight's [9] Nimbra 600 series switch can transport up to eight HD-SDI or 3G-SDI channels over a synchronous SONET/SDH network. There are several commercially available solutions for transport of compressed multi-channel video over the Internet, for example NTT Electronics [10] ES8000/DS8000 4K MPEG-2 encoder/decoder complemented with NA5000 IP interface unit or intoPIX's [11] system of PRISTINE PCI-E FPGA boards and JPEG 2000 IP cores. Several companies offer products for direct electro-optical conversion of video signals for remote transmissions. The DepthQ system of the Lightspeed Design [12] is specifically designed for 3D HD transmissions from robotic surgery.

The system described in this article differs from other solutions by providing uncompressed 3D HD transmission over a long-distance packet network (rather than dedicated fibre) with very small latency.

V. CONCLUSIONS

We have demonstrated that it is possible to transfer uncompressed 3D HD video signals from the robotic surgery for remote medical E-learning applications over a long-distance packet network in real time without observable visual disturbances.

The developed system can even be used for up to four simultaneous 3D HD transmissions bidirectionally. The next planned demonstration between Europe and Asia will take place on the 2011 APAN meeting in February 2011.

The performed experiments will have significance for further developments in the field of telemedicine and E-learning. Newly presented technology is a crucial contribution to building interactive remote sites. Since the video transmission is bidirectional, the audience has an option of interaction in the process of the transmitted surgical operation and be part of virtual reality.

This technology improves the quality of teaching and training and allows interactive participation in a process that takes place in a remote or normally inaccessible place. Deployment of this transmission technology depends on deployment of new multi-gigabit networks, which might seem to be an obstacle. However, experience from the previous period has shown that developments in networking infrastructure are much faster than previously expected.

REFERENCES

[1] The da Vinci Surgical System, Intuitive Surgical, http://www.intuitivesurgical.com/products/faq/index.aspx.
[2] "1.5 Gb/s Signal/Data Serial Interface", SMPTE 292-2008, Society of Motion Picture and Television Engineers, `http://www.smpte.org`.
[3] "Television - 3 Gb/s Signal/Data Serial Interface", SMPTE 424M-2006.
[4] "Television - 1920 x 1080 Image Sample Structure, Digital Representation and Digital Timing Reference Sequences for Multiple Picture Rates", SMPTE 274M-2008.
[5] "Dual Link 1.5 Gb/s Digital Interface for 1920 x 1080 and 2048 x 1080 Picture Formats", SMPTE 372-2009.
[6] Tom DeFanti, Cees de Laat, Joe Mambretti, Kees Neggers, Bill St. Arnaud. *TransLight: a global-scale LambdaGrid for e-science*, Communications of the ACM, Vol. 46, No. 11, Nov. 2003, pp. 34-41.
[7] GLIF: Linking the world with light, Informational brochure, `http://www.glif.is/publications/info/brochure.pdf`.
[8] Jiří Halák, Michal Krsek, Felix Nevřela. Sven Ubik, Petr Žejdl, *Real-time long-distance transfer of uncompressed 4K video for remote collaboration*, Future Generation Computing Systems, Elsevier, to be published.
[9] Net Insight AB, http://www.netinsight.se.
[10] NTT Electronics, http://www.ntt-electronics.com.
[11] intoPIX, http://www.intopix.com.
[12] DepthQ System, Lightspeed Design, http://www.depthq.com.
[13] ITU-T Recommendation G.114 - One-way transmission time, ITU-T Study Group 12, May 2003.

AUTHORS

Jiří Halák received his MSc. at the Czech Technical University and is working towards his PhD in Computer Science. He is also working as a researcher in CESNET, particularly in the field of programmable hardware.

Jiří Navrátil received his PhD in Computer Science from the Czech Technical University in Prague in 1984 for his work on analysis of timesharing systems. He has worked for 30 years at the Computing and Information Center of CTU Prague in High Performance Computing and Communications research. Since 2006 he is with CESNET in the group supporting E2E applications.

Pavel Pečiva works in the media applications for nearly 20 years. Since 2001 he works in MNUL Ústí nad Labem as technical support for education center of the hospital.

Milan Šárek studied computer science (PhD in 1990) at the Technical University in Brno. He has worked as a programmer, teacher and researcher in the field of medical informatics and computer networks. Since 2004 he is with CESNET.

Sven Ubik received his MSc. and Dr. in Computer Science from the Czech Technical University in 1990 and 1998, respectively. He is currently with the Research and development department of CESNET. His research interests include network monitoring, high-definition video, programmable hardware and optical networks. He is Cisco CCIE #14053.

Petr Žejdl received his MSc. at the Czech Technical University and is working towards his PhD in Computer Science. He is also working as a researcher in CESNET, particularly in the field of programmable hardware.

ACKNOWLEDGEMENT

CESNET work was supported by the research plan Optical National Research Network and its New Applications (MSM6383917201, 2004-2010) by the Czech Ministry of Education, Youth and Sports.

Special thanks belong to the team in the Masaryk Hospital.