



HAL
open science

Introduction to DICOM-RTV: a new standard for real-time video communication in hospitals

Saad El Jaouhari, Bernard Gibaud, Philippe Lemonnier, Guillaume Pasquier,
Eric Poiseau, Eric Guiffard, Patrick Hardy, Emmanuel Cordonnier

► **To cite this version:**

Saad El Jaouhari, Bernard Gibaud, Philippe Lemonnier, Guillaume Pasquier, Eric Poiseau, et al.. Introduction to DICOM-RTV: a new standard for real-time video communication in hospitals. Healthcom 2019 - IEEE International Conference on E-health Networking, Application & Services, Oct 2019, Bogotá, Colombia. pp.1-6. hal-02431873

HAL Id: hal-02431873

<https://hal.science/hal-02431873>

Submitted on 8 Jan 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Introduction to DICOM-RTV: a new standard for real-time video communication in hospitals

Saad EL JAOUHARI¹, Bernard GIBAUD², Philippe LEMONNIER¹, Guillaume PASQUIER¹,
Eric POISEAU³, Eric GUIFFARD¹, Patrick HARDY⁴ and Emmanuel CORDONNIER¹

¹B<>com, Cesson Sévigné, France, ²Univ Rennes, Inserm, LTSI - UMR 1099, F-35000 Rennes, France

³INRIA Rennes, Equipe SED, Rennes, France, ⁴Harmonic, Cesson Sévigné, France

Emails: ¹FirstName.LastName@b-com.com, ²bernard.gibaud@univ-rennes1.fr,

³eric.poiseau@inria.fr ⁴patrick.hardy@harmonicinc.com

Abstract—In order to overcome the challenges of managing real-time transfer of video, and/or audio, and associated medical metadata inside the medical theaters (e.g., operating room), a new DICOM communication service standard emerged. Its main objective is to deliver synchronized videos (potentially synchronized with their corresponding metadata) in real-time to surgeons during a surgery inside an operating room. Moreover, it allows, on one side, the transmission of real-time videos to subscribers with a quality of service comparable to the one inside the operating room, and on the other side, provides a standard that will allow the interoperability between the different medical equipment that produce/consume the media essences. This new DICOM extension is called DICOM Real-Time Video (DICOM-RTV). This paper is an introduction to this extension. It mainly presents the different challenges solved by this extension, illustrates it with relevant use cases, and provides the global architecture of the DICOM-RTV system.

Index Terms—DICOM, DICOM-RTV, Real-Time, Video, Metadata, Operating Room, Healthcare, Synchronization.

I. INTRODUCTION

DICOM, which stands for Digital Imaging and Communications in Medicine, is considered as the backbone of the image display in modern medicine, and currently the driving force behind the imaging workflow in hospitals around the world. Its main objective is to provide standards to manage all the chain from the medical image acquisition, the transfer, the processing and the interpretation, to the storage, in order to facilitate the interoperability between the different medical imaging equipment. Moreover, all the PACS (Picture Archiving and Communication System), which comprise the acquisition modalities, the storage and the viewing workstations, are implementing the DICOM standard. Additionally, DICOM is recognized as a standard by CEN (European Committee for Standardization) and ISO (International Organization for Standardization).

However, currently the standard does not handle the real-time video, and/or audio, and the associated metadata, captured from the medical theaters such as the Operating Room (OR). Generally, proprietary SDI/SDI over IP based solutions are being used to handle the real-time video. However, it breaks the interoperability objective of the DICOM standards because they neither transfer the corresponding medical metadata nor

enable the synchronization of multiple videos.

Hence, the objective of this paper is to introduce a DICOM standard called DICOM Real-Time Video (DICOM-RTV), prepared by DICOM Working Group 13 "Visible Light" and described in detail in the Supplement 202 [1]. The Supplement has been approved and in its final steps for publication. It defines a new IP-based DICOM service for broadcasting real-time video, together with the corresponding metadata in a synchronized way, to subscribers (e.g. display monitor), with a quality of service comparable to the one obtained via video cables. Moreover, by allowing the different devices inside the OR to communicate using standardized way (i.e. DICOM-RTV), the interoperability issue can be overcome. Finally, it upgrades the use of the SDI-based system in the OR by the IP-based one.

The proposed solution is based on the SMPTE ST 2110 family of standards, a widely used in the professional video (e.g., TV studios) equipment. It uses mainly the communication architecture specified in SMPTE ST 2110-10 [2]. As a first step, the current version of DICOM-RTV is restricted to the real-time streaming of uncompressed video. The underlying standards (the SMPTE ST2110 family) have first coped with uncompressed video, and are now addressing different flavors of compressed video in a consistent way, with standard supplements currently in the works. The proposed mechanism for conveying medical metadata along with the video (and audio) is fully compatible with this approach and it is anticipated that DICOM-RTV will be extended to support the upcoming compressed video essence transport schemes.

Furthermore, the proposed Supplement allows the transmission of several flows of data (e.g. video flows, video/audio flows, and medical data flows) in the same medium, by taking advantage of the multiplexing nature of IP, with a special focus on the medical metadata which is required to ensure the patient's safety (e.g., operating the right patient) and enabling advanced applications using video and data (e.g., image guided surgery).

This paper describes the main challenges, the choices that

have been made and an overview of the proposed solution to overcome these challenges. Hence, it is structured as follows: in section II, the main challenges solved by DICOM-RTV are listed. Then, a related work section III is presented, which discusses the literature works that deal with the real-time medical videos, mainly inside a medical infrastructure such as a hospital. Illustrative use cases are explained in section IV. Next, the proposed architecture is provided in section V, followed by a discussion and future challenges in section VI. Finally, the paper is concluded in section VII.

II. CHALLENGES

The DICOM-RTV Supplement is a response to the multiple challenges related to the handling of the real-time transmission and displaying of multimedia flow and the medical metadata of the patient being operated inside the OR.

Currently most of the ORs use solutions to handle the multimedia flow either via a cabled video infrastructure (SDI/DVI) or via proprietary video-over-IP requiring optical fibers, with often specific connectors. Additionally, very few solutions enable the transport of metadata, and if supported a proprietary format is used. Moreover, most of these solutions are not interoperable and very specific to the facility where they are being deployed. Hence, a first challenge is to be able handle such flow and to be interoperable with the different multimedia sources (e.g. endoscopes, cameras, etc.) and displaying devices (i.e. the different monitors in the OR).

The second challenge is the synchronization. It concerns both the synchronization between a video and a metadata flow, and the synchronization between two video flows. The currently proposed solution does not handle such synchronization since they are either SDI based with no synchronization mechanism or SDI over IP based claiming that very low latency implies an intrinsic synchronization of flows.

The first case concerns the synchronization between the real-time video and the real-time medical metadata gathered from the medical equipment which produces some metadata (e.g., parameters of the beam for ultrasound) or uses them for safety (e.g., patient's demographics shown over images on the device screen) in the OR. For instance, by synchronizing the real-time video flow of an laparoscope with its 3D position inside the patient, or with the gas pressure value inside the abdomen of the patient being operated during a laparoscopy in order to be able to better monitor the patient. The idea is to display both flows in one of the monitors present in the OR in order to provide the doctor with more visibility. As for the second case, the goal of synchronizing two video flows is to enhance the physicians precision during the surgery thanks to per-operative Augmented Reality application (e.g., for helping needle insertion in ultrasound guided biopsy).

Furthermore, with the generalization of image guided procedures, hybrid rooms are more and more used, combining

video based images (e.g., endoscopy) with radiology (e.g., C-Arms) in the same OR, the two solutions being provided by two different vendors having their own way of managing video.

Thus, the objective of the DICOM-RTV is to solve these challenges and to provide an interoperable solution that can be used with any equipment inside the OR.

III. RELATED WORK

In this section, a literature review regarding the solutions that manage real-time medical videos in the OR is presented. It shows that in last decade only a few works have been done in order to solve these challenges. In [3] the authors propose a solution to store real-time X-ray images of a patient's organ structure from a fluoroscope, as a part of the patient medical record. The video is captured, digitalized, converted and compressed in an MPEG2 format and integrated into the PACS in a DICOM file format. The proposed solution does not deal with the synchronization and the real-time display of the video. Another work presented in [4] takes advantage of the image assisted surgical treatment and developed a multimedia electronic patient record system that manages and distributes the real-time multimodality imaging and informatics in order to assist the surgeon during the different clinical phases of the operation (spinal surgery). They used images, clinical forms, waveforms, textual data for planning the surgery, two real-time imaging techniques, digital fluoroscopic and endoscopic video images, and more than half a dozen live vital signs of the patient during surgery in order to assist and monitor the surgery. All these data have to be acquired, displayed and archived in real-time. The acquired data are synchronized using a master clock and displayed live into a customized interface using an LCD screen in the OR. This paper demonstrates that from the technical point of view, what DICOM-RTV enables has been accessible for a while. But because OR is a multiple providers infrastructure, no such system have been deployed on the clinical side since. Only a urbanist approach, creating commonly adopted interoperability standards can make it possible to enable OR to go to this level of performance. In [5] the authors present a development of a radio channel model for an angiography room in a hospital. They use wireless communications in the 60 GHz band to send video data from an X-ray machine to a movable LCD screen in the OR. This paper is anticipating the next step of the OR, integrating wireless communications for flexibility (e.g., introducing non planned equipment during the procedure). DICOM-RTV will enable such evolution when compressed video is supported on the one hand, and when wireless network manages prioritization among the flows (slicing) and ultra-low latency, both supported by the 5G.

Other works, such as [6] [7] [8] [9], focus more on the telemedicine and the real-time collaboration around medical imaging, and not necessarily from the OR. The solutions

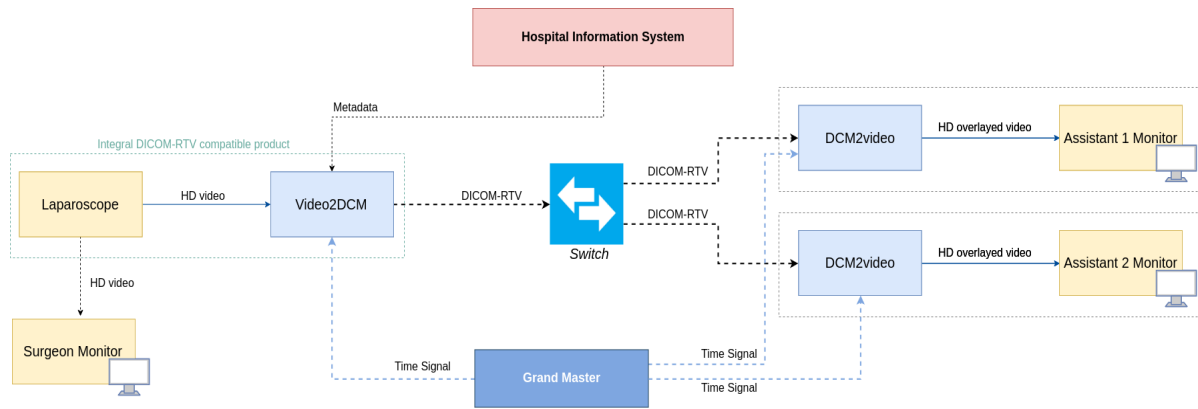


Fig. 1. Duplicating on additional monitors

propose real-time visio-conferencing while sharing, and discussing DICOM files (e.g. radiological images, Magnetic Resonance Imaging (MRI), Computed Tomography (CT), etc.). They do not provide solutions to manage the real-time videos in the OR as targeted by DICOM-RTV and do not provide solutions to synchronize the medical DICOM metadata with the video real-time flow. Tele-mentoring, compatible with lower quality video and higher latency (typically 500ms) is not the main use case targeted by DICOM-RTV in the first stage. Communication inside the OR with high quality video and low latency is the primary target of DICOM-RTV.

IV. USE CASES

In order to illustrate the interest of the proposed solution in the medical domain, and especially in the OR, several use cases are described in the informative annex of the DICOM-RTV Supplement.

A. Video duplication

During a surgery, several operators may be participating (i.e. surgeons, assistants, nurses, etc.). In this case, the display of the control images needs to be done in a way that they are displayed in front of each one of them. The purpose is to allow them to have an optimal view of the surgery and to allow them to anticipate the actions (e.g. provide instruments). Thus, the solution is to duplicate the images in other monitors. In this case a primary monitor is directly connected to the imaging system, and the secondary monitors display the duplicated images, as shown in Fig. 1.

The live video image has to be transferred to additional monitors with a minimal latency, without modifying the image itself (resolution, format, etc.). The latency between the two monitors should be compatible with collaborative activity for surgery where the surgeon is, for example, operating based on the primary monitor and the assistant is controlling the endoscope based on the secondary monitor. Equipments are synchronized with the Grand Master (which is the reference clock, using the IEEE 1588 standard, to lock

all the device clocks of the system).

As shown in Fig. 1, duplicating the video in a secondary monitor goes through the following steps:

- The video is gathered from a laparoscope, which provides real-time HD video. The laparoscope either contains the DICOM-RTV conversion capability, which converts the video to the DICOM format, via the video2DCM converter integrated in the product (as presented in Fig. 1), or sends the HD video to an external DICOM converter (video2DCM).
- An overlay can be added by the laparoscope to the HD video using metadata from an information system such as the Surgery Department Information System.
- The result is displayed in the surgeon monitor, which is the primary monitor.
- Then, in order to duplicate the video on a secondary monitor (i.e. assistant monitor), each video-frame and metadata, are timestamped at a common instant, as defined in the standard ST 2059-1 [10]. The latter provides a method of synchronization over IP networks.
- The resulting DICOM-RTV video is then transmitted through an IP switch to a destination DICOM-RTV converter (DCM2video), which uses the timestamp information of each flow, in order to re-synchronize the video, and the metadata (if any), convert it again from a DICOM-RTV format to a video format and then send it to the assistant monitor.

B. Merging two synchronized videos

Today, more and more doctors are using technological applications in order to help them perform safer and less invasive procedures during a surgery. Such applications are part of the Image Guided Surgery (IGS) domain. Thus, the objective of this use case is to use an Augmented Reality (AR) application in order to enrich live images captured from the OR by adding information as an overlay, which can be a 3D display of patient anatomy reconstructed from pre-operative MRI or CT scans, or projections of real-time intra-operative medical imaging (3D ultrasound typically).

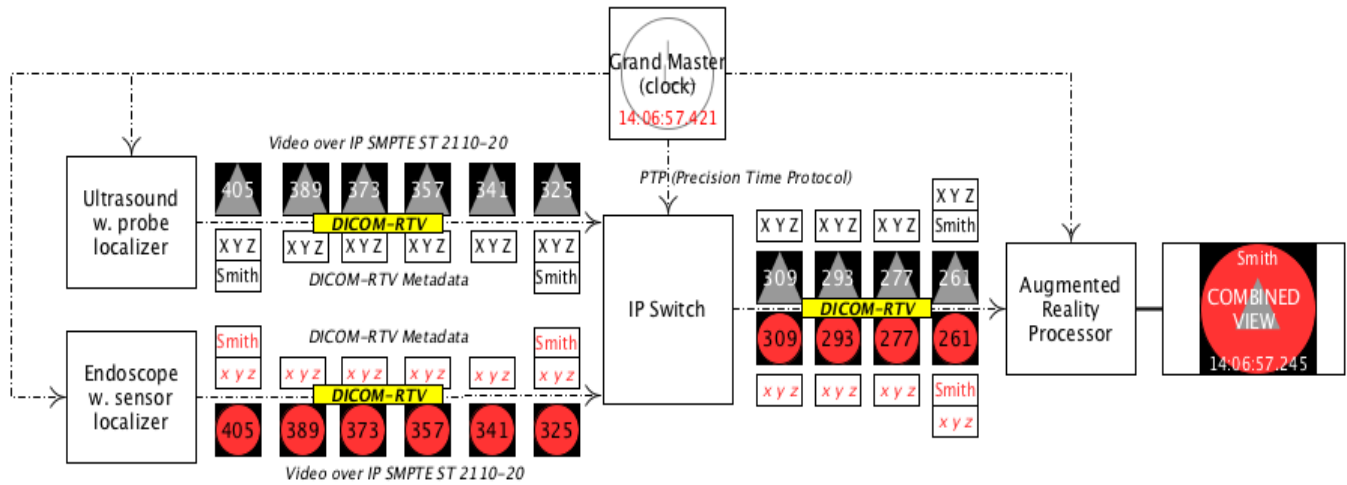


Fig. 2. Synchronizing and fusing two videos (endoscope and ultrasound), with the corresponding metadata (3D position), captured from an OR into a single video, with an estimated global latency of 176ms at the reception of the video on a monitor. The value is computed from the difference between the timestamp added to the image at 14:06:57 “245” ms and the time at the reception of the image retrieved from the same reference clock (Grand Master) 14:06:57 “421” ms, as in the example

For instance, the two video sources (e.g. an ultrasound and an endoscope), are first converted to the DICOM-RTV format (i.e. Video over IP SMPTE ST 2110-20 [11]), then overlaid with DICOM-RTV metadata (e.g. the 3D position of the ultrasound probe), timestamped, and finally sent over the network as shown in Fig. 2.

The receiver, which is also connected to the same network, captures the two videos and processes them via an “augmented reality processor” which synchronizes the two video frame-by-frame using their timestamps, and then sends the result to a displayer. The result shows a real-time combination image merging the primary live images (i.e. endoscopy) and the real-time secondary live images (i.e. ultrasound). Moreover, the result can be also exported as a new video source, through the DICOM-RTV protocol, to be displayed to the surgeon.

In this case, all video streams have to be transferred with an ultra-low latency and a very strict synchronization. Metadata associated with the video has to be updated at the frame rate (e.g., the 3D position of the US probe). The mechanisms used for generating augmented reality views or to detect and follow 3D position of devices are out of scope of this standard, only the method for conveying the multiple synchronized video/multi-frame sources along with the parameters, that may change at every frame, is specified [1].

Other interesting use cases are described, in the DICOM-RTV Supplement [1], such as recording and replaying the procedures and the synchronized videos for review by a senior.

V. PROPOSED ARCHITECTURE

As shown on Fig. 3, the DICOM-RTV communication is used to connect various video or multi-frame sources to various

destinations, through a standard IP switch, instead of using a video switch. In the future, the equipment producing video will support DICOM-RTV natively, but it is anticipated that the first implementations will rely on the use of converters to create a DICOM-RTV stream from the video stream (e.g., SDI) and associated metadata coming from information systems, through existing mechanisms (e.g., DICOM Worklist). Such converters have to be synchronized with the Grand Master which is delivering a very precise universal time. Similarly, the video receivers (e.g., monitors) are connected to the central switch via a converter which also has to be synchronized via the Grand Master. The different DICOM-RTV streams can be displayed, recorded, converted or combined together for different use cases. The medical metadata in the DICOM-RTV streams can be used to improve the quality of the whole system, as explained in the previous use cases. Such metadata can be for instance the insufflation parameters over the displayed laparoscopic video or, in the longer term, to display the procedure steps automatically detected from the video by AI techniques, to select the more relevant layout on the mosaic display monitor. The role each component in Fig. 3 is the following:

- **Operation Scheduler:** provides the patient’s metadata and related studies information through mechanisms such as DICOM Worklist, which contains the list of surgeries planned for a certain period of time, or HL7 ORM (Health Level Seven International General Order Message), containing order information for a certain procedure.
- **Medical video recorder:** allows the recording of the DICOM-RTV video flow in a specific format (similar to Material eXchange Format (MXF) for broadcast video). The recorded video can be for instance reviewed later by a senior or be source of excerpts of the procedure to be populated

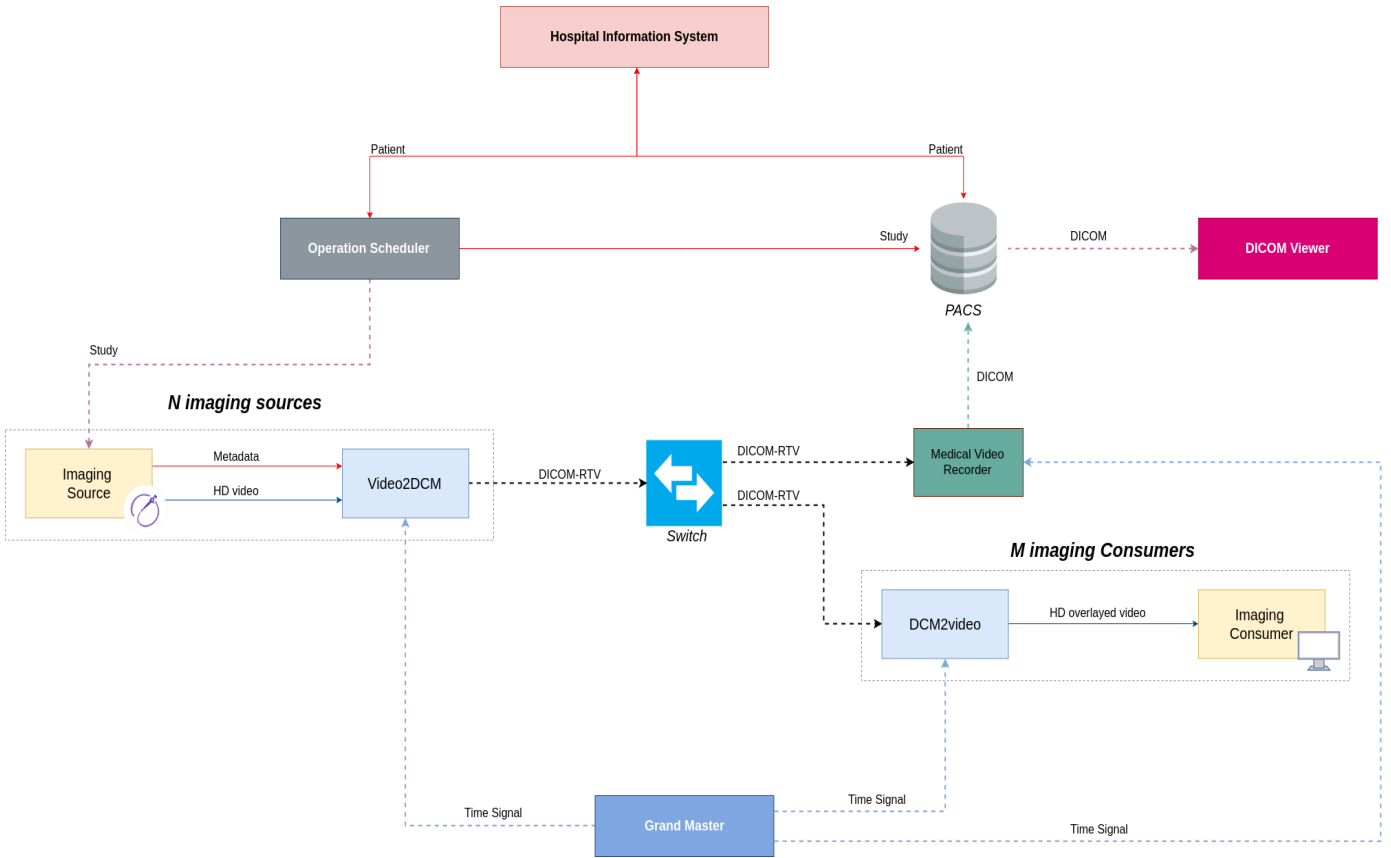


Fig. 3. Overview diagram of the operating room

- in the patients medical record, or just for safety reasons.
- **Video2DCM**: converts the RAW video to a DICOM-RTV video format (SMPTE ST 2110-20).
- **DCM2video**: convert the DICOM-RTV video format to a RAW video format.
- **Grand Master**: is a reference clock, implementing the Precision Time Protocol (PTP), which main role is to lock the clock of the DICOM-RTV equipment.
- **Medical Information System**: provides the metadata of the queried patient.
- **PACS**: allows the management of the medical images using archiving functionalities.
- **DICOM viewer**: allows the display of the medical images retrieved from the PACS.

As shown on Fig. 4, the DICOM Real-Time Video stream is comprised of typically three different flows (essences) for respectively video, audio and medical metadata information, using the intrinsic capability of IP to convey different flows on the same medium, multiplexing three kinds of blocks. There will be thousands of blocks for each video frame, hundreds for each audio sample and one for the medical metadata associated to each video frame, respectively represented as V (video), A (audio) and M (metadata) on the Fig. 5, which is the network view of the real-time streaming.

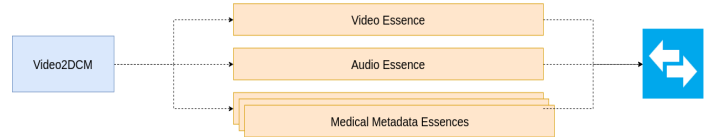


Fig. 4. Real-Time Video stream content overview

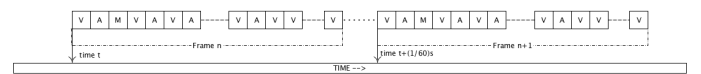


Fig. 5. Real-Time Video transmission details

VI. DISCUSSION AND FUTURE CHALLENGES

A real-live demonstration of the use of the DICOM-RTV standard and for a first time in the world, was made in collaboration with a medical team in the CHU (Centre Hospitalier Universitaire - University Hospital) of Rennes, France, during five photoselective vaporization of the prostate, a procedure used in the treatment of prostatic hyperplasia. The solution allowed the retransmission of medical images from different equipment in the OR in a perfectly synchronized way. DICOM-RTV allowed the display of two synchronized videos, one from a camera recording the hand gestures of the surgeon, and another one from an endoscope.

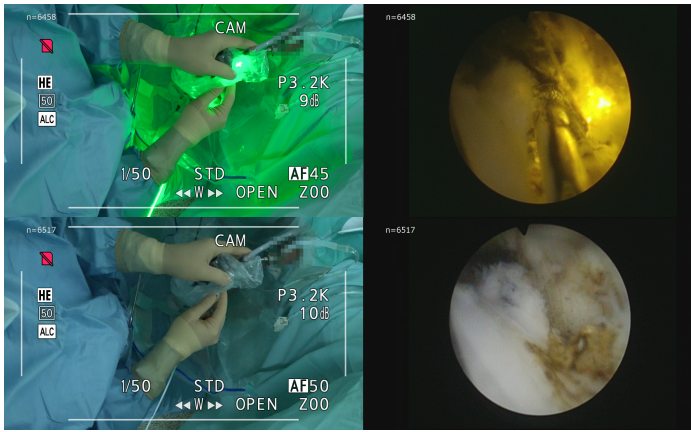


Fig. 6. Synchronized real-time videos using DICOM-RTV captured from the operating room during a surgery

The Fig. 6 is a screenshot of two synchronized frames of the real-time video, where the light in the right top video corresponds to the instant when the surgeon uses the laser. The added value of DICOM-RTV in this kind of procedure is to improve the surgical gestures accuracy and to share precise, real-time and synchronized videos with the different operators during a surgery, toward the next generation ORs.

The DICOM Standard Committee avoids reinventing the wheel and re-uses existing standards whenever they are relevant and applicable. Because of the quality requirement inside the OR, standards used in the professional video (e.g., TV studios) have been analyzed. Among several candidates, including SMPTE ST 2022-6 conveying SDI over IP, the SMPTE ST 2110 family of standards was chosen for several reasons: (i) it allows conveying separately video, audio and metadata, while synchronizing the flows, using a common architecture that can be re-used for other types of flow (e.g., medical metadata); (ii) it was rapidly adopted by users (e.g., TV studios) and vendors (e.g., video converters) and tested in test sessions for checking the actual interoperability; (iii) it conveys uncompressed high-quality video with possibility to define the resolution and the frame rate and compressed video transmission is in the roadmap. As a result of this choice, an informal link has been established between DICOM and VSF/SMPTE organizations, both being informed on the progress of the other. The pending SMPTE ST 2110-41 which aims to deal with fast metadata transmission, is for example elaborated while taking into account DICOM-RTV as an example of implementation (and the DICOM-RTV Supplement introduced a 128-byte preamble at the beginning of the payload to enable insertion of ST 2110-41 attributes, similarly to what was done for managing DICOM Files in DICOM Part 10).

The next challenges are to convince the community to adopt DICOM-RTV, by providing Proofs of Concepts (PoC) and performance results, and to extend DICOM-RTV to other kinds of sources (e.g., ultrasound, fluoroscope, operating microscope)

and dynamic metadata (e.g., insufflation parameters).

VII. CONCLUSION

By taking advantage of the SMPTE ST 2110, DICOM-RTV promises to deliver synchronized and high quality videos, together with their corresponding metadata, in real-time to the physicians (in particular the surgeons) and their assistants, inside the OR. The main objective is to enable a better management of imaging information, which impacts directly the patient safety and quality of care provided to the patient. Thus, this paper introduced the DICOM-RTV standard, and presented the main challenges that solves, together with some illustrative use cases and the global architecture of the DICOM-RTV system inside a medical infrastructure such as a hospital. It also discussed the different taken choices and the first world demonstration that was organized in the Rennes's university hospital. The latter represents a PoC of the DICOM-RTV Supplement.

ACKNOWLEDGMENT

Special thanks to DICOM WG13 and WG6 members for their contribution to the supplement 202 authoring and review.

REFERENCES

- [1] "Digital Imaging and Communications in Medicine (DICOM) Supplement 202: Real-Time Video," Jan. 2019.
- [2] "ST 2110-10:2017 - SMPTE Standard - Professional Media Over Managed IP Networks: System Timing and Definitions," *ST 2110-10:2017*, pp. 1–17, Nov 2017.
- [3] C. Mei-Yen, L. Chien-Tsai, S. Ying-Chou, L. Rheun-Chuan, C. Chian-Fa, M. T. Mu-Huo, and C. Cheng-Yen, "Design and Evaluation of a DICOM compliant Video-fluoroscopy Imaging System," in *2007 9th International Conference on e-Health Networking, Application and Services*, June 2007, pp. 248–251.
- [4] Documet, Jorge and Le, Anh and Liu, Brent and Chiu, John and Huang, H. K., "A multimedia electronic patient record (ePR) system for image-assisted minimally invasive spinal surgery," *International Journal of Computer Assisted Radiology and Surgery*, vol. 5, no. 3, pp. 195–209, May 2010. [Online]. Available: <https://doi.org/10.1007/s11548-009-0387-x>
- [5] M. Kyr, J. Simola, K. Haneda, K. Takizawa, H. Hagiwara, and P. Vainikainen, "Development of a channel model for 60 GHz radio systems in an angiography room," in *2011 IEEE Topical Conference on Biomedical Wireless Technologies, Networks, and Sensing Systems*, Jan 2011, pp. 39–42.
- [6] Li Ling, Yao Dezhong, Li Jianqing, Li Bin, and Wang Ling, "A multimedia telemedicine system," in *2005 IEEE Engineering in Medicine and Biology 27th Annual Conference*, Jan 2005, pp. 3746–3748.
- [7] C. Andrikos, G. Rassias, P. Tsanakas, and I. Maglogiannis, "An Enhanced Device-Transparent Real-Time Teleconsultation Environment for Radiologists," *IEEE Journal of Biomedical and Health Informatics*, vol. 23, no. 1, pp. 374–386, Jan 2019.
- [8] A. Puel, A. v. Wangenheim, M. I. Meurer, and D. D. J. d. Macedo, "BUCOMAX: Collaborative Multimedia Platform for Real Time Manipulation and Visualization of Bucomaxillofacial Diagnostic Images," in *2014 IEEE 27th International Symposium on Computer-Based Medical Systems*, May 2014, pp. 392–395.
- [9] K. A. Banitsas, P. Georgiadis, S. Tachakra, and D. Cavouras, "Using handheld devices for real-time wireless teleconsultation," in *The 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, vol. 2, Sep. 2004, pp. 3105–3108.
- [10] "ST 2059-1:2015 - SMPTE Standard - Generation and Alignment of Interface Signals to the SMPTE Epoch," *ST 2059-1:2015*, pp. 1–31, April 2015.
- [11] "ST 2110-20:2017 - SMPTE Standard - Professional Media Over Managed IP Networks: Uncompressed Active Video," *ST 2110-20:2017*, pp. 1–22, Nov 2017.