



Architecture for Augmented Reality Guidance Applications Using DICOM-RTV

Maxime Guinin¹, Guillaume Pasquier², Saad El Jaouhari², Patrick Hardy³, Julien Waechter¹, Flavien Bridault⁴, Alexandre Hostettler⁴, Jacques Marescaux⁴

¹ IHU Strasbourg, France

² bcom, Rennes, France

³ Harmonic, Rennes, France

⁴ IRCAD, Strasbourg, France

maxime.guinin@ihu-strasbourg.eu

Abstract

The recently published DICOM-Real Time Video (DICOM-RTV) standard allows real-time synchronization of transported video and medical metadata flows. In order to demonstrate the advantages of this synchronization by design, a demonstration of a DICOM-RTV Augmented Reality (AR) based solution is proposed. This solution consists in facilitating the surgeon's gesture by monitoring the positioning of the biopsy needle under ultrasound guidance and constructing virtual and augmented reality views of the needle's trajectory. In this paper, the platform architecture and the obtained results are presented. Furthermore, a comparison between our proposed solution and a setup that do not use DICOM-RTV is provided in order to show the difference in term of precision, synchronization and delay. The obtained results validate the contribution of real-time synchronization to improve the accuracy of surgical gesture aid applications such as AR.

1 Introduction

Published in September 2019, the DICOM-Real Time Video (DICOM-RTV) standard (Supplement 202: Real-Time Video, 2019) (El Jaouhari, et al., 2019) defines a new IP-based DICOM service for broadcasting real-time video and corresponding metadata in a synchronized way with a quality of service comparable to the one obtained via video cables. Moreover, by allowing the different devices inside the Operating Room (OR) to communicate using standardized way (i.e. DICOM-RTV), the interoperability issue can be overcome. In addition, this standard allows the transmission of several flows of data (e.g. video flows, video/audio flows, and medical data flows) in the same medium,

enabling advanced applications using video and data (e.g. image guided surgery). A special focus on the medical metadata is required to ensure patient safety.

On the other hand, medical AR is of growing interest in the OR for applications such as orthopedic surgery. It takes its main motivation from the necessity of providing medical imaging data and other patient information superimposed on the operation area in the most intuitive way. This requires real-time on-site visualization of heterogeneous data, and was one of the main goals of many medical solutions proposed in the literature. (Sutherland, 1968) proposed a tracked head-mounted display as a novel human-computer interface enabling viewpoint-dependent visualization of virtual objects. Later (Roberts, Strohbehn, Hatch, Murray, & Kettenberger, 1986) implemented the first medical AR system. Recently (Wen, Tay, Nguyen, Chng, & Chui, 2014), proposed a cooperative surgical system, guided by hand gestures and supported by an AR based surgical field.

In line with these issues, our research focused on an AR solution combining multiple DICOM-RTV streams. We started with a liver surgery AR application requiring a live ultrasound scan and camera. The use case consists in facilitating the clinician's gesture performing a tumor ablation under ultrasound (US) guidance. The developed solution aims to monitor the tools position (i.e. the needle and the US probe), construct virtual and AR views, visualize the needle's trajectory, the intersection between this trajectory and the ultrasound plane and the distance remaining to reach the target. The addition of a device for locating the tools is mandatory to track their respective positions. The position data, captured in real time, are sent through the medical metadata streams associated with the ultrasound and camera video. Thus, this paper demonstrates an AR application assisting the surgeon during a needle insertion manipulation using an accurate spatial and temporal image registration between the US imaging with the overhead camera video.

2 Materials & Methods

The overall system architecture is presented in Figure 1.

In this demo, the main data provider equipment used are:

- A US scanner (which captures an image of depth on a liver phantom), and constitutes the first video stream;
- An overhead camera constitutes the second video stream;
- An electromagnetic (EM) positioning device from trakSTAR of NDI (i.e., EM Tracking);

Two AR applications have been evaluated, as in Figure 1:

- One using Network Time Protocol (NTP) to synchronize the clock of the different devices (AR application 1)
- One using the DICOM-RTV standard requiring Precision Time Protocol (PTP) to synchronize the clock (AR application 2).

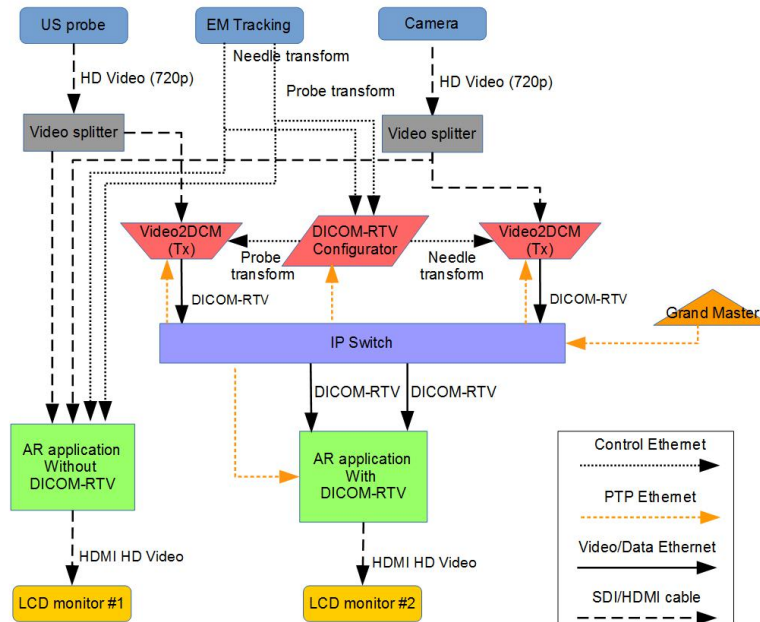


Figure 1: AR system architecture

They perform the same task, calculating the AR scene by merging the two incident images of two videos by using the associated dynamic metadata.

The AR application 1 takes directly as an input the two videos flows via SDI and their associated metadata via IP and shows the result on a LCD monitor (i.e., LCD monitor #0). The AR application 2 exploits as inputs two standardized DICOM-RTV components, each including a video flow and a metadata flow. The DICOM-RTV streams are provided over IP by dedicated transmitter device which operates the synchronization of the incoming components. The result is shown on a second LCD monitor (i.e., LCD monitor #2). For this second application, a set-up is deployed in order to encode and decode the two videos. It consists of:

- Two Transmitters (Tx), each mainly encapsulates the HD video and its associated metadata to the DICOM-RTV format (i.e., Video2DCM).
- A Receiver (Rx), to decapsulate the two DICOM-RTV streams and to feed the AR application.
- A DICOM-RTV Configurator, which configures the different DICOM-RTV devices, processes the different metadata (i.e. probe transform, needle transform) and distributes them to the corresponding transmitter.
- A Grand Master to distribute a master clock using PTP to synchronize the different clocks of the DICOM-RTV devices.

3 Results

After installation of the system, in the presence of fast and large movements (around 1 m/s), as shown in Figure 2, a significant gain is observed in the representation of needle movements on AR images (around 2 cm). In NTP mode, a bias leads to alignment errors in the images constituting the AR scene. The bias is due to the lower precision brought by the NTP protocol (few ms) and to the drift of the video frequency (typical accuracy of a video source frequency is ± 50 ppm). By consequence, the surgeon's gesture speed must be reduced to minimize the visual effect.

In DICOM-RTV mode, on the right side of the figure, the two needles images are perfectly synchronized in the same condition.



Figure 2: Demonstrating the interest of using DICOM-RTV in the AR-based application. NTP mode on the left side, using the DICOM-RTV protocol on the right side

4 Conclusion

The purpose of this setup was to show the advantages of the DICOM-RTV standard to produce augmented reality and reconstructed virtual reality images. As planned, better results were obtained using the AR application 2 based on DICOM-RTV and confirmed the interest of this new standard.

In AR application 1, the lack of synchronization generates a variable delay between the capture of the video and the sampling time of the position of the probe and the needle. This random delay cannot be compensated so it induces inaccuracy in the processing. This makes the conception of AR platforms tricky and complex. Our implementation proposes an approach based on the DICOM-RTV standard to simplify the conception of AR platforms regarding synchronization. We believe it is the base of new AR usages in OR, for example for orthopedic surgery. Such a system could also be adapted to record video and metadata streams and replay them later with perfect synchronization, which could be very helpful for offline AI training.

References

- El Jaouhari, S., Gibaud, B., Lemonnier, P., Pasquier, G., Poiseau, E., Guiffard, E., . . . Cordonnier, E. (2019). Introduction to DICOM-RTV: a new standard for real-time video communication in hospitals.
- Roberts, D. W., Strohbehn, J. W., Hatch, J. F., Murray, W., & Kettenberger, H. (1986). A frameless stereotaxic integration of computerized tomographic imaging and the operating microscope. *Journal of neurosurgery*, vol 65, no 4, 545-549.
- Supplement 202: Real-Time Video*. (2019). Digital Imaging and Communications in Medicine (DICOM).
- Sutherland, E. (1968). A head-mounted three dimensional display, in Proceedings of the December 9-11, fall joint computer conference, part I. ACM, pp. 757-764.
- Wen, R., Tay, W.-L., Nguyen, B. P., Chng, C.-B., & Chui, C.-K. (2014). Hand gesture guided robot-assisted surgery based on a direct augmented reality interface. *Computer Methods and Programs in Biomedicine*.