Real-Time Augmented Reality for Robotic-Assisted Surgery

Jørgensen, Martin Kibsgaard; Kraus, Martin

Published in:
The 3rd AAU Workshop on Robotics

Publication date:
2015

Document Version
Accepted author manuscript, peer reviewed version

Citation for published version (APA):
Jørgensen, M. K., & Kraus, M. (2015). Real-Time Augmented Reality for Robotic-Assisted Surgery. In Z.-H. Tan, S. Bai, T. Bak, M. Rehm, & E. A. Jochum (Eds.), *The 3rd AAU Workshop on Robotics: Proceedings* (pp. 19-23). Article 3 Aalborg Universitetsforlag.

Real-Time Augmented Reality for Robotic-Assisted Surgery

Martin Kibsgaard and Martin Kraus
Department of Architecture, Design and Media Technology
Aalborg University, Denmark
{kibsgaard, martin}@create.aau.dk

***Abstract*—Training in robotic-assisted minimally invasive surgery is crucial, but the training with actual surgery robots is relatively expensive. Therefore, improving the efficiency of this training is of great interest in robotic surgical education. One of the current limitations of this training is the limited visual communication between the instructor and the trainee. As the trainee's view is limited to that of the surgery robot's camera, even a simple task such as pointing is difficult. We present a compact system to overlay the video streams of the *da Vinci* surgery systems with interactive three-dimensional computer graphics in real time. Our system makes it possible to easily deploy new user interfaces for robotic-assisted surgery training. The system has been positively evaluated by two experienced instructors in robot-assisted surgery.**

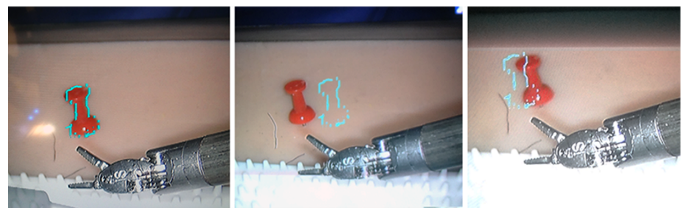

Fig. 2. Telestration. Left: drawing around a pen on the touch screen. Center: the drawing as displayed in the surgeon console's left display. Right: the drawing in the right display. The drawing gets offset for both eyes, making it appear at a different depth than the background. For some users this creates double vision and thus makes precise pointing impossible.

I. Introduction

The motivation for this work is to improve training efficiency for surgeons and medical students who are learning to use the *da Vinci* robotic surgery system for minimally invasive surgery. The training in robotic-assisted surgery is considered expensive, but it is also crucial for improving the outcome of operations [1]. A *da Vinci* system allows a surgeon to perform surgery inside a patient by controlling robotic arms through small incisions. A stereoscopic endoscope (camera) is also inserted to allow the surgeon to see the operating field. The basic setup of a *da Vinci* system is illustrated in Figure 1.

By observing training sessions and interviewing the instructors at Aalborg University Hospital (AUH), it became apparent that the limitation of the communication between the trainer and trainee is a significant problem in the training with the *da Vinci* system. The source of this problem is that the vision of the trainee sitting at the surgeon console is limited to that of the stereoscopic endoscope, which makes it difficult for the instructor to visually communicate with the trainee. Even a simple task such as pointing to a place in the operational field is currently difficult and anything more advanced is nearly impossible. If the trainee leans back from the console to gain vision of the instructor and/or operating room, the surgical system locks and the trainee loses vision of the surgical field. Often it is necessary for the trainee to leave the console for the instructor to take over to, for example, identify anatomy or demonstrate a skill.

Fig. 1. *da Vinci* surgery system. Left: surgeon operating the surgeon console. Center: assisting nurse operating the patient cart. Right: vision cart. Copyright 2015 Intuitive Surgical, Inc. [2]

Currently, the *da Vinci* system offers two methods of visual communication without disrupting operation of the robot: drawing on a touch screen (telestration) or showing additional video signals side-by-side with the endoscopic view (a feature called TilePro, see Figure 3). However, neither are being used during training at AUH. The 2D drawings from the telestration do not translate well to the stereoscopic display in the surgeon console (see Figure 2) [3], [4]. The drawings are displayed at a fixed depth in the stereoscopic view, which often results in double vision and, therefore, greatly limits the situations in which they are usable. The TilePro feature is not being utilised during training sessions as it reduces the size and resolution of the endoscopic view.

In this work, we present a system that uses hardware keying to overlay the stereoscopic video signal with computer graphics with minimal latency. The system allows the instructors to communicate more precisely with the trainee without interrupting the trainee's operation of the robot. We describe how our system improves on previous work in Section II. In Section III we describe the video hardware of the *da Vinci* systems and our system — including input and output devices. Evaluation of the system with two experienced instructors is described in Section IV. The system creates a foundation for multiple new user interfaces that we propose and discuss further in Section VI.

II. Previous Work

To solve the visual communication limitation, Galsgaard et al. [5] developed a system that overlays the stereoscopic video displayed in the surgeon console with 3D computer graphics. However, the 164 ms latency induced by the system made it difficult to operate the *da Vinci* robot and difficult to validate the benefits of the system. Ali et al. [3] created a similar system where they overlaid the stereoscopic video to create and evaluate 3D telestration. Like Galsgaard et al. [5] and other similar work [6], [7] (we assume add 400 ms delay), their overlay approach also involves data transfer with a CPU and multiple format conversions, which presumably introduces a noticeable delay of the video signal.

The previous works most similar to our system are [8] and [9]. Hattori et al. [8] employ an OCTANE MXE graphics workstation that can directly overlay computer graphics on an analog S-Video signal. It is, however, no longer in production and only the first generation *da Vinci* system uses S-Video for the stereoscopic video signal. Figl et al. [9] use two external video mixers to overlay graphics with minimal latency — also on the S-Video signal of the first generation system. Newer video mixers that support the digital HD video format of the newer *da Vinci* systems are available, but they are often bulky and expensive, making them less ideal for the scenario at AUH. An exception to this, is the Blackmagic Design ATEM Television Studio switcher; however, to reduce the delay to less than one frame (if possible at all) would require a more costly and bulky solution than the system we describe in Section III. It would require two switchers, SDI outputs from a computer and synchronization of the signals. A delay of one frame might appear very small, but at 60 Hz, it corresponds to more than 16 ms, while a 10 ms delay can already make a statistically significant difference in user performance [10].

Most other works that augment the video signal of the *da Vinci* systems use the TilePro feature to display both the original (undelayed) and the augmented video signal at the same time [11], [12], [13]. In this way, they avoid that the additional latency affects the control of the robot. However, as can be seen in Figure 3, this significantly reduces the size and resolution of both video signals. From [11] it appears that surgeons then tend to switch off the augmented video signal whenever possible. Thus, the TilePro setup is not ideal in a training situation.

None of the cited works appear to support the digital HD video format of the newest generation *da Vinci* systems, which require much faster processing because of the higher data rate (0.3 Gbit/s vs. 1.5 Gbit/s).

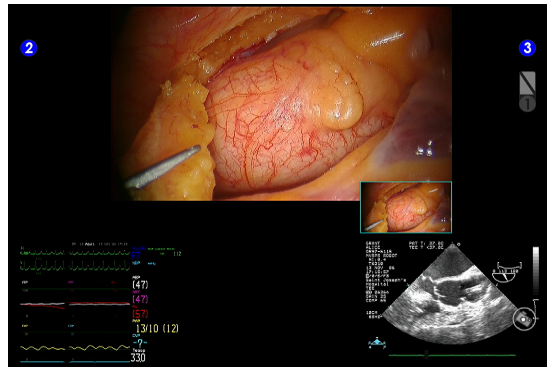

Fig. 3. TilePro feature of the *da Vinci S HD* system seen from the surgeons perspective. Here the feature is used to display physological information (left) and ultra sound video (right). Copyright 2015 Intuitive Surgical, Inc. [2]

III. Materials and Methods

A. Video Hardware

Initially, we investigated the video hardware of various *da Vinci* systems that were available to us (*da Vinci*, *da Vinci S HD*, and *da Vinci Si*). We examined the video signals using a Blackmagic Design DeckLink Duo card and a custom made format detection application. The pixel format of all the examined systems is YCbCr 4:2:2, which is a chroma subsampling format commonly used in live TV productions. The video components of the systems are connected by RG-59 coaxial cables with BNC connectors and are in most cases directly accessible. The video formats of each generation of the system are listed in Table I.

1) First generation da Vinci: The first generation *da Vinci* system uses S-Video (analog) and SD-SDI (standard definition serial digital interface). The SD-SDI video signal can be intercepted directly between the vision cart and the surgeon console. To overlay graphics on the first generation *da Vinci* system, a single DeckLink Duo card is sufficient, as this PCIe card supports internal keying on two SD-SDI signals at the same time. The stereoscopic video signal consists of two separate SDI signals: one for each eye. The setup for a first generation *da Vinci* system is illustrated in Figure 4.

2) da Vinci S HD and da Vinci Si: The newer *da Vinci S HD* and *da Vinci Si* systems use HD-SDI (high definition SDI) with a 1080i video format. On those systems, the video signal can be intercepted between the camera controller(s) and the synchronizer/CORE unit in the vision cart as illustrated in Figure 5. Additionally, the *da Vinci S HD* system has a redundant HD-SDI output on each camera controller.

For these newer systems, a more advanced video device is required to allow for keying in HD, for example, one DeckLink HD Extreme card for each video signal. We acquired two DeckLink HD Extreme cards (second generation) to support stereoscopic keying on a *da Vinci S HD* system. Recently, Blackmagic Design has released a less costly video card, DeckLink SDI 4K, that also supports internal HD keying.

TABLE I. Video formats of the da Vinci surgical systems. *i* indicates an interlaced format.

	Video Format	Refresh Rate	Year
da Vinci	PAL / NTSC	50i / 60i	1999
da Vinci S	PAL / NTSC	50i / 60i	2006
da Vinci S HD	1080 HD	59.94i	2006
da Vinci Si	1080 HD	59.94i	2009
da Vinci Xi	(1080 HD)	(59.94i)	2014

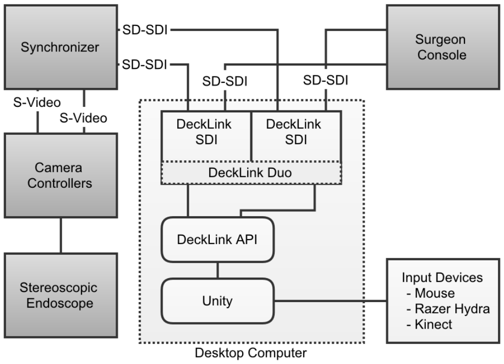

Fig. 4. System diagram of the setup used to overlay 3D graphics on the stereoscopic video of the first generation *da Vinci* system. Dark grey indicates *da Vinci* components.

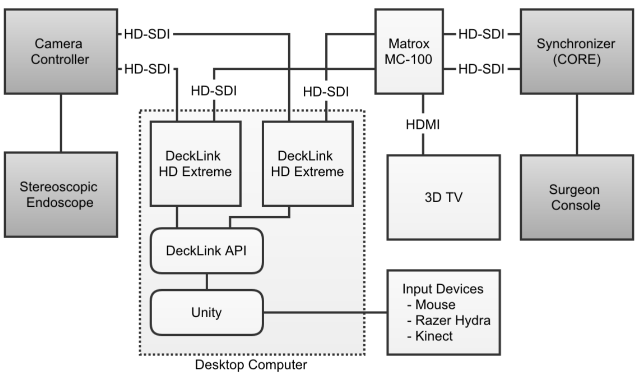

Fig. 5. System diagram of the setup used to overlay 3D graphics on the stereoscopic video of the *da Vinci S HD* and newer systems. Dark grey indicates *da Vinci* components.

B. Computer Graphics

For future applications and to make it possible to overlay three-dimensional (3D) graphics, we implemented a system in the free version of the game engine Unity, which we used to render 3D graphics that are output through the DeckLink API to the DeckLink card(s). We have created a custom graphical interface that allows developers to choose input, output or keying mode for each DeckLink device installed in the system. Furthermore, we implemented virtual instruments resembling those used on the *da Vinci* systems, which can be seen in Figure 6. Currently, the bottleneck of the system is the time it takes to transfer the rendered images from the graphics card to the system RAM, where the DeckLink cards can access them. A workstation graphics card, i.e. NVIDIA Quadro or AMD FirePro, could possibly improve the transfer rate as they have dedicated processing units for exactly that. We emphasize, however, that this bottleneck only delays the computer graphics, and it does not affect the stereo video signal of the endoscope.

C. Input and Output

For the trainer to be able to control the 3D graphics, we employed Razer Hydra game controllers. They support three translational degrees of freedom (DOF), three rotational DOFs and an analog trigger (one DOF); therefore, they are similar to the 7-DOF controls of the *da Vinci* systems [14]. Additionally, we implemented control with the Kinect for Windows v2, which supports skeleton tracking and hand gestures. To display the stereoscopic video signal to the trainer, we used a Matrox MC-100 converter, which converts two HD-SDI signals to 3D HDMI (see Figure 5). The latter signal is displayed by a 3D TV with passive stereo. To minimize latency for the 3D TV, it is important to turn off any processing performed by the TV (often achieved by switching the TV to *game mode*)

To compensate for any inaccuracies in the camera optics, the *da Vinci* systems have a calibration feature, where the two video streams are aligned by offsetting the images both horizontally and vertically. The offsets are introduced after we intercept the video signal and it is therefore necessary to adjust the signal for the 3D TV as well. Fortunately, the MC-100 has an on screen menu where it is possible to offset the images accordingly. Similarly, the virtual cameras have to be adjusted to match the offset, but in the opposite direction, for the computer graphics to be properly aligned.

D. System Setup

The core of our system is a regular desktop computer with the DeckLink cards installed. It has the following specifications: Intel Core i3-3240 3.4GHz CPU, Geforce 660 GTX GPU, 8 GB 1600 MHz RAM, 2×DeckLink HD Extremes (second generation), Microsoft Windows 8. The complete setup installed and tested at AUH is depicted in Figure 5.

To make the use of the *da Vinci* system less dependent on our system, we attached the redundant outputs of the camera controllers to loss-of-signal switchers. This has two benefits: it makes it possible to still use the robot without our system, and it makes it possible to immediately switch to the original video streams in case there are any problems with our system – simply by switching off our system.

IV. RESULTS

The proposed system is able to overlay the video streams of the *da Vinci* system with computer graphics. Like other work it supports the older generation *da Vinci* systems, and it is also

Fig. 6. Endoscopic video signal overlaid with a virtual instrument (blue) in real-time. This composition is shown to the left eye. A similar image is shown to the right eye, but with the real and virtual camera offset to the right.

compatible with the newer generations that utilize HD video signals. Our system is able to produce high quality computer graphics and overlay them in stereo at a rate matching the 1080i59.94 format. The overlaying induces less than 1 ms delay [15] on the original video signal. The computer graphics (e.g. virtual tools) are slightly delayed (less than 100 ms) compared to outputting on a regular computer monitor.

We evaluated our system with the help of two experienced robotic surgery instructors. Our system has twice been connected to the *da Vinci S HD* system normally used for training at Aalborg University Hospital. The tests showed that the system can improve visual communication between the instructor and a trainee. Especially, the overlaid virtual instruments were considered very useful by the experienced surgeon for showing advanced tasks to the trainees, while a simpler cursor was preferable for the assisting nurse.

Regarding user input devices, the Razer Hydra is useful for experienced surgeons to demonstrate certain skills to a trainee. The Kinect sensor has less DOFs, but is sufficient for pointing or drawing, which is often what is needed for the assisting surgeon or nurse. It is also possible to use the Kinect sensor in a sterile environment as it does not require any touching. However, because of the limited space in the training room at AUH, it is not convenient to use the Kinect sensor and a wireless gyroscopic mouse has since proved to be superior.

During the test we noticed that the perceived depth of the 3D graphics (virtual robotic instruments) did not entirely match the depth of the real environment. Furthermore, we learned that a clutch (allowing to move the controls without moving the instruments) is necessary to sufficiently simulate the real instruments. Since the test, the cameras have been adjusted to match properly with the endoscope used at AUH and clutching of the tools has been implemented. They have since been positively evaluated by several of the instructors at AUH.

The stability and usability of the system has since been tested for extended periods of time at multiple eight hour sessions. The only technical problem during the sessions was the 3D TV turning off, presumably, due to power fluctuations caused by the cauterization instruments. It did not affect the overlay system and the 3D TV was simply switched back on.

V. Conclusion

We have created a compact system that can be the foundation for multiple new types of user interfaces directly visible in the surgeon console during training for robotic-assisted surgery. Our implementation allows overlaying of stereo graphics in 1080 HD at 59.94i with less than 1 ms delay.

The developed software for keying graphics from Unity is available at homes.create.aau.dk/kibsgaard/.

VI. Future Work

Our system was developed to support future work and is currently used by students without requiring knowledge of the underlying APIs; i.e., the DeckLink API and OpenGL. Future work could further evaluate the use of overlaying virtual objects, images, videos, webcams, task lists, etc. during training with the *da Vinci* systems.

Acknowledgment

We thank Jane Petersson and Johan Poulsen from Aalborg University Hospital for all their help and sharing of expertise during the development and evaluation of our system.

References

[1] Health Quality Ontario, “Robotic-assisted minimally invasive surgery for gynecologic and urologic oncology,” *Ontario Health Technology Assessment Series*, vol. 10, no. 27, pp. 1–118, Dec. 2010. [Online]. Available: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3382308/

[2] Intuitive Surgical. (2014) Intuitive Surgical - Image Gallery - da Vinci Si System. http://www.intuitivesurgical.com/company/media/images/davinci_si_images.html.

[3] M. R. Ali, J. P. Loggins, W. D. Fuller, B. E. Miller, C. J. Hasser, P. Yellowlees, T. J. Vidovszky, J. J. Rasmussen, and J. Pierce, “3-d telestration: a teaching tool for robotic surgery,” *Journal of Laparoendoscopic & Advanced Surgical Techniques. Vol. 18, Issue 1*, pp. 107–112, 2008.

[4] C. Hasser, D. Larkin, B. Miller, G. Zhang, and W. Nowlan, “Medical robotic system providing three-dimensional telestration,” Patent EP1 965 699 A2, 2008, US Patent EP1965699 A2. [Online]. Available: http://www.google.com/patents/EP1965699A2

[5] B. Galsgaard, M. M. Jensen, F.-O. Matu, M. Thgersen, and M. Kraus, “Stereoscopic augmented reality system for supervised training on minimal invasive surgery robots,” *Proceedings of Virtual Reality International Conference (VRIC 2014)*, 2014.

[6] L.-M. Su, B. P. Vagvolgyi, R. Agarwal, C. E. Reiley, R. H. Taylor, and G. D. Hager, “Augmented reality during robot-assisted laparoscopic partial nephrectomy: Toward real-time 3d-ct to stereoscopic video registration,” *Urology*, vol. 73, no. 4, pp. 896 – 900, 2009. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S009042950801947X

[7] B. Vagvolgyi, L.-M. Su, R. Taylor, and G. D. Hager, “Video to ct registration for image overlay on solid organs,” *Proc. Augmented Reality in Medical Imaging and Augmented Reality in Computer-Aided Surgery (AMIARCS)*, pp. 78–86, 2008.

[8] A. Hattori, N. Suzuki, M. Hashizume, T. Akahoshi, K. Konishi, S. Yamaguchi, M. Shimada, and M. Hayashibe, “A robotic surgery system (da vinci) with image guided function–system architecture and cholecystectomy application,” *Studies in Health Technology and Informatics*, vol. 94, pp. 110–116, 2003.

[9] M. Figl, D. Rueckert, D. Hawkes, R. Casula, M. Hu, O. Pedro, D. P. Zhang, G. Penney, F. Bello, and P. Edwards, “Image guidance for robotic minimally invasive coronary artery bypass,” *Computerized Medical Imaging and Graphics: The Official Journal of the Computerized Medical Imaging Society*, vol. 34, no. 1, pp. 61–68, Jan. 2010.

[10] R. Azuma, Y. Baillot, R. Behringer, S. Feiner, S. Julier, and B. MacIntyre, “Recent advances in augmented reality,” *IEEE Comput. Graph. Appl.*, vol. 21, no. 6, pp. 34–47, Nov. 2001. [Online]. Available: http://dx.doi.org/10.1109/38.963459

[11] F. Volont, N. C. Buchs, F. Pugin, J. Spaltenstein, B. Schiltz, M. Jung, M. Hagen, O. Ratib, and P. Morel, “Augmented reality to the rescue of the minimally invasive surgeon. the usefulness of the interposition of stereoscopic images in the da vinci robotic console: Augmented reality in robotic surgery,” vol. 9, no. 3, pp. e34–e38. [Online]. Available: http://europepmc.org/abstract/med/23239589

[12] O. Ukimura, M. Aron, M. Nakamoto, S. Shoji, A. L. d. C. Abreu, T. Matsugasumi, A. Berger, M. Desai, and I. S. Gill, “Three-dimensional surgical navigation model with TilePro display during robot-assisted radical prostatectomy,” vol. 28, no. 6, pp. 625–630. [Online]. Available: http://online.liebertpub.com/doi/abs/10.1089/end.2013.0749

[13] A. Hughes-Hallett, P. Pratt, E. Mayer, S. Martin, A. Darzi, and J. Vale, “Image guidance for allTilePro display of 3-dimensionally reconstructed images in robotic partial nephrectomy,” vol. 84, no. 1, pp. 237–243. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0090429514003409

[14] K. Grande, R. S. Jensen, M. Kraus, and M. Kibsgaard, “Low-cost simulation of robotic surgery,” in *Proceedings of the Virtual Reality International Conference: Laval Virtual*, ser. VRIC ’13. ACM, 2013,

pp. 6:1–6:4. [Online]. Available: http://doi.acm.org/10.1145/2466816.2466823
[15] Yoshan, Developer Support, Blackmagic Design, Personal Communication, 2015.