

A Configurable System for Role-specific Video Imaging During Laparoscopic Surgery

Yuebing Jiang*, Timothy Perez†, Marios Pattichis*, Bilal Khan†

*Department of Electrical and Computer Engineering

†Department of Surgery

University of New Mexico

Albuquerque, NM 87106

Email: yuebing@ece.unm.edu, tperez@salud.unm.edu, pattichis@ece.unm.edu, bmkhan@salud.unm.edu

Abstract—In minimally invasive surgery, the surgeon and the assistant rely on a single laparoscopic video view for performing different clinical roles. The assistant is tasked with manipulating the camera view so as to maintain a global, *panoramic* view of the operation. The surgeon needs to remain focused on the operation, requiring a detailed *close-up* view. We use the term *role-specific video imaging* to describe the need to provide separate views for the assistant and the surgeon.

In this paper we introduce *role-specific video imaging* for laparoscopic surgery. The system is designed to be *configurable* in the sense that imaging parameters and algorithms can be adjusted in real-time so as to meet the specific needs that arise. The system was evaluated on 4 cases by two surgeons on a Linux-based 3.2.0 Kernel, with 4GB RAM, and Intel 3.4GHz I7 (2nd generation) microprocessor. Clinical evaluation of the different configuration modes has helped us determine that high-quality *role-specific imaging* can be achieved for zooming factors that are larger or equal to 2x2 with bilinear interpolation, while maintaining 30 frame per seconds for the panoramic and close-up views. In future work, in order to minimize interaction with the surgical team, the system will be upgraded to incorporate tracking of the operating instrument during surgery.

Index Terms—Laparoscopic, role-specific video, interpolation.

I. INTRODUCTION

More than just changing the size of the incision, laparoscopic surgery challenges the paradigms of operating room design, surgical team dynamics, and technical skills [1], [2]. Current laparoscopic imaging relies on a single view for the entire medical team. Yet, there are different and potentially conflicting needs and requirements among the members of the team. The surgeon needs a detailed, *close-up* view of the operative field while the assistant relies on a *panoramic* view.

A consequence of the constraints of current laparoscopic imaging is that the assistant manipulates instruments frequently out of visual field for operation. Functioning without visual feedback is inefficient and potentially dangerous. These challenges in laparoscopic surgery have been acknowledged, but little has been described to define performance of the assistant [3], nor ways to improve it. Methodology exists to quantify technical skill performance [4], [5] and optimize task efficiency by setting camera and monitor positions [6]. However, these methods have not been applied to assess performance of the assistant individually nor the dynamics of

surgeon/assistant interaction. Although the value of an adept assistant at surgery is time honored, this belief is mostly anecdotal and has not been the subject of rigorous study.

The application of a miniature video camera to a laparoscope birthed the field of minimally invasive surgery with its concomitant benefits of reduced post-operative pain and shorter recovery. Although the use of video technology allows for the entire team to visualize the abdominal cavity, a video laparoscope emulates the paradigm of a direct view laparoscope. A customized user interface can provide a detail view for the surgeon to do the operation, and a panoramic view for the assistant to manipulate the camera position.

Related biomedical imaging research has been reported in [7]. In [7], the authors prepared a laparoscopy display system which fused multiple input data sources (including pre-operative imagery and real-time imagery). This system met the requirement for real-time video display at 29.97fps throughput. However, the reported system did not produce multiple, real-time views as proposed in this paper.

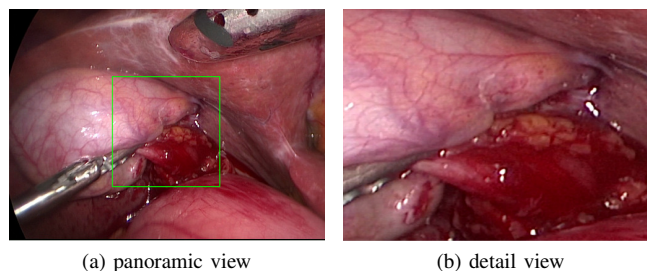


Fig. 1. Panoramic and close-up views for laparoscopic imaging. (a) Panoramic view for allowing the assistant to manipulate the camera. (b) Detailed view produced by real-time zooming (4x4 enlargement) for the surgeon.

In this paper, we introduce a configurable system that implements real-time, role-specific imaging. A summary of the basic ideas of our paper includes:

- **Role-specific imaging that includes panoramic and close-up views.** The basic idea is demonstrated in Fig. 1. Two separate video bitstreams are displayed in real time. The close-up is generated from the panoramic view using real-time interpolation.

- **Real-time configurable settings.** The system allows the users to dynamically adjust frame rates, interpolation methods, and zooming factors. We investigate the use of 4 different interpolation methods: (i) nearest-neighbor, (ii) bilinear interpolation, (iii) bi-cubic interpolation, and (iv) Lanczos-based interpolation.
- **Objective and subjective performance evaluation.** For objective evaluation, we report the video frame delays, achieved frame rates, and also estimate the video interpolation performance using both PSNR and SSIM [8]. For subjective evaluation, we report *mean opinion scores* for the different configuration settings.

The rest of the paper is organized as follows. In section II, we describe the different processing blocks, interpolation methods, and the user interface of our system. In section III, we provide results from objective and subjective evaluation of the system. A discussion of our results is also given in section III. Some concluding remarks are given in section IV.

II. METHODOLOGY AND IMPLEMENTATION

A. System Overview

The block diagram of the system is shown in Fig. 2. The video is acquired by a laparoscopy camera (Karl Storz Laparoscopic camera system). We refer to subsection III-A for details on the (simulated) acquisition and actual display resolutions. The camera video is then input to an Epiphan DVI2PCIe video capture card. The acquired video is used in generating the panoramic and close-up views in two separate monitors. For the panoramic view, the user can control the display frame rate by selecting a temporal reduction factor from 1 (original frame rate) to 6 (displaying 1 in every 6 frames). For the close-up view, the input video is cropped and interpolated to produce the zoomed image, while using the maximum, possible display rate. Here, we note that the display frame rate for the close-up view increases as we reduce the panoramic frame rate. In other words, computational time savings from slowing down the refresh rate for the panoramic view are used to maximize the refresh rate for the close-up view.

In section II-B, we describe the different interpolation methods that are used. We then describe the user interface for our approach in II-C.

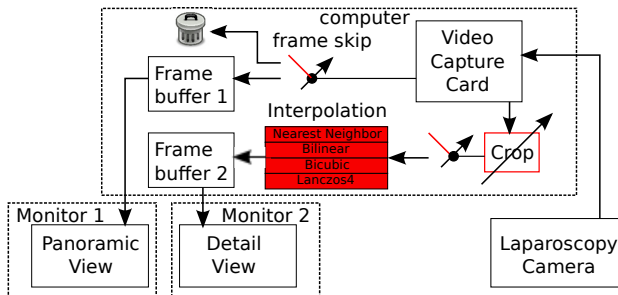


Fig. 2. Block-diagram for role-specific, laparoscopic imaging. Refer to Fig. 3 for the settings interface.

B. Interpolation Methods

We focus our attention on interpolation methods that are applicable for real-time video image processing. As a result, our approach is to focus on interpolation methods that require reasonable computational resources. Thus, in the order of increasing computational complexity, we have: nearest-neighbor, bilinear, bicubic, and Lanczos4 methods. We also note that efficient implementation of all of the interpolation methods are openly available in OpenCV [9].

For completeness, we also mention more sophisticated methods described in [10], [11]. In [10], the authors use anisotropic approach as a post-processing method to remove edge interpolation artifacts. More recently, in [11], the authors introduced a regularized, local linearized regression model for edge preserved interpolation. Unfortunately, these methods do not allow for real-time implementations or arbitrary zooming.

Nearest Neighbor Interpolation: In nearest neighbor interpolation, the interpolated pixel takes the value of its nearest neighbor. In terms of memory accesses, we process the image row-by-row. In an efficient implementation, a single image row is stored in the cache. The nearest-neighbor interpolation method requires the minimum number of memory accesses while not requiring the implementation of any arithmetic operations.

However, a significant limit of the approach comes the introduction of significant block artifacts at significant zooming levels. Thus, the method can be very effective for lower-level zooming.

Bilinear Interpolation: Bilinear interpolation requires the determination of the 4 nearest points. This is implemented in two phases. First, we interpolate along the columns. Second, we interpolate along the rows. The implementation is based on:

$$f_c(x+p, y) = f(x, y) \cdot (1-p) + f(x+1, y) \cdot p, \quad (1)$$

$$f(x+p, y+q) = f_c(x+p, y) \cdot (1-q) + f_c(x+p, y+1) \cdot q, \quad (2)$$

where $p, q \in (0, 1)$. Compared to nearest-neighbor interpolation, bilinear interpolation does not suffer from severe blocking artifacts. Bilinear interpolation provides a nice balance between computational efficiency and interpolation accuracy.

Bicubic Interpolation: Bicubic interpolation [12] uses 16 nearest neighbor points (4x4 neighborhood) to estimate the interpolated pixel. The approach is separable in the sense that we interpolate along the rows and columns separately. It is an extension of bilinear interpolation in the sense that it fits the data with a piecewise cubic model. Naturally, this higher-order model comes with requirements for more continuity in the image.

Lanczos4 Interpolation: Lanczos interpolation is based on the *sinc* function. Here, we note that the *sinc* function is the optimal choice for band-limited signals. However, unlike real images, band-limited signals are infinite. Also, the *sinc* function itself is infinite. Furthermore, if we model edges using

step functions, band-limited approximations will always produce ringing artifacts around each edge. Lanczos interpolation is thus based on truncating the sinc-function over a local, 8x8 window of the nearest neighboring pixels.

We present a summary of computational complexity for all methods in Table I. We have increased computational complexity from the nearest-neighbor to the Lanczos4 method. We note that parallel implementations (not considered here) tend to minimize differences among the different methods. Furthermore, we note the more restrictive assumptions that are made by more sophisticated methods: (i) piecewise constant for nearest neighbor, (ii) piecewise linear for bilinear, (iii) piecewise cubic for bicubic, and (iv) band-limitness for Lanczos4.

TABLE I

COMPUTATIONAL OVERHEAD FOR THE DIFFERENT INTERPOLATION METHODS IN TERMS OF THE REQUIRED ARITHMETIC OPERATIONS PER PIXEL. HERE, WE DO NOT TAKE INTO ACCOUNT MEMORY ACCESS ISSUES.

	Nearest neighbor	Bilinear	Bicubic	Lanczos4
Additions/Subtractions	0	3	17	15
Multiplies	0	2	22	40

C. User Interface

The control panel for the role-specific view system is shown in Fig. 3. Real-time measurements on the performance of the system is provided at the top of the interface. This includes the achieved frame rate in the number of frames per second, and the time delays through the different components of the system (see Fig. 2). The user can select the capture and display resolutions independently. Here, we note that the capture resolution is adjustable for simulation purposes. The interpolation method can be set to (i) nearest neighbor, (ii) bilinear, (iii) bicubic, or the (iv) the lanczos4 method.

The panoramic and close-up views are controlled independently. The user can select the panoramic view frame skip rate by using the combo box under “Panoramic Downsample”. Here, for a setting of “1”, no video frames are skipped. A setting of “6” will display one every 6 of the acquired video frames. The close-up view is specified in the panoramic view by using mouse clicks to specify the region of interest. In the example of Fig. 3, the close-up region is interpolated to 4 times the size of the original video acquisition resolution. We refer back to Fig. 1 for an example that shows both the panoramic and the the close-up views.

III. SYSTEM PERFORMANCE EVALUATION

A. Laparoscopic Video Dataset and Subjective Evaluation System Setup

We tested the performance of the system with 4 different, fully anonymized, pre-recorded videos of Laparoscopic Cholecystectomy. Two of the videos were recorded at 1920x1088 pixels per frame @ 30fps, and another two videos were recorded at 720x480 pixels per frame @ 30 fps. We used

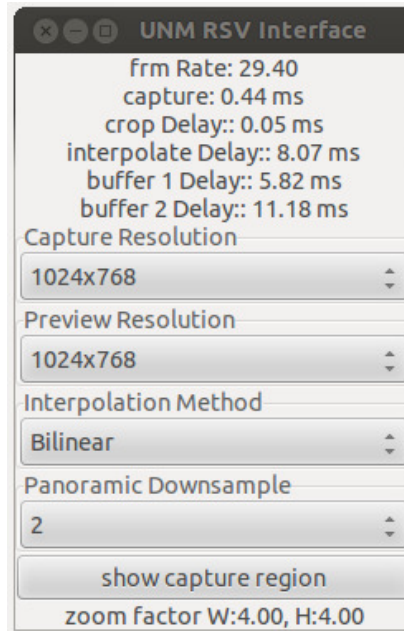


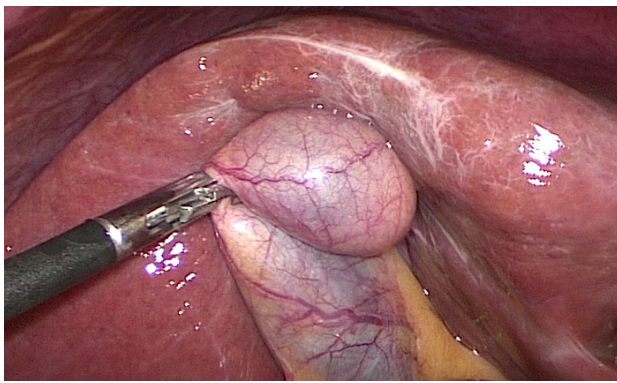
Fig. 3. System interface providing real-time performance measurements and user-adjustable options.

a separate desktop for simulating different video camera acquisition rates and resolutions. This allowed us to experiment with different options without requiring that we work in the operating room. On the other hand, we note that this approach limited our simulations to the display characteristics of the desktop that is used to simulate the laparoscopic camera. In particular, the original laparoscopic videos were normalized to the desktop display resolution of 1024x768.

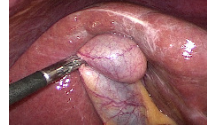
In terms of evaluating the performance, we need to make sure that the viewing conditions allow the doctors to view spatial frequencies that can be refreshed at frame rates that can exceed 30 fps at high contrast. In our experiments, the distance between the doctors and the monitor varied from 0.9-1.07 meters. Thus, assuming a viewing distance of 1 meter (39.37 inches) of a window of 11 inches x 8.5 inches that displays 1024x768 pixels, we get maximum display frequencies of angular resolution of over 60 cycles per degree for both vertical and horizontal frequencies. Thus, our display exceeds the practical limit of 50 cycles per degree of human visual perception [13], [14]. We do however note that this result is for the display of a single frequency component. The experiment does not consider masking artifacts from two or more frequency components. Such issues are addressed directly by asking the medical experts to evaluate the quality of the resulting videos.

B. System Description

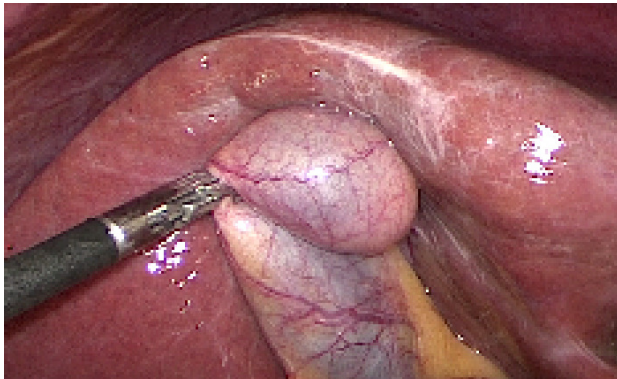
The overall system latency consists of components shown in Table II. We also refer back to Fig. 2 for a block diagram of the entire system. In terms of hardware, we have a Linux based 3.2.0 kernel with an Intel(R) Core(TM) i7-2600 CPU with



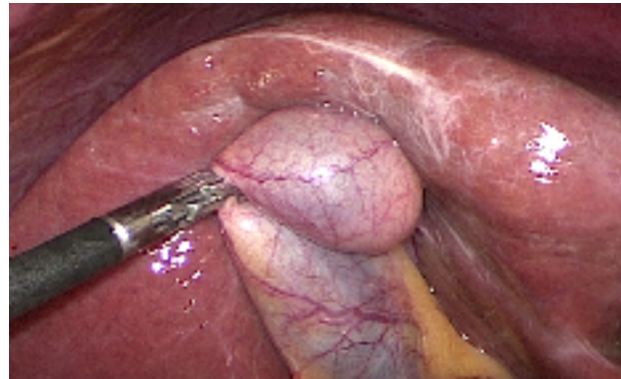
(a) Original input image.



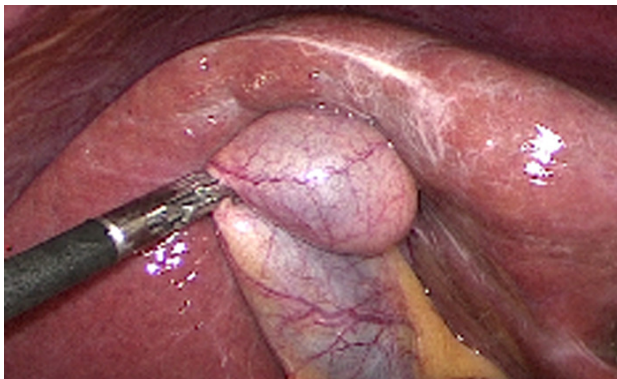
(b) Shrunk image



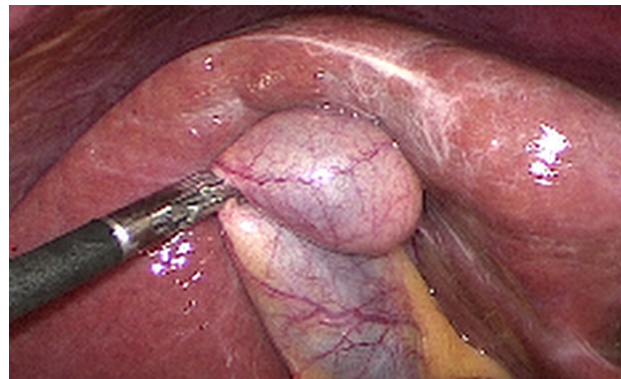
(c) nearest neighbor interpolation:SSIM=0.5481,PSNR=23.64 dB



(d) bilinear interpolation:SSIM=0.6934,PSNR=26.55



(e) bicubic:SSIM=0.6864,PSNR=26.02



(f) lanczos4:SSIM=0.6723,PSNR=25.79

Fig. 4. Performance of different interpolation methods based on 3x3 downsampling. (a) The original input image that is used for performance evaluation. (b) A 3x3 down-sampled version of the original input image of (a). (c) Nearest-neighbor interpolation from the down-sampled image of (b) to the original image size of (a). (d) Same as in (c) for bilinear interpolation. (e) Same as in (c) for bicubic interpolation. (f) Same as in (c) for lanczos4 interpolation.

4G DDR3 memory. The input video size is set to 1024x768 captured by a dvi2pcie grab card made by epiphan. First, the capture card grabs the input into frame memory. Second, a selected area is cropped from the frame memory. Third, interpolation maps the selected area into a second frame memory. Then, the frame memory contents are copied to the display buffers for the panoramic and close-up views.

Based on our usability study, the surgeon needs faster video updates of the close-up view, as opposed to potentially slower updates for the panoramic view that is seen by the assistant. Thus, we allow our system to update the panoramic view at a

rate that is downsampling the frame rate of the close-up view by an integer amount. We consider frame rates from 16 to 40 frames per second for the close-up view. These frame rates can be downsampled from 1 to 6 for generating the panoramic view. At such slower rates, we can see that clinical validation is essential.

We used one surgical video to perform a preliminary, quantitative study on the resulting video quality. For this example, we perform interpolation over 100 video frames that represent a 30 to 1 downsampled version of the original 3,000 video frames. To validate performance, we also reduce the frame size

TABLE II

EXECUTION TIMES FOR DIFFERENT COMPONENTS OF THE ROLE SPECIFIC VIEW SYSTEM. REFER TO FIG. 2 FOR THE DIFFERENT COMPONENTS OF THE SYSTEM. THE RESULTS REPRESENT SUMMARIES OVER ALL OF THE VIDEOS. ALSO, NOTE THAT THE RESULTS ONLY DEPEND ON THE VIDEO INPUT RESOLUTION THAT IS FIXED.

Procedure	Time (ms) mean \pm std. dev.
Video capture	0.427 \pm 0.013
Video crop	0.050 \pm 0
Nearest neighbor interp.	1.53 \pm 0.074
Bilinear interp.	8.38 \pm 0.304
Bicubic interp.	12.25 \pm 0.539
Lanczos4 interp.	24.17 \pm 0.487
Video display	10.38 \pm 0.047

TABLE III

ACHIEVED FRAME RATES FOR THE CLOSE-UP VIEW. WE PRESENT FRAME RATES IN TERMS OF NUMBER OF FRAMES PER SECOND. FRAME RATES VARY AS A FUNCTION OF THE INTERPOLATION METHOD AND THE FRAME-RATE DOWNSAMPLING FACTOR THAT IS USED TO GENERATE THE PANORAMIC VIEW.

Int. Method \rightarrow \downarrow Down. Fr. rate	Nearest neighbor	Bilinear	Bicubic	Lanczos4
1	24.87	22.05	21.08	16.93
2	31.88	27.90	26.26	19.94
3	34.41	30.66	28.41	21.08
4	36.62	31.82	29.55	21.72
5	37.98	33.53	30.65	22.28
6	39.11	34.72	31.32	22.91

of the original video frames and then reconstruct to the original size using interpolation. This standard quantitative approach allows us to measure the performance of the interpolation method against the acquired video. However, note that in the actual clinical setting, we will be increasing (not decreasing) the size of the original video frames. On the other hand, there is no ground truth on the zoomed videos used in real-practice. Furthermore, our approach of downsampling and reconstructing to the original size is considered a standard method for evaluating interpolation algorithms quantitatively. Also note that we perform proper clinical validation on the zoomed video frames as described below and it is re-assuring that both methods agree that bilinear interpolation is best. The quantitative results are summarized in Table IV. The subjective evaluation results over all of the videos are given in Table V.

From the results, it is clear that bilinear interpolation performed the best. Also, note that spatial zooming beyond 2x2 will produce noticeable artifacts. As discussed in the caption of Table V, the average score of the original videos was 4. Thus, we consider the average zoom-quality value of 3.125 as being acceptable here.

We summarize the video-update quality results in Table VI. It is very interesting to note the trade-off between the qualities of the panoramic and close-up views. By reducing the update rate of the panoramic view, we produce higher quality close-up views. This is also reflected in the achieved frame rates of Table III. In terms of video-update quality, the best results are

obtained for downsampling the panoramic view by a factor of 4. For this case, except for the Lanczos4, all methods achieve frame rates that are above 30 frames per second.

TABLE IV

MEAN RECONSTRUCTED PSNR/SSIM OVER 100 INTERPOLATED VIDEO FRAMES. THE PSNR VALUE IS EXPRESSED IN DB. SSIM VALUES ARE BOUNDED ABOVE BY 1.0.

Int. Method \rightarrow \downarrow Sp. Down. rate	Nearest neighbor	Bilinear	Bicubic	Lanczos4
2x2	25.24 dB / 0.7343	28.75 dB / 0.8485	28.45 dB / 0.8455	28.34 dB / 0.8392
3x3	23.77 dB / 0.6627	26.70 dB / 0.7844	26.23 dB / 0.7789	26.05 dB / 0.7677
4x4	22.70 dB / 0.6155	24.88 dB / 0.7199	24.32 dB / 0.7100	24.14 dB / 0.6974
5x5	22.78 dB / 0.6289	24.29 dB / 0.7019	23.77 dB / 0.6959	23.59 dB / 0.6824
6x6	22.34 dB / 0.6133	24.28 dB / 0.7060	23.72 dB / 0.7006	23.51 dB / 0.6858
7x7	21.87 dB / 0.5974	23.54 dB / 0.6805	22.97 dB / 0.6741	22.77 dB / 0.6608
8x8	21.19 dB / 0.5714	23.03 dB / 0.6672	22.42 dB / 0.6588	22.24 dB / 0.6461

TABLE V

SUBJECTIVE EVALUATION OF VIDEO IMAGE QUALITY FOR SPATIAL ZOOMING FOR THE CLOSE-UP VIEWS FOR ALL VIDEOS. SUBJECTIVE SCORES ARE GIVEN USING 5 (EXCELLENT, IMPERCEPTIBLE ISSUES), 4 (GOOD, PERCEPTIBLE BUT NOT ANNOYING ISSUES), 3 (FAIR, SLIGHTLY ANNOYING ISSUES), 2 (POOR, ANNOYING ISSUES), AND 1 (BAD, VERY ANNOYING ISSUES). WE NOTE THAT THE ORIGINAL, ACQUIRED VIDEO IMAGE DISPLAYS WERE GIVEN SCORES OF: 4,5,2,5 WHICH GIVES AN AVERAGE SCORE OF 4.

Int. Method \rightarrow \downarrow Sp. Down. rate	Nearest neighbor	Bilinear	Bicubic	Lanczos4
2x2	3	3.25	3.25	3.125
3x3	1.875	2.625	2.5	2.375
4x4	1.75	2	2	2

TABLE VI

SUBJECTIVE EVALUATION OF THE VIDEO-UPDATE QUALITY FOR THE CLOSE-UP VIEW. SUBJECTIVE SCORES ARE GIVEN FROM 1 (LOWEST) TO 5 (HIGHEST). WE PROVIDE SCORES FOR THE CLOSE-UP/PANORAMIC VIEWS.

Int. Meth. \rightarrow \downarrow Down. Fr. Rate	Nearest neighbor	Bilinear	Bicubic	Lanczos4
1	3.5 /3.75	3.875 /4	4.125 /4.125	3.5 /3.375
2	4.125 /3	3.625 /3.25	4.125 /2.375	3.625 /2.75
4	4.5 /2.375	4 /2	4.125 /1.75	4.125 /1.75

IV. CONCLUSION

In this paper, we presented a configurable Role-specific video system for laparoscopic surgery. The system can be used to optimize the user's experience by allowing for real-time configuration of the interpolation methods and the video-update frame rates. From the results, we can conclude that

bilinear interpolation outperforms nearest-neighbor, bicubic, and Lanczos4 method. Furthermore, by controlling the display frame rates of the close-up and panoramic views independently, we can produce a trade-off of video quality between the two. This allows us to use fast frame rates for the panoramic view when rapid changes are happening, or switch to fast frame rates for the close-up view for routine surgery. In order to reduce interaction with the surgical team, the system will be upgraded to also track the surgical instrument during operation.

REFERENCES

- [1] G. Ballantyne, "The pitfalls of laparoscopic surgery: Challenges for robotics and telerobotic surgery." *Surg Laparosc Endosc Percutan Tech.*, vol. 12(1), pp. 1–5, 2002, Feb.
- [2] E. Berber and A. E. Siperstein, "Understanding and optimizing laparoscopic videosystems," *Surgical Endoscopy*, vol. 15, pp. 781–787, 2001.
- [3] R. Sur, A. Wagner, D. Albala, and L. Su, "Critical role of the assistant in laparoscopic and robot-assisted radical prostatectomy," *Journal of Endourology*, vol. 22(4), pp. 587–590, 2008.
- [4] G. Fried and L. Feldman, "Objective assessment of technical performance," *World Journal Surgery*, vol. 32(2), pp. 156–160, 2008.
- [5] K. Moorthy, Y. Munz, S. Sarker, and A. Darzi, "Objective assessment of technical skills in surgery," *BMJ*, p. 327:1032, 2003.
- [6] L. Haveran, Y. Novitsky, D. Czerniach, G. Kaban, M. Taylor, K. Gallagher-Dorval, R. Schmidt, J. Kelly, and D. Litwin, "Optimizing laparoscopic task efficiency: the role of camera and monitor positions." *Surgical Endoscopy*, vol. 21(6), pp. 980–984, 2007 June.
- [7] C. Martin, Q. Han, D. Clarke, C. Carswell, A. Park, and W. Seales, "Implementation and analysis of scalable display architecture for laparoscopy," in *Computer-Based Medical Systems, 2009. CBMS 2009. 22nd IEEE International Symposium on*, aug. 2009, pp. 1–8.
- [8] W. Zhou, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.
- [9] G. Bradski and A. Kaehler, *Learning OpenCV: Computer Vision with the OpenCV Library*. Cambridge, MA: O'Reilly, 2008.
- [10] Y. Cha and S. Kim, "Edge-forming methods for color image zooming," *Image Processing, IEEE Transactions on*, vol. 15, no. 8, pp. 2315–2323, aug. 2006.
- [11] X. Liu, D. Zhao, R. Xiong, S. Ma, W. Gao, and H. Sun, "Image interpolation via regularized local linear regression," *Image Processing, IEEE Transactions on*, vol. 20, no. 12, pp. 3455–3469, dec. 2011.
- [12] R. Keys, "Cubic convolution interpolation for digital image processing," *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 29, no. 6, pp. 1153–1160, dec 1981.
- [13] J. C. Russ, *The Image Processing Handbook*, 5th ed. CRC Press., 2007.
- [14] T. N. Cornsweet, *Visual perception*. New York: New York, Academic Process, 1970.