SOFTWARE DEFINED SURGICAL STEREO MICROSCOPE

J. K. Rappel*, A. Lahiri†, and C.L. Teo*

*Department of Mechanical Engineering, National University of Singapore
†Department of Hand and Micro Surgery, National University Health System

Keywords: Software Defined Surgical Stereo Microscope, Graphics Pipeline, Digital Stereo Microscope, Surgical Stereopsis.

Abstract

A software defined surgical visual aid platform as an alternative to the existing fixed form optical surgical microscope is described. A software defined digital stereo microscope, similar in principle as an endoscope, decouples the light gathering components and functions from those for light processing and image visualization components and provides digital implementations for the latter. It comprises an array of zoom cameras replacing the microscope objective, a realtime graphics engine pipeline replacing the optical processing pipeline, a flat screen 3D display replacing the oculars and a set of surgeon and procedure presets replacing the initial mechanical setting up of the microscope. The user interfaces for focus and zoom control are retained through the familiar foot pedal and additional interaction devices such as 3D measurement tools are added on demand for analysis and documentation. It replicates and augments the functionality of the optical stereo microscope, but removes the usage limitations of the optical pipeline. To be an effective surgical microscope, the realtime nature of the optical stereo microscope and the hand-eye collocation ability must be retained in this alternative visual aid.

1 Introduction

Microsurgical tasks are generally performed using a fixed form surgical optical microscope [1,2]. Viewing through the optical microscope requires the viewer to rigidly and precisely maintain the alignment of the optical axis of the eyes with that of the microscope oculars and also limits the depth of field and field of view. As shown in Fig.1, the current Optical Stereo Microscopes used in surgery (SLM – Surgical Light Microscope) are bulky embodiments of highly specialized optical design [1-4] whose numerous high precision optical components are optically coupled with high precision of alignment resulting in high manufacturing and recalibration costs.

A software defined digital stereo microscope, similar in principle as an endoscope [5], decouples the light gathering components and functions from those for light processing and image visualization and provides digital implementations for the latter. This separation makes it easy to adapt new hardware and technologies.

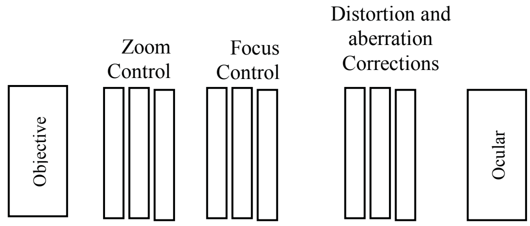

Fig.1 The Optical Stereo Microscope (SLM)

Recent advancements in plenoptic cameras using microlens arrays [6] suggests that many of the optical innovations in focusing and zoom may be replaced by simpler light gathering elements and flexible computational pipelines with much more precise and easy to use control. A graphics pipeline with a simple zoom lens optics [3] and light gathering objective optical elements can obtain the basic effects of an SLM such as magnification, clear focus, and stereoscopic viewing without the bulk and inflexibility of the SLM. With the use of graphics pipeline, high quality depth perception can be achieved without the usage difficulty [7] and the bulk of current SLM, which is of great utility in many applications of surgery. We term this innovation as SDSM, the software Defined Surgical stereo Microscope. According to the taxonomy in [8], SDSM is classified as a Window on World (WoW) with Video See Thru.

SDSM (Fig 2) described in this paper is a hand-eye collocated system, just like the modern SLM. Hand-eye collocation, placing hands in the line of sight of the gaze, is a means of achieving hand-eye coordination for better task performance. The term hand-eye coordination [9] refers to movements controlled with visual feedback and reinforced by hand contact with objects. SLMs offer hand-eye coordination by providing a conjoint visuomotor workspace. Any replacements to SLMs must maintain the hand-eye collocation.

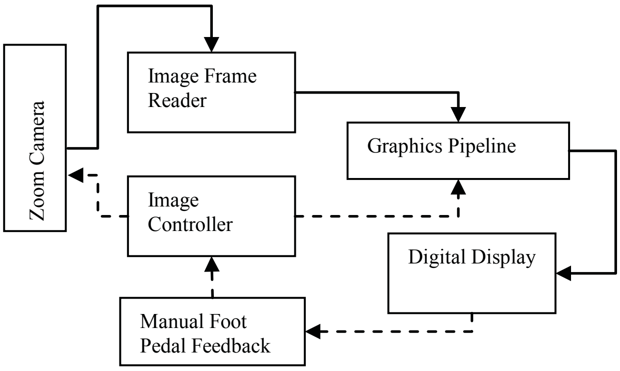

Fig.2 Software Defined Stereo Microscope (SDSM)

The paper is organized as follows. Section two describes the challenges in creating a Software Defined Stereo Microscope. In section 3, related work, especially the existing surgical visual aids and their pitfalls are discussed. In section 4, the software defined digital stereo microscope design and experiments used to characterize the microscope is discussed and in section 5 the results of the experiment are described along with the conclusions.

2 Problem

The adverse effects of delay and displacement of the visual feedback on task performance [9-17] is well known. To be an effective surgical visual aid, the realtime characteristics of the SLM should be reflected in the SDSM. The latency of the SDSM should be low enough to support not just the live view of the operating field but the live interaction with it. Interactions may be of three types: (i) Interactions to adjust the field of view requiring adjustments in the focus and zoom controlled by the foot pedal. The limit for the perception of instantaneous response to the user's interaction here is 100ms to thought disruption at 1000ms [39]. (ii) Interactions with the field of view such as suturing the tissue. Slow manual deliberate movements employed here have the highest frequency of 12 Hz. Both live interactions and live view are governed by the effective visual frame update rate [25] and capture-to-display-delay [16]. In order to provide realtime interactive view, the visual frame rate must be equal or higher than 25 FPS [18,37] and the rate must be held constant to avoid delay jitter. Capture-to-display delay of less than 75ms for realtime manual task performance [37,38] is an acceptable criteria by looking at the Fitt's Task performance variation and error rates. For slow movements such as surgical movements, open loop movements or prolonged movement larger delay may be tolerated [37]. (iii) A third type of interaction is to annotate the scene or augment the scene by altering the scene content to facilitate better task performance. This is not technically feasible with SLM. This interaction too should observe the realtime requirements to be effective.

Both these, visual frame update and capture-to-display delay, vary with software processing and the underlying hardware rates such as video camera's frame rate and refresh rate of the display. Fig 2 depicts the components that contribute to latency and the latency pipeline is shown in Fig.3.

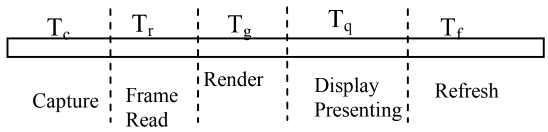

Fig.3 SDSM Processing Delay

Camera Delay (Tc) is the time it takes from the time the sensor is exposed to the scene to the time the digital image frame of the scene is made available to be read by SDSM software in to its buffers in the host system. The upper bound for this delay is governed by the camera frame rate that is usually published. For example, if the camera is generating a 720p frame at 60 FPS, then upper bound on the frame acquisition time is 16.6ms. Digital cameras use CCD or CMOS photovoltaic sensors and many components such as analog to digital conversion, shuttering, frame formation, filters, encoding, compression, interface protocol and resolution contribute to the achievable frame rate. A sensor must be chosen that provides enough resolution and frame rate minimizing shutter artifacts such as motion blur and smear. We can assume that $T_c = 1/f$ where f is the camera frame rate.

The second component of delay is in the reading of the image frame from the camera into the SDSM software buffers performed by the Image frame reader. This is read at specified intervals and when the frame is ready. The delay of this operation is governed by the transfer protocol being used such as USB 3.0, Fire Wire and LVDS, to name a few. Since the transfer of frames can overlap with the read operation if no compression is applied, this time can be ignored for uncompressed frame transfers. The transfer time is denoted by T_r.

The third component of the delay (T_g) occurs in the rendering pipeline which defines the colors for the pixels. This is dependent on the type of graphics hardware, the software organization being used and the degree of frame processing and augmentation required.

Fourth component of the delay is the delay in presenting the frame to the display hardware (T_q) waiting for the matching pair. For correct stereoscopic effect two frames, one for the left eye and another for the right eye must be presented. A stereo pair of the cameras that are currently active must each deliver a frame for presentation to the display hardware to happen.

For stereoscopic depth perception, two frames, one for the left eye and another for the right eye, must be presented. A stereo pair of the cameras that are currently active must each deliver a frame for presentation to the display hardware.

$$T_q = \max(T_r, T_g) \quad (1)$$

Let T be the reality to display (capture-to-display) delay.

$$T = T_c + T_r + T_g + T_q + 2/D \quad (2)$$

Where D be the display refresh rate, which is 120 Hz.

In addition, the rate of frames moving through the pipeline must be kept constant and it must be the maximum frame rate supported by the frame generator.

Another requirement is the faithful reproduction of reality in the display [19-22]. In SLM, reality was viewed through a magnifying glass with chromatic and geometric aberration corrected stereo. Reproduction of such reality is difficult since it requires forming the correct pair of images for stereoscopic viewing [19-22] and viewing image pairs on a flat screen in near field cause accommodation vergence conflict [34]. Non visual cues such as proprioception enhanced by hand-eye collocation may reduce the effect of accommodation vergence conflict. An extensive literature exists in the need and effect of hand-eye collocation [23-27].

Another related challenge is the realtime display of the effect of an operation. For example, using a realtime pointing device, the object of the physical world in the field of view is selected or annotated. Corresponding to the touch or annotation, a graphic is displayed. The display of the graphics must be in the correct video frame as it happened. Since video frames and position information are obtained as separate digital streams, their correct combination is a challenge.

Yet another challenge is that the impression of depth formed by binocular viewing, is subjective. Viewer also uses other cues such as haptic feedback to perceive the image depth [27-30]. Hence accuracy of the depth to the particular viewer is a challenge. The range of depth that can be presented to the viewer is a function of the inter-pupillary distance and the stereo fusion abilities which vary across individuals and with experience. In the case of SLMs, the oculars provided ability to adjust ocular separation to suit the inter-pupillary distance. For the SDSM, since the view is unrestricted, such adaptation of inter-pupillary distances does not happen. There must be provision to vary disparity based on the subjective depth perception abilities which can be stored as a surgeon preset.

3 Related Work

Two commonly used surgical visual aids are the surgical optical stereo microscopes used for higher magnification [1,2] and surgical loupe used for lower magnification [31]. A microscope has an objective lens for acquiring the image, an ocular for viewing the image. Between the ocular and objective, there are a number of optical elements for performing functions of focusing, zooming and optical quality enhancement. There are two main classes of optical microscopes [2,4] based on the design of the optical pathway: Greenough Stereo Microscopes that have two inclined objective lens assemblies and Common Main Objective Stereo Microscopes with a single common main objective lens assembly for both oculars. Surgeons view through the oculars to get a magnified view with depth perception through stereopsis. A surgeon must position his eyes just above the oculars of the microscope so that the exit pupil (the narrowest diameter area of the converged light exiting the eye piece before it spreads out and is about 2mm in diameter) be right in the pupil of the eye. Since human pupil is only about 3mm, this matching is often troublesome. This makes the margin of error for the head position and head tiny. If the alignment of the pupils are not exact, surgeons will only see a partial view of the operating site. Furthermore, the optical axis of the eyes must be aligned to the optical axis of the eye piece of the microscope. Each eye has just one position where it is comfortable for longer term – viz looking straight on. Other positions easily tire the eye.

While performing microsurgery, there is no reliable kinesthetic or haptic feedback as only small and slow movements are made to manipulate the delicate and small tissues using fine instruments. In addition, most of the monocular depth cues do not work with a fixed eye position and low depth of field. Surgeons must rely on the depth information obtained through stereopsis. Further eye muscle alignment is required to view through both the oculars to perceive depth through stereopsis. This is another reason for the fixed posture that causes fatigue to the surgeons.

Additionally, the microscope offers only low depth of field resulting in the surgeon making frequent adjustments of the microscope adding to the surgical time.

Another commonly used magnifying device is the surgical loupes [31]. This is a pair of magnifying glasses which can be worn and has a magnification range of 2X to 3.5X. The common surgical practice is to use loupes for surgeries requiring low magnification and optical stereo surgical microscopes for those requiring higher magnification. Surgical loupes are uncomfortable as slight movement of the head changes the magnified field of view and viewing away from the surgical site still produces magnified view.

Recently, a third option [20-26, 32, 33] of using an endoscope suspended from a stand is being attempted in Endoscope Assisted Microsurgery with some success in lab settings. Endoscope provides same magnification as the microscope with the exception of operating in a hand-eye separated fashion. The magnification is inversely proportional to the working distance since there is no separate zoom and focus functions in an endoscope. Typically, the working distance was reduced by an order of 10X in comparison with the microscope. Endoscopic images on the LCD display provided clear monocular views and had monocular depth cues. The eye-hand coordination was tedious and required training [32,33]. Though endoscope offered an alternative to remove the fixed head position, better depth cues are needed.

One method is to improve the endoscope with a pair of cameras and make the display to produce 3D. In addition, processing of the images, control of zoom and focus should be replaced in software. Our research has attempted to construct a software defined surgical stereo microscope system as a new alternative to SLM. It has only attempted to change the ocular, light processing elements into software components. Zoom and focus are still done in optical hardware controlled through software. Making them into software functions could be an extension of this research.

4 Software Defined Stereo Microscope

The basic SDSM was shown in Fig.2 which provides live stereo view, magnification and live interaction. It had only three hardware components, the camera, the display and the foot-pedal. An additional function is to annotate the video which is indicated in this discussion as a digital forceps.

All the functions of zooming, focusing, processing the image, stereo pair generation, rendering and annotation are done in software. An SDSM prototype used in surgery training is shown in Fig.4. The display which holds the camera is placed between the viewer eyes and the hands of the viewer. The workspace is at a distance of 150 – 200 mm when the cameras produce a width of field of 16 mm. The camera axes are held parallel to each other and is mounted to have two degrees of freedom. The translational motion adjusts the camera separation and the rotational motion allows the camera to be held perpendicular to the plane of the display surface.

The viewer wears shutter glasses synchronized with the display Vsync to give stereoscopic view of the work space. The work space view is magnified to have up to 9 times magnification matching the surgical microscope. A system that makes the SDSM may rely on threads of executions to use parallelism and concurrent execution. Since there are many IO and CPU events, separating the threads of execution can allow the CPUs be used by the threads of execution having completed its IO operation. Such threads of execution and their measured times are shown in Table 1.

Fig. 4 SDSM in Surgery Training

Thread	Function	Service Time
DIGITAL FORCEPS	Forceps thread, to provide coordinate information about the area of interest.	0.01ms per coordinate read.
PEDAL	Zoom, Focus control function read and Zoom Lens Control	1641.623 ms per pedal press
CAMERA	Reads video frames	0.084 ms
DREAD	Disk read of camera frames	0.0188 ms
DWRITE	Disk Write of camera frames	0.0183 ms
RENDER	Rendering thread that produces a render target view	1.045 ms
DISPLAY	Display of render target views	1.045 ms
WSTATS	Write Statistics	0.001 ms
RSTATS	Read Statistics	0.001 ms

Table 1 Threads of SDSM

Since most of these threads, except the RENDER thread, is IO bound there is enough CPU time for processing the compute requests on most modern computing platforms. The RENDER thread could be on a GPU. A quad core processor with Nvidia Quadro 4000 graphics card and 12 GB main memory system is used for the execution of the SDSM software components. The PEDAL thread processes the IO requests and issues IO requests to control the zoom and focus lens elements of the camera of choice. The camera is chosen by choosing the left and right pedals.

The render thread consumes the buffers created by the camera thread, the DREAD thread, and FORCEPS thread and submits to the graphics engine to produce a render view which is presented to the display by the DISPLAY thread.

DREAD and DWRITE threads supply or consume the frame video.

4.1. SDSM Delay Characterization

In Table 1, measurements of delay is shown against each thread. Using equation (1) and (2) and assuming 120 Hz display refresh rate, 30 FPS per camera, total delay from capture-to-display is derived from the delay presented in Table 1

$$T = 33.3 + 0.0188 + 1.045 + 1.045 + 1000*2/120 = 51.0304 \text{ ms}$$

The delay is significantly lesser than the tolerable limit specified. Hence it is a realtime device for the first two types of interaction and viewing. In Fig.6, the corresponding software data flow is shown. It is possible to optimize the rendering through parallel rendering using two graphics pipelines and reading multiple camera devices in parallel.

4.2. SDSM Frame Rate Characterization

The frame rate may be computed by reading the time stamp used in the video frames at the time of reading the frame. Subsequently, the frame rate is checked for its steady rate. It is done by instrumenting a counter in each of the execution thread. This is shown in the algorithm.

Algorithm SDSM Thread (K, L)

```
SetTimerTicks(k);
while (C = NextData() &&
           TimeNotElapsed())
                 ProcessData(C);
                 IncrementWorkCount(1);
 End while

if (TimeNotElapsed())
       k <-- max(k-1, K);
       IncrementIdleCount(1);
 End if
else k <-- min(k+1, L);
yield ( L-k);
```

The algorithm approximates the effect of thread scheduling, waiting for IO completion, and time quantum through the use of the scheduling variable k. Based on the experiments conducted in the microscope, certain processing time constants are determined for each of the processing threads. For each thread there is a minimum and maximum time determined and is denoted by (K, L) where K denotes the lower processing time and L denotes the upper limit on the processing time. The actual time of processing will vary between these two processing times. The initial values of these constants are set by using the video frame arrival rate, its processing time need from the shared compute resource and the rest of the time need. In this case, the frame arrival rate is 60 FPS with two active cameras at a time.

Using the work counter and the idle counter, the thread's effective usage may be estimated. The work counts must approach the production rates of the events that are processed in the thread to ensure realtime performance.

4.3. SDSM Live Interaction Characterization

Live interactive augmentation of the scene is modeled using the digital forceps. It is a magnetic position tracker registered with the workspace of the SDSM. It generates coordinate information at 240 Hz. Correctly rendering the position information with the video frame will ensure live interaction capability.

The forceps input results in a buffer of points denoting the forceps movement. The IO buffer will have all the movement data, but the FORCEPS thread would isolate movements that are meaningful. For example, movements after a button press so that it is not an accidental movement. The buffer data need to be used only when the frame is being displayed.

With the 120 Hz display, the forceps thread's buffer may be consumed in every two Vsync (the vertical sync pulse generated at 120 Hz). In 8.3 msec delay between the sync pulses, a large number of points could be read. However the forceps generates points by sampling the position at 240 Hz, which gives rise to process 2 points every frame refresh time. In fact between two points, no appreciable distance is moved as hand motion is of low frequency and amplitude. In two frame refresh time, 4 points are generated. They would match with the left and right frames of a video operating at 30 FPS. Within a frame delay (the time required to produce the next frame), 8 forceps positions are generated which will have matched the frame time stamp of the video frame. Utmost 8 points which are earlier than the time stamp of the video frame are selected for the frame. The time stamp of the previous video frame and the current video frame are compared to determine whether the position coordinates belongs to the current video frame. If the task is to point, the latest position coordinate that matches the time stamp is rendered into the video frame. If the task is to annotate, all 8 points that has matching timestamp are rendered in sequence into the video frame.

Another interaction is the interaction to manipulate the field of view. The PEADL thread is used infrequently, during setup and change in work space. The adjustments after the initial setup is incremental and is very minor. The PEADL thread collects the pedal pushes and outputs them as control commands to the command buffer maintained in the CPU. The Zoom Lens (embedded in the same thread as PEDAL) executes these commands to the camera zoom lens. Intermediate zoom and focus changes are seen in the video frames. This visual feed-back is treated as a 30 Hz interpretation of the control commands and it lasts for 48 – 50 frames before the operation is completed. The long delay is due to the slow movement of the lens movq 2 ction control. The stepper motor control currently used in the camera lens makes a micro step (256 microsteps makes a step, a step is 1.8 degrees of rotation of the rotor having 200 steps to make a revolution) in 200 msec. This causes the feeling that the foot pedal pushes respond slowly. It may be improved by faster controls with faster response times. We have selected the controls to keep the costs low in the prototyping phase. However, users did not experience significant delay as the multiple pushes they make can be executed as a single command of larger microstep. In order to avoid multiple detection of the pedal push, a sampling interval is defined which is set at 200 msec.

4.4. SDSM Characterization Experiment

The experimental setup (Fig.5) consists of a magnetic motion tracking device (Polhemus Liberty from Polhemus corporation) which has a position tracking resolution of 0.0013 mm with an accuracy of 0.8mm. Tracking can be programmed to sample the coordinates at 240 Hz. Since human movements are much slower than this, human movements can be captured for analysis. The motion tracking device is placed in the working space of the digital stereo microscope as shown in Fig. 7. This forms the digital forceps.

The microscope of Fig. 5 is programmed to show the realtime video of a cube being pointed at its corners. The experiment is conducted in two modes. In the first mode, the user sees the live video and interacts with the real object placed in the work space. The interaction is made a pointing task of pointing at the corners of the cube. There is some degree of haptic feedback due to the touching of the forceps tip on the cube. In the second mode, the user touches the corners of the cube using the video being played. The live view is augmented with the frames of the video having the cube in the workspace. The prerecorded video is blended in the graphics engine to create the augmented image.

A user may adjust the microscope using the foot pedal to gain focus to the cube placed in the work space and then touches the corners of the cube through the forceps. The user may adjust the microscope by using the foot pedal. However, in the second mode, foot pedal pushes will be to see the forceps tip focused. In summary, two fixed rate data streams, one from the pair of cameras that are currently active in the array of cameras and another from the digital forceps, and a variable rate data stream belonging to the PEDAL activities form the work load of the SDSM.

All the threads are instrumented with the event counters denoting the activity of the thread and idle invocations of the thread. If the system preserve the frame rate of the video, then it will have a constant activity count ratio (activity count / total event count).

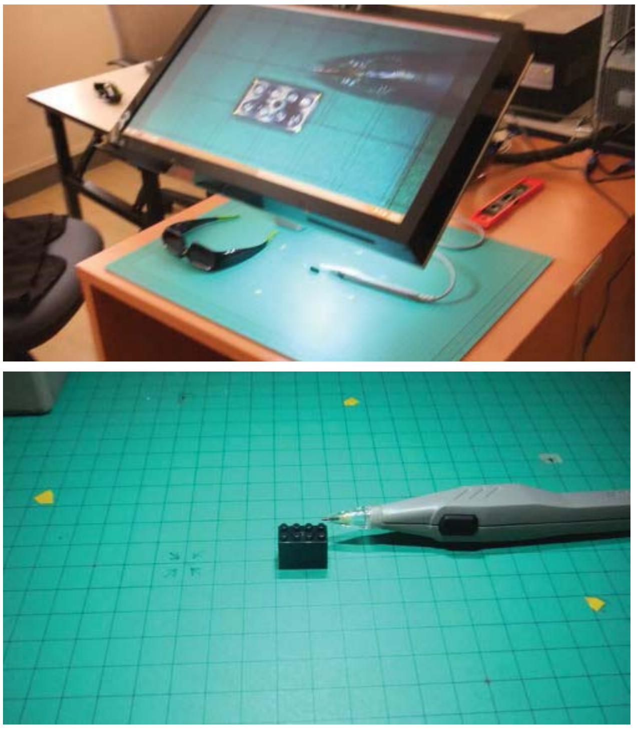

Fig.5 Experimental Setup

The experiment assesses the realtime nature of the SDSM to be compatible with the SLM counterpart. Hence the number of subjects that performed the movement is not important. A number of repetitions of the experiment is made to account for the natural load by other processes running in the system.

Using the RSTATS and WSTATS, the event counters are recorded against time interval at each change. When a new event counter is updated, the RSTATS will read all the activity and idle counter values and record them. This record is used for the analysis of the system.

5 Results and Conclusion

In Figure 6, the realtime nature of the processing is illustrated. The frame rate from both the cameras combined is at 60 frames/sec and the frame rate from the digital forceps is at 240 frames/sec. The 60 Hz camera frames gets rendered at the same rate. While presenting to the display, two video frames are presented simultaneously. Hence the display rate is at 30 frames/sec. Hence the display thread indicates moving the frame from the graphics pipeline to the display buffer queue for stereo presentation. The refresh is at 120 Hz and is not shown in the graph. A third data stream PDEAL is invoked by the pedal pushes which are results of the manual adjustments of focus and zoom while marking the cube corners.

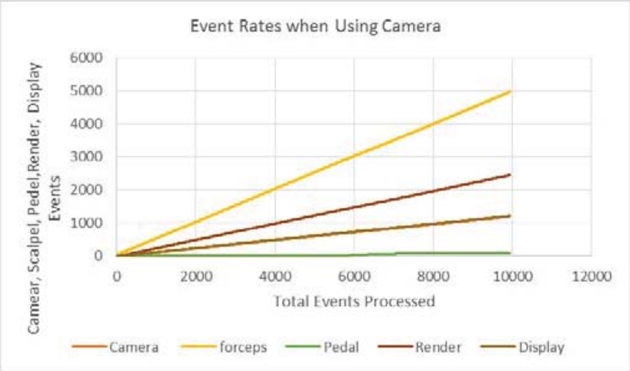

Fig.6 Realtime Events Processing on Live Interaction

The data rates for CAMERA, FORCEPS, RENDER and DISPLAY are held steady by the fixed slope of the line indicating the realtime processing of the frames. The data rate of the pedal events which would invoke another thread of execution is shown expanded in Fig. 7.

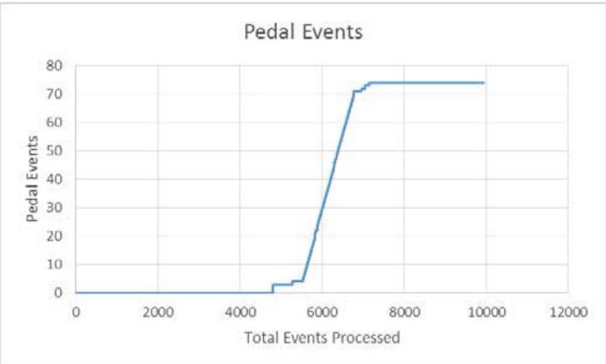

Fig.7 Foot Pedal Events

As seen from Fig.7, the configuration using pedal events lasted for roughly 2,000 events but did not cause any adverse effects to the processing of the camera or forceps events which are at high rates because their respective processing rates does not have any deviation. The slope of the event counter line is an indication of the relative number of events. There is one anomaly seen in Fig. 6. Though the rate of the FORCEPS line is four times the rate of the CAMERA lines, agreeing with the raw rates are 240 Hz and 60 Hz, the rate of the render line is almost double that of the camera. The rendering pipeline is in the GPU and does not check whether its inputs are changed or not. Hence it is allowed to render the previous frames again, though that did not produce useful work. Display is at half the frame rate as render, since it waits for two frames to get rendered. Hence, the rendering engine has the compute power to render at a faster speed than the current frame rate. In Fig.8, the render and display as functions of captured frames is shown. For every captured video frame almost two frames are rendered. This shows the additional render capacity available. Since the relationship is still linear, there is no significant delay jitter in the system and measured delays have very less variability.

The test is repeated with the video played from the attached hard disk. The results are shown in Fig.8. However the forceps is randomly moved since we are still progressing with the blending of the video frame and the realtime camera frame. The forceps will produce data at 240 Hz irrespective of the movements. The result is shown in Fig.8. From the figure it is seen that the video rates and the forceps rates are held constant. It may appear from the Figure that the video rate slightly reduced towards the end of the experiment. But it is due to that fact that the video is completely played and there are no video events to be counted. But the refresh of the video would continue. Hence there is a slight bend in the camera, render and display lines.

These results show realtime viewing and interactions are possible using SDSM. The capture delay can be at most 33ms. The camera read time is 0.084 ms and the render time is 1.045 ms. The display time is 1.03 ms. Bulk of the display time is to wait for the 2nd frame of the stereo pair to arrive. So it is assumed that both cameras are almost in sync.

Future enhancements include better time stamping of the cameras and the digital forceps, better synchronization of the devices and better registration of forceps generated coordinates with the workspace.

In conclusion, the SDSM is found to have good realtime response even though it is constructed in an inexpensive and flexible manner than the SLM. It rectifies the major limitations found in SLM and can be used in surgery training and practice. We will be conducting trials in surgery training using training systems before we test the utility in surgery practice inside the operating room.

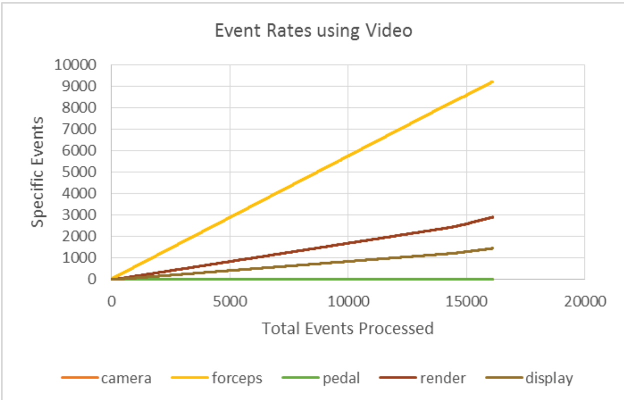

Fig.8 Realtime Events Processing on Video Annotation

Acknowledgements

We acknowledge support of the SPRING TECS POV grant TI/TECS/POV/12/2 which provided the funding for further product development based on the technology developed in this research. We also acknowledge Nidhin of Digital Surgicals for his help in conducting the experiments.

References

[1] G. Seward, Optical design of microscopes. 2010: SPIE.

[2] K.P. Zimmer, Optical designs for stereomicroscopes. in International Optical Design Conference. 1998. International Society for Optics and Photonics.

[3] K. Yamaji, "Design of zoom lenses," in Progress in Optics, E. Wolf, ed., (North Holland, Amsterdam, 1967), pp. 105-170.

[4] Schnitzler, H. and K.-P. Zimmer. Advances in stereomicroscopy. in Optical Systems Design. 2008. International Society for Optics and Photonics.

[5] M. Frey, Endoscopy and Microsurgery. 2001: Springer

[6] M. Levoy, R. Ng, A. Adams, M. Footer, M. Horowitz, "Light Field Microscopy", ACM Transactions on Graphics 25(3), Proc. SIGGRAPH 2006, pp. 1-11.

[7] J.K. Rappel, T.C. Leong, A. Lahiri, “A System for Analyzing Surgical Motion under a Surgical Visual Aid”, 9th Int. Nat. Conf. on Information, Communication and Signal Processig, Dec. 2013.

[8] Paul, and F. Kishino. "A taxonomy of mixed reality visual displays." IEICE TRANSACTIONS on Information and Systems, Vol. 77, No.12, pp. 1321-1329, 1994.

[9] R.J.Teather, R.S. Allison, W. Stuerzlinger, “Evaluating Visual/Motor Co-Location in Fish-Tank Virtual Reality”, IEEE Toronto International Conference - Science & Technology For Humanity, 2009

[10] O.Boyaci, A. Forte, S. A. Baset, and H. Schulzrinne,"vDelay: A Tool to Measure Capture-to-Display Latency and Frame Rate",11th IEEE International Symposium on Multimedia, 2009.

[11] R. Held, A. Efstathiou, M. Greene, “Adaptation to Displaced and Delayed Visual Feedback from the Hand”, Journal of Experimental Psychology Vol. 72, pp. 887–891, 1966

[12] A.S. Kornheiser, “Adaptation to Laterally Displaced Vision: a Review”, Psychological Bulletin Vol. 83, pp. 783–816, 1976

[13] I.S. MacKenzie, and C. Ware, Lag as a determinant of human performance in interactive systems. ACM CHI’93, 488-493.

[14] M.R. Mine, Characterization of End-to-End Delays in Head-Mounted Display Systems. University of North Carolina at Chapel Hill, 1993.

[15] A. Pavlovych, and W. Stuerzlinger, The tradeoff between spatial jitter and latency in pointing tasks. In Proc. EICS '09. ACM Press (2009), 187-196.

[16] R. Teather, A. Pavlovych, W. Stuerzlinger, and S. MacKenzie, Effects of tracking technology, latency, and spatial jitter on object movement, IEEE Symposium on 3D User Interfaces 2009, 43-50.

[17] C. Ware, and R. Balakrishnan, Reaching for objects in VR displays: lag and frame rate. ACM TOCHI 1, 4, 1994, 331-356.

[18] Y.C. Jessie Chen and J. E. Thropp, Review of Low Frame Rate Effects on Human Performance, IEEE Trans. on Systems, Man, and Cybernetics—PART A: Systems and Humans, VOL. 37, No. 6, Nov 2007, 1063-1076

[19] C. Ware , K. Arthur , Kellogg S. Booth, Fish tank virtual reality, Proceedings of the SIGCHI conference on Human factors in computing systems, p.37-42, April 24-29, 1993, Amsterdam, The Netherlands

[20] M. Wentink, J.J. Jakimowicz, L.M. Vos, D.W. Meijer, P.A. Wieringa, “Quantitative Evaluation of Three Advanced Laparoscopic Viewing Technologies: a Stereo Endoscope, an Image Projection Display, and a TFT Display”. Surgical Endoscopy Vol. 16, pp. 1237–1241, 2002

[21] A. Cushieri, "Visual display technology for endoscopic surgery", Minimally Invasive Surgery and their Allied Technology, Vol. 5, pp.427-434, 1996.

[22] G.B. Hanna, S.M. Shimi, A. Cushieri, “Task performance in endoscopic surgery is influenced by the location of the image display”, Ann. Surgery, Vol. 227, pp. 481-484, 1998.

[23] P. Breedveld, "Observation, manipulation, and eye-hand coordination problems in minimally invasive surgery", Proceedings of the 16th European Annual Conference on Human Decision Making and Manual Control, December 9–11, Kassel Germany, 1999

[24] O. Ben-Porat, M. Shoham, J. Meyer, “Control Design and Task Performance in Endoscopic Teleoperation”, Presence: Teleoperators & Virtual Environments Vol. 9, 256–267, 2000

[25] D. K. Boer, D.T.Jong, J. Dankelman, D. Gouma, Problems with Laparoscopic Instruments: Opinions Of Experts. Journal of Laparoendoscopic & Advanced Surgical Techniques Vol. 11, pp. 149–155, 2001

[26] R. R. Berguer, D.L. Forkey, W.D. Smith, “Ergonomic Problems Associated With Laparoscopic Surgery”, Surgical Endoscopy, Vol. 13, pp. 466–468, 1999

[27] Arsenault and C. Ware, "Eye-hand co-ordination with force feedback," in Proceedings of the SIGCHI conference on Human factors in computing systems The Hague, The Netherlands: ACM, pp.408-414, 2000.

[28] J.P. Rolland, F.A. Biocca., T. Barlow, A. Kancherla, “Quantification of Adaptation to Virtual-Eye Location in See-Thru Head-Mounted Displays”, Virtual Reality Annual International Symposium (VRAIS 1995), pp. 56–66 , 1995

[29] R. Arsenault, C. Ware, “The Importance of Stereo and Eye-Coupled Perspective for Eye-Hand Coordination in Fish Tank VR”, Presence, Teleoperators & Virtual Environments, Vol. 13, pp. 549–559, 2004

[30] D. Swapp, V. Pawar, C. Loscos, “Interaction With Co-Located Haptic Feedback in Virtual Reality”, Virtual Reality Vol. 10, pp. 24–30, 2006

[31] J.M Baker, and R.A. Meals, A practical guide to surgical loupes. The Journal of hand surgery, 1997. 22(6): p. 967-974.

[32] R.J. Franken, S.C. Gupta, J.C. Banis, S.V. Thomas, J.W. Derr, S.A. Klein, M. Kon, J.H. Barker, “Microsurgery without a microscope: laboratory evaluation of a three-dimensional on-screen microsurgery system”, Microsurgery vol. 16, No. 11, pp. 746-751, 1995

[33] M. El-Shazly, M. El-Sonbaty, A. Kamel, M. Zaki, A. Frick, and R. Baumeister, “Endoscopic-assisted microsurgery: microsurgery in the new millennium? A comparative experimental study,” British Journal of Plastic Surgery, vol. 56, no. 1, pp. 37-40, 1//, 2003.

[34] D.M. Hoffman, A.R. Girshick, K. Akeley, M.S. Banks, “Vergence-accommodation conflicts hinder visual performance and cause visual fatigue”, Jl. of Vision, Vol. 8, No. 3, pp. 1-30, 2008

[35] R. K. Reznick, H. MacRae,"Teaching Surgical Skills — Changes in the Wind", The New England Journal of Medicine, Vol. 355, pp. 2664-2669, 2006

[36] J. K. Rappel, H. Kazemi, T. Poston, B. Hai Lim, E. Burdet, and C. Leong Teo, “Assessing suturing techniques using a virtual reality surgical simulator,” Microsurgery, vol. 30, no. 6, pp. 479-86, Sep, 2010.

[37] I.S MacKenzie, C. Ware, Lag as a determinant of human performance in interactive systems, Proceedings of the ACM Conference on Human Factors in Computing Systems - INTERCHI '93, Newyork, 1993, 488-493.

[38] C. Ware and R. Balakrishnanan, Reaching for Objects in VR Displays:Lag and Frame Rate, ACM Trans. on Computer-Human Interaction, Vol. 1, No. 4, December 1994, Pages 331-356.

[39] Miller, R. B. (1968). Response time in man-computer conversational transactions. Proc. AFIPS Fall Joint Computer Conference Vol. 33, 267-277.