American Journal of Applied Sciences

Review

Remote Surgeon Hand Motion and Occlusion Removal in Mixed Reality in Breast Surgical Telepresence: Rural and Remote Care

[1]Krishna Shakya, [1]Suman Khanal, [1]Abeer Alsadoon, [1]P.W.C. Prasad, [2]Jeremy Hsu, [2]Anand Deva, [1]Manoranjan Paul and [3]A. Elchouemi

[1]School of Computing and Mathematics, Charles Sturt University, Sydney, Australia
[2]Faculty of Medicine and Health Sciences, Macquarie University, Australia
[3]Walden University, USA

Article history
Received: 22-09-2018
Revised: 15-12-2018
Accepted: 09-01-2019

Corresponding Author:
P.W.C. Prasad
School of Computing and Mathematics, Charles Sturt University, Sydney, Australia
Email: cwithana@studygroup.com

Abstract: Mixed reality in surgical telepresence has not been practically implemented as yet and research done, until now, is only theoretical. Aiming to implement a proposed new solution by merging virtual and augmented realities and bringing about improvements by removing occlusion and the noise caused by hand motions during surgery. The proposed system consists of an extended Camshift algorithm for the detection of the remote expert surgeon's hand during surgery and enhancement of an image synthesization algorithm to remove the occlusion by including two Red-Green-Blue-Depth (RGBD) cameras. The experimental results from 10 breast samples reduce the average accuracy error of the video frame image overlay from 1.44 mm to 1.28 mm with average processing time enhanced to 76 sec compared to the state-of-the-art method. The proposed system provides a satisfactory outcome in terms of accuracy and operating time, which allows surgeons in local and remote sites to collaborate more effectively.

Keywords: Virtual Reality, Mixed Reality, Diminished Reality, Augmented Reality, Telesurgery, Visualization, Tele Collaboration, Surgical Telepresence

Introduction

Any surgery is requiring expertise surgeon that has good knowledge and skill (DeSantis *et al*., 2017). In rural areas, the surgeons have less experience than surgeons in the city, which has made it necessary in the past for expert surgeons to travel to local sites; rural areas, to perform surgery. Furthermore, in the surgery preparation stage, the Computed Tomography (CT) scan reports of the patient should be produced and had to be shared with the expert surgeon for pre-planning (Bruellmann *et al*., 2013). This is costly and time-consuming process was later replaced by a video-guided monitoring system, the System for Tele-monitoring with Augmented Reality (STAR), which facilitated remote collaboration between local site; surgery site and remote site; expertise surgeon site (Andersen *et al*., 2016; Guo *et al*., 2014).

Mixed Reality (MR) has allowed expert surgeons to directly assist with their virtual hand in real live surgery at local sites (Shenai *et al*., 2011). One of the major components of this system is Augmented Reality (AR) which combines a real-world scene with a virtually rendered object through superimposing one upon the other. It provides a 3-Dimensional (3D) view by merging computer-generated 3D images with real environments captured in real time through camera recorded 2D videos (Wang *et al*., 2017; Milgram and Kishino, 1994). The remotely located expert surgeon can interact with the less expertise surgeon situated at a local site, acting as physical proxy in a Human-Computer Interaction (HCI) scenario (Przybyło, 2010; Liu *et al*., 2016).

In a Mixed Reality (MR), if the virtual object is added in the real environment, it is augmented reality and if the real object is added in a virtual environment, it creates augmented virtuality (Baker *et al*., 2015; Diaz *et al*., 2017; Pin, 2015). To visualize compound output in Mixed Reality, innovative implementations of devices like Google Glass (GG) (Chang *et al*., 2015), Virtual Reality (VR) or stereoscopic camera are being used by both, the local and the remote surgeon. In general, most mixed reality systems use proprietary software that does not reveal how the system was implemented.

(a) (b) (c)

Fig. 1: (a) Augmented Video (b) Virtual Presence (c) Mixed Visualization; [Above images are downloaded using Google Search engine, the images are free to use, share or modify, even commercially]

Mahesh in (Shenai *et al*., 2014) has described LD-Virtual Interactive Presence (VIP) as a novel paradigm for collaboration where multiple surgeons are digitally combined into a composite field. Similarly Anderson *et al*. (2016) has implemented a platform named VIPAR for virtual assistance but has only provided theoretical knowledge. Figure 1a depicts how the rendered CT scan data is superimposed on the human body during surgery in AR. In Fig. 1b, it shows how the expertise surgeon in remote site can interact with the surgeon of rural area in local site as a physical proxy of the mentoring patient using camera attach to its side. In Fig. 1c, we can see how the local surgeon is able to visualize merged mixed reality of the brain with a detail view of both augmented video and remote surgeon hand.

The main purpose of this paper is to create a MR with a 3D augmented video received from a local site showing the segmented virtual hand of the remote expertise surgeon. The idea here is to share mutual space between local and remote sites resulting in a mixed reality. the aim of this paper is to propose a new solution for the use of MR in surgical telepresence. A further aim is to implement an improved solution by merging virtual and augmented realities and bringing about improvements by removing occlusion and the noise caused by hand motions during surgery. The section called “Literature review” presents the collection of current solution and also describes the model of the stat of art method Shen *et al*. (2012). The section followed by "Proposed Model" include the flowchart and pseudocode of the proposed formula. The section followed by called “Results” presents the results of the state of art and proposed solutions. The last section called “Discussion” describes the comparison between current and proposed system results and provides the conclusion for the paper.

Related Work

Shenai *et al*. (2011) has introduced Virtual Interactive Presence and Augmented Reality (VIPAR) system, which consists of two binocular stereoscopic cameras that capture ‘view fields’ on both local and remote stations. Video analysis of 3 separate video clips yielded a mean compositing delay of 760±606 msec (when compared with the audio signal). Image resolution was adequate to visualize complex intracranial anatomy and provide interactive guidance. However, this method failed to identify smaller anatomical structures and depth mapping due to poor camera resolution, where depth mapping is the position of the remote surgeon’s hand in relation to the surgical field. A lack of depth mapping has the potential to create a significant problem in the surgical visual field if the view is distorted by bleeding.

To create a seamless video composition, Shen *et al*. (2012) has proposed a hybrid blending method which combines two video frames and generates a composite video. This technique is applied to the computer animation field where two different background images are merged. It uses a boundary-aware mesh generating algorithm in video frames to optimize the identification of 3D mean value coordinates and further implements alpha blending (a combination of alpha matting and a gradient domain algorithm) to refine the composition result. Hybrid video composition removes the artifacts, smudging and sawtooth from the final combined video. This solution has shown great reductions in processing time as it took just 120 sec for 100 video-frames of 1280*720 resolution. However, this system failed to detect foreground objects moving quickly and did not address occlusion removal after merging since it does not need to view occluded background images after merging. Nevertheless, taking these features into the proposed system will help to merge videos in real time.

Andersen *et al*. (2016) implemented a new mentoring method using the System for Telemonitoring with Augmented Reality (STAR) system. It is based on a transparent display which avoids focus shifting by placing annotations sent by the remote surgeon directly into the surgical field. Participants using STAR completed the 2 tasks with less placement error (45% and 68%) and with fewer focus shifts (86% and 44%),

but more slowly (19% for each task). Using this method has provided a way to monitor from a long distance but the expert surgeon can only communicate through Annotation and the audio channel which is not sufficiently interactive. It has Annotation placement errors, due to which the annotated area and anatomy could differ during operation. Therefore, due to all these limitations, this system does not offer opportunities for either improvement or use.

Similarly, Shenai *et al*. (2014) developed a Virtual Interactive Presence (VIP) system, which consists of remote and local stations. The local station captures a video feed from the site of a procedure and sends it to a remote station, where the capture participant's hand interacts with the video. The result is sent back to the operating field. Image processing in then implemented in C and Python's programming and merged using VIP. Image processing is a method of performing several operations like noise reduction, geometric transformations, image registration, etc. It is done to extract an important information from the image, for instance, visualizing (Observing object that is not clearly visible in the original image by sharpening and restoring), Image recognition (Distinguish an object within an image).

The researchers have solved the problem of a spatial relation between local and remote fields according to scale and space using a VIPAAR spatial alignment algorithm, i.e., the spatial relation of the surgeon's hands in relation to the surgical field. However, the work does not provide details of algorithms or a technique for merging. Therefore, this method is of no further interest.

Similarly, Shen *et al*. (2012) have proposed mixed reality interaction using a real-time Natural User Interfaces (NUI) engine for interaction, using a human hand in mixed reality. They have focused on mixed reality-based games and touch-less application. A video of a hand is capture using a web camera and they implement real-time processing using Graphic Processing Unit (GPU) programming. Alpha blending is used to create a mixed environment based on the blend factor alpha which signifies the opacity range to blend the hand with any background frame. To remove the noise, they have used dilation and erosion mask shifting. This is useful for the proposed solution and the concept of alpha blending has been implemented.

Walton *et al*. (2017) has proposed a way to render virtual objects in real time in mixed reality in the multimedia field. It has used a RGBD and a small fisheye camera to create environment maps, which generate a 3D geometric model of the scene, using a dense 3D model from an RGBD camera using a Simultaneous Localization And Mapping (SLAM) algorithm. The limitation of this solution is that geometric accuracy is higher when the virtual object is closer to the surrounding real scene. Due to this limitation, the system can be difficult to implement. Also, the object rendered is completely virtual, which is not the case in the proposed system.

Przybyło (2010) has proposed an extensive Continuously Adaptive Mean Shift (CAMSHIFT) algorithm, which detectsa hand in motion even during fast movement. This system converts an RGB image to a Hue-Saturation-Value (HSV) and collects motion information from the luminance components. However, tracking the object may be weaker due to an insufficient frame rate and the low sensitivity of the algorithm when only the finger is moved rather than a whole hand. Nonetheless, the extensive algorithm can be taken into consideration for implementation and has been included in the proposed solution.

Surgical guidance can be a significantly benefit MR Kersten-Oertel *et al*. (2013) using implementation of mixed reality image-guided surgery with different visualization techniques based on a Data, Visualization processing, View (DVV) taxonomy (Walton *et al*., 2017), whereby 'a set of data about the patient is sent to the end user for image processing and visualizing. The result is a new defined perception location, a display technique to create the visualization and an interactive tool which deals with design. Here researchers have identified 'Surface' as a better data representation, colour coding as best visualization processing and found psychophysical and human factor studies of visualization and interaction technique were the limitation.

Habert *et al*. (2017) introduced multi-layer visualization in the surgical scene in mixed reality, which improves the surgeon's visualization by recovering the occluded part through making the occluding object transparent and creating a different layer during surgery. A volumetric based image synthesization technique was used to recover the occlusion created by the expert hand during hand detection, by using multiple cameras from a different viewpoint. The experimental results show a recovery accuracy of between 45 to 99%. Two RGBD cameras on both sides at a diagonal angle are also used to calculate the depth image and wide-angle video image. The recovery of the occluded area is highly dependent on the RGBD camera settings, i.e., the higher the surgeon's hand with respect to the anatomy (Bradski, 1998), the higher the occlusion recovery on both sides. Using a synthesization technique provides user-adjustable multiple layer visualizations between hand, surgical instrument and patient. Adding this feature will solve occlusion problems in the proposed system.

State of the Art Solution

This part defines the currently existing system and limitations (highlighted in red Fig. 2). The model

proposed by Shen *et al*. (2012) was applied for combining videos using a hybrid video composition technique from the cinematography and computer animation field. The model provided a processing time of the video composition of 120 sec with a frame size of 1280*720 in 100 frames. The model is divided into three different sub-sections (Fig. 2) namely Source, Target and Composite Video. This model blends a given source video into a target video with high-performance in minimal computing time.

Target: The target in the given state-of-art model is considered as foreground image, which is extracted to merge with the background image i.e., source. A 2-Dimensional (2D) video is taken from the offline source, which means it is not a real-time video captured from a camera and then frames are extracted in the form of a series of frame cubes. These frame cubes are merged with the foreground image cubes from the composite video section using video composition.

Source: Similarly, the source is the background image where the target image is overlayed. Here, the frame cubes extracted from the offline video with a single camera starts capturing a video, which is divided into cubes as in the above target section. The composition of multiple image segments is called the frame cubes. Frame cubes mean the frames are extracted from the offline video that captured using single video camera and then will be divided into frame cubes. The composition of multiple image segments from the captured video is called frame cubes. To blend the source video with the target video, a source patch Region Of Interest (ROI) is marked. Where ROI is area of interest is the specific area within the source frame cubes where the target is composite. ROI in consecutive frames is automatically tracked using an optical flow algorithm. With the help of optical flow algorithm, it helps to identify the ROI of image objects between two consecutive image frames (Correia and Campilho, 2002). The boundary region of the source video is marked so that said source video can overlay in the target video, which is further achieved in the composite video subsection.

Composite Video: In this part, a hybrid video composition using mesh generated algorithm is used for generating a mesh of source and target video in the series of source and frame cubes. To generate a mesh, at first ROI of the foreground image is selected and 3D Mean-Value-Coordinate (MVC) of a source and target frame cubes are calculated. Where, 3D MVC is a process to calculate third coordinate of 3D image known as *z*-index of existing 2D image which only consists of *x* and *y* co-ordinate to create a 3D image. Once the mesh is generated, it is blended using hybrid alpha blending which combines alpha blending and gradient domain. The current system has enhanced alpha blending to hybrid blending which is the combination of mesh generation and alpha blending. It refines the generated composite video by removing the artefacts in less processing time with high accuracy.

Nevertheless, this model has mentioned few of the limitation. Firstly, the given algorithm could not handle foreground object if it is moving very fast and widely. When the foreground image is fast, it is hard to track and cage in a given mask. Secondly, this composition algorithm is not very effective when the boundary of an object keeps on changing very frequently due to the movement of object. When the boundary keeps changing the algorithm fail because of the conflict, it creates due to frequent change of ROI. Lastly, it has not mentioned anything about the recovery of an occlusion occurred after merging of a video. It requires strong image processing techniques for better occlusion recovery.

Hybrid Blending: Once the source and target video frames are extracted, ROI is selected in following frames and 3D mesh of video is generated using mesh generated algorithm. Then the generated mesh is blended using hybrid blending to create the composite video. Hybrid blending method is used to refine the result and eliminates the sawtooth and discoloration effects. The Hybrid-Blending equation that blends and refines the composite videos in the state of art method is presented in Equation 1 (Shen *et al*., 2012):

$$I_{final} = \alpha * \phi + (1 - \alpha) * B \quad (1)$$

Where:
I_{final} = Final composite Image
α = Opacity value
B = Background Image
ϕ = Constrain the blending degree

This system has proven accuracy with an overlay error of 1.44 mm and processing time of 74.9 sec. in 50 frames. It can also be improved further using the proposed technique to 1.28 mm to combine two videos in real time. The limitation to this system is that it could not detect the foreground frame cubes when it is in high motion as well as it does not recover the occluded part. The region of interest is selected in first frame cube of source video and it is tracked automatically in the consecutive image cubes, but the foreground image in high motion can cause blur in final composite video. The hybrid-blending method is used to merge and align two videos in real time is presented in Table 1 and flowchart of it is in Fig. 3.

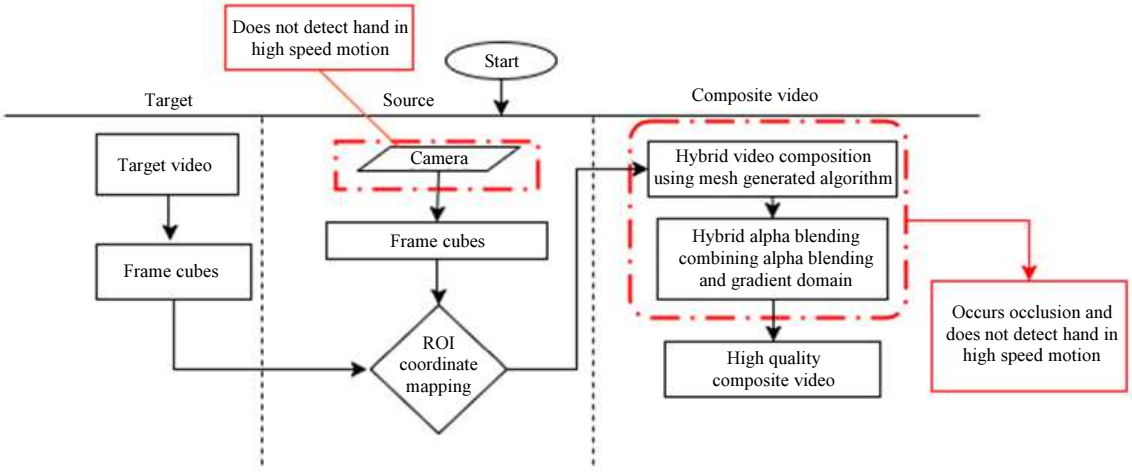

Fig. 2: Sate of Art video composition technique Shen *et al*. (2012); [The red border refers to the limitation of it]

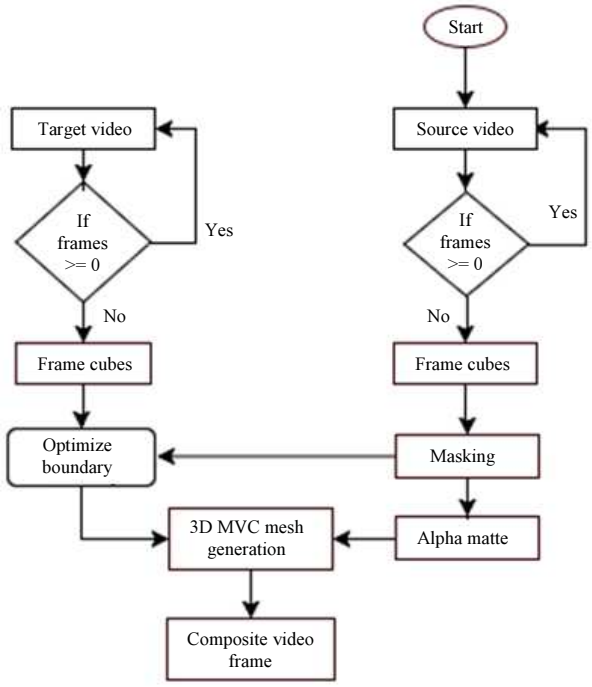

Fig. 3: Flowchart of video composition and hybrid blending

Table 1: Hybrid video composition method using mesh generated algorithm

Algorithm: State of Art Input: Augmented 3D video frame and hand video frame Output: Composite video
BEGIN Step-1: Camera captures the source image sequence and target image sequence is selected. Step-2: Region of Interest is selected to map the coordinate with target and remote source cubes. Step-3: Calculate hybrid blending using equation 1. END

Proposed Solution

Before proposing an ultimate solution, ranges of techniques related to mixed reality were reviewed for this article. Those articles had offered different solutions, which have both pros and cons. All of them were critically analyzed based on different factors. Although practical researches in surgical telepresence domain have not been done, a pre-existing technique for mixed reality has been reviewed to assess the feasibility of MR to perform long-distance surgical assistance. Out of all, one model presented by Shen *et al*. (2012) has been select as the best solution and the proposed solution is used as a base for it. The proposed solution (Fig. 4) takes necessary features from the state of art solution that proposed by Shen *et al*. (2012) and has some good features of second and third best solution for image synthesisation proposed by Habert *et al*. (2017) and extensive Continuously Adaptive Mean Shift (CAMSHIFT) algorithm proposed by Przybyło (2010) respectively to address the limitation we found in Shen *et al*. (2012) model. Furthermore, another feature is also added from system proposed by Habert *et al*. (2017) by implementing two Red-Green-Blue-Depth (RGBD) cameras (4 k camera (3840×2160)) to improve depth image and a wide-angle video image in a remote site. Image synthesisation is used to recover occluded area created by surgeon's hand and extensive Continuously Adaptive Mean Shift (CAMSHIFT) is used for detecting high motion hand even in the complex background. Equation 2:

$$I_{layers}(p)=\begin{cases}\alpha I_c(p)+\beta I_b(p)+\gamma I_{xray}(p) if\ \ p\epsilon\ foreground\\ (1-\delta)I_b(p)+\delta I_{xray}(p) else\end{cases} \quad (2)$$

Where:

I_b = First RGBD camera color image

I_c = Second RGBD camera color image extract which could be overlaid with transparency
I_{layer} = Multilayer image created by blending all the layers where $\alpha, \beta, \gamma, \delta$ = [0,1], $\alpha + \beta + \gamma = 1$ Where α =Alpha, β =Beta and γ = Gamma

This paper has proposed an enhanced hybrid blending synthesization equation to merge, refine and recover the occluded area parallel with high accuracy and proposed extensive Continuously Adaptive Mean Shift (CAMSHIFT) to detect a moving hand even in high speed. We have implemented it in breast surgery.

Algorithm History

The Alpha blending was introduced by Alvy Ray Smith on August 15, 1995. The technique was introduced to combine any two digital images with full alpha matte. The algorithm does not merge nicely when the source and target frame color is same. Hence, Shen *et al*. (2012) proposed hybrid alpha blending to refine sawtooth and discolor in composite video. Gary R. Bradski originally introduced CAMSHIFT algorithm in 1998 to detect and track human face and object. Later, it was modified to deal with dynamically varying color probability distribution capture from video frame sequence. The modified algorithm named as the Continuously Adaptive Mean Shift (CAMSHIFT) algorithm (Habert *et al*., 2017). The various researchers have implemented and modified CAMSHIFT in several sectors as per their need. Similarly, Przybyło (2010) has proposed extensive CAMSHIFT algorithm due to the fact they found the human visual system not only integrates the color but also motion.

Area of Improvement

Shen *et al*. (2012) proposed a hybrid blending algorithm. This algorithm consists of mesh generation and alpha blending algorithms. Our proposed algorithm consists of generation algorithm and the *Modified Volumetric Based Image Synthesization.* Alpha blending has enhanced by us. Proposing a new video-based blending technique for merging virtual hand and surgical scene with 3D mean value blending with Hybrid alpha blending. While merging the source and target video i.e., virtual hand with surgical scene the result might lead to artifacts. When the texture of the source and target are different, smudging and artifacts appear on composition video. Here the solution hybrid blending is used to remove the problem. We combine alpha matte with alpha blending to provide better result.

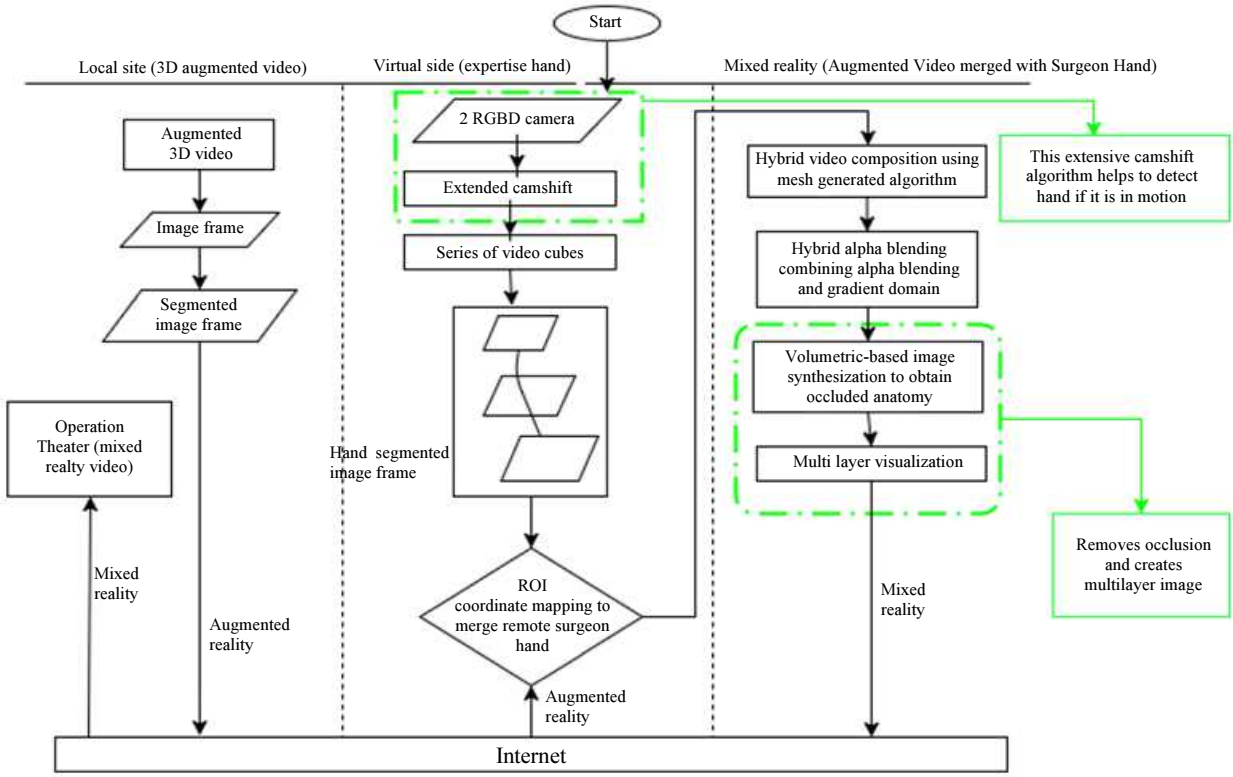

Fig. 4: Proposed Mixed Reality System using extensive CAMSHIFT and Enhanced Volumetric Synthesisation; [The green borders refer to the new parts in our proposed system]

The area of improvement is focused on reducing the occlusion occurred by the surgeon's hand motion. To improve the overall system in the proposed method, hand occlusion is undertaken first. The generated composited video of the both local and remote site from state-of-art does not consider about the recovery of the occluded area. The occlusion occurred by surgeon's hand and medical instruments are removed by image synthesisation technique in our proposed system. The synthesisation technique helps to recover the occluded area by merging the images captured by the 2RGBD camera from the side angle view. In state-of-the-art, Hybrid blending was used to combine videos from two different sources and it is now integrated with the modified image synthesisation technique by removing the unwanted *x*-ray image layers from the algorithm proposed by Habert *et al*. (2017) which will help to solve remove the occlusion.

Local Site Environment (Surgery Side)

This site generates 3D augmented video from the 3D CT scan data and images of patient anatomical structure. The 3D augmented video is converted to 3D augmented video frame and sends it to the remote site using standard Internet connection.

Remote-site Environment (Virtual Side) using Extensive CAMSHIFT Algorithm

Two RGBD cameras are used to capture the real-time video of expertise hand from the different angle. The depth (D) in RGBD represents the depth, which will maintain the depth perception during visualization in the surgical field. Videos captured by two cameras are processed by extensive CAMSHIFT for the motion detection of expertise hand in real time. The proposed extensive CAMSHIFT algorithm uses both color and motion captured by the camera to detect the hand in motion (Przybyło, 2010). Then videos are segmented into video frame and region of interest is selected and matched with the video frame came from the local site. Frames with ROI are blended using the Hybrid blending algorithm to combine two videos. Extensive CAMSHIFT algorithm is used here to remove the limitation of current solution, which does not detect moving image that is used for merging. The motion with the hand is calculated by simulating three different hand samples in Matlab using proposed extensive CAMSHIFT algorithm, which helps to improve the accuracy by 0.16 mm. Proposed Equation 3 is to detect the hand speed using extensive CAMSHIFT:

$$P = \begin{cases} 0 \\ \alpha.P_m + (1-\alpha).P_{color} \end{cases} BW = 0, BW = 1 \quad (3)$$

Where:

P = Final probability map
P_m = Motion probability Image
BW = Binary motion map
$Pcolor$ = Color probability image
α = Alpha values in RGB image

Remote Environment (Mixed Reality Side) using Enhanced Volumetric-Based Image Synthesization

Once merging is done between segmented hand from surgeon remote site and augmented video from a surgery local site the occlusion needs to be removed in MR. The current solution by Habert *et al*. (2017) could synthesize the view, the background occluded by surgical hand to visualizing it during transparency. In the beginning, two RGBD cameras are placed on the side of the C-arm, giving additional information from another viewpoint. Image captured from 2 RGBD camera angle is synthesized alongside alpha value which is generated during capturing process. After that, each camera synthesization and alpha value is multiplied. The raytracing is performed on every pixel to recover the complete occlusion of segmented hand. Additionally, a noise removal step is added using morphological opening on mask image of the hand. After that, the second raytracing was performed based on the segmentation using binary search resulting in a colour image of background synthesized combined with the colour image of first raytracing which creates an occlusion free image of expertise hand. Finally, 3D augmented video frame is merged with occlusion recovered frame by hybrid blending and create mixed reality.

The hybrid blending algorithm generates composite video without removing the occlusion using Equation 1. The image synthesization using Equation 2 is merged with hybrid blending which tries to eliminate the final video with occluded recovered video using Equation 4.

Proposed Enhanced Hybrid Blending Synthesization Algorithm

Calculate the volumetric based image Synthesization is using given enhanced Equation 4. The Mixed Reality can be created by the Equation 4:

$$I_{final}(p) = \begin{cases} (\alpha I_c(p) + \beta I_b(p)) \\ +(\alpha * \phi + (1-\alpha) * B) if\ p \epsilon foreground \\ (1-\delta) I_b(p) else \end{cases} \quad (4)$$

Where:

I_b = First RGBD camera color image
I_c = Second RGBD camera color image extract which could be overlaid with transparency
I_{final} = Multilayer image created by blending all the layers where α, β, γ, δ = [0,1], $\alpha + \beta + \gamma = 1$ where α = Alpha, β = Beta and γ = Gamma

where, ϕ calculates the constraint the blending degree calculated using Equation 5 (Habert *et al*., 2017):

$$\phi = \tau * \sum_{j=0}^{k}\sum_{i=0}^{n-1}\lambda_{ji}(x)(\varphi - \psi)(P_{ji}) + \psi \quad (5)$$

Where:
ψ = Sequence of foreground video frame cube
λ_{ji} = MVC according to specified boundary condition
K = constrain of the blending degree
τ = Function whose value is between 0-1
P = boundary condition

The above Equation 5 helps to calculate the constraint degree after calculating the mean value coordinate of given video to compose 3D video. With this approach, the local surgeon has full control for observing different layers having a choice to emphasize layer, they are interested. It has proposed a way to recover the occluded anatomical structure and view recovered image frame from live video frame capture from two-camera using image synthesized technique. The current technique of composition present in the state-of-the-art solution is unable to handle the foreground object, i.e., surgeon's hand that moves quite fast making hard to track and cage in a given mask. In addition, it did not consider about the occlusion recovery of background image occurred by foreground image during merging. However, the proposed technique is given by Habert *et al*. (2017) helps to recover the occluded area using image synthesization. Whereas, Equation 6 helped to segment the quick motion hand more accurately by using both of the color and motion information. Enhanced volumetric-based image synthesization has been integrated with hybrid video composition to refine, eliminate smudging and sawtooth artefacts and also recover a fair amount of an obstructed anatomical region from the composite video, which will allow the remote surgeon to guide virtually from long distance with better perception. Finally, when all these equations are combined and executed in Mixed Reality system, it works impeccably.

Why BSEC?

The CAMSHIFT algorithm, which uses the skin-color for detecting the human face and hand, is able to operate in a captured video in real time with a color probability distribution. Image synthesization is multiple layer visualizations with background recovery with a user-adjustable scheme for transparency between different layers. The state-of-the-art, Hybrid Blending we picked (In Equation 1) has only provided us with merging of type of videos from two different sources in one local site. It has a limitation during detecting hand or segmenting the object from source if it has higher pace. In addition, when this algorithm uses in surgical telepresence, occlusion recovery is significant, which needs to be fixed for better cooperation. Occlusion occurs when the surgeon's hand is placed over the patient anatomy during surgery; this could be fixed using both above-mentioned algorithms, i.e., extensive CAMSHIFT and Image synthesization. The proposed system takes the same steps hybrid blending, but it merges with the synthesization technique to reduce the processing time for running both extensive Camshift and Image synthesization algorithm separately. Both hybrid alpha blending and volumetric based image synthesization are merged to run in parallel manner helping to recover and blend into one single view with less processing time. The proposed solution does not take any *x*-ray image as input during the surgery in the virtual site so; it has been removed from the image synthesization equation. Similarly, the fast pace of hand is recovered using the extensive CAMSHIFT algorithm which runs in the beginning phase with the system which helps to better segmentation of hand. It captures the image using the RGBD camera using color and motion captured by the camera, where RGBD is a combination of RGB camera and its corresponding depth image. The RGB value is used for better segmentation of hand whereas; depth value is an image channel where each pixel denotes the distance between the corresponding image and image plane which keeps better accuracy.

Table 2 indicate the proposed algorithm. The proposed algorithm works well with the hybrid video composition technique. Both of the algorithms have been included within the system and integrated into the proposed equation to eliminate unnecessary variables based upon the equation results. The composite video of the proposed system is 1.29~1.45 mm which is an error in accuracy and the number of processed frame is 50 on average. The flowchart of the proposed extensive CAMSHIFT and volumetric based image synthesization algorithm is shown in Fig. 5 and 6 respectively.

Table 2: Proposed BSEC

Algorithm: Proposed BSEC system to remove occlusion and detect hand in motion Input: 3D augmented video frame and hand video Output: Multilayer mixed reality video
BEGIN Step-1: Two RGBD camera captures the hand image from the different viewpoint. Step-2: Convert RGB image frame to HSV conversion and calculate the motion of hand using equation 3. Step-3: Calculates the constraint the blending degree calculated using equation 5 Step-3: Calculate volumetric based image Synthesization is using given enhanced equation 4. Calculate the multilayer image created by blending all the layers END

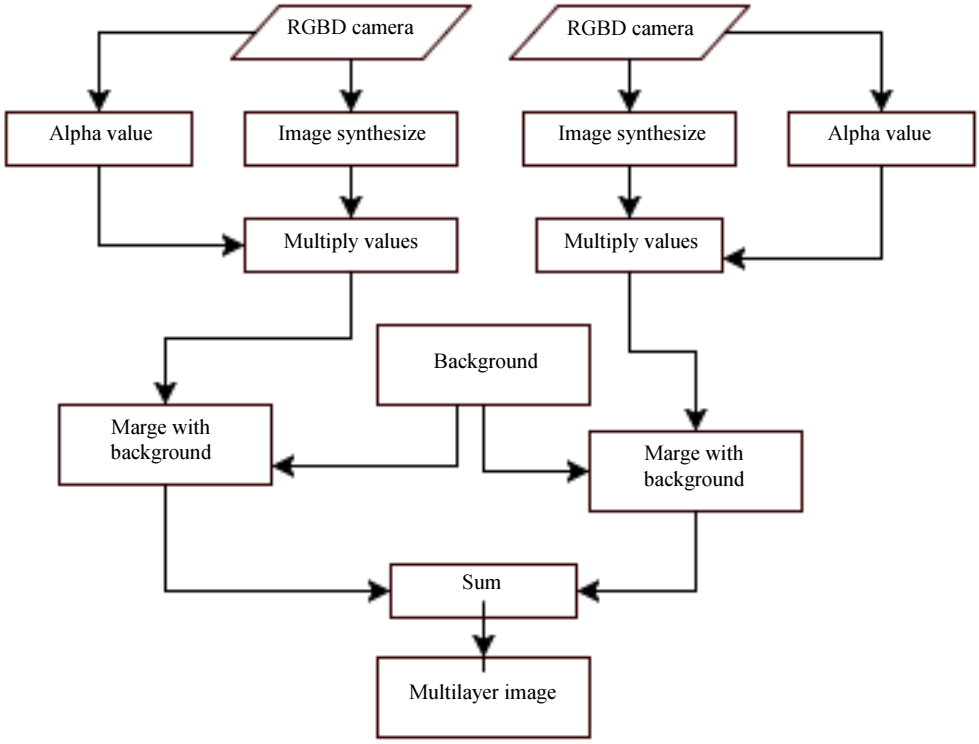

Fig. 5: Flowchart of Volumetric based image sythesization

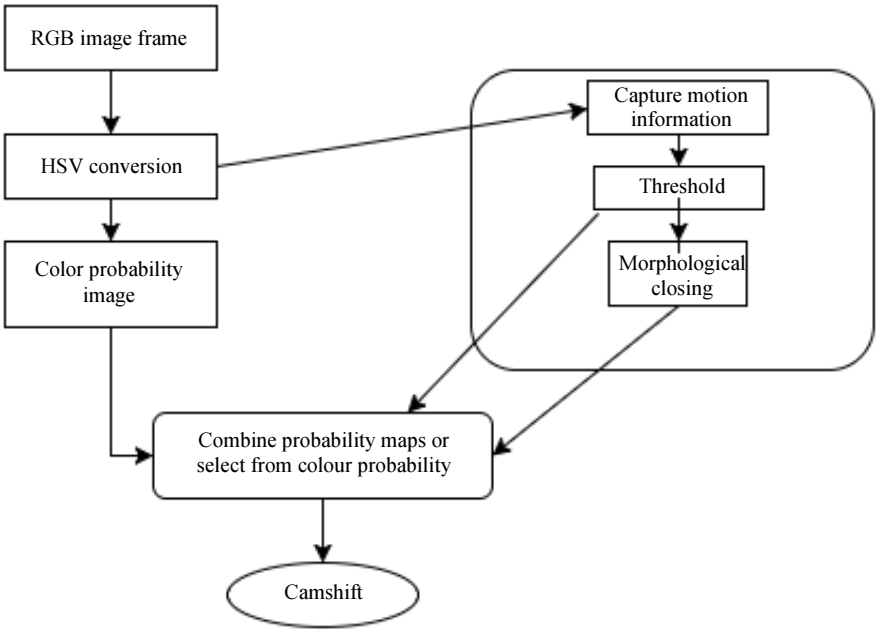

Fig. 6: Flowchart of extensive CAMSHIFT method

Experimental Results and Discussion

In this section, we present experimental results came from our current and proposed solution. We have used Matlab version R2017b with 10 samples of augmented videos of breast surgery; some hand image and hand videos as an input to our work. Breast surgery samples and hand image frame are collected

from online resources; our hand video is taken from Matlab using a laptop camera in the black background. Performance of our system is measure based on accuracy and processing time. All the experiments are done in a simulator called Matlab, which might result different from the real experiment and the result generated in 10 different breast and hand samples is shown in Table 3.

The experimental results from 10 breast samples reduce the average accuracy error of the video frame image overlay from 1.44 to 1.28 mm with average processing time enhanced to 76 sec compared to the state-of-the-art method.

The proposed system consists of three stages: Local site, Virtual site and Remote site. In the local site, 3D augmented video is produced and sent it to the remote through the Internet connection. In the remote site, the expert surgeon can see 3D augmented view received from local and it is used for guiding local surgeon by sending mixed reality. The mixed reality is created in remote site by merging remote surgeon hand with 3D augmented video frame. The produced mixed reality produced is sent back to the local surgeon in real time. The 3D augmented video frame received by the remote site surgeon, captured hand of a remote site surgeons is shown in Fig. 7a and 7b respectively. Two RGBD cameras are used to capture the hand; the motion with the hand is hard to detect so, Extensive Camshift Algorithm is used for the motion detection in real time. A frame is extracted from the video; ROI is used for mapping coordinate and hybrid blending is done to merge and align two frames. Then, proposed volumetric based image synthesization is used to remove the occlusion and create multilayer visualization and the overlayed image is shown in Fig. 5c.

Table 3: Accuracy and processing time results for breast samples

		Image Overlay										
				Current solution			Proposed solution					
Sample number	Sample details	Augmented Reality video	Original Hand	Processed Sample	Accuracy	Processing Time	Processed Sample	Accuracy	Processing Time	Processed Sample	Accuracy	Processing Time
1	Breasts (31, 5'6",140)				1.5 mm	77 sec		0.67 mm	0.49 sec		1.29 mm	78 sec
2	Breasts (34, 5'2",190)				1.46 mm	78 sec		0.67 mm	0.49 sec		1.33 mm	79 sec
	Breasts (40, 5'7",155)				1.45mm	64 sec		0.67 mm	0.49 sec		1.34 mm	65 sec
4	Breasts (25, 5'2",120)				1.54 mm	64 Sec		1.34 mm	0.67 sec		1.19 mm	65 sec
5	Breasts (46, 5'8",142)				1.78 mm	79 sec		1.34 mm	0.67 sec		1.45 mm	80 sec
6	Breasts (25, 5'3",160)				1.44 mm	76 sec		1.34 mm	0.67 sec		1.19 mm	77 sec
7	Breasts (45, 5'3", 116)				1.32 mm	89 sec		1.34 mm	0.67 sec		1.17 mm	90 sec
8	Breasts (48, 5'1",115)				1.53 mm	75 sec		1.23 mm	0.49 sec		1.43 mm	76 sec
9	Breasts (53, 5'6",120)				1.27 mm	74 sec		1.23 mm	0.49 sec		1.08 mm	75 sec
10	Breasts(21, 5'4",100)				1.31 mm	73 sec		1.23 mm	0.49 sec		1.30 mm	74

(a) (b) (c)

Fig. 7: (a) 3D Augmented View (b) Captured hand (c) Mixed Reality

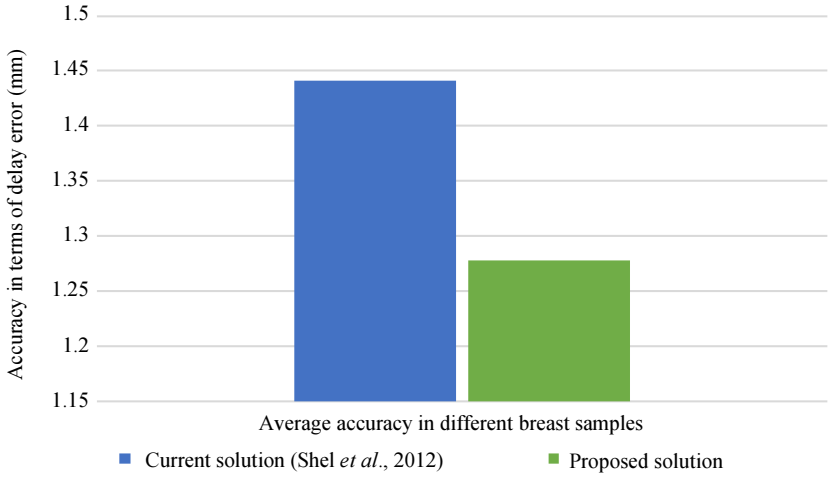

Fig. 8: Results of accuracy in image overlay

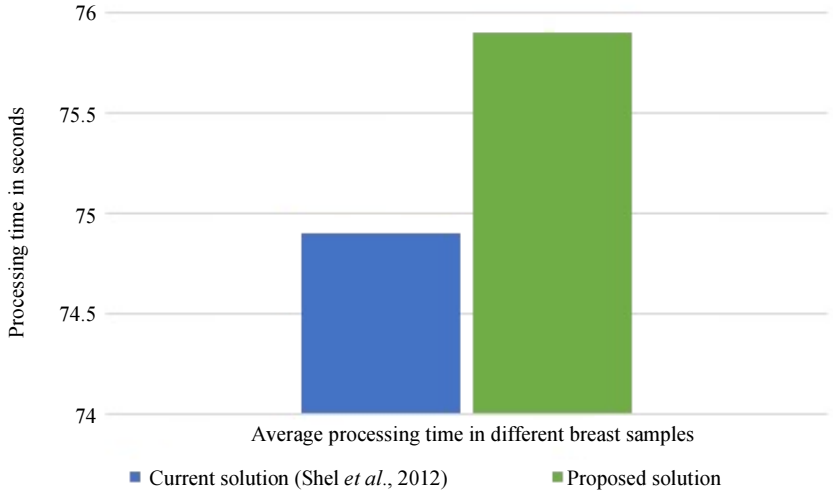

Fig. 9: Results of processing time for image overlay

The current and proposed algorithms were simulated in Matlab with the videos collected from the online sources and hand captured of own. The visualized hidden anatomy, such as the blood vessels and the tumors, will be an input as an augmented video to our proposed system. The augmented visualized video should be created in local site (surgery site) and sent to remote site (surgeon site). Sample dataset collected is from a different age group from 24 to 50 and weight from 99 lb to 190 lb. An augmented 3D video is used to experiment on both current and proposed equation algorithm. Accuracy is checked according to the overlay error in mixed reality using enhanced as well as processing time is calculated according to the processing time consumed by the system. Accuracy and processing time taken by the current and proposed solutions are displayed in the bar graph. Figure 8 shows the average accuracy error of current and proposed system in different ten samples. Similarly, Fig. 9 reflects the average processing time taken by proposed solution and current solution to generate the complete output. In Fig. 8 and 9, blue is indicated as current and green as a proposed solution. Figure 9 shows processing time comparison against the state-of-the-art method. The figure reveals that the proposed time only requires extra 1% computational time compared to the state-of-the-art method, however, it provides a significant improvement in terms of accuracy (Fig. 8).

The above results show the accuracy and processing time taken by the current solution and proposed novel solution for the breast telepresence. The difference between current and the proposed in accuracy and processing time is in Table 2. In above Table 2, the state-of-art has an accuracy error of 1.5 mm and it took 77

secs for processing 50 frames. The proposed system has used Extended Camshift and Volumetric Based Image Synthesization for detecting motion with the hand and removing occlusion in real time and has an accuracy error of 1.29 mm and 78 sec. The accuracy and processing time is calculated by simulating the temporal protocol in MATLAB systems. The 'Ruler' tool helps to measure the distance between two points. This measurement has done in pixel and converted to mm; where (1 pixel = 0.264583). Accuracy error distance has calculated between two pints (A and B); where A is the image point where the surgeon hand should indicate and B is where the remote hand has overlayed. We quantify the degree of improvement in processing time by running the state of art and proposed algorithms and the duration of running each algorithm. Furthermore, image overlay in real time with hand motion detection and occlusion removed does not cause a significant difference in processing time because of no parallel processing. The deformation of the soft tissues has not significant effect on the image overlay accuracy in breast surgeries. This is because of the breast is consist of fatty tissue and the possible deformation of surrounding dense soft tissues is minimal.

Conclusion and Future Work

To conclude, the mixed reality within the field of surgery is a great achievement, which is created by combining different techniques with further improvement. The implementation of two RGBD cameras and CAMSHIFT algorithm has improved to obtain a better segmentation of a hand for interaction even in the motion with the great accuracy. The hybrid alpha blending algorithm has successfully merged two videos from a different source and background seamlessly without smudging and sawtooth. It can also be implemented successfully during capturing the video in real-time. Volumetric based image synthesization technique has allowed us to recover and view occluded area-using transparency in a different level, which could even be adjustable by surgeons according to their needs. Motion threshold has suppressed the influence noise to the image during extraction of hand.

The implementation of such a system has been a field of a research on few surgeries; however, there are many areas where deep research still needs to be done since mixed reality is an emergent topic. Although few have proposed an idea about mixed reality, they have failed to implement practically. This research has explored different technique and integrated them to create a completely new system, which is both doable and practical. The accuracy error of the proposed system is 1.29 mm, which has shown significant improvement on the overall system. This justifies the addition of the proposed BSEC to overcome the different addressed limitations in the existing best solution.

The proposed algorithm was done in (MatLab simulation) to create a multilayer visualization with less accuracy error. Some proprietary software has been proposed to address this issue, but none of them has implemented and provided with an accurate solution. If the design and implementation work toward such anatomically challenging area, its potential application will be limitless and could be globally renowned.

Several improvements in future research can be done including transparency of local surgeon hand as user preferences, segmentation and occlusion. All these improvements can provide satisfying result in creating a mixed reality for telesurgery.

Acknowledgement

We are grateful to Mrs. Angelika Maag for proof reading and making corrections to this article. Without her support, it would have not been possible to submit this in the current form.

Author's Contribution

Krishna Shakya and Suman Khanal: They have completed this project as part of his MIT degree program.

Abeer Alsadoon: Worked on the setup of the experiments and gave important suggestions on design of experiments. In addition to, as Abeer Alsadoon is the main supervisor on this work, she provided the details guidelines and feedback in each step of this work.

P.W.C. Prasad: Made important revisions to most sections of the paper.

Anand Deva and Jeremy Hsu: Providing the idea of the topic and helped in modifying some parts.

Manoranjan Paul and A. Elchouemi: Done the technical review of the paper and provided technical guidance. Give the final review and approval for the manuscript to be submitted.

Acknowledgement

Ethics This article is original and contains unpublished material. The corresponding author confirms that all of the other authors have read and approved.

References

Andersen, D., V. Popescu, M.E. Cabrera, A. Shanghavi and G. Gómez *et al*., 2016. Avoiding focus shifts in surgical telementoring using an augmented reality transparent display. MMVR, 22: 9-14.

Baker, D.K., C.T. Fryberger and B.A. Ponce, 2015. The emergence of augmented reality in orthopaedic surgery and education. Emergence.

Bradski, G.R., 1998. Computer vision face tracking for use in a perceptual user interface. Intel Technol. J., Q2: 214-219.

Bruellmann, D.D., H. Tjaden, U. Schwanecke and P. Barth, 2013. An optimized video system for augmented reality in endodontics: A feasibility study. Clin. Oral Invest., 17: 441-448. DOI: 10.1007/s00784-012-0718-0

Chang, T.C., C.H. Hsieh, C.H. Huang, J.W. Yang and S.T. Lee *et al*., 2015. Interactive medical augmented reality system for remote surgical assistance. Applied Math. Inform. Sci., 9: 97-104.

Correia, M.V. and A.C. Campilho, 2002. Real-time implementation of an optical flow algorithm. Proceedings of the 16th International Conference on Pattern Recognition, (CPR'02), pp: 247-250.

DeSantis, C.E., J. Goding Sauer, A. Ma, L.A. Newman and A. Jemal, 2017. Breast cancer statistics, 2017, racial disparity in mortality by state. Cancer J. Clinicians, 67: 439-448. DOI: 10.3322/caac.21412

Diaz, R., J. Yoon, R. Chen, A. Quinones-Hinojosa and R. Wharen *et al*., 2017. Real-time video-streaming to surgical loupe mounted head-up display for navigated meningioma resection. Turk Neurosurg. DOI: 10.5137/1019-5149.JTN.20388-17.1

Guo, Y., O. Henao, T. Jackson, F. Quereshy and A. Okrainec, 2014. Commercial videoconferencing for use in telementoring laparoscopic surgery. MMVR, 18: 147-149.

Habert, S., M. Meng, P. Fallavollita and N. Navab, 2017. Multi-layer visualization for medical mixed reality.

Kersten-Oertel, M., P. Jannin and D.L. Collins, 2013. The state of the art of visualization in mixed reality image guided surgery. Computerized Med. Imag. Graph., 37: 98-112. DOI: 10.1016/j.compmedimag.2013.01.009

Liu, Y., Y. Wang, H. Sun, D. Cheng and D. Weng, 2016. Mixed and augmented reality innovations in Beijing. Institute of Technology.

Milgram, P. and F. Kishino, 1994. A taxonomy of mixed reality visual displays. IEICE Trans. Inform. Syst., 77: 1321-1329.

Pin, N., 2015. Vision-based three-dimensional hand interaction in markerless augmented reality environment.

Przybyło, J.M., 2010. Hand tracking algorithm for augmented reality systems. Automatyka/Akademia Górniczo-Hutnicza im. Stanisława Staszica w Krakowie, 14: 581-589.

Shen, Y., X. Lin, Y. Gao, B. Sheng and Q. Liu, 2012. Video composition by optimized 3D mean-value coordinates. Comput. Anim. Virtual Worlds, 23: 179-190. DOI: 10.1002/cav.1465

Shenai, M.B., M. Dillavou, C. Shum, D. Ross and R.S. Tubbs *et al*., 2011. Virtual Interactive Presence and Augmented Reality (VIPAR) for remote surgical assistance. Operative Neurosurgery, 68: ons200-ons207. DOI: 10.1227/NEU.0b013e3182077efd

Shenai, M.B., R.S. Tubbs, B.L. Guthrie and A.A. Cohen-Gadol, 2014. Virtual interactive presence for real-time, long-distance surgical collaboration during complex microsurgical procedures. J. Neurosurgery, 121: 277-284. DOI: 10.3171/2014.4.JNS131805

Walton, D.R., D. Thomas, A. Steed and A. Sugimoto, 2017. Synthesis of environment maps for mixed reality. Proceedings of the IEEE International Symposium on Mixed and Augmented Reality, Oct. 9-13, IEEE Xplore Press, Nantes, France, pp: 72-81. DOI: 10.1109/ISMAR.2017.24

Wang, R., Z. Geng, Z. Zhang, R. Pei and X. Meng, 2017. Autostereoscopic augmented reality visualization for depth perception in endoscopic surgery. Displays, 48: 50-60. DOI: 10.1016/j.displa.2017.03.003

Appendix

Appendix 1: Abbreviations for the terms used in the paper

VR	Virtual Reality
AR	Augmented Reality
MR	Mixed Reality
CT	Computed Tomography
3D	Three Dimensional
2D	Two Dimensional
HD	High-Definition
LUT	Look-up-table
GG	Google Glass
LD-VIP	Long Distance-Virtual Interactive Presence
HSV	Hue-Saturation Value
HIPAA	Health Insurance Portability and Accountability
MRI	Magnetic Resonance Imaging
HMD	Head Mounted Display
MVC	Mean Value Co-ordinates
CAMSHIFT	Continuously Adaptive Mean Shift