Multimedia Tools and Applications (2024) 83:49845–49874
https://doi.org/10.1007/s11042-023-17331-4

Mixed reality in surgical telepresence: a novel extended mean value cloning with automatic trimap generation and accurate alpha matting for visualization

Roshan Dallakoti[1] · Abeer Alsadoon[1,2,3] · P. W. C. Prasad[1,2] · Sarmad Al Aloussi[4] · Tarik A. Rashid[5] · Omar Hisham Alsadoon[6] · Ahmad Alrubaie[7] · Sami Haddad[8,9]

Received: 25 January 2021 / Revised: 9 December 2021 / Accepted: 29 September 2023 /
Published online: 4 November 2023

Abstract
The aim of this research is to propose an extended mean value cloning algorithm with automatic trimap generation and accurate alpha matting. This implementation improves the visualization accuracy of the merged video by reducing the discolored and smudging artefacts of the remote surgeon's boundary. It also makes the merge robust for the illumination changes by taking less processing time in real time surgery. The proposed system uses automatic trimap generation from the source video for accurate foreground extraction. Extended mean value cloning with gradient mixing is then applied for the cloning with optimized alpha matting for accurate and realistic video composition. The proposed system improved the visualization accuracy by providing almost 99.7% visibility of the pixels compared to the state-of-the-art solution, which provides 99.1% visibility of pixels. The overlay error was reduced from 0.93 mm to 0.63 mm. The processing time was also reduced. The proposed solution processed 8 frames per second, which is less time than the state-of-the-art solution, which processed 5 frames per second. The extended mean value cloning smooths the differences that presented in the target and source frames for seamless and realistic blending of pixels. The automatic trimap generation reduced the risk of false foreground selection and the generated optimal trimaps improved the alpha matte quality, which is optimized to reduce the smudging artefacts completely and to produce accurate visualization of the final merged image.

Keywords Video composition · Mean value cloning · Contour flow · Automatic trimap generation · Gradient mixing · Image matting

1 Introduction

Different complex surgeries of hard and soft tissues might require additional expertise. These complex surgeries can take advantage of real time assistance by using remote collaboration technologies, which can be accessed regardless of location. This real time video technology assisted in different complex surgeries, which has not been adopted

Extended author information available on the last page of the article

as a practical implementation due to technical barriers and lack of flexibility. These deficiencies can be satisfied by Mixed Reality in which Augmented Reality and Virtual reality are combined to achieve a virtual and real world mixed reality that combine to support remote collaboration [1]. To achieve remote collaboration in video assisted surgery, various video composition techniques have been proposed. However, the basic challenges during video composition are automatic extraction of the source video patch, cloning to the background target region of interest naturally, and removing the need for supervision in the real time video composition [2]. Cutting the source patch and passing into the target using interactive video editing tools or blending images together frame by frame are common methods that are used for video composition. Image matting is another technique in the video composition domain. Image matting is a method to accurately determine the foreground and the background in the image for the seamless merging of videos without smudging artefacts [3]. For image matting, trimap or scribble methods are required. Comparing scribble methods with the trimap-based method, the precision on the former is low compared to the latter. However, scribble methods reduce the processing time more significantly than the trimap [4]. Mean value cloning is a popular technique that is used for merging videos. Yehu, et al. [5] modified the mean value coordinates for merging two videos. Trimaps are the input source for the source video that is often required for image matting and merging with the background target. Traditionally, the trimaps are drawn manually which is time consuming and exhausting. For accurate visualization, manual selection of points as input for trimap generation is error prone [3].

The means to extract the source patch is by automatically generating the trimap and blending boundary pixels from the trimap to calculate the extended mean value coordinates. The extended mean value cloning was extended with gradient mixing to make it robust to the illumination challenges. Natural image matting was used as a preprocessing stage for the accurate foreground extraction, which was the remote surgeon's hand. In image matting technology, an image has foreground and background layers; image matting determines the pixels of the foreground layer [6]. The purpose of image matting is to find true source and target pixels and based on the finding, the undetermined pixels accurately. The alpha matte is generated for reducing the smudging artefacts around the surgeon's hand boundary in the final merged image. Many proposals suggest automatic trimap generation by using depth information and dilation for the foreground segmentation and unknown region detection, which is considered impractical. Henry and Lee [3] implemented automatic trimap generation without using the depth information, but by using image saliency, soft segmentation and fuzzy clustering. Li, et al. [4] considered the super pixels of different cluster images, This improved the fuzzy clustering with the density peak algorithm for better clustering results. Both Henry and Lee [3] and Li, et al. [4] proposed the approach to generate trimaps for accurate alpha matting without requiring d user intervention. Wang, et al. [2] used gradient mixing to optimize temporal spatial consistency optimizing the temporal coherency in all the frames. The blending boundary was optimized by making it color consistent between the target and source, which addresses the challenge of varying lighting conditions and motion difference. Cai, et al. [6] proposed the evaluation cost function to determine the best foreground and background sample pairs to determine where the unknown pixel belongs based not only on color but also on the spatial distance for the accurate natural image matting, which reduces the smudging artefacts from the final image. The main objective of this paper is to combine local surgery site generated augmented video with the remote surgeon's site gesture video seamlessly and to achieve a realistic mixed reality view. The real time final video composition must be able

to accurately show the surgeon's hand movements around the region of interest without blocking the background.

2 Literature review

Various relevant papers were studied to understand the concepts and techniques used for efficient video composition. The following analysis and discussion are from the comprehensive study.

2.1 Extraction of the source patch

Henry and Lee [3] enhanced the trimap generation by automating the generatuib process for image matting. Depth information was required for most automatic trimap generation methods and to segment the foreground object previously [7]. However, the proposed system does not require color depth information. Fuzzy c-means (FCM) was used to find the unknown region automatically. Only one image saliency algorithm was used for reducing the computational time and to reduce the risk of incomplete salient object segmentation. A lazy snapping agorithm is used to accurately segment the foreground. Finally, the unknown region of the trimap was generated by FCM. It generates the trimap automatically without the need for depth information. This method was applicable because it worked automatically on the RGB color images. Alpha mattes generated by the optimal tripmaps contain fewer artifacts and are processed faster. Automatic trimap generation improved the accuracy of the generation of the alpha mattes by reducing the artefacts and the alpha matte computational time. The working method of the proposed system relied on the accurate salient maps. Incorrect salient maps produced the incorrect trimap. If the image contains holes and complex objects in the background, then it is abstracted because the image is over segmented before calculating salient maps.

Pawin, et al. [8] enhanced the traditional background substraction using image intensities by using unit gradient vector based background substraction, which addresses the varying illumination intensities. The extended Sobel operators of size 5*5 were constructed because they were more effective for suppressing high frequency components in an image due to their powerful low pass filtering. The extended Sobel operator segments are more efficient than the traditional Sobel operator. Skin color segmentation was applied to segment the hand from full-color images and from the skin-colored background images [9]. Post processing reduced the limitation of background subtraction and skin color segmentation to provide accurate hand segmentation in real time. The binary images from the previous methods are integrated and an AND operation was performed to truly segment the foreground and the isolated pixels are removed. Both UGVs and Hue were invariant to varying image intensities due to dynamic lighting conditions. The camera with auto gain control function (AGC) supports the varying lighting conditions. The proposed method can accurately segment the hand foreground in 5–21 frames per second in varying lighting conditions. The proposed method is employed in the Raspberry Pi-embedded machine, which proves that the proposed algorithm works on a low-cost computer. Comparing the computational time, the proposed method requires more time (5–21 frames per second) and the final segmented hand has some noises when the lighting condition is highly irregular.

Li, et al. [4] proposed the simple and efficient high quality trimap generation technique for image matting by reducing the influence of human intervention. In the proposed

solution, the weight of each superpixel is determined to formulate the weight selection as an energy minimization problem. Similar weighted superpixels are incoporated prior together prior to foreground/background pixels and smoothness is applied to to the adjacent pixels. Different segmentation methods are used in an input image to generate a soft segmentation to make the system robust. Previously, only one segment was needed for obtaining the initial boundary of the foreground object [7]. Then, the fuzzy clustering method is used to divide the soft segmentation images to the foreground, background and unknown region. The rough trimap is created on fuzzy clustering the soft segmented maps from the different soft segmentation methods A final trimap is generated by weighting fusion of the various rough trimaps at the superpixel. Little intervention is required by the user to label the different parts of the image. The proposed solution addresses the pixels of different cluster images at the superpixel level and determines the weight of each superpixel level [10]. It provides more accurate fusion of segmented images to distinctly visualize the foreground, background and the unknown region. The proposed solution combines the density peak clustering (DPC) algorithm to improve the fuzzy c-means clustering (FCM) algorithm for better clustering of the pixels of the three regions. The trimap generated by the proposed method has smaller Root Mean Square Error (RMSE) and Mean Absolute Deviation (MAD). The accuracy of the trimap is also significantly better than that of the manually generated trimap. When the foreground color and the background color are close, the unknown region of the estimated trimap in the proposed solution is also smaller than the unknown region of the manually generated trimap. The more accurate trimap estimated by this method provides better matting results. Some user intervention is required in the selection of color in the input image.

2.2 Merging techniques for visualization

Venkata et al. [1] enhanced the mean value cloning algorithm with multi-layer visualization, which selects and synthesizes the images by using a volumetric image synthetization algorithm (VIS) to reconstruct the occluded area pixels for the visualization of region of interest. Traditional methods include the Poisson video composition methods, which are required to solve the 3D Poisson equation. To reduce the processing cost in real time, the mean value coordinates cloning technique is introduced for effective video composition as shown in the block diagram of [1] in figure one. Yehu, et al. [5]. Similarly, trimap propagation is used for object movement synchronization during the surgery. Alpha matting is used to remove the dirty noises in the blending boundary and the constraint coefficient is applied to the video to remove the discoloration. The video from the local surgery site is merged with the video from the remote expert surgeon's site to give the mixed reality view by accurately visualizing the surgeon's hand boundary with the surgical region of interest. This proposed system improves the overlay accuracy and the visualization in the merged image that can be used as an effective guide to the local surgeon throughout the surgery. This method can be used in developing countries where expertise is lacking. The proposed solution provided the final output video by merging two videos from the local and remote site without noises and discoloration. The solution reduced the overlay error from 1.3 mm to 0.9 mm and increased the visibility pixels to 99.1% from 98.4%. The processing time is similar to the current solution. The proposed system supports remote collaboration for real time surgeries by using mixed reality. This system converts the video into the frames before sending them through the Internet, which might be time consuming. Sending the video directly without extracting the frames significantly reduces the processing time and

reduces the noise that can be caused by the surgical blood flow at the time of remote collaboration. Also, the trimap is based on the user having to manually select the points in the screen.

Wang, et al. [2] introduced a new illumination guided video composition in a gradient domain, which improves the composition of the video for the varying illumination intensity. The user provides the inner and outer blending boundaries in the source video to generate the trimap. In this defined blending boundary, the gradients of the source patch and the target video are mixed for each frame [11]. An effective mean value coordinated interpolation is implemented for the smoothness of the difference between the source and target. The value of the interpolated coefficient controls the flow of color distribution. This strategy of optimized interpolation is effective for addressing the challenge of varying illumination conditions and for the indistinct object boundaries like smoke and dust making the image cloning consistent. Spatial–Temporal Consistent Boundary Computing also addresses the challenge of motion of objects in the video. Gradient mixing maintains the temporal spatial consistency and optimizes temporal coherency over the frames for seamless video composition. The proposed system had several challenging video sequences as input that resulted in high quality video composition. The blending region appearance harmonizes with the target brightness condition. It an produce the seamless and globally consistent composition results under varying illumination intensity that can be used in remote collaboration of the surgery using mixed reality. The solution might fail when there is inconsistent source and region of interest texture. When it is difficult to estimate the motion of the source object and the target background, more supervision is required to maintain the blending boundary for the accurate composition.

Hu, et al. [12] enhanced the background subtraction on pixel based for the sequence of video frames based on color-locality sensitive histograms (CLSH) for illuminant varying scenarios. Local histograms of each channel frame are computed and are merged to get the colored CLSH, which is then used to perform segmentation. Then, each pixel is classified as foreground or background comparing the pixel and the set of CLSH samples created by using Earth Mover's Distance (EMD). Blocked matting as real time matting is applied to smoothen the boundary of the video foreground and making it accurate [13]. Alpha matte is computed and applied only to the pixels that change state. This improves the speed of the matting in the real time video foreground segmentation and enhances performance. Similarly, local antialiasing is applied to the video boundary for smoothing and makes it realistic. For high-resolution videos, for example, for the frame size 1920×1080, processing time is longer for background segmentation; this longer time might not be as practical as real time video processing.

Wang, et al. [14] enhanced the existing POINTER system to increase the collaborative efficiency and to support more natural and intuitive interaction. The Basler camera streams live video at a resolution of 1280×960 pixels, which is rendered in 3D VR space in the plane so that the remote user sees a virtual copy of the local user's real task space in the Head Mount Display (HMD). The spatial position of the plane is fixed on the remote site so that the remote expert can move freely around the plane during remote collaboration. The hand gestures of the remote user is added on the plane. The Leap motion software development kit manipulates the hand gestures for stable and robust recognition, which is then projected into the real workspace in the local site for the local user to see. The Basler camera shows the video in the HMD in the remote site. The algorithm written in C/C++is imported to Unity3D in DLL format. This system supports more natural and intuitive interaction and improves the co-presence awareness and collaborative efficiency [15]. This calibration adjusts the position, rotation and scale of the live video rendered

in the VR space. The sharing of gestures significantly reduced the performance time and improved the performance, co-presence and remote interaction. However, the camera and projector in the local site is fixed, which will be misguiding for the complex operation and occlusion. Because most instructions are based on the remote expert, the local user might find it tiring to perform as long as the remote expert.

Basnet, et al. [16] has proposed the enhanced image reconstruction-based occlusion technique to reconstruct the occluded areas correctly and then remove them so that the tracker learning algorithm (TLD) does not slow the process. Tracking-learning-detection is used to track the object in the video frame. It uses a bounding box. The use of TLD with tracking and detection in real time video frames and the use of a bounding box reduces the search area and the processing time [17]. The TLD takes additional time to search the feature point after the occlusion and reinitializes increasing the processing time. The elimination of the conversion of 3D images to 2D image frames decreases the processing time by using an optical camera. The noise removed by the modified kernel nonlocal means (MKLMN) increases the image overlay accuracy. In the preoperative stages, the CT scan of the region of interest provides high quality images with details and information. In the intra-operative stage, the CT scan is segmented to create aspect graphs with various models used to match real time images from different perspectives. The optical camera takes live video of the real time surgery. The MKLMN filter removes the noise from the video frames. The hierarchy of frames is created, and the highest quality image is taken by the TLD for tracking and detecting the area of interest. In the case of occlusions, the TLD fails to find the feature. The TLD takes extra time to search the feature point after the occlusion and reinitializes increasing the processing time. Image reconstruction based on occlusion removal comes into play to reconstruct the occlusion correctly and remove the occlusion before the TLD fails. The image with the highest quality is sent for the pose refinement for removing the geometric error. The noise removed by MKLMN increases the image overlay accuracy. The image overlay was improved to 0.23 ~ 0.35 mm and increased the processing speed up to 8 ~ 12 frames per second.

Cai, et al. [6] has proposed the earthworm optimization algorithm (EWA) as an efficient metaheuristic algorithm to solve the sample optimization problem to find the best foreground–background pair. An optimized cost function is proposed to evaluate the true optimal true foreground pairs on a classic evaluation function. The classic evaluation function consist of both color and spatial measures for finding the optimal foreground–background sample pair for every undetermined pixel [13]. The authors have enhanced the EWA to find the true foreground–background colors for undetermined pixels. The algorithm is used to search the optimal sample pair for every undetermined pixel at the same time to improve the efficiency of the search. Furthermore, to improve the search ability and to escape from the local optimum, Cauchy mutation is used. This setup can efficiently find the optimal foreground–background sample pairs and improve the accuracy of matting for sampling-based image matting methods.

Hettig, et al. [18]) implemented Augmented Visualization Box (AVB) using a computer game engine for a real time global lighting model by customizing intraoperative lighting conditions. MeVisLab image processing software was used for 3D anatomical models [19]. The VR scene was created from the surface models. The surface textures were accurately done by using Voronoi noise and RGB color bending. For the AR visualization, the counter and transparency rendering were used. Post processing methods are used to calculate the edge of the contours. Virtual endoscopy with 30-degree optics, which is usually used in laparoscopy, was used for VR scene rendering. Visually occluded regions and internal anatomy was included by the AR scene. The final scene was created by superimposing the

AR to VR scene and rendered through the display. Through Augmented Visualization Box (AVB), the position and orientation of anatomical structures and the appearance of both AR and VR scenes can be customized. Because the datasets for both the AR and VR scene are the same, unnecessary registration procedures are reduced. The direct feedback loop is implemented so that the controlled registration is generated to further analyze and correct the input parameters by going to the backward steps. This helps the surgeons performing the surgery with the idea for the automatic registration of the 3D objects and helps in performing the surgery by reducing the time to resection the tumor. The proposed system is not accurate for registration without calibration. However, when used by the clinicians, there was positive feedback because it helped them to visualize the 3D models that would save time in the operating theatre while performing the surgery. While creating the VR scene, the simulation might not be as realistic as the real time surgery.

3 State of art

This section presents the features of the current system (highlighted inside the blue broken line in Fig. 1) and limitations (highlighted inside the red broken line in Fig. 1). Venkata, et al. [1] proposed enhanced mean value cloning with multilayer visualization to propose an Enhanced Multilayer Mean Value Cloning (EMLMV) algorithm to improve the overlay accuracy, visualization accuracy and the processing time, as shown in Table 1 The proposed algorithm includes the trimap propagation and alpha matting before merging the two source and target images for removing the smudging and discolored artefacts surrounded by the remote surgeon's hand [1]. Also, volumetric image synthetization is used to recover the background by dimming the foreground for multilayer visualization. It increases the accuracy by reducing the overlay error in merging images from 1.3 mm to 0.9 mm. The visualization error was also reduced by increasing the visibility from 98.4% to 99.1% (visibility of pixels). The processing time remains the same from 10 s per 50 frames to 11 s per 50 frames.

Surgical site (local site) The patient images and information are collected to produce the augmented video. This video is the input to the system that will be merged with the surgeon's hand at the remote site. The frames from the video are extracted and the region of interest is selected in the video frame to merge with the source from the remote site.

Remote site (expert surgeon site) The video from the camera is taken. The guidelines from the expert surgeon are converted to the video frames and the single image of frame is chosen to create a trimap. The trimap is created by selecting the foreground region (being the surgeon's hand), the background region (being the surgical area) and the remaining region (being the unknown region) with the help of the mouse. A trimap propagation technique is implemented to find the contour of the frames [1]. This state of art determines the motion flow between two consecutive frames and the contour flow of current stage is passed to the next stage, which helps to determine the object movements and update to the next iteration. The alpha matte is generated from the trimaps and extracts the foreground objects precisely and removes the smudging artefacts. Although the trimap generation and trimap propagation techniques reduce the visualization errors in the proposed system, the trimap must be created manually. The trimap created manually might not be accurate, resulting in the incorrect visualization of the foreground and background. The

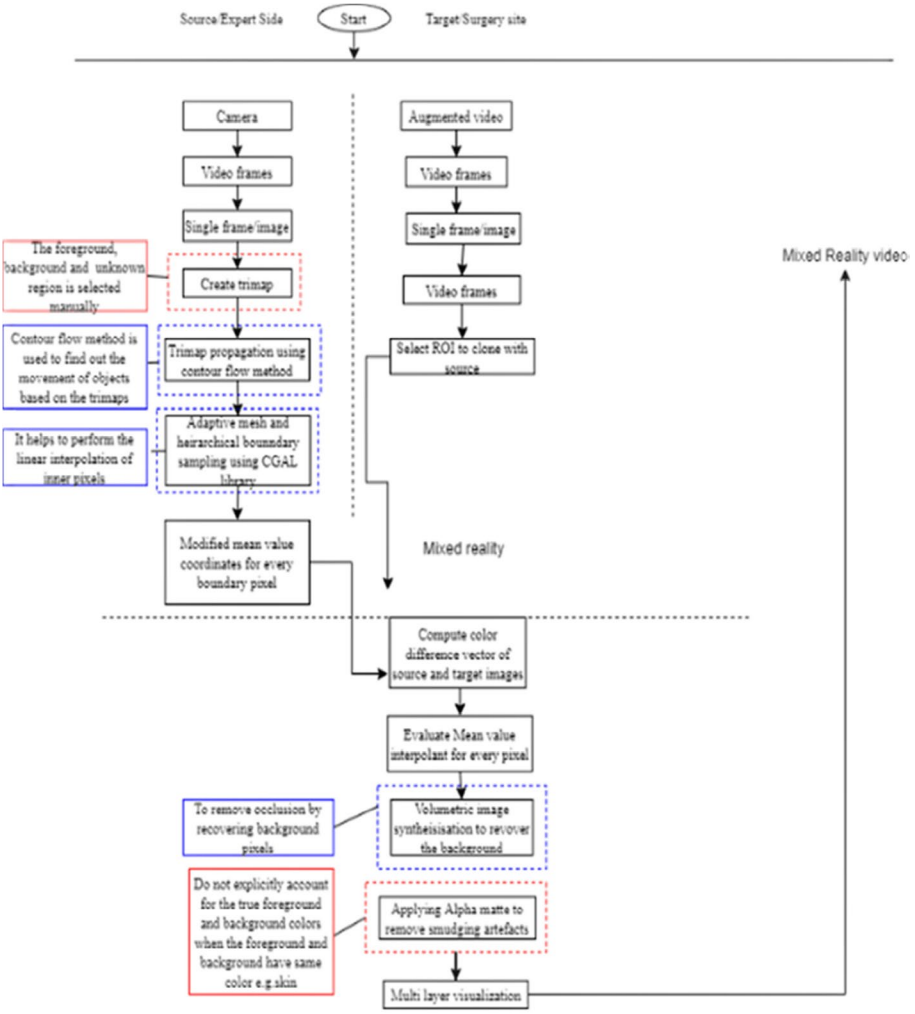


Fig. 1 Block diagram of the State of art system [1]

manual process of choosing the trimaps can be tedious and exhausting. To create an accurate trimap, the user needs to choose the best points; this process requires prior knowledge and experience to choose what the user wants. It is prone to errors in the case of manual trimap creation.

Video composition An adaptive mesh is constructed using the CGAL library based on the peak and low points from the generated trimap. Based on the hierarchical boundary sampling, the inner pixels of the expert surgeon's hand are ignored for the next iteration to reduce the processing time [1]. The iterative mean value cloning method is used to calculate mean values of each pixel until the last frame. The interpolation constraint coefficient is introduced to maintain the color variation in the both source and target images. Volumetric image synthetization is used to remove the occluded pixels to recover the background giving multilayer visualization. Alpha matte is applied to remove the smudging and

Table 1 Enhanced multi-layer means value algorithm

Input: boundary pixels list δ_p, boundary inner pixels list b_{in}, trimap with three regions source image and target image color vector I_t **Output**: merged video without artefacts
BEGIN 1. Generate the trimaps of the current and previous frames. 2. If frame 0 is not equal to frame N, then, calculate the contour flow between the two trimaps using the trimap propagation contour flow 3. Calculate the mean value coordinates of every inner pixel of generated source trimap and the boundary pixel generated in step 2. 4. Repeat steps 2 and 3 for every image to produce the cloned image. 5. Discoloured artefacts are removed in the cloned image using the interpolation constraint coefficient k valued from 0 to 255 which is considered as RGB channels. 6. To smoothen and remove the smudging artefacts, alpha matte is applied. 7. Volumetric image synthesisation is applied to remove occlusion and produce multi-layer visualization. END

discoloration artefacts. Finally, the refined and merged mixed reality video is sent to the local surgery site to guide the local surgeons.

The proposed method does not explicitly account for the foreground and background colors. Alpha matte generation in the proposed system is based on the sampling method that may produce a poor result when the foreground and background are complex and of the same color or contain a variety of textures. Some true samples that are not in the selected trimap may be missed, reducing the accuracy. The presented model reduced the overlay error from 0.9 mm from 1.3 mm and increased the visualization accuracy to 99.1% in terms of pixels, and the processing time was 50 frames per 10 s [1]. Trimap propagation using contour flow is implemented to determine the movement of objects based on trimaps. Figure 2 shows the Flowchart of the state of the art model.

The weight of the inner pixels for calculating mean value coordinates is calculated as:

$$W_i = \frac{1}{||b_i - p||^2} \tag{1}$$

where,
b_i is the boundary pixel of the blending region,
p is the inner pixel.

The final modified composite image without smudging and discoloration artefacts can be given as:

$$I_c = \alpha(p) + I_c(p)(1 - \alpha)I_t(p) \tag{2}$$

where,
I_c is the composite image.
$I_c(p)$ is the cloned region without smudging and discoloration artefacts.
α is the alpha channel or transparency i.e. between 0 and 1.
$I_t(p)$ is the target image.
p is the inner pixel.

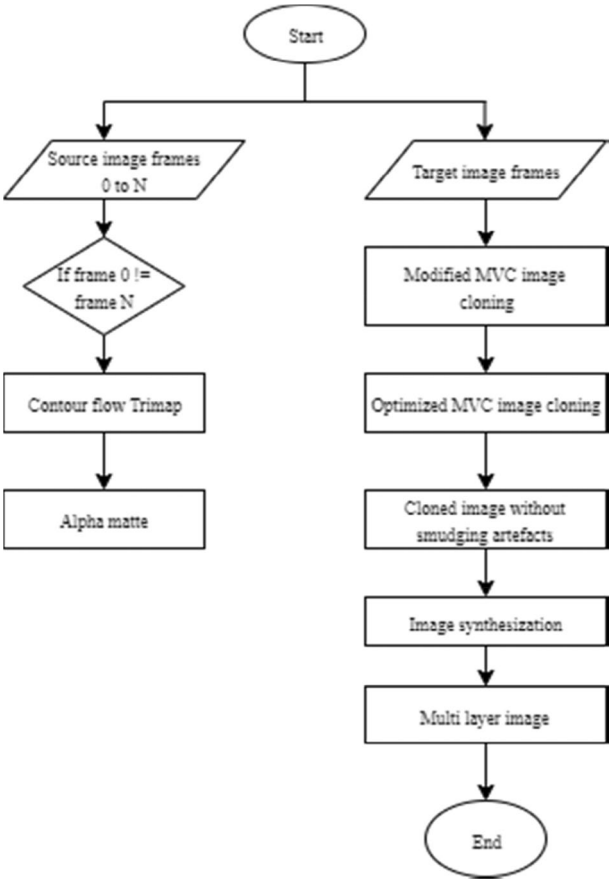


Fig. 2 State-of-the-art multi-layer mean value cloning

4 Proposed system

By studying the relevant papers in the field of video composition using mixed reality, accuracy, processing time, visualization, occlusion, varying illumination, trimap generation process, temporal consistency, multi-layer visualization are the factors that need to be considered. From the papers that were reviewed, [1] the state of art solution in the previous section considered as the basic for the proposed solution from the perspective of trimap propagation using contour flow method to detect the movement of objects based on the trimaps. This helps to determine the movement of objects and updates to the next iteration in the real time scenario, that is, surgery. Also, this paper provides the volumetric image synthetization for multilayer visualization. This allows the surgeon to visualize the background target image by recovering the occluded pixels from the foreground.

Henry and Lee [3] proposed a system to accurately generate trimaps automatically without user interaction. This system reduces the exhausting and manual process of creating the trimaps and reduce the processing time and reduce the risk of inaccuracy by incorrect trimaps. The alpha mattes generated by the trimaps generated by this proposed system has lesser artefacts and they were computed faster. Wang, et al. [2] has proposed spatial–temporal consistent boundary computing by optimizing the energy function to solve the problem

of varying lighting conditions and motion. Also, to create a seamless composition, the illumination guided gradients of the two input frames are mixed to maintain the temporal consistency and optimize the temporal coherency all over the frames. Cai, et al. [6] has proposed earthworm optimization algorithm (EWA) as an efficient metaheuristic algorithm to solve the sample optimization problem to find the best foreground background pair. An optimized cost function is proposed to evaluate the true optimal true foreground pairs on classic evaluation function. The algorithm is used to search the optimal sample pair for every undetermined pixel at the same time in order to improve the efficiency of the search. Furthermore, to improve the search ability and to escape from the local optimum, Cauchy mutation is used. This setup can efficiently find the optimal foreground–background sample pairs and improve the accuracy of matting for sampling-based image matting methods.

The proposed system consists of three stages: surgical site (local site), expert surgeon site (remote site) and video composition. The augmented video from the surgery site is combined with the video from the expert surgeon site to produce the common video giving the mixed reality view, as shown in Fig. 3.

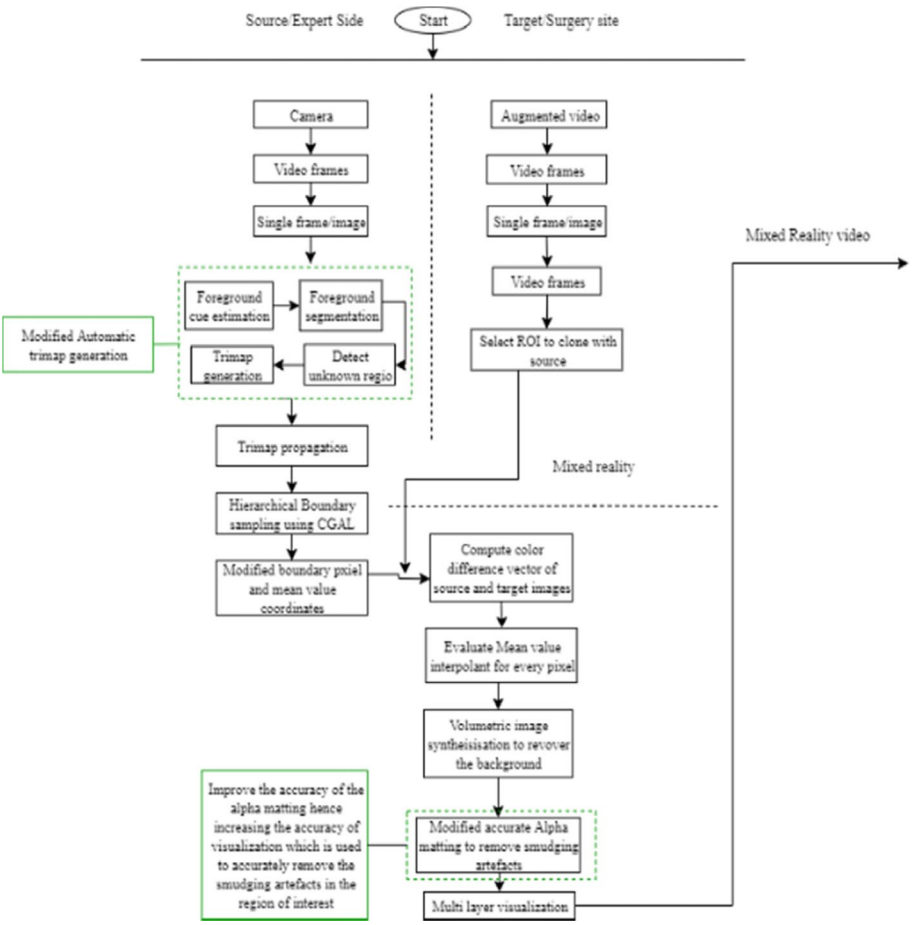


Fig. 3 Proposed MR system for telepresence surgeries

Surgical site (local site) The augmented video is produced from the patient information and images. Video frames from this video are extracted and for each frame, the region of interest (ROI) is selected to clone with the source patch through the Internet connection for video composition.

Remote site (expert surgeon site) The augmented video of the patient is received at the remote site. The camera present at the remote site captures the guidelines and actions of the surgeon to create a video. Video frames from the captured video is extracted to merge with the image frame of the patient. An automatic trimap is generated based on the segmented foreground, that is, the expert surgeon hand and unknown region detection by the combination of the techniques of image saliency, graph cut segmentation, fuzzy c-means clustering [3]. The automatic generation of trimaps saves time because the user is not required to manually create the trimap. The alpha mattes generated from the automatic optimal trimaps have less artefacts. Also, the trimaps generated by the preceding method compute the alpha mattes in less time, which significantly reduces processing time. The technique to generate the trimap does not depend on depth information. Hence, it works directly on RGB color images increasing the practical applicability. The trimap propagation technique is implemented to determine the contour flow between the two consecutive frames to detect the object movements in the real time [1].

Video composition An adaptive mesh is constructed using the CGAL library taking the peak and low points of the automatic generated trimap. Some pixels are ignored for the next iterations to reduce the processing time by using hierarchical boundary sampling [1]. A coherent blending boundary is computed to reduce the color mismatch between source and target pixels. Gradient mixing and enhanced mean value cloning are executed within the obtained boundary. The gradient mixing method is introduced to obtain consistent real composition, which is robust to the varying illumination changes and motion [2]. Also, an interpolation constraint coefficient is introduced to handle the color discrepancy that can occur for the scenes that have large color discrepancy between the sources and patch scenes. Similarly, volumetric image synthetization is used for the multi-layer visualization by recovering the occluded pixels. Accurate and efficient alpha matting is generated based on a discrete bio-inspired metaheuristic algorithm [6]. Because a sampling method is used to choose the pixels for each iteration, the objective is to find the true foreground and background pixels of the corresponding undetermined pixel so that the quality of the matting is optimized. It accurately helps to the remove the smudging artefacts due to discoloration in the region of interest, thus reducing the visualization error.

The mean value coordinate tangent formula was modified with the cotangent formula, which is used to smooth the triangular adaptive mesh that is constructed using hierarchical boundary sampling. The use of cotangent weight compensates the drawback of the irregular meshing completely. This adds the additional effect, called the remeshing effect, which smoothest the deformed surface additionally. The triangulation becomes more regular and the details are preserved. The computed geometric weights such as cotangent become smaller for the edge resulting in the numerical stability and the error is minimized. To prevent negative values that might occur, the extra constant is added to the angles between the boundary pixels.

4.1 Proposed mathematical model

The mean value interpolant for the boundary pixel can be given by [2].

$$w_i = \frac{\tan(\beta_{i-1}/2) + \tan(\beta_i/2)}{||b_i - p||} \tag{3}$$

where,

β_i is the angle between boundary pixels b_i and $b_{i+1,}$

$b_i - p$ is the space distance between boundary point p_i and inner point p

To smoothen the mesh and reduce the error in the frame-by-frame processing, the weight of the ith pixel as given by [5]

Equation 3 has been modified as:

$$Mw_i = \cot(\beta_{i-1}) + \cot(\beta_i) \tag{4}$$

Xie, et al. [20] proposed adding $\frac{\pi}{2}$ factor to prevent the negative value of the coordinate, Equation 4 has been modified by us as:

$$MW'_i = \cot\left(\beta_{i-1} + \frac{\pi}{2}\right) + \cot\left(\beta_i + \frac{\pi}{2}\right) \tag{5}$$

The boundary pixels are calculated from the unknown region of the automatically generated trimap using fuzzy clustering (FCM) on the region of interest proposed by [3]. The FCM is used locally with color so that the mixed pixels are determined based on the membership score in the clusters. The FCM is applied on the boundary region of the foreground object. The boundary region is divided into n*n patches and each patch is converted to the color value in terms of lightness (L), red, green (G) or blue yellow (B). Lightness (L) is discarded and to reduce the computational time, the FCM is applied only to patches containing the non-zero pixels. The formula for the ith pixel in a p_x patch for classifying the mixed pixel is given as:

$$p_x(i) = \begin{cases} 128, & if\ w_i^{max}(i) < w_{threshold}, \\ p_{SF}(i), & otherwise, \end{cases} \tag{6}$$

where,

$p_x(i)$ is the ith pixel in the patch p(x),

$p_{SF}(i)$ is the pixel intensity of the ith pixel of a patch p_{SF} with the same location as p_x in S_F foreground segmentation,

w_i is the set of all partial membership cluster scores for the ith pixel in p_x,

w_i^{max} is the maximum value in w_i, $w_{threshold}$ is the threshold membership score used to classify mixed pixels.

Therefore, the modified weight of the mean value interpolant is given by modifying Eq. 5 to:

$$MW''_i = \frac{MW'_i}{||p_{SF}(i) - \mathrm{p}||} \tag{7}$$

where $p_{SF}(i)$ is the new boundary pixel from automatically generated trimap,

p is the inner pixel.

The alpha matte of the corresponding undetermined pixel of the foreground background pair can be given by [6].

$$\alpha = \frac{(I_k - B_j)(F_i - B_j))}{\|F_i - B_j\|^2} \tag{8}$$

where,

I_k is the kth undetermined pixel in set $I, k = 1,2\ldots N_{I,}$

F_i is the ith pixel in the foreground sample set,

B_j is the jth pixel in the background sample set,

The alpha matte should not only consider the color measure, but should also consider the spatial measure for the true background and foreground pixels. The measurement of color distance determines how well the alpha matte fits into the linear combination of Eq. (8). However, sometimes overlapping of the color distribution of the known foreground and background regions cannot distinguish the true foreground and background colors based on the color distance in the unknown region. Therefore, spatial distance is measured. It measures the distance of the unknown pixel to the candidate sample pair to determine the closeness. The smaller value after the combination of color and spatial distance indicates that the undetermined pixel is closer to the sample foreground and background sample pair.

Therefore, Eq. 8 was modified to Eq. 9 [6]:

$$M\ \alpha = \frac{(I_k - B_j)(F_i - B_j))}{\|F_i - B_j\|^2} + S_k(F_i, B_j) \tag{9}$$

where.

$S_k(F_i, B_j)$ is the spatial distance to eliminate the ambiguity of the color distance is given by:

$$S_k(F_i, B_j) = \frac{\|X_{F_i} - X_k\|}{D_F} + \frac{\|X_{B_j} - X_k\|}{D_B} \tag{10}$$

where,

X_{F_i} is the spatial coordinate of the ith foreground sample,

X_{B_j} is the spatial coordinate of the jth background sample, and.

X_k is the spatial coordinate of the kth undetermined pixel.

D_F and D_B as normalization terms are the nearest distance of the kth undetermined pixel to the foreground boundary and background boundary, respectively.

Finally, enhanced equation is given by modifying Eq. 1 by:

$$EI_c = M\ \alpha\ I_c(p) + \left(1 - M\ \alpha\right) I_t(p) \tag{11}$$

where,

EI_c is the final enhanced composite image,

$M\ \alpha$ is the modified alpha channel or transparency i.e. [0,1].

$I_c(p)$ is the modified cloned region without smudging and discoloured artefacts,

$I_t(p)$ is the target image.

4.2 Area of improvement

Equations for the automatic generation of trimap were proposed by detecting the unknown region, which removes the user intervention to create the trimap and reduces the risk of inaccurate trimap generation. The proposed method generates the optimal trimaps that will

produce the improved alpha mattes by reducing the smudging artefacts and reduce the time to produce the alpha mattes. The alpha value was modified to generate the accurate alpha matte for reducing smudging artefacts by selecting the best foreground background sample pair of the cloned images obtained. The best foreground–background sample pair is calculated by the proposed evaluation function, which will give the true foreground and background samples that might have been missed in the trimap and helps to give accurate visualization. Similarly, the modified mean value cloning was extended with gradient mixing to make the system robust to illumination changes.

Why does clone value mean using automatic trimap generation for accurate image matting ? The proposed system uses mean value coordinates and image interpolation along with automatic trimap generation and efficient and accurate image matting to optimize cloning and visualization. The proposed system uses the unknown region detection technique to automatically generate a trimap to improve the accuracy of the generated trimap and helps to generate accurate alpha mattes in lesser time. The proposed solution considers the true foreground background sample for generating the alpha matte using the optimized energy evaluation function to select the best source and target pairs for image matting. The automatic trimap generation reduces user intervention for creating trimaps. To select the trimap manually requires experience and can be inaccurate, which might give the incorrect visualization. The proposed method solves this issue. Similarly, the alpha mattes from this trimap are generated in less time reducing the processing time. Finally, the proposed system considers the true source and target samples for generating the accurate alpha mattes that might have been missed during the trimap generation or for the complex image natures, which can be used to remove the smudging artefacts for accurate visualization. Table 2 shows the extended mean value algorithm and Fig. 4 shows the flowchart of the proposed Mean value cloning with automatic trimap generation for accurate image matting.

Table 2 Extended mean value cloning with automatic trimap generation for accurate image matting

Input: Input source image
Output: Realistic merged video without artefacts

BEGIN
Step 1: From the input image, calculate the foreground cue F_{cue}
Step 2: Foreground segmentation S_F from the foreground cue F_{cue} and corners as background cue and over segmented input images using lazy snapping technique.
Step 3: Boundary dilation of the segmented foreground to boundary region of interest. Apply FCM color clustering on the boundary region ROI_{FCM} on each patch divided to n*n patches.
$ROI_{FCM} = [(S_F \oplus SE_1) - (S_F \ominus SE_2)] \times I$
Step 4: If the threshold weight is less than 0.8, then it is categorised as unknown pixel which is the boundary pixel for calculating weight for the mean value coordinate.
For each patch in n*n I_{FCM} ,

$$p_x(i) = \begin{cases} 128, & if w_i^{max}(i) < w_{threshold}, \\ p_{SF}(i), & otherwise, \end{cases}$$

Step 5: Until the last frame, calculate the contour flow between the two trimaps using trimap propagation contour flow except the first frame 0.
$MoI_{cf}(p) = (Mox_{\text{current}}, Moy_{\text{current}})$
Step 6: Calculate the mean value coordinates of every inner pixel of the auto generated trimap and for boundary pixel

$$Mo\lambda_i(p) = \frac{W_i}{\sum_{j=0}^{m-1} W_j} i = 0,1,\text{ied}m - 1$$

Step 7: For every image, the above step 5 and step is iterated to produce the cloned image.
Step 8: Discoloured artefacts are removed using interpolation constraint coefficient
Step 9: True source and target sample is selected based on the energy evaluation function and Alpha matte is generated based on the calculated alpha value from the best sample pair to remove the smudging artefacts.
$I(mc) = \alpha(m)MI(cf)(p) + (1 - \alpha(m))I(\text{t})(p)$

Step 10: Multi-layer visualization is achieved using the volumetric image synthesisation by recovering the occluded pixels.
END

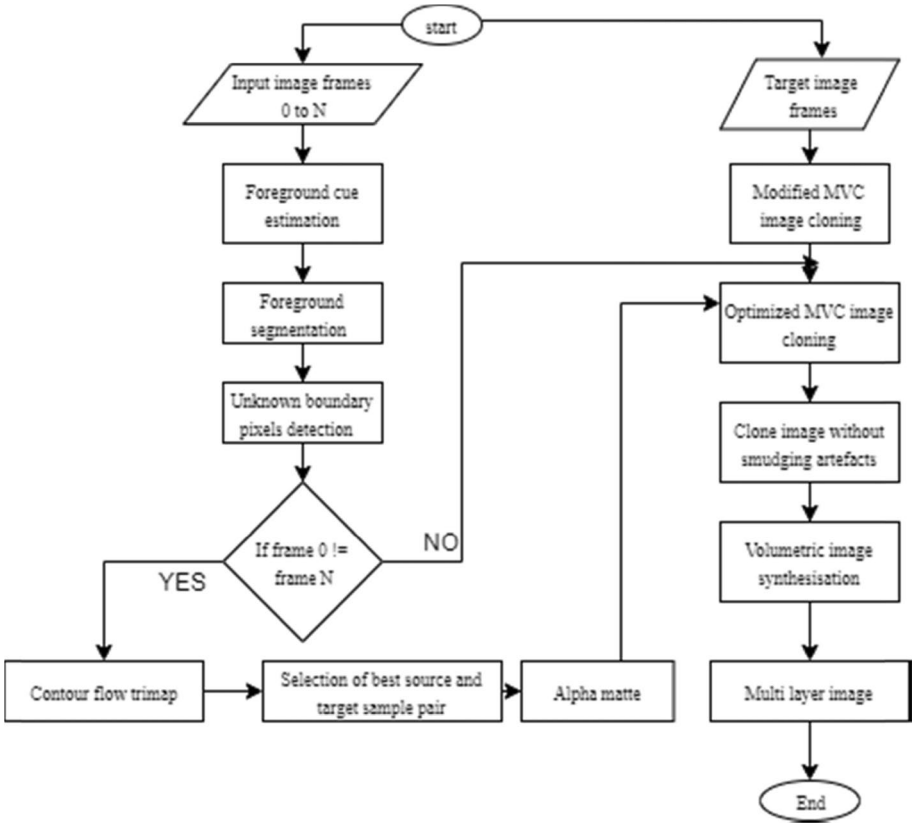


Fig. 4 The proposed mean value cloning with automatic trimap flow chart

5 Results and discussion

MATLAB R2018b was used for the implementation with 10 breast samples (soft tissue), 10 jaw samples (hard tissue) and 10 bowel samples (soft tissue) surgery videos. These sample videos are collected from the online resources based on different age groups and the type of sample. The lengths of the videos are from 6 to 90 s. Frames from the sample videos and the hand gesture frame are extracted from the sample videos and hand captured video that are input to the MATLAB source code. The resolution of images that are extracted for the input are 364 * 360 pixels for the jaw samples, 964 * 565 pixels for the breast samples and 670 * 420 pixels for the bowel samples. These collected samples are tested both for the state of the art and proposed systems shown in Table 1 and Table 2. The tested samples are preplanning of the breast surgery and intra operative of jaw and bowel surgery. These types provide the clarity and visualization of the external hand gesture on the patient region of interest before and after the surgery is performed.

Pre and intra operative stage: Expert surgical assistance is required in many rural parts of developing countries where expertise is lacking for performing the successful surgery. Telepresence surgery is regarded as the best method for the remote collaborative surgery where expertise can guide the surgical procedure irrespective of geographic location.

Venkata, et al. [1] used the enhanced multi-layer mean value cloning for video composition. The authors used manual trimap generation for the source patch that is to be merged with the target. Henry and Lee [3] proposed automatic trimap generation by unknown region detection to detect the trimap automatically for the accurate alpha matting, which increases the visualization accuracy and overlay accuracy. Similarly, alpha matte generation takes place quickly from the generated trimap reducing the processing time.

Figure 5 shows the pre-operative stage, and Figs. 6 and 7 show the intraoperative mixed view of the surgery. 5(a), 6(a), 7(a) are the patient image frames from the augmented video, which consists of pre-operative breast, intraoperative jaw and intraoperative bowel. 5(b), 6(b) and 7(b) are the expert surgeon hand captured by the camera and 5(c), 6(c) and 7(c) are the mixed reality view after the video composition from our proposed system which will be rendered in the screen for the local surgeons for guidance throughout the surgery.

Two samples of soft tissues (breast and bowel) and one sample of hard tissue (jaw) are tested for both the state of art and proposed solutions by using MATLAB. The output data is presented in Tables 3, 4, 5 and 6. The measures that are considered are overlay accuracy, processing time and visualization accuracy. These parameter measures are calculated for each sample of the data collected and its comparison is presented in the bar graphs in Figs. 8, 9, 10 and 11 respectively. Overlay error means the difference between the real scene and projected scene. A visualization error refers to the RGB color difference between

Fig. 5 Proposed system breast sample output

Fig. 6 Proposed system jaw sample output

Fig. 7 Proposed system bowel sample output

Table 3 Result table for accuracy of overlay error and processing time of soft tissue (breast)

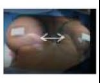
Sample Number	Sample Details	AR video (Patient Images)	Expert Hand	State of art System(Venkata et al., 2019)			Proposed System		
				Processed Sample	Overlay Accuracy	Processing time/frame in seconds	Processed Sample	Overlay Accuracy	Processing time/frame in seconds
1	Breasts (31,5'6'',140)				1.05mm	0.21		0.79mm	0.12
2	Breasts (34,5'2'',190)				0.79mm	0.22		0.55mm	0.12
3	Breast (40,5'7'', 155)				0.79mm	0.2		0.54mm	0.11
4	Breast(255'2'',120)				1.05mm	0.24		0.79mm	0.12
5	Breast (46,5'8'', 142)				0.79mm	0.19		0.55mm	0.11
6	Breast (25,5'3'',160)				0.79 mm	0.2		0.50mm	0.11
7	Breasts (53,5'6'',120)				1.32mm	0.2		0.80mm	0.1
8	Breasts (45, 5'3'',116)				1.05mm	0.21		0.79mm	0.12
9	Breasts(48,5'1'',115)				1.05mm	0.19		0.79mm	0.1
10	Breasts(21,5'4'',100)				0.79mm	0.18		0.50mm	0.1

Table 4 Accuracy of overlay error and processing time of hard tissue (jaw)

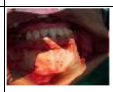
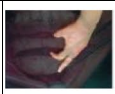
Sample Number	Sample Details	AR video (Patient Images)	Expert Hand	State of art System(Venkata et al., 2019)			Proposed System		
				Processed Sample	Overlay Accuracy	Processing time/frame in seconds	Processed Sample	Overlay Accuracy	Processing time/frame in seconds
1	Lower front teeth				1.05mm	0.23		0.79mm	0.12
2	Lower teeth				0.79mm	0.224		0.50mm	0.11
3	Lower front gums				1.05 mm	0.221		0.79mm	0.12
4	Lower gingiva				0.79mm	0.22		0.55mm	0.11
5	Upper side teeth				0.79mm	0.227		0.5 mm	0.12
6	Upper gingiva				1.05 mm	0.229		0.79mm	0.129
7	Lower gingiva				1.05mm	0.22		0.79mm	0.11
8	Lower gingiva				1.05mm	0.21		0.79mm	0.12
9	Lower molar				0.52mm	0.2		0.30mm	0.1
10	Lower gingiva				0.79mm	0.21		0.50mm	0.11

Table 5 Accuracy of overlay error and processing time of soft tissue (bowel)

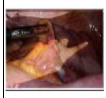
Sample Number	Sample Details	AR video(Patient Images)	Expert Hand	State of art System(Venkata et al., 2019)			Proposed System		
				Processed Sample	Overlay Accuracy	Processing time/frame in seconds	Processed Sample	Overlay Accuracy	Processing time/frame in seconds
1	Bowel 1				1.06mm	0.23		0.79mm	0.12
2	Bowel 2				0.80mm	0.221		0.50mm	0.121
3	Bowel 3				1.06mm	0.22		0.79mm	0.121
4	Bowel 4				0.80mm	0.227		0.50mm	0.122
5	Bowel 5				1.30mm	0.22		0.90mm	0.11
6	Bowel 6				0.80mm	0.21		0.51mm	0.1
7	Bowel 7				0.80mm	0.21		0.55mm	0.1
8	Bowel 8				1.06mm	0.23		0.80mm	0.12
9	Bowel 9				0.80mm	0.227		0.49mm	0.121
10	Bower 10				1.06mm	0.22		0.80mm	0.12

Table 6 Accuracy of visualization error results for breast samples

Sample Number	State of art processed sample	Proposed system processed sample	State of art processed breast sample RGB values			Proposed system processed breast sample RGB values		
			R	G	B	R	G	B
1			107	84	92	80	60	70
2			185	123	197	160	115	177
3			212	211	209	190	190	188
4			160	102	108	140	85	89
5			137	102	75	120	85	60
6			124	78	38	105	60	20

the original scene and the projected scene. The accuracy is measured by the built-in function imtool() in MATLAB. The image is passed as an input to the function, which displays the pixel position and RGB value of the particular pixel of an image in a window. These pixel positions and RGB values determine the accuracy. RGB values of both state of art and proposed solution is presented in Tables 5 and 6 and compared in Fig. 10. The overlay

Table 6 (continued)

7			210	196	88	190	180	70
8			208	128	63	190	115	50
9			148	119	119	130	100	100
10			231	176	138	211	150	120
Average			172.2	131.9	112.7	151.6	114	94.4

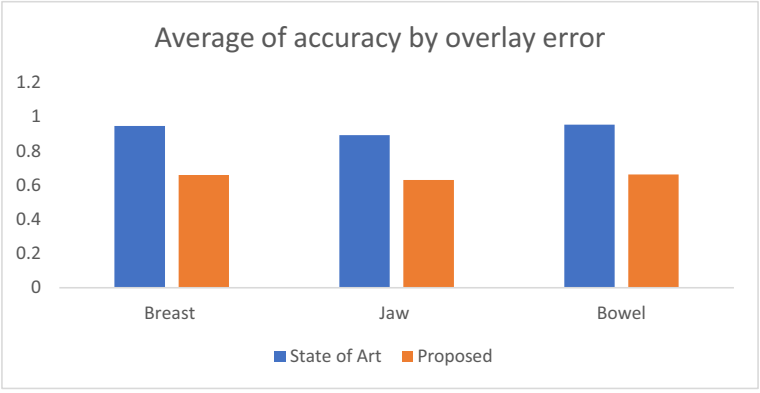


Fig. 8 Accuracy of overlay error in the state of art [1] and the proposed solution for breast, jaw and bowel samples. **a** The first two blue and red bars show the average overlay error in millimetres(mm) for the case of breast samples for the state of art and proposed systems respectively. **b** The second two blue and red bars show the average overlay error in millimetres (mm) for the case of jaw samples for the state of art and proposed systems respectively. **c** The third blue and red bars show the average overlay error in millimetres (mm) for the case of bowel samples for the state of art and proposed systems respectively

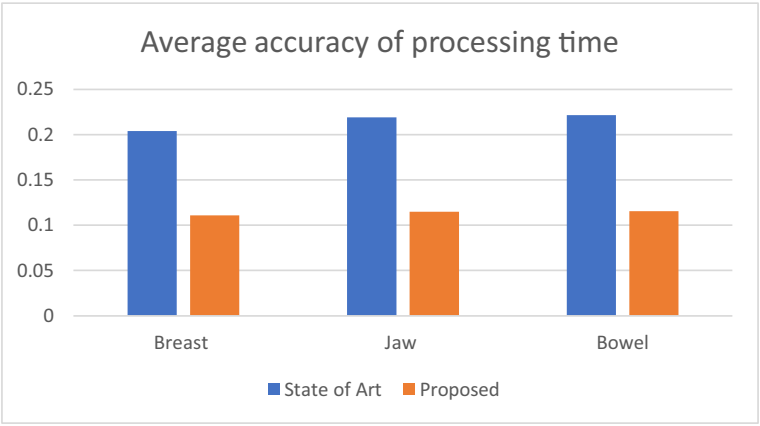


Fig. 9 Average accuracy of processing time in the state of art [1] and the proposed solution for breast, jaw and bowel samples. **a** The first two blue and red bars show the average processing time in seconds for the case of breast samples for the state of art and proposed systems respectively. **b** The second two blue and red bars show the average processing time in seconds for the case of jaw samples for the state of art and proposed systems respectively. **c** The third blue and red bars show the average processing time in seconds for the case of bowel samples for the state of art and proposed systems respectively

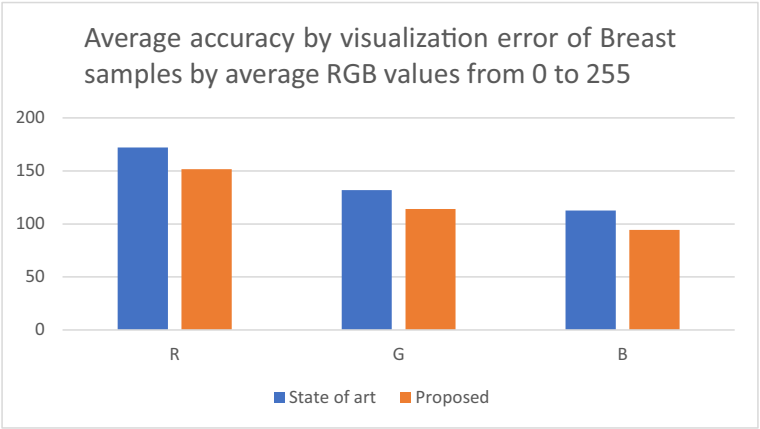


Fig. 10 Average accuracy of visualization for the current and proposed solution with breast samples in RGB. **a** The first two bars are the average red color value of the state of art and proposed solutions respectively, **b** The second two bars are the average green color value of the state of art and proposed solutions respectively, **c** The third two bars are the average blue color value of the state of art and proposed solutions respectively

error and the processing time is similarly shown in the Tables 3 and 4 and compared in Figs. 8 and 9, respectively.

The results of both the state of the art and proposed solutions are compared during video composition performed frame by frame. The visualization of the expert surgeon hand on the region of interest is improved. The foreground segmentation, that is, the surgeon's hand, is accurately merged with the background region of interest blending it seamlessly with less

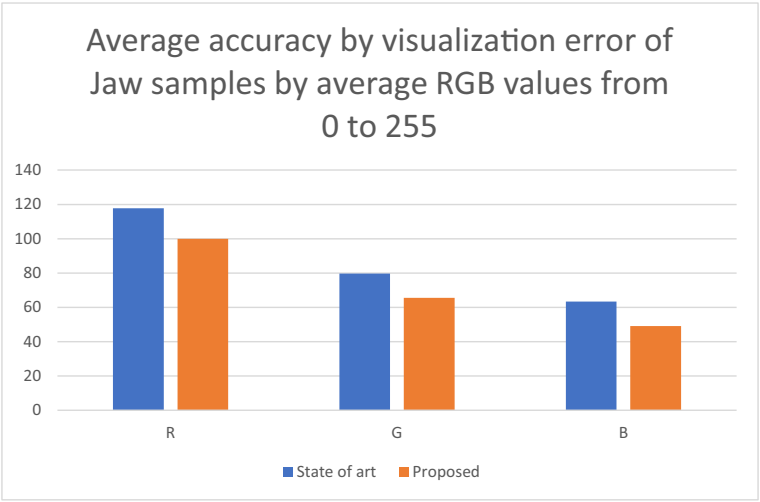


Fig. 11 Average accuracy of visualization for the current and proposed solution with bowel samples in RGB. **a** The first two bars are the average red color value of the state of art and proposed solutions respectively. **b** The second two bars are the average green color value of the state of art and proposed solutions respectively. **c** The third two bars are the average blue color value of the state of art and proposed solutions respectively

noise and smudging artefacts. Similarly, visualization is improved when the lighting conditions are varied, and the results show that the motion difference of the objects in the video frames are addressed. The overlay error and visualization error are reduced compared to the state of the art solution and the processing time was also reduced (Tables 7 and 8).

The results show that there is a significant difference between the state of art and the proposed systems. The comparison is visible in terms of accuracy of overlay error, visualization error and processing time. The overlay error in the proposed system was reduced from 0.93 mm to 0.67 mm. Similarly, the processing time from the state of the art system was reduced from 5 frames per second to 8 frames per second.

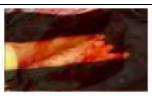
The proposed system also considered the RGB values for adjusting the pixel values, which can be calculated by using the MATLAB image tool. It displays the pixel information on the misplaced overlay and the difference between pixels can be calculated to determine the overlay error. The scale was based as 5 pixels to 1.32 mm. The interpolation coefficient, which is set to 0.08, is introduced. When there is greater brightness in the region of interest during surgery, this value can be increased so that the luminance between the source and target is consistent. Set it to a lower value for preserving the color and texture information of the surgeon's hand so that it is clearly visible during object movement and blood flow during the surgery. Similarly, the final output image was converted to RGB color vectors, as shown in the Tables 5 and 6, and then the comparison between the state of art and the proposed system RGB values is taken to measure the visualization error. The proposed system can modularize the color variation R from 55 to 210, G from 40 to 190 and B from 20 to 188. The average pixel value for the hand color is 122 (the average pixel value for the range 10 to 255) which performs well for all types of hands. The visualization of the proposed system was up to 99.1% for the current system; it was up to 99.7% for the proposed system.

Table 7 Accuracy of visualization error results of jaw samples

Sample Number	State of art processed sample	Proposed system processed sample	State of art processed breast sample RGB values			Proposed system processed breast sample RGB values		
			R	G	B	R	G	B
1			169	117	83	150	110	70
2			121	93	74	100	80	60
3			135	65	61	120	50	45
4			115	83	57	100	70	40
5			88	71	63	60	55	50

Automatic trimap generation required no user intervention to create a trimap, which reduced the processing time for real time video composition. The alpha matte generated from the automatic trimap took less time, which also reduced the processing time. The enhanced alpha matting that considers both color and spatial measures gave accurate natural image matting, which increases the visualization on the final merged image. Trimap propagation was used to transfer the flow from one frame to another frame, which addressed the motion difference and the interpolation coefficient factor helped to blend seamlessly making the luminance consistent for both source and target images. The processing time was reduced as the proposed system used the automatic trimap generation and the alpha matte generated from this trimap was accurate, resulting in the increased visualization. The enhanced natural alpha matting helped to seamlessly blend the target to background removing the smudging artefacts and discolored region. The processing time

Table 7 (continued)

6			87	52	39	70	35	25
7			160	135	112	140	120	100
8			112	65	51	105	50	35
9			72	55	45	55	40	30
10			118	61	48	100	45	35
Average			117.7	79.7	63.3	100	65.5	49

was reduced to 8 frames per second compared to 5 frames per second of the state of art. The visualization of pixels was up to 99.7% compared to 99.1% from the state-of-the-art solution. Similarly, the overlay accuracy was also reduced from the 0.9 mm to 0.5 mm. See Table 9.

6 Conclusion and future work

The proposed solution seamlessly blends the two videos without smudging and discolored artefacts and provides a realistic final video. The extended mean value cloning smoothes the differences present in the target and source frames for seamless and realistic blending of pixels. Automatic trimap generation reduces the risk of false foreground selection and the optimal generated trimaps improve the alpha matte quality, which is again optimized to reduce the smudging artefacts completely and produce the accurate visualization of the final merged image. This solution downgraded the overlay error from 0.93 mm to 0.67 mm and upgraded the visualization accuracy by increasing the visibility of pixels from 99.1%

Table 8 Accuracy of visualization error results of bowel samples

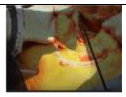
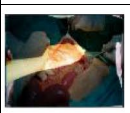
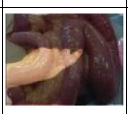
Sample Number	State of art processed sample	Proposed system processed sample	State of art processed breast sample RGB values			Proposed system processed breast sample RGB values		
			R	G	B	R	G	B
1			180	130	100	160	110	80
2			130	60	60	103	45	56
3			116	80	55	100	63	41
4			90	70	61	69	55	47
5			161	134	110	143	120	98
6			111	64	50	93	50	36
7			115	59	48	100	48	37
8			160	115	81	139	99	60
9			88	69	61	75	55	49
10			158	132	111	141	120	100
Average			130.9	91.3	73.7	112.3	76.5	60.4

Table 9 State of art and propose solution comparison

	Proposed Solution	State of the Art Solution
Name of the solution	Extended Mean value cloning with automatic trimap generation and accurate image matting	Enhanced multi-layer mean value cloning
Proposed equation	Modified Final composite image with modified alpha value is given by $I(mc) = \alpha(m)MI(cf)(p) + (1 - \alpha(m))I(\mathrm{t})(p)$	Final modified image composition is given by $I_{mc} = \alpha MI_c(p) + (1 - \alpha)I_t(p)$
Accuracy	• Overlay error from 0.93 mm to 0.63mm • Visibility of pixels from 99.1% to 99.7%	• Overlay error 0.9mm from 1.3mm • Visibility of pixels from 98.4% to 99.1%
Processing Time	8 frames in a second	5 frames in a second
Contribution 1	Automatic trimap generation without user intervention for automatically finding the unknown region	Manual trimap generation by user having to select two points in the video
Contribution 2	Selection of true source and target pairs for accurate image matting	Alpha matte is generated based on the trimap only

to 99.7%. The alpha matte generation process was fast, reducing the processing time from 5 frames per second to 8 frames per second. Currently, video composition occurs after the video from the local and remote site are converted to the frames. In the future, sending the video directly without converting to the frames will save time. This method would also removes noises that might occur due to blood flow during the surgery.

Funding Open Access funding enabled and organized by CAUL and its Member Institutions

Declarations

Conflicts of interests There is no conflict of interest.

References

1. Venkata HS et al (2019) A novel mixed reality in breast and constructive jaw surgical tele-presence. Comput Methods Programs Biomed 177:253–268
2. Wang J, Sheng B, Li P, Jin Y, Feng DD (2019) Illumination-guided video composition via gradient consistency optimization. IEEE Trans Image Process
3. Henry C, Lee S-W (2019) Automatic trimap generation and artifact reduction in alpha matte using unknown region detection. Expert Syst Appl 133:242–259
4. Li J, Yuan G, Fan H (2019) Generating trimap for image matting using color co-fusion. IEEE Access 7:19332–19354
5. S Yehu, W Lei, X Qiming, P Zhenyun, W Qicong (2015) A simple real-time image cloning algorithm based on modified mean-value coordinates, presented at the 2015 International Conference on Control, Automation and Information Sciences (ICCAIS)
6. Cai Z-Q, Lv L, Huang H, Liang Y-H (2019) A discrete bio-inspired metaheuristic algorithm for efficient and accurate image matting. Memet Comput 11(1):53–64
7. Donghyeon C, Sunyeong K, Yu-Wing T, In So K (2017) Automatic trimap generation and consistent matting for light-field images. IEEE Trans Pattern Anal Mach Intell 39(8):1504–1517
8. Pawin P, Jakkrit D, Toshiaki K, Itthisek N (2019) A real-time hand segmentation method using background subtraction and color information. Songklanakarin J Sci Technol (SJST) 41(2):436–444
9. Chaves-González JM, Vega-Rodríguez MA, Gómez-Pulido JA, Sánchez-Pérez JM (2010) Detecting skin in face recognition systems: A colour spaces study. Digit Signal Process 20(3):806–823
10. Jerripothula KR, Cai J, Yuan J (2016) Image co-segmentation via saliency co-fusion. IEEE Trans Multimedia 18(9):1896–1909
11. Chen T, Zhu JY, Shamir A, Hu SM (2013) Motion-aware gradient domain video composition. IEEE Trans Image Process 22(7):2532–2544
12. Hu Q, Sun H, Li P, Shen R, Sheng B (2018) Illumination-aware live videos background replacement using antialiasing optimization. Multimedia Tools Appl 77(18):24477–24497
13. Gastal ESL, Oliveira MM (2010) Shared sampling for real-time alpha matting. Computer Graphics Forum 29(2):575–584
14. Wang P et al (2019) 2.5DHANDS: a gesture-based MR remote collaborative platform. Int J Adv Manuf Technol 102(5–8):1339–1353
15. Anton D, Kurillo G, Bajcsy R (2018) User experience and interaction performance in 2D/3D telecollaboration. Futur Gener Comput Syst 82:77–88

16. Basnet BR, Alsadoon A, Withana C, Deva A, Paul M (2018) A novel noise filtered and occlusion removal: navigational accuracy in augmented reality-based constructive jaw surgery. Oral Maxillofac Surg 22(4):385–401
17. Kalal Z, Mikolajczyk K, Matas J (2012) Tracking-Learning-Detection. IEEE Trans Pattern Anal Mach Intell 34(7):1409–1422
18. Hettig J, Engelhardt S, Hansen C, Mistelbauer G (2018) AR in VR: assessing surgical augmented reality visualizations in a steerable virtual reality environment. Int J Comput Assist Radiol Surg 13(11):1717–1725
19. Perkins SL, Lin MA, Srinivasan S, Wheeler AJ, Hargreaves BA, Daniel BL (2017) A mixed-reality system for breast surgical planning, presented at the 2017 IEEE International Symposium on Mixed and Augmented Reality (ISMAR-Adjunct)
20. Xie Z-F, Shen Y, Ma L-Z, Chen Z-H (2010) Seamless video composition using optimized mean-value cloning. Vis Comput 26(6–8):1123–1134

Authors and Affiliations

Roshan Dallakoti[1] · Abeer Alsadoon[1,2,3] · P. W. C. Prasad[1,2] · Sarmad Al Aloussi[4] · Tarik A. Rashid[5] · Omar Hisham Alsadoon[6] · Ahmad Alrubaie[7] · Sami Haddad[8,9]

✉ Abeer Alsadoon
alsadoon.abeer@gmail.com

[1] School of Computing, Mathematics and Engineering, Charles Sturt University (CSU), Wagga Wagga, Australia

[2] School of Computer Data and Mathematical Sciences, Western Sydney University (WSU), Sydney, Australia

[3] Asia Pacific International College (APIC), Sydney, Australia

[4] Computer Technology and Information Management Department, Massasoit Community College, Brockton, MA, USA

[5] Computer Science and Engineering, University of Kurdistan Hewler, Erbil, KRG, Iraq

[6] Department of Islamic Sciences, Al Iraqia University, Baghdad, Iraq

[7] Faculty of Medicine, University of New South Wales, Sydney, Australia

[8] Department of Oral and Maxillofacial Services, Greater Western Sydney Area Health Services, Sydney, Australia

[9] Department of Oral and Maxillofacial Services, Central Coast Area Health, Gosford, Australia