Available online at www.sciencedirect.com

ScienceDirect

IFAC PapersOnLine 53-5 (2020) 792–796

From AR to AI: Augmentation Technology for Intelligent Surgery and Medical Treatments

Mei Zhang [1], Zhicheng Zhang[2], Xiao Wang[1], Hui Yu[3], Yifan Xia[4], Kanran Tan[5], Xiao Wang[6], Fei-Yue Wang[1]

1. The State Key Laboratory for Management and Control of Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China, and also with the Qingdao Academy of Intelligent Industries, Qingdao 266109, China (e-mail: mei.zhang@ia.ac.cn; feiyue.wang@ia.ac.cn; x.wang@ia.ac.cn).
2. The Department of the 7th Medical Center of PLA General Hospital, Beijing 100700, China (e-mail: dr_zhangzhicheng@126.com)
3. The School of Creative Technologies, University of Portsmouth, UK. (e-mail: hui.yu@port.ac.uk)
4. The School of Creative Technologies, University of Portsmouth, UK. (e-mail: Yifan.Xia@myport.ac.uk)
5. The Department of Computer Science, Johns Hopkins University, Baltimore, Maryland, USA. (e- mail: ktan7@jhu.edu)
6.The Department of Computer Science, Purdue University, West Lafayette, Indiana 47906, USA. (e-mail: wang3702@purdue.edu)}

Abstract: With the development of augmented reality (AR) technologies, more and more approaches are proposed for medical applications. With the help of AR technology, the doctor can highly improve the spatial perception and obtain more information beneath the displayed image. Generally, 3D reconstruction, registration, tracking and depth visualization are all important for an AR system, which directly determine the accuracy of the system. However, none of recent systems have achieved a perfect 3D vision for doctors. It is particularly challenging for the AR system to perform well for operation on highly deformable body tissues. Therefore, there is an urgent need to improve the 3D vision and intelligence of AR systems. In this paper, we provide a review of the recent development of augmentation technologies for surgery and medical treatment. And then, we introduce the parallel intelligence theory into AR systems, which can provide a feasible approach to enhance the efficiency.

Keywords: Potential impact of automation and open problems, Augmented reality, Medical treatment, Parallel theory, 3D technology

1. INTRODUCTION

In recent years, the augmented reality technology has achieved rapid development, and been applied in many areas, such as mechanical assembly, entertainment and education. Different from virtual reality (VR) technology that immersed users into a computer-synthesized virtual world, augmented reality (AR) makes the computer-generated virtual information be integral to the physical world. The characters of AR are highlighted in virtual and real objects alignment and good interactivity in real-time, which enables wide applications in medical practice. In 1986, Robets *et al.* (Roberts 1986) presented the first medical augmented reality system applied in neurosurgery. The system integrated segmented computed tomography (CT) images into the view of an operation microscope. The registration of CT image data is determined by the position of the microscope, which is controlled by a non-imaging ultrasonic tracking system. About two decades later, the applications of AR in endoscope surgery were proposed. Recently, the AR technology was widely applied commonly by assisted surgery systems for neurosurgery, laparoscopic surgery and so on (Bernhardt 2017). In short, it had achieved outstanding performance on three-dimensional visualization, process navigation and computer-human interaction.

The advantages of AR in medical applications can be summarized as follows:
(1) More extra information can be fused into the view of the user. Generally, different types of medical devices are needed to capture multiple data from patients, which are hard works for the doctor to imagine the structure of the tissue with extremely high accuracy. Obviously, AR provides a platform for better understanding of the anatomical context by integrating all the information into one display. Furthermore, the precious human experience can be visualized and coupled into the surgical workflow, which can highly improve the intelligence and efficiency of the treatment.
(2) Three-dimensional (3D) visualization of the tissues can be achieved. Generally, surgeons can only observe the outer side structure of the tissues during the operation with the naked eyes. With the help of augmentation technologies, 3D information of the underneath anatomical context captured by various sensors and device can be integrated together, which helps the surgeons to fully understand the tissues and the relative position between the surgical instrument and the operative site.
(3) Interaction between the computer and human can be implemented with an excellent presentation. The more information it is visualized, the more activities it can be handled by the doctor. Applying different kinds of sensors,

10.1016/j.ifacol.2021.04.174

users can finish a special interaction with the augmented information, which provides an extension of the spatial perception and manipulation ability for the surgeons.
According to the principle of superimposing the virtual images into the real object, augmented reality systems can generally be classified as video-based AR system, see-through AR system and projection-based AR system (Nicolau 2011). Generally, the registration, tracking and depth visualization are all key parts for the AR system.

2. CLASSIFICATION OF AUGMENTED REALITY SYSTEMS IN MEDICAL TREATMENT

2.1 Video-Based Augmented Reality System

The video-based AR system is implemented by overlaying the virtual image onto the real scene, which is generally performed on an external display device. The critical issue is to obtain an accurate registration between the real image and the augmented model. It is favourable to modulate the brightness and perform certain graphical effects by using this kind of AR systems. However, the field of view shown on the external display device is limited.
The video-based method is the easiest way to achieve augmented reality display. In 1998, Sato *et al.* (Sato 1998) registered 3D ultrasound tumor models to live video images of the patient's breast, which make the position information visible. In minimally invasive surgery, the AR technique is applied more and more commonly. In 2004, Nicolau *et al.* (Nicolau 2004) proposed a camera-based AR system for minimally invasive liver ablation. Several markers were distinguished to obtain the registration of the camera to the 3D model of the liver.
For minimal endoscopic surgery, the viewing angle of the image displayed on external display devices is dependent on the pose of the endoscope. Therefore, a tracker is needed to obtain the pose of targets, which helps to transform the coordinate of targets to world coordinates. Several methods were imposed to track the endoscope, the surgery tools and the patient. Souzakiet *et al.* (Souzakiet 2013) proposed a system to augment images from the CT and MRI data in thoracic endoscope surgery. Buchs *et al.* (Buchs 2013) added an optical tracking system into the image-guided robotic surgery hepatic lesions. After finishing the endoscopic calibration and patient registration, the virtual tumor model and the tools were overlaid onto the endoscopic image in real time. The doctor can have a clearer understanding of the entire surgical scene through the AR results.
The natural features of the body tissues can be utilized to track the video images, which can highly reduce the use of external-aided devices. Puerto-Souza *et al.*(Puerto-Souza 2013) investigated a feature points matching method to realize the video augmentation display. The 3D coordinates of feature points on 3D CT models were calculated, which were then used to obtain the registration with 2D images in the endoscopic view. Long-term tracking of critical tissues with high accuracy could be ensured by further processing the problem of tracking loss. Collins *et al.* (Collins 2014) utilized the outline information of the organ to provide constraint conditions for registration. Le Roy *et al.* (Le Roy 2019) aimed at an accurate registration for the deformable liver, which was obtained by using the anatomical landmarks and the natural silhouette.
With the development of electronic technologies, a mobile phone or panel computer can also be used as a display device. Cho *et al.* (Cho 2017) conducted a study on bone tumor resection using a flat panel or a mobile phone as a terminal for enhanced display. The positioning of these mobile devices and scenes was achieved by identifying a two-dimensional code.

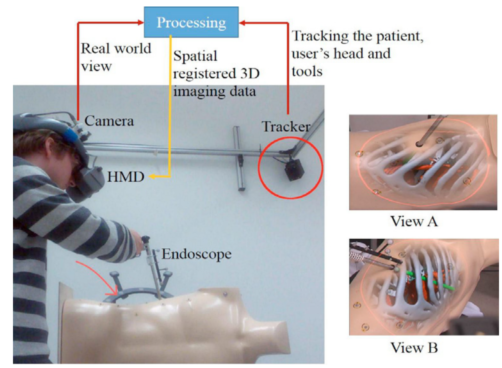

Fig. 1. Endoscopic surgery by use of a video-base HMD

The external display devices showing only 2D images, do not have the ability to provide 3D visual feedback for users, and thus fail to convey the 3D information of the surgery scene. The obvious advantage of 3D display is the enhancement in depth perception (Geng 2013). The head-mounted display (HMD) devices are the earliest applications to provide depth information in medical areas, which keeps a fixed distance between the display devices and the human eyes. The HMD system outputs two images with parallax to the left and right eyes, which automatically generates 3D visualization. The pose of the helmet needs to be detected in real-time, which is used to register the virtual model and the real scene. For the video-based HMD, additional miniature cameras are set on the helmet to record the image from the view of users. The pose of HMD, patient, endoscope and surgical tools are all obtained by the tracking system. The composite images are finally displayed on the screen of the HMD. If the speed of the tracking algorithm and image processing is not high enough, the update of the displayed image will be delayed from the motion of the human head, which will definitely affect the accuracy of surgery. Bichlmeier *et al.* (Bichlmeier 2007) developed a video-based HMD system to superimpose the anatomical structures images onto the patient. The surgical instruments were tracked and integrated into the HMD display. The reasonable depth cue motion parallax was considered to match the changes of the viewing angle. This system could be applied for medical AR training and education applications.
The AR systems with three-dimensional display devices are also explored, which can provide a natural 3D perception. Since the images with right parallax are directly provided for the two eyes of human in the 3D display, no tracking system is needed for the visual direction of human. It requires 3D-to-3D registration to build superimposed models. According to different design approaches, the three-dimensional display can be divided into spectacle display and glass-free display. The glass-free 3D display has received more and more attention due to its prominent characteristics. One of the most

popular 3D display devices is multi-view stereo display, which is functioned by attaching a cylindrical grating onto traditional 2D display devices. Different images with parallax are generated at a particular viewing angle, thereby the binocular parallax and motion parallax are both provided and multiplayer observation is allowed. Despite some weakness of limited resolution and spatial accuracy, it has vast application prospects in the operation room (Geng 2013). Abildgaard *et al.* (Abildgaard 2010) further illustrated that the autostereoscopic display had obvious advantages over traditional 2D display in medical image visualization. The visualization for tiny information is much easier to be distinguished by using the autostereoscopic display. Zhao *et al.* (Zhao 2016) proposed a floating autostereoscopic display system applying an integral videography (IV) display to realize 3D visualization with full parallax. Since complex processing were required to provide the 3D reconstruction for the real scene, only certain operations could utilize this kind of prototypes.

2.2 See-Through Augmented Reality System

The type of see-through AR systems was generally operated by introducing a semi-transmitting element into the field-of-view of users. The user can directly view the real scene through the semi-transmitting element, while the virtual images are simultaneously reflected into the view of human eyes. Comparing with the video-based AR system, more tactile feedback and a larger field of view can be achieved. But it is difficult to make the real scene be completely obscured by the perspective superimposed images.

The optical see-through HMD display is the most common system, in which a semi-transparent mirror is mounted. Different from the video-based HMD, the real scene can be naturally viewed through the optical elements of the see-through HMD. Therefore, the field-of-view is much larger and the problem of display delay of real scene can be avoided. The results of tracking and registration are very important to give the correct relationship among the real scene, display device and the eye in real time. If the head movement or real scene changes too fast, the delay of virtual models display may occur. Hamza-Lup *et al.* (Hamza-Lup 2007) proposed an optical see-through HMD system for lung surgery, which was implemented by integrating the optical tracking system and natural markers registration algorithms. Chen *et al.* (Chen 2015) applied an AR surgical navigation system to improve the safety and reliability of the surgery. Both the point-to-point registration and surface registration were adopted to determine the relationship of the virtual coordinate system and real coordinate system. After patient-to-display calibration, the virtual models aligned with the real scene could shift freely with the motion of head.

The semi-transparent elements can also be inserted into a stereoscope or microscope to implement augmented reality. But no calibration for the eye-to display is required. Different from the see-through HMD display, the display device generally set at a relatively fixed position to the patient. The first augmented stereo microscope was proposed by Edwards *et al.*(Edwards 1995). After patient-to-image registration, the 3D model is projected into the microscope eyepieces with correct superposition. The system was evaluated with a head phantom, which showed the accuracy with 0.3-0.5 mm. Berger *et al.* (Berger 1999) superimposed the photographic and angiographic images onto a bio-microscope. The natural features of retinal blood vessel were considered as excellent markers to track the retinal position. In 2019, Google introduced an AR device into a traditional microscope to diagnose cancer intelligently in real-time (Razavian 2019). The deep-learning algorithm was developed to distinguish the cancer cell in the specimen with an accuracy rate of more than 90%. The results were labelled and re-projected back to the optical path of the microscope. Although these systems above can provide good display, none of them overcome the depth problem of visualization.

The see-through AR display can also be implemented by directly fixing a semi-transmitting glass between the doctor and the patient. By using this AR system for surgery, no hand-eye coordinate problems occur.

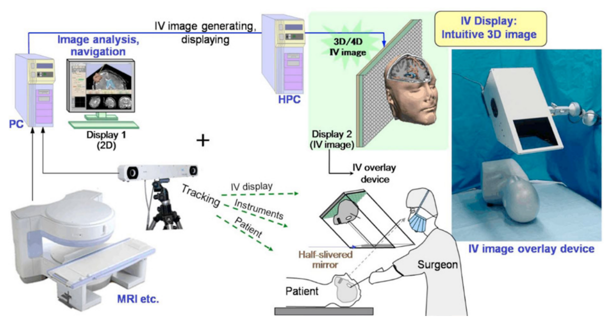

Fig. 2. The see-through AR system based on integral videography

Masamune *et al.* (Masamune 2002) placed a spectroscope on the CT scanner to superimpose the CT image onto the real scene. A registration method using triangular markers was further developed. However, this system only allowed for the observation at a specific angle. Liao *et al.* (Liao 2010) implemented an AR system applying integral videography, as shown in Figure 2. An optical tracking system was functioned to obtain the registration of the real patient body, the virtual images and the surgical instruments. An integrated imaging device and a semi-transparent mirror were fixed into the display part, which superimposed 3D images onto the patient in situ. No glasses and tracking devices for head motion were required in this system, while a 3D display with full parallax were achieved. The same display method also has been applied to dental surgery (Suenaga 2015). The disadvantage of this kind of AR system is the limited field-of-view and resolution constrained by the integrated imaging devices.

2.3 Projected-Based Augmented Reality System

The projected-based AR system is achieved when the virtual images are directly projected into a real scene. The key issues of this kind of AR systems include the registration of projection data to patients and the calibration, which need to be suitable for shape variation of the surface curvature (Shuhaiber 2004). Generally, virtual images are projected on the external surface of the patient, which is more suitable to project superficial anatomical structures, surface maps and so on. Therefore, the display information is directly observed without wearing any other devices and can serve for multiple persons simultaneously. However, since the projection image is 2D, the 3D information of the reconstructed models is inevitable lost.

Gavaghan *et al.* (Gavaghan 2011) developed a portable projection-based AR device applying the miniature laser projection technology, as shown in Figure 3. During the operation, images of 3D patient-specific models are directly projected onto the organ surface without any other intrusive hardware around. The most important work in this AR system is to accurately update the alignment of the virtual scene with the patient. Such a hand-held portable projection device can alleviate the installation problems of other display devices.

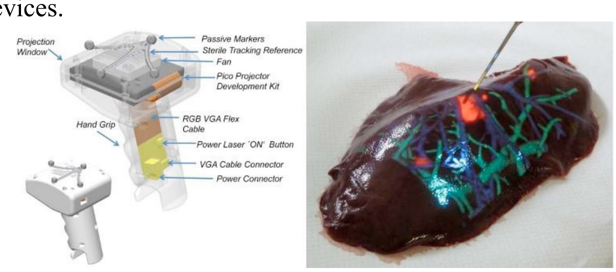

Fig. 3. The prototype and results of the portable projection-based AR system

3. DISCUSSION AND PERSPECTIVE

3.1 Further Improvements of Current AR Technologies

As an efficient concept used for medical treatment, augmented reality combined with other advanced technologies can help physicians to handle more information at one time and get better surgical navigation. Many AR technologies are introduced for pre-operative planning, which can help doctors to better understand the patient anatomy and the location of lesion. Recently, AR-based intra-surgical navigation systems have been increasingly introduced for facial contouring, maxillofacial surgery, bone tumor resection, neurosurgery and so on. It is very acceptable for physicians to enjoy any benefit by applying the augmented reality into the clinical workflow without changing the environment too much. Zhu *et al.* (Zhu 2017) developed an AR system to provide precise navigation for maxillofacial surgery. The reconstructed 3D model from CT data was superimposed onto the real environment, which was presented on the semi-transparent helmet. And 20 patients were surgically treated using the proposed AR system. It was shown from the clinical outcomes that the AR system could help surgeons to obtain intuitive visualization of interest regions with high accuracy. For the neurosurgeon and laparoscopic surgeon, it is very important to determine the tumor location and tumor resection boundary with no threat to surrounding tissues and blood vessel. Bourdel *et al.* (Bourdel 2017) applied AR techniques in laparoscopic gynecology to localize the myomas for 3 patients. The 3D model of the soft uterus was constructed using pre-operative MRI data, which then was registered with the intra-operative video. The AR technology greatly improved the accuracy of the localization for small myomas that was inadequate during laparoscopy. Consequently, AR technologies can greatly help the physicians to obtain trajectory planning, improved visualization, lesion localization and spatial construction of tissues, which can enhance the precision and the efficiency of the surgery.

However, as for current development of augmentation technologies some further improvements are still needed to enable physicians to fully and easily interact with virtual data.

（1） The visualization of accurate depth perception in a complex environment is very important. Especially in minimally invasive surgery, the field of view is limited, which would greatly deteriorate the spatial perception and localization ability of the physician. The issue of wrong depth perception will make the physician misunderstand the integration of the virtual and real images. Although some methods and technologies have been proposed to improve the depth sensation, some visual conflict cases still exist. It is hard to get the correct positional relation of objects, if the virtual images are simply overlaid on the real images. As we know, the depth cues are classified as psychological and physical cues. Four major psychological cues for the human brain used to obtain 3D sensation include linear perspective, shading, texture and prior knowledge. When focusing on targets with different distances, the convergence angle and accommodation of the two eyes are synchronously changed. However, for the type of AR like **HMDs**, the real focus planer and the depth of the virtual images presented do not match for human eyes, which cause the verge-accommodation conflict. Additionally, the shading and the blurriness of targets change with the variation of viewpoint in real time. However, it is not easy to match all the requirements. Currently, several types of AR visualization are provided to enhance the 3D sensation, such as transparent overlay, wireframe, virtual window, random points window and ghost window. We have made a comparison between different types of AR visualization implemented on a 3D grating LCD (Liquid Crystal Display) display device. The correctness and visibility of the five types were not identical for different people (Wang 2018). For further works, the AR system should provide real 3D visualization of the tissues without disturbing the actual surgical field.

（2） Intelligent approaches are needed to handle complex surgical situations. Multiple emergencies may occur during the operation process. But current systems are generally suitable to run in a defined condition, which can not provide a continuous help for physicians. Therefore, some further works are required to provide better solutions to physicians. On the one hand, highly accurate tracking and registration technologies are needed to update the information of patient in real time. On the other hand, some reliable processing is needed to intelligently predict the development and potential risk for the doctor, so as to provide effective guidance for doctors.

Various kinds of tracking and registration technologies have been proposed by researches. But different technologies have their own inherent advantages and drawbacks. For example, the optical tracking method can provide accurate results, while it is hard to work for flexible tube endoscope. The deformable tissue can be tracked by the intra-operative CT scanning. But it is still difficult to obtain the intra-operative deformation of intrahepatic. Applying the intra-operative ultrasound image, the deformation can be roughly estimated

(Lang 2009). Therefore, an effective fusion method integrating the above methods can be a potential solution to the tracking issue.

Generally, the operations like orthopaedic surgery and maxillofacial surgery may cause little deformation of tissues, so rigid registration methods work well for these situations. But for soft tissues or some surgical procedure, such as tissue cutting, the virtual tissue model constructed from preoperative data cannot be continuously aligned with the real scene without accurate deformable registration. Simpon *et al.* (Simpon 2013) proposed a method for deformation problems applying both rigid alignment and non-rigid correction. Usually, the registration provides surface matching results based on fiducially markers or skin surface identification. But it is much more important to obtain the internal deformation for correct superimposition of the invisible structure beneath the surface. In a short, we need to obtain not only the deformation parameters of the organ surface but also the internal deformation model to achieve the accurate non-rigid registration. Dumpuri *et al.* (Dumpuri 2010) tried to simulate the liver deformation patterns to match the surface images obtained. However, the deformation between different tissues is always independent. Therefore, it is critical to create a reliable model of the organ considering various types of information, including the deformation itself, adjacent viscera and so on. Based on the virtual model consistent with the actual condition, the trend of the disease is hopeful to be estimated. And some further surgical experiment can be done in the virtual scene to provide a precise guidance for doctors. Obviously, many further works are required to improve the performance of the current augmentation technologies.

3.2 Parallel AR Medical Diagnosis Theory

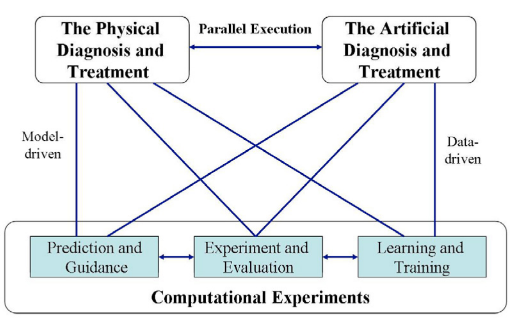

Fig. 4. The Basic framework of the parallel theory applied in AR medical treatment.

AR theory provides an expansion of the real world, which generally visualizes the integration of the virtual and real information. But the complexity and uncertainty are particularly seriously reflected in medical diagnosis. Therefore, a universal theoretical approach is needed to support a highly intelligent AR technique for the medical practices. The parallel intelligence theory, proposed by Wang F-Y (Wang 2004), is considered to be introduced into the AR technology to provide a fundamental solution to these challenges mentioned above. The parallel intelligence theory is aimed to resolve problems in modelling and control of complex systems, which is implemented by building a bridge between the real world and the artificial world. Obviously, the parallel intelligent theory and AR technologies are similar to work for the relationship between the virtual and real information. The most prominent advantage of the parallel intelligence theory is to predict and manage the development of the specific situations in real time applying the continuous feedback between the virtual system and physical system, which can highly improve the intelligent performance of the AR technology. The parallel intelligence theory has shown its effectiveness in many applications, such as in the area of the intelligent transportation, the manufacturing and the social computing. And some researches on parallel medicine (Wang 2013), parallel surgery and parallel gout have been proposed, which provide a feasible demonstration of the medical applications.

Artificial societies, computational experiments, and parallel execution (ACP) are three core parts of the parallel theory (Wang 2004). The basic frameworks of the parallel theory applied in AR medical treatment is shown in Figure 15. The physical surgical scene is stimulated in the artificial surgical system. Not only standard human tissues model and personal parameters but also the knowledge and skilled experience are included. The dynamic state of the physical diagnosis is needed to be accurately expressed, such as the respiratory model, which is the foundation of the whole system. Extensive learning and training are then run in the computational experiments module to predict the coming events. A large scale of diverse artificial data is produced, which further enhances the function of the artificial system. Finally, a real-time navigation is implemented via interaction in parallel execution modules. Based on the huge number of sampled data and the self-learning ability, this system can update the information and guide the surgery in a long term. Especially in some cases that emergencies usually occur in operations, the parallel intelligence medical diagnosis method can quickly provide resolutions for physicians. For example, due to the fast growth sick tissue, the pre-measured model does not quite match with the intra-operative images. Based on the function of intelligent description and prediction in the parallel AR medical theory, many possible cases and their treatment plans have been computed. Therefore, it is easy to provide a much more suitable presentation of the disease. In a short, the parallel AR medical diagnosis theory provides a multi-dimensional augmentation of the physical diagnosis and a deeply intelligent interaction between the physical and artificial diagnosis, which can fully enhance the potential power of the system.

Many approaches developed in different areas, such as artificial intelligence, big data processing, cloud computing, knowledge graph, will be applied in the application of the parallel AR diagnosis theory. But more works are further needed. It is still hard to accurately construct the dynamic model of physical diagnosis. In our body, There are still many unknown fields waiting for us to exploit, which affect the prediction of disease development. For the part of computational experiments , it is still hard to work for special disease with fewer samples. Some effective approaches are needed to be developed. Some non-invasive medical equipment is needed to quickly check the patient in real time, which is very important to give feedback between the artificial and physical diagnosis.