A Direct-View Graphic Fusion Interactive Technology for Surgical Application

Jia-Chong Ho, Yu-Hsuan Su, Yu-Hsiang Tsai, Chia-Hsun Tu, and Kuan-Ting Chen

jcho@itri.org.tw

[1] Electronic and Optoelectronic System Research Laboratories, Industrial Technology Research Institute, Chutung, Hsinchu, Taiwan 310, R.O.C.

Keywords: Image fusion, interactive system, transparent display.

ABSTRACT

A fusion interactive technology for surgical application in that transparent display, facing direction recognition, object recognition, and image fusion algorithms are included. The virtual object images can be displayed on the transparent display and superimposed on the affected areas of patients used in surgical operations to reduce the risk.

1 INTRODUCTION

AR technology, which has two types: video see-through type and optical see-through type, has been widely used in a variety of fields in recent years. The optical see-through technology is also called direct-view graphic (image) fusion, and what makes it different from video see-through type is that a transparent display is used and virtual object images are displayed on transparent display so that the users can look at the object in the actual environment with virtual image superimposed on it [1]. It is applicable for situations such as medical applications that require users to see the real object. In order to achieve virtual and real image information fusion with higher positional accuracy, the accuracy of object recognition and virtual-real fusion algorithm are critical. In this research, AI technology is used to identify and position the user's face and visual fiducial two-dimensional bar code is used for object positioning through the images captured by stereo camera. By acquiring the position and orientation of the face and object in space, the position of the virtual image to be shown on display can be calculated using virtual-real fusion algorithm.

Nowadays, some research institutes and companies are dedicated to importing AR technology into medical services for elevating the safety of medical treatment. Such as, Augmedics Inc. has developed a FDA-approved augmented reality head mounted display (HMD) equipped with various sensors for surgical image guidance using similar concepts [2-3]. However, because wearable AR devices still have some drawbacks, including inconvenience in putting on and taking off, too much weight on the neck, and leaning forward of the head because the center of gravity is moving forward.

In this paper, we proposed a direct-view image fusion interactive technology for surgical image guidance using virtual-real fusion algorithm. With this technology, the surgical operators can look at the actual affected area of the patient with virtual image information superimposed on it without wearing HMD device. We focus on how various functions can be properly integrated, along with its feasibility validation. The rest of the paper is organized as follows: the proposed interactive system, recognition module, and three-dimensional (3D) graphic fusion calculation are described in section 2. Experimental results are reported in section 3. Finally, concluding remarks are given in section 4.

2 THE PROPOSED INTERACTIVE SYSTEM

The proposed interactive system is illustrated in Figure 1. The user can see the real object through the transparent display. In a surgical scenario, prior to operations, the medical team will perform CT or magnetic resonance imaging (MRI) scans of the affected part of the patients to derive digital information of the lesion, which provides the medical team as preoperative, intraoperative, and postoperative reference. Because the files are all 3D images, 3D virtual environment is required in the system to perform 3D virtual and real information superimposition calculation, so that the CT or MRI images displayed on the transparent display can be fused with the actual affected part of the patients behind the transparent displays.

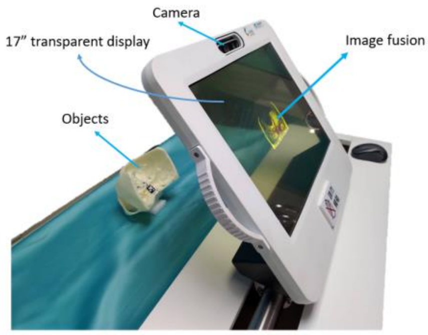

Fig. 1 The proposed interactive system.

2.1 System Structure

Figure 2 illustrates the software and hardware architecture of the system. The hardware comprises of two stereo cameras and a transparent display. The front camera captures the user's image, and the facial landmark detection algorithm recognizes the user's eye landmarks of the captured image, calculating the position and direction of the line of sight of the user. Meanwhile, a two-dimensional fiducial tag is attached to the surface of the target object to be observed as a positional reference, and the rear camera captures the tag image to derive position and orientation information in space. Then the derived position and orientation information is sent to the virtual-real image fusion algorithm to calculate the position and orientation of the virtual image to be displayed on the transparent display. In this way, the user can look at the target object with the fused virtual image superimposed through the transparent display.

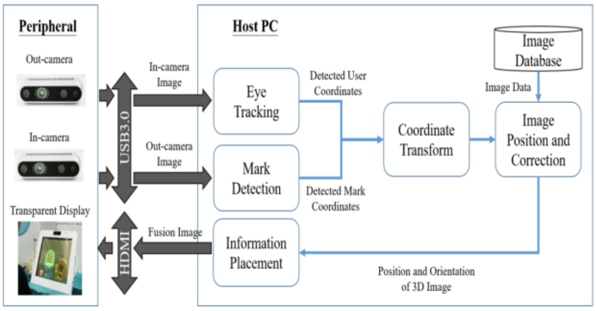

Fig. 2 The proposed interactive system structure.

2.2 Recognition Module

(1) Positioning of User

The gaze tracking module uses Intel RealSense D435i sensor, which can provide high-resolution color images and stable depth measurement, to capture images for facial landmark detection. This module detects the user's facial landmarks, using deep learning to determine the head posture and locate the center of the eyeball and pupil to calculate the gaze vector. In addition, this module does depth measurement via depth sensor to help locate the position. The calculation result is then sent to the host system through shared memory communication for further virtual-real image fusion calculation. The flow chart is shown in Figure 3.

In order to be applicable for surgical scenario, images of face with mask were added to the database for deep learning training to improve the accuracy of eye recognition. The recognition result is shown in Figure 4.

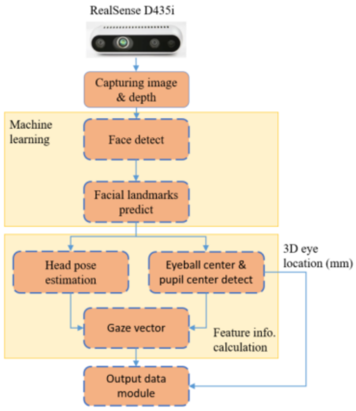

Fig. 3 The flow chart of user recognition module.

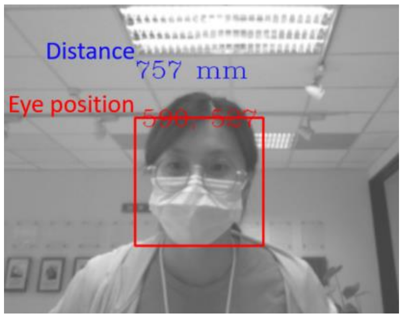

Fig. 4 The user's gaze direction with face mask.

(2) Positioning of Object

In surgical operation room, optical positioning tracking devices are often used to track instruments and patient's spatial positions. The positioning accuracy is required to be less than 2mm. The same specification requirements also works for direct-view image fusion interactive system. AprilTag tag tracking method has features of high accuracy and robustness[4]. In this study, Tag36h11 in AprilTag family was selected for positioning the target object, where 36 is the number of squares in the effective area of the pattern (number of bits), and 11 is the Minimum Hamming Distance which stands for minimum bits of codes required to be modified to become another ID Tag. Figure 5 shows the flow chart of the tag tracking.

Then, we design some experiments to measure the accuracy of the AprilTag positioning. After multiple positioning measurement experiments, the maximum error is 3.08mm, the minimum error is 0.553mm, and the average error is 1.718mm. The average value is less than the standard requirements for surgical applications.

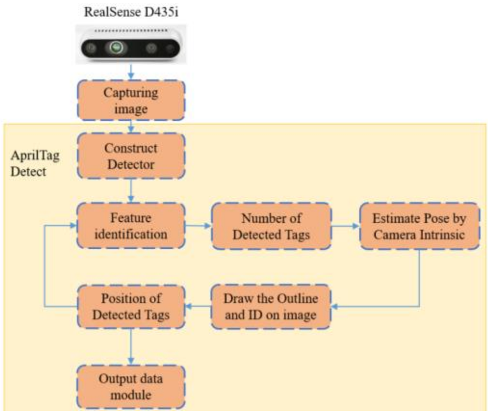

Fig. 5 The flow chart of tag tracking.

2.3 3D Graphic Fusion Calculation

The position and orientation data of the user and target object received by the two aforementioned positioning modules would be used to calculate the position of the virtual image to be presented on the transparent display through the virtual and real fusion algorithm. The algorithm converts the two different spatial coordinates information calculated by these modules into the same spatial coordinate system. The calculated positions of the image are required to be calibrated since there must be deviations in each physical camera position as they were installed.

The calculation can be divided into two cases based on whether user eyes and target object are located on the same sides or opposite sides relative to the normal line through the center of the panel. The schematic diagram of the calculation of the virtual-real fusion image position is illustrated in figure 6. The equations (1) and (2) each corresponds to the abovementioned cases respectively.

$$\begin{cases} x = \dfrac{d_2 \tan\theta_2 * d_1 + d_1 \tan\theta_1 * d_2}{d_1 + d_2} \\ \theta_r = \tan^{-1}\left|\dfrac{|d_2 \tan\theta_2 - x|}{d_2}\right| \end{cases} \tag{1}$$

$$\begin{cases} x = \dfrac{d_1 d_2 (\tan\theta_1 - \tan\theta_2)}{d_1 + d_2} \\ \theta_r = \tan^{-1}\left|\dfrac{d_1 d_2 \tan\theta_1 + {d_2}^2 \tan\theta_2}{d_1 d_2 + {d_2}^2}\right| \end{cases} \tag{2}$$

However, since there would be positional deviations during installation of both front and rear cameras in the mechanism. Therefore, it is necessary to calibrate the coordinates of the front and rear cameras. The calibration formula is shown in equation 3.

$$\begin{bmatrix} x_2 \\ y_2 \\ z_2 \\ 1 \end{bmatrix} = \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ y_1 \\ z_1 \\ 1 \end{bmatrix} \tag{3}$$

where (x_1, y_1, z_1) stands for the coordinates of the front camera, (x_2, y_2, z_2) stands for the coordinates of the rear camera, r_{xx} are the parameters of the rotation matrix, and t are the parameters of the displacement matrix.

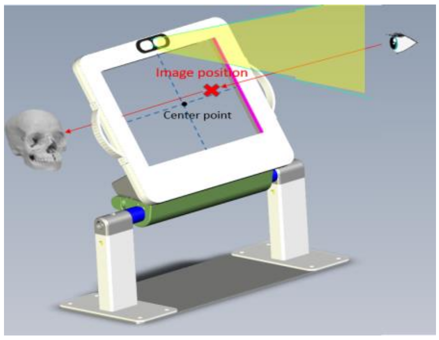

Fig. 6 Schematic diagram of the calculation of the virtual-real fusion image position in which the eyes and object are on the same sides or on the opposite sides.

3 Experiments

(1) Accuracy test

Accuracy test was executed to verify whether the accuracy of the virtual-real image fusion of this system can meet the needs of surgery and medical treatment or not. A sphere of 15cm diameter was used as a standard target object in the test , and a AprilTag was attached to the surface of the sphere as a fiducial point. Then a camera is placed at user's position to simulate user's eyes. The virtual and real target object images were simultaneously captured from the user's perspective by the camera, and the accuracy of the captured images was derived by comparing the visual and real object images. As can be seen in the figure 7, the error at the top, bottom, left and right edges of the measurement are less than 2mm.

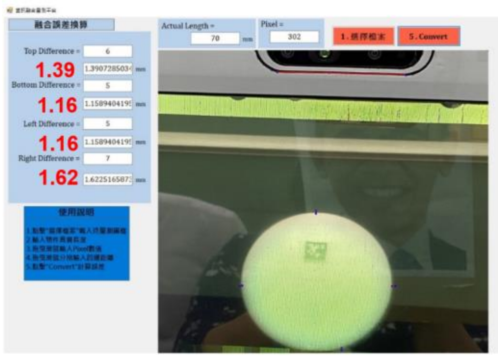

Fig. 7 Accuracy Test.

(2) Multi-perspective information fusion

In actual surgical situation, surgeons need to confirm their cutting positions when they perform surgical operations, thus the system provides the function of rotating the 3D virtual object in multiple angles based on the viewing angle of the operators, and fused the virtual images with the actual objects. Figure 8 shows the images of multi-perspective information fusion between $\pm 60°$.

Fig. 8 Multi-perspective information fusion.

4 CONCLUSIONS

We have developed a non-wearable virtual-real information fusion transparent display technology for surgical applications. Deep learning of facial landmark detection of wearing mask is implemented to track the position of surgeon’s eyes. The system uses high-precision tag positioning technology to accurately position the target object, and then uses the virtual-real information fusion algorithm to calculate the position and orientation angle of the 3D virtual object on the panel. The accuracy is less than 2mm.

At present, this system still has some issues to be overcome, such as excessive positional deviation of the virtual and real superimposed images at large viewing angles, parallax of users, serious panel reflection at high brightness environment, etc. Measures will be taken to improve these issues in the future.

REFERENCES

[1] K. K. Chen, Y. H. Huang, Y. H. Tsai, W. C. Chen, L. H. Wang, H. Y. Chen, G. Chen, J. C. Ho, and C. C. Lee, “Highly Transparent AMOLED Display with Interactive System”, Proc. SID Digest, Vol. 50, Issue 1, pp. 842-845 (2019).

[2] CA. Molina, N. Theodore, A K. Ahmed, E. M Westbroek, Y. Mirovsky, R. Harel, E. Orru', M. Khan, T. Witham, D. M Sciubba, “Augmented reality-assisted pedicle screw insertion: a cadaveric proof-of-concept study,” J. Neurosurg Spine. Vol. 31, Issue 1, pp. 1-8 (2019).

[3] G. Badiali, L. Cercenelli, S. Battaglia, E. Marcelli, C. Marchetti, V. Ferrari, and F. Cutolo, “Review on Augmented Reality in Oral and Cranio-Maxillofacial Surgery: Toward “Surgery-Specific” Head-Up Displays,” in IEEE Access, vol. 8, pp. 59015-59028 (2020).

[4] J. Kallwies, B. Forkel, and H. J. Wuensche, “Determining and Improving the Localization Accuracy of AprilTag Detection,” in Proceedings of the IEEE International Conference on Robotics and Automation (ICRA), pp. 8288-8294 (2020).