



HCI Based on Gesture Recognition in an Augmented Reality System for Diagnosis Planning and Training

Qiming Li, Chen Huang, Zeyu Li, Yimin Chen, Lizhuang Ma

► To cite this version:

Qiming Li, Chen Huang, Zeyu Li, Yimin Chen, Lizhuang Ma. HCI Based on Gesture Recognition in an Augmented Reality System for Diagnosis Planning and Training. 2nd International Conference on Intelligence Science (ICIS), Oct 2017, Shanghai, China. pp.113-123, 10.1007/978-3-319-68121-4_12 . hal-01820905

HAL Id: hal-01820905

<https://inria.hal.science/hal-01820905>

Submitted on 22 Jun 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

HCI based on Gesture Recognition in an Augmented Reality System for Diagnosis Planning and Training

Qiming LI^{1,2*}, Chen HUANG³, Zeyu LI^{3,4}, Yimin CHEN³, Lizhuang MA¹

¹ Department of Computer Science and Engineering, Shanghai Jiaotong University, Shanghai, 200240

² College of Information Engineering, Shanghai Maritime University, Shanghai, 201306

³ Department of Computer Science and Technology, Shanghai University, Shanghai, 200444

⁴ Computer Center, Ruijin Hospital, Shanghai Jiaotong University School of Medicine, Shanghai, 200025

{ Qiming LI, qmli@shmtu.edu.cn }

Abstract. An Augmented Reality System for Coronary Artery Diagnosis Planning and Training (ARS-CADPT) is designed and realized in this paper. As the characteristic of ARS-CADPT, the algorithms of static gesture recognition and dynamic gesture spotting and recognition are presented to realize the real-time and friendly Human-Computer Interaction (HCI). The experimental results show that, with the use of ARS-CADPT, the HCI is natural and fluent, which improves the user's immersion and improves the diagnosis and training effects.

Keywords: gesture recognition, augmented reality, human computer interaction

1 Introduction

Presently, 64 multi-slices computed tomographic coronary angiography technology has been considered as an effective way to diagnose coronary heart disease [1]. In the preoperative diagnosis planning process, the doctors are not accustomed to carry on the interactive diagnosis with computer by using the mouse and keyboard. 3D reconstruction based on Computed Tomography (CT) image sequence combined with augmented reality (AR) technology can effectively solve the above problems.

AR is a new technology that strengthens the user's perception of the real world by superimposing the virtual 3D information generated by the computer system onto the real scene. In fact, medicine is one of the earliest application fields of AR technology. State Andrei et al. [2] can draw a virtual 3D fetus on its abdomen position by ultrasonic scanning a pregnant woman. The doctor can understand the move and kick ability of the fetus through the Helmet-Mounted Displays (HMD) in 1994. AR technology can be used as an auxiliary means of surgical visualization. The 3D data of patient can be collected through Magnetic Resonance Imaging (MRI), CT or ultrasound images. According to the data, the corresponding virtual information can be rendered in

real time. Combining with the actual situation of patient, the doctors can get more complete information, and improve the operation finally [3]. Wu J. R. [4] implements a spine surgery AR system, in which the surgeons can make use of 3D virtual model of preoperative patients to carry out spinal surgery simulation practice. In minimally invasive surgery, AR technology enables doctors to obtain the clairvoyant ability and improve the quality of surgery [5]. AR can also be used for medical training. According to statistics, over 50% of the augmented virtual reality application system are used in medical training, the most of which are realized based on virtual reality (VR) technology [6]. The amount of application system based on AR is relatively less. The AR based aid medical training system [7] is used to achieve medical training and examination through human body modeling.

The natural and real-time HCI is one of the three important features of AR system [8]. However, the traditional interaction mode such as using the mouse and keyboard cannot meet the application requirements. People are eager to realize the HCI in a very natural way. Gesture is just the most natural and intuitive way of interaction in human communication except language. Therefore, HCI based on gesture recognition has become a hot research topic. Gestures are usually defined as hand shapes and movements produced by the combination of palms, fingers, and even arms. The task of HCI based on gesture recognition is: firstly, recognize the meaning of the gesture correctly according to the data captured in real time, then trigger the corresponding instruction, and make the system feedback finally.

An AR system used for coronary artery diagnosis planning and training is designed and realized in this paper, which is called ARS-CADPT in the following paragraph. The system is very complicated, but the HCI based on gesture recognition is mainly discussed in this paper. The operating user or the lecturer can interact with the 3D model of the coronary arteries in a natural and intuitive manner with the defined gestures, and can perform simulation measurement of radius of vessels, and thus achieve a comprehensive and intuitive presentation and an accurate and detailed explanation of the patient's situation. The interns or students can understand and study the patient's coronary detail situation on a large tiled screen.

2 System Architecture

2.1 The Hardware Architecture of ARS-CADPT

As shown in Fig.1, ARS-CADPT is constructed based on cluster architecture. It consists of several high-performance workstations, a parallel rendering and tiled display subsystem and a series of equipment for interaction data capture. Server 1 is used for 3D reconstruction of coronary artery based on CT images and storage the 3D coronary artery model database of all the previous patients. Server 2 is the surveillance and control center of the system. Server 3 is used for processing the lecturer's interaction data which is captured by the equipment such as Leap Motion, magnetic tracker, Microsoft Hololens, and so on. The display subsystem is consists of 5 parallel rendering nodes and a tiled screen, which is used for study, view and emulate for the student and intern users.

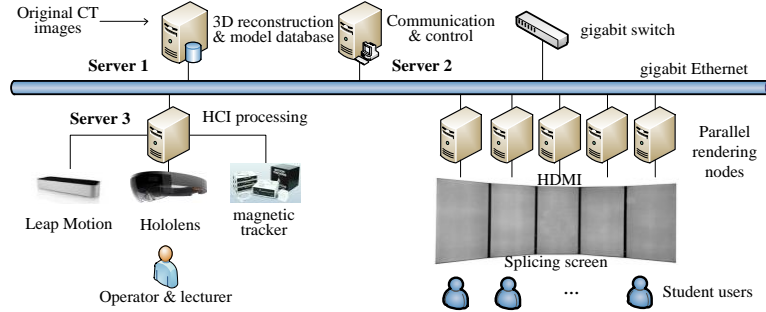


Fig. 1. The hardware of ARS-CADPT

2.2 The Software Framework of ARS-CADPT

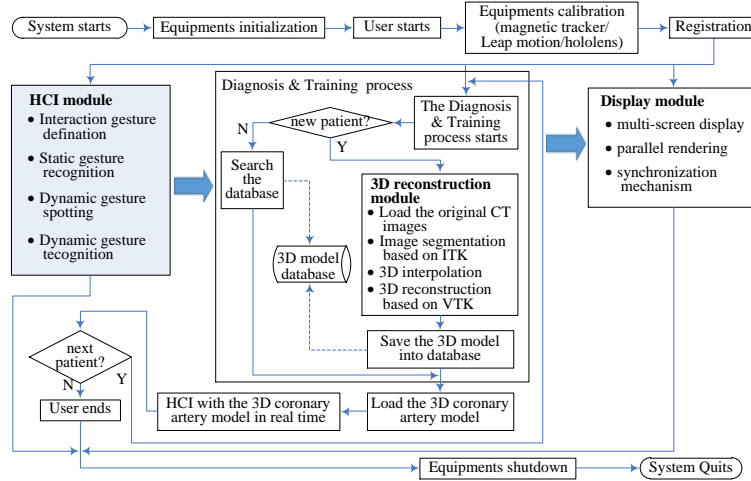


Fig. 2. The workflow and functions of the ARS-CADPT

The ARS-CADPT is realized based on multi-thread technique. The main thread is used for the diagnosis and training process, the HCI thread is used for the real-time interaction with the 3D coronary artery model, and the feedback is displayed to the users via the display thread. Therefore, the system mainly includes three function modules. 3D coronary artery reconstruction based on CT images, real-time HCI based on gesture recognition and synchronous display based on parallel rendering. Hereinto, the HCI module is the characteristic of the system. The interaction gestures used in the system are defined firstly. Then the algorithms for static gesture recognition, dynamic gesture spotting and recognition are proposed. The corresponding interaction operations are triggered according to the gesture recognition results finally.

3 Real-time HCI

As shown in Fig.2, the HCI module serves the main process of diagnosis and training. It is the bridge between the operating user and the system. The HCI in the system is accomplished based on the coordination of static and dynamic gesture recognition. The Leap Motion manufactured by Leap Company is used to capture the hand shape and motion trajectory.

3.1 Gesture Definition

Gesture includes static gesture and dynamic gesture. Considering a right-hander, the static gestures used in ARS-CADPT are defined in Table 1 and illustrated in Fig.3.

Table 1. The static gestures and the corresponding HCI functions in ARS-CADPT

Legend	Class		HCI function
Fig.3 a	1	System input	Left mouse button down
Fig.3 b	2		right mouse button down
Fig.3 c	3		capture the hand trajectory as the trajectory of cursor
Fig.3 d	4	model control	Control 3D model in full synchronous mode according to the hand trajectory captured in real time
Fig.3 e	5		Control 3D model in fine adjustment mode according to the recognition result of the real-time gesture trajectory
Fig.3 f	6		Radius measurement of vessel stenosis
Fig.3 g	7		In fine adjustment mode, control 3D model according to scale I
Fig.3 h	8		In fine adjustment mode, control 3D model according to scale II
Fig.3 i	9		In fine adjustment mode, control 3D model according to scale III
Fig.3 j	10		In fine adjustment mode, control 3D model according to scale IV



Fig. 3. The static gesture used in ARS-CADPT

The gestures can be divided into two categories: system input and model control. The former is used to achieve all kinds of system input instructions by gestures instead of mouse and keyboard, and the latter is used to manipulate and control the 3D models directly. Hereinto, the static gestures of class 1~3 are used for system input, and the static gestures of class 4~10 are used for model control.

In addition, the accurate measurement of vessel diameter, confirming the location and extent of stenosis and the situation of collateral circulation formation are objective gist for determining the diameter of stent during interventional treatment. However, in clinical practice, the measurement of vessel stenosis is mainly based on visual estimation, and its accuracy needs to be further improved. In ARS-CADPT, the oper-

ating user can interactively scale the virtual model freely and measure the radius of the blood vessel with the gesture shown in Fig.3f.

There are two ways of interacting with the 3D model in ARS-CADPT. The full synchronization mode is to make the movement of 3D model completely consistent with the hand of which the static gesture is class 4 (shown in Fig.3d), while the fine adjustment interaction mode is used for the slight, accurate and complex operations. The fine adjustment interaction mode is mainly realized by recognizing a series of dynamic gestures formed by the hand trajectory of which the static gesture is class 5 (shown in Fig.3e). Therefore, the dynamic gestures needed in ARS-CADPT are defined in Table 2, and some examples are illustrated in Fig.4. Hereinto, the translation gestures can be used for both system input and model control. The rotation and zoom gestures are used for model control only.

Table 2. The dynamic gestures and the corresponding HCI functions in ARS-CADPT

Gesture	Class		HCI function
Translation	Left	11	Next menu record item page etc. Move the 3D model along the X axis in the positive direction
	Right	12	Previous menu record item page etc. Move the 3D model along the X axis in the negative direction
	Up	13	Next menu record item page etc. Move the 3D model along the Y axis in the positive direction
	Down	14	Previous menu record item page etc. Move the 3D model along the Y axis in the negative direction
	Forward	15	Previous menu record item page etc. Move the 3D model along the Z axis in the negative direction
	Backward	16	Next menu record item page etc. Move the 3D model along the Z axis in the positive direction
Rotation	Left	17	Rotate the 3D model clockwise in the top view
	Right	18	Rotate the 3D model anti-clockwise in the top view
	Up	19	Rotate the 3D model anti-clockwise in the elevation view
	Down	20	Rotate the 3D model clockwise in the elevation view
	Forward	21	Rotate the 3D model clockwise in the left view
Zoom	Backward	22	Rotate the 3D model anti-clockwise in the left view
	in	23	Enlarge the 3D model proportionally along three axes
	out	24	Shrink the 3D model proportionally along three axes

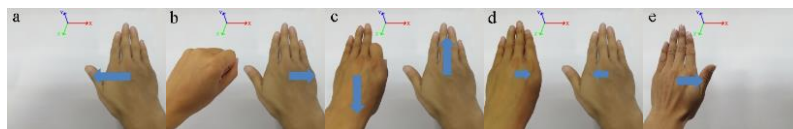


Fig. 4. Some examples of the dynamic gestures defined in ARS-CADPT: a. left translation (at default scale), b. right translation at scale I, c. right rotation at scale II, d. zoom out at scale IV, e. right translation by left hand (at default scale)

3.2 Static Gesture Recognition

The static gesture recognition algorithm based on rough sets theory was proposed. The static gesture recognition is considered as a decision table, denoted as $DT = (U, C \cup D, V, f)$. Hereinto, U is a nonempty finite set of all the static gesture instances, called universe. C and D are also nonempty finite sets, C is called condition-attribute set, and D is called decision-attribute set. $V = \bigcup_{a \in C \cup D} V_a$, V_a is the range of attribute a . $f: U \times A \rightarrow V$ is called the information function, which assigns a value to each attribute. The data of static gestures are mainly captured by Leap Motion. The distance between the fingertips and the distance between the fingertips and the palms center are considered as the major factors influencing the static gestures, which are belong to C . The distance is discretized into five values, so $V_C = \{1, 2, 3, 4, 5\}$. There is only one decision attribute: the static gesture (denoted as d), i.e. $D = \{d\}$. According to Table1, there is $V_d = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11(\text{undefined})\}$. To sum up, the decision table of static gesture recognition can be modeled as shown in Table 3.

Table 3. decision table of static gesture recognition

U	C								D
	1	2	...	10	i	ii	...	v	d
x_1	$v_{1,1}$	$v_{1,2}$...	$v_{1,10}$	$v_{1,i}$	$v_{1,ii}$...	$v_{1,v}$	$v_{1,d}$
x_2	$v_{2,1}$	$v_{2,2}$...	$v_{2,10}$	$v_{2,i}$	$v_{2,ii}$...	$v_{2,v}$	$v_{2,d}$
...
x_n	$v_{n,1}$	$v_{n,2}$...	$v_{n,10}$	$v_{n,i}$	$v_{n,ii}$...	$v_{n,v}$	$v_{n,d}$

Hereinto, x_j ($j=1, 2, \dots, n$) is the j -th static instance, $U = \{x_1, x_2, \dots, x_n\}$ is the set of static instance, and $v_{j,a}$ is the value of attribute a ($a \in C \cup D$) in the j -th static instance.

The decision table is constructed according to the selected sample set. Then, the attribute reduction algorithm based on Skowron discernibility matrix and discernibility function is adopted: Firstly, construct the discernibility matrix; Secondly, construct the discernibility function; Thirdly, simplify the discernibility function using the absorption law; Finally, the conjunctive normal forms in the minimal disjunctive normal form of the discernibility function are all the D-reduct of C .

The classical reduction algorithm of attribute values is based on the value core concept. At first, calculate the value core of every instance in the decision table after attribute reduction; then get the minimal reduct from the value core table; finally, obtain the decision rules.

At last, the rules can be used to recognize the user's static gesture in real time.

3.3 Dynamic Gesture Spotting

Pavlovic V. I. et al. [9] divide the movements of the hand into two categories. One is the gesture that conveys the user's intention, and the other is meaningless action. Therefore, the starting point and termination point of each dynamic gesture must be located in the acquired continuous gesture data stream. It is the premise and foundation of dynamic gesture recognition. However, the existing dynamic gesture recogni-

tion methods usually assume either known spatial spotting or known temporal spotting, or both [10], which is unrealistic in the practical applications.

According to the data captured by Leap Motion, a segment of right hand motion trajectory is drawn in Fig.5a. It can be seen the intervals of points are different. That means the speed is changing during the gestures. The curve shown in Fig.5b is the speed variation during the gestures in Fig.5a. It is clearly illustrated that the speed climbs up and then declines for several times. Each speed jump corresponds to a wave crest on the speed curve. There are five obvious wave crests which exactly correspond to five gestures. So, a simple method is to set a threshold. If speed is above the threshold, a gesture is detected. But this method would arouse some problems. One is that some noise points exist. Another is that the speed of dynamic gestures varies from person to person, and setting a threshold is not-so-flexible. In fact, the dynamic gestures defined in our system are all completed in a speed jump. Therefore, we could think that a dynamic gesture is generated only by judging an upward tendency of speed. Thirty data points are enough to represent the tendency from the experiment. We define the upward tendency, which is the speed at any moment is higher than before. The fuzzy set of the standard upward tendency is denoted as \underline{A} , and a new fuzzy set \underline{B} is obtained every time, and compute the close degree between \underline{A} and \underline{B} using Hamming close degree. The equation is:

$$N_H(\underline{A}, \underline{B}) = 1 - d(\underline{A}, \underline{B}) / n = 1 - \delta(\underline{A}, \underline{B}). \quad (1)$$

Where
$$d(\underline{A}, \underline{B}) = \sum_{i=1}^n |\mu_{\underline{A}}(x_i) - \mu_{\underline{B}}(x_i)|, \quad \mu(x_i) = \begin{cases} 1 & y_i > \text{Max} \\ 0 & y_i \leq \text{Max} \end{cases}, \quad \text{Max} = \begin{cases} y_i & y_i > \text{Max or } i=1 \\ \text{Max} & y_i \leq \text{Max} \end{cases}.$$

The lower the close degree is, the greater likelihood it is an upward tendency. If the close degree is less than 0.3, we believe the set \underline{B} represents the upward tendency.

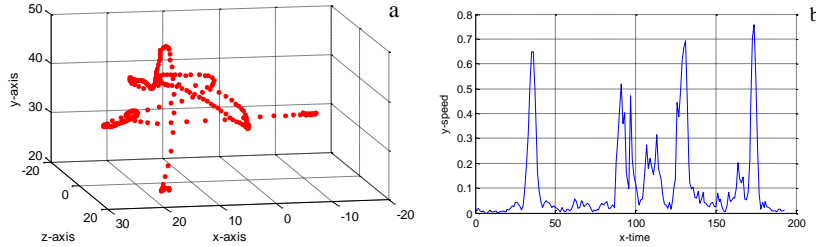


Fig. 5. A segment of dynamic gesture: a. Trajectory, b. Speed curve

3.4 Dynamic Gesture Recognition

After the gesture spotting, a series of independent dynamic gesture trajectories can be obtained in real time.

1. Single-hand gesture

Let $C = \{c_1, c_2, \dots, c_n\}$ be the set composed of the n classes dynamic gestures, and $A = \{a_1, a_2, \dots, a_m\}$ be the set composed of the m attributes influencing dynamic ges-

tures. s_k is the k -th gesture sample to be recognized. After s_k was recognized by all the attribute classifiers, a decision matrix is obtained and denoted as:

$$DM(s_k) = ((f_{11}^k, f_{21}^k, \dots, f_{m1}^k)^T, (f_{12}^k, f_{22}^k, \dots, f_{m2}^k)^T, \dots, (f_{1n}^k, f_{2n}^k, \dots, f_{mn}^k)^T). \quad (2)$$

Hereinto, the row vector $f_i = (f_{i1}^k, f_{i2}^k, \dots, f_{in}^k) (i=1, 2, \dots, m)$ is the recognition results of s_k by attribute classifier a_i with respect to the m classes, while the column vector $f_j = (f_{1j}^k, f_{2j}^k, \dots, f_{mj}^k)^T (j=1, 2, \dots, n)$ is the recognition results of s_k by all the attribute classifiers with respect to the c_j -th class. then the intersection of the function $f_j: A \rightarrow [0, 1]$ (if the output of classifier is not in the interval $[0, 1]$, it can be satisfied according to normalization) and the vector f_i, f_{ij}^k , indicates the degree of certainty that s_k is recognized as the c_j -th class by the attribute classifier a_i .

Let g be the fuzzy measure defined over $P(A)$, the power set of A . The fuzzy measure on the single-point set, i.e. fuzzy density $g_i = g(\{a_i\}) (i=1, 2, \dots, m)$ represents the degree of credibility that the attribute classifier a_i makes decision. If $X \in P(A)$, $g(X)$ represents the degree of credibility that the attribute classifier X makes decision. However, as the single attribute classifiers are designed for a certain attribute feature of dynamic gestures, they should have different degrees of credibility for different gestures, namely, the single attribute classifiers should have different fuzzy densities. Let $g_j = (g_{1j}, g_{2j}, \dots, g_{ij}, \dots, g_{mj})$ be the fuzzy density vector of class c_j , where g_{ij} represents the degree of credibility of the attribute classifier a_i with respect to class c_j . Then the fuzzy integral over A of the function f_j^k with respect to the fuzzy measure g_j is the overall objective estimate for s_k belonging to class c_j . In the way, for a certain gesture sample s_k , the system gives an integral value for every class, and the class that the greatest integral value corresponds with will be adopted as the recognition result.

2. Two-handed gesture

As for the two-handed dynamic gestures, the positions of the palm center at the beginning and end of the gesture are captured. Let $B^l = (b_x^l, b_y^l, b_z^l)$, $B^r = (b_x^r, b_y^r, b_z^r)$, $E^l = (e_x^l, e_y^l, e_z^l)$ and $E^r = (e_x^r, e_y^r, e_z^r)$ be the coordinates of both hands at the beginning and end of the gesture respectively, then the lengths of line segment $B^l B^r$ and line segment $E^l E^r$ (denoted as d_b and d_e respectively) and the angle between them (denoted as φ) can be calculated. By projecting φ to plane YOZ, XOZ and XOY respectively, the three direction angles (denoted as α, β and γ) can be calculated. At last, the current gesture can be recognized by the following rules:

IF $\varphi < \theta_1$ AND $(d_e - d_b) > \sigma$ ($< \sigma$), THEN $class(g) = \text{zoom in (out)}$

IF $\varphi > \theta_2$ AND $\max(\alpha, \beta, \gamma) = \alpha \mid \beta \mid \gamma$, THEN $class(g) = \text{Rotate around the X/Y/Z axis}$

Where θ_1, θ_2 and σ are thresholds predefined.

4 Experimental Results

The related experiments are carried out based on the system platform. Hereinto, the recognition rate for static gesture achieves an average performance of 97.3%. After the dynamic gesture spotting, the continuous dynamic gesture stream is divided into a set of isolated dynamic gestures. The system achieves an average performance of 92.4% for the dynamic gestures according to the algorithm proposed in section 3.4.

The system is used for coronary artery diagnosis planning and teaching. As shown in fig.6, the lecturer is controlling the 3D coronary artery model in a natural and intuitive manner with the defined gestures, and thus achieves a comprehensive and intuitive presentation and an accurate and detailed explanation of the patient's situation. At the same time, the student users or interns can study and understand the patient's coronary detail situation on a large tiled screen.



Fig. 6. System running instance: the lecturer is interacting with the 3D coronary artery model.

5 Conclusion and the Future Work

This paper presented an augmented reality system for coronary artery diagnosis planning and training. One of its advantages is to realize the real-time and friendly HCI by using the algorithms of static gesture recognition and dynamic gesture spotting and recognition. It can be concluded that the proposed solutions make the HCI more natural and convenient, make the explanation clearer and more intuitive, and finally achieve a better effect for the preoperative diagnosis planning and training.

It also can be concluded that the AR technology has great potential to apply to the computer-aided medical system. Some examples of AR-based surgical applications have been presented in the literatures [11, 12]. Meanwhile, there are still some technical challenges for further research and exploration. For example, the gesture set defined in ARS-CADPT is just a little subset of the human gesture set, and the gestures people used in daily life are much more complicated. This puts forward higher requirements to the gesture recognition algorithms. Moreover, there is still a lot of work to do before ARS-CADPT can be applied to real-time surgery.

Acknowledgement

This work is supported by Natural Science Foundation of China (Grant No.: 61472245) and Shanghai Municipal Natural Science Foundation (Grant No.: 14ZR1419700 and 13ZR1455600).

References

1. Miller J M, Rochitte C E, Dewey M, et al. Diagnostic Performance of Coronary Angiography by 64-row CT. *New England Journal of Medicine*, 359(359):2324-2336. (2008)
2. Andrei State, David T. Chen, Chris Tector, et al. Case Study: Observing a Volume Rendered Fetus within a Pregnant Patient. *IEEE Conference on Visualization*, pages: 364-368. (1994)
3. Tang S L, Kwok C K, Teo M Y, et al. Augmented reality systems for medical applications: Improving surgical procedures by enhancing the surgeon's 'view' of the patient. *IEEE Engineering in Medicine and Biology Magazine*, 17(3):49-58. (1998)
4. Wu J R, Wang M L, Liu K C, et al. Real-time advanced spinal surgery via visible patient model and augmented reality system. *Computer Methods & Programs in Biomedicine*, 113(3):869-881. (2014)
5. De Paolis L T, Aloisio G. Augmented reality in minimally invasive surgery. *Advances in Biomedical Sensing, Measurements, Instrumentation and Systems*, 55 LNEE:305-320 (2010)
6. Ivelina V. Alexandrova, Marcus Rall, Martin Breidt, et al. Animations of medical training scenarios in immersive virtual environments. *Workshop on Digital Media and Digital Content Management*, pages: 9-12. (2011)
7. Ana C M T G Oliveira, R Tori, W Brito, et al. Realistic simulation of deformation for medical training applications. *15th Symposium on Virtual and Augmented Reality*, pages: 272-275. (2013)
8. Azuma R T. A survey of augmented reality. *Presence-teleoperators and Virtual Environments*, 6(4):355-385. (1997)
9. Pavlovic V I, Sharma R, Huang T S. Visual Interpretation of Hand Gestures for Human-Computer Interaction: A Review. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7): 677-695. (1997)
10. Alon Jonathan, Athitsos Vassilis, Yuan Quan, et al. A unified framework for gesture recognition and spatiotemporal gesture segmentation, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(9):1685-1699. (2009)
11. Xiaojun Chen, Lu Xua, Yiping Wang, et al. Development of a surgical navigation system based on augmented reality using an optical see-through head-mounted display. *Journal of Biomedical Informatics*, 55 (C):124-131. (2015)
12. Xiaojun Chen, Lu Xu, Huixiang Wang, et al. Development of a surgical navigation system based on 3D Slicer for intraoperative implant placement surgery, *Medical Engineering & Physics*, 41:81-89. (2017)