Kinect-Based Gesture Recognition for Touchless Visualization of Medical Images

Jiaqing Liu[1*], Ryoma Fujii[1], Tomoko Tateyama[2], Yutaro Iwamoto[1], Yenwei Chen[1]

[1] Graduate School of Information Science and Engineering, Ritsumeikan University, Shiga, Japan.
[2] Department of Computer Science, Hiroshima Institute of Technology, Hiroshima Japan.

* Corresponding author. Tel.: 077-561-3003; email: gr0302kv@ed.ritsumei.ac.jp
Manuscript submitted July 10, 2017; accepted August 12, 2017.
doi:10.17706/ijcee.2017.9.2.421-429

Abstract: This paper proposes a novel touchless visualization system for computer aided surgery (CAS), which can control and manipulate the patient's 3D anatomy model without contact through the use of Kinect-based gesture recognition technology. Real-time visualization is important in surgery, particularly during the operation. But traditional input devices are reliant on physical contact, which are ill-suited for non-sterile conditions. The depth and skeleton information from Kinect are effectively utilized to produce markerless hand extraction. Based on this representation, histogram of oriented gradients (HOG) features and principal component analysis (PCA), are used to recognize hand gestures. We developed a new system, which can visualize 3D medical image with L form screen and 9 kinds of simple touchless single-handed interactions. Experiments show that the proposed system is able to achieve high accuracy.

Key words: Computer aided surgery(CAS), kinect, touchless visualization, medical image.

1. Introduction

With the development of medical imaging technologies, such as MR and CT, 3D medical images with high-resolution are become possible for assistance of diagnostics and surgery. Real-time imaging review is important in surgery, particularly during the operation. In traditional ways, however, a surgeon usually needs to use some physical contact devices such as mouse, keyboard or touch panel, which are ill-suited for non-sterile condition. So a touchless visualization system is helpful for supporting surgery. In 2014, Microsoft released a type of low-cost RGB-D camera, called Kinect. The Kinect brings a new generation of motion tracking with far greater accuracy and shorter response time which is considered as an ideal solution for touchless interactions. Several touchless interaction systems have been proposed for visualization of medical images in surgery operation room [1]-[3]. However, these systems still have some limitations: need two hands for interaction [1] or can only visualize 2D medical images [2]. In this paper, we developed a touchless visualization system using Kinect for assisting hepatic surgery, which can visualize 3D medical images with single-hand interactions. A preliminary version of this work was presented earlier in an international conference (InMed2016) [4]. The present work is an improved version of our previous work. In our new version, a robust and accurate Kinect-based gesture recognition system has been realized. We also performed a user study questionnaire to evaluate our system quantitatively.

2. Overview of Our Proposed System

Our proposed system is composed of two modules: visualization module and interaction module. In the

visualization module, we visualized the patient's anatomic models (3D surface rendering) such as the liver and its vessel structure based on Omega Space and Visualization ToolKit (VTK ver.7.1.0, from Kitware Inc.)[5]. In the interaction module, we designed 9 gestures to control the visualization mode. We apply a Kinect to obtain depth images of user's hand and recognize the hand gesture to control the visualization modes.

There are several kinds of modes including rotation, adjustment of opacity and switch between visible and invisible for three kinds of vessels. Fig. 1 summarized the major steps in the proposed system.

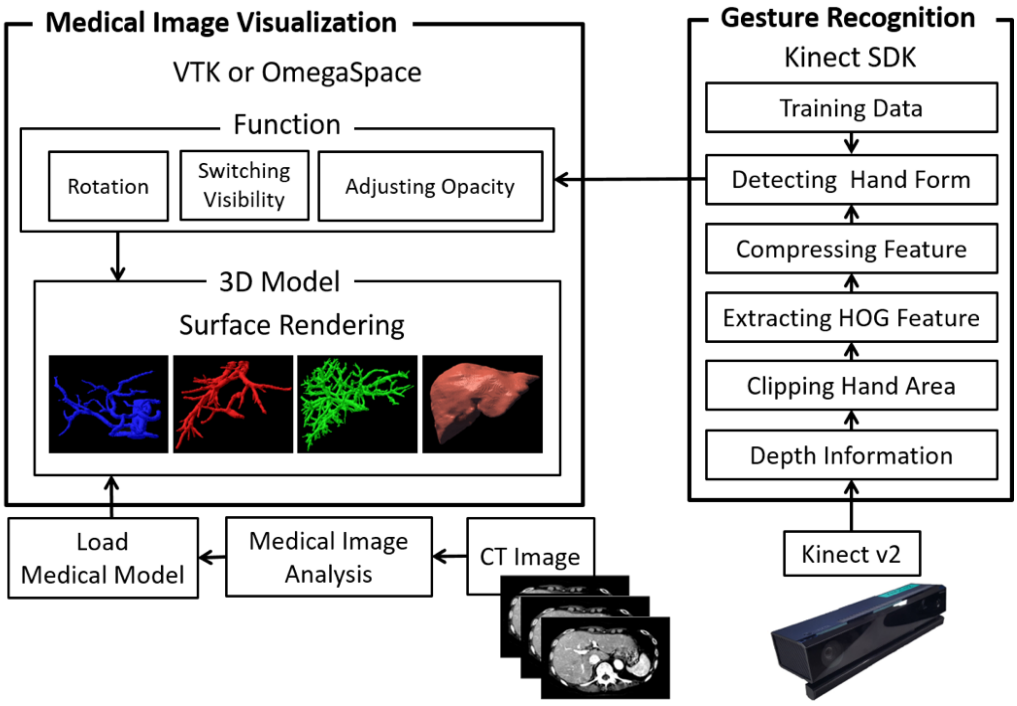


Fig. 1. Diagram of our proposed system.

3. Hand Gesture Recognition

In this section, we present our methods for the hand gesture recognition problem. We first briefly introduce hand depth image generation with Kinect and preprocessing approaches. Then we describe our general approach which uses HOG [6] for feature extraction and dimension reduction by PCA [7]. Finally, we show recognition results by a nonlinear SVM classifier.

3.1. Hand Depth Image Generation and Preprocessing

We utilize the depth information and skeleton tracking provided by the Kinect to generate the depth image of hand. First, we acquire a depth image of the user (Fig. 2(a)). Then, we do calibration between color and depth camera and using the right hand joint point as the center, chip out a 100×100 cm square region as a ROI of hand region (Fig. 2(b)). The depth image with a range from d-30 cm to d +5 cm is defined as a hand image, where d is the depth of the right hand joint point. The segmented hand image is shown in Fig. 2(c). Since the hand image has other regions' pixels with remained as noise, we apply an opening operator and a median filter to remove the noise (Fig. 2(d)). In practical applications, the extracted hand shapes usually have different depth values due to various distances from the camera to hand. We normalized the segmented depth image to a range of [0, 35] by $f'(x,y)= f(x,y)-d+30$, where $f(x,y)$ and $f'(x,y)$ are the depth values at pixel (x,y) before and after normalization, respectively.

(a) (b) (c) (d)

Fig. 2. The depth hand image for gesture recognition (the rad point is the right hand joint point detected by Kinect). (a) Depth image from Kinect. (b) Decision ROI of depth hand image. (c) Segmented depth hand image including noise. (d) Depth hand image excluding noise.

3.2. HOG for Feature Extraction

The histogram of oriented gradients (HOG) [6] are extracted from the acquired depth hand image. These features are used for gesture recognition. At first, we divide the hand image into small regions called "cells". The size of the cell is 10×10 (Fig. 3. (a)). The edge gradients and orientations are calculated at each cell's pixel. The gradient magnitude $m(x,y)$ is calculated using Eq. (1), where $f_x(x,y)$ and $f_y(x,y)$ are two gradient images in x and y, respectively, calculated by Eq. (3). The orientation $\theta(x,y)$ is calculated using Eq. (2). The orientations are quantified to nine bins. Each cell's histogram ($v(n)$, n=0, 1, 2, ..., 8) ranges from 0 to 180 degrees and each bin has 20 degrees. For each pixel, it's contribution (weight) to the histogram ($v(n)$) is given by the gradient magnitude.

$$m(x,y) = \sqrt{f_x(x,y)^2 + f_y(x,y)^2} \tag{1}$$

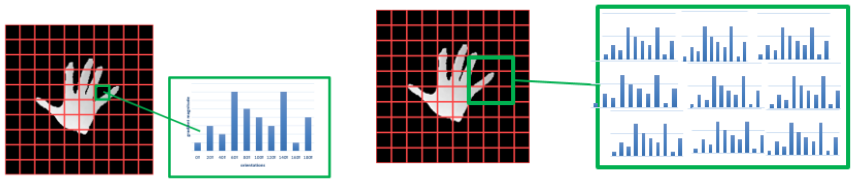
$$\theta(x,y) = \tan^{-1}\frac{f_y(x,y)}{f_x(x,y)} \tag{2}$$

$$\begin{aligned} f_x(x,y) &= I(x+1,y) - I(x-1,y) \\ f_y(x,y) &= I(x,y+1) - I(x,y-1) \end{aligned} \tag{3}$$

We group adjacent cells as a block for normalization. The block size is 3×3(Fig. 3. (b)). The number of block is 8×8. The normalization is done with Eq. (4).

$$v'(n) = \frac{v(n)}{\sqrt{\left(\sum_{k=1}^{q\times q\times N} v(k)^2\right)+1}} \tag{4}$$

where v is the histogram, $q\times q$ is the size of one block and N is the number of orientation bins. The set of these block histograms is used as the feature vector for hand gesture recognition. The dimension of feature vector is 8×8×q×q×N=5184.



(a) (b)

Fig. 3. (a) Gradient histogram for each cell. (b) Normalized gradient histograms for each block.

3.3. Principal Component Analysis of HOG Features

In this paper, we adopt training database which contains M=27000 images of depth hand gestures, were built by ourselves. As shown in previous section, the total number of features becomes L=5184 when extracting HOG features from all locations on the gird. Since the dimension of the feature vector is too large for real-time application, We utilize principal components analysis to reduce the dimensionality of the feature vectors. We donate HOG features as $\mathbf{h}_i = \left(h_{1,} h_{2,} \cdots h_{\mathrm{L},}\right)^T$. $\mathbf{m}$ is defined as the mean vector and is calculated by Eq.(5).

$$\mathbf{m} = \frac{1}{M}\sum_{\mathrm{i}=1}^{M} \mathbf{h}_{\mathrm{i}} \tag{5}$$

Covariance matrix S is calculated by Eq. (6).

$$S = \frac{1}{M}\sum_{\mathrm{i}=1}^{M} (\mathbf{h_i} - \mathbf{m})\,(\mathbf{h}_{\mathrm{i}} - \mathbf{m})^{\mathrm{T}} \tag{6}$$

We can calculate eigenvalue λ_i and eigenvector $\mathbf{u}_i$ from covariance matrix S by Eq. (7).

$$\mathbf{S}\mathbf{u}_i = \lambda_i \mathbf{u}_i \tag{7}$$

We sort eigenvalues in descending order to find the optimal subset. As shown in Fig.4, the accumulative proportion rate reaches 90% or more when adding the eigenvalues up to the 142, so we compress features to 142 dimensions. We donate the reduced HOG feature vector as $\mathbf{h}_{\mathrm{reduction}} = (\mathbf{u}_1^T\mathbf{h}, \mathbf{u}_2^T\mathbf{h}, \cdots \mathbf{u}_{142}^T\mathbf{h})^T$.

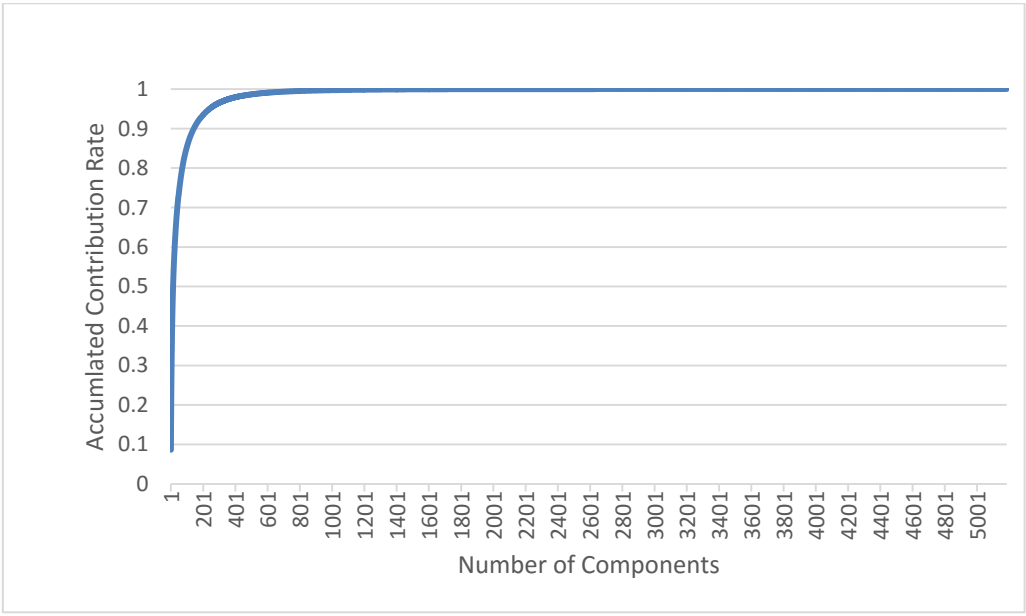


Fig. 4. Accumulative proportion rate.

3.4. Hand Gesture Recognition by a Nonlinear SVM Classifier

To recognize the hand gestures using HOG features, we trained a multi-class SVM classifier using LIBSVM [8]. The RBF kernel was used for nonlinear classification.

A test dataset was collected from 10 persons in advance, each person has 50 pieces of depth images of different hand shapes. Table 1 shows average accuracy rates, and Fig. 5 shows recognition results by box plot figure. X markers are defined as the mean of recognition rate.

Table 1. Recognition Rate (%)

	Hand open	Hand close	Grasp	Finger up	Finger down	Finger right	Finger left	Palm up	Palm down	Mean±SD
Male1	100	98	94	98	100	90	100	98	94	96.89±3.5
Male2	90	74	68	90	100	94	84	92	80	85.78±10.2
Male3	74	100	86	52	88	84	68	78	90	80.00±14.1
Male4	92	96	70	78	82	78	74	76	62	78.67±10.4
Male5	98	96	78	100	100	94	90	98	98	94.67±7.0
Male6	90	86	100	100	100	86	90	76	90	90.89±8.1
Female1	88	84	98	88	84	92	70	76	68	83.11±10.0
Female2	94	90	98	92	100	94	94	98	76	92.89±7.1
Female3	92	92	80	94	94	52	78	32	74	76.44±21.5
Female4	90	96	100	100	100	94	88	100	100	96.44±4.8
Mean±SD	90.80±7.0	91.20±8.0	87.20±12.5	89.20±14.8	94.80±7.4	85.80±13.0	83.60±10.7	82.40±20.6	83.20±13.1	

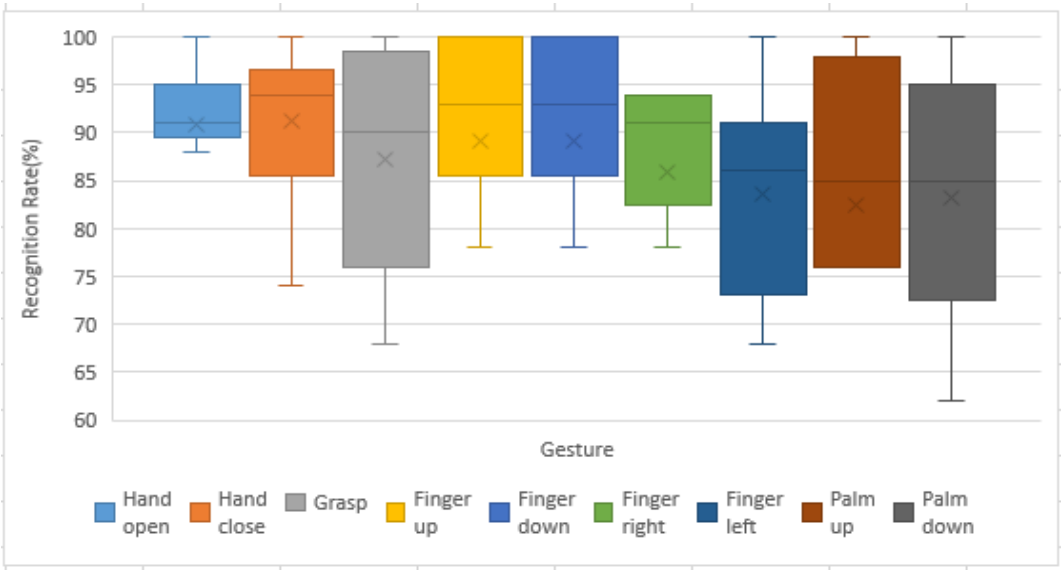
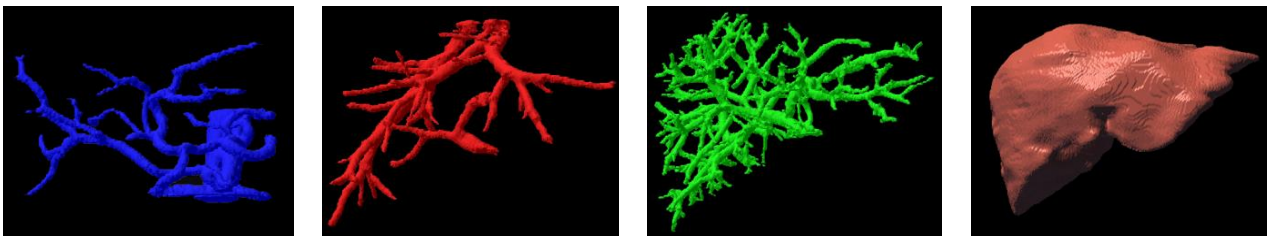


Fig. 5. Recognition results.

4. Visualization

In the visualization module, surface models of hepatic artery, hepatic portal vein, hepatic vein and liver parenchyma (Fig. 6.) are generated by converting volume data to a triangulated mesh surface by the use of marching cube algorithms. The volume data are segmented semi-automatically from CT images under the guidance of a physician [9], [10]. By visualizing these models together and changing the opacity of the liver, the surgeon can easily recognize the liver geometry, its vessels structures and locations during the surgery as shown in Fig. 7. Please refer [9], [10] for detailed information about CT data and segmented liver and vessel data.



(a) Hepatic artery (b) Hepatic portal vein (c) Hepatic vein (d) Liver

Fig. 6. Visualization for liver and vessels.

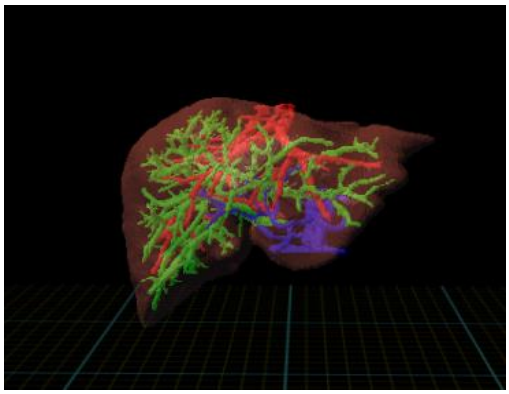


Fig. 7. Visualization of liver and its vessel structure (fused image).

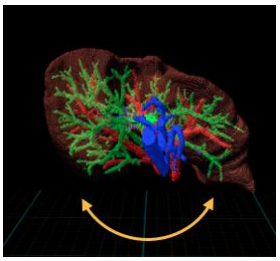
5. Touchless Visualization by Single-Handed Interactions

The proposed system adopts 9 gestures in this paper. Several settings are prepared to prevent unconsciously visualization operations. We just recognize the nearest person within 5*m* for capturing the best hand interactions. When the user's right hand is above waist for 45*cm*, the gesture is became available state. Surface rendering of liver and its vessels are visualized when the system is started. Three kinds of operations are available: rotation, opacity adjustment, display or non-display switching. With the right hand in the shape of grasp, models rotate along X direction when hand moves up and down, while models rotate along Y direction when hand moves right and left (Fig. 8 (a) (b)). We use the hand shape palm up and palm down to adjust the liver's opacity (Fig. 8 (c) (d) (e) (f)). We use finger left, finger up, finger down, finger left to switch display and non-display of hepatic artery, hepatic portal vein and hepatic vein (Fig. 8 (g) (h) (i) (j) (k) (l)). The surgeon can confirm the information of patient's liver and hepatic vessels based on single-handed interactions (Table 2).

Table 2. Visualization Mode Controlled by Gestures

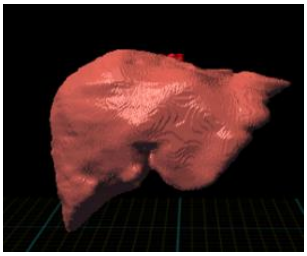
Gesture	Visualization mode
Hand close	idle state
Grasp right, left	Rotation of models
Palm up, down	Adjustment of liver's opacity
finger up, down, left	visualizing of Hepatic artery, portal vein, Hepatic vein

(a) Grasp



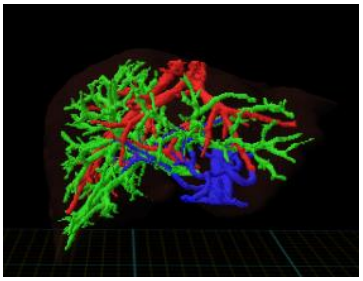
(b) Rotation of models

(c) Palm up



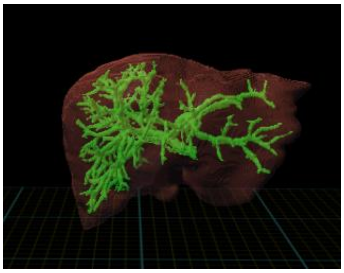
(d) Liver's opacity up

(e) Palm down

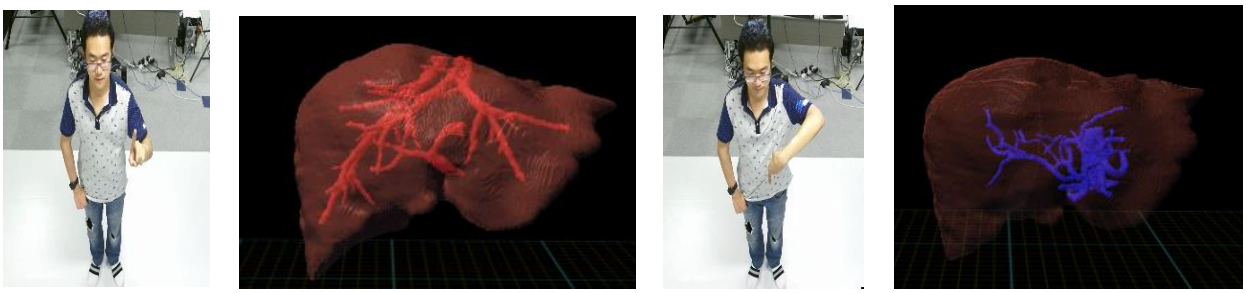


(f) Liver's opacity down

(g) Finger left



(h) Visualizing of Hepatic vein



(i)Finger up (j) Visualizing of Hepatic portal vein (k) Finger Down (l) Visualizing of Hepatic artery

Fig. 8. Visualization operation by single-handed touchless interactions.

(a) (b)Rotation mode, (c) (d) (e) (f)Opacity change mode, (g) (h) (i) (j) (k) (l)Switch display mode

6. Users Evaluation

After using our system, a total of 15 participants completed a subjective user satisfaction questionnaire as shown in Table 3. Evaluation items are divided into visualization, rotation of models, adjustment of opacity, display or non-display operation and etc. Each item is evaluated in 5 levels for each subject. The evaluation decreases as approaching 1, and becomes higher as closer to 5. It can be seen that most of operations are satisfying, but some operations such as the switch display still need to be improved.

Table 3. Questionnaire Results

		1	2	3	4	5	Mean±SD
		weak	←	normal	→	strong	
visulization	three-dimensionality		1	10	1	3	3.40±0.9
		bad	←	normal	→	good	
	clearness			1	8	6	4.33±0.6
	reality		1	3	8	3	3.87±0.8
	easy to understand			3	5	7	4.27±0.8
Rotation of models	intuitive		2	2	6	5	3.93±1.0
	Smoothness		1	3	4	7	4.13±0.9
	accuracy		1	4	6	4	3.87±0.9
	Fatigue		3		5	7	4.07±1.1
Adjustment of opacity	intuitive	1	1	4	3	6	3.08±1.2
	Smoothness		1	5	3	6	3.93±1.0
	accuracy		2	6	3	4	3.80±1.0
	Fatigue	1	1	4	3	6	3.80±1.2
switch display or non-display	intuitive		1	6	3	5	3.80±1.0
	Smoothness		1	6	3	5	3.80±1.0
	accuracy		3	4	7	1	3.40±0.9
	Fatigue		3	3	3	6	3.80±1.2
others	Wearing comfort of glasses		2	7	4	2	3.40±0.9
	range of movement			4	6	5	4.07±0.8
	Immersion feeling			2	8	5	4.20±0.6
	easy to learn			2	7	6	4.27±0.70

7. Conclusion

We developed a touchless visualization system for medical volume. By using our system, surgeons can operate and check the information of patient body without touching devices. The system can visualize the liver and its vessel structure together and the visualization is controlled by single-handed interactions. We use the depth images obtained by Kinect and HOG feature extraction to recognize hand shape. User evaluations show that most of operations are satisfying. In the future work, we are going to improve some operations such as switch display using some deep learning methods to make the method robust enough to work in surgery. Moreover, experiments involve surgeon will be conducted.

Acknowledgment

This work is supported in part by the Grant-in Aid for Scientific Research from the Japanese Ministry for Education, Science, Culture and Sports (MEXT) under the Grant Nos. 15K16031, 17H00754, 17K00420; and in part by the MEXT Support Program for the Strategic Research Foundation at Private Universities, Grand No.S1311039 (2013-2017). We would like to thank Dr. Kaibori of Kansai Medical University for providing medical data and helpful advice on this research.

References

[1] Gallo, L. (2011). *Controller-Free Exploration of Medical Image Data: Experiencing the Kinect.* National Research Council of Italy Institute for High Performance Computing and Networking.

[2] Yoshimitsu, K., Muragaki, Y., Iseki, H., *et al.* (2014). Development and initial clinical testing of "OPECT": An innovative device for fully intangible control of the intraoperative image-displaying monitor by the surgeon. *Neurosurgery*, *Suppl 1,* 46-50.

[3] Ruppert, G. C., Coares, C., Lopes, V., *et al*. (2012). Touchless gesture user interface for interactive image visualization in urological surgery. *World J. Urol., 30,* 687-691.

[4] Fujii, R., Tateyama, T., Chen, Y. W., *et al*. (June 15-17, 2016). A touchless visualization system for medical volumes based on Kinect gesture recognition. In Y. W., Chen, *et al.* (Eds.), *Proceedings of Innovation in Medicine and Healthcare 2016* (pp. 209-215). Springer.

[5] Kitware. *Visualization Tool Kit (VTK).* Retrieved from http://www.vtk.org/

[6] Dalal, N., & Triggs, B. (2005). Histograms of oriented gradients for human detection. *Proceedings of 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)* (pp. 886-893), San Diego, CA, USA.

[7] Vladimir, N. V. (Nov. 1999). *The Nature of Statistical Learning Theory, Statistics for Engineering and Information Science* (2nd ed.). Springer.

[8] Chang, C. C., & Lin, C. J. (2016). *LIBSVM: A Library for Support Vector Machines*. National Taiwan University. Retrieved December 22, 2016 from https://www.csie.ntu.edu.tw/~cjlin/libsvm/

[9] Kaibori, M., Chen, Y. W., Matsui, K., Ishizaki, M., Tsuda, T., Nakatake, R., Sakaguchi, T., Matsushima, H., Miyawaki, K., Shindo, T., Tateyama, T., & Kwon, A. H. (2013). Novel liver visualization and surgical simulation system. *J Gastrointest Surg., 17,* 1422-1428.

[10] Tateyama, T., Kaibori, M., Chen, Y. W., *et al*. (2013). Patient-specified 3D-visualization for liver and vascular structures and interactive surgical planning system. *Medical Imaging Technology, 31,* 176-188.

Jiaqing Liu was born in 1993, received a B.E. degree in 2012 from Northeastern University, China. Now he is a graduate student in Graduate School of Information Science and Engineering, Ritsumeikan University in Japan. His research interests include image processing and analysis, virtual reality, body detection and deep learning.

Ryoma Fujii was born in 1992, received a B.E. and a M.E. degrees in 2015 and 2017 from Ritsumeikan University in Japan. His research interests include image processing, medical image visualization and virtual reality.

Tomoko Tateyama received a B.E. degree, a M.E degree in 2003, a Ph.D degree from University of the Ryukyus, Okinawa Japan in 2001, 2003 and 2009, respectively. She was an assistant researcher in 2009-2012, and assistant professor in 2013-2015 at Ritsumeikan University. She is currently an assistant professor in Hiroshima institute of Technology, Hiroshima Japan. Her research interests include Medical image analysis visualization, computer anatomical model, pattern recognitions, computer graphics and vision, virtual reality, development computer aided surgery/diagnosis system. And she is the IEEE SMC and EMBC member.

Yutaro Iwamoto received the B.E. and M.E., and D.E. degree from Ritsumeikan University, Kusatsu in Japan in 2011 and 2013, and 2017, respectively.

He is currently an assistant professor at Ritsumeikan University, Kusatsu, Japan. His current research interests include medical image processing and computer vision, and deep learning.

Yenwei Chen received a B.E. degree in 1985 from Kobe Univ., Kobe, Japan, a M.E. degree in 1987, and a D.E. degree in 1990, both from Osaka University, Osaka, Japan. From 1991 to 1994, he was a research fellow with the Institute for Laser Technology, Osaka. From October 1994 to March 2004, he was an associate professor and a professor with the Department of Electrical and Electronic Engineering, University of the Ryukyus, Okinawa, Japan. He is currently a professor with the College of Information Science and Engineering, Ritsumeikan University, Kyoto, Japan. He is also a chair professor with the College of Computer Science and Technology, Zhejiang University, China. He is an associate editor of International Journal of Image and Graphics (IJIG) and an editorial board member of the International Journal of Knowledge based and Intelligent Engineering Systems. His research interests include pattern recognition, image processing and machine learning. He has published more than 200 research papers in these fields.