Image Interpretation System for Informed Consent to Patients by Use of a Skeletal Tracking

Naoki Kamiya[1], Hiroki Osaki[2], Jun Kondo[2], Huayue Chen[3], and Hiroshi Fujita[4]

[1]*Department of Information and Computer Engineering, Toyota National College of Technology, Toyota, JAPAN*
[2]*Advanced Course of Computer Science, Toyota National College of Technology, Toyota, JAPAN*
[3]*Department of Anatomy, Graduate school of Medicine, Gifu University, Gifu, JAPAN*
[4]*Department of Intelligent Information, Division of Regeneration and Advanced Medical Sciences, Graduate School of Medicine, Gifu University, Gifu, JAPAN*

E-mail n-kamiya@toyota-ct.ac.jp

Abstract

In recent years, a variety of medical images are used for diagnosis. However, it is difficult to understand these images by patient without anatomical knowledge. In particular, CT images consist of large number of slice images. It causes a lack of understanding about positional information in their body. This will adversely affect the first-person informed consent (IC). We developed an image interpretation system for the patient. Our purpose of this system is patient participation in clinical site. This system is based on the skeletal tracking for recognizing patient's hand movements and determines the relative position of the body. The skeletal tracking has been implemented using the Microsoft Kinect sensor with official SDK (KinectSDK-v1.0-bata2-x86). We performed association of tracking marker and a CT number. Finally, the calculated CT images are displayed on the iPad handled with the patient. We tested the effectiveness of the proposed system in cooperation with volunteer subjects without anatomical knowledge. As a result, they were able to properly identify the proposed position of the cross-sectional image than before.

1. Introduction

In the field of health care, due to the advancement of medical equipment, the imaging device such as MRI, CT and PET/CT is often used for diagnosis. The tomographic image is a high-definition. However, the large amounts of images can be obtained in an imaging session. For example, in the torso CT images were used in this study is composed of a cross-sectional images over 1,000 per patient. In such images, a doctor select the number of images used in the diagnosis from within the huge image, then presented and explains to the patient. Normally, all slice images or the image list of the required range is displayed for the patient. Therefore, some patients can understand its spatial position roughly, in general, it is considered to be difficult for an intuitive understanding for patient from a few images of tomographic image without anatomical knowledge. In addition, the importance of the informed consent (IC) has been increasing recently [1, 2]. Informed consent means the patient's medical care based on their free will. It cannot be achieved if without their "understanding" based on the accurate information.
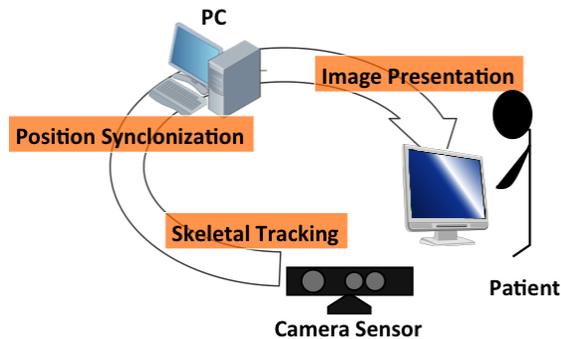
Fig. 1. Conceptual diagram of the proposed method

Therefore, the system must be able to understand easily for the tomographic images by the patient without anatomical knowledge. The image presentation method can be considered variety of forms, we develop the image interpretation system that can be understood intuitively the patient himself. In this study, we generate the above-mentioned system using an environmental mounted sensor. Using the Microsoft Kinect, we measure the coordinates of each part of the body by skeletal tracking. The skeletal tracking using Kinect have attracted much attention in the medical research field because of the no need to add equipment to the patient and can be introduced at low cost. As an example research using Kinect in the medical field, such as machine control by obtaining the body position and movements [3].

In this study, we develop a system to display the corresponding CT images with the behavior of the patient's hands in real time by use of the skeletal position estimation based on the skeletal tracking. Figure 1 shows the conceptual diagram of the proposed method. This system consists of three parts, the patient, the sensor, and a personal computer (PC). The patient does not require a contact-type sensor, though it is non-invasive, the patients can be understood their internal structure of the body by intuitive behavior. It is a system that can be understood in association with the position of their own body images that are presented by the doctor. Evaluation is carried out by an observer study. Evaluate the patient's level of understanding as to whether or not improved by the system use.

2. Methods

In our previous study, we proposed a concept model of this image presentation system [4]. This previous system also uses the Kinect as the environment-
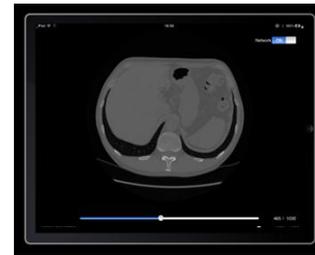

Fig. 2. The system interface of our previous system [4].

mounted sensor. In addition, we use the tablet device, "iPad" for presenting images to the patient. We have developed the system acquires the spatial position of the iPad, to present a corresponding tomographic image in any cross-section. Figure 2 shows the interface of our previous system. In this system was useful in understanding cross-sectional tomographic image. However, it is necessary to operate the tablet device by the patient. Because of it, such as elderly, will focus on the operation of the iPad, which prevented an intuitive understanding about image which presented by the doctor.

In this paper, we develop a system using only the relative position of the body without iPad to make the decision of the displayed cross-section image. Therefore, patients need not to understand how to operate the iPad. In addition to the cross-sectional plane, we achieve presentation another anatomical-plane; sagittal and coronal plane. The patient can understand three-dimensional positional information of their body not only cross-sectional view. In this way, we suppose that to improve understanding for the patient about the organ locations. In particular, when presented with a massive lesion, it has a three-dimensional spread, such as a depth direction. Therefore, it is considered that can understand the spatial spread of the lesion region in the body by three cross-sectional views in the proposed system.
Figure 3 shows the flow chart of our proposed method. The system consists of four parts. The details are described in the following sections.

2.1. Development Environment
In this paper, we developed image interpretation system under the following conditions. Using the Microsoft's Kinect as a environment-mounted sensor. A computer is a generic 32-bit personal computer.

2.2. Input Images
Tomographic image is composed of a large number of images; it is difficult to understand for the patient.

Therefore, CT image is used as input image in this study. However, the algorithm of the proposed method is not intended for a particular modality. In principle, it can be applied to various images, such as MRI and PET/CT.

Here, we use torso CT image as input image, which imaged from pubic bone to shoulder. The image is composed one case per 1000 cross-sectional images. Similarly, the sagittal plane and coronal plane is composed of about 500 images. At the same time, skeletal image is also used as input image for the position synchronization described in Section D. The skeletal image is segmented each bone by use of skeletal topology [5].

2.3. Skeletal Tracking

Based on the skeletal tracking, make the acquisition of the distinctive position on the patient's body. As mentioned above, to achieve the skeletal tracking by using a KinectSDK. Control point obtained by the skeletal tracking is used to enter the position where the patient wants to view and also used to alignment of the image.

Figure 4 shows the three images acquired from the Kinect. (a) is the normal pictured image, (b) is the depth image and (c) is the skeletal tracking image. The depth image changes the gray value depending on the depth information obtained from the depth sensor, and the human area is also recognized and colored independently. The skeletal tracking image is an image connected by the straight line between each landmark points. In the KinectSDK, a total of 20 landmarks can be automatically recognized upper two people at the same time. Figure 5 shows the landmarks correspondence with body position.

At the input of the image position want to observe by the patient is achieved by calculating the relative position between landmarks. The relative position as
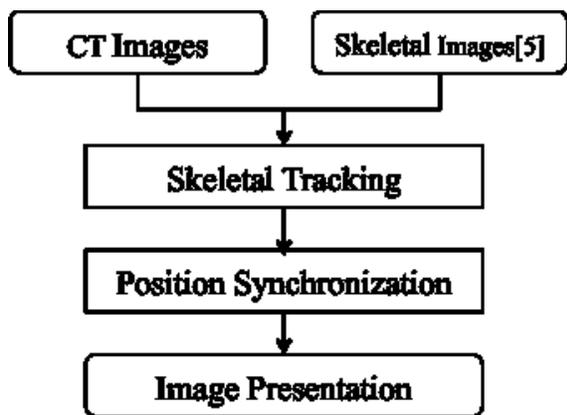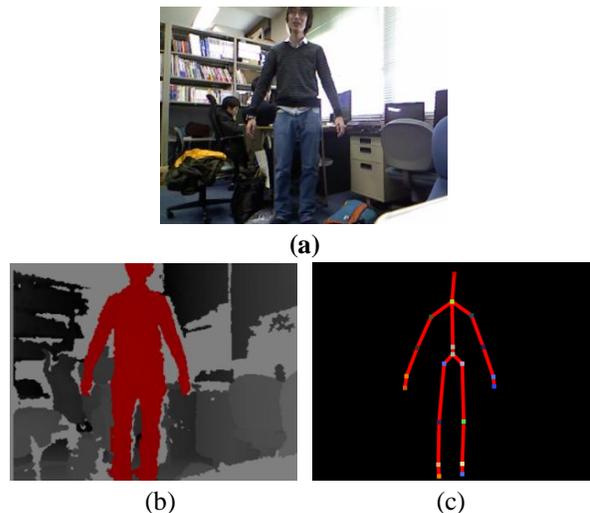

(a)

(b)   (c)

Fig. 4. Image acquired from Kinect: (a) normal, (b) depth, and (c) skeletal tracking image.

the difference value, calculated using the right hand (shown HAND_RIGHT in fig.5) and other three landmarks. In cross-section, use the relative position of the umbilicus (show HIP_CENTER) and the neck (SHOULDER_CENTER) to the position of the right hand. Similarly, in the sagittal plane is used the right shoulder landmark (SHOULDER_RIGHT). In coronal plane, use the depth information of the right hand to select the display image from the depth changing.

2.4. Position Synchronization

Make the selection of the patient's CT slices corresponding to the body position that was entered in the previous section because of the large slice numbers in torso CT images. In this
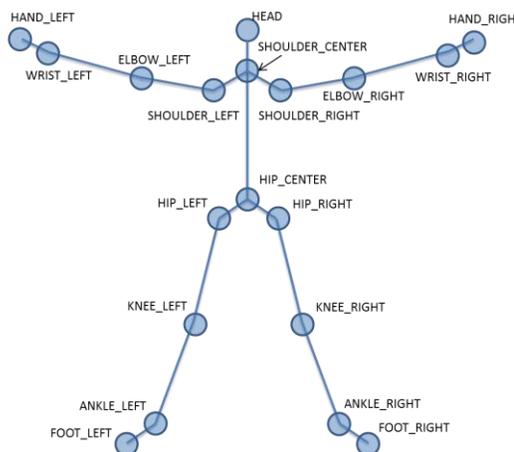

Fig. 3. Flowchart of our proposed method

Fig. 5. Landmark positions on the human body.
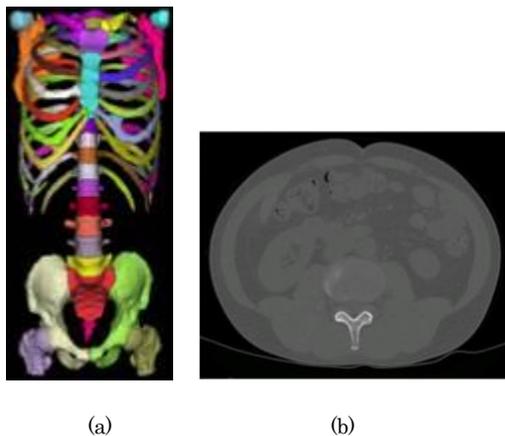
(a)   (b)

Fig. 6. Umbilicus slice detection: (a) skeletal image, (b) detected umbilicus slice.

paper, an umbilicus position in the CT image is recognized with relative ease, so we recognize umbilicus position in the CT images automatically and synchronize the umbilicus position (HIP_CENTER) recognized by the KinectSDK in the skeletal image.

The recognition of the umbilicus cross-section in the CT image makes use of the skeletal image, which is one of the input images. The axial position of the umbilicus is located at the top of the pelvis, and also located at the lowest end of the costal bone. We determine the umbilicus position between these two slices. Figure 6 shows the image of the skeletal image (a) and detected umbilicus slice (b). In figure 6 (a), each color represents a skeleton recognized separately [5]. Based on this umbilicus position, we determine the display image using the right hand position obtained in Section C. The cross-sectional position in input CT images to be synchronized with the approximate spatial location of the right hand in the following method.

In the axial plane, the shoulder and hip landmarks define the upper limit and lower limit, respectively. Then, in the torso CT image, they should be mapped to the top and bottom of the image. Then, to calculate where to locate the right hand between the top and bottom for determine the display slice.

In the sagittal plane, umbilicus and right shoulder landmarks define the center and the right endpoint, respectively. Then, in the torso CT
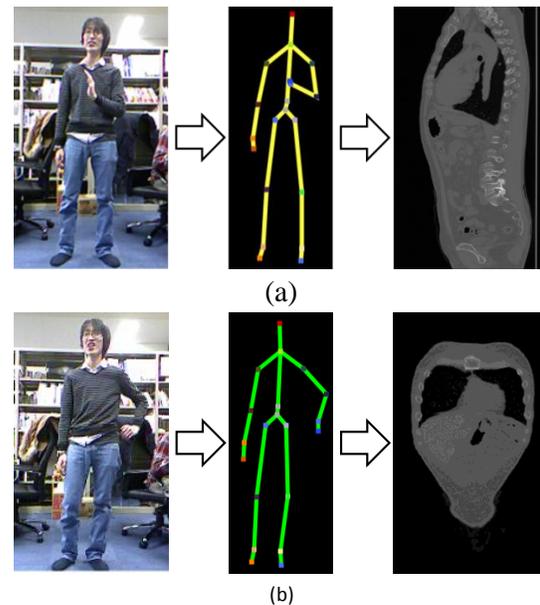

(a)

(b)

Fig. 7. Synchronized slice image: (a) sagittal and (b) coronal plane.

image, they should be mapped to the umbilicus and right edge of the image.

In the coronal plane, the patient's umbilicus landmark is defined to the reference position. Then, determine the image due to changes in the relative position of the right hand against the umbilicus landmark. Empirically, track the movement of the hand in the range of around 15cm from the umbilicus. When indicating the location closest to the camera, assign to the foreground image.

Figure 7 shows the synchronized slice image between patient's hand movement and corresponding CT slices. (a) shows the sagittal plane, and (b) shows the coronal plane.

3. Results and Discussion

For the verification of whether or not be able to grasp in association with body position, the tomographic image obtained from this system. Conduct experiments on the 10 subjects without anatomical knowledge. Evaluate the changes in the level of understanding about the location on the body before and after use in the proposed system.

First, we presented five cross-sectional images randomly. These images are selected from, shoulder, lung, xiphoid, umbilicus and pelvis cross-section. Subjects had to answer the thinking position on the body to the number line. Next, the subjects freely

Table 1. Response error before and after use of the system.

	Average	Standard Deviation
Before	14.94	35.91
After	9.49	6.64
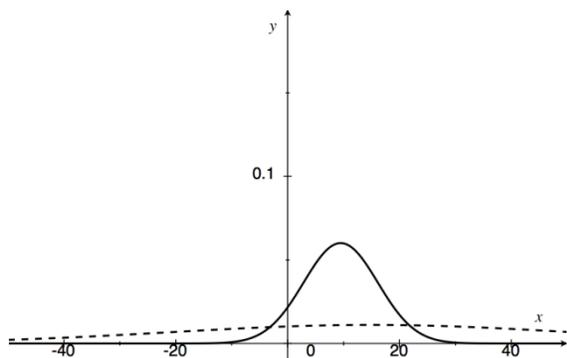
Fig. 8. Response error before and after use of the system: dotted line indicates the system before use, solid line indicates the system after use.

observe the CT images using this system. Then, the same five cross-sectional images are displayed again randomly, they answer on the number line once again. Finally, determine the error between correct answer and input answer on the number line.

Table 1 shows the average of response error in a five cross-sectional view. The use of the system, the average value of the error was reduced to 9.49 from 14.94. In addition, standard deviation of the error is significantly decreased from 35.91 to 6.94. Figure 8 shows the same results approximated to the normal distribution. After using this system, from the shape of the graph, many subjects lead to correct answer. However, it can be confirmed that the convergence has been shifted in a positive direction. It is considered from an error of scale or system error of display position.

Therefore, it is considered to increase the level of understanding for the presented CT images have been achieved.

4. Conclusion

In this study, we developed a novel image presentation system to support informed consent. This system can be understood intuitively for the patient to tomographic images like CT. Compared with our conventional method [4], eliminate the need for "iPad" as displaying device, and possible to observe two another sections; sagittal and coronal. The usefulness of the system has been shown in the experiment for the people without anatomical knowledge. In the future, it is necessary to achieve the reconstruction of arbitrary cross-section and to consider the application to medical education.

References

[1] R. Volpe, "Patients' expressed and unexpressed needs for information for informed consent," Journal of Clinical Ethics, vol. 21, no. 1, pp. 45-57, 2010.

[2] Y. Ivashkov, V. Norman, "Informed consent and the ethical management of the older patient," Anesthesiol Clinical, vol. 27, no. 3, pp. 569-580, 2009.

[3] AP.Bo, M. Hayashibe, P. Poignet, "Joint angle estimation in rehabilitation with inertial sensors and its integration with Kinect," Proceeding of IEEE Eng. Med. Biol. Soc., pp.3479-3483, 2011.

[4] J. Kondo, N. Kamiya, H. Osaki, T. Hara, C. Muramatsu, and H. Fujita, "Interactive system for next-generation medical care system using iPad and Kinect," 97th Scientific assembly and annual meeting of the radiological society of north America, radiology informatics series: mobile computing devices (scientific formal (paper) presentations), MSVR31-05, p. 91, 2011.

[5] X. Zhou, T. Hayashi, M. Han, H. Chen, T. Hara, H. Fujita, R. Yokoyama, M. Kanematsu, and H. Hoshi, "Automatic segmentation and recognition of the bone structure in non-contrast torso CT images using implicit anatomical knowledge," Proceeding of SPIE, 7259, 72593S, 2009. Doi: 10.1117/12.812945.