Copyright ⓒ 2015 KSII

MultiView-Based Hand Posture Recognition Method Based on Point Cloud

Wenkai Xu[1], Ick-Soo Lee[2], Suk-Kwan Lee[3], Bo Lu[4], Eung-Joo Lee[5]
[1] Department of Information Communications and Engineering, University of Tongmyong
Busan, 608-711 Korea
[e-mail: xwk6298@hotmail.com]
[2] Department of Management, College of Kyungsang
Busan, 611-701 Korea
[e-mail: isle@bsks.ac.kr]
[3] Department of Information Security, University of Tongmyong
Busan, 608-711 Korea
[e-mail: skylee@tu.ac.kr]
[4] Institute of Electronic Commerce and Modern Logistics, Dalian University
Dalian, 116622 China
[e-mail: lubo_documents@hotmail.com]
[5] Department of Information Communications and Engineering, University of Tongmyong
Busan, 608-711 Korea
[e-mail: ejlee@tu.ac.kr]
*Corresponding author: Eung-Joo Lee

Received July 21, 2014; revised March 8, 2015; accepted May 3, 2015; published July 31, 2015

Abstract

Hand posture recognition has played a very important role in Human Computer Interaction (HCI) and Computer Vision (CV) for many years. The challenge arises mainly due to self-occlusions caused by the limited view of the camera. In this paper, a robust hand posture recognition approach based on 3D point cloud from two RGB-D sensors (Kinect) is proposed to make maximum use of 3D information from depth map. Through noise reduction and registering two point sets obtained satisfactory from two views as we designed, a multi-viewed hand posture point cloud with most 3D information can be acquired. Moreover, we utilize the accurate reconstruction and classify each point cloud by directly matching the normalized point set with the templates of different classes from dataset, which can reduce the training time and calculation. Experimental results based on posture dataset captured by Kinect sensors (from digit 1 to 10) demonstrate the effectiveness of the proposed method.

Keywords: Hand posture recognition, depth information, noise reduction, 3D point cloud, Kinect

This research was supported by the Busan Metropolitan City, Korea, under the 2015 BB21 program grants and Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education, Science and Technology(NRF-2011-0023118).

http://dx.doi.org/10.3837/tiis.2015.07.0014 **ISSN : 1976-7277**

1. Introduction

Human-computer interaction (HCI) plays an important role in various applications today. It exists the desire to realize more natural forms of interaction between humans and machines. The user should ideally interact with machines without the need of cumbersome devices (such as colored markers or gloves) or apparatus like remote controls, mouse and keyboards. Hand gestures can give an alternative and easy means of communication with machines and could revolutionize the way we use technology in our everyday activities.

Hand gesture recognition from visual images has a number of potential applications in HCI, machine vision, virtual reality (VR), machine control in the industrial field, and so on. Most conventional approaches to hand gesture recognition have employed data gloves. But, for more natural interface, hand gesture must be recognized from the visual images as in the communication between humans without using any external devices. Our research is intended to find a high-efficiency approach to improve the algorithm of hand detecting and hand gesture recognition.

A typical example of this trend is the gaming industry and the launch of Microsoft's new product-Kinect. Other domains, where gesture recognition is needed, include but are not limited to sign language recognition, virtual reality environments and smart home system. Recovering the full kinematic parameters of the skeleton of the hand over time, commonly known as the hand-pose estimation problem, is challenging for many reasons: high dimensionality of the state space, self-occlusions, insufficient computational resource, uncontrolled environments, rapid hand motion and noise in the sensing device, etc.

Extensive researches have been conducted on hand posture recognition making use of 2D digital image [1, 2]. However, it is still ongoing research as most papers do not provide a complete solution to the previously mentioned problems. As the first step of hand gesture recognition, hand detection and tracking are usually implemented by skin color or shape based segmentation, which can be derived from RGB image [3]. However, because of the intrinsic vulnerability against background clutters and illumination variations, hand gesture recognition on 2D RGB images usually requires a clean and simple background, which limits its applications in the real world.

With the rapid development of RGB-Depth (RGB-D) sensors, it becomes possible to obtain the 3D point cloud of the observed scene and offers great potential for real-time measurement of static and dynamic scenes. This means some of the common monocular and stereo vision limitations are partially resolved due to the nature of the depth sensor. Compared to the traditional RGB camera, research on 3D depth map has significant advantages for its availability to discover strong clues in boundaries and 3D spatial layout even in cluttered background and weak illumination. Particularly, those traditional challenging tasks such as object detection and segmentation become much easier with the depth information [4].

The recent progress in depth sensors such as Microsoft's Kinect device has generated a new level of excitement in gestures recognition. Several researchers have proposed some approaches based on depth information for this issue [5, 6]. Depth image generated by depth sensor is a simplified 3D description, however most of current methods only treat depth image as an additional dimension of information and implement recognition process in 2D space. Ren et al. employed a template matching based approach to recognize hand gestures through a histogram distance metric of Finger Earth Mover Distance (FEMD) through a near-convex

estimation [5]. Bergh and Van Gool [6] used a Time of Flight (ToF) camera combined with a RGB camera to successfully recognize four hand gestures by simply using small patches of hands. However, their method only considered the outer contour of fingers but ignored the palm region that also provides crucial shape and structure information for complex hand gestures. Most of these methods explicitly use the sufficient 3D information conveyed by the depth maps.

The contribution of this paper is to leverage multi-viewed 3D point cloud for hand posture recognition. Unless using 2D descriptor, sufficient 3D information can be obtained after point registration, which can get rid of the obstruction of self-occlusion and rotation. Moreover, users will interact with the two sensors much more naturally rather than making a command scrupulously by using the conventional method. The experimental results demonstrate the method proposed in this paper gives effective and robust performance.

2. Related Work

Human hand is a highly deformable articulated object with a total of about 27 degrees of freedom (DOFs) [7]. As a consequence the hand can adopt a variety of static postures that can have distinct meanings in human communication.

Hand posture recognition techniques consist of two stages: hand detection and hand pose classification. First the hand is detected in the image and segmented. Afterwards information is extracted that can be utilized to classify the hand posture. This classification allows it to be interpreted as a meaningful command [8].

A first group of hand poses recognition researchers focus on these so-called “static” hand poses. A second research domain is the recognition of “dynamic” hand gestures, in which not the pose but the trajectory of the hand is analyzed. This article focuses on the static hand poses. For more information on dynamic gestures see [3, 9], etc.

Hand detection techniques can be divided into two main groups [3]: data-glove based and vision-based approaches. The former sensors are attached to a glove to detect the hand and finger positions. The latter requires only a camera, so they are relatively low cost and are minimally obtrusive for the user. The vision based approaches can detect the hand using information about the depth, color, etc. Once the hand is detected, hand posture classification methods for vision-based approaches can be divided into three categories: low level features, appearance based approaches and high-level features.

Numerous researchers raised the thought that full reconstruction of the hand is not necessary for gesture recognition. Therefore, these methods only use low-level image features that are fairly robust to noise and can be extracted quickly. An example of low-level features used in hand postures recognition is the radial histogram. Appearance-based methods use a collection of 2D intensity images to model the hand. These images can for example be acquired by Principal Component Analysis (PCA). Some researchers who use appearance based approaches are [10-12]. Methods relying on high-level features use a 3D hand model [17-19]. High-level features can be derived from the joint angles and pose of the palm. Most model-based approaches create a 3D model of a hand by defining kinematic parameters and project the 3D model onto a 2D space. The hand posture can be estimated by finding the kinematic parameters of the model that result in the best match between the projected edges and the edges extracted from the input image. Other approaches reconstruct the hand posture as a “voxel” model, based on images obtained by a multi-viewpoint camera system. The joint

angles are then estimated by directly matching the three dimensional hand model to the measured voxel model.

One advantage of 3D hand model based approaches is that they allow a wide range of hand gestures if the model has enough DOFs. However, these methods are also given disadvantages. The database required to cover different poses under diverse views is very large, complicated invariant representations have to be used. The initial parameters have to be close to the solution at each frame. Moreover the fitting process is highly sensitive to noise. Due to the complexity of the 3D structures used, these methods may be relatively slower than the other approaches. Of course, this problem must be suppressed in order to assure on-line performance.

3. Methodology of Hand Posture Recognition

3.1 Environment Setting

As we mentioned above, it is not ample to obtain a robust hands segmentation by using a RGB camera as variable illuminations, self-occlusions and uncontrolled background [1, 2]. For using the 3D information that very useful for hand detection and segmentation sufficiently, a stereo vision-based system with two Kinect sensors is established in our study (as **Fig. 1**).

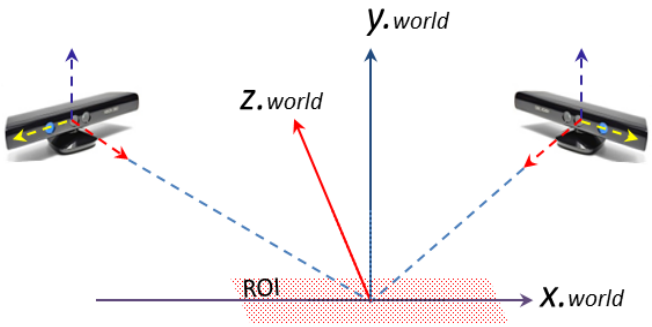


Fig. 1. Stereo vision-based system with two Kinect sensors

A default value of yaw angle is set for each Kinect for keeping the view on the same horizontal plane, and an angle of 45 degree with the X-axis world coordinate system provides the most information while performing gestures that have inevitable occlusion, within a certain distance from each Kinect, each one can see "half" of the hand, hence capturing more structural information. The red shaded region is considered as "region of interest (ROI)", which is the effective area of operation and is actually a projection of cuboid (mentioned in the section 3.3). Furthermore, $X.world, Y.world$ and $Z.world$ indicate the coordinate axes of the world coordinate system.

Based on the designed system, users will interact with the two sensors much more naturally rather than making a command scrupulously as before. Moreover, the invaluable point cloud could be obtained by normalization operation (discussed in the section 3.3) by the devised system.

3.2 Preprocessing and Hand Segmentation Based on Depth Information

Although Kinect sensor has the advantage of rapid imaging and obtaining RGB image and depth image simultaneously, unfortunately, the quality of depth image is poor and noisy since

it is intended for gesture and body motion recognition purpose. This noise is mainly caused by the design of its hardware with IR projector. To hand posture recognition method in this paper, obtaining accurate point cloud is crucial to further processing, so we should perform the noise reduction firstly.

Filtering is perhaps the most fundamental operation of image processing and computer vision. In the broadest sense of term "filtering", the value of the filtered image at a given location is a function of the values of the input image in a small neighborhood of the same location. In reference [13], authors proposed a bilateral filtering for edge-preserving smoothing; the generality of bilateral filtering is analogous to that of traditional filtering, which called domain filtering. To bilateral filtering, weight coefficient w consist of weight of spatial domain w_s and weight of grey domain w_r,

$$w = w_s \times w_r \tag{1}$$

Here,

$$w_s = exp(-\frac{(i-x)^2+(j-y)^2}{2\sigma_s^2}) \tag{2}$$

$$w_r = exp(-\frac{(I(i,j)-I(x,y))^2}{2\sigma_r^2}) \tag{3}$$

Where $(i,j) \in \Omega$ is the neighborhood of (x,y), $I(x,y)$ indicates the grey value of depth map at (x,y); σ_s and σ_r are standard deviation of Gaussian function, which determine the performance of bilateral filtering. Supposing the size of the input image is with 640×480; Ω is with 11×11, frame rate of the input image is 30fps, computer has to perform over 1.15×10^{10} exponent arithmetic per second; it greatly limits the speed of filtering. As we have known the noise range of depth image captured by Kinect sensor is within 3mm, namely we regard two pixels as on two depth planes if their depth difference is greater than 3mm when the depth values of these two pixels exist.

We replace (3) with a simple binary function in this paper as

$$w_r = \begin{cases} 1 & |I(x,y) - I(i,j)| \le 3, I(x,y) \neq 0 \\ 0 & |I(x,y) - I(i,j)| > 3, I(x,y) \neq 0 \end{cases} \tag{4}$$

By using formula (2) and (4), the depth information can be effectively obtained and the noise also be removed successfully. Moreover, proposed improved bilateral filtering can reduce the calculation time greatly. **Fig. 2** shows the performance of proposed filtering.

(a) (b)

Fig. 2. Depth noise reduction and smoothing: (a) raw depth map, (b) depth map after preprocessing

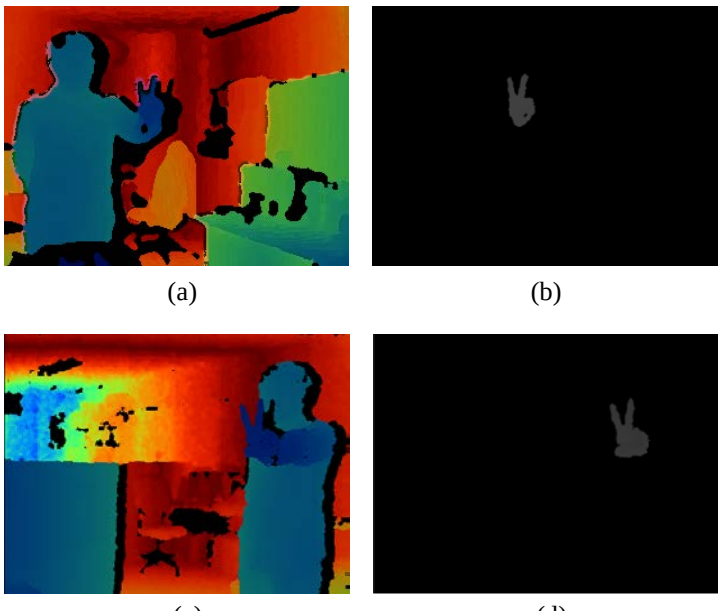
As we discussed above, before each kind of hand posture recognition algorithm can be evaluated effectively the hand needs to be segmented from the input images. As the color information is sensitive to lighting conditions, background and shift, in our study color (RGB) information cannot be taken into consideration, we expect to create a hand posture recognition system that without influence by brightness, shift and highly robust, against the noise.

A reasonable assumption i to ensure the hand is always the most front body part facing the camera, so we either inherit this heuristic rule to pre-processing 3D depth maps to segment hand regions based on depth information.

Given the point cloud $\Phi = \{p_1, p_2, \ldots, p_n\}$in the world coordinate system, we need to extract the points that belong to the user's hand. We required that during this process, the hand will be the closest object to each Kinect sensor. The coordinate of the closest point is written as $p_{closest} = (X, Y, Z)$. The subset $\Phi' = \{hp_1, hp_2, \ldots, hp_n\}$ will be searched in Φ for hand region, where the following conditions controlled [19].

$$\begin{cases} p_{closest}.X \le hp_n.X \le p_{closest}.X + 0.15m \\ p_{closest}.Y \le hp_n.Y \le p_{closest}.Y + 0.15m \\ p_{closest}.Z \le hp_n.Z \le p_{closest}.Z + 0.2m \end{cases} \tag{5}$$

The subset of point cloud Φ' is guaranteed to be contained in a box with volume $0.15*0.15*0.1=0.00125m^3$. The value 0.15 and 0.2 for width, height and depth of the bounding box, were determined empirically to ensure it can contain hands of various sizes. And we eliminate the forearm part by performing connected component analysis and selecting the largest remaining to sub-segment as input to the point cloud registration. The hand segmentation results from each sensor are shown in **Fig. 3**.



(a) (b)

(c) (d)

Fig. 3. Hand segmentation results: (a) and (c) are depth images captured from two sensors, (b) and (d) are hand region segmentation results.

3.3 Hand Posture Recognition Algorithm

Based on the system we designed, two Kinect sensors can obtain most of valid depth information from two perspectives; the registration of two point cloud subset is to assign correspondences between two sets of points and to recover the transformation that maps one point set to the other for addressing the issue of rotation and self-occlusion.

In this study, we leverage probabilistic algorithm for the point set registration from two Kinect sensors. Since the posture recognition contains only point cloud templates corresponding to the users' hands shape, we adopt rigid point set registration approach for lowering computational complexity and reducing the sensitivity to noise rather than no-rigid method [20].

Let $X_{3\times N} = (x_1, \dots, x_N)$ be the first point set containing N points from sensor 1 and $Y_{3\times M} = (y_1, \dots, y_M)$ be the second point set containing M points form sensor 2. We consider the points in Y as the GMM centroids, and points in *X* as the data points generated by the GMM. The GMM probability density function is

$$p(x) = \sum_{m=1}^{M} P(m)p(x|m) \tag{6}$$

where

$$p(x|m) = \frac{1}{(2\pi\sigma^2)^{3/2}} exp^{-\frac{\|x-sRm-t\|^2}{2\sigma^2}} \tag{7}$$

Here, *s* is the scaling factor, *R* is the rotation matrix and *t* is the translation vector; $P(m) = 1/M$ and σ^2 is the isotropic covariance.

In order to estimate *s*, *R* and *t*, we maximize the likelihood function or equivalently minimize the negative log-likelihood function as

$$E(s, t, R, \sigma^2) = -\sum_{n=1}^{N} log \sum_{m=1}^{M} P(m)p(x_n|m) \tag{8}$$

Expectation Maximization (EM) algorithm can be used to find these parameters in (8) and we use toolbox in [14] for fast implementation. The registration result of point set from two sensors is shown in **Fig. 4(a~c)**.

Furthermore, Least Square Method (LSM) is used to fit this new surface and the normal vector of the plane is $n = (a, b, c)$. We intend to correct the orientation of the surface, namely to rotate the fitting surface to parallel to the *X*-*Y* plane. This assumes that φ_1 is the rotation angle around the *Y-axis* and φ_2 is the rotation angle around the *X-axis*, then

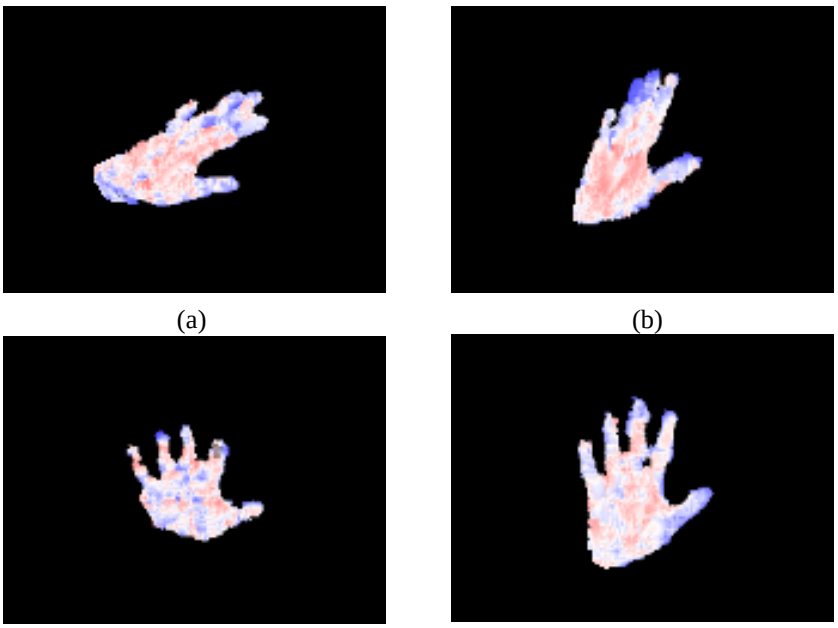
$$\varphi_1 = arctan(\frac{a}{c}) \tag{9}$$

$$\varphi_2 = arctan(\frac{b}{c}) \tag{10}$$

Given one point of the posture point cloud is $p = (px_i, py_i, pz_i)^T$, then the point $p' = (px_i', py_i', pz_i')^T$ after rotation is

$$p' = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\varphi_2 & \sin\varphi_2 \\ 0 & -\sin\varphi_2 & \cos\varphi_2 \end{bmatrix} \begin{bmatrix} \cos\varphi_1 & 0 & -\sin\varphi_1 \\ 0 & 1 & 0 \\ \sin\varphi_1 & 0 & \cos\varphi_1 \end{bmatrix} p \tag{11}$$

Fig. 4(d) indicates the normalized hand posture point cloud after rotation.



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

Fig. 4. Point cloud registration and normalization: (a) point cloud image from sensor1, (b) point cloud image from sensor2, (c) point cloud registration result, (d) normalized point cloud using (11).

Through the process above we can obtain a point cloud image, which is robust and approximate to front viewed hand posture. Moreover, the obstacle of scale, self-occlusion and rotation is relieved, which provides convincing guarantee for the performance of hand posture recognition. To determine which class the obtained point cloud falls into, Eq. (8) is utilized again to calculate the pair-wise similarity of p' (normalized point cloud image) and D_i (dataset, $i = 10$). As $\sigma^2_{p'}(D_i) = \sigma^2(p', D_1)$ reflects the difference between two matching point sets, namely, the smaller σ^2 is the higher similarity between two point sets. The discrimination criterion we set is explained as

$$C(p') = arg\,min_i[\sigma^2_{p'}(D_1), \ldots, \sigma^2_{p'}(D_i), \ldots, \sigma^2_{p'}(D_{10})] \tag{12}$$

As we utilize the accurate reconstruction and classify each point cloud by directly matching the normalized point set with the templates of different classes from dataset, which can reduce the training time and calculation.

4. Experimental Results and Analysis

The hand posture recognition system we proposed in running on the hardware environment of Intel (R) Core (TM) i5 (3.4GHz), two Kinect sensors, GTX 780 (GPU) graphic card and software environment of window 7 (64bit) and Visual Studio 2010. GPU acceleration via OpenCL is used to reduce the computation time of point cloud registration for real-time performance.

The dataset of hand posture is captured by Kinect sensor and it contains 50 multi-viewed point clouds and color information of 10 digit postures (decimal digit from 1 to 10). Five volunteers participated in the data collection and each individual performed all the 10 digits with the hand posture (See **Fig. 5**).

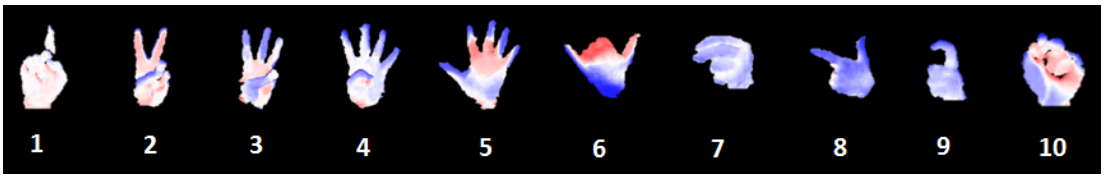


Fig. 5. Examples of hand posture database captured by Kinect sensor on the front

For obtaining 3D point clouds accurately and further processing, it is necessary to perform pre-processing firstly. In this paper, an improved bilateral filter is proposed for noise reduction and smoothing; coefficients σ_s and σ_r directly influence the performance of denoising. Based on the analysis in section 3.2, weight w_r is set as 0 when the depth difference is larger than 3mm, otherwise w_r is set as 1 and σ_s is set as 4 in our experiments.

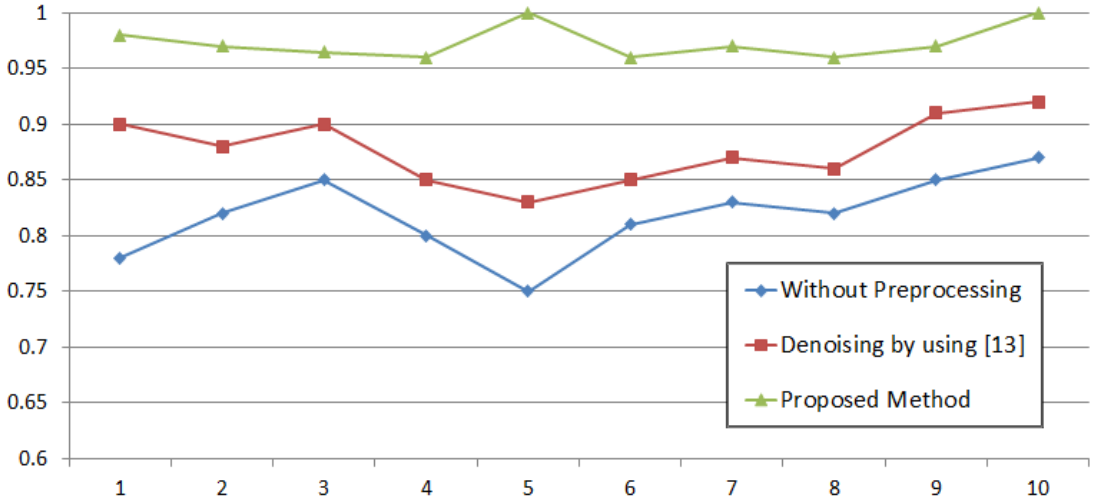


Fig. 6. Comparisons of our proposed method with the method of without preprocessing and denoising using method in [13]

Compared to original bilateral filter **[**13**]**, the improved method proposed in this paper reduces the computational time to 270ms rather than 1750ms by the former and also reduces the depth noise efficiently (See **Fig. 2**). For verifying the validity of the proposed noise reduction method, we perform the experiments by using the method without preprocessing, by

original bilateral filter and by proposed method in this paper, the comparison result is shown as **Fig. 6**.

Furthermore, we compare the proposed method in this study with the contour-matching method [15] and 2D image based HOG features [16] on hand posture dataset, the recognition accuracies of three methods based on 100 experiments are shown in **Fig. 7**.

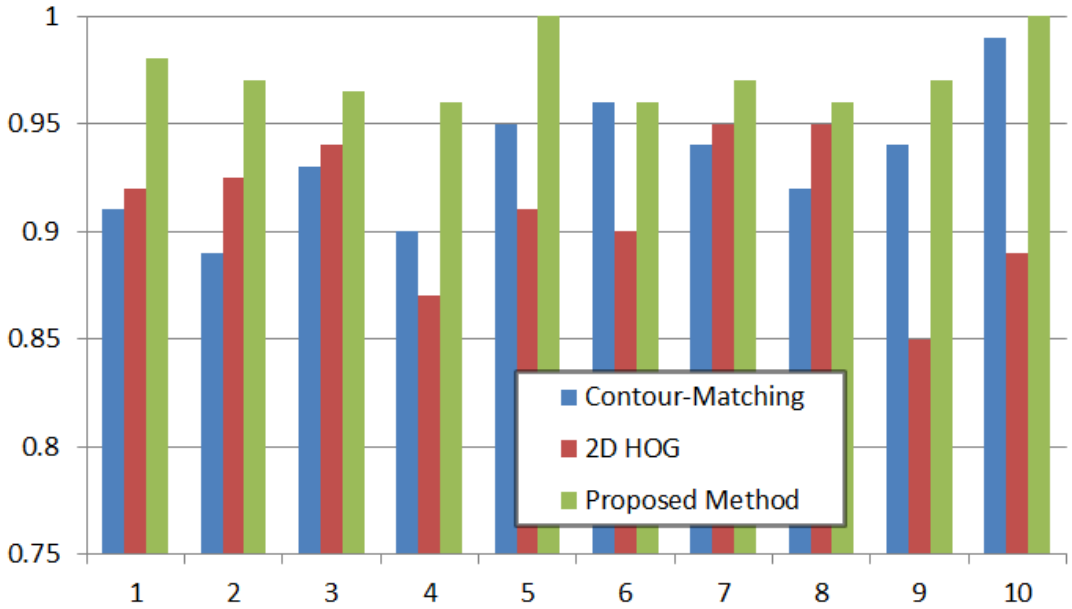


Fig. 7. Comparisons of our proposed method with the contour-matching method and conventional 2D HOG method

The hand posture recognition performance is further evaluated using confusion matrix as shown in **Table 1~3**, the classification class with the maximum score over the 10 classifiers is chosen when classifying an arbitrary posture, which indicated as Eq. (8).

From these comparisons, the proposed method considerably outperforms the contour-matching based method and 2D HOG descriptor. Our method explicitly captures the 3D surface properties such as folded thumb in palm rather than only used contour information. The average accuracy of recognition over the confusion matrix using the proposed method reaches 97.35%, which is higher than other two methods.

Table 1. Confusion matrix for contour-matching method

Class	1	2	3	4	5	6	7	8	9	10
1	91	5	1	0	0	0	0	1	0	2
2	7	89	2	0	0	0	2	0	0	0
3	0	2	93	3	0	0	0	0	2	0
4	0	0	7	90	3	0	0	0	0	0
5	0	0	0	5	95	0	0	0	0	0
6	2	0	0	0	0	96	0	2	0	0

7	0	0	0	0	0	0	94	0	2	4
8	0	5	0	0	0	3	0	92	0	0
9	2	0	0	0	0	0	0	0	94	4
10	1	0	0	0	0	0	0	0	0	99

Table 2. Confusion matrix for 2D HOG method

Class	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
1	92	2	0	0	0	0	0	0	0	6
2	5	92	1	0	0	0	0	2	0	0
3	0	4	94	2	0	0	0	0	0	0
4	0	0	7	87	6	0	0	0	0	0
5	0	0	0	8	91	1	0	0	0	0
6	2	0	0	0	0	90	0	8	0	0
7	0	0	0	0	0	0	95	0	0	5
8	3	0	0	0	0	2	0	95	0	0
9	6	0	0	0	0	0	0	4	85	5
10	1	0	0	0	0	0	4	0	6	89

Table 3. Confusion matrix for proposed method in this paper

Class	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
1	98	0	0	0	0	0	0	0	2	0
2	1	97	1	0	0	0	0	1	0	0
3	1	2	96	1	0	0	0	0	0	0
4	0	0	1	96	3	0	0	0	0	0
5	0	0	0	0	100	0	0	0	0	0
6	0	0	0	0	0	96	0	1	0	3
7	3	0	0	0	0	0	97	0	0	0
8	2	0	0	0	0	0	0	96	0	2
9	2	0	0	0	0	0	0	0	97	1
10	0	0	0	0	0	0	0	0	0	100

5. Conclusion

Hand posture recognition acts a very important role in HCI and it has appealed many efforts invested from the research field of computer vision in recent decades for its strong potential in numerous applications. However, hand gesture recognition is still a challenging work due to the wide range of poses and considerable intra-class variations, scaling, viewpoint change and hand articulations. It is unable to achieve the satisfactory effect depends on just color information.

In this paper, we proposed a hand posture recognition approach based on multi-viewed 3D point cloud, which explicitly used the sufficient 3D information conveyed by the depth maps and made users' operation more freely. Unless using 2D descriptor, sufficient 3D information can be obtained after point registration, which can get rid of the obstruction of self-occlusion and rotation. Moreover, we utilize the accurate reconstruction and classify each point cloud by directly matching the normalized point set with the templates of different classes from dataset, which can reduce the training time and calculation. Experimental results based on posture dataset captured by Kinect sensors (from digit 1 to 10) demonstrate the effectiveness of the proposed method.

Additionally, we would like to expand our hand posture recognition system in order to accommodate more challenging postures from other domains in future work.

References

[1] N. Pugeault, R. Bowden, "Spelling it out: Real-time ASL fingerspelling recognition," in *Proc. of IEEE International Conference on Computer Vision Workshops*, pp. 1114-1119, 2011. Article (CrossRef Link)

[2] J. Ravikiran, Mahesh, Kavi, Mahishi, Suhas, R. Dheeraj, S. Sudheender, Pujari, Nitin V, "Finger detection for sign language recognition," in *Proc. of International Multi Conference of Engineers & Computer Scientists*, pp. 18-20, 2009. Article (CrossRef Link)

[3] W. Xu, E. Lee, "Continuous Gesture Trajectory Recognition System based on Computer Vision," *International Journal of Applied Mathematics and Information Science*, vol. 6, no. 2S, pp.339s-346s, 2012. Article (CrossRef Link)

[4] N. Silberman, D. Hoiem, P. Kohil, R. Fergus, "Indoor Segmentation and Support Inference from RGBD Image," in *Proc. of European conference on Computer Vision*, pp. 746-760, 2012. Article (CrossRef Link)

[5] W. Xu, E. Lee, "A New NUI Method for Hand Tracking and Gesture recognition Based on User Experience," *International Journal of Security and Its Applications*, vol. 7, no. 2, pp. 148-158, 2013. Article (CrossRef Link)

[6] M. V. Bergh, L. V. Gool, "Combining RGB and ToF cameras for real-time 3D hand gesture interaction," in *Proc. of , IEEE Workshop on Application of Computer Vision*, pp. 66-72, 2011. Article (CrossRef Link)

[7] P. Garg, N. Aggarwal, S. Sofat, "Vision based hand gesture recognition," *World Academy of Science, Engineering and Technology*, vol. 49, pp. 972-977, 2009. Article (CrossRef Link)

[8] G. Murthy, R. Jadon, "A review of vision based hand gestures recognition," *International Journal of Information Technology*, vol. 2, no. 2, pp. 405-410, 2009. Article (CrossRef Link)

[9] C. Shan, T. Tan, Y. Wei, "Real-time hand tracking using a mean shift embedded particle filter," *Pattern Recognition*, vol. 40, no. 7, pp. 1958-1970, 2007. Article (CrossRef Link)

[10] G. Murthy, R. Jadon, "Hand gesture recognition using neural networks," in *Proc. of IEEE Advance Computing Conference*, pp. 134-138, 2010. Article (CrossRef Link)

[11] B. Stenger, "Template-based hand pose recognition using multiple cues," in *Proc. of 7th Asian Conference on Computer Vision*, pp. 551-560, 2006. Article (CrossRef Link)

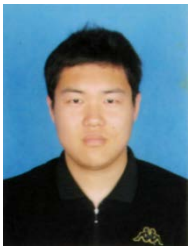
[12] E. Sanchez-Nielsen, L. Anton-Canalis, M. Hernandez-Tejera, "Hand gesture recognition for

human-machine interaction," *Journal of WSCG*, vol. 12, no. 1-3, pp. 2-6, 2003. Article (CrossRef Link)
[13] C. Tomasi, R. Manduchi, "Bilateral Filtering for Gray and Color Images," in *Proc. of the 1998 IEEE International Conference on Computer Vision*, pp. 839-846, 1998. Article (CrossRef Link)
[14] A. Myronenko, Xubo Song, "Point set registration: Coherent point drift," *Pattern Analysis and Machine Intelligence*, vol. 32, no. 12, pp. 2262-2275, 2010. Article (CrossRef Link)
[15] Z. Ren, J. Yuan, Z. Zhang, "Robust hand gesture based on finger-earth mover's distance with a commodity depth camera," in *Proc. of 19th ACM International Conference on Multimedia*, pp. 1093-1096, 2011. Article (CrossRef Link)
[16] X. Yang, C. Zhang, Y. Tian, "Recognizing Actions Using Depth Motion Maps-based Histograms of Oriented Gradients," in *Proc. of 20th ACM International Conference on Multimedia*, pp. 1057-1060, 2012. Article (CrossRef Link)
[17] L. Bretzner, I. Laptev, T. Lindeberg, "Hand gesture recognition using multi-scale color features, hierarchical models and particle filtering," in *Proc. of 5th IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 423-428, 2002. Article (CrossRef Link)
[18] C. Zhang, X. Yang, Y. Tian, "Histogram of 3D Facets: A Characteristic Descriptor for Hand Gesture Recognition," in *Proc. of 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition*, pp. 1-8, 2013. Article (CrossRef Link)
[19] W. Xu, E. Lee, "A Novel Method for Hand Posture Recognition Based on Depth Information Descriptor," *KSII TRANSACTIONS ON INTERNET AND INFORMATION SYSTEMS*, vol. 9, no. 2, pp. 763-774, 2015. Article (CrossRef Link)
[20] A. Myronenko, Xubo Song, "Point Set registration: Coherent point drift," *PAMI, IEEE Transactions on*, vol. 27, no. 12, pp. 2262-2275, 2010. Article (CrossRef Link)

Wenkai Xu received his B. S. at Dalian Polytechnic University in China (2006-2010) and Master degree at Tongmyong University in Korea (2010-2012). Currently, he is studying in Department of Information and Communications Engineering Tongmyong University for PH. D. His main research areas are image processing, computer vision, biometrics and pattern recognition.

Ick-Soo Lee received his B.S. degree in computer engineering from Dongseo University, Busan, Korea, in 2001 and the M.S. degrees in computer engineering from Kyungpook National University, Daegu, Korea, in 2003. He is currently a Ph.D candidate of department of information and communication engineering from Tongmyong University, Busan, Korea.

Suk-Hwan Lee received the B.S., M.S., and Ph.D. degrees in Electrical Engineering from Kyungpook National University, Korea in 1999, 2001, and 2004 respectively. He worked at Electronics and Telecommunications Research Institute in 2005. He is currently an Associate Professor in the Department of Information Security at Tongmyong University, which he started in 2005. He works as an Editor of Korea Multimedia Society Journal, is a member of IEEE, IEEK, IEICE and also is an officer of IEEE R10 Changwon section. His research interests include multimedia signal processing, multimedia security, digital signal processing, bio security, and computer graphics.



Bo Lu received the B. S. in School of Economics and Management, Harbin Institute of Technology (2002-2006), and M. S. and Ph. D in School of Port and Logistics Management, Tongmyong University, respectively (2006-2011). From 2011, he works in Institute of E-commercial and Modern Logistics in Dalian University as a vice professor, and works in Management College, University of Chinese Academy of Science, China as Postdoctoral. He has enrolled the Liaoning Provincial “BaiQianWan” Talents Project, the first batch of Liaoning Provincial Social Science Youth Talents, expert of Dalian Municipal Science and Technology Bureau. Special Research Fellow of China Society of Logistics and the Executive member of the council of Liaoning Province Economic and Trade Development Association. His main research interests include logistics management, port management and prediction science.

Eung-Joo Lee received his B. S., M. S. and Ph. D. in Electronic Engineering from Kyungpook National University, Korea, in 1990, 1992, and Aug. 1996, respectively. Since 1997, he has been with the Department of Information & Communications Engineering, Tongmyong University, Korea, where he is currently a professor. From 2000 to July 2002, he was a president of Digital Net Bank Inc. From 2005 to July 2006, he was a visiting professor in the Department of Computer and Information Engineering, Dalian Polytechnic University, China. His main research interests include biometrics, image processing, and computer vision.