

A Novel Method for Hand Posture Recognition Based on Depth Information Descriptor

Wenkai Xu¹, Eung-Joo Lee²

¹Department of Information Communications and Engineering, University of Tongmyong
Busan, 608-711 Korea
[e-mail: xwk6298@hotmail.com]

²Department 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 31, 2014; revised December 27, 2014; accepted January 21, 2015;
published February 28, 2015*

Abstract

Hand posture recognition has been a wide region of applications in Human Computer Interaction and Computer Vision for many years. The problem arises mainly due to the high dexterity of hand and self-occlusions created in the limited view of the camera or illumination variations. To remedy these problems, a hand posture recognition method using 3-D point cloud is proposed to explicitly utilize 3-D information from depth maps in this paper. Firstly, hand region is segmented by a set of depth threshold. Next, hand image normalization will be performed to ensure that the extracted feature descriptors are scale and rotation invariant. By robustly coding and pooling 3-D facets, the proposed descriptor can effectively represent the various hand postures. After that, SVM with Gaussian kernel function is used to address the issue of posture recognition. Experimental results based on posture dataset captured by Kinect sensor (from 1 to 10) demonstrate the effectiveness of the proposed approach and the average recognition rate of our method is over 96%.

Keywords: Hand posture recognition, 3-D information, coding facet, pooling facets, SVM

1. Introduction

Tremendous technology shift has played a dominant role in all disciplines of science and technology. During the last decades, the technology of hand gesture recognition has been a very attractive research topic of Human Computer Interaction (HCI). The main purpose of researching hand gesture recognition is to use this nature manner to implement HCI. With the rapid development of computer technology, the research on new HCI, which suits human's communication custom, has obtained the great evolvement. The hand gesture acts as one of the common communication methods in human life. So, hand gesture recognition has become the research focus. Nevertheless, because of the variety, confusing, otherness in both temporal field and spatial field uncertainty in vision and that the hand is complicated plasmodium, the research has become a multi-knowledge inter across problem filled with challenge.

A gesture is spatiotemporal pattern which maybe static, dynamic or both. Static morphs of the hands are called postures and hand movements are called gestures. Hand posture recognition from visual images has a number of potential applications in HCI, machine vision, virtual reality (VR), machine control in the industrial area, and so on. Most conventional approaches to hand posture recognition have employed data gloves [1]. But, for more natural interface, hand posture must be recognized from 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 posture recognition.

Extensive researches have been conducted on hand gesture recognition making use of 2-D digital image [2, 3, 4, 5, 6]. 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 inferred from RGB image [7]. However, because of the intrinsic vulnerability against background clutters and illumination variations, hand gesture recognition based on 2-D 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 3-D 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 3-D depth map has significant advantages for its availability to discover strong clues in boundaries and 3-D 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 [8].

The recent progress in depth sensors such as Microsoft's Kinect device has generated a new level of excitement in gestures recognition. With the depth information, a skeleton tracking system has been developed by Microsoft and hand gesture recognition based on depth maps has gained growing interests, but it does not handle hand gestures which typically involve palm and finger motions. Several researchers have proposed some approaches based on depth information for this issue [9]-[13]. Depth image generated by depth sensor is a simplified 3-D description, however most of current methods only treat depth image as an additional dimension of information and still implement the recognition process in 2-D 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 [11, 12]. Bergh and Van Gool [13] 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 important shape and structure information for complex hand gestures. Most of these methods explicitly use the sufficient 3-D information conveyed by the depth maps.

In this paper, we proposed a novel feature descriptor to explicitly encode the 3-D shape information from depth maps based on 3-D point cloud. After hand region segmentation by using depth information (distance from Kinect sensor), a 3-D local support surface associated with each 3-D cloud point is defined as a 3-D facet. By robust coding and pooling these facets, SVM with Gaussian kernel function is utilized to classify the posture from dataset. Compared with contour-matching method and 2-D HOG method, our proposed method verifies the advantage of effectiveness.

The rest of this paper is organized as follow: In Section 2, we review the previous work in hand posture recognition. Section 3 addresses the issue of hand posture recognition based on depth information descriptor. The detail of hand segmentation and normalization is presented firstly in this section, and then cell feature and pooling feature descriptor are proposed to implement the classification by SVM algorithm. Section 4 presents experimental evaluation of classification accuracy using the method proposed in this paper compared to methods of state-of-the-art. Finally, the conclusion and the future work is provided in Section 5.

2.Related Work

The human hand is a highly deformable articulated object with a total of about 27 degrees of freedom (DOFs) [14]. As a consequence, the hand can adopt a variety of static posture that can have distinct meanings in human communication. A first group of hand posture 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 [7] and [15], etc.

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 [16].

Hand detection techniques can be divided into two main groups [7]: data-glove based [1] and vision based approaches. The former uses sensors 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.

Many 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 2-D intensity images to model the hand. These images can for example be acquired by

Principal Component Analysis (PCA). Some algorithms based on appearance are presented in [4], [5], [6]. S. Gupta et al. proposed method using 15 local Gabor filters and the features are being reduced by PCA to overcome small sample size problem. Classification of the gestures as per their classes will be done with the help of one against one multiclass SVM [17].

Methods relying on high-level features use a 3-D hand model. High-level features can be derived from the joint angles and pose of the palm. Most model-based approaches create a 3-D model of a hand by defining kinematic parameters and project the 3-D model onto a 2-D space [18]. 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. P. Breuer et al. present a gesture recognition system for recognizing hand movements in near realtime in [19], the measured data is transformed into a cloud of 3-D-points after depth keying and suppression of camera noise by median filtering.

One advantage of 3-D hand model based approaches is that they allow a wide range of hand gestures if the model has enough DOFs. However, these methods also have 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 3-D 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. Hand Posture Recognition Approach Using 3-D Information

In this study, the first work is to create a point cloud using Kinect sensor for utilizing sufficient 3-D information conveyed by the depth map. There are two main reasons why we can ignore color information: first, the hand is relatively flat textured and second, we cannot assume consistent lighting conditions. In fact, our system can operate in complete darkness. While ignoring color information, we acknowledge that it also could provide useful information for solving the hand posture estimation problem.

3.1 Hand Region Segmentation Based on Depth Information

As we discussed above, the hand needs to be extracted (segmentation) from the input images before each kind of recognition algorithm can be worked out. For this, we use an effective rule based on the closest points to the camera and a region of interest around them. For this segmentation method, the assumption is made that the user's hand is the object closest to the Kinect camera.

Given the point cloud set $\Phi = \{p_1, p_2, \dots, p_n\}$ 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 the Kinect sensor. The coordinate of the closest point is written as $p_{closest} = (X, Y, Z)$. The subset $\Phi' = \{hp_1, hp_2, \dots, hp_n\}$ will be searched in Φ for hand area, where the following conditions hold.

$$p_{closest}.X \leq hp_n.X \leq p_{closest}.X + 0.15m \quad (1)$$

$$p_{closest}.Y \leq hp_n.Y \leq p_{closest}.Y + 0.15m \quad (2)$$

$$p_{closest}.Z \leq hp_n.Z \leq p_{closest}.Z + 0.2m \quad (3)$$

The subset Φ' is guaranteed to be contained in a cuboid 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. Assuming the segmentation process completes successfully, the subset contains points pertaining to the hand and part of the forearm. The hand segmentation result is shown as Fig. 1.

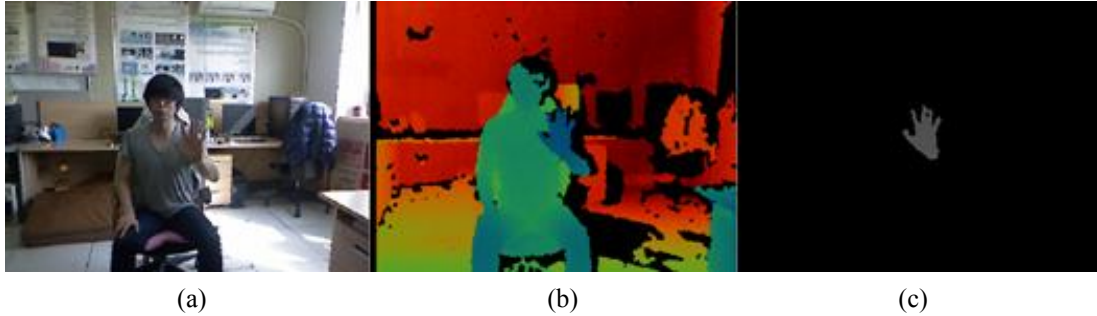


Fig. 1. Hand segmentation from depth map: (a) RGB color image, (b) depth image, (c) hand segmentation results.

3.2 Hand Image Normalization

As the problem of hand self-occlusion, in-plane rotation is a challenge to hand posture recognition, scale and orientation normalization will be performed in this step so that the extracted feature descriptors are scale and rotation invariant. For orientation normalization, we first estimate the hand orientation parameters, and then rotate the hand point cloud in such way that the palm plane will be parallel to the image plane and the hand will point upward.

The normalization algorithm consists of three steps as described below:

- Scale normalization: The hand region on P is scaled to fit into a predefined rectangle with size of $150*150$.
- Out-Plane normalization: Fit a plane P to the hand point cloud, and compute a rotation that will rotate P to be parallel to the image plane. This step is very useful when the visible surface of the hand is approximately planar. If not, such normalization could lead to an overstretched image with holes, and we do not perform out-plane normalization in this case. See Fig. 2(a) and 2(b).
- In-Plane normalization: We project all the points onto P and compute the principal direction. We then compute an in-plane rotation matrix so that the principal direction points upward after rotation. See Fig. 2 (b) and 2 (c).

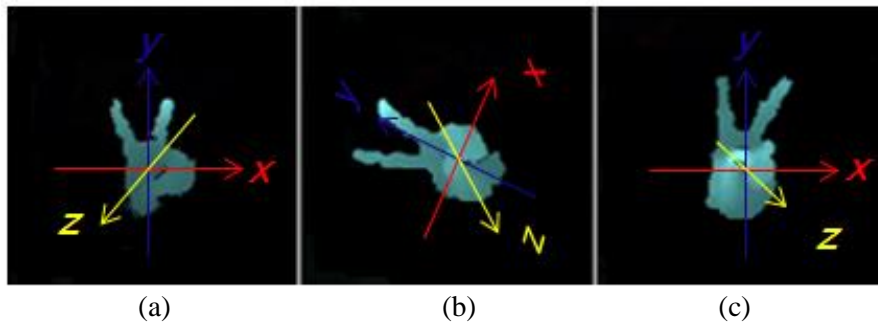


Fig. 2. Orientation and scale normalization: (a) out-plane rotation image, (b) in-plane rotation image, (c) normalization image.

Given one point of the posture point cloud is $P = (px_i, py_i, pz_i)^T$, the transformation matrix is written as TM , then the point $P' = (px_i', py_i', pz_i')^T$ after normalization is

$$P' = P * TM \quad (4)$$

Here, TM is written as matrix form as

$$TM = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos(\alpha) & -\sin(\alpha) & 0 \\ 0 & \sin(\alpha) & -\cos(\alpha) & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} \cos(\beta) & 0 & \sin(\beta) & 0 \\ 0 & 1 & 0 & 0 \\ -\sin(\beta) & 0 & \cos(\beta) & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} \cos(\gamma) & -\sin(\gamma) & 0 & 0 \\ \sin(\gamma) & \cos(\gamma) & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Where α, β, γ represents the rotation angle around the X -axis, Y -axis, Z -axis, respectively.

After the normalization procedure, we obtain the rotation parameters and a depth map of the normalized hand point cloud. This image is called “Hand-Image”, which along with the rotation parameters, will be used at the feature generation stage.

3.3 Extraction of Cell Feature and Pooling Feature

The 3-D surface properties, such as bumps and grooves provide significant information, especially when the outer contour is not sufficient or discriminative to perform classification. As shown in the first two pictures in Fig. 3, the 3-D surfaces of the thumb constitute of an informative region to differentiate the two hand gestures that share similar visual patterns.

A 3-D facet defined as a 3-D plane which can be represented by $[n_x, n_y, n_z, d]^T$, where the first three coefficients are the normal vector $n = [n_x, n_y, n_z]^T$ of a 3-D facet and the forth one d is the Euclidean distance from the plane to the origin coordinate. Although all four coefficients are needed to determine a local surface, in this paper we only concentrate on the distance rather than the absolute orientation of a local surface. Thus, we code a 3-D facet only using its relative distance d . The procedure of coding each 3-D facet is illustrated in Fig. 3, we call it “Cell feature”.

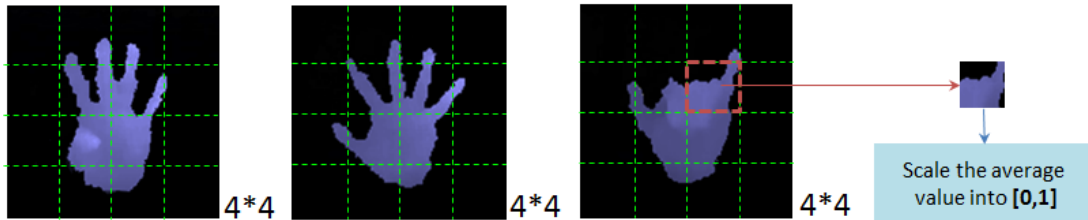


Fig. 3. 4*4 Cell feature

For cell feature, we compute the occupied area of each cell as well as the average depth for the non-empty cells, and then we scale the average value into $[0, 1]$ by

$$D_{ave} = \frac{(D_{(x,y)} - D_{min})}{(D_{max} - D_{min})} \quad (5)$$

Where D_{ave} means the average of depth value in one cell, D_x denotes the depth value at (x, y) , D_{max} and D_{min} represents the maximum and minimum depth value in one cell, respectively.

A cloud point of 3-D depth map can be mapped onto a 2-D depth image. Therefore, each cloud point corresponds to a pixel in 2-D depth image, which means the pixels on 2-D depth image originate from those 3-D points that locate in the front surface. Therefore, we can completely assert that the d attribute of all the occupancy features are non-negative. Proposed cell feature can alleviate the problem of similar contour since we consider the interior region of the folded thumb, which makes the depth average value of each cell more discriminative.

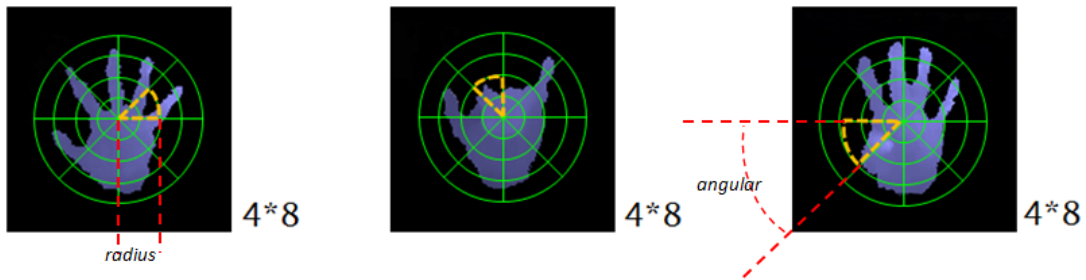


Fig. 4. The concentric spatial pooling scheme (The entire hand region is quantized as 4bins in radius and 8 bins in angular).

After coding the 3-D facets of the depth map, a concentric spatial pooling scheme [20] is used to group these coded 3-D facets from the entire hand posture region to generate the feature descriptor, as illustrated in Fig. 4. In the step of concentric spatial pooling, we divide the normalized hand region into 32 spatial bins, namely, four radius quantization bins and 8 angular quantization bins. We also compute the average depth value for each bin as the feature descriptor is therefore with the dimension of $(4*4)*(4*8) = 16*32 = 512$.

3.4 Hand Posture Recognition Based on Support Vector Machine (SVM)

After these descriptors are generated as above, the Support Vector machines (SVM) with Radial Basis Function (RBF) kernel is used as the classifier in this study. SVM is a supervised learning technique for optimal modeling of data. It learns decision function and separates data class to the maximum width. SVM learner defines the hyper-plane for the data and maximum margin is found between these hyper-planes. Because of the maximum separation of hyper-plane, it is also considered as a margin classifier. Margin of the hyper-plane is the minimum distance between hyper-plane and the support vectors and this margin is maximized in it [17, 21].

SVM is a well-suited classifier where features are large in number because they are robust to the curse of dimensionality. Kernel function is the computation of the inner product $\Phi(x) \cdot \Phi(y)$ directly from the input. One of the characteristics of using the kernel is that there is no need to explicitly represent the mapped feature space. For optimizing classification issue in this paper, we need to identify the best parameters for the SVM. In Table 1, we list three different SVM kernels with various parameter settings and each SVM is tested on our hand posture database. From Table 1, Gaussian Kernel Function with parameter $\sigma = 3$ performs the best performance (the average accuracy is over 96%) for hand posture recognition. Therefore, SVM using Gaussian Kernel Function with $\sigma = 3$ is leveraged to classify the posture descriptor in this paper, which obtained at last section.

Table 1. Experiment results using 3 kernel types with different parameters

Kernel Type	Degree	σ	Accuracy of trial number (1-10) (%)									
			1	2	3	4	5	6	7	8	9	10
Linear	—	—	91	90	92	89	90	92	93	93	91	94
Polynomial	2	—	92	93	91	92	93	94	95	95	91	93
Polynomial	3	—	93	92	93	90	94	95	94	92	93	92
Polynomial	4	—	94	93	92	94	91	92	93	90	94	91
Gaussian	—	1	93	92	90	92	95	95	90	94	93	95
Gaussian	—	2	95	92	94	93	95	95	92	93	94	95
Gaussian	—	3	95	94	96	96	98	97	97	94	96	98
Gaussian	—	4	94	91	95	92	93	94	95	92	94	93
Gaussian	—	5	93	91	93	94	90	91	92	91	93	92

4. Experimental Classification Results and Analysis

The hand posture recognition system we designed is running on the hardware environment of Intel (R) Core (TM) i5 (3.40 GHz), a Kinect sensor, and the software environment of Windows 7 and Visual Studio 2010.

The dataset for training and testing is captured by a Microsoft sensor and it contains 1000 depth maps of hand gestures (decimal digit from 1 to 10) from 10 subjects with 10 samples for each hand gesture. We use recordings of nine persons for training and the remaining person for testing. Such experiment is repeated for 10 random subdivisions of the data. Before computing the descriptor, we normalize the hand region into an image patch with the fixed size of 150×150 to reduce the information loss in quantization.



Fig. 5. Hand posture database used in this study, which captured by Kinect sensor (the first row are the depth images and the second row are the corresponding color images)

For verifying the performance of Cell+Pooling feature, we perform comparison experiments with cell feature, pooling feature and the mixture of these two features, respectively. The comparison result is indicated in **Fig. 6**, X-axis expresses the categories of 10 postures and Y-axis represents the recognition rate. From **Fig. 6**, the hybrid descriptor that

combines the cell feature and pooling feature is superior to individual ones. Because of Cell+Pooling feature explicitly utilizes the sufficient 3-D information conveyed by the depth maps than the other two, thereby differentiates each hand posture effectively.

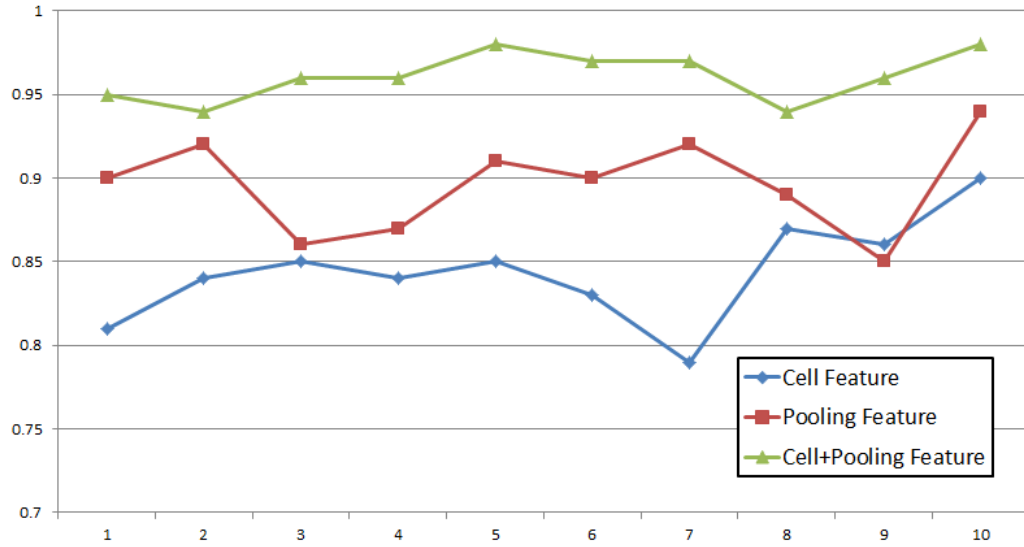


Fig. 6. Comparisons of Cell Feature, Pooling Feature and Cell+Pooling Feature

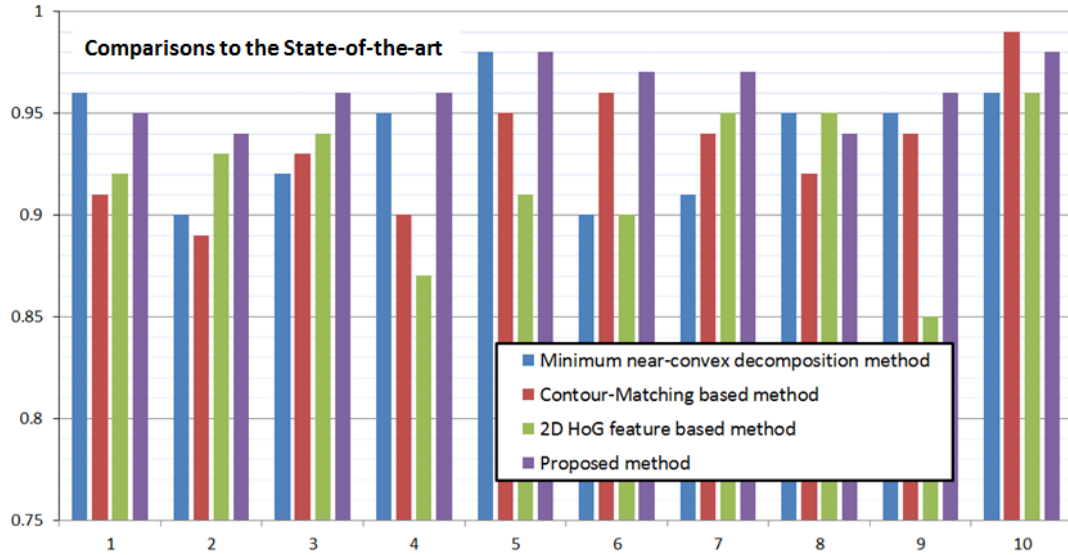


Fig. 7. Comparisons of our proposed method with other three methods presented previously

We compare the proposed method in this paper with the Minimum near-convex decomposition method [11], contour-matching method [20] and conventional 2-D image based HOG [22] on our hand posture dataset. Methods proposed in [11, 20] only considered the outer contour of fingers but ignored the palm region that also provides important shape and structure information for complex hand posture. Moreover, in the implementation of HOG, we evenly separate the normalized patches into 8*8 non-overlapping cells and each cell has eight orientation bins. The feature vectors of four different normalizations, namely,

L1-norm, L2-norm, L1-sqrt and L2-Hys are concatenated as the final HOG representation as [23]. The HOG is descriptor is therefore with the dimension of $8*8*8*4=2048$. The posture recognition accuracies of three methods are shown in Fig.7.

From these comparisons, the proposed method considerably outperforms the contour matching based method [11, 20] and the traditional 2-D HOG descriptor [22]. Compare to the contour-matching method in [20], our method explicitly captures the 3-D surface properties such as folded thumb in palm rather than only outer contour information. Meanwhile, compared to the HOG, the feature descriptor is with the dimension of 512 in proposed method rather than 2048. Therefore, our method greatly decreases computation complexity and decoding time.

The performance of hand posture recognition is further evaluated using a confusion matrix as shown in Fig. 8. The classification class with the maximum score over the 10 classifiers was chosen when classifying an arbitrary posture (Max-Win rule). This metric always results in a single classification (correct or incorrect), and no false positive cases. If the maximum score points to the incorrect class, then we said that the posture was misclassified. The confusion matrix (see Fig. 8) was created by comparing the scores obtained by each classifier applied to a given testing image, and selecting the maximum score from all 10 classifiers. The average accuracy of correct classification over the confusion matrix using the proposed method in this paper reaches 96.1%, which is higher than the use of the other method mentioned above.

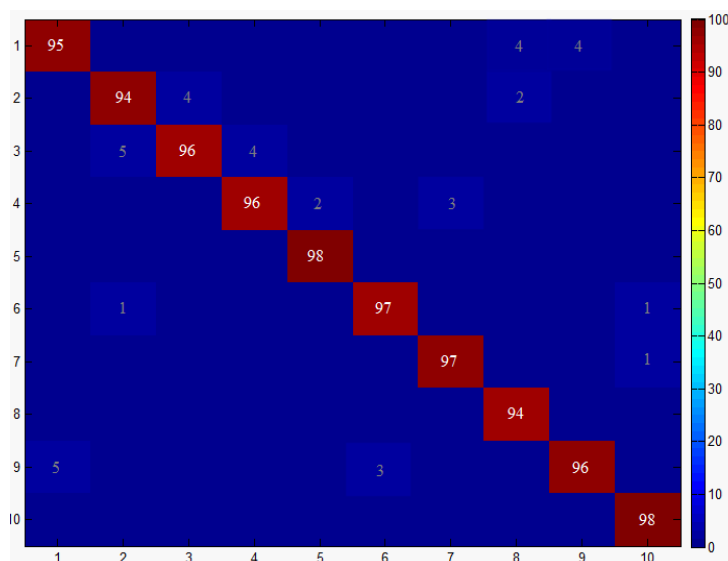


Fig. 8. Confusion matrix computed when using the method proposed in this paper

In general, proposed method performs 2.3% and 2.0% higher accuracy than [11] and [20] on our posture database, and 4.3% higher than 2-D HOG descriptor. The confusion matrix indicates the accuracies of classification are all over 96%, which demonstrates the effectiveness of our method.

5. Conclusion

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 and illumination variations. In this

paper, a normalized 3-D descriptor conveyed by the depth maps is used to explicitly capture and model discriminative surface information. Unless using color information (RGB), depth information obtained from IR sensor is not affected by the condition of illumination, the system proposed in this study can work well even in darkness.

We implement the hand posture recognition by using SVM with linear kernel function after coding the facets and spatial pooling. Compared to contour-matching method and 2-D HOG method, the experimental results verified the effectiveness of the proposed approach based on using the sufficient 3-D information, which including the shape and structure.

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] Brandon Paulson, Danielle Cummings and Tracy Hammond, "Object interaction detection using hand posture cues in an office setting," *International Journal of Human-Computer Studies*, vol. 69, issues.1-2, pp. 19-29, 2011. [Article \(CrossRef Link\)](#).
- [2] N. Pugeault and R. Bowden, "Spelling it out: Real-time ASL fingerspelling recognition," in *Proc. of IEEE International Conference on Computer Vision Workshops (ICCVW)*, pp. 1114-1119, 2011. [Article \(CrossRef Link\)](#).
- [3] Ravikiran J., Kavi Mahesh, SuhasMahishi , Dheeraj R., Sudheender S., Nitin V.Pujari, "Finger detection for sign language recognition," in *Proc. of International MultiConference of Engineers & Computer Scientists*, pp. 18-20, January 2009. [Article \(CrossRef Link\)](#).
- [4] G. Murthy and R. Jadon, "Hand gesture recognition using neural networks," in *Proc. of IEEE Advance Computing Conference*, pp. 134-138, Feb. 19-20, 2010. [Article \(CrossRef Link\)](#).
- [5] B.Stenger, "Template-based hand pose recognition using multiple cues," in the *Proc. of 7th Asian Conference on Computer Vision*, pp. 551-560, January 13-16, 2006. [Article \(CrossRef Link\)](#).
- [6] E. Sanchez-Nielsen, L. Anton-Canalis and M. Hernandez-Tejera, "Hand gesture recognition for human-machine interaction," *Journal of WSCG*, vol. 12, no. 1-3, pp. 2-6, 2004. [Article \(CrossRef Link\)](#).
- [7] WenkaiXu and EungJoo Lee, "Continuous Gesture Trajectory Recognition System based on Computer Vision," *International Journal of Applied Mathematics & Information Science*, vol.6, no.2S, pp.339s-346s, 2012. [Article \(CrossRef Link\)](#).
- [8] N. Silberman, D. Hoiem, P. Kohil and R. Fergus, "Indoor Segmentation and Support Inference from RGBD Image," in *Proc. of 12th European Conference on Computer Vision*, pp. 746-760, 2012. [Article \(CrossRef Link\)](#).
- [9] X. Liu and K. Fujimura, "Hand Gesture Recognition using Depth Data," in *Proc. of International Conference on Automatic Face and Gesture Recognition*, pp. 529-534, 2004. [Article \(CrossRef Link\)](#).
- [10] Myung-CheolRoh, Ho-Keun Shin and Seong-Whan Lee, "View-independent human action recognition with Volume Motion template on single stereo camera," *Patter Recognition Letter*, vol. 31, no. 7, pp. 639-647, May 2010. [Article \(CrossRef Link\)](#).
- [11] Z. Ren, J. Yuan and Z. Zhang, "Robust hand gesture recognition 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 \(CrossRef Link\)](#).
- [12] Z. Ren, J. Yuan, C. Li and W. Liu, "Minimum near-convex decomposition for robust shape representation," in *Proc. of IEEE International Conference on Computer Vision*, pp. 303-310, 2011. [Article \(CrossRef Link\)](#).
- [13] M. Van den Bergh and L. Van Gool, "Combining RGB and ToF cameras for real-time 3-D hand gesture interaction," in *Proc. of IEEE Workshop on Applications of Computer Vision*, pp. 66-72, 2011. [Article \(CrossRef Link\)](#).
- [14] P. Garg, N. Aggarwal and S. Sofat, "Vision based hand gesture recognition," *World Academy of Science, Engineering and Technology*, vol. 3, pp. 972-977, January 28, 2009.

- [Article \(CrossRef Link\)](#).
- [15] C. Shan, T. Tan and Y. Wei, "Real-time hand tracking using a mean shift embedded particle filter," *Pattern Recognition*, vol. 40, no. 7, pp. 1958-1970, 2007. [Article \(CrossRef Link\)](#).
- [16] G. Murthy and R. Jadon, "A review of vision based hand gestures recognition," in *Proc. of International Journal of Information Technology*, vol. 2, no. 2, pp. 405-410, 2009. [Article \(CrossRef Link\)](#).
- [17] Shikha Gupta, JafreezalJaafar, Wan Fatimah Wan Ahmad, "Static Hand Gesture Recognition Using Local Gabor Filter," *Procedia Engineering*, vol. 41, pp. 827-832, 2012. [Article \(CrossRef Link\)](#).
- [18] Y. Li, "Hand gesture recognition using Kinect," in *Proc. of IEEE 3rd International Conference on Software Engineering and Service Science*, pp. 196-199, June 22-24, 2012. [Article \(CrossRef Link\)](#).
- [19] P. Breuer, C. Eckes and S. Muller, "Hand Gesture Recognition with a novel IR Time-of-Flight Range Camera-A pilot study," in *Proc. of 3rd International Conference of MIRAGE*, France, pp. 247-260, March 28-30, 2007. [Article \(CrossRef Link\)](#).
- [20] Z. Ren, J. Yuan and Z. Zhang, "Robust gesture based on finger-earth mover's distance with a commodity depth camera," in *Proc. of the 19th ACM International Conference on Multimedia*, pp. 1093-1096, 2011. [Article \(CrossRef Link\)](#).
- [21] C.J. C. Bruges, "A Tutorial on Support Vector Machines for Pattern Recognition," *Data Mining and Knowledge Discovery*, vol. 2, Issue 2, pp. 121-167, 1998. [Article \(CrossRef Link\)](#).
- [22] Hui Li, "Static Hand Gesture Recognition Based on HOG with Kinect," in *Proc. of 2012 4th International Conference on Intelligent Human-Machine Systems and Cybernetics*, Nanchang, China, pp. 271-273, 26-27 Aug. 2012. [Article \(CrossRef Link\)](#).
- [23] X. Yang, C. Zhang and Y. Tian, "Recognizing Actions Using Depth Motion Maps-based Histograms of Oriented Gradients," in *Proc. of International Conference on ACM Multimedia*, pp. 1057-1060, 2012. [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.



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.