

A Statistical-Topological Feature Combination for Recognition of Isolated Hand Gestures from Kinect Based Depth Images

Soumi Paul, Hayat Nasser, Mita Nasipuri, Phuc Ngo, Subhadip Basu, Isabelle
Debled-Rennesson

► **To cite this version:**

Soumi Paul, Hayat Nasser, Mita Nasipuri, Phuc Ngo, Subhadip Basu, et al.. A Statistical-Topological Feature Combination for Recognition of Isolated Hand Gestures from Kinect Based Depth Images. IWCI, Jun 2017, Plovdiv, Bulgaria. 10256, LNCS. <10.1007/978-3-319-59108-7>. <hal-01517817>

HAL Id: hal-01517817

<https://hal.inria.fr/hal-01517817>

Submitted on 3 May 2017

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

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

A Statistical-Topological Feature Combination for Recognition of Isolated Hand Gestures from Kinect Based Depth Images

Soumi Paul¹, Hayat Nasser^{2,3}, Mita Nasipuri¹, Phuc Ngo^{2,3}, Subhadip Basu¹,
Isabelle Debled-Rennesson^{2,3}

¹ Department of Computer Science & Engineering,
Jadavpur University, Kolkata 700032, India

`soumip@research.jdvu.ac.in, {subhadip, mnasipuri}@cse.jdvu.ac.in`

² Université de Lorraine, LORIA, UMR 7503, Vandoeuvre-lès-Nancy, F-54506, France

³ CNRS, LORIA, UMR 7503, Vandoeuvre-lès-Nancy, F-54506, France

`{hayat.nasser, hoai-diem-phuc.ngo, isabelle.debled-rennesson}@loria.fr`

Abstract. Reliable hand gesture recognition is an important problem for automatic sign language recognition for the people with hearing and speech disabilities. In this paper, we create a new benchmark database of multi-oriented, isolated ASL numeric images using recently launched Kinect V2. Further, we design an effective statistical-topological feature combinations for recognition of the hand gestures using the available V1 sensor dataset and also over the new V2 dataset. For V1, our best accuracy is 98.4% which is comparable with the best one reported so far and for V2 we achieve an accuracy of 92.2% which is first of its kind.

Keywords: Hand Gesture Recognition, Sign Language, Kinect, Statistical-Topological Features, Depth Data, Discrete Curve, Polygonal Simplification

1 Introduction

Hand gesture recognition is of great importance due to its potential applications in contactless human-computer interaction (HCI). In particular, reliable hand gesture recognition is crucial for many applications, including automatic sign language recognition for the HCI of hearing and speech impaired persons. Some of these techniques require wearing of an electronic glove [5] so that the key features of hand can be accurately measured, but the device is somewhat costly and inconvenient for domestic applications. Another class of methods uses optical markers [8] instead of electronic gloves but it requires rather complex configuration. On the other hand, affordable depth-based systems are coming up with promising results in the field of depth-based hand gesture recognition. Microsoft Kinect [15] is one such RGB-D sensor providing synchronized color and depth images. It was launched as a gaming device, but computer vision research community has taken interest into it and extended it for a lower cost replacement for

traditional 3D cameras, such as stereo cameras and time-of-flight (TOF) cameras. In just two years after Kinect V1 was released, a large number of scientific papers with technical demonstrations have started appearing in diverse publication venues. Recently, the new version of the Kinect sensor V2 has been launched with more accurate depth sensing technology. The work presented in this paper involves both V1 and V2 sensors for recognition of isolated depth images of ASL.

Literature Survey and Our Contributions. Many vision based hand gesture recognition algorithms have been proposed in the past years and comprehensive reviews can be found in [9]. Methods based on skin color model [21] and hand shape model [20] have also been proposed. However, they are not robust in the dynamic environment and rely significantly on the models. The recent development of depth cameras, such as Microsoft Kinect [15], Creative Senz3D or Mesa Swiss-Ranger etc., opens up new avenues for hand gesture recognition. Therefore, how the depth information can be efficiently utilized and how the depth camera can be incorporated in the hand gesture recognition system is an active topic of research [17]. In early studies, hand detection mainly relies on vision-based features which was sensitive to variations of skin colors and lighting. On the other hand, depth camera offers a much simpler way of isolating hands by depth thresholding. After the hand localization and segmentation, various hand features can be extracted from either the depth maps, e.g. Histogram of 3D Facets (H3DF) [22], or the corresponding color images such as Histogram of Oriented Gradients (HOG) [3], which will then be used for hand gesture recognition.

We have already explored that depth based image recognition has more advantage over vision based systems. Within depth based systems, recent trend is to use low cost device like Kinect V1 to input images, which gives a color image of 640 x 480 resolution and depth map of 320 x 240 at 30 FPS. Whereas recently launched V2 has better RGB resolution 1920 X 1080 and depth resolution 512 x 424 at 30 FPS. Not only that, the field of view (i.e., the solid angle through which the detector is sensitive to electromagnetic radiation) has also been expanded, skeleton joint point has also been upgraded from 20 to 26 and most importantly, with USB 3.0, the speed has been increased to get more support for real time applications. A detailed comparison of Kinect V1 and V2 can be found in [16].

With the use of depth data we can detect hand gestures robustly in the cluttered background independent of lighting conditions. So the objective of our proposed work is to, 1) create a new challenging benchmark of multi-oriented, isolated ASL numeric image dataset using recently launched Kinect V2, 2) design an effective statistical-topological feature combinations for recognition of the hand gestures using the available V1 sensor dataset and also the new V2 dataset.

2 Tools to Study Discrete Contours

In this section, we recall a method of contour simplification based on selected dominant points. They are computed thanks to a discrete structure, named

adaptive tangential cover (ATC) [11], well adapted to analyse irregular noisy contours.

2.1 Adaptive Tangential Cover [11]

An **adaptive tangential cover (ATC)** is composed of a sequence of maximal straight segments, called maximal blurred segments, of the studied contour. The notion of maximal blurred segment has been introduced in [4] as an extension of arithmetical discrete line [14] with a width parameter for noisy or disconnected digital contours.

Definition 1. An **arithmetical discrete line** $\mathcal{D}(a, b, \mu, \omega)$, with a main vector (b, a) , a lower bound μ and an arithmetic thickness ω (with $a, b, \mu, \omega \in \mathbb{Z}$ and $\text{gcd}(a, b) = 1$) is the set of integer points (x, y) verifying $\mu \leq ax - by < \mu + \omega$.

Definition 2. A set S_f is a **blurred segment of width ν** if the discrete line $\mathcal{D}(a, b, \mu, \omega)$ containing S_f has the vertical (or horizontal) distance $d = \frac{\omega - 1}{\max(|a|, |b|)}$ equal to the vertical (or horizontal) thickness of the convex hull of S_f , and $d \leq \nu$ (see Fig. 1).

Let C be a discrete curve and $C_{i,j}$ a sequence of points of C indexed from i to j . Let denote the predicate “ $C_{i,j}$ is a blurred segment of width ν ” as $BS(i, j, \nu)$.

Definition 3. $C_{i,j}$ is called a **maximal blurred segment (MBS) of width ν** and denoted $MBS(i, j, \nu)$ iff $BS(i, j, \nu)$, $\neg BS(i, j + 1, \nu)$ and $\neg BS(i - 1, j, \nu)$ (see Fig. 1).

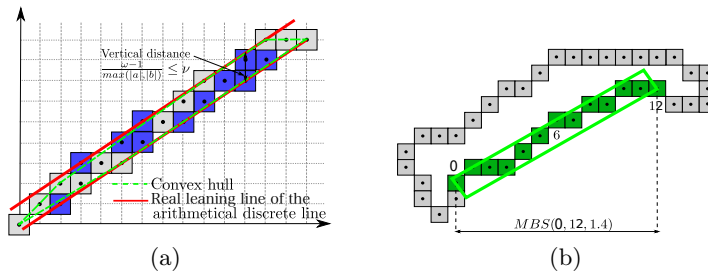


Fig. 1. (a) Example of arithmetical discrete line $\mathcal{D}(2, -3, -5, 5)$ (grey and blue points) and a blurred segment of width $\nu = 1.4$ (grey points) bounded by D . (b) Maximal blurred segment of width $\nu = 1.4$ (green points) - color figure online.

An *ATC* consists of MBS's of different widths, which are a function of the noise perturbations of the studied contour. In particular, we use the local noisy estimator, namely **meaningful thickness** [6, 7], to determine the significant

width locally at each point of the contour. This meaningful thickness is used as an input parameter to compute the *ATC* with appropriate widths w.r.t. noise. A non-parametric algorithm is developed in [11] to compute the *ATC* of a given discrete curve. In the *ATC*, the obtained *MBS* decomposition of various widths transmits the noise levels and the geometrical structure of the given discrete curve (see Fig. 2(a,c)).

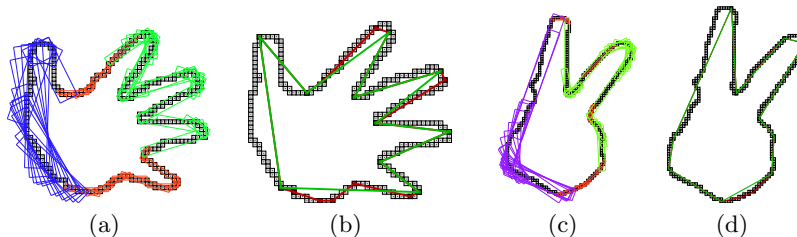


Fig. 2. (a,c): adaptive tangential cover , (b,d): polygonal representation (in red) using the dominant points and polygonal simplification results (in green) - color figure online.

2.2 Polygonal Simplification [10–12]

Dominant points are significant points on a curve with local maximum curvature. Such points contain a rich information which allows to characterize and describe the curve. Issued from the dominant point detection proposed in [10, 12] and the notion of *ATC*, an algorithm is developed in [11] to determine the dominant points of a given noisy curve C . The main idea is that the candidate dominant points are localized in the common zones of successive *MBS* of the *ATC* of C . An angle measure m is used to determine the dominant points with local extreme curvature in the common zones. More precisely, this measure m is the angle between the considered point and the two left and right endpoints of the left and right *MBS* involved in the studied common zone. When the considered point varies, m becomes a function of it. A dominant point is defined as a local minimum of m . Dominant points are illustrated in Fig. 2 (b,d) in red lines. Red lines represent the polygonal representation of the shape.

First goal of finding the dominant points is to have an approximate description of the input curve, called **polygonal simplification**. Dominant points are sometimes redundant or stay very near, which is presumably undesirable in particular for polygonal simplification. So, we associate to each detected dominant point a **weight**, i.e, the ratio of integral sum of square errors and the angle with the two dominant point neighbours, indicating its importance with respect to the approximating polygon of the curve. Polygonal simplification is illustrated in Fig. 2 (b,d) with green lines.

3 Feature Descriptors

In this work, we are using a combination of topological, statistical and geometric features. In the following, we describe each feature descriptor in details.

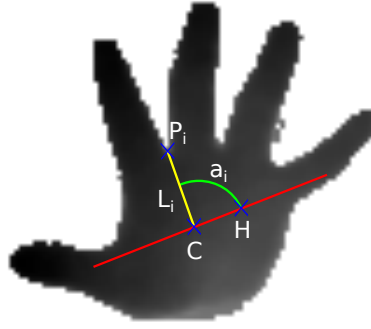


Fig. 3. Histogram of Contour Angles and Contour Distances

3.1 Histogram of Contour Angles (HoCa) and Contour Distances (HoCd)

Suppose there are m contour points $(X_1, Y_1), \dots, (X_m, Y_m)$ and n convex hull points $(X'_1, Y'_1), \dots, (X'_n, Y'_n)$.

The image moments with pixel intensity $f(x, y)$ at location (x, y) are given by

$$M_{ij} = \sum_{x,y} x^i y^j f(x, y). \quad (1)$$

For a contour or a set of points without any associated intensity, $f(x, y)$ is taken to be 1.

In our case, let M_{ij}^c denote the moments of the contours and let M_{ij}^h be the moments of the convex hull points. Then the center $C = (X_c, Y_c)$ of the contour points is given by

$$X_c = \frac{M_{10}^c}{M_{00}^c}, Y_c = \frac{M_{01}^c}{M_{00}^c}, \quad (2)$$

and the center $H = (X'_c, Y'_c)$ of the convex hull is given by

$$X'_c = \frac{M_{10}^h}{M_{00}^h}, Y'_c = \frac{M_{01}^h}{M_{00}^h}. \quad (3)$$

Then we fix the line segment L_{CH} between the points C and H . Let L_i be the line segment joining C to the contour point $P_i = (X_i, Y_i)$ and let a_i be the angle between L_{CH} and L_i formed at the point C , for $i = 1, \dots, m$.

Now we create a histogram with 10 bins from these m angles a_1, \dots, a_m and get 10 descriptors say h_1, \dots, h_{10} .

Let $|L_i|$ be the length of the line segment L_i , $i = 1, \dots, m$. We create a histogram with 10 bins from these m lengths and thus get 10 descriptors l_1, \dots, l_{10} .

In Fig. 3 we explain the above set of descriptors.

3.2 Moments

We use different kinds of moments as follows.

Raw or spatial moments We use Equation (1) on the entire image, where (x, y) denotes a pixel location and $f(x, y)$ denotes the corresponding greyscale intensity value. From this, we generate 10 descriptors M_{ij} , with $0 \leq i + j \leq 3$. More explicitly, we use $M_{00}, M_{01}, M_{10}, \dots, M_{03}, M_{30}$.

Central moments First we calculate the spatial moments as above. Then we define

$$\bar{x} = \frac{M_{10}}{M_{00}} \text{ and } \bar{y} = \frac{M_{01}}{M_{00}}. \quad (4)$$

Now, the central moments are given by

$$\mu_{ij} = \sum_{x,y} (x - \bar{x})^i (y - \bar{y})^j f(x, y), \quad (5)$$

where (x, y) denotes a pixel location and $f(x, y)$ denotes the corresponding greyscale intensity value. From this, we generate 7 descriptors μ_{ij} , with $2 \leq i + j \leq 3$. More explicitly, we use $\mu_{11}, \dots, \mu_{03}, \mu_{30}$.

Central standardized or normalized or scale invariant moments These moments are normalized versions of the central moments, defined as follows:

$$\nu_{ij} = \frac{\mu_{ij}}{(M_{00})^{\frac{i+j}{2} + 1}}. \quad (6)$$

From this, we generate 7 descriptors $\nu_{11}, \nu_{12}, \nu_{02}, \nu_{20}, \nu_{21}, \nu_{03}, \nu_{30}$.

3.3 Geometric Descriptors from Polygonal Simplification of Shape Contours

We propose different descriptors from the selected dominant points (DP), obtained with the method presented in Sec.3.3, applied on a shape S . The polygon obtained with the DP is called $DP(S)$. Let $per(DP(S))$ and $area(DP(S))$ be

respectively the perimeter and the area of $DP(S)$. Let $ch(DP(S))$ be the convex hull of $DP(S)$. The following descriptors give indications about compactness and convexity of the shape S :

$$\frac{per(DP(S))^2}{area(DP(S))} \quad (7)$$

$$\frac{per(ch(DP(S)))^2}{area(ch(DP(S)))} \quad (8)$$

$$\frac{area(ch(DP(S)))}{area(DP(S))} \quad (9)$$

We compute descriptors that indicate if the contour of the shape S is regular or contains big irregularities :

- mean value of angle between two successive segments formed by three successive dominant points of $DP(S)$
- variance value of angle between two successive segments formed by three successive dominant points of $DP(S)$
- minimum distances from the centroid of S to the dominant points of S
- maximum distances from the centroid of S to the dominant points of S
- the difference between minimum and maximum distances
- variance of segment lengths in $DP(S)$
- number of peaks detected in S

We should notice that the centroid of S is the centroid of the polygonal simplification of the shape S . All our geometric descriptors are based on the dominant points selected by the polygonal simplification process. We do not work with all points of the contour S .

Moreover we detect the number of “peaks” in each shape contour.

A peak is a dominant point located in a convex part of the shape contour with an angle greater than a given threshold. We fix this threshold angle to 1,38. Fig.4 shows that in this convex part, point P_{i+1} is considered as a peak point because its angle is greater than the threshold and angle at point P_{i+2} is less than the threshold.

4 Dataset and Pre-processing for Experimental Evaluation

This section describes the collection of dataset and pre-processing that we performed to extract strong features.

4.1 Dataset from Kinect V1

As a benchmark dataset for Kinect V1, we are using the NTU dataset [13]. This dataset is collected from 10 subjects and it contains 10 gestures for numbers 0 to 9. Each subject performs 10 different poses for the same gesture. Thus this database has 1000 cases in total and it consists of a color image and the corresponding depth map.

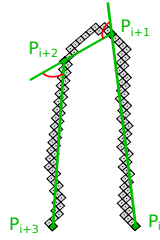


Fig. 4. Convex part of a shape contour shows the angle at each dominant points ($P_i, P_{i+1}, P_{i+2}, P_{i+3}$) to determine peak points.

4.2 Development of New Benchmark Dataset from Kinect V2

We are proposing a new hand gesture dataset using Kinect V2. In this dataset we have collected standard ASL hand gestures from 10 subjects for numbers 0 to 9. Each subject performs 10 different poses for the same gesture. Thus in total the dataset has $10 \text{ people} \times 10 \text{ gestures/people} \times 10 \text{ cases/gesture} = 1000$ cases, each of which consists of a contour map and the corresponding depth map. This dataset is a real-life dataset, which is collected in cluttered backgrounds. Besides, for each gesture, the subject poses with variations in hand orientation (almost 180 degree variation), scale, articulation, etc.

4.3 Data Collection

For NTU dataset collected from Kinect V1, we have Annotated the ROI region in RGB images. Then after depth thresholding, we have cropped the exact ROI regions from depth images by comparing it with the RGB images. For V1 dataset, it was not a problem because depth resolution is 320×240 and RGB resolution is 640×480 which is just the double. Whereas, in our proposed database collected from Kinect V2, RGB resolution is 1920×1080 and Depth resolution is 512×424 . So in our database, we collected a joint color-depth hand gesture along with wrist joint point. This gives us the flexibility of locating the ROI in the whole frame. Gesture samples are shown in Fig. 5 which are labeled from 0 to 9. It should be noted that this dataset is a real life dataset collected in uncontrolled environment with different illumination and different orientations.

4.4 Hand Localization and Segmentation

In previous depth camera-based approaches [17], the hand is required to be the front-most object from the depth camera. Moreover, a black belt on the gesturing hands wrist is also required in some cases [13], which is rather inconvenient for real world applications. In our system, we relax these restrictions by utilizing the rather stable joints from Kinect's skeleton tracking. The Kinect joints are directly used to locate the hands, wrists and elbows. By assuming that the hand

is visible to the camera without any occlusion, it allows us to quickly separate the hands from background objects using depth information alone. Using the hand joint point as the center, a pair of color texture and depth map blocks is extracted first and then the hand shape is segmented quickly using a depth threshold value.

4.5 Noise Removal

In practical applications, the extracted hand gestures usually have different scales due to various distances from the camera to hand, or different rotations caused by the body postures. Moreover, different people’s hands always have distinct characteristics even for the same gesture. Hence it is necessary to perform some pre-processing to normalize and align the shape representation before recognition. However, the palm size varies from one person to another, which affects the recognition between different subjects. So instead of depth thresholding in our dataset, We have done histogram thresholding calculating mean and standard deviation of the depth values and extracted gray scale cropped hand shapes, shown in Fig. 5. In Fig. 5, it also can be seen that the hand shapes from different persons are correctly segmented, even when the hands are cluttered by the face or background.

5 Performance Evaluation

Now we evaluate the performance of the proposed system from mean accuracy, time efficiency and comparisons with other methods. We are using Random Forest classifier on extracted features using the Weka version 3.6.12 (c) 1999-2014 [19] machine learning tool. We performed all experiments on an Intel i7-6500U CPU @2.50 GHz processor with 8 GB RAM and 64-bit Windows 8.1 Pro OS.

In our experiments, leave- p -out (LpO) cross-validation (CV) is conducted to evaluate the recognition performance, where with M instances, $p < M$ subjects are used for testing and the remaining for training. This process is repeated for every combination of subjects so that the average accuracy can be computed. For NTU dataset, we only calculated LOO CV to compare the results. In our dataset, two values of (1 and 100) are considered, which are respectively referred to as leave-one-out CV (LOO CV) and 10-fold CV. Experiments based on these two CVs are presented in next section.

5.1 Results and Comparison

Here we produce comparative results (see Table 1) between our proposed work and other different algorithms on standard V1 NTU hand digit dataset. Our mean accuracy is 98.4% which outperforms all other works mentioned in the table. Fig. 5 shows our proposed benchmark V2 dataset containing ASL gestures

Table 1. Comparison between Mean Accuracy of Shape Contexts, Skeleton Matching, HOG, H3DF, FEMD and our methods on the NTU HAND DIGIT DATASET

Algorithms	Mean Accuracy
Skeleton Matching[1]	78.6%
Shape Context with bending cost[2]	79.1%
Shape Context without bending cost[2]	83.2%
HOG[3]	93.1%
Thresholding Decomposition+FEMD[13]	93.2%
Near-convex Decomposition+FEMD[13]	93.9%
H3DF[22]	95.5%
Current Work	98.4%

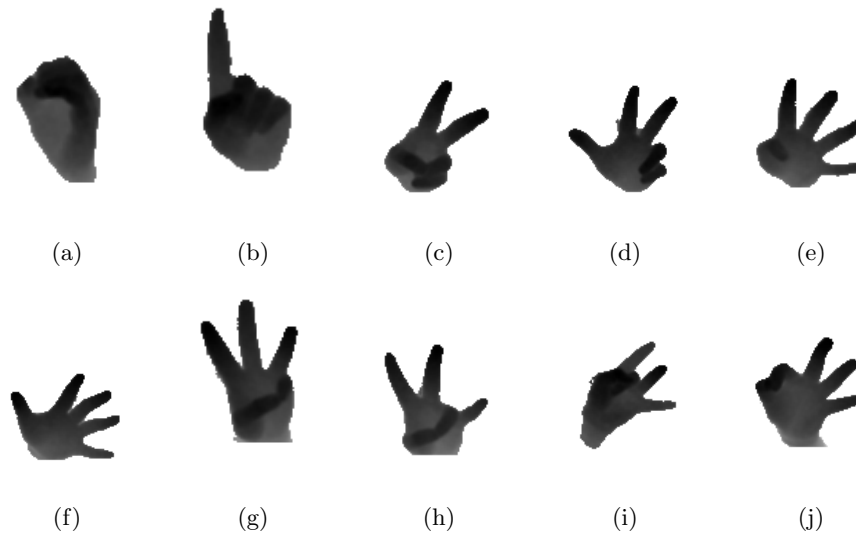


Fig. 5. Gesture samples (0-9) of ASL captured in Kinect V2

0 to 9 captured using Kinect V2. On this dataset, we get an accuracy of 92.2% with the same feature set as V1.

Fig. 6 illustrates the confusion matrix of hand gesture recognition on standard V1 dataset and our benchmark V2 dataset (Fig. 5) using our feature set presented in Sec. 4.

6 Conclusion

In this work, a hand gesture recognition system from depth data using topological-statistical features for contactless HCI has been proposed. It is based on an efficient feature extraction and classification which recognizes gestures based on the shape and depth features. The effectiveness of the system is illustrated by

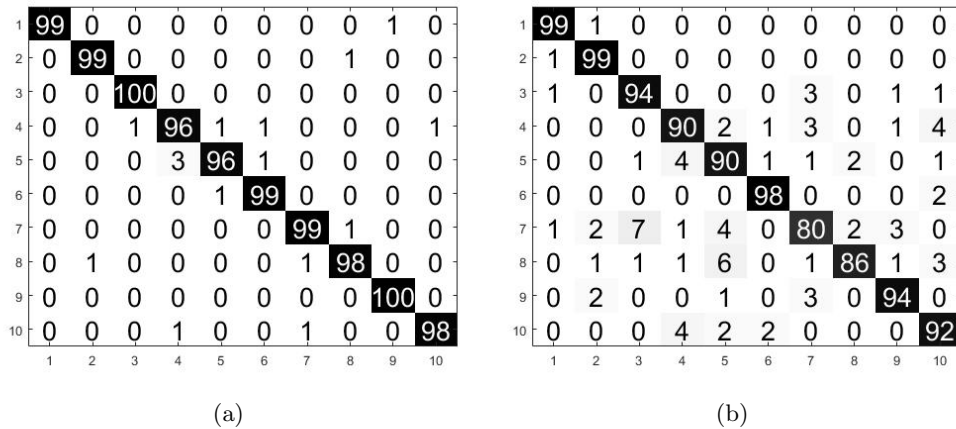


Fig. 6. The confusion matrices of hand gesture recognition using with our feature set: (a) LOO CV on NTU hand digit dataset, (b) 10-fold CV on our own V2 benchmark dataset

extensive experiments on two real-life datasets, NTU hand digit dataset [13] and our own V2 dataset (achieving high mean accuracies 98.4%, 92% respectively). In comparison with previous distance measures such as FEMD [13], shape context[2], Skeleton Matching[1], H3DF [22], HOG [3], our feature set achieves comparable performance for hand gesture recognition. Recently, another work of SPEMD [18] has been proposed with 99.1% (LOO CV) accuracy on their own dataset. We are still to compare our features with their dataset. Our proposed descriptors are computationally efficient and thus suitable for fast gesture recognition. As this work has only been done using depth map, so our future research will focus on exploring robust color features and extending it to dynamic hand gesture, body posture and generic object recognition.

References

1. Bai, X., Latecki, L.J.: Path similarity skeleton graph matching. *IEEE transactions on pattern analysis and machine intelligence* 30(7), 1282–1292 (2008)
2. Belongie, S., Malik, J., Puzicha, J.: Shape matching and object recognition using shape contexts. *IEEE transactions on pattern analysis and machine intelligence* 24(4), 509–522 (2002)
3. Dalal, N., Triggs, B., Schmid, C.: Human detection using oriented histograms of flow and appearance. In: *European conference on computer vision*. pp. 428–441. Springer (2006)
4. Debled-Rennesson, I., Feschet, F., Rouyer-Degli, J.: Optimal blurred segments decomposition of noisy shapes in linear time. *Computers & Graphics* 30(1), 30–36 (2006)

5. Dewaele, G., Devernay, F., Horaud, R.: Hand motion from 3d point trajectories and a smooth surface model. In: European Conference on Computer Vision. pp. 495–507. Springer (2004)
6. Kerautret, B., Lachaud, J.O.: Meaningful Scales Detection: an Unsupervised Noise Detection Algorithm for Digital Contours. *Image Processing On Line* 4, 98–115 (2014)
7. Kerautret, B., Lachaud, J.O., Said, M.: Meaningful Thickness Detection on Polygonal Curve. In: Proceedings of the 1st International Conference on Pattern Recognition Applications and Methods. pp. 372–379. SciTePress (2012)
8. Kry, P.G., Pai, D.K.: Interaction capture and synthesis. In: ACM Transactions on Graphics (TOG). vol. 25, pp. 872–880. ACM (2006)
9. Mitra, S., Acharya, T.: Gesture recognition: A survey. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* 37(3), 311–324 (2007)
10. Ngo, P., Nasser, H., Debled-Rennesson, I.: Efficient dominant point detection based on discrete curve structure. In: International Workshop on Combinatorial Image Analysis IWCIA), Kolkata, India, November. LNCS, vol. 9448 (2015)
11. Ngo, P., Nasser, H., Debled-Rennesson, I., Kerautret, B.: Adaptive tangential cover for noisy digital contours. In: Discrete Geometry for Computer Imagery - 19th IAPR International Conference, DGCI 2016, Nantes, France. LNCS, vol. 9647, pp. 439–451 (2016)
12. Nguyen, T.P., Debled-Rennesson, I.: A discrete geometry approach for dominant point detection. *Pattern Recognition* 44(1), 32–44 (2011)
13. Ren, Z., Yuan, J., Meng, J., Zhang, Z.: Robust part-based hand gesture recognition using kinect sensor. *IEEE transactions on multimedia* 15(5), 1110–1120 (2013)
14. Reveillès, J.P.: Géométrie discrète, calculs en nombre entiers et algorithmique (1991), thèse d'état. Université Louis Pasteur, Strasbourg
15. Shotton, J., Sharp, T., Kipman, A., Fitzgibbon, A., Finocchio, M., Blake, A., Cook, M., Moore, R.: Real-time human pose recognition in parts from single depth images. *Communications of the ACM* 56(1), 116–124 (2013)
16. Soumi Paul, S.B., Nasipurit, M.: Microsoft kinect in gesture recognition: A short review. *International Journal of Control Theory and Applications* 8(5), 2071–2076 (2015)
17. Suarez, J., Murphy, R.R.: Hand gesture recognition with depth images: A review. In: RO-MAN, 2012 IEEE. pp. 411–417. IEEE (2012)
18. Wang, C., Liu, Z., Chan, S.C.: Superpixel-based hand gesture recognition with kinect depth camera. *IEEE Transactions on Multimedia* 17(1), 29–39 (2015)
19. WEKA: Fibonacci notes (1996), <http://www.cs.waikato.ac.nz/ml/weka/downloading.html>
20. Wu, Y., Lin, J., Huang, T.S.: Analyzing and capturing articulated hand motion in image sequences. *IEEE transactions on pattern analysis and machine intelligence* 27(12), 1910–1922 (2005)
21. Yang, M.H., Ahuja, N., Tabb, M.: Extraction of 2d motion trajectories and its application to hand gesture recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 24(8), 1061–1074 (2002)
22. Zhang, C., Yang, X., Tian, Y.: Histogram of 3d facets: A characteristic descriptor for hand gesture recognition. In: Automatic Face and Gesture Recognition (FG), 2013 10th IEEE International Conference and Workshops on. pp. 1–8. IEEE (2013)