Hand Number Gesture Recognition Using the Recognized Hand Parts in Depth Images

Dong-Luong Dinh[a], Hyun Jae Jeon[b], Sang Beom Nam[b], Sungyoung Lee[a], and Tae-Seong Kim[b,*]

[a]Departement of Computer Engineering, Kyung Hee University,

1 Seocheon-dong, Giheung-gu, Yongin-si, Gyeonggi-do, Republic of Korea

E-mail address: {luongdd, sylee} @oslab.khu.ac.kr

[b]Departement of Biomedical Engineering, Kyung Hee University,

1 Seocheon-dong, Giheung-gu, Yongin-si, Gyeonggi-do, Republic of Korea

E-mail address: tskim@khu.ac.kr

Abstract

In this paper, we present a novel approach of recognizing hand number gestures using the recognized hand parts in depth images. Our proposed approach is divided into two stages: (i) hand parts recognition by random forests (RFs) and (ii) rule-based hand number gesture recognition. In the first stage, we create a database of synthetic hand depth silhouettes and their corresponding hand parts-labeled maps and train RFs with the database (DB). Then with the trained RFs, we recognize or label hand parts in depth silhouettes. In the second stage, based on the information of the recognized or labeled hand parts, hand number gestures are recognized according to our derived rules of features. In our experiments, we quantitatively and qualitatively evaluated our hand parts recognition system with synthetic and real data. Then, we tested our hand number gesture recognition system with real data. Our results show the average recognition rate of 97.80 % over the ten hand number gestures from five subjects.

Keyword: Depth Images, Recognized Hand Parts, Hand Number Gestures, Gesture Recognition System

1. Introduction

Among various gesture recognition methodologies using bodies, faces, or hands, hand-based gesture or sign recognition is one of efficient and natural ways for human machine interfaces. Potential applications of hand gesture or sign recognition include human computer interfaces, automatic sign language interpretations [1], remote controls of electronic appliances [2, 3], and smart robotics. One of key functions in these applications is to recognize hand poses, movements, or both which convey meaningful information of users' intention to give instructions to a machine. Among hand gesture recognition of signs such as letters, numbers, or symbols, hand number gesture recognition is an important task to interface between humans and machines. Although various approaches have been proposed for hand number gesture recognition, they can be divided into two commonly used approaches: one is inertial sensor glove-based and the other vision-based approaches [4].

The sensor glove-based approach uses an array of sensors which are attached to a hand glove that transduces fingers and hand actions into electrical signals to recognize the hand number gestures. In [5], a sensor glove was used to process and convey the degree of flexion in each finger of the hand with a tri-axis accelerometer placed on the back of the hand to provide its orientation information. A neural network method was employed to recognize hand number gestures. In [6, 7], six accelerometers on a hand glove were used to get relative angles between fingers and the palm of the hand. Each number gesture was recognized by a fuzzy rule-based classification system. Another method proposed in [8] introduced a 3D input device with inertial sensors to recognize number drawing gestures in 3D. The signals of angular velocity and acceleration were considered as motion features. Fisher discriminant analysis was used for hand number gesture recognition. Although, these methods reported some success in hand number gesture recognition, these studies require of wearing gloves and positioning sensor on the glove or hand. Therefore, motion sensor- or glove-based recognition approaches of hand gestures, in general, are not considered as a natural and convenient way in spite of high sensitivity [9].

Vision-based approaches uses imaging sensors or cameras to get gesture features in colors, shapes, orientations, texture, contours, motions, or distances [4, 11]. Most methodologies for color image based-hand number gesture recognition have relied on the basic motion or skin color information. In [12], the authors proposed a method to recognize hand number gestures of Persian sign language from 1 to 10 using the skeleton information of hand silhouettes. The endpoints of the skeleton were extracted as fingertips in recognizing ten gestures. In [13], the authors presented a recognition methodology for nine hand sign gestures using Hidden Markov Models with the orientation and contour features extracted from hand silhouettes. Another method proposed in [14, 15] used a color glove in which different colors were imprinted on five fingers of the hand glove. Then finger recognition was performed based on color information to classify hand number gestures. However, these studies exhibited limited success in hand number gesture recognition since the color and shape features are not always reliable and color images are generally sensitive to noise, lighting conditions, cluttered backgrounds, and especially occlusions.

Recently, with the introduction of new depth imaging sensors, some studies have been focusing on using depth features for hand gesture recognition [16]. In [17], a method was presented for recognizing hand number gestures by template matching using Chamfer Distance, measuring shape similarity based on the orientation, size, shape, and position features from depth images. In [18], the hand gesture recognition combined two types of hand pose features, namely, distance and curvature features by computing the hand contours. Finally, support vector machine was employed to recognize hand sign gestures. In these approaches, the same or similar feature of color images are extracted from depth images, thereby resulting in marginally improved recognition rate. In another proposed method [19], the finger parts were detected by hand shape decomposition and hand number gesture recognition was done via template matching by minimizing dissimilarity distance using their proposed Finger-Earth Mover's distance metric. However, this approach suffered from the ambiguity in the length of fingers, resulting in reduced recognition rate. In general, these previous depth image-based approaches suffer from no information of finger parts, in spite of advantages of depth images such as low sensitivity to lighting conditions.

In this study, we have developed a novel approach of recognizing hand number gestures by recognizing or labeling hand parts in depth images. Our proposed approach consists of two main processes: hand parts recognition by random forests (RFs) classifier and rule-based hand number gesture recognition. The main advantage of our proposed approach is that the state of each finger gets directly identified through the recognize hand parts and then number gestures are recognized based on the state of each finger. We have tested and validated our approach on synthetic and real data. Our experimental tests achieved 97.8 % in recognition of ten hand number gestures with five subjects. Our hand number gesture recognition system should be useful in human and machine interaction applications.

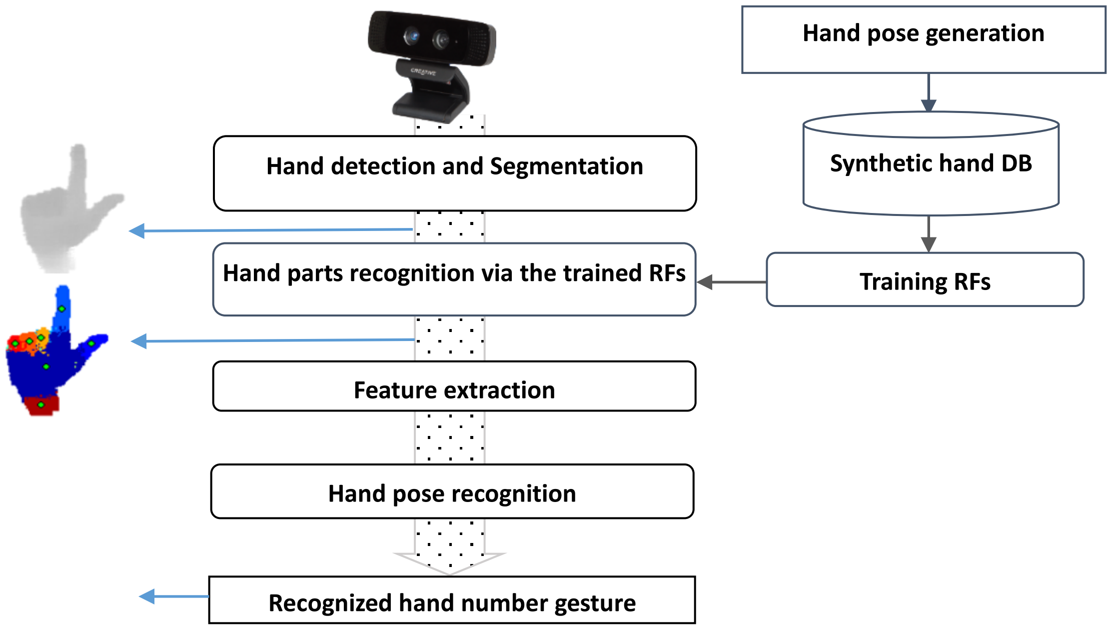


Fig. 1 The framework of our hand number gesture recognition system.

2. Methodology

Overall process of our proposed system for hand number gesture recognition, shown in Fig. 1, consists of two main parts: in the first part, a synthetic hand database (DB), which contains pairs of depth maps and their corresponding hand parts-labeled maps, was generated. Then, the DB was used to training RFs. In the recognition stage, a depth image was first captured from a depth camera and then a hand depth silhouette was extracted by removing the background. Next, the hand parts of a depth silhouette were recognized using the trained RFs. Next, a set of features was extracted from the labeled hand parts. Finally, based on the extracted features, hand number gestures were recognized by our rule-based approach.

2.1 Synthetic hand DB generation

To recognize hand parts from a hand depth silhouette via RFs, the synthetic hand DB, which contains pairs of depth images and their corresponding hand part-labeled maps, is needed to train RFs. We created such the DB with a synthetic hand model using 3Ds-max, a commercial 3D graphic package [21]. To identify hand parts, twelve labels were assigned to each hand model as shown in Fig. 3. The five fingers including the thumb, index, middle, ring, and pinky fingers, were represented by ten corresponding labels including five front and five back sides. The front parts were coded with the index numbers of 2, 3, 4, 5 and 6. Likewise, the five back sides were coded with the index numbers of 7, 8, 9, 10 and 11, respectively. This is critical in our recognition system to identify the open and close state of each finger. In addition, the palm and wrist parts were given the index number of 1 and 12. Our DB comprised 5,000 pairs covering the ten different hand number gestures as shown in Fig. 2. Among them, a set of 3,000 pairs was used in training and a set of the remaining pairs was used in testing. A set of 500 pairs in the DB represents each hand number gesture captured under many different view angles. A set of samples of 3-D hand model and the map of the corresponding labeled hand parts are shown in Fig. 3. Images in the DB had a size of 240 x 320 with 16-bit depth values.

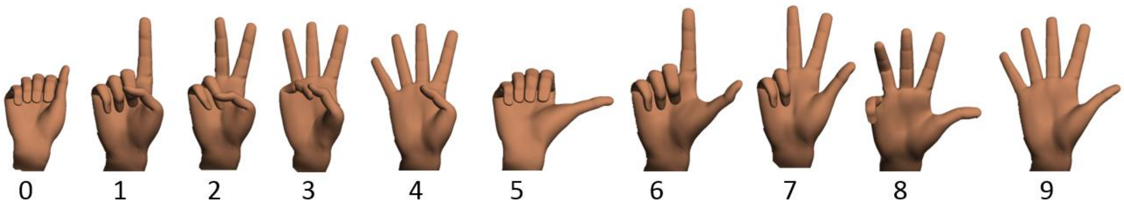
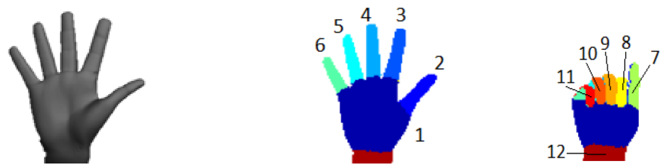


Fig. 2. Hand number gestures representing numbers from 0 to 9.



(a) (b)

Fig. 3 Hand model: (a) 3D hand model in 3Ds-max and (b) Hand parts with twelve labels

2.2 RFs for pixel-based classification

In training, we used an ensemble of 21 decision trees. The maximum depth of trees was 20. Each tree in RFs was trained with different pixels sampled randomly from the DB. A subset of 500 training sample pixels was drawn randomly from each synthetic depth silhouette. A sample pixel was extracted as done in [20], to obtain 800 candidate features. At each splitting node in the tree, a subset of 28 candidate features was considered. For pixel classification, each pixel *p* of a tested depth silhouette was extracted to obtain the candidate features. For each tree, starting from the root node, if the value of the splitting function was less than a threshold of the node, *p* went to left and otherwise *p* went to right. The optimal threshold for splitting the node was determined by maximizing the information gain in the training process. The probability distribution over 12 hand parts was computed at the leaf nodes in each tree. Final decision to label each depth pixel for a specific hand part was based on the voting result of all trees in the RFs.

2.3 Hand parts recognition

To recognize hand parts of each hand depth silhouette, all pixels of each hand depth silhouette were classified by the trained RFs to assign a corresponding label out of the 12 indices. A centroid point was withdrawn from each recognized hand part and represents each hand part.

2.4 Feature extraction and rules for hand number gesture recognition

From the recognized hand parts, we extracted a set of features. In our labeling, each finger was represented by two different labels: one label for its front side corresponding to the open state of the finger and another for its back side corresponding to the close state of the finger. From the information of the recognized hand parts, we identify the open or close states of each finger. The states of the five labeled fingers were identified and saved as features, namely f_{Thumb}, f_{Index}, f_{Middle}, f_{Ring}, and f_{Pinky} respectively.

$$f_{Fingers}(i) = \begin{cases} 1: Open & for\,Label = 2,3,4,5, or\ 6 \\ 0: Close & for\ Label = 7,8,9,10, or\ 11 \end{cases} \tag{1}$$

Table 1: Recognition rules of the number signs based on the states of the five fingers

f_{Thumb}	f_{Index}	f_{Middle}	f_{Ring}	f_{Pinky}	**Number representation (ID)**
0	0	0	0	0	0
0	1	0	0	0	1
0	1	1	0	0	2
0	1	1	1	0	3
0	1	1	1	1	4
1	0	0	0	0	5
1	1	0	0	1	6
1	1	1	0	0	7
1	1	1	1	0	8
1	1	1	1	1	9

0=Close and 1= Open.

To recognize hand number gestures, we have derived a set of recognition rules. The set of five features from the states of all fingers is used to encode the meaning of the ten hand number poses. The derived recognition rules are given in Table 1.

3. Results

3.1 Results of hand parts recognition

To evaluate our hand parts recognition quantitatively, we tested on a set of 2,000 hand depth silhouettes containing various number poses over the ten hand number poses. The average recognition rate of the hand parts was 96.60 %.

Then, we assessed the hand parts recognition on real data qualitatively. Since the ground truth labels are not available, we only performed visual inspection on the recognized hand parts. A representative set of the recognized hand parts are shown in Fig. 4.

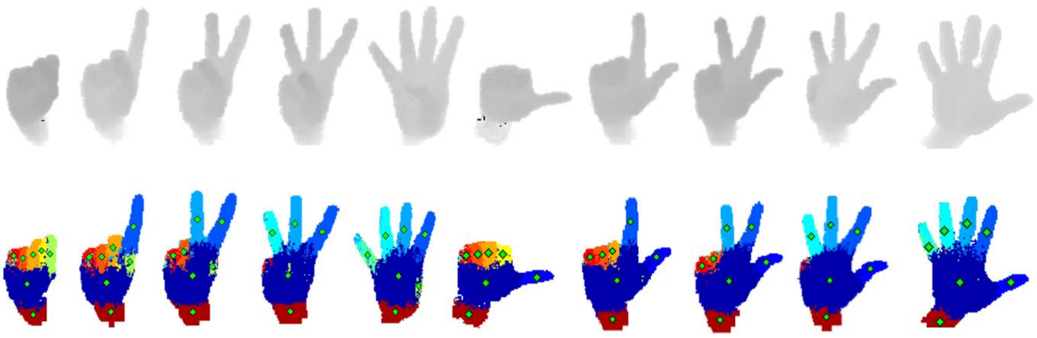


Fig. 4. A set of representative results of the recognized hand parts via the trained RFs on real data.

3.2 Results of hand number sign gesture recognition

Table 2: The confusion matrix of number gestures recognition using our proposed system

ID	0	1	2	3	4	5	6	7	8	9
0	100									
1	2	98								
2			100							
3				90	10					
4					100					
5				2		98				
6							98	2		
7								98	2	
8									100	
9					2			2		96

To test our proposed hand sign gesture recognition methodology, a set of hand depth silhouettes was acquired from five different subjects. Each subject was asked to make 100 hand number poses. Table 2 shows the recognition results of the ten hand number gestures in a form of confusion matrix. The mean recognition rate of 97.80 % was achieved.

4. Conclusion

In this paper, we have presented a novel hand number gesture recognition system with the labeled hand parts via the trained RFs from a hand depth silhouette. We have achieved the mean recognition rate of 97.80 % over the ten hand number gestures from five subjects. Our presented work should be applicable for recognizing sign language gestures.

Acknowledgments

This research was supported by the MSIP (Ministry of Science, ICT & Future Planning), Korea, under the ITRC (Information Technology Research Center) support program supervised by the NIPA (National IT Industry Promotion Agency (NIPA-2013-(H0301-13-2001)).

References

[1] M. Elmezain, A. Al-Hamadi, S. S. Pathan, and B. Michaelis, “Spatio-temporal feature extraction-based hand gesture recognition for isolated American Sign Language and Arabic numbers,” In proc. of International Symposium on Image and Signal Processing and Analysis, pp. 254–259, 2009.

[2] C. H. Wu and C. H. Lin, “Depth-based hand gesture recognition for home appliance control,” In proc. of IEEE 17th International Symposium on Consumer Electronics, pp. 279–280, 2013.

[3] A. Shimada, T. Yamashita, and R. Taniguchi, “Hand gesture based TV control system — Towards both user- & machine-friendly gesture applications,” In proc. of 19th Korea-Japan Joint Workshop on Frontiers of Computer Vision (FCV), pp. 121–126, 2013.

[4] G. R. S. Murthy and R. S. Jadon, “A review of vision based hand gestures recognition,” Int. Journal of Information Technology and Knowledge Management, vol. 2, no. 2, pp. 405–410, 2009.

[5] M. E. Cabrera, J. M. Bogado, L. Fermin, R. Acuna, and D. Ralev, “Glove-based Gesture Recognition System,” In proc. of International Conference on Climbing and Walking Robots and the Support Technologies for Mobile Machines, pp. 747–753, 2012.

[6] T. D. Bui, and L. T. Nguyen, “Recognizing Postures in Vietnamese Sign Language With MEMS Accelerometers,” IEEE Sensors Journal, vol. 7, no. 5, pp. 707–712, 2007.

[7] Y. Tabata, T. Kuroda, and K. Okamoto, “Development of a glove-type input device with the minimum number of sensors for Japanese finger spelling,” In proc. of Int. Conf. Disability, Virtual Reality & Associated Technologies, pp. 305–310, 2012.

[8] J. K. Oh, S. J. Cho, W. C. Bang, W. Chang, E. S. Choi, J. Yang, J. K. Cho, and D. Y. Kim, "Inertial sensor based recognition of 3-D character gestures with an ensemble classifiers," Int. Workshop on Frontiers in Handwriting Recognition, pp. 112–117, 2004.

[9] L. Dipietro, A. M. Sabatini, and P. Dario, "A Survey of Glove-Based Systems and Their Applications," IEEE Trans. on System, Man, and Cybernetics–PART C: Application and reviews, vol. 38, no. 4, pp. 461–482, 2008.

[10] X. Ruize, Z. Shengli, W. J. Li, "MEMS accelerometer based nonspecific-user hand gesture recognition," Sensors Journal, IEEE, vol. 12, no. 5, pp. 1166–1173, 2012.

[11] H. S. Hasan and S. A. Kareem, "Human-computer interaction using vision-based hand gesture recognition system: a survey," Neural computing and application, 2013.

[12] A. Barkoky, and N. M. Charkari, "Static hand gesture recognition of Persian sign numbers using thinning method," In proc. of Int. Conf. on Multimedia Technology (ICMT), pp. 6548–6551, 2011.

[13] R. L. Vieriu, B. Goras, and L. Goras, "On HMM static hand gesture recognition," In proc of 10th International Symposium on Signals, Circuits and Systems (ISSCS), 2011.

[14] B. Swapna, F. Pravin, and V. D. Rajiv, "Hand Gesture Recognition System for Numbers Using Thresholding," Communications in Computer and Information Science, vol. 250, pp. 782–786, 2011.

[15] L. Lamberti and F. Camastra, "Handy: A real-time three color glove-based gesture recognizer with learning vector quantization," Expert Systems with Applications, vol. 39, pp. 10489–10494, 2012.

[16] J. Suarez and R. R. Murphy, "Hand gesture recognition with depth images: A review," In proc. of the 21st IEEE International Symposium on Robot and Human Interactive Communication, pp. 411–417, 2012.

[17] X. Liu and K. Fujimura, "Hand gesture recognition using depth data," In proc. of IEEE International Conference on Automatic Face and Gesture Recognition, pp. 529–534, 2004.

[18] F. Dominio, M. Donadeo, G. Marin, P. Zanuttigh, and G. M. Cortelazzo, "Hand Gesture Recognition with Depth Data," In proc. of ACM/IEEE International Workshop on Analysis and Retrieval of Tracked Events and Motion in Imagery Stream, pp. 9–16, 2013.

[19] Z. Ren, J. Yuan, J. Meng, and Z. Zhang, "Robust part-based hand gesture recognition using kinect sensor," IEEE Trans. on Transactions and Multimedia, vol. 15, no. 5, pp. 1110–1119, 2013.

[20] J. Shotton, A. Fitzgibbon, M. Cook, T. Sharp, M. Finocchio, R. Moore, A. Kipman, and A. Blake, "Real-time human pose recognition in parts from single depth images," In Proc. CVPR, 2011.

[21] Autodesk 3Ds MAX, 2012.