Data Article

HGM-4: A new multi-cameras dataset for hand gesture recognition

Vinh Truong Hoang

Ho Chi Minh City Open University, Vietnam

A R T I C L E   I N F O

A B S T R A C T

Gesture recognition technology is rapidly growing in the recent years due to the demands of many application such as computer game and sport, human robot interaction, assistant systems, sign language interpretation and e-commerce. One of the most important of gesture recognition is hand-gesture recognition. For example, it can be used to control all devices (television, radio, air-condition, and doors) by just hand gestures for smart home application. The HGM-4 dataset is built for hand gesture recognition (the full dataset is available from: https://data.mendeley.com/datasets/jzy8zngkbg/4) which contains total 4,160 color images (1280×700 pixels) of 26 hand gestures captured by four cameras at different position. The training and testing set are defined to create a benchmark framework for comparing the experimental results.

E-mail address: vinh.th@ou.edu.vn

Specifications table

Subject	Computer Vision, Pattern Recognition, Artificial Intelligence
Specific subject area	hand-gesture recognition, image classification, biometric recognition, sign language
Type of data	Image (1280×700 pixels) in RGB color space
How data were acquired	This dataset contains images that were taken by 4 cameras at different positions by laptop camera, indoor condition.
Data format	RAW
Parameters for data collection	Hand-gesture images are removed background semi-automatically.
Description of data collection	This dataset consists of 4,160 images of 26 gestures acquired by 4 different cameras.
Data source location	Ho Chi Minh City Open University, Ho Chi Minh City, Vietnam
Data accessibility	Mendeley Data
	https://data.mendeley.com/datasets/jzy8zngkbg/4
	http://dx.doi.org/10.17632/jzy8zngkbg.4

Value of the data

- This dataset is constructed for hand-gesture recognition which contains 26 different gestures corresponding to 26 letters of sign language.
- This is the first dataset containing 4 cameras images for hand-gesture in contrast with the rest pubic datasets.
- Hand gesture recognition might be used from this dataset in supervised and semi-supervised learning context.
- This dataset can be applied to study the hand-gesture recognition problems under multiple views. The potential applications can be used for sign language interpretation, contactless device control.
- We propose three strategies of experimental protocol with one, two and three training sets per gesture. The image from 4 cameras were combined as (training, testing) couples with all possible combination. For example, all the images captured from 1 camera are used for testing while the images from the remaining 3 cameras are used as a training set as the first strategy. This decomposition makes HGM-4 as a first benchmark dataset for multiple cameras hand-gesture recognition.

1. Data

Gesture recognition allows to interpret an image or sequence of images, i.e., video into a meaningful description. Among them, hand gesture recognition is the active research topics in machine vision and human robot interaction and has a wide range of potential applications such as video games, medical systems, wearable devices, and multimedia systems [11,12]. Many different approaches exist based on image analysis can be found in literature. Chansri and Srinonchat [1] present hand gesture of Thai sign language under a complex background using fusion of depth and color video. Maqueda et al. [6] a robust vision-based hand-gesture recognition system using volumetric spatiograms of local binary patterns. Dinh et al. [2] presents hand gesture interface for appliances control in smart home environments based on synthetic hand depth and random forests classifier. Dominio et al. [3] extract and divide the acquired hand images into palm and finger regions. Then, four different image descriptors are extracted and an SVM classifier is associated to recognize the performed gestures. Guan et al. [4] introduce a method by fusing information from multiple cameras to provide reliable hand pose estimation. Just and Marcel, [5] present a comparative study of hand gesture recognition in an isolated, complex, dynamic environment based on Hidden Markov Model. Tavakoli et al. [13] introduce a method to classify hand gestures on wearable devices that use EMG sensors as an input source.

Fig. 1. Illustration of one hand-gesture by two different views under different cameras.

Table 1
Summary of the recent published hand-gesture dataset in the literature.

Dataset Name	Number of views	Number of gestures	Total images	Resolution	Publicly
FEMD [7]	1	12	1,000	640 × 480	No
Interact Play [5]	2	16	16,000	-	Yes
IMHG [12]	2	8	836	640 × 480	Yes
HGM-4	4	26	4,000	1280 × 700	Yes

There are a few hand gesture databases available to the research community. Most of the database consist of one hand gestures. Just and Marcel [5] present the first dataset for both one- and two-handed gestures. Recently Poon et al. [9,10] present a new study for bimanual (two-hands) gesture recognition to overcome the drawback of hand-hand self-occlusion. Fig. 1 illustrates an example of this phenomena in case of one hand gesture of a single gesture captured by two different cameras in front of and below the hand. Pisharady and Saerbeck [8] present a complete review methods and databases in vision-based hand gesture recognition in 26 publicly available hand gesture databases. All the reviewed databases are based on single view. By analyzing the recent published hand gesture datasets in literature (see Table 1), we see that, there is a few public hand gesture datasets dealing with multi-view cameras. The IMHG [12] is a public dataset with front view and side view for each gesture. Motivated from this idea, we propose the novel, publicly available HGM-4 for one hand gesture dataset. An illustration of 26 gestures images of HGM-4 dataset are shown in Table 2. These gestures represent the alphabet letter of Vietnamese sign language. This dataset can be applied for dealing with contactless device control application or sign language interpretation since the cameras can be disposed at any positions.

2. Experimental Design, Materials, and Methods

This data is available online at Mendeley Repertory. It is organized in four main folders: CAM _Left, CAM_Right, CAM_Front and CAM_Below. Each main folder contains 26 sub-folders corresponding 26 classes of hand-gestures. Each sub-folder (from A to Z) has exactly 40 colored images with 1280 × 700 pixels. Table 3 presents the properties of HGM-4 dataset. Each gesture is performed by 5 persons. Four cameras have been used to capture hand gesture at four different positions. The cameras setup of our method is illustrated in Fig. 2. We have one monitor and four fixed cameras. We have 5 volunteers and each one performs 26 hand gestures. Each person performs hand gesture in front of monitor and above the keyboard. Four images are then

Table 2

The 26 classes of hand gesture of HGM-4 dataset.

Gesture	Illustration	Gesture	Illustration	Gesture	Illustration
A		J		S	
B		K		T	
C		L		U	
D		M		V	
E		N		W	
F		O		X	
G		P		Y	
H		Q		Z	
I		R			

Table 3

Properties of HGM-4 dataset.

Camera Position	Total images	Number of gestures	Number of images per gesture	Number of acting persons
CAM _Left	1,040	26	8	5
CAM_Right	1,040	26	8	5
CAM_Front	1,040	26	8	5
CAM_Below	1,040	26	8	5

Fig. 2. Camera setup: each hand-gesture (in front of screen and above the keyboard) is captured at the same time by four cameras.

Fig. 3. Illustration of three distinct images of the same gesture captured by below camera.

Fig. 4. Original image and its segmentation with removed background by Otsu's method and enhanced by technical expert.

captured for each gesture simultaneously. The first gesture is performed at the middle of four cameras. After acquiring each picture, the volunteer moves the hand with the same gesture in order to have 8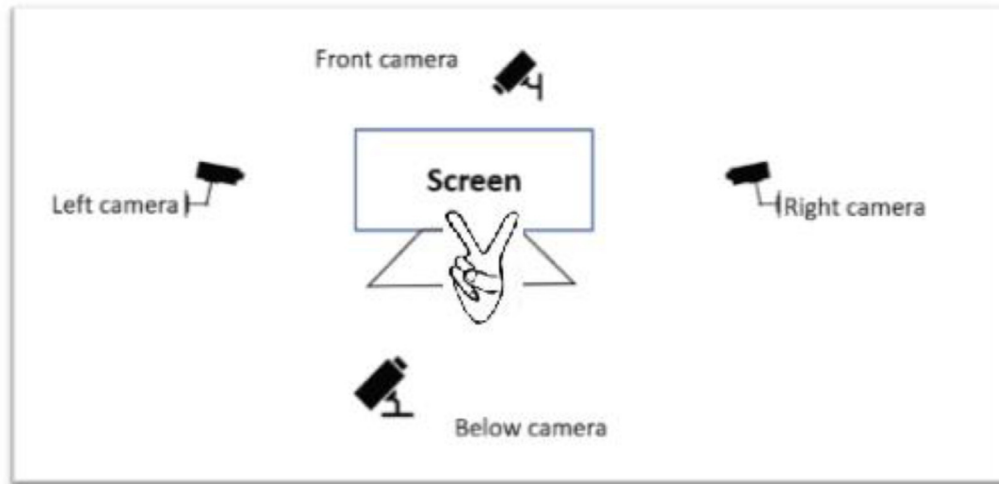 different images at different scales. A new movement must not rotate the hand compared with the first performance. Fig. 3 illustrates three distinct images of the same gesture captured by below camera. It is worth to note that this screen is used to control and view images from four cameras. A technician will take four images after verifying quality and resolution.

All images are segmented to remove background as illustrated in Fig. 2 by The Otsu's method [14]. This approach returns a single intensity threshold that separate pixels into two classes, foreground, and background (as illustrated in Fig. 4). The automated background removal tool is applied automatically based on selecting bimodal histogram. We use Matlab program to perform this task. However, it does not give a perfect result. In some cases, it still contains the pixels of another object or pixels of hand is removed unintentionally. Our technician verifies each image and enhance the background removal again by Photoshop program.

The standard protocol is computed by the average accuracy over 4 decomposition. In each time, all the images from one camera are used for testing set while images from the rest cameras are used for training sets. The purpose is to learn, and model a given physical images captured from different condition such as view, distance, or cameras. Three configurations are proposed, with one, two and three training sets per gesture with all possible combinations is listed in Table 4 such as:

Table 4

All combination possible to create training and testing set for experiment.

For one training set		For two training sets		For three training sets	
Training	Testing	Training	Testing	Training	Testing
CAM _Left	CAM _Right	CAM _Front	CAM _Left	CAM _Right	CAM _Left
	CAM _Front	CAM _Below	CAM _Right	CAM _Front	
	CAM _Below			CAM _Below	
CAM_Right	CAM _Left	CAM _Left	CAM _Front	CAM _Left	CAM_Right
	CAM _Front	CAM _Right	CAM _Below	CAM _Front	
	CAM _Below			CAM _Below	
CAM_Front	CAM _Right	CAM_ Below	CAM _Front	CAM _Right	CAM_Front
	CAM_Left	CAM _Left	CAM _Right	CAM_Left	
	CAM _Below			CAM _Below	
CAM_Below	CAM _Right,	CAM _Front	CAM_ Below	CAM _Right	CAM_Below
	CAM_Left,	CAM _Right	CAM _Left	CAM_Left	
	CAM _Front	CAM _Left	CAM _Right	CAM _Front	
		CAM _Front	CAM _Below		
		CAM _Right	CAM _Left		
		CAM _Below	CAM _Front		

Conflict of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.dib.2020.105676.

References

[1] C. Chansri, J. Srinonchat, Hand Gesture Recognition for Thai Sign Language in Complex Background Using Fusion of Depth and Color Video, Procedia Computer Science 86 (2016) 257–260, doi:10.1016/j.procs.2016.05.113.

[2] D.-L. Dinh, J.T. Kim, T.-S. Kim, Hand Gesture Recognition and Interface via a Depth Imaging Sensor for Smart Home Appliances, Energy Procedia 62 (2014) 576–582, doi:10.1016/j.egypro.2014.12.419.

[3] F. Dominio, M. Donadeo, P. Zanuttigh, Combining multiple depth-based descriptors for hand gesture recognition, Pattern Recognition Letters 50 (2014) 101–111, doi:10.1016/j.patrec.2013.10.010.

[4] Haiying Guan, Jae Sik Chang, Longbin Chen, R.S. Feris, M. Turk, Multi-view Appearance-based 3D Hand Pose Estimation, in: 2006 Conference on Computer Vision and Pattern Recognition Workshop (CVPRW'06), New York, NY, USA, IEEE, 2006 154–154, doi:10.1109/CVPRW.2006.137.

[5] A. Just, S. Marcel, A comparative study of two state-of-the-art sequence processing techniques for hand gesture recognition, Computer Vision, and Image Understanding 113 (2009) 532–543, doi:10.1016/j.cviu.2008.12.001.

[6] A.I. Maqueda, C.R. del-Blanco, F. Jaureguizar, N. García, Human–computer interaction based on visual hand-gesture recognition using volumetric spatiograms of local binary patterns, Computer Vision and Image Understanding 141 (2015) 126–137, doi:10.1016/j.cviu.2015.07.009.

[7] Z. Ren, J. Yuan, Z. Zhang, Robust Hand Gesture Recognition Based on Finger-Earth Mover's Distance with a Commodity Depth Camera, in: Proceedings of the 19th ACM International Conference on Multimedia, Association for Computing Machinery, New York, NY, USA, 2011, pp. 1093–1096, doi:10.1145/2072298.2071946.

[8] P.K. Pisharady, M. Saerbeck, Recent methods and databases in vision-based hand gesture recognition: A review, Computer Vision, and Image Understanding 141 (2015) 152–165, doi:10.1016/j.cviu.2015.08.004.

[9] G. Poon, K.C. Kwan, W.-M. Pang, Real-time Multi-view Bimanual Gesture Recognition, in: 2018 IEEE 3rd International Conference on Signal and Image Processing (ICSIP), Shenzhen, IEEE, 2018, pp. 19–23, doi:10.1109/SIPROCESS.2018.8600529.

[10] G. Poon, K.C. Kwan, W.-M. Pang, Occlusion-robust bimanual gesture recognition by fusing multi-views, Multimed Tools Appl 78 (2019) 23469–23488, doi:10.1007/s11042-019-7660-y.

[11] Y. Zhang, C. Cao, J. Cheng, H. Lu, EgoGesture: A New Dataset and Benchmark for Egocentric Hand Gesture Recognition, IEEE Trans. Multimedia. 20 (2018) 1038–1050, doi:10.1109/TMM.2018.2808769.

[12] D. Shukla, The IMHG dataset: A Multi-View Hand Gesture RGB-D Dataset for Human-Robot Interaction, (n.d.) 4.
[13] M. Tavakoli, C. Benussi, P. Alhais Lopes, L.B. Osorio, A.T. de Almeida, Robust hand gesture recognition with a double channel surface EMG wearable armband and SVM classifier, Biomedical Signal Processing and Control 46 (2018) 121–130, doi:10.1016/j.bspc.2018.07.010.
[14] N. Otsu, A threshold selection method from gray-level histograms, IEEE Transactions on Systems, Man and Cybernetics 9 (1979) 62–66.