2018 13th IEEE International Conference on Automatic Face & Gesture Recognition

Holoscopic 3D Micro-Gesture Database for Wearable Device Interaction

Yi Liu, Hongying Meng, Mohammad Rafiq Swash, Yona Falinie A. Gaus, Rui Qin
Department of Electronic and Computer Engineering
Brunel University London
London, UK
{*yi.liu,hongying.meng,rafiq.swash,yonafalinie.abdgaus,rui.qin*}*@brunel.ac.uk*

***Abstract*—With the rapid development of augmented reality (AR) and virtual reality (VR) technology, human-computer interaction (HCI) has been greatly improved for gaming interaction of AR and VR control. The finger micro-gesture is a hot research focus due to the growth of the Internet of Things (IoT) and wearable technologies and recently Google has developed a radar based micro-gesture sensor which is Google Soli. Also, there are a number of finger micro-gesture techniques have been developed using Time of Flight (ToF) imaging sensors for wearable 3D glasses such as Atheer mobile glasses. The principle of holoscopic 3D (H3D) imaging mimics fly's eye technique that captures a true 3D optical model of the scene using a microlens array, however, there is a limited progress of holoscopic 3D systems due to the lack of high quality public available database. In this paper, holoscopic 3D camera is used to capture high quality holoscopic 3D micro-gesture video images and a new unique holoscopic 3D micro-gesture (HoMG) database is produced. HoMG database recorded the image sequence of 3 conventional gestures from 40 participants under different settings and conditions. For the purpose of H3D micro-gesture recognition, HoMG has a video subset of 960 videos and a still image subset with 30635 images. Initial micro-gesture recognition on both subsets has been conducted using the traditional 2D image and video features and popular classifiers and some encouraging performance has been achieved. The database will be available for the research communities and speed up the research in the area of holoscopic 3D micro-gesture.**

***Keywords*-Holoscopic 3D; Micro-gesture; database; wearable;**

I. Introduction

Gesture is a remarkable interaction way for Human Computer Interaction (HCI), which is a conventional non-verbal communication method. It is one type of pervasive body language that can be used for communication. However, with the development of the gaming interaction and wearable device, precise finger gesture has more advantages than body gesture, especially for control devices [1]. The finger movement is one of the micro-gestures that can accurately manipulate the device. Kinect and RBG-D camera were popular sensors for gaming in the Augmented Reality (AR) and Virtual Reality (VR) community with its low-cost been a major advantage, as well as its immersive user experience and usability [2]. There displays support the 3D gesture systems and need free space to support flexible interaction [3]. However, these systems lack the ability to capture quality and accurate objects which could be seen as one of its major drawbacks[4]. Recently, some new research from Leap Motion [5] and Google Soli Project [6] created new techniques for 3D detection that has huge potentials for success. Holoscopic 3D (H3D) imaging system is a novel potential technique which can satisfy the higher demand for user interactive experience. Detection of precision 3D micro-gesture can make use of the wide view coverage of the Holoscopic 3D camera to capture accurate finger movement [7].

H3D system supports the RGB high quality dynamic and static data, and it is renowned for high accuracy and true 3D to excellence than traditional 3D capture devices. H3D system has been very successful in 3DTV and display area. However, this technology has not been used widely for capturing and recognition of the finger gesture. This paper aims to use the H3D imaging system to create a unique 3D micro-gesture database, further to promote the gesture recognition.

II. Related Work

HCI appeared early in 1983 [8], which use multiple modalities such as voice, gestures (e.g. body, hand, arm, finger, etc.). For example, Siri [9] is a very popular voice-based interface. However,the natural gesture is another way to interact with the computer. The trend of the HCI is user experience of intuitionistic and effective [3]. The gesture is a touchless, non-intrusive method for HCI, and it is represented as the diverse type of the gestures [10]. Manipulative type of gesture appears the most popular one from the previous literature. The aim is to control entity being manipulated through the actual movements of the gesturing hand and arm [11]. Hand as a direct input device is more and more popular, as one of the outstanding interaction methods.

The Kinect and RGB-D camera were very popular in recent years due to the benefits of Kinect and RGB-D camera that have low cost and wide availability as a sensor [12] to capture gestures. However, RGB-D camera suffers from the underside artifacts such as the edge inaccuracies, low object remission [4]. The Kinect sensor offers the information of

978-1-5386-2335-0/18/$31.00 ©2018 IEEE
DOI 10.1109/FG.2018.00129

(a) Button (b) Dial (c) Slider

Figure 1. Three different types of finger micro-gestures studied in HoMG database.

the depth measurement and creates coordinates of the 3D objects. Although the abundant development toolkits can support the human body recognition, the weakness is its lacking ability to capture the flexible and robust mechanism to perform high-level gesture [13].

Leap Motion (LM) [14] is a device that can be used to detect the hand and finger dynamic movements through its API software. The API has the robust pre-processing function which can reduce the complexity of the user control. However, LM is a monocular video sensor which is a challenging for capturing the abundant dynamic hand gestures and finger micro movements [15].

Holoscopic 3D camera is a single aperture sensor not only to represent the real-time and represents a true volume spatial optical model of the object scene but also to record the viewing natural continuous parallax 3D objects within a wide viewing zone [16]. It provides a new way to capture micro-gestures.

Isaac et al. [3] presents a review summary of 21 gesture datasets from previous research and public datasets, in which 7 databases are for hands. Most datasets are recorded using to the Kinect or RGB-D camera as the sensor.

In order to the support the diversification of the gesture recognition and encourage the development of the human computer interaction, we propose a new 3D gesture database included the three ubiquitous micro gestures that are most the popular ones used in the Google Soli project. Those are intuitive and unobtrusive manipulative gesture. This database does not only include the continuous dynamic data but also contained the abundant static data to support the 3D micro-gesture recognition.

III. Database Construction

A. Micro-gestures

There are many micro-gestures that can be used for control in AR and VR applications. In this research, three intuitive micro-gestures are selected references to the Google Soli project [6] as shown in Fig. 1. The three gestures are based on the human intuitiveness when they try to control display. For instance, the button gesture executes the submission function, dial gesture shows that user wants to slightly adjust the current situation, and the slider gesture is to express the slide up or down to adjust the volume and options. This three gesture belong to the manipulative type of the gesture, which are used to touchless control the devices or simulation console.

B. H3D imaging technology

H3D imaging technology is a success for use in the 3D cinema, 3D-capable televisions and broadcasters. The H3D camera used here is built from the 3D Vivant Project (3D Live Immerse Video-Audio Interactive Multimedia) [17] and the purpose is to capture high quality 3D images. The developed camera includes micro-lens array, relay lens, and digital camera sensors. The principle of the holoscopic 3D imaging is shown in Fig. 2. The 3D holoscopic image's spatial sampling is determined by the number of lenses. It shows that the captured 2D lenslet array views are slight different angle than its neighbor and reconstructed image in relay [17]. The detailed parameters of the camera are shown in Fig. 3.

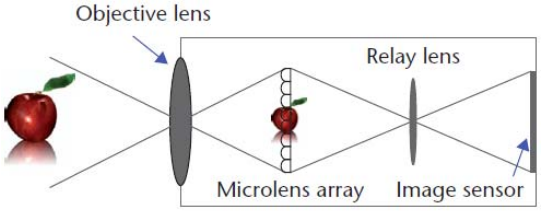

Figure 2. Principle of the holoscopic 3D camera. The microlens array is placed between objective and relay lens to produce fly eye style images. [17]

Holoscopic 3D camera sensor has unique optical components which support the continuous parallax RGB image system, contain the depth information viewpoint images. The figure shows the H3D imaging having the full color with full parallax. H3D imaging is comprised of the 2D array of micro images.

The H3D sensor is a crucial requirement for the capture of the objects. This database uses the H3D imaging system to support the dynamic and static RGB data. And its not only can record the continuous motion, but repetitive lens

array can extract different angles viewpoint images. The uniqueness to encourage the innovation of gesture capture and recognition.

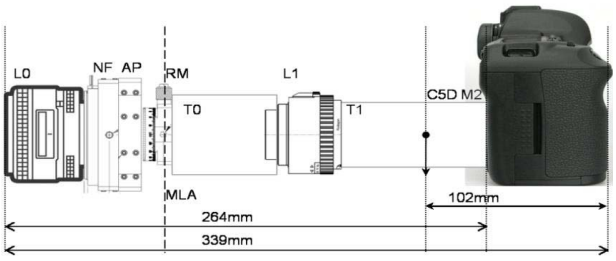

Figure 3. Assembled holoscopic 3D camera.

C. Recording Setup

Figure 4. Data acquisition setup.

The recording H3D gesture place, as in Fig.4. A green screen room is used for the recording where it can offer clear and professional recording background to reduce noise. Before the recording, the holoscopic 3D camera adapter and surface are set up in advance. Canon 5D camera is used and the camera settings are configured to ISO200, shutter 1/250. Holoscopic 3D camera adapter is calibrated and the lens is corrected.

Considering the influence of distances, angles, and backgrounds, we prepared 4 positions for participants. Two positions are the close and far locations where the objective lens set to 45cm and 95cm. The other two positions are from the left and right hand side for the convenience of the participants. In the closed position, we set a hollow frame to help the participant to find the 3D micro-gesture capture zone.

We remind the participants the gesture name while record their finger movements. The participants can perform their micro-gesture at their own speed. The recording is done during the time around 15 minutes for one participant.

We prepared two different colour backgrounds, two different distances of close and far from the end of the camera lens to gesture area. The recorded imaging resolution is 1086x1902, and the micro lens is 28x28. Participants are successively stand each pre-established position to play three gestures around 3-5 seconds. The three gestures are involved button, dial,and slider.

Table I
DETAILED INFORMATION ABOUT THE DATA ACQUISITION.

Parameters	**Detailed Information**
Micro-gesture	Button (B), Dial (D), Slider (S)
Participants	Male (33), Female (17)
Hand	Right (R), Left (L)
Distance	Close (45cm), Far (95cm)
Background	Green (G), White (W)
Camera	Canon 5D
Image resolution	1902 x 1086
Lens array	28 x 28
Shutter speed	1/250
Film speed	ISO200
Frame rate	25
Recording length	Between 2 and 20 Sec.

D. Participants

In total, 40 participants attended the recordings including 17 female participants and 33 male participants, who all read the participant information sheet guidance and sign the research ethics application forms before the recording. There is no any limitation of age and race for the participants and We respect the participants will. Some participants wear married rings and watches during the recording on finger movements. These increase the datas noise and bring more challenges. We recorded the participants double hands in order to increase the diversity of the data. The detailed information about the data acquisition is summarized in Table I.

E. HoMG Database

For the data collection, the recordings from 40 participants are selected to make the HoMG database. The recordings were done under different conditions. One participant has recorded 24 videos. In total, 960 videos are included in the database.

For micro-gesture recognition, it can be done based on single image or can be done from a short video. So this database was divided into two subsets: image based and video based micro-gesture subsets.

1) Video subset: There are 40 subjects and each subject has 24 videos due to the different setting and three gestures. For each video, the frame rate is 25 frames per second and length of videos are from few seconds to 20 seconds and not equally. The whole dataset was divided into 3 parts. 20

subjects for the training set, 10 subjects for development set and another 10 subjects for testing set. In this way, the micro-gesture recognition is person independent.

2) Image subset: Video can capture the motion information of the micro-gesture and it is a good way for micro-gesture recognition. However, it needs more data and takes a long time. It is very interesting to see whether it is possible to recognize the micro-gesture from a single image with high accuracy.

From each video recording, the different number of frames were selected as the still micro-gesture images. In total, there are 30635 images selected. The whole dataset was split into three partitions: a Training, Development, and Testing partition. There are 15237 images in the training subsets of 20 participants with 8364 in close distance and 6853 in the far distance. There are 6956 images in the development subsets of 10 participants with 3077 in close distance and 3879 in far distance. There are 8442 images in the testing subsets of 10 participants with 3930 in close distance and 4512 in far distance.

The summary of the HoMG database is listed in the Table II.

Table II
THE SUMMARY OF THE HOMG DATABASE.

Partition	Subjects	Image Set	Video Set
Training	20	16763	480
Development	10	6560	240
Testing	10	7291	240

IV. INITIAL INVESTIGATION ON MICRO-GESTURE RECOGNITION

The initial investigation is carried out independently based micro-gesture recognition study from video and image separately. We would like to see how high performance can be achieved from each.

A. Video based micro-gesture recognition

There are many good features that can be extracted from each video to capture the movement of the fingers. Here LBPTOP [18] and LPQTOP [19] are selected. These features can not only calculate the distribution of the local information of each frame, but also the distribution of finger movements along to the time. From each video, the frame size was reduced to 66x38 from 1920x1080 firstly, then a feature vector with the dimension of 768 is extracted using LBPTOP and LPQTOP for the classification. For the classification, it is a three class classification problem. There are lots of classifiers available. Here, most popular ones such as k-NN, Support Vector Machines (SVM) and Naive Bayes classifiers are chosen for comparison purpose.

Table III shows the accuracy using three different classifiers under different distance on video based micro-gesture recognition. From this table, it can be seen that LPQTOP is better than LBPTOP for feature extraction. SVM is better than k-NN and Naive Bayes classifiers in most cases. In general, the accuracy on close distance is better than far distance because the detailed information of the finger movement can be captured. For the testing set, both training and development sets were used for training. Overall, 66.7% accuracy can be achieved even use the feature extraction methods from all videos in the testing set.

Table III
RECOGNITION ACCURACY (%) OF VIDEO BASED MICRO-GESTURE RECOGNITION ON DEVELOPMENT (DEV.) AND TESTING SETS USING K-NN, SVM AND NAIVE BAYES CLASSIFIERS.

Dataset	Distance	Feature	Classifier		
			k-NN	*SVM*	*Bayes*
Dev.	Close	LBPTOP	53.3	68.3	52.5
		LPQTOP	56.7	66.7	63.3
	Far	LBPTOP	40.8	53.3	47.5
		LPQTOP	50.8	55.8	49.2
	All	LBPTOP	44.5	52.9	47.9
		LPQTOP	47.9	**60.4**	51.3
Test	Close	LBPTOP	56.7	53.3	40.8
		LPQTOP	67.5	73.3	65.8
	Far	LBPTOP	55	55	50.8
		LPQTOP	51.7	65.8	58.3
	All	LBPTOP	53.3	59.5	45.4
		LPQTOP	60.4	**66.7**	57.5

Table IV
RECOGNITION ACCURACY (%) OF IMAGE BASED MICRO-GESTURE RECOGNITION ON DEVELOPMENT (DEV.) AND TESTING SETS USING K-NN, SVM AND NAIVE BAYES CLASSIFIERS UNDER DIFFERENT DISTANCE CONDITIONS.

Dataset	Distance	Feature	Classifier		
			k-NN	*SVM*	*Bayes*
Dev.	Close	LBP	40.9	44.3	46.0
		LPQ	43.4	45.0	42.8
	Far	LBP	35.9	32.1	37.4
		LPQ	36.7	52.6	47.5
	All	LBP	41.0	35.0	39.6
		LPQ	32.9	**51.6**	50.6
Test	Close	LBP	49.7	33.6	45.4
		LPQ	44.1	46.4	39.7
	Far	LBP	50.9	37.7	47.2
		LPQ	34.6	51.6	50.0
	All	LBP	44.7	48.9	44.7
		LPQ	46.8	**50.9**	46.8

B. Image based micro-gesture recognition

For each image, 2D texture features such LBP [20] and LPQ [21] were extracted to represent each image. These two features captured the edge and local information of the 2D image in different ways and form a histogram feature vector with the dimension of 256. Popular classification methods such as k-NN, SVM and Naive Bayes classifiers were used for recognising the three different micro-gestures.

Table IV should the experimental results on video based micro-gesture recognition by training on the training set and

tested on the development, and testing subsets. From this table, it can be seen that for most of the classifications, around 50% accuracy can be achieved.

V. Conclusions and Future Works

A. Conclusions

This paper introduces a unique holoscopic 3D micro-gesture database (HoMG), which is recorded under different settings and conditions from 40 participants. The data recording uses the similar the H3D system of fly viewing to capture the participants precise finger movements. The H3D imaging system supports robust 3D depth micro lens array to capture dynamic and static information. The HoMG database has 3 unobtrusive manipulative gestures in two different backgrounds, two different distances, left and right hands. These micro-gestures can be used to control multifarious displays. This database would speed up the research in this area.

The database is further divided into video and image subsets. Initial investigation of micro-gesture recognition is conducted. For the comparison, video based method achieved better performance as it has dynamic finger movement information in the data. However, this method needs much more data and computing time. Image based method is convenient for the user and might have more applications, especially on the portable devices. Even with the standard 2D feature extraction methods and basic classification methods, 66.7% recognition accuracy can be achieved for micro-gesture videos and over 50.9% accuracy for micro-gesture images. This baseline methods and results will give a foundation for other researchers to explore their methods.

B. Future works

From the initial investigation, it can be seen that the recognition accuracy can reach around 66% even just using the 2D image processing methods. For 3D image processing methods, such as extracting the different viewing point images and extract 3D information of the micro-gesture, high accuracy will be achieved. This will be our future works. In addition, more type of gestures can be added into the dataset for wide applications.

Acknowledgment

We gratefully acknowledge the support of NVIDIA Corporation with the donation of the Tesla K40 GPU used for this research.

References

[1] R. Huslschmid, B. Menrad, and A. Butz, "Freehand vs. micro gestures in the car: Driving performance and user experience," in *2015 IEEE Symposium on 3D User Interfaces (3DUI)*, March 2015, pp. 159–160.

[2] Z. Zhang, "Microsoft kinect sensor and its effect," *IEEE MultiMedia*, vol. 19, no. 2, pp. 4–10, Feb 2012.

[3] I. Wang, M. B. Fraj, P. Narayana, D. Patil, G. Mulay, R. Bangar, J. R. Beveridge, B. A. Draper, and J. Ruiz, "Eggnog: A continuous, multi-modal data set of naturally occurring gestures with ground truth labels," in *2017 12th IEEE International Conference on Automatic Face Gesture Recognition (FG 2017)*, May 2017, pp. 414–421.

[4] F. Garcia, D. Aouada, T. Solignac, B. Mirbach, and B. Ottersten, "Real-time depth enhancement by fusion for RGB-D cameras," *IET Computer Vision*, vol. 7, no. 5, pp. 1–11, October 2013.

[5] G. Marin, F. Dominio, and P. Zanuttigh, "Hand gesture recognition with jointly calibrated leap motion and depth sensor," *Multimedia Tools Appl.*, vol. 75, no. 22, pp. 14 991–15 015, Nov. 2016.

[6] J. Lien, N. Gillian, M. E. Karagozler, P. Amihood, C. Schwesig, E. Olson, H. Raja, and I. Poupyrev, "Soli: Ubiquitous gesture sensing with millimeter wave radar," *ACM Trans. Graph.*, vol. 35, no. 4, pp. 142:1–142:19, Jul. 2016.

[7] M. R. Swash, O. Abdulfatah, E. Alazawi, T. Kalganova, and J. Cosmas, "Adopting multiview pixel mapping for enhancing quality of holoscopic 3D scene in parallax barriers based holoscopic 3D displays," in *2014 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting*, June 2014, pp. 1–4.

[8] C. Zhang, X. Yang, and Y. Tian, "Histogram of 3D facets: A characteristic descriptor for hand gesture recognition," in *2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*, April 2013, pp. 1–8.

[9] J. R. Bellegarda, "Spoken Language Understanding for Natural Interaction: The Siri Experience," in *Natural Interaction with Robots, Knowbots and Smartphones*, J. Mariani, S. Rosset, M. Garnier-Rizet, and L. Devillers, Eds. New York, NY: Springer New York, 2014, pp. 3–14.

[10] P. K. Pisharady and M. Saerbeck, "Recent methods and databases in vision-based hand gesture recognition: A review," *Computer Vision and Image Understanding*, vol. 141, pp. 152–165, 2015.

[11] M. Karam and m. c. Schraefel, "A Taxonomy of Gestures in Human Computer Interactions," *Technical Report, Eletronics and Computer Science.*, pp. 1–45, 2005.

[12] G. Ren and E. O'Neill, "3D marking menu selection with freehand gestures," in *2012 IEEE Symposium on 3D User Interfaces (3DUI)*, March 2012, pp. 61–68.

[13] R. Ibañez, Á. Soria, A. Teyseyre, and M. Campo, "Easy gesture recognition for Kinect," *Advances in Engineering Software*, vol. 76, pp. 171–180, 2014.

[14] L. E. Potter, J. Araullo, and L. Carter, "The leap motion controller: A view on sign language," in *Proceedings of the 25th Australian Computer-Human Interaction Conference: Augmentation, Application, Innovation, Collaboration*, ser. OzCHI '13. New York, NY, USA: ACM, 2013, pp. 175–178.

[15] W. Lu, Z. Tong, and J. Chu, "Dynamic hand gesture recognition with leap motion controller," *IEEE Signal Processing Letters*, vol. 23, no. 9, pp. 1188–1192, Sept 2016.

[16] B. Kaufmann and M. Akil, "3D images compression for multi-view auto-stereoscopic displays," in *International Conference on Computer Graphics, Imaging and Visualisation (CGIV'06)*, July 2006, pp. 128–136.

[17] A. Aggoun, E. Tsekleves, M. R. Swash, D. Zarpalas, A. Dimou, P. Daras, P. Nunes, and L. D. Soares, "Immersive 3D

holoscopic video system," *IEEE MultiMedia*, vol. 20, no. 1, pp. 28–37, Jan 2013.

[18] G. Zhao and M. Pietikainen, "Dynamic texture recognition using local binary patterns with an application to facial expressions," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 6, pp. 915–928, June 2007.

[19] B. Jiang, M. Valstar, B. Martinez, and M. Pantic, "A dynamic appearance descriptor approach to facial actions temporal modeling," *IEEE Transactions on Cybernetics*, vol. 44, no. 2, pp. 161–174, Feb 2014.

[20] T. Ojala, M. Pietikäinen, and D. Harwood, "A comparative study of texture measures with classification based on featured distributions," *Pattern Recognition*, vol. 29, no. 1, pp. 51–59, 1996.

[21] V. Ojansivu and J. Heikkilä, *Blur Insensitive Texture Classification Using Local Phase Quantization*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, pp. 236–243.