DYNAMIC HAND GESTURE RECOGNITION USING SPATIAL-REMPORAL FETURES

Huong Giang Doan[1,2], Duy Anh Vu[1], Hai Vu[1] , Thanh Hai Tran[1]

[1] International Research Institute MICA HUST - CNRS/UMI - 2954 - INP Grenoble
[2] Industrial Vocational College Hanoi

{huong-giang.doan, hai.vu, thanh-hai.tran}@mica.edu.vn, duyanhbkhn@gmail.com

ABSTRACT *- Hand gesture recognition has been studied for a long time. However it still is a challenge field. Furthermore, the hand gesture recognition to control devices in smart-house such as television, light, fan, camera, door,... that requires height recognition accuracy. This paper proposed a simple and effective model to recognize a defined hand gesture set. Analyzing Spatial - Temporal characteristics that includes cycle pattern, different in length, non-synchronization phase of gesture, hand shape changing during temporal dimension and direction features of hand movements. After that hand gestures recognition is based on this Spatial - Temporal features.*

Keywords *- Human Computer Interaction, Hand gesture recognition, Spatial-Temporal hand gesture recognition, Principal Componel Analysis.*

I. INTRODUCTION

Hand gesture recognition has been a very active research topic in the area of computer vision. It has been widely applied in a large variety of practical applications in real world which includes security and surveillance, content-based video analysis, Human-Computer Interaction (HCI) and animation. The mainly HCI, e.g recognizing hand/body gestures to control game consoles [1]. Samsung smart-TV can manipulate TV-functions using dynamic hand gestures. Omron introduces the smart-TV integrated facial and hand recognition. PointGrab [2] proposed an unique solution based on shape and motion recognition including gesture, face and user behavior analytic. Consumer electronics and mobile devices like WiSee system [3]. However, Hand gesture recognition is still a challenging problem due to the complexity of hand shapes in gestures, multi-trajectory gestures, background condition and motion blurring and changing of light conditions. Recent research has been motivated to explore more efficient multi-modal gesture recognition methods [4][5][6][7]. Furthermore, how to recognize human gestures using multi-modal information in an efficient way is still a hot topic.

In my research, we try to solve some following problems:

- The first is following a hand gesture database set to control equipment that actions change following cycling and repeating of hand posture. They include on/off, go left and go right gestures. Moreover, each person implement is different from with other person about postures and velocity.
- We propose to use multi-modal data for this gestures recognition which are separated into spatial and temporal features.

Specifically, a novel approach using depth and RGB image from the Kinect sensor is proposed for multi-modal gesture recognition. The general framework of the proposed approach is illustrated in Fig. 1. Consequently, the proposed system is feasible to deploy practical applications. The rest of paper is organized as follows: Sec. II. briefly survey related works. Sec. III. describes the proposed framework. Sec. IV. describes the experimental results and finally Sec. V. concludes and suggests further research directions.

II. RELATED WORKS

In term of the deploying applications using hand gestures, [8] proposes a static hand language recognition system to support the hearing impaired people; [9] uses hand postures to control a remote robot in mechanical systems; Similar systems have been deployed for game simulations such as [5][10]. The fact, there are uncountable solutions for a vision-based hand posture recognition system. Readers can refer good surveys such as [12][13] for technical details. Roughly speaking, according to the hand gesture recognition technical, some methods has been implemented that are Neural Network [14], Hidden Markov Models (HMMs) [15][16][17][18], Dynamic Time Warping (DTW) [19][20] and Conditional Random Fields (CRFs) [21][22][23], Or according to features combining unitizing that are none, late fusion [24] or early fusion. Another hand, according to the data inputs that the existing methods of gesture recognition or action recognition can be roughly divided into four categories: RGB video based, depth video based, skeleton data based and multi-modal data based.

In this study, we pursue a hand posture recognition system for controlling devices (e.g., televisions, lighting systems) in a smart-room. Therefore, we briefly survey recent trends that feasibly deploy to home appliances. Microsoft Xbox-Kinect is a success commercial product recognizing hand/body gestures to control game consoles [1]. Many technology companies launch smart-devices using like-Kinect sensors (e.g., Asus Kinect, softKinect). For instance, Samsung smart-TV can manipulate TV-functions using dynamic hand gestures. Omron introduces the smart-TV integrated facial and hand recognition. PointGrab [2] proposed an unique solution based on shape and motion recognition including gesture, face, and user behavior analytic. Increasingly, in-air gesture recognition is being incorporated into consumer electronics and mobile devices like WiSee system [3]. WiSee focuses on gesture recognition and shows how to extract gesture information from wireless transmissions. It promises new trend for home appliances because this technique can operate in non-line-of-sight scenario, that is a limitation of the vision-based system.

Controlling equipment is high accuracy equipment. Normally, setting the commands to control that are cyclical and repetitive. In the fact, the end_users doesn't implement the standard dynamic hand gestures. There someone doesn't implement the true time (faster and slower) or another ones doesn't implement standard phases (start, change, stop). Different from above works, in this study we focus on resolve problems with our database. The proposed system effective combine the spatial and temporal features to recognize the dynamic hand gestures that changes both hand shape and temporal.

III. PROPOSED APROACH

A. Proposed framework

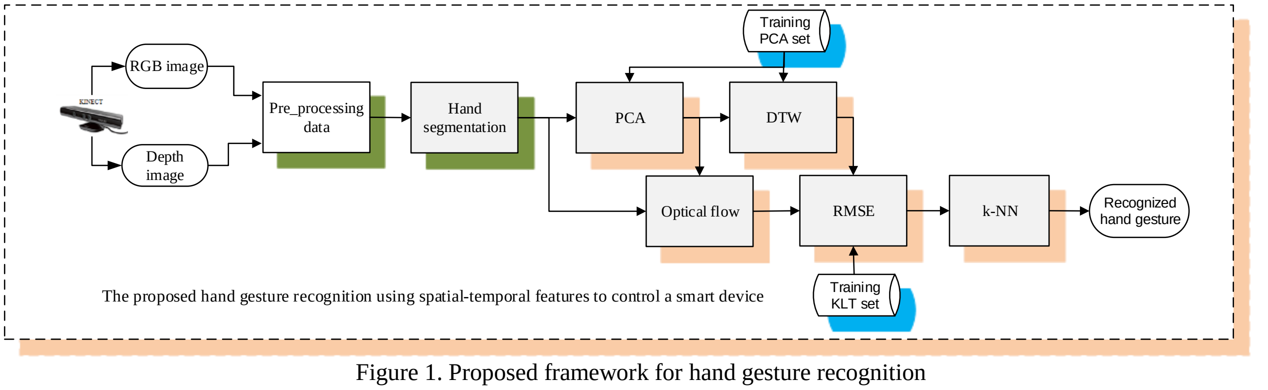


Figure 1. Proposed framework for hand gesture recognition

Hand gestures to control electronics equipment that requipments high accuracy recognition. Normally, controlling data for each activity of the hand gesture is a consecutive sequence of successive frames. Five hand gestures are defined and built by us to control electronics equipment that includes: increasing, decreasing, left, right and on/off. Those gestures are cycling and repeating. We consider dynamic hand gestures are an action that includes many consecutive hand postures in temporal and spatial. The problem is to find out the hand postures in each gesture, a connection of this hand posture about temporal and spatial to identify hand gesture. The implementation of sticking the whole process of hand gestures activities aimed at collecting more comprehensive information of each activity and provide appropriate identification for each gesture. Therefore, to describe the states of a processing, find out the relationship between them and a method of evaluating an activity with the same model with it or not. Here, an overview of the proposed framework is shown in Fig. 1. By using a fixed the Kinect sensor, a RGB image and a depth image are concurrently wrapped. The main flow-work for detecting hand from (I, D) images consists of a series of the cascaded steps, as shown in Fig. 1.

B. Pre-processing data

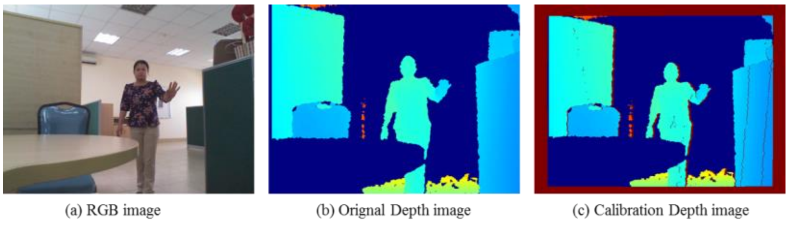


Figure 2. Calibration RGB-D images from the Kinect sensor

Depth and RGB images from the Kinect sensor are not measured from the same coordinates. In the past, there were many researches have mentioned this problems as [13]. In our work, we utilized calibration method of Microsoft to repair the depth images. The result showed in Fig. 2a is RGB image(I), Fig.2b is original depth image and Fig. 2c show result of calibration depth image (D).

C. *Hand segmentation*

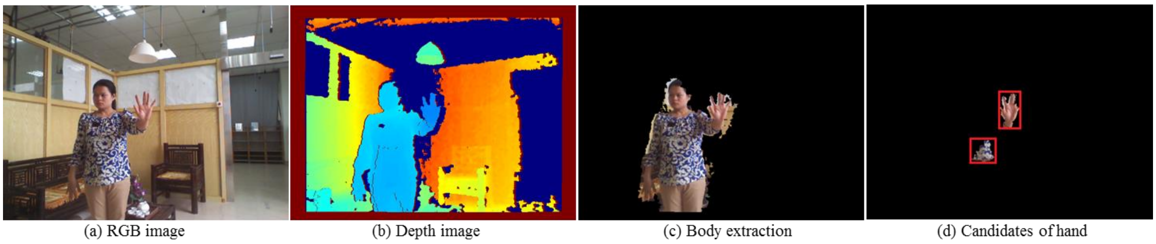


Figure 3. Hand region detection

Because the sensor and the environment are fixed, we firstly detect human regions using background subtraction techniques. Both depth and RGB images can be used for the background subtraction. Because, the depth data is less sensitive with illumination. Therefore, we use depth images for background subtraction. Among numerous techniques of the background subtractions, we adopt Gaussian Mixture Model (GMM) [36] because this technique have been shown to be the best suitable for our system. Figure 3(a-c) shows results of the background subtraction. Given a region of human body (as shown in Fig. 3c), we continuously extract candidates of the hand (as shown in Fig. 3d) and a hand segmentation result X (as shown in Fig. 4) is taken out after pruning hand region that presents detail in [24]

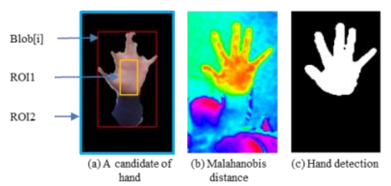
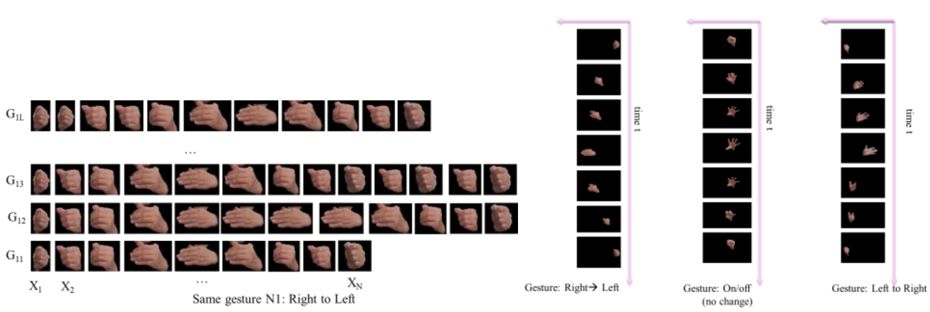


Figure 4. Hand segmentation

D. *Characteristics of hand gesture*

Hand gesture to be represented by a sequence of consecutive frames that improves hand images from the hand segmentation step. Hand gesture set is $G = \{G_i | i = 1,2,3,4,5\}$ which G_1 is on/off hand gesture set, G_2 is increase hand gesture set, G_3 is decrease hand gesture set, G_4 is go_left hand gesture set, G_5 is go_right hand gesture set. Each gesture $G_i = \{X_{ij} | j = 1,2,\dots,m\}$ is defined by postures, hand gestures changes in shape, speed and phase:

- Postures in a gesture are cycling and repeating (Fig. 5(a): the first stage, middle stage and final stage), first stage and final stage is nearly the same.
- Postures in a gesture is not the same in the states number of gesture (Fig. 5a: postures number (X_{ij}) in hand gesture G_4 are difference lengths).
- Hand gestures are non-synchronization in phase (Fig. 5(a): first states long and short).
- Hand gesture is executed by many different people and each of them does different velocity at different times.
- Direction of hand gestures are distinguishing (Fig. 5(b))



(a) (b)

Figure 5. Spatial-Temporal features of hand gestures

Many approaches about hand gesture recognition that has been proposed in the related works. Hand gestures are mentioned in our database that has features: cyclical, repetitive about the shape, size of the gesture, the direction of movement in space, velocity. We divide these characteristics into two groups which can distinguish between various spatial-temporal data representing hand gestures. The first group is performed on a space invariant that includes presentations of cyclical repetition of shape, size. The second group is featured performers on a time interval invariant that includes the direction of movement and speed. With this unique group identification process we split into two hand gesture recognition parts: the spatial features recognition and temporal feature recognition which will be presented in detail in the following section.

E. Spatial hand gesture

Spatial features of a hand gesture which we focus on resolve some problems: changing of hand shapes, no synchronization and difference of length.

1. Hand shape in space features with PCA technical

To determine spatial of gestures, there're many approach as using PCA (Principal Component Analysis) space, matching algorithms as shape matching, contour shape, shape context, to warding the real-time hand gesture recognition so this step is low time cost that's still enough accuracy. We proposed using CPA method to reduce dimensions and present hand features space that has largest variable data for spatial features recognition. After hand image being segmented that will be converted to a gray image. Segmentation images of hand are not the same size, so it is resized into X ($p_{i,j}; i = 0 \div 64, j = 0 \div 64$). To reduce image distortion, size of image is larger than 64x64 pixels that will be scaled to 64 pixels. Size of image is smaller than 64x64pixels that will be scaled the same ratio and based on the center of the images. A row of X is template data, a column of X is feature. Data in rows of X aren't similar in amplitude so normalization is implemented by the 1 standard deviation that presents in (1):

$$X^* = \{x^*_{ij}\};\ x^*_{ij} = \frac{x_{ij} - g_j}{\sqrt{n}\sigma_j};\ g_j = \frac{\sum_{i=1}^{n} x_{ij}}{n} \qquad (1)$$

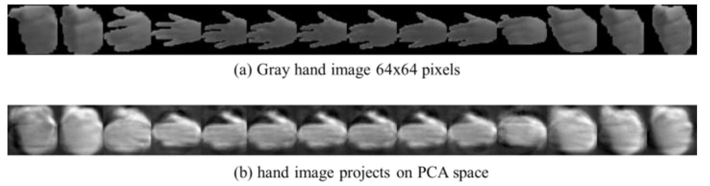
σ_j is standard deviation in column j of X matrix, n is rows number of X matrix. Hand images (X^*) is reshaped into matrix one row and 64x64cols that is Y ($q_i; i = 1 \div 4096$) that dates are still very big (4096 dimensions), as present in (2):

$$X = \begin{bmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,64} \\ x_{2,1} & x_{2,2} & \cdots & x_{2,64} \\ \vdots & \vdots & \vdots & \vdots \\ x_{64,1} & x_{64,2} & \cdots & x_{64,64} \end{bmatrix} => X^* = \begin{bmatrix} x^*_{1,1} & x^*_{1,2} & \cdots & x^*_{1,64} \\ x^*_{2,1} & x^*_{2,2} & \cdots & x^*_{2,64} \\ \vdots & \vdots & \vdots & \vdots \\ x^*_{64,1} & x^*_{64,2} & \cdots & x^*_{64,64} \end{bmatrix} \Rightarrow Y = [q_1 \quad q_2 \quad \cdots \quad q_{4096}] \qquad (2)$$

So using PCA to reduce correlation data between component from Y helps to reduce computational workload and still enough information that will be implement in training phase and testing phase:

- The first is training phase: a training hand gesture includes M hand postures $G_i = [Y_0 \quad Y_1 \quad \cdots \quad Y_M]^T$, a training hand gesture set includes N hand gestures: $G = [G_0 \quad G_1 \quad \cdots \quad G_N]^T$ that is input of PCA step and takes out feature space. In this research, PCA space is setup by 20 dimensions and feature space information PCA include: μ covariance matrix of Y_i vector in G_i, eigenvalues is λ, eigenvectors is e and H^* is projects of H in PCA space that fators of training hand gesture set will be saved out file. This file will be read when system restart to have the feature space.
- The second is testing phase: a gesture G has n postures Y_i ($i = 1 \div n$), in PCA space this gesture will be presented Y_i^*($i = 1 \div n$) as presents (3) and result of hand gesture in PCA space illustrates as Fig. 6:

$$G = \begin{bmatrix} Y^1 \\ Y^2 \\ \cdots \\ Y^n \end{bmatrix} = \begin{bmatrix} q_1^1 & q_2^1 & \cdots & q_{4096}^1 \\ q_1^2 & q_2^2 & \cdots & q_{4096}^2 \\ & & & \\ q_1^n & q_2^n & \cdots & q_{4096}^n \end{bmatrix} \Rightarrow (PCA) \Rightarrow G^* = \begin{bmatrix} Y^{*1} \\ Y^{*2} \\ \cdots \\ Y^{*n} \end{bmatrix} = \begin{bmatrix} q_1^{*1} & q_2^{*1} & \cdots & q_{20}^{*1} \\ q_1^{*2} & q_2^{*2} & \cdots & q_{20}^{*2} \\ & & & \\ q_1^{*n} & q_2^{*n} & \cdots & q_{20}^{*n} \end{bmatrix} \qquad (3)$$



(a) Gray hand image 64x64 pixels

(b) hand image projects on PCA space

Figure 6. Go left hand gesture before and after projects on PCA space

2. Synchronizing hand gestures with DTW technical

As Sec. D., in a dynamic hand gesture class, each person implements gestures with difference lengths. That has a difference number of postures and some gestures aren't the same phase with other gestures (hand closing and hand opening state). The phase synchronization between two dynamic hand gestures that is necessary. Many techniques have been devised to perform these tasks as using DTW (Dynamic Time Wrapping), HMM (Hidden Markov Model), ... Our proposed that matching is implemented by DTW algorithm that is applied to solve the problem enable matched two samples signal having with different lengths for small errors and real time.

DTW method optimums match between the two time series with some constraints. Two sequence hand gesture are stretched non-linear along the time axis order to determine the similarity between them. The simplest version of DTW can be implemented by this presentation: two hand gestures $G_1\{X_{11}, X_{12},\ldots X_{1n}\}$ and $G_2\{X_{21}, X_{22},\ldots X_{2m}\}$ that length is n and m. DTW method will indicate that each X_{1i} in the hand gesture G1 is matched positions in hand gesture G2. So, optimum matching pair is determined. This synchronization with some results as shown in Fig. 7. below and that the PCA feature space using has good performed information of the hand shape. Those results will be utilized that is an input for the temporal hand gesture step.

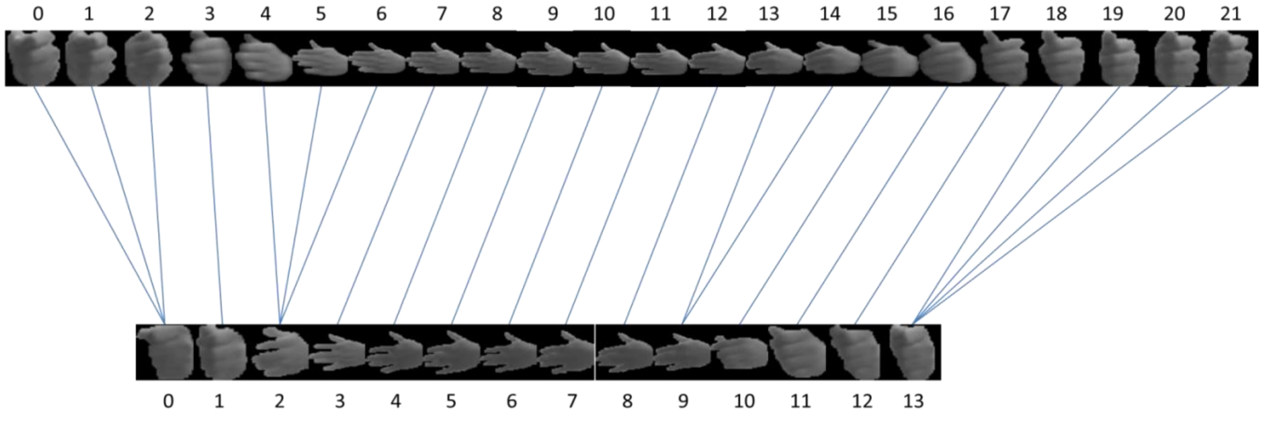


Figure 7. Result of DTW

F. *Temporal features of hand gestures*

1. Hand gesture trajectory with KLT:

Many proposed methods in last years that are temporal hand gesture recognition. After hand localization and giving results of DTW in the previous step, building and matching movement sequence. In our proposed method, the hand motion trajectory is implemented by KLT that combines the optical flow method of Lucas–Kanade[25] and the good feature points segmentation method of Shi–Tomasi[26]. The algorithm determines optical flow of Lucas-Kanade that based on three assumptions: the invariance of light intensity, the movement of hands in two consequence frames is small and Cohesion of space (the neighboring points on the same surface of the hand is the same motion). KLT help to trajectory of feature points of hand or calculates optical flow of hand between two sequence postures. At the first frame of hand gesture, feature points of hand posture will be segmented and this feature points will be trajectoried by the next posture to the end posture of gesture. So, each feature point creates a trajectory. If optical flow of two sequence postures less than 1, features is seem not movement and if optical flow of two sequence postures more than 50, features is seem not reliable. So this no movement points and no reliable points will be removed. If the feature points are less than threshold in a frame, that frame will segment some new feature points and this new feature points will trajectory in the next frame. Our research utilizes 20 feature points between two sequence posture hands and implements continuously from the start posture to the end posture of each gesture.

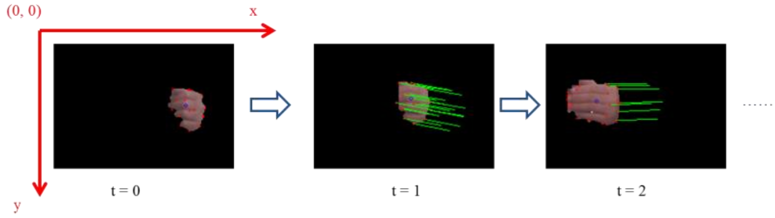


Figure 8. Optical flow of the go-left hand gesture

Giving optical flows of postures in each gesture, trajectory of feature points is built. A gesture has n frame (n posture) $f_1, f_2,..., f_N$, each f_i has K feature points at $p_{i,j}$, average of K feature points on x and y is $\overline{p_{\iota,J}}$. So a gesture presents by average of trajectories is $T = \left[\overline{p_{\iota,J}^1} \quad \ldots \quad \overline{p_{\iota,J}^n}\right]$. Which is more robust than trajectory base on centroid point of posture because the hand segmentation step has many noises on background. This Fig .9. illustrates trajectories of 20 feature points and an average trajectory of the go-right hand gesture in spatial-temporal coordinate. Red circles presents the feature point coordinates $p_{i,j}$ at frame t (t = 0 ÷ (n-1)). Blue square is presented by $\overline{p_{\iota,J}}$.

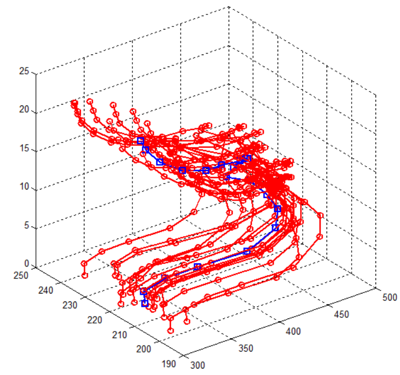


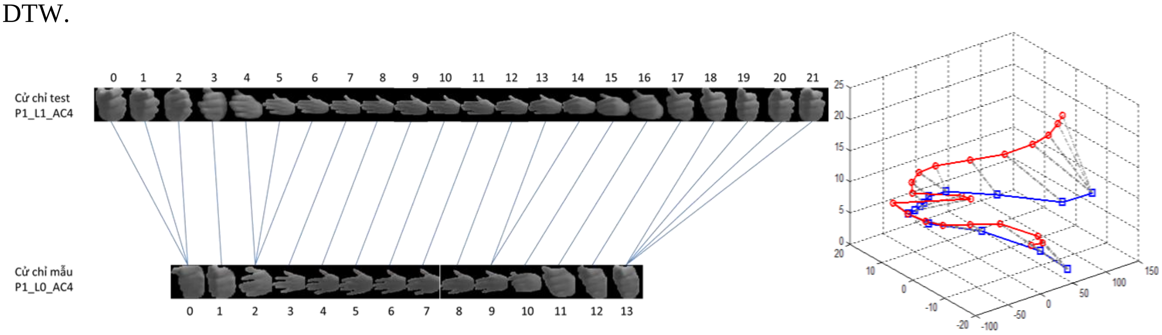
Figure 9. Trajectory of the go-right hand gesture

2. Similarity hand gestures:

Giving average trajectory of a dynamic hand gestures. Because hand position in each gesture isn't the same coordinate, so the first we have to normalize $T = \left[\overline{p_{i,j}^1} \quad \ldots \quad \overline{p_{i,j}^n}\right]$ on x and y dimension by $T^* = \left[\overline{p_{i,j}^1} - (\bar{x}, \bar{y}) \quad \ldots \quad \overline{p_{i,j}^n} - (\bar{x}, \bar{y})\right]$. In hand gesture recognition, a training set is P and a test set is T that is similar estimated by RMSE distance which calculates an error at $p_{i,j}$ of P and T which is presented by following (4) formula:

$$RMSE(T, P) = \sqrt{\frac{\sum_{i=1}\left(p_i(x,y) - q_i(x,y)\right)^2}{n}} \qquad (4)$$

But length of T and P aren't the same so a direction calculation doesn't implement (Fig .10a). Thanks to the link between postures (T,P) that are DTW results so estimating RMSE will become feasible (Fig .11a). The experimental results in Sec.IV will show the using RMSE is simple and separation results between the gestures is clear. If the RMSE values are smaller, two gestures (T, P) is more similar. This Fig .11a. illustrates the removing link of two trajectories, the blue path is trajectory of training gesture, the red path is trajectory of testing gesture, the gray path are links DTW.



(a) *(b)*

Figure 10. All trajectory link of two hand gesture (T,P); (a) DTW results (b) RMSE results

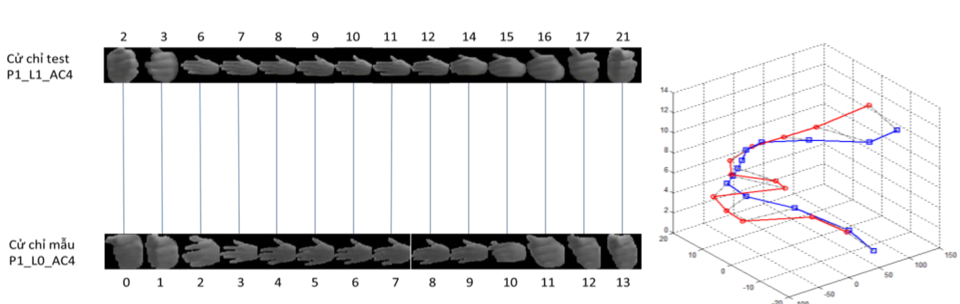


Figure 11. Trajectory link of two hand gesture (T,P) remove repetition link; (a) DTW results (b) RMSE results

IV. EXPRIMENTAL RESULT

We evaluate the proposed method in the effectiveness utilizing of the spatial-temporal features that can be distinguished between two hand gestures based on three following assessments: (1) utilizing cross-correlation and the average squared error RMSE; (2) changes shape but not change trajectory; (3) change trajectory but not changes shape. Using five dataset to evaluate that includes on-off (AC1), up (AC2), down (AC3). left (AC4) and right (AC5). Each dataset includes 10 templates that are implemented by two people (P1, P2), each person implements one command in 5 times (L1, L2, …, L5). So dataset has 50 templates. Hand gestures are labeled by: [P+order number]_[L+times]_[class lable:ACx]. Ex: P1_L1_AC4, P2_L3_AC5,…

1. Utilizing cross-correlation and the average squared error RMSE

In the first evaluation, the measurements performed using the cross-correlation value and RMSE between hand gesture test T and hand gesture training P. This assessment takes out that if only using cross-correlation between T and P that will be very difficult to distinguish between the two hand gestures; while combined DTW allows standardized length T and P, the assessment showed that RMSE is easy to distinguish between two different hand gestures. Fig .10. illustrates this:

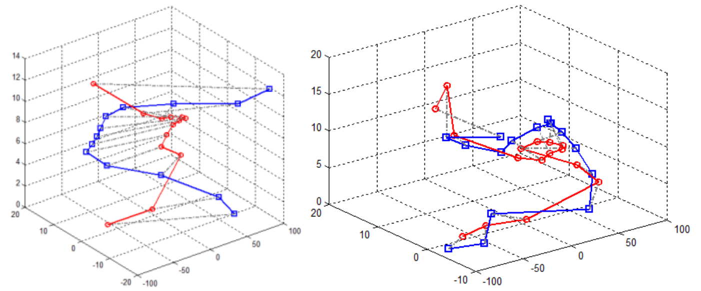


Figure 10. DTW and RMSE results between two hand gestures; (a) Two difference hand gestures; (b) Two same hand gestures

Table 1,2 illustrates results of cross-correlation and RMSE between about two hand gestures:

Table 1: P1_L3_AC5 with P1_L0_AC4

Cross-correlation	0.8522	0.6020
RMSE	83.1950	

Table 2: P1_L3_AC5 with P1_L0_AC5

Cross-correlation	0.8135	0.5921
RMSE	49.2284	

Table 3: Results between the test set T and traing set P

	P1_L1_AC4	P1_L1_AC5	P1_L3_AC4	P1_L3_AC5	P2_L2_AC4	P2_L2_AC5	P2_L3_AC4	P4_L3_AC4
P1_L0_AC4	**32.55**	86.41	**28.42**	83.19	**18.04**	82.9	**51.12**	**39.31**
P1_L0_AC5	101	**41.25**	120.72	**49.22**	102.16	**32.16**	112.79	118

Table 3 illustrates that if only utilizing cross-correlation between two gestures (T, P) that is difficult to distinguish two hand gestures even the trajectory are identified; Because of the difference length of two trajectory (T,P) so the cross-correlation calculating isn't effectiveness. Our proposed method utilized DTW results to take out hand posture links and combine RMSE that clearly distinguish between two difference hand gestures. As shown in Table 3, based on the minimum values (average is 36.5) when comparing testing hand gesture with training hand gesture that has 100% accurate results.

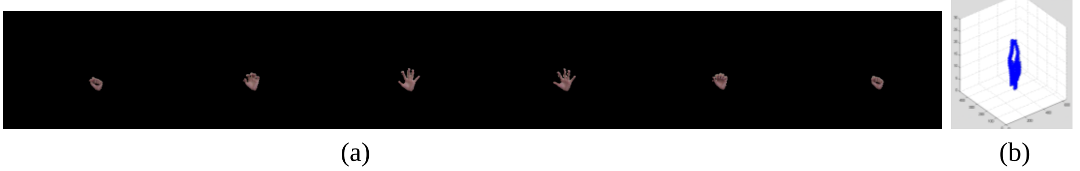
2. Changes shape but not change trajectory

This estimation is to check distinguishing of a dynamic hand gesture when the trajectory doesn't change but hand posture changes in time. Clearly illustration is distinguishing of close-open-close hand gesture that is on_off command. This dynamic hand gesture only changes hand shape but position of hand doesn't changing. Comparison results between on_off1 and on_off2 hand gestures with go_right and go_left hand gestures that is presented in following Table 4:

Table 4: RMSE of the changing and no changing trajectory of hand gestures

	on_off_1	on_off_2
AC4 (P1_L0_AC4)	44.73	43
AC5 (P1_L0_AC5)	59.05	66.49

While RMSE of on_off_1 and on_off_2 gesture is only 20.5 that helps to distinguish gestures.



(a) (b)

Figure 10. "on_off" hand gesture (a) changing on hand postures; (b) no changing on trajectory

3. Change trajectory but not changes shape

This estimation is to check distinguishing of a hand gesture when the trajectory changes but hand shape doesn't change in time. Those dynamic hand gestures are go_right_1 and go_right_2. It changes only position and doesn't change hand shape that result is illustrated in Fig .11:

Table 5: RMSE of the changing and no changing hand shape of hand gestures

	go_right_1	go_right_2
AC4 (P1_L0_AC4)	63.06	58.59
AC5 (P1_L0_AC5)	125.66	117.84

(a) (b)

Figure 11. "go_right_1" hand gesture (a) no changing on hand shape; (b) changing on trajectory (right_left)

The go_right_1 and go_right_2 that are the same trajectory with AC4 and not the same with AC5. So, RMSE value with AC4 are smaller than with AC5. Moreover, this values (63,06 and 58,59) are still more than average value of true hand gestures in table 3 (only 36.5). This shows that our system is still detecting and distinguishing if a person doesn't true implements a command.

Table 6 following presents confusion matrix of recognition results with fine dynamic hand gestures (left, right, up, down, on-off) that implements by 5 people and each dynamic hand gesture implements in 5 times. Each set has 25 dynamic hand gestures, evaluation method is "Leave-p-out-cross-validation" method (p=5) to separates training and testing data. Ultilizing k_NN with RMSE distance to recognize dynamic hand gestures with mean time cost is (178 ± 12)ms and the accuracy rate at (96% ± 2.5).

Table 6: Confusion matrix of five dynamic hand gesture recognitions

Dynamic hand gesture	AC5	AC4	AC3	AC2	AC1
AC5	**25**	0	0	0	0
AC4	0	**24**	0	0	**1**
AC3	0	0	**23**	0	**2**
AC2	0	0	0	**24**	**1**
AC1	0	0	0	0	**25**

V. CONCLUSION

This report described a vision-based hand gesture recognition system. Our proposed method utilized basic techniques of vision-based to be applied to recognition of hand gestures with based techniques: PCA, DTW, optical flow and RMSE. How to perform relatively simple, and effective. Initial results solve the requirement problems of our dynamic hand gesture recognition with accuracy approximates 96% and time rate approximates 178ms/frame. Thus, it is feasible to implement my recognition system to control the TV or indoor lighting system. In the future, we will continue improving, reducing time rate and perfecting the gesture recognition system that is built a controlling system equipments in the room using by the hand gesture recognition.

VI. REFERENCES

[1] Microsoft Kinect for Windows, http://www.microsoft.com/enus/kinectforwindows., November 2013.
[2] PointGrab Company, PointGrab brings gesture control to home appliances, (2013).
[3] P.Qifan, S.Gupta, S.Gollakota, and S.Patel, "Whole-Home Gesture Recognition Using Wireless Signals", In Proceedings of the 19th Annual International Conference on Mobile Computing and Networking, (2013).
[4] I.Bayer and T.Silbermann, "A multi modal approach to gesture recognition from audio and video data", In Proceedings of the 15th ACM on International conference on multimodal interaction, 461-466, (2013).
[5] Q.Chen, C.Joslin, and N.D.Georganas, "A Dynamic Gesture Interface for Virtual Environments Based on Hidden Markov Models", In Proceedings of IEEE International Workshop on Haptic Audio Visual Environments and their Applications, (2005).
[6] X.Chen and M.Koskela, "Online rgb-d gesture recognition with extreme learning machines", ACM on International conference on multimodal interaction, pp.467-474, (2013).
[7] S.Escalera, J.Gonz'alez, X.Bar´o, M.Reyes, O.Lopes, I.Guyon, V.Athitsos, and H.Escalante, "Multi-modal gesture recognition challenge 2013: Dataset and results", In Proceedings of the 15th ACM on International conference on multimodal interaction, pp.445–452, (2013).

[8] J.Choi and B.Seo, "Robust Hand Detection for Augmented Reality Interface", In Proceedings of the 8th International Conference on Virtual Reality Continuum and its Applications in Industry, (2009).
[9] F.Picard and P.Estraillier, "Motion Capture System Contextualization Application to Game Development", In Proceedings of Computer Games: AI, Animation, Mobile, Interactive Multimedia, Educational, Serious Games, (2009).
[10] MPD.Silva, V.Courboulay, and A.Prigent, "Gameplay experience based on a gaze tracking system", EURASIP Journal on Applied Signal Processing, (2007).
[11] K.Nandakumar, K.W.Wan, S.M.A.Chan, W.Z.T.Ng, J.G.Wang, and W.Y.Yau, "A multi-modal gesture recognition system using audio, video, and skeletal joint data", In Proceedings of the 15th ACM on International conference on multimodal interaction, pp.475-482, (2013).
[12] X.Zabulis, H.Baltzakis, and A.Argyros, "Vision-based Hand Gesture Recognition for Human Computer Interaction", Lawrence Erlbaum Associates, (2009).
[13] S.Rautaray and A.Agrawal, "Vision based hand gesture recognition for human computer interaction", a survey, Artificial Intelligence Review (2012).
[14] X.Deyou, "A Network Approach for Hand Gesture Recognition in Virtual Reality Driving Training System of SPG", International Conference on Pattern Recognitioin, pp.519-522, (2006).
[15] M.Elmezain, A.Al-Hamadi, and C.Michaelis, "Real-Time Capable System for Hand Gesture Recognition Using Hidden Markov Models in Stereo Color Image Sequences", Journal of WSCG 16, pp.65–72, (2008),.
[16] M.Elmezain, A.Al-Hamadi, and J.Appenrodt, "A Hidden Markov Model-based continuous gesture ecognition system for hand motion trajectory", International Conference on Pattern Recognition, pp.1–4, (2008).
[17] D.Kim, J.Song, and D.Kim, "Simultaneous Gesture Segmentation and Recognition Based on Forward Spotting Accumlative HMMs", Journal of Pattern Recognition Society 40, pp.1–4, (2007).
[18] A.Oikonomopoulos, I.Patras, and M.Pantic, "Spatiotemporal salient points for visual recognition of human actions", IEEE Transactions, pp.710–719, (2005).
[19] K.Takahashi, S.Sexi, and R.Oka, "Spotting Recognition of Human Gestures From Motion Images", In Technical Report IE92-134, pp.9–16, (1992),.
[20] J.Alon, V.Athitsos, Y.Quan, and S.Sclaroff, "A Unified Framework for Gesture Recognition and Spatiotemporal Gesture Segmentation", IEEE Transactions on Pattern Analysis and Machine Intelligence 31, pp. 1685–1699, (2009).
[21] H.Yang, S.Scharoff, and S.Lee, "Sign Language Spotting with a Threshold Model Based on Conditional Random Fields", IEEE Transaction on Pattern Analysis and Machine Intelligence 31, pp.1264–1277, (2009),.
[22] J.Lafferty, A.McCallum, and F.Pereira, "Conditional Random Fields: Probabilistic Models for Segmenting and Labeling sequence Data", Internation Conference on Machine Learning, pp.282–289, (2001).
[23] M.Elmezain, A.Al-Hamadi and B.Michaelis, "Discriminative models-based hand gesture recognition", International Conference on Machine Vision, pp.123–127, (2009).
[24] Huong-Giang Doan, Hai Vu, Thanh Hai Tran, Eric Castelli, "Improvements of RGB-D Hand Posture Recognition Using an User-Guide Scheme", In Proceeding(s) of the 7th IEEE International Conference on Cybernetics and Intelligent Systems (CIS) and the 7th IEEE International Conference on Robotics, Automation and Mechatronics (RAM), (2015).
[25] Bruce D. Lucas and Takeo Kanade, "An Iterative Image Registration Technique with an Application to Stereo Vision", Proc. International Joint Conference on Artificial Intelligence, pages 674–679, 1981.
[26] J.Shi and C.Tomasi, "Good Features to Track", IEEE Conference on Computer Vision and Pattern Recognition, pages 593–600, (1994).
[27] J.Listgarten, "Analysis of sibling time series data: alignment and difference detection", Ph.D. thesis, University of Toronto, (2006).
[28] J.Rob Hyndman, B.Anne Koehler, "Another look at measures of forecast accuracy", International Journal of Forecasting 22(4), pp.679 – 688, (2006).

NHẬN DẠNG CỬ CHỈ ĐỘNG CỦA BÀN TAY SỬ DỤNG CÁC ĐẶC TRƯNG KHÔNG GIAN VÀ THỜI GIAN

Đoàn Thị Hương Giang, Vũ Duy Anh, Vũ Hải, Trần Thị Thanh Hải

Tóm tắt*— Nhận dạng cử chỉ tay đã được nghiên cứu trong thời gian dài, Tuy nhiên nó vẫn còn là một lĩnh vực còn nhiều thách thức. Hơn nữa, nhận dạng cử chỉ tay để điều khiển các thiết bị phòng thông minh như tivi, quạt, camera, cửa,... đòi hỏi phải có độ chính xác nhận dạng cao. Bài báo này đề xuất một mô hình đơn giản và hiệu quả để nhận dạng bộ cơ sở dữ liệu đã được định nghĩa để điều khiển các thiết bị điện. Việc phân tích các đặc trưng không gian và thời gian bao gồm tính chu kỳ lặp lại của các cử chỉ tĩnh trong mỗi cử chỉ động, sự khác nhau về độ dài của các cử chỉ động, sự không đồng bộ về pha giữa các cử chỉ động, sự thay đổi về hình trạng tay và các đặc trưng về hướng cũng như sự di chuyển của tay của bộ cơ sở dữ liệu đã định nghĩa. Sau đó nhận dạng các cử chỉ dựa trên các đặc trưng không gian và thời gian này.*