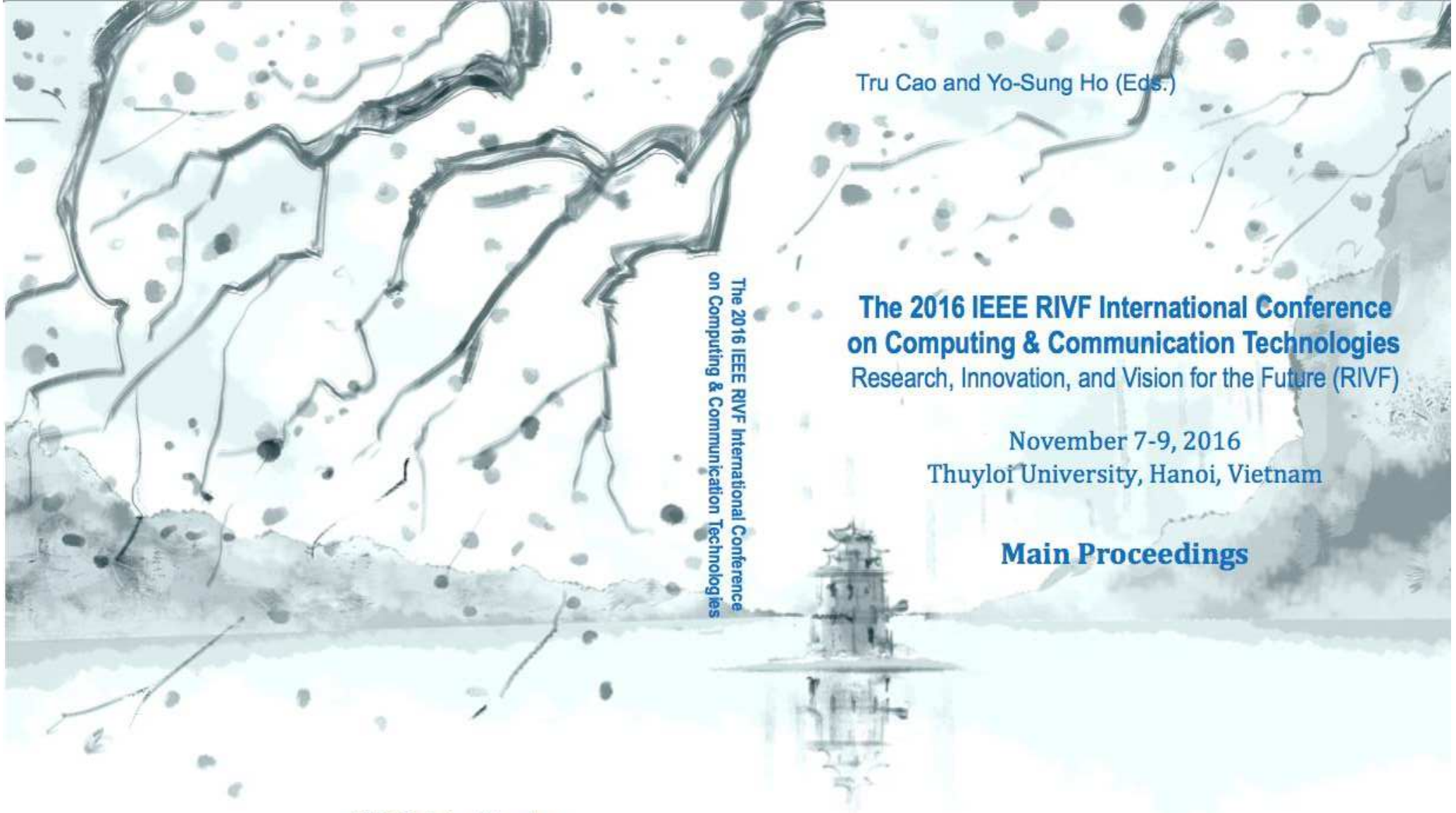

Tru Cao and Yo-Sung Ho (Eds.)

The 2016 IEEE RIVF International Conference on Computing & Communication Technologies

Research, Innovation, and Vision for the Future (RIVF)

November 7-9, 2016
Thuyloi University, Hanoi, Vietnam

Main Proceedings

The 2016 IEEE RIVF International Conference on Computing & Communication Technologies

IEEE Catalog Number
Part number: CFP1656A-PRT
ISBN: 978-1-5090-4133-6

IEEE

2016 IEEE RIVF
International Conference on Computing & Communication Technologies

Research, Innovation, and Vision for the Future (RIVF)

November 07-09, 2016
Thuyloi University, Hanoi, Vietnam

Main Proceedings

Organizer
IEEE Vietnam Section

Technical Sponsors
IEEE Communications Society
IEEE Computational Intelligence Society

Editors
Tru Cao and Yo-Sung Ho

IEEE Catalog Number (Print): CFP1656A-PRT
ISBN (Print): 978-1-5090-4133-6

IEEE Catalog Number (USB): CFP1656A-USB
ISBN (USB): 978-1-5090-4132-9

IEEE Catalog Number (Xplore compliant): CFP1656A-ART
ISBN (Xplore compliant): 978-1-5090-4134-3

Phase Synchronization in a Manifold Space for Recognizing Dynamic Hand Gestures from Periodic Image Sequence

Huong-Giang Doan*†, Hai Vu*, Thanh-Hai Tran*,
*International Research Institute MICA, Hanoi University of Science and Technology
†Industrial Vocational College Hanoi
Email: {huong-giang.doan,hai.vu,thanh-hai.tran}@mica.edu.vn

***Abstract*—Phase synchronization issue, that is caused by spotting gestures from video stream, varying frame-rates, speed of subject's implementation, should be overcome in developing Human-Computer Interaction (HCI) application using dynamic hand gestures. This paper tackles an interpolation technique to efficiently solve this issue. We firstly propose a new representation of dynamic hand gestures space that consists of both spatial and temporal features extracted from the hand gestures. The spatial features are extracted based on a manifold learning technique (ISOMAP) that takes into account non-linear features (e.g., poses of hand, illumination conditions, hand-shape differences). The temporal features handle hand movements thanks to Kanade-Lucas-Tomasi (KLT), good feature points tracking algorithm. We then propose an efficient interpolation scheme on the constructed space of hand gestures. This scheme ensures inter-period phase continuity as well as normalizes length of the hand gestures. We examine the proposed method with three different large datasets of dynamic hand gestures. Evaluation results confirm that the best accuracy rate achieves at 98% that is significantly higher than results from previous works (at 94%). The proposed method suggests a feasible and robust solution addressing technical issues in developing HCI application using the hand gestures to control home appliance devices.**

I. Introduction

Designing and developing dynamic hand gesture recognition systems have been stimulated for home appliances [1] thanks to recent advantages of machine learning algorithms as well as achievement of RGB and Depth sensors (e.g., Microsoft Kinect [2]). To achieve robust and accurate systems, there are technical issues that should be overcome. Phase synchronization, that ensures consistent matching of a pair between probe and gallery image sequences, is one of the typical problems. Inconsistent phasing appears due to spotting gestures from the video stream. Although many video segmentation techniques could be applied [1], [3], [4], most of them require pre-determined thresholds to cut starting and ending points. As a result, gesture image sequences are different in length. Another reason could be the variation in speed of subject's implementation and or video stream is captured at different frame rates. To solve this issue, Dynamic Time Warping (DTW) usually is utilized for registering two temporal sequences. Other approaches are inspired by temporal interpolation techniques so that the interpolated videos are normalized by a pre-determined length of the sequence. This paper argues two above-mentioned techniques addressing the phase synchronization issues, which gives better performance for a dynamic hand gesture recognition system.

In this paper, the problem of phase synchronization is constrained on *periodic* hand gesture image sequences. The periodicity hypothesis is based on characteristics of the designed gestures which an end-user changes hand shapes in a cyclical pattern during a hand movement in our previous work [1]. Along a so-called hand-path (trajectory), moving direction represents the meaning of a gesture corresponding to a control command to the home appliance. They are five dynamic gestures: on/off, left, right, up and down (e.g., Fig. 2 shows on/off gesture). For such dataset, the practical issues relating to the phase synchronization consist of: too short gestures (only a few postures) or too long gestures (with many hand postures), non-uniform sampling rate or motion fluctuations appearing in a periodic hand gesture. These issues could make a phase registration technique discarding useful cues due to inter-period phase continuity. In our previous work [1], DTW algorithm was adopted to align two hand-shape image sequences. A pair of hand shapes is compared through a (dis)similarity measurement in which hand shapes are projected into a PCA (Principal Component Analysis) space. Because of frame-to-frame comparisons, DTW-based approaches could not infer phase continuity. As a consequence, recognition rate is degraded due to missed-alignment of two sequences. In this study, we deal with the phase synchronization issues by inspiring a temporal interpolation technique. The key idea is that a hand gesture sequence is normalized by a pre-defined length of the sequence. The proposed interpolation scheme is based on evaluating similarities on whole sequences. Instead of aligning a pair of hand shapes, we solve the missed phase for the whole sequence of frames to take into account the inter-period phase continuity.

To this end, hand shapes are exploited through an isometric feature mapping algorithm (ISOMAP [5]). ISOMAP is one of the most basic manifold learning algorithms that tries to preserve a different geometrical properties at all scales. We firstly represent hand gesture sequences in a new space whose dimensions are combined from the most important features extracted from ISOMAP and the hand trajectory. Given a dynamic gesture, we deploy the proposed interpolation scheme on each dimension of this new space to reconstruct a new

978-1-5090-4134-3/16/$31.00 ©2016 IEEE

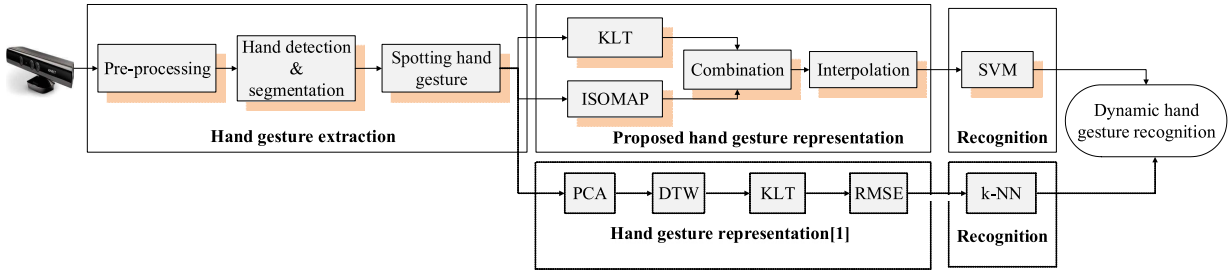

Fig. 1. The proposed framework of hand gesture recognition

image sequence with the pre-determined number of frames. The support vector machine (SVM) technique [6] is utilized to assign gesture label of the interpolated sequence. We evaluate performance of the proposed approaches by comparing with DTW-based approaches on three datasets presented in [1] and [7]. The achieved performance is very competitive.

The rest of paper is organized as follows. Section II surveys related techniques of the dynamic hand gestures recognition. Section III describes the proposed techniques. Section IV reports the experimental results. Finally, Section V concludes works and suggests further research directions.

II. Related work

There are uncountable solutions for developing a vision-based hand posture/gesture recognition system in the literature. Readers can refer good surveys such as [8], [9], [10], [11]. For detecting and recognizing hand gestures from a video stream, most of the related works have to deal with common issues such as the complexity of hand shapes, a variation of gesture trajectories, cluttered background, light conditions, changing velocity, and missed phases in the dynamic gestures. The phase synchronization issue has been particularly interested in many relevant works. For example in [12], [13], [14], the authors proposed to use Dynamic Time Warping (DTW) technique. The DTW is adopted from time series analysis domain. Additionally, Hidden Markov Model (HMM) and its variant are preferred to solve state issues what appear in an image sequence of the hand gestures.

The temporal and/or spatial interpolation has been developed and considered in many topics of computer vision such as video editing, video compression or matching/recognizing dynamic action, and so on. The interpolation idea has been deployed based on the example-based methods. These approaches typically try to generate a high frame-rate video from a single/multiple low frame-rate video [15], [16]. Solution [17] enhanced the spatial and/or temporal resolution of videos. Another spatial-temporal resolution issue [18], [19] from two sequences that ones with high resolution and low frame rate, the other with low resolution and high frame rate. Temporal interpolation is utilized in image space [20]. In particular, an interpolation approach [21], [22] deals with low frame-rate videos for gait recognition. They proposed techniques for creating a periodic temporal super resolution. In this work, we inspire a temporal resolution technique from raw dynamic gestures. The interpolated video sequences are considered as normalized sequences in terms of length of the sequence. Its performance is compared with conventional phase synchronization technique that is based on DTW algorithm.

III. Proposed approach

Our proposed framework, as shown in Fig. 1, composes of three main components: hand gesture extraction, hand gesture representation and recognition. Each component consists of a series of cascaded procedures. First, to extract the hand gestures from a video stream, we rely on techniques presented in [1]. To make paper be consolidated, we summarize briefly these steps in Sec. III-A. In the upper panel of Fig. 1, we propose a new approach for gesture representation with phase synchronization. To be more easily comparing with [1], the corresponding steps in [1] are shown in the lower panel. Instead of using PCA as [1] for a linear dimension reduction, we use non-linear technique (ISOMAP). We then combine the most important features extracted from ISOMAP space (spatial features) with movement features based on KTL (temporal features) to create gesture representation. This feature vector is finally interpolated to obtain the one with a desirable temporal resolution. Finally, SVM classifier is applied to predict the label of gesture. In Sec. III-B, we present in detail sub-steps of our proposed hand gesture representation.

A. Brief summary of hand gesture extraction

1) Pre-processing: Depth and RGB data captured from the Kinect sensor [2] that are not measured from the same coordinate system. In the literature, the problem of calibrating depth and RGB data has been mentioned in several works for instance [9]. However, the calibration method of Microsoft is utilized due to its availability and ease to use.

2) Hand detection and segmentation from still images: First, human body region is separated from background using a background subtraction technique. A Gaussian Mixture Model [23] is adopted because this technique is real-time computation and achieve reliable results in [1]. From extracted human body, hand candidates are continuously extracted based on the distribution of depth image [24].

3) Hand gesture spotting: A dynamic hand gesture is an image sequence consisting of consecutive hand postures. Length of such sequences may vary in time. In a real application, it is a critical task to determine the starting and ending

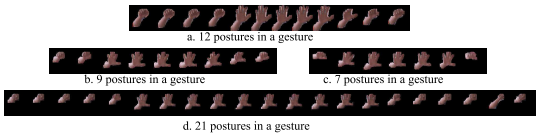

Fig. 2. Examples of gesture spotted from continuous sequences of frames.

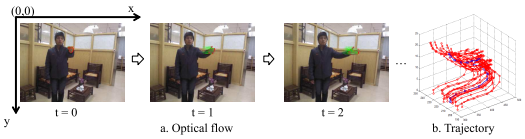

Fig. 3. Optical flow and Trajectory of the go-right hand gesture.

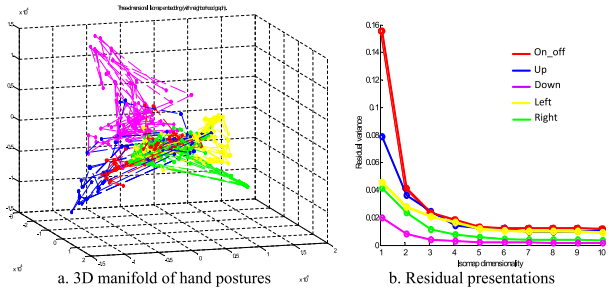

Fig. 4. a) 3D manifold of hand postures belonging to five gesture classes. b) Residual R_d of each hand gesture class. Each color presents one gesture class.

instances of a hand gesture. In this study, all pre-defined gestural commands have the same hand shape/posture at starting (opened hand) and ending times (closed hand). Moreover, hand shapes appearance within a gesture are underlying a cyclical pattern. These properties are utilized to deploy gesture spotting techniques. Fig. 2 shows some examples in which hand gestures are spotted in different length or duplicating frame at starting and ending points. (e.g., the sequences in Fig. 2(b)(c) are more shorter sequence than in Fig. 2(a)(d), there are many duplicating frames at start/ending point in Fig. 2(d)).

B. Hand representation from spatial and temporal features

In common situation, a dynamic hand gesture consists of consecutive hand postures which are variant in both temporal and spatial dimensions. We propose a new representation of dynamic hand gesture combining both spatial and temporal features. Sections below will explain how these features are extracted.

1) Temporal features extraction for characterizing hand movement: In the last years, many methods have been proposed for extracting temporal features of human actions. Aiming at stable and real-time algorithm, we select KLT (Kanade-Lucas-Tomasi) tracker to extract hand movement trajectory. This technique combines the optical flow method of Lucas-Kanade [25] and the good feature points detection method of Shi-Tomasi [26]. It was widely utilized in the literature for object tracking, motion representation. In this study, KLT is done through following steps: First, we detect good feature points on the current frame of the image sequence. Then we track these good feature points in the next frame. This is repeated until the end of the frame sequence. Connecting tracked points in the consecutive frames creates a trajectory of the hand movement. Because many trajectories are generated at a certain time, we select twenty most significant ones to represent a gesture. Each trajectory composes of K good feature points $\{p_1, p_2, ..., p_K\}$. Each point p_i has coordinates (x_i, y_i). Taking average of all points gives a average trajectory $G = [\overline{p_1}, \overline{p_2}, ..., \overline{p_K}]$. This average trajectory represents a representative gesture's direction. Those are the temporal features extracted from an image sequence of a hand gesture. Therefore, the temporal feature Tr_N^G of each dynamic hand gesture G is presented as the following (1);

$$Tr_N^G = \{(x_1, y_1), (x_2, y_2), ..., (x_N, y_N)\} \quad (1)$$

Figure 3(a) illustrates tracked points from several frames. Fig .3(b) illustrates trajectories of twenty feature points and the average trajectory of a $Right$ gesture in spatial-temporal coordinate. Red circles present coordinates of the good feature points p_i at frame i $(i = [1, N])$. Blue squares represent the average $\overline{p_i}$.

2) Spatial features extraction for characterizing hand shapes: A hand gesture composes of many hand postures, and as the results, a hand gesture creates a high-dimension feature space. In our previous work [1], we used PCA, a linear dimension reduction technique, to reduce the dimension of a static hand posture. However, dynamic hand gestures are rendered from various poses of hand, illumination conditions, hand-shape differences. These factors which non-linear features require more sophisticated algorithms to properly extract features/manifold from an original feature space. Difference from [1], in this study, we construct a low-dimension space of hand postures by utilizing a manifold learning technique, that aims to deploy a non-linear dimensionality reduction [27]. The manifold learning algorithms try to preserve a different geometrical property of the underlying manifold. Popular non-linear manifold learning algorithms include the isometric feature mapping (ISOMAP) algorithm [5], the locally linear embedding (LLE) algorithm [28], and the Laplacian eigenmaps (LE) algorithm [29]. The ISOMAP technique is selected in this study because it captures better intrinsic structures of hand postures. The ISOMAP technique preserves the best geodesic distances between any two data points in the original high-dimensional space [5]. It is simple to implement and runs much faster than other manifold learning techniques.

Given a set of N segmented postures $\boldsymbol{X} = \{X_1, X_2, ..., X_i, ..., X_N\}$ whose element X_i is of different size. We resize all of them to same size of (100x100) pixels then reshape each to a row vector of size $(1, 10000)$. The input set of hand postures becomes $\boldsymbol{Z} = \{Z_1, Z_2, ..., Z_i, ..., Z_N\}$ where $Z_i = reshape(X_i', 1, 10000), X_i' = resize(X_i, 100, 100)$.

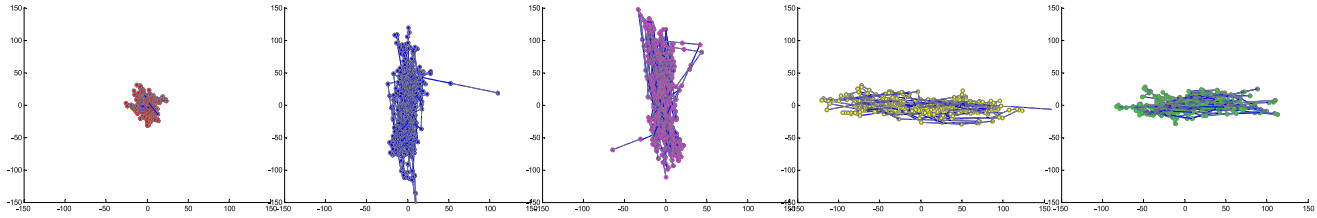

Fig. 5. Distribution of dynamic hand gestures in the low-dimension.

The ISOMAP algorithm takes $\boldsymbol{Z}$ as input and computes the corresponding coordinate vectors $\boldsymbol{Y} = \{Y_i \in R^d, i = 1, ..., N\}$ in the d-dimensional manifold space $(d << D)$. The ISOMAP algorithm comprises the following three steps: constructing neighborhood graph, computing the pairwise geodesic distances and building d-dimensional embedding. An important issue concerning the ISOMAP algorithm is how to determine the dimension d of ISOMAP space. The residual variance R_d is used to evaluate the error of dimensionality reduction between the geodesic distance matrix G and the Euclidean distance matrix in the d-dimensional space D_d. We analyze the residual error R_d as shown in Fig. 4(b) and observe that when $d > 3$, the residual error does not reduce significantly for all five gesture classes. Therefore, $d = 3$ is chosen to extract three most significant dimensions for hand posture representations. Three first components in the manifold space are extracted as spatial features of each hand shape/posture. A dynamic hand gesture then is represented as following:

$$\boldsymbol{Y}_N^G = \{(Y_{1,1}, Y_{1,2}, Y_{1,3}), (Y_{2,1}, Y_{2,2}, Y_{2,3}), ..., (Y_{N,1}, Y_{N,2}, Y_{N,3})\} \quad (2)$$

Figure 4(a) illustrates 3-D manifolds of five different hand gestures. Each manifold traces a closed non-linear curve in the embedded space. The curves intersect at starting and ending points of each dynamic gesture. The fact that all five classes have the similar starting and ending hand shape so they are projected at the same regions in the manifold space. Other parts of the curves are distinguished from one gesture to others. Inter-class variances obviously are more increased through the proposed manifold space.

3) Combination of the spatial and temporal features: The extracted spatial and temporal features are combined to completely represent dynamic hand gestures. We define a dynamic hand gesture G^{TS} with N images as (3). Where (x_i, y_i) are taken from (1); $Y_{i,j}$ taken from (2); $i = 1..N; j = 1..3$:

$$\boldsymbol{G}^{TS} = \begin{bmatrix} P_1 & P_2 & ... & P_n \end{bmatrix} = \begin{bmatrix} x_1 & x_2 & ... & x_N \\ y_1 & y_2 & ... & y_N \\ Y_{1,1} & Y_{2,1} & ... & Y_{N,1} \\ Y_{1,2} & Y_{2,2} & ... & Y_{N,2} \\ Y_{1,3} & Y_{2,3} & ... & Y_{N,3} \end{bmatrix} \quad (3)$$

Figure 5(a)-(e) illustrates new representations in 3-D space of five different hand gestures. In comparison with Fig. 4,

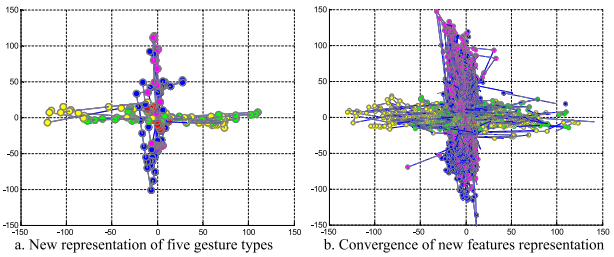

Fig. 6. Five dynamic hand gestures in the 3D dimension.

separation between five gestures are clearer than that is presented in Fig. 4. The main reason is that the extracted temporal features are embedded in this new space/representation. Fig. 6 confirms inter-class variances when whole dataset is projected in the proposed space. In particularly, cyclic patterns of the hand gestures are presented as closed-circles. The $Turn\ on\_off$ gestures consist circles around the zero point. The Up dynamic gestures are presented in cyan. The $Down$ dynamic gestures are presented in magenta curves. The $Left$, $Right$ dynamic gestures are presented in red, and green curves, respectively.

C. Phase synchronization using hand posture interpolation

By utilizing both spatial and temporal features to represent a dynamic hand gesture, comparison between two gestures, e.g., one from gallery and another from probe, could be straightforward implementations (e.g., using cross-correlation, RMSE measurements). However, inter-period phase would be discarded. In other words, periodic pattern of image sequence has been omitted. To overcome this issue, we deploy an interpolation scheme so that hand gesture sequences have same length, and maximize inter-period phase continuity. We propose a scheme based on piecewise interpolation and similarity measurement between two adjacent points in the proposed hand gesture space. Supposing M is the desired length for each gesture, given $\boldsymbol{G}^{TS} = \{P_1, P_2, ..., P_N\}$ at time instances $(t_1, t_2, ..., t_N)$ respectively, a distance vector of $\boldsymbol{G}^{TS}$ is calculated by $\boldsymbol{D_{inter}} = \{d_i; (i = 1, ..., N-1)\}$ where $d_i = ||P_i - P_{i+1}||_2$ is Euclidean distance between two consecutive postures P_i and P_{i+1}.

In case a dynamic gestures consists of N hand postures which is lower than the pre-defined length M $(N < M)$,

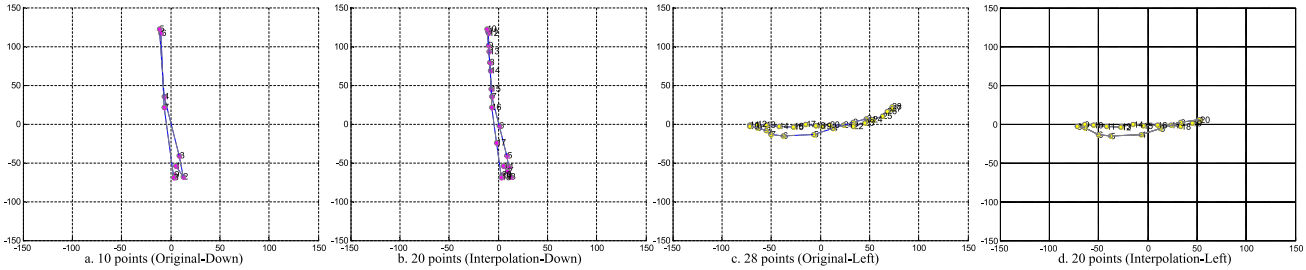

Fig. 7. a, c) Original hand gestures. b,d) corresponding interpolated hand gestures

we find a maximal distance from vector $\boldsymbol{D_{inter}}$ ($d_{max} = \max(D_{inter})$). This furthest point is the first priority to do the interpolation. Then the interpolated point is inserted between them. Denoting $P_i = [x_i, y_i, Y_{i,1}, Y_{i,2}, Y_{i,3}]^T$ and $P_{i+1} = [x_{i+1}, y_{i+1}, Y_{i+1,1}, Y_{i+1,2}, Y_{i+1,3}]^T$ are two furthest points, a new point P^* is inserted as defined in (4):

$$P^* = [\frac{x_{i+1}-x_i}{2}, \frac{y_{i+1}-y_i}{2}, \frac{Y_{i+1,1}-Y_{i,1}}{2}, \frac{Y_{i+1,2}-Y_{i,2}}{2}, \frac{Y_{i+1,3}-Y_{i,3}}{2}]^T \quad (4)$$

The length of the new sequence after inserting P^* is $N+1$. This procedure is iterated until the sequence length reaches M postures.

In case $N > M$, we find a minimal value of vector $\boldsymbol{D_{inter}}$ ($d_{min} = \min(D_{inter})$) between two nearest points, supposing P_i, P_{i+1}. We then eliminate one from these two points as follows (5):

$$P_{removed} = \begin{cases} P_i & [(d_{i-1} < d_{i+1}) \& (i \neq N-1)] or [(i=1)] \\ P_{i+1} & [(d_{i-1} > d_{i+1}) \& (i \neq 1)] or [(i=N-1)] \end{cases} \quad (5)$$

Figure 7 presents some results of the interpolation procedure with the same length of sequence M is equal to 20. The number of postures in Fig. 7 (a) is equal to 10. In Fig. 7 (c), the number of postures is up to 28. Fig. 7 (b),(d) are two interpolated hand gestures after applying the interpolation procedure.

D. SVM-based hand gesture recognition

Dynamic hand gesture recognition is performed by using multi-class SVM classifier. The input of multi-class SVM classifier is feature vectors extracted from interpolated sequence $F(1, 5 * M)$ as defined in (6):

$$F = [x_1, y_1, Y_{1,1}, Y_{1,2}, Y_{1,3}, ..., x_M, y_M, Y_{M,1}, Y_{M,2}, Y_{M,3}]^T \quad (6)$$

The output of multi-class SVM will be one value among $\{0, 1, 2, 3, 4\}$ corresponding to the gesture elements: $\{Turn\ On\_off, Up, Down, Left, Right\}$.

IV. Experimental results

The proposed framework is warped by a C++ program and a Matlab program on a PC Core i5 3.10GHz CPU, 4GB RAM. We evaluate performance of the hand gesture recognition on three different datasets: $MSRGesture3D$ [7]; $MICA1$ and $MICA2$ [1]. We conduct two evaluations: The performance of the proposed method when temporal resolution M is changed,

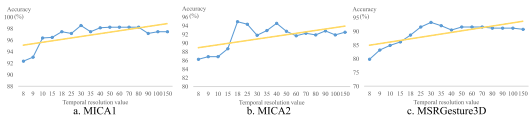

Fig. 8. The dynamic hand gesture recognition results with the difference interpolation thresholds.

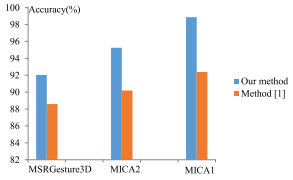

Fig. 9. The best dynamic hand gesture recognition results

and the accuracy rate of the hand gesture recognition system using optimal values of M. The accuracy rate is the ratio between the numbers of true positives per total number of hand gestures used in testing.

A. Influence of temporal resolution on recognition accuracy

In this evaluation, we test the accuracy rate with various values of the temporal resolution threshold M. M is fluctuated from 8 to 150 frames for each dynamic hand gesture. The accuracy rates are illustrated in Fig. 8(a)-(c), that show results on $MICA1$, $MICA2$ and $MSRGesture3D$ datasets, respectively. As shown, if M value is small, hand gesture recognition result is degraded. Performance are saturated when M is equal to 18 frames per one dynamic gesture for $MICA1$ and $MICA2$ dataset, and M is equal to 30 frames for $MSRGesture3D$ dataset [7].

B. Evaluation of hand gesture recognition

We follow *Leave-p-out-cross-validation* method, with p equals 1. It means that gestures of one subject are utilized for testing and the remaining subjects are utilized for training. The recognition results are given in Fig. 9. Averagely, by using the optimal parameters M, as suggested in Fig. 8, the proposed method obtains the accuracy rate at 92.03 $\pm$ 5.16 % with $MSRGesture3D$, 97.95 $\pm$ 3.09 % with $MICA1$ and 94.95 $\pm$ 4.65 % with $MICA2$. For the public dataset $MSRGesture3D$, this performance is slightly higher than

the results in the current state-of-art works (e.g., [30] reported 87.7%, [31] is 88.5%, [32] ups to 92.45% and [33] obtained 94.72%. For the same datasets $MICA1$and $MICA2$, the proposed method is clearly better than [1]. [1] obtained 92.45 $\pm$ 2.75 % with $MICA1$ and 90.08 $\pm$ 1.83 % with $MICA2$. Fig. 9 visually compares these results in which blue columns present the results of the proposed method and red columns present the results of the DTW-based approaches. Obviously, the proposed method that ensures the inter-period phase continuity, is over perform the DTW-based approaches [1], that discarded useful cues along temporal dimension.

V. Discussion and Conclusion

Discussion

Our proposed system has a few limitations. The current results could not confirm that this method is adaptive with dynamic hand gestures that include only one, two frames or too many duplicates frames at a certain phase, such as starting or ending point. Moreover, the proposed method needs to be more confirmed that it is a robust and tolerance system with changing of subject positions and/or difference hand directions. Consequently, these limitations suggest us directions to future works.

Conclusion This paper described a new representation hand gesture that combines the temporal and spatial features. The spatial feature is achieved by the manifold learning and the temporal feature is obtained by the KLT algorithm. We also presented the interpolation method for a high temporal resolution from a low temporal resolution of the hand gestures. This interpolation resolves the phase synchronization issue to enhance recognition rate of the dynamic hand gestures. The experimental results confirmed that accuracy of the proposed algorithm is enhanced. Which result is higher than the previous methods. The proposed technique is feasible to deploy real applications to control home appliance devices.

VI. Acknowledgment

This research is funded by Hanoi University of Science and Technology under the Project T2016-LN-27 “Development of a multi-modal system for controlling light appliance”.

References

[1] H. Doan, H. Vu, and T. Tran, “Recognition of hand gestures from cyclic hand movements using spatial-temporal features,” in *SoICT, Hue, Vietnam*, 2015, pp. 260–267.

[2] “http://www.microsoft.com/en-us/kinectforwindows.”

[3] M. Elmezain, A. Al-Hamadi, and B. Michaelis, “A novel system for automatic hand gesture spotting and recognition in stereo color image sequences,” *WSCG*, vol. 17, no. 1-3, pp. 89–96, 2009.

[4] P. Morguet and M. Lang, “Spotting dynamic hand gestures in video image sequences using hidden markov models,” in *ICIP*, 1998, pp. 193–197 vol.3.

[5] J. B. Tenenbaum, V. de Silva, and J. C. Langford, “A global geometric framework for nonlinear dimensionality reduction,” *Science*, vol. 290, no. 5500, pp. 2319–2323, 2000.

[6] C. J. C. Burges, “A Tutorial on Support Vector Machines for Pattern Recognition,” vol. 43, pp. 1–43, 1997.

[7] “http://research.microsoft.com/en-us/um/people/zliu/actionrecorsrc/.”

[8] X. Zabulis, H. Baltzakis, and A. Argyros, *Vision-based Hand Gesture Recognition for Human Computer Interaction*. Lawrence Erlbaum Associates, 2009.

[9] S. Rautaray and A. Agrawal, “Vision based hand gesture recognition for human computer interaction: a survey,” *Artif. Intel. Rev.*, vol. 43, pp. 1–54, 2015.

[10] 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,” *WSCG*, vol. 16, pp. 65–72, 2008.

[11] H. Yang, S. Scharoff, and S. Lee, “Sign Language Spotting with a Threshold Model Based on Conditional Random Fields,” *TPAMI*, vol. 31, pp. 1264–1277, 2009.

[12] K. Takahashi, S. Sexi, and R. Oka, “Spotting Recognition of Human Gestures From Motion Images,” *In Technical Report IE92-134*, pp. 9–16, 1992.

[13] S. S. Jambhale and A. Khaparde, “Gesture recognition using dtw & piecewise dtw,” in *ICECS*, 2014, pp. 1–5.

[14] K. Barczewska and A. Drozd, “Comparison of methods for hand gesture recognition based on Dynamic Time Warping algorithm,” *FedCSIS*, pp. 207–210, 2013.

[15] M. Shimano, T. Okabe, I. Sato, and Y. Sato, “video temporal super-resolution based on self-similarity,” *ACCV*, vol. 6492, pp. 93–106, 2010.

[16] E. Shechtman, Y. Caspi, and M. Irani, “Space-Time Super-Resolution,” vol. 27, no. 4, pp. 531–545, 2005.

[17] A. Gupta, P. Bhat, M. Dontcheva, O. Deussen, B. Curless, and M. Cohen, “Enhancing and experiencing spacetime resolution with videos and stills,” *ICCP*, 2009.

[18] K. Watanabe, Y. Iwai, H. Nagahara, M. Yachida, and T. Suzuki, “Video synthesis with high spatio-temporal resolution using motion compensation and spectral fusion,” *IEICE Transactions*, vol. 89-D, no. 7, pp. 2186–2196, 2006.

[19] K. Watanabe, Y. Iwai, T. Haga, and M. Yachida, “A fast algorithm of video super-resolution using dimensionality reduction by DCT and example selection,” in *ICPR, Florida, USA*, 2008, pp. 1–5.

[20] S. T. L. C. A. G. M. M, “View and Time Interpolation in Image Space,” *Computer Graphics Forum 2008*, vol. 27, no. 7, pp. 1781–1787, 2008.

[21] Y. Makihara, A. Mori, and Y. Yagi, “Periodic Temporal Super Resolution Based on Phase Registration and Manifold Reconstruction,” *Ipsj*, vol. 3, no. 1, pp. 134–147, 2011.

[22] N. Akae, Y. Makihara, and Y. Yagi, “Gait recognition using periodic temporal super resolution for low frame-rate videos,” *IJCB*, 2011.

[23] C. Stauffer and W. Grimson, “Adaptive background mixture models for real-time tracking,” in *Proceedings of CVPR*, 1999.

[24] H.-G. Doan, V.-T. Nguyen, H. Vu, and T.-H. Tran, “A combination of user-guide scheme and kernel descriptor on rgb-d data for robust and realtime hand posture recognition,” *Eng. Appl. Artif. Intell.*, vol. 49, no. C, pp. 103–113, Mar. 2016.

[25] B. D. Lucas and T. Kanade, “An iterative image registration technique with an application to stereo vision,” in *Proc. IJCAI*, 1981, pp. 674–679.

[26] J.Shi and C.Tomasi, “Good features to track,” in *Proc. IJCAI*, 1994, pp. 593–600.

[27] T. Lin and H. Zha, “Riemannian manifold learning,” *TPAMI*, vol. 30, no. 5, pp. 796–809, May 2008.

[28] S. T. Roweis and L. K. Saul, “Nonlinear Dimensionality Reduction by Locally Linear Embedding,” *Science*, vol. 290, pp. 2323 – 2326, 2000.

[29] M. Belkin and P. Niyogi, “Laplacian eigenmaps for dimensionality reduction and data representation,” *Neu. Comput.*, vol. 15, no. 6, pp. 1373–1396, Jun. 2003.

[30] A. Kurakin, Z. Zhang, and Z. Liu, “A real time system for dynamic hand gesture recognition with a depth,” in *EUSIPCO*, 2012, pp. 1975–1979.

[31] J. Wang, Z. Liu, J. Chorowski, Z. Chen, and Y. Wu, “Robust 3D Action Recognition with Random Occupancy Patterns,” *ECCV*, pp. 872–885, 2012.

[32] O. Oreifej and Z. Liu, “HON4D: Histogram of oriented 4D normals for activity recognition from depth sequences,” *CVPR*, pp. 716–723, 2013.

[33] X. Yang and Y. Tian, “Super Normal Vector for Action Recognition Using Depth Sequences,” *CVPR*, pp. 804–811, 2014.