Fusion Machine Learning Strategies for Multi-modal Sensor-based Hand Gesture Recognition

Huong-Giang Doan
Control and Automation Faculty
Electric Power University
Hanoi, Vietnam
giangdt@epu.edu.vn

Ngoc-Trung Nguyen
Control and Automation Faculty
Electric Power University
Hanoi, Vietnam
trungnn@epu.edu.vn

Abstract-Hand gesture recognition has attracted the attention of many scientists, because of its high applicability in fields such as sign language expression and human machine interaction. Many approaches have been deployed to detect and recognize hand gestures, like wearable devices, image information, and/or a combination of sensors and computer vision. However, the method of using wearable sensors brings much higher accuracy and is less affected by occlusion, lighting conditions, and complex background. Existing solutions separately utilize sensor information and/or only use sensor information processing and decision-making algorithms over conventional threshold comparison algorithms and do not analyze data or utilize machine learning algorithms. In this paper, a multi-modal solution is proposed that combines information for measuring the curvature of the fingers and sensors for measuring angular velocity and acceleration. The provided information from the sensors is normalized and analyzed and various fusion strategies are used. Then, the most suitable algorithm for these sensor-based multiple modalities is proposed. The proposed system also analyzes the differences between gestures and actions that are almost similar but in fact, they are just normal moving gestures.

Keywords-hand glove; acceleration sensor; hand gesture recognition; flex sensor; human-machine interaction

I. INTRODUCTION

Hand gestures are one of the most natural ways of interaction between people (e.g. sign language [1-6, 24]) or Human Computer Interaction–HCI [7-9]. Conveying messages in sign language through hand gestures has attracted the attention of many researchers and technology developers. As a consequence, many hand gesture recognition approaches have been proposed, such as using the change of radar waves when the user changes the state of the hand [10-11, 25], utilizing image information [12-14], or sensors physically attached to the hand [15-17]. The radar-based method [10] is easy and convenient for end-users because of independent equipment but it is dramatically affected by environmental noises and the number of antennas [27]. There are many moving objects in the normal environment that can also cause changes in feedback waves such as movement of other people, or other body parts. For the image-based approach [12], hand gesture recognition systems often face many challenges such as high time cost for hand detection and hand recognition, effects of illumination, occlusion, complex background conditions, etc. The hand mounted sensor-based method [15], although limited due to device dependence that has much higher accuracy compared to other methods and especially it face to many criteria when the actual control system requires high accuracy. Thus, this solution is preferred in many special cases.

Hand-mounted sensors were applied to measure the change in the hand shapes and hand movements that were quite efficient because of their accuracy and real-time response. Many solutions have been proposed, such as the electronic gloves [17-18]. These gloves have integrated flex sensors, mounted on the fingers, to collect the hand's curvature changing. However, this solution only measures the change in the flexure of the fingers, but it cannot obtain the movement of the hand. In [19-20], the authors proposed a method to attach the sensors on the hand for receiving velocity and angular acceleration. This method achieves the alternation of hand movement, but it is not possible to obtain the changes of hand shapes. In [21-23], the authors combined both curvature and velocity sensors, but the classification of hand gesture categories was performed by the conventional comparison structures with simple instruction sets of microcontrollers (e.g. ARV, 8051 with compare and check condition instructions), without evaluation or survey changing of parameters of sensors or machine learning algorithms. Quantitative evaluations were not implemented to compare the defined activities with normal human activities.

In this paper, a new solution is proposed, one that combines both hand shape and hand movement sensors with various fusion strategies of features. The obtained results have success rates of up to 99.87% for the stable hand gesture dataset and 97.59% when both mobile and immobile hand gesture datasets are considered. These results are higher than the 97.4% and 86.3% for seen and unseen users of 12 gestures in [28] that fed the data of finger's curvatures to a convolution neural network. There are no published hand gesture datasets based on mounted sensors and the published ones do not provide both hand's curvatures and position at the same time. So, in this research, we collected two new hand gesture databases, one with stationary hand and one with moving hand.
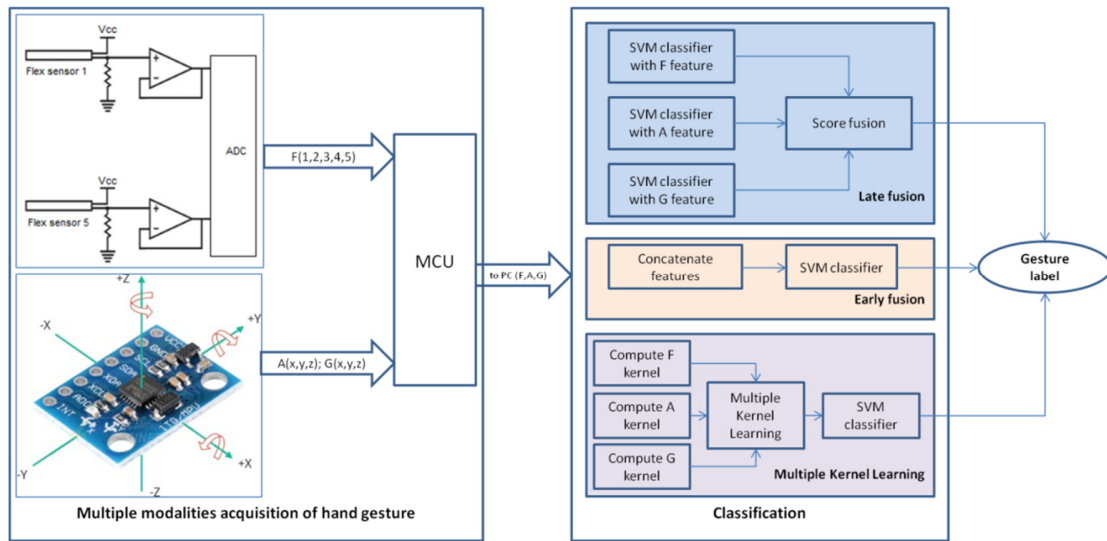
Fig. 1.   Proposed framework for hand gesture recognition.

The data flows will be fed into classifiers with various fusion strategies of machine learning techniques to find the most optimal and suitable classification solution.

II. PROPOSED METHOD

Our proposed framework for hand gesture recognition is illustrated in Figure 1. In this research, multiple modalities of hand $(l \in [1, L], L = 3$ corresponding to $l \in [A, G, F])$ are captured by two streams:

- Finger's curvature: five flex sensors are used to present the curvature variation of fingers.

- Hand's motion tracking and angle: sensor MPU6050 is integrated by 3-axis MEMS of gyroscope and 3-axis MEMS of accelerometer.

All features are processed and synchronized by an MCU (Microcontroller Unit) and are transferred to a PC. Finally, various classification strategies were used, either single or multi-modalities. The cascade steps in our proposed framework will be presented in detail at the following Sections.

A. Hardware Design for Multiple features of the Hand Glove

The detailed hardware design of the electronic glove is illustrated in Figure 2. Five flex sensors are permanently mounted along the fingers of the glove to collect the curvature based on the change of the corresponding resistance value. The resistors' data are preprocessed to convert into voltage value before conversion from analog to digital with a 10-bit ADC (Analog to Digital Converter) resolution, corresponding to a range of values from 0 to 1023. The values of flex sensors from the thumb to the little finger are denoted by F_1 to F_5 $((F_1, F_2, F_3, F_4, F_5)/F(1,2,3,4,5)$ corresponding to the curvature data of the five fingers. At the same time, the values of the 3-axis MEMS gyroscope and 3-axis MEMS accelerometer of the MPU6050 sensor were collected. The resolution of ADCs is 16 bits with a range from 0 to 65535, denoted by $(A_x, A_y, A_z)/A(x, y, z)$ and $(G_x, G_y, G_z)/G(x, y, z)$. After the MCU collects the two data streams (flex sensors and

MPU sensor) at the same time, all data are packaged and sent to the PC via the USB port according to the UART standard.
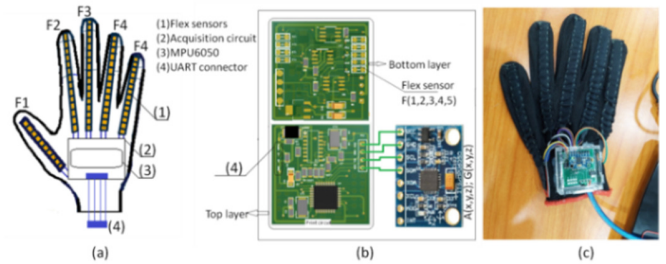

Fig. 2.   Harware design of the hand glove.

In addition, each time, the information of the hand gesture is transferred by a feature vector which is composed by eleven elements in total: $A(x, y, z)$, $G(x, y, z)$, and $F(1,2,3,4,5)$, respectively. This feature is presented in detail in (1) and will be utilized in various strategies in detail in the next section:

$$F_{Ges} = [F_G^1, F_G^2, \ldots, F_G^M], (M = 11)$$
$$= [A(x, y, z); G(x, y, z); F(1,2,3,4,5)] \quad (1)$$
$$= [A_x, A_y, A_z, G_x, G_y, G_z, F_1, F_2, F_3, F_4, F_5]$$

B. Hand Gesture Dataset

In this paper, twelve hand postures are used, as illustrated in Figure 3. Each hand gesture is captured when the end-user is immobile or in movement. In practice, the system will recognize the hand gestures at any time and any place, i.e. hand morphologies during normal human activities that could have similar characteristics to the previously defined dataset. In addition, the feature of hand shape is only based on the curvature of the fingers that are collected by the five flex sensors. It is apparent that the accelerometer sensors and velocity sensors provide useful information about the movement and the direction of the hand's movement.
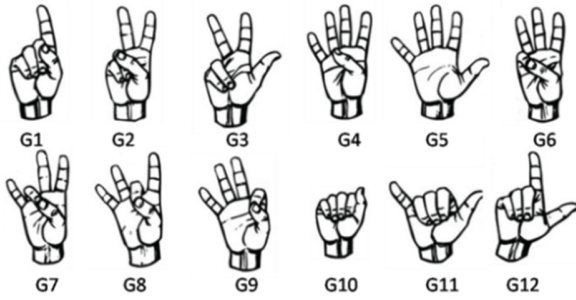
Fig. 3.   The twelve hand postures of our dataset.

Figure 4 illustrates hand tracking with three accelerator parameters $A(x, y, z)$ and three velocity features $G(x, y, z)$ of four hand gestures in two cases: (1) the end-user raises the hand in front of his/her body and performs a gesture (blue, white, violet and grey colors in Figure 4) and (2) the end-user implements the same hand shape but hand and body are moving (green, yellow, red and orange colors in Figure 4). It is clear that feature A(x,y,z) or feature G(x,y,z) of the same hand gesture represent the hand's movement. It could effectively separate between immobile hand mobility hands while they do not have meaning for stable gestures.
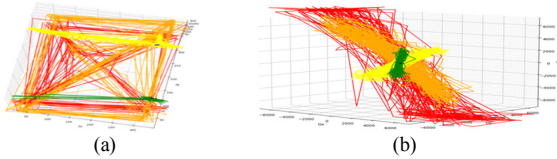

Fig. 4.   (a) A(x,y,z) and G(x,y,z) elements of the same hand gesture.

In practice, the users are usually immobile when they want to control the device and often put their hands in front of their body. Meanwhile, the end-user's hand shapes exist at both stable and moving situations. Therefore, we collected 12 gestures as shown in Figure 3 when the hand is in front of the user's face, when the hand is immobile and during normal movement. Each category will be separately labeled by Ges_k^1 (gesture of class k, hand is frontal of body and stable) and Ges_k^2 (gesture of class k, hand is at any position and movement). We divided the data into two datasets named HandGlove1 and HandGlove2. HandGlove1 is composed by $[Ges_k^1](k \in [1, K], K = 12)$ and HandGlove2 by $[Ges_k^1; Ges_k^2](k \in [1, K], K = 12)$. This means that HandGlove2 dataset has a duplicated number of categories up to $K = 24$ classes. Each gesture was collected three times and each time consists of about 200 samples. A total of 15 adults $(User_j (j \in [1, E], E = 15))$, including 7 females and 8 males were invited to collect our dataset at different days and at various times of the day. These databases were normalized and classifiers were used.

C. Data Processing

As described above, the data collected from the flex sensors are denoted by $F(1,2,3,4,5)$ and have a value range of 0 to 1023. The data from the MPU6050 sensors include $A(x, y, z)$ and $G(x, y, z)$ have a range of values from 0 to 65535. A sensor component has a different value range and 11

components must be processed to obtain the same range. Each characteristic component F_{Ges}^i of the 12 gestures $[Ges_k^1]$ and/or $[Ges_k^1; Ges_k^2](k \in [1, K], K = 12)$ of 15 collectors $(User_j (j \in [1, E], E = 15))$ is normalized. Firstly, the maximum value $F_{Ges\_max}^i$ and the smallest value $F_{Ges\_min}^i$ can be determined:

$$F_{Ges\_max}^i = Max \left(F_{Ges}^i \text{ of all } (Ges_k; User_j) \right) (k \in [1, K]); (j \in [1, E]) \quad (2)$$

$$F_{Ges\_min}^i = Min \left(F_{Ges}^i \text{ of all } (Ges_k; User_j) \right) (k \in [1, K]); (j \in [1, E]) \quad (3)$$

Each component $F_{G\_nor}^i$ will be recalculated as (4):

$$F_{Ges\_nor}^i = \frac{F_{Ges}^i - F_{Ges\_min}^i}{F_{Ges\_max}^i - F_{Ges\_min}^i} \quad (4)$$

As a result, the feature vector could be presented by: $F_{Ges} = [F_{Ges\_nor}^1, F_{Ges\_nor}^2, ..., F_{Ges\_nor}^{11}]$. Figure 6 illustrates the t-SNE (Distributed Stochastic Neighbor Embedding) for all the 12 gestures. Figure 6(a) uses $[F_{Ges\_nor}^7; ...; F_{Ges\_nor}^{11}]$ elements, Figure 6(b) utilizes $[F_{Ges\_nor}^4; ...; F_{Ges\_nor}^{11}]$, and Figure 6(c) utilizes $[F_{Ges\_nor}^1; ...; F_{Ges\_nor}^{11}]$. This Figure shows that, if only using the curvature feature from the flex sensors, the data distribution of the gestures is not segregated than using combinations of the motion features. The composition of 6 motion features enables gesture types to be separated into more distinct spatial domains. This result is only qualitative but not quantitative. These normalized features will be put into classifiers with single patterns and combined features as described in Section III.D.

D. Hand Gesture Classification

Multi-modal data stream is presented above with F, A, and G which are combined using different fusion techniques. In this research, three fusion techniques are investigated: late fusion, early fusion, and Multiple Kernel Learning (MKL) [2]. In the following, we will briefly survey these methods that are utilized in the classification block of Figure 1.

1) Early Fusion

For early fusion, all normalized feature vectors of the 3 modalities are concatenated into a final feature vector as presented in (5):

$$F_f = [F_{G\_nor}^1, F_{G\_nor}^2, ..., F_{G\_nor}^M] \quad (5)$$

Then, the feature vector F_f is used as an input of a final multiple SVM classifier to predict the hand gesture label.

2) Late Fusion

For late fusion, each modality is used with multiple SVM classifiers C_{SVM}^l with $l \in [1, L](L = 3)$. The outputs of the classification step are the score vectors $S^l = [s_1^l, s_2^l, ..., s_K^l]$. Next, the maximum operator for these score vectors S^l is applied to obtain the final score vector as presented in (6):

$$S = [s_1, s_2, ..., s_K] = \frac{S^1 + S^2 + S^3}{3} = [\frac{s_1^1 + s_1^2 + s_1^3}{2}, \frac{s_2^1 + s_2^2 + s_2^3}{2}, ..., \frac{s_K^1 + s_K^2 + s_K^3}{2}] \quad (6)$$

Then, the final decision is obtained as illustrated in (7):

$$gesture\ label = argmax_{k \in [1, K]}(s_k) \quad (7)$$

3) Multiple Kernel Learning

MKL is an algorithm that combines a set of base kernels that could represent different similarity measures of different sources of data. As a result, each feature vector from the lth ($l \in [1, L]$) modality utilizes a kernel function to compute the corresponding kernel matrix as shown in (8):

$$Kernel_K^v = similarity[(F_{Ges}^A, F_{Ges}^G), (F_{Ges}^A, F_{Ges}^F), (F_{Ges}^F, F_{Ges}^G)] \quad (8)$$

where $v \in [1, V], (V = 3)$. Then, kernel matrices $Kernel_K^v$ are used to combine for $Kernel$ factor as presented in (9):

$$Kernel = func_\mu(\{Kernel_K^v\}_{v=1}^V|_\mu) \quad (9)$$

where $func$ denotes the function form. The combination coefficient μ sees that the values are bound to the predefined rules or are optimized by the learning process of MKL [1, 2]. In this research, EasyMKL [1] with a binary margin maximization MKL algorithm is chosen which uses the convex summation function. The coefficients μ are restricted to be non-negative and sum to 1. A final kernel machine classifier will decide the label based on the combined kernel matrix $Kernel$ by a SVM classifier.

III. EXPRIMENTIAL RESULTS

In this paper, two datasets (HandGlove1 and HandGlove2) were utilized. The "leave-one-subject-out cross-validation" protocol is followed [3]. This means that, we took samples from one subject for testing and samples from the remaining subjects for training. Then we compute the average accuracy of every experiment.

A. Single Modality for Hand Gesture Recognition

In this section, we used HandGlove1 dataset in separating the following data types: Five components are the finger curvature sensors $F(1,2,3,4,5)$, 3 components measure the angle of the hand $A(x,y,z)$, and 3 components show the angular velocity $G(x,y,z)$. Each type is firstly normalized. It is then put into the SVM classifier. The classification results on each modality are presented in Table I. Table I shows that the finger's curvatures of the hand gesture dataset (Figure 3) obtained the highest recognition result (92.15%) while $A(x,y,z)$ cue has the lowest at only 36.48%. Although the result of the $G(x,y,z)$ modality is higher than the cue modality, but it still is ineffective for the defined dataset with 42.76%. This result shows that each data type gives low results, not suitable to control in practice.

TABLE I.   SINGLE MODALITY FOR HAND GESTURE RECOGNITION ACCURACY (%) WITH SVM CLASSIFIER

Modality	Accuracy (%)
$A(x,y,z)$	36.48
$G(x,y,z)$	42.76
$F(1,2,3,4,5)$	92.15

B. Hand Gesture Recognition on Various Fusion Strategies

In this part, we will investigate different strategies (early fusion, late fusion, and MKL) on the three streams. Evaluations are conducted on HandGlove1 dataset. The results are shown in Figure 5.
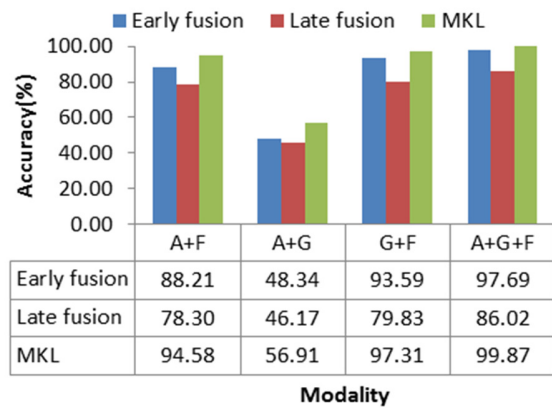

Modality	A+F	A+G	G+F	A+G+F
Early fusion	88.21	48.34	93.59	97.69
Late fusion	78.30	46.17	79.83	86.02
MKL	94.58	56.91	97.31	99.87

Fig. 5.   Hand gesture recognition on various fusion strategies.

a) F(1,2,3,4,5)   b) F(1,2,3,4,5) and A(x,y,z)

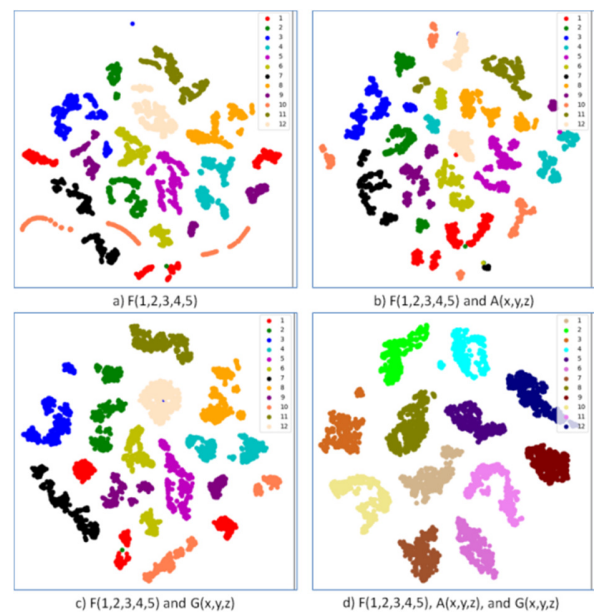c) F(1,2,3,4,5) and G(x,y,z)   d) F(1,2,3,4,5), A(x,y,z), and G(x,y,z)

Fig. 6.   T-SNE of 12 gestrures with various modalities.

Figure 5 shows that:

- KML obtains the highest accuracy at 94.58%, 90.18%, 97.31%, and 99.87% with combination of *A* and *F*, *A* and *G*, *G* and *G*, and *A+G+F* respectively. The late fusion method reached the lowest accuracies on all data combinations (around 75% and the smallest at 46.17% for *A+G* modality). We can see that early fusion is quite simple while its results are lower than KML's. The late fusion method requires many classifiers but its accuracy is far smaller than KML's. Therefore, the KML solution will be utilized in the rest of the testing.

- Combinations of the 3 modalities are the highest on all classifiers and account for 97.69%, 86.02% and 99.87% for early fusion, late fusion, and KML respectively. These results can be explained through Figure 6 that shows the distribution of data using the t-SNE method [26]. When *G* is used, the hand gesture classes distribute in the overlap domains. When *F* and *A* are combined as shown in Figure

6(b) or *F* and *G* as illustrated in Figure 6(c) the distribution of the categories has improved but the separation is not clear enough. In Figure 6(d), *A*, *G*, and *F* are utilized and the data distribution domains of the 12 hand gestures are completely separated.

- Association of the three modalities accounts for slightly higher results in HandGlove1 dataset because the gestures in this dataset have negligible displacement. The hand rotations are almost the same and the hand is raised in front of the body. However, the system is trained by HandGlove1 dataset, and then this trained model is deployed in real application. As a result, the online system has many mistakes. To see the effectiveness of *A* and *G* modalities, we will be evaluated in the below Section IV.C.

C. Distinction between Command Action and Normal Gestures

In this section, the KML fusion strategy will be used for testing on the HandGlove2 dataset. All gestures belonging to the group Ges_k^2 are labeled with labels from 13 to 24. The recognition accuracy is considered only with gestures with labels from 1 to 12. These gestures are defined and expected to be correctly recognized by the system. The user raises his/her hand in front of the face and the state of the hand is stationary. Only the hand shape changes according to the specified gesture means. In this test, a combination of different modalities will be used, such as: *A* and *F*, *G* and *F*, and *A*, *G*, and *F*. The evaluation results are shown in Figure 7.

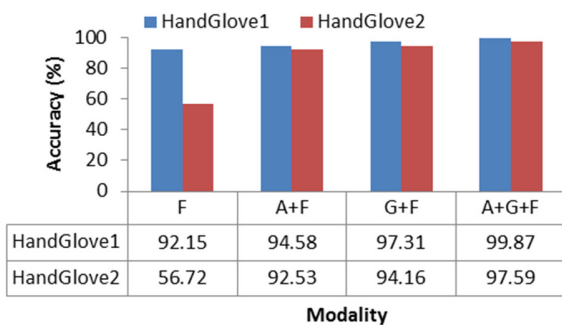| Modality	F	A+F	G+F	A+G+F
HandGlove1	92.15	94.58	97.31	99.87
HandGlove2	56.72	92.53	94.16	97.59

Fig. 7.   Efficient of modalities combination for hand gesture recognition accuracy (%) in a real environment.

In HandGlove1 dataset, the recognition results of the ways of combining data have not a big gap from 92.15% to 99.87%. Hand gestures are distinguished with other hand postures. In addition, end-user's hand and body are immobile so that the curvature of the fingers could provide enough cues of hand shapes in this dataset, so, the recognition results are quite good. However, in HandGlove2 database, the recognition accuracy is low when the 5 flex sensors are utilized. This is because of for the same changing shapes, i.e. the same five *F* elements, one of them moves like normal hand movement (different *A* and *G* elements). Gestures 1 to 12 have the same hand shape with corresponding gestures 13 to 24 (Figure 8). Therefore, using only the curvatures of the fingers does not give enough information to distinguish between a control operation and a normal operation as shown in Figure 8. For example, gesture 1

(red color) and gesture 13 (slow brown color) in Figure 8 have the same hand shape but hand gesture 1 is immobile while gesture 13 is in movement. Distributions of these gestures overlap in Figure 8(a), but are reparative in Figure 8(b). Thus, there is a big gap between single modal (56.72% with *F*) and multi-modal (97.59% with *A*, *G*, and *F*) in the second row of Figure 7.

In HandGlove2 dataset, combining *F* and *A* or *F* and *G* showed a significant improvement in accuracy, increasing up to 33% and 36% respectively (Figure 7). Especially, the coherency of the 3 components (*A*, *G*, and *F*) obtained the highest recognition accuracy, which was increased up to 97.59% in this dataset. The proposed framework could apparently distinguish between the 12 hand gestures in both immobile and mobile states. The efficiency of the system dramatically increased and practical applications can now be deployed.
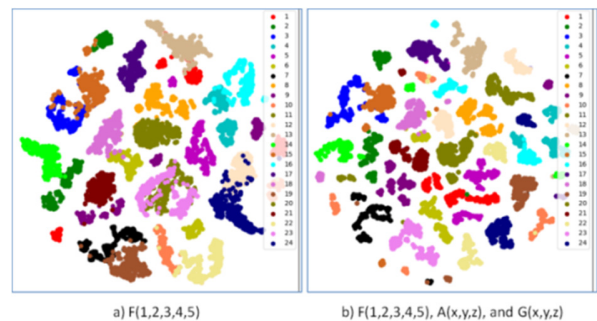

a) F(1,2,3,4,5)   b) F(1,2,3,4,5), A(x,y,z), and G(x,y,z)

Fig. 8.   T-sne of 24 gestrures with various modalities.

IV.   CONCLUSION AND DISCUSSION

This paper presents a comparative analysis of recent fusion strategies for static hand gesture recognition on three modalities of an electronics hand glove. Among the evaluated fusion strategies (late fusion, early fusion, and KML), KML performed better, achieving the highest recognition accuracy. The evaluation results on two datasets show great interest in combining multi-modalities (5 curvature fingers from flex sensors, 3-axis gyroscope and 3-axis accelerometer from MPU6050 sensor) to increase accuracy. In most cases, multimodal KML models can achieve an accuracy rate above 97%. This performance is remarkable and promises a feasible solution for deploying gesture-based applications in practice. Finally, it was found that there is a notable gap in recognition accuracy between the same hand shapes but different normal hand movements. The last remark opens up new research directions that require further investigation on data combination with image information, using deep convolutional neural networks, and online learning. Once these bottlenecks are resolved, the development of a gesture-based interface in practical applications is straightforward.

REFERENCES

[1] F. Aiolli, G. Da San Martino, and A. Sperduti, "A Kernel Method for the Optimization of the Margin Distribution," in *Artificial Neural Networks - ICANN 2008*, Berlin, Heidelberg, Germany, 2008, pp. 305–314, https://doi.org/10.1007/978-3-540-87536-9_32.

[2] T.-H. Tran, H.-N. Tran, and H.-G. Doan, "Dynamic Hand Gesture Recognition from Multi-modal Streams Using Deep Neural Network," in *Multi-disciplinary Trends in Artificial Intelligence*, 2019, pp. 156–167, https://doi.org/10.1007/978-3-030-33709-4_14.

[3] Dang-Manh Truong was with Hanoi University of Science Technology, Vietnam, D.-M. Truong, H.-G. Doan, T.-H. Tran, H. Vu, and T.-L. Le, "Robustness Analysis of 3D Convolutional Neural Network for Human Hand Gesture Recognition," *International Journal of Machine Learning and Computing*, vol. 9, no. 2, pp. 135–142, Apr. 2019, https://doi.org/10.18178/ijmlc.2019.9.2.777.

[4] P. Das, T. Ahmed, and Md. F. Ali, "Static Hand Gesture Recognition for American Sign Language using Deep Convolutional Neural Network," in *2020 IEEE Region 10 Symposium (TENSYMP)*, Dhaka, Bangladesh, Jun. 2020, pp. 1762–1765, https://doi.org/10.1109/TENSYMP50017.2020.9230772.

[5] J. Zamora-Mora and M. Chacón-Rivas, "Real-Time Hand Detection using Convolutional Neural Networks for Costa Rican Sign Language Recognition," in *2019 International Conference on Inclusive Technologies and Education (CONTIE)*, San Jose del Cabo, Mexico, Jul. 2019, pp. 180–1806, https://doi.org/10.1109/CONTIE49246.2019.00042.

[6] N. Sarhan and S. Frintrop, "Transfer Learning For Videos: From Action Recognition To Sign Language Recognition," in *2020 IEEE International Conference on Image Processing (ICIP)*, Jul. 2020, pp. 1811–1815, https://doi.org/10.1109/ICIP40778.2020.9191289.

[7] S. Veluchamy, L. R. Karlmarx, and J. J. Sudha, "Vision based gesturally controllable human computer interaction system," in *2015 International Conference on Smart Technologies and Management for Computing, Communication, Controls, Energy and Materials (ICSTM)*, Avadi, India, Feb. 2015, pp. 8–15, https://doi.org/10.1109/ICSTM.2015.7225383.

[8] H.-G. Doan, H. Vu, T.-H. Tran, and E. Castelli, "Improvements of RGB-D hand posture recognition using an user-guide scheme," in *2015 IEEE 7th International Conference on Cybernetics and Intelligent Systems (CIS) and IEEE Conference on Robotics, Automation and Mechatronics (RAM)*, Siem Reap, Cambodia, Jul. 2015, pp. 24–29, https://doi.org/10.1109/ICCIS.2015.7274542.

[9] S. Lee, K. Park, J. Lee, and K. Kim, "User Study of VR Basic Controller and Data Glove as Hand Gesture Inputs in VR Games," in *2017 International Symposium on Ubiquitous Virtual Reality (ISUVR)*, Nara, Japan, Jun. 2017, https://doi.org/10.1109/ISUVR.2017.16.

[10] N. E. A. Rashid, Y. A. I. M. Nor, K. K. M. Sharif, Z. I. Khan, and N. A. Zakaria, "Hand Gesture Recognition using Continuous Wave (CW) Radar based on Hybrid PCA-KNN," in *2021 IEEE Symposium on Wireless Technology Applications (ISWTA)*, Shah Alam, Malaysia, Dec. 2021, pp. 88–92, https://doi.org/10.1109/ISWTA52208.2021.9587404.

[11] G. Zhang, S. Lan, K. Zhang, and L. Ye, "Temporal-Range-Doppler Features Interpretation and Recognition of Hand Gestures Using mmW FMCW Radar Sensors," in *2020 14th European Conference on Antennas and Propagation (EuCAP)*, Copenhagen, Denmark, Mar. 2020, https://doi.org/10.23919/EuCAP48036.2020.9135694.

[12] P. Molchanov, X. Yang, S. Gupta, K. Kim, S. Tyree, and J. Kautz, "Online Detection and Classification of Dynamic Hand Gestures with Recurrent 3D Convolutional Neural Networks," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, Jun. 2016, pp. 4207–4215, https://doi.org/10.1109/CVPR.2016.456.

[13] J. J. Raval and R. Gajjar, "Real-time Sign Language Recognition using Computer Vision," in *2021 3rd International Conference on Signal Processing and Communication (ICPSC)*, Coimbatore, India, Feb. 2021, pp. 542–546, https://doi.org/10.1109/ICSPC51351.2021.9451709.

[14] J. Fink, B. Frénay, L. Meurant, and A. Cleve, "LSFB-CONT and LSFB-ISOL: Two New Datasets for Vision-Based Sign Language Recognition," in *2021 International Joint Conference on Neural Networks (IJCNN)*, Shenzhen, China, Jul. 2021, https://doi.org/10.1109/IJCNN52387.2021.9534336.

[15] X. Chu, J. Liu, and S. Shimamoto, "A Sensor-Based Hand Gesture Recognition System for Japanese Sign Language," in *2021 IEEE 3rd Global Conference on Life Sciences and Technologies (LifeTech)*, Nara, Japan, Mar. 2021, pp. 311–312, https://doi.org/10.1109/LifeTech52111.2021.9391981.

[16] K. Liu, C. Chen, R. Jafari, and N. Kehtarnavaz, "Multi-HMM classification for hand gesture recognition using two differing modality sensors," in *2014 IEEE Dallas Circuits and Systems Conference (DCAS)*, Richardson, TX, USA, Jul. 2014, https://doi.org/10.1109/DCAS.2014.6965338.

[17] S. Zhu, A. Stuttaford-Fowler, A. Fahmy, C. Li, and J. Sienz, "Development of a Low-cost Data Glove using Flex Sensors for the Robot Hand Teleoperation," in *2021 3rd International Symposium on Robotics Intelligent Manufacturing Technology (ISRIMT)*, Changzhou, China, Sep. 2021, pp. 47–51, https://doi.org/10.1109/ISRIMT53730.2021.9596972.

[18] A. K. Panda, R. Chakravarty, and S. Moulik, "Hand Gesture Recognition using Flex Sensor and Machine Learning Algorithms," in *2020 IEEE-EMBS Conference on Biomedical Engineering and Sciences (IECBES)*, Langkawi Island, Malaysia, Mar. 2021, pp. 449–453, https://doi.org/10.1109/IECBES48179.2021.9398789.

[19] S. S. Kumar, R. Gatti, S. K. N. Kumar, N. Nataraja, R. P. Prasad, and T. Sarala, "Glove Based Deaf-Dumb Sign Language Interpreter," in *2021 International Conference on Recent Trends on Electronics, Information, Communication Technology (RTEICT)*, Bangalore, India, Dec. 2021, pp. 947–950, https://doi.org/10.1109/RTEICT52294.2021.9573990.

[20] Z. Zou, Q. Wu, Y. Zhang, and K. Wen, "Design of Smart Car Control System for Gesture Recognition Based on Arduino," in *2021 IEEE International Conference on Consumer Electronics and Computer Engineering (ICCECE)*, Guangzhou, China, Jan. 2021, pp. 695–699, https://doi.org/10.1109/ICCECE51280.2021.9342137.

[21] X. Chen *et al.*, "A Wearable Hand Rehabilitation System With Soft Gloves," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 2, pp. 943–952, Oct. 2021, https://doi.org/10.1109/TII.2020.3010369.

[22] S. Mugala, P. Jjagwe, and J. Asiimwe, "Glove Based Sign Interpreter for Medical Emergencies," in *2019 IST-Africa Week Conference (IST-Africa)*, Nairobi, Kenya, Feb. 2019, https://doi.org/10.23919/ISTAFRICA.2019.8764816.

[23] P. Telluri, S. Manam, S. Somarouthu, J. M. Oli, and C. Ramesh, "Low cost flex powered gesture detection system and its applications," in *2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA)*, Coimbatore, India, Jul. 2020, pp. 1128–1131, https://doi.org/10.1109/ICIRCA48905.2020.9182833.

[24] H. V. T. Chi, D. L. Anh, N. L. Thanh, and D. Dinh, "English-Vietnamese Cross-Lingual Paraphrase Identification Using MT-DNN," *Engineering, Technology & Applied Science Research*, vol. 11, no. 5, pp. 7598–7604, Oct. 2021, https://doi.org/10.48084/etasr.4300.

[25] N. T. T. Vu, N. P. Tran, and N. H. Nguyen, "Recurrent Neural Network-based Path Planning for an Excavator Arm under Varying Environment," *Engineering, Technology & Applied Science Research*, vol. 11, no. 3, pp. 7088–7093, Jun. 2021, https://doi.org/10.48084/etasr.4125.

[26] M. Wattenberg, F. Viégas, and I. Johnson, "How to Use t-SNE Effectively," *Distill*, vol. 1, no. 10, Oct. 2016, Art. no. e2, https://doi.org/10.23915/distill.00002.

[27] A. A. Alzamil, "Assessment of Uplink Massive MIMO in Scattering Environment," *Engineering, Technology & Applied Science Research*, vol. 10, no. 5, pp. 6290–6293, Oct. 2020, https://doi.org/10.48084/etasr.3743.

[28] O. Luzanin and M. Plancak, "Hand gesture recognition using low-budget data glove and cluster-trained probabilistic neural network," *Assembly Automation*, vol. 34, no. 1, pp. 94–105, Jan. 2014, https://doi.org/10.1108/AA-03-2013-020.