Dynamic hand gesture recognition based on 3D pattern assembled trajectories

Said Yacine Boulahia, Eric Anquetil, Franck Multon, Richard Kulpa

Dynamic hand gesture recognition based on 3D pattern assembled trajectories

Said Yacine Boulahia[1], Eric Anquetil[1], Franck Multon[1] and Richard Kulpa[1]
[1] IRISA/INRIA, Rennes. France
e-mail: said-yacine.boulahia@irisa.fr, eric.anquetil@irisa.fr,
franck.multon@irisa.fr, richard.kulpa@irisa.fr

***Abstract*— Over the past few years, advances in commercial 3D sensors have substantially promoted the research of dynamic hand gesture recognition. On a other side, whole body gestures recognition has also attracted increasing attention since the emergence of Kinect like sensors. One may notice that both research topics deal with human-made motions and are likely to face similar challenges. In this paper, our aim is thus to evaluate the applicability of an action recognition feature-set to model dynamic hand gestures using skeleton data. Furthermore, existing datasets are often composed of pre-segmented gestures that are performed with a single hand only. We collected therefore a more challenging dataset, which contains unsegmented streams of 13 hand gesture classes, performed with either a single hand or two hands. Our approach is first evaluated on an existing dataset, namely DHG dataset, and then using our collected dataset. Better results compared to previous approaches are reported.**

***Keywords*— 3D dynamic hand gesture recognition, Leap Motion dataset, Skeleton joint data, Real-time recognition.**

I. Introduction

The increasing demand of touch-less interfaces encourages the research on hand gesture recognition. In particular, this is crucial for applications such that virtual and augmented reality, sign language recognition, gaming, robotics, etc.

Previous approaches dealing with 3D hand gesture recognition can be gathered into two main categories: static and dynamic hand gesture recognition. The former focus only on the posture by extracting hand silhouettes or the hand regions of interest. The latter consider instead the progression of hand joint positions across time. As it brings a greater amount of information and is more adapted for interaction interfaces, we consider in this paper the dynamic hand gesture recognition using skeleton data.

Encouraged by recent advances in terms of sensor technologies, such as the Leap Motion device, various approaches are being proposed. Compared to 2D cameras, these sensors are more robust to common low-level issues in RGB imagery like background subtraction and light variation. However, hand gesture recognition is still a very difficult computer vision topic. This is mainly due to the fact that the hand is an object with a complex topology and has many possibilities to perform the same gesture. Furthermore, as with any human-made trajectory, hand joint trajectories are also subject to inter-class variabilities and intra-class confusions which should be addressed when considering dynamic hand gesture recognition.

978-1-5386-1842-4/17/$31.00

We therefore propose in this paper to consider dynamic hand gesture recognition in the light of previous advances of whole body action recognition. In particular, we propose to evaluate the suitability of a skeleton-based action recognition feature-set, namely HIF3D [1], to model hand gestures by associating each finger to a body arm or foot. The considered feature-set is particularly interesting as it was conceived based on a handwriting features-set. By doing so, we want to go a step further toward the fusion of ways the human-made trajectories issued either from whole body actions or hand gestures, are modelled.

Our second contribution in this paper is the collection of a new dataset to allow deeper evaluations of dynamic hand gesture recognition systems. In fact, we noticed that existing datasets are only composed of very short clips (around 30 frames) during which the recorded gestures are performed with only a single hand. We also found that those recorded gestures are perfectly performed without missing parts as it is the case in real-life use. Last, previous datasets contain only pre-segmented gestures which prevents from evaluating segmentation methods. Our recorded data aim to address these gaps.

The remainder of this paper is organized as follows: Section II reviews recent related work in the literature. In Section III, we briefly introduce the HIF3D feature-set that was initially designed for modelling skeleton-based actions. We then present in Section IV our dynamic hand gesture recognition approach for pre-segmented gestures. The new collected dataset is described in Section V. Results obtained over an existing dataset, namely DHG, and our new dataset are presented and discussed in Section VI. Finally, in Section VII conclusions and future research directions are drawn.

II. Related work

Several approaches dealing with dynamic hand gesture recognition have been developed so far. Unlike the static approaches based on hand description on a single image, dynamic methods exploit the temporal character of hand motion, by considering the gesture as a time series of hand position or shape. Early methods relied mostly on RGB data. For instance, Monnier et al. [10] proposed to fuse body skeleton-

based and Histogram of Oriented Gradients (HOG) features and then performed gestures classification using a boosted cascade classifier. Escobedo-Cardenas and Camara-Chavez [7] also suggested to combine intensity, depth and skeleton joint data captured by Kinect sensor for enhancing dynamic hand gesture recognition. In [8], authors presented the MSR-3D hand gesture database containing 12 dynamic American Sign Language. They also proposed a recognition algorithm based on an action graph to represent the dynamic part of a gesture.

However, most of those methods relied on Kinect data which only contain poor information of the hands position. This is particularly insufficient as each finger could be a valuable source of information when recognizing hand gestures. The emergence of Leap Motion like devices has changed the paradigm regarding hand gestures recognition. In fact, with such devices, skeleton data became much more attractive as they consist in detailed information of finger positions.

The approach proposed by [9] is among the first methods that considered hand gesture recognition using Leap Motion data. They proposed to represent a hand as a combination of fingertips angles, elevation and distance. Leap Motion is also considered for sign language recognition as proposed by Elons et al. [6] to recognize Arabic signs. More recently, De Smedt et al. [3] developed a new skeleton-based approach for 3D hand gesture recognition. Authors built their descriptor by exploiting the geometric shape of the hand, which was encoded by a Fisher vector. Noticing a lack of datasets, authors presented also a new Dynamic Hand Gesture (DHG) dataset, which contains sequences of hand skeleton in addition to the depth image.

As revealed in the previous review, the dynamic hand gesture recognition using skeleton data is an insufficiently explored area despite the promising applications that it can allow. In contrast, whole body action recognition using skeleton data has become popular as Shotton et al. [13] proposed a real-time method to accurately predict the 3D positions of body joints from depth images. Since then, several skeleton-based approaches have been proposed to recognize whole body actions. Therefore, instead of addressing from scratch the problem of dynamic hand gesture recognition, our first contribution in this paper is to extend an existing feature-set initially conceived for skeleton-based action recognition, namely HIF3D [1]. In fact, extending HIF3D to model dynamic hand gestures is particularly interesting as this feature-set was inspired by an efficient handwriting features. The goal is thus to draw closer to one another the ways of representing whole body and hand based gestures using skeleton data. Furthermore, in the context of dynamic hand gesture recognition using skeleton data, we noticed a lack of datasets that consider realistic conditions such as using two hands, noisy/missing data and also unsegmented streams. We therefore present below a new Leap Motion based dynamic hand gesture (LMDHG) dataset, which provides sequences of hand skeleton considering these realistic conditions. On the one hand, this dataset will allow the evaluation of existing approaches when neither the number of hands (one or two) used is not known beforehand. On the other side, it will open new scientific axes to consider such as the recognition in unsegmented streams.

III. Preliminaries

Our dynamic hand gesture recognition approach is based on a representation that was initially conceived for modelling whole body actions. This is motivated by the fact that both hand gestures and whole body actions consist in a hierarchy of skeleton joints evolving across time, i.e. human-made trajectories. To model these motions, one should often address similar intra-class and inter-class variations. Besides, the fact that whole body actions have been extensively addressed during the past few years, provides us with several feature-sets and representations that can efficiently model human motions. Exploring the aptitude of such representations to model dynamic hand gestures is a promising research trend which would allow the fusion of systems addressing both topics.

To this end, we retained in this study an existing feature set, namely HIF3D [1]. This choice is mainly motivated by three reasons. First, it has shown state-of-the-art results in the field of skeleton-based action recognition. Second, it is composed of only 89 features which makes it very compact compared to other feature-sets. Last, as it was derived from a handwriting feature-set [4], we expect that it is a good candidate to model 3D hand-based trajectories.

In particular, HIF3D's features report information about the spatial relationship of different trajectories but also information about the pattern shape. In fact, some of HIF3D's features (41 features) consist in a straightforward 3D adaptation of some handwriting descriptors while the remaining 49 features were specifically designed for 3D actions but still encode the same information as the one carried by equivalent features in handwriting area. For more details about HIF3D features we refer the reader to [1].

IV. Proposed dynamic hand gesture representation

Instead of representing a hand gesture with low-level features such as raw positions or angles, we propose to extract the higher level HIF3D features on the hands trajectories. In fact, similarly to the whole body, hands are a set of hierarchical joints. Despite of the trajectories amplitude produced either by the whole body or with hands only, the resulting pattern could be only considered as a 3D motion produced by a human. We therefore propose to draw a parallel between fingers and whole body joints trajectories so as to exploit the valuable improvement in terms of whole body modelling. This is particularly interesting as the employed HIF3D features are themselves the result of a transfer operated from 2D hand-drawn symbols to model 3D whole body gestures. By doing so, we want to go a step further toward the fusion of ways human-made trajectories are modelled.

In particular, we consider as input data the raw 3D positions of each finger tip along with the palm and wrist positions (Figure 1-a). The positions of each such joint are provided in a reference centred on the recording device (for instance the Leap Motion). When performing a given hand gesture, the successive positions of each joint constitute a 3D trajectory. These trajectories are then assembled to form a single 3D pattern S consisting in a single 3D trajectory: $S = \{s_1, ..., s_T, s_{T+1}, ..., s_{2T}, s_{2T+1}, ..., s_{3T}, s_{3T+1}, ..., s_n\}$, where T is the length of each single trajectory and $n = K \times T$ is the number of points in S. K is the number of considered joints (K=7 for a unique hand dataset or K=14 for a two hands dataset). Each point $s_i = (x_i, y_i, z_i)$ is located in the three-dimensional space.

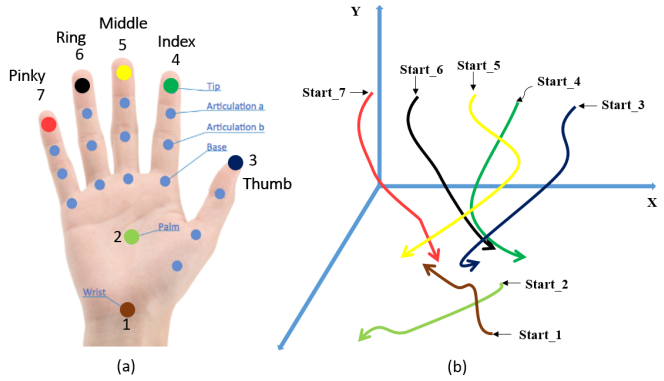

Fig. 1. (a) Selected joints for our dynamic hand gesture representation. (b) Illustration of a 3D pattern resulting from trajectories assembling.

An illustration of the obtained 3D pattern S is provided in Figure 1-b. For instance, the first and the second assembled trajectories correspond to the wrist joint and the palm joint, respectively. We retain the following order for assembling trajectories: wrist, palm, thumb, index, middle, ring and pinky.

The obtained 3D pattern is then passed to the HIF3D features extractor. The resulting features report high level information about both local and global shape of the 3D pattern. Among features that report local information there are for instance the first features $\mathbf{f_1}$, $\mathbf{f_2}$ and $\mathbf{f_3}$ (equation 1) which describe the starting of a gesture according to the X, Y and Z axes. In fact, positions of the first and last points constitute important features for distinguishing patterns in many situations.

$$\mathbf{f_1} = \frac{x_1 - c_x}{l} + \frac{1}{2}, \quad \mathbf{f_2} = \frac{y_1 - c_y}{l} + \frac{1}{2}, \quad \mathbf{f_3} = \frac{z_1 - c_z}{l} + \frac{1}{2} \quad (1)$$

with x_1, y_1 and z_1 the 3D coordinates of the first point composing the 3D pattern S. c_x, c_y and c_z are the coordinates of the center of the box bounding S.

Some other HIF3D's features like the convex hull based feature, referenced as $\mathbf{f_{89}}$, report a global information. In fact, the $\mathbf{f_{89}}$ feature capture the 3D shape of the resulting pattern by considering its convex hull volume V_H and the total length L of the 3D pattern S (equation 2).

$$\mathbf{f_{89}} = \frac{L^3}{V_H} \quad (2)$$

Furthermore, as HIF3D features allow only to express the hand shape variation, i.e. the spatial information, we need to additionally capture the temporal information. In fact, some gestures, such as reversed gestures which have the same spatial characteristics, could only be distinguished based on their temporal sequencing information. We therefore extract HIF3D features according to the so called Temporal Pyramid (TP). This is a commonly used technique in hand gestures recognition [15, 3] and consists in extracting features on overlapping sub-sequences resulting from a temporal split of the global gesture sequence. The temporal split could be performed at different levels and we propose to conduct it at only two levels as illustrated in Figure 2.

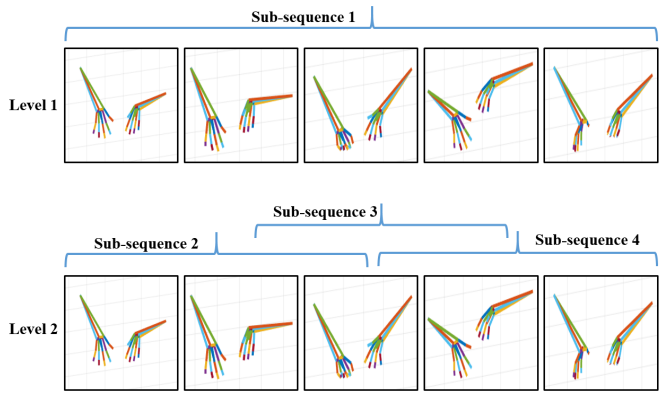

Fig. 2. Illustration of our representation using two temporal levels.

Thus, the combination of each window, which leads the final representation, consists in a concatenation of all the features extracted on the whole sequence and on the three overlapping sub-sequences. This leads a representation of only 356 elements, which is a very reduced dimension compared to state-of-the-art representations that can contain thousands of features.

As far as the classification is concerned, we employ the support vector machine (SVM) classifier. This supervised learning classifier is widely used in computer vision classification problems. In particular, we choose a polynomial kernel along with experimentally optimized parameters. We make use of the implementation contained in the LIBSVM package [2].

V. Collection of a new dynamic hand gesture dataset

For better evaluating existing approaches using segmented and unsegmented hand gestures, we collected a new dynamic hand gesture dataset, namely the LeapMotion Dynamic Hand Gesture (LMDHG) dataset [1]. In the following, we describe

[1]https://www-intuidoc.irisa.fr/en/english-leap-motion-dynamic-hand-gesture-lmdhg-database/

the collected dataset and then we present a baseline approach used to get preliminary results on unsegmented sequences.

A. LMDHG dataset

Different from existing dynamic hand gesture datasets, LMDHG contains unsegmented sequences of hand gestures performed with either one hand or both hands (an example is shown in Figure 3). There were 21 participants, each participant performed at least one sequence, resulting in 50 sequences. Each sequence contains 13 ± 1 class gestures leading to a total of 608 gesture instances. At the end of each gesture, the participant was asked to keep his hands above the Leap Motion before performing an other gesture. We labelled this no-gesture as an idle class. Furthermore, the order of classes in each sequence is aleatory and is different from one sequence to another.

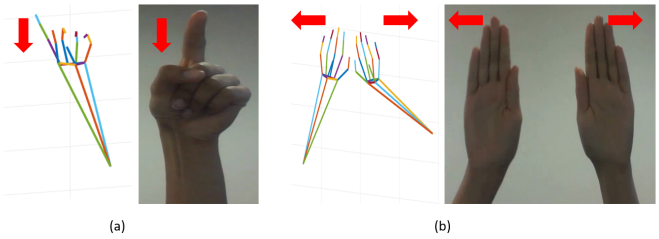

Fig. 3. (a) Line gesture performed with one hand. (b) Zoom gesture performed with two hands.

Each frame contains the 3D coordinates of 46 joints (23 joints for each hand). If one of the hands is not tracked then the position of its joints are set to zero. The Leap Motion sensor is used to collect our dataset. We provide also the labels and ground truth start and end of each gesture class in each sequence.

The gestures composing our dataset are listed in Table 1. They have been chosen to evaluate the robustness of hand gesture recognition systems against different challenges. We emphasized our main challenges compared to existing hand gesture datasets: (1) Evaluate robustness to gestures performed with one or two hands; (2) Evaluate robustness to noisy/ missing data; (3) Address the detection and recognition of hand gestures in unsegmented streams.

B. Baseline approach for unsegmented hand gesture recognition

To allow getting some preliminary results in the unsegmented case using the collected dataset, we present in this section a baseline approach. In fact, the approach uses a temporal sliding window of size T which is determined using the training set (equation 3).

$$T = \frac{1}{K} \sum_{i=1}^{i=K} T_i \tag{3}$$

Gesture	#Hands	tag name
Point to	1	HG1
Catch	1	HG2
Shake with two hands	2	HG3
Catch with two hands	2	HG4
Shake down	1	HG5
Shake	1	HG6
Draw C	1	HG7
Point to with two hands	2	HG8
Zoom	2	HG9
Scroll	1	HG10
Draw Line	1	HG11
Slice	1	HG12
Rotate	1	HG13

Table 1. List of the gestures included in the LMDHG dataset.

with K the size of the training dataset and T_i the temporal duration of each instance i of this training set.

During the testing step, features are extracted along the sliding window as done for pre-segmented gestures (Section IV). An SVM classifier is then used to provide the most probable class, noted $Class\_i$, along with its confidence score $Score(Class\_i)$. If the predicted class is different from the idle class and its confidence score overpasses a given confidence threshold θ, then the approach outputs that this detected class is occurring (equation 4).

$$Output = \begin{cases} \textbf{Class\_i} \quad , & \text{if} \quad Score(Class\_i) \geq \theta \\ & \& \quad Class\_i \neq idle \\ \textbf{No decision} \quad , & \text{otherwise} \end{cases} \tag{4}$$

VI. Experimental results

In this section we first evaluate our approach on an existing dynamic hand gesture dataset, namely DHG dataset [3], that has been recently published. We then provide preliminary results on the collected LMDHG dataset using segmented and unsegmented sequences.

A. DHG results

DHG is a recent dynamic hand gesture dataset presented in [3]. It contains instances of 14 pre-segmented hand gestures performed in two ways: using one finger and the whole hand. Each gesture is performed between 1 and 10 times by 28 participants in 2 ways, resulting in 2800 instances. Each frame of sequences contains a depth image, the coordinates of 22 joints both in the 2D depth image space and in the 3D world space forming a full hand skeleton.

Different protocols could be conducted on this dataset. We retain two protocols that consider only 3D skeleton data. According to the first protocol, gestures are grouped into

14 classes such that the number of fingers is not taken into account. In the second protocol, gestures are grouped into 28 classes such that two gestures of the same motion but are performed respectively with one finger and whole hand, belong to two different categories. The second protocol is obviously more challenging as the inter-class variability is lower than in the first protocol.

To allow a fair comparison with previous methods, we use 1960 instances for training the recognition model and 840 instances for evaluation. Table 2 reports the results of our approach along with those of previous methods in the cases of 14 and 28 gestures .

Method	14 gestures (%)	28 gestures (%)
HoWR [3]	35.61	-
SoCJ [3]	63.29	-
HoHD [3]	67.64	-
Oreifej and Liu [12, 14]	78.53	74.03
Devanne et al. [5, 14]	79.61	62.00
SoCJ + HoHD [3]	82.29	-
Guerry *et al.* [14]	82.90	71.90
SoCJ + HoHD + HoWR [3]	83.07	80.00
Ohn-Bar and Trivedi [11, 14]	83.85	76.53
De Smedt et al. [3, 14]	88.24	**81.90**
Our	**90.48**	80.48

Table 2. Comparison between our approach and previous approaches considering 14 and 28 gestures on DHG dataset.

Our approach outperforms previous methods when considering 14 gestures with a final score of 90.48%. Furthermore, we obtain a score of 80.48% when considering the more challenging case of 28 gestures. In the latter case we achieve the second best score which is very close to the state-of-the-art performance of [3], namely 81.90%.

Several conclusions could be drawn based on the achieved results. First, we have shown that recognizing 3D actions and 3D dynamic hand gestures share similar properties as they have been addressed similarly. Results also attest to the merits of HIF3Ds in modelling hand gestures and also the interest of considering high level features instead of using raw data or low level features. Last, it is important to note that our approach uses a subset of 7 joints out of 22 and an interesting perspective would be to study the impact of considering more joints as done by previous methods. For a detailed view of our results when using 14 gestures, we provide the confusion matrix in Figure 4.

B. LMDHG results

In this section the aim is to provide some preliminary results on the collected dataset when segmenting or no the sequences. For both cases we used 70% of the dataset to train the model, namely sequences 1 to 35, while 30% of the sequences are used for testing (sequences from 36 to 50). According to this split, most of people in the test set did not participate to collect the training set.

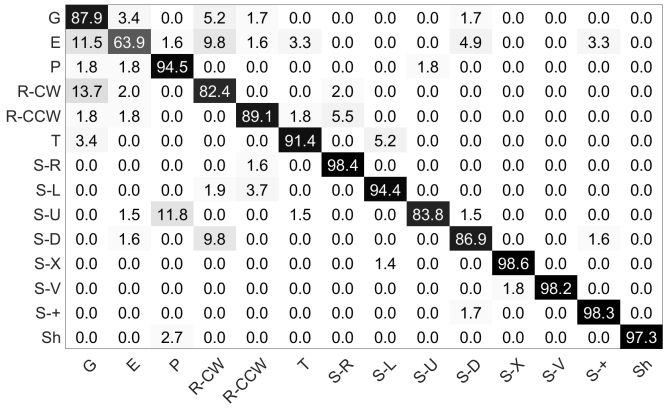

Fig. 4. Our confusion matrix on DHG dataset using 14 gestures.

For the segmented case, the overall score achieved with our approach equals to 84.78%. This score is particularly interesting as the LMDHG dataset contains noisy and incomplete gestures. Furthermore, as the dataset contains some gesture classes performed using one hand and some others which need two hands, the result testifies also the robustness of our approach against such problem. We provide the confusion matrix in Figure 5.

For evaluating our approach with unsegmented sequences, we use the F_{score} measure defined as:

$$F_{score} = 2 * \frac{Precision * Recall}{Precision + Recall} \tag{5}$$

Precision indicates how often the gesture is actually present when the system claims it is. Recall measures how many true gestures are recognized by the system.

Using the baseline unsegmented approach we achieve an overall F_{score} of 54.11%. This score shows the difficulty of recognizing gestures in unsegmented streams given that the same representation achieved a substantially higher accuracy in the segmented case. Besides the fact that the starts and

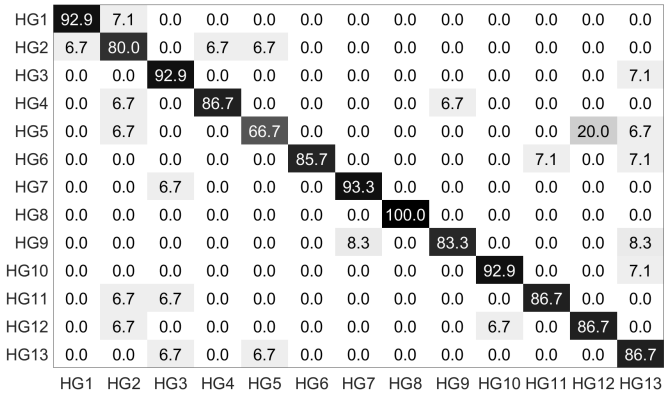

Fig. 5. Our confusion matrix on LMDHG dataset.

ends of gestures are not known beforehand, our collected dataset contains gestures of very different temporal sizes even though belonging to the same class. This makes it particularly difficult to determine an appropriate size of the sliding window. Furthermore, in some sequences, two consecutive gestures are often separated by meaningless motion instead of the conventional idle class motion. This is particularly a source of false positive detections which decreases the overall score. In future work, we plan to design a more robust segmentation method which takes into account the difficulties highlighted above.

VII. Conclusion

We proposed in this paper a new dynamic hand gesture recognition approach using skeleton joint data. The proposed approach is based on the HIF3D feature-set that was initially conceived for modelling whole body actions. This feature-set is particularly interesting as it was inspired by a handwriting recognition method. By doing so, our ultimate goal is to fuse the way hand gestures and whole body gestures are represented and recognized.

Our second contribution in this paper consists in the new hand gestures dataset LMDHG. Different from the publicly available datasets, LMDHG is composed of unsegmented streams of 13 hand gestures using either one or two hands. This new dataset would particularly allow considering the more challenging problem of real-time hand gesture detection an recognition in unsegmented streams.

The proposed approach was first evaluated using an existing pre-segmented gesture dataset, namely DHG. We achieved a state-of-the-art performance when using 14 gestures and the second best result when considering 28 gestures. We also reported results using our collected dataset both in the segmented and unsegmented case.

Future work will probably focus on the segmentation challenges highlighted during our experiments by applying a multiscale approach in which sliding windows of a couple of different lengths would be used.

References

[1] Said Yacine Boulahia, Eric Anquetil, Richard Kulpa, and Franck Multon. Hif3d: Handwriting-inspired features for 3d skeleton-based action recognition. In *Proceedings of the IEEE International Conference on Pattern Recognition*, 2016.

[2] Chih-Chung Chang and Chih-Jen Lin. Libsvm: a library for support vector machines. *ACM transactions on intelligent systems and technology (TIST)*, 2(3):27, 2011.

[3] Quentin De Smedt, Hazem Wannous, and Jean-Philippe Vandeborre. Skeleton-based dynamic hand gesture recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 1–9, 2016.

[4] Adrien Delaye and Eric Anquetil. Hbf49 feature set: A first unified baseline for online symbol recognition. *Pattern Recognition*, 46(1):117–130, 2013.

[5] Maxime Devanne, Hazem Wannous, Stefano Berretti, Pietro Pala, Mohamed Daoudi, and Alberto Del Bimbo. 3-d human action recognition by shape analysis of motion trajectories on riemannian manifold. *IEEE transactions on cybernetics*, 45(7):1340–1352, 2015.

[6] AS Elons, Menna Ahmed, Hwaidaa Shedid, and MF Tolba. Arabic sign language recognition using leap motion sensor. In *Computer Engineering & Systems (ICCES), 2014 9th International Conference on*, pages 368–373. IEEE, 2014.

[7] E Escobedo-Cardenas and G Camara-Chavez. A robust gesture recognition using hand local data and skeleton trajectory. In *Image Processing (ICIP), 2015 IEEE International Conference on*, pages 1240–1244. IEEE, 2015.

[8] Alexey Kurakin, Zhengyou Zhang, and Zicheng Liu. A real time system for dynamic hand gesture recognition with a depth sensor. In *Signal Processing Conference (EUSIPCO), 2012 Proceedings of the 20th European*, pages 1975–1979. IEEE, 2012.

[9] Giulio Marin, Fabio Dominio, and Pietro Zanuttigh. Hand gesture recognition with jointly calibrated leap motion and depth sensor. *Multimedia Tools Appl.*, 75 (22):14991–15015, 2016.

[10] Camille Monnier, Stan German, and Andrey Ost. A multi-scale boosted detector for efficient and robust gesture recognition. In *ECCV Workshops (1)*, pages 491–502, 2014.

[11] Eshed Ohn-Bar and Mohan Trivedi. Joint angles similarities and hog2 for action recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 465–470, 2013.

[12] Omar Oreifej and Zicheng Liu. Hon4d: Histogram of oriented 4d normals for activity recognition from depth sequences. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 716–723, 2013.

[13] Jamie Shotton, Toby Sharp, Alex Kipman, Andrew Fitzgibbon, Mark Finocchio, Andrew Blake, Mat Cook, and Richard Moore. Real-time human pose recognition in parts from single depth images. *Communications of the ACM*, 56(1):116–124, 2013.

[14] Quentin De Smedt, Hazem Wannous, Jean-Philippe Vandeborre, J. Guerry, B. Le Saux, and D. Filliat. 3D Hand Gesture Recognition Using a Depth and Skeletal Dataset. In Ioannis Pratikakis, Florent Dupont, and Maks Ovsjanikov, editors, *Eurographics Workshop on 3D Object Retrieval*. The Eurographics Association, 2017. ISBN 978-3-03868-030-7. doi: 10.2312/3dor.20171049.

[15] Chenyang Zhang, Xiaodong Yang, and YingLi Tian. Histogram of 3d facets: A characteristic descriptor for hand gesture recognition. In *IEEE International Conference on Automatic Face and Gesture Recognition*, 2013.