

3D Multistroke Mapping (3DMM): Transfer of Hand-Drawn Pattern Representation for Skeleton-Based Gesture Recognition

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

▶ **To cite this version:**

Said Yacine Boulahia, Eric Anquetil, Richard Kulpa, Franck Multon. 3D Multistroke Mapping (3DMM): Transfer of Hand-Drawn Pattern Representation for Skeleton-Based Gesture Recognition. 12th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2017), May 2017, Washington, DC., United States. pp.462 - 467, 10.1109/FG.2017.63 . hal-01555452

HAL Id: hal-01555452
https://hal.science/hal-01555452

Submitted on 4 Jul 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

3D Multistroke Mapping (3DMM): Transfer of hand-drawn pattern representation for skeleton-based gesture recognition

Said Yacine Boulahia*, Eric Anquetil, Richard Kulpa, Franck Multon
Université Européenne de Bretagne, France
INSA de Rennes, Avenue des Buttes de Coesmes, F-35043 Rennes
INRIA-IRISA, CNRS UMR 6074, Campus de Beaulieu, F-35042 Rennes

***Abstract*— Exergames involve using the fullbody to interact with an immersive world, which raises the challenge of capturing, processing and recognizing the action of the user even for cheap mocap systems such as the Microsoft Kinect. In fact, these recent technological advances have renewed interest in skeleton-based action recognition. Our review of related literature reveals that the issues encountered are not the result of random processes, which could simply be studied by using statistical tools, but are instead due to the fact that the pattern to be recognized, i.e. an action, was produced by a human being. 2D hand-drawn symbols are further examples of patterns resulting from a human motion. Therefore, the main contribution of this paper is to examine the validity of transferring the expertise of hand-drawn symbol representation to better recognize actions based on skeleton data. Principally, we propose a new action representation, namely the 3DMM, as an initial case-study illustrating how such transfer could be conducted. The experimental results, obtained over two benchmarks, confirm the soundness of our approach and encourage more thorough examination of the transfer.**

I. INTRODUCTION

Recognition of human actions has recently become an active research topic in computer vision. It has great potential in applications such as video surveillance, sport video analysis, human-computer interaction, motion retrieval, computer animation and so forth.

Recent advances in sensing technology have renewed interest in skeleton-based human action recognition and, since then, various skeleton-based recognition methods have flourished. In particular, we noticed that the best ones are not limited only to the pattern recognition aspect of the problem. In fact, it becomes clear that the many issues observed in human action patterns, such as the inter-class similarity and the intra-class variability, should not be considered as merely resulting from random processes. These issues are governed instead by kinematic constraints that must be considered while modelling these human motions.

This key observation is therefore, specifically and even exclusively due to the fact that the pattern to be represented and recognized, i.e. an action, was produced by a human being. 2D hand-drawn symbols are further examples of patterns produced by a human motion, since they result from the movement of a human hand, and the issues encountered during their representation are accordingly due to more than random processes. Interestingly, the proposed methods for the recognition of online 2D hand-drawn patterns are well ahead as far as the consideration of human-related constraints in the conception of their representations is concerned.

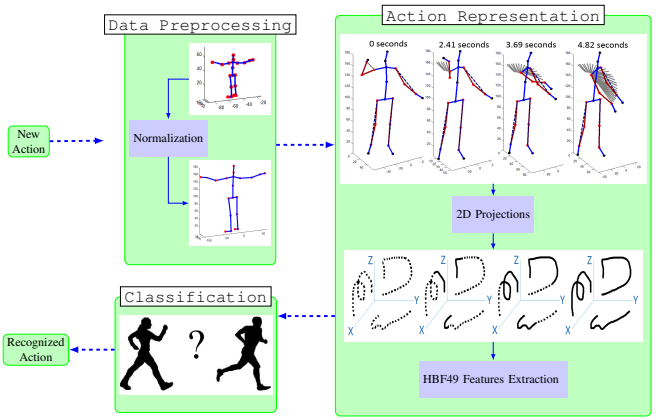


Fig. 1. Major steps constituting the proposed skeleton-based action recognition approach.

Our work is therefore guided by the following question: *"Is it possible to represent and recognize 3D skeleton-based human actions by transferring the rich expertise acquired in representing online hand-drawn symbols, in terms of reasonable dimensionality, high performance and the consideration of kinematic constraints?"*. A positive answer to this approach would not only enhance the research about action recognition, by avoiding the repetition of trial and error, but would even allow for the emergence of transversal recognition approaches which could recognize both hand-drawn symbols and 3D human actions.

Considering this research trend, we put forward in this paper a new skeleton-based action representation, namely the **3DMM**, in the form of an initial case-study illustrating how such a transfer can be conducted. This approach consists of mapping 3D trajectory data into 2D feature space; hence, the name: 3D Multistroke Mapping (**3DMM**). We purposely built on an existing online hand-drawn feature-set, namely the HBF49 [1], given its ability to deal with patterns of a diverse nature.

The general pipeline of the **3DMM** approach is composed of three stages (Figure 1). The first step, called data preprocessing, consists of applying a series of operations to the input skeleton data in order to specifically ensure against the invariance of different morphologies. The second step deals with the actual representation transfer of the action.

This consists of projecting the processed joint trajectories on each of the Cartesian planes in order to extract the so-called HBF49 features. Finally, the classification was carried out based on two classifiers namely Multilayer Perceptron (MLP) and Support Vector Machines (SVM).

The remainder of this paper is organized as follows. Section II, concerns the presentation of the work related to human action recognition focusing on the approaches based on skeleton data. Here, the HBF49 feature-set is briefly presented. Section III, centers around, the presentation of the **3DMM** approach as an initial case-study of such a transfer. The experimental results obtained over two skeleton-based benchmarks, including HDM05 [2] and UTKinect-Action datasets [3], are presented in Section IV. Section V concludes this paper and discusses future work.

II. Related Work

Our aim in this paper is to explore a new trend consisting of how skeleton-based actions could be represented based on previous expertise of online hand-drawn symbol recognition.

On the one hand, it is important to highlight that a hand-drawn symbol does not only refer to a handwritten text. In fact, it covers a wider field which, in addition to handwriting, includes sketch diagrams, signatures, free drawings and pen-based control commands.

Early online hand-drawn recognition approaches focused on recognizing only single-stroke symbols. A stroke is the trace of a pen-tip movement which starts at pen-down and ends at pen-up. However, multistroke symbol recognizers have to take into account shape variations, differences in stroke ordering, and a varying number of strokes.

In particular, Delaye *et al.* [1] conceived a new feature-set, called HBF49 (Heterogeneous Baseline Feature Set). Compared to most feature-sets proposed in other hand-drawn symbol recognition approaches, this set, composed of 49 features, has a great advantage regarding its low dimensionality. Furthermore, according to its authors, the HBF49 feature-set aims at recognizing hand-drawn symbols in a very wide range of different contexts. It is able to describe any kind of symbol, either monostroke or multistroke and includes some features that are sensitive to orientation and stroke order. Last, the high performance achieved by the HBF49 over different kind of benchmarks is another reason why we retained this feature-set.

On the other hand, action recognition based on skeleton data is attracting an increasing attention among the computer vision community. In fact, a skeletal representation can not only capture the essential structure of a subject in an easily-understood and compact way, but it is also insensitive to variations in viewpoint, human body scale and motion speed.

The histogram-based representations, such as the Histograms of 3D Joint Locations (HOJ3D) [3] or the Histogram of Oriented Displacements (HOD)[4], are among the first techniques used to model human actions based on skeleton data. However using only histograms made it difficult to distinguish gestures that are the reverse of each other.

An other approach consisted then in projecting the joint trajectories into a more adapted representation spaces. For instance, Vemulapalli *et al.* [5] have explicitly modelled displacement between body parts as curves belonging to the special Euclidean group SE(3).

Recently, Chaudhry *et al.* [6] introduced a new research trend by using the 3D raw coordinates to build a bio-inspired representation through leveraging findings in the area of static shape encoding in the neural pathway of the primate cortex. Similarly Zhang and Parker [7] conceived a bio-inspired representation, called BIPOD, by spatially decomposing 3D human skeleton trajectories and projecting them onto three anatomical planes, and then encoding high-order temporal dependencies.

Globally, previous skeleton-based approaches aimed to find the best way for extracting the most discriminant information to represent an action. One can notice that recent approaches, such as the bio-inspired representations [6], [7], tend to take into account the physiological aspect of this particular pattern recognition problem. In our current work, we aim to go a step forward as far as the consideration of kinematic constraints is concerned. In fact, we think that an interesting research trend into 3D skeleton-based human action recognition could be based on extending the achievements realized in online 2D symbol recognition. To this end, we have retained the features of the HBF49 introduced above since it was designed for the recognition of most kinds of hand-drawn symbols while considering, at the same time, the kinematic aspects of the problem. The approach resulting from this transfer is outlined in Section III.

III. Proposed approach

This section deals with our proposed action recognition approach, namely: 3D Multistroke mapping (**3DMM**). Our goal is to show that the proposed 2D to 3D transfer may be promising, and, hence, we do not attempt to find the best way to conduct this transfer. In this regard, we present successively the three main steps composing the **3DMM** approach, namely data preprocessing, action representation and classification which are illustrated in Figure 1.

A. Data preprocessing

Preprocessing mainly aims to tackle the anthropometric differences between individuals. For example, walking movements can differ in speed and footsteps since, in general, subjects have different limbs' length. To strengthen such anthropometric invariance, we retained the data transformation that was initially proposed for motion editing in computer animation by [8] and later adapted for gesture recognition by [9].

According to this proposition, the 3D trajectories of 12 joints associated with movement of arms and feet, including the shoulders, elbows, wrists, hips, knees and ankles should be considered in order to capture the information necessary for an action (Figure 2). The 3D position j_i^t of each joint j_i at time t is given according to the coordinate system centred in the hip joint of the skeleton.

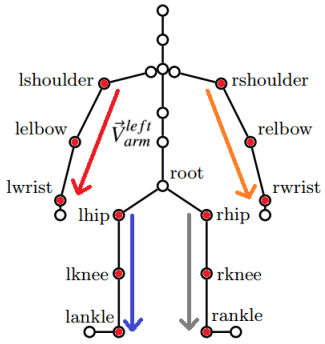


Fig. 2. Selected joints and the associated morphology-independent vectors.

Given these coordinates, four vectors are computed corresponding to each body part, namely, Left Arm, Right Arm, Left Leg and Right Leg. For instance the Left Arm vector noted $\vec{V}_{LeftArm}(t)$ relates to the two end joints of the left-up body part which are Left Shoulder j^t_{LSh} and Left Wrist j^t_{LWr}, at time t (Eq.1).

$$\vec{V}_{LeftArm}(t) = \overrightarrow{j^t_{LSh} j^t_{LWr}} \tag{1}$$

The three other vectors namely, $\vec{V}_{RightArm}(t)$, $\vec{V}_{LeftLeg}(t)$ and $\vec{V}_{RightLeg}(t)$ are computed in a similar way. All these vectors are represented in Figure 2.

After that, total-extension normalization is performed for each vector to make it independent from body size. That is to say, the vector $\vec{V}_{LeftArm}(t)$ for instance is normalized by the total arm length to obtain the morphology independent vector of the left-up body part at time t which we refer to as $\vec{V}^{MI}_{LeftArm}(t)$. This normalization is expected to reduce the influence of the subject morphology. The used formula is given in Eq.2, where j^t_{LElbow} refers to the Left Elbow joint position at time t.

$$\vec{V}^{MI}_{LeftArm}(t) = \frac{\vec{V}_{LeftArm}(t)}{||j^t_{LSh} j^t_{LElbow}|| + ||j^t_{LElbow} j^t_{LWr}||} \tag{2}$$

For any given action, the variation across time of the vectors introduced gives rise to four morphology-independent trajectories used as input data to the next step, namely action representation.

B. Action representation

In the following passages, we shall introduce the assumption that led to the proposed transfer, which we refer to as multistroke assumption and then, based on this, explain the feature extraction procedure. Finally, we aim to justify the feature selection used to obtain the final action representation.

1) Multistroke assumption and feature extraction: In this step, the four preprocessed 3D trajectories have to be projected on to the three planes, namely, (XOY), (YOZ) and (ZOX). As shown in Figure 3(a), this projection yields four trajectories in each plane. Then, instead of designing a new set of features as was done in existing approaches, we opted for using the previously mentioned HBF49 set which has already proved itself to be very efficient in the field of 2D trajectory recognition.

After projection, the hand-drawn HBF49 features are extracted separately from each plane. In each such a plane the feature extraction might be performed either according to the monostroke or the multistroke strategy. (1) In monostroke feature extraction, each trajectory is considered as a 2D monostroke "symbol" and is, thus, processed independently of the others. (2) In multistroke feature extraction strategy, the separate projected trajectories are considered as different strokes composing the same 2D "symbol", and thus we extract the features as for a multistroke hand-drawn symbol. We therefore followed a multistroke modelling which considers, in fact, a 3D action as three projected multistroke "symbols" where each stroke is related to one of the four body parts (limbs) previously identified (Figure 3(b)).

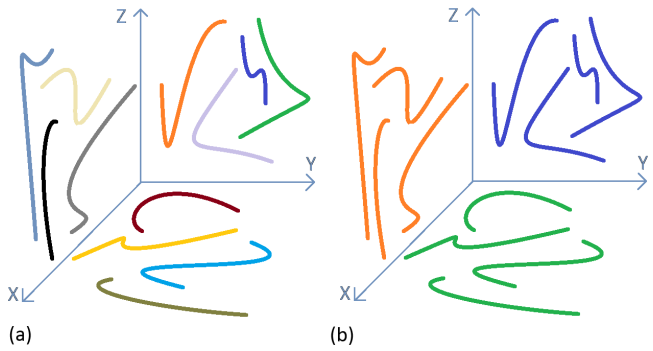


Fig. 3. For a motion involving 4 limbs, each trajectory is projected on to the three planes. (a) In the monostroke version, each single trajectory is considerd alone, hence the different colour for each such a trajectory. (b) In the multistroke version, a symbol is composed of all the projected trajectories belonging to the same plane, and are coloured the same in this illustration.

As has been discussed in related work, the HBF49 can characterize both monostroke and multistroke symbols with the same number of features. That is, the multistroke variant generates 49 features per plane and thus results in an action representation composed of 147 features ($3 * 49$). Compared to monostroke, multistroke preserves correlation and offers a smaller feature-set. In our work, we have retained, therefore, this variant for the action representation step; hence, the name of the approach : 3D Multistroke Mapping (**3DMM**). We consider that the extension of the multistroke concept, found in hand-drawn symbol recognition, used to handle several skeleton trajectories simultaneously, is an original approach which addresses the plurality of trajectories often faced in action recognition.

Furthermore, since the HBF49 features do not capture the temporal dependence inside a pattern sequence, we built our representation according to a multilevel temporal split of the sequence. This way of integrating temporality has been commonly adopted in many action recognition approaches using different variants as in the works of [10], [11], [4]. In our approach we adopted a slightly different variant for this temporal split. The top level representation is computed

over the entire sequence according to the multistroke scheme defined above. This yields the first 147 features. The lower levels are computed over smaller overlapping windows of the entire sequence. For the purpose of this paper, we have limited the number of levels to two, which brings the total length of the representation to 588 ($147 * 1 + 147 * 3$).

2) Feature selection: As we extract the same features according to each of the three projection planes, it is likely that some of the generated features are found to be redundant and can be dropped. The **3DMM** approach, therefore, carries out a features selection step before moving on to the classifier learning step, using a fairly widespread selection algorithm called One-R [12].

C. Action classification

The last step of our **3DMM** recognition approach concerns classification which was undertaken with two popular algorithms, namely, Multilayer Perceptron (MLP) and Support Vector Machines (SVM).

These two standard classifiers were employed with default configuration. For the MLP classifier, we used one hidden layer where the number of nodes equals the number of classes plus the number of used features. For the SVM classifier, we used a Polynomial kernel along with experimentally fixed parameters: the polynomial degree is set to 3, the gamma parameter is set to 10^2 and the C parameter is set to 10. However, it should be noted that it may be relevant to optimize these parameters in order to further improve the recognition performance.

IV. Experimental validation

We have evaluated the performance of the **3DMM** approach on two challenging and publicly available skeleton-based benchmarks including HDM05 [2] and UTKinect [3]. In this section, we first describe these databases. Then, we compare the performance of our approach with that of previous state-of-the-art action recognition techniques, insisting on the fact that this comparison is made with approaches that used only skeleton-data.

A. Datasets

The HDM05 dataset contains around one hundred motion classes, including various walking and kicking motions, cartwheels, jumping jacks, grabbing and depositing motions, squatting motions and so on. Each motion class contains 10 to 50 different instances of the same type of motion, covering a broad spectrum of semantically meaningful variations. Each sequence was captured using a motion capture system at a rate of 30 frames per second. We used the same action classes and test settings as in [13], where the data of three subjects was used for training (the actors bd, mm and tr) and the data of the others for testing (the actors bk and dg), resulting in 250 mocap sequences.

The second benchmark used in our evaluation is the UTKinect-Action dataset which was collected using depth sequences. This dataset contains 10 types of human actions which take place in indoor settings including: walking, sitting down, standing up, picking up, carrying, throwing, pushing, pulling, waving and clapping hands. Each action was collected from 10 different people who repeated each motion twice. Altogether, the dataset contains 6220 frames of 200 action samples. The length of sample actions ranges from 5 to 120 frames. This dataset is particularly interesting for our study due to the significant variation between different instances of the same action and the great variation in the duration of the actions.

B. Results and discussion

In this section, we present and discuss the results obtained firstly, on the HDM05 dataset and then, on the UTKinect-Action dataset.

1) HDM05 motion capture dataset results: To investigate the performance of the proposed approach, we conducted a series of experiments on the HDM05 motion capture dataset. We first used the same combination of subjects in the training and test datasets as proposed by [13] in which the data of three subjects was used for training (the actors bd, mm and tr) and the data of the others for testing (the actors bk and dg).

We built a two-Level representation by proceeding as follows: over each window of the lower level we extracted the best selected features that were determined on the top level. Depending on the classifier used, MLP or SVM, we extracted a different number of features on each window, namely 20 features when using the MLP and 100 features when using SVM. As result, we finished up with two temporal representations composed of 80 ($20*1+20*3$) and 400 ($100*1+100*3$) features respectively. Table I presents not only our results with and without temporal splitting but also those obtained by several state-of-the-art approaches.

Method	#Features	Reco. rate (%)
MIJA/MIRM + LCSS [14]	-	85.23
SMIJ + Nearest neighbour [13]	-	91.53
LDS + SVM [6]	-	91.74
Skeletal Quads + SVM [11]	9360	93.89
Cov3DJ + SVM [10]	43710	95.41
BIPOD + SVM [7]	-	96.70
HOD + SVM [4]	1116	97.27
3DMM + SVM + Level = 1	100	**91.74**
3DMM + MLP + Level = 1	20	**92.66**
3DMM + SVM + Level = 2	400	**94.49**
3DMM + MLP + Level = 2	80	**94.49**

TABLE I
COMPARISONS BETWEEN **3DMM** APPROACH, WITH AND WITHOUT TEMPORAL SPLIT, AND PREVIOUS APPROACHES ON THE HDM05 DATASET.

On this dataset, our approach, with the two-Level temporal representation and the SVM classifier, achieves an average accuracy of 94.49% while it achieves only 91.74% using the same classifier but without temporal splitting. A similar observation is noticed when using the MLP classifier which achieves 94.49% with temporal splitting and 92.66% without it. Compared to the most recent approaches, we achieved very decent results and our approach even outperforms some of the more elaborate approaches such as the SMIJ [13]

Method	Walk	Sit	Stand	Pick	Carry	Throw	Push	Pull	Wave	Clap	OverAll (%)
LTI + HMM [15]	**63.16**	100	100	100	83.33	**61.11**	90	100	85	85	86.76
Grassmann + SVM [16]	100	80	100	100	100	**60**	**65**	85	100	95	88.5
HOJ3D + HMM [3]	96.5	91.5	93.5	97.5	97.5	**59**	81.5	92.5	100	100	90.95
DS-SRC + Nearest neighbour [17]	90	100	95	85	100	**75**	90	95	100	80	91
STFC + SVM [18]	90	95	95	100	**65**	90	95	100	100	85	91.5
HSOM + VMM [19]	-	-	-	-	-	-	-	-	-	-	94.5
3DMM + SVM + Level = 1 (90 features)	85	100	100	95	**90**	**100**	95	100	100	95	**96**
3DMM + MLP + Level = 1 (40 features)	90	95	100	95	**90**	**95**	100	100	100	95	**96**
3DMM + SVM + Level = 2 (360 features)	85	100	100	95	**90**	**100**	95	100	100	95	**96**
3DMM + MLP + Level = 2 (160 features)	90	95	100	95	**90**	**95**	100	100	100	95	**96**

TABLE II

COMPARISONS BETWEEN **3DMM** APPROACH AND PREVIOUS APPROACHES ON THE UTKINECT-ACTION DATASET ACCORDING TO THE LOSEQOCV PROTOCOL. RECOGNITION RATE (%) GIVEN FOR EACH ACTION CLASS.

or the Skeletal Quads [11] representations. Despite the fact that the **3DMM** approach is an initial case-study for the transfer from hand-drawn symbol representation to model whole-body actions, the high accuracy obtained with such an approach already testifies to the soundness of the transfer. The results achieved confirm the maturity of the hand-drawn features, here HBF49, in terms of capturing the discriminant information of human-produced patterns.

Moreover, our approach competes with the most efficient skeleton-based approaches that have previously been evaluated on this dataset such as Cov3DJ [10], BIPOD [7] or HOD [4]. Compared to these sophisticated approaches, the **3DMM** is to be considered as an initial case-study for adapting the existing hand-drawn symbol recognition works. Even though, its performance is close to these state-of-the-art results. Furthermore, while these representations, referred to above, suffer from a very high dimensionality (Table I), our proposition is composed of a reduced number of features (only 80 features with the MLP classifier versus more than 1000 features in previous approaches). The effectiveness of our approach is therefore fostered by its simplicity (reduced dimensionality). According to the very promising results of the **3DMM**, we estimate that the main hypothesis of this paper, concerning the possible transfer of hand-drawn symbol recognition expertise to efficiently represent 3D whole-body actions, has been largely confirmed.

Other experiments were carried out on the HDM05 dataset to verify the ability of the **3DMM** approach to handle morphological variabilities. In fact, data of different subjects used to make up the training and test dataset can affect the classification accuracy of a specific algorithm. Therefore, to remove this bias, we evaluated our approach using all possible combinations of three subjects out of five (C_5^3 = 10) in different sets of training and unseen test datasets respectively. Each combination was evaluated 10 times with and without a temporal splitting of the actions. The results show an average accuracy, obtained by using the SVM and MLP classifiers without the temporal splitting, as 83.86% and 83.17% respectively. On the other side, the use of the temporal based representation by means of the SVM and MLP classifier respectively brought accuracies to 90.73% and 92.99%. This evaluation is different from the ones conducted by previous approaches where only one specific combination of subjects was used in the training and test datasets respectively. Our results show the significant strength of the **3DMM** approach and its robustness to morphological variations in the training dataset. Since no previous approach has been evaluated on this dataset according to this last protocol, our results may even serve as a reference for future action recognition approaches.

2) UTKinect-Action dataset results: Finally, we compared the **3DMM** approach to previous approaches on the UTKinect-Action dataset. For a fair comparison with the previous representations, we followed the same experimental procedure initially proposed by [3] which consisted of performing the Leave One Sequence Out Cross Validation (LOSeqOCV) on the 200 sequences constituting this dataset. For each iteration, a single sequence was used for the test while 199 sequences were used for learning. In addition to the calculation of an average rate, we separately calculated the recognition rate for each of the 10 action classes. The results obtained by our approach with and without temporal splitting are presented in Table II.

Our experiments have proved undoubtedly that the **3DMM** approach is effective on all action classes, unlike previous approaches which collapsed significantly for some of them, for instance in, the "Throw" class for [3], [16], [15] and the "Carry" class for [18]. Such high rates of accuracy on all types of actions naturally lead to an overall average greater than all previous skeleton-based approaches evaluated on the UTKinect-Action dataset, that is 96% with and without temporal splitting. Compared to previous benchmarks, this result is especially important given the noisy nature of this dataset which was collected by a Kinect. Compared to current approaches evaluated on this benchmark, according to the LOSeqOCV protocol, the **3DMM** approach reduces the error rate by 27.27%, i.e. ($\frac{1.5}{5.5}$).

Similarly to [18], a cross-subject test was also performed, in which the data of half of the subjects were used for training and the remaining for testing. This test is different from LOSeqOCV, where the majority of the subjects were used for training. The recognition rates are presented in Table III. It is noted that we do not compare with [5], as they use a different cross-subject protocol.

Method	Reco. rate (%)
STFC [18]	85
Joint features [20]	87.90
3DMM + MLP + Level = 1 (70 features)	**90.32**
3DMM + SVM + Level = 1 (100 features)	**90.46**
3DMM + MLP + Level = 2 (280 features)	**90.81**
3DMM + SVM + Level = 2 (400 features)	**91.51**

TABLE III

COMPARISON BETWEEN **3DMM** APPROACH AND THE BEST TESTED APPROACH ON UTKINECT-ACTION DATASET, ACCORDING TO A CROSS-SUBJECT-VALIDATION.

This last experiment, based on cross-subjects evaluation, tells a great deal about the greater strength of the **3DMM** approach when faced with subject morphology variability. As presented in Table III, the most accurate result is 91.51% which was obtained by means of a two-Level representation and the SVM classifier. With our first attempt to transfer hand-drawn symbol recognition techniques, we have outperformed all previous approaches that have used only skeleton data from the UTKinect according to this protocol.

V. CONCLUSION AND FUTURE WORK

We have presented, in this paper, a novel research trend consisting in the transfer of hand-drawn symbol recognition expertise to represent 3D actions. We have based this proposition on the observation that patterns produced by a human motion, in particular 2D hand-drawn symbols and 3D actions, share several important properties. One of the most important similarity we have highlighted is the fact that both human performances are driven by similar motion control laws. Since this observation is exclusive to the domain of hand-drawn symbols and 3D actions, both being the result of a human motion, we hypothesized that both recognition problems could be addressed in similar ways.

The 3D Multistroke Mapping (**3DMM**) approach is the outcome of such a transfer. To this end, we built on an existing online feature-set to represent hand-drawn symbols, namely the HBF49. This choice was motivated by the ability of the HBF49 to deal with many application contexts and its robustness to patterns of diverse nature. We have emphasized the fact that the extension of the multistroke concept found in the field of hand-drawn symbol recognition and used to handle simultaneously several joints trajectories, is an original way of addressing the plurality of trajectories often faced in action recognition.

The experimental results carried out on two challenging datasets, namely HDM05 and the UTKinect-Action datasets show that the **3DMM** approach is a promising human action representation which competes with state-of-the-art methods based only on skeleton data. Future work will probably focus on addressing multi-person interactions on the basis of existing hand-drawn representations.

REFERENCES

[1] A. Delaye and E. Anquetil, "Hbf49 feature set: A first unified baseline for online symbol recognition," *Pattern Recognition*, vol. 46, no. 1, pp. 117–130, 2013.

[2] M. Müller, T. Röder, M. Clausen, B. Eberhardt, B. Krüger, and A. Weber, "Documentation mocap database hdm05," 2007.

[3] L. Xia, C.-C. Chen, and J. Aggarwal, "View invariant human action recognition using histograms of 3d joints," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 20–27, 2012.

[4] M. A. Gowayyed, M. Torki, M. E. Hussein, and M. El-Saban, "Histogram of oriented displacements (hod): describing trajectories of human joints for action recognition," in *Proceedings of the International Joint Conference on Artificial Intelligence*, pp. 1351–1357, 2013.

[5] R. Vemulapalli, F. Arrate, and R. Chellappa, "Human action recognition by representing 3d skeletons as points in a lie group," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 588–595, 2014.

[6] R. Chaudhry, F. Ofli, G. Kurillo, R. Bajcsy, and R. Vidal, "Bio-inspired dynamic 3d discriminative skeletal features for human action recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 471–478,2013.

[7] H. Zhang and L. E. Parker, "Bio-inspired predictive orientation decomposition of skeleton trajectories for real-time human activity prediction," in *Proceedings of the IEEE International Conference on Robotics and Automation*, pp. 3053–3060, 2015.

[8] R. Kulpa, F. Multon, and B. Arnaldi, "Morphology-independent representation of motions for interactive human-like animation," in *Computer Graphics Forum*, vol. 24, pp. 343–351, 2005.

[9] A. Sorel, R. Kulpa, E. Badier, and F. Multon, "Dealing with variability when recognizing user's performance in natural 3d gesture interfaces," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 27, no. 08, 2013.

[10] M. E. Hussein, M. Torki, M. A. Gowayyed, and M. El-Saban, "Human action recognition using a temporal hierarchy of covariance descriptors on 3d joint locations," in *Proceedings of the International Joint Conference on Artificial Intelligence*, vol. 13, pp. 2466–2472, 2013.

[11] G. Evangelidis, G. Singh, and R. Horaud, "Skeletal quads: Human action recognition using joint quadruples," in *Proceedings of the IEEE International Conference on Pattern Recognition*, pp. 4513–4518, 2014.

[12] R. C. Holte, "Very simple classification rules perform well on most commonly used datasets," *Machine learning*, vol. 11, no. 1, pp. 63–90, 1993.

[13] F. Ofli, R. Chaudhry, G. Kurillo, R. Vidal, and R. Bajcsy, "Sequence of the most informative joints (smij): A new representation for human skeletal action recognition," *Journal of Visual Communication and Image Representation*, vol. 25, no. 1, pp. 24–38, 2014.

[14] H. Pazhoumand-Dar, C.-P. Lam, and M. Masek, "Joint movement similarities for robust 3d action recognition using skeletal data," *Journal of Visual Communication and Image Representation*, vol. 30, pp. 10–21, 2015.

[15] L. L. Presti, M. La Cascia, S. Sclaroff, and O. Camps, "Gesture modeling by hanklet-based hidden markov model," in *Proceedings of the Asian Conference on Computer Vision*, pp. 529–546, 2014.

[16] R. Slama, H. Wannous, M. Daoudi, and A. Srivastava, "Accurate 3d action recognition using learning on the grassmann manifold," *Pattern Recognition*, vol. 48, no. 2, pp. 556–567, 2015.

[17] I. Theodorakopoulos, D. Kastaniotis, G. Economou, and S. Fotopoulos, "Pose-based human action recognition via sparse representation in dissimilarity space," *Journal of Visual Communication and Image Representation*, vol. 25, no. 1, pp. 12–23, 2014.

[18] W. Ding, K. Liu, F. Cheng, and J. Zhang, "Stfc: Spatio-temporal feature chain for skeleton-based human action recognition," *Journal of Visual Communication and Image Representation*, vol. 26, pp. 329–337, 2015.

[19] W. Ding, K. Liu, F. Cheng, and J. Zhang, "Learning hierarchical spatio-temporal pattern for human activity prediction," *Journal of Visual Communication and Image Representation*, vol. 35, pp. 103–111, 2016.

[20] Y. Zhu, W. Chen, and G. Guo, "Fusing spatiotemporal features and joints for 3d action recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 486–491, 2013.