HIF3D: Handwriting-Inspired Features for 3D Skeleton-Based Action Recognition

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

2016 23rd International Conference on Pattern Recognition (ICPR)
Cancún Center, Cancún, México, December 4-8, 2016

HIF3D: Handwriting-Inspired Features for 3D skeleton-based action 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
Email: {said-yacine.boulahia, eric.anquetil, richard.kulpa, franck.multon}@irisa.fr

***Abstract*—Action recognition based on human skeleton structure represents nowadays a prosper research field. This is mainly due to the recent advances in terms of capture technologies and skeleton extraction algorithms. In this context, we observed that 3D skeleton-based actions share several properties with handwritten symbols since they both result from a human performance. We accordingly hypothesize that the action recognition problem can take advantage of trial and error already carried out on handwritten patterns. Therefore, inspired by one of the most efficient and compact handwriting feature-set, we propose in this paper a skeleton descriptor referred to as Handwriting-Inspired Features (HIF3D). First of all a data preprocessing is applied to joint trajectories in order to handle the variabilities among actor's morphologies. Then we extract the HIF3D features from the processed joint locations according to a time partitioning scheme so as to additionally encode the temporal information over the sequence. Finally, we selected the Support Vector Machine (SVM) to achieve the classification step. Evaluations conducted on two challenging datasets, namely HDM05 and UTKinect, testify the soundness of our approach as the obtained results outperform the state-of-the-art algorithms that rely on skeleton data.**

***Index Terms*—Human action recognition, Skeleton-based features, HIF3D, RGB-D data, Handwriting-Inspired Features, Joint trajectory modelling.**

I. Introduction

Recognizing human actions is an active area of research in computer vision and pattern recognition. It has great potential in applications such as surveillance, sport analysis, human-computer interaction and entertainment. Despite the large amount of studies that has been conducted and many promising advances, it is far from being a solved problem.

Technically, an action is a sequence generated by a human subject during the performance of a task. Action recognition deals with the process of labelling such motion sequence with respect to the depicted motions. The often cited experiment of Johansson [1] showed that humans can recognize actions by observing only the main joints of a human body. This observation motivated the emergence of a plethora of skeleton-based approaches which initially used as inputs 2D images captured by a single RGB camera from which it was needed to extract the skeleton structure. Unfortunately, such 3D extraction from 2D video sensors was difficult since the monocular RGB data is highly sensitive to various factors like illumination changes, variations in view-point, occlusions and background clutter. As an alternative line of work numerous researchers have started using motion capture (MoCap) systems to extract 3D joint positions by using markers and high precision camera array. While these marker-based equipments provide accurate measurements of body poses and joint locations, they are often tedious and very expensive.

More recently, the release of the Microsoft Kinect sensor and the seminal algorithm of Shotton et al. [2] largely eases the task of extracting 3D joint positions. This advance resulted in a renewed interest towards skeleton-based human action recognition. Particularly, there has been since then a proliferation of works which propose new set of features to represent action relying on skeleton information.

In this context, we observed that skeleton-based human actions share several properties with handwritten symbols since they both result from a human performance. For instance, in both cases one needs to model the progress of body part trajectories and to handle the intra-class variability. Such an observation suggests that 3D action difficulties may be tackled by adapting solutions already proposed for handwriting recognition. Despite the evident similarity between those two problems and the abundance of solutions in handwriting literature, no previous work considered the transposal of handwriting recognition achievements to the recognition of 3D actions.

We therefore intend in this paper to explore the validity of such transposal by conceiving a new set of features, referred here to as Handwriting-Inspired Features (HIF3D), which are based on one of the most efficient and recent handwriting feature-set introduced by Delaye and Anquetil [3]. To that end, we first apply for joint input data the preprocessing suggested by [4] in order to tackle morphological variabilities. After that we build our action representation by extracting the HIF3D features according to a temporal partitioning scheme so as to integrate the performing order of subactions. Finally a Support Vector Machine (SVM) is trained on the output representations to achieve the classification step.

The rest of the paper is organized as follows. In Section II, we provide the background supporting the proposed representation. In Section III, we introduce our set of Handwriting-Inspired Features (HIF3D) to model human skeletal motion. In Section IV we test the proposed representation on two datasets, namely HDM05 [5] and UTKinect [6], and conclude in Section V.

978-1-5090-4846-5/16/$31.00 ©2016 IEEE

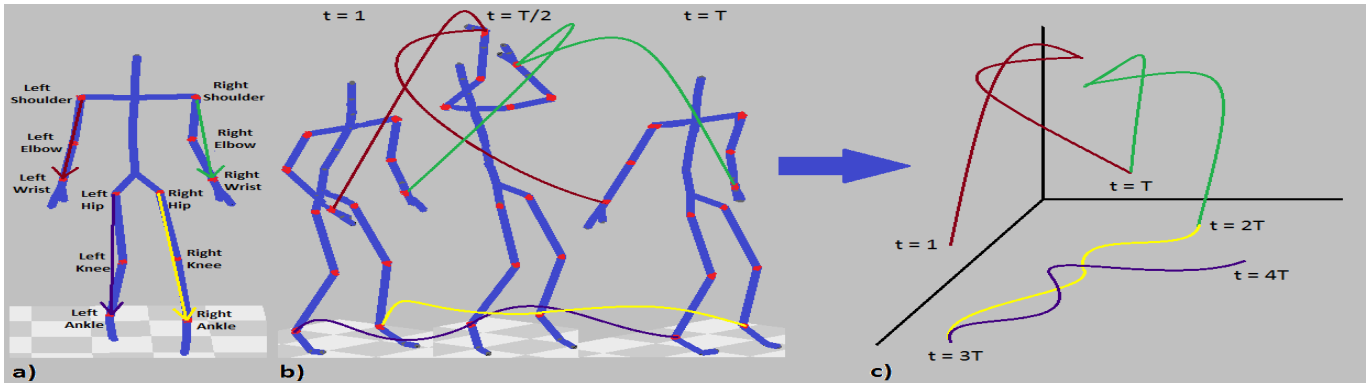

Fig. 1. (a) Selected joints with the associated normalized vectors (collored arrows) and (b) illustration of the four morphology-independent trajectories and (c) the 3D multistroke pattern resulting from the trajectories assembling.

II. Preliminaries

In this section we first provide a short overview of the previous skeleton-based representations. Then we introduce the handwriting features proposed by [3], namely HBF49. Last we briefly outline the preprocessing setup suggested by [4] which we apply for raw data to tackle morphological variabilities.

A. Action recognition approaches

Human action recognition approaches can be broadly grouped into three categories namely: template matching, statistical, and structural methods [7]. Approaches of the first category measure the similarity between the action to be recognized and a stored template while taking into account all allowable distortions. Several works followed this trend among which we can find the Spatio Temporal Feature Chain (STFC) [8] and the Linear Dynamical Systems (LDS) [9]. In the statistical approaches, each action is represented in terms of d features or measurements and is viewed as a point in a d-dimensional space. One such method is the Skeletal Quad proposed by Evangelidis et al. [10] which locally encodes the relation of joint quadruples leading to a vector of 1560 dimensions per level. Last, in the structural recognition approaches an action is decomposed in primitive elements and then a formal analogy is drawn between the structure of that action and the syntax of a language. For instance the SMIJ representation of Ofli et al. [11] first determines the subsets of joints activated when performing an action and then expresses the whole action in terms of the mined joints.

As for other pattern recognition domains, this categorisation is also found for handwriting approaches. We particularly follow in this paper a statistical approach. However the transposal from handwriting to action recognition can similarly be explored by following approaches of the other categories.

B. Handwritten symbol encoding: HBF49

Handwriting recognition is a widely established research field. It attracted great attention since the early 1950s which resulted in a great variety of approaches. Globally, those approaches either dealt with single stroke or multi-stroke handwritten symbols. A stroke is the writing from pen down to pen up. The HBF49 features proposed by Delaye and Anquetil [3] are aimed to be a generic representation of multistroke symbols, without consideration of drawing constraints or domain specificities. HBF49 is a very compact representation that includes only 49 features which belong to different families so as to exhaustively cover the aspects of handwritten symbols encountered in the literature. They can be roughly grouped into two categories: dynamic and visual features. Dynamic features were used to model the writing process, while the visual ones focused on the appearance of the writing result. After extensive evaluations on many (eight) benchmarking datasets, authors affirmed that the HBF49 feature-set is able to deal with several application contexts and is robust with respect to patterns of diverse nature. For more details about the features and the statistical validation of the authors we refer the reader to [3].

C. Morphology-independent preprocessing

Data variations induced by different morphologies is one of the specific problem found in action recognition and not in handwriting area. To increase the robustness of our representation to such variations, we adopt the morphology-independent preprocessing of Kulpa et al. [4]. Globally, authors proposed to only consider twelve main joints to compute four trajectories corresponding to each body part, namely LeftArm, RightArm, LeftLeg and RightLeg. Next, these trajectories are normalized by the total-length of the related body part in order to make them independent from the subject's morphology which results in four morphology independent trajectories (Figure 1-b). The twelve joints to be considered are shoulders, elbows, wrists, hips, knees and ankles (Figure 1-a).

After this preprocessing, we further noted that an action and a handwritten symbol are structured the same. In fact a handwritten symbol is composed of strokes (or segments) which are the 2D trajectories drawn to build the desired symbol. The processed action can then also be considered as a 3D multistroke pattern by assuming that each of the four processed trajectories is a 3D stroke (Figure 1-c) . This similarity is adopted in this paper and the four trajectories previously obtained are assembled to get one single pattern.

III. Handwriting-Inspired Human Action Representation

In the following, we present in detail our proposed HIF3D representation. A first section introduces a series of useful notations. Next we present the first subset of features, named extended features, since they are a 3D adaptation of the descriptors contained in HBF49. Then we formulate the newly features which are specific to the 3D human action but still encode the same information as the one carried by equivalent features in handwriting area. Last we explain the temporal partitioning construction that we adopted to extract the HIF3D features on different time windows.

A. Notations

- A pattern S is a sequence of 3D points resulting from the assembling of four morphology-independent trajectories. $S = \{s_1, ..., s_T, s_{T+1}, ..., s_{2T}, s_{2T+1}, ..., s_{3T}, s_{3T+1}, ..., s_n\}$, where T is the length of each single trajectory (or stroke) and $n = 4 \times T$ is the number of points in *S*. Each point $s_i = (x_i, y_i, z_i)$ is located in the three-dimensional space,
- $||.||$ denotes the Euclidean distance between points,
- $L = L_{1,n}$ is the total length of *S*,
- s_m is the middle-path point,
- x_{max} is the abscissa of the rightmost point of *S*, x_{min}, y_{max}, y_{min}, z_{max} and z_{min} are the other extreme coordinates,
- B is the bounding box of *S*, it is the cuboid parallel to the axis, defined by x_{min}, x_{max}, y_{min}, y_{max}, z_{min}, z_{max},
- $\mathbf{w} = x_{max} - x_{min}$ is the width of B, $\mathbf{h} = y_{max} - y_{min}$ is the height of B, and $\mathbf{d} = z_{max} - z_{min}$ is the depth of B, (if $\mathbf{w}$, $\mathbf{h}$ or $\mathbf{d}$ are null, their value is set to 1),
- $l = max(\mathbf{w}, \mathbf{h}, \mathbf{d})$,
- c_x, c_y and c_z are the coordinates of the center of B,
- $\mu(\mu_x, \mu_y, \mu_z) = (1/n) \sum_{i=1}^{n} s_i$ is the center of gravity of the pattern *S*.

B. Set 1: Extended features

They represent the features which can directly be extended from 2D trajectory to 3D one. We retain in **HIF3D** a total of 41 extended features which are described as follows.

Starting and ending points: The first three features are

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

Similarly we obtain $\mathbf{f_4}$, $\mathbf{f_5}$ and $\mathbf{f_6}$ by replacing in formula (1) x_1, y_1 and z_1 with the ending point coordinates x_n, y_n and z_n respectively.

First point to last point vector: In the 3D case we compute the length of the vector $\vec{v} = \overrightarrow{s_1 s_n}$ and its directional cosines.

$$\mathbf{f_7} = ||\vec{v}||, \quad \mathbf{f_8} = \frac{\vec{v}.\vec{u_x}}{||\vec{v}||}, \quad \mathbf{f_9} = \frac{\vec{v}.\vec{u_y}}{||\vec{v}||}, \quad \mathbf{f_{10}} = \frac{\vec{v}.\vec{u_z}}{||\vec{v}||} \tag{2}$$

Closure: It permits to highlight differences between closed and elongated patterns. It is defined as:

$$\mathbf{f_{11}} = \frac{||\vec{v}||}{L} \tag{3}$$

Angle of initial vector: The initial vector relates the first and the third points: $\vec{w} = \overrightarrow{s_1 s_3}$. We retained the directional cosines:

$$\mathbf{f_{12}} = \frac{\vec{w}.\vec{u_x}}{||\vec{w}||}, \quad \mathbf{f_{13}} = \frac{\vec{w}.\vec{u_y}}{||\vec{w}||}, \quad \mathbf{f_{14}} = \frac{\vec{w}.\vec{u_z}}{||\vec{w}||} \tag{4}$$

Inflections: They relate the positioning of the middle-path point s_m to that of the middle point of segment $s_1 s_n$

$$\mathbf{f_{15}} = \frac{1}{\mathbf{w}}(x_m - \frac{x_1 + x_n}{2}), \quad \mathbf{f_{16}} = \frac{1}{\mathbf{h}}(y_m - \frac{y_1 + y_n}{2}), \quad \mathbf{f_{17}} = \frac{1}{\mathbf{d}}(z_m - \frac{z_1 + y_n}{2}) \tag{5}$$

Proportion of downstrokes trajectory: In handwriting recognition downstrokes are the portions of drawing trajectories oriented towards the bottom of the writing surface. Following [3] we extended this concept for 3D trajectories:

$$\mathbf{f_{18}} = \sum_{k=1}^{p_x} LX_k, \quad \mathbf{f_{19}} = \sum_{k=1}^{p_y} LY_k, \quad \mathbf{f_{20}} = \sum_{k=1}^{p_z} LZ_k \tag{6}$$

with p_x, p_y, p_z the number of downstrokes and LX_k, LY_k, LZ_k their length along the X, Y and Z axes.

Bounding box diagonal angles: We measure the three ratios of the box sides:

$$\mathbf{f_{21}} = \arctan\left(\frac{\mathbf{h}}{\mathbf{w}}\right), \quad \mathbf{f_{22}} = \arctan\left(\frac{\mathbf{d}}{\mathbf{h}}\right), \quad \mathbf{f_{23}} = \arctan\left(\frac{\mathbf{w}}{\mathbf{d}}\right) \tag{7}$$

Trajectory length: These features carry an orientation-independent information:

$$\mathbf{f_{24}} = L, \qquad \mathbf{f_{25}} = \frac{\mathbf{w} + \mathbf{h} + \mathbf{d}}{L} \tag{8}$$

Deviation: This is an other orientation-independent feature which evaluates the average distance from points of *S* to the center of gravity μ:

$$\mathbf{f_{26}} = \frac{1}{n} \sum_{i=1}^{n} ||\overrightarrow{s_i \mu}|| \tag{9}$$

Average direction: These features compute a directional information by averaging pairwise directions of segments defined in the trajectory of *S*:

$$\mathbf{f_{27}} = \frac{1}{n-1} \sum_{i=1}^{n-1} \arctan\left(\frac{x_{i+1} - x_i}{z_{i+1} - z_i}\right) \tag{10}$$

The two other features $\mathbf{f_{28}}$ and $\mathbf{f_{29}}$ are computed by substituting in formula (10) the couples (x_i, z_i) with respectively (y_i, x_i) and (z_i, y_i) .

Absolute angle histogram: These features are based on eight angle histograms (h_1-h_8) that accounts for the number of

segments oriented in eight directions. For each segment the orientation is given by:

$$\alpha_i = \arccos\left(\frac{x_{i+1} - x_i}{\sqrt{(x_{i+1} - x_i)^2 + (y_{i+1} - y_i)^2}}\right) \quad (11)$$

The first four features $\mathbf{f_{30}}$ - $\mathbf{f_{33}}$ are computed as the sum of contributions from all angles α_i to opposite directional bins.

$$\mathbf{f_{30}} = \frac{h_1 + h_5}{n}, \quad .. \quad \mathbf{f_{33}} = \frac{h_4 + h_8}{n} \quad (12)$$

We obtain eight other features namely $\mathbf{f_{34}}$ - $\mathbf{f_{37}}$ and $\mathbf{f_{38}}$ - $\mathbf{f_{41}}$ by following the previous procedure and substituting the couples (x_i, y_i) with respectively (y_i, z_i) and (z_i, x_i) in formula (11).

C. Set 2: Newly features

The second subset of features still carry the characteristic information identified for handwritten pattern but have different formulations since the original 2D formulas can not be directly applied for the 3D case.

Curvature and perpendicularity: We denote θ_i the angle defined by consecutive segments within the same stroke:

$$\theta_i = \arccos\left(\frac{\overrightarrow{s_{i-1}s_i}.\overrightarrow{s_is_{i+1}}}{||\overrightarrow{s_{i-1}s_i}||||\overrightarrow{s_is_{i+1}}||}\right) \quad (13)$$

The curvature and perpendicularity are defined as:

$$\mathbf{f_{42}} = \sum_{i=2}^{n-1} \theta_i, \quad \mathbf{f_{43}} = \sum_{i=2}^{n-1} \sin^2(\theta_i) \quad (14)$$

We obtain two other features $\mathbf{f_{44}}$ and $\mathbf{f_{45}}$ by substituting θ_i in formula (14) with ϕ_i which is the angle defined by consecutive planes π within the same stroke (formula 15).

$$\phi_i = \angle(\pi_{i-1,i,i+1}, \pi_{i,i+1,i+2}) \quad (15)$$

k-perpendicularity and k-angle: By introducing a parameter k, the previous angle θ_i is extended to θ_i^k as :

$$\theta_i^k = \arccos\left(\frac{\overrightarrow{s_{i-k}s_i}.\overrightarrow{s_is_{i+k}}}{||\overrightarrow{s_{i-k}s_i}||||\overrightarrow{s_is_{i+k}}||}\right) \quad (16)$$

This determines another measure of local angles:

$$\mathbf{f_{46}} = \sum_{i=k+1}^{n-k} \sin^2(\theta_i^k), \quad \mathbf{f_{47}} = \max_{i=k+1}^{n-k} \theta_i^k \quad (17)$$

And similarly we obtain features $\mathbf{f_{48}}$ and $\mathbf{f_{49}}$ by extending the ϕ_i angle to ϕ_i^k and substitute it in formula (17):

$$\phi_i^k = \angle(\pi_{i-k,i,i+k}, \pi_{i,i+k,i+2k}) \quad (18)$$

For our experiments, we fixed the k to 2 similarly to [3].

Relative angle histogram: First, the relative local angles are computed from smoothing by linear combination θ_i and θ_i^k :

$$\psi_i^k = \gamma\theta_i + (1 - \gamma)\theta_i^k \quad (19)$$

where we retain the same empirical values of $\gamma = 0.25$ and $k = 2$ as fixed by [3]. Next the contributions of ψ_i^k angles are cumulated in four histogram bins uniformly distributed in $[0, \pi]$. Last, four features $\mathbf{f_{50}}$ - $\mathbf{f_{53}}$ are obtained from the histogram divided by n. We identically compute four other features $\mathbf{f_{54}}$ - $\mathbf{f_{57}}$ by considering the angles χ_i^k obtained from ϕ_i and ϕ_i^k:

$$\chi_i^k = \gamma\phi_i + (1 - \gamma)\phi_i^k \quad (20)$$

3D zoning histogram: We define a regular 3D partition of the bounding box B into $3 \times 3 \times 3$ voxels resulting in twenty-seven zoning features. Similar to [3], histograms are built by computing a fuzzy weighted contribution from each point to its eight neighbouring voxels, where the weights are proportional to the distance from the point to the voxels center $c_{j,k,l}$.

$$\mathbf{f_{58}} = \frac{1}{n}\sum_{i=1}^{n} \mu_{111}(s_i), .. \quad \mathbf{f_{84}} = \frac{1}{n}\sum_{i=1}^{n} \mu_{333}(s_i) \quad (21)$$

with $0 \leq \mu_{jkl}(s_i) \leq 1$ is the contribution of point s_i to the voxel with center $c_{j,k,l}$ for each $1 \leq j, k, l \leq 3$

3D moments invariants: Instead of Hu moments used in 2D, we adopted common 3D invariants [12]. To do so we first compute inertia central moments in 3D:

$$m_{pqr} = \sum_{i=1}^{n} (x_i - \mu_x)^p (y_i - \mu_y)^q (z_i - \mu_z)^r \quad (22)$$

The moments are then normalized, for guaranteeing scale independence:

$$\nu_{pqr} = \frac{m_{pqr}}{m_{000}^{\gamma}}, \qquad \gamma = 1 + \frac{p + q + r}{3} \quad (23)$$

The three invariant features are computed as reported in [12]:

$$\begin{aligned}
\mathbf{f_{85}} &= \nu_{200} + \nu_{020} + \nu_{002}, \\
\mathbf{f_{86}} &= \nu_{200}\nu_{020} + \nu_{200}\nu_{002} + \nu_{020}\nu_{002} - \nu_{110}^2 - \nu_{101}^2 - \nu_{011}^2, \\
\mathbf{f_{87}} &= \nu_{200}\nu_{020}\nu_{002} + 2\nu_{110}\nu_{101}\nu_{011} \\
&\quad - \nu_{002}\nu_{110}^2 - \nu_{020}\nu_{101}^2 - \nu_{200}\nu_{011}^2
\end{aligned} \quad (24)$$

Convex Hull features: The last two features capture the 3D shape of the resulting pattern by considering its convex hull. The convex hull H of S is first computed by means of the quickhull algorithm [13] and then we deduce its volume V_H. The two related features are the convex hull volume normalized by the bounding box volume, and the compactness:

$$\mathbf{f_{88}} = \frac{V_H}{\mathbf{w} * \mathbf{h} * \mathbf{d}}, \qquad \mathbf{f_{89}} = \frac{L^3}{V_H} \quad (25)$$

D. Temporal partitioning construction

As introduced previously, the proposed HIF3D features do not capture the temporal dependence inside an action sequence. Therefore, and similar to spatial partitioning in scene recognition [14], we extract the HIF3Ds features according to a multilevel split of the sequence. The top level HIF3D is computed over the entire sequence. The lower levels are computed over smaller overlapping windows of the entire sequence. The final representation is the concatenation of the computed HIF3Ds over all considered levels. Two levels are used in this paper as shown in Figure 2. Similar temporal construction is commonly adopted for action recognition as in the works of [10, 15].

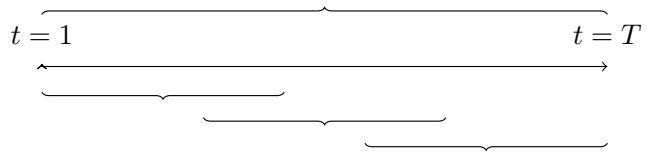

Fig. 2. Illustration of the Temporal partitioning construction adopted in our representation. The features extraction at the l^{th} level covers $\frac{T}{3^l}$ frames of the sequence, where T is the length of the entire sequence.

IV. EXPERIMENTATION

In this section we compare the proposed HIF3D features to state-of-the-art skeletal representations on two publicly available datasets including HDM05 [5] and UTKinect [6].

Since the aim is to measure the effectiveness of the proposed representation, we deliberately retained a popular algorithm for the classification step, namely Support Vector Machines (SVM). We use for all the experiments a temporal partitioning of two levels (Level = 2). In the following we first consider the HDM05 and then the UTKinect datasets.

A. Mocap Database HDM05

HDM05 is an optical marker-based dataset [5] which 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.

Several studies have already been conducted on the HDM05 dataset. For our evaluation we adopt the experimental setup of [11] that suggests a set of 11 actions. The actions are performed by 5 subjects, while each subject performs each action a couple of times ; this suggests a set of 249 sequences. As with [11], we use a cross-subject splitting with 3 (the actors bd, mm and tr) and 2 subjects (the actors bk and dg) in training and testing sets respectively, thus having 139 training and 110 testing examples at our disposal. Table I reports

TABLE I
COMPARISON OF THE HIF3D PERFORMANCE WITH THE STATE-OF-THE-ART RESULTS ON **HDM05 DATASET**.

Method & Year	Recognition rate (%)
SMIJ + SVM, 2014 [11]	84.47
MIJA/MIRM + LCSS, 2015 [16]	85.23
LDS + SVM, 2013 [9]	91.74
Skeletal Quads + SVM, 2014 [10]	93.89
Cov3DJ + SVM, 2013 [15]	95.41
BIPOD + SVM, 2015 [17]	96.70
HIF3D + SVM + Level = 2	**98.17**

the quantitative results of the proposed HIF3D representation over the HDM05 dataset. Our approach achieves an average accuracy of 98.17% with a temporal hierarchy of two levels only, bringing the representation length to a total of $4 * 89 = 356$ features. Furthermore, Table I shows that the proposed approach outperforms existing skeleton-based representations and obtains a state-of-the-art score over this dataset. Besides, our representation is far simpler than all previous approaches since firstly it does not require all joints data and mostly it is size reduced. While one of the best result previously reported [15] was achieved by means of a three level hierarchy with 1830 features for each time partition, our representation is composed instead of two levels with 89 features per partition. Therefore, the effectiveness of our approach is fostered by its simplicity compared to other propositions.

B. UTKinect-Action Dataset

To validate the applicability of our HIF3D representation on marker-free dataset, we conduct another set of experiments on the UTKinect-Action which was captured using a stationary Kinect sensor [6]. It consists of 10 actions performed by 10 different subjects. Each subject performed all actions twice. Altogether, there are 199 action sequences. The 3D locations of 20 joints are provided with the dataset. This is a challenging dataset due to high intra-class variations.

On this second dataset we first evaluated our representation according to the Leave-One-Sequence-Out (LOSeqO) protocol proposed by [6]. It consists in testing one single sequence while the other sequences are used for learning. The results of the experiment are presented in Table II. According to this protocol we attain an average recognition accuracy of 94%. Table II also shows that our approach improves over the current state-of-the art. This second experiment confirms the effectiveness of the proposed HIF3D representation when operating with marker-free capture systems.

TABLE II
COMPARISON OF THE HIF3D PERFORMANCE WITH PREVIOUS APPROACHES ON **UTKINECT DATASET** ACCORDING TO THE LEAVE-ONE-SEQUENCE-OUT PROTOCOL.

Method & Year	Recognition rate (%)
LTI + HMM, 2014 [18]	86.76
Grassmann + SVM, 2015 [19]	88.5
HOJ3D + HMM, 2012 [6]	90.95
STFC + SVM, 2015 [8]	91.5
HIF3D + SVM + Level = 2	**94**

We carried a further evaluation on this dataset according to the cross-subjects scheme where we used all possible combinations of five subjects out of ten as different sets of training and unseen test datasets respectively (C_{10}^5 = 252 rounds). Results reported in Table III show that we reach an average recognition accuracy of 90.96%. The proposed handwriting-inspired representation is accordingly able to efficiently recognize actions in real conditions where tested sequences belong to different subjects than those used for training, hence the great inter-subject discrimination power of the features. Moreover, the obtained score outperforms state-of-the-art results and thus testify the soundness of the inspiration from previous handwriting work to represent skeleton-based actions.

V. CONCLUSION

We introduced in this paper a novel skeleton-based representation of 3D human action. Motivated by the recent advances

TABLE III
COMPARISON OF THE HIF3D PERFORMANCE WITH PREVIOUS APPROACHES ON **UTKINECT DATASET**, ACCORDING TO A CROSS-SUBJECT VALIDATION PROTOCOL.

Method & Year	**Recognition rate (%)**
STFC + SVM, 2015 [8]	85
Fusing features + Random Forests, 2013 [20]	87.90
HIF3D + SVM + Level = 2	**90.96**

in handwriting recognition and the several similarities shared with human action recognition, the proposed Handwriting-Inspired Features (HIF3D) aim to extract from skeleton trajectories the same characteristic information as the one mined from handwritten patterns. To that end, first we preprocess twelve joint positions among the skeleton input data in order to get a morphology-independent 3D multistroke symbol. After that, we extract the HIF3D features from the obtained 3D pattern according to a two-level time partitioning of the actions in order to encode both their spatial and temporal aspects. Last we build a multi-class classifier based on linear SVMs. Experiments conducted on two challenging benchmarks have shown very promising results. In particular we outperform state-of-the-art approaches according to different testing schemes while operating with a much more compact set of features. This provides strong evidence in favour of the HIF3D soundness. The future work will focus on extending this representation for early gestures recognition by considering handwriting temporal segmentation methods.

REFERENCES

[1] G. Johansson, "Visual perception of biological motion and a model for its analysis," *Attention, Perception, & Psychophysics*, vol. 14, no. 2, pp. 201–211, 1973.

[2] J. Shotton, T. Sharp, A. Kipman, A. Fitzgibbon, M. Finocchio, A. Blake, M. Cook, and R. Moore, "Real-time human pose recognition in parts from single depth images," *Communications of the ACM*, vol. 56, no. 1, pp. 116–124, 2013.

[3] 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.

[4] R. Kulpa, F. Multon, and B. Arnaldi, "Morphology-independent representation of motions for interactive human-like animation," in *Computer Graphics Forum*, vol. 24, no. 3. Wiley Online Library, 2005, pp. 343–351.

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

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

[7] A. K. Jain, R. P. Duin, and J. Mao, "Statistical pattern recognition: A review," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 22, no. 1, pp. 4–37, 2000.

[8] 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.

[9] 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*, 2013, pp. 471–478.

[10] G. Evangelidis, G. Singh, and R. Horaud, "Skeletal quads: Human action recognition using joint quadruples," in *ICPR 2014-International Conference on Pattern Recognition*, 2014.

[11] 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.

[12] F. A. Sadjadi and E. L. Hall, "Three-dimensional moment invariants," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, no. 2, pp. 127–136, 1980.

[13] C. B. Barber, D. P. Dobkin, and H. Huhdanpaa, "The quickhull algorithm for convex hulls," *ACM Transactions on Mathematical Software (TOMS)*, vol. 22, no. 4, pp. 469–483, 1996.

[14] S. Lazebnik, C. Schmid, and J. Ponce, "Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories," in *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, vol. 2. IEEE, 2006, pp. 2169–2178.

[15] 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 *IJCAI*, vol. 13, 2013, pp. 2466–2472.

[16] 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.

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

[18] L. L. Presti, M. La Cascia, S. Sclaroff, and O. Camps, "Gesture modeling by hanklet-based hidden markov model," in *Computer Vision–ACCV 2014*. Springer, 2014, pp. 529–546.

[19] 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.

[20] Y. Zhu, W. Chen, and G. Guo, "Fusing spatiotemporal features and joints for 3d action recognition," in *Proceedings of the IEEE Conf. on CVPR-W*, 2013, pp. 486–491.