Extending the Sigma-Lognormal Model of the Kinematic Theory to Three Dimensions

Roman Schindler*, Manuel Bouillon*, Réjean Plamondon† and Andreas Fischer*‡
*DIUF department, DIVA group, *University of Fribourg*, Fribourg, Switzerland
†Laboratoire Scribens, *École Polytechnique de Montréal*, Montréal, Québec H3T 1J4, Canada
‡iCoSys Institute, *University of Applied Sciences Western Switzerland*, Fribourg, Switzerland
Email: {*roman.schindler,manuel.bouillon,andreas.fischer*}*@unifr.ch, rejean.plamondon@polymtl.ca*

***Abstract*—The Kinematic Theory of rapid human movements and its Sigma-Lognormal model enables to model human gestures, in particular complex handwriting patterns such as words, signatures and free gestures. This paper investigates the extension of the theory and its Sigma-Lognormal model from two dimensions to three, taking into account new acquisition modalities (motion capture), multiple subjects, and unconstrained motions. Despite the increased complexity and the new acquisition modalities, we demonstrate that the Sigma-Lognormal model can be successfully generalized to describe 3D human movements. Starting from the 2D model, we replace circular with spherical motions to derive a representation of unconstrained human movements with a new 3D Sigma-Lognormal model. First experiments show a high reconstruction quality with an average signal-to-noise ratio (SNR) of 18.52 dB on the HDM05 dataset. Gesture recognition using dynamic time warping (DTW) achieves similar recognition accuracies when using original and reconstructed gestures, which confirms the high quality of the proposed model.**

Keywords*—*Kinematic Theory of rapid human movements, Sigma-Lognormal model, trajectory reconstruction, 3D motion recognition, dynamic time warping

I. Introduction

The Kinematic Theory of rapid human movements is used to analyze human movements as a process depending on the neuromuscular parameters of the human body. Applied to handwriting, it can be used to study writer expertise [1], [2], to verify genuine signatures [3], [4], or to synthesize artificial handwriting [5], [6], to name just a few applications.

In this paper, we investigate the extension of the Sigma-Lognormal model to three dimensions to analyze 3D human movements. This modeling of 3D motions with the Kinematic Theory of rapid human movements could potentially lead to a wide new range of applications including, for example, neuromuscular disorder analysis, movement synthesis for robotics and computer games, and person identification by gait analysis.

We pursue a natural extension of the model by replacing circular with spherical motions for individual lognormal strokes (which are hypothesized to act along a pivot). While the estimation of the lognormal parameters from the velocity profile remains the same as in 2D, we integrate polar angles in addition to azimuthal angles and adapt the estimation of the angular parameters, accordingly.

The resulting 3D Sigma-Lognormal model is empirically tested on the HDM05 motion capture database, which contains different motion classes, such as walking, dancing, kicking, etc. We reconstruct the movements with our proposed model and assess the quality of the reconstruction with a signal-to-noise ratio (SNR). Furthermore, we conduct a classification experiment based on dynamic time warping (DTW), both, with the original and with the reconstructed trajectory, as an additional assessment of the reconstruction quality.

The remainder of this paper is organized as follows. Section II presents the Kinematic Theory of rapid human movements and the 2D Sigma-Lognormal model. Section III introduces the equations of the proposed 3D extension. In Section IV, we present the results of our empirical evaluation. Section V concludes and discusses future work.

II. Kinematic Theory of Rapid Human Movements

This section briefly presents the Kinematic Theory of rapid human movements with the original Sigma-Lognormal model [7] in two dimensions, which we will extend to three dimensions in the next section.

A. The Sigma-Lognormal Model

The kinematic theory postulates that any movement is the combination of movement primitives, so-called strokes, with lognormal speed. Those strokes are initiated at time t_0 in the central nervous system with an intended distance D. They are then actuated with a log time delay μ and a log response time σ. The Kinematic Theory formulates the speed of a stroke as

$$|\vec{v}(t)| = \frac{D}{\sqrt{2\pi}\cdot\sigma(t-t_0)}\exp\left(-\frac{[\ln(t-t_0)-\mu]^2}{2\sigma^2}\right) \quad . \tag{1}$$

Assuming the movement acts along a pivot, the angular position of each stroke is given by:

$$\theta(t) = \theta_s + \frac{\theta_e-\theta_s}{D}\int_0^t |\vec{v}(\tau)|d\tau \quad , \tag{2}$$

where θ_s is the start angle and θ_e is the end angle.

Simplest rapid movements are composed of two strokes, the strokes of the agonist and of the antagonist action needed to execute the movement. Those two strokes are combined according to the Delta-Lognormal model ($\Delta\Lambda$). More complex movements in two dimensions are described as the sum of the different strokes according to the Sigma-Lognormal model

$(\Sigma\Lambda)$ [8]. In the general case, 2D movements (like handwriting) are described as a vector sum of strokes

$$\vec{v}(t) = \sum_{i=1}^{n} \vec{v_i}(t) \quad , \tag{3}$$

where n is the number of strokes.

Accordingly, the velocity and position of the movement in x- and y-direction are

$$v_x(t) = \sum_{i=1}^{n} |\vec{v_i}(t)| \cos(\theta_i(t)) \quad , \tag{4}$$

$$v_y(t) = \sum_{i=1}^{n} |\vec{v_i}(t)| \sin(\theta_i(t)) \quad , \tag{5}$$

$$x(t) = \int_0^t v_x(\tau) d\tau \quad , \tag{6}$$

$$y(t) = \int_0^t v_y(\tau) d\tau \quad . \tag{7}$$

In summary, a 2D movement can be represented as a combination of lognormal strokes $(l_1, \ldots, l_n)$ with six parameters each:

$$l = (t_0, D, \mu, \sigma, \theta_s, \theta_e) \quad . \tag{8}$$

For more details about the Kinematic Theory of rapid human movements, we refer the reader to [9]–[12].

B. Stroke extraction and parameters estimation

In order to represent a movement with the Sigma-Lognormal model, the input is first preprocessed, and then the different strokes are extracted with their respective parameters.

The preprocessing is minimal, but required to enable the correct modeling of the whole movement. The movement is artificially stopped at the beginning and at the end of the motion by artificially holding the respective positions for $200ms$ [3]. Reducing the speed at the border of the trajectory to zero improves the extraction of the first and the last lognormal stroke. If the movements have been acquired with different devices, it is recommended to interpolate the input trajectories at a common sampling rate, i.e. $200Hz$ [3]. If necessary, a Chebyshev filter can also be applied to remove high-frequency components if some noise was introduced by the acquisition device.

From the input trajectory, we get the observed velocity $\vec{v}_o(t)$, and then three steps are applied to extract the different strokes as illustrated in Figure 1. First, the local minima and maxima of the speed profile $|\vec{v}_o(t)|$ are used to detect the biggest stroke l. Second, the parameters $l = (t_0, D, \mu, \sigma, \theta_s, \theta_e)$ of this stroke are estimated based on an initial analytical solution using the Robust XZERO algorithm [13]. These initial solutions are then refined by means of non-linear least squares curve fitting. Third, the extracted stroke is added to the estimated model and its estimated velocity $\vec{v}_e(t)$ is subtracted from the observed velocity $\vec{v}_o(t)$. This three-step process is repeated until the signal-to-noise ratio (SNR) cannot be further improved.

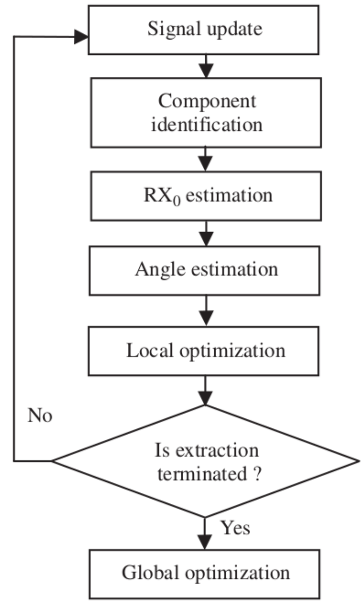

Fig. 1. Workflow of the Sigma Lognormal parameters estimation process. Illustration from O'Reilly [3]

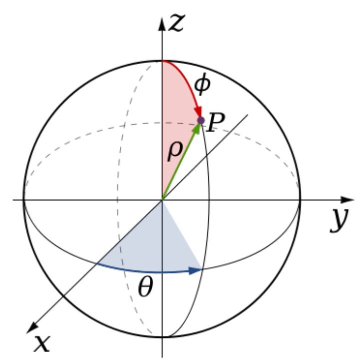

Fig. 2. Illustration of the movement 3D model with $\phi(t)$ in addition to the 2D representation with $\theta(t)$ and $\rho = |\vec{v}(t)|$.

C. Model quality assessment

The quality of the reconstructed movement can be used to assess the quality of the model by means of a signal-to-noise ratio (SNR) between the observed movement $\vec{v}_o(t)$ and the reconstructed movement $\vec{v}_r(t)$

$$SNR = 10 \cdot \log \left(\frac{\int_{t_s}^{t_e} |\vec{v}_o(\tau)|^2 d\tau}{\int_{t_s}^{t_e} |\vec{v}_o(\tau) - \vec{v}_r(\tau)|^2 d\tau} \right) \quad , \tag{9}$$

where t_s is the starting time and t_e the ending time of the movement.

III. Extension to three dimensions

The Kinematic Theory of rapid human movements assumes that strokes act along a pivot. The strokes can be described in the two dimensional plane with the distance to the origin ρ and one angle θ. In order to model human motions in three dimensions, an additional angle ϕ is required as shown in Figure 2.

When adding a third dimension, the velocity value from Equation 1 does not change, but the angular position of Equation 2 now depends on a second angle ϕ

$$\phi(t) = \phi_s + \frac{\phi_e - \phi_s}{D} \int_0^t |\vec{v}(\tau)| d\tau \quad . \tag{10}$$

The velocity Equations 4 and 5 become

$$v_x(t) = \sum_{i=1}^{M} |\vec{v}_i(t)| \sin(\phi_i(t)) \cos(\theta_i(t)) \quad , \tag{11}$$

$$v_y(t) = \sum_{i=1}^{M} |\vec{v}_i(t)| \sin(\phi_i(t)) \sin(\theta_i(t)) \quad , \tag{12}$$

and the velocity in the z-direction is

$$v_z(t) = \sum_{i=1}^{M} |\vec{v}_i(t)| \cos(\phi_i(t)) \quad . \tag{13}$$

The position of the movement in x- and y-direction in Equations 6 and 7 do not change, and the position in z-direction is

$$z(t) = \int_0^t v_z(\tau)\, \mathrm{d}\tau \quad . \tag{14}$$

In summary, a 3D movement can be represented as a combination of lognormal strokes $(l_1, \ldots, l_n)$ with eight parameters each

$$l = (t_0, D, \mu, \sigma, \theta_s, \theta_e, \phi_s, \phi_e) \quad . \tag{15}$$

The estimation process of Figure 1 does not change, only the angle estimation step is extended to estimate the two new parameters. The new angle parameters ϕ_s and ϕ_e can be estimated in a similar way as the original parameters θ_s and θ_e with

$$\phi\_n(t) = \arccos \frac{v_{o,z}(t)}{v_o(t)} \quad , \tag{16}$$

$$\Delta\phi = \frac{\phi\_n(t_{4\_n}) - \phi\_n(t_{2\_n})}{l(t_4) - l(t_2)} \quad , \tag{17}$$

$$\phi_s = \phi\_n(t_{3\_n}) - \Delta\phi(l(t_3) - l(t_1)) \quad , \tag{18}$$

$$\phi_e = \phi\_n(t_{3\_n}) - \Delta\phi(l(t_5) - l(t_3)) \quad , \tag{19}$$

where t_i are the times of the points p_i as follows:

- p_1 Lognormal stroke beginning
- p_2 First inflexion point
- p_3 Local maximum of the lognormal stroke
- p_4 Second inflexion point
- p_5 Lognormal stroke ending

IV. Experiments and Results

In this section, we present the results of our empirical evaluation conducted with the proposed 3D extension of the Sigma-Lognormal model. First, we measure the signal-to-noise ratio (SNR) on the HDM05 dataset to assess the model quality of the 3D gestures. Second, we investigate the impact of using synthetic gestures for a dynamic time warping (DTW) classifier to explore the potential for applications in the field of 3D action recognition.

TABLE I. Comparison of the signal-to-noise ratio (SNR) mean (μ) and standard deviation (σ) between 2D and 3D data.

	μ (dB)	σ (dB)
2D projection in (x,y) plane	19.19	3.77
3D motions	18.52	4.09

A. Dataset

For evaluating the 3D Sigma-Lognormal model, we use the HDM05 dataset [14]. HDM05 is a motion capture (*mocap*) dataset that contains the trajectories of various points on the skeleton that was recorded with a suit containing more than 40 markers. The input data are the 3D trajectories of the suit markers recorded at 120Hz, which allows to evaluate the 3D Sigma-Lognormal model with a high precision. The dataset is composed of roughly 100 classes that were performed 10 to 50 times by 5 subjects, amounting to 1,457 samples in total.

From the skeleton data, we compute the trajectories of wrists and ankles relative to shoulders and hips, respectively, which yield four limb trajectories [15]. Those limb vectors are then normalized by the limb length to get a motion representation that is independent of the morphology of the subject [16]. This allows to get a simplified representation of the skeleton, that is independent of the position of the skeleton but that preserves the main characteristics of the movements. We also use 11 selected actions [17] for our experiments (deposit floor, elbow to knee, grab high, hop both legs, jog, kick forward, lie down floor, rotate both arms backward, sneak, squat, throw basketball), resulting in 249 samples.

Each limb trajectory is represented as a sequence of (x, y, z) coordinates, from which we can extract the sequences of velocities (v_x, v_y, v_z) and accelerations (a_x, a_y, a_z), which gives a total of 36 features. All those features are computed with second order regression [18], and they are normalized by a z-score normalization over all sampling points [4].

B. Trajectory Reconstruction Quality

To evaluate the quality of the representation of the 3D motions with the Sigma-Lognormal model, we measure the signal-to-noise ratio (SNR) between the original and reconstructed trajectories [7]. To reconstruct trajectories, we extract the Sigma-Lognormal model parameters from the 36 features of the input samples, and then we try to reconstruct the motions (with the 36 features) from the Sigma-Lognormal model parameters.

Figure 3 shows the original (in blue) and reconstructed (in red) trajectories of the wrists and ankles of a kick motion sample. Visually, the original and reconstructed trajectories look very much alike. Figure 4 shows the original (in blue) and reconstructed (in red) velocity profiles of the four limbs of the same kick motion sample. Again, the reconstruction is very close to the original.

In order to quantify the quality of the reconstruction, we measure the SNR between the original and reconstructed trajectories [7]. Figure 5 presents the SNR distribution for all the four reconstructed trajectories of the 249 samples (hence 996 reconstructions). The mean SNR is $18.52dB$ (with a standard deviation of $4.09dB$) which suggests a high quality

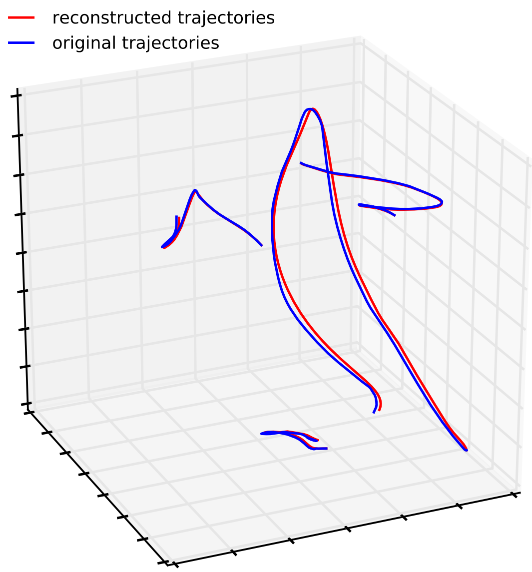

Fig. 3. Original trajectories (blue) and reconstructed (red) trajectories of wrists and ankles in a kick motion sample.

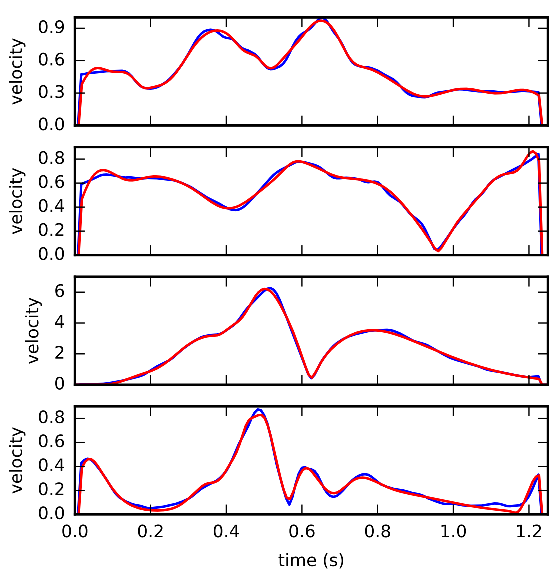

Fig. 4. Original (blue) and reconstructed (red) velocity profiles of right wrist, left wrist, right ankle and left ankle of the kick motion of Figure 3.

of the 3D Sigma Lognormal modeling. The general quality of the 3D reconstructions is as good as the quality of the 2D reconstructions, which we can obtain by discarding the $z-$dimension, as shown in Table I.

C. Motion Recognition Results

In order to further investigate the quality of the proposed 3D Sigma-Lognormal model, we use reconstructed movements to perform action recognition on the HDM05 dataset. We use dynamic time warping (DTW) to compute a distance between two movements. To avoid unusual warping paths and to speed up the computation, a Sakoe-Chiba band [22] is employed.

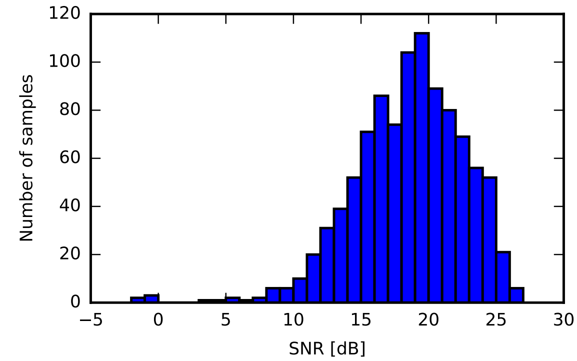

Fig. 5. Distribution of signal-to-noise ratio (SNR) values of the reconstructed 3D trajectories of the HDM05 dataset (996 samples).

TABLE II. COMPARISON OF THE DIFFERENT COMBINATIONS OF INPUT DATA (FOUR SUBJECTS AS REFERENCES AND ONE SUBJECT FOR TESTING).

DTW input data	Accuracy (%)	SD (%)
Position	93.61	4.14
Velocity	**96.44**	4.54
Acceleration	91.51	5.56
Position + Velocity	95.35	3.39
Velocity + Acceleration	95.76	3.33
Position + Acceleration	95.43	2.81
Position + Velocity + Acceleration	95.34	4.17

The classification of a test sample is done by computing the DTW distance to a set of reference samples and predicting the class of the nearest reference sample.

First, we study the importance of the position, velocity and acceleration profiles. Table II presents the recognition accuracy obtained when using the original movements of four subjects as reference samples and those of the remaining subject as test samples. The average accuracy and the standard deviation (SD) over the five experiments are indicated. The best results are obtained when using only the velocity profile. In fact, adding either the position or acceleration (or both) yield lower recognition accuracies. Since the velocity profile performed best, we proceeded using only 12 velocity features (3 dimensions and 4 limbs).

Next, we compare our classification results with other state-of-the-art methods. Table III shows the results of the proposed velocity-based DTW classifier on the same train/test split that is typically used in the literature [17]. The results indicate that our classifier is able to reach state-of-the-art performance.

Finally, we compare the recognition accuracy obtained when using reconstructed movements as reference samples,

TABLE III. COMPARISON OF RECOGNITION ACCURACY ON THE HDM05 BENCHMARK (USING THE PROPOSED TRAIN-TEST SPLIT [17]).

	Accuracy (%)
Cov3DJ + SVM [19]	95.41
HOD + SVM [20]	97.27
BIPOD + SVM [21]	96.70
HIF3D + SVM + Level = 2 [15]	98.17
Our approach (DTW + velocity profile)	97.25

TABLE IV. COMPARISON OF THE RECOGNITION ACCURACY (%) FOR ORIGINAL OR RECONSTRUCTED REFERENCE SAMPLES.

	Number of training subjects			
	1	2	3	4
Originals	72.93	78.70	86.53	96.44
Reconstructions	72.84	78.57	86.56	96.06

while still testing on the original test samples. We use five-fold cross-subject splitting and vary the number of subjects whose movements are used as reference sequences. The samples of all remaining subjects are used for testing. Each setting is run five times, accordingly, and the results are averaged. Table IV shows the results obtained when using original movements as reference samples or when using reconstructed movements (always using original samples for testing). The accuracy obtained with reconstructed samples is very similar to the one obtained with the original samples and a paired t-test shows no significant difference between the two sets of results ($p > 0.05$, *n.s.*). This result confirms the high model quality of the proposed 3D Sigma-Lognormal model.

V. CONCLUSION

In this paper, we presented a 3D extension of the Sigma-Lognormal model to represent unconstrained 3D human movements. First experiments show a good model quality, as demonstrated by a high signal-to-noise ratio (SNR) of reconstructed motions and the fact that similar classification results have been achieved with original and reconstructed reference samples.

The best classification results were achieved using only the velocity profile. This observation is consistent with the underlying hypothesis of the Kinematic Theory that the velocity is the main control variable used by the central nervous system to plan and execute a movement. It is also consistent with the basic property of the velocity vector, which is tangent to the trajectory, allowing learning by observing [23].

A promising line of future research is the synthesis of 3D movements, which is expected to support classification systems with artificial but realistic training data. In general, there is a wide range of applications that could emerge from modeling 3D motions with the Kinematic Theory in biometrics, biomedicine, and robotics.

ACKNOWLEDGMENTS

This work has been supported by the NSERC-Canada Grant RGPIN-2015-06409 to R. Plamondon, by the European Union H2020 program with the grant number 731861, and by the National Natural Science Foundation of China with the grant number 61650110512.

REFERENCES

[1] R. Plamondon, C. O'Reilly, C. Rémi, and T. Duval, "The lognormal handwriter: learning, performing, and declining," *Frontiers in Psychology*, vol. 4, Dec. 2013.

[2] C. O'Reilly and R. Plamondon, "Design of a neuromuscular disorders diagnostic system using human movement analysis," in *Information Science, Signal Processing and their Applications (ISSPA), 2012 11th International Conference on*, pp. 787–792, IEEE, 2012.

[3] C. O'Reilly and R. Plamondon, "Development of a Sigma-Lognormal Representation for On-line Signatures," *Pattern Recogn.*, vol. 42, pp. 3324–3337, Dec. 2009.

[4] A. Fischer and R. Plamondon, "Signature verification based on the kinematic theory of rapid human movements," *IEEE Transactions on Human-Machine Systems*, vol. 47, no. 2, pp. 169–180, 2017.

[5] A. Almaksour, E. Anquetil, R. Plamondon, and C. O'Reilly, "Synthetic Handwritten Gesture Generation Using Sigma-Lognormal Model for Evolving Handwriting Classifiers," 2011.

[6] D. Martín-Albo, R. Plamondon, and E. Vidal, "Training of on-line handwriting text recognizers with synthetic text generated using the kinematic theory of rapid human movements," in *Frontiers in Handwriting Recognition (ICFHR), 2014 14th International Conference on*, pp. 543–548, IEEE, 2014.

[7] C. O'Reilly and R. Plamondon, "Development of a Sigma–Lognormal representation for on-line signatures," *Pattern Recognition*, vol. 42, no. 12, pp. 3324–3337, 2009.

[8] R. Plamondon and M. Djioua, "A multi-level representation paradigm for handwriting stroke generation," *Human movement science*, vol. 25, no. 4, pp. 586–607, 2006.

[9] R. Plamondon, "A kinematic theory of rapid human movements. Part I. Movement representation and generation," *Biological cybernetics*, vol. 72, pp. 295–307, Feb. 1995.

[10] R. Plamondon, "A kinematic theory of rapid human movements. Part II. Movement time and control," *Biological cybernetics*, vol. 72, pp. 309–20, Feb. 1995.

[11] R. Plamondon, "A kinematic theory of rapid human movements: Part III. Kinetic outcomes," *Biological Cybernetics*, vol. 78, no. 2, pp. 133–145, 1998.

[12] R. Plamondon, C. Feng, and A. Woch, "A kinematic theory of rapid human movement. Part IV: a formal mathematical proof and new insights," *Biological Cybernetics*, vol. 89, no. 2, pp. 126–138, 2003.

[13] M. Djioua and R. Plamondon, "A new algorithm and system for the characterization of handwriting strokes with delta-lognormal parameters," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 11, pp. 2060–2072, 2009.

[14] M. Müller, T. Röder, M. Clausen, B. Eberhardt, B. Krüger, and A. Weber, "Documentation Mocap Database HDM05," 2007.

[15] S. Y. Boulahia, E. Anquetil, R. Kulpa, and F. Multon, "HIF3d: Handwriting-Inspired Features for 3d Skeleton-Based Action Recognition," in *23rd International Conference on Pattern Recognition (ICPR 2016)* (IEEE, ed.), (Cancun, Mexico), Dec. 2016.

[16] 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, Wiley Online Library, 2005.

[17] 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.

[18] J. Fierrez, J. Ortega-Garcia, D. Ramos, and J. Gonzalez-Rodriguez, "HMM-based on-line signature verification: Feature extraction and signature modeling," *Pattern recognition letters*, vol. 28, no. 16, pp. 2325–2334, 2007.

[19] 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, pp. 2466–2472, 2013.

[20] 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 *IJCAI*, pp. 1351–1357, 2013.

[21] 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*, pp. 3053–3060, IEEE, 2015.

[22] H. Sakoe, S. Chiba, A. Waibel, and K. F. Lee, "Readings in speech recognition," *San Francisco, CA, USA: Morgan Kaufmann Publishers Inc*, pp. 159–165, 1990.

[23] R. Plamondon and A. Alimi, "Speed/accuracy trade-offs in target-directed movements," *Behavioral and Brain Sciences*, vol. 20, no. 2, pp. 279–349, 1997.