



Approximation of Curvature and Velocity for Gesture Segmentation and Synthesis

Sylvie Gibet, Pierre-François Marteau

► To cite this version:

Sylvie Gibet, Pierre-François Marteau. Approximation of Curvature and Velocity for Gesture Segmentation and Synthesis. (Revised Selected Papers), May 2007, Lisboa, Portugal. pp.13-23, 10.1007/978-3-540-92865-2_2 . hal-00502860

HAL Id: hal-00502860

<https://hal.science/hal-00502860>

Submitted on 16 Jul 2010

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.

Approximation of Curvature and Velocity for Gesture Segmentation and Synthesis

Sylvie Gibet¹ and Pierre-François Marteau²

¹ IRISA, Campus de Beaulieu, F-35042 Rennes Cedex
Sylvie.Gibet@irisa.fr

² VALORIA, Université de Bretagne Sud, Université européenne de Bretagne,
Campus de Tohannic, rue Yves Mainguy,
F-56017 Vannes, France
Pierre-Francois.Marteau@univ-ubs.fr

Abstract. This paper describes a new approach to analyze hand gestures, based on an experimental approximation of the shape and kinematics of compressed arm trajectories. The motivation of such a model is on the one hand the reduction of the gesture data, and on the other hand the possibility to segment gestures into meaningful units, yielding to an analysis tool for gesture coding and synthesis. We show that the measures of the inverse of the distance between adaptive samples and velocity estimated at these points are respectively correlated to the instantaneous curvature and tangential velocity directly computed on motion capture data. Based on these correlation results, we propose a new way to automatically segment hand gestures. We show also that this approach can be applied to a global analysis / synthesis framework, useful for automatic animation of virtual characters performing sign language gestures.

1 Introduction

Numerous approaches have been developed for the representation of human gesture. These studies sensibly differ, whether the emphasis is placed on the search of structural features used for example for gesture transcription or coding, or signal-based features used for example to represent the movement kinematics.

The first class of studies relies most of the time on a thorough knowledge of gestures, leading to a semantic, or a symbolic description using different types of codes. For example, the Laban notation was defined for describing dance or martial art gestures [1]. In the same way, HamNoSys notation was defined to represent sign languages gestures [2]. Other studies propose a segmentation of human gestures into a sequence of discrete phases or motion units of different types, based for example on velocity or acceleration profiles [3, 4, 5].

The second class of studies is related to the analysis of human motion, which has become feasible with the recent development of new technologies for motion capture. The analysis is supported by signal processing or statistical techniques, and has yielded to data-based methods for gesture recognition, motion retrieval, or computer animation. The representation and understanding of motion followed two distinct approaches: the first one can be characterized by the identification of regularities in

motion, expressed in terms of analytical motion laws [6], and the second one is dedicated to the characterization of the variability in motion, due for instance to variations in styles [7]. Among the motion laws, we retain two specific ones which can be used for trajectory segmentation. In particular, the two-third power law, expressing a power relation between velocity and curvature [8] was proposed for segmenting three dimensional unconstrained drawing movements, on the basis of abrupt changes of the velocity gain factor. Another segmentation hypothesis was based on the observation that endpoint trajectories of human arm movements tend to be piecewise planar [9]. These segmentation hypotheses are largely discussed in the neuroscience community.

One major problem in representing gesture from recorded data is that these data are multidimensional. Motion capture data generally consist indeed of sampled trajectories for each degree-of-freedom characterizing the position and orientation of the different joints of the human skeleton. Consequently, direct access and use of motion data is rather time consuming. Moreover, motion data are produced by systems which are redundant, which means that there is an excess of degrees of freedom in the motion representation. In order to efficiently represent gesture, it is therefore useful to compress the original information. Different categories of techniques are classically used for the dimensionality reduction. One category consists in projecting motion into other bases, using for example the Principal Component Analysis [10]. This method extracts the main axes which maximize the variance of the motion. Another category of methods proposes a compression of trajectories by curves and surfaces approximations. Few works concern motion trajectories. Polygonal approximation provides characteristics points to represent the geometry of the trajectory. These points, which correspond to local curvature extrema, can be connected by line segments. This method has been used by [11] for non-uniform sub-sampling of motion time-series. Another method proposes curve approximation using active contours [12]. These methods are developed for dance gesture recognition.

Independently of the data reduction method, we propose in this paper to characterize gesture trajectories expressed in a reduced space by approximated measures of curvature and velocity. Curvature mainly represents the geometry of the trajectories, and velocity represents the kinematics. The data reduction method is presented in [13].

We are mainly interested by structured gestures conveying meaningful information, such as French sign language gestures. These gestures generally use the visuo-gesture channel. In other words, they use the 3D space to sketch specific patterns expressed by Cartesian trajectories. We assume that hand gestures are well characterized by their shape (change of curvature), as well as by their kinematics (change of velocity). In sign language, the signer can indeed draw the shape of the symbol as an icon representing some feature of the object or the activity to be symbolized (Fig. 1). Expressive gestures may also implicitly contain some velocity or acceleration profiles. In particular variations in velocity are responsible for the aggregation of samples in some areas of the trajectories.

We propose here to study both these spatial and kinematics characteristics in the reduced representation space. In order to illustrate the method, we implement it on arm end-point trajectories. Nevertheless, the method can be extended to multidimensional motion trajectories the dimension of which is higher than three. Basing our work on a compressed representation of trajectories [13], we define approximations of adaptive velocity and curvature. We show that these approximations can be strongly

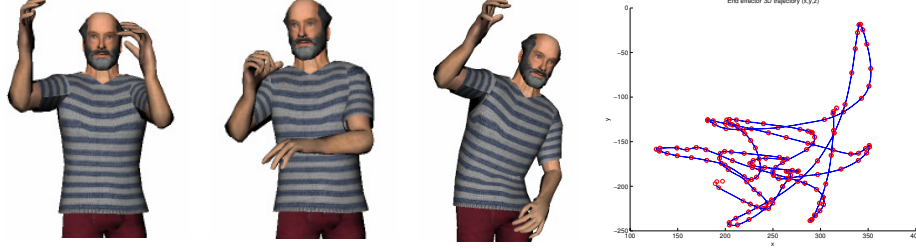


Fig. 1. Three first pictures: sign language gestures using the visuo-gesture channel; right: corresponding 3D end-point trajectory of the right arm with the target location

related to curvature and tangential velocity, not only in 3D space, but in any multidimensional space. These measures provide new tools to automatically analyse gestures, and can be applied to noisy data. An interpretation is given for the segmentation of sign language gestures and its possible use for gesture synthesis.

The paper is mainly composed of four sections. Section 2 gives an overview of the analysis method. Section 3 presents an evaluation of the method applied on 3D arm end-point trajectories, in terms of correlation and compression rate. After illustrating some results about the segmentation of sign language trajectories, section 4 presents some results concerning a global analysis / synthesis framework. The paper concludes and gives some perspectives in section 5.

2 Analysis of Arm Movements

The gestures consist of raw data composed of 3D Cartesian trajectories, each trajectory representing the evolution with time of one coordinate x , y , or z expressing the position of a specific joint. For our study, we consider $X(t)$ as constituted of time-series in $3.p$ dimensions, represented by spatial vectors $X(t) = [x_1(t), y_1(t), z_1(t), x_2(t), y_2(t), z_2(t), \dots, x_p(t), y_p(t), z_p(t)]$. In practice, we deal with the sampled trajectory at a constant frequency of 120 Hz: $X(n)$ where n is the time-stamp index.

In this analysis method, we rely on the method *DPPLA* algorithm, an implementation of a reduced search dynamic programming approach described in [15], which provides a linear piecewise curve approximation, based on an optimization process. This algorithm finds samples in the time series $X(t)$, not regularly located in time.

The method *DPPLA* consists in seeking an approximation $X_{\hat{\theta}}$ of $X(n)$, θ being the set of discrete time location $\{n_i\}$ of the segment endpoints. The selection of the optimal set of parameters $\hat{\theta} = \{\hat{n}_i\}$ is performed using a recursive dynamic programming algorithm. The result of this method is the optimal identification of discrete X_{T_i} key-points – we call them spatial targets – delimitating the segments, for a given compression rate ρ .

The time complexity of the optimal dynamic programming algorithm is $O(k.n^2)$ [14] where n is the number of samples, and k the number of segments. *DPPLA*, relying on the reduced search dynamic programming algorithm, has a $O(n^2/k)$ time complexity. Note that recently a multi-resolution algorithm based on *DPPLA* in $O(n)$ time

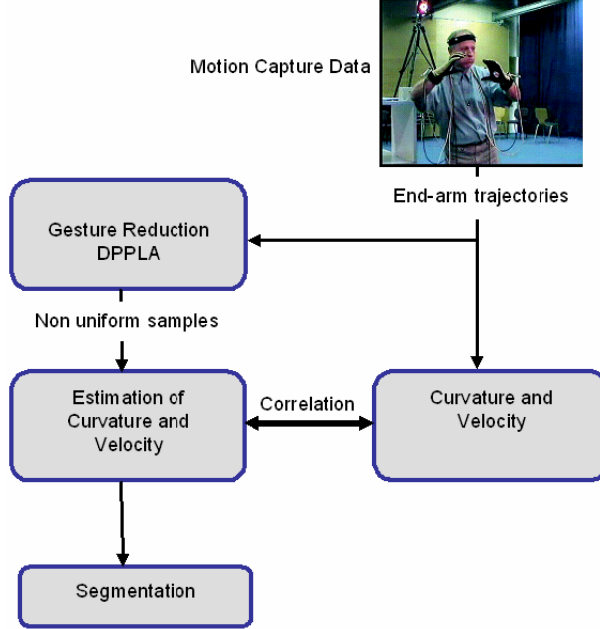


Fig. 2. An approximation of the end-point trajectory using the non uniform sampling algorithm

complexity has been proposed relaxing a bit more the optimality [16, 17]. For detailed description of the method, see [13].

The general approach is illustrated in figure 2. We propose to extract an estimation of the curvature and the velocity from the non uniform samples obtained thanks to the *DPPLA* algorithm. This estimation is evaluated through a correlation measure between the estimated measures and the analytical calculus of both the curvature and the velocity.

In this paper our work is based on 3D end point trajectories $X(t) = [x(t), y(t), z(t)]$, the coordinates being calculated in the shoulder frame. For any smooth trajectory parameterized with t , we express the instantaneous velocity $v(t)$ and the absolute value of the instantaneous curvature $\kappa(t)$:

$$v(t) = \|\dot{X}(t)\| = \sqrt{\dot{x}^2 + \dot{y}^2 + \dot{z}^2} \quad (1)$$

$$\kappa = \frac{\|\dot{X}(t) \times \ddot{X}(t)\|}{\|\dot{X}(t)\|^3} \quad \text{and} \quad R(t) = \frac{1}{|\kappa|} \quad (2)$$

where R is the radius of curvature. The curvature measures how fast a curve is changing direction at a given point.

These variables have been extensively studied for a variety of goal-directed experimental tasks. In particular, a number of regularities have been empirically observed for end-point trajectories of the human upper-limb, during 2D drawing movements.

However, for 3D movements with great spatial and temporal variations, it can be difficult to directly extract significant features from these signals. Moreover, computing the radius of curvature raises a problem, when the velocity is too high, or when there are

inflection points in the trajectories. In particular for noisy data the radius of curvature may be difficult to compute. Finally, for higher dimensions, the curvature is not defined, prohibiting its use in the angular space in particular.

We propose to approximate these velocity and curvature by empirical measures calculated from the adaptive samples identified through the *DPPLA* algorithm.

We define the target-based velocity by the expression:

$$V_{T_{gi}}(n_i) = \frac{\|X(n_{i+1}) - X(n_{i-1})\|}{n_{i+1} - n_{i-1}} \quad (3)$$

where n_{i+1} and n_{i-1} are temporal indices of the associated targets Tg_{i+1} and Tg_{i-1} .

As the targets are not regularly located, the addition effect of this measure, homogeneous to a velocity, is to filter the raw data. The filtering depends on the compression rate.

We define as well the inverse distance between adjacent targets as:

$$\kappa_{T_{gi}}(n_i) = \frac{1}{\|X(n_i) - X(n_{i-1})\|} \quad (4)$$

With this formulation, we assume that this last quantity might be linked to a measure of aggregation points on the trajectory: when the movement velocity decreases, the distance between original samples decreases and the curvature appears to be important. Therefore, $\kappa_{T_{gi}}(n_i)$ expresses a spatial quantity which might be correlated to curvature at time-index n_i .

In the next section, we will study the correlation between the target-based approximations and the instantaneous values. We will also study the influence of the compression parameter k of the compression algorithm.

3 Analysis of 3D Endpoint Arm Data

One deaf signer performed the gestures. He signed sequences of French sign language gestures representing several versions of bulletin weather performed with different styles, relative to the subject's dynamics and emotional state. The sequences were composed of 12 phrases; the whole duration was about 30 s. The subject was asked to perform the gestures with variations of the geometry (large vs. small amplitude), kinematics (high vs. low speed) and dynamics (smooth vs. jerky).

Raw data are first filtered by a low pass Butterworth filter with a cutoff frequency of 10.0 Hz. We consider sequences of about 10000 frames.

The analysis of correlation is achieved, on the one hand between the log of target-based velocity and the log of its instantaneous value, and on the other hand between the inverse of the distance between targets and the instantaneous curvature. The results concerning the velocity are shown in figure 3 (a).

They illustrate an excellent correlation between the two variables, thus allowing us to use target-based velocity as a good approximation of instantaneous velocity. We may also compute the acceleration of arm end-point trajectories on the basis of this target-based velocity. The correlation between the log of the inverse target distances and the log of its instantaneous curvature is also very good, as illustrated in figure 3 (b). The points with abrupt changes are located at the same place, but the target-based signal

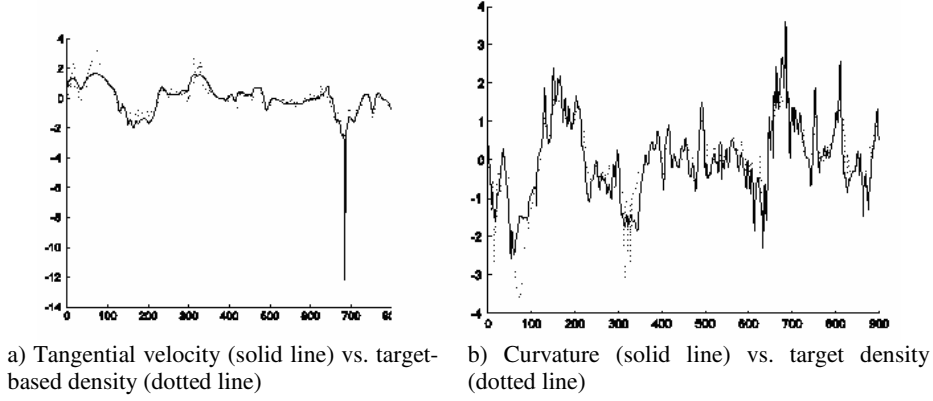


Fig. 3. Correlation for 3D end-point trajectories of arm movements; a) Correlation between instantaneous tangential velocity (solid line) and target-based velocity (dotted line); b) Correlation between instantaneous curvature (solid line) and inverse target density (dotted line); for each signal x , we computed: $(\log(x) - \text{mean}(\log(x)))/\text{std}(\log(x))$.

seems less noisy than the original one. This makes possible to approximate curvature as the inverse of target density.

The influence of the compression factor characterizing the adaptive sampling algorithm is analyzed at the light of the correlation coefficient. The results can be seen in figure 4. It shows that for the target-based velocity, the correlation coefficient remains very close to 1, independently of the compression rate (from 50% to 95%). For the target-based acceleration, the correlation coefficient is very good (0.9), for a compression rate varying until 70%. Beyond this limit, the correlation coefficient abruptly falls. The correlation coefficient is lower for the inverse distance, but still high (.85), even for a high compression rate (until 80%). These results support the assumption that target-based variables can be used without a significant loss of data for the analysis of 3D end-point trajectories.

4 Gesture Segmentation and Data-Driven Synthesis

Studies on gesture [3,4] showed that human gestures can be segmented into distinct phases. Some researches assumed that objective measures can be used to segment hand movement. In particular, Kita et al. showed that abrupt changes of direction, accompanied by a velocity discontinuity indicate phase boundaries in hand trajectories. These observations have been exploited by [5], who proposed a new distance metric to detect phase boundaries, based on the sign of the first and second derivatives of endpoint trajectories. The analysis method described above can be used for automatically segmenting 3D arm motion. Moreover, it can be used for a compact gesture representation and for data-driven synthesis.

4.1 Segmentation

Our segmentation is based on the observation that phase boundaries may appear when the radius of curvature becomes very small, and the velocity decreases at the same

time, indicating a change of direction. Our segmentation algorithm is based on the product variable $v(t) \cdot \kappa(t)$, and on its approximation, based on the approximated target-based variables : $v_{T_{gi}}(n_i) \cdot \kappa_{T_{gi}}(n_i)$.

A color-coding method allows to quantify the variations of the variable, according to an equally distribution of its values. The meaning of this coding is presented in table 1.

Table 1. Coding values for the color coding

coding	Variable values	Interpretation
black	---	lowest values
blue	--	very low values
cyan	-	low values
green	0	average values
yellow	+	high values
magenta	++	very high values
red	+++	highest values

The color-coding is reported on 3D trajectories, as can be seen in figure 4.

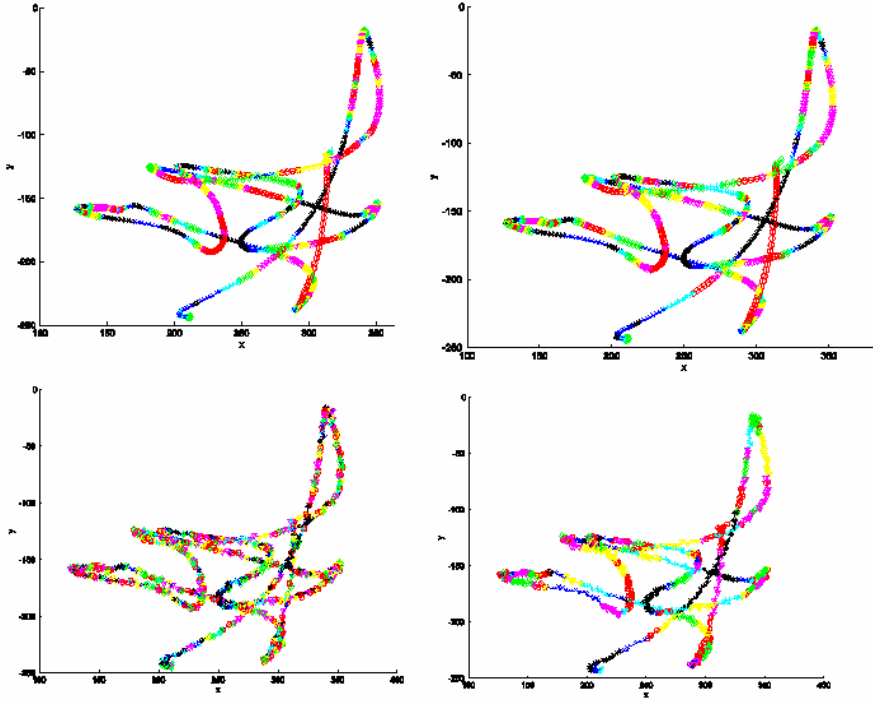


Fig. 4. Example of end-point trajectories segmentation (in the xy plane) using a color-coding of quantified variables (different color or gray levels); up-left: segmentation using the product $\kappa(t) \cdot v(t)$; up-right: segmentation using the product $\kappa_{T_{gi}}(t) \cdot v_{T_{gi}}(t)$; down-left: using the product $\kappa(t) \cdot v(t)$ for noisy data; down-right: segmentation using the product $\kappa_{T_{gi}}(t) \cdot v_{T_{gi}}(t)$ for noisy data.

When the velocity is very low, the color is green (clear gray). In the contrary, when the velocity is high and the curvature low, the color is red (dark gray). The level of quantification indicates the size of the segmental units. A great similarity can be observed between the segmentation of the curve $v(t). \kappa(t)$ and $v_{Tgt}(n_i). \kappa_{Tgt}(n_i)$ (see figure 4 up-left and up-right).

We observe also in figure 4 (down-left and down-right) that the method is well adapted to noisy data, the segmentation using the target-based measures of curvature and velocity being more filtered, while keeping the main motion phases.

4.2 Gesture Analysis / Synthesis Framework

The analysis algorithm described above can be used for representing in a compact way gesture trajectories. These trajectories can be just 3D end-point trajectories, or multidimensional trajectories. In the latter case, the trajectories may be represented by angular postures or Cartesian positions at each joint of the articulated chains.

We define an analysis / synthesis framework which separates the adaptive sampling (DPPLA) and segmentation off-line process and the in-line data-based synthesis process, as illustrated in figure 5.

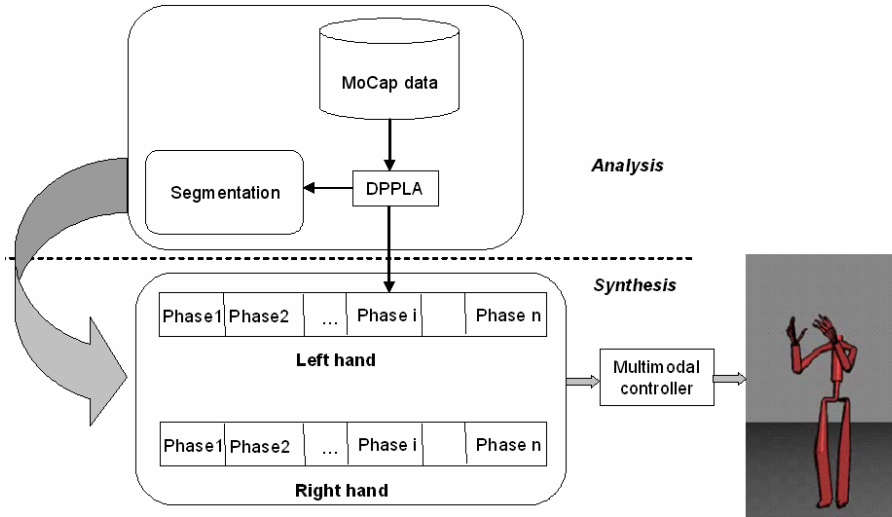


Fig. 5. Analysis / synthesis framework: up) Off-line analysis using *DPPLA* algorithm and the derived automatic segmentation; down) In-line synthesis using the phase description and the adaptive sampling within each phase.

The adaptive sampling (*DPPLA* algorithm) is used for automatic segmentation, as presented in section 4.1. This process can be used jointly with a manual segmentation process to identify phases within the gesture trajectories.

For each phase, the synthesis is achieved by an inverse kinematics process, applied for the arms on adaptive samples of end-point trajectories, and a tracking process for the rest of the body. The inverse kinematics is based on a sensory motor approach

with biological relevance, as described in [18] and [19]. When applied to 3D end-point trajectories (hand motion), the discrete representation which is provided by the *DPPLA* algorithm is directly used to deal with the co-articulation between the motion segments.

The synthesis is achieved for body, hand and arm gestures. This will be soon extended to a complete multimodal system, including the synthesis of the head motion and gaze direction.

Note that the segmentation method can be extended to multidimensional trajectories, representing angular or Cartesian trajectories. It has already been experimented for 6D trajectories, representing wrist and elbow Cartesian trajectories. The results are similar to the ones obtained with 3D trajectories: we are able to identify segments along the sequence, with a varying compression factor. But the necessity to take into account more degrees of freedom still needs to be experimented and evaluated, both for segmentation and synthesis purposes.

5 Conclusion and Future Perspectives

This paper presented a method for computing an approximation of the curvature and velocity characterizing arm trajectories. This method is applied on compressed data, obtained from a non uniform sub-sampling algorithm which automatically extracts discrete target patterns from raw data, for a given compression rate. Given a desired trajectory, we already showed that the targets patterns represented in an optimal way the original trajectories.

We showed that the target-based approximations are correlated with the instantaneous tangential velocity and curvature. They can therefore be used as an alternative means of representing both the geometry and the kinematics of end-point trajectories. Moreover, this representation can be adjusted by adapting the compression rate, according to its influence on the correlation. The results obtained for 3D trajectories are very promising. This method for analyzing the shape and kinematics of gesture trajectories leads to a new analysis tool for multidimensional data.

These empirical approximations provide a significant way to segment gestures. The measure proposed in this paper, in terms of the product of the target-based velocity by the target-based curvature, gives us indeed an original means of delimitating segments which are more or less short, depending on our algorithm parameterization. In order to affirm that these segments represent meaningful components, we should compare them with those obtained through manual segmentation, for larger sequences of gestures containing multiple representations of motion segments. Nevertheless, for gestures composed of chunks whose kinematics strongly discriminate them (with different profiles of acceleration and curvature), it would be interesting to use our segmentation algorithm as an alternative method to the classical geometrical ones.

Manual and automatic segmentation are combined in a more general analysis / synthesis framework. This leads to the proposition of a data-based synthesis process, which associates tracking and inverse kinematics, according to the phases implicitly contained in gestures.

In future works, we will use this approach to identify meaningful units of semantic gestures (for instance sign language gestures). We will thus determine an optimal

compression rate, by temporally aligning the proposed segmentation with a semantically interpretable segmentation. Other variables should also be tested for segmentation, and confronted to manual segmentation. Furthermore, we intend to experiment this approach for dealing with co-articulation aspects of gestures. The evaluation of the synthesis methods will be conducted both for the understanding of sign language gestures and for the realism of the produced gestures.

References

1. Maletic, V.: *Body, Space, Expression: The Development of Rudolf Laban's Movement and Dance Concepts*. Mouton de Gruyter, New York, NY, USA (1987)
2. Prillwitz, S., Leven, R., Zienert, H., Hanke, T., Henning, J., et al.: HamNoSys, version 2.0 - Hamburg Notation System for Sign Languages, an introductory guide. In: *International studies on Sign Language and communication of the Deaf*, vol. 5. Signum Press, Hamburg (1989)
3. Kendon, A.: How gestures can become like words. In: Poyatos, F. (ed.) *Crosscultural perspectives in nonverbal communication*, Toronto, Hogrefe, pp. 131–141 (1988)
4. McNeill, D.: *Hand and Mind – What Gestures Reveal about Thought*. The University of Chicago Press, Chicago (1992)
5. Kita, S., van Gijn, I., van der Hulst, H.: Movement Phase in Signs and Co-speech Gestures, and their Transcriptions by Human Coders. In: Wachsmuth, I., Fröhlich, M. (eds.) *GW 1997. LNCS*, vol. 1371, pp. 23–35. Springer, Heidelberg (1998)
6. Gibet, S., Kamp, J.-F., Poirier, F.: Gesture Analysis: Invariant Laws in Movement. In: Camurri, A., Volpe, G. (eds.) *GW 2003. LNCS*, vol. 2915, pp. 1–9. Springer, Heidelberg (2004)
7. Héloir, A., Gibet, S., Multon, F., Courty, N.: Captured Motion Data Processing for Real Time Synthesis of Sign Language. In: Gibet, S., Courty, N., Kamp, J.-F. (eds.) *GW 2005. LNCS*, vol. 3881, pp. 168–171. Springer, Heidelberg (2006)
8. Viviani, P., Terzuolo, C.: Trajectory determines movement dynamics. *Neuroscience* 7, 431–437 (1982)
9. Soechting, J.F., Terzuolo, C.A.: Organization of arm movements in three dimensional space. Wrist motion is piecewise planar. *Neuroscience* 23, 53–61 (1987)
10. Alexa, M., Müller, W.: Representing Animations by Principal Components. *Computer Graphics Forum* 19(3) (2000)
11. Chenevière, F., Boukir, S., Vachon, B.: A HMM-based dance gesture recognition system. In: *Proceedings of the 9th international workshop on systems, signals and image processing*, Manchester, UK, pp. 322–326 (2002)
12. Boukir, S., Chenevière, F.: Compression and recognition of dance gestures using a deformable model. *Pattern Analysis and Applications (PAA) Journal* 7(3), 308–316 (2004)
13. Marteau, P.F., Gibet, S.: Adaptive sampling of motion trajectories for discrete task-based analysis and synthesis of gesture. In: Gibet, S., Courty, N., Kamp, J.-F. (eds.) *GW 2005. LNCS*, vol. 3881, pp. 168–171. Springer, Heidelberg (2006)
14. Pikaz, A., Dinstein, I.: An algorithm for polygonal approximation based on iterative point elimination. *Pattern Recognition Letters* 16, 557–563 (1995)
15. Kolesnikov, A., Franti, P.: Reduced-search dynamic programming for approximation of polygonal curves. *Pattern Recognition Letters* 24, 2243–2254 (2003)

16. Marteau, P.F., M  nier, G.: Adaptive multiresolution and dedicated elastic matching in linear time complexity for time series data mining. In: Sixth International Conference on Intelligent Systems Design and Applications (IEEE ISDA 2006), Jinan Shandong, China, October 16-18 (2006)
17. Marteau, P.F., M  nier, G.: Speeding up Simplification of Polygonal Curves using Nested Approximations. *Journal of Pattern Analysis and Application*, 1–8 (June 3, 2008), doi: 10.1007/s10044-008-0133-y
18. Gibet, S., Marteau, P.F.: A Self-Organized Model for the Control, Planning and Learning of Nonlinear Multi-Dimensional Systems Using a Sensory Feedback. *Journal of Applied Intelligence* 4, 337–349 (1994)
19. Gibet, S., Marteau, P.F., Julliard, F.: Models with Biological Relevance to Control Anthropomorphic Limbs: A Survey. In: Wachsmuth, I., Sowa, T. (eds.) *GW 2001. LNCS (LNAI)*, vol. 2298, pp. 105–119. Springer, Heidelberg (2002)