HAL
open science

Analysis of Human Motion, Based on the Reduction of Multidimensional Captured Data - Application to Hand Gesture Compression, Segmentation and Synthesis

Sylvie Gibet, Pierre-François Marteau

To cite this version:

Sylvie Gibet, Pierre-François Marteau. Analysis of Human Motion, Based on the Reduction of Multidimensional Captured Data - Application to Hand Gesture Compression, Segmentation and Synthesis. (AMDO 2008), Jul 2008, Port d'Andratx, Spain. pp.72-81, 10.1007/978-3-540-70517-8_8 . hal-00502849

HAL Id: hal-00502849

https://hal.science/hal-00502849

Submitted on 15 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.

Analysis of Human Motion, based on the Reduction of Multidimensional Captured Data– Application to Hand Gesture Compression, Segmentation and Synthesis

Sylvie Gibet[1], Pierre-François Marteau[2]

[1]IRISA/Bunraku, Campus of Beaulieu, F-35042 RENNES CEDEX
[2]VALORIA, University of Bretagne Sud, Campus de Tohannic, BP 573, F-56017 VANNES CEDEX
sylvie.gibet@irisa.fr, pierre-francois.marteau@univ-ubs.fr

Abstract. This paper describes a method to analyze human motion, based on the reduction of multidimensional captured motion data. A Dynamic Programming Piecewise Linear Approximation model is used to automatically extract in an optimal way key-postures distributed along the motion data. This non uniform sub-sampling can be exploited for motion compression, segmentation, or re-synthesis. It has been applied on arm end-point motion for 3D or 6D trajectories. The analysis method is then evaluated, using an approximation of the curvature and the tangential velocity, which turns out to be robust to noise and can be calculated on multidimensional data.

1 Introduction

The representation and the thorough understanding of human motion is a crucial and challenging problem which has been raised in many scientific areas, including animation of virtual characters, analysis of motor performances in sport motion and for disabled people. In recent years, the huge development of new technologies for motion capture has made the analysis of human motion feasible, and yielded data-based methods for gesture analysis, retrieval, and computer-generated animation.

One major problem in representing gesture from recorded data is that these data are multidimensional and direct use of them is rather expensive and fastidious. Another problem is the lack of flexibility. Computing motion from real motion chunks necessitates indeed the elaboration of large data sets, and the development of data-driven methods for tracking, adapting or generating new motion. Finally, finding the best motion representation is a central problem, depending on the application. As these processes operate on multidimensional data, one way to characterize gesture is to compress the original information into relevant samples and to use this data reduction to efficiently retrieve or reconstruct the motion, or to identify meaningful motion units. The automatic extraction of key frames (postures or key points) is also an efficient way to synthesize new gestures, which takes into account the spatial variability of gestures and the co-articulation effects.

In this paper we consider motion captured data consisting of sampled trajectories that characterize the evolution with time of the position and orientation of the human joints. For human gestures, these joint trajectories present specific profiles that can be readable through the analysis of shape (curvature) and kinematics (velocity). 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 cinematic characteristics in a reduced representation space. We use an adaptive sub-sampling algorithm, called DPPLA (Dynamic Programming Piecewise Linear Approximation), which identifies in an optimal manner a set of targets located on the trajectories. This target-based representation of trajectories is applied to automatic segmentation of 3D arm end-point trajectories and to motion reconstruction using inverse kinematics. An evaluation process is defined, using an approximation of the curvature and velocity along the motion sequences. We show that these approximations are strongly correlated to curvature and tangential velocity, not only in the 3D space, but in multi dimensional space. These measures provide a way to automatically analyse gestures. Furthermore, the method can then be extended to multidimensional motion trajectories.

The paper is mainly composed of six sections. Section 2 gives an overview of the related works. After describing in section 3 the adaptive non uniform sampling algorithm (DPPLA) used for data reduction, section 4 proposes an analysis method using an approximation of curvature and velocity. Section 5 presents some results related to segmentation of 3D arm end-point trajectories and the way the obtained segments may be used for synthesis. The paper concludes and gives some perspectives in section 6.

2 Related works

Numerous techniques have been developed for the analysis of human motion captured data. These studies differ considerably, whether the emphasis is placed on data reduction for retrieval, segmentation, recognition, or synthesis purposes.

There are many different mathematical approaches for curves and surfaces approximations, which tend to reduce the dimensionality of the motion data. Few works concern motion trajectories. Polygonal approximation provides characteristics points to represent the shape of the trajectory. These points, which correspond to local curvature extrema, can be connected by line segments. This method has been used by [1] for non-uniform sub-sampling of motion time-series. Another method proposes curve approximation using active contours [2]. These methods are developed for dance gesture recognition.

Other methods have been proposed to the problem of approximating multidimensional curves using piecewise linear simplification and dynamic programming in $O(kn^2)$ complexity [3]. Some efficient algorithms [3-4] (in $O(n\log(n))$ complexity) have been proposed.

The objective in this paper is not so much to find out the best data reduction method, but to define an adaptive method with a pre-defined compression rate that can be applied to multidimensional data. Moreover, we propose an analysis tool, expressed in the reduced space by measures that approximate the curvature and the velocity of motion trajectories.

3 Non uniform sampling algorithm (DPPLA)

The motion consists 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 $X(t) = [x_1(t), y_1(t), z_1(t)\ x_2(t), y_2(t), z_2(t) \ldots x_p(t), y_p(t), z_p(t)]$. In practice, we deal with sampled trajectories at a constant frequency of 120 Hz: *X(n)* where *n* is the time-stamp index.

The approach consists in seeking an approximation $X_{\hat{\theta}}$ of *X(n)*, θ being the set of discrete time location $\{n_i\}$ of the segments endpoints. The selection of the optimal set of parameters $\hat{\theta} = \{\hat{n}_i\}$ is performed using the adaptive dynamic programming presented below.

Let us define $\theta(k)$ as the parameters of a piece wise approximation containing *k* segments, and $\delta(k,i)$ as the minimal error between the best piecewise linear approximation containing *k* segments and covering the discrete time window $\{1,..,i\}$:

$$\delta(k,i) = \underset{\theta(k)}{Min}\left\{\sum_{n=1}^{i}\left\|X_{\theta(k)}(n) - X(n)\right\|^2\right\}$$

$\delta(k,i)$ can be decomposed as follows:

$$\delta(k,i) = \underset{n_k \le i}{Min}\{d(n_k,i) + \delta(k-1,n_k)\}$$

$$where \quad d(n_k,i) = \sum_{n=n_k}^{i}\left\|Y_{k,i}(n) - X(n)\right\|^2$$

$$and \quad Y_{k,i} = (X(i) - X(n_k)).\frac{n-n_k}{i-n_k} + X(n_k)$$

The end of the recursion gives the optimal piecewise linear approximation, e.g. the set of discrete time locations of the extremity of the linear segments:

The result of this method is the optimal identification of discrete X_{Ti} key-points – we call them spatial targets – delimitating the segments, for a given compression rate. The complexity of the algorithm is $O(n^2/k)$ where *n* is the number of samples, and *k* the number of segments, but can be decreased down to $O(n)$ if optimality is somehow relaxed [6].

An example of the application of the algorithm on 3D arm end-point data is illustrated in figure 1.

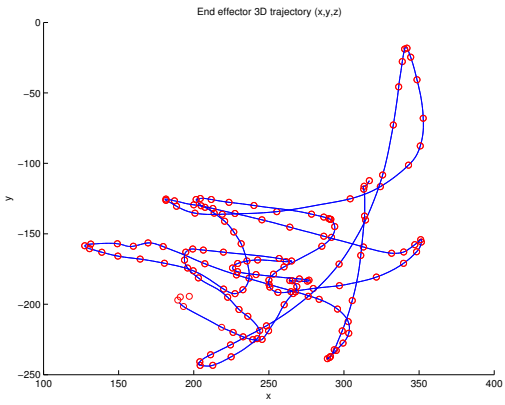

Fig. 1 Approximation of the end-point trajectory using the non uniform sampling algorithm

4 Approximation of curvature and velocity using DPPLA model

First of all we worked 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 expressed the instantaneous velocity $v(t)$ and the absolute value of the instantaneous curvature $\kappa(t)$:

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

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

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, comput-

ing the radius of curvature raises a problem, when the velocity is too high, or when there are inflexion 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:

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

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 also the inverse distance between adjacent targets as:

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

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, $\kappa_{Tgi}(n_i)$ expresses a spatial quantity which might be correlated to curvature at time-index n_i.

This approximation has been experimented on arm end-point motion. 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 2 (left). 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 2 (right). The points with abrupt changes are located at the same place, but the target-based signal seems less noisy than the original one. This makes possible to approximate curvature as the inverse of target density.

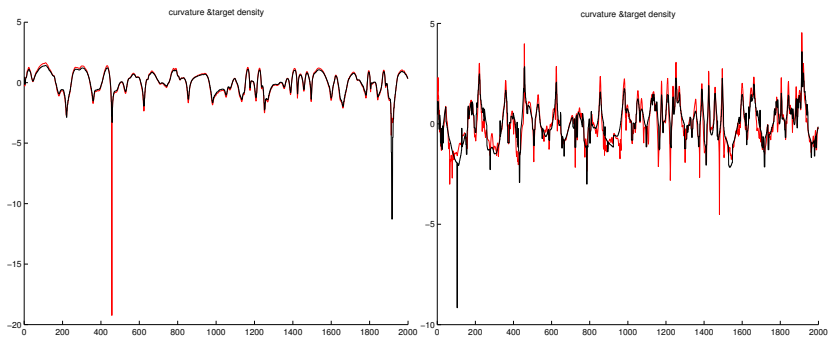

Fig. 2. Correlation for 3D end-point trajectories of arm movements; left: correlation between target-based velocity (black) and instantaneous tangential velocity (red); right: correlation between target density (black) and instantaneous curvature (red);

For each signal x, we computed: $(\log(x) - \text{mean}(\log(x)))/\text{std}(\log(x))$

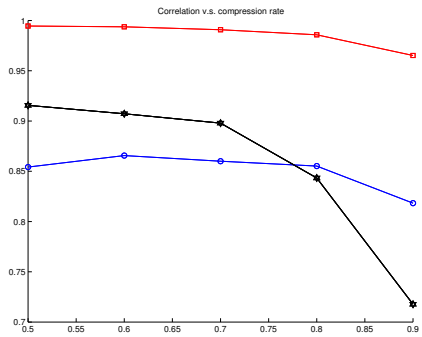

Fig. 3. Correlation coefficient versus compression rate of the adaptive sampling algorithm; (circle): curvature vs. inverse target distance; (star): acceleration vs. target based acceleration; (square): tangential velocity vs. target based velocity

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 3. 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.

5 Gesture segmentation and data-driven synthesis

In this section, we explain how the non uniform sampling algorithm as well as the derived target-based cinematic measures (curvature and velocity) can be applied to both the automatic segmentation of gestures and the data-driven synthesis of gestures.

5.2 Segmentation

Studies on gesture [7] 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 [8], 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 the 3D arm motion. Moreover, it can be used for a compact gesture representation and for data-driven synthesis.

Our segmentation is based on the observation that phase boundaries might occur when the radius of curvature becomes very small, and the velocity decreases at the same time, indicating a change of direction. The segmentation algorithm is based on the product variable $v(t).\kappa(t)$, and on its approximation, based on the approximated target-based variables : $v_{Tgi}(n_i).\kappa_{Tgi}(n_i)$.

A color-coding method allows us 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.

Table1. 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. 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 $v_{Tgi}(n_i).\kappa_{Tgi}(n_i)$ (see figure 4 left and right).

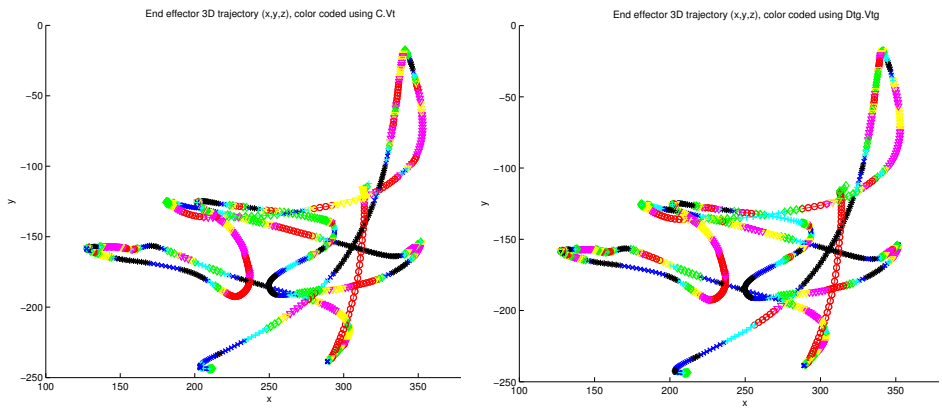

Fig. 4. Example of end-point trajectories segmentation (in the *xy* plane) using a color-coding of quantified variables (different gray levels); left: segmentation using the product $\kappa(t).v(t)$; right: segmentation using the product $\kappa_{Tgi}(t).v_{Tgi}(t)$; A great similarity between the two sequences can be observed.

5.2 Gesture synthesis by inverse kinematics

When applied to 3D end-point trajectories (hand motion), the discrete representation which is provided by the *DPPLA* algorithm gives a non uniform flow of 3D targets. Figure 5 shows the distribution of targets along the gesture sequence (top). The reconstruction error between the sub-sampled trajectory and the original one is shown on figure 5 (bottom).

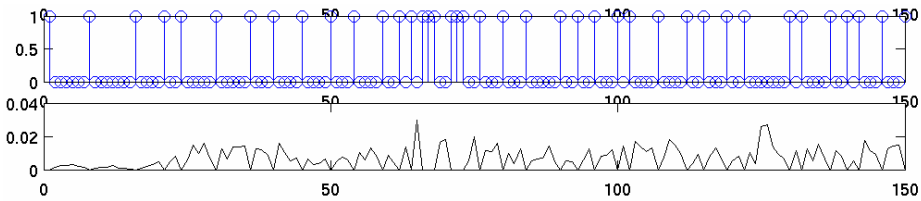

Fig. 5 (top) Motion separation points assigned by the DPPLA algorithm. The x-axis corresponds to the frame number, and the vertical bars specify the target points assigned by the algorithm; (bottom) Reconstruction error between target-based and simulated trajectory

These targets can be directly used as input of an animation engine which automatically computes the angular parameters of the articulated chain, given the end-extremity position. This kinematics inversion process can be achieved in two ways, as illustrated in Figure 6.

The first method consists in reconstructing the trajectory from the non uniform target flow, and then using a classical inverse kinematics (IK) [9, 10] model through a tracking process (Fig. 6 (left)). The reconstruction can be simply achieved by linear or cubic interpolation.

The second method directly uses discrete time-stamped targets, which are fed into a GSMM controller [12, 13]. These controllers provide a means of dealing with co-articulation. In the near future we intend to evaluate the quality of the produced movements according to specific compression rates and to compare the two methods.

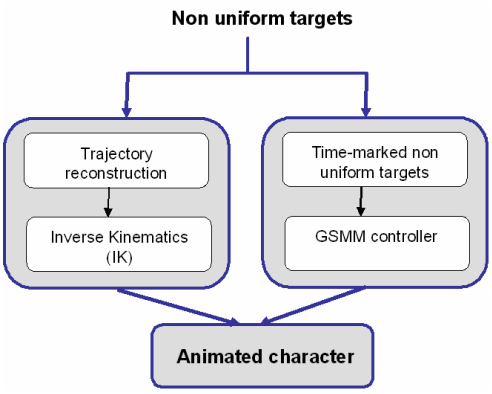

Fig. 6 Synthesis from non uniform targets, using a classical IK algorithm (left), or an adaptive GSMM controller (right)

6 Conclusion and future work

This paper presented a method for analyzing human motion based on an adaptive subsampling technique that leads to reduce multidimensional data. This technique extracts discrete target patterns from raw data, for a given compression rate. From a 3D trajectory, we showed that the target-based trajectory can be reconstructed while keeping the main spatial and cinematic characteristics of the original trajectory.

The evaluation of the compression algorithm is performed using an approximation of the curvature and the velocity calculated along the motion sequence. We showed indeed that the target-based approximations are correlated with the instantaneous tangential velocity and curvature. They can therefore be used as an alternative to represent both the shape 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.

This method has proved to be efficient for 3D or 6D arm-trajectories. It is therefore an interesting method for the analysis of multidimensional data. These empirical approximations provide a significant way to automatically segment gestures. The measure that we propose, characterized by the product of the target-based velocity by the target-based curvature, gives us an original means of delimitating segments which

depends on our algorithm parameterization (compression rate). A study in progress intends to determine how this automatic segmentation might help a manual segmentation process for communication gestures.

The sub-sampling of the end-point trajectory also provides a way to reduce the information flow entering an inverse kinematics process. In the near future, we will show how to exploit this sub-sampling to deal with spatial variability and co-articulation inherent to motion.

References

1. 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, June 2002, pp. 322-326
2. Boukir S., Chenevière F.: Compression and recognition of dance gestures using a deformable model, Pattern Analysis and Applications (PAA) Journal, Springer-Verlag, Vol. 7, No 3, (2004) 308-316.
3. Perez J.C., Vidal E.: Optimum polygonal approximation of digitized curves, Pattern Recognition Letters, Vol. 15. (1994) 743-750
4. Goodrich M.T.: Efficient piecewise-linear function approximation using the uniform metric. Proceedings of the tenth annual symposium on Computational geometry Stony Brook, New York, United States, (1994) 322 – 331
5. Agarwal P.K., Har-Peled S., Mustafa N.H., Wang Y.: Near-Linear Time Approximation Algorithms for Curve Simplification Proceedings of the 10th Annual European Symposium on Algorithms (2002).
6. Marteau, P.F., Ménier, G., Adaptive multiresolution and dedicated elastic matching in linear time complexity for time series data mining, Sixth International Conference on Intelligent Systems Design and Applications (IEEE ISDA 2006), Jinan Shandong, China, 16-18 October, 2006.
7. Kita, S., van Gijn, I., van der Hulst, H. Movement phase in signs and co-speech gestures, and their transcriptions by human coders. Gesture and Sign Language in Human-Computer Interaction, GW 1997, Bielefeld, Germany, Lecture Notes in Computer Science, I. Wachsmuth & al ed., Springer, Vol. 1371, pp. 23-35, 1998.
8. A. Majkowska, V. Zordan, and P. Faloutsos. Automatic slicing for hand and body animations. Eurographics/ ACM SIGGRAPH Symposium on Computer Animation (2006), pp. 1-8, M.P. Cani, J. O'Brien (Ed.)
9. D. Chi, M. Costa, L. Zhao, and N. Badler: The EMOTE model for Effort and Shape, ACM SIGGRAPH '00, New Orleans, LA, (2000) 173-182
10. S. Carvalho , R.Boulic, . Thalmann. "Interactive Low-Dimensional Human Motion Synthesis by Combining Motion Models and PIK", Journal of Computer Animation and Virtual Worlds,Wiley, 2007
11. S. Gibet S., P.F. Marteau P.F. A self-organised model for the control, planning and learning of nonlinear multivariable systems using a sensori- feedback, Journal of Applied Intelligence, Kluwer Academic Publisher, vol 4, Boston, 337-349, 1994.
12. S. Gibet S., P.F. Marteau. Expressive Gesture Animation Based on Non Parametric Learning of Sensory-Motor Models, CASA 2003, IEEE Computer Animation and Social Agents, May 7-9, 2003.