Fast Motion retrieval with the distance input space

Thibaut Le Naour, Nicolas Courty, Sylvie Gibet

Fast Motion retrieval with the distance input space

Thibaut Le Naour*, Nicolas Courty*, Sylvie Gibet*

* VALORIA, University of Bretagne Sud, 56000 Vannes, France

Abstract. Accessing and querying large motion databases can be a time consuming operation since motions are by essence big multivariate time series. However, with the help of good dimensionality reduction methods, this problem can be alleviated. Yet, it appears that the intitial motion representation space if of great importance with regards to the final performances. This paper advocates the idea that the distance representation space (*i.e.* ensemble of joint-to-joint distances of the skeleton), though bigger, can yield a powerful representation of the motion. In order to validate this idea, comparisons with rotation and position based representations are experimented in a motion retrieval task performed with the HDM05 motion database. The final resulting retrieval algorithm is very fast with satisfying performances.

1 Introduction and related work

There is a large number of tehniques to represent a motion. Associated to these representations, motion retrieval algorithms needs efficient similarity measures. Tradionally, these similarity measures can be separated in two categories: numerical and logical. Among numerical approaches, dimensionality reduction has the advantage to offer a representation smaller than original motion without losing major information. Alexa and Muller[**?**] are the first ones to introduce PCA technique, which consists in to project a maximum of information on the axis which describe the most significant of the motion. In 2007, Xiang and Zhu[**?**] presents a new non-linar method: isometric feature mapping based on geodesic distances. Kovar and al [**?**] proposes a new method based on the distances between point clouds of motion postures. He uses dynamic time warping algorithm (DTW) in particulary to bring out spatial and temporal relations between movements. DTW is an efficient algorithm but it is expensive in time calculation. Other approaches like geometric representation associated to logical laws provide an abstract characterization of the motion. Müller et al.[**?**] pre-process the motion database by dividing motions according to qualitative geometric features. Later motion template defined by matrices (row represent spatial information and column temporal data) were introduced in [**?**]. More recently Liang et al.[**?**] presented an original method based on kinematic and dynamic constraints to retrieve a motion.

In this paper, we highlight the importance of using relevant representations of movements to achieve very fast motion retrieval. We show that a distance-based motion representation, coupled to the inner product in the latent space leads to good recognition rates with a very low computational cost. Our results are tested for movements contained in the database HDM05.

2 Motion representation

We propose to use the spatial relationships between joint positions in order to characterize a motion. More specifically, a motion is expressed as the evolution of Euclidean distances between skeleton joints over time. This space will be referred to as the distance space. For each posture, we extract a graph $\boldsymbol{G} = (\boldsymbol{V}, \boldsymbol{E}, d)$ where $\boldsymbol{V}$ is the set of vertices and $\boldsymbol{E}$ is the set of edges which are changing over time. The segments composing the skeleton being fixed during the animation, we do not retain them to express the movement. d is therefore the set of distances of $V \times V \rightarrow \mathbb{R}_+$. We express in particular the Euclidean distance between the point $\boldsymbol{x}_i$ and the point $\boldsymbol{x}_j$, for every $(\boldsymbol{x}_i, \boldsymbol{x}_j) \in E$. The spatial structure of the graph can be expressed by the set of equations:

$$\forall(\boldsymbol{x}_i, \boldsymbol{x}_j) \in E, \|\boldsymbol{x}_i - \boldsymbol{x}_j\| = d_{ij}, \tag{1}$$

where $\|.\|$ is the Euclidian norm. Finally, a motion is represented by a $n \times m$ matrix $\boldsymbol{M}$ where n is number of distance and m number of frame composing motion. Figure 1 illustrates this graph for one posture. Contrary to positional or rotational representations, it is interesting to observe that an important amount of data is thus necessary to represent a posture. For example, in a case of a skeleton with 31 joints, 465 distances values are necessary to define the posture, where only $31 * 3 = 93$ values are needed in the position space. Information characterizing motion being most of the time redundant (joint orientations, positions, or distances), statistical methods have been introduced to reduce the representation space. In particular, classical methods use Principal Component Analysis (PCA) to decompose motion into an orthogonal subspace, in which each principal components drive a decreasing amount of variance. This approach allows to compress motion information, and to give a simplified and significant representation of motion through the first eigenaxis. **We postulate here that the high-dimensional space induced by distances will adapt particularly well to such linear dimensionality reduction techniques, and bring eigenaxis that will capture well the essence of motions.** This assumption is inspired by the classical kernel methods theory that first embed data into high-dimensional before conducting linear statistical techniques.

3 Methodology and results

In order to retrieve motion which are similar to an example motion in the database, we use a similarity measure which is based on the first axis of the PCA method applied on the matrix of distances characterizing each movement. More precisely, for two motions represented by the matrices $\boldsymbol{M}_i$ and $\boldsymbol{M}_j$, the similarity measure is defined by the inner product between the principal vectors

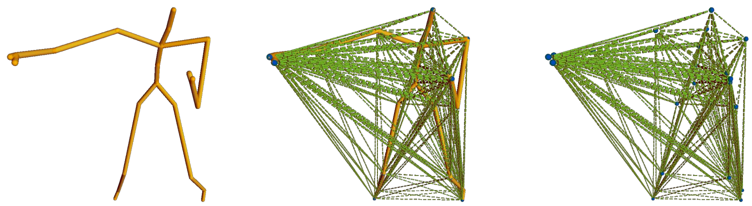

Fig. 1. Creation of the graph associated to a posture: in yellow the skeleton and in green dotted line the edges of variable lengths.

$\boldsymbol{V}_i$ and $\boldsymbol{V}_j$, with $\boldsymbol{V}_i$ is the first eigen vector extracted through a singular value decomposition (SVD) applied to $\boldsymbol{M_i}$. **Let us note here that this measure allows to work independently of the temporal length of the considered sequences.**

Our method has been tested on the database HDM05[**?**], which contains a collection of 497 motions distributed in 78 different classes. After removing the example-motion from the database, we define for each movement a criteria by searching the nearest neighbors of a movement and testing its class. Hence, one entry $c_{i,j}$ of the confusion matrix will be the number of times a motion of class i has been associated to a motion of class j normalized by the number of motions in class i. Of course, the best retrieval algorithm would give high values along the diagonal. The test is completed for all the movements, without any prior treatment. The results are illustrated in figure 2. It shows the confusion matrices when searching a motion represented by the first eigen vector resulting from the PCA, applied on distances (a), positions (b), and orientations (c). Furthermore, the searching is performed for the first two nearest neighbors. We observe a strong coherency between searched movements and the results in the case of movements represented by distances, and inversely there is no correlation for movements represented by positions or rotations. This results show the pertinence to use distances to represent a motion than positions and rotations. The efficiency rate of our algorihm so the number of success motion retrieve by 2-nearest neighbors algorithm for all motion, is 68.8% and of success without any prior processing of the database (13.7% for positions and 13.9% for rotations).

The comparison time is negligible since the comparison of a motion with the rest of the database corresponds to the inner product of two 465-values vectors.

4 Conclusion

Our objective was to provide a fast motion retrieval method, capable of finding quickly motions similar to the example motion in a large motion database. Our

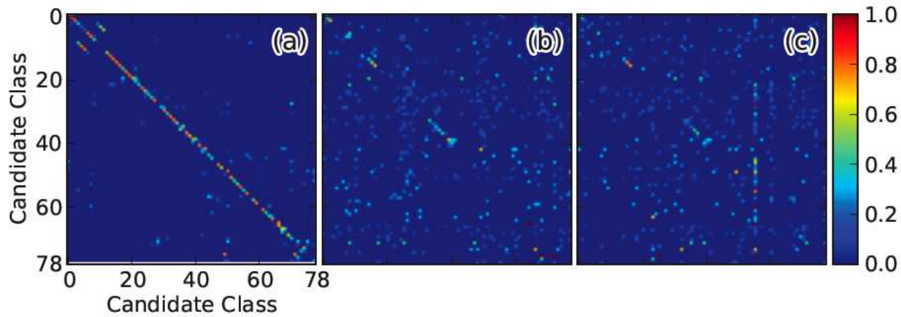

Fig. 2. Confusion matrices obtained using distance (a), positional (b) and rotational (c) representation space of the motions

method is based on a motion representation in the distance space, a motion being characterized by the evolution over time of distances between the skeleton joints. The motion is then reduced by using a principal component analysis. With such an approach, searching a motion within the database is performed simply and quickly through the inner product of the first PCA-axis for each pair of movements. We showed that it is interesting to represent a motion in distance space, in particulary because our technique need more information than position and rotation space to describe a motion. Moreover our technique which can be carried out in real time on non-pretreated motions from the database HDM05, is promising, with a success rate close to 70%. Future works will consider thorough comparisons with state-of-the-art algorithms in motion retrieval.