Signal Processing 120 (2016) 691–701

Contents lists available at ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

Adaptive multi-view feature selection for human motion retrieval

Zhao Wang [a], Yinfu Feng [b], Tian Qi [b], Xiaosong Yang [a,*], Jian J. Zhang [a]

[a] *Bournemouth University, Poole Dorset BH12, United Kingdom*

[b] *School of Computer Science, Zhejiang University, Hangzhou 310027, PR China*

ARTICLE INFO

Article history:
Received 7 September 2014
Received in revised form
21 November 2014
Accepted 24 November 2014
Available online 3 December 2014

Keywords:
Human motion retrieval
Feature selection
Multi-view learning
Trace ratio minimization problem

ABSTRACT

Human motion retrieval plays an important role in many motion data based applications. In the past, many researchers tended to use a single type of visual feature as data representation. Because different visual feature describes different aspects about motion data, and they have dissimilar discriminative power with respect to one particular class of human motion, it led to poor retrieval performance. Thus, it would be beneficial to combine multiple visual features together for motion data representation. In this article, we present an Adaptive Multi-view Feature Selection (AMFS) method for human motion retrieval. Specifically, we first use a local linear regression model to automatically learn multiple view-based Laplacian graphs for preserving the local geometric structure of motion data. Then, these graphs are combined together with a non-negative view-weight vector to exploit the complementary information between different features. Finally, in order to discard the redundant and irrelevant feature components from the original high-dimensional feature representation, we formulate the objective function of AMFS as a general trace ratio optimization problem, and design an effective algorithm to solve the corresponding optimization problem. Extensive experiments on two public human motion database, i.e., HDM05 and MSR Action3D, demonstrate the effectiveness of the proposed AMFS over the state-of-the-art methods for motion data retrieval. The scalability with large motion dataset, and insensitivity with the algorithm parameters, make our method can be widely used in real-world applications.

© 2014 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/3.0/).

1. Introduction

Motion capture (MOCAP) is a prevalent and powerful technique to record the movement of human or other objects. It has been widely used in various areas such as computer animation, computer games, film production, sports medicine and athletic training. Since the commercial MOCAP systems are very expensive and the motion capture process is time-consuming, it becomes critical to reuse the abundant pre-captured motion data. With the rise of some novel motion capture systems and technologies like depth-based MOCAP [1,2], it brings explosive growth of motion data. A huge mass of human motion data have been accumulated in these years. Therefore, an efficient and effective human motion retrieval algorithm plays an important role in managing and accessing these available data.

Due to the high complexity and diversity of human motion, human motion retrieval is a very challenging task in computer animation and multimedia analysis communities. Similar to other multimedia retrieval tasks like image

[*] Corresponding author.
E-mail addresses: zwang@bournemouth.ac.uk (Z. Wang), fyf200502@hotmail.com (Y. Feng), qteat@zjt.edu.cn (T. Qi), xyang@bournemouth.ac.uk (X. Yang), jzhang@bournemouth.ac.uk (J.J. Zhang).

http://dx.doi.org/10.1016/j.sigpro.2014.11.015
0165-1684/© 2014 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/3.0/).

retrieval and document retrieval, feature representation is the corner-stone of human motion retrieval algorithm and system. A compact and discriminative motion feature representation cannot only significantly improves the performance of the retrieval algorithm but also dramatically reduces the consuming time of the algorithm. As we known, there is a big gap between the low-level visual feature and the high-level semantic meaning, so single low-level visual feature is usually unable to fully characterize all aspects of the motion data. In other words, it would be beneficial to fuse multiple visual features for motion data representation.

However, many researchers tended to use a single type of visual feature as motion data representation in the past years, which leads to under-performing retrieval performance [3,4]. In contrast, recent studies have shown that using multi-view features can dramatically boost the algorithms' performance in computer vision and machine learning [5–7]. The core problem in these methods is how to exploit these multi-view features that has specific statistic property and lies on some low-dimensional subspace space in the feature fusion phase. An intuitive way is to concatenate these multi-view features into a high-dimensional feature representation, which would bring in the risk of dimensional explosion as well as the redundancy of information. Because each visual feature characterizes different aspects of human motion and has dissimilar discriminative power with respect to a specific class of motion, it is wise to jointly take these multi-view features into account and exploit the embedded complementary information. In addition, we notice that although multi-view feature learning is a hot topic in computer vision and machine learning, less research effort have been made in the problem of multi-view feature learning and feature selection for human motion retrieval.

To this end, we propose an Adaptive Multi-View Feature Selection (AMFS) method in this article, which can automatically assign multi-view features with adaptive feature weights and select out a compact and discriminative feature subset from the original high-dimensional multi-view features. With the selected low-dimensional feature representation, it not only improves the motion data retrieval accuracy but also speeds up the whole motion data retrieval processing.

The remainder of this article is organized as follows. In Section 2, some recent research work on motion feature representation and multi-view feature selection for motion data related tasks will be introduced. Then, the details of our proposed AMFS and its optimization method will be presented in Section 3. Finally, the experiments and some discussions are given in Sections 4 and 5, respectively.

2. Related work

To solve the human motion retrieval problem, we resort to adaptive multi-view feature selection to learn a compact yet discriminative feature representation from multi-view features. Our method relates to two topics: human motion retrieval and multi-view feature selection. Thus, in this section we briefly review some related work on these two topics one by one.

2.1. Human motion retrieval

With the explosive increase of multimedia data, extensive research effort has been dedicated to multimedia analysis, such as image and video annotation and retrieval [8,9]. To reveal and exploit the geometric structure of multimedia data, many graph-based models have been proposed in recent years. They can be used as geometric image descriptors [10–12] to enhance image categorization. Besides, these methods can be used as image high-order potential descriptors of superpixels [13–15]. Further, graph-based descriptors can be used as a general image aesthetic descriptors to improve image aesthetics ranking, photo retargeting and cropping [16,13].

Human motion data is a special kind of multimedia data, it has many similar properties with image and videos. Thus, many traditional multimedia retrieval approaches can be employed to process the motion data. In general, the main motion retrieval procedure includes three steps: (1) visual feature extraction; (2) training a search/ranking model; (3) similar human motion retrieval or ranking. To speed up the searching procedure, some data indexing techniques such as kd-tree and R-tree are used in approximate searching [17–19]. Indeed, the success of human motion retrieval heavily relies on the motion feature representation. Therefore, many researchers have focused their attentions on construction and learning efficient motion feature representation for content-based human motion retrieval.

As we known, textural labels [20] such as *jump* or *kick* have been used to search motion clips at the beginning. However, it requires the user to label all motion data in database in advance, which is very tedious and time-consuming. Moreover, due to the high ambiguity and subjectivity of keyword, it is difficult to choose a correct and meaningful labels for a given human motion sequence. In fact, how to measure the similarity between two texture labels is still an open problem in text mining.

To overcome the shortcomings of label-based methods, the content-based methods that extract numeric features from motion data and use them to measure the similarity between two human motion have been proposed [3,4]. Based on the used feature of the algorithm, we divide the exiting work into two categories: frame/pose based methods and sequence/clip based methods.

To describe the local motion properly, a number of frame/pose based motion features have been developed. Chen et al. [4] presented a geometric pose feature (GPF) which is effe; ctive in encoding pose similarity. Raptis et al. [57] proposed an approach that using joint angles as feature to recognize dance actions. Xia et al. [21] had shown a histogram of 3D joints descriptors, after generating a dictionary, the temporal motion model is created via HMM. Following a similar paradigm, Kapsouras and Nikolaidis [22] used joints orientation angles and angles forward differences as local features to create a code book and generate a Bag of Visual Word (BOVW) to represent the action. A Gaussian Mixture Model (GMM) was applied by Qi et al. [23] to represent character poses, wherein the motion sequence is encoded, and then DTW and a string matching algorithm are applied for motion comparison. Barnachon et al. [24] used histograms of action poses to represent the action and employed DTW for comparing and recognizing actions. Besides, these two approaches [23,24] have shown the potential for online action recognition. Moreover, sparse representation can also be applied in motion data retrieval [25]. Generally, to use pose/frame based features, it requires an integration step such as BOVW [22], DTW [23] or manifold embedding [10,26] to represent the whole motion clip.

The sequence/clip based motion feature utilizes some global properties of the motion sequence, e.g. moving path or trajectories. Muller et al. [3,27] presented the motion template (MT) that the motions of the same class can be represented by an explicit interpretable matrix using a set of Boolean geometric feature. To overcome the tolerate temporal variance problem in MT training process, dynamic time warping (DTW) was employed in their work. However, it still requires the training data in same class to have same number of cycles, otherwise MT will not be well trained. Moreover, both the training and comparing process of MT is time-consuming. Gowayyed et al. [28] proposed a trajectory feature named histograms of oriented displacements (HOD). Each 3D trajectory is represented by the HOD of its 2D projection to *xy*, *xz* and *yz* planes. A temporal pyramid method, where the trajectory is computed using the whole, halves and then quarters of the motion sequence, is applied to preserve the motion temporal information. Compared with previous trajectory features [29,30], HOD provides a fixed-length feature representation for variable length motion sequences, which benefits further analysis. Xiao et al. [31] proposed a probability graph model for motion retrieval. However their work requires selecting representative frames from motion sequence before generating the probability graph.

Recently, a new technique called sketch-based interface searching, has been introduced in motion retrieval. Chao et al. [32] used a hand drawing sketch as query to describe the body/joint trajectories. For both the input query and the motion clips in database, spherical harmonic is used to encode the joint/body trajectories for motion representation. Choi et al [33] proposed a sketch based motion retrieval framework that using skeleton stick figure as input query, where the motion clips in the database are converted to sequences of stick figures. Then the retrieval work is realized by matching stick figures. However, it requires special skills and experiences for drawing stick figure, which brings in some new difficulties for the user.

2.2. Multi-view feature selection

The traditional feature selection methods such as Laplacian Score [34] and Feature Ranking [35] are developed to select features that preserve the data property or data manifold structure. However, these learning methods were designed for the single-view feature data scenario, and they did not consider the correlations and complementary information between different features when they applied to deal with multi-view data that comprise multiple features in feature representation. Thus, some multi-view feature selection methods were developed for various applications like multimedia data ranking, annotation, recognition and retrieval [5,36,6,37–40], where the information and correlations between different features have been exploited. To integrate multimodel feature, Zhang et al. [41] presented the feature correlation hypergraph (FCH), where shared entropy is proposed to capture the high order correlation among multimodal features. Multimodel graph learning methods can also be implemented to video annotation [42] and image ranking [43]. Moreover, some metrics [44,45] have been developed specially for multi-view feature based methods, which measure the similarities and dissimilarities more accurately than the traditional Euclidean distance metric when using multi-view features.

However, as motion data belongs to a special kind of multimedia data, few multi-view feature based researches are reported on this topic. Some of previous multi-view feature selection approaches [6] can handle multiple real data applications, but no enough experiments on motion data have been taken and shown. Chen et al. [4] used feature selection method to boost their proposed geometric pose descriptor. However, his work focuses on the task of pose recovery from the monocular image. Gao et al. [46] developed a multi-modality action recognition approach based on collaborative representation [47]. Tang and Leung [7] proposed to select suitable kinds of features for each query and then split them. Their work [46,7] did not consider correlations between the original features, where some prior researches [5,6,38] have already shown that leveraging shared information is beneficial to improve the recognition accuracy. Therefore, we are going to develop a multi-view feature selection method for motion data retrieval that takes the discriminating power of each feature synergistically.

To this end, the local graph of each type of feature is constructed separately to flexibly model the visual similarity under each type of feature view. Then, these learned graphs are united with a non-negative view-weight vector to jointly fusion the original multiple features. Finally, we propose the objective function of MFS for adaptively learning feature weight and selecting out a compact and discriminative feature components from the original multiple high dimensional feature representation. The feature selection process brings in twofold advantages: (1) compared with the original multiple high dimensional feature representation, the selected feature subset becomes lower and compact, which leads to less computational cost; (2) after feature selection, only the most correlated feature components are hold thus the redundancy information is discarded, which potentially leads a better performance. The flowchart of our proposed AMFS is shown in Fig. 1.

3. AMFS algorithm and optimization method

3.1. Detail of AMFS

Suppose a human motion dataset consists of N motion clips, we denote them as $\{x_1, x_2, \ldots, x_N\}$ and represent them using M types of visual features. For any motion clips, $x_i^{(v)} \in \mathbb{R}^{d_v \times 1}$ is the v-th visual feature representation of the motion clip x_i, where d_v is the dimension of the v-th visual feature. Thus, the v-th feature data matrix of this dataset can be denoted as $X^{(v)} = [x_1^{(v)}, x_2^{(v)}, \ldots, x_N^{(v)}] \in \mathbb{R}^{d_v \times N}$. The multi-view feature data matrix of this dataset is denoted as $X = [X^{(1)}; X^{(2)}; \ldots; X^{(M)}] \in \mathbb{R}^{D \times N}$, where $D = \sum_{v=1}^{M} d_v$. Since different visual feature describes different aspects about motion

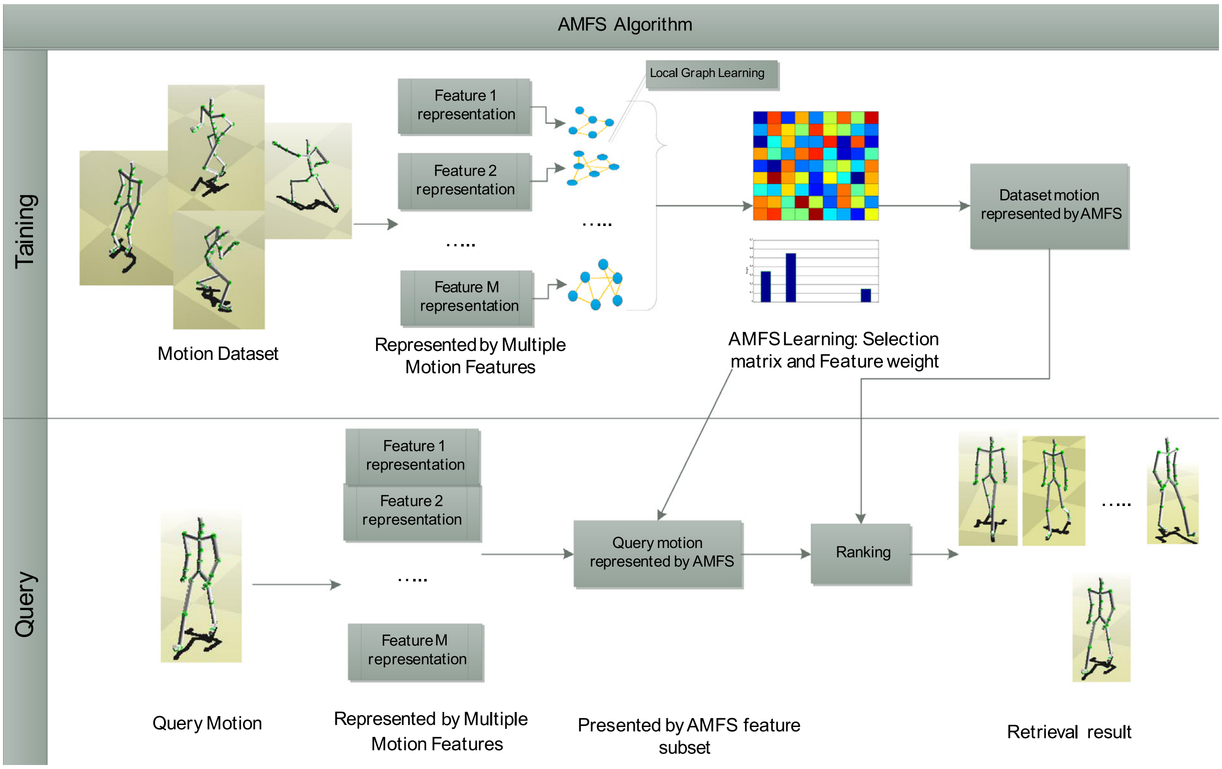

Fig. 1. The flowchart of our proposed AMFS motion retrieval framework. It comprises of two parts: (1) AMFS learning part, which learns a compact and discriminative feature representation from original high dimensional multi-view features. Therefore, the motion data in database can be represented by the learned compact features; (2) motion retrieval part, where the query motion is firstly represented by the original multiple high dimensional features, and then is translated to the same compact feature space with the learned feature selection matrix. After that, many existing ranking algorithms can be directly applied to the task of motion retrieval in our framework.

data, and they have dissimilar discriminative power with respect to one particular class of human motion. Thus, it would be beneficial to combine multiple visual features together for motion data representation. If we just concatenate all of these multiple features together, it does not only lack of physical meaning but also fail to exploit the complementary information between different visual feature. More worse, it would bring in some redundant information in feature representation. To overcome these problems, we hope to select out a compact yet discriminative feature subset $Y \in \mathbb{R}^{d \times N}, d < D$ from the original multiple features X via feature selection method. That is to say, we hope to get

$$Y = W^T X, \quad Y \in \mathbb{R}^{d \times N}, \quad W \in \{0,1\}^{D \times d} \tag{1}$$

where W is a feature selection matrix that picks out the most representative feature elements from X.

Since the local geometric structure information is important in real-world applications [48], we hope to preserve the local information of motion data during feature selection. To this end, a local regressive model is designed to character the local information of motion data [49]. We assume that for each data point, its selected low-dimensional feature can be expressed as a linear function mapping from its original visual feature.

Taken a data point $x_i^{(v)}$ in the v-th feature view $X^{(v)}$, where $X^{(v)} = [x_1^{(v)}, x_2^{(v)}, \ldots, x_N^{(v)}] \in \mathbb{R}^{d_v \times N}$ for example, we first find its k nearest neighbors $N_k(x_i^{(v)}) = \{x_i^{(v)}, x_{i1}^{(v)}, x_{i2}^{(v)} \ldots, x_{i(k-1)}^{(v)}\}$, wherein $x_{ip}^{(v)}, p = 1, 2, \ldots, k-1$ is the k-nearest neighbors of $x_i^{(v)}$ in $X^{(v)}$. If we denote $y_j \in Y_{N_i}$ is the corresponding feature selection result for $x_j \in N_k(x_i^{(v)})$, the local regressive model can be formulated as below:

$$y_i^{(v)} = f_i(x_i^{(v)}) = (w_i^{(v)})^T x_i^{(v)} + b_i^{(v)}, \quad i = 1, 2, \ldots, N \tag{2}$$

where $w_i^{(v)} \in \mathbb{R}^{d_v \times d}$, $b_i^{(v)} \in \mathbb{R}^d$.

In real-world applications, the local geometric structure of visual data is approximately linear [48]. Besides, it is computational efficiency for practical applications using linear method. Therefore, a linear method is chosen in this work rather than other complex nonlinear models [49]. That is to say, a linear regression model $f_i(x_i^{(v)}) = (w_i^{(v)})^T x_i^{(v)} + b_i^{(v)}$ could be used for each data point $x_j^{(v)} \in N_k(x_i^{(v)})$ for mapping from its original visual feature to the corresponding low-dimensional feature. Then, the local prediction error can be represented by $\| f_i(x_j^{(v)}) - y_j \|^2$, that is

$$\| (w_i^{(v)})^T x_j^{(v)} + b_i^{(v)} - y_j^{(v)} \|^2 \tag{3}$$

And, the local model error can be computed by summing the local predict error of each data point in $N_k(x_i^{(v)})$. Thus, we can minimize the local model error to get a best mapping function, that is

$$\min \sum_{x_j^{(v)} \in N_k(x_i^{(v)})} \| (w_i^{(v)})^T x_j^{(v)} + b_i^{(v)} - y_j \|^2 \tag{4}$$

In order to avoid over-fitting, a regularization term is imposed, so the local model error is formulated as

$$\sum_{x_j^{(v)} \in N_k(x_i^{(v)})} \| (w_i^{(v)})^T x_j^{(v)} + b_i^{(v)} - y_j \|^2 + \mu \| w_i^{(v)} \|_F^2 \tag{5}$$

Now, considering the v-th feature view rather than a single data point, let us denote $X_i(v) = [x_i^{(v)}, x_{i1}^{(v)} \ldots, x_{i(k-1)}^{(v)}] \in \mathbb{R}^{d_v \times k}$, $Y_{N_i} = [y_i, y_{i1}, \ldots, y_{i(k-1)}] \in \mathbb{R}^{d \times k}$. The local model error can be written as

$$\| (X_i^{(v)})^T w_i^{(v)} + 1_k (b_i^{(v)})^T - Y_{N_i}^T \|_F^2 + \mu \| w_i^{(v)} \|_F^2 \tag{6}$$

Then the objective function of minimizing the local model error can be formulated as

$$\min_{w_i^{(v)}|_{i=1}^N, b_i^{(v)}|_{i=1}^N, Y_{N_i}^{(v)}|_{i=1}^N} \sum_{i=1}^N \| (X_i^{(v)})^T w_i^{(v)} + 1_k (b_i^{(v)})^T - Y_{N_i}^T \|_F^2 + \mu \| w_i^{(v)} \|_F^2 \tag{7}$$

By setting the derivatives of (7) to the zero w.r.t $b_i^{(v)}$, we have

$$\begin{aligned} &(w_i^{(v)})^T X_i^{(v)} 1_k + k b_i^{(v)} - Y_{N_i} 1_k = 0 \\ &\Rightarrow b_i^{(v)} = \frac{1}{k}\left(Y_{N_i} 1_k - (w_i^{(v)})^T X_i^{(v)} 1_k \right) \end{aligned} \tag{8}$$

By setting the derivatives of (7) to the zero $w_i^{(v)}$ and substituting (9), we have

$$\begin{aligned} &X_i^{(v)} (X_i^{(v)})^T w_i^{(v)} + X_i^{(v)} (1_k (b_i^{(v)})^T - Y_N^T) + \mu w_i^{(v)} = 0 \\ &\Rightarrow w_i^{(v)} = [X_i^{(v)} H (X_i^{(v)})^T + \mu I]^{-1} X_i^{(v)} H Y_{N_i}^T \end{aligned} \tag{9}$$

where $H = I - (1/k) 1_k 1_k^T$ is the centralized matrix. Note that $H = HH^T = H^T$. Substituting (9) and (10) to (7), then we can get

$$\begin{aligned} &(X_i^{(v)})^T w_i^{(v)} + 1_k (b_i^{(v)})^T - Y_{N_i}^T \\ &= H (X_i^{(v)})^T [X_i^{(v)} H (X_i^{(v)})^T + \mu \cdot I]^{-1} X_i^{(v)} H Y_{N_i}^T - H Y_{N_i}^T \end{aligned} \tag{10}$$

Thus, the original objective function (8) can be formulated as

$$\begin{aligned} \min_{Y_{N_i}^{(v)}|_{i=1}^N} \sum_{i=1}^N & \| H((X_i^{(v)}))^T [X_i^{(v)} H (X_i^{(v)})^T + \mu \cdot I]^{-1} X_i^{(v)} H Y_{N_i}^T - H Y_{N_i}^T \|_F^2 \\ & + \mu \| [X_i^{(v)} H (X_i^{(v)})^T + \mu \cdot I]^{-1} X_i^{(v)} H Y_{N_i}^T \|_F^2 \end{aligned} \tag{11}$$

Let us denote

$$\begin{aligned} \mathcal{J} = \sum_i^N & \| H (X_i^{(v)})^T [X_i^{(v)} H (X_i^{(v)})^T + \mu \cdot I]^{-1} X_i^{(v)} H Y_{N_i}^T - H Y_{N_i}^T \|_F^2 \\ & + \mu \| [[X_i^{(v)} H([X_i^{(v)})^T + \mu \cdot I]^{-1} X_i^{(v)} H Y_{N_i}^T \|_F^2 \\ = & \operatorname{tr}\{ Y_{N_i} [H (X_i^{(v)})^T (X_i^{(v)} H (X_i^{(v)})^T + \mu \cdot I)^{-1} X_i^{(v)} H - H]^2 Y_{N_i}^T \} \\ & + \mu \operatorname{tr}\{ Y_{N_i} [H (X_i^{(v)})^T (X_i^{(v)} H (X_i^{(v)})^T + \mu \cdot I^{-2} X_i^{(v)}] H] Y_{N_i}^T \} \\ = & \operatorname{tr}\{ Y_{N_i} [H - H (X_i^{(v)})^T (X_i^{(v)} H (X_i^{(v)})^T + \mu \cdot I)^{-1} X_i^{(v)} H] Y_{N_i}^T \} \end{aligned} \tag{12}$$

Therefore, the objective function (12) is equivalent to

$$\min_{Y_{N_i}|_{i=1}^N} \sum_{i=1}^N \operatorname{tr}\{ Y_{N_i} L_i^{(v)} Y_{N_i}^T \} \tag{13}$$

where $L_i^{(v)} = H - H (X_i^{(v)})^T (X_i^{(v)} H (X_i^{(v)})^T + \mu \cdot I)^{-1} X_i^{(v)} H$, $L_i^{(v)} \in \mathbb{R}^{k \times k}$. Let us denote a selection matrix $S_i^{(v)} \in \mathbb{R}^{N \times k}$ that $(S_i^{(v)})_{pq} = 1$ when data point $x_p^{(v)}$ is the qth neighbor of $x_i^{(v)}$, otherwise $(S_i^{(v)})_{pq} = 0$. Hence, the local feature selection results of $x_i^{(v)}$ and its neighbors can be represented as $Y_{N_i} = Y S_i^{(v)}$. Then, we can get the objective function as

$$\min_Y \sum_{i=1}^N \operatorname{tr}\{ Y S_i^{(v)} L_i^{(v)} (Y S_i^{(v)})^T \} \tag{14}$$

Now, let us focus on the whole feature rather than looking at the single data point and its neighbors. The local graph for the vth feature then can be built as

$$L^{(v)} = [S_1^{(v)}, S_2^{(v)} \ldots, S_N^{(v)}] diag(L_1^{(v)}, L_2^{(v)} \ldots, L_N^{(v)}) [S_1^{(v)}, S_2^{(v)} \ldots, S_N^{(v)}]^T \tag{15}$$

Thus, we formulate the objective function for the v-th feature as

$$\min_Y \operatorname{tr}\{ Y L^{(v)} Y^T \} \tag{16}$$

Here, let us move on to consider all the features rather than a single feature. Eq. (1) in beginning can be represented as a combination of selection from all the features:

$$\begin{aligned} Y &= \begin{bmatrix} W_{d_1 \times d} \\ W_{d_2 \times d} \\ \vdots \\ W_{d_M \times d} \end{bmatrix} \begin{bmatrix} X^{(1)} \\ X^{(2)} \\ \vdots \\ X^{(M)} \end{bmatrix} = [W_{d_1 \times d}^T, W_{d_2 \times d}^T, \ldots, W_{d_M \times d}^T] \begin{bmatrix} X^{(1)} \\ X^{(2)} \\ \vdots \\ X^{(M)} \end{bmatrix} \\ &= \sum_{v=1}^M W_{d_v \times d}^T \cdot X^{(v)} = \sum_{v=1}^M Y^v \end{aligned} \tag{17}$$

Since each feature characterizes different aspects of the motion and hold different intrinsic discriminative power. In order to keep the locality from each feature view and take the synergy effect of all features, a non-negative weight vector α is introduced to unite all the features together.

$$\begin{aligned} \min_Y & \sum_{v=1}^M \alpha_v \operatorname{tr}(Y L^{(v)} Y^T) \\ & = \operatorname{tr}\left(Y \sum_{v=1}^M \alpha_v L^{(v)} Y^T \right) \end{aligned} \tag{18}$$

where $\sum_{v=1}^M \alpha_v = 1, \alpha_v \geq 0$. The value range of α is between 0 and 1 since prior research [49] has shown that negative of α is lack of physical meaning and the non-negative constraint would make result much closer to the idea solution. One thing need to be pointed out is that, the solution to α in last equation is $\alpha_v = 1$ corresponding to minimize $\operatorname{tr}(Y L^{(v)} Y^T)$ over different feature views, and $\alpha_v = 0$ otherwise. This solution means that only one best feature is selected, which does not meet our objective. Thus the method utilized in [50,5,6] is adopted for this problem, α_i is set as α_i^r with $r > 1$. Therefore, each feature would has a particular contribution for final Y. From (1), we get the improved objective function:

$$\begin{aligned} & \min_{W,\alpha} \operatorname{tr}\left(W^T X \sum_{v=1}^M \alpha_v^r L^{(v)} X^T W \right) \\ & \text{s.t.} \sum_{v=1}^M \alpha_v = 1, \quad \alpha_v \geq 0, \quad W \in \{0,1\}^{D \times d} \end{aligned} \tag{19}$$

All of above shows how we are going to preserve the local information from each feature. Moreover, we are also going to keeps the global information at same time. Inspired by

PCA, we get

$$\min_{W} \operatorname{tr}(W^T X H_X X^T W) \tag{20}$$

Therefore, the final objective function is written in a race ratio minimization form, which is shown below:

$$\min_{W,\alpha} \frac{\operatorname{tr}(W^T X \sum_{v=1}^{M} \alpha_v^r L^{(v)} X^T W)}{\operatorname{tr}(W^T X H_X X^T W)}$$
$$\text{s.t.} \quad \sum_{v=1}^{M} \alpha_v = 1, \quad \alpha_v \geq 0, \quad W \in \{0,1\}^{D\times d} \tag{21}$$

3.2. Optimization method

Inspired by [51], we develop an iterative approach to solve the optimization of our trace ratio represented objective function. It is clearly a nonlinearly constrained optimization problem. The convergence proof of proposed optimization method is omitted due to paragraph limitation, where similar proof can be found in [50,5,36]. The feature weight vector α is firstly initialized as $\alpha_v = 1/M$, W is set with a random selected matrix, then W and α will be iteratively updated individually, while the other variable holding constant. Detail of the procedure is shown below.

Algorithm 1. Optimize W with fixed α.

```
Initialize feature weight α:
    $\alpha_v = 1/M$;
Local and global structure learning:
    set $A = X(\sum_{v=1}^{M} \alpha_v^r L^{(v)})X^T, B = XH_XX^T$ ;
Solving trace ratio
    $\lambda = \operatorname{tr}(W^T AW)/\operatorname{tr}(W^T BW)$ ;
Calculate score for each feature element
    $sc_i = w_i^T(A-\lambda B)w_i, w_i = [0_{i-1}, 1, 0_{D-i}], i = 1, \dots, D$;
Select relevant feature elements to update W
    Sort components according to $sc_i$ in descent order, update $W$
    with the first $d$ components
Repeat until convergence, output W
```

While W is fixed, the objective function is only related to α.

$$\min_{\alpha} \quad \operatorname{tr}\left(W^T X \sum_{v=1}^{M} \alpha_v^r L^{(v)} X^T W\right)$$
$$\text{s.t.} \quad \sum_{v=1}^{M} \alpha_v = 1, \alpha_v \geq 0 \tag{22}$$

For the objective function shown above, we apply the Langrange multiplier to solve it. Following the similar procedure in [50], finally we can get

$$\alpha_v = \frac{\left(\dfrac{1}{\operatorname{tr}(W^T X L^{(v)} X^T W)}\right)^{1/(r-1)}}{\sum_{v=1}^{M}\left(\dfrac{1}{\operatorname{tr}(W^T X L^{(v)} X^T W)}\right)^{1/(r-1)}} \tag{23}$$

Algorithm 2. Optimize α with fixed W.

```
for $v = 1$ to $M$ do
        $f_v = \operatorname{tr}(W^T XL^{(v)} X^T W)^{-1/(r-1)}$;
end for
$F = \sum_{v=1}^{M} f_v$
    for$v = 1$ to $M$ do
    $\alpha_v = f_v/F$;
end for
output $\alpha$
```

The updating of W and α should be recursively continued until it achieves convergence. Then, the local minimum solution for the selection matrix W and feature weight α can be obtained.

For Eq. (21), it needs to be pointed out that the parameter r can modulate the smoothness difference between graphes [42]. If $r \to 1$, the difference between each graph would be expanded, only α_v of the smoothest graph would be close to 1, which is the "select best feature" which is mentioned at the explanation of Eq. (18). If $r \to \infty$, the effect of difference is reduced and α_v for each feature would be close to each other. Thus, the choice of parameter r depends on the complementation of the multi-view features. The value of r should be enlarged to explore the synergistic effect of multi-view features while rich complementation exists. Otherwise r should be small to keep the performance of the "best" feature.

4. Experiment

4.1. Experiment setup

4.1.1. Experimental data preparation

The proposed AMFS method has been tested on a commonly used public mocap database HDM05 dataset [52] and MSR-Action 3D (MSR3D) dataset [53] to prove the efficiency of AMFS for motion retrieval. HDM05 dataset consists 2061 motion sequences belonging to 130 classes performed by 5 actors. Since many of original classes are close to each other, e.g. *punchLFront1Reps and punchLFront2Reps* we have combined similar action classes and finally generate 25 classes for 2061 motion clips. To further challenge our method, we choose the training and testing action sets that performed by different actors, e.g. 2 for training and 3 for testing. MSR3D has 567 motion sequences that belongs to 20 action types with 10 actors, while each actor performs each action 2 or 3 times. The 20-joints locations extracted by [54] are used instead of original depth image of MSR3D, where we randomly choose half of motion clips from each action class to be training data and the others are used as testing data.

4.1.2. Feature selection

At the beginning of AMFS, different kinds of features should be chosen to form the multi-feature representation. The power of AMFS is based on the complementarity between different types of features. During our experiments, we find that increasing features of similar type makes little effect on the final retrieval performance. Thus, we decided to choose only one typical sequence based feature HOD [28] and one pose based feature GPF [4] that both specially developed for motion data. Besides, for GPF, after feature extraction, a codebook is learnt via k means, then bag of visual word (BOVW) method *LLC* [55] is applied to generate a fixed length GPF-BOVW feature representation.

4.1.3. Performance evaluation and comparison

In this experiment, firstly, we compare the performance of our AMFS with the original two features on HDM05 dataset and MSR3D dataset. To exclude the factor of classification/ ranking method, the simple Euclidean distance is used to rank the result for all the methods. Moreover, a retrieval performance test is also taken on both two datasets. It examines the performance of AMFS and two basic features. Moreover, several existing multi-view feature combination methods, Laplacian Score [34], Feature Ranking [35] and Unsupervised Discriminative Feature Selection (UDFS) [36,56], are also implemented as comparison. At last, we also test AMFS's parameter sensitivity and examined its convergence speed.

4.2. Performance experiment result

In Fig. 2, it provides the retrieval result for two example query motion with 4 methods: AMFS, HOD, GPF-BOVW and Laplacian score. It shows the top three ranking results for every query motion given by each methods. AMFS generally outperforms the competitors.

The results in Fig. 3 show that our AMFS consistently outperforms using single features individually on both HDM05 and MSR3D datasets. The overall precision and detail performance on single action class provide the evidence that

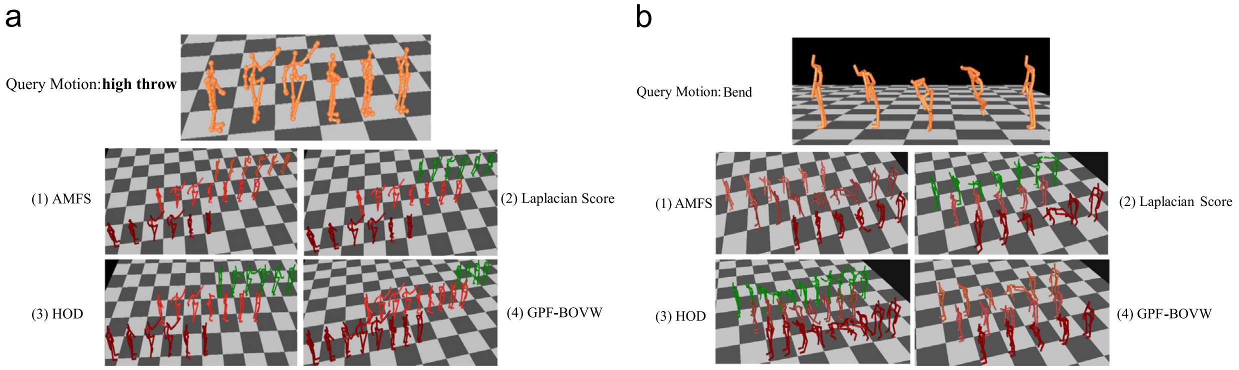

Fig. 2. Example of motion retrieval results with AMFS, Laplacian score, HOD and GPF-BOVW. The top three ranking results (from bottom to top) are given by each method, where red stands for the same action class with input and green stands for different action class. (a) High throw and (b) bend. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this article.)

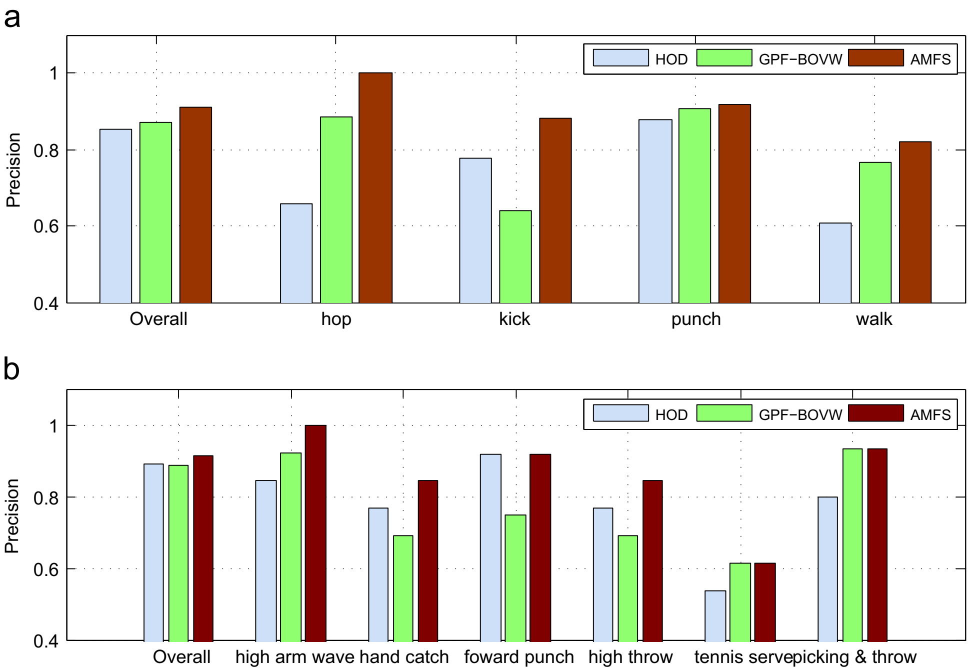

Fig. 3. Performance comparison of our AMFS algorithm and single features on HDM05 and MSR Action3D. Overall accuracy and detail result of some example action classes are shown. Actions: (a) HDM05 *hop, kick, punch, walk* and (b) MSR Action3D *high arm wave, hand catch, forward punch, high throw, tennis serve, pick up and throw.*

AMFS can combine the original features synergistically. The comparison of AMFS, and other multi-feature combination methods shows that AMFS is a effective method to generate relevance feature subset for human motion retrieval.

Fig. 4 shows the retrieval performance of AMFS, two single feature algorithms and three multi-feather combination methods on two datasets. AMFS generally outperforms other methods. The comparison of AMFS and the competitors shows that AMFS is benefited from the complementary of different features, which provides a better achievement.

4.3. Algorithm parameter sensitivity

The result of the sensitivity test for parameter μ and r is reported in Fig. 5, where Fig. 5a is the test on HDM05 dataset and Fig. 5b is the test on MSR Action3D dataset. The variance range of parameter r is from 2 to 16 and the variance range of μ is $[0.01, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5]$. The result demonstrates that our AMFS is robust to the parameter changing, meanwhile, it can still achieve a high retrieval accuracy even the parameter is not initialized carefully. Besides, the effect of parameter k, which is the parameter used to build the feature local graph, has also been tested. The result is shown in Fig. 6.

As we mentioned in Section 3.2 that the objective function is solved using an iterative approach, the speed of reaching convergence is crucial for the computational efficiency in the real world motion retrieval. The result of convergence examination on HDM05 dataset is given in Fig. 7. We noticed that AMFS can reach convergence within 20 iterations. Thus, it demonstrates that our optimization approach for AMFS is efficient.

4.4. Discussions

Generally the AMFS improves the motion retrieval performance. It simultaneously discovers the intrinsic relation

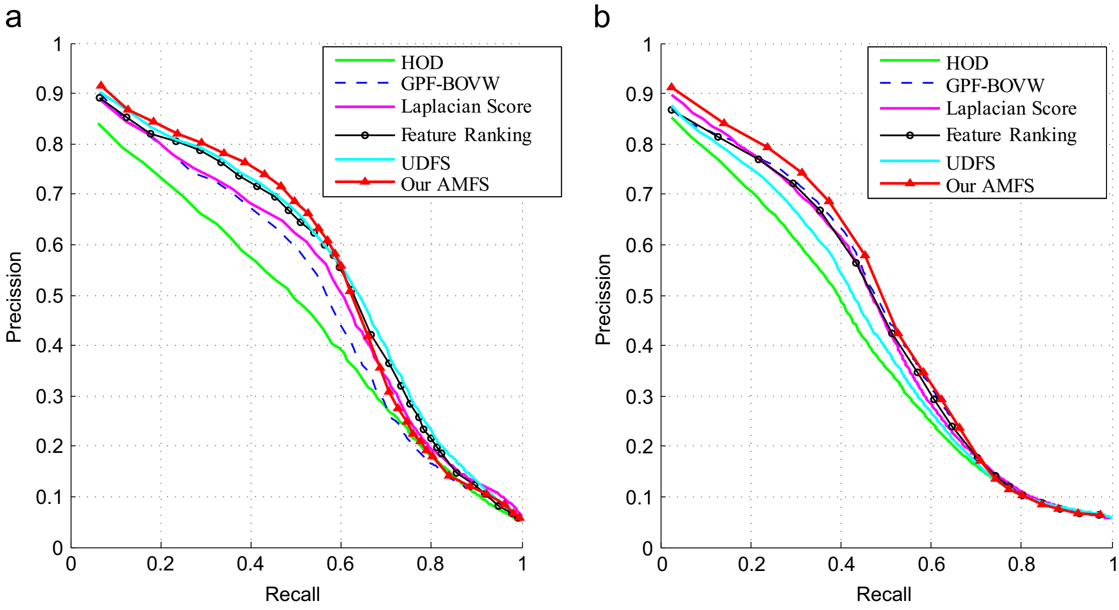

Fig. 4. Retrieval performance of our AMFS algorithm: (a) test on HDM05 dataset and (b) test on MSR3D dataset.

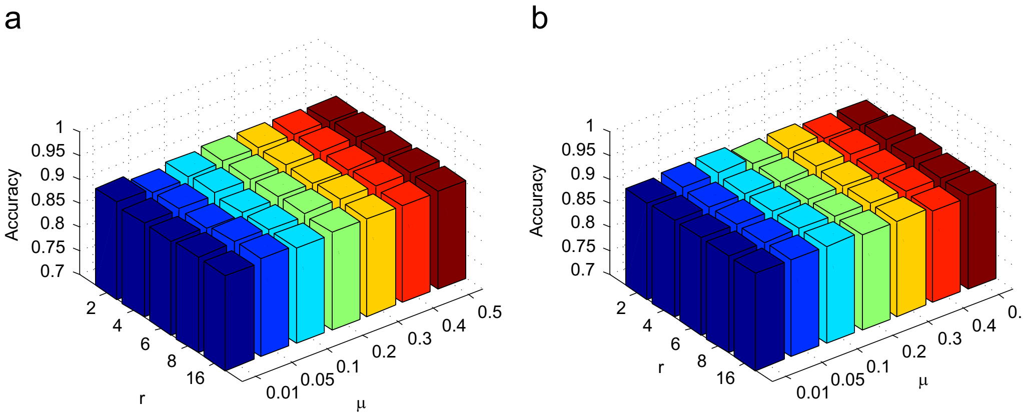

Fig. 5. Performance variations of our AMFS algorithm with different parameter r and μ: (a) test on HDM05 dataset and (b) test on MSR3D dataset.

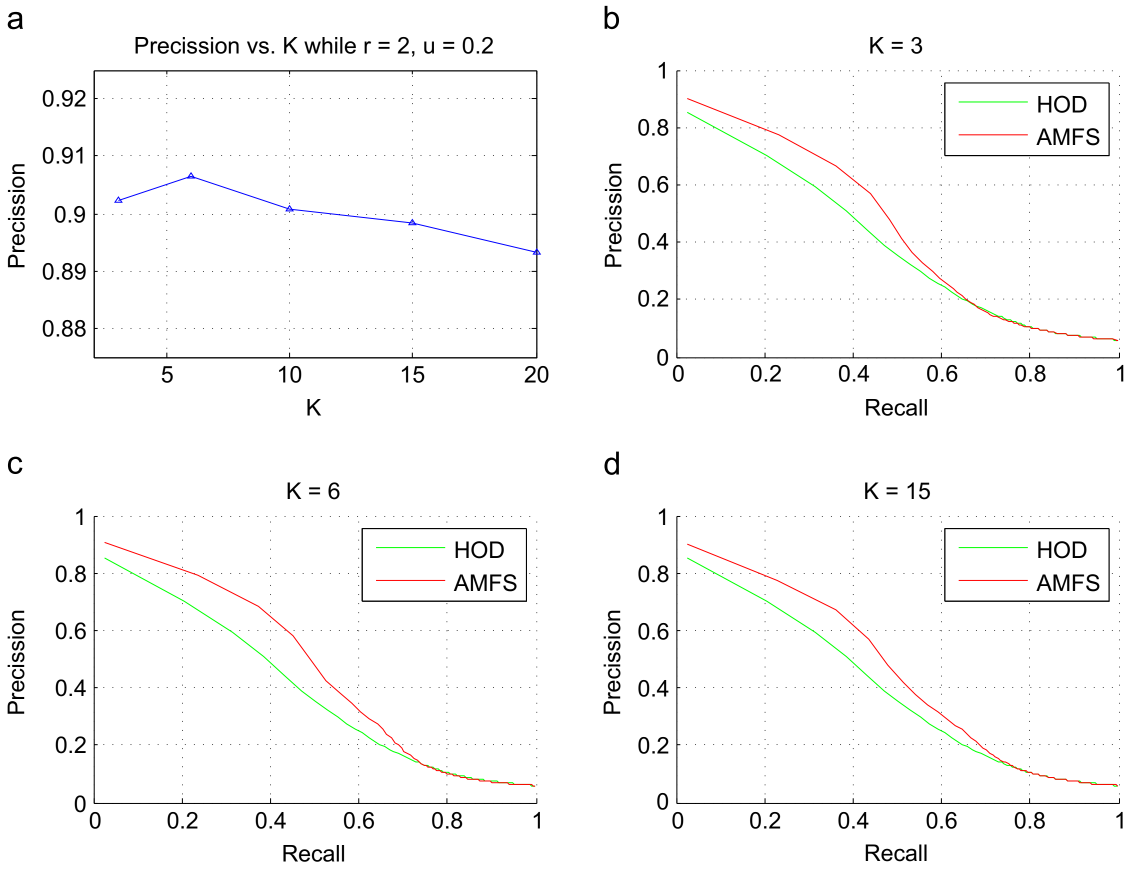

Fig. 6. Parameter sensitivity test for k on HDM05 dataset.

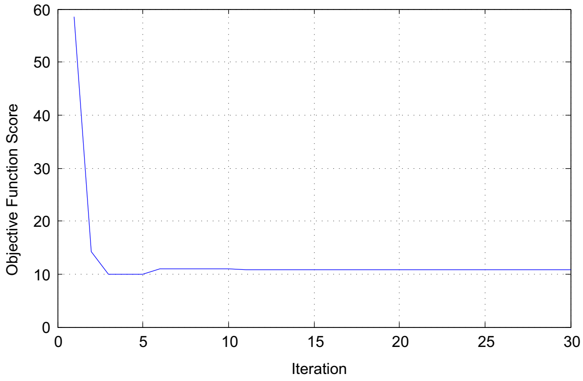

Fig. 7. Illustration of the convergence of AMFS on HDM05 dataset.

between visual categories in a compact and relevance selected feature subspace by leveraging the underlying data structure and similarity between different feature views. It is fast, insensitive to parameters and robust to large scale data.

Comparing with previous work on motion features that only using small subset motion data, e.g. 251 actions from HDM05 [15,28], the whole HODM05 dataset is employed for performance testing. The performance on the larger dataset shows that our AMFS is able to handle the challenge from real-world big data. Besides, AMFS can directly be used if a new effective motion feature is proposed in the future. Any new motion features can be easily integrated into our framework. All of these properties make it applicable for real world problem solving.

However, one thing still need to be pointed out is that, as indicated in [37], AMFS cannot guarantee the higher performance from single features if the data does not hold a manifold structure. Moreover, we find that added similar type of features will bring little effect to the performance. Besides, fluctuation could occur during experiment when the number of features increased, this is a problem need to be solved in future.

5. Conclusion

In this paper, a multi-view feature selection method has been proposed to categorize motion data. We build a trace-ratio objective function and provide an iterative optimization approach. The AMFS discovers the intrinsic relations between visual words in each motion feature subspace to improve the retrieval performance. The performance of AMFS method has been evaluated on the HDM05 and MSR3D motion datasets. In order deal with the real-world problem, all the 2061 actions in HDM05 dataset are employed for performance test, and we choose different actors of motion for training and testing. The experimental results show that the AMFS method can improve the performance by combining different motion features synergistically and it has potential to be implemented in the real large motion dataset retrieval.

Acknowledgements

This research is supported by the grant of the Sino-UK Higher Education Research Partnership for Ph.D. Studies Project funded by the Department of Business, Innovation and Skills of the British Government and Ministry of Education of China, the Zhejiang Provincial Natural Science Foundation of China (LY13F020001) and Zhejiang Province Public Technology Applied Research Projects (No. 2014C33090).

References

[1] Z. Zhang, Microsoft kinect sensor and its effect, IEEE Multimed. 19 (2) (2012) 4–10.

[2] H.P. Shum, E.S. Ho, Y. Jiang, S. Takagi, Real-time posture reconstruction for microsoft kinect, IEEE Trans. Cybern. 43 (5) (2013) 1357–1369.

[3] M. Müller, T. Röder, Motion templates for automatic classification and retrieval of motion capture data, in: Proceedings of the 2006 ACM SIGGRAPH/Eurographics Symposium on Computer Animation, Eurographics Association, 2006, pp. 137–146.

[4] C. Chen, Y. Zhuang, F. Nie, Y. Yang, F. Wu, J. Xiao, Learning a 3D human pose distance metric from geometric pose descriptor, IEEE Trans. Vis. Comput. Graph. 17 (11) (2011) 1676–1689.

[5] T. Xia, D. Tao, T. Mei, Y. Zhang, Multiview spectral embedding, IEEE Trans. Syst. Man Cybern. Part B, Cybern.: A Publ. IEEE Syst. Man Cybern. Soc. 40 (6) (2010) 1438–1446.

[6] Y. Feng, J. Xiao, Y. Zhuang, X. Liu, Adaptive unsupervised multi-view feature selection for visual concept recognition, in: Computer Vision–ACCV 2012, Springer, 2013, pp. 343–357.

[7] J.K. Tang, H. Leung, Retrieval of logically relevant 3D human motions by adaptive feature selection with graded relevance feedback, Pattern Recognit. Lett. 33 (4) (2012) 420–430.

[8] M. Wang, X.-S. Hua, Active learning in multimedia annotation and retrieval: a survey, ACM Trans. Intell. Syst. Technol. (TIST) 2 (2) (2011) 10.

[9] M. Wang, R. Hong, G. Li, Z.-J. Zha, S. Yan, T.-S. Chua, Event driven web video summarization by tag localization and key-shot identification, IEEE Trans. Multimed. 14 (4) (2012) 975–985.

[10] L. Zhang, Y. Han, Y. Yang, M. Song, S. Yan, Q. Tian, Discovering discriminative graphlets for aerial image categories recognition, IEEE Trans. Image Process.: A Publ. IEEE Signal Process. Soc. 22 (12) (2013) 5071–5084.

[11] L. Zhang, M. Song, X. Liu, J. Bu, C. Chen, Fast multi-view segment graph kernel for object classification, Signal Process. 93 (6) (2013) 1597–1607.

[12] L. Zhang, M. Song, X. Liu, L. Sun, C. Chen, J. Bu, Recognizing architecture styles by hierarchical sparse coding of blocklets, Inf. Sci. 254 (2014) 141–154.

[13] L. Zhang, Y. Yang, Y. Gao, Y. Yu, C. Wang, X. Li, A probabilistic associative model for segmenting weakly-supervised images, IEEE Trans. Image Process. (T-IP) 23 (9) (2014) 4150–4159.

[14] L. Zhang, Y. Gao, Y. Xia, K. Lu, J. Shen, R. Ji, Representative discovery of structure cues for weakly-supervised image segmentation, IEEE Trans. Multimed. 16 (2) (2014) 470–479.

[15] F. Ofli, R. Chaudhry, G. Kurillo, R. Vidal, R. Bajcsy, Sequence of the Most Informative Joints (SMIJ): a new representation for human skeletal action recognition, in: CVPR Workshops, IEEE, 2012, pp. 8–13.

[16] L. Zhang, M. Song, Q. Zhao, X. Liu, J. Bu, C. Chen, Probabilistic graphlet transfer for photo cropping, IEEE Trans. Image Process. 22 (2) (2013) 802–815.

[17] C. Böhm, S. Berchtold, D.A. Keim, Searching in high-dimensional spaces: index structures for improving the performance of multimedia databases, ACM Comput. Surv. (CSUR) 33 (3) (2001) 322–373.

[18] E. Keogh, T. Palpanas, V.B. Zordan, D. Gunopulos, M. Cardle, Indexing large human-motion databases, in: Proceedings of the Thirtieth International Conference on Very Large Data Bases, vol. 30, VLDB Endowment, 2004, pp. 780–791.

[19] B. Krüger, J. Tautges, A. Weber, A. Zinke, Fast local and global similarity searches in large motion capture databases, in: Proceedings of the 2010 ACM SIGGRAPH/Eurographics Symposium on Computer Animation, Eurographics Association, 2010, pp. 1–10.

[20] A. Yoshitaka, T. Ichikawa, A survey on content-based retrieval for multimedia databases, IEEE Trans. Knowl. Data Eng. 11 (1) (1999) 81–93.

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

[22] I. Kapsouras, N. Nikolaidis, Action recognition on motion capture data using a dynemes and forward differences representation, J. Vis. Commun. Image Represent. 25 (6) (2014) 1432–1445.

[23] T. Qi, J. Xiao, Y. Zhuang, H. Zhang, X. Yang, J. Zhang, Y. Feng, Real-time motion data annotation via action string, Comput. Anim. Virtual Worlds 25 (3–4) (2014) 293–302.

[24] M. Barnachon, S. Bouakaz, B. Boufama, E. Guillou, Ongoing human action recognition with motion capture, Pattern Recognit. 47 (1) (2014) 238–247.

[25] L. Zhou, Z. Lu, H. Leung, L. Shang, Spatial temporal pyramid matching using temporal sparse representation for human motion retrieval, Vis. Comput. 30 (6–8) (2014) 845–854.

[26] J. Wang, Z. Liu, Y. Wu, J. Yuan, Learning actionlet ensemble for 3D human action recognition, IEEE Trans. Pattern Anal. Mach. Intell. 36 (5) (2014) 914–927.

[27] M. Müller, A. Baak, H.-P. Seidel, Efficient and robust annotation of motion capture data, in: Proceedings of the 2009 ACM SIGGRAPH/Eurographics Symposium on Computer Animation, ACM, 2009, pp. 17–26.

[28] M.A. Gowayyed, M. Torki, M.E. Hussein, M. El-Saban, Histogram of oriented displacements (HOD): describing trajectories of human joints for action recognition, in: Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence, AAAI Press, 2013, pp. 1351–1357.

[29] S. Wu, Y. Li, J. Zhang, A hierarchical motion trajectory signature descriptor, in: IEEE International Conference on Robotics and Automation (ICRA), IEEE, 2008, pp. 3070–3075.

[30] J. Yang, Y. Li, K. Wang, A new descriptor for 3D trajectory recognition via modified CDTW, in: IEEE International Conference on Automation and Logistics, IEEE, 2010, pp. 37–42.

[31] Q. Xiao, J. Li, Y. Wang, Z. Li, H. Wang, Motion retrieval using probability graph model. Computational Intelligence and Design (ISCID), 2013 Sixth International Symposium on (book title), IEEE 2 (2013) 150–153.

[32] M.-W. Chao, C.-H. Lin, J. Assa, T.-Y. Lee, Human motion retrieval from hand-drawn sketch, IEEE Trans. Vis. Comput. Graph. 18 (5) (2012) 729–740.

[33] M.G. Choi, K. Yang, T. Igarashi, J. Mitani, J. Lee, Retrieval and visualization of human motion data via stick figures, in: Computer Graphics Forum, vol. 31, Wiley Online Library, 2012, pp. 2057–2065.

[34] X. He, D. Cai, P. Niyogi, Laplacian score for feature selection, Adv. Neural Inf. Process. Syst. (2005) 507–514.

[35] Z. Zhao, H. Liu, Spectral feature selection for supervised and unsupervised learning, in: Proceedings of the 24th International Conference on Machine learning, ACM, 2007, pp. 1151–1157.

[36] Y. Yang, H.T. Shen, Z. Ma, Z. Huang, X. Zhou, l2, 1-Norm regularized discriminative feature selection for unsupervised learning, in: IJCAI Proceedings of the International Joint Conference on Artificial Intelligence, vol. 22, 2011, p. 1589.

[37] Y. Yang, J. Song, Z. Huang, Z. Ma, N. Sebe, A.G. Hauptmann, Multi-feature fusion via hierarchical regression for multimedia analysis, IEEE Trans. Multimed. 15 (3) (2013) 572–581.

[38] S. Wang, Z. Ma, Y. Yang, X. Li, C. Pang, A.G. Hauptmann, Semi-supervised multiple feature analysis for action recognition, IEEE Trans. Multimed. 16 (2) (2014) 289–298.

[39] Y. Han, F. Wu, D. Tao, J. Shao, Y. Zhuang, J. Jiang, Sparse unsupervised dimensionality reduction for multiple view data, IEEE Trans. Circuits Syst. Video Technol. 22 (10) (2012) 1485–1496.

[40] Y. Han, Y. Yang, Y. Yan, Z. Ma, N. Sebe, X. Zhou, Semisupervised feature selection via spline regression for video semantic recognition, IEEE Trans. Neural Netw. Learn. Syst. 1 (2014) p. 99.

[41] L. Zhang, Y. Gao, C. Hong, Y. Feng, J. Zhu, D. Cai, Feature correlation hypergraph: exploiting high-order potentials for multimodal recognition, IEEE Trans. Cybern. 44 (8) (2013) 1408–1419.

[42] M. Wang, X.-S. Hua, R. Hong, J. Tang, G.-J. Qi, Y. Song, Unified video annotation via multigraph learning, IEEE Trans. Circuits Syst. Video Technol. 19 (5) (2009) 733–746.

[43] M. Wang, H. Li, D. Tao, K. Lu, X. Wu, Multimodal graph-based reranking for web image search, IEEE Trans. Image Process. 21 (11) (2012) 4649–4661.

[44] M. Wang, X.-S. Hua, J. Tang, R. Hong, Beyond distance measurement: constructing neighborhood similarity for video annotation, IEEE Trans. Multimed. 11 (3) (2009) 465–476.

[45] J. Yu, M. Wang, D. Tao, Semisupervised multiview distance metric learning for cartoon synthesis, IEEE Trans. Image Process. 21 (11) (2012) 4636–4648.

[46] Z. Gao, J.-M. Song, H. Zhang, A.-A. Liu, Y.-B. Xue, G.-P. Xu, Human action recognition via multi-modality information, J. Electr. Eng. Technol. 9 (2) (2014) 739–748.

[47] D. Zhang, M. Yang, X. Feng, Sparse representation or collaborative representation: which helps face recognition?, in: 2011 IEEE International Conference on Computer Vision (ICCV), IEEE, 2011, pp. 471–478.

[48] S.T. Roweis, L.K. Saul, Nonlinear dimensionality reduction by locally linear embedding, Science 290 (5500) (2000) 2323–2326.

[49] Y. Yang, F. Nie, D. Xu, J. Luo, Y. Zhuang, Y. Pan, A multimedia retrieval framework based on semi-supervised ranking and relevance feedback, IEEE Trans. Pattern Anal. Mach. Intell. 34 (4) (2012) 723–742.

[50] M. Wang, X.-S. Hua, X. Yuan, Y. Song, L.-R. Dai, Optimizing multigraph learning: towards a unified video annotation scheme, in: Proceedings of the 15th International Conference on Multimedia, ACM, 2007, pp. 862–871.

[51] Y. Jia, F. Nie, C. Zhang, Trace ratio problem revisited, IEEE Trans. Neural Netw. 20 (4) (2009) 729–735.

[52] M. Müller, T. Röder, M. Clausen, B. Eberhardt, B. Krüger, A. Weber, Documentation mocap database HDM05, Technical report, No. CG-2007-2, ISSN 1610-8892, Universität Bonn, June 2007. http://resources.mpi-inf.mpg.de/HDM05/.

[53] W. Li, Z. Zhang, Z. Liu, Action recognition based on a bag of 3d points, in: 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), IEEE, 2010, pp. 9–14.

[54] J. Shotton, T. Sharp, A. Kipman, A. Fitzgibbon, M. Finocchio, A. Blake, M. Cook, R. Moore, Real-time human pose recognition in parts from single depth images, Commun. ACM 56 (1) (2013) 116–124.

[55] X. Wang, L. Wang, Y. Qiao, A comparative study of encoding, pooling and normalization methods for action recognition, in: Computer Vision–ACCV 2012, Springer, 2013, pp. 572–585.

[56] Z. Ma, F. Nie, Y. Yang, J.R. Uijlings, N. Sebe, A.G. Hauptmann, Discriminating joint feature analysis for multimedia data understanding, IEEE Trans. Multimed. 14 (6) (2012) 1662–1672.

[57] Raptis Michalis, Kirovski Darko, Hoppe Hugues, Real-time classification of dance gestures from skeleton animation, Proceedings of the 2011 ACM SIGGRAPH/Eurographics Symposium on Computer Animation, ACM (2011) 147–156, http://dx.doi.org/10.1145/2019406.2019426.