Neurocomputing 129 (2014) 585–595

Contents lists available at ScienceDirect

Neurocomputing

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

Classification and translation of style and affect in human motion using RBF neural networks

S. Ali Etemad [a,*], Ali Arya [b]

[a] Department of Systems and Computer Engineering, Carleton University, Ottawa, ON, Canada
[b] School of Information Technology, Carleton University, Ottawa, ON, Canada

ARTICLE INFO

Article history:
Received 18 November 2012
Received in revised form
9 September 2013
Accepted 22 September 2013
Communicated by Qingshan Liu
Available online 8 October 2013

Keywords:
Motion capture
Radial basis functions
Neural networks
Style translation
Classification

ABSTRACT

Human motion can be carried out with a variety of different affects or styles such as happy, sad, energetic, and tired among many others. Modeling and classifying these styles, and more importantly, translating them from one sequence onto another has become a popular problem in the fields of graphics, multimedia, and human computer interaction. In this paper, radial basis functions (RBF) are used to model and extract stylistic and affective features from motion data. We demonstrate that using only a few basis functions per degree of freedom, successful modeling of styles in cycles of human walk can be achieved. Furthermore, we employ an ensemble of RBF neural networks to learn the affective/stylistic features following time warping and principal component analysis. The system learns the components and classifies stylistic motion sequences into distinct affective and stylistic classes. The system also utilizes the ensemble of neural networks to learn motion affects and styles such that it can translate them onto neutral input sequences. Experimental results along with both numerical and perceptual validations confirm the highly accurate and effective performance of the system.

1. Introduction

Generally, human motion consists of a hierarchy of themes, each of which is composed of a different set of spatiotemporal features [1]. While the primary themes are related to specific actions like walking or running, the secondary themes relate to the *style* in which those actions are performed. They contain variations caused by individual characteristics of the actor such as gender, age, emotions, energy, mood, health, and even inherited characteristics. These stylistic variations, or secondary features, are constantly present in motion data. However, they are extremely difficult for machine learning approaches to extract and analyze due to their personalized nature, small spatiotemporal significance, and often, lack of sufficient and consistent training data.

Recently, as a direct result of developments in computational capacities, animation of human motion has received significant amount of attention, and along with it, the study of stylistic variations has become subject to extensive research [2,3]. A state of the art problem in this field is style translation (conversion), or in other words, the capability to computationally alter the secondary features of a motion sequence to achieve different secondary themes. A robust translation system can eventually be employed for synthesizing stylistic animation of characters from base sequences. For example, given one base walk sequence, a vast variety of different walks can be created which can be employed for different characters. Such systems can widely be utilized in various multimedia and human–computer interaction (HCI) applications [10], as well as human–robot interaction purposes (HRI) [11,12]. More specifically, virtual worlds and digital games can employ such systems to synthesize a variety of character animations, eliminating the need for large databases. For example, given a neutral motion sequence (a sequence not showing any particular secondary feature), a variety of stylistic sequences can be synthesized. These synthesized sequences can range from feminine, masculine, energetic, tired, young, old, healthy, and wounded, styles to more complex ones such as limping. Robotics, and robot motion in particular, is another realm which cam employ motion style synthesis. Using automated systems, robot motion can be generated and controlled without the need for predefined motion trajectories. Overall, such systems will eventually enable significant data re-use for digital and physical applications. Another field where motion style translation can be utilized is physiology and psychology studies. These systems will enable researchers to better understand the processes that our bodies and minds employ to generate and perceive different types of motion [13].

In this paper, we tackle the problem of modeling motion style features extracted from human motion data. We demonstrate that radial basis functions (RBFs) are powerful means for this purpose. We then employ this concept and train an ensemble of Gaussian RBF neural networks. Training is carried out in two different modes: the first mode acts as a classifier while the second one is utilized for translating the style features from one class onto

* Corresponding author.
E-mail addresses: ali.etemad@carleton.ca (S. Ali Etemad), arya@carleton.ca (A. Arya).

0925-2312/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.neucom.2013.09.001

another. For example, given an input neutral walk, a tired walk sequence is achieved. The motion data, prior to application in the system, require two critical pre-processing steps: time warping and dimensionality reduction. These modules are described in detail. Finally, multiple experimental results with high classification and style translation accuracy, as well as significant generalization capability, demonstrate the effectiveness of our approach towards modeling stylistic features.

2. Related work

Here we review the different methods used for recognition and translation of motion styles. Generally, on the topic of action recognition, there has been significant amount of work done such as [14–16], among many others. While these processes focus on classification of the actions rather than actor attributes (which is the goal of this research), they can be applied if modified or trained differently. We then review some of the literature focused on style recognition. Finally, works on style translation are reviewed.

K-nearest neighbor (KNN) classifiers have been employed, successive to dimensionality reduction [17], other pre-processing techniques [18], or using the raw data [19]. Through such approaches, a distance measure of the test action sequence with respect to the training set is often computed and minimized. Support vector machines (SVM) have shown to be another effective method [20]. Through these method, the feature space is divided to classify the different action classes using a vector or plane supported by a number of vectors. Relevance vector machines (RVM), which are a modified probabilistic version of SVM, have also been successfully utilized for this purpose [21]. Hidden Markov models (HMM) and other probabilistic means have been widely used to model and classify actions. Hidden states are used in HMMs to model different states of the motion sequence. As an example, [22] is one of the works that have utilized these models for action classification. Finally, artificial neural networks (ANN) have been largely ignored for action classification. In one of the few approaches using ANNs, in [23] a self-organizing neural model is employed. In [24], optical flow-based motion history maps were utilized to train a back-propagation multilayer perceptron (MLP) for action classification.

Classification of styles using motion capture data has not been widely addressed in the literature. This may be in part due to such data being scarce and expensive to record or purchase compared to video motion data. The CMU dataset (http://mocap.cs.cmu.edu/) does contain many motion capture sequences, and it has been used for classification of actions. However, motion capture datasets that contain many *repeated* takes of stylistic actions, which can be used for classification of style and affect, are not widely and readily available. As a result, to the best of our knowledge, there are not many studies of style using publically available datasets to which we can compare our results. Table 1 presents some recent studies on style/affect using motion capture data. These studies do not use publicly available datasets. In [26,27], the utilized method based on PCA and linear discriminant functions (LDF), performs gender classification with an accuracy of 92.7%. Venture [45] uses inverse kinematics (IK) and similarity criteria to categorize neutrality, joy, anger, and sadness. Karg et al. [46,47] employ linear discriminant analysis (LDA), PCA, and KPCA for feature extraction prior to classifying neutral, angry, sad, and happy walks using KNN, Naive Bayes, and SVM. Based on the feature extraction method and classifier used, a wide range of accuracies from 40% to 100% is achieved. Livne et al. [48], use transfer learning and linear regression to classify gender and weight from motion capture data and video pose tracking successive to PCA and Fourier transform (FT). In this study, an accuracy of up to 98% is acquired.

Most methods proposed for style synthesis and translation result in a particular set of style features being transferred from an origin sequence onto a destination sequence. This approach often involves relative editing of sequences using linear operations in the form of interpolation and extrapolation [4], successive to proper temporal alignment of critical features, which is usually achieved through dynamic time warping (DTW) [5]. DTW maximizes the alignment between two trajectories, minimizing a relative distance function. This will ensure that style features are extracted from and added to the correct postures of the sequence. More advanced methods such as probabilistic models [6], system identification processes [7], and database methods [8] have also been introduced and implemented for style translation and synthesis. Signal processing techniques, such as [9], on the other hand, provide us with a better understanding of how affective/stylistic features are added onto regular motion signals. For example, in [28] we proposed a technique which employs frequency minimization of temporal cues for reconstructing the neutral component of stylistic motion trajectories with spatiotemporal cubic splines. This work was based on the assumption that stylistic features appear as high frequency add-ons. Other frequency-based techniques have also been utilized in the past. In [9], Bruderlin and Williams conclude that when filtering motion trajectories, lower frequency gains are manifested as decreases in intensity of performed actions. They show that middle band frequency gains result in exaggerated movements and finally by increasing higher frequencies, nervous twitches are synthesized. In [29] it was shown that the high-frequency components can depend on low-frequency ones. Pullen and Bregler [30], later used motion capture data to add texture to keyframed animation. The model utilizes correlations among separate body parts. Furthermore, their proposed technique adds mid and high-level frequency alterations to keyframed or synthesized signals through texturing. Finally, learning-based methods have also been utilized for style translation. For example, Grochow et al. [49] use Gaussian processes and Liu et al. [50] utilize inverse optimization as tools for learning motion styles.

In style translation, a common characteristic among the aforementioned techniques, despite their high accuracy and practicality, is high dependence on the training data with little or often no generalization. For example, when interpolating/extrapolating sequences, the extracted features are extremely reliant on the source and target sequences. In other words, when using traditional techniques, the features translated onto the neutral sequences are directly those of the stylistic sequence used in the process. Similar arguments are true for most other available style synthesis and translation systems where the inability to dynamically synthesize

Table 1
Studies focusing on classification of style using motion capture data.

Authors and reference	Classified feature	Method
Troje, [26,27]	Gender	LDF (+PCA, FT)
Venture [45]	Affect	IK similarity
Karg et al. [46,47]	Affect	KNN, Naive Bayes, SVM (+PCA, KPCA, LDA)
Livne et al. [48]	Gender, weight	Transfer learning and linear regression (+PCA, FT)
Our method	Affect, age, energy	KNN, SVM, RBFNN (+PCA)

the style features has so far been one of the shortcomings. Additionally, lack of a single reliable tool capable of performing both classification and synthesis of styles, is evidently a problem yet to be solved.

3. Background and data

According to [1], a sequence of human motion can be represented by a sum of a primary action, weighted sum of secondary features (SF), and random noise. This model, therefore, can be denoted by:

$$\mathbf{Y} = \mathbf{P} + \sum_{i=1}^{r} \mathbf{w}_i \times \mathbf{S}_i + \mathbf{e}\,, \tag{1}$$

where $\mathbf{P}$ represents the primary action (main action class) and $\mathbf{S}$ is the set of SFs, $\mathbf{W}$ is the set of weights r members, and $\mathbf{e}$ represents random noise in the data. For simplification purposes, in the course of this study, we assume that $\mathbf{S}$ and $\mathbf{w}$ only have one member ($r=1$) and therefore combinational styles such as young-tired or energetic-feminine are not taken into account. Subsequently $\mathbf{w}$ will equal to 1 after normalization of the weight. Accordingly, from Eq. (1), we have $\mathbf{Y}=\mathbf{P}+\mathbf{S}+\mathbf{e}$.

Assuming sequence 1 with $\mathbf{Y}_1=\mathbf{P}_1+\mathbf{S}_1+\mathbf{e}_1$, sequence 2 with $\mathbf{Y}_2=\mathbf{P}_2+\mathbf{S}_2+\mathbf{e}_2$, and the two primary action classes being identical ($\mathbf{P}_1=\mathbf{P}_2$), we can conclude $\Delta\mathbf{Y}=(\mathbf{S}_2-\mathbf{S}_1)+(\mathbf{e}_2-\mathbf{e}_1)$. By selecting one of the sequences to be stylistically neutral ($\mathbf{S}_1=0$) and through minimizing the noise in the data, the SF set of the second sequence can be extracted. Adding this to a different neutral (or even stylistic sequence) will result in style translation. Similarly, interpolation of two sequences with similar primary actions and different secondary themes gives $(1-\alpha)\mathbf{Y}_1+\alpha\mathbf{Y}_2=\mathbf{P}_1+(1-\alpha)\mathbf{Y}_1+\alpha\mathbf{Y}_2+(1-\alpha)\mathbf{e}_1+\alpha\mathbf{e}_2$. By minimizing the noise, we have blended the secondary themes.

We used a Vicon MX40 motion capture system to record some of the data. The motion capture system tracks the exact locations of light-reflecting markers placed on a special suit using high spatiotemporal infrared cameras. Multiple walking sequences were performed by 5 different actors and in different styles (secondary themes). Actors were first asked to perform (act) young and old style walks. Additionally, energetic sequences were captured at the beginning of each session. Successive to all the walking performances as well as additional exercise, tired walks were performed and recorded. As illustrated in Fig. 1, the system then generates a model for the corresponding actor and outputs the motion matrix. The model illustrated in this figure is visualized using the WX Motion Viewer (http://cgg.mff.cuni.cz/~semancik/research/wxmv/). In addition to our own recorded data, we used the HDM05 [44] dataset (http://www.mpi-inf.mpg.de/resources/HDM05/) to test our system. Similar to our dataset, this HDM05 contains many motion capture data recorded using a marker-based motion capture system. It contains multiple actions performed by 5 actors. Each actor has performed several regular, happy, and sad walk cycles, each containing four or more steps, among other walks such as sideways, left/right turns, and etc.

Motion capture data can be represented by a number of consecutive postures variable with time. We represent each posture with a finite number of markers corresponding to different regions of the body. The motion matrix can either be characterized through the location of markers at each frame or instance of time, joint angles, or through other means. In other words,

$$\mathbf{D} = [\mathbf{p}^1\ \mathbf{p}^2 \cdots \mathbf{p}^m]^T, \tag{2}$$

where $\mathbf{D}$ is the motion matrix. Accordingly, $\mathbf{p}$ represents each posture with $\mathbf{p}\in\mathbb{R}^{3l}$ where l is the number of markers representing each posture in three dimensions ($l\in\mathbb{N}$). m is the number of frames and $m\in\mathbb{N}$. The ith posture $\mathbf{D}_i$ can be represented by $\mathbf{p}^i=[\theta^i_{1,x}\ \theta^i_{1,y}\ \theta^i_{1,z}\cdots\theta^i_{l,x}\ \theta^i_{l,y}\ \theta^i_{l,z}]$, where θ is the joint angle, with $\theta\in\mathbb{R}$ and $0\le\theta<360$. The trajectory of the jth degree of freedom (DOF) is described by $\theta_j=[\theta^1_j\ \theta^2_j\cdots\theta^m_j]^T$. Also, a displacement vector $\mathbf{d}$ is defined where $\mathbf{d}\in\mathbb{R}^3$, to indicate the center mass of the actor in 3-dimensional Cartesian space at each time step. We can hereby conclude that a motion sequence with l markers can be represented by $3l+3$ trajectories or DOFs.

The HDM05 dataset contains excess DOFs with respect to our data, which belong to fingers, toes, and other insignificant joints. We first remove these extra joints from the model and data. We then segment the sequences containing regular forward walks in different styles. The segmentation is very influential in the process, as they will affect training of the ANNs. Segments are carefully selected such that they start and end in similar postures. A total of 48 segmented sequences are achieved in neutral, happy, and sad themes (16 in each class).

4. Modeling secondary features using Gaussian RBFs

A radial function $\phi:\mathbb{R}^s\to\mathbb{R}$ is defined by $\phi(t)=\varphi(r)$ where $r=\|t\|$ given $\varphi:[0,\infty)\to\mathbb{R}$ is a univariate function and $\|\cdot\|$ is a norm operator such as the Euclidean norm. Accordingly we define a Gaussian radial kernel $\varphi:\mathbb{R}^s\times\mathbb{R}^s\to\mathbb{R}$, by

$$\varphi(t;\mu,\sigma^2)=\phi(\|t-\mu\|)=\exp\left\{\frac{-\|t-\mu\|^2}{2\sigma^2}\right\}, \tag{3}$$

where μ and σ^2 denote the mean variance respectively. Hence, the kth dimension of the SF set, can be approximated by:

$$\sum_{j=0}^{M}\alpha_{k,j}\varphi_j(t;\mu_j,\sigma_j^2), \tag{4}$$

where $\boldsymbol{\alpha}$ is the amplitude of each RBF used to model each DOF and M is the number of RBFs used for each DOF. Fig. 2 illustrates SF

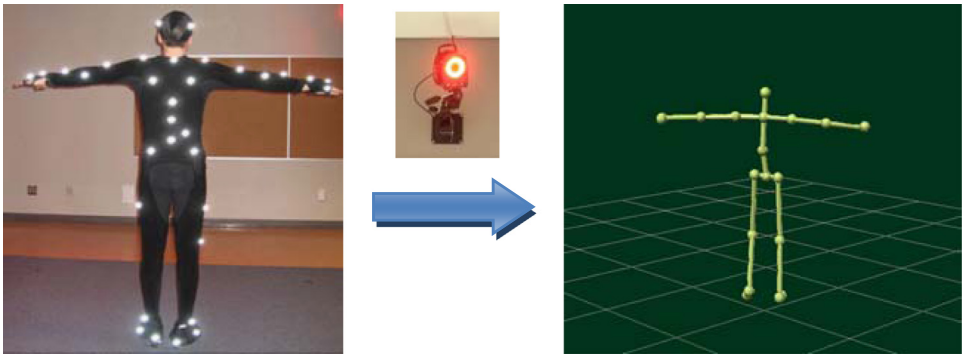

Fig. 1. A motion capture session is presented where a 3D model is created using the system.

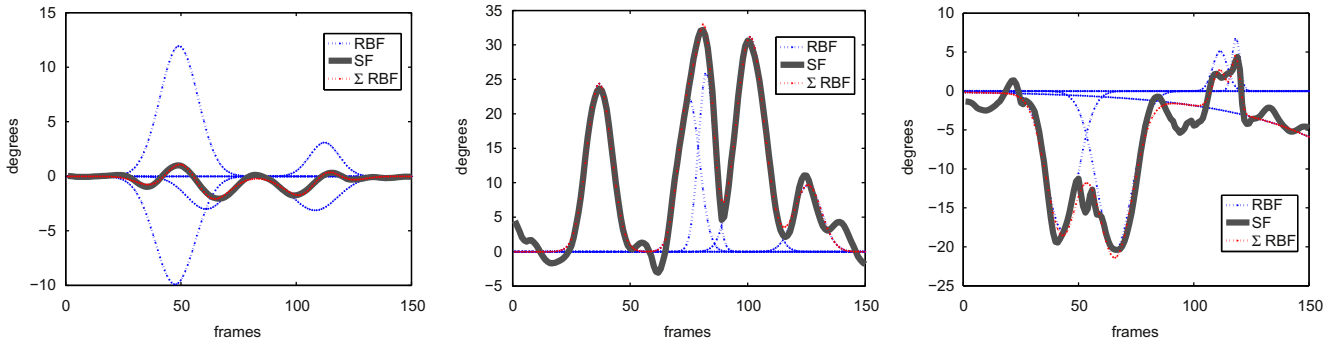

Fig. 2. Three different SF trajectories of an energetic walk approximated using 5 RBFs.

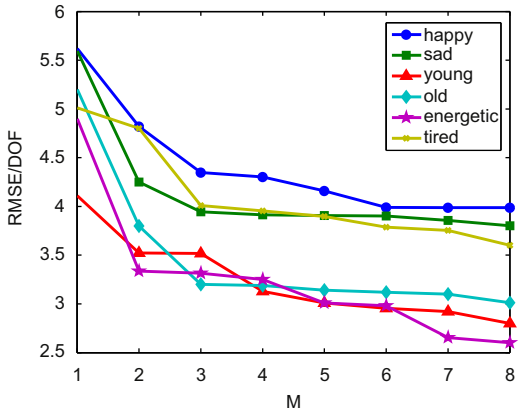

Fig. 3. The average RMSE/DOF vs. number of RBFs used to approximate the SF sets of 15 happy, sad, young, old, energetic, and tired walking sequences.

trajectories of the 2nd, 12th, 27th, DOF of an energetic walk, modeled with a weighted sum of 5 Gaussian RBFs ($M=5$). The linear least squares method has been used to decompose the trajectories into the basis functions. We extracted the SF set using linear subtraction of an energetic walk from a neutral walk successive to time warping. The warping details are presented in Section 5.1.

By assigning different values for M, we can control the precision of the approximated trajectory and the amount of residues. Fig. 3 illustrates the average RMSE/DOF vs. number of RBFs used to approximate the SF sets of 15 happy, sad, young, old, energetic, and tired walking sequences. We have employed 1 to 8 RBFs to model the SF sets. This approximation can extend well beyond 8 and as expected the residues decrease as more RBFs are used and approximations become more accurate. Perceptually, however, increasing the number of RBFs beyond a certain point is of little significance. To perceptually evaluate the notion of modeling SFs using RBFs, we extracted and modeled the SF sets of 6 stylistic walks using $M=\{1,\ 3,\ 5,\ 7\}$ RBFs. The modeled SFs are then added back onto the neutral portion of the original sequences. Each sequence was animated for of 5 human subjects. Table 2 presents the percentage of the audience who were able to correctly identify the secondary themes of the re-synthesized sequences when compared to the original sequences. Based on the results, only a few RBFs, as little as $M=3$, is sufficient for approximating SFs. The perception error rate, however, significantly increases when only 1 RBF/DOF is employed. When $M=7$ there is no perceived variation with respect to the original sequences, indicating that it is not required to extend the computations beyond $M=7$.

When approximating trajectories using RBFs, some methods may result in sub-optimal models. In Fig. 2 (middle), for example, 2 RBFs are used to model the middle peak which occurs around frame 80. One of these RBFs, however, could have successfully modeled that portion of the trajectory, leaving the other RBF to model and eliminate some of the residues. Proper constraints and approach towards this type of approximation, however, can optimize the outcome such that $(\partial RMSE/\partial M)\leq 0$ would always hold true. Even in cases where a trajectory is already perfectly modeled ($RMSE=0$), adding a new RBF should not increase the RMSE since $\sigma^2=0$ or $\alpha=0$ can be selected.

5. Classification and translation system overview

Successive to illustrating that RBFs can accurately model SF sets, through the following sections, we focus on employing RBFs for classification and translation of SFs. We utilize RBF neural networks (RBFNN) for this purpose. In order to modify and adapt the data for use in two different systems with the aim of classification and style translation, pre-processing is required. Two pre-processing steps are often carried out when dealing with motion data: time warping [7] and principal component analysis (PCA) [25]. The former is performed to temporally align the sequences while the latter is carried out for dimensionality reduction. Through the following sub-sections, we describe these two steps and their influence on the data. Fig. 4 presents the overall schematic of the classification/translation system.

5.1. Correlation optimized time warping

For both classification and translation, the training and test data need to be temporally warped. The goal of this process is to align corresponding features of the data. Usually, the practical and simple DTW method [31] is employed for this purpose. Several other warping techniques, such as [32–34] among others, have been presented and explored which can be employed based on the application at hand. In the case of this study, we utilize correlation optimized time warping (CoTW) [35,37]. This technique has previously illustrated good performance in different aspects such as peak shape and area preservation [36]. In addition, distance measures used in DTW and most other warping methods seem to be improper means for finding the similarities between motion trajectories. Alternative methods, such as Pearson's correlation coefficient (PCC), have shown to be more appropriate for motion sequences [37] and make more contextual sense, since the overall shapes of motion sequences captured using PCC are of more importance compared to distance measurements.

Table 2
Residual error rates and audience identification results for approximated secondary themes.

	$M=1$		$M=3$		$M=5$		$M=7$	
	RMSE/DOF	Perception	RMSE/DOF	Perception	RMSE/DOF	Perception	RMSE/DOF	Perception
Happy	5.62	0.40	4.35	0.80	4.16	0.80	3.99	0.80
Sad	5.60	0.40	3.94	0.80	3.91	1.00	3.86	1.00
Young	4.11	0.60	3.52	1.0	3.01	1.00	2.92	1.00
Old	5.20	0.40	3.20	0.80	3.14	1.00	3.10	1.00
Energetic	4.91	0.60	3.31	1.0	3.01	1.00	2.65	1.00
Tired	5.09	0.60	4.01	0.80	3.90	0.80	3.75	1.00
Average	5.09	0.50	3.72	0.87	3.52	0.93	3.38	0.97

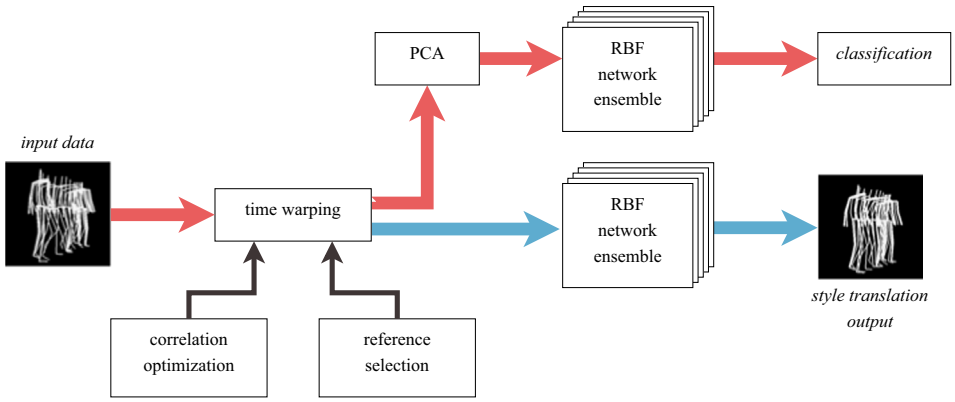

Fig. 4. Overview of the classification and style translation system. The original data are first warped. PCA is applied when performing classification. The ensemble of RBF networks are then trained based on the two modes resulting in either classification or translation of style and affect.

CoTW operates by first selecting a reference trajectory. Then, for two trajectories u and v, each with a length of n, objective function:

$$\rho(u,v)=\frac{\sum_{i=1}^{n}(u_i-\mu_u)(v_i-\mu_v)}{\sqrt{\sum_{i=1}^{n}(u_i-\mu_u)^2\sum_{i=1}^{n}(v_i-\mu_v)^2}}, \quad (5)$$

is calculated and maximized. This is done through segmenting the non-reference trajectory (or trajectories) and linearly stretching/compressing each segment by a constraint parameter. In cases where compression is required, uniform temporal down-sampling is carried out, while for stretching, linear interpolation achieves the desired temporal length. The entire process is carried out using dynamic programming [35].

Fig. 5(a) illustrates two sets of walking sequences both plotted as spatiotemporal trajectories. Blue trajectories belong to sequence (1) (reference) and the pink ones belong to sequence (2). The figure also illustrates the correlation map between them (b), calculated using PCC. In both representations, misalignments are visible. Fig. 5(c) presents the same two sequences when sequence 2 is warped to be aligned with the other. This increased alignment can also be realized from the diagonal line in Fig. 5(d) which is the maximum correlation path.

Warping multiple signals (for example, the trajectories of the 6th DOF from 10 different walking sequences) together with the aim of reaching global alignment is often a difficult task that may need iterative alterations between the signals. A more simple and practical way is to assign a particular trajectory as the reference and align all other trajectories accordingly. The reference trajectory could be selected through a variety of different methods such as random selection, average-based, or even PCA-based techniques. These methods, however, may not present the best solution, resulting in the selection of a trajectory with sub-optimal number of peaks. Therefore, in this paper, we implement the very effective and simple technique proposed in [38]. In this method, the trajectory that is the most similar to all other signals and maintains the most simplicity (less deviation) is selected as the reference. Similarity is again measured using Eq. (5).

A very important point to consider is that when aligning several motion sequences, for trajectory $i=n$, a particular sequence might be selected as the reference, while for trajectory $i=n'$, a different sequence might be the more suitable reference. In such cases, aligning the dataset (multiple sequences) will result in loss of correlation within the different DOFs of each single sequence (except for the reference sequence). This problem will manifest itself as artifacts, one of which is commonly known as footskating [39]. In order to prevent this from happening, we compute the global similarity values between the sequences. Though this may cause some particular trajectories to be improperly set as the reference, synchronization is preserved and no artifacts are introduced.

Using CoTW, the input–output datasets composed of multiple motion sequences are warped and aligned. Fig. 6 presents all corresponding trajectories of the 60th DOF of neutral walk sequences. The set is composed of 14 samples and the red trajectory is the reference selected using the aforementioned approach. The figure shows proper alignment of all the trajectories with respect to the reference.

5.2. PCA

Periodic full-body motion contains a significant amount of redundancy in its many DOFs. As a result, PCA has shown to be

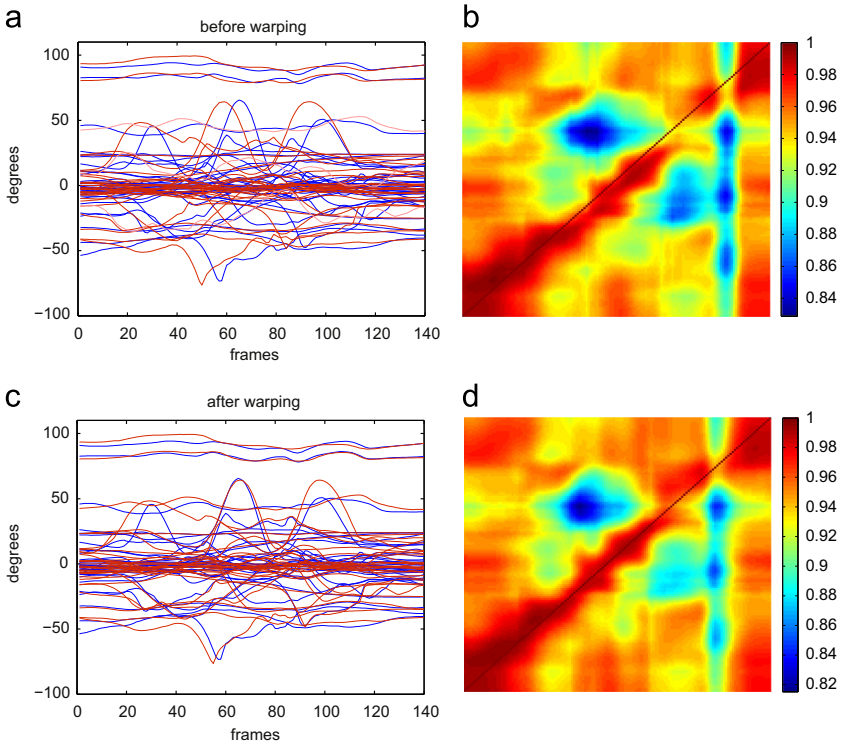

Fig. 5. Two motion sequences before and after warping. Original trajectories are illustrated with blue and red lines in (a) while (b) shows the correlation map where the diagonal line indicates the ideal alignment. The trajectories are mostly aligned as shown in (c) where a more optimum correlation path is achieved as shown in (d). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

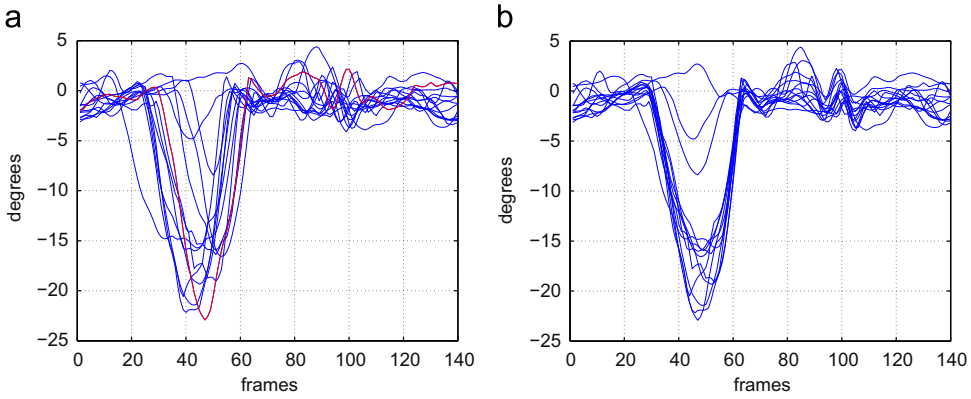

Fig. 6. 14 trajectories of the 60th DOF ($i=60$) from a neutral walk dataset are presented in (a). The red curve is the global optimum reference selected by the system. Temporally warped and aligned versions of the same trajectories using CoTW are illustrated in (b). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

a very effective means of reducing dimensionality in this type of data [25]. In other words, due to high correlation between different parts of the body, there is no need to utilize the entire dimensionality of the data. Therefore, we employ PCA and compute the (principal components) PCs for representing the sequences in the dataset. Thus, lower dimension classifiers can be used. For style translation, however, PCA will not be used since the entire sequence needs to be reconstructed and the output PCs would be difficult to directly interpret and animate.

As mentioned earlier, posture $\mathbf{D}_i$ of the sequence is described by $\mathbf{p}^i=[\theta_1^i\ \theta_2^i\cdots\ \theta_{3l}^i]$ and the trajectory of the jth DOF is described by $\theta_j=[\theta_j^1\ \theta_j^2\cdots\ \theta_j^m]^T$. Accordingly, the average of the jth trajectory is calculated through $\overline{\theta}_j=(1/m)\sum_{i=1}^{m}\theta_j^i$ and constructs the mean data matrix $\overline{\mathbf{D}}$ in which all arrays in each DOF are composed of the average value corresponding to that marker. Accordingly, each vector is zero-centered by:

$$\tilde{\mathbf{D}}_i=\mathbf{D}_i-\overline{\mathbf{D}}_i,\ \text{for } i=1 \text{ to } n \tag{6}$$

Using the formulation above, we calculate the covariance matrix by

$$\Gamma=\frac{1}{3l}\sum_{i=1}^{3l}\tilde{\mathbf{D}}_i\,\tilde{\mathbf{D}}_i^T \tag{7}$$

The eigenvalues and eigenvectors of Γ are then computed. Subsequently, assuming $\mathbf{Q}=[\mathbf{q}_1\ \mathbf{q}_2\cdots\mathbf{q}_r]$ is the set of eigenvectors which correspond to the r largest eigenvalues, we obtain eigenmotion features by projecting $\mathbf{D}$ onto the eigenmotion space using:

$$\mathbf{X}=\mathbf{Q}^T\mathbf{D}, \tag{8}$$

where the redundancy in the data is significantly eliminated. Fig. 7 illustrates that for the different secondary themes in our dataset,

around 94% of the sequence can be represented using only 6 PCs. This rate climbs to over 99% for 15 PCs where regularly 60 vectors were required to demonstrate the data. However, it is critically important to note that the joint angle matrix and the 3D displacement vector are of different nature (degrees and meters). Therefore, the displacement vector must first be removed and PCA applied to the remaining joint angle matrix.

In addition to dimensionality reduction, PCA results in features which are more distinct and easier to classify. In Fig. 8, we illustrate an energetic and a tired sequence in the PC subspace. In (a) we illustrate PC 1 vs. PC 2 in which blue and red clusters have emerged. While for action classification, the first few PCs would suffice, for style recognition, higher PCs would also be informative as they contain smaller variations in the data which most likely correspond to SFs. Fig. 8(b) illustrates PC 1 vs. PC 8 for the same sequences where the two classes are recognizable in the subspace.

6. RBF network for classification and translation

In Section 4, we illustrated that the SFs can be accurately approximated using a weighted sum of only a few RBFs. Computing the set optimum parameters $\{\alpha, \mu, \sigma^2\}$, however, is a challenging problem. In order to calculate the parameter set and employ them for classification and synthesis of SFs, RBF networks have been proposed and utilized [40].

The input–output relation for RBFs can be demonstrated using the network illustrated in Fig. 9. In this formation, L and K dimensional input–outputs are assumed and the hidden neuron activation functions are in the form of $\varphi(t;\mu,\sigma^2)$ with R members.

To train the RBF network, we use the orthogonal least squares technique [41]. Through this algorithm, the network learns by iteratively adding RBF neurons, minimizing the sum of output residues. The weights of the network are calculated by solving:

$$\Phi^T(\Phi\mathbf{w}^T - \mathbf{T}) = 0, \tag{9}$$

where $\mathbf{T}$ is the input set, Φ is the function output set, and the solution is $\mathbf{w}^T = (\Phi^T\Phi)^{-1}\Phi^T\mathbf{T}$. Here, $\mathbf{w}$ corresponds to α in the aforementioned RBF model. It should also be noted that there are alternative training methods for RBF networks which can be employed.

In order to train the system for classification and translation of secondary themes, we construct ensembles of RBF networks. While entire sequences can be trained using a single network, learning all DOFs of a sequence, which often contain phase shifts with respect to one another, is difficult and confusing for ANNs [42]. Therefore, we design an ensemble of networks, which includes a different and separate network for each DOF. Each network classifies each DOF to the best of its ability, and accordingly, a subsystem classifies the entire motion sequence based on majority vote [43].

For classification, our experiments indicate that it is difficult for a single ANN, or even an ensemble of ANNs, to learn different themes of a single primary class of action. For example, when we experimented with an ensemble learning both young and old themes, we observed high confusion rates. This is because the weights

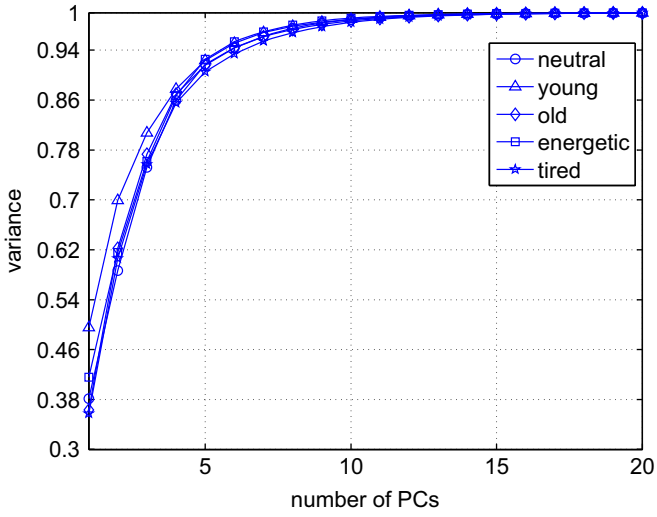

Fig. 7. The amount of variance captured using PCA. Approximately 94% of the information is captured in the first 6 PCs.

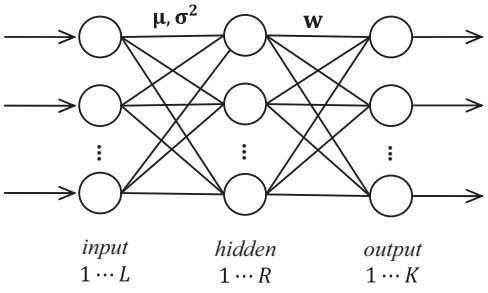

Fig. 9. RBF network layout.

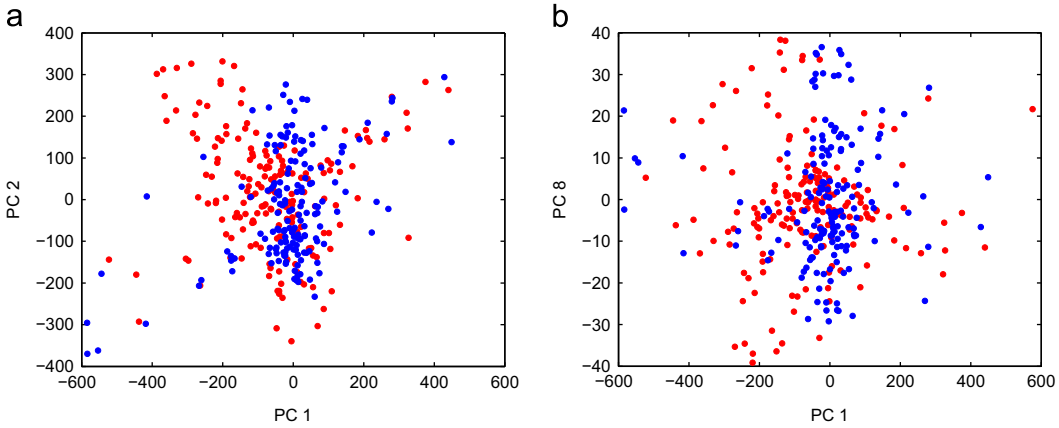

Fig. 8. Visualization of PC subspace for an energetic (red) and tired (blue) walking sequence. In (a), by using the first two PCs, distinct clusters are formed. Similarly, distinct clusters are visible in (b) where PCs 1 and 8 are employed. These higher order PCs are likely to be informative and beneficial for being used in SF-related classifiers. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

corresponding to a young walk cannot accurately represent the old walk, resulting is high error rates, especially for blind test data. To overcome this issue, we generate a different ensemble for each secondary theme: happy, sad, energetic, tired, young, and old. For classification, coefficients of the PCs are used. Each ensemble learns the relationship between the coefficients of the first and subsequent 9 PCs of a particular theme. $\mathbf{C}_1^{\tau}$ represents the set of first PC coefficients of the sequence with theme τ used as the training input while $\mathbf{C}_i^{\tau}$ is the set of the ith PC coefficients of the same sequence used as output. Using the classifier following the training phase, $\mathbf{C}_1^{x}$ is fed to all classifiers of the 6 different thematic ensembles. The ensemble that outputs $\mathbf{C}_i^{\gamma}$ satisfying:

$$\arg\min_{\gamma} \left\| \mathbf{C}_i^{x} - \mathbf{C}_i^{\gamma} \right\|_2, \tag{10}$$

for the majority of $i = \{2, \ldots, 10\}$, determines the secondary theme of the sequence. In this equation γ represents the theme of the ensemble.

Translation of styles is carried out using a different ensemble of RBFNNs. Similar to the classifier module, we train this set of networks as 6 different non-connected anticipators, one for each secondary theme. The networks do not use the PC vectors, but rather the warped motion data. Each anticipator is composed of one individual neural network per DOF. The ensembles learn the relationship between a DOF of a neural set and the corresponding DOF of a stylistic set. As a result, the networks learn how a neutral sequence can be transformed to a stylistic sequence, hence style translation. When it is desired for a neutral walk to take on a specific theme, for example energetic, the neutral sequence is fed to the set of networks that are trained with energetic walks as their outputs.

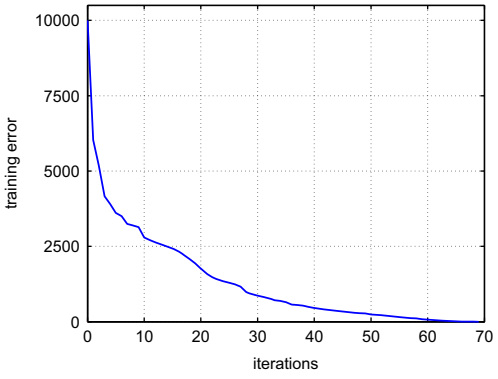

Fig. 10. A sample RBFNN learning curve.

7. Results and discussion

As mentioned earlier, two datasets of multiple walking sequences recorded using motion capture systems are utilized. For our own dataset, several actors were asked to perform neutral and stylistic walking sequences. Ethics approvals for data acquisition as well as perceptual studies were secured. 75 walking sequences, 15 in each of the 5 neutral, young, old, tired, and energetic categories, make up our dataset. Additionally, 48 segmented sequences, 16 in each of the 3 neutral, happy, and sad categories from the HDM05 dataset are employed. The HDM05 skeleton model is modified as described in Section 3.

Other motion capture data such as those available in the Carnegie Mellon University dataset (*http://mocap.cs.cmu.edu/*) are publically available and accessible. However, in order for such data to be readily usable for ANN purposes, the data need to be very consistent. Ideally, the sequences need to be controlled cycles with similar model

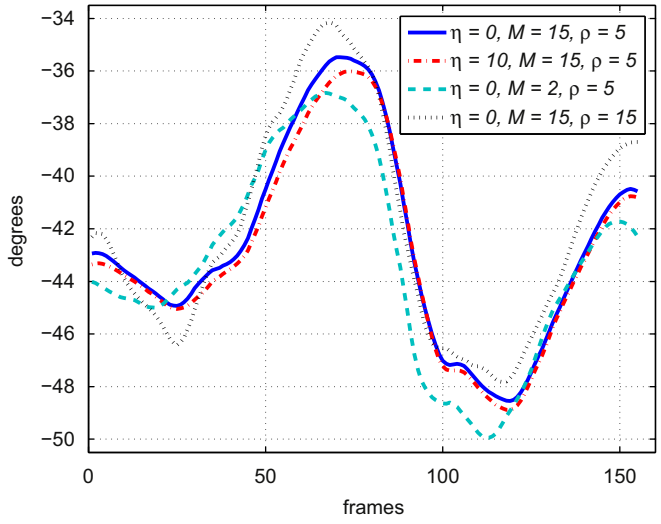

Fig. 11. A sample style translation output with different network parameters.

Table 3
Successful classification percentages for the KNN classifier with 4 different distance measures and the SVM classifier with 4 different kernels compared to RBFNN both with and without PCA.

Classifier	Type	Happy	Sad	Young	Old	Energetic	Tired	Average
KNN	KNN-E	100.0	100.0	93.3	100.0	100.0	100.0	98.9
	KNN-C_1	100.0	100.0	100.0	100.0	93.3	100.0	98.9
	KNN-C_2	93.8	93.8	100.0	100.0	93.3	100.0	96.8
	KNN-C_3	93.8	87.5	100.0	93.3	100.0	100.0	95.8
	PCA+KNN-E	75.0	81.2	86.7	93.3	93.3	93.3	87.1
	PCA+KNN-C_1	68.7	75.0	93.3	86.6	93.3	100.0	86.1
	PCA+KNN-C_2	68.7	75.0	86.7	93.3	86.7	93.3	83.9
	PCA+KNN-C_3	68.7	75.0	86.7	93.3	93.3	100.0	86.2
SVM	SVM-L	93.8	100.0	100.0	100.0	93.3	93.3	96.7
	SVM-Q	100.0	100.0	86.7	93.3	100.0	93.3	95.5
	SVM-C	100.0	100.0	86.7	100.0	100.0	93.3	96.7
	SVM-R	56.2	87.5	93.3	86.7	100.0	60.0	80.6
	PCA+SVM-L	62.5	62.5	93.3	80.0	86.7	86.7	78.6
	PCA+SVM-Q	68.7	75.0	86.7	100.0	93.3	86.7	85.1
	PCA+SVM-C	68.7	68.7	80.0	100.0	80.0	80.0	79.6
	PCA+SVM-R	56.2	50.0	80.0	60.0	60.0	60.0	61.0
RBFNN	RBFNN	68.7	81.2	93.3	60.0	86.7	60.0	75.0
	PCA+RBFNN	87.5	93.8	93.3	93.3	100.0	93.3	93.5

structures but performed multiple times. Furthermore, for the goal put forth in this research, i.e. classification/translation of SF, they need to be carried out with different SF types. To the best of our knowledge, at the time that we conducted this research, there were no publically available datasets from which multiple repeated and controlled affective/stylistic sequences (such as those described in the above paragraph) could be derived. Hence, we used segments from the HDM05 and also recorded our own data.

The ANN ensembles are generated and trained in MATLAB as described in Section 6. In general, the ensembles train well and expected learning curves are achieved. Fig. 10 presents a sample learning curve from the system.

When classification is carried out, 15-fold and 16-fold cross validations are used for our dataset and the HDM05 dataset respectively. Naturally, styles such as happy, energetic, and young are often confused with one another even when observed by human subjects. Similarly, confusion rates for sad, tired, and old are quite high. It is therefore fair to expect for ANNs to show high error rates should a 6-class system be used. Therefore, we use binary-classes for evaluation of the outputs and do not mix the affect, energy, and age related themes.

We evaluate the RBFNN by comparing its performance with KNN and SVM classifiers. For KNN, 4 different distance measures, namely, Euclidean (E), city-block (C_1), cosine (C_2), and correlation (C_3) are used. For SVM, we used 4 different kernel functions, namely, linear (L), quadratic (Q), cubic (C), and RBF (R).

For the 3 classifiers, both PC coefficient vectors and raw data are used, successive to time warping. Table 3 presents the results where several trends are observed. For KNN and SVM, the raw data are better features compared to PCs with an average of 97.6% and 92.4% vs. 85.8% and 76.1% respectively. Also, KNN generally outperforms SVM. For KNN, the distance measure doesn't show a significant effect. For SVM, on the other hand, the RBF kernel performs with the least classification accuracy. Compared to the other classifiers, the raw and warped data are not classified accurately with the RBFNN. When trained with PC coefficients, on the other hand, the results are highly accurate and almost on par with the other classifiers. Meanwhile, different style classes do not show any particular effect and no particular theme is easier to classify compared to others. Finally compared to other methods from the literature, discussed in Section 2 (Table 1), the performance of our approach is relatively high and acceptable.

Style translation is carried out using the ensemble of RBFNNs trained with neutral sequences as inputs and stylistic ones as outputs. The different parameters of the system do not significantly affect the translation process. A sample trajectory transformed from neutral to energetic with different training termination rates ($\eta = 0$ and $\eta = 10$), number of RBFs ($M = 15$ and $M = 2$), and different RBF spreads ($\rho = 5$ and $\rho = 15$), show comparable results. This is shown in Fig. 11 where the differences in the output trajectories are only a few degrees and thus insignificant. Similarly, other DOFs of the input data show reasonable resilience towards the mentioned parameters. Nevertheless, the different number of RBF neurons affects the translation performance for blind data (not been used in the training process). When many RBFs (more than 15) are utilized, over-fitting occurs. This issue manifests itself as unnatural motion in the output

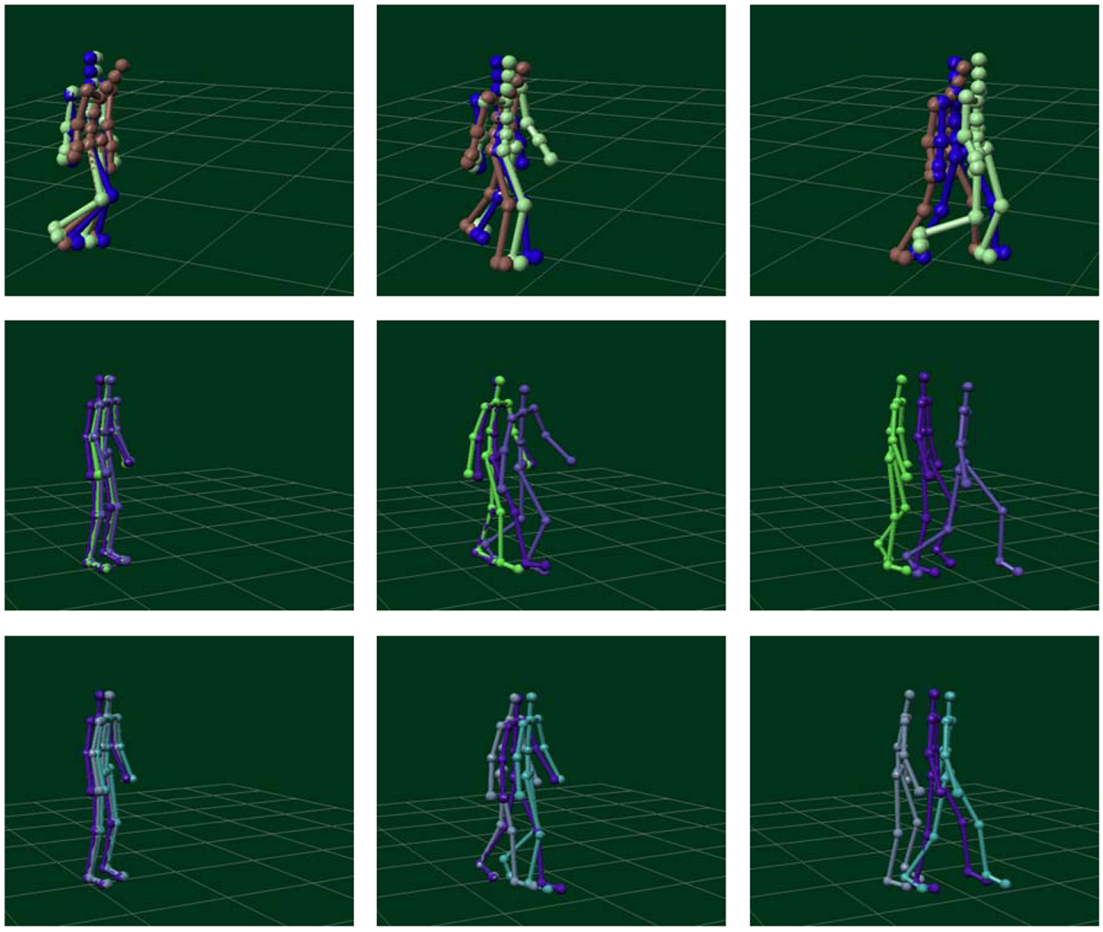

Fig. 12. Inputs and style translation outputs obtained using the RBFNN system. Top row shows happy, neutral, and sad; middle row shows young, neutral, and old; and the bottom row illustrates energetic, neutral, and tired walks. In all three cases, the neutral walks were used as the network inputs and the illustrated affective/stylistic walks were obtained as outputs.

Table 4
Confusion rates for style translation outputs.

		Perceived										
		Happy	**Neutral**	**Sad**		**Young**	**Neutral**	**Old**		**Energetic**	**Neutral**	**Tired**
System output	**Happy**	0.7	0.2	0.1	**Young**	0.9	0.1	0.0	**Energetic**	0.7	0.2	0.1
	Sad	0.0	0.2	0.8	**Old**	0.0	0.0	1.0	**Tired**	0.0	0.1	0.9

sequences. Our experiments show that 5 to 10 RBFs are sufficient for proper generalization of the problem.

The neutral input and transformed outputs are animated using WX Motion Viewer and illustrated in Fig. 12. Stylistic postures are evident in the outcome. Top row shows neutral, happy, and sad sequences. The middle row presents neutral, young, and old walks, and the bottom row illustrates the neutral, energetic, and tired sequences. Raised and increased sway in the arm motion as well as longer steps are demonstrative of the successful style synthesis for happy, young, and energetic styles. In the sad walk, the head is tilted downwards, the arms are lowered, and motion is slowed down. An accompanying video submitted with this manuscript, animated using blender (http://www.blender.org/), better illustrates the outcome. No post processing is carried out on the output data. Very little footskating indicates the accurate performance of our method. Table 4 presents the confusion matrix for the outputs perceived by 10 human participants. The table shows very low confusion rates, denoting successful style translation. Further investigations show that even when actions outside the initial learning input class (neutral) are fed to the ensembles, style translation is carried out with adequate accuracy. For example, when an old walk is used as the input of a neutral-to-young ensemble, the output is transformed to the young style. Artifacts, however, are observed in some joints. Considering the pre-existence of un-related types of SF in these input sequences, the relatively high quality outputs are very promising.

According to the results presented in this paper, the classification mode of our proposed system is on par with KNN and SVM. While KNN shows slightly better performance, our method benefits from higher generalization. Moreover, when trained without the full dimensionality (without PCA), our method is capable of style translation, which is itself a complicated and important problem in the field.

8. Conclusions

In this paper we demonstrate that RBFs are effective tools for modeling affective/stylistic features in motion capture data. It was illustrated that few RBF/DOF can accurately model the SF sets and the remaining spatiotemporal residues were perceptually and numerically insignificant. Subsequently, we presented a novel method using an ensemble of RBFNNs for classification of style classes successive to time warping and PCA. The method performed with high accuracy when compared to other classifiers such as KNN and SVM with different distance measures and kernels. We then trained a separate ensemble which learned stylistic features and added them onto neutral sequences. This approach was used for the process called style translation. High quality animation was generated. Sufficient spatiotemporal quality as well as perceptually sound translation of styles, demonstrated the effectiveness of the proposed system. A major advantage of the proposed method is that it benefits from generalization capabilities of neural networks for both classification and style translation. Finally, our approach is one of the few techniques that uses a single tool, i.e. RBFNN, for both classification and translation of affective/stylistic features in human motion.

Acknowledgments

This work was supported in part by The Natural Sciences and Engineering Research Council of Canada (NSERC) and Ontario Centres of Excellence (OCE).

Some of the data used in this project was obtained from HDM05.

Appendix A. Supplementary material

Supplementary data associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.neucom.2013.09.001.

References

[1] S.A. Etemad, A. Arya, Modeling and transformation of 3D human motion, in: Proceedings of the Fifth International Conference on Computer Graphics Theory and Applications, 2010, pp. 307–315.
[2] R. Cowie, E. Douglas-Cowie, N. Tsapatsoulis, G. Votsis, S. Kollias, W. Fellenz, J.G. Taylor, Emotion recognition in human–computer interaction, IEEE Signal Processing Magazine. 18 (1) (2001) 32–80.
[3] W. Ma, S. Xia, J.K. Hodgins, X. Yang, C. Li, Z. Wang, Modeling style and variation in human motion, in: Proceedings of the ACM SIGGRAPH/Eurographics Symposium on Computer Animation. 2010, 21–30.
[4] C. Rose, M.F. Cohen, B. Bodenheimer, Verbs and Ad-verbs: multidimensional Motion Interpolation, IEEE Computer Graphics and Applications. 18 (5) (1998) 32–40.
[5] H. Sakoe, S. Chiba, Dynamic programming algorithm optimization for spoken word recognition, IEEE Transactions on Acoustics Speech and Signal Processing 26 (1) (1978) 43–49.
[6] M. Brand, A. Hertzmann, Style machines, in: Proceedings of ACM SIGGRAPH. 2000, 183–192.
[7] E. Hsu, K. Pulli, J. Popovic, Style translation for human motion, in: Proceedings of ACM SIGGRAPH. 2005, 1082–1089.
[8] X. Wu, M. Tournier, L. Reveret, Natural character posing from a large motion database, IEEE Computer Graphics and Applications. 31 (3) (2011) 69–77.
[9] A. Bruderlin, L. Williams, Motion signal processing, in: Proceedings of ACM SIGGRAPH. 1995, 97–104.
[10] J. Lee, J. Chai, P.S.A. Reitsma, J.K. Hodgins, N.S. Pollard, Interactive control of avatars animated with human motion data, ACM Transactions on Graphics (SIGGRAPH 2002) 21 (3) (2002) 491–500.
[11] S. Schaal, Is imitation learning the route to humanoid robots? Trends in Cognitive Sciences 3 (6) (1999) 233–242.
[12] S. Calinon, A. Billard, Stochastic gesture production and recognition, model for a humanoid robot, in: International Conference on Intelligent Robots and Systems. 2005.
[13] D.R. Saunders, D. Williamson, N.F. Troje, Gaze patterns during perception of direction and gender from biological motion, Journal of Vision 10 (11) (2010) 1–10. (9).
[14] F. Lv, R. Nevatia, Single view human action recognition using key pose matching and Viterbi Path searching, in: IEEE Conference on Computer Vision and Pattern Recognition, 2007, pp. 1–8.
[15] S. Ali, A. Basharat, M. Shah, Chaotic invariants for human action recognition, in: 11th IEEE International Conference on Computer Vision, 2007, pp. 1–8.
[16] S. Ali, M. Shah, Human action recognition in videos using kinematic features and multiple instance learning, IEEE Transactions on Pattern Analysis and Machine Intelligence 32 (2) (2009) 288–303.
[17] L. Wang, D. Suter, Learning and matching of dynamic shape manifolds for human action recognition, IEEE Transactions on Image Processing 16 (6) (2007) 1646–1661.
[18] Y. Wang, H. Jiang, M.S. Drew, Z.-N. Li, G. Mori, Unsupervised discovery of action classes, in: Proceedings of the Conference on Computer Vision and Pattern Recognition. 2, 2006, pp. 1654–1661.
[19] M. Blank, L. Gorelick, E. Shechtman, M. Irani, R. Basri, Actions as space–time shapes, in: Proceedings of the International Conference on Computer Vision. 2, 2005, pp. 1395–1402.

[20] C. Schüldt, I. Laptev, B. Caputo, Recognizing human actions: a local SVM approach, in: Proceedings of the International Conference on Pattern Recognition. 3, 2004, pp. 32–36.
[21] I. Laptev, B. Caputo, C. Schüldt, T. Lindeberg, Local velocity-adapted motion events for spatio-temporal recognition, Computer Vision and Image Understanding 108 (3) (2007) 207–229.
[22] F. Lv, R. Nevatia, Single view human action recognition using key pose matching and Viterbi path searching, in: Proceedings of the Conference on Computer Vision and Pattern Recognition. 2007, pp. 1–8.
[23] Y. Kuniyoshi, M. Shimozaki, A self-organizing neural model for context-based action recognition, in: First International IEEE EMBS Conference on Neural Engineering, 2003, pp. 442–445.
[24] S.A. Etemad, P. Payeur, A. Arya, Automatic Temporal Location and Classification of Human Actions based on Optical Features, in: Second International IEEE Congress on Image and Signal Processing, 2009, pp. 1–5.
[25] N.F. Troje, Decomposing biological motion: a framework for analysis and synthesis of human gait patterns, Journal of Vision 2 (2002) 371–387.
[26] N.F. Troje, The little difference: Fourier based gender classification from biological motion, in: R.P. Würtz, M. Lappe (Eds.), Dynamic Perception, Aka Press, Berlin, 2002, pp. 115–120.
[27] N.F. Troje, Retrieving information from human movement patterns, in: i.n.:T. F. Shipley, J.M. Zacks (Eds.), Understanding Events: How Humans See, Represent, and Act on Events, Oxford University Press, 2008, pp. 308–334.
[28] S.A. Etemad, A. Arya, Separation and extraction of energy variants from human motion using temporal minimization, in: Proceedings of IEEE Conference on Virtual Environments, Human–Computer Interfaces and Measurement Systems, 2011, pp. 73–77.
[29] J.S.D. Bonet, Multiresolution sampling procedure for analysis and synthesis of texture images, in: Proceedings of ACM SIGGRAPH, 1997, pp. 361–368.
[30] K. Pullen, C. Bregler, Motion capture assisted animation: texturing and synthesis, ACM Transactions on Graphics 21 (3) (2002) 501–508.
[31] H. Sakoe, S. Chiba, Dynamic programming algorithm optimization for spoken word recognition, IEEE Transactions on Acoustics Speech and Signal Processing 26 (1) (1978) 43–49.
[32] E. Hsu, M. da Silva, J. Popovic, Guided time warping for motion editing, in: Proceedings on ACM SIGGRAPH/Eurographics Symposium on Computer Animation, 2007, pp. 45–52.
[33] A. Witkin, Z. Popovic, Motion warping, in: Proceedings of ACM SIGGRAPH. 1995, 105–108.
[34] F. Zhou, F. De La Torre, Canonical time warping for alignment of human behavior, in: Proceedings of Neural Information Processing Systems, 2009.
[35] G. Tomasi, F. van den Berg, C. Andersson, Correlation optimized warping and dynamic time warping as preprocessing methods for chromatographic data, Journal of Chemometrics 18 (2004) 231–241.
[36] V. Pravdova, B. Walczak, D.L. Massart, A comparison of two algorithms for warping of analytical signals, Analytica Chimica Acta 456 (2002) 77–92.
[37] S.A. Etemad, A. Arya, A. Customizable Time warping method for motion alignment, in: Proceedings of the Seventh IEEE International Conference on Semantic Computing (ICSC'13), (2013) 387–388.
[38] T. Skov, F. van den Berg, G. Tomasi, R. Bro, Automated alignment of chromatographic data, Journal of Chemometrics. 20 (11–12) (2006) 484–497.
[39] E. Lyard, N. Magnenat-Thalmann, A simple footskate removal method for virtual reality applications, The Visual Computer 23 (9–11) (2007) 689–695.
[40] E.J. Hartman, J.D. Keeler, J.M. Kowalski, Layered neural networks with Gaussian Hidden units as universal approximations, Neural Computation Summer 2 (2) (1990) 210–215.
[41] S. Chen, C.F.N. Cowan, P.M. Grant, Orthogonal least squares learning algorithm for radial basis function networks, IEEE Transactions on Neural Networks 2 (2) (1991) 302–309.
[42] H. Lee, S. Hong, E. Kim, Neural network ensemble with probabilistic fusion and its application to gait recognition, Neurocomputing 72 (7–9) (2009) 1557–1564.
[43] L. Kuncheva, Atheoretical study on six classifier fusion strategies, IEEE Transactions on Pattern Analysis and Machine Intelligence 24 (2002) 281–286.
[44] M. Müller, T. Röder, M. Clausen, B. Eberhardt, B. Krüger, A. Weber, A, Documentation Mocap Database HDM05, Technical Report, No. CG-2007–2, 2007.
[45] G. Venture, Human characterization and emotion characterization from gait, in: Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) , 2010, pp. 1292–1295.
[46] M. Karg, K. Kühnlenz, M. Buss, Recognition of affect based on gait patterns, IEEE Transactions on Systems, Man, and Cybernetics Part B: Cybernetics 40 (4) (2010) 1050–1061.
[47] M. Karg, R. Jenke, W. Seiberl, K. Kuuhnlenz, A. Schwirtz, M. Buss, A comparison of PCA, KPCA and LDA for feature extraction to recognize affect in gait kinematics, in: Third International IEEE Conference on Affective Computing and Intelligent Interaction and Workshops, 2009, pp. 1–6.
[48] M. Livne, L. Sigal, N.F. Troje, D.J. Fleet, Human attributes from 3D pose tracking, Computer Vision and Image Understanding 116 (2012) 648–660.
[49] K. Grochow, S.L. Martin, A. Hertzmann, Z. Popović., Style-based inverse kinematics, ACM Transactions on Graphics 23 (3) (2004) 522–531.
[50] K. Liu, A. Hertzmann, Z. Popović., Learningphysics-based motion style with nonlinear inverse optimization, ACM Transactions on Graphics 24 (3) (2005) 1071–1081.

S. Ali Etemad received his B.Sc. from Isfahan University of Technology, Iran, in 2007 and his M.A.Sc. in Electrical and Computer Engineering from the department of Systems and Computer Engineering, Carleton University, Ottawa, Canada, in 2009. He is currently working towards his Ph.D. at Carleton and is a research assistant and a teaching assistant. He has also worked as a contract instructor for the department. His current areas of research are interactive multimedia and human–computer interaction, with a focus on human motion and affective computing, using machine learning, pattern recognition, artificial intelligence, image/video processing, and perceptual techniques. He is a recipient of multiple scholarships and awards, including the Ontario Graduate Scholarship, multiple departmental scholarships/awards, and others. He is a reviewer for several journals and has been a technical committee member for various conferences.

Ali Arya received his Bachelor's degree in Electrical Engineering from Tehran Poly-technique, Iran, in 1989 and his Ph.D. in Computer Engineering from the University of British Columbia, Canada, in 2003. He has more than 10 years of industry experience as project manager and software engineer, has worked as instructor and researcher in the University of British Columbia and Simon Fraser University in Vancouver, Canada, and joined the School of Information Technology, Carleton University, Ottawa, Canada, in 2006 where he is currently an Associate Professor of Interactive Multimedia and Design. Ali's major research areas are Computer Graphics and Animation, Multimedia Systems, Human–Computer Interaction, Virtual Worlds and Collaborative Environments, Artificial Intelligence, and Digital Art. He is a senior member of IEEE, on the editorial board of International Journal of Computer Games Technology and Journal of Systemics and Informatics, and member of technical and program committees of many conferences in the area of multimedia systems. His research has been funded by NSERC, SSHRC, OCE, and industry partners.