Discriminative functional analysis of human movements

Ali-Akbar Samadani a,⁎, Ali Ghodsi b, Dana Kulić a

a Department of Electrical and Computer Engineering, University of Waterloo, Canada
b Department of Statistics and Actuarial Science, University of Waterloo, Canada

ARTICLE INFO

Article history:
Available online 7 January 2013

Keywords:
Human movement time-series analysis
Dimensionality reduction for human movement analysis
Affective movement analysis
Human behaviour analysis

ABSTRACT

This paper investigates the use of statistical dimensionality reduction (DR) techniques for discriminative low dimensional embedding to enable affective movement recognition. Human movements are defined by a collection of sequential observations (time-series features) representing body joint angle or joint Cartesian trajectories. In this work, these sequential observations are modelled as temporal functions using B-spline basis function expansion, and dimensionality reduction techniques are adapted to enable application to the functional observations. The DR techniques adapted here are: Fischer discriminant analysis (FDA), supervised principal component analysis (PCA), and Isomap. These functional DR techniques along with functional PCA are applied on affective human movement datasets and their performance is evaluated using leave-one-out cross validation with a one-nearest neighbour classifier in the corresponding low-dimensional subspaces. The results show that functional supervised PCA outperforms the other DR techniques examined in terms of classification accuracy and time resource requirements.

© 2013 Elsevier B.V. All rights reserved.

1. Introduction

The perception of human behaviour arises from a combination of observable functional and expressive cues. The functional cues communicate explicit information about the nature of the activity being performed (e.g., knocking), while the expressive cues communicate information about the feeling and intention of the demonstrator performing the activity (e.g. knocking angrily). Humans infer and ascribe affective meaning to an observed motion even if none is intended (Wallbott, 1998; Blake and Shiffrar, 2007). Eliminating the expressive cues from human behaviour analysis can result in misinterpretation of the demonstrated activity. Therefore, it is important to study the contribution of the expressive cues in human activity recognition and develop a system that can infer the affective expression encoded in human movements.

There are many challenges to developing such a system due to the subconscious nature of the affective movements, and interpersonal differences in conveying and perceiving affective expressions. Human movements can be subtle and are understood by humans sometimes without being consciously noticed and often without explicit consciousness of the features that communicate affective expressions (Cowie, 1535). Furthermore, human movements are highly variable in terms of intensity, timing, and the flexibility of body, even when the same demonstrator repeats a single movement multiple times (i.e., stochastic nature of the human movement execution (Faisal et al., 2008). It is unlikely, therefore, that humans can precisely tell us how to generate or recognize a specific affective movement.

Many of the works on human movement analysis focus on characterizing general human movement, rather than focusing on the affective component of movement. The main categories of approaches to human movement analysis include: dynamical systems (e.g., Nakanishi et al., 2004), neural networks (e.g., Ogata et al., 2005), and dimensionality reduction (e.g., Inamura et al., 2004; Losch et al., 2007). The work presented in this paper falls under the dimensionality reduction category. Statistical dimensionality reduction (DR) techniques transform high-dimensional data to a lower-dimensional subspace spanned by a set of feature transformations suitable for discriminative analysis. The resulting lower-dimensional embedding also helps to visualize the high-dimensional data, which in turn aids interpretation of a given dataset along its intrinsic degrees of freedom represented by the dimensions of the reduced subspace. DR techniques can be categorized into supervised (e.g., FDA) or unsupervised (e.g., PCA), and linear (e.g., PCA) or nonlinear (e.g., Isomap) techniques. DR techniques are often used for human movement analysis. For instance, principal component analysis (PCA) (Santello et al., 2002; Urtasun et al., 2006), Fisher discriminant analysis (FDA) (Dick and Brooks, 2003) and independent component analysis (Ivanenko et al., 2005) have been used to obtain a lower dimensional representation of human movements. In (Jenkins and Matarica, 2004), spatio-temporal Isomap (ST-Isomap) is used to embed motion...
capture data into a lower dimensional subspace for the purpose of extracting motion vocabularies. Application of Hidden Markov Models (HMM) in motion modeling is frequently reported in the literature (e.g. Bernardin et al., 2005; Kulic et al., 2008; Iba et al., 1999). HMMs are well suited for modeling the stochastic nature of human movements. However, HMM-based motion modeling suffers from reduced accuracy as the similarity between the motions increases. To be able to distinguish between similar motions, a more complex HMM is required, which comes at the expense of increased memory and time resource requirements (Kulic et al., 2008).

Unlike research into functional activity perception from human motion, research on affect perception focuses on movement features independent from the functional aspect of the movement. Research into human perception of affective movement has reported on multiple, sometimes conflicting movement features which may be important for affect perception. Mainly, velocity, acceleration, and jerk (rate of change of acceleration) are suggested as the key contributing movement features is conveying affect (Roether et al., 2009). For instance, depression is associated with slow movements and elation is characterized by fast and expansive movements (Argyle, 1988). In (Crane and Gross, 2007), velocity, head orientation, shoulder, and elbow range of motion are found as significant features in affect perception from gait. Boone et al. tested the ability of children (4–8 years) and adults (17–22 years) to correctly perceive affect encoded in expressive dance movements (Boone and Cunningham, 1998). The perception rate was above chance for both children and adults and they report the use of the following six movement cues by participants for perception of affective expressions: changes in tempo (anger), directional changes in face and torso (anger), frequency of arms up (happiness), duration of arms away from torso (happiness), muscle tension (fear), duration of time leaning forward (sadness). In a similar study, Camurri et al. (2003) tested human movement perception in four emotional categories (anger, fear, grief and happiness) conveyed through the same dance movement. They report that human observers were able to detect the transmitted emotions through the dance movement. They also suggest that the duration of the movement, quantity of the movement (the amount of observed movement relative to the velocity and movement energy represented as a rough approximation of the physical momentum) and contraction index (measured as the amount of body contraction/expansion) play key roles in the perception of affect from dance movement. In (Pollick et al., 2001), different affective expressions conveyed through arm and hand shaking and knocking movements displayed as point-light animations were recognized by human observers. Faster movements were perceived as conveying high arousal emotions (arousal is an emotion dimension that corresponds to the level of activation, mental alertness, and physical activity – shortly the “call to action” (Russell and Mehrabian, 1977). Among three affective hand movements used in a perceptual study (anger, happiness, sadness), only angry movements were reliably perceived (Samadani et al., 2011). Furthermore, the arousal level for different affective hand movements used in (Samadani et al., 2011) was correctly perceived. These studies suggest that body movements convey critical information about their demonstrators’ affective state.

The results of these perceptual studies have been applied to implement automated approaches for affect recognition from motion. Berntardt and Robinson used speed, acceleration, jerk, and distance of hand from body to distinguish between neutral, happy, angry, and sad knocking movements using support vector machines (SVMs) (Berntardt and Robinson, 2007). The knocking movements were performed by 30 individuals and a recognition rate of 50% was achieved. Camurri et al. used decision trees to distinguish between choreographed dance movements expressing anger, fear, grief, and joy and report a correct recognition rate of 40% on test data (Camurri et al., 2004). Boredom, confusion, delight, flow, and frustration were detected at a rate of 39% from sitting using a combination of classifiers including Bayesian, SVM, decision trees, artificial neural network (ANN), and k-nearest neighbours (KNN) (D’Mello and Graesser, 2009). In (Kapur et al., 2005), five different classification approaches: logistic regression, decision trees, naive bayes, ANN, and SVM were evaluated for automatic recognition of affective body movements conveying anger, sadness, joy, and fear, using a dataset collected with an optical motion capture system. Mean marker velocity and acceleration values and standard deviations of the marker position, velocity and acceleration were manually selected as the movement features. ANN and SVM were reported as the most accurate classifiers with a demonstrator-specific recognition rate as high as 92%. Gunes and Piccardi (2009) studied the temporal/phase synchrony of affective face and body expressions for enhancing the automatic recognition of affect using a video-recorded movement dataset. They considered 12 expressions: anxiety, boredom, uncertainty, puzzlement, neutral/positive/negative surprise, anger, disgust, fear, happiness and sadness. Different classification approaches including SVM, decision trees, ANN, and Adaboost were applied and the best inter-individual recognition rate (77%) was obtained using Adaboost with decision tree classifiers. ANN was used in (Jansen et al., 2008) to recognize affective expressions (neutral, happy, sad, angry) from gait using kinetic features (measured using a force plate) and kinematic features (joint angle trajectory and angular velocity of arm, hip, knee, and ankle). Person-specific recognition of 98.5% and between-individual recognition of around 80% were reported. In (Rett, 2009), Bayesian nets are used to model movements based on relationships between Laban movement analysis (LMA) descriptors (Laban, 1947) and physical movement characteristics (e.g., acceleration, and curvature). They report an inter-individual recognition rate of 77% for an expressive movement dataset consisting of the following movements: lunging for a ball, maestro (conducting an orchestra), stretching to yawn, making the Ok-sign, pointing, waving bye bye, reaching for someone’s hand to shake, and waving sagittally (approach sign) (Rett, 2009). In (Karg et al., 2010), PCA, kernel PCA, linear discriminant analysis, and general discriminant analysis were used for extracting a set of features to improve automatic recognition of discrete emotions (anger, sadness, happiness, and neutral) from gait using ANN, naive bayes and SVM classifiers. An inter-individual recognition accuracy of 69% was observed for discrete emotions.

Despite the diverse literature on affective movement recognition, the movement features critical to affect recognition are not yet precisely known, and the selection of features for affective movement recognition is usually done in an ad hoc manner. The current paper presents a systematic approach for automatic identification of the features most salient for affective movement recognition from raw movement measurements. Affective body movements are defined by a collection of joint angles or Cartesian positions evolving over time. Therefore, it is important to consider the temporal characteristics of the movements in designing an automatic affect recognition model. This study presents an approach to capture and represent both spatial and temporal features of the affective movements, which are then used to obtain a discriminative subspace for affective movement recognition using adapted dimensionality reduction (DR) techniques.

To apply DR techniques to sequential observations such as affective movements, a fixed-length representation of these observations is needed. An approach for fixed-length representation of sequential observations is basis function expansion (BFE). BFE expresses the sequential observations as temporal functions computed as a linear combination of a fixed number of basis functions (e.g., Fourier basis function). After transforming sequential
observations into basis function space using BFE, DR techniques need to be adapted to enable application to the resulting functional datasets. In (Ramsay, 1997), functional principal component analysis (FPCA) is introduced, an approach for applying PCA to functional datasets. In (Rossi et al., 2005), an extension of radial basis function networks and multiplexer perceptrons to functional inputs is presented. Kernel-based functional nonparametric regression is introduced in (Ferraty and Vieu, 2006) and applied on chemometrics, speech recognition and econometrics. In (Blau et al., 2005), functional k-nearest neighbour classification is introduced and tested with labelled speech samples. Functional fitting of gesture trajectories is exploited in (Bandra et al., 2009) to extract local features (corners of curvature functions), which are used along with global features (gross geometric and structural characteristic of gestures) for gesture matching using different pairwise distance functions.

In this work, affective movements are represented by a collection of functional features using BFE. Then, an approach similar to Ramsay (1997) is employed to extend Fisher discriminant analysis of functional features using BFE. Then, an approach similar to Fisher discriminant analysis of functional features using BFE. Then, an approach similar to Fisher discriminant analysis of functional features using BFE.

2. Proposed approach

In this work, we present a systematic approach for fixed-length time-series representation and automatic recognition of affective movements using BFE and DR techniques, respectively. Sequential observation data such as affective movements are different from classical multivariate datasets in which all the datapoints are conventionally represented by a vector of d discrete dimensions (features). Affective movements are inherently characterized as multidimensional time-series data that might vary in length due to temporal variability in human movement execution. Each movement observation Xi is described by a collection of features evolving over a temporal interval Ti. For instance, if we have n movements defined by m time-series features, then a movement Xi at a time-frame t ∈ Ti is defined as:

\[ X_i(t) = (f_{i1}(t), f_{i2}(t), \ldots, f_{im}(t)), \quad \text{for} \quad \{t = 1, 2, \ldots, T_i\}, \]

where \( f_{ij}(t) \) is the value of the jth time-series feature of movement Xi at time t. To enable the application of DR techniques, the movement trajectories need to be represented by an equal number of discrete features, i.e., a fixed-size vectorial representation. In the following a principled approach for fixed-length representation of the movement observations based on basis function expansion (BFE) is presented.

2.1. Basis function expansion

BFE is a common method for representing sequential observations as temporal functions computed at every time step t (Ramsay, 1997). In BFE, time-series features are computed as a weighted linear combination of a set of basis functions \( \phi_k(t) \), where \( K \) is the total number of basis functions. Considering a multivariate time-series observation \( X(t) = \{x_1(t), \ldots, x_n(t)\}^{T} \) with \( x_i(t) \) being a processed multivariate time-series movement sequence, which is now represented as a vector.

\[ f_j = \sum_{k=1}^{K} c_{jk} \phi_k(t) \quad \text{for} \quad j \in \{1, 2, \ldots, n\}, \]

\( \Phi \) is a matrix of size \( T \times K \) containing basis function values \( \phi_k(t) \), and \( c_{jk} \) is a vector of basis function coefficients for the jth time-series feature of \( X(t) \). The BFE coefficients of individual features are then concatenated in a single vector to represent the multivariate time-series observation \( X \). This forms a dataset of \( X = [x_1, x_2, \ldots, x_n]^T \) with \( x_i \) being a processed multivariate time-series movement sequence, which is now represented as a vector.

2.2. Functional dimensionality reduction

To enable the use of DR techniques on the class of functional datasets, this paper develops an adaptation of statistical DR techniques to functional datasets. First, the functional formulation of PCA proposed by Ramsay Ramsay (1997) is reviewed. An extension of this methodology is then proposed for functional Fisher discriminant analysis (FFDA), functional supervised PCA (FSPCA), and functional Isomap (Flisomap).

2.2.1. Functional PCA

PCA provides a very informative way to interpret variation in the data by projecting it to a lower-dimensional space formed by p principal components (PCs; directions of maximum variation in the data), where \( p < d \) (the dimensionality of the data) (Jolliffe, 2002). Unlike conventional PCA, in FPCA (Ramsay, 1997), the PCs are a set of orthonormal eigenfunctions expressed by a weighted linear combination of basis functions. In multivariate PCA, the eigenequation to be solved is:

\[ SW = \lambda W, \]

where S is the sample covariance matrix, W is a set of eigenvectors with \( \lambda \) being the eigenvalues. For functional observations \( x_i \), the variation in the dataset is approximated using the bivariate covariance function \( \nu \):

\[ \nu(s, t) = \frac{1}{n-1} \sum_{i=1}^{n} (x_i(s) - \bar{x}(s))(x_i(t) - \bar{x}(t)), \]

where \( \bar{x} \) is the mean function.
where \( x \) is the mean of the sequential observations in the dataset and is obtained by:

\[
x(t) = \frac{1}{n} \sum_{i} x_i(t). \tag{4}
\]

In order to obtain FPCA feature transformation \( w_1(t) \), the cost function that corresponds to the variance of the FPCA embedding (FPCA scores) along \( w(t) \) should be maximized:

\[
\max_{w} \sum_{t} |S_{t}|^2 \\
\text{s.t.: } \int w_t^2(s) ds = |w_t|^2 = 1, \tag{5}
\]

where \( S_t \) is the FPCA embedding for a time-series observation \( x(t) \) along \( w(t) \) and is derived using:

\[
S_{t} = \int w_t(s) x(s) ds. \tag{6}
\]

Similarly, the functional feature transformation \( w_r(t) \) can be obtained using Eq. (5), introducing an additional orthogonality constraint on eigenfunctions:

\[
\int w_t(s) w_t(s) ds = 0. \tag{7}
\]

This procedure is continued to obtain as many functional feature transformations as required. In the following, a solution for the above maximization is described. In FPCA, the eigenfunction to be solved is:

\[
\int \nu(s, t) W(s) ds = \lambda W(t). \tag{8}
\]

\( W \) is a matrix of size \( K \times p \) with its columns representing functional PCs or eigenfunctions \( w_r(t); r = \{1, 2, \ldots, p\} \). Suppose that we have a time-series dataset \( X = \{x_1, x_2, \ldots, x_n\} \) that contains \( n \) time-series observations represented by vectors \( x_i = c_i^T \Phi \). Using BFE, \( w(t) \), the rth eigenfunction of (PC) of FPCA applied on \( X \) is expressed as:

\[
w_r(t) = \sum_{k=1}^{K} b_{rk} \phi_k \rightarrow w_r = b_r^T \Phi, \tag{9}
\]

where \( \Phi \) is a matrix of size \( T \times K \) containing basis function values \( \phi_k(t) \), and \( b_r \) represents coefficients corresponding to basis functions \( \phi_k \) used to obtain a functional estimation of \( w(t) \). \( T \) is the number of sample points in the original time-series observation. Let \( C \) be a matrix of size \( K \times n \) containing the coefficients for BFE of time-series observations. Each column of \( C \) represents BFE coefficients for a single time-series observation \( x_i \). The covariance function in Eq. (3) can be re-written as:

\[
\nu(s, t) = (n - 1)^{-1} \Phi^T C C^T \Phi \tag{10}
\]

Hence, the corresponding eigenfunction is:

\[
(n - 1)^{-1} \Phi^T \Phi \left(C^T C \Phi \right) \Phi^T (s) b_r ds = \lambda \Phi^T (t) b_r. \tag{11}
\]

Eq. (11) holds for all the temporal arguments \( t \), and therefore can be reduced to:

\[
(n - 1)^{-1} \Phi^T C M b_r = \lambda b_r, \tag{12}
\]

where \( M = \int \Phi^T (s) \Phi (s) ds \). Furthermore, the constraint \( |w_r|^2 = 1 \) from Eq. (5) is equivalent to:

\[
|w_r|^2 = b_r^T M b_r = 1. \tag{13}
\]

If we define, \( u_r = M^{1/2} b_r \), we can rewrite the FPCA eigen Eq. (12) as:

\[
(n - 1)^{-1} \Phi^T C M^{1/2} u_r = \lambda u_r \tag{14}
\]

s.t.: \( u_r^T u_r = 1 \)

Eq. (14) is a generalized eigenvalue problem and can be solved for eigenfunctions \( u_r \). The resulting \( u_r \) can be used to compute \( b_r \), which in turn is used to compute the eigenfunction \( w_r(t) \) using Eq. (9). Unlike multivariate PCA, where the maximum number of PCs is equal to the dimensionality of the dataset, here the maximum number of PCs is \( T - 1 \) if \( K > T \), and \( K \) otherwise; where \( T \) is the number of discrete sampling points in the original time-series observations and \( K \) is the total number of basis functions used to expand these observations. In the case of a concatenated multivariate time-series dataset obtained as explained in the previous section, the eigenfunction corresponding to each individual time-series feature can be obtained by breaking down the resulting eigenfunctions from FPCA into pieces of length equal \( K \).

2.2.2. Functional FDA

For multivariate datasets of dimensionality \( d \), Fisher discriminant analysis (Fisher, 1936) projects a \( K_f \)-class dataset into a \( K_f - 1 \) dimensional space in an attempt to maximize the distance between projected means and minimize the variance of each projected class. From this intuition, the FDA objective function in a \( K_f \)-class multivariate problem can be formulated as:

\[
\max_{w} \frac{\text{Tr}(W^T S_b W)}{\text{Tr}(W^T S_w W)} \quad \text{s.t.: } \text{Tr}(W^T S_w W) = 1, \tag{15}
\]

where \( W \) is a weight matrix of size \( (d \times K_f - 1) \), and \( S_b \) and \( S_w \) are the between-class and within-class covariances, respectively. Introducing a Lagrange multiplier \( \lambda \) and setting the first derivative of the Lagrange function with respect to \( W \) to zero results in a generalized eigenvalue problem:

\[
S_b^{-1} S_w W = \lambda W. \tag{16}
\]

Therefore, \( W \) is a matrix of eigenvectors associated with the eigenvalues of \( S_b^{-1} S_w \). Here, a similar approach to FPCA is employed to adapt FDA for sequential observations estimated using BFE (i.e., functional observations). Assume that there are \( n \) multivariate time-series movements \( X_{b_i} \), belonging to \( K_f \) classes and each movement \( X_i \) is defined over a temporal sequence \( t = \{1, 2, \ldots, T_i\} \). As before, these movements are estimated as temporal functions using a fixed number of basis functions resulting in a set of functional observations \( X = \{x_1, x_2, \ldots, x_n\} \). The within-class covariance for the \( k_f \)th class of the functional observations is computed as:

\[
\nu_g(s, t) = (n_g - 1)^{-1} \Phi^T C_{k_f} C_{k_f}^T \Phi, \tag{17}
\]

where \( n_g \) is the number of movements belonging to the \( k_f \)th class, \( \Phi \) is a matrix of size \( T_i \times K_f \) containing basis function values \( \phi_k(t) \), and \( C_k \) is a matrix carrying the coefficients corresponding to the basis functions \( \phi_k \), of the movements in class \( k_f \). Similar to conventional FDA, the between-class covariance \( (S_b) \) can be estimated by subtracting the within-class covariances \( (S_w) \) from the total covariance \( (S_t) \). The total covariance \( S_t \) is defined using Eq. (3). Now, if we solve for the FDA optimization problem by introducing a Lagrange multiplier \( \lambda \), we get the below eigenfunction for the functional observations:

\[
\int S_b W(s) ds = \lambda \int S_w W(s) ds, \tag{18}
\]

where \( W(s) \) is a matrix containing the eigenfunctions (columns) associated with the \( K_f - 1 \) largest eigenvalues \( \lambda \). The rth eigenfunction is \( w_r(t) = \sum c_{rk} \phi_k(t), r = \{1, 2, \ldots, K_f - 1\} \). The set of the top \( K_f - 1 \) eigenfunctions is represented in matrix form as: \( W = \Phi B \). Using the expressions obtained for the within-class covariances and the between-class covariance, Eq. (18) can be written for eigenfunction \( w_r \) as follows:
with a linear kernel on reduced subspace). In the following, supervised PCA is formulated using Eq. (6).

Suppose that there are two random variables then the cross-covariance operator between the resulting feature projected into a higher-dimensional feature (Hilbert) space and random variables. For this measure, first, each random variable is mapped to the Hilbert spaces is a column vector of 1's and are the normalizing terms for between-class and within-class covariances. Similar to FPCA, the associated constraint to the eigen Eq. (20) is:

\[
|w|^2 = b_i^T M b_i = 1.
\]

Let \( u_i = M^{1/2} b_i \), therefore eignequation (20) can be written as:

\[
\begin{align*}
&z_b C_s C_b^T M^{1/2} u_i = \lambda z_w C_w C_w^T M^{1/2} u_i, \\
&(z_w C_w C_w^T)^{-1} (z_b C_b C_b^T)^{1/2} u_i = \lambda u_i.
\end{align*}
\]

The above generalized eigenvalue problem can be solved for eigenfunctions \( u_i \), which is used to compute the FFDA feature transformations \( w_i \). Subsequently these feature transformations will be used to embed the movements into the lower dimensional embedding of FFDA using Eq. (6).

### 2.2.3. Functional Supervised PCA based on HSIC (FSPCA)

In cases where there are non-linear relations between two random variables, non-linear dependency measures need to be used to explore the correlation between these random variables. The Hilbert Schmidt independence criterion (HSIC) [Gretton et al., 2005] provides an efficient tool to examine dependencies between two random variables. For this measure, first, each random variable is projected into a higher-dimensional feature (Hilbert) space and then the cross-covariance operator between the resulting feature spaces is used to derive the HSIC measure between the given variables. Suppose that there are two random variables \( x \) and \( y \) mapped to the Hilbert spaces \( F \) and \( G \) using the mapping function \( \phi(x) \) and \( \psi(y) \), respectively. Let \( K(x, x') = \langle \phi(x), \phi(x') \rangle_F \) and \( L(y, y') = \langle \psi(y), \psi(y') \rangle_G \) be unique kernels associated with Hilbert spaces \( F \) and \( G \), respectively. The empirical estimate of the HSIC measure between them is expressed as:

\[
\text{HSIC}(x, y) = \frac{1}{n^2} \text{Tr}(KHLH),
\]

where \( H \) is a constant matrix used to centerize \( K \) and \( L \). In this work, linear and Gaussian RBF kernels are used.

\[
H = 1 - \frac{1}{n} ee^T,
\]

where \( e \) is a column vector of 1's and \( n \) is the dimensionality of the random variables.

Suppose \( X \) is a matrix carrying a set of multivariate datapoints in its columns and \( Y \) is a vector carrying the datapoint labels. HSIC-based supervised PCA [Barshan et al., 2011] is a nonlinear dimensionality reduction technique aiming to find a set of feature transformations \( W \), that maximize \( \text{Tr}(KHKB) \), i.e., it finds a \( W \) such that the transformation \( W^TX \) is highly dependent on \( Y \). \( T = \{1, 2, \ldots, p\} \), where \( p \) is the desired dimensionality of the reduced subspace. In the following, supervised PCA is formulated with a linear kernel on \( X \) and any type of kernel on \( Y \).

\[
\begin{align*}
&\omega_i^X \xrightarrow{\text{linear kernel}} \omega_i^X W^T X, \\
&\omega_i^Y \xrightarrow{\text{any kernel}} b_i, \\
&\max(\text{Tr}(W^TXHKBW^T)) \text{ s.t. } \omega_i^X = 1.
\end{align*}
\]

Similar to classical PCA, the constraint \( \omega^\top W = 1 \) is introduced to make the optimization problem well-posed. It can be shown that the solution to the maximization problem (25) are the largest eigenvalues of \( XHKBX \) and their corresponding eigenvectors [Barshan et al., 2011]. A special case of the above formulation is when \( B = I \), which results in the conventional PCA. Here, a modification of supervised PCA is presented that allows application of this nonlinear DR technique to the functional data set. A modification of Eq. (25) to accommodate the basis function representation of the observations is as follows:

\[
\omega_i^X \xrightarrow{\text{linear kernel}} \omega_i^X W^T b_i, \quad \omega_i^Y \xrightarrow{\text{any kernel}} b_i, \\
\max(\text{Tr}(W^TXHKBW^T)) \text{ s.t. } \omega_i^X = 1.
\]

Similar to FPCA and FFDA, the above maximization can be formulated as a generalized eigenvalue by introducing a Lagrange multiplier \( \lambda \). Then, the resulting eigenvalue can be solved for \( u_i \), which will be used to obtain \( w_i \). Finally, the FSPCA embedding can be obtained using Eq. (6).

#### 2.2.4. Functional Isomap (F-Isomap)

Isomap is a non-linear extension of the Multidimensional Scaling (MDS) [Chen et al., 2008] dimensionality reduction technique, which performs MDS on the geodesic space of the non-linear data manifold, preserving the pairwise geodesic distances between datapoints in the reduced subspace [Tenenbaum et al., 2000]. Isomap embedding is performed in three steps: (1) finding neighbours of each point (e.g. \( k \)-nearest neighbours) and constructing a graph \( M \), which represents the neighbourhood relationships, (2) computing the geodesic pairwise distance between all the points and (3) embedding the data with MDS, based on the geodesic distances between the datapoints. To enable the application of Isomap to sequential observations, first functional estimations of these observations are obtained and then the \( k \)-nearest neighbours for each observation are investigated. Next, the geodesic distance between two sequential observations is computed as the shortest path between the two observations in the neighbourhood graph, which passes through neighbour observations:

\[
d_N(a, b) = \min_0 \sum_{i=0}^{n-1} d(x_i, x_{i+1}).
\]

In Eq. (27), \( O \) includes two or more connected movements in the neighbourhood graph, with \( x_1 = a \) and \( x_n = b \). \( x_i \) and \( x_{i+1} \) are \( k \)-nearest neighbours and \( x_i = C_i^T \Phi \) (functional representation of the movement). \( I \) is the number of movements in the geodesic path between \( a \) and \( b \), including \( a \) and \( b \). Next, matrix \( N \) is formed with entries corresponding to the pairwise geodesic distances \( d_N \). Eigenvector decomposition is performed on \( N \) to obtain the eigenfunctions \( W_r(s) = \omega_i^X \Phi \) corresponding to the top \( p \) eigenvalues. The functional Isomap lower-dimensional embedding is computed using the top eigenfunctions as follows:

\[
S_p(x) = \sqrt{\lambda_r} W_r.
\]

where \( S_p(x) \) is the lower-dimensional embedding of \( x \), \( S_p(x) \) is the \( r \)th dimension of \( S(x) \), \( W_r \) is the \( r \)th component of the \( r \)th eigenfunction \( W_r \), with \( \lambda_r \) being the corresponding eigenvalue for \( W_r \). Unlike other DR techniques in this work, Isomap embedding does not provide a parametric transformation that can be used for transforming previously-unseen high-dimensional observations into the low-dimensional space. In [Bengio et al., 2004], a non-parametric estimation of the Isomap low-dimensional transformation is introduced to test the Isomap embedding. It is shown that the Isomap embedding for a test observation \( x \), denoted as \( S_p(x) \) can be approximated as:
The hand movement dataset considers one movement type, performed by multiple demonstrators (Kleinsmith et al., 2006). Three sets of movements of closing and opening the hand, mainly involving phalangeal and carpo-metacarpal joint movements. Three different expressions were considered: sadness, happiness and anger. For each expression, 5 trials were performed. 

\[ S_r(x_i) = \frac{1}{\sqrt{N}} \sum_{j=1}^{N} w_j \tilde{K}(x_i, x_j), \]  

(29)

where \((a, w_i)\) are the eigenvalue-eigenvector pairs obtained from performing eigenvalue decomposition on the neighbourhood matrix \(N\). \(\tilde{K}(a, b)\) is a kernel function that produces the neighbourhood matrix \(N\) for Isomap embedding and is defined as:

\[ \tilde{K}(a, b) = -\frac{1}{2} \left( d_{2}(a, b) - E_x [d^2(a, b)] - E_{x'} [d^2(a, x')] + E_{x'} [d^2(x, x')] \right). \]  

(30)

In the case of sequential observations, the eigenvector \(w_i\) is approximated using BFE as \(b_i \Phi\). Considering \(X = \{x_1, x_2, \ldots, x_n\}\) as a set of observations, the Isomap embedding for the test observation \(x_t\) is approximated as:

\[ S_r(x_t) = \frac{1}{\sqrt{N}} \sum_{j=1}^{N} (b_j \varphi(t)) (E_x [d^2(x, x_j)] - d^2(x, x_j)), \]  

(31)

where \(E_x\) is an average over the training observations. Detailed discussion and proof for Isomap testing can be found in (Bengio et al., 2004). Here, Eq. (29) is used to test the F-Isomap embedding.

3. Experiments

Two affective movement datasets are used to assess the proposed approaches; a simple dataset consisting of a single hand movement performed by a single demonstrator (Kleinsmith et al., 2006) and a much larger dataset consisting of a variety of full-body movements performed by multiple demonstrators (Kleinsmith et al., 2006). The hand movement dataset considers one movement type, closing and opening the hand, mainly involving phalangeal and carpo-metarcal joint movements. Three sets of movements of 10 trials were collected, where each set conveys a different expression. Three different expressions were considered: sadness, happiness and anger. For each expression, 5 trials were performed on the right hand and 5 on the left hand. A demonstrator performed the hand movements while wearing a Dataglove (ShapeHand from Measurand, 2009). Videos of these movements are available in reference Samadani (2011). The Cartesian coordinates for the root joint (wrist) and three joints A, B and C along each finger (Fig. 1) were collected at 84 frames per second.

Next, a challenging dataset of full-body affective movements was used to further assess the discriminative and computational qualities of the functional DR techniques. This dataset contains 183 acted full-body affective movements obtained from thirteen demonstrators who freely expressed movements conveying anger, happiness, fear, and sadness (Kleinsmith et al., 2006); hence, creating a dataset with different within-class movements (movement differing in structure and physical execution, while expressing a same affect). There are 32 markers attached to body landmarks and their 3D Cartesian coordinates are collected using a motion capture system; hence a total of 96 time-series Cartesian trajectories describe a movement. There are 46 sad, 47 happy, 49 fearful, and 41 angry movements in the full-body dataset.

For both datasets, the affective movements are preprocessed through BFE before the application of functional DR techniques. Two hundred B-splines of 4th degree are chosen to represent the affective movement time series using BFE. BFE is performed using MATLAB code provided in (Ramsay, 2008). Next, FPCA, FFDA, FSPCA, and FIsomap are applied to obtain discriminative lower-dimensional embeddings of the transformed affective movements (i.e., functional estimation of the movements). For the FSPCA, two types of kernels are applied to the movement labels; a linear kernel and the Gaussian radial basis function (GRBF) kernel. MATLAB code provided in (Tenenbaum et al., 2000) is modified to generate the FIsomap embedding. For FPCA, FFDA and FSPCA, Eq. (6) is used to obtain the lower-dimensional embedding for test observations. For FIsomap, lower-dimensional embeddings of test observations are computed using Eq. (31). The performance of the functional DR techniques in discriminating between affective movements is examined with leave-one-out cross validation (LOOCV) using one-nearest-neighbour classification error.

4. Results

The two-dimensional embeddings of the affective hand movements obtained by FPCA, FFDA, FSPCA and FIsomap are shown in Fig. 2. The LOOCV training and testing recognition rates for different functional DR techniques along with their training time for the full-body dataset are shown in Table 1.

The functional DR techniques are next applied to a more challenging full-body movement dataset. Table 2 shows the LOOCV training and testing recognition rates obtained using the 1NN classifier in the resulting reduced spaces for the full-body dataset. For the full-body movements, 3D subspaces of the functional DR techniques are used to compute LOOCV recognition rates due to their discriminative advantage over 2D subspaces. The 2D embedding of training and testing full-body movements are shown in Fig. 3 to illustrate the ability of the functional DR techniques to discriminatively embed the high-dimensional affective movements in a low dimensional space.

---

1 The markers are placed on the following body landmarks: left front head, right front head, left back head, right back head, top chest, centre chest, left front waist, right front waist, left back waist, right back waist, top of spine, middle of back, left outer metatarsal, right outer metatarsal, left toe, right toe, left shoulder, right shoulder, left outer elbow, right outer elbow, left hand, right hand, left wrist inner near thumb, right wrist inner near thumb, left wrist outer opposite thumb, right wrist outer opposite thumb, left knee, right knee, left ankle, right ankle, left heel, right heel.
5. Discussion

FFDA performs perfectly on the training observations, as it collapses different observations belonging to the same affective class to nearly a single point (FFDA embeddings in Figs. 2 and 3). However, it fails to accurately separate the test observations (low FFDA testing recognition rates reported in Tables 1 and 2). The FFDA overfitting effect observed for the hand movements (Fig. 2) and for the full-body movements (Fig. 3) is due to the small number of high-dimensional observations used (smaller number of observations than the dimensionality of the observation), and confirms earlier findings that FDA performs poorly on high dimensional problems when few training points are available (Martinez and Kak, 2001).

Separated clusters of hand movements belonging to the same affective class are clearly observable in the reduced subspaces of FPCA, F-Isomap, and FSPCA (Fig. 2). These distinct clusters correspond to the movements performed on the left and right hands. Furthermore, in the FPCA-Linear and F-Isomap embeddings, angry and happy hand movements overlap to some extent, while sad hand movements form distinct clusters (Fig. 2).

FIsomap achieves good performance on the training hand movements, however, poor performance is obtained on testing exemplars. As Isomap does not provide a parametric transformation that can be used for evaluating the generalizability of the resulting reduced subspace to out-of-sample movements, an approximation of Isomap out-of-sample embedding proposed in (Bengio et al., 2004) is adapted here to test the generalizability of the Flsomap to unseen movements. This approximation might be the reason for the poor testing performance of the Flsomap (Table 1). Furthermore, the performance of Isomap deteriorates if the datapoints belong to disjoint underlying manifolds, which might be the case here (Geng et al., 2005).

For the full-body dataset, among the functional DR techniques, FSPCA-GRBF embedding shows dense and more distinct one-class clusters of movements in the resulting lower-dimensional space. By visual inspection, it is easy to associate different subintervals of dimensions of the FSPCA-GRBF subspace to distinct affective movements. For instance, lower values of the first dimension of the FSPCA-GRBF subspace are occupied by sad movements, whereas happy movements are distributed along the higher values of the first dimension. As can be seen from the LOOCV recognition rates (Table 2), FSPCA-GRBF results in the highest training and testing recognition rates. FPCA, F-Isomap, and FSPCA-Linear embeddings show a large overlapping between full-body movements from different classes resulting in poor discrimination between the training and testing affective movements (Table 2).

As overviewed in Section 1, in the affective movement recognition literature, the automatic inter-individual emotion recognition rates range from 40% to 77% (Bernhardt and Robinson, 2007; Camurri et al., 2004; D’Mello and Graesser, 2009; Gunes and Piccardi, 2009; Karg et al., 2010), depending on the number of intended emotions, number of demonstrators, and the amount of within-class variations in the movements. Using apex postures from 108 of the movements in the full-body dataset used in our study, Kleinsmith et al. (2006) tested human perception of the intended emotions. The overall recognition rate was 54.7% with the least recognized postures being fearful ones (49.9% recognition rate) and the most recognized ones being the sad postures (63.4% recognition rate). The FSPCA-GRBF applied on the full-body dataset achieves overall training recognition rate of 59.1% (sad: 65.9%, happy: 63.9%, fearful: 50.9%, angry: 55.6%) and testing recognition rate of 53.6% (sad: 60.9%, happy: 61.7%, fearful: 44.9%, angry: 46.4%) which are comparable to human recognition rates on the same dataset as reported in the Kliensmith et al. perceptual study (Kleinsmith et al., 2006).

If we consider the extent to which each class in the reduced subspaces is spread as a measure of quality of the embeddings of DR techniques, by visual inspection, one can argue that FSPCA-GRBF kernel results in the most compact embedding of the classes for both hand and full-body movements. In the case of FFDA, despite of the compact embedding of each class to a single point, as discussed above, due to overfitting, poor embedding of the test observations is observed (FFDA embeddings in Fig. 3). The compact embeddings of the high-dimensional movements also facilitate the interpretation of the reduced subspace dimensions, as distinct subintervals of these dimensions can be associated to distinct affective movement classes (i.e., in FSPCA-GRBF embedding shown in Fig. 2, sad movements are uniquely characterized by lower values of dimension 1).

Next, the resulting functional transformations obtained by FPCA, FFDA, and FSPCA for the hand dataset are plotted as perturbations of the overall mean of the feature; \( \mu(t) \pm \varepsilon \mu(t) \), where \( \mu(t) \) is the functional feature mean across the movements, \( \varepsilon \) is the perturbing constant, and \( \mu(t) \) is the functional transformation corresponding to that feature. Fig. 4 shows examples of perturbation plots for the hand movements resulting from FPCA, FFDA, and FSPCA techniques: (a) \( Y \)-trajectory of joint C of the middle finger corresponding to the first dimension of the reduced subspaces and (b) \( Z \)-trajectory of joint A of the thumb corresponding to the second dimension of the reduced subspaces. The perturbation plots for other hand coordinates can be obtained similarly. These perturbation plots help to evaluate the importance of different movement functional features in constructing the discriminative reduced spaces either as a whole or over subintervals. For the
Y-trajectory of joint C of the middle finger, the perturbation plots for FPCA, and FSPCA with linear and GRBF kernels are quite similar while being different from the one corresponding to FFDA embedding (Fig. 4(a)). According to Fig. 4(a), for the Y-trajectory of joint C of the middle finger, the functional feature variations mainly at the beginning and at the end of the temporal interval of the movement execution play an important role in producing FPCA and FSPCA subspaces. If a functional feature has little effect in producing the discriminative subspaces, it appears as overlapping of the functional feature mean with its positive and negative functional transformation perturbations as is the case in the FFDA perturbation plot for the Y-trajectory of joint C of the middle finger. Therefore, the contribution of the Y-trajectory of joint C of the middle finger in constructing FFDA embedding is not significant. As can be seen in Fig. 2, the 2D subspaces generated using FPCA and FSPCA are similar in their first dimension; sad movements are embedded along the lower-values, while happy and angry movements are characterized by higher values of the first dimension; hence, similar perturbation plots for the first dimension of the FPCA and FSPCA subspaces are obtained. Differences between FPCA and FSPCA embeddings occur along the second dimension.

An example of the perturbation plots for the Z-trajectory of joint A of the thumb corresponding to the second dimension of the reduced subspaces produced by FPCA, FFDA, and FSPCA is shown in Fig. 4(b). For the Z-trajectory of joint A of the thumb, the perturbation plot for FPCA demonstrates that the functional feature variations mainly at the beginning and at the end of the temporal interval of the movements play an important role in producing the FPCA embedding. For the FFDA embedding, the functional feature transformation introduces a highly variable trend of weights starting at one-third of the movement feature and attenuating toward the end (FFDA perturbation plot in Fig. 4(b)). This demonstrates the FFDA search for a direction in high-dimensional

Table 2
Leave-one-out cross validation training and testing recognition rates for functional DR techniques applied on the affective full-body movements. Highest training and testing recognition rates are highlighted.

<table>
<thead>
<tr>
<th></th>
<th>Training recognition rate (%)</th>
<th>Testing recognition rate (%)</th>
<th>Elapsed training time (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>FPCA</td>
<td>44.0</td>
<td>43.2</td>
<td>0.70</td>
</tr>
<tr>
<td>FFDA</td>
<td>100</td>
<td>36.6</td>
<td>564.59</td>
</tr>
<tr>
<td>F-Isomap</td>
<td>47.0</td>
<td>43.7</td>
<td>1.81</td>
</tr>
<tr>
<td>FSPCA-Linear</td>
<td>44.2</td>
<td>43.7</td>
<td>1.68</td>
</tr>
<tr>
<td>FSPCA-GBRF</td>
<td>59.1</td>
<td>53.6</td>
<td>1.71</td>
</tr>
</tbody>
</table>

Fig. 3. Affective full-body movement embedding for (a) training data and (b) testing data in the resulting 2D subspaces.
movement space that maximally separates different movement classes, while forming compact classes along that direction through weighting individual basis functions constructing the functional feature. The contribution of the Z-trajectory of joint A of the thumb in constructing the FSPCA-Linear embedding is not significant. The perturbation plot for the FSPCA-GRBF shows that the entire Z-trajectory of joint A of the thumb plays an important role in constructing the discriminative FSPCA-GRBF embedding. Therefore, FPCA and the two variations of FSPCA DR techniques result in a different functional feature transformation for discriminative embedding of the affective movements. The superiority of the FSPCA techniques over the FPCA in the discriminative analysis (Table 1) is likely due to the fact that FSPCA benefits from movement labels in constructing the discriminative lower-dimensional subspace.

Among the functional DR techniques covered here, FFDA is the most computationally expensive and this is due to the requirement for computing the overall covariance as well as individual class covariances. The least computationally expensive is FPCA followed by FSPCA. The computational complexity of the Flsomap algorithm depends on the computation of pairwise geodesic distances (Tenenbaum et al., 2000).

The presented discriminative approach is not limited to pairwise comparison (such as the case in Bandera et al., 2009), instead, it systematically identifies a subspace where discriminative analysis on all the movements can be performed at once. Furthermore, newly observed movements can be classified by embedding them in the resulting lower-dimensional space. The by-product of the identified lower-dimensional spaces are the features spanning these spaces, which are the critical movement features in discriminating between different affective movements. Therefore, there is no need for hand-picking and estimating movement features that might be important for affective movement recognition (as done in Rett, 2009; Bandera et al., 2009).

6. Conclusion and future work

Different DR techniques are used in the context of functional data analysis to find a discriminative low-dimensional embedding of a set of sequential affective movement observations. First, FDA, supervised PCA and Isomap DR techniques are adapted to enable application to the sequential functional observations. The sequential observations are first modelled as temporal functions using BFE with a fixed number of B-spline basis functions. Then, functional versions of the DR techniques, FPCA, FFDA, FSPCA and Flsomap, are applied on the BFE representation of the affective movements and corresponding lower-dimensional embeddings are obtained.
Leave-one-out cross validation using 1NN classifier was applied in the reduced embeddings and training and testing recognition rates were computed as the measure of performance for the functional DR techniques. Overall, considering the testing recognition rate for the both datasets (testing recognition rate demonstrates the capability to discriminate unseen movements) and elapsed training time as assessment criteria, the FSPCA-GRBF outperforms other functional DR techniques tested here. Furthermore, for the full-body dataset, considering the large number of freely expressed affective movements demonstrated by 13 different actors, FSPCA-GRBF technique shows promising performance when compared with the perceptual and automatic inter-personal affective movement recognition studies.

The presented movement recognition approach is particularly useful since it uses a minimal set of systematically obtained feature transformations (dimensions spanning the lower-dimensional subspaces), rather than trying to recognize the movements in their original high-dimensional time-series format, which is most likely characterized by many redundant and irrelevant features to the recognition task. BFE is an efficient way to represent the high-dimensional and variable-length sequential observations as functions estimated by weighted linear combinations of a fixed number of basis functions, which satisfies the discriminative DR techniques' requirement for fixed-length vectorial representation of the sequential observations. BFE can also be regarded as an intermediate dimensionality reduction step as it produces a smoother down-sampled version of the original temporal observations.

The long term goal of this project is to develop a systematic approach for identifying a subset of features that can be used to distinguish between different movements (i.e., to optimize movement recognition). Using more variants of movements conveying an affective expression will help identify a more generalized set of features (feature transformations) and movement qualities associated with that affective expression, which consequently will help to develop a more robust and accurate computational model for human affective movement recognition.

Acknowledgement

We would like to thank Dr. Andrea Kleinsmith for making their full-body dataset (Kleinsmith et al., 2006) available for us. This work is supported by the National Science and Research Council of Canada (NSERC).

References


Bernardini, K., Ogawa, K., Ikeuchi, K., Dillmann, R., 2005. A sensor fusion approach for identifying a subset of features that can be used to distinguish between different movements (i.e., to optimize movement recognition).


