Temporal Extension of Laplacian Eigenmaps for Unsupervised Dimensionality Reduction of Time Series

M. Lewandowski, J. Martinez-del-Rincon, D. Makris and J.-C. Nebel
Digital Imaging Research Centre, Kingston University, London, UK
Email: M.Lewandowski@kingston.ac.uk

***Abstract*—A novel non-linear dimensionality reduction method, called Temporal Laplacian Eigenmaps, is introduced to process efficiently time series data. In this embedded-based approach, temporal information is intrinsic to the objective function, which produces description of low dimensional spaces with time coherence between data points. Since the proposed scheme also includes bidirectional mapping between data and embedded spaces and automatic tuning of key parameters, it offers the same benefits as mapping-based approaches. Experiments on a couple of computer vision applications demonstrate the superiority of the new approach to other dimensionality reduction method in term of accuracy. Moreover, its lower computational cost and generalisation abilities suggest it is scalable to larger datasets.**

***Keywords*-temporal Laplacian Eigenmap; dimensionality reduction; manifold learning; time-series; human motion**

I. Introduction

With the exponential increase of data production driven by applications such as the internet, computer vision, medical imaging, speech recognition and genomics, powerful tools are required by scientists to allow the analysis of these data. Since real datasets are usually highly dimensional and nonlinear, nonlinear dimensionality reduction techniques have become essential in the exploration of large volumes of multivariate data.

These methods can be classified in two main categories: mapping-based and embedding-based. Mapping-based approaches, such as Gaussian process latent variable model (GPLVM) [1], use probabilistic nonlinear functions to map the embedded space to the data space. On the other hand, embedded-based approaches such as Laplacian Eigenmaps (LE) [2] and Isomap [3], estimate the structure of the underlying manifold by approximating each data point according to their local neighbours on the manifold.

As many datasets are time series, quality of embedded spaces can be improved by taking into account the temporal dependencies between points. Spatio-temporal Isomap (ST-Isomap) [4] empirically alters the original weights in the graph of local neighbours to emphasise similarity between temporal related points. Similarly, back constraint GPLVM (BC-GPLVM) [5] is able to include temporal coherence constraints to ensure the smoothness of the mapping between spaces. Another mapping-based approach, i.e. Gaussian process dynamical model (GPDM) [6], integrates time information by associating nonlinear, autoregressive dynamics to the embedded space.

When dealing with time series, all these temporal extensions of dimensionality reduction methods generate better quality embedded spaces than the initial approaches. However, they also suffer from their original limitations: embedding-based techniques are very sensitive to the choice of manually set parameters such as neighbourhood size, whereas mapping-based approaches are so time-consuming that they are limited to applications which do not rely on large training sets.

Many computer vision applications, such as body pose tracking and action recognition, require the creation of low dimensional models learned from large time series datasets. Since accuracy is highly correlated to training set sizes, mapping-based approaches are usually not suitable. On the other hand, embedding-based methods suffer from lack of robustness. In order to deal with both limitations, we propose a novel embedding-based method called temporal Laplacian Eigenmaps (TLE) where temporal information is integral part of the objective function and neighbourhood sizes are derived automatically from data analysis. This is achieved by introducing two types of intuitive temporal graphs which are incorporated into the LE framework. This produces description of low dimensional spaces which integrates time coherence between data points. Our method is particularly suitable for time series data which include data repetition; otherwise it is equivalent to standard LE.

The structure of this paper is organised as follows. After introducing the TLE algorithm, it is validated qualitatively and quantitatively on real datasets of human motion. Then, we apply our method to an action recognition task. Finally, conclusions and future work are presented.

II. Methodology

Temporal Laplacian Eigenmaps algorithm is an unsupervised nonlinear method for dimensionality reduction which learns manifolds designated for time series data. Given a set of data points $Y = \{y_i\}_{(i=1..n)}$ distributed on a manifold in a high dimensional space ($y_i \in R^D$), TLE is able to discover their low dimensional representation $X = \{x_i\}_{(i=1..n)}$, $x_i \in R^d$ with $d \ll D$ by preserving the temporal structure

of the data manifold instead of its local geometry as standard LE does [2].

The temporal similarity between data points is maintained implicitly during dimensionality reduction by building new types of neighbourhood graphs (Fig. 1) which express temporal dependencies. Consequently, local temporal neighbours are placed nearby in the embedded space without the need to enforce any artificial constraints as in [4]. Two types of temporal neighbourhoods are defined for each data point P_i:

- Adjacent temporal neighbours (A): the *2m* closest points in the sequential order of input (Fig. 1a):

$$A_i \in \{P_{i-m}, ..., P_{i-1}, P_i, P_{i+1}, ..., P_{i+m}\} \quad (1)$$

- Repetition temporal neighbours (R): the q_i points similar to P_i, extracted from the q_i repetitions, F_{ik}, of time series fragment, F_i, defined by *2s* adjacent temporal neighbours (Fig. 1b):

$$R_i \in \{F_{i,1}(C), ..., F_{i,q_i}(C)\} \quad (2)$$

where $F_{i,k}(C)$ returns the centre point of $F_{i,k}$.

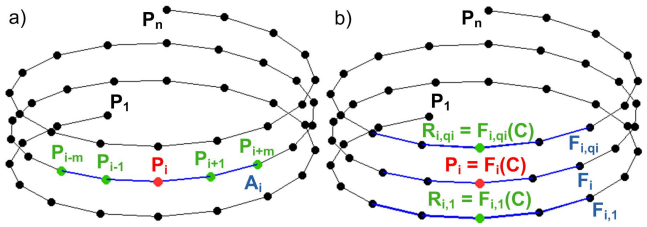

Figure 1. Description of temporal neighbours (green dots) of a given data point, P_i, (red dots) in a) adjacent and b) repetition graphs.

The process of dimensionality reduction can be summarised by the following steps.

First, weights *W* are assigned to the edges of each graph $G \in \{A, R\}$ using the standard LE formulation:

$$W_{ij}^G = \begin{cases} exp(\|y_i - y_j\|^2) & \text{i,j connected} \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

Then we introduce the following extended cost function to combine information from both graphs:

$$\text{argmin}_X = X^T L_A X + X^T L_R X \quad (4)$$

$$\text{subject to: } X^T D_A X + X^T D_R X = I \quad (5)$$

where $D^G = diag\{D_{11}^G, D_{22}^G, , D_{nn}^G\}$ is a diagonal matrix with entries: $D_{ii}^G = \sum_{j=1}^{n} W_{ij}^G$, and $L_G = D^G - W^G$ is the Laplacian matrix. The minimum of the objective function can be found by applying Lagrange multipliers to Eq. 4 subject to the constraint expressed by Eq. 5:

$$\wedge(X, \lambda) = X^T(L_A + L_R)X - \lambda(I - X^T(D_A + D_R)X) \quad (6)$$

$$(L_A + L_R)X = \lambda(D_A + D_R)X \quad (7)$$

The embedded space *X* is spanned by the eigenvectors given by the *d* smallest nonzero eigenvalues λ using the generalised eigenvalue problem (Eq. 7).

The selection of *2m* adjacent neighbours, where *m=1*, is straightforward since it is based on the data temporal order (Eq. 1). The size of the repetition neighbourhood, *q*, corresponds to the number of times a state is repeated in the training set. While ST-Isomap considers this as prior knowledge, we overcome this constraint by introducing a procedure to automatically determine the optimal repetition neighbourhood:

1) Associate to each data point, P_i, *2s* adjacent temporal neighbours, where s=5, to create the local trajectory, F_i, centred on P_i.
2) Search for similar trajectories $F_{i,k}$, according to the dynamic time warping metric (DTW) [7] and a similarity of 1.5 standard deviations by sliding a warping window through the entire training set.
3) Extract from each similar trajectory, $F_{i,k}$, the data point which corresponds to P_i, i.e. the centre of $F_{i,k}$. The extracted points define P_i's temporal repetition neighbourhood.

III. Validation of Temporal LE approach

The proposed algorithm is evaluated through a comparative analysis of performance produced by standard dimension reduction methods, i.e. LE, Isomap and BC-GPLVM, and their respective improved temporal versions, i.e. TLE, ST-Isomap and GPDM.

Since activity independent techniques have been used to produce 3D posture estimates [8]–[10] and analysis of such sequences allows activity identification [11], initial estimates can be refined using learned motion models. Here, MoCap data of repeated actions provided by the HumanEva dataset [12] are converted into quaternions, which produces sequences of 52-dimension feature vectors. Then, each action space is reduced to their intrinsic dimensionality which is 2 according to eigenvalue-based estimator [13]. Finally, 3D pose estimates are refined using a nearest neighbour approach: estimates are projected to the embedded space and the nearest neighbour is projected back to the posture space [14]. The reconstruction error of the refined 3D poses is calculated using the groundtruth provided in the HumanEva dataset [12].

In order to evaluate embedding-based methods using this framework, we have included a mapping function which allows projecting data between high and low dimensional spaces. This is achieved using unsupervised Radial Basis Function network [14].

Here, we consider three different subjects performing two actions (walking and jogging). To measure the performances of the different methods, experiments are conducted using cross-validation taking either one or two subjects for training leaving respectively two or one subjects for testing. Initial

pose estimates are simulated by introducing a Gaussian noise to groundtruth poses to obtain an average error of 80mm. Quantitative results are calculated by averaging 5 test sequences. Unlike TLE and mapping-based approaches, all other embedding-based methods require manual parameter tuning. In this study, we used the default parameters provided with the Matlab implementations of BC-GPLVM [5] and GPDM [6]. In the case of spectral methods, extensive testing was conducted to determine the optimal settings for each experiment. In addition, the number of nontrivial neighbours required for ST-isomap [4] was calculated using the TLE estimation procedure.

Performance analysis confirms the generalisation abilities of the methods integrating temporal constraints since data from a second subject improves their accuracy (Table I). Conversely, performances of Isomap and LE worsen. Fig. 2a, 2b, 2c shows these embedding-based methods fail to produce a unique ellipse to represent the 2-subject walking cycle in the embedded space. Among temporal methods, BC-GPLVM and TLE benefit the most from additional training samples (accuracy +12%). On the other hand, GPDM's dynamic model seems to be able to optimise most of its parameters from a single subject. Consequently, TLE and BC-GPLVM are the most successful approaches. However, TLE is not only more precise and produces a better quality embedded space (Fig. 2b, 2f), but also significantly faster, even when the cost of the proposed automatic parameter estimation procedure is added (Fig. 3 last columns). This is very important because this shows that, unlike BC-GPLVM, TLE has the ability to learn models from much larger training sets which should conduce to even better results.

Table I
REFINEMENT ACCURACY.

Accuracy (std) in mm	Walking: 1 subject	Walking: 2 subjects	Jogging: 1 subject	Jogging: 2 subjects
Isomap	73.3 (4)	81.9 (22)	79.2 (5)	87.6 (12)
ST-Isomap	70.9 (12)	69.6 (7)	77.6 (10)	74.0 (9)
BC-GPLVM	70.5 (7)	62.4 (5)	77.0 (4)	68.8 (3)
GPDM	67.1 (9)	66.2 (8)	74.1 (4)	72.2 (5)
LE	72.5 (5)	86.0 (5)	76.7 (5)	90.8 (3)
TLE	**64.9 (4)**	**57.3 (2)**	**71.8 (3)**	**63.5 (3)**

IV. APPLICATION TO ACTIVITY RECOGNITION

In the previous section, we have demonstrated the superior performance of TLE on a human pose estimation scenario. Here, we integrate our technique within a standard human action recognition framework [15] to perform video annotation. In addition to the static 2916-dimension feature vector, which is built using implicit distance functions over extracted silhouettes, we include dynamics characteristics provided by optical flow [16]. Then, using the generalisation

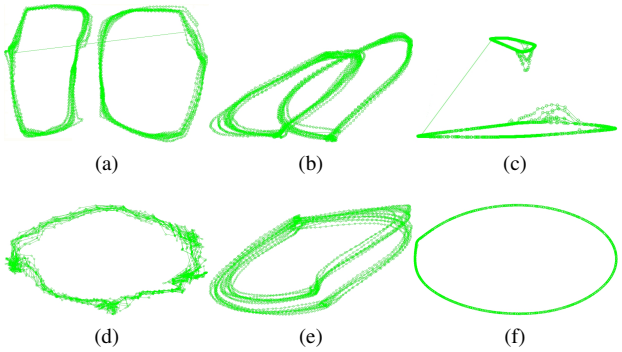

Figure 2. Embedded spaces for walking (2 subjects) using a) Isomap, b) BC-GPLVM, c) LE, d) ST-Isomap, e) GPDM and f) TLE.

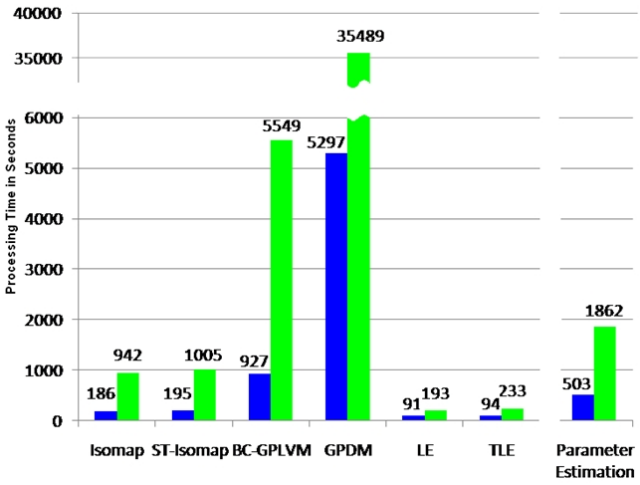

Figure 3. Training times based on either 1 (blue) or 2 subject (green) walking sequences (parameter estimation is manual for all embedding-based methods).

power of TLE, we produce a single 4-dimension descriptor per action instead of the action and subject dependent descriptors required by the standard framework [15]. Finally, action classification is performed by applying the sum rule of the following three metrics: the modified Hausdorff distance [17], curve dissimilarity function [18] and optical flow variation.

Performance of our system is evaluated using the Weizmann human action dataset [17] which consists of 9 different subjects repeating several times 10 actions. This provides 240 instances of simple motions, such as bending and waving. Quantitative results are calculated using the popular nine-fold cross validation schema already used by [17], [19], [20].

Action recognition results are presented in Table II. Usage of TLE improves accuracy of the standard framework [15] to 100% which is achieved by the most recent state of the art methods. Since TLE's generalisation property handles stylistic variations displayed by different people, this scheme is scalable to a larger subject population.

Table II
COMPARISON TO PREVIOUS RESULTS ON THE WEIZMANN DATASET.

Name	Accuracy	Comments
TLE + [15]	100%	Model per action
Blackburn [15]	95%	Model per action per subject
Blank [17]	100%	No action model
Yeffet [19]	100%	Model per action
Schindler [20]	100%	Model per action
Jhuang [21]	98.8%	Model per action
Wang [22]	97.8%	Model per action

V. CONCLUSION

This paper introduces a novel embedded-based dimensionality reduction approach, temporal LE, dedicated to time series. Its main contribution is the inclusion of temporal information including repetitions into the LE framework without requiring the manual tuning of parameters. As demonstrated in 3D pose recovery and action recognition applications, TLE ensures temporal coherence which improves the generalisation properties of the produced embedded spaces. In addition, the method is computationally efficient, which provides the data scalability which lacks from mapping-based dimensionality reduction approaches.

ACKNOWLEDGMENT

The authors would like to thank Odest Chadwicke Jenkins from Brown University and Eraldo Ribeiro from Florida Institute of Technology for helpful discussions and sharing their codes.

REFERENCES

[1] N. Lawrence, "Gaussian process latent variable models for visualisation of high dimensional data," *Proc. NIPS*, vol. 16, 2004.

[2] M. Belkin and P. Niyogi, "Laplacian eigenmaps and spectral techniques for embedding and clustering," *Proc. NIPS*, vol. 14, pp. 585–591, 2001.

[3] J. Tenenbaum, V. Silva, and J. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, no. 5500, pp. 2319–2323, 2000.

[4] O. Jenkins and M. Matarić, "A spatio-temporal extension to isomap nonlinear dimension reduction," *Proc. ICML*, pp. 441–448, 2004.

[5] N. Lawrence and J. Quinonero-Candela, "Local Distance Preservation in the GP-LVM Through Back Constraints," *Proc. ICML*, pp. 513–520, 2006.

[6] J. Wang, D. Fleet, and A. Hertzmann, "Gaussian process dynamical models," *Proc. NIPS*, vol. 18, pp. 1441–1448, 2006.

[7] L. Rabiner and B.-H. Juang, *Fundamentals of Speech Recognition*. Prentice-Hall, Inc., 1993.

[8] P. Kuo, J. Nebel, and D. Makris, "Camera Auto-Calibration from Articulated Motion," *Proc. AVSS*, pp. 135–140, 2007.

[9] M. Lee and I. Cohen, "A Model-Based Approach for Estimating Human 3D Poses in Static Images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 6, 2006.

[10] G. Mori, X. Ren, A. Efros, and J. Malik, "Recovering Human Body Configurations: Combining Segmentation and Recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 7, 2006.

[11] R. W. Poppe, *Discriminative Vision-Based Recovery and Recognition of Human Motion*. PhD Thesis, 2009.

[12] L. Sigal and M. Black, "HumanEva: Synchronized Video and Motion Capture Dataset for Evaluation of Articulated Human Motion," *Brown Univertsity*, 2006.

[13] K. Fukunaga and D. Olsen, "An algorithm for finding intrinsic dimensionality of data," *IEEE Trans. on Computers*, vol. C-20, no. 2, pp. 176–183, 1971.

[14] M. Lewandowski, D. Makris, and J.-C. Nebel, "Automatic Configuration of Spectral Dimensionality Reduction Methods for 3D Human Pose Estimation," *Workshop on Visual Surveillance*, 2009.

[15] J. Blackburn and E. Ribeiro, "Human Motion Recognition Using Isomap and Dynamic Time Warping," *Lecture Notes in Computer Science*, vol. 4814, pp. 285–298, 2007.

[16] B. Lucas and T. Kanade, "An Iterative Image Registration Technique with an Application to Stereo Vision," *Proc. IJCAI*, vol. 3, p. 3, 1981.

[17] L. Gorelick, M. Blank, E. Shechtman, M. Irani, and R. Basri, "Actions as Space-Time Shapes," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 12, p. 2247, 2007.

[18] M. Frenkel and R. Basri, "Curve Matching Using the Fast Marching Method," *EMMCVPR*, pp. 35–51, 2003.

[19] L. Yeffet and L. Wolf, "Local Trinary Patterns for Human Action Recognition," *Proc. ICCV*, 2009.

[20] K. Schindler and L. van Gool, "Action Snippets: How Many Frames Does Human Action Recognition Require?" *Proc. CVPR*, pp. 1–8, 2008.

[21] H. Jhuang, T. Serre, L. Wolf, and T. Poggio, "A Biologically Inspired System for Action Recognition," *Proc. ICCV*, vol. 1, no. 2, pp. 1–8, 2007.

[22] L. Wang and D. Suter, "Recognizing Human Activities from silhouettes: Motion subspace and Factorial Discriminative Graphical Model," *Proc. CVPR*, pp. 1–8, 2007.