

# Probabilistic Feature Extraction from Multivariate Time Series using Spatio-Temporal Constraints

Michał Lewandowski, Dimitrios Makris, and Jean-Christophe Nebel

Digital Imaging Research Centre, Kingston University, London, United Kingdom  
{m.lewandowski,d.makris,j.nebel}@kingston.ac.uk

**Abstract.** A novel nonlinear probabilistic feature extraction method, called Spatio-Temporal Gaussian Process Latent Variable Model, is introduced to discover generalised and continuous low dimensional representation of multivariate time series data in the presence of stylistic variations. This is achieved by incorporating a new spatio-temporal constraining prior over latent spaces within the likelihood optimisation of Gaussian Process Latent Variable Models (GPLVM). As a result, the core pattern of multivariate time series is extracted, whereas a style variability is marginalised. We validate the method by qualitative comparison of different GPLVM variants with their proposed spatio-temporal versions. In addition we provide quantitative results on a classification application, i.e. view-invariant action recognition, where imposing spatio-temporal constraints is essential. Performance analysis reveals that our spatio-temporal framework outperforms the state of the art.

## 1 Introduction

A multivariate time series (MTS) is a sequential collection of high dimensional observations generated by a dynamical system. The high dimensionality of MTS creates challenges for machine learning and data mining algorithms. To tackle this, feature extraction techniques are required to obtain computationally efficient and compact representations.

Gaussian Process Latent Variable Model [5] (GPLVM) is one of the most powerful nonlinear feature extraction algorithms. GPLVM emerged in 2004 and instantly made a breakthrough in dimensionality reduction research. The novelty of this approach is that, in addition to the optimisation of low dimensional coordinates during the dimensionality reduction process as other methods did, it marginalises parameters of a smooth and nonlinear mapping function from low to high dimensional space. As a consequence, GPLVM defines a continuous low dimensional representation of high dimensional data, which is called latent space. Since GPLVM is a very flexible approach, it has been successfully applied in a range of application domains including pose recovery [2], human tracking [20], computer animation [21], data visualization [5] and classification [19].

However, extensive study of the GPLVM framework has revealed some essential limitations of the basic algorithm [5, 6, 8, 12, 19, 21, 22]. First, since GPLVM

aims at retaining data global structure in the latent space, there is no guarantee that local features are preserved. As a result, the natural topology of the data manifold may not be maintained. This is particularly problematic when data, such as MTS, have a strong and meaningful intrinsic structure. In addition, when data are captured from different sources, even after normalisation, GPLVM tends to produce latent spaces which fail to represent common local features [8, 21]. This prevents successful utilisation of the GPLVM framework for feature extraction of MTS. In particular, GPLVM cannot be applied in many classification applications such as speech and action recognition, where latent spaces should be inferred from time series generated by different subjects and used to classify data produced by unknown individuals. Another drawback of GPLVM is its computationally expensive learning process [5, 6, 21] which may converge towards a local minima if the initialization of the model is poor [21].

Although recent extensions of GPLVM, i.e. back constrained GPLVM [12] (BC-GPLVM) and Gaussian Process Dynamical Model [22] (GPDM), allow satisfactory representation of time series, the creation of generalised latent spaces from data issued from several sources is still an unsolved problem which has never been addressed by the research community. In this paper, we define ‘style’ as the data variations between two or more datasets representing a similar phenomenon. They can be produced by different sources and/or repetitions from a single source. Here, we propose an extension of the GPLVM framework, i.e. Spatio-Temporal GPLVM (ST-GPLVM), which produces generalised and probabilistic representation of MTS in the presence of stylistic variations. Our main contribution is the integration of a spatio-temporal ‘constraining’ prior distribution over the latent space within the optimisation process.

After a brief review of the state of art, we introduce the proposed methodology. Then, we validate qualitatively our method on a real dataset of human behavioral time series. Afterwards, we apply our method to a challenging view independent action recognition task. Finally, conclusions are presented.

## 2 Related work

Feature extraction methods can be divided into two general categories, i.e., deterministic and probabilistic frameworks. The deterministic methods can be further classified into two main classes: linear and non linear methods. Linear methods like PCA cannot model the curvature and nonlinear structures embedded in observed spaces. As a consequence, nonlinear methods, such as Isomap [17], locally linear embedding [13] (LLE), Laplacian Eigenmaps [1] (LE) and kernel PCA [14], were proposed to address this issue. Isomap, LLE and LE aim at preserving a specific geometrical property of the underlying manifold by constructing graphs which encapsulate nonlinear relationships between points. However, they do not provide any mapping function between spaces. In contrast, kernel PCA obtains embedded space through nonlinear kernel based mapping from a high to low dimensional space. In order to deal with MTS, extensions of Isomap [3], LE [8] and the kernel based approach [15] were proposed.

Since previously described methods do not model uncertainty, another class of feature extraction methods evolved, the so-called latent variable models (LVM). They define a probability distribution over the observed space that is conditioned on a latent variable and mapping parameters. Consequently, it produces a probabilistic generative mapping from the latent space to the data space. In order to address intrinsic limitations of probabilistic linear methods, such as probabilistic principal component analysis [18] (PPCA), Lawrence [5] reformulated the PPCA model to the nonlinear GPLVM by establishing nonlinear mappings from the latent variable space to the observed space. From a Bayesian perspective, the Gaussian process prior is placed over these mappings rather than the latent variables with a nonlinear covariance function. As a result, GPLVM produces a complete joint probability distribution over latent and observed variables.

Recently, many researchers have exploited GPLVM in a variety of applications [2, 5, 19–21], thus designing a number of GPLVM-based extensions which address some of the limitations of standard GPLVM. Lawrence [5, 6] proposed to use sparse approximation of the full Gaussian process which allows decreasing the complexity of the learning process. Preservation of observed space topology was supported by imposing high dimensional constraints on the latent space [12, 21]. BC-GPLVM [12] enforces local distance preservation through the form of a kernel-based regression mapping from the observed space to the latent space. Locally linear GPLVM [21] (LL-GPLVM) extends this concept by defining explicitly a cylindrical topology to maintain. This is achieved by constructing advanced similarity measures for the back constrained mapping function and incorporation of the LLE objective function [13] into the GPLVM framework to reflect a domain specific prior knowledge about observed data. Another line of work, i.e. GPDM [22], augments GPLVM with a nonlinear dynamical mapping on the latent space based on the auto-regressive model to take advantage of temporal information provided with time series data.

Current GPLVM based approaches have proven very effective when modelling of MTS variability is desired in the latent space. However, these methods are inappropriate in a context of recognition based applications where the discovery of a common content pattern is more valuable than modelling stylistic variations. In this work, we tackle this fundamental problem by introducing the idea of a spatio-temporal interpretation of GPLVM. This concept is formulated by incorporating a constraining prior distribution over the latent space in the GPLVM framework. In contrast to LL-GPLVM and BC-GPLVM, we aim at implicitly preserving a spatio-temporal structure of the observed time series data in order to discard style variability and discover a unique low dimensional representation for a given set of MTS. The proposed extension is easily adaptable to any variant of GPLVM, for instance BC-GPLVM or GPDM.

### 3 Methodology

Let a set of multivariate time series  $Y$  consists of multiple repetitions (or cycles) of the same phenomenon from the same or different sources and all data points

$\{y_i\}_{(i=1..N)}$  in this set are distributed on a manifold in a high dimensional space ( $y_i \in R^D$ ). ST-GPLVM is able to discover low dimensional representation  $X = \{x_i\}_{(i=1..N)}$  ( $x_i \in R^d$  with  $d \ll D$ ) by giving a Gaussian process prior to mapping functions from the latent variable space  $X$  to the observed space  $Y$  under a constraint  $L$  to preserve the spatio-temporal patterns of the underlying manifold. The entire process is summarized in figure 1.

Initially the spatio-temporal constraints  $L$  are constructed. These constraints are exploited twofold. First they are used to better initialise the model by discovering a low dimensional embedded space which is close to the expected representation. Secondly, they constrain the GPLVM optimisation process so that it converges faster and maintains the spatio-temporal topology of the data. The learning process is performed using the standard two stage maximum a posteriori (MAP) estimation used in GPLVM. Latent positions  $X$ , and the hyperparameters  $\Phi$  are optimised iteratively until the optimal solution is reached under the introduced constraining prior  $p(X|L)$ . The key novelty of the proposed methodology is its style generalisation potential. ST-GPLVM discovers a coherent and continuous low dimensional representation by identifying common spatio-temporal patterns which result in discarding style variability among all conceptually similar time series.

### 3.1 Gaussian Process Latent Variable Model

GPLVM [5] was derived from the observation that a particular probabilistic interpretation of PCA is a product of Gaussian Process (GP) models, where each of them is associated with a linear covariance function (i.e. kernel function). Consequently, the design of a non-linear probabilistic model could be achieved by replacing the linear kernel function with a non-linear covariance function. From a Bayesian perspective, by marginalising over the mapping function [5], the complete joint likelihood of all observed data dimensions given the latent positions is:

$$p(Y|X, \Phi) = \frac{1}{(2\pi)^{DN/2} |K|^{D/2}} \exp(-0.5 \text{tr}(K^{-1} Y Y^T)) \quad (1)$$

where  $\Phi$  denotes the kernel hyperparameters and  $K$  is the kernel matrix of the GP model which is assembled with a combination of covariance functions:  $K = \{k(x_i, x_j)\}_{(i,j=1..N)}$ . The covariance function is usually expressed by the nonlinear radial basis function (RBF):

$$k(x_i, x_j) = \alpha \exp(-\frac{\gamma}{2} \|x_i - x_j\|^2) + \beta^{-1} \delta_{x_i x_j} \quad (2)$$

where the kernel hyperparameters  $\Phi = \{\alpha, \beta, \gamma\}$  respectively determine the output variance, the variance of the additive noise and the RBF width.  $\delta_{x_i x_j}$  is the Kronecker delta function.

Learning is performed using two stage MAP estimation. First, latent variables are initialized, usually using PPCA. Secondly, latent positions and the

hyperparameters are optimised iteratively until the optimal solution is reached. This can be achieved by maximising the likelihood (1) with respect to the latent positions,  $X$ , and the hyperparameters,  $\Phi$  using the following posterior:

$$p(X, \Phi | Y) \propto p(Y | X, \Phi) p(X) p(\Phi) \quad (3)$$

where priors of the unknowns are:  $p(X) = N(0, I)$ ,  $p(\Phi_i) \propto \prod_i \Phi_i^{-1}$ . The maximisation of the above posterior is equivalent to minimising the negative log posterior of the model:

$$-\ln p(X, \Phi | Y) = 0.5((DN+1) \ln 2\pi + D \ln |K| + \text{tr}(K^{-1}YY^T) + \sum_i \|x_i\|^2) + \sum_i \Phi_i \quad (4)$$

This optimization process can be achieved numerically using the scaled conjugate gradient [11] (SCG) method with respect to  $\Phi$  and  $X$ . However, the learning process is computationally very expensive, since  $O(N^3)$  operations are required in each gradient step to inverse the kernel matrix  $K$  [5]. Therefore, in practice, a sparse approximation to the full Gaussian process, such as 'fully independent training conditional' (FITC) approximation [6] or active set selection [5], is exploited to reduce the computational complexity to a more manageable  $O(k^2N)$  where  $k$  is the number of points involved in the lower rank sparse approximation of the covariance [6].

### 3.2 Spatio-Temporal Gaussian Process Latent Variable Model

The proposed ST-GPLVM relies on the novel concept of a spatio-temporal constraining prior which is introduced into the standard GPLVM framework in order to maintain temporal coherence and marginalise style variability. This is achieved by designing an objective function, where the prior  $p(X)$  in Eq. 3 is replaced by the proposed conditioned prior  $p(X|L)$ :

$$p(X, \Phi | Y, L) \propto p(Y | X, \Phi) p(X | L) p(\Phi) \quad (5)$$

where  $L$  denotes the spatio-temporal constraints imposed on the latent space. Although  $p(X|L)$  is not a proper prior, conceptually it can be seen as an equivalent of a prior for a given set of weights  $L$  [21]. These constraints are derived from graph theory, since neighbourhood graphs have been powerful in designing nonlinear geometrical constraints for dimensionality reduction using spectral based approaches [1, 13, 17]. In particular, the Laplacian graph allows preserving approximated distances between all data points in the low dimensional space [1]. This formulation is extensively exploited in our approach by constructing cost matrix,  $L$ , which emphasizes spatio-temporal dependencies between similar time series. This is achieved by designing two types of neighbourhood for each high dimensional data point  $P_i$  (figure 2):

- Temporal neighbours (T): the  $2m$  closest points in the sequential order of input (figure 2a):  $T_i \in \{P_{i-m}, \dots, P_{i-1}, P_i, P_{i+1}, \dots, P_{i+m}\}$

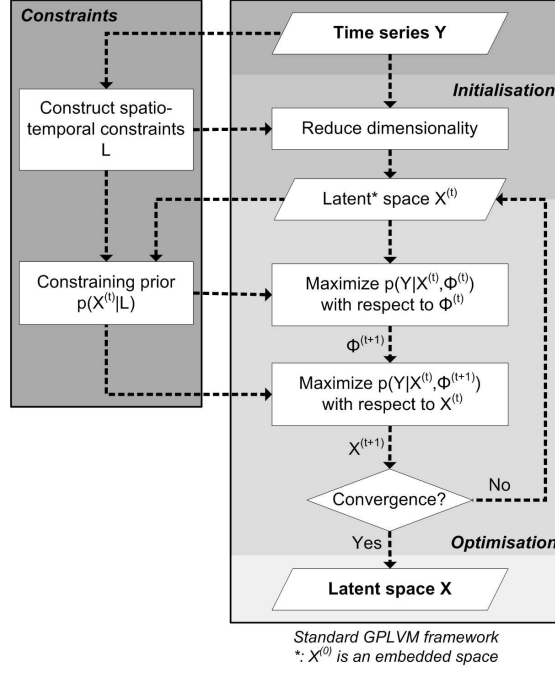
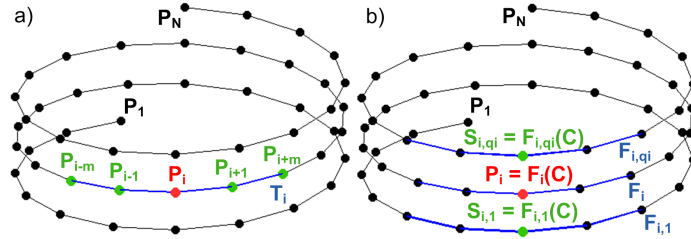


Fig. 1: Spatio-Temporal Gaussian Process Latent Variable Model pipeline.

- Spatial neighbours (S): let's associate to each point,  $P_i$ ,  $2s$  temporal neighbours which define a time series fragment  $F_i$ . The spatial neighbours,  $S_i$ , of  $P_i$  are the centres of the  $q_i$  time series fragments,  $F_{ik}$ , which are similar to  $F_i$  (figure 2b):  $S_i \in \{F_{i,1}(C), \dots, F_{i,q_i}(C)\}$ . Here  $F_{i,k}(C)$  returns the centre point of  $F_{i,k}$ . The neighbourhood  $S_i$  is determined automatically using either dynamic time warping [8] or motion pattern detection [7].

Fig. 2: Temporal (a) and spatial (b) neighbours (green dots) of a given data point,  $P_i$ , (red dots).

Neighbourhood connections defined in the Laplacian graphs implicitly impose points closeness in the latent space. Consequently, the temporal neighbours allow to model a temporal continuity of MTS, whereas spatial neighbours remove style variability by aligning MTS in the latent space.

The constraint matrix,  $L$ , is obtained, first, by assigning weights,  $W$ , to the edges of each graph,  $G \in \{T, S\}$ , using the standard LE heat kernel [1]:

$$W_{ij}^G = \begin{cases} \exp(\|y_i - y_j\|^2) & \text{i,j connected} \\ 0 & \text{otherwise} \end{cases} \quad (6)$$

Then, information from both graphs are combined  $L = L_T + L_S$  where  $L_G = D^G - W^G$  is the Laplacian matrix.  $D^G = \text{diag}\{D_{11}^G, D_{22}^G, \dots, D_{NN}^G\}$  denotes a diagonal matrix with entries:  $D_{ii}^G = \sum_{j=1}^N W_{ij}^G$ . The prior probability of the latent variables, which forces each latent point to preserve the spatio-temporal topology of the observed data, is expressed by:

$$p(X|L) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\text{tr}(X L X^T)}{2\sigma^2}\right) \quad (7)$$

where  $\sigma$  represents a global scaling of the prior and controls the 'strength' of the constraining prior. Note that although distance between neighbours (especially spatial ones) may be large in  $L$ , it is infinite between unconnected points.

The maximisation of the new objective function (5) is equivalent to minimising the negative log posterior of the model:

$$-\ln p(X, \Phi|Y, L) = 0.5(D \ln |K| + \text{tr}(K^{-1} Y Y^T) + \sigma^{-2} \text{tr}(X L X^T) + C) + \sum_i \Phi_i \quad (8)$$

where  $C$  is a constant:  $(DN + 1) \ln 2\pi + \ln \sigma^2$ . Following the standard GPLVM approach, the learning process involves minimising Eq. 8 with respect to  $\Phi$  and  $X$  iteratively using SCG method [11] until convergence.

ST-GPLVM is initialised using a nonlinear feature extraction method, i.e. temporal LE [8] which is able to preserve the constraints  $L$  in the produced embedded space. Consequently, compared to the standard usage of linear PPCA, initialisation is more likely to be closer to the global optimum. In addition, the enhancement of the objective function (3) with the prior (7) constrains the optimisation process and therefore further mitigates the problem of local minima. The topological structure in terms of spatio-temporal dependencies is implicitly preserved in the latent space without enforcing any domain specific prior knowledge.

The proposed methodology can be applied to other GPLVM based approaches, such as BC-GPLVM [12] and GPDM [22] by integrating the prior (7) in their cost function. The extension of BC-GPLVM results in a spatio-temporal model which provides bidirectional mapping between latent and high dimensional spaces. Alternatively, ST-GPDM produces a spatio-temporal model with an associated nonlinear dynamical process in the latent space. Finally, the proposed extension is compatible with a sparse approximation of the full Gaussian process [5, 6] which allows reducing further processing complexity.

#### 4 Validation of ST-GPLVM approach

Our new approach is evaluated qualitatively through a comparative analysis of latent spaces discovered by standard non-linear probabilistic latent variable models, i.e. GPLVM, BC-GPLVM and GPDM and their extensions, i.e., ST-GPLVM, ST-BC-GPLVM and ST-GPDM, where the proposed spatio-temporal constraints have been included.

Our evaluation is conducted using time series of MoCap data, i.e. repeated actions provided by the HumanEva dataset [16]. The MoCap time series are firstly converted into normalized sequences of poses, i.e. invariant to subject's rotation and translation. Then each pose is represented as a set of quaternions, i.e. a 52-dimension feature vector. In this experiment, we consider three different subjects performing a walking action comprising of 500 frames each. The dimensionality of walking action space is reduced to 3 dimensions [20,22]. During the learning process, the computational complexity is reduced using FITC [6] where the number of inducing variables is set to 10% of the data. The global scaling of the constraining prior,  $\sigma$ , and the width of the back constrained kernel [12] were

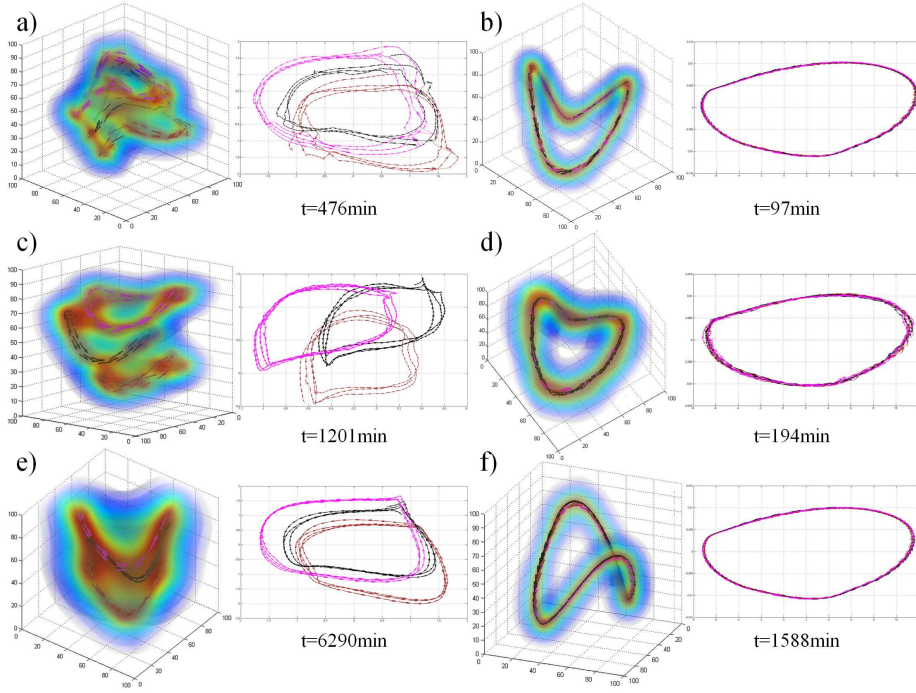


Fig. 3: 3D models learned from walking sequences of 3 different subjects with corresponding first 2 dimensions and processing times: a) GPLVM, b) ST-GPLVM, c) BC-GPLVM, d) ST-BC-GPLVM, e) GPDM and f) ST-GPDM. Warm-coloured regions correspond to high reconstruction certainty.



set empirically to  $10^4$  and 0.1 respectively. Values of all the other parameters of the models were estimated automatically using maximum likelihood optimisation. The back constrained models used a RBF kernel [12].

The learned latent spaces for the walking sequences with the corresponding first two dimensions and processing times are presented in figure 3. Qualitative analysis confirms the generalisation property of the proposed extension. Standard GPLVM based approaches discriminate between subjects in the spatially distinct latent space regions. Moreover, action repetitions by a given subject are represented separately. In contrast, the introduction of our spatio-temporal constraint in objective functions allows producing consistent and smooth representation by discarding style variability in all considered models. In addition, the extended algorithms converge significantly faster than standard versions. Here, we achieve a speed-up of a factor 4 to 6.

## 5 Application of ST-GPLVM to activity recognition

We demonstrate the effectiveness of the novel methodology in a realistic computer vision application by integrating ST-GPLVM within the view independent human action recognition framework proposed in [7]. Here, the training data comprises of time series of action images obtained from evenly spaced views located around the vertical axis of a subject. In order to deal with this complicated scenario, the introduced methodology is extended by a new initialisation procedure and a new advanced constraining prior. The learning process of action recognition framework is summarised in figure 4.

First, for each view ( $z=1..Z$ ), silhouettes are extracted from videos, normalized and represented as 3364-dimensional vectors of local space-time saliency features [7]. Then, spatio-temporal constraints  $L_z$  are calculated. During initialisation, style invariant one-dimensional action manifolds  $X_z$  are obtained using temporal LE [8] and subsequently all these view-dependent models are combined to generate a coherent view invariant representation of the action [7]. The outcome of this procedure reveals a torus-like structure which is used to initialise GPLVM and encapsulates both style and view. Finally, the latent space and parameters of model are optimised jointly under a new combined prior  $p(X|L)$ . This prior is derived by taking into account constraints associated with each view:

$$p(X|L) = \prod_{z=1}^Z \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\text{tr}(X_z L_z X_z^T)}{2\sigma^2}\right) \quad (9)$$

where  $L$  is a block diagonal matrix formed by all  $L_z$ . Action recognition is performed by maximum likelihood estimation. Performance of the system is evaluated using the multi-view IXMAS dataset [23] which is considered as the benchmark for view independent action recognition. This dataset is comprised of 12 actions which are performed 3 times by 12 different actors. In this dataset, actors' positions and orientations in videos are arbitrary since no specific instruction was given during acquisition. As a consequence, the action viewpoints are

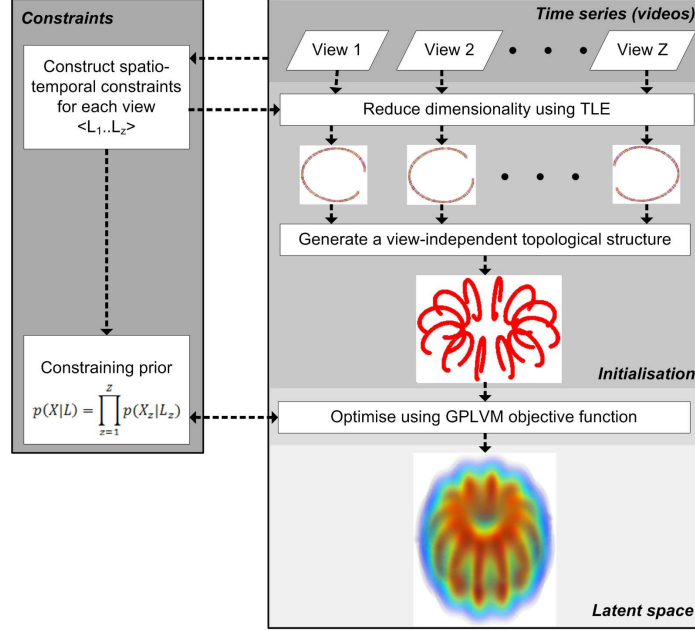


Fig. 4: Pipeline for generation of probabilistic view and style invariant action descriptor.

arbitrary and unknown. Here, we use one action repetition of each subject for training, whereas testing is performed with all action repetitions. Experiments are conducted using the popular leave-one-out schema [4,23,24]. Two recognition tasks were evaluated using either a single view or multiple views. In line with other experiments made on this dataset [9,10,24], the top view was discarded for testing. The global scaling of the constraining prior and the number of inducing variables in FITC [6] were set to  $10^4$  and 25% of the data respectively. Values of all the other parameters of the models were estimated automatically using maximum likelihood optimisation.

Action recognition results are compared with the state of the art in table 1 (top view excluded). Examples of learned view and style invariant action descriptors using ST-GPLVM are shown in figure 5. Although different approaches may use slightly different experimental settings, table 1 shows that our framework produces the best performances. In particular, it improves the accuracy of the standard framework [7]. The confusion matrix of recognition for the 'all-view' experiment reveals that our framework performed better when dealing with motions involving the whole body, i.e. "walk", "sit down", "get up", "turn around" and "pick up". As expected, the best recognition rates 78.7%, 80.7% are obtained for camera 2 and 4 respectively, since those views are similar to those used for training, i.e. side views. Moreover, when dealing with either different,

Table 1: Left, average recognition accuracy over all cameras using either single or multiple views for testing. Right, class-confusion matrix using multiple views.

%	Subjects /Actions	Average accuracy	
		Single view	All views
Weinland [23]	10 / 11	63.9	81.3
Yan [24]	12 / 11	64.0	78.0
Junejo [4]	10 / 11	74.1	-
Liu [9]	12 / 13	71.7	78.5
Liu [10]	12 / 13	73.7	82.8
Lewandowski [7]	12 / 12	73.2	83.1
<b>Our</b>	12 / 12	<b>76.1</b>	<b>85.4</b>

check watch	0.83	0.14	0.03	0	0	0	0	0	0	0	0
cross arms	0.17	0.69	0.11	0	0	0	0	0.03	0	0	0
scratch head	0.03	0.03	0.72	0	0	0	0	0.14	0	0	0
sit down	0	0	0	0.97	0	0	0	0	0	0	0.03
get up	0	0	0	0	1.00	0	0	0	0	0	0
turn around	0	0	0	0	0	1.00	0	0	0	0	0
walk	0	0	0	0	0	0	1.00	0	0	0	0
wave hand	0	0.03	0.19	0	0	0	0	0.67	0	0	0.03
punch	0.06	0	0	0	0	0	0	0	0.75	0	0.17
kick	0	0	0	0	0	0	0	0.08	0.89	0	0
point	0	0	0	0	0	0	0	0	0.19	0	0.81
pick up	0	0	0	0.08	0	0	0	0	0	0	0.92

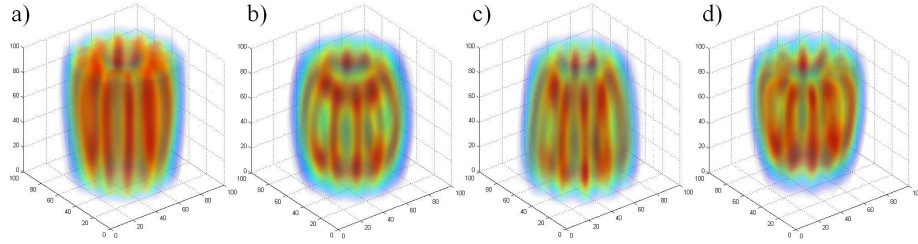


Fig.5: Probabilistic view and style invariant action descriptors obtained using ST-GPLVM for a) sit down, b) cross arms, c) turn around and d) kick.

i.e. camera 1, or even significantly different views, i.e. camera 3, our framework still achieves good recognition rate, i.e. 75.2% and 69.9% respectively.

## 6 Conclusion

This paper introduces a novel probabilistic approach for nonlinear feature extraction called Spatio-Temporal GPLVM. Its main contribution is the inclusion of spatio-temporal constraints in the form of a conditioned prior into the standard GPLVM framework in order to discover generalised latent spaces of MTS. All conducted experiments confirm the generalisation power of the proposed concept in the context of classification applications where marginalising style variability is crucial. We applied the proposed extension on different GPLVM variants and demonstrated that their Spatio-Temporal versions produce smoother, coherent and visually more convincing descriptors at a lower computational cost. In addition, the methodology has been validated in a view independent action recognition framework and produced state of the art accuracy. Consequently, the concept of consistent representation of time series should benefit to many other applications beyond action recognition such as gesture, sign-language and speech recognition.

## References

1. Belkin, M., Niyogi, P.: Laplacian eigenmaps and spectral techniques for embedding and clustering. *Proc. NIPS* 14, 585–591 (2001)
2. Ek, C., Torr, P., Lawrence, N.: Gaussian process latent variable models for human pose estimation. *Machine Learning for Multimodal Interaction* pp. 132–143 (2007)
3. Jenkins, O., Matarić, M.: A spatio-temporal extension to isomap nonlinear dimension reduction. *Proc. ICML* pp. 441–448 (2004)
4. Junejo, I., Dexter, E., Laptev, I., Pérez, P.: Cross-view action recognition from temporal self-similarities. *Proc. ECCV* 12 (2008)
5. Lawrence, N.: Gaussian process latent variable models for visualisation of high dimensional data. *Proc. NIPS* 16 (2004)
6. Lawrence, N.: Learning for larger datasets with the Gaussian process latent variable model. *Proc. AISTATS* (2007)
7. Lewandowski, J., Makris, D., Nebel, J.C.: View and style-independent action manifolds for human activity recognition. *Proc. ECCV* 6316 (2010)
8. Lewandowski, M., Martinez-del Rincon, J., Makris, D., Nebel, J.C.: Temporal extension of laplacian eigenmaps for unsupervised dimensionality reduction of time series. *Proc. ICPR* (2010)
9. Liu, J., Ali, S., Shah, M.: Recognizing human actions using multiple features. *Proc. CVPR* (2008)
10. Liu, J., Shah, M.: Learning human actions via information maximization. *Proc. CVPR* (2008)
11. Möller, M.: A scaled conjugate gradient algorithm for fast supervised learning. *Neural Networks* 6(4), 525–533 (1993)
12. N.D., L., Quinonero-Candela, J.: Local Distance Preservation in the GP-LVM Through Back Constraints. *Proc. ICML* pp. 513–520 (2006)
13. Roweis, S., Saul, L.: Nonlinear dimensionality reduction by locally linear embedding. *Science* 290(5500), 2323–2326 (2000)
14. Schölkopf, B., Smola, A., Müller, K.: Kernel principal component analysis. *ICANN* pp. 583–588 (1997)
15. Shyr, A., Urtasun, R., Jordan, M.: Sufficient dimension reduction for visual sequence classification. *Proc. CVPR* (2010)
16. Sigal, L., Black, M.: HumanEva: Synchronized Video and Motion Capture Dataset for Evaluation of Articulated Human Motion. Brown University (2006)
17. Tenenbaum, J., Silva, V., Langford, J.: A global geometric framework for nonlinear dimensionality reduction. *Science* 290(5500), 2319–2323 (2000)
18. Tipping, M., Bishop, C.: Probabilistic principal component analysis. *Journal of the Royal Statistical Society, Series B* 61, 611–622 (1999)
19. Urtasun, R., Darrell, T.: Discriminative Gaussian process latent variable model for classification. *Proc. ICML* pp. 927–934 (2007)
20. Urtasun, R., Fleet, D., Fua, P.: 3d people tracking with gaussian process dynamical models. *Proc. CVPR* 1, 238–245 (2006)
21. Urtasun, R., Fleet, D., Geiger, A., Popović, J., Darrell, T., Lawrence, N.: Topologically-constrained latent variable models. *Proc. ICML* (2008)
22. Wang, J., Fleet, D., Hertzmann, A.: Gaussian process dynamical models. *Proc. NIPS* 18, 1441–1448 (2006)
23. Weinland, D., Boyer, E., Ronfard, R.: Action recognition from arbitrary views using 3d exemplars. *Proc. ICCV* 5(7), 8 (2007)
24. Yan, P., Khan, S., Shah, M.: Learning 4D action feature models for arbitrary view action recognition. *Proc. CVPR* 12 (2008)