2014 IEEE International Conference on Acoustic, Speech and Signal Processing (ICASSP)

FUNCTIONAL RELEVANT MULTICHANNEL KERNEL ADAPTIVE FILTER FOR HUMAN ACTIVITY ANALYSIS

A. Álvarez-Meza, G. Castellanos-Domínguez *

Signal processing and Recognition Group
Universidad Nacional, Manizales-Colombia

J.C. Príncipe

Dep. of Electrical and Computer Engineering
University of Florida, Gainesville, FL

ABSTRACT

A multichannel kernel adaptive filtering framework is presented that highlights relevant channels for the task of analyzing Motion Capture (MoCap) data. Functional relevance analysis is performed over input multichannel data by computing the pair-wise channel similarities to describe the main behavior of the considered applications. Particularly, the well-known Kernel Least Mean Square filter is enhanced using a correntropy-based similarity criterion between channel pairs. Besides, two sparseness criteria are studied to extract a sample subset that constructs a learning model displaying a good trade-off between filter complexity and accuracy. The proposed approach allows devising complex relationship among multi-channel time-series, revealing dependencies among the channels and the process time-structure. The method is tested in a well-known MoCap data set. Results show that our framework is an adequate alternative for finding functional relevance amongst multi-channel time-series.

Index Terms— multichannel data, functional relevance, adaptive filtering, MoCap data.

1. INTRODUCTION

In machine learning applications it is difficult to interpret the available information due to its complexity and its large amount of extracted features. Mostly, input data hold different structures varying over time. So, along with spatial statistical relationship, we have to deal with data time structure. Besides, the instantaneous random variables are hardly ever independently distributed, i.e., stochastic processes possess a time structure [1]. Moreover, since input data often consist of multi-channel time-series, there is a need for making clear inter–channel relationship. Regarding this, complete information to solve such kind of problems resides in the joint probability density function of multivariate data, but it is never done due to its high-dimensionality [2]. On the other hand, approaches based on kernel adaptive filtering aim to reveal nonlinear structure of the time-series by means of kernel functions, however, for multi-channel series, they commonly assume inter channel independence [3, 4].

As a real-world application of interest, Motion Capture (MoCap) videos, which are related to human activity recordings, contain a set of channels interacting each other according to human behavior. Such a data is commonly used for nonmechanical analysis and for tracking human 3D articulated motions [5]. Particularly, estimating 3D location and orientation of the human joints is notoriously difficult, because it is a high-dimensional problem and is riddled with ambiguities coming from noise, monocular imagery, and occlusions. Some machine vision based approaches employ prior models to highlight such temporal and spatial relationships among time-series to enhance 3D prediction accuracy. Linear models (e.g. PCA) are among the simplest priors [6], however, they restrict model expressiveness leading into inaccuracies when learning complex motions. Other works aim to deal with such a complexity by means of non-linear dimensionality reduction techniques [7, 8], aiming to preserve the manifold local structure, but tend to fail when manifold assumptions are violated, e.g., in the presence of noise, or multiple activities. Additionally, other frameworks are based on probabilistic latent variable models [5], which impose some priors according to the estimated covariance that governs data distribution. Nonetheless, they require many samples to properly model the hidden process, generating complex functions that tend to infer biased models. Overall, when analyzing MoCap most of proposed algorithms do not consider directly both time-structure and joints dependencies of data, assuming that the channels are independent each other.

Here, inspired on the Kernel Least Mean Square - KLMS adaptive filter [4, 3], and using a correntropy based similarity function [9], a functional relevance framework for multichannel models learning online is introduced. KLMS is extended to take advantage of input data pair-wise similarities describing the main behavior of considered time-series. Additionally, two significance measures are studied: kernel and Minimum Description Length based to update the filter model along the time. Using those criteria, an input subset is extracted to construct a learning model that preserves a trade-off between filter complexity and accuracy. As the first approach,

*Under grants provided by a Ph.D. scholarship funded by Colciencias.

our algorithm is tested on a well-known MoCap database for tracking 3D human joints. Since our goal is to demonstrate the proposed approach capability to highlight the main human poses of a given video, we consider both the temporal data structure given by the adaptive filter scheme, and the inter-channel relationships given by the proposed functional relevance estimation stage.

2. MULTI-CHANNEL FUNCTIONAL RELEVANCE

Let $\boldsymbol{X} \in \Re^{p \times T}$ be a multi-channel input matrix, with p channels and T samples, where each row vector $\boldsymbol{x}$ of $\boldsymbol{X}$ can be treated as a real-valued time-series. Here, to discover both the spatial and temporal relationships among channels, each channel is analyzed using a sliding window of size L, where L is the embedding dimension [10]. Therefore, the embedding data representation $\boldsymbol{H} \in \Re^{p \times L \times n}$ is calculated as $\boldsymbol{h}_t^i = \left[X_{i,L(t-1)+1}, \ldots, X_{i,L(t-1)+L}\right]$, where $\boldsymbol{h}_t^i \in \Re^{1 \times L}$ is the i-th row of the t-th matrix $\boldsymbol{H}_t \in \Re^{p \times L}$ in $\boldsymbol{H}$, with $i = 1, \ldots, p$ and $t = 1, \ldots, n$ ($n = T/L$). So, our goal is to find from $\boldsymbol{H}$ the spatial relationships into each time embedding sliding window. To this end, given a matrix $\boldsymbol{H}_t$ with row vectors $\boldsymbol{h}_t^i$, the similarity among channels can be estimated by $\kappa_S(\boldsymbol{h}_t^i, \boldsymbol{h}_t^j) = \langle \phi(\boldsymbol{h}_t^i), \phi(\boldsymbol{h}_t^i) \rangle$, where $\kappa_S(\cdot,\cdot)$ is a kernel function [11]. Through the so called "kernel trick", $\phi(\cdot)$ may not need to directly computed. Then, the well-known Gaussian kernel is considered as $\kappa_S(\boldsymbol{h}_t^i, \boldsymbol{h}_t^j) = \exp(-\|\boldsymbol{h}_t^i - \boldsymbol{h}_t^j\|_2^2 / 2\sigma_S^2)$, being $\sigma_S \in \Re^+$ the kernel band-width. Therefore, the relationship among channels at instant t-th is extracted as a pair-wise feature representation, which is the multichannel relevant matrix $\boldsymbol{S} \in \Re^{p \times p \times n}$, with elements $S_t^{i,j} = \kappa_S(\boldsymbol{h}_t^i, \boldsymbol{h}_t^j)$, where $\boldsymbol{S}_t \in \Re^{p \times p}$ is the matrix of the t-th time window.

Regarding this, in multichannel activity data there are many channels that are not relevant for a given task, but become relevant for another. In this sense, instead of using all the channels all the time, it is preferable to highlight such pair-wise similarities from $\boldsymbol{S}$ that become relevant for the studied process. Thus, a measure of similarity on the spatio-temporal data based on correntropy is introduced [9]. Correntropy is a kind of localized measure to estimate how similar two random variables are: when two random variables are very close, correntropy equals the 2-norm distance, which evolves to 1-norm distance if two random variables get further apart, even falls to zero-norm as they are far apart. Hence, we propose to estimate the similarity between two spatio-temporal matrices as

$$\kappa_K(\boldsymbol{S}_t, \boldsymbol{S}_{t'}) = \frac{1}{p^2} \sum_{i,j} \exp\left(-\frac{|S_t^{i,j} - \boldsymbol{S}_{t'}^{i,j}|^2}{2\sigma_K^2}\right), \tag{1}$$

where $\sigma_K \in \Re^+$ and $S_t^{i,j}$ can be considered as a functional connection weight between channels i and j, and $\boldsymbol{S}_t$ is symmetric with respect to the main diagonal. Now, to highlight relevant functional connections among channels in an online fashion while considering the data temporal structure, a kernel adaptive filtering framework will be extended to deal with the correntropy based spatio-temporal relationships.

2.1. Kernel adaptive filtering for multichannel data

Given a sequence of input-output pairs $\{(\boldsymbol{S}_1, y_1), \ldots, (\boldsymbol{S}_T, y_T)\}$, and based on risk minimization analysis, our main goal is to compute the continuous mapping $y_t = f(\boldsymbol{S}_t)$, where $y_t \in \Re$ is a given output signal that is related to the interactions immersed in $\boldsymbol{S}_t$. Indeed, from the above proposed multichannel analysis, the learning function can be defined as $f : \mathbb{S} \times \mathbb{S} \to \Re$, where $\mathbb{S} \subseteq \Re^{p \times p}$. A kernel adaptive filter is a kernel sequential estimator of f, such that f_t is updated on the basis of the last estimate f_{t-1} and the current input-output pair $\{\boldsymbol{S}_t, y_t\}$ [3]. In this sense, Kernel Least Mean Square - KLMS appears as an extension of the well-known Least Mean Square - LMS based adaptive filter [4]. KLMS aims to exploit the kernel mapping from an input space to a RKHS, being able to deal with nonlinear relationships among samples. Thus, input data is mapped into RKHS, on which the LMS adaptive filtering is applied as

$$\begin{cases} f_1 = 0 \\ e_t = y_t - f_{t-1}(\boldsymbol{S}_t) \\ f_t = f_{t-1} + \eta e_t \kappa_K(\boldsymbol{S}_t, \cdot) \end{cases}, \tag{2}$$

where $\hat{y}_t = f_{t-1}(\boldsymbol{S}_t) = \sum_{r=1}^{t-1} \alpha^r \kappa_K(\boldsymbol{S}_t, \boldsymbol{S}_r)$, $0 < \eta < 1$ is the filter step size, and $\alpha^r = \eta e_r$. Here, $\kappa_K(\cdot,\cdot)$ is the proposed correntropy based similarity function in (1). Note that KLMS uses all learned observations to estimate the output of a new input, however, it involves a complex function that may lead to over fitting, not mentioning its high computational load. In this regard, two KLMS based quantization criteria will be extended to obtain an adequate trade-off between system complexity and accuracy performance.

The first quantization criterion is inspired on the Quantized Kernel Least Mean Square - QKLMS algorithm [3], which aims to discover the main model structure along the time by computing the Euclidean distance on the original input space between a given sample and the codebook. As an alternative, here we propose to estimate the similarity between the new sample and the system model taking advantage of the RKHS. So, let $\boldsymbol{C}_{t-1} \in \mathbb{S}^{c_{t-1}}$ be the KLMS system codebook at $t-1$ instant with network size c_{t-1}. Provided a new sample the quantization measure Ψ_K is estimated as

$$\Psi_K(\boldsymbol{S}_t, \boldsymbol{C}_{t-1}) = \max_r \kappa_K(\boldsymbol{S}_t, \boldsymbol{C}_{t-1}^r), \tag{3}$$

quantizing $\boldsymbol{S}_t$ by $\boldsymbol{C}_{t-1}^{r^*} \in \mathbb{S}$ based on the threshold $\gamma \in \Re^+$.

Now, the second quantization criteria is based on the Minimum Description Length - MDL as a measure of model simplicity, which implies the model with the least description length as the best [12]. We proposed to analyze both complexity and accuracy of the system by extending the MDL algorithm including functional relevant representations. Thus, given a new sample, the costs of adding this data into center dictionary and merging it into the nearest center are compared. For this purpose, a sliding window of size L_w is studied yielding the MDL based cost function Ψ_M as follows:

$$\Psi_M(\boldsymbol{S}_t,\boldsymbol{C}_{t-1}) = \log(L_w) + \log(\sum_{l=t-L_w}^{t} \hat{e}_l^2)^{\frac{L_w}{2}} - \log(\sum_{l=t-L_w}^{t} \bar{e}_l^2)^{\frac{L_w}{2}}, \tag{4}$$

being $\hat{e}_l = e'_l - \eta e_t \kappa_K(\boldsymbol{S}_t,\boldsymbol{S}_l)$ the prediction error after adding a new center and $\bar{e}_l = e'_l - \eta e_t \kappa_K(\boldsymbol{C}^{r^*}_{t-1},\boldsymbol{S}_l)$ is the error after merging the new data using (3). Note that MDL approximations are inferred according to the last model at t time instant $\{\boldsymbol{C}_{t-1},\boldsymbol{\alpha}_{t-1}\}$. Such a model is used to estimate the outputs into the subwindow, that is, $e'_l = y_l - \sum_{r=1}^{c_{t-1}} \alpha^r_{t-1}\kappa_K(\boldsymbol{S}_l,\boldsymbol{C}^r_{t-1})$. So, if $\Psi_M(\boldsymbol{S}_t,\boldsymbol{C}_{t-1}) > 0$ in (4), $\boldsymbol{S}_t$ must be merged, otherwise, it is added to the coodebook. In Algorithm 1, the proposed Functional Relevant based extension of Quantized Kernel Least Mean Square - FRKLMS is presented. Note that when MDL cost function is used, $\gamma = 0$.

Algorithm 1: FRKLMS

Input: $\boldsymbol{S}_t \in \mathbb{S}, y_t \in \Re$, $0 < \eta < 1$, $\sigma_K > 0, \gamma \geq 0$
Output: $\hat{y}_t \in \Re, \boldsymbol{C}_t \in \mathbb{S}^{c_t}, \alpha_t \in \Re^{c_t}$
$\boldsymbol{C}_1 = \{\boldsymbol{S}_1\}, \alpha_1 = \{\eta y_1\}$;
while $\{\boldsymbol{S}_t, y_t\}$ $(t > 1)$ available **do**
  $\hat{y}_t = \sum_{r=1}^{c_{t-1}} \alpha^r_{t-1}\kappa_K(\boldsymbol{S}_t,\boldsymbol{C}^r_{t-1})$
  $e_t = y_t - \hat{y}_t$
  $r^* = \arg\max_r \kappa_K(\boldsymbol{S}_t,\boldsymbol{C}^r_{t-1})$
  if $\Psi_{K/M}(\boldsymbol{S}_t,\boldsymbol{C}_{t-1}) > \gamma$ **then**
    $\boldsymbol{C}_t = \boldsymbol{C}_{t-1}$ $\alpha^{r^*}_{t-1} = \alpha^{r^*}_{t-1} + \eta e_t$ $\boldsymbol{\alpha}_t = \boldsymbol{\alpha}_{t-1}$
  else
    $\boldsymbol{C}_t = \{\boldsymbol{C}_{t-1},\boldsymbol{S}_t\}$ $\boldsymbol{\alpha}_t = \{\boldsymbol{\alpha}_{t-1},\eta e_t\}$
return

3. EXPERIMENTAL SET-UP AND RESULTS

To test capability of the proposed approach in finding the main dynamics of multi-channel time-series, a 3D human pose task is studied. Particularly, the well-known CMU Motion Capture Database - MoCap is used. Data hold 12 Vicon infrared MX-40 cameras recording at 120 Hz with 4 megapixel-resolution images. Subjects wear a black jump suit with 38 markers taped on while the infra-red cameras see the markers. Taken images by several cameras are triangulated to get 3D data representation. Then, motion activity of each subject is recorded in a BVH format video. For the concrete testing, the following activities are studied: walking (subject 02 video 01), jumping (subject 02 video 04), basketball (subject 06 video 15), and dancing (subject 05 video 11). Hence, we obtain 2D data by projecting from the 3D format into 2D. Thus, provided that MoCap video is given, an input multi-channel matrix $\boldsymbol{X} \in \Re^{p\times T}$ is obtained, where $p = 38 \times 2 = 64$ corresponds to the synthesized 2D angle coordinates, and T represents the frame sequence number.

Regarding the third coordinate $\boldsymbol{y} \in \Re^{T\times 1}$, it will be inferred based on the proposed FRKLMS by learning the function $\hat{\boldsymbol{y}} = f_{t-1}(\boldsymbol{S}_t)$. In this case, each input frame $\boldsymbol{x}_t$ is normalized with respect to the Hips joint, that is, this joint must be always centered at the $(0,0,0)$ position. So, each input matrix $\boldsymbol{X}$ is used to find the multi-channel time-series embedding matrix $\boldsymbol{H}$, fixing empirically the sliding window size value as $L = 25$ with $80[\%]$ of overlapping. After that, functional relevant based representation among channels is estimated using the Gaussian kernel with σ_S fixed as the mean value of the input Euclidean distances among $\boldsymbol{h}^i_t$ vectors. Lastly, the functional relevance based adaptive filter that is described in Algorithm 1 is performed. It is worth nothing that the correntropy based temporal relationships are computed using the σ_K value as the mean Frobenious norm among $\boldsymbol{S}_t$ matrices. Note that proposed FRKLMS is tested using both quantization criterion, kernel and MDL based ($\mathrm{FRKLMS_K}$ and $\mathrm{FRKLMS_{MDL}}$). In addition, the filter parameters are fixed empirically as $\gamma = 0.9$, $\eta = 0.75$, and $L_w = 5$. To evaluate system robustness against different testing noise conditions, the input data $\boldsymbol{X}$ is corrupted with additive white Gaussian noise to get different Signal to Noise Ratio conditions - SNR, $SNR = \{2,5,10,15,25\}[dB]$. As a baseline approach, the KLMS is applied [3], but for this case of multi-channel time-series, each row vector of $\boldsymbol{X}$ is employed as the filter input. Additionally, KLMS is also quantized using kernel and MDL based measures ($\mathrm{KLMS_K}$ and $\mathrm{KLMS_{MDL}}$). In Fig. 1, some $\mathrm{FRKLMS_K}$ visual results are shown for the studied videos (free of noise conditions). Performance of both compared algorithms is measured according to the mean Relative Error - RE and the final network size c_T, as shown in Fig. 2.

4. DISCUSSION

As seen in Figs. 1(d), 1(h), 1(l), and 1(p), the different relationships among channels for each activity are highlighted. Overall, there are some channels which share high similarity according to the given human pose, encoding the main functional relevance of the studied movement. Likewise, relationships are highlighted when analyzing the temporal kernel matrices (Figs. 1(c), 1(g), 1(k), and 1(o)). Note that other kernel functions can be used to compute the functional relationships, e.g., linear, polynomial, Laplacian, etc. In any case, an application task at hand can be adapted to a concrete kernel function depending on the available user's prior knowledge.

In this work, after visual inspection of Fig. 1(c), one can notice how the cyclic pattern of the walking movement is inferred by the filter. This fact is corroborated by the KPCA projection of the $\boldsymbol{K}$ matrix, as shown in Fig. 1(b). Moreover, such a cyclic behavior is related with the main poses of the gait, which are revealed by the filter final codebook, as seen

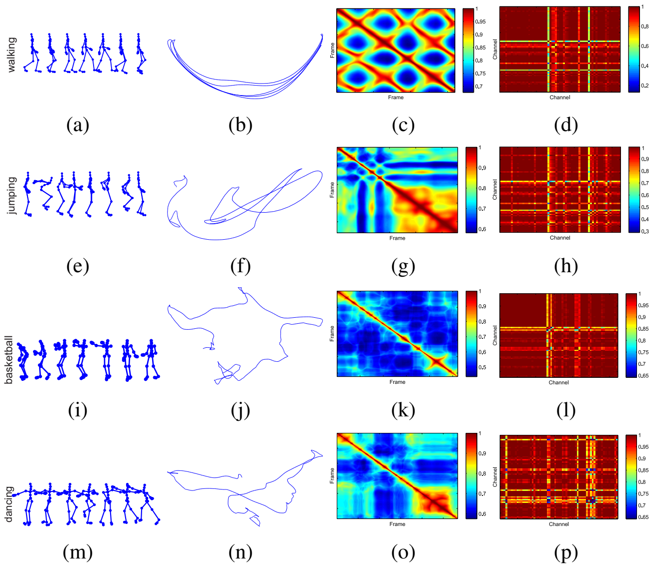

Fig. 1. Visual results. **First column**: Main codebook (MoCap space). **Second column**: KPCA projection. **Third column**: Kernel relationships among functional connectivity. **Fourth column**: Functional connectivity matrix example.

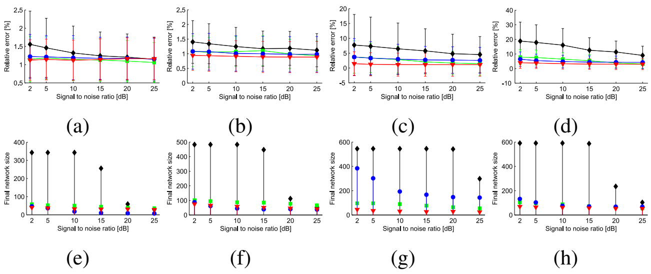

Fig. 2. Learning results (—KLMS_{K}, —KLMS_{MDL}, —FRKLMS_{K}, —$\text{FRKLMS}_{\text{MDL}}$). **First column**: walking. **Second column**: jumping. **Third column**: basketball. **Fourth column**: dancing.

in Fig. 1(b). Similar behavior is observed in the jumping video, as seen in Fig. 1(g) that shows the temporal relationships among functional relevant channel interactions. Here, the filter discovers two jumping cycles and one static state at the end of the video, since the subject jumps twice and then just stands for a while. In fact, the KPCA projection allows corroborating this behavior, as seen in Fig. 1(f). Regarding to more complex activities, in Fig. 1(k) tree main functional assemblies can be seen for the basketball video. In this case, the proposed approach is able to discover the preparing, shooting, and standing up states of the basketball activity (see main poses in Fig. 1(i)). Additionally, the KPCA projection reveals above mentioned behavior in Fig. 1(j), where no cyclic connections (circular shapes) are obtained. In turn, attained results reveal similar behavior for dancing video. In this case, however, there are more assemblies than for the basketball video because of activity complexity (see Fig. 1(m)). In fact, some cyclic patterns are discovered by the filter as seen in the kernel temporal relationships shown in Fig. 1(o). Those patterns can be identified visually in Fig. 1(n).

On the other hand, in most of the cases, proposed FRKLMS gets better performance than the baseline KLMS in terms of obtained RE results, as shown in Fig. 2. Particularly, the four studied approaches attain an acceptable performance on the walking video, since this is a smooth activity, not requiring a complex function to infer it. Yet, FRKLMS estimates a suitable learning function supplying the lowest number of samples, as seen in Figs. 2(a) and 2(e). The same results are obtained again for jumping, where FRKLMS outperforms the KLMS algorithm for both considered significance based alternatives (see Figs. 2(b) and 2(f)). Similarly, FRKLMS outperforms again the baseline algorithm for basket and dancing videos, as seen in Figs. 2(c), 2(d), 2(g), and 2(h). This advantage can be explained based on the proposed functional relevant representation, which highlights the joint relationships as an assembly, before applying the recursive kernel based filter. Thus, the temporal structure and the statistical dependencies among channels are suitable discovered.

Regarding to the employed significance measures, overall, traditional MDL based one obtains a suitable performance with low complexity, while the kernel based one tends to build more complex functions, decreasing the system performance for low SNR conditions. Thus, considering both the prediction error and the input data similarities into the quantization stage, as in MDL, allows to attain a better performance than only considering the input data similarities.

5. CONCLUSIONS

A functional relevance analysis approach is proposed to analyze multi-channel time-series. From the introduced representation, a kernel based adaptive filtering framework is extended to take advantage of the input data pair-wise similarities using a correntropy based function [1]. In addition, two significance measures are studied: kernel and MDL based, which are used to extract an inputs subset that allows to learn a model preserving a trade-off between filter complexity and accuracy. Our approach is tested on a well-known MoCap database for tracking 3D human joints, from which some videos are used to track human activities. According to the attained results, our framework provides an adequate alternative for finding functional relevant dependencies as an assembly into multi-channel time series. As a result, the system accuracy improves in comparison with the baseline KLMS [3], which does not consider directly inter channel dependencies. It is important to note that even when our approach is able to predict a given output time-series, it is also useful to interpret and to analyze visually complex relationships into multi-channel time-series. As future work, it would be interesting to perform further analysis for adapting the kernel parameters and for generating different codebooks along the time to deal with non stationary processes.

6. REFERENCES

[1] Ignacion Santamaría, Puskal Pokharel, and Jose Principe, "Generalized correlation funtion: definition, properties, and application to blind equalization," *IEEE Trans. on Signal Processing*, vol. 54, no. 6, pp. 2187–2197, 2006.

[2] Lin Li, I.M. Park, S. Seth, J.C. Sanchez, and J.C. Principe, "Functional connectivity dynamics among cortical neurons: A dependence analysis," *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, vol. 20, no. 1, pp. 18–30, 2012.

[3] Songlin Zhao, Pingping Zhu, and Principe Jose, "Quantized kernel least mean square algorithm," *IEEE Trans. on Neural Networks and Learning Systems*, vol. 23, pp. 22–32, 2012.

[4] Weifeng Liu, P.P. Pokharel, and J.C. Principe, "The kernel least-mean-square algorithm," *Signal Processing, IEEE Transactions on*, vol. 56, no. 2, pp. 543–554, 2008.

[5] Angela Yao, Juergen Gall, Luc Van Gool, and Raquel Urtasun, "Learning probabilistic non-linear latent variable models for tracking complex activities," in *NIPS*, 2011.

[6] R. Urtasun, D.J. Fleet, A. Hertzmann, and P. Fua, "Priors for people tracking from small training sets," in *Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on*, 2005, vol. 1, pp. 403–410 Vol. 1.

[7] Tobias Jaeggli, Esther Koller-Meier, and Luc Gool, "Learning generative models for multi-activity body pose estimation," *International Journal of Computer Vision*, vol. 83, no. 2, pp. 121–134, 2009.

[8] A.M. Álvarez-Meza, J. Valencia-Aguirre, G. Daza-Santacoloma, C.D. Acosta-Medina, and G. Castellanos-Domínguez, "Video analysis based on multi-kernel representation with automatic parameter choice," *Neurocomputing*, vol. 100, no. 0, pp. 117 – 126, 2013.

[9] Weifeng Liu, Puskal P. Pokharel, and Jose C. Principe, "Correntropy: Properties and applications in non-gaussian signal processing," *IEEE Transactions on Signal Processing*, vol. 55, no. 11, pp. 5286–5298, November 2007.

[10] Ana Rita Teixeira, Ana Maria Tomé, M. Boehm, Carlos G. Puntonet, and Elmar W. Lang, "How to apply nonlinear subspace techniques to univariate biomedical time series," *IEEE Transactions on Instrumentation and Measurement*, vol. 58, no. 8, pp. 2433–2443, 2009.

[11] Bernhard Scholkopf and Alexander J. Smola, *Learning with Kernels*, The MIT Press, Cambridge, MA, USA, 2002.

[12] Ya. I. Molkov, D. N. Mukhin, E. M. Loskutov, A. M. Feigin, and G. A. Fidelin, "Using the minimum description length principle for global reconstruction of dynamic systems from noisy time series," *Phys. Rev. E*, vol. 80, pp. 046207, Oct 2009.