

Emotional Expression Classification using Time-Series Kernels

András Lőrincz¹, László A. Jeni², Zoltán Szabó¹, Jeffrey F. Cohn^{2,3}, and Takeo Kanade²

¹Eötvös Loránd University, Budapest, Hungary, {andras.lorincz, szzoli}@elte.hu

²Carnegie Mellon University, Pittsburgh, PA, laszlo.jeni@ieee.org, tk@cs.cmu.edu

³University of Pittsburgh, Pittsburgh, PA, jeffc@cs.cmu.edu

Abstract

Estimation of facial expressions, as spatio-temporal processes, can take advantage of kernel methods if one considers facial landmark positions and their motion in 3D space. We applied support vector classification with kernels derived from dynamic time-warping similarity measures. We achieved over 99% accuracy – measured by area under ROC curve – using only the ‘motion pattern’ of the PCA compressed representation of the marker point vector, the so-called shape parameters. Beyond the classification of full motion patterns, several expressions were recognized with over 90% accuracy in as few as 5-6 frames from their onset, about 200 milliseconds.

1. Introduction

Because they enable model based prediction and timely reactions, the analysis and identification of spatio-temporal processes, including concurrent and interacting events, are of great importance in many applications. Spatio-temporal processes have intriguing features. Consider three examples. First is a feature-length movie, which is a set of time series of pixel intensities in which the number of pixels is on the order of 100,000 and the values follow each other at a rate of 30 fps for a time interval of 2 hours or so. Second are financial time series. Currency exchange rates, stocks, and many other finance related data are heavily affected by both common underlying processes and by one another and exhibit large coupled fluctuations over 9 orders of magnitude [16]. Third, and a focus of the current paper, are facial expression time series. The motion patterns of landmark points of a face, such as mouth and eye corners comprise a ‘landmark space’. The dynamics of change in this space can reveal emotion, pain, and cognitive states, and regulate social interaction. For effective human-computer interaction, automated facial expression analysis is important.

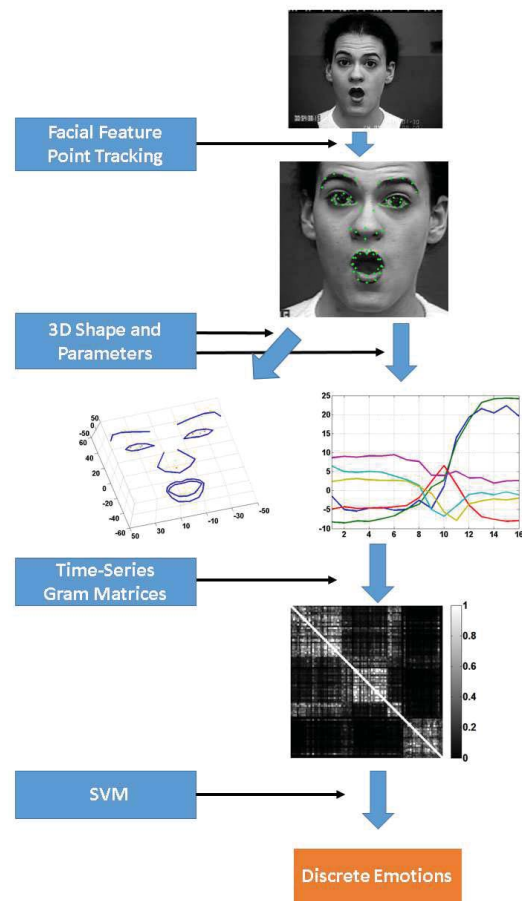


Figure 1. Overview of the system.

In all three types of time series, i.e., movies, financial data, and facial expressions, we are dealing with spatio-temporal processes. For each, kernel methods hold great promise.

The demands for characterizing such processes pose spe-

cial challenges because very different signals may represent the same process when the process is viewed from different distances (in the case of the movie), different time scales (market data), or different viewing angles (facial expressions). Such invariances and distortions need be taken into account.

Early attempts to use support vector methods for the prediction of time series were very promising [15] even in the absence of algorithms compensating for temporal distortions. In other areas, such as speech recognition, time warping algorithms have been developed early in order to match slower and faster speech fragments, see, e.g., [6] and the references therein. Dynamic time warping is one of the most efficient methods that offer the comparison of temporally distorted samples [18].

Recently, the two methods, i.e., dynamic time warping and SVMs have been combined and show considerable performance increases in the analysis of spatio-temporal signals [5, 4]. Efficient methods using independent component analysis [12], Haar filters [22], hidden Markov models [1, 2, 20] have been applied for problems related to the estimation of emotions and facial expressions. Here, we study the efficiency of novel dynamic time warping kernels [5, 4] for emotional expression estimation.

Our contributions are as follows: We show that (1) time series kernel methods are highly precise for emotional expression estimation using landmark data only and (2) they enable early and reliable estimation of expression as soon as 5 frames from expression onset, i.e., around 200 ms.

The paper is organized as follows. First, in the Methods section we review how landmark points are observed in 3 dimensions, sketch the two spatio-temporal kernels that we applied, and describe support vector machine (SVM) principles. Section 3 is about our experimental studies. It is followed by our Discussion and Summary.

2. Methods

2.1. Facial Feature Point Localization

To localize a dense set of facial landmarks, Active Appearance Models (AAM) [14] and Constrained Local Models (CLM) [19] are often used. These methods register a dense parameterized shape model to an image such that its landmarks correspond to consistent locations on the face.

Of the two, person specific AAMs have higher precision than CLMs, but they must be trained for each person before use. On the other hand, CLM methods can be used for person-independent face alignment because of the localized region templates.

In this work we use a 3D CLM method, where the shape model is defined by a 3D mesh and in particular the 3D vertex locations of the mesh, called landmark points. Consider the shape of a 3D CLM as the coordinates of 3D vertices

that make up the mesh:

$$\mathbf{x} = [x_1; y_1; z_1; \dots; x_M; y_M; z_M], \quad (1)$$

or, $\mathbf{x} = [\mathbf{x}_1; \dots; \mathbf{x}_M]$, where $\mathbf{x}_i = [x_i; y_i; z_i]$. We have T samples: $\{\mathbf{x}(t)\}_{t=1}^T$. We assume that – apart from scale, rotation, and translation – all samples $\{\mathbf{x}(t)\}_{t=1}^T$ can be approximated by means of the linear principal component analysis (PCA).

In the next subsection we briefly describe the 3D Point Distribution Model and how the CLM method estimates the landmark positions.

2.1.1 Point Distribution Model

The 3D point distribution model (PDM) describes non-rigid shape variations linearly and composes it with a global rigid transformation, placing the shape in the image frame:

$$\mathbf{x}_i(p) = s\mathbf{P}\mathbf{R}(\bar{\mathbf{x}}_i + \Phi_i\mathbf{q}) + \mathbf{t}, \quad (2)$$

($i = 1, \dots, M$), where $\mathbf{x}_i(p)$ denotes the 3D location of the i^{th} landmark and $\mathbf{p} = \{s, \alpha, \beta, \gamma, \mathbf{q}, \mathbf{t}\}$ denotes the parameters of the model, which consist of a global scaling s , angles of rotation in three dimensions ($\mathbf{R} = \mathbf{R}_1(\alpha)\mathbf{R}_2(\beta)\mathbf{R}_3(\gamma)$), a translation \mathbf{t} and non-rigid transformation \mathbf{q} . Here $\bar{\mathbf{x}}_i$ denotes the mean location of the i^{th} landmark (i.e. $\bar{\mathbf{x}}_i = [\bar{x}_i; \bar{y}_i; \bar{z}_i]$ and $\bar{\mathbf{x}} = [\bar{\mathbf{x}}_1; \dots; \bar{\mathbf{x}}_M]$) and \mathbf{P} denotes the projection matrix to 2D:

$$\mathbf{P} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}. \quad (3)$$

We assume that the prior of the parameters follow a normal distribution with mean $\mathbf{0}$ and variance $\mathbf{\Lambda}$ at a parameter vector \mathbf{q} :

$$p(\mathbf{p}) \propto N(\mathbf{q}; \mathbf{0}, \mathbf{\Lambda}), \quad (4)$$

From \mathbf{x}_i points PCA provides $\bar{\mathbf{x}}$ in (2) and $\mathbf{\Lambda}$ in (4).

2.1.2 Constrained Local Model

CLM is constrained through the PCA of PDM. It works with *local experts*, whose opinion is considered independent and are multiplied to each other:

$$J(\mathbf{p}) = p(\mathbf{p}) \prod_{i=1}^M p(l_i = 1 | \mathbf{x}_i(\mathbf{p}), \mathcal{I}) \rightarrow \min_{\mathbf{p}}, \quad (5)$$

where $l_i \in \{-1, 1\}$ is a stochastic variable telling whether the i^{th} marker is in its position or not, $p(l_i = 1 | \mathbf{x}_i(\mathbf{p}), \mathcal{I})$ is the probability that for image \mathcal{I} and for marker position \mathbf{x}_i (being a function of parameter \mathbf{p} , i.e., for $\mathbf{x}_i(\mathbf{p})$) the i^{th} marker is in its position.

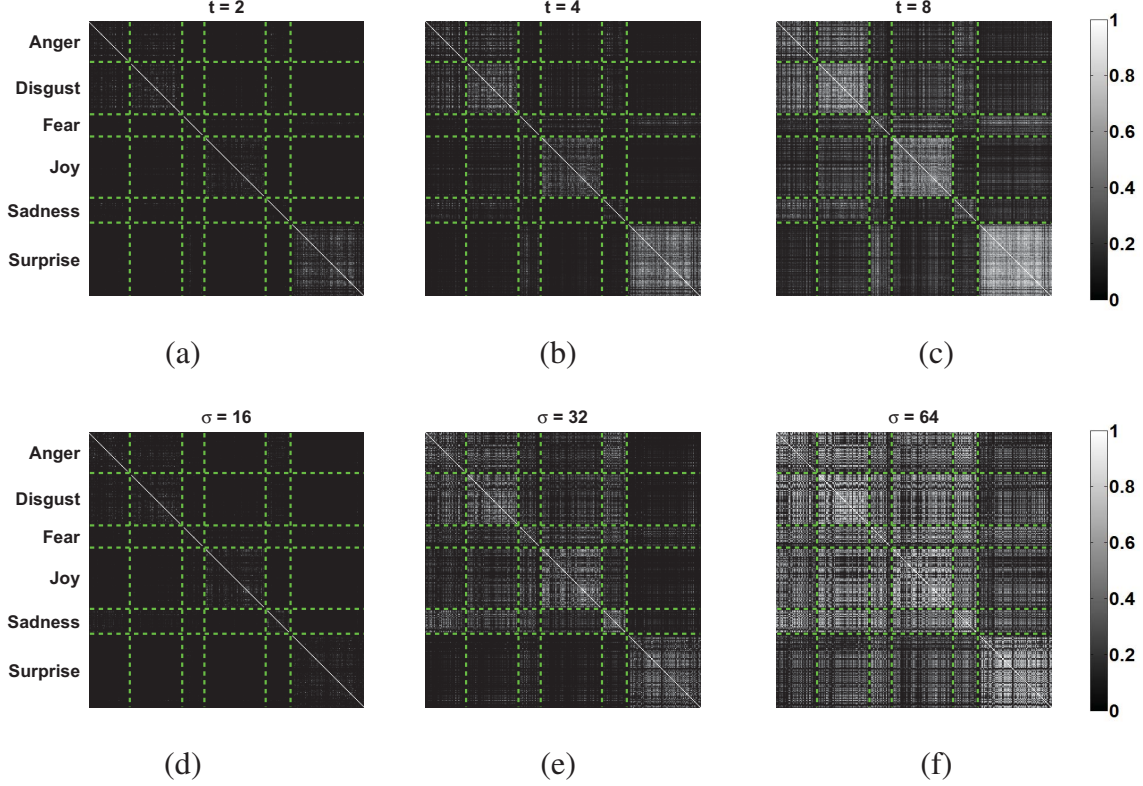


Figure 2. Gram matrices induced by the pseudo-DTW kernel (a-c) and the GA kernel (d-f) with different parameters. The rows and columns represent the time-series grouped by the emotion labels (the boundaries of the different emotional sets are denoted with green dashed lines). The pixel intensities in each cell show the similarity between two time-series.

The interested reader is referred to [19] for the details of the CLM algorithm.¹

2.2. Time-series Kernels

Kernel based classifiers, like any other classification scheme, should be robust against invariances and distortions. Dynamic time warping, traditionally solved by dynamic programming, has been introduced to overcome temporal distortions and has been successfully combined with kernel methods. Below, we describe two kernels that we applied in our numerical studies: the Dynamic Time Warping (DTW) kernel and the Global Alignment (GA) kernel.

2.2.1 Dynamic Time Warping Kernel

Let $\mathcal{X}^{\mathbb{N}}$ be the set of discrete-time time series taking values in an arbitrary space \mathcal{X} . One can try to align two time series $x = (x_1, \dots, x_n)$ and $y = (y_1, \dots, y_m)$ of lengths n and m , respectively, in various ways by distorting them. An alignment π has length p and $p \leq n + m - 1$ since the two series have $n + m$ points and they are matched at least

one point of time. We use the notation of [4]. An alignment π is a pair of increasing integral vectors (π_1, π_2) of length p such that $1 = \pi_1(1) \leq \dots \leq \pi_1(p) = n$ and $1 = \pi_2(1) \leq \dots \leq \pi_2(p) = m$, with unitary increments and no simultaneous repetitions. In turn, for all indices $1 \leq i \leq p - 1$, the increment vector of π belongs to a set of 3 elementary moves as follows

$$\begin{pmatrix} \pi_1(i+1) - \pi_1(i) \\ \pi_2(i+1) - \pi_2(i) \end{pmatrix} \in \left\{ \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right\} \quad (6)$$

Coordinates of π are also known as warping functions. Let $A(n, m)$ denote the set of all alignments between two time series of length n and m . The simplest DTW 'distance' between x and y is defined as

$$DTW(x, y) \stackrel{def}{=} \min_{\pi \in A(n, m)} D_{x, y}(\pi) \quad (7)$$

Now, let $|\pi|$ denote the length of alignment π . The *cost* can be defined by means of a local divergence ϕ that measures the discrepancy between any two points x_i and y_j of vectors x and y .

$$D_{x, y}(\pi) \stackrel{def}{=} \sum_i^{|\pi|} \phi(x_{\pi_1(i)}, y_{\pi_2(i)}) \quad (8)$$

¹We used the CLM software of Saragih, which is available here <https://github.com/kylemcdonald/FaceTracker>.

The squared Euclidean distance is often used to define the divergence $\phi(x, y) = \|x - y\|^2$. Although this measure is symmetric, it does not satisfy the triangle inequality under all conditions – so it is not rigorously a distance – and cannot be used directly to define a positive semi-definite kernel. This problem can be alleviated by projecting matrix $D_{x,y}(\pi)$ to a set of symmetric positive semi-definite matrices. There are various methods for accomplishing such approximations. They called distance substitution [7]. We applied the alternating projection method of [8] that finds the nearest correlation matrix. Denoting the new matrix by $D\hat{T}W(x, y)$, the modified DTW distance induces a positive semi-definite kernel as follows

$$k_{DTW}(x, y) = e^{-\frac{1}{t}D\hat{T}W(x,y)}, \quad (9)$$

where t is a constant.

The full procedure can be summarized as follows: (1) take the samples, (2) compute the Euclidean distances for each sample pair, (3) build the matrix from these sample pairs, (4) find the nearest correlation matrix, (5) use it to construct a kernel, and (6) compute the Gram matrix of the support vector classification problem. Fig. 2 (a)-(c) show Gram matrices induced by the pseudo-DTW kernel with different t parameters.

2.2.2 Global Alignment Kernel

The Global Alignment (GA) kernel assumes that the minimum value of alignments may be sensitive to peculiarities of the time series and intends to take advantage of all alignments weighted exponentially. It is defined as the sum of exponentiated and sign changed costs of the individual alignments:

$$k_{GA}(x, y) \stackrel{def}{=} \sum_{\pi \in A(n,m)} e^{-D_{x,y}(\pi)}. \quad (10)$$

Equation (10) can be rewritten by breaking up the alignment distances according to the local divergences: similarity function κ is induced by divergence ϕ :

$$k_{GA}(x, y) \stackrel{def}{=} \sum_{\pi \in A(n,m)} \prod_{i=1}^{|\pi|} e^{-\phi(x_{\pi_1(i)}, y_{\pi_2(i)})} \quad (11)$$

$$\stackrel{def}{=} \sum_{\pi \in A(n,m)} \prod_{i=1}^{|\pi|} \kappa(x_{\pi_1(i)}, y_{\pi_2(i)}), \quad (12)$$

where notation $\kappa = e^{-\phi}$ was introduced for the sake of simplicity. It has been argued that k_{GA} runs over the whole spectrum of the costs and gives rise to a smoother measure than the minimum of the costs, i.e., the DTW distance [5]. It has been shown in the same paper that k_{GA} is positive

definite provided that $\kappa/(1 + \kappa)$ is positive definite on \mathcal{X} . Furthermore, the computational effort is similar to that of the DTW distance; it is $\mathcal{O}(nm)$. Cuturi argued in [4] that global alignment kernel induced Gram matrix do not tend to be diagonally dominated as long as the sequences to be compared have similar lengths.

In our numerical simulations, we used local kernel $e^{-\phi_\sigma}$ suggested by Cuturi, where

$$\phi_\sigma \stackrel{def}{=} \frac{1}{2\sigma^2} \|x - y\|^2 + \log \left(2 - e^{-\frac{\|x-y\|^2}{2\sigma^2}} \right). \quad (13)$$

Fig. 2 (d)-(f) show Gram matrices induced by the GA kernel with different σ parameters.

2.3. Time-series Classification using SVM

Support Vector Machines (SVMs) are very powerful for binary and multi-class classification as well as for regression problems [3]. They are robust against outliers. For two-class separation, SVM estimates the optimal separating hyper-plane between the two classes by maximizing the margin between the hyper-plane and closest points of the classes. The closest points of the classes are called support vectors; the optimal separating hyper-plane lies at half distance between them.

We are given sample and label pairs $(\mathbf{x}^{(i)}, y^{(i)})$ with $\mathbf{x}^{(i)} \in \mathbb{R}^m$, $y^{(i)} \in \{-1, 1\}$, and $i = 1, \dots, K$. Here, for class '1' and for class '2' $y^{(i)} = 1$ and $y^{(i)} = -1$, respectively. We also have a feature map $\phi : \mathbb{R}^m \rightarrow \mathcal{H}$, where \mathcal{H} is a Hilbert-space. The kernel implicitly performs the dot product calculations between mapped points: $k(x, y) = \langle \phi(x), \phi(y) \rangle_{\mathcal{H}}$. The support vector classification seeks to minimize the cost function

$$\min_{\mathbf{w}, b, \xi} \frac{1}{2} \mathbf{w}^T \mathbf{w} + C \sum_{i=1}^K \xi_i \quad (14)$$

$$y^{(i)} (\mathbf{w}^T \phi(\mathbf{x}^{(i)}) + b) \geq 1 - \xi_i, \quad \xi_i \geq 0. \quad (15)$$

where ξ_i ($i = 1, \dots, K$) are the so called slack variables that generalize the original SVM concept with separating hyper-planes to soft-margin classifiers that have outliers that can not be separated.

3. Experiments

3.1. Cohn-Kanade Extended Dataset

In our simulations we used the Cohn-Kanade Extended Facial Expression (CK+) Database [13]. This database was developed for automated facial image analysis and synthesis and for perceptual studies. The database is widely used to compare the performance of different models. The database contains 123 different subjects and 593 frontal image sequences. From these, 118 subjects are annotated with

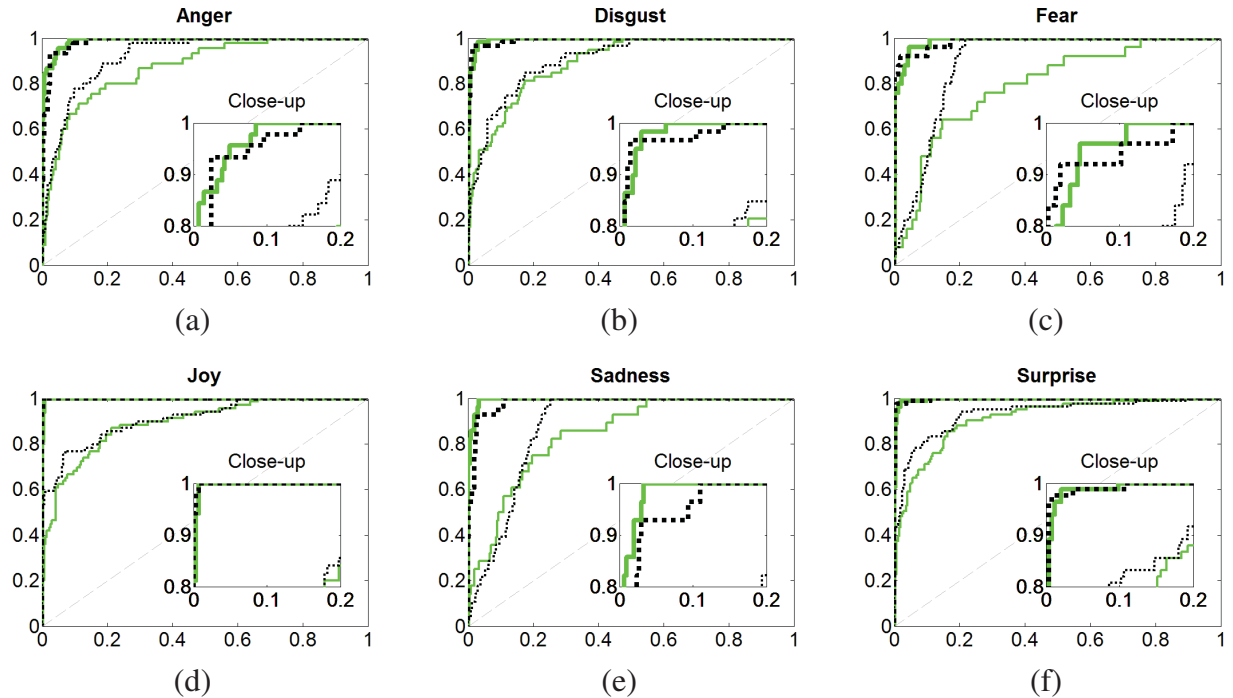


Figure 3. ROC curves of the different emotion classifiers: (a) anger, (b) disgust, (c) fear, (d) joy, (e) sadness and (f) surprise. Thick lines: performance using all frames of the sequences. Thin lines: performance using the 6th frames of the sequences. Solid (dotted) line: results for pseudo DTW (GA) kernel.

the seven universal emotions (anger, contempt, disgust, fear, happy, sad and surprise). Action units are also provided with this database for the apex frame. The original Cohn-Kanade Facial Expression Database distribution [11] had 486 FACS-coded sequences from 97 subjects. CK+ has 593 posed sequences with full FACS coding of the peak frames. A subset of action units were coded for presence or absence. For these sequences the 3D landmarks and shape parameters were provided by the CLM tracker itself.

3.2. Emotional Expression Classification

In this set of experiment we studied the two kernel methods on the CK+ dataset. We measured the performances of the methods for emotion recognition.

First, we tracked facial expressions with the CLM tracker and annotated all image sequences starting from the neutral expression to the peak of the emotion. The CLM estimates the rigid and non-rigid transformations. We removed the rigid ones from the faces and represented the sequences as multi-dimensional time-series built from the non-rigid shape parameters.

We calculated Gram matrices using the pseudo DTW and the GA kernels and performed leave-one-subject out cross validation to maximally utilize the available set of training data. For both kernels, we searched for the best parameter (t in the case of pseudo-DTW kernel and σ in the case of GA kernel) between 2^{-5} and 2^{10} on a logarithmic scale

with equidistant steps and selected the parameter having the lowest mean classification error. The SVM regularization parameter (C) was searched within 2^{-5} and 2^5 in a similar fashion. If the pseudo-DTW kernel based Gram matrix was not positive semi-definite then we projected it to the nearest positive semi-definite matrix using the alternating projection method of [8].

The result of the classification is shown in Fig. 3. Performance is nearly 100% for expressions with large deformations in the facial features, such as disgust, happiness and surprise.

To the best of our knowledge, classification performance with time-series kernels is better than the best available results to date, including spatio-temporal ICA, boosted dynamic features, and non-negative matrix factorization techniques. For detailed comparisons, see Table 1.

3.3. Early Expression Classification

Encouraged by the results of the first experiment, we decided to constrain the maximum length of the sequences used in training and testing in order to estimate performance in the early phase of the emotion events.

We cropped the the time series between 2 and 16 frames and trained kernel SVMs for one-vs-all emotion classification. Figure 4 shows the classification performance as a function of the maximum length of the sequences. According to the figures, 3-to-4 frames can reach 80% AUC-ROC

Table 1. (a) Comparisons with hand-designed spatio-temporal Gabor filters (Wu et al. 2010 [21]), learned spatio-temporal ICA filters (Long et al. 2012 [12]) and Sparse Non-negative Matrix Factorization (NMF) filters (Jeni et al. 2013 [9]) on the first 6 frames. (b) Comparisons with boosted dynamic features (Yang et al. 2009 [22]) on the last frames of the sequences.

(a)								(b)							
Method	Anger	Disg.	Fear	Joy	Sadn.	Surp.	Average	Method	Anger	Disg.	Fear	Joy	Sadn.	Surp.	Average
Wu et al. [21]	0.829	0.677	0.667	0.877	0.784	0.879	0.786	Yang et al. [22]	0.973	0.941	0.916	0.991	0.978	0.998	0.966
Long et al. [12]	0.774	0.711	0.692	0.894	0.848	0.891	0.802	Long et al. [12]	0.933	0.988	0.964	0.993	0.991	0.999	0.978
Jeni et al. [9]	0.817	0.908	0.774	0.938	0.865	0.886	0.865	Jeni et al. [9]	0.989	0.998	0.977	0.998	0.994	0.994	0.992
This work, DTW	0.873	0.893	0.793	0.892	0.843	0.909	0.867	This work, DTW	0.991	0.994	0.987	0.999	0.995	0.996	0.994
This work, GA	0.921	0.905	0.887	0.910	0.871	0.930	0.904	This work, GA	0.986	0.993	0.986	1.000	0.984	0.997	0.991

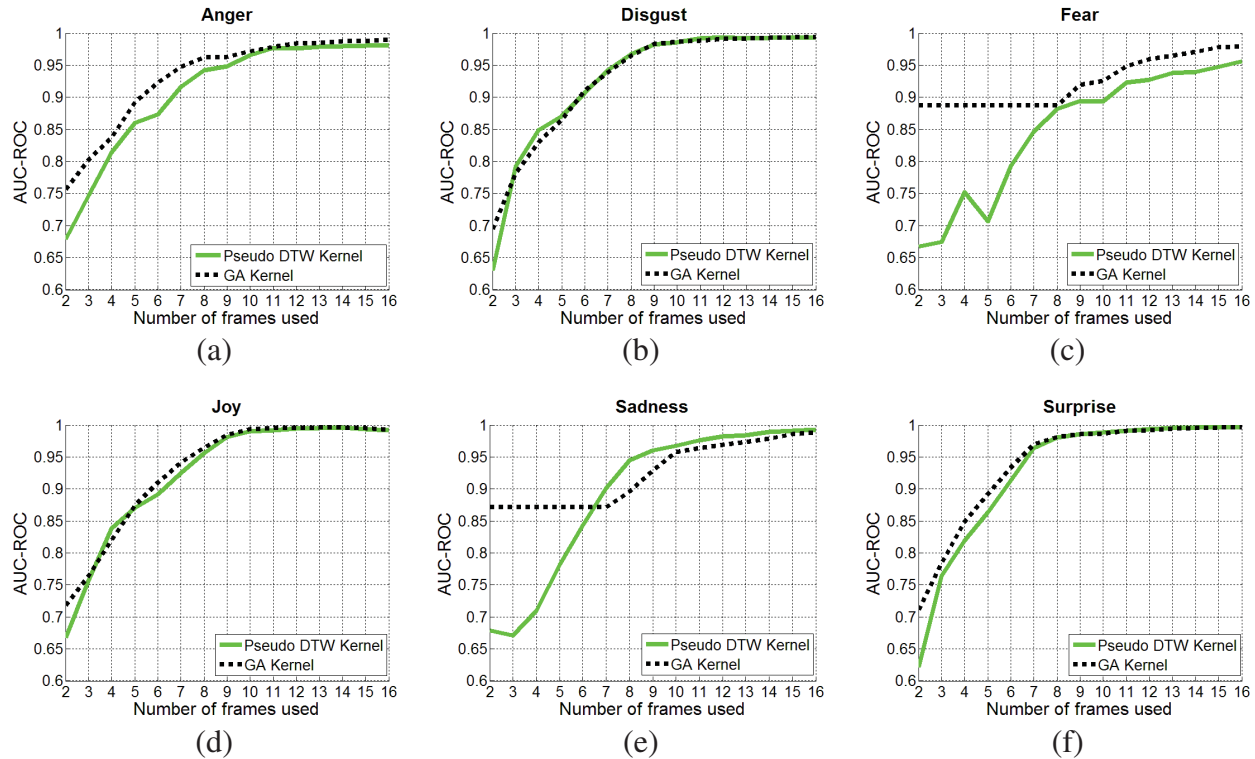


Figure 4. Area Under ROC curve values of the different emotion classifiers: (a) anger, (b) disgust, (c) fear, (d) joy, (e) sadness and (f) surprise. Solid (dotted) line: results for the pseudo DTW (GA) kernel.

performance, whereas 5-to-6 frames are sufficient for about 90% performance.

4. Discussion and Summary

We have studied time-series kernel methods for the analysis of emotional expressions. We used the well known 3D CLM method with the available open source C++ implementation of Jason Saragih. Compared to previous results, we found superior performances both on the first 6 frames and on the last few frames of the sequences collected from the CK+ database. It is notable that we used only shape information and neglected the textural one, since the 3D CLM model can compensate for the head poses making the method robust against head pose variations [10].

The NMF method [9] that deeply exploits textural features comes close to our method and one expects that mixing the two methods may unite the advantages of the two approaches, namely the robustness of the shape based method against pose variations and light conditions, and sometimes strong textural changes under small landmark position variations. Also, textural changes are less sensitive to the estimation noise of the landmark positions.

We achieved highly promising results at early times of the series: 3-to-4 frames reached 80% AUC-ROC performance, whereas 5-to-6 frames were sufficient for about 90% performance. Such early detection enables timely response in human-computer interactions and collaborations. Furthermore, the early frames of the series have smaller AUC values and should make emotion estimation more robust.

In sum, time-series kernels are very promising for emotion recognition. There is a number of potential improvements to our method, such as (i) joined texture and shape based facial expression recognition using for example, probabilistic SVMs, and (ii) novel DTW optimization methods, like lower bounding or the UCR suite approach [17] that can make the proposed system tractable for real-time analysis.

5. Acknowledgments

We are grateful to Jason Saragih for providing his CLM code for our work.

References

- [1] K. Bousmalis, L.-P. Morency, and M. Pantic. Modeling hidden dynamics of multimodal cues for spontaneous agreement and disagreement recognition. In *Proceedings of the Ninth IEEE International Conference on Automatic Face and Gesture Recognition (FG 2011)*, pages 746–752, 2011. 2
- [2] K. Bousmalis, S. Zafeiriou, L.-P. Morency, and M. Pantic. Infinite hidden conditional random fields for human behavior analysis. *IEEE Transactions on Neural Networks and Learning Systems*, 24(1):170–177, 2013. 2
- [3] C.-C. Chang and C.-J. Lin. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2:27:1–27:27, 2011. Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>. 4
- [4] M. Cuturi. Fast global alignment kernels. In *Proceedings of the International Conference on Machine Learning*, 2011. 2, 3, 4
- [5] M. Cuturi, J.-P. Vert, . Birkenes, and T. Matsui. A kernel for time series based on global alignments. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2007)*, volume 2, pages 413–416, 2007. 2, 4
- [6] J. R. Deller, J. G. Proakis, and J. H. Hansen. *Discrete-time processing of speech signals*. IEEE New York, NY, USA:, 2000. 2
- [7] B. Haasdonk and C. Bahlmann. Learning with distance substitution kernels. In *Pattern Recognition*, pages 220–227. Springer, 2004. 4
- [8] N. J. Higham. Computing the nearest correlation matrix a problem from finance. *IMA Journal of Numerical Analysis*, 22(3):329–343, 2002. 4, 5
- [9] L. A. Jeni, J. M. Girard, J. F. Cohn, and F. De La Torre. Continuous AU intensity estimation using localized, sparse facial feature space. In *2nd International Workshop on Emotion Representation, Analysis and Synthesis in Continuous Time and Space (EmoSPACE)*, 2013. 6
- [10] L. A. Jeni, A. Lőrincz, T. Nagy, Z. Palotai, J. Sebők, Z. Szabó, and D. Takács. 3d shape estimation in video sequences provides high precision evaluation of facial expressions. *Image and Vision Computing*, 30(10):785–795, 2012. 6
- [11] T. Kanade, J. F. Cohn, and Y. Tian. Comprehensive database for facial expression analysis. In *Proceedings of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition (FG 2000)*, pages 46–53, 2000. 5
- [12] F. Long, T. Wu, J. R. Movellan, M. S. Bartlett, and G. Littlewort. Learning spatiotemporal features by using independent component analysis with application to facial expression recognition. *Neurocomputing*, 93:126 – 132, 2012. 2, 6
- [13] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews. The extended Cohn-Kanade dataset (CK+): A complete dataset for action unit and emotion-specified expression. In *Proceedings of the Computer Vision and Pattern Recognition Workshops (CVPRW 2010)*, pages 94–101, 2010. 4
- [14] I. Matthews and S. Baker. Active appearance models revisited. *International Journal of Computer Vision*, 60(2):135–164, 2004. 2
- [15] K.-R. Müller, A. J. Smola, G. Rätsch, B. Schölkopf, J. Kohlmorgen, and V. Vapnik. Predicting time series with support vector machines. In *Artificial Neural Networks-ICANN'97*, pages 999–1004. Springer, 1997. 2
- [16] T. Preis, J. J. Schneider, and H. E. Stanley. Switching processes in financial markets. *Proceedings of the National Academy of Sciences*, 108(19):7674–7678, 2011. 1
- [17] T. Rakthanmanon, B. Campana, A. Mueen, G. Batista, B. Westover, Q. Zhu, J. Zakaria, and E. Keogh. Searching and mining trillions of time series subsequences under dynamic time warping. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 262–270. ACM, 2012. 7
- [18] H. Sakoe and S. Chiba. Dynamic programming algorithm optimization for spoken word recognition. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, 26(1):43–49, 1978. 2
- [19] J. M. Saragih, S. Lucey, and J. F. Cohn. Deformable model fitting by regularized landmark mean-shift. *International Journal of Computer Vision*, 91(2):200–215, 2011. 2, 3
- [20] M. F. Valstar and M. Pantic. Combined support vector machines and hidden markov models for modeling facial action temporal dynamics. In *Human-Computer Interaction*, pages 118–127. Springer, 2007. 2
- [21] T. Wu, M. Bartlett, and J. R. Movellan. Facial expression recognition using gabor motion energy filters. In *Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on*, pages 42–47, 2010. 6
- [22] P. Yang, Q. Liu, and D. N. Metaxas. Boosting encoded dynamic features for facial expression recognition. *Pattern Recognition Letters*, 30(2):132–139, 2009. 2, 6