Continuous AU Intensity Estimation using Localized, Sparse Facial Feature Space

László A. Jeni, Jeffrey M. Girard, Jeffrey F. Cohn and Fernando De La Torre

Abstract— **Most work in automatic facial expression analysis seeks to detect discrete facial actions. Yet, the meaning and function of facial actions often depends in part on their intensity. We propose a part-based, sparse representation for automated measurement of continuous variation in AU intensity. We evaluated its effectiveness in two publically available databases, CK+ and the soon to be released Binghamton high-resolution spontaneous 3D dyadic facial expression database. The former consists of posed facial expressions and ordinal level intensity (absent, low, and high). The latter consists of spontaneous facial expression in response to diverse, well-validated emotion inductions, and 6 ordinal levels of AU intensity.**

In a preliminary test, we started from discrete emotion labels and ordinal-scale intensity annotation in the CK+ dataset. The algorithm achieved state-of-the-art performance. These preliminary results supported the utility of the part-based, sparse representation. Second, we applied the algorithm to the more demanding task of continuous AU intensity estimation in spontaneous facial behavior in the Binghamton database. Manual 6-point ordinal coding and continuous measurement were highly consistent. Visual analysis of the overlay of continuous measurement by the algorithm and manual ordinal coding strongly supported the representational power of the proposed method to smoothly interpolate across the full range of AU intensity.

I. INTRODUCTION

Most approaches to automatic facial expression analysis seek to detect discrete facial actions or discrete emotion-specified expressions. [1], [2], [3]. Yet, facial actions and expressions can vary in intensity as well as in kind. Differences in intensity have proved important in distinguishing the different possible meanings and functions of facial actions that are otherwise similar. For instance, differences in intensity, alone or in combination with other actions, have distinguished between posed and spontaneous smiles [4] and between smiles perceived as polite versus those perceived as embarrassed [5].

Efforts to detect or measure the intensity of facial actions have begun. One approach is based on the Facial Action Coding System (FACS: [6]). FACS defines ordinal-level variation in action unit intensity. Action units (AU) are anatomically based, discrete facial actions that may occur alone or in combinations to represent nearly all possible facial expressions. Mahoor and his colleagues [7], [8] and Kaltwand and Pantic [9] proposed frameworks to detect AU intensity.

László A. Jeni and Fernando De La Torre are with The Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, 15213, USA. laszloaj@andrew.cmu.edu, ftorre@cs.cmu.edu

Jeffrey M. Girard and Jeffrey F. Cohn are with the Department of Psychology, University of Pittsburgh, Pittsburgh, PA, 15260, USA. jmg174@pitt.edu, jeffcohn@cs.cmu.edu

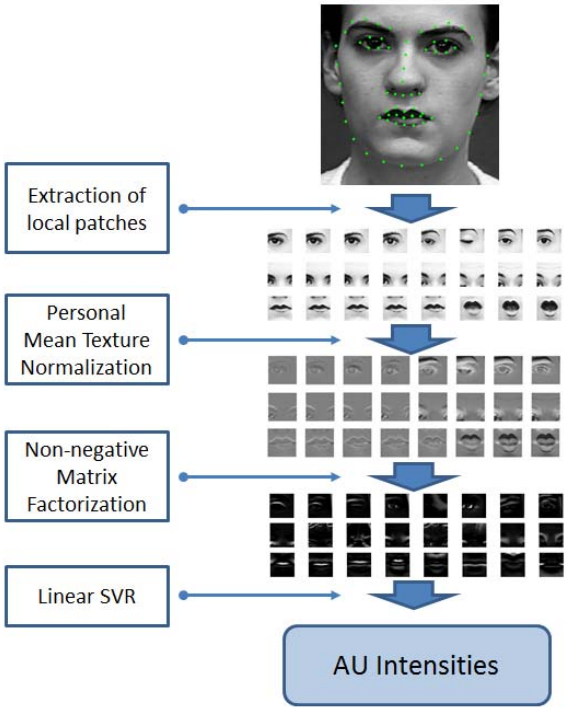


Fig. 1. Overview of the system.

Related approaches have been applied to the problem of measuring intensity variation in pain expressions [10] and [9]. The latter are based on Prkachin and Solomons Pain Intensity Metric [11]. Ordinal pain scores are computed as the sum of specific AU intensities or directly from facial features [12]. These approaches all seek to detect ordinal level intensity.

Another approach seeks to detect continuous variation in latent dimensions of interpersonal behavior. The interpersonal emotion circumplex conceptualizes dimensions of valence, activity, and dominance [13]. Nicolaou [14] proposed an Output-Associative Relevance Vector Machine regression framework to detect continuous variation in dimensions of the emotion circumplex. Gunes [15] reviewed work in continuous detection of emotion-circumplex dimensions from facial expression, as well as from other modalities. These approaches all assume that latent psychological constructs (e.g., valence) can be measured from holistic representations of visible facial expression or other modalities. Whether or not emotion can be measured directly or can only be inferred is outside the scope of the current work [16].

We propose an approach to continuous measurement of discrete facial actions (AU). The approach is inspired by recent research into the primate visual system [17], which emphasizes the role of part-based representations in early visual processing and both part-based and holistic representations in later, higher-level, visual processing. First, we acquire person-independent dense facial marker registration by the means of a Constrained Local Model (CLM) and extract localized patches around the markers (66-point face mesh). Then we build a sparse, part-based representation of the patches by applying non-negative matrix factorization. This decomposition is appealing for two reasons: (a) the filters represent highly localized parts of the face that are robust to noise and structural perturbations, and (b) from a computational standpoint, convolution with a sparse filter can be done more efficiently than with a non-sparse filter [18].

Preliminary results in two different databases suggest that this representation gives rise to reliable estimation of continuous AU intensity. Fig. 1 presents an overview.

The rest of the paper is organized as follows. Theoretical components are described in Section II. Datasets and their properties are reviewed in Section III. Experimental results on AU intensity are detailed in Section IV. Discussion and a summary conclude the paper (Section V).

II. METHODS

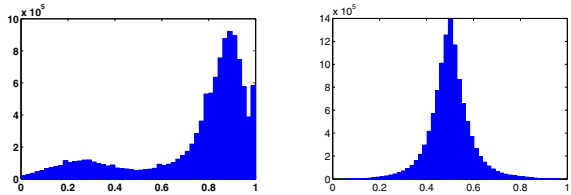
Our proposed method consists of three main steps. First, we estimate landmark positions on face images using the CLM method. We describe the details of this technique in Section II-A. Then we extract local texture patches around the markers (Section II-B). To do that, we remove the rigid transformation from the acquired shape using Procrustes normalization (Section II-B.1); extract AU-related regions from the normalized image (Section II-B.2); and then normalize the acquired patches by removing the personal mean (Section II-B.3). Using the normalized patches we build a component based representation using non-negative matrix factorization. The details of this method are described in Section II-C. SVM based classification and regression methods are reviewed in Section II-D.

A. Constrained Local Model

CLM methods are generative parametric models for person-independent face alignment. In this work we were using a 3D CLM method, in which the shape model is defined by a 3D mesh and, in particular, the 3D vertex locations of the mesh, which are referred to as landmark points. Consider the shape of a 3D CLM as the coordinates of 3D vertices of the M landmark points:

$$\mathbf{x} = (x_1, y_1, z_1, ..., x_M, y_M, z_M)^T, \tag{1}$$

or, $\mathbf{x} = (\mathbf{x}_1, \ldots, \mathbf{x}_M)^T$, where $\mathbf{x}_i = (x_i, y_i, z_i)^T$. 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).



(a) (b)

Fig. 2. Histogram of the pixel intensity values: (a) before and (b) after the Personal Mean Normalization.

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

$(i = 1, \ldots, M)$, where $\mathbf{x}_i(p)$ denotes the 3D location of the i^{th} landmark and $p = \{s, \alpha, \beta, \gamma, \mathbf{q}, \mathbf{t}\}$ 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. $\bar{\mathbf{x}}_i = [\bar{x}_i, \bar{y}_i, \bar{z}_i]^T$ and $\bar{\mathbf{x}} = [\bar{\mathbf{x}}_1; \ldots; \bar{\mathbf{x}}_M]$) and $\mathbf{P}$ denotes the projection matrix to 2D.

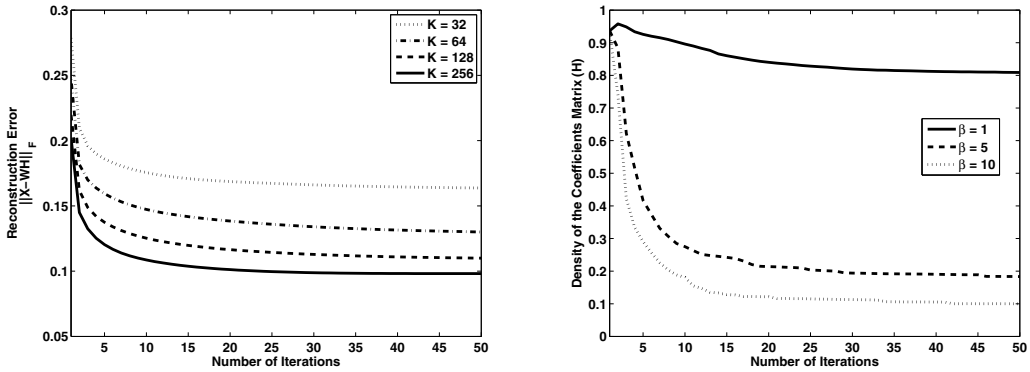
The interested reader is referred to [19] for the details of the CLM algorithm.

B. Extracted Localized Patches

1) Procrustes Normalization: For any shape $\mathbf{x}$, Procrustes transformation applies translation, uniform scaling and rotation to match the reference shape $\mathbf{x}_r$ in Euclidean norm. The minimum of this cost function is called the Procrustes distance. We applied it for 2D transformations.

2) Patch Selection: Following the tracking and registration of facial landmark points, normalized texture patches are extracted around the mouth, eyes, and nose regions. The points selected are based on their relevance to specific AU. For instance, AU 15 (depressor anguli) pulls the lip corners down and causes pouching of the soft tissue below them. To detect AU 15, then, we wish to extract patches from these regions. The lateral lip corners, as well as the nasolabial furrow above them, are relevant to AU 12 (zygomatic major), which pulls the lip corners obliquely in smiling. Point selection is informed by our knowledge of the functional facial anatomy of expression.

3) Personal Mean Texture Normalization: In the next step, personal mean texture normalization is performed. We calculate an average texture for each subject and each patch (the so called personal mean texture) and compute the differences between each texture patch and the personal mean texture. This step is important, because it removes individual differences in texture. Fig. 2 shows the pixel intensity histograms of a texture sequence before and after personal mean normalization. Note, that we normalize the intensities to [0,1] interval and shift the mean to 0.5 to preserve the non-negativity property of the data for the matrix factorization step.



(a) (b)

Fig. 3. Characteristics of the NMF method: (a): reconstruction error with different number of components retained [K]. (b) Density of the coefficients matrix with different regularization parameters [β].

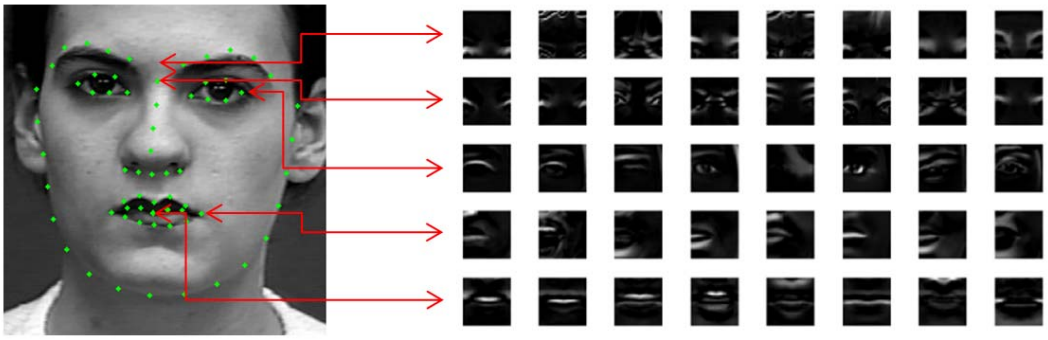


Fig. 4. Some of the NMF filters and the corresponding locations.

C. Non-negative Matrix Factorization

Given a non-negative data matrix $\boldsymbol{X} = [x_1, ...x_n] \in \mathbb{R}_+^{d\times n}$, each column of X is a data point. NMF aims to find two non-negative matrices $\boldsymbol{W} \in \mathbb{R}_+^{d\times m}$ and $\boldsymbol{H} \in \mathbb{R}_+^{m\times n}$ which minimize the following cost function

$$\mathbf{J} = \frac{1}{2}\|\mathbf{X} - \mathbf{WH}\|_F^2, s.t. \mathbf{W} \geq 0, \mathbf{H} \geq 0, \qquad (3)$$

where $\|.\|_F$ is the Frobenius norm.

This formulation is obtained from the log-likelihood function under the assumption of Gaussian error. $\mathbf{W}$ called basis matrix and $\mathbf{H}$ called coefficient matrix. The interpretation is that each data point is a non-negative combination of the basis vectors.

This formulation has issues of scale-invariance and non-unique solution, which imply that the non-negativity constrained on the least squares is insufficient in some cases. Sparsity is a popular regularization principle in statistical modelling [20].

In this work, we regularized the basis matrix and enforced sparsity on the coefficients by minimizing the following cost function:

$$\mathbf{J} = \frac{1}{2}\|\mathbf{X} - \mathbf{WH}\|_F^2 + \alpha\|\mathbf{W}\|_F^2 + \beta\sum_{i=1}^{n}\|\mathbf{h}_i\|_1^2, \qquad (4)$$

where h_i is the i^{th} column of $\mathbf{H}$. For minimizing this function we used an alternating active set method [21]. Fig. 3 shows the characteristics of the factorization with different regularization parameters. As we increase the sparsity constraint (parameter β), the density of the coefficients matrix decreases.

The interested reader is referred to [22] for a detailed survey of the NMF techniques. Fig. 4 depicts some of the acquired filters and the corresponding locations on the face.

D. Support Vector Machines for AU Intensity Estimation

Our Support Vector Machines make another building block of the algorithms, since after we factorize the texture patches, we perform an SVM-based classification using the different emotions as the class labels and a regression procedure for AU intensity estimation.

Support Vector Machines (SVMs) are powerful for both binary and multi-class classification and regression. SVMs 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.

We are given sample and label pairs $(\mathbf{x}^{(k)}, y^{(k)})$ with $\mathbf{x}^{(k)} \in \mathbb{R}^m$, $y^{(k)} \in \{-1, 1\}$, and $k = 1, ..., K$. Here, for class '1' and for class '2' $y^{(k)} = 1$ and $y^{(k)} = -1$, respectively. We also have a set of feature vectors $\boldsymbol{\phi}(= [\phi_1; \ldots; \phi_J])$:

$\mathbb{R}^m \rightarrow \mathbb{R}^J$, where J might be infinite. The support vector classification seeks to minimize the cost function

$$\min_{\mathbf{w},b,\xi} \frac{1}{2}\mathbf{w}^T\mathbf{w} + C\sum_{k=1}^{K}\xi_k \tag{5}$$

$$y^{(k)}(\mathbf{w}^T\boldsymbol{\phi}(\mathbf{x}^{(k)}) + b) \geq 1 - \xi_k, \;\; \xi_k \geq 0. \tag{6}$$

where ξ_k $(k = 1, \ldots, K)$ 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.

Support vector regression (SVR) has a very similar form to support vector machine. For details on SVR techniques, the interested reader is referred to the literature, e.g., to [23] and the references therein.

In every AU intensity experiment we used the L_2-loss L_2-regularized least-squares SVM (LS-SVM). This least squares SVM formulation for the linear case modifies (5) and (6) to a loss function

$$J(\mathbf{w}, b) = \frac{1}{2}\mathbf{w}^T\mathbf{w} + \frac{\mu}{2}\sum_{k=1}^{K}\left(y^{(k)} - \left(\mathbf{w}^T\boldsymbol{\phi}(\mathbf{x}^{(k)}) + b\right)\right)^2 \tag{7}$$

that should be minimized.

III. DATASETS

We used two major databases that differ in type of expression, duration, and granularity of FACS intensity annotation. Cohn-Kanade [24], Cohn-Kanade Extended [25] and Enhanced Datasets [26] include relatively brief (mean duration = 20 frames) sequences of posed facial behavior. Intensity coding has three ordinal levels. The Binghamton University Spontaneous 3D Dynamic Facial Expression Database includes longer (approximately 600 frames) sequences of spontaneous facial behavior in response to diverse emotion inductions, and intensity is annotated with six levels of ordinal intensity for the subset of AU [27].

A. Datasets based on the Cohn-Kanade Facial Expression Database

The Cohn-Kanade Extended Facial Expression (CK+) Database [25] 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 subjects and 593 frontal image sequences. From these, 118 subjects are annotated with the seven universal emotions (anger, contempt, disgust, fear, happy, sad and surprise). The image sequences are annotated with 68 landmark points. Action units are also provided with this database for the apex frame. The original Cohn-Kanade Facial Expression Database distribution [24] 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.

TABLE I

DISTRIBUTION OF AU INTENSITIES IN THE CK ENHANCED DATASET.

Cohn-Kanade Enhanced							
Int	AU1	AU2	AU4	AU5	AU6	AU7	AU9
0	6793	7327	6460	7654	7165	7186	7891
1	536	336	763	311	546	556	257
2	1128	794	1234	492	746	715	309
Int	AU12	AU15	AU17	AU23	AU24	AU25	AU27
0	6742	7579	6622	7871	7727	4950	7716
1	378	330	716	177	235	482	111
2	1337	548	1119	409	495	3025	630

TABLE II

DISTRIBUTION OF AU INTENSITIES IN THE BINGHAMTON DATASET.

Binghamton Spontaneous Database		
Int	AU12	AU14
0	9	502
1	1902	2966
2	5970	3701
3	6133	2684
4	3770	1352
5	1540	6

To provide additional ground-truth, the RPI ISL research group manually re-annotated the original CK dataset [26] frame-by-frame. This annotation contains temporal segments of 14 Action Units. The presence of 14 AUs were annotated with three intensities: absent, present but with low intensity, and present. This dataset is referred to as the CK Enhanced Dataset. The distribution of AU intensities can be seen in Table I.

B. Binghamton Spontaneous Dynamic Facial Expression Database

The Binghamton dataset [27] includes video from various emotion elicitation procedures. Forty-one participants (56% female, 48.7% white, average age 20.2 years) were filmed with a frontal camera (520x720 pixel resolution) while engaging in eight tasks designed to elicit emotions such as anxiety, surprise, embarrassment, fear, pain, anger, and disgust. Example tasks include being surprised by a loud sound, submerging a hand in ice water, and smelling rotten meat. For each task, the 20-second segment with the highest AU density was identified; this segment then was coded for AU onset (start) and offset (end) by certified and reliable FACS coders.

After AU 12 and AU 14 onsets and offsets were coded, short videos were made of each AU event. An event is defined as contiguous frames from onset (start) to offset (end) of the AU. Certified and reliable FACS coders then coded frame-level intensity using custom software. This coding involved assigning each video frame a numerical value with 0 representing the absence of the AU and 1 through 5 representing trace through maximum intensity, as defined by FACS. The distribution of AU intensities can be seen in Table II.

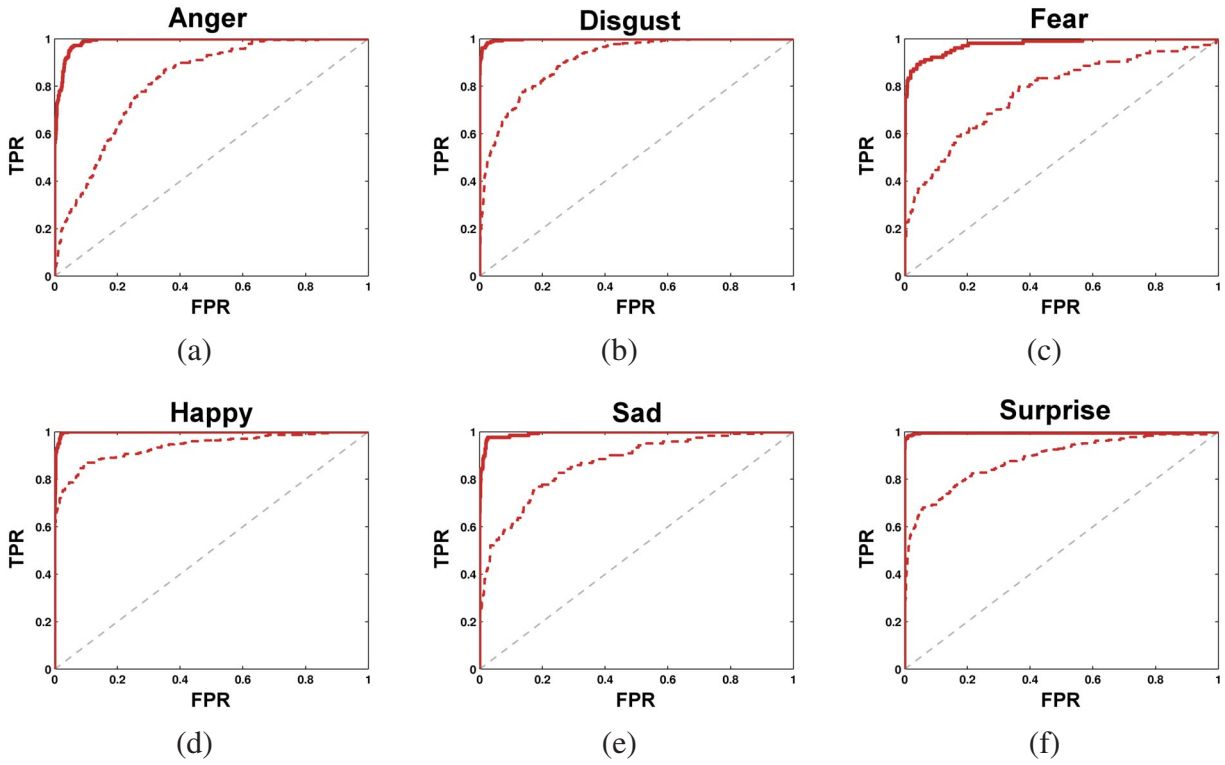


Fig. 5. ROC curves of the different emotion classifiers: (a) anger, (b) disgust, (c) fear, (d) happy, (e) sad and (f) surprise. The solid lines show the performance on the last frames of the sequences, the dashed lines represent the performance on the 6th frame.

TABLE III

COMPARISONS WITH HAND-DESIGNED SPATIO-TEMPORAL GABOR FILTERS (WU ET AL. 2010 [28]) AND LEARNED SPATIO-TEMPORAL ICA FILTERS (LONG ET AL. 2012 [29]) ON THE FIRST 6 FRAMES.

	Anger	Disgust	Fear	Happy	Sad	Surprise	Average
Wu et al. 2010	0.72	0.70	0.64	0.84	0.83	0.87	0.77
Long et al. 2012	0.77	0.71	0.69	0.89	0.85	0.89	0.80
This work	0.82	0.91	0.77	0.94	0.87	0.89	0.86

TABLE IV

COMPARISONS WITH BOOSTED DYNAMIC FEATURES (YANG ET AL. 2009 [30]) ON THE LAST FRAMES OF THE SEQUENCES.

	Anger	Disgust	Fear	Happy	Sad	Surprise	Average
Yang et al. 2009	0.92	0.96	0.84	0.98	0.95	1.00	0.94
Long et al. 2012	0.93	0.99	0.96	0.99	0.99	1.00	0.98
This work	0.99	1.00	0.98	1.00	0.99	0.99	0.99

IV. EXPERIMENTS

We executed a number of evaluations to judge the strength of the proposed representation. Studies concern (i) discrete emotion classification, (ii) the performances of continuous AU intensity estimation on posed and (iii) spontaneous datasets.

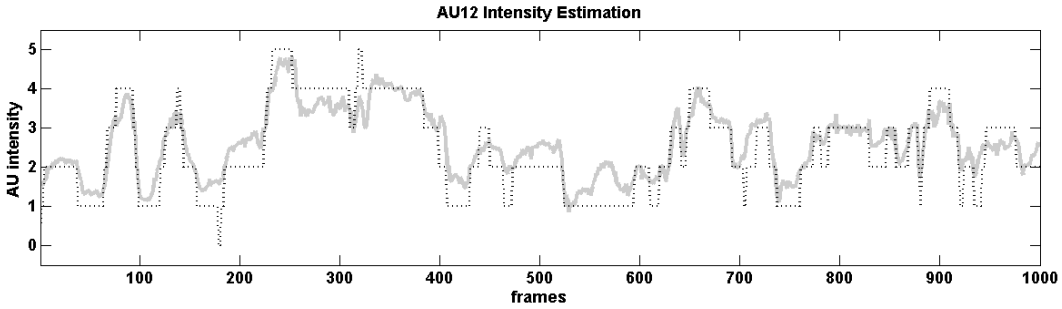
A. Preliminary Study on Discrete Emotion Classification

In this experiment we used the CK+ dataset with the discrete emotion labels. We trained a binary SVMs using the leave-one-subject-out cross validation method. To test the performance of recognizing emotions with low intensity, in one experiment we used only the first 6th frames of the sequences. In the other experiment we used the last frame of each sequence which correspond to the apex of the expression. The acquired ROC performance curves of the classification are depicted in Fig. 5.

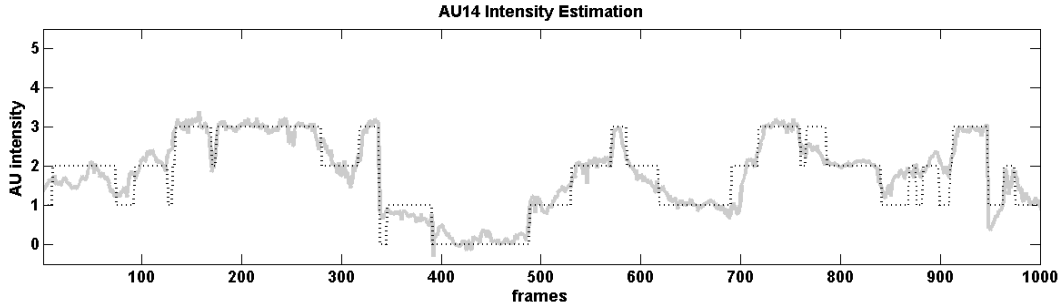
TABLE V

AU INTENSITY ESTIMATION PERFORMANCE ON THE CK ENHANCED DATASET. MSE = MEAN SQUARED ERROR, PCC = PEARSON'S CORRELATION COEFFICIENT.

AU	MSE	PCC
1	0.115	0.913
2	0.082	0.932
4	0.163	0.890
5	0.109	0.895
6	0.121	0.903
7	0.152	0.844
9	0.097	0.907
12	0.099	0.934
15	0.155	0.898
17	0.147	0.886
23	0.185	0.828
24	0.207	0.833
25	0.170	0.918
27	0.048	0.939



(a)



(b)

Fig. 6. The continuous SVR output for (a) AU12 (lip corner puller) and (b) AU14 (dimpler) intensity estimation. The dotted lines show the ground truth intensities.

These results from the NMF based feature representation are better than the available best result that uses spatio-temporal ICA and boosted dynamic features, see our comparisons in Tables III and IV.

B. Study on AU Intensity Estimation on Posed Data

A number of works demonstrate that emotion recognition from estimation of AU intensities - i.e., following the practice of FACS coders - is comparable or better than direct emotion recognition. Also, for situation analysis, recognition of facial expressions beyond the basic emotions is of high relevance and such information is encoded into the AU intensities. In turn, we studied AU intensity estimation on posed expression data.

Encouraged by the results of the previous experiment, we decided to use the data of the Enhanced CK database and tuned Support Vector Regressors for estimating AU intensities. In this dataset 14 AUs were annotated frame-by-frame with three intensities in the Enhanced CK database, namely (a) absent, (b) present but with low intensity, and (c) present. We assigned 0, 1 and 2 values for these levels, respectively. This case can be approximated with the L_2-loss L_2-regularized least-squares SVM (LS-SVM).

We evaluated the performance of the estimation using Mean Squared Error (MSE) and Pearson's correlation coefficient (PCC). The results can be seen in Table V.

The intensities of AUs with large distortions (like AU12 or AU27) can be estimated with high correlations, whereas more subtle AUs, like AU24 (lip pressor) have slightly higher estimation errors.

TABLE VI
AU INTENSITY ESTIMATION PERFORMANCE ON THE BINGHAMTON DATASET. MSE = MEAN SQUARED ERROR, PCC = PEARSON'S CORRELATION COEFFICIENT.

AU	MSE	PCC
12	0.485	0.818
14	0.210	0.837

C. Study on AU12/AU14 Intensity Estimation on Spontaneous Data

The major part of CK+ and CK Enhanced datasets contains posed facial expressions that may considerably differ from the spontaneous ones [31]. Spontaneous expressions, which may differ in intensity, timing, and co-occurrence with other actions [32], [5], [2].

In this experiment we used the Binghamton dataset [27] which has frame-level intensity coding for AU12 (lip corner puller) and AU14 (dimpler) on a six-point scale: 0 representing the absence of the AU and 1 through 5 representing trace through maximum intensity, as defined by FACS.

We used LS-SVM for the linear estimation of the intensities of these two AUs using leave-one-subject-out cross validation method. The results for AU12 and AU14 intensity estimation can be seen in Table VI. The estimation error of AU12 intensity is higher than in the previous experiment, that used a posed dataset.

The output of the support regressor for two different

subjects can be seen in Fig. 6. As illustrated, high precision intensity estimation can be achieved including in difficult cases like AU14.

V. DISCUSSION AND SUMMARY

We are interested in situation analysis and human-computer interaction that call for continuous AU intensity estimation.

We studied facial expression recognition and continuous AU intensity estimation using the localized, sparse representation. We found state-of-the-art performance in emotion-specified expression recognition. In comparison with spatio-temporal ICA based method [29] and boosted dynamic features [30], better results were obtained. We then used the method in the task of continuous AU intensity estimation and demonstrated the performance of the proposed approach on both posed and spontaneous datasets.

Our results compare favorably to other methods in the literature, however direct comparison with other studies is difficult, because of differences in datasets, reliability in coding (which is often unknown) and performance metrics.

In sum, part-based, sparse information is very efficient for facial expression analysis. Further improvements are expected for methods that include temporal information [14]. Such improvements are necessary for situation analysis and human-computer interaction.

VI. ACKNOWLEDGMENTS

This work was supported in part with support from NIH grant R01 MH096951 to the University of Pittsburgh.

REFERENCES

[1] Y. Tian, T. Kanade, and J. Cohn, "Facial expression analysis," *Handbook of face recognition*, pp. 247–275, 2005.

[2] Z. Zeng, M. Pantic, G. Roisman, and T. Huang, "A survey of affect recognition methods: Audio, visual, and spontaneous expressions," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 31, no. 1, pp. 39–58, 2009.

[3] F. D. la Torre and J. F. Cohn, "Ri: Medium: Unsupervised and weakly-supervised discovery of facial events," in *ed: NSF*, 2010.

[4] J. F. Cohn and K. L. Schmidt, "The timing of facial motion in posed and spontaneous smiles," *J. Wavelets, Multi-resolution and Information Processing*, vol. 2, pp. 1–12, 2004.

[5] Z. Ambadar, J. Cohn, and L. Reed, "All smiles are not created equal: Morphology and timing of smiles perceived as amused, polite, and embarrassed/nervous," *Journal of Nonverbal Behavior*, vol. 33, pp. 17–34, 2009. [Online]. Available: http://dx.doi.org/10.1007/s10919-008-0059-5

[6] P. Ekman, W. Friesen, and J. Hager, *Facial Action Coding System (FACS): Manual*. Salt Lake City (USA): A Human Face, 2002.

[7] N. Zaker, M. H. Mahoor, W. I. Mattson, D. S. Messinger, and J. F. Cohn, "Intensity measurement of spontaneous facial actions: Evaluation of different image representations," in *International Conference on Development and Learning*, 2012.

[8] M. Mahoor, S. Cadavid, D. Messinger, and J. Cohn, "A framework for automated measurement of the intensity of non-posed facial action units," 2009, pp. 74–80.

[9] S. Kaltwang, O. Rudovic, and M. Pantic, "Continuous pain intensity estimation from facial expressions," in *Advances in Visual Computing*, ser. Lecture Notes in Computer Science, vol. 7432. Heidelberg: Springer, 2012, pp. 368–377.

[10] Z. Hammal and J. F. Cohn, "Temporal coordination of head motion in couples with history of interpersonal violence," in *IEEE International Conference on Automatic Face and Gesture Recognition*, 2013.

[11] K. Prkachin and P. Solomon, "The structure, reliability and validity of pain expression: Evidence from patients with shoulder pain," *Pain*, vol. 139, pp. 267–274, 2008.

[12] P. Lucey, J. F. Cohn, K. M. Prkachin, P. Solomon, and I. Matthews, "Painful data: The unbc-mcmaster shoulder pain expression archive database," *Image, Vision, and Computing Journal*, vol. 30, pp. 197–205, 2012.

[13] R. Woodworth and H. Schlosberg, *Experimental Psychology*. Holt, 1954. [Online]. Available: http://books.google.com/books?id=yL9zr_pGXugC

[14] M. A. Nicolaou, H. Gunes, and M. Pantic, "Output-associative rvm regression for dimensional and continuous emotion prediction," in *IEEE International Conference on Automatic Face and Gesture Recognition*, 2011.

[15] H. Gunes and M. Pantic, "Automatic, dimensional, and continuous emotion recognition," *International Journal of Synthetic Emotions*, vol. 1, pp. 68–99, 2010.

[16] J. F. Cohn, "Foundations of human-centered computing: Facial expression and emotion," in *International Joint Conference on Artificial Intelligence*, 2006.

[17] W. A. Freiwald, D. Y. Tsao, and M. S. Livingstone, "A face feature space in the macaque temporal lobe," *Nature Neuroscience*, vol. 12, pp. 1187–1196, 2009.

[18] J. Ngiam, P. W. W. Koh, Z. Chen, S. A. Bhaskar, and A. Y. Ng, "Sparse filtering," in *Advances in Neural Information Processing Systems 24*, J. Shawe-Taylor, R. Zemel, P. Bartlett, F. Pereira, and K. Weinberger, Eds., 2011, pp. 1125–1133.

[19] J. M. Saragih, S. Lucey, and J. F. Cohn, "Deformable model fitting by regularized landmark mean-shift," *International Journal of Computer Vision*, vol. 91, pp. 200–215, 2011.

[20] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society*, vol. 58, pp. 267–288, 1994.

[21] H. Kim and H. Park, "Nonnegative matrix factorization based on alternating nonnegativity constrained least squares and active set method," *SIAM Journal on Matrix Analysis and Applications*, vol. 30, pp. 713–730, 2008.

[22] Y. Wang and Y. Zhang, "Non-negative matrix factorization: a comprehensive review," *IEEE Transactions on Knowledge and Data Engineering*, vol. PP, pp. 1–20, 2012.

[23] J. Shawe-Taylor and N. Cristianini, *Kernel Methods for Pattern Analysis*. UK: Cambridge University Press, 2004.

[24] T. Kanade, J. F. Cohn, and Y. Tian, "Comprehensive database for facial expression analysis," in *IEEE International Conference on Automatic Face and Gesture Recognition*. IEEE, 2000, pp. 46–53.

[25] 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 *3rd IEEE Workshop on CVPR for Human Communicative Behavior Analysis*. IEEE, 2010, pp. 94–101.

[26] Y. Tong, W. Liao, and Q. Ji, "Facial action unit recognition by exploiting their dynamic and semantic relationships," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, pp. 1683–1699, 2007.

[27] X. Zhang, L. Yin, A. Horowitz, M. Reale, S. Canavan, P. Liu, and J. F. Cohn, "A 3d spontaneous dynamic facial expression database," in *IEEE International Conference on Automatic Face and Gesture Recognition*. IEEE, 2013, pp. –.

[28] T. Wu, M. S. Bartlett, and J. R. Movellan, "Facial expression recognition using gabor motion energy filters," in *Workshop on CVPR for Human Communicative Behavior Analysis*. IEEE, 2010, pp. 42–47.

[29] 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*, vol. 93, pp. 126–132, 2012.

[30] P. Yang, Q. Liu, and D. N. Metaxas, "Boosting encoded dynamic features for facial expression recognition," *Pattern Recognition Letters*, vol. 30, pp. 132–139, 2009.

[31] T. McLellan, L. Johnston, J. Dalrymple-Alford, and R. Porter, "Sensitivity to genuine versus posed emotion specified in facial displays," *Cognition and Emotion*, vol. 24, pp. 1277–1292, 2010.

[32] K. L. Schmidt, Z. Ambadar, J. F. Cohn, and L. I. Reed, "Movement differences between deliberate and spontaneous facial expressions: Zygomaticus major action in smiling," *Journal of Nonverbal Behavior*, vol. 30, pp. 37–52, 2006.