Multi-output Laplacian Dynamic Ordinal Regression for Facial Expression Recognition and Intensity Estimation

Ognjen Rudovic¹, Vladimir Pavlovic² and Maja Pantic¹,³
¹ Comp. Dept., Imperial College London, UK
² Dept. of Computer Science, Rutgers University, USA
³ EEMCS, University of Twente, Netherlands
{o.rudovic,m.pantic}@imperial.ac.uk http://ibug.doc.ic.ac.uk
vladimir@cs.rutgers.edu http://seqam.rutgers.edu

Abstract

Automated facial expression recognition has received increased attention over the past two decades. Existing works in the field usually do not encode either the temporal evolution or the intensity of the observed facial displays. They also fail to jointly model multidimensional (multi-class) continuous facial behaviour data; binary classifiers - one for each target basic-emotion class - are used instead. In this paper, intrinsic topology of multidimensional continuous facial affect data is first modeled by an ordinal manifold. This topology is then incorporated into the Hidden Conditional Ordinal Random Field (H-CORF) framework for dynamic ordinal regression by constraining H-CORF parameters to lie on the ordinal manifold. The resulting model attains simultaneous dynamic recognition and intensity estimation of facial expressions of multiple emotions. To the best of our knowledge, the proposed method is the first one to achieve this on both deliberate as well as spontaneous facial affect data.

1. Introduction

Facial behavior is believed to be the most important source of information when it comes to affect, attitude, intentions, and social signals interpretation [2]. Automatic facial expression recognition has therefore been an active topic of research for more than two decades [17, 25]. Most systems developed so far attempt automatic recognition of prototypic facial expressions of six basic emotions (anger, happiness, fear, surprise, sadness, and disgust). The main criticism that these works received from both cognitive and computer scientists is that the methods are not applicable in real-world situations, where subtle changes in both appearance and temporal evolution of facial expressions typify the displayed facial behavior [2, 1]. Current works in the field usually do not encode either the intensity of the observed facial appearance changes or the evolution of these changes in time [25]. Instead, current approaches usually apply six binary classifiers - one for each target prototypic facial expression of emotion - that code input face imagery as either belonging to the target class or not.

Exceptions to this trend include a small number of works on automatic coding of facial imagery in terms of either temporal segments of facial actions (e.g., [22, 11, 14, 18]), or temporal segments of prototypic expressions of emotions (e.g., [7, 10]), or a small number of prototypic facial expression intensity levels (e.g., [5]). Some of the past works in the field have proposed methods that could be used for recognition of facial expression temporal segments and/or intensity levels (e.g., [23, 20]), but did not actually report any quantitative results for that task. Most of these works use temporal graphical models being either generative (e.g., Hidden Markov Models (HMM), [7, 22, 11]) or discriminative (e.g., CRFs [10]) trained for recognition of temporal segments of a target facial expression. However, most of these approaches fail to jointly model different emotions, making the models suboptimal for the emotion modeling task.

A method that does not conform to this rule is the H-CORF model [9], which has been successfully used for simultaneous recognition of multiple emotion-related expressions and their intensities. Yet, despite improvements over other dynamic models (e.g., HMM or standard CRF), H-CORF relies on linear feature models. Such 'simple' feature representation is usually not discriminative enough for recognition and intensity estimation of facial behaviour due to the large variation in expressions and their intensity among different subjects.

In this paper, we propose to model topology of the input data by a low-dimensional manifold that preserves discriminative information about various facial expressions of
emotions and ordinal relationships between their intensities while being largely invariant to intra- and inter-subject variations. We incorporate this topology into the H-CORF framework for dynamic ordinal regression by constraining H-CORF parameters to lie on this nonlinear manifold. To keep the model computationally tractable we adopt a locally linear approximation of the otherwise nonlinear parameter manifold. This manifold approximation is then coupled to and jointly estimated with the H-CORF. In this manner we directly find the most discriminative features for dynamic recognition of emotions and their intensities. To the best of our knowledge, this is the first method that performs simultaneous recognition of multiple facial expressions of emotions and ordinal relationships between their intensities, and (iii) intrinsic topology of multidimensional continuous facial affect data, encoded by an ordinal manifold structure.

In what follows, we consider a $K$-class classification problem, where we let $c \in \{1, \ldots, K\}$ be the nominal category (i.e., emotion class). Each nominal category $c$ is assumed to have $R$ different ordinal scales (i.e., emotions intensities), denoted as consecutive integers $h_r \in \{1, \ldots, R\}$ that keep the ordering information. The observations, denoted by $x = x_1 \ldots x_T$ and where the sequence length $T$ can vary from instance to instance, serve as input covariates for predicting both $c$ and $h$. If not stated otherwise, we assume a fully supervised setting: we are given a training set of $N$ data triplets $D = \{(c^i, h^i, x^i)\}^N_{i=1}$, which are i.i.d. samples from an underlying but unknown distribution.

The remainder of the paper is organized as follows. We give a short overview of the models for dynamic ordinal regression in Sec. 2. We describe the manifold learning approach employed in the proposed model in Sec. 3. The proposed model for dynamic ordinal regression is described in Sec. 4. Sec. 5 shows the experimental results. Sec. 6 concludes the paper.

2. Dynamic Ordinal Regression

The goal of ordinal regression is to predict the label $h$ of an item represented by a feature vector\footnote{We use the notation $x$ interchangeably for both a sequence observation $x = \{x_t\}$ and a vector, which is clearly distinguished by context.} $x \in \mathbb{R}^p$ where the output indicates the preference or order of this item: $h = 1 < h = 2 < \ldots < h = R$. Modeling of the item orders can be accomplished by means of standard static ordinal regression models (e.g [21, 4, 3]), which, in contrast to multi-class classification models, preserve ordering relationships between different labels. Nevertheless, static models for ordinal regression ignore temporal correlations between the labels, which is of essence when dealing with sequential data. In what follows, we describe two recently proposed dynamic models for ordinal regression [10, 9].

2.1. Conditional Ordinal Random Field (CORF)

CORF [10] is an extension of standard CRF [12] to structured output ordinal regression setting. It models the distribution of a set (sequence) of ordinal variables $h = \{h_r\}, h_r \in \{1, \ldots, R\}$, conditioned on inputs $x$. As in standard CRF, the distribution $P(h|x)$ has a Gibbs form clamped on the observation $x$:

$$P(h|x; \theta) = \frac{1}{Z(x; \theta)} e^{s(x; h; \theta)},$$

where $Z(x; \theta) = \sum_{h \in \mathcal{H}} e^{s(x; h; \theta)}$ is the normalizing partition function ($\mathcal{H}$ is a set of all possible output configurations), and $\theta$ are the parameters\footnote{For brevity, we often drop the dependency on $\theta$ in our notation.} of the score function $s(\cdot)$, defined as

$$s(x, h; \theta) = \sum_{r \in V} \Psi_r^{(V)}(x, h_r; \{a, b, \sigma\}) + \sum_{e \in E} \Psi_e^{(E)}(x, h_r = k, h_s = l),$$

thus, summing up the influence of the node features ($\Psi_r$) and the edge features ($\Psi_e$) on the model output. However, in contrast to standard CRF, CORF employs the modeling strategy of static ordinal regression methods (see [3]) to define the node features $\Psi_r$. Specifically, the probabilistic ranking likelihood, $P(h = c|f(x)) = P(f(x) \in [b_{c-1}, b_c))$, is used, where $f(x) = a^T \phi(x)$ is the linear model in the induced feature space $\phi(x)$. Thus, $a$ projects the features $\phi(x)$ on a single line divided into $R$ bins, with the binning parameters $b = [-\infty = b_0, \ldots, b_R = +\infty]^T$, which satisfy the ordinal constraints ($b_i < b_{i+1}, \forall i$). Under the Gaussian noise assumption, the ranking likelihood becomes

$$P(h = c|f(x)) = \Phi\left(\frac{b_c - f(x)}{\sigma}\right) - \Phi\left(\frac{b_{c-1} - f(x)}{\sigma}\right),$$

where $\Phi(\cdot)$ is the standard normal cdf, and $\sigma$ is the parameter that controls the steepness of the likelihood function [3]. The ranking likelihood in (3) is used to set the node potential at node $r$ in the CORF model as $\Psi_r^{(V)}(x, h_r = c) = \log P(h = c|f(x))$, while the edge features, $\Psi_e^{(E)}(x, h_r, h_s)$, are set as

$$I(h_r = k \land h_s = l) \mathbb{R}_x^R \otimes |\phi(x_r) - \phi(x_s)|.$$

$I(\cdot)$ is the indicator function that returns 1 (0) if the argument is true (false), and $\otimes$ denotes the Kronecker product. Finally, the parameters of the CORF model are stored in $\theta = \{a, b, \sigma, u\}$, and $\phi(x) = [1, x^T]^T$, as in [10].
2.2. Multi-output CORF (M-CORF)

The output of the CORF model introduced in the previous section comprises ordinal variables $h_c \in \{1 \ldots R\}$ corresponding to a single class. To deal with multiple classes, $c = \{1, \ldots, K\}$, H-CORF [9] combines $K$ independent CORF models by employing the modeling strategy of Hidden CRF (H-CRF) [16], resulting in a new score function

$$s(c, x, h; \Omega) = \sum_{k=1}^{K} I(c = k) \cdot s(x, h; \theta_k)$$

(5)

where $s(x, h; \theta_k)$ is defined by (2), and $\Omega = \{\theta_k\}_{k=1}^{K}$, where $\theta_k = \{a_k, b_k, \sigma_k, u_k\}$ for $k = 1 \ldots K$, are the parameters of H-CORF. With the new score function, the joint and class conditional distributions are given by

$$P(c, h|x) = \frac{\exp(s(c, x, h))}{Z(x)},$$

(6)

$$P(c|x) = \sum_h P(c, h|x) = \sum_h \frac{\exp(s(c, x, h))}{Z(x)}$$

(7)

Evaluation of the class-conditional $P(c|x)$ depends on the partition function $Z(x) = \sum_c \sum_h \exp(s(c, x, h))$ and the class-latent joint posteriors $P(c, h|x, h)$. Both can be computed from independent consideration of $K$ individual CORFs. Note that the H-CORF model treats ordinal variables $h$ as latent variables, and, thus, does not employ the corresponding labels during training. In what follows, we also consider a fully supervised setting in which labels for both classes $c$ and ordinal variables $h$ are known. To distinguish this setting from standard H-CORF, we call it Multi-output CORF (M-CORF).

Shared-paramater M-CORF: SM-CORF

In M-CORF, each CORF component is assigned an independent set of parameters $\{\theta_k\}$. However, since our classes are related\(^3\), it seems natural to use some shared parameters that 'couple' individual CORF components so that similarities across them can be exploited. Furthermore, the parameter sharing should constrain the parameters to a more plausible region of the parameter space. This is achieved by modeling intensities of different emotions on a common real line, divided by the binning parameters $b$, which are shared among all classes (emotions). We call this model the Shared-parameter M-CORF (SM-CORF) model, where a set of parameters $\{b, \sigma\}$ is shared among all CORF components, while ordinal projections $a_k$ and transition matrix $u_k$ are emotion-specific. In the same way, we define the Shared parameter H-CORF (SH-CORF) model.

3This comes from the fact that temporal segments of each emotion class are labeled as neutral, onset and apex, where neutral should be common for all emotions.

3. Manifold for Ordinal Regression

The goal of standard manifold learning is to discover a latent space in which topology of the input features $x$, sometimes also informed by labels of $x$, is preserved. Such data representation may be more discriminative and better suited for modeling of dynamic ordinal regression. In what follows, we first describe an unsupervised method for manifold learning. We then extend this method to obtain a manifold that satisfies ordinal constraints. Finally, we show how this ordinal manifold can be incorporated into the HCORF framework for dynamic ordinal regression.

3.1. Locality Preserving Projection (LPP)

Locality Preserving Projection (LPP) [8] is the optimal linear approximation to the eigenfunctions of the Laplace Beltrami operator on the manifold, which is capable of discovering nonlinear manifold structure. It uses the notion of the Laplacian of the graph to compute a transformation matrix which maps the data points to a subspace. Formally, it first constructs an undirected graph $G = (V, E)$, where each edge is associated with a weight $W_{ij}$. The elements of the weight matrix can, for instance, be computed by means of the heat kernel [8]

$$W_{ij} = \exp(-\sigma_w^{-2}\|x_i - x_j\|^2),$$

(8)

where $\sigma_w$ is the width of the kernel. Based on the weight matrix $W$, it computes the graph Laplacian as $L = D - W$, where $D$ is a diagonal matrix with $D_{ii} = \sum_j W_{ij}$. To obtain the embeddings, the relationship between latent and observed variables is modeled as $z = F x$, where $F$ is the projection matrix. By imposing the orthonormal constraints $(x^T D x = I)$, $F$ is found in a closed form as a solution to the generalized eigenvalue problem:

$$x^T L x F = \lambda x^T D x F,$$

(9)

where the column vectors $F_i, i = 1, \ldots, D_z$, with $D_z$ being the dimension of the manifold, are the eigenvectors corresponding to the minimum eigenvalue solutions ($\lambda_i$) of (9). Thus, the projection $F$ defines the manifold on which inputs $x$ vary more smoothly.

3.2. Supervised Ordinal LPP (SO-LPP)

To obtain a manifold that is better adjusted to emotion classification, [19] proposed a Supervised Locality Preserving Projections (S-LPP) algorithm, which also encodes class information when computing the weight matrix in (8). We extend this algorithm by also encoding the ordering of the class labels, so as to preserve the smooth transitions between different emotion intensities on the manifold. We call this algorithm Supervised Ordinal LPP (SO-LPP) since its proximity matrix $W$ is defined as an ordinal weight matrix.
using the latent variables into the SM-CORF model by constraining its parameters. Furthermore, we use the prior in (12) to obtain a prior over the projection matrix $F$ as

$$p(F|x) = \frac{1}{Z_F} \exp \left( -\frac{\alpha}{2} F x L x^T F^T \right)$$

The role of this prior is to enforce smoothness constraints on the manifold in which we intend to model ordinal regression. Note that these constraints are different from temporal constraints imposed by dynamic features in the SM-CORF model, since the former aim at preserving the topology of our input data.

By using the prior in (12), the likelihood function of the SM-CORF model given by (6) and by assuming a Gaussian prior over the model parameters $\Omega$, $P(\Omega) = \mathcal{N}(\Omega|0, \gamma I)$, we obtain the posterior distribution

$$P(F, \Omega|c, h, x) \propto P(c, h|x, F, \Omega) P(F|x) P(\Omega).$$

If we use the maximum a posteriori (MAP) strategy to estimate the projection matrix $F$ and the model parameters $\Omega$, the topology of our input data will be seamlessly integrated into the model [26]. We call this model the Laplacian SM-CORF (LSM-CORF) model. The importance of the GMRF-based prior in (13) can best be seen in terms of the graphical structure of the resulting model. Namely, this prior introduces an additional graphical structure into the SM-CORF model. Specifically, the graphical structure of the SM-CORF model alone has the form of a chain representing the explicit dependencies only between the labels of the neighbouring nodes. On the other hand, the graphical structure of the GMRF is richer in the sense that it captures dependencies between the labels over the whole dataset.

The MAP estimate of $(F, \Omega)$ can be obtained by minimizing the following objective function:

$$\arg \min_{F, \Omega} \sum_{i=1}^{n} - \ln P(c_i, h_i|x_i, F, \Omega)$$

$$+ \frac{\lambda}{2} F x L x^T F^T + \frac{\gamma}{2} \|\Omega\|^2 + \text{const.}$$

The penalty term $F x L x^T F^T$ has the role of manifold regularization in the LSM-CORF model, while $\lambda$ and $\gamma$ control the complexity of the projection matrix $F$ and the ordinal regression model learned in the latent space, respectively. The Laplacian SH-CORF (LSH-CORF) model is obtained by replacing the likelihood term in (15) with (7) and by using the same analogy as before.

### 4. Laplacian SM-CORF (LSM-CORF)

In this section, we incorporate topology of our input data into the SM-CORF model by constraining its parameters to lie on the ordinal manifold. This is achieved by enforcing the latent variables $u \equiv \phi(x) = Fx$ to be a Gaussian Markov Random Field (GMRF) w.r.t. graph $L$ (see [26] for details). Based on the GMRF representation, we obtain a prior over the latent variables $U = [u_1 u_2 \ldots u_N]$ as

$$p(U) = \prod_{i=1}^{n} p(u_i) = \frac{1}{Z_U} \exp \left( -\frac{\alpha}{2} tr(ULU^T) \right)$$

where $Z_U$ is a normalization constant and $\alpha > 0$ is a scale parameter. Furthermore, we use the prior in (12) to obtain a
matrix $F_0$ via the SO-LPP algorithm, and set the parameters $\Omega$ as in [9]. We then alternate between steps 1-2 in Alg.1 until convergence. Model learning in LSH-CORF requires some additional steps and this is described in Alg.1. The initial projection matrix $F_0$ (i.e., the graph Laplacian) is obtained by dividing each training sequence $x_i$ into $R$ segments with approximately equal length, and by labeling each segment with the corresponding intensity level (i.e. the segment number). After one iteration of BFGS (step 2 in Alg.1), we use the new parameters $(F, \Omega)$ to compute the likelihood of each intensity level $h_i$, where $i = 1, \ldots, R$. These likelihoods are then used to update graph Laplacian in (11). The steps 1-4 in Alg.1 are repeated until convergence of the evidence function.

Algorithm 1 Model Learning in LSH-CORF

Require: $\{c^i, h^i_0, x^i\}_{i=1}^N$ and $(F_0, \Omega_0)$
1. Evaluate the evidence in (15) and calculate the gradients w.r.t. $(F, \Omega)$.
2. Feed the evidence and gradients to the BFGS method.
3. Calculate $P(h = i | F, x, \Omega) = \sum_c P(c, h = i | F, x, \Omega)$, where $i = 1, \ldots, R$.
4. Update graph Laplacian based on (11).
5. Repeat (1-4) until convergence of the evidence in (15).

5. Experiments

In this section we demonstrate the performance of the proposed method in the task of facial expression recognition and its intensity estimation from the frontal view facial images. We use image sequences from two publicly available datasets: the BU-4DFE dataset [24] and the Cohn-Kanade (CK) dataset [13]. Both datasets contain image sequences of different subjects displaying facial expressions of six prototypic emotions: Anger, Disgust, Fear, Happiness, Sadness and Surprise. We select 120 image sequences that come from 30 subjects from BU-4DFE, and 167 image sequences from 98 subjects from CK. All image sequences start with a neutral face evolving to the apex of the target display. Image sequences from the BU-4DFE dataset are sub-sampled so that the sequence lengths in both datasets are about 20-frame long on average. Each image sequence is annotated as one of six prototypic emotions ($c = \{1, \ldots, 6\}$), and each frame is manually labeled into three ordinal categories: neutral ($h = 1$) $\prec$ onset ($h = 2$) $\prec$ apex ($h = 3$).

In this study, we use the locations of a set of characteristic facial points as the input features. In the case of BU-4DFE, we use 39 facial points extracted using the particle-filter-based tracker [15]. For CK, we use 20 facial points extracted using the particle-tree-based tracker [15]. Fig. 6 depicts examples of the tracked sequences. The tracked points are later registered to a reference face and normalized w.r.t. the first frame in each image sequence. Finally, the PCA reduction preserving 95% of the total energy is applied to the input features, giving rise to the 16-dimensional inputs, for BU-4DFE, and to the 24-dimensional inputs, for CK, which are denoted by $x$.

We perform two sets of experiments. In the fully supervised setting, we compare the performance of our LSM-CORF model with: (1) Multi-output Hidden Markov Model (M-HMM), used as the baseline, (2) M-CORF and (3) SM-CORF. In the unsupervised setting, we perform the same experiments using the ‘hidden’ models (H-HMM/H-CORF/SH-CORF/LSH-CORF), all of which treat the intensity levels as latent variable. The M-HMM model is obtained by combining the outputs of standard HMM models trained independently for each emotion category using one-shot learning with $h$ hidden states. In the unsupervised case (H-HMM), the initial estimates of the hidden states $h$ are set as in LSH-CORF (Sec. 4.1). The emotion/level prediction for a given test sequence is accomplished using Viterbi decoding. Note that in this paper we do not include comparison with regular CRFs and static ordinal regression, since the state-of-the-art H-CORF [9] model has been shown to outperform those models in the target task.

In all our experiments, we apply 10-fold cross validation procedure, where each fold contains image sequences of different subjects. We report the accuracy using the mean error rate ($\frac{1}{N} \sum_n I(c_n \neq \tau_n)$) for facial expression recog-
The SH/SM-CORF models exhibit superior performance to the H/M-CORF models in the tasks of facial expression recognition and its intensity estimation. The ordinal manifold dimensionality which resulted in the best performance of the approach in question is used for the training/testing. Here we also include the results obtained by standard H/M-CORF models, which use an independent set of parameters for each emotion class, and are trained/tested using the original PCA-based feature vectors.

Table 1. **BU-4DFE dataset.** The performance of the compared approaches per emotion category. The ordinal manifold dimensionality which resulted in the best performance of the approach in question is used for the training/testing. Here we also include the results obtained by standard H/M-CORF models, which use an independent set of parameters for each emotion class, and are trained/tested using the original PCA-based feature vectors.

<table>
<thead>
<tr>
<th>Method</th>
<th>Mean Error Rate for Facial Expression Recognition</th>
<th>Mean Absolute Loss for Facial Expression Intensity Prediction</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Angry</td>
<td>Disgust</td>
</tr>
<tr>
<td>M-HMM</td>
<td>20.7</td>
<td>51.4</td>
</tr>
<tr>
<td>M-CORF</td>
<td>33.3</td>
<td>33.3</td>
</tr>
<tr>
<td>SM-CORF</td>
<td>58.3</td>
<td>15.8</td>
</tr>
<tr>
<td>LSM-CORF</td>
<td>31.6</td>
<td>15.7</td>
</tr>
<tr>
<td>H-HMM</td>
<td>23.0</td>
<td>40.3</td>
</tr>
<tr>
<td>H-CORF</td>
<td>36.1</td>
<td>37.7</td>
</tr>
<tr>
<td>SH-CORF</td>
<td>40.0</td>
<td>41.6</td>
</tr>
<tr>
<td>LSH-CORF</td>
<td>26.6</td>
<td>16.6</td>
</tr>
</tbody>
</table>

Figure 3. **BU-4DFE dataset.** Facial expression recognition and its intensity estimation achieved by (a) SM-CORF and (b-c) LSM-CORF, in the 3D ordinal manifold learned by the proposed OS-LPP. In SM-CORF, the embeddings remain unchanged during optimization of $\Omega = \{a_c, b, \sigma, u_a\}$, while in LSM-CORF, $\Omega$ and the embedding matrix $F$ are jointly optimized. Both algorithms converged in less than 80 iterations. Below each image, the error rates for facial expression recognition (in %) and mean abs. loss for the intensity estimation (obtained after the number of iterations (it)) are shown. Different colors in the images depict the embeddings of facial expressions of different emotion categories, and ($\ast, \ast, \ast$) correspond to their intensity levels (i.e., neutral, onset and apex), respectively.

5.1. Experiments on the BU-4DFE dataset

To select an optimal manifold for ordinal regression, we test the performance of the compared approaches w.r.t. the size of the ordinal manifold obtained as explained in Sec. 3. The average test errors for facial expression recognition and its intensity estimation are shown in Fig. 1. Here we only report results for SH/SM-CORF models since the regular H/M-CORF models often contain Smiles (of embarrassment) and acted Screams which are sources of confusion with Happiness and Surprise.

We also observed the manifold learning during the model estimation phase. For visualization purpose, we model the ordinal manifolds in 3D. Fig. 3 depicts adaptation of the LS/H/SM-CORF models to the corresponding manifolds. Fig. 3(a) shows the SM-CORF model estimated on the ‘fixed’ manifold, while Fig. 3(b)-(c) show how this manifold changes during estimation of the proposed LSM-CORF model, which jointly estimates the manifold and the CORF parameters. As can be seen from Fig. 3(a), the SM-CORF
5.2. Experiments on the Cohn-Kanade dataset

Fig. 4 shows the performance of the compared approaches w.r.t. the size of the ordinal manifold, while Table 2 shows the performance per emotion category obtained by using optimal ordinal manifolds to train/test the methods. The LSM-CORF model consistently outperforms other models, both in supervised and unsupervised setting. Interestingly, the proposed LSH-CORF model still accurately predicts emotion intensities, which is, in part, contributed to its modeling of the data topology. The confusion matrices in Fig. 5 similarly reflect superior performance of our LSH-CORF model compared to H-CORF [9], which we found to be prone to data overfitting.

6. Conclusions

Modeling the intrinsic topology of the facial affect data is important for deducing discriminative features for dynamic recognition of emotions and their intensity. Standard generative models like HMMs and discriminative models like H-CORF [9] use simple linear feature representation that is unable to capture such topology. In contrast, the proposed LSM-CORF model incorporates this topology into the H-CORF framework, giving rise to a linear approximation of the otherwise non-linear model for dynamic ordinal regression. As evidenced by the results, the proposed method attains effective simultaneous dynamic recognition and intensity estimation of multiple emotions on both deliberately and spontaneously displayed facial expressions.

Acknowledgments

We are grateful to Minyoung Kim for his help throughout the course of this work. This material is based upon work supported by the European Research Council under the ERC Starting Grant agreement no. ERC-2007-StG-203143 (MAHNOB), and by the National Science Foundation under Grant No. IIS 0916812.

References

Table 2. **CK dataset.** The performance of the compared approaches per emotion category. The ordinal manifold dimensionality which resulted in the best performance of the approach in question is used for the training/testing.

<table>
<thead>
<tr>
<th>Method</th>
<th>Mean Error Rate for Facial Expression Recognition</th>
<th>Mean Absolute Loss for Facial Expression Intensity Prediction</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Angry</td>
<td>Disgust</td>
</tr>
<tr>
<td>M-HMM</td>
<td>50.1</td>
<td>22.0</td>
</tr>
<tr>
<td>M-CORF</td>
<td>38.7</td>
<td>32.0</td>
</tr>
<tr>
<td>SM-CORF</td>
<td>35.5</td>
<td>16.1</td>
</tr>
<tr>
<td>LSM-CORF</td>
<td>23.5</td>
<td>12.3</td>
</tr>
<tr>
<td>H-HMM</td>
<td>60.0</td>
<td>22.0</td>
</tr>
<tr>
<td>H-CORF</td>
<td>42.6</td>
<td>1.0</td>
</tr>
<tr>
<td>SH-CORF</td>
<td>32.5</td>
<td>8.00</td>
</tr>
<tr>
<td>LSM-CORF</td>
<td>28.7</td>
<td>9.20</td>
</tr>
</tbody>
</table>

Figure 6. **Continuous prediction of naturalistic expression of emotion.** The shown images are subsampled from the test sequence by factor four. The graphs in between show the estimated probabilities for various facial expressions of emotions and their intensities obtained by the proposed LSM-CORF model. The models are trained using the data from BU-4FE (**left**), and CK (**right**), some examples of which are shown in the bottom row.


