

# Discriminant Multi-Label Manifold Embedding for Facial Action Unit Detection

Anıl Yüce, Hua Gao and Jean-Philippe Thiran

Signal Processing Laboratory (LTS5), École Polytechnique Fédérale de Lausanne, Switzerland

anil.yuce@epfl.ch, hua.gao@epfl.ch, jean-philippe.thiran@epfl.ch

**Abstract**—This article describes a system for participation in the Facial Expression Recognition and Analysis (FERA2015) sub-challenge for spontaneous action unit occurrence detection. The problem of AU detection is a multi-label classification problem by its nature, which is a fact overseen by most existing work. The correlation information between AUs has the potential of increasing the detection accuracy. We investigate the multi-label AU detection problem by embedding the data on low dimensional manifolds which preserve multi-label correlation. For this, we apply the multi-label Discriminant Laplacian Embedding (DLE) method as an extension to our base system. The system uses SIFT features around a set of facial landmarks that is enhanced with the use of additional non-salient points around transient facial features. Both the base system and the DLE extension show better performance than the challenge baseline results for the two databases in the challenge, and achieve close to 50% as F1-measure on the testing partition in average (9.9% higher than the baseline, in the best case). The DLE extension proves useful for certain AUs, but also shows the need for more analysis to assess the benefits in general.

## I. INTRODUCTION

Facial Action Unit (AU) recognition and its applications have become one of the most important fields in facial analysis and computer vision research almost 40 years after their first standardization by the Facial Action Coding System [1]. Also with more commercial tools available, it is being more commonly used for, for example, psychology and psycho-pathology research, in addition to personal use such as Human Computer Interaction, leisure and so on since many years now. However, it is still far from being treated as a solved problem, especially due to difficulties caused by low intensity and short duration of the facial muscle contractions during spontaneous behavior, uncontrolled scene configurations during natural interactions and subjective appearance variations among many other reasons. The Facial Expression Recognition and Analysis Challenge 2015 (FERA2015) [2] is important since it addresses these factors within the datasets used. Both SEMAINE [3] and BP4D [4] databases include spontaneous facial expressions that contain AUs of various intensities and durations that were recorded in a mostly unconstrained manner.

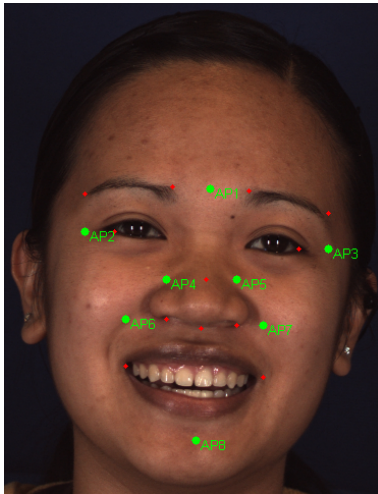
Since the first FERA challenge (FERA 2011 [5],[6]) many advances have been proposed. Variants of Local Binary Patterns (LBP) are still popular for static and dynamic 2D or 3D action unit detection because of their efficiency, e.g. [7],[8]. SIFT (Scale Invariant Feature Transform) descriptors have also been used efficiently within various frameworks ([9],[10]). We participate in the challenge with a system

that also uses SIFT as features on an enhanced set of facial landmarks that includes points around transient facial features with support vector machine (SVM) classification.

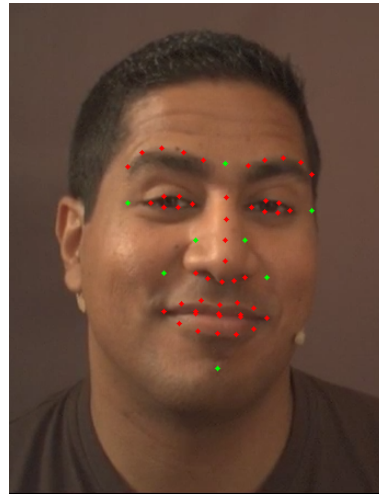
The AUs generally occur in combinations during natural behaviour and these combinations form a correlation pattern between them. However, there is not a lot of work in the literature on the use of multi-label information for AU detection rather than treating them as independent labels. This information may prove valuable since AU recognition is actually a multi-label problem, i.e. a data point belongs to multiple labels. A well-known work that uses this information in a temporal manner is the one by Tong et al. [11]. Mahoor et al. have combined Laplacian Eigenmaps, which is a locality preserving method for embedding the data on a lower dimensional manifold, with spectral regression to learn separate sub-spaces for AUs to detect their intensity [12]. There have also been other attempts to use manifold learning type of projections for facial expression recognition, e.g. [13]. In this work we also propose an extension to our system that uses Discriminant Laplacian Embedding (DLE) [15]. DLE is a method that combines the concept of Laplacian Eigenmaps [14] and a multi-label adaptable variant of Linear Discriminant Analysis (LDA), which constitutes the discriminative part of the system. It has been applied successfully for classification of multi-label data [15], and this is the first time it is applied on the AU detection problem.

The challenge data is composed of three partitions: training, development and test. The test set is the one that is used for ranking and it is not available to the participants. Therefore, in contrast to the development set, the results used for ranking are obtained in a *blind fashion* via sending a running software to the organizers without possibility of parameter tuning etc. We show that our proposed system performs successfully on both of the challenging datasets used (BP4D and SEMAINE) and outperforms the challenge baseline for both the development and test sets. In addition, we present the initial analysis on the effect of the proposed DLE extension on both the development and test partitions. We show improvement for certain AUs, while also observing that the performance is quite data dependent.

In the rest of the paper we first explain the proposed base system in Sec. II, then the proposed extension of multi-label DLE in Sec. III. Sec. IV presents our results on the development and test partitions of the two challenge datasets in comparison with the baseline results as well as between the two proposed methods. In Sec. V we present our



(a) 8 Additional facial landmarks (green) and the landmarks (red) used in their computation on a sample image from the BP4D database.



(b) The full 57 points mask on a sample image from the SEMAINE database.

Fig. 1. Facial landmarks obtained from the face tracker and 8 additional points.

conclusions on the challenge results and effectiveness of the proposed method and list potential improvement methods.

## II. PROPOSED SYSTEM OVERVIEW

Common to any automatic facial analysis system, the initial step for AU detection is to locate the face region and facial landmarks in the images, for which we employ a face tracker based on the supervised descent method (SDM) [16], similarly to the baseline paper [2]. The SDM starts with an initial guess and estimates the shape using a cascade of regression models that are learned at each step using local texture features (e.g. SIFT) extracted from the landmarks estimated in the previous step. The tracker in the end provides 49 landmarks on the face for each frame in the tracked video sequence. Using the locations of these 49 points we estimate the position of eight additional points that contain important local appearance information related to certain AUs. The positions of these points are calculated after an affine warping performed to correct for the in-plane rotation. They mostly mark the non-salient landmarks on the face, which generally appear as result of a muscle contraction, compared to the main set of 49, which locate the salient facial features. The locations of these additional points (*AP*) are illustrated on an example image from the BP4D database in Fig. 1(a) as well as the landmarks that are used in their calculation.

*AP1* is located at the center of the inner most points of the two eyebrows and locates a critical region mostly for *AU4* (brow lowerer) but also for *AU1* (inner brow raise). *AP2* and *AP3* are located around the crow-feet wrinkles close to the two eyes and their positions are calculated using the corners of the eye-brows and the center of the eye. These wrinkles are important indicators of *AU6* and are potentially correlated to *AU7* (eye-lid tightener). *AP4* and *AP5* are located on each side of the nose and mainly

added to include appearance information that occurs during AUs such as *AU10* (lip raiser) or *AU9* (nose wrinkler, not included in this challenge). The positions of *AP6* and *AP7* are calculated as the  $x$  position of the corresponding corner of the mouth and the  $y$  location of the nose tip. These points mark the nasolabial furrows, whose appearance change with action units like *AU6* (cheek raiser) and *AU10*. Finally, *AP8* is the point located on the chin that is obtained such that it is equidistant to the nose tip with *AP1*, assuming a vertical symmetry on the face. This point is mainly important for *AU17* (chin raiser) but also contains information for other AUs that cause a shape and appearance change on the chin. These eight points provide more coverage on the face and thus additional appearance information. Note that the locations of these points are invariant to pose change since they are calculated relatively to the 49 points obtained from the SDM face tracker after correction for the in-plane rotation.

After aligning the face using the eye locations and scaling to a fixed size of 200 by 200 pixels we extract local appearance features around the 57 landmarks in total (Fig. 1(b)) using SIFT [17]. SIFT features have recently been successfully applied on AU detection on the GEMEP-FERA and CK+ databases ([9],[18]). The SIFT descriptors are extracted in the 32 by 32 local neighborhood around each landmark, resulting in a feature vector of size  $128 \times 57 = 7296$ . The dimensionality is reduced using PCA and the number of features to retain for use in the classification is selected separately for each AU such that the F1 measure is maximized on the development sets.

Finally these features are used to train an L1-regularized linear-SVM for each of the 14 action units in the challenge. The training is performed on a custom training set that is the combination of the neutral and peak frames of each sequence in the CK+ database [19], non-speech frames of the training

partition of the GEMEP-FERA database [5] and examples from the SEMAINE [3] and BP4D [4] training partitions down-sampled such that there is a certain minimum number of examples of each AU that is present in each sequence. The resulting combination is a set of 6713 data points and each AU retains a positive/negative sample ratio of at least 10%. The  $C$  parameters of the linear-SVMs are learned through a 5-fold cross-validation within this training set. For the final system submission we also learn a threshold for the distance to the separating hyper-plane in a way that maximizes the F1-score on the development set. This threshold can be an effective biasing parameter between the precision and recall and may depend on factors such as database recording conditions or subjective appearance differences. However, the results reported on the development set still use the default 0 as the decision threshold in order to have a fair comparison.

### III. DISCRIMINANT LAPLACIAN EMBEDDING FOR MULTI-LABEL DATA

The problem of AU detection has rarely been treated as a multi-label problem. In this work we apply the multi-label DLE method proposed in [15] to investigate the advantage of using the mutual information between AUs instead of treating them independently. DLE makes use of the similarities of samples in the training data in terms of both the features and labels, allowing the integration of the correlation between multiple labels.

The method can be summarized as the combination of LDA and Laplacian Embedding [14] in a multi-label setting in order to utilize the locality information of the data in a supervised manner. Given a data matrix  $X$  of  $n$  samples and the corresponding label matrix  $Y$  of size  $n \times P$  ( $P$  is the number of different labels, meaning  $Y$  contains a binary vector for each sample indicating the existence of each AU in our case) the embedding is performed by solving the eigenvalue problem:

$$(A^{-\frac{1}{2}}S_w^{-\frac{1}{2}}S_bS_w^{-\frac{1}{2}}A^{-\frac{1}{2}})U = \Lambda U \quad (1)$$

$\Lambda$  being the set of eigenvalues and  $U$  the combination of eigenvectors, that will be used to project the data matrix  $X$ .  $S_w$  and  $S_b$  are the within-class and between-class scatter matrices, defined similarly to those in standard LDA (explained in detail later on) and  $A = X LX^T$ , with  $L$  being the graph Laplacian [14].  $L = D - W$ ,  $D$  being the diagonal of  $W$  and  $W$  is defined as the "Label Correlation Enhanced Pairwise Similarity" in the work that we have adopted [15] and is formulated as the weighted sum of the *feature similarity matrix*,  $W_x$  and the *label similarity matrix*,  $W_L$ :

$$W = W_x + \mu W_L \quad (2)$$

$W_x$  is the  $n \times n$  pairwise similarity matrix, similar to most embedding algorithms, and is calculated through the Gaussian kernel similarity function (aka. heat kernel). The bandwidth of the kernel function  $\sigma$  is fixed as the average of all absolute pairwise differences in the training set.

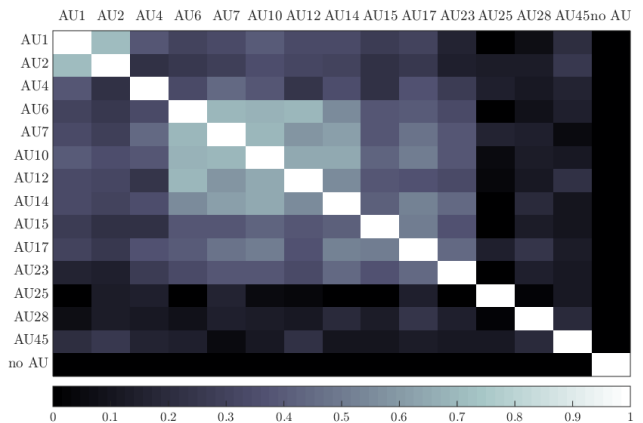


Fig. 2. Action Unit Correlation Matrix on the Training Set

$W_L$ , on the other hand, is calculated using the pairwise similarities between the label vectors  $\mathbf{y}$  of each sample and for two samples  $i$  and  $j$  is formulated as:

$$W_L(i, j) = \frac{\mathbf{y}_i^T C \mathbf{y}_j}{\|\mathbf{y}_i\| \|\mathbf{y}_j\|} \quad (3)$$

$C$  is the  $P \times P$  label-correlation matrix calculated from the training data. Embedding  $C$  in  $W_L$  allows weighting the pairwise label similarities by how correlated two labels are and thus results in placing two samples that have co-existence in highly correlated labels close to each other in the final embedding space. To give an example, if two samples are both labeled as  $AU1 = 1$  and  $AU2 = 1$  these samples will be close to each other in the final space because of the high correlation between  $AU1$  and  $AU2$  (c.f. Fig. 2), whereas they would have been placed further if we had not incorporated this correlation of labels. Fig. 2 shows the correlation between every AU, where the high correlation between certain AUs can be marked with a lighter color. Samples from the BP4D database were excluded in the computation of  $AU25$  and  $AU45$  correlations, since these were not annotated. In addition to the labels defining whether each of the 14 AUs under question within this challenge exist or not, we add a 15th binary label to include the cases where none of these AUs exist. This additional label, of course is not correlated to any of the 14 AUs (Fig. 2).

Finally,  $\mu$  in (2) is the balance parameter between the pairwise feature and label similarity matrices and was optimized on the development partition of the challenge data separately for each AU.

The second component of the DLE is the multi-label LDA. The standard LDA aims to project the data on a lower dimensional space in which the distance between samples with different labels are maximized and samples with the same labels are densely placed close to each other. This is performed by maximizing the between-cluster scatter and minimizing the within-class scatter. In the multi-label case the corresponding matrices are defined as the sum of the

single-label scatter matrices for each type of label  $p$  ([15]):

$$S_b = \sum_{p=1}^P S_b^p, S_b^p = \left( \sum_{i=1}^n Y_{ip} \right) (\mathbf{m}_p - \mathbf{m})(\mathbf{m}_p - \mathbf{m})^T \quad (4)$$

$$S_w = \sum_{p=1}^P S_w^p, S_w^p = \sum_{i=1}^n Y_{ip} (\mathbf{x}_i - \mathbf{m}_p)(\mathbf{x}_i - \mathbf{m}_p)^T \quad (5)$$

where  $\mathbf{m}_p$  is the mean of all samples belonging to the label  $p$  and  $\mathbf{m}$  is the multi-label global mean:

$$\mathbf{m} = \frac{\sum_{p=1}^P \sum_{i=1}^n Y_{ip} \mathbf{x}_i}{\sum_{p=1}^P \sum_{i=1}^n Y_{ip}} \quad (6)$$

These two kinds of projections allow us both to learn a correlation enhanced lower-dimensional manifold and a multi-label discrimination of the data. Projecting the data on the embedding space defined by  $U$ , which is learned by solving (1), we obtain a lower dimensional feature vector as input to the classifier. These features are more discriminative while still containing the correlations between the AUs in addition to the locality properties of the original data. In our tests, the data matrix  $X$  of size  $n$  by  $d$  is obtained by projecting the full data matrix on the PCA space. This initial dimension reduction step enables removing most of the redundancy in the data and thus, allows for a more efficient embedding. The final number of eigenvectors to be used is chosen by optimizing the F1-score, which is a balanced compromise between precision and recall, on the development partitions for both datasets and separately for each AU.

#### IV. RESULTS ON THE FERA-15 CHALLENGE DATA

The FERA 2015 challenge consists of three sub-challenges: AU occurrence detection, AU intensity recognition with prior occurrence knowledge and full AU occurrence and intensity detection. The challenges are on two spontaneous facial expression databases with AU annotations: SEMAINE database [3] and BP4D database [4]. The participants are provided with a training and development set for each database and are asked to run their programs on independent test sets, which they do not see and they have no prior knowledge about, except that they are part of the same database. The test partitions of both datasets contain subjects from the development and training sets as well as unseen subjects. For more details on the databases and partitions the reader is referred to the challenge baseline paper [2].

We have participated in the AU occurrence detection sub-challenge and in this section report our results on the development and test partitions. The fact that these databases are recorded spontaneously makes the task more challenging since the recorded subjects act completely naturally without any instructions about their facial behaviour. This causes more variant AU appearance and occurrences in terms of intensity and combination compared to databases with constraints on the action units or facial expressions performed (e.g. [19], [5], [20]).

#### A. Results on the Development Set

We first test our proposed system and the Discriminant Laplacian Embedding extension on the development set. No image from the development sets of the 2 datasets were included in the training of the classifiers or learning the PCA and embedding bases. However, the development set allows us to obtain the number of features that are optimal for each action unit and also to learn the threshold for the SVM decision value to be used on the test partition. This threshold was kept at 0 in order to obtain a fair comparison with the baseline results and also between the two proposed methods. The DLE was applied to the PCA projected matrix of dimension  $d = 1000$ .

Table I shows the average Overall Accuracy (OA) and F1 measures obtained using the two systems (PCA-SIFT and DLE-SIFT) in comparison with the baseline results. Tables II and III present our results for each AU on the BP4D and SEMAINE development partitions, respectively in terms of OA, Area Under ROC Curve (AUC) and F1. As can be seen from Table I both systems achieve significantly better performance than the baseline systems (with geometric and LGBP-TOP [7] features) on the development set. For a detailed per-AU comparison the reader is referred to the baseline paper [2]. The increased accuracy compared to the baseline shows the efficacy of the chosen features and also the advantage of the enhanced set of facial landmarks. The enhanced set indeed results in an average F1-score increment of 0.7% and 2.6% on the BP4D and SEMAINE development partitions respectively, compared to the 49 point standard set, tested under the same conditions.

The advantage of using a DLE with multi-label information over standard PCA, on the other hand, is not very clear. Although for some AUs the method is more efficient, in average the improvement remains marginal. The difference is clearer when tested on the SEMAINE database, which suggests that the success of the method may depend on the data distribution or the similarity of the distribution of data between the training and test sets. On the SEMAINE database the clearest improvement is on  $AU_{28}$ , while on BP4D it is on  $AU_4$  and  $17$ . The same improvement not appearing on the two databases further suggests that the data distribution is an important factor. More training data will probably provide better variability and thus better generalization of the success of embedding.

#### B. Challenge Results for AU Occurrence Detection on the Unseen Test Set

This section presents the results we have obtained on the test partitions, which constitute the main challenge. Tables IV and V show the F1-scores obtained on the BP4D and SEMAINE test partitions respectively using the two proposed systems and in comparison with the challenge baseline results obtained with the geometric and appearance features. The first observation is that both of the proposed systems clearly outperform the challenge baseline on the test set except for some AUs, namely AUs 2, 12, 14 and 45, for which geometric features are apparently more effective.

TABLE I  
COMPARISON WITH AVERAGE BASELINE RESULTS ON THE DEVELOPMENT PARTITIONS

	SEMAINE		BP4D	
	OA	F1	OA	F1
Baseline Geometric	0.735	0.351	0.712	0.580
Baseline Appearance	0.680	0.298	0.639	0.539
PCA-SIFT (Prop. 1)	0.793	0.417	<b>0.735</b>	0.589
DLE-SIFT (Prop. 2)	<b>0.802</b>	<b>0.435</b>	<b>0.735</b>	<b>0.591</b>

TABLE II  
RESULTS ON THE BP4D DEVELOPMENT PARTITION

AU	PCA-SIFT			DLA-SIFT		
	OA	AUC	F1	OA	AUC	F1
1 (Inner Brow Raiser)	0.717	0.674	0.395	0.694	0.695	<b>0.41</b>
2 (Outer Brow Raiser)	0.669	0.563	<b>0.284</b>	0.664	0.563	0.262
4 (Brow Lowerer)	0.791	0.786	0.472	0.805	0.78	<b>0.509</b>
6 (Cheek Raiser)	0.809	0.888	<b>0.802</b>	0.801	0.873	0.783
7 (Lid Tightener)	0.698	0.765	<b>0.761</b>	0.691	0.756	0.746
10 (Lip raiser)	0.734	0.795	0.781	0.729	0.789	<b>0.789</b>
12 (Lip Corner Puller)	0.859	0.933	<b>0.877</b>	0.839	0.914	0.857
14 (Dimpler)	0.606	0.699	0.611	0.598	0.706	<b>0.616</b>
15 (Lip Corner Depressor)	0.732	0.779	<b>0.447</b>	0.728	0.769	0.43
17 (Chin Raiser)	0.642	0.724	0.573	0.70	0.761	<b>0.604</b>
23 (Lip Tightener)	0.831	0.782	0.486	0.838	0.783	<b>0.487</b>
Average	0.735	0.762	0.589	0.735	0.763	<b>0.591</b>

TABLE III  
RESULTS ON THE SEMAINE DEVELOPMENT PARTITION

AU	PCA-SIFT			DLA-SIFT		
	OA(%)	AUC(%)	F1(%)	OA(%)	AUC(%)	F1(%)
2 (Outer Brow Raiser)	0.804	0.753	<b>0.308</b>	0.822	0.732	0.306
12 (Lip Corner Puller)	0.671	0.677	0.480	0.682	0.731	<b>0.512</b>
17 (Chin Raiser)	0.957	0.889	<b>0.394</b>	0.957	0.886	0.303
25 (Lips Part)	0.757	0.74	0.482	0.725	0.74	<b>0.494</b>
28 (Lip pucker)	0.978	0.906	0.509	0.982	0.947	<b>0.672</b>
45 (Blink)	0.591	0.683	<b>0.329</b>	0.649	0.668	0.324
Average	0.793	0.775	0.417	0.803	0.784	<b>0.435</b>

As weighted average on the two databases, the best F1 score (**0.499**) is obtained with the PCA-SIFT system and is 0.099 higher than the challenge baseline with appearance features (improved by 24.8%) and 0.054 higher than that with geometric features (improved by 12.3%).

TABLE IV  
F1-SCORES ON THE BP4D TEST PARTITION

AU	Our Results		Baseline	
	PCA-SIFT	DLE-SIFT	Geo.	App.
1	<b>0.261</b>	0.226	0.188	0.180
2	0.167	0.149	<b>0.185</b>	0.159
4	<b>0.283</b>	0.233	0.197	0.225
6	<b>0.729</b>	0.697	0.645	0.671
7	0.785	<b>0.802</b>	0.799	0.751
10	<b>0.802</b>	0.742	0.801	0.799
12	0.779	0.784	<b>0.801</b>	0.792
14	0.625	0.599	<b>0.72</b>	0.666
15	<b>0.348</b>	0.223	0.238	0.139
17	<b>0.380</b>	0.325	0.311	0.245
23	<b>0.441</b>	0.424	0.320	0.239
Average	<b>0.508</b>	0.473	0.473	0.442

For BP4D, the accuracies obtained are much lower compared to the development set. This is expected as the few parameters (number of features, decision thresholds and  $\mu$

TABLE V  
F1-SCORES ON THE SEMAINE TEST PARTITION

AU	Our Results		Baseline	
	PCA-SIFT	DLE-SIFT	Geo.	App.
2	0.655	0.663	0.569	<b>0.755</b>
12	<b>0.769</b>	0.759	0.595	0.517
17	0.215	<b>0.255</b>	0.091	0.066
25	<b>0.623</b>	0.613	0.445	0.400
28	0.251	<b>0.262</b>	0.250	0.009
45	0.325	0.347	<b>0.396</b>	0.209
Average	0.481	<b>0.483</b>	0.391	0.326

in 2) that we have were optimized on the development set. However, on the SEMAINE test partition we obtain better results than the development set, which is possibly an indicator that the SEMAINE test and development partitions are more similar to each other compared to BP4D and that our classifiers are able to generalize well enough to this unseen dataset. As explained in the definitions of the sets [2], the test set of BP4D was indeed recorded at a different time, possibly under different conditions.

Our results show that the DLE system achieves a marginal increase in the mean accuracy on the SEMAINE database, compared to the system with standard PCA. Better results

are obtained for AUs 2, 17, 28 and 45. However, this is not the case for the BP4D, with higher F1 measure only for AUs 7 and 12. This probably implies once again that the DLE is highly dependent on the data distribution and that the BP4D development and test partitions contain more variation of AU combinations than that is contained in the training set compared to the SEMAINE database. More tests with more variability and a higher number of training data is needed to reach a conclusion on the benefits of the method, which will be performed next as an extension to this work. Better tuning of the parameters may also greatly increase the accuracy. Another possible cause of the problem is the very low rank of the between-class scatter matrix. Our further work will include using the Laplacian of a dissimilarity matrix instead of the LDA terms in the formulation.

## V. CONCLUSIONS

We have presented an AU detection system that uses SIFT features obtained from an enhanced facial landmarks configuration that includes points around transient facial features and also an extension that uses Multi-label Discriminant Laplacian Embedding with integrated correlation between AUs. The system is applied to the FERA 2015 sub-challenge for spontaneous AU occurrence detection on the SEMAINE and BP4D databases. We obtain a significant increase of accuracy compared to the challenge baseline with both the proposed systems and in the best case obtain 49.85% F1 score as average on the test sets of the two databases.

Comparing the efficiency of the proposed extension of DLE, we observe better results on certain AUs, but only a small increase in the average accuracy is obtained. Using multi-label information in AU detection is a difficult task mainly due to the large number of labels and huge variability in terms of their co-occurrence. In this work, taking into account the different performances on the two test partitions we reach the conclusion that the success of the multi-label DLE might also be data dependent and needs further analysis. The high accuracy of the method on some AUs suggests that the DLE is worthy of more investigation. Future work includes integrating a dissimilarity matrix instead of the LDA related matrices, which are generally of very low rank. Another extension for further improvement would be to include the temporal adjacency of data points in the embedding or classification scheme.

## REFERENCES

- [1] P. Ekman and W.V. Friesen, Facial action coding system: A technique for the measurement of facial movement, 1977.
- [2] M. Valstar, J. Girard, T. Almaev, G. McKeown, M. Mehu, L. Yin, M. Pantic, J. Cohn, FERA 2015 - Second Facial Expression Recognition and Analysis Challenge, Face and Gesture Recognition, Proc. IEEE Int'l Conf on, 2015.
- [3] G. McKeown, M. Valstar, R. Cowie, M. Pantic, and M. Schroder, The semaine database: Annotated multimodal records of emotionally colored conversations between a person and a limited agent, IEEE Transactions on Affective Computing, 3:517, 2012.
- [4] X. Zhang, L. Yin, J. F. Cohn, C. S., M. Reale, A. Horowitz, and J. M. Girard, Bp4d-spontaneous: a high-resolution spontaneous 3d dynamic facial expression database. Image and Vision Computing, 32(10):692706, 2014.
- [5] M. F. Valstar, B. Jiang, M. Mehu, M. Pantic, and K. Scherer, The First Facial Expression Recognition and Analysis Challenge, Automatic Face and Gesture Recognition, Proc. of the IEEE International Conference on, 2011.
- [6] M.F. Valstar, et al., Meta-analysis of the first facial expression recognition challenge. Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on 42.4 (2012): 966-979, 2012.
- [7] T. Almaev, and M. F. Valstar, Local gabor binary patterns from three orthogonal planes for automatic facial expression recognition. Affective Computing and Intelligent Interaction (ACII), 2013 Humaine Association Conference on. IEEE, 2013.
- [8] N. Bayramoglu, Z. Guoying, and M. Pietikainen, CS-3DLBP and geometry based person independent 3D facial action unit detection. Biometrics (ICB), 2013 International Conference on. IEEE, 2013.
- [9] Xiaoyu Ding, Wen-Sheng Chu, De la Torre, F., Cohn, J.F., Qiao Wang, Facial Action Unit Event Detection by Cascade of Tasks. Computer Vision (ICCV), 2013 IEEE International Conference on , 2013.
- [10] Y. Zhu, et al., Dynamic cascades with bidirectional bootstrapping for action unit detection in spontaneous facial behavior. Affective Computing, IEEE Transactions on 2.2 (2011): 79-91, 2011.
- [11] Tong, Yan, Wenhui Liao, and Qiang Ji, Facial action unit recognition by exploiting their dynamic and semantic relationships. Pattern Analysis and Machine Intelligence, IEEE Transactions on 29.10 (2007): 1683-1699, 2007.
- [12] M. H. Mahoor, et al., A framework for automated measurement of the intensity of non-posed facial action units. Computer Vision and Pattern Recognition Workshops, 2009. CVPR Workshops 2009. IEEE Computer Society Conference on. IEEE, 2009.
- [13] C. Shan, G. Shaogang, and P. W. McOwan, A comprehensive empirical study on linear subspace methods for facial expression analysis. Computer Vision and Pattern Recognition Workshop, 2006. CVPRW'06. Conference on. IEEE, 2006.
- [14] M. Belkin, N. Partha, Laplacian eigenmaps for dimensionality reduction and data representation. Neural computation 15.6: 1373-1396, 2003.
- [15] H. Wang, H. Huang and C. Ding, Discriminant Laplacian Embedding, AAAI. 2010.
- [16] X. Xiong and F. De la Torre, Supervised descent method and its applications to face alignment, Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on, pp. 532-539. 2013.
- [17] D. Lowe, Distinctive image features from scale-invariant keypoints, Inter. Journal of Computer Vision 60, 91110, 2004.
- [18] F. Ringeval et al., Prediction of asynchronous dimensional emotion ratings from audiovisual and physiological data, Pattern Recognition Letters (2014).
- [19] P. Lucey et al., The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression, Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on. IEEE, 2010.
- [20] Pantic, Maja, et al., Web-based database for facial expression analysis, Multimedia and Expo, IEEE International Conference on., 2005.