Facial Action Unit Recognition with Sparse Representation

Mohammad H. Mahoor[1], Mu Zhou[1], Kevin L. Veon[1], S. Mohammad Mavadati[1], and Jeffrey F. Cohn[2]

[1]Department of Electrical and Computer Engineering, University of Denver, Denver, CO 80208
[2]Department of Psychology, University of Pittsburgh, Pittsburgh, PA 15260

Emails: mmahoor@du.edu, mu.zhou09fall@gmail.com, kevin.veon@du.edu, seyedmohammad.mavadati@du.edu, and jeffcohn@pitt.edu

***Abstract*- This paper presents a novel framework for recognition of facial action unit (AU) combinations by viewing the classification as a sparse representation problem. Based on this framework, we represent a facial image exhibiting the combination of AUs as a sparse linear combination of basis constituting an overcomplete dictionary. We build an overcomplete dictionary whose main elements are mean Gabor features of AU combinations under examination. The other elements of the dictionary are randomly sampled from a distribution (e.g., Gaussian distribution) that guarantees sparse signal recovery. Afterwards, by solving L_1-norm minimization, a facial image is represented as a sparse vector which is used to distinguish various AU patterns. After calculating the sparse representation, the classification problem is simply viewed as a rank maximal problem. The index of the maximal value of the sparse vector is regarded as the class label of the facial image under test. Extensive experiments on the Cohn-Kanade facial expressions database demonstrate that this sparse learning framework is promising for recognition of AU combinations.**

Keywords- sparse representation; L1-norm minimization; facial expressions recognition; FACS-AU detection.

I. INTRODUCTION

Compressive Sampling (CS) is one of the evolving techniques that can highly impact the frontiers of imaging science. Conventional approaches for image acquisition and reconstruction follow the basic principle of the Nyquist frequency sampling theory. However, the emerging theory of CS says that this conventional wisdom is inaccurate. Based on this new idea, which is supported by a strong mathematical foundation [13, 14], it is possible to represent a signal (e.g., image) accurately and sometimes exactly from a number of samples far smaller than the desired resolution of the image. This sampling paradigm is also called *Sparse Representation* (SR) meaning that one can represent a signal using linear combination of only few nonzero coefficients taken nonadaptively from coefficients describing an image. In fact, SR theory states that an image can be represented by a sparse linear combination of some redundant basis, which constitute an overcomplete dictionary.

Lately, the theory and applications of sparse representation attracted much attention in computer vision and pattern recognition literature [15-18, 21, 32]. In particular, sparse representation has been applied to face recognition [15], facial expression recognition [32], object recognition [16], image denoising [17], super resolution image processing [18], and object tracking [21].

Particularly related to this paper, Wright et al. [15] casted the problem of face recognition as a linear regression model using the new theory of sparse representation. They represented a facial image as sparse combination of multiple given facial images with known identities. They claimed that the face recognition problem under the effect of occlusion and image corruption can be addressed using the fact that these errors are often sparse with respect to the standard pixel basis. Extensive experiments on publically available databases such as Yale database showed that this approach is not sensitive to the choice of features and outperforms support vector machine (SVM), nearest neighbor (NN), and nearest subspace for face recognition.

In [32], Ying et al. utilized sparse representation for facial expressions recognition. They designed two classifiers in the sparse domain using two different sets of image features: raw gray scale pixel values and local binary patterns. The final expression recognition was then performed by fusing the results of the two classifiers. Although they used sparse representation for image representation, they did not justify (neither theoretically nor experimentally) the design of their overcomplete dictionary and why L_1-norm minimization should work for their problem. One important step in CS theory is to design a dictionary that guarantees perfect or near perfect signal reconstruction via L_1-norm minimization.

In this paper we turn our attention to the problem of facial expressions recognition based on sparse representation and in particular recognition of AU combinations described by the Facial Action Coding System (FACS). Among many pattern recognition problems, automatic facial expression recognition has remained an active research area in computer vision and pattern recognition [1-3]. The problem is challenging due to the complexity and variety of facial expressions. In fact, human facial expressions reveal subtle facial muscle movements. FACS, invented by Ekman [11], is a standard technique for describing facial activities and expressions. FACS provides a description of all possible visually detectable facial variations in terms of 44 AUs.

The existing computer vision techniques for automatic AU recognition focus primarily on single AU recognition. These approaches can be divided into static approaches [6-7, 10, 22, 26, 27] and temporal approaches [12, 31]. Under the static approaches, Bartlett et al. [6] employed SVM and AdaBoost to realize spontaneous AU detection. As an example of temporal approaches, Tong et al. [12] proposed a unified probabilistic framework based on the dynamic Bayesian network (DBN) to represent single facial motions. Recently, Simon et al. [29] developed a segment-based approach called kSeg-SVM that incorporates benefits of both approaches and avoids their limitations. We refer our reader to [2] for detailed discussions on recent approaches for AU recognition.

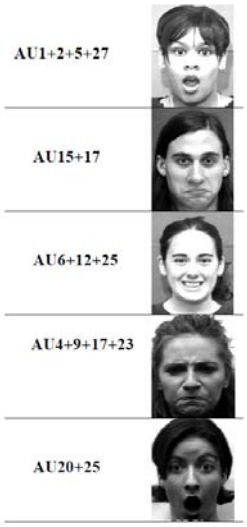


Figure 1. AU combinations under examination.

To the best of our knowledge, less attention has been given to the problem of recognizing AU combinations which occur frequently in real life. Thus, we focus our attention on recognizing AU combinations in facial images. A closer look at the Cohn-Kanade facial expressions database [24] reveals that five combinations of AUs, AU1+2+5+27, AU15+17, AU6+12+25, AU4+9+17+23, and AU20+25 are exhibited more often by the subjects enrolled in the Cohn-Kanade database. Figure 1 shows an example of these combinations of AUs. These combinations consist of the most significant AU combinations that represent basic expressions. Although single AU detection is possible, only certain combinations are representative of human emotion. Hence, detecting single AUs may not be directly relevant to detecting human emotions.

In this paper, we present a novel sparse learning approach for recognition of combined facial action units. We propose to build an overcomplete dictionary whose elements consist of mean Gabor features extracted for each class of AU combinations augmented with a matrix of randomly chosen elements (i.e., Gaussian distribution or $\{-\sqrt{3},0,+\sqrt{3}\}$ numbers randomly selected with probabilities 1/6, 2/3, and 1/6, respectively [13, 14]). In order to represent variations in facial appearances caused by facial expressions, Gabor filters at different orientations and scales are applied at the position of facial labels in images extracted by active appearance model (AAM) [20] training. Afterwards, a set of Gabor features describing a facial image with an unknown combination of AUs is represented as a linear combination of dictionary entries. The corresponding sparse coefficients can be calculated by L_1-norm minimization, which guarantees the perfect or near-perfect reconstruction of the sparse representation of the interest signal. This sparse signal will have a maximal response to one of the trained columns in the dictionary. Classification is performed by selecting the largest response of the trained classes after sparse reconstruction.

The sparse representation is usually learned by solving an L_1-norm minimization process [13, 14]. Mathematically speaking, if we have an overdetermined system of linear equations (the number of observations/equations is more than the number of unknown variables), $Y = Ax$, where size of A is $m \times n$ and $m > n$, then such a problem can be solved using the least square error technique (also known as L_2-norm minimization). However, in sparse representation the number of observations is less than the number of unknowns ($m < n$). In this situation, we deal with an underdetermined system of linear equations. Although L_2-norm minimization will result in a solution to this system, the resulting signal will not be sparse as desired. If we assume that the solution is sparse, L_1-norm minimization can find the exact solution for this problem if the overcomplete dictionary matrix, A, satisfies the Restricted Isometry Property (RIP) [13] that is discussed later in this paper.

To summarize, the contributions of this paper are as follows:

1. The problem of feature classification is casted as a sparse linear combination of basis constituting an overcomplete dictionary, which is solved using L_1-norm minimization.
2. The proposed sparse learning framework is used for recognition of AU combinations to tackle the challenging problem of facial expressions recognition.

The remainder of this paper is organized as follows. We present our framework on sparse feature representation in Section II followed by a theoretical discussion for L_1-norm minimization and the structure of the dictionary in Section III. We give and discuss the experimental results using the Cohn-Kanade facial expressions database in Section IV. We finally conclude the paper in Section V.

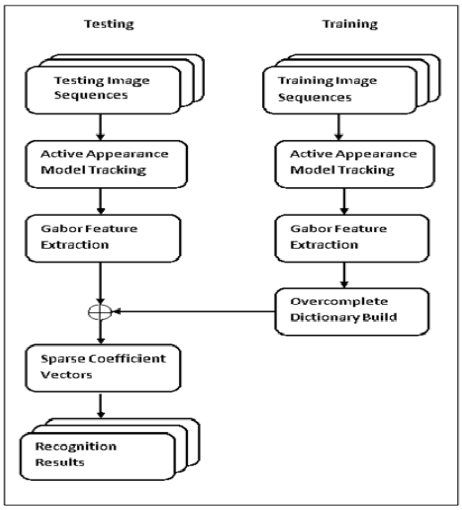


Figure 2. The framework of proposed approach for sparse representation and recognition of AU combinations.

II. SPARSE REPRESENTATION OF FACIAL FEATURES

In this section we describe our framework for recognition of AU combinations in facial images. The structure of the proposed framework is illustrated in Figure 2, which is separated into two phases: training phase and testing phase.

A. Training phase

First, every facial image is represented using Gabor filters. Gabor filters are shown to be a good choice for facial image representation and feature extraction [4, 6, 22 28]. We apply Gabor filters in six orientations and five wavelengths at the location of facial landmark points extracted using AAM. A facial image is represented by a feature vector containing the Gabor features extracted at these locations. Afterwards, an overcomplete dictionary is constructed using the mean Gabor features for each class augmented with randomly selected numbers.

Each AU combination is treated as a class represented by a template feature vector constructed through averaging the extracted Gabor features of the facial images with the same class label. For instance, if we are examining c classes of AU combinations then c number of mean Gabor features, $\overline{g} = \{\overline{g}_1, \overline{g}_2, ..., \overline{g}_c\}$, are generated from training images and used to generate the c first columns of the dictionary. The residual columns of the dictionary, $W = \{w_1, w_2, w_3 \ldots w_{n\text{-}c}\}$ are either randomly selected from the i.i.d. Gaussian distribution, $N(0, 1/m)$, or randomly selected from $\{-\sqrt{3}, 0, +\sqrt{3}\}$ with probabilities, 1/6, 2/3, and 1/6, respectively [13,14]. Figure 3 illustrates the structure of the dictionary used in this work for recognition of five AU combinations. The dictionary, $A = [\overline{g}_1, \ldots, \overline{g}_c, w_1, \ldots w_{n\text{-}c}]$, constructed in the training phase, is used for facial expressions classification in the testing phase.

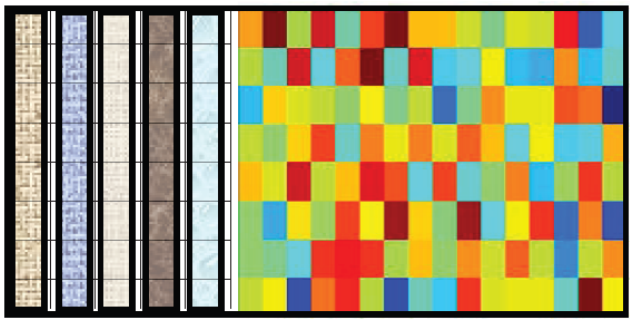


Figure 3. Structure of the Dictionary. The first five columns represents the mean Gabor feature vectors associated to five class of AU combinations under examination and the random color pattern represents the randomly selected coefficients from i.i.d. Gaussian distribution or from $\{-\sqrt{3}, 0, +\sqrt{3}\}$ numbers.

B. Testing phase

For a given facial image with unknown facial expressions, we again represent the facial image using Gabor filters applied at the location of the facial landmark points extracted using the AAM. We then state that this feature vector, y, can be linearly and sparsely represented in terms of the elements of the overcomplete dictionary A: $y = Ax$, where the size of A is $m \times n$ and $m \ll n$. The solution to this underdetermined system of linear equations is obtained using the L_1-norm minimization.

After calculating the sparse vector x, the classification problem is simply viewed as a rank maximal problem. The index of the maximal value of vector x is regarded as class label and the main contributor to this linear representation framework.

Figure 4 summarizes our proposed algorithm for building the overcomplete dictionary and then using it for facial AU combinations recognition using sparse representation. In this paper, we consider primarily the classification of five AU combinations, AU1+2+5+27, AU15+17, AU6+12+25, AU4+9+17+23, and AU20+25, which frequently occur in human facial expressions (see Figure 1).

In the following section we describe how the underdetermined system of linear equations is solved using L_1-norm minimization.

Algorithm: AU Recognition by Sparse Learning

1. **Input:** Mean Gabor feature $\{\bar{g}_1, \bar{g}_2, \bar{g}_3, \bar{g}_4, \bar{g}_5\}$ for five AUs classes, and an unknown input facial Gabor feature y.
2. Dictionary construction:
$A = [\bar{g}_1, \bar{g}_2, \bar{g}_3, \bar{g}_4, \bar{g}_5, w_1, w_2, w_3 \ldots w_{n\text{-}5}]$
where w_i are random vectors generated using Gaussian distribution or numbers randomly selected from $\{-\sqrt{3}, 0, +\sqrt{3}\}$.
3. Solve the L_1-norm minimization problem:
$\hat{x} = \arg\min \|x\|_1$ s.t. $y = Ax$
4. Obtain the associated sparse coefficients:
$\hat{x} = [\hat{x}_1, \hat{x}_2, \hat{x}_3, \hat{x}_4, \hat{x}_5, t_1, t_2, \ldots t_{n\text{-}5}]$
5. **Output**: AUs class assignment by:
$\hat{C} = \arg\max_i [\hat{x}_1, \hat{x}_2, \hat{x}_3, \hat{x}_4, \hat{x}_5]$

Figure 4: Our algorithm for AU recognition using sparse-based learning on bases of an overcomplete dictionary.

III. SPARSE LEARNING BY L_1-NORM MINIMIZATION

A. L0/L1-norm Minimization Equivalence

Let us assume that $A \in R^{m \times n}$ is an overcomplete dictionary of n prototype elements, where $m \ll n$, and argue that a signal $y \in R^m$ can be represented as a sparse linear combination of these dictionary elements. The calculation of this sparse representation can be mathematically described as follows:

$$\min \|x\|_0 \text{ subject to } y = Ax \qquad (1)$$

where $\|.\|_0$ denotes the L_0-norm, which simply counts the number of non-zero entries in vector x. However, solving this L_0-norm minimization is known to be an NP-hard problem, which requires computational search of all subsets of non-zero coefficients [13]. CS theory proposes an alternative way to handle this issue by converting the L_0-norm minimization into L_1-norm minimization. CS theory guarantees that the L_0-norm minimization and the L_1-norm minimization have exactly the same solution if the true solution x is sufficiently sparse and the matrix A satisfies the RIP. If so, the solution x can be efficiently recovered by a tractable L_1-norm minimization:

$$\min \|x\|_1 \text{ subject to } y = Ax \qquad (2)$$

This is a convex optimization problem that conveniently can be solved using linear programming techniques such as basis pursuit [21]. In this paper, we utilize the solution of Equation (2) as a sparse representation of facial features and then for classification of AU combinations.

Sometimes, different types of transforms such as Fourier transform, wavelet, curvelet, and/or ridgelet [19] are utilized to build the basis of the overcomplete dictionary. However, in this paper we utilize random numbers for generating the overcomplete dictionary as discussed in the following.

B. A Discussion on the Dictionary

An overcomplete matrix $M \in R^{m \times n}$ satisfies the Restricted Isometry Property with respect to an isometry constant $\delta_k > 0$ over all k-sparse vectors. For any k-sparse vector $x \in R^n$, the following property must hold:

$$(1 - \delta_k) \| x \|_{L_2}^2 \leq \| Mx \|_{L_2}^2 \leq (1 + \delta_k) \| x \|_{L_2}^2 \qquad (3)$$

This property essentially requires that every set of columns with cardinality less than k approximately behave like an orthonormal system. An important result is that if the columns of the sensing matrix M are approximately orthogonal, then the exact recovery phenomenon occurs [13, 14]. Interestingly, if matrix M is formed by one of the following operations:

i) *random sampling of i.i.d entries from a Gaussian distribution N(0,1/m).*
ii) *random sampling of i.i.d. entries from* $\{-\sqrt{3}, 0, +\sqrt{3}\}$ *numbers randomly selected with probabilities, 1/6, 2/3, and 1/6, respectively.*

then matrix M satisfies the RIP with an overwhelming probability, $1 - O(e^{-\gamma m})$, if

$$k \leq C.m / \log(\frac{n}{m}) \qquad (4)$$

Both C and γ are very small positive constants. This provides a hint on what would be the appropriate size of the overcomplete dictionary generated using random numbers. In this paper, we study and compare both aforementioned random sampling distributions for generating the overcomplete dictionary and then classification of AU combination. The dictionary utilized in this work is composed of the matrix of mean Gabor features, $\bar{g}$, and the random matrix M that satisfies RIP : $A = [\bar{g}, M]$. In this paper, we built the dictionary **A** and empirically showed that the augmented matrix still satisfies the RIP. For more detail see Section IV.

IV. EXPERIMENTS AND RESULTS

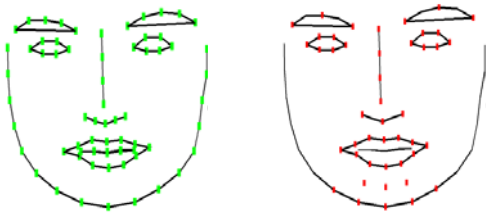
A. Facial Expressions Database

Our proposed framework was evaluated on the Cohn-Kanade facial expressions database [24]. The database contains facial image sequences from 97 subjects between the ages of 18 and 50 years exhibiting single AUs and combinations of AUs. The image sequences begin with a neutral face and end with maximum intensity facial expression. The images are gray-scale with a resolution of 640×480 pixels. For our experiments, we choose the last 3 or 4 frames with maximal AU intensity from 84 subjects (406

images) which exhibit the five AU combinations.

B. Facial Feature Extraction

We used AAM for extracting facial landmark points and facial image registration. Sixty eight facial landmark points were extracted using AAMs that were specifically trained for each subject on about 3% to 5% of image frames [20]. Since 68 facial features are very dense and some of the facial points are very close to each other compared to the size of Gabor filter masks (the smallest Gabor filter used in this work has a size of 5×5 pixels), we selected a subset of 42 points out of the 68 AAM points. We added three points located in the chin area, resulting in a total of 45 labels used to apply the Gabor filters. These extra points are calculated as the mid points between each pair of the lowest points in the lip and the face boundary. These three points are particularly helpful for describing AU17 using Gabor filters. Figure 5 illustrates the 68 facial landmark points and the 45 selected facial points.



(a) (b)

Figure 5. (a) 68 AAM facial landmarks (b) 45 selected feature points.

The Gabor filters are selected at five scales, $\{1, \sqrt{6}/2, 3/2, 3\sqrt{6}/4, 3\}$ and six orientations, {0, π/6, 2π/6, 3π/6, 4π/6, 5π/6}. Therefore, with five wavelengths, six orientations, and 45 landmarks, we have a total m=5×6×45=1350 Gabor features for each image.

Prior to extracting Gabor features, the raw 2-D facial images are processed to normalize the image variations due to the effect of lighting, scale, and head pose. For lighting normalization, first the contrast of the images is normalized using histogram equalization. Then the intensity values of each image are normalized to have zero mean and unit variance. For pose and scale normalization, eye coordinates are used to align the faces such that the coordinates of the two centers of the eyes in each individual image are registered to fixed locations.

C. Results

The experiment is divided into to two phases. The training phase contains computation of mean Gabor features $\bar{\mathrm{g}} = \{\bar{g}_1, \bar{g}_2, \bar{g}_3, \bar{g}_4, \bar{g}_5\}$ of five classes of AU combinations. We randomly select 30 images from 10 subjects for each class as training data to generate the mean Gabor feature corresponding to AU1+2+5+27, AU15+17, AU6+12+25, AU4+9+17+23, and AU20+25. These five mean Gabor feature vectors are utilized to build the dictionary matrix A whose other columns are generated either by selecting randomly from a Gaussian distribution $N(0,1/m)$ or from numbers $\{-\sqrt{3},0,+\sqrt{3}\}$ as discussed in Section III.

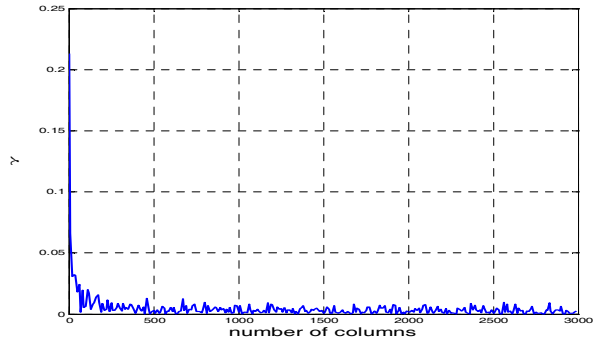


Figure 6. Parameter γ versus the number of columns of the dictionary A.

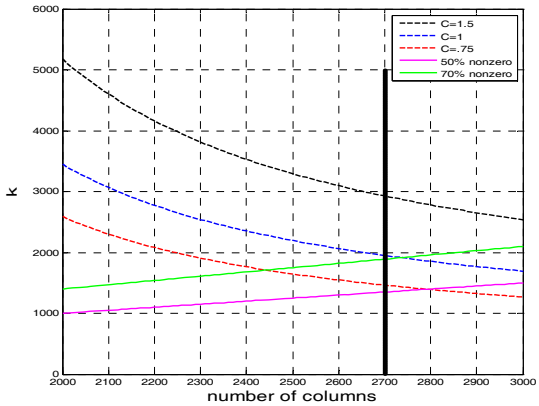


Figure 7. This figure studies the condition in Equation (4) empirically. We calculated the two sides of this equation for different values of C and k as the number of columns, n, increases. As this figure shows, if n = 2705, we ensure that the condition in equation (4) is satisfied.

We ran some experiments to study the effect of the number of randomly generated columns of matrix A on the satisfaction of RIP. Our experiments show that as long as the number of classes, c, satisfies $c \ll n$, matrix A still satisfies RIP with an overwhelming probability.

We calculated parameter γ for different numbers of columns of matrix A. As Figure 6 shows, γ rapidly approaches zero as n increases (c=5 is fixed). Since m=1350 < n, then γ is guaranteed to be close to zero. This means that matrix A will satisfy the RIP Equation (4). Equation (4) basically relates the number of nonzero coefficients, k, with the number of measurements, m, and the number of columns of matrix A. We empirically calculated the right side of Equation (4), which is the upper bound on k, for different values of C to choose the optimum value for n. As Figure 7 shows, if we assume that C=1 (dotted blue curve) and only 50% of the coefficients are nonzero (solid magenta line), then Equation (4) is always satisfied.

Based on this experiment, we chose the number of columns in matrix A to be n = 5+2×1350 = 2705. This means that

matrix A has a size of 1350×2705, which is an overcomplete dictionary and satisfies Equation (4). Even if 70% of the coefficients are nonzero (solid green line), Equation (4) is still satisfied for the chosen $n = 2705$.

Figure 8 shows a sample sparse representation, x, for an image in class AU15+17 (class 2). As this figure shows, the vector x is sparse meaning that the coefficients are mostly zero or close to zero (less than 50% of the coefficients are nonzero). Among the nonzero coefficients, one coefficient has a larger value compared to others. This large coefficient corresponds to the 2nd column of the dictionary representing the weight of class two (i.e., the mean Gabor feature vector of AU15+17). As we discussed in Section III, every facial image under test is assigned to the class corresponding to the index of its maximal value of its first five elements of sparse vector. Our experiments also show that the maximal coefficient always occurs among the first $c=5$ coefficients of vector x.

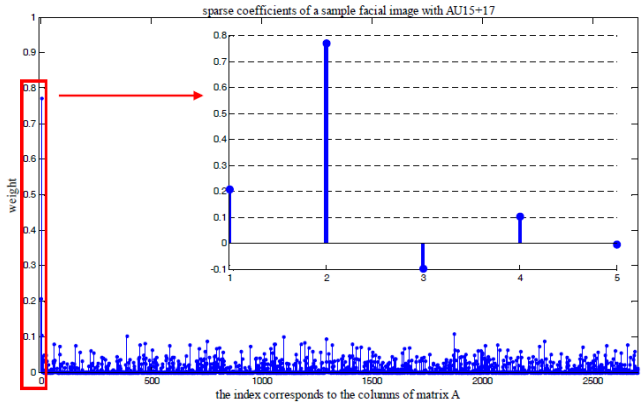


Figure 8. Sparse coefficients of a facial image containing AU15+17. Less than 50% of the coefficients are nonzero.

The remaining images (256 images) of other subjects (total of 74 subjects) were utilized for testing our developed AU classification approach. We extracted the Gabor feature vector of each test image and solved the L_1-norm optimization defined by Eq. (2) using the L1-magic package [30]. We use the standard basis pursuit problem solver using a primal-dual algorithm described in L1-magic user manual [30].

Table 1 shows the confusion matrices obtained as a result of recognition based on our approach. As the table shows, despite of some overlaps between classes AU15+17 and AU4+9+17+23, and classes AU6+12+25 and AU20+25, our method can distinguish between these AUs combinations significantly. Comparing Table 1a with Table 1b shows that the overcomplete dictionary constructed from random selection of $\{-\sqrt{3},0,+\sqrt{3}\}$ outperforms the overcomplete dictionary constructed from i.i.d. Gaussian distribution. The recognition rate of the first three classes is the same for both distributions.

Table 1. Confusion Matrix for AU combinations classification using sparse learning (a) the random pattern in the dictionary generated is generated by random selection of numbers from $\{-\sqrt{3},0,+\sqrt{3}\}$ and (b) generated using the Gaussian distribution.

(a)

AUs Combination	1+2+5+27	15+17	6+12+25	4+9+17+23	20+25
1+2+5+27	**79**	0	0	0	0
15+17	0	**36**	0	2	0
6+12+25	0	0	**45**	0	0
4+9+17+23	0	3	3	**48**	6
20+25	2	0	0	0	**32**
Total images	**81**	**39**	**48**	**50**	**38**

(b)

AUs Combination	1+2+5+27	15+17	6+12+25	4+9+17+23	20+25
1+2+5+27	**79**	0	0	0	0
15+17	0	**36**	0	4	2
6+12+25	0	0	**45**	0	2
4+9+17+23	0	3	3	**46**	7
20+25	2	0	0	0	**27**
Total images	**81**	**39**	**48**	**50**	**38**

We compared the performance of our sparse learning approach with SVM-classification and nearest neighbor classification. For both our sparse learning approach and nearest neighbor classification we used mean Gabor features as training samples. We employed libSVM [31] for C-SVM classification and used all training samples instead of the mean Gabor features. Table 2 illustrates the recognition rate of our approach and compares the results with the C-SVM and NN classifiers.

Table 2. Comparison between SVM, NN and sparse representation for recognition of AU combinations.

Method	**Sparse-Representation**		**SVM**	**NN**
AU Com.	Gaussian Dist.	$-\sqrt{3},0,+\sqrt{3}$ rand numbers		
1+2+5+27	97.5%	97.5%	98.7%	97.5%
15+17	92.3%	92.3%	84.6%	74.35%
6+12+25	93.7%	93.7%	93.7%	93.7%
4+9+17+23	96.0%	92.0%	90.0%	96.0%
20+25	84.2%	71.0%	39.5%	55.5%
Overall rate	**93.8%**	**91.0%**	**85.04%**	**86.7%**

Table 2 shows that our approach exhibits comparable performance as the other two classification methods for recognition of three classes of AU combinations (AU1+2+5+27, AU6+12+25, and AU4+9+17+23), but performs significantly better than them for classes AU15+17 and AU20+25.

Table 2 also shows that our approach outperforms both the C-SVM classification and nearest neighbor classification approaches overall. The overall recognition rate of C-SVM is 85.04%, which was achieved by utilizing a polynomial kernel. Nearest neighbor proves to be slightly better with an overall recognition rate of 86.7%. Both approaches are surpassed by our framework, which obtains an overall recognition rate of 93.8%. Except for AU1+2+5+27, where C-SVM is slightly better than our approach, the recognition rate of our approach is equal or significantly better than the C-SVM classification technique.

V. CONCLUSION AND FUTURE WORK

This paper presented a novel sparse learning approach for AU combination classification. We represented facial images by extracting Gabor features at the location of facial landmark points extracted using AAM. Then, we developed an overcomplete dictionary to efficiently and stably recognize the combination of facial AUs using L_1-norm minimization. The experiments on the Cohn-Kanade database showed promising results compared to the C-SVM and NN techniques for classification of five combinations of AUs. In the future, we plan to extend this approach to more AU combinations to establish a systematic framework for spontaneous facial expression classification. Also, it will be interesting to evaluate more general video based object recognition problems.

ACKNOWLEDGEMENT

This research was funded by the grant IIS-0957983 from the National Science Foundation.

REFERENCES

[1] Y. Tian,T. Kanade and J.F Cohn. Facial expression analysis.Handbook of face recognition, 247–276. Springer, New York , 2005.

[2] Z. Zeng, M. Pantic, G. Roisman, and T. S. Huang, T.S. A Survey of Affect Recognition Methods: Audio, Visual, and ICMI2007, 126-133, 2007.

[3] B. Fasel and J. Luettin, .Automatic facial expression analysis: a survey. IEEE Journal of Pattern Recognition,259-275, 2003.

[4] M. Lyons and S. Akamatsu. Coding Facial Expressions with Gabor Wavelets. IEEE.International Conference on Automatic Face and Gesture Recognition, 200-205,April, 1998.

[5] T. Ojala and M. Pietikainen. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns.IEEE Transactions on Pattern Analysis and Machine Intelligence, 2002.

[6] Bartlett, M.S., Littlewort, G., Frank, M.G., Lainscsek, C., Fasel, I., and Movellan, J.Recognizing Facial Expression: Machine Learning and Application to Spontaneous Behavior. IEEE International Conference on Computer Vision and Pattern Recognition. 568-573,2005.

[7] M. H. Mahoor, S. Cadavid, D. S. Messinger, and J. F. Cohn, "A Framework for Automated Measurement of the Intensity of Non-Posed Facial Action Units", 2nd IEEE Workshop on CVPR for Human communicative Behavior analysis (CVPR4HB),Miami Beach, June 25, 2009.

[8] Y. Chang, C. Hu, and M. Turk, Probabilistic expression analysis on manifolds, International Conference on Computer Vision and Pattern Recognition, Washington DC, June 2004.

[9] Peng Yang, Qingshan Liu, and Dimitris. N. Metaxas, Boosting Coded Dynamic Features for Facial Action Units and Facial Expression Recognition, IEEE International Conference on Computer Vision and Pattern Recognition, 2007.

[10] M. Pantic and L. Rothkrantz. Facial action recognition for facial expression analysis from static face images. IEEE Transactions on Systems, Man, and Cybernetics, pages 1449– 461, 2004.

[11] P. Ekman, W. Friesen, Facial Action Coding System: Manual, Consulting Psychologist Press, Palo Alto, 1978.

[12] Yan Tong, Jixu Chen and Qiang Ji, A Unified Probabilistic Framework for Spontaneous Facial Action Modeling and Understanding, IEEE Transactions on Pattern Analysis and Machine Intelligence, 258-274, Vol. 32, No. 2, February, 2010.

[13] E. Candes, J. Romberg, and T. Tao. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. IEEE Transaction on Information Theory,52(2), 2006.

[14] D. Donoho. Compressed sensing. IEEE Transaction on Information Theory, 52(4), 2006.

[15] John Wright, Allen Yang, Arvind Ganesh, Shankar Sastry, and Yi Ma. Robust Face Recognition via Sparse Representation.IEEE Transaction on Pattern Analysis and Machine Intelligence, vol. 31, no. 2, February 2009.

[16] J. Marial, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman. Discriminative learned dictionaries for local image analysis. IEEE International Conference on Computer Vision and Pattern Recognition ,2008.

[17] M. Elad and M. Aharon. Image denoising via sparse and redundant representations over learned dictionaries. IEEE Transaction Image Processing, 54(12),2006.

[18] Jianchao Yang, John Wright, Yi Ma, Thomas Huang. Image Super-Resolution as sparse representation of Raw Image Patches, IEEE International Conference on Computer Vision and Pattern Recognition, 2008.

[19] S. Mallat. A wavelet tour of signal processing, second edition. Academic Press, New York, 1999.

[20] Jing Xiao, Simon Baker, Iain Matthews, and Takeo Kanade, "Real-Time Combined 2D+3D Active Appearance Models," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, June, 2004, pp. 535 - 542.

[21] X. Mei and H. Ling, Robust Visual Tracking using L1 Minimization, Proceeding of ICCV 2009.

[22] Ying-li Tian , Takeo Kanade , Jeffrey F. Cohn. Evaluation of Gabor-Wavelet-Based Facial Action Unit Recognition in Image Sequences of Increasing Complexity. IEEE International Conference on Automatic Face and Gesture Recognition, 229-234,2002.

[23] M. Aharon, M. Elad, and A. Bruckstein. K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation.IEEE Transactions on Signal Processing, Vol.54, No. 11, Novermber 2006

[24] T. Kanade, J. Cohn, and Y. Tian. Comprehensive database for facial expression analysis. In International Conference on Face and Gesture Recognition, 46–53, March, 2000.

[25] C.-C. Chang and C.-J. Lin. Libsvma library for support vector machines. http://www.csie.ntu.edu.tw/ cjlin/libsvm/.

[26] Y. Tian, J. F. Cohn, and T. Kanade. Facial expression analysis. In S. Z. Li and A. K. Jain, editors, Handbook of face recognition. New York, New York: Springer, 2005.

[27] S. Lucey, I. Matthews, C. Hu, Z. Ambadar, F. De la Torre, and J. Cohn. AAM derived face representations for robust facial action recognition. In International Conference on Automatic Face and Gesture Recognition, 2006.

[28] M. H. Mahoor and M. Abdel-Mottaleb, A Multi-modal Approach for Face Recognition Based on Ridge Images and Attributed Relational Graph, IEEE Transactions on Information Forensics and Security, Volume 3, Issue 3, pp.431- 440, Sept. 2008.

[29] T. Simon, F De La Torre, J. F. Cohn, Action Unit Detection with Segment-based SVMs IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2010.

[30] http://www.acm.caltech.edu/l1magic/

[31] L. Shang and K. Chan. Nonparametric discriminant HMM and application to facial expression recognition. In Conference on Computer Vision and Pattern Recognition, 2009.

[32] Z. Ying, Z. Wang, and M. W. Huang, Facial Expression Recognition Based on Fusion of Sparse Representation, Advanced Intelligent Computing Theories and Applications. With Aspects of Artificial Intelligence, Lecture Notes in Computer Science, 2010, Volume 6216/2010, pp. 457-464.