DOI: 10.2352/ISSN.2470-1173.2016.11.IMAWM-455

Joint and Discriminative Dictionary Learning for Facial Expression Recognition

Sriram Kumar, Behnaz Ghoraani, Andreas Savakis

Abstract

Dictionary Learning and sparse coding methods have been widely used in computer vision with applications to face and object recognition. A common challenge when performing expression recognition is that face similarities may confound the expression recognition process. An approach to deal with this problem is to learn expression specific dictionaries, so that each atom corresponds to one expression class. However, even when employing expression specific dictionaries, it is likely that two atoms from two sub-dictionaries share common characteristics due to facial similarities. In this paper, we consider a joint dictionary that captures common facial attributes, and class-specific dictionaries that are used to classify different expressions. We investigate three dictionary learning methods for sparse representation classification: one that learns a global dictionary based on K-SVD, one that learns expression specific dictionaries based on Fisher Discrimination Dictionary Learning (FDDL), and one that learns a shared as well as expression specific dictionaries based on Dictionary Learning Separating Commonality and Particularity (DL-COPAR). We demonstrate the effectiveness of the shared dictionary learning approach on the extended Cohn-Kanade database where DL-COPAR outperforms FDDL and K-SVD by a significant margin.

Introduction

Facial expression recognition has many applications such as human-computer interaction, driver monitoring, health and wellness, entertainment, surveillance and others. Recognizing facial expressions is a challenging task, and sometimes similarities in facial appearance may interfere with the recognition of facial expressions. In this paper, we propose a sparse representation classification approach with joint and discriminative dictionary learning in order to overcome the difficulty of confounding face expression and identity.

The pioneering work of Ekman et al. [1], identified six universal expressions shown in Figure 1, and introduced a method to quantify facial actions and expressions based on action units. The Facial Action Coding System (FACS) was proposed to quantify facial actions based on muscle movements, so that each expression can be represented as a combination of action units. Numerous facial expression recognition methods have been presented in the literature [2, 3, 4]. These methods can be broadly categorized into geometric and appearance based. Common geometric methods include Active Shape Model (ASM) or Active Appearance Model (AAM) [5]. Appearance based methods work with local or holistic facial appearance. They often compute intermediate representations of images using features such as Gabor wavelets [6] and Local Binary Patterns (LBP) [7]. Gabor wavelets generate features that correspond to edges at various frequencies and orientations inspired from the human visual system. LBP features capture texture variations and are capable of handling severe changes in illumination.

Most of the expression recognition pipelines begin high dimensional representations of facial features, and use dimensionality reduction techniques such as Principal Component Analysis (PCA) and manifold learning. Dimensionality reduction benefits the classification process by reducing the data size and organizing the data in a space that improves classification accuracy. Manifold learning techniques have been utilized for expression recognition [8] among other facial analysis tasks. Sparse Representation (SR) classification techniques have demonstrated good performance in face recognition [9] and expression recognition [10], [11]. Manifold based Sparse Representation (MSR) [11] combines manifold learning and sparse representations to tackle the problem of coefficient contamination due to facial identity in expression recognition.

Recent developments in dictionary learning methods have shown that learning a dictionary from data is beneficial because it produces better and more efficient representations [12, 13]. In [14], discriminative dictionary learning is proposed by using the class label information. In [15], the authors proposed a discriminative approach that exploits the coherence between atoms in the dictionary to learn a shared/common dictionary and class-specific dictionaries. Other dictionary learning methods include [16-18].

A joint/common dictionary is considered to address the issue of similarities in elements across dictionaries. The proposed approach is effective for dealing with a common problem in expression recognition where the learned system classifies faces that are similar in appearance rather than classifying the expression. In this context, learning an expression specific and a shared dictionary plays an important role in classifying expression with high accuracy. By detecting shared features, we learn sub-dictionaries whose atoms are not correlated with other dictionaries.

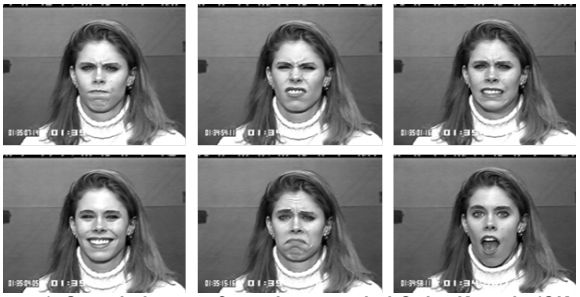


Figure 1. Sample images from the extended Cohn-Kanade (CK+) facial expression dataset illustrating (top to bottom, left to right) anger, disgust, fear, happy, sad, surprise.

In the next sections, dictionary learning is overviewed for sparse representation classification using K-SVD [12], Fisher Discrimination Dictionary Learning (FDDL) [14], and Dictionary Learning Separating Commonality and Particularity (DL-COPAR) [15]. Results are reported on the extended Cohn-Kanade (CK+) database.

DOI: 10.2352/ISSN.2470-1173.2016.11.IMAWM-455

Joint Dictionary Learning

Sparse Representation

In a sparse representation framework, a sample $\boldsymbol{y}$ in $\mathbb{R}^d$ space is represented on a dictionary of samples $\mathbf{X} \in \mathbb{R}^{d \times p}$ via the sparse coefficients $\boldsymbol{a}$, as follows:

$$\boldsymbol{y} = \mathbf{X}\boldsymbol{a} \tag{1}$$

The coefficient vector $\boldsymbol{a}$ is sparse if the dictionary $\mathbf{X}$ is overcomplete, i.e., $d \ll p$, where d is the dimensionality of the data and p is the number of atoms in the dictionary. The coefficient optimization problem is expressed using the L_1 norm in a regression framework as follows

$$\boldsymbol{a}^* = argmin\|\boldsymbol{a}\|_1 \; s.t. \; \mathbf{X}\boldsymbol{a} = \boldsymbol{y} \tag{2}$$

where $\|\boldsymbol{a}\|_1 = \sum|a_i|$. It has been shown that the L_1 norm induces sparsity and is robust to outliers. The optimization problem is cast in terms of a cost function as,

$$\boldsymbol{a}^* = argmin\| \mathbf{X}\boldsymbol{a} - \boldsymbol{y}\|_2^2 + \lambda\|\boldsymbol{a}\|_1 \tag{3}$$

where λ is a parameter regulating the amount of sparsity that we want to enforce.

This optimization problem can be solved using greedy algorithms such as Orthogonal Matching Pursuit (OMP) and Least Angle Regression (LAR) [19]. The optimization problem in Eq. (2) is a convex relaxation problem that was originally formulated using the L_0 norm. Donoho et al. [20] showed that using the L_1 norm makes the problem tractable and promotes sparsity in the coefficients. The benefit of using L_1 minimization is that the problem can be efficiently solved using convex optimization algorithms, and furthermore the signal y can be represented efficiently using a small number of dictionary elements.

During sparse representation classification, the approach taken is based on minimum reconstruction error. However, a drawback of this approach is that the dictionary is not optimized and when the number of dictionary samples is very large, the SR process can become time consuming and even unstable. It has been shown in [12] that better sparse representation is achieved when the dictionary is learned from the data instead of using a pre-defined dictionary.

In this paper, we consider three dictionary learning methods for expression recognition: K-SVD, Fisher Discriminative Dictionary Learning (FDDL) and Dictionary Learning Separating Commonality and Particularity (COPAR). A brief overview of these methods is presented next.

K-SVD

K-SVD [12] was introduced as a means to learn an overcomplete dictionary of manageable size from the training data, where each new dictionary element is a linear combination of training samples. K-SVD is an iterative technique, where at each iteration, training samples are first sparsely coded using the current dictionary estimate, and then dictionary elements are updated one at a time while keeping the remaining atoms fixed. Rubinstein *et al.* [14] implemented an efficient implementation of K-SVD using Batch Orthogonal Matching Pursuit. The objective function that K-SVD tries to solve is given in Eq. (4).

$$\{\mathbf{D}, \boldsymbol{a}\} = argmin\|\boldsymbol{x} - \mathbf{D}\boldsymbol{a}\|_2^2 \; s.t. \|\boldsymbol{a}\|_0 \leq \delta \tag{4}$$

where $\mathbf{D}$ is the learned dictionary and $\boldsymbol{a}$ are the sparse coefficients. The norm on $\boldsymbol{a}$ is the L_0 norm which counts the number of non-zero elements in the coefficient vector, and the parameter δ controls the amount of sparsity in the coefficient vector $\boldsymbol{a}$.

For classification, a simple regression based classifier is learned, which is a transformation matrix $\mathbf{C}$ that estimates the class for a given test sample $\boldsymbol{y}_i$. To solve for $\mathbf{C} \in \mathbb{R}^{P \times C}$, we define $\mathbf{H}$ as a sparse ground truth matrix, $\mathbf{H} \in \mathbb{R}^{C \times N}$ and N is the number of samples. Each column of $\mathbf{H}$ corresponds to a training sample, where the c^{th} element is set to 1 if y_i belongs to that class, or 0 otherwise. The problem is formulated as

$$\hat{\mathbf{C}} = argmin \|\mathbf{H} - \mathbf{C}^T\mathbf{A}\|_2^2 \tag{5}$$

The above can be solved directly via ridge regression which has an analytic solution given as follows:

$$\mathbf{C} = (\mathbf{A}\mathbf{A}^T)^{-1}\mathbf{A}\mathbf{H}^T \tag{6}$$

Fisher Discrimination Dictionary Learning (FDDL)

Meng *et al.* [14] proposed a discriminative dictionary learning framework that jointly learns a dictionary and discriminative sparse codes using Fisher's Discrimination criterion. Instead of learning one global dictionary for all classes, a class specific dictionary $\mathbf{D} = [\mathbf{D}_1, \mathbf{D}_2, \ldots, \mathbf{D}_c]$ is learned, where $\mathbf{D}_i$ is the class-specified sub-dictionary associated with class i, and c is the total number of classes. For classification with such dictionary $\mathbf{D}$, it is possible to use the minimum reconstruction error criterion, as done with sparse representation classification [8]. Sparse codes are obtained for each of the class specific dictionaries and the reconstruction error for each sub-dictionary is computed to determine the class label. This is often referred to as local sparse coding. The objective function $J_{(\mathbf{D},\mathbf{A})}$ is a function of the dictionary $\mathbf{D}$ and the sparse codes $\mathbf{A}$ given by

$$J_{(\mathbf{D},\mathbf{A})} = argmin_{(\mathbf{D},\mathbf{A})}\Big\{\sum_{i=1}^{c} r(\mathbf{X}_i, \mathbf{D}, \mathbf{A}_i) + \lambda_1\|\mathbf{A}\|_1 + \lambda_2 f(\mathbf{A}) + \eta\|\mathbf{A}\|_F^2\Big\} \tag{7}$$

where $f(\mathbf{A})$ is the term that promotes discrimination in the sparse codes based on Fisher's Discrimination criterion and is defined as,

$$f(\mathbf{A}) = tr\big(\mathbf{S}_{\boldsymbol{W}}(\mathbf{A}) - \mathbf{S}_{\boldsymbol{B}}(\mathbf{A})\big) + \eta\|\mathbf{A}\|_F^2 \tag{8}$$

where $\mathbf{S}_{\boldsymbol{W}}$ and $\mathbf{S}_{\boldsymbol{B}}$ are scatter matrices of the within class and between class sparse codes respectively and $\eta\|\mathbf{A}\|_F^2$ enforces convexity and is regulated by parameter η. The term $r(\mathbf{X}_i, \mathbf{D}, \mathbf{A}_i)$ is defined below.

$$r(\mathbf{X}_i, \mathbf{D}, \mathbf{A}_i) = \|\mathbf{X}_i - \mathbf{D}\mathbf{A}_i\|_F^2 + \left\|\mathbf{X}_i - \mathbf{D}_i\mathbf{A}_i^i\right\|_F^2 + \sum_{\substack{j=1 \\ j \neq i}}^{c}\left\|\mathbf{D}_j\mathbf{A}_i^j\right\|_F^2 \tag{9}$$

In addition to penalizing the reconstruction error, $r(\mathbf{X}_i, \mathbf{D}, \mathbf{A}_i)$ ensures that dictionary atoms corresponding to one sub-dictionary are representative of that dictionary alone and don't contribute to

DOI: 10.2352/ISSN.2470-1173.2016.11.IMAWM-455

the other sub-dictionaries. When learning class specific dictionaries, the summation term of $\left\|\mathbf{D}_j\mathbf{A}_i^j\right\|_F^2$ in $r(\mathbf{X}_i, \mathbf{D}, \mathbf{A}_i)$ enforces that the sparse codes are representative of that class alone and do not include other classes.

The sparse codes can be computed using the Iterative Projections Method (IPM) [21]. Since the objective function given in Eq. (7) is not jointly convex, an iterative optimization process is used, where the dictionaries are kept constant while the sparse coefficients are learned, and the sparse coefficients are kept constant while the dictionaries are learned. During the dictionary update process, each atom in a sub-dictionary is updated separately keeping the other atoms constant [14]. A detailed algorithm of the dictionary learning process and convexity of the Eq. (8) is given in [14].

Dictionary Learning Separating Commonality and Particularity (DL-COPAR)

Shu *et al.* [18] proposed a dictionary learning framework that jointly learns class specific sub-dictionaries, while simultaneously learning a shared/common dictionary. It is observed that in dictionary learning, dictionary atoms in one sub-dictionary may be correlated with atoms in another dictionary, which makes the class specific dictionary less discriminative. A shared or common dictionary contains atoms that have high coherence with atoms in class-specific sub-dictionaries. An incoherence term is introduced to learn the common dictionary $\mathbf{D}_c$ and the objective function of the DL-COPAR framework is given as:

$$J = \sum_{c=1}^{C} \left\{ \begin{matrix} \|\mathbf{X}_c - \mathbf{D}\mathbf{A}_c\|_F^2 + \left\|\widetilde{\mathbf{Q}}_{/c}^T\mathbf{A}_c\right\|_F^2 \\ +\left\|\mathbf{X}_c - \mathbf{D}\widetilde{\mathbf{Q}} \;\; \widetilde{\mathbf{Q}}^T\mathbf{A}_c\right\|_F^2 + \lambda\emptyset(\mathbf{A}_c) \end{matrix} \right\} + \eta \sum_{c=1}^{C+1}\sum_{\substack{j=1 \\ j\neq c}}^{C+1} \mathcal{Q}(\mathbf{D}_c, \mathbf{D}_j) \tag{10}$$

where,

$$\mathbf{Q}_c = \left[q_1^c, \dots, q_j^c, \dots, q_{K_c}^c\right] \in \mathbb{R}^{K\times K_c}$$
$$\widetilde{\mathbf{Q}}_{/c} = \left[\mathbf{Q}_1, \dots, \mathbf{Q}_{c-1}, \mathbf{Q}_{c+1}, \dots, \mathbf{Q}_C\right]$$

The dictionary D includes the class specific dictionary and the common dictionary. It is given as $\mathbf{D} = [\mathbf{D}_1, \dots, \mathbf{D}_c, \dots, \mathbf{D}_C, \dots, \mathbf{D}_{C+1}] \in \mathbb{R}^{d\times K}$. The sub-dictionaries $\mathbf{D}_c \in \mathbb{R}^{d\times K_c}$ and $\mathbf{D}_{c+1} \in \mathbb{R}^{d\times K_{C+1}}$ represent the class specific and common dictionary respectively, where K_c and K_{C+1} denote the number of atoms in the class specific and common dictionaries respectively. The j^{th} column of $\mathbf{Q}_c$ is a vector of zeros except for a value of 1 at the j^{th} location, i.e. the column is expressed as $q_j^c = [0, \dots, 1, \dots, 0]^T$. Hence, $\mathbf{Q}_c{}^T\mathbf{Q}_c = \mathbf{I}$. A new matrix $\widetilde{\mathbf{Q}}_c = [\mathbf{Q}_c, \mathbf{Q}_{c+1}]$ is obtained from $\mathbf{Q}_c, \mathbf{Q}_{c+1}$ corresponding to the selection operator for the class specific dictionary and the common dictionary.

The first term in Equation (10) is the reconstruction error and the regulating parameter λ controls the amount of sparsity. The term $\emptyset(\mathbf{A}_c) = \sum_{i=1}^{N_c}\|a_i^c\|_1$ is the regularization on the class specific sparse codes and $A_c = \left[a_1^c, \dots, a_{N_c}^c\right]$. The third term in Equation (10) is the class specific reconstruction error. In order to ensure that the sparse codes are representative of their respective class specific dictionary alone, the term $\left\|\widetilde{\mathbf{Q}}_{/c}^T\mathbf{A}_c\right\|_F^2$ is added to the objective function. This term enforces the coefficients to be set to zero if they do not correspond to the c^{th} class specific and common dictionary.

The incoherence term in Eq. (10) is defined as $\mathcal{Q}(\mathbf{D}_i, \mathbf{D}_j) = \left\|\mathbf{D}_i^T\mathbf{D}_j\right\|_F^2$. In [17], the incoherence term was used to eliminate similar atoms among sub-dictionaries. But in this formulation the incoherence term is used to eliminate common patterns between class specific and common dictionaries. The dictionaries and the coefficients are learned by alternatively optimizing the cost function of Eq. (10). The atoms in the dictionaries are updates that are obtained iteratively one atom at a time [18]. The coefficients are learned by keeping the dictionaries constant in Eq. (10). The equation then reduces to the LASSO problem, which can be efficiently solved by the feature-sign algorithm [23].

Experimental Results

In our experiments, we evaluate three dictionary leaning methods, K-SVD [12], FDDL [14] and DL-COPAR [15] on the extended Cohn-Kanade (CK+) facial expression dataset [22].

Figure 2. Sample images from the extended Cohn-Kanade (CK+) facial expression dataset after cropping and normalization (top to bottom, left to right) disgust, surprise, anger, happy, sad, fear.

Dataset

The extended CK+ [22] expression dataset contains 118 subjects in 327 video sequences exhibiting the expressions of anger, disgust, fear, happiness, sadness, surprise, and contempt. Since the images can be of any resolution, the images in the sequences were preprocessed to standard size before feature extraction. In the first set of experiments, five expressions were considered: happy, sad, angry, surprise and fear. For each expression sequence, the last six frames were extracted which contained the onset of the expression to the peak expression. The ground truth information provided in the database contains landmark points for each image. These landmark points were used to create a bounding box and then normalize the image with respect to eye distance and resize the normalized image to a fixed size. The images were resized to 24×21, and normalized such that they had unit L_2 norm. Samples images for each expression are shown in Fig. 2.

Three dictionary learning methods, namely K-SVD, FDDL and DL-COPAR, were evaluated by performing four-fold cross validation. In our experiments with K-SVD, a dictionary with 150 atoms was learned. For FDDL, expression specific dictionaries with 30 atoms each were learned. For DL-COPAR a common dictionary with 10 atoms and expression specific dictionaries with 30 atoms each were learned.

For classification with K-SVD, a linear regression classifier based on sparse codes was utilized and classification was performed based on minimum reconstruction error. A local classification scheme was used for FDDL and DL-COPAR, where

©2016 Society for Imaging Science and Technology
DOI: 10.2352/ISSN.2470-1173.2016.11.IMAWM-455

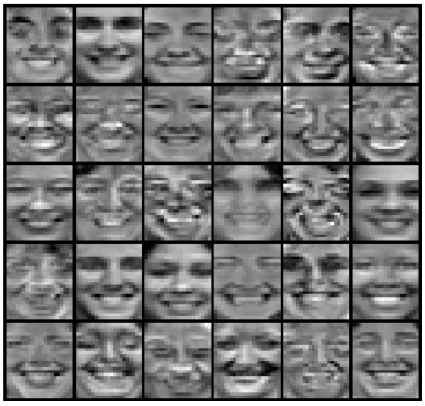
sparse codes were generated for each expression specific dictionary and the reconstruction error was computed. An expression class label was assigned based on minimum reconstruction error.

TABLE 1. CLASSIFICATION RESULTS ON THE CK+ DATASET WITH FIVE EXPRESSIONS USING FOUR FOLD CROSS VALIDATION

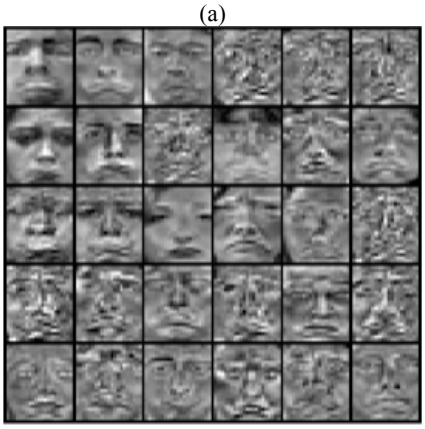
Method	Number of Expressions	Accuracy (in %)
K-SVD FDDL DL-COPAR	Five Expressions	0.93±0.003 0.95±0.01 0.99±0.01
DL-COPAR	Six Expressions	0.9811±0.0062

Table 1 reports the average and the standard deviation of the expression recognition results. The advantage of the common dictionary is reflected in the improved performance of DL-COPAR over K-SVD and FDDL. Sample basis atoms from DL-COPAR are shown in Figures 3 and 4.

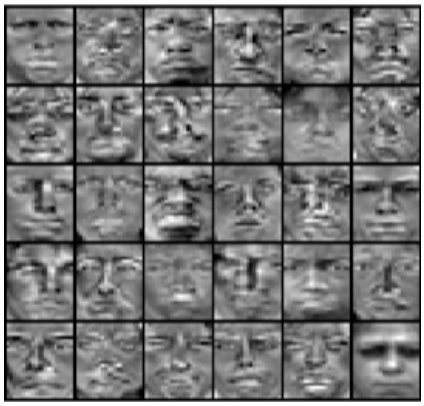
In our second experiment, we applied the DL-COPAR method for the recognition of six expressions: happy, sad, angry, surprise, fear and disgust. The expression classification results are shown in Table 1. A relatively small drop (<1%) in performance is observed when going from five to six expressions. The effectiveness of the DL-COPAR framework over the other methods is attributed to the shared/common dictionary that is learned due to the incoherence term in the dictionary learning process.



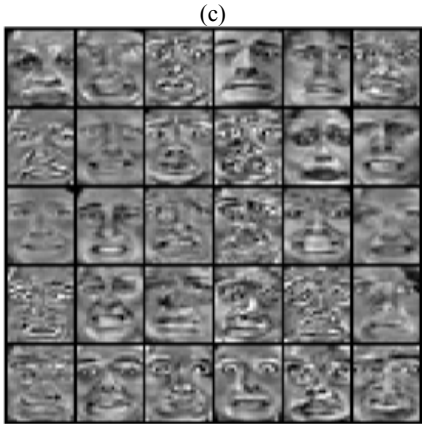
(a)



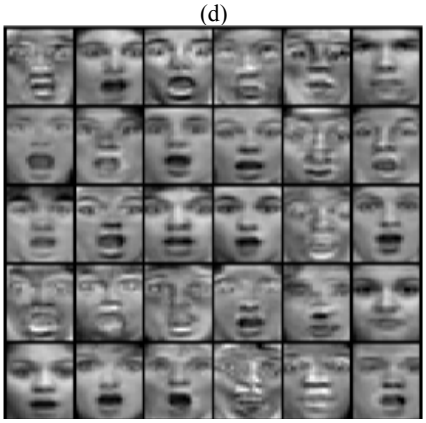
(b)



(c)



(d)



(e)

Figure 3. Sample atoms from each of the expression specific dictionaries of DL-COPAR (a) Happy, (b) Sad, (c) Angry, (d) Fear, (e) Surprise.

Figure 4. Sample dictionary atoms from the common/shared dictionary using DL-COPAR dictionary learning.

DOI: 10.2352/ISSN.2470-1173.2016.11.IMAWM-455

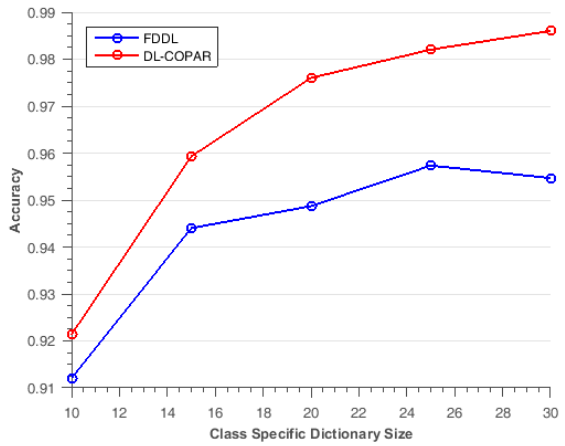


Figure 5. Comparison between FDDL and DL-COPAR with varying class specific dictionary size

Another experiment was performed to investigate the effect of class specific dictionary size on classification accuracy. The expression recognition performance of the FDDL and DL-COPAR frameworks were compared after varying the size of the expression specific dictionaries. The recognition accuracy for the two methods is plotted in Figure 5. The class specific dictionary size was increased from 10 to 30 in increments of 5. The results illustrate that initially the performance of both algorithms is comparable, up to a sub-dictionary size of 15. As the number of atoms in the expression specific dictionaries increases, the DL-COPAR dictionary learning method outperforms FDDL significantly.

Conclusion

In this paper, we explored three Dictionary Learning frameworks for facial expression recognition. The shared/common dictionary learning framework (DL-COPAR) performs exceedingly well in comparison to FDDL and K-SVD. This can be attributed to the fact that learning expression specific dictionaries along with a shared/common dictionary enhances expression recognition performance by decoupling facial identity and expression during the dictionary learning process due to the incoherence term. Excellent classification performance was obtained using DL-COPAR on extended CK+ dataset, with 99% accuracy for five expressions and 98.1% accuracy for six expressions.

References

[1] P. Ekman, and W. V. Friesen. *Manual for the facial action coding system*. Consulting Psychologists Press, 1978.

[2] Z. Zeng, M. Pantic, G. Roisman, and T. Huang, "A Survey of Affect Recognition Methods: Audio, Visual, and Spontaneous Expressions," IEEE Transactions on Pattern Analysis and Machine Intelligence, 31(1): pp. 39 - 58, 2009.

[3] F. De la Torre and J. F. Cohn, "Facial Expression Analysis," *Guide to Visual Analysis of Humans: Looking at People*, Springer, 2011.

[4] R. W. Ptucha, A. Savakis, "Facial Expression Recognition", *IGI Global Encyclopedia of Information Science and Technology*, 3rd Edition, IGI Global, 2013.

[5] T. F. Cootes, G. J. Edwards, and C. J. Taylor, "Active appearance models," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 23, pp. 681-5, 2001.

[6] I. Buciu, C. Kotropoulos, and I. Pitas, "ICA and Gabor representation for facial expression recognition," International Conference on Image Processing, 2003.

[7] X. Feng, M. Pietikainen, and A. Hadid, "Facial expression recognition based on local binary patterns," Pattern Recognition and Image Analysis, vol. 17, 2007.

[8] C. Shan, S. Gong, and P. W. McOwan, "Appearance manifold of facial expression," in Computer Vision in Human-Computer Interaction. ICCV 2005 Workshop on HCI. 5, Berlin, Germany, 2005

[9] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and M. Yi, "Robust face recognition via sparse representation," IEEE Trans. Pattern Anal. Mach. Intell., vol. 31, no. 2, pp. 210–227, Feb. 2009.

[10] S. Zafeiriou and M. Petrou, "Sparse representations for facial expressions recognition via l1 optimization," in IEEE Conference on Computer Vision and Pattern Recognition, 2010.

[11] R. Ptucha, and A. Savakis, "Manifold based sparse representation for facial understanding in natural images," *Image and Vision Computing* 31.5 (2013): 365-378, 2013.

[12] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," IEEE Trans. Signal Process., vol. 54, no. 11, pp. 4311–4322, Nov. 2006.

[13] R. Rubinstein, M. Zibulevsky, and M. Elad, "Efficient implementationof the K-SVD algorithm using batch orthogonal matching pursuit," Dept. Comput. Sci., Israel Inst. Technol., Haifa, Israel, Tech. Rep. CS-2008-08, 2008.

[14] M. Yang, L. Zhang, X. Feng, and D. Zhang, "Fisher discrimination dictionary learning for sparse representation," in Proc. IEEE Int. Conf. Computer Vision Patt. Rec (CVPR), Nov. 2011, pp. 543–550.

[15] S. Kong and D. Wang, "A dictionary learning approach for classification: Separating the particularity and the commonality," in Proc. ECCV, 2012, pp. 186–199.

[16] Z. Jiang, Z. Lin, and L. S. Davis, "Learning a discriminative dictionary for sparse coding via label consistent K-SVD," in Proc. IEEE Conf. CVPR, Jun. 2011, pp. 1697–1704.

[17] I. Ramirez, P. Sprechmann, and G. Sapiro, "Classification and clustering via dictionary learning with structured incoherence and shared features," In CVPR, 2010.

[18] Q. Zhang and B. Li, "Discriminative K-SVD for dictionary learning in face recognition," in Proc. IEEE Conf. CVPR, Jun. 2010, pp. 2691–2698.

[19] Efron, Bradley, et al. "Least angle regression." *The Annals of statistics* 32.2 (2004): 407-499.

[20] D. L. Donoho, M. Elad, and V. N. Temlyakov, "Stable recovery of sparse overcomplete representations in the presence of noise," IEEE Trans. Inf. Theory, vol. 52, no. 1, pp. 6–18, Jan. 2006.

[21] L. Rosasco, A. Verri, M. Santoro, S. Mosci, and S. Villa. Iterative Projection Methods for Structured Sparsity Regularization. MIT Technical Reports, MIT-CSAIL-TR-2009-050, CBCL-282, 2009.

[22] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews, "The extended Cohn-Kanade dataset (CK+): A complete dataset for action unit and emotion-specified expression," in Proc.

DOI: 10.2352/ISSN.2470-1173.2016.11.IMAWM-455

IEEE Conf. Computer Vision Pattern Recognition Workshops, pp. 94–101, 2010.

[23] Lee, H., Battle, A., Raina, R., & Ng, A. Y. (2006). Efficient sparse coding algorithms. In *Advances in neural information processing systems* (pp. 801-808).

Author Biography

Sriram Kumar is a graduate student pursuing the MS degree in Electrical Engineering at Rochester Institute of Technology. He graduated Fist Class in Electrical and Electronic Engineering in 2013 from Anna University, India. During fall 2015, he worked as a computer vision intern at PerceptiMed, Mountain View, CA. His research interests are in computer vision, including expression recognition, domain adaptation and robust learning.

Dr. Behnaz Ghoraani joined the Department of Biomedical Engineering at Rochester Institute of Technology in 2012 and started her research group "The Biomedical Signal and Image Analysis Lab". She completed her Ph.D. at Ryerson University (2006-2010), Toronto, Canada, and was a postdoctoral fellow in Faculty of medicine at University of Toronto (2010-2012). Dr. Ghoraani's research interests include non-stationary signal analysis and time-frequency analysis specialized in biomedical applications.

Andreas Savakis is professor of Computer Engineering at Rochester Institute of Technology, where he served as department head from 2000 to 2011. He received the B.S. and M.S. degrees from Old Dominion University and the Ph.D. from North Carolina State University, all in Electrical Engineering. Before joining RIT, he was with the Kodak Research Labs. His research interests span Image Processing, Image Understanding and Computer Vision.