Dictionary Based Approach for Facial Expression Recognition from Static Images

Krishan Sharma(✉) and Renu Rameshan

School of Computing and Electrical Engineering, Indian Institute of Technology, Mandi, Himachal Pradesh, India
krishan_sharma@students.iitmandi.ac.in, renumr@iitmandi.ac.in

Abstract. We present a simple approach for facial expression recognition from images using the principle of sparse representation using a learned dictionary. Visual appearance based feature descriptors like histogram of oriented gradients (HOG), local binary patterns (LBP) and eigenfaces are used. We use Fisher discrimination dictionary which has discrimination capability in addition to being reconstructive. The classification is based on the fact that each expression class with in the dictionary spans a subspace and these subspaces have non-overlapping directions so that they are widely separated. Each test feature point has a sparse representation in the union of subspaces of dictionary formed by labeled training points. To check recognition performance of the proposed approach, extensive experimentation is done over Jaffee and CK databases. Results show that the proposed approach has better classification accuracy than state-of-the-art techniques.

Keywords: Histogram of oriented gradient · Local binary pattern · Eigenfaces · Dictionary learning · Sparse representation · Facial expression recognition

1 Introduction

Facial expressions play an important role in human communication. Mehrabian's [1] study of non verbal communication shows that 55% of information transfers through facial expressions. Ekman [2] linked facial expressions to basic emotions (disgust, happy, sad, fear, surprise, anger).

An automatic facial expression recognition system can be used for human computer interaction by making computers more receptive to human needs but it has always been a difficult and challenging task for the computer vision community researcher from past several years. Recently many researchers are working on this problem and fast growth is seen. In fact many digital cameras in market now uses facial expression recognition algorithms but recognition error rate is still high due to large pose, scale and rotation variation, different lighting condition and occlusion.

S. Mukherjee et al. (Eds.): ICVGIP 2016, LNCS 10481, pp. 39–49, 2017.
https://doi.org/10.1007/978-3-319-68124-5_4

Sparse representation and dictionary learning techniques are successfully applied to various image processing and computer vision tasks like compression, denoising, super-resolution and classification etc. Mairal et al. [3] proposed a method to learn separate dictionaries for different classes and classification performed is based on reconstruction error, however this is a very inefficient and time consuming process when number of training classes are large.

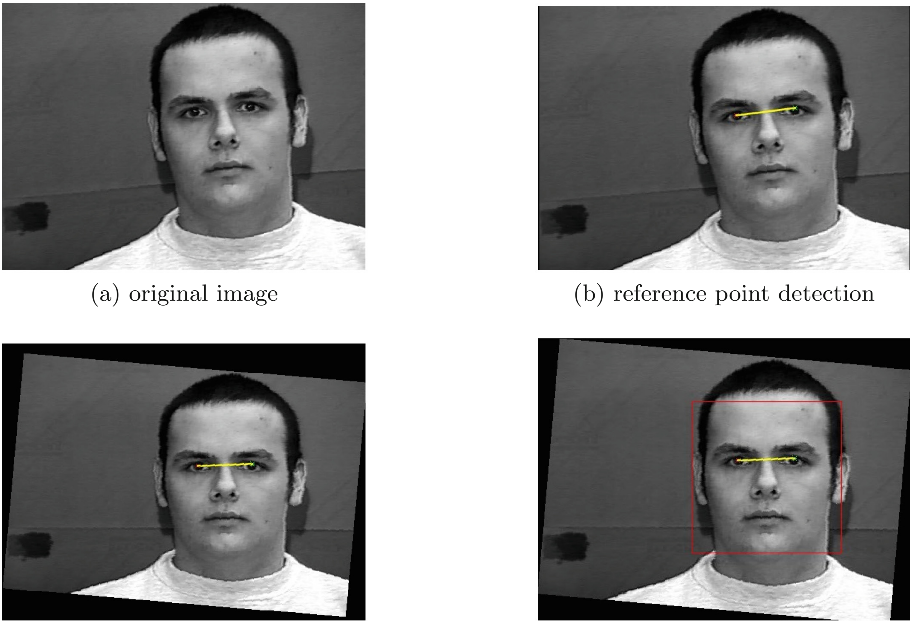
Aharon et al. [4] proposed KSVD based method to learn an overcomplete dictionary that is purely reconstructive but lacks discrimination capabilities hence not suitable for classification. The discriminative KSVD (DKSVD) [5] learns a dictionary that is based on extending the KSVD algorithm by incorporating the classification error into the objective function, discrimination is obtained by iteratively updating dictionary atoms based on linear predictive classifier. Liu et al. [6] combined Gabor features with DKSVD for facial expression recognition. Sparse coding trees [7] are supervised classification trees for expression recognition that use node-specific dictionaries and classifiers to direct input based on classification results in the feature space at each node. Guo [8] proposed a smile expression classification technique based on biologically inspired feature and patched based dictionary learning. Cottor [9] also proposed a region based weighted voting of sparse representation classifiers for facial expression recognition.

The aim of this paper is to develop a new approach which uses appearance based features along with Fisher discrimination dictionary [10] for classification. Appearance based features include histogram of oriented gradients (HOG) [11], local binary patterns (LBP) [12] and eigenfaces [13] that encode visual variations of the face images. The idea behind using the Fisher discrimination dictionary is that each expression class with in the dictionary forms a class specific subdictionary that spans a subspace and these subspaces for all different classes are widely separated and each test feature point has a sparse representation in the union of subspaces of dictionary formed by labeled training feature points.

Our contribution is using Fisher discrimination dictionary in feature space for achieving facial expression classification with an improved performance. The proposed classifier has a performance superior to the state of the art techniques. The dictionary is learned in such a way that subspaces formed by each subdictionary are orthogonal to each other. Each test image is then classified by finding the maximum projection of its feature point in these subspaces. Result shows that HOG features perform best with Fisher discrimination dictionary. The paper is structured as follows. Section 2 describes the proposed approach. Experimental analysis and results are given in Sect. 3. Section 4 describes the conclusion and future work.

2 The Proposed Approach

The proposed facial expression recognition algorithm as shown in Fig. 2 consists of the following steps, *viz.* Region of interest and reference point detection (Sect. 2.1), face normalization (Sect. 2.2), feature extraction (Sect. 2.3), learning dictionary from training data (Sect. 2.4), classification through dictionary


(a) original image (b) reference point detection

(c) face normalization (d) face ROI detection

Fig. 1. Face image registration

(Sect. 2.5). The face region of interest (ROI) from any input profile face image is detected using the Viola & Jones face detection algorithm proposed in [14]. The detected face is normalized by first detecting two reference points for image registration. Center of both eyes are chosen as the reference points for image registration. After normalization, appearance based image features are extracted and classification is performed with learned dictionary.

2.1 Reference Point Detection

The major problems which occur in expression recognition from face images are affine (scaling, translation and rotation) variations. To overcome these problems, face images are normalized first by detecting two reference points *viz* e1 and e2. e1 and e2 are denoted as left eye and right eye center respectively which are detected automatically using template matching (Fig. 1b). Five different templates for each eye are made and co-ordinates to the reference points are assigned based on maximum voting strategy over correlation values.

2.2 Face Normalization

Face normalization (Fig. 1c) is required to overcome the issues caused by affine transformation such as rotation and scaling. It is required that all the face images

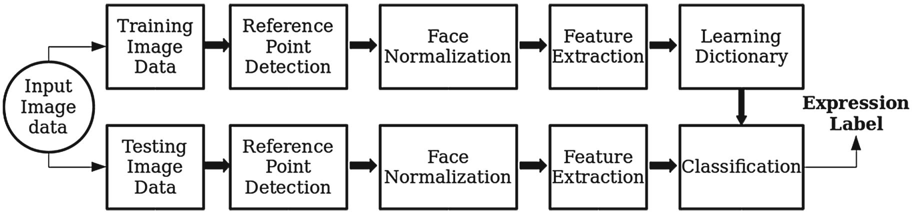


Fig. 2. Flow diagram of proposed approach

in the database should be of same size and properly registered around the given reference points i.e. reference points in all the database images should lie at same co-ordinate locations.

Let (a_1^0, b_1^0) and (a_2^0, b_2^0) denotes the co-ordinates of reference points e1 and e2 respectively in the original image I_0. Reference point co-ordinates $e1'$ and $e2'$ in the registered image template I_1 are (a_1^1, b_1^1) and (a_2^1, b_1^1) respectively. Image I_0 is scaled to Image I_s with respect to reference point e1 by a scaling factor $K = \frac{d_0}{dist(e1,e2)}$, where d_0 is a predefined distance between the reference points ($e1'$ and $e2'$) of the reference image template obtained by scaling matrix $[T_s]$. Image I_s is rotated to Image I_r with respect to e1 using rotation matrix $[T_r]$ by an angle $\theta = \frac{b_2^0 - b_1^0}{a_2^0 - a_1^0}$, where θ is the angle between horizontal axis and line joining e1 and e2. Image I_r is translated to I_t by $t_x = a_1^1 - a_1^0$ and $t_y = b_1^1 - b_1^0$ in x and y direction respectively using translation matrix $[I_t]$. The image transformation matrix from original image I_0 to the register image I_1 is given as:

$$T = \underbrace{\begin{bmatrix} cos\theta & -sin\theta & a_1^o \cdot (1 - cos\theta) + b_1^o \cdot sin\theta \\ sin\theta & cos\theta & b_1^o \cdot (1 - cos\theta) - a_1^o \cdot sin\theta \\ 0 & 0 & 1 \end{bmatrix}}_{\text{Rotation Matrix } [T_r\ ] \ w.r.t. \text{ point } e_1} \underbrace{\begin{bmatrix} K & 0 & a_1^o \cdot (1 - K) \\ 0 & K & b_1^o \cdot (1 - K) \\ 0 & 0 & 1 \end{bmatrix}}_{\text{Scaling Matrix } [T_s] \ w.r.t. \text{ point } e_1} \underbrace{\begin{bmatrix} 1 & 0 & t_x \\ 0 & 1 & t_y \\ 0 & 0 & 1 \end{bmatrix}}_{\text{Translation } [T_t]} \quad (1)$$

2.3 Feature Extraction

In this work, three different types of feature extraction methods are used which exploit different key properties of facial expressions. All of them are discussed here in detail.

Histogram of Oriented Gradients: The well known HOG descriptor [11] is used to extract the facial expression features. The prime advantage of using the HOG representation is that it captures edge or gradient structure that is a very good characteristic of facial expressions. HOG features are also invariant to geometric and photometric transformations. Contrast normalization in HOG makes it invariant to shadowing and illumination changes.

Local Binary Patterns: LBP features [12] are used to encode spatial variations from face image which is the effective representation of facial features as it is robust to rotation and illumination variations. The binary patterns are generated at a pixel by thresholding its neighboring pixels and the threshold is selected based on the value at the center pixel. Histograms of these patterns represent an image as the feature vector.

Eigenfaces: Eigenfaces [13] are the set of eigenvectors obtained from the covariance matrix of training data. Face images can be represented in a low dimension subspaces of original face space by projecting the face image over eigenvectors (eigenfaces) sorted by their eigenvalue.

2.4 Learning Dictionary from Training Data

Fisher discrimination dictionary proposed by Yang et al. [10] is used for classification purpose due to its high discrimination and reconstruction capabilities. We are giving a brief overview of Fisher discrimination dictionary in this section. Unlike the previous dictionary learning methods which learn either the shared dictionary or different class specific dictionaries, this method learns a structured dictionary $\mathbf{D} = [\mathbf{D}_1, \mathbf{D}_2, \ldots, \mathbf{D}_j, \ldots \mathbf{D}_c] \in \mathbb{R}^{n*p}$, where c, n and p denotes the total number of classes, feature vector dimension and total number of dictionary atoms respectively. $\mathbf{D}_j \in \mathbb{R}^{n*p_j}$ is the class specific sub-dictionary for j^{th} class with p_j dictionary atoms.

While learning the Fisher dictionary, three constraints are enforced. First constraint is that each class specific sub-dictionary should have well representation of training data belong to that class and poor representation belong to other classes. Second constraint is that representation coefficients should have large between class scatter and small within class scatter and third constraint is the high sparsity of representation coefficients.

Let $\mathbf{Y} = [\ \mathbf{Y}_1, \mathbf{Y}_2, \ldots, \mathbf{Y}_j, \ldots \mathbf{Y}_c] \in \mathbb{R}^{n*N}$ denotes the input training data with N examples, where $\mathbf{Y}_j \in \mathbb{R}^{n*N_i}$ are the data samples belongs to j^{th} class. Similarly, sparse coefficients are defined as $\mathbf{X} = [\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_j, \ldots \mathbf{X}_c] \in \mathbb{R}^{p*N}$ such that $\mathbf{Y} \approx \mathbf{DX}$. It is desirable that dictionary $\mathbf{D}$ should be discriminative and reconstructive simultaneously along with sparse coding coefficients. Therefore, cost function can be defined as:

$$
\begin{aligned}
J_{\{\mathbf{D},\mathbf{X}\}} = & \underbrace{\min_{\mathbf{D},\mathbf{X}} \sum_{j=1}^{K} (||\mathbf{Y}_j - \mathbf{D}\mathbf{X}_j||_F^2 + ||\mathbf{Y}_j - \mathbf{D}_j \mathbf{X}_j^j||_F^2)}_{\text{reconstruction term}} \\
& + \underbrace{\lambda_1 (tr(S_w(\mathbf{X})) - tr(S_b(\mathbf{X})) + \eta ||\mathbf{X}||_F^2)}_{\text{discrimination term}} \\
+ \quad & \underbrace{\lambda_2 ||\mathbf{X}||_1}_{\text{sparsity constraint}} \qquad s.t. \quad ||\mathbf{D}_i \mathbf{X}_j^i||_F^2 \le \epsilon_f \quad \forall j \neq i
\end{aligned}
\tag{2}
$$

where $S_w(\mathbf{X})$ and $S_b(\mathbf{X})$ are within and between class scatter respectively. λ_1, λ_2 and η are parameters. $||\mathbf{X}||_F^2$ is added to make the discriminative term convex.

This cost function is non-convex jointly with $\mathbf{D}$ and $\mathbf{X}$. To make it convex, this is optimized with respect to one variable keeping other fixed. Thus the optimization method is an iterative algorithm with alternate sparse coding and dictionary update stages. More details can be found in [10].

2.5 Classification Through Dictionary

Let a test sample $\mathbf{y}$ belongs to class $\mathbf{Y}_j$ is represented sparsely as $\mathbf{x}$ over learned dictionary $\mathbf{D}$. Let $\boldsymbol{\mu}_j$ denotes the mean vector of representation coefficients $\mathbf{X}_j$ which is the sparse representation of $\mathbf{Y}_j$ over the dictionary $\mathbf{D}$. For test sample $\mathbf{y}$ to belong to $\mathbf{Y}_j$, it should be represented well by class specific sub-dictionary $\mathbf{D}_j$ i.e. representation residual error should be least with $\mathbf{D}_j$ and simultaneously representation coefficient should be closest to $\boldsymbol{\mu}_j$. Sparse representation $\hat{\mathbf{x}}$ of test sample $\mathbf{y}$ over dictionary $\mathbf{D}$ can be solved as:

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}}\{||\mathbf{y} - \mathbf{D}\mathbf{x}||_2^2 + \gamma||\mathbf{x}||_1\} \tag{3}$$

where γ is a regularization parameter between reconstruction term and sparsity constraint term. $||.||_p$ is p^{th} norm where p = 1 or 2.

$\hat{\mathbf{x}} = [\hat{\mathbf{x}_1}, \hat{\mathbf{x}_2}, \ldots, \hat{\mathbf{x}_c}]$; where $\hat{\mathbf{x}_j}$ denotes the representation coefficient over class specific sub-dictionary $\mathbf{D}_j$. If test sample $\mathbf{y}$ belongs to j^{th} class then representation residual error $||\mathbf{y} - \mathbf{D}\hat{\mathbf{x}}_j||_2^2$ will be less than $||\mathbf{y} - \mathbf{D}\hat{\mathbf{x}}_i||_2^2$ for $j \neq i$ and also $\hat{\mathbf{x}}_j$ will be near to mean vector $\boldsymbol{\mu}_j$ than $\boldsymbol{\mu}_i$ for $j \neq i$. By incorporating both the representation residual error and representation coefficient term in classification, metric for classification can be defined as:

$$e_j = ||\mathbf{y} - \mathbf{D}_j\hat{\mathbf{x}}_j||_2^2 + \alpha.||\hat{\mathbf{x}} - \boldsymbol{\mu}_j||_2^2 \tag{4}$$

where α is the weight balance term. Class label for test sample y is given as:

$$classlabel\{y\} = argmin_j\{e_j\} \tag{5}$$

2.6 Subspace Analysis of Learned Dictionary

Fisher dictionary can be seen as structural combination of different class specific sub-dictionaries where each sub-dictionary is spanning a subspace within dictionary space. It is desirable that each pair of sub-dictionaries should be maximally incoherent making each pair of subspaces orthogonal to each other. To make each sub-dictionary more discriminative and to promote the incoherence between sub-dictionaries, Yang et al. added $(||D_i X_j^i||_F^2 \leq \epsilon_f \quad \forall j \neq i)$ as a penalty term to the cost function. This constraint enforces that j^{th} class data can not be represented by i^{th} class sub-dictionary hence should make each sub-dictionary maximally incoherent with each other i.e. subspaces should be mutually orthogonal. Definition of principal angles is given in Definition 1. Definition shows that subspaces are orthogonal if any of the principal angle is $\pi/2$.

Definition 1 [15]. *Let $\boldsymbol{S}_i$ and $\boldsymbol{S}_j$ are the subspaces spanned by sub-dictionaries $\boldsymbol{D}_i$ and $\boldsymbol{D}_j$ respectively. Let columns of $\boldsymbol{Q_{Si}} \in \mathbb{R}^{n*k}$ and $\boldsymbol{Q_{Sj}} \in \mathbb{R}^{n*k}$ are the orthonormal bases for these subspaces. The principal angles ($\theta_l \in [0, \pi/2]$; l=1,...,k) between these subspaces can be defined as*

$$\begin{aligned} cos\ \theta_l = \max_{\boldsymbol{u_l} \in span(\boldsymbol{Q_{Si}})} \max_{\boldsymbol{v_l} \in span(\boldsymbol{Q_{Sj}})} & \boldsymbol{u_l}^T \boldsymbol{v_l} \\ s.t.: \quad & ||\boldsymbol{u_l}||_2 = ||\boldsymbol{v_l}||_2 = 1 \\ \boldsymbol{u_l}^T \boldsymbol{u_m} = 0; \quad & m = 1, 2, \ldots, l-1 \\ \boldsymbol{v_l}^T \boldsymbol{v_m} = 0; \quad & m = 1, 2, \ldots, l-1 \end{aligned} \tag{6}$$

Total angle between subspaces S_i and S_j is defined as

$$\theta_T = cos^{-1}(\prod_{l=1}^{k} cos\ \theta_l) \tag{7}$$

The principal angles (in degree) in HOG feature space between 13-dimensional subspaces for sad and neutral expressions in Jaffee Database are [20.45 33.77 40.51 44.43 50.46 51.83 55.93 61.81 68.66 70.97 76.82 82.80 84.36]. These angles show that though the two subspaces are not exactly orthogonal, yet they do not have any common direction i.e. these subspaces are widely separated but not maximally separated.

3 Experimental Analysis

3.1 Databases

The Japanese Female Facial Expression (JAFFEE) Database. This database [16] consists of 213 grayscale images with seven different facial expression (happy, sad, surprise, fear, disgust, anger, neutral) posed by 10 Japanese female models. Each expression image has been rated on 6 emotion adjectives by 60 Japanese subjects (Fig. 3).

Fig. 3. Jaffee database [16] sample images of anger, disgust, fear, happy, neutral, sad and surprise emotion respectively

Cohn-Kanade AU-Coded Expression Database. This database [17] includes 486 sequences from 97 posers. Each sequence begins with a neutral expression and proceeds to a peak expression. The peak expression for each sequence in fully FACS coded and given an emotion label. The emotion label refers to what expression was requested rather than what may actually have been performed. Subjects in the released portion of the Cohn-Kanade AU-coded facial expression database are 97 university students. They ranged in age from 18 to 30 years. Sixty-five percent were female, fifteen percent were African-American, and three percent were Asian or Latino. Image sequences from neutral to target display were digitized into 640 by 480 or 490 pixel arrays with 8-bit precision for grayscale values (Fig. 4).

Fig. 4. CK database [17] sample images of anger, disgust, fear, happy, sad and surprise emotion respectively

3.2 Testing Strategy

The image database is divided into two parts, one-third data is used for testing and rest for training. First we extract the face from raw image data using Viola & Jones face detection algorithm then each face is register by two reference points which are center of both the eyes and image is normalized to 120*96 pixel size. Fisher dictionary is learned by feature extracted from training data. For testing, 3-fold cross validation is performed over both datasets.

Anger	Disgust	Fear	Happy	Neutral	Sad	Surprise
AN	DI	FE	HA	NE	SA	SU

3.3 Results and Discussion

The overall recognition results for both Jaffee and CK database are shown in Tables 1 and 2 respectively. Three different features are used to learn Fisher discrimination dictionary. Experimentation results show that HOG features with Fisher dictionary gives best recognition results. This is due to the fact that

HOG captures edges or gradient structures. These structures represent the characteristics of facial expressions. Recognition performance of eigenface features over CK database is poor due to large variation in face skin color of individual subject varying from very dark to fair. Tables 3 and 4 show the comparison of our approach with the state of art techniques and hence we can conclude that our results are comparable with previous approaches.

Table 1. Recognition Performance (%) over Jaffee database

	LBP Features							Eigenface Features							HOG Features						
	AN	DI	FE	HA	NE	SA	SU	AN	DI	FE	HA	NE	SA	SU	AN	DI	FE	HA	NE	SA	SU
AN	93.33	6.67	0	0	0	0	0	93.34	0	0	3.33	3.33	0	0	96.67	3.33	0	0	0	0	0
DI	10	90	0	0	0	0	0	3.33	80	6.67	0	0	6.67	3.33	6.67	90	0	0	0	3.33	0
FE	0	3.33	96.67	0	0	0	0	0	6.67	83.33	3.33	0	6.67	0	0	0	100	0	0	0	0
HA	0	0	0	96.67	0	0	3.33	0	0	3.33	66.7	23.33	6.67	0	0	0	0	96.67	0	3.33	0
NE	3.33	0	0	0	93.33	0	3.33	0	0	3.33	6.67	90	0	0	0	0	0	0	100	0	0
SA	6.67	3.33	13.33	0	0	76.67	0	6.67	3.33	6.67	0	3.33	76.67	3.33	0	3.33	3.33	3.33	0	90	0
SU	0	0	0	0	3.33	0	96.67	0	6.67	3.33	0	0	3.33	86.67	0	0	3.33	3.33	0	0	93.34
Rec. Rate	91.90							82.38							95.23						

Table 2. Recognition performance (%) over CK database

	LBP Features							Eigenface Features							HOG Features						
	AN	DI	FE	HA	NE	SA	SU	AN	DI	FE	HA	NE	SA	SU	AN	DI	FE	HA	NE	SA	SU
AN	93.33	6.67	0	0	0	0	0	93.34	0	0	3.33	3.33	0	0	96.67	3.33	0	0	0	0	0
DI	10	90	0	0	0	0	0	3.33	80	6.67	0	0	6.67	3.33	6.67	90	0	0	0	3.33	0
FE	0	3.33	96.67	0	0	0	0	0	6.67	83.33	3.33	0	6.67	0	0	0	100	0	0	0	0
HA	0	0	0	96.67	0	0	3.33	0	0	3.33	66.7	23.33	6.67	0	0	0	0	96.67	0	3.33	0
NE	3.33	0	0	0	93.33	0	3.33	0	0	3.33	6.67	90	0	0	0	0	0	0	100	0	0
SA	6.67	3.33	13.33	0	0	76.67	0	6.67	3.33	6.67	0	3.33	76.67	3.33	0	3.33	3.33	3.33	0	90	0
SU	0	0	0	0	3.33	0	96.67	0	6.67	3.33	0	0	3.33	86.67	0	0	3.33	3.33	0	0	93.34
Rec. Rate	91.90							82.38							95.23						

Table 3. Recognition performance (%) of state-of-the-art techniques over JAFFEE database

Method	Recognition accuracy
Optical flow + SVM [18]	87.5
Gabor + DKSVD [5]	94.3
2DPCA + SVM [19]	94
LBP + DKSVD [5]	78.6
2D locality preserving projections [20]	95
LGIP + SVM [21]	92.4
Proposed approach (HOG + FDD)	**95.23**

Table 4. Recognition performance (%) of state of art techniques over CK database

Method	Recognition accuracy(%)
Optical flow + SVM [18]	89.2
Muscle based feature + SVM [22]	85.9
Differential AAM + KNNS [23]	86.5
Rankboost with l1 regularization [24]	88
AU Dictionary + Sparse coding [25]	88.2
Proposed approach (HOG + FDD)	**93.33**

4 Conclusion and Future Work

In this paper, we have described a Fisher discrimination dictionary based approach to recognize facial expressions from static images. Various appearance based descriptors like histogram of oriented gradient (HOG), local binary pattern (LBP) and eigenfaces are used and results show that HOG features along with Fisher discrimination dictionary gives the best recognition performance. For future, this work can be extended to recognize facial expression from videos using temporal information.

References

1. Mehrabian, A.: Nonverbal Communication. Transaction Publishers, Piscataway (1972)
2. Ekman, P., Rolls, E., Perrett, D., Ellis, H.: Facial expressions of emotion: an old controversy and new findings [and discussion]. Philos. Trans. R. Soc. B: Biol. Sci. **335**(1273), 63–69 (1992)
3. Mairal, J., Bach, F., Ponce, J., Sapiro, G., Zisserman, A.: Discriminative learned dictionaries for local image analysis. In: IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2008. IEEE, pp. 1–8 (2008)
4. Aharon, M., Elad, M., Bruckstein, A.: K-SVD: an algorithm for designing overcomplete dictionaries for sparse representation. IEEE Trans. Sig. Process. **54**(11), 4311 (2006)
5. Zhang, Q., Li, B.: Discriminative K-SVD for dictionary learning in face recognition. In: 2010 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, pp. 2691–2698 (2010)
6. Liu, W., Song, C., Wang, Y.: Facial expression recognition based on discriminative dictionary learning. In: 2012 21st International Conference on Pattern Recognition (ICPR). IEEE, pp. 1839–1842 (2012)
7. Chen, H.C., Comiter, M.Z., Kung, H.T., McDanel, B.: Sparse coding trees with application to emotion classification. In: 2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 77–86, June 2015
8. Guo, L.: Smile expression classification using the improved BIF feature. In: 2011 Sixth International Conference on Image and Graphics (ICIG), pp. 783–788, August 2011

9. Cotter, S.F.: Weighted voting of sparse representation classifiers for facial expression recognition. In: 2010 18th European Signal Processing Conference. IEEE, pp. 1164–1168 (2010)
10. Yang, M., Zhang, L., Feng, X., Zhang, D.: Fisher discrimination dictionary learning for sparse representation. In: 2011 International Conference on Computer Vision. IEEE, pp. 543–550 (2011)
11. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2005), vol. 1, pp. 886–893. IEEE (2005)
12. Ojala, T., Pietikainen, M., Maenpaa, T.: Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Trans. Pattern Anal. Mach. Intell. **24**(7), 971–987 (2002)
13. Turk, M.A., Pentland, A.P.: Face recognition using eigenfaces. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Proceedings CVPR 1991. IEEE, pp. 586–591 (1991)
14. Viola, P., Jones, M.J.: Robust real-time face detection. Int. J. Comput. Vis. **57**(2), 137–154 (2004)
15. Harandi, M., Sanderson, C., Shen, C., Lovell, B.C.: Dictionary learning and sparse coding on Grassmann manifolds: an extrinsic solution. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 3120–3127 (2013)
16. Lyons, M.J., Akamatsu, S., Kamachi, M., Gyoba, J., Budynek, J.: The Japanese Female Facial Expression (JAFFE) Database (1998)
17. Lucey, P., Cohn, J.F., Kanade, T., Saragih, J., Ambadar, Z., Matthews, I.: The extended Cohn-Kanade dataset (CK+): a complete dataset for action unit and emotion-specified expression. In: 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Workshops. IEEE, pp. 94–101 (2010)
18. Liao, C.-T., Chuang, H.-J., Duan, C.-H., Lai, S.-H.: Learning spatial weighting via quadratic programming for facial expression analysis. In: 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Workshops. IEEE, pp. 86–93 (2010)
19. Oliveira, L., Mansano, M., Koerich, A., de Souza Britto Jr., A.: 2D principal component analysis for face and facial-expression recognition. Comput. Sci. Eng. **13**(3), 9–13 (2011)
20. Cheng, F., Yu, J., Xiong, H.: Facial expression recognition in Jaffe dataset based on Gaussian process classification. IEEE Trans. Neural Netw. **21**(10), 1685–1690 (2010)
21. Lubing, Z., Han, W.: Local gradient increasing pattern for facial expression recognition. In: 19th IEEE International Conference on Image Processing (ICIP) (2012)
22. Benli, K.S., Eskil, M.T.: Extraction and selection of muscle based features for facial expression recognition. In: 2014 22nd International Conference on Pattern Recognition (ICPR), pp. 1651–1656, August 2014
23. Cheon, Y., Kim, D.: Natural facial expression recognition using differential-aam and manifold learning. Pattern Recogn. **42**(7), 1340–1350 (2009)
24. Yang, P., Liu, Q., Metaxas, D.N.: Rankboost with L1 regularization for facial expression recognition and intensity estimation (2009)
25. Taheri, S., Qiu, Q., Chellappa, R.: Structure-preserving sparse decomposition for facial expression analysis. IEEE Trans. Image Process. **23**(8), 3590–3603 (2014)