Facial Expression Recognition using Histogram of Oriented Gradients with SVM-RFE Selected Features

Sumeet Saurav[1, 2], Sanjay Singh[2], and Ravi Saini[2]

[1] Academy of Scientific & Innovative Research (AcSIR), Ghaziabad, India
[2] CSIR-Central Electronics Engineering Research Institute, Pilani, India
sumeet@ceeri.res.in

Abstract. This study is an attempt towards improving the accuracy and execution time of a facial expression recognition (FER) system. The algorithmic pipeline consists of a face detector block, followed by a facial alignment and registration, feature extraction, feature selection, and classification blocks. The proposed method utilizes histograms of oriented gradients (HOG) descriptor to extract features from expressive facial images. Support vector machine recursive feature elimination (SVM-RFE), a powerful feature selection algorithm is applied to select the most discriminant features from high-dimensional feature space. Finally, the selected features were fed to a support vector machine (SVM) classifier to determine the underlying emotions from expressive facial images. Performance of the proposed approach is validated on three publicly available FER databases namely CK+, JAFFE, and RFD using different performance metrics like recognition accuracy, precision, recall, and F1-Score. The experimental results demonstrated the effectiveness of the proposed approach in terms of both recognition accuracy and execution time.

Keywords: Facial expression recognition (FER), Histogram of oriented gradients (HOG), Feature selection, Support vector machine (SVM) classifier.

1 Introduction

Psychological study has revealed facial expression as one of the most powerful ways through which humans communicate their emotions, cognitive states, intensions, and opinions to each other [1]. It is a well-known fact that facial expression contain nonverbal communication cues, which helps to identify the intended meaning of the spoken words in face-to-face communication. Therefore, there is a huge demand of an efficient and robust facial expression recognition (FER) system for a real-world human computer interaction (HCI) system. An automated FER technology equipped with robots can talk to children and take care of elderly people. This technology can also be used in hospitals to monitor patients, which will in turn save precious time and money. Additionally, FER technology can be applied in a car to identify the fatigue level of the drives which will avoid accidents and save lives.

Recognition of facial expression of a person either using a static image or sequence of images coming from a video stream is a well-studied problem since last decades.

However, the presented techniques available in literature have not yet achieved the desired performance (in terms of recognition accuracy and computation time) leading towards real-world deployment of FER systems. This may be attributed towards lack of efficient discriminative feature extractor coupled with robust classifier having real-time computing capability. Available works in literature on FER could be classified either based on appearance features or geometrical features. Since, this work has made use of appearance-based FER system, therefore, a brief review of some of the related works available in literature has been discussed below.

A comprehensive review of FER system for person-independent FER based on Local Binary Pattern (LBP) has been discussed by the authors in [2]. Other works using LBP features coupled with Kernel Discriminant Isometric map (KDIsomap) and Discriminant Kernel Locally Linear Embedding (DKLLE) has been discussed in [3] and [4] respectively. The authors in [5] have proposed an automatic FER in which the LBP features were extracted from the salient facial patches and classified using support vector machine (SVM) classifier. In order to overcome the noise susceptibility of LBP, a new descriptor called Local Ternary Pattern (LTP) was proposed by [6]. Inspired by the usefulness of LTP a new descriptor called Gradient Local Binary Pattern (GLTP) [7] was proposed for automated FER. Recently, an improved version of the GLTP called Improved GLTP (IGLTP) has also been reported in [8]. A new feature descriptor called the Compound Local Binary Pattern (CLBP) has been proposed for the purpose of FER by the authors in [9]. A novel local feature descriptor called local directional number pattern (LDN) has been proposed for the purpose of face analysis and expression recognition by the authors in [10]. Another very popular descriptor called Weber local descriptor (WLD) has also been utilized for the purpose of facial expression recognition. In [11], the authors have used the multi-scale version of this descriptor for extracting facial traits which were then classified using a support vector machine-based classifier. A novel technique called Weber Local Binary Image Cosine Transform (WLBI-CT) has been proposed by authors in [12]. Apart from the texture-based information, some works have also utilized shape-based information extracted using Histogram of Oriented Gradients (HOG). Use of HOG descriptor for facial expression recognition has been possibly first discussed in [13]. Automated facial expression recognition based on histogram of oriented gradient feature vector differences has been proposed by the authors in [14]. A comprehensive study on the application of histogram of oriented gradients (HOG) descriptor in the facial expression recognition problem has been done in [15]. In this work, the authors have investigated the importance of different HOG parameters and their impact on the classification accuracy of the facial expression recognition. Inspired by the effectiveness of the HOG descriptor for preserving the local information using orientation density distribution and gradient of the edge, the authors in [16], have proposed a novel technique which consists of transforming the HOG features to frequency domain thereby making this descriptor one of the most suitable to characterize illumination and orientation invariant facial expressions. The HOG descriptor is also used in conjunction with other descriptors. For instance, in [17], facial expression recognition with fusion features extracted from the silent facial areas using LBP and HOG has been proposed.

This study presents an algorithmic pipeline leading towards improvement in recognition accuracy and execution time of a FER system. Motivated by the success of

Histogram of Oriented Gradients (HOG) in facial expression recognition task, we have also used this descriptor for the purpose of extracting facial traits from different expressions. Since the extracted features are usually of large size, therefore, in order to overcome the limitation of the curse of dimensionality and extensive computation, we have used an embedded feature selection algorithm called support vector machine recursive feature elimination (SVM-RFE) for the purpose of removing irrelevant and redundant features from the original feature space. The selected features extracted from different expressive facial images are classified using a support vector machine (SVM) classifier.

The remainder of this paper is organized as follows: In section 2, the proposed facial expression recognition is described. Experimental results and discussion have been described in section 3. Final conclusion has been made in section 4.

2 Proposed Methodology

Facial expression recognition from a generic image requires an algorithmic pipeline that involves different operating blocks as shown in Fig.1. The red arrow indicates the path followed during the training phase and the green one during the testing phase of the pipeline. First step detects the human face in the image under investigation which is then aligned and registered to a standard size of 65 x 59 pixels as recommended in [15]. Here we have used the Multiblock-Local Binary Pattern (MB-LBP) features based version of the Viola and Jones face detector. The pre-trained cascade classifier used has been obtained from the work mentioned in [18]. For facial landmarks detection, a robust technique called Supervised Descent Method (SDM) also known as Intraface (IF) is used [19]. Using different facial landmark coordinates obtained from IF, the facial image is registered. This registration step as mentioned in [15] makes it sure that the position of the eyes in different images have the same position. This helps the HOG descriptor in extracting features from different facial images with similar spatial reference. The vector of features extracted by HOG is then passed through an embedded feature selection (FS) block called support vector machine recursive feature elimination (SVM-RFE) proposed by the authors in [20]. SVM-RFE feature selection block selects the most discriminant features that separates the expressions and gives optimal features with reduced dimension using the iterative algorithm shown in Fig. 2. SVM-RFE uses criteria derived from the coefficients in SVM models to assess features, and recursively removes features that have small criteria. We have used the linear version of the SVM-RFE in our experiments based on One-Versus-Rest (OVR) approach for selecting features from different facial expressions. The selected features are finally used by SVM classifier to classify the facial emotions by means of One-Versus-One (OVO) multi-class classification strategy.

3 Experimental Results and Discussions

In this section, we describe various experiments performed in this work. All the experiments were carried out on a laptop with 2.50 GHz Core i5 processor and 4 GB of RAM, running under Windows 8 working framework. The proposed system is simulated using Matlab 2015b tool.

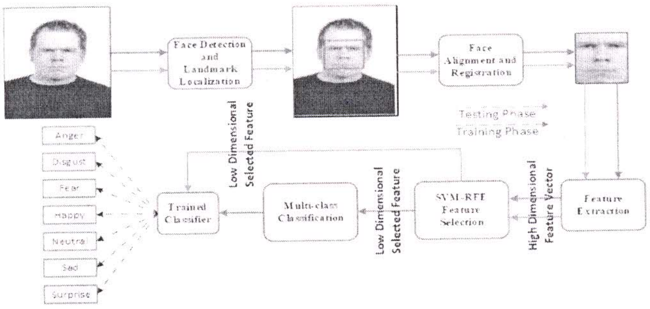
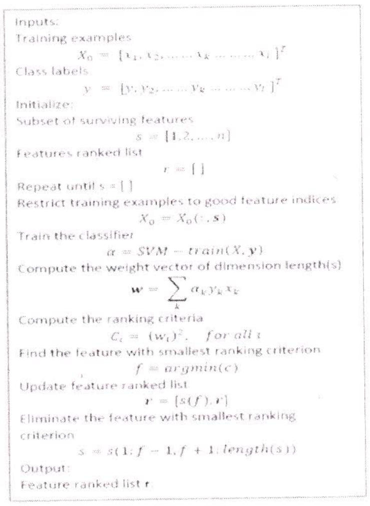

Face Detection and Landmark Localization
Face Alignment and Registration
Testing Phase
Training Phase
Low Dimensional Selected Feature
Anger
Disgust
Fear
Happy
Neutral
Sad
Surprise
Trained Classifier
Multi-class Classification
Low Dimensional Selected Feature
SVM-RFE Feature Selection
High Dimensional Feature Vector
Feature Extraction

Fig. 1. Proposed facial expression recognition algorithmic pipeline



```
Inputs:
Training examples
        $X_0 = [x_1, x_2, \ldots\ldots x_k \ldots\ldots x_l]^T$
Class labels
        $y = [y_1, y_2, \ldots\ldots y_k \ldots\ldots y_l]^T$
Initialize:
Subset of surviving features
        $s = [1, 2, \ldots, n]$
Features ranked list
        $r = [\,]$
Repeat until s = [ ]
Restrict training examples to good feature indices
        $X_0 = X_0(:, s)$
Train the classifier
        $\alpha = SVM - train(X, y)$
Compute the weight vector of dimension length(s)
        $\mathbf{w} = \sum_k \alpha_k y_k x_k$
Compute the ranking criteria
        $C_i = (w_i)^2, \quad for\ all\ i$
Find the feature with smallest ranking criterion
        $f = argmin(c)$
Update feature ranked list
        $\mathbf{r} = [s(f), \mathbf{r}]$
Eliminate the feature with smallest ranking
criterion
        $s = s(1: f - 1, f + 1: length(s))$
Output:
Feature ranked list r
```

Fig. 2. SVM-RFE feature selection pseudo-code

3.1 Datasets

Three publicly available benchmark FER datasets namely the CK+ [22], JAFFE [23], and the RFD [24] were used for conducting experiments. The CK+ database is an extended version of the CK database which contains both male and female subjects. In this study, we used a total of 407 images obtained with the following distribution among the considered classes of expressions: anger (An: 45), disgust (Di: 59), fear (Fe: 50), happy (Ha: 69), sad (Sa: 56), neutral (Ne: 60), and surprise (Su: 68). The 6-

expression version of the database excludes the neural expression images from the above distribution. The second dataset named Rebounds Faces Database (RFD) consists of facial images from 8 expressions (anger, disgust, fear, happiness, contemptuous, sadness, surprise and neutral) filmed using 67 subjects looking at three directions with 5 different face orientations. Three categories of expressions obtained from the database has been used in our experiments. The first category called RFD category 1, consists of images comprising 7 expressions (anger, contempt, disgust, fear, happy, sad, and surprise) consisting of a total of 469 images with 67 images coming from each facial expression. The second category called RFD category 2 also consists of 7 prototypic expression (anger, disgust, fear, happy, neutral, sad, and surprise) with similar distribution to that of RFD category 1 images. Finally, the third category of expressions called RFD 8 obtained from the database consists of all the eight prototypic expressions (anger, contempt, disgust, fear, happy, neutral, sad, and surprise) with a total of 536 images. Experiments were also carried out using the Japanese female facial Expression (JAFFE) dataset. This dataset contains 7 different prototypic facial expression images: anger, disgust, fear, happy, neutral, sad, and surprise. It consists of 10 female subjects performing different facial expressions with a total of 213 images.

3.2 Experimental results on CK+, RFD, and JAFFE database

Performance of SVM-RFE selected features in terms of different performance metrics classified using OVO linear SVM classifier [21] on CK+, JAFFE, and RFD databases have been shown in Table 1-Table 4.

Table 1. Performance analysis with different feature size on CK+ 7 database

No. of Features	Avg. Acc.	Avg. Prec.	Avg. Recall	Avg. F1-Score
33	92.38	91.59	92.02	91.73
66	95.05	94.40	94.55	94.45
98	96.56	96.07	96.33	96.14
129	97.05	96.78	96.85	96.81
159	97.79	97.46	97.60	97.51
188	98.77	98.49	98.72	98.59
218	98.28	97.94	98.20	98.03
247	98.27	97.94	98.15	98.02

The experiments were performed using unsigned HOG features extracted using the parameter setting as in [15]: cell size 7 x 7 pixels, block size 2 x 2 with one cell overlap in both horizontal and vertical direction, and number of histogram bins equal to 7. In all the experiments, a 10-fold cross-validation testing procedure has been used wherein, average accuracy (Avg. Acc.), average precision (Avg. Prec.), average recall (Avg. Recall), and average F1-Score (Avg. F1-Score) denotes average score of the 10-folds of these performance metrics.

Table 2. Performance analysis with different feature size on JAFFE 7 database

No. of Features	Avg. Acc.	Avg. Prec.	Avg. Recall	Avg. F1-Score
33	92.49	92.54	92.53	92.47
66	92.49	92.50	92.57	92.48
98	94.84	94.85	94.85	94.83
129	95.31	95.29	95.35	95.30
159	97.65	97.63	97.79	97.67
188	97.18	97.17	97.35	97.22
218	97.18	97.17	97.35	97.22
247	96.71	96.69	96.95	96.76

Table 3. Performance of the proposed approach on RFD 7 category 1 with different feature size

No. of Features	Avg. Acc.	Avg. Prec.	Avg. Recall	Avg. F1-Score
32	93.39	93.39	93.38	93.38
65	95.95	95.95	95.94	95.92
94	95.74	95.74	95.83	95.73
125	96.38	96.38	96.43	96.38
151	95.49	95.52	95.56	95.51
177	97.42	97.44	97.49	97.44
203	98.72	98.72	98.73	98.72
236	98.72	98.72	98.73	98.72

Table 4. Performance of the proposed approach on RFD 7 category 2 with different feature size

No. of Features	Avg. Acc.	Avg. Prec.	Avg. Recall	Avg. F1-Score
32	93.60	93.60	93.67	93.58
63	95.74	95.74	95.76	95.73
94	96.16	96.16	96.26	96.16
125	95.95	95.95	96.01	95.95
149	97.44	97.44	97.55	97.44
179	98.08	98.08	98.16	98.07
209	98.08	98.08	98.16	98.07
235	97.87	97.87	97.88	97.86

From the above tables, we find that with increase in feature dimension, there is an increase in the values of different performance metrics. However, after reaching an optimal feature dimension, the performance starts degrading. This clearly indicates that not all the HOG extracted features are significant. Confusion matrices corresponding to the optimal number of selected features classified using OVO linear SVM classifier has been shown in Table 5-Table 7 for CK+ 7, JAFFE 7, and RFD 7 category 1 databases respectively. As could be seen for CK+ database, the classifier successfully classified all the expression images except the anger and neutral class. Moreover, in case of JAFFE database, some of the images from the disgust class got misclassified into sad and fear class. Also, in case of RFD, the classifier performed well in classifying most of the expressions except surprise and anger.

Table 5. Confusion matrix with optimal selected features on CK+ 7 database

	An	Di	Ha	Ne	Sa	Su	Fe
An	**91.11**	4.44	0	4.44	0	0	0
Di	0	**100**	0	0	0	0	0
Ha	0	0	**100**	0	0	0	0
Ne	1.67	0	0	**98.33**	0	0	0
Sa	0	0	0	0	**100**	0	0
Su	0	0	0	0	0	**100**	0
Fe	0	0	0	0	0	0	**100**

Table 6. Confusion matrix with optimal selected features on JAFFE 7 database

	An	Di	Ha	Ne	Sa	Su	Fe
An	**100**	0	0	0	0	0	0
Di	0	**93.10**	0	0	3.44	0	3.44
Ha	0	0	**100**	0	0	0	0
Ne	0	0	0	**100**	0	0	0
Sa	0	0	3.33	0	**96.77**	0	0
Su	0	0	3.33	0	0	**96.67**	0
Fe	0	0	0	0	3.12	0	**96.88**

Table 7. Confusion matrix with optimal selected features on RFD 7 category 1 database

	An	Co	Di	Fe	Ha	Sa	Su
An	**97.01**	0	0	1.49	1.49	0	0
Co	0	**100**	0	0	0	0	0
Di	0	0	**100**	0	0	0	0
Fe	0	0	0	**100**	0	0	0
Ha	0	0	0	0	**98.51**	0	1.49
Sa	0	0	0	0	0	**98.51**	1.49
Su	0	0	0	0	1.49	1.49	**97.01**

Performance comparison of the proposed approach on different FER datasets with optimal unsigned HOG features has been shown in Table 8. Also, experimental evaluation on these databases using the signed version of the HOG descriptor is listed in Table 9. The unsigned features contain orientation bins evenly spaced over 0^0-180^0 whereas in the case of signed features the range is 0^0-360^0 .Comparing these two tables one can find that, on CK+ and JAFFE database both signed and unsigned version of the HOG descriptor performed equally well. However, on RFD database signed HOG descriptor has a lead.

Table 8. Performance of the proposed approach using unsigned HOG features

Database	Optimal Feature Size	Avg. Acc.	Avg. Prec.	Avg. Recall	Avg. F1-Score
CK+ 6	85	99.71	99.63	99.72	99.67
CK+ 7	188	98.77	98.49	98.72	98.59
RFD 7 category 1	203	98.72	98.72	98.72	98.72
RFD 7 category 2	179	98.08	98.08	98.16	98.07
RFD 8	172	94.22	94.22	94.33	94.18
JAFFE 6	83	97.81	97.79	97.95	97.82
JAFFE 7	153	97.65	97.63	97.79	97.67

Table 9. Performance of the proposed approach using signed HOG features

Database	Optimal Feature Size	Avg. Acc.	Avg. Prec.	Avg. Recall	Avg. F1-Score
CK+ 6	131	99.71	99.72	99.64	99.67
CK+ 7	208	99.02	98.92	99.03	98.97
RFD 7 category 1	147	98.93	98.93	98.95	98.94
RFD 7 category 2	150	98.29	98.29	98.32	98.29
RFD 8	194	97.95	97.95	97.93	97.93
JAFFE 6	100	97.27	97.24	97.38	97.27
JAFFE 7	223	97.18	97.20	97.25	97.21

Performance comparison of the proposed approach with different state-of-the-art approaches having similar database distribution and testing procedure has been shown in Table 10. From the table, we find that the proposed approach attained similar/better performance compared to different approaches available in the literature.

4 Conclusion

In this paper, we presented an efficient algorithmic pipeline for FER system. The algorithmic pipeline consists of a face detection unit, face alignment & registration unit followed by features extraction, feature selection and classification units. Signed and unsigned versions of the HOG descriptor have been used for extracting features from the facial image of size 65 x 59 pixels. Feature selection using SVM-RFE has been employed to select significant features from the high dimensional HOG features. Finally, a multi-class SVM classifier has been used to classify the selected features into their respective expression categories. After performing a significant number of experiments, we found that using only the SVM-RFE selected features, the proposed approach achieved state-of-the-art results on CK+, JAFFE, and RFD FER databases with reduced processing time, computational resources, and memory requirements, as there is multi-fold reduction in the feature dimension compared to the original feature dimension. Thus, the proposed approach could be effectively used for real-time implementation of a FER system on an embedded platform.

Table 10. Comparison of the proposed approach with different state-of-the-art approaches

References	Database	Avg. Prec.	Avg. Recall	Avg. Acc.	Avg. F1-Score
[5]	JAFFE 6	92.63	91.80	91.80	92.22
Proposed	JAFFE 6	97.79	97.95	97.81	97.82
Proposed	JAFFE 7	97.63	97.79	97.65	97.67
[8]	CK+ 6	99.40	94.10	94.10	94.10
[15]	CK+ 6	95.80	95.90	98.80	95.80
[5]	CK+ 6	94.69	94.10	94.09	94.39
[2]	CK+ 6	----	----	95.50	----
Proposed	CK+ 6	99.63	99.72	99.71	99.67
[2]	CK+ 7	----	----	93.40	----
[15]	CK+ 7	94.30	94.10	98.50	94.10
Proposed	CK+ 7	98.92	99.03	99.02	98.97
[15]	RFD 7 category 1	94.90	94.90	98.50	94.80
Proposed	RFD 7 category 1	98.93	98.95	98.93	98.94
Proposed	RFD 7 category 2	98.29	98.32	98.29	98.29
[15]	RFD 8	93.00	92.90	98.20	92.90
Proposed	RFD 8	97.95	97.93	97.95	97.93

References

1. Knapp, Mark L., Judith A. Hall, and Terrence G. Horgan. Nonverbal communication in human interaction. Cengage Learning, 2013.
2. Shan, Caifeng, Shaogang Gong, and Peter W. McOwan. "Facial expression recognition based on local binary patterns: A comprehensive study." Image and Vision Computing 27.6 (2009): 803-816.
3. Zhao, Xiaoming, and Shiqing Zhang. "Facial expression recognition based on local binary patterns and kernel discriminant isomap." Sensors 11.10 (2011): 9573-9588.
4. Zhao, Xiaoming, and Shiqing Zhang. "Facial expression recognition using local binary patterns and discriminant kernel locally linear embedding." EURASIP journal on Advances in signal processing 2012.1 (2012): 20.
5. Happy, S. L., and Aurobinda Routray. "Automatic facial expression recognition using features of salient facial patches" IEEE transactions on Affective Computing 6.1 (2015): 1-12.
6. Tan, Xiaoyang, and Bill Triggs. "Enhanced local texture feature sets for face recognition under difficult lighting conditions." IEEE transactions on image processing 19.6 (2010): 1635-1650.