# Continuous Emotion Recognition: Sparsity Analysis

**Neeru Rathee**

**Abstract** Continuous emotion recognition is the key concern of researchers working in the field of facial behavior analysis and human–computer interaction. An attempt for continuous emotion recognition is made along with the sparsity analysis. In the presented work, Gabor filters are used to extract features from facial images. Before applying Gabor filters, preprocessing is done on facial images so as to reduce the variance due to illumination, scaling, and rotation. The Gabor magnitude as well as Gabor phase is used to represent facial features. The features were applied to relevance vector regression for continuous emotion recognition. The sparsity analysis is done by analyzing the support vectors to ensure its application in real time. The proposed approach was evaluated on extended Cohn-Kanade database. The results represent the efficacy of the presented approach.

**Keywords** Relevance vector regression · Emotion recognition · Facial expression analysis · Gabor filters · Sparsity

## 1 Introduction

Emotion results in changes in facial expressions corresponding to the different moods. Ekman and Friesen [5] defined six basic emotions: anger, fear, disgust, sad, surprise, and happy. Earlier facial expression recognition was a key concern of researchers, but the pioneering work of [2] introduced facial expressions as a challenge to the researcher working in the field of computer vision.

Continuous emotion recognition can be achieved either directly (recognizing the emotion directly using facial features) or indirectly (first recognizing the facial action unit and then recognizing the emotion). The accuracy of both the methods is controversial due to the dependence of emotion recognition methods on several

N. Rathee (✉)
Maharaja Surajmal Institute of Technology, (affiliated to Guru Gobind Singh Indraprastha University), Janakpuri C-4, New Delhi, India
e-mail: neeru1rathee@gmail.com



P.K. Sa et al. (eds.), *Progress in Intelligent Computing Techniques: Theory, Practice, and Applications*, Advances in Intelligent Systems and Computing 518,
DOI 10.1007/978-981-10-3373-5_26


**Fig. 1** Block diagram representation of continuous emotion recognition

factors such as feature extraction method, preprocessing applied, and the classification algorithm.

The emotion recognition approaches proposed so far cater for the emotion recognition accuracy by presenting the simulation results and have not commented on the sparsity of the data. The sparsity can be roughly estimated by computing the relevance or support vectors used by the regression method (used for continuous emotion estimation). The larger the vectors, the lesser is the sparsity, and ultimately, it will lead to large computation load. In the presented approach, sparsity analysis is presented for continuous emotion recognition.

In the presented approach, face is extracted from the given image frames using the approach mentioned in [13]. The extracted face images are preprocessed so that the effect of illumination variation and scaling may be reduced. The preprocessed images are applied to Gabor filters to extract features. The features so extracted are applied to relevance vector regression (RVR) for continuous emotion recognition. The motivation behind using RVR is the use of relevance vector machine for facial expression recognition [3].

The organization of the paper is as follows: The preprocessing step is briefly explained in Sect. 2. Section 3 gives details of Gabor feature extraction. The basic RVR is described in brief in Sect. 4. The database used to evaluate the proposed approach is mentioned in Sect. 5. The experimental results are discussed in briefly in Sect. 6. Section 7 lists conclusion and future scope. The block diagram of the proposed approach is shown in Fig. 1.

## 2 Preprocessing

Emotion recognition methods are highly dependent on facial feature extraction methods. The face images are extracted from the given image by computing the integral image from the given image and then adopting the method introduced by [12]. The two important tasks performed in preprocessing include histogram equalization and scaling. Histogram equalization helps in removing the effects of non-uniform illumination by normalizing the contrast to the complete range of available pixels. Due to varying distance between subject and camera, and varying image resolution, the actual information is suppressed. To remove these effects, scaling is done and the images are resized to a fix size.

## 3 Gabor Feature Extraction

The pixel-based features or texture features are extracted by many techniques. The most popular techniques include LBP, HOG, and Gabor wavelets. The Gabor features are extracted by applying a bank of filters to an image. To extract the detailed information, features were extracted at 8 different orientations and 5 spatial frequencies being frequently used for extraction of texture features. To extract the texture information, filter bank with different characteristic frequencies and orientations is implemented for feature extraction. This resulted in decomposition of an image [9], which is popularly used for facial feature extraction [7].

## 4 Relevance Vector Regression

Relevance vector regression, proposed by Tipping [11], is applied for continuous emotion recognition. The linear model in feature space is given by

$$f(x,w) = \sum_{n=1}^{N} w_n.K(x,x_n) + w_0 \tag{1}$$

where $w_0$ represents the bias term and $K(x,x_n)$, $n = 1, 2.....N$ represents a set of non-linear transformations. RVR uses a loss function $L(t, f(x,w)$ that is used to measure the quality of the estimation. RVR uses a loss function called $\varepsilon$-insensitive loss function.

$$L(t, f(x,w) = \begin{cases} 0 & p = |t - f(x,w)| - \varepsilon \leq 0 \\ |t - f(x,w)| & otherwise \end{cases}$$

RVR uses $\varepsilon$-insensitive loss to perform linear regression in the high-dimensional feature space while reducing model complexity by minimizing $w^2|$. This is done by introducing slack variables $\xi_n$, $\xi_n^*$, $n = 1, 2, ...N$ to measure the deviation of training samples outside the $\varepsilon$-sensitive zone. So, RVR is formulated as minimization of the following function:

$$min\ \frac{1}{2}||w||^2 + C\sum_{n=1}^{N}(\xi_n + \xi_n^*) \tag{2}$$

subjected to the condition

$$\begin{cases} t_n - f(x_n,w) \leq \varepsilon + \xi_n^* \\ f(x_n,w) - t_n \leq \varepsilon + \xi_n \\ \xi_n, \xi_n^* \geq 0, n = 1, 2, .....N \end{cases}$$

The constant $C > 0$ determines the trade-off between the flatness of $f$ and the values up to which deviations greater than $\varepsilon$ are tolerated.

## 5 Database

We experimented with the Extended Cohn-Kanade (CK+) [6] dataset as this dataset is the most frequently used in the literature. The CK+ dataset contains 327 image sequences of seven expressions: happy, sad, fear, surprise, disgust, surprise, and contempt, performed by 118 subjects. Recordings in this dataset end at the apex of the expression. We have used only six basic expressions defined by [4] and had not included images of 'contempt' expression in our experiments. The first frame of every sequence represents a neutral expression and last frame represents emotion expression. The regression analysis is done on the complete sequence of an emotion.

## 6 Experimental Results

The facial images were preprocessed and Gabor filters were applied to extract facial features. To remove the redundant data, independent component analysis (ICA) was applied on the extracted features. The ICA mapped data was applied to RVR and SVR for training. The three most popular kernels have been adopted in our experiments: radial basis function kernel, linear kernel, and polynomial kernel.

To train the RVR, cross-validation (CV) is adopted due to availability of smaller data. The two main types of cross-validation are leave one subject out CV and tenfold CV. In leave one subject out CV, features of one subjects are used for testing and rest are used for training. In tenfold CV, the whole data is distributed in 10 parts equally. Out of 10 parts, nine parts are used for training and one for testing. A set of ten experiments is repeated to compute tenfold CV [1]. The performance of the proposed approach is also affected by image size. In preprocessing, images were rescaled to different sizes, which resulted in variation of performance of the proposed approach.

Performance criterion: The measures that have been used for evaluating the proposed approach include mean squared error (MSE), number of relevance vectors, and correlation coefficient (Corr). MSE represents the difference between actual and expected values. To avoid this, squared correlation coefficient was used as a measure of performance. Correlation coefficient refers to the degree of similarity between two datasets. Sparsity refers to the nonzero values that represent the sparsity of the data. The number of relevance vectors represents the sparsity of the data. For an efficient system, MSE should be low, Corr should be high, and number of support vectors should be low.

The Gabor features were extracted after resizing the image to three different sizes and performance was computed for three different feature sets. The results are represented in Table 1.

**Table 1** Performance evaluation of RVR on different image sizes

| Sr.no. | Image size | MSE | Correlation | Number of relevance vectors |
|---|---|---|---|---|
| 1 | 70 × 70 | 1.132 | 0.754 | 385 |
| 2 | 80 × 80 | 1.041 | 0.773 | 296 |
| 3 | 90 × 90 | 1.184 | 0.753 | 319 |

**Table 2** Performance evaluation of the SVR on different image sizes

| Sr.no. | Image size | MSE | Correlation | Number of support vectors |
|---|---|---|---|---|
| 1 | 70 × 70 | 1.112 | 0.779 | 435 |
| 2 | 80 × 80 | 1.001 | 0.792 | 308 |
| 3 | 90 × 90 | 1.165 | 0.760 | 382 |

From the above table, it may be concluded that image size 80 × 80 is the optimum size. The number of relevance vectors is directly related to sparsity of the data so the sparsity analysis along with other measures for support vector regression was also computed and presented in Table 2.

The experimental results represent that the number of support is minimally required for 80 × 80 resolution. Along with this, MSE is minimum and correlation is maximum at this particular resolution.

## 7 Conclusion

An approach for continuous emotion recognition is presented by extracting Gabor features from facial images. To make the proposed system applicable in real time, sparsity analysis is also presented by computing the number of support vectors during regression. The two regression techniques, SVR and RVR, have been explored for sparsity analysis. Along with sparsity, accuracy is also compared for both the techniques at various resolutions of images. The experimental results shows that accuracy of both the techniques is almost the same. The optimum resolution is 80 × 80 in terms of MSE, correlation coefficient, and number of support vectors for RVR. In the presented approach, only Gabor features have been explored. The combination of features may result in better accuracy. Moreover, the proposed approach may be evaluated on recent popular database to prove its robustness.

## References

1. Ahonen, T., Hadid, A., Pietikainen, M.: Face description with local binary patterns: Application to face recognition. Pattern Analysis and Machine Intelligence, IEEE Transactions on 28(12), 2037–2041 (2006)
2. Darwin, C.: The expression of the emotions in man and animals. Oxford University Press (1998)
3. Datcu, D., Rothkrantz, L.J.: Facial expression recognition with relevance vector machines. In: Multimedia and Expo, 2005. ICME 2005. IEEE International Conference on. pp. 193–196 (2005)
4. Ekman, P., Friesen, W.: Facial Action Coding System: A Technique for the Measurement of Facial Movement. Consulting Psychologists Press, Palo Alto (1978)
5. Ekman, P., Friesen, W.V.: Measuring facial movement. Environmental psychology and nonverbal behavior 1(1), 56–75 (1976)
6. Kanade, T., Cohn, J.F., Tian, Y.: Comprehensive database for facial expression analysis. In: Automatic Face and Gesture Recognition, 2000. Proceedings. Fourth IEEE International Conference on. pp. 46–53 (2000)
7. Li, Y., Mavadati, S.M., Mahoor, M.H., Zhao, Y., Ji, Q.: Measuring the intensity of spontaneous facial action units with dynamic bayesian network. Pattern Recognition 48(11), 3417–3427 (2015)
8. Littlewort, G., Bartlett, M.S., Fasel, I.R., Susskind, J., Movellan, J.R.: Dynamics of facial expression extracted automatically from video. Image and Vision Computing 24, 615–625 (2006)
9. Tian, Y.l., Kanade, T., Cohn, J.F.: Evaluation of Gabor-wavelet-based facial action unit recognition in image sequences of increasing complexity. In: Automatic Face and Gesture Recognition, 2002. Proceedings. Fifth IEEE International Conference on. pp. 229–234. IEEE (2002)
10. Tipping, M., Bishop, C.: Variational relevance vector machine (Apr 12 2005), US Patent 6,879,944
11. Tipping, M.E.: Sparse bayesian learning and the relevance vector machine. The journal of machine learning research 1, 211–244 (2001)
12. Viola, P., Jones, M.: Robust Real-time Object Detection. International Journal of Computer Vision (2001)
13. Viola, P.A., Jones, M.J.: Rapid Object Detection using a Boosted Cascade of Simple Features. In: Computer Vision and Pattern Recognition. vol. 1, pp. 511–518 (2001)