Facial Expression Recognition using Learning Vector Quantization

Gert-Jan de Vries[1,2], Steffen Pauws[1], and Michael Biehl[2]

[1] Philips Research - Healthcare, High Tech Campus 34, 5656 AE Eindhoven, The Netherlands
[2] Johann Bernoulli Institute for Mathematics and Computer Science, University of Groningen, The Netherlands

Abstract. Although the detection of emotions from facial video or images has been topic of intense research for several years, the set of applied classification techniques seems limited to a few popular methods. Benchmark datasets facilitate direct comparison of methods. We used one such dataset, the Cohn-Kanade database, to build classifiers for facial expression recognition based upon Local Binary Patterns (LBP) features. We are interested in the application of Learning Vector Quantization (LVQ) classifiers to this classification task. These prototype-based classifiers allow to inspect of prototypical features of the emotion classes, are conceptually intuitive and quick to train. For comparison we also consider Support Vector Machine (SVM) and observe that LVQ performances exceed those reported in literature for methods based upon LBP features and are amongst the overall top performing methods. Most prominent features were found to originate, primarily, from the mouth region and eye regions. Finally, we explored the specific LBP features that were found most influential within these regions.

1 Introduction

Many applications, especially in human-computer interaction can benefit from facial expression recognition of its users [15], ranging from affective content selection to adaptive system behavior to the affective state of the user. For many of these applications, users have restricted range of motion, enabling the recognition of affect through unobtrusive measurement using video or photo cameras. Examples of such systems range from affective music players [29] to intelligent car safety systems [10] and air traffic control [14]. Although performances of facial expression recognizers are already high, there is still room for improvement, which will optimize user satisfaction and allow for integration of such systems in daily life.

As addressed by Van den Broek et al. [2], publicly available data sets that can be used as benchmarks are scarce in affective computing. For facial emotion recognition, however, such benchmark databases are available. Kanade, Cohn and Tian published a "Comprehensive Database for Facial Expression Analysis" in 2000 [7], later known as the Cohn-Kanade database. It consists of image

sequences displaying the faces of participants who were instructed to show a range of "facial displays" consisting of at least one Action Unit (AU) [3]. The participants were university students between 18 and 50 years of age, 69% of them female, and represented of a mix of ethnicities. The image sequences are labeled per still image with AUs that are active, which can be translated to emotion labels using a set of rules provided by Ekman, Friesen and Hager [3]. For 100 of the participants at least one of the prototypic emotions (Anger, Disgust, Fear, Joy, Sadness, and Surprise) has been recorded and can be used for the classification of emotions from facial expressions.

We observe that Support Vector Machine (SVM) is a very popular technique applied at this boundary of affective computing and computer vision. While SVMs have been applied successfully to various classification tasks, there are various reasons to investigate how alternative classification methods perform as affective classifiers. Learning Vector Quantization (LVQ) methods have been successfully applied in many settings [12], including facial expression recognition on other datasets [1], but to the best of our knowledge, not to the task of recognizing facial expressions from the Cohn-Kanade database. This type of classifier has several benefits, such as low computational complexity resulting in fast training times, conceptually intuitive nature, and possibility to inspect relevant features without performing additional analyses. In order to put our work into perspective, we performed a comprehensive literature review of methods applied to the Cohn-Kanade database which will be treated in the next section. After that, we present the methods used for our affective classifiers and results obtained. Finally, we present a discussion and conclusion.

2   Cohn-Kanade database

The Cohn-Kanade database has been widely used to develop and validate techniques for facial emotion recognition. To obtain an overview of techniques and their performances, we have searched the literature systematically using Web of Science [21]. The search terms "Cohn AND Kanade" resulted in 153 publications. We selected 43 publications for full analysis by excluding e.g., those using temporal information (video). In these papers we identified 199 classification schemes for 6 or 7 emotion classes, which were trained using the Cohn-Kanade database [7] and for which a performance was reported.

We applied the following criteria for further filtering: accuracy of a model should be reported; and it should be validated using data from at least 50 participants, which is half of the available participants in the Cohn-Kanade database. These criteria were satisfied by 96 models, of which Table 1 shows a summary. First of all, it shows that the task of classifying unseen faces, using per person (pp)-cross-validation, is more difficult than classifying unseen instances of known faces (using cross-validation). Most studies concern the 7-class problem that considers the expressions Anger (A), Disgust (D), Fear (F), Joy (J), Sadness (Sa), Surprise (Su) and a Neutral (N) expression. On average, the cross-validation performances reach 90.75% accuracy; exceptionally high performance

is reported [27] when using ensembles of SVMs: 99.40% accuracy, 3% more than the second best published result.

Table 1. Meta analysis of 150 models from literature.

Nr. of Classes	Classes[1]	Validation method	Nr. of models	Accuracy		
				min	mean	max
7	A,D,F,J,N,Sa,Su	cross	31	78.90%	90.75%	99.40%
7	A,D,F,J,N,Sa,Su	pp-cross	26	73.40%	85.94%	94.88%
6	A,D,F,J,Sa,Su	cross	20	82.52%	89.21%	96.70%
6	A,D,F,J,Sa,Su	pp-cross	17	76.12%	86.54%	96.40%
6	A,D,F,J,Sa,Su	single split	2	83.05%	87.43%	91.81%

[1] Abbreviations used: A(nger), D(isgust), F(ear), J(oy), (N)eutral, (Sa)dness, (Su)rprise

We focus on the most difficult cross validation type (pp-cross-validation) that assesses the performance on classifying emotions from unseen faces. Table 2 shows the 26 classifiers that are published for the 7-class problem, showing that accuracy ranges from 73.40% [19] to 94.88% [28]. The most frequently used feature-type is Local Binary Patterns (LBP) followed by Gabor features. Highest performance is obtained by methods that use projections of the original images to some lower dimensional space, such as KDIsomap, Linear Discriminant Analysis (LDA) and Principal Component Analysis (PCA). Slightly inferior are the methods based upon feature extraction such as Local Directional Patterns (LDP) and LBP and less successful are methods based upon Gabor features. The most popular classification techniques are SVM and LDA.

In this work, we will explore the application of LVQ classifiers to this classification problem and will use the open box nature of these classifiers to gain more insight into the classification problem. We will use LBP-features because they have been used most often, can be obtained relatively efficiently and have been demonstrated to give good performance.

3   Methods

From the Cohn-Kanade database we selected 310 image sequences, coming from 95 subjects, that could be labeled as one of the emotions Anger, Disgust, Fear, Joy, Sadness or Surprise. For each sequence, the neutral face and three peak frames, i.e., those with highest emotional intensity, were used for emotional expression recognition. Following Shan et al. [19, 5] and Tian [22], we used the distance between manually annotated location of the eyes to rotate, crop and scale the images to 108x147 pixels, which were used as input to the further preprocessing. First, the images were rotated to ensure horizontal alignment of the eyes. The distance between the eyes (d_{eyes}) was determined and then the images were cropped such that they measured $2d_{eyes}$ by $3d_{eyes}$, and finally they were resized to 108x147 pixels.

Table 2. Literature overview of studies that classify 7 emotion classes using the Cohn-Kanade database and validated using participant wise cross validation, grouped by feature type.

Reference	Features	Classifier type	Accuracy	#pp used in validation	#images used in validation
Zhao et al., 2011[28]	KDIsomap	SVM	94.88%	96	1409
Zhao et al., 2011[28]	KIsomap	SVM	75.81%	96	1409
Zhao et al., 2011[28]	KLDA	SVM	93.32%	96	1409
Zhao et al., 2011[28]	LDA	SVM	90.18%	96	1409
Zhao et al., 2011[28]	KPCA	SVM	92.59%	96	1409
Zhao et al., 2011[28]	PCA	SVM	92.43%	96	1409
Jabid et al., 2010[6]	LDP	SVM	93.40%	96	1632
Jabid et al., 2010[6]	LDP	Template matching	86.90%	96	1632
Jabid et al., 2010[6]	LBP	SVM	88.90%	96	1632
Shan et al., 2009[19]	LBP	SVM	88.90%	96	1280
Lajevardi & Hussain, 2010[9]	LBP	LDA	88.40%	100	?
Zavaschi et al., 2013[27]	LBP	SVM	84.30%	100	1281
Shan et al., 2009[19]	LBP	Linear programming	82.30%	96	1280
Jabid et al., 2010[6]	LBP	Template matching	79.10%	96	1632
Shan et al., 2009[19]	LBP	Template matching	79.10%	96	1280
Shan et al., 2009[19]	LBP	LDA	73.40%	96	1280
Shan et al., 2009[19]	LBP	LDA&ANN	73.40%	96	1280
Zavaschi et al., 2013[27]	LBP&Gabor	SVM ensemble	88.90%	100	1281
Zavaschi et al., 2013[27]	LBP&Gabor	SVM	79.20%	100	1281
Lajevardi & Hussain, 2010[9]	HLACLF	LDA	91.60%	100	?
Lajevardi & Hussain, 2010[9]	HLAC	LDA	89.90%	100	?
Lajevardi & Hussain, 2010[9]	Gabor	LDA	89.70%	100	?
Jabid et al., 2010[6]	Gabor	SVM	86.80%	96	1632
Shan et al., 2009[19]	Gabor	SVM	86.80%	96	1280
Lu et al., 2006[11]	Gabor	NKFDA	85.59%	93	≤ 651
Zavaschi et al., 2013[27]	Gabor	SVM	78.70%	100	1281

As discussed in the previous section, the most frequently used feature type is LBP, which is also our choice. We derived the LBP-features from the scaled images in the following way:

Per grey valued pixel (i_c) the LBP-value is calculated by comparing the pixel to its eight neighbors, resulting in a binary string of which the decimal value is taken, according to:

$$LBP(x_c, y_c) = \sum_{n=0}^{7} s(i_n - i_c)\, 2^n \tag{1}$$

where s is the Heaviside step function. The $2^8 - 1$ possible outcomes are reduced to $L = 59$ by regarding only those LBP values with at most 2 bitwise transitions when considered as a circular pattern, as proposed by Ojala et al. [13]. The patterns in this subset are termed "Uniform" patterns and represent bright and dark spots, corners and edges.

The images are divided into $(6*7 = 42)$ regions R_j of size 18x21 pixels, where per region a histogram $H_i = \sum_{x,y} \delta_{LBP(x,y),i}$ (with $(x,y) \in R_j, i = 0, \ldots, L-1$) is built. These histograms are placed next to each other, forming a single vector of length $N = 42 * 59 = 2478$.

Learning Vector Quantization (LVQ) comprises a family of classifiers that is of open box nature, that is, they provide direct insight into the information learned by the classifier. LVQ, initially proposed by Kohonen [8], defines prototypes in the same (mathematical) space as the data to represent the classes. These prototypes are directly interpretable as they show characteristics of classes in terms of the features chosen. During training, samples are presented sequentially, and for each sample the closest prototypes are updated by moving them towards or away from the presented sample. Several variants have been proposed, amongst which Robust Soft Learning Vector Quantization (RSLVQ) [18]. It introduces soft prototype assignments which corresponds to a soft window around the decision boundary [26]. In addition, we apply SVM [23], the most popular technique in this domain of computer vision applications. Next to the "linear SVM", i.e., the large margin perceptron classifier [24], we also applied SVM with a Radial Basis Function (RBF) kernel, which however, did not improve upon the results. For further reference, we also include a baseline classifier that always assigns the label corresponding to the class with the highest prior probability.

Validation was performed using 10x10-fold participant-wise cross validation, i.e., 10 repetitions of randomly chosen 10-fold cross validation, where participants are strictly separated in training and test data. In this way, the performances obtained reflect the generalization performances to unseen participants. We applied the classifiers to the 7 class facial expression recognition task, compared their generalization performances, and inspected the confusion matrices. Finally, we inspected the prototypes trained by RSLVQ, with specific attention for the relevances it (implicitly) assigns to the features. To this end, we considered differences between the 'Neutral' prototype and other prototypes.

4 Results

The results, given in Table 3 show that our classifiers reach over 91% accuracy. RSLVQ competes well with SVM, the latter reaching insignificantly better performance, especially when using the LBP features with overlapping regions. On the other hand, the results over the 10 times 10-folds are slightly more stable for RSLVQ.

Table 3. Generalization performances on 10x10 fold participant-wise cross validation per subtask.

	Method	Accuracy
7 class	Baseline	21.6%
	RSLVQ	$91.2\% \pm 0.5\%$
	SVM	$91.4\% \pm 0.5\%$

Confusion matrices of SVM and RSLVQ are available in Tables 4 and 5. Differences between the confusions made by both classifiers are small and for both we observe that most errors correspond to misclassifying various emotions as 'Neutral'. This might suggest that the classifiers have most difficulty with low-intensity instances of emotions (other than Neutral) while the emotions themselves are quite well separable. Most difficult emotions are Fear, of which 13% is misclassified as Joy, and Anger, which is often confused with Neutral and Sadness.

Table 4. Confusion matrix (averaged over 10x10-fold cross validation) for 7class classification by SVM. Entries are percentages per actual emotion.

Actual \ Predicted	A	D	F	J	N	Sa	Su
Anger	78.8	3.3	0	0.1	10.7	7.1	0
Disgust	3.6	90.3	0	0	2.3	3.7	0
Fear	0	0	78.6	12.6	6.8	1.9	0.1
Joy	0	0	0.2	99	0.9	0	0
Neutral	0.5	0	0	1	94.5	2.7	1.3
Sadness	2.6	0.3	0	0	9	85.6	2.5
Surprise	0	0	0	0	1	0	99

In order to inspect the (implicit) relevances assigned by RSLVQ, we summed up all absolute pairwise differences between the prototype representing 'Neutral' faces and the other emotions. The difference vector of two prototypes corresponds to the direction in feature space along which the two classes are discriminated. The absolute value of its components can be interpreted as to measure the relevance of the corresponding feature. Figure 1 shows this information aggregated per region as used in the building of the LBP histograms. It indicates that most

Table 5. Confusion matrix (averaged over 10x10-fold cross validation) for 7class classification by RSLVQ. Entries are percentages per actual emotion.

Actual \ Predicted	A	D	F	J	N	Sa	Su
Anger	81.8	5.9	0	0.6	7.9	3.7	0
Disgust	2.1	93.3	0	0	1.8	2.8	0
Fear	1.2	0.4	79.4	13.5	2.3	2.2	1
Joy	0.4	0	0.7	97.7	1.3	0	0
Neutral	1	0.2	0	1.5	93.2	2.4	1.8
Sadness	3.9	0.8	0	0.3	7.8	84	3.2
Surprise	0	0	0	0	1.4	0	98.6

informative to the classifier are the regions around the mouth, followed by the eyes and eye-brows.
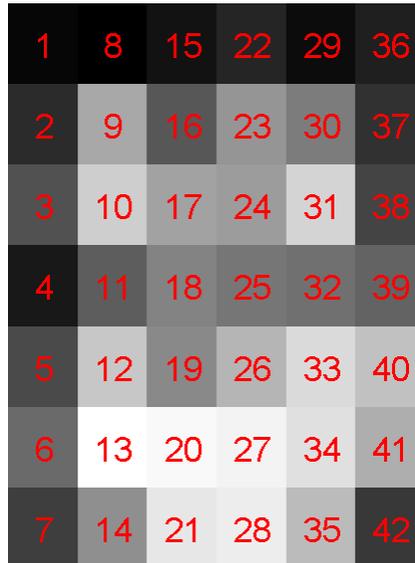
Fig. 1. Relevance of image regions in the RSLVQ classifier (white levels indicate relevance).

The feature vectors we used represent the frequency of observing certain textural elements within 42 different regions of the face. Figure 2 shows the LBP-features linked to the 48 most relevant histogram entries. We see that, out of the 42 regions, the regions around the mouth are best represented. Regions 20 and 27 represent the (upper side of the) mouth and the LBP-features represented in the top 48 indicate the importance of textural components that are lighter at the top than on the bottom. Similarly, regions 21 and 28 represent the chin and lower side

of the mouth, from which LBP-features that indicate lighter bottom and darker top are present. Finally, we observe that regions 13 and 34, corresponding to the left and right side of the mouth, are mostly represented by textural components that have darker right and left sides, respectively. These observations seem to indicate that opening of the mouth, which is accompanied by dark pixels in the center of the mouth-region, is the most important distinction between various emotions.
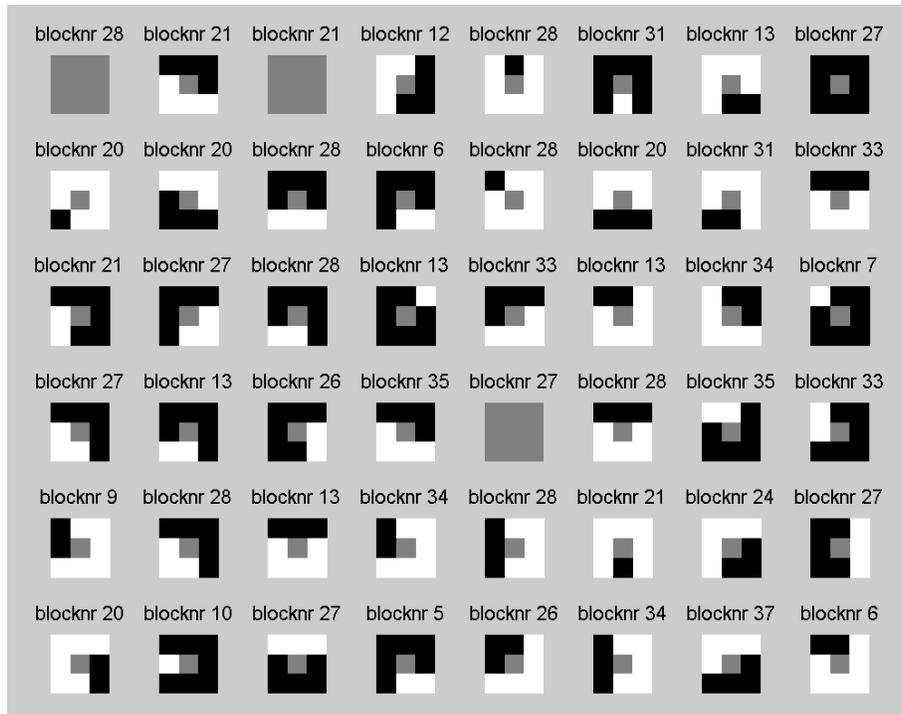

Fig. 2. Top 48 most relevant LBP-features used the RSLVQ classifier; ordered from let to right, top to bottom. The block numbers refer to the regions as numbered in Figure 1.

Figure 3 shows the aggregated relevances per region for each of the emotions in isolation, i.e., representing the difference to the 'Neutral' emotion. We observe that the relevances for Surprise are quite distributed, and more expressive around the central mouth regions and chin. Sadness shows even higher relevance of the chin areas; Joy is most different from Neutral in the outer and upper mouth regions, while Fear differs in the outer and central mouth regions. Finally, the relevances of Anger and Disgust are more scattered, but in comparison to the other emotions have relatively high contributions of the features from the eyes, brows and forehead.
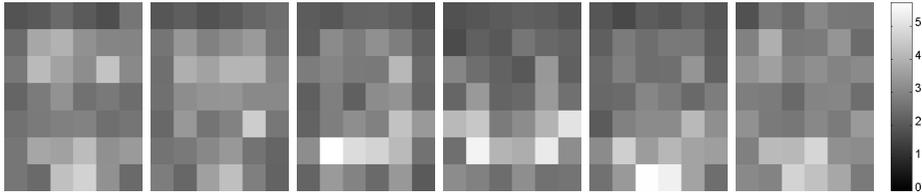
Fig. 3. Relevance of image regions in the RSLVQ classifier for Anger, Disgust, Fear, Joy, Sadness and Surprise (from left to right).

5   Discussion

The results we obtained show high accuracy on the tasks of classifying facial expressions, represented as LBP feature vectors, into emotions. We used two different representations, one using non-overlapping regions in the facial pictures, the other using overlapping regions. The non-overlapping regions yield more intuitively interpretable feature vectors, while the overlapping regions contain more information, but also increase the dimensionality of the feature vectors almost by a factor 4. For the 7-class classification task, which includes the neutral face as a class, our classifiers reached an accuracy of 91.4% for SVM and 91.2% for RSLVQ. Four techniques [28] from one paper show better performances when using SVM and different feature sets. In comparison to the methods that use LBP features, our classifiers perform better by 3.3 percentage points.

Despite the marginally lower performance of RSLVQ compared to SVM, the prototype based method might still be preferable because of significantly lower computational complexity due to the inherent ability to handle multi-class classification problems while SVM requires to train many more classifiers in an ensemble scheme, i.e., k classifiers in a one-vs-all or $k(k-1)/2$ in a one-vs-one scheme for k-class classification. Especially when considering parameter optimization, this computational advantage might be of interest.

The prototype based classifiers we used enabled us to inspect the prototypes and infer which features are considered most relevant by the classifiers. The mouth region turned out to be most influential. Within this region the LBP features that correspond to various mouth openings were most important. While eye-brows are known to be activated in many different emotions [4], and they are found to be the prominent facial elements to highlight prosody [20], our results suggest that for automated facial expression recognition, the mouth-region is more important. The regions representing the eye-brows and forehead, however, do help our classifier in distinguishing especially Anger and Disgust from the other emotions. Shan and colleagues [19] used AdaBoost in combination with pattern matching to determine the most influential LBP histograms from an exhaustive set of 16640 facial regions and identified most discriminant regions around the eyes and mouth. With our approach, we obtained these indications of relevance directly from the trained classifier, rather than performing additional and computationally intensive analyses.

We also observed that not only the occurrence frequencies of uniform LBP features are relevant for the classification, but also the frequency of non-uniform patterns, which were joined together in one bin in each histogram representing a photo region, were represented in the list of most influential features. Moreover, the 1st, 3rd and 29th most influential features were such non-uniform patterns. On the other hand, the use of uniform patterns rather than all LBPs reduced the feature space with more than a factor 4 and helps keeping the search space manageable.

6   Conclusion

We have performed a comprehensive literature overview of attempts to classify facial expressions from the Cohn-Kanade database and observed that generalization performances on the 7 class task average at 85.9%. Maximum reported accuracies on this task were 94.9%. While being the most popular, or at least most frequently used, type of features, LBP features reached only up to 88.9%.

To the best of our knowledge, we have applied LVQ classifiers for the first time to the task of facial expression recognition using the Cohn-Kanade database. The generalization accuracies obtained (91.2% for 7-class classification) show that RSLVQ is among the most successful classifiers overall and outperforms all reported efforts using LBP features. As a reference we used SVM, which showed even slightly better performances (91.4% for 7-class) but, in contrast to RSLVQ, does not allow for direct inspection of the knowledge learned and used by the classifiers and requires significantly more time to train and optimize. By inspecting the prototypes trained by RSLVQ we noticed that the most prominent features originate from the mouth region, followed by the eye-regions. The specific LBP features that are used most prominently by the classifier confirm that mouth opening/closing is discriminative for various emotions.

In the present work, we have used implicit relevances obtained from difference vectors of RSLVQ prototypes. Other LVQ variants can be designed that explicitly train relevance vectors along with prototypes; examples are Generalized Matrix Learning Vector Quantization (GMLVQ)[16] and Matrix Robust Soft Learning Vector Quantization (MRSLVQ)[17]. Future work includes the application of such methods to the challenge of facial expression recognition. Another interesting future extension of the current work is to observe how our methods perform on spontaneous emotions. Although being more challenging, recent developments[25] indicate that results obtained in one setting can be transferred successfully to the other. Finally, the literature review we performed indicates that performances might be further improved by considering different feature sets such as LDP or Scale-Invariant Feature Transform (SIFT).

The high performances obtained indicate that implementation in consumer products becomes more and more feasible. Natural choices of first applications include real time behavior adaptation of laptops or tablets to their users' emotions. By, for example, being able to distinguish frustration from happiness, human-computer interaction can be greatly improved because it allows for de-

tection of suboptimal interactions and adapt at real time by offering alternative actions when frustration is detected. The flexibility of LVQ to directly handle multi-class classification (i.e., without requiring classification schemes such as 'one-vs-all' that are needed by binary classifiers), allows for quick training times and opens up the ability to train user specifics and personalize the model by learning at real time. Such personalized systems should be able to obtain even better performances for facial expression recognition.

Acknowledgements

References

1. Bashyal, S., Venayagamoorthy, G.K.: Recognition of facial expressions using Gabor wavelets and learning vector quantization. Engineering Applications of Artificial Intelligence 21(7), 1056–1064 (Oct 2008)
2. van den Broek, E., Janssen, J., Westerink, J., Healey, J.: Prerequisites for affective signal processing (asp). In: Encarnao, P., Veloso, A. (eds.) Biosignals 2009: Proceedings of the Second International Conference on Bio-Inspired Systems and Signal Processing. pp. 426–433. INSTICC Press, Portugal (2009)
3. Ekman, P., Friesen, W.V., Hager, J.C.: Facial Action Coding System [E-book]. Research Nexus, Salt Lake City, UT (2002)
4. Ekman, P.: About brows: Emotional and conversational signals. In: von Cranach, M., Foppa, K., Lepenies, W., Ploog, D. (eds.) Human Ethology: Claims and limits of a new discipline, pp. 169–248. Cambridge University Press (1979)
5. Gritti, T., Shan, C., Jeanne, V., Braspenning, R.: Local features based facial expression recognition with face registration errors. In: 8th IEEE International Conference on Automatic Face Gesture Recognition, 2008. FG '08. pp. 1–8 (2008)
6. Jabid, T., Kabir, M., Chae, O.: Facial expression recognition using local directional pattern (LDP). In: 2010 17th IEEE International Conference on Image Processing (ICIP). pp. 1605–1608 (2010)
7. Kanade, T., Cohn, J., Tian, Y.: Comprehensive database for facial expression analysis. In: Fourth IEEE International Conference on Automatic Face and Gesture Recognition, 2000. Proceedings. pp. 46–53 (2000)
8. Kohonen, T.: Improved versions of learning vector quantization. In: International Joint Conference on Neural Networks. vol. 1, pp. 545–550. IEEE (1990)
9. Lajevardi, S.M., Hussain, Z.M.: Novel higher-order local autocorrelation-like feature extraction methodology for facial expression recognition. IET Image Processing 4(2), 114–119 (2010)
10. Lisetti, C.L., Nasoz, F.: Affective intelligent car interfaces with emotion recognition. In: Proceedings of the 11th International Conference on Human Computer Interaction. p. 41. ACM, Las Vegas, USA (2005)

11. Lu, H., Wang, Z., Liu, X.: Facial expression recognition using NKFDA method with gabor features. In: The Sixth World Congress on Intelligent Control and Automation, 2006. WCICA 2006. vol. 2, pp. 9902–9906 (2006)
12. Neural Networks Research Centre, Helsinki: Bibliography on the self-organizing maps (SOM) and learning vector quantization (LVQ). Otaniemi: Helsinki Univ. of Technology. Available on-line: `http://liinwww.ira.uka.de/bibliography/Neural/SOM.LVQ.html` (2002)
13. Ojala, T., Pietikinen, M., Menp, T.: Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Transactions on Pattern Analysis and Machine Intelligence 24(7), 971–987 (2002)
14. Pantic, M., Sebe, N., Cohn, J.F., Huang, T.: Affective multimodal human-computer interaction. In: Proceedings of the 13th Annual ACM International Conference on Multimedia. pp. 669–676. MULTIMEDIA '05, ACM, New York, NY, USA (2005)
15. Peter, C., Beale, R.: Affect and Emotion in Human-Computer Interaction - From Theory to Applications. No. 4868 in Lecture Notes in Computer Science, Springer, Berlin, Heidelberg (2008)
16. Schneider, P., Biehl, M., Hammer, B.: Adaptive relevance matrices in learning vector quantization. Neural Computation 21(12), 3532–3561 (2009)
17. Schneider, P., Biehl, M., Hammer, B.: Distance learning in discriminative vector quantization. Neural computation 21(10), 2942–2969 (2009)
18. Seo, S., Obermayer, K.: Soft learning vector quantization. Neural Computation 15, 1589–1604 (2003)
19. Shan, C., Gong, S., McOwan, P.W.: Facial expression recognition based on local binary patterns: A comprehensive study. Image and Vision Computing 27(6), 803–816 (2009)
20. Swerts, M., Krahmer, E.: Facial expression and prosodic prominence: Effects of modality and facial area. Journal of Phonetics 36(2), 219–238 (2008)
21. Thomson Reuters: Web of science (2014), http://apps.webofknowledge.com/
22. Tian, Y.L.: Evaluation of face resolution for expression analysis. In: Conference on Computer Vision and Pattern Recognition Workshop, 2004. CVPRW '04. pp. 82–82 (2004)
23. Vapnik, V.: Statistical learning theory. Wiley (1998)
24. Vapnik, V.N.: The Nature of Statistical Learning Theory. Springer-Verlag New York, Inc., New York, NY, USA (1995)
25. Wan, S., Aggarwal, J.K.: Spontaneous facial expression recognition: A robust metric learning approach. Pattern Recognition 47(5), 1859–1868 (2014)
26. Witoelar, A.W., Ghosh, A., de Vries, J.J.G., Hammer, B., Biehl, M.: Window-based example selection in learning vector quantization. Neural Computation 22(11), 2924–2961 (2011)
27. Zavaschi, T.H., Britto Jr., A.S., Oliveira, L.E., Koerich, A.L.: Fusion of feature sets and classifiers for facial expression recognition. Expert Systems with Applications 40(2), 646–655 (2013)
28. Zhao, X., Zhang, S.: Facial expression recognition based on local binary patterns and kernel discriminant isomap. Sensors (Basel, Switzerland) 11(10), 9573–9588 (2011)
29. van der Zwaag, M.D., Janssen, J.H., Westerink, J.H.: Directing physiology and mood through music: Validation of an affective music player. IEEE Transactions on Affective Computing 4(1), 57–68 (2012)