Material based on: de Vries, J. J. G., Pauws, S. C. and Biehl, M.: submitted, Facial expression recognition using learning vector quantization, Pattern Recognition

Hoofdstuk 5
EMOTION FROM A FACIAL PERSPECTIVE

Abstract

The detection of emotions from facial video or images has been topic of research for several years, nevertheless 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 are found most influential within these regions.

5.1 Introduction

In human history, facial expressions have grown as important element of inter-human communication. Especially since mankind developed social emotions, estimated to date back to 2 million years ago at the time of the early *Homo Erectus* (Dubreuil 2010), the face became the primary means for emotional expression. Many applications, especially in human-computer interaction can benefit from facial expression recognition of its users (Peter and Beale 2008), ranging from affective content selection to adaptive system behavior to the affective state of the user. For many of such applications, users have restricted range of motion, enabling the measurement of affect through unobtrusive measurement using video or photo cameras. Examples of such systems range from affective music players (van der Zwaag et al. 2012) to intelligent car safety systems (Lisetti and Nasoz 2005) and air traffic control (Pantic et al. 2005).

Various studies have been performed using data obtained through video, using either still images, or a temporal sequence of images; with varying levels of success. Table 5.1 shows a brief summary of ten emotion classification studies from video. It can be seen that various parts of the body have been used to derive features from, that the number of features used varies heavily and performance ranges from 20% to 98% on various tasks. With exception of the 20% performance by Cottrell and Metcalfe (1991), the performances can be considered relatively high. Because the number of classes, number of participants, prior probability of classes and methods used for validation vary between these studies, their performance cannot be compared directly. In order to overcome this problem we choose a standard database for facial emotion recognition enabling us to benchmark our classifier performance.

As addressed by van den Broek et al. (2009), 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 Analysisïn 2000 (Kanade et al. 2000), 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). 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 et al. (2002). 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 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. LVQ methods have been successfully applied in many settings (Neural Networks Research Centre, Helsinki 2002), 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

Table 5.1: *Review of ten machine learning studies employing facial characteristics to recognize emotions.*

Reference	Input[1]	Ss[2]	Feat.[3]	Technique	Targets	Perf[4]
Cottrell & Metcalfe (1991)	Face [S]	20	4096	ANN	8 emotions	20%
Essa & Pentland (1995; 1997)	FACS [ST]	8			5 emotions	98%
Yacoob & Davis (1996)	Face [ST]	32	16		7 emotions	70%
Lien et al. (2000)	FACS [ST]	100	38	LDA, HMM	9 action units	80%
Cohen et al. (2003)	Motion Units [ST]	53	12	BN	7 emotions	83%
Pantic & Patras (2006)	FACS [ST]	19	24		27 action units	87%
Littlewort et al. (2006)	Face [S]	100	900	LDA, SVM	7 emotions	93%
Gunes & Piccardi (2009)	Head, hands, body [ST]	10	172	DT, BN, SVM, ANN, AdaBoost	12 emotions	85%
Sanchez et al. (2010)	Face [ST]	52	84	SVM	6 emotions	83%
Xiao et al. (2011)	Face [S]	$\leq$100	4320	kNN, SVM	6 emotions	97%

[1] Abbreviations used: S(patial), T(emporal)
[2] Number of subjects
[3] Number of features
[4] Performance (accuracy)

discussion and conclusion.

5.2 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

Table 5.2: *Meta analysis of 150 models from literature.*

Nr. of Classes	Classes[1]	Validation method	Number of models	Accuracy min	mean	max
7	A,D,F,J,N,Sa,Su	cross-validation	31	78.90%	90.75%	99.40%
7	A,D,F,J,N,Sa,Su	pp-cross-validation	26	73.40%	85.94%	94.88%
6	A,D,F,J,Sa,Su	cross-validation	20	82.52%	89.21%	96.70%
6	A,D,F,J,Sa,Su	pp-cross-validation	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

(Thomson Reuters 2014). 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 (Kanade et al. 2000) 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 5.2 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). Many studies concern the 6-class problem, considering the expressions Anger (A), Disgust (D), Fear (F), Joy (J), Sadness (Sa), and Surprise (Su). Alternatively, a 7-class problem which also includes a Neutral (N) expression has been investigated in a majority of studies. With exception of one study (Zavaschi et al. 2013) which reports exceptionally high performance of ensembles of SVMs: 99.40% accuracy, 3% more than the second best published result, the latter appears more difficult when judged by maximum performance. With respect to mean performances, however, this does not hold. This might be explained by the majority of research focussing on the 7-class problem.

Table 5.3: *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.*

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

Table 5.4: *Literature overview of studies that classify 6 emotion classes using the Cohn-Kanade database and validated using participant wise cross validation, grouped by feature type.*

Citation	Features	Classifier type	Accuracy	#pp used in validation	#images used in validation
Jabid et al. (2010b)	LDP	SVM	96.40%	96	1224
Jabid et al. (2010b)	LDP	Template matching	89.20%	96	1224
Li et al. (2009)	SIFT	SVM	96.33%	90	300
Jabid et al. (2010b)	LBP	SVM	92.60%	96	1224
Shan et al. (2009)	LBP	SVM	92.60%	96	960
Shan et al. (2009)	LBP	Linear programming	89.60%	96	960
Jabid et al. (2010b)	LBP	Template matching	84.50%	96	1224
Shan et al. (2009)	LBP	Template matching	84.50%	96	960
Shan et al. (2009)	LBP	LDA	79.20%	96	960
Shan et al. (2009)	LBP	LDA&ANN	79.20%	96	960
Jabid et al. (2010b)	Gabor	SVM	89.80%	96	1224
Shan et al. (2009)	Gabor	SVM	89.80%	96	960
Fazli et al. (2009)	Gabor & PCA&LDA	PNN	89.00%	70	192
Wang and Yin (2007)	facial feature points	LDA	82.68%	53	864
Wang and Yin (2007)	facial feature points	QDC	81.96%	53	864
Wang and Yin (2007)	facial feature points	SVC	77.68%	53	864
Wang and Yin (2007)	facial feature points	NBN	76.12%	53	864

We focus on the most difficult cross validation type (pp-cross-validation) that assesses the performance on classifying emotions from unseen faces. Table 5.3 shows the 26 classifiers that are published for the 7-class problem, showing that accuracy ranges between 73.40% (Shan et al. 2009) and 94.88% (Zhao and Zhang 2011). The most used feature-type is LBP followed by Gabor features. Highest performance is obtained by methods that use (non-)linear projections of the original images to some lower dimensional space, such as KDIsomap, LDA and PCA. Slightly inferior are the methods based upon feature extraction such as LDP and LBP and less successful are methods based upon Gabor features. The most popular classification techniques are SVM and LDA.

Table 5.4 shows the 16 classifiers found for the 6-class problem showing performances ranging from 76.12% (Wang and Yin 2007) to 96.40% (Jabid et al. 2010b). Similar trends as for the 7-class problem can be observed where LBP features are most

used. LDP features reach highest performance closely followed by Scale-Invariant Feature Transform (SIFT) features (Li et al. 2009) and LBP. Again, Gabor features perform worse, followed by facial features points; and SVM dominates among the types of classifier used.

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.

Figure 5.1: *Examples of cropped images (Kanade et al. 2000) representing (from left to right) Anger, Disgust, Fear, Joy, Sadness and Surprise.*

5.3 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. (Shan et al. 2009, Gritti et al. 2008) and Tian (2004), 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 $2d_{\text{eyes}}$ by $3d_{\text{eyes}}$, and finally they were resized to 108x147 pixels. Figure 5.1 shows examples of resized images from several participants. As discussed in the previous section, the most used feature type is LBP, which is the one of 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:

$$LBP(x_c, y_c) = \sum_{n=0}^{7} s(i_n - i_c)\, 2^n \tag{5.1}$$

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. (2002). The patterns in this subset are termed "Uniform"patterns and represent bright and dark spots, corners and edges.

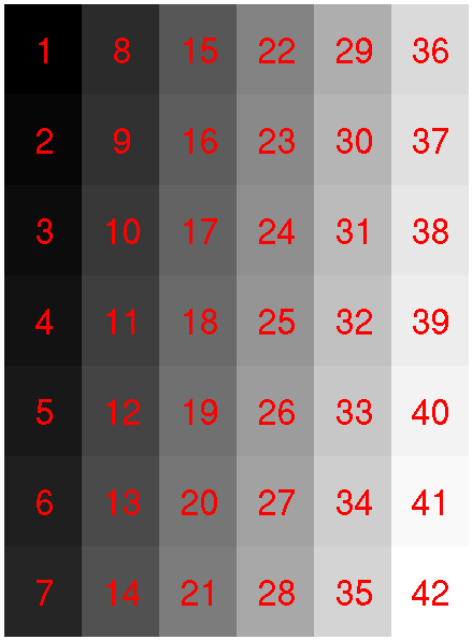

Figure 5.2: *Subdivision of an image into 42 regions. The numbers indicate the ordering used in the concatenation of the feature vectors.*

The images are divided into $(6 * 7 = 42)$ regions R_j of size 18x21 pixels, as indicated by Figure 5.2, 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 $N = 42 * 59 = 2478$. Another variant uses

$11*13 = 143$ overlapping regions, resulting in a vector of length $N = 143*59 = 8437$.

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 both 6- and 7-class facial expression recognition tasks, compared their generalization performances, and inspected the confusion matrices. Finally, we inspected the prototypes trained by Robust Soft Learning Vector Quantization (RSLVQ), with specific attention for the relevances it (implicitly) assigns to the features. To this end, we inspected differences between the 'Neutral' prototype and other prototypes.

5.4 Results

The results, given in Table 5.5 and 5.6 show that for the subtask of classifying 7 emotion classes, our classifiers reach over 91% accuracy. RSLVQ competes well with SVM, the latter reaching slightly higher 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. For the 6-class classification task, we observe a similar pattern: very good performance of RSLVQ (93.3%) and even slightly better accuracy by SVM (up to 94.5%). Tables 5.7 and 5.8 report the training performances obtained for the 7-class and 6-class tasks, respectively. We observe that both SVM and RSLVQ manage to achieve perfect training performance.

Table 5.5: *Test performances on* 10×10 *fold participant-wise cross validation on the 7-class facial expression data.*

	LBP overlap		LBP no overlap	
Method	Hyper-parameter	Accuracy	Hyper-parameter	Accuracy
Baseline		21.6%		21.6%
kNN	$k = 11$	$73.8\% \pm 0.7\%$	$k = 11$	$72.2\% \pm 0.7\%$
SVM - linear	$C = 0.1$	$92.2\% \pm 0.5\%$	$C = 0.1$	$91.4\% \pm 0.5\%$
Means		$83.1\% \pm 0.6\%$		$82.8\% \pm 0.4\%$
GLVQ		$84.0\% \pm 0.3\%$		$82.9\% \pm 0.4\%$
RSLVQ	$v_{soft} = 5$	$91.3\% \pm 0.3\%$	$v_{soft} = 0.6$	$91.2\% \pm 0.5\%$

Table 5.6: *Test performances on* 10×10 *fold participant-wise cross validation on the 6-class facial expression data.*

	LBP overlap		LBP no overlap	
Method	Hyper-parameter	Accuracy	Hyper-parameter	Accuracy
Baseline		26.2%		26.2%
kNN	$k = 11$	$79.6\% \pm 0.7\%$	$k = 11$	$78.7\% \pm 0.7\%$
SVM - linear	$C = 0.01$	$94.5\% \pm 0.6\%$	$C = 0.1$	$94.0\% \pm 0.4\%$
Means		$88.5\% \pm 0.4\%$		$87.2\% \pm 0.3\%$
GLVQ		$85.4\% \pm 0.4\%$		$84.5\% \pm 0.3\%$
RSLVQ	$v_{soft} = 5$	$93.2\% \pm 0.5\%$	$v_{soft} = 0.9$	$93.3\% \pm 0.2\%$

Table 5.7: *Training performances on* 10×10 *fold participant-wise cross validation on the 7-class facial expression data.*

	LBP overlap		LBP no overlap	
Method	Hyper-parameter	Accuracy	Hyper-parameter	Accuracy
Baseline		21.6%		21.6%
kNN	$k = 11$	$87.7\% \pm 0.0\%$	$k = 11$	$88.0\% \pm 0.0\%$
SVM - linear	$C = 0.1$	$100.0\% \pm 0.0\%$	$C = 0.1$	$100.0\% \pm 0.0\%$
Means		$90.9\% \pm 0.1\%$		$90.3\% \pm 0.1\%$
GLVQ		$93.9\% \pm 0.1\%$		$93.3\% \pm 0.1\%$
RSLVQ	$v_{soft} = 5$	$100.0\% \pm 0.0\%$	$v_{soft} = 0.6$	$100.0\% \pm 0.0\%$

Table 5.8: *Training performances on* 10×10 *fold participant-wise cross validation on the 6-class facial expression data.*

	LBP overlap		LBP no overlap	
Method	Hyper-parameter	Accuracy	Hyper-parameter	Accuracy
Baseline		26.2%		26.2%
kNN	$k = 11$	$87.3\% \pm 0.1\%$	$k = 11$	$87.9\% \pm 0.0\%$
SVM - linear	$C = 0.01$	$100.0\% \pm 0.0\%$	$C = 0.1$	$100.0\% \pm 0.0\%$
Means		$93.7\% \pm 0.1\%$		$93.1\% \pm 0.1\%$
GLVQ		$93.5\% \pm 0.1\%$		$93.6\% \pm 0.0\%$
RSLVQ	$v_{soft} = 5$	$100.0\% \pm 0.0\%$	$v_{soft} = 0.9$	$100.0\% \pm 0.0\%$

Confusion matrices of SVM and RSLVQ are available in Tables 5.9 and 5.10. 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 mistaken with Neutral and Sadness. We also inspected the confusion matrices for 6-class classification (see Tables 5.11 and 5.12) and noticed that the misclassifications as Neutral are mainly compensated by increased class-wise accuracy of Sadness and slight increases for the other emotions, of which Surprise can be detected flawlessly.

In order to inspect the (implicit) relevances assigned by RSLVQ, we summed up

Table 5.9: *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

Table 5.10: *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

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 5.3 shows this information aggregated per region (as used in the building of the LBP histograms). It indicates that most informative to the classifier are the regions around the mouth, followed by the eyes and eye-brows.

The feature vectors we used represent the frequency of observing certain textural elements within 42 different regions of the face. Figure 5.4 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 lower side of the mouth, from which LBP-

Table 5.11: *Confusion matrix (averaged over 10x10-fold cross validation) for 6class classification by SVM. Entries are percentages per actual emotion.*

Actual \ Predicted	A	D	F	J	Sa	Su
Anger	83.7	3.2	0	2.5	10.6	0
Disgust	3.1	93.3	0	0	3.6	0
Fear	0	0	83.5	14	2.3	0.1
Joy	0.3	0	0.2	99.4	0.2	0
Sadness	3.4	0.4	0	0.1	94.1	2
Surprise	0	0	0	0	0	100

Table 5.12: *Confusion matrix (averaged over 10x10-fold cross validation) for 6class classification by RSLVQ. Entries are percentages per actual emotion.*

Actual \ Predicted	A	D	F	J	Sa	Su
Anger	82.5	6	0	2.4	9.1	0.1
Disgust	1	96.4	0	0	2.6	0
Fear	0.3	0	79.9	15.4	2.6	1.8
Joy	0.4	0	0.6	99	0	0
Sadness	3	1	0	0.1	91.8	4.1
Surprise	0	0	0	0	0	100

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.

Figure 5.5 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 Neutal 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.

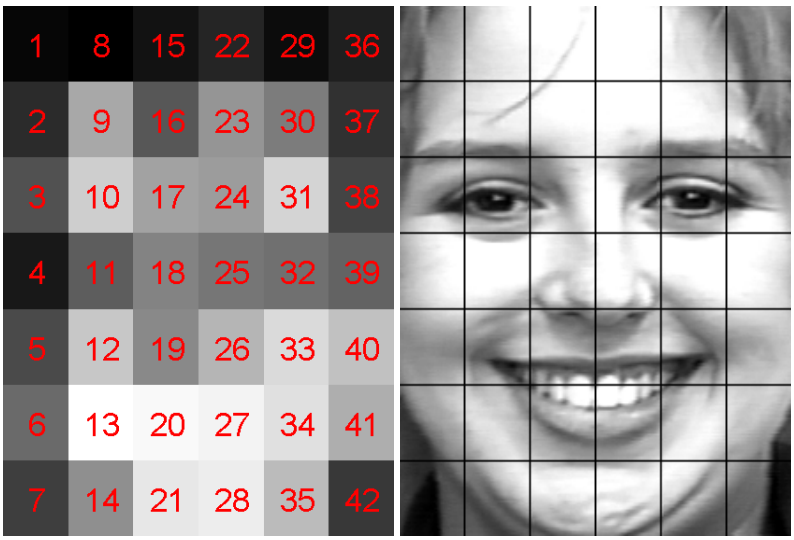

Figure 5.3: *Relevance of image regions in the RSLVQ classifier (left;white levels indicate relevance), and example picture for reference (right).*

5.5 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. Using non-overlapping regions, the 6-class classification task was best performed by SVM with 94.0% accuracy. Second best was RSLVQ with 93.2%. Slightly better performances were obtained using the overlapping regions (maximum accuracy of 94.5%). When comparing these to the performances reported in literature on the same data set and classification task, we observe that there are two attempts that gained better performance. Both use SVM and either use LDP (Jabid et al. 2010b) or SIFT features (Li et al. 2009). The SIFT-based study, however uses only a subset of 300 pictures, indicating that they left out pictures, which might be suitably chosen, rather than the full data set. If we compare our method with other methods using LBP features, we outperform them by 1.9 percentage points.

For the 7-class classification task, which includes the neutral face as a class, our

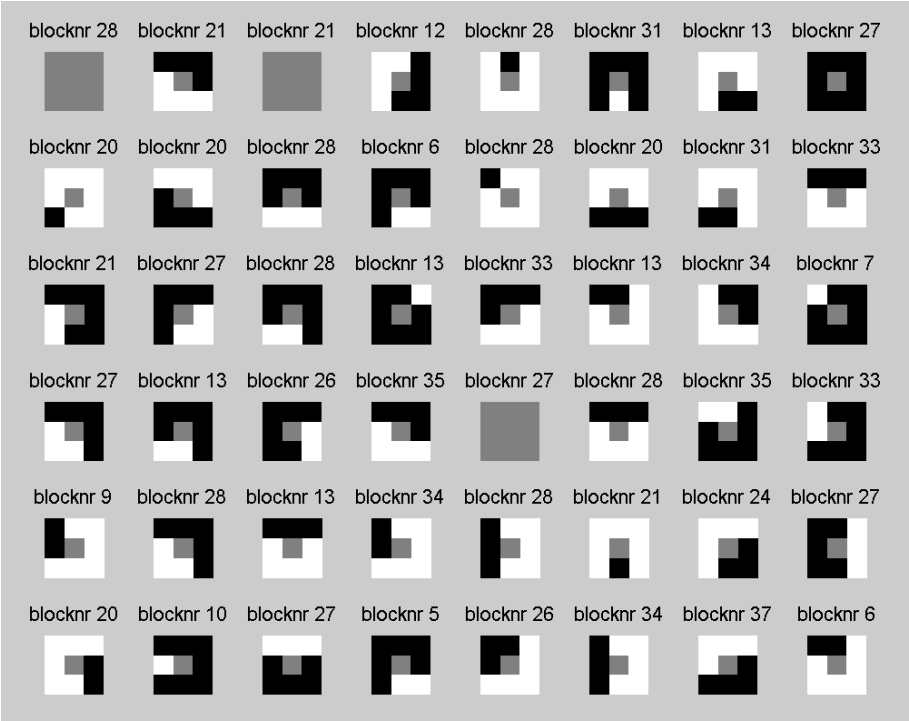

Figure 5.4: *Top 48 most relevant LBP-features used the RSLVQ classifier; ordered from let to right, top to bottom. The block numbers refer to the regions as numbered in Figure 5.3.*

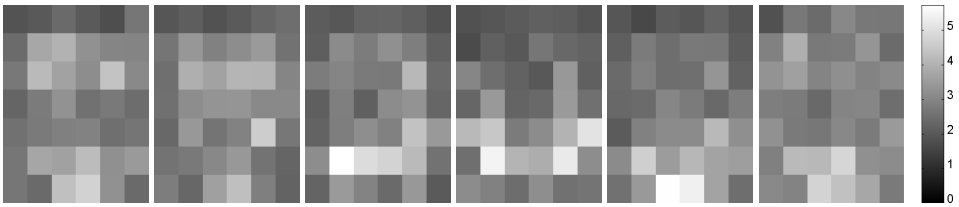

Figure 5.5: *Relevance of image regions in the RSLVQ classifier for Anger, Disgust, Fear, Joy, Sadness and Surprise (from left to right).*

classifiers reached an accuracy of 91.4% for SVM and 90.9% for RSLVQ. Again, slightly better performances were obtained using the overlapping regions (maximum accuracy of 92.2%). Four techniques (Zhao and Zhang 2011) 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.

The prototype based classifiers we used enabled us to inspect the prototypes and infer which features are considered most influential 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 (Ekman 1979), and they are found to be the prominent facial elements to highlight prosody (Swerts and Krahmer 2008), 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 (Shan et al. 2009) 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.

5.6 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 and 6 class tasks average at 85.9% and 86.5%, respectively. Maximum reported accuracies on these tasks were 94.9% and 96.4%. While being the most popular, or at least most frequently used, type of features, LBP features reached only up to 88.9% and 92.6%, respectively for 7 and 6 classes.

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 (91.3% for 7-class, and 93.3% for 6-class classifi-

cation) 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 (92.2% for 7-class, and 94.5% for 6-class classification) but, in contrast to RSLVQ, does not allow for direct inspection of the knowledge learned and used by the classifiers. 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) (Schneider et al. 2009a) and Matrix Robust Soft Learning Vector Quantization (MRSLVQ) (Schneider et al. 2009b). 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 (Wan and Aggarwal 2014) 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 SIFT.

The high performances obtained indicate that implementation in consumer products starts to become 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 detection of suboptimal interactions and adapt at real time by offering alternative actions when frustration is detected. The limited complexity of LVQ, that can 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.