SMILING AND NEUTRAL FACIAL DISPLAY RECOGNITION WITH THE LOCAL BINARY PATTERNS OPERATOR

Karolina Nurzynska*

Institute of Informatics, Silesian University of Technology,

ul. Akademicka 16, 44-100 Gliwice, Poland,

Karolina.Nurzynska@polsl.pl

Bogdan Smolka

Institute of Automatic Control, Silesian University of Technology,

ul. Akademicka 16, 44-100 Gliwice, Poland,

Bogdan.Smolka@polsl.pl

Abstract

The lack of smile as well as its diminished frequency are one of the key factors indicating many kinds of mood disorders (e.g. depression, dysphoria) whereas the difficulty in proper emotion identification is characteristic for people suffering from schizophrenia or autism. Therefore, automatic analysis of emotions by computer systems improves preparing the medical diagnosis but also influences positively the quality of life of those who already suffer.

In this work the problem of automatic classification between smiling and neutral facial display is discussed. Due to the fact that in the literature there are many suggested solutions for emotion recognition based on facial image features, we start this work with an overview, which enables comparison of existing approaches. Next, the problem of adequate image division schema and mask choice is addressed. Since the length of the feature vector causes many problems, this issue is also investigated. The experiments were performed on three databases: *Cohn-Kanade*, *Feret*, and *Combined*, which was specially designed for this research. The obtained results lead to the conclusion that the signs of enjoyment are most accurately classified with application of linear SVM, when the image is described with Local Binary Patterns (LBP) or uniform LBP texture operator. Moreover, applying additional mask shortens the feature vector and keeps the result at the same level. Finally, when Principal Component Analysis (PCA) was used the correct recognition rate slightly diminished by a few percent, but significantly decreased the computational cost.

1 Introduction

Smiling face on the first place is associated with pleasure, comfort, and security, however there are many types of smile, not only these presented to reflect delight.[1] People smile in many social situations just to fit to the expected behaviours, but they also smile to conceal the true emotions like embarrassment, anger, or fear. A number of research in psychology have found that only Duchenne smile (which is characterised by the zygomatic major pulling the lip corners upwards towards the cheekbones, and the orbicularis oculi, which raises the cheek and gathers the skin inwards from around the eye socket) is a sign of enjoyment.[2] Since these emotions are so often expressed in daily life, the changes in its intensity, frequency and ability of recognition are measured as symptoms of many mood disorders, e.g. depression disorders, bipolar disorders, anxiety syndromes, dysphoria. Additionally, for individuals suffering from schizophrenia or autism the correct recognition of others emotion is hampered.

Observation of dynamic changes in facial display is widely used in the medical research concerning depression. The smile intensity and duration is correlated with psychological disorders.[3] There is also a noticeable relationship between non verbal behaviour (like diminishing of head movements and the number of Duchenne's smiles) and severity of depression.[4] Furthermore, the smile attenuation is seen as a key predictor of depression in patients with facial neuromuscular disorders.[5] It is also known that such persons are less likely to express positive affect when presented with positive-mood inductions than non-depressed subjects.[6] Finally, the work[7] stated that only changes in smile are correlated with the mood, while analysing results from cardiovascular measures, reports of emotion brought no conclusions.

It has been found that patients with schizophrenia have problems with correct understanding of emotions presented by others,[8].[9] Additionally autistic adults are impaired on the genuine and fake smile recognition since they tend to omit the eye contact.[10] But not only illness may cause problems with correct emotion understanding. Individuals who easily react on others negative emotions are better in veracity of smile recognition, while those, who tend to be easily contagioned with good emotion make more errors.[11] In case of deception, presenting emotions inconsistent with real feelings aims to create and maintain in another a belief that the communicator himself considers false.[12] Yet, it was proven that offenders smile significantly less frequently when lying.[13] Moreover, the work[14] suggests that controlling one feature in facial expression display (e.g. brow frowning) manifests itself by control of other half of the face, that gives hopes to find means to distinguish the disguise.

Many research, in medicine and psychology, concerning the emotion problem are based on the works of Ekman and Friesen,[15] which resulted in creation of Facial Action Coding System, (FACS). It enables to determine which muscles action is necessary in human face to perform a particular facial gesture. Those muscles sets were enumerated and called Action Units, enabling detailed description of the emotion. In order to annotate gathered experimental data, authors of[4, 6, 14, 16] used certified experts to work over the data, which is time consuming, tiresome, and impossible to use for each patient.

Therefore, the automatic emotion recognition, with special respect for smile detection and specification of its genuineness, is finding a broad application in the medicine. Moreover, the automatic analysis of behaviour is seen as one of the methods available in telemedicine[17] which slowly expands and is believed to be a way to manage problems

of dementia or depression.[18] For instance, *SimSensei* application[19] enables online emotional state analysis, and automatic detection of heart rate from variations in the luminance of someone's face was presented by.[20] Next, there is a method for stroke detection developed, which except of slurring of speech, checks the asymmetry in smile.[21] All those issues may be addressed by a computer software.

This paper addresses the problem of smiling and neutral facial display recognition. Since there are many approaches in the literature which deal with this issue, but are incomparable due to different classification schemas exploited, as well as various data sets used, we start from comparison of existing techniques in order to choose the best classification method. Next, the influence of chosen image division schema and applied mask on the recognition accuracy is investigated. Since the length of calculated feature vector is considerable, the application of principal component analysis (PCA) in order to reduce the feature space is proposed.

The paper is structured as follows. Section 2 gives a brief overview of undertaken works to automatically analyse emotions. Then Section 3 presents chosen texture operators used for facial emotion description. Next, in Section 4 details concerning the experiment set-ups are given and obtained results are described in Section 5. The work is concluded in Section 6.

2 State of the art

Automatic emotion recognition was addressed for many years in the literature. First works concerning this issue started in 80s and used optical flow for muscle movements tracking to recognize presented in the image sequence emotion[22] and was later developed in,[23].[24] Since the movement of lips corners in the direction of eyes is associated with smiling, the active shape models were also used for this emotion recognition.[25]

More recently, the local binary patterns were exploited to describe the facial display and discriminate between basic emotions (like: happiness, sadness, anger, etc.),[26],[27] and similar approaches are presented in,[28],[29].[30] There are also other solutions which apply Gabor filters with GentleSVM,[31] or extreme learning machine as a classifier.[32] Next, not all researchers use the whole face image for this emotion recognition. In some cases only the lips region was analysed,[33][34] and in other the eyes were examined.[35] There is also interesting work which not only concentrates on smile detection, but also tries to evaluate its intensity[36] whereas[37] provides general overview of face emotion description and detection. Finally,[38] compares existing techniques and tries to find out, whether they are sufficiently reliable to work in real life situations. This work points out, that most developed methods for smile classification use for testing databases prepared in laboratory scenarios, which tend to be too superficial. Yet, according to experiments performed with utilization of data collected from the Internet[38] it was concluded that it should be possible to discriminate between emotions, however there is still a lot of room for further work.

3 Image description

Texture operators belong to the group of very popular technique used for the description of image content. These methods calculate special features on the basis of pixel values gathered according to the particular rules. First order

features[39] are derived from the image histogram and describe, between others, its mean, variance, or energy. Second order features are represented by the co-occurrence matrices introduced by Haralick et al.[40] This method computes a set of features, for instance features describing image quality as contrast, dissimilarity, or homogeneity using a square matrix, which contains the number of occurrences of each possible combinations of two grey-scale pixel values. Next, the run-length codes[41] are gathered from the matrix which collects the information about the number of adjoining pixels in the image which have similar intensities. Finally, local binary patterns method,[42] (LBP), is the youngest approach presented in this comparison, however it proved already to be very powerful. Here, for each pixel in the image a code is determined, then all those codes are collected in a form of a histogram. Since it exhibited very good performance in case of emotion recognition,[30] we decided to apply this approach in presented research.

3.1 Local binary patterns

Local binary pattern histogram gathers the information about the number of occurrences of each possible code in the image. The code is calculated for each pixel (see Fig. 1), but the neighbourhood is considered as well:[42–44]

$$\mathrm{LBP}_{P,R}(x_c, y_c) = \sum_{p=0}^{P-1} s(g_p - g_c) \cdot 2^p, \tag{1}$$

where g_c is a grey-scale value of a pixel of interest, which coordinates in the monochromatic image I are (x_c, y_c), g_p $(p = 0, \ldots, P-1)$ is an intensity value of each point in the local processing window, $s(z)$ is a threshold function, which concentrates on the difference sign:

$$s(z) = \begin{cases} 1, & z \geq 0 \\ 0, & z < 0. \end{cases} \tag{2}$$

The points are sampled regularly in the neighbourhood on the circle of radius R:

$$x_p = x_c + R\cos(2\pi p/P), \quad y_p = y_c - R\sin(2\pi p/P), \quad p = 0, \ldots, P-1. \tag{3}$$

Since the coordinates may not be integers, the intensity value is interpolated by a weighted mean of neighbouring pixels intensities.

As the LBP values assigned to the image pixels for $P = 8$ are in the range $[0, 255]$, therefore they can be visualized in form of 8-bit images. Figure 2 depicts exemplary images (a), corresponding maps of LBP values (b) and the LBP histograms (c).

3.2 Uniform patterns

The LBP histogram length is correlated with the sampling points number, P, due to its relationship with the number of histogram bins: 2^P. Therefore, the bigger the number of points in the area, the longer is the histogram, which becomes problematic for classification. Therefore, in the further research[45] the code composition and its influence on the image description was addressed. It was noticed, that when examining the code as a sequence of bits, the

number of transition between 1s and 0s plays an important role. In consequence, the valid binary sequences, called uniform patterns, were defined as those which have up to two bit transitions, eg. 00111000, and for them separate bins are prepared (see Fig. 3). For those sequences with more transitions, for instance 01010101, one collective bin is used. Such an approach shortens the histogram length to $P(P-1)+3$ bins. In literature this procedure is named uniform local binary patterns, ULBP.

3.3 Rotation Invariant LBP

Another approach concentrates on the finding that when the texture rotates, the data in the histogram also moves around the bins,[46].[45] That might be omitted when so-called ROR operator is applied, which treats the LBP code as a bit sequence and searches for such rotation of bits, which results in the smallest numerical code value:

$$\mathrm{RILBP}_{P,R} = \min_j \mathrm{ROR}(\mathrm{LBP}_{P,R}, j), \quad \mathrm{ROR}(\delta, j) : \ \delta(j) = (j-1) \bmod n, \tag{4}$$

and δ is a n bit sequence. In other words, for two sequences 01110000 and 00011100 the smallest rotation invariant code is 00000111 and it is achieved for $j = 4$ and $j = 2$, respectively. More examples are presented in Fig. 4. Application of rotation invariant LBP, (RILBP) enables histogram length shortening to 36 elements for $P = 8$. There are of course other approaches, like this presented in[47] which exploits Zernike moments to improve the performance.

3.4 Rotation Invariant and Uniform LBP

The rotation may also influence uniform patterns, therefore it is not surprising that a rotation invariant version of this operator, named RIULBP, was also suggested. Here, the patters which have up to two bit transitions between 1s and 0s are subject to ROR operator application, which reduces the final histogram length to $P + 2$ bins.

4 Experiment set-up

Before the classification between smiling and neutral facial displays may take place, some preliminary preprocessing is necessary. The visualization of algorithm's steps is depicted in Fig. 5. For the input image the face region is determined using face detector presented in.[48] Then a feature vector describing the face should be calculated. However, computing LBP histogram for whole image is too general. Since it is possible to distinguish which muscles are responsible for the emotion visible on the face,[15] it is obvious that each part of the facial display conveys different information and should be treated separately, when the feature vector is generated. Therefore, several image division schemas and image masks were exploited in our experiments. Applying each of them consists of building the feature vector as a concatenation of LBP histograms calculated for each of its part. Besides mentioned preprocessing, the feature vector keeps immoderately long, therefore the principal component analysis, is applied for decreasing the dimensions of feature space. Finally, the classification process is performed.

4.1 Image division schemas

When considering possible image division schemas, it would be convenient to describe the image in as detailed manner as possible. However, the LBP histogram should be filled densely, what influences the minimal possible size of the sub-image. Moreover, the smaller is the sub-image, the longer gets the final feature vector, which is unwanted characteristic. Therefore, in order to bring together all mentioned demands, considering that the facial display image resolution is 112×150 pixels, following image division schemas were proposed: 4×5, 4×10, and 7×10, which detailed description is presented in Tab. 1.

4.2 Masks

Regardless of the chosen image division schema, we also studied whether it is possible to remove some facial display area from the consideration, without loss of the classification accuracy. In verify this assumption several masks were designed and those most promising are presented in Fig. 6.

4.3 Image databases

In order to evaluate the performance of smile recognition, several databases are exploited in the literature. They vary in image quality, number of subjects, image resolution, etc. Therefore, results achieved on one image set are incomparable with those gathered for the second one. In consequence, in this work all proposed set-ups are rated on three different databases so as to achieve better insight into its performance and accuracy.

Basing on the findings enabling FACS definition, the *Cohn-Kanade* AU-Coded Facial Expression Database,[26,49] was developed, which includes image sequences from neutral to target display of basic emotions (e.g., joy, surprise, anger, fear, disgust).

In the presented work, images from 82 sequences were chosen which depicted the neutral and smiling expression. The image resolution is 640×480 pixels with 8-bit precision for grey-scale values. The images differ in lighting conditions and the subjects do not wear glasses or other covering elements, as well as do not have beards. The same set of subjects represents neutral and smiling expressions. Some examples are presented in Fig. 7(a).

The Defence Advanced Research Project Agency (DARPA) and the National Institute of Standards and Technology (NIST) prepared *Feret* database[50] for facial recognition system evaluation. In our experiments, 62 images were selected in order to represent the neutral and smiling expression. The subjects vary in each group. Moreover, it is possible to have different images for the same person in one group. Figure 7(b) depicts some examples of this database.

In previously described datasets, as well as other considered: *Nottingham originals*, *Iranian*, *Utrecht*, and *Pain*,[51] the number of subjects is small. In order to assure greater diversity of subjects, a collective *Combined* database was created. It gathers images from all the mentioned datasets, that in total gives 712 images equally divided between the two emotion groups.

5 Results and discussion

In order to verify the recognition accuracy of the most popular classifiers, to choose the division schema which is the most beneficial and find out whether it is possible to diminish the feature vector length several experiments were performed.

5.1 Classifier comparison

The aim of this experiment was to compare the accuracy of several commonly used classifiers for face expression recognition. First from the considered was the k-nearest neighbourhood, (kNN), which is an example of the non-parametric pattern recognition method. For a tested object it bases the decision on the majority voting among k closest neighbours from the training set. Various distance metrics were examined: *Euclidean*, *cityblock*, *cosine*, and *correlation*. Moreover, the k parameter varied from 5 to 13 for odd numbers only. Second, was the support vector machine approach, (SVM), which maps the data into higher dimension space, where a planar division between classes is possible. For the calculation various kernels were considered: linear, Gaussian Radial Basis Function with scaling factor taking following values $0.01, 0.25, 0.50, 0.75, 1.00, 2.00$, and polynomial with order of $2, 3, 4$, and 5.[52] Finally, the template matching technique,[30] (TM), was also used. In this case, for each class an average feature vector is calculated and the kNN with $k = 1$ decides to which class the tested object belongs to, exploiting chi square statistics as a distance metric. For all possibilities 10-fold cross-validation approach was applied. The image description used 7×10 division mask. The LBP parameters were following: $R = 1$ and $P = 8$.

Table 2 (section on the top) gathers the best results for *Cohn-Kanade* database achieved for each of the mentioned methods with distinction between applied texture operators. When comparing the results achieved for each image description technique, one can easily notice, that the linear SVM usually overcomes other solutions and the best recognition performance reaches 95% for LBP and ULBP image description techniques. The accuracy gained by the kNN for $k = 9$ is not far behind. Yet, the choice of distance metric influences the result considerably. From those exploited in this research, *cityblock* and *cosine* are worth mentioning for good performance. The TM approach proved to give worse accuracy, however still 93% for LBP operator were recorded. Finally, in all cases the rotation invariant methods efficiency is poor in this application.

Similarly as for previous database, Tab. 2 (section in the middle) collects the best outcome achieved for each type of classifier when *Feret* dataset was tested. The efficiency of presented methods is comparable with previous one, however generally the accuracy is lower. The best score, 88%, was recorded for linear SVM when LBP and ULBP operators were used. Once again, kNN technique with $k = 9$ and *cityblock* distance metric gives good results, however, they are incomparable with SVM ones. The TM method is independent from the texture operator used, but the results have the same accuracy as those of kNN.

Finally, Tab. 2 (section on the bottom) summarises the classification accuracy for the *Combined* database. Here, the linear SVM for LBP operator outperforms other solutions reaching 88.63%. However, ULBP is not far behind - 86.68%. In accordance with previous findings, kNN with $k = 9$ and *cityblock* metric is a little bit less accurate.

TM method provides the worst outcome, especially when the rotation invariant techniques for image description are applied.

From the performed experiment one can see that the best classifier for emotion recognition is linear SVM, which outperforms both other techniques and other SVM kernels considered in this work. As was mentioned, the highest accuracy in case of kNN classifier was recorded for $k = 9$ for the *cityblock* distance metric. However, there were cases where $k = 11$ brought better results, yet the difference was negligible. The TM classifier is the most unstable one. Hence the authors believe that it could overcome other classifiers when optimal parameters are found, however generally it gives the worst performance from all considered techniques.

Four texture operators were compared during this experiment. Regardless the chosen combination, two of them: ULBP and LBP give the best performance. Similar performance of those methods is not surprising, since their definition does not differ significantly. However, when the time and memory consumption plays the most important role, the ULBP should be applied, because it generates much shorter feature vectors. Surprising is the fact, that the rotation invariant version of the image descriptor performance is much worse taking into account that the data never rotates. The unsatisfactory results probably are connected with the reduction of feature vector length. In[52] more databases are compared.

5.2 Image division influence on the performance

The goal of the next experiment was to evaluate the influence of image division schema on the classification performance. Here, three different image division plans were considered from the most general (4×5) by the average (4×10) to the most detailed one (7×10). Since in previous experiment ULBP and LBP image descriptors proved to outperform the others, only these two techniques were considered. Additionally, the correlation between the recognition accuracy and the radius parameter of LBP technique were investigated with constant sampling point value, $P = 8$. The classification was performed with linear SVM method using 10-fold cross-validation approach.

Table 3 gathers results for *Cohn-Kanade*, *Feret*, and *Combined* databases. The differences in classification accuracy between those datasets are similar as in previous experiment. Once again, it is difficult to state which texture operator performs better. Moreover, it is difficult to notice any pattern suggesting the relation between the radius of the LBP operator and the correct classification rate. In case of the image division schema it seems that a little bit better performance is observed for 4×5 and 4×10 when *Cohn-Kanade* database is considered. When one looks at the outcomes of *Feret* database it seems that 4×10 and 7×10 are a little bit more precise. Similar findings are valid, when results for *Combined* database are evaluated.

In consequence of such a small differences between suggested image division approaches, it was decided that the masks should be created for 4×5 division schema, which is characterised by smaller number of sub-images, and 7×10 division schema, which is the most detailed approach.

The second part of this experiments investigates the influence of applied mask on the smiling versus neutral facial display recognition. Five different masks presented in Fig. 6 were considered. The experiments were performed for each database and the results are presented in Tabs. 4 and 5. In order to have better insight into the mask approach, tables presenting classification accuracy when masks for 4×5 and 7×10 image division schema are created, should

be considered together. The best outcome for the *Cohn-Kanade* image database was recorded for 4×5 v2. mask when ULBP operator with $R = 3$ was applied (see Tabs. 4 and 5 - top section). Yet, there are also many other combinations of parameters with very good performance, all of them are marked by bold font. When comparing the results gathered for *Feret* dataset (see Tabs. 4 and 5 - middle section) application of 7×10 v3. mask and LBP operator brings the best results. Similar observation may be drawn from classification accuracy obtained for *Combined* database (Tabs. 4 and 5 - bottom section).

Moreover, it is worth pointing out that application of masks did not diminish the classification performance, however the feature vector length was reduced significantly, as is summarised in Tab. 6. When comparing results presented in Tab. 3 and Tabs. 4 and 5 it is visible that for *Cohn-Kanade* dataset the classification keeps 97.56% of accuracy, for *Feret* case it grows from 88.71% without masks to 89.52%, and for *Combined* database it stays at the range of 89.34%.

5.3 Feature space dimension reduction with PCA

The last experiment addressed the problem of feature vector length shortening with application of PCA. As presented in Tab. 6 the number of histogram elements varies between 590 and 10240. Application of mask removes some sub-images without loss of information, therefore it was interesting to find whether more general rules might be applied. For this experiment two image division schemas were chosen: 4×5 and 7×10 and the most promising mask from previous experiment: 7×10 v3. The LBP parameters were set to $R = 1$ and $P = 8$, while for classification kNN method was applied with *cityblock* distance metric and $k = 9$.

The top section in Tab. 7 gathers results achieved for *Cohn-Kanade* database. Regardless of used texture operator and image description schema, the classification accuracy diminishes by few (up to 4%) percent when feature space reduction was applied. This situation differs reasonably from those recorded for *Feret* and *Combined* databases (middle and bottom section in Tab. 7). Exploiting PCA in classification procedure for *Feret* image set reduced the accuracy up to 19%. However, when 99% of information is exploited to build a feature vector for ULBP operator with mask and LBP operator with 4×5 image division schema the loss of correct classification rate is only 5%. In case of *Combined* database the variation of results is lower, and for 4×5 image division schema and 99% of information used ranges from 3% when ULBP to 5% for LBP operator is used. Although, the deterioration of the general performance is noticeable, it is worth to point out how it influenced the feature vector length (see Tab. 8). The resulting number of elements in feature vector is presented as a percentage in regard of original length, due to its variation between experiments with the same set-ups. When looking at these numbers one can see that for LBP texture operator the reduction is much higher. However, when comparing the number of valid eigenvectors, it becomes similar for those two approaches. Next, though higher reduction of feature space in case of LBP, the classification performance keeps on the same level with ULBP. Finally, it might be concluded that for *Cohn-Kanade* database application of PCA might bring some benefits, due to decrease of the calculation load without big impact on the correct classification rate.

6 Conclusions

This paper presents the research results focused on recognition between smiling and neutral facial display. It addresses the problem of classification technique influence on the recognition accuracy and verifies which of the local binary pattern versions: LBP, ULBP, RILBP, and RIULBP is the most beneficial for image description. Furthermore, the image division schemas are discussed in combination with several masks used to emphasise the most crucial face parts. Finally, obtained feature space dimensions are diminished with application of PCA without dramatic loss of efficiency.

According to achieved results, the linear SVM proved to be the most accurate classifier and LBP and ULBP the most beneficial image descriptors. The correct classification ratio on *Cohn-Kanade* database reached 95% and 89% for *Feret* and *Combined* image data set. The performed experiments revealed that application of the division schema does not influence the results much, however, after mask application it could be concluded that removing some less important part of the image when calculating the feature vector does not diminish the classification performance, but considerably decreases the computational cost. Finally, application of PCA proved that there is a place for considerable feature space dimension reduction, yet it is correlated with worsening of the performance.

This work shown also that it is possible to create an accurate system for automatic recognition of smiling and neutral facial displays in image sequences. That should support and simplify the research concerning mood disorders, but also enable creation of software based tools to support medical treatment of these syndromes. In our further works, we plan to concentrate on the problem of recognition of genuine and posed smile.

Acknowledgement

This work has been supported by the Polish National Science Centre (NCN) under the Grant: DEC-2012/07/B/ST6/01227 and was performed using the infrastructure supported by POIG.02.03.01-24-099/13 grant: GeCONiI - Upper-Silesian Center for Scientific Computation.

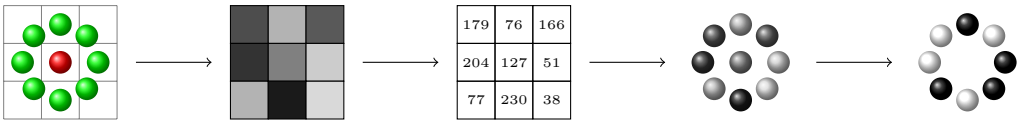


Figure 1: Idea of LBP code calculation for a neighbourhood with parameters: $R = 1$ and $P = 8$. From the left: neighbourhood, pixel grey-scale values in the region, pixel intensities, values in the sampled points, binarized output.

(a) Test images (b) LBP maps (c) Corresponding LBP histograms

Figure 2: Local binary patterns visualization.

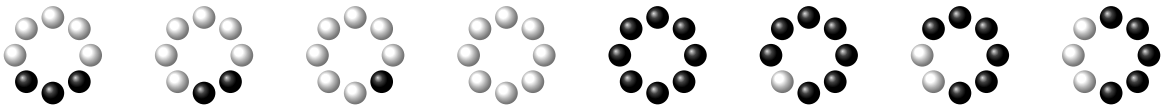


Figure 3: Examples of uniform patterns.

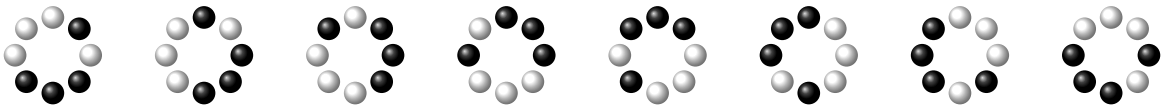


Figure 4: Examples of rotated pattern.

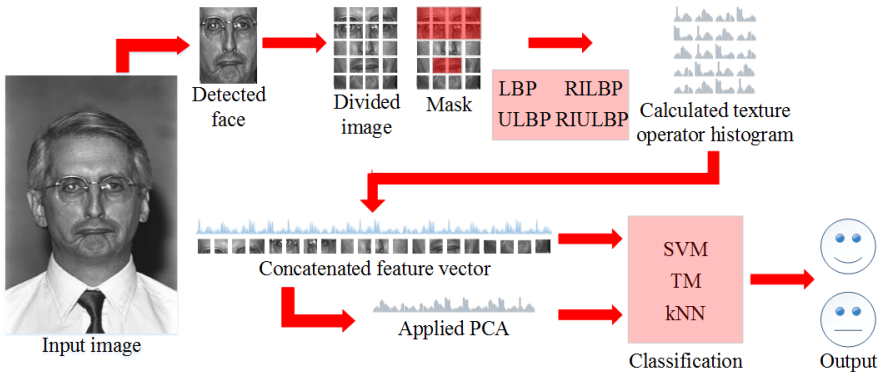
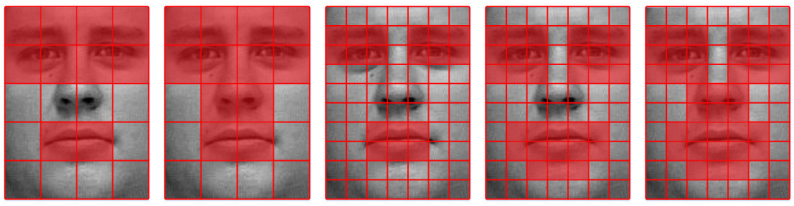


Figure 5: Experiment set-up.



(a) 4x5 v1. (b) 4x5 v2. (c) 7x10 v1. (d) 7x10 v2. (e) 7x10 v3.

Figure 6: Examined masks setups.

(a) *Cohn-Kanade* database. (b) *Feret* database.

Figure 7: Examples of neutral expression are presented in the top row, whereas the smiling ones in the bottom row.

Type	4×5	4×10	7×10
Number of sub-images	20	40	70
Resolution	28×30	28×15	16×15
Elements in histogram	840	420	240

Table 1: Description of image division schemas.

Database	Classifiers	Methods	ULBP	LBP	RIULBP	RILBP
Cohn-Kanade	kNN	Euclidean	90.85	89.63	86.59	86.59
		cityblock	93.29	**95.73**	**87.20**	85.37
		cosinus	93.90	94.51	85.37	87.20
		correlation	91.46	92.07	85.98	87.80
	SVM	linear	**95.73**	**95.12**	**87.20**	**89.63**
	TM	chi	89.02	92.68	85.98	87.80
Feret	kNN	Euclidean	69.35	68.55	70.16	73.39
		cityblock	80.65	77.42	73.39	71.77
		cosinus	74.19	75.81	74.19	71.77
		correlation	74.19	74.19	77.42	76.61
	SVM	linear	**88.71**	**87.90**	**81.45**	**80.65**
	TM	chi	75.03	75.00	75.81	76.61
Combined	kNN	Euclidean	80.08	77.42	77.98	77.70
		cityblock	81.91	81.49	77.56	77.42
		cosinus	77.56	79.38	**78.54**	76.58
		correlation	78.40	79.38	**78.96**	77.42
	SVM	linear	**86.68**	**88.63**	77.98	**81.49**
	TM	chi	79.66	81.63	61.57	62.41

Table 2: Performance accuracy when different classifiers are applied.

	Division schemas								
	4×5			4×10			7×10		
	$R=1$	$R=2$	$R=3$	$R=1$	$R=2$	$R=3$	$R=1$	$R=2$	$R=3$
Cohn-Kanade									
ULBP	**96.34**	95.73	**96.34**	**96.95**	**96.34**	94.52	95.73	95.12	92.68
LBP	**97.56**	95.73	94.51	**96.34**	95.71	95.09	95.12	95.09	92.68
Feret									
ULBP	85.48	86.29	85.48	86.29	**87.90**	85.48	**88.71**	**87.90**	**87.90**
LBP	83.87	86.29	**87.90**	**87.10**	**87.10**	**87.90**	**87.90**	83.55	83.87
Combined									
ULBP	85.69	84.99	83.31	87.24	85.83	86.54	86.68	85.83	83.31
LBP	85.83	86.82	86.40	87.80	**88.92**	**89.34**	**88.63**	87.24	86.82

Table 3: Influence of image division schema on facial display classification.

	Division schemas					
	4×5 v1.			4×5 v2.		
	$R=1$	$R=2$	$R=3$	$R=1$	$R=2$	$R=3$
Cohn-Kanade						
ULBP	**96.32**	93.25	95.09	95.09	95.71	**97.55**
LBP	95.71	95.09	93.25	**96.32**	**96.32**	93.87
Feret						
ULBP	75.81	83.87	82.26	78.23	84.68	80.65
LBP	76.61	87.90	85.48	83.06	85.48	84.68
Combined						
ULBP	78.96	80.08	77.28	82.3	81.07	78.40
LBP	80.50	80.93	82.89	81.07	82.75	84.01

Table 4: Smiling vs. neutral facial display classification with 4×5 masks.

	Division schemas								
	7×10 v1.			7×10 v2.			7×10 v3.		
	$R=1$	$R=2$	$R=3$	$R=1$	$R=2$	$R=3$	$R=1$	$R=2$	$R=3$
Cohn-Kanade									
ULBP	**96.32**	95.09	92.02	**96.32**	95.71	95.71	**96.93**	**96.93**	**96.32**
LBP	93.87	92.64	93.87	93.87	95.09	94.48	**96.32**	**96.32**	**96.32**
Feret									
ULBP	81.45	83.06	75.81	76.90	85.48	85.48	86.29	87.10	85.48
LBP	86.29	83.87	81.45	**88.71**	86.29	84.68	**89.52**	87.10	**88.71**
Combined									
ULBP	83.03	80.93	78.12	86.40	84.01	82.75	86.40	84.29	84.15
LBP	85.97	84.71	85.55	86.82	86.12	86.82	86.96	87.10	**89.34**

Table 5: Smiling vs. neutral facial display classification with 7×10 masks.

	Feature vector length							
	4×5			4×10	7×10			
Texture operator		v1.	v2.			v1.	v2.	v3.
LBP	5120	2560	3072	10240	17920	4608	7936	9472
ULBP	1180	590	708	2360	4130	1062	1829	2183
	Percent of the original feature vector length							
	100	50	60	100	100	26	44	53

Table 6: Feature vector length for each division schema and applied mask.

ULBP			*PCA*	LBP		
4×5	7×10	mask	Information [%]	4×5	7×10	mask
Cohn-Kanade						
94.48	95.09	94.48	100	95.09	94.48	93.87
91.41	93.87	90.80	99	93.87	92.64	91.41
92.02	91.41	90.80	98	92.02	94.48	90.80
92.64	92.64	93.25	97	93.87	90.80	92.02
Feret						
78.23	75.81	79.03	100	76.61	83.06	86.29
65.32	65.32	73.39	99	71.77	64.52	72.58
66.13	65.32	68.55	98	62.10	62.10	69.35
64.52	70.16	69.35	97	63.71	67.74	74.19
Combined						
77.70	81.63	83.59	100	80.22	82.33	82.61
74.61	70.27	75.32	99	75.32	69.57	73.63
74.05	69.14	76.44	98	73.35	67.60	73.49
74.19	66.34	76.44	97	73.35	68.02	74.05

Table 7: *Cohn-Kanade* database. Results of classification after applying PCA on LBP texture operator.

ULBP				LBP		
4×5	7×10	mask	PCA	4×5	7×10	mask
1180	4130	2183	100%	5120	17920	9472
Length is given as a percent of the original feature vector length						
33	15	26	99%	9	3	6
25	13	23	98%	7	3	6
20	12	20	97%	6	3	5

Table 8: Feature vector length changes due to application of PCA.

References

[1] P. Ekman and W.V. Friesen. Felt, false, and miserable smiles. *Journal of Nonverbal Behavior*, 6(4):238–252, 1982.

[2] P. Ekman, R.J. Davidson, and W.V. Friesen. The Duchenne smile: Emotional expression and brain physiology: Ii. *Journal of Personality and Social Psychology*, 58(2):342–353, 1990.

[3] S. Scherer, G. Stratou, M. Mahmound, J. Boberg, J. Gratch, A. Rizzo, and L.-P. Morency. Automatic Behavior Descriptors for Psychological Disorder Analysis. In *IEEE Conference on Automatic Face and Gesture Recognition*, Shanghai, China, April 2013.

[4] J.M. Girard, J.F. Cohn, M.H. Mahoor, S.M. Mavadati, Z. Hammal, and D.P. Rosenwald. Nonverbal social withdrawal in depression: Evidence from manual and automatic analyses. *Image Vision Comput.*, 32(10):641–647, 2014.

[5] J.M. VanSwearingen, J.F. Cohn, and A. Bajaj-Luthra. Specific impairment of smiling increases the severity of depressive symptoms in patients with facial neuromuscular disorders. *Aesthetic Plastic Surgery*, 23(6):416–423, 1999.

[6] L.I. Reed, M. Sayette, and J.F. Cohn. Impact of depression on response to comedy: A dynamic facial coding analysis. *Journal of Abnormal Psychology*, 116(4):804–809, 2007.

[7] Tsunemoto Suzuki. [The effects of hypnosis on emotional responses of depressed students in frustrating situations]. *Shinrigaku kenkyu : The Japanese journal of psychology*, 73(6):457–463, February 2003.

[8] C.G. Kohler, E.A. Martin, N. Stolar, F.S. Barrett, R. Verma, C. Brensinger, W. Bilker, R.E. Gur, and R.C. Gur. Static posed and evoked facial expressions of emotions in schizophrenia. *Schizophrenia Research*, 105(1-3):49–60, October 2008.

[9] T. Souto, A. Baptista, D. Tavares, C. Queirós, and A. Marques. Facial emotional recognition in schizophrenia: preliminary results of the virtual reality program for facial emotional recognition. *Revista de Psiquiatria Clinica*, 40(4):129–134, 2013.

[10] Z.L. Boraston, B. Corden, L.K. Miles, D.H. Skuse, and S.J Blakemore. Brief report: Perception of genuine and posed smiles by individuals with autism. *Journal of Autism and Developmental Disorders*, 38(3):574–580, 2008.

[11] V. Manera, E. Grandi, and L. Colle. Susceptibility to emotional contagion for negative emotions improves detection of smile authenticity. *Frontiers in Human Neuroscience*, 7(6), 2013.

[12] J. Masip, E. Garrido, and C. Herrero. Defining deception. *Anales de Psicologia*, 20(1):147–171, 2004.

[13] S. Porter, N. Doucette, M. Woodworth, J. Earle, and B. MacNeil. Halfe the world knowes not how the other halfe lies: Investigation of cues to deception exhibited by criminal offenders and non-offenders. *Legal and Criminological Psychology*, 13:27–38, 2008.

[14] C.M. Hurley and M.G. Frank. Executing facial control during deception situations. *Journal of Nonverbal Behavior*, 35(2):119–131, 2011.

[15] P. Ekman and W. Friesen. Facial action coding system: a technique for the measurement of facial movement. *Consulting Psychologists Press*, 1978.

[16] D.S. Messinger, M.H. Mahoor, S-M. Chow, and J.F. Cohn. Automated measurement of facial expression in infant-mother interaction: a pilot study. *Infancy: the official journal of the International Society on Infant Studies*, 14(3):285–305, 2009.

[17] M. Valstar. Automatic behaviour understanding in medicine. In D. Heylen and A. Vinciarelli, editors, *Proceedings of the 2014 Workshop on Roadmapping the Future of Multimodal Interaction Research including Business Opportunities and Challenges, RFMIR@ICMI 2014, Istanbul, Turkey, November 16, 2014*, pages 57–60. ACM, 2014.

[18] D. Sato, A. Sugiura, and K. Yonemura. Relation between influence of smile on brain and stroop effect. In R. Magjarevic and J.H. Nagel, editors, *World Congress on Medical Physics and Biomedical Engineering 2006*, volume 14 of *IFMBE Proceedings*, pages 2868–2871. Springer Berlin Heidelberg, 2007.

[19] J. Pavulus. Kinect-powered depression detector is amazing and creepy. http://www.technologyreview.com/view/513126/kinect-powered-depression-detector-is-amazing-and-creepy/, March, 8 2015.

[20] Hao-Yu Wu, M. Rubinstein, E. Shih, J. Guttag, F. Durand, and W.T. Freeman. Eulerian video magnification for revealing subtle changes in the world. *ACM Trans. Graph. (Proceedings SIGGRAPH 2012)*, 31(4), 2012.

[21] P. Peck. Got a minute? you could diagnose stroke. http://www.webmd.com/stroke/news/20030213/got-minute-you-could-diagnose-stroke, March, 8 2015.

[22] K. Mase. An application of optical flow - extraction of facial expression. In *IAPR Workshop on Machine Vision Applications*, pages 195–198, 1990.

[23] M.S Bartlett, J.C Hager, P. Ekman, and T.J. Sejnowski. Measuring facial expressions by computer image analysis. *Psychophysiology*, 36(2):253–263, 1999.

[24] I.A. Essa and A.P. Pentland. Coding, analysis, interpretation, and recognition of facial expressions. *IEEE Trans. Pattern Anal. Mach. Intell.*, 19(7):757–763, July 1997.

[25] A. Lanitis, C.J. Taylor, and T.F. Cootes. Automatic interpretation and coding of face images using flexible models. *IEEE Trans. Pattern Anal. Mach. Intell.*, 19(7):743–756, July 1997.

[26] J.F. Cohn, A.J. Zlochower, J. Lien, and T. Kanade. Automated face analysis by feature point tracking has high concurrent validity with manual facs coding. *Psychophysiology*, 36(2):35–43, 1999.

[27] J. Cohn, A.J. Zlochower, J.J. Lien, and T. Kanade. Feature-point tracking by optical flow discriminates subtle differences in facial expression. In *Proceedings of the 3rd IEEE International Conference on Automatic Face and Gesture Recognition (FG '98)*, pages 396 – 401, April 1998.

[28] Shu Liao, Wei Fan, A.C.S. Chung, and Dit-Yan Yeung. Facial expression recognition using advanced local binary patterns, tsallis entropies and global appearance features. In *Image Processing, 2006 IEEE International Conference on*, pages 665–668, Oct 2006.

[29] M. Pantic and L.J.M. Rothkrantz. Automatic analysis of facial expressions: The state of the art. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22:1424–1445, 2000.

[30] C. Shan, Sh. Gong, and P.W. McOwan. Facial expression recognition based on local binary patterns: A comprehensive study. *Image Vision Comput.*, 27(6):803–816, May 2009.

[31] Qi Wu, Xunbing Shen, and Xiaolan Fu. The machine knows what you are hiding: An automatic micro-expression recognition system. In S. DMello, A. Graesser, B. Schuller, and J.-C. Martin, editors, *Affective Computing and Intelligent Interaction*, volume 6975 of *Lecture Notes in Computer Science*, pages 152–162. Springer Berlin Heidelberg, 2011.

[32] Le An, Songfan Yang, and Bir Bhanu. Efficient smile detection by extreme learning machine. *Neurocomputing*, 2014.

[33] D. Freire, M. Castrillón, and O. Dniz. Smile detection using local binary patterns and support vector machines. In *VISAPP (1)*, pages 398–401. INSTICC Press, 2009.

[34] David Freire, Luis Antn, and Modesto Castrilln. A study for the self similarity smile detection. In J. Fierrez, J. Ortega-Garcia, A. Esposito, A. Drygajlo, and M. Faundez-Zanuy, editors, *Biometric ID Management and Multimodal Communication*, volume 5707 of *Lecture Notes in Computer Science*, pages 97–104. Springer Berlin Heidelberg, 2009.

[35] H. Dibeklioglu, R. Valenti, A.A. Salah, and T. Gevers. Eyes do not lie: Spontaneous versus posed smiles. In *ACM International Conference on Multimedia*, 2010.

[36] Keiji Shimada, Tetsu Matsukawa, Yoshihiro Noguchi, and Takio Kurita. Appearance-based smile intensity estimation by cascaded support vector machines. In Reinhard Koch and Fay Huang, editors, *ACCV Workshops (1)*, volume 6468 of *Lecture Notes in Computer Science*, pages 277–286. Springer, 2010.

[37] F. De la Torre and J.F. Cohn. *Guide to Visual Analysis of Humans: Looking at People*, chapter Facial Expression Analysis. Springer, 2011.

[38] J. Whitehill, G. Littlewort, I. Fasel, M. Bartlett, and J. Movellan. Toward practical smile detection. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 31(11):2106–2111, Nov 2009.

[39] M. Petrou and P.G. Sevilla. *Image processing: dealing with texture*. John Wiley & Sons Inc., 2006.

[40] R.M. Haralick, K. Shanmugam, and Its'Hak Dinstein. Textural features for image classification. *IEEE Transactions on Systems, Man and Cybernetics,*, SMC-3(6):610–621, Nov 1973.

[41] F. Albregtsen, B. Nielsen, and H.E. Danielsen. Adaptive gray level run length features from class distance matrices. In *Pattern Recognition. Proc. 15th International Conf. on*, volume 3, pages 738–741 vol.3, 2000.

[42] T. Ojala, M. Pietikäinen, and D. Harwood. A comparative study of texture measures with classification based on featured distributions. *Pattern Recognition*, 29(1):51–59, 1996.

[43] T. Ojala, M. Pietikäinen, and T. Mäenpää. A generalized local binary pattern operator for multiresolution gray scale and rotation invariant texture classification. In *Proc. 2nd Intern. Conf. on Advances in Pattern Recognition*, pages 397–406, London, UK, 2001. Springer.

[44] T. Ojala, M. Pietikäinen, and T. Mäenpää. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 24(7):971–987, Jul 2002.

[45] M. Pietikäinen, A. Hadid, G. Zhao, and T. Ahonen. *Computer Vision Using Local Binary Patterns*, volume 40 of *Computational Imaging and Vision*. Springer, 2011.

[46] T. Ahonen, J. Matas, Ch. He, and M. Pietikäinen. Rotation invariant image description with local binary pattern histogram fourier features. In *Proceedings of the 16th Scandinavian Conference on Image Analysis*, SCIA '09, pages 61–70, Berlin, Heidelberg, 2009. Springer-Verlag.

[47] Yu Wang, Yongsheng Zhao, and Yi Chen. Texture classification using rotation invariant models on integrated local binary pattern and zernike moments. *EURASIP J. Adv. Sig. Proc.*, 2014:182, 2014.

[48] M. Kawulok and J. Szymanek. Precise multi-level face detector for advanced analysis of facial images. *Image Processing, IET*, 6(2):95–103, March 2012.

[49] J.-J. J. Lien, T. Kanade, J.Y. Cohn, and C. Li. Detection, tracking, and classification of action units in facial expression. *Journal of Robotics and Autonomous Systems*, 31:131–146, July 1999.

[50] FERET database. http://www.itl.nist.gov/iad/humanid/ feret/feret_master.html, July 2014.

[51] Smile databases. http://pics.psych.stir.ac.uk, July 2014.

[52] K. Nurzynska and B. Smolka. Optimal classification method for smiling vs. neutral facial display recognition. *Journal of Medical Informatics and Technology*, 23:87–94, 2014.