PCA Application in Classification of Smiling and Neutral Facial Displays

Karolina Nurzynska[1(✉)] and Bogdan Smolka[2]

[1] Institute of Informatics,
Silesian University of Technology, ul. Akademicka 16
44-100 Gliwice, Poland
karolina.nurzynska@polsl.pl
[2] Institute of Automatic Control,
Silesian University of Technology, ul. Akademicka 16
44-100 Gliwice, Poland
bogdan.smolka@polsl.pl

Abstract. Psychologists claim that the majority of inter-human communication is transferred non-verbally. The face and its expressions are the most significant channel of this kind of communication. In this research the problem of recognition between smiling and neutral facial display is investigated. The set-up consisting of proper local binary pattern operator supported with an image division schema and PCA for feature vector length diminishing is presented. The achieved results using k-nearest neighbourhood classifier are compared on a couple of image datasets to prove successful application of this approach.

1 Introduction

Emotions depicted on human face are the most important method of non-verbal communication between people. Regardless how well one can distinguish between different emotional states of others seeing only their facial display, it is rather easy to deceive the person. The consequences of this mistake might be unpleasant in personal life but also very severe when security measures are considered. Therefore, creation of automatic systems, which could deal with this issue, is an urgent research problem.

The research concentrating on facial expression recognition started in the last decade of XX century. First works used optical flow to follow the movement of muscles displayed in the image sequence [16]. Then this technique was supported with other means, e.g. holistic approach, explicit measurements, as is described in [5,9]. Next, another solution how to describe a subtle motion in image sequences is given in [6,7]. The last research concentrates on Facial Action Coding System [8] describing which muscles activate during a given emotional state. Finally, active shape models were exploited to describe the emotions in [13]. Although, eigenfaces introduced in [22] are commonly used for face recognition, more recently the local binary patterns [4] were applied to solve this problem [3,10]. Moreover, this texture operator is also exploited for emotion classification [21]. This technique gives very good outcome as reported in [14,19].

S. Kozielski et al. (Eds.): BDAS 2015, CCIS 521, pp. 398–407, 2015.
DOI: 10.1007/978-3-319-18422-7_35

In this research some existing algorithms for facial display classification into smiling and neutral expression are compared using similar experiment set-up and classification method. Moreover, the application of principal component analysis for the reduction of the feature vector length, which for considered methods is very long, is a novel contribution. Following section 2 describes the local binary patterns which are used for image description. Section 3 gives detailed information about the experiment set-up. Then, the results of performed experiments are described in section 4 and the conclusions are drawn in section 5.

2 Local Binary Patterns

Local binary patterns, (LBP), describe a texture in a form of a histogram. Each histogram bin represents a numerical value corresponding to a small texture area [20]. There are a couple of algorithms which might be used for the bin value calculation. Depending on the chosen approach, the distribution of the values in histogram changes as well as the length. In the following sections the two most popular versions of this texture operator are presented.

2.1 Basic LBP Operator

This operator was introduced by Ojala et al. [17] with assumption that the texture might be well described by its local pattern and its strength. Firstly, this approach assumed a 3×3 pixel neighbourhood in the monochromatic image I. Yet soon it was generalised to a circular neighbourhood, which radius R and number of sampling points P were parametrized [18]. Thus for each pixel (x, y) the neighbourhood is defined as follows:

$$\begin{aligned} g_p &= I(x_p, y_p), \\ p &= 0, ..., P-1, \\ x_p &= x + R\cos(2\pi p/P), \\ y_p &= y - R\sin(2\pi p/P), \end{aligned} \tag{1}$$

where g_p is a grey-scale value of the p-th sampled point. The gathered information from the local texture T, represents a joint distribution t, of the pixel: $T = t(g_c, g_0, g_1, ..., g_{P-1})$, for $g_c = I(x, y)$. Instead of storing the original pixel values, more information may be derived from the differences between the sampled points and the central value g_c. Moreover, since the central value is statistically independent, it might be omitted. Then the distribution is given as $t(g_0 - g_c, g_1 - g_c, ..., g_{P-1} - g_c)$. Still, there are many possibilities to code the information, therefore only the sign of the differences is recorded in the generic local binary patterns:

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

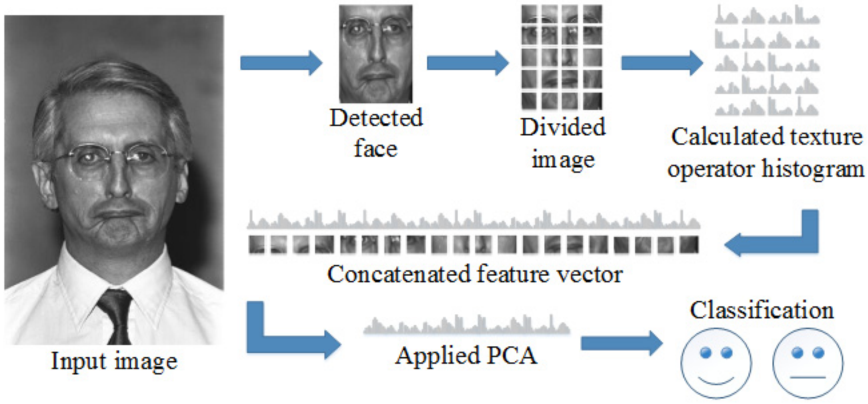


Fig. 1. Experiment set-up

where

$$s(z) = \begin{cases} 1, & z \geq 0, \\ 0, & z < 0. \end{cases} \quad (3)$$

2.2 Uniform Patterns

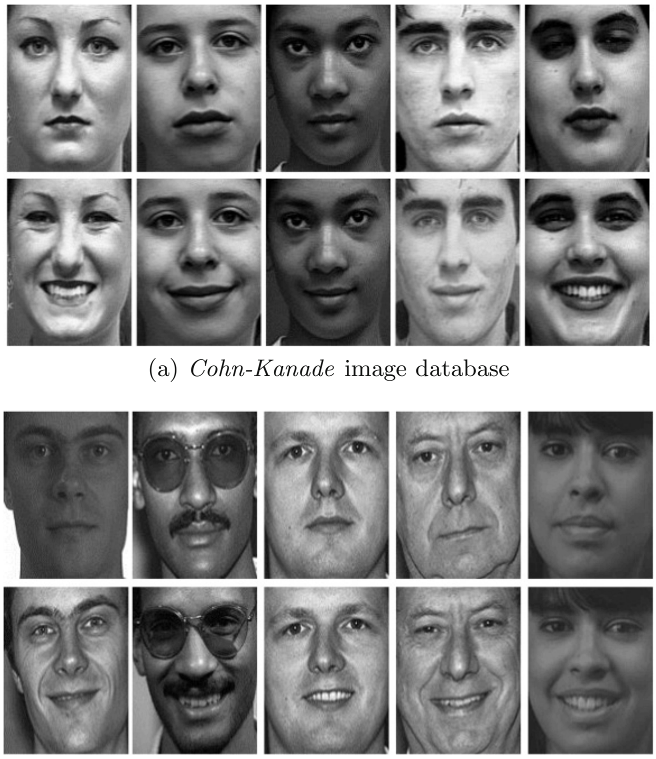
The histogram generated with LBP contains 2^P different values (for $P = 8$ that gives 256 elements), hence as a feature vector is rather long. Moreover, it was observed that some patterns store more valid information than others. It was concluded [20] that the most descriptive labels are when there are up to two transition between 0 and 1 in the operator written as a series of bits, e.g. 01110000. This patterns were named a 'uniform' patterns, ($uLBP$) and have separate bins, while all other with higher number of transitions are considered non-uniform and are placed in one bin. This approach diminishes the number of histogram bins to $P(P-1)+3$ for P sampled points.

3 Experiment Set-up

In order to start classification between smiling and neutral facial displays it is necessary to perform some preliminary processing of the input image. The steps applied in presented framework are depicted in Fig. 1 and described below. Moreover, an overview of used images datasets is also presented.

3.1 Image Databases

The discussed approach for emotion classification was evaluated on three different databases (some exemplary images are presented in Fig. 2). First of them is the Cohn-Kanade AU-Coded Facial Expression Database [6,15] which consists



(a) *Cohn-Kanade* image database

(b) *Feret* image database

Fig. 2. Examples of images gathered in the databases. Neutral facial displays are in the top row, whereas smiling one in the bottom row.

of image sequences presenting the change of facial display from neutral to representing the basic emotion, like smile, anger, etc. From 82 sequences, images with neutral and smiling facial display were selected. These are monochromatic images with 640×480 pixel resolution. The lighting condition varies in the images, the subjects do not wear any covering elements, such as glasses or beards (see Fig. 2a).

Second database was created from images collected in *Feret* data set prepared for face recognition purposes [1]. Here, 62 images were selected to represent each emotional state. The subjects differ between the groups, however it is possible to have several pictures of one subject in the group. The images are also monochromatic with varying lighting condition. Some of the subjects are wearing glasses or have beards. Refer to Fig. 2b for some examples.

Due to a small number of facial expression in each of the accessed databases, it is difficult to generalise the obtained results, therefore a database containing images from several existing data sets was prepared. It contains data from previously described databases but also from *Iranian*, *Nottingham originals*, *Pain*, and *Utrecht* data sets [2]. There are 712 images in total for both expressions. Some of these images are in colour, other are monochromatic, the resolution also varies.

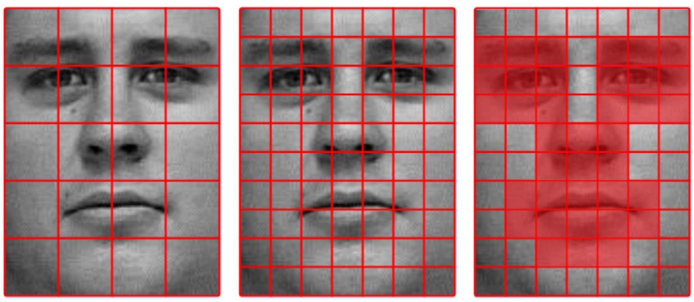
3.2 Face Image Acquisition

The first step of image preprocessing converts all images into grey-scale ones. Then, in order to describe emotion with texture operator, it is necessary to find and normalize the face region in the image. For this purpose a face detector described in [12] is applied. Since the original images are taken in various resolution and also the head size and distance from the camera changes between images, it was decided to resize the detected face into a constant resolution of 112×150 pixels for further processing.

3.3 Image Division

The emotions drawn on the face are related to its certain parts. According to research conducted by Ekman [8] it is possible to distinguish muscles which activate when displaying an emotion in the face. In this research the Face Action Coding System was developed, which binds the muscle movement with an emotion depicted in the face. Therefore, calculating a global *LBP* histogram for the whole image would be too general to describe precisely the visual information. In consequence, the image is divided into several sub-images, and the final feature vector is a concatenation of histograms calculated for each of them, as it is presented in Fig. 1. Moreover, the sub-images which present the region of muscle activation (in case of the smile: eyes and lips) convey more information, therefore it is possible to consider only these sub-images when building a feature vector by applying adequate masks.

In the presented work, three approaches for image division were considered. One of them divided the images into 20 sub-images using a mask with 4 columns



(a) 4×5 (b) 7×10 (c) mask

Fig. 3. Sub-images used to build a feature vector

and 5 rows. The second approach were more detailed and divided the images into 70 sub-images with 7 columns and 10 rows. Finally, in the second approach a mask was applied to analyse only the parts where the changes really took place as depicted in Fig. 3.

3.4 Principal Component Analysis

The length of the feature vector calculated to describe the image, depending on the chosen image division schema and texture operator applied, can result in hundreds up to thousands of elements. This severely complicates the classification process. Therefore, an attempt to diminish the number of elements by selecting those which convey the most descriptive information was undertaken.

In order to reduce the dimensionality of a large data set a principal component analysis is used [11], PCA. It is a mathematical projection, which transforms the data space description from the original one appointed by each feature in the feature vector into a new one, where the data variability is considered for space description. Then, the first principal component (eigenvector) is a vector corresponding to the direction in which the highest data variability is measured, next the second principal component reflects the less variable direction orthogonal to the previous one, and so on. Moreover, for each eigenvector, it is possible to define its importance (eigenvalue) with respect to whole data set. Choosing the most informative eigenvectors and summing up corresponding eigenvalues enables to claim which part of the initial data is described when only the picked up components are used. When this accuracy of information seems sufficient for the task solution, the data from eigenspace is transformed back using only the selected eigenvectors for calculation, what reduces the dimensionality of the original data space.

In presented research, the PCA parameters were obtained on the training set and then used to transform the feature vectors in testing data set. In experiments the results achieved for original features vectors are compared to those recorded when 97%, 98%, or 99% of initial information is exploited.

4 Results and Discussion

The experiments were performed on three previously described image data sets. The research were conveyed for two different image division schemas and the mask approach. Tables contain also information whether change of one of the texture operator parameters (radius) influences the results. The second parameter (number of sampling points) was constant because of its correlation with histogram length. The classification was performed using the k-nearest neighbourhood technique with cityblock distance metric and k equal to 9.

Tables 1 and 2 gather results achieved for the *Cohn-Kanade* database. The performance is comparable for all used texture operator. It is interesting to observe, that the radius does not influence the results when feature vector of full length is considered (last row of the tables). However, the results gathered with PCA application vary when the radius and applied image division

Table 1. *Cohn-Kanade* database - *LBP* texture operator

	$R=1$			$R=2$			$R=3$		
Information	4×5	7×10	mask	4×5	7×10	mask	4×5	7×10	mask
97%	93.87	90.80	92.02	93.87	90.18	94.48	94.48	93.25	88.34
98%	92.02	94.48	90.80	92.64	90.80	96.32	93.25	93.87	87.73
99%	93.87	92.64	91.41	92.64	90.80	91.41	94.48	94.48	91.41
100%	95.09	94.48	93.87	96.93	95.09	94.48	96.32	95.09	94.48

Table 2. *Cohn-Kanade* database - *uLBP* texture operator

	$R=1$			$R=2$			$R=3$		
Information	4×5	7×10	mask	4×5	7×10	mask	4×5	7×10	mask
97%	92.64	92.64	93.25	94.48	92.64	93.25	93.25	93.25	90.80
98%	92.02	91.41	90.80	94.48	88.96	92.64	95.71	91.41	90.80
99%	91.41	93.87	90.80	93.87	88.34	92.02	93.25	93.87	90.18
100%	94.48	95.09	94.48	96.93	95.09	95.71	96.93	96.32	96.32

schema changes. When considering *LBP*, the correct classification rate of 96% was recorded when 98% of original information was exploited with the set-up using mask and $R = 2$. On the other hand, for *uLBP* the performance is worse around 2% when compared to full feature vector used in case of set-ups using 4×5 image division and mask.

The situation is different when *Feret* (results presented in tables 3, 4) and *All* data sets are considered (outcomes gathered in tables 5, 6). First of all, since the images vary much more, the general accuracy of correct recognition rate for the original feature vector length is lower (refer to the last row in each of the tables). The best result for *Feret* database achieves 86% for *LBP* texture

Table 3. *Feret* database - *LBP* texture operator

	$R=1$			$R=2$			$R=3$		
Information	4×5	7×10	mask	4×5	7×10	mask	4×5	7×10	mask
97%	63.71	67.74	74.19	66.94	58.06	65.32	70.97	69.35	66.13
98%	62.10	62.10	69.35	65.32	64.52	62.90	64.52	65.32	61.29
99%	71.77	64.52	72.58	70.16	67.74	62.90	63.71	68.55	62.90
100%	76.61	83.06	86.29	75.81	78.23	82.26	74.19	81.45	83.87

Table 4. *Feret* database - *uLBP* texture operator

	$R=1$			$R=2$			$R=3$		
Information	4×5	7×10	mask	4×5	7×10	mask	4×5	7×10	mask
97%	64.52	70.16	69.35	75.00	64.52	67.74	61.29	61.29	65.32
98%	66.13	65.32	68.55	67.74	62.10	67.74	75.81	62.90	58.87
99%	65.32	65.32	73.39	69.35	62.90	74.19	67.74	62.90	61.29
100 %	78.23	75.81	79.03	76.61	82.26	81.45	77.42	83.06	84.68

Table 5. *All* database - *LBP* texture operator

	$R = 1$			$R = 2$			$R = 3$		
Information	4×5	7×10	mask	4×5	7×10	mask	4×5	7×10	mask
97%	73.35	68.02	74.05	74.33	61.01	69.00	74.61	59.19	64.52
98%	73.35	67.60	73.49	76.02	60.45	69.00	72.79	59.33	65.50
99%	75.32	69.57	73.63	76.30	61.85	71.81	74.61	57.92	65.22
100%	80.22	82.33	82.61	81.35	82.19	83.03	83.73	83.03	81.91

Table 6. *All* database - *uLBP* texture operator

	$R = 1$			$R = 2$			$R = 3$		
Information	4×5	7×10	mask	4×5	7×10	mask	4×5	7×10	mask
97%	74.19	66.34	76.44	77.14	59.05	69.28	76.72	58.91	67.60
98%	74.05	69.14	76.44	76.58	59.19	69.28	76.16	56.94	63.96
99%	74.61	70.27	75.32	76.58	61.85	69.85	77.00	59.61	62.55
100%	77.70	81.63	83.59	81.07	83.03	84.01	82.61	80.50	79.66

operator when mask is applied and $R = 1$, whereas for *All* database 83% score is recorded in several configurations. Next, application of PCA analysis diminishes the performance considerably in almost every case. There are, however, few exceptions. In *Feret* data set described with *uLBP* for image division 4×5 when $R = 2$ when 97% of original information is used and for the same image division schema when $R = 3$ and 98% of original information is applied.

The exploitation of PCA enables diminishing the feature vector length considerably. Table 7 presents the lengths of the feature vector when each of the texture operator is applied for each image description schema. Since the length of the feature vector after PCA application differs slightly in each experiment (even for similar set-up) instead of numeric length the part of the original feature vector length is given in percentage. One can see that in case of *LBP* the reduction is much higher when the other texture operator is applied. But when looking at the number of eigenvectors used in both cases they are rather similar.

From the performed experiments one can draw a conclusion that in case of *Cohn-Kanade* database the application of PCA might bring some benefits with

Table 7. Feature vector length

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

feature vector shortening without big impact on the correct classification accuracy. In other cases, however, application of PCA should be applied with care.

5 Conclusions

In this paper, a framework enabling classification between smiling and neutral facial displays was presented. The face was described using two texture operators: *LBP* and *uLBP*. Moreover, three different approaches to face description were given. Two of them consisted of image division schemas, whereas the third used specially designed mask. For each set-up the classification performance was calculated for the reference databases: *Cohn-Kanade*, *Feret*, and *All*. The presented results show that, although application of PCA reduced the feature vector length reasonably, it also in most cases is responsible for correct classification accuracy loss, therefore it should be used carefully.

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: GCONiI - Upper-Silesian Center for Scientific Computation.

References

1. Feret database (July 2014), http://www.itl.nist.gov/iad/humanid/feret/feret_master.html
2. Smile databases (July 2014), http://pics.psych.stir.ac.uk
3. Ahonen, T., Hadid, A., Pietikainen, M.: Face description with local binary patterns: Application to face recognition. IEEE Trans. Pattern Anal. Mach. Intell. 28(12), 2037–2041 (2006), http://dx.doi.org/10.1109/TPAMI.2006.244
4. Ahonen, T., Matas, J., He, C., Pietikäinen, M.: Rotation invariant image description with local binary pattern histogram Fourier features. In: Salberg, A.-B., Hardeberg, J.Y., Jenssen, R. (eds.) SCIA 2009. LNCS, vol. 5575, pp. 61–70. Springer, Heidelberg (2009)
5. Bartlett, M.S., Hager, J.C., Ekman, P., Sejnowski, T.J.: Measuring facial expressions by computer image analysis. Psychophysiology 36(2), 253–263 (1999)
6. Cohn, J.F., Zlochower, A.J., Lien, J., Kanade, T.: Automated face analysis by feature point tracking has high concurrent validity with manual facs coding. Psychophysiology 36(2), 35–43 (1999)
7. Cohn, J.F., Zlochower, A.J., Lien, J.J., Kanade, T.: Feature-point tracking by optical flow discriminates subtle differences in facial expression. In: Proc. of the 3rd. Intern. Conf. on Face & Gesture Recog., Washington, DC, USA, pp. 396–401 (1998), http://dl.acm.org/citation.cfm?id=520809.796049
8. Ekman, P., Friesen, W.: Facial action coding system: A technique for the measurement of facial movement. Consulting Psychologists Press (1978)
9. Essa, I.A., Pentland, A.P.: Coding, analysis, interpretation, and recognition of facial expressions. IEEE Transactions on Pattern Analysis and Machine Intelligence 19(7), 757–763 (1997)

10. Heusch, G., Rodriguez, Y., Marcel, S.: Local binary patterns as an image preprocessing for face authentication. In: 7th International Conference on Automatic Face and Gesture Recognition, FGR 2006, pp. 6–1 (April 2006)
11. Jolliffe, I.T.: Principal Component Analysis. Springer (2002)
12. Kawulok, M., Szymanek, J.: Precise multi-level face detector for advanced analysis of facial images. IET Image Processing 6(2), 95–103 (2012)
13. Lanitis, A., Taylor, C.J., Cootes, T.F.: Automatic interpretation and coding of face images using flexible models. IEEE Trans. on Pattern Analysis and Machine Intelligence 19(7), 743–756 (1997)
14. Liao, S., Fan, W., Chung, A.C.S., Yeung, D.Y.: Facial expression recognition using advanced local binary patterns, tsallis entropies and global appearance features. In: 2006 IEEE International Conference on 2006 IEEE International Conference on Image Processing, pp. 665–668 (October 2006)
15. Lien, J.J.J., Kanade, T., Cohn, J.Y., Li, C.: Detection, tracking, and classification of action units in facial expression. Journal of Robotics and Autonomous Systems 31, 131–146 (1999)
16. Mase, K.: An application of optical flow – extraction of facial expression. In: IAPR Workshop on Machine Vision Applications, pp. 195–198 (1990)
17. Ojala, T., Pietikäinen, M., Harwood, D.: A comparative study of texture measures with classification based on featured distributions. Pattern Recognition 29(1), 51–59 (1996), http://www.sciencedirect.com/science/article/pii/0031320395000674
18. Ojala, T., Pietikäinen, M., Mäenpää, T.: A generalized local binary pattern operator for multiresolution gray scale and rotation invariant texture classification. In: Singh, S., Murshed, N., Kropatsch, W.G. (eds.) ICAPR 2001. LNCS, vol. 2013, pp. 397–406. Springer, Heidelberg (2001), http://dl.acm.org/citation.cfm?id=646260.685274
19. Pantic, M., Rothkrantz, L.J.M.: Automatic analysis of facial expressions: The state of the art. IEEE Trans. Pattern Anal. Mach. Intell. 22(12), 1424–1445 (2000), http://dx.doi.org/10.1109/34.895976
20. Pietikäinen, M., Zhao, G., Hadid, A., Ahonen, T.: Computer vision using local binary patterns 40, 13–9 (2011)
21. Shan, C., Gong, S., McOwan, P.W.: Facial expression recognition based on local binary patterns: A comprehensive study. Image Vision Comput. 27(6), 803–816 (2009), http://dx.doi.org/10.1016/j.imavis.2008.08.005
22. Turk, M.A., Pentland, A.P.: Face recognition using eigenfaces. In: Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 1991, pp. 586–591 (June 1991)