JOURNAL OF MEDICAL INFORMATICS & TECHNOLOGIES Vol. 23/2014, ISSN 1642-6037

local binary patterns,
support vector machines,
k-nearest neighbourhood, template matching

Karolina NURZYŃSKA[1], Bogdan SMOŁKA[2]

OPTIMAL CLASSIFICATION METHOD FOR SMILING VS NEUTRAL FACIAL DISPLAY RECOGNITION

Human face depicts what happens in the soul, therefore correct recognition of emotion on the basis of facial display is of high importance. This work concentrates on the problem of optimal classification technique selection for solving the issue of smiling versus neutral face recognition. There are compared most frequently applied classification techniques: k-nearest neighbourhood, support vector machines, and template matching. Their performance is evaluated on facial images from several image datasets, but with similar image description methods based on local binary patterns. According to the experiments results the linear support vector machine gives the most satisfactory outcomes for all conditions.

1. Introduction

Emotion recognition is one of the popular issues studied in the image understanding domain. For instance, it can be applied to automatic evaluation of customers service quality. Yet, much more interesting and challenging is the possibility of discrimination between spontaneous and posed emotion expression. That in future may enable a lie detection and could be used in security systems.

In pioneering works, a movement of facial muscles was extracted by the optical flow technique [11]. Next, this method was supported by other means [7] to improve the performance. [3] gives an overview of facial expression recognition, which compares the optical flow performance with holistic approach and explicit measurements. Finally, similar way is used in [5] to describe subtle motions in facial displays. This research exploits the Facial Action Coding System [6], describing muscle activation related to emotion expression, and compares recognition results with regard to the place, where the movement is noticed.

Since local binary pattern, (*LBP*), texture operator was so successfully applied for face recognition, it should not be surprising that soon it was exploited for facial expression classification, too. The work [17] presents results of emotion classification on the *Cohn-Kanade* database with very good performance. In this approach the uniform version of this texture operator is exploited and classified with application of support vector machine, (*SVM*), with different

[1] Institute of Informatics, Silesian University of Technology, Akademicka 16, 44-100 Gliwice, Poland.
[2] Institute of Automatic Control, Silesian University of Technology, Akademicka 16, 44-100 Gliwice, Poland.

kernels. Moreover, authors suggest a *boosted-LBP* to improve the classification accuracy. Finally, the work [9] compares their solution for facial expression classification based on *LBP* with other approaches, such as: gradient maps, Tsallis entropy of Gabor filtered images using *Jaffe* database. Combination of suggested methods gives 94% accuracy. Yet more comparison concerning facial expression recognition can be found in [15].

The presented work addresses the problem of optimal classification method choice for smiling and neutral facial display recognition. Yet, to authors knowledge, their influence on classification accuracy for the same image description parameters never was considered. Therefore, using well known *LBP* operator for facial expression description, we evaluate the most commonly used classifiers on few image datasets.

2. Local binary patterns

Local binary pattern, (*LBP*), is an operator which describes a texture in form of a histogram. Each histogram bin corresponds to a numerical value characterising a small texture area [16]. Nevertheless the straightforward construction of the operator, it has been proven to give very good results in various applications.

Basic LBP operator. Having a monochrome image $I(x, y)$, for each pixel (x, y) a grey level value is given as $g_c = I(x, y)$ [12], [13], [14]. Next, in order to define the circular neighbourhood of radius R around point (x, y) and with P sampling points following formulas are defined:

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

Since the local texture (T) of the image $I(x, y)$ is a joint distribution (t) of the neighbourhood pixels one can write:

$$T = t(g_c, g_0, g_1, ..., g_{P-1}). \tag{2}$$

Then subtracting the central value from the neighbourhood may be applied without loss of information. As the central pixel is statistically independent from the differences, the factorization can take place. Moreover, the grey level value in the centre of the texture does not contain any useful information, hence can be omitted. Calculated texture descriptor is still difficult to apply due to huge number of possible permutations of values. Therefore, only the signs are considered:

$$t(s(g_0 - g_c), s(g_1 - g_c), ..., s(g_{P-1} - g_c)), \tag{3}$$

and a following thresholding function is given:

$$s(z) = \begin{cases} 1, & z \geq 0 \\ 0, & z < 0. \end{cases} \tag{4}$$

Finally, the generic local binary pattern is calculated as a summing of the thresholded differences weighted by consecutive powers of two:

$$LBP_{P,R}(x_c, y_c) = \sum_{p=0}^{P-1} s(g_p - g_c)2^p. \tag{5}$$

The basic *LBP* texture operator consists of distribution of the values calculated according to the Eq. 5 for each pixel in the image.

Uniform Patterns. It was noticed [16] 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, eg. 01110000, and those patterns are defined as a 'uniform' patterns, (*uLBP*). Other with higher number of transitions are considered non-uniform and are given a single label. This approach diminished the number of different labels to $P(P-1)+3$ for P bits.

Rotational Invariance. A specific *uLBP* pattern is given as $U_P(n,r)$, where pair (n,r) specifies a uniform pattern, here n describes the number of 1-bits and r the rotation [2], [16] of the pattern. Moreover, the patterns are stored in an array, which enables easy recalculation of the pattern between different angles:

$$U_P(n,r) \quad \text{rotated} \quad \text{by} \quad \alpha = a\frac{360}{P} \quad \doteq U_P(n, r+a \mod P). \tag{6}$$

Rotation Invariant LBP. Following preceding consideration, rotation of texture is responsible of *LBP* histogram translation [16]. Yet, it is possible to normalize such histogram by rotation invariant mapping:

$$riLBP_{P,R} = \min_i ROR(LBP_{P,R}, i); \tag{7}$$

where $ROR(x,i)$ is a circular bitwise right rotation of a bit sequence x by i steps.

Rotation Invariant and Uniform LBP. One can also imagine the combination of rotation invariance applied for uniform local binary pattern, what will give a new texture operator called *riuLBP*.

3. Classifiers

In case of face expression recognition several classification approaches are used, however three of them are applied most frequently. These are: k-nearest neighbourhood, support vector machine, and template matching based on a chi square statistics. In presented research the accuracy and influence of classifier parameters are investigated.

K-nearest neighbourhood classification, (*kNN*), approach is a non-parametric pattern recognition method, which decides about element assignment on the basis of majority vote of k closest neighbours from the training set. The k parameter should be the smallest possible integer. There are several possible ways to calculate the distance between elements. In conducted research following metrics were considered: Euclidean, cityblock, cosine, and correlation. Moreover, k parameter took odd values in range from 5 to 13.

Template matching, (*TM*), is another approach for image classification. However, in presented case it works with *kNN* classifier. As described in [17] a template is calculated for each class of face expression, then the input image is classified with the *kNN* technique to the closest template. Each emotion template is calculated as an averaged feature vector. Next, a chi square statistics, χ^2, for two feature vectors S and M is used for distance calculation:

$$\chi^2(S,M) = \sum_i \frac{(S_i - M_i)^2}{S_i + M_i}. \tag{8}$$

Support vector machine. Different approach to classification represents support vector machine, (*SVM*), which aims to discover a data model useful for two-class problem solution. In order to solve the problem, the exemplary objects are treated as points in the space mapped in such a way, that it is possible to define a clear gap between objects from different categories. The goal is to achieve the gap as broad as possible. Then the testing examples are mapped to the same space where it is possible to predict to which class they belong to depending on which side of the gap they are. *SVM*, in addition to the linear, performs also a non-linear classification using several kernels which map the input data into high-dimensional feature space. In presented research following kernels were applied: linear kernel, Gaussian Radial Basis Function kernel with scaling factor taking following values $0.01, 0.25, 0.50, 0.75, 1.00, 2.00$, and polynomial kernel with order of $2, 3, 4$, and 5.

4. Experiments and results

4.1. Experiments setup

The presented algorithms for smiling versus neutral facial display classification were tested on several databases, which differ in acquisition schema, image resolution and other parameters. In order to perform the classification, images from all databases were normalized to monochrome images with 8-bits and which resolution is 112×150. The only exception are images from the *Pain-Cropped* database, which collection is already normalized, however differently. Nevertheless, this gives good opportunity to see the influence of image normalization on the algorithms performance.

Calculating one *LBP* histogram for whole image is insufficient for the task of emotion recognition. Therefore, the common procedure assumes to divide the image into smaller parts, calculate *LBP* histogram for each of them and finally concatenate all histograms in one, as it is presented in Fig. 1, [1]. This solution enables better description of the most interesting parts of the face, which are the neighbourhood of the mouth and eyes. During the experiments a division schema with 7 columns and 10 rows was exploited. The *LBP* radius was equal to 1 and the number of sampling points was set to 8.

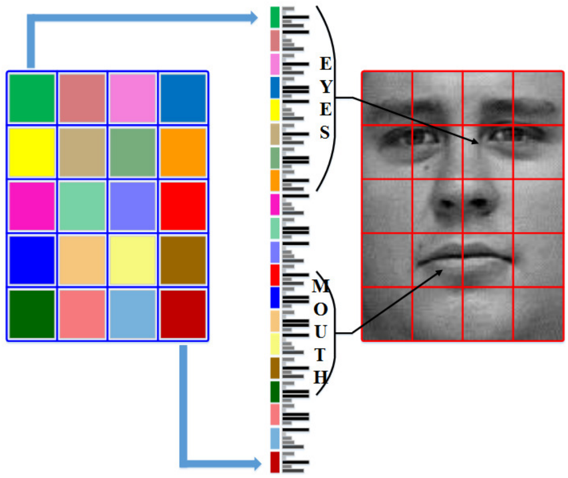

Fig. 1. Creation of feature vector with the *LBP* texture operator.

Due to a very small number of objects in the database the K-fold cross-validation approach was used for classification efficiency calculation. In all experiments K parameter was set to 10. The elements in databases were randomly selected for each from the K subsets.

4.2. Database overview

The most well-known database prepared for emotion classification is *Cohn-Kanade* AU-Coded Facial Expression Database [4], [10], which includes image sequences from neutral to target display of basic emotions. The available images were recorded by the camera placed in front of the subject. Image sequence for each emotion consists of different number of frames, which digital resolution is 640×480 or 490 pixels with 8-bit precision for greyscale values.

For the described research, images from 82 sequences were chosen which present the neutral and smiling expression. In the final subset there are 69% female, 15% African-American, and 2% Asian. The images differ in lighting conditions. The subjects do not were glasses or other

Table 1. *Cohn-Kanade* database classification results.

Classifiers	Params	$uLBP$	LBP	$riuLBP$	$riLBP$
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

Fig. 2. *Cohn-Kanade* database image examples.

Table 2. *Feret* database classification results.

Classifiers	Params	$uLBP$	LBP	$riuLBP$	$riLBP$
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

Fig. 3. *Feret* database image examples.

covering elements, as well as do not have beards. The same set of subjects represent neutral and smiling expressions. Some examples are presented in Fig. 2.

The best classification performance for each classification approach are presented in Table 1. One can see that the *linear SVM* usually overcomes other solutions and the best performance reaches 95% for the *LBP* and *uLBP* image description technique. The accuracy gained by the *kNN* for $k = 9$ is not far behind. However, one can see that the way of distance calculation between feature vectors influences classification result. In case of the mentioned texture operators the best are *cityblock* and *cosinus* distance metrics. It is interesting to point out that generally, the performance of rotation invariant techniques is poor. Finally, the *TM* classification schema results fall worse in comparison with other techniques, yet still for the *LBP* it reaches almost 93%.

Another very popular database is *Feret* [18]. It was possible to choose, from the corpus consisting of 14051 8-bit greyscale images of human heads with views ranging from frontal to left and right profiles, 62 frontal images for neutral and smiling expression each. The subjects differ in smiling and neutral groups. Moreover, it is possible to have different images for the same person in one group. Fig. 3 presents examples of this database.

The results presented in Table 2 similarly as in previous case show the best outcome achieved for each type of classification. One can easily notice that, although, the accuracy is weaker, similar behaviour of the classification techniques is observed. As it was before, the best results, around 88%, are achieved for *linear SVM* when the *LBP* and *uLBP* texture operator is applied. Then the *kNN* classifier with the *cityblock* distance metric and $k = 9$ gives good result. However, they are incomparable with the *SVM* performance. In case of *TM* classification schema the result is comparable with *kNN* outcome, yet it is stable nevertheless used texture operator.

Another considered image set was the *Pain* database [19], which consists of 599 posed basic expressions of 13 female and 10 male subjects. The images are in colour with 720×576 resolution. The experiments were conveyed on subset of this database which contained 48 frontal neutral and 48 smiling images; some examples are given in Fig. 4. There is also a subset of images with a cropped version collected for 12 female individuals. The images are normalized with fixed eye position, converted to monochromatic images with 181×241 resolution. Since for the best classifiers chosen in previous experiments in case of the cropped version almost all results reached 95-100% accuracy, they are not presented.

Table 3. *Pain* database classification results.

Classifiers	Params	$uLBP$	LBP	$riuLBP$	$riLBP$
kNN	Euclidean	100.00	97.87	96.81	95.74
	cityblock	98.94	96.81	95.74	95.74
	cosinus	97.87	98.94	97.87	95.74
	correlation	100.00	98.94	97.87	97.87
SVM	linear	100.00	100.00	100.00	97.87
TM	chi	97.87	98.94	92.55	92.55

Fig. 4. *Pain* database image examples.

Table 4. *Iranian* database classification results.

Classifiers	Params	$uLBP$	LBP	$riuLBP$	$riLBP$
kNN	Euclidean	72.22	70.83	75.00	76.39
	cityblock	76.39	80.56	79.17	73.61
	cosinus	69.44	68.06	75.00	75.00
	correlation	73.61	70.83	73.61	76.39
SVM	linear	72.22	63.89	80.56	77.78
TM	chi	72.22	65.28	73.61	69.44

Fig. 5. *Iranian* database image examples.

Table 5. *Nottingham* database classification results.

Classifiers	Params	$uLBP$	LBP	$riuLBP$	$riLBP$
kNN	Euclidean	70.63	67.46	72.22	73.02
	cityblock	73.81	70.63	74.60	76.98
	cosinus	69.05	69.05	76.19	72.22
	correlation	64.29	69.84	73.81	68.25
SVM	linear	76.98	79.37	76.98	76.98
TM	chi	75.40	71.43	75.40	67.46

Fig. 6. *Nottingham* database image examples.

Table 6. *Utrecht* database classification results.

Classifiers	Params	$uLBP$	LBP	$riuLBP$	$riLBP$
kNN	Euclidean	59.09	59.09	61.36	59.09
	cityblock	65.91	68.18	63.64	62.12
	cosinus	64.39	64.39	59.09	59.85
	correlation	63.64	62.12	63.64	59.85
SVM	linear	67.42	-	66.67	63.64
TM	chi	65.91	65.15	59.09	57.58

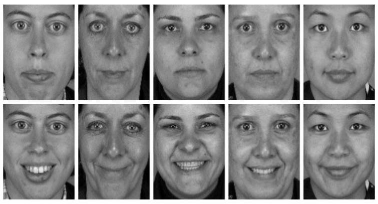

Fig. 7. *Utrecht* database image examples.

Table 3 gathers experiment results for the *Pain* dataset. Here, the classification outcomes vary between 92-100%. However, very good are those results, once again it is possible to observe that applying *linear SVM* brings the best performance. Yet, the other classification techniques are not much worse, especially when considering the small number of elements in the tests.

For comparison three other image databases were also considered: *Iranian* [19] (Table 4, Fig. 5), *Utrecht* [19] (Table 6, Fig. 7) and *Nottingham* originals [19] (Table 5, Fig. 6). The first two datasets consist of colour images, which resolution is 1200×900 pixels. There were chosen 36 frontal neutral and smiling images from the middle-east female subjects collected in the *Iranian* set. The subset containing the considered emotions in *Utrecht* image database consisted of 53 images, which represented both male and female. Finally, the *Nottingham* database contained 63 frontal images, which resolution was 288×384 for monochromatic image.

Tables 4-6 present classification accuracy results. For all this databases the performance is

Table 7. Number of subjects in each dataset.

DB Name	Elems.
Cohn-Kanade	164
Feret	124
Iranian	72
Nottingham	126
Pain	96
Pain-Cropped	24
Utrecht	106
All	712

Table 8. *All* database classification results.

Classifiers	Params	$uLBP$	LBP	$riuLBP$	$riLBP$
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

weaker and does not reach above 80%. In case of the *Nottingham* database it might result from the low resolution of input data. Yet, in the other image set it must result from the occlusion artefacts, e.g. glasses, on the images. Anyway, still the *linear SVM* gives the best outcome, loosing sometimes with *kNN* in case of the *Iranian* database.

4.3. Discussion

Since each from presented above databases consist of considerably small number of examples a collective database was created. Table 7 summarises the database volume, whereas Table 8 gathers the classification accuracy. The best performance was achieved by the *linear SVM* classifier for *LBP* (88.63%) and then *uLBP* (86.68%) texture operator. A little bit less accurate is the *kNN* classifier with $k = 9$. Once again, the best results for the mentioned texture operators are for the *cityblock* distance metrics. Similar results are noticed for *TM* classifier, however, when the rotation invariant version of texture operator is applied its performance is getting very low.

It can be claimed that the best technique for emotion classification is application of *linear SVM*. In presented research *SVM*s with other kernels and many different parameters were also considered. However, their performance were not satisfactory. It is also worth to mention, that in some cases it was difficult to train the *SVM* due to the considerable length of the feature vector.

The higher accuracy of the *kNN* classifier in most cases were achieved for $k = 9$. There were some cases, where $k = 11$ brought better results, yet the difference was negligible. In case of the *LBP* and *uLBP* the best results were achieved for the *cityblock* metric. Yet, when rotation invariant version are taken under consideration, it is difficult to choose particular metric.

The *TM* classifier is the most unstable one. There are cases where its accuracy is comparable or even better that *kNN*. So the authors believe that it could overcome other classifiers when good parameters are found, however generally it gives the worst performance from all considered techniques.

The classification techniques were tested on four available texture operators. In all experiments the best performance was achieved for the *LBP* and *uLBP* texture operator. It is not surprising that these methods give so similar results, since their definitions do 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. Astonishing is the fact, that nevertheless the image database the rotation invariant version of the image descriptor enables classification with accuracy around 70%, where the data never rotates. The unsatisfactory results probably are connected with the reduction of feature vector length.

5. Conclusions

This work presents results of experiments aiming at optimal classifier selection for the problem of emotion recognition on the basis of facial display. Three most frequently exploited classifiers are considered: k-nearest neighbourhood, support vector machines, and template matching. Their performance is evaluated on eight image databases: *Cohn-Kanade*, *Feret*, *Pain*, *Pain-Cropped*, *Iranian*, *Nottingham*, *Utrecht*, and *All*. Moreover four variation of the basic *LBP* operator were considered.

According to the gathered results, *linear SVM* outperforms other classifiers in almost all cases. However, application of *kNN* with parameter k set to 9 and distance metric *cityblock* in some settings is not far behind. Due to various outcomes of *TM* classifier, it is not suggested for further use, although, one should have in mind that with proper parameters it may give comparable results with others.

ACKNOWLEDGEMENT

This work has been supported by the Polish National Science Centre (NCN) under the Grant: DEC-2012/07/B/ST6/01227.

Bibliography

[1] AHONEN T., HADID A., PIETIKÄINEN M., Face Recognition with Local Binary Patterns, ECCV (1), 2004, pp. 469-481.
[2] AHONEN T., MATAS J., HE CH., PIETIKÄINEN M., Rotation Invariant Image Description with Local Binary Pattern Histogram Fourier Features, Proc. 16th Scandinavian Conf. on Image Anal.,Oslo, Norway, 2009, pp. 61-70.
[3] BARTLETT M.S, HAGER J.C, EKMAN P., SEJNOWSKI T.J., Measuring facial expressions by computer image analysis, Psychophysiology, 1999, Vol. 36, No. 2, pp. 253-263.
[4] 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, 1999, Vol. 36, pp. 35-43.
[5] COHN J., ZLOCHOWER A., LIEN J-J. J, KANADE T., Feature-point tracking by optical flow discriminates subtle differences in facial expression, Proc. 3rd IEEE Intern. Conf. on Autom. Face and Gesture Recog., 1998, pp. 396-401.
[6] EKMAN P. AND FRIESEN W.,Facial Action Coding System: A Technique for the Measurement of Facial Movement, Consulting Psychologists Press, 1978.
[7] ESSA I. A., PENTLAND A. P., Coding, Analysis, Interpretation, and Recognition of Facial Expressions, IEEE Trans. Pattern Anal. Mach. Intell., 1997, Vol. 19, No. 7, pp. 757-763.
[8] HEUSCH G. AND RODRIGUEZ Y. AND MARCEL S., Local binary patterns as an image preprocessing for face authentication, 7th International Conference on Automatic Face and Gesture Recognition, 2006, pp. 6-14.
[9] LIAO SH., FAN W., CHUNG, AC. S., YEUNG D.-Y., Facial Expression Recognition using Advanced Local Binary Patterns, Tsallis Entropies and Global Appearance Features, IEEE International Conference on Image Processing, 2006, pp. 665-668.
[10] LIEN J-J. J., KANADE T., COHN J., LI C., Detection, Tracking, and Classification of Action Units in Facial Expression, Journal of Robotics and Autonomous Systems, 1999.
[11] MASE K.,An Application of Optical Flow – Extraction of Facial Expression, IAPR Workshop on Machine Vision Applications, 1990, pp. 195-198.
[12] OJALA T., PIETIKÄINEN M., HARWOOD D., A comparative study of texture measures with classification based on featured distributions, Pattern Recognition, 1996, pp. 51-59.
[13] OJALA T., PIETIKÄINEN M., MÄENPÄÄ T., A generalized Local Binary Pattern operator for multiresolution gray scale and rotation invariant texture classification, Advances in Pattern Recognition, ICAPR 2001 Proceedings, Lecture Notes in Computer Science 2013, Springer, pp. 397-406.
[14] OJALA T., PIETIKÄINEN M., MÄENPÄÄ T., Multiresolution Gray-Scale and Rotation Invariant Texture Classification with Local Binary Patterns, IEEE Trans. Pattern Anal. Mach. Intell., 2002, Vol. 24, No. 7, pp. 971-987.
[15] PANTIC M., ROTHKRANTZ L. J. M.,Automatic analysis of facial expressions: the state of the art, Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2000, Vol. 22, pp. 1424-1445.
[16] PIETIKÄINEN M., ZHAO G., HADID A., AHONEN T., Computer Vision Using Local Binary Patterns, Computational Imaging and Vision, Springer, 2011, No. 40, pp. 13-49.
[17] SHAN C., GONG SH., MCOWAN P. W., Facial expression recognition based on Local Binary Patterns: A comprehensive study, Image and Vision Computing, 2009, Vol. 27, No. 6, pp. 803-816.
[18] FERET database *http://www.itl.nist.gov/iad/humanid/feret/feret_master.html* 2014.07.04
[19] Iranian, Nottingham, Pain, Utrecht database *pics.psych.stir.ac.uk* 2014.07.04