Improving Face Recognition Methods Based on POEM Features

Ladislav Lenc[1,2] and Pavel Král[1,2]

[1]*Dept. of Computer Science & Engineering, Faculty of Applied Sciences, University of West Bohemia, Plzeň, Czech Republic*

[2]*NTIS - New Technologies for the Information Society, Faculty of Applied Sciences, University of West Bohemia, Plzeň, Czech Republic*

llenc,pkral@kiv.zcu.cz

Keywords: POEM, Face Recognition, Local Features, UFI

Abstract: POEM descriptors has been successfully used for face recognition. The usual way how the descriptor is utilized consists in constructing POEM features in the rectangular non-overlapping regions covering the whole image. The features created in the regions are then concatenated into one long vector representing the face. We propose an enhancement of this method using automatic key-point identification strategies. In our approach, the image features are created in the detected key-points. We also employ a more complex matching procedure that compares the features individually. This method is efficient particularly when the number of training samples is small and therefore neural network based methods fail, because they do not have enough training data. The proposed approach is evaluated on three standard face corpora. We also study the influence of several parameters of the method on the overall performance. The obtained results show that the combination of POEM features with the automatic point identification and a more sophisticated matching algorithm brings significant improvement over the baseline method.

1 INTRODUCTION

Facial recognition is an intensively studied research field that finds its use in many practical applications. It is also one of the most useful biometric identification methods. There are many successful approaches based on local features that solve the face recognition problem.

This work concentrates on the patterns of oriented edge magnitudes (POEM) descriptors. It has been proven that POEM has a great ability to capture important information and it was successfully used for face recognition (Vu et al., 2012; Lenc, 2016). Similarly as other local descriptors, such as local binary patterns (LBP) and others, image representations are usually constructed by a concatenation of histograms of POEM values computed in rectangular image regions. This concept is known as a histogram sequence (HS). A significant improvement of the HS was proposed in (Lenc and Král, 2016)In this approach, LBP features are created in automatically detected points and a more sophisticated matching algorithm is used. The results of this approach show an improvement over the original methods using LBP and HS.

In this work, we propose a novel face recognition method which uses automatically detected key-points together with POEM features. We also aim at a more detailed evaluation of the influence of some important parameters of the key-point detection method. Compared to the original key-point detection method we employ also oriented FAST and rotated BRIEF (ORB) key-point detection algorithm.

We also evaluate several algorithms for key-point reduction. Some key-point detectors tend to find too many key-points which brings redundancy and increased computational costs. It is thus beneficial to use algorithms that can reduce the number of key-points while preserving the informative value of the key-point set. Another contribution of this work is the introduction of a new key-point reduction scheme. The proposed approach is efficient particularly when the number of training samples is small and therefore methods based on neural nets fail because of insufficient amount of training data.

The method is evaluated on three standard face corpora, namely AR, UFI and LFW. AR (Martinez and Benavente, 1998) database represents a well-controlled dataset, while UFI (Lenc and Král, 2015) and LFW (Huang et al., 2007) corpora are much more challenging because they contain real-world images with few training examples.

The rest of the paper is organized as follows. The

following section describes the relevant face recognition methods. Section 3 details the proposed approach. Section 4 presents the corpora used for evaluation and the following section presents experimental results realized on these data. The last section concludes the paper and proposes some future research directions.

2 Related Work

The LBP operator was originally utilized for texture classification in (Ojala et al., 1994). It is a variant of the texture unit (Wang and He, 1990). Its computation is based on a small local neighbourhood of a given pixel. By comparing the central pixel with its 8 neighbouring ones we create an 8 bit code representing the pixel. The bits in the code are set to 1 if the values of the given neighbouring pixels are greater than the value of the central one. The rest of the bits is set to 0.

The popularity of face recognition methods based on local descriptors has begun mainly due to the work of Ahonen et al. (Ahonen et al., 2004; Ahonen et al., 2006). It has been proven that these simple descriptors, initially used for texture classification, are very useful also in the face recognition field. This method has introduced the histogram sequence (HS) representation of face images. The utilization of HS ensures that histograms computed in corresponding parts of two face images are compared. The reduction of possible codes using only uniform patterns brought a speed-up of the method while preserving very good recognition accuracy.

A plethora of various, more or less sophisticated, extensions of LBP was proposed in the following years. We can mention e.g. local ternary patterns (LTP) (Tan and Triggs, 2010), dynamic threshold local binary patterns (DTLBP) (Li et al., 2012) and completed local binary patterns (CLBP) (Guo et al., 2010). All of these methods bring some small improvements over the basic LBP. The main advantage is a better handling of lower quality images that are affected by varying lighting conditions and noise.

Three- and Four-patch LBP variations were proposed in (Wolf et al., 2008). The codes are constructed by comparison of three or four patches respectively. The more sophisticated computation brings better robustness. The algorithm works very well on face recognition using the LFW dataset.

Local derivative patterns (LDP) differ in utilizing features of higher orders than LBP. Again, the main advantage over LBP is better accuracy in the case of challenging illumination conditions. A great success was made by the authors of the POEM descriptor (Vu et al., 2012). POEM based features outperformed many other image descriptors and succeeded also in the face verification task.

Further improvements of the basic descriptor based methods were proposed in (Lenc and Král, 2014) and (Lenc and Král, 2016). The main novelty consists in using automatically detected points for the feature construction. There is also an improved matching algorithm that allows reaching better accuracies in comparison with basic methods using HS. Another possible improvement lies in weighting. Each region in the face can have a different weight which again increases the accuracy. An approach utilizing genetic algorithm to set-up the weights was proposed in (Lenc, 2016).

A thorough description of more algorithms that were proposed is beyond the scope of this paper, therefore, for the further reading please refer to (Nanni et al., 2012).

3 POEM-based Face Recognition

The first step in the proposed algorithm is the key-point identification. Two methods for this tasks are described in Sections 3.1 an 3.2. The second step is the key-point reduction described in Section 3.3. Follows the description of the POEM algorithm in Section 3.4. The final building block of the processing pipeline is the image matching that is detailed in Section 3.5.

3.1 Gabor Wavelet Key-point Identification

We use the Gabor wavelets based method described in (Lenc and Král, 2014). Gabor filters utilized in this work are computed using eqs. (1) and (2) which describe the real and the imaginary part of the wavelet respectively (computed in point (r,c)).

$$g(r,c;\lambda,\theta,\psi,\sigma,\gamma) = \exp\left(-\frac{\acute{r}+\gamma^2\acute{c}^2}{2\sigma^2}\right)\cos\left(2\pi\frac{\acute{r}}{\lambda}+\psi\right) \quad (1)$$

$$g(r,c;\lambda,\theta,\psi,\sigma,\gamma) = \exp\left(-\frac{\acute{r}+\gamma^2\acute{c}^2}{2\sigma^2}\right)\sin\left(2\pi\frac{\acute{r}}{\lambda}+\psi\right) \quad (2)$$

where $\acute{r} = r\cos\theta + c\sin\theta$, $\acute{c} = -r\sin\theta + c\cos\theta$, λ is the wavelength of the cosine factor, θ represents the orientation of the filter and ψ is a phase offset, σ and γ are parameters of the Gaussian envelope,

σ is the standard deviation of the Gaussian and γ defines the ellipticity (aspect ratio) of the function. The magnitude is then computed from the real and imaginary parts.

We use a set of 40 Gabor filters that are applied on the face image. The filter responses are scanned using a square sliding window W of the size $w \times w$. The window centre (r_0, c_0) is considered to be a key-point iff:

$$R_j(r_0, c_0) = \max_{(r,c) \in W} R_j(r, c) \tag{3}$$

$$R_j(r_0, c_0) > \frac{1}{wi * hi} \sum_{r=1}^{wi} \sum_{c=1}^{hi} R_j(r, c) \tag{4}$$

where $j = 1, ..., N_G$ (N_G is the number of Gabor filters) and wi and hi are image width and height respectively.

We will refer as *REAL* the method which uses only the real part and *MAGNITUDE* the one that computes the magnitude from the real and the imaginary parts.

3.2 ORB Key-point Identification

ORB (Rublee et al., 2011) method was proposed as an alternative to the patented SIFT and SURF algorithms. It provides key-points that are computed using the FAST algorithm. BRIEF binary descriptors are used to create robust features that are successfully used for image matching and other applications. We use only the key-point detection part of this algorithm.

The FAST algorithm (Rosten and Drummond, 2006) detects corner key-points. Its main strength is its speed and it is thus often used in real-time applications. FAST uses a circle with 3 pixel radius and centre in the candidate key-point. The point P with the intensity I_P is a candidate if the intensities of N contiguous points on the circle are all:

1. greater then $I_P + t$
2. smaller then $I_P - t$

where t is a threshold. Parameter N and threshold t influence the number of resulting key-point candidates. N is usually set to 12. The ORB algorithm additionally uses Harris corner filter to select the best candidates.

3.3 Key-point Reduction

The key-points detected by one of the described algorithms can be used directly without any reduction. Unfortunately, the resulting number of key-points and features is very high. It has been shown (Lenc and Král, 2014) that this point number can be significantly reduced using clustering with no significant information loss. Therefore, we evaluate three ways how to construct the resulting points. All of them are based on the K-means clustering. We refer as *key-points* the points identified by the particular key-point detection algorithm. *Feature points* are the points where features are created.

3.3.1 Face Specific Position

This method constructs the feature points for each image independently. The set of key-points extracted by the key-point identification algorithm is used as the input of the clustering algorithm. The cluster centres are then directly used as the feature points. We will refer this method as FS-POEM.

3.3.2 Person Specific Position

The second method takes all images of one person and puts the key-points extracted from all images together. The points are then clustered and the cluster centres are used as the feature points. While testing, we extract key-points from the unknown image and cluster it to obtain the feature points. This method will be referred as PS-POEM.

3.3.3 Global Position

The global position method uses the same set of feature points for each image. A representative subset of the image gallery is used for key-point extraction. All resulting points are clustered and the centres are again used as the feature points. The points are thus the same for all images. No clustering is performed for the unknown images. Only the features are created in the determined feature points. We will denote this method as GL-POEM.

3.4 POEM Features

POEM descriptor was proposed in (Vu et al., 2012). It is based on gradients computed in each image pixel. The gradients are usually computed by one of the well-known edge detection convolution operators such as Sobel or Scharr. The approximation using these operators allows computing gradients in both x and y direction and subsequently compute the gradient magnitude and orientation.

The gradient orientations are then discretized. The usual number of orientations proposed in the original paper is 3. It is denoted d. Vector of the length d is thus used as a representation of each pixel. It is a histogram of gradient values in a small square

neighbourhood of a given pixel called *cell*. Figure 1 depicts the meaning of *cell* and *block* terms.

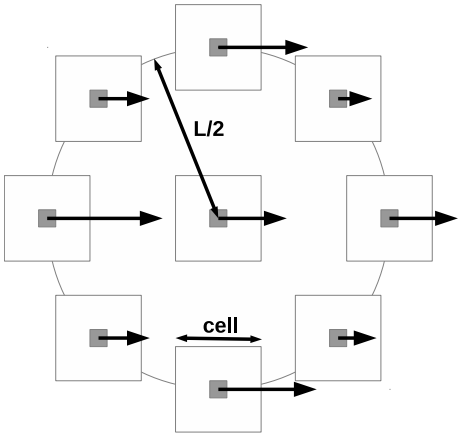

Figure 1: Computation of POEM descriptor. Squares around the pixels are called *cells* while the surroundings with diameter L is called *block*. Arrows represent the accumulated gradients.

The final encoding is similar to LBP as depicted in Figure 2. It is done in a round neighbourhood with diameter L called *block*. The algorithm assigns either 0 or 1 value to the 8 neighbouring pixels by Equation 5.

$$B_i = \begin{cases} 0 & \text{if } g_i < g_c \\ 1 & \text{if } g_i \geq g_c \end{cases} \tag{5}$$

where B_i is the binary value assigned to the neighbouring pixel $i \in \{1,..,8\}$, g_i denotes the gray-level value of the neighbouring pixel i and g_c is the gray-level value of the central pixel. The resulting values are then concatenated into an 8 bit number. Its decimal representation is used to create the feature vector.

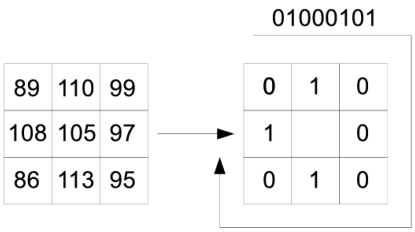

Figure 2: Computation of the final POEM value.

It is computed for each gradient orientation and thus the descriptor is d times longer than in the case of LBP.

3.5 Matching Algorithm

In the proposed algorithm we do not concatenate the feature vectors. Instead, the distance of the feature sets is computed using the algorithm utilized in (Lenc and Král, 2016).

The feature vectors are compared using the histogram intersection (HI). The advantage of this method is its simplicity and fast computation. Moreover, in our preliminary experiments it outperformed some more sophisticated methods such as χ^2 statistic in terms of accuracy. HI computation is described by eq. (6).

$$HI(f,r) = 1 - \sum_i \min(f_i, r_i) \tag{6}$$

where i is the number of histogram bins. This form is interpreted as a distance measure. 0 value thus means the same histograms.

The distance of two face representations is computed by eq. (7).

$$sim(F,R) = \sum_{f_i} \min_{r_j \in N(f_i)} (HI(f_i, r_j)) \tag{7}$$

where $N(f_i)$ is the neighbourhood of the feature f_i defined by the $distanceThreshold$ that specifies the maximum distance within that the features are compared. It means that for each feature of the face F we find the closest one within the neighbourhood $N(f_i)$ from the face R. The distance of the two faces is computed as a sum of these minimum distances.

The recognized face $\hat{F}$ is given by the following equation:

$$\hat{F} = \arg\min_R (sim(F,R)) \tag{8}$$

4 Corpora

4.1 AR Face Database

AR Face Database[1] (Martinez and Benavente, 1998) was created at the Univerzitat Autonòma de Barcelona. This database contains more than 4,000 colour images of 126 individuals. The images are stored in a *raw* format and their size is 768×576 pixels. The individuals are captured under significantly different lighting conditions and with varying expressions. Another characteristic is a possible presence of glasses or scarf.

4.2 Labeled Faces in the Wild

We use the cropped version of the well-known Labeled faces in the wild (LFW) dataset (Huang et al.,

[1] http://www2.ece.ohio-state.edu/ aleix/ARdatabase.html

2007). This version was first utilized in (Sanderson and Lovell, 2009). The main reason for the cropping is the presence of a background in the original images that may add information and improve performance in some cases. The preprocessing should ensure more fair conditions for testing.

The extraction method places a bounding box around the faces and resizes the resulting area to 64×64 pixels. The bounding box is placed to the same location in every image.

We use the identification scenario proposed in (Xu et al., 2014). It uses a subset of 86 people with 11 to 20 images per person. 7 images of each person are used for the gallery and the rest is used as the probe set. The total numbers of images are 602 and 649 for gallery and probe set respectively.

4.3 Unconstrained Facial Images

The Unconstrained facial images (UFI) dataset was proposed in (Lenc and Král, 2015). It is a real-world database created from photographs acquired by reporters of a news agency. It thus shows significant variances in the image quality, face orientation, face occlusion etc. The database is designated for the identification task. It comes with two image sets. The *Cropped images* dataset contains preprocessed faces extracted from photographs while the *Large images* includes a variable amount of background. We utilize the cropped version while the other partition is intended to be used with complete face recognition systems including the face localization stage.

The images have resolution of 128×128 pixels. The total number of individuals is 605. In average 7.1 images of each person are in the gallery set. The total number of gallery images is 4316. The probe set contains just one image for every individual.

5 Experiments

The first experiment was carried out to compare the key-point identification methods. No clustering is performed in this case and the detected key-points are directly used as the feature points. The size of the sliding window is set to 25 according to (Lenc and Král, 2016). Table 1 shows the comparison of the methods on all utilized datasets. We report the accuracy and also the number of identified key-points.

The best results are obtained using the real part of Gabor wavelets. It is partly due to the larger number of points that are found by this method. On the other hand, ORB achieved the worst results. It is evident mainly on LFW where it also finds very low number of key-points. A possible reason for the lower number of key-poits is lower resolution of the images.

Table 2 compares 3 key-point identification methods with two cluster counts, namely 50 and 100. All key-point reduction types are examined.

The results show that both Gabor wavelet based methods have significantly better accuracy than ORB. The real part of Gabor wavelet is mostly sufficient and achieves slightly better accuracies than magnitude. Moreover, it is computationally less expensive. Based on these experiments, we choose the real part of Gabor wavelet as the best approach and use it in all following experiments.

The next experiment performs a more fine-grained evaluation of how the number of clusters influences the overall accuracy. The three utilized datasets are tested with all three key-point reduction methods.

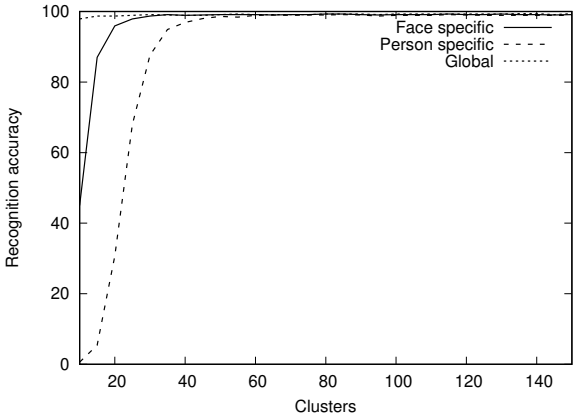

Figure 3: Recognition accuracy in dependence on cluster count evaluated on AR database.

Figure 3 shows the results for the AR database. It is obvious that this database is relatively easy and even a very low number of feature points is enough to reach high accuracy. Cluster counts higher than 40 are sufficient for all key-point reduction method. The global position approach allows using even lower numbers. The highest accuracy of 99.4% is reached using the Global position. However, the results of the two other methods are very close and both are higher than 99%.

Figure 4 reports the results for the challenging LFW database. In this case, the minimal number of clusters is around 60. The recognition accuracy then slightly increases. The highest accuracy of 56.1% is obtained with person specific position and 120 clusters.

The curves for the UFI dataset depicted in Figure 5 have a very similar shape as in the case of the LFW. The best result is 74.2% for the Global Position method with 135 clusters.

Table 1: Comparison of different key-point determination methods.

Dataset	UFI		AR		LFW	
Method	**ACC (%)**	**Points**	**ACC (%)**	**Points**	**ACC (%)**	**Points**
REAL	72.2	330	98.9	413	54.7	352
MAGNITUDE	71.4	275	98.7	333	53.2	290
ORB	68.3	271	95.9	253	41.1	83

Table 2: Comparison of different cluster counts.

Key-point reduction	Face Specific			Person Specific			Global		
Dataset	**UFI**	**AR**	**LFW**	**UFI**	**AR**	**LFW**	**UFI**	**AR**	**LFW**
Method/clusters	**ACC (%)**			**ACC (%)**			**ACC (%)**		
REAL/50	71.1	99.1	52.5	69.1	98.6	51.8	71.7	99.1	51.6
REAL/100	71.7	99.1	54.9	72.4	98.9	55.3	73.4	99.3	53.8
MAGNITUDE/50	71.1	98.7	54.5	65.3	96.3	46.1	71.6	99.1	50.4
MAGNITUDE/100	71.7	98.9	54.4	71.7	98.9	54.2	73.1	99.1	53.9
ORB/50	61.7	94.7	40.8	54.5	92.3	34.1	70.4	98.6	46.8
ORB/100	64.5	95.1	41.0	64.3	95.6	39.0	71.7	98.7	50.5

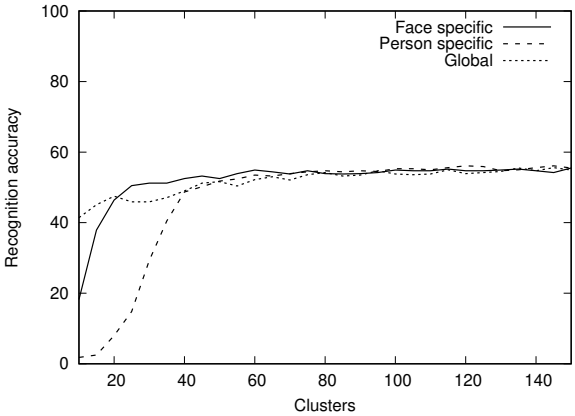

Figure 4: Recognition accuracy in dependence on cluster count evaluated on LFW database.

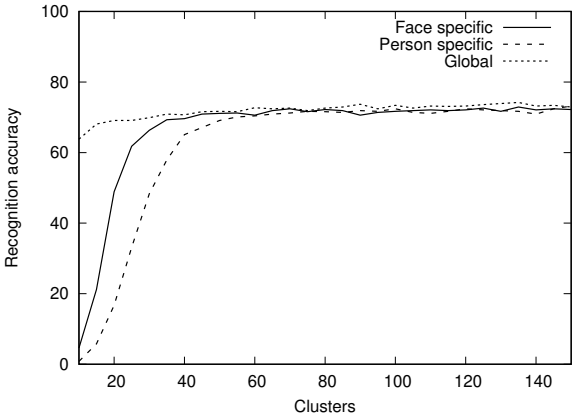

Figure 5: Recognition accuracy in dependence on cluster count evaluated on UFI database.

Table 3 brings the comparison of our results with previously reported LBP based methods. We used the $LBP_{8,2}$ variant. All methods are evaluated with 100 clusters to allow fair comparison.

Table 3: Accuracy (in %) of the proposed methods in comparison with LBP-based approaches.

Method / Dataset	UFI	AR	LFW
FS-LBP	64.0	98.9	52.1
PS-LBP	63.3	98.7	51.0
GL-LBP	63.0	98.7	51.8
FS-POEM	71.7	99.1	54.9
PS-POEM	72.4	98.9	**55.3**
GL-POEM	**73.4**	**99.3**	53.8

The comparison indicates that POEM is superior mainly on the UFI dataset. The results for the AR database are comparable while the difference on the LFW dataset is around 3% in average. We can state that face specific method is superior together with LBP. However, there are differences in the POEM based method. In this case, person specific and global methods perform better than the face specific one.

Table 4 brings a comparison of our best results obtained on the UFI dataset with the results reported in the literature. The results for FS-LBP and SIFT are recomputed because the results reported in the literature were obtained on older versions of the UFI dataset.

6 Conclusion

In this paper we have proposed an extension of the POEM-based face recognition method. It combines automatic detection of feature points and a better matching algorithm with POEM features. We have

Table 4: Accuracy of the proposed method on the UFI dataset in comparison with other approaches reported in the literature.

Method / Dataset	ACC (%)
FS-LBP (Lenc and Král, 2016)	63.96
POEM-HS (Lenc, 2016)	65.95
POEM-HS weighted (Lenc, 2016)	68.93
SIFT (Lenc and Král, 2012)	58.68
M-BNCC (Gaston et al., 2017)	74.55
GL-POEM (proposed)	74.20

also evaluated several aspects of the method and their influence on the resulting accuracy.

The methods were tested on three standard face corpora. The results are consistently better than those of the previously published methods using automatically detected points together with LBP features. Moreover, we were able to reach state-of-the-art accuracy on the UFI dataset.

One of possible improvements is adding weighting also to this method with dynamic feature points. Based on the results of weighting together with methods using HS for face representation, it could bring further increase of the recognition accuracy.

Acknowledgements

This work has been partly supported by the project LO1506 of the Czech Ministry of Education, Youth and Sports and by the Cross-border Cooperation Program Czech Republic - Free State of Bavaria ETS Objective 2014-2020 (project no. 211).

REFERENCES

Ahonen, T., Hadid, A., and Pietikäinen, M. (2004). Face recognition with local binary patterns. In *Computer vision-eccv 2004*, pages 469–481. Springer.

Ahonen, T., Hadid, A., and Pietikainen, M. (2006). Face description with local binary patterns: Application to face recognition. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, (12):2037–2041.

Gaston, J., Ming, J., and Crookes, D. (2017). Unconstrained face identification with multi-scale block-based correlation. pages 1477–1481.

Guo, Z., Zhang, L., and Zhang, D. (2010). A completed modeling of local binary pattern operator for texture classification. *IEEE Transactions on Image Processing*, 19(6):1657–1663.

Huang, G. B., Ramesh, M., Berg, T., and Learned-Miller, E. (2007). Labeled faces in the wild: A database for studying face recognition in unconstrained environments. Technical report, Technical Report 07-49, University of Massachusetts, Amherst.

Lenc, L. (2016). Genetic algorithm for weight optimization in descriptor based face recognition methods. In *Proceedings of the 8th International Conference on Agents and Artificial Intelligence*, pages 330–336. SCITEPRESS-Science and Technology Publications, Lda.

Lenc, L. and Král, P. (2012). Novel matching methods for automatic face recognition using sift. In *IFIP International Conference on Artificial Intelligence Applications and Innovations*, pages 254–263. Springer.

Lenc, L. and Král, P. (2014). Automatically detected feature positions for lbp based face recognition. In *IFIP International Conference on Artificial Intelligence Applications and Innovations*, pages 246–255. Springer.

Lenc, L. and Král, P. (2015). Unconstrained Facial Images: Database for face recognition under real-world conditions. In *14th Mexican International Conference on Artificial Intelligence (MICAI 2015)*, Cuernavaca, Mexico. Springer.

Lenc, L. and Král, P. (2016). Local binary pattern based face recognition with automatically detected fiducial points. *Integrated Computer-Aided Engineering*, 23(2):129–139.

Li, W., Fu, P., and Zhou, L. (2012). Face recognition method based on dynamic threshold local binary pattern. In *Proceedings of the 4th International Conference on Internet Multimedia Computing and Service*, pages 20–24. ACM.

Martinez, A. and Benavente, R. (1998). The AR Face Database. Technical report, Univerzitat Autonòma de Barcelona.

Nanni, L., Lumini, A., and Brahnam, S. (2012). Survey on LBP based texture descriptors for image classification. *Expert Systems with Applications*, 39(3):3634–3641.

Ojala, T., Pietikainen, M., and Harwood, D. (1994). Performance evaluation of texture measures with classification based on kullback discrimination of distributions. In *Pattern Recognition, 1994. Vol. 1-Conference A: Computer Vision & Image Processing., Proceedings of the 12th IAPR International Conference on*, volume 1, pages 582–585. IEEE.

Rosten, E. and Drummond, T. (2006). Machine learning for high-speed corner detection. In *European conference on computer vision*, pages 430–443. Springer.

Rublee, E., Rabaud, V., Konolige, K., and Bradski, G. (2011). Orb: An efficient alternative to sift or surf. In *Computer Vision (ICCV), 2011 IEEE international conference on*, pages 2564–2571. IEEE.

Sanderson, C. and Lovell, B. C. (2009). Multi-region probabilistic histograms for robust and scalable identity inference. In *International Conference on Biometrics*, pages 199–208. Springer.

Tan, X. and Triggs, B. (2010). Enhanced local texture feature sets for face recognition under difficult lighting conditions. *IEEE transactions on image processing*, 19(6):1635–1650.

Vu, N.-S., Dee, H. M., and Caplier, A. (2012). Face recogni-

tion using the POEM descriptor. *Pattern Recognition*, 45(7):2478–2488.

Wang, L. and He, D.-C. (1990). Texture classification using texture spectrum. *Pattern Recognition*, 23(8):905–910.

Wolf, L., Hassner, T., Taigman, Y., et al. (2008). Descriptor based methods in the wild. In *Workshop on Faces in'Real-Life'Images: Detection, Alignment, and Recognition*.

Xu, Y., Fang, X., Li, X., Yang, J., You, J., Liu, H., and Teng, S. (2014). Data uncertainty in face recognition. *IEEE transactions on cybernetics*, 44(10):1950–1961.