

Pixelwise Local Binary Pattern Models of Faces Using Kernel Density Estimation

Timo Ahonen and Matti Pietikäinen

Machine Vision Group, University of Oulu, Finland
{tahonen,mkp}@ee.oulu.fi

Abstract. Local Binary Pattern (LBP) histograms have attained much attention in face image analysis. They have been successfully used in face detection, recognition, verification, facial expression recognition etc. The models for face description have been based on LBP histograms computed within small image blocks. In this work we propose a novel, spatially more precise model, based on kernel density estimation of local LBP distributions. In the experiments we show that this model produces significantly better performance in the face verification task than the earlier models. Furthermore, we show that the use of weighted information fusion from individual pixels based on a linear support vector machine provides with further improvements in performance.

1 Introduction

In the last years, different problems in face image analysis, such as face detection, face recognition and facial expression recognition have received very much attention in computer vision research. These problems are interesting from the viewpoints of basic research aiming to efficient descriptors for facial images and of applications such as surveillance and human-computer interaction [1].

Recently, a face descriptor based on Local Binary Patterns [2] was proposed [3]. In its original form, the method is based on dividing a facial image into small non-overlapping rectangular regions of equal size and describing each region using a local binary pattern histogram. A global description of the face is then formed by concatenating the histograms.

This representation has successfully been applied to face recognition [3], face detection [4], facial expression recognition [5], demographic classification [6] and also other tasks such as general object recognition [7].

Some of the limitations of that method have also been discussed in the literature and different extensions and modifications have been proposed. In [8], Zhang *et al.* used AdaBoost learning algorithm for selecting a set of local blocks and their weights. Then, the LBP methodology was applied to the obtained blocks yielding in smaller feature vector length. Rodriguez and Marcel noticed that the method as such does not suit well for the face verification task and proposed an approach based on adapted, client-specific LBP histograms [9].

Even though the LBP labels are invariant to monotonic mappings of gray scale, illumination changes still pose a challenge to the local binary pattern

based representations. To alleviate this problem, Tan and Triggs proposed a pre-processing chain for illumination normalization. Furthermore, they introduced a three-level version of local patterns and a Hausdorff-like distance between query and gallery label images to replace histogram based approach. [10]

Other proposed enhancements include multi-scale block LBP which considers mean gray values from larger pixel blocks than original LBP [11], using patterns at multiple scales for representation [12] and combining LBP representation with other information such as Gabor filter based recognition [13], [14], [15].

In this work we propose a novel model for estimating local LBP distributions. This new method is based on kernel density estimation in the xy -space. We also propose the use of support vector machine for combining the information coming from different parts of the face as such learning method can be expected to result in higher performance than simple summation.

2 The Local Binary Pattern Operator

The local binary pattern operator [2] was originally designed for texture description. It has nevertheless shown very good performance also in many other application areas, and recently it has been used most widely in facial image description. LBP features are simple and fast to compute, yet LBP histograms have been shown to be discriminative features for face recognition even under some mis-alignment of the facial image [3], [12].

The basic LBP operator labels the pixels of an image by thresholding each 3×3 pixel neighborhood of the input image with the center value, multiplying the thresholded values by powers of two and summing them. Usually the histogram of the resulting labels is then used as an image descriptor.

The operator is extended to use neighborhoods of different sizes [2] by using a circular neighborhood around the center pixel and bilinearly interpolating the pixel values. This allows any radius and number of pixels in the neighborhood.

In the following, the general LBP operator is denoted by $LBP_{a,b}$, where a is the number of sampling points and b is the radius of the circular neighborhood. The operator outputs a labelled image, denoted by $f_l(x, y)$, whose value at integer coordinates (x, y) is one of the possible LBP labels $0, \dots, N - 1$. In this work we do not consider the uniform pattern and rotation invariant extensions of the LBP operator [2], thus the number of different labels produced by $LBP_{a,b}$ is $N = 2^a$.

2.1 Estimating LBP Histograms for Face Description

Most algorithms applying local binary patterns for face description make use of histograms of LBP labels computed within small image blocks. In [3], Ahonen *et al.* proposed dividing the facial image into R non-overlapping rectangular regions each of equal size and computing the histogram $H^r(i)$ independently within each block r

$$H^r(i) = n^r \sum_{x,y} I \{f_l(x, y) = i\} I \{(x, y) \in \text{Region}(r)\}, \quad (1)$$

where $r = 1, \dots, R, i = 0, \dots, N - 1$ and

$$I\{A\} = \begin{cases} 1, & A \text{ is true} \\ 0, & A \text{ is false.} \end{cases}$$

The normalization constants n^r are set so that the sum of the histogram equals 1, and in case of rectangular regions of size of $w \times h$ pixels, the normalization constant is $n^r = 1/(wh)$.

As the amount of training data per client in face verification applications is limited in many cases, Rodriguez and Marcel [9] proposed using maximum a posteriori (MAP) adaptation [16] to achieve more reliable estimate of LBP distributions. In this context, MAP adaptation can be reduced to modelling the client-specific distribution of LBP labels by a weighted sum of histogram obtained from this client and a world model:

$$\hat{H}_C^r(i) = \alpha H_W^r(i) + (1 - \alpha) H_C^r(i) \quad (2)$$

In this equation, $H_W^r(i)$ denotes the histogram for region r of the world model learned from training data independent of the client set, $H_C^r(i)$ is the model for client C and α is a weighting factor that is chosen through cross-validation in the development set.

The problem of model (1) is that finding the optimal division of face image area into R regions is a non-trivial task. Ahonen *et al.* proposed using a grid of rectangular block of equal size which was experimentally shown to be an effective way to encode facial information [3]. So far, also other works applying LBP histograms for face description have used this block based approach, either using a manually fixed grid (*e.g.* [3], [4], [5]) or by finding the division and weights denoting the relative importance of each region through some optimization procedure such as AdaBoost (*e.g.* [6], [8]) In some other works, the histogram based model has been omitted and LBP is used as preprocessing for linear discriminant analysis [17] or the query and gallery label images are compared directly using a Hausdorff like distance measure [10].

Another problem with model (1) is that when the values of $H^r(i)$ are used as estimates of probability distribution of LBP labels for pixels within the region r it is questionable whether the estimate is reliable near the borders of the region.

To address these two problems of the block based approach, we propose replacing it with Kernel Density Estimation (KDE) [18] in xy -space. KDE allows computing an estimate of probability distribution of LBP labels at any image coordinates (x, y) as a normalized sum of kernels positioned at occurrences of each LBP label. Here we use the symmetric Gaussian kernel, so the kernel density estimate of LBP distribution at location (x, y) becomes

$$H^{(x,y)}(i) = n^{(x,y)} \sum_{s,t} I\{f_l(s,t) = i\} \exp \left\{ -\frac{(x-s)^2 + (y-t)^2}{2\sigma^2} \right\}. \quad (3)$$

The difference between block based and kernel density estimation of LBP local occurrence probability is illustrated in Fig. 1.

Again, the normalizing constants $n^{(x,y)}$ are set so that $\sum_i H^{(x,y)}(i) = 1$. If there are more than one label image available, as is the case in computing the

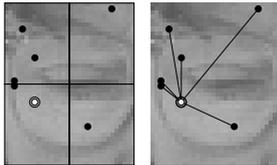


Fig. 1. Comparison of block based and kernel density estimation of LBP histograms in xy -space. Left: In block based estimation the probability of occurrence of a specific LBP label in arbitrary coordinates (x, y) is the number of this label in the region (x, y) belongs to divided by the total number of pixels in the region. Right: In kernel density estimation, the probability estimate is the sum of kernels positioned at occurrences of the label.

world model and sometimes also the client models, this means simply having more samples in the xy -space. In practice the resulting model can be computed by applying (3) to each of the images independently and then averaging over the obtained values.

2.2 Generative Model for LBP Based Face Verification

In this section we briefly review the generative model for LBP based face verification [9]. For full details and derivations, we refer to the original paper.

The biometric identity verification problem is the task of accepting or rejecting the claimed identity C of a subject using a biometric sample X from the subject. The decision is made based on a model θ_C of the client.

In the following, we assume that the sample X consists of the LBP labels computed from the facial image of the subject, i.e. $X = \{f_l(x, y)\}$, where the LBP label image $f_l(x, y)$ is obtained by applying a selected LBP operator to the input face image $f(x, y)$. Furthermore, if block based estimation of LBP probabilities is used, the LBP label image is further divided into local regions and we denote the k -th observation (pixel in the LBP label image) in r -th local region by l_k^r and in this case $X = \{l_k^r\}$. Now, assuming independence of the observations, the probability of observing features $X = \{l_k^r\}$ given model θ_C is

$$P(X|\theta_C) = \prod_{r=1}^R \prod_{k=1}^K \hat{H}_C^r(l_k^r), \quad (4)$$

The decision problem for accepting the claimed identity can then be expressed as a binary function of the log-likelihood ratio

$$\Lambda(X, C) = \log P(X|\theta_C) - \log P(X|\theta_W) \quad (5)$$

$$= \sum_{r=1}^R \sum_{k=1}^K \left[\log(\hat{H}_C^r(l_k^r)) - \log(H_W^r(l_k^r)) \right] \quad (6)$$

The claim of identity C is accepted if $\Lambda(X, C) \geq \tau$ and rejected if $\Lambda(X, C) < \tau$. The decision threshold τ is usually learnt from a development set (see next section).

To increase the performance of the verification system, we propose replacing the block based client models $\hat{H}_C^r(i)$ and world model $H_W^r(i)$ of LBP distributions with the kernel density estimates $\hat{H}_C^{(x,y)}(i)$ and $H_W^{(x,y)}(i)$ computed using (3). The log likelihood ratio then becomes

$$A(X, C) = \sum_{x,y} \left[\log(\hat{H}_C^{(x,y)}(f_l(x, y))) - \log(H_W^{(x,y)}(f_l(x, y))) \right] \quad (7)$$

By denoting $\lambda_C(x, y) = \log(\hat{H}_C^{(x,y)}(f_l(x, y))) - \log(H_W^{(x,y)}(f_l(x, y)))$ we arrive at the decision function

$$D(X, C) = \text{sign} \left\{ \sum_{x,y} \lambda_C(x, y) - \tau \right\}, \quad (8)$$

and now the claim is accepted if $D(X, C) = 1$ and rejected if $D(X, C) = -1$.

2.3 Support Vector Machines for Combining Likelihoods

In essence, the decision function (8) is a sum of log likelihood ratios corresponding to individual pixels. As earlier studies with LBP based face recognition have shown, different facial areas contribute unequally in terms of recognition. In [3], a weighted sum of Chi square distances of LBP histograms from different facial regions outperformed unweighted sum. In that work the weights for individual regions were set based on the performance of each region when used for recognition alone. In [8], both the weights and the region parameters were obtained through Adaboost learning.

Inspired by those results, we propose replacing the sum in decision function (8) by a weighted sum, i.e. a general linear discriminant function, obtaining

$$D(X, C) = \text{sign} \left\{ \sum_{x,y} w^{(x,y)} \lambda_C(x, y) - \tau \right\}. \quad (9)$$

One can easily see that (8) is a special case of this decision function when $w^{(x,y)} = 1$ for all (x, y) .

A convenient way to compute the weights $w(x, y)$ and the threshold τ is to resort to the linear Support Vector Machine [19]. Previously SVMs have been proposed for information fusion for person verification by, *e.g.* Ben-Yacoub *et al.* who used SVM to combine information from face and voice modalities in multi-modal person verification [20]. In face analysis, Heisele *et al.* used a support vector machine for fusing information from component-wise SVMs in face detection and recognition [21].

The linear SVM is obtained by finding such weights $w(x, y)$ that the minimum distance between the decision boundary and the training samples is maximized, i.e. it is a maximum margin classifier. The SVM can be seen as a structural risk minimization solution to the linear classification problem, and thus it is less prone to overtraining than some other learning methods. This is especially



Fig. 2. Example from the BANCA databases. From left: 1. controlled scenario 2. degraded scenario, 3. adverse scenario.

important in our case, since the input dimension is rather high with respect to the number of training samples. Despite that, in our experiments it was observed that the SVM did not show significant overfitting but it performed well also on unseen data.

3 Experiments

The performance of the proposed facial image representation in the face verification task was tested using the BANCA English database [22]. In the experimental setup we followed the procedures of [9] to make the results with kernel density estimation of LBP distributions comparable to block-based model used in that work.

3.1 BANCA Database

To test the performance of the proposed algorithm, we use the English section of BANCA database [22]. The BANCA English database consists of 52 subjects (26 male and 26 female) captured at 12 sessions. These 12 sessions were further divided into 3 scenarios: controlled, degraded and adverse. Fig. 2 shows example images from the different scenarios.

The images of the BANCA database can be used with different protocols defining which part of the data is used for training the subject models and which part is used for testing. We used the following five protocols in the experiments: Mc (Controlled data is used for client training and testing), Ud (Controlled data for training, degraded for testing), Ua (Controlled data for training, adverse for testing), P (Controlled data for training, all data for testing), G (All data for training, all data for testing). A more detailed description of the protocols can be found in [22].

The two possible cases that can occur in identity verification are true access, i.e. the claimed identity of subject is true and impostor access, which means that the claimed identity is false. Furthermore, the two possible error types are false acceptance, in which the system wrongly accepts an impostor and false rejection, in which the system wrongly rejects a true client.

Given a set of test true and impostor access cases, two error measures depending on the decision threshold τ can be defined. The false acceptance rate $P_{FA}(\tau)$ is $P_{FA}(\tau) = \frac{n_{FA}(\tau)}{n_I}$, where $n_{FA}(\tau)$ is the number of false acceptances given decision threshold τ and n_I is the total number of impostor access test cases. The false rejection rate $P_{FR}(\tau)$ is defined as $P_{FR}(\tau) = \frac{n_{FR}(\tau)}{n_T}$, where $n_{FR}(\tau)$ is the number of false rejections and n_T is the total number of true client test cases. In our experiments, we measure the system performance using half total error rate $HTER(\tau)$, which is $HTER(\tau) = \frac{P_{FA}(\tau) + P_{FR}(\tau)}{2}$.

The performance measure $HTER(\tau)$ depends on the decision threshold τ . To avoid bias in the reported results, the threshold should not be computed using the evaluation set but the performance should be assessed by first finding a threshold that produces lowest $HTER$ on a development set and then reporting the $HTER$ on an evaluation set obtained using this threshold. Also, other parameters for the method to be evaluated should be selected in a similar fashion. For this reason, the BANCA database is divided into 2 groups of 26 subjects (13 women, 13 men) each. The decision threshold τ and other parameters can then be selected by optimizing the performance on group 1. Next, group 2 is used to evaluate the performance with these parameters and vice versa. The average $HTER$ of g1 (using g2 as development) and g2 (using g1 as development) is reported as performance measure for each protocol.

3.2 Experimental Setup

The images in the BANCA database were first geometrically normalized to size 84×68 pixels using the eye coordinates provided with the images.

In the 1st experiment, the $LBP_{4,2}$ operator was applied to the normalized images which resulted in labelled images of size 80×64 pixels. The world model $H_W^{(x,y)}(i)$ and client models $H_C^{(x,y)}(i)$ were then constructed using equation (3) and client model adaptation was done using (2). The parameters σ for KDE and α for model adaptation as well as the decision threshold τ were obtained by minimizing the $HTER$ in a development set. Using these parameters, the models for the evaluation set were classified using decision rule (8).

In the 2nd experiment the procedure was the same as in experiment 1 with the difference that the weighted fusion rule (9) was used instead of (8). In each protocol and group, the same parameters σ and α as in experiment 1 were used and the weights were obtained through training a SVM in the development set. Again, the $HTER$ obtained with these weights on the separate evaluation set was reported.

In experiments 3 and 4 we tested the effect of illumination normalization. It has been reported that the performance of LBP based face recognition can be enhanced with normalization procedure consisting of gamma correction, difference of gaussian filtering and contrast equalization [10]. In experiments 3 and 4 this preprocessing of [10] was applied to the images prior to applying LBP but otherwise the procedure was kept completely the same.

3.3 Results

The half total error rates of the proposed method under different protocols of the BANCA database are reported in Table 1.

Table 1. Half Total Error rates for different methods on BANCA protocols Mc, Ud, Ua, P and G

Method	Mc	Ud	Ua	P	G
0. LBP-block [9]	7.3	10.7	22.6	19.2	5.0
Best in [9]	4.9	10.7	17.6	14.8	5.0
1. LBP-KDE	4.3	6.4	18.1	17.6	4.0
2. Weighted LBP-KDE	3.7	6.6	15.1	11.6	3.0
3. PP+LBP-KDE	3.0	5.9	8.7	10.1	1.9
4. Weighted PP+LBP-KDE	3.2	7.1	10.4	9.1	1.5

For comparison, the error rates for LBP face verification with block based estimation of LBP distributions [9] is reported on the 1st row of the table. In that work, LBP was compared to two other methods (LDA/NC and DCT-mod2/GMM), and the 2nd row of the table shows the best result among these three methods.

The results from Experiment 1 show that switching from block based estimation (0) to kernel density estimation (1) improves the accuracy of LBP face verification in all the BANCA protocols. When it is further combined with weighted fusion rule trained with SVM (2), it produces better results than the comparison methods in all the protocols. With KDE the error rates drop in best cases to about half of those of LBP-block.

As expected, illumination normalization preprocessing (3 and 4) is useful also in case of BANCA images and especially in protocols Ua and P with test images containing severe illumination changes compared to training images. In the Mc protocol the effect of preprocessing is smaller as both the training and testing images have been taken under controlled conditions.

When assessing the result of support vector machine based weighted fusion of log-likelihoods from individual pixels, it can be noted that in case of non-preprocessed images, introducing weights gives better results in all the protocols except Ud. On the other hand, with preprocessed images, the weights obtained with SVM actually worsen the results in three protocols (Mc, Ud and Ua). In the P and G protocols the development set is larger than in MC, Ud and Ua, and in those two protocols the SVM performs best, which indicates that the SVM training clearly benefits from the higher amount of training data available.

4 Conclusions

In this work we proposed a novel method for estimating the local distributions of LBP labels for face description. The method is based on kernel density estimation

in xy -space, and it provides much more spatial accuracy than the earlier, block-based methods.

The kernel density estimation based method alleviates two problems related to computing histograms within local blocks. First, finding an optimal way to divide the facial area into local blocks has proven to be a difficult problem. Second, the block histogram estimates of LBP label probabilities are likely to be unreliable near the block boundaries. Both these problems are avoided when using the proposed method.

We also proposed and evaluated the use of Support Vector Machine in information fusion of log-likelihoods corresponding to individual pixels for the binary classification task for identity verification. The SVM is a learning method for finding the weights corresponding to usefulness of information coming from that face area for the verification task. As it is less prone to overfitting than many other learning methods, it is well suited for this task as the input dimension is rather high.

In the experiments we showed that kernel density estimation clearly improves the verification performance in the BANCA dataset in comparison to block based method. Moreover, the SVM was shown to further enhance the results.

Here we applied kernel density estimation to LBP description of faces for the face verification task. However, we believe that this same methodology is applicable to other tasks such as face detection and alignment or in other application areas where local estimation of local binary pattern distributions are used. Also, the method is likely to be useful with other local descriptors than LBP. Further research includes exploring these areas, investigating other methods for fusing information from individual pixels, and experiments with larger datasets.

Acknowledgments. This work has been performed with partial support from the MOBIO project (contract number IST-214324), which belongs to the 7th Framework Research Programme of the European Union. The authors want to thank Dr. Abdenour Hadid for helpful comments in preparation of the manuscript.

References

1. Li, S.Z., Jain, A.K. (eds.): Handbook of Face Recognition. Springer, Heidelberg (2005)
2. Ojala, T., Pietikäinen, M., Mäenpää, T.: Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Trans. Pattern Analysis and Machine Intelligence* 24(7), 971–987 (2002)
3. Ahonen, T., Hadid, A., Pietikäinen, M.: Face description with local binary patterns: Application to face recognition. *IEEE Trans. Pattern Analysis and Machine Intelligence* 28(12), 2037–2041 (2006)
4. Hadid, A., Pietikäinen, M., Ahonen, T.: A discriminative feature space for detecting and recognizing faces. In: Proc. Conf. Computer Vision and Pattern Recognition (CVPR 2004), vol. 2, pp. 797–804 (2004)
5. Shan, C., Gong, S., McOwan, P.W.: Robust facial expression recognition using local binary patterns. In: Proc. IEEE Int. Conf. on Image Processing (ICIP 2005), vol. II, pp. 914–917 (2005)

6. Yang, Z., Ai, H.: Demographic classification with local binary patterns. In: Lee, S.-W., Li, S.Z. (eds.) ICB 2007. LNCS, vol. 4642, pp. 464–473. Springer, Heidelberg (2007)
7. Zhang, H., Gao, W., Chen, X., Zhao, D.: Object detection using spatial histogram features. *Image Vision Comput.* 24(4), 327–341 (2006)
8. Zhang, G., Huang, X., Li, S.Z., Wang, Y., Wu, X.: Boosting local binary pattern (LBP)-based face recognition. In: Li, S.Z., Lai, J.-H., Tan, T., Feng, G.-C., Wang, Y. (eds.) SINOBIOMETRICS 2004. LNCS, vol. 3338, pp. 179–186. Springer, Heidelberg (2004)
9. Rodriguez, Y., Marcel, S.: Face authentication using adapted local binary pattern histograms. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006. LNCS, vol. 3954, pp. 321–332. Springer, Heidelberg (2006)
10. Tan, X., Triggs, B.: Enhanced local texture feature sets for face recognition under difficult lighting conditions. In: Zhou, S.K., Zhao, W., Tang, X., Gong, S. (eds.) AMFG 2007. LNCS, vol. 4778, pp. 168–182. Springer, Heidelberg (2007)
11. Liao, S., Zhu, X., Lei, Z., Zhang, L., Li, S.Z.: Learning multi-scale block local binary patterns for face recognition. In: Lee, S.-W., Li, S.Z. (eds.) ICB 2007. LNCS, vol. 4642, pp. 828–837. Springer, Heidelberg (2007)
12. Chan, C.H., Kittler, J., Messer, K.: Multi-scale local binary pattern histograms for face recognition. In: Lee, S.-W., Li, S.Z. (eds.) ICB 2007. LNCS, vol. 4642, pp. 809–818. Springer, Heidelberg (2007)
13. Yan, S., Wang, H., Tang, X., Huang, T.S.: Exploring feature descriptors for face recognition. In: *Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP 2007)* (2007)
14. Zhang, W., Shan, S., Gao, W., Chen, X., Zhang, H.: Local gabor binary pattern histogram sequence (LGBPHS): A novel non-statistical model for face representation and recognition. In: *Proc. Tenth IEEE Int. Conf. on Computer Vision (ICCV 2005)*, vol. 1, pp. 786–791 (2005)
15. Tan, X., Triggs, B.: Fusing Gabor and LBP feature sets for kernel-based face recognition. In: Zhou, S.K., Zhao, W., Tang, X., Gong, S. (eds.) AMFG 2007. LNCS, vol. 4778, pp. 235–249. Springer, Heidelberg (2007)
16. Gauvain, J.L., Lee, C.H.: Maximum a posteriori estimation for multivariate gaussian mixture observations of markov chains. *IEEE Trans. Speech and Audio Processing* 2(2), 291–298 (1994)
17. Heusch, G., Rodriguez, Y., Marcel, S.: Local binary patterns as an image preprocessing for face authentication. In: *Proc. IEEE Int. Conf. on Automatic Face and Gesture Recognition (FG 2006)*, pp. 9–14 (2006)
18. Scott, D.W.: *Multivariate Density Estimation: Theory, Practice, and Visualization*. Wiley-Interscience, Hoboken (1992)
19. Vapnik, V.N.: *Statistical Learning Theory*. Wiley-Interscience, Hoboken (1998)
20. Ben-Yacoub, S., Abdeljaoued, Y., Mayoraz, E.: Fusion of face and speech data for person identity verification. *IEEE Trans. Neural Networks* 10(5), 1065–1074 (1999)
21. Heisele, B., Serre, T., Poggio, T.: A component-based framework for face detection and identification. *Int. Journal of Computer Vision* 74(2), 167–181 (2007)
22. Bailly-Bailli re, E., Bengio, S., Bimbot, F., Hamouz, M., Kittler, J., Mari thoz, J., Matas, J., Messer, K., Popovici, V., Por e, F., Ru z, B., Thiran, J.P.: The BANCA database and evaluation protocol. In: Kittler, J., Nixon, M.S. (eds.) AVBPA 2003. LNCS, vol. 2688, pp. 625–638. Springer, Heidelberg (2003)