A Discriminative Feature Space for Detecting and Recognizing Faces

Abdenour Hadid, Matti Pietikäinen and Timo Ahonen
Machine Vision Group
Infotech Oulu and Dept. of Electrical and Information Engineering
P.O. Box 4500 FIN-90014 University of Oulu, Finland
`http://www.ee.oulu.fi/mvg`

Abstract

In this paper, we introduce a novel discriminative feature space which is efficient not only for face detection but also for recognition. The face representation is based on local binary patterns (LBP) and consists of encoding both local and global facial characteristics into a compact feature histogram. The proposed representation is invariant with respect to monotonic gray scale transformations and can be derived in a single scan through the image. Considering the derived feature space, a second-degree polynomial kernel SVM classifier was trained to detect frontal faces in gray scale images. Experimental results using several complex images show that the proposed approach performs favorably compared to the state-of-the-art methods. Additionally, experiments with detecting and recognizing low-resolution faces from video sequences were carried out, demonstrating that the same facial representation can be efficiently used for both detection and recognition.

1. Introduction

Developing face detection and recognition systems involves two crucial aspects: facial representation and classifier design. The aim of facial representation is to derive a set of features from raw face images which minimizes the intra-class variations and maximizes the extra-class ones. In face recognition, a class refers to the different face instances of an individual while face detection can be viewed as a two-class recognition problem in which a pattern is classified as being a ”face” or ”nonface”. Obviously, if inadequate features are adopted, even the most sophisticated classifiers will fail to accomplish the given recognition task. Therefore, it is important to derive features which:
(i) Discriminate different classes well while tolerating within-class variations,
(ii) Can be easily extracted from raw face images in order to allow fast processing,
(iii) Lie in a low dimensional space (short vector length) in order to avoid a computationally expensive classifier.

Naturally, it is not obvious to find features which simultaneously meet all these criteria because of the large variability in facial appearances due to different factors such as scale, orientation, pose, facial expressions, lighting conditions, presence of glasses etc. Many features have been proposed and applied to face image analysis. For instance, Gabor wavelet features are used for face recognition in the elastic bunch graph matching algorithm (EBGM) [19]. Although good results have been obtained, the algorithm uses complex analysis to extract a large set of Gabor wavelet coefficients. Other features, such as those using PCA [16] and LDA [3] subspaces, have also been considered. Such features are simple to compute, but their discriminative power is limited [11]. To overcome the main limitation of the PCA representation, Local Feature Analysis (LFA) is presented in [10]. For face detection, the normalized pixel values [6] [20] and Haar-like features [18] are the most considered ones. Heisele et al. reported that gradient and wavelet features do not perform better than the normalized pixel values [6], while Viola and Jones developed an efficient face detection system using Haar-like features and AdaBoost as a fast learning algorithm [18].

Despite the fact that in most face recognition scenarios a detection phase precedes recognition, not much work has addressed the issue of defining the same discriminative features for both tasks. The existing facial representations are dedicated either for detection or for recognition. To the best of our knowledge, the attempts to unify these tasks mainly consisted of using the recognition results as feedback to the detector and/or vice versa, rather than using the same features for both tasks. In this paper, we introduce a novel discriminative feature space which can be used for both detection and recognition. Our approach is based on local binary patterns (LBP) [8] and consists of dividing the facial image into a set of regions from which LBP feature histograms (representing texture contents within regions) are computed and concatenated into a single histogram. It should be noted that the basic LBP features have performed very well in various applications, including texture classification and segmentation, image retrieval and surface inspection. The idea of using them here is motivated by the fact that face images can be seen as a composition of micro-patterns, such as those shown in Fig. 2, which can be well described by LBP.

In our recent work, we obtained outstanding results in face recognition using an LPB-based method in which the face image was divided into many small non-overlapping blocks [1]. Our extensive experiments clearly showed the superiority of the LBP-based approach over all considered methods (PCA, Bayesian intra/extrapersonal classifier, EBGM) on FERET tests which included testing the robustness of the method against different facial expressions, illumination changes and aging of the subjects. A limitation of the facial representation in [1] is that it cannot be used for small-sized face images common in many face detection and recognition problems. For example, in the FERET tests the images have a resolution of 130x150 pixels and they were typically divided into 49 blocks, leading to a feature vector of thousands of elements.

In many applications, such as the detection and recognition of faces from photographs or video sequences, the faces can be on the order of 20x20 pixels. In this paper, we will propose an LBP-based representation which is suitable for low-resolution images and has a short feature vector needed for fast processing. Using these features a second-degree polynomial kernel SVM classifier is trained to detect frontal faces in complex gray scale images. Additionally, experiments with detecting and recognizing low-resolution faces from video sequences are carried out to demonstrate that the same facial representation can be efficiently used for both detection and recognition.

2 Face Description with Local Binary Patterns

The original LBP operator labels the pixels of an image by thresholding the 3*3 neighborhood of each pixel with the value of the center pixel and considering the results as a binary number. Fig. 1 shows an example of LBP calculation [8]. The 256-bin histogram of the labels computed over a region can be used as a texture descriptor. Each bin (LBP code) can be regarded as a micro-texton. Local primitives which are codified by these bins include different types of curved edges, spots, flat areas etc. Fig. 2 shows some examples.

The LPB operator has been extended to consider different neighborhood sizes [8]. For example, the operator $LBP_{4,1}$ uses only 4 neighbors while $LBP_{16,2}$ considers the 16 neighbors on a circle of radius 2. In general, the operator $LBP_{P,R}$ refers to a neighborhood size of P equally spaced pixels on a circle of radius R that form a circularly sym-

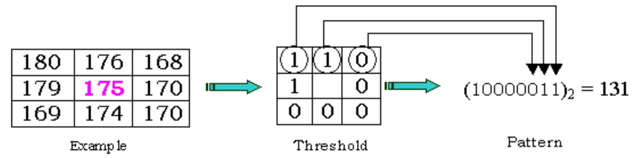

Figure 1: **Example of LBP calculation**

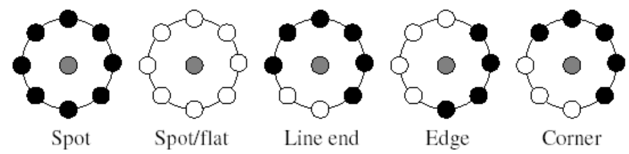

Figure 2: **Examples of texture primitives which can be detected by LBP (white circles represent ones and black cirlces zeros)**

metric neighbor set. $LBP_{P,R}$ produces 2^P different output values, corresponding to the 2^P different binary patterns that can be formed by the P pixels in the neighbor set. It has been shown that certain bins contain more information than others. Therefore, it is possible to use only a subset of the 2^P local binary patterns to describe the textured images. Ojala et al. [8] defined these fundamental patterns (called also "uniform" patterns) as those with a small number of bitwise transitions from 0 to 1 and vice versa. For example, 00000000 and 11111111 contain 0 transition while 00000110 and 01111000 contain 2 transitions and so on. Accumulating the patterns which have more than 2 transitions into a single bin yields an LPB descriptor, denoted $LBP^{u2}_{P,R}$, with less than 2^P bins.

An LBP description computed over the whole face image encodes only the occurrences of the micro-patterns without any indication about their locations. To overcome this effect, we introduced in [1] a representation which consists of dividing the face image into several (e.g. 49) non-overlapping blocks from which the local binary pattern histograms are computed (using the $LBP^{u2}_{8,2}$ operator) and concatenating them into a single histogram. In such a representation, the texture of facial regions is encoded by the LBP while the shape of the face is recovered by the concatenation of different local histograms. However, this representation is more adequate for larger sized images (such as the FERET images) and leads to a relatively long feature vector typically containing thousands of elements. Therefore, we will propose here a new facial representation which is efficient for low-resolution images. A specific of this representation is the use of overlapping regions and a 4-neighborhood LBP operator ($LBP_{4,1}$) to avoid statistical unreliability due to long histograms computed over small regions. Additionally, we will enhance the holistic descrip-

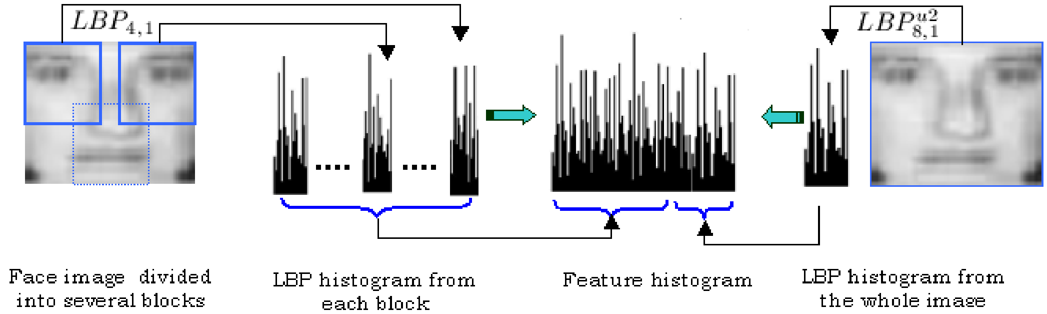

Figure 3: **Facial representation: a face image is represented by a concatenation of a global and a set of local LBP histograms**

tion of a face by including the global LBP histogram computed over the whole face image. We considered 19x19 as the standard resolution in our experiments and derive the LBP facial representation as follows:

We divide a 19x19 face image into 9 overlapping regions of 10x10 pixels (overlapping size=4 pixels). From each region, we compute a 16-bin histogram using the $LBP_{4,1}$ operator and concatenate the results into a single 144-bin histogram. Additionally, we apply $LBP_{8,1}^{u2}$ to the whole 19*19 face image and derive a 59-bin histogram which is added to the 144 bins previously computed. Thus, we obtain a (59+144=203)-bin histogram as a face representation (see Fig. 3).

3 Application to Face Detection

3.1 Background

In recent years, many methods for detecting faces have been proposed [20]. Among these methods, those relying on training sets to capture the large variability in facial appearances have attracted much attention and demonstrated the best results. Generally, these methods scan an input image at all possible locations and scales then classify the subwindows either as face or nonface. For instance, Sung and Poggio [15] considered Gaussian clusters to model the distribution of face and nonface patterns and then used a multilayer perceptron for detection. Rowley et al. [13] used a retinally connected neural network when scanning an image to decide whether the examined small windows are faces or not. In [12], the authors proposed a SNOW learning structure for face detection. Viola and Jones [18] presented an efficient detection scheme using Haar-like features and AdaBoost as a fast training algorithm. Schneiderman and Kanade [14] described a naive Bayes classifier to estimate the joint probability of local appearances and positions of subregions of the face at multiple resolutions. In [9], frontal faces are detected by SVM with a polynomial kernel.

Most of these algorithms use raw pixel data as features. Such inputs can achieve surprisingly high recognition rates [6], but they are sensitive to all kinds of changes and require complex classifiers. Instead, Viola and Jones used Haar-like features and the AdaBoost learning algorithm to overcome problems caused by the large number of extracted features.

3.2 Our Approach: Using LBP Features

We considered LBP features as a facial representation and built a face detection system. As discussed in Section 1, there are two essential issues for developing such a system: facial representation and classifier design. We chose an SVM (support vector machine) classifier [17] since it is well founded in statistical learning theory and has been successfully applied to various object detection tasks in computer vision. In short, given training samples (face and nonface images) represented by their extracted LBP features, an SVM classifier finds the separating hyperplane that has maximum distance to the closest points of the training set. These closest points are called support vectors. To perform a nonlinear separation, the input space is mapped onto a higher dimensional space using Kernel functions. In our approach, to detect faces in a given target image, a 19x19 subwindow scans the image at different scales and locations. We considered a downsampling rate of 1.2 and a moving scan of 2 pixels. At each iteration, the LBP representation LBP_x is computed from the subwindow and fed to the SVM classifier to determine whether it is a face or not. The classifier decides on the ”faceness” of the subwindow according to the sign of the following function:

$$F(LBP_x) = Sgn(\sum_{i=1}^{l} \alpha_i y_i K(LBP_x, LBP_{t_i}) + b) \quad (1)$$

where LBP_{t_i} is the LBP representation of the training sample t_i, y_i is 1 or -1 depending on whether t_i is a positive or negative sample (face or nonface), l is the number of samples, b is a scaler (bias), and $K(.,.)$ the second degree polynomial kernel function defined by:

$$K(LBP_x, LBP_{t_i}) = (1 + LBP_x \cdot LBP_{t_i})^2, \quad (2)$$

where α_i are the parameters of the SVM classifier, recovered by solving the following quadratic programming problem:

$$Max(\sum_{i=1}^{l} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{l} \alpha_i \alpha_j y_i y_j K(LBP_{t_i}, LBP_{t_j})) \quad (3)$$

$$Subject\ to \underbrace{\left\{ \sum_{i=1}^{l} \alpha_i y_i = 0 \right\}}_{0 \le \alpha_i \le C} \quad (4)$$

where C is the cost of constrain violation during the training process which is fixed to 5 in our experiments.

Additionally, given the results of the SVM classifier, we perform a set of heuristics to merge multiple detections and remove the false ones. For a given detected window, we count the number of detections within a neighborhood of 19x19 pixels (each detected window is represented by its center). The detections are removed if their number is less than 3. Otherwise, we merge them and keep only the one with the highest SVM output.

3.3 Training Data

Because we are building an appearance-based detection scheme, large training sets of images are needed in order to capture the variability of facial appearances. For this purpose, we collected a set of 1000 face images. To increase the number of samples, we added the mirror images to obtain a set of 2000 face patterns. Some examples of these training samples are shown in Fig. 4. Additionally,

Figure 4: **Examples of training face images**

to enable the system to also detect rotated faces (in-plane rotation), we extended the set by rotating the training face images by $\pm 18^o$, $\pm 12^o$, $\pm 6^o$. Fig. 5 shows an example of a face image and the artificially generated samples. Overall, we obtained a training set of 12000 faces.

To collect nonface patterns, we used the ”boostrap” strategy in five iterations [15]. First, we randomly extracted 6000 patterns from a set of natural images which do not

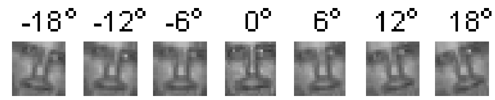

Figure 5: **Examples of virtual faces, generated by rotating a face image**

contain faces (an example is shown in Fig. 6). Then, at each iteration we trained the system, run the face detector, and collected all those nonface patterns that were wrongly classified as faces and used them for training. Overall, we obtained $6000 + 14560$ nonface patterns as negative training examples.

Figure 6: **Example of a natural image from which negative training samples are extracted**

3.4 Experimental Results

From the collected training sets, we extracted the proposed facial representations (as described in Section 2). Then, we used these features as inputs to the SVM classifier and trained the face detector. The system was run on several images from different sources to detect faces. Fig. 7 shows some detection examples. It can be seen that most of the upright frontal faces are detected. For instance, Fig. 7.G shows perfect detections. In Fig. 7.F, only one face is missed by the system. This missing is due to occlusion. A similar situation is shown in Fig. 7.A in which the missed face is due to a large in-plane rotation. Since the system is trained to detect only in-plane rotated faces up to $\pm 18^o$, it succeeded to find the slightly rotated faces in Fig. 7.C, Fig. 7.D and Fig. 7.H and failed to detect largely rotated ones (as those in 7.E and 7.C). A false positive is shown in Fig. 7.E while a false negative is shown in Fig. 7.D. Notice that this false negative is expected since the face is pose-angled (i.e not in frontal position). These examples summarize the main aspects of our detector using images from different sources.

In order to report quantitative results and compare them against those of the state-of-the-art algorithms, we considered the test images from the MIT-CMU sets [13] that are used with the Bayesian Discriminating Features (BDF) method in [7]. There are 80 images containing 227 im-

Figure 7: **Detection examples in several images from different sources. The images A, B, F, G and H are from the subset of MIT-CMU tests. They belong to the 80 images considered for comparison. The images C, D and E are from the World Wide Web. Note: excellent detections of upright faces in A, F and G; detections under slight in-plane rotation in A, C and H; missed face in F because of occlusion, missed faces in C, E and A because of large in-plane rotation; missed face in D because of a pose-angled face; and a false detection in E**

ages. Some of these images are shown in Figs. 7.(A, B, F, G and H). Table 1 presents the performance of our face detector and those of two other approaches: BDF [7] and Schneiderman-Kanade [14]. We can see (from the 1st, 2nd and 6th rows of Table 1) that our approach compares favorably against the comparative approaches. The LBP-based method succeeded in detecting 221 faces without any false positives. Some missing faces are mainly due to occlusion (see an example in Fig. 7.F) and large in-plane rotation (Fig. 7.A). Notice that a reason for the lower performance of the Schneiderman-Kanade approach is that this system is more general, not only dedicated to frontal faces but also to faces in different poses.

Additionally, if the detected faces are to be fed to a recognition/verification phase, then one may tolerate some false detections since it is unlikely that these images will be accepted as those of an individual (therefore they will be rejected). In such a context and by tolerating only 13 false detections, our face detector performed even better as it succeeded in detecting 222 faces among 227 (the system detected the missed face in Fig. 7.A despite the large in-plane rotation). The 7th row of Table 1 presents this performance.

In order to further investigate the discriminative power of the LBP based facial representation, we implemented a similar SVM based face detector using different features as inputs and then compared the results to those obtained using the proposed LBP features. We chose the normalized pixel features as inputs since it has been shown that such features perform better than the gradient and wavelet based ones when using an SVM classifier [6]. We trained the system using the same training samples as described in Section 3.3. Table 1 (3rd row) shows the performance of using normalized pixel values as features. Although the results are quite good as 213 faces among 227 were detected, still our approach (i) performed better (comparison between the 3rd row and 6th row in Table 1) (ii) used a shorter feature vector (203 versus 361) (iii) did not require histogram equalization (iv) and required a smaller number of support vectors (408 versus 512).

Analyzing the LBP features and investigating the usefulness of dividing the facial images into regions, we noticed that computing the LBP features only from the whole images (59 bins) yields a low detection rate of 20.7% (see 4th row in Table 1). This is expected since such a representation encodes only the occurrences of the micro-patterns without any indication about their locations. However, dividing the facial images into regions and calculating the LBP features from these regions yielded a good result (see the 5th row in Table 1). Combining both representations further enhances the detection performance.

Table 1: **Comparative performance of detecting 227 faces in 80 images**

Method	Face detected	False detections	Detection rates
Schneiderman-Kanade(1.0, 1.0) [14]	218	41	96.0 %
BDF Method [7]	221	1	97.4 %
Normalized Pixel features [6]	213	6	93.8 %
LBP representation : $LBP^{u2}_{8,1}$ (59 bins)	47	184	20.7 %
LBP representation : $LBP_{4,1}$ (144 bins)	211	9	92.9 %
LBP representation : $LBP_{4,1}$+$LBP^{u2}_{8,1}$ (203 bins)	**221**	**0**	97.4 %
LBP representation : $LBP_{4,1}$+$LBP^{u2}_{8,1}$ (203 bins)	222	13	97.8 %

4 Application to Face Recognition from Video Sequences

In this section, we investigate the usefulness of the same representation in discriminating faces of different subjects in addition to discriminating faces from non-face patterns. For this purpose, we chose the problem of face recognition from video sequences because it is a challenging task as the images are generally of low resolution.

4.1 Experimental Data

For our experiments, we chose the MoBo (motion of body) video database [4] which contains 96 face sequences of 24 different subjects walking on a treadmill. Four different walking situations are considered: slow walking, fast walking, incline walking and carrying a ball. Each sequence consists of 300 frames. First, we applied the proposed detection scheme to detect faces in the video frames. Exploiting the fact that there is only one face in each frame yielded a detection rate of 100% without any false positives. Fig. 8 shows two example frames and also the detected faces. All the extracted face images were normalized to 19x19 pixels and considered in face recognition experiments.

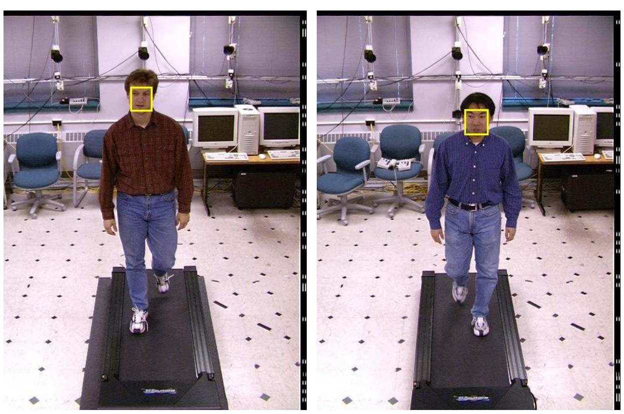

Figure 8: **Detection examples from MoBo frames**

4.2 Face Recognition Experiments

An appearance-based face recognition system was built to compare the performances of three different approaches: Principal Component Analysis (PCA) [16], Linear Discriminant Analysis (LDA) [3] and LBP. For the LBP based approach, we considered the same features as in face detection (203-bin histograms) and adopted χ^2 (Chi-square) as a dissimilarity metric for comparing a target face histogram S to a model histogram M:

$$\chi^2(S,M) = \sum_{i=0}^{202} \frac{(S_i - M_i)^2}{S_i + M_i} \tag{5}$$

Since the data consists of face sequences in 4 different situations, we considered one situation for training and the others for testing. We report the average recognition rates for the 4 combinations: 1 training situation and 3 testing situations. For a given situation, we extracted K face models from each training face sequence G:

$$G = \{G_{face_1}, G_{face_2}, ..., G_{face_N}\}, \tag{6}$$

where $N = 300$. To determine the identity of a test face sequence B

$$B = \{B_{face_1}, B_{face_2}, ..., B_{face_N}\}, \tag{7}$$

we used the probabilistic voting strategy over all frames in B. The probabilistic voting strategy consists of combining the recognition confidences in every frame to decide on the person identity in the target video B.

The extraction of the face models is performed using an unsupervised learning algorithm. The approach is based on applying the locally linear embedding algorithm (LLE) to the raw feature data and then performing K-means clustering in the obtained low dimensional space. We extracted $K = 5$ face models from each training sequence and used them in the appearance-based face recognition. Details about the model extraction process can be found in [5].

Table 2 presents the recognition rates for the different approaches. The results clearly show that the proposed facial representation is efficient for face recognition as it outperformed the PCA and LDA based approaches. This confirms that the same representation can be used for both detection and recognition.

Table 2: **Comparative performance of recognizing 19x19 face images using PCA, LDA and LBP**

Method	PCA	LDA	LBP
Recognition Rate	76.3%	69.5 %	**83.9 %**

5 Discussion and Conclusion

A novel discriminative feature space was proposed which is efficient not only for face detection but also for recognition. The representation is based on local binary patterns and consists of encoding, into an enhanced feature histogram, both shape and texture information of the facial image. The representation can be derived in a single scan through the image and does not require histogram equalization since it is invariant with respect to monotonic gray scale transformations.

Considering the proposed representation, we trained a 2nd degree polynomial kernel SVM to detect frontal faces in gray scale images. The experimental results clearly showed the validity of our approach which compared favorably against the state-of-the-art algorithms. Additionally, by comparing our results to those obtained using normalized pixel values as inputs to the SVM classifier, we confirmed the efficiency of an LBP-based facial representation. Indeed, the results showed that: (i) the proposed LBP features are more discriminative than the normalized pixel values (comparison between the 3rd row and 6th row in Table 1), (ii) The proposed representation is more compact as, for 19x19 face images, we derived a 203-element feature vector while the raw pixel features yield a vector of 361 elements, and (iii) Our approach required a smaller number of support vectors (408 versus 512). Although we focused the experiments on detecting frontal faces, one may adopt a similar facial representation to perform multi-view face detection by considering also non-frontal training face images. We are currently investigating this issue.

In the derivation of the facial representation, we computed local LBP feature histograms from image regions and concatenated the results to the global LBP histogram computed over the whole face image. An alternative is to consider a coarse-to-finer detection strategy. The idea consists of defining a two-level hierarchical LBP representation. In the first level, LBP histograms extracted from the whole image are used to train an SVM classifier. The goal of this first classifier is to filter the patterns and keep only the most probable face images. Then, a finer face representation, extracted from the image regions, can be used to select the face patterns among these candidates.

When dividing the facial images into several regions, we gave an equal weight to the contribution of each region. However, one may use different weights, depending on the role of the given regions in detection/recognition. For instance, since the eye regions are important for recognition, a high weight can be attributed to the corresponding regions. Such a procedure enhanced the facial representation in [1].

We adopted the proposed facial representation to recognize low-resolution faces from video sequneces. First, we applied our face detector to segment faces from the video frames. Using the information that there is only one face in each frame yielded a detection rate of 100 % without any false positives. We used these extracted faces and built an appearance based face recognition system. We compared the recognition results obtained with our facial representation to those obtained with PCA and LDA approaches. The results clearly showed the superiority of the LBP-based recognition (83.9% versus 76.3% and 69.5% for PCA and LDA, respectively). This confirmed the validity and efficiency of using the same facial representation for both detection and recognition. This property is interesting both from practical and cognitive points of view. A unified feature space allows combining face detection and recognition within a single framework in which the two tasks (detection and recognition) can support each other. From the cognitive point of view, there appears to be no evidence that different features are used for detecting and recognizing faces. Although these two tasks are unified in the human visual system, it is still unclear how this is done.

Our main goal in this paper was to show the high discriminative power of the proposed facial representation. Therefore, we have not focused on speeding up the face detector. In our future work, we plan to investigate the use of different, simpler and faster classifiers than SVMs for face detection. For instance, Viola and Jones [18] have suggested sub-linear weak classifiers combined via boosting, while Elad et al. used a Maximal Rejection Classifier(MRC) [2].

Acknowledgment

We would like to thank Dr. R. Gross for providing us the MoBo database. This research was sponsored by the Academy of Finland and the Finnish Graduate School in Electronics, Telecommunications and Automation (GETA).

References

[1] T. Ahonen, A. Hadid, and M. Pietikäinen. Face recognition with local binary patterns. In *Proc. the 8th Eu-*

ropean Conference on Computer Vision (to appear), May 2004.

[2] M. Elad, Y. Hel-Or, and R. Keshet. Rejection based classifer for face detection. *Pattern Recognition Letters*, 23:1459–1471, 2002.

[3] K. Etemad and R. Chellappa. Discriminant analysis for recognition of human face images. *Journal of the Optical Society of America*, 14:1724–1733, 1997.

[4] R. Gross and J. Shi. The CMU Motion of Body (MoBo) database. Technical Report CMU-RI-TR-01-18, Robotics Institute, Carnegie Mellon University, June 2001.

[5] A. Hadid and M. Pietikäinen. Selecting models from videos for appearance-based face recognition. In *Proc. the 17th International Conference on Pattern Recognition (to appear)*, August 2004.

[6] B. Heisele, T. Poggio, and M. Pontil. Face detection in still gray images. Technical Report 1687, Center for Biological and Computational Learning, MIT, 2000.

[7] C. Liu. A bayesian discriminating features method for face detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25:725–740, 2003.

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

[9] E. Osuna, R. Freund, and F. Girosit. Training support vector machines: an application to face detection. In *Proc. Computer Vision and Pattern Recognition*, pages 130–136, June 1997.

[10] P. Penev and J.Atick. Local feature analysis: a general statistical theory for object representation. *Network: Computation in Neural Systems*, 7:477–500, 1996.

[11] P. Phillips, H. Moon, S. A. Rizvi, and P. J. Rauss. The FERET evaluation methodology for face-recognition algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22:1090–1104, 2000.

[12] D. Roth, M. Yang, and N. Ahuja. A snow-based face detector. In *Neural Information Processing Systems*, 2000.

[13] H. A. Rowley, S. Baluja, and T. Kanade. Neural network-based face detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(1):23–38, 1998.

[14] H. Schneiderman and T. Kanade. Probabilistic modeling of local appearance and spatial relationships for object recognition. In *Proc. Computer Vision and Pattern Recognition*, pages 45–51, June 1998.

[15] K.-K. Sung and T. Poggio. Example-based learning for view-based human face detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20:39–51, 1998.

[16] M. Turk and A. Pentland. Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 3:71–86, 1991.

[17] V. Vapnik, editor. *Statistical Learning Theory*. Wiley, New York, 1998.

[18] P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. In *Proc. Computer Vision and Pattern Recognition*, pages 511–518, 2001.

[19] L. Wiskott, J.-M. Fellous, N. Kuiger, and C. von der Malsburg. Face recognition by elastic bunch graph matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19:775–779, 1997.

[20] M.-H. Yang, D. J. Kriegman, and N. Ahuja. Detecting faces in images: a survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24:34–58, 2002.