On the Spatial Distribution of Local Non-parametric Facial Shape Descriptors

Olli Lahdenoja[1,2], Mika Laiho[1], and Ari Paasio[1]

[1] University of Turku, Department of Information Technology
Joukahaisenkatu 3-5, FIN-20014, Turku, Finland
[2] Turku Centre for Computer Science (TUCS)

Abstract. In this paper we present a method to form pattern specific facial shape descriptors called basis-images for non-parametric LBPs (Local Binary Patterns) and some other similar face descriptors such as Modified Census Transform (MCT) and LGBP (Local Gabor Binary Pattern). We examine the distribution of different local descriptors among the facial area from which some useful observations can be made. In addition, we test the discriminative power of the basis-images in a face detection framework for the basic LBPs. The detector is fast to train and uses only a set of strictly frontal faces as inputs, operating without non-faces and bootstrapping. The face detector performance is tested with the full CMU+MIT database.

1 Introduction

Recently, significant progress in the field of face recognition and analysis has been achieved using partially or fully non-parametric local descriptors which provide invariance against changing illumination conditions. These descriptors include Local Binary Pattern (LBP) [1] which was originally proposed as a texture descriptor in [2] and its extensions such as Local Gabor Binary Pattern (LGBP) [3]. In MCT (Modified Census Transform [4]) the means for forming the descriptor are very similar to LBP, hence it is also called modified LBP. The iLBP method for extending the neighborhood of the MCT for multiple radius was presented in [5].

The above mentioned methods for local feature extraction have been applied also to face detection [6] and facial expression recognition [7] (also using a spatio-temporal approach). In face detection, for MCT [4] a cascade of classifiers was used and in [5] a multiscale strategy for iLBP features in a cascade was proposed. In [6] an SVM approach was adopted using the LBPs as features for face detection.

Although the above mentioned (discrete, i.e. non-continuously valued) local descriptors have become very popular, the individual characteristics of each descriptor has not been intensively studied. In the work of [8], MCT and LBP were compared among some other face normalization methods in face verification performance point of view using the eigenspace approach. In [9] the LBPs

A.-B. Salberg, J.Y. Hardeberg, and R. Jenssen (Eds.): SCIA 2009, LNCS 5575, pp. 351–358, 2009.

were seen as thresholded oriented derivative filters and compared to e.g. Gabor filters.

In this paper we present a systematic procedure for analyzing the local descriptors aiming at finding possible redundancies and improvements as well as deepening the understanding of these descriptors. We also show that the new basis-image concept, which is based on a simple histogram manipulation technique can be applied to face detection based on discrete local descriptors.

2 Background

The fundamental idea of LBP, LGBP, MCT and their extensions is to compare intensity values in a local neighbourhood in a way which produces a representation which is invariant to intensity bias changes and the distribution of the local intensities. In a short period of time after [1] in which a clear improvement in face recognition rates was obtained against many state-of-the-art reference algorithms, very impressive recognition results with the standard FERET database among many other databases have been achieved.

A main characteristic of these methods is that they use histograms to represent a local facial area and classification is performed between the extracted histograms, the bins of which describe discrete micro-textural shapes. The LBP (which is also included in LGBP) is clearly a more commonly used descriptor than MCT, possibly because of reduced dimension of the histogram description (by a factor of two) and further histogram length reduction methods, such as the usage of only uniform patterns [2].

While the main difference between MCT and LBP is that in MCT instead of center pixel the mean of all pixels is used as reference intensity (and that the center pixel is included into resulting pattern), the difference between LGBP and LBP is that in LGBP, Gabor filtering is first applied in different frequencies and orientations, after which the LBPs are extracted for classification. LGBP provide a significant improvement in face recognition accuracy compared to basic LBP, but due to many different Gabor filters (resulting in many histograms) the dimensionality of the LGBP feature vectors becomes extremely high. Therefore dimensionality reduction, e.g. PCA and LDA are applied after feature extraction.

3 Research Methods and Analysis

3.1 Constructing the Facial Shape Descriptors

We used the normalized FERET [10] gallery data set (consisting of 1196 intensity faces) as inputs for histogramming which aimed at constructing a representative set of images (so called basis-images) which describe the response of each individual local pattern (e.g. LBP, MCT, LGBP) to different facial locations (and hence, the shape of these locations). Also, some tests were performed with full 3113 intensity image data containing the fb and dup 1 sets. The construction of the basis-images is described in the following.

In a histogram perspective, a pattern histogram is constructed for each spatial face image location (x-y pixel) through the whole input intensity image set. These histograms are then placed to their corresponding spatial (x-y pixel) locations where they were extracted from, and all the other bins in the histograms, except the bin under investigation are ignored. Thus, the resulting basis-image of a certain pattern consists of a spatial arrangement of bin magnitudes for that pattern. The spatial (x-y) size of a basis image is the same as that of each individual input intensity image. This technique results N basis-images for which N is the total number of patterns (histogram bins). Then each basis image is (separately) normalized according to its total sum of bins. The normalization removes the bias which results from the differences in total number of occurrences of each pattern in facial area and shows the pattern specific shape distribution clearly. These basis-images represent the shape distribution of individual patterns among the facial area on average.

Although the derivation of the basis-images is simple, we consider the existence of these continuously valued images a non-trivial case. This is because, especially LBPs, are usually considered as texture descriptors despite of wide range of applications, instead of descriptors with a certain larger scale shape response.

3.2 Analyzing the Properties of Local Descriptors

We conducted tests on LBP and MCT (and some initial tests with LGBP) in order to find out their responses to facial shapes. Neighborhood with a radius of 1 and sample number of 8 was used in the experiments (i.e. 8-neighborhood), but the method allows for choosing any radius. The basis-images of all uniform LBP descriptors are shown in Figure 1. Also, the four basis-images in the upper right corner represent examples of non-uniform patterns. The uniformity of a LBP refers to the total number of circular 0-1 and 1-0 transitions of the LBP (patterns with uniformity of 0 or 2 are considered as uniform patterns, in general).

It seems that as the uniformity level increases (i.e. non-uniform patterns are considered, see Figure 1) the distribution becomes less spatially detailed. However the patterns that are 'near' to uniform patterns seem to give a more detailed response (e.g. non-uniform pattern 0001011) than patterns far from uniformity criterion (e.g. pattern 00101010). In [11] it was observed, that rounding non-uniform patterns into uniform using a hamming distance measure between them resulted in lower error rates in face recognition. With larger data set (of 3113 input intensity faces) many non-uniform patterns seemed to occur in eye center region. By examining the basis-images it seems that non-uniform patterns can not describe facial shapes in as discriminative manner as uniform patterns (which has previously only empirically been verified). Also, as the uniformity level increases the patterns become more rare, as expected.

When studying the distribution of MCT (Modified Census Transform, also called mLBP), we noticed that with the test set used, uniform patterns formed clear spatial shapes similarly to LBPs, while many non-uniform patterns were very rare (i.e. only distinct occurrences). Hence, we propose using the same

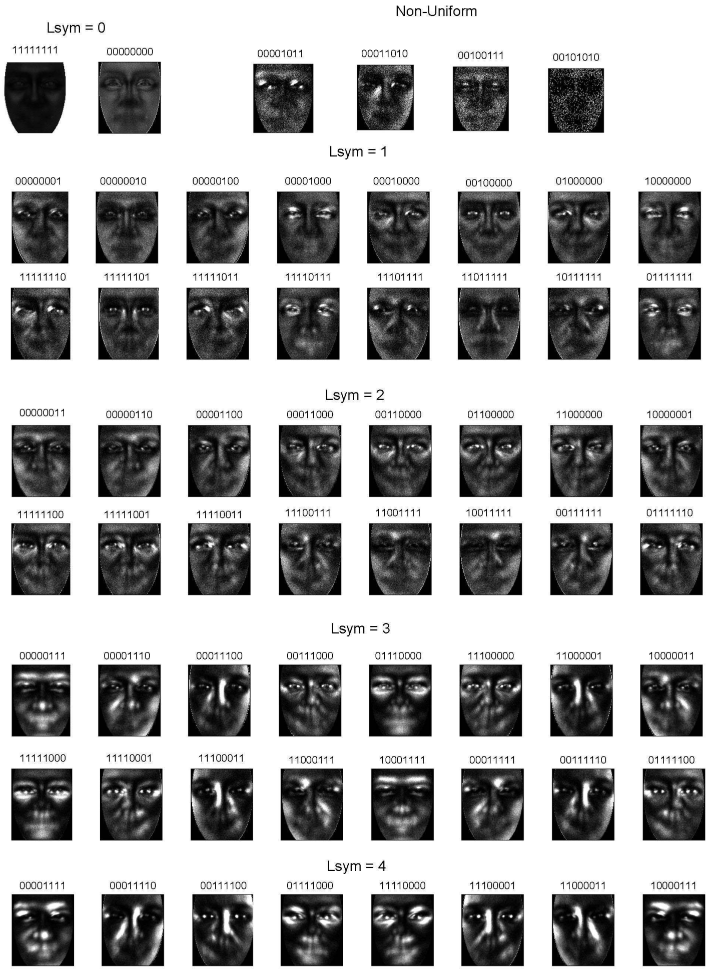

Fig. 1. Selected LBP basis-images

concept of uniform patterns that have been used for LBPs, also with MCTs in face analysis.

In [12] so called symmetry levels for uniform LBPs were presented. Symmetry level L_{sym} of an uniform LBP is defined as the minimum between the total amount of ones and total amount of zeros in a pattern. It was observed in [12] that as the symmetry level of an uniform LBP increases, also the average discriminative efficiency of the LBP increases. This was verified in tests with face recognition using the FERET database. Interestingly, the basis-images of uniform patterns can be divided into classes by their symmetry levels. The spatial distinction between pattern occurrence probabilities gets larger (as occurrence probabilities also mean histogram bin magnitudes, which are now represented as brightness values in Figure 1). Hence, there is a connection between the shape of the basis-images and the discriminative efficiency of the patterns so that as the basis-images become more spatially varied, also the discriminative efficiency of those patterns in face recognition increases [12]. It is also interesting to notice, that the LBPs with a smaller symmetry level seem to give the largest response in the eye regions.

4 Applying Basis-Images for Face Detection

4.1 Motivation

Although the face representation with basis-images is illustrative for examining the response of each pattern to different facial shapes, it can also be used as such in a more quantitative manner. We examined the discriminative power of the basis-image representation in face detection framework since this allows implementing a very compact face detector which requires a negligible time and effort for training or collecting training samples. The training time for the classifier was less than a minute with a P4 processor PC and Matlab.

The simple structure of the classifier and training might be beneficial in certain application environments (e.g. special hardware). However if a state-of-the-art detection rate would be required some of the more complicated procedures (e.g. using also non-faces and bootstrapping) would be necessary. At this point uniform basis-images were used with the basic LBP. However, also MCT and LGBP could be applied in a similar manner for constructing a face detector straightforwardly. The latter methods would lead to a higher dimension of the face description (i.e. more basis-images would have to be used for complete face representation) but might also improve the detection rate and FPR.

4.2 Classification Principle

The face detector implemented operates with a 21x21 search window size. It is slided through all image scales (scaling performed with bilinear subsampling). First the input image is formed for all scales and for each scale the LBP transform is applied. For a certain search window position and scale the LBPs within the search window are replaced by the magnitudes of the corresponding basis-images of these same LBPs in the current spatial locations. For example, if we are in a search window position (x, y) (positions vary between 1 and 21 in x and y directions) we read the LBP of that position (e.g. '00001111') of the input image and use it to find the basis-image of the LBP '00001111', after which the value of that basis-image in the same position (x, y) is summed into accumulator. The 'faceness' measure is then formed by accumulating the magnitudes of the (normalized) basis-image look-ups within the search window area (note that the basis-image concept allows for the normalization procedure). The 'faceness' measure is finally compared against a fixed threshold (determined empirically), which determines whether the sample belongs to class face or non-face. In the current implementation we use 59 basis images, i.e. one for each uniform LBP, and one for describing all the remaining LBPs.

The operations can be performed in cascade, for example, simply by subsampling certain x-y search window positions at a time (possibly first determining which positions belong to the most important ones) and applying a proper threshold for each stage. We tested using two stages to achieve a detection speed of about 4-8 fps with P4 processor and 320x240 resolution in Matlab. However, the detection results reported in this paper were performed without a cascade. In that case the detection speed was approximately 1-2 fps. The search window step in both x and y directions was two in the tests performed. In the experiments a pre-processing step for the input test images (in full scale) and also to basis-images was performed. Both images were low-pass filtered with a 3x3 averaging mask.

4.3 Experimental Results

A detection rate of 78.7% was obtained with 126 false detections with full CMU+MIT database consisting total of 507 faces in cluttered scenes. The total amount of patches searched was about $96.4 * 10^6$ which results in false positive rate in the order of $1.3*10^{-6}$. A maximum of 18 scales were used with scale downsampling factor between 1 and 1.2. Many of the faces that were not detected were not fully frontal, hence it explains part of the moderate recognition rate compared to more advanced detectors, which can easily achieve more than 90% detection rates (however, a more versatile set of input samples for classification is provided with them).

We also tested the detection performance with an easier (more frontal faces) subset of the CMU+MIT set which has been used e.g. in [6]. With this subset there were a total of 227 faces in 80 images. We obtained a detection rate of 87.7% (including drawn faces) with 53 false detections. The total amount of patches searched was about $44.4 * 10^6$ which results in false positive rate in the order of $1.2 * 10^{-6}$ with this set. Hence, the discriminative efficiency (FPR, False Positive Rate) shows a relatively good performance considering the simplicity of the detection framework. In the Figures 2 and 3 some detection results are shown.

Fig. 2. Example detection results with the CMU+MIT database

Fig. 3. Example detection results with the CMU+MIT database

5 Discussion

The idea of basis-images could possibly be extended into other face analysis applications. For example, it might be possible to construct person specific basis images if enough face samples would be present. This could be used for increasing the performance of a face recognition system. In facial expression analysis using a proper alignment procedure it could be possible to capture different expressions to different basis-image sets and use these for recognition and illustration. Also, the effect of global illumination on non-parametric local descriptors could be studied using the basis-image framework.

6 Conclusions

In this paper we presented a method for analyzing local non-parametric descriptors in spatial domain, which showed that they can be seen as orientation selective shape descriptors which form a continuously valued holistic facial pattern representation. We established a dependency between the spatial variability of the resulting LBP basis-images and the symmetry level concept presented in [12]. Through the analysis of basis-images we propose that uniform patterns could be beneficial also with MCTs as with LBPs. We also tested the discriminative power of the basis-image representation in face detection, thus resulting in a new kind of face detector implementation, showing a moderate discriminative efficiency (FPR, False Positive Rate).

References

1. Ahonen, T., Hadid, A., Pietikainen, M.: Face Recognition with Local Binary Patterns. In: Pajdla, T., Matas, J. (eds.) ECCV 2004. LNCS, vol. 3021, pp. 469–481. Springer, Heidelberg (2004)
2. Ojala, T., Pietikainen, M., Maenpaa, T.: Multiresolution Gray-scale and Rotation Invariant Texture Classification with Local Binary Patterns. IEEE Transactions on Pattern Analysis and Machine Intelligence 24(7), 971–984 (2002)
3. 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: Tenth IEEE International Conference on Computer Vision, ICCV, October 2005, vol. 1, pp. 786–791 (2005)
4. Froba, B., Ernst, A.: Face detection with the modified census transform. In: Sixth IEEE International Conference on Automatic Face and Gesture Recognition, May 2004, pp. 91–96 (2004)
5. Jin, H., Liu, Q., Tang, X., Lu, H.: Learning Local Descriptors for Face Detection. In: IEEE International Conference on Multimedia and Expo., ICME, July 2005, pp. 928–931 (2005)
6. Hadid, A., Pietikainen, M., Ahonen, T.: A Discriminative Feature Space for Detecting and Recognizing Faces. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition CVPR, Washington, DC, vol. 2, pp. 797–804 (2004)
7. Feng, X., Pietikainen, M., Hadid, A.: Facial expression recognition with local binary patterns and linear programming. Pattern Recognition and Image Analysis 15(2), 546–548 (2005)
8. Ruiz-del-Solar, J., Quinteros, J.: Illumination Compensation and Normalization in Eigenspace-based Face Recognition: A comparative study of different preprocessing approaches. Pattern Recognition Letters 29(14), 1966–1978 (2008)
9. Ahonen, T., Pietikainen, M.: Image description using joint distribution of filter bank responses. Pattern Recognition Letters 30(4), 368–376 (2009)
10. Phillips, P.J., Wechsler, H., Huang, J., Rauss, P.: FERET Database and Evaluation Procedure for Face Recognition Algorithms. Image and Vision Computing 16, 295–306 (1998)
11. Yang, H., Wang, Y.: A LBP-based Face Recognition Method with Hamming Distance Constraint. In: Proceedings of Fourth International Conference on Image and Graphics (ICIG 2007), pp. 645–649 (2007)
12. Lahdenoja, O., Laiho, M., Paasio, A.: Reducing the feature vector length in local binary pattern based face recognition. In: IEEE International Conference on Image Processing, ICIP, September 2005, vol. 2, pp. 914–917 (2005)