Learning Discriminative LBP-Histogram Bins for Facial Expression Recognition

Caifeng Shan and Tommaso Gritti
Philips Research, High Tech Campus 36, Eindhoven 5656 AE, The Netherlands
{caifeng.shan, tommaso.gritti}@philips.com

Abstract

Local Binary Patterns (LBP) have been well exploited for facial image analysis recently. In the existing work, the LBP histograms are extracted from local facial regions, and used as a whole for the regional description. However, not all bins in the LBP histogram are necessary to be useful for facial representation. In this paper, we propose to learn discriminative LBP-Histogram (LBPH) bins for the task of facial expression recognition. Our experiments illustrate that the selected LBPH bins provide a compact and discriminative facial representation. We experimentally illustrate that it is necessary to consider multiscale LBP for representing faces, and most discriminative information is contained in uniform patterns. By adopting SVM with the selected multiscale LBPH bins, we obtain the best recognition performance of 93.1% on the Cohn-Kanade database.

1 Introduction

Machine analysis of facial expressions, enabling computers to analyze and interpret facial expressions as humans do, has many applications such as human-computer interaction and computer animation; so it has attracted much attention in last two decades [9, 3].

A vital step for successful facial expression analysis is deriving an effective facial representation from original face images. Two types of features [15], geometric features and appearance features, are usually considered for facial representation. Geometric features deal with the shape and locations of facial components (including mouth, eyes, brows, and nose), which are extracted to represent the face geometry [16]. Appearance features present the appearance changes (skin texture) of the face (including wrinkles, bulges and furrows), which are extracted by applying image filters to either the whole face or specific facial regions [6]. The geometric features based facial representations commonly require accurate and reliable facial feature detection and tracking, which is difficult to accommodate in real-world unconstrained scenarios, e.g., under head pose variation. In contrast, appearance features suffer less from issues of initialization and tracking errors, and can encode changes in skin texture that are critical for facial expression modeling. However, most of the existing appearance-based facial representations still require face registration based on facial feature detection, e.g., eye detection.

As an efficient non-parametric method summarizing the local structure of an image, Local Binary Patterns (LBP) has been introduced for facial representation recently [1, 13] (see Section 2 for details about LBP). The most important properties of LBP features

BMVC 2008 doi:10.5244/C.22.27

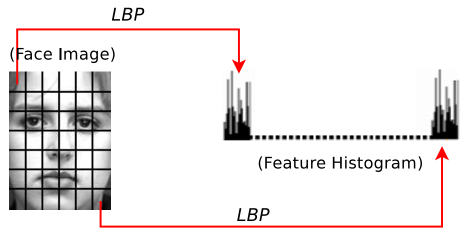

Figure 1: A face image is divided into sub-regions from which LBP histograms are extracted and concatenated into a single, spatially enhanced feature histogram.

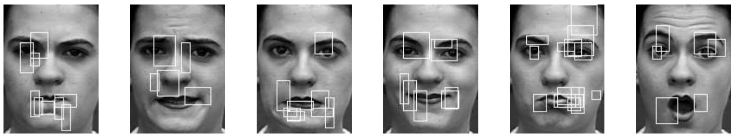

Figure 2: The sub-regions selected by AdaBoost for each facial expression. From left to right: Anger, Disgust, Fear, Joy, Sadness, and Surprise.

are their tolerance against monotonic illumination changes and their computational simplicity. In the original LBP-based facial representation [1, 13], as shown in Figure 1, face images are first equally divided into non-overlapping sub-regions to extract the LBP histograms within each sub-region, which are then concatenated into a single, spatially enhanced feature histogram. Possible criticisms of this method are that dividing the face into a grid of sub-regions is somewhat arbitrary, as sub-regions are not necessary well aligned with facial features, and that the resulting facial representation suffers from fixed size and position of sub-regions. To address these, in [18, 12], by shifting and scaling a sub-window over face images, many more sub-regions are obtained, and then Adaboost [11] is adopted to select the most discriminative sub-regions in term of LBP histogram. Figure 2 shows the selected sub-regions for each facial expression.

In most of the existing work, LBP histograms are extracted from local facial regions as the region-level description, where the *n*-bin histogram is utilized as a whole. However, not all bins in the LBP histogram are necessary to contain useful information for facial representation. It is helpful and interesting to have a closer look at the local LBP histogram at the bin level, to identify the discriminative LBP-Histogram (LBPH) bins for better facial representation. To our best knowledge, this problem has not been investigated in the existing work. In this paper, we propose to learn discriminative LBPH bins for the task of facial expression recognition. Adaboost (Section 3) is adopted to learn LBP features at the bin level. Our experiments (Section 4) illustrate that the selected LBPH bins provide a much more compact facial representation, reducing feature length greatly, while producing better facial expression recognition performance. We experimentally verify the validity of uniform patterns for facial representation from the point view of machine learning. We also evidently show that it is necessary to consider multiscale LBP for facial representation. By adopting Support Vector Machine (SVM) with the selected multiscale LBPH bins, we obtain the recognition performance of 93.1% on the Cohn-Kanade database, which is comparable to the best performance reported so far on the database.

Related Work — As a powerful means of texture description, LBP features have been widely exploited in many applications (see a comprehensive bibliography related to LBP methodology online[1]). For facial image analysis, LBP features have been extensively exploited recently. Rodriguez and Marcel [10] proposed a generative approach for face authentication, which considers local LBP histograms as probability distributions and represents a generic face model by a collection of LBP histograms. Chan *et al.* [2] presented to extract multi-scale LBP histograms from each local regions, which has shown promising performance for face recognition. Liao *et al.* [5] proposed Multi-scale Block LBP for face recognition, in which the computation is done based on average values of block sub-regions, instead of individual pixels. Recently Zhao and Pietikäinen [19] introduced the volume LBP and LBP from three orthogonal planes for dynamic texture recognition, showing promising performance on dynamic facial expression recognition by combining appearance and motion. More recently Tan and Triggs [14] introduced an extension of LBP called Local Ternary Patterns (LTP) that is less sensitive to noise in near-uniform regions.

2 Local Binary Patterns

The original LBP operator, introduced by Ojala *et al.* [7], labels the pixels of an image by thresholding a 3×3 neighborhood of each pixel with the center value and considering the results as a binary number. Formally, given a pixel at (x_c, y_c), the resulting LBP can be expressed in the decimal form as

$$LBP(x_c, y_c) = \sum_{n=0}^{7} s(i_n - i_c) 2^n \tag{1}$$

where n runs over the 8 neighbors of the central pixel, i_c and i_n are the gray-level values of the central pixel and the surrounding pixel, and $s(x)$ is 1 if $x \geq 0$ and 0 otherwise.

Ojala *et al.* [8] later made two extensions of the original operator. Firstly, the operator was extended to use neighborhood of different sizes, to capture dominant features at different scales. Using circular neighborhoods and bilinearly interpolating the pixel values allow any radius and number of pixels in the neighborhood. The notation (P, R) denotes a neighborhood of P equally spaced sampling points on a circle of radius of R. Secondly, they proposed to use a small subset of the 2^P patterns, produced by the operator $LBP(P, R)$, to describe the texture of images. These patterns, called *uniform patterns*, contain at most two bitwise transitions from 0 to 1 or vice versa when considered as a circular binary string. For example, 00000000, 001110000 and 11100001 are uniform patterns. The uniform patterns represent local primitives such as edges and corners. It was observed that most of the texture information was contained in the uniform patterns. Labeling the patterns which have more than 2 transitions with a single label yields an LBP operator, denoted $LBP(P, R, u2)$, which produces much less patterns without losing too much information.

After labeling an image with a LBP operator, a histogram of the labeled image $f_l(x, y)$ can be defined as

$$H_i = \sum_{x,y} I(f_l(x, y) = i), \qquad i = 0, \ldots, L-1 \tag{2}$$

[1] *http://www.ee.oulu.fi/mvg/page/lbp_bibliography*

where L is the number of different labels produced by the LBP operator and $I(A)$ is 1 if A is true and 0 otherwise.

LBP based Facial Representation — Each face image can be seen as a composition of micro-patterns which can be effectively detected by the LBP operator. Ahonen *et al.* [1] introduced a LBP based face representation for face recognition. To consider the shape information of faces, they divided face images into M small non-overlapping regions $R_0, R_1, \ldots, R_M$ (as shown in Figure 1). The LBP histograms extracted from each sub-region are then concatenated into a single, spatially enhanced feature histogram defined as:

$$H_{i,j} = \sum_{x,y} I(f_l(x,y) = i) I((x,y) \in R_j) \tag{3}$$

where $i = 0, \ldots, L-1$, $j = 0, \ldots, M-1$. The extracted feature histogram describes the local texture and global shape of face images. The face representation has also been proved effective for facial expression recognition [13].

To address the limitations of the above scheme including arbitrary sub-region division and fixed size/position of sub-regions, Adaboost later was adopted to learn the most discriminative sub-regions (in term of LBP histogram) from a large pool of sub-regions generated by shifting and scaling a sub-window over face images [18, 12]. Some examples of selected sub-regions for facial expressions are shown in Figure 2. In the existing work, the LBP histograms are always extracted from sub-regions, and used as a whole for the regional description. However, not all bins in the LBP histogram are discriminative for facial representation. In the next section, we propose to learn discriminative LBP-Histogram (LBPH) bins for better facial representation.

3 Learning Discriminative LBP-Histogram Bins

We adopt Adaboost to learn the discriminative LBPH bins. Adaboost [11] learns a small number of weak classifiers whose performance is just better than random guessing, and boosts them iteratively into a strong classifier of higher accuracy. The process of Adaboost maintains a distribution on the training samples. At each iteration, a weak classifier which minimizes the weighted error rate is selected, and the distribution is updated to increase the weights of the misclassified samples and reduce the importance of the others. Since Viola and Jones [17] introduced the Adaboost-based face detector, Adaboost has been successfully used in many computer vision problems. For example, Adaboost was adopted to select Gabor wavelet features [6] and spatial-temporal features [16] for facial expression recognition.

As the traditional Adaboost works on two-class problems, the multi-class problem here is accomplished by using the one-against-rest technique, which trains Adaboost between one expression with all others. For each Adaboost learner, the images of one expression were positive samples, while the images of all other expressions were negative samples. The weak classifier is designed to select the single LBPH bin which best separates the positive and negative examples. Similar to [17], a weak classifier $h_j(x)$ consists of a feature f_j which corresponds to each LBPH bin, a threshold θ_j and a parity p_j indicating the direction of the inequality sign:

$$h_j(x) = \begin{cases} 1 & \text{if } p_j f_j(x) \leq p_j \theta_j \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

4 Experiments

We carry out experiments on the Cohn-Kanade database [4], one of the most comprehensive databases in the current facial expression research community. The Cohn-Kanade database consists of 100 university students aged from 18 to 30 years, of which 65% were female, 15% were African-American, and 3% were Asian or Latino. Subjects were instructed to perform a series of 23 facial displays, six of which were based on description of basic emotions (i.e., Anger, Disgust, Fear, Joy, Sadness, and Surprise). Image sequences from neutral to target display were digitized into 640×490 pixel arrays with 8-bit precision for grayscale values. For our experiments, we selected 320 image sequences of basic emotions from the database. The sequences come from 96 subjects, with 1 to 6 emotions per subject. For each sequence, the neutral face and three peak frames were used for prototypic expression recognition, resulting in 1,280 images (108 Anger, 120 Disgust, 99 Fear, 282 Joy, 126 Sadness, 225 Surprise, and 320 Neutral). To test the algorithms' generalization performance, we adopt 10-fold cross-validation scheme in our experiments, reported the average recognition results.

Following the existing work [13, 10], we scaled the faces to a fixed distance between the two eyes. Facial images of 110×150 pixels were cropped from original frames based on the two eyes location.

4.1 Experiment: Limited Sub-regions

In [13], face images of 110×150 pixels were divided into 42 (6×7) sub-regions of 18×21 pixels (as shown in the bottom right corner of Figure 4), and then the 59-label $LBP(8,2,u2)$ operator was adopted to extract LBP features. These parameter settings were suggested in [1]. We started our experiments with the same setting. Thus each face image was described by a LBP histogram of 2,478 (42×59) bins. We adopted Adaboost to learn discriminative LBPH bins and boost a strong classifier. We plot in the left side of Figure 3 the recognition performance of the boosted strong classifier as a function of the number of features selected. With the 200 selected LBPH bins, the boosted strong classifier achieves recognition rate of 85.3%, which is much better than 79.1% produced by Template Matching using all 2,478 bins in [13].

Spatial Distribution of Features Selected — Figure 4 shows the spatial distribution of the top 200 features selected in the 10-fold cross-validation experiments. As can be observed, different facial expressions have different distribution patterns. For example, for "Disgust", most discriminative features locate in the eye inner corners, while most discriminative features for "Joy" are distributed in the mouth corners, which indeed reenforce human observation. Overall, discriminative features for facial expression classification mostly distribute in eyes and mouth regions.

Uniform Patterns — It was observed that most of the texture information was contained in the *uniform patterns* [8], so uniform patterns was used to reduce the length of LBP histograms. Based on this, the 59-label $LBP(8,2,u2)$ operator, instead of 256-label $LBP(8,2)$, was widely used for facial representation. Here we verify the validity of uniform patterns for facial representation from a point view of machine learning. By using the $LBP(8,2)$ operator, each face image was represented by a LBP histogram of 10,752 (42×256) bins. We plot in the left side of Figure 3 the recognition performance

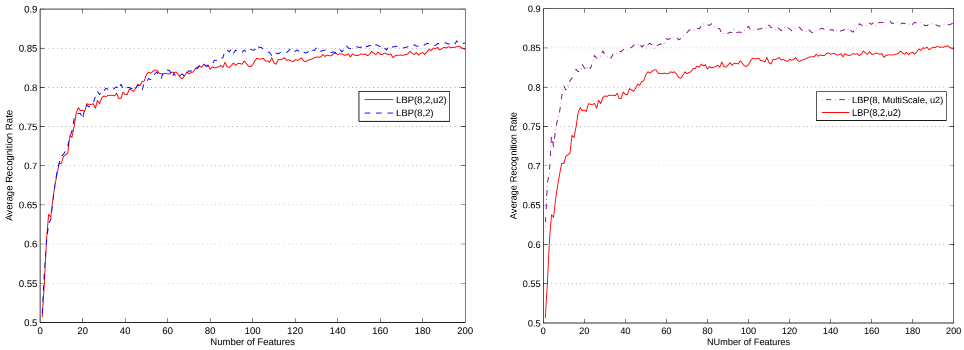

Figure 3: Average recognition rate of the boosted strong classifiers, as a function of the number of feature selected. *Left*: $LBP(8,2,u2)$ vs $LBP(8,2)$; *Right*: $LBP(8,2,u2)$ vs $LBP(8, Multiscale, u2)$.

of the boosted strong classifier. We can see that the boosted strong classifier of $LBP(8,2)$ performs similarly with that of $LBP(8,2,u2)$, which illustrates that the non-uniform patterns do not provide more discriminative information for facial expression recognition. To further verify this, we took a closer look at the learned LBPH bins of $LBP(8,2)$, and found that 91.1% of them are uniform patterns. Therefore, with boost learning we experimentally verify that most of discriminative information for facial expression recognition was contained in the uniform patterns.

Multi-scale LBP — By varying the sampling radius R, LBP of different resolutions can be obtained. The multiscale LBP has provided better performance than single scale LBP for texture classification [8] and face recognition [2]. Here we also investigate multiscale LBP for facial expression recognition. We applied the $LBP(8,R,u2)(R = 1,\cdots,8)$ to extract multiscale LBP features, resulting a LBP histogram of 19,824 (42×59×8) bins for each face image. We then run Adaboost to learn discriminative LBPH bins from the multiscale feature pool. We plot in the right side of Figure 3 the recognition performance of the boosted strong classifier. As can be observed, the boosted strong classifier of multiscale $LBP(8,R,u2)(R = 1,\cdots,8)$ produces consistently better performance than that of single scale $LBP(8,2,u2)$, providing recognition rate of 88.6% with the 200 selected LBPH bins. Thus the multiscale LBP brings more discriminative information for facial expression recognition. Figure 5 shows the scale distribution of the selected LBPH bins in the 10-fold cross-validation experiments. We can see that the scale distribution of features selected for different expressions are different. Overall, discriminative LBPH bins distribute at all scales, especially scales $R = 3,4,6,7,8$. Our experimental results suggest that multiscale LBP features should be considered for facial expression recognition.

We summarize our experiment results in Table 1, where we also include experimental results reported in [13]. As can be observed, the boosted strong classifiers with 200 selected LBPH bins outperform the template matching method using all 2,478 bins [13]. The boosted strong classifier with 200 multiscale LBPH bins provides comparable result to the SVM classifier using all 2,478 bins in [13].

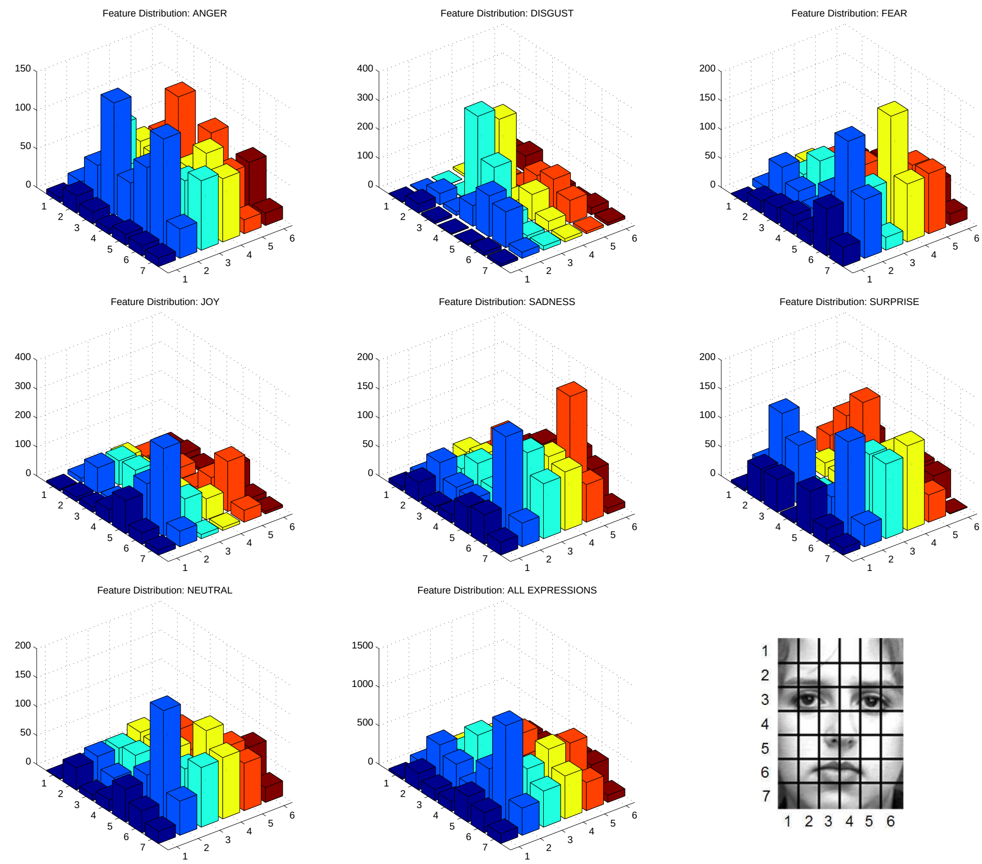

Figure 4: Spatial distribution of the selected LBPH bins (an example face image divided in sub-regions is included in the bottom right corner for illustration).

4.2 Experiment: More Sub-regions

In the above experiments, only 42 equally divided sub-regions were considered for feature selection. By shifting and scaling a sub-window over face images, we can get many more sub-regions, which potentially contain more complete and discriminative information about face images. We shifted the sub-window with the shifting step of 14 pixels vertically and 12 pixels horizontally. The sub-window was scaled as 14, 21, or 28 pixels (height) and 12, 18, or 24 pixels (width) respectively. In total 725 sub-regions were obtained. By using multiscale $LBP(8,R,u2)(R=1,\cdots,8)$, a histogram of 342,200 ($725\times59\times8$) bins was extracted from each face image.

To improve computation efficiency, we adopted a coarse to fine feature selection scheme: We first run Adaboost to select LBPH bins from each single scale $LBP(8,R,u2)$, then applied Adaboost to the selected LBPH bins at different scales to obtain final feature selection results. We plot in the left side of Figure 6 the recognition performance of the boosted strong classifiers as a function of the number of features selected. We can see

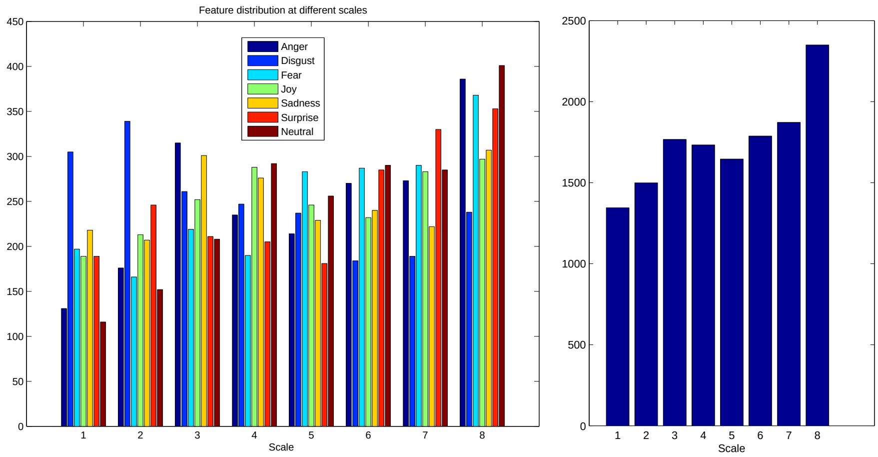

Figure 5: Scale distribution of the selected LBPH bins. *Left*: distribution of each class; *right*: overall distribution.

Methods	Recognition Rate
$LBP(8,2,u2)$ + Adaboost	85.3%
$LBP(8,2)$ + Adaboost	85.9%
$LBP(8,Multiscale,u2)$ + Adaboost	88.6%
$LBP(8,2,u2)$ + Template Matching [13]	79.1%
$LBP(8,2,u2)$ + SVM (Polynomial) [13]	88.4%

Table 1: Recognition performance of different methods using LBP features extracted from equally divided sub-regions.

that the final boosted strong classifier of multiscale LBP provides better performance than that of each single scale. Among strong classifiers of single scales, it seems that scales $(R = 3,4,5,6)$ perform better, while the performance of scales $(R = 1,8)$ is poor.

Feature Distribution — The scale distribution of final selected multiscale LBPH bins is shown in the right side of Figure 6. We can observe that most discriminative LBPH bins come from scales $(R = 3,4,5,8)$. The scale distribution is different from that we obtained in Figure 5, and this is because the features were selected from many more sub-regions. Regarding the spatial distribution of final selected features, Figure 7 shows the top 20 sub-regions that contain most LBP bins selected for each facial expression. We can see that each facial expression has its unique spatial distribution of selected features. As we discussed before, most of discriminative features distribute in eyes and mouth regions.

SVM classification — Finally we adopted SVM to recognize facial expressions using the selected LBPH bins. In [6], SVM using Gabor features selected by Adaboost (AdaSVM) achieves the best performance (93.3%) reported so far on the Cohn-Kanade database. We compare our LBP-based methods with the Gabor-based methods [6] in Table 2. We used the SVM implementation in the library SPIDER[1], and the multi-class classification was

[1]Public available at *http://www.kyb.tuebingen.mpg.de/bs/people/spider/index.html*

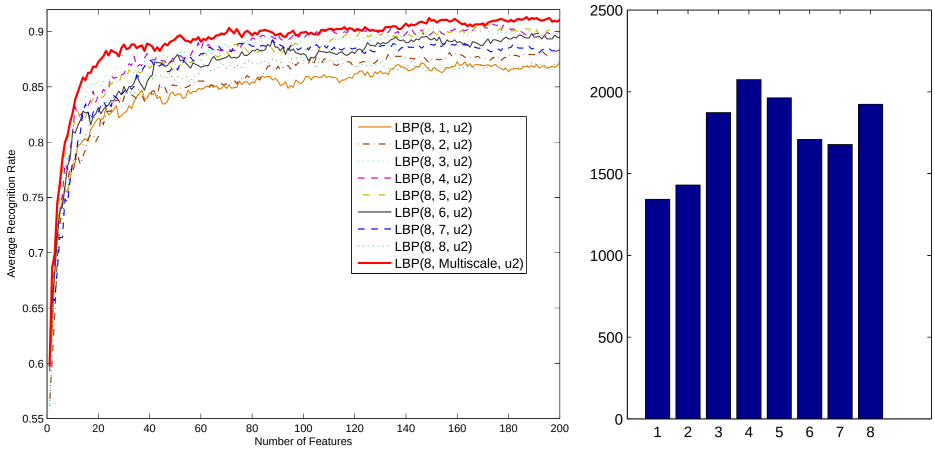

Figure 6: *Left*: Average recognition rate of boosted strong classifiers, as a function of the number of feature selected. *Right*: Scale distribution of the selected LBPH bins

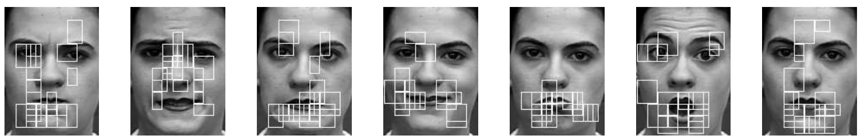

Figure 7: The top 20 sub-regions that contain most LBPH bins selected. From left to right: Anger, Disgust, Fear, Joy, Sadness, Surprise, and Neutral.

accomplished by using the one-against-rest technique. It can be observed that the boosted LBPH bins produce comparable results with the boosted Gabor features [6].

Features	Recognition Rates		
	Adaboost	AdaSVM(Linear)	AdaSVM(RBF)
LBP	91.3%	93.0%	93.1%
Gabor[6]	90.1%	93.3%	93.3%

Table 2: Comparison between the boosted LBPH bins and Gabor wavelet features.

5 Conclusions

In this paper, we propose to learn discriminative LBP-Histogram (LBPH) bins for the task of facial expression recognition using Adaboost. Our experiments illustrate that the selected LBPH bins provide a compact and discriminative facial representation. We experimentally verify the validity of uniform patterns for facial representation. We also evidently illustrate that it is necessary to consider multiscale LBP. By adopting SVM with the selected multiscale LBPH bins, we obtain the recognition performance of 93.1% on the Cohn-Kanade database.

References

[1] T. Ahonen, A. Hadid, and M. Pietikäinen. Face recognition with local binary patterns. In *European Conference on Computer Vision (ECCV)*, pages 469–481, 2004.

[2] C. Chan, J. Kittler, and K. Messer. Multi-scale local binary pattern histograms for face recognition. In *International Conference on Biometrics (ICB)*, pages 809–818, 2007.

[3] B. Fasel and J. Luettin. Automatic facial expression analysis: a survey. *Pattern Recognition*, 36:259–275, 2003.

[4] T. Kanade, J.F. Cohn, and Y. Tian. Comprehensive database for facial expression analysis. In *IEEE International Conference on Automatic Face & Gesture Recognition (FG)*, pages 46–53, 2000.

[5] S. Liao, X. Zhu, Z. Lei, L. Zhang, and S. Z. Li. Learning multi-scale block local binary patterns for face recognition. In *International Conference on Biometrics (ICB)*, pages 828–837, 2007.

[6] G. Littlewort, M. Bartlett, I. Fasel, J. Susskind, and J. Movellan. Dynamics of facial expression extracted automatically from video. *Image and Vision Computing*, 24(6):615–625, June 2006.

[7] T. Ojala, M Pietikäinen, and D. Harwood. A comparative study of texture measures with classification based on featured distribution. *Pattern Recognition*, 29(1):51–59, 1996.

[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(7):971–987, 2002.

[9] M. Pantic and L. Rothkrantz. Automatic analysis of facial expressions: the state of art. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(12):1424–1445, 2000.

[10] Y. Rodriguez and S. Marcel. Face authentication using adapted local binary pattern histograms. In *European Conference on Computer Vision (ECCV)*, pages 321–332, 2006.

[11] R. E. Schapire and Y. Singer. Improved boosting algorithms using confidence-rated predictions. *Maching Learning*, 37(3):297–336, 1999.

[12] C. Shan, S. Gong, and P. W. McOwan. Conditional mutual information based boosting for facial expression recognition. In *British Machine Vision Conference (BMVC)*, volume 1, pages 399–408, Oxford, UK, September 2005.

[13] C. Shan, S. Gong, and P. W. McOwan. Robust facial expression recognition using local binary patterns. In *IEEE International Conference on Image Processing (ICIP)*, volume 2, pages 370–373, Genoa, Italy, September 2005.

[14] X. Tan and B. Triggs. Enhanced local texture feature sets for face recognition under difficult lighting conditions. In *IEEE International Workshop on Analysis and Modeling of Faces and Gestures (AMFG)*, pages 168–182, 2007.

[15] Y. Tian, T. Kanade, and J. Cohn. *Handbook of Face Recognition*, chapter 11. Facial Expression Analysis. Springer, 2005.

[16] M. Valstar and M. Pantic. Fully automatic facial action unit detection and temporal analysis. In *IEEE Conference on Computer Vision and Pattern Recognition Workshop*, page 149, 2006.

[17] P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 511–518, 2001.

[18] G. Zhang, X. Huang, S. Z. Li, Y. Wang, and X. Wu. Boosting local binary pattern (lbp)-based face recognition. In *Chinese Conference on Biometric Recognition*, pages 179–186, 2004.

[19] G. Zhao and M. Pietikäinen. Dynamic texture recognition using local binary patterns with an application to facial expressions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(6):915–928, 2007.