

# Gender Classification by LUT Based Boosting of Overlapping Block Patterns

Rakesh Mehta<sup>1</sup>(✉), Manuel Günther<sup>2</sup>, and Sébastien Marcel<sup>2</sup>

<sup>1</sup> Tampere University of Technology, Tampere, Finland  
rakesh.mehta@tut.fi

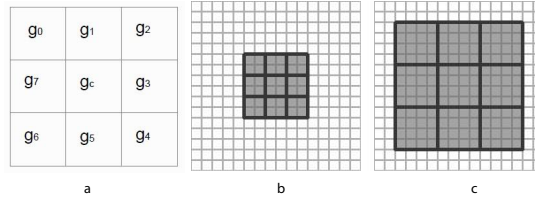
<sup>2</sup> Idiap Research Institute, Martigny, Switzerland  
{manuel.guenther,marcel}@idiap.ch

**Abstract.** The paper addresses the problem of gender classification from face images. For feature extraction, we propose discrete Overlapping Block Patterns (OBP), which capture the characteristic structure from the image at various scales. Using integral images, these features can be computed in constant time. The feature extraction at multiple scales results in a high dimensionality and feature redundancy. Therefore, we apply a boosting algorithm for feature selection and classification. Look-Up Tables (LUT) are utilized as weak classifiers, which are appropriate to the discrete nature of the OBP features. The experiments are performed on two publicly available data sets, Labeled Faces in the Wild (LFW) and MOBIO. The results demonstrate that Local Binary Pattern (LBP) features with LUT boosting outperform the commonly used block-histogram-based LBP approaches and that OBP features gain over Multi-Block LBP (MB-LBP) features.

## 1 Introduction

Gender classification is an important task that finds its applications in a number of areas such as security, surveillance, criminology, multimedia, etc. A number of biometric cues have been used for this purpose, such as face images [1], person gait [2], fingerprints [4], speech or the combination of face and fingerprints [3]. In this paper, we focus on gender classification from face images, as these can be easily captured and does not require special devices, but only an ordinary camera.

A number of approaches have been proposed for gender classification from the face images. The classification pipeline from images consists of two main steps: feature extraction and classification. A variety of features have been used for gender classification, such as Haar [5], SIFT [6], Local Binary Patterns (LBP) [7], Shape Context [6], Local Directional Patterns (LDP) [9], Interlaced Derivative Pattern (IDP) [8], or Local Circular Patterns (LCP) [10]. Some approaches have also used a combination of these features to capture discriminative information from the images [6,7]. Originally, LBP is a local texture descriptor. However, because of its high discriminative power and computational simplicity, it has been applied in face processing tasks such as face detection [19], face recognition



**Fig. 1.** (a) The notation used for LBP (b) The MB-LBP blocks for block size  $2 \times 2$  (c) MB-LBP blocks for block size  $4 \times 4$

[16], emotion recognition [18], and also for gender classification LBP and its variants have shown promising results.

For classification, Linear Discriminant Analysis (LDA), Nearest Neighbor (NN), Radial Basis Function network and Support Vector Machines (SVM) were studied [11]. The results show that SVM outperforms the other algorithms only by a small margin. On the other hand, when the number of features are high, boosting has shown good performance [5], [26]. The advantage of boosting compared to the other classifiers is that it makes the classification process faster, as only a relatively small number of discriminative features are required during the testing phase. For a comparative study of different approaches, we refer the readers to [17].

LBP features are computed by taking the difference of the pixels in a restricted neighborhood and, therefore, they are not able to capture information at different scales. To incorporate information at higher scales, Multi-Block LBP (MB-LBP) was proposed [15], which takes difference of block sums instead of pixels. However, as the block size increases, the distance between the adjacent blocks becomes too large to capture the relation with respect to its neighborhood. Thus, to efficiently include the relation from the surrounding even at larger scales, in this paper we propose Overlapping Block Patterns (OBP), which capture local image gradient information at different scales from the image. Using integral images, these features can be computed in constant time.

To select only a relatively small number of discriminative features, we apply boosting. We use Taylor Boosting [12], which is a generalized algorithm that supports a family of loss functions and first and second order optimization. Taylor Boosting has been combined with Look-Up Tables (LUT) as weak classifiers [20] to design classification algorithms specifically for non-metric and discrete features such as OBP. The experiments on two challenging face image databases demonstrate that the proposed OBP features outperform not only LBP, but also MB-LBP features. The proposed algorithm also has the advantage of low complexity, as the features are computed in constant time and the classification using boosting is only linearly dependent on the number of weak classifiers.

The remaining of this paper is organized as follows: In Section 2, we start with discussing LBP and MB-LBP features, then we present the novel OBP features. Finally, gender classification using LUT based boosting is discussed. The experiments are performed on two public data sets following their standard

protocols and the results are reported in Section 3. Finally, Section 4 concludes the paper.

## 2 OBP based Gender Classification

In this section we first discuss the LBP features and its extension MB-LBP. Then, we propose the novel OBP features that capture discriminative information of blocks from the immediate neighborhood. Finally, we present the LUT based boosting algorithm, which can be used with discrete features like LBP, MB-LBP and OBP.

### 2.1 LBP and MB-LBP

The original LBP operator captures the relation of a pixel with its neighboring  $3 \times 3$  pixels. It is computed by taking the difference of a pixel around it neighbors as:

$$F(g_c) = \sum_{i=0}^7 s(g_i - g_c) \cdot 2^i \quad (1)$$

with:

$$s(x) = \begin{cases} 1 & : x \geq 0 \\ 0 & : x < 0 \end{cases} \quad (2)$$

where  $g_c$  denotes the central pixel, for which the LBP feature is computed, and  $g_i$  are the neighboring pixels. Fig. 1(a) shows the notation used for the position of the pixels for LBP. If an LBP feature is to be computed at location  $(x, y)$  in image  $I$ , the parameters of (1) and (2) are given as  $g_c = I(x, y)$ ,  $g_0 = I(x - 1, y - 1)$ ,  $g_1 = I(x, y - 1) \dots g_7 = I(x - 1, y)$ . Although fast to compute and discriminative, these features are sensitive to noise as they are computed by taking the difference of the individual pixels. Another limitation of these features is their inability to capture the information at variable scale size.

To overcome these limitations, MB-LBP features have been proposed which use rectangular blocks, instead of the pixel values. The relation between the adjacent blocks is given by (1) and (2), however, in this case the parameters  $g_c$  and  $g_i$  represent the sum of the blocks instead of individual pixel intensity. The sum of the pixels intensities for a block of size  $L \times L$  is represented as:

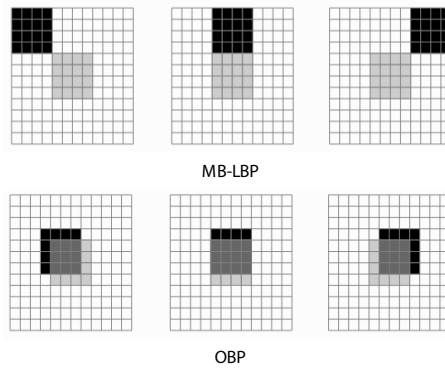
$$S_L(x, y) = \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} I(x + i, y + j), \quad (3)$$

where  $(x, y)$  represent the upper left corner of the block and it is used to denote its position. Thus, for MB-LBP the parameters of (1) are defined as:  $g_c = S_L(x, y)$ ,  $g_0 = S_L(x - L, y - L)$ ,  $g_1 = S_L(x, y - L)$ ,  $\dots$   $g_7 = S_L(x - L, y)$ . The blocks for MB-LBP feature for different block sizes ( $L = 2, 4$ ) are shown in Fig. 1.

The key idea behind LBP based features is to compute the first order derivative of the central values along the eight different neighboring directions and encode the sign of these derivatives. Without loss of generality, in the following we consider the derivative along a single direction for LBP and MB-LBP features at a position  $(x, y)$  in image  $I$ , but the same discussion applies for all other directions as well. The derivative is considered along the direction of  $g_0$ , thus it is given as  $g_c - g_0$ , which corresponds to the first bit during the computation of the LBP or MB-LBP feature value. For LBP feature computation, derivative along the direction of  $g_0$  pixel can be represented as:

$$g_c - g_0 = I(x, y) - I(x - 1, y - 1). \tag{4}$$

Hence, only the variations in the immediate neighborhood around the point  $(x, y)$  are observed.



**Fig. 2.** Difference in block positioning between MB-LBP and OBP

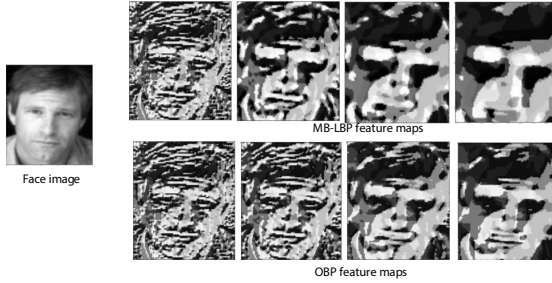
In case of MB-LBP feature computation, the derivative along  $g_0$  is given as:

$$g_c - g_0 = S_L(x, y) - S_L(x - L, y - L). \tag{5}$$

Comparing (4) and (5) it can be observed that both take the difference of function at a point with its neighbor. However, in (5) the difference of the function is taken at step size  $L$ . Although the direction of the derivative in MB-LBP is similar to LBP, it is not taken from the immediate neighborhood like in case of LBP. For larger block size this step increases substantially and it does not provide the accurate estimate of the derivative at that point any longer.

## 2.2 OBP

The proposed OBP features capture the relation of the block with respect to its immediate neighborhood. Similar to the LBP and MB-LBP features, the derivative is computed along eight directions corresponding to the adjacent neighboring blocks and the sign of the derivative is encoded by multiplying with unique



**Fig. 3.** The face image and the feature maps generated from this image using OBP and MB-LBP features at scales  $L = 1, 3, 5, 7$  (left to right)

weight. However, unlike the MB-LBP features the derivative is taken at finer rate, which captures the details around the block in a more efficient manner.

To compute the OBP feature at coordinate  $(x, y)$  in image  $I$ , the parameters  $g_c$  and  $g_i$  of (1) are defined as  $g_c = S_L(x, y)$ ,  $g_0 = S_L(x - 1, y - 1)$ ,  $g_1 = S_L(x, y - 1)$ ,  $\dots$   $g_7 = S_L(x - 1, y)$ . It should be noted that the only difference in the definition of these parameters between MB-LBP and OBP is the step size of  $L$  for MB-LBP and 1 for OBP. To further illustrate this difference, the blocks and their positions are explained in Fig. 2 for block of size  $4 \times 4$ . It can be observed that the neighboring blocks overlap with the center block in case of OBP. The light gray color represents the central blocks, black indicates the neighboring block and the dark gray represents the overlapping area. For block size  $L$  the overlap is  $L - 1$ , which is the maximum possible overlap between two blocks in any direction. This overlap helps in capturing the subtle difference from the neighborhood of the blocks and was empirically found to obtain the best classification performance.

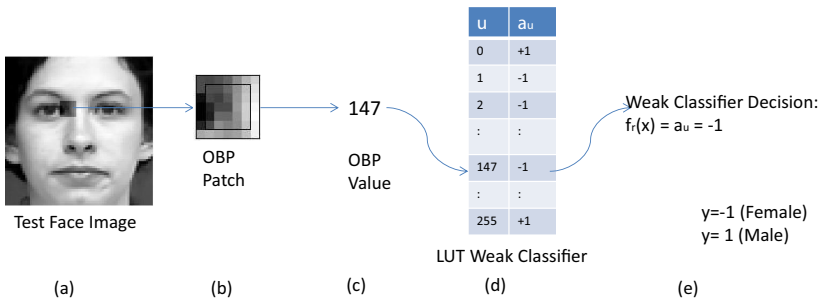
Fig. 3 shows the face image and the MB-LBP and OBP feature images for different block sizes  $L = 1, 3, 5$  and 7. It can be observed that for small sizes, both the MB-LBP and OBP capture the fine details and the feature images are similar. In fact, for block size  $L = 1$ , LBP, MB-LBP and OBP are identical. However, as the block size increases, more differences between the feature images can be observed. The OBP feature images at higher scales show finer details when compared to the MB-LBP images. Furthermore, at higher scales the border pruning effect in case of the MB-LBP features is evident. The resulting feature image for MB-LBP features is always  $3L - 1$  pixels shorter in each dimension because features cannot be computed for the first  $L$  and last  $2L - 1$  pixels in each dimension. In opposition, for OBP only  $L + 1$  features are removed at the border, the difference can best be seen in the feature image at scale 7 in Fig. 3.

In order to capture the micro-structure at different scales, we extract OBP features at various block sizes. We restrict the features to square shape because we empirically observed that they are sufficient to capture the block based information. However, the features can easily be generalized to rectangular shape, which results in a substantial increase in the number of features and, therewith,

in the time required for training – without any significant improvement of the classification performance.

### 2.3 LUT based Taylor Boosting

For gender classification, features are extracted at multiple block sizes from the face images. This results in a large number of features even for small images. Most of these features are redundant and correlated. In order to select a small number of highly discriminative features, we apply boosting using the Taylor Boosting algorithm [12]. The main characteristics of the applied boosting algorithms are: 1) Taylor Boosting has shown to perform better than the more commonly used AdaBoost algorithm, 2) Look-Up Tables (LUT) are used as the weak classifier, which is appropriate to the proposed discrete OBP features, 3) the algorithm is very general and support different kind of loss functions and optimization strategies



**Fig. 4.** Decision of a single weak LUT classifier using OBP features: (a) Test image with the feature position (b) the extracted block and its immediate neighborhood (c) the OBP feature value computed for this feature (d) the LUT of the weak classifier (e) the weak decision for the weak classifier

Let the labeled training samples be represented as  $\{(\mathbf{x}_n, y_n)_{n=1:N}\}$ , where  $y_n \in \{-1, 1\}$  denotes the class of the samples  $\mathbf{x}_n$ . The goal of the boosting algorithm is to build a strong classifier  $F$  as a weighted linear combination of  $R$  weak classifiers  $f_r$ :

$$F(\mathbf{x}) = \sum_{r=1}^R \alpha_r f_r(\mathbf{x}) \tag{6}$$

by minimizing some loss function  $L(F) = \sum_n l(y_n, F(\mathbf{x}_n))$ . Boosting can be interpreted as gradient descent in the functional space, where at each iteration a weak classifier is added to minimize the local loss. Taylor Boosting was proposed as a generalized boosting algorithm that can support first and second order optimization and a variety of loss functions. In this paper, we use Taylor Boosting

**Data:** Training samples, weak classifier  $f_i$ , loss  $L(F)$   
**Result:** Strong classifier  $F$   
 initialization  $F = 0$   
**for**  $r = 1$  to  $r \leq R$  **do**  
     Select the weak classifier  $f_r = f(\mathbf{x}; d, a)$   
      $d^* = \arg \min_d - \sum_u \left| \sum_{n, \mathbf{x}_n = u} L'_n(F) \right|$   
      $a_u = - \sum_{n, \mathbf{x}_n, d^* = u} L'_n(F)$   
     Compute weight using line search:  $\alpha_r = \arg \min_{\alpha} L(F + \alpha f_r)$   
     Update strong classifier:  $F \leftarrow F + \alpha_r f_r$   
**end**  
**return**  $F$

**Algorithm 1.** Taylor Boosting using LUT as weak classifier.

with Look-Up Table as weak classifiers and logistic loss to optimize expectation loss formulation [20].

An important contribution of the paper is to introduce Look-Up Tables (LUT) as the weak classifier with the Taylor Boosting algorithm for gender classification. Commonly, the decision stump is used as the weak classifier in boosting [22]. Decision stump is parametrized by a threshold and assumes that the features are metric in nature. However, the proposed LBP based features are non-metric because each LBP value represents a distinct pattern that is discrete and bounded, feature values lie in the range of [0, 255]. Therefore, a stump classifier cannot be used as weak classifier for those features. Instead, we use a LUT as a weak classifier  $f_r(\mathbf{x})$ , which selects a single feature and uses it for predicting the class of the test sample. The weak classifier  $f_r(\mathbf{x})$  is parametrized by a feature index  $d$  and a look-up table  $a$ . Therefore, the parametrized form of the weak classifier is represented as  $f_r(\mathbf{x}; d, a)$ , and on each round of boosting the parameters  $d, a$  and the weight  $\alpha_r$  of the weak classifier are determined. The feature index represents size  $L$  of the OBP feature and its position  $(x, y)$  in the image, while the LUT has 256 entries and each entry  $a_u$  corresponds to a specific discrete value of the feature. The value  $a_u$  gives the weak classification decision for a sample based only on the feature index  $d$ . The training algorithm for the binary classification problem, e. g., gender classification is given in Algorithm 1. For more detailed analysis of the boosting algorithm please refer to [20].

Fig. 4 shows the classification using LUT for a single weak classifier during testing. The specific feature corresponding to the classifier is extracted from the image and the feature value is computed. This feature value is looked up in the LUT to make the decision for the current weak classifier. The final decision incorporates a linear combination of these individual decisions (Eq. (6)) and compares the result with a certain threshold.

### 3 Experiments

Experiments are conducted on two publicly available large face data sets: MOBIO [23] and Labeled Faces in Wild (LFW) [24]. In this section, we first provide brief information about the data sets and the protocols used for testing. Afterwards, the experimental setup is specified and, finally, the results are presented.

#### 3.1 Data set and Protocol

To evaluate the performance in the uncontrolled setting, we use the data sets Labeled Faces in the Wild (LFW) and MOBIO.

LFW consists of 13,233 images of celebrities acquired in uncontrolled environments leading to high variability in terms of pose, illumination and facial expressions. Experiments are conducted using the BeFIT protocol,<sup>1</sup> in which the images are randomly divided into five subsets and leave-one-out cross validation is performed. Thus, in each split, four subsets of images are used for training and the fifth is used for testing. The final performance is reported as the average accuracy over the five rounds.

The MOBIO data set consists of 61 hours of audio-visual data of 150 people captured within 12 different sessions. This data set is challenging as the images are captured using mobile device and, therefore, the samples are noisy. Here, we use only facial images that have been extracted from these videos [23]. The total number of images is above 27,000, evenly divided into subsets for training, development and testing. The training samples are used for learning the strong classifier, the development set is used to determine the threshold to minimize the error and the final unbiased accuracy is reported on the test set.

The evaluation metrics used in our work are classification accuracy, true positive rate (TPR) and true negative rate (TNR) defined as [17]:

$$TPR = \frac{TP}{P}, \quad TNR = \frac{TN}{N}, \quad \text{Accuracy} = \frac{TP + TN}{P + N}, \quad (7)$$

where  $TP$  and  $TN$  are the number of samples correctly classified as positive (i. e. male) or negative (i. e. female), respectively, and  $P$  and  $N$  are the total numbers of positive and negative samples. Furthermore, we used a variant of the receiver operating characteristic (ROC) curve that plots the fraction of males classified correctly over the fraction of females classified incorrectly.

#### 3.2 Experimental Setup

The images from both data sets are cropped, rotated and geometrically normalized based on the eye locations, which are hand-labeled<sup>2</sup> for the MOBIO

<sup>1</sup> <http://fipa.cs.kit.edu/431.php>

<sup>2</sup> <http://www.idiap.ch/resource/biometric>



database and automatically detected<sup>3</sup> for LFW. The size of the images is set to  $36 \times 36$  in all the experiments, based on the results from [17], where the best accuracy was reported for this size. The inter-ocular distance is set to half of the images width, having eye positions located at 14 pixels from top, see Fig. 4(a) for an example. After obtaining the normalized face images, a photometric normalization [21] is performed.

In our experiments on gender classification, the results are reported using four different algorithms:

- PCA + LDA: Principle Component Analysis (PCA) is applied on the raw image pixels to retain 98% of the variance and Linear Discriminant Analysis (LDA) is used as classifier. Although it seems a naive approach, it has shown promising results for gender classification [13].
- LBP<sub>PHS</sub> + LDA: Each face image is divided into  $6 \times 6$  blocks each of size  $6 \times 6$  pixels. The uniform LBP features are computed for each block and histograms are obtained. The histograms are concatenated to form the feature vector, whose dimensionality is reduced with PCA as explained above. Again, LDA is used as the classifier.
- MB-LBP + Boosting: Multi-Block LBP features are extracted from the face images using square blocks of sizes  $1 \times 1$  to  $7 \times 7$ . Strong LUT classifiers are obtained after 800 rounds of boosting.
- OBP + Boosting: Overlapping Binary Pattern features are extracted from the face images using square blocks of sizes  $1 \times 1$  to  $7 \times 7$ . Strong LUT classifiers after 800 rounds of boosting are used.

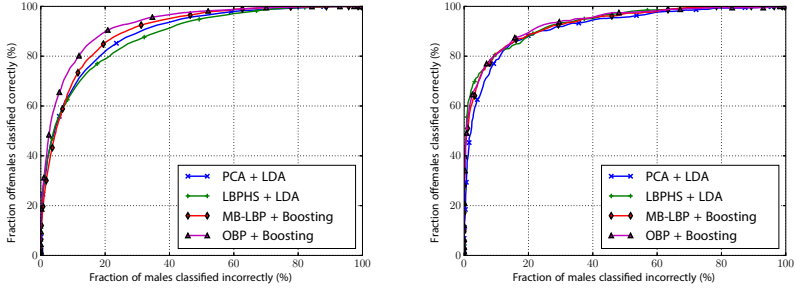
The development of the algorithms for this paper is done using the open source machine learning toolbox Bob [25]. The code to reproduce the results of the paper is publicly available<sup>4</sup>.

### 3.3 Results

The results for both data sets and all tested algorithms are presented in Tables 1 and 2. They show the accuracies of the different approaches along with the TPR and TNR. For MOBIO it can be observed that the PCA + LDA achieves an accuracy of 80.04%, which is outperformed by the 81.58% of LBP<sub>PHS</sub> + LDA. Clearly, LBP features work better than simple pixel values as used by PCA. Further, the MB-LBP features along with LUT based boosting achieves an accuracy of 83.24%, which is significantly higher than both the PCA and LBP<sub>PHS</sub> features with LDA. The gain in performance can be attributed to the fact that the features are extracted at different scales and only discriminative features are utilized during classification. The OBP features further gain 2.11% compared to the MB-LBP features. Here, we would like to point out that the accuracy is not the mean of TPR and TNR because the number of male and female samples are different (see (7)).

<sup>3</sup> <http://lear.inrialpes.fr/people/guillaumin/data.php>

<sup>4</sup> The source code is available at <https://pypi.python.org/pypi/bob.paper.SCIA2015>.



**Fig. 5.** ROC curves of all methods on the MOBIO and LFW test set

**Table 1.** Classification Results on the MOBIO test set

Method	Accuracy	TPR	TNR
PCA + LDA	80.04 %	78.02 %	83.95 %
LBPHS + LDA	81.58 %	<b>86.56 %</b>	71.89 %
MB-LBP + Boosting	83.24 %	85.82 %	78.21 %
OBP + Boosting	<b>85.35 %</b>	86.42 %	<b>83.28 %</b>

**Table 2.** Average classification Results on LFW

Method	Accuracy	TPR	TNR
PCA + LDA	85.92 %	86.08 %	85.39 %
LBPHS + LDA	88.06 %	88.73 %	<b>85.76 %</b>
MB-LBP + Boosting	90.42 %	97.12 %	67.32 %
OBP + Boosting	<b>90.81 %</b>	<b>97.49 %</b>	67.79 %

The performance on LFW is in general higher than on MOBIO, most probably because the LFW protocol includes more training data. Additionally, the samples are less noisy and the illumination is not as adverse as in MOBIO. Nevertheless, a similar trend can be observed in the performances of the algorithms. The PCA and LBPHS features perform inferior to the MB-LBP and OBP features. The best overall performance is again achieved by OBP + Boosting, it gains more 5% and 2% over PCA + LDA and LBPHS + LDA, respectively. The OBP features outperform MB-LBP features by a small margin, which indicates that overlapping of features helps in capturing more discriminative information at higher scales.

To visualize the results on different operating points, the ROC curves for all the approaches are plotted and shown in Figures 5. From the ROC curves it can be observed that there is a substantial performance improvement by the LUT based approaches over the LDA based approaches. Further, among the LUT boosting methods, OBP features constantly outperform MB-LBP.

The testing using the proposed algorithm can be performed in real time because evaluating the strong classifier for a sample simply involves a weighted sum of the responses of the weak classifiers. Thus, the classification of a sample using  $R$  rounds of boosting involves  $R$  feature computations, evaluation of the weak classifier response for each feature and, finally, summation of these responses. As discussed earlier, the feature computation using integral images can be implemented at constant time. Evaluation of  $R$  weak classifiers involves  $R$  elementary LUT indexing operations and summations. Therefore, both the feature extraction and classification times are linear in the number of weak classifiers  $R$ . Note that the number of features involved in boosting are much less than LBPHS based approaches. For instance, in our experiments the boosting based approaches involves 800 feature computation, whereas the feature vector length for the LBPHS is 2124 ( $= 6 \times 6 \times 59$ ).

## 4 Conclusion

A new method for gender classification from face images is presented. Novel steps are introduced at both the feature extraction and the classification stage. Discrete Overlapping Block Features (OBP) features are proposed to overcome the limitation of LBP and MB-LBP features. These features are fast to compute, discriminative in nature and can capture information at variable size. OBP are combined with Look-Up Tables based Taylor Boosting. Experiments are performed on two large publicly available data sets using their standardized protocols. The results demonstrate that OBP features consistently outperform the MB-LBP features and the combination of these features with boosting outperforms LBP histogram based approaches. The proposed algorithm can be used in real time as the complexity is linear in the number of weak classifiers, which makes it fast and efficient.

**Acknowledgments.** The research leading to this paper has received funding from the Swiss National Science Foundation under the National Center of Competence in Research IM2 ([www.im2.ch](http://www.im2.ch)) and the FP7 European project BEAT ([www.beat-eu.org](http://www.beat-eu.org)).

## References

1. Gao, W., Ai, H.: Face gender classification on consumer images in a multiethnic environment. In: Tistarelli, M., Nixon, M.S. (eds.) ICB 2009. LNCS, vol. 5558, pp. 169–178. Springer, Heidelberg (2009)
2. Li, X., Maybank, S.J., Yan, S., Tao, D., Xu, D.: Gait Components and their Application to Gender Recognition. *IEEE Transactions on Systems, Man, and Cyber.* 145–155 (2008)
3. Jain, A.K., Nandakumar, K., Lu, X., Park, U.: Integrating faces, fingerprints, and soft biometric traits for user recognition. In: Maltoni, D., Jain, A.K. (eds.) BioAW 2004. LNCS, vol. 3087, pp. 259–269. Springer, Heidelberg (2004)

4. Badawi, A., Mahfouz, M., Tadross, R., Jantz, R.: Fingerprint-based gender classification. In: International Conference on Image Processing, Computer Vision, and Pattern Recognition, pp. 41–46 (2006)
5. Shakhnarovich, G., Viola, P.A., Moghaddam, B.: A Unified Learning Framework for Real Time Face Detection and Classification. *IEEE Int. Conf. on FG* (2002)
6. Wang, J.-G., et al.: Boosting dense SIFT descriptors and shape contexts of face images for gender recognition. In: *CVPRW* (2010)
7. 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)
8. Shobeirinejad, A., Gao, Y.: Gender classification using interlaced derivative patterns. In: *IEEE International Conference on Pattern Recognition* (2010)
9. Jabid, T., Kabir, M.H., Chae, O.: Gender classification using local directional pattern (LDP). In: *IEEE International Conference on Pattern Recognition* (2010)
10. Wang, C., Huang, D., Wang, Y., Zhang, G.: Facial image-based gender classification using local circular patterns. In: *IEEE ICPR*, pp. 2432–2435 (2012)
11. Moghaddam, B., Yang, M.-H.: Learning Gender with Support Faces. *IEEE Transactions on PAMI*, 707–711 (2002)
12. Saberian, M.J., Masnadi-Shirazi, H., Vasconcelos, N.: Taylorboost: first and second-order boosting algorithms with explicit margin control. In: *CVPR* (2011)
13. Bekios-Calfa, J., Buenaposada, J.M., Baumela, L.: Revisiting Linear Discriminant Techniques in Gender Recognition. *IEEE Transactions on PAMI*, 858–864 (2011)
14. Trefný, J., Matas, J.: Extended set of Local Binary Patterns for Rapid Object Detection. *Computer Vision Winter Workshop* (2010)
15. Liao, S.C., Zhu, X.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)
16. Ahonen, T., Hadid, A., Pietikinen, M.: Face recognition with local binary patterns. In: *Computer Vision*, pp. 469–481. Springer (2004)
17. Mäkinen, E., Raisamo, R.: Evaluation of Gender Classification Methods with Automatically Detected and Aligned Faces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 541–547 (2008)
18. Zhao, G., Pietikinen, M.: Dynamic Texture Recognition using Local Binary Patterns with an Application to Facial Expressions. *IEEE TPAMI* (2007)
19. Jin, H., Liu, Q., Lu, H., Tong, X.: Face detection using improved LBP under bayesian framework. In: *IEEE ICIG*, pp. 306–309 (2004)
20. Atanasoaei, C.: Multivariate Boosting with Look-up-Tables for Face Processing. Ph.d thesis, EPFL (2012)
21. 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)
22. Viola, P., Jones, M.J.: Robust Real-time Face Detection. *International Journal of Computer Vision*, 137–154 (2004)
23. McCool, C., Marcel, S., Hadid, A., et al.: Bi-modal Person Recognition on a Mobile Phone: Using Mobile Phone Data. *IEEE ICME Workshop on Hot Topics in Mobile Multimedia* (2012)
24. Huang, G.B., Ramesh, M., Berg, T., Learned-Miller, E.: Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments. Technical Report. University of Massachusetts, Amherst (2007)

25. Anjos, A., El Shafey, L., Wallace, R., Günther, M., McCool, C., Marcel, S.: Bob: a free Signal Processing and Machine Learning Toolbox for Researchers. In: ACM Multimedia, pp. 1449–1452 (2012)
26. Wu, B., Haizhou, A., Huang, C.: LUT-based adaboost for gender classification. In: Audio and Video Based Biometric Person Authentication. Springer, Berlin Heidelberg (2003)