International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2013): 6.14 | Impact Factor (2013): 4.438

Face Recognition with Local Binary Patterns, Spatial Pyramid Histograms and Nearest Neighbor Classification

Abhijeet Tayde[1], A. S. Deshpande[2]

[1]Department of Electronics and Telecommunication, JSPM's I.C.O.E.R., Wagholi, Pune(412207) Maharastra, India

[2]Professor, Department of Electronics and Telecommunication, JSPM's I.C.O.E.R., Wagholi, Pune(412207) Maharastra, India

Abstract: ***Face recognition algorithms generally assume that face images are well aligned and have a similar pose yet in many different practical applications it is impossible to meet these certain conditions. Thus extending face recognition to unconstrained face images has become an active area for research. At this end, histograms of Local Binary Patterns (LBP) have proven to be highly discriminative descriptors for face recognition. Most LBP-based algorithms use a rigid descriptor matching strategy that's not robust against pose variation and misalignment. Here two algorithms are proposed for face recognition which are designed to deal with pose variations and misalignment. It also incorporate an illumination normalization step that increases robustness against lighting variations. The proposed algorithms use descriptors based on histograms of LBP and perform descriptor matching with spatial pyramid matching (SPM) and Naive Bayes Nearest Neighbor (NBNN) respectively. The main contribution is the inclusion of flexible spatial matching schemes; it uses an image-to-class relation to provide an improved robustness with respect to intra-class variations. The comparison is compulsory between the accuracy of the proposed algorithms against Ahonen's original LBP-based face recognition system and two baseline holistic classifiers on four standard datasets. Results indicate that the algorithm based on NBNN outperforms the other solutions and does so more markedly in presence of pose variations.***

Keywords: face recognition; local binary patterns; naïve Bayes; nearest neighbor; spatial pyramid.

1. Introduction

Most face recognition algorithms are designed to work best with well aligned, illuminated and frontal pose face images. In many possible applications, however its not possible to meet these certain conditions. Some examples are surveillance, automatic tagging and human robot interaction. Therefore, there have been many recent efforts to develop algorithms that perform well with unconstrained face images.

In this the of use local appearance descriptors such as Gabor jets, SURF, SIFT, HOG and histograms of Local Binary Patterns have become increasingly common. Algorithms that use local appearance descriptors are more robust against occlusion, expression variation, pose variation and small sample sizes than traditional holistic algorithms.

In this work will focus on descriptors based on Local Binary Patterns (LBP), as they are simple computationally efficient and have proved to be highly effective features for face recognition. Nonetheless the methods described can be readily adapted to operate with alternative local descriptors. Within LBP-based algorithms, most of the face recognition algorithms using LBP follow the approach proposed by Ahonen et al. In this approach the face image is divided into a grid of small of non overlapping regions where a histogram of the LBP for each region is constructed. The similarity of two images is computed by summing the similarity of histograms from corresponding regions. The drawback of the previous method is it assumes that a given image region corresponds to the same part of the face in all the faces in the dataset. This is only possible if the face images are fully frontal, scaled, and aligned properly. In addition to this, LBP are invariant against monotonic grayscale transformations. They are still affected by illumination changes that induce non monotonic gray-scale changes such as self shadowing.

Here two algorithms for face recognition propose and compare which are specially designed to deal with moderate pose variations and misaligned faces. These algorithms are based on earlier techniques from the object recognition literature: spatial pyramid matching and Naive Bayes Nearest Neighbors(NBNN). Our main contribution in this is the inclusion of flexible spatial matching schemes based on an "image-to-class" relation which provides an improved robustness with respect to intra-class variations. These matching schemes use spatially dependent variations of the "bag of words" models with LBP histogram descriptors. As a further refinement, also incorporate a state of the art illumination compensation algorithm to improve robustness against illumination changes.

2. Algorithms

By summarizing the main steps of the algorithms used. Then by describing each step in detail. The proposed face recognition process consists of four main parts:

1) Preprocessing:
It begin by applying the Tan and Triggs' illumination normalization algorithm to compensate for illumination variation in the face image. No further preprocessing such as face alignment is performed.

2) LBP operator application:
In this second stage LBP are computed for each pixel, creating a fine scale textural description of the image.

3) Local feature extraction:
Local features are created by computing histograms of LBP over local image regions.

4) Classification:
Each face image in test set is classified by comparing it against the face images in the training set. The comparison performed using the local features obtained in the previous step. The first two steps are shared by all the algorithms.

A. Preprocessing

Illumination accounts for a large part of the variation in appearance of face images. Various preprocessing methods have been created to compensate for variation. We have chosen to use the method proposed by Tan and Triggs since it is simple, efficient and has been shown to work well with local binary patterns.

The algorithm consists of four steps:

1) Gamma correction to enhance the dynamic range of dark regions and compress light areas and highlights. We use γ= 0:2.
2) Difference of Gaussians (DoG) filtering that acts as a "band pass" partially suppressing high frequency noise and low frequency illumination variation. For the width of the Gaussian kernels we use σ_0 = 1:0 and σ_1 = 2:0.
3) Contrast equalization to rescale image intensities in order to standardize intensity variations. The equalization is performed in two steps:

$$I(x,y) \leftarrow \frac{I(x',y')}{mean\left(\left(|I(x',y')|^{a}\right)\right)^{1/a}} \quad (1)$$

where I(x, y) refers to the pixel in position (x,y) of the image I and $\tau$and a are parameters. We use a = 0:1 and τ= 10.

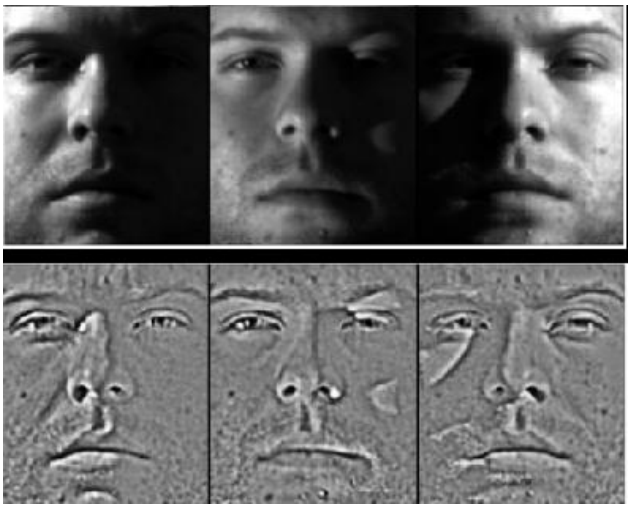

Figure 1: The upper row shows three images of a subject from the Yale B dataset under different lighting conditions. The bottom row shows the same images after processing with Tan and Triggs' illumination normalization algorithm. Appearance variation due to lighting is drastically reduced.

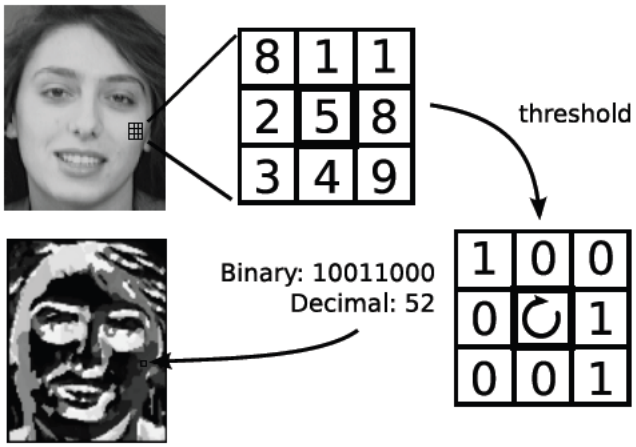

Figure 2: The LBP operator thresholds each pixel against its neighboring pixels and interprets the result as a binary number. In the bottom image each gray-level value corresponds to a different local binary pattern.

4) Compress all values into the range (0; 1) with a hyperbolic tangent function:

$$I(x,y) \leftarrow 0.5 \tanh\left(I\left(x', \frac{y'}{\tau}\right)\right) + 0.5 \quad (2)$$

The values of the parameters $\gamma, \sigma_0, \sigma_1, a \; and \; \tau$ are those suggested by Tan and Triggs. Figure 1 illustrates the effects of the illumination compensation.

B. Local Binary Patterns

Local binary patterns were introduced by Ojala et al, as a fine scale texture descriptor. In its simplest form, an LBP description of a pixel is created by thresholding the values of the 3*3 neighborhood of the pixel against the central pixel and interpreting the result as a binary number. The process is illustrated in figure 2. The LBP operator is generalized by allowing larger neighborhood radii r and different number of sampling points s. And the parameters are indicated by the notation LBPs,r. For example, the original LBP operator with radius of 1 pixel and 8 sampling points is LBP8;1.

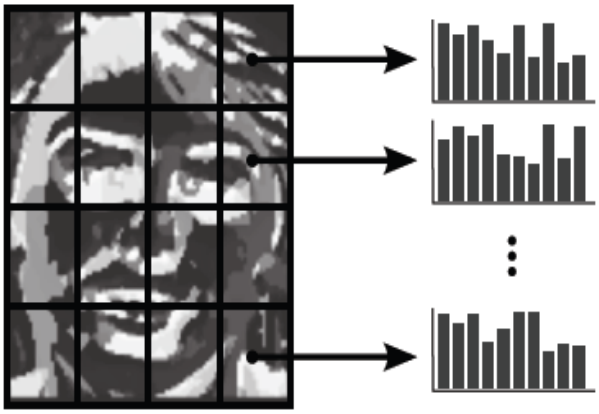

Figure 3: LBP descriptors are built by partitioning the LBP face image into a grid and computing LBP histograms over each grid cell. These histograms may then be concatenated into a vector or treated as individual descriptors.

Another important extension is the definition of "uniform patterns". An LBP is defined as uniform if it contains at

most two 0-1 or 1-0 transitions when viewed as a circular bit string. Thus the 8-bit strings 01100000 and 00000000 are uniform, while 01010000 and 00011010 are not. Ojala observed that when using 8 sampling points, uniform patterns accounted for nearly 90% of the patterns in their image datasets. Therefore, little information is lost by assigning all non uniform patterns to a single arbitrary number. Since only 58 of the 256 possible 8 bit patterns are uniform, and enables significant space savings when building LBP histograms. To indicate the usage of two-transition uniform patterns the superscript u2 is added to the LBP operator notation. Hence the LBP operator with a 2 pixel radius, 8 sampling points and uniform patterns is known as LBPu2 8;2. The success of LBP has inspired several variations. These include local ternary patterns elongated local binary patterns, multi scale LBP, centralized binary patterns and patch based LBP among others. During this work we use LBPu2 8;2, which was chosen by Ahonen in their pioneering work applying LBP to face recognition. This descriptor has been used by itself or in combination with other features, by most methods that use LBP for face recognition.

C. Face description and recognition

In order to build the description of a face image we follow the basic methodology proposed by Ahonen. Once the LBP operator is applied to the face image, the face image is divided into regions and a histogram of LBP is computed for each region. The final description of each face is a set of local histograms and this process is illustrated. Given the face description, different recognition schemes are possible. As mentioned earlier, Ahonen's original method is not very robust to pose variations and face misalignment. Here we have to explore two additional approaches to counter this problem which are based on spatial pyramid matching and the Naive Bayes Nearest Neighbor schemes. The following sections present more details on the face description and recognition systems used by each method.

1) Ahonen system:

In Ahonen's system, each face image is partitioned into a grid of non-overlapping square regions. A LBP histogram is computed independently for each region. Then all the resulting histograms are concatenated together into a large vector. Ahonen call this vector a "spatially enhanced histogram", since the order of histograms that compose it implicitly encode spatial information. This method tends to produce fairly high dimensional vectors. For example, if an image is divided into an 8 *8 grid and the LBPu2 8;2 operator is used (so the histograms have length 59) the spatially enhanced histogram has length 8* 8* 59 = 3776. In order to perform face recognition, each face image in the training and test sets is converted to a spatially enhanced histogram. Then ordinary nearest neighbor classification is performed with a histogram distance measure such as x^2 or histogram intersection. In this, we use the x^2 to measure distance between histograms.

$$X^2(x,y) = \sum_{i=1}^{D} \frac{(x_i - y_i)^2}{(x_i + y_i)} \quad (3)$$

where D is the dimensionality of the spatially enhanced histograms. In tests this measure performed slightly better than histogram intersection. We have not tested the weighted variations of this distance that Ahonen also explore in their work.

2) Spatial Pyramid Match:

One of the parameters for Ahonen's system is the size of the regions. Though Ahonen report that their algorithm is relatively robust to small variations of this parameter, the election of a region size is somewhat arbitrary and is subject to aliasing effects. Furthermore, Ruiz del Solar report that while using larger regions is more robust against face misalignment, it has less discriminative power. This has motivated us to explore the combination of multiple LBP histograms at various resolutions as an alternative to the Ahonen grid representation.

In order to create the multi-resolution LBP histogram use the spatial pyramid histogram approach introduced by Lazebnik which is based on the pyramid histogram of Grauman. Lazebnik successfully used spatial pyramid histograms to match sets of quantized SIFT descriptors for the task of object recognition. In a similar task, Bosch, use spatial pyramid histogram of intensity gradients to compute shape similarity. The process of building the spatial pyramid histogram is similar to building Ahonen's spatially enhanced histograms at various resolutions and concatenating the results. More precisely, a spatial pyramid histogram with L levels is built by first creating the level 0 histogram with the LBP over theentire image. Next, the image is divided in four equal sized regions and a level 1 LBP histogram is computed for each region. The process is repeated by recursively subdividing each region and computing level l histograms in each region until the desired level L is reached. A simple calculation shows that there will 22l level l histograms and that by summing this number over l = 0,……..L a spatial pyramid histogram with L levels will have a total of $2^{2L+2} - \frac{1}{3}$histograms. As in Ahonen's method all these histograms are concatenated together into a large vector 1. For example, if we describe a face image with a three level spatial pyramid (L = 3) and LBPu2 8;1, the resulting vector has length ($2^{2L+2} - \frac{1}{3}$)* 59 = 5015.

For classification a nearest neighbor classifier is used, as in the Ahonen system. However, to compare histograms use a distance based on the Pyramid Match Kernel with some of the modifications used by Bosch instead of plain x^2. The motivation behind this distance is that matches among histograms at coarser resolutions should be given less weight because it is less likely than they come from corresponding face parts. Specifically, if we have two spatial pyramids x and y and we denote by _l the sum of the distance between all the histograms at level l then the distance is calculated as

$$d(x,y) = \frac{\delta_0}{2^L} + \sum_{l-1}^{L} \frac{\delta_l}{2^{2L-l+1}} \quad (4)$$

3) Naive Bayes Nearest Neighbor:

While expecting spatial pyramid histograms to be more robust to face misalignment and pose variation than Ahonen's spatially enhanced histograms, they still have a rigid approach to spatial matching. As in Ahonen's method, when two face images are compared each local feature in one image is compared against the local feature found at the

same position in the other image. This suggests a more flexible spatial matching approach, where in local features from one image are allowed to be matched to local features found in different positions from other images.

This idea evokes the "bag of visual words" approach that has proved successful in object recognition and scene classification. However, it seems unwise to discard all spatial information given that it clearly is useful for visual recognition, as shown by work incorporating spatial information into the bag of words model. Another disadvantage of the bag of words model is that it requires a codebook creation stage which tends to lose discriminative information.

In this we test an intermediate approach introduced by Boiman in the context of visual object recognition using local descriptors. Since the method is based on the Nearest Neighbor classifier and makes a naive Bayes assumption it is named "Naive Bayes Nearest Neighbor" (NBNN). NBNN assumes images are represented by sets of local features. Boiman's work uses a combination of various visual descriptors including SIFT and Shape Contexts. For this we use the aforementioned LBP histograms over local regions as descriptors. To make the algorithms comparable use the same grid-based regions as the Ahonen method. Nonetheless instead of concatenating the histograms of each region into a single vector, each histogram is kept separate. To keep track of spatial information the histograms are augmented with the (x; y) coordinates of the center of its region. Therefore under this scheme each face is not described by a single vector as in the previous two approaches but by a set of vectors.

Supposing the LBP descriptors have been extracted for all face images in the training set, the NBNN classification procedure for a test face image P is summarized in algorithm 1. One of the intuitions behind this algorithm is that instead of minimizing an "image-to-image" distance (as the othernearest neighbor classifiers in this paper) it minimizes an "image-to-class" distance by aggregating the descriptors from all the images of each subject. Suppose we have a probe image P and wish to find gallery subject $\hat{G}$ it belongs to with the maximum a posteriori (MAP) criterion. If we assume the priors p(G) to be uniform, we have

$$\hat{G} = \arg max\, p(G/P) = \arg max\, p(P/G) \quad (5)$$

Applying log,

$$\hat{G} = \arg max \sum_{i=1}^{n} \log P\left(\frac{d_i}{G}\right) \quad (6)$$

Set this parameter by cross-validating in a small in house face dataset. We found $\alpha = 1$ to be a good choice and used this value with all the datasets. Since not all datasets use the same image size to make the influence of α commensurate across datasets linearly scale all (x, y) coordinates so the upper left corner of the image is at (0, 0) and the lower right corner is at (1, 1).

The flexible spatial matches used by NBNN are advantageous in datasets with misalignment and pose variations. However, this flexibility comes at a computational cost. If we denote the number of descriptors per image by nD, the number of training images per subject by ns and the number of subjects in the training set by nG, it is clear that each query takes O(ns .n2 D. nG) time using linear nearest neighbor search 2.

This lead us to test a slight variation of NBNN, which dub Restricted Naive Bayes Nearest Neighbor (RNBNN). In RNBNN the restricted descriptor matches to be from the same position in the image. This is equivalent to using a very large value for α and reduces the computational cost to O(ns.nD. nG),the same as Ahonen's method. While RNBNN should perform worse than NBNN in unconstrained face images, it still reaps the benefits of aggregating the descriptors from the same subject which allows it to use the training data more fully than Ahonen's method. Moreover, when images are well aligned it may actually perform better than NBNN by avoiding descriptor mismatches (i.e. matching descriptors from different facial regions).

An intermediate approach between ordinary NBNN and RNBNN is to restrict descriptor matches to be from a predefined spatial neighborhood in the image, thus reducing computational cost by making less distance comparisons. These tests suggest this method has a very similar accuracy to ordinary NBNN. Since it can be considered as a simple speed optimization with respect to NBNN do not present further results on this approach.

3. Experiments and Results

1) Datasets

We perform experiments on four datasets: AT&T-ORL, Yale, Georgia Tech and Extended Yale B. These datasets differ in the degree of variation of pose, illumination and expression present in their face images. The main characteristics of each dataset are summarized in table I. Regarding the image size, cropping, and alignment of the datasets:

- For AT&T-ORL we used the original images at 112 * 92.
- For Yale the face area was extracted with Viola Jonesdetector implementation from OpenCV and resized to 128 * 128.
- The cropped version of the Georgia Tech dataset was used and the images were resized to 156* 111.
- For Extended Yale B, the manually cropped and aligned subset from [36] was used at the original size of 192*168.

Table1: Summary of all Datasets

Dataset	No. sub-jects	Total images	Variation	Ref.
AT&T-ORL	40	400	pose, expression, eye glasses	[34]
Yale	15	165	expression, eye glasses, lighting	[18]
Georgia Tech	50	750	pose, expression, scale, orientation	[35]
Ext. Yale B (frontal)	38	2414	lighting	[36]

Table 2: Results for AT&T-ORL Dataset

Method	With TT (%)	Without TT (%)
AH	95	95.45
SPM	96.7	97.16
NBNN	98.4	**99.35**
RNBNN	96.82	95.6
Eig	50.95	93.3
Fish	64.32	92.58

Table 3: Results for YALE Dataset

Method	With TT (%)	Without TT (%)
AH	97.91	84.05
SPM	96.96	82.65
NBNN	**98.18**	86.81
RNBNN	97.39	88.45
Eig	57.72	74.94
Fish	67.84	91.25

2) Evaluation Methodology

Comparison of the three algorithms we have described in this paper and add the results of two classic holistic algorithms, Eigenfaces and Fisherfaces as a baseline. For each algorithm show the results with and without the DoG illumination normalization. For each dataset use approximately half of the subjects per class as training set and the rest as test. Specifically, 5, 5, 7 and 31 training images were used for the AT&TORL, Yale, Georgia Tech and Extended Yale B datasets respectively. The reported accuracy is the average over 10 runs, with a different training and test set partition used in each run.

3) Algorithm parameters:

The major parameter for the LBP-based algorithms is the the size of regions used for LBP histograms i.e. the characteristics of the grid used to partition the images. We tested 6*6, 7*7 and 8*8 grids in a small in-house face dataset. We found 88 to give slightly better results for the Ahonen and NBNN algorithms, so we use this grid size for all the datasets.

For the spatial pyramid algorithm we chose a three level pyramid (L = 3), because this gives an 8*8 grid at the finest level. This makes the results for this algorithm more comparable to the results on the other two. For the holistic algorithms the major parameter is the dimensionality of the subspace on which the data is projected. For the Eigenfaces algorithm varied the dimensionality D from 10 to 150 in increments of 10 and report the best accuracy. This was obtained with D = 50 for AT&T-ORL, D = 30 for Yale, D = 50 for Georgia Tech and D = 120 for Extended Yale B. In the Fisherface algorithm we varied dimensionality from 5 to the maximum dimensionality supported by the algorithm which is one less than the number of classes in the dataset. In all the datasets the best results were obtained by setting D to the largest value possible.

Regarding these experiments we make a few observations:

- NBNN is the clear winner in the less constrained datasets such as Georgia Tech. It also has the best performance in Yale and AT&T-ORL. However, in Extended Yale B with illumination normalization it falls behind the holistic algorithms (though it performs better than them with no illumination normalization). This is explained by the fact that Extended Yale B subset is a very well aligned dataset which only varies illumination, a situation where holistic algorithms and Fisherfaces in particular, work well.
- RNBNN performed somewhat better than the Ahonen algorithm, specially when illumination normalization is not used. As expected, the performance of RBNN suffers in less constrained datasets. On the other hand, in the well aligned Yale B dataset it actually worked better than ordinary NBNN and was the best algorithm with no illumination normalization.
- Spatial pyramid histograms perform slightly better than Ahonen's method in the less constrained datasets. However, it performed slightly worse in the well aligned Extended Yale B dataset as well as the Yale dataset. This suggests that most of the discriminative power of the pyramids is in the highest level.
- In face datasets with large illumination variations (Yale and Extended Yale B) Tan and Triggs' illumination normalization algorithm boosts the accuracy of LBP based classifiers significantly. Holistic classifiers only benefited in Extended Yale B. In the rest the illumination normalization lowers their accuracy to a surprising degree. We found that in these cases the decrease was inversely proportional to the width of the DoG bandpass filter. In face datasets with little or no lighting variation, LBP based perform slightly worse with Tan and Triggs' algorithm while the holistic algorithms still perform significantly worse.
- The behavior of RNBNN and NBNN in the Extended Yale B dataset with no illumination normalization is interesting; they outperform the other LBP-based algorithms by a 20% margin. This is a consequence of aggregating the descriptors for each class because it allows each face region to be matched to a similarly illuminated face region from the training set, in a certain sense inferring a new face by "composing pieces" from various face images.

4. Conclusions

Our main result is that the NBNN algorithm improves performance substantially with respect to the original LBP based algorithm when used in relatively unconstrained face datasets. NBNN also outperforms the original LBP algorithm even when faces are frontal and well aligned, though by a smaller margin. This improvements may be attributed to the flexible spatial matching scheme and the use of the "image-to-class" distance, which makes a better use of the training data than the "image-to-image" distance.

References

[1] J. Wright and G. Hua, "Implicit elastic matching with random projections for Pose-Variant face recognition," *in Proc. CVPR, 2009*.
[2] P. Dreuw, P. Steingrube, H. Hanselmann, and H. Ney, "SURFFace: face recognition under viewpoint consistency constraints," in British *Machine Vision Conference*, London, UK, Sep. 2009.
[3] L. Wolf, T. Hassner, and Y. Taigman, "Descriptor based methods in the wild," in *Proc. ECCV*, 2008.
[4] J. Ruiz-del-Solar, R. Verschae, and M. Correa, "Recognition of faces in unconstrained environments: A comparative study," *EURASIP Journal on Advances in Signal Processing*, vol. 2009, pp. 1–20, 2009.
[5] J. Zou, Q. Ji, and G. Nagy, "A comparative study of local matching approach for face recognition," *Image Processing, IEEE Transactions* on, vol. 16, no. 10, pp. 2617–2628, 2007.
[6] X. Tan and B. Triggs, "Fusing gabor and LBP feature sets for Kernel-Based face recognition," in *Analysis and Modeling of Faces and Gestures*, 2007, pp. 235–249.
[7] H. Bay, T. Tuytelaars, and L. V. Gool, "Surf: Speeded up robust features*," Lecture notes in computer science*, vol. 3951, p. 404, 2006.
[8] D. G. Lowe, "Distinctive image features from Scale-Invarian keypoints," *Int. J. Comput. Vision*, vol. 60, no. 2, pp. 91–110, 2004.
[9] M. Bicego, A. Lagorio, E. Grosso, and M. Tistarelli, "On the use of SIFT features for face authentication," in Proceedings of the 2006 *Conference on Computer Vision and Pattern Recognition Workshop*. IEEE Computer Society, 2006, p. 35.
[10] A. Albiol, D. Monzo, A. Martin, J. Sastre, and A. Albiol, "Face recognition using HOG-EBGM*," Pattern Recogn. Lett.*, vol. 29, no. 10, pp. 1537–1543, 2008.
[11] T. Ojala, M. Pietikainen, and T. Maenpaa, "Gray scale and rotation invariant texture classification with local binary patterns," *Lecture Notes in Computer Science*, vol. 1842, p. 404420, 2000.
[12] T. Ahonen, A. Hadid, and M. Pietikainen, "Face description with local binary patterns: Application to face recognition*," IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 12, pp. 2037–2041, 2006.
[13] Y. Rodriguez and S. Marcel, "Face authentication using adapted local binary pattern histograms," *Lecture Notes in Computer Science*, vol. 3954, p. 321, 2006.
[14] S. Lazebnik, C. Schmid, and J. Ponce, "Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories," in Proceedings of the *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Volume 2. IEEE Computer Society*, 2006, pp. 2169–2178.