Face Recognition with Local Binary Patterns, Spatial Pyramid Histograms and Naive Bayes Nearest Neighbor classification

Daniel Maturana, Domingo Mery and Álvaro Soto
Departamento de Ciencias de la Computación
Pontificia Universidad Católica
Santiago, Chile
Email: {dimatura, dmery, asoto}@uc.cl

***Abstract*—Face recognition algorithms commonly assume that face images are well aligned and have a similar pose – yet in many practical applications it is impossible to meet these conditions. Therefore extending face recognition to unconstrained face images has become an active area of research.**

To this end, histograms of Local Binary Patterns (LBP) have proven to be highly discriminative descriptors for face recognition. Nonetheless, most LBP-based algorithms use a rigid descriptor matching strategy that is not robust against pose variation and misalignment.

We propose two algorithms for face recognition that are designed to deal with pose variations and misalignment. We 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. Our contribution is the inclusion of flexible spatial matching schemes that use an image-to-class relation to provide an improved robustness with respect to intra-class variations.

We compare the accuracy of the proposed algorithms against Ahonen's original LBP-based face recognition system and two baseline holistic classifiers on four standard datasets. Our 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; naive Bayes; nearest neighbor; spatial pyramid.**

I. Introduction

Most face recognition algorithms are designed to work best with well aligned, well illuminated, and frontal pose face images. In many possible applications, however, it is not possible to meet these 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 [1]–[4].

In this context, the of use local appearance descriptors such as Gabor jets [5], [6], SURF [7], SIFT [8], [9], HOG [10] and histograms of Local Binary Patterns [11] 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 [4], [5].

In this work we 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 [3], [4], [12], [13]. Nonetheless, the methods described in this paper 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 [12]. 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 then computed by summing the similarity of histograms from corresponding regions.

One drawback of the previous method is that 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, while LBP are invariant against monotonic gray-scale transformations, they are still affected by illumination changes that induce non monotonic gray-scale changes such as self shadowing [17].

In this paper, we propose and compare two algorithms for face recognition that are specially designed to deal with moderate pose variations and misaligned faces. These algorithms are based on previous techniques from the object recognition literature: spatial pyramid matching [14], [15] and Naive Bayes Nearest Neighbors (NBNN) [16]. Our main contribution 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, we also incorporate a state of the art illumination compensation algorithm to improve robustness against illumination changes [17].

This paper is organized as follows. Section II discusses the details of our approach. Section III-C shows the results of applying our methodology to standard datasets. Finally, section IV presents the main conclusions of this work.

II. ALGORITHMS

We start by summarizing the main common steps of the algorithms used in this work. Then we describe each step in detail. The proposed face recognition process consists of four main parts:

1) Preprocessing: We begin by applying the Tan and Triggs' illumination normalization algorithm [17] to compensate for illumination variation in the face image. No further preprocessing, such as face alignment, is performed.
2) LBP operator application: In the 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 is performed using the local features obtained in the previous step.

The first two steps are shared by all the algorithms. The algorithms we explore in this work vary in how they perform the last two steps, as we detail in section II-C.

A. Preprocessing

Illumination accounts for a large part of the variation in appearance of face images [18]. Various preprocessing methods have been created to compensate for this variation [19]. We have chosen to use the method proposed by Tan and Triggs [17] 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 $\gamma = 0.2$.
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 $\sigma_0 = 1.0$ and $\sigma_1 = 2.0$.
3) Contrast equalization to rescale image intensities in order to standardize intensity variations. The equalization is performed in two steps:

$$\mathrm{I}(x,y) \leftarrow \frac{\mathrm{I}(x',y')}{(\mathrm{mean}(|\,\mathrm{I}(x',y')|^a))^{1/a}}$$
$$\mathrm{I}(x,y) \leftarrow \frac{\mathrm{I}(x',y')}{(\mathrm{mean}(\min(\tau, |\,\mathrm{I}(x',y')|)^a))^{1/a}}$$

where $\mathrm{I}(x,y)$ refers to the pixel in position (x,y) of the image I and τ and a are parameters. We use $a = 0.1$ and $\tau = 10$.

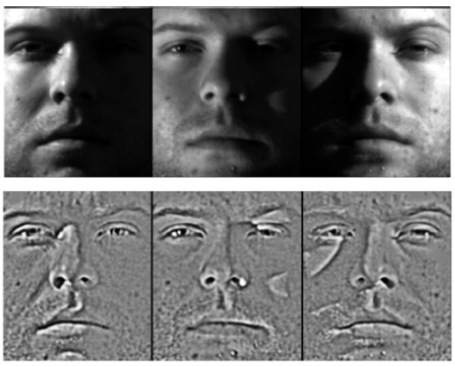

Figure 1. The upper row shows three images of a subject from the Yale B dataset under different lighting conditions. The bottom row shows the same images after processing with Tan and Triggs' illumination normalization algorithm. Appearance variation due to lighting is drastically reduced.

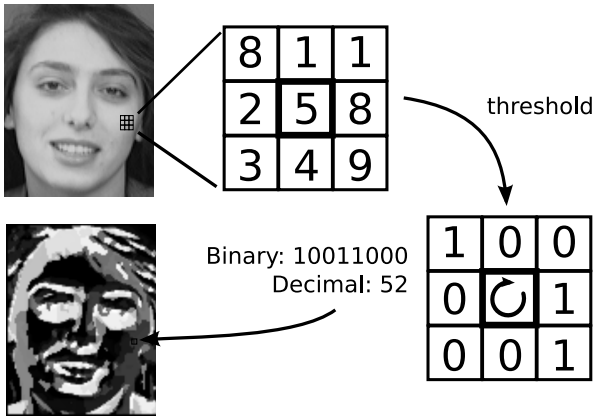

Figure 2. The LBP operator thresholds each pixel against its neighboring pixels and interprets the result as a binary number. In the bottom image each gray-level value corresponds to a different local binary pattern.

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

$$\mathrm{I}(x,y) \leftarrow 0.5\tanh(\mathrm{I}(x',y')/\tau) + 0.5$$

The values of the parameters γ, σ_0, σ_1, a and τ 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 [20] 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.

In [11] the LBP operator is generalized by allowing larger neighborhood radii r and different number of sampling points s. These parameters are indicated by the notation $LBP_{s,r}$. For example, the original LBP operator with radius of 1 pixel and 8 sampling points is $LBP_{8,1}$. Another

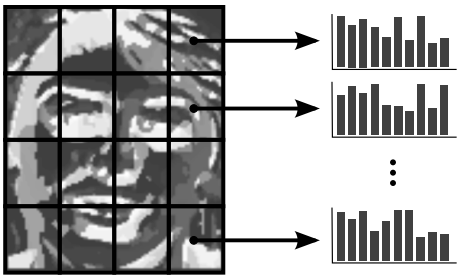

Figure 3. LBP descriptors are built by partitioning the LBP face image into a grid and computing LBP histograms over each grid cell. These histograms may then be concatenated into a vector or treated as individual descriptors.

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, this 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 $LBP^{u2}_{8,2}$.

The success of LBP has inspired several variations. These include local ternary patterns [17], elongated local binary patterns [21], multi scale LBP [22], centralized binary patterns [23] and patch based LBP [3], among others.

In this work we use $LBP^{u2}_{8,2}$, which was chosen by Ahonen et al [12] 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 (e.g. [3], [6], [24], [25]).

C. Face description and recognition

In order to build the description of a face image we follow the basic methodology proposed by Ahonen [12]. 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. This process is illustrated in 3.

Given the face description, different recognition schemes are possible. As mentioned in the introduction, Ahonen's original method is not very robust to pose variations and face misalignment. Here, we explore two additional approaches to counter this problem, which are based on spatial pyramid matching [14] and the Naive Bayes Nearest Neighbor [16] schemes.

In the following sections we 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. An LBP histogram is computed independently for each region. Then, all the resulting histograms are concatenated together into a large vector. Ahonen et al 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 $LBP^{u2}_{8,2}$ 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 under this scheme, each face image in the training and test sets is converted to a spatially enhanced histogram via the process described above. Then ordinary nearest neighbor classification is performed with a histogram distance measure such as χ^2 or histogram intersection [26]. In this work we use the χ^2 to measure distance between histograms:

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

where D is the dimensionality of the spatially enhanced histograms. In our preliminary tests this measure performed slightly better than histogram intersection. We have not tested the weighted variations of this distance that Ahonen et al 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 et al 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 et al [4] 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 we use the spatial pyramid histogram approach introduced by Lazebnik et al [14], which is based on the pyramid histogram of Grauman [27]. Lazebnik successfully used spatial pyramid histograms to match sets of quantized SIFT descriptors for the task of object recognition. In a similar task, Bosch et al [28], 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 the

entire 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 2^{2l} level l histograms and that by summing this number over $l = 0, \ldots, L$ a spatial pyramid histogram with L levels will have a total of $(2^{2L+2} - 1)/3$ 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 $LBP^{u2}_{8,1}$, the resulting vector has length $((2^{2*3+2} - 1)/3) * 59 = 5015$.

For classification a nearest neighbor classifier is used, as in the Ahonen system. However, to compare histograms we use a distance based on the Pyramid Match Kernel [27] with some of the modifications used by Bosch [28] instead of plain χ^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 δ_l the sum of the distance between all the histograms at level l (we use χ^2, as in [28]) then the distance is calculated as

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

3) Naive Bayes Nearest Neighbor: While we expect 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, wherein 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 (e.g. [15], [29]). 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 [14], [30]. Another disadvantage of the bag of words model is that it requires a codebook creation stage which tends to lose discriminative information, as shown in [16].

In this paper we test an intermediate approach, introduced by Boiman et al [16] 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 [8] and Shape Contexts [32]. In this paper we use the aforementioned LBP histograms over local regions as descriptors. To make the algorithms comparable we 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.

Algorithm 1 NBNN algorithm

```
{Input: probe face image P}
{Output: gallery subject Ĝ}
Extract descriptors d_1, ... d_n from test image P
for i = 1 to n do
  for each training subject G do
    NN_G(d_i) ← NN of d_i among images of G
  end for
end for
Ĝ ← arg min_G Σ_{i=1}^{n} ||d_i − NN_G(d_i)||^2
```

One of the intuitions behind this algorithm is that instead of minimizing an "image-to-image" distance (as the other nearest neighbor classifiers in this paper) it minimizes an "image-to-class" distance by aggregating the descriptors from all the images of each subject. This intuition is justified by the following reasoning, presented in [16]. Suppose we have a probe image P and we 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

$$\hat{G} = \arg\max_G p(G|P) = \arg\max_G p(P|G)$$

We assume the image descriptors to be independent given the subject g they belong to (Naive Bayes assumption):

$$p(P|G) = p(d_1, \ldots, d_n) = \prod_{i=1}^{n} p(d_i|G)$$

Applying log,

$$\hat{G} = \arg\max_G \sum_{i=1}^{n} \log p(d_i|G) \qquad (1)$$

[1] We modify slightly the construction process of the pyramid used by Lazebnik in order to emphasize the similarities with Ahonen's grid spatially enhanced histogram, but by modifying the kernel function appropriately the results are equivalent. In particular, instead of treating the LBP "channels" separately we interleave them.

Rewriting the right hand side using the law of total probability,

$$\hat{G} = \arg\max_G \sum_d p(d|P) \log p(d|G)$$

where we sum over the space of all possible descriptors d. By subtracting the constant term $\sum_d p(d|P) \log p(d|P)$ on the right side (which does not affect $\hat{G}$) and rearranging,

$$\begin{aligned}\hat{G} &= \arg\max_G \left(\sum_d p(d|P) \log \frac{p(d|G)}{p(d|P)} \right) \\ &= \arg\min_G \mathrm{KL}\left(p(d|P) \| p(d|G)\right)\end{aligned}$$

where $\mathrm{KL}(\cdot\|\cdot)$ is the Kullback-Leibler divergence between two distributions. Thus in this case the MAP criterion is equivalent to minimizing the KL divergence between the descriptor distributions of P and the descriptor distribution of the subject $\hat{G}$ (i.e. the "image-to-class" distance).

We have not specified how to calculate (1), and in particular $p(d|G)$. The NBNN approach is to approximate the Parzen likelihood estimator for $p(d|G)$ with the r nearest neighbors NN_j, $j = 1 \dots r$ of d belonging to G:

$$p_{NN}(d|G) = \frac{1}{L} \sum_j^r \mathrm{K}(d - d^G_{NN_j}) \tag{2}$$

where K is the Gaussian Parzen kernel function: $\mathrm{K}(d - d^G_j) = \exp(\frac{1}{2\sigma^2}\|d - d^G_j\|^2)$. If $r = 1$, corresponding to a single nearest neighbor approximation, (2) becomes a simple expression and the constant factors such as σ^2 may be ignored. Then (1) becomes:

$$\hat{G} = \arg\min_G \sum_{i=1}^n \|d_i - NN_G(d_i)\|^2$$

which is the expression used in algorithm 1.

Boiman et al incorporate spatial information into this scheme by appending (x, y) pixel coordinates to each descriptor, scaled by a factor α. Thus the squared euclidean distance between two descriptors d_1 and d_2 at positions (x_1, y_1) and (x_2, y_2) becomes

$$\sum_i (d_{1i} - d_{2i})^2 + \alpha \left((x_1 - x_2)^2 + (y_1 - y_2)^2\right)$$

The value of α determines the weight given to spatial information when matching descriptors. If the value is set to 0 spatial information is completely disregarded. This may be beneficial when dealing with very large pose variations but probably increases the chances of mismatches. In the other extreme, setting α to a very large value forces descriptors to be matched exclusively with descriptors from the same spatial location, as in Ahonen's method.

We 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 we 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, as we show in section III-C. However, this flexibility comes at a computational cost. If we denote the number of descriptors per image by n_D, the number of training images per subject by n_s and the number of subjects in the training set by n_G, it is clear that each query takes $O(n_s \cdot n_D^2 \cdot n_G)$ time using linear nearest neighbor search [2].

This lead us to test a slight variation of NBNN, which we dub Restricted Naive Bayes Nearest Neighbor (RNBNN). In RNBNN we restrict 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 $O(n_s \cdot n_D \cdot n_G)$, 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. Our 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 we do not present further results on this approach.

III. Experiments and results

A. Datasets

We perform experiments on four datasets: AT&T-ORL [34], Yale [18], Georgia Tech [35] and Extended Yale B [36].

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 Jones detector 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.

[2]Using spatial index data structures such as cover trees [33] the complexity can be reduced to $O(n_D \log(n_s \cdot n_D) \cdot n_G)$.

Table I
SUMMARY OF FACE DATASETS

Dataset	No. subjects	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]

- For Extended Yale B, the manually cropped and aligned subset from [36] was used at the original size of 192×168.

B. Evaluation methodology

We compare the three algorithms we have described in this paper and add the results of two classic holistic algorithms, Eigenfaces [37] and Fisherfaces [18] as a baseline. For each algorithm we show results with and without the DoG illumination normalization.

For each dataset we 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&T-ORL, 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.

1) 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 8×8 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 we 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.

Table II
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 III
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

Table IV
RESULTS FOR GEORGIA TECH DATASET

Method	With TT (%)	Without TT (%)
AH	72.9	75.1
SPM	76.07	77.67
NBNN	87.97	**92.67**
RNBNN	76.52	81.2
Eig	6.5	71.3
Fish	16.4	53.05

Table V
RESULTS FOR EXTENDED YALE B DATASET

Method	With TT (%)	Without TT (%)
AH	95.7	73.72
SPM	93.8	72.97
NBNN	97.15	93.2
RNBNN	99.31	94.79
Eig	99.88	60.11
Fish	**99.98**	92.23

C. Results and discussion

Tables II, III, IV and V summarize accuracy of each classifier on the four datasets. For economy of space we use the abbreviations "AH" for Ahonen's system, "SPM" to refer to spatial pyramid matching, "NBNN" for Naive Bayes Nearest Neighbor, "RNBNN" for Restricted Naive Bayes Nearest Neighbor, "Eig" for Eigenfaces, "Fish" for Fisherfaces and "TT" for Tan and Triggs' illumination normalization.

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.

IV. Conclusions and future work

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.

A. Future work

One of the drawbacks of NBNN is the increase in computational cost relative to the original LBP based algorithm. Since this cost is caused by the large amount of nearest neighbor queries it would be beneficial to speed up nearest neighbor queries with spatial index data structures such as cover trees [33] or locality sensitive hashing [38].

Another interesting avenue of research is to complement or replace the $LBP^{u2}_{8,2}$ histogram descriptors with other local descriptors, such as SIFT [8], SURF [7] or one of the many LBP variations. Furthermore, we are currently exploring strategies to learn a discriminative LBP-like descriptor from the data itself.

It would also be of interest to find a better alternative to the grid-based regions used in this paper. The grid partition has no natural relation to the shape of the face and suffers from quantization effects. One possibility is to detect "interesting" facial regions (such as the eyebrows, nose and mouth) and extract descriptors in these selected regions.

Acknowledgment

This work was partially funded by FONDECYT grant 1095140 and LACCIR Virtual Institute grant No. R1208LAC005 (http://www.laccir.org).

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, "SURF-Face: 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-Invariant 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.

[15] A. Bosch, A. Zisserman, and X. Muoz, "Scene classification via pLSA," in *Computer Vision ECCV 2006*, 2006, pp. 517–530.

[16] O. Boiman, E. Shechtman, and M. Irani, "In defense of Nearest-Neighbor based image classification," in *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, 2008, pp. 1–8.

[17] X. Tan and B. Triggs, "Enhanced local texture feature sets for face recognition under difficult lighting conditions," *Lecture Notes in Computer Science*, vol. 4778, p. 168, 2007.

[18] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. fisherfaces: recognition using class specific linearprojection," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 19, no. 7, pp. 711–720, 1997.

[19] J. Ruiz-del-Solar and J. Quinteros, "Illumination compensation and normalization in eigenspace-based face recognition: A comparative study of different pre-processing approaches," *Pattern Recognition Letters*, 2008.

[20] T. Ojala, M. Pietikinen, and D. Harwood, "A comparative study of texture measures with classification based on featured distributions," *Pattern Recognition*, vol. 29, no. 1, pp. 51–59, 1996.

[21] S. Liao and A. Chung, "Face recognition by using elongated local binary patterns with average maximum distance gradient magnitude," in *Computer Vision ACCV 2007*, 2007, pp. 672–679.

[22] S. Liao, X. Zhu, Z. Lei, L. Zhang, and S. Li, "Learning multi-scale block local binary patterns for face recognition," in *Advances in Biometrics*, 2007, pp. 828–837.

[23] X. Fu and W. Wei, "Centralized binary patterns embedded with image euclidean distance for facial expression recognition," in *International Conference on Natural Computation*, vol. 4. Los Alamitos, CA, USA: IEEE Computer Society, 2008, pp. 115–119.

[24] S. Marcel, Y. Rodriguez, and G. Heusch, "On the recent use of local binary patterns for face authentication," *International Journal on Image and Video Processing Special Issue on Facial Image Processing*, 2007.

[25] G. Zhang, X. Huang, S. Li, Y. Wang, and X. Wu, "Boosting local binary pattern (LBP)-Based face recognition," in *Advances in Biometric Person Authentication*, 2005, pp. 179–186.

[26] M. Swain and D. Ballard, "Indexing via color histograms," in *Computer Vision, 1990. Proceedings, Third International Conference on*, 1990, pp. 390–393.

[27] K. Grauman and T. Darrell, "The pyramid match kernel: Discriminative classification with sets of image features," in *Proceedings of the Tenth IEEE International Conference on Computer Vision - Volume 2*. IEEE Computer Society, 2005, pp. 1458–1465.

[28] A. Bosch, A. Zisserman, and X. Munoz, "Representing shape with a spatial pyramid kernel," in *Proceedings of the 6th ACM international conference on Image and video retrieval*. Amsterdam, The Netherlands: ACM, 2007, pp. 401–408.

[29] J. Sivic, B. C. Russell, A. Efros, A. Zisserman, and W. T. Freeman, "Discovering object categories in image collections," in *Proc. ICCV*, vol. 2, 2005.

[30] E. B. Sudderth, A. Torralba, W. T. Freeman, and A. S. Willsky, "Describing visual scenes using transformed dirichlet processes," *Advances in Neural Information Processing Systems 18*, pp. 1299—1306, 2005.

[31] ——, "Learning hierarchical models of scenes, objects, and parts," in *Proceedings of the Tenth IEEE International Conference on Computer Vision - Volume 2*. IEEE Computer Society, 2005, pp. 1331–1338.

[32] S. Belongie and J. Malik, "Matching with shape contexts," in *Content-based Access of Image and Video Libraries, 2000. Proceedings. IEEE Workshop on*, 2000, pp. 20–26.

[33] A. Beygelzimer, S. Kakade, and J. Langford, "Cover trees for nearest neighbor," in *Proceedings of the 23rd international conference on Machine learning*. Pittsburgh, Pennsylvania: ACM, 2006, pp. 97–104.

[34] F. Samaria and A. Harter, "Parameterisation of a stochastic model for human face identification," in *Applications of Computer Vision, 1994., Proceedings of the Second IEEE Workshop on*, 1994, pp. 138–142.

[35] A. V. Nefian, M. Khosravi, and M. H. Hayes, "Real-Time detection of human faces in uncontrolled environments," *Proceedings of SPIE Conference on Visual Communications and Image Processing*, vol. 3024, pp. 211—219, 1997.

[36] A. S. Georghiades, P. N. Belhumeur, and D. J. Kriegman, "From few to many: Illumination cone models for face recognition under variable lighting and pose," *IEEE Trans. Pattern Anal. Mach. Intelligence*, vol. 23, no. 6, pp. 643–660, 2001.

[37] M. Turk and A. Pentland, "Eigenfaces for recognition," *Journal of cognitive neuroscience*, vol. 3, no. 1, pp. 71–86, 1991.

[38] A. Andoni and P. Indyk, "Near-optimal hashing algorithms for approximate nearest neighbor in high dimensions," *Commun. ACM*, vol. 51, no. 1, pp. 117–122, 2008.