Combining different local binary pattern variants to boost performance

Loris Nanni\textsuperscript{a,}\textsuperscript{*}, Sheryl Brahnam\textsuperscript{b}, Alessandra Lumini\textsuperscript{a}

\textsuperscript{a}Department of Electronic, Informatics and Systems (DEIS), Università di Bologna, Via Venezia 52, 47023 Cesena, Italy
\textsuperscript{b}Computer Information Systems, Missouri State University, 901 S. National, Springfield, MO 65804, USA

\textbf{A R T I C L E  I N F O}

\textbf{Keywords:}
Image analysis
Texture descriptors
Support vector machine
Local binary patterns
Local binary pattern histogram Fourier features

\textbf{A B S T R A C T}

This paper focuses on the combination of variants of local binary patterns (LBP), widely considered the state of the art among texture descriptors, using the same radius and the same number of neighborhoods. We report new experiments exploring several LBP-based descriptors and propose a set of variants for the representation of images. Our experiments are of two main types. In the first set, the Fourier transform is used to extract features starting from the histogram of uniform patterns. In these experiments we test different methods of extracting features from the histogram and each method is used to train a set of support vector machines (SVMs) which are then combined. In the second set of experiments, features are extracted from histograms using different definitions of uniform patterns. These are used to train SVMs, and the results are then combined. Our results show that descriptors extracted from LBP using the same radius and the same number of neighborhoods can be combined to improve classifier performance.

\textcopyright 2010 Elsevier Ltd. All rights reserved.

1. Introduction

Recent advances in image processing and pattern recognition methods have fostered the development of large databases of digital images. This in turn has had the circular effect of encouraging more research in image processing and pattern recognition. Medical research, in particular, has expanded with the growth of new image databases and techniques. One area of particular promise involves searching large databases of medical images of patients to identify shared sets of salient features – a task that has the potential of helping scientists better understand the genesis and progress of many diseases that are inadequately understood today. Other potential medical applications of these new image technologies include developing novel methods for detecting and diagnosing disease. One new area of research, for example, involves the application of face recognition techniques. It is well known that many diseases produce facial abnormalities and interrupt normal facial expression formation; thus face recognition technology could be utilized as an early diagnostic tool (see Brahnam, Nanni, & Randall, 2007b; Ekmann, Huang, Sejnowski, Hager, & Golomb, 1992). Recent work along this line includes that of Fang, Fang, Huang, and Tuceryan (2006) and Gunarathne and Sato (2003), who successfully developed face recognition systems to detect neurological disorders, and Brahnam, Chuang, Shih, and Slack (2005, 2007a), who used face recognition technology to detect pain (a strong indicator of disease) in neonates.

Texture-based methods, especially those using local binary patterns (LBP) (Ojala, Pietikainen, & Maenpaa, 2002), are proving particularly powerful and are now being employed in many state-of-the-art image classification systems. In the field of medicine, for instance Oliver, Lladó, Freixenet, and Martí (2007) successfully use LBP to represent significant micro-patterns in mammogram images, and Unay and Ekin (2008) introduced a powerful search and retrieval method that used LBP with support vector machines (SVMs) to find relevant slices in brain magnetic resonance volumes. Also of note is the work of Keramidas, Iakovidis, Maroulis, and Dimitropoulos (2008), who investigated the use of textural features extracted from thyroid ultrasounds, and Nanni and Lumini (2008b, 2008c), who used LPB in a system that automated cell phenotype image classification.\footnote{A fairly complete listing of biomedical work that has explored LBP in image classification systems can be found at http://www.ee.oulu.fi/mvg/page/lbp_bibliography#biomedical.}

In the field of face classification, LBP has been extensively explored in the work of Nanni and Lumini (2007) and Ahonen, Hadid, and Pietikainen (2006). Other important papers that have investigated LPB in other application areas, for example smart guns and fingerprint identification, include the work of Nanni and Lumini (2008) and Shang and Veldhuis (2007).

Most investigations of LPB have utilized uniform LPB patterns. Recently, several variants of LPB have been proposed (see, for example, Nanni, Lumini, & Brahnam, 2009). Interesting work where a few non-uniform features were concatenated in a feature vector obtained using uniform patterns is reported in Zhou, Wang, and Wang (2008).
In this paper we study two different fusion approaches using different LPB variants as the descriptors and support vector machines as the classifiers. In both methods, we combine descriptors extracted using the same radius and the same number of neighborhoods. In the first set of experiments reported in this paper, we propose some variants of what is called local binary pattern histogram Fourier (LBP-HF) features (Ahonen, Matas, He, & Pietikäinen, 2009). The fusion among the proposed variants is then studied. In the second set of experiments, different definitions of uniform patterns are used to extract different descriptors. These descriptors are then used to train a pool of classifiers, which are combined using sum rule.

The remainder of this paper is organized as follows. In Section 2 we provide an overview of LBP as a descriptor. In Section 3 we detail our proposed approaches. In Section 4, we describe the datasets used in our experiments. In Section 5, we report experimental results. Finally, in Section 7, we provide a few concluding remarks and directions for future research.

2. Descriptors based on LBP

Since the mid 1990s, LBP has been the focus of much texture-based research. Prof. Pietikainen’s research group (Ojala, Pietikäinen, & Harwood, 1996) was one of the first to explore the advantages of using LBP as a texture descriptor. His work was followed by many other investigations, including that of Nanni and Lumini (2007) and Shang and Veldhuis (2007).

LBP is a simple operator. It is calculated by computing the binary differences between the gray value of a given pixel $x$ and the gray values of its $P$ neighboring pixels on a circle of radius $R$ around $x$. The LBP operator is rotation invariant when the smallest value of $P - 1$ bitwise shift operations on the binary pattern is selected. A uniform pattern is defined as one where the number of transactions in the sequence between 0 and 1 is less than or equal to two, with the number of possible uniform patterns being $2^{P-1}$. The LBP feature vector is extracted from each cell and is the histogram of dimension $P + 2$ (a single bin for non-uniform patterns).

LBP has proven to be such a powerful local texture operator because it possesses several desirable properties (Ojala et al., 2002). Of particular importance is the fact the LBP operator is low in computational complexity and handles variations in illumination better than most other local operators.

Despite these advantages, LBP poses several significant challenges. One concerns the loss of anisotropic structural information that occurs when solving the rotation invariant problem in conventional LBP. Anisotropic information is vital in many domains, most notably face recognition. To preserve this information (Liao & Chung, 2007), proposed an elliptical neighborhood definition. Another important challenge with conventional LBP concerns its sensitivity to noise in the near-uniform image regions. To overcome this problem (Tan & Triggs, 2007), proposed using local ternary patterns (LTP). In this method the difference between a given pixel $x$ and its neighbor $u$ is encoded by 3 values, rather than two, according to a threshold $\tau$: $1$ if $u > x + \tau$; $-1$ if $u < x - \tau$; else 0. As illustrated in Fig. 1, the ternary pattern is represented as two binary patterns by considering its positive and negative components. The histograms that are computed from the binary patterns are then concatenated to form the LTP feature vector.

Other variants of a 3-valued coding scheme have been proposed. For example Ahonen and Pietikäinen (2007) and Iakovidis, Keramidas, and Maroulis (2008), apply a fuzzy thresholding function to overcome the noise sensitivity of LBP. Similarly Hafiane, Seetharaman, Palaniappan, and Zavidovique (2008) map the intensity space to LBP by thresholding a given pixel against the median value of its neighborhood.

Yet another problem with conventional LBP is the size of the extracted histogram. One method for reducing dimensionality is that proposed by Heikkinä, Matti Pietikäinen, and Schmid (2008). In this work, a given pixel is compared with center-symmetric pairs of pixels. This produces only $2^d$ binary patterns given 8 neighbors rather than the $2^d$ different binary patterns produced by normal LBP.

Other improvements of conventional LBP use preprocessing methods to enhance classification performance (Nanni & Lumini, 2008; Zhang, Shan, Chen, & Gao, 2007). Gabor wavelets are combined with the LBP operator to represent face images, for instance, in the work of Zhang et al. (2007). This method of representing faces, however, is high in dimensionality since multiple Gabor transformations must be performed. Zhang et al., attempt to overcome this problem by applying dimensionality reduction to the output of the LBP operators.

In Ahonen et al. (2009) a rotation invariant image descriptor based on uniform local binary patterns is proposed. The discrete Fourier transform is used to extract a class of features that are invariant to rotation of the input image starting from the histogram rows of the uniform patterns. Ahonen et al. (2009) call these descriptors local binary pattern histogram Fourier (LBP-HF) features.

3. Methods

Our aim is to show that it is possible to improve classification performance by combining descriptors that are extracted using the same radius and the same number of neighbors. In the first set of experiments, we propose some variants of the LBP-HF features described above (Ahonen et al., 2009). We compare these features with features that were extracted using discrete cosine transform (DCT) and Daubechies wavelets (Daubechies, 1992).

A discrete cosine transform (DCT) expresses a sequence of finitely many data points in terms of a sum of cosine functions oscillating at different frequencies. A DCT is a Fourier-related transform that is similar to the discrete Fourier transform (DFT), except that it uses real numbers.

Daubechies filters maximizes the smoothness of wavelets by maximizing the rate of decay (scaling function) of its Fourier transform. This is accomplished using a cascading algorithm. For one dimensional wavelet decomposition, the first step starts with a signal and produces two sets of coefficients: approximation coefficients (scaling coefficients) and detail coefficients (wavelet coefficients). The approximation coefficients are then split into

![Fig. 1. An example of splitting an ternary code into positive and negative LBP codes.](image-url)
two parts using the same algorithm, and these are likewise replaced by approximation coefficients and detail coefficients. This process is then repeated.

In our experiments the following approaches are tested:

- FF, the original method, where from each DFT the first half of the coefficients are retained.
- DC, where from each DCT the first half of the coefficients are retained.
- W + F, where the histogram is decomposed by Daubechies wavelet before DFT and then the method FF is performed; \(^2\)
- W + D, where the histogram is decomposed by Daubechies wavelet before DCT and then the method DC is performed;
- W + FF, the histogram is decomposed by Daubechies wavelet before DFT and then the method FF is performed, with all coefficients retained;
- W + FD, the histogram is decomposed by Daubechies wavelet before DCT and then the method DC is performed, with all coefficients retained.

Moreover, we compare results using a noise robust version of LBP (NLB), as well as standard LBP. In the noise robust version, the difference between a given pixel \(x\) and its neighbor \(y\) is 1 if \(y \geq x + 1\). In our experiments \(\tau = 3\).

The second set of experiments is based on the definition of uniform patterns. In the standard LBP method, a pattern is considered uniform if the number of transactions in the sequence between 0 and 1 is less than or equal to \(K\), where \(K = 2\). Here we change the definition of uniform patterns by changing the value of \(K\). For each different uniform pattern definition, a different classifier is trained.

In both sets of experiments, we use support vector machines (SVM) (Cristianini & Shawe-Taylor, 2000) as the classifier. SVM finds the equation of a hyperplane that maximally separates all the points between two classes. In the case where a linear decision boundary does not exist, kernel functions are used to project the data points onto a higher-dimensional feature space which can be separated by a hyperplane. Commonly used kernels include polynomial kernels and radial basis function kernels.

The features used for training SVMs are linearly normalized to \([0, 1]\). In our experiments, we have modified for our purposes the original LBP and LBP-HF code available at http://www.ee.oulu.fi/mvg/page/lbp_matlab.

### 4. Datasets

We examine the merits of using LBP variants by running extensive experiments using several benchmark medical databases: the first Infant COPE database of neonatal facial images, where the problem is to detect pain in a neonate’s facial features; the 2D HeLa dataset and Locate endogenous dataset, where the problem is to classify cell phenotype images starting from fluorescence microscope images; the Herlev University Hospital Pap smear dataset, where the problem is to classify the status of pap smear cells between normal or abnormal; and the Amino-acids database, where the problem is to classify amino-acids that either bind or do not bind multiple Human Leukocyte Antigens. In addition to the medical databases, we conduct experiments using the DaimlerChrysler Pedestrian database, where the problem is to detect images containing pedestrians; and the difficult Subjective Human Face Trait database, where the problem is to match the social impressions faces make on the average human observer by mapping a given face to a set of binary trait classes.

In the remainder of this section, we describe each of these databases and the evaluation protocols that were used in our experiments.

#### 4.1. Infant COPE database and evaluation protocols (Pain)

The infant classification of pain expressions (COPE) database and study design were first described in Brahnam et al. (2005). This database is a collection of 204 facial images of 26 neonates. The images capture the facial expressions of the infants as they are experiencing the pain of a heel lance and three nonpain stressors.

Depending on the type of stressor administered and the state of the infant, the facial images in the infant COPE database are divided into the following categories: (1) Rest, (2) Cry, (3) Air Stimulus, (4) Friction, and (5) Pain. Images were first taken of the infants in their initial state of being either peaceful (rest) or crying. The infants were then systematically subjected to the three stressors and the pain stimulus. The stressors produced facial expressions that were indicative of distress and thus similar in expression to those produced as a result of experiencing pain. The first stressor involved the bodily disturbance of being moved from one crib to another. The second stressor was a puff of air on the nose, and the third stressor was a friction stimulus, where the infants’ heels were rubbed in preparation for the state mandatory blood exam. The rubbing occurred for 10–15 s and was administered with a cotton ball soaked in 70% alcohol. The pain stimulus was the puncture on the heel with a lancet followed by repeated squeezing of the heel as the blood samples were taken.

A total of 204 photographs were taken and categorized as follows: rest (67) cry (18), air stimulus (23), friction (36), and pain (60). Fig. 2 shows two example sets of the five neonatal expressions. For complete details of the experimental design, see Brahnam et al. (2007b).

In our experiments, we label this dataset Pain, and the classification protocol used was the following: (1) the images were divided by subject; (2) the images of subject, s, formed the testing set while the remaining subjects formed the training set; (3) this procedure was then repeated for each subject. The images were resized to 100 × 120 pixels to reduce computational complexity. From each image, 64 overlapping cells of dimension 25 × 25 were created at steps of 11 pixels. A different SVM was then trained on each of these cells, and the resulting 64 classifier decisions were summed.

#### 4.2. Pap smear dataset (pap)

This dataset of 917 pap smear samples was collected at the Herlev University Hospital using a digital camera and microscope (Jantzen, Norup, Dounias, & Bjerregaard, 2005). Two skilled cytotechnicians manually classified each cell into the two classes of normal versus abnormal (see Fig. 3 for an example). In the case of a disagreement, a medical doctor examined the cells and made a final classification.

In our experiments, we label this dataset Pap. To calculate the area under the ROC-curve, a 5-fold cross validation technique

---

\(^2\) The W + F method is as follows:

\[
\begin{align*}
&k = 1; \\
&\text{for } j = 1: \text{length(map.orbits)} \\
&b = \text{inputvectors}(;, \text{map.orbits}(j)+1); \\
&\text{if } (\text{size(b, 2) > 1}) \\
&[l, c] = \text{dwt}(b, \text{‘db4’}, \text{‘mode’, ‘sym’}); \\
&b = \text{ff.toolbox}(); \\
&b = \text{abs}(b); \\
&b = b(:, 1): \text{floor}(\text{size(b, 2)/2+1}); \\
&\text{end} \\
&\text{featurevectors3(k:k+size(b, 2)-1)=b;} \\
&k = k + \text{size(b, 2)}; \\
&\text{end}
\end{align*}
\]
was employed, where each dataset was randomly divided into 4/5ths for training and 1/5th for testing.

4.3. DaimlerChrysler pedestrian dataset (Ped)

The subset extracted from the DaimlerChrysler pedestrian dataset (Munder & Gavrila, 2006)\(^3\) contains 4,900 images, some of which contain pedestrian images. This dataset is difficult to classify because the it was purposely seeded with images of non-pedestrian samples where a shape-based pedestrian detector resulted in a low confidence match. Examples of images with pedestrians (left) and non-pedestrians (right) are displayed in Fig. 4.

From the original dataset, we extracted a set of 4,900 images. We label this dataset Ped, and the experimental results were obtained using a 5-fold cross validation, where each dataset is randomly split into 4/5ths for training and 1/5th for testing.

4.4. Amino acids (AM)

The amino acids dataset contains peptides from five HLA-A2 molecules (Bozic, Zhang, & Brusic, 2005). In Table 1, we list those where the five HLA-A2 molecules either bind (B) or do not bind (NB) multiple human leukocyte antigen (HLA).

The representation of the peptide/protein is obtained by examining physicochemical properties of the amino acids available in the amino acid index database (Kawashima & Kanehisa, 2000).\(^4\) The twenty amino-acids are sorted given the value of the selected protein. A ranking value (Feng & Wang, 2008) is determined by weighing the position of the amino-acid in the sequence. The first amino-acid is given the ranking value of 1, and the last is given a value of 1/20. If there are no two amino-acids with the same values, then the sequence is determined based on the number of different values in the sequence. Two amino-acids with the same value have the same rank. As an example, if the 20 bases are sorted according to a given physicochemical property \(P\): \(N < K < R < Y < F =\)

\(^3\) The DaimlerChrysler dataset is available at http://www.science.uva.nl/research/isladc-ped-class-benchmark.html.

\(^4\) Available at www.genome.jp/dbget/aaindex.html.
Q < S < H < M < W < G = L < V < E < I < A < D < T < P < C, then the corresponding weights are: \( \text{rank}_P(N) = \frac{1}{18} \), \( \text{rank}_P(K) = \frac{2}{18} \), \( \text{rank}_P(R) = \frac{3}{18} \), …, and \( \text{rank}_P(C) = 1 \).

A matrix is formed by collecting all the pairs of amino-acids that compose the sequence of the peptide/protein into a square matrix, \( \text{OM} (P) \), having dimensions \( l \times l \), where \( l \) is the length of the sequence. For each pair of elements, \( s \) and \( t \) of the sequence, the corresponding entry \( \text{OM} (P)_{st} \) of this matrix is given by \( \frac{\text{rank}_P(s) + \text{rank}_P(t)}{2} \). The diagonal values of the matrix are \( \text{OM} (P)_{ss} = \text{rank}_P(s) \).

In our experiments, this dataset is labeled \( AM \), and results were obtained using a 5-fold cross validation (each dataset is randomly split into 4/5ths for training and 1/5th for testing) to calculate the area under the ROC-curve.

4.5. Subjective human face trait dataset (Traits)

This dataset, described in Brahnam and Nanni (2009), is used to classify faces according to the social impressions they make on the average human observer. It contains a set of 111 artificially constructed faces (using the composite software program FACES, produced by InterQuest and Micro-Intel) that were proven to exhibit strong human consensus within the bipolar extremes of the following six trait categories: intelligence, maturity, warmth, sociality, dominance, and trustworthiness. Each of the trait databases is divided into two classes, high and low. An example of a high versus low faces in the database of warmth is provided in Fig. 5.

The classifiers are trained to match the bipolar extremes of the faces in each of the six trait dimensions, with the performance measured by the Area Under the ROC curve and averaged across all six dimensions. In our experiments, we label this dataset \( Traits \), and results were obtained using a 5-fold cross validation (each dataset is randomly split into 4/5ths for training and 1/5th for testing) to calculate the area under the ROC-curve.

5. Experimental results

The performance measure adopted in the experiments reported in this paper is the area under the ROC-curve (AUC) (Fawcett, 2004). AUC is a scalar measure of classification performance. It can be interpreted as the probability that the classifier will assign a higher score to a randomly picked positive sample than to a randomly picked negative sample.

In the first set of experiments, see Tables 2 and 3, we compare the different variants of LBP-HF (note: in the \( AM \) dataset \( \tau = 0.1 \), instead of \( \tau = 3 \), as in the other datasets) with both the LBP and NLB histograms tested. We also report in Tables 2 and 3, the results obtained by:

- \( U \), standard LBP where uniform patterns (non rotation invariant) are used;
- \( FA \), all the LBP and NLB variants are combined (LBP-FF, …, LBP-U, NLB-FF, …, NLB-U);
- \( FB \), all the LBP variants are combined (LBP-FF, …, LBP-U);
- \( FT \), all the NLB variants are combined (NLB-FF, …, NLB-U).

Fig. 4. Original images (pedestrian at left, non-pedestrian at right).

Fig. 5. Samples from the two classes (high and low) of warmth. The top row faces were rated significantly higher in warmth. The bottom row faces were rated significantly lower in warmth.

<table>
<thead>
<tr>
<th>HLA-A2</th>
<th>B</th>
<th>NB</th>
</tr>
</thead>
<tbody>
<tr>
<td>0201</td>
<td>224</td>
<td>378</td>
</tr>
<tr>
<td>0202</td>
<td>619</td>
<td>2361</td>
</tr>
<tr>
<td>0204</td>
<td>641</td>
<td>2162</td>
</tr>
<tr>
<td>0205</td>
<td>648</td>
<td>2346</td>
</tr>
<tr>
<td>0206</td>
<td>621</td>
<td>2349</td>
</tr>
</tbody>
</table>

Table 1
Number of binders (B) and non-binders (NB) in training and testing sets for HLA-A2.
The row AV reports the average rank of that method evaluated for all the datasets. The rank is the position of a descriptor in relation to the other descriptors based in its classification performance.

Given the results reported in Tables 2 and 3, we can make the following conclusions:

- NLB outperforms LBP;
- FF and DC obtain a similar performance;
- In the datasets tested, U outperforms LBP-HF, except in the Pain dataset where U does not work well;
- In the Traits dataset U works very well (better than the variants of the LBP-HF), in this dataset the fusion is not useful (the performance is comparable with U);
- FT outperforms each stand-alone NLB variants, and in the same way FB outperforms each stand-alone LBP variants;
- The best method is FA, where all the descriptors are combined.

It is well-known in the literature that LTP outperforms LBP. In general, our experiments confirm this. However, we note that LTP, depending on the dataset, works as well as or worse than LBP when coupled with HF. In our opinion this is due to the problem of the curse of dimensionality: the large number of features extracted, coupled with the small number of training samples available in most of the benchmarks used, actually degraded the performance of a classifier. In future work we want to examine other classifiers using a random subspace ensemble of unstable classifiers, such as Levenberg–Marquardt neural networks, since it has been demonstrated (for example, in Nanni & Lumini, 2008b), that using a random subspace ensemble instead of a stand-alone classifier reduces the problem incurred by the curse of dimensionality.

As experimental confirmation of our hypothesis, we report the results using the full DaimlerChrysler pedestrian dataset (see Table 4). This dataset is divided into five fully disjoint sets, three for training and two for testing. Each single set contains 4800 pedestrian examples and 5000 non-pedestrian examples. For each experiment, three different training set are generated, each by selecting two out of the three training sets (thus, we have a very large training set). Testing all three combinations of the training

<table>
<thead>
<tr>
<th>Dataset</th>
<th>LBP</th>
<th>FA</th>
<th>FB</th>
</tr>
</thead>
<tbody>
<tr>
<td>FF</td>
<td>0.89</td>
<td>0.89</td>
<td>0.86</td>
</tr>
<tr>
<td>DC</td>
<td>0.86</td>
<td>0.87</td>
<td>0.87</td>
</tr>
<tr>
<td>W+F</td>
<td>0.88</td>
<td>0.88</td>
<td>0.88</td>
</tr>
<tr>
<td>W+D</td>
<td>0.88</td>
<td>0.88</td>
<td>0.88</td>
</tr>
<tr>
<td>W+FF</td>
<td>0.90</td>
<td>0.90</td>
<td>0.90</td>
</tr>
<tr>
<td>W+FD</td>
<td>0.90</td>
<td>0.90</td>
<td>0.90</td>
</tr>
<tr>
<td>U</td>
<td>0.96</td>
<td>0.96</td>
<td>0.96</td>
</tr>
<tr>
<td>Traits</td>
<td>0.78</td>
<td>0.77</td>
<td>0.73</td>
</tr>
<tr>
<td>AV</td>
<td>10.00</td>
<td>10.80</td>
<td>11.40</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Dataset</th>
<th>LBP</th>
<th>FA</th>
<th>FT</th>
</tr>
</thead>
<tbody>
<tr>
<td>FF</td>
<td>0.89</td>
<td>0.88</td>
<td>0.91</td>
</tr>
<tr>
<td>DC</td>
<td>0.91</td>
<td>0.91</td>
<td>0.91</td>
</tr>
<tr>
<td>W+F</td>
<td>0.92</td>
<td>0.92</td>
<td>0.92</td>
</tr>
<tr>
<td>W+D</td>
<td>0.91</td>
<td>0.91</td>
<td>0.91</td>
</tr>
<tr>
<td>W+FF</td>
<td>0.93</td>
<td>0.93</td>
<td>0.93</td>
</tr>
<tr>
<td>W+FD</td>
<td>0.90</td>
<td>0.90</td>
<td>0.90</td>
</tr>
<tr>
<td>U</td>
<td>0.79</td>
<td>0.79</td>
<td>0.79</td>
</tr>
<tr>
<td>Traits</td>
<td>0.77</td>
<td>0.77</td>
<td>0.77</td>
</tr>
<tr>
<td>AV</td>
<td>8.60</td>
<td>8.80</td>
<td>9.20</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Dataset</th>
<th>LTP</th>
<th>DC</th>
<th>FA</th>
<th>FT</th>
</tr>
</thead>
<tbody>
<tr>
<td>FF</td>
<td>0.85</td>
<td>0.87</td>
<td>0.88</td>
<td>0.88</td>
</tr>
<tr>
<td>LBP</td>
<td>0.85</td>
<td>0.87</td>
<td>0.88</td>
<td>0.88</td>
</tr>
<tr>
<td>FT</td>
<td>0.85</td>
<td>0.87</td>
<td>0.88</td>
<td>0.88</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Dataset</th>
<th>LBP</th>
<th>FA</th>
<th>FB</th>
</tr>
</thead>
<tbody>
<tr>
<td>FF</td>
<td>0.80</td>
<td>0.78</td>
<td>0.82</td>
</tr>
<tr>
<td>DC</td>
<td>0.84</td>
<td>0.72</td>
<td>0.53</td>
</tr>
<tr>
<td>W+F</td>
<td>0.86</td>
<td>0.75</td>
<td>0.64</td>
</tr>
<tr>
<td>W+D</td>
<td>0.82</td>
<td>0.71</td>
<td>0.51</td>
</tr>
<tr>
<td>W+FF</td>
<td>0.72</td>
<td>0.57</td>
<td>0.54</td>
</tr>
<tr>
<td>W+FD</td>
<td>10.00</td>
<td>13.20</td>
<td>15.00</td>
</tr>
</tbody>
</table>
sets on both test sets yields six different ROC areas. We used the same system proposed in Nanni and Lumini (2008d). First the images are normalized to reduce the lighting effects, then each image is divided into two sub-images: a lower and upper area. From each area we extract several feature vectors. These are used to train a pool of SVM classifiers combined by Sum Rule. The results obtained by LBP-HF and LTP-HF variants are reported in Table 4. As expected LTP outperforms LBP.

In the second set of experiments, Tables 5 and 6, we compare performance by changing the values of K in the definition of the uniform patterns (the rotation invariant version is used). The method named FUS is the fusion among the four classifiers trained using our descriptors obtained with \( K = 2, K = 4, K = 6, \) and \( K = 8. \) The classifier trained with \( K = 2 \) has a weight in the weighted sum rule of 6, while the other classifiers have a weight of 1, due to the higher performance of the standard descriptor (i.e., \( K = 2. \))

In this set of experiments, we use LTP (with \( \tau = 3, \) except in the AM dataset, where \( \tau = 0.1, \)) when standard rotation invariant features are used, and NLB, when HF variants are used (because of the problem of the curse of dimensionality).

Given the results reported in Tables 5 and 6, we make similar conclusions to those we made regarding the LBP-HF variants tests:

- Both LTP/NLB outperform LBP;
- In the Traits dataset U, when \( K > 2, \) the performance is very low, so in this dataset the fusion is not useful (performance is comparable with standard methods when \( K = 2); \)
- The fusion (FUS) method outperforms the stand-alone methods.

We think it is very interesting that combining LBP descriptors using the same geometric loci, radius, and neighborhood improves performance when compared to the stand-alone methods.

6. Conclusion

This paper compares several new variants of the well-known local binary pattern based texture descriptors. In particular, variants of the LBP-HF features are studied. In addition, experiments are performed that investigate the performance of other definitions of the uniform patterns (only rotation invariant versions are considered). Each method is used to train a different SVM classifier and its performance is then compared to different fusions among these classifiers. The main conclusion drawn is that fusion outperforms stand-alone methods. Moreover, it is interesting to note that descriptors taken from the same geometric loci using the same radius and the same number of neighborhoods results in a boost in performance with respect to the stand-alone methods.

In future studies, we plan to examine weighed combinations of descriptors based on different loci of points that are extracted from different neighborhoods (obtained by varying \( P \) and \( R \) ) in order to evaluate the potentiality of LBP-based ensembles. Another possible future work is to study the performance of the proposed texture descriptors when the feature extraction is performed from images that have been pre-processed using various methods (e.g., Gabor filters).

Acknowledgments

The authors thank T. Ahonen, T. Ojala, M. Pietikäinen and T. Mäenpää for sharing their code.

References


