Local binary patterns variants as texture descriptors for medical image analysis

Loris Nanni a,*, Alessandra Luminia a, Sheryl Brahnam b

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

1. Introduction

Recent imaging technologies have permitted researchers to collect a large number of digital images in the medical domain. For this reason, developing a reliable method for automated search and retrieval of images from medical databases is a very important research topic.

An interesting example is the patient-to-patient search, where, given an image as query, several images from different patients are compared and an automatic system retrieves relevant cases. In this way the physicians can study the most similar clinical cases and improve the quality of their diagnosis. Other interesting works in bioimaging include [1], where a classification system for thyroid ultrasound texture is presented; [2], where a system for cell phenotype image classification is proposed; [3], where potential use of face classification in medical diagnosis for pain detection is studied; [4], where abnormalities in facial configurations of adults are used to detect the presence of fetal alcohol syndrome (FAS); and [5], where a mesh-based approach is proposed that estimates the asymmetries in facial actions to determine the presence of facial motion dysfunction in patients with Bell’s palsy.

Obviously it is not possible to cite all the relevant papers on image-based techniques in medical image analysis. In this paper we focus our attention on one of the most used texture descriptors: the local binary patterns (LBP). LBP have recently proven useful in describing medical images. In [6] the LBP are used to describe images of brain magnetic resonance (MR) volumes: given a query image the system retrieves relevant slices. In [7] the LBP are used for representing salient micro-patterns in mammographic mass detection and to train a support vector...
machine (SVM) with the aim of distinguishing between the true recognized masses and the ones which actually are normal parenchyma. LBP are widely considered the state of the art among texture descriptors, and several variants have been proposed in the last few years to improve their discriminative power. In this work, we review the existing variants and propose some novel variants to LBP-based texture descriptors.

To assess the performance of the proposed systems, three medical image datasets are used:

1. Infant COPE database, where the aim is to classify pain states from neonatal images.
2. 2D-HeLa dataset, where the aim is to classify the protein localization starting from fluorescence microscope images.
3. Pap smear dataset, where the aim is to classify smear cells as either normal or abnormal.

The paper is organized as follows. In Section 2 an overview of the existing approaches based on LBP is given. In Section 3 the proposed variants are detailed. In Section 4 a description of the datasets used as case studies in this work is given, and related works for each of the three classification problems are described. In Section 5 the experimental results are reported. Finally, in Section 6, some concluding remarks and directions for future research are presented.

2. LBP and its variants

LBP is a local texture operator proposed by [8], which has a low computational complexity and a low sensitivity to changes in illumination. The LBP have been tested in several applications, for computational complexity and a low sensitivity to changes in illumination. The LBP have been tested in several applications, for example, in texture classification [8], face recognition [10], smart gun [11], fingerprint identification [12], and automated cell phenotype image classification [2].

To calculate the LBP operator the local binary difference between the gray value of a pixel x and the gray values of neighborhood of x placed on a circle of radius R is evaluated. To obtain a rotation invariant descriptor [9], P = 1 bitwise shift operations on the binary pattern are performed, and the smallest value is selected. A pattern is defined “uniform” if the number of transactions between “0” and “1” of the sequence is less or equal to two, with the number of different types of uniform patterns that can occur being P + 1. To describe a given image, the histogram of dimension P + 2 is extracted. It contains the occurrence of the P + 1 types of uniform patterns, and the number of non-uniform patterns.

The circular neighborhood definition allows to obtain a rotation invariant descriptor, but in some problems the anisotropic structural information is an important information source. To exploit this anisotropic structural information, an elliptical neighborhood definition has been used [13] for a face recognition system. This variant to the standard LBP has been named elliptical binary pattern (EBP).

Another variant has been proposed by [14] to solve the problem of the sensitivity to noise in near-uniform image regions. This method, called local ternary patterns (LTP), proposed a 3-valued coding that includes a threshold around zero for the evaluation of the local gray-scale difference. The idea of a 3-valued coding is proposed also in [15,16], where a fuzzy thresholding function is used to make the LBP operator more robust to noise. In [17] another technique, called median binary patterns, is proposed for obtaining a noise resistant texture descriptor. In this technique, the localized binary pattern is obtained by thresholding the pixels against their median value within a neighborhood. A similar idea is proposed with the improved LBP [18,19], a variant of LBP which compares the neighborhood pixels’ intensity values against the local mean pixel intensity, instead of the intensity of the central pixel, to reduce the effect of noise.

Finally, in [20] center-symmetric LBP are studied in order to obtain a shorter LBP histogram. Instead of comparing each pixel with the center pixel, only center-symmetric pairs of pixels are compared. An example is shown in Fig. 1. While the standard LBP produces 2^8 different binary patterns (considering 8 neighbors), the center-symmetric LBP produces 2^4 binary patterns.

Other interesting works in this area include [12,21], where the LBP descriptors are coupled with some preprocessing methods for improving the classification performance. For instance, in [21] the LBP descriptors are extracted from the images convolved with the Gabor filters.

3. Novel LBP variants

In this section we describe the texture descriptors proposed in this paper, which are obtained by considering different shapes for the neighborhood calculation and different encodings for the evaluation of the local gray-scale difference.

3.1. Neighborhood topology

Since in the literature the circular neighborhood definition proposed in the original version of LBP has already been replaced by an elliptical one in order to better deal with anisotropic structural information, we test the loci of points detailed in Table 1 and in Fig. 2 for the neighborhood definition.

All the loci of points (with the exception of the circle) are considered, using different main directions, by introducing a rotation angle of β.

3.2. Encoding for the evaluation of the local gray-scale difference

The original LBP descriptor uses a binary encoding to represent the difference between the gray-level of a pixel and each of its neighborhoods. Some variants have been proposed that use three-value encodings (LTP, from [14]) to reduce the sensitivity to noise in near-uniform image regions. In this work, we suggest using a

![Fig. 1. The center-symmetric LBP creates only 2^4 binary patterns.](image)

Table 1

<table>
<thead>
<tr>
<th>Loci of points</th>
<th>Equation</th>
<th>Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Circle</td>
<td>x^2 + y^2 = r^2</td>
<td>r is the circle radius</td>
</tr>
<tr>
<td>Ellipse</td>
<td>x^2 / a^2 + y^2 / b^2 = 1</td>
<td>a, b are the semimajor and semiminor axis lengths</td>
</tr>
<tr>
<td>Parabola</td>
<td>y = −(b^2 / 4c)x^2 + 2c</td>
<td>c is the distance between vertex and focus</td>
</tr>
<tr>
<td>Hyperbola</td>
<td>x^2 / a^2 − y^2 / b^2 = 1</td>
<td>a, b are the semimajor and semiminor axis lengths</td>
</tr>
<tr>
<td>Archimedean spiral</td>
<td>r = a + bθ</td>
<td>(r, θ) represent the polar coordinates; the parameter a turns the spiral, while b controls the distance between successive turnings</td>
</tr>
</tbody>
</table>

five-value encoding, named quinary, in order to obtain a more robust descriptor. In this variant the difference between the gray value of the center pixel $x$ from the gray values of one of its neighborhood $u$ can assume 5 values instead of the 2 values as in standard LBP. In our studies, all the configurations for encodings (binary, ternary and quinary) are coupled to all the neighborhood topologies. The calculation of the encoding is performed according to the following rules:

- **Binary encoding ($B$)**: the difference $d$ between $x$ and $u$ is encoded by 2 values:
  $$d = \begin{cases} 1 & u > x \\ 0 & \text{otherwise} \end{cases}$$

- **Ternary encoding ($T$)**: the difference $d$ is encoded by 3 values according to a threshold $t$:
  $$d = \begin{cases} 1 & u > x + t \\ 0 & x - t \leq u < x + t \\ -1 & \text{otherwise} \end{cases}$$

- **Quinary encoding ($Q$)**: the difference $d$ is encoded by 5 values according to two thresholds $t_1$ and $t_2$:
  $$d = \begin{cases} 2 & u \geq x + t \\ 1 & x + t_1 \leq u < x + t_2 \\ 0 & x - t_1 \geq u < x + t_1 \\ -1 & x - t_2 \leq u < x - t_1 \\ -2 & \text{otherwise} \end{cases}$$

The ternary pattern is split into two binary patterns by considering its positive and negative components, as illustrated in Fig. 3. The quinary pattern is split into four binary patterns (Fig. 4) according to the following binary function $b_c(x)$, $c \in \{-2, -1, 1, 2\}$:

$$b_c(x) = \begin{cases} 1 & x = c \\ 0 & \text{otherwise} \end{cases}$$

The first binary pattern is obtained by considering $c = 2$, the second considering $c = 1$ and so on for $c = -1$ and $c = -2$. Finally, the histograms computed from the binary patterns are concatenated.

### 4. Related works and datasets for the three application problems

#### 4.1. Pain expression detection

Pain expression detection is a promising medical application of face recognition. It has important potential diagnostic value, especially for populations, such as neonates, that are incapable of articulating their pain experiences [22]. At present, health professionals must infer pain in neonates by examining various physiological and behavioral indicators that are strongly associated with pain. Face recognition is particularly relevant in neonatal pain assessment, since research has shown that facial expressions of pain provide the most reliable and accurate source of information regarding an infant's state [23]. Even while the infant is sleeping, pain triggers discernable facial activity: eye squeeze, deepening naso-labial furrow, and a bulging of the forehead. By observing an angular opening of the mouth and a taut tongue, even neonatal cries of pain can be distinguished from other crying states.

Work on using face classification in this problem domain was first proposed in [24]. Additional research in this area includes [22,25].
Augmenting neonatal pain assessment with machine recognition systems would help professionals to maintain objectivity in their observations. Even though neonatal expressions of pain are clearly discernable, health professionals have difficulty identifying these facial signals. Observer bias is a major obstacle to proper assessment. The personality of the health care professional and desensitization to pain due to repeated exposure to patient suffering are some of the major factors involved in observer bias [26,27]. Adding machine evaluations of facial displays in neonatal pain assessment would provide a continuous source of objective monitoring.

As described in detail in [25], the initial infant classification of pain expressions (COPE) database contains 204 facial images of 26 neonates experiencing the pain of a heel lance and three nonpain stimuli: a bodily disturbance (movement from one crib to another), air stimulus applied to the nose, and friction on the heel using cotton soaked in alcohol. The air stimulus typically provoked an eye squeeze that is similar to that produced by pain. The friction and bodily disturbance stimuli produced facial displays indicative of various levels of discomfort, as well as crying expressions that were not provoked by pain. The pain stimulus was part of a state mandatory blood exam. In addition to the above, images were taken of the infants at rest. Of the 204 photographs taken, 67 are rest, 18 are cry, 23 are air stimulus, 36 are friction, and 60 are pain. Fig. 5 provides two example sets of the five neonatal expressions. Since pain detection is a two state problem, the rest, cry, air stimulus, and friction images were combined to form a class of 144 nonpain images.

The following evaluation protocol is used in the classification experiments: the images are divided by subject, the testing set contains images of a given subject and all the other subjects used for the training (this procedure is repeated for each subject). The performance measure adopted in the experiments reported in this paper is the area under the ROC-curve [28]. The area under the ROC-curve (AUC) is a scalar measure of classification performance, which 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.

4.2. Cell phenotype image classification

Cell phenotype image classification is a bioimaging problem which is concerned with finding the spatial distribution of a protein within a cell. Classifying protein subcellular patterns is important since the knowledge of the subcellular location of a protein is useful in understanding its specific function and in describing the cell behavior under different conditions [29].

Since the 1990s several research groups, most notably [29–31] and [2,32], as well as a number of independent researchers [33–35], began to research automated cell phenotype image classification starting from fluorescent microscope images. Recent work focused on training generic classifiers using image descriptors [36,37]. The more common image statics used for this problem are the following: Haralick texture measures [38], Zernike moments...
Several studies have focused on fusion approaches at either the feature or the score level. For example, in [36–39] the feature vector is formed by concatenating various descriptors of interest. In [38] such ensemble methods as Adaboost, bagging, and mixtures-of-experts proved valuable using the 2D-HeLa dataset. In [40] a multi-resolution approach is employed that fuses classifiers trained using descriptors extracted from different resolution spaces. In [2,32] fusion is performed at the score level. In [32] weak detectors trained on a large set of randomly extracted features are fused with traditional knowledge-based strong detectors, and in [2] ensembles of neural networks descriptors are trained using random subspace.

The 2D-HeLa dataset\(^2\) [40] consists of 862 single-cell images (16 bit gray scale of size 512 by 382 pixels). The description of the dataset in terms of number of classes and samples per class is reported in Table 2. Some sample images from the dataset are reported in Fig. 6. The performance measure adopted in the experiments reported in this paper is the accuracy, since this is a multi-class problem. All the experimental results have been 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 accuracy.

### 4.3. Pap smear diagnosis

The Papanikolaou test, or pap smear, was first introduced by Georgios Papanikolaou in the first half of the 20th century [41] and is routinely used today to diagnose cervical cancer. Cells scraped

---

<table>
<thead>
<tr>
<th>Table 3</th>
<th>A summary of the pap smear dataset: classes and number of samples per each class.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class</td>
<td>Pap smear dataset</td>
</tr>
<tr>
<td>Normal Superficial squamous epithelial</td>
<td>74</td>
</tr>
<tr>
<td>Normal Intermediate squamous epithelial</td>
<td>70</td>
</tr>
<tr>
<td>Normal Columnar epithelial</td>
<td>98</td>
</tr>
<tr>
<td>Total normal</td>
<td>242</td>
</tr>
<tr>
<td>Abnormal Mild squamous non-keratinizing dysplasia</td>
<td>182</td>
</tr>
<tr>
<td>Abnormal Moderate squamous non-keratinizing dysplasia</td>
<td>146</td>
</tr>
<tr>
<td>Abnormal Severe squamous non-keratinizing dysplasia</td>
<td>197</td>
</tr>
<tr>
<td>Abnormal Squamous cell carcinoma in situ intermediate</td>
<td>150</td>
</tr>
<tr>
<td>Total abnormal</td>
<td>675</td>
</tr>
<tr>
<td>Total</td>
<td>917</td>
</tr>
</tbody>
</table>

from the cervix are examined under a microscope by cyto-
technicians. In the last 50 years, this simple test has saved millions
of women’s lives. Although successful, Pap smears are not always
perfectly analyzed. Automating cell analysis could help increase
accuracy and speed up the process by reducing time spent on
routine tasks [42]. Work in this area is relatively recent. In [42] an
algorithm is proposed that detects cancer using 20 features
extracted from images of single human cells.

The database collected at the Herlev University Hospital [41] by
means of a digital camera and microscope contains 917 samples
that belong to seven different classes belonging to two super-
classes (normal/abnormal). Each sample (cell) is manually
classified by two skilled cyto-technicians, with difficult samples
inspected by a specialist. The description of the dataset in terms of
number of classes and samples per class is reported in Table 3.
Some sample images from the dataset are reported in Fig. 7. All the
experimental results have been 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 AUC, which is the
performance measure adopted for this two-class problem.

5. Experimental results

The three problems introduced in Section 4 are solved by
designing a classification method based on the LBP variants
proposed in Section 3 as feature extractor and SVM as classification
approach [43].

The feature extraction parameters setting used for the LBP
variants is reported in Table 4. The setting does not depend on the
studied problem, but only on the dimension of the considered
neighborhood. As suggested in [9] we test two values for \( P: P = 10 \)
and \( P = 18 \). The other parameters (i.e. \( \tau, \tau_1 \) and \( \tau_2 \)) have been fixed
for all the experiments without optimization, anyway empirical
tests have shown very similar performance for different values
for \( \tau, \tau_1 \) and \( \tau_2 \).

A linear SVM is used as the classifier in the 2D-HeLa dataset and
in the PAP dataset, while in the Cope dataset the radial basis
function SVM is used. No preprocessing is performed in the Cope
dataset, while in the 2D-HeLa dataset and in the PAP dataset we
perform a contrast-limited adaptive histogram equalization.3

The first group of tests is aimed at comparing the loci of points
detailed in Section 3.1 and the different encoding representations
detailed in Section 3.2. The results reported in the following
Figs. 8–10 are obtained fixing the main direction of all the loci to a
rotation angle of \( \beta = 0^\circ \) and the number of neighboring pixels to
\( P = 10 \).

In order to select the best configuration we have first selected
the best encoding, which is the quinary encoding on the three
datasets, then the best neighborhood topology, which is, on
average, the ellipse.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Parameters values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Loci of points</td>
<td>( P = 10 ) ( P = 18 )</td>
</tr>
<tr>
<td>Circle</td>
<td>( r = 1 ) ( r = 2 )</td>
</tr>
<tr>
<td>Ellipse</td>
<td>( a = 2; b = 1 ) ( a = 3; b = 2 )</td>
</tr>
<tr>
<td>Parabola</td>
<td>( c = 1 ) ( c = 2 )</td>
</tr>
<tr>
<td>Hyperbola</td>
<td>( a = 1; b = 1 ) ( a = 2; b = 2 )</td>
</tr>
<tr>
<td>Archimedean spiral</td>
<td>( r = 1 ) ( r = 2 )</td>
</tr>
<tr>
<td>Encodings</td>
<td>Ternary encoding ( \tau = 3 ) Quinary encoding ( \tau_1 = 2; \tau_2 = 5 )</td>
</tr>
</tbody>
</table>

Therefore, in the second group of tests we compare the
performance varying the number of neighboring pixels \( \{ P \in \{ 10, 18 \} \} \) and the rotation angle \( \{ \beta \in \{ 0^\circ, 45^\circ, 90^\circ, 135^\circ \} \} \) only for the
EQP variant. In Figs. 11–13 the performance of EQP on the three
datasets are reported. As can be observed in these figures,
increasing the value of \( P \) strongly improves performance.

Finally, the third group of tests is aimed at comparing the EQP
variant with the best LBP variants reported in the literature and to
show that the combination of descriptors created with different
parameter settings can be useful to create stronger descriptors. In
particular, in Table 5 the following methods are evaluated4:

- EQP, the EQP variant with \( P = 18 \) and the most performing value
  for \( \beta \).

---

3 adaptihisteq.m function from MATLAB 7.4.

4 All the LBP variants here proposed are based on modifications of the “official”
LBP code shared by T. Ojala, M. Pietikäinen and T. Maenpää. LPQ is implemented as the matlab code shared by the authors [44].
EQP, a variant of the EQP18 approach obtained concatenating the features extracted at four different rotation angles $\beta = \{0^\circ, 45^\circ, 90^\circ, 135^\circ\}$.

LBP, the standard (rotation invariant) LBP descriptor [9] with $P = 18$ and $R = 2$.

EBP, the elliptical binary pattern variant [13] with $P = 18$ and the most performing value for $\beta$.

LTP, the local ternary pattern variant [14] with $P = 18$.

LPQ, the descriptor recently proposed in [44].

ILBP, the improved local binary pattern [18,19] with $P = 18$.

CSLBP, the center-symmetric local binary pattern [20] with $P = 18$.

ENS, the combination by sum rule among five classifiers, where each is trained by quinary encoded features extracted using one of the five neighborhood topologies. Each feature vector is obtained setting $P = 18$ and obtained concatenating the features extracted at different rotation angles $\beta = \{0^\circ, 45^\circ, 90^\circ, 135^\circ\}$.

The aim of this table is not an unfair comparison among single descriptors (LBP, LTP) and ensembles (EQP, ENS, ...). We proposed some combinations of descriptors in order to point out that the use of some LBP variants (EQP in this case, but not only this one) can be very useful for obtaining a good descriptor that works very well in different datasets (face in COPE, sub-cellular location in 2D-HeLa, and tissue in PAP).

The results reported in Table 5 should be interpreted as follows:

- EQP, one of the best variants proposed in this work, performs, on average, better than the other LBP variants presented in the literature (with the same parameter optimization, i.e. the choice of $\beta$ when applicable).
- The performance of EQP is only marginally better than the performance of EQP, but it is obtained without the optimization of the parameter $\beta$.
- ENS, differently from the other approaches, is a combination of classifiers trained by descriptors created from different loci. The comparison among ENS and the other methods in Table 5 is not fair, but the good performance obtained by ENS suggests the possibility of considering different neighborhood configurations to create perturbations useful in the creation of ensembles.
- Particularly interesting are the results on the 2D-HeLa dataset, where the proposed texture descriptor obtained the highest performance among the single texture descriptors tested on the 2D-HeLa dataset: in [45] several descriptors were compared and the best performance obtained by a single feature set was lower than 90%.
- Even fixing the parameters $P$ and $R$, thus extracting information from the same neighborhood, it is possible to obtain different sets of LBP-based features which can give different performance.
- The advantage of EQP with respect to other variants, which makes this choice very good for practitioners, is the good performance on all the three datasets, which are very different from each other.

Finally, in order to compare the computational load of the different LBP variants, a comparison among the computation time...
References


