A simple method for improving local binary patterns by considering non-uniform patterns

Loris Nanni [a,*], Sheryl Brahnam [b], Alessandra Lumini [c]

[a] Department of Information Engineering, Università di Padova, Via Gradenigo, 6 - Padova, Italy
[b] Computer Information Systems, Missouri State University, 901S. National, Springfield, MO 65804, USA
[c] DEIS, Università di Bologna, Via Venezia 52, 47521 Cesena, Italy

ABSTRACT

The basic idea behind LBP is that an image is composed of micropatterns. A histogram of these micropatterns contains information about the local features in an image. These micropatterns can be divided into two types: uniform and non-uniform. In standard applications using LBP, only the uniform patterns are used. The non-uniform patterns are considered in only a single bin of the histogram that is used to extract features in the classification stage. Non-uniform patterns have undesirable characteristics: they are of a high dimension, partially correlated, and introduce unwanted noise. To offset these disadvantages, we explore using random subspace, well-known to work well with noise and correlated features, to train features based also on non-uniform patterns. We find that a stand-alone support vector machine performs best with the uniform patterns and random subspace with histograms of 50 bins performs best with the non-uniform patterns. Superior results are obtained when the two are combined. Based on extensive experiments conducted in several domains using several benchmark databases, it is our conclusion that non-uniform patterns improve classifier performance.

1. Introduction

Because of rapid advances in imaging technology, large databases of digital images are being assembled in many fields resulting in new classification problems. This is particularly the case in the field of medicine. One promising area of new research is finding methods for automatically extracting relevant medical images of multiple patients sharing salient features. This capability would prove extremely helpful to experts unraveling the causes and progress of many diseases that are currently poorly understood. Another new area of research involves the potential use of face classification in medical diagnosis [8]. For example, [19,26] have demonstrated the value of building machine systems that detect abnormalities reflective of disease in patient facial structures and expressions. Since a number of environmental and genetic syndromes are known to produce facial feature abnormalities and many neurological disorders are associated with anomalous facial expression formation, face recognition systems applied to medical diagnosis will become an exciting new area of future research [18].

Many methods currently being devised to search image databases and classify faces use texture-based descriptors, with Local Binary Patterns (LBP), first proposed by [53], widely considered the state of the art among texture descriptors. LBP has several attractive properties: it has proven to be a powerful discriminator, is low in computational complexity, and is less sensitive to changes in illumination than many descriptors. Since LBP is very resistant to lighting changes, it makes a good choice for encoding fine details.

Since the 1990s, LBP texture descriptors have been the focus of considerable research and have demonstrated their superiority as texture descriptors in several comparative studies [2], Heikkilä and Pietikäinen [68], [53]. Some examples of research using LBP in medicine include the work of [63], who introduced a novel search and retrieval method for finding relevant slices in brain MR (magnetic resonance) volumes. In the field of mammographic mass detection, [55] use LBP for representing significant micropatterns. Once the LBP are extracted, Support Vector Machines (SVMs) classify true masses from normal parenchyma. In [41], textural features extracted from thyroid ultrasounds are investigated, and a number of researchers have explored using LBP in automated cell phenotype image classification (see, for example, [51]). In the area of face recognition, LBP has been explored, for example, in the work of [49] and [2]. Other interesting areas of research using LBP include work on smart guns [59] and fingerprint identification [50]. A large repository of papers that explore LBP can be found at http://www.ee.oulu.fi/mvg/page/lbp_bibliography#biomedical.

The basic idea behind LBP is that an image is composed of micropatterns. LBP is the first-order circular derivative of patterns

* Corresponding author.
E-mail addresses: loris.nanni@unipd.it (L. Nanni),
sbrahnam@missouristate.edu (S. Brahnam),
alessandra.lumini@unibo.it (A. Lumini).

that is generated by concatenating the binary gradient directions. A histogram of these micropatterns contains information about the distribution of edges and other local features in an image. As discussed in detail in Section 2, there are two types of LBP patterns: uniform and non-uniform patterns. Normally only the uniform patterns are used for extracting features since the feature vector extracted from non-uniform patterns introduces noise and possesses high dimensionality. One of the few works that we are aware of that deeply explores non-uniform patterns is that of [67], who incorporated a few non-uniform patterns in a feature vector obtained using mostly uniform patterns. Combining uniform patterns with a few non-uniform patterns was shown in [67] to improve performance. Non-uniform patterns thus appear to contain useful information.

Other interesting works are [44,27]. In [44] rotation invariant patterns are selected, instead of the uniform patterns. They propose choosing patterns that represents 80% of the patterns in the training data. In [27] the hierarchical multiscale LBP is presented. It is an approach that improves performance by extracting information from the non-uniform bins. The hierarchical multiscale LBP is based on a multiresolution approach that utilizes three different radii: (1) The LBPs for biggest radius are extracted first, then (2) for the "nonuniform" patterns, the counterpart LBPs of smaller radius is extracted, finally, (3) among the new LBPs, the "non-uniform" patterns and "uniform" patterns are extracted using an even smaller radius. This procedure is iterated until the smallest radius is used to extract features.

The aim of our research is to develop a simple and practical method for tackling non-uniform patterns. After several experiments conducted on diverse datasets, we have concluded that the problems associated with non-uniform patterns can be partially solved using a random subspace (RS) classifier [31]. RS is good at handling noise and correlation, and the dimensionality on non-uniform patterns can be reduced by extracting histograms using fewer bins. In our experiments, the best results were obtained by combining a stand-alone SVM classifier, trained using standard uniform patterns, with a RS classifier trained using the non-uniform patterns.

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 approach using non-uniform patterns. In Section 4, we describe the several datasets used in our experiments. In Section 5, we report experimental results. Finally, in Section 6, we provide a few concluding remarks and directions for future research.

2. Background on LBP

Given an image $\mathbf{I}$ of size $n \times m$ grayscale pixels and we denote with $\mathbf{I}(q)$ the gray level of the qth pixel of the image $\mathbf{I}$, the LBP operator is calculated at each pixel by evaluating the binary differences of the values of a small circular neighborhood (with radius R) around the value of a central pixel q_c (see Fig. 1).

Formally, the LBP operator is defined as follows:

$$LBP(P,R) = \sum_{p=0}^{P-1} s(I(q_p) - I(q_c))$$

where P is the number of pixels in the neighborhood, q_p is a pixel of the neighborhood of q_c, R is the radius, and $s(x) = 1$ if $x \geq 0$, otherwise 0.

The LBP operator is made rotation invariant by selecting the smallest value of $P-1$ bitwise shift operations on the binary pattern. A pattern is considered uniform if the number of transactions in the sequence between 0 and 1 is less than or equal to two. It was observed by Ojala et al. [53] that certain patterns seem to be fundamental properties of texture, providing the vast majority of patterns, sometimes over 90%. These patterns are called "uniform" because they have one thing in common: at most two one-to-zero or zero-to-one transitions in the circular binary code. Non uniform patterns are considered the patterns that contains the main part of the noise of the images. Notice that a region with no transitions is a background or a flat region of the image.

The number of possible uniform patterns is $P+1$. The feature vector extracted from each cell is the LBP histogram of dimension $P+2$, with the extra bin containing all the non-uniform patterns. We have used, in the experimental section, the most used parameters setting reported the literature: ($P=16$; $R=2$) and ($P=8$; $R=1$).

A drawback in using the circularly symmetric neighborhood to solve the rotation invariant problem in conventional LBP is the loss of anisotropic structural information. This information is important in many problems, for example, in face recognition. For this reason, [43] proposed an elliptical neighborhood definition that preserves anisotropic structural information.

Another problem with conventional LBP is that it is sensitive to noise in the near-uniform image regions. This can be overcome by using Local Ternary Patterns (LTP) as proposed by [62].

In LTP the difference between a central pixel q_c and its neighbor q_p is encoded by three values according to a threshold τ and the following function $t(\mathbf{I}, q_p, q_c, \tau)$: 1 if $\mathbf{I}(q_p) \geq \mathbf{I}(q_c) + \tau$; -1 if $\mathbf{I}(q_p) \leq \mathbf{I}(q_c) - \tau$; else 0. The ternary pattern is then split into two binary patterns by considering its positive and negative components, as illustrated in Fig. 2, according to the following binary function $b_c(x)$, $c \in \{-1, 1\}$:

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

Finally, the histograms that are computed from the binary patterns are concatenated to form the feature vector.

There are other variants of a three-valued coding scheme. Ahonen and Pietikäinen [1,35] apply a fuzzy thresholding function
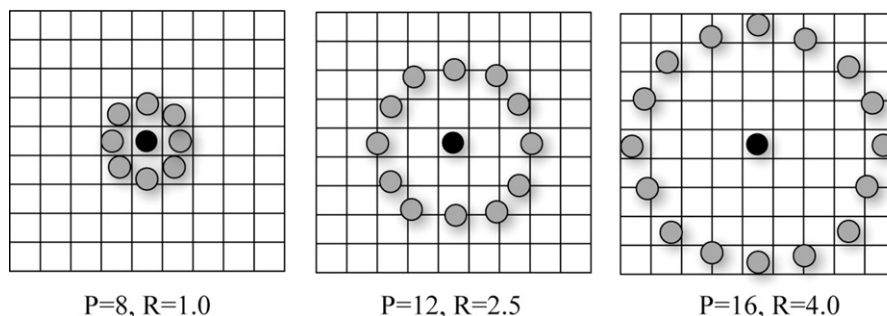

P=8, R=1.0   P=12, R=2.5   P=16, R=4.0

Fig. 1. Three circularly symmetric neighborhood sets (gray) around a central pixel (black). If a point does not fall exactly on a pixel grid, the value is interpolated.
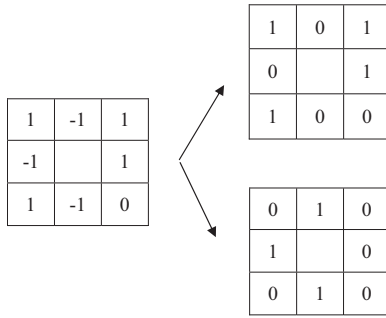
Fig. 2. An example of splitting a ternary code into positive and negative LBP codes.

center-symmetric LBP compares:
1. n_0 with n_4
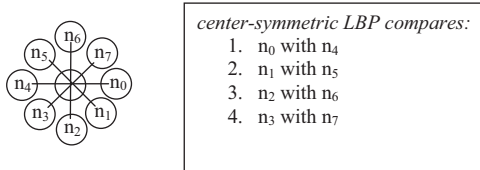2. n_1 with n_5
3. n_2 with n_6
4. n_3 with n_7

Fig. 3. The center-symmetric LBP creates only 2^4 binary patterns.

to make LBP more noise resistant. Using median binary patterns is another method for making LBP more robust in terms of noise. In [28], for example, the intensity space is mapped to LBP by thresholding a given pixel against the median value of its neighborhood.

Another problem with LBP is the rather large size of the extracted histogram. Using center-symmetric local binary patterns, [30] managed to reduce the LBP histogram dimension by comparing a given pixel with center-symmetric pairs of pixels. Instead of comparing each pixel with the center pixel, only center-symmetric pairs of pixels are compared. An example is shown in Fig. 3. Given 8 neighbors, center-symmetric LBP produces only 2^4 binary patterns in contrast to the 2^8 different binary patterns produced by normal LBP.

Finally, some interesting work that combines LBP descriptors with preprocessing methods to improve classification performance includes the work of [50,66]. For instance, in [66] Gabor wavelets are combined with the LBP operator to represent face images. Unfortunately, this method of representing faces is high in dimensionality because of the multiple Gabor transformations that are performed. Zhang et al. [66] overcome this problem to some degree by applying dimensionality reduction to the output of the LBP operators.

3. Proposed approach

As noted in the introduction, [67] demonstrated that using non-uniform patterns increases system performance. Despite these benefits, uniform patterns are almost exclusively used for extracting features. The primary reason non-uniform features are avoided is because they produce a feature vector of high dimensionality. Non-uniform patterns are also highly correlated.

After conducting a number of experiments using non-uniform patterns, we conclude that the problems of high dimensionality and noise can be partially solved by using the following: (1) a random subspace classifier instead of a stand-alone method and (2) histograms using fewer bins. The Random Subspace (RS) Method [31] modifies the training dataset by generating K (where $K = 100$ in this paper) new training sets containing only a random subset of k features (here $k = 50\%$ of the number of original features). Classifiers are then trained on these modified training sets then are combined by a given decision rule. We have chosen

$K = 100$ since in several papers (e.g., Ho [31]) is shown that RS obtains the best performance with $K > 50$.

The random subspace ensemble method is a three step process, below is a step by step outline of the random subspace approach:

1. Given a d-dimensional data set $D = \{(x_j, t_j) \,|\, 1 \le j \le m\}, x_j \in R^d$; $t_j \in C = \{1, \ldots, c\}$ where m is the number of training patterns, x_j is the jth training pattern, t_j is the label of the jth pattern and C is the set of classes, K new projected k-dimensional data sets $D_i = \{(P_i(x_j), t_j) \,|\, 1 \le j \le m\}$ are generated $(1 \le i \le K)$, where P_i is a random projection. P_i is obtained by random selecting, through the uniform probability distribution, a subset of k features from the whole pool of d features;
2. Each new data set D_i is used to train a given classifier h_i, $1 \le i \le K$;
3. The final classifier h is obtained by aggregating (i.e., combining) the base classifiers $h_1, \ldots, h_K$ through a given decision rule (here we have used the mean rule ([42])).

The classifier we use in our experiments is the support vector machine (SVM) [64,15]. SVM is a two-class prediction method that finds the equation of a hyperplane that separates all the points in the two classes while simultaneously maximizing the distance between the hyperplane and the two classes. In those cases where a linear decision boundary does not exist, a kernel function can be used to project the data onto a higher-dimensional feature space that can be separated by a hyperplane. Some typical kernels used in SVM include polynomial kernels and radial basis function kernels. It should be noted that the features used for training SVM are linearly normalized to [0 1]. We have tested (LibSVM toolbox): linear SVM; radial basis function (rbf) SVM; polynomial SVM. The parameters of each method are tuned in each dataset by a grid search using an internal 5-fold cross validation in the training set.

As reported in Section 5, we found that when only the uniform LBP patterns are used, the best performance is obtained using a stand-alone SVM classifier. The best classification results, however, are obtained by combining the stand-alone SVM classifier trained with standard uniform LBP patterns with a random subspace ensemble trained also with the non-uniform patterns. In our experiments, we have modified the original LBP code found at http://www.ee.oulu.fi/mvg/page/lbp_matlab.[1]

Hence our goal is to enhance performance obtained by an SVM trained with the rotation invariant uniform patterns by extracting a set of features from all the rotation invariant patterns. Below is a step by step outline of our approach:

Step 0: the rotation invariant uniform patterns (named SET-A) and all the rotation invariant patterns are extracted. From the rotation invariant patterns a histogram using fewer bins is extracted (named SET-B).

Step 1: A support vector machine is trained and tested using the features of SET-A obtaining the set of class similarity named SCORE-A.

Step 2: A random subset of 50% features are then selected from the SET-B.

Step 3: A support vector machine is trained and tested using the features selected in Step 2.

Step 4: Steps 2 and 3 are performed 100 times.

Step 5: The 100 classifier results are then combined using the mean rule obtaining the set of class similarity named SCORE-B.

---

[1] we have extracted the uniform patterns with *getmapping*011(16,'riu2') while the non-uniform pattern are extracted by *getmapping*011(16,'ri').

The mean rule selects as final score the mean of the scores of the pool of the classifiers that belong to the ensemble.

Step 6: The final score is given by the sum of SCORE-A and SCORE-B.

We want to stress that usually SVM is not suited for building an ensemble since it is a "strong" classifier, so perturbations of training set (e.g., random subspace or bagging) are not suited, several times but not always, for building an ensemble of SVMs. In this work we create the ensemble using different feature sets extracted from the images, since the two feature set are slightly uncorrelated they can be used for building an ensemble of SVMs, other examples of ensemble of SVMs built using different feature sets are ([36,61]).

4. Datasets

4.1. Infant COPE database[2] and evaluation protocols

First described in [7], the Infant COPE (Classification Of Pain Expressions) database, is a collection of 204 facial photographs of 26 neonates that were taken while they were experiencing the pain of a heel lance and three nonpain stressors.

Based on the type of stressor and the state of the infant, the facial images in the infant COPE database are classified into the following categories: (1) Rest, (2) Cry, (3) Air Stimulus, (4) Friction, and (5) Pain. The method for collecting images is the following. First, images were taken of the infant's initial state. These were divided into the two categories of rest and cry. The infants were then subjected to three stressors and a pain stimulus. The objective of the stressors was to produce facial expressions that were similar to those of pain. The infants were moved from their crib to a crib where the camera was set up. This move offered a stressor that often resulted in additional crying images and facial expressions that indicated an adverse reaction to the disturbance. After a timed interval, the infants were then administered the second stressor: a puff of air on the nose. The intention of the air puff stimulus was to provoke an eye squeeze that is similar to the eye squeeze found in neonatal facial expressions of pain. For the third stressor, the infants' heels were rubbed for 10 to 15 s with a cotton ball soaked in 70% alcohol. This friction stressor resulted in additional crying and disturbance reactions. The pain category contains images of the infant's initial reactions to the pain stimulus: the puncture on the heel of a lance followed by repeated squeezing of the heel as blood samples were taken for a state mandatory blood exam.

Of the 204 photographs taken, 67 are rest, 18 are cry, 23 are air stimulus, 36 are friction, and 60 are pain. For complete details of the experimental design, see [8]. The following evaluation protocol is used in the Infant COPE classification experiments. The images are divided by subject. The images of a given subject, s, form the testing set while the remaining subjects form the training set. This procedure is repeated for each subject. In order to reduce the computational requirements of the approach, the faces in the original images of size 3008×2000 pixels are extracted from the rest of the background and resized to 91×113 pixels. From each image, 64 overlapping cells of dimension 25×25 are created at steps of 11 pixels (see Fig. 4). A different classifier is trained on each of these cells, and the 64 classifier decisions are then combined.

The performance measure adopted in the experiments reported in this paper is the area under the ROC-curve [20].

Fig. 4. Original image (left), rotated, cropped and scaled image (right) with two overlapping sub-windows obtained by a single vertical and horizontal shifting of 11 pixels.

Table 1
A summary of the 2D HeLa dataset: classes and number of samples per each class.

Class	2D HeLa dataset
Actinfilaments	98
Endsome	84
ER	86
Golgi giantin	87
Golgi GPP130	85
Lysosome	91
Microtubules	91
Mitochondria	73
Nucleolus	80
Nucleus	87
Total	862

4.2. 2D HeLa dataset[3]

The 2D HeLa dataset is composed by 862 single-cell images (16 bit greyscale of size 512 by 382 pixels) [9].[4] In Table 1, we report the class labels and number of samples per class. The protocol used in our experiments was a 5-fold cross validation technique, the dataset was randomly divided, five times, into 80% for training and 20% for testing and then the average performance is reported.

4.3. Pap smear dataset

The pap smear database of 917 samples was collected at the Herlev University Hospital by means of a digital camera and microscope [37]. Skilled cyto-technicians and doctors manually classified each cell into one of two classes (Normal vs Abnormal). Each cell was examined by two cyto-technicians. The medical doctor examined cells that were difficult to classify.

To calculate the area under the ROC-curve, a 5-fold cross validation technique was employed, the dataset was randomly divided, five times, into 80% for training and 20% for testing and then the average performance is reported.

4.4. DaimlerChrysler pedestrian dataset

The DaimlerChrysler pedestrian dataset ([48][5]) has proven to be a difficult dataset to classify. This is because the non-pedestrian samples include a number of images 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. 5.
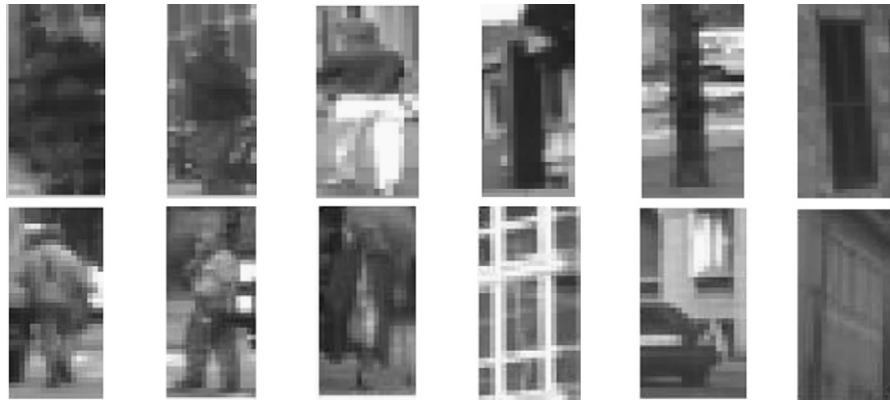

Fig. 5. Original images (Pedestrian at left, Non-Pedestrian at right).

Table 2
A summary of the LOCATE dataset: classes and number of samples per each class.

Class	Locate endogenous dataset
Actin-cytoskeleton	50
Cytoplasm	0
Endosomes	49
ER	50
Golgi	46
Lysosomes	50
Microtubule	50
Mitochondria	50
Nucleus	50
Peroxisomes	57
PM	50
Total	**502**

Table 3
Number of Binders (B) and Non-Binders (NB) in training and testing sets for HLA-A2.

HLA-A2	B	NB
0201	224	378
0202	619	2361
0204	641	2162
0205	648	2346
0206	621	2349

From the original dataset, we have extracted a set of 4900 images. The experimental results were obtained using a 5-fold cross validation, the dataset was randomly divided, five times, into 80% for training and 20% for testing and then the average performance is reported.

4.5. LOCATE mouse protein sub-cellular localization endogenous database

The LOCATE mouse protein sub-cellular localization endogenous database contains approximately 50 images, with each image containing somewhere between 1 and 13 cells per class ([21].[6] The description of the dataset in terms of number of classes and samples per class is reported in Table 2. The experimental results were obtained using a 5-fold cross validation protocol, the dataset was randomly divided, five times, into 80% for training and 20% for testing and then the average performance is reported.

4.6. Amino acids

The amino acids dataset contains peptides from five HLA-A2 molecules [6]. As detailed in Table 3, the five HLA-A2 molecules either bind (B) or non-bind (NB) multiple Human Leukocyte Antigen (HLA).

The matrix representation of the peptide/protein is obtained by considering a selected physicochemical property of the amino-acids,

which can be obtained by the amino acid index database ([40].[7] The physicochemical properties of amino-acids are critical factors that affect the protein's function/structure. First, the 20 amino-acids that make up a protein sequence are sorted according to the amino acid index value of the selected protein. Then a ranking value, which weighs the position of the amino-acid in the sequence, is assigned to each [22]. The ranking rule is the following: the first amino-acid (that is the one with the highest value) is given the value 1, the last amino-acid (that is the one with the lowest value) is given the value 1/20, that is, assuming no two amino-acids have the same value. Two amino-acids with the same value have the same rank. Thus, for example, if the 20 bases of a protein have two pairs of amino acids with the same value and the amino acids are sorted according to the following given physicochemical property P: $N < K < R < Y < F = Q < S < H < M < W < G = L < V < E < I < A < D < T < P < C$, then the corresponding weights are: $\mathrm{rank}_P(N) = 1/18$, $\mathrm{rank}_P(K) = 2/18$, $\mathrm{rank}_P(R) = 3/18$, …, and $\mathrm{rank}_P(C) = 1$.

The ordering relationships between all the pairs of amino-acids that compose the sequence of the peptide/protein are collected into a square matrix, named $OM(P)$ 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 $OM(P)_{s,t}$ of this matrix is given by $[\mathrm{rank}_P(s) + \mathrm{rank}_P(t)]/2$. Therefore, the diagonal values of the matrix are $OM(P)_{s,s} = \mathrm{rank}_P(s)$.

The our experiments, results have been obtained using a 5-fold cross validation (the dataset was randomly divided, five times, into 80% for training and 20% for testing and then the average performance is reported) to calculate the area under the ROC-curve.

4.7. Weizmann dataset

In this paper we use the 10-class Weizmann dataset ([25].[8] This dataset has become a popular benchmark for the task of

---

[6] The LOCATE mouse protein sub-cellular localization endogenous database is available at http://locate.imb.uq.edu.au/.

[7] Available at www.genome.jp/dbget/aaindex.html.
[8] Available at http://www.wisdom.weizmann.ac.il/~vision/SpaceTimeActions.html.

Fig. 6. Some samples of the 10 action classes in the Weizmann dataset.
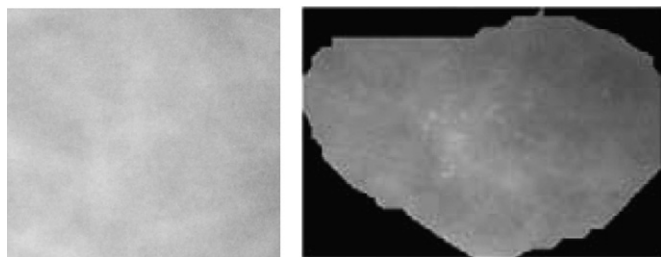

Fig. 7. Examples from the breast cancer dataset (left, normal tissue; right, abnormal tissue).

matching images portraying human actions to a set of class labels. The 10 actions in the 10-class Weizmann dataset are illustrated in Fig. 6. The actions are performed by 9 actors.

In our experiments, the descriptors are extracted from the mask images, also included in the dataset, using a background subtraction algorithm (e.g., http://maven.smith.edu/~nhowe/research/code/) [32]. Once a mask image is extracted from the original image, the image is divided into 49 subregions. For each subregion, a feature vector is extracted. These 49 feature vectors are then concatenated to represent a given frame. Experimental results were obtained using a 5-fold cross validation protocol, the dataset was randomly divided, five times, into 80% for training and 20% for testing and then the average performance is reported.

4.8. Breast cancer

Breast cancer is the major cause of cancer-related deaths among adult women. It is known that the best prevention method is precocious diagnosis. For our experiments in this domain, we used the Digital Database for Screening Mammography (DDSM), a publicly available database of digitized screen-film mammograms. This database has two classes: benign and malignant tissues. We selected the same 273 malignant and 311 benign images used in [39]. It is very important to note that in this dataset the images have very different dimensions, so we had to normalize the histograms extracted from each image to obtain a good

performance. Example images from the Breast dataset are shown in Fig. 7. The our experiments, results have been obtained using a 5-fold cross validation (the dataset was randomly divided, five times, into 80% for training and 20% for testing and then the average performance is reported) to calculate the area under the ROC-curve.

5. Experimental results

In Table 4, we report our experimental results using the datasets and performance indicators described in Section 4. Notice that in the 2-classes problems we have used the area under the ROC curve (AUC) since it has been shown [58] that AUC is empirically and theoretically better than accuracy, due to the fact that accuracy does not consider the scores of the classifiers. The area under the ROC curve is a scalar measure to evaluate 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.

In each table the best results in each datasets are bolded, if there are more bold values in each dataset means that there is overlap among the values of "mean ± standard deviation" of the best approaches. In the first set of experiments, we used LBP with $P=16$ and $R=2$. Histograms with 100 bins were used for extracting non-uniform features. As noted in Section 3, the classifier used in our experiments is SVM. In Table 4, a stand-alone SVM is represented as SA, and random subspace (with each subspace containing 50% of the original features) is represented as RS. The sum rule was used with RS. As seen in Table 4, we ran experiments using SA and RS with both the uniform and non-uniform feature textures.

Under the column heading *concatenation*, we report experiments using a feature vector obtained by concatenating both the uniform and the non-uniform features. Under the heading sum rule, we report fusion experiments that combine the two methods (SA and RS) using the sum rule. FUS1 is the sum rule combining uniform+SA with non-uniform+RS. FUS2 is the weighted sum rule combining uniform+SA and non-uniform+RS, where the weight of uniform+SA is 2 and the weight of non-uniform+RS is 1.

Table 4
Performance obtained by LBP with $P=16$ and $R=2$.

| | Uniform | | Non-uniform | | Concatenation | | Sum rule | | | |
	SA	RS	SA	RS	SA	RS	SA	RS	FUS1	FUS2
2D-Hela	0.827	0.776	0.712	0.821	0.767	0.854	0.801	0.853	0.855	**0.871**
PAP	0.749	0.730	0.799	**0.808**	**0.800**	**0.807**	0.793	0.780	0.789	0.778
Infant COPE	0.849	**0.874**	0.765	0.796	0.842	0.860	0.849	**0.868**	0.853	0.853
Pedestrian	0.835	**0.868**	0.739	0.784	0.821	**0.860**	0.847	**0.866**	0.855	0.850
Locate endogenous	0.833	0.774	0.776	0.823	0.809	**0.870**	0.855	0.829	0.853	0.847
Amino-acids	0.637	0.651	0.639	0.641	**0.686**	**0.689**	0.676	**0.671**	0.667	0.663
Weizmann	0.877	0.888	0.833	0.855	0.877	0.877	**0.911**	0.900	0.900	0.900
Breast	**0.894**	0.879	0.869	0.864	**0.900**	**0.905**	0.900	0.881	**0.897**	**0.899**
Average	0.813	0.805	0.766	0.799	0.813	**0.840**	0.829	0.831	0.833	0.833

Table 5
Performance obtained by LTP with $P=16$ and $R=2$.

| | Uniform | | Non-uniform | | Concatenation | | Sum rule | | | |
	SA	RS	SA	RS	SA	RS	SA	RS	FUS1	FUS2
2D-Hela	**0.920**	0.909	0.730	0.827	0.767	0.864	0.864	0.881	0.891	**0.920**
PAP	0.829	0.814	0.818	0.831	0.829	**0.844**	**0.837**	**0.830**	**0.837**	**0.837**
Infant COPE	0.925	0.925	0.844	0.878	0.847	0.898	0.922	0.922	0.922	0.925
Pedestrian	0.918	**0.951**	0.844	0.871	0.886	0.918	0.928	**0.951**	0.931	0.927
Locate endogenous	**0.913**	0.907	0.772	0.833	0.851	0.900	0.894	0.896	**0.906**	**0.916**
Amino-acids	0.741	0.757	0.766	0.774	**0.805**	**0.816**	0.789	0.796	0.790	0.777
Weizmann	**0.977**	**0.977**	0.911	0.911	0.933	0.933	0.933	0.933	0.933	0.944
Breast	**0.951**	0.945	0.881	0.909	0.912	0.935	**0.953**	0.946	**0.956**	**0.959**
Average	**0.897**	**0.898**	0.821	0.854	0.854	0.888	0.890	0.894	**0.896**	**0.901**

Table 6
Performance obtained varying the number of bins for the non-uniform patterns.

	$h=25$	$h=50$	$h=100$	Standard LTP	Standard LBP
2D-Hela	0.911	**0.923 ± 0.004**	**0.920**	**0.920 ± 0.005**	0.827 ± 0.010
PAP	0.822	**0.861 ± 0.008**	0.837	0.829 ± 0.011	0.749 ± 0.010
Infant COPE	0.928	**0.928 ± 0.007**	0.925	**0.925 ± 0.007**	0.849 ± 0.009
Pedestrian	0.930	**0.936 ± 0.003**	0.927	0.918 ± 0.006	0.835 ± 0.009
Locate endogenous	0.904	**0.922 ± 0.004**	0.916	0.913 ± 0.007	0.833 ± 0.010
Amino-acids	**0.773**	0.766 ± 0.009	**0.777**	0.741 ± 0.012	0.637 ± 0.011
Weizmann	0.944	**0.977 ± 0.002**	0.944	**0.977 ± 0.002**	0.877 ± 0.008
Breast	0.949	**0.960 ± 0.002**	**0.959**	0.951 ± 0.004	0.894 ± 0.007
Average	0.895	**0.909 ± 0.005**	0.901	0.897 ± 0.007	0.813 ± 0.009

SA under this heading (Sum rule) is the fusion between uniform-SA and nonuniform-SA, while RS is the fusion between uniform-RS and nonuniform-RS.

A similar set of experimental results is reported in Table 5 for LTP with $P=16$ and $R=2$. Histograms with 100 bins were used for extracting the non-uniform features. The threshold τ used in LTP is 3 for all the datasets where the elements of the image belong to [0,255], while in the amino-acids/Action dataset τ was set to 0.1 since in these datasets the elements of the matrix are [0,1].

Given the results reported in Tables 4 and 5, we can make the following conclusions. First, LTP clearly outperforms LBP. In all the tested approaches LTP outperforms LBP, LTP obtains an average performance of slightly lower than 0.9 while LBP obtains an average performance of 0.8. Second, RS with uniform patterns does not improve performance with respect to SA; however, RS with the non-uniform patterns does. We suspect that this behavior is due to the undesirable characteristics mentioned above regarding non-uniform patterns: they are correlated, high-dimensional and introduce noise. Finally, we observe that the best fusion method is FUS2, which combines both uniform and non-uniform patterns. It should be noted that the uniform

patterns are considered more important since a weight of 2 is given to them in FUS2.

In Table 6 we conduct more refined experiments using FUS2, where h, the number of bins used for extracting the non-uniform features, is varied. It is clear that the best results are obtained with $h=50$. In some datasets the performance difference between FUS2 with $h=50$ and a standard LTP is only marginal, but in other datasets FUS2 with $h=50$ proves superior. We can conclude, therefore, that in all seven datasets the performance of FUS2 with $h=50$ is equal to or better than that obtained by LTP.

In Table 7 we compare the performance of standard LBP, standard LTP, and FUS2[9] with $h=50$ using $P=8$ and $R=1$. With these values for P and R, FUS2 outperforms LTP on all seven datasets. As a result of these tests, we conclude that non-uniform patterns combined with uniform patterns results in a very superior texture descriptor.

As final tests we compare the different approaches here tested by Wilcoxon Signed-Rank test, which is considered in [17] the

---

[9] It is based on LTP feature extractor.

Table 7
Performance obtained by LBP and LTP with $P=8$ and $R=1$.

	Standard LBP	Standard LTP	FUS2 with $h=50$
2D-Hela	0.758 ± 0.014	0.838 ± 0.012	0.844 ± 0.011
PAP	0.715 ± 0.014	0.779 ± 0.014	0.785 ± 0.013
Infant COPE	0.876 ± 0.011	0.883 ± 0.011	0.885 ± 0.010
Pedestrian	0.869 ± 0.010	0.945 ± 0.006	0.959 ± 0.005
Locate endogenous	0.670 ± 0.013	0.686 ± 0.012	0.814 ± 0.010
Amino-acids	0.604 ± 0.014	0.714 ± 0.012	0.749 ± 0.011
Weizmann	0.944 ± 0.004	0.944 ± 0.003	0.977 ± 0.002
Breast	0.888 ± 0.007	0.912 ± 0.004	0.921 ± 0.004
Average	$0.791 \pm \pm 0.011$	0.838 ± 0.009	0.867 ± 0.008

Table 8
SVM parameters.

	Kernel	Parameters C: Cost of the constrain violation Gamma: parameters of the radial based kernel
2D-Hela	Linear	$C=100$
PAP	Linear	$C=10$
Infant COPE	Rbf	$C=0.5$ Gamma$=100$
Pedestrian	Rbf	$C=1$ Gamma$=1000$
Locate endogenous	Linear	$C=100$
Amino-acids	Rbf	$C=1$ Gamma$=10$
Weizmann	Linear	$C=100$
Breast	Linear	$C=10$

best statistical measure to compare classifiers, to demonstrate our thesis. We have compared the following approaches using the Wilcoxon Signed-Rank test:

- LTP versus LBP, LTP wins (i.e., we reject the null hypothesis, level of significance 0.05, and accept that the two methods have significant different performance) against LBP (both with $P=8$ $R=1$ and with $P=16$ $R=2$);
- FUS2[9] (with $h=50$) versus LTP, FUS2 (with $h=50$) wins against LTP (both with $P=8$ $R=1$ and with $P=16$ $R=2$).

From the analysis of the experimental results it is clear the advantage of the proposed approach with respect to the considered baseline methods evaluated in this work.

Finally, in Table 8, we report the SVM kernel and the SVM parameters used in each dataset.

6. Conclusion

This paper focused on a new texture descriptor, based on non-uniform patterns in Local Binary Pattern and Local Ternary Patterns, for extracting features from a matrix. A new effective approach for feature extraction is designed that is based on the fusion of classifiers trained considering the uniform patterns and classifiers trained considering only the non-uniform patterns. The experimental results show that the proposed approach produces a reliable set of features for training a machine learning classifier (the support vector machines used in this paper). The feature extractor proposed and compared in this work has been tested on a broad spectrum of datasets: the Infant COPE database of neonatal facial images; the 2D HeLa dataset and the Locate endogenous dataset of fluorescence microscope images; the Pap smear dataset of smear cells images; the Pedestrian dataset of pedestrian images; the amino-acids dataset of amino-acids that bind/non-bind multiple Human Leukocyte Antigen; the breast cancer dataset of normal and abnormal breast tissue; the

Weizmann dataset of human action video. In each of these datasets, our proposed fusion method outperformed standard LBP and LTP.

A 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 the different methods (e.g., Gabor filters).

References

[1] T. Ahonen, M. Pietikäinen, Soft histograms for local binary patterns. In: Processing Finnish Signal Processing Symposium (FINSIG 2007), Oulu, Finland, 2007.
[2] T. Ahonen, A. Hadid, M. Pietikainen, Face description with local binary patterns: application to face recognition, IEEE TPAMI 28 (12) (2006) 2037–2041.
[6] I. Bozic, G.L. Zhang, V. Brusic, Predictive vaccinology: optimization of predictions using support vector machine classifiers, Intelligent Data Engineering and Automated Learning 2005, pp. 375–381 2005.
[7] S. Brahnam, C.-F. Chuang, F. Shih, M. Slack, SVM classification of neonatal facial image of pain, In: The Proceedings of the 6th International Workshop on Fuzzy Login and Applications (WILF 2005), University of Milan, Crema, Italy 2005.
[8] S. Brahnam, L. Nanni, S. Randall, Neonatal facial pain detection using NNSOA and LSVM, In: Proceedings of the The 2008 International Conference on Image Processing, Computer Vision, and Pattern Recognition (IPCV'08), Las Vegas, Nevada, 2008 2008.
[9] A. Chebira, Y. Barbotin, C. Jackson, T. Merryman, G. Srinivasa, R.F. Murphy, J. Kovačević, A multiresolution approach to automated classification of protein subcellular location images, BMC Bioinformatics 8 (2007) 210.
[15] N. Cristianini, J. Shawe-Taylor, An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods, Cambridge University Press, 2000.
[17] J. Demsar, Statistical comparisons of classifiers over multiple data sets, Journal of Machine Learning Research 7 (2006) 1–30.
[18] P. Ekman, T.S. Huang, T.J. Sejnowski, J.C. Hager, B. Golomb, Final Report to NSF of the Planning Workshop on Facial Expression Understanding, Human Interaction Lab, University of California, San Francisco, CA, 1992.
[19] J. Fang, S. Fang, J. Huang, Tuceryan M. Digital geometry image analysis for medical diagnosis, presented at Proceedings of the 2006 ACM Symposium on Applied Computing, 2006.
[20] T. Fawcett, ROC Graphs: Notes and Practical Considerations for Researchers, Technical Report, HP Laboratories, Palo Alto, USA, 2004.
[21] J.L. Fink, R.N. Aturaliya, M.J. Davis, F. Zhang, K. Hanson, M.S. Teasdale, R.D. Teasdale, LOCATE: A Protein Subcellular Localization Database. Nucleic Acids Research, 34, 2006 (database issue).
[22] J. Feng, T.-M. Wang, Characterization of protein primary sequences based on partial ordering, Journal of Theoretical Biology (2008)http://dx.doi.org/10.1016/j.jtbi.2008.07.007.
[25] L. Gorelick, Blank Moshe, E. Shechtman, et al., Actions as space-time shapes, IEEE Transactions on Pattern Analysis and Machine Intelligence 29 (12) (2007) 2247–2253.
[26] P. Gunaratne, Y. Sato, Estimation of asymmetry in facial actions for the analysis of motion dysfunction due to paralysis, International Journal of Image and Graphics 3 (2003) 639–652.
[27] Z. Guo, L. Zhang, D. Zhang, X.Q. Mou, Hierarchical multiscale lbp for face and palmprint recognition, ICIP (2010). 2010.
[28] A. Hafiane, G. Seetharaman, K. Palaniappan, B. Zavidovique, Rotationally invariant hashing of median binary patterns for texture classification, ICIAR 2008 (2008) 619–629.
[30] M. Heikkilä, M. Matti Pietikäinen, C. Schmid, Description of interest regions with local binary patterns, Pattern Recognition (2009) 425–436.
[31] T.K. Ho, The random subspace method for constructing decision forests, IEEE Transactions on Pattern Analysis and Machine Intelligence 20 (8) (1998) 832–844.
[32] N.R. Howe, A. Deschamps, Better Foreground Segmentation Through Graph Cuts, arXiv:cs/0401017, 2004.
[35] D.K. Iakovidis, E. Keramidas, D. Maroulis, Fuzzy local binary patterns for ultrasound texture characterization, Proceedings of the Image Analysis and Recognition, 5th International Conference (ICIAR 2008), Lecture Notes in Computer Science 5112, Springer, pp. 750–759, 2008.
[36] Y. Fu, Y. Xian, Image classification based on multi-feature combination and PCA-RBaggSVM, IEEE International Conference on Progress in Informatics and Computing (PIC), 2011.

[37] J. Jantzen, J. Norup, G. Dounias, B. Bjerregaard, Pap-smear benchmark data for pattern classification. In: Proceedings of the NiSIS 2005, Nature inspired Smart Information Systems (NiSIS), EU co-ordination action, Albufeira, Portugal, 2005 (pp. 1–9).

[39] G.B. Junior, A. Cardoso de Paiva, A.C. Silva, A.C. Muniz de Oliveira, Classification of breast tissues using moran's index and geary's coefficient as texture signatures and svm, Computers in Biology and Medicine 39 (12) (2009) 1063–1072.

[40] S. Kawashima, M. Kanehisa, AAindex: amino acid index database, Nucleic Acids Research 20 (1) (2000) 374.

[41] E.G. Keramidas, D.K. Iakovidis, D. Maroulis, N. Dimitropoulos, Thyroid texture representation via noise resistant image features, Proceedings of the 21st IEEE International Symposium on Computer-Based Medical Systems (CBMS 2008), 2008, 560–565.

[42] J. Kittler, M. Hatef, R. Duin, J. Matas, On combining classifiers, IEEE Transactions on Pattern Analysis and Machine Intelligence 20 (3) (1998) 226–239.

[43] S. Liao, A.C.S. Chung, Face recognition by using Elongated local binary patterns with average maximum distance gradient magnitude, Computer Vision — ACCV 2007 (2007) 672–679.

[44] S. Liao, M.W.K. Law, A.C.S. Chung, Dominant local binary patterns for texture classification, IEEE Transactions on Image Processing 18 (5) (2009) 1107–1118. May, 2009.

[48] S. Munder, D.M. Gavrila, An experimental study on pedestrian classification, IEEE Transactions on Pattern Analysis and Machine Intelligence 28 (11) (2006) 1863–1868.

[49] L. Nanni, A. Lumini, Region Boost learning for 2D+3D based face recognition, Pattern Recognition Letters 28 (15) (2007) 2063–2070.

[50] L. Nanni, A. Lumini, Local binary patterns for a hybrid fingerprint matcher, Pattern Recognition (11) (2008) 3461–3466.

[51] L. Nanni, A. Lumini, A reliable method for cell phenotype image classification, Artificial Intelligence in Medicine 43 (2) (2008) 87–97.

[53] T. Ojala, M. Pietikainen, T. Maeenpaa, Multiresolution gray-scale and rotation invariant texture classification with local binary patterns, IEEE Transactions on Pattern Analysis and Machine Intelligence 24 (7) (2002) 971–987.

[55] A. Oliver, X. Lladó, J. Freixenet, J. Martí, False positive reduction in mammographic mass detection using local binary patterns, In: Proceedings of the Medical Image Computing and Computer-Assisted Intervention (MICCAI 2007), Brisbane, Australia, Oct 29–Nov 2, Lecture Notes in Computer Science 4791, Springer, Vol I, 2007, pp. 286–293.

[58] Z.C. Qin, ROC analysis for predictions made by probabilistic classifiers, Proceedings of the 4th International Conference on Machine Learning and Cybernetics 5 (2006) 3119–3312.

[59] X. Shang, R. Veldhuis, Local absolute binary patterns as image preprocessing for grip-pattern recognition in smart gun, In: Proceedings of the 1st IEEE International Conference on Biometrics: Theory, Applications, and Systems, 2007 pp. 1–6.

[61] M. Tahir, Khan, A Majid, Protein subcellular localization of fluorescence imagery using spatial and transform domain features, Bioinformatics 28 (1) (2012) 91–97.

[62] X. Tan, B. Triggs, Enhanced local texture feature sets for face recognition under difficult lighting conditions, In: Analysis and Modelling of Faces and Gestures, vol 4778 of LNCS, pp 168–182. Springer, 2007.

[63] D. Unay, A. Ekin, Intensity versus texture for medical image search and retrieval Proceedings of the 5th IEEE International Symposium on Biomedical Imaging: From Nano to Macro (ISBI 2008), 2008, 241–244.

[64] V.N. Vapnik, The Nature of Statistical Learning Theory, Springer-Verlag, New York, 1995.

[66] W. Zhang, H. Shan, X. Chen, W. Gao, Local gabor binary patterns based on mutual information for face recognition, International Journal of Image and Graphics 4 (2007) 777–793.

[67] H. Zhou, R. Wang, C. Wang, A novel extended local binary pattern operator for texture analysis, Information Sciences 22 (2008) 4314–4325.

[68] M. Heikkilä, M. Pietikäinen, A texture-based method for modeling the background and detecting moving objects, IEEE Transactions on. Pattern Analysis and Machine Intelligence 28 (2) (2006) 657–662.

Loris Nanni received his Master Degree cum laude on June-2002 from the University of Bologna, and the April 26th 2006 he received his Ph.D. in Computer Engineering at DEIS, University of Bologna. His research interests include pattern recognition, bioinformatics, and biometric systems (fingerprint classification and recognition, signature verification, face recognition).

Alessandra Lumini received her Master Degree from the University of Bologna, Italy, on March 26th 1996. In 1998 she started her Ph.D. studies at DEIS- University of Bologna and in 2001 she received her Ph.D. degree in Computer Engineering for her work on "Image Databases". Now she is an Associate Researcher at University of Bologna. She is interested in pattern recognition, bioinformatics, Biometric Systems, Multidimensional Data Structures, Digital Image Watermarking and Image Generation.

Sheryl Brahnam received her Ph. D. (Computer Science) from the Graduate Center at the City University of New York (2002). Now she is an Associate Professor at Missouri State University. She is interested in pattern recognition, face recognition, medical image analysis.