Multimed Tools Appl
DOI 10.1007/s11042-016-4174-8

Texture classification based on curvelet transform and extreme learning machine with reduced feature set

Sanae Berraho[1] · Samira El Margae[1] · Mounir Ait Kerroum[1] · Youssef Fakhri[1]

Received: 5 January 2016 / Revised: 11 November 2016 / Accepted: 18 November 2016

Abstract In this work, a novel approach for texture classification is proposed. We present a highly discriminative and simple descriptor to achieve feature learning and classification simultaneously for texture classification. The proposed method introduces the application of digital curvelet transform and explores feature reduction properties of locality sensitive discriminant analysis (LSDA) in conjunction with extreme learning machine (ELM) classifier. The image is mapped to the curvelet space. However, the curse of dimensionality problem arises when using the curvelet coefficients directly and therefore a reduction method is required. LSDA is used to reduce the data dimensionality to generate relevant features. These reduced features are used as the input to ELM classifier to analytically learn an optimal model. In contrast to traditional methods, the proposed method learns the features by the network itself and can be applied to more general applications. Extensive experiments conducted in two different domains using two benchmark databases, illustrate the effectiveness of the proposed method. In addition, empirical comparisons of the proposed method against curvelet transform in conjunction with traditional dimensionality reduction tools show that the suggested method does not only lead to a more reduced feature set, but it also outperforms all the compared methods in terms of accuracy.

Keywords Texture classification · Digital curvelet transform · LSDA · Extreme learning machine · Dimension reduction

These authors contributed equally

✉ Sanae Berraho
berraho.sanae@gmail.com

✉ Samira El Margae
elmargaesamira@gmail.com

[1] LaRIT Laboratory, team Network, Telecommunications and Artificial intelligence, Faculty of Science, University Ibn Tofail, BP 242, Kenitra, Morocco

Published online: 08 December 2016

1 Introduction

Image analysis is an important field of computer vision responsible for extracting information from images. Texture classification is a very important field of image analysis. For this reason, the literature has been presented methods that provide more and more highly discriminative texture descriptors. Surely texture is a fundamental part of visual features and recently texture classification has received considerable attention, because it can be applied to many vision-related applications such as object recognition, remote sensing, content-based image retrieval, and biological and medical applications. However, it is still very challenging and a difficult task for designing a high accuracy and robust texture classification system for real world applications due to different complexities such as the wide range of natural texture types, the large intra-class variations in texture images, different lighting conditions, positions, variations in scale and rotations caused by arbitrary viewing and illumination conditions, the demands of few computations [36].

The multiresolution analysis has proved to be useful in textural feature extraction. In these methods the texture image is transformed by the use of the respective transform to local spatial/frequency representation by wrapping the image with appropriate band-pass filters tuned to specific parameters. Most commonly employed multiresolution analysis techniques include the Gabor transform [26] and the wavelet transform [32, 50]. Although wavelet transform has obvious advantages in dealing with point singularities, it fails to efficiently represent objects with highly anisotropic elements, such as lines and curvilinear structure. Motivated by the need of representation the lines, curves and edges, Candes and Donoho developed curvelet multi-resolution analysis in 2000 [11, 12]. Curvelet transform is not only a multi-scale transform, but also directionally sensitive and highly anisotropic, which are more efficiently than those traditional transforms in representing singularities and edges along curves.

Curvelet transform is gaining popularity in different research areas, like signal processing, image analysis. It has been successfully applied in image denoising [48], image compression, image fusion [14], contrast enhancement [47], image deconvolution [47], high quality image restoration [47], astronomical image representation [48] etc. Examples of two applications, contrast enhancement and denoising. Readers are suggested to go through the referred works for further information on various applications of the curvelet transform. Recently, curvelets have also been employed to address several pattern recognition problems, such as face recognition [40, 53], optical character recognition [38], finger-vein pattern recognition [52] and palmprint recognition [18].

An important challenge in many problems of machine learning, computer vision and pattern recognition is to extract a small number of good features. This leads to avoid the "curse of dimensionality" problem [34] and therefore enhance the performance of classification in term of computational coast. A common way to overcome this problem is to use dimensionality reduction techniques. Two of the most popular techniques for this purpose are principal component analysis (PCA) [35] and linear discriminant analysis (LDA) [2]. PCA is an unsupervised method, it aims to project the data along the direction of maximal variance. LDA is supervised, it searches for the project axes on which the data points of different classes are far from each other while requiring data points of the same class to be close to each other. Both of them are spectral methods, i.e., methods based on eigenvalue decomposition of either the covariance matrix for PCA or the scatter matrices (within-class scatter matrix and betweenclass scatter matrix) for LDA. Intrinsically, these methods try to estimate the global statistics, i.e. mean and covariance [7]. However, in some real-world applications such as breast cancer diagnosis, the number of samples is usually much smaller than the

dimensionality. Then the covariance matrix of each class may not be accurately estimated. In this case, the generalization capability on testing samples can not be guaranteed. Also, these methods effectively see only a linear manifold that based on the euclidean structure. They fail to capture the structure which lies on a nonlinear submanifold hidden in the image space. Therefore, some manifold learning algorithms have attracted more attention, they aim at preserving the local nonlinear structure of the data, examples of these algorithms being isometric feature mapping (Isomap) [49], locally linear embedding (LLE) [44], and locality preserving projections (LPP) [43]. These algorithms are all unsupervised and neglect the class label information, which will impair the recognition accuracy [13]. More recently, Cai et al. proposed an efficient supervised manifold learning algorithm called locality sensitive discriminant analysis (LSDA) [7]. By discovering the local manifold structure, LSDA finds a projection which maximizes the margin between data points from different classes at each local area. Particularly, the data points are mapped into a subspace in which the nearby points with the same label are close to each other while the nearby points with different labels are far apart. Several researches have been performed in pattern recognition based on LSDA [7, 23, 51].

The ingrained limitations of existing classification algorithms include large sensitivity to view point variations, number of prototypes and slow classification speed [42]. In this work, we propose a new feature extraction approach for texture classification. We deal with the problem of extracting a reduced set of discriminative features from curvelet transform. The objective is to investigate the application of curvelet on images in conjunction with LSDA and ELM classifier and to compare it with the application of some commonly used dimension reduction methods. Our objective is to find an optimal representation of texture that perform well in several domains. Extensive experiments were carried out on two databases of different domains: digital database for screening mammography (DDSM) [27] and German traffic sign recognition benshmark database [46].

The remainder of this paper is organized as follows. Section 2 describes the proposed method. Section 3 contains experimental results and discussion while Section 4 concludes the article.

2 The proposed approach

This paper deals with the problem of presenting images with a reduced and discriminative feature vector. The proposed method consists of the following steps: feature extraction, dimensionality reduction and classification (Fig. 1).

2.1 Feature extraction

Curvelet transform was developed by Candes and Donoho in 1999. Its development was motivated by the need of image analysis [8]. The transform represents an image at different

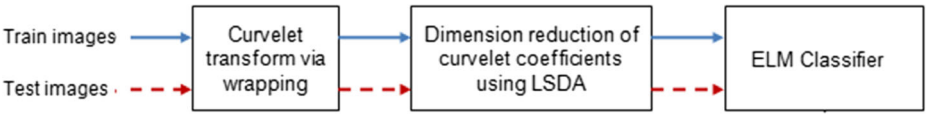

Fig. 1 The basic model of the proposed algorithm

scales and angles [8]. It provides an efficient representation of smooth objects with discontinuities along curves. As an extension of the wavelet concept, the curvelet transform was not only successfully applied in several image processing fields but also outperformed wavelet-based methods [15, 37]. Curvelet will be superior over wavelet in following cases [12]:

– Optimally sparse representation of objects with edges.
– Optimal image reconstruction in severely ill-posed problems.
– Optimal sparse representation of wave propagators.

The obvious disadvantage of the first generation of curvelet transform is its rather slow performance. In the past few years curvelet construction has been redesigned in order to make it simpler to understand and use. This second generation curvelet transform [12] introduced in 2006, compared to the first generation curvelet transform, has simpler concept, faster speed and lower degree of redundance [10]. Curvelet transform is multiscale and multidirectional. Curvelets exhibit highly anisotropic shape obeying parabolic-scaling relationship (they take the shape of elongated needles at finer scales).

In order to implement curvelet transform, first a 2D-FFT of the image is taken. Then the 2D fourier frequency plane is divided into parabolic wedges. Finally an inverse FFT of each wedge is taken to find the curvelet coefficients at each scale j and angle l. Figure 2a shows the division of wedges of the Fourier frequency plane. The wedges are the result of partitioning the fourier plane in radial (concentric circles) and angular divisions. Concentric circles are responsible for decomposition of the image in multiple scales (used for bandpassing the image) and angular divisions correspond to different angles or orientation. Hence, to address a particular wedge one needs to define the scale and angle first. In the spatial domain, each wedge corresponds to a particular curvelet at that given scale and angle. Figure 2b represents curvelets in spatial cartesian grid associated with a given scale and angle [12].

To perform fast discrete curvelet transform, two implementations were proposed [12]. These implementations differ in the way curvelet at a given scale and angle are translated with respect to each other. The first one is based on unequally spaced fast fourier transforms (USFFT). The second one is based on the wrapping of specially selected fourier samples.

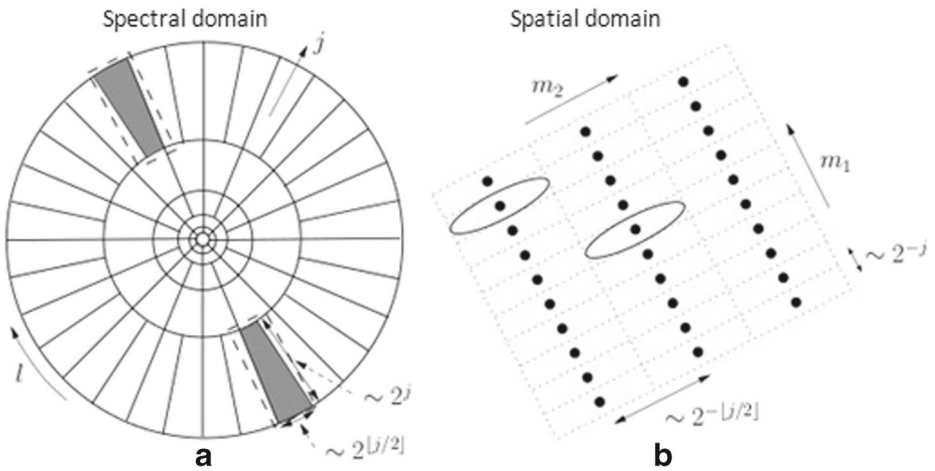

Fig. 2 Curvelets in Fourier frequency (**a**) and spatial domain (**b**)

Curvelet transform via wrapping is chosen for this paper as it is the fastest curvelet transform currently available.

Let's take the original $n \times n$ image as $f(n_1, n_2)$, where n_1, n_2 are the rectangular coordinates. $\hat{f}[n_1, n_2]$ denote its 2D-FFT. The implementation steps of 2D curvelet transform are described as follows [12]:

1. Apply 2D-FFT and obtain fourier samples $\hat{f}[n_1, n_2]$, where $n/2 \leqslant n_1, n_2 \leqslant n/2$.
2. For each scale j and angle l, calculate the product $\widetilde{U}_{j,l}[n_1, n_2]\hat{f}[n_1, n_2]$
3. Wrap this product around the origin and obtain $\hat{f}[n_1, n_2] = W(\widetilde{U}_{j,l}\hat{f})[n_1, n_2]$, where n_1 and n_2 are in the range of $0 \leqslant n_1 \leqslant L_{1,j}$ and $0 \leqslant n_2 \leqslant L_{2,j}$ separately.
4. Apply the inverse 2D-FFT to each $\tilde{f}_{j,l}$, and obtain the discrete coefficients (Fig. 3).

2.2 Dimensionality reduction

Locality sensitive discriminant analysis [7] characterizes the local geometrical structure of the data manifold by building a nearest neighbor graph. However, in order to improve the discriminative ability of the low-dimensional features, the class label information is incorporated into the feature extraction process. The goal of LSDA [7] is that the similar data points of original space need more closely in low-dimensional space, and after the dimensionality reduction, the distance of different types of data points in the low-dimensional space should be as large as possible.

Given N data points $X = \{x_1, x_2, ..., x_N\} \subset R^{D \times N}$ in a D dimensional space. These data points belong to C class, each class contains n_c, $c = 1, 2, ..., C$ samples. The first step of LSDA is to construct a nearest neighbor graph G by putting an edge between each sample and its k nearest neighbors. Let $N(x_i) = \{x_i^1, ..., x_k^1$, be the set of k nearest neighbors of x_i, the weight matrix of G in LSDA is given by:

$$W_{i,j} = \begin{cases} 1 \text{ if } x_i \in N(x_j) \text{ or } x_j \in N(x_i), \\ 0, \text{ otherwise} \end{cases} \tag{1}$$

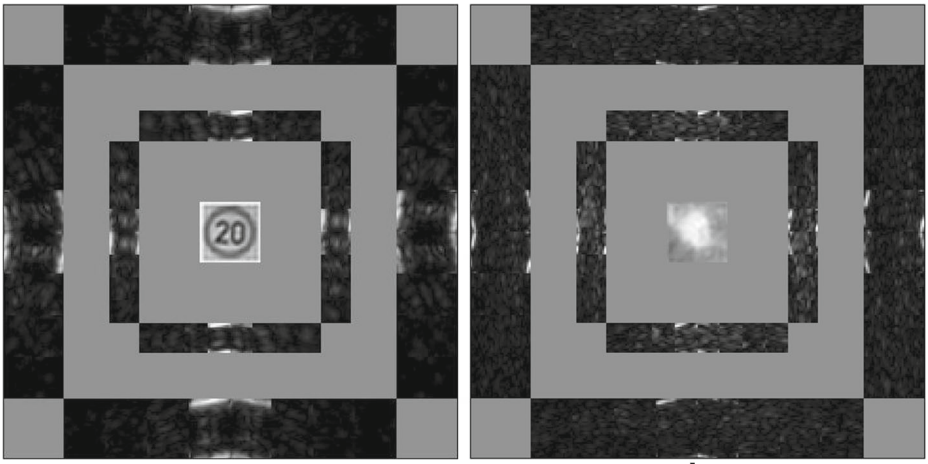

Fig. 3 Curvelet coefficients: (**a**) Traffic sign image (**b**) Mammogram image

The nearest neighbor graph G with weight matrix W characterizes the local geometry of the data manifold. Then, for the sake of discovering the local discriminant structures, the nearest neighbor graph is partitioned into two parts: a within-class graph (G_w) and a between-class graph (G_b). For each sample $x_i (i = 1, ..., N)$, the set of its k nearest neighbors can be split into two subset $N_w(x_i)$ and $N_b(x_i)$. $N_w(x_i)$ is the set of neighbors sharing the same label with x_i, and $N_b(x_i)$ is the set of neighbors with different labels from x_i. The definitions of $N_w(x_i)$ and $N_b(x_i)$ are as follows:

$$N_w(x_i) = \{x_i^j | l(x_i^j) = l(x_i), 1 \leqslant j \leqslant k\} \tag{2}$$

$$N_b(x_i) = \{x_i^j | l(x_i^j) \neq l(x_i), 1 \leqslant j \leqslant k\} \tag{3}$$

where $l(x_i)$ is the class label of x_i. Clearly, $N_w(x_i) \cap N_b(x_i) = \emptyset$ and $N_w(x_i) \cup N_b(x_i) = N(x_i)$.

In LSDA, the adjacent weight matrices of G_b and G_w, i.e. W_b and W_w, are defined as:

$$W_{b,ij} = \begin{cases} 1 \text{ if } x_i \in N_b(x_j) \text{ or } x_j \in N_b(x_i), \\ 0, \text{ otherwise} \end{cases} \tag{4}$$

$$W_{w,ij} = \begin{cases} 1 \text{ if } x_i \in N_w(x_j) \text{ or } x_j \in N_w(x_i), \\ 0, \text{ otherwise} \end{cases} \tag{5}$$

It is clear to see $W = W_b + W_w$ and the nearest neighbor graph G can be thought of as a combination of within-class graph G_w and between-class graph G_b.

The objective of LSDA is to obtain a low-dimensional feature space in which the nearby points with the same label are close to each other while the nearby points with different labels are far apart (Fig. 4). Thus, a reasonable criterion is to optimize the following two functions:

$$min \frac{1}{2} \sum_{ij} (y_i - y_j)^2 W_{w,ij} \tag{6}$$

$$max \frac{1}{2} \sum_{ij} (y_i - y_j)^2 W_{b,ij} \tag{7}$$

where $y_i \in R^{d \times 1} d << D$ is the feature extraction result of x_i. Suppose the low dimensional features of the input data can be obtained by a transformation matrix A, that is, $y_i = A^T x_i$, $i = 1, 2, ..., N$. After a series of deductions, the objective function of LSDA can be derived as:

$$\begin{aligned} &\arg\max A^T X(\beta L_b + (1-\beta) W_w) X^T A \\ &s.t. A^T X D_w X^T A = 1 \end{aligned} \tag{8}$$

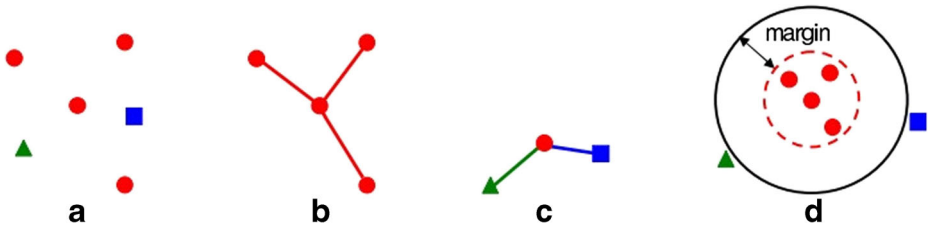

Fig. 4 **a** The center point has five neighbors. The points with the same color and shape belong to the same class. **b** The within-class graph connects nearby points with the same label. **c** The between-class graph connects nearby points with different labels. **d** After Locality Sensitive Discriminant Analysis, the margin between different classes is maximized [7]

where D_w and D_b are diagonal matrices whose entries are column (or row, since W_w and W_b are symmetric) sums of W_w and W_b, that is, $D_{w,ii} = \sum_j W_{w,ii}$ and $D_{b,ii} = \sum_j W_{b,ii}$. $L_b = D_b - W_b$ is the Laplacian matrix of G_b and β is a regulative parameter with $0 \leqslant \beta \leqslant 1$. Finally, the transformation matrix A can be obtained by maximizing the generalized eigenvalues problem:

$$A^T X(\beta L_b + (1-\beta)W_w)X^T A = \lambda X D_w X^T A \tag{9}$$

2.3 Classification

Different pattern classifiers have been applied for object recognition, including After dimensionality reduction, we used Extreme learning machine (ELM) classifier for classification. ELM as a new learning algorithm for single layer feedforward neural networks (SLFNs) as shown in Fig. 5, was first introduced by Huang el al. [29, 30] ELM seeks to overcome the challenging issues faced with the traditional SLFNs learning algorithms such as slow learning speed, trivial parameter tuning and poor generalization capability. ELM has demonstrated great potential in handling classification and regression tasks with excellent generalization performance. The learning speed of ELM is much faster than conventional gradient based iterative learning algorithms of SLFNs like back propagation algorithm while obtaining better generalization performance. ELM has several significant features [30] which distinguish itself from the traditional learning algorithms of SLFNs:

- ELM is easily and effectively used by avoiding tedious and time-consuming parameter tuning.
- ELM has extremely fast learning speed.
- ELM has much better generalization performance than the gradient based iterative learning algorithms in most cases.
- ELM is much simpler and without being involved in local minima and over-fitting.
- ELM can be used to train SLFNs with many non-differentiable activation functions.

A standard ELM classifier, whose M hidden nodes use infinitely differentiable activation functions, could approximate arbitrary samples with zero error [29], which means given a training data $\vartheta = \{(x_i, t_i)\}_{i=1}^N$, the output function of the single hidden layer feedforward neural network (SLFN) with L hidden neurons can be expressed as:

$$f(x_i) = \sum_{j=1}^{L} \beta_j h_j(a_j, b_j, x_i) = h(x_i)\beta, \quad i = 1, ..., N, \tag{10}$$

Where $\beta = [\beta_1, ..., \beta_L]^T$ is the output weight matrix, $h(x_i) = [h_1(a_1, b_1, x_i), ..., h_j[(a_L, b_L, x_i)]$ is the network output corresponding to the training sample x_i, $h_j(.)$ is a nonlinear piecewise continuous function, and $a_j \in R^d$ and $b_j \in R (j = 1, 2, ..., L)$ are parameters of the jth hidden node, respectively. The purpose of network training is to find suitable network parameters to minimize the error function $\|H\beta - T\|_2$, where:

$$H = \begin{bmatrix} h(x_1) \\ . \\ . \\ . \\ h(x_N) \end{bmatrix} \tag{11}$$

and

$$T = \begin{bmatrix} t_1^T \\ . \\ . \\ . \\ t_N^T \end{bmatrix} \tag{12}$$

are the hidden layer output matrix and the target output, respectively. The illustration of a SLFN with L hidden neurons is depicted in Fig. 5.

Other than updating the network parameters iteratively as done in conventional gradient decedent algorithms, ELM employs random hidden node parameters and the tuning free training strategy for feedforward neural networks. The learning is then transferred to solving a linear system which has been well suited via the minimal norm least square approach [30]. As shown in the universal approximation capability theorems [31], ELM is flexible with hidden activation functions. Almost any nonlinear piecewise continuous functions and their linear combinations work well in ELM algorithms [30, 31]. Thanks to these advantages, ELM has shown superiority of the fast learning speed and reasonable generalization performance over SVM and its variants [29, 30].

3 Experimental study

3.1 Datasets

In this section we describe the benchmark datasets used to evaluate the proposed approach. As stated previously, we aim to prove that our proposed method could be promising in different domains. In fact, the proposed approach was evaluated using two different databases from two different search areas, namely, breast cancer diagnosis and traffic sign recognition:

3.1.1 Digital database for screening mammography

The experiments carried out in this work used the Digital Database for Screening Mammography (DDSM) [27], which is is currently the largest public mammogram database. It is constituted of mammographic images and its corresponding technical and clinical information, including exam dates, age of patients, digitalization equipment, lesion types (according to BI-RADS [5]), and existent pathologies.

In this research, for data consistency purposes, all images are collected from the same type of scanner and resolution. We chose the scanner type Howtek 960 because a large number of cases are digitized by this type in DDSM [27]. From this data set 920 digital mammograms were randomly selected in cranio-caudal (CC) and medio lateral oblique (MLO) views. There are five classes used as main classes, which are normal mammogram, benign mass, malign mass, benign calcification, malign calcification.

Extraction of regions of interest (ROI) is an essential step before feature extraction, however it is not in the scope of this work. So, the manual extraction of ROIs is adopted for an efficient comparison. In DDSM database, an important part of the whole mammographic image comprises the pectoral muscle and the background with a lot of noise. Based on the physician annotation provided in the data set, we apply a manual cropping to images. Thus, all the background information and most of the noise are eliminated. The ROI is a

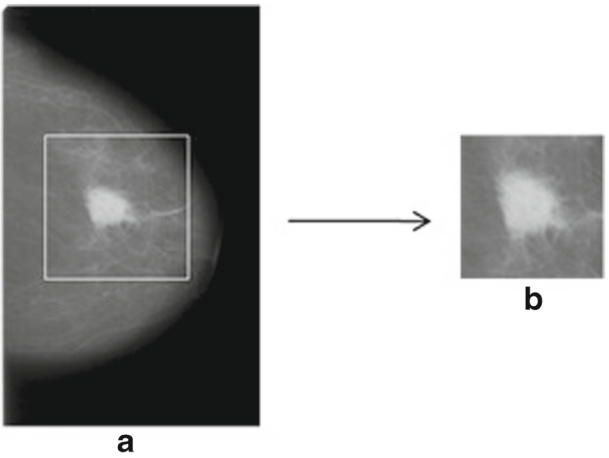

Fig. 5 Original image (3601 × 6404), (**b**) Cropped image (128 × 128)

rectangular area of (128 × 128) pixels. This size was experimentally chosen to ensure that the ROI covers the entire abnormality. Also most breast cancer diagnosis research works use ROIs of (128 × 128) [3, 16, 19]. An example of cropping is given in Fig. 5. Figure 6 illustrates, examples of ROIs extracted from the DDSM database and used in the experiments.

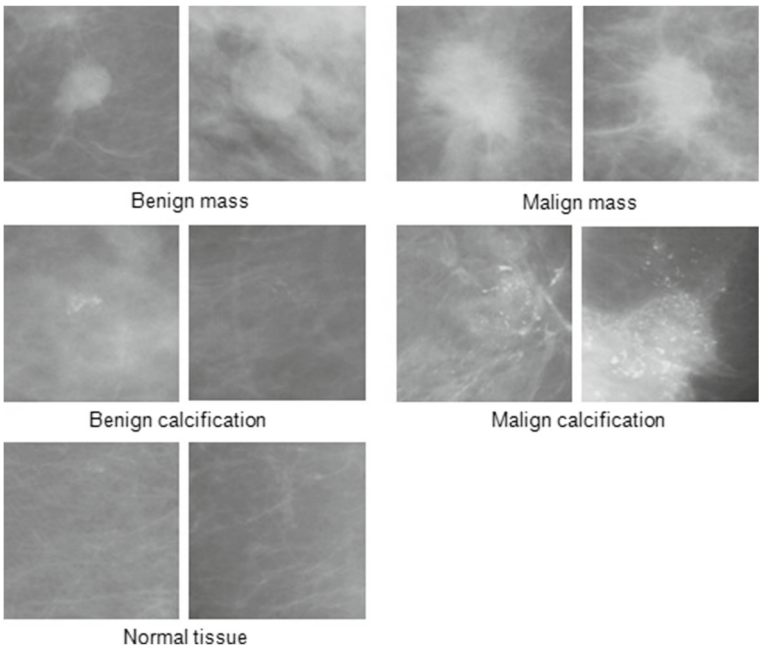

Fig. 6 A sample of the used ROIs from the DDSM database

3.1.2 German traffic sign recognition benchmark database

The database, German Traffic Sign Recognition Benchmark(GTSRB) was created from approximately 10 hours of video that was recorded while driving on different road types in Germany during daytime [46]. The GTSRB dataset contains 51,839 sign images in total, partitioned into 39,209 training images and 12,630 test images. The signs are in 43 classes, including 8 speed limit signs, 4 other prohibitory signs, 4 derestriction signs, 8 mandatory signs, 15 danger signs, and 4 unique signs. The size of the traffic signs varies between 15×15 and 222×193 pixels [46]. Figure 7 presents some sample images taken from German dataset.

In the GTSRB dataset, the size of the traffic signs varies across very different scales. Hence as a preprocessing step, all the images have been resized to 128×128 pixels.

3.2 Experiment settings

In the experimental level, the main problem covered is to distinguish between different images classes. The proposed procedure was tested in Matlab R2012a, on a 64-bit PC with an i5 microprocessor with 4 cores, 8 GB of RAM and a hard disk of 320 GB. Experiments were conducted on two different datasets. We begin our experiments on the DDSM database, which is a very large film screen mammographic dataset. Then we conducted the same experiments on the GTSRB database for traffic sign recognition.

Once the images are cropped and resized as described in the previous section, the second phase deals with feature extraction from the ROIs of the set of images. The curvelet transform is used to represent the ROIs in multiscale decomposition levels. The ROIs of images are transformed into 4 different scales (2,3,4 and 5) in order to select the best scale to strike a balance between the speed and performance of the system. Curvelet coefficients derived from a specific scale are then concatenated to form one very big feature vector (data fusion).

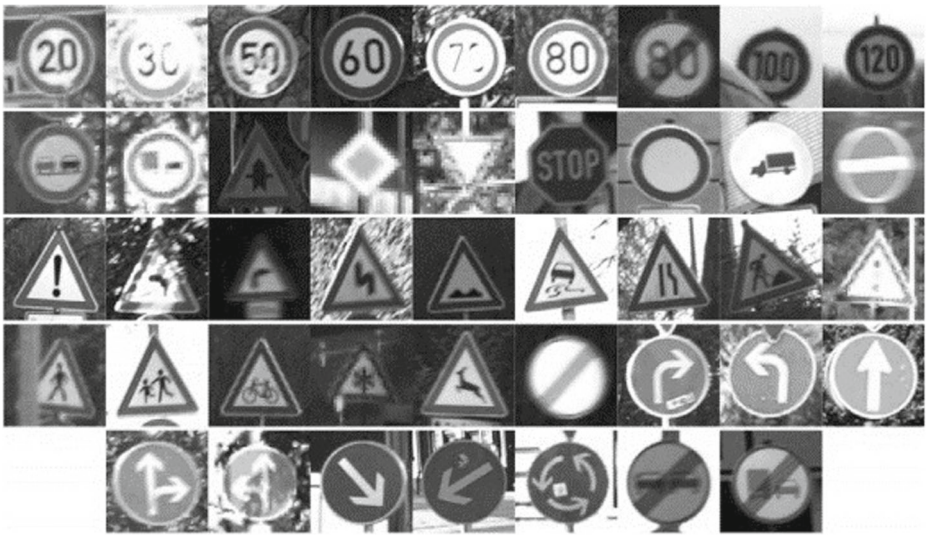

Fig. 7 Example of sample images from GTSRB database [46]

Extracted curvelet coefficients can then be directly fed to the ELM classifier to perform the identification task. An image of size 128 × 128 when decomposed using curvelet transform at scale 2 and angle 8 will produce coefficients of slightly larger size. Working with such large number of feature vectors is not suitable as the complexity of the algorithm will be extensively high.

The subspace analysis LSDA is used in order to extract the most relevant features, and to reduce the size of the data by using only the principal discriminant features. The new image to be identified is projected into d-dimensional subspace. In real cases, it is very difficult to select the suitable nearest neighborhood parameter k for LSDA. In this work we decide the suitable value of k experimentally.

In order to establish the reliability of the proposed method, comparison with the application of digital curvelet transform in conjunction with well-established existing dimensionality reduction tools like PCA and LDA is performed.

After the curvelet sub-images are projected to desired feature space, features are input to an ELM to analytically learn an optimal model. The only parameter to be determined is the number of hidden nodes. Since there is no instructive criterion that states how many hidden nodes should be used for a specific task like traffic sign recognition and the training of ELM classifier is extremely fast, here we decide the reasonable number experimentally.

3.3 Results and discussion

The First analysis investigated the effect of numbers of ELM hidden neurons using various activation functions. This was carried out to find the suitable number of hidden neurons and the best performing activation function for ELM classifier. Fast learning and testing speed offered by ELM allows us to repeat the experiments several times. So every experiment is executed 20 times for each database and average results are reported. Different numbers of hidden nodes of ELM are chosen with different types of activation functions to fid the best number. We firstly took as 10 and then increased to the input sample size by steps of 10. We searched ELM with the best recognition rate.

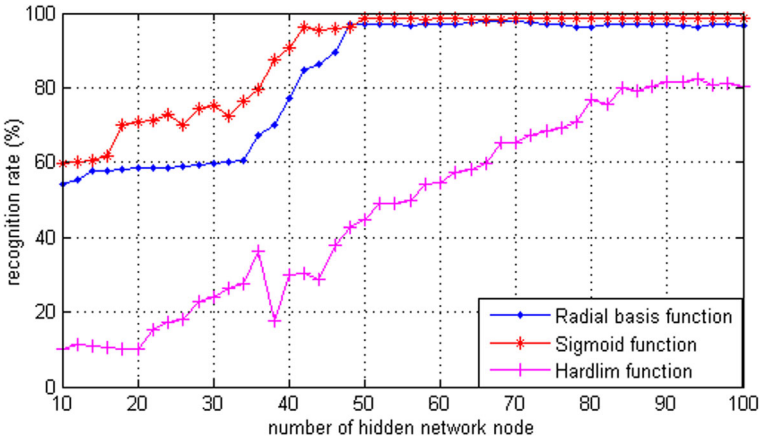

Fig. 8 The curve of recognition accuracy regarding different activation functions and different numbers of hidden nodes : GTSRB database

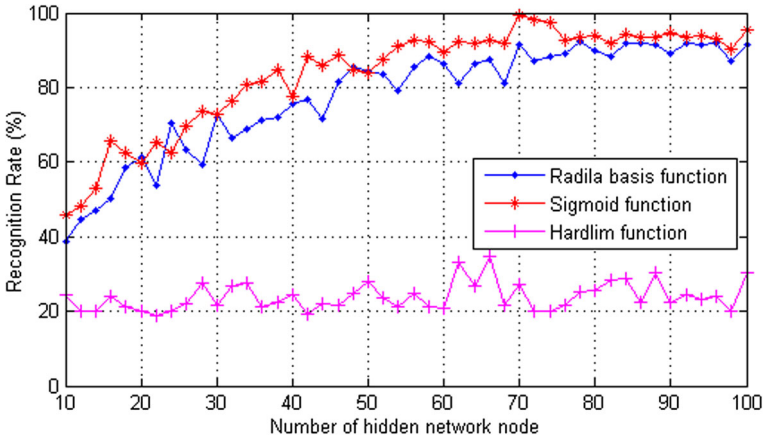

Fig. 9 The curve of recognition accuracy regarding different activation functions and different numbers of hidden nodes : DDSM database

Figures 8 and 9 illustrate the curve of recognition accuracy regarding different activation functions and different numbers of hidden nodes for GTSRB and DDSM databases respectively.

As shown in Fig. 8, the recognition rate generally increases as there are more hidden nodes used. The performances in terms of recognition rate of the sigmoid and RBF functions become better, contrary to the performance of hardlim function. Since the performance of the sigmoid function is the best among the three activation functions [45], it is chosen as the activation function for the ELM model. So far the highest performance of the recognition system 98.57 % was achieved with sigmoid function when using only 50 hidden nodes.

In the experimental work on DDSM database, it is demonstrated that using 70 hidden nodes is sufficient to produce the best performance with a sigmoid activation function 99.50 % as shown in Fig. 9.

In the previous section, we have presented a theoretical overview of curvelet transform and explained why it can be expected to work better than the traditional wavelet transform. Curvelet based feature extraction takes the raw or the prepared ROI as input. The images are then decomposed into curvelet sub-bands in different scales and orientations. We have then to deal with two major questions to bring in improvements in the classifcation performance:

1. What is the best performing scale of curvelet transform decomposition?
2. and what is the suitable number of features for the used dimensionality reduction technique (LSDA)?

Tables 1 and 2 show the classification accuracy of the proposed curvelet-LSDA method at curvelet transform $scale = \{2, 3, 4, 5\}$, $angle = 8$ on the selected images from DDSM and GTSRB datasets respectively. The dimension of features is varied to display how classification accuracy changes with selection of number of features.

As can be seen from these tables, the classification accuracies of our proposed curvelet-LSDA are consistently promising. For DDSM database with the increase in feature dimension at the beginning, the classification rates are also improved. However, the trend is not maintained for all the dimensions. When they achieve their top results (99.50 % at

Table 1 Curvelet-LSDA based mammogram diagnosis using different scales and dimensions: DDSM database

		Dimension														
		10	20	30	40	50	60	70	80	90	100	110	120	130	140	150
Scale 2	AC.	59.13	73.48	94.50	98.23	98.90	99.25	99.50	99.23	98.88	98.57	98.12	97.34	97.24	97.32	96.56
	ET.	19.5	19.44	19.88	19.99	19.93	19.94	19.89	20.17	20.12	21.63	20.89	20.13	20.45	21.65	21.88
Scale 3	AC.	57.30	79.41	94.56	97.44	98.96	99.56	99.13	98.91	98.91	98.26	97.60	96.74	96.30	95.87	95.43
	ET.	41.19	40.38	40.94	40.94	41.77	41.47	42.60	42.91	42.87	41.95	42.03	42.46	42.58	42.78	42.92
Scale 4	AC.	52.17	76.30	91.96	94.78	98.91	99.62	98.70	98.91	98.48	98.26	97.60	97.39	96.74	96.52	96.09
	ET.	46.48	47.27	48.69	51.96	52.13	52.19	51.54	52.15	50.46	53.26	53.38	54.39	53.25	54.42	53.16
Scale 5	AC.	51.52	72.17	94.57	96.52	99.62	98.91	98.70	98.70	98.26	98.04	97.83	97.39	96.96	96.52	95.43
	ET.	57.86	48.27	48.78	50.36	53.78	54.21	49.34	49.35	52.56	54.16	54.28	55.34	54.25	56.42	54.86

AC: Classification accuracy (%)

ET : Execution time for each ROI (ms)

dim=70 for scale 2, 99.56 % at dim=60 for scale 3, 99.62 % at dim=60 for scale 4 and 99.62 % at dim=50 for scale 5), the classification rates begin to decrease with the increase in the dimension.

Also, for GTSRB database, the classification rates are high at the beginning (98.57 % for scale 2 at dim=40, 96.27 % for scale 3 at dim=50 and 97.80 % for scales 4 at dim= 60 and 96.90 for scale 5 at dim=70). Moreover, after reaching the the highest value at each scale,the classification rates begin to decrease with the increase in the dimension.

For both databases a scale value of 2 has been used in order to strike a balance between the speed and performance of the system. It has been observed that higher scale has shown only marginal improvement or even no improvement in terms of classification accuracy. Hence, using more scales will only increase the execution time without improving significantly the classification accuracy.

Table 2 Curvelet-LSDA based traffic sign recognition using different scales and dimensions: GTSRB database

		Dimension														
		10	20	30	40	50	60	70	80	90	100	110	120	130	140	150
Scale 2	AC.	72.30	90.7	95.3	98.57	97.98	97.03	96.8	96.07	95.70	94.84	94.01	93.80	93.55	92.13	91.82
	ET.	18.6	18.21	18.70	18.95	19.01	19.23	19.50	19.78	19.92	19.96	19.98	19.99	19.99	20.10	20.4
Scale 3	AC.	62.20	73.21	83.40	95.34	96.27	96.05	95.84	95.16	94.80	94.33	93.18	93.03	92.20	91.08	90.29
	ET.	22.15	22.48	22.99	23.07	23.33	23.50	23.78	23.79	23.80	23.81	23.90	23.93	23.95	23.91	23.98
Scale 4	AC.	68.40	76.50	86.20	95.97	97.20	97.80	95.90	93.33	92.13	91.01	90.88	90.75	90.02	89.82	89.07
	ET.	26.40	26.37	26.45	26.50	26.90	26.98	27.07	27.19	27.41	27.59	27.63	27.80	27.91	28.03	28.11
Scale 5	AC.	71.30	75.33	88.21	90.25	95.20	95.77	96.90	95.36	95.01	94.38	91.87	91.51	90.05	88.50	87.10
	ET.	32.12	32.23	32.33	33.36	33.78	33.25	33.38	33.35	33.46	33.40	33.51	33.59	33.67	33.70	33.81

AC: Classification accuracy (%)

ET : Execution time for each ROI (ms)

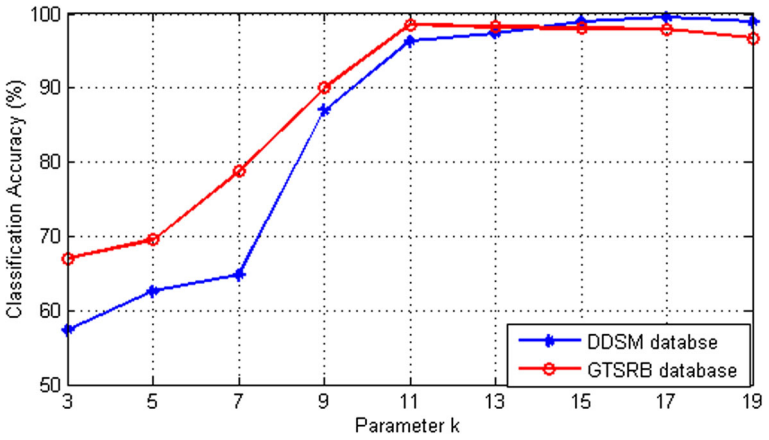

Fig. 10 The impact of K parameter value on classification accuracy

In the second experiment, the impact of the LSDA neighborhood size k on the performance of curvelet-LSDA algorithm is also tested. We set the neighborhood parameter as $k = \{3, 5, 7, 9, 11, 13, 15, 17, 19\}$. The maximal classification rates under different k values are given in Fig. 10. From this figure, it can be seen that the proposed curvelet-LSDA obtains the best recognition result when the neighborhood size parameter is $k = 17$ forDDSM database and $k = 11$ for GTSRB database. This may due to that small neighborhood size cannot well reflect the local geometry of the data manifold.

In the following experiment the performance of the proposed curvelet-LSDA is evaluated and compared with curvelet based PCA, LDA and PCA+LDA techniques. Number of features is varied to display how classification rate changes with dimension of features. Curves of classification rate versus feature dimensionality for DDSM and GTSRB databases are shown in Figs. 11 and 12.

We note here that LDA is not applicable directly when the number of the samples is much smaller than the dimension of the sample space, case of the selected sets from DDSM and GTSRB databases. This is called a small sample size problem [22]. In this case, LDA is often performed on a PCA transformed space.

For the two databases, curvelet-LSDA framework shows significant performance improvement over the PCA based and LDA based schemes. Though, LDA is usually expected to do better than PCA, it is not reflected in our experimental study as we have worked with small databases. Also curvelet based PCA+LDA is able to achieve higher classification rate in comparison with curvelet-PCA.

So we can observe that compared with the unsupervised PCA, supervised algorithms (LSDA and LDA) can achieve better recognition results. The reason is that the class label information can improve the discriminative ability of the data in low-dimensional feature space. Also as can be seen from Figs. 11 and 12, the classification rates of our proposed curvelet-LSDA are consistently better than other considered methods regardless of the dimensions.

In order to establish the reliability of the proposed curvelet-LSDA method in conjunction with ELM classifier, we have compared the performance of ELM classifier against well-established existing classification techniques like SVM, k-Nearest Neighbors (k-NN).

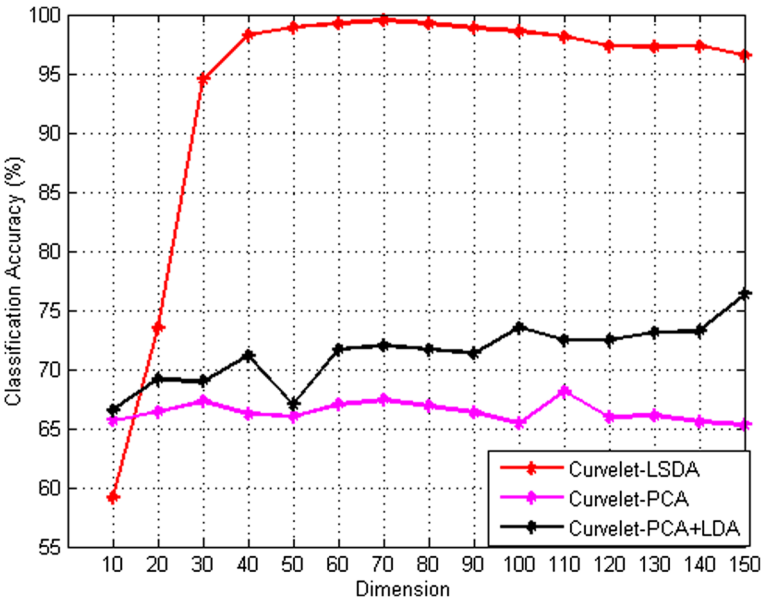

Fig. 11 Recognition accuracy vs. Dimension on the DDSM database

In this experiment, we compared the parameter optimization time, the testing time, the recognition rate and we choose the optimal parameters for each classifier that fit for correct recognition.

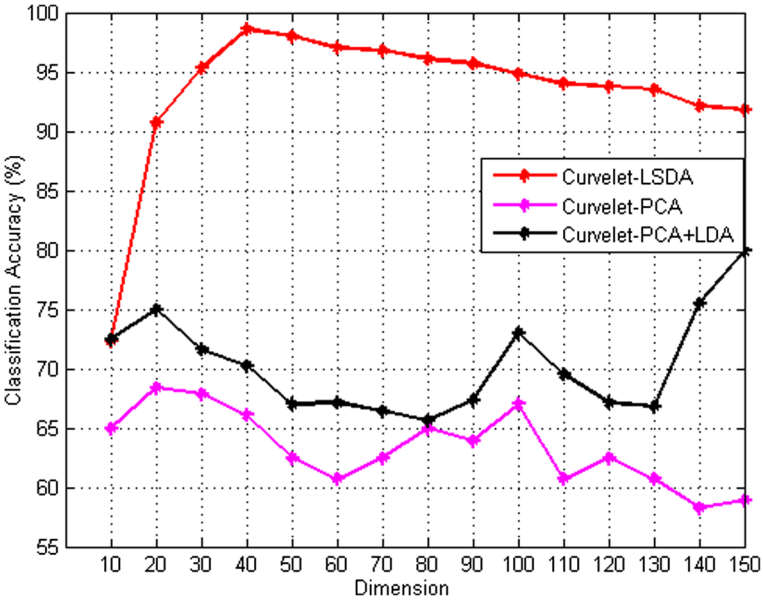

Fig. 12 Recognition accuracy vs. Dimension on the GTSRB database

Table 3 Recognition rates (%), time complexity and tuning parameters for k-NN, SVM, and ELM for color GTSRB and DDSM databases

	GTSRB					DDSM				
	P.O.T. (s)	Ts.T.(s)	Tr.T. (s)	O.P.	R.R(%)	P.O.T.	Ts.T. (s)	Tr.T. (s)	O.P	R.R(%)
SVM	80	2.13	4.03	2^3; 2^5	98.10	50	2.23	4.25	2^4; 2^6	97.90
fKNN	26.08	2.33	—-	$k = 1, d = City - block$	97.43	21.23	2.42	—-	$k = 1, d = City - block$	99.13
fELM	19.11	0.063	0.12	$N = 50$	98.57	16.35	0.047	0.15	$N = 58$	99.50

POT: Parameter Optimization Time

Ts.T. : Testing Time

Tr.T. : Training Time

O.P. : Optimal Parameters

R.R : Recognition Rate

In general, RBF kernel is a reasonable first choice when we want to implement SVM [28]. This kernel have its own parameters which need to be optimized in order to improve the performance rate of the classifier. We decide the kernel parameters using the grid search method [28].

For successful implementation of k-NN, it requires two parameters to tune: k and the distance metric (d). Selection of k values is one of the essential works to make the classification more successful. Therefore we used also the grid search method to determine the optimal value of K.

As can be seen from Table 3, the results are performed for both databases.
SVM is the most time consuming method in with regard to the parameter optimization time for both databases. For K-NN there are two major problems [4].

1. There is no output trained model to be used; the algorithm has to use all the training examples on each test, therefore its time complexity is linear.
2. Its classification performance depends on choosing the optimal number of neighbors (k), which is different from one data sample to another.

The advantage of ELM is clearly seen by taking both the recognition rate, training time and parameter optimization time into consideration. Another advantage of ELM is that the recognition performance is less dependent on parameter tuning. It is due to the simplicity of the traffic sign model and less number of tuning parameters of the ELM learning algorithm. The only parameter to be determined is the number of hidden nodes [29].

Table 4 compared the performance of two different feature sets (features extracted with curvelet transform and features extracted with the proposed curvelet-LSDA) on the DDSM and GTSRB datasets, respectively, for mammogram diagnosis and traffic sign recognition. The recognition accuracy and the feature sets size were given. This shows that the proposed curvelet-LSDA yields the highest accuracy (99.50 % for DDSM database and 98.57 % for GTSRB database) with only 70 and 40 features. Curvelet-LSDA scheme is able to achieve over 10 % higher classification rate in comparison to curvelet transform.

Finally, Tables 5 and 6 present a comparison in terms of accuracy of some representative approaches reported in literature for traffic sign recognition and breast cancer diagnosis, respectively. To our knowledge, very few research works discussion the use of multiresolution approaches for traffic sign recognition, Especially, this is the first work that exploits the use of curvelet transform in conjunction with LSDA and ELM. Also for mammogram diagnosis there are few research works discussing the use of curvelet transform. Yet, comparing the published work on these fields remains a difficult task. The accuracy of a classification system depends strongly on the used experimental protocol e.g. the used benchmark database, the number of ROIs, the nature of abnormalities to detect, etc. Moreover some studies concentrate on the complete chain or focus on the classification or the detection part only. The number of images is also an important factor, because the classification tool

Table 4 The classification rates of the images classification system via Curvelet transform without and with LSDA

	Curvelet		Curvelet-LSDA	
	Dimension	Accuracy	Dimension	Accuracy
DDSM	23609 (scale 2)	87.5 %	70	99.50 %
GTSRB	23609 (scale 2)	89 %	40	98.57 %

Table 5 Comparison with some similar traffic sign recognition representative approaches of the state of the art, in terms of recognition rate

Reference	Datset	Features and reduction technique	Classifier	Recognition rate (%)
Gonzalez et al. [25]	GTSRB	LDA	Multi-Layer Perceptron	96.5
Fleyeh and Davami [21]	Private	PCA	SVM	97.9
Cai et al. [6]	Private	Dual-tree complex wavelet transform+PCA	Nearest Neighbour	81.80
	Private	Dual-tree complex wavelet transform+LPP	Nearest Neighbour	80.40
	Private	Dual-tree complex wavelet transform+2D-PCA	Nearest Neighbour	92.15
	Private	Dual-tree complex wavelet transform+2D-ICA	Nearest Neighbour	97.24
Huang et al. [33]	GTSRB	Histogram of oriented gradient (HOG)	ELM	95.95
Our method	GTSRB	Curvelet transform+LSDA	ELM	**98.57**

Table 6 Comparison with some similar breast cancer diagnosis representative approaches of the state of the art, in terms of recognition rate

Reference	Dataset	Features and reduction technique	Classifier	Recognition rate (%)
GEDIK et al. [24]	mini-MIAS	Curvelet transform, PCA, LDA	SVM, KNN	98
Eltoukhy et al. [19]	mini-MIAS	Biggest Curvelet Coefficients	Euclidean distance	98.59
Menaka et al. [41]	mini-MIAS	Gray Level Co-occurrence Matrices	ELM	97.11
Dhahbi et al. [17]	mini-MIAS	curvelet moments	KNN	91.27
Eltoukhy et al. [20]	mini-MIAS	wavelet/curvelet transform	SVM	97.30
Malar et al. [39]	mini-MIAS	Multilevel Wavelet Analysis	Phase Encoded Complex-valued ELM	95.41
Our method	DDSM	Curvelet transform, LSDA	ELM	99.50

presents better generalization capabilities depending on the number of samples in the training step. The tables give the used database, features and classifiers used in the original works. The result obtained with our method is also included.

The results obtained from Tables 5 and 6 showed that the proposed method provide better performance in comparison with some approaches which were used in the same context explored in this work. Generally, the curvelet transform in conjunction with LSDA has reached an important classification rate. Tt is evident that proposed method is better than existing methods as the accuracy is sufficiently improved with reduction in number features.

4 Conclusion

The main contribution of this article is to extract a reduced set of discriminative features for images texture classification. We propose a new discriminative feature extraction method for images description based on curvelet transform and LSDA in combination with ELM classifier. Particularly, there is no previous work discussing the use of LSDA as a dimensionality reduction tool either for mammogram analysis or for traffic sign recognition. Experiments are conducted on two different types of databases, DDSM for film screen mammography and GTSRB for traffic sign recognition and the results validated the proposed approach consistently. Our findings confirm the utility of dimensionality reduction step since in the tested feature sets, the use of a reduced feature subset increases the accuracy values. Furthermore, The most prominent property of LSDA is the complete preservation of both discriminant and local geometrical structure in the data. Hence, using the LSDA approach with curvelet features can significantly improve the recognition rate in comparison with using the PCA and LDA methods especially in case with a small number of training samples.

References

1. Amayeh G, Amayeh S, Manzuri MT (2008) Fingerprint images enhancement in curvelet domain. In: International Symposium on Visual Computing. Springer, Berlin, pp 541–550
2. Belhumeur PN, Hepanha JP, Kriegman DJ (1997) Eigenfaces vs. fisherfaces: recognition using class specific linear projection. IEEE Trans Pattern Anal Mach Intell 19:711–720
3. Berraho S, Ait Kerroum M, Fakhri Y (2015) A hybrid feature extraction scheme based on DWT and uniform LBP for digital mammograms classification. International Review on Computers and Software (IRECOS), Vol. 10
4. Bhatia N, Vandana A (2010) Survey of nearest neighbor techniques. (IJCSIS) Int J Comput Sci Inf Secur 8:302–305
5. BI-RADS (2003) Breast imaging reporting and data system (BI-RADS) breast imaging atlas (4th ed.) American College of Radiology
6. Cai ZX, Gu MQ (2013) Traffic sign recognition algorithm based on shape signature and dual-tree complex wavelet transform. J Cent South Univ 20:433–439
7. Cai D, He X, Zhou K, Han J, Bao H (2007) Locality sensitive discriminant analysis. In: IJCAI, pp 708–713
8. Candes E, Donoho D (1999) Curvelets: A surprisingly effective nonadaptive representation for objects with edges. Department of Statistics, Stanford University, 1–10
9. Candes E, Donoho D (2000) Curvelets, multiresolution representation, and scaling laws. In: International Symposium on Optical Science and Technology. International Society for Optics and Photonics, pp 1–12

10. Candes E, Donoho D (2001) Curvelets, multiresolution representation and scaling laws. In: SPIE Wavelet applications in signal and image processing
11. Candes E, Donoho D (2004) New tight frames of curvelets and optimal representations of objects with piecewise c2 singularities. Commun Pure Appl Math 57:219–266
12. Candes E, Demanet L, Donoho D, Ying L (2006) Fast discrete curvelet transforms. Multiscale Model Simul 5:861–899
13. Chen HT, Chang HW, Liu TL Local discriminant embedding and its variants. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol 2, pp 846–853
14. Choi M, Kim RY, Kim MG (2004) The curvelet transform for image fusion. International Society for Photogrammetry and Remote Sensing. ISPRS 35:59–64
15. Dettori L, Semler L (2007) A comparison of wavelet, ridgelet, and curvelet-based texture classification algorithms in computed tomography. Comput Biol Med 37:486–498
16. Dhahbi S, Barhoumi W, Zagrouba E (2015) Breast cancer diagnosis in digitized mammograms using curvelet moments. Comput Biol Med 64:79–90
17. Dhahbi S, Barhoumi W, Zagrouba E (2015) Breast cancer diagnosis in digitized mammograms using curvelet moments. Comput Biol Med 4:79–90
18. Dong K, Feng G, Hu D (2005) Digital curvelet transform for palmprint recognition. Lect Notes Comput Sci, Springer 3338:639–645
19. Eltoukhy M, Faye I, Samir B (2010) Breast cancer diagnosis in digital mammogram using multiscale curvelet transform. Comput Med Imaging Graph 34:269–276
20. Eltoukhy M, Ibrahima F, Samir B (2012) A statistical based feature extraction method for breast cancer diagnosis in digital mammogram using multiresolution representation. Comput Biol Med 42: 123–128
21. Fleyeh H, Davami E (2011) Eigen-based traffic sign recognition. IET Intell Transp Syst 5:190–196
22. Fukunaga K (2013) Introduction to statistical pattern recognition. Academic press
23. Gao Q, Liu J, Cui K, Zhang H, Wang X (2014) Stable locality sensitive discriminant analysis for image recognition. Neural Netw 54:49–56
24. Gedik N, Atasoy A (2013) A computer-aided diagnosis system for breast cancer detection by using a curvelet transform. Turk J Electr Eng Comput Sci 21:1002–14
25. Gonzalez-Reyna SE, Avina-Cervantes JG, Ledesma-Orozco SE, Cruz-Aceves I, de Guadalupe Garcia-Hernandez M (2013) Traffic Sign Recognition Based on Linear Discriminant Analysis. In: Mexican International Conference on Artificial Intelligence, pp 185–193. Springer, Berlin
26. Grigorescu SE, Petkov N, Kruizinga P (2002) Comparison of texture features based on Gabor filters. IEEE Trans Image Process 11:1160–1167
27. Heath M, Bowyer K, Kopans D, Moore R, Kegelmeyer WP (2000) The digital database for screening mammography. In: Proceedings of the 5th international workshop on digital mammography, pp 212–218
28. Hsu CW, Chang CC, Lin CJ (2003) A practical guide to support vector classification
29. Huang G, Zhou H, Ding X, Zhang R (2012) Extreme learning machine for regression and multiclass classification. IEEE Trans Syst Man Cybern B Cybern 42:513–529
30. Huang G, Zhu Q, Siew C (2006) Extreme learning machine: theory and applications. Neurocomputing 70:489—501
31. Huang GB, Chen L, Siew CK (2006) Universal approximation using incremental constructive feedforward networks with random hidden nodes. IEEE Transactions on Neural Networks 17:879–892
32. Huang K, Aviyente S (2008) Wavelet feature selection for image classification. IEEE Trans Image Process 17:1709–1720
33. Huang Z, Yu Y, Gu J (2014) A novel method for traffic sign recognition based on extreme learning machine. In: 2014 11th World Congress on Intelligent Control and Automation (WCICA), pp 1451–1456
34. Jain AK, Duin RPW, Mao J (2000) Statistical pattern recognition: a review. IEEE Trans Pattern Anal Mach Intell 22:4–37
35. Jolliffe IT (1986) Principal component analysis. Springer, New York
36. Kong S, Wang D (2012) Multi-level feature descriptor for robust texture classification via locality-constrained collaborative strategy. arXiv:1203.0488
37. Li L, Zhanga X, Zhanga H, Hea X, Xua M (2015) Feature extraction of non-stochastic surfaces using curvelets. Precis Eng 39:212—219
38. Majumdar A (2007) Bangla basic character recognition using digital curvelet transform. J Pattern Recognit Res 2:17–26

39. Malar E, Kandaswamy A, Gauthaam M (2013) Multiscale and Multilevel Wavelet Analysis of Mammogram Using Complex Neural Network. In: International Conference on Swarm, Evolutionary, and Memetic Computing. Springer International Publishing, pp 658–668
40. Mandal T, Majumdar A, Wu QMJ (2007) Face recognition by curvelet based feature extraction. Proc Int Conf Image Anal Recognit 4633:806–817
41. Menaka K, Karpagavalli S (2014) Mammogram classification using Extreme Learning Machine and Genetic Programming. In: International Conference on Computer Communication and Informatics (ICCCI), pp 1–7
42. Mohammed AA, Minhas R, Wu QJ, Sid-Ahmed M (2011) Human face recognition based on multidimensional PCA and extreme learning machine. Pattern Recogn 44:2588–2597
43. Niyogi X (2004) Locality preserving projections. In: Neural information processing systems, vol 16, p 153
44. Roweis ST, Saul LK (2000) Nonlinear dimensionality reduction by locally linear embedding. Science 290:2323–2326
45. Skrzypek J, Karplus W (1992) Neural networks in vision and pattern recognition. World Scientific
46. Stallkamp J, Schlipsing M, Salmen J, Igel C (2012) Man vs. computer: Benchmarking machine learning algorithms for traffic sign recognition. Neural Netw 32:323–332
47. Starck JL, Donoho D, Candes E (2001) Very high quality image restoration by combining wavelets and curvelets. In: International Symposium on Optical Science and Technology, pp 9–19
48. Starck JL, Donoho D, Candes E (2003) Astronomical image representation by the curvelet transform. Astron Astrophys 398:785–800
49. Tenenbaum JB, De Silva V, Langford JC (2000) A global geometric framework for nonlinear dimensionality reduction. Science 290:2319–2323
50. Van de Wouwer G, Scheunders P, Van Dyck D (1999) Statistical texture characterization from discrete wavelet representations. IEEE Trans Image Process 8:592–598
51. Yi Y, Zhang B, Kong J, Wang J (2015) An improved locality sensitive discriminant analysis approach for feature extraction. Multimed Tool Appl 74:85–104
52. Zhang Z, Ma S, Han X (2006) Multiscale feature extraction of finger-vein patterns based on curvelets and local interconnection structure neural network. In: 18th International Conference on Pattern Recognition (ICPR'06), vol 4, pp 145–148
53. Zhang J, Zhang Z, Huang W, Lu Y, Wang Y (2007) Face recognition based on curvefaces. Proc Natural Computation

Sanae Berraho was born in Rabat, Morocco, on May 2, 1989. She received her BS degree in networks and telecommunications from Faculty of Sciences University, Mohammed V-Agdal, Rabat, Morocco in 2010. She received her Ms Degree in networks and systems from Faculty of science University, Ibn Tofail, Kenitra, Morocco in 2012.She is a PhD candidate currently at Laboratory LaRIT, team Imagery and multimedia, University Ibn Tofail,Faculty of Science, Kenitra, Morocco. Her research interest includes image processing and pattern recognition.

Samira El Margae was born in Rabat, Morocco, on September 25, 1988. She received her BS degree in networks and telecommunications from Faculty of Sciences University, Mohammed V-Agdal, Rabat, Morocco in 2010. She received her Ms Degree in networks and systems from Faculty of Science University, Ibn Tofail, Kenitra, Morocco in 2012. She is a PhD candidate currently at Laboratory LaRIT, team Imagery and multimedia, University Ibn Tofail,Faculty of Science, Kenitra, Morocco. Her research interest includes image processing and pattern recognition.

Mounir Ait Kerroum received his Master's Degree (DESA) and his PhD in Computer sciences and Telecommunications in 2003 and 2010, respectively from the Faculty of Sciences, University Mohammed V, Rabat, Morocco. He joined the ENCG (Ecole Nationale de commerce et de Gestion), Ibn Tofail University, Kenitra, Morocco, as an Assistant professor in 2010 and he was promoted to Associate Professor in November 2014 at the same university. Actually, he is the Head of Research Team in networks, telecommunications and artificial Intelligence, LaRIT laboratory (Laboratoire de Recherche en Informatique et Télécommunications). His current research interests include Pattern recognition, Artificial Intelligence, Data mining, hyperspectral and medical image processing.

Youssef Fakhri was born in Skhirat, Morocco January 2, 1979. He received his Bachelor's Degree in Electronic Physics in 2001 and his Master's Degree (DESA) in Computer and Telecommunication from the Faculty of Sciences, University Mohammed V, Rabat, Morocco, in 2003 where he developed his Master's Project at the ICI company, Morocco.He received a PhD in 2007 from the University Mohammed V - Agdal, Rabat, Morocco in collaboration with the Polytechnic University of Catalonia (UPC), Spain. He joined the Faculty of Sciences of Kénitra, Department of Computer Science and Mathematics, Ibn Tofail University, Morocco, as an Associate Professor in Mars 2009, he is the head of Networks and Telecommunications research team at LaRIT, Associate Researcher at the Laboratory for Research in Computing and Telecommunications (LRIT) in the Faculty of Sciences of Rabat, and Member of Pole of Competences STIC Morocco. His current research interests include QoS in wireless communication, WSN, wireless.