CrossMark

ORIGINAL ARTICLE

Partition selection with sparse autoencoders for content based image classification

Rik Das[1] · Ekta Walia[2]

Abstract Managing colossal image datasets with large dimensional hand-crafted features is no more feasible in most of the cases. Content based image classification (CBIC) of these large image datasets calls for the need of dimensionality reduction of features extracted for the purpose. This paper identifies the escalating challenges in the discussed domain and introduces a technique of feature dimension reduction by means of identifying region of interest in a given image with the use of reconstruction errors computed by sparse autoencoders. The automated process identifies the significant regions in an image for feature extraction. It not only improves the dimension of useful features but also contributes to increased classification results compared to earlier approaches. The reduction in number of one kind of features easily makes space for the inclusion of other features whose fusion facilitates improved classification performance compared to individual feature extraction techniques. Two different datasets, i.e. Wang dataset and Corel 5K dataset have been used for the experiments. State-of-the-art classifiers, i.e. Support Vector Machine and Extreme Learning Machine are used for CBIC. The proposed techniques are evaluated and compared in the context of both the classifiers and analysis of results suggests the appropriateness of the proposed methods for real time applications.

1 Introduction

Content based image classification (CBIC) has emerged as a research theme of importance due to escalating application of image data in assorted domains including medicine, entertainment, education, defence, etc. [1]. Radical improvement in accuracy has been observed in object recognition systems with the advent of various low level features [2]. Nevertheless, the efficiency of these systems turns out to be questionable in case of large image datasets because of the high computational overhead and elevated storage requirements. One of the reasons for the above mentioned issues has been the dimension of feature vectors extracted to represent an image. The hand-crafted feature extraction techniques in contemporary literatures have considered extraction of image features by manipulating the entire image surface [3]. However, the whole image may not be necessary to create a distinct feature vector to effectively sample the image categories using fine-grained features [4]. Literature suggests the effectiveness of image blocks for local feature extraction. Significant performance enhancement in local invariant face recognition has been observed by adaptive selection of image blocks [5]. Hence, it has become imperative to locate the region of interest (ROI) for extracting useful features to facilitate effective CBIC. Image data is growing on a daily basis, and considering the aforesaid inflation, it is essential for image-based real time applications to reduce the memory requirements and computational expenses. The authors have

✉ Rik Das
rikdas@xiss.ac.in

[1] Department of Information Technology, Xavier Institute of Social Service, Ranchi, India

[2] Department of Computer Science, University of Saskatchewan, Saskatoon, Canada

attempted to utilize partial image for feature extraction by including those image blocks having significant contribution in offering credible features. The process of insignificant image blocks removal has been carried out before feature extraction which has considerably reduced the computational overhead during feature extraction and has also contributed in feature dimension reduction. A recent work [6] has readily addressed the issue by modelling a technique of error histogram analysis of image partitions via an $m/p/m$ autoencoder. The autoencoder has been chosen to be a shallow one, (where, $p < m$) in order to identify relevant image partitions portraying higher amount of reconstruction error. This facilitates the identification of insignificant partitions which do not have much involvement in image identification process. The technique has primarily extracted features from all partitions of a segmented image and has further identified features extracted from insignificant image segments which will not contribute considerably to image recognition process [6]. Features of such insignificant segments are discarded and the remaining features are used for retrieval of medical images. However, to the best of our knowledge, removal of insignificant partitions before feature extraction by calculating reconstruction error using sparse autoencoders has been first introduced in this paper and is not being carried out in the aforesaid work. Also, as opposed to [6], where intra class image retrieval problem is addressed, for the first time, we introduce sparse autoencoder based framework for feature vector dimensionality reduction in the application of CBIC. We observe that the classification accuracy with features extracted from the entire image without segmentation has been much less compared to the novel approach of significant partition selection for feature extraction. Dimension of feature vector also gets reduced as it is extracted only from chosen partitions and not from the entire image. It is observed, that selection of partitions based on reconstruction errors using sparse autoencoders has significantly contributed in enhancing classification accuracy. Gradual lessening of the number of partitions for feature extraction with reduced feature dimension has attenuated the classification accuracy, but the level of accuracy still remains higher compared to classification accuracy with features extracted from images with no partitions. In a similar fashion, reduction in dimension of fused features for partitioned images has outperformed the classification accuracy of fused features for their counterparts with no partitions. Thus, the soundness of the proposed approach lies in the consistency in classification accuracy even after ensuring decreased feature dimension. The scientific rigor of the approach has been established by testing the proposed technique with two public datasets namely, Wang dataset and Corel 5K. The proposed approach has exhibited similar behaviour when scaled up with 5 times increased number of

images in Corel 5K compared to Wang dataset. Two different feature extraction techniques, i.e. Uniform Local Binary Patterns (*ULBP*) [7] and colour histogram [8] of *RGB* image are used to extract features from relevant partitions as well as randomly selected partitions of segmented images each of which leads to different classification accuracies. Due to the reduction in number of partitions for which features are extracted, it becomes easier to fuse more relevant features such as colour histogram with *min–max* normalization [9] and create a *fusion* based feature input to the classifiers. As expected, the fusion based approach has revealed superior classification accuracy compared to any of the individual approaches. The proposed technique has also surpassed classification accuracy recorded for two state-of-the-art techniques of feature extraction, viz., Histogram of Oriented Gradients (HOG) [10] and Scale-invariant feature transform (SIFT) [11]. This has established the partition based method to be of general interest to increase the classification accuracy. Two different datasets, namely, Wang Dataset (1000 images with 10 categories) and Corel 5K Dataset (5000 images with 50 categories) are used for evaluation purpose. Thus, on the whole, 6000 images are considered for the experimental work.

The objectives of this work can be summarized as follows:

- Enabling the use of sparse autoencoders to identify significant image partitions for dimension reduction of hand-crafted features.
- Formation of fusion based feature vector with features extracted using two different techniques.
- Using Support Vector Machine (SVM) and Extreme Learning Machine (ELM) for content based image classification with reduced features as well as fused features.
- Comparison of classification performance of proposed technique with reduced and fused feature sets.
- Comparison of the proposed approach with state-of-the-art techniques.

This paper is organized as follows. Background and conceptualization of the research work has been discussed in the Introduction part. Contemporary literature survey has been carried out in Sect. 2. The proposed framework for feature extraction is explained in Sect. 3. Classification results have been analysed and compared in Sect. 4. Finally, the research outcomes have been discussed in Sect. 5.

2 Related work

CBIC has embraced widespread literature that includes varied approaches over last 20 years [12]. Present work has introduced a novel approach of selecting the ROI in an

image for feature extraction using sparse autoencoder. Therefore, the paper is in correlation with visual feature extraction from image datasets for CBIC and also with the role of autoencoders to assist content based feature extraction. The work also facilitates fusion methodologies as it demonstrates the usefulness of a feature ensemble technique for increased classification accuracy along with dimensionality constraints. The following subsections contain survey of some of the relevant works carried out in the aforesaid areas.

2.1 Classification technique

Supervised classification utilizes the feature vectors extracted from image as input data. Researchers have developed different techniques based on Artificial Intelligence (Logical/Symbolic techniques), Perceptron and Statistics (Bayesian Networks, Instance-based techniques) [13]. The logic based algorithms are based on decision trees and rule based classifiers [14]. The decision trees sort the feature values of the instances for classification purpose. The rule based classifiers represent each class by disjunctive normal form (DNF) in classification rules. Perceptron-based techniques are either single layered perceptrons or multi-layered perceptrons. The single layered perceptron has connection weights/predictors to convert the input feature vectors to weighted inputs which are evaluated on the basis of an adjustable threshold which is maintained for the output predictions. The perceptron-like methods are binary which convert a multiclass problem to a set of multiple binary classification problems [15]. Multi layered perceptrons constitute of multi-layer neural network by joining a large number of units (neurons) together in a pattern of connections. It comprises of three different units including input units, for receiving the information meant for processing; output units, to divulge the results of processing; and hidden units, which remain in between the input and the output units. A Feed-forward network permits one way movement of signals, from input to output. An emerging learning algorithm named Extreme Learning Machine (ELM) has been extensively used for the generalized single hidden layer feed-forward neural networks [16]. The significance of ELM learning architecture is the random generation of the hidden node parameters and analytical computation of the output weights. The usefulness of ELM has been further stretched to kernel learning which has revealed the suitability of ELM for an extensive variety of classification. The statistical models are based on an explicit underlying probabilistic approach which signifies a probability that an instance belongs in each class, instead of classifying it merely. Another contemporary classification technique is Support Vector Machine (SVM) that introduces the notion of a "margin" which in turn signifies either side of a hyperplane to differentiate between two different classes [17]. Derivation of an optimum separating hyperplane can be carried out by minimizing the squared norm of the separating hyperplane for linear separation of two classes.

2.2 Feature extraction

Efficient feature extraction is fundamental necessity for the success of CBIC. The high level representation of image data can be achieved by means of feature extraction from low-level data (pixels). Fundamentally, the visual features are arranged in three different hierarchies, namely, low-level features (primitive), middle-level features (logical) and high-level features (abstract) [18]. The low-level features comprises of texture, colour, shape etc. which has been considered by most of the early systems for image classification. Recently, a surge of activity is witnessed for the development of mid-level features involving sub-images, bag-of-words approach etc. and high-level image features comprising of semantics. It is essential for the features to express the image content correctly for proper classification outcome. If the features are unable to signify the image content correctly then images can hardly be categorized.

Colour is considered to be one of the essential and meaningful features that are widely used in CBIC and object recognition [19]. Number of descriptors has been designed for extraction of colour features. Occurrence of colour in an image can be well expressed by colour histogram [20]. Significant colour distribution in the ROI can be described by means of Dominant colour descriptor [19]. Probability of locating a pair of colours based on a specified distance is depicted by colour correlogram [21, 22]. Colour moments have revealed better image identification compared to conventional methods [23]. Global features are popularly extracted by building a grey level histogram consisting of 256 bins. However, the main shortcomings of histogram based technique are probable allocation of similar colour intensities to different bins and lack of any spatial information. Calculation of local histogram has also been performed by splitting the images into partitions to overcome the aforementioned issues. Another important feature of an image is the texture feature which has a vital role for feature extraction in CBIC. The different forms of texture feature extraction techniques are of statistical, geometrical or model based nature. The statistical way of representing textures is characterized by grey-level co-occurrence matrix (GLCM), Markov random field Gabor wavelet, edge histogram descriptor (EHD) etc. [24, 25]. A realistic approach to signify the grey level textures by utilizing patterns is carried out by implementing Local Binary Patterns (*LBP*) [26]. The low computational

overhead of *LBP* along with its invariance to resolution changes has enabled it to become the state-of-the-art texture descriptor.

2.3 Autoencoders

Decoding and encoding of inputs can be carried out with minimum error by means of a special type of neural network known as autoencoders [27]. An autoencoder is trained to encode an input into some kind of representation which in turn is capable of reconstructing the same input from that representation [28]. The inputs are distributed into an arrangement of hidden layer activations followed by the successful training of an autoencoder. This process is termed as feature learning for the autoencoder. The dimension of feature is equivalent to the number of neurons in the hidden layer. A sparse autoencoder adds a penalty to the error function which ensures activation of each hidden neuron in response to only a certain type of input, not all of them.

2.4 Fusion techniques

Feature fusion enables the ensemble of diverse features extracted by means of different feature extraction techniques. Fusion of complementary features leads to enhancement of classification rate and is categorized into four different categories; early fusion, late fusion, hybrid fusion and intermediate fusion. The classifier is provided with a single input by early fusion as it combines the features of different techniques before learning. Late fusion enables decision fusion of the classification decisions obtained from different classifiers for each feature extraction technique. The two aforesaid techniques are represented in a combined form by the Hybrid fusion technique. Multiple features are integrated by means of Intermediate fusion by considering a joint model for decision to yield superior prediction accuracy [29]. Several techniques have been proposed for fusion of various complementary handcrafted features such as in [30–41].

Following observations are made based on the survey of the contemporary literature:

- None of the techniques except in [5] have considered feature extraction from a ROI in an image.
- Although improvement in classification accuracy is observed with fusion based techniques, but increase in dimension of features due to fusion has not been dealt with.
- Large feature size due to above two limitations has contributed in increased computational overhead for classification.

The authors have readily addressed the observations made in the literature review and have designed a method to handle the aforesaid issues. The novel method has outperformed the traditional approaches and has revealed improved outcomes.

3 Proposed framework

We introduce a threshold based partition selection technique by means of a shallow autoencoder [42]. We prefer the autoencoder guided approach of selective removal of small image partitions which are not relevant for feature extraction process and will not significantly contribute to image classification. It helps with reduction of feature dimension. The autoencoder comprises of generic *m/p/m* architecture for encoding m inputs into p positions and then further reconstructs m outputs back from p positions. A shallow network signifies the use of an autoencoder with $p < m$ which enables the autoencoder to compress the input to reduced dimension.

A recent approach performed fast intra-class medical image retrieval using significant image partitions [6]. Initially, each medical image has been divided into several partitions and in the training phase of classification, *LBP* features have been extracted from all image partitions and have been stored in feature matrix. Reconstruction error has been computed for all the partitions and the errors have been stored in an error histogram for each class. During testing phase, *LBP* features have been computed for the testing image's partitions which are then classified by SVM. Based on their predicted class and the corresponding error recorded in error histogram, the partitions have been either discarded or retained for intra-class image retrieval task. Here, we extract 59 *ULBP* features to encode image texture. However, in contrast to the existing technique, we eradicate the irrelevant image partitions before feature extraction and thus prohibit their use in the feature extraction stage itself. We never extract features from all image partitions which minimizes the computational overhead of our classification method. Classification is carried out with SVM and ELM both for performance comparison. The reduction in number of partitions for which features are computed, allows for a feature fusion based approach for CBIC. Therefore, colour histogram based features are also extracted from the significant image partitions in addition to *ULBP* features, and are used in classification. Feature extraction using colour histogram is carried out with *RGB* images where the dimension of extracted features is dependent on the number of input bins for each colour component. However, the number of bins must be same for all three colour components. The dimension of extracted features is $nbins^3$ where *nbins* denote the number of input bins and 3 is the number of colour components. In this work, 4 bins are chosen for each colour component resulting in feature dimension of 64.

Each image in a dataset is divided into $n \times n$ partitions of equal dimensions. We use a shallow network with *m/p/m* autoencoder to locate the image partitions that exhibit high error for decoding. The process of selecting the partitions is carried out by comparing the reconstruction error of the autoencoded partitions. Thus, the error turns out to be useful to locate the partitions relevant for feature extraction in a particular image. Image partitions with complex structures like edges, structures etc. will have high decoding error compared to image partitions having smooth regions with no significant gradient change. Classification results will have least contribution from partitions having minimum decoding error.

As mentioned earlier, the $n \times n$ partitions in an image I are to be analyzed for ROI by means of autoencoder to locate the partitions for feature extraction. A threshold is therefore set for reduction of image partitions used for feature extraction and further for classification. The threshold t is chosen such that $t \in [0, 1)$, is fraction of number of image partitions for a particular image. Therefore, the approach has minimized as many as $[t \times n \times n]$ partitions which are not useful for feature extraction. The partitions below an error threshold are removed and the rest are retained for feature extraction.

The step-by-step process for the proposed framework is given as Algorithm 1 in Fig. 1. The partitions of an image are shown in Fig. 2.

Further, we propose the use of fused features which is more feasible now with the limited number of partitions being used in classification. We have designed the fusion of features extracted from different number of selected partitions using autoencoder's error as shown in Figs. 1 and 2. The process of fusion is depicted in Fig. 3.

4 Experimental results and analysis

The experiments are conducted with Neural Network Toolbox of MATLAB R2015b on a machine with Intel core i5 processor and 4 GB RAM. We have used two different datasets which are briefly described in the following subsections. Further, the reconstruction error generated by the autoencoder is discussed. Thereafter, we discuss the results obtained from various classification techniques. Finally, we report and compare the results of classification obtained from diverse feature extraction techniques.

4.1 Datasets used

4.1.1 Wang dataset

Wang et al. have provided this widely used public dataset with 1000 images [40]. Dimension of every image is 256 × 384 or 384 × 256 pixels and the images are divided into 10 different categories with 100 images in each category-Tribals, Sea Beaches, Gothic Structures, Buses, Dinosaur, Elephants, Flowers, Horses, Mountains and Food. A sample collage for Wang's dataset has been given in Fig. 4.

4.1.2 Corel 5K dataset

Corel 5K is another benchmark dataset in computer vision experiments which has 50 different categories of images with dimensions: 128 × 192 or 192 × 128 pixels [43]. Some of the image categories are beer, wolf, lion, elephant, tiger, mountains, vegetables, faces etc. A sample collage of the Corel dataset is given in Fig. 5. For the sake of simplicity, the images of both datasets are resized to square sizes.

4.2 Error measurement for partition selection

Shallow neural network architecture of an autoencoder results in some level of error when the input values provided to the input neurons are compressed by the neurons in the hidden layer and are further reconstructed back into same number of output values as that of the input values. The number of grey values in each partition created by partitioning an image into *nxn* partitions is fed as input to the autoencoder. The number of input neurons and output neurons are equal to the number of grey values and the number of neurons in the hidden layer is lesser than the neurons in input/output layers as $p < m$. The value of p is chosen to be 10 in our work which is much less than the input value m. For example, the value of m is 4096 in case when the 256 × 256 Wang dataset image is divided into 16 equal partitions and is 1024 in case when the 128 × 128 Corel 5K dataset image is divided into 16 equal partitions. The reconstructed input values produced from the compressed hidden layer is available at the output layer and the error of reconstruction is measured with mean squared error (MSE) method as given in Eq. (1).

$$\text{MSE} = \frac{1}{m} \sum_{y=1}^{\sqrt{m}} \sum_{x=1}^{\sqrt{m}} \left[I(x,y) - I'(x,y) \right]^2 \tag{1}$$

where, $I(x,y)$ and $I'(x,y)$ are the autoencoder's input and output grey values of a partition. À flow-chart of feature extraction process is given in Fig. 6.

The threshold of partition selection is initialized at the beginning and based on the higher MSE values, the partitions are chosen from a set of consecutive partitions. The partitions with lower MSE values within a set are discarded and not considered for feature extraction.
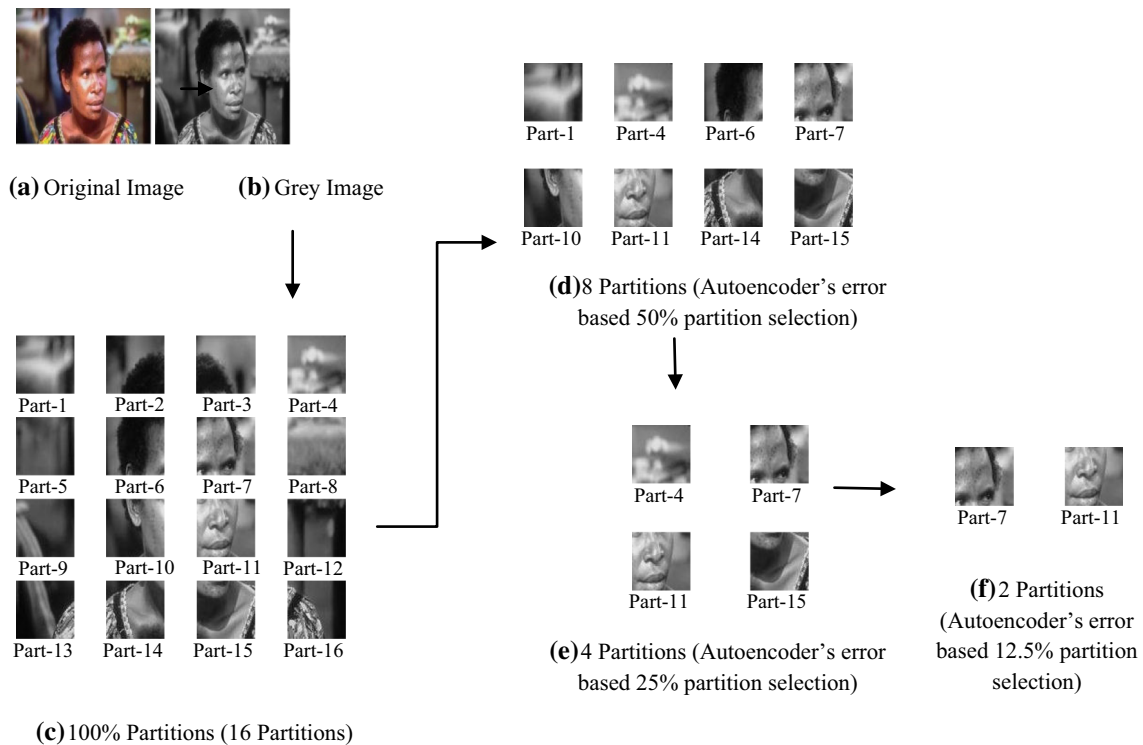
Algorithm 1

Begin

1. Input a grey scale Image I.

2. Initialize number n in order to segment the image I, into $n \times n$ partitions.

3. For each of the partitions (P) Repeat

    3.1 /* Calculate the reconstruction error generated by autoencoder using procedure *autoEncode*/

    $$error \leftarrow autoEncode(P)$$

4. Initialize the threshold $t \in [0,1)$ based on the percentage number of partitions to be retained.

5. Compare reconstruction error of nt consecutive partitions where $nt = \left\lceil \dfrac{1}{t} \right\rceil$.

5.1. For each set of error values $E_1...E_{nt}$ of nt consecutive partitions $P_1...P_{nt}$, Repeat

5.1.1/* Identify maximum error value using procedure *sort*/

    $$\max\_Error \leftarrow sort(E_1,....E_{nt})$$

    5.1.2. Retain the partition P_R having *max_Error*.

6. Discard the remaining $\dfrac{n \times n}{nt}$ partitions.

7. For each retained partition $P_{R,}$ Repeat

    7.1/* Extract *ULBP/Colour Histogram feature vector(s) (fv)* using procedure *extract<technique>features*/

    $$fv \leftarrow extractULBPfeatures\ (P_R)\,/\,extractHISTOGRAMfeatures\ (P_R)$$

8. Use the *feature vector(s)* to perform classification with 10-fold cross validation/ leave-one-out using SVM or ELM classifiers.

End

Fig. 1 Proposed framework
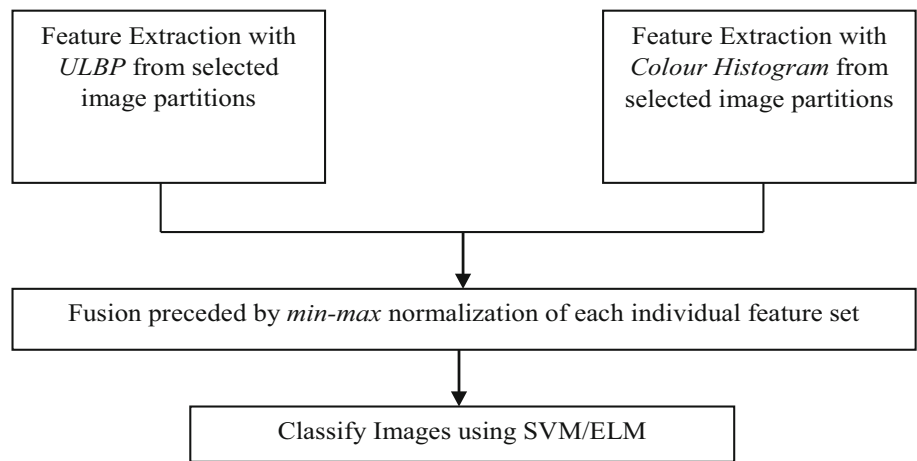
4.3 Classification techniques

Two different classifiers, namely, SVM and ELM, have been used for the purpose of CBIC with the novel approach of feature extraction proposed in this work. The classification is performed using *tenfold cross validation* and *leave-one-out* validation. For *tenfold cross validation*, the entire dataset is divided into 10 *subsets*. For *leave-one-out*, it is divided into N (where N is the number of images in the dataset) subsets. Training set for *tenfold cross validation* comprises of 9 *subsets* and the one *remaining subset* is considered as the testing set. In case of *leave-one-out cross* validation, the training set comprises of $N - 1$ subsets and the testing set is the remaining *subset*. The method is repeated for 10 trials in case of *tenfold cross* validation and for N trials in case of *leave-one-out*. In each of the trials, the testing and training subsets are changed in round robin fashion. The average classification accuracy obtained in 10 *trials* for *tenfold cross validation* and for N trials for *leave-one out* validation measures the performance of classification. The classification accuracy is calculated by averaging the ratio of number of correctly classified images n_c to the total number of images in the dataset N, as in Eq. (2).

(a) Original Image   **(b)** Grey Image

(d) 8 Partitions (Autoencoder's error based 50% partition selection)

(c) 100% Partitions (16 Partitions)

(e) 4 Partitions (Autoencoder's error based 25% partition selection)

(f) 2 Partitions (Autoencoder's error based 12.5% partition selection)

Fig. 2 Partition selection for feature extraction with reconstruction error of autoencoder

Fig. 3 Fusion of features for classification

Feature Extraction with *ULBP* from selected image partitions

Feature Extraction with *Colour Histogram* from selected image partitions

Fusion preceded by *min-max* normalization of each individual feature set

Classify Images using SVM/ELM

$$\text{Accuracy} = \frac{1}{N} \sum_{k} \frac{n_c}{N} \tag{2}$$

where, k is the number of trials, which is 10 in case of *tenfold cross validation* and N in case of *leave-one-out cross validation*.

4.4 Classification results

4.4.1 Results on Wang dataset

For the sake of simplicity, the images are resized using bicubic interpolation to 256×256, 128×128, 64×64 and 32×32 pixels. We divide the images into 16 partitions in each of the cases considered. Thus, each of the 16 partitions is of 64×64, 32×32, 16×16 and 8×8 pixels respectively. Initially, *tenfold cross* validation is carried out to measure the classification accuracy as depicted in Tables 1, 2, 3 and 4. For the purpose of comparison, classification is first carried out with *ULBP* features extracted from entire image without dividing it into partitions. Further, the classification is carried out by dividing the image into 16 partitions and gradually reducing the partitions for feature extraction based on the proposed framework. Further, we also compare the results with

Fig. 4 Three sample images of each class of Wang's dataset

Fig. 5 Sample collage for Corel 5K dataset

random selection of partitions. SVM classifier with Radial Basis Function (*RBF)* kernel is used here for the classification purpose.

Results in Table 1 can be used to compare the *tenfold cross validation* classification accuracy obtained with feature vectors extracted from images without segmentation
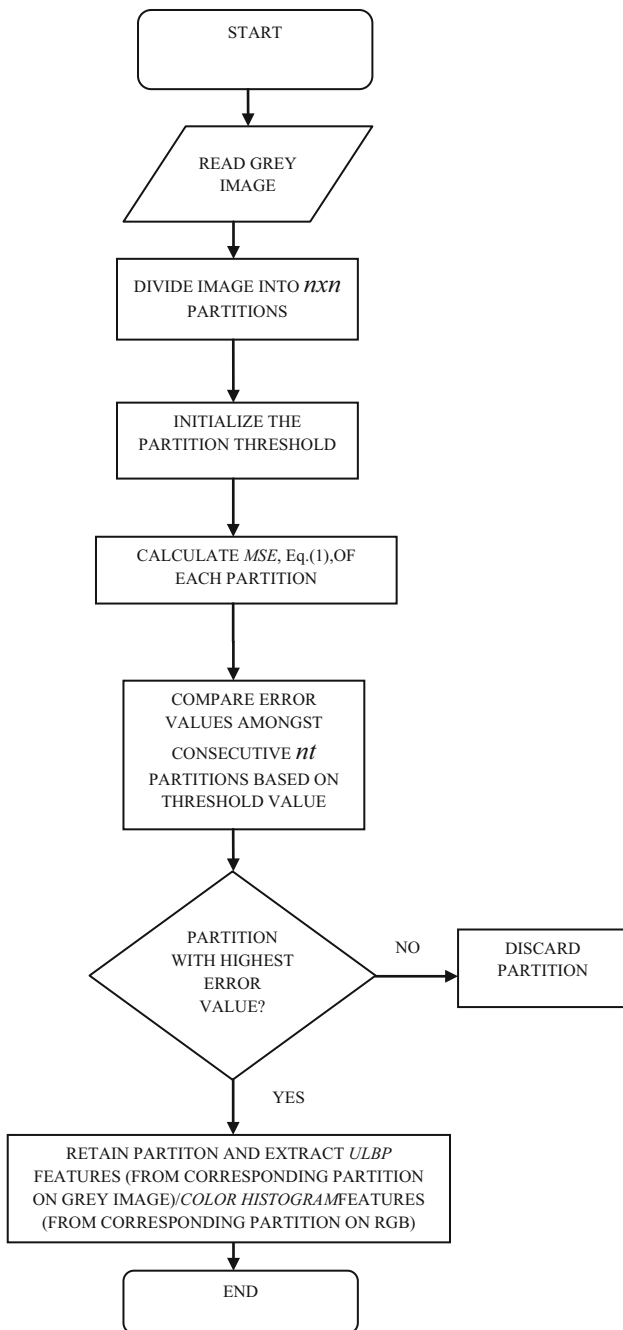
Fig. 6 Flowchart for feature extraction process

and with segmentation into 16 partitions. The classification accuracy with *ULBP* features extracted globally from images without segmentation is observed to be 53.3% for Wang dataset. Further, the images are segmented into 16 partitions and evaluated by extracting features from each of the partitions locally using *ULBP*, i.e. 100% feature dimension, which results in a classification accuracy of 80.2% using SVM classifier with *RBF* kernel. The same setup for classification is maintained and feature dimension reduction is carried out by selecting partition threshold

$t \in [0, 1)$. Each of the 16 partitions is autoencoded and MSE is computed. Based on the threshold, *nt* consecutive partitions are compared, and the partitions that exhibit higher MSE values are retained. Thus, feature dimension reduction is carried out by lowering the threshold limit ranging from 0.5 to 0.125 of total number of partitions created for each image. The classification accuracies obtained for features reduced with different thresholds 0.5, 0.25 and 0.125 are 73.4, 62.6 and 52.9% respectively.

The results discussed above reveal that the classification accuracy for segmentation based feature extraction using *ULBP* is found to be higher when all the 16 segments are used for feature extraction and also for all threshold based segment selection ranging from 0.5 to 0.25 compared to feature extraction from image without segmentation from the whole image.

Classification performance with *ULBP* features extracted from 100% of the image segments (16 partitions) reduces by 6.8%, when compared to the classification performance with dimension reduction of features by 50% (8 partitions), which is further reduced by 10.8% when features are reduced from 50 to 25% (4 partitions) and drops by 9.7% when dimension is reduced from 25 to 12.5% (2 partitions).

Finally, extraction of *ULBP* features has been carried out by randomly selecting image partitions from 16 different partitions created in the image. The threshold range has been maintained as $t \in [0, 1)$. The accuracy values for 50, 25 and 12.5% of feature dimension generated from randomly selected blocks are 58, 50.4 and 40.8% respectively. All the accuracy values are less than the corresponding accuracy values obtained for features extracted from autoencoded blocks..

Different image dimensions 128×128, 64×64 and 32×32 pixels, have been considered further for the experiments to demonstrate consistency of the observations made in Table 1. Each of the different image dimensions have been divided into 16 equal partitions and SVM classifier with *RBF* kernel is used for classification. The results are displayed in Tables 2, 3 and 4.

An analysis of the results presented in Tables 2, 3 and 4 indicates that classification performance exhibits similar characteristics for different image dimensions except 32×32 pixels images. In case of 32×32 pixels images, the classification accuracy with segmented images is higher only up to 0.5 threshold value compared to that of image without segmentation.

Therefore, the following inferences can be drawn from the above analysis:

- Local feature extraction from segmented image using *ULBP* has higher classification accuracy compared to image without segmentation.

Table 1 *Tenfold cross validation* accuracy of SVM (RBF kernel) with *ULBP* on 256 × 256 pixels images of Wang dataset

Partition selection with proposed framework				Random partition selection		
Threshold	Number of partitions	Number of *ULBP* features	Accuracy (in %)	Number of partitions	Number of *ULBP* features	Accuracy (in %)
Nil	Nil	59	53.3	–	–	–
–	16	16 × 59	80.2	–	–	–
0.5	8	8 × 59	73.4	8	8 × 59	58
0.25	4	4 × 59	62.6	4	4 × 59	50.4
0.125	2	2 × 59	52.9	2	2 × 59	40.8

Table 2 *Tenfold cross validation* accuracy of SVM (RBF kernel) with *ULBP* on 128 × 128 pixels images of Wang dataset

Partition selection with proposed framework				Random partition selection		
Threshold	Number of partitions	Number of *ULBP* features	Accuracy (in %)	Number of partitions	Number of features	Accuracy (in %)
Nil	Nil	59	57.9	–	–	–
–	16	16 × 59	79.5	–	–	–
0.5	8	8 × 59	73.9	8	8 × 59	55.7
0.25	4	4 × 59	63.5	4	4 × 59	48.9
0.125	2	2 × 59	51.3	2	2 × 59	37.7

Table 3 *Tenfold cross validation* accuracy of SVM (RBF kernel) with *ULBP* on 64 × 64 pixels images of Wang Dataset

Partition selection with proposed framework				Random partition selection		
Threshold	Number of partitions	Number of features	Accuracy (in %)	Number of partitions	Number of features	Accuracy (in %)
Nil	Nil	59	50.3	–	–	–
–	16	16 × 59	74.6	–	–	–
0.5	8	8 × 59	67.6	8	8 × 59	50.7
0.25	4	4 × 59	57.7	4	4 × 59	43.3
0.125	2	2 × 59	43.7	2	2 × 59	34.7

Table 4 *Tenfold cross validation* accuracy of SVM (RBF kernel) with *ULBP* on 32 × 32 pixels images of Wang dataset

Partition selection with proposed framework				Random partition selection		
Threshold	Number of partitions	Number of features	Accuracy (in %)	Number of partitions	Number of features	Accuracy (in %)
Nil	Nil	59	48.9	–	–	–
–	16	16 × 59	65.3	–	–	–
0.5	8	8 × 59	56.4	8	8 × 59	40.4
0.25	4	4 × 59	41.8	4	4 × 59	34
0.125	2	2 × 59	31.3	2	2 × 59	27.3

- Radical reduction in feature dimension by means of autoencoded error based partition selection with predefined partition threshold leads to decrease in classification accuracy, which in most cases (always with 50% dimension reduction) is still higher than accuracy observed without segmentation.
- Feature reduction carried out with random selection of image partition has less accuracy compared to

autoencoded error based partition selection as the ROI is not properly identified in case of random selection.

Henceforth, the autoencoded error based partition selection and random partition selection are carried out once again on the dataset images for feature extraction using colour histogram technique. The image dimension in this case is considered to be 256×256 pixels and it has been divided into 16 partitions.

Later, the features extracted using colour histogram and ULBP methods are used for classification with the ELM classifier using RBF kernel. Also, the features extracted with two different feature extraction techniques are fused using min–max normalization and provided as input to SVM and ELM classifiers. In case of Wang dataset, leave-one-out validation is also carried out to evaluate classification accuracy. For the Wang dataset, the performance of SVM on individual and fused features is depicted in Table 5.

Table 6 depicts the performance of ELM classifier using leave-one-out cross validation with individual and fused features of Wang dataset.

4.4.2 Results on Corel 5K dataset

We perform tenfold cross validation on the bigger Corel 5K dataset using the individual and fused features. For the sake of simplicity, the images are resized using bi-cubic interpolation to 128×128 pixels. The results obtained with SVM and ELM classifiers, both with RBF kernel, are presented in Tables 7 and 8 respectively.

The results in Tables 7 and 8 reinforce the performance pattern of ELM and SVM observed with Wang dataset in Tables 1 through 6. In general, ELM classifier outperforms SVM in terms of classification accuracy.

Therefore, with our experiments, we analyse that the RBF kernel exhibits superior performance with ELM classifier in comparison to SVM classifier. Also, dimension reduction facilitates the fusion of colour histogram and ULBP features for classification.

4.4.3 Comparison with state-of-the-art techniques

Finally, the proposed technique of feature extraction based on partition selection has been compared with respect to two different baseline algorithms; viz., Histogram of Oriented Gradients (HOG) [10] and Scale-invariant feature transform (SIFT) [11]. The experiment is carried out with Wang dataset. Different parameters of comparisons have been considered including comparison of feature dimension, classification accuracy, individual feature extraction and classification time as well as total time consumed for feature extraction and classification. Each of these comparisons is documented in Table 9.

4.4.3.1 Feature dimension Comparison of feature dimension shown in Table 9 has revealed maximum feature dimension for HOG (Cell size = 32) features compared to all other feature extraction techniques (SIFT, ULBP, Colour Histogram and Fusion). SIFT has feature dimension much smaller compared to HOG (Cell

Table 5 *Leave-one-out* cross validation accuracy of SVM (RBF kernel) with Colour Histogram, ULBP and fused features on 256×256 pixels images of Wang dataset

Features	Partition selection with proposed framework				Random partition selection		
	Threshold	Number of partitions	Number of features	Accuracy (in %)	Number of partitions	Number of features	Accuracy (in %)
Colour histogram	Nil	Nil	64	48.60	–	–	–
	–	16	16×64	69.90	–	–	–
	0.5	8	8×64	61.50	8	8×64	53.60
	0.25	4	4×64	58.70	4	4×64	35.10
	0.125	2	2×64	46.30	2	2×64	41.40
ULBP	Nil	Nil	59	31.70	–	–	–
	–	16	16×59	70.50	–	–	–
	0.5	8	8×59	59.70	8	8×59	47.20
	0.25	4	4×59	56.70	4	4×59	45
	0.125	2	2×59	50.10	2	2×59	25.30
Fused feature	Nil	Nil	$59 + 64$	58.60	–	–	–
	–	16	$16 \times (59 + 64)$	82.10	–	–	–
	0.5	8	$8 \times (59 + 64)$	77.70	8	$8 \times (59 + 64)$	69.20
	0.25	4	$4 \times (59 + 64)$	61.60	4	$4 \times (59 + 64)$	61.60
	0.125	2	$2 \times (59 + 64)$	52.60	2	$2 \times (59 + 64)$	48.50

Table 6 *Leave-one-out* cross validation accuracy of ELM (RBF kernel) with *ULBP*, colour histogram and fused feature son 256 × 256 pixels images of Wang dataset

Partition Selection	Features	No. of partitions				
		Nil (%)	16 (%)	8 (%)	4 (%)	2 (%)
Proposed framework (classification accuracy in percentage)	*ULBP*	78.2	81.1	81	76.3	65.9
	Colour histogram	79.1	82.8	79.4	74.5	24.7
	Fused features	81.6	88.2	85	80.6	74.1
Random selection (classification accuracy in percentage)	*ULBP*	–	–	62.6	59.1	54.9
	Colour histogram	–	–	71	66.8	59.7
	Fused features	–	–	74.9	73.2	69

Table 7 *Tenfold cross* validation accuracy of SVM (RBF kernel) with *ULBP*, colour histogram and fused features on 128 × 128 pixels images of Corel5K Dataset

Partition selection	Features	No. of Partitions				
		Nil (%)	16 (%)	8 (%)	4 (%)	2 (%)
Proposed framework (classification accuracy in percentage)	*ULBP*	11.2	57.9	36.3	30.3	18.06
	Colour histogram	8.3	46.7	45.6	40.5	22.3
	Fused features	19.2	70.7	51.3	48.8	34.2
Random selection (classification accuracy in percentage)	*ULBP*	–	–	16.3	8.7	5.6
	Colour histogram	–	–	15.5	10.6	7.3
	Fused features	–	–	40.9	22.7	12.8

Table 8 *Tenfold cross* validation accuracy of ELM (RBF kernel) with *ULBP*, colour histogram and fused features on 128 × 128 pixels images of Corel 5K Dataset

Partition selection	Features	No. of partitions				
		Nil (%)	16 (%)	8 (%)	4 (%)	2 (%)
Proposed framework (classification accuracy in percentage)	*ULBP*	50	71.1	64.8	53.2	36.06
	Colour histogram	49.9	65.1	63.6	57.9	39.6
	Fused features	64.7	73.5	72.2	71.5	57.5
Random selection (classification accuracy in percentage)	*ULBP*	–	–	36.7	29.8	24.8
	Colour histogram	–	–	44.6	38.06	13.1
	Fused features	–	–	52.3	51.2	43.4

size = 32/64), ULBP and Colour Histogram computed for images with 16 and 8 partitions and also for colour histogram of images with 4 selected partitions. HOG (Cell size = 64) has lower feature dimension compared to ULBP, Colour Histogram and fused features computed for images with 16 partitions. As expected, all the fused features for images having 16 and 8 selected partitions have larger dimension compared to SIFT and HOG (Cell size = 64) features. However, ULBP feature dimension for images with 4 selected partitions and dimension of ULBP, Colour histogram and fused features extracted from 2 selected image partitions is small compared to both the state-of-the-art HOG and SIFT.

4.4.3.2 Classification accuracy Table 9 also shows comparison of the classification accuracy of the proposed and state-of-the-art feature extraction techniques. Two different classifiers, viz., SVM and ELM have been used for the classification purpose. For the proposed technique 16, 8, 4 and 2 partitions are considered. For 16 and 8 partitions, ULBP, colour histogram and fused features have higher accuracy than all the state-of-the-art techniques i.e. HOG (cellsize = 32/64) and SIFT with both the classifiers. In case of 4 partitions, classification with SVM and ELM has higher accuracies for ULBP, colour histogram and fused features compared to SIFT. But classification with HOG (cell size = 64) features has outclassed the 4

Table 9 Comparison of proposed technique with State-of-the-art techniques for Wang Dataset images of 256 × 256 pixels

Types of techniques	Image pre-processing criteria	Name of techniques	Feature dimension	Classifiers	Accuracy (tenfold cross validation)	Feature extraction time (in s)	Classification time (in s)	Total time (in s)
State-of-the-art	Cell size = 64	HOG	576	SVM	70.7	21.87	1.2	23.07
			576	ELM	73.2	21.87	0.04	21.91
	Cell size = 32	HOG	2304	SVM	63.7	41.82	31.04	72.86
			2304	ELM	73.5	41.82	0.1	41.92
	No. of keypoints = 250	SIFT	250	SVM	21.8	7596.42	5.87	7602.29
			250	ELM	47	7596.42	0.02	7596.44
Proposed techniques	No. of partitions = 16	ULBP	944	SVM	80.2	29.98	1.32	31.3
			944	ELM	82.4	29.98	0.042	30.022
		Colour histogram	1024	SVM	80	3999.31	1.21	4000.52
			1024	ELM	83.5	3999.31	0.045	3999.355
		Fusion	1968	SVM	90.2	4029.29	3.31	4032.60
			1968	ELM	89.5	4029.29	5.8	4035.09
	No. of partitions = 8	ULBP	472	SVM	73.4	4580.17	1.2	4581.37
			472	ELM	81.5	4580.17	0.041	4580.211
		Colour histogram	512	SVM	76.6	5946.08	1.05	5947.13
			512	ELM	82.9	5946.08	0.035	5946.115
		Fusion	984	SVM	87.5	10,526.25	2	10,528.25
			984	ELM	85.2	10,526.25	0.049	10,526.3
	No. of partitions = 4	ULBP	236	SVM	62.6	4285.9	0.7	4286.6
			236	ELM	74.2	4285.9	0.03	4285.93
		Colour histogram	256	SVM	64.5	5174.06	0.84	5174.9
			256	ELM	74	5174.06	0.02	5174.08
		Fusion	492	SVM	71.5	9459.96	0.92	9460.88
			492	ELM	80.2	9459.96	0.03	9459.99
	No. of partitions = 2	ULBP	118	SVM	52.9	4025.47	0.63	4026.1
			118	ELM	66.5	4025.47	0.02	4025.49
		Colour histogram	128	SVM	49.1	4695.11	0.73	4695.84
			128	ELM	61.8	4695.11	0.02	4695.13
		Fusion	246	SVM	67.1	8720.58	0.78	8721.36
			246	ELM	72.4	8720.58	0.03	8720.61

partitions ULBP and colour histogram features for SVM in contrast to that of ELM. Though, the fused feature (only with dimension 492 much less than HOG) for 4 partitions has shown greater classification accuracy compared to both the state-of-the-art techniques.

ULBP, colour histogram and fused features extracted only from 2 selected partitions of images underperform as compared to HOG (Cell size = 32/64) features. However, they still outperform SIFT.

The above analysis has indicated that in most of the test cases, the proposed method has surpassed the classification accuracies with both the state-of-the-art feature extraction techniques, viz., HOG (Cell size = 32/64) and SIFT. However, HOG features extracted with cell size = 32 has shown superior classification results compared to proposed technique for 4 and 2 partitions. The

image has 64 partitions in case of HOG cell size = 32 which is four times the maximum partitions of 16 made in images for proposed technique of feature extraction. Yet, the proposed technique have surpassed classification accuracy with HOG (Cell size = 32) in several occasions. Further, when cell size for HOG feature extraction is increased to 64, the number of image partitions comes down to 16 which equals to the maximum number of image partitions made in the proposed technique. It has been observed that classification results of proposed technique of feature extraction (i.e. fused features) from 16 image partitions which is equal to that of HOG (Cell size = 64) as well as from 8 and 4 image partitions which are half and one-fourth respectively compared to that of image partitions for HOG (Cell size = 64) achieves better performance.

Therefore, the efficacy and robustness of the proposed method have been well established and found to be superior compared to that of the state-of-the-art techniques.

4.4.3.3 Comparison of speed Table 9 also shows comparison of the time consumed for feature extraction, classification and total time by different techniques. SIFT has been used to extract features from images without partitions, whereas, ULBP and Colour Histogram have been used to extract features from selected image partitions as per the proposed framework. SIFT is the costliest in terms of the feature extraction time, however, HOG takes minimum feature extraction time amongst all the individual techniques. It is worth mentioning here that for the case of 16 partitions, since no selection is made the feature extraction time for *ULBP* is much less as compared to *ULBP* feature extraction for 8, 4 and 2 partitions, where feature extraction time includes partition selection as well. There is a general trend in decrease of feature extraction times for ULBP, Colour Histogram and fused features going from 8 to 4 to 2 partitions.

Comparison of classification time consumed with two different classifiers, viz., SVM and ELM is also shown in Table 9. The classification process has been carried out with features extracted by the proposed approach and the state-of-the-art techniques. It is found that classification time with SVM and ELM for HOG features extracted from images is higher with respect to the classification time consumed with other features extracted using SIFT, ULBP and Colour Histogram. SIFT has been recorded to have lesser or equal classification time for ELM compared to classification time with features extracted with ULBP, Colour Histogram and fused features from 16, 8 4 and 2 selected partitions. Conversely, for classification with SVM it has been observed that SIFT has consumed greater than or equal time to that of with ULBP, Colour Histogram and fused features from 16, 8 4 and 2 selected partitions for Colour Histogram and Fusion. In accordance with [16], ELM is almost always found to be very fast at classification as compared to SVM.

A comparison of time consumed collectively during the process of feature extraction and classification is also given in the last column of Table 9. It is observed that combined time of feature extraction and classification with SVM as well as ELM using SIFT is higher than combined time for ULBP and colour histogram (with SVM and ELM respectively) from 16, 8, 4 and 2 different partitions of image. However, HOG is very inexpensive than the proposed feature extraction techniques in terms of the total time for feature extraction and classification time.

Therefore the envisioned objectives are fulfilled and results can be summarized as follows:

- Partition selection with the proposed framework consistently outperforms random partition selection, indicating that significant image partitions are successfully identified using sparse autoencoders.
- Due to reduced dimensions of individual features, fusion based features are successfully created and they consistently outperform individual features for content based image classification.
- ELM with RBF kernel achieves better classification accuracy than SVM with the same kernel.
- Dimension reduction results in smaller drop in the classification accuracy with ELM than with SVM.
- The proposed approach has outperformed the state-of-the-art techniques in terms of classification accuracy.

5 Conclusion

Image data has an unquestionable influence in current context in almost every real-world applications. Plethora of research work is conducted with an effort to design efficient techniques to identify the desired image information with minimum computational overhead. But most of the techniques yield hefty feature size to represent the corresponding image which in turn makes the identification process time consuming. Moreover, entire image is considered for feature extraction which may not be essential for effective feature extraction. This paper identifies the limitations of the existing systems and has presented a novel methodology to locate the region of interest (ROI) in an image for extraction of feature vectors. Identifying the ROI drastically reduces the feature size with least impact on accuracy for content based image classification. The authors have innovatively used sparse autoencoder to identify the significant image regions by comparing reconstruction errors of different image partitions, which leads to feature vector dimension reduction. Experiments are conducted for different image dimensions and with two popular classifiers, namely SVM and ELM. Early fusion of feature vectors extracted using two different techniques boosts up the classification results with reduced feature dimensionality. The proposed approach has also outperformed the classification performance with features extracted by the state-of-the-art techniques such as HOG and SIFT. The framework presented in this work is useful for real time applications that call for resourceful management of image data and its content based classification.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Miao RH, Tang JL, Chen XQ (2015) Classification of farmland images based on color features. J Vis Commun Image Represent 29:138–146

2. Zand M, Doraisamy S, Halin AA, Mustaffa MR (2015) Texture classification and discrimination for region-based image retrieval. J Vis Commun Image Represent 26:305–316

3. Ping Tian D (2013) A review on image feature extraction and representation techniques. Int J Multimed Ubiquitous Eng 8(4):385–396

4. Lin Y, Lv F, Zhu S, Yang M, Cour T, Yu K, Huang T (2011) Large-scale image classification: fast feature extraction and SVM training. In: 2011 IEEE conference on computer vision and pattern recognition (CVPR). IEEE, pp 1689–1696

5. Ekenel HK, Stiefelhagen R (2006) Block selection in the local appearance-based face recognition scheme. In: Conference on computer vision and pattern recognition workshop, 2006. CVPRW'06. IEEE, pp 43–43

6. Camlica Z, Tizhoosh HR, Khalvati F (2015) Autoencoding the retrieval relevance of medical images. In: 2015 International conference on image processing theory, tools and applications (IPTA). IEEE, pp 550–555

7. Topi M, Timo O, Matti P, Maricor S (2000) Robust texture classification by subsets of local binary patterns. In: Proceedings of the 15th international conference on pattern recognition, 2000, vol 3. IEEE, pp 935–938

8. Swain MJ, Ballard DH (1991) Color indexing. Int J Comput Vision 7(1):11–32

9. Walia E, Pal A (2014) Fusion framework for effective color image retrieval. J Vis Commun Image Represent 25(6):1335–1348

10. Dalal N, Triggs B (2005) Histograms of oriented gradients for human detection. In: IEEE computer society conference on computer vision and pattern recognition, 2005. CVPR 2005, vol 1. IEEE, pp 886–893

11. Lowe DG (1999) Object recognition from local scale-invariant features. In: The proceedings of the seventh IEEE international conference on computer vision, 1999, vol 2. IEEE, pp 1150–1157

12. Dunham MH (2006) Data mining: introductory and advanced topics. Pearson Education India, Bengaluru

13. Kotsiantis SB, Zaharakis I, Pintelas P (2007) Supervised machine learning: a review of classification techniques. Informatica 31:249–268

14. Weise T, Chiong R (2010) Evolutionary data mining approaches for rule-based and tree-based classifiers. In: 2010 9th IEEE international conference on cognitive informatics (ICCI). IEEE, pp 696–703

15. Dasgupta S, Kalai AT, Monteleoni C (2009) Analysis of perceptron-based active learning. J Mach Learn Res 10(Feb):281–299

16. Huang G, Huang GB, Song S, You K (2015) Trends in extreme learning machines: a review. Neural Netw 61:32–48

17. Huang X, Shi L, Suykens JA (2014) Support vector machine classifier with pinball loss. IEEE Trans Pattern Anal Mach Intell 36(5):984–997

18. Zhou XS, Huang TS (2000) CBIR: from low-level features to high-level semantics. In: Proceedings of SPIE—the international society for optical engineering, vol 3974, pp 426–431

19. Talib A, Mahmuddin M, Husni H, George LE (2013) A weighted dominant color descriptor for content-based image retrieval. J Vis Commun Image Represent 24(3):345–360

20. Brunelli R, Mich O (2001) Histograms analysis for image retrieval. Pattern Recogn 34(8):1625–1637

21. Rasheed W, An Y, Pan S, Jeong I, Park J, Kang J (2008) Image retrieval using maximum frequency of local histogram based color correlogram. In: International conference on multimedia and ubiquitous engineering, 2008. MUE 2008. IEEE, pp 62–66

22. Huang J, Kumar SR, Mitra M, Zhu WJ, Zabih R (1997) Image indexing using color correlograms. In: Proceedings of the 1997 IEEE computer society conference on computer vision and pattern recognition, 1997. IEEE, pp 762–768

23. Lu TC, Chang CC (2007) Color image retrieval technique based on color features and image bitmap. Inf Process Manag 43(2):461–472

24. Wang XY, Yu YJ, Yang HY (2011) An effective image retrieval scheme using color, texture and shape features. Comput Stand Interfaces 33(1):59–68

25. Park DK, Jeon YS, Won CS (2000) Efficient use of local edge histogram descriptor. In: Proceedings of the 2000 ACM workshops on multimedia. ACM, pp 51–54

26. Yuan LUO, Wu CM, Zhang Y (2013) Facial expression feature extraction using hybrid PCA and LBP. J China Univ Posts Telecommun 20(2):120–124

27. Wang W, Huang Y, Wang Y, Wang L (2014) Generalized autoencoder: a neural network framework for dimensionality reduction. In: Proceedings of the IEEE conference on computer vision and pattern recognition workshops, pp 490–497

28. Bengio Y, Courville A, Vincent P (2013) Representation learning: a review and new perspectives. IEEE Trans Pattern Anal Mach Intell 35(8):1798–1828

29. Zhu Q, Shyu ML (2015) Sparse linear integration of content and context modalities for semantic concept retrieval. IEEE Trans Emerg Top Comput 3(2):152–160

30. ElAlami ME (2011) A novel image retrieval model based on the most relevant features. Knowl Based Syst 24(1):23–32

31. Hiremath PS, Pujari J (2007) Content based image retrieval using color, texture and shape features. In: International conference on advanced computing and communications, 2007. ADCOM 2007. IEEE, pp 780–784

32. Yue J, Li Z, Liu L, Fu Z (2011) Content-based image retrieval using color and texture fused features. Math Comput Model 54(3):1121–1127

33. Banerjee M, Kundu MK, Maji P (2009) Content-based image retrieval using visually significant point features. Fuzzy Sets Syst 160(23):3323–3341

34. Jalab HA (2011) Image retrieval system based on color layout descriptor and Gabor filters. In: 2011 IEEE conference on open systems (ICOS). IEEE, pp 32–36

35. Shen GL, Wu XJ (2013) Content based image retrieval by combining color, texture and CENTRIST. IEEE Int Workshop Signal Process 1:1–4

36. Irtaza A, Jaffar MA, Aleisa E, Choi TS (2014) Embedding neural networks for semantic association in content based image retrieval. Multimed Tools Appl 72(2):1911–1931

37. Rahimi M, Moghaddam ME (2015) A content-based image retrieval system based on color ton distribution descriptors. SIViP 9(3):691–704

38. ElAlami ME (2014) A new matching strategy for content based image retrieval system. Appl Soft Comput 14:407–418

39. Subrahmanyam M, Wu QJ, Maheshwari RP, Balasubramanian R (2013) Modified color motif co-occurrence matrix for image indexing and retrieval. Comput Electr Eng 39(3):762–774

40. Walia E, Vesal S, Pal A (2014) An effective and fast hybrid framework for color image retrieval. Sens Imaging 15(1):93

41. Subrahmanyam M, Maheshwari RP, Balasubramanian R (2012) Expert system design using wavelet and color vocabulary trees for image retrieval. Expert Syst Appl 39(5):5104–5114

42. Wang Y, Yao H, Zhao S, Zheng Y (2015) Dimensionality reduction strategy based on auto-encoder. In: Proceedings of the 7th international conference on internet multimedia computing and service. ACM, p 63

43. Liu GH, Yang JY (2013) Content-based image retrieval using color difference histogram. Pattern Recogn 46(1):188–198