IJCSE International Journal of Computer Sciences and Engineering Open Access

Research Paper Vol.-6, Issue-8, Aug 2018 E-ISSN: 2347-2693

Feature Extraction Using Principal Component Analysis and Discrete Wavelet Transform for Image Classification

B. Kumar

Dep. of Computer Science & IT, MJP Rohilkhand University, Bareilly, India.

Corresponding Author:bkumar@mjpru.ac.in, Tel: +91 581 2520310

Available online at: www.ijcseonline.org

Accepted: 15/Aug/2018, Published: 31/Aug/2018

Abstract— Feature extraction is an important part of any image classification scheme. It provides more informative and compact values derived from the original data. In this paper two conventional and widely used techniques known as principal component analysis (PCA) and discrete wavelet transform (DWT) are used for feature extraction. Both techniques are based on entirely different approaches. The results for the two techniques are analyzed and compared. The classification is performed with a benchmark classifier support vector machine. The experiments are carried out on a publically available datasets. The results have shown that DWT has performed better than PCA under the tested scenario.

Keywords— Classification, Feature extraction, Discrete Wavelet Transform (DWT), Principal Component Analysis (PCA).

I. INTRODUCTION

Image classification is an important research area having applications in face recognition, character recognition, medical imaging, remote sensing, forensic, and object detection, etc. A classification tool and feature extraction are two major components in almost all such applications. The classification can be divided into two broad categories: supervised and unsupervised. The supervised classifiers need a set of samples with known labels. This set is called training set. By using training samples, the classifier is first trained through a learning or training process. During training process the classifier learns about the characteristics of the different classes in terms of some parameters. Once the training process is over, the classifier can classify the new unknown samples based on the knowledge it acquired. However, the performance of these classifiers highly depends on the quality and quantity of the training samples. Support vector machine (SVM) [1], random forest, and artificial neural networks, etc. are the well-known classifiers in this category. In the absence of good quality samples, the unsupervised classifiers provide the alternative. The distance based classifiers are prominent in this category. Usually the supervised classifiers provide better accuracy than their unsupervised counterparts.

Each element in digital images has some information mainly related to the colors. Bur for a particular application, not the entire image provide relevant information. The raw image is usually transformed to obtain more information about the color, shape, and texture, etc. about the objects in the image. This process is known as feature extraction. A number of feature extraction techniques are available that work on different aspects of the image to extract different kind of information. In this paper two unsupervised feature extraction techniques discrete wavelet transform (DWT) [2] and principal component analysis (PCA) [3] are analyzed and compared with the help of experimental results. For simplification, the images containing only single object are considered in the experiments.

The remaining paper is organized as follows. In Section II, some research work related to problem is discussed, both feature extraction techniques DWT and PCA are briefly discussed in Section III. The methodology used is `presented in the Section IV, the experimental results are analyzed in Section V, and Section VI concludes the work with remarks on future work.

II. RELATED WORK

Khalil et al. [4] compared three features extraction techniques including gray level co-occurrence matrix (GLCM), histogram of oriented gradients (HOG), and local binary patterns (LBP) for the classification of MR brain images using k-nearest neighbor classifier. They found that HOG performed better than two other techniques for MR brain images. Öztürk and Akdemir [5] compared classification results of GLCM, LBP, local binary GLCM, grey level run length matrix (GLRLM), and segmentation based fractal texture analysis (SFTA) feature extraction techniques using different classifiers. The results for histopathological images illustrated that SFTA is the best among all the tested techniques.

Yue et al. [6] used color features for being rotation, translation, and scale invariant. The RGB color space was converted to HSV color space. They also computed texture using GLCM and fused with color features to develop more robust system. Das et al. [7] reduced the feature set to 12 irrespective of the dimensionality by feature extraction using binarization of the image. Boukharouba and Bennia [8] developed a feature extraction method with Freeman chain code for the images of handwritten characters. The method is independent of the size of the image and therefore no normalization is required. Bala and Kaur [9] used texton XOR patterns to collect image structure to retrieve the query and database images. The operations were performed in HSV color space. Gao et al. [10] proposed quaternion moments for color object classification. Quaternion moments are faster than the moments defined in polar coordinate system. They are implemented in RGB color space without losing original information. These features are good for color as well as grey scale images.

III. FOUNDATION

Usually better results are obtained with extracted features as compared to the raw image. The useful features are extracted by transforming the image with the help of some mathematical techniques. PCA and DWT are two such techniques, which are discussed in this section.

A. Principal Component Analysis

Principal component analysis (PCA) [3] is a statistical technique that can be used as an unsupervised feature extraction method. It diagonalizes the covariance matrix of the image by eigenvalue decomposition. The eigenvalue is a kind of measure of the variance of the variable. The higher eigenvalue indicates the greater contribution to the information. Therefore, only those variables are considered as features that correspond to higher eigenvalues. These orthogonal variables are known as principal components. PCA produces the uncorrelated principal components with decreasing variance. Only those principal components form the feature set that account for the most of the variance. Usually the first few principal components retain the most of the variance.

Let us consider an image with *n* number of observed data elements. The mean value *m* is calculated as follows

$$m = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{1}$$

The mean value *m* used to get covariance matrix *C*

$$C = \frac{1}{n-1}\sum_{i=1}^{n} (x_i - m)\,(x_i - m)^T \tag{2}$$

The following eigenvalue equation is solved to obtain transformation matrix

$$\lambda V = CV \tag{3}$$

Where λ is the matrix of eigenvalues and *V* is the matrix of eigenvectors. The principal components corresponding to some top nonzero eigenvalues form the transformation matrix. With the help of transformation matrix, the desired features are produced.

B. Discrete Wavelet Transform

Wavelets are effective mathematical tools for signal processing in time-frequency domain [11] at different scales, resolutions, and shifts. The wavelet represents a signal in terms of their basis functions known as mother wavelet. It is capable of obtaining time information as well as frequency information. Its digital counterpart is known as discrete wavelet transform (DWT). In DWT, wavelet is represented by a set of low-pass and high-pass filters [9]. For a dyadic DWT, the signal is passed through the filters followed by decimation by a factor of two. Mathematically, the 1-D DWT is expressed as follows

$$W_f^{\Psi}(i,j) = \sum_{x=0}^{n-1} f(x)\Psi_{i,j}^{*}(x) \tag{4}$$

where $W_f^{\Psi}(i,j)$ is the wavelet coefficient of the signal $f(x)$, i and j are the integers representing scale and translation parameters, and Ψ is the mother wavelet.

In a single step wavelet decomposition, the output of high-pass filter is known as wavelet details and output of low-pass filter is called approximation. The process can be performed repeatedly. The number of levels for decomposition depends on the specific signal or image. The use of wavelet detail and approximation depends on the application. There are number of mother wavelets including Haar, Daubechies, and Symmlet, etc. The Haar is one of the most widely used and simple mother wavelet [12, 13] defined as

$$\Psi(x) = \begin{cases} 1, & 0 \le x \le 1/2 \\ -1, & 0 \le x \le 1/2 \\ 0, & \text{otherwise} \end{cases} \tag{5}$$

It resembles the step function. For multi-dimensional DWT, the 1-D DWT is performed multiple times across multiple scales.

IV. METHODOLOGY

The methodology used in this work is illustrated in Fig. 1. It is based on the supervised classification approach. Before classifying the queried image, the classifier needs to be trained. The training process requires a set of images with given labels. Such a set of images with their given labels is known as training set. Once the training process is over, the classifier is ready to predict the labels for new input images. The performance of supervised classifiers heavily bank on the training process. Therefore, it is a crucial step in the given methodology and training set should be prepared carefully with proper representation of all the classes.

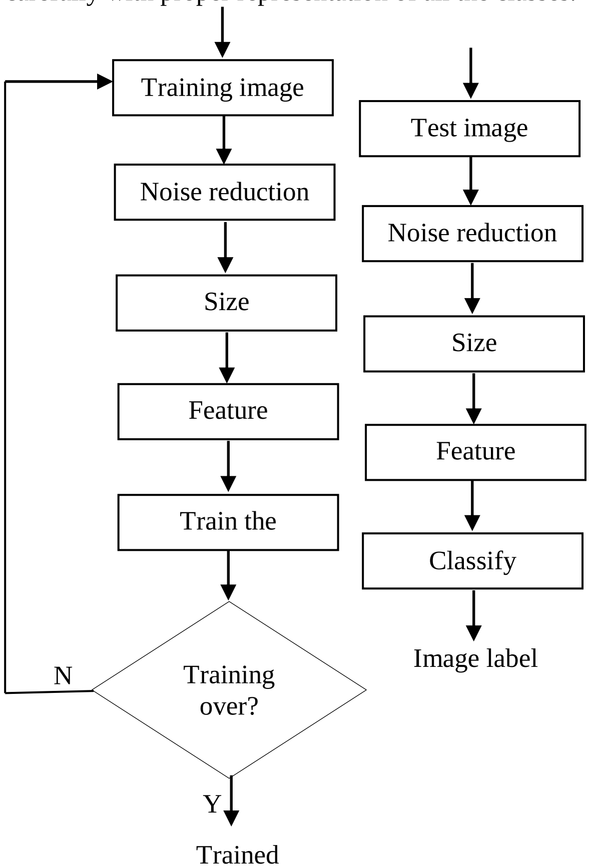

Fig. 1 The classification process: (a) Training (b) Test.

(a) (b)

The methodology therefore can be divided into two parts: training and testing. The first step in both cases is denoising the input image to reduce the undesirable elements. Since different images from different sources may vary in size. Therefore normalization is used to bring all the images to a chosen standard size. In the next step, feature extraction is applied to extract the discriminative information. The extracted feature set is used as input to the classifier both in training and test phases. The trained classifier provides the label for the queried image.

V. Results and Discussion

The experiments are carried out on the images from a publicly available Corel dataset. The sample images are shown in Fig 2. The images are divided in five classes: Gun, Tree, Ship, Dinosaur, and Rose. For each class the 40% samples are for training purpose and remaining 60% samples are used in the classification process. The SVM is used as a classifier. The implementation is done in Matlab and SVM is implemented with Matlab interface of the well-known LIBSVM [14] tool. The results presented in terms of classification accuracy which is determined as the percentage of the correctly classified images during testing phase. The overall classification accuracy (OA) is calculated as follows

$$OA = \frac{No. of\ correctly\ classified\ images}{Total\ no. of\ tested\ image} \times 100 \tag{6}$$

(a)

(b)

(c)

(d)

(e)

Fig 2. Sample images from the dataset: (a) Gun (b) Tree (c) Ship (d) Dinosaur (e) Rose.

(b) (b) (c)

Fig 3. Principal components of an image from class Gun: (a) First (b) Second (c) Third

Fig 4. The components of an image from class Gun obtained by DWT decomposition.

The principal components of an image from class Gun are shown Fig 3 and components obtained by 2-D DWT decomposition are shown in Fig 4. It can be observed that the first principal component contains the highest amount of information. DWT provides four sub-images after one level of decomposition, out of which, the approximate component can be used as a feature.

The classification results are given in Table 1. It can be observed that for the most of the classes, DWT provides better results. Only for class ‘Gun’, the PCA based approach has produced better results. In both cases, the classification accuracies for class ‘Ship’ is lowest. This could be due to more background features. The results for class ‘Dinosaur’ are best.

Table 1. Classification accuracies (%) in terms of OA.

Class	DWT	PCA
Bonsai	64.12	62.88
Gun	64.56	65.41
Ship	64.14	60.67
Dinosaur	67.75	66.42
Rose	66.37	65.16

Overall, it is observed that the classification accuracy is better for those objects that do not have many other features or objects in the background.

VI. CONCLUSION

In this work, PCA and DWT techniques were used for image feature extraction for the purpose of the classification. The SVM was used as a classifier. The results for both feature extraction techniques were compared and analyzed in terms of classification accuracy. It is found that for the given methodology, the DWT based approach produced better results than PCA. Both techniques resulted in better results for those images that do not have many features in background. Therefore, the future work includes the removal of background from the image before object classification.

REFERENCES

[1] G. Mountrakis, J. Im, and C. Ogole, “*Support vector machines in remote sensing: A review*,” ISPRS Journal of Photogrammetry and Remote Sensing, vol. 66, pp. 247–259, 2011.

[2] L.M. Bruce, C.H. Koger, and J. Li, “Dimensionality reduction of hyperspectral data using discrete wavelet transform feature extraction,” IEEE Transactions on Geoscience and Remote Sensing, vol. 40, no. 10, pp. 2331-2338, 2002.

[3] Joliffe, I., "*Principal Component Analysis*," Springer-Verlag, New York, 2002.

[4] M. Khalil, H. Ayad, and A. Adib, "*Performance evaluation of feature extraction techniques in MR-Brain image classification system*," Procedia Computer Science, vol. 127, pp. 218-225, 2018.

[5] Ş. Öztürk and B. Akdemir, "Application of feature extraction and classification methods for histopathological image using GLCM, LBP, LBGLCM, GLRLM and SFTA," Procedia Computer Science, vol. 132, pp. 40-46, 2018.

[6] J. Yue, Z. Li, L. Liu, and Z. Fu, "*Content-based image retrieval using color and texture fused features*," Mathematical and Computer Modelling, vol. 54, pp. 1121-1127, 2011.

[7] R. Das, S. Thepade, S. Bhattacharya, and S. Ghosh, "Retrieval architecture with classified query for content based image recognition," Applied Computional Intelligence and Soft Computing,, vol. 2016, pp. 1-9, 2016.

[8] A. Boukharouba and A. Bennia, "Novel feature extraction technique for the recognition of handwritten digits," Applied Computing and Informatics, vol. 13, pp. 19-26, 2015.

[9] A. Bala and T. Kaur, "Local texton XOR patterns: A new feature descriptor for content based image retrieval," Engineering Science and Technology, An International Journal, vol. 19, pp. 101-112, 2016.

[10] L. Guo, M. Dai, and M. Zhu, "Quaternion moment and its invriants for color object classification," Information Science, vol. 273, pp. 132-143, 2014.

[11] Y. Qian, M. Ye, and J. Zhou, "Hyperspectral image classification based on structured sparse logistic regression and three-dimensional wavelet texture features," IEEE Transactions On Geoscience And Remote Sensing, vol. 51, no. 4, pp. 2276-2291, 2012.

[12] J. Li, "Wavelet-based feature extraction for improved endmember abundance estimation in linear unmixing of hyperspectral signals," IEEE Transactions On Geoscience And Remote Sensing, vol. 42, no. 3, pp. 644-649, 2004.

[13] L. M. Bruce, C. H. Koger, and J. Li, "Dimensionality reduction of hyperspectral data using discrete wavelet transform feature extraction," IEEE Transactions nn Geoscience And Remote Sensing, vol. 40, no. 10, pp. 2331-2338, 2002.

[14] Chang, C.-C., and Lin, C.-J., "*LIBSVM: a library for support vector machines,"* ACM Transactions on Intelligent Systems and Technology, vol. 2, no. 3, pp. 27:1-27:27, 2011.

Author Profile

B. Kumar is MTech in Computer Science & Technology and PhD. He is working as Associate Professor in the Department of Computer Science & Information Technology at MJP Rohilkhand University, Bareilly. He is a member of IEEE. He has published several research papers in reputed international journals including Thomson Reuters (SCI & Web of Science) and conferences including ACM, IEEE, and Springer. His research interests include wireless networks, image processing, pattern recognition, and remote sensing.

