Applied Mathematics & Information Sciences
An International Journal

http://dx.doi.org/10.18576/amis/140418

Adaptive Multimodal Feature Fusion for Content-Based Image Classification and Retrieval

*Samy Bakheet**, *Mahmoud Mofaddel, Emadedeen Soliman and Mohamed Heshmat*

Faculty of Computers and Information (FCI), Sohag University, P. O. Box 82524 Sohag, Egypt

Received: 2 Feb. 2020, Revised: 23 Apr. 2020, Accepted: 4 May 2020
Published online: 1 Jul. 2020

Abstract: Content-Based Image Retrieval (CBIR) is a potential application of computer vision to the image retrieval problem to search images from large-scale image databases according to a user's request in terms of a query image. Semantic gap remains an endemic and awkward challenge for the development of high accuracy CBIR systems. It arises due to the inherent difference between the digital representation of images by machine and high-level semantic concepts of images. In this paper, we introduce an adaptive feature fusion framework for Content-Based Image Classification and Retrieval (CBICR) based on stacked random forests for feature fusion, where salient multimodal features, including low-level visual features (e.g., color, edge histogram, Hu moments, etc.) are automatically extracted from image regions and adaptively fused together. Then, a particular sampling and classification mechanisms of Random Forests are exploited to adaptively fuse the utilized features together. To assess the effectiveness of the proposed method, various experiments are carried out on a large scale dataset of real and synthetic images. The results demonstrate desirable performance of the proposed method in terms of efficiency, effectiveness, and robustness.

Keywords: CBICR, Multi-modal feature fusion, Random Forests

1 Introduction

In the past two decades or so, with the high performance and the large capacities of smartphones, the number of images over the internet has increased rapidly. With this large scale of images, the world is in need of a robust way to classify or retrieve images based on its content. Many methods and approaches have been developed to produce efficient CBICR systems [1,2]. Sarwar et al. [3] presented a novel method to make a Bag Of Visual Words (BOVW) more robust via support vector machine learning technique for CBVIR system. Karakasis et al. [4] used affine moment invariants with BOVW to improve CBVIR system based on local features. Pradhan et al. [5] developed a three-level hierarchical CBIR framework from color and shape features, based on an adaptive Tetrolet Transform and new feature histograms.

Random forest approach has recently had a wide area in image classification. Large varieties of methods have been improved to propose a random forest model for classification [6,7]. Random forest is a classifier consisting of groups of multiple decision tree using ensemble classification method [8]. An ensemble is a finite collection of models used to obtain average predictive accuracy rather than using any single model in the ensemble collection [9,10]. Random forest works on a set of data such as image feature vectors.

Multi-feature types are usually fused in a single feature vector for learning or to quantify an image for content-based visual image retrieval [11]. In recent years, a lot of methods have been proposed in computer vision and image classification applications, which adopt multi-feature fusion [12,13,14]. Hu et al. [15] presented a new fusion of multi-features by mixing texture, shape and color with a diffusion process to improve the original distance matrix for content-based image retrieval based on multi-feature fusion. Paul et al. [16] proposed an improved method for classification. where the authors opted on removing unimportant features to reduce the number of used decision trees. Kulkarni et al. [17] used a Random Forest Classifier (RFC) to classify pixels in Landsat Thematic Mapper imagery to design an algorithm for land cover classification. The smart choice of image feature vectors that we make in the paper can lead to high accurate classification results.

The remainder of the paper is organized as follows. we presents an overview of Decision Tree (DT) related

* Corresponding author e-mail: samy.bakheet@fci.sohag.edu.eg

classification models including the Random Forest (RF) in Section 2. Section 3 gives a detailed description of the proposed CBICR framework, while experimental results are presented and discussed in Section 4. Section 5 is devoted to conclusion and further research.

2 Decisions Tree and Random Forests

Decision tree is highly powerful machine learning and classification tool, which is built on a binary tree. Therefore, it can perform the classification task fast, in time O(log N). Fig. 1 shows an example of a decision tree to classify an image based on its color channel value.

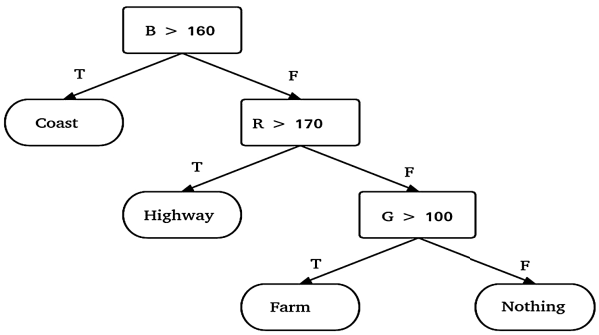

Fig. 1 A simple example of decision tree

Random forest which consists of multiple decision trees is an even more accurate classifier. Fig. 2 shows an example for Multi-decision tree. Any RFC starts with constructing multiple decision trees connected together at the root of the forest which used to make predictions where each decision tree votes on its prediction final classification. These votes are tabulated label classifier and the label with the most votes is for the final classification [8]. Random forests are built on the ensemble method that uses Jensenś Inequality [18] which states that the convex combined (average) ensemble will have an error less than or equal to the average error of the individual models.

3 Proposed Methodology

In this section, the proposed methodology is described in details, which is established with two main purposes, i.e. image classification and content-based image retrieval. In the first part of the section, we describe the image classification module, while the CBIR module will be explained at the end of the section.

3.1 Image Classification

Machine learning algorithms, such as Random Forests (RFs), have demonstrated excellent performance for solving various data classification and retrieval tasks. The algorithmic procedure of the proposed method is depicted in Fig. 3, where the algorithm loops over each image in the dataset to collect all labels and extract features for training and testing. As features are more convenient with a dataset, as classification become more accurate. Fusing multi-modal features is seen as one of the most active hacks in machine learning algorithms [19,20]. It should be noted that selecting appropriate features is highly dependent on the underlying dataset [21,22]. In our method, three types of features are extracted and fused together, i.e. color, Haralick texture and shape features, where the Color Channel Statistics (CCS) and Color Channel Histogram (CCH) are extracted from five regions of the image.

Color channel statistics: HSV color space defines colors with three particular channels: Hue (H), Saturation(S) and Value (V). The extracted features include the mean and standard deviation of the intensity distribution of the three channels. This generates a feature vector of $2 \times 3 = 6$ dimensions, which is indeed of significantly low dimension and contains only the most discriminative information [23].

Haralick texture features: These features are used to describe the texture of an image. Haralick et al. [24] proposed the Gray Level Cooccurrence Matrix (GLCM) and 14 statistical measures of textural features which were used for robust classification of rocks, but they have been recently deployed in numerous feature representation applications, such as electronic shopes and medical images [25]. Extraction process of these features involves two main steps. The first is to compute GLCM, and the features are extracted in the second step, where the matrix quantifies texture by counting how many pairs of adjacent pixels with specific values appear in an image as shown in Fig. 4, while the direction of adjacency is represented as in Fig. 5. These features are simply computed from the GLCM, such as contrast, homogeneity, entropy, etc. Texture features are a smart selection to quantify and represent touch, appearance, or consistency of a surface. Textures are always used to compare roughness or smoothness [26,27,28]. This may encourage us to use Haralick texture features to represent the distinctions between rocks, sand, and bricks in the used dataset. The GLCM matrix is defined as follows

$$G = \begin{bmatrix} P(1,1) & P(1,2) & \dots & P(1,N_g) \\ P(2,1) & P(2,2) & \dots & P(2,N_g) \\ \vdots & \vdots & \ddots & \vdots \\ P(N_g,1) & P(N_g,2) & \dots & P(N_g,N_g) \end{bmatrix} \quad (1)$$

where $N_g \times N_g$ is the size of the square GLCM matrix. The Haralick features used in our work are presented from Eqs. (2) to (14) [24]. Angular Second Moment

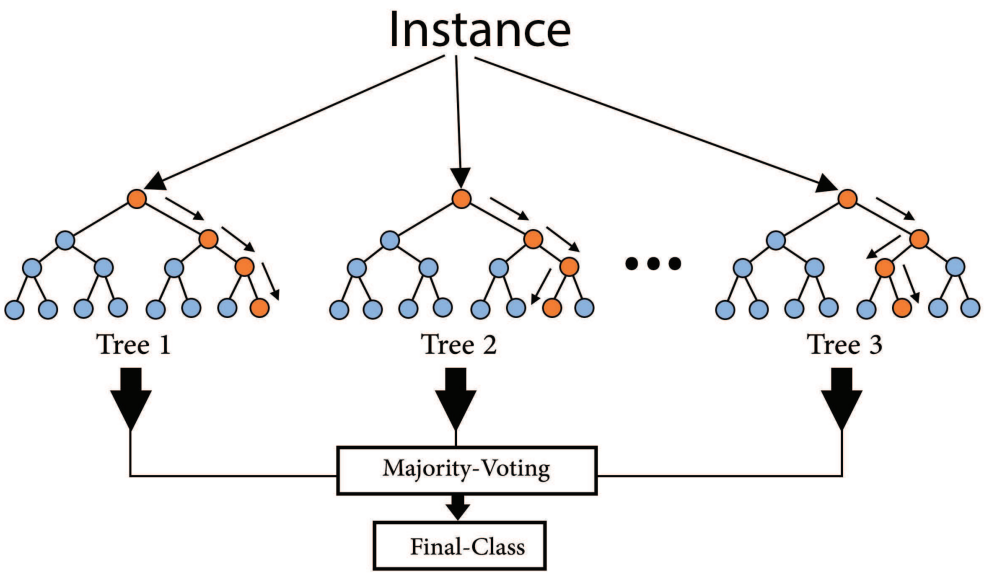

Fig. 2 Classification using decision trees in the Random Forest

(ASM) is defined as:

$$ASM = \sum_i \sum_j P(i,j)^2 \tag{2}$$

Homogeneity or Inverse Deference Moment (IDM) that reflects decreases in ASM is given by

$$IDM = \sum_i \sum_j \frac{1}{1+(i-j)^2} P(i,j) \tag{3}$$

Entropy (E) as a measure of randomness of intensity image is defined as:

$$E = -\sum_i \sum_j P(i,j) \log(P(i,j)) \tag{4}$$

Sum Entropy (S_E) is given as:

$$S_E = -\sum_{i=2}^{2N_g} P_{x+y}(i) \log\{P_{x+y}(i)\} \tag{5}$$

Difference Entropy (D_E):

$$D_E = \sum_{i=0}^{N_g-1} P_{x-y}(i) \log\{P_{x-y}(i)\} \tag{6}$$

Correlation (C) that define the relations between pairs of variables is given as:

$$C = \frac{\sum_i \sum_j (ij) P(i,j) - \mu_x \mu_y}{\sigma_x \sigma_y} \tag{7}$$

where μ_x, μ_y, σ_x and σ_y are the means and standard deviations of P_x and P_y, respectively. Contrast (Ç) is the gradient of color that distinguishes between objects is defined as:

$$(Ç) = \sum_{n=0}^{N_g-1} n^2 \{ \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} P(i,j) \}, |i-j| = n \tag{8}$$

Sum of Squares is the Variance (σ^2):

$$\sigma^2 = \sum_i \sum_j (i-\mu)^2 P(i,j) \tag{9}$$

Sum Variance (S_{σ^2}):

$$S_{\sigma^2} = \sum_{i=2}^{2N_g} (i - S_E)^2 P_{x+y}(i) \tag{10}$$

Difference Variance (D_{σ^2}):

$$D_{\sigma^2} = \sum_{i=0}^{N_g-1} i^2 P_{x-y}(i) \tag{11}$$

Sum Average (S_A):

$$S_A = \sum_{i=2}^{2N_g} i P_{x+y}(i) \tag{12}$$

Information Measures of Correlation 1 ($IMoC_1$):

$$IMoC_1 = \frac{HXY - HXY1}{max\{HX, HY\}} \tag{13}$$

Information Measures of Correlation 2 ($IMoC_2$):

$$IMoC_2 = (1 - exp[-2.0(HXY2 - HXY)])^{\frac{1}{2}} \tag{14}$$

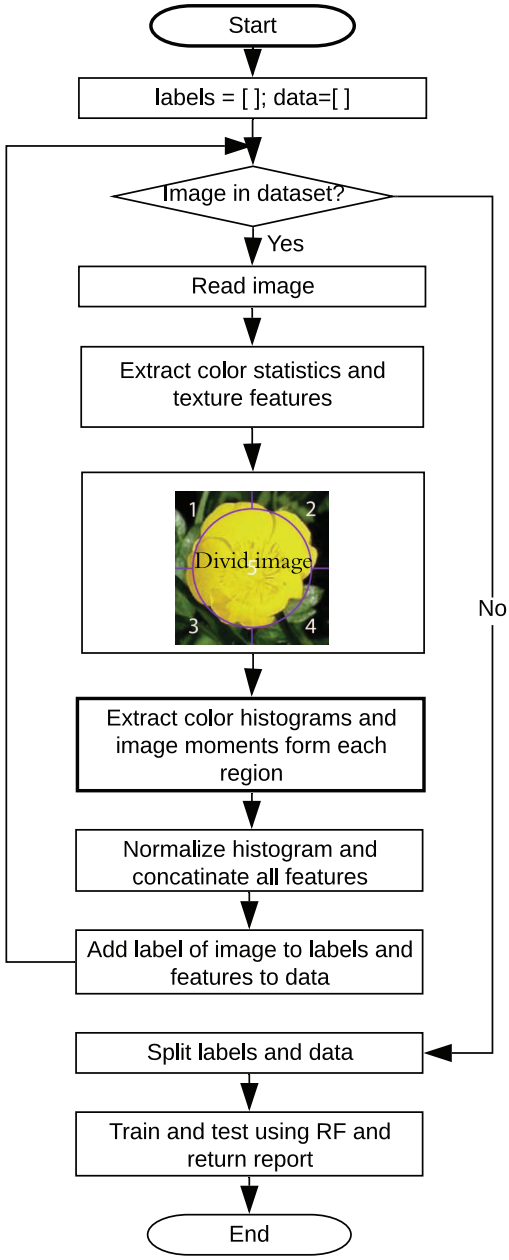

Fig. 3 A flowchart for the proposed CBICR method

where HY and HX are the entropies of P_x and P_y, respectively:

$$HXY = -\sum_i \sum_j P(i,j) log(P(i,j)) \quad (15)$$

$$HXY1 = -\sum_i \sum_j P(i,j) log\{P_x(i)P_y(j)\} \quad (16)$$

$$HXY2 = -\sum_i \sum_j P_x(i)P_y(j) log\{P_x(i)P_y(j)\} \quad (17)$$

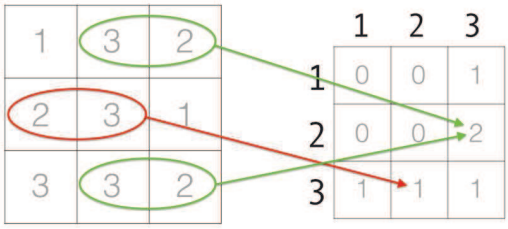

Fig. 4 Sample computation of GLCM matrix from an input image

Fig. 5 Four directional neighborhood of a pixel (i.e. left to right, top to bottom, top-left to bottom-right, and top-right to bottom-left)

Dividing image: The major challenge with any image classification algorithm is that each pixel in an image has its special properties of position and color value. If we can determine the position with the values of some pixels, it surely conveys most information related to the contents of the image [29,30]. To overcome this challenge, Hu moments and color histogram with an appropriate number of bins for each color channel are extracted from five different regions of the image, and then concatenated with other features. An example of an image divided into five regions of interest is shown in Fig. 3.

HSV color histogram: Histograms are usually used to give a rough sense of the density of pixel intensities in an image. It is important to carefully select the number of bins for the histogram descriptors using an appropriate color space [31,32]. In our experiments, the five regions of interest in the image are modeled using their HSV histograms with 5 bins for the Hue channel, 2 bins for the saturation channel, and 3 bins for the value channel. This generates a feature vector of $5 \times 2 \times 3 \times 5 = 150$ dimensions.

Shape features: Moments features are often used to quantify shape of an object in an image [4]. In 1962, Ming-Kuei Hu [33] proposed seven moment features to specify the shape of an object in an image. These moments are invariant to changes in rotation, translation, and scaling. The seven-moment features expressed by Eqs. (18) to (24) are computed on the selected five

regions of interest (ROIs) in a given image to produce a feature vector of $7 \times 5 = 35$ dimensions.

$$\eta_1 = (\mu_{20} + \mu_{02}) \tag{18}$$

$$\eta_2 = (\mu_{20} - \mu_{02})^2 + 4\mu_{11}^2 \tag{19}$$

$$\eta_3 = (\mu_{30} + 3\mu_{12})^2 + (3\mu_{21} - \mu_{30})^2 \tag{20}$$

$$\eta_4 = (\mu_{30} + \mu_{12})^2 + (\mu_{21} + \mu_{03})^2 \tag{21}$$

$$\begin{aligned}\eta_5 = {} & (\mu_{30} - 3\mu_{12})(\mu_{30} + \mu_{12})((\mu_{30} + \mu_{12})^2 \\ & - 3(\mu_{21} + \mu_{03})^2) + (3\mu_{21} - \mu_{03})(\mu_{21} + \mu_{03}) \\ & (3(\mu_{30} + \mu_{12})^2 - (\mu_{21} + \mu_{03})^2)\end{aligned} \tag{22}$$

$$\begin{aligned}\eta_6 = {} & (\mu_{20} - \mu_{02})((\mu_{30} + \mu_{12})^2 - (\mu_{21} + \mu_{03})^2) \\ & + 4\mu_{11}(\mu_{30} + 3\mu_{12})(\mu_{21} + \mu_{03})\end{aligned} \tag{23}$$

$$\begin{aligned}\eta_7 = {} & (3\mu_{21} - \mu_{03})(\mu_{30} + \mu_{12})((\mu_{30} + \mu_{12})^2 \\ & - 3(\mu_{21} + \mu_{03})^2) - (\mu_{30} - 3\mu_{12})(\mu_{21} + \mu_{03}) \\ & (3(\mu_{30} + \mu_{12})^2 - (\mu_{21} + \mu_{03})^2)\end{aligned} \tag{24}$$

where μ_{pq} is the (p+q) order central geometric moments of an image $f(x,y)$ and calculated by

$$\mu_{pq} = \sum_x \sum_y (x - \bar{x})^p (y - \bar{y})^q I(x,y) \tag{25}$$

while $\bar{x}$ and $\bar{y}$ are the -x and -y coordinates of the centroid of the shape and are calculated by

$$\bar{x} = \frac{m_{10}}{m_{00}}, \quad \bar{y} = \frac{m_{01}}{m_{00}} \tag{26}$$

where η_{pq} is the (p+q)th order geometric moments of an image $f(x,y)$, and is calculated by

$$\eta_{pq} = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x^p y^q f(x,y) dx dy \tag{27}$$

To propose a single feature vector representing each image in the dataset, all features are concatenated in a single feature vector of dimensions: $6 + 13 + 150 + 35 = 204$. The next step is to split labels and features for training and testing using an appropriate number of decision trees for the RFC. It is strongly recommended to save the returned data from random forest model in a separated file for later use to reduce learning time and memory buffer size. After that, test data are used to test and evaluate the accuracy learning ability of the proposed method. The final goal of the second phase is to automatically classify the test images in the dataset into their respective categories. This stage involves all steps related to extracting all features from test images, such as color, Haralick textures, and Hu moments. After that, the feature vector of a given test image is evaluated with the classifier model built in the training phase to predict the class label of the test image.

3.2 Content-based Image Retrieval

The proposed Content-Based Image Retrieval (CBIR) system is designed based on the previous classification model, which can retrieve all relevant images to a given query image based on the content of an image rather than meta-data [34]. The following two algorithms present the major procedural steps involved in the developed CBIR system. Algorithm 1 is to find relevant images for each category in the dataset. This stage is executed only once and the retrieved relevant images for each category are saved for later use, but the other stage given by Algorithm 2 aims to predict the nearest category for a given query image and all relevant images are then returned, where *categories_Relevants* is a dictionary to record the relevant images and *model* is a pre-trained model able to predict a category for a given image.

Algorithm 1: The CBIR algorithm is to find relevant images for each category using the pre-trained model.

```
categories_Relevants ← {}
foreach category in categories do
    relevants ← [ ]
    foreach image in dataSet do
        image_category = model.predict(image)
        if category == image_category then
            relevants.append(image)
        end
    end
    categories_Relevants[category] ← relevants
end
```

Algorithm 2: This algorithm is to predict a category for the query and retrieve all relevant images for the query image

```
Function Find_Relevants(model,
  categories_Relevants, query_image) :
    category ← model.predict (query_image)
    relevant_images ← categories_Relevants[category]
    return relevant_images
```

3.3 Performance Evaluation

For evaluation purposes, various standard metrics, such as precision, recall, f-score, and accuracy [35] are used to

evaluate performance of the proposed CBICR method. These metrics are given by the following four equations:

$$perisicion = \frac{TP}{TP+FP} \tag{28}$$

$$recall = \frac{TP}{TP+FN} \tag{29}$$

$$f-score = \frac{recall \times precision}{recall + precision} \tag{30}$$

$$accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{31}$$

where True Positive (TP) occurs when the correct category is predicted, while True Negative (TN) occurs when an image is correctly predicted as not belonging to a category. Similarly, False Positive (FP) occurs when an image is incorrectly predicted to be belonging to a category, while False Negative (FN) happens when an image is incorrectly predicted as not belonging to a category.

4 Experimental Results

The proposed method for image classification and retrieval is experimented and evaluated on two large-scale real-world datasets. In the following experiments, two variants of the proposed method are described and evaluated; one with compact features (FRFs) and the other with extended features (MFRFs). In other words, in the first variant of the proposed method, color histograms or moments are not included in constructing features.

4.1 Results on Dataset #1

The proposed method is first experimented and evaluated on a large-scale dataset containing a total of 1680 images from five categories, i.e. Coast, Highway, Forest, Street, and Flower. The first four categories have been collected over the internet[1] and the fifth one is a mix of Folwers17 dataset[2] [36]. Each one of the five categories includes 336 images. In experiments, the available image dataset is divided into two subsets with 75% for training and the remaining 25% for testing the model performance, which are selected randomly from the whole dataset.

Fig. 6 exhibits that the highest precision achieved by the proposed method is 0.85, when using compact features and 99% of image data for learning, while the same precision can be archived when using the extended features and only 20% of images for learning. Table 1 indicates comparative results for the two variants of the proposed method, when using 25% of all images in the dataset as testing data. The results in Fig. 7 highlight the potential capability of the proposed method to accurately classify images in a wide range of illumination conditions.

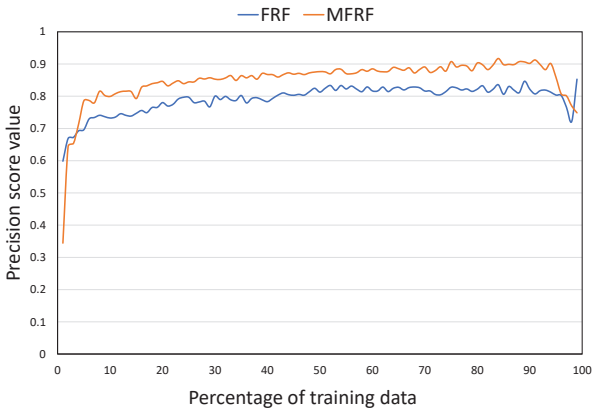

Fig. 6 Average precision versus percentage of learning data methods on dataset #1.

Table 1 Comparative results by the proposed method on dataset #1.

Labels	Precision	
	FRF	MFRF
Coast	0.70	0.80
Forest	0.93	0.95
Highway	0.91	0.98
Flower	0.80	0.98
Street	0.81	0.88
Avrage:	0.83	0.91

4.2 Results on Dataset #2

In this set of experiments, the proposed CBICR method is experimented and evaluated on a challenging dataset comprising roughly 1000 images from 10 categories, i.e. Coast, Forest, Highway, Flower, Street, Car, Airplanes, Face, Guitar, and Motorbike. The first five categories are a subset from the previous experiment, while the rest five categories are from the Clatech101 dataset[3] [37]. The proposed method can perform classification and retrieval tasks robustly, achieving an overall precision of 0.89,

[1] https://gurus.pyimagesearch.com/protected/code/image_classification/dt_and_rf.zip

[2] http://www.robots.ox.ac.uk/ vgg/data/flowers/17/

[3] http://www.vision.caltech.edu/Image_Datasets/Caltech101/

Fig. 7 Examples of correctly classified images from different categories

where HSV color histogram is quantized to $5 \times 4 \times 4$ bins. It is noteworthy that the MFRFs version of the method consistently performs better than its FRFs counterpart, as shown in Fig. 8. Table 2 summarizes the obtained results of classification, where a set of 75% of total images in the dataset was used for training.

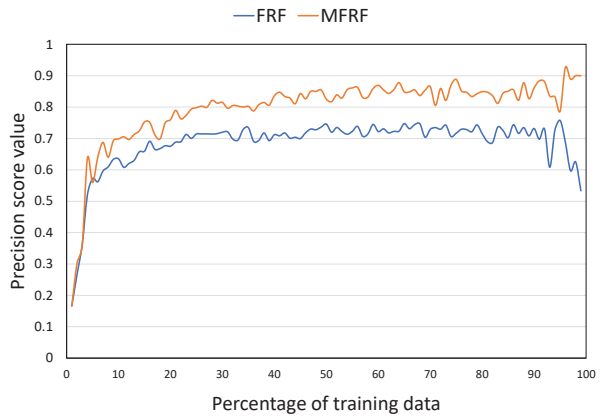

Fig. 8 Average precision versus percentage of learning data methods on dataset #2

Table 2 Comparative results of the proposed method on dataset #2.

Labels	Precision	
	FRF	MFRF
Airplanes	0.90	0.80
Car	0.94	.94
Coast	0.67	0.85
Face	0.84	0.87
Flower	0.58	0.94
Forest	0.77	0.97
Guitar	0.56	0.88
Highway	0.63	0.87
Motorbike	0.86	0.86
Street	0.64	0.84
Avrage:	0.72	0.89

4.3 Results of CBIR System

We tested the CBIR system on dataset #2 of 10 categories. The results are summarized in the Table 3. Fig. 9 shows a sample of successfully retrieved relevant images for a sample of three query images from different categories, where the green-bordered images are the query images and the red-bordered image is irrelevant retrieved image.

The results in Tables 2 and 3 demonstrate that the precision value for the CBIR module is less than that of the classification module. The reason might be due to the mistake occurred in classifying the data. On the other side, the retrieval model predicts the query coast image as

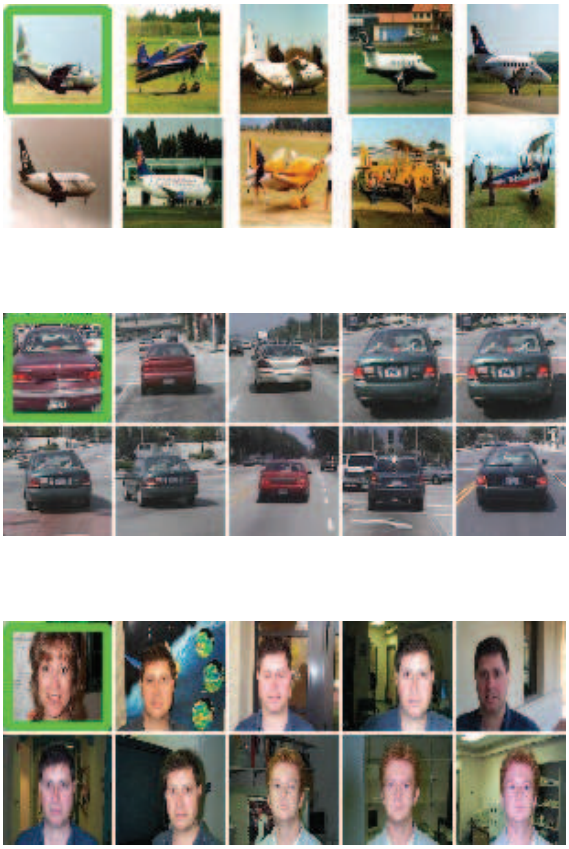

Fig. 9 Retrieval results of different categories; query images are shown at the left-top corners with green borders.

Table 3 Performance evaluation of the CBIR system on dataset #2

Label	Accuracy	Precision	Recall	f-score
airplanes	0.98	0.87	0.93	0.90
cars	1.00	0.94	1.00	0.97
coast	0.95	0.77	0.82	0.79
faces	0.98	0.87	1.00	0.97
flower	0.96	0.79	0.70	0.74
forest	0.98	0.93	0.93	0.93
guitars	0.87	0.52	0.34	0.41
highway	0.95	0.72	0.70	0.70
motorbike	0.96	0.80	0.86	0.83
street	0.96	0.78	0.86	0.82
Average	0.96	0.80	0.82	0.80

a highway image due to the similarity in the color of these two classes (see Fig. 9). All algorithms of the proposed CBICR method have been implemented in Python codes using sklearn package and run on a laptop with Intel core i5-4300U CPU @ 1.90GHz $\times$ 4 with 8GB RAM on Ubuntu 18.04.2 LTS OS 64bit.

5 Conclusions and Future Work

In this paper, an adaptive feature-fusion framework for Content-Based Image Classification and Retrieval (CBICR) based on stacked Random Forests has been presented. In this framework, a set of multi-modal features is conveniently extracted and adaptively fused. The experimental results obtained on a combined large-scale dataset of real and synthetic images have unveiled that the set of extracted multi-modal features does not only have a positive impact on the performance of the proposed CBICR approach, but it also narrows the semantic gap between low-level visual image features and high-level semantic concepts. Future work will focus on further developing and improving the performance of the proposed CBICR framework through application of a bag-of-visual words (BOVW) based approach as a baseline and development of an improved CBICR framework using quantization.

References

[1] S. Sadek, A. Al-Hamadi, B. Michaelis, and U. Sayed, Efficient Region-Based Image Querying, *International Journal of Computing*, **2**, 1–6 (2010).

[2] S. Sadek, A. Al-Hamadi, and B. Michaelis, Toward real-world activity recognition: An SVM based system using fuzzy directional features, *WSEAS Trans Inf Sci Appl*, **10**, 116-127 (2013).

[3] A. Sarwar, Z. Mehmood, T. Saba, K. A. Qazi, A. Adnan, and H. Jamal, A novel method for content-based image retrieval to improve the effectiveness of the bag-of-words model using a support vector machine, *Journal of Information Science*, **45**, 117–135 (2019).

[4] E. G. Karakasis, A. Amanatiadis, A. Gasteratos, and S. A. Chatzichristofis, Image moment invariants as local features for content based image retrieval using the bag-of-visual-words model, *Pattern Recognition Letters*, **55**, 22–27 (2015).

[5] J. Pradhan, S. Kumar, A. K. Pal, and H. Banka, A hierarchical CBIR framework using adaptive tetrolet transform and novel histograms from color and shape features, *Digital Signal Processing*, **82**, 258–281 (2018).

[6] S. Sadek, A. Al-Hamadi, B. Michaelis, and U. Sayed, Image retrieval using cubic splines neural networks, *International Journal of Video and Image Processing and Network Security (IJIPNS)*, **9**, 17–22 (2009).

[7] T. K. Ho, The random subspace method for constructing decision forests, *IEEE transactions on pattern analysis and machine intelligence*, **20**, 832–844 (1998).

[8] L. Breiman, Random forests, Machine learning, **45**, 5–32 (2001).
[9] S. Sadek, A. Al-Hamadi, B. Michaelis, and U. Sayed, *A New Method for Image Classification Based on Multi-level Neural Networks*, in Proc. Proceedings of International Conference on Signal and Image Processing (ICSIP'09), 197-200 (2009).
[10] T. G. Dietterich, *Ensemble methods in machine learning*, in Proc. International workshop on multiple classifier systems, 1–15 (2000).
[11] S. Bakheet and A. Al-Hamadi, Chord-length shape features for license plate character recognition, *J. Russ. Laser Res.*, **41**, 156-170 (2020).
[12] F. Li, J. Wang, R. Lan, Z. Liu, and X. Luo, Hyperspectral image classification using multi-feature fusion, *Optics & Laser Technology*, **110**, 176–183 (2019).
[13] B. Tu, N. Li, L. Fang, D. He, and P. Ghamisi, Hyperspectral Image Classification with Multi-Scale Feature Extraction, *Remote Sensing*, **11**, 534–549 (2019).
[14] K. Li, C. Zou, S. Bu, Y. Liang, J. Zhang, and M. Gong, Multi-modal feature fusion for geographic image annotation, *Pattern Recognition*, **73**, 1–14 (2018).
[15] J. Zhou, X. Liu, W. Liu, and J. Gan, Image retrieval based on effective feature extraction and diffusion process, *Multimedia Tools and Applications*, **78**, 6163–6190 (2019).
[16] A. Paul, D. P. Mukherjee, P. Das, A. Gangopadhyay, A. R. Chintha, and S. Kundu, Improved random forest for classification, *IEEE Transactions on Image Processing*, **27**, 4012–4024 (2018).
[17] A. D. Kulkarni, and B. Lowe, Random forest algorithm for land cover classification, *International Journal on Recent and Innovation Trends in Computing and Communication*, **4**, 58–63 (2016).
[18] M. Kuczma, *Inequalities*, in An introduction to the theory of functional equations and inequalities, 2nd ed., A. Gilányi, Ed. Springer Science: Business Media, Basel· Boston· Berlin, 202–214, (2009).
[19] J. Li, B. Zhang, G. Lu, and D. Zhang, Generative multi-view and multi-feature learning for classification, *Information Fusion*, **45**, 215–226 (2019).
[20] S. Bakheet, A. Al-Hamadi, M. A. Mofaddel, Recognition of human actions based on temporal motion templates, *British Journal of Applied Science & Technology*, **20**, 1-11 (2017).
[21] J. Wu, B. Zhang, J. Zhou, Y. Xiong, B. Gu, and X. Yang, Automatic Recognition of Ripening Tomatoes by Combining Multi-Feature Fusion with a Bi-Layer Classification Strategy for Harvesting Robots, *Sensors*, **19**, 612–633 (2019).
[22] M. A. Mofaddel, S. Bakheet, and R. Youssef, Adaptive Fingerprint Image Enhancement Based On Cascading Filtering, *International Journal of Engineering and Information Systems (IJEAIS)*, **6**, 157–161 (2017).
[23] S. A. Mehre, A. K. Dhara, M. Garg, N. Kalra, N. Khandelwal, and S. Mukhopadhyay, Content-Based Image Retrieval System for Pulmonary Nodules Using Optimal Feature Sets and Class Membership-Based Retrieval, *Journal of digital imaging*, **32**, 362–385 (2019).
[24] R. M. Haralick, K. Shanmugam and I. Dinstein, Textural features for image classification, *IEEE Transactions on systems, man, and cybernetics*, **SMC-3**, 610–621 (1973).
[25] S. Bakheet, An SVM Framework for Malignant Melanoma Detection Based on Optimized HOG Features, *Computation*, **5**, 1–14 (2017).
[26] G. Madaan, Various Approaches of Content Based Image Retrieval Process: A Review, International Journal of Scientific Research in Computer Science, *Engineering and Information Technology*, **3**, 711-716 (2018).
[27] G. Wei, H. Cao, H. Ma, S. Qi, W. Qian, and Z. Ma, Content-based image retrieval for lung nodule classification using texture features and learned distance metric, *Journal of medical systems*, **42**, 13–19 (2018).
[28] R. Marmo, S. Amodio, R. Tagliaferri, V. Ferreri, and G. Longo, Textural identification of carbonate rocks by image processing and neural network: Methodology proposal and examples, *Computers & geosciences*, **31**, 649–659 (2005).
[29] D. Li, and Y. Fang, An algorithm to cluster data for efficient classification of support vector machines, *Expert Systems with Applications*, **34**, 2013–2018 (2008).
[30] S. Bakheet, A fuzzy framework for real-time gesture spotting and recognition, *J. Russ. Laser Res.*, **38**, 61-75 (2017).
[31] K. Hu, X. Niu, S. Liu, Y. Zhang, C. Cao, F. Xiao, W. Yang, and X. Gao, Classification of melanoma based on feature similarity measurement for codebook learning in the bag-of-features model, *Biomedical Signal Processing and Control*, **51**, 200–209 (2019).
[32] M. A. Mofaddel, and S. Sadek, *Adult Image Content Filtering: A Statistical Method Based on Multi-Color Skin Modeling*, in Proc. IEEE International Symposium on Signal Processing and Information Technology (ISSPIT'10), 366-370 (2010).
[33] M. Hu, Visual pattern recognition by moment invariants, *IRE transactions on information theory*, **8**, 179–187 (1962).
[34] N. Jain, and S. S. Salankar, *Content-Based Image Retrieval Using Color and Texture Features Through Ant Colony Optimization*, in Proc. Computing, Communication and Signal Processing, 1029–1037 (2019).
[35] L. Semler, L. Dettori, and J. Furst, *Wavelet-based texture classification of tissues in computed tomography*, in Proc. 18th IEEE Symposium on Computer-Based Medical Systems (CBMS'05), 265–270 (2005).
[36] M-E Nilsback, and A. Zisserman, *A visual vocabulary for flower classification*, in Proc. 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06), 1447–1454 (2006).
[37] K. Grauman, and T. Darrell, *The pyramid match kernel: Discriminative classification with sets of image features*, in Proc. Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1, 1458–1465 (2005).

Samy Bakheet received the doctorate degree (Dr.-Ing.) in Neuro-Information Technology (NIT) from Otto-von-Guericke University Magdeburg, Germany in 2013. His current research interests are geared towards high-level recognition problems in computer vision, such as human activity/event recognition, human pose estimation, object/scene recognition, etc. He has authored more than 50 technical papers in well-reputed refereed journals and international conference/symposium proceedings in the fields of computer vision, pattern recognition, machine learning, medical imaging and robotics.

Mahmoud A. Mofaddel received his PhD degree from Rostock University, Germany in 1999. Currently, he is working as an associate Professor at Faculty of Computers and Information, Sohag University, Egypt. He has authored/co-authored more than 23 scientific papers. His research interests include high performance computing, and image processing.

Mohamed Heshmat received his B.Sc. and M.Sc. degrees from south valley University, Sohag branch, Sohag, Egypt in 2002 and from Sohag University, Sohag, Egypt, 2010, respectively, and his Ph.D. degree from Sohag University, Sohag, Egypt and Bauhaus-University, Weimar, Germany. His research interests include computer vision, 3D data acquisition, object reconstruction, image segmentation, image enhancement, image recognition.

Emadedeen Soliman received his MS degree in Computer Science from Sohag University, Egypt in 2014. His research interests include various multimedia related applications, such as visual surveillance, content-based image retrieval (CBIR), video summarization, and semantic video annotation.