Content-Based Image Retrieval Using Hybrid Feature Extraction Techniques

B. Akshaya, S. Sruthi Sri, A. Niranjana Sathish, K. Shobika, R. Karthika and Latha Parameswaran

Abstract Images consist of visual components such as color, shape, and texture. These components stand as the primary basis with which images are distinguished. A content-based image retrieval system extracts these primary features of an image and checks the similarity of the extracted features with those of the image given by the user. A group of images similar to the query image fed is obtained as a result. This paper proposes a new methodology for image retrieval using the local descriptors of an image in combination with one another. HSV histogram, Color moments, Color auto correlogram, Histogram of Oriented Gradients, and Wavelet transform are used to form the feature descriptor. In this work, it is found that a combination of all these features produces promising results that supersede previous research. Supervised learning algorithm, SVM is used for classification of the images. Wang dataset is used to evaluate the proposed system.

B. Akshaya · S. Sruthi Sri · A. Niranjana Sathish · K. Shobika · R. Karthika (✉)
Department of Electronics and Communication Engineering, Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Coimbatore, India
e-mail: r_karthika@cb.amrita.edu

B. Akshaya
e-mail: akshayabalan96@gmail.com

S. Sruthi Sri
e-mail: shruthisridhar.10@gmail.com

A. Niranjana Sathish
e-mail: niranjana.sathish97@gmail.com

K. Shobika
e-mail: shobi2197@gmail.com

L. Parameswaran
Department of Computer Science and Engineering, Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Coimbatore, India
e-mail: p_latha@cb.amrita.edu

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_58

1 Introduction

With a splurge of visual data, it is a vastly cumbersome task to scan tens of thousands of images manually. The content-based image retrieval system extracts the basic features of every image in a dataset and compares the same with those of the image provided by the user. General flow states that post the query matching, the system ranks the images in a descending order of similarity with the given input query and the output is all the images that are ranked highest. Every image will have three basic components: Shape, Color, and Texture. A CBIR relies on these extracted features, individually or combinations of them, to extract and run similarity algorithms on them.

While CBIR completely relies on the contents of the image itself, image retrieval using the metadata of images can also be done. There is textual data associated with each image and traditional methods of retrieval, such as retrieval using keywords, can be done. While the annotation process is time consuming and laborious, there is also an additional limitation with this system which pertains to lack of standardization. Meaning, no two users perceive an image in the same way. Since there is no standard way to perceive an image, multiple users will give multiple annotations which are not likely to match. This is will also result in large amounts of junk data which is undesirable. Thus, CBIR is more suitable for large amounts of visual data.

In this paper, a combination of color, texture, and shape features are extracted in order to maximize the accuracy of extraction and efficacy of output. A proposed method uses best of multiple techniques to improve quality of output and reduce error margins.

In order to enhance the results of a robust dataset (Wang dataset), the images are classified by Supervised Vector Machines (SVM) algorithm. SVM is the most efficient supervised learning algorithm for image recognition, face recognition, speech recognition, and face detection. It is reliable, accurate and is most efficient for binary classification.

2 Literature Survey

The properties of Hue, Saturation, and Intensity values color space is analyzed in [1]. The values are varied and the visual perception is studied. The saturation value is used to decide if the Hue or the Intensity of the pixel is closer to human perception.

Various CBIR tools are compared in [2]. From this comparison, it is observed that most of the systems use color and texture features. Shape feature is not as common. Layout feature is very rarely used. Retrieval techniques based on a single feature worked well only for a specific set of images.

One of the most commonly used color feature in CBIR system is color histogram, [3–5]. Color Histogram concentrates only on the proportion of the number of various types of colors in an image, but does not focus on the spatial location of the colors.

Noise is not handled efficiently by histograms because they are very sparse. To overcome the drawback of this feature, features such as color-correlogram and color moments are applied. Preprocessing the images will increase the accuracy.

In [6], it is observed that color and texture features are used. Support Vector Machine (SVM) and Euclidean distance are applied to retrieve similar images.

Different approaches of different combinations of color, shape and texture retrieval are compared in [7]. When color (color histogram) and texture features (standard wavelet) are combined, accuracy was enhanced but the feature set was inadequate. On merging color, texture and shape feature (Color moment, Gabor filter, Gradient Vector Flow), the strong feature set was created.

Ecosembles is formed by concatenating different combinations of weak feature views (color, shape, texture, etc.), which must be extracted from images. Histograms are extracted from a group of images with varying numbers of bins for each histogram. By examining the Intelligence, Surveillance, and Reconnaissance data, it is understood that Ecosembles performed slightly better than GIST descriptors coupled with SVM. This is observed in [8].

In [9], images are retrieved separately using the features like color Histogram, Gabor and wavelet transform for texture, and Shape information from Phase congruency (edge detection for any change in illumination and contrast in the image). A combination of these produced an accuracy of 96.4%.

A system for biometric security for CBIR is developed based on the extraction of Shape (moment invariant), Color (Histogram), and Texture (Gabor wavelet) in [10].

The color and texture features are concatenated where Wavelet-Based Color Histogram (WBCH) method is used. The precision of this proposed method is found to be better. The computational steps are reduced with the help of wavelet transform, thereby increasing the retrieval speed [11].

In [12], a relative study on several features like merged color histogram and Gabor transform is performed.

With only statistical entities of the first order such as mean and standard deviation, Gabor wavelet showed better classification results. This method is proven to be slightly superior to the co-occurrence matrix which is usually used for texture classification [13].

A CBIR system in which the features like HOG, SIFT, SURF, and color histogram are used to extract the features of the image and formed a collection of local feature vectors is observed in [14].

3 Proposed Work

The primary goal of the proposed work is to retrieve similar images from a database as a response to passing a single query image. Low-level features of images are used in this approach. The prime motive of this work is to obtain an efficient and less complex image retrieval system. To achieve this, a comparative study on image retrieval using different features is performed. Initially, images are retrieved with

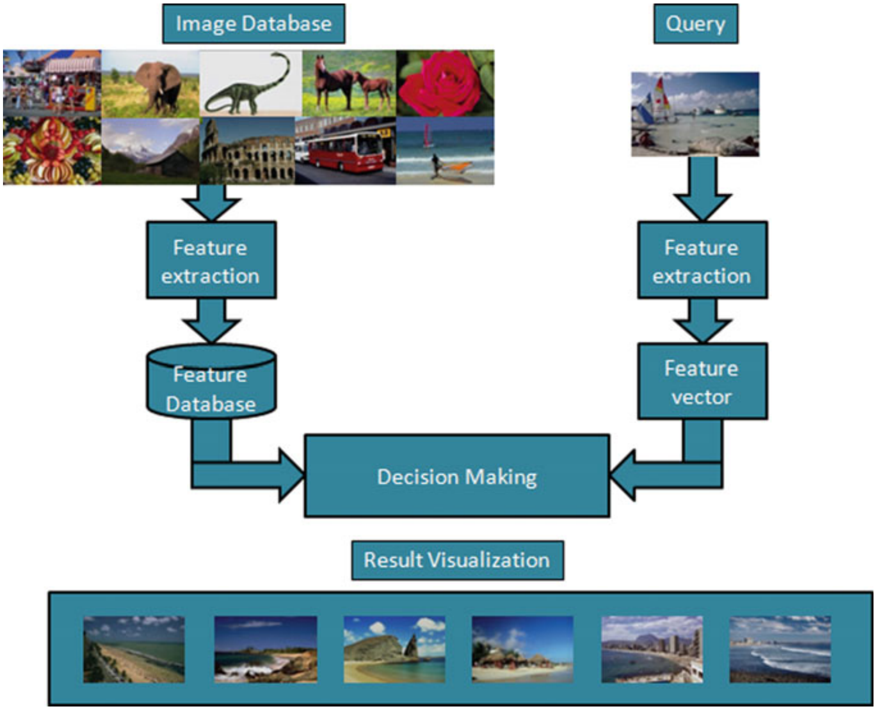

Fig. 1 CBIR system architecture

color, shape and texture features separately. Further, all three features are combined and results are compared.

In this approach, Support Vector Machine (SVM) algorithm is used for classification. The dataset used here is Wang dataset [15], which consists of 1000 images consisting of 10 different classes. Each class has 100 images. The classes include Africa, Rose, Beach, Monument, Bus, Horse, Food, Dinosaur, Scenery, and Elephant. The sizes of the images are either 384×256 or 256×384.

The schematic representation of the CBIR is shown in Fig. 1.

4 Feature Extraction

This is the most crucial stage in CBIR. The feature of each image is extracted and stored in a vector. The process is as described below.

4.1 Color Feature

HSV Histogram. Color Histogram is a schematic impression of the distribution of colors in an image. Although simple and straightforward, Color Histogram is sensitive to change in brightness and does not account for spatial information. Hence, images are converted from Red, Blue, and Green (RGB) toHue, Saturation, Value (HSV) color space. The histogram is computed for the HSV color model as it has an advantage that it separates the chrominance and luminance of an image. Each HSV component is quantized into (8 * 2 * 2) bins which gives a vector of 32 dimensions.

Color Auto Correlogram. A color auto correlogram, which is an indexed table of color pairs and their probabilities, includes spatial correlation of colors and is easy to compute. In color auto correlogram, color distribution is calculated as a function of the distance between two pixels. The image is quantized into 4 * 4 * 4 bins, which give a feature vector of 64 dimensions.

Color Moment. Color moment is the simplistic distribution of color in the image. Likeliness of two images can be compared with the help of color moments. The mean and standard deviation moments are calculated. Mean is the average color of the image and the standard deviation is calculated by taking the square root of the variance. Thus, the first two moments of each channel (RGB) is extracted to form a vector of six dimensions.

In order to compensate for the deficits of each technique, a combination of three techniques are used in order to leverage their best outcomes for color feature extraction.

Mean:

$$E_i = \sum_{j=1}^{N} \frac{1}{N} P_{ij} \tag{1}$$

N Number of pixels in the image
P_{ij} value of the jth pixel in the image at ith color channel.

Standard deviation:

$$\sigma = \sqrt{\left(\frac{1}{N} \sum_{j=1}^{N} \left(P_{ij} - E_i\right)^\wedge 2\right)} \tag{2}$$

E_i mean value for ith color channel of the image.

4.2 Shape Feature

Wavelets are a more general way to represent and analyze multiresolution images. Thus, shape feature extraction is accomplished using wavelet transform and Histogram of Oriented Gradients.

Histogram of Oriented Gradients (HOG). The HOG technique counts the number of occurrences of a specific orientation in each part of the image. It forms a feature vector of 1-N length, where N is the length of the HOG feature. The mean is used to form a feature vector.

Wavelet Transform. The wavelet transform is a multiresolution filtering technique that eliminates noise efficiently. The DWT (2 discrete wavelet transform) is used for detection of edges. Coiflet wavelet is applied with a 3 level decomposition and the mean and standard deviation is used to form a feature vector of 40 dimensions.

4.3 Texture Feature

Gabor Wavelet. Gabor Wavelet is used for the extraction of texture feature. The Gabor representation minimizes uncertainty in space and frequency dimensions and the micro-features extracted characterize texture information. Gabor wavelet filters are applied to each image spanning across four scales and six orientations. This produces a vector of 48 dimensions.

All features from the aforementioned steps are concatenated to form a feature vector of 192 dimensions.

A query image is an input for feature extraction and the feature vector is stored.

5 Classifier

5.1 Support Vector Machines (SVM)

Post the feature extraction process, all the images in the database are classified using SVM. It is a supervised learning algorithm that is used for classification and regression analysis. The approach used is "one-versus-one", where $\frac{n!}{(n-k)!k!}$ binary classifiers have to be trained for a k-way problem. It differentiates the samples of a pair of classes at a time. When a query image is given as input, a voting scheme is applied to all $\frac{n!}{(n-k)!k!}$ classifiers. The predicted output by the classifier is the class that gets the highest number of '+1' predictions.

6 Result Analysis

Various experiments were performed to show the efficiency of the proposed method. The system receives a single query image and returns 20 similar images from the database. The result is tested using test images from each class. Of 100 images in each class, 95 images are trained and 5 images are used for testing. For each query image, there exist 95 relevant images.

Performance evaluation can be done using numerous metrics. In this paper, precision and recall have been used for performance evaluation.

Precision

Precision is the ratio of retrieved relevant images to the total number of images retrieved.

$$\text{Precision} = \frac{\text{no. of relevant images retrieved}}{\text{total no. of images retrieved}} \tag{3}$$

Recall

Recall is the measure of how many number of truly relevant results are retrieved. A high recall implies that the algorithm has returned most of the relevant images.

$$\text{Recall} = \frac{\text{no. of relevant images retrieved}}{\text{no. of relevant images in the database}} \tag{4}$$

Table 1 shows the average values of the precision and recall of Wang dataset. It shows the retrieval performance when color, texture, and shape features are used in isolation. The average value is taken for each class. Table 2 shows the values based on a combination of color, texture, and shape feature.

Table 3 shows a comparison of the existing technique to the proposed technique. It can be seen that the color feature is more effective compared to shape and texture when the features are used in isolation. It is clear that the proposed system where the features are combined offers the highest value of precision and recall (Figs. 2, 3, 4 and 5).

The results demonstrate that the output obtained by using color, shape and texture features separately do not match with the query image whereas a combination of the three produced much accurate results. The performance measure was calculated based on precision and recall.

7 Conclusion

The goal of this paper is to retrieve images from a database with reliable accuracy by using multiple techniques in tandem with one another. The proposed work introduces an integrated approach to CBIR which helps retrieve similar images from a

Table 1 Retrieval result using color, shape, and texture separately

Class	Color		Shape		Texture	
	Precision	Recall	Precision	Recall	Precision	Recall
Africa	0.8	0.16	0.4	0.08	0.6	0.12
Beach	0.8	0.16	0.6	0.12	0.6	0.12
Monument	0.8	0.16	0.4	0.08	1	0.2
Bus	0.8	0.16	0.6	0.12	0.6	0.12
Dinosaur	0.8	0.16	1	0.2	1	0.2
Elephant	0.8	0.16	0.8	0.16	0.2	0.04
Rose	1	0.2	0.8	0.16	1	0.2
Horse	1	0.2	0.8	0.16	0.4	0.08
Mountain	0.8	0.16	0.6	0.12	0.6	0.12
Food	0.6	0.12	0.4	0.08	0.6	0.12
Mean	0.82	0.164	0.64	0.128	0.66	0.132

Table 2 Retrieval result combining color, shape, and texture

Class	Precision	Recall
Africa	0.8	0.16
Beach	0.8	0.16
Monument	0.8	0.16
Bus	1	0.21
Dinosaur	1	0.21
Elephant	0.8	0.16
Rose	1	0.21
Horse	1	0.21
Mountain	0.8	0.16
Food	0.8	0.16

Table 3 Comparison of the existing and proposed method

	Color	Shape	Texture	Color, shape and texture
Precision	0.82	0.64	0.66	0.88
Recall	0.164	0.128	0.132	0.18

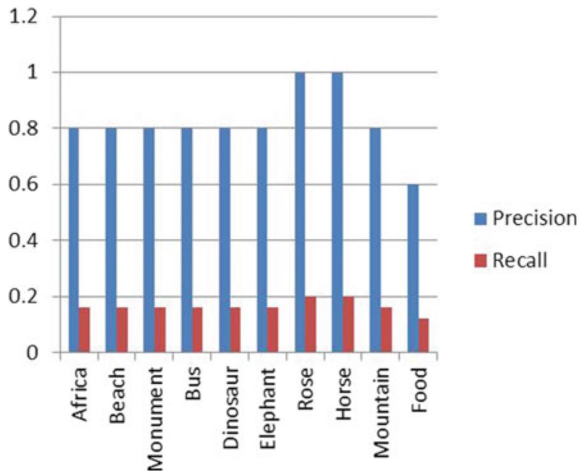

Fig. 2 Precision and recall using color

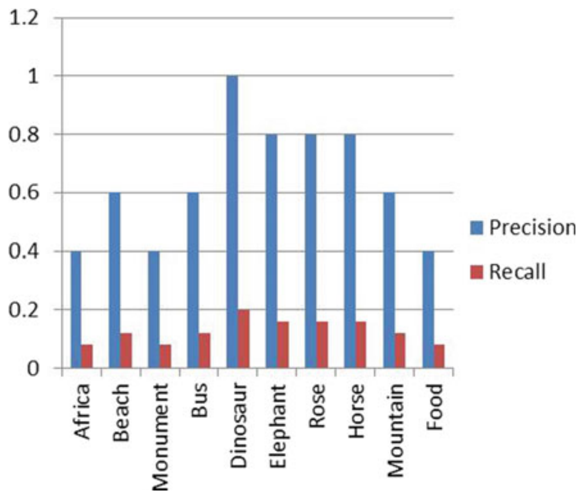

Fig. 3 Precision and recall using shape

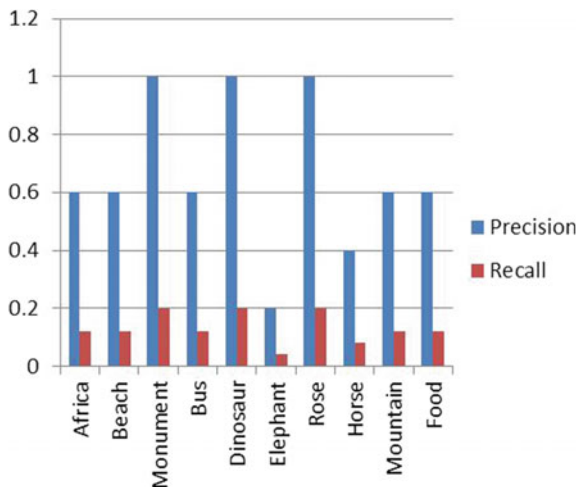

Fig. 4 Precision and recall using texture

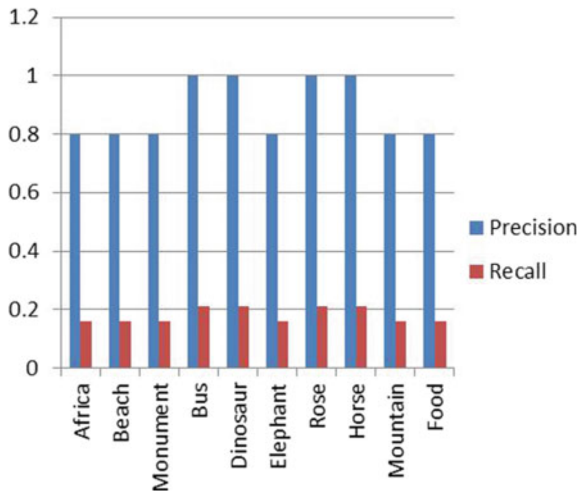

Fig. 5 Precision and recall combining all features

database using SVM. A comparative study on retrieval results was performed and the results using color, shape, and texture features in tandem with one another has given the precision value of 0.88 and recall value of 0.18 for the Wang dataset, which superseded previous research on CBIR. Future refinement of the work will involve research using a bag of words as a feature for larger datasets.

References

1. Sural S, Qian G, Pramanik S (2002) Segmentation and histogram generation using the HSV color space for image retrieval. In: IEEE international conference on image processing
2. Shirazi S, Khan NUA, Umar AI, Razzak MI, Naz S, AlHaqbani B (2016) Content-based image retrieval using texture color shape and region. Int J Adv Comput Sci Appl (IJACSA) 7(1):418–426
3. Iqbal Q, Aggarwal JK (2002) CIRES: a system for content-based retrieval in digital image libraries. In: Seventh international conference on control, automation, robotics, and vision (ICARCV), Singapore
4. Gonde AB, Maheshwari RP, Balasubramanian R (2013) Modified curvelet transform with vocabulary tree for content based image retrieval. Digit Signal Proc 23(1):142–150
5. Manjunath BS, Ohm JR, Vasudevan VV, Yamada A (2001) Color and texture descriptors. IEEE Trans Circ Syst Video Technol 11(6):703–715
6. Giri A, Meena YK (2014) Content based image retrieval using integration of color and texture features. Int J Adv Res Comput Eng Technol (IJARCET) 3(4)
7. Pandey D, Shivpratapkushwah A (2016) Review on CBIR with its advantages and disadvantages for low-level features. Int J Comput Sci Eng 4(7):161–167
8. Rosebrock A, Oates T, Caban J (2013) Ecosembles: a rapidly deployable image classification system using feature-views. In: 12th International conference on machine learning and applications
9. Prakash KSS, Sundaram RMD (2007) Combining novel features for content based image retrieval. In: EURASIP conference focused on speech and image processing
10. Aravind G, Andan HM, Singh T, Joseph G (2015) Development of biometric security system using CBIR and EER. In: IEEE international conference on communication and signal processing (ICCSP)
11. Singha M, Hemachandran K (2012) Content based image retrieval using color and texture. Signal Image Process Int J (SIPIJ) 3(1):39
12. Raghupathi G, Anand RS, Dewal ML (2010) Color and texture features for content based image retrieval. In: Second International conference on multimedia and content based image retrieval
13. Arivazhagan S, Ganesan L, Priyal SP (2006) Texture classification using Gabor wavelets based rotation invariant features. Pattern Recogn Lett 27(16):1976–1982
14. Bagyammal T, Parameswaran L (2015) Context based image retrieval using image features. Int J Adv Inf Eng Technol (IJAIET) 9(9)
15. Li J, Wang JZ (2003) Automatic linguistic indexing of pictures by a statistical modeling approach. IEEE Trans Pattern Anal Mach Intell 25(9):1075–1088. http://wang.ist.psu.edu/docs/related/