CBIR using Bucketing and Histogram

Sangita Chaudhari
A.C.Patil College of Engineering,Kharghar

Nitin P. Kshirsagar
G.V.Acharya Institute of Engineering,Shelu

Jitendra S. Saturwar
JDIET, Yavatmal

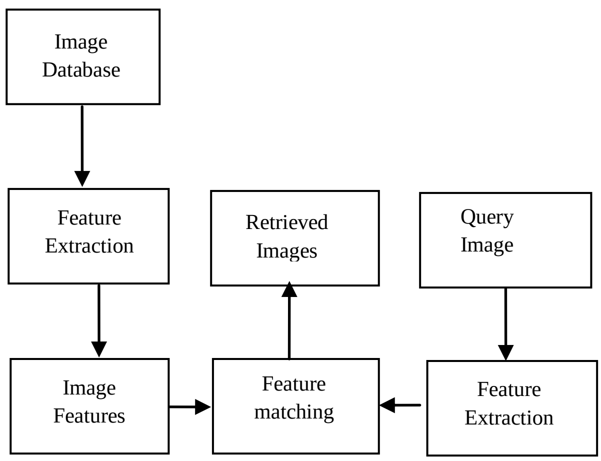
ABSTRACT
In content based image retrieval (CBIR) system, Colour histogram is widely used. Although colour histograms are efficient and insensitive to changes in camera viewpoint, images with different appearance may result in similar histograms. In this paper, we have proposed histogram based matching by diving images into several buckets (grids). This will overcome the shortcoming of histogram matching of whole image. The proposed algorithm is evaluated using performance metrics and it is observed that proposed algorithm (with bucketing) gives better performance than that of without bucketing retrieval algorithm.

General Terms
Pattern Recognition, Security

Keywords
Bucketing, Content based image retrieval, histogram matching.

1. INTRODUCTION
Advances in data storage and image acquisition technologies have enabled the creation of large image datasets. To manage these datasets, it is necessary to develop appropriate information systems. Image searching is one of the most important services that need to be supported by such systems. In general, two different approaches have been applied to allow searching on image collections: one based on image textual meditate and another based on image content information. The first retrieval approach is based on attaching textual metadata to each image and uses traditional database query techniques to retrieve them by keywords [1]. However, these systems require a previous annotation of the database images, which is a very laborious and time-consuming task. In second approach called content-based image retrieval, image processing algorithms are used to extract feature vectors that represent image properties such as color, texture, and shape. In this approach, it is possible to retrieve images similar to one chosen by the user. One of the main advantages of this approach is the possibility of an automatic retrieval process, contrasting to the effort needed to annotate images.
Colour histograms are frequently used to compare images. Examples of their use in multimedia applications include scene break detection and querying a database of images [2]. Colour histograms are popular because they are trivial to compute, and tend to be robust against small changes in camera viewpoint [3].

2. PRINCIPLE OF CBIR
Content-based image retrieval, also known as query by image content and content-based visual information retrieval is the application of computer vision to the image retrieval problem, that is, the problem of searching for digital images in large databases. Content-based search makes use of the contents of the images themselves, rather than relying on human-input metadata such as captions or keywords. In CBIR, features are extracted from each image that is stored in the database and compared to the features of the query image [4].
Basic idea behind CBIR is that, when building an image database, feature vectors from images (the features can be color, shape, texture, region or spatial features, features in some compressed domain, etc.) are to be extracted and then store the vectors in another database for future use. When given a query image its feature vectors are computed. If the distance between feature vectors of the query image and image in the database is small enough, the corresponding image in the database is to be considered as a match to the query. The search is usually based on similarity rather than on exact match and the retrieval results are then ranked accordingly to a similarity index.



3. IMAGE RETRIEVAL USING HISTOGRAM MATCHING
We describe a method which imposes additional constraints on histogram based matching. In histogram refinement the pixels of a given bucket will be subdivided into classes based on local features, including texture, orientation, distance from the nearest edge, relative brightness, etc.
Histogram refinement prevents pixels in the same bucket from matching each other if they do not fall into the same class. Pixels in the same class can be compared using any standard method for comparing histogram buckets (such as Euclidean Distance). This allows fine distinctions that cannot be made with colour histograms.
The initial stage we first convert the image into grey scale image and then we calculate the histogram for gray scale image. The next step is to find the Euclidean distance of the histogram of the respective images. Based on Euclidean distance we have an option whether to compare the images with bucketing or without bucketing. When we select without bucketing option then directly the Euclidean distance of the

histogram of different images is compared. And when we use the bucketing method we have choice of bucketing which varies from 4 to 36 buckets. In this method we divide the images into the given number of buckets and the histogram of those buckets are calculated and compared with the bucketed histogram of the query image. Then we sort the image according to the minimum Euclidean distance calculated and thus the best eight results are displayed. The proposed system is formalised by following steps:

1. Convert RGB images from databases to gray scale images.
2. Divide each image into M x M buckets (grids).
3. Calculate histogram for each bucket.
4. Repeat step 1 to 3 for query image.
5. For every bucket from query image and database image, calculate Euclidean distance using following equation.

$$d_{i,j} = \sqrt{\sum_{k=0}^{255} (hist_{ki} - hist_{kj})^2} \qquad (1)$$

   Where $d_{i,j}$ is distance between bucket B_i of query image(I_Q) and bucket B_j of database image(I_D). $hist_{ki}$ and $hist_{kj}$ are corresponding histograms of bucket B_i and B_j.
6. Calculate similarity distance by adding all Euclidean distances between query image and database image.

$$SD(I_Q, I_D) = \sum_{K=1}^{M \times M} d_k \qquad (2)$$

7. Sort similarity distances in ascending order and retrieve first N images.

Two images with identical color histograms can have different grayscale histograms; thus, Gray scale histograms create a finer distinction than color histograms. This is particularly important for large image databases, in which many images can have similar color histograms.

4. EXPERIMENTAL RESULTS

The experiment was conducted on a set of 1000 images from Corel database. The database consists of 100 images of every class. The images are taken from different viewpoints but under approximately constant illumination conditions. Figure 2 shows some sample images from the database. The image retrieval system is implemented using JAVA. Figure 3 shows top eight results for query response for without bucketing. Figure 4 and 5 shows results for 4 and 9 bucketing based histogram matching algorithm.

4.1 Performance Evaluation of CBIR

The level of retrieval accuracy achieved by a system is important to establish its performance. If the outcome is satisfactory and promising, it can be used as a standard in future research works. In CBIR, precision-recall is the most widely used measurement method to evaluate the retrieval accuracy[5].

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

$$\text{Pr}\,ecicion = \frac{Number\ of\ relevent\ images\ retrieved}{Total\ number\ of\ images\ retrieved} \qquad (3)$$

Recall is defined as the ratio of the number of relevant images retrieved to the total number of relevant images in the whole database

$$\text{Re}\,call = \frac{Number\ of\ relevent\ images\ retrieved}{Total\ number\ of\ relevant\ images\ \text{in}\ database} \qquad (4)$$

Error Rate is defined as the ratio of the number of non-relevant retrieved images to the total number of images in the whole database.

$$Error\ Rate = \frac{Number\ of\ non-relevent\ images\ retrieved}{Total\ number\ of\ images\ retrieved} \qquad (5)$$

Table 1. Performance Evaluation of CBIR

No. of Buckets	Precision	Recall	Error Rate
0	0.46	0.46	0.53
4	0.71	0.71	0.28
9	0.75	0.75	0.24

5. CONCLUSION

In this paper, histogram based content based image retrieval system is presented. Histogram matching is done by diving images into several buckets (grids). This will overcome the shortcoming of histogram matching of whole image. The proposed algorithm is evaluated against performance metrics and it is observed that proposed algorithm (with bucketing) gives better performance than that of without bucketing retrieval algorithm.

6. REFERENCES

[1] Gerard, S. and Buckely, C. 1988.Term-Weighting Approaches in Automatic Text Retrieval. Information Processing and Management, 24(5): 513-523.

[2] Pentland, A., Picard, R. and Sclaro, S. 1996. Photobook: Content-based manipulation of image databases. International Journal of Computer Vision, 18(3):233-254.

[3] Kumar, A., and Esther, J. 2011. Comparative Study on CBIR based by Color Histogram, Gabor and Wavelet Transform. International Journal of Computer Applications 17(3):37-44.

[4] Suhasini, P., Krishna, K. and Krishna V. 2009. CBIR using color histogram processing. Journal of Theoretical and Applied Information Technology. 6(1):116-122.

[5] Koselka, M., Laaksonen, J., Laakso, S. and Oja, E. 2000. Evaluating the Performance of Content-Based Image Retrieval Systems. In proceedings of 4th International Conference on Advances in Visual Information Systems.

Fig. 2: Random Images from each of the 10 Classes Corel Database

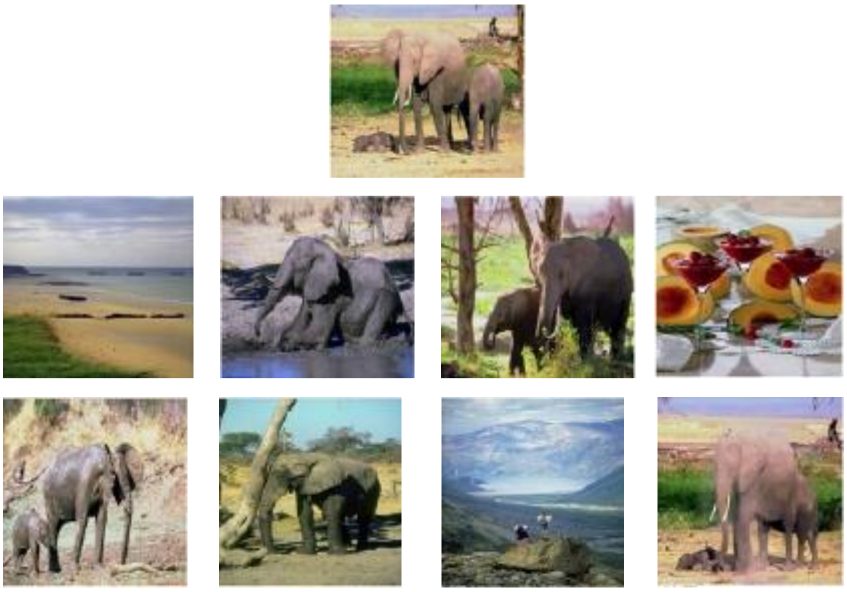


Fig. 3: Query Response for CBIR without Bucketing

Fig. 4: Query Response for CBIR with 4 Bucketing

Fig. 5: Query Response for CBIR with 9 bucketing