Colorado Technical University

Technical Report
Computer Science

Using image color vectors to index and retrieve images from an image database

Paul V. Thor
Professor of Computer Science
Colorado Technical University
pthor@coloradotech.edu

Bo I. Sandén
Professor of Computer Science
Colorado Technical University
bsanden@acm.org

Mary Jane Willshire
Visiting Professor of Computer Science
Colorado Technical University
willshir@up.edu

Technical Report Number
CTU-CS-2001-003

Using image color vectors to index and retrieve images from an image database

Paul V. Thor, Bo I. Sandén and Mary Jane Willshire

Abstract

Along with image databases comes the need to classify and index images for later retrieval. For classification and indexing purposes, information about the image must be gleaned from either it or its storage model. This paper proposes the Image Color Vector Number (ICVN) as one indexing scheme. ICVN looks at the color trends when an image is traversed according to a set pattern. It is used in a manner similar to the color histogram, which shows the distribution of pixels in an image across a gray scale.

The paper examines the effectiveness of ICVN alone and in conjunction with the color histogram as an index to an image database. The results show that the ICVN is a viable alternative to the color histogram. The ICVN technique combined with the color histogram shows a marked improvement over a color histogram alone when tested with typical digital scenery images.

1. Introduction

Techniques for storing and retrieving images as well as general image processing are areas of increasing interest. The image database is a relatively new capability that came into existence with the advent of graphics files and the ability to digitize images. With the development of the image database comes the need to classify or index images so they can later be retrieved. For classification and indexing purposes, information about the image must be gleaned from either it or its storage model.

Finding the right image in a database remains an intensely human activity. Typical queries are "find companies whose stock prices move similarly", "find images of sunsets", and "find X-rays that contain something that has the texture of a tumor" [2]. It is highly desirable to automate the activity of finding the right image in a database.

Many techniques have been used to determine the differences among images. The most obvious differentiation that exists between images is the image content and its spatial arrangement. Image detection methods often use computer vision techniques that involve edge detection and enclosing blocks [8]. Other factors that can be used to distinguish among images are the amount of color content or texture and the size and location of a predominant feature. Object overlapping orders, the apparent distance of a feature from an identifiable boundary, the lighting scheme and the amount of abstraction have also been used. Most current clipart/image databases fall back on a textual description of the image content to direct image retrieval. This paper puts forward a technique for image abstraction and differentiation, the computation of a color vector number, which when used in conjunction with the color histogram representation of an image, improves the accuracy of retrieval of a set of similar images in an image database.

In order to compute the image color vector number, a color image is represented as an array of pixels where each pixel is described by a 32-bit integer. In the RGB model of digital color representation [6, 12, 3], a pixel can be described by the intensity of each of three-color channels, blue green and red. The intensity of each color is represented by a number between 0 and 255. The average of these three numbers represents the gray scale intensity of a pixel. In the ICVN technique, the image is partitioned, the colors in each partition examined and averaged, and the vector number is computed. These values are used to determine the changes in the color patterns over the image. The details for the generation of the ICVN are given in Section 3 of this paper.

2. Image indexing and retrieval

"Retrieval based on visual object appearance involves four categories of information items: features; feature space; feature groups; and image space" [5]. A feature is a characterization of a specific property of the image obtained through an attribute analysis transformation algorithm. A feature is typically represented as a set of numbers and often referred to as a feature vector. Feature vectors are most often obtained by image projection, applying a function to the feature input, and distance calculations between two given features. The features of visual objects can be represented as points in the feature space. For example, a feature called "texture" of an image region may be represented by the three numbers: randomness, periodicity and directionality [5].

The features of an image occupy a position in feature space. Features from different images with similar values are located in the same region of the space. Operations performed on feature space include finding the hypercube that encloses selected feature points, finding all points in a given hypercube and finding the k nearest neighbors of a given point.

Most current research focuses on methods for using locations, sizes and arrangements of the image's compositional elements as the basis for retrieval [5]. One kind of image database system uses explicit spatial data structures and operators. These databases attempt to catalog spatial relationships between compositional objects in the image. The query language has operators for specifying object relationships in terms of size and relative locations. These systems tend to work well for constrained domains, but are often baffled when attempting to find images of clouds on a blue sky at the top of an image [5].

Other systems divide the image into regions and ask the user to specify what regions are important for spatial matching. Image features for each region of importance are determined, and similarity with the target is determined by means of distance functions on the features. One problem with these schemes is to determine how much spatial deviation is allowed in a good match.

A third approach segments an image into like-sized regions. Retrievals take the form of "Retrieve all images containing regions that have the color of object A, the texture of object B and the shape of object C, and lie in the upper third of the image" [4]. The ICVN is an extension of this idea. For each segment of the image the average color specification is determined and the patterns that result when comparing the color in one segment to the color in another segment form the basis for an abstraction of the image. This image representation, that is generally unique to each image, can be compared to the same representation generated from other images.

The ICVN idea stems from the color histogram, but encompasses more information. While the color histogram merely generates a vector of intensity magnitudes, the ICVN, in essence captures the color variations as one traverses the image segments resulting in an image's color variation signature. This signature, when combined with the color histogram, facilitates the differentiation/similarity of images in a more refined way. The cost of determining the ICVN is minimal as it uses image information that is necessary for determining the color histogram. The combination has been shown to be much more effective for identifying image similarity than using the color histogram alone.

2.1 Determining similarity

There are two broad approaches to determining similarity of visual images, the metric approach and the transformation approach [11]. The metric approach uses a difference metric to compare image objects. There is only one notion of similarity. This approach makes it easy to develop an index and allows for efficient computation. In the transformational approach, the user has options as to what should be considered factors for comparison. Weights for the factors can be applied and then the resulting information applied to a transformation process for final encoding that facilitates comparison or indexing. For example, if night pictures are desired, a larger weighting toward darker intensities may be applied to the color histogram. The color histogram and the ICVN are a form of transformation applied to the color information in the image. The idea is similar to hashing a key value to a legal index in a hash table.

One traditional way to determine likeness is to use the image's color features. It is a typical metric approach. A color histogram looks at the gray scale intensity value of each pixel. It shows how many pixels there are at each possible intensity value from 0-255 and results in a near-unique vector for almost all images. The search engine uses some concept of distance between histograms to determine similarity. The color histogram method is fast, making real-time implementation possible. It is immune to distortion including deformation, translation, rotation, occlusion, and scaling [9]. Although images retrieved by means of color histograms share the overall color tone, they differ greatly in semantic content [7]. Furthermore, color histograms do not capture the spatial relationships of the color regions [10]. An advantage of using the ICVN is that it does capture a relationship between the colors in adjacent or distant regions of the image.

2.2 Evaluation of the results of image retrieval

The evaluation of image retrieval is based on subjective judgment. There are two criteria, goodness of retrieval and effectiveness [4]. Goodness of retrieval is a judgment by the user as to how well the system performs in response to a query. Users measure goodness primarily "based on how much of the retrieved data is good, rather than on how much of the relevant data is retrieved" [4].

Effectiveness of a query to find the images similar to a given image is calculated based on the concepts of the ideal set (I), of those images that make the best solution to the query from a human viewpoint. The actual set (A) contains those objects actually returned. Two generally used measures of effectiveness are recall and precision. Recall (R) is the number of correctly selected objects divided by the size of the ideal set, that is:

$$R = |I \cap A| / |I| \quad (1)$$

Precision (P) is the number of correctly selected objects divided by the total number of objects returned, that is:

$$P = |I \cap A| / |A| \quad (2)$$

Effectiveness (E) is defined as the harmonic mean of precision and recall, that is

$$E = 2 / (1/R + 1/P) = 2\,|I \cap A| / (|I| + |A|) \qquad (3)$$

Clearly, $0 \leq R \leq 1$, $0 \leq P \leq 1$ and $0 \leq E \leq 1$. $R = P = E = 0$ when no images from the retrieval set are returned. $R = 1$ implies that all the images in the ideal set are in the actual set. $P = 1$ implies that no images outside the ideal set are selected. $E = 1$ implies that the actual and ideal sets are identical.

Database retrieval algorithms can be evaluated using a recall versus precision graph. This way, one can "evaluate quantitatively both the quality of the overall answer set and the breadth of the retrieval algorithm" [1]. These are the measures used to evaluate the contribution of the ICVN to the retrieval.

3. The image color vector number

This paper proposes a metric approach to determining image similarity. Each image is partitioned into patches. The average intensity of each color channel is computed for each patch and stored in a vector. This vector including the variations between patches in average color value for each color channel is used to index an image. By extracting and comparing the average color values that are encountered in different patches as an image is traversed, a distinctive index or code for the image is obtained that can be used to differentiate images in a database. The index is called the image color vector number (ICVN). The ICVN of an image reflects the color trends when the image is traversed according to a set pattern.

To compute the ICVN, the image is partitioned into patches, and a number is obtained by comparing adjacent patches when the image is traversed according to a given pattern. Typically, a rectangular image is partitioned by means of a grid, and the resulting regions traversed line by line from the top left corner. Using zeros and ones to represent an increasing or decreasing average color channel value between two adjacent areas, a distinctive number is generated with one integer per patch. The combined ICVN may typically contain one integer value per patch. If the red, green, and blue channels were increasing, a 3 (1+1+1) is assigned to this transition location. If only one color channel, is increasing, for instance green, a 1 (0+1+0) is assigned at this position. In this research, the image is divided into 64 equal sized patches. While the ICVN is obviously insufficient to recreate the image, it is unlikely that two images would generate identical ICVNs. The ICVN can be used either by itself or in combination with a color histogram to increase the effectiveness of the retrieval of images similar to a target image.

For example, given the image depicted in figure 1, the image is subdivided into 64 patches as shown. Each patch is processed to find its average color and an abstraction of the image is obtained as seen in figure 2. The abstract image is used to compute the ICVN for the original image.

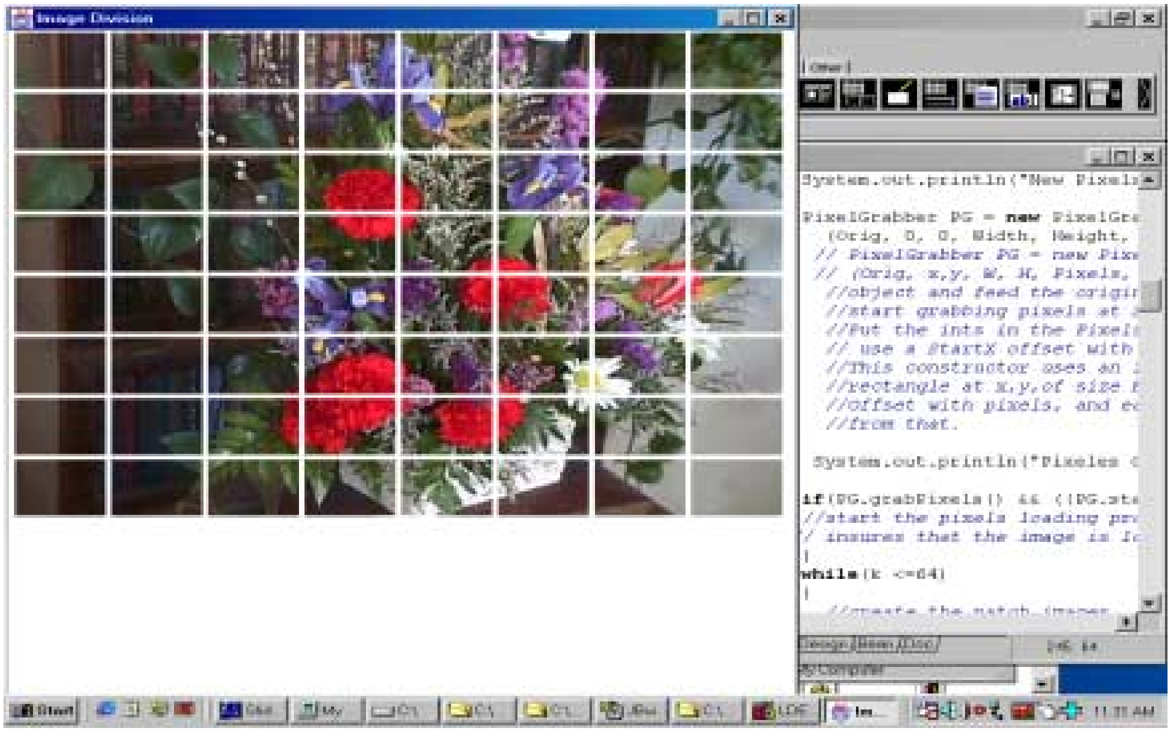
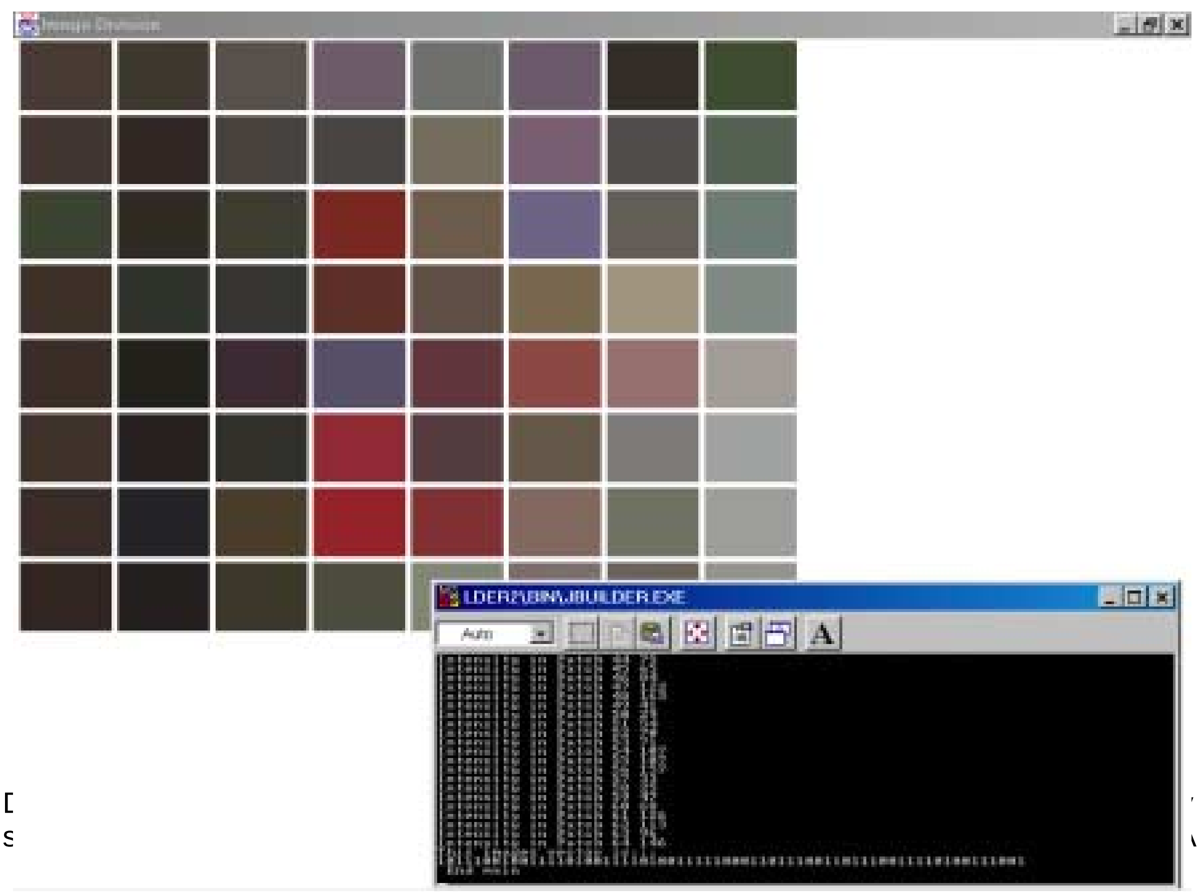


Figure 1. Screen shot of a Partitioned Image – above.
Figure 2. The Average Color of each Patch - below



more comprehensive approach is to find the average intensity of each primary color. If the gray scale value is included, this results in four vector numbers associated with a single image, one vector number for each color channel and one for the gray scale.

3.1 Evaluation of the effectiveness of the color vector approach

A program that computed all the color vectors, a color histogram for each image as well as an image index made up of the ICVN and channel vectors was implemented and run first against a set of primitive images consisting of simple shapes and two or three colors covering varying areas, and then against a database of 103 scenery images.

For the purposes of this test, the ICVNs of an image were calculated by storing for each patch the average red, green, blue and gray intensities in each color channel between this patch and the next.

Images were compared patchwise. For each color channel and each patch, if the difference in average intensity between the pictures was within a certain range, then the change value corresponding to that channel was given a heavy weight. As the change in color intensity increased, less weight was given to the change value.

Testing with the primitive images indicated that the program was sensitive to the amount of colored areas as well as the actual colors involved. For instance, a blue square on a yellow background would cause all other images that had a yellow background and the same blue square to be retrieved, regardless of the position of the square in the image. As the sensitivity threshold was reduced, other similar images such as different color combinations were brought into the result set as well as images with the same color scheme but with the total area of the square broken into many squares. Further reduction of the threshold brought in still other images with additional color areas or where the blue square was much larger than in the test image.

Using the color histogram alone on these primitive images was not successful at all. Invariably, the technique selected all the primitive images in the index. This may have been resolved with higher comparison thresholds.

For the testing with the scenery database, 33 images were selected at random and used as control images. Five evaluators were asked to pick those images in the database that were intuitively closest to each control image. An ideal set was created for each of the control images by including those images that at least three of the evaluators had chosen. The program was run with each of the control images to find like images. This was done with three schemes: image vector alone, color histogram alone and a combination. The ideal sets in combination with the actual sets returned by the program allowed the effectiveness of the program to be quantified.

The program was run in each of the three modes: vector only, vector and histogram, and histogram only. For each mode and each of the 33 test cases, it was run at each of ten threshold settings. With a low threshold, a complex image with some blue sky caused many images with blue sky to be selected. With a higher threshold, the blue sky played a lesser role, and images with similar complexity and color schemes across the picture were selected. With an even higher threshold, only the control image was selected.

The program was also run in such a way that a) only the patchwise color difference and b) only the color variation between patches was used. Each of these was less effective than their combination, and so those approaches were abandoned.

3.2 Results

While the original idea of one ICVN to abstract an image for the purposes of indexing an image database is valid, it became necessary to use the raw information from each patch to actually do the comparisons. Therefore the actual index record being stored is made up of the four vectors as well as the color histogram information and a pointer to the actual image. This record still represents less than 2% of the information contained in the actual image. The effectiveness of searching this index for similar images is addressed next.

In each trial, the program was given a target image and then, based on set parameters, the program retrieved images from the database that its internal algorithm determined were similar to the target image. The Average Precision was calculated using equation (2). Recall that the higher the precision value, the better the performance of the algorithm. Overall the tests conducted, the average Precision is as follows:

Combined approach: 54.22%
Using the ICVNs only: 52.68%
Using the Color Histogram only: 50.59%

This shows that the combined approach has a higher precision value than either of the other approaches alone. The threshold setting did have an impact on the retrieval results. This can be seen by considering the graphs given in figures 3, 4 and 5. These figures show the average effectiveness, calculated using equation (3), by threshold for the three retrieval schemes. For threshold 4 and above, the combined approach was more effective than the color histogram alone. At lower thresholds, virtually the entire database was selected by each scheme, meaning that the effectiveness would be nearly identical, and thus these threshold levels are too low to be of interest. At higher thresholds, the schemes differentiated themselves until threshold 10 where each scheme generally selected the target image alone. For this reason, the target image was eliminated from the calculations. The effectiveness of the combined scheme and of the vector alone was similar. In addition, each was better than the color histogram alone when reasonable threshold levels are considered (that is, above level 4).

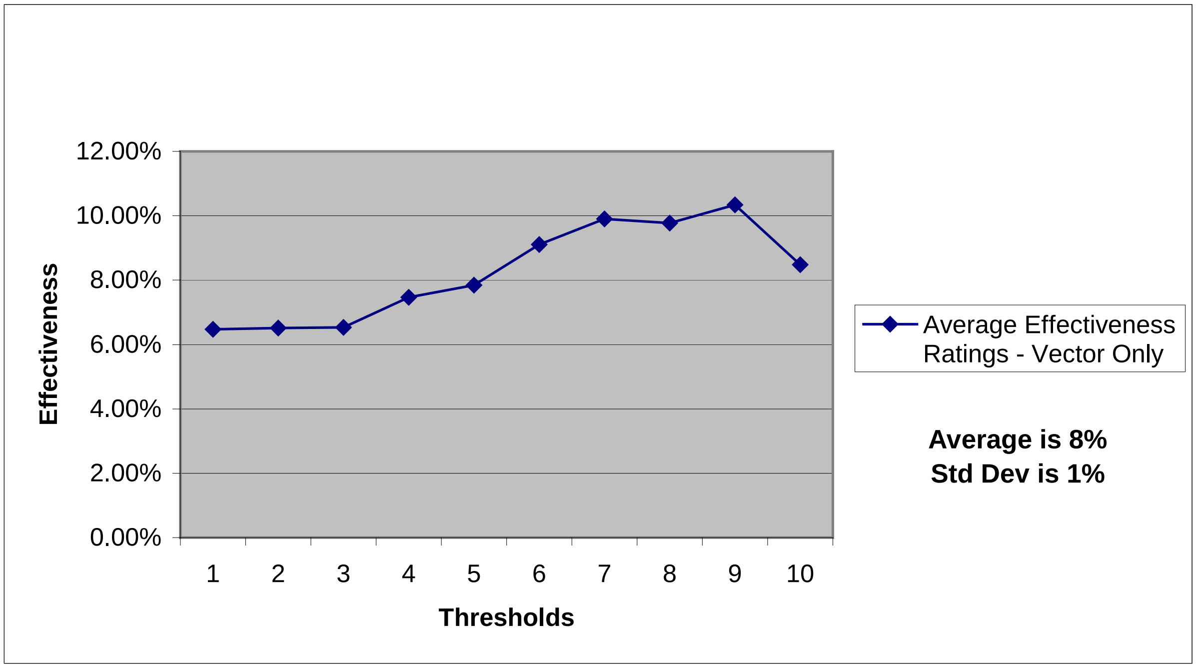


Figure 3. Average Effectiveness Ratings by Thresholds – Vector Scheme Alone

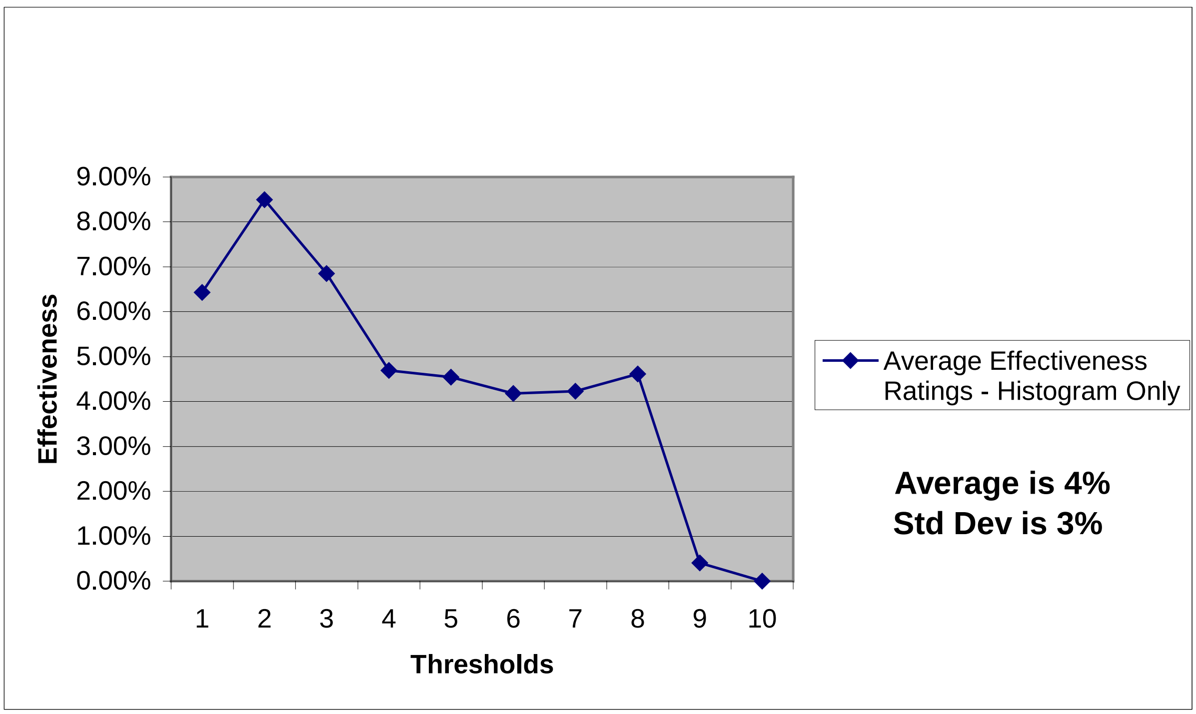


Figure 4. Average Effectiveness Ratings by Threshold – Histogram Only Scheme

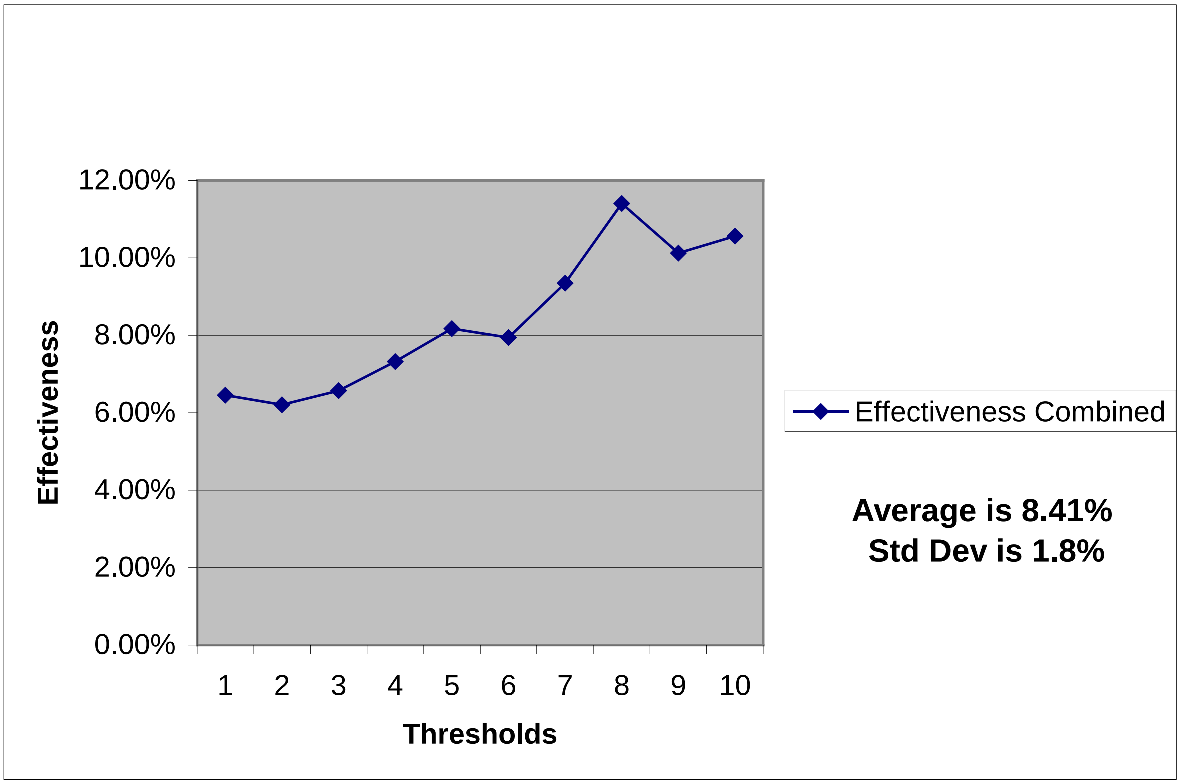


Figure 5. Average Effectiveness Ratings by Threshold for the Combined Scheme

Finally, consider the recall values as calculated using equation (1). Figure 6 shows that recall dropped off considerably with increasing threshold level for the color histogram only scheme while it dropped roughly linearly from 90% at threshold 1 to around 10% at threshold 10 for both the color vector and the combination. In all cases, the combined average recall was higher than either of the single approaches, further supporting the contribution of the ICVN to the retrieval effort.

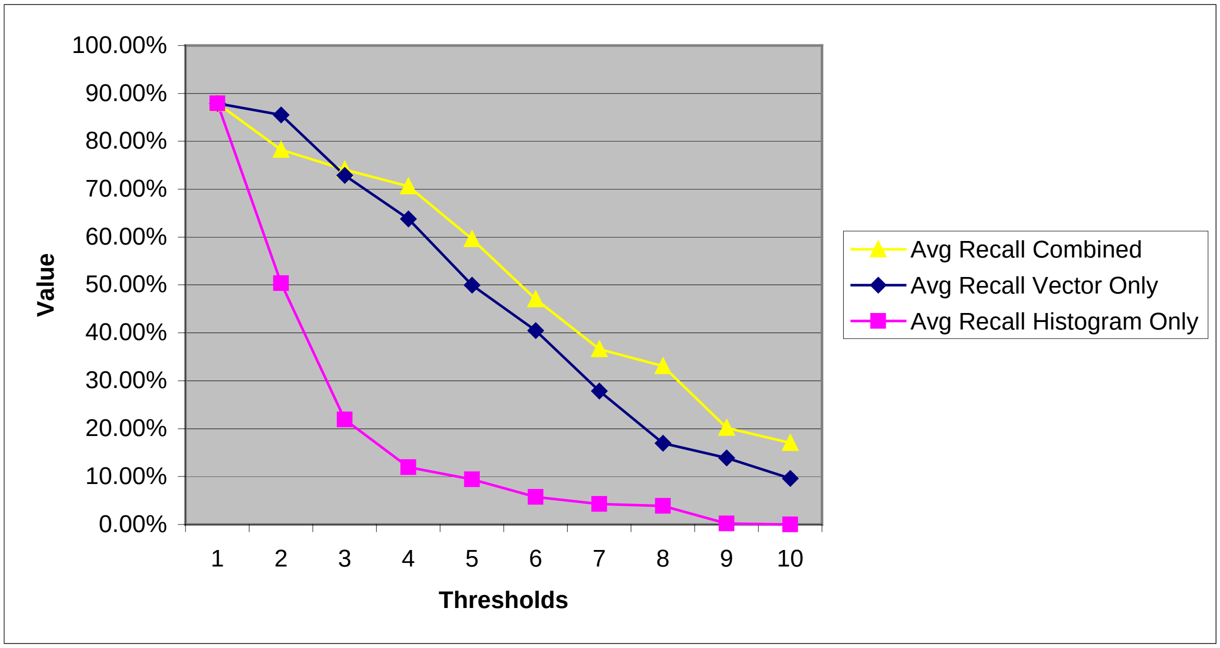


Figure 6. Average Recall At Each Threshold

Based on the above discussion concerning Recall and Precision, it was concluded that both the ICVN and Combined approaches were better than the color Histogram alone approach for finding like images. There was not sufficient difference to determine whether or not the ICVN alone was superior to the Combined approach.

4. Conclusion

The color vector number scheme for searching and retrieving similar images from an image database proved to add value to the color histogram technique for image retrieval. The ICVN takes up less than 2% of the image's storage size, yet yields an average precision of 54.22%. A comparison with the 50% average precision achieved in human interaction intensive, text-based retrieval makes the ICVN scheme competitive with other image indexing and searching schemes.

A large patch size tended to dilute the color information by transforming it to an intensity level. On the other hand, making the patches smaller would make the retrieval process more computing intensive. Finding an optimum patch size was left for further research.

Another issue for future investigation was the traversal of the image patches. Possibilities include traversing only part of an image and with smaller and smaller patch sizes. The traversal order could also be examined. This research only looked at traversing comparing side-by-side patches in a left to right, top to bottom sequence.

References

[1] Baeza Yates Ricardo and Berthier Ribeiro Neto. *Modern Information Retrieval*, ACM Press and Addison Wesley, 1999.

[2] Faloutsos, Christos, 1999. *Multimedia IR: Indexing and Searching*, from Modern Information Retrieval, edited by: Baeza Yates Ricardo and Berthier Ribeiro Neto, ACM Press and Addison Wesley, 1999.

[3] Gonzalez, Rafael C. and Richard E. Woods, 1992. *Digital Image Processing*, Addison Wesley Inc.1992.

[4] Gupta, Amarnath and Ramesh Jain, 1997. “Visual Information Retrieval,” Communications of the ACM, May 1997/Vol. 40. No. 5.

[5] Gupta, Amarnath, Simone Santini and Ramesh Jain, 1997. “In Search of Information in Visual Media,” Communications of the ACM, December 1997/Vol. 40. No. 12.

[6] Klette, Reinhard and Piero Zamperoni, 1996. *Handbook of Image Processing Operations*, John Wiley & Sons, 1996.

[7] Lipson, Pamela R., 1996. *Context and Configuration Based Scene Classification*, MIT Dissertation, 1996.

[8] Ma, W. Y. and B. S. Manjunath, 1997.*Edge Flow: A Framework for Boundary Detection and Image Segmentation*, U.C. Santa Barbara, from IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Jun 1997. IEEE Computer Society Press, 1997.

[9] Park, In Kyu, Il Dong Yun, and Sang Uk Lee, 1997. “Models and Algorithms for Efficient Color Image Indexing,” from IEEE Proceedings: Content Based Image and Video Libraries, Jun 1997. IEEE Computer Society, 1997.

[10] Ratan, A. Lakshmi and W.E.L. Grimson, “Training Templates for Scene Classification using a Few Examples,” AI Lab, M.I.T. from Proceeding IEEE Workshop on Content Based Access of Image and Video Libraries, June 20, 1997.

[11] Subrahmanian, V. S., 1998. *Principles of Multimedia Database Systems*, Morgan Kaufmann Publishers, Inc., 1998.

[12] Umbaugh, Scott E., 1998. *Computer Vision and Image Processing – a practical approach using CVIPtools*, Prentice Hall, Inc., 1998.