

## ***eID: A System for Exploration of Image Databases***

Daniela Stan, Ishwar K. Sethi

Intelligent Information Engineering Laboratory

Department of Computer Science & Engineering

Oakland University

Rochester, Michigan 48309-4478

Phone: (248) 370-2137

Fax: (248) 370-4625

[dstan@oakland.edu](mailto:dstan@oakland.edu), [isethi@oakland.edu](mailto:isethi@oakland.edu)

## Abstract

The goal of this paper is to describe an exploration system for large image databases in order to help the user understand the database as a whole, discover hidden relationships, and formulate insights with minimum effort. While the proposed system works with any type of low-level feature representation of images, we describe our system using color information. The system is built in three stages: the feature extraction stage in which images are represented in a way that allows efficient storage and retrieval results closer to the human perception; the second stage consists of building a hierarchy of clusters in which the cluster prototype, as the *electronic identification (eID<sup>\*</sup>)* of that cluster, is designed to summarize the cluster in a manner that is suited for quick human comprehension of its components. Besides summarizing the image database to a certain level of detail, an *eID* image will be a way to provide access either to another set of *eID* images on a lower level of the hierarchy or to a group of perceptually similar images with itself. As a third stage, the multi-dimensional scaling technique is used to provide us with a tool for the visualization of the database at different levels of details. Moreover, it gives the capability for semi-automatic annotation in the sense that the image database is organized in such a way that perceptual similar images are grouped together to form perceptual contexts. As a result, instead of trying to give all possible meanings to an image, the user will interpret and annotate an image in the context in which that image appears, thus dramatically reducing the time taken to annotate large collection of images.

**Keywords:** k-means clustering, multi-dimensional scaling, image annotation, semantic retrieval

---

\* In a formal definition, an *electronic IDentification (eID)* is the most similar image to the other images from a corresponding cluster; that is, the image in the cluster that maximizes the sum of the squares of the similarity values to the other images of that cluster.

## 1. INTRODUCTION

The increasing rate at which images are generated in many application areas gives rise to the need of image retrieval systems to provide an effective and efficient access to image databases, based on their visual content. There are two main approaches to access an image database: a query driven methodology allows the user to specify either a text query (keywords, annotations, etc) or an image query; a browsing driven methodology allows users to navigate through the database until they identify an image of interest and then, initiate a search using that image as the query image. The query driven methodology is more appropriate for experts or users who do not have any difficulties in formulating a query, while browsing driven methodology is important for users who are not very familiar with the image domain characteristics and they would like first to get deeper insights about the image database before formulating their query. In both cases, the users should not only see and accept the computer's retrieval results, but also easily understand and interpret them. Consequently, in order to create a powerful knowledge discovery environment in which the user's conceptualization of a query corresponds to what actually the system processes, human mind's exploration capabilities should be integrated with the processing power of computers.

Our approach is meant to generate visualization of an image database from finer to finer details, to help the user understand the database as a whole, discover hidden relationships, and formulate insights with minimum effort. To understand better our approach, imagine discovering a particular topic in a book with minimum reading: you read the table of content and if you are interested in a particular chapter, you read the titles' sections of the corresponding chapter and so on until you find the pages containing the information that you want to read. The same modality of exploration we want to create for an image database having thousands or millions of items: in our case, the table of content will give a general view of the image database and will consist of the most representative images, referred as *eID* images. Besides summarizing the image database to a certain level of details, an *eID* image, just by a 'click on' action, will be a way to provide access either to another set of electronic identification images at a regional/lower view layer (like titles' subsections of a book chapter) or to a group of perceptual similar images at a image/terminal layer. Moreover, instead of presenting the image database as linear lists of *eIDs* at different

levels of details, we build two-dimensional maps in which the similarity ranked order of *e*IDs is preserved. This brings a salient feature to our exploration system, in addition to that of browsing a collection to find a satisfactory query to initiate the image retrieval process. It gives the system the capability of semi-automatic annotation in the sense that the image database is organized in such a way that perceptual similar images are grouped together to form perceptual contexts later used for annotation. As a result, instead of trying to give all possible meanings to an image, the user will interpret and annotate an image in the context in which that image appears.

## **2. RELATED WORK AND OUR APPROACH**

Since the early 90's, many Content-based Image Retrieval (CBIR) Systems<sup>†</sup> have been built. The QBIC (Query by Image Content) system developed by IBM (Niblack et al., 1994, Flickner et al., 1995) supports query based on example images, user-constructed sketches, drawings, selected color and texture patterns, and text-based query. Virage, developed by Excalibur Technologies Corp. (Bach et al., 1996), is a CBIR system similar to QBIC that has an additional characteristic of combining queries using different features: users can adjust the weights associated with the visual features according to their significance. Some other CBIR systems that employ different techniques and allow similar type of queries are Photobook by Pentland et al. (1996), MARS (Multimedia Analysis and Retrieval System), by Mehrotra et al. (1997a and 1997b), RetrievalWare by James (1993), VisualSEEk by Smith and Chang (1996a), and Netra by Ma and Manjunath (1997). More recent research emphasis is given to the integration of the human component in order to improve the retrieval performance. For example, the interactive version, FourEyes, of Photobook developed at MIT Media Labs (Minka and Picard, 1996) proposes a "society of models" approach in order to incorporate the human's perception. MARS developed at University of Illinois at Urbana-Champaign is based on relevance feedback architecture that provides the user with a friendly interface in order to evaluate the current retrieval result given by the system (Rui et al. 1997a and 1997b).

There are a few earlier research efforts on image databases' exploration. The system described by Craver et al. (1998) is based on a new data structure built on the multi-linearization of image attributes for efficient organization and fast visual browsing of the images. The systems proposed by Sethi and Coman (1999),

---

<sup>†</sup> CBIR = Content-Based Image Retrieval Systems that permit image searching based on features automatically extracted from the images themselves (pixel data) in Kato, 1992.

and Zhang and Zhong (1995) are based on Hierarchical Self-Organizing Map and are used to organize a complete image database into a 2-D grid. MacCuish et al. (1996) and Rubner et al. (1997) use Multi-Dimensional Scaling to organize images returned in response to a query and for direct organization of a complete database. Active browsing is proposed by Chen et al. (1998) by integrating relevance feedback into the browsing environment and thus, users can modify the database organization to suit a desired task.

Through the paper, we will be using the acronyms MDS<sup>♦</sup> to denote the Multi-Dimensional Scaling approach and SOM<sup>♥</sup> to denote the Self-Organizing Map approach.

We propose the system shown in Figure 1. The first stage is the feature extraction stage in which images are represented in a way that allows efficient storage and retrieval results that correspond to the human perception. Image features are often very high dimensional or the similarity metrics are too complex to have efficient indexing structures.

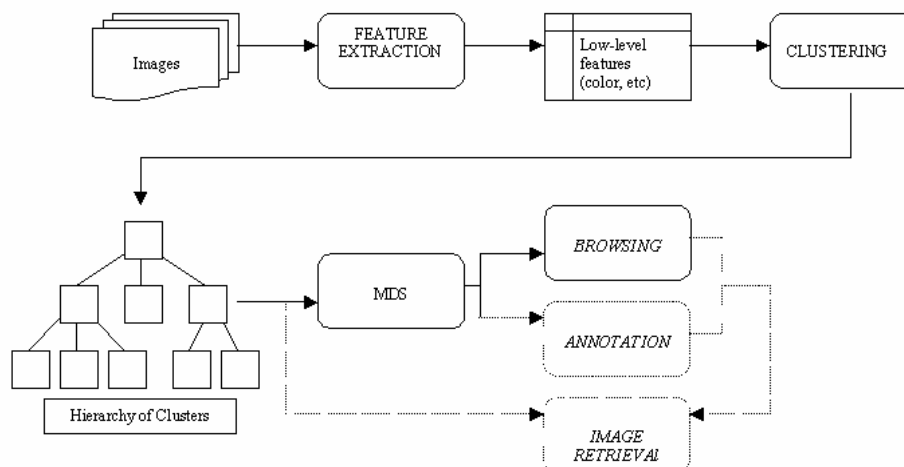


Figure 1: The diagram of *eID* system. The dashed lines represent extended functionalities integrated into our system.

<sup>♦</sup> MDS = Multi-Dimensional Scaling is the search for a low dimensional space, in which points in the space represent the objects, one point representing one object, and such that the distances between points in this space match as well as possible the original dissimilarities between objects in the high dimensional space in Cox and Cox (1994).

<sup>♥</sup> SOM = Self-Organizing Map is an artificial neural network based on competitive and cooperative learning that preserves the topology of the input space when mapping to a 2-dimensional network space. A detailed mathematical formulation of the SOM solution may be found in Kohonen (1997).

The existing multi-dimensional indexing techniques concentrate only on how to identify and improve indexing techniques that are scalable to high dimensional feature vectors in image retrieval (Rui et al., 1999). The other nature of feature vectors in Image Retrieval, i.e. non-Euclidean similarity measures, cannot be explored using structures that have been developed based on Euclidean distance metrics such as the k-d trees, the R-d trees and its variants. Therefore, in building the hierarchy of clusters as the second stage, we use an adaptation of k-means clustering technique as an effective indexing module that solves both high dimensionality and non-Euclidean nature of some color feature spaces. The hierarchy of clusters will give the capability of image database organization at different levels of detail and the cluster prototype, as the electronic identification of that cluster, will summarize the cluster in a manner that is suited for quick human comprehension of its components.

We choose to use the Multi-Dimensional Scaling to visualize the image database. The reason for using the MDS approach instead of the SOM approach is that we want to provide the user with a global view of the image database and to preserve the image similarities, as they are commonly perceived by humans when mapping from the original feature space to the new 2-dimensional space. It is known that SOM preserves the topology, i.e., the local neighborhood relations (Kaski et al., 1998), which in terms represents local views of the image database. The disadvantage of MDS approach is that it does not impose any hierarchical structure; however, this is eliminated by first organizing the image database as a hierarchy of clusters and then visualizing every level of the hierarchy using a 2-dimensional MDS map. The beauty of our approach is that, instead of applying MDS for direct organization of the complete database as in Rubner et al. (1997), at every level of the hierarchy we apply MDS on the corresponding eID images to get global views of the image database at different levels of detail. Consequently, the disadvantage of being very computationally intensive when applied to very large image databases is eliminated since it is applied only on the set of most representative images.

In measuring the effectiveness of a visualization system, we need to realize that the final evaluation will be done by a human. The human eye can only distinguish between a relatively small number of images at once; this number is miniscule in comparison to the thousands of images that are found in large image databases. Therefore, having a system that presents the user with all the images in the collection is obviously very limited. In our proposed eID system, the number of image points is reduced due to the use

of cluster prototypes being displayed instead of the entire image collection. The data density will also be improved since the cluster prototypes represent clusters of similar images that were formed by maximizing the dissimilarities between the different groups of images. Since the user will be presented with much less information, it should be much easier and simpler to explore the image database: having a general view of the database, the user obtains *details on demand* by just ‘click on’ actions performed on the eIDs.

The remainder of the paper is organized as follows. Section 3 describes the eID system. Section 4 presents how the proposed system can be used for image annotation and retrieval. Section 5 considers the effectiveness of the system and the considerations are expounded with experiments on a database of 2100 images. Finally, we conclude with some final comments and a note on future work.

### **3. THE EID SYSTEM DESCRIPTION**

#### **3.1. FEATURE EXTRACTION**

This stage requires making choices about representing or coding the pixel values that serve as inputs to the data discovery stage. The representation choices have a great bearing on the kinds of patterns that will be discovered by the next stage.

While the proposed procedure works with any type of low-level feature representation of images, we describe our system using color information. Color feature is one of the most widely used visual features in Image Retrieval. It is relatively robust to background complication and independent of image size and orientation (Rui et al., 1999). Some representation studies of color perception and color spaces can be found in McCamy et al. (1976), Miyahara (1998) and Wang et al. (1997). In this paper, we use the Color-WISE representation by Sethi et al. (1998) in which the representation is guided primarily on three factors. First, the representation must be closely related to human visual perception since a user determines whether a retrieval operation in response to an example query is successful or not. Color-WISE uses the HSV (hue, saturation, value) color coordinate system that correlates well with human color perception and is commonly used by artists to represent color information present in images. Second, the representation must encode the spatial distribution of color in an image. Because of this consideration, Color-WISE system relies on a fixed partitioning scheme. This is in contrast with several proposals in the literature suggesting color-based segmentation to characterize the spatial distribution of color information (Smith and

Chang, 1996b). Although the color-based segmentation approach provides a more flexible representation and hence more powerful queries, we believe that these advantages are outweighed by the simplicity of the fixed partitioning approach. In the fixed partitioning scheme, each image is divided into  $M \times N$  overlapping blocks as shown in Figure 2.

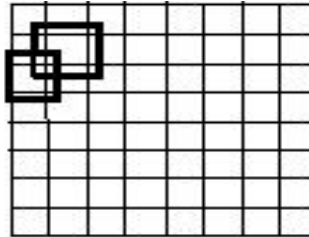


Figure 2: The fixed partitioning scheme with overlapping blocks

The overlapping blocks allow a certain amount of ‘fuzzy-ness’ to be incorporated in the spatial distribution of color information, which helps in obtaining better performance. Three separate local histograms (hue, saturation and value) for each block are computed. The third factor considered by the Color-WISE system is the fact that the representation should be as compact as possible to minimize storage and computational efforts. To obtain a compact representation, Color-Wise system extracts from each local histogram the location of its area-peak. Placing a fixed-sized window on the histogram at every possible location, the histogram area falling within the window is calculated. The location of the window yielding the highest area determines the histogram area-peak; this value represents the corresponding histogram. Thus, a more compact representation is obtained and each image is reduced to  $3 \times M \times N$  values ( $3$  represents the number of histograms). Fig. 3 shows two images and their Color-WISE representations using  $8 \times 8$  blocks.

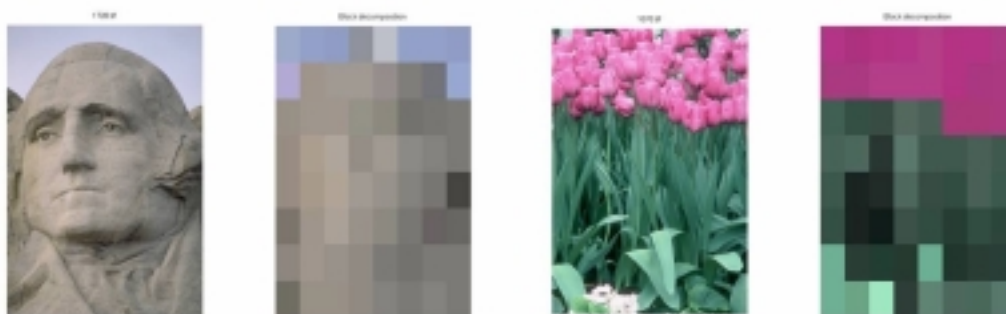


Figure 3: Two examples of color images and their Color-WISE approximations

## 3.2. HIERARCHY OF CLUSTERS

We use a variation of K-means clustering from Duda and Hart (1973) to build a hierarchy of clusters. The variation of K-means algorithm is required since the color triplets (hue, saturation, and value) derived from RGB space by non-linear transformation, are not evenly distributed in the HSV space; the representative of a cluster calculated as a centroid also does not make much sense in such a space. Instead of using the Euclidean distance, we need to define a measure that is closer to the human perception in the sense that the distance between two color triplets is a better approximation to the difference perceived by humans.

We present the used similarity metric and how the cluster representatives are calculated in Sections 3.2.1 and 3.2.2, respectively. The used splitting criterion and cluster validity index are explained in Section 3.2.3.

### 3.2.1 Color similarity metric

Clustering methods require that an index of proximity or associations be established between pairs of patterns according to Jain and Dubes (1998); a proximity index is either a similarity or dissimilarity. The more two images resemble each other, the larger a similarity index and the smaller a dissimilarity index will be.

Since our retrieval system is designed to retrieve the most similar images with a query image, the proximity index will be defined with respect to similarity. Different similarity measures have been suggested in the literature to compare images (Faloutsos et al, 1994, Swain and Ballard, 1991).

We are using in our clustering algorithm the similarity measure that, besides the perceptual similarity between different bins of a color histogram, recognizes the fact that human perception is more sensitive to changes in hue values (Sethi et al., 1998). It also recognizes that human perception is not proportionally sensitive to changes in hue value.

Let  $q_i$  and  $t_i$  represent the block number  $i$  in two images  $Q$  and  $T$ , respectively. Let  $(h_{q_i}, s_{q_i}, v_{q_i})$  and  $(h_{t_i}, s_{t_i}, v_{t_i})$  represent the dominant hue-saturation-value triplet value of the selected block in the two

images  $Q$  and  $T$ . The block similarity (Fig.4) is defined by the following relationship in Sethi et al. (1998):

$$S(q_i, t_i) = \frac{1}{1 + a * D_h(h_{q_i}, h_{t_i}) + b * D_s(s_{q_i}, s_{t_i}) + c * D_v(v_{q_i}, v_{t_i})} \quad (1)$$

The functions  $D_h$ ,  $D_s$  and  $D_v$  measure the block-similarity with respect to hue, saturation and value components. The constants  $a$ ,  $b$  and  $c$  define the relative importance of hue, saturation and value when evaluating the similarity. Since human perception is more sensitive to hue, a higher value is assigned to  $a$  than to  $b$ .

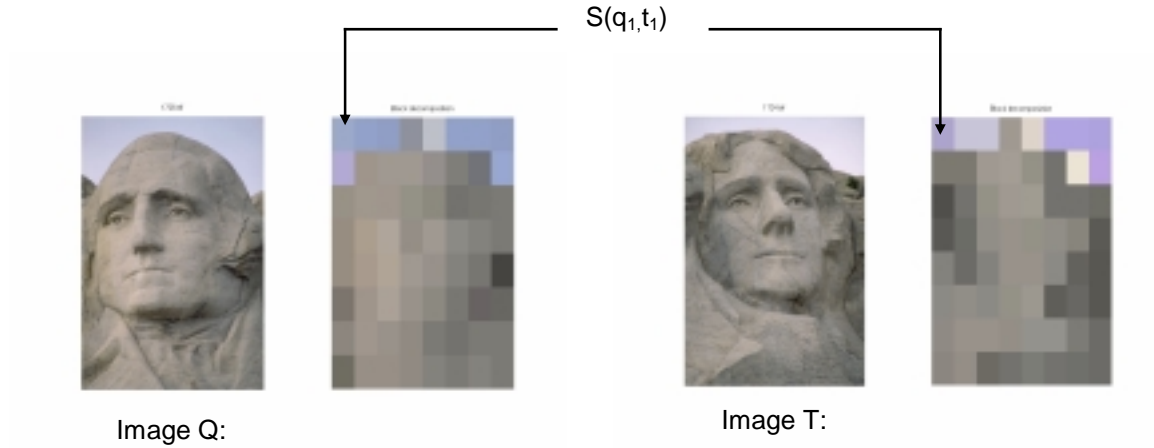


Figure 4: Block-level similarity calculation between two images

The following function is used to calculate  $D_h$ :

$$D_h(h_{q_i}, h_{t_i}) = 1 - \cos^k \left( \left( \frac{1}{256} \right) * \|h_{q_i} - h_{t_i}\| * \frac{\pi}{2} \right) \quad (2)$$

The function  $D_h$  explicitly takes into account the fact that hue is measured as an angle. Through empirical evaluations,  $k=2$  provides a good non-linearity in the similarity measure to approximate the subjective judgment of the hue similarity.

The saturation similarity (Sethi et al., 1998) is calculated by:  $D_s(s_{q_i}, s_{t_i}) = \frac{\|s_{q_i} - s_{t_i}\|}{256}$  (3)

The value similarity is calculated by using the same formula as for saturation similarity. Using the similarities between the corresponding blocks from image  $Q$  and image  $T$ , the similarity between two images is calculated by the following expression (Sethi et al., 1998):

$$S(Q, T) = \frac{\sum_{i=1}^{3 \times M \times N} m_i S(q_i, t_i)}{\sum_{i=1}^{3 \times M \times N} m_i} \quad (4)$$

The quantity  $m_i$  in the above expression represents the masking bit for block  $i$  and  $M \times N$  stands for the number of blocks.

### 3.2.2 Cluster prototypes as electronic IDentifications

The cluster prototypes are designed to summarize the clusters in a manner that is suited for quick human comprehension of the database. They will inform the user about the approximate region in which clusters and their descendants are found. Having the cluster prototypes as electronic identifications (*eID*) giving access to their descendants by ‘click on’ actions, the system will provide users with summaries of the image collection and details on demand.

We define the cluster prototype to be the most similar image to the other images from the corresponding cluster; in another words, the cluster representative is the *clustroid* point in the feature space, i.e., the point in the cluster that maximizes the sum of the squares of the similarity values to the other points of the cluster. If  $C$  is a cluster, its clustroid  $M$  is expressed as:

$$M = \arg \left( \max_{I \in C} \sum_{J \in C} S^2(I, J) \right) \quad (5)$$

Here  $I$  and  $J$  stand for any two images from a cluster  $C$  and  $S(I, J)$  is their similarity value calculated using (4). We use *arg* to denote that the clustroid is the argument (image) for which the maximum of the sums is obtained.

### 3.2.3 Splitting Criterion

To build a partition for a specified number of clusters  $K$ , a splitting criterion is necessary to be defined.

Since the hierarchy aims to support similarity searches, we would like nearby feature vectors to be collected in the same or nearby nodes. Thus, the splitting criterion in our algorithm will try to find an optimal partition that is defined as one that maximizes the criterion sum-of-squared-error function:

$$J_e(K) = \sum_{k=1}^K w_k \sum_{I \in C_k} S^2(I, M_k), \text{ where } w_k = \frac{I}{n_k} \quad (6)$$

$M_k$  and  $I$  stand for the clustroid and any image from cluster  $C_k$ , respectively;  $S^2(I, M_k)$  represents the squared of the similarity value between  $I$  and  $M_k$ , and  $n_k$  represents the number of elements of cluster  $C_k$ .

The reason of maximizing the criterion function comes from the fact that the proximity index measures the similarity; that is, the larger a similarity index value is, the more two images resemble one another.

Once the partition is obtained, in order to validate the clusters, i.e. whether or not the samples form one more cluster, several steps are involved. First, we define the null hypothesis and the alternative hypothesis as follows:  $H_0$ : there are exactly  $K$  clusters for the  $n$  samples, and  $H_A$ : the samples form one more cluster. According to the Neyman-Pearson paradigm (Rice, 1995) a decision as to whether or not to reject  $H_0$  in favor of  $H_A$  is made based on a statistics  $T(n)$ . The statistic is nothing else than the cluster validity index that is sensitive to the structure in the data:

$$T(n) = \frac{J_e(K)}{J_e(K+1)} \quad (7)$$

To obtain an approximate critical value for the statistic that is the index is large enough to be 'unusual', we use a threshold that takes into account that, for large  $n$ ,  $J_e(K)$  and  $J_e(K+1)$  follow a normal distribution.

Following these considerations, we consider the threshold  $\tau$  defined in Duda and Hart (1973) as:

$$\tau = 1 - \frac{2}{\pi * d} - \alpha * \sqrt{\frac{2 * \left(1 - \frac{8}{\pi^2 * d}\right)}{n * d}} \quad (8)$$

The rejection region for the null hypothesis at the  $p$ -percent significance level is:

$$T(n) < \tau \tag{9}$$

The parameter  $\alpha$  in (8) is determined from the probability  $p$  that the null hypothesis  $H_0$  is rejected when it is true and  $d$  is the sample size. The last inequality provides us with a test for deciding whether the splitting of a cluster is justified.

### 3.3. IMAGE DATABASE VISUALIZATION USING MULTIDIMENSIONAL SCALING

Summarizing the previous section, we extracted the most representative images, the eIDs, at different levels of details and thus, the difference between presenting only eIDs and presenting the whole collection is not purely quantitative, but also qualitative. However, eIDs and images are still represented in a high dimensional feature space,  $F$  ( $\dim(F) = 3 \times M \times N$ ), that is difficult to interpret or visualize. Therefore, it is necessary to find a spatial representation of them in a low-dimensional space.

We use Multi-dimensional Scaling technique to map the high-dimensional feature map to a 2-dimensional visualization space  $\vartheta$ . By definition (Cox and Cox, 1994), the technique consists of the search  $\phi$  for a low-dimensional space (Fig.5), in which the points represent the objects (one point representing one object), and such that the distances  $d_{rs}$  between points match as well as possible the dissimilarities  $\delta_{rs}$  between objects in the original space;  $r, s$  stand for the names of two different objects.

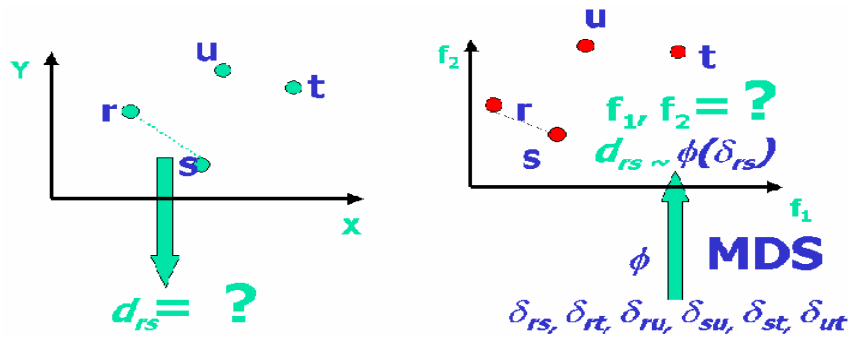


Figure 5: Graphical interpretation of MDS: the geometrical space from the left side is meant to illustrate the usual situation when, given the coordinates of the points in a space, the task is to find the distances between them. The MDS is solving the reverse problem: given the distances between objects, it finds the

configuration of points in a lower dimensional space ( $f_1, f_2$  represent the axes of the lower space obtained by MDS) such as the distances between points in the new space “match” as much as possible the distances between objects in the original space.

In our application, the definition of an object in the original feature space depends on the exploration task:

- If the task is to browse the image database to get a general view about the image database, an object is an eID image and the non-metric MDS approach is applied for every level of the hierarchy.
- If the task is to retrieve images the most similar with a query image, then the objects are the retrieved images and the approach is used to map the query results in a 2-dimensional space.
- If the task is to annotate images, the approach is used for the visualization of every cluster and thus, an object is an image of a particular cluster.

For every task, it is important that the configuration of points in the lower space to be closer to the human perception judgments rather than actual measured distances. Consequently, the dissimilarities between images are calculated by the inverse of non-Euclidian measure defined by formula (4) and thus, a non-metric approach of MDS is used to find the configuration of points in the 2-dimensional space. The non-metric model (introduced by Kruskal in 1964) relaxes the above definition of MDS and attempts to only preserve the rank orders of the dissimilarities; the transformation to a lower dimensional space can be arbitrary, but must obey the monotonic constraint:

$$\delta_{rs} < \delta_{tu} \Rightarrow \phi(\delta_{rs}) \leq \phi(\delta_{tu}), \text{ for all } r, s, t, u \text{ objects in the space} \quad (10)$$

In the next section, we show how our system can be used as a semi-automatic tool for image annotation and retrieval.

#### **4. EXTENDED FUNCTIONALITIES OF THE EID SYSTEM**

The use of clustering and the MDS approach at the cluster level gives the system the capability of creating and visualizing contexts (groups of similar images) in which meaningful perceptual impressions can be formed about the content and the similarity among images. As a result, instead of trying to give all possible meanings to an image, the system interprets an image in the context in which that image appears. Since we

are using only color information in the current implementation of the *eID* system, the meanings associated with images are those that can be derived from color information (such as “sunset”, “arid”, “landscape”, and “marine”). By adding more low-level features to the system (such as texture, shape), additional keywords can be learnt by the system.

There are two scenarios:

- Textual information (keywords, labels) is provided with images to describe their visual content. In this case, the most frequent keyword in the corresponding group of image points will be considered the semantic interpretation (Fig. 6). We refer the reader to Stan and Sethi (2001) to learn more about this kind of association between images and keywords.



Figure 6: Cluster visualization using MDS approach: images most likely having same semantic interpretation are close in the exploration space creating contexts for annotation; choosing the relevance region to contain these images, a rule for automatic annotation is extracted for the “sunset” concept.

- There is no textual information provided with the images. In this case, our system acts like a semi-automatic tool for annotation: different clusters of similar images are presented to the person who is in charge with the annotation. The advantage is that the annotator does not have to search the whole database to look for similar images with the image that has to be annotated. The group

of images in which the image appears will offer a context in which the image can be interpreted and annotated; if these images have the same semantic interpretation, *that semantic meaning can be associated to the corresponding cluster and also to the most representative low-level features for that cluster*. Formally, by most representative features for a cluster, we mean the low-level features with the lowest standard deviation values (Stan and Sethi, 2001). The relationship between image data, the most representative features, and the most likely visual concepts learned by the system can be visualized in Fig. 7.

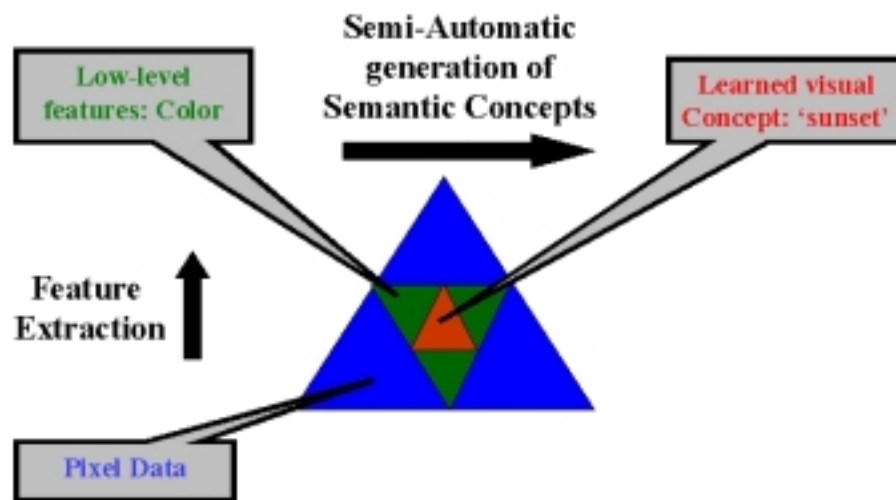


Figure 7: The triangle relationship between data, information, and knowledge: low-level features (information) are extracted purely from pixel data, and knowledge (learned visual concepts) is discovered from the most important low-level features and image contexts

Two of the key benefits of annotation are: semantic browsing and queries using keywords. Since the presented Content-based Image Retrieval (CBIR) system was developed using a hierarchy of clusters, an additional feature of our system will be to enable semantic browsing. On a graphical interface, the keywords may serve as a conceptual summary of the image database; they will function as *landmarks* in the sense that they will help orient and direct the user by providing pointers during the browsing process.

Its counterpart, queries using text information is much easier to use and understand for users who are not familiar with image domain characteristics. For example, if a user made a query using the keyword “sunset”, the system should retrieve images that have “sunset” in the image tag; the user would then choose

an image from the retrieved ones, which will be the new query by image example, whose result will be the closest to the users' conceptualization of the original textual query. Therefore, it helps reduce the lack of coincidence between what user expects the system should retrieve and the actual retrieved images.

## 5. EXPERIMENTAL RESULTS

In this section, we present our preliminary results obtained in evaluating the *eID* system on a general purpose database of 2100 images. Every image is partitioned in 8 by 8 blocks; for every block, three-color histograms are calculated (hue, saturation, and value) and the locations of the histogram's area peaks represent the tri-value of the corresponding block. Therefore, the color vector representation of each image is a one-dimensional array with  $3*8*8$  elements,  $[H_1, H_2, \dots, H_{64}, S_1, S_2, \dots, S_{64}, V_1, V_2, \dots, V_{64}]$  where the index stands for the block number when counted in the image from left to right and from top to bottom. Figure 8 shows two images and their color representation in a matrix form.

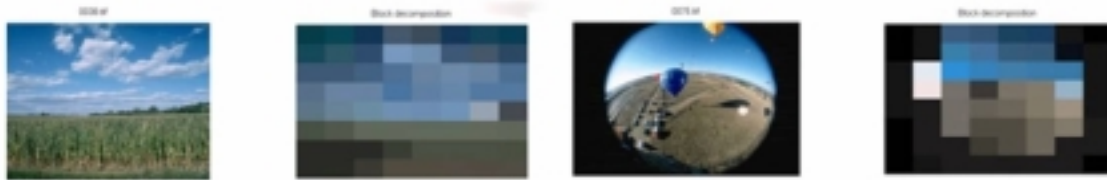


Figure 8: Two examples of original images and their Color-WISE representation

The corresponding feature vectors of the 2100 images are randomly split in two sets: the training set is comprised of 67% of the database while the testing set is 33% of the image database. The training set was used to derive the hierarchy of clusters and to learn the mappings between the low-level features and textual descriptors as described in Stan and Sethi (2001). The testing set was used to validate the performance of the annotation and retrieval of the *eID* system.

In order to obtain the first level (global layer) of the hierarchy, we apply k-means clustering algorithm for  $k = 2, 3, \dots$  and at each consequent  $k$ , the cluster validity is checked to ensure that the number of elements in every cluster is a moderate one and the sum-of-squared-error criterion is satisfied as depicted in (6). The values of the constants  $a$ ,  $b$  and  $c$  used to calculate the similarities (4) are experimentally determined as being 2.5, 0.5 and 3, respectively. Comparing the values of the test statistic (7) and the values of the threshold (8) with respect to inequality (9), the possible number of clusters for different small values of the

significance level is obtained. Following the theoretical part from section 3.2.3, since the value for  $K = 31$  is greater than the threshold for consecutive small values of  $p$ , we choose the value of  $K$  to be 30. Next, we obtain lower levels (regional layer) of the hierarchy by considering the candidate nodes (represented by  $eIDs$ ) for splitting clusters having at least 30 images (at least 2% out of the training set) by applying the k-means clustering algorithm again. The minimum number of elements in every cluster is decided as a compromise between the size of the terminal nodes and the number of upper nodes in the hierarchy. In the current implementation, we end up with a search tree having 81 nodes ( $eIDs$ ) and 4 levels: the global layer has 30  $eIDs$  and they provide a general view about the image database; the regional layers have 40 and 11  $eIDs$ , respectively and they are the keys to obtain details on demand about the image collection. The average number of images per terminal node is equal to 40.

Let us consider the task of browsing the image collection. Initially, the user is presented with the first level of the hierarchy representing the global view of the collection (Fig.9).



Figure 9: The first layer of the hierarchy giving the global view of the image database. (The numbers above each image represent the cluster ID at the first level and the clustroid ID in the image collection, respectively).

Since we used only color information, the global view shows the different color compositions present in the database; it can be noticed that the database consists mostly of images with dominating blue color. The use of MDS for visualization of the hierarchy levels allows the user to see the perceptual similarity between clustroids and the fact that the color composition of the database changes gradually from blue to green and further, to brown, yellow and red.

Further, let us assume that the user is interested to see more images that are similar to the *eID* representing the concept “sunset”, which represents the set of keywords *sunset*, *sunrise*, *dusk*, and *dawn*. If the user clicks on the “sunset” *eID*, its representatives will be displayed on the next level. Fig. 10 shows the representatives of the clusters in which the “sunset” cluster was split and anyone of these four *eIDs* will point to leaves (images that are most similar to a selected clustroid) in the hierarchy as shown in Fig. 11-14. All four *eIDs* have similar semantic meaning (*sunset/sunrise/dusk/dawn*) with the *eID* parent.



Figure 10: Set of *eIDs* representing the concept “sunset” on the second level displayed after the user clicks on the clustroid ID 1701 in Figure 9.



Figure 11: Set of images/leaves representing the concept “sunset” on the third level displayed after the user clicks on the clustroid ID 1701 in Figure 10

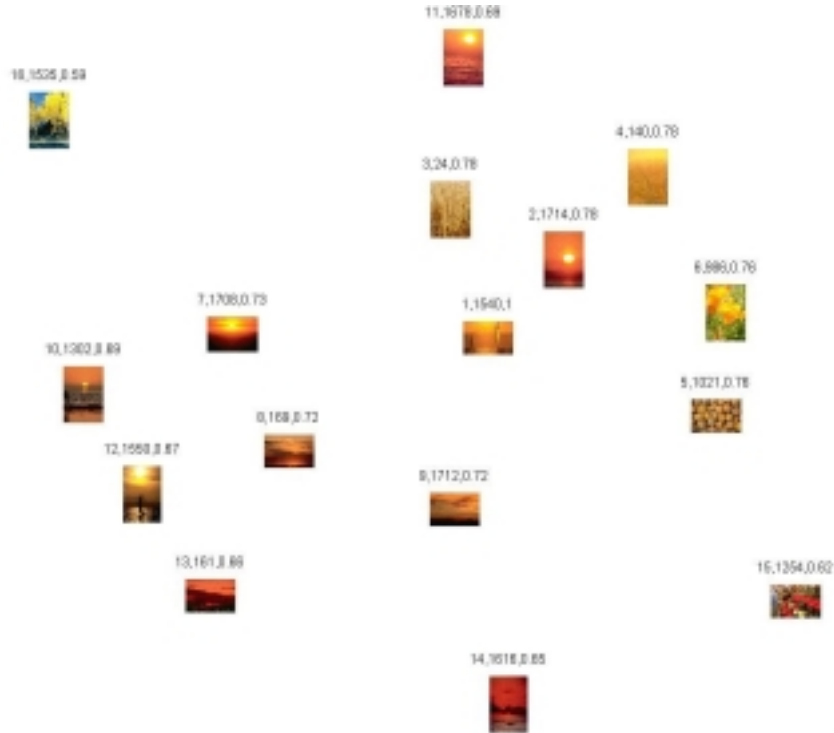


Figure 12: Set of images/leaves representing the concept “sunset” on the third level displayed after the user clicks on the clustroid ID 1540 in Figure 10



Figure 13: Set of images/leaves representing the concept “sunset” on the third level displayed after the user clicks on the clustroid ID 1696 in Figure 10



Figure 14: Set of images/leaves representing the concept “sunset” on the third level displayed after the user clicks on the clustroid ID 1345 in Figure 10

If the task is to do annotation, instead of searching the whole collection to assign a label to an image, the terminal nodes are searched and thus, on average, only 2% of the images are searched for semantic similarity at once. For instance, Fig.11-14 can be considered as four different contexts for image annotation. We are currently in the process of measuring the performance of the manual annotation (using the contexts produced by the hierarchy) with respect to time per image to annotate it and the agreement on annotation given by different people.

In parallel, we have annotated the images by looking at the entire collections and the annotations were used to derive the rules between low-level features and semantic concepts. For each cluster that is a terminal node, the components of the feature vectors are ordered in increasing order of standard deviations and the most frequent keywords were calculated as well as shown in Stan and Sethi (2001). We extracted rules for some semantic concepts such as “sunset”, “landscape”, “arid”, and “marine”. For illustration purposes, we present results from the clusters in which the optimal textual characterization is “sunset”. For the entire hierarchy and set of rules, we refer the reader to Stan (2002).

There are 152 “sunset” images in the database out of which 107 belong to the training set and 45 to the testing set. The hierarchy of clusters has six clusters whose most frequent concept is “sunset”; they cover 56% of the “sunset” images from the training set, which is depicted in Table 1.

Table 1: Statistics for the “sunset” clusters; the notation for the Cluster ID stands for the cluster number at different levels of the hierarchy: for example, 09\_03\_02 denotes a cluster on the third level being a sub-cluster of cluster 09\_03 whose parent is cluster 09 on the first level of the hierarchy.

Cluster ID	# of Images	# of "sunset"	% of "sunset" images
12_01	27	19	70.37%
12_02	13	5	38.46%
12_03	16	11	68.75%
12_04	10	6	60.00%
20_01	19	14	73.68%
09_03_02	9	5	55.56%
<b>Total # of images</b>	94	60	N/A

For every cluster, using the mapping function from Stan and Sethi (2001) between the most significant low-level features and the concept “sunset”, a mapping rule is derived. Tables 2 shows the accuracy of the annotation for the cluster, training, and test sets using the rules whose number of low-level features is given in the same table.

Table 2: Rule sizes and accuracy for *Sunset/Sunrise/Dusk/Dawn* Annotation

Precision (%)	12_01	12_02	12_03	12_04	20_01	9_03_02	Average/ rule
<b>Cluster Data</b>	70.59	80.00	80.00	100.00	80.00	83.33	82.32
<b>Training Set</b>	78.26	69.23	61.11	72.73	80.00	80.00	73.56
<b>Test Set</b>	84.21	80.00	40.00	33.00	100.00	71.43	68.11
<b>Nr. Features(%)</b>	23.44	23.44	41.67	10.42	15.63	15.63	21.70

For example, the rule derived from Cluster 12\_04 has the classification precision of 100% for the cluster data, 72.73% for the training set, and 33% for the test set for a number of features equal to 20. As we can see from Table 2, we obtain high accuracy results for the cluster and training set (82.32% and 73.56%, respectively); even for the test set, we obtain relatively high accuracy (68.11%) on average per rule using only about 22% of original features.

Higher accuracy can be obtained if the number of features per rule is increased, however the number of images retrieved will decrease. In deciding on the number of features, we looked to find rules whose size will produce reasonable (more than 70%) classification accuracy on the cluster data. The results show that our approach has good accuracy in terms of precision of annotation. On the other hand, there might be images that do not receive any keyword assignment by being assigned to clusters whose textual characterization is different than their meaning. In our case, approximately 76 out of 107 “sunset” images in the training set and 29 out of 45 in the test set were covered by the above rules. More images will receive a keyword assignment if a hierarchy of clusters with more levels of details will be developed.

Fig. 15 shows the most important features and the most representative image for the rule derived from cluster 12\_04. Some images annotated as “sunset” by the same rule can be seen in Fig. 16.

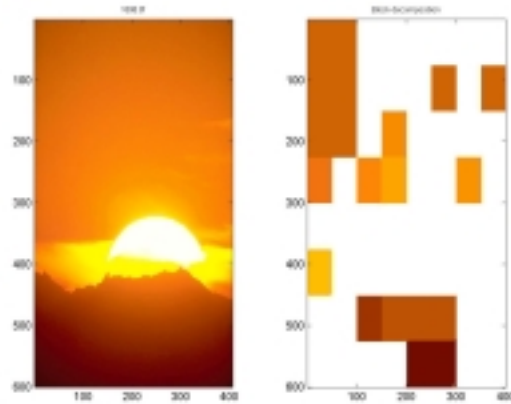


Figure 15: Example of a rule (right image) graphically represented and the  $eID$  (left image) which most closely matches the rule derived from cluster 12\_04.



Figure 16: Sample images indexed as “sunset” by the rule from Figure 15

Let us consider the case of retrieval by query image. When compared to a retrieval system that uses global color histogram for image representation and histogram intersection (Swain and Ballard, 1991) as a similarity metric, our system gives better retrieval results. The explanation resides on the fact that the  $eID$  system encodes the spatial distribution of color in an image, and uses the block-level similarity measure that takes into account the human sensitivity with respect to the hue component. Fig. 17 - 20 show the retrieval results displayed in a linear list when the top left image from every figure is chosen as a query image; the color-wise approach is shown in the first row, while the histogram intersection approach is shown on the second row.



Figure 17: Retrieval results displayed as a linear list. The query image is the on the first position in the list.



Figure 18: Retrieval results displayed as a linear list. The query image is the on the first position in the list.



Figure 19: Retrieval results displayed as a linear list. The query image is the on the first position in the list.



Figure 20: Retrieval results displayed as a linear list. The query image is the on the first position in the list.

In order to show the order of perceptual similarity among the retrieved images, we use the MDS technique again (Fig. 21, Fig.23, Fig.24 and Fig.25).

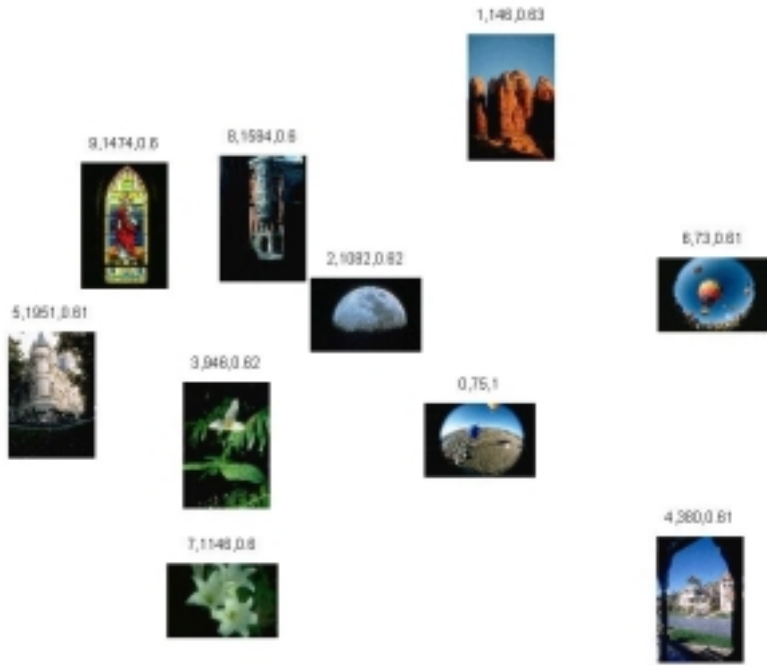


Figure 21: The retrieval results from Fig. 17 displayed using MDS; the query image has the ID of 75; the numbers above the images represent the rank order of the retrieved images, the image ID, and the similarity value, respectively.



Figure 22: The same results as shown in Fig. 21, but using the block representation which is used to calculate the similarity value.



Figure 23: The retrieval results for the query image 466 from Fig. 18 displayed using MDS.



Figure 24: The retrieval results for the query image 532 from Fig. 19 displayed using MDS.



Figure 25: The retrieval results for the query image 415 from Fig. 20 displayed using MDS.

Since the MDS technique cannot be applied for the histogram intersection (the symmetry of the distance matrix is not satisfied), we present the perceptual grouping only for the eID system. We notice that it is possible to visualize the results not only based on the similarity reflected by the color composition, but also by the semantic meaning. This is the real value of MDS when used for visualization of either the different levels of the hierarchy or the retrieval results. Images with similar color content are grouped together in the visualization space which will make the process of annotation more efficient, and at the same time, better understanding the computer's results.

For understanding of the retrieval results, in Fig. 22, we show the block decomposition of the most similar images with the query image from Fig. 21. We notice that the block decomposition consists of black boundary blocks for all retrieved images; furthermore, central blocks of these images give the position of the images in the visualization space. For example, images with the IDs 946 and 1146 are near each other in the visualization space because they have similar color blocks in the middle of the partition.

## 5. SUMMARY

As an overview, our proposed system allows users to explore the image database in order to browse, annotate, and formulate a textual or image query. We built a hierarchy of clusters, whose prototypes were represented at every level using 2-dimensional MDS maps such that the similarities' rank order from the original low-level feature space were preserved. These maps were used to browse the image collection for finer to finer details. Since the maps represent only the cluster prototypes and their number is small, the user is presented with summary views of the image database and *details on demand* are obtained by 'click on' actions performed on the *e*IDs. From point of view of annotation, 2-dimensional MDS maps are used again but for visualizing image contexts in which meaningful perceptual impressions can be formed about the content and the similarity among images.

As future work, we want to implement our system using other low-level features in addition to those based on color information and experiment it on larger image databases.

## REFERENCES

- Bach, J.R., Fuller, C., Gupta, A. et al. (1996). The Virage Image search engine: An open framework for image management. *Proceedings of SPIE Storage and Retrieval for Image and Video Databases*, 76-87
- Chen, J.-Y., Bouman, C.A., & Dalton, J.C. (1998). Similarity pyramids for browsing and organization of large image databases. *Human Vision and Electronic Imaging III*, 563-575.
- Cox, T. & Cox, M. (1994). *Multidimensional Scaling*. Chapman & Hall.
- Craver, S., Yeo, B.-L., & Yeung, M.M. (1998). Image browsing using data structure based on multiple space-filling curves. *Proceedings of the Thirty-six Asilomar Conference on Signals, Systems, and Computers*.
- Duda, R.O. & Hart, P.E. (1973). *Pattern classification and scene analysis*. John Wiley & Sons, Inc.
- Faloutsos, C., Equitz, W., & Flickner, M. (1994). Efficient and Effective Querying by Image Content. *Journal of Intelligent Information Systems* 3, 231-262.

- Flickner, M., Sawhney, H., Niblack, W. et al (1995). Query by image and video content: The QBIC system. *IEEE Computer*.
- Jain, A.K. & Dubes, R.C. (1998). Algorithms for Clustering Data. *Prentices Hall Advanced Reference Series*.
- James D. (1993). Content-based retrieval in multimedia imaging. Proceedings of SPIE Storage and Retrieval for Image and Video Databases.
- Kaski, S., Lagus, K., Honkela, T., & Kohonen, T. (1998). Statistical aspects of the WEBSOM system in organizing document collections. *Computer Science and Statistics 29*, 281-290.
- Kohonen, T. (1997). *Self-Organizing Maps*. Second Edition. Boston: Springer-Verlag.
- Kato T. (1992). Database architecture for content-based image retrieval. *Proceedings of SPIE: Image Storage and Retrieval Systems*, 112-123.
- Ma, W.Y. & Manjunath, B.S. (1997). Netra: A toolbox for navigating large image databases. *Proceedings of IEEE Int. Conf. on Image Processing*.
- MacCuish, J., McPherson, A., Barros J. & Kelly, P. (1996). Interactive layout mechanisms for image database retrieval. *Proceedings of SPIE/IS&T Conf. on Visual Data Exploration and Analysis, III*, (2656), pp. 104-115.
- [McCamy, C.S., Marcus, H., & Davidson, J.G. (1976). A color-rendition chart. *Journal of Applied Photographic Engineering*, 2(3).
- Mehrotra, S., Chakrabarti, K., Ortega, M., Rui, Y. & Huang, T.S. (1997a). Multimedia analysis and retrieval system. *Proceedings of the 3rd International Workshop on Information Retrieval Systems*.
- Mehrotra, S., Rui, Y., Ortega, M., & Huang, T.S. (1997b). Supporting content-based queries over images in MARS. *Proceedings of IEEE Int. Conference on Multimedia Computing and Systems*.
- Minka, T.P. & Picard, R. W. (1996). Interactive learning using a “society of models”. *Proceedings of IEEE Computer Vision and Pattern Recognition*, 447-452.

- Miyahara, M. (1998). Mathematical transformation of (r,g,b) color data to Munsell (h,s,v) color data. *Proceedings of SPIE Visual Communications and Image Processing*, vol. 1001.
- Niblack, W., Barber, R. & et al (1994). The QBIC project: Querying images by content using color, texture, and shape. *Proceedings of SPIE Storage and Retrieval for Image and Video Databases*.
- Pentland, A., Picard, R.W., & Sclaroff, S. (1996). Photobook: Content-based manipulation of image databases. *International Journal of Computer Vision*.
- Rice, J.A. (1995). *Mathematical Statistics and Data Analysis*. Duxbury Press.
- Rubner, Y., Guibas, L., & Tomasi, C. (1997). The earth's mover distance, multi-dimensional scaling, and color-based image retrieval. *Proceedings of the ARPA Image Understanding Workshop*.
- Rui, Y., Huang, T.S., & Chang, S.F. (1999). Image Retrieval: Past, Present, and Future. *Journal of Visual Communication and Image Representation*, 10, 1-23.
- Rui, Y., Huang, T.S., Mehrotra, S., and Ortega, M. (1997a). A relevance feedback architecture in content-based multimedia information retrieval systems. *Proceedings of IEEE Workshop on Content-based Access of Image and Video Libraries*, in conjunction with IEEE CVPR'97.
- Rui, Y., Huang, T.S., Mehrotra, S., and Ortega, M. (1997b). Automatic matching tool selection using relevance feedback in MARS. *Proceedings of Second International Conference on Visual Information Systems*.
- Sethi, I. K., Coman, I., Day, B. et al (1998). Color-WISE: A system for image similarity retrieval using color. *Proceedings of the SPIE: Storage and Retrieval for Image and Video Databases*, 3132, 140-149.
- Sethi, I.K. & Coman, I. (1999). Image retrieval using hierarchical self-organizing feature maps. *Pattern Recognition Letters* 20, 1337-1345.
- Sethi, I.K., Coman, I., & Stan, D. (2001). Mining association rules between low-level image features and high-level concepts. *Proceedings of the SPIE: Data Mining and Knowledge Discovery III*, 279-290.

- Smith, J.R. & Chang, S.-F. (1996a). VisualSEEk: A fully automated content-based query image system. *Proceedings of ACM Multimedia '96*.
- Smith, J.R. & Chang, S.F. (1996b). Tools and Techniques for Color Image Retrieval. *Proceedings of the SPIE: Storage and Retrieval for Image and Video Databases IV*, vol. 2670, 381-392.
- Stan, D. & Sethi, I.K. (2001). Mapping low-level image features to semantic concepts. *Proceedings of SPIE: Storage and Retrieval for Media Databases*, 172-179.
- Stan, D. (2002). eID: A system for Exploration of Image Databases. *Ph.D. Thesis*, Oakland University.
- Swain, M.J. & Ballard, D.H. (1991). Color Indexing. *International Journal of Computer Vision*, vol. 7(1), 11-32.
- Wang, J., Yang, W.-J., & Acharya, R. (1997). Color clustering techniques for color-content-based image retrieval from image databases. *Proceedings of IEEE Conf. on Multimedia Computing and Systems*.
- Zhang, H. & Zhong, D. (1995). A scheme for visual feature based image indexing. *Proceedings of SPIE/IS&T Conference on Storage and Retrieval for Image and Video Databases III*, vol. 2420, 36-46.