Local vs. Global Histogram-Based Color Image Clustering

Bongani Malinga, Daniela Raicu, Jacob Furst
DePaul University
School of Computer Science
Chicago, IL
bmalinga@students.depaul.edu, draicu@cs.depaul.edu, jfurst@cti.depaul.edu

Abstract

In this paper, we present two image clustering techniques to automatically group color images that correlate with semantic concepts. This work goes towards satisfying the ever growing need for techniques that are capable of automatically generating semantic concepts for images from their visual features. We present two techniques and evaluate their relative performances based on the perceptual similarity among images that are clustered together. The first technique is based on the localized histogram information while the second approach uses global histogram information to characterize the images. Experiments using a 2100 image database are presented to show the relative effectiveness of the presented systems. Preliminary results show that the local histogram approach gives better clustering results and its characterization of images is more closely related to the way in which humans perceive images.

1 Introduction

In recent years, there has been a growing interest in developing effective methods for content-based image clustering and retrieval [1,2,7,8,13,14]. This interest has been motivated by the need to efficiently manage large image databases and efficiently run image retrievals to get the best results without exhaustively searching the global database each time. This leads to huge savings in time and money, especially in fields where the bulk of working databases are image files or any kind of media whose contents cannot be adequately described by simple keywords or short texts. As digital media become more popular, corporations and individuals gather an increasingly large number of digital images. As a collection grows to more than a few hundred images, manual (visual) clustering becomes impossible and the need for formalized clustering and retrieval becomes crucial. These clustering techniques are also necessary for handling very large un-clustered or untagged pre-existing databases. The systems presented in this paper satisfy this need by automatically grouping images from their low-level visual features.

Images are typically read as RGB models and then transformed into the HSV color model. The RGB [3] color model is composed of the primary colors Red, Green, and Blue. They are considered the "additive primaries" since the colors are added together to produce the desired color. The HSV [3] color model was used in this work. This color model defines colors in terms of three constituent components; hue, saturation and value. The hue and saturation components are intimately related to the way human eye perceives color because they capture the whole spectrum of colors. The Value represents intensity of a color, which is decoupled from the color information in the represented image.

Clustering [5] is an unsupervised way of data grouping using a given measure of similarity. Clustering algorithms attempt to organize unlabeled feature vectors into clusters or groups such that samples within a cluster are more similar to each other than to samples belonging to different clusters. Since there is no information given about the underlying data structure or the number of clusters, there is no single solution to clustering; neither is there a single similarity measure to differentiate all clusters. Because of this reason, there is no theory that describes clustering uniquely. Even though there is an increasing interest in the use of clustering methods in image processing, clustering has a rich history in many disciplines such as biology, psychology and marketing. A number of books on clustering have been published [5, 6], in addition to some useful and influential papers [4, 8, 13, 14].

One commonality in all clustering techniques is that there are three main stages that have to be addressed in series in order to perform effective clustering: image acquisition, feature extraction, and calculation of pattern similarity measures between image representations [8]. With many possible variations to these stages, no one method has been proven to yield the optimal results. Researchers have explored the field of image clustering with many different combinations of the variations in these stages. This research project explores and compares two of the methods as a means to perform image clustering such that each cluster provides essentially the same information about the images in it.

The remainder of the paper is organized as follows: Section 2 describes the local histogram-based approach; Section 3 describes the global histograms approach. Section 4 presents and compares the effectiveness of these two approaches where the considerations are based on experiments performed using a database of 2100 color images; finally, we conclude with some final comments and a note on future work.

2 Local histogram-based approach

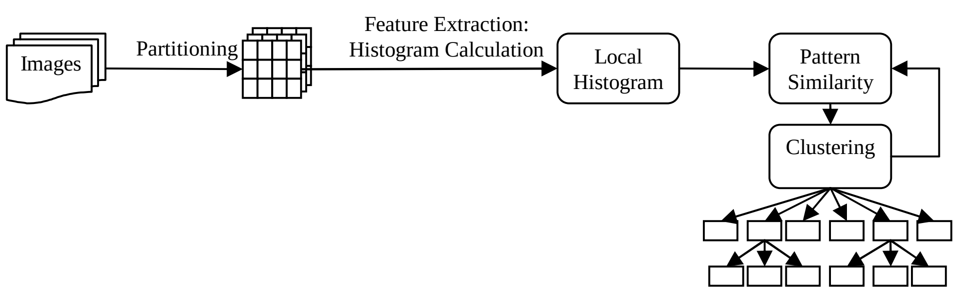

Figure 1: Main steps of the local histogram based system

The goal of this approach [8] is to partition the image into M x N local blocks, extract the features for each block and calculate the pattern similarity measures that are used in the applied variation of the K-Means clustering technique [8]. Figure 1 shows the stages featured in this clustering approach.

2.1 Image partitioning

The representation of images 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. Therefore the image representation must encode the spatial distribution of color in an image. Because of this consideration, this approach relies on a fixed partitioning scheme to capture the special distribution of color. This is in contrast with several proposals in the literature [13] suggesting color-based segmentation to characterize the spatial distribution of color information. Although the color-based segmentation approach provides a more flexible representation and hence more powerful queries, these advantages are outweighed by the simplicity of the fixed partitioning approach. In the fixed partitioning scheme, each image is divided into M x N overlapping blocks as shown in Figure 2.

The overlapping blocks allow a certain amount of fuzziness to be incorporated in the spatial distribution of color information, which helps in obtaining a better performance.

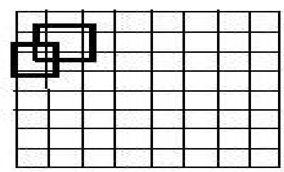

Figure 2: The fixed partitioning scheme with overlapping blocks.

2.2 Feature extraction

Three separate local histograms (hue, saturation and value) for each block (Figure 2) are computed. 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. For each partition in the image, the maximum histograms values for the H, S and V components are used to represent the partitions. Thus, a more compact representation is obtained and each image is reduced to 3 x M x N numbers (3 represents the number of histograms).

2.3 Pattern similarity measure

The more two images resemble each other, the larger a similarity index should be. Different similarity measures have been suggested in the literature to compare images [10, 11, 12]. The clustering algorithm uses 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 [14,], which captures all the colors. It also recognizes that human perception is not proportionally sensitive to changes in hue value, which captures the brightness of colors.

Let q_i and t_i represent the block number *i* in images ***Q*** and ***T***, respectively. Let $\left(h_{q_i}, s_{q_i}, v_{q_i}\right)$ and $\left(h_{t_i}, s_{t_i}, v_{t_i}\right)$ represent the dominant hue-saturation-value triple of two image blocks ***q*** and ***t***, respectively. The block similarity is defined by the following relationship [8]:

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

Here, D_h D_s and D_v represent the functions that measure similarity in hue, saturation and hue. The constants a, b and c define the relative importance of hue, saturation and value in similarity components. Since human perception is more sensitive to hue, a higher value is assigned to a than to b. The following function [8] is used to calculate D_h:

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

The function D_h explicitly takes into account the fact that hue is measured as an angle. Through empirical evaluations, a value of k equal to two provides a good non-linearity in the similarity measure to approximate the subjective judgment of the hue similarity. The saturation similarity is calculated by [8]:

$$D_s\left(s_{q_i}, s_{t_i}\right) = \frac{\left\|s_{q_i} - s_{t_i}\right\|}{256} \tag{3}$$

The value similarity is calculated by using the same formula as for saturation similarity. Using the similarities between the corresponding blocks from the images *Q* and *T*, the similarity between the images is calculated by the following expression [8]:

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

2.4 Clustering

A hierarchy of clusters is developed to build an effective clustering module that solves both high dimensionality and non-Euclidean nature of the used color space. At every level of the hierarchy, the variation of K-Means clustering uses a “clustroid” as a cluster-prototype [8] and the non-Euclidean similarity metric defined by formula (4) to measure the image similarity. For this approach, the term “variation” is used based on the consideration that normal K-means uses the means of the clusters as their prototypes. The resultant clusters are further divided into other disjoint sub-clusters performing organization of information at several levels, going for finer and finer distinctions. The adaptation of the k-means algorithm is required because the color triplets derived from RGB space are not evenly distributed in the HSV space.

3 Global histogram-based approach

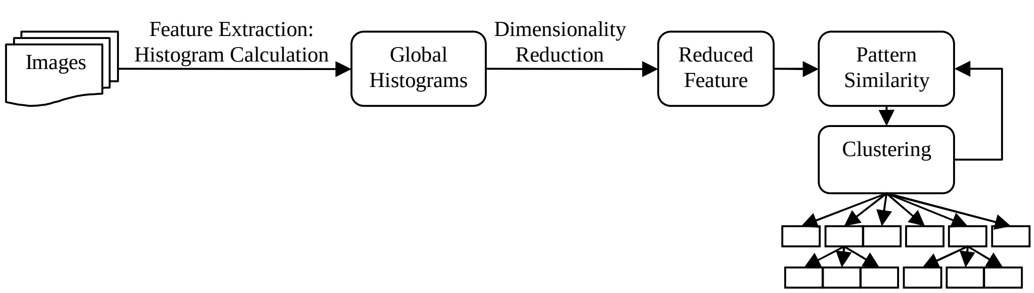

Figure 3: Stages in Image Clustering

The objective of this approach is to study the effectiveness of using the global HSV color space histograms of images as the descriptors in image clustering. The goal is to calculate the

HSV global histograms for all the images, reduce the dimensions of the image descriptor vectors using Principal Component Analysis and calculate the similarity measures between the images. The clustering results are then analyzed to see if the results have any semantic meaning. Figure 3 shows the general stages in this clustering technique from image acquisition all the way to the final clusters. All parts of these stages are implemented in Matlab™ [3].

3.1 Feature extraction

Since images are 3D arrays, there is a need to further process them in order to represent these images in a more compact way. The color image features, which are the color histograms for each corresponding image (one histogram for hue, one for saturation and one for value), are calculated. The histograms for each of the dimensions use 256 bins. These histograms are then concatenated for each image, so that the result is a feature vector with 768 elements, the first 256 being the bins for the hue histogram, the next 256 bins being the ones for the saturation histogram and the rest for the intensity. The image set with 2100 images generates an image description matrix file having 2100 rows and 768 columns.

3.2 Dimensionality reduction: PCA

The image histogram vector can be used as the final image descriptor for clustering. However, the data could be processed even further in order to extract the most important features and provide an opportunity to reduce the feature space from 768 elements to a lesser number of elements. Principal Component Analysis (PCA) is applied on the data and several different sets of the top principal components are retained as the descriptors for clustering. The challenge is to experiment with a different number of components in order to get the best clustering results.

PCA is a multi-variate statistical analysis method used for feature reduction. The new features (principal components) are linear transformations of the original variables (which are normally correlated) and obtained such that they are uncorrelated or orthogonal to each other [9]. Since the image bin descriptors are correlated, they can be represented by a smaller number of principal components because of the redundancy of the descriptors. Therefore, instead of using the entire set of descriptors for clustering, the PCA method will provide the most important principal components to be used for finding the similarities among images.

3.3 Pattern similarity measure

There are a large number of distance measures that have been used in the context of color image processing. Common ones include: Manhattan, Euclidean, histogram intersection and Mahalanobis [10,11,12]. It is widely accepted that the choice of proximity measure can have a profound effect on the performance of any clustering technique. It is also significant for Content-Based Image Retrieval (CBIR) as when querying a multimedia database, we are normally interested in the nearest neighbors, as determined by the implemented distance measure. However, often the choice of distance measure is made without much thought and this presents an area of improvement in image clustering and retrieval systems. Some researchers have implemented different distance measures to accommodate the non-linearity of the color spaces [8,11,12].

In this approach the Euclidean distances are calculated between each of the images' feature representation and the feature representations of the assigned / updated cluster centers. Let ***h*** and ***g*** represent two color histograms. The Euclidean distance between the color histograms ***h*** and ***g*** can be computed as:

$$d^2(h,g) = \sum_A \sum_B \sum_C (h(a,b,c) - g(a,b,c))^2 \qquad (5)$$

where A, B and C represent the three color channels (H,S,V) and a, b and c are the corresponding histogram components of the channels. In this distance formula, there is only comparison between the identical bins in the respective histograms. Two different bins may represent perceptually similar colors but are not compared cross-wise. All bins contribute equally to the distance.

3.4 Clustering

The k-means clustering technique is applied to both the original data (with all 768 descriptors) and the PCA reduced interpretation of the data until no more transfers can occur between clusters. K-Means clustering finds a grouping of the measurements that minimizes the within-cluster sum-of-squares. Each measurement, represented by a vector, is grouped so that it is assigned to one of a fixed number of clusters.

One requirement for applying the K-Means clustering technique is that the value of k must be known. However in this case, as in most other systems, the value of k is unknown and determining this value is often as critical as creating the clusters themselves. In order to determine the appropriate value of k, the clustering technique is executed iteratively with k ranging from two to a reasonable maximum number. Using the results of the iterations, the intra-cluster and inter-cluster statistics are calculated. The intra-cluster similarity measure f^2, for the k clusters is calculated using:

$$f^2 = \sum_{k=1}^{k} \sum_{i=1}^{m_k} (v_i^{(k)} - m^{(k)})^2 \qquad (6)$$

where i is the i th element in each cluster, m_k is the number of elements in cluster k, $m^{(k)}$ is the cluster mean of cluster k and v is the value of the i th observation in cluster k.

The inter-cluster similarity measure b^2, for the k clusters is calculated using:

$$b^2 = \left[\sum_{k=1}^{k} n_k \cdot (m^{(k)})^2 \right] - n \cdot m^2 \qquad (7)$$

where n_k is the number of elements in cluster k, $m^{(k)}$ is the cluster mean of cluster k, n is the number of observations and m is the overall center of the dataset. The ratio of these quantities, f^2/b^2 is then calculated for each set of k clusters. The trend of this ratio gives a very good indication of the value of k suitable for the dataset.

4 Experimental results

For the first approach, the color vector representation of each image has 3*8*8 elements since each image is partitioned into 8*8 overlapping blocks and the image color content is characterized by the three HSV components. We rescale hue and saturation to values between 0 and 255. First, we apply the K-Means algorithm to derive a two-level hierarchy of clusters, and the cluster validity is checked for every cluster. The values of the constants (a, b and c) in formula (1) are experimentally chosen as being 2.5, 0.5 and 0, respectively. The value, or brightness, is not used in the distance calculation to avoid discriminating between the same colors with different brightness. We end up with a hierarchy containing 30 clusters at the first level and 70 clusters at the second level. Detailed results of the local histogram approach can be found in [8].

For the second approach, PCA was applied to the descriptor matrix and Figure 4 shows the cumulative percentage variance captured by the principal components (PC's) and the scree plot. It can be seen from the graph that only about 485 of the 768 principal components were needed to capture all the variability in the data. This was also verified by the similarity of the clustering results obtained from using all, versus retaining 485 principal components.

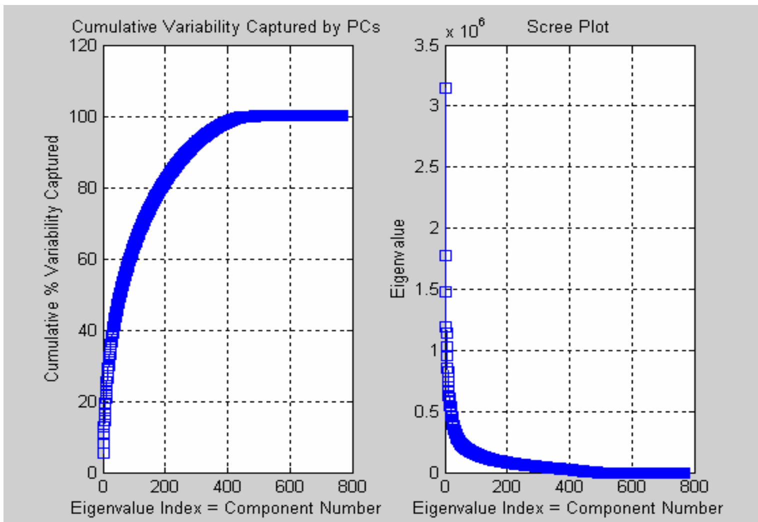

Figure 4: Cumulative Variance Captured and Scree plot

The clustering algorithm was executed several times, each time with the same initial center assignments across a different number of principal components in the data matrix. No significant deterioration was observed in the clustering results as the number of retained principal components was reduced. However, it was observed that the fewer the number of retained principal components, the slower the convergence rate of the clustering algorithm.

To determine the value of k, the b^2, f^2 statistics and their ratio were calculated over a range of k. Figure 5 shows the results of there quantities as k ranged from 2 to 40. As Figure 5 depicts, it was reasonable to choose the value of k anywhere between 20 and 40. However, for the purposes of comparison with the results in [8], the number of clusters (k) at the first level of the hierarchy was set to 30. The same approach was used to create the second level of the hierarchy, where the bigger clusters (with > 50 images) were further divided into smaller classes.

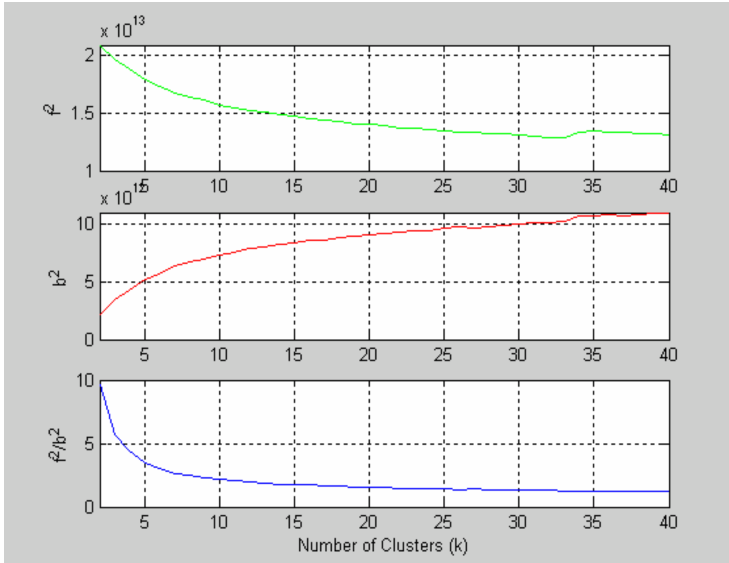

Figure 5: Intra-cluster and inter cluster statistics

At the first level of the hierarchy, good “within-cluster” semantic relations were observed for smaller clusters but for the bigger clusters (with greater than 50 images), there was more randomness to the images and less semantic relations between the images. Table 1 shows the cluster index and the corresponding number of images for the clusters for the first level of the hierarchy.

Cluster Index	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Cluster Size	87	41	122	17	35	119	78	64	81	83	65	47	154	57	11
Cluster Index	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
Cluster Size	23	62	43	30	51	26	80	47	104	120	200	121	56	55	15

Table 1: Cluster index and size for first level of hierarchy

The second level of the hierarchy was created by further clustering the images within the bigger clusters obtained at the first level. This did not yield any significant results because no meaningful sub-clusters were observed visually. This can be explained by the fact that global features can only differentiate at global color distribution level and images that are grouped the same cluster have similar global distributions but may have different spatial distributions. To try and 'fine tune' the clusters for better semantic meaning results without more information (spatial) will not be beneficial.

Figures 6 and 7 show some images contained in two of the clusters that resulted from this approach. Considering the wide variation of image color in the database, this technique showed decent performance, especially at the first level of the hierarchy, given that most of the images within clusters had some discernable visual commonality.

Figure 6: Sample of images for smaller cluster (Cluster # 30)

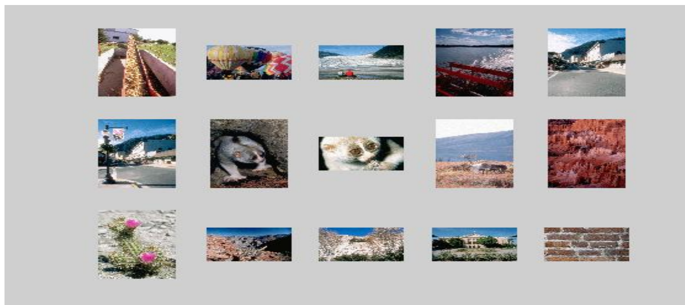

Figure 7: Sample of images from bigger cluster (cluster # 13)

5 Conclusion

In this paper, two techniques for unsupervised clustering of color images were presented. In the first approach, images were partitioned into a fixed number of equal blocks and then represented in terms of their local (block -wise) HSV histogram peaks. An adaptation of k-means clustering using a non-Euclidean similarity metric was applied to discover the natural patterns of the data in the low-level feature space. In the second approach, global image histograms were derived and Principal Component Analysis (PCA) was used to reduce the size of the image descriptor matrix. The k-means clustering algorithm using the Euclidean similarity metric was used to cluster the images. Experiments using a 2100 image database were presented to show the relative effectiveness of the presented systems.

For good results, the representation of images 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. The comparison of the results of these two approaches support this statement because the local histogram approach encoded the spatial distribution of color in an image by applying a fixed partitioning scheme and it yielded better results than the global histogram approach. This conclusion is based on visually assessing the semantic meanings of within-group images from both techniques. This also stems from the observation that no significant benefits were observed when the second hierarchy level of clusters was created for the global approach.

The drawback of using image histogram features is that information about object location, shape, and texture is discarded. This means that images clustered or retrieved by using the global color histogram may not be semantically related even though they might share similar color distributions. Using the local histogram approach improves this situation but combining color histograms with other image features such as shape and texture should improve the overall image clustering performance.

6 Future work

Visual assessment of results is very subjective and becomes cumbersome for large image databases. The primary objective of this project was to compare the performance of two clustering techniques and there still remains a need in the future to derive an objective performance measure that would yield absolute results; especially if there is no a priori knowledge about the classes or annotations contained in the image database.

References

[1] Daniela S. Raicu, Ishwar K. Sethi, "Synobins: An Intermediate Level Towards Annotation and Semantic Retrieval", *EURASIP JASP* (Accepted 2005)

[2] Yixin Chen, James Z. Wang and Robert Krovetz, "CLUE: Cluster-based Retrieval of Images by Unsupervised Learning," *IEEE Transactions on Image Processing*, vol. 14

[3] The Mathworks: Matlab Image Processing Toolbox

[4] S. Krishnamachari and M. Abdel-Mottaleb, "Hierarchical clustering algorithm for fast image retrieval" *IS&T/SPIE Conference on Storage and Retrieval for Image and Video databases VII*, pages 427-435, San-Jose, CA, Jan 1999.

[5] Everitt, B. (1993). *Cluster Analysis*, 3rd Edition. Edward Arnold, London.

[6] Author E. Backer, E. (1995). *Computer-assisted reasoning in cluster analysis*, Prentice Hall International (UK) Ltd. Hertfordshire, UK.

[7] D. Stan and I. K. Sethi, "Image retrieval using a hierarchy of clusters" Proceedings of Second *International Conference on Advances in Pattern Recognition*, pp. 377-386, 2001.

[8] D. Stan and I. K. Sethi, "Mapping low-level image features to semantic concepts", *Proceedings of SPIE: Storage and Retrieval for Media Databases*, pp. 172-179, 2001.

[9] D.P. Berrar, W. Dubitzky, M. Granzow, eds. "Singular value decomposition and principal component analysis", *A Practical Approach to Microarray Data Analysis*, Norwell, MA, 2003. pp. 91-109.

[10] P Howarth and S Ruger, "Fractional distance measures for content-based image retrieval" *European Conf on Information Retrieval* (ECIR, Santiago de Compostela, Spain, Mar 2005), pages 447.

[11] Y. Rubner and C. Tomasi, "Texture Metrics", *Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics*, pages 4601-4607, October 1998.

[12] Y. Rubner, C. Tomasi, and L. Guibas, "The Earth Mover's Distance as a metric for image retrieval", *Technical Report STAN-CS-TN-98-86*, Computer Science Department, Stanford University, September 1998.

[13] J. R. Smith and S. F. Chang, "Tools and Techniques for Color Image Retrieval," in *Proceedings of the SPIE: Storage and Retrieval for Image and Video Databases IV*, 2670, pp. 381-392, 1996.

[14] I. K. Sethi, I. Coman, B. Day et al., "Color-WISE: A system for image similarity retrieval using color," *Proceedings of the SPIE: Storage and Retrieval for Image and Video Databases*, 3132; pp. 140-149, 1998.