Adaptive Binning and Dissimilarity Measure for Image Retrieval and Classification

Wee Kheng Leow and Rui Li
School of Computing, National University of Singapore
3 Science Drive 2, Singapore 117543, Singapore
leowwk, lir@comp.nus.edu.sg

Abstract

Color histogram is an important part of content-based image retrieval systems. It is a common understanding that histograms that adapt to images can represent their color distributions more efficiently than histograms with fixed binnings. However, among existing dissimilarity measures, only the Earth Mover's Distance can compare histograms with different binnings. This paper presents a detailed quantitative study of fixed and adaptive binnings and the corresponding dissimilarity measures. An efficient dissimilarity measure is proposed for comparing histograms with different binnings. Extensive test results show that adaptive binning and dissimilarity produce the best overall performance, in terms of good accuracy, small number of bins, no empty bin, and efficient computation, compared to existing fixed binning schemes and dissimilarity measures.

1. Introduction

Histograms are often used to estimate the distributions of color over an image. There are two methods of generating histograms: *fixed binning* and *adaptive binning*. Typically, a fixed binning method induces histogram bins by partitioning the color space into rectangular bins. Once the bins are derived, they are fixed and the same binning is applied to all images. On the other hand, adaptive binning adapts the bins to the actual distributions of the images. As a result, different binnings are induced for different images.

It is a common understanding that adaptively binned histograms can represent the distributions more efficiently than histograms with fixed binning [7]. But, the comparative advantages have not been quantitatively studied. Moreover, among existing dissimilarity measures, only the Earth Mover's Distance can be used to compare histograms with different binning schemes [7].

This paper presents a detailed quantitative study of fixed and adaptive binnings and the corresponding dissimilarity measures. An efficient dissimilarity measure is proposed for comparing histograms with different binnings. Extensive test results show that the combination of adaptive binning and dissimilarity measure yields the best overall performance compared to existing fixed binning and dissimilarity measures in image classification and retrieval tasks.

2. Related Work

There are two types of *fixed binning* schemes: *regular partitioning* and *clustering*. The first method simply partitions the axes of a target color space into regular intervals, thus producing rectangular bins [3, 9, 10]. The second method partitions a color space into a large number of rectangular cells, which are then clustered by a clustering algorithm such as the k-means [2, 5, 13].

Adaptive binning is similar to color space clustering in that k-means clustering or its variants are used to induce the bins [8]. However, the clustering algorithm is applied to the colors in an image instead of the colors in an entire color space. Therefore, adaptive binning produces different binning schemes for different images.

Note that different binning schemes require different color quantization methods. For regular partitioning, a color is quantized to the centroid of the rectangular bin containing the color, producing a rectangular tessellation of the color space. On the other hand, for color space clustering and adaptive clustering, a color is quantized to the centroid of its nearest cluster, thus producing a Voronoi tessellation of the color space. We shall call the histograms produced by the three methods *regular*, *clustered*, and *adaptive histograms*.

Among commonly used dissimilarity measures, Earth Mover's Distance (EMD) is the only one that can compare histograms with different binnings [7]. Puzicha et al. performed a systematic evaluation of the performance of various dissimilarity measures in classification, segmentation, and retrieval tasks [7]. They concluded that dissimilarities such as χ^2, Kullback-Leibler divergence, and Jessen difference divergence[1] performed better than other measures for larger images, while EMD, Kolmogorov-Smirnov, and Cramer/von Mises performed better for smaller images. The study of Puzicha et al. focused mainly on measuring the performance of dissimilarity measures.

This paper complements the Puzicha et al. study in the following ways: (1) It provides a quantitative evaluation of the performance of the three types of binning schemes (Section 3). (2) It proposes a dissimilarity measure that can compare histograms with different binnings (Section 4). Since the dissimilarity measure does not require an optimization procedure, it can be computed more efficiently than EMD. (3) This paper proposes different methods for benchmarking the combined performance of binning and dissimilarity measure in classification and retrieval tasks (Section 5). These benchmarking tests more closely resemble the retrieval of complex images with one or more regions of interests than those in [7].

3. Adaptive Binning

Adaptive binning can be achieved by the k-means clustering algorithm or its variants. This section describes an adaptive variant that can automatically determine the appropriate number of clusters required. The algorithm can be summarized as follows:

Adaptive Clustering

Repeat
  For each pixel p,
    Find the nearest cluster k to pixel p.
    If no cluster is found or distance $d_{kp} \geq S$,
      create a new cluster with pixel p;
    Else, if $d_{kp} \leq R$,
      add pixel p to cluster k.
  For each cluster i,
    If cluster i has at least N_m pixels,
      update centroid $\mathbf{c}_i$ of cluster i.
    Else, remove cluster i.

The distance d_{kp} between the centroid $\mathbf{c}_k$ of cluster k and pixel p with color $\mathbf{c}_p$ is defined as the CIE94 color-difference equation:

$$d_{kp} = \left[\left(\frac{\Delta L^*}{k_L S_L} \right)^2 + \left(\frac{\Delta C^*_{ab}}{k_C S_C} \right)^2 + \left(\frac{\Delta H^*_{ab}}{k_H S_H} \right)^2 \right]^{1/2} \tag{1}$$

where ΔL^*, ΔC^*_{ab}, and ΔH^*_{ab} are the differences in lightness, chroma, and hue between $\mathbf{c}_k$ and $\mathbf{c}_p$, $S_L = 1$, $S_C = 1 + 0.045\,\bar{C}^*_{ab}$, $S_H = 1 + 0.015\,\bar{C}^*_{ab}$, and $k_L = k_C = k_H = 1$ for reference conditions. The variable $\bar{C}^*_{ab}$ is the geometric mean between the chroma values of $\mathbf{c}_k$ and $\mathbf{c}_p$, i.e., $\bar{C}^*_{ab} = \sqrt{C^*_{ab,k} C^*_{ab,p}}$. The CIE94 color-difference equation is used instead of the Euclidean distance in CIELAB because recent psychological studies show that CIE94 is more perceptually uniform than does Euclidean distance [6, 11].

The adaptive clustering algorithm groups a pixel p into its nearest cluster if it is near enough ($d_{kp} \leq R$). On the other hand, if the pixel p is far enough ($d_{kp} \geq S$) from its nearest cluster, then a new cluster is created. Otherwise, it is left unclustered and will be considered again in the next iteration. This clustering algorithm, thus, ensures that each cluster has a *maximum radius* of R and that the clusters are separated by the distance of approximately S called the *nominal cluster separation*. The value of S is defined as a multiple γ of R, i.e., $S = \gamma R$. Reasonable values of γ range from 0 (for complete overlapping of the clusters) to 2 (for non-overlapping of clusters). Since the algorithm creates a cluster only when a color is far enough from all existing clusters, it can determine the number of clusters required to effectively represent the colors in an image. It also ensures that each cluster has a significant number of (at least N_m) pixels; otherwise, the cluster is removed. In the current implementation, N_m is fixed at 10.

This adaptive clustering algorithm is similar to that of Gong et al. [4]. Both algorithms ensure that the clusters are not too large in volume and not too close to each other. However, our adaptive algorithm is simpler than that in [4]. Moreover, it does not require seed initialization, and can automatically determine the appropriate number of clusters.

4. Weighted Correlation

The Earth Mover's Distance (EMD) is currently the only dissimilarity measure for histograms with different binnings. It performs linear optimization which is computationally expensive. This section defines a dissimilarity measure that can be computed without optimization.

Let $\{h_i\}$, $i = 1, \ldots, n$, denote the bin counts of histogram $\mathcal{H}$ with centroids $\{\mathbf{c}_i\}$. Define histogram $\mathcal{H}'$ in a similar manner. Since $\mathcal{H}$ and $\mathcal{H}'$ have different binnings, their bin centroids are not identical.

To understand how to measure the dissimilarity between $\mathcal{H}$ and $\mathcal{H}'$, let us first consider how to measure the similarity between two bins h_i and h'_j with different bin centroids $\mathbf{c}_i$ and $\mathbf{c}'_j$. Let $h(\mathbf{x})$ and $h'(\mathbf{x})$ denote the actual distributions of colors, where $\mathbf{x}$ denote the 3D color coordinates. Then, the similarity s_{ij} between the two distributions can be defined as the correlation between them:

$$s_{ij} = \int h(\mathbf{x}) h'(\mathbf{x}) d\mathbf{x} \,. \tag{2}$$

[1] The formula that Puzicha et al. [7] called "Jeffreys divergence" is more commonly known as "Jessen difference divergence" in Information Theory literature [1, 12].

Equation 2 is integrated over the 3D space, and is very tedious and time-consuming to compute even for normal distributions. To simplify the computation, let us assume that the distributions are uniform within the bins and 0 outside. Then, Eq. 2 has to be integrated over the intersecting volume only, yielding:

$$s_{ij} = \int \frac{h_i}{V} \frac{h'_j}{V'} d\mathbf{x} = \frac{V_s}{VV'} h_i h'_j \tag{3}$$

where V and V' are the volumes of the bins and V_s is the volume of intersection. Therefore, the similarity between two bins can be defined as the weighted product of the bin counts h_i and h'_j, with the weight w'_{ij} defined in terms of the volume of intersection V_s.

In a perceptually uniform color space such as CIELAB, color similarity is roughly isotropic. That is, the histogram bins are spherical. From solid geometry, the volume of intersection V_s between two equal-sized spherical bins of radius R, separated by a distance d between their centroids, can be derived as:

$$V_s = V - \pi R^2 d + \frac{\pi}{12} d^3 \tag{4}$$

where $V = 4\pi R^3/3$ is the volume of a sphere. For mathematical convenience, the cluster separation d can be specified as a multiple of R, i.e., $d = \alpha R$, $0 \leq \alpha \leq 2$, and the weight w'_{ij} can be defined as

$$w'_{ij} = \frac{V_s}{V} = \begin{cases} 1 - \frac{3}{4}\alpha + \frac{1}{16}\alpha^3 & \text{if } 0 \leq \alpha \leq 2 \\ 0 & \text{otherwise.} \end{cases} \tag{5}$$

Note that w'_{ij} is bounded between 0 and 1, and is symmetric: $w'_{ij} = w'_{ji}$. It can be shown empirically that Eq. 5 closely approximates a Gaussian function of the form $\exp(-d^2/\sigma^2)$ with an appropriate σ.

The dissimilarity $d(\mathcal{H}, \mathcal{H}')$ between histograms $\mathcal{H}$ and $\mathcal{H}'$ can be defined as the following *weighted correlation*:

$$d(\mathcal{H}, \mathcal{H}') = 1 - \sum_{i=1}^{n} \sum_{j=1}^{n'} w'_{ij} h_i h'_j \,. \tag{6}$$

with the following normalization:

$$\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} h_i h_j = \sum_{i=1}^{n'} \sum_{j=1}^{n'} w''_{ij} h'_i h'_j = 1 \tag{7}$$

where w_{ij} and w''_{ij} are the corresponding weights between the bin centroids.

5. Quantitative Evaluation

Quantitative evaluation of the adaptive clustering and weighted correlation was performed with three tests. The first test evaluated the accuracy of adaptive clustering in retaining color information. The second and third tests evaluated the combined performance of adaptive clustering and weighted correlation in image retrieval and classification.

5.1. Color Retention

In this test, the performance of the adaptive clustering was compared with those of regular partitioning and color space clustering. The colors of the images were assumed to be represented in the sRGB space and the target color space was CIELAB.

Test Setup

The adaptive clustering was tested with cluster radius R ranging from 7.5 to 22.5 and nominal cluster separation factor γ ranging from 1.1 to 1.5. For regular partitioning, the L^*-axis of the CIELAB space was partitioned into l equal intervals (l = 8, 10, 12, 14, 16), and the a^*- and b^*-axes were partitioned into m equal intervals (m = 5, 8, 10 and $m \leq l$). The centroids of the bins were mapped back to the sRGB space and bins with illegal sRGB values were discarded. For color space clustering, the CIELAB space was partitioned into $32\times32\times32$ equal partitions and the bin centroids were clustered using the same adaptive clustering algorithm, with $7.5 \leq R \leq 20$ and $1.1 \leq \gamma \leq 1.5$.

As the test images, 100 visually colorful images were randomly selected from the Corel 50,000 photo collection. The images had sizes of either 256×384 or 384×256. Color histograms were generated for each image using the three binning methods.

The performance of the three binning methods were measured by three indicators, namely, the number of bins or clusters produced, the number of empty bins, and the mean color error measured as the mean difference between the actual colors and the quantized colors (in CIE94 units). These performance indicators were averaged over all the images.

Color Error

Experimental results show that the larger the bin volume (or cluster radius R) and the larger the bin separation γ, the smaller is the number of bins and the larger is the mean color error. Figure 1 shows that regular partitioning produced slightly larger mean color error compared to color space clustering, while adaptive clustering produced the smallest error. Given a fixed number of bins, regular and clustered histograms have errors that are about twice those of adaptive histograms.

Empty Bins

Figure 2 shows the average percentage of empty bins in the regular and clustered histograms. With a large number of bins, both histograms have 50% or more empty bins. With

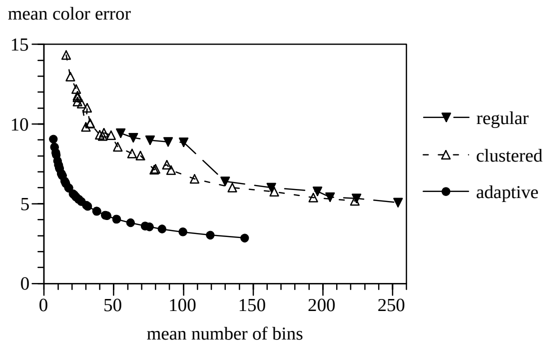

Figure 1. Comparison of mean color errors of regular, clustered, and adaptive histograms.

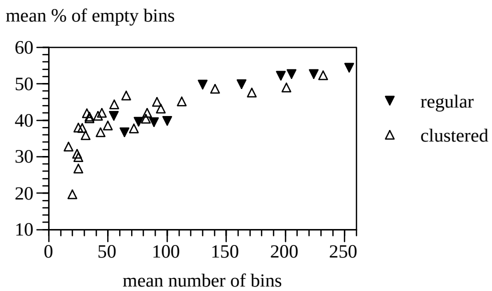

Figure 2. Average percentage of empty bins in regular and clustered histograms.

a small number of bins, clustered histograms have as few as 20% empty bins. The adaptive histograms have no empty bins. These test results show that adaptive histograms can retain color information more accurately with fewer bins than do regular and clustered histograms.

Discussion

Existing systems typically use 64-bin clustered histograms or more than 150 bins for regular histograms. Their respective mean color errors are about 8 and 6, with 45% and 50% empty bins. In comparison, 64-bin adaptive histograms can achieve a color error of about 3.5, lower than human acceptability threshold [11], with no empty bins.

In the subsequent tests, the parameter values of clustered and adaptive binning methods were fixed at $R = 10$ and $\gamma = 1.5$ because this combination yielded good color retention with small number of bins. With these parameter values, the adaptive binning method produced an average of 37.8 bins with a mean color error of 4.53, and the color space clustering method produced 80 bins, a mean color error of 7.19, and 42% empty bins. In principle, the mean color error of color space clustering can be reduced to, say, below 5 so that it is comparable to that of adaptive binning. However, this will require the clustered histograms to have much more than 250 clusters—a value that is both impractical and beyond our experimental range. It was not necessary to test regular partitioning further because its performance was similar to that of color space clustering.

5.2. Image Retrieval

This test assessed the combined performance of binning schemes and dissimilarity measures in image retrieval.

Test Setup

In the image retrieval test of Puzicha et al. [7], random samples of pixels were extracted from the test images. Samples that were drawn from the same image should have similar distributions and were regarded as belonging to the same class. This kind of test samples is useful for testing the performance of various similarity measures in computing global similarity between two images.

A different kind of test samples was prepared for our tests. Each of the 100 images used in the color retention test (Section 5.1) was regarded as forming one query class. These images were scaled down and each embedded into 20 different host images, giving a total of 2000 composite images at each scaling factor. The scaled images were used as query images, and the composite images that contained the same embedded images were regarded as relevant. This test paradigm should be useful for testing the combined performance of binning schemes and dissimilarity measures in retrieving images that contain a particular target region or color distribution of interest. We feel that this test more closely resembles the retrieval of complex images containing one or more regions of interests compared to that in [7]. In the test, scaling factors for image width/height of 1/4, 1/2, and 3/4 were used. These values gave rise to embedded images with area scaling factors of 1/16, 1/4, and 9/16 compared to the original images.

The test was performed with L_2 (i.e., Euclidean) distance, Jessen difference divergence (JD), EMD, and the weighted correlation (WC) measure described in Section 4. L_2 served as the base case. JD and EMD were reported in [7] to yield good performance, respectively, for large and small sample sizes. Other dissimilarity measures evaluated in [7] were expected to yield similar results and were therefore omitted. Both L_2 and JD could be tested only with clustered histograms. The program for EMD was downloaded from Rubner's web site (http://robotics.stanford.edu/ rubner), and was tested only with adaptive histograms due to its longer execution time. The CIE94 distance was used as EMD's ground distance because it is more perceptual uniform than Euclidean distance in the CIELAB space. WC was tested with both clustered and adaptive histograms.

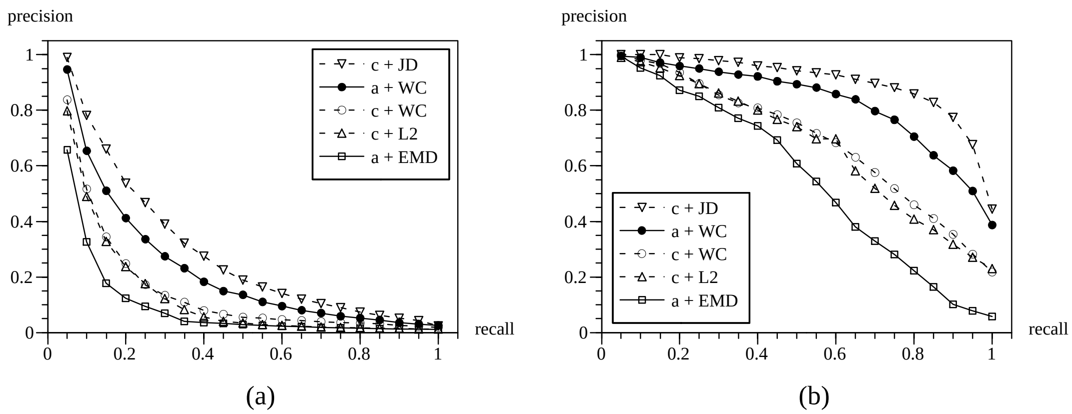

Figure 3. Precision-recall curves of various combinations of binning methods (c: clustered, dashed line; a: adaptive, solid line) and dissimilarities (JD: Jessen difference divergence, WC: weighted correlation, L2: Euclidean, EMD: Earth Mover's Distance). (a) Scaling = 1/2, (b) scaling = 3/4.

Results and Discussion

Figure 3 plots the precision-recall curves of the image retrieval results for width/height scaling factors of 1/2 and 3/4. The curves for scaling factor of 1/4 are not shown because all combinations of binnings and dissimilarity measures performed poorly. They all had very low precision of less than 0.2 at recall rate of 0.1, and their precision dropped to about 0.01 at recall rate of 0.3 and above.

All five combinations performed significantly better for the larger scaling factor of 3/4 than for 1/2. For both scaling factors, clustered histograms together with JD (c + JD) performed best, with the adaptive histograms and WC (a + WC) combination following closely behind. The a + WC combination performed significantly better than c + WC, which had roughly the same performance as c + L_2. These results show that, given the same dissimilarity measure, adaptive histograms perform better than clustered histograms because they can describe color information more accurately and yet use fewer bins (Section 5.1).

Somewhat surprisingly, EMD (with adaptive histograms) performed poorer than L_2. Compared to the results in [7], which show that EMD performed better for small sample sizes, it is noted that our smallest scaling factor of 1/4 corresponds to an image size of 6144 pixels, which is far larger than the sample sizes used in [7]. Moreover, the adaptive histograms have an average of 37.8 bins, and they correspond to medium sized histograms in [7]. These parameter values may have obscured the strengths of EMD in extreme cases of small sample sizes and small number of bins. On the other hand, our choice of the number of bins, which was supported by the color retention test (Section 5.1), and the sample sizes should better resemble the retrieval of complex images with multiple regions.

5.3. Image Classification

This test assessed the combined performance of binning schemes and dissimilarity measures in image classification.

Test Setup

The composite images generated in the retrieval tests (Section 5.2) were used for image classification test. The composite images that contained the same embedded image were considered as belonging to the same class. This would correspond to the practical application in which images containing the same region are considered as identical.

The k-nearest-neighbor classifier with leave-one-out procedure was applied on each of the 2000 composite images. Odd values of $k = 1, 3, 5, 7, 9$ were chosen to remove the possibility of ties. Classification error, averaged over all 2000 images, were computed for each combination of binning scheme, dissimilarity measure, and k value.

Results and Discussion

Figure 4 shows the classification performance for width/height scaling factors of 1/2 and 3/4. The curves for 1/4 scaling are not shown because all combinations of binnings and dissimilarity measures performed poorly.

All five combinations performed significantly better for the larger scaling factor of 3/4 than for 1/2. Moreover, their classification accuracies increased with increasing number of nearest neighbors k. Similar to the image retrieval results, c + JD gave the best performance for both scaling factors, with a + WC following closely behind. The a + WC combination performed better than c + WC, and c + L_2 again had the lowest accuracy. These results again show that, given the same dissimilarity measure, adaptive histograms perform better than clustered histograms. Unlike

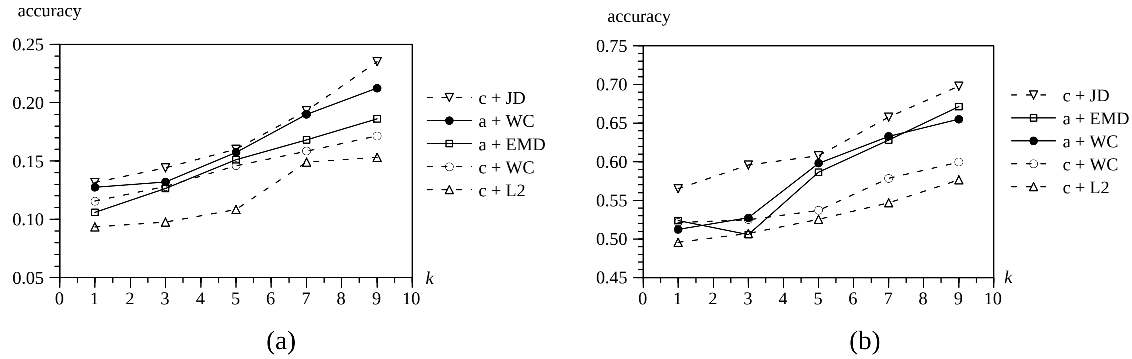

Figure 4. Classification accuracy of various combinations of binning methods and dissimilarities.

in the retrieval tests, the performance of a + EMD was very good in the classification tests. The classification accuracy of a + EMD closely matched that of a +WC, especially for the larger scaling factor of 3/4.

6. Conclusions

This paper presented an adaptive clustering method and a dissimilarity measure for comparing histograms with different binnings. The adaptive clustering algorithm is an adaptive variant of the k-means clustering algorithm and it can determine the number of clusters required to effectively describe the colors in an image. The dissimilarity measure computes a weighted correlation between two histograms, and the weights are defined in terms of the volumes of intersection between overlapping spherical clusters. Since this measure does not require optimization, it executes more efficiently than Earth Mover's Distance (EMD) does.

Extensive tests were performed to evaluate the performance of adaptive clustering and weighted correlation (WC) on color retention, image retrieval, and image classification tasks. Compared to fixed binning schemes, adaptive clustering can retain color information more accurately with fewer bins and no empty bin. The combined performance of adaptive clustering and WC is comparable to that of Jessen difference divergence and better than those of L_2 and EMD. Given the same dissimilarity measure, adaptive clustering performs better than fixed binning. Therefore, the combination of adaptive clustering and weighted correlation achieve the best overall performance of good accuracy, small number of bins, no empty bin, and efficient computation.

Acknowledgments

This research is supported by NUS ARF R-252-000-072-112 and NSTB UPG/98/015.

References

[1] J. Burbea and C. R. Rao. Entropy differential metric, distance and divergence measures in probability spaces: A unified approach. *J. Multivariate Analysis*, 12:575–596, 1982.

[2] G. Ciocca and R. Schittini. A relevance feedback mechanism for content-based image retrieval. *Infor. Proc. and Management*, 35:605–632, 1999.

[3] I. J. Cox, M. L. Miller, S. O. Omohundro, and P. N. Yianilos. PicHunter: Bayesian relevance feedback for image retrieval. In *Proc. ICPR '96*, pages 361–369, 1996.

[4] Y. Gong, G. Proietti, and C. Faloutsos. Image indexing and retrieval based on human perceptual color clustering. In *Proc. CVPR '98*, 1998.

[5] J. Hafner, H. S. Sawhney, W. Esquitz, M. Flickner, and W. Niblack. Efficient color histogram indexing for quadratic form distance functions. *IEEE Trans. PAMI*, 17:729–736, 1995.

[6] M. Melgosa. Testing CIELAB-based color-difference formulas. *Color Research and Application*, 25(1):49–55, 2000.

[7] J. Puzicha, J. M. Buhmann, Y. Rubner, and C. Tomasi. Empirical evaluation of dissimilarity for color and texture. In *Proc. ICCV '99*, pages 1165–1172, 1999.

[8] Y. Rubner, C. Tomasi, and L. J. Guibas. A metric for distributions with applications to image databases. In *Proc. ICCV '98*, 1998.

[9] S. Sclaroff, L. Taycher, and M. L. Cascia. Image-Rover: A content-based image browser for the world wide web. In *Proc. IEEE Workshop on Content-Based Access of Image and Video Libraries*, 1997.

[10] J. R. Smith and S.-F. Chang. Single color extraction and image query. In *Proc. ICIP '95*, 1995.

[11] T. Song and R. Luo. Testing color-difference formulae on complex images using a CRT monitor. In *Proc. of 8th Color Imaging Conference*, 2000.

[12] I. J. Taneja. New developments on generalized information measures. In P. W. Hawkes, editor, *Advances in Imaging and Electron Physics*, volume 91. Academic Press, 1995.

[13] A. Vailaya, A. Jain, and H. J. Zhang. On image classification: City images vs. landscapes. *Pattern Recognition*, 31:1921–1935, 1998.