The Analysis and Applications of Adaptive-Binning Color Histograms

Wee Kheng Leow and Rui Li

Dept. of Computer Science, National University of Singapore,

3 Science Drive 2, Singapore 117543

leowwk@comp.nus.edu.sg, lir@comp.nus.edu.sg

Corresponding Author:

Wee Kheng Leow

Dept. of Computer Science, National University of Singapore,

3 Science Drive 2, Singapore 117543

leowwk@comp.nus.edu.sg

phone: +65-6874-6540

fax: +65-6779-4580

Abstract

Histograms are commonly used in content-based image retrieval systems to represent the distributions of colors in images. It is a common understanding that histograms that adapt to images can represent their color distributions more efficiently than do histograms with fixed binnings. However, existing systems almost exclusively adopt fixed-binning histograms because, among existing well-known dissimilarity measures, only the computationally expensive Earth Mover's Distance (EMD) can compare histograms with different binnings. This article addresses the issue by defining a new dissimilarity measure that is more reliable than the Euclidean distance and yet computationally less expensive than EMD. Moreover, a mathematically sound definition of mean histogram can be defined for histogram clustering applications. Extensive test results show that adaptive histograms produce the best overall performance, in terms of good accuracy, small number of bins, no empty bin, and efficient computation, compared to existing methods for histogram retrieval, classification, and clustering tasks.

Keywords: Color histograms, adaptive binning, histogram-based dissimilarity measures, image retrieval, image classification, image clustering.

1 Introduction

In content-based image retrieval systems, histograms are often used to represent the distributions of colors in images. There are two general 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 [8, 9, 21, 25, 32, 35, 38]. Once the bins are derived, they are fixed and the same binning scheme is applied to all images. On the other hand, adaptive binning adapts to the actual distributions of colors in images [3, 11, 22, 27, 31]. As a result, different binnings are induced for different images.

It is a common understanding that adaptively-binned histograms can represent the distributions of colors in images more efficiently than do histograms with fixed binning [11, 27, 31]. However, existing systems almost exclusively adopt fixed-binning histograms because among existing well-known dissimilarity measures, only the Earth Mover's Distance (EMD) can compare histograms with different binnings [27, 31]. But, EMD is computationally more expensive than other dissimilarity measures because it requires an optimization process.

Another major concern is that fixed-binning histograms have been regarded as vectors in a linear vector space, with each bin representing a dimension of the space. This convenient vector interpretation makes it possible to apply various well-known algorithms, such as clustering, Principle Component Analysis, and Singular Value Decomposition to process and analyze histograms [10, 26, 34]. Unfortunately, this approach is not satisfactory because the algorithms are applied in a linear vector space, which assumes the Euclidean distance as the measure of vector difference. And Euclidean distance has been found to be less reliable than other measures for computing histogram dissimilarity [5, 27, 33]. As a result, the

effectiveness and reliability of the approach is compromised.

Adaptive histograms cannot be conveniently mapped into a linear vector space because different histograms may have different bins. Although Multidimensional Scaling (MDS) [4] can be used to recover the Euclidean coordinates of the histograms from pairwise distances between them, it is computationally expensive to apply MDS on a large number of (say, more than 100) histograms. Moreover, MDS incurs an error in recovering the coordinates, further compromising the effectiveness of adaptive histograms in practical applications.

To address the above issues, this article proposes a new dissimilarity measure for adaptive color histograms (Section 5) that is more reliable than the Euclidean distance and yet computationally less expensive than the Earth Mover's distance. Moreover, a mathematically sound definition of mean histogram can be defined for histogram clustering applications. Extensive test results (Section 6) show that the use of adaptive histograms produces the best overall performance, in terms of good accuracy, small number of bins, no empty bin, and efficient computation, compared to existing methods in histogram retrieval, classification, and clustering tasks.

2 Related Work

There are two types of *fixed binning* schemes: *regular partitioning* and *color space clustering*. The first method simply partitions the axes of a target color space into regular intervals, thus producing rectangular bins. Typically, one of the three color axes is regarded as conveying more important information and is partitioned into more intervals than are the other two axes. For example, VisualSeek [35] partitions the HSV space into $18 \times 3 \times 3$ color bins and

4 grey bins, producing 166 bins. PicHunter [9] also partitions the HSV space in a similar manner. The CIELUV space has also been used [21, 32] because it is more perceptually uniform than RGB space [2]. In partitioning these spaces, bins that correspond to illegal RGB colors are usually discarded.

The second method partitions a color space into a large number of cells, which are then clustered by a clustering algorithm such as the k-means. For example, QBIC [13] partitions the RGB space into $16 \times 16 \times 16$ cells, maps the cells to a modified Munsell HVC space, and then clustered the cells into k clusters. Vailaya et al. [38] apply a similar method but map the RGB cells into the HSV space, where 64 bins are produced. Quicklook [8] maps sRGB [1] cells into CIELAB and clusters them into 64 bins.

Adaptive binning is similar to color space clustering in that k-means clustering or its variant is used to induce the bins. However, the clustering algorithm is applied to the colors in an image instead of the colors in an entire color space [3, 11, 22, 27, 31]. Therefore, adaptive binning produces different bins for different images.

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, with 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 [27, 31]. Puzicha et al. performed a systematic evaluation of the performance of various dissimilarity measures in classifica-

tion, segmentation, and retrieval tasks [27]. They concluded that dissimilarities such as χ^2, Kullback-Leibler divergence, and Jeffreys divergence performed better than other measures for large sample size (i.e., number of pixels sampled in an image), while EMD, Kolmogorov-Smirnov, and Cramer/von Mises performed better for small sample size. In an earlier paper, they presented another similar study for the tasks of texture segmentation and retrieval [28]. Manjunath and Ma also performed a benchmark study for texture retrieval [19].

The study of Sebe et al. [33] shows that the Euclidean distance is justified from a maximum likelihood perspective when the additive noise distribution is Gaussian. However, their experiments on image retrieval, stereo matching, and motion tracking suggest that real noise distributions are better modeled by the Cauchy distribution than Gaussian and Exponential. Consequently, the Cauchy metric achieves greater accuracy than the Euclidean distance, sum of absolute difference, and Kullback relative information.

Brunelli and Mich [5] proposed the concept of *histogram capacity* to quantify the effectiveness of histograms as image descriptors and histogram dissimilarities for image retrieval. Their analysis results permit the design of scalable image retrieval systems that make optimal use of computational and storage resources.

This article complements the above studies in the following ways: (1) It provides a quantitative evaluation of the performance of the three types of binning schemes (Section 3). (2) It defines a new dissimilarity measure that can compare histograms with different binnings (Section 5). Since the dissimilarity measure does not require an optimization procedure, it can be computed more efficiently than EMD. (3) It proposes different methods for benchmarking the combined performance of binning and dissimilarity measure in image retrieval, classification, and clustering tasks (Sections 6.3, 6.4, 6.5). These benchmarking tests more

closely resemble the retrieval of complex images with one or more regions of interests than those in [27].

3 Adaptive Binning

Adaptive binning of the colors in an image can be achieved by an appropriate vector quantization algorithms such as k-means clustering or its variants [24]. This section describes an adaptive variant of k-means 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 a pixel p with color $\mathbf{c}_p$ is defined as the CIE94 color-difference equation [2]:

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

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 or CIELUV because recent psychological studies show that CIE94 is more perceptually uniform than does Euclidean distance [2, 12, 14, 23, 36].

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 adequately 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. [11] 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 [11]. Moreover, it does not require seed initialization, and can automatically determine the appropriate number of clusters.

In practice, for efficiency sake, the algorithm is repeated for only 10 iterations. When the algorithm terminates, some colors may still be unclustered. During color quantization or histogram generation, these unclustered colors are quantized to the colors of their nearest clusters. Empirical tests show that having a small amount of unclustered colors during the clustering process does not produce significant error in the color quantization results. For instance, our test results show that 5% unclustered colors contribute to only a 1% increase in the mean error of color quantization compared to the case in which all the colors are clustered. In fact, leaving some colors unclustered makes the algorithm more robust against noise colors that differ significantly from other main colors in the image. These noise colors typically occur at abrupt color discontinuities in the images.

4 Overview of Histogram Similarity

Before discussing the mathematics of adaptive histograms, let us motivate the mathematical formulation by first describing a possible definition of similarity measure for adaptive color histograms. To begin, let us first consider two adaptive histograms H and H', each having only one bin located at $\mathbf{c}$ and $\mathbf{c}'$, with bin counts h and h', respectively. Let $f(\mathbf{x})$ and $f'(\mathbf{x})$ denote the actual density distributions of colors in and around the two bins, where $\mathbf{x}$ denote 3D color coordinates. Then, the similarity $\zeta(H, H')$ between the two distributions can be

defined, as is commonly practiced, as the correlation between them:

$$\zeta(H, H') = \int f(\mathbf{x})\, f'(\mathbf{x})\, d\mathbf{x}\ . \tag{2}$$

Equation 2 is integrated over the 3D space. It is very tedious and time-consuming to compute the integration even if normal distributions are assumed for $f(\mathbf{x})$ and $f'(\mathbf{x})$. 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 $\mathcal{V}$ only, yielding:

$$\zeta(H, H') = \int_{\mathcal{V}} \frac{h}{V}\frac{h'}{V'}\, d\mathbf{x} = \frac{V_s}{VV'} hh' \tag{3}$$

where V and V' are the volumes of the bins and V_s is the volume of intersection. Therefore, the similarity between two distributions can be defined as the weighted product of the bin counts h and h', with the weight $w(\mathbf{c}, \mathbf{c}')$ defined in terms of the volume of intersection V_s. The weight $w(\mathbf{c}, \mathbf{c}')$ can be interpreted as the similarity between the two bins.

In an appropriate color space that is perceptually uniform, such as CIELAB, spherical bins of the same radius can be adopted for ease of computation of bin similarity. The adoption of spherical bins is supported by the use of appropriate color-difference equations such as CIE94, CMC, and BDF, all defined in the CIELAB color space [2]. Recent psychological tests have confirmed that these color-difference equations are more perceptually uniform than does Euclidean distance in the CIELAB and CIELUV spaces [2, 12, 14, 23, 36].

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. The bin separation d can be specified as a

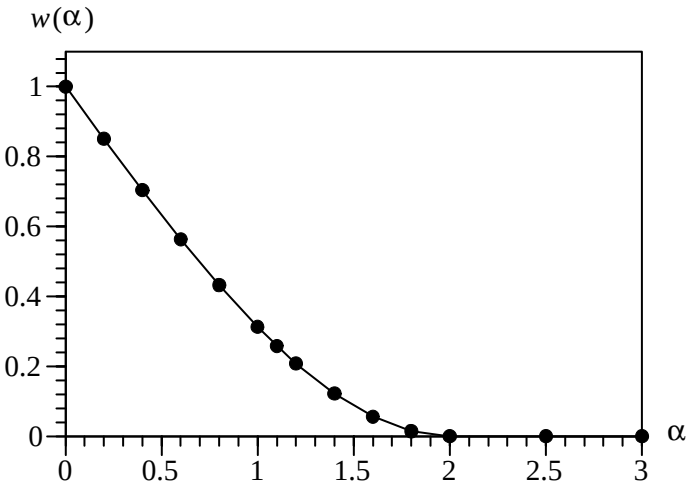


Figure 1: A plot of bin similarity $w(\alpha)$ against bin separation ratio α.

multiple of R, i.e., $d = \alpha R$, and the weight $w(\mathbf{c}, \mathbf{c}')$ can be defined as

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

This definition of the weight is simpler than that derived in Eq. 3 because it is dependent only on the distance between the bin centroids and is independent of the bin volumes. Figure 1 shows that the function $w(\alpha)$ decreases at a faster-than-linear rate with increasing α.

For histograms with more than one bin, the similarity $\zeta(H, H')$ can thus be defined as follows:

$$\zeta(H, H') = \sum_{i=1}^{n} \sum_{j=1}^{n'} w(\mathbf{c}_i, \mathbf{c}'_j) h_i h'_j \ . \tag{6}$$

In practice, it is useful to bound the value of similarity measure so that its inverse, the dissimilarity measure, is also bounded. To achieve this goal, it is necessary to normalize the bin counts h_i and h'_j. So, a proper definition of histogram normalization is needed. In the next section, we shall provide a rigorous mathematical treatment of histogram normalization

and similarity.

5 Adaptive Color Histograms

An adaptive color histogram $H = (n, \mathcal{C}, \mathcal{H})$ is defined as a 3-tuple consisting of a set $\mathcal{C}$ of n bins with bin centroids $\mathbf{c}_i$, $i = 1, \ldots, n$, and a set $\mathcal{H}$ of corresponding bin counts $h_i > 0$. Given two adaptive histograms $G = (m, \{\mathbf{b}_i\}, \{g_i\})$ and $H = (n, \{\mathbf{c}_i\}, \{h_i\})$, define the *weighted correlation* $G \cdot H$ as in Eq. 6:

$$G \cdot H = \sum_{i=1}^{m} \sum_{j=1}^{n} w(\mathbf{b}_i, \mathbf{c}_j)\, g_i\, h_j \ . \tag{7}$$

A histogram H can be normalized into $\overline{H}$ by dividing each bin count by the histogram norm $\|H\| = \sqrt{H \cdot H}$. The similarity $s(G, H)$ between G and H is then defined as the weighted correlation between their normalized forms: $s(G, H) = \overline{G} \cdot \overline{H}$. We will prove in Theorem 2 that $s(G, H)$ is bounded between 0 and 1 (under a mild condition). So, the dissimilarity $d(G, H)$ can be defined as $d(G, H) = 1 - s(G, H)$.

Mean histogram is very useful in applications such as clustering. It is defined in terms of an operation called the *merging* of histograms, which is, in turn, defined in terms of the addition and union of histograms.

Definition 1 (Addition) *The addition of histograms $G = (n, \mathcal{C}, \{g_i\})$ and $H = (n, \mathcal{C}, \{h_i\})$ with identical set of bins $\mathcal{C}$ is $G + H = (n, \mathcal{C}, \{g_i + h_i\})$.*

Histogram addition is defined only for histograms with an identical set of bins.

Definition 2 (Union) *The union of histograms $G = (m, \mathcal{B}, \mathcal{G})$ and $H = (n, \mathcal{C}, \mathcal{H})$ with disjoint sets of bins, i.e., $\mathcal{B} \cap \mathcal{C} = \emptyset$, is $G \cup H = (m + n, \mathcal{B} \cup \mathcal{C}, \mathcal{G} \cup \mathcal{H})$.*

Note that the notation $\mathcal{G} \cup \mathcal{H}$ is used to mean the collection of the bin counts of the two histograms, which allows for duplicates, instead of the usual set union. Although it is possible to define the union operator on histograms with common bins, the above definition is more convenient for the following histogram merging operator which combines addition and union into a single operation.

Definition 3 (Merging) *Let histogram $G = X \cup Y$ and $H = X' \cup Z$ such that X and X' have the same set of bins and X, Y, and Z have disjoint sets of bins. Then, the merged histogram $G \uplus H = (X \cup Y) \uplus (X' \cup Z) = (X + X') \cup Y \cup Z$.*

That is, two histograms are merged by collecting all the bins and adding the bin counts of identical bins. Note that it is always possible to express two histograms G and H in the form given in Definition 3 for histogram merging to be well-defined.

In order that histogram dissimilarity is well defined (i.e., $d(G, H) \geq 0$), histogram similarity has to be bounded from above by the value 1. It turns out that this is guaranteed if an equivalent form of *Cauchy-Schwarz inequality* holds for adaptive histograms:

$$(G \cdot H)^2 \leq (G \cdot G)\,(H \cdot H) \tag{8}$$

or equivalently

$$G \cdot H \leq \|G\|\|H\| \ . \tag{9}$$

Unfortunately, the Cauchy-Schwarz inequality does not hold for adaptive histograms in general. Here, we provide a necessary condition and a sufficient condition for Cauchy-Schwarz inequality, supplemented by examples and discussion about practical issues.

Theorem 1 (Cauchy-Schwarz Inequality) *Given histograms $G = (m, \{\mathbf{b}_i\}, \{g_i\})$ and $H = (n, \{\mathbf{c}_i\}, \{h_i\})$, define*

- $\Phi = \dfrac{\sum\limits_{i,j,k,l} w(\mathbf{b}_i, \mathbf{c}_k)\, w(\mathbf{b}_j, \mathbf{c}_l)}{\sum\limits_{i,j,k,l} w(\mathbf{b}_i, \mathbf{b}_j)\, w(\mathbf{c}_k, \mathbf{c}_l)}$
- $\lambda = \min\limits_{i,j,k,l} g_i g_j h_k h_l = \left(\min\limits_i g_i \ \min\limits_k h_k\right)^2$
- $\Lambda = \max\limits_{i,j,k,l} g_i g_j h_k h_l = \left(\max\limits_i g_i \ \max\limits_k h_k\right)^2$.

Then, a necessary condition for Cauchy-Schwarz inequality $(G \cdot H)^2 \leq (G \cdot G)\,(H \cdot H)$ *to hold is* $\Phi \leq \Lambda/\lambda$*, and a sufficient condition is* $\Phi \leq \lambda/\Lambda$.

Proof. (Necessary condition) From the definition of weighted correlation,

$$(G \cdot H)^2 = \left(\sum_{i,k} w(\mathbf{b}_i, \mathbf{c}_k)\, g_i h_k\right)^2 = \sum_{i,j,k,l} w(\mathbf{b}_i, \mathbf{c}_k)\, w(\mathbf{b}_j, \mathbf{c}_l)\, g_i g_j h_k h_l \tag{10}$$

and

$$(G \cdot G)\,(H \cdot H) = \sum_{i,j} w(\mathbf{b}_i, \mathbf{b}_j)\, g_i g_j \sum_{k,l} w(\mathbf{c}_k, \mathbf{c}_l)\, h_k h_l = \sum_{i,j,k,l} w(\mathbf{b}_i, \mathbf{b}_j)\, w(\mathbf{c}_k, \mathbf{c}_l)\, g_i g_j h_k h_l \ . \tag{11}$$

Thus,

$$\sum_{i,j,k,l} w(\mathbf{b}_i, \mathbf{c}_k)\, w(\mathbf{b}_j, \mathbf{c}_l)\, \lambda \leq (G \cdot H)^2$$

and

$$(G \cdot G)(H \cdot H) \leq \sum_{i,j,k,l} w(\mathbf{b}_i, \mathbf{b}_j)\, w(\mathbf{c}_k, \mathbf{c}_l)\, \Lambda \ .$$

If Cauchy-Schwarz inequality holds, then

$$(G \cdot H)^2 \leq (G \cdot G)(H \cdot H)$$

which implies that $\Phi \leq \Lambda/\lambda$.

(Sufficient condition) If the sufficient condition is satisfied, then

$$(G \cdot H)^2 \leq \sum_{i,j,k,l} w(\mathbf{b}_i, \mathbf{c}_k)\, w(\mathbf{b}_j, \mathbf{c}_l)\, \Lambda \leq \sum_{i,j,k,l} w(\mathbf{b}_i, \mathbf{b}_j)\, w(\mathbf{c}_k, \mathbf{c}_l)\, \lambda \leq (G \cdot G)(H \cdot H) \ .$$

Thus, Cauchy-Schwarz inequality holds. □

Note that there is a gap between the necessary condition and the sufficient condition (Fig. 2). Histograms that satisfy Cauchy-Schwarz inequality must satisfy the necessary condition but may or may not satisfy the sufficient condition. The following examples illustrate the gap.

Example 1: Consider histograms $G = (n, \{\mathbf{b}_i\}, \{g_i\})$ and $H = (n, \{\mathbf{c}_i\}, \{h_i\})$ with the same number of n bins such that $w(\mathbf{b}_i, \mathbf{b}_i) = w(\mathbf{c}_i, \mathbf{c}_i) = w(\mathbf{b}_i, \mathbf{c}_i) = 1$ for all i while all other weights $w(\cdot) = 0$. Then, the weighted correlation $G \cdot H$ reduces to the inner product of two vectors, which is well-known to satisfy Cauchy-Schwarz inequality [20]. But, $\Phi = 1 > \lambda/\Lambda$ for most G and H, except for the case $\lambda = \Lambda$ in which the histograms have uniform bin counts.

For a more concrete example, take $G = (2, \{\mathbf{b}_i\}, \{1, 1/2\})$ and $H = (2, \{\mathbf{c}_i\}, \{1, 1/4\})$ with the above weights. Then, $\Phi = 1 > \lambda/\Lambda = 1/64$, i.e., the histograms violate the sufficient condition. But, Cauchy-Schwarz inequality still holds:

$$(G \cdot H)^2 = \frac{81}{64} < (G \cdot G)(H \cdot H) = \frac{85}{64} \ .$$

Example 2: Consider $G = (2, \{\mathbf{b}_i\}, \{1, 1/2\})$ and $H = (2, \{\mathbf{c}_i\}, \{1, 1/4\})$ with weights $w(\mathbf{b}_i, \mathbf{b}_i) = w(\mathbf{c}_i, \mathbf{c}_i) = 1$ for all i, $w(\mathbf{b}_i, \mathbf{b}_j) = w(\mathbf{c}_i, \mathbf{c}_j) = 0$ for $i \neq j$, and $w(\mathbf{b}_i, \mathbf{c}_j) = 1$ for all i, j. Then, $\Phi = 4 \leq \Lambda/\lambda = 64$, i.e., the histograms satisfy the necessary condition. But,

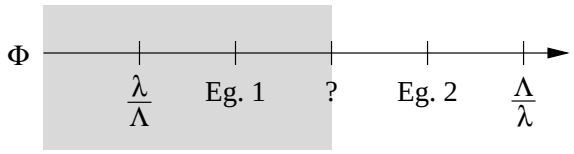


Figure 2: Cauchy-Schwarz inequality is satisfied when Φ falls in the shaded region. $\Phi \leq \Lambda/\lambda$ gives the necessary condition and $\Phi \leq \lambda/\Lambda$ gives the sufficient condition.

Cauchy-Schwarz inequality does not hold:

$$(G \cdot H)^2 = \frac{255}{64} > (G \cdot G)(H \cdot H) = \frac{85}{64} \ .$$

These examples are summarized in Fig. 2.

In practice, the bins of a histogram are sparsely distributed and with minimum overlap between them. A bin of G overlaps at most a small number of, say μ, bins of H significantly, and $\mu \ll \min(m, n)$. Moreover, the weight function is inversely related to the distance between the bins (Eq. 5, Fig. 1). As a result, the numerator of Φ has at most μ^2 significant terms. On the other hand, the denominator of Φ has at least $m \times n$ significant terms because $w(\mathbf{b}_i, \mathbf{b}_i) = w(\mathbf{c}_k, \mathbf{c}_k) = 1$ for all i, k. Therefore, Φ tends to be much smaller than Λ/λ, the necessary condition. This explains the observation that the histogram dissimilarities that we computed in our tests and applications are all non-negative.

The boundedness of histogram similarity and dissimilarity follows directly from Cauchy-Schwarz inequality:

Theorem 2 (Boundedness of Similarity) *For any histograms G and H satisfying Cauchy-Schwarz inequality, $0 \leq s(G, H) \leq 1$ and $0 \leq d(G, H) \leq 1$.*

Proof. Since the bin similarity w and bin counts g_i and h_i are non-zero, $s(G, H) \geq 0$ and

$d(G, H) \leq 1$. From Cauchy-Schwarz inequality,

$$(G \cdot H)^2 \leq (G \cdot G)\,(H \cdot H) = \|G\|^2 \|H\|^2 \ .$$

Therefore,

$$s(G, H) = \frac{G \cdot H}{\|G\|\|H\|} \leq 1 \text{ and } d(G, H) \geq 0 \ . \qquad \square$$

Next, let us provide a mathematically sound definition for a mean histogram.

Definition 4 (Mean Histogram) *M is a mean histogram of H_i, $i = 1, \ldots, N$, if M maximizes the total similarity $S(M)$:*

$$S(M) = \sum_{i=1}^{N} s(M, H_i)\,. \tag{12}$$

This definition is equivalent to saying that the mean histogram minimizes the total distance $\sum_i d(M, H_i)$, which is consistent with the usual definition of mean.

Now, we can show how to compute the mean histogram that is mathematically sound.

Theorem 3 (Mean Histogram) *$\biguplus_i \overline{H}_i$ is a mean of histograms H_i, $i = 1, \ldots, N$.*

Proof. Let M denote $\biguplus_i \overline{H}_i$. Total similarity $S(M)$ between M and H_i is

$$S(M) = \sum_i s(M, H_i) = \sum_i \overline{M} \cdot \overline{H}_i\,.$$

From the definition of weighted correlation (Eq. 7) and histogram merging (Definition 3), it is easy to show that

$$\sum_i \overline{M} \cdot \overline{H}_i = \overline{M} \cdot \overline{\biguplus_i} \overline{H}_i\,.$$

Therefore,

$$S(M) = \overline{M} \cdot M = \|M\| .$$

Now, consider any histogram M' that is arbitrarily close to but different from M, i.e., $s(M', M) < 1$. Total similarity $S(M')$ between M' and H_i is

$$\begin{aligned} S(M') &= \sum_i s(M', H_i) = \sum_i \overline{M'} \cdot \overline{H}_i \\ &= \overline{M'} \cdot \biguplus_i \overline{H}_i = \overline{M'} \cdot M = \|M\| \, \overline{M'} \cdot \overline{M} \\ &< \|M\| = S(M) \ . \end{aligned}$$

Therefore, M maximizes the total similarity $S(M)$. □

Notice that the computation of mean histogram based on other non-Euclidean distances may require an optimization procedure that is computationally expensive in general. In contrast, the computation of mean histogram based on adaptive histogram dissimilarity is as straightforward as that of a Euclidean mean, and yet is applicable to histograms with different binnings.

The usual definition of mean divides the sum by the number of items that are added together:

$$M = \frac{1}{N} \sum_i H_i \ . \tag{13}$$

But, it is applicable only to fixed-binning histograms. For adaptive histograms, this division is not necessary because the "division" is performed within the merging operation, i.e., histograms are normalized before they are merged to produce the mean histogram.

An implication of Theorem 3 is that histogram merging is equivalent to histogram averaging. If histograms that are very different from each others are merged, we expect to obtain a mean histogram that is not similar to any of the histograms that are merged. Such

a mean may not be useful in practice. An analogy is the mixing of the color pigments of red, green, and blue, and yielding grey which is very different from the original colors. On the other hand, merging similar histograms yields a mean that is similar to the histograms that are merged.

The merging of many histograms together may result in a merged histogram with a large number of bins. So, it might be useful to merge similar bins so as to reduce the number of bins. This procedure can be performed by applying the adaptive binning algorithm (Section 3).

6 Performance Evaluation

Four types of tests were conducted to evaluate the performance of adaptive color histograms and weighted correlation dissimilarity measure: color retention, image retrieval, image classification, and image clustering.

6.1 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 [1], and the target color space was CIELAB.

6.1.1 Test Setup

Adaptive clustering algorithm 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 $32 \times 32 \times 32$ 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.

6.1.2 Color Error

Experimental results (Fig. 3a) show 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 about twice the amount of error as do adaptive histograms. As the bin volume (or cluster radius R) and the bin separation γ of adaptive histograms increase, the number of bins decreases but the mean color error increases. Figure 3(a) shows that beyond a sufficiently large number of bins, the decrease in error with increasing number of bins becomes insignificant.

6.1.3 Empty Bins

Figure 3(b) 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 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.

6.1.4 Visual Quality

Figure 4 shows two sample images quantized using adaptive binning and achieved a mean color error of 5 CIE94 units or less. Visual inspection reveals that the color-quantized images look indistinguishable from the original images except at regions where banding occurs such as clear blue sky. This is the result of quantizing the gradually varying colors into discrete bins. This observation matches recent psychological study [36] very well, which shows that that human's color acceptability threshold is 4.5. That is, two colors with a color difference of less than 4.5 are regarded as practically identical. Note that the acceptability threshold is slightly larger than the perceptibility threshold of 2.2 [36], which is the threshold below which two colors are perceptually indistinguishable.

6.2 Discussion

Existing image retrieval systems (Section 2) 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 (Figure 3). In comparison, 64-bin adaptive histograms

can achieve a color error of about 3.5, lower than human acceptability threshold [36], 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 could 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.

6.3 Image Retrieval

This test assessed the combined performance of binning schemes and dissimilarity measures in image retrieval. The weighted correlation dissimilarity (WC) described in Section 5 was compared with three existing dissimilarity measures, namely L_2 (Euclidean), Jessen Difference Divergence (JD)[1], and Earth Mover's Distance (EMD).

[1]The formula that Puzicha et al. [27] called "Jeffreys divergence" is more commonly known as "Jessen difference divergence" in Information Theory literature [6, 7, 37]. Jeffreys divergence, as given in the literature [6, 7, 15, 18, 37], takes the form $\sum_i (g_i - h_i) \log(g_i/h_i) = \sum_i [g_i \log(g_i/h_i) + h_i \log(h_i/g_i)]$.

- L_2 (Euclidean) distance:

$$d(G,H) = \left(\sum_i (g_i - h_i)^2 \right)^{1/2} \tag{14}$$

- Jessen difference divergence (JD):

$$d(G,H) = \sum_i \left(g_i \log \frac{g_i}{m_i} + h_i \log \frac{h_i}{m_i} \right) \tag{15}$$

  where $m_i = (g_i + h_i)/2$.

- Earth Mover's distance (EMD) [30]:

$$d(G,H) = \frac{\sum_{i,j} f_{ij}\, d(\mathbf{b}_i, \mathbf{c}_j)}{\sum_{i,j} f_{ij}} \tag{16}$$

  where $d(\mathbf{b}_i, \mathbf{c}_j)$ denotes the dissimilarity between bins $\mathbf{b}_i$ and $\mathbf{c}_j$, and f_{ij} is the optimal flow between G and H such that the total cost $\sum_{i,j} f_{ij}\, d(\mathbf{b}_i, \mathbf{c}_j)$ is minimized, subject to the constraints:

$$f_{ij} \geq 0 \ , \ \ \sum_i f_{ij} \leq h_j \ , \ \ \sum_j f_{ij} \leq g_i \ , \ \ \sum_i \sum_j f_{ij} = \min(\sum_i g_i, \sum_j h_j) \ . \tag{17}$$

  The dissimilarity $d(\mathbf{b}_i, \mathbf{c}_j)$ between two bins is typically defined as a monotonic increasing function of the ground distance between the bins.

Among these dissimilarity measures, L_2 served as the base case of the performance evaluation. JD and EMD are reported in [27] to yield good performance, respectively, for large and small sample sizes. Other dissimilarity measures evaluated in [27] are expected to yield similar results and are therefore omitted.

WC is tested with both clustered and adaptive histograms whereas 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, and was taken as the dissimilarity between two bins. This arrangement produced a total of five combinations of binning schemes and dissimilarity measures.

6.3.1 Test Setup

In the image retrieval test of Puzicha et al. [27], 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 dissimilarity 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 6.1) was regarded as forming a different 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 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 [27]. 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), Earth Mover's distance (EMD), and the weighted correlation (WC) dissimilarity of adaptive histograms.

6.3.2 Results and Discussion

Figure 5 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 of binning schemes and dissimilarity measures 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 6.1).

Somewhat surprisingly, EMD (with adaptive histograms) performed poorer than L_2. Compared to the results of Puzicha et al. [27], 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 [27]. Moreover, the adaptive histograms have an average of 37.8 bins, and they correspond to medium sized

histograms in [27]. These parameter values may have obscured the strengths of EMD in extreme cases of small sample sizes and small number of bins. However, our choices of the number of bins, which are supported by the color retention test (Section 6.1), and the sample sizes are expected to match practical retrieval of complex images with multiple regions better than those of Puzicha et al. [27].

6.4 Image Classification

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

6.4.1 Test Setup

The composite images generated in the retrieval tests (Section 6.3) 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. Classification error, averaged over all 2000 images, were computed for each combination of binning scheme, dissimilarity measure, and k value.

6.4.2 Results and Discussion

Figure 6 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 had the lowest accuracy. These results again show that, given the same dissimilarity measure, adaptive histograms perform better than clustered histograms. Unlike 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.

6.4.3 Spatial Precision

To further investigate the cause of EMD's inconsistent performance, another performance index called *spatial precision* [29] was computed. Spatial precision measures, for a given image I, the proportion of images within a given distance d from I that belong to the same class as I. The distance d is usually defined in terms of the distance to the k-th nearest image of the same class as I. Figure 7 plots the spatial precision averaged over all 2000 images for each k. The spatial precision of the dissimilarity measures is smaller for a smaller image scaling factor and decreases with increasing value of k.

The result shows that as the value of k (i.e., the neighborhood size) increases, more negative samples that belong to other classes are included in the neighborhood. However, given the large number of classes (100) in the test, it is possible that only a small number of negative samples from each class is included. As a result, the majority class can still be the

correct class even when there are many negative samples. This is especially true for EMD since its spatial precision decreases faster than those of other dissimilarity measures.

In the tests that we conducted, CIE94 distance was used as the ground distance of EMD because it is more perceptually uniform than Euclidean distance in the CIELAB space. It may be possible to improve EMD's performance by using a different, hopefully more appropriate ground distance. For instance, Puzicha et al. used $1 - g$, where g is a Gaussian, as the ground distance in their performance evaluation [27]. Nevertheless, the appropriate choice of the ground distance can only be determined empirically, and this implies that EMD is less convenient to use and less robust than the other dissimilarity measures.

6.5 Image Clustering

Clustering of images or image regions have been used in content-based image retrieval as a preprocessing step of image or region classification. In the case of clustering based on fixed color histograms of images, a typical method is to regard each histogram bin as representing a dimension in a high-dimensional space in which clustering is performed (e.g., [10]). This method implicitly assumes that Euclidean distance is a reliable measure of histogram dissimilarity, which we have shown to be false in the previous sections. Although there are other more reliable similarities such as JD, no easy way of computing mean histogram based on these similarities exists—because there is no easy way to compute a histogram that minimizes the sum distance to the histograms in a cluster. An alternative may be to apply k-medoid clustering algorithms [16, 17] but they are computationally more expensive than k-means clustering. In the case of adaptive histograms, k-means clustering is still applicable

because a simple and mathematically correct method of computing mean histogram exists, i.e., histogram merging (Section 5).

As an example, we describe a version of k-means clustering for adaptive histograms.

k-means clustering for adaptive histograms

Pick k histograms as the centroids M_i of clusters C_i.

Repeat

For each histogram H_j,

Group H_j to the nearest cluster C_i:

$d(H_j, M_i) \leq d(H_j, M_l)$ for all l.

For each cluster C_i,

Compute new centroid M_i of cluster C_i:

$$M_i = \biguplus_{H_j \in C_i} \overline{H}_j,$$

Merge bins of M_i by applying the adaptive binning algorithm (Section 3).

6.5.1 Test Setup

400 composite images from 20 classes (20 from each class) were randomly chosen from the images generated for the retrieval test. The composite images that contained the same embedded image should be closer to each other then to the other images. Three sets of tests were performed using the following combinations of color histograms and dissimilarity measures: (1) fixed clustered histograms with Euclidean distance and Euclidean mean (c + L_2), (2) fixed clustered histograms with JD for cluster assignment and Euclidean mean for com-

puting cluster centroid (c + JD/L_2), and (3) adaptive histograms with weighted correlation dissimilarity and histogram merging (a + WC). For the first two cases, an ordinary k-means clustering was used. For the third case, the k-means clustering for adaptive histograms was used. For each case, separate clustering tests were conducted with the number of clusters ranging from 5 to 40.

6.5.2 Results and Discussion

Clustering performance is measured in terms of the *cluster spread* and *cluster homogeneity*. The cluster spread Ω is the effective radius of a cluster normalized by its distance to the nearest neighboring cluster:

$$\Omega = \frac{1}{k}\sum_{i=1}^{k}\omega_i \tag{18}$$

$$\omega_i = \frac{\frac{1}{|C_i|}\sum_{H_j \in C_i} d(M_i, H_j)}{\min_{j \neq i} d(M_i, M_j)} \tag{19}$$

where M_i is the mean histogram of cluster C_i, $d(\cdot)$ is the CIE94 distance, and k is the number of clusters. It measures the compactness of the clusters and the amount of overlaps between the clusters. The smaller the cluster spread, the more compact are the clusters and the less are the overlaps between them.

The cluster homogeneity Θ measures the proportion of histograms in a cluster that belong to the majority class of the cluster:

$$\Theta = \frac{1}{k}\sum_{i=1}^{k} P(L(C_i)\,|\,C_i) \tag{20}$$

where $L(C_i)$ denotes the majority class of cluster C_i and $P(L(C_i) \mid C_i)$ is the conditional probability of $L(C_i)$ given C_i. If the cluster homogeneity is less than $1/n$, then the cluster must contain histograms that belong to at least $n+1$ classes. Therefore, the smaller the n, the larger than Θ, and the more homogeneous is the cluster.

Figure 8 compares the cluster spread and cluster homogeneity of the three test cases at different number of clusters. For all three cases, clustering performance improved significantly when the number of clusters k increased from 5 to 20. At $k > 20$, the cluster spreads of c + L_2 and c + JD/L_2 improved slightly with increasing k but their cluster homogeneity decreased. Notice that performing cluster assignment with JD did not improve clustering performance significantly because the computation of mean histogram was based on L_2 instead of the more reliable JD.

In contrast, the cluster spread and homogeneity of a + WC stabilized at $k > 20$, and were better than those of c + L_2 and c + JD/L_2 for all k. In other words, a + WC produced more compact and more homogeneous clusters that were more widely spaced out than did c + L_2 and c + JD/L_2. Moreover, its performance is more stable than those of the other two cases. This result indicates that a + WC is more effective and reliable for practical applications in which the optimal number of clusters k is often unknown.

7 Conclusions

This paper presented an adaptive color clustering method and a dissimilarity measure for comparing histograms with different binnings. The color clustering algorithm is an adaptive variant of the k-means clustering algorithm and it can determine the number of clusters

required to adequately 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 bins. Since this measure does not require optimization, it executes more efficiently than does Earth Mover's Distance (EMD).

Extensive tests were performed to evaluate the performance of adaptive histograms on color retention, image retrieval, image classification, and image clustering. Compared to fixed binning schemes, adaptive color clustering can retain color information more accurately with fewer bins and no empty bin. The combined performance of adaptive color clustering and weighted correlation dissimilarity (WC) is comparable to that of Jessen difference divergence and better than those of L_2 and EMD for image retrieval and image classification tasks. For image clustering, a variant of k-means clustering algorithm is adapted to cluster adaptive histograms. Test results show that WC performs better than L_2 and the JD/L_2 combination because it allows the clustering algorithm to produce more compact and more homogeneous clusters that are widely spaced out. In conclusion, the adaptive histograms achieve the best overall performance in terms of accuracy, small number of compact and homogeneous bins, no empty bin, and efficient computation for image retrieval, classification, and clustering tasks.

Acknowledgments

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

References

[1] IEC 61966-2.1. *Default RGB Colour Space - sRGB*. International Electrotechnical Commission, Geneva, Switzerland, 1999. see also www.srgb.com.

[2] R. S. Berns. *Billmeyer and Saltzman's Principles of Color Technology*. John Wiley & Sons, 3rd edition, 2000.

[3] E. Binaghi, I. Gagliardi, and R. Schettini. Image retrieval using fuzzy evaluation of color similarity. *Int. J. PR and AI*, 8:945–968, 1994.

[4] I. Borg and P. Groenen. *Modern Multidimensional Scaling*. Springer-Verlag, New York, 1997.

[5] R. Brunelli and O. Mich. Histograms analysis for image retrieval. *Pattern Recognition*, 34(8):1625–1637, 2001.

[6] 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.

[7] J. Burbea and C. R. Rao. On the convexity of some divergence measures based on entropy functions. *IEEE Trans. Information Theory*, 28(3):489–495, 1982.

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

[9] 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.

[10] C. Y. Fung and K. F. Loe. Learning primitive and scene semantics of images for classification and retrieval. In *Proc. ACM Multimedia '99*, pages II: 9–12, 1999.

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

[12] S.-S. Guan and M. R. Luo. Investigation of parametric effects using small colour differences. *Color Research and Application*, 24(5):331–343, 1999.

[13] 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.

[14] T. Indow. Predictions based on munsell notation. I. perceptual color differences. *Color Research and Application*, 24(1):10–18, 1999.

[15] H. Jeffreys. *Theory of Probability*. Oxford, 2nd edition, 1948.

[16] L. Kaufmann and P. J. Rousseeuw. Clustering by means of medoids. In Y. Dodge, editor, *Statistical Data Analysis based on the L_1 Norm and Related Methods*, pages 405–416. Elsevier Science, 1987.

[17] L. Kaufmann and P. J. Rousseeuw. *Finding Groups in Data: An Introduction to Cluster Analysis*. John Wiley, 1990.

[18] S. Kullback and R. A. Leibler. On information and sufficiency. *Annals of Mathematical Statistics*, 22:79–86, 1951.

[19] B. Manjunath and W. Ma. Texture features for browsing and retrieval of image data. *IEEE Trans. PAMI*, 18(8):837–842, 1996.

[20] J. E. Marsden and M. J. Hoffman. *Elementary Classical Analysis*. W. H. Freeman, 2nd edition, 1993.

[21] B. M. Mehtre, M. S. Kankanhalli, A. Desai, and G. C. Man. Color matching for image retrieval. *Pattern Recognition Letters*, 16:325–331, 1995.

[22] B. M. Mehtre, M. S. Kankanhalli, and W. F. Lee. Content-based image retrieval using a composite color-shape approach. *Information Processing & Management*, 34(1):109–120, 1998.

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

[24] N. M. Nasrabad and R. A. King. Image coding using vector quantization: A review. *IEEE Trans. Communications*, 36(8):957–971, 1988.

[25] W. Niblack, R. Barber, W. Equitz, M. Flickner, E. Glasman, D. Petkovic, P. Yanker, C. Faloutsos, and G. Taubin. The QBIC project: Querying images by content using color, texture, and shape. In *Proc. SPIE Conf. on Storage and Retrieval for Image and Video Databases*, volume 1908, pages 173–181, 1993.

[26] Z. Pečenović. Image retrieval using latent semantic indexing. Final Year Graduate Thesis, École Polytechnique Fédérale de Lausanne, Switzerland, 1997.

[27] 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.

[28] J. Puzicha, T. Hofmann, and J. Buhmann. Nonparametric similarity measures for unsupervised texture segmentation and image retrieval. In *Proc. CVPR '97*, pages 267–272, 1997.

[29] K. Rodden, W. Basalaj, D. Sinclair, and K. Wood. A comparison of measures for visualising image similarity. In *Proc. Challenges of Image Retrieval*, 2000.

[30] Y. Rubner. *Perceptual Metrics for Image Database Navigation*. PhD thesis, Computer Science Dept., Stanford U., 1999.

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

[32] S. Sclaroff, L. Taycher, and M. La 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.

[33] N. Sebe, M. S. Lew, and D. P. Huijsmans. Toward improved ranking metrics. *IEEE. Trans. on PAMI*, 22(10):1132–1143, 2000.

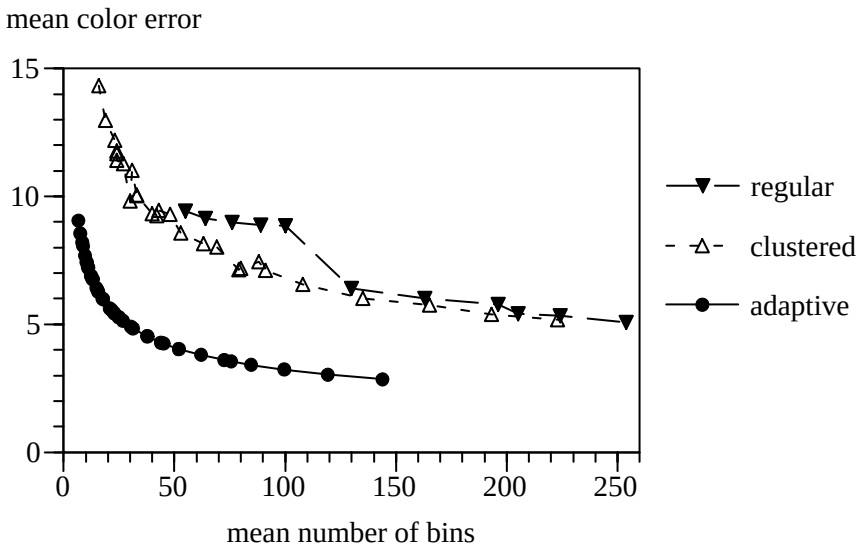
[34] G. Sheikholeslami, W. Chang, and A. Zhang. Semantic clustering and querying on heterogeneous features for visual data. In *Proc. ACM Multimedia '98*, pages 3–12, 1998.

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

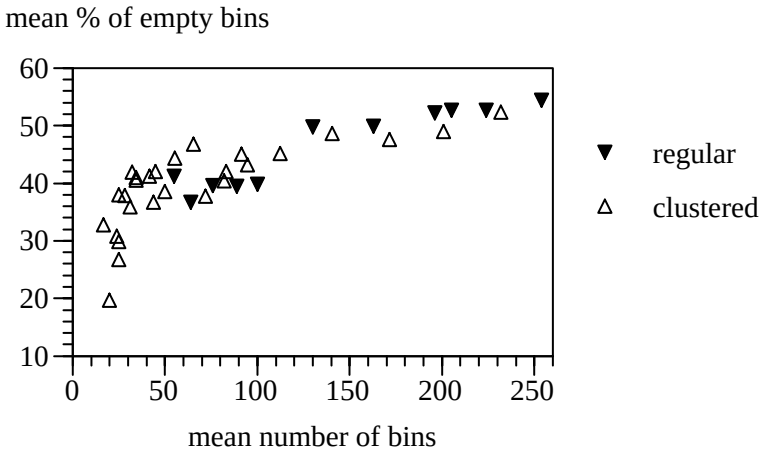
[36] T. Song and R. Luo. Testing color-difference formulae on complex images using a CRT monitor. In *Proc. of 8th Color Imaging Conference*, 2000.

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

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

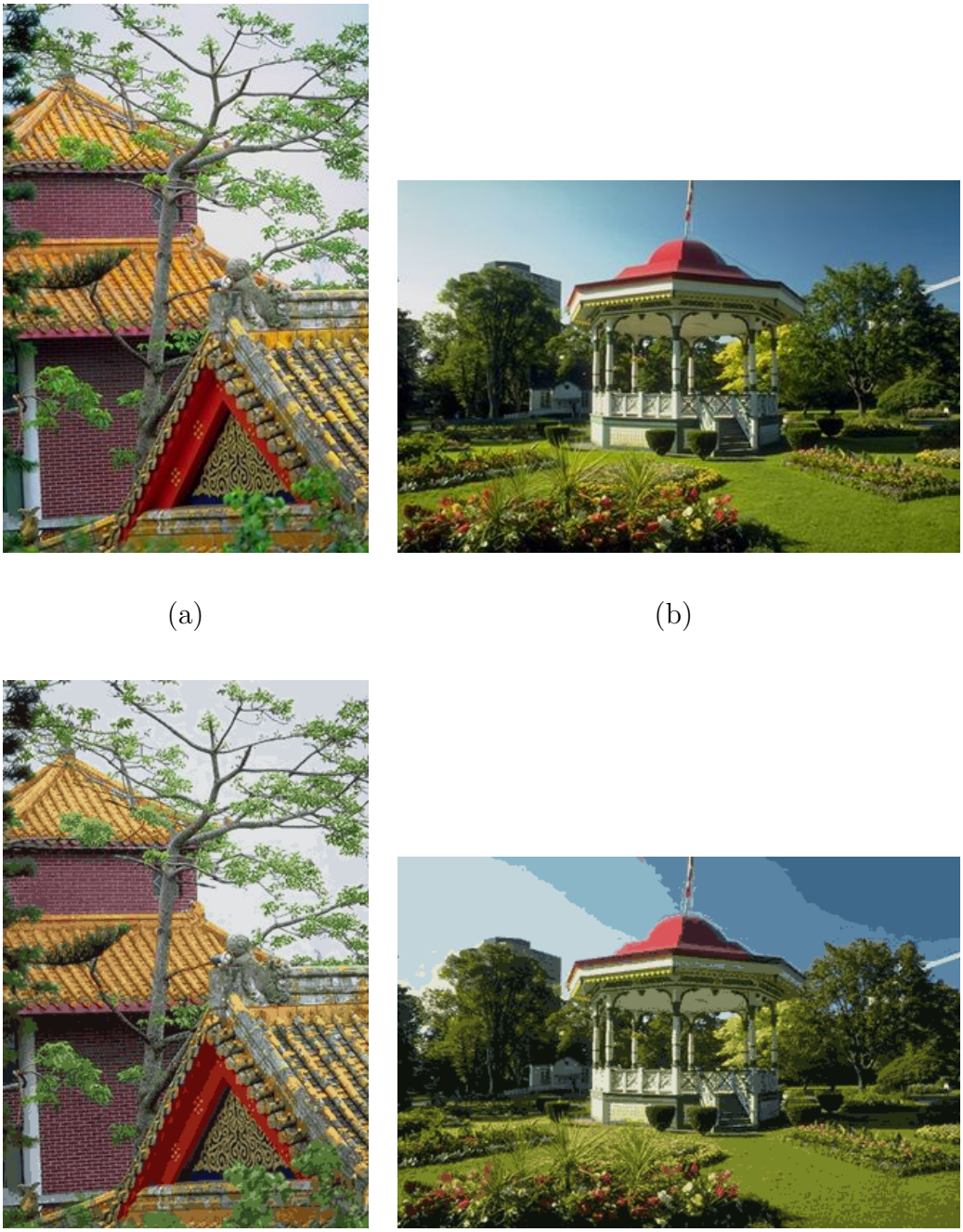


(a)



(b)

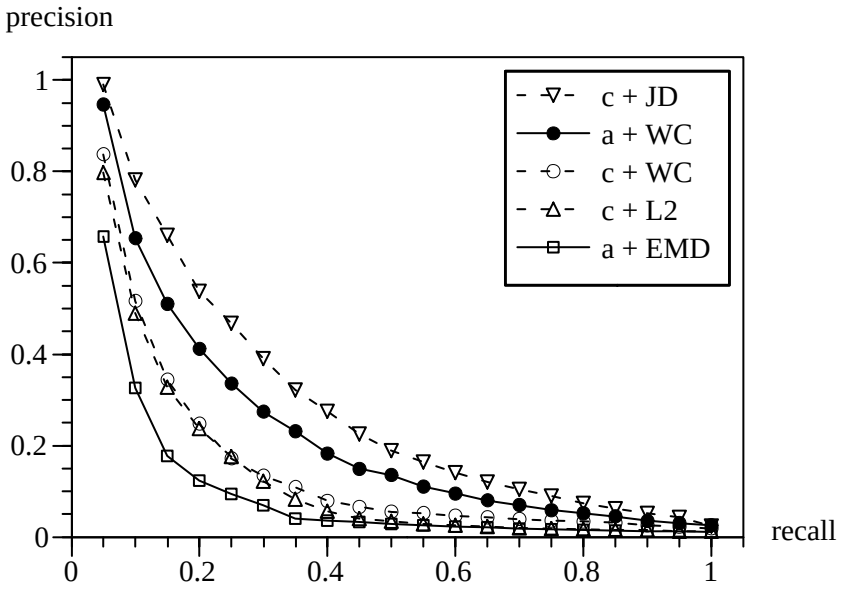
Figure 3: Color clustering performance. (a) Mean color errors of regular, clustered, and adaptive histograms. (b) Average percentage of empty bins in regular and clustered histograms.



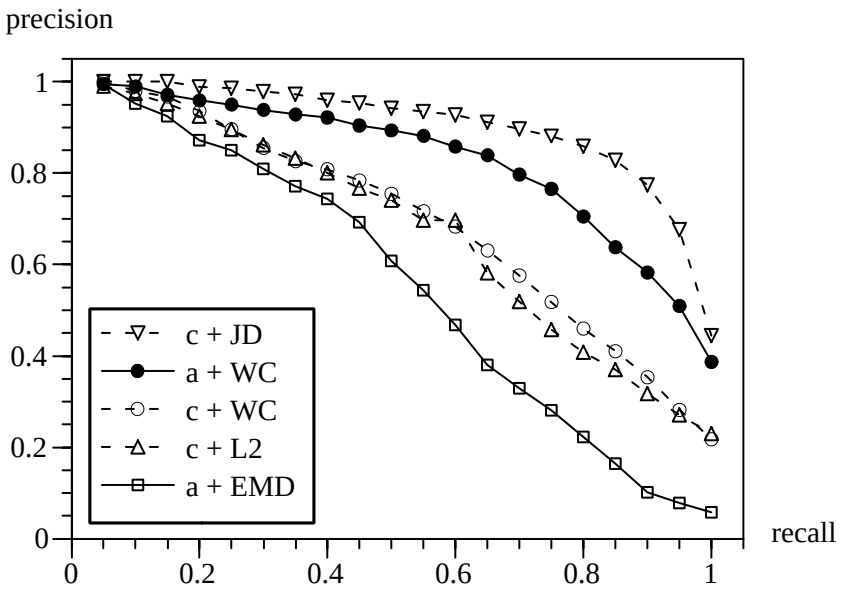
(a) (b)

(c) (d)

Figure 4: Color quantization results. The original images contain (a) 71599 colors and (b) 46218 colors. The color-quantized images contain only (c) 39 colors and (d) 31 colors, and are visually indistinguishable from the original images except for the regions where banding occurs.

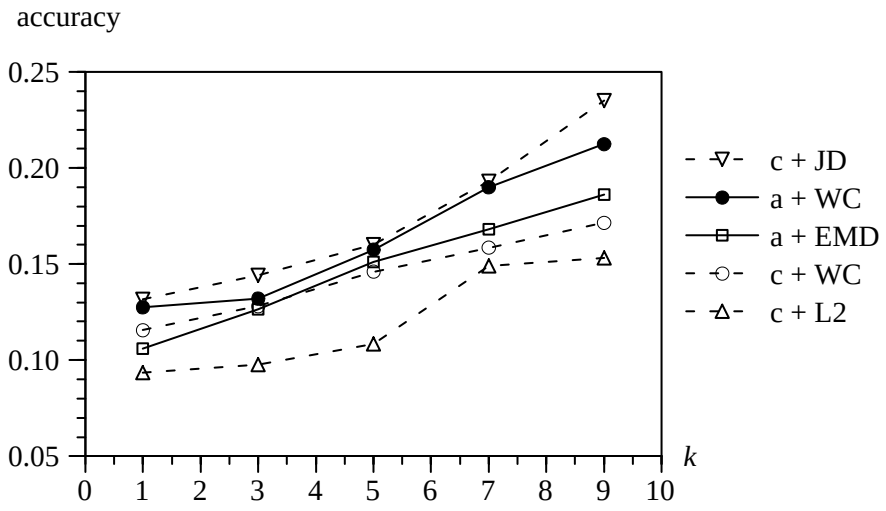


(a)

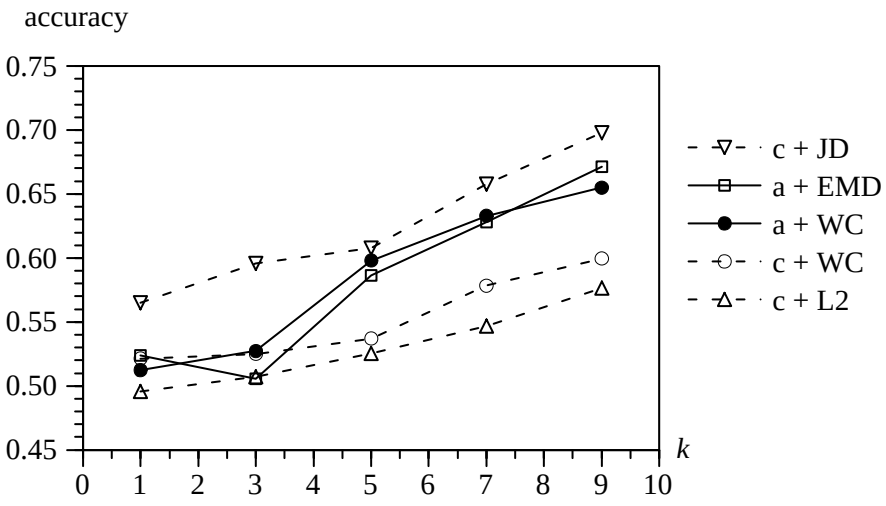


(b)

Figure 5: 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.

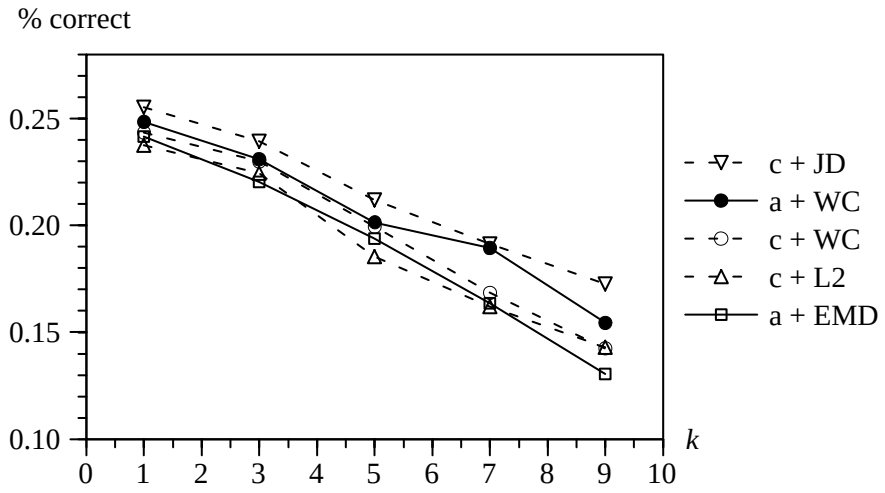


(a)

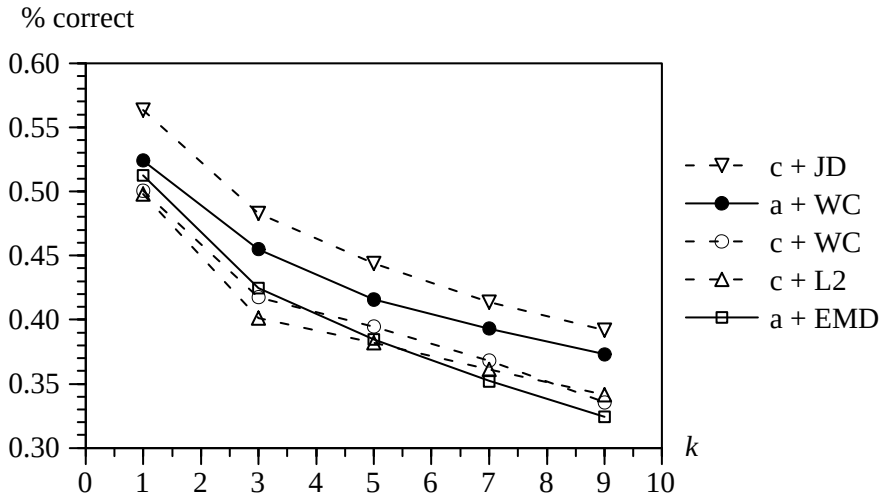


(b)

Figure 6: Classification accuracy of various combinations of binning methods and dissimilarities. (a) Scaling = 1/2, (b) scaling = 3/4.

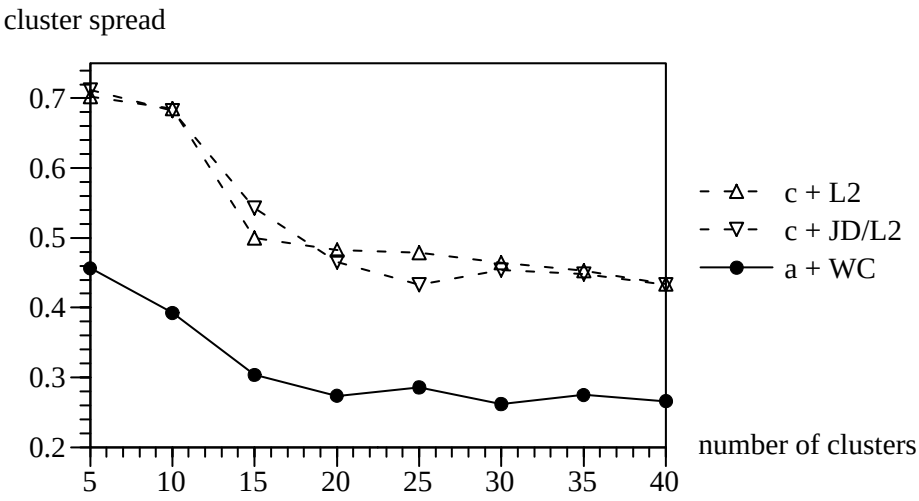


(a)

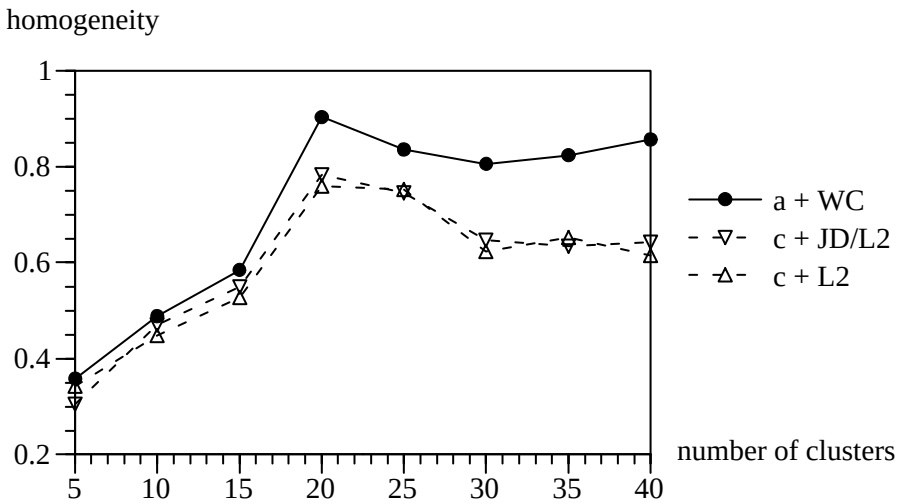


(b)

Figure 7: Spatial precision of various combinations of binning methods and dissimilarity measures. (a) Scaling = 1/2, (b) scaling = 3/4.



(a)



(b)

Figure 8: Comparison of (a) cluster spread and (b) cluster homogeneity between the three test cases.