The Algebra and Analysis of Adaptive-Binning Color Histograms

Wee Kheng Leow
Technical Report Number TRB8/02
Dept. of Computer Science, National University of Singapore,
3 Science Drive 2, Singapore 117543
leowwk@comp.nus.edu.sg

August 27, 2002

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.

Another major concern is that fixed-binning histograms have been regarded as vectors in a linear vector space so that powerful algorithms such as clustering, PCA, and SVD can be applied to process and analyze the histograms. Unfortunately, this approach is not satisfactory because the natural distance measure in linear vector space, the Euclidean distance, is less reliable than other non-Euclidean dissimilarities are in measuring histogram dissimilarity, thus compromising the effectiveness and reliability of the approach.

This technical report addresses the above issues in a mathematically sound manner. In this article, adaptive histograms are formally defined and are provided with a well-defined set of operations. Properties of the operations on adaptive histograms are analyzed, leading to mathematically sound definitions of similarity, dissimilarity, and mean of histograms. Extensive test results in [10] have shown 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 and classification tasks.

Index terms: Color histograms, adaptive binning, histogram-based dissimilarity measures.

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 [4, 5, 12, 15, 19, 21, 23]. 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 [2, 7, 13, 17, 18]. 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 [7, 17, 18]. 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 [17, 18]. 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 [6, 16, 20]. 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 dissimilarities in measuring histogram dissimilarity [17]. 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) [3] can be used to recover the Euclidean coordinates of the histograms from pairwise distances between them, this procedure is computationally expensive to apply on a large number (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, the application of histograms must be laid on a solid mathematical foundation. It is with this aim in mind that this technical report defines the algebra (i.e., set of operators) of a class of adaptive histograms that are useful for content-based image retrieval and other related applications (Section 3). With the algebra defined, we can analyze the properties of adaptive histograms, giving rise to well-founded definitions of similarity, dissimilarity, and mean histogram (Section 4), which are essential for many powerful procedures such as classification and clustering. Unlike Earth Mover's Distance, the dissimilarity measure defined in this article does not require an optimization process, thus making it much more efficient to compute. Detailed quantitative evaluation of the performance of adaptive histograms for image retrieval and classification have already be presented in [10].

2 An Overview

Before discussing the mathematics of adaptive histograms, let us motivate the mathematical formulation by first describing a possible definition of dissimilarity 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 distributions of colors in and around the two bins, where $\mathbf{x}$ denote

3D color coordinates. Then, the similarity $S(H, H')$ between the two distributions can be defined, as is commonly practiced, as the correlation between them:

$$S(H, H') = \int f(\mathbf{x})\, f'(\mathbf{x})\, d\mathbf{x}\ . \tag{1}$$

Equation 1 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. 1 has to be integrated over the intersecting volume $\mathcal{V}$ only, yielding:

$$S(H, H') = \int_{\mathcal{V}} \frac{h}{V} \frac{h'}{V'}\, d\mathbf{x} = \frac{V_s}{VV'} hh' \tag{2}$$

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 [1]. Recent psychological tests have confirmed that these color-difference equations are more perceptually uniform than does Euclidean distance in the CIELAB and CIELUV spaces [1, 8, 9, 14, 22].

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:

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

where $V = 4\pi R^3/3$ is the volume of a sphere. The bin separation d can be specified as a multiple of R, i.e., $d = \alpha R$, and the weight $w(\mathbf{c}, \mathbf{c}')$ can be defined as

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

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 $S(H, H')$ can thus be defined as follows:

$$S(H, H') = \sum_{i=1}^{n} \sum_{j=1}^{n'} w(\mathbf{c}_i, \mathbf{c}'_j) h_i h'_j\ . \tag{5}$$

In the following sections, we shall provide the mathematical foundation for this form of similarity measure.

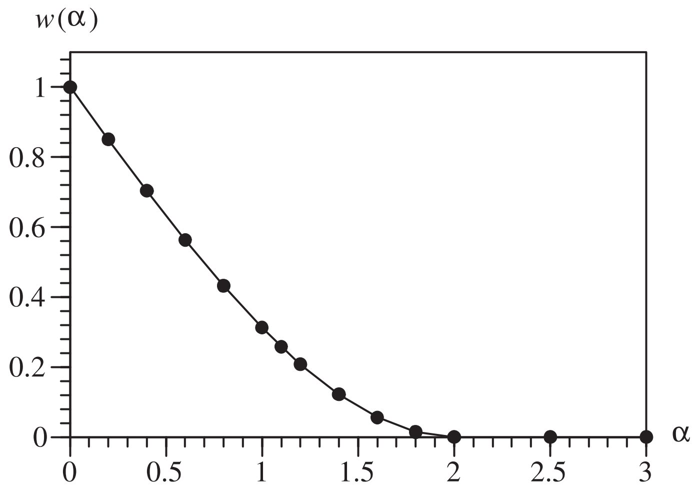

Figure 1: A plot of bin similarity $w(\alpha)$ against bin separation ratio α.

3 Algebra of Adaptive Histograms

To operate on and to analyze adaptive histograms, it is necessary to first provide a well-defined set of operators on these histograms. This section presents general definitions of adaptive histograms and operations on them.

3.1 Bins and Histograms

First, let us give a formal definition to histogram bins.

Definition 1 (Bin) *A bin is a region in an N_f-dimensional feature space identified by a vector $\mathbf{c}$ denoting the centroid of the region.*

Definition 2 (Bin Similarity) *The similarity $w(\mathbf{b}, \mathbf{c})$ between bins $\mathbf{b}$ and $\mathbf{c}$ is given by a monotonic function inversely related to the distance $d(\mathbf{b}, \mathbf{c})$ between them. Bin similarity is symmetric $w(\mathbf{b}, \mathbf{c}) = w(\mathbf{c}, \mathbf{b})$ and bounded $0 \leq w(\mathbf{b}, \mathbf{c}) \leq 1$.*

For mathematical convenience, Euclidean distance may be used for $d(\mathbf{b}, \mathbf{c})$. Now, adaptive histograms can be defined as follows.

Definition 3 (Adaptive Histogram) *An adaptive histogram $H = (n, \mathcal{C}, \mathcal{H})$ is a 3-tuple consisting of a set $\mathcal{C}$ of n bins $\mathbf{c}_i$, $i = 1, \ldots, n$, and a set $\mathcal{H}$ of corresponding bin counts $h_i \geq 0$. The set of bins of H is also denoted as $\mathcal{C}(H)$.*

Definition 4 (Null Histogram) *A null histogram $O = (n, \mathcal{C}, \{h_i\})$ has only empty bins, i.e., $h_i = 0$ for all $\mathbf{c}_i \in \mathcal{C}$, $i = 1, \ldots, n$.*

3.2 Operators

Next, we define appropriate operators on adaptive histograms.

Definition 5 (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 identical set of bins.

Definition 6 (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 subsumes addition and union.

Definition 7 (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 7 for histogram merging to be well-defined.

Definition 8 (Identity Element) *For any histogram H and null histogram O, $H + O = H \cup O = H \uplus O = H$.*

Strictly speaking, $H \cup O$ and $H \uplus O$ result in histograms containing empty bins collected from O. Definition 8 allows empty bins to be ignored or considered removed. Conversely, empty bins can be arbitrarily added to a histogram without affecting its mathematical properties.

The next two operations are analogous to, but different from, scalar product and inner product of vectors.

Definition 9 (Scalar Product) *The scalar product of a non-negative scalar k and a histogram $H = (n, \mathcal{C}, \{h_i\})$ is $kH = (n, \mathcal{C}, \{kh_i\})$.*

Definition 10 (Weighted Correlation) *The weighted correlation between histograms $G = (m, \{\mathbf{b}_i\}, \{g_i\})$ and $H = (n, \{\mathbf{c}_i\}, \{h_i\})$, denoted as $G \cdot H$, is defined as follows:*

$$G \cdot H = \sum_{i=1}^{m} \sum_{j=1}^{n} w(\mathbf{b}_i, \mathbf{c}_j)\, g_i h_j \ . \tag{6}$$

Weighted correlation is non-negative $G \cdot H \geq 0$ and commutative $G \cdot H = H \cdot G$ because the bin counts g_i and h_j are non-negative and the bin similarities $w(\mathbf{b}_i, \mathbf{c}_j)$ are non-negative and symmetric. The null histogram O is totally uncorrelated to any non-null histogram H: $H \cdot O = 0$.

Having defined the adaptive histograms and operators on them, we now prove the algebraic properties of the operators.

Theorem 11 (Commutativity, Associativity, Linearity, Distributivity)

$$
\begin{aligned}
G \; op \; H &= H \; op \; G && (7)\\
(F \; op \; G) \; op \; H &= F \; op \; (G \; op \; H) && (8)\\
k(G \; op \; H) &= kG \; op \; kH && (9)\\
F \cdot (G \; op \; H) &= F \cdot G + F \cdot H && (10)
\end{aligned}
$$

where op denotes either the $+$, $\cup$, or $\uplus$ operator.

Proof. The proofs of commutativity (Eq. 7), associativity (Eq. 8), and linearity (Eq. 9) are straightforward and are omitted. The proof of distributivity (Eq. 10) is given below. Let $F = (l, \{\mathbf{a}_i\}, \{f_i\})$, $G = (m, \{\mathbf{b}_i\}, \{g_i\})$, and $H = (n, \{\mathbf{c}_i\}, \{h_i\})$. For histogram addition, let $m = n$ and $\{\mathbf{b}_i\} = \{\mathbf{c}_i\}$. Then,

$$
\begin{aligned}
F \cdot (G + H) &= \sum_i \sum_j w(\mathbf{a}_i, \mathbf{b}_j) f_i \, (g_j + h_j)\\
&= \sum_i \sum_j w(\mathbf{a}_i, \mathbf{b}_j) f_i \, g_j + \sum_i \sum_j w(\mathbf{a}_i, \mathbf{c}_j) f_i \, h_j = F \cdot G + F \cdot H \; .
\end{aligned}
$$

For histogram union, let $\{\mathbf{b}_i\} \cap \{\mathbf{c}_i\} = \emptyset$. Then,

$$
F \cdot (G \cup H) = \sum_i \sum_j w(\mathbf{a}_i, \mathbf{b}_j) f_i \, g_j + \sum_i \sum_k w(\mathbf{a}_i, \mathbf{c}_k) f_i \, h_k = F \cdot G + F \cdot H \; .
$$

For histogram merging, G and H can be written in the form $G = X \cup Y$ 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,

$$
\begin{aligned}
F \cdot (G \uplus H) &= F \cdot ((X + X') \cup Y \cup Z) = F \cdot X + F \cdot X' + F \cdot Y + F \cdot Z\\
&= F \cdot (X \cup Y) + F \cdot (X' \cup Z) = F \cdot G + F \cdot H
\end{aligned}
$$

□

The next operator is the norm operator, which is required to provide a well-defined concept of histogram similarity and dissimilarity.

Definition 12 (Norm) *The norm $\|H\|$ of a histogram H is defined as $\sqrt{H \cdot H}$.*

Definition 13 (Normalized Histogram) *The normalized histogram $\overline{H}$ of a histogram H is defined as $\overline{H} = kH$ where $k = 1/\|H\|$, or in short, $\overline{H} = H/\|H\|$.*

Note that $\|\overline{H}\| = 1$.

Theorem 14 (Scalar Product) *For any non-negative scalar k and histogram H, $\overline{kH} = \overline{H}$.*

Proof. $\|kH\|^2 = kH \cdot kH = k^2 H \cdot H = k^2 \|H\|^2$. Therefore,

$$
\overline{kH} = \frac{kH}{\|kH\|} = \frac{kH}{k\|H\|} = \frac{H}{\|H\|} = \overline{H}.
$$

□

This theorem says that all scalar multiples of a histogram have the same normalized form.

Corollary 15 (Scalar Product) *For any non-negative scalar k and histogram H, $\|kH\| = k\|H\|$.*

Proof. See the proof of Theorem 14. □

4 Analysis of Adaptive Histograms

Recall that a bin is a region in an N_f-dimensional feature space. A histogram is a collection of bins and the corresponding bin counts. It is thus a distribution of the bins in the N_f-dimensional space. In practice, the bins are usually sparsely distributed (as a result of appropriate color clustering, for instance). Consequently, the number of neighbors that a bin overlaps is much smaller than the total number of bins in the histogram. This notion of sparse distribution of bins is used in the analysis of the properties of adaptive histograms.

4.1 Histogram Similarity

Definition 16 (Histogram Similarity) *The similarity $s(G, H)$ between histograms G and H is defined as the weighted correlation between their normalized forms: $s(G, H) = \overline{G} \cdot \overline{H}$. The dissimilarity $d(G, H)$ between them is defined as $d(G, H) = 1 - s(G, H)$.*

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:

$$(G \cdot H)^2 \leq (G \cdot G)\,(H \cdot H) \tag{11}$$

or equivalently

$$G \cdot H \leq \|G\| \|H\| \ . \tag{12}$$

Unfortunately, its complete proof in general is exceedingly difficult to furnish. Instead, a necessary condition and a sufficient condition for Cauchy-Schwarz inequality are provided, supplemented by examples and discussion about practical issues.

Theorem 17 (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,

$$(G \cdot H)^2 = \left(\sum_{i,k} w(\mathbf{b}_i, \mathbf{c}_k)\, g_i h_k \right)^2 = \sum_{i,j,k,l} w(\mathbf{b}_i, \mathbf{c}_k)\, w(\mathbf{b}_j, \mathbf{c}_l)\, g_i g_j h_k h_l \tag{13}$$

and

$$(G \cdot G)\,(H \cdot H) = \sum_{i,j} w(\mathbf{b}_i, \mathbf{b}_j)\, g_i g_j \sum_{k,l} w(\mathbf{c}_k, \mathbf{c}_l)\, h_k h_l = \sum_{i,j,k,l} w(\mathbf{b}_i, \mathbf{b}_j)\, w(\mathbf{c}_k, \mathbf{c}_l)\, g_i g_j h_k h_l \;. \quad (14)$$

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. Histograms that satisfy Cauchy-Schwarz inequality must satisfy the necessary condition but may or may not satisfy the sufficient condition. Conversely, histograms that satisfy the sufficient condition also satisfy Cauchy-Schwarz inequality (Fig. 2). 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 [11]. 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 $H = (2, \{\mathbf{b}_i\}, \{1, 1/4\})$ 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 $H = (2, \{\mathbf{b}_i\}, \{1, 1/4\})$ 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, Cauchy-Schwarz inequality does not hold:

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

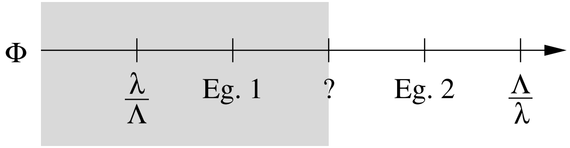

Figure 2: Cauchy-Schwarz inequality is satisfied when Φ falls in the shaded region. $\Phi \leq \Lambda/\lambda$ gives the necessary condition and $\Phi \leq \lambda/\Lambda$ gives the sufficient condition.

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. 4, 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 18 (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,

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

□

In addition, the triangle inequality of histogram merging operation also follows from Cauchy-Schwarz inequality:

Theorem 19 (Triangle Inequality) *For any histograms G and H satisfying Cauchy-Schwarz inequality,*

$$\|G \uplus H\| \leq \|G\| + \|H\| \;.$$

Proof. The proof follows from Cauchy-Schwarz inequality:

$$\begin{aligned}
\|G \uplus H\|^2 &= (G \uplus H) \cdot (G \uplus H) = G \cdot G + 2G \cdot H + H \cdot H \\
&\leq G \cdot G + 2\left((G \cdot G)\,(H \cdot H)\right)^{1/2} + H \cdot H \\
&= \|G\|^2 + 2\|G\|\|H\| + \|H\|^2 = (\|G\| + \|H\|)^2
\end{aligned}$$

□

The last concept to be introduced in this section is the equivalence of histograms. It is an important concept that formalizes the notion of replacing a histogram by another without loss of information.

Definition 20 (Histogram Equivalence) *Histogram G is equivalent to histogram H, denoted as $G \equiv H$, if $s(X, G) = s(X, H)$ for all X.*

In other words, G is indistinguishable from H as far as other histograms X are concerned. Note that $s(X, G) = s(X, H)$ does not imply $G = H$. For example, $kH \neq H$ for $k \neq 1$. But, $\overline{kH} = \overline{H}$ (Theorem 14), which implies $s(X, kH) = s(X, H)$. So, kH and H are equivalent but not identical.

4.2 Histogram Space

Adaptive histograms can be regarded as points in a *histogram space* $\mathcal{S}$, which can be approximately recovered from the dissimilarity between the histograms using methods such as multidimensional scaling (MDS) [3]. The dimensionality of $\mathcal{S}$ is not determined by the number of bins in the histograms but by the goodness of fit of MDS. In general, the larger the number of dimensions, the better is the fit. The construction of the histogram space can be impractical when the number of histograms is large, say, more than 100. It turns out that we can operate on the histograms, such as computing the mean histogram, without knowing explicitly the coordinates of the histograms in $\mathcal{S}$. Therefore, there is no need to explicitly construct the space $\mathcal{S}$.

Definition 21 (Bounded Space) *The space $\mathcal{S}$ of histograms is referred to as a bounded space if Cauchy-Schwarz inequality holds for any histograms G and H in $\mathcal{S}$.*

Next, we prove closedness of bounded histogram space under two most important operations: scalar multiplication and histogram merging. The closedness theorems allow us to operate on the histograms without worrying that histogram dissimilarity may become negative and, thus, ill-defined.

Theorem 22 (Closedness) *The bounded histogram space $\mathcal{S}$ is closed under scalar multiplication.*

Proof. Suppose that histograms $X, H \in \mathcal{S}$. Then, for any scalar $k \geq 0$,

$$(X \cdot kH)^2 = k^2(X \cdot H)^2 \leq k^2(X \cdot X)\,(H \cdot H) = (X \cdot X)\,(kH \cdot kH)\ .$$

Hence, Cauchy-Schwarz inequality holds and $\mathcal{S}$ is closed under scalar multiplication □

Theorem 23 (Closedness) *The bounded histogram space $\mathcal{S}$ is closed under the merging of histograms $H_i \in \mathcal{S}$, $i = 1, \ldots, N$, if $s(H_i, H_j) \geq s(X, H_i)$ and $s(H_i, H_j) \geq s(X, H_j)$ for any i, j, and any $X \in \mathcal{S}$.*

Proof. By Theorem 11,

$$(X \cdot \biguplus_i H_i)^2 = \left(\sum_i X \cdot H_i\right)^2 = \sum_i \sum_j (X \cdot H_i)\,(X \cdot H_j) \tag{15}$$

and

$$(X \cdot X)\,(\biguplus_i H_i \cdot \biguplus_i H_i) = \sum_i \sum_j (X \cdot X)\,(H_i \cdot H_j)\ . \tag{16}$$

It is given that, for any i, j and $X \in \mathcal{S}$,

$$s(H_i, H_j) \geq s(X, H_i) \quad (17)$$
$$s(H_i, H_j) \geq s(X, H_j)\ . \quad (18)$$

Multiplying Eq. 17 and 18 yields

$$\frac{(H_i \cdot H_j)\, s(H_i, H_j)}{\|H_i\| \|H_j\|} \geq \frac{(X \cdot H_i)\,(X \cdot H_j)}{\|X\|^2\, \|H_i\| \|H_j\|}\ . \quad (19)$$

Arranging terms in Eq. 19 yields, for any i, j,

$$(X \cdot X)\,(H_i \cdot H_j) \geq (X \cdot X)\,(H_i \cdot H_j)\, s(H_i, H_j) \geq (X \cdot H_i)\,(X \cdot H_j) \quad (20)$$

because $s(H_i, H_j) \leq 1$. Therefore, summing the terms in Eq. 20 over all i, j and comparing with Eqs. 15, 16 yields

$$(X \cdot \biguplus_i H_i)^2 \leq (X \cdot X)\,(\biguplus_i H_i \cdot \biguplus_i H_i)$$

and the space $\mathcal{S}$ is closed under histogram merging. □

The condition given in the above theorem may seem restrictive. However, it does not severely limit the application of adaptive histograms in practice. What it means is that only histograms that are similar to each other should be merged. The following theorems provide explanations of why this should be the case.

First, we give a general definition of an average histogram.

Definition 24 (Mean Histogram) *M is a mean histogram of H_i, $i = 1, \ldots, N$, if M maximizes the total similarity $S(M)$:*

$$S(M) = \sum_{i=1}^{N} s(M, H_i) \quad (21)$$

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.

Notice that the computation of mean histogram based on Euclidean distance is straightforward:

$$M = \frac{1}{N} \sum_i H_i \quad (22)$$

because the Euclidean mean minimizes the sum-squared distance between the histograms and the mean. But, it is applicable only to fixed-binning histograms. On the other hand, the mean computation based on other non-Euclidean distances requires an optimization procedure that is computationally expensive in general. In contrast, the computation of mean histogram based on adaptive histogram dissimilarity (Definition 16) is as straightforward as that of a Euclidean mean, and yet is applicable to histograms with different binnings.

Theorem 25 (Mean Histogram) $\biguplus_i \overline{H}_i$ *is a mean of histograms H_i, $i = 1, \ldots, N$, in the bounded histogram space.*

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

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

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 $S(M)$. □

The usual definition of mean divides the sum by the number of items that are added together (e.g., Eq. 22). For adaptive histograms, this division is not necessary because histogram similarity and dissimilarity are defined in terms of normalized histograms and all scalar multiples of a histogram have the same normalized form (Theorem 14).

An implication of Theorem 25 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. This result ties in neatly with Theorem 23 which advises that only similar histograms should be merged.

The merging of several histograms may result in a histogram with a large number of bins. So, it might be useful to merge similar bins so as to reduce the number of bins.

Theorem 26 (Bin Merging) *Let $G_i = (1, \{\mathbf{b}_i\}, \{g_i\})$, $i = 1, \ldots, n$, and $\delta = \max_{i,j} \|\mathbf{b}_i - \mathbf{b}_j\|$. Define $M = (1, \{\mathbf{c}\}, \{h\})$ such that*

$$\mathbf{c} = \frac{1}{n} \sum_i \mathbf{b}_i \ , \quad h = \sum_i g_i \ .$$

Then, $M \equiv \bigcup_i G_i$ as $\delta \to 0$.

Proof. Let $X = (m, \{\mathbf{u}_i\}, \{x_i\})$ be any histogram. Then,

$$X \cdot M = \sum_j w(\mathbf{u}_j, \mathbf{c}) \, x_j h = \sum_j w(\mathbf{u}_j, \mathbf{c}) \, x_j \sum_i g_i \ .$$

As $\delta \to 0$, $\mathbf{b}_i \to \mathbf{c}$ and $w(\mathbf{u}_j, \mathbf{b}_i) \to w(\mathbf{u}_j, \mathbf{c})$ for each i. In the limit,

$$X \cdot M = \sum_i \sum_j w(\mathbf{u}_j, \mathbf{b}_i) \, x_j g_i = \sum_i X \cdot G_i = X \cdot \bigcup_i G_i \ .$$

Moreover,

$$\|\bigcup_i G_i\|^2 = \sum_i \sum_j w(\mathbf{b}_i, \mathbf{b}_j)\, g_i g_j \ .$$

As $\delta \to 0$,

$$\|\bigcup_i G_i\|^2 \to \sum_i \sum_j g_i g_j = \left(\sum_i g_i\right)^2 = \|M\|^2 \ .$$

Thus, $s(X, M) = s(X, \bigcup_i G_i)$, i.e., $M \equiv \bigcup_i G_i$ as $\delta \to 0$. □

Note that $M \neq \bigcup_i G_i$ but $M \equiv \bigcup_i G_i$ when $\delta \to 0$. In this theorem, we choose $\mathbf{c}$ to be the mean of bins $\mathbf{b}_i$ so that M is as similar to $\bigcup_i G_i$ as possible. For bins that are closely spaced (i.e., $\delta \approx 0$), an equivalent bin is obtained by merging the bins. The merged bin can replace the bins that are merged without loss of information.

5 Conclusion

To conclude, let us summarize the properties of adaptive histograms presented in this article and compare them with the properties of a linear vector space ([11], Definition 1.6.2 and Theorem 1.6.6). The properties for any histograms F, G, H and null histogram O, with non-negative scalar k include the following:

I. Properties of operators:

   (a) commutativity: $G \uplus H = H \uplus G$
   (b) additive identity: $H \uplus O = H$
   (c) associativity: $(F \uplus G) \uplus H = F \uplus (G \uplus H)$
   (d) distributivity: $k(G \uplus H) = kG \uplus H$
   (e) multiplicative identity: $kH = H$ for $k = 1$
   (f) associativity: $k(G \cdot H) = (kG) \cdot H$
   (g) distributivity: $(k + l) \cdot H = kH + lH$

II. Properties of weighted correlation (analogous to vector inner product):

   (a) positivity: $H \cdot H \geq 0$
   (b) nondegeneracy: $H \cdot H = 0$ iff $H = O$
   (c) distributivity: $F \cdot (G \uplus H) = (F \cdot G) + (F \cdot H)$
   (d) multiplicativity: $(kG) \cdot H = k(G \cdot H)$
   (e) symmetry: $G \cdot H = H \cdot G$

III. Properties of norm:

   (a) positivity: $\|H\| \geq 0$
   (b) nondegeneracy: $\|H\| = 0$ iff $H = O$

(c) multiplicativity: $\|kH\| = k\|H\|$

(d) triangle inequality: $\|G \uplus H\| \leq \|G\| + \|H\|$

IV. Properties of dissimilarity:

(a) positivity: $d(G, H) \geq 0$

(b) nondegeneracy: $d(G, H) = 0$ iff $G = H$

(c) symmetry: $d(G, H) = d(H, G)$

V. Cauchy-Schwarz inequality:

$(G \cdot H)^2 \leq (G \cdot G)(H \cdot H)$ or alternatively $G \cdot H \leq \|G\|\|H\|$

A space that satisfies the properties in Groups II and III, for all real k, is known as an inner product space and a normed space, respectively [11]. A space that satisfies Group IV properties as well as triangle inequality for distance, i.e., $d(F, H) \leq d(F, G) + d(G, H)$, is called a metric space [11].

In comparison, adaptive histograms do not have well-defined negatives because histograms do not have negative bin counts. As a result, the space of adaptive histograms is almost, but not exactly, a normed inner product space because Group II and III properties hold only for non-negative real scalar k. If we were to accept negative bin counts, the histogram norm as given in Definition 12 would be ill-defined, unless we accept complex norm, because $H \cdot H$ could be negative.

We feel that the algebra of adaptive histograms defined in this article provides a good compromise between mathematical soundness and practical usefulness. Adaptive histograms, together with an appropriate set of operators, form a space that possesses properties equivalent to those of a linear vector space, though a histogram space is not exactly identical to a linear vector space. In particular, similarity, dissimilarity, and mean of histograms have well-defined properties, making adaptive histograms useful for practical applications. At the same time, the computation of histogram similarity, dissimilarity, and mean are simple, straightforward, and does not require an optimization process. In contrast, the computation of Earth Mover's Distance and mean histogram based on other non-Euclidean distances require optimization procedures. The performance of adaptive histograms for image retrieval and classification have already been demonstrated in [10]. Its performance for clustering and image pyramid will be illustrated in an upcoming article.

Acknowledgments

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

References

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

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

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

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

[5] 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.

[6] 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.

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

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

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

[10] W. K. Leow and R. Li. Adaptive binning and dissimilarity measure for image retrieval and classification. In *Proc. IEEE CVPR*, 2001.

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

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

[13] 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.

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

[15] 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.

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

[17] 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.

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

[19] 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.

[20] 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.

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

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

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