Bin Ratio-Based Histogram Distances and Their Application to Image Classification

Weiming Hu, Nianhua Xie, and Ruiguang Hu

(National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing 100190)
{wmhu, nhxie, rghu}@nlpr.ia.ac.cn

Haibin Ling

(Department of Computer and Information Science, Temple University, Philadelphia, USA)
hbling@temple.edu

Qiang Chen

(IBM Research, Australia, Level 5, Lygon street, Carlton, VIC, Australia, 3053)
qiangchen@au1.ibm.com

Shuicheng Yan

(Department of Electrical and Computer Engineering, National University of Singapore, Singapore 117576)
eleyans@nus.edu.sg

Stephen Maybank

(Department of Computer Science and Information Systems, Birkbeck College, Malet Street, London WC1E 7HX)
sjmaybank@dcs.bbk.ac.uk

Abstract: Large variations in image background may cause partial matching and normalization problems for histogram-based representations, i.e. the histograms of the same category may have bins which are significantly different, and normalization may produce large changes in the differences between corresponding bins. In this paper, we deal with this problem by using the ratios between bin values of histograms, rather than bin values' differences which are used in the traditional histogram distances. We propose a bin ratio-based histogram distance (BRD), which is an intra-cross-bin distance, in contrast with previous bin-to-bin distances and cross-bin distances. The BRD is robust to partial matching and histogram normalization, and captures correlations between bins with only a linear computational complexity. We combine the BRD with the ℓ_1 histogram distance and the χ^2 histogram distance to generate the ℓ_1 BRD and the χ^2 BRD, respectively. These combinations exploit and benefit from the robustness of the BRD under partial matching and the robustness of the ℓ_1 and χ^2 distances to small noise. We propose a method for assessing the robustness of histogram distances to partial matching. The BRDs and logistic regression-based histogram fusion are applied to image classification. The experimental results on synthetic datasets show the robustness of the BRDs to partial matching, and the experiments on seven benchmark datasets demonstrate promising results of the BRDs for image classification.

Index terms: Histogram bin ratio, Histogram distance, Image classification

1. Introduction

Histogram-based representation is widely applied to many pattern recognition tasks, such as image or scene classification, visual appearance modeling, and visual action recognition, because of its simplicity and rich

discriminative information. In the bag-of-words model [8, 9, 23, 35, 38, 43], an image is represented using a histogram of the visual words obtained by quantizing visual patches, where each bin value in the histogram represents the probability of observing the corresponding word. Then, these histograms are used for image classification, object detection, and action recognition, etc. An efficient and effective measure of the distance (dissimilarity) between histograms plays an important role in histogram-based applications.

1.1. Related work

Currently, there exist several histogram distances [11, 21, 23, 24, 25, 35, 36, 43] which can be classified into bin-to-bin distances and cross-bin distances.

The bin-to-bin distances between two histograms are based on the differences of the corresponding bins in the histograms. Let $\mathbf{h} = \{h_i\}_{i=1}^n$ be a histogram for occurrence statistics with n bins where h_i represents the value of the i-th bin. The ℓ_1 and ℓ_2 distances between two histograms $\mathbf{h}^A$ and $\mathbf{h}^B$ are $\left\|\mathbf{h}^A - \mathbf{h}^B\right\|_1$ and $\left\|\mathbf{h}^A - \mathbf{h}^B\right\|_2$, respectively, where $\|\cdot\|_1$ and $\|\cdot\|_2$ are, respectively, the vector ℓ_1 and ℓ_2 norms. The histogram intersection [34] between two histograms $\mathbf{h}^A$ and $\mathbf{h}^B$ is $\sum_{i=1}^n \min(h_i^A, h_i^B)$ [13, 17, 22, 40]. When the areas of the two histograms are equal, the histogram intersection is equivalent to the ℓ_1 distance. The χ^2 distance [29] between two histograms $\mathbf{h}^A$ and $\mathbf{h}^B$ is [8, 23, 24, 35, 43]:

$$d_{\chi^2}(\mathbf{h}^A, \mathbf{h}^B) = 2\sum_{i=1}^{n} \frac{(h_i^A - h_i^B)^2}{h_i^A + h_i^B}. \tag{1}$$

The Bhattacharyya coefficient $B(\mathbf{h}^A, \mathbf{h}^B)$ between histograms $\mathbf{h}^A$ and $\mathbf{h}^B$ is

$$B(\mathbf{h}^A, \mathbf{h}^B) = \sum_{i=1}^{n} \sqrt{h_i^A h_i^B}. \tag{2}$$

The Bhattacharyya distance $D_B(\mathbf{h}^A, \mathbf{h}^B)$ between $\mathbf{h}^A$ and $\mathbf{h}^B$ is defined as: $D_B(\mathbf{h}^A, \mathbf{h}^B) = -\ln(B(\mathbf{h}^A, \mathbf{h}^B))$. The Jeffrey divergence $D_{Jd}(\mathbf{h}^A, \mathbf{h}^B)$ between histograms $\mathbf{h}^A$ and $\mathbf{h}^B$ is defined as:

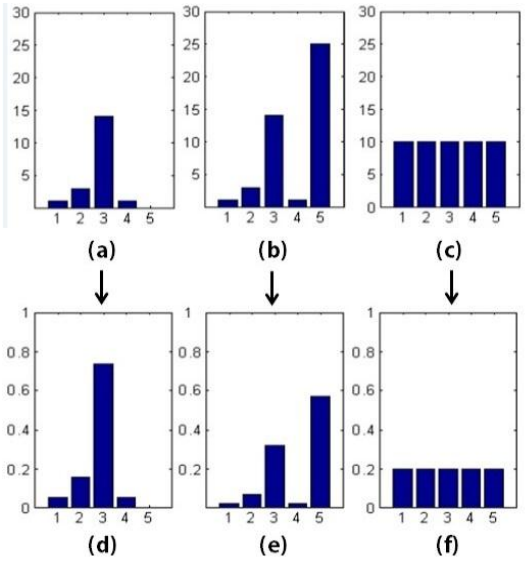
$$D_{Jd}(\mathbf{h}^A, \mathbf{h}^B) = \sum_{i=1}^{n} \left(h_i^A \ln\left(\frac{2h_i^A}{h_i^A + h_i^B}\right) + h_i^B \ln\left(\frac{2h_i^B}{h_i^A + h_i^B}\right)\right). \tag{3}$$

These bin-to-bin distances are widely used, because they are simple, efficient, and easy to implement.

The cross-bin distances [14, 20, 27, 30] allow cross bin comparison between two histograms to gain a more robust measure of their similarities. Rubner et al. [30] proposed a cross-bin distance, called the earth mover's distance (EMD), which is the first order Wasserstein distance. It reduces distance calculation to a transportation problem. Zhang et al. [43] showed that the EMD has an outstanding performance on various datasets. However, the time complexity of the EMD is $O(n^3 \log(n))$, which is very high. When the dimension of the feature vectors is large, the number of temporary variables required to compute the EMD is so large that internal memory overflows may be produced. As a result, performance of the EMD cannot be tested on large image datasets.

Although the existing histogram distances, either the bin-to-bin distances or the cross-bin distances, are effective in many applications, they still have limitations which are discussed as follows.

The first limitation is the effect of partial matching on bin values. The histograms of two images of the same category may have bins whose values are significantly different, due to various amounts of background clutter which are irrelevant to the foreground object or due to occlusions of the foreground object by other objects. Histograms are often normalized in visual recognition to adapt to large scale changes. However, normalization may produce large changes in the differences between corresponding bins in these histograms. As a result, it may be difficult to classify images using histograms. Fig. 1 shows an example of the partial matching problem. In the figure, (a) and (b) are histograms of two images in the same category, (c) is a reference histogram with a uniform distribution, and (d), (e), and (f) are the normalized histograms corresponding to (a), (b), and (c), respectively. While bins 1 to 4 are exactly the same in the histograms shown in (a) and (b), bin 5 is significantly different due to a large amount of background clutter in the image from which histogram (b) is computed. The table (g) shows that, before normalization, the distance between the histograms shown in (a) and (b) is smaller than the distance between the histograms shown in (a) and (c). The table (h) shows that, after normalization, the typical bin-to-bin distances, i.e., the ℓ_1 distance, the histogram intersection, the χ^2 distance, the Bhattacharyya distance, and the Jeffrey divergence, indicate that the histograms shown in (d) and (f) are more similar than the histograms shown in (d) and (e). The EMD is also strongly affected by partial matching and histogram normalization, because it depends on bin difference values. As a result, the partial matching problem influences the measures of similarities between images.



dist.	dis(a, b)	dis(a, c)
L_1	25	39

(g)

dist.	dis(d, e)	dis(d, f)
L_1	1.1364	1.0737
$\frac{1}{HI}$	2.3158	2.1591
χ^2	1.5873	1.3690
EMD	1.2560	0.8842
Bhatta.	0.4199	0.2655
Jeffrey	0.5097	0.3964
BRD	2.3376	2.8831

(h)

Fig. 1. An example of the effects of partial matching on the distances between histograms: (a), (b) and (c) show three histograms before normalization, where histograms in (a), (b), and (c) are [1, 3, 14, 1, 0], [1, 3, 14, 1, 25], and [10, 10, 10, 10, 10], respectively; (d), (e), and (f) are the corresponding normalized histograms; (g) shows the ℓ_1 distances between histograms shown in (a) and (b) and between histograms shown in (a) and (c) before normalization; (h) shows the distances between histograms shown in (d) and (e) and between histograms shown in (d) and (f), calculated using the ℓ_1 distance, the histogram intersection, the χ^2 distance, the Bhattacharyya distance, the Jeffrey divergence, the EMD, and the BRD.

Current histogram distances do not consider the correlations between pairs of bins in a histogram. These bin

correlations capture co-occurrences of visual words in the bag-of-words model. For example, the visual words "eye" and "mouth" usually appear together in face images, and the ratio of the frequencies of the visual words "eye" and "mouth" remains stable. Many techniques [1, 2, 16, 19, 21, 31, 33, 39] take into account joint word distributions and model the spatial co-occurrence of visual words. For instance, Agarwal and Triggs [1] proposed a hyper-feature which exploits spatial co-occurrence statistics of features. Li et al. [19] proposed a Markov stationary feature which uses Markov chain models to characterize the spatial co-occurrence of histogram patterns. Both Ling et al. [21] and Nguyen et al. [49, 50, 52] took into account the spatial distribution of code words by modeling the weak geometric context of images. They encoded the spatial co-occurrence statistics into the bag-of-features model by defining the proximity distribution kernel of quantized local features. Specifically, the co-occurrence statistics were encoded at low level with respect to the detected key points. The key-points-based features have an advantage over the conventional ones, e.g. Gabor features, because they are invariant to many geometric distortions and transformations. Other novel and robust image descriptors were also developed in [51, 53] for applications to visual recognition tasks. Inspired by the use of co-occurrence statistics in low level feature extraction in Ling and Soatto's work and Nguyen et al.'s work, we regarded that it is interesting to encode co-occurrence correlations between the different bins of a histogram in a non-parametric way in a histogram similarity measure. Fig. 1 is an example of co-occurrence correlations between different bins: the histogram bin correlations between the first four bins in histogram (a) are repeated in histogram (b). These histogram bin correlations are useful to produce a more accurate measure of similarity between histograms. The Mahalanobis distance, which is covariance-based, encodes correlations between different bins of a histogram. It is scale-invariant. But, when the dimension of the feature vectors is large, the covariance matrix is usually singular and does not have an inverse. The Moore-Penrose pseudo-inverse matrix may be used as an approximation. But its computation is very costly.

1.2. Our work

In this paper, we address the above challenges, and propose a bin-ratio-based histogram distance [41]. Bin ratios are defined as the ratios between histogram bin values. Given an *n*-bin histogram $\mathbf{h} \in \Re^n$, we define its ratio matrix as $H = (h_i / h_j) \in \Re^{n \times n}$. It contains the ratios defined by all the pairs of bins in the histogram. Given two histograms, we define their bin ratio-based distance (BRD) as the sum of the squared normalized differences over all the elements **of** their ratio matrices. The BRD is combined with the ℓ_1 distance and the χ^2 distance, to form two new measures: the ℓ_1 BRD and the χ^2 BRD. These BRDs (i.e., the BRD, the ℓ_1 BRD, and the χ^2 BRD) are applied to image classification. Logistic regression, which is a type of probabilistic statistical fusion model, is used to fuse multiple histogram distances for improving the accuracy of image classification.

The main contributions of our work are summarized as follows:

- The bin-ratio information in histograms is used to construct a new histogram distance, the BRD. In contrast with the existing histogram distances, the BRD is more robust to the effects of partial matching

resulting from background clutter and occlusions, as bin ratios of histograms describing the same object have **a** higher similarity. As an example, in Fig. 1(h) the BRD between the histograms in (d) and (e) is less than the BRD between the histograms in (d) and (f). The bin ratios capture correlations between pairs of bins. The BRD includes cross-bin information about the same histogram and forms a new type of histogram distance: the intra-cross-bin distance. The BRD has a linear computational complexity comparable to the complexity of the bin-to-bin distances, and much lower than the complexity of the cross-bin distances.

- The BRD is flexible and can be easily combined with other histogram measures to benefit from their advantages. In particular, we propose the ℓ_1 BRD and the χ^2 BRD which combine the properties of the BRD and the properties of the ℓ_1 distance and the χ^2 distance.
- We propose a method for assessing the robustness of histogram distances to partial matching. We also propose image classification methods based on the BRDs and the logistic regression fusion.

Extensive experimental results show the robustness of the BRDs to partial matching, and illustrate very promising results when the ℓ_1 BRD and the logistic regression-based histogram fusion are used to classify natural images.

The rest of the paper is organized as follows: Section 2 proposes the BRD. Section 3 presents the ℓ_1 BRD and the χ^2 BRD. Section 4 describes the assessment of the robustness of histogram distances to partial matching. Section 5 describes kernel-based image classification using the BRDs and logistic regression, and reports the experimental results. Section 6 concludes the paper.

2. Bin Ratio-Based Histogram Distance

Histogram bin ratios are unchanged by normalization although bin values are changed. It is intuitive that the ratios of bins for the foregrounds in the images in the same category are overall stable. The bin correlations, i.e., joint frequencies of visual words, are included in the ratios between bins. These observations motivate the construction of a new histogram distance based on the ratio relations between bins, in order to yield more robust image classification results.

The ℓ_2 normalization and the ℓ_1 normalization are two typical histogram normalization methods. If the Euclidean distance measure or the cosine distance measure is used, the ℓ_2 normalization is more appropriate. If the ℓ_1 distance measure or the χ^2 distance measure is used, the ℓ_1 normalization is more appropriate. While the ℓ_1 normalization is popular for histogram statistics, the ℓ_2 histogram normalization is widely used in the computer vision community. For example, Felzenszwalb et al. [42] explicitly pointed out that the ℓ_2 histogram normalization was applied to the HoG (Histogram of Oriented Gradients) feature [37]. We use the ℓ_2 histogram normalization and the square distance measure to define the BRD.

An ℓ_2 normalized histogram with *n* bins is a column vector $\mathbf{h} \in \Re^n$, such that

$$\|\mathbf{h}\|_2^2 = \sum_{k=1}^{n} h_k^2 = 1. \tag{4}$$

To capture pairwise relations between bins, we define the ratio matrix $H \in \Re^{n \times n}$ of **h**:

$$H = \left(\frac{h_j}{h_i} \right)_{1 \le i, j \le n} = \begin{pmatrix} \frac{h_1}{h_1} & \frac{h_2}{h_1} & \frac{h_3}{h_1} & \bullet & \frac{h_n}{h_1} \\ \frac{h_1}{h_2} & \frac{h_2}{h_2} & \frac{h_3}{h_2} & \bullet & \frac{h_n}{h_2} \\ \bullet & \bullet & \bullet & \bullet & \bullet \\ \frac{h_1}{h_n} & \frac{h_2}{h_n} & \frac{h_3}{h_n} & \bullet & \frac{h_n}{h_n} \end{pmatrix} = \begin{pmatrix} \frac{\mathbf{h}^T}{h_1} \\ \frac{\mathbf{h}^T}{h_2} \\ \bullet \\ \frac{\mathbf{h}^T}{h_n} \end{pmatrix} \tag{5}$$

where each matrix element h_j / h_i is the ratio of a bin value h_j to another bin value h_i. These bin ratios are usually stable for histograms describing the same object.

The *i*-th row $\mathbf{h}^T / h_i$ in the ratio matrix represents the ratios of all the bin values to the value of the *i*-th bin. A squared distance $d(\mathbf{p}, \mathbf{q})$ between two ℓ_2 normalized histograms **p** and **q** is defined as:

$$d(\mathbf{p}, \mathbf{q}) = \sum_{i=1}^{n} \left\| \frac{\mathbf{q}}{q_i} - \frac{\mathbf{p}}{p_i} \right\|_2^2 = \sum_{i=1}^{n} \sum_{j=1}^{n} \left(\frac{q_j}{q_i} - \frac{p_j}{p_i} \right)^2 = \|P - Q\|_2^2 \tag{6}$$

where *P* and *Q* are the ratio matrices for **p** and **q**, respectively. The distance between two histograms is thus computed as the squared ℓ_2 norm of the differences between their ratio matrices.

The distance shown in (6) is unstable when p_i or q_i are zero or close to zero: very small changes in the value of p_i or q_i can produce large differences in the distance. To avoid this problem, we propose to introduce a normalization term $1/q_i + 1/p_i$ into (6). On dividing by this normalization term, the influence of the denominators p_i and q_i in (6) is reduced. Thus, the bin ratio-based squared distance of the *i*-th row between **p** and **q** is defined by:

$$d_{BRD,i}(\mathbf{p}, \mathbf{q}) = \left\| \frac{\frac{\mathbf{q}}{q_i} - \frac{\mathbf{p}}{p_i}}{\frac{1}{q_i} + \frac{1}{p_i}} \right\|_2^2 = \sum_{j=1}^{n} \left(\frac{\frac{q_j}{q_i} - \frac{p_j}{p_i}}{\frac{1}{q_i} + \frac{1}{p_i}} \right)^2 = \sum_{j=1}^{n} \left(\frac{p_i q_j - p_j q_i}{p_i + q_i} \right)^2. \tag{7}$$

Using this normalization, dividing by p_i or q_i is replaced with multiplying by p_i or p_i. The numerator $p_i q_j - p_j q_i$ still represents ratio difference, and the denominator $p_i + q_i$ is similar to the normalization term in the χ^2 distance. Using (7), we define the squared bin ratio-based distance (BRD) $d_{BRD}(\mathbf{p}, \mathbf{q})$ between histograms **p** and **q** as:

$$d_{BRD}(\mathbf{p}, \mathbf{q}) = \sum_{i=1}^{n} d_{BRD,i}(\mathbf{p}, \mathbf{q}) = \sum_{i=1}^{n} \sum_{j=1}^{n} \left(\frac{p_i q_j - p_j q_i}{p_i + q_i} \right)^2. \tag{8}$$

In contrast to the ℓ_1 and ℓ_2 distances between *n*-dimensional vectors, the BRD is defined using $n \times n$ ratio matrices of vectors. It thus contains more information than the ℓ_1 and ℓ_2 distances [2, 6, 18]. The assumption in the BRD is that the ratio relations between histogram bins are overall kept for images in the same category. The BRD criterion is effective for dealing with the noise (deformation or perturbation) which does not completely destroy bin ratio relations.

The calculation of $d_{BRD}(\mathbf{p},\mathbf{q})$, as given by (8), has quadratic computational complexity $O(n^2)$. As shown in Annex 1, the BRD $d_{BRD}(\mathbf{p},\mathbf{q})$ can be reformulated as:

$$d_{BRD}(\mathbf{p},\mathbf{q}) = n - \|\mathbf{p}+\mathbf{q}\|_2^2 \sum_{i=1}^{n} \frac{p_i q_i}{(p_i+q_i)^2}. \tag{9}$$

Using (9), the BRD is calculated in a linear time complexity $O(n)$, because both the terms $\|\mathbf{p}+\mathbf{q}\|_2^2$ and

$$\sum_{i=1}^{n} \frac{p_i q_i}{(p_i+q_i)^2} \tag{10}$$

on the right hand side of (9) have linear complexity, and the combination of these two items takes only a constant time. It is noted that the ℓ_1 normalization can be used for the BRD. But if so, the corresponding computational complexity is $O(n^2)$. It is noted that (10) is a reweighted correlation measure between two histograms. It is interesting that it contains the terms $p_i q_i$ which are also included in the Battacharrya distance and it contains the terms $p_i + q_i$ which are included in the χ^2 distance as normalization terms.

While the above BRD is robust to partial matching and histogram normalization, it can be unstable if there are noisy bins with small values. When one of p_i or q_i is zero and the other is small (usually corresponding to small noise) $d_{BRD,i}(\mathbf{p},\mathbf{q}) = 1$. When both p_i and q_i are zero, $d_{BRD,i}(\mathbf{p},\mathbf{q})$ is undefined in that it corresponds to "0/0". These effects show that the BRD is sensitive to small noise. By contrast, many classical histogram measures handle small noise effectively, as they are essentially based on differences between bins. Therefore, we combine the BRD with other histogram distance measures, in order to improve its stability to small noise.

3. The ℓ_1 BRD and χ^2 BRD

In contrast with the multiple-kernel methods [10, 15] which use multiple histogram distances, we explore different combination between histogram distance measures. We first combine the BRD with the widely used ℓ_1 distance, which is known to be robust to outliers and small noise but sensitive to partial matching. Two common combination rules are sum and product. We choose the product as the combination rule, because the sensitivity of the BRD to small noise arises from the denominator, and multiplying a term to the BRD may reduce the effect of the denominator in the BRD. The ℓ_1 BRD of the *i*-th row between histograms **p** and **q** is defined as:

$$d_{\ell_1-BRD,i}(\mathbf{p},\mathbf{q}) = |p_i - q_i| d_{BRD,i} = |p_i - q_i| - \frac{|p_i - q_i| p_i q_i}{(p_i + q_i)^2} \|\mathbf{p}+\mathbf{q}\|_2^2 \tag{11}$$

where Equation (E) in Annex 1 is substituted into (11). The ℓ_1 BRD between **p** and **q** is defined as:

$$d_{\ell_1-BRD}(\mathbf{p},\mathbf{q}) = \sum_{i=1}^{n} d_{\ell_1-BRD,i}(\mathbf{p},\mathbf{q}) = \|\mathbf{p}-\mathbf{q}\|_1 - \|\mathbf{p}+\mathbf{q}\|_2^2 \sum_{i=1}^{n} \frac{|p_i - q_i| p_i q_i}{(p_i + q_i)^2}. \tag{12}$$

It is seen from (12) that when bin value p_i or q_i is zero, the product of the BRD and the ℓ_1 distance reduces to the ℓ_1 distance which is robust to small noise. This ensures that the ℓ_1 BRD which is naturally better suited to partial matching than the ℓ_1 distance is more robust than the original BRD in the presence of small noise.

Similarly, the χ^2 distance is combined with the BRD to generate the χ^2 BRD. The χ^2 BRD of the *i*-th row between **p** and **q** is defined as:

$$d_{\chi^2-BRD,i}(\mathbf{p},\mathbf{q}) = 2\frac{(p_i - q_i)^2}{p_i + q_i} d_{BRD,i}(\mathbf{p},\mathbf{q}). \tag{13}$$

The χ^2 BRD between **p** and **q** is then given by:

$$d_{\chi^2-BRD}(\mathbf{p},\mathbf{q}) = \sum_{i=1}^{n} 2\frac{(p_i - q_i)^2}{p_i + q_i} d_{BRD,i}(\mathbf{p},\mathbf{q}) = d_{\chi^2}(\mathbf{p},\mathbf{q}) - 2\|\mathbf{p}+\mathbf{q}\|_2^2 \sum_{i=1}^{n} \frac{(p_i - q_i)^2 p_i q_i}{(p_i + q_i)^3} \tag{14}$$

where

$$d_{\chi^2}(\mathbf{p},\mathbf{q}) = 2\sum\nolimits_{i=1}^{n} \frac{(p_i - q_i)^2}{p_i + q_i} \tag{15}$$

is the χ^2 distance between **p** and **q**.

It is noted that the ℓ_1 BRD and the χ^2 BRD still have linear computational complexity $O(n)$. This makes them suitable for large scale tasks.

4. Robustness to Partial Matching

In the following, we address the evaluation of the robustness of histogram distances to partial matching using synthetic data.

We use three histograms $\mathbf{h}^A$, $\mathbf{h}^B$, and $\mathbf{h}^C$:

- The first histogram $\mathbf{h}^A$ is obtained from an ideal object model. For example, for object recognition, $\mathbf{h}^A$ is the histogram of visual words generated from an image containing nothing but the foreground object, i.e., $\mathbf{h}^A$ is not affected by background clutter and occlusion.
- The second histogram $\mathbf{h}^B$ is obtained from an image that contains both the foreground object and background clutter. In addition, parts of the object may be occluded.
- The third $\mathbf{h}^C$ is a reference histogram which contains the least information for the application. It can be the average histogram of the samples. Ideally, the average histogram is just the uniform histogram.

Given a histogram distance measure $d(.,.)$, we check whether $d(\mathbf{h}^A,\mathbf{h}^B) < d(\mathbf{h}^A,\mathbf{h}^C)$ for each sample. Using experimental datasets, the probability that $d(\mathbf{h}^A,\mathbf{h}^B) < d(\mathbf{h}^A,\mathbf{h}^C)$ is computed. This probability is used to estimate the robustness of distance $d(.,.)$. In the method, comparison for each sample is carried out first, and then the statistics of the comparison results are calculated and used to describe the robustness of a histogram distance.

We simulate the following different cases of partial matching to test the robustness of histogram distances:

- We randomly generate the histogram of the foreground object and the histogram of the background to simulate pure background clutter without any occlusions, i.e., homogeneous background noise.
- The foreground and background histograms are generated to simulate random histogram contamination.
- A new image is synthesized by combining a foreground object image and a background image according to **a** random occlusion relation such that partial matching occurs between the histogram of the foreground object image and the histogram of the synthetic image.

These cases cover synthetic histogram data and the histograms data extracted from synthesized images. The histogram data are useful for exploring the properties of histogram distances. In the first case, we supply a simplified theoretical analysis and a partially theoretical motivation to support the claim that the BRDs are robust to partial matching. The use of simulated data in the second case is useful because a very large number of samples are produced to test the robustness of histogram distances. The third case in which real images are used to simulate partial matching can partially reveal the properties of histogram distances obtained from real images, although the set of the real images only cover partially the histogram space.

In the following, we consider first synthetic histograms which are generated in the first and second cases, and then the histograms of synthetic background images.

4.1. Synthetic histograms

We evaluate the robustness of the histogram distances in the context of background clutter, but without occlusion. The following three simple four-bin histograms are defined:

$$\begin{cases} \mathbf{h}^A = (1 \quad u \quad v \quad 0) \\ \mathbf{h}^B = (1 \quad u \quad v \quad e) \\ \mathbf{h}^C = (1 \quad 1 \quad 1 \quad 1) \end{cases} \tag{16}$$

where the parameters *u*, *v*, and *e* satisfy $1 \le u \le v$ and $0 < e$. In (16), the values of the first three bins in $\mathbf{h}^A$ are kept in $\mathbf{h}^B$. The fourth bin simulates the effects of background noise.

Let $S = d(\mathbf{h}^A,\mathbf{h}^C) - d(\mathbf{h}^A,\mathbf{h}^B)$. Our aim is to find out how *S* changes when background noise e increases. The ℓ_1 histogram normalization is used for the ℓ_1 distance and the χ^2 distance. The ℓ_2 normalization is used for the BRD, the ℓ_1 BRD, and the χ^2 BRD. Annex 2 gives the explicit formulae for the considered distances.

From the derivations in Annex 2, it is seen that for each fixed pair (u,v), $d(\mathbf{h}^A,\mathbf{h}^B)$ increases, whatever the

chosen distance *d*, when background noise *e* increases, and $d\left(\mathbf{h}^A,\mathbf{h}^C\right)$ is independent of *e*. It follows that $S=d\left(\mathbf{h}^A,\mathbf{h}^C\right)-d\left(\mathbf{h}^A,\mathbf{h}^B\right)$ is strictly monotonically decreasing when *e* increases. This indicates that the more the background noise, the less accurate the match to the foreground object.

Let e_D be the root of the equation $S=0$ in which *e* acts as a variable. Because S is strictly monotonically decreasing, when $e<e_D$, $S>0$, i.e. $d\left(\mathbf{h}^A,\mathbf{h}^C\right)>d\left(\mathbf{h}^A,\mathbf{h}^B\right)$, and when $e>e_D$, $S<0$ i.e. $d\left(\mathbf{h}^A,\mathbf{h}^C\right)<d\left(\mathbf{h}^A,\mathbf{h}^B\right)$. Therefore, we can use the value of e_D to estimate the robustness of the histogram distance to partial matching. Let e_{d_1} and e_{d_2} be, respectively, the roots of $S=0$ for the distance d_1 and the distance d_2. If $e_{d_1}>e_{d_2}$, d_1 is more robust to partial matching than d_2.

We calculated the values of e_D for different histogram distances with different pairs (u,v), where $1\le u\le v\le 100$. Table 1 shows the probability of $e_{d_1}>e_{d_2}$ for distances d_1 and d_2 among different values of *u* and *v*. We compared the proposed BRDs with the ℓ_1 distance and the χ^2 distance. For the one dimensional histograms used in the experiments, the histogram intersection is equivalent to the ℓ_1 distance [34]. Similarly, when there is not any prior information on the cost matrix of the EMD, the EMD is also equivalent to the ℓ_1 distance. Therefore, the observations in the experiments can be generalized to the EMD and the histogram intersection. Among all the 5050 pairs of (u,v) excluding (1,1), the values of e_D for the BRD are always larger than those for other distances. This means that the BRD is more robust to homogeneous background noise than other distances. This is because the BRDs embed bin correlation information which remains stable against background clutter.

Table 1. The results of comparison between different histogram distances for simulated homogeneous background noise: in each row one of the distances is compared with other distances

	$>\ell_1$	$>\chi^2$	>BRD	$>\ell_1$ BRD	$>\chi^2$ BRD
ℓ_1	N/A	61.07%	0	2.24%	3.74%
χ^2	38.93%	N/A	0	0.34%	2.30%
BRD	99.98%	99.98%	N/A	99.98%	99.98%
ℓ_1 BRD	97.76%	99.66%	0	N/A	27.29%
χ^2 BRD	96.26%	97.70%	0	72.71%	N/A

In real applications, the background clutter may include occlusion, and the background noise is often very complex, influencing the values of a number of bins. So, we explore partial matching when the histograms are randomly corrupted.

We assume that the background corrupts the foreground object randomly, e.g., with occlusion. Let the vector $\mathbf{h}_{back}=\omega(h_1^b \quad h_2^b \quad \cdots \quad h_i^b \quad \cdots \quad h_n^b)$ be the histogram of the background, where *n* is the number of bins and ω is a parameter controlling the influence of background information. We define the histogram $\mathbf{h}^A$ of the foreground object, the histogram $\mathbf{h}^B$ of the image, and the reference histogram $\mathbf{h}^C$ as follows:

$$\begin{cases} \mathbf{h}^A = \left(h_1^A \quad h_2^A \quad \cdots \quad h_i^A \quad \cdots \quad h_n^A\right) \\ \mathbf{h}^B = \mathbf{h}^A + \mathbf{h}_{back} \\ \mathbf{h}^C = \left(1 \quad 1 \quad \cdots \quad 1 \quad \cdots \quad 1\right) \end{cases} \tag{17}$$

Each bin value in $\mathbf{h}^A$ and $\mathbf{h}_{back}$ is randomly sampled from a uniform distribution over [0, 1]. Because in real applications only a part of the histogram bins are strongly perturbed by background data, we randomly set some bin values in $\mathbf{h}_{back}$ to zero. Let $r \in [0, 1]$ be the fraction of the bins influenced by background. When r=1, the background contains all types of visual words and influences every bin in $\mathbf{h}^A$. Given ω and r, each pair $\mathbf{h}^A$ and $\mathbf{h}_{back}$ are randomly generated, and it is checked whether the inequality $d(\mathbf{h}^A, \mathbf{h}^B) < d(\mathbf{h}^A, \mathbf{h}^C)$ holds for distance d. This process is repeated for a number of times, and the probability that the inequality holds is calculated and used as a measure of the robustness of d. Many probabilities can be recorded when ω and r change.

In the experiments, the number of bins was set to 100. Given a value of ω and a value of r, 10000 samples were used to check whether $d(\mathbf{h}^A, \mathbf{h}^B) < d(\mathbf{h}^A, \mathbf{h}^C)$ holds, and then the probability of $d(\mathbf{h}^A, \mathbf{h}^B) < d(\mathbf{h}^A, \mathbf{h}^C)$ was obtained. Fig. 2 shows the results for the ℓ_1 distance, the χ^2 distance, the Bhattacharyya distance, the Jeffrey divergence, the EMD, the BRD, the ℓ_1 BRD, and the χ^2 BRD. The figure reveals the following useful points:

- When ω=1, almost every histogram distance d yields a probability of 1 for $d(\mathbf{h}^A, \mathbf{h}^B) < d(\mathbf{h}^A, \mathbf{h}^C)$. This means that when the background noise is small, all histogram distances give a correct classification.
- When ω increases above 1, the performance for all the histogram distances falls rapidly. This shows that large background noise has a strong negative effect on the accuracy of classification.
- When r=1, the probability that $d(\mathbf{h}^A, \mathbf{h}^B) < d(\mathbf{h}^A, \mathbf{h}^C)$ is a maximum for each distance d. When r=0.4 or 0.6, the probability that $d(\mathbf{h}^A, \mathbf{h}^B) < d(\mathbf{h}^A, \mathbf{h}^C)$ is low. This is mainly due to the histogram normalization. When r=1, all the bins of $\mathbf{h}^B$ have larger values than the corresponding bins in $\mathbf{h}^A$. Then after normalization, the distance between $\mathbf{h}^A$ and $\mathbf{h}^B$ decreases. When r=0.4, 40% of the bins in $\mathbf{h}^B$ are larger than the corresponding bins in $\mathbf{h}^A$ and the rest have the same values as the corresponding bins in $\mathbf{h}^A$. After normalization, the 60% unchanged bins in $\mathbf{h}^B$ are decreased and the 40% increased bins may be increased or decreased. The result is that the distance between $\mathbf{h}^A$ and $\mathbf{h}^B$ is relatively large and matching the histograms becomes less accurate. This situation is the most common in real applications, because the background usually does not contain all the visual words of the foreground object.
- The ℓ_1 BRD and the χ^2 BRD are significantly more accurate than all other distances. This result is different from the results for the homogeneous background. This is because when the background corrupts the foreground object randomly, the sensitiveness of the BRD to small noise is exposed clearly. These results demonstrate the effectiveness of the bin level combination between the BRD and the ℓ_1 and χ^2

distances.

- The different histogram distances have their own characteristics. For example, the BRD is robust to homogenous background noise and the χ^2 BRD is robust to random background noise. No distance measure can outperform all its competitors in all cases.

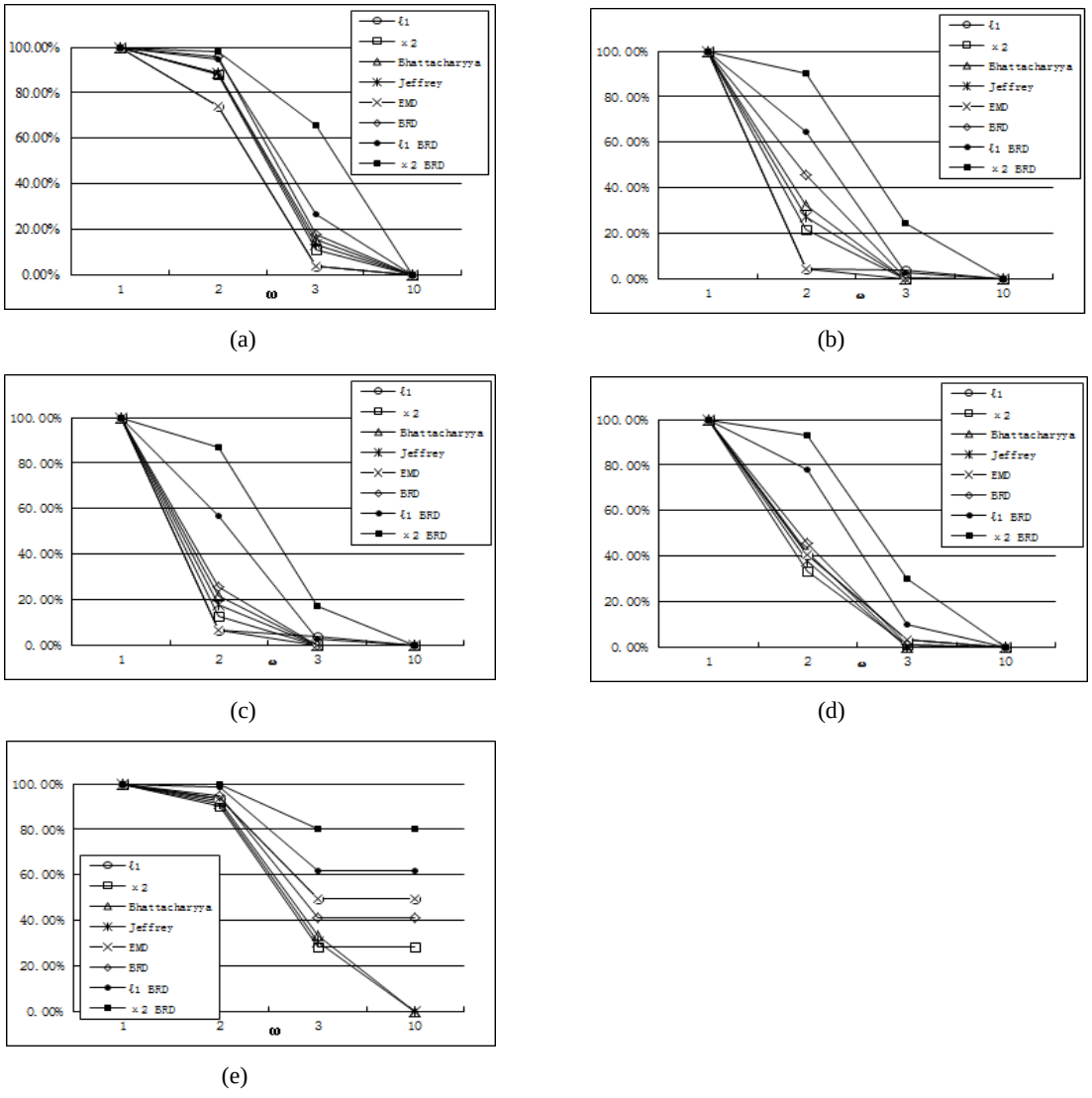


Fig. 2. The probabilities that $d(\mathbf{h}^A,\mathbf{h}^B) < d(\mathbf{h}^A,\mathbf{h}^C)$ for each histogram distance *d*, given random histogram corruption for a range of r and ω: (a) *r*=0.2; (b) *r*=0.4; (c) *r*=0.6; (d) *r*=0.8; (e) *r*=1. The x-coordinate indicates the values of ω, and the y-coordinate indicates the probabilities expressed as percentages.

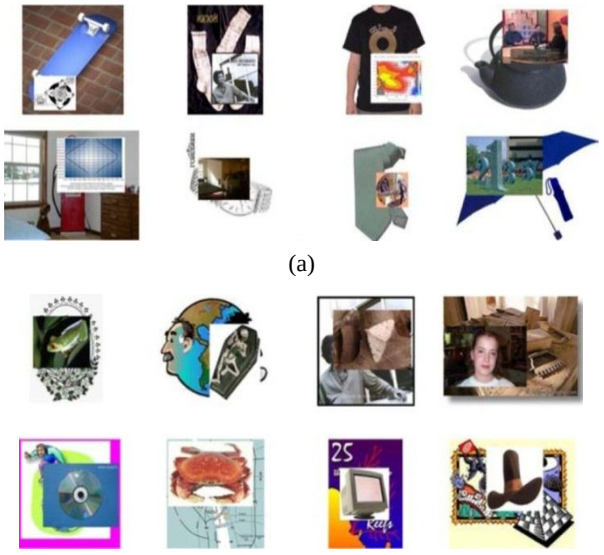
4.2. Synthetic background images

From a real image dataset which consists of object images and background images, object images and foreground images are selected and combined to produce synthetic images in the following two ways:

- The background image is placed onto the foreground image such that the foreground object is partly occluded by the background image.
- The foreground object image is placed randomly onto the background image, and then the background image is regarded as clutter.

The histogram $\mathbf{h}^A$ of each foreground image and the histogram $\mathbf{h}^B$ of each synthetic image are constructed. The reference histogram $\mathbf{h}^C$ is set to the average histogram of all the foreground images. We calculate the probability that $d(\mathbf{h}^A,\mathbf{h}^B) < d(\mathbf{h}^A,\mathbf{h}^C)$ for each distance function d using a large number of synthetic images. This probability is used to measure the robustness of d.

In the simulation, we selected 320 object images with 16 categories from the Caltech256 dataset [12], where each category consists of 20 images and 467 background images. Each synthetic image was obtained by randomly combining a foreground image and a background image from the dataset. Half of the synthetic images were obtained by placing the background image onto the foreground image, as shown in Fig. 3 (a). The image size ratio γ of the background image to the foreground object image was randomly chosen from [0.1, 0.3]. The foreground object image was fixed, and the background image was resized according to γ and randomly placed onto the foreground image. The limited range for γ ensures that there are no large occlusions. Half of the synthetic images were obtained by randomly placing the foreground object image onto the background image, as shown in Fig. 3 (b). The ratio γ was randomly chosen from [1.5, 4] to avoid too large clutter.



(a)

(b)

Fig. 3. Examples of synthetic images: (a) the foreground image is occluded by the background image; (b) the foreground image is placed randomly onto the background image.

For each foreground image, we repeated the above synthetic process 100 times to construct 100 synthetic images. The classic bag-of-words model was used to map each image into a histogram. Scale invariant feature transform (SIFT) features were extracted from images. The k-means method was employed to cluster the feature vectors of the images into 200 clusters where each cluster corresponded to a visual word. For each image, the word the closest to each feature component was found and the frequency of each word was counted to form the word histogram. Then, the histogram $\mathbf{h}^A$ of each foreground image, the histogram $\mathbf{h}^B$ of each synthetic image, and the reference histogram $\mathbf{h}^C$ were constructed. The probability that $d(\mathbf{h}^A,\mathbf{h}^B) < d(\mathbf{h}^A,\mathbf{h}^C)$ for each distance function d

was calculated using the synthetic images. The results are shown in Fig. 4. It is seen that the χ^2 BRD has the most accurate results, and the ℓ_1 BRD has accuracy **close** to the χ^2 BRD. The BRDs yield much more accurate results than the ℓ_1 distance, the χ^2 distance, the Bhattacharyya distance, the Jeffrey divergence, and the EMD whose cost matrix was calculated using the ℓ_2 distances of codebook cluster centers.

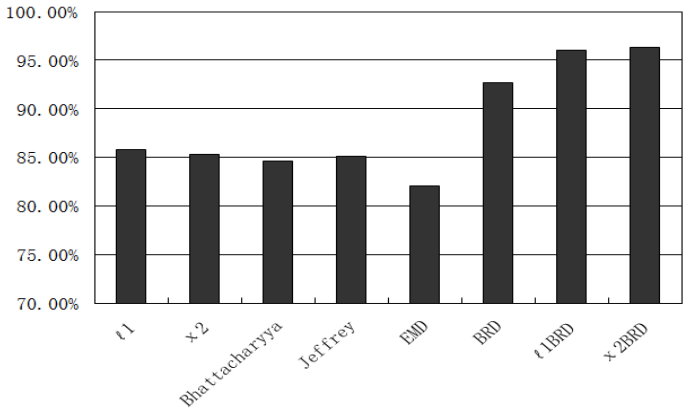


Fig. 4. The probabilities of $d(\mathbf{h}^A,\mathbf{h}^B) < d(\mathbf{h}^A,\mathbf{h}^C)$ for each distance function *d* with a synthetic background experiment, where the x-coordinate indicates different histograms and the y-coordinate indicates the probabilities.

In the above three simulations, as reported in Sections 4.1 and 4.2, the robustness of the different histogram distances to partial matching was thoroughly assessed. It is concluded that the BRDs are more robust to partial matching than the traditional five distances functions: the ℓ_1 distance, the χ^2 distance, the Bhattacharyya distance, the Jeffrey divergence, and the EMD.

5. Kernel-Based Image Classification

To use bin ratio-based histogram distances (BRDs) for image classification [4, 7, 28, 44, 45], we combine BRDs with the standard bag-of-words model. We follow the kernel-based framework in [43], i.e. we build the kernels of the BRDs using the extended Gaussian kernels [5]:

$$K(\mathbf{p},\mathbf{q}) = \exp\left(-\frac{1}{A}d(\mathbf{p},\mathbf{q})\right) \tag{19}$$

where $d(\mathbf{p}, \mathbf{q})$ is a squared distance between $\mathbf{p}$ and $\mathbf{q}$ which are histograms of two images, and A is a scaling parameter that can be determined by cross-validation. In [43], it is shown that when A is the mean of all the distances between samples, the ℓ_1 distance, the χ^2 distance, and the EMD empirically perform most accurately. It is shown in our experiments that the BRD, the ℓ_1 BRD, and the χ^2 BRD perform empirically most accurately when A is set to twice the mean of all the distances between samples. Currently, it is not known if BRDs-based kernels are Mercer kernels. Nevertheless, in our experiments, these kernels have always produced positive definite Gram matrices. It is noted that some widely used kernels, e.g. the EMD-kernel, are also not known to be Mercer kernels [43]. Some non-Mercer kernels also work effectively in real applications [5].

We use logistic regression to fuse different histogram distances in a simple manner, and evaluate the associated

classification results. Let $\vec{x} = (s_1, s_2, \cdots, s_I)$, where s_i (i=1,2,…,I) is the probability output of the i-th classifier and I is the number of classifiers. The logistic regression for information fusion is represented by:

$$f(\vec{x}) = \frac{1}{1+e^{-(\vec{w}*\vec{x})}} \tag{20}$$

where parameter vector w is estimated using the training samples. The label for each test sample is determined by the output of the logistic regression. We use logistic regression to fuse the results of the ℓ_1 BRD, the χ^2 distance, the Bhattacharyya distance, and the Jeffrey divergence for image classification.

We used the following seven benchmark datasets to test the performance of the BRDs and the logistic regression-based histogram fusion for image and scene classification: the Scene-15 dataset [17], PASCAL VOC 2008 [7], PASCAL VOC 2005 [8], PASCAL VOC 2011, 17 Oxford Flowers [24], 102 Oxford Flowers [25], and Caltech-256. In the following, we first give an initial description of the different datasets and their setups, then provide a global synthesis of all the experiments, and finally describe the local analysis on individual datasets.

5.1. Datasets and setups

1) Scene-15 dataset: This dataset [17] is a combination of several earlier datasets [9, 17, 26]. It contains 4485 scene images from 15 categories, with 200 to 400 images per category. In [11], histogram intersection was used on this dataset and a kernel codebook technique was used in comparison with standard codebook. For fair comparison, we closely followed the experimental setup in [11]. For each image in this dataset, a SIFT descriptor was sampled on a regular grid with space of eight pixels between neighboring grids. Each SIFT feature component was calculated on a 16×16 patch. We applied the histogram intersection, the χ^2 distance, the ℓ_1 distance**,** and the proposed BRDs to the feature vectors of the images in this dataset. The dataset was randomly split into the training set and the test set. A codebook vocabulary was generated using k-means on the training set. Normal codebook (hard assign) and kernel codebook (allowing for code word uncertainty) were used respectively. For each type of codebook, the SVM was employed for the kernels. For multi-class classification, we used the one-versus-all scheme in the Libsvm. Five-fold cross-validation was applied to the training set to tune the parameters. The accuracy for classifying the test set was calculated by averaging the accuracies of each category. The classification process was repeated for 10 rounds. The average accuracies of the different distances over 10 rounds were reported.

2) PASCAL visual object classes (VOC) 2008: This dataset [7, 8, 23] consists of twenty object categories with 8465 images derived from the Internet. The backgrounds in the images are usually very complex. A single image may contain multiple objects, and thus have multiple labels. The whole dataset has 2111 training images, 2221 validation images, and 4133 test images. Category labels are only released for the training and validation images. The labels of the test images are unknown to all the users. The results on the test set must be sent to the PASCAL organizers who report the accuracy of the results. We followed the framework of the winner of PASCAL 2008, Tahir et al. [35], except that we used the ℓ_1 BRD instead of the χ^2 distance in [35]. The feature vectors [32] of the

training images were clustered using k-means to generate a vocabulary of 4000 words. The ℓ_1 BRD was used to measure histogram distances. The extended Gaussian kernel and the SRKDA method in [3] were used to classify images. The parameters were estimated on the validation set and then used on the test set.

3) PASCAL VOC 2005: We tested our method on the difficult classification test set (test 2) in the PASCAL VOC 2005 dataset [8]. This test set is similar to PASCAL VOC 2008, but it only contains the following four categories: motorbike, bicycle, car, and persons. There are 1543 images in the test set. An image may include persons and a motorbike, and thus may have multiple labels. The best score in the competition to classify the images in the test set was achieved by Zhang and Schmid [8] using the χ^2 distance with an extended Gaussian kernel. Later, Zhang et al. [43] obtained a similar result using the EMD distance. We followed Zhang and Schmid's experimental setup in [8]. The Harris-Laplace detector and the SIFT descriptor were used to extract features. We obtained 1000 visual words by clustering the training samples using k-means. For fair comparison, the standard codebook was employed instead of the kernel codebook. We used the ℓ_1 BRD instead of the χ^2 distance in [8]. As in [8], the extended Gaussian kernel and the SVM classifier [4] were used. The parameters of the SVM were determined using two-fold cross-validation on the training set.

4) PASCAL VOC 2011: We used the 2011 version of the PASCAL VOC classification dataset to make the comparison. The PASCAL VOC 2012 classification dataset is the same as that used in 2011. No new data have been added. In the dataset there are 10994 images in 20 classes. The classification results must be uploaded to the PASCAL official website to obtain the information about their accuracy. We followed the experimental setup of the winner of the PASCAL VOC 2011 classification challenge. For each image, SIFT, LBP (Local Binary Pattern), and HOG features were extracted using dense sampling and the detector of points of interest. The features were then aggregated into the holistic Bag-of-Words (BoW) image representations. Various patch features were extracted using multiple image segmentations to form the image-level BoW representations. The detection features were obtained using the deformable part model for different object classes. The resulting detection kernel was combined with the visual feature kernel by weighted summation. Lasso prediction, the SVM, and the regression classifier were combined into one classifier. Kernel regression was utilized to fuse all the confidences from these three classifiers.

5) 17 Oxford Flowers: This dataset [24] contains images from 17 flower categories. There are 80 images per category. For each category, 40 images were used for training, 20 images for validation and 20 images for testing [24]. We used the same experimental setup as for PASCAL 2008 except that we used the standard SVM instead of the SRKDA. Thirty channels of features were used and combined by averaging the histogram distances of each channel. For each feature, a kernel codebook of 4000 code words was used. We classified the images in this dataset in three independent experiments and reported the average accuracy and variance.

6) 102 Oxford Flowers: This dataset [25] contains 8189 images from 102 flower categories with 40-250 images per category. For each category, there are 10 training images and 10 validation images, and the remaining

images are the test images. We used the same experimental setup as for the above 17 Oxford flowers dataset, and this setup is also the same as that used by the winners of PASCAL 2008, Tahir et al. [35]. Specifically, the feature vectors of the training images were clustered to generate a vocabulary of 4000 words. Kernel codebook was used for vector quantization. The parameters were estimated on the validation set and further used for the test set.

7) Caltech-256: The classical benchmarked Caltech-256 dataset, which was used to evaluate the robustness of the BRDs to partial matching, was also used to evaluate image classification performance. As suggested by the builders of the dataset, for each category 30 samples were selected for training, and 25 samples were selected for testing. A dense sampling was used to generate local patches where each patch corresponds to a point of interest. Then, each image patch was further represented by the RGB-SIFT descriptor. Three different image division modes were used to represent each image: the whole image without subdivision (1x1), 4 image parts obtained by dividing the image into 4 quarters (2x2) and 3 image parts obtained by dividing the image into three horizontal bands (1x3). The lengths of the feature vectors for the three division modes are 2000, 6000, and 8000, respectively. A kernel was constructed for each division mode, and the average of the three kernels was input to the SVM-based classifier. For each image category, a vocabulary of 2000 words was generated by clustering and a binary classifier was designed. A total of 256 binary classifiers were obtained.

5.2. Global synthesis

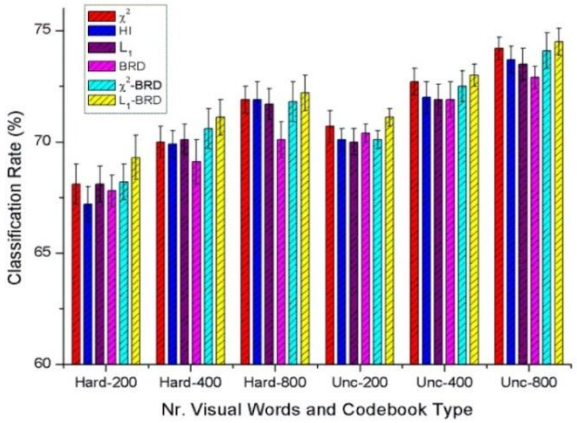


Fig. 5. Classification results for the various histogram distances on the Scene-15 dataset over different vocabulary sizes and codebook types. The terms “Hard” and “Unc” refer to hard assignment and uncertain assignment.

Fig. 5 shows the results, on the Scene-15 dataset, of the χ^2 distance, the ℓ_1 distance, the BRD, the χ^2 BRD, the ℓ_1 BRD, and the histogram intersection with the following codebook sizes: 200, 400, and 800, and with normal codebook (hard assign) or kernel codebook (code word uncertainty). Table 2 compares the classification precision of our method on the test set of the PASCAL VOC 2008 dataset with those of Tahir et al.’s method [35] and with the highest classification precisions for each image category from all the competitors of PASCAL 2008. Table 3 compares the results, on the PASCAL VOC 2005 data, of our ℓ_1 BRD and logistic regression with the results of the Bhattacharyya distance, the Jeffrey divergence and the Mahalanobis distance, and the results of the state-of-the-art

methods in [8, 21, 43], where the logistic regression fuses the ℓ_1 BRD, the χ^2 distance, the Bhattacharyya distance, and the Jeffrey divergence. Table 4 shows the results, on the PASCAL VOC 2011 dataset, of the methods based on the ℓ_1 BRD kernel, and the χ^2 distance kernel, the Bhattacharyya distance kernel, and the Jeffrey divergence kernel, and the result of the winner of the challenge. Table 5 summarizes the recognition accuracies, on the 17 Oxford flowers dataset, of the ℓ_1 BRD-based method, the χ^2 distance-based method, the Bhattacharyya distance-based method, the Jeffrey divergence-based method, and the Mahalanobis distance-based method, the top-down attention-based method [46], the excellent methods in [24, 25, 36], and the logistic regression-based method. Table 6 shows the recognition rates, on the 102 Oxford flower dataset**,** of the ℓ_1 BRD-based method, the methods based on the competing histogram distances, the logistic regression-based method, and Nilsback and Zisserman's method [25]. Table 7 shows the results, on the Caltech256 dataset, of the ℓ_1 BRD, the competing histogram distances, the logistic regression, the method in [12], the method based on the dictionary learning on single manifold (DLSM) in [47], the method based on the dictionary learning on multiple manifolds (DLMM) in [47], and the method in [11]. The experimental setups for the different histogram distances are exactly the same to avoid bias. From these tables and figure, the following global properties are revealed:

- The results of the ℓ_1 BRD are more accurate than or comparable to the state-of-the-art results in all the datasets.
- Our ℓ_1 BRD yields more accurate results than the χ^2 distance, the Bhattacharyya distance, the Jeffrey divergence, and the Mahalanobis distance.
- The logistic regression-based information fusion overall improves the classification accuracies of the individual histogram distances. So, there is room for improvement of the accuracy of the ℓ_1 BRD.
- The Mahalanobis distance yields much less accurate results. This is because the feature vectors are very sparse, and the covariance matrix is unable to describe the distribution of the features.
- The results clearly show the effectiveness of the ℓ_1 BRD.

5.3. Local analysis on individual datasets

1) Scene-15 dataset: On this dataset, the results from our re-implementation of the histogram intersection are **close** to the results in [11]. The performance of the BRD by itself is comparable to the performance of histogram intersection, although the BRD is sensitive to small noise. This indicates that bin-ratios contain rich discriminative information. The ℓ_1 BRD yields the largest average classification rates over each vocabulary size and each codebook type. This demonstrates the effectiveness of the combination of the BRD and the ℓ_1 distance. The BRD and the χ^2 BRD are robust to different types of background, but for complex backgrounds in the set of real images, the ℓ_1 BRD yields more accurate results.

2) PASCAL VOC 2008: It is shown that the ℓ_1 BRD outperforms the winner's method in 14 out of 20 categories, and it has a better average performance. Since we strictly followed the winner's method except for the histogram distance, the results on the PASCAL 2008 dataset clearly demonstrate the superiority of the proposed bin ratio-based distance. In 11 out of the 20 categories, the precisions of our method are higher than the best precisions obtained by all the competitors. In 9 out of 20 categories, the ℓ_1 BRD did not achieve the best results. This is because the best results for different image categories may be due to particular choices of features and classification strategies, and the features used in our method and the assumption of the BRD may not be most suitable for these 9 categories.

Table 2. Precisions of the winner's method, the best precisions per category among all competitors of PASCAL 2008, and the precisions of our method on the test set of PASCAL 2008.

Category	Winner [35]	Best achieved [7]	ℓ_1 BRD
Aeroplane	79.5%	**81.1%**	79.7%
Bicycle	54.3%	54.3%	**56.3%**
Bird	61.4%	**61.6%**	61.1%
Boat	64.8%	**67.8%**	66.5%
Bottle	30.0%	30.0%	**30.6%**
Bus	52.1%	52.1%	**56.5%**
Car	**59.5%**	**59.5%**	58.9%
Cat	59.4%	**59.9%**	58.1%
Chair	48.9%	48.9%	**49.4%**
Cow	33.6%	33.6%	**34.9%**
Dining table	37.8%	40.8%	**43.5%**
Dog	46.0%	**47.9%**	47.0%
Horse	66.1%	67.3%	**67.5%**
Motorbike	64%	**65.2%**	62.9%
Person	86.8%	**87.1%**	86.6%
Potted plant	29.2%	31.8%	**33.2%**
Sheep	42.3%	42.3%	**42.7%**
Sofa	44.0%	45.4%	**45.7%**
Train	**77.8%**	**77.8%**	76.2%
TV/monitor	61.2%	64.7%	**64.8%**
Mean accuracy	54.9%	N/A	**56.1%**

Table 3. Correct classification rates at equal error rates on test set 2 in the PASCAL challenge 2005

	Motor	Bike	Person	Car	Average
Winner [8]	**79.8%**	72.8%	71.9%	72%	74.1%
Winner (EMD) [43]	79.7%	68.1%	75.3%	74.1%	74.3%
Ling and Soatto [21]	76.9%	70.1%	72.5%	78.4%	74.5%
ℓ_1 BRD	79.1%	**75.4%**	73.9%	78.2%	76.7%
Bhattacharyya	75.3%	74.1%	72.6%	78.5%	75.1%
Jeffrey divergence	76.5%	73.5%	74.3%	74.2%	74.6%
Mahalanobis	41.3%	71.9%	62.2%	75.5%	62.7%
Logistic regression	77.3%	72.5%	**75.5%**	**83.2%**	**77.1%**

3) PASCAL VOC 2005: On this dataset, our method obtains the most accurate result in one of the four categories. For other categories, the results of the ℓ_1 BRD are comparable to the best results. This indicates that the

proposed ℓ_1 BRD is not sensitive to different image categories. In contrast with Ling and Soatto's method [21] in which the spatial co-occurrence statistics are considered in the feature extraction stage, the ℓ_1 BRD obtains more accurate results in the categories of Motor, Bike, and Person by more than 1.4%, but very slightly less accurate results on the category of Car with a decrease of only 0.2%. The ℓ_1 BRD improves on the average classification of Ling and Soatto's method by 2.2%. This indicates that our method is effective to consider correlations between pairs of histogram bins. It is noted that the ℓ_1 BRD does not yield the most accurate results for some image categories. One of the reasons is that if the noise completely destroys the bin ratio relations in the histograms of the images in the same category, the ℓ_1 BRD may not be accurate enough to compute the distances between histograms.

4) PASCAL VOC 2011: Although many non-trivial adjustments [48] used by the winner cannot be duplicated by us, the result of the ℓ_1 BRD-based method, which is more accurate than the results of the χ^2 distance-based method, the Bhattacharyya distance-based method, and the Jeffrey divergence-based method, is still comparable to the winning result of the PASCAL VOC 2011 challenge.

Table 4. The average precisions of different methods on the PASCAL VOC 2011 dataset

Methods	Recognition rate
ℓ_1 BRD	77.17%
χ^2 distance	76.82%
Bhattacharyya	72.50%
Jeffrey divergence	76.61%
Winner	**78.56%**

5) 17 Oxford Flowers: The ℓ_1 BRD yields a larger average recognition rate and a smaller standard deviation than the method in [25], which is in turn more accurate than the methods in [24, 36]. The result of the top-down attention-based method [46] is slightly higher than that of our BRD-based method. This is because, in the top-down attention-based method, the hue features were included in the process of producing bag of words based on the SIFT features. When the logistic regression was used to fuse different histogram distances, an average recognition rate which is larger than that of the top-down attention-based method was obtained.

Table 5. The average recognition rates of different methods on the 17 Oxford flowers dataset

Methods	Recognition rate (%)
Nilsback and Zisserman [24]	71.76±1.76
Varma and Ray [36]	82.55±0.34
Nilsback and Zisserman [25]	88.33±0.3
Top-down attention [46]	91.00
χ^2	87.45±1.13
ℓ_1 BRD	89.02±0.60
Bhattacharyya	87.05±3.47
Jeffrey divergence	87.75±3.06
Mahalanobis	24.61±1.36
Logistic regression	**91.47±2.04**

6) 102 Oxford Flowers: It is seen that our ℓ_1 BRD-based method yields a higher recognition rate than the method of Nilsback and Zisserman who produced the dataset. Logistic regression explicitly improves the classification accuracies of the individual distances.

Table 6. The average recognition rates of different methods on the 102 Oxford flower dataset.

Methods	Recognition rate
Nilsback and Zisserman [25]	72.80%
χ^2	79.68%
ℓ_1 BRD	80.45%
Bhattacharyya	79.01%
Jeffrey divergence	79.43%
Mahalanobis	17.30%
Logistic regression	**81.60%**

7) Caltech-256: The result of the ℓ_1 BRD is comparable to the recently published results. The logistic regression yields the most accurate result.

Table 7. The average recognition rates of different methods on the Caltech-256 dataset

Methods	Recognition rate
ℓ_1 BRD	45.57%
χ^2 distance	44.89%
Bhattacharyya	44.02%
Jeffrey divergence	45.46%
Mahalanobis	23.16%
Logistic regression	**46.43%**
Method in [12]	34.10%
DLSM [47]	35.12%
DLMM [47}	36.22%
Result in [11]	27.20%

6. Conclusion

In this paper, we have proposed a group of bin ratio-based histogram distances, i.e., the BRD, the ℓ_1 BRD, and the χ^2 BRD. These are new types of histogram distance, namely intra-cross-bin distances, while previous histogram distances have been bin-to-bin distances or cross-bin distances. These BRDs contain the correlations between pairs of histogram bins, while maintaining a linear computational complexity. They are robust to partial matching and histogram normalization. The ℓ_1 BRD and the χ^2 BRD can overcome the sensitiveness of the BRD to small bin values or noise. The robustness of the BRDs to partial matching is demonstrated using synthetic datasets. We have compared BRDs experimentally with several state-of-the-art histogram distance measures on seven benchmark datasets for image classification. Among these histogram distances, the ℓ_1 BRD overall generates the most accurate results in the benchmark datasets.

References

1. A. Agarwal and B. Triggs, "Hyperfeatures - Multilevel Local Coding for Visual Recognition," INRIA Research Report RR-5655, 2006.
2. A. C. Berg, T. L. Berg, and J. Malik, "Shape Matching and Object Recognition Using Low Distortion Correspondences," *in Proc. of Computer Vision and Pattern Recognition*, vol. 1, pp. 26-33, 2005.
3. D. Cai, X. He, and J. Han, "Efficient Kernel Discriminant Analysis via Spectral Regression," *In Proc. of IEEE International Conference on Data Mining*, pp. 427-432, Oct. 2007.
4. C.-K. Chiang, C.-H. Duan, S.-H. Lai, and S.-F. Chang, "Learning Component-Level Sparse Representation Using Histogram Information for Image Classification," *in Proc. of IEEE International Conference on Computer Vision*, pp. 1519-1526, 2011.
5. O. Chapelle, P. Haffner, and V. Vapnik, "Support Vector Machines for Histogram-Based Image Classification," *IEEE Trans. on Neural Networks*, vol. 10, no. 5, pp. 1055-1064, Sep. 1999.
6. O. Duchenne, F. Bach, I. Kweon, and J. Ponce, "A Tensor-Based Algorithm for High-Order Graph Matching," *in Proc. of Computer Vision and Pattern Recognition Workshops*, pp. 1980-1987, 2009.
7. V. Ablavsky and S. Sclaroff, "Learning Parameterized Histogram Kernels on the Simplex Manifold for Image and Action Classification," *in Proc. of IEEE International Conference on Computer Vision*, pp. 1473-1480, 2011.
8. M. Everingham, A. Zisserman, C. Williams, L. van Gool, M. Allan, C. Bishop, O. Chapelle, N. Dalal, T. Deselaers, and G. Dorko, "The 2005 Pascal Visual Object Classes Challenge," *Machine Learning Challenges*, Lecture Notes in Computer Science, vol. 3944, pp. 117-176, 2006.
9. L. Fei-Fei and P. Perona, "A Bayesian Hierarchical Model for Learning Natural Scene Categories," *in Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 524-531, vol. 2, June 2005.
10. P. Gehler and S. Nowozin, "On Feature Combination for Multiclass Object Classification," *in Proc. of IEEE International Conference on Computer Vision*, pp. 221-228, Sep. 2009.
11. J.C.V. Gemert, C.J. Veenman, A.W.M. Smeulders, and J.M. Geusebroek, "Visual Word Ambiguity," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 32, no. 7, pp. 1271-1283, July 2010.
12. G. Griffin, A. Holub, and P. Perona. "Caltech-256 Object Category Dataset," Technical Report: CaltechAUTHORS: CNS-TR-2007-001, California Institute of Technology, 2007.
13. K. Grauman and T. Darrell, "The Pyramid Match Kernel: Efficient Learning with Sets of Features," *Journal of Machine Learning Research*, vol. 8, no. 4, pp. 725-760, April 2007.
14. J. Hafner, H. Sawhney, W. Equitz, M. Flickner, and W. Niblack, "Efficient Color Histogram Indexing for Quadratic Form Distance Functions," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 17, no. 7, pp. 729-736, July 1995.
15. G.R.G. Lanckriet, N. Cristianini, P. Bartlett, L.E. Ghaoui, and M.I. Jordan, "Learning the Kernel Matrix with Semi-Definite Programming," *Journal of Machine Learning Research*, vol. 5, pp. 27-72, Jan. 2004
16. G. Lang and P. Seitz, "Robust Classification of Arbitrary Object Classes Based on Hierarchical Spatial Feature-Matching," *Machine Vision and Applications*, vol. 10, no. 3, pp. 123-135, 1997.
17. S. Lazebnik, C. Schmid, and J. Ponce, "Beyond Bags of Features: Spatial Pyramid Matching for Recognizing Natural Scene Categories," *in Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, vol. 2, pp. 2169-2178, 2006.
18. M. Leordeanu and M. Hebert, "A Spectral Technique for Correspondence Problems Using Pairwise Constraints," *in Proc. of IEEE International Conference on Computer Vision*, vol. 2, pp. 1482-1489, Oct. 2005.
19. J. Li, W. Wu, T. Wang, and Y. Zhang, "One Step Beyond Histograms: Image Representation Using Markov Stationary Features," *in Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1-8, June 2008.
20. H. Ling and K. Okada, "Diffusion Distance for Histogram Comparison," *in Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, vol. 1, pp. 246-253, June 2006.
21. H. Ling and S. Soatto, "Proximity Distribution Kernels for Geometric Context in Category Recognition," *in Proc. of IEEE International Conference on Computer Vision*, pp. 1-8, Oct. 2007.
22. S. Maji, A.C. Berg, and J. Malik, "Classification Using Intersection Kernel Support Vector Machines is Efficient," *in Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1-8, June 2008.
23. M. Marszalek, C. Schmid, H. Harzallah, and J.V.D Weijer, "Learning Object Representations for Visual Object Class Recognition," *in Proc. of Visual Recognition Challange workshop: PASCAL VOC 2007, in Conjunction with ICCV*, 2007.
24. M.-E. Nilsback and A. Zisserman, "A Visual Vocabulary for Flower Classification," *in Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, vol. 2, pp. 1447-1454, 2006.
25. M.-E. Nilsback and A. Zisserman, "Automated Flower Classification over a Large Number of Classes," *in Proc. of Indian Conference on Computer Vision, Graphics and Image Processing*, pp. 722-729, Feb. 2008.
26. A. Oliva and A. Torralba, "Modeling the Shape of the Scene: a Holistic Representation of the Spatial Envelope," *International Journal of Computer Vision*, vol. 42, no. 3, pp. 145-175, 2001.
27. O. Pele and M. Werman, "Fast and Robust Earth Mover's Distances," *in Proc. of IEEE International Conference on Computer Vision*, pp. 460-467, Sep. 2009.
28. F. Li, G. Lebanon, and C. Sminchisescu, "Chebyshev Approximations to the Histogram $\chi 2$ Kernel," *in Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2424-2431, 2012.

29. J. Puzicha, T. Hofmann, and J. Buhmann, "Non-Parametric Similarity Measures for Unsupervised Texture Segmentation and Image Retrieval," *in Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 267-272, June 1997.
30. Y. Rubner, C. Tomasi, and L.J. Guibas, "The Earth Mover's Distance as a Metric for Image Retrieval," *International Journal of Computer Vision*, vol. 40, no. 2, pp. 99-121, Nov. 2000.
31. S. Savarese, J. M. Winn, and A. Criminisi, "Discriminative Object Class Models of Appearance and Shape by Correlations," *in Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, vol. 2, pp. 2033-2040, 2006.
32. K.E.A.V.D. Sande, T. Gevers, and C.G.M, Snoek, "Evaluation of Color Descriptors for Object and Scene Recognition," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 32, no. 9, pp. 1582-1596, Sep. 2010.
33. C. Schmid, "Weakly Supervised Learning of Visual Models and Its Application to Content-Based Retrieval," *International Journal of Computer Vision*, vol. 56, no. 1-2, pp. 7-16, Jan.-Feb. 2004.
34. M. Swain and D. Ballard, "Color Indexing," *International Journal of Computer Vision*, vol. 7, no. 1, pp. 11-32, 1991.
35. M.A. Tahir, J. Kittler, K. Mikolajczyk, F. Yan, K.E.A.V.D. Sande, and T. Gevers, "Visual Category Recognition Using Spectral Regression and Kernel Discriminant Analysis," *in Proc. of IEEE International Workshop on Subspace*, Kyoto, Japan, pp. 178-185, Jan 2009.
36. M. Varma and D. Ray, "Learning the Discriminative Power Invariance Trade-Off," *in Proc. of IEEE International Conference on Computer Vision*, pp. 1-8, Oct. 2007.
37. N. Dalal and B. Triggs, "Histograms of Oriented Gradients for Human Detection," *in Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, vol. 1, pp. 886-893, June 2005.
38. J. Willamowski, D. Arregui, G. Csurka, C. Dance, and L. Fan, "Categorizing Nine Visual Classes Using Local Appearance Descriptors," *ICPR Workshop on Learning for Adaptable Visual Systems*, Cambridge, UK, pp. 21-31, Aug. 2004.
39. Z. Wu, Q. Ke, M. Isard, and J. Sun, "Bundling Features for Large Scale Partial-Duplicate Web Image Search," *in Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 25-32, 2009.
40. J. Wu and J.M. Rehg, "Beyond the Euclidean distance: Creating Effective Visual Codebooks Using the Histogram Intersection Kernel," *in Proc. of IEEE International Conference on Computer Vision*, pp. 630-637, Sep. 2009.
41. N. Xie, H. Ling, W. Hu, and X. Zhang, "Use Bin-Ratio Information for Category and Scene Classification," *in Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2313-2319, June 2010.
42. P. Felzenszwalb, R. Girshick, D. McAllester, and D. Ramanan, "Object Detection with Discriminatively Trained Part Based Models," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 32, no. 9, pp. 1627-1645, Sep. 2010.
43. J. Zhang, M. Marszalek, S. Lazebnik, and C. Schmid, "Local Features and Kernels for Classification of Texture and Object Categories: a Comprehensive Study," *International Journal of Computer Vision*, vol. 73, no. 2, pp. 213-238, June 2007.
44. A. Opelt, M. Fussenegger, A. Pinz, and P. Auer, "Weak Hypotheses and Boosting for Generic Object Detection and Recognition," *in Proc. of European Conference on Computer Vision*, pp. 71-84, 2004.
45. G. Csurka, C. Dance, L. Fan, J. Williamowski, and C. Bray, "Visual Categorization with Bags of Keypoints," *in Proc. of ECCV Workshop on Statistical Learning in Computer Vision*, pp. 1-22, 2004.
46. F.S. Khan, J. van de Weijer, and M. Vanrell, "Modulating Shape Features by Color Attention for Object Recognition," *International Journal of Computer Vision*, vol. 98, no. 1, pp. 49-64, 2012.
47. B.-D. Liu, Y.-X. Wang, Y.-J. Zhang, and B. Shen, "Learning Dictionary on Manifolds for Image Classification," *Pattern Recognition*, vol. 46, no. 7, pp. 1879-1890, 2013.
48. Q. Chen, Z. Song, Y. Hua, Z. Huang, and S. Yan, "Hierarchical Matching with Side Information for Image Classification," *In Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3426-3433, June 2012.
49. H.-G. Nguyen, R. Fablet, and J.-M. Boucher, "Visual Textures as Realizations of Multivariate Log- Gaussian Cox Processes," *in Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2945-2952, June 2011.
50. H.-G. Nguyen, R. Fablet, A. Ehrhold, and J.-M. Boucher, "Keypoint-Based Analysis of Sonar Images: Application to Seabed Recognition," *IEEE Trans. on Geoscience and Remote Sensing*, vol. 50, no. 4, pp. 1171-1184, April 2012.
51. H.-G. Nguyen, R. Fablet, and J.-M. Boucher, "Multivariate Log-Gaussian Cox Models of Elementary Shapes for Recognizing Natural Scene Categories," *in Proc. of IEEE International Conference on Image Processing*, Brussels Belgium, pp. 665-668, Sep. 2011.
52. H.-G. Nguyen, R. Fablet, and J.-M. Boucher, "Spatial Statistics of Visual Keypoints for Texture Recognition," *in Proc. of European Conference on Computer Vision*, Crete-Greece, pp.764-777, Sep. 2010.
53. H.-G. Nguyen, R. Fablet, and J.-M. Boucher, "Invariant Descriptors of Sonar Textures from Spatial Statistics of Local Features," *in Proc. of International Conference on Acoustics, Speech and Signal Processing*, pp. 1674-1677, March 2010.

Annex 1: Simplification of the BRD

In this annex, we reformulate the BRD to show that it can be calculated in a linear time complexity. Starting from (7) in the main text, $d_{BRD,i}$ is rewritten as:

$$
\begin{aligned}
& d_{BRD,i}(\mathbf{p},\mathbf{q}) \\
&= \frac{\sum_{j=1}^{n}\left(q_j p_i - q_i p_j\right)^2}{\left(p_i+q_i\right)^2} \\
&= \frac{\sum_{j=1}^{n}\left(q_j^2 p_i^2 + q_i^2 p_j^2 - 2q_j p_i q_i p_j\right)}{\left(p_i+q_i\right)^2} \\
&= \frac{p_i^2\sum_{j=1}^{n} q_j^2 + q_i^2\sum_{j=1}^{n} p_j^2 - 2p_i q_i \sum_{j=1}^{n} q_j p_j}{\left(p_i+q_i\right)^2}
\end{aligned}
\quad . \tag{A}
$$

According to the ℓ_2 normalization in (4) in the main text,

$$
\sum_{j=1}^{n} q_j^2 = 1 \text{ and } \sum_{j=1}^{n} p_j^2 = 1. \tag{B}
$$

Substitution of (B) into (A) yields:

$$
\begin{aligned}
& d_{BRD,i}(\mathbf{p},\mathbf{q}) \\
&= \frac{p_i^2 + q_i^2 - 2p_i q_i \sum_{j=1}^{n} p_j q_j}{\left(p_i+q_i\right)^2} \\
&= \frac{\left(p_i+q_i\right)^2 - p_i q_i\left(2 + 2\sum_{j=1}^{n} p_j q_j\right)}{\left(p_i+q_i\right)^2}
\end{aligned}
\quad . \tag{C}
$$

Substitution of

$$
2 = \sum_{j=1}^{n} q_j^2 + \sum_{j=1}^{n} p_j^2 \tag{D}
$$

into (C) yields:

$$
\begin{aligned}
& d_{BRD,i}(\mathbf{p},\mathbf{q}) \\
&= \frac{\left(p_i+q_i\right)^2 - p_i q_i \sum_{j=1}^{n}\left(p_j+q_j\right)^2}{\left(p_i+q_i\right)^2} \\
&= 1 - \frac{p_i q_i}{\left(p_i+q_i\right)^2}\left\|\mathbf{p}+\mathbf{q}\right\|_2^2
\end{aligned}
\quad . \tag{E}
$$

Using Equation (E), the BRD $d_{BRD}(\mathbf{p},\mathbf{q})$ is written as

$$
\begin{aligned}
& d_{BRD}(\mathbf{p},\mathbf{q}) \\
&= \sum_{i=1}^{n}\left(1 - \frac{p_i q_i}{\left(p_i+q_i\right)^2}\left\|\mathbf{p}+\mathbf{q}\right\|_2^2\right) \\
&= n - \left\|\mathbf{p}+\mathbf{q}\right\|_2^2 \sum_{i=1}^{n} \frac{p_i q_i}{\left(p_i+q_i\right)^2}
\end{aligned}
\quad . \tag{F}
$$

Using (F), the BRD is calculated in a linear time complexity $O(n)$.

Annex 2: Explicit representation of histogram distances in Section 4.1

For conciseness in the notation, we define $w_1 = 1+u+v$ and

$$
w_2 = \left(1+u^2+v^2\right)^{1/2}. \tag{G}
$$

For the ℓ_1 distance and the χ^2 distance, we have:

$$
d_{\ell_1}\left(\mathbf{h}^A,\mathbf{h}^B\right) = \frac{2}{1+w_1/e} \tag{H}
$$

$$d_{\ell_1}\left(\mathbf{h}^A,\mathbf{h}^C\right)=\left|\frac{1}{4}-\frac{1}{w_1}\right|+\left|\frac{1}{4}-\frac{u}{w_1}\right|+\left|\frac{1}{4}-\frac{v}{w_1}\right|+\frac{1}{4} \tag{I}$$

$$d_{\chi^2}\left(\mathbf{h}^A,\mathbf{h}^B\right)=\frac{2e}{e+2w_1} \tag{J}$$

$$d_{\chi^2}\left(\mathbf{h}^A,\mathbf{h}^C\right)=2-\frac{4}{4+w_1}-\frac{4u}{4u+w_1}-\frac{4v}{4v+w_1}. \tag{K}$$

For the BRDs, we have:

$$d_{BRD}\left(\mathbf{h}^A,\mathbf{h}^B\right)=4-\frac{6w_2}{w_2+\sqrt{w_2^2+e^2}} \tag{L}$$

$$d_{BRD}\left(\mathbf{h}^A,\mathbf{h}^C\right)=4-\left(\frac{2w_2+w_1}{2w_2^2}\right)\left(\frac{1}{\left(\frac{1}{w_2}+\frac{1}{2}\right)^2}+\frac{u}{\left(\frac{u}{w_2}+\frac{1}{2}\right)^2}+\frac{v}{\left(\frac{v}{w_2}+\frac{1}{2}\right)^2}\right) \tag{M}$$

$$d_{\ell_1-BRD}\left(\mathbf{h}^A,\mathbf{h}^B\right)=\frac{w_1}{w_2}\frac{\left(\sqrt{w_2^2+e^2}-w_2\right)^2}{\sqrt{w_2^2+e^2}\left(w_2+\sqrt{w_2^2+e^2}\right)}+\frac{e}{\sqrt{w_2^2+e^2}} \tag{N}$$

$$\begin{aligned} &d_{\ell_1-BRD}\left(\mathbf{h}^A,\mathbf{h}^C\right)\\ &=\left|\frac{1}{w_2}-\frac{1}{2}\right|\frac{(u_1)^2+(v-1)^2+1}{(w_2+2)^2}+\left|\frac{u}{w_2}-\frac{1}{2}\right|\frac{(u_1)^2+(v-1)^2+u^2}{(w_2+2u)^2}\\ &+\left|\frac{v}{w_2}-\frac{1}{2}\right|\frac{(u_1)^2+(v-1)^2+v^2}{(w_2+2v)^2}+\frac{1}{2} \end{aligned} \tag{O}$$

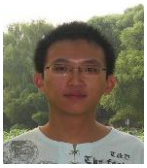
$$d_{\chi^2-BRD}\left(\mathbf{h}^A,\mathbf{h}^B\right)=2\left(\frac{w_1}{w_2}\frac{\left(\sqrt{w_2^2+e^2}-w_2\right)^3}{\sqrt{w_2^2+e^2}\left(w_2+\sqrt{w_2^2+e^2}\right)^2}+\frac{e}{\sqrt{w_2^2+e^2}}\right) \tag{P}$$

$$\begin{aligned} &d_{\chi^2-BRD}\left(\mathbf{h}^A,\mathbf{h}^C\right)\\ &=\frac{(2-w_2)^2\left((u-1)^2+(v-1)^2+1\right)}{w_2(2+w_2)^3}+\frac{(2u-w_2)^2\left((u-v)^2+(u-1)^2+u^2\right)}{w_2(2u+w_2)^3}\cdot\\ &+\frac{(2v-w_2)^2\left((u-v)^2+(v-1)^2+v^2\right)}{w_2(2v+w_2)^3}+1 \end{aligned} \tag{Q}$$

Acknowledgments

This work is partly supported by the 973 basic research program of China (Grant No. 2014CB349303), the National 863 High-Tech R&D Program of China (Grant No. 2012AA012504), and the Natural Science Foundation of Beijing (Grant No. 4121003).

Weiming Hu received the Ph.D. degree from the Department of Computer Science and Engineering, Zhejiang University in 1998. From April 1998 to March 2000, he was a postdoctoral research fellow with the Institute of Computer Science and Technology, Peking University. Now he is a professor in the Institute of Automation, Chinese Academy of Sciences. His research interests are in visual motion analysis, recognition of web objectionable information, and network intrusion detection.



Nianhua Xie received the B.E. degree in automation engineering from the Beijing Institute of Technology, Beijing, China, in 2005. In 2011, he received the Ph.D degree in the Institute of Automation, Chinese Academy of Sciences, Beijing, China. In July 2011, he joined the Beijing Sogou Technology Development Company as a business Ad

researcher. His research interests include image processing, computer vision, and machine learning.

Ruiguang Hu received both the Bachelor and Master degrees from the College of Optoelectronic Engineering at Chongqing University in 2006 and 2009. Now he is a PhD candidate in the National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing, China. His research interests include computer vision, machine learning, image classification, object recognition, saliency detection, information fusion, and Internet content security.

Haibin Ling received the BS degree and the MS degree from Peking University, China, in 1997 and 2000, respectively, and the PhD degree from the University of Maryland, College Park, in 2006. From 2006 to 2007, he worked as a postdoctoral scientist at the University of California Los Angeles. After that, he joined Siemens Corporate Research as a research scientist. Since Fall 2008, he has been an assistant professor at Temple University. His research interests include computer vision, medical image analysis, human computer interaction, and machine learning.

Qiang Chen is currently a Post-doc Research Fellow with the Electrical and Computer Engineering Department of National University of Singapore, where he received his Ph.D. degree in 2013. He received his B.E degree from the Department of Automation, University of Science and Technology of China (USTC) in 2006 and and M.S. degree from Department of Automation, Shanghai Jiaotong University in 2009. His research interests include computer vision and pattern recognition.

Shuicheng Yan is currently an Associate Professor with the Department of Electrical and Computer Engineering, National University of Singapore, Singapore, and the Founding Lead of the Learning and Vision Research Group. Dr. Yan's research areas include computer vision, multimedia, and machine learning. He has authored or coauthored over 300 technical papers over a wide range of research topics. He is an Associate Editor of the IEEE Transactions on Circuits and Systems for Video Technology and ACM Transactions on Intelligent Systems and Technology.

Stephen Maybank received a BA in Mathematics from King's college Cambridge in 1976 and a PhD in computer science from Birkbeck college, University of London in 1988. Now he is a professor in the Department of Computer Science and Information Systems, Birkbeck College. His research interests include the geometry of multiple images, camera calibration, visual surveillance.