Diffusion Distance for Histogram Comparison

Haibin Ling
Center for Automation Research,
Computer Science Department,
University of Maryland
College Park, MD, 20770, USA
hbling@umiacs.umd.edu

Kazunori Okada
Imaging and Visualization Department,
Siemens Corporate Research, Inc.
755 College Rd. E.
Princeton, NJ, 08540, USA
kazunori.okada@siemens.com

Abstract

In this paper we propose diffusion distance, a new dissimilarity measure between histogram-based descriptors. We define the difference between two histograms to be a temperature field. We then study the relationship between histogram similarity and a diffusion process, showing how diffusion handles deformation as well as quantization effects. As a result, the diffusion distance is derived as the sum of dissimilarities over scales. Being a cross-bin histogram distance, the diffusion distance is robust to deformation, lighting change and noise in histogram-based local descriptors. In addition, it enjoys linear computational complexity which significantly improves previously proposed cross-bin distances with quadratic complexity or higher. We tested the proposed approach on both shape recognition and interest point matching tasks using several multi-dimensional histogram-based descriptors including shape context, SIFT, and spin images. In all experiments, the diffusion distance performs excellently in both accuracy and efficiency in comparison with other state-of-the-art distance measures. In particular, it performs as accurately as the Earth Mover's Distance with much greater efficiency.

1. Introduction

Histogram-based local descriptors (HBLDs) are used widely in various computer vision tasks such as shape matching [1, 22, 12, 2], image retrieval [14, 15], and texture analysis [9]. HBLDs are very effective for these tasks because distributions capture rich information in local regions of objects. However, in practice, HBLDs often suffer from distortion problems due to deformation, illumination change and noise, as well as the *quantization effect* [20]. Fig. 1 demonstrates an example with shape context [1]. The deformation between (a) and (b) makes their shape context histograms significantly different.

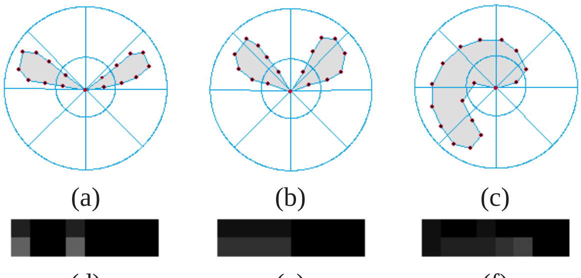


Figure 1. An example of deformation problem on shape context histograms. (a), (b) and (c) show three different shapes shown over log-polar bins. (d), (e) and (f) show the corresponding histograms of (a), (b) and (c) using the same 2D bins, respectively.

The most often used *bin-to-bin* distances between HBLDs (e.g. χ^2 statistics, L_2 distance and Kullback-Leibler divergence) assume that the histograms are already aligned, so that a bin in one histogram is only compared to the corresponding bin in the other histogram. These methods are sensitive to distortions in HBLDs as well as quantization effects. For example in Fig. 1, they falsely state that (b) is closer to (c) than to (a). *Cross-bin* distances, such as the *Earth Mover's Distance* (EMD) [20], allow bins at different locations to be (partially) matched and therefore alleviate the quantization effect. However, most of the cross-bin distances are only efficient for one-dimensional histograms (including EMD), which unfortunately limits their application to the multi-dimensional HBLDs such as shape context [1], SIFT [14], etc.

Targeting this problem, we propose a new dissimilarity distance between HBLDs, *diffusion distance*. The new approach models the difference between two histograms as a temperature field and considers the diffusion process on the field. Then, the integration of a norm on the diffusion field over time is used as a dissimilarity measure between the histograms. For computational efficiency, a Gaussian pyramid is used to discretize the continuous diffusion process.

The diffusion distance is then defined as the sum of norms over all pyramid layers. The new distance allows cross-bin comparison. This makes it robust to distortions such as deformation, lighting change and noise that often causes problems for HBLDs. Experimentally we observed that the diffusion distance performs as accurate as EMD. On the other hand, due to the exponentially decreasing layer sizes in the Gaussian pyramid, the new approach has a linear time complexity, which is much faster than previously used cross-bin distances with quadratic complexity or higher.

In summary, the diffusion distance is the main contribution of this paper. It is robust to distortion and quantization effects in comparing histogram-based local descriptors, while it is much more efficient than previously proposed cross-bin approaches. In our experiments on both shape features (shape context [1]) and image features (SIFT [14], shape context [1] and spin image [9, 7]), our method outperformed other state-of-the-art methods.

The rest of the paper is organized as follows. Sec. 2 reviews related work. Sec. 3 presents the proposed diffusion distance and discusses its relationship to EMD and previously proposed pyramid-based approaches. Sec. 4 describes experiments comparing the diffusion distance to other methods on shape matching and interest point matching tasks. Sec. 5 concludes the paper.

2. Related Work

Dis/similarity measures between histograms can be categorized into bin-to-bin and cross-bin distances. Our approach falls into the latter category. In the following, we discuss the cross-bin distances that are most related to our study.

The Earth Mover's Distance (EMD) proposed by Rubner et al. [20] defines the distance computation between distributions as a transportation problem. EMD is very effective for distributions with sparse structures, e.g., color histograms in the CIE-Lab space in [20]. However, the time complexity of EMD is larger than $O(N^3)$ where N is the number of histogram bins. This prevents its application to multi-dimensional histogram-based descriptors such as the HBLDS.

Indyk and Thaper [6] proposed a fast (approximative) EMD algorithm by embedding the EMD metric into a Euclidean space. The embedding is performed using a hierarchical distribution analysis. EMD can be approximated by measuring the L_1 distance in the Euclidean space after embedding. The time complexity of the embedding is $O(Nd\log\Delta)$, where N is the size of feature sets, d is the dimension of the feature space and Δ is the diameter of the union of the two feature sets to be compared. The embedding approach is effectively applied to retrieval tasks [6] and shape comparison [2].

Most recently, Grauman and Darrell [3] proposed using the *pyramid matching kernel* for feature set matching. In [3], a pyramid of histograms of a feature set is extracted as a description of an object. Then the similarity between two objects is defined by a weighted sum of histogram intersections [21] at each scale.

Our work differs from the above works in several ways. First, we model the similarity between histograms with a diffusion process. Second, we focus on comparing histogram-based local descriptors such as shape context [1] and SIFT [14], while the above works focus on feature distributions in the image domain. The difference between the proposed approach and the pyramid matching kernel in [3] is studied in Sec. 3.

Previously, we proposed a fast EMD algorithm, EMD-L_1 [13], for histogram comparison. EMD-L_1 utilizes the special structure of the L_1 ground distance on histograms for a fast implementation of EMD. Therefore it still solves the transportation problem, which is fundamentally different from the motivation of this paper. The diffusion distance is much faster than EMD-L_1 and performs similarly in the case of large deformations. However, in a preliminary experiment with only small quantization errors, EMD-L_1 performed better than the diffusion distance. More comprehensive comparisons between them remains as an interesting future work.

Other histogram dissimilarity measures and an evaluation can be found in [19]. In [19], the authors also describe two other cross-bin distances: early work by Peleg et al. [17] and a heuristic approach, *quadratic form* distance [16, 4].

The diffusion process has widely been used for the purpose of data smoothing and scale-space analysis in the computer vision community. Some earlier work introducing this idea can be found in [23, 8]. These works axiomatically demonstrated that a PDE model of the linear heat dissipation or diffusion process has Gaussian convolution as a unique solution. More recent well-known diffusion-based methods include anisotropic diffusion for edge-preseving data smoothing [18] and automatic scale selection with γ-normalized Laplacian [11]. It also provides a theoretical foundation to other vision techniques such as Gaussian pyramids and the SIFT feature detector [14]. Despite its ubiquitousness, to the best of our knowledge, this is the first attempt to exploit the diffusion process to compute a histogram distance.

3. The Diffusion Distance Between Histograms

3.1. Modelling Histogram Difference with a Diffusion Process

Let us first consider 1D distributions $h_1(x)$ and $h_2(x)$. It is natural to compare them by their difference, denoted as $d(x) = h_1(x) - h_2(x)$. Instead of putting a metric on d

directly, we treat it as an isolated temperature field $T(x,t)$ at time $t=0$, i.e. $T(x,0)=d(x)$. It is well known that the temperature in an isolated field obeys the heat diffusion equation

$$\frac{\partial T}{\partial t} = \frac{\partial^2 T}{\partial x^2} \tag{1}$$

It has a unique solution

$$T(x,t) = T_0(x) * \phi(x,t) \tag{2}$$

given initial condition $T_0(x)$

$$T(x,0) = T_0(x) \doteq d(x) \tag{3}$$

where $\phi(x,t)$ is the Gaussian filter

$$\phi(x,t) = \frac{1}{(2\pi)^{1/2}t} \exp\{-\frac{x^2}{2t^2}\} \tag{4}$$

Note that the mean of the difference field is zero, therefore $T(x,t)$ becomes zero everywhere when t increases. In this sense, $T(x,t)$ can be viewed as a process of histogram value exchange which makes h_1 and h_2 equivalent. Intuitively, the process *diffuses* the difference between two histograms, therefore a dissimilarity can be extracted by measuring the process. A distance between h_1 and h_2 is defined as

$$\widehat{K}(h_1,h_2) = \int_0^{\bar{t}} k(|T(x,t)|)dt \tag{5}$$

where $\bar{t}$ is a positive constant upper bound of the integration, which can be ∞ as long as the integration converges. $k(.)$ is a norm that measures how $T(x,t)$ differs from 0. In this paper, we use the L_1 norm because of its computational simplicity and good performance in our pilot studies.

Next we will show how $\widehat{K}$ handles deformation with a simple 1D example.

Assume a simple case where $h_1(x) = \delta(x)$ and $h_2(x) = \delta(x-\Delta)$, as shown in Fig. 2 (a) and (b). This means the histogram is shifted by $\Delta \geq 0$. The initial value of $T(x,t)$ is therefore $T_0 = \delta(x) - \delta(x-\Delta)$, as shown in Fig. 2 (c). The diffusion process becomes

$$\begin{aligned} T(x,t) &= (\delta(x) - \delta(x-\Delta)) * \phi(x,t) \\ &= \phi(x,t) - \phi(x-\Delta,t) \end{aligned} \tag{6}$$

Use the L_1 norm for $k(.)$,

$$\begin{aligned} k(|T(x,t)|) &= \int_{-\infty}^{\infty} |\phi(x,t) - \phi(x-\Delta,t)|dx \\ &= 2\int_{-\infty}^{\Delta/2} (|\phi(x,t) - \phi(x-\Delta,t)|)dx \\ &= 2\left(\int_{-\infty}^{\Delta/2} \phi(x,t)dx - \int_{-\infty}^{-\Delta/2} \phi(x,t)dx\right) \\ &= 2\left(2\int_{-\infty}^{\Delta/2} \phi(x,t)dx - 1\right) \end{aligned} \tag{7}$$

From (5) and (7), it is clear that $k(.)$ and $\widehat{K}$ are monotonically increasing with Δ. This suggests that $\widehat{K}$ indeed measures the degree of deformation between two histograms.

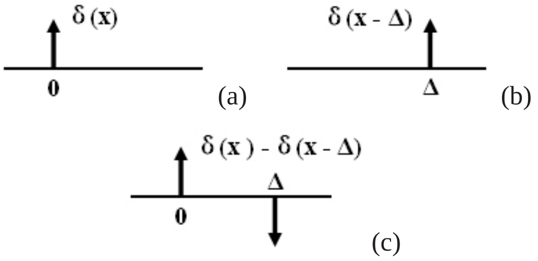


Figure 2. Two histograms with shift Δ between them and their difference. (a) h_1. (b) h_2. (c) $d = h_1 - h_2$.

3.2. Relation to the Earth Mover's Distance

From the above discussion, it is clear that $\widehat{K}$ is a cross-bin distance, which allows comparison between bins at different locations. In this subsection we will discuss its relation with EMD [20], which is another effective cross-bin histogram distance.

Given two histograms h_1 and h_2, EMD models h_1 as a set of supplies and h_2 as a set of demands. The minimum *work* to transport all supplies to demands is used as the distance between h_1 and h_2. In other word, EMD measures the dissimilarity between histograms with a transportation problem [20].

Note that bins of h_1 and h_2 share same lattice locations, which means that it takes zero work to transport supplies from a bin in h_1 to the same bin in h_2. This leads to an intuitive interpretation of EMD with the difference $d = h_1 - h_2$: EMD is the minimum work of exchanging values in d to make d vanish everywhere.

This provides an intuition about the difference between EMD and $\widehat{K}$. EMD seeks the exchanging scheme which has the minimum work, while $\widehat{K}$ measures a more "natural" exchanging scheme, i.e. diffusion process. While EMD has been successfully applied to several vision tasks (e.g. [20, 2]), the diffusion-based distances have not been evaluated with any vision tasks. Our conjecture is that they may fit to different tasks. In our experiments (see Sec. 4) on the HBLDs suffering large deformation, both approaches perform quite similarly. Below we demonstrate an example, in which $\widehat{K}$ performs better than EMD.

Consider three one-dimensional histograms h_1, h_2 and h_3 as illustrated in the left of Fig. 3. h_2 is shifted from h_1 by Δ, while h_3 can not be linearly transformed from h_1. We want to compare h_1 to h_2 and h_3. Subtracting h_2 and h_3 from h_1, we get the differences d_{12}, d_{13} as shown in the right of Fig. 3. It is clear that the EMD between h_1 and h_2 are the same as the EMD between h_1 and h_3. Perceptually, however, h_1 seems to be more similar to h_2 than to h_3.

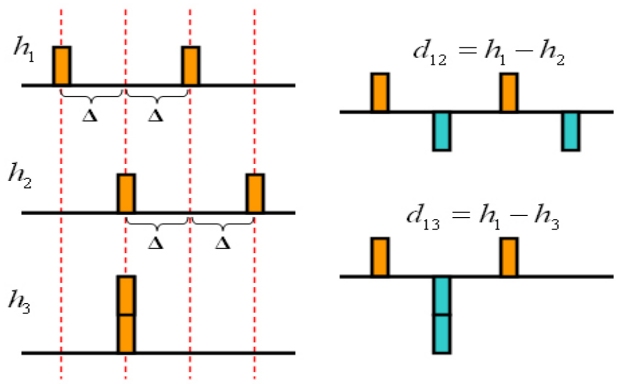


Figure 3. Left: Three 1D histograms. Right: The differences between them.

Fig. 4 shows the diffusion process $T(x,t)$ at $t = 0, 6, 12$. From the figure we see that $k(|T(x,t)|)$ for h_1 and h_2 is always smaller than that for h_1 and h_3. Therefore, $\widehat{K}(h_1,h_2) < \widehat{K}(h_1,h_3)$. This is more consistent with our perception.

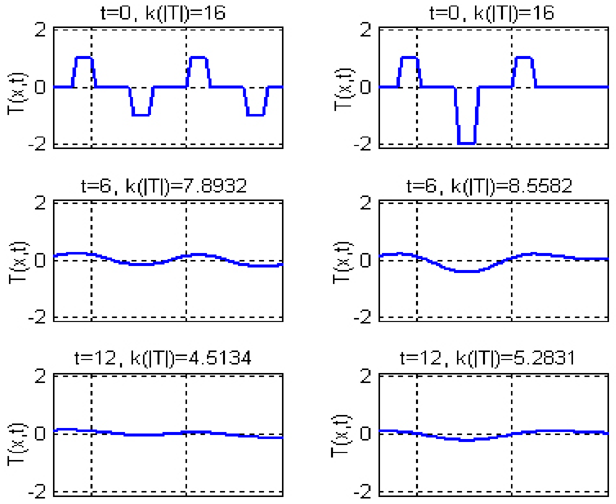


Figure 4. The diffusion process of the difference d_{12} (left column) and d_{13} (right column). Each row shows the diffusion result at a different time t. $k(|T|)$ is measured using the L_1 norm; the values show that d_{12} decays faster than d_{13}.

3.3. Diffusion Distance

It is straightforward to extend previous discussions to higher dimensions. Consider two m-dimensional histograms $h_1(\mathbf{x})$ and $h_2(\mathbf{x})$, where $\mathbf{x} \in \mathbb{R}^m$ is a vector. The definition of $\widehat{K}(h_1,h_2)$ is the same as in Sec. 3.1, except that equations (1) and (4) are replaced by (8) and (9), respectively.

$$\frac{\partial T}{\partial t} = \nabla^2 T \tag{8}$$

$$\phi(\mathbf{x},t) = \frac{1}{(2\pi)^{m/2}t}\exp\{-\frac{\mathbf{x}^\top\mathbf{x}}{2t^2}\} \tag{9}$$

Now the problem is how to compute $\widehat{K}$. Direct computation of equation (7) is expensive. Instead, we use an alternative distance function based on the Gaussian pyramid. The Gaussian pyramid is a natural and efficient discritization of the continuous diffusion process $T(\mathbf{x},t)$. It is justified because smoothing allows subsampling without aliasing. With this idea, we propose the *diffusion distance* $K(h_1,h_2)$ as

$$K(h_1,h_2) = \sum_{l=0}^{L} k(|d_l(\mathbf{x})|) \tag{10}$$

where

$$d_0(\mathbf{x}) = h_1(\mathbf{x}) - h_2(\mathbf{x}) \tag{11}$$

$$d_l(\mathbf{x}) = [d_{l-1}(\mathbf{x}) * \phi(\mathbf{x},\sigma)] \downarrow_2 \quad l = 1,...,L \tag{12}$$

are different layers of the pyramid. The notation "$\downarrow_2$" denotes half size downsampling. L is the number of pyramid layers and σ is the constant standard deviation for the Gaussian filter ϕ.

Note that as long as $k(.)$ is a metric, $K(h_1,h_2)$ forms a metric on histograms. In particular, in this paper we choose $k(.)$ as the L_1 norm, which makes the diffusion distance a true metric. Equation (10) is then simplified as

$$K(h_1,h_2) = \sum_{l=0}^{L} |d_l(x)| \tag{13}$$

The computational complexity of $K(h_1,h_2)$ is O(N), where N is the number of hitogram bins. This can be easily derived by two facts. First, the size of d_l exponentially reduces. Second, only a small Gaussian filter ϕ is required which makes the convolution take time linear in the size of d_l for each scale l.

3.4. Relation to the Pyramid Matching Kernel

The diffusion distance (13) is similar to the pyramid matching kernel (PMK) recently proposed by Grauman and Darrell [3] in that both methods compare histograms by summing the distances over all pyramid layers.

As mentioned in the related work section, our approach focuses on histogram-based local descriptors, while PMK focuses on feature set matching. The two methods have the following differences.

First, when comparing each pyramid layer, PMK counts the number of newly matched feature pairs via the difference of histogram intersection [21]. This is particularly effective for handling occlusions for feature set matching. However, this is not an effective strategy for HBLDs because they are usually normalized. In contrast, we employ the L_1 norm to compare each pyramid layer.

Second, PMK uses varying weights for different scales by emphasizing finer scales more. This is reasonable for feature set matching as mentioned in [3]. However in the diffusion distance, uniform weights are used - this seems more natural and performs better than non-uniform weights in our preliminary experiments.

Third, the diffusion distance uses Gaussian smoothing before downsampling according to the underlying diffusion process.

Fourth, PMK requires random shifting when extracting histograms from feature sets to alleviate quantization effects. The proposed method avoids such a strategy by using the intuitive cross-bin referencing imposed by the diffusion.

4. Experiments

In this section the diffusion distance is tested for two kinds of vision tasks using HBLDs. The first experiment is for shape features, where the diffusion distance is used to compare shape context [1] in a data set with articulated objects. The second experiment is for interest point matching on a data set with synthetic deformation, illumination change and heavy noise. Both experiments demonstrate that the proposed method is robust for quantization problems.

4.1. Shape Matching with Shape Context

This subsection compares the diffusion distance for shape matching with shape context (SC) [1] and the inner-distance shape context (IDSC) [12]. Shape context is a shape descriptor that captures the spatial distribution of landmark points around every interest key point [1]. IDSC is an extension of SC using the shortest path distance instead of Euclidean distance. In [12], SC and IDSC are used for contour comparison with a dynamic programming (DP) scheme. We use the same framework, except for replacing the χ^2 distance with the diffusion distance and EMD (with Rubner's code[1]) for measuring dissimilarity between (inner-distance) shape contexts.

The experiment is conducted on an articulated shape database tested in [12]. The database contains 40 images from 8 different objects. Each object has 5 images articulated to different degrees (see Figure 5). This data set is designed for testing articulation, which is a special and important case of deformation. [12] shows that the original shape context with χ^2 distance does not work well for these shapes. The reason is that the articulation causes a large deformation in the histogram.

We use exactly the same experimental setup as used in [12]: 200 points are sampled along the outer contours of every shape; 5 log-distance bins and 12 orientation bins are used for shape context histograms. The same dynamic programming matchings are used to compute distances between pairs of shapes. The recognition result is evaluated as following: For each image, the 4 most similar matches are chosen from other images in the dataset. The retrieval result is summarized as the number of 1st, 2nd, 3rd and 4th most similar matches that come from the correct object. Table 1 shows the retrieval results using the shape context. It demonstrates that the diffusion distance works much better than the χ^2 distance.

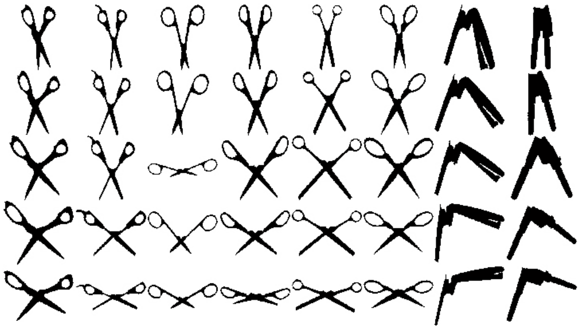


Figure 5. Articulated shape database. This dataset contains 40 images from 8 objects. Each column contains five images from the same object with different articulation.

Table 1. Retrieval result on the articulated dataset with shape context [1]. The running time (in seconds) of using χ^2 was not reported in [12].

Distance	Top 1	Top 2	Top 3	Top 4	Time
χ^2 [12]	20/40	10/40	11/40	5/40	N/A
EMD [20]	37/40	33/40	24/40	16/40	1355s
Diffu. Dist.	34/40	27/40	19/40	14/40	67s

Table 2. Retrieval result on the articulated dataset with the inner-distance shape context [12]. The running time (in seconds) of using χ^2 was not reported in [12].

Distance	Top 1	Top 2	Top 3	Top 4	Time
χ^2 [12]	40/40	34/40	35/40	27/40	N/A
EMD [20]	39/40	38/40	26/40	28/40	1143s
Diffu. Dist.	40/40	36/40	37/40	23/40	68s

Table 2 shows the results for inner-distance shape context. In this case, though the inner-distance is already insensitive to articulation, the diffusion distance still improves the result. From the tables we also see that the diffusion distance works similarly to EMD, while being more efficient.

4.2. Image Feature Matching

This subsection describes the experiment for interest point matching with several state-of-the-art image descriptors. The experiment was conducted on two image data sets. The first data set contains ten image pairs with synthetic deformation, noise and illumination change, see Fig. 6 for some examples. The second one contains six image pairs with real deformation and lighting changes, some of them

[1] http://ai.stanford.edu/∼rubner/emd/default.htm

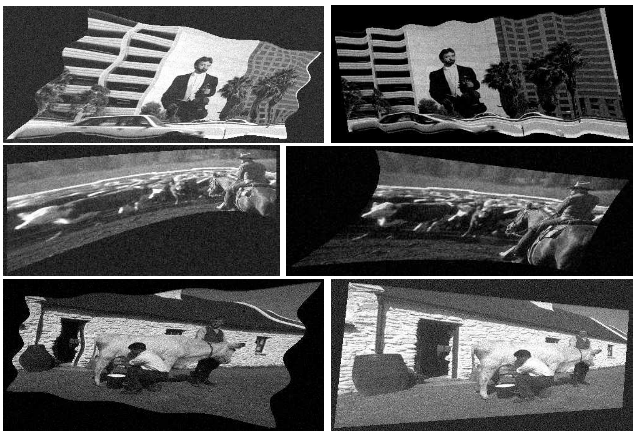


Figure 6. Some synthetic image pairs with synthetic deformation, illumination change and noise.

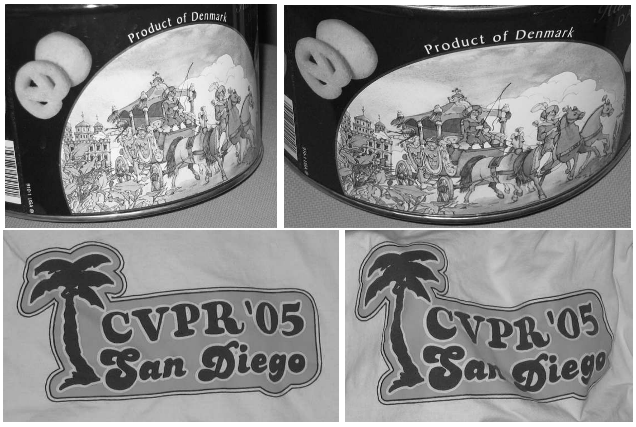


Figure 7. Some real image pairs containing deformation and lighting change. Two pairs of images with large lighting change are not shown here due to copyright issues. They are available at http://www.cs.umd.edu/˜hbling/Research/Publication/data/RD-cvpr06.zip.

are shown in Fig. 7. The experimental configuration and results are described below.

Dissimilarity measures. We tested the diffusion distance along with several popular bin-to-bin distances, as well as cross-bin distances. The bin-to-bin distances include the χ^2 statistics, the symmetric Kullback-Leibler divergence (KL), symmetric Jensen-Shannon(JS) divergence [10], L_2 distance and Bhattacharyya distance (BT). Cross-bin distances include EMD, EMD-L_1 and quadratic form(QF). For EMD, we use Rubner's online code with L_2 ground distance. The quadratic form distance is implemented according to [20]. For the diffusion distance, we set the Gaussian standard deviation $\sigma = 0.5$ and use a window of size 3×3 ($3 \times 3 \times 3$ for 3D histograms). We did not compare with PMK [3] because it requires random shifting when building a initial histogram (zero-th layer) and it uses the intersection focusing on un-normalized histograms extracted from feature sets.

Interest point. We use Harris corners [5] for the matching experiments. The reason for this choice is that, due to the large deformation, noise and lighting change, it is hard to apply other interest point detectors. On the other hand, we focus more on comparing descriptors than the interest points. For the synthetic data set, we pick 200 points per image pair with the largest cornerness responses. To compute the descriptors, a circular support region around each interest point is used. The region diameter is 41 pixels, which is similar to the setting used in [15]).

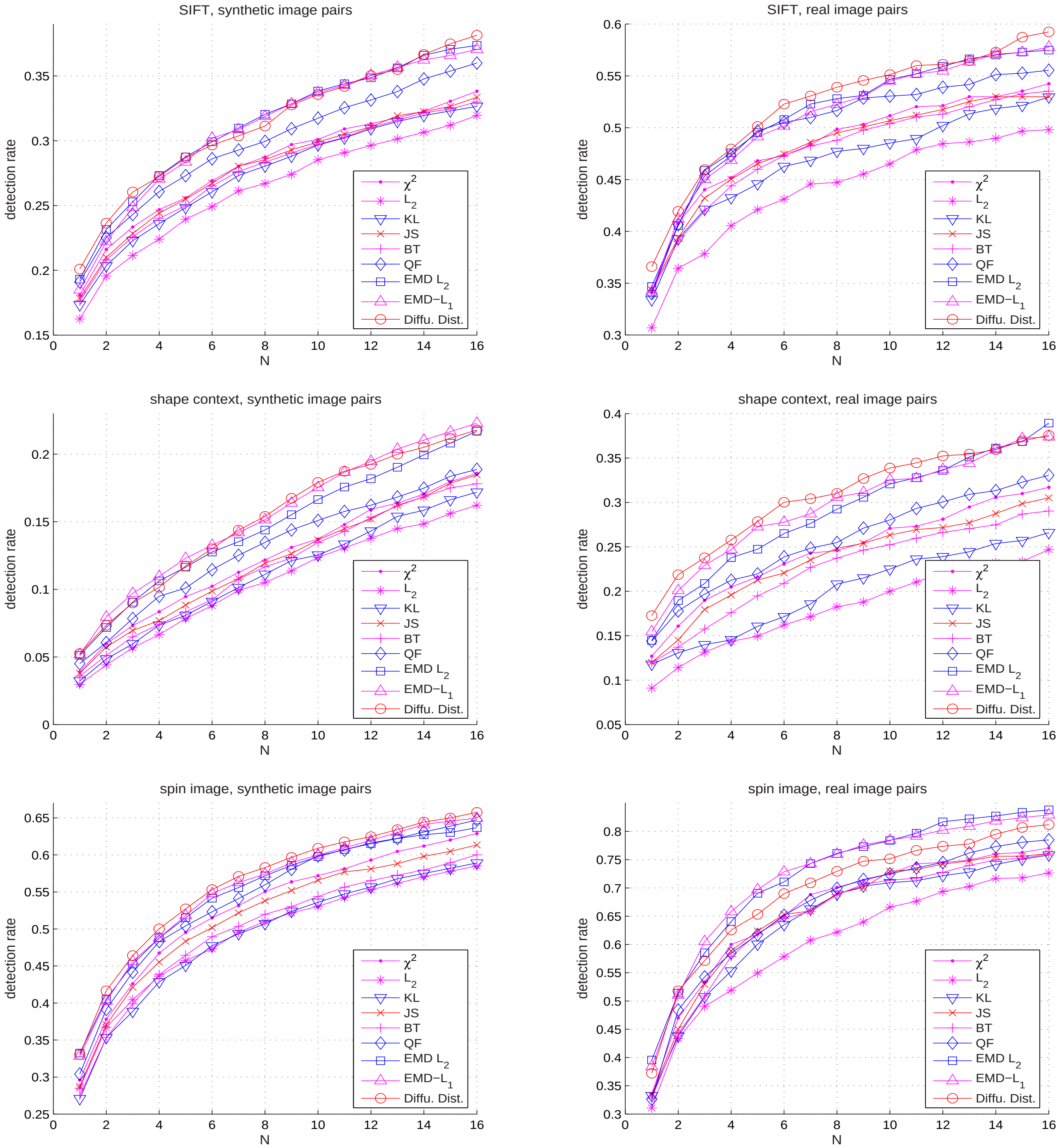
Descriptors. We tested all the distances on three different histogram-based descriptors. The first one is SIFT proposed by [14]. It is a weighted three-dimensional histogram, 4 bins for each spatial dimensions and 8 bins for gradient orientation. The second one is the shape context [1]. The shape context for images is extracted as a two-dimensional histogram counting the local edge distribution in a similar way to [15]. In our experiment, we use 8 bins for distance and 16 bins for orientation. The third one is the spin image [9, 7] which measures the joint spatial and intensity distribution of pixels around interest points. We use 8 distance bins and 16 intensity bins.

Evaluation criterion. For each pair of images with their interest points, we first find the ground-truth correspondence. This is done automatically for the synthetic data set and manually for the real image pairs. Then, for efficiency we removed those points in Image 1 with no correct matches (this also makes the maximum detection rate to 1). After that, every interest point in Image 1 is compared with all interest points in Image 2 by comparing the SIFT extracted on them. The detection rate among the top N matches is used to study the performance. The detection rate r is defined similarly to [15] as $r = \frac{\text{\# correct matches}}{\text{\# possible matches}}$.

Experiment results. A Receiver Operating Characteristic (ROC) based criterion is used to show the detection rates versus N that is the number of most similar matches allowed. The ROC curves on synthetic and real image pairs are shown in Fig. 8. In addition, the running time of each method is recorded. The average running time over real image pairs is summarized in Table 3. From these results, we see that the cross-bin distances work better than bin-to-bin distances. EMD, EMD-L_1 and the diffusion distance perform consistently better than the quadratic form distance. For efficiency, it is clear that the diffusion distance is much faster than all three other cross-bin distances - this is due to its linear computational complexity.

5. Conclusion and Future Work

We model the difference between two histograms as an isolated temperature field. Therefore the difference can


SIFT, synthetic image pairs

SIFT, real image pairs

shape context, synthetic image pairs

shape context, real image pairs

spin image, synthetic image pairs

spin image, real image pairs

Figure 8. ROC curves for interest point matching experiments. Left column is for synthetic image pairs and right for real image pairs. First row is for experiments with SIFT [14], second row for shape context [1], and third row for spin image [9, 7]

be studied with a diffusion process. Combining this idea and the connection between a diffusion process and the Gaussian pyramid, we proposed a new distance between histograms, diffusion distance. We show that the diffusion distance is robust for comparing histogram-based local descriptors since it alleviates deformation problems as well as quantization effects that often occur in real vision problems. In the experiments on both shape features and image features, the proposed approach demonstrates very promising performance in both accuracy and efficiency in comparison

Table 3. Average time (in seconds) for interest point matching between a real image pair. SC is short for shape context and SI for spin image.

Approach	SIFT [14]	SC [1]	SI [9, 7]
χ^2	0.055	0.047	0.042
L_2	0.007	0.009	0.01
KL	0.161	0.229	0.2
JS	0.317	0.284	0.299
BT	0.044	0.034	0.047
QF	3.622	3.625	3.675
EMD(L_2)	603.955	418.419	468.955
EMD-L_1	6.041	3.693	3.74
Diffu. Dist.	0.909	0.117	0.112

with other state-of-the-art histogram distances.

We are interested in deepening our understanding of how the diffusion process models the histogram difference, including further theoretical analysis of the deformation problem and the relationship between the diffusion process and other cross-bin distances, especially the Earth Mover's Distance. We are also interested in applying the proposed approach to other histogram comparison problems aside from local descriptors.

Acknowledgments

We would like to thank Visvanathan Ramesh and David Jacobs for stimulating discussions and invaluable comments. We would also thank Yossi Rubner for the transportation simplex code. The main part of the work was done when H. Ling was an intern at Siemens Corporate Research. He also thanks NSF grant (ITR-03258670325867) for supporting his PhD research.

References

[1] S. Belongie, J. Malik and J. Puzicha. "Shape Matching and Object Recognition Using Shape Context", *IEEE Trans. on PAMI*, 24(24):509-522, 2002. 1, 2, 5, 6, 7, 8

[2] K. Grauman and T. Darrell, "Fast Contour Matching Using Approximate Earth Mover's Distance", *CVPR*, I:220-227, 2004. 1, 2, 3

[3] K. Grauman and T. Darrell. "The Pyramid Match Kernel: Discriminative Classification with Sets of Image Features". *ICCV*, 2005. 2, 4, 5, 6

[4] J. Hafner, H. S. Sawhney, W. Equitz, M. Flickner, and W. Niblack. "Efficient color histogram indexing for quadratic form distance functions", *IEEE Trans. on PAMI*, 17(7):729-736, 1995. 2

[5] C. Harris and M. Stephens, "A combined corner and edge detector", *Alvey Vision Conference*, 147-151, 1988. 6

[6] P. Indyk and N. Thaper, "Fast Image Retrieval via Embeddings", *In 3rd Workshop on Statistical and computational Theories of Vision*, Nice, France, 2003 2

[7] A. Johnson, M. Hebert, "Using spin images for efficient object recognition in cluttered 3D scenes". *IEEE Trans. on PAMI*, 21(5):433-449, 1999. 2, 6, 7, 8

[8] J. J. Koenderink. "The structure of images", *Biol. Cybern.*, 50:363-370, 1984. 2

[9] S. Lazebnik, C. Schmid, and J. Ponce, "A sparse texture representation using affine-invariant regions," *IEEE Trans. PAMI*, 27(8):1265-1278, 2005. 1, 2, 6, 7, 8

[10] J. Lin. "Divergence measures based on the Shannon entropy". *IEEE Trans. Inform. Theory*, 37(1):145-151, 1991. 6

[11] T. Lindeberg. "Feature Detection with Automatic Scale Selection", *IJCV*, 30(2):79-116, 1998. 2

[12] H. Ling and D. W. Jacobs, "Using the Inner-Distance for Classification of Articulated Shapes", *CVPR*, II:719-726, 2005. 1, 5

[13] H. Ling and K. Okada. "EMD-L_1: An Efficient and Robust Algorithm for Comparing Histogram-Based Descriptors", *ECCV*, 2006, to appear. 2

[14] D. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," *IJCV*, 60(2), pp. 91-110, 2004. 1, 2, 6, 7, 8

[15] K. Mikolajczyk and C. Schmid, "A Performance Evaluation of Local Descriptors," *IEEE Trans. on PAMI*, 27(10):1615-1630, 2005. 1, 6

[16] W. Niblack, R. Barber, W. Equitz, M. Flickner, E. Glasman, D. Pektovic, P. Yanker, C. Faloutsos, and G. Taubin. "The QBIC project: querying images by content using color, texture and shape". *In Proc. of SPIE Storage and Retrieval for Image and Video Databases*, pp.173-187, 1993. 2

[17] S. Peleg, M. Werman, and H. Rom. "A Unified Approach to the Change of Resolution: Space and Gray-level", *IEEE Trans. on PAMI*, 11:739-742, 1989. 2

[18] P. Perona and J. Malik. "Scale-Space and Edge Detection Using Anisotropic Diffusion". *IEEE Trans. on PAMI*, 12(7):629-639, 1990. 2

[19] Y. Rubner, J. Puzicha, C. Tomasi, and J. M. Buhmann. "Empirical Evaluation of Dissimilarity Measures for Color and Texture", *CVIU*, 84:25-43, 2001. 2

[20] Y. Rubner, C. Tomasi, and L. J. Guibas. "The Earth Mover's Distance as a Metric for Image Retrieval", *IJCV*, 40(2):99-121, 2000. 1, 2, 3, 5, 6

[21] M. J. Swain and D. H. Ballard. "Color Indexing", *IJCV*, 7(1):11-32, 1991. 2, 4

[22] A. Thayananthan, B. Stenger, P. H. S. Torr and R. Cipolla, "Shape Context and Chamfer Matching in Cluttered Scenes", *CVPR*, I:1063-6919, 2003. 1

[23] A. P. Witkin. "Scale-space filtering", *IJCAI*, pp.1019-1022, 1983. 2