Topology-Preserved Diffusion Distance for Histogram Comparison

Wang Yan[†], Qiqi Wang[‡], Qingshan Liu[†], Hanqing Lu[†], and Songde Ma[†]
[†]National Laboratory of Pattern Recognition
Institute of Automation, Chinese Academy of Sciences
{wyan, qsliu, luhq, masd}@nlpr.ia.ac.cn
[‡]Institute of Computational and Mathematical Engineering
Stanford University
qiqi@stanford.edu

Abstract

In most previous works, histograms are simply treated as n-dimensional arrays or even reshaped into vectors when measuring the distances between them. However many histograms have their intrinsic topologies, such as HSV histogram (cone), shape context (polar), orientation histogram (circle). The topologies are important for so-called cross-bin distance, because they determine the similarities between histogram bins, and influence the cross-bin distances between histograms. In this paper, we proposed the topology-preserved diffusion distance to take the topology into account. This method extracts the distance by measuring the heat diffusion process defined on the topology of the histogram. Moreover, a fast implementation with time complexity $O(N)$ is developed. Experiments on image retrieval and interest point matching show the effectiveness and efficiency of the proposed method.

1 Introduction

Histograms are widely used in many applications of image analysis and computer vision, such as interest point matching [8, 9], shape matching [2], image retrieval [12] and texture analysis [11]. They are very effective due to the rich information captured by the distribution. However, it is well known that histogram is sensitive to the changes of illumination and viewpoints, as well as quantization effects [2], therefore the design of a robust histogram distance is a challenging task.

According to the type of bin correspondence, histogram distance is divided into two categories [12], i.e. bin-to-bin and cross-bin distance. The former just compares each bin in one histogram to the corresponding bin in the other. The Minkowski distance (such as L_1 and L_2), histogram intersection, and χ^2 statistics belong to this category. These distances are sensitive to distortions, and suffer from the quantization effect. In contrast, the cross-bin distances allow the cross-bin comparison, and therefore are more robust to distortions. Quadratic Form distance (QF) [4], Earth Mover's Distance (EMD) [12], EMD-L_1 [7], EMD-Embedding [5], Pyramid Matching Kernel (PMK) [3] and diffusion distance [6] fall into this category.

Almost all of the previous works simply treated the histogram as an n-d interval. However in practice, many histograms have their special topological structures. For example, HSV colour histogram has a cone-shaped structure, orientation histogram is a circle, and shape context is based on the polar coordinate system. The simple treatment as an interval results in great distortions of the similarities between some bins, and then degrades the accuracy of the cross-bin distance. Take 1-d orientation histogram as an example. It's often represented as an interval $[0, 2\pi)$, though it's a circle actually. Given a small positive ε, two orientations 0 and $2\pi - \varepsilon$ are almost the same. However, with the traditional representation, the two locate at two extremes of the interval, respectively. The distance between them is almost the longest, which means the smallest similarity. It contradicts with human perception. The similar contradictions also exist in HS colour histogram with the first dimension for Hue and the second for Saturation, which is usually represented as a 2-d interval $[0, 1) \times [0, 1]$. Compared to the polar representation, the distances between colours locate at different sides of the line $H = 0$ are enlarged improperly, and the same for the distances between colours with small saturations. Similar problems exist in some other histograms, such as Scale-Invariant Feature Transform (SIFT) [8] and shape context [2], when they are represented as n-d intervals.

In the paper, we proposed the topology-preserved diffusion distance for histogram matching, which is inspired by Ling and Okada's work [6]. In their work, the cross-bin relations are simulated by the heat diffusion on the n-d interval, and the distance is the integral of the diffusion process. Different from [6], the proposed method solves the diffusion process on the histogram's intrinsic topology, rather than the interval. By preserving of the topology, it's more consistent with human perception. Sophisticated numerical method for Partial Differential Equation (PDE) is used to handle the non-trivial topology. Compared to the convolution in [6], it has solid mathematical background, such as the error bound and the numerical stability. The time complexity of the distance is $O(N)$, where N is the number of bins. The experiments are conducted on image retrieval and interest point matching. The proposed distance is compared with other state-of-the-art methods, and hypothesis tests are conducted to show its superior performance.

The rest of the paper is organized as follows. Section 2 discusses the related works. Our work is described in Section 3. Experiments are reported in Section 4 and then conclusion is drawn in Section 5.

2 Related Works

In this section, we briefly review the cross-bin distances, because our method belongs to this category. For more comprehensive discussion, please refer to [11, 12].

QF [4] is an early proposed cross-bin distance. Given two histograms h_1 and h_2, the distance is defined as

$$QF(h_1, h_2) = (h_1 - h_2)^T \mathbf{A}(h_1 - h_2), \tag{1}$$

where $\mathbf{A} = [a_{ij}]$ is the weight matrix and the weights a_{ij} denote similarities between bins i and j. In the comparison of colour histograms [4], the topology is taken into account by defining

$$a_{ij} = 1 - d_{ij}/d_{\max}, \tag{2}$$

where d_{ij} is the L_2 distance between colours i and j, and $d_{\max} = \max_{i,j}(d_{ij})$. QF makes each bin in one histogram to correspond to all the bins in the other, and thus tends to

overestimate the mutual similarity without a pronounced mode [12]. Different from QF, Our method use the diffusion process to simulate the cross-bin relations, and the bin in one histogram dynamically corresponds to some neighbouring bins in the other.

EMD dynamically selects the correspondences by solving a transportation problem. Although it achieves good performances in image retrieval [12] and texture analysis [11], its computation is costly, and usually large than $O(N^3)$, where N is the number of bins. Several fast approximations have been proposed. [5] embeds the EMD metric into a Euclidean space, and the EMD can be approximated by the L_1 distance in the space after embedding. Its time complexity is $O(Nd \log \Delta)$, where N is the number of features, d is the dimension of the feature space and Δ is the diameter of the union of the two feature sets. PMK [3] is proposed for feature set matching. First, a pyramid of histograms of a feature set is extracted, and then the similarity between two feature sets is defined by a weighted sum of histogram intersections at each level of the pyramid. EMD-L_1 [7] utilizes the special structure of the L_1 ground distances on histograms for a fast implementation of EMD.

The major difference between our method and the EMD related distances above is that the topology of the histogram is not considered in the latter. EMD uses ground distances defined on the n-d interval, and the other approximate methods are all developed for this specific type of ground distance. Although EMD may handle non-trivial topology by using properly defined ground distance, it's costly to compute ($> O(N^3)$). Our method is much faster ($O(N)$). Besides the major difference, our method differs from PMK in another two ways. First, PMK focuses on feature distributions in the image domain [3], while ours focuses on comparison of histogram-based descriptors, such as SIFT. Second, PMK uses intersection to allow partial matching, which is important for handling occlusions for feature set matching. In contrast, we employ the L_1 distance, because the histograms are all normalized.

Diffusion distance [6] measures histogram distance by heat diffusion. The difference of two histograms h_1 and h_2 is treated as the initial condition of a heat diffusion process $u(\mathbf{x}, t)$, and the distance is defined as

$$K(h_1, h_2) = \int_0^T \|u(\mathbf{x}, t)\|_1 \, dt, \tag{3}$$

where T is a constant, and $\|\cdot\|_1$ represents the L_1 norm. [6] convolutes the initial condition with a Gaussian window iteratively to approximate the diffusion, and sums up the L_1 norms after each convolution to approximate the integral. The bin correspondences are implicitly determined by the diffusion. Its time complexity is $O(N)$, where N is the number of bins.

Similar to the diffusion distance, our method is also defined as the integral of the diffusion process. However, there are some significant differences. First, we define diffusion process on the histogram's intrinsic topological structure, while diffusion distance solves the process on an n-d interval. Second, we utilize numerical methods for PDE, i.e. finite volume method [1] and backward Euler scheme [10], to solve the diffusion process. In contrast, diffusion distance uses convolution to approximate the diffusion, which cannot handle the non-trivial topology.

3 Our Work

In this section, we first introduce the numerical method for heat diffusion equation, and then present the topology-preserved diffusion distance. At last, a fast implementation is described.

3.1 Numerical Method for Heat Diffusion Equation

We discretize the heat diffusion equation with Neumann boundary condition

$$\frac{\partial u(\mathbf{x},t)}{\partial t} = \nabla \cdot \nabla u(\mathbf{x},t), \quad \mathbf{x} \in \Omega, \tag{4}$$

$$\frac{\partial u(\mathbf{x},t)}{\partial \mathbf{x}} = 0, \quad \mathbf{x} \in \partial\Omega, \tag{5}$$

and then solve it numerically. The approach is briefly introduced as follows.

First, the spatial derivative $\nabla \cdot \nabla u(\mathbf{x},t)$ is discretized by finite volume method [1]. With division $\mathscr{D}$, the domain Ω is divided into N cells $\{c_k\}_{k=1}^N$, and the solution u is approximated in each cell as a constant, i.e.

$$u(\mathbf{x},t) \approx u_k(t), \quad \mathbf{x} \in c_k. \tag{6}$$

Integrating both sides of (4) over cell c_k, and using Gauss theorem and the boundary condition, we can approximate (4) and (5) with the spatial discretized equation

$$V_k \frac{\mathrm{d}u_k}{\mathrm{d}t} = \sum_{j \in \mathscr{N}_k} \alpha_{kj}(u_j - u_k), \tag{7}$$

where $\mathscr{N}_k$ is the set of neighbours of the cell c_k, and V_k and α_{kj} are constants related to the topology of domain Ω and the division $\mathscr{D}$ only.

By including the solutions of all cells, (7) can be rewritten in matrix form

$$\mathbf{M}\frac{\mathrm{d}\mathbf{u}}{\mathrm{d}t} = \mathbf{A}\mathbf{u}, \tag{8}$$

where diagonal matrix $\mathbf{M}$ and operator matrix $\mathbf{A}$ consists of $\{V_k\}_{k=1}^N$ and $\{a_{kj}\}_{k,j=1}^N$, respectively, and column vector $\mathbf{u} = [u_1, u_2, \ldots, u_N]^T$ consists of solutions in all cells.

Second, the time domain $[0,T]$ is discretized into a series of time steps $0 = t_0 < t_1 < \cdots < t_L = T$. Using the backward Euler scheme [10] to approximate the time derivative, the linear ordinary differential equation (8) becomes completely algebraic equation

$$\mathbf{M}\frac{\mathbf{u}^{(k)} - \mathbf{u}^{(k-1)}}{\Delta t_k} = \mathbf{A}\mathbf{u}^{(k)}, \quad k = 1, 2, \ldots, L, \tag{9}$$

where $\mathbf{u}^{(k)} = \mathbf{u}(t_k)$ is the solution at the k-th time step, and $\Delta t_k = t_k - t_{k-1}$. In numerical computation, we usually use fixed time step $\Delta t_k = \Delta t$. Defining matrix $\mathbf{B} = (\mathbf{M} - \Delta t \mathbf{A})^{-1}\mathbf{M}$, we can simply advance solution by

$$\mathbf{u}^{(k)} = \mathbf{B}\mathbf{u}^{(k-1)}. \tag{10}$$

Further more, we can get the solution at any time point directly by

$$\mathbf{u}^{(m)} = \mathbf{B}^m \mathbf{u}^{(0)}. \tag{11}$$

Due to the properties of the backward Euler scheme [10], our discretization (9) is stable for any positive time step Δt. The accuracies of both the spatial and temporal discretization are first-order. Therefore, the error in the numerical solution is $O(\Delta t) + O(\Delta \mathbf{x})$, where Δt is the size of the time step, and $\Delta \mathbf{x}$ is the size of the cells.

3.2 Topology-Preserved Diffusion Distance

Some notions are introduced first. A normalized histogram h is a probability density function defined on domain Ω, which is embedded in a normed space X. The topology of h is actually the topology of Ω. For example, the domain of colour histogram for Hue and Saturation is a disk embedded in the 2-d plane. The histogram $\hat{h}$ often referred in computer vision is the discrete version of h. It corresponds to a division $\mathscr{D}$, which divides Ω into cells $\{c_i\}_{i=1}^N$. The integral of h over a cell is the value of the corresponding bin in $\hat{h}$. We use "^" to represent discrete histogram and other related functions.

To compute the topology-preserved diffusion distance between two histograms, the heat diffusion equation with their difference as the initial condition is solved first. And then, the distance is extracted by integrating the L_1 norm of the process along time. Given two histograms, $h_1(\mathbf{x})$ and $h_2(\mathbf{x})$, their corresponding initial condition is

$$u(0, \mathbf{x}) = h_1(\mathbf{x}) - h_2(\mathbf{x}). \tag{12}$$

Given the solution of heat diffusion equation (4) with conditions (5) and (12), the topology-preserved diffusion distance is defined as

$$K(h_1, h_2) = \int_0^T \int_\Omega |u(\mathbf{x}, t)| \, d\mathbf{x} \, dt. \tag{13}$$

If Ω is an n-d interval and the division $\mathscr{D}$ is uniform, (13) reduces to the diffusion distance.

The method introduced in Section 3.1 is used to compare discrete histograms. Given two histograms $\hat{h}_1$ and $\hat{h}_2$, (4) and (5) are spatial discretized according to their common division $\mathscr{D}$, and the initial condition is

$$\mathbf{u}^{(0)} = \hat{h}_1 - \hat{h}_2. \tag{14}$$

We can get the discretized temperature field $\mathbf{u}(t)$ at any time t by (11). Since the integral over Ω can be approximated by L_1 norm, and the integral along time can be approximate by summation, (13) can be rewritten as

$$\hat{K}(\hat{h}_1, \hat{h}_2) = \sum_{i=0}^L \|\mathbf{u}(T_i)\|_1 \tag{15}$$

where $T_0 < T_1 < \ldots < T_L$ are time points. L is usually set to 2 or 3. The time complexity of this distance is $O(LN^2)$, where N is the number of bins. In the next section, a fast implementation is introduced, and its complexity is $O(LN)$.

A toy example is given in Figure 1 to illustrate the advantage of the proposed method. In the three Hue-Saturation histograms in Figure 1(a), only one bin in each is nonzero.
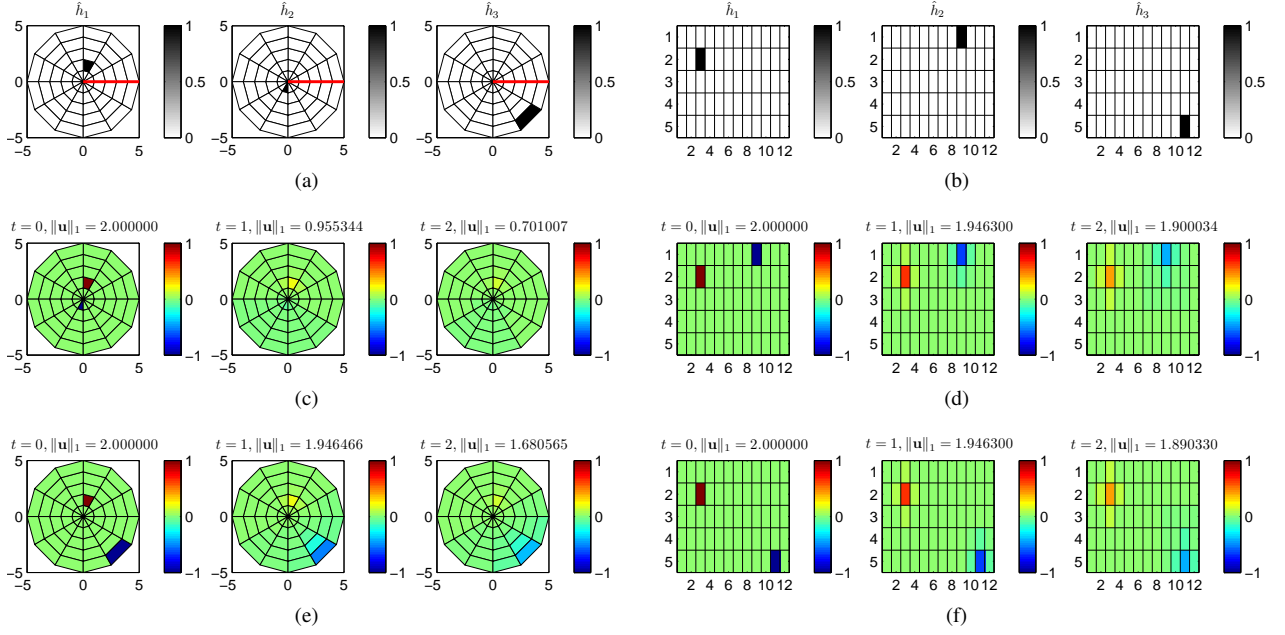
Figure 1: Toy example to show the advantage of the proposed method. (a) Histograms on disks. (b) Histograms on rectangles. (c) Diffusion process of $\hat{h}_1$ and $\hat{h}_2$ on the disk. (d) Diffusion process of $\hat{h}_1$ and $\hat{h}_2$ on the rectangle. (e) Diffusion process of $\hat{h}_1$ and $\hat{h}_3$ on the disk. (f) Diffusion process of $\hat{h}_1$ and $\hat{h}_3$ on the rectangle. Time points and L_1 norms of the temperature fields are shown above the images.

Intuitively, the similarity between $\hat{h}_1$ and $\hat{h}_2$ is larger than the one between $\hat{h}_1$ and $\hat{h}_3$, because the ground distance between the nonzero bins in the former pair is smaller. Cutting along the red line in Figure 1(a), i.e. $H = 0$, and performing some transformation, we get the common histograms in Figure 1(b). The diffusion processes on both disk and rectangle with different initial conditions are illustrated by Figure 1(c), (e), (d) and (f) respectively. The L_1 norms above the images show that the process in Figure 1(c) decays faster than the one in Figure 1(e). But there's no similar phenomenon in Figure 1(d) and (f). In fact, the L_1 norm of the last image in Figure 1(d) is even slightly larger than the corresponding one in Figure 1(f). The topology-preserved distances of Figure 1(c) and (e) are 3.6564 and 5.6270, respectively. This is consistent with the intuition. In contrast, the diffusion distances of Figure 1(d) and (f) are 3.2331 and 2.8826, respectively. Obviously, the diffusion distance fails in this case.

3.3 A Fast Implementation

Because of the linearity, the diffusion process with initial condition (12) can be viewed as the difference of two sub-processes, which use two histograms as the initial conditions respectively. The same holds in the discrete case. Plug (14) and (11) into (15), we get

$$\hat{K}(\hat{h}_1, \hat{h}_2) = \sum_{i=0}^{L} \left\| (\mathbf{B}^{m_i} \hat{h}_1) - (\mathbf{B}^{m_i} \hat{h}_2) \right\|_1, \tag{16}$$

where $m_i = \lfloor T_i/\Delta t \rfloor$. Since the division $\mathscr{D}$, the domain Ω, the time step Δt and the time points $T_0 < T_1 < \ldots < T_L$ are all predetermined, $\mathbf{B}$ can be computed in advance. Therefore both vectors, i.e. $\mathbf{B}^{m_i}\hat{h}_1$ and $\mathbf{B}^{m_i}\hat{h}_2$, can be computed at feature extraction step. The online computation only includes the differences of the vectors and the L_1 norms, and thus the online complexity is $O(LN) = O(N)$.

4 Experiments

The proposed methods are tested on natural image retrieval and interest point matching. Seven distances are compared, including L_1, L_2, χ^2, QF, EMD, Diffusion Distance (Diffusion) and Topology-Preserved Diffusion Distance (Topology). The weight matrix of QF is determined according to [4]. For the diffusion distance, we set $\sigma = 0.5$ as [6], and use 3×3 window for image retrieval and $3 \times 3 \times 3$ window for interest point matching. L_2 ground distance on the n-d interval is used in EMD. For the proposed method, we empirically choose time points $\{0,1,2\}$ for image retrieval and $\{0,2,4\}$ for interest point matching.

4.1 Natural Image Retrieval

This experiment is performed on the widely used Corel-5000 database [13], which consists of 5000 images. 8×8 HS colour histogram is used as the only feature. 1000 images (10 categories) with relatively significant colour characteristics are selected as the queries. For each query, the nearest 100 images are returned.

The average precisions of different distances are plotted in Figure 2 with respect to the scope. The time costs of different distances are shown in Table 1. EMD outperforms all the other methods, but its time cost is too high. The proposed method places the second, with much smaller time cost. L_1 and diffusion distance perform almost the same, and they are both the third. Although topology is taken into account, QF is worse than L_1, which is only a bin-to-bin distance. It confirms the analysis in Section 2, i.e. the static correspondence limits QF's performance. χ^2 and L_2 are the last.

Distance	Topology	Diffusion	L_1	χ^2	L_2	QF	EMD
Times (s)	18.0	14.1	6.3	13.4	7.2	238.4	8023.4

Table 1: Time costs in image retrieval

To further confirm the improvement, hypothesis tests are conducted. For a specific scope and a specific distance, the average precisions of 10 categories are treated as i.i.d. samples drawn from some distribution. The proposed method is compared with the others using these samples. Since the distribution is unknown, non-parametric Wilcoxon's signed rank test (one-sided) for two related samples is adopted. The p-values of the tests are listed in Table 2. Except EMD, all the others are small than 0.05, which means the improvements over the corresponding methods are all statistically significant.
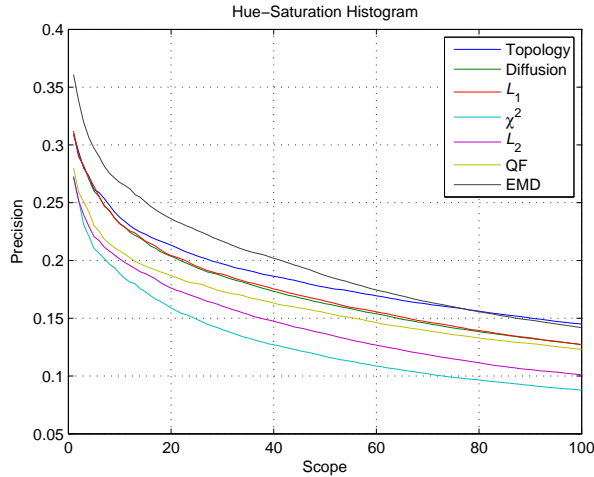
Figure 2: Retrieval precisions with respect to the scope in image retrieval

Scope	Diffusion	L_1	χ^2	L_2	QF	EMD
20	0.0469	0.0371	0.0039	0.0020	0.0020	0.5566
40	0.0020	0.0059	0.0039	0.0020	0.0020	0.6250
60	0.0039	0.0273	0.0039	0.0020	0.0137	0.7695
80	0.0098	0.0117	0.0059	0.0020	0.0420	0.6250
100	0.0039	0.0059	0.0039	0.0020	0.0322	0.6953

Table 2: p-values of hypothesis tests in image retrieval

4.2 Interest Point Matching

This experiment is performed on the Affine Covariant Regions Dataset [9], which consists of 40 image pairs with known plane projective transforms. We extract SIFT like descriptors from the interest regions detected by the Hessian-Affine detector [9]. The descriptor differs from SIFT by ignoring the tri-linear interpolation [8] and by being normalized by L_1 norm. The number of local descriptors varies from 200 to 4000 per image depending on the content.

The evaluation strategy in [9] is utilized. For each pair of images, the ground truth correspondences are first determined by the known transform. Then, we use the threshold-based strategy to match descriptors, i.e. two descriptors are matched if the distance between them is below a threshold. Varying the threshold, a Receiver Operating Characteristic (ROC) curve can be obtained. For some image pairs, it's hard to obtain the complete ROC curve with any distance because the precision keeps low. It's probably due to the limitations of the detector and/or the descriptor. For this reason, 21 image pairs are selected, and ROC curves in Figure 3 of different methods are the averages on these pairs.

Compared to image retrieval, similar ranking are shown in Figure 3. EMD is the best, followed by the topology-base diffusion distance. The diffusion distance and L_1 place the third, and then QF, L_2 and χ^2. The margin between Topology and Diffusion (or L_1) is
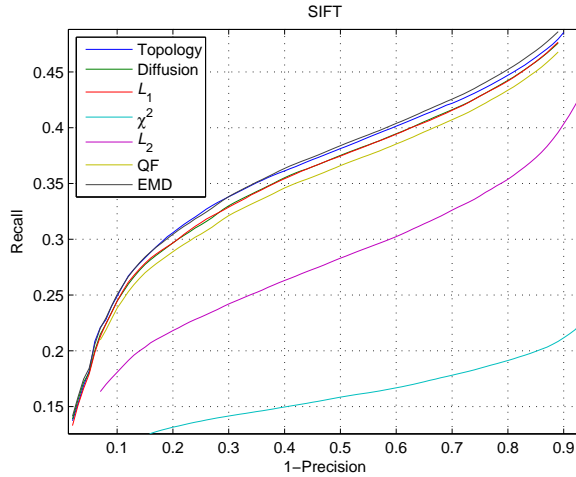
Figure 3: ROC curves in interest point matching

1−Precision	Diffusion	L_1	χ^2	L_2	QF	EMD
0.2	7.9802e-005	1.2267e-004	5.9570e-005	5.9570e-005	7.1872e-005	0.7823
0.4	0.0033	4.1887e-004	5.9570e-005	5.9570e-005	6.4356e-004	0.5829
0.6	6.1791e-004	5.4342e-004	5.9570e-005	5.9570e-005	3.5792e-005	0.8392
0.8	0.0037	4.1887e-004	5.9570e-005	5.9570e-005	5.0872e-005	0.5929

Table 3: p-values of hypothesis tests in interest point matching

roughly 1%. In spite of the superior performance, the computation of EMD costs about 300 hours. In contrast, our method uses only about 10 minutes, and the diffusion distance uses about 7 minutes.

The same hypothesis tests are conducted. For a specific precision and a specific distance, the recalls of different image pairs are treated as i.i.d. samples, on which the comparisons are based. The p-values are listed in Table 3. Again, the improvements over the other methods are significant, except EMD. Compared to Table 2, the p-values are smaller, which means the improvements are more significant in the sense of statistics, in spite of the smaller margins showed in Figure 3.

5 Conclusions

In this paper, we extend the diffusion distance by combining the idea of topology preserving. The proposed method defines the diffusion process on the topology of the histogram, and measures the distance by integrating the L_1-norm of the process along time. It outperforms most existing histogram distances by preserving the topology, and also outperforms topology-based QF by utilizing the diffusion process. Among the methods with complexities lower than $O(N^2)$, the proposed one is the most accurate. Moreover, it's also very efficient with the complexity $O(N)$.

Acknowledgement

References

[1] T. Barth and M. Ohlberger. *Finite Volume Methods: Foundation and Analysis*, volume 1 of *Encyclopedia of Computational Mechanics*, chapter 15. John Wiley & Sons, West Sussex, 2004.

[2] S. Belongie, J. Malik, and J. Puzicha. Shape matching and object recognition using shape context. *IEEE Trans. PAMI*, 24(24):509–522, 2002.

[3] K. Grauman and T. Darrell. The pyramid matching kernel: Discriminative classification with sets of image features. In *Proc. Int'l Conf. on Computer Vision*, 2005.

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

[5] P. Indyk and N. Thaper. Fast image retrieval via embeddings. In *Proc. Third Workshop on Statistical and Computational Theories of Vision*, 2003.

[6] H. Ling and K. Okada. Diffusion distance for histogram comparison. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, 2006.

[7] H. Ling and K. Okada. EMD-L_1: An efficient and robust algorithm for comparing histogram-based descriptors. In *Proc. European Conf. on Computer Vision*, 2006.

[8] D. Lowe. Distinctive image features from scale-invariant keypoints. *Int'l J. Computer Vision*, 60(2):91–110, 2004.

[9] K. Mikolajczyk and C. Schmid. A performance evaluation of local descriptors. *IEEE Trans. PAMI*, 27(10):1615–1630, 2005.

[10] P. Moin. *Fundamentals of Engineering Numerical Analysis*. Cambridge University Press, 2001.

[11] Y. Rubner, J. Puzicha, C. Tomasi, and J. M. Buhmann. Empirical evaluation of dissimilarity measures for color and texture. *Computer Vision and Image Understanding*, 84(1):25–43, 2001.

[12] Y. Rubner, C. Tomasi, and L. J. Guibas. The earth mover's distance as a metric for image retrieval. *Int'l J. Computer Vision*, 40(2):99–121, 2000.

[13] H. Tong, J. He, M. Li, C. Zhang, and W. Ma. Graph based multi-modality learning. In *Proc. ACM Multimedia*, 2005.