A COMPLEX NETWORK BASED FEATURE EXTRACTION FOR IMAGE RETRIEVAL

Jieqi Kang, Shan Lu, Weibo Gong, Patrick A. Kelly

University of Massachusetts, Amherst
Electrical and Computer Engineering Department
151 Holdsworth Way, Amherst, MA, 01003

ABSTRACT

In this paper, we propose a complex network based low-level feature for image retrieval systems based on the graph representation of the image and the mathematical theory of diffusion over manifolds. We show that the proposed image feature is invariant to non-structural changes on images and performs well in hand written digits classification task. We also show that performance of image retrieval with existing low-level features could be improved by combining with the proposed feature.

1. INTRODUCTION

Fast content based image retrieval (CBIR) has been a subject of intensive research for many years. A typical image retrieval task includes several sub-tasks such as feature extraction and classification. The critical initial step in most image retrieval systems is extraction of appropriate low-level features.

Researchers have designed a variety of image features based on intensity, color, shape and texture statistics or descriptors. Recent surveys [1, 2, 3, 4] have briefly introduced or evaluated many of the low-level features appearing in the state of the art CBIR literature [5, 6]. These features have all been derived from reasoning and/or intuition about useful information for classification. However, no low-level feature can completely capture essential structural information. Recent research [7, 8] suggest graphs as the form that organizes information in the brain. Hence, it seems reasonable to design a low-level feature that is based on a graph representation for the image.

The idea of using graphs in image processing is not entirely new [9] and there are many published examples of graph similarity testing. In [10] and [11], the authors treat every pixel as a vertex in a graph and connect the neighboring pixels according to the intensity differences. In [12, 13], researchers build graphs based on descriptors such as the edge location and segmented region. For graph similarity testing, in a recent survey [14] the authors introduce most of the existing methods. However, these methods are either not scalable or not always effective [15]. Some other methods such as vertex ranking require prior knowledge of the nodes mapping between the two compared graphs [16], which is not suitable for comparing image-generated graphs that do not have any vertex label.

In this work, we consider a feature vector based on graph heat content. Heat content is the summation of heat diffusion on the graph over time. The asymptotic behavior of the heat content has already been used as a property to understand the connectivity structure of the graph [17, 15]. In [18, 19] the author points out that Monte-Carlo simulations of diffusion can be effective in testing the similarity of complex graphs and that this may have implications on concept abstraction mechanisms in human and animal intelligence. Heat content can be estimated using a random walk on graph. In [20], the authors show that a random walk based similarity algorithm is effective in differentiating large graphs with the same heavy tail degree distribution.

Our proposed approach exhibits the following features. First, our method uses a graph representation of the image and summarizes the graph structure into a single heat content feature vector which is invariant to image rotation and small distortion. Second, the length of the feature vector can be very short since the initial behavior of heat content is the most important. Third, the heat content feature performs better in classification compared to some widely-used low-level features and can be combined with other features to further improve the retrieval performance.

The rest of the paper is organized as follows. In Section 2, we discuss the fast feature extraction for a complex network. In Section 3, we propose the general algorithm of graph generation and feature extraction. In Section 4, three experiments are shown to illustrate the performance of the heat content feature. Section 5 is the conclusion and suggestion for future work.

2. HEAT CONTENT FEATURE EXTRACTION OF COMPLEX NETWORK

2.1. Heat equation and heat content

Our method is based on a fast feature extraction algorithm for complex networks. Let $G = (V, E)$ denote a weighted graph generated by an image. V is the vertex set and $E \subseteq V \times V$ is the edge set. The adjacency matrix is denoted as $A = [a_{uv}]$

978-1-4799-5751-4/14/$31.00 ©2014 IEEE

and a_{uv} is the edge weight from u to v. We define the out-degree matrix $D = diag[d_u]$ to be $d_u = \sum_v a_{uv}$. The graph Laplacian and the normalized graph Laplacian [21] of the graph G are defined as $L = D - A$ and $\mathcal{L} = D^{-1/2}LD^{-1/2}$.

We partition all the vertices in V into two subsets iD and ∂D and $V = iD \cup \partial D$. iD is the set of all interior nodes and ∂D is the set of all boundary nodes. With the definition of boundary and interior nodes, the heat equation of the graph is associated with the normalized graph Laplacian and can be defined as

$$\begin{cases} \frac{\partial h_t}{\partial t} = -\mathcal{L}h_t \\ h_t(u,v) = 0 \text{ for } u \in \partial D, \end{cases} \tag{1}$$

with initial condition $h_0(u,u) = 1$ if $u \in iD$.

Intuitively, $h_t(u,v)$ is the amount of heat that flows from vertex u to vertex v at time t. All heat that flows to the boundary vertices is absorbed and the rest remains. The interior heat is rich in information because the graph structure affects the heat diffusion pattern, which determines the total heat. We denote the the number of interior vertices as $|iD|$ and the $|iD| \times |iD|$ heat matrix H_t is defined as $H_t(u,v) = h_t(u,v)$ for $u, v \in iD$.

For convenience, we still use $\mathcal{L}$ to denote the interior part of the original Laplacian. We denote $\Lambda = diag[\lambda_i]$ as the diagonal eigenvalue matrix and $\Phi = [\phi_1, \cdots, \phi_n]$ as the eigenvector matrix in which ϕ_i is the corresponding eigenvector of the eigenvalue λ_i. We diagonalize $\mathcal{L}$ as $\mathcal{L} = \Phi\Lambda\Phi^{-1}$, where $\Phi^{-1} = [\pi_1, \pi_2, \cdots, \pi_n]'$. The solution to the heat equation is $H_t = e^{-\mathcal{L}t} = \Phi e^{-\Lambda t}\Phi^{-1}$ and for each entry of H_t, we have $H_t(u,v) = \sum_{i=1}^{|iD|} e^{-\lambda_i t}\phi_i(u)\pi_i(v)$. The heat content $Q(t)$ is defined as

$$Q(t) = \sum H_t(u,v) = \sum_u \sum_v \sum_{i=1}^{|iD|} e^{-\lambda_i t}\phi_i(u)\pi_i(v). \tag{2}$$

A discrete-time heat content vector with a given length could be seen as a type of feature of the corresponding image. Research has already proved that small changes in the graph will not affect the Laplacian spectrum too much [22]. Thus the heat content feature should be invariant to perturbations and small distortions of the original image.

2.2. Fast heat content estimation

Directly finding the eigenvalues and the eigenvectors of the graph is very time comsuming. We have to estimate the heat content in a more affordable way. One approach to estimate the heat content is to use a matrix multiplication method. We could approximate the continuous heat content by the summation of a discrete time random walk on graph. Suppose we have the transition matrix defined as $M = D^{-1}A$ and a lazy random walk transition matrix as $M_L = (1-\delta)I + \delta M$. In a lazy random walk, a random walker either moves out with probability δ or remains at the current node with probability $1 - \delta$. For any given time $t = k\delta$, the distribution of the random walker can be calculated theoretically as $P_t = M_L^k P_0 = (I - \frac{t}{k}L_r)^k P_0$ where P_0 is the initial distribution and L_r is the random walk Laplacian [23]. Notice that when $k \to \infty$ ($\delta \to 0$ while keeping $k\delta = t$) the distribution will converge to $e^{-L_r t}P_0$. Thus we have $M_L^k \to e^{-L_r t}$. The heat content $H_t = e^{\mathcal{L}t}$ is a little different, but we can use the relationship between the random walk Laplacian and the normalized Laplacian to connect them. We have $L_r = D^{-1}L = D^{-1/2}\mathcal{L}D^{1/2}$. By using the notations in previous subsection, the random walk Laplacian can be expressed as $L_r = (D^{-1/2}\Phi)\Lambda(\Phi^{-1}D^{1/2}) = (D^{-1/2}\Phi)\Lambda(D^{-1/2}\Phi)^{-1}$. So each entry in matrix M_L^k converges to $M_L^k(u,v) \to e^{-L_r t} = \sum_{i=1}^{|iD|} e^{-\lambda_i t}\phi_i(u)\pi_i(v)\sqrt{d_u/d_v}$. We then have the following estimation of heat content by matrix multiplication

$$\hat{Q}(k) = \sum_u \sum_v M_L^k(u,v)\sqrt{\frac{d_u}{d_v}}. \tag{3}$$

We can also use a Monte Carlo method to estimate the heat content. We will discuss the details in section 3.

3. HEAT CONTENT ESTIMATION BASED IMAGE FEATURE EXTRACTION

3.1. Graph generation algorithm for grayscale image

To generate a graph from an image, our approach is to treat each pixel in the image as a node in the graph. We only consider grayscale images in this paper.

First, every pixel p_i is represented by one vertex $v_i \in V$ in our graph $G = (V, E)$. Then the weight of each edge $e = (v_i, v_j) \in E$ is defined by the following equation

$$A(i,j) = \begin{cases} \frac{\epsilon_1 + f(I_i, I_j)}{d(p_i, p_j)} & (I_i \neq I_j) \\ \frac{\epsilon_2}{d(p_i, p_j)} & (I_i = I_j) \end{cases} \tag{4}$$

in which I_i, I_j are intensities of pixels p_i and p_j. $d(p_i, p_j)$ is a distance measure of pixel p_i to p_j. A monotonically increasing function $f(I_i, I_j)$ calculates the difference between the intensity of pixel p_i and p_j. ϵ_1 and ϵ_2 are two small positive numbers. We could either connect every pixels pair or only connect all the neighboring pixels for less computational complexity. The intuition behind this graph generation approach is that human vision tends to focus on high-contrast places. Our approach emphasizes the close high-difference pixel pairs by giving the corresponding graph edges larger weights.

3.2. Random walk based heat content estimation

Suppose we have already generated a graph $G = (V, E)$ from an image I with the boundary B defined as a set of all the nodes corresponding to pixels on the natural boundary of the image. The major steps of our Monte Carlo heat content estimation algorithm are:

(1) According to the weight matrix A, compute the degree of each node D and the corresponding random walk matrix P.

(2) At the initial time, give every vertex the same number k of random walkers to simulate a uniform initial condition. The initial size of random walkers on the graph is $s = nk$.

(3) In each step, every random walker stays at its current vertex v_i with probability δ $(0 \leq \delta \leq 1)$ and with probability $1 - \delta$ goes out. The probability of going to neighbor v_j should equal to p_{ij} in the random walk matrix P. If $v_j \in B$, for which B represent all the boundary vertices, the random walker is then deleted.

(4) After every step, a heat content value is calculated by $\hat{Q}(t) = \sum_{i=1}^{s} 1 \times (\frac{d_o}{d_c})_i^{1/2}$ in which d_o and d_c are the degree of origin node v_o and current node v_c of the ith random walker.

(5) Running the algorithm for T/δ steps to get the heat content $\hat{Q}(t)$ from 0 to T.

4. THE RESULTS OF THE EXPERIMENT

Three experiments were performed to examine the performance of our heat content feature. The first experiment is an example to show the performance of the heat content feature under image variations such as rotation, stretching and affine transforms. The second experiment is a retrieval experiment for spectrograms. The purpose of this experiment is to check the robustness of the heat content feature for similar-structure but largely distorted spectrogram images due to different reading speed and background environment. In the last experiment, we apply the heat content feature in a classification task for hand written digits dataset MNIST and compare it with some widely used low-level features.

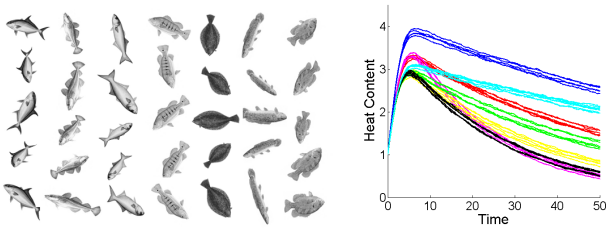
4.1. Heat content of affine transformed images

We set up a dataset containing seven types of fish images. The original fish image is from [24]. To each type of fish image, five affine transforms under random but bounded (the variations are less than 30%) coefficients were applied to generate five similar fish images for each type. Figure 1(a) shows the result of fish images after affine transform.

The simulated heat contents of all of the fish images are plotted in figure 1(b). The heat content curves of the same color corresponds to the same type of the fish images. Even from the beginning part ($t < 10$), the heat content curves can successfully cluster the same types of fishes into one group. The result illustrates that the heat content feature is robust to rotation, stretching and affine transform of images, which is a useful property for image feature extraction.

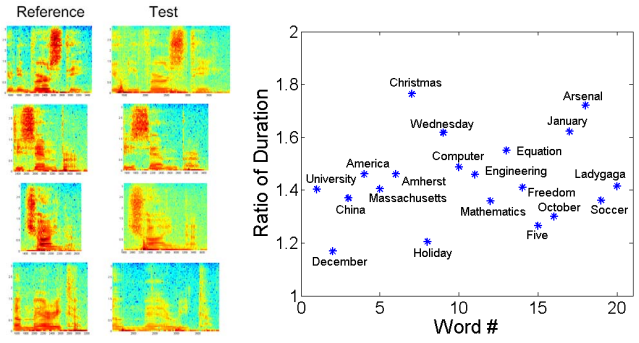
4.2. Spectrogram retrieval experiment

In this experiment, we test our feature by a spectrogram retrieval experiment. A spectrogram is an image, which is the result of a short time fourier transform (STFT) of an audio sample. We record 40 single-word ".wav" audio samples and generate all the spectrograms. The STFT window sizes and the overlapping lengths are set to be the typical settings. Then the 40 spectrograms are divided into 2 groups. The first group is the reference group in which 20 words are read at normal speed. The second group is the test group in which all the previous words are read again but at slower speed in a slightly different recording environment.



(a) Fish image examples (b) Heat content curves

Fig. 1. Affine transformed fish image examples and corresponding heat content curves. In the graph generation process, $f(I_i, I_j) = |I_i - I_j|$, d is the square of distance, $\epsilon_1 = \epsilon_2 = 1$. In the Monte Carlo estimation, $\delta = 0.1$ and the number of initial random walkers per vertex is 1.



(a) Spectrograms (b) Ratio of durations

Fig. 2. The spectrograms and duration ratios between test samples and reference samples

Figure 2(a) shows the spectrograms of several word pairs in the reference and test group. The test images are distorted from the reference images, but the major structure still remains. Figure 2(b) illustrates the exact ratio of time durations on the vertical scale between the 20 slower test samples and the 20 original reference samples. The distortion percentage is from 20% to 80%. For each spectrogram S, we convert the image to grayscale, and then divided the image into 16 blocks. A 10-steps heat content vector h is computed for each block. Figure 3 shows the Euclidean distance matrix of the combined heat content vectors between all the reference and test samples. We can see that every test sample is correctly matched to its reference.

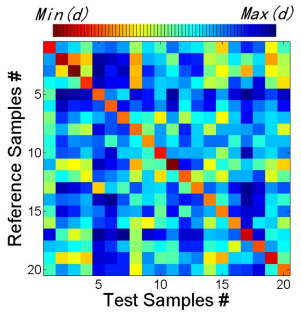


Fig. 3. Distance matrix for the test and reference spectrograms

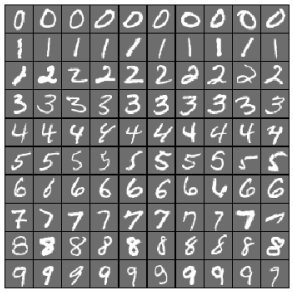


Fig. 4. Hand-written digit image examples in MNIST database

4.3. Hand written digits classification experiment

The MNIST database [25] is a benchmark widely used in image classification algorithm comparison. It contains a training group of $60,000$ images and a testing group of $10,000$ images. One property of this dataset is that all the images are hand written digits with standard size and contrast, which is very "similar" already. Figure 4 shows some example images.

We first compute the heat content feature for every image based on forty-nine 10×10 random position overlapping local image blocks. Five steps of heat content is used for each image block. Table 1 compares the classification error rates of the heat content feature and other blockwise similar-size low-level features including the intensity histogram, intensity moments, Gabor coefficients, gray-level co-occurrence matrix (GLCM) and edge directions histogram. All the features are normalized and the classification algorithm is a k-nearest-neighbors classifier with L2 norm as the distance measure. The result shows that although the heat content feature alone is not the best, it is still better than some single low-level features. More importantly, if we add the heat content feature to any other feature and form a combined feature, the error rate always drops. This result provides preliminary evidence that the heat content feature contains some useful image information which is not represented by the existing low-level image features.

Our next experiment is to simulate the real image retrieval task by combining three types of features (Intensity, texture and shape). We apply logistic and linear kernel support vector machine (SVM) [26] classifiers to execute the classification. Table 2 shows that in all situations the performance of the combined feature with the heat content is better than the original combination in both classifiers. This result further illustrates that the heat content feature contains unique and useful image information as a new type of low-level feature which has the potential to improve the feature exaction step for current image retrieval systems.

Feature	Alone	with **HC**
Intensity histogram (Histogram)	10.76%	**6.89%**
Intensity moments (Moments)	8.94%	**7.04%**
Gabor coefficients (Gabor)	3.05%	**2.94%**
Gray-level co-occurrence (GLCM)	6.92%	**4.49%**
Edge directions (Edge)	3.68%	**3.15%**
Heat content (HC)	6.54%	N/A

Table 1. Classification error rate (k-NN classifier)

Feature	Logistic	SVM
Histogram + Gabor + Edge	2.12%	1.47%
Histogram + Gabor + Edge + **HC**	**2.01%**	**1.31%**
Histogram + GLCM + Edge	2.58%	1.62%
Histogram + GLCM + Edge + **HC**	**2.28%**	**1.54%**
Moments + Gabor + Edge	2.06%	1.30%
Moments + Gabor + Edge + **HC**	**1.90%**	**1.24%**
Moments + GLCM + Edge	2.36%	1.48%
Moments + GLCM + Edge + **HC**	**2.20%**	**1.41%**
Combined feature	1.82%	1.29%
Combined feature with **HC**	**1.78%**	**1.22%**

Table 2. Classification error rate (logistic/SVM classifier)

5. CONCLUSION

Feature extraction is the first and most fundamental step of a fast image retrieval system. In this paper, we propose a graph-based image representation associated with a fast graph similarity comparison algorithm based on the asymptotic behavior of the heat content. The heat content feature is shown to be a robust, easily computed image feature and has the potential to be an effective and efficient low-level feature for image retrieval. The heat content feature can also be combined with other existing low-level features to create a complex visual signature which may improve the following classification/retrieval tasks. Although we still need further experiments and analysis to thoroughly understand the advantages and drawbacks of the heat content feature, our preliminary results show that heat content computation could be a useful component of image retrieval tasks.

In future work we will first consider how to improve the graph generation algorithm. We also plan to design a visual signature including the modified heat content feature which can handle the color information of image and to test its performance on large scale image retrieval tasks such as the COREL database.

6. ACKNOWLEDGEMENTS

This work is supported in part by the United States National Science Foundation CNS-1065133 and CNS-1239102, and Army Research Office Contract W911NF-08-1-0233 and W911NF-12-1-03.

7. REFERENCES

[1] R. Datta, D. Joshi, J. Li, and J. Z. Wang, "Image retrieval: Ideas, influences, and trends of the new age," *ACM Computing Surveys*, vol. 40(2), pp. 5:1–5:60, 2008.

[2] I. Felci Rajam and S. Valli, "A survey on content based image retrieval," *Life Science Journal*, vol. 10(2), pp. 2475–2487, 2013.

[3] T. Deselaers, D. Keysers, and H. Ney, "Features for image retrieval: An experimental comparison," *Information Retrieval*, vol. 11, pp. 77–107, 2008.

[4] P. Howarth and S. Ruger, "Evaluation of texture features for content-based image retrieval," *Image and Video Retrieval, Lecture Notes in Computer Science*, pp. 326–334, 2004.

[5] Y.G. Jiang, Q. Dai, J. Wang, C.W. Ngo, X.Y. Xue, and S.F. Chang, "Fast semantic diffusion for large-scale context-based image and video annotation," *IEEE Transactions on Image Processing*, vol. 21(6), pp. 3080–3091, 2012.

[6] L. Zhang, L. Wang, and W. Lin, "Generalized biased discriminant analysis for content-based image retrieval," *IEEE Transactions on Systems, Man, and Cybernetics-Part B: Cybernetics*, vol. 42(1), pp. 282–290, 2012.

[7] E. Bullmore and O. Sporns, "Complex brain network: graph theoretical analysis of structural and functional systems," *Nature Reviews Neuroscience*, vol. 10, pp. 186–198, 2009.

[8] M. Rubinov and O. Sporns, "Complex network measures of brain connectivity: Uses and interpretations," *NeuroImage*, vol. 52, pp. 1059–1069, 2010.

[9] I. K. Park, II D. Yun, and S. U. Lee, "Color image retrieval using hybrid graph representation," *Image and Vision Computing*, vol. 17, pp. 465–474, May 1999.

[10] P.F. Felzenszwalb, "Efficient graph-based image segmentation," *International Journal of Computer Vision*, vol. 59(2), pp. 167–181, 2004.

[11] L. Grady, "Random walks for image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28(11), pp. 1768–1783, 2006.

[12] A.R. Backes, D. Casanova, and O.M. Bruno, "A complex network-based approach for boundary shape analysis," *Pattern Recognition*, vol. 42, pp. 54–67, 2009.

[13] Z. Harchaoui and F. Bach, "Image classification with segmentation graph kernels," in *IEEE Conference on Computer Vision and Pattern Recognition*, Minneapolis, USA, June 2007, pp. 1–8.

[14] D. Koutra, A. Parikh, A. Ramdas, and J. Xiang, "Algorithms for graph similarity and subgraph matching," 2011, Available at `https://www.cs.cmu.edu/~jingx/docs/DBreport.pdf`.

[15] P. Mcdonald and R. Meyers, "Isospectral polygons, planar graphs and heat content," *Proceedings of the American Mathematical Society*, vol. 131(11), pp. 3589–3599, 2003.

[16] P. Papadimitriou, A. Dasdan, and H. Carcia-Molina, "Web graph similarity for anomaly detection," *Journal of International Service and Applications*, vol. 1(1), pp. 19–30, 2010.

[17] P. Mcdonald and R. Meyers, "Diffusion on graphs, poisson problems and spectral geometry," *Transaction of the American Mathematical Society*, vol. 354(12), pp. 5111–5136, 2002.

[18] W. Gong, "Can one hear the shape of a concept?," in *Proceedings of the 31st Chinese Control Conference (Plenary Lecture)*, Hefei, China, July 2012, pp. 22–26.

[19] W. Gong, "Transient response functions for graph structure addressable memory," in *Proceedings of the 52th IEEE Conference on Decision and Control*, Florence, Italy, December 2013, pp. 6978 – 6985.

[20] S. Lu, J. Kang, W. Gong, and D. Towsley, "Complex network comparison using random walks," in *Proceedings of the Companion Publication of the 23rd International Conference on World Wide Web Companion*, Seoul, Korea, 2014, pp. 727–730.

[21] F. Chung, *Spectral Graph Theory*, American Mathematical Society, 1997.

[22] S. Butler, "Interlacing for weighted graphs using the normalized laplacian," *Electronic Journal of Linear Algebra*, vol. 16, pp. 90–98, 2007.

[23] R. I. Kondor and J. Lafferty, "Diffusion kernels on graphs and other discrete structures," in *In Proceedings of the ICML*, 2002, pp. 315–322.

[24] International Game Fish Association, "Igfa fish database," Available at `http://www.igfa.org/fish/fish-database.aspx`.

[25] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86(11), pp. 2278–2324, 1998.

[26] C. C. Chang and C. J. Lin, "LIBSVM: A library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, pp. 27:1–27:27, 2011, Software available at `http://www.csie.ntu.edu.tw/~cjlin/libsvm`.