21st International Conference on Pattern Recognition (ICPR 2012)
November 11-15, 2012. Tsukuba, Japan

Local Tangent Space Based Manifold Entropy for Image Retrieval

Yi Wang[1], Hongyu Li[1,2], Junyu Niu[2*], Lin Zhang[1]
[1]*School of Software Engineering, Tongji University, Shanghai, China*
[2]*School of Computer Science, Fudan University, Shanghai, China*
{1096170028,hyli,cslinzhang}@tongji.edu.cn, jyniu@fudan.edu.cn

Abstract

This paper proposes a new manifold entropy function based on local tangent space (LTS). With this entropy function, we further propose a framework for image retrieval. The retrieval is treated as searching for ordered cycles by categories in image datasets. The optimal cycles can be found by minimizing our manifold entropy of images.

1. Introduction

Most of manifold based image retrieval methods consider the image feature space as an embedded manifold and try to find the mapping between the feature space and the manifold [6]. In [2], however, the retrieval is treated as searching for an ordered cycle in an image database through minimizing a manifold entropy function. This framework has a clear advantage over pervious manifold based methods: it can directly rank and return relevant images and does not need to learn a mapping from the observed feature space to the unclear semantic manifold space.

One major problem in [2] is that the implicit assumption that the embedded manifold is 1-dimensional is too oversimplified and restricted to be convincing since the intrinsic dimension of embedded manifold in feature space is generally greater than 1. Therefore, the discrete curvature of 1-dimentional curves cannot faithfully depict the local geometric properties of manifold. To overcome such problem, we propose a new entropy function on the basis of local tangent space (LTS) [1] in this work.

Local tangent space provides a good linear approximation for local geometry of nonlinear manifold. Our starting point is to use the subspace distance [3] between local tangent spaces to estimate geometric similarity between pairwise points on the nonlinear image manifold.

2. Manifold Entropy

2.1. Local Tangent Space

In differential geometry, a tangent space at a point intuitively contains the possible directions in which one can tangentially pass through this point. The elements of the tangent space are called tangent vectors. All the tangent spaces have the same dimension, equal to the dimension of the manifold.

The LTS at a given point is constructed from the local neighborhood of the given point. Given a data set composed of N m-dimensional points, $X = [x_1, \dots, x_N]$, $x_i \in R^m$, the LTS at each point can be obtained through computing the principal components in each local neighborhood. For each x_i, let $X_i = [x_{i_1}, \dots, x_{i_k}]$ be a matrix consisting of its k-nearest neighbors in terms of Euclidean distance. Let Q_i be an $m \times d$ matrix forming an orthonormal basis of the LTS of x_i. Q_i can be computed as the d left singular vectors of $X_i(I - \frac{1}{k}ee^T)$ corresponding to its d largest singular values, where I is an identical matrix and e is an m-dimensional column vector of all ones.

The local tangent space provides a low-dimensional linear approximation of the local geometric structure of the nonlinear manifold.

2.2. LTS Based Manifold Entropy

As discussed in [2], the manifold entropy is mainly constructed with two cues, the spatial position and the local geometric feature of each point. In this work, we propose to use the local tangent space as the local geometric feature at a given point. The subspace distance between the LTSs serves as a metric of geometric similarity between points to depict the manifold entropy.

In particular, given a set of data in an m-dimensional space $X = \{x_i | x_i \in R^m, i = 1,2, \dots, N\}$, we first define a *cycle* C of length N as a closed path without self-loops. Each datum in this cycle is connected with two neighbors and the corresponding connection order O is symbolized as,

$$X = \{o_1, o_2, \dots, o_N, o_1\}$$

where the entry corresponds to the index of data. Then the manifold entropy of the set X with the order O is calculated as the average of entropy on every point in the cycle C,

$$E(X, O) = \frac{1}{N} \sum_{i=1}^{N} e(X, O, i). \quad (1)$$

In this formula, $e(X, O, i)$ is composed of two parts: the spatial component $p(X, O, i)$ and the geometric component $g(X, O, i)$ as,

$$e(X, O, i) = p(X, O, i) + g(X, O, i). \quad (2)$$

The spatial component is measured by the Euclidean distance,

$$p(X, O, i) = d^2\left(o_i, o_{(i+1)}\right). \quad (3)$$

And the Euclidean distance is only computed for continuous points o_i and $o_{(i+1)}$ in the cycle C with the order O.

The geometric component is measured with a subspace distance defined in [3]. The subspace distance $d_s(S_a, S_b)$ between the p-dimensional subspace S_a and the r-dimensional subspace S_b is defined as,

$$d_s(S_a, S_b) = \sqrt{\max(p, r) - \sum_{i=1}^{p} \sum_{j=1}^{r} \left(u_i^T v_j\right)^2} \quad (4)$$

where u_i and v_j are respectively the i-th and j-th vector of the orthonormal basis of S_a and S_b.

In this work, LTSs of points in the same cycle C are of the same rank, $p = r$. Only keeping r, hence the geometric component of the LTS based manifold entropy is computed as,

$$g(X, O, i) = d_s\left(S_{o_i}, S_{o_{(i+1)}}\right) = \sqrt{r - \sum_{i=1}^{r} \sum_{j=1}^{r} \left(u_i^T v_j\right)^2}. \quad (5)$$

The manifold entropy E stated above is essentially a quantity of measuring the uncertain state of the cycle C and contains the smoothness and sharpness of the path with the connection order O. In addition, from the viewpoint of pattern recognition, the entropy is also a metric of disorder and similarity of the data.

3. Image Retrieval Framework

From the discussion in [2], we know that misclassifying a point into a category must lead the sharp increase of the entropy of this category with an optimal cycle. On the other side, if a point is correctly grouped into a category, the entropy of this category with an optimal cycle will only change a little.

Thus, the image retrieval framework works in the following way, as shown in Figure 1. Firstly, each category is respectively trained to obtain an optimal cycle (model). Then when a query image comes, we insert it at a position in one of these obtained cycles through comparing the entropy increase caused by inserting this image into each optimal cycle. After that, we rank the nearest neighbors of the query image at that position along the cycle according to Euclidean distance between images. Finally, the relevant images are returned.

3.1. Training Procedure

According to the definition of manifold entropy, to find an optimal cycle, we need to minimize the entropy,

$$O^* = arg\ min\ E(X, O). \quad (6)$$

In this study, we approximate the global minimum of the entropy through a simplified tabu search method [4]. The procedure of tabu search requires the design of neighborhood map and tabu list. The neighborhood map represents the transformation from a cycle to another. The tabu list is a short-term memory containing the representation of transformations, which is generally used to prevent previous transformation from being repeated.

3.2. Query Procedure

When a query image q comes, our strategy of retrieval is to find the "best" position of q in all optimal cycles. We mean by "best" that the increase of entropy ΔE due to adding q to the current position is the least among all positions. To reduce the disturbance of acquisition noise and computation error, we propose to introduce the idea of k-NN in our framework. Specifically, we pick the first k best positions from all cycles as k candidates, and then rank all cycles according to the following criterion,

$$r(C_i, q) = \frac{count(C_i)}{average(C_i)}, \quad (7)$$

where $count(C_i)$ is the number of candidates belonging to category C_i and $average(C_i)$ denotes the average entropy increase that candidates from category C_i result in. $average(C_i)$ can be formulated as,

$$average(C_i) = \frac{\sum_{C_i} \Delta S^*}{count(C_i)},$$

$$\Delta S_j^* = \Delta S_j \times \beta \qquad j = 1,2, \dots, k, \quad (8)$$

where parameter β is a weight for considering the position of each candidate in the candidate order.

The rank $r(C_i, q)$ implies the confidence of image q belonging to category C_i. Obviously, high confidence requires more $count$ and less $average$.

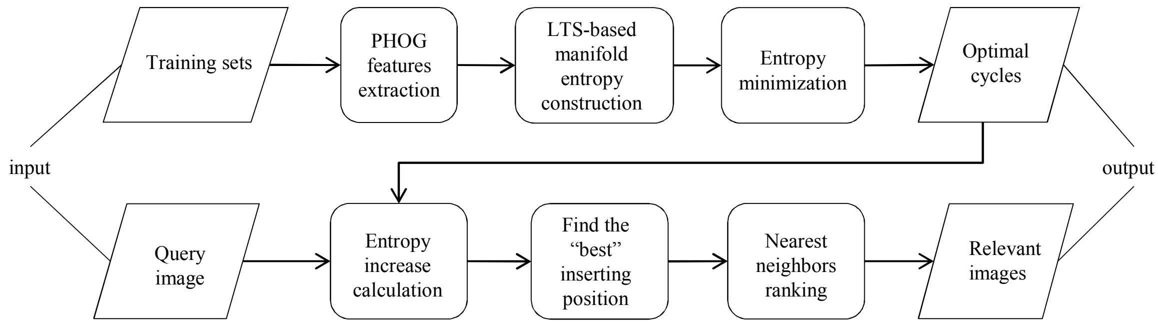

Figure 1. Flow chart of the framework. First row: training procedure. Second row: query procedure.

4. Experimental Results

This section presents the results of experiments conducted to evaluate and analyze the performance of the proposed method for image retrieval. Two datasets are employed for image retrieval. 2,452 images of 23 semantic categories from Caltech101 Object_ Categories build the first dataset. The second dataset is downloaded from an online shopping site, composed of 1,506 images of 16 categories.

4.1. Parameter Choice

In our algorithm, we have to preset two parameters: the tabu list size and the coefficient β in equation (8).

Actually, the tabu list size is an important key parameter of tabu search method. It greatly affects the optimization speed and the quality of the solution. The fixed size of the tabu list in [2] may not help to find the best solution due to different scale of each category. Therefore, the tabu list size in this study is set to be the rounded square root of each problem scale.

The premise of equation (8) is that candidates with more entropy increase should be penalized. So the coefficient β is computed as,

$$\beta = \log_2(j+1)\,. \qquad j = 1,2,\dots,k$$

Obviously, the j-th entropy increase is enlarged logarithmically proportional to the position j.

4.2. Performance Analysis

This study adopts the Pyramid of Histograms Orientation Gradients (PHOG) [5] as the content of images since PHOG pays more attention to high-level features. Once image features such as PHOG are extracted, we can compute the local tangent space at each feature point.

To evaluate the performance of our approach, 10-fold cross-validation experiments were conducted on two datasets mentioned above. Firstly, the training set for each category was trained to obtain an optimal cycle respectively. Then each image from the validation set, in turn, was used as a query example to launch the retrieval. In this work, experiments were completed on a machine with an NVIDIA GTX 260 GPU card.

Figure 2 shows some examples of retrieval on the Caltech101 dataset using our framework. These examples give an intuition of visual similarity captured with our method. We can find that the proposed image retrieval method can perform particularly well for two categories, "winsdor_chair" and "stegosaurus". Only few irrelevant images are returned for other categories.

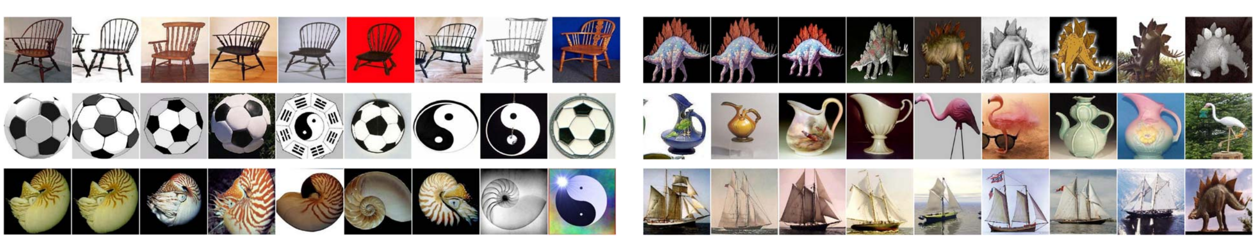

Figure 2. Image retrieval on the Caltech101 dataset. Most left column: query images. Other columns: the top 8 retrieved images. Category names from top-left to bottom-right: winsdor_chair, stegosaurus, schooner, nautilus, ewer, soccer

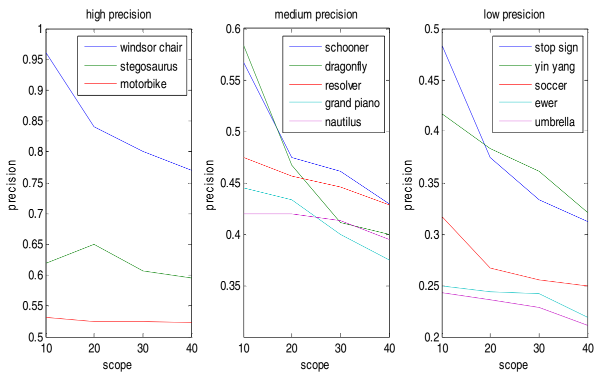

Figure 3. Performance of our image retrieval framework on the Caltech101 dataset.

Figure 3 presents the average precision for each category of the Caltech101 dataset. In this context, the scope denotes the number of top returned images, and the precision is the ratio of the number of top relevant images to the scope. When the scope is 10, the average precisions of these 23 categories all exceed 50%, with the highest 95%. Essentially, excellent categories, such as "windsor_chair" and "stegosaurus", represent coherent natural scenes with less clutter. The low precision for categories "ewer" and "soccer" is partly the result of the ambiguity in visual content. However, our method can overall improve the average precision 10% over the work in [2].

In addition, based on the proposed framework, we build a prototype of an item search engine for online shopping site. The second dataset were used to test the item search and Figure 4 presents some search results of two items. Despite the color difference, the shape of the returned images overall looks similar. Combined with powerful image features, our framework can handle different feature-related item search requests.

5. Conclusion

In this paper, we propose an entropy function based on local tangent space. Subspace distance between local tangent spaces provides a good estimation of geometric similarity between high-dimensional points on nonlinear manifold. In our framework, the retrieval is treated as searching for ordered cycles by categories in image datasets. The optimal cycles can be found by minimizing our entropy function through tabu search. Relevant images are returned as the nearest neighbors along the best optimal cycle. Experimental results on Caltech 101 demonstrate that the proposed method is quite promising.

6. Acknowledgment

This work was partially supported by Natural Science Foundation of China Grant 60903120, 863 Project 2009AA01Z429, Shanghai Natural Science Foundation Grant 09ZR1434400, and Innovation Program of Shanghai Municipal Education Commission.

(a) backpack

(b) short

Figure 4. Item search for an online shopping site

7. References

[1] Z. Zhang and H. Zha, "Principal Manifolds and Nonlinear Dimension Reduction via Local Tangent Space Alignment," in Technical Report CSE-02019, CSE, Penn State Univ., 2002.

[2] C. Zhang, H. Li, Q. Guo, J. Jia, and I.-F. Shen, "Fast active tabu search and its application to image retrieval," in *IJCAI'09*, pages 1333-1338, 2009.

[3] Guido Zuccon, Leif A. Azzopardi and C. J. van Rijsbergen, "Semantic Spaces: Measuring the Distance between Different Subspaces," in Quantum Interaction 2009, LNAI 5494, pages 225-236, 2009.

[4] F. Glover, "Future paths for integer programming and links to artificial intelligence," in *Computers and Operations Research*, volume 13, pages 533-549, 1986.

[5] A. Bosch, A. Zisserman, and X. Munoz, "Representing shape with a spatial pyramid kernel," in *CIVR'07*, pages 401-408, 2007.

[6] D. Cai, X. He, and J. Han, "Regularized regression on image manifold for retrieval," in MIR '07: Proceedings of the international workshop on Workshop on multimedia information retrieval, pages 11–20. NewYork, NY, USA, 2007. ACM.