Entropy Based Image Semantic Cycle for Image Classification

Hongyu Li[1,2], Junyu Niu[3,⋆], and Lin Zhang[2]

[1] Electronic Engineering Department, Fudan University, Shanghai, China
[2] School of Software Engineering, Tongji University, Shanghai, China
[3] School of Computer Science, Fudan University, Shanghai, China
{hongyuli,jyniu}@fudan.edu.cn, cslinzhang@tongji.edu.cn

Abstract. This paper proposes a novel framework for image classification with an entropy based image semantic cycle. Entropy minimization leads to an optimal image semantic cycle where images are connected in the semantic order. For classification, the training step is to find an optimal image semantic cycle in an image database. In the test step, the suitable position of an unknown image in this cycle is first found. Then, the class membership is determined through recognizing the nearest neighbors at this position. Experimental results demonstrate that the proposed framework achieves higher classification accuracy.

Keywords: Entropy, Image Semantic Cycle, Image Classification.

1 Introduction

As an integral part of image management and retrieval, image classification aims to classify an unknown image by the object category, has been the subject of many recent papers [1,2,3,4,5]. Most state-of-the-art image classification work focuses on the description of images. One of the popular approaches to image classification is to model images with bag-of-features (BOF) [6,7] and classify them with general classifiers, such as the Support Vector Machine (SVM) method [2]. Although it has been reported that SVM classifiers have a good performance in image classification, the classification accuracy still needs to improve for practical applications.

Motivated by [8], this paper proposes a novel framework for image classification. The core of this framework is to design an entropy based image semantic cycle, i.e., a classifier in essence, for an image database. The training procedure in the proposed framework employs the prior information of image class membership to find the optimal image semantic cycle, different from the idea in [8] where image retrieval is an unsupervised learning problem and does not require any priori.

For a test image, the strategy of classification here is to find its best position in the trained optimal image semantic cycle, where the "best" means that the entropy increases the least when inserting the image into this position.

⋆ Corresponding author.

T. Huang et al. (Eds.): ICONIP 2012, Part V, LNCS 7667, pp. 533–540, 2012.

2 Entropy Based Cycle

Statistical mechanics views entropy as the amount of uncertainty, or mixedupness in the phrase of Gibbs. In the field of computer vision, the concept of entropy is also widely used, e.g., the literature [9] and [10].

2.1 Entropy Definition

Given a set of vertices $V = \{v_i | i = 1, \cdots, n\}$, there exists a *cycle* graph C, a closed path without self-intersections, among them in which each vertex has degree 2. Suppose the cycle starts from vertex v_1 and ends at vertex v_n, we can write the connection order of vertices in this cycle as

$$O = \{o_1, o_2, \ldots, o_n, o_1\},$$

where the entry corresponds to the index of vertices. The path along the connection order O implies the relationship between vertices and embodies the degree to which the cycle is disordered and mixed up. If the path is smooth and desirable, the cycle should be clean and in order; otherwise, the cycle is disordered.

Like in statistical mechanics, to measure the degree to which the cycle is disordered and mixed up, this paper proposes a novel definition of entropy descriptor. The entropy definition is quite general, and is expressed in terms of the spatial position and local discrete curvature of data points.

In particular, for a set of vertices V, if the spatial coordinate x_i of each vertex v_i and the connection order O are known, the entropy of a cycle C is represented as the average of the entropy on every point in the cycle as,

$$E(V, O) = \frac{1}{n} \sum_{i=1}^{n} e(V, O, i). \tag{1}$$

$e(V, O, i)$ is defined as follows,

$$e(V, O, i) = 1 - \exp(-\frac{G(V, O, i)}{\sigma}), \tag{2}$$

where the symbol $\exp(\cdot)$ denotes an exponential function and σ is an empirical constant representing the variance. $G(V, O, i)$ is composed of two parts: the spatial distance $D(V, O, i)$ and local discrete curvature $L(V, O, i)$ as,

$$G(V, O, i) = D^2(V, O, i) + \alpha L^2(V, O, i), \tag{3}$$

weight coefficient α is used to adjust the contribution of L to the entropy. The definition of these two parts is same as that in [8].

The 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. Therefore, for a set of vertices V, the implicit *cycle* can be formally defined as a triplet in this paper,

$$C = (V, O, E).$$

In addition, from the viewpoint of pattern recognition, the entropy is also a metric of distinction and similarity of the data. That is, the smaller the entropy, the more ordered the path and the cycle, the more similar the continuous points along the order.

2.2 Finding an Optimal Cycle Using Tabu Search

From the definition of entropy, if the cycle connecting vertices are ordered enough, the entropy is supposed to be quite small. Other possible paths incline towards the more disorder and the higher entropy. Therefore, to find an optimal cycle, we need to search for the cleanest order O, which can be achieved through minimizing the entropy defined above,

$$\min_{O} E(V, O). \tag{4}$$

That is, the order O^* for the optimal cycle C^* is obtained as follows,

$$O^* = \arg\min E(V, O). \tag{5}$$

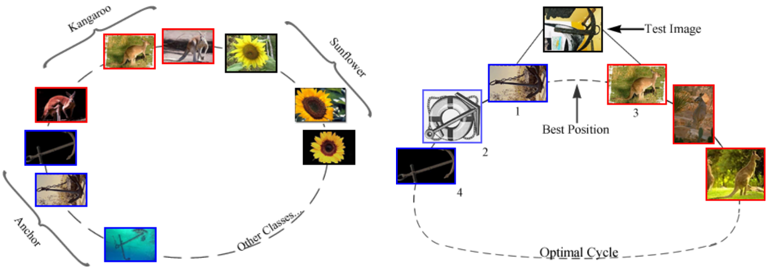
Finding a globally optimal solution to equation (5) is an NP-hard problem and completely impossible in practice as there exist $\frac{(n-1)!}{2}$ possible combinations of order for $n(n \geq 4)$ vertices. In this study, we approximate the global minimum of the entropy through a simplified tabu search method [8].

3 Framework for Image Classification

In the proposed framework for image classification, we view an image as a point in the image feature space, and train the images in the way of searching for the optimal cycle through entropy minimization. In this context, the connection order O represents the relevance of images in the entropy based cycle. If local or global features of images are extracted and used to represent images, the entropy based cycle is essentially *an image semantic cycle* containing the content of images. This paper adopts the Pyramid of Histograms Orientation Gradients (PHOG) [11] as the content of images for PHOG pays more attention to such high-level features as edge or contour. In the following experiments, the PHOG descriptor turns an image into a 252-dimensional vector.

3.1 Training Algorithm

The fact that continuous points along the optimal cycle usually belong to the same category except for those boundary points enlightens us that the nearest neighbors of an image along the optimal image semantic cycle should be more similar and relevant except for those boundary images between different categories. Therefore, to do image classification, the key step in our framework is to search for the optimal image semantic cycle, which is actually our training procedure.



(a) initialization with category (b) grouping the test image

Fig. 1. An example of image classification

Since classification is a supervised learning problem, the categories of training images are already known and can be made full use of during the training. Therefore, the initialization of the image semantic cycle in the training does not have to be completely random. The idea of "locally random" initialization is brought up. That is, images in the same class are randomly linked together and hold a separate section of the initial cycle, different classes (sections) are randomly placed in the initial cycle. The idea of initialization with category information is illustrated in Fig.1(a). Images from category *kangaroo* and category *sunflower* are put together respectively and different categories, *kangaroo, sunflower, anchor etc.* are randomly placed in the cycle. This initialization strategy considers the known class membership during training, the speed of searching for the optimal image semantic cycle is thus obviously improved.

Given an image database with category information, the training procedure is briefly described as in Table 1.

Table 1. Training Algorithm

Step 1	Initialize the image semantic cycle with a *locally random* order O.
Step 2	Construct the entropy descriptor with Eq.(1) in the image feature space.
Step 3	Minimize the entropy with tabu search, and search for the optimal image semantic cycle with the order O^*.

3.2 Test Algorithm

The classification strategy here is to find the best position of a test image Q in the optimal image semantic cycle and determine the class membership with its nearest neighbors along this optimal cycle. By "best", we mean that the entropy increase of the optimal image semantic cycle will be the least if inserting the test image into this position.

The principle of grouping the test image is borrowed from in the idea of k-NN. More specifically, k nearest neighbors of image Q are first picked from the optimal cycle; then image Q is considered as a member of the class to which most neighbors belong. The measure of determining the nearest neighbors can be the Euclidean distance for simplicity. For example, the 4 nearest neighbors of the test image in Fig.1(b) come from two different categories respectively, three from *anchor* and one from *kangroo*. As a consequence, the test image is grouped into category *anchor*.

Given a test image Q, the test procedure in the proposed framework is briefly described as in Table 2.

Table 2. Test Algorithm

Step 1	Calculate the entropy variation ΔE when inserting the unknown image Q at each position in the optimal image semantic cycle O^*.
Step 2	Find the best position for image Q from the optimal cycle O^*. The measure of determining the best position is that ΔE must be the least if inserting Q at this position.
Step 3	Select k nearest neighbors of Q along the optimal cycle according to Euclidean distance.
Step 4	Assign image Q to the class from which most neighbors come.

3.3 Dynamic Update

An obvious advantage of the proposed framework is that the dynamic update of the trained model, the image semantic cycle, is quite simple and easy. Suppose that we already had an optimal image semantic cycle, $C = (V, O, E)$, and now a new image Q comes to this cycle. To update this cycle, the image Q is first added to V. As a result, the new vertex set V' is equal to the union of V and Q, i.e., $V' = union(V, Q)$. Secondly, find the best position of Q in the current order O. Since the "best" means that the entropy variation ΔE is the least at this position, the new entropy E' remains the least if inserting Q at this best position, $E' = E + \Delta E$. In other words, the new cycle C' after inserting will keep the optimal property. At this time, the connection order O of images is also changed to O'. Therefore, the new image semantic cycle C' is reformulated as, $C' = (V', O', E')$.

In fact, the training procedure can also be completed in the way of dynamic update. A subset of images are first trained to get the optimal image semantic cycle, which is very quick due to the small number of images. The other images gradually come and dynamically update the image semantic cycle until all images join in the training procedure. This dynamic update for training is essentially a procedure of incremental learning, which thus has the potential application to video tracking.

4 Experimental Results

Since the calculation of entropy of each element is independent, parallel computing is a good way to speed up such an optimization problem. Like the work [8], this study also makes full use of the ability of graphic processing unit (GPU) in parallel computing and implements the training procedure with the CUDA

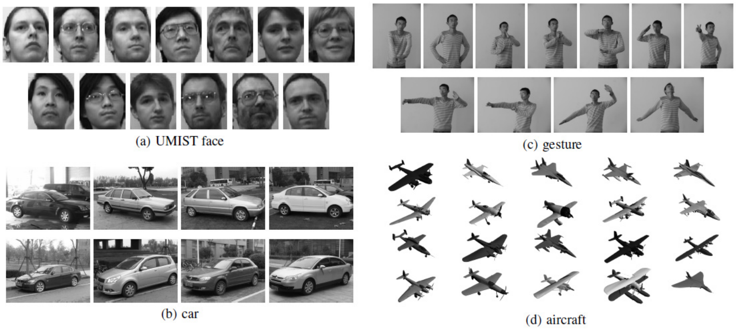


Fig. 2. Example images of four datasets used in our experiments

programming technology. The convergence speed of entropy minimization can become faster 100 times with GPU than with CPU. In the parallel way, the test procedure is as quick as realtime and is hardly affected by the data scale.

This section presents the results of experiments conducted to evaluate and analyze the performance of the proposed framework for image classification. The datasets used are first presented, followed by the comparison of the proposed framework with the SVM based framework.

4.1 Datasets

Five datasets were tested respectively for image classification. The first dataset is the UMIST face database [12]. To track the smooth change of shooting angles, 360 face images belonging to 13 persons were manually picked in our experiments from the UMIST face. The car and gesture datasets were acquired by ourselves. The former contains 14 different types of cars, each of which has about 30 photos. The latter was generated in the laboratory environment, composed of 11 different human gestures and 20 images for each gesture. Another dataset concerns 700 images of 7 aircraft models. The images were produced through projecting three-dimensional models into different planes, which is equivalent to changing the viewpoint through rotating the models. Some example images of the above four datasets are shown in Fig.2. The last dataset includes 3286 images of 32 categories with significant variance in shape, picked out from the Caltech-101 Object Categories dataset[13].

4.2 Classification Accuracy

To evaluate the performance of the proposed approach to image classification, 10-fold cross-validation experiments were here conducted on five datasets mentioned above. To compare with the state of the art, the SVM based classifier [14] was also tested on same datasets due to its advantages in speed and excellent performance in image classification. All test results are presented in Table 3, where the second and third columns respectively represent the number of classes

Table 3. Classification accuracy on different datasets

datasets	class	image	Ours(%)	SVM(%)
UMIST	13	360	98.89(0.77)	85.28(3.13)
car	14	400	99.25(2.08)	74.00(2.79)
gesture	11	220	99.60(1.11)	78.64(5.85)
aircraft	7	700	99.00(1.35)	74.57(4.84)
Caltech-101	32	3286	88.88(2.94)	78.77(2.98)

and images in each dataset. The last two columns are the average classification accuracy respectively obtained using our approach and SVM. The value in parentheses denotes the confidence interval. Although it is well known that SVM has the strong ability of generalization in image classification, the proposed approach obviously has the better performance than SVM, higher accuracy and smaller confidence interval.

The accuracy for the first four datasets is very high and almost reaches 100%. Such good classification performance on the UMIST face and gesture datasets demonstrates that the proposed framework is quite promising in the application of face and gesture analysis. The success in classifying the projection images of aircraft models makes possible the future image-based model retrieval.

In spite of the existence of background clutter in the Caltech-101 (or car) images, our method can still work well with the average accuracy 88.88% (99.25%) and confidence interval 2.94 (2.08), much better than SVM with the average accuracy 78.77% (74.00%) and confidence interval 2.98 (2.79).

To sum up, no matter how different the viewpoint of images is or how cluttered the image background is, the performance of our method is always better than the SVM based classifier, which proves the feasibility and robustness of the proposed framework for image classification.

5 Conclusion

In this paper, we propose an entropy based image semantic cycle for image classification. The training problem in classification is treated as searching for an ordered image semantic cycle for an image database. The optimal cycle is found by minimizing the entropy through tabu search. For classification, the test image is assigned to the class to which most neighbors belong. The use of GPU in our framework yields a very considerable speedup. Experiments demonstrate that the proposed method is feasible and robust to the cluttered background and the viewpoint variation of images.

Acknowledgement. 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.

References

1. Bosch, A., Zisserman, A., Munoz, X.: Image classification using random forests and ferns. In: ICCV, pp. 1–8 (2007)
2. Goh, K.S., Chang, E., Cheng, K.T.: SVM binary classifier ensembles for image classification. In: CIKM 2001, pp. 395–402 (2001)
3. Wang, J., Yang, J., Yu, K., Lv, F., Huang, T.S., Gong, Y.: Locality-constrained linear coding for image classification. In: CVPR, pp. 3360–3367 (2010)
4. Gao, S., Tsang, I.W.H., Chia, L.T., Zhao, P.: Local features are not lonely - laplacian sparse coding for image classification. In: CVPR, pp. 3555–3561 (2010)
5. Cai, H., Yan, F., Mikolajczyk, K.: Learning weights for codebook in image classification and retrieval. In: CVPR, pp. 2320–2327 (2010)
6. Sivic, J., Zisserman, A.: Video google: A text retrieval approach to object matching in videos. In: ICCV, pp. 1470–1477 (2003)
7. Csurka, G., Dance, C.R., Fan, L., Willamowski, J., Bray, C.: Visual categorization with bags of keypoints. In: Workshop on Statistical Learning in Computer Vision, ECCV, pp. 1–22 (2004)
8. Zhang, C., Li, H., Guo, Q., Jia, J., Shen, I.F.: Fast active tabu search and its application to image retrieval. In: IJCAI, pp. 1333–1338 (2009)
9. Wang, W., Wang, Y., Huang, Q., Gao, W.: Measuring visual saliency by site entropy rate. In: CVPR, pp. 2368–2375 (2010)
10. Lin, R.S., Ross, D.A., Yagnik, J.: Spec hashing: Similarity preserving algorithm for entropy-based coding. In: CVPR, pp. 848–854 (2010)
11. Bosch, A., Zisserman, A., Munoz, X.: Representing shape with a spatial pyramid kernel. In: CIVR 2007, pp. 401–408 (2007)
12. Graham, D.B., Allinson, N.M.: Characterizing virtual eigensignatures for general purpose face recognition. Face Recognition: From Theory to Applications 163, 446–456 (1998)
13. Fei-Fei, L., Fergus, R., Perona, P.: Learning generative visual models from few training examples: An incremental bayesian approach tested on 101 object categories. Comput. Vis. Image Underst. 106, 59–70 (2007)
14. Chang, C., Lin, C.: LIBSVM: a library for support vector machines (2001)