Visual words assignment on a graph via minimal mutual information loss

Yue Deng
http://media.au.tsinghua.edu.cn/dengyue.html
Yanjun Qian

Yipeng Li

Qionghai Dai
http://media.au.tsinghua.edu.cn/qhdai.html
Guihua Er

Department of Automation
Tsinghua University
Beijing, China

Abstract

Visual codewords assignment plays an important role in many Bag of Features (BoF) models for image understanding and visual recognition. It allocates image descriptors to the most similar codewords in the pre-configured visual dictionary to generate descriptive histogram for the consequent categorization. Nevertheless, existing assignment approaches, *e.g*. nearest neighbors strategy and Gaussian similarity, suffer from two problems:1) too strong Euclidean assumption and 2) neglecting the label information of the local features. Accordingly, in this paper, we propose an assignment method to simultaneously consider the above two issues in a unified model via graph learning and information theoretic criterions. For learning, the proposed model can be efficiently solved in a closed-form with the reasonable graph topology invariant approximation. Moreover, the learned projections enable us to extend the assignment ability to the out-of-sample visual features beyond the initial training graph. Experiments on our own manifold dataset and two benchmarks verify the effectiveness of the proposed graph assignment method.

1 Introduction

In computer vision, a long standing but still challenging problem is how to accurately categorize different images based on their intrinsic contents. Most of the existing frameworks to solve this problem involve two steps [3] [4]: 1) image representation and 2) image classification. The second step is a well defined machine learning task and can be effectively solved by some strong classification machines, e.g. SVM. Therefore, the prominent task in image categorization is subject to how to effectively generate visual vectors[1] to represent different images. For image representation, prevalent algorithms include topic models, *e.g*. Latent Dirichlet Allocation (LDA) [10][2], Codeword similarity assignment [13][15][18] and sparse coding [19]. In this paper, we mainly focus on the codeword similarity based methods for

[1]We call the generated information as visual *vectors* to distinguish

image categorization because it usually achieves higher accuracy than the topic models and greatly releases the heavy computational burdens of sparse coding methods.

Codeword similarity methods construct a histogram to describe the original image by assigning each local image feature to a codeword in the dictionary, a.k.a. codebook. Codebook generation is the first step of our framework which produces the representative codewords by clustering the training local image features into different clustering centers. The second step is codeword assignment which allocates different image features to the selected codewords according to some strategy and similarity. For codeword assignment, albeit many variations have been proposed around the years, hard assignment [15] and soft assignment with Kernels [18] are two of the most influential works. Although hard and soft methods show promising results on a number of practical applications, we argue that these two methods suffer two problems that deserves consolidated investigations.

For typical assignment, the first problem is that they are all implemented in the Euclidean space while neglecting the intrinsic distribution of massive image features. It is widely known that it is not possible to lose sight of the nonlinearity of data distribution when a large amount of data are available[7] [1] [8]. However, in the Euclidean space, all the feature similarity are judged by the linear metric which is not sufficient to reveal the nonlinear structure of data. Accordingly, inspired by many previous works on manifold learning, in this paper, we propose to use a graph to model the distribution of data. On the feature graph, different nonlinear similarity, *e.g.* geodesic distance and commute time of random walk, can be used as a nature metric to evaluate the feature similarity during assignment.

Besides, typical codeword assignment methods only consider the feature similarity while neglecting the label information contained in each local feature. Since the images are from different categories, the local image features also have their own label information of the image category. However, this critical property is usually neglected for both codebook selection and codeword assignment. Accordingly, inspired by the previous works on information theoretical clustering [14], we consider using the mutual information to reveal such a mutual relationship between local image features and their corresponding image labels.

Accordingly, we design the Graph Assignment with minimal Mutual Information Loss (GAMIL) model to address the aforementioned properties for codeword assignment. In a nutshell, our algorithm considers finding orthogonal projections to preserve the original nonlinearity of the training image features in an embedding space with minimal mutual information loss. Then, all the in-sample and out-of-sample local image features can be assigned to the codewords in the embedding space. In order to effectively solve the GAMIL model, we use the Epanechnickov Kernel [9] to estimate the probability density function (pdf) of the feature distribution in the embedding space. With such a kernel, the GAMIL can be well solved in a closed-form which greatly reduce the computational complexity of optimization.

In order to verify the performance of the proposed method, we compare our GAMIL with other state-of-the-art assignment methods on three datasets. We publish our own Multi-view human body dataset, a laboratory synthetic dataset, to investigate the advantages of GAMIL to handle data under manifold distributions. The improvements of GAMIL are also validated on two benchmarks of Scene-15 [10] and Caltech-101 [11].

2 Graph assignment with minimal mutual information loss

In this part, we will introduce the proposed GAMIL model. We define the local image features set as $S = \{(f_1, l_1), (f_2, l_2)...(f_n, l_n)\}$, where $f_i \in \mathbb{R}^p$ is the image feature, *e.g.* dense sift, extracted on the original image and $l_i \in \{1, 2, ...C\}$ is the category label of the image that f_i is extracted from. C is the number of image categories. The quantity of samples in the training set S is quite huge since these image features are densely sampled from a many images. For example, for the Scene-15 dataset, the quantity of training features is more than $320,000$. With the increasing of data quantity, the manifold structure of data will come out.

2.1 Describing feature similarity via graph

(a) ground truth   (b) assignment via ED.   (c) assignment via GD.   (d) assignment via CT.
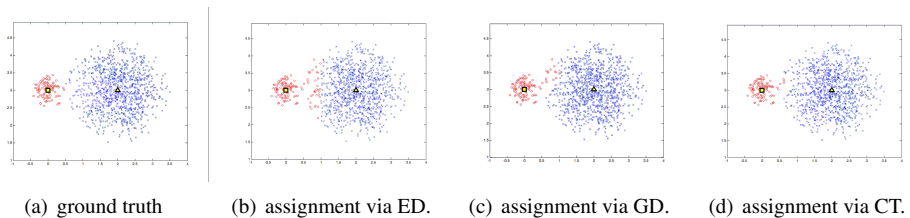
Figure 1: A toy assignment via different methods, i.e. Euclidean Distance(ED), Geodesic Distance(GD) and Commute Time(CT).

Therefore, in this paper, we propose to use a graph to model the samples in S. Using the manifold structure, it is possible to model the linearity among data by the locality similarity; and the global nonlinearity can be evaluated by some graph metric on the manifold. For graph construction, we use the KNN method that one node f_i is connected with its nearest k neighbors. Two widely known graph metric to evaluate the nonlinear relationships between data for assignment are geodesic distance [12] and commute time [17]. Geodesic distance refers to the shortest distance between two nodes on the graph and commute time records the steps of a random walk to travel around a pair of nodes. An intuitive toy example that shows the advantages of graph assignment is provided in Fig.1. In the figure, the yellow markers represent the codewords and the features are assigned to the codeword with the nearest distance strategy by different metrics. Obviously, the manifold assignment methods by geodesic distance and commute time generally outperform the result assigned by Euclidean distance.

But the above-motioned method on a graph is only suitable to the in-sample features. For practical usage of codeword assignment, it is desirable to extend the assignment ability to the out-of-sample data. One trivial solution to solve this problem is to construct such a manifold for any testing image and then to calculate the graph metric from the feature nodes to the codeword. However, the computational cost is extremely huge for real time usage. Fortunately, inspired by [8], we propose to embed the graph into an Euclidean space with a linear projection matrix. In the embedded space, the original graph metric is well preserved by the Euclidean distance. According to [8], we give the graph metric guided transformation,

$$\min_{\Omega} tr(\Omega^T F(D-W)F^T \Omega) \quad s.t. \quad \Omega^T FDF^T \Omega = I \tag{1}$$

where $F = [f_1, ..., f_n] \in \mathbb{R}^{p \times n}$ is the feature matrix; $\Omega \in \mathbb{R}^{p \times q}, q < p$ is the linear projection matrix. $W = [w_{ij}]$ is the weight matrix obtained on the graph which records the similarity between any two nodes i, j on the graph and $D = diag(\sum_i W_{ij})$. $tr(\cdot)$ represents the trace of a matrix. Eq.1 projects all the training samples into the subspace with a linear projection and the clustering centers in the embedding space are selected as codewords. For an out-of-sample feature, we first project it to the subspace and then assign it to each codeword via Euclidean similarity. It is because the Euclidean distance in the embedding space represents the original nonlinear graph similarity on the manifold.

2.2   Minimal mutual information loss criterion

In the previous section, we propose to use a graph structure to reveal the similarity among the training features. But it worths noting that each feature in the set S also contains the label information. During training, we know where the image feature comes from. Therefore, it will be better if we can also take the label information into the assignment process. Accordingly, the problem changes to be how to evaluate the relationship between the features F and their labels L. Fortunately, owing to the previous work [14], we know that the relationship of feature and label is always judged by the mutual information, i.e. $I(F;L)$. In probability theory and information theory, the mutual information [6] of two random variables is a quantity that measures the mutual dependence of the two random variables. It measures how much knowing one of these variables reduces uncertainty about the other. Informally, in our case, $I(F;L)$ can be interpreted as how much the uncertainty is reduced about the label L if we know the feature F. Therefore, for discriminative learning, a large mutual information score is desired.

The proposed graph assignment method projects the high dimensional feature in a low dimensional space ($q < p$). Ideally we hope that the mutual information on the original graph should be kept the same in the embedding space, i.e. $I(F;L) = I(\Omega^T F, L)$. Unfortunately, reducing the dimensionality of data from high to low of course causes information loss. Therefore, instead of mutual information preservation, we propose to use the minimal mutual information loss criterion, i.e. to minimize $I(F;L) - I(\Omega^T F;L)$. $I(F;L)$ is a fixed value because the original F and L are both known. Therefore, the optimization changes to minimizing $-I(Y = \Omega^T F;L)$. So, the remaining problem is how to calculate the mutual information in the embedding space. Starting from the basic definitions in information theory, we know $I(Y;L) = H(L) - H(L|Y)$, where $H(\cdot)$ means the entropy. The probability of labels, i.e. $P(L)$ is usually assumed to be uniform distribution or it can be estimated from training samples. No matter which prior for the label information is used, the entropy $H(L)$ is a fixed value and can thus be dropped from the optimization. Therefore, by considering both the information loss and graph similarity, the optimization for our GAMIL model is given,

$$\min_{\Omega} \underbrace{tr(\Omega^T F(D - W)F^T \Omega)}_{\text{graph assignment}} + \underbrace{\alpha H(L|Y = \Omega^T F)}_{\text{mutual information loss}} \quad s.t. \Omega^T FDF^T \Omega = I, \quad (2)$$

where α is a user specified parameter which trades off the graph assignment and mutual information loss.

We will stop here to make some remarks on the second term in (2). Although $H(L|Y)$ is obtained from minimal information loss criterion, we can also give the physical explanations to it directly. $H(L|Y = \Omega^T F)$ is the entropy for the labels conditioned on knowing the features in the embedding space. In the ideal case that the entropy is optimally minimized,

the uncertainty of L by giving Y reduced to the minimum. The minimal entropy implies the maximal determination of labels by seeing the features. However, in (2), the entropy can not be arbitrarily minimized. We can only find the best Ω that seeks the balance between the entropy minimization and graph similarity preservation.

2.3   Model learning

The GAMIL model in (2) is not easily to be learned since we do not know the true probability distribution for the labels in the embedding space. One way to make it tractable is to estimate the entropy and probability in a discrete form from the training data. Therefore, we get,

$$H(L|Y) = \sum_{y \in Y} p(y)H(L|Y=y) = - \sum_{f \in F} p(\Omega^T f) \sum_{l \in L} p(l|y=\Omega^T f) \log p(l|y=\Omega^T f) \quad (3)$$

To explicitly get the entropy, we should estimate probability density function (pdf) of $p(\Omega^T f)$ and $p(l|Y=\Omega^T f)$. We calculate $p(\Omega^T f)$ using the kernel density estimation that

$$p(\Omega^T f_i) = \frac{1}{N} \sum_{j \neq i} \frac{1}{\sigma} K\left(\frac{\Omega^T f_i - \Omega^T f_j}{\sigma}\right) \quad (4)$$

, where $K(\cdot)$ is the selected kernel density and σ is the parameter. There are many the kernel density functions, $e.g.$ the widely used Gaussian window. However, in this paper, we prefer to use the Epanechnickov Kernel [9] because it may help us simplify the final optimization. The Epanechnickov Kernel $K_\varepsilon(x) = \frac{3}{4}(1-x^2) \ \forall |x| < 1$ and 0 otherwise. The probability

$$p(l|y_i = \Omega^T f_i) = \frac{1}{\|N_\sigma(y_i)\|} \sum_{y_j \in N_\sigma(y_i)} \delta(l_j = l) \quad (5)$$

, where $\delta(\cdot)$ is the Kronecker delta function and $N_\sigma(\cdot)$ represents the $\sigma - ball$ nearest neighbors. By taking $K_\varepsilon(\cdot)$ and $p(l|y_i)$ into (3) and with some simple algebra, we get that,

$$\frac{3}{4} \sum_{y_i} \sum_{y_j \in N_\sigma(y_i)} \sigma^2 H(L|y_i) - \frac{3}{4} \sum_{y_i} \sum_{y_j \in N_\sigma(y_i)} H(L|y_i)\|y_i - y_j\|^2$$
$$= \frac{3}{4} \sum_{y_i} \sum_{y_j \in N_\sigma(y_i)} \sigma^2 H(L|y_i) - \frac{3}{4} \sum_{ij} H_{ij}\|y_i - y_j\|^2 \quad (6)$$

In (6), $H_{ij} = H(l|y_i)$ if $y_i \in N_\sigma(y_j)$ and zero otherwise. Up to now, we have shown how to estimate $H(L|Y)$ from the training data. However, the optimization is still intractable by such an estimation. This is because the estimation of $H(L|y_i)$ involves a nonlinear functionality of finding the nearest neighbors in the embedding space. Ways to solve such an intractable optimization may use some methods like Gibbs sampling which is computational heavily. Therefore, in this paper, we introduce an approximation strategy to make GAMIl model solved efficiently.

In our approach, we note that the first term in (2) is a graph preservation term, which preserves the topology of the original graph into the embedding subspace. Accordingly, we assume that after the graph projection, the neighboring topology on the original graph is kept the same in the embedding space. Of course, this assumption does not strictly hold for all the nodes. But we believe, such topology invariant rule is applied to most of the nodes

in the embedding space. This graph topology preservation property is guaranteed by the first embedding term in GAMIL. Based on such an assumption, we can use the conditional entropy $H(l|f_i)$ on the original graph to replace the entropy $H(l|y_i)$ in the embedding space. When taking such an approximation into (6), the first term becomes a constant and H_{ij} in the second term is fixed. Therefore, the GAMIL model subject to the optimization ,

$$\min_{\Omega} tr(\Omega^T F(D-W)F^T\Omega) - \alpha tr(\Omega^T F(D_H-H)F^T\Omega) \quad s.t. \ \Omega^T FDF^T\Omega = I, \quad (7)$$

where $D_H = diag(\sum_i(H_{ij}))$. Eq.7 is a standard convex function with quadratic objective. If we define $L_W = D-W$ and $L_H = D_H - H$, we can write the Lagrangian multiplier of 7 in the form of,

$$L(\Omega,\Lambda) = tr(\Omega^T F(L_W - \alpha L_H)F^T\Omega) + <\Lambda, I - \Omega^T FDF^T\Omega>, \quad (8)$$

where $<A,B> = tr(AB^T) = tr(A^T B)$ is the inner product and Λ is the Lagrangian multiplier which is a diagonal matrix. By setting $\frac{\partial L}{\partial \Omega} = 0$, we get,

$$F(L_W - \alpha L_H)F^T\Omega = FDF^T\Omega\Lambda. \quad (9)$$

Therefore, the optimal solution to Ω is subject to such a general eigenvalue decomposition. Since the objective in GAMIL is to be minimized, so the eigen-vectors corresponding to the smallest eigenvalues are used in the projection matrix. After getting the optimal projection matrix Ω^{opt}, we can assign any image features to the codewords in the embedding space.

3 Experiments

3.1 Experimental setup

The experimental setup of this work substantially follows Lazebnik et al. [15]. We randomly pick a number of images per class for training, and the left are for testing. In order to get reliable results, each experiment is repeated for 10 times (otherwise notice).

To describe an image, we use a grid-based method to extract the dense sift features. The dense SIIF features [16] are extracted on a 16×16 pixel patches sampled every 8 pixels. For our GAMIL method, the codebook is generated in the embedding space by K-means algorithm.

For classification, we use the SVM with a histogram intersection kernel. The libSVM toolbox [5] is adopted to make a one-versus-one approach for multi-class classification. A 10-fold cross-validation procedure is implemented on the training set to find an optimal parameter for SVM.

3.2 Multiview human body categorization

Before starting the evaluations on benchmark dataset, we would like first to publish our own dataset on Multiview Human Bodies (MHB). We have setup a multi-camera 3D studio to capture multi-view images for human actors. Compared with natural image datasets, our dataset could better highlight the manifold property among data. This multi-view dataset consists of images from 20 cameras which are evenly placed on a ring. Our dataset contains
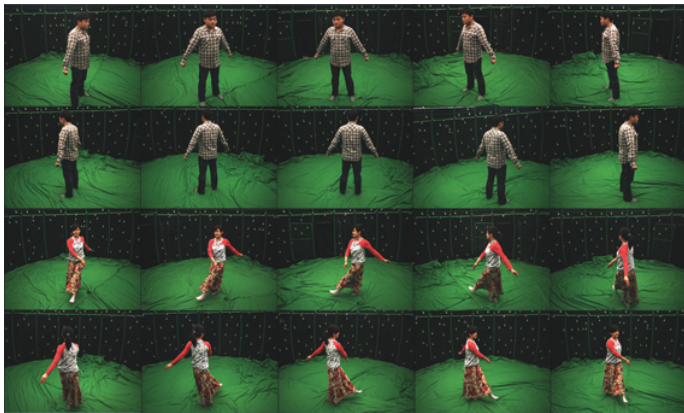
Figure 2: The images in the Multiview human body dataset.

8 actors/actress. In Fig. 2, we show one boy and one girl as demos. For categorization, a rectangle which contains the whole human body is extracted from each image.

In this dataset, we will verify why manifold properties can be emphasized by the proposed GAMIL model. On the manifold, we test two widely used graph metric, i.e. geodesic distance [12] (GD) and commute time [17] (CT). We do not use other manifold similarity e.g. locality Euclidean distance or locality linearity, since these similarity cannot reveal the relationship between unconnected nodes. First, we will only use graph assignment method while drop the information theoretic term. In a nutshell, we set $\alpha = 0$ in (2) only to verify the manifold assignment ability. We compare it with the hard and soft assignment methods in Euclidean space. For this dataset, 5 images for each person are used to train the GAMIL model and the rest 15 images are for test. The dimensionality of projection matrix Ω^T is fixed as 128×50, and size of the vocabulary is 200. Training and testing procedures are randomly repeated for 50 times and the average results are reported in the first row in Table.1.

Table 1: Image categorization results on the MHB dataset

	CT	GD	Hard	Soft
$\alpha = 0$	88.3	89.2	85.8	86.1
$\alpha = 1$	89.2	90.8	-	-

Then, we consider taking the mutual information loss term into the optimization and set $\alpha = 1.0$ in the GAMIL model. The corresponding results are provided in the second row of the table.

From the results, our GAMIL model outperforms the results of hard and soft assignment methods that use Euclidean similarity. By incorporating the manifold metric into the assignment, we get more than 3% improvements than both the hard and soft method. This point can be verified from the first row in Table.1. Moreover, the classification can be further improved by considering the mutual information loss term into the optimization. The highest classification rate (90.8%) was obtained by GAMIL model with the geodesic distance.

For the two graph metric, it is concluded that geodesic distance slightly outperforms the commute time. But, one significant advantage of commute time is due to its computational

efficiency. According to [17], the calculation of commute time is fast which only requires to solve a general inverse problem. But the computational costs of the geodesic distance required are quite heavy. Therefore, to trade off the effectiveness and efficiency, we choose the commute time as the graph metric in the following experiments on the benchmark datasets.

3.3   Results on benchmarks

In this part, we test our algorithm on two benchmark datasets. We report our GAMIL algorithm on Scene-15 [10] and Caltech-101[11] databases. In Scene-15 dataset, we use 100 images per category as training samples and the rests are for test. For Caltech-101 dataset, 15 images in each category are used for training and the rests are for test. To learn the GAMIL model, we also fix the projection dimension of Ω as 128×50. Both the codeword selection and assignments are implemented in the embedding space. For these two datasets, the experiments are repeated for 10 times and the average classification results are recorded. To improve the classification accuracy, following the suggestions in [15], we use a 2-level pyramid. For each pyramid level, the codewords number are fixed as 200. The nonlinear metric used in the GAMIL model is the commute time.

For comparison purpose, our GAMIL model will be compared with assignment algorithms with the same experimental setup. Besides, we will also compare our method with other image categorization methods that do not follow the way of codeword assignment. We will compare GAMIL with sparse coding method [19] and info-loss clustering [14]. The categorization accuracies are reported in Table.2.

Table 2: The comparisons of GAMIL model with other state-of-the-arts on two benchmarks

Algorithms	Scene-15	Caltech-101
Hard	76.3	56.4
Soft	78.2	59.5
GAMIL	**80.7**	64.3
Info-loss[14]	74.7	-
Sparse coding[19]	80.3	**67.0**

In the table, the first three rows report the results of three codeword assignment algorithms. These three results are obtained with the same experimental setup that we use the same codewords and the same parameters to conduct the experiments. Our reproducing results on hard and soft results is also similar to the reported result in [19]. From these three assignment results, our GAMIL achieves the best performance on both the two benchmarks.

We also compare the GAMIL with info-loss clustering method [14]. In [14], Lazebnik *et al.* proposed a general co-clustering method based on mutual information loss strategy. The co-clustering algorithm is applied to codewords generation for image categorization. Although both their works and ours use the mutual information loss criterion, the goals are definitely different. [14] focus on clustering while we want to make assignment. From the comparison on scene-15 dataset, our GAMIL outperforms the result reported in [14]. But the main contribution of [14] is not specific for scene categorization. It is a machine learning paper and scene categorization is just an application. However, we still pit them as a comparison here since both of us use mutual information loss criterion.

Finally, we compare GAMIL model with the sparse coding methods. In the Scene-15 dataset, our GAMIL slightly outperforms the results of sparse coding. In the large Caltech-

101 dataset, the sparse coding method achieves the highest accuracy. Nevertheless, we will argue that the time costs of GAMIL and sparse coding are not in a same order of magnitude. To assign each feature to the codewords, GAMIL just needs to calculate inner products which can be finished in almost real time. However, sparse coding requires to solve ℓ_1 minimization for each feature on the image. Therefore, sparse coding is far more computational heavy than GAMIL. GAMIL can produce comparable good result is less time.

4 Conclusions and future works

This paper presents a codeword assignment method to generate visual vectors for image categorization. We contribute on mainly two points: 1) we use a graph structure to model the massive image features and 2) the proposed algorithm takes the label information of the image features during assignment. To the best of our knowledge, this is the first time to assign codewords via a manifold way and it is also the first time to consider discriminative information into assignment. Owing to these two critical criterions, our algorithm outperforms most other assignment methods on both the MHB dataset and two benchmarks.

However, there are also some points that deserve future justifications. First, for the mutual information estimation, we use the non-differentiable Epanechnickov kernel to estimate the probability density of the feature distribution. Such a kernel and its corresponding approximations enable us to directly obtain the closed-form solution to the GAMIL optimization. However, this solution is only an approximation to the original problem. In our future work, we will consider using a continuous kernel, *e.g.* Gaussian kernels to model the feature probability and try to solve the optimization exactly via alternating methods. Besides, the experiments discussed in this paper are almost on natural images. We will generalize the effectiveness of the proposed method to more complicated image categorization applications, *e.g.* on medical images and the SAR images.

5 Acknowledgement

References

[1] M. Belkin and P. Niyogi. Laplacian eigenmaps and spectral techniques for embedding and clustering. In *NIPS*, 2001.

[2] D. M. Blei, A. Y. Ng, and M. I. Jordan. Latent dirichlet allocation. In *JMLR*, 2003.

[3] A. Bosch, A. Zisserman, and X. Muoz. Scene classification using a hybrid generative/discriminative approach. *TPAMI, IEEE*.

[4] Y.L. Boureau, F. Bach, Y. LeCun, and J. Ponce. Learning weights for codebook in image classification and retrieval. In *CVPR*, 2010.

[5] Chihchung Chang and Chihjen Lin. Libsvm: A library for support vector machines.

[6] T.M. Cover and J.A. Thomas. Elements of information theory, second ed. *Wiley*, 2006.

[7] Y. Deng, Q. Dai, and Z. Zhang. Graph lalapce for occluded face completion and recognition. *Image Processing, IEEE Transactions on*, (99), 2011.

[8] Yue Deng, Qionghai Dai, Ruiping Wang, and Zengke Zhang. Commute time guided transformation for feature extraction. In *CVIU*, 2012.

[9] V. A. Epanechnikov. Nonparametric estimation of a multidimensional probability density. *Theory of Probability and its Applications*, 1969.

[10] L. Fei-Fei and P. Perona. A bayesian hierarchical model for learning natural scene categories. In *CVPR*, 2005.

[11] Li Fei-Fei, R. Fergus, and P. Perona. Learning generative visual models from few training examples: An incremental bayesian approach tested on 101 object categories. In *CVPR*, 2004.

[12] Vin de Silva Joshua B. Tenenbaum and John C. Langford. A global geometric framework for nonlinear dimensionality reduction. *Science*, 2000.

[13] F. Jurie and B. Triggs. Creating efficient codebooks for visual recognition. In *ICCV*, 2005.

[14] S. Lazebnik and M. Raginsky. Supervised learning of quantizer codebooks by information loss minimization. *TPAMI*, 2009.

[15] S. Lazebnik, C. Schmid, and J. Ponce. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In *CVPR*, 2006.

[16] D.G. Lowe. Object recognition from local scale-invariant features. In *ICCV*, 1999.

[17] Huaijun Qiu and E.R. Hancock. Clustering and embedding using commute times. In *TPAMI*, 2007.

[18] Jan C. van Gemert, Cor J. Veenman, Arnold W.M. Smeulders, and Jan Mark Geusebroek. Visual word ambiguity. In *TPAMI*, 2010.

[19] Jianchao Yang, Kai Yu, Yihong Gong, and T. Huang. Linear spatial pyramid matching using sparse coding for image classification. In *CVPR*, 2009.