

Clustering Using Class Specific Hyper Graphs[★]

Shengping Xia^{1,2,★★} and Edwin R. Hancock²

¹ ATR Lab, School of Electronic Science and Engineering, National University of Defense Technology, Changsha, Hunan, P.R. China 410073

² Department of Computer Science, University of York, York YO1 5DD, UK

Abstract. The aim in this paper is to develop a method for clustering together image views of the same object class. Local invariant feature methods, such as SIFT, have been proven effective for image clustering. However, they have made either relatively little use or too complex use of geometric constraints and are confounded when the detected features are superabundant. Here we make two contributions aimed at overcoming these problems. First, we rank the SIFT points (R-SIFT) using visual saliency. Second, we use the reduced set of R-SIFT features to construct a specific hyper graph (CSHG) model of holistic-structure. Based on the CSHG model, a two stage clustering method is proposed. In which, images are clustered according to the pairwise similarity of the graphs, which is a combination of the traditional similarity of local invariant feature vectors and the geometric similarity between two graphs. This method comprehensively utilizes both SIFT and geometric constraints, and hence combines both global and local information. Experiments reveal that the method gives excellent clustering performance.

1 Introduction

Clustering uses features to organize data in a number of groups. One can use clustering to organize large collections of images. In this paper we aim to form clusters of images that are similar - similar in semantics as well as similar in visual appearance. Generally, there are two modalities for image clustering. The first is appearance feature based, the second is graph based.

1.1 Appearance Based Clustering

Traditionally, images are represented using feature vectors extracted from global appearance. Many kinds of central clustering methods have been proposed by using these feature vectors. These clustering methods can be categorized into partitioning methods, hierarchical methods, density-based methods, grid-based methods. The major difficulty in applying such techniques lies in obtaining feature vectors, namely building appropriate representations of images. For high dimensional vectors, the more scattered the data distribution is, the poorer the performance of clustering. A common problem is that the different views of an object present very different appearance. An example is furnished by the images of the cat from the COIL100 data set shown in Fig.1. More examples of

* Supported by ATR Key Lab Foundation project 91408001020603.

★★ Corresponding author.



Fig. 1. Example images of small cat in COIL100 data set

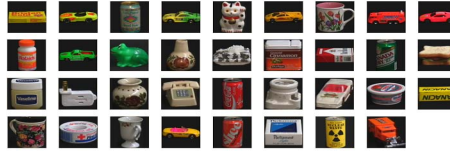


Fig. 2. Example images of objects from COIL 100

various objects in this data set are shown in Fig.2. A global appearance based feature vector might be inappropriate in these cases. It is difficult to build a global feature for clustering these into their individual categories by using global appearance based feature vector methods.

Hence, local invariant feature based methods[3][6][8], e.g. SIFT, PCA-SIFT, GLOH and SURF, have recently been developed and have been successfully and widely used for object classification [3][7], scene classification [2][5] and video clustering or retrieval [9]. According to a comprehensive test on several large image databases, SIFT appears to be the most appealing descriptor for practical uses, and has become a standard of comparison[7][2].

Local features play the role of "visual words" predictive of certain "topic" or object class. For example, an eye is highly predictive of a face being present in the image. If our visual dictionary contains words that are sufficiently discriminative when taken individually, then it is possible to achieve a high degree of success for whole image classification, i.e. identification of the object class contained in the image without attempting to segment or localize that object, simply by looking which visual words/features are present, regardless of their spatial layout.

The success of the orderless feature model may be explained with the help of an analogy to bag-of-words models for text document classification [10]. From this point, some Bayesian text models, such as probabilistic Latent Semantic Analysis (pLSA) [5], Latent Dirichlet Analysis (LDA)[9] and transformed Dirichlet Process (TDA)[11] have been adapted and used to model scene categories. Where images are represented by the frequency of "visual words". The *Bag-of-Features*(BOF) models, which represent an image as an orderless collection of local features, have recently demonstrated impressive levels of performance[7]. In all of these methods central clustering methods can be utilized to find distinguishable visual features.

However, there are two basic drawbacks for the local invariant feature based methods. Firstly, in most cases too many orderless feature points are extracted from each image. As a result the exhaustive search over features is time consuming and too many false positives are produced. One way of overcoming these problems is to use visual saliency [4] to select a subset of the salient points for analysis. Secondly, in the BOF model, local features play the role of "visual words" in an orderless way. This representation overlooks geometric and structural constraints on feature arrangement. This is

likely to compromise recognition, clustering or retrieval performance. A more versatile and expressive tool for representing structured data are attributed graphs (hereinafter simply referred to as graphs). Hence there has been a recent interest in modeling and learning structural relationships using graph based method.

1.2 Graph Based Clustering

Graph clustering is an important, yet relatively under-researched topic in computer vision [14][15]. The importance of the topic stems from the fact that it can be used as a tool for learning the class-structure of data abstracted in terms of relational graphs. Problems of this sort are posed by a multitude of unsupervised learning tasks in knowledge engineering, pattern recognition and computer vision. The process can be used to structure large data-bases of relational models or to learn equivalence classes. Broadly speaking, there are two approaches to the problem.

The first of these is to use center clustering methods. Here a class prototype is maintained, and the aim is to find clusters of graphs that are close to the prototype. This prototype may be either an explicit representation of full graph-structure or a proxy representation based on features that succinctly express the graph characteristics. The learning of explicit cluster prototypes is a difficult task due to (1) missing powerful analytical techniques for data analysis and (2) high computational complexity when measuring structural similarities between graphs.

Considering the former problem, it is not clear how to embed dynamic data structures such as graphs into a vector space without loss of structural information. Therefore it is not straightforward to provide well defined characterizations of basic concepts such as the mean or variance of a set of graphs. The latter problem of measuring the structural similarity is more generally referred to as the graph matching problem. However, taking into account the practical importance, investigations on clustering of graphs has been recently revitalized[19][20] [21][22]. However, It is well known that the graph matching problem is NP-complete[18]. These algorithms are computationally intensive and have hence only been demonstrated on rather small graphs.

The lack of analytical methods in the domain of graphs led to the formulation of clustering problem as pairwise clustering [23][17]. The process of pairwise clustering is somewhat different to the more familiar one of central clustering. Whereas central clustering aims to characterize cluster membership using the cluster mean and variance, in pairwise clustering it is the relational similarity of pairs of images which is used to establish cluster membership. Although less well studied than central clustering, there has recently been renewed interest in pairwise clustering.

Recently, Torsello and Hancock [12][17] have constructed the class-prototype through tree-union and have performed clustering using a mixture of tree-unions controlled by a description length criterion [16]. These ideas are taken further by Escolano and Lozano [13] who extended the methodology to graphs rather than trees, and used an EM algorithm for clustering. When posed in a pairwise setting, the graph-clustering problem involves measuring the pairwise distances of the graphs. Provided that the data set consists of N samples, pairwise clustering requires the solution of $O(N^2)$ NP-complete graph matching problems which might be intractable for large data sets. Thus, clustering large

set of graphs is still widely unexplored and one of the most challenging problems in structural pattern recognition.

To summarize, local invariant feature based methods, which represent an image as an orderless collection of local features, have demonstrated impressive performance, though overlooking geometric and structural constraints upon feature arrangement. While the graph based method provides an elegant way of encapsulating within-class structural variations, it is relatively impoverished in terms of the attribute information encoded. For this reason, we have proposed a class specific hyper graph (CSHG) model to combine both SIFT and geometric constraints in a companion paper [1] submitted to the same workshop. The CSHG model includes two basic parts.

Firstly, we rank SIFT-points using visual saliency for each image or ROI. The highly ranked salient points are used to construct attributed Delaunay graphs. Secondly, the CSHG model is used to develop a prototype structure through the merging of salient graphs. This structure encapsulates both variations in structure, and edge and node attributes, including the information contained in the ranked local invariant feature. In the CSHG model, a new method on measuring the pairwise graph similarity is proposed.

Based on the CSHG model, we develop a new two stage clustering method in this paper. In the first stage a sample of the graphs in a trained CSHG model are clustered to form initial clusters using constraints based on the pairwise similarity. The pairwise graph matching is sketched using the SIFT matching, which is not NP-complete. Because it utilizes an RSOM [24] (a SOM-based incremental clustering tree) central clustering method in learning a CSHG model, the clustering procedure requires the solution of $O(N \lg N)$ pairwise SIFT matching problem. Hence our model can be used to deal with large datasets. The remaining graphs are either recognized as one of the existing clusters or rejected from consideration.

The structure of this paper is as follows. In Section 2 we present some preliminaries. In Section 3 we describe the framework for modeling and clustering. Section 4 provides some experimental results. This paper finishes in Section 5 with the conclusions.

2 Preliminaries

See [1] Section 2.

3 CSHG Modeling and Clustering

3.1 CSHG Modeling

See [1] Section 3.

3.2 Clustering Based on a Trained CSHG

Suppose a CSHG model has been trained from a batch of images. In the CSHG model, the relationships of the graphs $ONNG_{K\tau}^*\{G_l\}$ have been encoded and will be updated with additional graphs. The number of the items in $ONNG_{K\tau}^*\{G_l\}$ is largely influenced by the threshold \mathcal{R}_τ^* . In our system, \mathcal{R}_τ^* is set as 5, and the number K of the highly

ranked graphs in $ONNG_{K\tau}^*\{G_l\}$ is set as 14. We expand the definition of the siblings of a graph (Def. 2.7 in [1]) to include the role of \mathcal{R}_τ^* as follows.

Def. 3.1 Siblings of a Graph(\mathcal{S}): Siblings of G_l is re-defined as:

$$\mathcal{S}\{G_l\} = \{G_q \mid CMG(G_l, G_q) = 1, G_q \in ONNG_{K\tau}^*\{G_l\} \triangleq \mathcal{S}_{\mathcal{R}_\tau^*}\{G_l\}\}. \quad (1)$$

For any graph $G_l \in$ CSHG, we can obtain its siblings within $ONNG_{K\tau}^*\{G_l\}$. For each graph $G_q \in \mathcal{S}\{G_l\}$, their corresponding siblings can also be obtained. In this way, we can iteratively obtain a series of graphs according to the sibling relationships. This is called a family tree of a graph G_l .

Def. 3.2 Family Tree of a Graph within g Generations(FTOG): Given a CSHG model, the family tree of a graph(FTOG) G_l within g generations, denoted as $\mathcal{F}\{G_l, g\}$, is defined as:

$$\mathcal{F}\{G_l, g\} = \mathcal{F}\{G_l, g - 1\} \bigcup_{G_q \in \mathcal{F}\{G_l, g-1\}} \mathcal{S}\{G_q\}. \quad (2)$$

where G_l is non-redundant graph of the CSHG model, and g is the number of the generations of a family tree. If $g = 1$, $\mathcal{F}\{G_l, 1\} = \mathcal{F}\{G_l, 0\} \cup \mathcal{S}\{G_l\}$ and $\mathcal{F}\{G_l, 0\} = \{G_l\}$. If $\lim_{g \rightarrow \infty} \mathcal{F}\{G_l, g\} \triangleq \mathcal{F}\{G_l, \infty\} =$ CSHG, CSHG is called all-linked.

Def. 3.3 Clan of a CSHG: Given a CSHG model, $\exists G_{l_i}, i = 1, 2, \dots, \mathbb{C}$, subject to $\mathcal{F}\{G_{l_i}, g\} \not\subseteq \mathcal{F}\{G_{l_j}, g\}, i \neq j, i, j \in \{1, 2, \dots, \mathbb{C}\}$, and for each family tree, there are at least N_τ graph members, e.g. $N_\tau = 3$. A clan of CSHG within g generations is defined as

$$\mathcal{C}\{g\} = \{ \mathcal{F}\{G_{l_i}, g\}, i = 1, 2, \dots, \mathbb{C} \}. \quad (3)$$

We can regard a family tree of a graph within g generations ($\mathcal{F}\{G_l, g\}$) as a cluster if the number of graphs is greater than or equal to N_τ . The clan of a CSHG is a cluster separation of the whole graph set. Corresponding to \mathcal{R}_τ^* , $\mathcal{F}\{G_l, g\}$ and $\mathcal{C}\{g\}$ are separately denoted as $\mathcal{F}_{\mathcal{R}_\tau^*}\{G_l, g\}$ and $\mathcal{C}_{\mathcal{R}_\tau^*}\{g\}$.

If $\mathcal{R}_\tau^* = 0$, it is easy to obtain that $\mathcal{F}\{G_l, \infty\} =$ CSHG. If \mathcal{R}_τ^* is set as the number of the nodes in a graph, namely the selected number of R-SIFT points for an image, there might always be only one item in a family tree $\mathcal{F}\{G_l, g\}$. Especially, in the training stage, the threshold $\mathcal{R}_{\tau_0}^*$ is set to 20 to check for the redundancy. Consequently, the selection of the value of \mathcal{R}_τ^* is a key point for a CSHG model. Here we will continue to present a two stage clustering method.

Firstly, if we set a relatively large \mathcal{R}_τ^* , e.g. $\mathcal{R}_{\tau_1}^* = 12, \forall G_l \in$ CSHG, $\mathcal{F}_{\mathcal{R}_{\tau_1}^*}\{G_l, \infty\}$ normally converges to a stable graph set and each $\mathcal{F}_{\mathcal{R}_{\tau_1}^*}\{G_l, \infty\}$ is regarded as the seed members of a cluster. In this way, a clan of a CSHG model can be obtained, denoted as $\mathcal{C}_{\mathcal{R}_{\tau_1}^*}\{g\}$. Suppose there are N_c family trees obtained and each of which are uniquely labeled. These are the initial seed clusters in a CSHG model, which is denoted as follows:

$$\mathcal{C}_{\mathcal{R}_{\tau_1}^*} = \{ \mathcal{F}_{\mathcal{R}_{\tau_1}^*}\{G_l, \infty\} \triangleq \{ \mathcal{F}_l \mid l \in \{L_1, L_2, \dots, L_{N_c}\} \}. \quad (4)$$

After the first stage, generally there are still some graphs which have not been clustered. As a second stage, the un-clustered graphs will be either recognized as one of the clusters in $\mathcal{C}_{\mathcal{R}_{\tau_1}^*}$ or discarded from being clustered using the following method.

For an un-clustered graph G_l ,

$$\text{if } \exists \text{ only one clustered graph } G_q, \text{ s.t. } \mathcal{R}^*(G_l, G_q) \geq \mathcal{R}_{\tau_2}^*, \text{ then } L_{G_l} = L_{G_q}. \quad (5)$$

where $G_q \in ONNG_{K\tau}^*\{G_l\}$, L_{G_l} is the cluster label of G_l , and L_{G_q} is the cluster label of G_q . Otherwise, G_l remains un-clustered.

Furthermore, for any graph G_l of the remaining un-clustered graphs, we can obtain its K -clustered nearest neighbor graphs $ONNG_{K\tau}^*\{G_l\}$ with a threshold $\mathcal{R}_{\tau_3}^*$, e.g. $\mathcal{R}_{\tau_3}^* = 5$.

$$ONNG_{K\tau_3}^*\{G_l\} = \{G_q \mid \mathcal{R}^*(G_l, G_q) \geq \mathcal{R}_{\tau_3}^*, \text{ and } \exists \mathcal{F}_{L_i} \text{ s.t. } G_q \in \mathcal{F}_{L_i}\}. \quad (6)$$

According to the K graphs G_{q_i} , $i = 1, 2, \dots, K$, we use a weighted voting method to determine an FTOG $\mathcal{F}_{L_{i0}}$ that the graph G_l belongs to.

$$L_{i0} = \arg \max_{\mathcal{F}_{L_i}} \{ \overline{W}^* \{ \mathcal{F}_{L_i} \} = \frac{W^* \{ \mathcal{F}_{L_i} \}}{\sum_{\mathcal{F}_{L_i}} W^* \{ \mathcal{F}_{L_i} \}} \geq 2/3 \}. \quad (7)$$

where $W^* \{ \mathcal{F}_{L_i} \} = \sum_{G_q \in \mathcal{F}_{L_i}} \{ \mathcal{R}^*(G_l, G_q) \}$, $\mathcal{F}_{L_i} \in \mathcal{C}_{\mathcal{R}_{\tau_1}^*}$ and $\overline{W}^* \{ \mathcal{F}_{L_i} \}$ satisfy the following conditions

$$E_{\overline{W}^*} = - \sum_{\mathcal{F}_{L_i}} \{ \overline{W}^* \{ \mathcal{F}_{L_i} \} \cdot \ln \{ \overline{W}^* \{ \mathcal{F}_{L_i} \} \} \} < E_{\tau}. \quad (8)$$

where E_{τ} is set as 0.4 according to the experiments. We borrowed this equation from the information entropy theorem. If no such family tree exists, or if $\forall G_q \in \text{CSHG}$, $\mathcal{R}^*(G_l, G_q) < \mathcal{R}_{\tau_3}^*$, G_l is abstained from being clustered, which is also called a rejection.

In an incremental clustering process, for a new graph G_l , if Equation (9) holds, G_l will be incremented to the CSHG model. For clustering, $\mathcal{F}_{L_{i1}}$ and $\mathcal{F}_{L_{i2}}$ are merged into the family tree $\mathcal{F}_{\mathcal{R}_{\tau_1}^*}\{G_l, \infty\}$ of graph G_l . G_l is called a joint graph of the cluster-merging.

$$\exists G_{q_1}, G_{q_2} \in \mathcal{S}_{\mathcal{R}_{\tau_1}^*}\{G_l\}, \text{ s.t. } G_{q_1} \in \mathcal{F}_{L_{i1}}, G_{q_2} \in \mathcal{F}_{L_{i2}}, \text{ and } \mathcal{F}_{L_{i1}}, \mathcal{F}_{L_{i2}} \in \mathcal{C}_{\mathcal{R}_{\tau_1}^*} \quad (9)$$

where $\mathcal{S}_{\mathcal{R}_{\tau_1}^*}\{G_l\} = \{G_q \mid \text{CMG}(G_l, G_q) = 1, G_q \in ONNG_{K\tau_1}^*\{G_l\}\}$ are the clustered siblings with threshold $\mathcal{R}_{\tau_1}^*$ of graph G_l , and G_q is a graph that has been clustered in the S1 of Algorithm 1.

To summarize, the principal steps for CSHG based clustering are listed in Table 1, referred as Algorithm 1.

Table 1. CSHG based matching and incremental training

Algorithm 1: CSHG Based Clustering	
Input: a CSHG model	
Output: updated CSHG model, Clan of the CSHG model	
S1	Obtaining kernel clusters $\mathcal{C}_{\mathcal{R}_{\tau_1}^*}$, namely the clan of the CSHG with threshold $\mathcal{R}_{\tau_1}^*$;
S2	Recognizing un-clustered graphs G_l according to Equation (5).
S3	Recognizing un-clustered graphs G_l according to Equation (7). <End>

4 Experiments

We use all 72 images of 50 objects from the COIL100 (The COIL set contains objects placed on a black background with no clutter, the view angle changes from 0° to 355° with a step size of 5°, and there are 72 images for each object.). For each image, 6 noised images are generated with multiple noise-parameters. There are 25200 images. We aim to cluster these images into corresponding classes.

For this clustering task, firstly, the 3600 initial noise free images are modeled with a CSHG model and then clustered. Secondly, the remaining 21600 noisy images are incrementally modeled and clustered using the CSHG model obtained from the noise free images.

Table 2. Clustering of 3600 initial images of 50 objects in COIL100 dataset

ID of Cluster	ID of Object in COIL	Num. of Ini. Images	Num. of 1 st stage Clustered	Num. of Finally Abstained
31 clusters	1,6,7,8,14,15,16,19,24,28,30,37,39,46,49,53,55,58,60,62,66,69,72,79,81,87,89,91,93,96	72 images per calss	totally 2232	0
32	13	72	33	3
33	13	72	28	3
34	23	72	71	1
35	26,71	144	133	6
36	27	72	70	2
37	29,32	144	144	0
38	31,54	144	144	0
39	44,59	144	144	0
40	57	72	71	1
41	61,92	144	144	0
42	67,84,98	216	216	0
43	90	72	36	2
44	90	72	29	2
45	99	72	69	1
Total		3600	3564	21
Percent		100%	99.42%	0.58%

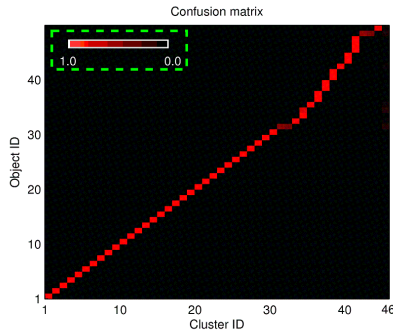


Fig. 3. Confusion matrix of clustering of 3600 initial images of 50 objects from the COIL100 dataset

For the 50 objects, the 3600 initial images are clustered into 45 clusters as shown in Table 2. The confusion matrix for the clustering result is shown in Fig.3, where the Y-axis is the order of the object ID in COIL 100. The X-axis is the cluster ID and 46 corresponds to the rejected images.

There are 35 clusters that separately correspond to 35 different objects without confusion. Example images from these 35 object-classes are shown in Fig. 2. In addition, 13 objects are assigned to 6 clusters as shown in Fig. 4. Each group of objects in Fig. 4 (A)(B)(C)(E) are actually identical in shape but color. Since it only uses gray scale information in SIFT, our method fails in this case. We hence regard these objects in the four groups as being correctly clustered according to shape.

For 2 groups of overlapping-clustered objects shown in Fig. 4(D)(F), there is only one joint graph. Finally 2 objects are assigned into 4 clusters, shown in Fig. 5. The graphs of the front view images are very different from the graphs of the back view images, while the graphs of the side view images are too simple to be joint graphs of the two clusters of the front and the back. However these clusters will be merged into one object in the supervised learning stage while assigning a same name to the front cluster and the back cluster. Finally, 0.42% of all these initial images are discarded from being clustered, namely the rejection rate is 0.42%.



Fig. 4. 6 groups of objects are overlapping-clustered into 6 clusters



Fig. 5. Identical objects are assigned into separated clusters

Table 3. Clustering of various images with multiple threshold restriction

Image source	$\mathcal{R}^* \geq \mathcal{R}_{\tau_0}^*$ ($\mathcal{R}_{\tau_0}^* = 20$)	$\mathcal{R}^* \geq \mathcal{R}_{\tau_1}^*$ ($\mathcal{R}_{\tau_1}^* = 12$)	$\mathcal{R}^* \geq \mathcal{R}_{\tau_2}^*$ ($\mathcal{R}_{\tau_2}^* = 8$)	$\mathcal{R}^* \geq \mathcal{R}_{\tau_3}^*$ ($\mathcal{R}_{\tau_3}^* = 5$)	$\mathcal{R}^* < \mathcal{R}_{\tau_3}^*$ ($\mathcal{R}_{\tau_3}^* = 5$)
Initial	40.27%	84.32%	94.00%	99.42%	0.58%
Noise0.01	31.86%	79.55%	92.61%	99.17%	0.83%
Noise0.03	22.55%	72.51%	90.53%	98.94%	1.06%
Noise0.05	15.21%	59.56%	88.58%	98.65%	1.35%
Noise0.07	9.31%	38.68%	86.83%	98.36%	1.54%
Noise0.09	6.17%	27.75%	84.61%	98.25%	1.75%
Noise0.11	3.37%	19.16%	81.32%	97.80%	2.20%

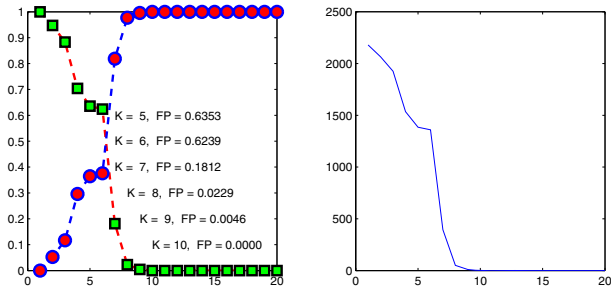


Fig. 6. Clustering test result of 2300 various negative sample images. These images are randomly selected from three data sets: Caltech-101, Caltech-256 and Google images. We use the CSHG model of 50 objects obtained above to test how these negative sample images are clustered. For the left sub-figure, Y-axis corresponds to clustering rate or rejection rate and X-axis corresponds to the threshold of \mathcal{R}_τ^* . The green-face-square marker corresponds to the clustering rate, the red-face-circle marker corresponds to the rejection rate. For the right sub-figure, the plot is the number of images with different threshold of \mathcal{R}_τ^* .

Our second experiment involves adding multiple kinds of noise to each image. The noise processes consists of a mixture of salt and pepper noise, speckle noise, and Gaussian noise. The fraction of added noise increases from 0.01 to 0.11 with a step 0.02. For each case, 3600 noisy images are generated. Based on the CSHG model and clustering result obtained from the initial images, the noisy images are incrementally clustered. The results are shown in Table 4. The average correct clustering rate for all these 21600 images is 98.45%, The remaining 1.55% were rejected and were not correctly clustered. This is a significant property of our method. If a graph of an image is not similar enough to the graphs of the CSHG model, it should not be clustered.

After clustering, we have constructed a compact multiple-object example CSHG model for all of these 50 objects. Hence we have also explored how negative images are recognized using this CSHG model. We randomly selected 2300 images from three data sets: Caltech-101, Caltech-256 and Google images. In Fig. 6, we investigate the false classification rate and correct rejection rate with varying threshold \mathcal{R}_τ^* for these negative images. From Fig. 6, there are 2.29% of all the 2300 images that are false clustered or recognized according to S1 of Algorithm 1. However the average false clustering rate for each class is only 0.046%. And there are 36.37% of the images that are correctly rejected according to the restriction that is $\mathcal{R}^* < \mathcal{R}_{\tau_3}^*$. The remaining 61.34% are rejected according to S3 of Algorithm 1 since Equation (7) poses a very strict condition especially on the negative samples.

5 Conclusions

In paper [1], we demonstrated how to construct a CSHG model. This model comprehensively captures both global and local information through utilizing both SIFT and geometric constraints. Based on this model, we proposed a two stage generalized pairwise clustering method, which combines the recognition process into the clustering

process, to capture a more holistic and compact representation of the objects. Experiments reveal that the method gives excellent clustering performance. The model can be extended to object recognition and content based image retrieval.

References

1. Xia, S.P., Hancock, E.R.: 3D Object Recognition Using Hyper-Graphs and Ranked Local Invariant Features. In: S+SSPR 2008 (to appear, 2008)
2. Bosch, A., Zisserman, A., Muoz, X.: Scene Classification Using a Hybrid Generative/Discriminative Approach. *IEEE Trans. PAMI* 30(4), 1–16 (2008)
3. Lowe, D.G.: Distinctive image features from scale-invariant key points. *IJCV* 60(2), 91–110 (2004)
4. Elazary, L., Itti, L.: Interesting objects are visually salient. *Journal of Vision* 8(3), 1–15 (2008)
5. Li, F.F., Perona, P.: A Bayesian hierarchical model for learning natural scene categories. *CVPR* 2, 524–531 (2005)
6. Bay, H., Tuytelaars, T., Gool, L.V.: SURF: Speeded Up Robust Features. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) *ECCV 2006, Part I. LNCS*, vol. 3951, pp. 404–417. Springer, Heidelberg (2006)
7. Zhang, J., Marszablek, M., Lazebnik, S., Schmid, C.: Local Features and Kernels for Classification of Texture and Object Categories. *IJCV* 73(2), 213–238 (2007)
8. Yan, K., Sukthankar, R.: PCA-SIFT: A more distinctive representation for local image descriptors. *CVPR* 2, 506–513 (2004)
9. Sivic, J., Zisserman, A.: VideoGoogle: A text retrieval approach to object matching in videos. *ICCV* 2, 1470–1477 (2003)
10. Nigam, K., Lafferty, J., McCallum, A.: Using maximum entropy for text classification. In: *IJCAI Workshop on Machine Learning for Information Filtering*, pp. 61–67 (1999)
11. Sudderth, E.B., Torralba, A., Freeman, W.T., Willsky, A.S.: Describing Visual Scenes Using Transformed Objects and Parts. *Int. J. Comput. Vis.* 77, 291–330 (2008)
12. Torsello, A., Hancock, E.R.: Graph Embedding using Tree Edit Union. *Pattern Recognition* 40, 1393–1405 (2007)
13. Bonev, B., Escolano, F., Lozano, M.A., Suau, P., Cazorla, M.A., Aguilar, W.: Constellations and the Unsupervised Learning of Graphs. In: Escolano, F., Vento, M. (eds.) *GbrPR 2007. LNCS*, vol. 4538, pp. 340–350. Springer, Heidelberg (2007)
14. Rizzi, S.: Genetic operators for hierarchical graph clustering. *Pattern Recognition Letters* 19, 1293–1300 (1998)
15. Segen, J.: Learning graph models of shape. In: Laird, J. (ed.) *Proceedings of the Fifth International Conference on Machine Learning*, pp. 29–35 (1988)
16. Torsello, A., Hancock, E.R.: Learning shape-classes using a mixture of tree-unions. *IEEE Trans. PAMI* 28(6), 954–967 (2006)
17. Torsello, A., Robles-Kelly, A., Hancock, E.R.: Discovering Shape Classes using Tree Edit Distance and Pairwise Clustering. *International Journal of Computer Vision* 72, 259–285 (2007)
18. Garey, M., Johnson, D.: *Computers and intractability: A guide to the theory of NP-completeness*. W.H. Freeman and Company, New York (1979)
19. Günter, S., Bunke, H.: Self-organizing map for clustering in the graph domain. *Pattern Recognition Letters* 23, 401–417 (2002)
20. Lozano, M.A., Escolano, F.: Em algorithm for clustering an ensemble of graphs with comb matching. In: Rangarajan, A., Figueiredo, M.A.T., Zerubia, J. (eds.) *EMMCVPR 2003. LNCS*, vol. 2683, pp. 52–67. Springer, Heidelberg (2003)

21. Luo, B., Wilson, R.C., Hancock, E.R.: Spectral embedding of graphs. *Pattern Recognition* 36(10), 2213–2223 (2003)
22. Jain, B.J., Wysotzki, F.: Central Clustering of Attributed Graphs. *Machine Learning* 56, 169–207 (2004)
23. Hofmann, T., Buhmann, J.M.: Pairwise data clustering by deterministic annealing. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 19(1), 1–14 (1997)
24. Xia, S.P., Liu, J.J., Yuan, Z.T., Yu, H., Zhang, L.F., Yu, W.X.: Theory and Algorithm of Machine Learning Based on RSOM Tree Model. *ACTA Electronica sinica* 33(5), 937–944 (2005)