3D Object Recognition Using Hyper-Graphs and Ranked Local Invariant Features

Shengping Xia[1,2,⋆] and Edwin R. Hancock[2]

[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. Local invariant feature-based methods such as SIFT have been proven highly effective for object recognition. 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 class specific hyper graph (CSHG) which comprehensively utilizes local SIFT and global geometric constraints. Moreover, it efficiently captures multiple object appearance instances. We show how the CSHG can be learned from example images for objects of a particular class. Experiments reveal that the method gives excellent recognition performance, with a low false-positive rate.

1 Introduction

Recently, local invariant feature extraction (LIFE) methods[5] [8][10], such as SIFT, PCA-SIFT,GLOH and SURF, have proven successful and have been widely used for object recognition [5] [9], scene classification [1][2] [7] and video retrieval [12][13]. According to comprehensive experimental analysis on several large image databases[9], SIFT appears to be the most effective descriptor for practical uses, and has become a standard of comparison. SIFT is scale invariant, rotation invariant, and is also illumination invariant for a limited range of light source directions. However, there are two basic drawbacks to its use in object representation. Firstly, in most cases the features delivered are super-abundant and exceed the number that can be effectively used for the purposes of matching or recognition. As a result the exhaustive search of features is time consuming and too many false positive matches are often produced. One way of overcoming these problems is to use visual saliency [6] to select a subset of the features for analysis. Secondly, in the Bag-of-words[7] approach local features usually play the role of "visual words" that are predictive of a certain object class.

However, this representation overlooks both geometric and structural constraints on feature arrangement. By discarding this information recognition performance is compromised. A more versatile and expressive tool for representing structured data are attributed graphs (hereinafter simply referred to as graph). When recognition is attempted using graph-based abstractions, it is difficult to account for natural structural variations.

⋆ Corresponding author.

N. da Vitora Lobo et al. (Eds.): SSPR&SPR 2008, LNCS 5342, pp. 117–126, 2008.

One way of overcoming the problem is to construct a class prototype by merging example structures together. For instance, Torsello and Hancock [14] have constructed the class-prototype through tree-union and have performed clustering using a mixture of tree-unions controlled by a description length criterion [15]. These ideas are taken further by Escolano and Lozano [16] who extended the methodology to graphs rather than trees, and used an EM algorithm for clustering. However, although this work provides an elegant way of encapsulating within-class structural variations it is relatively impoverished in terms of the attribute information encoded. For this reason in this paper we turn to hypergraphs. Our aim is to develop a prototype structure through the merging of salient graphs. This structure encapsulates both variations in structure, and edge and node attributes.

In this paper we make two contributions to appearance based object recognition. Firstly, we rank SIFT-points using visual saliency and select only the most salient points to construct object models. Secondly, we use a class specific hypergraph (CSHG) to model objects compactly. The hypergraphs are based on multiple Delaunay graphs. Each of these is constructed from the selected feature points for a single prototype image of an object. In this way, the object models can be constructed with a minimum of object views.

2 Preliminaries

Before we detail our method, we commence by providing some preliminary definitions.

2.1 Graph and Class Specific Hyper Graph

Def. 2.1 ***Attributed Graph*** G [4]**:** An attributed graph is a 2-tuple $G = (V, E)$, where V is the (finite) set of vertices (also called nodes), $E \subseteq V \times V$ is the set of edges.

Def. 2.2 ***Spectral Eigenvalue Vector*** (S_G)**:** Let A be the adjacency matrix of G and D be the diagonal degree matrix. The spectral decomposition of the normalized Laplacian matrix $\hat{L} = D^{-1/2}(D - A)D^{-1/2}$ [4] is $\hat{L} = \Phi\Lambda\Phi^T$, where $\Lambda = diag(\lambda_1, \lambda_2, ..., \lambda_{\|G\|})$, $\lambda_1 \leq \lambda_2 \leq ... \leq \lambda_{\|G\|}$ is the diagonal matrix with the ordered eigenvalues as elements and $\Phi = (\phi_1, \phi_2, ..., \phi_{\|G\|})$ is the matrix with the correspondingly ordered eigenvectors as columns. If $e = (1, 1, 1, .., 1)^T$ is the all-ones-vector, then the ordered eigenvalues can be represented by the spectral vector $S_G = \Lambda e$.

Def. 2.3 ***Maximum Common Subgraph*** (MCS)[4]**:** G is a maximum common subgraph of G_i and G_j, denoted by $MCS(G_i, G_j)$, if it is a common subgraph of G_i and G_j and there is no other common subgraph having more nodes than G . We denote any subgraph of $MCS(G_i, G_j)$ as $CS(G_i, G_j)$.

Def. 2.4 ***Correct Matching of Graph*****(CMG):** $CMG(G_i, G_j), G_i, G_j \in HV$, is defined as:

$$CMG(G_i, G_j) = \begin{cases} 1, & \|MCS(G_i, G_j)\| \geq n_\tau, \quad (1) \\ 0, & else. \quad (1') \end{cases}$$

where n_τ is threshold on the number of correctly matched nodes. If $\|MCS(G_i, G_j)\| \geq n_\tau$, then the two graphs are a CMG pair.

Using ideas from basic graph theory [4] and hypergraph theory [3], we construct a class specific hyper graph (CSHG), which we use to capture structural variations for objects of a particular class.

Def. 2.5 ***Class Specific Hyper Graph*** **(CSHG):** A CSHG is defined as 2-tuple $CSHG = (HV, HE)$, where 1) the hyper vertex $HV = \{ G_i, i = 1, 2, ..., M \}$ is a finite set of graphs forming the nodes of the CSHG; 2) the hyper-edge HE is the edge set of the CSHG. There is an edge between two nodes G_i and G_j, denoted by $Edge(G_i,G_j)$, if and only if $CMG(G_i, G_j) = 1$.

Def. 2.6 ***Simple Class Specific Hyper Graph*****:** Given $G_i \in HV$, $\forall\, G_j\, (G_j \neq G_i) \in HV$, if $Edge(G_i, G_j) \equiv \emptyset$, G_i is isolated. If all G_i are isolated, then the graph is a simple CSHG.

Def. 2.7 ***Siblings of a Graph($\mathscr{S}$)*****:** The siblings of G_i are defined as $\mathscr{S}\{G_i\} = \{ G_j | CMG(G_i, G_j) = 1, G_j \in HV\}$. For a graph G_i, there may be more than 2 siblings, and any two siblings may not form a CMG pair.

Def. 2.8 ***A Path of CSHG*****:** A path between G_i and G_j is defined as a set of graphs

$$\mathscr{P}\{G_i, G_j\} = \{G_i, G_{j1}, G_{j2}, ...G_{jK}, G_j\}. \tag{2}$$

where $CMG\{ G_i , G_{j1} \}=1$, $CMG\{ G_j , G_{jK} \} =1$ and for any two consecutive graphs $CMG\{ G_{j(l)}, G_{j(l+1)} \} =1$, $l = 1, 2, ..., K - 1$. If $K = 0$, the path degenerates to a CMG pair.

Def. 2.9 ***Redundant Graph*****:** A graph G_t is redundant for an existing CSHG model if

$$\exists\, G_i, G_j, s.t. \mathscr{P}\{G_i, G_j\} \neq \emptyset, \mathscr{S}\{G_t\} \subseteq \bigcup \mathscr{S}\{G_l \mid G_l \in \mathscr{P}\{G_i, G_j\}\}. \tag{3}$$

The special case is that $\mathscr{S}\{G_t\} = \{G_i, G_j \} = \mathscr{P}\{ G_i, G_j \}$.

2.2 SIFT and Its Ranking for Visual Saliency

Local Invariant Feature Extraction. The SIFT algorithm is described in detail in [5]. Let $\overrightarrow{X}^t=(x_1^t, x_2^t)^T$ be the coordinates of the t-th SIFT point, $\overrightarrow{R}^t = (r^t, \alpha^t)^T$ be a vector with the corresponding gradient magnitudes and orientations as components, and $\overrightarrow{U}^t=(u_1^t, u_2^t, ..., u_n^t)^T$ be the descriptors of SIFT feature. The three vectors are concatenated to form a SIFT feature-vector $V^t = ((\overrightarrow{X}^t)^T, (\overrightarrow{R}^t)^T, (\overrightarrow{U}^t)^T)^T$ and the set of feature-vectors is $\mathcal{V} = \{V^t, t = 1, 2, ..., m\}$.

Def. 2.10 ***Positive Match between SIFT Descriptors*****:** Given two sets of SIFT points $\mathcal{V}_i = \{ V_i^{t_i},\ t_i = 1, 2, ..., m_i \}$ and $\mathcal{V}_j = \{ V_j^{t_j},\ t_j = 1, 2, ..., m_j \}$, for a descriptor $U_i^{t_i}$ of $\mathcal{V}_i$, $\exists\, U_j^{t_{j1}} \in \mathcal{V}_j$, subject to

$$U_j^{t_{j1}} = \arg\min d(U_j^{t_j}, U_i^{t_i}),\ U_j^{t_j} \in \mathcal{V}_j \ and\ d(U_j^{t_j}, U_i^{t_i}) \leq \varepsilon_\tau. \tag{4}$$

$$U_j^{t_{j2}} = \arg\min d(U_j^{t_j}, U_i^{t_i}),\ U_j^{t_j} \in \mathcal{V}_j - V_j^{t_{j1}}\ . \tag{5}$$

$$and\ \gamma = d(U_j^{t_{j1}}, U_i^{t_i}) / d(U_j^{t_{j2}}, U_i^{t_i}) \leq \gamma_\tau. \tag{6}$$

where $d(\bullet) = \| \bullet \|_2$ is Euclidean distance between two vectors. ε_τ is a threshold of Euclidean distance and γ_τ is a threshold of the ratio of γ defined above. When the condition above is satisfied, then $U_i^{t_i}$ and $U_j^{t_{j1}}$ are said to have a positive match.

In our graph based model, the coordinates are used to construct Delaunay graph $G = (\mathcal{V}, E)$, with node-set $\mathcal{V}=\{V^t,\ t = 1, 2, ..., m\}$. It is important to emphasize that in our CSHG model the nodes are attributed with information such as locations and descriptors of SIFT points. We now expand our definition of $CS(G_i,G_j)$ to include the role of SIFT points.

Def. 2.11 ***Common subgraph*****:** $CS(G_i,G_j)$ satisfies the conditions:

1) For each node in $CS(G_i,G_j)$, the corresponding descriptors of G_i and G_j are positive matches.

2) We obtain two weighted Delaunay graphs $G_i^{'}$ and $G_j^{'}$ using the feature coordinations of G_i and G_j. For $CS(G_i,G_j)$, the Euclidean distance of spectral vectors $S_{G_i^{'}}$ and $S_{G_j^{'}}$ must be less than the threshold λ_τ.

Ranking Local Invariant Features through Boosting. The method to rank SIFT features, referred to as R-SIFT, is outlined in the schematic diagram given in Figure 1. The method uses an image synthesis method to locate feature points that are robust to random image perturbations

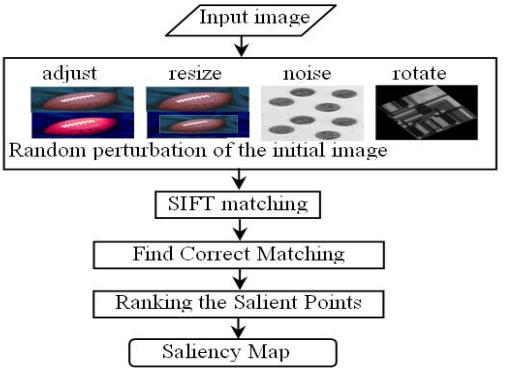

Fig. 1. Diagram of R-SIFT algorithm

We input an initial image to the image synthesis system. For the initial image the feature set is $\mathcal{V}_{ini}$ and the points belonging to the set, i.e. V_{ini}^t, are ordered according to the gradient magnitudes of $\overrightarrow{R}_{ini}^t$. We synthesized new images using a Monte Carlo method which resizes, rotates and adds noise to the original image.

These synthesized images are then processed to extract features and the features are matched with those in the initial input image. The fraction of correctly matching points (FCM) for each SIFT point in the original image is η^t. and the gradient magnitudes of $\overrightarrow{R}^t = (r^t, \alpha^t)^T$ is unified according to the maximum-minimum transform in the group of $\overrightarrow{R}_{ini}$, which is denoted as ζ^t. Our measure used for ranking the feature points is

$$\rho^t = \eta^t \times \zeta^t. \tag{7}$$

3 Constructing a CSHG from R-SIFT Features

We select the first T R-SIFT points from an image with known class identity. We denote the points by the set $\mathcal{V}$=$\{V^t, t = 1, 2, ..., T\}$ which we use to construct a Delaunay graph G. In our experiment, T is set as 40. If there are less than T points, then all the available SIFT points are selected. In this way, a set of graphs $G_l,\ l = 1, 2, ..., N$ are obtained and these are used to construct the hyper-vertex set HV of a CSHG.

According to Def. 2.9, a redundant graph means that there are sufficient similar local regions that it may be discarded without significant loss of information. In practice redundant graphs are iteratively discarded from the set $G_l,\ l = 1, 2, ..., N$, until an irreducible subset is located and used to construct a CSHG. To bootstrap the process, prior to training, a group of graphs can be randomly selected without checking for redundancy.

For convenience, we use the notation $CSHG\{\bullet\}$ for the item set in a CSHG. Hence, $CSHG\{G_l\}$ is the set of graphs and $CSHG\{\overrightarrow{U}^t\}$ is the set of R-SIFT feature descriptors. Given a new graph $G_l = (\mathcal{V}_l, E_l)$, $\mathcal{V}_l$=$\{ V_l^t \mid V_l^t$=$((\overrightarrow{X}_l^t)^T, (\overrightarrow{R}^t{}_l)^T, (\overrightarrow{U}_l^t)^T)^T, t = 1, 2, ..., T_l\}$, and a trained CSHG model. For each descriptor $\overrightarrow{U}_l^t$, the ε nearest neighbors, denoted as $NN_\varepsilon\{U_l^t\}$, are those descriptors $\overrightarrow{U}_q^j$ in a CSHG which satisfy the following restrictions:

$$NN_\varepsilon\{U_l^t\} = \{ U_q^j \mid \|U_l^t - U_q^j\|_2 \le \varepsilon_\tau,\ U_q^j \in CSHG\{\bullet\}\}. \tag{8}$$

where ε_τ (normally less than 1) is the Euclidean distance threshold for accepting two descriptors as neighbors. If the the similarity of two the descriptors is defined as

$$\mathcal{S}(U_q^j, U_l^t) = \exp\{-\beta \cdot \|U_l^t - U_q^j\|_2\}. \tag{9}$$

where $\beta \in [0, 1]$. So a modified set of nearest neighbors is

$$NN_\varepsilon\{U_l^t\} = \{ U_q^j \mid \mathcal{S}(U_q^j, U_l^t) \ge \exp\{-\varepsilon_\tau\},\ U_q^j \in CSHG\{\bullet\}\}. \tag{10}$$

The ε nearest neighbor graphs of G_l in a CSHG denoted by $NNG_\varepsilon\{G_l\}$, are

$$NNG_\varepsilon\{G_l\} = \{ G_q \mid \forall U_l^t \in G_l,\ U_q^j \in NN_\varepsilon\{U_l^t\},\ U_q^j \in G_q \text{ and } G_q \in CSHG\{\bullet\}\}. \tag{11}$$

The cardinality of the common subset of G_l and G_q is a measure of how well the two graphs are matched, denoted as $\|CS(G_l,G_q)\|$. In order to find the best matched graphs, combining $\mathcal{S}(U_q^j, U_l^t)$, a ranking-oriented similarity of two graphs G_t and G_q, denoted as $\mathcal{R}(G_l,G_q)$, is defined as

$$\mathcal{R}(G_l, G_q) = \|CS(G_l, G_q)\| \times \prod_{t=1,2,...,T_l} \mathcal{S}(U_q^j, U_l^t). \tag{12}$$

Each graph $G_q \in NNG_\varepsilon\{G_l\}$ can be placed in descending order according to $\mathcal{R}(G_l,G_q)$, and this order is denoted by $ONNG_\varepsilon\{G_l\} = \{ G_{tq},\ tq = 1, 2, ..., \|NNG_\varepsilon\{G_l\}\|\}$. Hence, the first K graphs of the ordered set $ONNG_\varepsilon\{G_l\}$ are the $K-$ nearest neighbor graphs of G_l, and these are denoted as $NNG_K\{G_l\}$. If the number of ε nearest neighbor graphs

$NNG_\varepsilon\{G_l\}$ is large, we may select $NNG_K\{G_l\}$ instead for the purposes of recognition, clustering, retrieval or tracking.

Up to now, geometric constraints have not been utilized. As a result $ONNG_\varepsilon\{G_l\}$ may contain serious errors due to mismatched feature points. In fact, the false positive rate becomes rather high if we make decisions according to just $NNG_K\{G_l\}$ or $ONNG_\varepsilon\{G_l\}$.

To overcome this problem suppose that for a matched graph pair G_l and $G_q \in NNG_K\{G_l\}$, the coordinates of the matched features are as X_l and X_q. To match these points we use a Procrustes alignment procedure [11]. To this end we construct the orthogonal matrix

$$R = \arg\min \|X_l \cdot \Omega, X_q\|_F, \ \textit{subject to } \Omega^T \cdot \Omega = I. \tag{13}$$

where $\| \bullet \|_F$ denotes the Frobenius norm. The norm is minimized by the nearest orthogonal matrix

$$R^* = \Psi \cdot \Upsilon^*, \ \textit{subject to } X_l^T \cdot X_q = \Psi \ \cdot \ \Sigma \ \cdot \ \Upsilon^*. \tag{14}$$

where $\Psi \cdot \Sigma \cdot \Upsilon^*$ is the singular value decomposition of matrix $X_l^T \cdot X_q$. The goodness-of-fit criterion is the root-mean-squared error, denoted as $e(X_l, X_q)$. The best case is $e(X_l, X_q) = 0$. The error e can be used as a measure of geometric similarity between the two groups of points. Suppose there exists mismatched pairs of feature points, then $e(X_l, X_q)$ will be greater than 0. If we discard one pair of points from X_l and X_q, denoted $X_l - X_l^i$ and $X_q - X_q^i$, a set $e(X_l - X_l^i, X_q - X_q^i),\ i = 1, 2, ..., \|CS\,(G_l,G_q)\|$ can be obtained. The maximum decrease of $e(X_l - X_l^i, X_q - X_q^i)$ is defined as

$$\triangle e(\|CS\,(G_l, G_q)\|) = e(X_l, X_q) - \min\{e(X_l - X_l^i, X_q - X_q^i)\} \tag{15}$$

if $\triangle e(\|CS\,(G_l, G_q)\|)/\ e(X_l, X_q) > \epsilon$, e.g. $\epsilon = 0.1$, the corresponding pair X_l^i and X_q^i is discarded as a mismatched feature pair. This leave-one-out procedure can be proceed iteratively, and is referred as the iterative Procrustes matching of G_l and G_q.

After the iterative Procrustes method has been applied, the similarity ranking of G_l and G_q can be redefined as

$$\mathscr{R}^*(G_l, G_q) = \mathscr{R}(G_l, G_q) \ \times \ (\ \exp(-\ e(X_l, X_q))\ )^\kappa. \tag{16}$$

where κ is a parameter used to amplify the influence of the geometric dissimilarity of X_l and X_q, and is adaptively determined according to the number of mismatched feature pairs discarded by iterative Procrustes matching. Similarly, according to the ranking index $\mathscr{R}^*(G_l, G_q)$, ordered graph sets $ONNG_\varepsilon^*\{G_l\}$ and $ONNG_K^*\{G_l\}$ can be obtained. Moreover, we define a highly ranked subset $ONNG_{K\tau}^*\{G_l\}$ of $ONNG_K^*\{G_l\}$ as follows,

$$ONNG_{K\tau}^*\{G_l\} = \{\ \mathscr{R}^*(G_l, G_q) \geq \mathscr{R}_\tau^*,\ G_q \ \in ONNG_K^*\{G_l\}\ \}. \tag{17}$$

only those $G_q \in ONNG_{K\tau}^*\{G_l\}$ will be utilized. In Section 4, we will show how $\mathscr{R}_\tau^*$ is determined.

To summarize the above, suppose a CSHG is constructed and we wish to locate the best matched graph for a new image so as to incrementally refine the CSHG. This involves three steps. First, construct a graph G_l based on R-SIFT features. Second, find

the best matched graph set $ONNG^*_{K\tau}\{G_l\}$ from the CSHG. Third, check whether G_l is redundant or not. If not, G_l will be incremented to the CSHG.

For multiple objects, CSHG's are trained separately and can be aggregated together as a multi-object CSHG.

4 Experiments and Discussions

This section is divided into two parts. We commence in Section 4.1 by illustrating the effectiveness of the R-SIFT method. Section 4.2 reports results on using the CSHG for object recognition for single 3D object.

4.1 Experiments on R-SIFT

Figure 2 consists of a series of panels each containing three images. From left to right the leftmost is the initial image, the middle images shows the first 20 R-SIFT points (deep blue circles) and their Delaunay graph (red lines), the rightmost images shows the complete set of detected SIFT points. In the first two panels for human faces, the first 10 R-SIFT correspond to the mouth, eyes, nose, face, jowl and hair. In the remaining examples (e.g. the dog, the deer, the polar bear and the flower) the R-SIFT features are also meaningful. The result is encouraging, since most of the highly ranked points correspond to those known to be visually salient, and hence the ranking process seems to be effective.

Fig. 2. Ranked salient features of various images

4.2 CSHG Modeling and Recognition for Single 3D Object

The CSHG differs from alternative recognition methods in two important respects. First, negative samples are not prerequisites in the training of the CSHG model. Second, object models can be constructed individually for different objects. In this section, we will demonstrate how the CSHG's can be constructed for different objects and then used for object recognition. We will use object 100 (the small car) from the COIL-100 data set, shown in Figure 3, to illustrate some of the steps of the process.

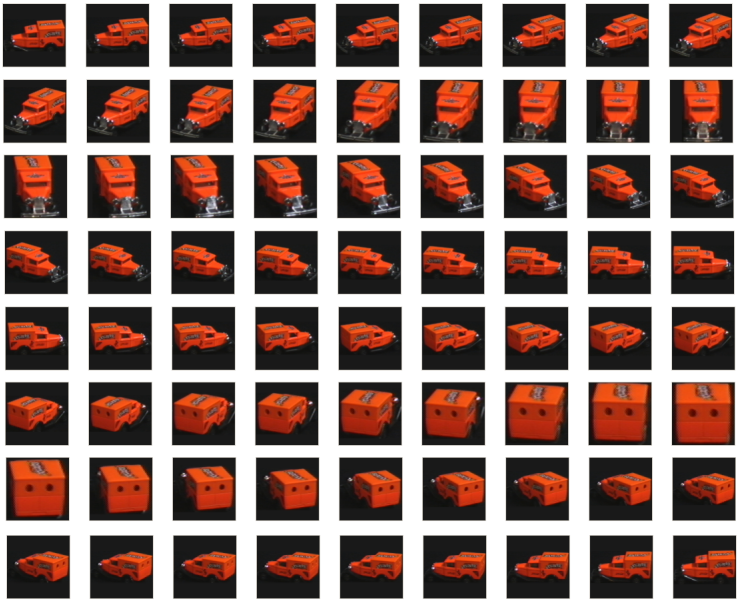

Fig. 3. Images of different views of the small car(obj100) in COIL data set

For the object, we randomly selected 56 of the available 72 images as a training set. Each of the 56 training images were processed to extract R-SIFT points. For each of the 56 training images, there are at most 40 R-SIFT points and these are used to construct a Delaunay graph. After traing, 42 graphs are categorized as non-redundant and used to construct a CSHG model. Their sibling relations are also learned and encoded.

Each of the 72 object views are used to generate noisy test images by adding salt and pepper noise, speckle noise or Gaussian noise. For the case of salt and pepper noise, a noise density d (i.e. that approximately $d*100\%$ pixels of an image affected) varies from 0.01 to 0.12 with increments of 0.01. For each noise setting 360 different images are generated. In our experiments we vary the recognition threshold $\mathscr{R}_\tau^*$ from 0 to 15. The correct recognition rates (denoted by r_g) for the case of salt and pepper noise are shown in Figure 4 (Left). If the threshold $\mathscr{R}_\tau^*$ is set as 7, the average of correct recognition rates ($\overline{r}_g$) is greater than 0.95. If, on the other hand, $\mathscr{R}_\tau^*$ is set as 6 then $\overline{r}_g$ is greater than 0.97, and if $\mathscr{R}_\tau^*$ is set as 8, then $\overline{r}_g$ decreases to less than 0.9.

The second noise process considered is pure speckle. Here multiplicative uniformly distributed random noise with mean 0 and variance v is added to the initial images. The multiplicative noise variance v is varied from 0.05 to 0.16. The recognition results obtained are shown Figure 4 (Right). The third noise process considered is an additive mixture of salt and pepper noise, speckle noise, and Gaussian noise. The variances of the all three processes are set to 0.03 , 0.04 or 0.05, and the recognition results obtained are shown in Figure 5(Left). The results are very similar to those shown in Figure 4, and hence the nature of the noise process does not appear to be a critical issue in determining performance.

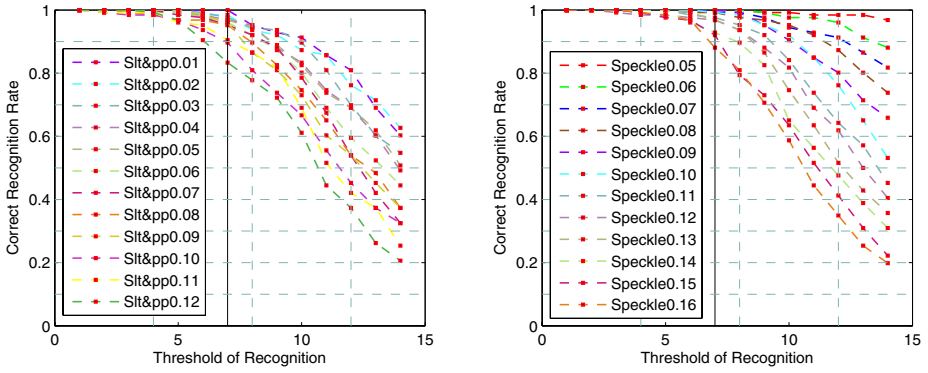

Fig. 4. Recognition results of different generated image data sets

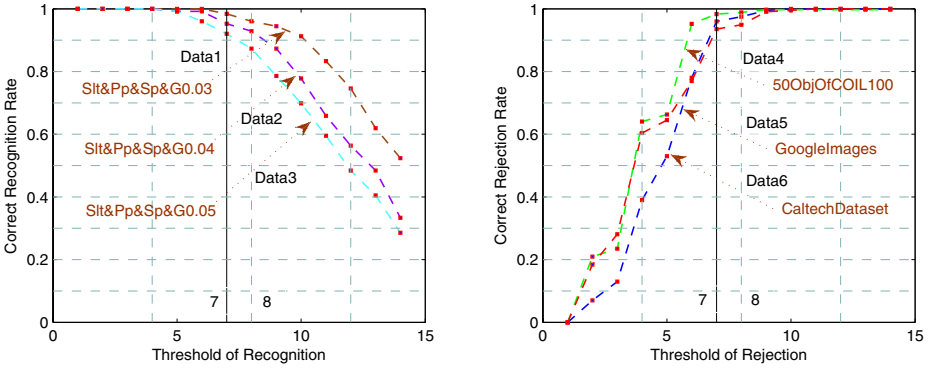

Fig. 5. Recognition results for different data sets

We have also explored how the true rejection rate (denoted by r_j) depends on the threshold. We randomly selected negative test samples from the COIL100 data set, Caltech101 data set and Google images. For COIL100, we randomly selected 36 images for 50 random objects (giving 1800 images). Similarly, we randomly selected 500 images from Caltech101 and 500 images from Google images. The rejection results are shown in Figure 5(Right). If $\mathscr{R}_\tau^*$ is set as 7, the average correct rejection rate ($\overline{r}_j$) for the three cases is greater than 0.95. On the other hand, if $\mathscr{R}_\tau^*$ is set to 8, $\overline{r}_j$ is greater than 0.98 and if $\mathscr{R}_\tau^*$ is set to 6, then $\overline{r}_j$ decreases to 0.83. Hence to compromise, we set $\mathscr{R}_\tau^*$ to 7.

5 Conclusions

We have shown how CSHG models can be obtained separately for multiple objects. Objects are efficiently represented with R-SIFT points. The CSHG model comprehensively utilizes both SIFT and geometric constraints, and hence combines both global

and local information. The framework of this paper can be extended to pose estimation, action recognition and event recognition.

At the moment the construction of our class prototype is purely structural. In the future we aim to pose its construction in an information theoretic setting by casting the process into a description length setting.

References

1. Bosch, A., Zisserman, A., Muoz, X.: Scene Classification Using a Hybrid Generative/Discriminative Approach. IEEE Trans. PAMI 30(4), 1–16 (2008)
2. Bosch, A., Muoz, X., Marti, R.: Which is the best way to organize/classify images by content? Image and Vision Computing 25, 778–791 (2007)
3. Berge, C.: Hypergraphs. North-Holland, Amsterdam (1989)
4. Chung, F.: Spectral Graph Theory. American Mathematical Society (1997)
5. Lowe, D.G.: Distinctive image features from scale-invariant key points. IJCV 60(2), 91–110 (2004)
6. Elazary, L., Itti, L.: Interesting objects are visually salient. Journal of Vision 8(3), 1–15 (2008)
7. Li, F.F., Perona, P.: A Bayesian hierarchical model for learning natural scene categories. CVPR 2, 524–531 (2005)
8. Bay, H., Tuytelaars, T., Gool, L.V.: SURF: Speeded Up Robust Features. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006. LNCS, vol. 3951, pp. 404–417. Springer, Heidelberg (2006)
9. 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)
10. Yan, K., Sukthankar, R.: PCA-SIFT: A more distinctive representation for local image descriptors. CVPR (2), 506–513 (2004)
11. Schonemann, P.: A generalized solution of the orthogonal Procrustes problem. Psychometrika 31, 1–10 (1966)
12. Sivic, J., Zisserman, A.: VideoGoogle: A text retrieval approach to object matching in videos. ICCV 2, 1470–1477 (2003)
13. Sivic, J., Russell, B.C., Efros, A.A., Zisserman, A., Freeman, W.T.: Discovering objects and their location in images. ICCV 1, 370–378 (2005)
14. Torsello, A., Hancock, E.R.: Graph Embedding using Tree Edit Union. Pattern Recognition 40, 1393–1405 (2007)
15. Torsello, A., Hancock, E.R.: Learning Shape Classes using a Mixture of Tree Union. IEEE Trans. PAMI 28, 954–967 (2006)
16. 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)