Sketch Recognition by Ensemble Matching of Structured Features

Yi Li
http://www.eecs.qmul.ac.uk/~yl303

Yi-Zhe Song
http://www.eecs.qmul.ac.uk/~yzs

Shaogang Gong
http://www.eecs.qmul.ac.uk/~sgg

School of Electronic Engineering and
Computer Science,
Queen Mary, University of London,
London E1 4NS, UK

Abstract

Sketch recognition aims to automatically classify human hand sketches of objects into known categories. This has become increasingly a desirable capability due to recent advances in human computer interaction on portable devices. The problem is nontrivial because of the sparse and abstract nature of hand drawings as compared to photographic images of objects, compounded by a highly variable degree of details in human sketches. To this end, we present a method for the representation and matching of sketches by exploiting not only local features but also global structures of sketches, through a star graph based ensemble matching strategy. Different local feature representations were evaluated using the star graph model to demonstrate the effectiveness of the ensemble matching of structured features. We further show that by encapsulating holistic structure matching and learned bag-of-features models into a single framework, notable recognition performance improvement over the state-of-the-art can be observed. Extensive comparative experiments were carried out using the currently largest sketch dataset released by Eitz et al. [15], with over 20,000 sketches of 250 object categories generated by AMT (Amazon Mechanical Turk) crowd-sourcing.

1 Introduction

Sketch has been used as a basic visual communication media since pre-historic times and is becoming ubiquitous in recent years due to the popularity of touch sensitive devices such as smart phones and tablets. Neuroscience studies also suggest that sketching is the elementary way for human to conceptualize visual information [12, 22], and is more expressive when retrieving images than raw text [4, 7]. This has resulted in a growing interest in automatic sketch recognition systems [15] and sketch-based image retrieval (SBIR) systems [4, 9, 14]. However, automatically interpreting human sketches consistently is nontrivial. This is because (1) sketches are typically sparse in details without color and texture, as compared to natural images that possess richer and denser visual details; (2) the geometric structures of object sketches often have variable complexity and are inconsistent in details; (3) inter-class variations of sketches can be small whilst their intra-class differences can be large, making classification more ambiguous.

Current methods for sketch recognition and sketch-based image retrieval address the sparseness problem by densely sampling features [14, 15], using larger patches [13, 15] or sampling features along the edges [9]. They all employ a bag-of-features (BOF) representation of object sketches without considering their spatial structures. For instance, although [15] proposed a method for learning a sketch model per object category using Support Vector Machines (SVM) by utilizing a very large crowd-sourced database in order to address the inter-class similarity and intra-class variety problem, they do not exploit structural information. A characteristic of sketch is that its basic strokes do not necessarily exhibit strong discriminative cues in isolation. Instead, the structures of a sketch both locally and holistically contain informative visual cues for discriminating different sketches. We argue that a desirable sketch representation and matching model requires a mechanism to encode its structure in relation to the sparse features so that the model is capable of capturing sufficient information for inter-class discrimination whilst being flexible enough to cope with intra-class structural variations (holistic) and subtle structural detail differences (local).

In general, there are a number of ways for representing holistic image structures including commonly employed spatial pyramid matching [11] or spatial BOF [3]. On the other hand, a more robust and potentially flexible approach to image structure representation is a star graph model [2]. A standard star graph, also known as an ensemble, has an assigned center with each feature point represented as a node in the graph and connected to the center. An ensemble model encodes both direction and distance of each feature point to the graph center therefore provides a richer and more flexible representation of spatial structures than other alternative schemes. In this work, we exploit a star graph as a structured feature representation to encode both local features and the holistic structure of a sketch. We also exploit the ensemble matching for computing the distance metric when comparing star graph representations of different sketches [18]. Moreover, we examine and evaluate alternative local features for the star graph and use K nearest neighbors (KNN) classification to evaluate the effectiveness of different features and the star graph representation for sketch recognition. We show in the experiments (Section 4) that by encoding holistic structures alone, recognition performance comparable to that of a BOF based alternative [15] can be obtained.

While star graphs can represent holistic structure of sketches well, current BOF approaches [9, 14, 15] have the benefit of being able to better capture subtle structural details. For instance, diary cows have similar holistic structure to that of a horse and only differ to the latter in body local texture patterns (Figure 4(c)). To that end, we further propose an unified framework to address both holistic structural variations and local detail differences. More specifically, we introduce a separate category filtering process as a first step prior to ensemble matching. This filtering process not only preserves sketch categories having similar local patterns, but also speeds up the ensemble matching process since less candidate categories are considered during matching. We follow [15] who trained SVM classifiers using a BOF representation per sketch category. However, instead of classifying into just the top one category, the classifier is used to select top N sketch categories that are most similar to the query, where N is much smaller than the total number of categories. We demonstrate the advantages of this unified approach against both a standard star graph based ensemble matching model and the local BOF based model [15] with extensive comparative evaluations in the experiments (Section 4).

The contributions of this work are: (1) introduce a model for encoding holistic structure of sketches by exploiting a star graph and ensemble matching, this is specifically designed to overcome the feature sparsity problem encountered in sketch recognition; (2) propose an unified sketch recognition framework that addresses both holistic structural changes and

subtle local detail variations; (3) examine and evaluate different local feature representations for sketch recognition using star graph and ensemble matching, and demonstrate the performance advantage of our proposed model over alternative sketch recognition techniques published in the literature.

2 Related Work

The general concept of a structured feature representation has been adopted elsewhere including scene category recognition [11], landmark images retrieval [3] and sketch-based retrieval of drawings [7, 19]. However, most of these structure encoding methods are fairly specific to the problem domains they were designed to address, therefore are not readily applicable to the problem of representing sketches. Many techniques are designed for the image domain and work with BOF representations. For instance, the spatial pyramid matching method [11] utilizes a series of increasingly sparser grids over the image, then obtains a weighted sum of the number of matches yielded at each grid level. It is designed for scene categorization and optimized for capturing frequently emerged representative patterns of certain scene categories. However, such a scheme is not directly suitable for sketches due to large deformations and variations in sketches, and potential poor structural information captured by fixed-position grid cells due to the sparsity and uncertain distribution of sketch details. Alternatively, 2D features can be projected from a 2D feature space onto certain lines, or circles in order to group the features by sectors [3]. This concept of 1D encoding of local features by their 2D locations works well for landmark images when a dominant direction(s) may be readily obtained, a property cannot be found generally in sketches.

There are several methods proposed for structured representations of sketches, in which topological relationships between sketch parts were exploited for improving matching accuracy [7, 19]. However, such schemes are restricted to a certain type of sketches, namely technical drawings (e.g. CAD), when sketch parts can be relatively easily segmented into well-defined topological parts. This is not generally the case for free-hand drawn sketches. It has been shown that grouping and segmenting human free drawn sketches consistently and reliably is a challenging problem and remains an open problem to be resolved [20].

Many local feature representations commonly found in the image domain have been investigated for sketches. Eitz et al. [14] offers a detailed account of many popular features including angular radial partitioning (ARP), edge histogram descriptor (EHD), and histogram of oriented gradients (HOG), and investigated their performance for SBIR where a BOF representation was adapted. It is found that HOG generally outperforms others, but the performance of which is sensitive to patch size and codebook size of the BOF model. Very recently, an evaluation was conducted by Hu and Collomosse [8], where several local features including gradient field HOG (GF-HOG), multi-resolution HOG (MR-HOG), scale-invariant feature transform (SIFT), self-similarity (SSIM), shape context (SC) and structure tensor (ST) are investigated within a BOF model for SBIR. HOG based features were again found to outperform others.

Sketch recognition shares some common characteristics with shape matching [1, 5]. However, significant differences exist. Most shape matching techniques work on enclosed 2D outline contours of objects without inner structural details and/or with more consistent shape characteristics extracted from object images. On the other hand, hand-drawn object sketches are relatively free (less regularized) with internal feature details of a sketch being important for discriminating sketches of different objects, especially on large-scale datasets where ambiguities commonly exist [15].
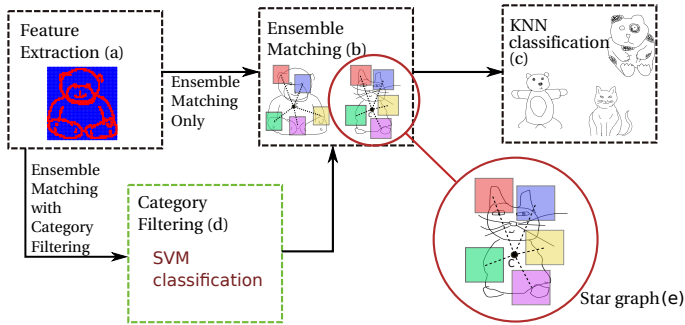
Figure 1: Four components (a-d) of our approach with working flows and visualization of a star graph (e).

3   Methodology

The proposed sketch recognition model consists of four main components (shown in Figure 1). They are: local feature extraction, structure representation by star graph ensemble matching, KNN classification and category filtering via SVM classification. We shall describe each component in details as follows.

3.1   Feature Extraction

We sample feature points densely along sketch strokes while ignoring background. The descriptors and locations of the features are then used to form the star graph model which encodes both holistic geometrical structures and local features of the sketch. We consider three types of local features as candidates for constructing a model representation.

Histogram of Oriented Gradients (HOG): computed on a dense grid of uniformly divided cells. The gradients in each cell are interpolated into several orientation bins, which are then formed into a histogram reflecting the cell sequence. This feature has commonly been reported to have best performance with sketches [9, 13, 14, 15].

Self-Similarity (SSIM): proposed by [18] to capture the salient geometric structure in the image computed on repeated patch patterns other than pixels. It is closely related to ensemble matching with the potential for different visual domains, like photos, paintings and drawings.

Local Binary Pattern (LBP) : labels the texture of each pixel by thresholding the neighborhood of the pixel into a binary number that is within a fixed pattern set. Then the pixels' textures are formed into a histogram of the fixed set of patterns [16]. LBP is robust to monotonic gray-scale changes and has good discriminative power.

3.2   Ensemble Matching

The matching of two ensembles is essentially a graph matching problem, where each ensemble is encoded as a star graph. More precisely, we denote a star graph as $G = (\mathcal{V}, \mathcal{E}, \mathcal{A})$, where $\mathcal{V}$, $\mathcal{E}$, $\mathcal{A}$ represent respectively a set of nodes, edges and attributes of the graph. In particular, $\mathcal{V} = \{v_i\}_{i=1}^{N_s} \cup c$ is the set of all N_s sample points $\{v_i\}_{i=1}^{N_s}$ and the center c, and $e_i \in \mathcal{E}$ is the implicit link between v_i and c. Moreover, $\mathbf{a}_{ic} \in \mathcal{A}$ represents the geometrical relationship between v_i and c, and $\mathbf{a}_i \in \mathcal{A}$ denotes the corresponding feature descriptor of v_i, which can be one of the three candidate feature descriptors, HOG, SSIM and LBP, to be evaluated for star graph ensemble matching.
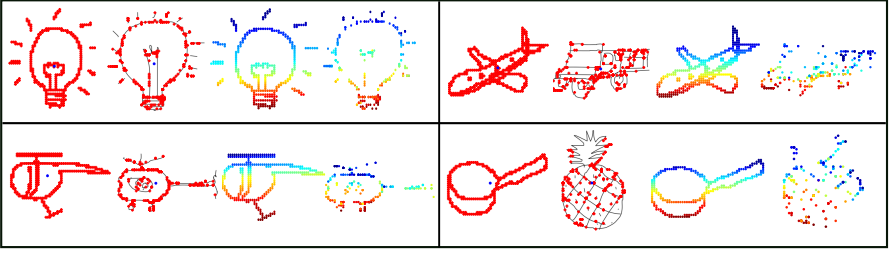
Figure 2: Ensemble matching visualization of four pairs of sketches: first two images of each cell show the matched points (in red); the last two illustrate matching correspondences where matched points have the same color.

The computation of the similarity between ensemble q (query) and t (target) is formulated as follow [2]:

$$P(G^q, G^t) = \prod_i P(\mathbf{a}_i^t | \mathbf{a}_i^q) P(\mathbf{a}_{ic}^t | \mathbf{a}_{ic}^q) \tag{1}$$

where $G^q = (\mathcal{V}^q, \mathcal{E}^q, \mathcal{A}^q)$ and $G^t = (\mathcal{V}^t, \mathcal{E}^t, \mathcal{A}^t)$ are their corresponding star graphs. $P(\cdot, \cdot)$ denotes the normalized distance metric value and is considered as probability. The feature similarity term $P(\mathbf{a}_i^t | \mathbf{a}_i^q)$ accounts for the similarity between features and is computed with a sigmoid function:

$$P(\mathbf{a}_i^t | \mathbf{a}_i^q) = \frac{1}{1 + \exp\left(-\|\mathbf{a}_i^t - \mathbf{a}_i^q\|_1\right)} \tag{2}$$

on the L_1 distance of two features according to [18]. The feature location correlation term $P(\mathbf{a}_{ic}^t | \mathbf{a}_{ic}^q)$ stands for the location correlation between two features and is computed with Equation (3) following [2].

$$P(\mathbf{a}_{ic}^t | \mathbf{a}_{ic}^q) = \exp\left(-(\mathbf{a}_{ic}^t - \mathbf{a}_{ic}^q)^T S_L^{-1} (\mathbf{a}_{ic}^t - \mathbf{a}_{ic}^q)\right) \tag{3}$$

where S_L is a constant covariance matrix to allow for some deviations in patch locations.

A visualization of matching results for several pairs of ensembles is shown in Figure 2, where matched points are assigned the same color. It can be seen that points having similar appearance and location correspondences are well matched.

We modify traditional ensemble matching [18] in several minor ways. First, we employ a two steps ensemble matching algorithm to accelerate the matching process. It first finds the most similar D target features $\{\mathbf{a}_j^t\}_{j=1}^D$ for each feature in the query according to Equation (2) (D is much smaller than the total feature amount in the target), then calculates location correlations only for these D features. Second, a partial matching penalty factor w is added to penalize the points not being matched in the target, which is defined as the proportion of points being matched in the target. Third, the sum rule is employed to obtain the overall matching score instead of the product rule employed in [2], as it is proven to be the most resilient to estimation errors [10]. The similarity between the query and the target is then:

$$P(G^q, G^t) = w * \sum_i \max_j P(\mathbf{a}_j^t | \mathbf{a}_i^q) P(\mathbf{a}_{jc}^t | \mathbf{a}_{ic}^q) \tag{4}$$

In addition, we swap the query and the target after obtaining the matching score and add up the swapped score to the previous score, so the matching score between two sketches becomes constant comparing from either side. We also get rid of the center estimation and

multi-scale matching, due to the single object nature and pre-scaling procedure, respectively. The center of an ensemble is set to its center of mass.

3.3   KNN classification

Given a query sketch ensemble, q, represented by a star graph, G^q, and a test set of sketches $\mathcal{T} = \{\mathcal{T}^m\}_{m=1}^M$, $\mathcal{T}^m = \{G^{ml}\}_{l=1}^L$, where $\mathcal{T}^m$ is a set of L sketches belonging to the mth category, each represented by a star graph G^{ml}. M is the total number of categories in the dataset and the total sketch number is $M \times L$ (assuming L is constant for all categories). Matching is then performed following Equation (1), and the distance function for KNN classification is the reciprocal of Equation (1):

$$D(G^q, G^{ml}) = \frac{1}{P(G^q, G^{ml})} = \frac{1}{\sum_i P(\mathbf{a}_i^{ml}|\mathbf{a}_i^q)P(\mathbf{a}_{ic}^{ml}|\mathbf{a}_{ic}^q)} \tag{5}$$

For each query ensemble q, its K nearest neighbors are found using the metric of Equation (5). And its category is assigned to the one where the majority of the K nearest neighbors belong to.

3.4   Category Filtering

We employ SVM classifiers to filter sketch categories prior to ensemble matching, therefore keep N categories closest to the query in term of local details other than holistic structure. That is, we reduce $\mathcal{T}$ to $\hat{\mathcal{T}} = \{\hat{\mathcal{T}}^i\}_{i=1}^N$, where $N \ll M$.

More specifically, we represent a sketch by a n-dimensional BOF histogram $\mathbf{h}$. A set of SVM classifiers are trained with respect to the number of sketch categories in a training dataset. For a probe sketch image to be classified, the following voting function classifies a given probe sketch image into the ith category:

$$c^i(\mathbf{h}) = \sum_j w_j^i K(\mathbf{s}_j^i, \mathbf{h}) + b \tag{6}$$

where K is a kernel function, $\mathbf{s}_j^i$ are the support vectors, w_j^i are weights, and b is the bias. $c^i(\mathbf{h})$ is therefore the classification response measuring similarity between the probe and the ith category. All parameters are obtained during SVM training, and the RBF kernel was used for K.

For each category i, a classifier c^i is trained using the sketches from category i as positive examples and sketches from all the rest categories as negative examples. Given a probe sketch, we compute classification responses for all classifiers, $\{c^i(\mathbf{h})\}_{i=1}^M$, and keep the top N categories as the closest categories in terms of representative local features.

We deploy k-means clustering to generate a visual codebook $\mathcal{U} = \{\mathbf{u_i}\}_{i=1}^V$ [15, 17]. We then represent a feature by assigned weights to each word $\mathbf{u_i}$ using a Gaussian kernel to improve classification performance. σ of the Gaussian kernel is essential to the final performance and is optimized using cross-validation in this work.

4   Experiments

In this section, we demonstrate the effectiveness of our proposed sketch recognition method and compare its performance against BOF based models [15]. Moreover, we also compare the benefit of the ensemble matching with category filtering approach over a star graph ensemble matching based model. From the experiments, it is evident that structured feature

representation using a star graph with ensemble matching can significantly improve sketch recognition rate. In addition, category filtering can further enhance ensemble matching by selecting a much smaller subset of the most relevant categories according to the local structural details presented in the probe sketch image. Among the three local feature types considered, our experiments show that the HOG features remain to be the most effective for structured feature representation using star graph ensemble matching.

4.1 Dataset and settings

Dataset Eitz et al. [15] provided a large-scale human sketch dataset, currently the largest sketch benchmark dataset publicly available with 250 object categories and 20,000 sketches. The dataset was generated by AMT crowd-sourcing using 1,350 unique non-expert drawers participated in the experiment, therefore make it diverse in terms of drawing style and sophistication level. In our experiments, to be comparative with [15], each sketch is normalized to 256×256 pixels, by extracting the bounding box with the sketch and then scaling and locating it in the center of the 256×256 area.

Features Three types of features are evaluated with the proposed frameworks. They are HOG, SSIM and LBP. Self similarity feature is computed with VGG's implementation [6]. The 'var_noise' parameter is set to 50,000, 5 radical bins and 12 angular bins were used. A 51×51 grid (every 5 pixels) is used to extract the sample points. VGG's saliency checking, homogeneity checking and second-nearest neighbor checking were all disabled, for their unsuitability for sketches. A customized homogeneity checking is introduced to keep sample points all along the sketch contour and these sample points are adopted by the rest two features. HOG is computed using the VLFeat [21] implementation, where each patch is divided into 4×4 cells, and the orientation parameter is set to 4. LBP is computed using CMV's implementation [16], and the default settings are kept.

Cross validation 4-fold cross validation is employed for all the experiments reported in this work, as compared to the 3-fold cross validation carried out in [15]. The whole dataset is separated into growing subsets (20,40,60,80 sketches per category), with 4-fold cross validation performed on each subset.

Bag-of-features representation For all the features referred, the BOF representations are obtained from a codebook containing 1,000 visual words. 1,000,000 features are randomly sampled to train the codebook with k-means clustering. The σ parameter for the Gaussian kernel is searched between $[0.001, 1]$ and set to 0.5 for HOG and SSIM, and 0.01 for LBP.

K **nearest neighbors parameter settings** For KNN classification on BOF, K is searched among $\{1, 2, 3, 4, 5\}$, and the distance metric is searched in $\{cityblock, euclidean, cosine, correlation\}$. The parameter settings we chosen for HOG, SSIM, LBP are $[5, cityblock], [5, cityblock], [5, cityblock]$ respectively.

SVM parameter settings The γ and C parameters of SVM are searched between $[2^{-2}, 2^{8}]$. A coarse grid search is performed first with an interval of 2^2, followed by a fine grid search with an interval of $2^{0.25}$. The best parameter settings for HOG, SSIM, LBP are $[2^{6.25}, 2^{3.25}], [2^7, 2^3], [2^7, 2^4]$ respectively.

4.2 Evaluation against Unstructured Feature Representation Model

Table 1 shows quantitative comparisons between the proposed matching method using HOG features against that of [15]. We further breakdown the comparisons into two categories, unsupervised and supervised, based on whether explicit training was involved in any part of

	Eitz et al. [15]	**Ours**
Unsupervised	44%	**53.3%**
Supervised	56%	**61.5%**

Table 1: Result comparisons.

the recognition process. We compare two methods in the unsupervised category: KNN classification with 'soft' assignment reported by Eitz et al. [15], and ensemble matching without category filtering of ours. Supervised result for Eitz et al. refers to their result of SVM classification with 'soft' assignment, while for ours refers to ensemble matching with SVM based category filtering result. It is worth pointing out that using structure encoding by star graph and utilizing ensemble matching, our unsupervised method is capable of giving comparable recognition performance (53.3%) to the supervised learning model (56%) reported by [15]. Moreover, the performance from our ensemble matching with SVM based category filtering approach improves significantly against [15], that is from 56% to 61.5%, where the reported human-to-human recognition rate on the same dataset is 73% [15]. The following sections will offer more detailed analysis of our results.

4.3   Effect of Ensemble Matching

We compare our ensemble matching only approach against BOF based KNN sketch recognition method [15], using three features stated in Section 3.1. The KNN parameters are stated in Section 4.1. Figure 3(a) offers quantitative results for the three features. It is evident that for all three feature types, ensemble matching provides significant improvement than the KNN classification on BOF. It is also evident that HOG features yield the best performance. Figure 4(b) shows three top retrieval sketches of BOF(HOG) and Ens(HOG). The KNN on BOF results normally have weak structure correspondence while ensemble matching results generally have very tight structure correspondence. This somewhat explains the significant improvement in recognition shown in Figure 3(a), and demonstrates the importance of structured feature representation and the effectiveness of the proposed ensemble matching method.
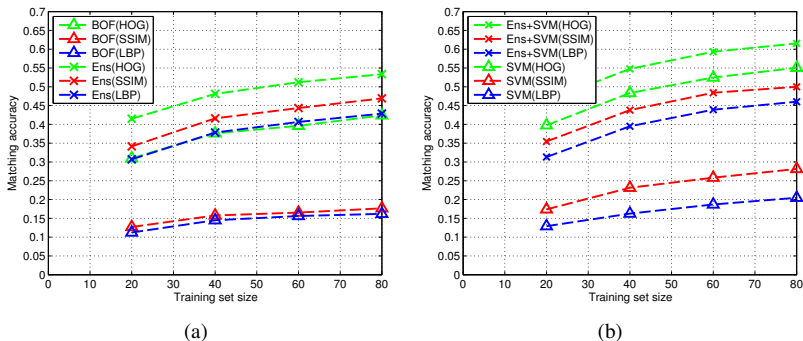

(a)   (b)

Figure 3: (a) quantitative results of ensemble matching (Ens(*)) and KNN classification on bag-of-features (BOF(*)) for HOG, SSIM and LBP; (b) recognition accuracies of our ensemble matching with category filtering approach (Ens+SVM(*)) and SVM classification (SVM(*)) on HOG, SSIM and LBP.
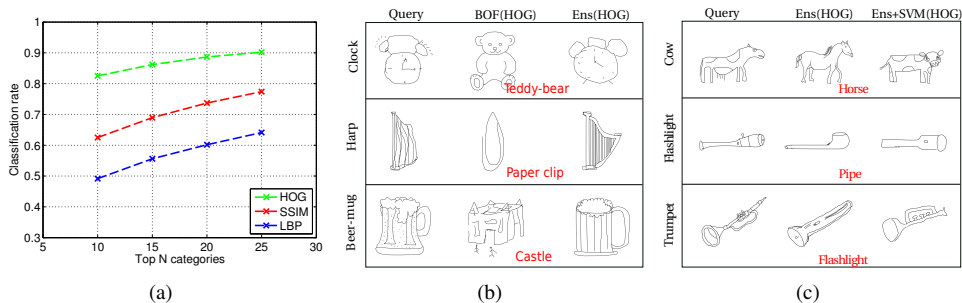
Figure 4: (a) classification rates of category filtering when N is $\{10, 15, 20, 25\}$ for HOG, SSIM and LBP; (b) qualitative comparisons of top retrieval results of three sketch probes on KNN classification on bag-of-features (BOF(HOG)) and ensemble matching only method (Ens(HOG)). (c) qualitative comparisons of top retrieval results of three sketch probes on ensemble matching only (Ens(HOG)) and ensemble matching with category filtering (Ens+SVM(HOG)).

4.4 Effect of Category Filtering

To evaluate the effect of category filtering given a probe sketch image, SVM models are trained per category using the parameter settings stated in Section 4.1. A probe is then classified by each model to generate corresponding classification response. The categories are then ranked by the classification responses, among which top N are kept. We empirically tested the N value variation among $\{10,15,20,25\}$, and found the best choices for HOG, SSIM and LBP features are 10, 25, 25 respectively. Figure 4(a) illustrates how N correlates with classification rate. It can be seen that our filtering step can generally preserve the correct categories, especially using HOG. After the category filtering step, ensemble matching is utilized for sketch classification. The recognition results using the three different feature types are shown in Figure 3(b).

To train SVM models of [15], we use the same features involved in the star graph, and quantize them into BOF representations. The SVM models are trained per category using parameters stated in Section 4.1. And the query is assigned to the category whose SVM model yields the highest classification response. Different recognition accuracies for the three features are shown in Figure 3(b).

From Figure 3(a) and 3(b), it is evident that the unified framework outperforms both ensemble matching only and SVM classification using BOF, and HOG yields the best performance. Figure 4(c) shows the top retrieval examples of ensemble matching without (Ens(HOG)) and with category filtering (Ens+SVM(HOG)). Without category filtering, ensemble matching would mismatch the query due to the high similarity in the holistic structure and ignoring the subtle structural details. However, with category filtering, those mistakes are minimised to some extent. Note that similar qualitative results for SVM(*) are not provided since SVM does category classification only.

5 Conclusions

Holistic geometrical structures of sketches are helpful source of information for sketch recognition, yet have been largely ignored by existing methods. In this work, a star graph based ensemble matching method is presented to model holistic geometrical structures of a sketch. By integrating ensemble matching into sketch recognition, the structure informa-

tion of a sketch can be effectively represented using a star graph. Extensive comparative experiments and staged analysis on the effect of different components of our method are carried out using the largest human sketch dataset currently available in the public domain. One limitation of a standard star graph based ensemble matching is that it stresses more on holistic structure than subtle local detail differences of the sketches. To overcome this limitation, an unified ensemble matching with multi-SVM classification based category filtering is designed to simultaneously benefit from both holistic structure and subtle local details of sketches. Our experiments show that such a combined approach is able to outperform significantly the current state-of-the-art using supervised training of SVMs given a bag-of-features representation [15]. We also show that HOG features are most effective for sketch recognition when encoded in a star graph representation, as compared to self-similarity and LBP local features.

References

[1] S. Belongie, J. Malik, and J. Puzicha. Shape matching and object recognition using shape contexts. *PAMI*, 24(4):509–522, 2002.

[2] O. Boiman and M. Irani. Detecting irregularities in images and in video. *IJVC*, 74(1): 17–31, 2007.

[3] Y. Cao, C. Wang, Z. Li, L. Zhang, and L. Zhang. Spatial-bag-of-features. In *CVPR*, pages 3352–3359, 2010.

[4] Y. Cao, H. Wang, C. Wang, Z. Li, L. Zhang, and L. Zhang. Mindfinder: interactive sketch-based image search on millions of images. In *International Conference on Multimedia*, pages 1605–1608, 2010.

[5] Y. Cao, Z. Zhang, I. Czogiel, I. Dryden, and S. Wang. 2d nonrigid partial shape matching using mcmc and contour subdivision. In *CVPR*, pages 2345–2352, 2011.

[6] K. Chatfield, J. Philbin, and A. Zisserman. Efficient retrieval of deformable shape classes using local self-similarities. In *Workshop on Non-rigid Shape Analysis and Deformable Image Alignment, ICCV*, pages 264–271, 2009.

[7] M.J. Fonseca, A. Ferreira, and J.A. Jorge. Sketch-based retrieval of complex drawings using hierarchical topology and geometry. *Computer-Aided Design*, 41(12):1067–1081, 2009.

[8] R. Hu and J. Collomosse. A performance evaluation of gradient field hog descriptor for sketch based image retrieval. *CVIU*, 117:790–806, 2013.

[9] R. Hu, M. Barnard, and J. Collomosse. Gradient field descriptor for sketch based retrieval and localization. In *ICIP*, pages 1025–1028, 2010.

[10] J. Kittler, M. Hatef, R.P.W. Duin, and J. Matas. On combining classifiers. *PAMI*, 20 (3):226–239, 1998.

[11] S. Lazebnik, C. Schmid, and J. Ponce. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In *CVPR*, pages 2169–2178, 2006.

[12] D. Marr. *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. 1982.

[13] E. Mathias, H. Kristian, B. Tamy, and A. Marc. An evaluation of descriptors for large-scale image retrieval from sketched feature lines. *Computers & Graphics*, 34(5):482–498, 2010.

[14] E. Mathias, H. Kristian, B. Tamy, and A. Marc. Sketch-based image retrieval: Benchmark and bag-of-features descriptors. *TVCG*, 17(11):1624–1636, 2011.

[15] E. Mathias, H. James, and A. Marc. How do humans sketch objects? *ACM TOG (Proceedings SIGGRAPH)*, 31(4):44:1–44:10, 2012.

[16] T. Ojala, M. Pietikäinen, and T. Mäenpää. A generalized local binary pattern operator for multiresolution gray scale and rotation invariant texture classification. In *Proceedings of the Second International Conference on Advances in Pattern*, pages 397–406, 2001.

[17] J. Philbin, O. Chum, M. Isard, J. Sivic, and A. Zisserman. Lost in quantization: Improving particular object retrieval in large scale image databases. In *CVPR*, pages 1–8, 2008.

[18] E. Shechtman and M. Irani. Matching local self-similarities across images and videos. In *CVPR*, pages 1–8, 2007.

[19] P. Sousa and M.J. Fonseca. Sketch-based retrieval of drawings using spatial proximity. *Journal of Visual Language and Computing*, 21(2):69–80, 2010.

[20] Z. Sun, C. Wang, L. Zhang, and L. Zhang. Free hand-drawn sketch segmentation. In *ECCV*, pages 626–639, 2012.

[21] A. Vedaldi and B. Fulkerson. VLFeat: An open and portable library of computer vision algorithms. http://www.vlfeat.org/, 2008.

[22] D.B. Walther, B. Chai, E. Caddigan, D.M. Beck, and F.-F Li. Simple line drawings suffice for functional mri decoding of natural scene categories. *Proceedings of the National Academy of Sciences*, 108(23):9661–9666, 2011.