Making Better Use of Edges via Perceptual Grouping

Yonggang Qi[†] Yi-Zhe Song[⋆] Tao Xiang[⋆] Honggang Zhang[†]
Timothy Hospedales[⋆] Yi Li[⋆] Jun Guo[†]
[†]Beijing University of Posts and Telecommunications [⋆]Queen Mary University of London
{qiyg,zhhg,guojun}@bupt.edu.cn {yizhe.song,t.xiang,t.hospedales,yi.li}@qmul.ac.uk

Abstract

We propose a perceptual grouping framework that organizes image edges into meaningful structures and demonstrate its usefulness on various computer vision tasks. Our grouper formulates edge grouping as a graph partition problem, where a learning to rank method is developed to encode probabilities of candidate edge pairs. In particular, RankSVM is employed for the first time to combine multiple Gestalt principles as cue for edge grouping. Afterwards, an edge grouping based object proposal measure is introduced that yields proposals comparable to state-of-the-art alternatives. We further show how human-like sketches can be generated from edge groupings and consequently used to deliver state-of-the-art sketch-based image retrieval performance. Last but not least, we tackle the problem of free-hand human sketch segmentation by utilizing the proposed grouper to cluster strokes into semantic object parts.

1. Introduction

The human visual system is so powerful that we can easily derive structure from chaos. The Gestalt school of psychologists refer to this phenomenon as perceptual organization [49, 50], where visual elements are grouped based on a set of simple rules, such as proximity and continuity [52]. It is commonly acknowledged that early human visual processing operates by first performing edge detection followed by perceptual organization to group edges into object-like structures [7, 41].

Inspired by these psychological discoveries, extracting edges has long been regarded as key to solving the vision problem. This motivation has resulted in extensive prior art, from the simple gradient-driven Canny edges [8] to more sophisticated methods [30, 38, 27, 29, 33] that exploit multiple features and more elaborate algorithms. Despite these great strides, the produced edges remain noisy and therefore can not be directly used in higher-level applications.

In order to better exploit noisy edges, researchers have investigated means of grouping edges into continuous contours [18, 1, 53] and organizing them hierarchically in terms of probabilistic edge maps [3, 2]. These advances have led to successful application of edges in higher-level vision tasks such as object detection [40, 55], object proposal generation [2, 54] and sketch-based image retrieval (SBIR) [16, 25]. All of these studies either implicitly or explicitly utilize perceptual grouping. However, most are limited to just one Gestalt principle locally and relies on additional features/cues (e.g., color and texture) to alleviate the limitation.

In this paper, we propose a grouper that utilizes multiple Gestalt principles synergistically, and works with edges alone. More specifically, a novel multi-label graph-cut algorithm is used to group edges according to two key Gestalt principles, i.e. continuity and proximity. This is realized by a novel strategy of computing relative penalties between labels and edges using a learning to rank algorithm. Our new perceptual edge grouping framework produces better edge groups which can be used to improve a variety of higher-level tasks including: (i) object proposal generation, where multi-cue edge grouping is exploited for the first time, (ii) SBIR, where the proposed edge grouping helps in generating more human-like edge maps and (iii) free-hand human sketch segmentation, where the grouper is directly used to group strokes to form object parts. All three applications benefits significantly from the unique characteristic of perceptual grouping – it generates object-like structures, which is the ultimate goal of object proposal algorithms and reason behind better SBIR and sketch segmentation.

Our contributions are summarized as follows: (i) A novel perceptual edge grouping algorithm is introduced by embedding learning to rank into a multi-label graph-cut framework. (ii) We propose a novel object proposal generation method based on our perceptual edge grouping framework that improves over existing approaches. (iii) Our grouper facilitates a novel SBIR system that outperforms existing alternatives. (iv) We demonstrate a novel human free-hand sketch segmentation algorithm that breaks sketches into semantic parts.

2. Related Work

Perceptual Edge Grouping There are numerous methods for grouping edge fragments. Some [51, 46, 45] model the problem heuristically where grouping costs or saliency measurements are hand-crafted by intuition. Although they work in certain situations, they may not generalize to others. Others [39, 18, 34] use probabilistic framework, where grouping functions are based on cue statistics, whereas a few specifically examine the global closure property of edges [19, 32]. Without exception, they all work on relatively simple images with plain background. The problem of discovering complete edge groups of an object in a complex real images remains unsolved.

Perceptual grouping has historically played dominant role in edge grouping. [6, 17] importantly verified natural statistics of Gestalt principles for edge grouping. This finding inspires plenty of subsequent work [21, 43, 26] on edge grouping using Gestalt principles. Crucially, although unary Gestalt principles have proven to be useful for edge grouping when used alone [45, 36, 1], very few studies [44] investigate how multiple principles can be exploited jointly in a single framework. This is challenging due to the problem of Gestalt confliction [49, 50], or cross-cue discrepancy. In this paper, we introduce a probabilistic model to fuse two Gestalt principles, i.e., proximity and continuity, as cues for edge grouping, effectively solving the Gestalt confliction problem.

Objectness Objectness [2] is regarded as a key preprocessing step for object detection nowadays. It aims to generate a set of candidate detection boxes (object proposals) that likely contain objects. Significant speed-ups in object detection can be achieved if high recall can be guaranteed with 10^4 or less window boxes. Most current objectness methods are based in some way on segmentation (hence multiple cues such as color and texture), e.g., gpbUCM [22], objectness [2], SelectiveSearch [48] and MCG [4], with a few exceptions such as BING [10] and CPMC [9]. Inspired by the most recent work of EdgeBoxes [54] which yields impressive results by simply measuring the number of edges wholly enclosed in a detection box, we facilitate objectness computation by exploiting the edge structure given by our perceptual edge grouping algorithm. Different to [54] that examines continuity of edge fragments, our edge groups are formed by integrating both proximity and continuity through learning to rank, therefore offering better cues to form object proposals.

SBIR Closely correlated with the explosion in the availability of touchscreen devices, sketch-based image retrieval (SBIR) has become an increasingly prominent research topic in recent years. It conveniently sits between text-based image retrieval (TBIR) that uses textual keywords as search query, and content-based image retrieval (CBIR) that asks users to supply an exemplar image instead. Most prior work on SBIR [14, 24, 16] generally operates as follows: first extract edges from a natural image to approximate a sketch, then local features (e.g., HOG) are extracted on the resulting edge map and the sketch, then finally query and edge-map features are matched (e.g., with KNN). However, very few studies [37] specifically consider the role of sketch generation to bridge the semantic gap. Moreover, none specifically studies how quality of edge maps can influence SBIR performance. In this paper, we demonstrate that our perceptual edge grouping based method for sketch generation reduces the cross-modal gap between edges generated from images and human sketches, and is thus more suitable for SBIR compared to the traditional edge descriptors.

Sketch Segmentation Segmenting free-hand human sketches have profound applications in object modeling, image retrieval and 2D diagram understanding. Parsing sketches is difficult, due to (i) the lack of visual cues on sketches (black and white lines) and (ii) sketches are often abstract therefore hard to learn a model from. As a result, most work relies on auxiliary information. Sun et al. [47] performed segmentation with the help of a million of clipart images. They employed a local and greedy merging strategy based on the proximity. In comparison, our grouper works with multiple Gestalt cues under a global optimization framework. Recently, a data-driven approach [28] was proposed that performs segmentation and labeling simultaneously. It works by first learning from an existing database of 3D models, and segmenting sketches by optimally fitting 3D components onto sketches using mixed integer programming. Our sketch segmentation method works independently of auxiliary dataset and produces plausible segmentation results. Since sketches are essentially clean (and abstract) edges, segmentation quality can also be used as an ideal measure of grouping performance.

3. Perceptual Edge Grouping

We first present how perceptual edge grouping can be cast into a graph partitioning problem. There are two stages: graph construction by ranking, followed by graph partitioning. Given an image, in the first stage, a graph of edges is constructed by using primal RankSVM [23] to integrate two key Gestalt principles, i.e., proximity and continuity. Then a multi-label graph-cut algorithm [5] is used to partition the graph, thus producing an optimal edge grouping.

3.1. Edge Graph Construction

Let us denote an edge graph constructed from image I as $G(\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ represents a set of edges and $\mathcal{E}$ is a set of links that each of them connects a pair of edges in $\mathcal{V}$. There is a score associated with each link, denoted as e_{ij}, which indicates the likelihood that the pair of edges being grouped together.

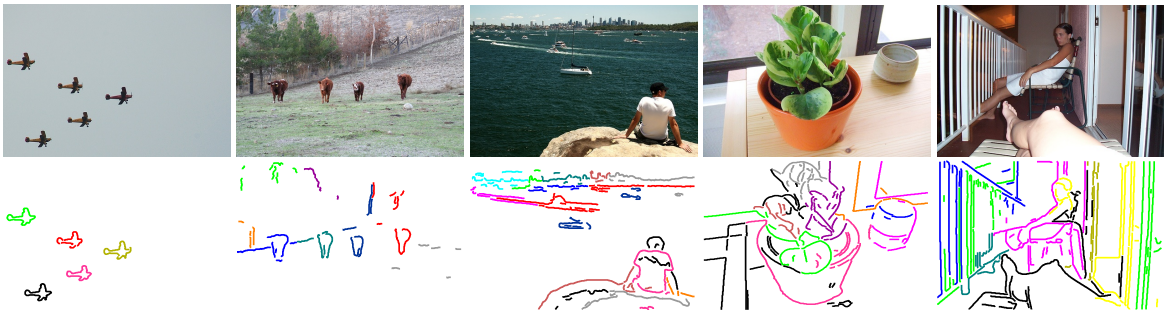

Figure 1. Quantitative examples of our edge grouping results (Bottom) on both simple and complex scenes (Top). The grouping results are color-coded, showing that edges in the same object are grouped together.

Specifically, given an image I, a state-of-the-art edge detector, Structured Edges [12], is used to extract edge map M. Then the set of edges $\mathcal{V}$ are obtained by cutting edges at junction points on edge map M. To estimate the link score e_{ij} in $\mathcal{E}$, a RankSVM model is used to score each pair of edges in $\mathcal{V}$. More precisely, we formulate the problem of estimating the link weight/score e_{ij} as a learning to rank problem. Given any edge $v_i \in \mathcal{V}$, the link score e_{ij} of the edge pair (v_i, v_j) should be greater than the score e_{ik} of another pair of edges (v_i, v_k), if v_i is more likely to be grouped with v_j rather than v_k. Let "edge v_i is preferred to group with v_j rather than v_k" be specified as "$(v_i, v_j) \succ (v_i, v_k)$". The goal is to learn a ranking function $F(\boldsymbol{x}) = \boldsymbol{\omega}^T \boldsymbol{x}$ that outputs a score such that $F(\boldsymbol{x}(v_i, v_j)) > F(\boldsymbol{x}(v_i, v_k))$ for any $(v_i, v_j) \succ (v_i, v_k)$. $\boldsymbol{x}(v_i, v_j)$ is the feature of edge pair (v_i, v_j) and $\boldsymbol{\omega}$ refers to a weight vector adjusting by learning algorithm.

Two key Gestalt principles, proximity and continuity, are employed for representing each edge pair (v_i, v_j), that is, $\boldsymbol{x}(v_i, v_j)$ is a 2-dimensional feature vector, where dimensions correspond to a measurement of continuity (i.e. slope trend) and proximity (i.e. geometry distance), respectively. To learn how to use these two Gestalt principles together to group edges, we train on a subset of a large scale human-drawn sketch database [13]. Edges of each sketch in the training set are manually segmented into semantic-groups. Then preference of edge pairs is obtained as $P = (\hat{\mathcal{V}}^+, \hat{\mathcal{V}}^-)$: For any edge $v_i \in \mathcal{V}$, a positive pair $\hat{\mathcal{V}}^+$ is formed by v_i and another edge $v_j \in \mathcal{V}$ in the same group, denoted as (v_i, v_j), while a relative negative pair $\hat{\mathcal{V}}^-$ is formed by the same edge v_i and a edge v_k from different group, as (v_i, v_k). Therefore, the ranking objective is to fulfill $F(\boldsymbol{x}(v_i, v_j)) > F(\boldsymbol{x}(v_i, v_k))$, i.e. $\boldsymbol{\omega}^T \boldsymbol{x}(v_i, v_j) > \boldsymbol{\omega}^T \boldsymbol{x}(v_i, v_k)$. Given all the preferences, the problem is formulated as follows:

$$\begin{aligned} &minimize: L(\omega, \xi_k) = \frac{1}{2}||\boldsymbol{\omega}||^2 + C\sum \xi_k \\ &subject\ to: \forall P: \boldsymbol{\omega}^T(\boldsymbol{x}_{\hat{\mathcal{V}}^+} - \boldsymbol{x}_{\hat{\mathcal{V}}^-}) \geqslant 1 - \xi_k \\ &\qquad \forall k: \xi_k \geqslant 0 \end{aligned} \tag{1}$$

The above optimization problem can be solved by performing SVM classification on pairwise difference vectors $(\boldsymbol{x}_{\hat{\mathcal{V}}^+} - \boldsymbol{x}_{\hat{\mathcal{V}}^-})$. Given the ranking function $F(\boldsymbol{x}) = \boldsymbol{\omega}^T \boldsymbol{x}$ to weight all the links in graph $G(\mathcal{V}, \mathcal{E})$, we next present how to partition the resulting graph for edge grouping.

3.2. Graph Partition

Edge grouping is treated as a graph partition problem on the edge graph $G(\mathcal{V}, \mathcal{E})$. It is formulated as a min-cut/max-flow optimization problem [5], which seeks to minimize an overall energy function defined as:

$$E(v_L) = \sum_{v_i \in \mathcal{V}} D(v_i, v_L) + \sum_{\{v_i, v_j\} \in N} S(v_i, v_j) \tag{2}$$

$$\begin{aligned} where,\ D(v_i, v_L) &= sigmoid(F(\boldsymbol{x}))^{-1} \\ &= sigmoid(\boldsymbol{\omega}^T \boldsymbol{x}(v_i, v_L))^{-1} \end{aligned} \tag{3}$$

$$S(v_i, v_j) = d(v_i, v_j)^{-1} \tag{4}$$

where v_L is a set of edges that serve as cluster centers, it is randomly initialized, then optimized by graph-cuts. N is the set of pairs of neighboring edges in $\mathcal{V}$. $D(v_i, v_L)$ is the data cost energy measuring the fitness between a edge v_i and the assigned cluster center v_L. The higher the fitness, the lower the cost or penalty. For example, suppose candidate edge pair v_i and v_j tends to group with each other, i.e., they sit close in Gestalt space (high response in continuity and proximity). The resulting ranking score $F(\boldsymbol{x}) = \boldsymbol{\omega}^T \boldsymbol{x}(v_i, v_j)$ tends to be large; an inverse sigmoid function on $F(\boldsymbol{x})$ is thus used which leads to a small value in data cost $D(v_i, v_j)$ to fit into the graph-cuts framework (Eq. 3). $S(v_i, v_j)$ is the smoothness cost as defined in Eq. 4, which measures the spatial correlation between neighboring edges. Edges with a smaller distance have higher probability of belonging to the same group. Between two neighboring edges v_i and v_j, the smoothness energy is defined by the inverse Euclidean Hausdorff-distance between them, i.e., $d(v_i, v_j)^{-1}$.

In summary, we seek the optimal edge grouping by cutting edge graph $G(\mathcal{V}, \mathcal{E})$ constructed above into groups with minimum energy $E(v_L)$ (Eq. 2). Some example edge grouping on complex scenes are shown in Figure 1.

4. Edge Grouping for Objectness

In this section, we describe our perceptual-grouping approach to objectness. Given an image, similar to [54], we adopt the efficient Structured Edge [12] algorithm for edge detection. We then perform our previously described perceptual edge grouping algorithm on the edge map to find edge groups. The key step for objectness is then a scoring procedure for each candidate bounding box based on the enclosed grouped edges. Finally, the ranked bounding boxes are output as the object proposals.

4.1. Generating Bounding Boxes

Similar to [54], we generate candidate bounding boxes in a sliding window manner over position, scale and aspect ratio. The step size for each of the three variables is determined by a parameter α representing the Intersection over Union (IoU) with neighboring boxes. For example, one step size in position, scale and aspect ratio will generate a set of bounding boxes with an IoU of α. In this work, we set $\alpha = 0.65$ as in [54]. The scale ranges from a minimum box area of 1000 pixels to the full image, and the aspect ratio from 1/3 to 3.

Algorithm 1: Closure of edge groups G_b

Input: Bounding box $b = (x, y, b_w, b_h)$, (x, y) specifies the position of box b, b_w and b_h are the width and height of box b; Edge groups G_b is a set of grouped edges in box b.

Output: Closure of G in b: $C(G_b)$

for *each collum c and row r in box b* **do**
- compute the maximum distance formed by the edge pixel at this collum c: $D_c = max_{row}(G_b, c) - min_{row}(G_b, c)$.
- compute the maximum distance formed by the edge pixel at this row r: $D_r = max_{col}(G_b, r) - min_{col}(G_b, r)$.

compute convexhull: $convx(G_b) = (\sum_c D_c + \sum_r D_r)/2$;

compute box area: $A_b = b_w \times b_h$;

compute closure: $C(G_b) = cvhull(G_b)/A_b$;

4.2. Scoring Bounding Boxes

Given a set of generated bounding boxes B, we describe how to measure the probability that an object is contained in a candidate bounding box $b \in B$. Recall that the result of edge grouping should correspond to an object or part thereof. A reasonable criterion to measure objectness is therefore: (i) the closure area of an edge group should occupy the candidate bounding box as much as possible, since object boundaries often form closed regions [19, 32], (ii) the structural complexity of edge groups under a candidate box is relatively simple, since ideally each edge group should correspond to exactly one object (or their parts).

Specifically, given a bounding box b, we denote G_b as the grouping results of edges covered in b. Closure $C(G_b)$ is defined as the ratio of convex hull of the edge group $cvhull(G_b)$ to the area covered by bounding box A_b, i.e. $cvhull(G_b)/A_b$. Algorithm 1 shows how to obtain closure corresponding to a candidate bounding box. The structural complexity of edges is defined as the number of groups per unit area of the bounding box n/A_b, where n is the number of edge groups in box b. To this end, our scoring function for the box b, with width b_w and height b_h, is:

$$P_b = \frac{C(G_b)}{n/A_b} = \frac{cvhull(G_b)}{n(b_w \times b_h)^2} \tag{5}$$

4.3. Post Processing

Inspired by previous work on objectness, there are two necessary post processing steps for generating better proposals: refinement and Non-Maximal Suppression (NMS). Since the strategy of generating candidate bounding boxes may misalign them to objects, refinement is used to further improve P_b in a greedy iterative search manner by justifying position, scale and aspect ratio of bounding boxes, similar to that performed in [54]. In addition, to generate a relatively small set of objectness proposals with *high recall*, we employ Non-Maximal Suppression (NMS) to prune the candidate list. A box b_i is removed if there exists another box b_j with a greater score, where the IoU of b_i and b_j is more than a threshold β.

5. Edge Grouping for SBIR

In this section, we present how edge grouping is used for sketch-based image retrieval (SBIR). This is a challenging task because of the cross-domain gap between objects in sketch and real images. Sketch generation, or converting real images into sketch-like images is necessary to make sketch-based image retrieval feasible. Given elementary edge detection, edge grouping is the key step in sketch generation. This is then followed by a filtering process to generate the sketch. To perform SBIR, histogram of oriented gradients (HOG) H^I is extracted for each machine generated sketch, and query sketch H^s. Gallery images are then ranked according to the χ^2 histogram distance $d(H^s, H^I)$ between a query sketch and the synthesized sketch.

Next we describe how sketches can be generated from real images using the proposed perceptual grouping framework. There are three main stages for automatic sketch generation as follows.

Extracting edge map globalPb [3] is used for edge detection since it delivers the best semantic edges. The

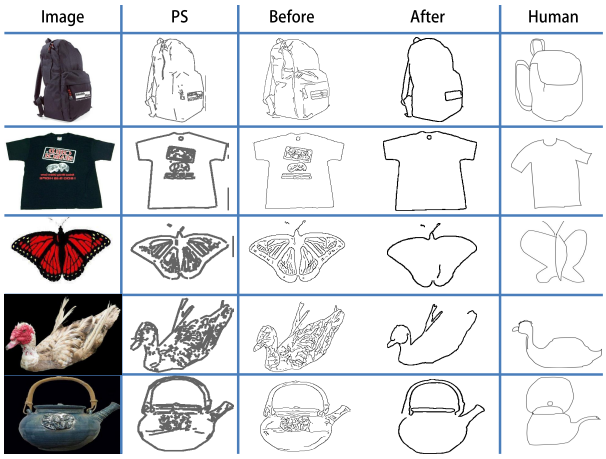

Figure 2. Sketch examples. From left to right: original image, primal sketch [20], input to our model (before), and our generated sketch (after), human free hand drawn sketch. We can observe that sketches generated using our methods keep a similar level of details as those from humans.

edge map is further transformed to edge fragments $\mathcal{V} = \{v_1, v_2, \ldots, v_n\}$ by cutting at points of high curvature.

Edge grouping We then perform the proposed perceptual grouping framework on edge fragments $\mathcal{V}$, aims to group salient edges together, hence separate them from noise. Importantly, our RankSVM and graph-cut model is learned to fuse cues and resolve Gestalt confliction. This results in a better grouping result from considering the two Gestalt principles simultaneously, i.e., continuity and proximity.

Sketching by group-based filtering Given a set of edge groups after boundary grouping, our goal is to filter redundancy to generate human-drawing-like sketches. Inspired by [51] which finds salient contours by ratios measuring gaps, continuation and length among contour segments, we formulate an energy function to measure the coarseness level of edge groups. Only groups with low level coarseness are kept as the generated sketch. More specifically, for a group of boundaries $G_i \in G$, the energy function is formulated as $E(G_i) = \frac{|h|}{S}$, where $|h|$ indicates the number of high curvature turning points in group G_i and S represents the total length of all curve segments in group G_i. With this procedure, sketches are generated from real images. Some examples are shown in Figure 2, along with sketches generated by other methods for comparison.

6. Stroke Grouping for Sketch Segmentation

In this section, we present how the proposed grouping framework is capable of sketch segmentation at object part level. More specifically, the goal is to group human drawn strokes of an object into semantic parts automatically.

Given a sketch $S = \{s_1 \ldots s_n\}$ which consists of n strokes, we aim to achieve sketch segmentation by exploiting our perceptual grouping framework. Intuitively, similar to grouping edge fragments for objectness, strokes belonging to the same object part should sit close in the Gestalt space. Therefore, for each pair of strokes (s_i, s_j), the previously learned ranking function $F(\boldsymbol{x}) = \boldsymbol{\omega}^{\boldsymbol{T}}\boldsymbol{x}(s_i, s_j)$ (see Section 3.2) is used to score how likely they are belonging to the same part. Afterwards, this score is fed into the overall grouping framework by Eq. 3. This is followed by solving the optimization problem in Eq. 2, which gives us the final discovered object parts (i.e., stroke groups).

7. Experiments

In this section, we present experimental results of our approaches for objectness, SBIR and sketch segmentation based on our perceptual edge grouping framework.

7.1. Object Proposal Generation

Dataset and Settings The Pascal VOC 2007 test dataset, which has about 5000 unconstrained real images in 20 categories, is used to evaluate our proposed objectness approach for generating object proposals. To score a correct match, we use fraction of a ground truth annotation covered by a candidate box, informally called intersection over union (IoU). An object is successfully discovered by the candidate box if IoU above a threshold. We compare against five state-of-the-art objectness methods, namely MCG [4], SelectiveSearch [48], objectness [2], EdgeBoxes [54], and BING [10]. MCG ranks candidate boxes by merging segments upon the multi-scale hierarchical segmentation, SelectiveSearch carefully designs features and scores formulations to greedily merge low-level superpixels into regions, and objectness ranks candidate boxes based on a combination of cues, e.g., saliency, color, location, how much such windows overlap with low-level segments, etc. These three are segmentation based methods. In contrast, EdgeBoxes and BING work without segmentation. EdgeBoxes discovers objects by counting the number of edges wholly enclosed in box, while BING trains a linear classifier over edge features which is then applied in a sliding window manner.

Results The results of top 10^4 proposals are shown in Table 1 where performance is evaluated using three metrics: Recall, AUC, and Jaccard index as in [54]. It can be seen that: (i) Our approach achieves the best recall (over 90%) when IoU threshold is from 0.5 to 0.7. This is significant because as mentioned previously, since missed objects will never be rediscovered, it is critical for object proposal method to have high *recall* of objects. (ii) Overall the performance is better than the closely related EdgeBoxes method [54], demonstrating the usefulness of perceptual grouping for edge-based object proposal generation. (iii) Although overall slightly better results are obtained by MCG and SelectiveSearch, they are built upon multiple-cues

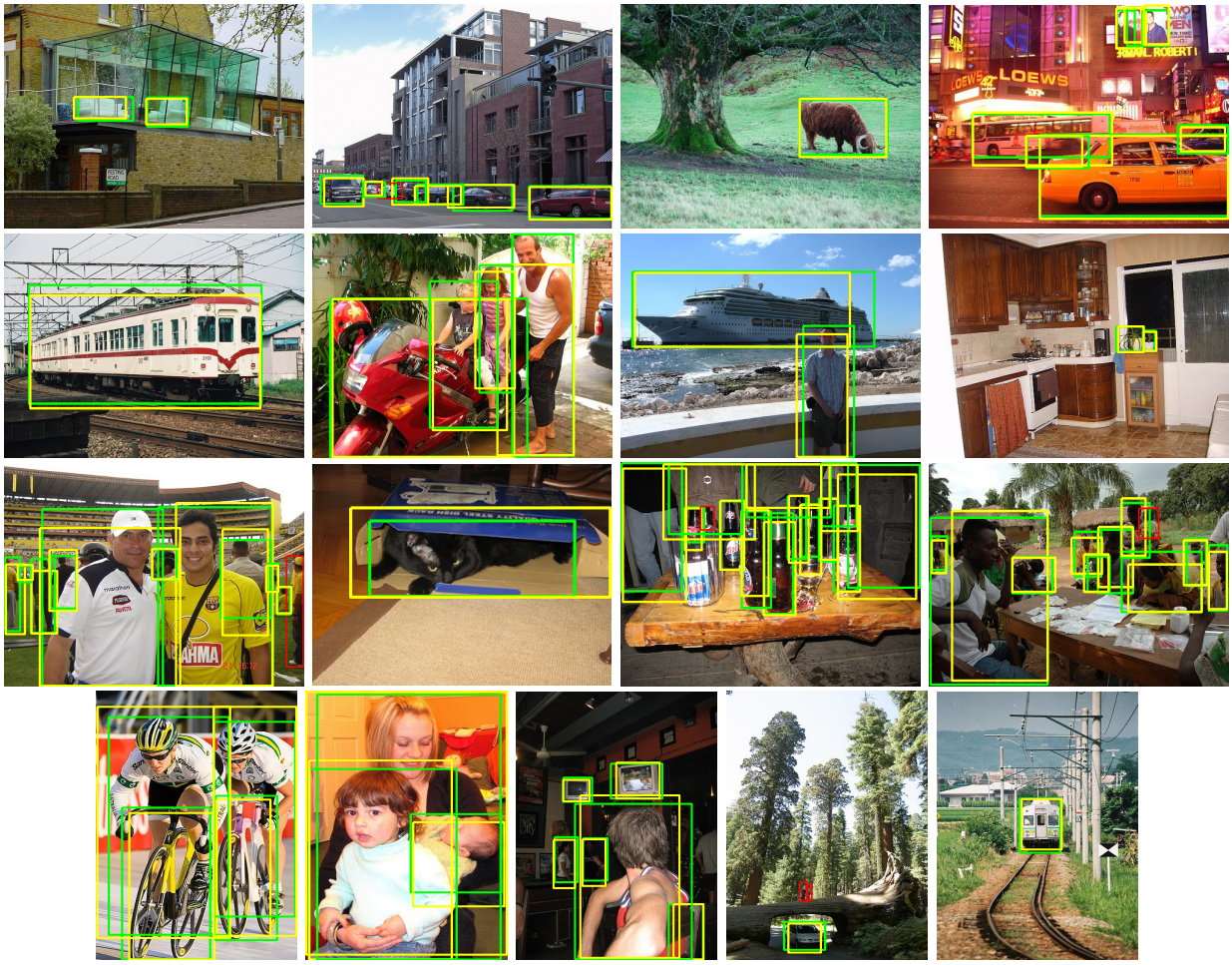

Figure 3. Qualitative examples of our edge grouping based objectness measure. An IoU threshold of 0.65 is used to determine if correctly discover an object here. Yellow boxes are the best produced candidates of our method. Ground truth bounding boxes are shown in green and red, while green indicates correctly discovering an object and red indicates failure of the detection.

	IoU=0.5			IoU=0.6			IoU=0.7		
Methods	Recall(%)	AUC(%)	J_i(%)	Recall(%)	AUC(%)	J_i(%)	Recall(%)	AUC(%)	J_i(%)
PerceptualEdge (Ours)	95.84	28.25	79.47	92.90	18.77	80.23	83.87	9.83	81.72
PerceptualEdge-proximity	92.25	28.25	78.69	88.62	17.42	79.65	81.97	8.82	80.77
PerceptualEdge-continuity	90.53	25.41	78.06	86.15	16.57	79.22	77.70	8.30	80.68
BING [10]	96.39	15.33	65.90	67.65	6.79	70.03	27.34	2.19	78.01
EdgeBoxes50 [54]	93.13	21.06	72.61	83.62	12.13	74.50	54.04	5.09	79.42
EdgeBoxes [54]	94.77	28.13	79.68	91.15	18.82	80.64	81.65	10.06	82.32
MCG [4]	93.61	31.44	83.60	87.42	22.39	85.60	77.56	14.08	88.16
Objectness [2]	83.92	15.79	68.81	69.90	7.98	71.41	34.74	2.60	77.48
Sel.Search [48]	91.51	31.04	83.91	85.82	22.15	85.81	77.17	13.96	88.09

Table 1. Comparison of top 10^4 proposals with state-of-the-art on Recall, AUC and Jaccard index at instance level(J_i). J_i is defined as the mean best overlap for all the ground truth instances in the test set, to reflect the quality of generated bounding boxes.

including color and texture, whilst our method exploits edge cue only. Our approach to objectness is thus more generally applicable to situations where no additional cues are available, e.g., images in black and white or sketches. It can also be seen that recall drops when only single Gestalt cues are used for grouping, which further confirms the benefits of using multiple cues. Some qualitative results are shown in Figure 3.

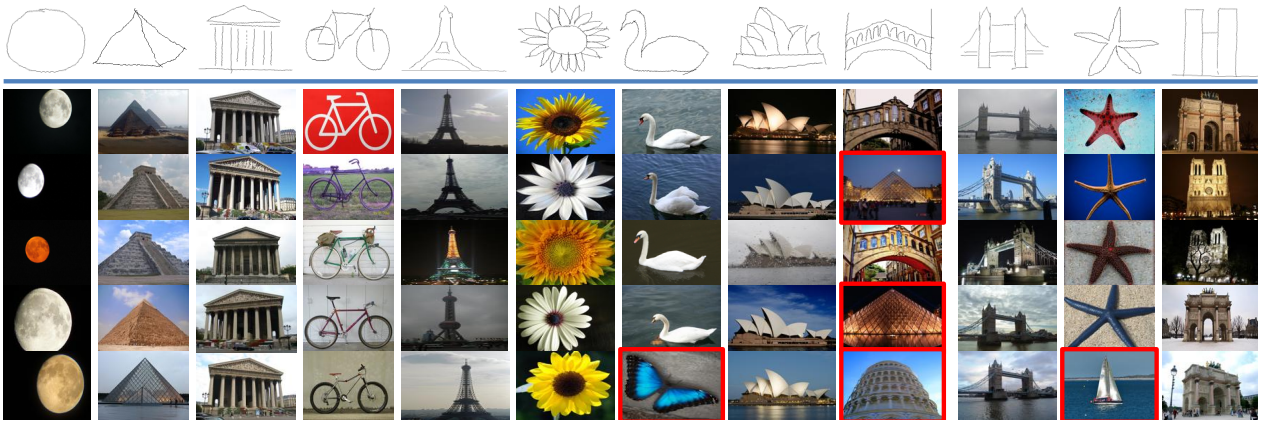

Figure 4. Example query sketch, and their top ranking results (from top to bottom) in the Flick15K dataset. Red boxes show false positives.

Methods	Vocabulary size	MAP
PeceptualEdge (Ours)	non-BoW	**0.1837**
PeceptualEdge-proximity	non-BoW	0.1602
GF-HOG [25]	3500	0.1222
HOG [11]	3000	0.1093
SIFT [31]	1000	0.0911
SSIM [42]	500	0.0957
ShapeContext [35]	3500	0.0814
StructureTensor [15]	500	0.0798
PeceptualEdge-continuity	non-BoW	0.0789
StructureTensor [15]	non-BoW	0.0735

Table 2. SBIR results comparison (MAP) against single-cue grouping and state-of-the-art alternatives.

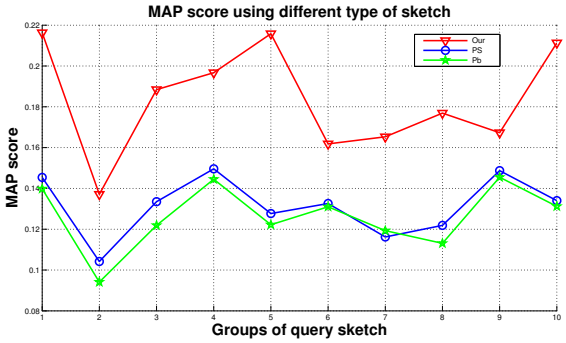

Figure 5. MAP performance comparison of our generated sketch, Primal sketch [20] (PS) and Pb [3].

7.2. Sketch-Based Image Retrieval

Dataset The Flickr15k dataset [25] serves as the benchmark for our sketch-based image retrieval system. This dataset consists of: (i) about 15k photographs sampled from Flickr and manually labeled into 33 categories based on shape; and (ii) 330 free-hand drawn sketch queries drawn by 10 non-expert sketchers. In our case, we use the real images in Flickr15k as retrieval candidates, and the 330 sketches without the semantic tags to serve as queries.

Settings We compare our proposed SBIR system using perceptual-grouping based sketch generation (PerceptualEdge) with a state-of-the-art non-BoW method, StructureTensor [15]; and six other BoW methods: Gradient Field HOG (GF-HOG) [25] which is the state-of-the-art BoW-based method, SIFT [31], Self Similarity (SSIM) [42], Shape Context [35], HOG [11] and the Structure Tensor [15]. Similar to [25], (i) for the non-BoW baseline (non-BoW StructureTensor), we compute the standard HOG descriptor over all edge pixels of the query sketch and real images to be retrieved, and then the ranking retrieval results are obtained based on the distance between them; (ii) for the six BoW-based baseline methods, all employ a BoW strategy but with different feature descriptors. E.g., for GF-HOG, features of GF-HOG are extracted over all edge pixels of Canny edge map, then a BoW vocabulary $\mathcal{V}$ is formed via k-means, and a frequency histogram H^I is built for each real image using the previously learned vocabulary $\mathcal{V}$. Similarly, a frequency histogram H^s of the query sketch is constructed using the same vocabulary $\mathcal{V}$. Real images are then ranked according to histogram distance to the query $d(H^s, H^I)$. In addition to the baselines, we also compare two single-cue (proximity only, continuity only) variants of our approach to demonstrate the importance of integrating multiple grouping cues.

Because most previous work on SBIR relies on one edge detector (e.g. Canny) and just focuses on feature extraction [25, 42, 35, 11], the problem of how sketch generation effects retrieval performance has been largely ignored. We therefore further investigate that how different types of sketch generator contribute to retrieval performance. In particular, we offer comparison of three sketch generation techniques, namely Pb [3], PS [20] and our proposed approach.

Results Quantitative and qualitative results are shown in Table 2 and Figure 4, respectively. Table 2 reports the Mean Average Precision (MAP), produced by averaging the Average Precision (AP) over all the 330 sketch queries. We

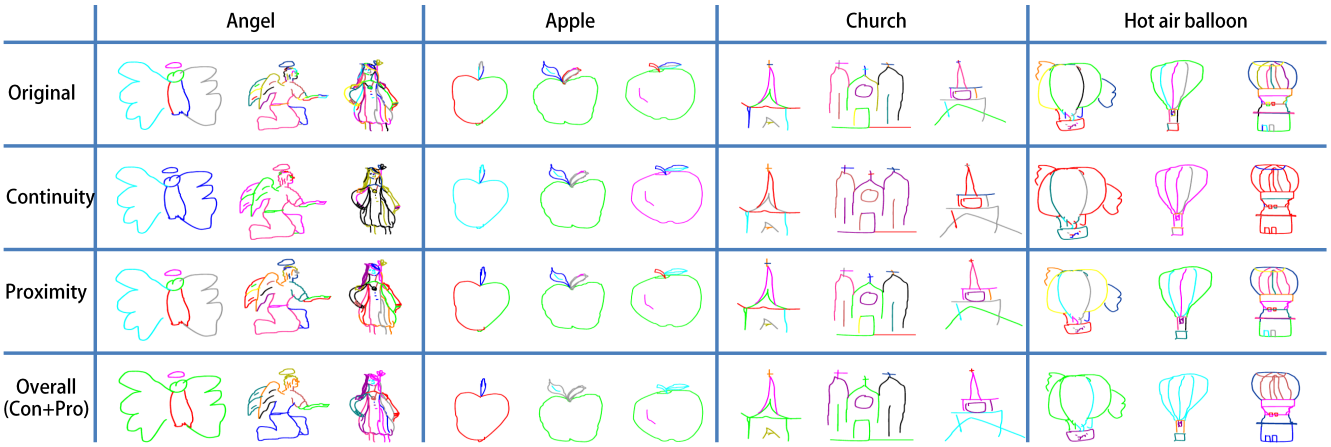

Figure 6. Example sketch segmentation results of four object categories. Left to right: angel, apple, church and hot air ballon; top to bottom: original sketch, continuity only, proximity only and overall result using both cues.

can observe from Table 2 that our proposed SBIR method achieves 0.1837 MAP, outperforms all the baseline methods. In particular, the proposed method offers an over 2-fold improvement compared to the state-of-the-art non-BoW method (i.e. non-BoW StructureTensor). It is also worth noting that retrieval performance drops when only one grouping cue is considered, and the proximity cue seems to have played a more dominant role. Figure 4 presents several sketch queries and their retrieval results over the Flickr15k dataset. We can observe that the returned top ranking images correspond closely to the query sketches shape. Although there are some inaccuracies (e.g. between starfish and sailing boat), the majority of results are relevant. Finally, Figure 5 compares the retrieval performance when different types of sketch generator are used in the same SBIR system. It clearly shows that our proposed sketch generator is superior to the other competitors over all the 10 groups of query sketches.

7.3. Sketch Segmentation

Dataset We employ a subset of the large scale human-drawn sketch database [13] to perform the sketch segmentation experiment. To cover the diversity of the sketch database, we chose 25 categories, including "apple" and "hot air balloon" which are simple cases, and "angel" and "church" which exhibit relatively complex structures. The original sketch stroke information provided by the database is used as input to perceptual grouping instead of edge fragments.

Settings In line with the experiments on objectness and SBIR, we conduct sketch segmentation experiments using either proximity or continuity alone, or both synergistically. Since no ground truth is available, we offer qualitative illustrations instead. Because sketches are essentially clean (and abstract) edges, segmentation quality can be used an ideal qualitative measure for the grouping performance. Quantitative measures of the proposed grouper can be found in objectness and SBIR experiments instead.

Results The qualitative results of our sketch segmentation method are illustrated in Figure 6. It can be observed that our grouper is able to generate highly plausible segmentations of objects. For simple categories such as "apple" and "hot air balloon", segmentations clearly exhibit a top-to-bottom structure. On complex objects such as "church" and "angel", our grouper also produces reasonable object parts. It is interesting to note that on "angel" where object structure is highly complex, our segmentation still depicts parts that are semantically meaningful, such as wings, body, arms and legs. In contrast, segmentations using single Gestalt perform consistently worse, again confirms the benefits of our multiple Gestalt framework.

8. Conclusion

In this paper, we have proposed a unified approach for perceptual edge grouping. Two commonly used Gestalt principles are integrated through a ranking strategy to construct an edge graph, followed by a multi-label graph cut algorithm partitions the graph. Based on the proposed edge grouper, three applications on object proposal generation, SBIR, and free-hand sketch segmentation are demonstrated. The experimental results validate the effectiveness of our perceptual edge grouping framework for these tasks.

Acknowledgment

This work was partially supported by National Natural Science Foundation of China under Grant No.61273217, 61175011, 61171193, 61402047, 61511130081 and the 111 project under Grant No.B08004.

References

[1] N. Adluru, L. J. Latecki, R. Lakämper, T. Young, X. Bai, and A. D. Gross. Contour grouping based on local symmetry. In *ICCV 2007*.

[2] B. Alexe, T. Deselaers, and V. Ferrari. Measuring the objectness of image windows. *TPAMI*, 2012.

[3] P. Arbelaez, M. Maire, C. Fowlkes, and J. Malik. Contour detection and hierarchical image segmentation. *TPAMI 2011*.

[4] P. A. Arbeláez, J. Pont-Tuset, J. T. Barron, F. Marqués, and J. Malik. Multiscale combinatorial grouping. In *CVPR*, 2014.

[5] Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *TPAMI 2001*.

[6] E. Brunswik and J. Kamiya. Ecological cue-validity of'proximity'and of other gestalt factors. *The American journal of psychology*, 1953.

[7] D. C. Burr, M. C. Morrone, and D. Spinelli. Evidence for edge and bar detectors in human vision. *Vision research*, 29, 1989.

[8] J. Canny. A computational approach to edge detection. *TPAMI*, 1986.

[9] J. Carreira and C. Sminchisescu. CPMC: automatic object segmentation using constrained parametric min-cuts. *TPAMI*, 2012.

[10] M. Cheng, Z. Zhang, W. Lin, and P. H. S. Torr. BING: binarized normed gradients for objectness estimation at 300fps. In *CVPR*, 2014.

[11] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *CVPR 2005*.

[12] P. Dollár and C. L. Zitnick. Structured forests for fast edge detection. In *ICCV*, 2013.

[13] M. Eitz, J. Hays, and M. Alexa. How do humans sketch objects? *SIGGRAPH 2012*.

[14] M. Eitz, K. Hildebrand, T. Boubekeur, and M. Alexa. A descriptor for large scale image retrieval based on sketched feature lines. In *SBIM 2009*.

[15] M. Eitz, K. Hildebrand, T. Boubekeur, and M. Alexa. An evaluation of descriptors for large-scale image retrieval from sketched feature lines. *Computers & Graphics 2010*.

[16] M. Eitz, K. Hildebrand, T. Boubekeur, and M. Alexa. Sketch-based image retrieval: Benchmark and bag-of-features descriptors. *TVCG 2011*.

[17] J. H. Elder and R. M. Goldberg. Ecological statistics of gestalt laws for the perceptual organization of contours. *Journal of Vision*.

[18] J. H. Elder, A. Krupnik, and L. A. Johnston. Contour grouping with prior models. *TPAMI*, 2003.

[19] J. H. Elder and S. W. Zucker. Computing contour closure. In *ECCV*, 1996.

[20] C. en Guo, S. C. Zhu, and Y. N. Wu. Primal sketch: Integrating structure and texture. *CVIU 2007*.

[21] W. Geisler, J. Perry, B. Super, and D. Gallogly. Edge co-occurrence in natural images predicts contour grouping performance. *Vision research, 2001*.

[22] C. Gu, J. J. Lim, P. Arbelaez, and J. Malik. Recognition using regions. In *CVPR*, 2009.

[23] R. Herbrich, T. Graepel, and K. Obermayer. Large margin rank boundaries for ordinal regression. *Advances in neural information processing systems*, 1999.

[24] R. Hu, M. Barnard, and J. P. Collomosse. Gradient field descriptor for sketch based retrieval and localization. In *ICIP 2010*.

[25] R. Hu and J. P. Collomosse. A performance evaluation of gradient field hog descriptor for sketch based image retrieval. *CVIU 2013*.

[26] M. Kaschube, F. Wolf, T. Geisel, and S. Löwel. The prevalence of colinear contours in the real world. *Neurocomputing, 2001*.

[27] I. Kokkinos. Boundary detection using f-measure-, filter- and feature- (f^3) boost. In *ECCV*, 2010.

[28] R. W. H. Lau. Data-driven Segmentation and Labeling of Freehand Sketches. *SIGGRAPH Asia*, 2014.

[29] M. Leordeanu, R. Sukthankar, and C. Sminchisescu. Generalized boundaries from multiple image interpretations. *TPAMI*, 2014.

[30] J. J. Lim, C. L. Zitnick, and P. Dollár. Sketch tokens: A learned mid-level representation for contour and object detection. In *CVPR*, 2013.

[31] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *IJCV 2004*.

[32] S. Mahamud, L. R. Williams, K. K. Thornber, and K. Xu. Segmentation of multiple salient closed contours from real images. *TPAMI*, 2003.

[33] J. Mairal, M. Leordeanu, F. Bach, M. Hebert, and J. Ponce. Discriminative sparse image models for class-specific edge detection and image interpretation. In *ECCV*, 2008.

[34] D. R. Martin, C. Fowlkes, and J. Malik. Learning to detect natural image boundaries using local brightness, color, and texture cues. *TPAMI*, 2004.

[35] G. Mori, S. J. Belongie, and J. Malik. Efficient shape matching using shape contexts. *TPAMI 2005*.

[36] G. Papari and N. Petkov. Adaptive pseudo-dilation for gestalt edge grouping and contour detection. *TIP 2008*.

[37] Y. Qi, J. Guo, Y. Li, H. Zhang, T. Xiang, and Y.-Z. Song. Sketching by perceptual grouping. In *ICIP 2013*.

[38] X. Ren and L. Bo. Discriminatively trained sparse code gradients for contour detection. In *NIPS*, 2012.

[39] X. Ren, C. Fowlkes, and J. Malik. Learning probabilistic models for contour completion in natural images. *IJCV*, 2008.

[40] K. Schindler and D. Suter. Object detection by global contour shape. *Pattern Recognition*, 41, 2008.

[41] R. M. Shapley and D. J. Tolhurst. Edge detectors in human vision. *The Journal of physiology*, 229, 1973.

[42] E. Shechtman and M. Irani. Matching local self-similarities across images and videos. In *CVPR 2007*.

[43] M. Sigman, G. A. Cecchi, C. D. Gilbert, and M. O. Magnasco. On a common circle: natural scenes and gestalt rules. *PNAS, 2001*.

[44] Y.-Z. Song, B. Xiao, P. M. Hall, and L. Wang. In search of perceptually salient groupings. *TIP 2011*.

[45] J. S. Stahl and S. Wang. Edge grouping combining boundary and region information. *TIP*, 2007.

[46] J. S. Stahl and S. Wang. Globally optimal grouping for symmetric closed boundaries by combining boundary and region information. *TPAMI*, 2008.

[47] Z. Sun, C. Wang, L. Zhang, and L. Zhang. Free hand-drawn sketch segmentation. In *ECCV*. Springer, 2012.

[48] J. R. R. Uijlings, K. E. A. van de Sande, T. Gevers, and A. W. M. Smeulders. Selective search for object recognition. *IJCV*, 2013.

[49] J. Wagemans, J. H. Elder, M. Kubovy, S. E. Palmer, M. A. Peterson, M. Singh, and R. von der Heydt. A century of Gestalt psychology in visual perception: I. Perceptual grouping and figure-ground organization. *Psychological bulletin 2012*.

[50] J. Wagemans, J. Feldman, S. Gepshtein, R. Kimchi, J. R. Pomerantz, P. A. van der Helm, and C. van Leeuwen. A century of Gestalt psychology in visual perception: II. Conceptual and theoretical foundations. *Psychological Bulletin 2012*.

[51] S. Wang, T. Kubota, J. M. Siskind, and J. Wang. Salient closed boundary extraction with ratio contour. *TPAMI 2005*.

[52] M. Wertheimer. *Laws of organization in perceptual forms*. Harcourt, Brace & Jovanovitch, London, 1938.

[53] Q. Zhu, G. Song, and J. Shi. Untangling cycles for contour grouping. In *ICCV*, 2007.

[54] C. L. Zitnick and P. Dollár. Edge boxes: Locating object proposals from edges. In *ECCV*, 2014.

[55] C. L. Zitnick and D. Parikh. The role of image understanding in contour detection. In *CVPR*, 2012.