# SKETCHING BY PERCEPTUAL GROUPING

*Yonggang Qi*[†] *Jun Guo*[†] *Yi Li*[⋆] *Honggang Zhang*[†] *Tao Xiang*[⋆] *Yi-Zhe Song*[⋆]

[†] School of Information and Communication Engineering, BUPT, Beijing, China
[⋆] School of EECS, Queen Mary, University of London, London E1 4NS, U.K.

## ABSTRACT

Sketch is used for rendering the visual world since prehistoric times, and has become ubiquitous nowadays with the increasing availability of touchscreens on portable devices. However, how to automatically map images to sketches, a problem that has profound implications on applications such as sketch-based image retrieval, still remains open. In this paper, we propose a novel method that draws a sketch automatically from a single natural image. Sketch extraction is posed within an unified contour grouping framework, where perceptual grouping is first used to form contour segment groups, followed by a group-based contour simplification method that generate the final sketches. In our experiment, for the first time we pose sketch evaluation as a sketch-based object recognition problem and the results validate the effectiveness of our system over the state-of-the-arts alternatives.

***Index Terms*—** Sketch, Contour grouping, Sketch-based retrieval

## 1. INTRODUCTION

There exists plenty of prior work on sketches in computer vision, from the pioneering work of David Marr on primal sketches [1] to sketch-based image retrieval [2][3]. Recent work on sketches includes free hand-drawn sketch segmentation [4], and using sketches as feature descriptors [5]. Nonetheless, how to make machines draw sketches as humans do is still an open problem. Solving such a problem opens the doors for many applications such as sketch-based image retrieval, e.g. human drawn sketches are used to retrieve natural images containing the same object category and vice versa.

Early work on automatic sketching takes a contour detection and object segmentation approach [6, 7, 8, 9], which aims to produce curves that perfectly depicts an image, or constitute global object profiles. Abelaez et al. [6] proposed a general framework to transform the output of any contour detector into a hierarchical region tree with the intention to generate object contours that are most similar to human object segmentation. Zhu et al. [7] exploited the inherent topological 1D structure of salient contours. The grouping is performed by eigen-decomposition of contour grouping graphs. However, object contours or profiles are intrinsically different from human sketches. In particular, compared with object contours, human sketches exhibit higher variances in terms of style, viewpoint and abstraction level (see Fig. 1). Importantly, an object contour is often only a subset of human sketch which typically includes additional details inside the contour. In this paper we aim to address the problem of automatically generating sketches from just a single image.

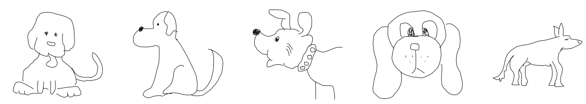

**Fig. 1**. Example human sketches of a “dog” [10]

Recently the problem of generating sketches rather than object segmentations out of natural images started to attract attention. Marvaniya et al. [11] proposed to sketch an object given a set of images of the same object category. The essential idea behind this work is to discover repeatable salient contours across the set of images of the same object class. In contrast, our method only requires a single image. This makes our method much more generally applicable. The work by Guo et al. [12] is probably the most related work to ours and represents the current state-of-the-art in sketch generation from a single image. It combines two generative models learned from natural image statistics, sparse coding model and Markov random field model, for representing geometric structures and stochastic textures respectively to produce sketch. In contrast, in the work we show that by exploiting perceptual grouping principles, a much simpler method can be developed which is able to generate better sketches.

More specifically, we treat automatic sketch generation as a perceptual grouping and filtering problem. The concept of perceptual grouping originates from psychological studies. The gestalt psychologists were the first to draw attention to the phenomenon that human visual system is very powerful in that it can easily find sense from chaos [13]. Computer vision researchers often borrow concepts found in perceptual grouping to tackle problems such as image segmentation [14], contour grouping [9] and object recognition [8]. In this paper, we use the principle of good continuation both to form contour groups, and importantly to filter the formed contour segment groups to generate sketches. Although being intu-

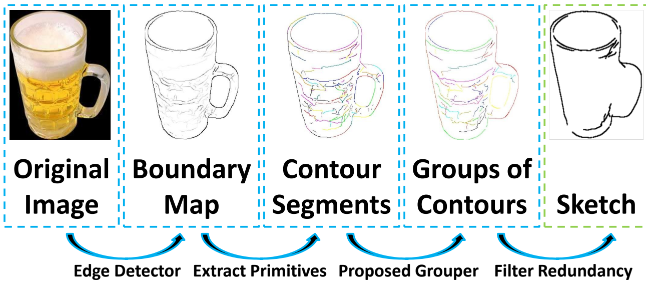


**Fig. 2**. Overview of sketch drawn process.

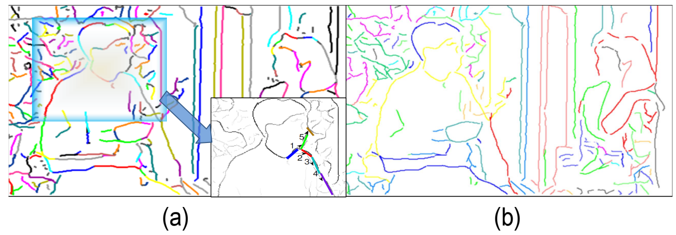


**Fig. 3**. (a) Color-coded Roughly Straight Contour Segments and T-step forecasting(bottom right). (b) Grouping Result.

itive, applying the good continuation principle for grouping is not straightforward. This is because when measured locally, continuation between neighboring contour segments are susceptible to noise which may lead to over-segmentation.To overcome this limitation, a novel T-step forecasting strategy is formulated to measure the continuation of line segmentation over a local support region for robustness. The continuation measure is encapsulated in a graph cut framework for contour grouping, which is followed by group filtering to generate the final sketch. Essentially, the underlying hypothesis is that perceptual grouping is able to find sense out of chaos, leaving only signals corresponding to sketches of human resemblance. Thus, we filter using information accumulated in the grouping process [15]. The entire process is shown in Fig.2.

Existing work [7, 11] evaluates sketching from natural images by comparing computer generated sketches with those generated by humans from the same images. This evaluation strategy is clearly not appropriate for evaluating how well computer generated sketches matches human ones. This is because when a human draw a sketch of an object category, he or she does not need to copy from natural images. Therefore matching automated sketching with human sketching without references to natural images is more relevant to the sketch-based retrieval applications. However, since different people can draw the same object categories very differently (see Fig. 1), choosing which one to match becomes a problem. To solve this problem, we evaluate the quality of automated sketching by measuring how well an sketch classifier learned from a large dataset of human sketches [10] can classify computer generated sketches. Our results show that the sketches generated using our method outperforms a number of state-of-the-arts alternatives.

In summary, our contributions are: (1) For the first time, perceptual grouping principles are applied to contour grouping for automated sketch generation. (2) We formulate a novel measure of the gestalt rule of good continuation. By using by a forecasting strategy to extend the continuation measure beyond neighboring contour segments, this measure is more robust to contour extraction noise. (3) A novel evaluation strategy is devised which is able to evaluate the automatically generated sketches quantitatively and directly compare against human drawn sketches of object categories.

## 2. A MULTI-LABEL GRAPH-CUTS BASED MODEL FOR DRAWING SKETCH

We cast the problem of contour grouping into a graph optimization problem. Given a graph $G = (V, E)$, this optimization problem is solved by a multi-label graph-cuts algorithm [16][17]. Similar to [18], there are two kinds of vertices $V$ in $G$: one is a set of contour segments $Q$ given by dividing the image boundary map into small segments; the other is a set of possible labels $L$ assigned to contour segments. Consequently, edges $E$ in graph $G$ are also divided into two types: n-links between neighboring contour segments and t-links between contour segments and labels. The overall energy function is then defined as:

$$E(L) = \sum_{q \in Q} D(q, L) + \sum_{\{q,p\} \in N} V_{qp} \tag{1}$$

where $p, q$ correspond to neighboring individuals in $Q$. The first term of the equation is the data cost which measures the fitness between the contour segment $q$ and the possible assigned label $L$, the second term is the smoothness cost which indicates the spatial correlation between neighboring contour segments $q, p$.

### 2.1. Extracting grouping primitives

Roughly straight contour segments serve as the grouping primitives as shown in Fig.3(a). Inspired by [8], boundaries are first detected by the Berkeley natural boundary detector [19], then chained according to the connectedness criterion to form the final contour segments. In contrast to [8], rather than linking the roughly straight contour segments into complex junctions, i.e. L-junctions (k=2), T-junctions (k=3), and higher order junctions, we use individual straight contour segments as our grouping primitives and leave second-level grouping to our proposed grouper.

### 2.2. T-step forecasting for measuring continuation

Good continuation is a very important grouping principle for human to perform contour grouping [13]. Based on this principle contour segments are grouped to form smooth contours. To measure contour continuation in a computational framework, the easiest approach is to treat it as a local cue and examine whether any neighboring contour segments form a smooth contour.

However, it is often not the optimal grouping strategy if one only considers two neighboring contour segments. For example, as shown in Fig.3(a) bottom right, the blue segment will be more likely to be grouped together with the green segment for that the continuity between blue segment and red segments is poor, while if we look at a bigger scale, grouping the blue and red segments together becomes a better choice, because they will form a longer smooth curve (1 to 4). In order to utilize continuation more globally, we introduce a forecasting strategy. More specifically, when measuring the continuity between two contour segments, not only the geometric relationship between these two but all nearby segments are considered. That is, for every pair of contour segments, along with the direction that they point to, we will walk T more steps to confirm the continuity of pair of contour segments. T is set to 5 in this work. Algorithm 1 summarizes this procedure.

**Algorithm 1:** T-step forecasting

**Input**: neighboring segments p and q.
**Output**: T-step forecasting continuity between p and q.
Compute curvature difference: $C = curvratio(q, p)$;
**while** $i < T$ **do**
  Find neighboring contour segments set $\Omega$ to p;
  **for** *each* $o \in \Omega$ **do**
    compute $\hat{C} = curvratio(p, o)$ ;
  find $o = argmax_o \hat{C}$, set $C_i = \hat{C}_{max}$;
  update $i = i + 1$, $p = o$;
The overall continuity: $C_{\{q,p\} \in N}(q, p) = C \prod_{i=1}^{T} C_i$;

## 2.3. Contour grouping by graph cut

The optimal label $L$ is given by solving the optimization problem formulated by eq. (1), where the energy function to be optimized has a data cost item and a smoothness term.

The data cost term measures the fitness between the line segment $q$ and the possible assigned label $L$. The higher the fitness, the lower the cost or penalty. Given the possible groupings, the data cost item is naturally obtained correspondingly. More specifically, in our case:

$$D(q, L) = 1 - \frac{1}{|p|} \sum_{p \in L} C_{\{q,p\} \in N}(q, p) \tag{2}$$

where $|p|$ is the number of contour segments in the possible group.

The smoothness cost term measures the spatial correlation between neighboring elements. Elements with a smaller distance have higher probability of belonging to the same gestalt group. Between two neighboring elements $q$ and $p$, the smoothness energy is defined by the inverse Euclidean Hausdorff-distance between them, which is similar to the one used in [20].

$$V_{qp} = d(q, p)^{-1} \tag{3}$$

Given the two terms the optimization problem eq. (1) is solved using a method similar to [16]. A color-coded grouping result of Fig.3(a) can be found in Fig.3(c).

## 2.4. Sketching by group-based filtering

Human draw sketches in an abstract way which is usually geometric far from real models [10], while in this paper, it is based on a more simpler assumption: important groups of contours are perceptually salient. Inspired by [21] which finds salient contours by ratio contour that measures gaps, continuation and length among contour segments, we propose a energy function to analysis the coarseness level of groups of contour segments. More specifically, our energy function is defined as:

$$E(G) = \frac{|h|}{L} = \frac{\sum h|\{curvratio(h) > t\}}{\int_G dx} \tag{4}$$

where $h$ are the high curvature change points among contour segments in a group of contour segments $G$, $|h|$ is the number of these points, and $L$ is the length of the group of contour segments. A threshold $t$ is used to determine how many groups are removed after the filtering. We simply used $max(E)/2$ for each image in this work.

# 3. EXPERIMENTS ON SKETCH BASED OBJECT RECOGNITION

To evaluate the quality of our automatically generated sketches, we design a sketch-based object recognition experiment.

**Training Set –** All 250 categories from the large dataset of human sketches [10] are utilized as the training set. There is a total of 20,000 sketches used for training, with 80 sketches in each category. The sketches in the dataset are all free-hand drawn sketches collected from large amount of humans. One can argue that sketches in this dataset capture how human draw sketches in general due to the size of the dataset and the large number of object classes.

**Testing Set –** Six categories are selected from Caltech 256 [22] which also feature in the human sketch dataset. There are 80 images in each category, giving a total of 480 sketches generated from the 480 natural images automatically. These sketches are used as the testing queries in our sketch recognition experiments.

**Object recognition model –** To represent the sketches, a bag-of-features representation is adopted. Following [10], we first

**Fig. 4**. Sketch examples. From Left to Right: original image, Canny, Pb, primal sketch, our sketch, human free hand drawn sketch.

extract local features, then construct a visual vocabulary using k means clustering. This vocabulary is used for quantizing features of the new sketches and SVM is used for classification. Specifically, for each category, all the sketches in this category are used as the positive examples and the others categories' sketches as the negative examples. Then we can learn a binary SVM classifier with RBF kernel to make the decision on whether a new sketch belongs to this category.

**Competitors –** We compare the recognition performance of our sketches (**Our**), against sketches produced by canny edge detector (**Canny**), Pb boundary detector (**Pb**) [6], Primal Sketch (**PS**)[12]. Among these three alternatives, Canny is a baseline representing how sketch can be generated by simple edge detection. The Pb contour detector [6] represents the current state-of-the-arts in object contour detection. Evaluating this method shows that how well object contours can be used to approximate object sketches. The PS method [12] is the only existing method on automated sketching using a single natural image.

**Results –** The qualitative results are shown in Fig. 4 and the quantitative results for rank-1 and rank-10 classification experiments are shown in Table 1 and Table 2, respectively. The tables show that all techniques give quite low classification rate in the rank-1 experiment, yet both Pb and our method come on top. This unsatisfactory performance is partially because of a 250-class classifier is used (classification by chance gives 0.4%). But it also shows that (1) the automatically generated sketches still exhibit a high degree of noise that confuse the classifiers and (2) the human perception of an object category has huge variations across people and differs significantly from the extracted object contours of natural images. However, the rank-10 classification results start to show the differences between the different compared approaches. In particular, it shows clearly that our sketching method outperforms the alternatives. As expected, the result by Canny is very poor – without any filtering the detected edges are too noisy and contain too much unnecessary details to be useful. Compared to the two other alternatives, the averaged recognition rate over the six categories using our methods is 22.92% higher than that state-of-the-art Primal Sketch (PS) and Pb. The improvement is particularly notable for challenging categories such as dog, elephant and beer-mug which have greater intra-class variations than the other three categories (e.g. there are far greater number of different types of dogs than ipod). For example, on beer-mug, we achieved 42.5% classification rate, compared to 15% using PS. Similarly, a three-fold increase in classification accuracy was obtained on the dog class. On elephant, the increase becomes 10-fold. Although not designed for sketch-based retrieval applications, the contour detection method Pb yielded more competitive results. Nevertheless Table 2 shows that its performance is consistently inferior to ours except on Car-tire where slightly better result was obtained. We can see from Fig. 4 that in general sketches generated using our methods keep a similar level of details as those from human. In contrast, both Pb and PS retain too much details. However, our approach can not guarantee closed contour sketches as human draws. An additional refine processing, e.g. the method proposed in [23], might be necessary in future work.

| | **Canny** | **Pb** | **PS** | **Our** |
|---|---|---|---|---|
| Airplane | 1.25 | **7.50** | 1.25 | 3.75 |
| Car-tire | 1.25 | **7.50** | 2.50 | 5.00 |
| Elephant | 0 | 1.25 | 0 | **3.75** |
| Ipod | **10.00** | 8.75 | 2.50 | 8.75 |
| Beer-mug | 2.50 | 7.50 | 2.50 | **11.25** |
| Dog | 0 | 0 | **1.25** | 0 |
| Average | 2.50 | **5.42** | 1.67 | **5.42** |

**Table 1**. Rank 1 classification rate (%)

| | **Canny** | **Pb** | **PS** | **Our** |
|---|---|---|---|---|
| Airplane | 10.00 | 26.25 | 15.00 | **28.75** |
| Car-tire | 15.00 | **23.75** | 8.75 | 22.50 |
| Elephant | 0 | 2.50 | 1.25 | **12.50** |
| Ipod | 13.75 | 15.00 | 15.00 | **20.00** |
| Beer-mug | 15.00 | 35.00 | 15.00 | **42.50** |
| Dog | 0 | 3.75 | 3.75 | **11.25** |
| Average | 8.96 | 17.71 | 9.79 | **22.92** |

**Table 2**. Rank 10 classification rate (%)

## 4. CONCLUSION

In this paper, we proposed a novel approach for automatic sketch generation from a single natural image. We casted sketch extraction into a perceptual contour grouping and filtering problem, and by exploiting simple perceptual grouping principles, we are able to develop an effective automated sketching algorithm to simulate how human draw object categories. A sketch-based object recognition experiment confirms the usefulness of our automatic drawn sketches for sketch retrieval tasks.

## 5. REFERENCES


[1] D. Marr and E. Hildreth, "Theory of Edge Detection," *Proceedings of the Royal Society of London. Series B, Biological Sciences*, vol. 207, no. 1167, pp. 187–217, 1980.

[2] Mathias Eitz, Kristian Hildebrand, Tamy Boubekeur, and Marc Alexa, "Sketch-based image retrieval: Benchmark and bag-of-features descriptors," *IEEE transactions on visualization and computer graphics*, 2011.

[3] Mathias Eitz, Ronald Richter, Kristian Hildebrand, and Marc Alexa, "Sketch-based shape retrieval," *ACM Transactions on Graphics, Proceedings SIGGRAPH (to appear)*, vol. 31, no. 4, 2012.

[4] Zhenbang Sun, Changhu Wang, Liqing Zhang, and Lei Zhang, "Free hand-drawn sketch segmentation," in *Proceedings of the 12th European conference on Computer Vision - Volume Part I*, Berlin, Heidelberg, 2012, ECCV'12, pp. 626–639, Springer-Verlag.

[5] Andrea Vedaldi and Andrew Zisserman, "Self-similar sketch.," in *ECCV (2)*, Andrew W. Fitzgibbon, Svetlana Lazebnik, Pietro Perona, Yoichi Sato, and Cordelia Schmid, Eds. 2012, vol. 7573 of *Lecture Notes in Computer Science*, pp. 87–100, Springer.

[6] Pablo Arbelaez, Michael Maire, Charless Fowlkes, and Jitendra Malik, "Contour detection and hierarchical image segmentation.," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 5, pp. 898–916, 2011.

[7] Qihui Zhu, Gang Song, and Jianbo Shi, "Untangling cycles for contour grouping," in *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*, oct. 2007, pp. 1 –8.

[8] V. Ferrari, L. Fevrier, F. Jurie, and C. Schmid, "Groups of adjacent contour segments for object detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 1, pp. 36–51, 2008.

[9] R. Kennedy, J. Gallier, and Jianbo Shi, "Contour cut: Identifying salient contours in images by solving a hermitian eigenvalue problem," in *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, june 2011, pp. 2065 –2072.

[10] Mathias Eitz, James Hays, and Marc Alexa, "How do humans sketch objects?," *ACM Transactions on Graphics (Proceedings SIGGRAPH)*, vol. 31, no. 4, pp. 44:1–44:10, 2012.

[11] Smit Marvaniya, Sreyasee Bhattacharjee, Venkatesh Manickavasagam, and Anurag Mittal, "Drawing an automatic sketch of deformable objects using only a few images," in *ECCV Workshops (1)*, 2012, pp. 63–72.

[12] Cheng en Guo, Song Chun Zhu, and Ying Nian Wu, "Primal sketch: Integrating structure and texture," *Computer Vision and Image Understanding*, vol. 106, no. 1, pp. 5–19, 2007.

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

[14] Jianbo Shi and J. Malik, "Normalized cuts and image segmentation," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 22, no. 8, pp. 888 –905, aug 2000.

[15] Yi-Zhe Song, Bai Xiao, Peter M. Hall, and Liang Wang, "In search of perceptually salient groupings.," *TIP*, 2011.

[16] A. Delong, A. Osokin, H.N. Isack, and Y. Boykov, "Fast approximate energy minimization with label costs," in *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, june 2010, pp. 2173 –2180.

[17] Y. Boykov and V. Kolmogorov, "An experimental comparison of min-cut/max- flow algorithms for energy minimization in vision," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 26, no. 9, pp. 1124 –1137, sept. 2004.

[18] Y. Boykov, O. Veksler, and R. Zabih, "Fast approximate energy minimization via graph cuts," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 23, no. 11, pp. 1222 –1239, nov 2001.

[19] D. R. Martin, C. C. Fowlkes, and J. Malik, "Learning to detect natural image boundaries using local brightness, color, and texture cues," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 26, no. 5, pp. 530–549, May 2004.

[20] Liangliang Nan, Andrei Sharf, Ke Xie, Tien-Tsin Wong, Oliver Deussen, Daniel Cohen-Or, and Baoquan Chen, "Conjoining gestalt rules for abstraction of architectural drawings.," *ACM Trans. Graph.*, vol. 30, no. 6, pp. 185, 2011.

[21] Song Wang, Toshiro Kubota, Jeffrey Mark Siskind, and Jun Wang, "Salient closed boundary extraction with ratio contour," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 4, pp. 546–561, 2005.

[22] G. Griffin, A. Holub, and P. Perona, "The Caltech-256," Tech. Rep., California Institute of Technology, 2007.

[23] Dian Gong and Grard G. Medioni, "Probabilistic tensor voting for robust perceptual grouping.," in *CVPR Workshops*. 2012, pp. 1–8, IEEE.