Neurocomputing 165 (2015) 338–349

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Im2Sketch: Sketch generation by unconflicted perceptual grouping

Yonggang Qi [a,*], Jun Guo [a], Yi-Zhe Song [b], Tao Xiang [b], Honggang Zhang [a], Zheng-Hua Tan [c]

[a] School of Information and Communication Engineering, BUPT, Beijing, China
[b] School of EECS, Queen Mary University of London, London, UK
[c] Department of Electronic Systems, Aalborg University, Aalborg, Denmark

ARTICLE INFO

Article history:
Received 20 October 2014
Received in revised form
26 January 2015
Accepted 9 March 2015
Communicated by Ran He
Available online 17 March 2015

Keywords:
Gestalt confliction
RankSVM
Perceptual grouping
Sketch generation
Sketch-based image retrieval

ABSTRACT

Effectively solving the problem of sketch generation, which aims to produce human-drawing-like sketches from real photographs, opens the door for many vision applications such as sketch-based image retrieval and non-photorealistic rendering. In this paper, we approach automatic sketch generation from a human visual perception perspective. Instead of gathering insights from photographs, for the first time, we extract information from a large pool of human sketches. In particular, we study how multiple Gestalt rules can be encapsulated into a unified perceptual grouping framework for sketch generation. We further show that by solving the problem of Gestalt confliction, i.e., encoding the relative importance of each rule, more similar to human-made sketches can be generated. For that, we release a manually labeled sketch dataset of 96 object categories and 7680 sketches. A novel evaluation framework is proposed to quantify human likeness of machine-generated sketches by examining how well they can be classified using models trained from human data. Finally, we demonstrate the superiority of our sketches under the practical application of sketch-based image retrieval.

1. Introduction

There exists plenty of prior work on sketch in computer vision, from sketch recognition [1,2] to sketch-based image retrieval (SBIR) [3,4]. Recently, sketch research has gained much momentum due to the proliferation of touch sensitive devices. Nonetheless, how to automatically produce sketches using machines as humans do is still an open problem [1,5,6]. Solving this problem importantly opens the door for many vision applications, especially for SBIR since better sketch conversion essentially closes the domain gap between query sketches and gallery photographs.

This paper sets out to tackle the problem of sketch generation via learning from established theories found in human visual cognition studies. Human visual system is very powerful so that we can easily find sense from chaos. In Neuroscience, one of the most critical problems is understanding how human brains perceive visual objects. In particular, perceptual grouping, a concept introduced by the Gestalt school of psychologists [7], advocates that human perceives certain elements of the visual world as going together more strongly than the others. Max Wertheimer, a pioneer in the Gestalt school, pointed out the significance of perceptual grouping and further listed a series of rules, such as proximity, similarity and continuation [8]. His work has consequently triggered plenty of research specifically aimed at understanding human visual systems [9,10].

We treat sketch generation as a perceptual grouping and filtering task. Essentially, the underlying hypothesis is that *perceptual grouping is able to find sense from chaos therefore leaving only signals corresponding to sketches of human resemblance*. More specifically, our choice of utilizing an elementary grouping process for sketch generation is justified as follows: (i) sketches exhibit a lack of visual cues (black and white drawing versus textured image regions), making conventional vision algorithms sub-optimal, (ii) sketch despite abstract (e.g., a stickman human figure) is very iconic – it is often structural variations of strokes that capture object appearance, last but not least, (iii) sketches are the simplest form of depictions of human visual impressions that can be rendered by hand, therefore act as ideal basis for applying/testing theories found in human visual cognition [3].

Traditionally, applying perceptual grouping in vision applications incurs two critical design considerations: (i) how to combine multiple Gestalt rules into a single globally optimized framework, (ii) how to encode the relative importance of each rule. For the former problem, although unary Gestalt principle has been proven to be useful for contour grouping when used alone [11–13], very few work [14] attempt to investigate how they can be exploited jointly in a single framework. The latter problem, often referred as Gestalt confliction, remains unaddressed to date [7,15]. Despite

* Corresponding author.
E-mail address: qiyg@bupt.edu.cn (Y. Qi).

http://dx.doi.org/10.1016/j.neucom.2015.03.023
0925-2312/

being the subject of investigation in the fields of psychology, little is known about how Gestalt confliction work in human vision systems [16], thus shedding little light on how to design a computer vision system.

In this paper, we first propose a unified grouping framework that is able to work with multiple Gestalt principles simultaneously. We then show how Gestalt confliction can be accounted for by learning from a dataset of pre-segmented human sketches. In particular, a multi-label graph-cuts [17–19] perceptual grouping framework is developed to group stroke segments while utilizing the learned importance of different Gestalt principles. It follows that, upon generating sketch from photograph, the same learned perceptual grouping framework can be used to form groups of image boundary segments, which are further filtered to produce human-drawing-like sketches. More specifically, a learning to rank strategy based on RankSVM [20] is proposed to learn the relative importance between two Gestalt principles, namely proximity and continuity. We learn from a subset of a large scale human-drawn sketch dataset [1], where each sketch is pre-segmented into semantic groups. The entire process of the proposed sketch generation framework is shown in Fig. 1.

To evaluate the quality of automatically generated sketches, we present a novel approach that recognizes them by sketch classifiers trained from human data. Prior works [5,21] evaluate sketching performance by comparing computer generated sketches with tracings produced by humans. This evaluation strategy importantly does not account for likeness to human-made sketches because (i) sketches are abstract depictions that are fundamentally different from tracing of image boundaries, (ii) humans often sketch without reference to real photographs of objects, (iii) sketches exhibit much more intra-class variability, due to different levels of drawing skills and individual visual impressions. By measuring how well a human sketch classifier [1] recognizes machine-generated sketches, we essentially examine how closely they resemble human-made ones. Our results show that the sketches generated using our method outperform a number of state-of-the-arts alternatives.

To further demonstrate the quality of our sketches, we demonstrate its effectiveness for SBIR. Experimental results confirm a positive performance boost on the largest SBIR dataset to date [4], when compared with state-of-the-art alternatives. It importantly shows that our sketch generation algorithm is able to bridge the domain gap between sketches and natural images. As can be seen from Fig. 2, the proposed sketch converter yields cleaner sketch that matches better to the query when encoded using common descriptor such as Histogram of Oriented Gradients (HOG).

The contributions of this paper can be summarized as follows:

(1) We apply perceptual grouping as means for automated sketch generation and propose a learning to rank strategy to learn the relative importance among two Gestalt principles.

(2) A novel evaluation strategy is devised to quantitatively evaluate human likeness of sketches.

(3) We demonstrate the effectiveness of sketch generation in SBIR and show a performance boost when compared with state-of-the-art alternatives.

(4) A new dataset containing 96 object categories and 7680 sketches is released, where each sketch is segmented into semantic parts by human.

2. Related work

2.1. Perceptual grouping

Perceptual grouping is one particular kind of organization phenomenon. Historically, grouping is stated as the fact that observers perceive some elements of visual field as “going together” more strongly than others [7]. Wertheimer first laid out the problem of perceptual grouping [8] by asking what stimulus factors influence the perceived grouping of discrete elements. Several work have been triggered since to investigate the problem of perceptual grouping. To date, many Gestalt principles, such as proximity, continuity, symmetry, parallelism and closure [7], have been discovered by researchers, and plenty of computer vision applications [22–27] rely on these principles to work. However, the problem of Gestalt confliction, that is how multiple Gestalt principles work collectively, remains relatively unaddressed to date. Very few work attempts to investigate this problem explicitly, however, early evidence [28] suggests that an effective solution to the problem will likely boost grouping performance. In fact, recently the problem of Gestalt confliction is tackled by Nan et al. [29] to simplify architectural drawings with conjoining Gestalt rules. Although achieving good performance, the method cannot be directly applied to natural images as it relies on clear geometric properties that are hard to extract from images. This paper aims to formulate a general framework to learn the relative importance of different Gestalt principles explicitly, thus develop an algorithm to congregate them for sketch generation on natural images.

2.2. Sketch generation

Sketch generation aims to convert images into human-drawing-like sketches. Application such as SBIR benefits from better sketches since it is easier to match two entities from the same domain other than across domains. Early work on automatic sketching takes a contour detection and object segmentation approach [21,30,31], which aims to produce curves that perfectly depict an image, or constitutes global object profiles. Abelaez et al. [30] proposed a general framework to transform the output of any contour detector into a hierarchical region tree with

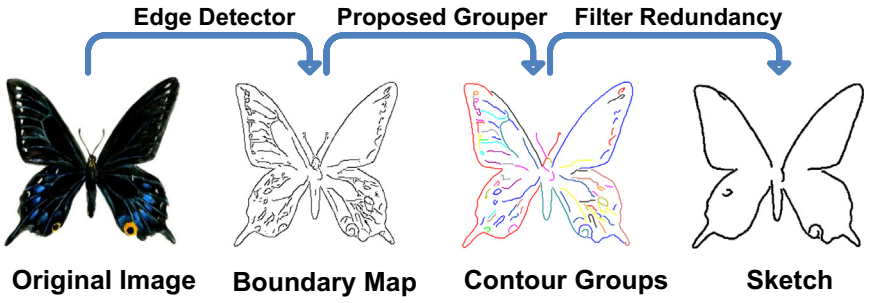


Fig. 1. Overview of sketch generation. Given a real image of ‘butterfly’, we first extract the corresponding boundary map, then the proposed perceptual grouping framework is employed to produce contour groups, followed by a filtering procedure to generate a ‘butterfly’ sketch consequently.

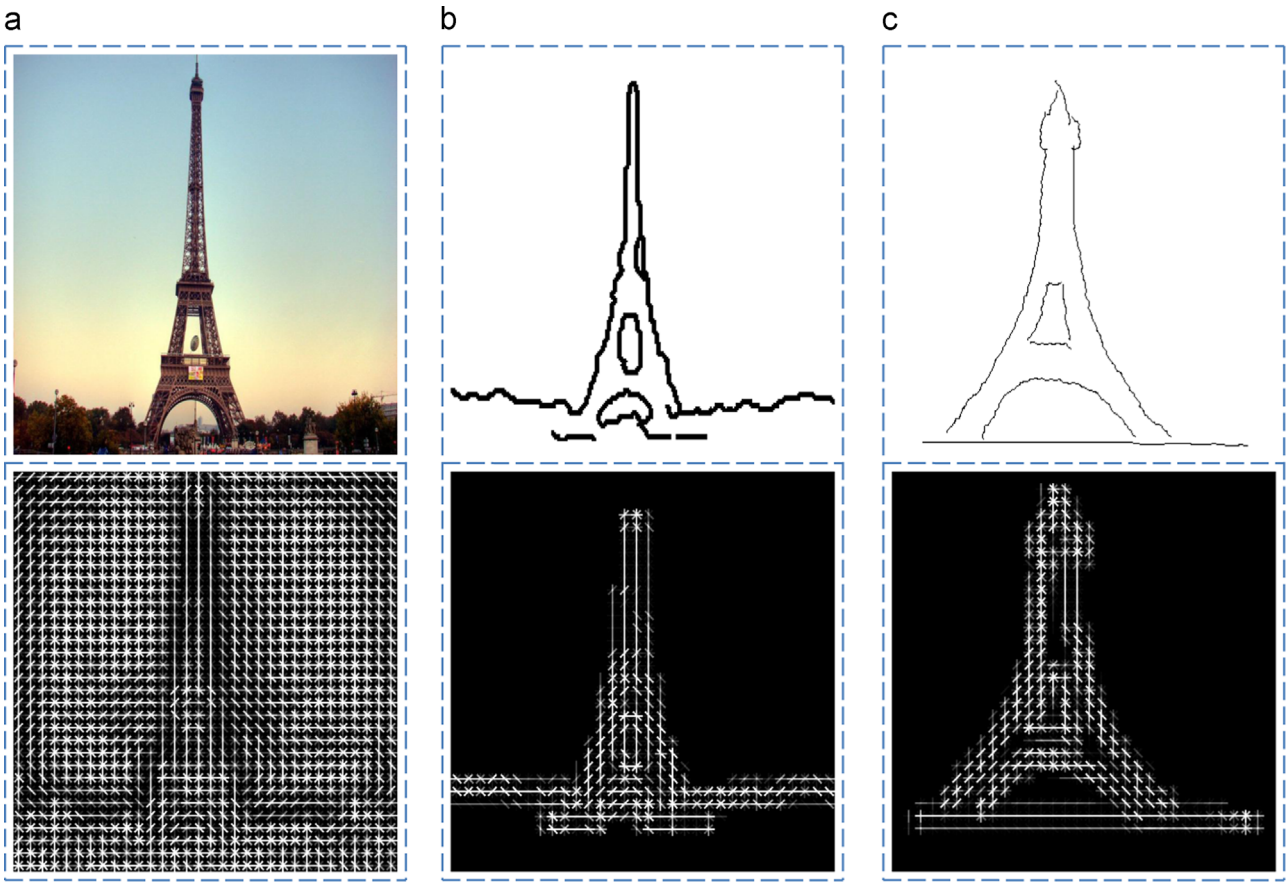


Fig. 2. Left to right: visualizations of original image, machine generated sketch, human sketch and their HOG feature descriptors. It shows clearly that the proposed sketch converter yields cleaner sketch that matches better to the human sketch query when encoded using common descriptor such as Histogram of Oriented Gradients (HOG). (a) Real image, (b) generated sketch, and (c) human sketch.

the intention to generate object contours that are most similar to human object segmentation. Zhu et al. [21] exploited the inherent topological 1D structure of salient contours. Grouping is performed by eigen-decomposition of contour grouping graphs. However, object contours or profiles are fundamentally different from human sketches. In particular, compared with object contours, human sketches exhibit higher variances in terms of style, viewpoint and abstraction level. Importantly, an object contour is often only a subset of human sketch which typically includes additional details inside the contour.

Recently, Marvaniya et al. [5] 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 work only requires a single image to work, making it more generally applicable. The work by Guo et al. [6] 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 and Markov Random Field, to encode geometric structures and stochastic textures, respectively. Our study shows that by exploiting perceptual grouping principles, a much simpler method can be developed which is able to generate better sketches.

2.3. Sketch-based image retrieval

Query by Visual Example (QVE) has attracted much attentions in recent years with SBIR being one of the main driving forces. There are three main reasons behind this: (i) a sketch speaks a "hundred" of words, which makes it a more efficient and precise query (e.g., shape, pose, style of a handbag) than a text-based image retrieval system [4,35], (ii) words are not always the most convenient way to describe the exact object people want to search, especially when it comes to fine-grained object details, (iii) the exploding availability of touch-screen devices that is fast changing the way how people input search query.

Most prior work on SBIR [33,34,3] primarily operate as follows: first edges are extracted to approximate sketch, then local features (e.g., HOG) are extracted on the resulting edge maps, finally features of a query sketch and the approximated sketches from natural image are matched using KNN. However, very few work [32] specifically study the role of sketch generation to bridge the semantic gap. In this paper, we demonstrate that our perceptual grouping based method for sketch generation is more suitable for SBIR compared to the traditional edge descriptors.

3. A multi-label graph-cuts model for grouping

In this section, a multi-label graph-cuts [18] based model for perceptual grouping is described which conjoins different Gestalt principles. In particular, two Gestalt principles, continuity and proximity, are utilized, however the framework can be extended to work with more principles. Specifically, we formulate the problem of grouping as a min-cut/max-flow optimization problem. We later show in Section 4 how Gestalt confliction can be learned and easily embedded into this framework to deal with Gestalt conflict-tion by re-weighting datacost according to the type of Gestalt principles.

3.1. Potential groupings

To solve this grouping problem, we need to specify the relationship between primitives and the possible labels, thus to compute the data cost item required in the subsequent multi-label graph-cuts framework. The possible labels are obtained by discovering potential groupings, which are sets of primitives, indicating the possibility to assign a label to one primitive.

In our case, potential groupings are defined by two Gestalt principles: continuity and proximity. Given a set of primitives, Q, which consists of n primitives in a sketch, each primitive $q_i \in Q$ will in turn serve as one where all other primitives $q_j \in Q$ are compared against with. More specifically, each set of potential groupings is defined as follows:

Continuity Gestalt is defined by detecting groups which would form a continuous curve:

$$L_i^{con} = \bigcup \{q_i, q_j\} \,|\, \{R_s\{q_i, q_j\} > t_s\} \tag{1}$$

where $R_s\{q_i, q_j\}$ indicates the slope trend difference between q_i and q_j. Hence the set of continuity groups is the union over all meta-continuity Gestalt, which can be denoted as:

$$L^{con} = \bigcup_{\forall i} L_i^{con} \tag{2}$$

Proximity Gestalt is defined by detecting groups where primitives are close enough to each other:

$$L_i^{pro} = \bigcup \{q_i, q_j\} \,|\, \{R_p\{q_i, q_j\} > t_p\} \tag{3}$$

where $R_p\{q_i, q_j\}$ indicates the spatial distance between q_i and q_j. Similarly, the set of proximity groups is defined as

$$L^{pro} = \bigcup_{\forall i} L_i^{pro} \tag{4}$$

where t_s and t_p in the above equations are fixed thresholds for determining whether a pair of primitives should be grouped into a potential group or not when applying one of the two Gestalt principles as grouping criterion in turn. Given the potential groupings, it follows that the data cost item can be naturally obtained, as detailed in the next section.

3.2. Multi-label graph-cuts model

The problem of grouping is formulated as a min-cut/max-flow optimization problem, where the overall energy function is defined as

$$E(L) = \sum_{q_i \in Q} D(q_i, L) + \sum_{\{q_i, q_j\} \in N} V_{\{q_i, q_j\}} + \sum_{l \in L} F_l \tag{5}$$

where N is the set of pairs of neighboring elements in Q, and $L = \{L^{con}, L^{pro}\}$ as defined in Section 3.1. D is the data cost energy, V is the smoothness cost energy, and F is the label cost energy. Detailed definitions for each of these three terms are given as follows:

Data cost represents the fitness between primitive q_i and the possible assigned label L_i. The higher the fitness, the lower the cost or penalty. More specifically, continuity and proximity data costs are defined as follows:

Continuity data cost measured between q_i and L_i^{con} is defined according to the average continuity over all the pairs of primitives:

$$D(q_i, L_i^{con}) = 1 - \frac{1}{|L_i^{con}|} \sum_{q_j \in L_i^{con}} R_s\left\{q_i, q_j\right\} \tag{6}$$

where L_i^{con} $(i = 1, 2, \ldots, n)$ is a label, which represents potential groupings found by continuity principle, $|L_i^{con}|$ is the number of primitives in the potential group L_i^{con}. We use one minus the average continuity value for that, the more the q_i fits to continuity, the lower the penalty. The same reason applies to the definition of proximity data cost.

Proximity data cost defined between q_i and L_i^{pro} based on the average proximity over all pairs of primitives:

$$D(q_i, L_i^{pro}) = 1 - \frac{1}{|L_i^{pro}|} \sum_{q_j \in L_i^{pro}} R_p\left\{q_i, q_j\right\} \tag{7}$$

where L_i^{pro} $(i = 1, 2, \ldots, n)$ is a label, representing potential grouping found by proximity Gestalt principle. $|L_i^{pro}|$ is the number of pairs of primitives in this proximity induced grouping.

Smoothness cost defines 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_i and q_j, the smoothness energy is defined by the inverse Euclidean Hausdorff-distance between them, which is the same as the one used in [29]

$$V_{\{q_i, q_j\}} = d(q_i, q_j)^{-1} \tag{8}$$

Label cost penalizes overly complex models and encouraging the explanation of the input sketch with fewest and cheapest labels. We define F_l as a non-negative label cost, which measures the Gestalt affinity for each specific Gestalt principle, for label l. More specifically, for continuity Gestalt, label cost is measured by the average continuity between every pair of primitives, i.e.,

$$F_l = \sum_{q_i, q_j \in l} R_s\{q_i, q_j\}, \quad l \in L^{con} \tag{9}$$

for proximity Gestalt, label cost is measured by the average distance between every pair of primitives, i.e.,

$$F_l = \sum_{q_i, q_j \in l} R_p\{q_i, q_j\}, \quad l \in L^{pro} \tag{10}$$

Upon solving the optimization problem defined in Eq. (5), each primitive will be assigned with an optimal group label.

4. Learning Gestalt confliction

Given the grouping framework detailed above, in this section, we aim to introduce a human-annotated sketch dataset containing 7680 sketches in 96 categories, and show how these annotations can be utilized to learn Gestalt confliction using a learning to rank strategy. We further demonstrate how the learned confliction information can be embedded into the general grouping framework laid out in the previous section.

4.1. Dataset

We randomly select 96 object categories (80 sketches per category, 7680 sketches in total) from a large scale human drawn sketch dataset [1].[1] 15 participants (7 male, 8 female) are then asked to manually label strokes in each sketch into groups of semantic parts. Essentially, instead of segmenting images into semantic regions as performed in common segmentation datasets [30], we produce a sketch segmentation dataset where each semantic part is assigned a unique label. A subset of annotated sketches are illustrated in Fig. 3, where semantic parts are color-coded.

As previously mentioned, having such a dataset is critical to our ultimate aim of learning from human visual perception. Conceptually, sketches are employed as object depictions in the human brain. Recent Neuroscience work [3] also indicates that simple,

[1] We were not able to work on the full dataset in [1] due to the cost sensitive nature of the labeling task involved.

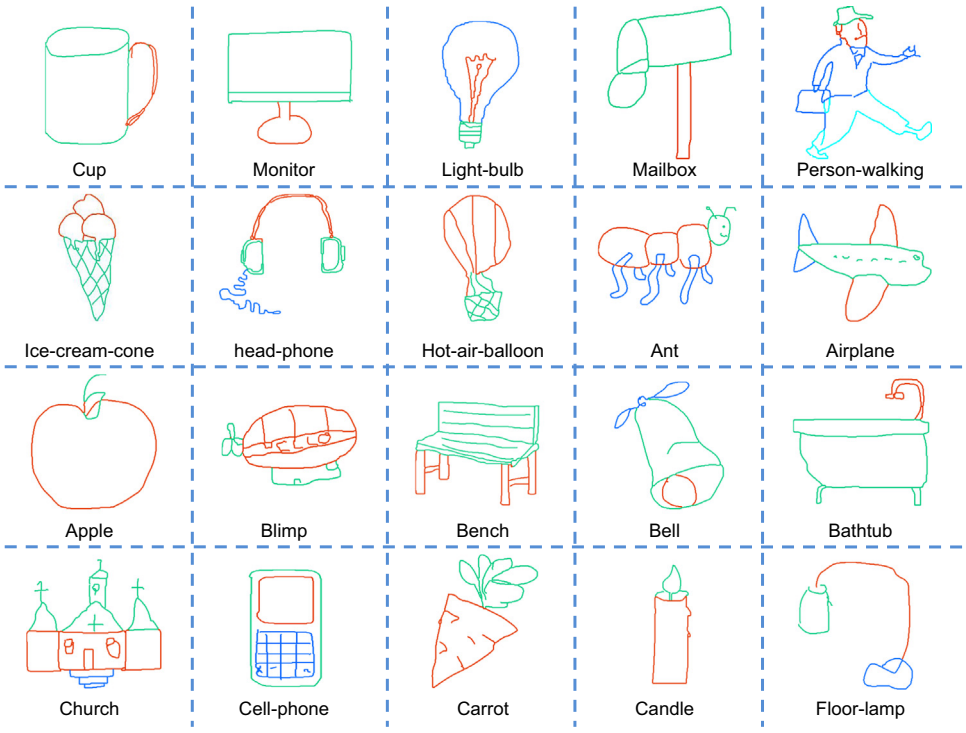


Fig. 3. Example sketches from the human-labeled sketch dataset. All the strokes in each sketch are manually labeled into groups of semantic parts. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)

abstracted sketches activate our brain in similar ways to real stimuli (e.g., natural images). In the next section, we present how Gestalt confliction can be learned from this dataset.

4.2. Learning to rank for Gestalt untangling

We cast the problem of learning the importance of different Gestalt principles into a RankSVM [36] model. RankSVM is widely used in Computer Vision problems such as person re-identification and gait recognition [37]. In our case, every stroke is treated as a query to retrieve all other strokes in the sketch. Consequently, we can rank them according to whether the query and every other primitive belong to the same group – rank 1 is assigned if so, rank 2 otherwise. Essentially, the ranking model is to learn a weighted distance using the similarity measured by the two Gestalt principles (i.e., continuity and proximity), so that this ranking order is maintained as much as possible across all training sketches. In this paper, we use the Matlab implementation of primal RankSVM supplied by Chapelle [38]. More specifically, our training set is composed of in the following:

Set of primitives denoted as $Q=\{q_1,q_2,\ldots,q_{|Q|}\}$, where $|Q|$ is the number of primitives in Q. $|Q|$ varies according to the complexity of the sketch, i.e., more primitives in complex sketches.

Pair of primitives written as (q_i,q_j) each represented as a 2-dimensional feature vector $\mathbf{x}(q_i,q_j)$, which indicates the difference between the pair of primitives. In particular, each dimension in $\mathbf{x}(q_i,q_j)$ corresponds to a Gestalt principle, i.e., the first dimension of vector $\mathbf{x}(q_i,q_j)$ corresponds to continuity and the second one corresponds to proximity. $\mathbf{x}(q_i,q_j)$ is defined as

$$\mathbf{x}(q_i,q_j)=\begin{bmatrix} x(q_i,q_j)_{con} \\ x(q_i,q_j)_{pro} \end{bmatrix}$$

where $x(q_i,q_j)_{con}=R_s\{q_i,q_j\}$, $x(q_i,q_j)_{pro}=R_p\{q_i,q_j\}$, R_s and R_p are slope trend and geometry distance between the two primitives, respectively. Therefore, $x(q_i,q_j)_{con}$ and $x(q_i,q_j)_{pro}$ measure the continuity and proximity between q_i and q_j, respectively.

Pair relationship: Every single primitive q_i is labeled by a relevance indicator $y(q_i,q_j)$ which represents its relationship to another primitive q_j. In our case, we define y with a value 1 when a primitive q_i is grouped together with another primitive q_j, and -1 otherwise. Thus, for primitive q_i, all the other primitives are divided into two sets according to its relevance indicator with q_i:

$$Q(q_i)^+=\{q_1^+,q_2^+,\ldots,q_{|Q(q_i)^+|}\}$$

where $y(q_i,q_j^+)=1$ for all $q_j^+\in Q(q_i)^+$ $(j=1,2,\ldots,|Q(q_i)^+|)$, and $|Q(q_i)^+|$ represents the number of elements relevant to q_i in the set, similarly,

$$Q(q_i)^-=\{q_1^-,q_2^-,\ldots,q_{|Q(q_i)^-|}\}$$

where $y(q_i,q_j^-)=-1$ for all $q_j^-\in Q(q_i)^-$ $(j=1,2,\ldots,|Q(q_i)^-|)$, and $|Q(q_i)^-|$ represents the number of elements irrelevant to q_i in the set.

Following the above, positive pairs $\hat{Q}^+=(q_i,q_j^+)$ and relative negative pairs $\hat{Q}^-=(q_i,q_j^-)$ are formed. Accordingly, preference pairs $P=(\hat{Q}^+,\hat{Q}^-)$ are produced. With the constraints P, we can learn the ranking function, $f(q_i,q_j)=\omega^T\mathbf{x}(q_i,q_j)$, where ω refers to a 2-dimensional weight vector indicating the significance of each Gestalt principle.

Specifically, we obtain ω in the learning function by solving the following optimization problem:

$$\omega=\underset{\omega}{\operatorname{argmin}}\frac{1}{2}||\omega||^2+C\sum_{k=1}^{|P|}l(\omega^T(\hat{Q}^+-\hat{Q}^-)) \tag{11}$$

where k is the index of the preference pairs, $|P|$ is the total number of preference pairs used for training, C is a positive importance weight on the ranking performance and is automatically selected by cross validation on the training set. l is the hinge loss function.

Table 1
Inverse importance (α) of Gestalt rules learned by RankSVM. Proximity with the smallest weight is the dominant principle when conflicts with continuity.

Gestalt	**Continuity**	**Proximity**
α	0.8525	0.1475

Although the efficient primal RankSVM algorithm [38] is adopted, the amount of training data is still too big to be computationally tractable. To handle this problem, we subsample from each training data. That is, in each sketch image, we randomly choose two primitives in every group, and utilize all selected primitives to form positive pairs and relative negative pairs. Finally, approximately 7 million preference pairs are formed for learning the RankSVM model. On a standard Linux server with CPU@2.53 GHz 12G RAM, the entire training procedure costs approx. 5 h.

We finish by converting ω into α by normalizing each dimension of ω to $[0,1]$. It follows that α has two values each corresponds to one Gestalt principle. α is essentially the inverse importance of each Gestalt principle. We use α instead of ω to facilitate later analysis and most importantly enable easy embedding into the grouping framework, as detailed in the next section.

Table 1 shows the learned α value for the two Gestalt principles studied. Note that smaller value corresponds to higher significance. This table shows clearly that proximity is much more important than continuity. Interestingly, this finding is in tune with results from psychology studies [28], which suggests that humans also rely more on proximity.

4.3. Unconflicted grouping

Taking into consideration of the learned Gestalt confliction, the multi-label graph-cuts model (Section 3) can be further improved. More specifically, Eq. (5) is updated as follows:

$$E(L)=\sum_{q_i\in Q}\alpha D(q_i,L)+\sum_{\{q_i,q_j\}\in N}V_{\{q_i,q_j\}}+\sum_{l\in L}F_l \tag{12}$$

where the only difference between Eq. (5) and Eq. (12) is that data cost D is re-weighted according to the pre-learned relative importance of continuity and proximity. Essentially, Gestalt confliction is taken into account for better grouping – the more important one the Gestalt principle is, the less it contributes to the overall energy E (the goal is to minimize E). In our case, according to the learned inverse importance (α) shown in Table 1, proximity plays a dominant role when it conflicts with continuity. We will demonstrate the effectiveness of the proposed approach in Section 6.

5. Sketch generation

In this section, we introduce how sketches can be generated from real images using the proposed perceptual grouping framework. There are primarily three stages (Fig. 1) for automatic sketch generation: (i) extracting boundary map to produce curve segments as grouping primitives, (ii) grouping boundary by the proposed grouper while accounting for Gestalt confliction, (iii) filtering away redundancy by coarseness analysis of boundary groups. Details of each stage are presented as follows.

5.1. Extracting boundary map

Given a real image I, we first perform contour extraction using a state-of-the-art edge detection algorithm [30] to obtain a boundary map B from I. Afterwards, the boundary map B is further transformed to several curve segments $Q=\{q_1,q_2,\ldots,q_n\}$ using a method derived from psychological studies how humans perform the same task [7,15].

5.2. Unconflicted grouping of boundary

To filter away redundancy, we perform the proposed unconflicted perceptual grouping framework on curve segments $Q=\{q_1,q_2,\ldots,q_n\}$, which aims to group the salient curve segments together, and thus separate them from noise. In particular, by embedding the learned Gestalt confliction information into the multi-label graph-cuts model, a better grouping result can be obtained by considering two Gestalt principles simultaneously, i.e., continuity and proximity, to facilitate the following filtering process. By minimizing the objective function defined in Eq. (12), the optimal solution L produces a set of curve segments groups $G=\{G_1,G_2,\ldots,G_m\}$. Based on the result of boundary grouping, only groups of salient boundaries are maintained according to a consequent coarseness analysis procedure.

5.3. Sketching by group-based filtering

Given a set of groups G after boundary grouping, our goal is to filter away redundancy to generate human-like sketches. Inspired by [39] which finds salient contours by ratio contour that measures gaps, continuation and length among contour segments, we propose a energy function to analyze the coarseness level of groups of curve segments. Therefore, only groups with low level coarseness are maintained as the generated sketch. More specifically, for a group of boundaries $G_i\in G$, the energy function is formulated as

$$E(G_i)=\frac{|h|}{S}=\frac{\sum h|\{\text{curvratio}(h)>t\}}{\int_{G_i}dx} \tag{13}$$

where h indicates the high curvature change points on the curve segments in group G_i, $|h|$ is the number of these points. S represents the total length of all curve segments in group G_i. A threshold is used to determine how many groups should be kept, and is automatically chosen using cross-validation.

5.4. Complexity analysis

To generate a sketch from a real image, the most time consuming process by far is boundary grouping. The worst case time complexity for grouping boundaries by our graph-cuts algorithm is $O(mn^2)$ where n is the number of nodes and m is the number of edges in the graph [17]. In practice, the average time spent for generating a sketch is approx. 11.25 s on a Linux server with CPU@2.53 GHz 12G RAM.

6. Experiments and analysis

We first conduct an experiment to evaluate the effectiveness of our proposed grouping framework, especially with and without the learned Gestalt confliction. Then we demonstrate how human-drawing likeness of machine-generated sketch can be measured using a novel sketch recognition experiments.

6.1. Unconflicted Gestalt grouping

Experimental Settings: A subset of human sketch dataset [1] is used, which includes 96 categories and 80 sketches in each category. Half for learning the importance of different Gestalt principles, and the other half for testing under the grouping

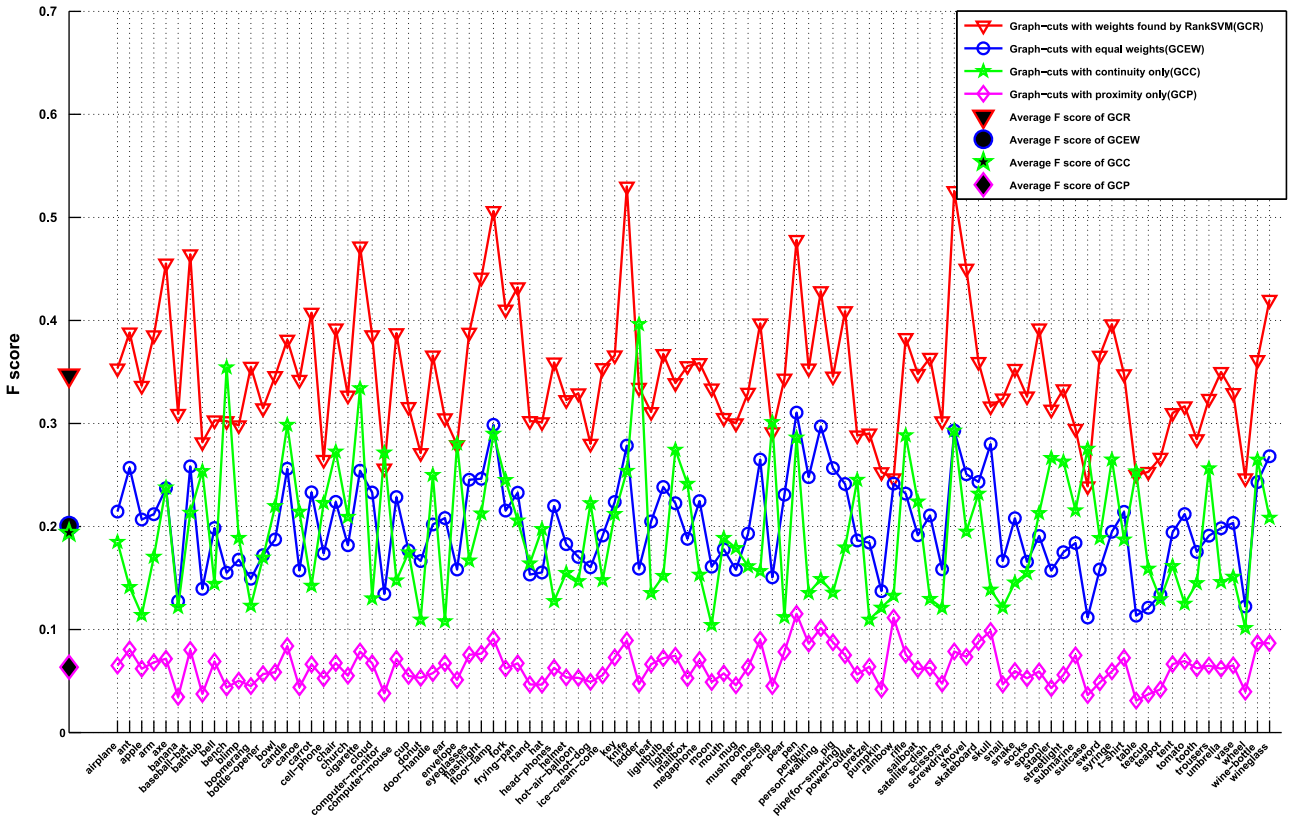


Fig. 4. Grouping performance comparison over 7680 sketches. The *x*-axis shows the 96 object categories, and the *y*-axis shows the average *F* score obtained by different methods on each category. It shows clearly that the proposed perceptual grouping method consistently outperforms the others.

evaluation. We run our experiment 10 times by randomly selecting the set of half sketch images each time, and α shown in Table 1 is averaged over all 10 trials.

Evaluation metric: *F*-measure is widely used for evaluation in the field of information retrieval, and it is adopted to evaluate the grouping performances in our problem. More specifically, for each test sketch with n curve segments, the grouping obtained using our graph-cuts algorithm is represented as an affinity matrix. Given the ground truth (human grouping result), another affinity matrix is constructed. Thus we compute the *F*-measure using these two matrices to evaluate how well the estimated grouping matches the ground truth grouping.

We compare the grouping result obtained using our algorithm with the learned importance weight (GCR) against those by using the same graph-cuts algorithm but either (1) use an equal weight to the two principles (GCEW), or (2) use one of the two principles alone, i.e., continuity (GCC) and proximity (GCP) serving as the only principle for grouping in turn. The contour grouping results on each of the 96 categories are reported in Fig. 4. It shows clearly that over all of the 96 categories, our algorithm with the learned weighting consistently outperforms the algorithm with equal weight assigned to the two Gestalt principles. On average, an increase of 14.57% in the *F*-measure score is obtained. This result demonstrates that the learned weighting not only supports the psychology study findings, but also has practical use in solving computer vision problems. Also, we can see that the performance of our algorithm is better than any Gestalt principle used alone. And even though naively combining the two Gestalt principles, the grouping result is still better than using continuity or proximity alone. Moreover, it is interesting to note that when a single Gestalt principle is used, continuity is the best option. Our learned weighting suggests that a larger weight should be given to proximity if the combination is to yield any improvement on the grouping performance.

6.2. Measuring human-drawing likeness of sketches

To evaluate the human likeness of our automatically generated sketches, we design a novel sketch-based object recognition experiment. The idea is that, human free-hand sketch trained classifiers should be able to recognize machine generated sketches, therefore quantifying human-likeness.

Training and testing Set: All 250 categories from the large dataset of human sketches [1] 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 collections of free-hand drawn sketches come from many participants. 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 categories. A subset category of Caltech 256 [40] is used for testing, which also features in the human sketch dataset [1]. Specifically, there are fifteen categories are randomly selected: 'airplane' (80 images), 'car-tire' (79 images), 'elephant' (73 images), 'ipod' (80 images), 'beer-mug' (73 images), 'dog' (79 images), 'backpack' (81 images), 't-shirt' (80 images), 'binoculars' (81 images), 'duck' (58 images), 'frying-pan' (63 images), 'baseball-bat' (48 images), 'butterfly' (56 images), 'mailbox' (34 images) and 'teapot' (80 images). Consequently, it gives a total of 1045 sketches generated from the 1045 corresponding natural images automatically. These sketches are used as the testing queries in our sketch recognition experiments.

Training sketch classifier: Following [1], we represent each human free-hand sketch using Bag-of-Words (BoW) coupled with Histogram of Oriented Gradients (HOG) features. Firstly, to achieve scale

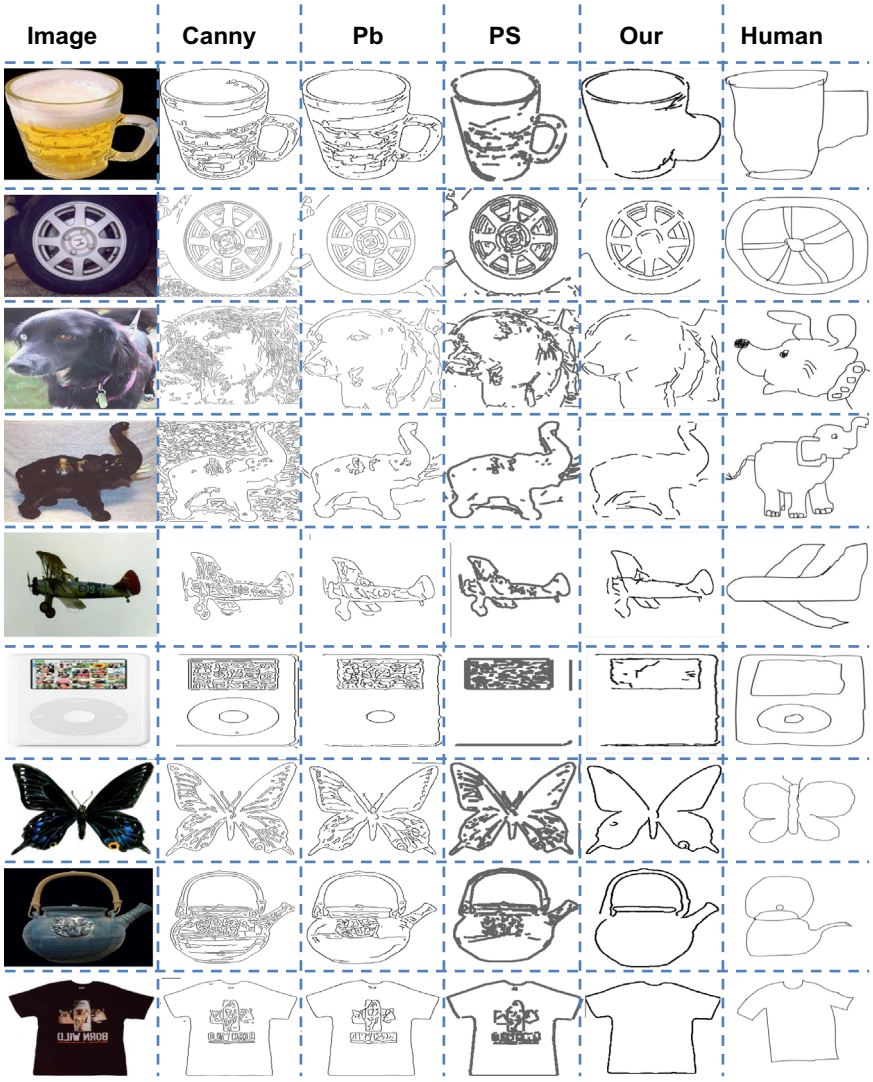


Fig. 5. Sketch examples. From left to right: original image, Canny, Pb, primal sketch, our sketch, and human free hand drawn sketch. We can observe that sketches generated using our methods keep a similar level of details as those from human.

invariant, sketches are all scaled into a 256×256 sketch image, then we extensively sample patches over it, i.e., 784 patches per sketch, to extract HOG features. After collecting a large number of sample patches, they are utilized for building vocabulary by k-means clustering ($k=500$). The vocabulary is then used for quantizing features of any new sketch. To learn a SVM classifier for each category, all the sketches in this category are used as positive examples and the other categories' sketches as negative ones. It follows that we can learn a binary SVM classifier with RBF kernel to make the decision on whether a new sketch belongs to this category.

Baselines: We compare the recognition performance of our sketches (*Our*), against sketches produced by canny edge detector (*Canny*), Pb boundary detector (*Pb*) [30], Primal Sketch (*PS*) [6]. Among these three alternatives, Canny is a baseline representing how sketch can be generated by simple edge detection. The Pb contour detector [38] 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. PS [6] is the only existing method for automated sketching using a single natural image.

A few qualitative results are illustrated in Fig. 5. As can be seen, our machine generated sketches consistently generate better sketches than alternatives and offer close resemblance to those produced by human. Table 2 summarizes quantitative results for rank-10 classification experiments. We can observe that our sketching method outperforms the alternatives. As expected, the

Table 2
Rank 10 classification rate (%). In general, sketch generated by our method have better chance to be recognized by classifiers over nearly all the categories.

Object	**Canny**	**Pb**	**PS**	**Ours**
Airplane	10.00	26.25	15.00	**26.25**
Car-tire	15.00	23.75	8.75	**29.25**
Elephant	0	2.50	1.25	**16.50**
Ipod	13.75	15.00	15.00	**15.00**
Beer-mug	15.00	**35.00**	15.00	30.25
Dog	0	3.75	3.75	**17.72**
Backpack	0	19.75	39.51	**43.21**
T-shirt	5.00	26.25	23.75	**28.75**
Binoculars	6.17	27.16	23.46	**39.51**
Duck	0	1.72	0	**5.17**
Frying-pan	0	28.57	39.68	**76.19**
Baseball-bat	2.08	18.75	16.67	**45.83**
Butterfly	0	16.07	10.71	**26.79**
Mailbox	2.94	5.88	11.76	**17.65**
Teapot	0	13.75	11.25	**23.75**
Average	4.66	17.61	15.71	**29.45**

result by Canny[2] is very poor – without any filtering the detected edges are too noisy and contain too much unnecessary details to be useful, which even lead to total failure (0% recognition rate) on the following categories: 'elephant', 'dog', 'backpack', 'duck', 'frying-pan', 'butterfly' and 'teapot'. Compared to the other two alternatives, the averaged recognition rate over the fifteen categories using our methods is 29.45%, which offers an approximately two-fold improvement over state-of-the-art method, i.e., Primal Sketch (PS) (15.71%), and obviously outperforms Pb with an improvement of recognition rate of 11.84%. The improvement is particularly notable for challenging categories such as 'dog', 'elephant' and 'duck' which have greater intra-class variations than the other categories (e.g., there are far greater number of different types of 'dogs' than 'ipod'). For example, a five-fold increase in classification accuracy was obtained on the 'dog' class. On 'elephant', the increase becomes more than 10-fold. Similarly, on 'duck', although we just achieved 5.17% classification accuracy, it still outperforms other two alternatives, particularly, PS gives zero recognition accuracy on this category. It is interesting that 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 'Beer-mug'. An explanation to this is that, Beer-mugs are relatively simple (object centered with plain background), making final sketches to maintain too few key curves after the filtering process, hence performing slightly worse than Pb. We can see from Fig. 5 that in general sketches generated using our methods keep a similar level of details as those from human.

7. Sketch-based image retrieval

In this section, we present a novel application of sketch-based image retrieval (SBIR) which aims to retrieve natural images by a human drawn sketch query. It is a challenging task because images contain the same objects, but come from different domains (i.e., sketch and real image) produce distinct representations of objects. Therefore, we deal with this problem by converting real images into sketch-like images, which makes sketch-based image retrieval possible (see Fig. 2). More specifically, Histogram of Oriented Gradients (HOG) H^I is extracted for each machine generated sketch, and similarly for the query sketch H^s. Upon retrieval, gallery images are ranked according to histogram distance $d(H^s, H^I)$ between every pair of query sketch and real image.

7.1. Dataset

Flickr15k, which is proposed in [4], serves as the benchmark for our sketch-based image retrieval system. It is currently the largest and most commonly used benchmark for SBIR. It contains approx. 15k photographs sampled from Flickr and manually labeled into 33 categories, and 330 free-hand drawn sketch queries drawn by 10 non-expert sketchers. In our experiment, we utilize images in Flickr15k as retrieval candidates, and the 330 sketches without semantic tags to serve as queries.

7.2. Experimental settings

We compare our proposed sketch-based image retrieval based on sketch generation (SBIR-SG(non-BoW)) with state-of-the-art non-BoW method, i.e., StructureTensor(non-BoW) [41], and six other BoW-based methods, i.e., Gradient Field HOG (GF-HOG) [4] which is the state-of-the-art BoW-based method, SIFT [42], Self-Similarity (SSIM) [43], Shape Context [44], HOG [45] and the Structure Tensor [41]. Similar to [4], (i) for the non-BoW baseline method (i.e., non-BoW Structure-Tensor), we compute the standard HOG descriptor over all edge pixels of query sketch and real images to be retrieved, then the ranking retrieval results are obtained based on the distance between them; (ii) for the six BoW-based baseline methods, all of them employ a BoW strategy but with different feature descriptors, e.g., for the method of GF-HOG, features of GF-HOG are extracted over all local pixels of Canny edge map, then a BoW vocabulary $\mathcal{V}$ is formed via k-means, therefore, a frequency histogram H^I is built for representing each real image by using the previously learned vocabulary $\mathcal{V}$, similarly, a frequency histogram H^s of the query sketch is constructed by using the same vocabulary $\mathcal{V}$. In the end, real images are then ranked according to histogram distance $d(H^s, H^I)$.

In addition, because most of the previous work on SBIR rely on edge detector (e.g., Canny) to work and just focus on the feature extraction [4,43–45], the problem of how sketch generation effects retrieval performance has been largely ignored, we further investigate that how different types of sketch generator contribute to the retrieval performance. In particular, we offer comparison of four sketch generation techniques, namely Canny, Pb, PS and our proposed approach.

7.3. Results and discussion

Quantitative and qualitative results are shown in Table 3 and Fig. 6, respectively. Table 3 reports the Mean Average Precision (MAP) value, which is produced by averaging the Average Precision (AP) over all the 330 sketch queries. MAP is computed by exploiting a widely used implementation for MAP scoring distributed via the TRECVID benchmark. We can observe from Table 3 that our proposed SBIR method achieves 0.1659 MAP score, which 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). In addition, Fig. 6 presents several sketch queries and their retrieval results over 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. The reason behind false positive results returned by our system is that, with only black and white lines, the shape of the generated sketch is coincidentally very close to the query sketch. Furthermore, Fig. 7 demonstrates

Table 3
MAP results comparison. Our proposed method outperforms over all competitors including one state-of-the-art non-BoW method, StructureTensor(non-BoW), and six other BoW-based methods, GF-HOG, HOG, SIFT, SSIM, ShapeContext and StructureTensor.

Methods	Distance measures	Vocabulary size	MAP
SBIR-SG(non-BoW)	Chi^2	–	**0.1659**
StructureTensor(non-BoW)	Tensor distance	–	0.0735
GF-HOG	Histogram Intersection	3500	0.1222
HOG	Chi^2	3000	0.1093
SIFT	Chi^2	1000	0.0911
SSIM	Chi^2	500	0.0957
ShapeContext	Chi^2	3500	0.0814
StructureTensor	Chi^2	500	0.0798

[2] The canny generated sketches are produced using default parameters supplied by MATLAB.

Fig. 6. Example query sketch, and their top ranking results (ranking from left to right) over the Flick15K dataset. Red boxes show the returned irrelevant results. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)

the retrieval performance comparison when different types of sketch generator are utilized in the same SBIR system. It clearly shows that our proposed sketch generator superior than the other competitors over all the 10 groups of sketch query sketch, and it is predictable that Canny perform the worst due to the over-complex generated sketch.

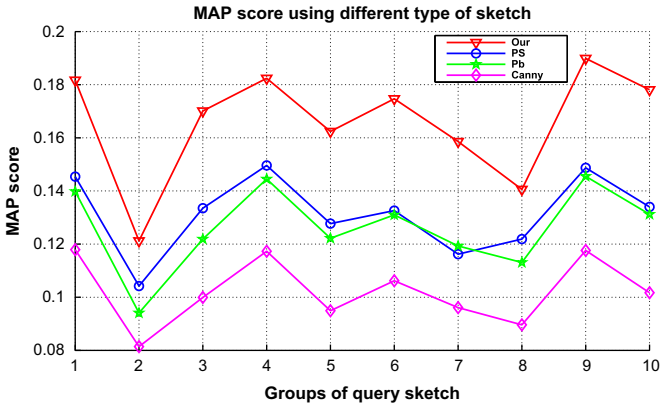


Fig. 7. MAP performance comparison of our generated sketch, Primal sketch (PS), Pb and Canny. From top to bottom: Our, PS, Pb and Canny. It clearly shows that our proposed sketch generator superior than the other competitors over all the 10 groups of sketch query sketch, and it is predictable that Canny perform the worst due to the over-complex generated sketch.

8. Conclusion and future work

In this paper, we presented 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 two commonly used perceptual grouping principles, i.e., continuity and proximity, we were able to develop an effective automated sketching algorithm to simulate how human draw objects. Furthermore, we were able to show that grouping performance could be improved by investigating the relative importance of diverse Gestalt principles. In doing so, a new dataset with human annotations was proposed which makes possible to learn Gestalt confliction. A sketch-based object recognition experiment confirmed the effectiveness of the sketch generation algorithm. Finally, a simple and novel sketch-based image retrieval application was introduced which validated the effectiveness of automatic sketch generation for sketch-based image retrieval.

While the ultimate goal of this work is to generate sketches with the same drawing style as humans, current results still do not closely resemble human. In future work, we intend to design a learning strategy to map realistic edges to impressionistic lines as those drawn by humans to solve this problem.

Acknowledgments

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

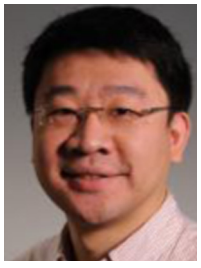
References

[1] M. Eitz, J. Hays, M. Alexa, How do humans sketch objects? ACM Trans. Graph 31 (4) (2012) 44.
[2] Y. Li, Y.-Z. Song, S. Gong, Sketch recognition by ensemble matching of structured features, in: BMVC 2013.
[3] M. Eitz, K. Hildebrand, T. Boubekeur, M. Alexa, Sketch-based image retrieval: Benchmark and bag-of-features descriptors, IEEE Transactions on Visualization and Computer Graphics 17 (11) (2011) 1624–1636.
[4] R. Hu, J.P. Collomosse, A performance evaluation of gradient field HOG descriptor for sketch based image retrieval, Computer Vision and Image Understanding 117 (7) (2013) 790–806.
[5] S. Marvaniya, S. Bhattacharjee, V. Manickavasagam, A. Mittal, Drawing an automatic sketch of deformable objects using only a few images, in: ECCV Workshops 2012.
[6] C. Guo, S.C. Zhu, Y.N. Wu, Primal sketch: Integrating structure and texture, Computer Vision and Image Understanding 106 (1) (2007) 5–19.
[7] J. Wagemans, J.H. Elder, M. Kubovy, S.E. Palmer, M.A. Peterson, M. Singh, R. von der Heydt, A century of Gestalt psychology in visual perception: I. Perceptual grouping and figure-ground organization, Psychological bulletin 138 (6) (2012) 1172–1217.
[8] M. Wertheimer, Laws of organization in perceptual forms, Harcourt, Brace & Jovanovitch, London, 1938.
[9] A. Amir, M. Lindenbaum, A generic grouping algorithm and its quantitative analysis, IEEE Trans, Pattern Anal. Mach. Intell 20 (2) (1998) 168–185.
[10] X. Ren, J. Malik, Learning a classification model for segmentation., in: ICCV 2003.
[11] J.S. Stahl, S. Wang, Edge grouping combining boundary and region information, IEEE Transactions on Image Processing 16 (10) (2007) 2590–2606.
[12] G. Papari, N. Petkov, Adaptive pseudo dilation for gestalt edge grouping and contour detection, IEEE Transactions on Image Processing 17 (10) (2008) 1950–1962.
[13] N. Adluru, L. J. Latecki, R. Lak¨amper, T. Young, X. Bai, A. D. Gross, Contour grouping based on local symmetry., in: ICCV 2007.
[14] Y. Song, X. Bai, P.M. Hall, L. Wang, In search of perceptually salient groupings, IEEE Transactions on Image Processing 20 (4) (2011) 935–947.
[15] J. Wagemans, J. Feldman, S. Gepshtein, R. Kimchi, J.R. Pomerantz, P.A. van der Helm, C. van Leeuwen, A century of Gestalt psychology in visual perception: II, Conceptual and theoretical foundations., Psychological bulletin 138 (6) (2012) 1218–1252.
[16] M. Kubovy, M. van den Berg, The whole is equal to the sum of its parts: A probabilistic model of grouping by proximity and similarity in regular patterns, Psychological Review 115 (1) (2008) 131–154.
[17] Y. Boykov, V. Kolmogorov, An experimental comparison of mincut/max-flow algorithms for energy minimization in vision, IEEE Trans. Pattern Anal. Mach. Intell 26 (9) (2004) 1124–1137.
[18] Y. Boykov, O. Veksler, R. Zabih, Fast approximate energy minimization via graph cuts, IEEE Trans. Pattern Anal. Mach. Intell 23 (11) (2001) 1222–1239.
[19] A. Delong, A. Osokin, H.N. Isack, Y. Boykov, Fast approximate energy minimization with label costs, International Journal of Computer Vision 96 (1) (2012) 1–27.
[20] O. Chapelle, S.S. Keerthi, Efficient algorithms for ranking with svms, Inf. Retr. 13 (3) (2010) 201–215.
[21] Q. Zhu, G. Song, J. Shi, Untangling cycles for contour grouping, in: ICCV 2007.
[22] L.R. Williams, K.K. Thornber, A comparison of measures for detecting natural shapes in cluttered backgrounds, International Journal of Computer Vision 34 (2-3) (1999) 81–96.
[23] J. H. Elder, S. W. Zucker, Computing contour closure, in: ECCV 1996.
[24] Y. Ming, H. Li, X. He, Connected contours: A new contour completion model that respects the closure effect, in: CVPR 2012.
[25] S. M. Bileschi, L. Wolf, Image representations beyond histograms of gradients: The role of gestalt descriptors, in: CVPR 2007.
[26] Z. Ma, A. Leijon, Bayesian estimation of beta mixture models with variational inference, IEEE Trans, Pattern Anal. Mach. Intell 33 (11) (2011) 2160–2173.
[27] Z. Ma, P.K. Rana, J. Taghia, M. Flierl, A. Leijon, Bayesian estimation of dirichlet mixture model with variational inference, Pattern Recognition 47 (9) (2014) 3143–3157.
[28] J.H. Elder, R.M. Goldberg, Ecological statistics of Gestalt laws for the perceptual organization of contours, J.Vis. 2 (4) (2002) 324–353.
[29] L. Nan, A. Sharf, K. Xie, T.-T. Wong, O. Deussen, D. Cohen-Or, B. Chen, Conjoining gestalt rules for abstraction of architectural drawings., SIGGRAPH 2011.
[30] P. Arbelaez, M. Maire, C. Fowlkes, J. Malik, Contour detection and hierarchical image segmentation, IEEE Trans. Pattern Anal. Mach. Intell 33 (5) (2011) 898–916.
[31] R. Kennedy, J. Gallier, J. Shi, Contour cut: Identifying salient contours in images by solving a hermitian eigenvalue problem, in: CVPR 2011.
[32] H. Zhang, K. Zhao, Y. Song, J. Guo, Text extraction from natural scene image: A survey, Neurocomputing (2013).
[33] M. Eitz, K. Hildebrand, T. Boubekeur, M. Alexa, A descriptor for large scale image retrieval based on sketched feature lines, in: SBIM 2009.
[34] R. Hu, M. Barnard, J. P. Collomosse, Gradient field descriptor for sketch based retrieval and localization, in: ICIP 2010.
[35] Y. Qi, J. Guo, Y. Li, H. Zhang, T. Xiang, Y.-Z. Song, Sketching by perceptual grouping, in: ICIP 2013.
[36] T. Joachims, Optimizing search engines using clickthrough data, in: KDD 2002.
[37] R. Mart´ın-F´elez, T. Xiang, Gait recognition by ranking, in: ECCV 2012.
[38] O. Chapelle, Training a support vector machine in the primal, Neural Computation 19 (5) (2007) 1155–1178.
[39] S. Wang, T. Kubota, J.M. Siskind, J. Wang, Salient closed boundary extraction with ratio contour, IEEE Trans. Pattern Anal. Mach. Intell 27 (4) (2005) 546–561.
[40] G. Griffin, A. Holub, P. Perona, The Caltech-256, Tech. rep., California Institute of Technology, 2007.
[41] M. Eitz, K. Hildebrand, T. Boubekeur, M. Alexa, An evaluation ofdescriptors for large-scale image retrieval from sketched feature lines, Computers & Graphics 34 (5) (2010) 482–498.
[42] D.G. Lowe, Distinctive image features from scale-invariant keypoints, International Journal of Computer Vision 60 (2) (2004) 91–110.
[43] E. Shechtman, M. Irani, Matching local self-similarities across images and videos, in: CVPR 2007.
[44] G. Mori, S.J. Belongie, J. Malik, Efficient shape matching using shape contexts, IEEE Trans. Pattern Anal. Mach. Intell 27 (11) (2005) 1832–1837.

[45] N. Dalal, B. Triggs, Histograms of oriented gradients for human detection, in: CVPR 2005.

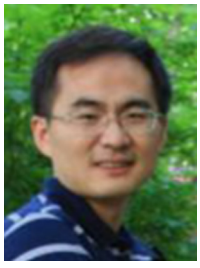
Yonggang Qi is currently a Ph.D. candidate at School of Information and Communication Engineering, Beijing University of Posts and Telecommunications. His research interest is computer vision, particularly focus on perceptual grouping and object recognition. He was a visiting Ph.D. student at Department of Electronic Systems at Aalborg University, Aalborg, Denmark, in 2013.

Jun Guo received his Ph.D. from Tohuku-Gakuin University, in 1993. He is currently the Vice-President of BUPT, a Distinguished Professor at Beijing University of Posts and Telecommunications, and the Dean of the School of Information and Communication Engineering. He is mainly engaged in the research of pattern recognition, web searching, and network management. He has more than 200 publications at International top journals and conferences, including SCIENCE, IEEE Trans. on PAMI, IEICE Trans, ICPR, ICCV, SIGIR. He received numerous international and national awards, including 3 IEEE International Awards, the second prize of Beijing scientific and technological progress, the second prize of the Ministry of Posts and Telecommunications scientific and technological progress.



Yi-Zhe Song is a Lecturer (Assistant Professor) at School of Electronic Engineering and Computer Science, Queen Mary, University of London. He researches into computer vision, computer graphics and their convergence, particularly perceptual grouping, image segmentation (description), cross-domain image analysis, non-photorealistic rendering, with a recent emphasis on human sketch representation, recognition and retrieval. He received both the B.Sc. (first class) and Ph.D. degrees in Computer Science from the Department of Computer Science, University of Bath, UK, in 2003 and 2008, respectively; prior to his doctoral studies, he obtained a Diploma (M.Sc.) degree in Computer Science from the Computer Laboratory, University of Cambridge, UK, in 2004. Prior to 2011, he worked at University of Bath as a Research and Teaching Fellow. He is an Associate Editor of Neurocomputing and member of IEEE and BMVA.

Tao Xiang received the Ph.D. degree in electrical and computer engineering from the National University of Singapore, in 2002. He is currently a Reader (Associate Professor) in the School of Electronic Engineering and Computer Science, Queen Mary University of London. His research interests include computer vision, machine learning, and data mining. He has published over 100 papers in international journals and conferences and co-authored a book, Visual Analysis of Behaviour: From Pixels to Semantics.



Honggang Zhang received the B.S. degree from the Department of Electrical Engineering, Shandong University, in 1996, the Master and Ph.D. degrees from the School of Information Engineering, Beijing University of Posts and Telecommunications (BUPT), in 1999 and 2003, respectively. He worked as a visiting scholar in School of Computer Science, Carnegie Mellon University (CMU) from 2007 to 2008. He is currently an Associate Professor and Director of web search center at BUPT. His research interests include image retrieval, computer vision and pattern recognition. He published more than 30 papers on TPAMI, SCIENCE, Machine Vision and Applications, AAAI, ICPR, ICIP. He is a Senior Member of IEEE.

Zheng-Hua Tan received the B.Sc. and M.Sc. degrees in Electrical Engineering from Hunan University, Changsha, China, in 1990 and 1996, respectively, and the Ph. D. degree in electronic engineering from Shanghai Jiao Tong University, Shanghai, China, in 1999. He is an Associate Professor in the Department of Electronic Systems at Aalborg University, Aalborg, Denmark, which he joined in May 2001. He was a Visiting Scientist at the Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, USA, an Associate Professor in the Department of Electronic Engineering at Shanghai Jiao Tong University, and a Postdoctoral Fellow in the Department of Computer Science at Korea Advanced Institute of Science and Technology, Daejeon, Korea. His research interests include speech and speaker recognition, noise-robust speech processing, multimedia signal and information processing, human–robot interaction, and machine learning. He has published extensively in these areas in refereed journals and conference proceedings. He has served as an Editorial Board Member/Associate Editor for Elsevier Computer Speech and Language, Elsevier Digital Signal Processing and Elsevier Computers and Electrical Engineering. He was a Lead Guest Editor for the IEEE Journal of Selected Topics in Signal Processing. He has served/serves as a Program Co-chair, Area and Session Chair, Tutorial Speaker and Committee Member in many major international conferences.