Image and Vision Computing 46 (2016) 64–73

Contents lists available at ScienceDirect

Image and Vision Computing

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

Sketch retrieval via local dense stroke features☆

Chao Ma [a,*,1], Xiaokang Yang [a], Chongyang Zhang [a], Xiang Ruan [b], Ming-Hsuan Yang [c]

[a] Institute of Image Communication and Network Engineering, Shanghai Jiao Tong University, Shanghai 200240, PR China
[b] Omron Corporation, Kyoto, Japan
[c] Electrical Engineering and Computer Science, University of California, Merced, CA 95344, United States

ARTICLE INFO

Article history:
Received 13 August 2014
Received in revised form 11 July 2015
Accepted 27 November 2015
Available online 22 January 2016

Keywords:
Sketch retrieval
Stroke feature
Poisson based histogram of orientation

ABSTRACT

Sketch retrieval aims at retrieving the most similar sketches from a large database based on one hand-drawn query. Successful retrieval hinges on an effective representation of sketch images and an efficient search method. In this paper, we propose a representation scheme which takes sketch strokes into account with local features, thereby facilitating efficient retrieval with codebooks. Stroke features are detected via densely sampled points on stroke lines with crucial corners as anchor points, from which local gradients are enhanced and described by a quantized histogram of gradients. A codebook is organized in a hierarchical vocabulary tree, which maintains structural information of visual words and enables efficient retrieval in sub-linear time. Experimental results on three data sets demonstrate the merits of the proposed algorithm for effective and efficient sketch retrieval.

1. Introduction

Sketch-based image retrieval, which deals with the problem of retrieving similar images from a large database based on a hand-drawn query, has received considerable attention in recent years [1–8]. Sketches, originating from the contour or skeleton of an object, have long been proposed as an effective intermediate representation for describing essential shape information of objects [9] with numerous applications. In this work, we define a *sketch* as a collection of hand-drawn stroke lines, which can be closed or open as shown in Fig. 1, to describe an object of interest.

As sketches are hand-drawn with free styles to represent objects, sketch retrieval is challenging due to several factors. First, there exist large intra-class differences, as a result of experiential and cognitive differences among individuals, e.g., giraffe sketches drawn by two individuals are likely to be significantly different in terms of shapes (see Fig. 1). Second, there exist small inter-class differences, due to loss of visual details (i.e., texture and appearance), e.g., the sketch of an apple may look similar to that of an orange. Therefore, the key issue for sketch retrieval lies in an effective scheme to represent sketches that takes both inter-class and intra-class differences into consideration.

Recent work on sketch retrieval mainly focuses on retrieving natural images (*sketch-to-image*) on large database [10,3,4,5,7,6,8,11,12], while considerably less attention is paid to retrieving sketches (*sketch-to-sketch*). With the increasing capacity of the sketch dataset (e.g., the TU Berlin dataset [8] is created by crowd sourcing with 20,000 sketches of 250 object categories), it is of great importance to resolve the problem of retrieving sketches on large scale database. Due to large intra-class and small inter-class differences between sketches, it is ineffective to retrieve them simply using shape retrieval algorithms [13–16], where shapes are derived from natural objects with regular and simple contours (rather than hand-drawn). Unlike these simple shapes, sketches are hand-drawn with significant disparities on the number and length of stroke lines even for the same class.

In this paper, we propose an algorithm for efficient and effective sketch matching with focus on *sketch-to-sketch* rather than *sketch-to-image* retrieval based on one hand-drawn query. We represent a sketch image by local features that are distributed evenly on stroke lines. For efficient query and match, local features of a sketch image are described by a quantized histogram of gradients and stored hierarchically in a vocabulary tree. Each sketch image is then represented by the index of tree nodes instead of storing all of the local feature descriptors in a long vector. We show that a straightforward bag-of-words approach with local corner features for sketch retrieval is not effective. Instead, the proposed algorithm focuses on stroke lines of a sketch image with crucial corner points and evenly sampled points, which performs more robustly for sketch retrieval. In addition, the proposed representation scheme facilitates integration with other spatial kernels [17] to capture spatial information of local features and usage of inverted index on tree nodes to speed up quantization of local features. We evaluate the proposed algorithm on three large data sets of hand-drawn sketches. Experimental results on these data sets with more than 20,000 sketch images show that the proposed algorithm performs favorably against state-of-the-art methods in terms of retrieval accuracy and execution time.

☆ This paper has been recommended for acceptance by Seong-Whan Lee, PhD.

* Corresponding author.
E-mail addresses: chaoma@sjtu.edu.cn (C. Ma), xkyang@sjtu.edu.cn (X. Yang), sunny_zhang@sjtu.edu.cn (C. Zhang), gen@omm.ncl.omron.co.jp (X. Ruan), mhyang@ucmerced.edu (M.-H. Yang).

[1] C. Ma was sponsored by China Scholarship Council and took a two-year study in University of California at Merced.

http://dx.doi.org/10.1016/j.imavis.2015.11.007
0262-8856/

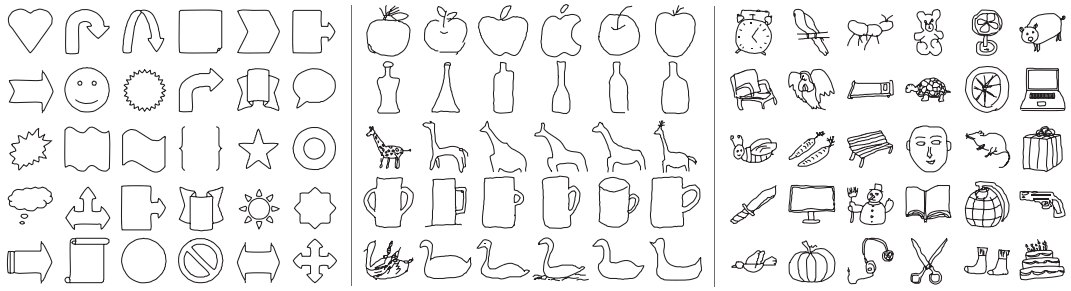

Fig. 1. Sketch images. From left to right: Office icon library, hand-drawn ETHZ shape [1] (apple, bottle, giraffe, mug and swan) and TU Berlin sketch [8] data set. Notice that the office icons have minor inter-class differences (e.g., the right arrows) while the ETHZ shapes have large intra-class differences (e.g., the giraffes and swans).

Compared to early results of this work [18], we show effectiveness of the proposed local features which use edge information of foreground and background regions to better represent sketches (Section 3.2); we analyze the histogram distribution of the number of the stroke points each sketch contains to set the optimal number of dense stroke features (Section 4.1); and we present more experimental results and discuss the application scenarios of the proposed sketch retrieval method (Figs. 9 and 11).

2. Related work and problem context

Existing methods on primal sketches focus on representation schemes based on primitive features such as edges as well as curves. In [19,20], sketches are stored in the form of multiple strokes and retrieved by using the shape of each stroke and the spatial relationship between them. When sketches are simple close-formed hand-drawings, the Fourier transformed boundary is used as shape feature for representation [21,22]. By applying the 2-D Fourier transform on a polar shape image, an adapted Fourier descriptor is proposed to represent sketches in terms of contour [23]. In [10], Rui et al. review sketch-based image retrieval with focus on the contour-based and region-based representation schemes. However, due to simplicity of representation for sketches, these methods are ineffective to represent and index complex sketches of a large scale database.

In recent years, much attention has been paid to sketch retrieval due to its wide applications for intelligent human computer interfaces. Ferrari et al. construct the ETHZ shape database [1] and k-adjacent segments to detect objects in images based on hand-drawn examples where image edges are partitioned into contour segments and organized in chains. In addition, shape modeling [24], Chamfer matching [2], partial shape matching [11,5], and discriminative latent shape models [3] have been applied to sketch-based object detection and localization. However, these methods mainly focus on retrieving objects in images using one good query sketch (i.e., sketch to images), and thus they are less effective for complex sketch retrieval (i.e., sketch to sketches) when there exist large intra-class and small inter-class differences.

To retrieve object images from a large database, feature descriptors are commonly extracted for indexing and matching sketches. The descriptor-based representations in the literature can be roughly categorized as either global or local. In [25], Chalechale et al. exploit angular and spatial distributions of edge pixels to represent holistic features, which is similar to the shape context information [26]. Shao et al. [27] instead extract key points along stroke lines to account for shape difference between sketches. In [4], Cao et al. propose an edge descriptor for sketch based image retrieval. As the underlying matching method is based on Chamfer distance with focus on global geometric information, the proposed edge-based descriptors are less effective in describing complex sketches. On the other hand, local feature based methods are more robust to represent complex sketches. In [6,7], Eitz et al. leverage the bag-of-words formulation with SIFT descriptors for sketch-based image retrieval (i.e., sketch to images). Hu et al. [12] also present a bag-of-words approach based on multiple descriptors and histogram of image gradients for sketch-based image retrieval. Both these methods use grid-based sampling methods to locate local features and the k-means clustering algorithm to learn codebooks for following indexing scheme. In contrast, we focus more on selecting the most representative local features that are evenly distributed on strokes including crucial corner points and describing local features via a coarsely quantized histogram of gradients. We note that existing methods focus on sketch *classification* [4,6,12,25] (i.e., sketch to images with object types) or detection (i.e., sketch to images with object locations) [24,1–5], rather than *retrieval* (i.e., sketch to sketches with ranking) based on one query.

3. Proposed algorithm

We present the proposed algorithm for sketch retrieval via stroke features, which consists of three components: selecting the most representative stroke points, describing stroke features using a quantized histogram of gradients, and representing sketch images using a hierarchical vocabulary tree for matching. Fig. 2 shows the main steps of the proposed method. In the training phase, we extract all the local features of sketch images and store these local features in a hierarchical vocabulary tree similar to [28], where each sketch is indexed by the frequency of tree nodes to which its local features belong. In the retrieval phase, each query sketch is represented via its stroke features and the same vocabulary tree. For efficient retrieval, this vocabulary tree can be easily integrated with an inverted indexing method, which tallies the identities (labels) of training sketch images that have local features belonging to each node. Retrieval can thus be carried out by counting the hit frequency between a query sketch and the inverted list with identities of training images.

3.1. Densely sampled stroke points

In this work, we use local stroke features to represent sketches instead of contour segments [1,2,3,5]. While several key point detectors such as difference of Gaussian (DoG) [29], Hessian operator [30] and Harris–Laplace detector [31] can be used for locating local features, they are designed mainly for finding salient points. As salient key points are usually sparsely distributed over an image, it is of great importance to capture their spatial relationship for better object representation (rather than simple bag-of-words approaches). Grid-based as well as random sampling methods have also been proposed to locate local features [32,33]. As sketch images consist of strokes with no textural information, it is essential to select the most representative stroke points for local features. Since corners and end pixels of strokes always encode important geometric information of a sketch, they are used as anchor positions for dense sampling to encode shape information properly.

In addition, sketches with complex shapes are not compactly represented by grid-based sampling well (e.g, local features detected by grid points may capture few stroke points). Thus, we propose to extract evenly distributed stroke points based on anchor corners. For sketch retrieval, each image is normalized to a canonical size and the Harris