

Free-hand sketch recognition by multi-kernel feature learning[☆]



Yi Li^{*}, Timothy M. Hospedales, Yi-Zhe Song, Shaogang Gong

Computer Vision Lab, School of Electronic Engineering and Computer Science, Queen Mary University of London, Mile End Road, London E1 4NS, United Kingdom

ARTICLE INFO

Article history:

Received 3 July 2014

Accepted 6 February 2015

Available online 19 February 2015

Keywords:

Sketch recognition

Star graph

Ensemble matching

Multiple kernel learning

Attributes

ABSTRACT

Free-hand sketch recognition has become increasingly popular due to the recent expansion of portable touch-screen devices. However, the problem is non-trivial due to the complexity of internal structures that leads to intra-class variations, coupled with the sparsity in visual cues that results in inter-class ambiguities. In order to address the structural complexity, a novel structured representation for sketches is proposed to capture the holistic structure of a sketch. Moreover, to overcome the visual cue sparsity problem and therefore achieve state-of-the-art recognition performance, we propose a Multiple Kernel Learning (MKL) framework for sketch recognition, fusing several features common to sketches. We evaluate the performance of all the proposed techniques on the most diverse sketch dataset to date (Mathias et al., 2012), and offer detailed and systematic analyses of the performance of different features and representations, including a breakdown by sketch-super-category. Finally, we investigate the use of attributes as a high-level feature for sketches and show how this complements low-level features for improving recognition performance under the MKL framework, and consequently explore novel applications such as attribute-based retrieval.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

Throughout human civilization, sketch has been used as a basic form of communication. Examples of human sketches from ancient times can still be found in pre-historic cave art and pictograms nowadays. With the rapid emergence of portable touchscreen devices, sketches became much easier to obtain and are often a few finger sweeps away. This movement consequently led to an ever growing interest in free-hand sketch analysis from the computer vision community, where researchers investigated the feasibility of utilizing sketches in many novel tasks such as automatic sketch recognition [1] and sketch-based image retrieval (SBIR) [2–5]. It has even been argued that sketches are more expressive than raw text when retrieving images [6,3], and are able to capture visual memory of natural scenes [7].

Nevertheless, the task of automatically recognizing free-hand human sketches remains nontrivial, mainly due to the relatively large intra-class variations and inter-class ambiguities as opposed to images and other forms of sketches traditionally studied (e.g., CAD (Computer-Aided Design) drawings [8,9]). More specifically: (i) sketches generally capture complex structures in abstract forms,

a characteristic that is more evident in free-hand sketches where the depicting process is heavily unconstrained in terms of style and drawing ability; (ii) sketches, unlike conventional images, are naturally sparse in visual cues (e.g., without color and texture), this consequently makes applications of traditional image-oriented algorithms nontrivial. These unique properties of free-hand sketches ultimately render traditional shape/contour matching techniques inapplicable [10,11]. Fig. 1 offers a visual comparison of inputs used for shape matching (Fig. 1(a)) and human free-hand sketches (Fig. 1(b)). As can be seen, sketches are generally abstract, lack visual cues and it is often subtle internal structure differences that disambiguate one category from another. Meanwhile, differences between users in choices of abstraction and detail results in large intra-class variations.

Prior work on sketches typically addresses the feature sparsity challenge by densely sampling on a grid [1,4] or along the edges [2], or utilizing larger patches [1,12]. And most sketch recognition [1] or SBIR methods [2–5], uniformly employ a BoF (Bag-of-features) representation, in which holistic structure information is lost. Techniques taking account of structural information are commonly found in the image domain. Notable ones include for example, spatial pyramid matching [13] and spatial BoF [14]. However, these methods either use a rigid grid scheme [13] or presume that a dominant direction (either linear or circular) exists for the arrangement of the image features. But due to the abstractness, deformation and large variations of sketches, those schemes are not flexible enough to capture sketch structure. To the best of

[☆] This paper has been recommended for acceptance by Barbara Caputo.

^{*} Corresponding author.

E-mail addresses: yi.li@qmul.ac.uk (Y. Li), t.hospedales@qmul.ac.uk

(T.M. Hospedales), yizhe.song@qmul.ac.uk (Y.-Z. Song), s.gong@qmul.ac.uk (S. Gong).

our knowledge, only a few methods [6,15] employ a hierarchical topology graph to encode holistic structure of sketches. However, they are limited to working with CAD and clip-art [1] drawings which are clean and topologically separable, whereas sketch segmentation itself is an open problem under active research [16].

To address the complexity of internal structures, we propose a mid-level representation to capture the holistic structure of sketches. More specifically, we employ a star graph to encode both local features and holistic structure of a sketch and exploit ensemble matching as a similarity measure. A standard star graph, also known as an ensemble, has an arbitrarily assigned center with all the graph nodes connected to it by edges. The nodes represent feature points in the image and store the corresponding feature descriptors. A star graph encodes both direction and distance of each node to the center in the edges' weights. Detailed comparisons of different features and representations are performed on support vector machine (SVM) classifiers, and the results clearly show the advantage of the proposed star graph representation.

Furthermore, although different features or representations have different levels of performance, we argue that all features contain some potentially complementary information, at least for some classes, and should ideally be used together. We therefore address the cue sparsity problem via Multiple Kernel Learning (MKL), aiming at fully utilizing the discriminative power of all features and eliminating the both bias imposed by any single feature, as well as the design challenge of selecting the 'best' feature for an application. Our experiment on the human free-hand sketch dataset with the most categories to date [1] confirms state-of-the-art performance of MKL on sketch recognition over approaches [1,17] employing BoF representation of popular features including Histogram of Oriented Gradient (HOG) [18], Self-Similarity (SSIM) [19] and Daisy [20], and the star graph representation constructed on HOG feature. Somewhat more subtly, but equally importantly, the same strategy addresses the open design challenge of deciding which similarity metric [5] to use in a given application.

The dataset [1] we evaluate on has as many as 250 categories. In order to show how different representations benefit certain categories more clearly, we introduce the concept of super-categories, which is defined as a superset of basic categories that share a higher-level semantic property (e.g., animal, plant). We found that although the star graph is generally best, different representations tend to favor different super-categories. By using all the features together, MKL obtains the best performance overall on all super-categories.

An interesting finding from the super-category analysis is that the confusions inside super-categories are much bigger than those

between super-categories. This is especially true for large super-categories such as animal and vehicle. It is hypothesized that higher-level semantic properties shared among categories (e.g., spots on the body of a giraffe or butterfly) could help to remove ambiguity within a super-category – a hypothesis that was found to be successful in the image domain [21–23]. Inspired by Lampert et al. [21], this work performs a preliminary study on how sketch attributes can benefit sketch recognition by constructing an attribute kernel within the MKL framework. The experiment is carried out on the animal super-category with classic animal attributes from [21] as well as additional attributes obtained from WordNet [24]. Experimental results show attributes to be effective in improving recognition performance inside super-categories.

Finally, going beyond simple recognition/retrieval of sketch categories, we show how the high-level semantic nature of attribute features can be used to enable novel applications. We demonstrate attribute-based retrieval (query by description rather than category; e.g., stripy), and joint category-attribute retrieval (find a long-leg ant, etc). The attribute classifiers can be further used to offer semantic-level rankings to sketches, for instance telling which zebra is stripier.

Our preliminary work [17] introduced the star graph: a spatially structured representation to model the structural complexity of sketches. By further combining with a category filtering step, it significantly improved the state-of-the-art sketch recognition performance [1]. This paper extends that work to further improve the sketch recognition performance and demonstrate some new interesting applications on sketches. More specifically: (i) we propose a MKL model to utilize multiple features, representations and similarity metrics, including star graph, to address the visual cue sparsity problem and surpass prior state-of-the-art performance; (ii) we demonstrate for the first time sketches attributes, and their value for both sketch recognition and enabling new retrieval applications.

2. Related work

2.1. Towards free-hand sketches

There are several categories of sketches each possessing different sophistication levels and characteristics, they include CAD drawings [8,9], artistic sketches (clip-art drawings) [25,26], and free-hand human sketches [2–4]. CAD drawings are generated by designers using professional software, where standard building blocks are used to construct more sophisticated entities. As a result, CAD drawings often show clear topological structure, and

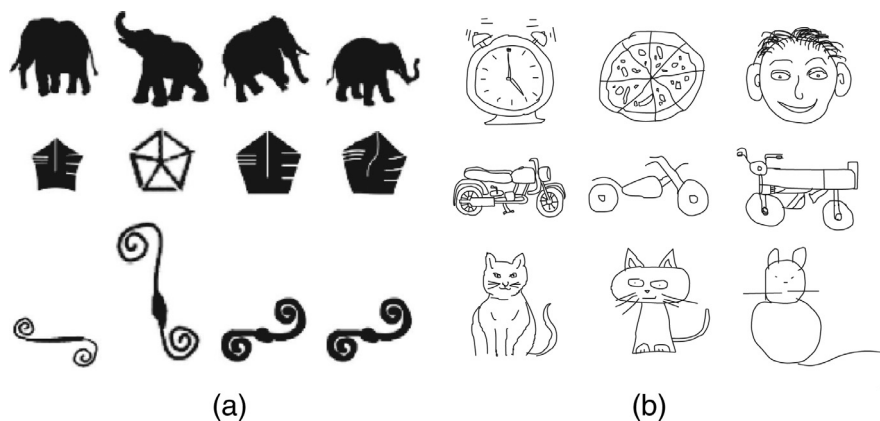


Fig. 1. (a) Typical inputs of shape matching are generally silhouettes with quite simple internal structures [11]; three rows each corresponding to one object category; (b) human sketches generally have more complicated internal structures. Human sketches often exhibit similar silhouettes, however it is the internal structures that make them different, e.g., the alarm clock, the pizza, and the face shown in the first row in (b), and since people have different drawing styles, abstraction level and completeness of internal structures for the same object categories tend to differ, see motorbike and cat.

Download English Version:

<https://daneshyari.com/en/article/525721>

Download Persian Version:

<https://daneshyari.com/article/525721>

[Daneshyari.com](https://daneshyari.com)