Learning to classify human object sketches

Mathias Eitz*
TU Berlin

James Hays†
Brown University

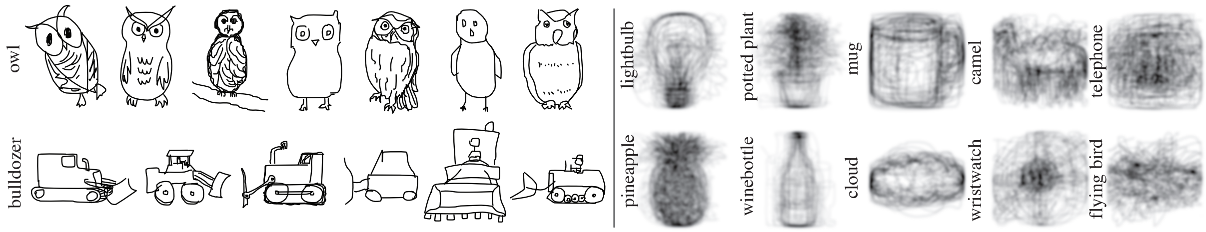

Figure 1: *Left: sample sketches from two categories, right: averages of 30 sketches per category.*

Abstract

We present ongoing work on object category recognition from binary human outline sketches. We first define a novel set of 187 "sketchable" object categories by extracting the labels of the most frequent objects in the LabelMe dataset. In a large-scale experiment, we then gather a dataset of over 5,500 human sketches, evenly distributed over all categories. We show that by training multi-class support vector machines on this dataset, we can classify novel sketches with high accuracy. We demonstrate this in an interactive sketching application that progressively updates its category prediction as users add more strokes to a sketch.

1 Introduction

Sketching is a common means of visual communication often used for conveying rough visual ideas as in architectural drawings, design studies, comics, or movie storyboards. There exists significant prior research on retrieving images or 3d models based on sketches. The assumption in all of these works is that, in some well-engineered feature space, sketched objects will resemble their real-world counterparts. But this fundamental assumption is often violated – most humans are not faithful artists. Instead people use *shared, iconic* representations of objects (e.g. stick figures) or they make dramatic simplifications or exaggerations. Because the relationship between sketched and real objects is so abstract, to recognize sketched objects one must *learn* from a training database of real sketches. Because people represent the same object using differing degrees of realism and distinct drawing styles (see Fig. 1, left), we believe that a successful approach can only be based on a dataset that provides a sufficiently *dense* sampling of that space, i.e. we need a large training dataset of sketches. Although both shape and proportions of a sketched object may be far from that of the corresponding real object, and at the same time sketches are an impoverished visual representation, humans are amazingly accurate at interpreting such sketches. In this paper we demonstrate our ongoing work on trying to teach computers classify sketched objects just as humans do effortlessly.

2 Dataset & Classification

To classify sketches we address four main tasks: 1) defining a set of object categories – ideally those would represent the most common objects in our world; 2) creating a dataset of sketches with diverse samples for each category; 3) defining low-level features for representing the sketches and finally 4) training classifiers from our dataset such that we can accurately recognize novel sketches. We have defined a list of common object categories by computing the 1,000 most frequent objects from the LabelMe dataset and collapsing semantically similar categories. This resulted in 187 object categories, mainly containing common objects such as airplane, house, cup, and horse. In a large-scale user experiment, we asked humans to sketch such objects given only the category name. We instructed them to a) draw sketches that would be "clearly recognizable to other humans" as belonging to a given category, b) use outlines only and c) avoid context around the actual object. Currently, the dataset contains 30 sketches in each category for a total of about 5,500 sketches. We have performed this experiment using Amazon Mechanical Turk. For learning on this dataset we perform two main steps: we employ a bag-of-features approach [Sivic and Zisserman 2003] and use SIFT-like descriptors to represent sketches as histograms of visual words [Eitz et al. 2011]. We train one-vs-all classifiers using support vector machines with RBF kernels. We find the best model by performing grid search over the parameters space of the SVM and use 5-fold cross-validation to avoid overfitting. The best-performing model achieves an accuracy of about 37%. This is a very reasonable result considering that chance lies at about 0.54%. We additionally demonstrate the subjectively very good performance of the resulting model in an interactive application that progressively visualizes classification results as the user adds strokes to a sketch (please see the accompanying video).

3 Conclusion

We have demonstrated that – given a large dataset of sketches – reasonable classification rates can be achieved, limited primarily (we believe) by the bag-of-features representation which does not encode any spatial information. Clearly, constructing better features, and extending and analyzing the dataset are promising areas for future work. Finally, the large dataset in itself (which we plan to provide as a free resource) as well as the semantic sketch classification will be highly beneficial for applications such as sketch-based image and 3D model retrieval.

References

EITZ, M., HILDEBRAND, K., BOUBEKEUR, T., AND ALEXA, M. 2011. Sketch-based image retrieval: benchmark and bag-of-features descriptors. *IEEE Trans. Vis. Comp. Graph.*. Preprints.

SIVIC, J., AND ZISSERMAN, A. 2003. Video Google: A Text Retrieval Approach to Object Matching in Videos. In *IEEE ICCV*.

*e-mail: m.eitz@tu-berlin.de
†e-mail: hays@cs.brown.edu