Fast and Accurate Image Matching with Cascade Hashing for 3D Reconstruction

Cheng Guan

Jun 30, 2018

1. Elements of the Comparison

Last time, I learn the main introduction of the paper. Today I read the paper deeply.The authors use statitis of 7 datasets (5 for scenes and 2 for objects) here to describe the Fig. 1, followed by their model evaluation part. See supplementay for details on human data gathering protocols on datasets.
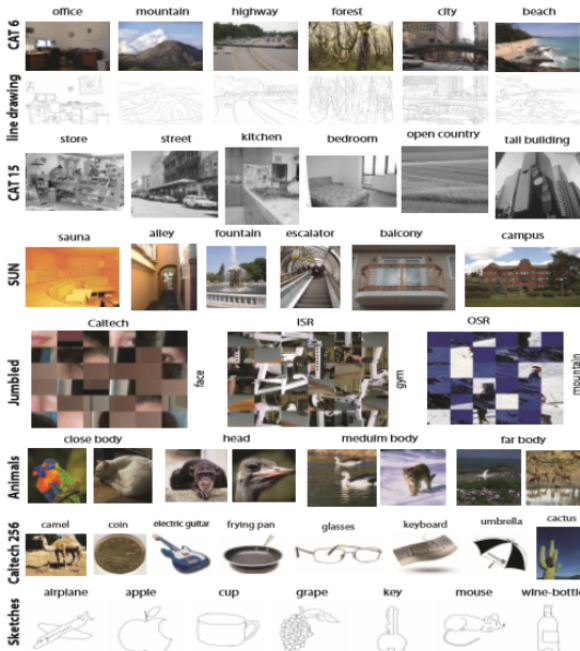

Figure 1. Sample images from scene and object datasets used in this study

1.1. Human Studies

6-CAT:[6] includes 3,866 color photographs: beaches (553), city (648), forest (730), highway (563), mountain (720), and offices (652) with resolution of 800×600 pixels. Line drawings of about 80 images per category (475 total) were taken from a small part of the Lotus Hill Dataset, onto which contours were traced by trained artists.

15-CAT:[3] contains 4,578 gray-level images of size about 300×200 including coast(360), forest(328), highway (260), mountain (374), street (392), tall buildings (356), bedroom (216), kitchen (210), street (292), tall buildings (356), inside city (308), open country (410 images), suburb (241),industrial (311), living room (289), industry (308), and store (315). The first 8 categories were called the 8-CAT dataset.

SUN:[7] is an extensive scene database containing 899 categories and 130,519 images. The authors experiment with 397 well-sampled categories to evaluate state-of-the-art algorithms and their accuracy versus humans.

Animals:[5] contains 600 target images(256×256 pixels; 150 *close body*, 150 *far body*, 150 *head*, and 150 *medium body*) and 600 distractors. The advantage of the data set is that it provides the accuracy of the on $90°$ and $180°$ rotated image and allows the invariance analysis of the model.

Jumbled images:[4] Humans can recognize jumbled images as those in Fig. 1, 4th row. Indeed, Parikh[4] showed a a majority-vote accumulation over human classification of the individual blocks is a good predictor of human responses of the entire jumbled images.

Caltech-256:[2] This dataset, one of the most challenging dataset of object recognition, some shortcomings of CalTeC10-101 data set are corrected by introducing different lighting conditions, poses, background, size and camera systematics.

Sketch images:[1] This dataset contains non-expert sketches of everyday objects such as teapot or car. There are 20,000 unique sketches evenly distributed over 250 categories (*i.e.*, 80 images per category). In perceptual research, humans can correctly identify the object category 73.1% of time (chance is 0.4%). Given a random sketch, participants are asked to select the best fitting category from the set of 250 object categories.

References

[1] M. Eitz, J. Hays, and M. Alexa. How do humans sketch objects? *ACM TOG*, 2012. 1

[2] G. Griffin, A. Holub, and P. Perona. Caltech-256 object category dataset. *California Institute of Technology*, 2007. 1

[3] S. Lazebnik, C. Schmid, and J. Ponce. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In *CVPR*, 2006. 1

[4] D. Parikh. Recognizing jumbled images: The role of local and global information in image classification. In *ICCV*, 2011. 1

[5] T. Serre, A. Oliva, and T. Poggio. A feedforward architecture accounts for rapid categorization. *PNAS*, 2007. 1

[6] D. Walther, B. Chai, E. Caddigan, D. Beck, and L. Fei-Fei. Simple line drawings suffice for functional MRI decoding of natural scene categories. *PNAS*, 2011. 1

[7] J. Xiao, J. Hays, K. A. Ehinger, A. Oliva, and A. Torralba. SUN database: Large-scale scene recognition from abbey to zoo. In *CVPR*, 2010. 1