Finding "good" features for natural scene classification

Eamon Caddigan[1,2], Dirk B. Walther[2], Barry Chai[3], Diane M. Beck[1,2], Li Fei-Fei[3,4]

[1]Department of Psychology, University of Illinois at Urbana-Champaign, [2]Beckman Institute, University of Illinois at Urbana-Champaign, [3]Computer Science Department, Princeton University, [4]Psychology Department, Princeton University

Natural scene categorization by humans

- Readily assign basic-level categories (Tversky & Hemenway, 1983)
- Rapid and efficient (Thorpe et al., 1996)
- Needs little visual attention (Fei-Fei et al., 2002)
- Require localized spectral information (Loschky et al., 2007)

Natural scene categorization by computers

- Power spectrum (magnitude of spatial frequencies; Oliva & Torralba, 2001)
- Bag of words (local texture descriptors; Fei-Fei & Perona, 2006)
- Spatial pyramid to vary localization (Lazebnik et al., 2006)

An integrative approach for feature analysis

Dataset

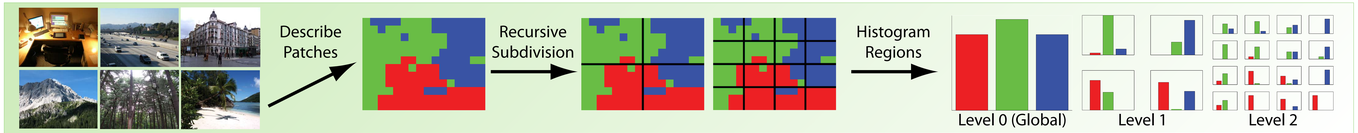
Six natural categories: beaches, cities, forests, highways, mountains, and offices.

100 images per category; 50 rated as "good", 50 rated as "bad" (see Torralbo et al., poster 33.303).

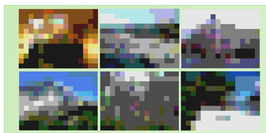
Method

Hold patch size (16 x 16 pixels), grid spacing (8 pixels), and classifier (Linear SVM) constant.

Measure the following for each feature:
- The degree of localization resulting in the highest classification rate
- The correlation between classification errors and human behavior
- The difference between good and bad category exemplars

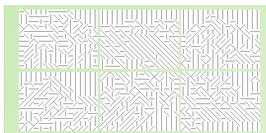


Color

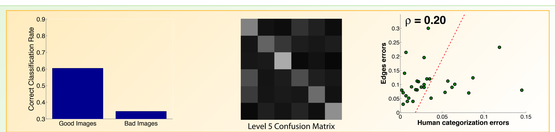


- Average color in each patch
- 5 histogram bins per color channel (R, G, and B) for 125 bins in total

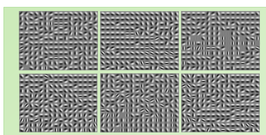
Edges



- Detect Canny edges in different orientations
- The gradient was calculated in four directions
- Non-maximum suppression and hysteresis thresholding traced good edges

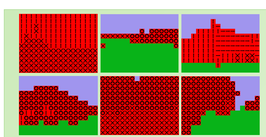


Gabor filters

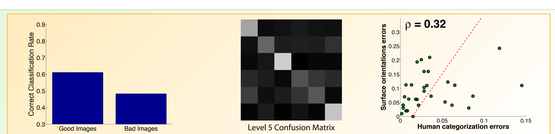


- This feature captures the dominant orientation and frequency of each patch
- Gabor filters at 4 scales and 8 orientations
- Find the filter that exhibited the strongest response when centered over the patch

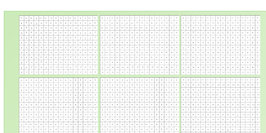
Surface orientations



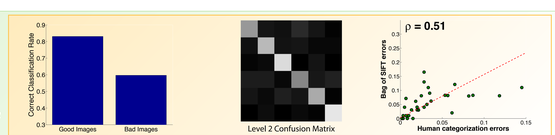
- Reconstructed using Hoiem, et al. (2005)
- Histogrammed into seven bins: ground plane, sky, and vertical surfaces that are one of: solid, porous, frontal, 45 degrees out-of-plane, or 135 degrees out-of-plane
- Surface orientations capture 3D scene layout



Bag of SIFT

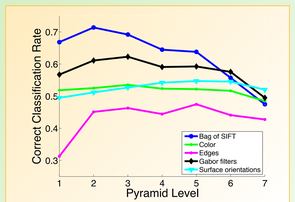


- Same approach taken by Lazebnik, et al., but restricted to a single level of the pyramid
- 200 word dictionary created by clustering all patch descriptors
- SIFT features exhibit invariance to rotation and scale, similar to neurons in primate IT (Lowe, 2004)



Feature localization

- All features benefit from some localization
- No improvement for any feature beyond level 6



Good and bad images

- Classification is consistently worse for images rated as "bad"

Correlating error patterns

- One way to compare human and computer scene classification
- Features with high classification rates can show different error patterns than humans (e.g., Gabor filters)

References

Fei-Fei, L., & Perona, P. (2005). A Bayesian Hierarchical Model for Learning Natural Scene Categories. IEEE Comp. Vis. Patt. Recog.

Fei-Fei, L., VanRullen, R., Koch, C., & Perona, P. (2002). Rapid natural scene categorization in the near absence of attention. Proc. Natl. Acad. Sci, 99, 8378-8383.

Hoiem, D., Efros, A. A., & Hebert, M. (2005). Geometric Context from a Single Image. ICCV.

Lazebnik, S., Schmid, C. & Ponce, J. (2006). Beyond Bags of Features: Spatial Pyramid Matching for Recognizing Natural Scene Categories. IEEE Comp. Vis. Patt. Recog.

Loschky, L. C., Sethi, A., Simons, D. J., Pydimarri, T. N.,Ochs, D., & Corbeille, J. L. (2007). The importanceof information localization in scene gist recognition. Journal Of Experimental Psychology: Human Per-ception & Performance, 33, 1431–1450.

Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. Int. Journal of Computer Vision, 60 (2), 91-110.

Oliva, A., & Torralba, A. (2001). Modeling the shape of the scene: A holistic representation of the spatial envelope. Int. Journal of Computer Vision, 42, 145-175.

Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. Nature, 381, 520-522.

Torralbo, A., Chai, B., Caddigan, E., Walther, D., Beck, D., & Fei-Fei, L. (2009). Categorization of good and bad examples of natural scene categories. Presented at the 9th annual meeting of the Vision Sciences Society.

Tversky, B., & Hemenway, K. (1983). Categories of Environmental Scenes. Cognitive Psychology, 15, 121-149.