

# Large Image Databases and Small Codes for Object Recognition

**Rob Fergus**  
Courant Institute,  
New York University  
fergus@cs.nyu.edu

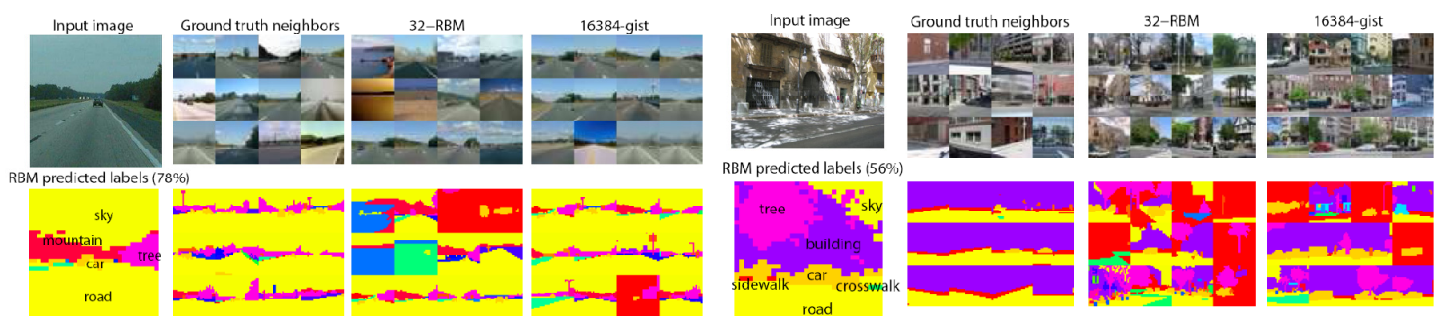
**Antonio Torralba**  
CSAIL, MIT  
torralba@csail.mit.edu

**Yair Weiss**  
Hebrew University  
yweiss@cs.huji.ac.il

**William T. Freeman**  
CSAIL, MIT  
billf@csail.mit.edu

With the advent of the Internet, billions of images are now freely available online and constitute a dense sampling of the visual world. Using a variety of non-parametric methods, we explore this world with the aid of a large dataset of 79,302,017 images collected from the Web. Motivated by psychophysical results showing the remarkable tolerance of the human visual system to degradations in image resolution, the images in the dataset are stored as 32x32 color images. Each image is loosely labeled with one of the 75,062 non-abstract nouns in English, as listed in the Wordnet lexical database. Hence the image database gives a comprehensive coverage of all object categories and scenes. The semantic information from Wordnet can be used in conjunction with nearest-neighbor methods to perform object classification over a range of semantic levels minimizing the effects of labeling noise. For certain classes that are particularly prevalent in the dataset, such as people, we are able to demonstrate a recognition performance comparable to class-specific Viola-Jones style detectors.

In the second part of the talk, we present efficient image search and scene matching techniques that are not only fast, but also require very little memory, enabling their use on standard hardware or even on handheld devices. Our approach uses the Semantic Hashing idea of Salakhutdinov and Hinton [1], based on Restricted Boltzmann Machines [2] to convert the Gist descriptor (a real valued vector that describes orientation energies at different scales and orientations within an image) to a compact binary code, with a few hundred **bits** per image. Using our scheme, it is possible to perform real-time searches on our Internet image database using a single large PC and obtain recognition results comparable to the full descriptor. Using our codes on high quality labeled images from the LabelMe database gives surprisingly powerful recognition results using simple nearest neighbor techniques, as shown below:



Two examples of non-parametric recognition using the LabelMe database. Top row, from L to R: query image; ground-truth neighbors using manual object annotations; neighbors found using our compact 32-bit binary code; neighbors using original Gist descriptor (16384 bits). Bottom, L to R: voted segmentation of query image (with % pixels correctly labeled); manual annotations of neighboring images for each of the three sets.

## References

- [1] R. R. Salakhutdinov and G. E. Hinton. Semantic hashing. In *SIGIR workshop on Information Retrieval and applications of Graphical Models*, 2007.
- [2] G. E. Hinton and R. R. Salakhutdinov. Reducing the dimensionality of data with neural networks. *Nature*, 313(5786):504–507, July 2006.