

Alexander Freytag, Johannes Rühle, Paul Bodesheim, Erik Rodner, and Joachim Denzler

Live experiment
at the poster!

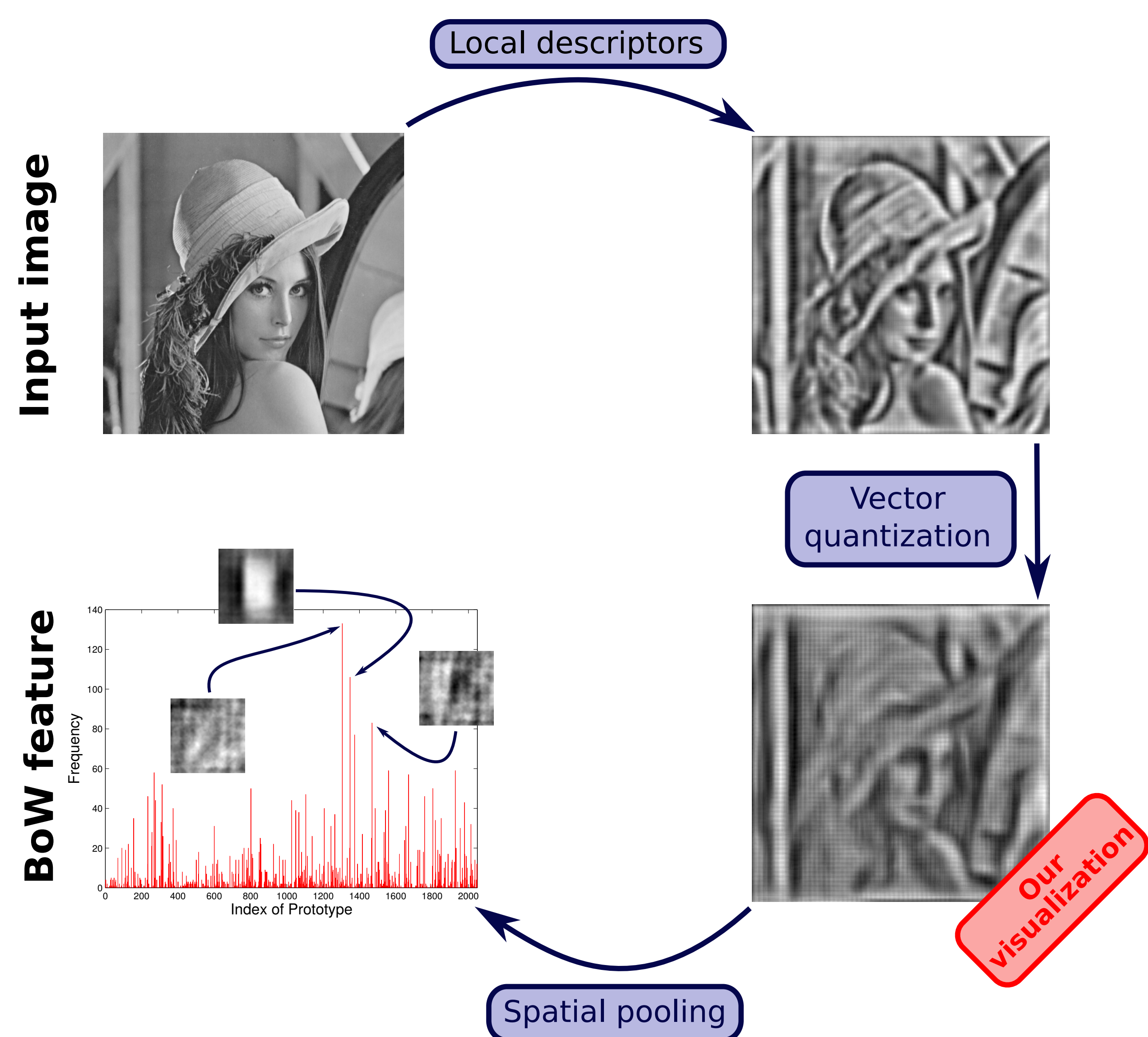
Computer Vision Group, Friedrich Schiller University of Jena
Ernst-Abbe-Platz 2, 07743 Jena, Germany

<http://www.inf-cv.uni-jena.de/>

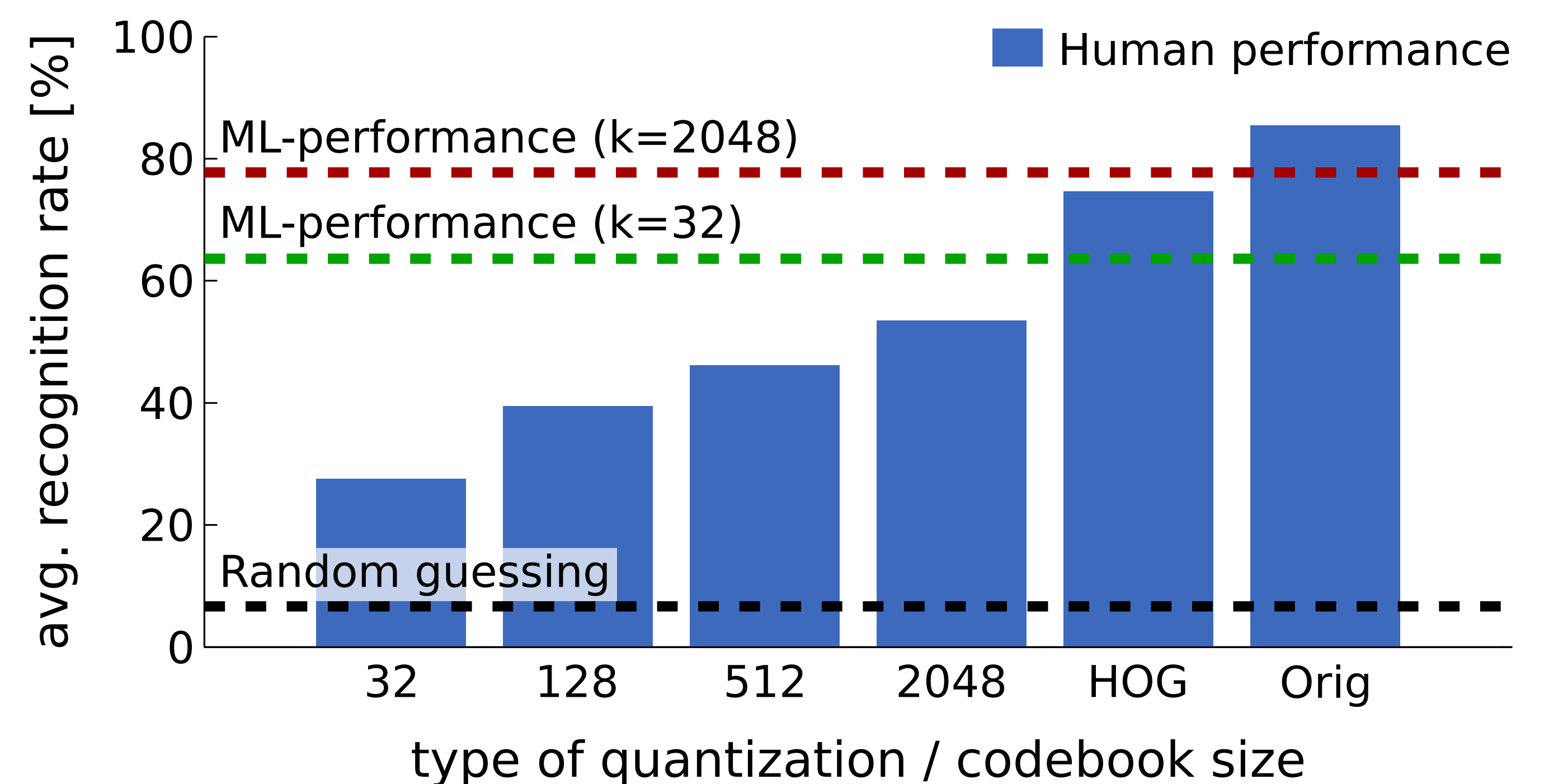
Online experiment:
<http://goo.gl/Vffr9y>

Motivation and goal

- Understand quantization effects in bag-of-words models.
- Measure the loss of visual information for image categorization with respect to the quantization level.



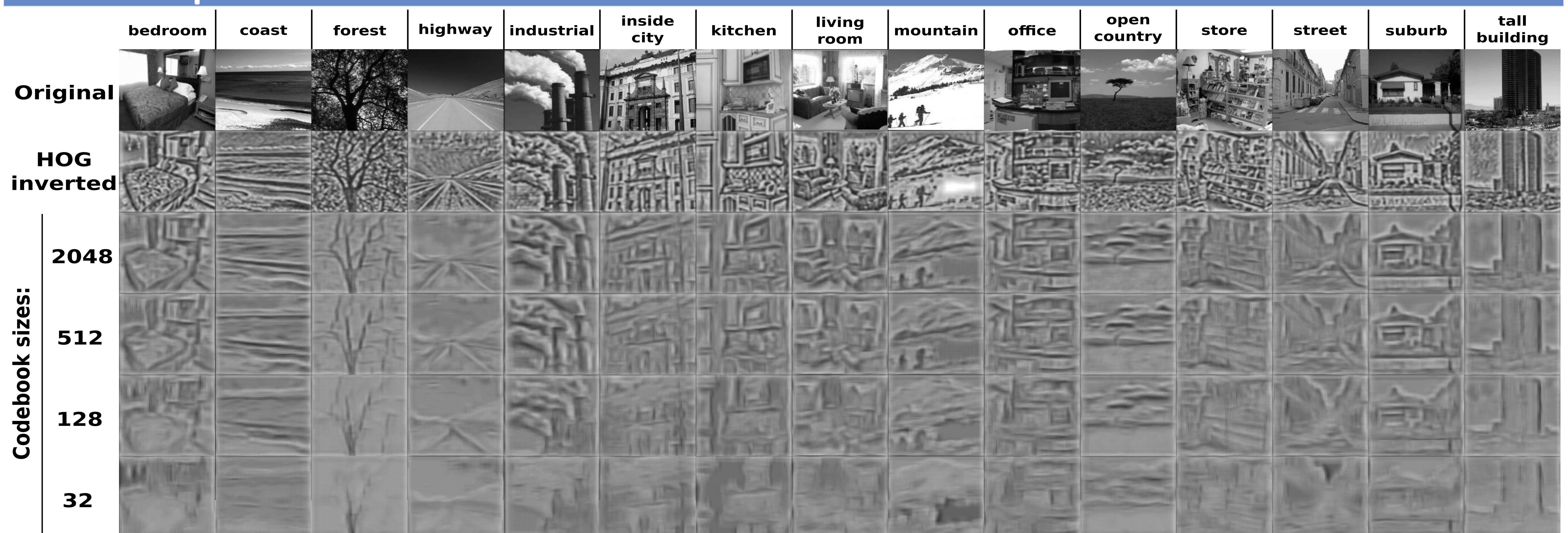
Comparing human and machine performance (15 scenes dataset)



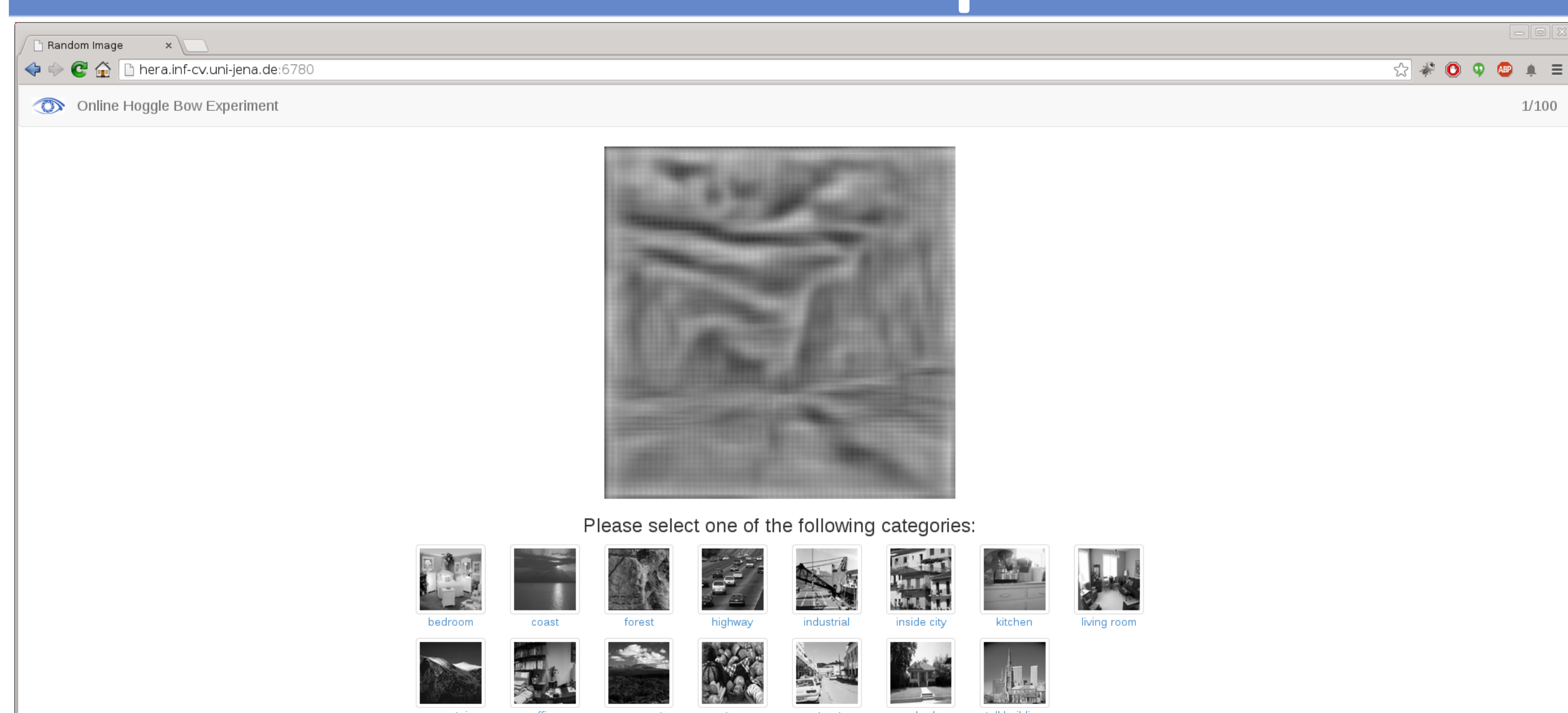
Method details

- Extract local feature descriptors (histograms of oriented gradients, HOG) of the grayscale image from patches on a dense grid
- Cluster descriptors to compute visual codebook, e.g., with k-means
- Quantization of each descriptor with codebooks of different sizes
- Feature inversion technique proposed by Vondrick et al. (ICCV, 2013) to obtain grayscale patches from histogram descriptors (see project page: <http://web.mit.edu/vondrick/ihog/>)
- Replace original image patches by inverted descriptors

Effects of quantization



Website for our human experiment



Conclusions

- Humans perform worse than machine learning approaches when being restricted to visual information present in quantized local features rather than seeing the original input images.
- Early stages of local feature extraction seem to be most crucial for achieving human performance on original images.
- Large codebook sizes in the order of thousands of prototypes are essential not only for good machine learning performance, but more interestingly, also for human image understanding.