Toward Robust Material Recognition for Everyday Objects

Diane Hu
http://cseweb.ucsd.edu/~dhu/
University of California, San Diego
San Diego, CA, USA

Liefeng Bo
http://www.cs.washington.edu/homes/lfb/
University of Washington
Seattle, WA, USA

Xiaofeng Ren
http://www.cs.washington.edu/homes/xren/
Pervasive Computing Center, Intel Labs
Seattle, WA, USA

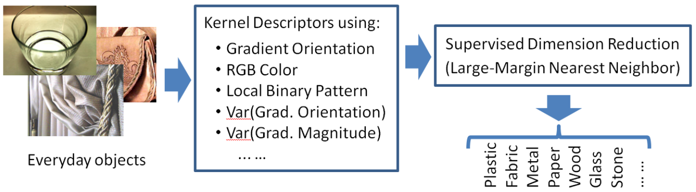


Figure 1: We study material recognition in real-world images. We take a discriminative approach by utilizing a rich set of material-motivated features based on kernel descriptors [1], dimension-reduced with large margin distance learning. We show extensive evaluations on the Flickr dataset [2] as well as new datasets using ImageNet and macro photos.

Everyday Material Recognition. Material recognition is a challenging problem and is fundamental to visual perception. Recent research pushes material recognition from lab settings (such as CuRET) into the real-world. The MIT dataset [2] selects Flickr photos as samples for common materials, demonstrating the difficulties of material recognition. We study material recognition for everyday objects utilizing a rich set of local features under a single framework of Kernel Descriptors [1]. We evaluate both standard features (shape and color) as well as material-motivated features (variances of gradient orientation and magnitude). We use large-margin distance learning [5] to reduce descriptor dimensions by a factor of 30. We provide insights into questions such as "how hard is real-world material recognition?", "what are the best features for material?", and "how does material recognition relate to object recognition?".

Kernel Descriptors and Extensions. Kernel descriptors [1] are a family of patch-level features that are recently introduced for visual recognition. The basic idea of kernel descriptors is that the similarity between patches can be formulated as a match kernel, and highly non-linear match kernels can be well approximated through kernel PCA, leading to kernel descriptors over local image patches. More specifically, the shape kernel descriptor from [1] is based on the local binary pattern:

$$K_s(P,Q) = \sum_{z\in P}\sum_{z'\in Q} \tilde{s}_z \tilde{s}_{z'} k_b(b_z, b_{z'}) k_p(z,z')$$

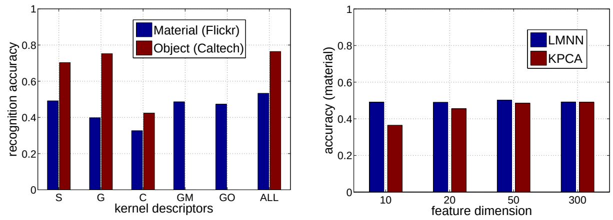
where P and Q are two patches, z denotes the 2D position of a pixel in the patch , $\tilde{s}_z$ is the standard deviation of pixel values (3×3 local window), and b_z is the local binary pattern [4] around z. $k_b(b_z,b_{z'}) = \exp(-\gamma_b \|b_z - b_{z'}\|^2)$ and $k_p = \exp(-\gamma_p\|z-z'\|^2)$ are gaussian kernels.

Motivated by material characteristics, we introduce two new kernel descriptors using variances of gradient orientation and magnitude. Let σ_z^o, $\sigma_{z'}^o$ be the standard deviation of gradient orientation around z and z', and σ_z^m and $\sigma_{z'}^m$ the standard deviations of gradient magnitude. We define

$$\begin{aligned} K_{GO}(\cdot,\cdot) &= \sum_z\sum_{z'} \tilde{s}_z\tilde{s}_{z'} k_{go}(\sigma_z^o, \sigma_{z'}^o) k_b(b_z,b_{z'}) k_p(z,z') \\ K_{GM}(\cdot,\cdot) &= \sum_z\sum_{z'} \tilde{s}_z\tilde{s}_{z'} k_{gm}(\sigma_z^m, \sigma_{z'}^m) k_b(b_z,b_z) k_p(z,z') \end{aligned}$$

where k_{go} and k_{gm} are Gaussian kernels that measure the similarity of the variance of gradient orientation and gradient magnitude, respectively.

Material Recognition Results on Flickr. We evaluate the five kernel descriptors on the Flickr dataset (Figure 2). The two new kernel descriptors, capturing local variations, work surprisingly well on material recognition. Combining descriptors, We improve the state of the art from from 45% to **54%**, a large step toward robust material recognition for everyday objects. For comparison, we also show object recognition results on Caltech 101. The results are largely similar, with interesting differences such as the relative performance of shape vs gradient (texture) descriptors.



(a) (b)

Figure 2: Results on MIT/Flickr [2]. (a) Material recognition accuracies of five kernel descriptors with LMNN=10; with object recognition accuracies on Caltech 101 from [1]. (b) Accuracy vs dimension in large margin nearest neighbor for the shape descriptor, compared to kPCA.

Material vs Object Recognition. To further understand real-world material recognition and its connections to object recognition, we collect two new datasets, one using ImageNet and the other using macro photos. For the ImageNet-Material7 dataset, we choose 7 common material categories and, for each material, choose 10 object categories that are commonly associated with that material. The dataset includes 100 images for each object category, 1000 total for each material category, i.e. 10 times the size of the Flickr dataset. For the macro-lowres dataset, we use 30 physical objects commonly found in everyday life and take two types of images: one type is low-res, low-quality, and taken with a cheap webcam at VGA resolution. The other type is DSLR photos (24M pixels) with a high quality macro lens that provides 1:1 magnification.

In the paper, we show many experimental results on exploring the relations between material and object recognition, such as: (1) evaluating the same set of descriptors on both tasks; (2) comparing material recognition accuracies when the system is trained on same or different objects; (3) feeding object predictions into material recognition and vice versa; (4) comparing recognition accuracies using macro vs lowres images.

(a) (b)

Figure 3: (a) Examples from the ImageNet-Material7 dataset with joint material and object labels, from the *fabric* and *plastic* material categories. (b) An example from the macro-lowres dataset. Shown is a macro photo and a lowres webcam photo of a physical object (towel).

[1] L. Bo, X. Ren, and D. Fox. Kernel descriptors for visual recognition. In *NIPS*, 2010.

[2] Ce Liu, Lavanya Sharan, Edward H. Adelson, and Ruth Rosenholtz. Exploring features in a bayesian framework for material recognition. In *CVPR*, 2010.

[3] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. ImageNet: A Large-Scale Hierarchical Image Database. In *CVPR*, 2009.

[4] T. Ojala, M. Pietikainen, and T. Maenpaa. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *PAMI*, 2002.

[5] K.Q. Weinberger and L.K. Saul. Distance metric learning for large margin nearest neighbor classification. *JMLR*, 2009.