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

Abstract

Material recognition is a fundamental problem in perception that is receiving increasing attention. Following the recent work using Flickr [16, 23], we empirically study material recognition of real-world objects using a rich set of local features. We use the Kernel Descriptor framework [5] and extend the set of descriptors to include material-motivated attributes using variances of gradient orientation and magnitude. Large-Margin Nearest Neighbor learning is used for a 30-fold dimension reduction. We improve the state-of-the-art accuracy on the Flickr dataset [16] from 45% to **54**%. We also introduce two new datasets using ImageNet and macro photos, extensively evaluating our set of features and showing promising connections between material and object recognition.

1 Introduction

Perceiving and recognizing material properties of surfaces and objects is a fundamental aspect of visual perception. Understanding materials enables us to interact with a world full of novel objects and scenes. A robust material recognition solution will find a wide range of uses such as in context awareness and robotic manipulation.

Visual perception of materials is traditionally studied as texture classification: close-up views of material samples are carefully captured in a lab with varying viewing angles and illuminations, such as in the CUReT dataset [9]. Near-perfect accuracy has been reported (e.g. in [25]). However, studies have also shown that material recognition in the real world is much more challenging and is far from solved [8].

Most recently, researchers have been trying to push material recognition into the real-world. The Flickr dataset from MIT [23] selects photos from Flickr as samples for common material categories, demonstrating the difficulties of material recognition and coming much closer to real-world conditions and applications. State-of-the-art material recognition [16] combines a large number of heterogeneous features in a Bayesian framework. However, there have been virtually no comparative studies. Efforts are clearly needed to explore questions such as "how hard is real-world material recognition?", "what are the best features for material recognition?", and "how does material recognition connect to object recognition?".

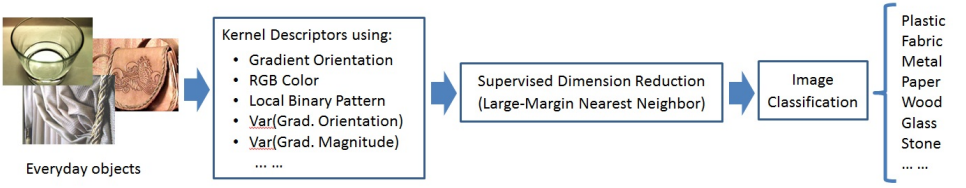


Figure 1: We study material recognition in real-world images. We take a discriminative approach by utilizing a rich set of material-motivated features based on kernel descriptors [5], dimension-reduced with large margin distance learning. We show extensive evaluations on the Flickr dataset [16] as well as new datasets using ImageNet and macro photos.

In this work we seek empirical answers to these questions by exploring state-of-the-art visual features and discriminative learning techniques. Figure 1 shows an overview of our approach. We utilize a rich set of local features under a single framework of Kernel Descriptors [5]. We study and evaluate both standard features (shape and color) as well as material-motivated features (variances of gradient orientation and magnitude) that aim at capturing different aspects of material properties. We use large-margin distance learning [26] to reduce descriptor dimensions by a factor of 30, and use efficient match kernels [4] to compute image-level features for SVM classification.

We show extensive results and comparisons on the MIT Flickr dataset. Our approach greatly improves the state of the art on real-world material recognition, pushing the accuracy on the Flickr dataset from 45% to **54**%. Moreover, We evaluate how the five kernel descriptors, individually or combined, perform on the material recognition task, and compare that to the object recognition task.

We also evaluate our approach on two new datasets: a much larger dataset of 7000 images collected from ImageNet, covering 7 material and 67 object categories, as well as a macro dataset photographing physical objects with both a low-res webcam and a high-res macro lens. These datasets allow us to evaluate, in greater depth, real-world material recognition for everyday objects. In addition, we show interesting connections between material recognition and object recognition that lead to promising future directions.

2 Related Work

Material perception is a fundamental problem in vision [1] that has been studied from a number of perspectives. One line of approach seeks to model detailed underlying reflectance properties of materials using the Bidirectional Reflectance Distribution Function (BRDF) and Bidirectional Texture Function (BTF) [9, 21].

Material recognition is typically viewed as a texture classification problem. The work on 3D textons [15] directly addresses material recognition using multiple aligned images of varying viewpoint and lighting conditions. In recent years, the CUReT dataset [9] has become a standard. CUReT and similar benchmarks have motivated a lot of progress on texture research, including joint modeling of filter responses [24] and non-parametric patch-based models [25], with results approaching 100%. On the other hand, there are a lot of challenges not revealed in these benchmarks, such as scale variation and intra-class variation [7, 8].

A more interesting development in material perception recognition is to move from lab

conditions into the real world. Sharan et al [23] presented the Flickr material dataset which used Flickr photos, captured under unknown real-world conditions, to cover common material categories found in daily life. The work of Liu et al [16] presented extensive experiments and analysis on this new challenging task. Our work is directly motivated by these efforts to bring material recognition closer to real-world scenarios.

Material recognition and object recognition are often perceived as closely related problems and their solutions share many things in common [3]. The work of Zhang et al [28] is a good example of how state-of-the-art visual features and classification techniques are applied to both object and material recognition. There are a few interesting examples in which material and object recognition interact, such as separating material from shape and illumination [18] and using reflectance features for object recognition [19].

In image classification, extracting visual features is crucial. SIFT features [17] have been the most popular and are used in almost all modern visual recognition systems. Recently, Bo et al [5] proposed a set of *kernel descriptors* which uses kernel approximations to outperform SIFT on many benchmarks. We heavily use the kernel descriptor framework in our system.

One separate line of interest is the learning of distance metrics and supervised dimensionality reduction. There has been a lot of work on distance metric learning (LMNN) [2, 27] and dimensionality reduction [22] and its applications in vision [12]. We use large margin nearest neighbor learning [26] – which has been successful in many problem settings – to dramatically reduce the dimensionality and computational costs of our kernel descriptors.

3 Material Recognition with Kernel Descriptors

In this work, we want to take one step further toward real-world material perception by considering material recognition to be a part of the bigger picture of visual recognition. In particular, we want to study material recognition in the framework of state-of-the-art object recognition systems, both in terms of processing pipelines and techniques.

The most popular and successful local descriptors are orientation histograms including SIFT and HOG, which are robust to minor transformations of images. Recent work on kernel descriptors [5] show that orientation histograms are equivalent to a certain type of match kernel over image patches. Based on this novel view, a family of kernel descriptors are proposed, which are able to turn pixel attributes into patch-level features. It has been shown that kernel descriptors outperform SIFT on many recognition tasks. In this work we use and extend a family of visual features, the *kernel descriptors* [5], as the basis for visual recognition. We evaluate the three kernel descriptors in [5] in the context of material recognition and also introduce two new kernel descriptors, based on the variance of gradient magnitude and orientation, to capture local variations found in texture patterns.

3.1 Kernel Descriptors for Material Recognition

Kernel descriptors 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.

A standard set of three descriptors, based on three pixel-level features, have been introduced and evaluated for object recognition (see [5] for details):

1. **Shape**: using local binary patterns, the shape kernel descriptor captures brightness variation patterns in a 3x3 neighborhood. Local binary patterns are then combined with a position kernel and a gradient-based weighting scheme to produce a patch-level shape descriptor that is very effective for shape-based recognition. We find that the shape descriptor is particularly useful for material recognition, possibly because of the high-frequency variations in textures.

2. **Gradient**: the gradient kernel descriptor uses gradient magnitude and orientation as the underlying similarity measure. It is again integrated with a position kernel to produce patch-level features. Somewhat surprisingly, we find that the gradient kernel descriptor is a lot less effective for material recognition than for object recognition, possibly because shape per se is less important in the material case.

3. **Color**: the color kernel descriptor uses RGB color values as the underlying pixel-level feature. The color descriptor is very useful for materials such as foliage or water. It is also superior to a color histogram representation, as it behaves much more like bag-of-words where it implicitly discovers recognition-related structures in the color space.

Motivated by visual characteristics of different materials, we here introduce two novel kernel descriptors that aim at capturing local variations:

1. **Variance of gradient orientation.** We also use the standard deviation of the gradient orientation. The intuition behind the variance-of-gradient-orientation descriptor is to capture the difference between a sharp corner versus a soft corner. For instance, metal is often made to have sharp corners because of molding properties. On the other hand, plastic objects are a lot more likely to have round corners.

2. **Variance of gradient magnitidue.** We use the standard deviation of the gradient magnitude in a 3x3 neighborhood as the underlying pixel-level feature. The motivation behind this descriptor is that we try to differentiate between a hard edge (e.g. occlusion boundary) versus a soft edge (e.g. smooth illumination change). A soft edge will have more smooth transitions in the gradient across a neigborhood. Intuitively, metal tends to have hard edges and glass, with transparency, would tend to have smooth edges.

More specifically, the shape kernel descriptor is based on the local binary pattern [20]:

$$K_s(P,Q) = \sum_{z\in P}\sum_{z'\in Q} \tilde{s}_z \tilde{s}_{z'} k_b(b_z, b_{z'}) k_p(z,z') \tag{1}$$

where P and Q are patches from two different images, z denotes the 2D position of a pixel in an image patch (normalized to $[0,1]$), $\tilde{s}_z = s_z/\sqrt{\sum_{z\in P} s_z^2 + \varepsilon_s}$, s_z is the standard deviation of pixel values in the 3×3 local window around z, ε_s a small constant, and b_z is a binary column vector that binarizes the pixel value differences in the 3×3 local window around z. $k_b(b_z, b_{z'}) = \exp(-\gamma_b \|b_z - b_{z'}\|^2)$ is a Gaussian kernel that measure the similarity between the census transform around each pixel z and z'; and finally, $k_p = \exp(-\gamma_p \|z - z'\|^2)$ is a gaussian kernel over two-dimensional pixel coordinates. Together, $K_s(P,Q)$ measures similarity of shape between two patches.

Following the kernel descriptor framework, we extended the shape kernel descriptor by adding a linear kernel over the variance of the pixel gradient orientation values; a second

descriptor was constructed by adding a linear kernel over the variance of the pixel gradient magnitude values.

Let σ_z^o be the standard deviation of gradient orientation around z (using angles), and $\sigma_{z'}^o$ the standard deviation of orientation at z'. Let σ_z^m and $\sigma_{z'}^m$ be the standard deviations of gradient magnitude. The gradient orientation kernel K_{GO}, and gradient magnitude kernel, K_{GM}, are as follows:

$$K_{GO}(\cdot,\cdot) = \sum_z \sum_{z'} \tilde{s}_z \tilde{s}_{z'} k_{go}(\sigma_z^o, \sigma_{z'}^o) k_b(b_z, b_{z'}) k_p(z, z') \tag{2}$$

$$K_{GM}(\cdot,\cdot) = \sum_z \sum_{z'} \tilde{s}_z \tilde{s}_{z'} k_{gm}(\sigma_z^m, \sigma_{z'}^m) k_b(b_z, b_z) k_p(z, z') \tag{3}$$

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

Match kernels are computationally expensive when image patches are large. Kernel descriptors provide a way to extract compact low-dimensional features from match kernels by projecting infinite dimensional features to a low dimensional space [5]. The gradient orientation kernel descriptor has the following form:

$$F_{GO}^t(P) = \sum_{n=1}^{d_{go}} \sum_{i=1}^{d_b} \sum_{j=1}^{d_p} \alpha_{nij}^t \left\{ \sum_{z \in P} \tilde{s}_z k_{go}(\sigma_z^o, u_n) k_b(b_z, v_i) k_p(z, w_j) \right\} \tag{4}$$

where $\{u\}_{n=1}^{d_{go}}$, $\{v\}_{i=1}^{d_b}$ and $\{w\}_{j=1}^{d_p}$ are uniformly sampled from their support region, d_{go}, d_b and d_p are the sizes of basis vectors for k_{go}, k_b and k_p, and α_{nij}^t are projection coefficients computed by applying KPCA to the joint basis vector set: $\{\phi_{go}(u_1) \otimes \phi_b(v_1) \otimes \phi_p(w_1), \cdots, \phi_{go}(u_{d_{go}}) \otimes \phi_b(v_{d_b}) \otimes \phi_p(w_{d_p})\}$ ($\otimes$ is Kronecker product), with t going through the dimensions of the descriptor.

Shape, gradient, color, gradient orientation, gradient magnitude kernel descriptors are strong in their own right and complement one another. Their combination turns out to be always (much) better than the best individual feature. In section 4 we will show how two novel kernel descriptors improve material recognition.

3.2 Supervised Dimensionality Reduction

Because kernel descriptors use unsupervised dimensionality reduction and have to support all possible tasks after this step, the resulting dimensionality needs to be pretty high (for instance, 300 dimensions) in order for accuracy to also remain high. Here, we use supervised dimensionality reduction to reduce the dimensionality by more than a factor of 30 (from 300) without sacrificing recognition accuracy. In fact, in many cases we also see recognition accuracy improve. To achieve this, we choose to use Large Margin Nearest Neighbor (LMNN) [26], as it has shown to be successful in many computer vision applications. LMNN is best known as a method of distance metric learning for nearest neighbor classification. Given a set of training examples $\{(x_i, y_i\}_{i=1}^N$, LMNN optimally learns a Mahalanobis distance metric $\mathbf{L}^\top \mathbf{L}$ such that "neighboring" examples with the same label are pushed closer

together, and "imposter" examples with dissimilar labels are pulled further apart

$$\begin{aligned} &\sum_{ij}^{N} \eta_{ij} \|\mathbf{L}(x_i - x_j)\|^2 + C \sum_{ij}^{N} \eta_{ij}(1 - y_{il})\xi_{ijl} \\ &s.t. \|\mathbf{L}(x_i - x_l)\|^2 - \|\mathbf{L}(x_i - x_j)\| \leq 1 - \xi_{ijl} \\ &\xi_{ijl} \geq 0 \\ &i, j, l = 1, \cdots, N \end{aligned} \tag{5}$$

where η_{ij} is 1 if the input x_j is a target neighbor of input x_i and 0 otherwise.

Nearest neighbor classification then defines the distance between any two examples x_i and x_j using the learned Mahalanobic distance metric $\mathbf{L}$:

$$d_{\mathbf{L}}(x_i, x_j) = \|\mathbf{L}(x_i - x_j)\|^2 \tag{6}$$

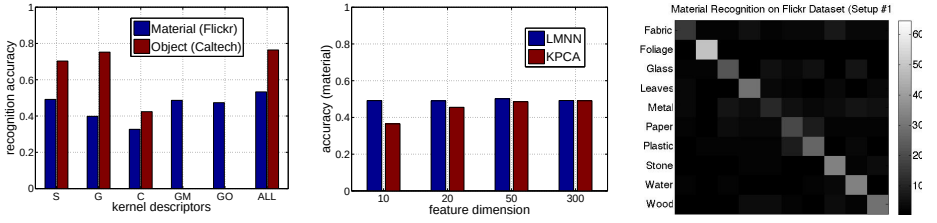
The matrix $\mathbf{L}$ is used simply to transform the space in a way that increases nearest neighbor classification performance. However, initializing $\mathbf{L}$ to be rectangular essentially forces the resulting transformed features into a lower dimensional space. Because the learning of the distance metric makes use of training labels, LMNN is regarded as a successful method for supervised dimensionality reduction.

In our experiments, we obtain a feature vector for each patch in an image using the kernel descriptors described above. To apply LMNN, we represent each image as single feature vector formed from the average of all of its patches. We choose this averaging scheme because raw image patches seemed too diverse and devoid of contextual information; when given as input to LMNN, computation was slow (given the massive number of total patches), and proved to have too many constraints for LMNN to learn a useful transformation. We then use these per-image feature vectors as input into LMNN to obtain a rectangular distance metric that should push similar images (with the same class label) closer together, and dissimilar images (with different labels) further apart. We show in our experimental results that the combination of applying LMNN to our kernel descriptors maintains very high performance, even at extremely low dimensions. For the classification model on top of these features, we have experimented with a number of them including spatial pyramid matching [13, 14], naive Bayes nearest neighbor [6], and the efficient match kernel (EMK) [4]. For efficiency and performance reasons we choose EMK in our experiments.

4 Experimental Evaluations

We use three datasets in our experiments: the MIT Flickr dataset [23]; and two new datasets we have collected, one using ImageNet and one using macro images. We evaluate all five kernel descriptors on the three datasets. We use large margin distance learning (LMNN) [26] to reduce dimensions from 300 to 10. For the Flickr/MIT dataset, we have observed large variations (up to 2%) due to the small size of the data, and we report average accuracies over 5 trials. For the ImageNet and Lowres-Macro datasets, we report accuracies over 5 trials as well, by randomly splitting training and testing data.

The Flickr/MIT Dataset. The Flickr dataset is collected at MIT [23] by Sharan et al using Flickr photos. This includes 1000 images in 10 common material categories, ranging from fabric, paper, and plastic, to wood, stone, and metal. Liu et al [16] reported state-of-the-art results by exploring a large set of heterogeneous features and latent Dirichlet allocation. In all our experiments on Flickr, we do five trials and report the average accuracies.



(a) (b) (c)

Figure 2: Results on MIT/Flickr. (a) Material recognition accuracies of five kernel descriptors with LMNN=10; with object recognition accuracies on Caltech 101 from [5]. (b) Accuracy vs dimension in LMNN for the best performing descriptor (shape), compared to kPCA. (c) Confusion matrix for material recognition using combined features.

Figure 2 (a) shows our results on the Flickr dataset of five kernel descriptors and their combination. The shape descriptor does very well on material recognition, and so does the two novel variance kernel descriptors we introduce. They reach 49% accuracy by themselves, higher than the best accuracy reported in [16] (45%), and much higher than any of the single features tested there (35% for SIFT). By combining 5 descriptors, the accuracy is **54**%, a significant improvement over [16] toward practical material recognition.

In Figure 2(a) we also include the results from [5] on Caltech 101 [11] on the three standard kernel descriptors. Interestingly, the gradient descriptor does about the same on Caltech comparing to the shape descriptor, but much worse on the material task. This may be due to its inability to capture high-frequency variations or the lack of information in large-scale shapes (e.g. long, smooth contours). Relatively speaking, color does better on material than on object recognition. On both Flickr and Caltech, combinations of features only offer marginal improvements over the best single feature.

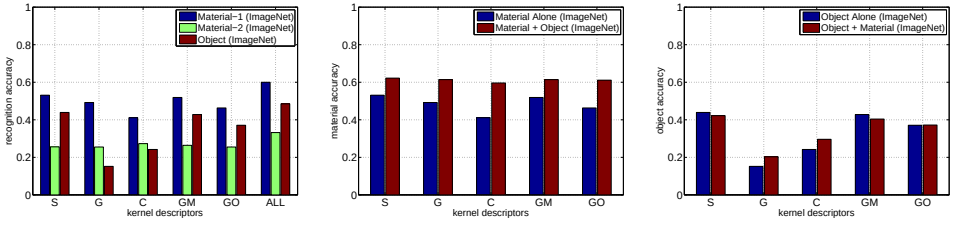
Figure 2(b) shows the evaluation of the supervised dimensionality reduction using large margin nearest neighbor (LMNN), on the best performing descriptor (shape). LMNN performs very well, as there is no drop in accuracy going from 300 dimensions to 10. In comparison, kernel PCA is much less effective, losing 12% accuracy when reducing to 10 dimensions. This reduction greatly improves computational efficiency of kernel descriptors.

Figure 2(c) shows the confusion matrix of the combined classification. Foliage is among the easiest to classify, no doubt partly due to its unique color. Metal, on the other hand, is the hardest, consistent with the findings from [16]. It is followed by fabric and paper, which, interestingly, differs from that in [16]. In particular, we do fairly well on wood, possibly due to the combination of color and texture cues used.

ImageNet-Material7 Dataset. We use ImageNet [10], the vast publicly available image database, to collect a much larger material dataset that focuses on everyday objects. We chose 7 common material categories, a subset of those in the Flickr dataset (water, leather and stone are uncommon in household settings and were left out). For each material, we picked 10 object categories that are commonly associated with that material. [1]. For instance, the material *plastic*, contained object categories including bottles, keyboards, and (plastic) cans. Figure 3 shows several examples from this dataset. One key difference between this dataset and the Flickr dataset is that we do not manually filter out photos that do not "fit" into a material recognition experiment. 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.

[1] We only choose object categories that have a unique material type and leave to future work cases that are not.

Figure 3: Examples from our ImageNet-Material7 dataset. (Left) Examples from the *fabric* material category, which consists of 10 object categories such as curtains, bags, and rugs. (Right) Examples of the *plastic* category, covering 10 object categories such as bottles, keyboards and plastic cans. The dataset covers 7 material categories, 10 objects per material, about 7000 images total. This dataset has both object and material labels for images. It covers a wide variety of objects and scenes, more challenging than the Flickr dataset.



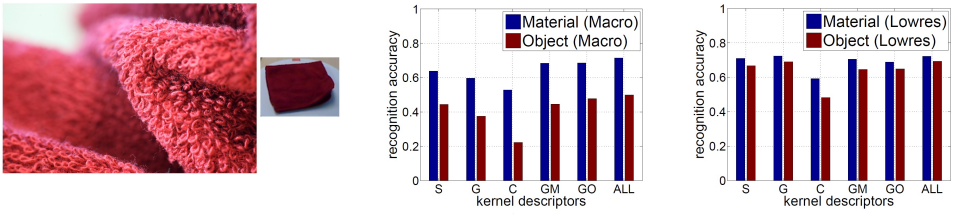
(a) (b) (c)

Figure 4: Results on ImageNet-Material7. (a) Accuracy of material recognition, comparing to object recognition. Two setups for material recognition: **material 1** uses all the objects in both training and testing (different images), and **material 2** uses half of the object categories (5 per material) in training and half in testing. (b) Does object recognition help material recognition? Using outputs from an object recognizer, accuracy on material recognition (setup 1) is significantly increased. (c) Does material recognition help object recognition? Assuming that much of the high accuracy on material recognition in Setup 1 is due to object information (such as shape), we expect that it does not help much – and it is indeed the case.

We set up two experiments: **material 1**, which uses the same set of objects in training and testing for material recognition. That is, for each object category with 100 images, we randomly choose 50 for training and 50 for testing. The training knows what objects to expect in testing, so we can imagine that a lot of object information, for example boundary shape, would be included in the models. We also set up **material 2**, which does not allow object categories to overlap in training and testing. This way, we minimize the impact of object knowledge, such as shape, on material recognition.

Figure 4 (a) shows the accuracies for the two experiments on material recognition, as well as results on the object recognition task. As we expect, **material 2**, deprived of object information, has much lower accuracy than **material 1**. The two novel variance kernel descriptors again perform very well. Interestingly, color does relatively well on **material 2**, outperforming shape and gradient descriptors. Combining the five descriptors provides a much larger improvement compared to what we achieve on the Flickr dataset, possibly because the dataset is larger and leaves more room for learning interactions between descriptors.

Because the ImageNet material dataset has both material and object labels, we can explore interesting questions such as how material and object recognition tasks interact with each other. As a first step, we do the following experiment under **material 1**: to evaluate



(a) (b) (c)

Figure 5: Results on lowres+macro dataset. (a) An example of the macro and lowres images (of a towel). (b) Material and object recognition accuracy on the macro images. (c) Material and object recognition accuracy on the lowres images.

how much object recognition helps material recognition, we first run the object recognizer (using the same setup with kernel descriptors). Then we use the (soft SVM) outputs from the object recognizer and add them as input to the material recognizer. Similarly we can run the material recognizer first, and then feed the outputs into the object recognizer.

The results are very encouraging, as shown in Figure 4 (b) and (c). For each of the five kernel descriptors, by incorporating object recognition outputs, we improve the material recognition accuracy by about 10%. This may be due to the fact that, instead of training all the objects (of a certain material) altogether, the system knows how to split the big set into consistent subsets in object recognition training. It thus has a better chance to discover distinctive features. On the other hand, material information does not help object recognition much, probably because in this dataset object information is a lot more specific than material, and there is also sufficient training data.

Lowres-Macro Dataset. We used 30 physical objects commonly found in everyday life and took two types of images: one type is low-res, low-quality, and taken with a cheap webcam at 640x480 resolution. The other type is 24 M pixels (reduced to 6M for easier processing), taken by a DSLR with a high quality macro lens that provides 1:1 magnification. Five photos were taken for each object at different viewing angles. We compare material recognition using both types of images.

Figure 5 (a) is an example of the macro versus lowres images in the dataset, where close-up macro images are very distinctive from lowres object images. Recognition results are shown in Figure (b) and (c). Overall, the results look similar, and the two new variance-based kernel descriptors do well. There are several interesting things worth noting: (1) object recognition is generally much better with the lowres images, showing the benefits of whole-object shape information; (2) the two variance-based descriptors seem to capture texture details better than the gradient descriptor, doing better on macro; (3) color does better on lowres images, possibly because of the spatial information encoded in the color kernel descriptor. We believe such a dataset with both macro and lowres images of physical objects can be very useful in exploring the synergies between material and object recognition.

5 Discussion

We have presented empirical studies of material recognition in the setting of photos of real-world objects. We use a rich set of local features under the kernel descriptor framework, combined with large margin nearest neighbor learning. We show that supervised dimensionality reduction works very well, reducing descriptor dimensions from 300 to 10. We report

large improvements on the challenging Flickr material dataset [16], from 45% to **54**%.

We have done extensive evaluations on three material datasets of everyday objects, exploring the use of various kernel descriptors in material recognition and comparing them to the case of object recognition. Shape and color descriptors demonstrate different characteristics for handling material and object recognition. The two novel variance kernel descriptors work very well for both recognition tasks.

Our work makes solid progress toward robust material recognition and its applications under real-world conditions. The interaction between material and object recognition is of particular interest. We have shown that the same set of local features and classification techniques work well for both recognition tasks, and there are interesting interactions when considering the two tasks together. This opens up many promising directions for future work such as to explore multi-task learning.

References

[1] E.H. Adelson. On seeing stuff: the perception of materials by humans and machines. In *Proceedings of the SPIE*, volume 4299, pages 1–12, 2001.

[2] A. Bar-Hillel, T. Hertz, N. Shental, and D. Weinshall. Learning distance functions using equivalence relations. In *ICML*, 2003.

[3] S. Bileschi and L. Wolf. A unified system for object detection, texture recognition, and context analysis based on the standard model feature set. In *BMVC*, 2005.

[4] L. Bo and C. Sminchisescu. Efficient match kernels between sets of features for visual recognition. In *NIPS*, 2009.

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

[6] O. Boiman, E. Shechtman, and M. Irani. In defense of nearest-neighbor based image classification. In *CVPR*, 2008.

[7] B. Caputo, E. Hayman, and P. Mallikarjuna. Class-specific material categorisation. In *ICCV*, volume 2, 2005.

[8] B. Caputo, E. Hayman, M. Fritz, and J.O. Eklundh. Classifying materials in the real world. *Image and Vision Computing*, 28(1):150–163, 2010.

[9] K.J. Dana, B. van Ginneken, S.K. Nayar, and J.A.N.J. Koenderink. Reflectance and Texture of Real-World Surfaces. *ACM Transactions on Graphics*, 18(1):1–34, 1999.

[10] 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.

[11] L. Fei-Fei, R. Fergus, and P. Perona. Learning generative visual models from few training examples: An incremental bayesian approach tested on 101 object categories. *Computer Vision and Image Understanding*, 106(1):59–70, 2007.

[12] A. Frome, Y. Singer, F. Sha, and J. Malik. Learning globally-consistent local distance functions for shape-based image retrieval and classification. In *CVPR*, 2007.

[13] K. Grauman and T. Darrel. The pyramid match kernel: Discriminative classification with sets of image features. In *ICCV*, pages 1458–65, 2005.

[14] S. Lazebnik, C. Schmid, and J. Ponce. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In *CVPR*, pages 2168–78, 2006.

[15] T. Leung and J. Malik. Representing and recognizing the visual appearance of materials using three-dimensional textons. *Int'l. Journal of Computer Vision*, 43(1):29–44, 2001. ISSN 0920-5691.

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

[17] D. Lowe. Distinctive image features from scale-invariant keypoints. *Int'l. Journal of Computer Vision*, 60(2):91–110, 2004.

[18] S.G. Narasimhan, V. Ramesh, and S.K. Nayar. A class of photometric invariants: Separating material from shape and illumination. In *ICCV*, 2003.

[19] S.K. Nayar and R.M. Bolle. Reflectance based object recognition. *Int'l. Journal of Computer Vision*, 17(3):219–240, 1996. ISSN 0920-5691.

[20] T. Ojala, M. Pietikainen, and T. Maenpaa. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 971–987, 2002. ISSN 0162-8828.

[21] S.C. Pont and J.J. Koenderink. Bidirectional texture contrast function. *Int'l. Journal of Computer Vision*, 62(1):17–34, 2005. ISSN 0920-5691.

[22] B. Scholkopf, A. Smola, and K.R. Muller. Kernel principal component analysis. *Artificial Neural Networks*, pages 583–588, 1997.

[23] L. Sharan, R. Rosenholtz, and E. Adelson. Material perception: What can you see in a brief glance? *Journal of Vision*, 9(8):784, 2009. ISSN 1534-7362.

[24] M. Varma and A. Zisserman. A statistical approach to texture classification from single images. *Int'l. Journal of Computer Vision*, 62(1):61–81, 2005. ISSN 0920-5691.

[25] M. Varma and A. Zisserman. A statistical approach to material classification using image patch exemplars. *IEEE Trans. PAMI*, 31(11):2032–47, 2009. ISSN 0162-8828.

[26] K.Q. Weinberger and L.K. Saul. Distance metric learning for large margin nearest neighbor classification. *The Journal of Machine Learning Research*, 10:207–244, 2009. ISSN 1532-4435.

[27] E.P. Xing, A.Y. Ng, M.I. Jordan, and S. Russell. Distance metric learning with application to clustering with side-information. *Advances in neural information processing systems*, pages 521–528, 2003. ISSN 1049-5258.

[28] J. Zhang, S. Lazebnik, and C. Schmid. Local features and kernels for classification of texture and object categories: a comprehensive study. *Int'l. Journal of Computer Vision*, 73, 2007.