Visual Material Traits: Recognizing Per-Pixel Material Context

Gabriel Schwartz
Drexel University
gbs25@drexel.edu

Ko Nishino
Drexel University
kon@drexel.edu

Abstract

Information describing the materials that make up scene constituents provides invaluable context that can lead to a better understanding of images. We would like to obtain such material information at every pixel, in arbitrary images, regardless of the objects involved. In this paper, we introduce visual material traits to achieve this. Material traits, such as "shiny," or "woven," encode the appearance of characteristic material properties. We learn convolution kernels in an unsupervised setting to recognize complex material trait appearances at each pixel. Unlike previous methods, our framework explicitly avoids influence from object-specific information. We may, therefore, accurately recognize material traits regardless of the object exhibiting them. Our results show that material traits are discriminative and can be accurately recognized. We demonstrate the use of material traits in material recognition and image segmentation. To our knowledge, this is the first method to extract and use such per-pixel material information.

Figure 1. Materials like the plastic in these images exhibit a wide range of appearances depending on the object and scene, making extraction of material information without the use of object information challenging. We propose to locally recognize visual material traits, distinct appearances of material properties such as "translucent," to provide contextual cues for challenging vision tasks including material category recognition and segmentation.

1. Introduction

Information regarding what an object is made of, *i.e.*, its material, can provide crucial cues for image understanding. If a robot detects soft dirt or a smooth metal surface ahead, it can adjust its movement in advance. Material can sometimes be the only discerning factor between different objects. For example, material information can enable an object detection method to distinguish between a person and a stone statue despite the similarities in their shapes.

Ideally, we would like to extract information regarding the underlying materials at each pixel, without any prior knowledge of what objects are in the scene. This per-pixel material information may potentially facilitate many image understanding methods including image segmentation, object detection, and recognition. For this, we argue that material information should be extracted without higher-level knowledge of the scene, in particular about the object. Only by disentangling visual cues of materials and objects may we exploit material estimates to aid the recognition of objects and the scene. If our material estimates depend on prior knowledge of the object, they are precluded from use in any object recognition or scene understanding process.

Extracting material information in the form of general material categories, such as fabric or metal, has proven difficult [9, 13, 19]. As shown in Figure 1, the appearance of a single material category exhibits large intra-class variability. Each image contains a sample of plastic material, but the material appearance varies based on the object and scene conditions. Recently, Sharan *et al*. [19] introduced a framework that directly recognizes material categories by extracting features and providing a single prediction for the material of an entire image. In their work, they show that material categories are strongly intertwined with features such as their edge slices and edge ribbons. These features convey object-specific information, such as boundary contours. In fact, when their method is run on globally scrambled images, thus removing object-specific information, accuracy drops from 57.1% to 42.6%. Hu *et al*. [9] also show

Input Image Organic Fuzzy

Figure 2. Example material trait recognition. Non-masked pixels in (b) and (c) correspond to pixels with high probability ($p > 0.5$) of exhibiting the given trait. Note that the recognized material traits appear consistently across regions of related materials.

that explicitly supplying object-specific information (output from object recognition) significantly increases material category recognition accuracy. Our method achieves 49.2% accuracy while explicitly excluding object information.

Instead of looking at materials at the category level, we may view material information as a form of visual attribute. Patterson and Hays [18] include a variety of specific materials, *e.g.* asphalt, in their scene-wide attribute framework. Existing techniques recognize a single set of attributes describing an entire image or region. Even in methods that use local features, the framework makes only a single global prediction [5].

How can we then extract material information at each pixel regardless of the object? Looking at the images in Figure 1, one can see that plastic tends to have properties that are associated with a distinct visual appearance, such as "smooth and translucent." To extract material information, we exploit the fact that each material exhibits a certain set of characteristic properties that are shared across appearances of that material in different objects. These properties can include tactile ones such as "hard," or purely visual ones such as "shiny." We propose to model the local visual appearance of these characteristic material properties as a novel intermediate representation: visual material traits.

Though each material trait has an intuitive meaning, some can be challenging to quantify; for example, what makes a soft material look soft? Instead of focusing on hand-tuning a large set of designed features, our framework learns a set of image features in an unsupervised fashion to best represent material trait appearance. We supplement the unsupervised features with a small set of well-known low-level features to describe the space of material trait appearance more completely. By using a randomized decision forest for supervised material trait recognition, we are able to recognize material traits at every pixel in an image. Figure 2 shows the per-pixel recognition results for two material traits on two images from the dataset of Martin *et al.* [16]. The traits are accurately recognized everywhere, even in the the Koala image, despite the fact that the training data included no Koalas or other animals.

Our results show that visual material traits can be recognized accurately in challenging image datasets, as high as 93.1% with an average accuracy of 78.4%. To express more complex concepts, such as material categories, we may treat the distribution of material traits in a region as an image descriptor. Furthermore, material traits learned from one dataset can be recognized and used to extract material information from an entirely different set. This is in contrast with past methods [9, 19] that train and test on images taken from a single source. These results show that the representation generalizes well. We also demonstrate the use of material traits in mid-level image understanding tasks by augmenting segmentation algorithms with per-pixel material information. Our results show that material traits can provide valuable material information to processes for which it was previously unavailable.

2. Related Work

The recent work of Fleming *et al.* [7] is most closely related to our efforts in material trait recognition. In their experiments, they found that perceptual qualities (material traits) are highly correlated with the material classes of the Flickr materials database [20]. Their study, however, relies on human-provided subjective estimates of the presence of these qualities. We directly recognize material traits.

Visual attributes have been widely used in object and scene recognition, but largely at the image or scene level. Ferrari and Zisserman [6] introduced a generative model for certain pattern and color attributes such as "dots," or "stripes." The attributes described in their model focus on texture and color and do not encode any material properties. The model also requires a segmented image as input; the attributes thus cannot provide per-pixel information to applications, such as segmentation, that require it. Kumar *et al.* [10] propose a face search engine with their attribute-based FaceTracer framework. FaceTracer uses SVM and AdaBoost to recognize attributes within fixed facial regions. Farhadi *et al.* [5] apply attributes to the object recognition task. Their results show an improvement in accuracy over a basic approach using texture features. Recently, Patterson and Hays [18] showed that they can recognize a variety of visual attributes, including general material categories. Lampert *et al.* [11] show that attributes transfer information between disjoint sets of classes. In both cases, these visual attributes are single scene-wide detections and not localized.

Adelson [1] first proposed material categories as a dis-

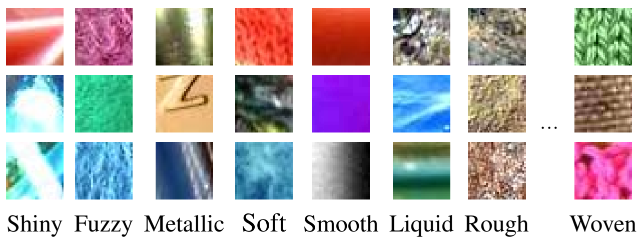


Figure 3. Successfully recognized material traits. These image patches were recognized by our framework as exhibiting the indicated material traits. Even at the patch level, we can see the characteristic visual appearances of each material trait.

tinct concept from textures or objects. Sharan *et al*. [20] introduced a new image database containing images from the photo sharing website Flickr; Liu *et al*. [13] also created a framework to recognize these material categories using a modified LDA probabilistic topic model. Hu *et al*. [9] showed improved performance on the Flickr database using kernel descriptors and large-margin nearest neighbor distance metric learning. Their experiments showed that providing explicit object detection information to material category recognition results in a large improvement in accuracy. Sharan *et al*. [19] show that without information associated with the objects, performance degrades significantly (from 57.1% to 42.6%). Specifically, they note that material category recognition depends heavily on non-local features such as edge contours. In our framework, we explicitly avoid relying on this object information.

Wang *et al*. [22] and Liu *et al*. [14] demonstrate very high material classification accuracy and do so at the per-pixel level. Both methods require a physical measurement apparatus (a dome of lights in fixed positions) to collect material reflectance data. Lombardi and Nishino [15] estimate BRDF parameters from single images but require geometry information. Our goal is to achieve passive extraction of material information from single ordinary images.

3. Representing Material Traits

Figure 3 shows examples of the visual material traits recognized by our framework. Even at the local level of the example images, each visual material trait corresponds to the appearance of a characteristic material property. Ideally, recognition of these material traits will enable us to extract crucial material information from any image.

The key contribution of our material traits is their ability to encode per-pixel material information without relying on object-specific features. Material traits provide a compact, local, and discriminative encoding of material properties. To obtain a representation for these material traits, we must avoid introducing any dependence on object information in the recognition process. We accomplish this by learning the best convolutional features to describe material trait patches in an unsupervised setting. Convolutional features are ideal for this purpose as they can be applied at any point in an image, and do not encode object boundary contours. We supplement these unsupervised features with selected low-level features to describe appearance patterns that cannot be learned by the unsupervised model.

3.1. Convolutional Material Trait Features

Expressing the appearance of material traits poses a challenge. While intuitive, traits such as “fuzzy” can be hard to quantify. While we may attempt to do so using only existing designed features, the space of images that may be represented using these features is incomplete (as shown by our feature selection results).

Rather than rely solely on handcrafted features, we determine features associated with each material trait through unsupervised feature learning. Unsupervised learning builds a generative model for images by finding simple components that can be combined to reproduce them. Constraints, such as sparsity, force optimal model components to also act as discriminative features for classification.

Our goal is to recognize per-pixel, object-independent visual material traits. To this end, we choose to learn convolutional features so that we may extract them at any pixel in an image. By operating in fixed local neighborhoods, convolutional features ensure that we do not encode object boundary contours. These boundary contours are the primary source of undesired object-dependent features in previous frameworks [9, 19].

We build upon the convolutional auto-encoder (CAE) model [17] to learn the feature kernels. The model defines images as the weighted sum of convolution kernel responses. Optimal filters under the model are defined by the following objective function:

$$C = T_r + \alpha T_w + \beta T_s \,. \tag{1}$$

The objective contains three terms: a reconstruction error term T_r, a weight-decay (smoothness) term T_w, and a sparsity term T_s. The weight-decay and sparsity terms have corresponding weights α and β, and each term acts as a constraint to help produce useful features.

Reconstruction error for N images is the squared-difference between the input images $\mathbf{I}$ and their reconstructions $\mathbf{R}$ using the learned features,

$$T_r = \frac{1}{N}\sum_{i=1}^{N} \|\mathbf{I}_i - \mathbf{R}_i\|_2^2 \,. \tag{2}$$

Since the features are convolution kernels, the reconstructed images $\mathbf{R}$ are described in terms of the encoding in feature space $\mathbf{E}_i$ by

$$\mathbf{E}_i = h\left(\mathbf{W} * \mathbf{I}_i + b_e\right), \quad (3)$$

$$\mathbf{R}_i = \mathbf{W}' * \mathbf{E}_i + b_r, \quad (4)$$

$$h\left(x_i\right) = \begin{cases} 0 & \text{if } x < 0 \\ x_i & \text{if } 0 \leq x_i \leq 1 \\ 1 & \text{if } x > 1 \end{cases} \quad (5)$$

with $*$ representing convolution with a set of filters $\mathbf{W}$, along with bias terms b_e and b_r for the encoding and reconstruction, respectively. Some formulations force the reconstruction filters $\mathbf{W}'$ to be the transpose of the encoding filters $\mathbf{W}$. We, however, found that allowing them to be separately optimized resulted in more diverse features.

The non-linear encoding function $h\left(x_i\right)$ in Equation 3 contains a linear region between 0 and 1. If allowed, the combination of small encoding weights and large decoding weights could force any inputs to encode solely into this linear region. Such an encoding would result in a trivially perfect reconstruction. Weight decay, $T_w = \|\mathbf{W}\|_2^2 + \|\mathbf{W}'\|_2^2$, is a term that prevents this trivial solution by ensuring that the weights do not take on exceedingly large values.

By definition, discriminative image features do not appear everywhere in an image. Figure 3 shows that certain material traits, particularly "shiny," exhibit strong local appearance cues. Sparsity constraints express this property well. Sparse features are features that are only present in a small fraction of the possible locations in each image, as measured by their presence in the encoding $\mathbf{E}_i$. As in Lee *et al*. [12], we enforce sparsity by penalizing the difference between mean filter activations and a small constant p:

$$T_s = \left\| p - \frac{1}{N} \sum_{i=1}^{N} \mathbf{E}_i \right\|_2^2 . \quad (6)$$

To further constrain the learning process and obtain a discriminative feature set, we force a fixed number of the features to be oriented first-order Gaussian filters. Learning these filters alone will satisfy both sparsity and reconstruction constraints, but their discriminative power is limited. As shown in Table 1, edge filters are selected roughly half as often as the CAE-learned features.

We optimize the full objective function using L-BFGS with automatically-generated symbolic gradient evaluation.

Figure 4 shows a selection of the top convolution filters by the CAE, ranked by average presence in the corresponding material trait images. The filters were learned from whitened material trait image patches. The top filters appear to represent the presence or absence of specific local texture patterns. For comparison, the non-ranked features on the right exhibit far less texture variation.

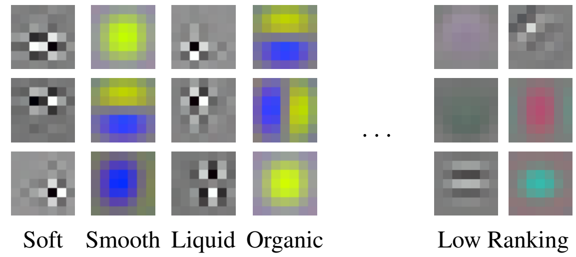


Figure 4. These 7×7px. convolution filters learned by the CAE represent the top three filters for the listed material traits, ranked by average presence in the testing images. The filters represent characteristic local texture and color patterns. The six filters on the right do not rank in the top three for any material trait. They exhibit significantly less texture variation than the top filters.

3.2. Supplemental Features

Cybenko [3] showed that artificial neural networks, including auto-encoders such as the CAE, are capable of approximating any continuous function defined on $\mathbb{R}^n$. There are, however, local features such as HOG that are not continuous and thus cannot be learned by the CAE. These discrete features may encode important properties of material traits, such as the strong local patterns in woven material. To address this, we supplement the learned features with Local Binary Patterns (LBP), HOG features and color histograms. We do not use other low-level features, such as the edge slices and ribbons of Sharan *et al*. [19], as they encode object-specific information and cannot be extracted on a per-pixel basis.

The results of our feature selection process show that these additional features supplement rather than replace the CAE-learned features. As will be shown in Table 1 in the following analysis of feature selection, CAE features are selected on average as often as any of the supplemental features. Furthermore, our analysis in Table 2 shows that the CAE features play a crucial role in the application of material traits.

3.3. Groupwise Feature Selection

We would like to obtain a feature set that generalizes well to new datasets. To avoid overfitting and improve generalization, we perform feature subset selection on the supplemental and CAE-learned features. Our final feature set contains a small number of groups of conceptually related features. Rather than separate the groups into individual elements, we select the best combination of groups to recognize each trait. This process takes advantage of the fact that two individually useless features can have predictive power when grouped together [8]. We are able to exhaustively evaluate all combinations of groups (CAE features, oriented edges, HOG, LBP, color histograms), selecting those that maximize performance on a validation set. Feature groups are not further divided, thus, for example, either all HOG

Trait	CAE	Oriented	HOG	LBP	Color Histograms
Shiny	•				•
Fuzzy		•		•	
Transparent	•	•	•		
· · · (13 Material Traits)					
Total Uses	7	4	6	9	7

Table 1. Selected features for material traits. As "fuzziness" is characterized by fine edge patterns, oriented filters and LBP are useful. Since we define "shiny" only on areas that exhibit specular highlights, it follows that color histograms and learned convolutional filters are important features for this material trait.

features are included or none are.

Table 1 shows the results of our feature selection process[1]. Features are selected fairly evenly and, as the full table shows, in disjoint sets. A particular case of note is the "shiny" material trait. Since we focus on recognizing visual material traits without dependence on object-specific information, "shiny" is synonymous with specular highlights. This may be seen clearly in Figure 3. While there are visual cues, such as contoured reflections on a car body, that may lead an observer to call a material "shiny," these features are specific to the object and do not directly indicate the material trait. As a result of this, color histograms and learned convolutional filters prove to be more useful features for this material trait.

4. Recognizing Material Traits

For training and testing, we annotate images in the Flickr Materials Database (FMD) [20] with masks indicating regions that exhibit each material trait. From these regions, we extract 45,500 annotated patches[2]. We use balanced sets of positive and negative examples to train randomized decision forest (RDF) classifiers for each material trait. Though we use the same dataset as methods that include object information, our feature set and recognition process explicitly avoid object dependence.

Figure 2 shows the recognition results for two material traits on an image from the Berkeley Segmentation Dataset (BSDS) [16]. Note that the main object in the image, a Koala, was not present in the Flickr dataset. The FMD does not, in fact, contain any animals or any examples of animal fur. Despite this, characteristic properties of the fur and plants are accurately recognized.

Figure 5 contains recognition accuracies for each of the 13 material traits. Since we predict material traits independently, and the training and testing data are balanced, random chance performance is 50% accuracy. Most material traits are recognized very accurately, however, some are challenging. "Metallic" and "transparent" have the two

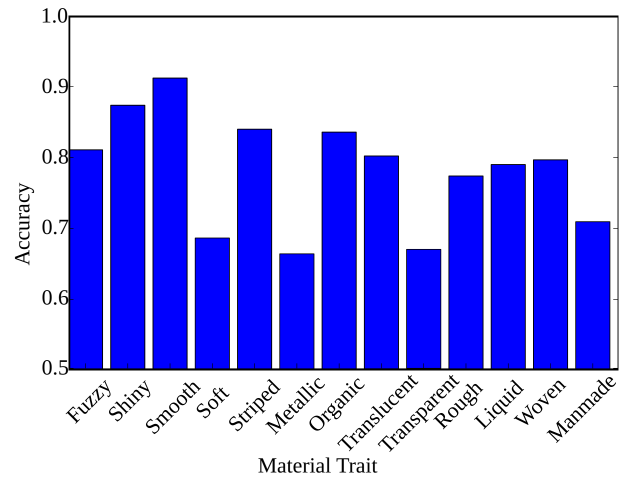


Figure 5. Visual material trait recognition accuracy. Material traits are recognized via binary classification on a balanced training and testing set, thus random chance accuracy is 50%. Most traits are recognized well. Difficult material traits, such as metallic and transparent, are challenging due to their object- and environment-dependent appearances. Average accuracy is 78.4%.

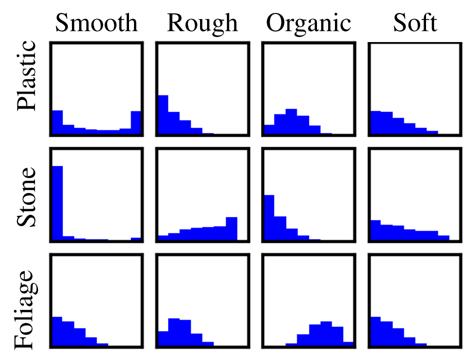


Figure 6. Material trait distributions. We compute the class-conditional distributions for each material trait given each material category. These are stored as histograms, examples of which are shown above. Plastic is most often smooth, while stone is extremely rarely smooth.

lowest recognition rates (66.4% and 67.0%). The appearance of these material properties depends heavily on the environment surrounding the object. In the case of a reflective metal surface or a clear glass sphere, the appearance is determined entirely by the object and its environment. As we explicitly avoid object dependence, we cannot expect to model these particular material traits with the same level of accuracy as others. Despite this, "metallic" and "transparent" are still recognized better than chance.

Material traits, as a form of visual attribute, should represent a discriminative set of appearances. To investigate this, we compute the class-conditional distributions of material traits given material categories. We use the ten categories of the FMD for this test. For each image in each category, we sample material traits uniformly across the masked material region in the image. Figure 6 shows selected distributions

[1] A full list of all material traits and their corresponding features may be found in our supplemental material.

[2] Our implementation uses, but is not restricted to, 32×32px. patches.

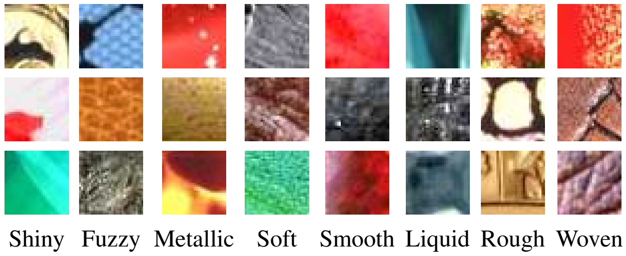


Figure 7. Our framework produced false-positive detections of material traits in these patches. For the challenging metallic trait, it is clear that color plays a strong role. The misclassifications generally have a metallic color even though the material is not metal. In some rare cases such as "smooth" there are missing annotations and thus the false positives are actually true positives.

from the set $\{p\left(t_i|m_j\right)|i \in 1 \ldots 13, j \in 1 \ldots 10\}^3$. The resulting distributions do, in fact, represent the characteristic properties of their respective material categories. Stone is often rough but very rarely smooth (there are a small number of polished stone examples in the training data), plastic is smooth, and foliage is organic. As material traits are purely visual, they can occasionally produce false positives, as seen in $p\left(\text{soft}|\text{stone}\right)$. While stone is not soft, porous stones may have a soft appearance.

Figure 7 shows a set of false positive material trait recognition results. "Shiny," with its characteristic bright highlights, is prone to be recognized in over-exposed image regions. Results for "metallic" show that color is a strong cue for this material trait. Though the patches are metallic in color, the material is not in fact metallic. These are limitations of the representation. There are a few cases where the material trait annotations are incomplete, generally for the pervasive "smooth" material trait.

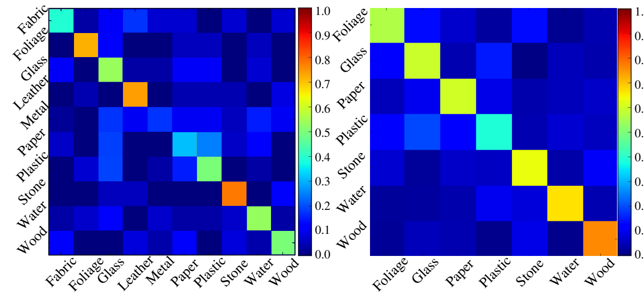
5. Using Visual Material Traits

Our analysis shows that we may accurately recognize material traits. The material trait distributions also show that material traits encode discriminative material information. Each material category exhibits characteristic class-conditional material trait distributions. From these results, we expect to be able to inform higher-level processes with material information from material traits. Material trait distributions allow us to recognize material categories in arbitrary images without dependence on prior object knowledge. We also demonstrate a preliminary application of material traits to the problem of segmentation.

5.1. Material Categories from Visual Material Traits

Sharan *et al.* [19] showed that material category recognition depends on object-specific information. Despite this,

[3]Please see our supplemental material for a full visualization of all class-conditional material trait distributions.



(a) Flickr (b) ImageNet

Figure 8. Confusion matrices showing true class vs. predicted class on the Flickr Material Database and ImageNet images. Average accuracy is 49.2% in (a) and 60.5% in (b). Though metal and glass both have an appearance that is environment-dependent, glass is more accurately classified. This is likely due to the tendency of glass to create characteristic local distortions.

our class-conditional trait distributions suggest that the information encoded in material traits does provide a discriminative set of features for material category recognition. We rely on these visual material trait distributions to encode and recognize material categories.

We recognize material categories via material traits by training a randomized decision forest (RDF) classifier on the material trait distributions. Distributions are computed from material traits recognized in uniformly sampled random patches. We select features and train material trait classifiers using half of the FMD for training, then predict their class-conditional distributions. We further supplement the distributions, in a cascade fashion, with the output of a RDF classifier trained to directly predict the material category of a patch using our feature set. The cascade process is responsible for improvements in the more recognizable categories such as foliage (11% improvement), with minor changes in other categories. Accuracy without the cascade process is 46.5%, only a 2.7% reduction.

Using the computed class-conditional distributions, we train an SVM classifier with a histogram intersection kernel to recognize material categories. The histogram intersection kernel, defined as

$$k\left(\mathbf{x}, \mathbf{y}\right) = \sum_i \min\left(x_i, y_i\right), \tag{7}$$

for histogram feature vectors $\mathbf{x}$ and $\mathbf{y}$ with elements x_i and y_i, measures the similarity between the two normalized histograms [2]. As the material trait distributions are histograms, they are ideally suited for the histogram intersection kernel SVM.

Figure 8 shows the average and per-class accuracy for our method on the FMD. We split the dataset of 1000 images in half for training and testing. Our accuracy (49.2%) does not surpass the final results of Sharan *et al.* (57.1%) but again, their method relies heavily on features that encode the shape of the objects. We do find that our

Wood (Stone) Plastic (Glass) Paper (Foliage)

Figure 9. Three misclassified ImageNet images, with true classes for each prediction is in parentheses. The left two are a result of confusing appearances (striped and translucent are more often associated with wood and plastic respectively) while the rightmost is due to the bounding box poorly fitting the object.

method achieves higher accuracy than that of theirs (42.6%) when object context is removed. These results show that material traits provide important information to the material recognition process.

To demonstrate the ability of material traits to generalize well between datasets, we collected a second set of material images from a different source: ImageNet [4]. ImageNet obtains images from a variety of sources; they are thus more diverse than solely Flickr images. We collected 3480 images from ImageNet via searches for each material category. Images without bounding boxes were discarded.

To evaluate the use of material traits for material recognition on this ImageNet dataset, we first train material trait classifiers on the full set of FMD images. We then split the ImageNet images evenly into training and test sets and compute the distributions of recognized material traits on the training and test sets. We train an SVM classifier with the histogram intersection kernel of Equation 7 using the distribution of material traits on the training set.

Figure 8 shows the average accuracy for our method on this dataset. The average accuracy of 60.5% on ImageNet images shows that material traits encode material information that depends on neither the particular type of object exhibiting a material, nor the scene context in which that material appears. While Hu *et al*. [9] do not provide an exact value, visual inspection of their results indicates an accuracy of roughly 60% as well.

Figure 9 contains three misclassification examples from ImageNet images. The stone in the first image has brown color stripes characteristic of wood. The glass in the second image looks translucent due to condensation, and translucent is a trait associated with plastic more than glass. The final image is a misclassification due to localization. The ImageNet database only provides object bounding boxes, not masks. This box contains mostly smooth regions and light colors, traits representative of paper.

We ran a set of tests, summarized in Table 2, to examine the impact of each major component of the material trait and category recognition process. The first row, accuracy when performing direct category recognition, with all features, without material traits, shows that the trait representation provides crucial information for the material recognition process. By excluding either CAE-learned features or supplemental features (HOG, LBP, Color Histograms) from the trait recognition process, we see that both feature sets are necessary in order to best represent material categories.

FS	Traits	SF	CAE	Accuracy
		•	•	34.2%
•	•	•		43.5%
	•		•	42.5%
•	•	•	•	49.2%

Table 2. Performance breakdown. FS: feature selection, SF: supplemental features, CAE: convolutional auto-encoder features. For the first row we performed direct material category recognition using the concatenation of all feature sets. This shows that the trait representation is indeed providing crucial information.

5.2. Segmenting Images with Visual Material Traits

Segmenting images is a challenging process partially because the concept of a good segmentation is subjective. In the Berkeley Segmentation Dataset (BSDS) benchmark of Martin *et al*. [16], evaluation relies on multiple human segmentations as ground truth, since each one is a potentially correct solution. Visual material traits, with their accurate encoding of characteristic and intuitive material properties, should contribute valuable contextual cues to this process.

As an investigation of the potential for image segmentation via material traits, we augment the Normalized Cuts (NCuts) algorithm of Shi and Malik [21] with material trait information. In their method, they treat image segmentation as a graph partitioning problem and show that the optimal solution can be obtained from the solution to a generalized eigensystem (specifically, the eigenvector $\mathbf{y}_2$ corresponding to the second-smallest eigenvalue):

$$(\mathbf{D} - \mathbf{W})\,\mathbf{y} = \lambda \mathbf{D} \mathbf{y}\,, \tag{8}$$

where $\mathbf{W}$ is a matrix of weights representing pairwise pixel similarities and $\mathbf{D}$ is a diagonal matrix containing the sum of all weights for a given pixel. We add an additional term,

$$\exp\left\{\frac{-\left\|\mathbf{t}_i - \mathbf{t_j}\right\|_2^2}{\sigma_T}\right\}\,, \tag{9}$$

to the similarity score function used to obtain $\mathbf{W}$. $\mathbf{t}_i$ represents the predicted per-trait probabilities for pixel i in the image and σ_T is a scaling parameter. This term should cause pixels that exhibit similar material traits to be grouped together in the segmentation.

Figure 10 shows images segmented using the original NCuts algorithm and our modified version. The first example shows that material traits can help discriminate between regions exhibiting different material properties (fuzzy grass and rocks). The expanded border around the penguin in the second segmentation is likely due to the fact that the

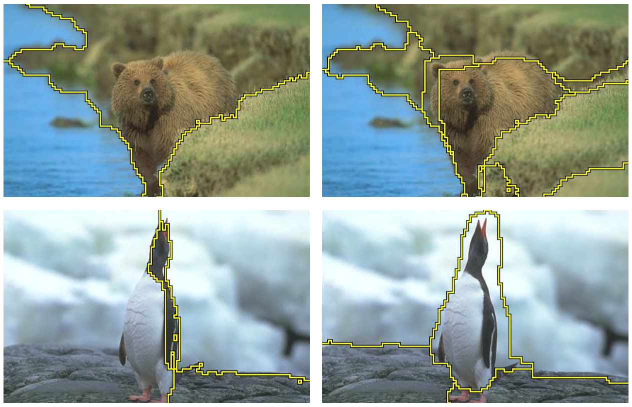


Figure 10. Comparing segmentation with and without material traits. Images on the left were segmented using the original NCuts algorithm, while those on the right were segmented with our modified version. Material traits can indicate the difference between fuzzy grass in the foreground and rocks in the background, despite the fact that they have similar colors.

traits are recognized in part using learned convolution kernels. The size of these kernels is likely to be an important parameter for good segmentations. These results show that contextual cues from material traits can indicate regions of similar materials that should be merged, or regions that should be split despite similar color or texture.

6. Conclusion

We introduce the recognition of visual material traits as a representation of object-independent, per-pixel material information. Our results show that material traits contribute useful information to the material category recognition process. Furthermore, we show that material traits generalize well to new and diverse datasets.

Our preliminary image segmentation results show that the per-pixel recognition of material traits can contribute useful material information to new applications in which it was previously unavailable.

Results from material category recognition, segmentation, and from recognition of the material traits alone, demonstrate that visual material traits form a compact and discriminative representation for crucial object-independent material information. We expect material traits to prove useful in future exploration of image understanding.

References

[1] E. H. Adelson. On Seeing Stuff: The Perception of Materials by Humans and Machines. In *SPIE*, pages 1–12, 2001.

[2] A. Barla, F. Odone, and A. Verri. Histogram intersection kernel for image classification. In *Proceedings of the International Conference on Image Processing*, 2003.

[3] G. Cybenko. Approximation by superpositions of a sigmoidal function. *Mathematics of Control Signals and Systems MCSS*, 2(4):303–314, 1989.

[4] 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.

[5] A. Farhadi, I. Endres, D. Hoiem, and D. Forsyth. Describing Objects by their Attributes. In *CVPR*, pages 1778–1785, 2009.

[6] V. Ferrari and A. Zisserman. Learning Visual Attributes. In *NIPS*, pages 433–440, 2007.

[7] R. W. Fleming, C. Wiebel, and K. Gegenfurtner. Perceptual qualities and material classes. *Journal of Vision*, 13(8):1–20, 2013.

[8] I. Guyon and A. Elisseeff. An Introduction to Variable and Feature Selection. *Journal of Machine Learning Research*, 3(7-8):1157–1182, 2003.

[9] D. Hu, L. Bo, and X. Ren. Toward Robust Material Recognition for Everyday Objects. In *BMVC*, pages 48.1–48.11, 2011.

[10] N. Kumar, P. Belhumeur, and S. Nayar. FaceTracer : A Search Engine for Large Collections of Images with Faces. In *ECCV*, pages 340–353, 2008.

[11] C. Lampert, H. Nickisch, and S. Harmeling. Learning to Detect Unseen Object Classes by Between-Class Attribute Transfer. In *CVPR*, pages 951–958.

[12] H. Lee, C. Ekanadham, and A. Ng. Sparse Deep Belief Net Model for Visual Area V2. *NIPS*, pages 1–8, 2008.

[13] C. Liu, L. Sharan, E. H. Adelson, and R. Rosenholtz. Exploring Features in a Bayesian Framework for Material Recognition. In *CVPR*, pages 239–246, 2010.

[14] C. Liu, G. Yang, and J. Gu. Learning Discriminative Illumination and Filters for Raw Material Classification with Optimal Projections of Bidirectional Texture Functions. In *CVPR*, 2013.

[15] S. Lombardi and K. Nishino. Single Image Multimaterial Estimation. In *CVPR*, pages 238–245, 2012.

[16] D. Martin, C. Fowlkes, D. Tal, and J. Malik. A Database of Human Segmented Natural Images and its Application to Evaluating Segmentation Algorithms and Measuring Ecological Statistics. In *ICCV*, pages 416–423.

[17] J. Masci, U. Meier, D. Cireşan, and J. Schmidhuber. Stacked Convolutional Auto-Encoders for Hierarchical Feature Extraction. In *ICANN*, pages 52–59, 2011.

[18] G. Patterson and J. Hays. SUN Attribute Database: Discovering, Annotating, and Recognizing Scene Attributes. In *CVPR*, 2012.

[19] L. Sharan, C. Liu, R. Rosenholtz, and E. H. Adelson. Recognizing Materials Using Perceptually Inspired Features. *International Journal of Computer Vision*, 2013.

[20] L. Sharan, R. Rosenholtz, and E. Adelson. Material Perception: What Can You See in a Brief Glance? *Journal of Vision*, 9(8):784, 2009.

[21] J. Shi and J. Malik. Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8):888–905, 2000.

[22] O. Wang, P. Gunawardane, S. Scher, and J. Davis. Material Classification Using BRDF Slices. In *CVPR*, pages 2805–2811, 2009.