# Automatically Discovering Local Visual Material Attributes[1]

Gabriel Schwartz, Ko Nishino
Department of Computer Science, Drexel University

Shape cues play an important role in computer vision, but shape is not the only information available in images. Materials, such as fabric and plastic, are discernible in images even when shapes, such as those of an object, are not. We argue that it would be ideal to recognize materials without relying on object cues such as shape. This would allow us to use materials as a context for other vision tasks, such as object recognition.

In this paper, we derive a framework that allows us to discover locally-recognizable material attributes from perceptual material distances. Discovered attributes exhibit the same desirable properties as known semantic material traits [1], despite the fact that they are discovered using only partial supervision. We obtain perceptual distances from a simple crowdsourcing process that can be readily extended to large datasets.

Our attribute discovery process consists of three main components: perceptual distance measurement, attribute space definition, and attribute classifier learning. First, we measure pairwise perceptual distances between materials based on simple binary decisions on image patches collected through Amazon Mechanical Turk (AMT). Given these pairwise perceptual distances, we define a category-attribute space that preserves these distances while exhibiting desired attribute properties. Finally, we train a set of classifiers to reproduce the category-attribute space as predictions on image patches.

Directly measuring perceptual material distances based on human annotations would be an ideal source of data. One could potentially ask annotators to give a measure of the difference between two images, but this would require some form of calibration to compare answers between different annotators. We instead propose to reduce the question to a binary one: "Do these patches look different or not?" The underlying assumption is that if two image patches look similar, they do so as a result of at least one shared visual material property.

Discovering attributes given only a desired set of pairwise distances poses a challenge. A straightforward approach would be to directly train classifiers to predict attributes that encode the distances. This would be a highly under-constrained problem as we do not even know which attributes to associate with which categories. We instead separate the discovery process into two steps. First, we find a category-attribute mapping that encodes the desired distance matrix. We define the mapping as a category-attribute association matrix $\mathbf{A}$. We constrain this mapping such that the resulting attributes are actually recognizable. Given a mapping, we may then define a set of attribute classifiers that reproduces this mapping on image patches.

To discover realizable attributes, we encode our own prior knowledge that recognizable attributes exhibit a particular distribution and sparsity pattern. We observe that semantic attributes, specifically visual material traits, have a Beta-distributed association with material categories. Generally, a material category will either strongly exhibit a trait or it will not exhibit it at all. We would like the values in $\mathbf{A}$ to be Beta-distributed to match the that of known material trait associations. We achieve this by minimizing the KL divergence between the distribution of values in $\mathbf{A}$ (approximated by a kernel density estimate) and the desired Beta distribution.

For attribute classification, we use a general two-layer non-linear model. This affords us the flexibility to place the correct constraints on the classifier learning process. This is essential given the lack of full supervision. For both attribute discovery and classifier learning, we formulate the problems as non-linear optimization over the space of model parameters (category-attribute mapping and classifier weights).

Figure 1 shows an example of three discovered attributes. We visualize the attributes on a sample image by predicting their per-pixel probabilities. We can clearly see that the attributes highlight distinct image regions. Figure 2 demonstrates that we do indeed discover an attribute space that separates material categories. Quantitatively, we verify that our discovered attributes encode material properties using logic regression. Using only 30 attributes and a simple classifier, we are able to recognize known material traits [1] as well as methods that use a much larger feature set. We also show that the discovered attributes can be used to locally recognize materials with accuracy comparable to using fully supervised material traits.

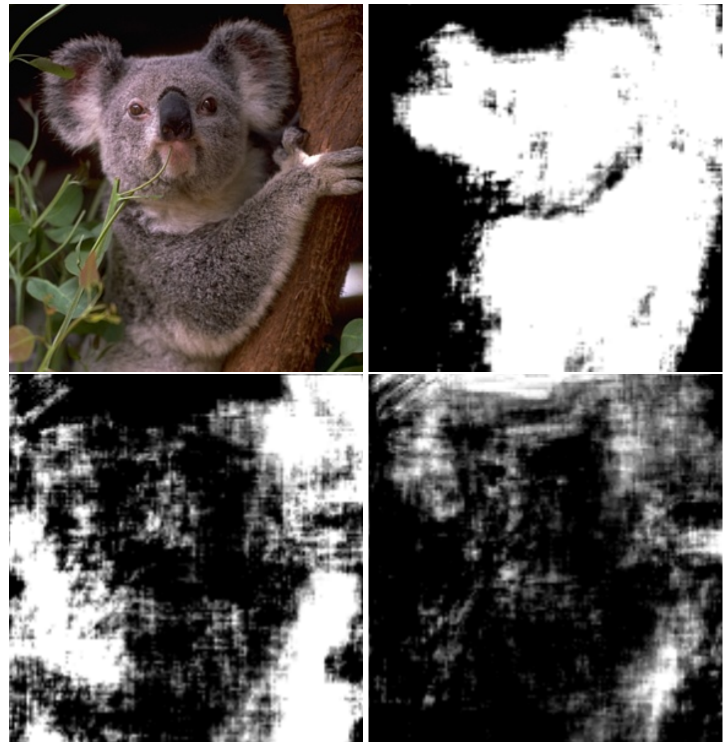

Figure 1: Predicted probabilities for three discovered attributes. Discovered attributes exhibit the desired properties of material traits despite the fact that we only rely on local, weakly-supervised information.

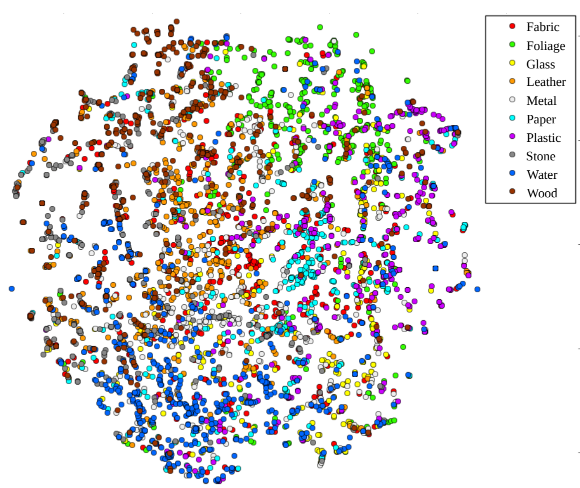


Figure 2: Separation of categories in the discovered attribute space. By using our discovered attributes as the input space for 2D t-SNE, we see that material categories form clusters in the attribute space.

[1] Supported by ONR N00014-14-1-0316 and NSF IIS-0964420, 1353235 and 1421094.

This is an extended abstract. The full paper is available at the Computer Vision Foundation webpage.

[1] Gabriel Schwartz and Ko Nishino. Visual Material Traits: Recognizing Per-Pixel Material Context. In *Color and Photometry in Computer Vision (Workshop held in conjunction with ICCV'13)*, 2013.