Material-Specific Chromaticity Priors

Jeroen Put
jeroen.put@uhasselt.be

Nick Michiels
nick.michiels@uhasselt.be

Philippe Bekaert
philippe.bekaert@uhasselt.be

Hasselt University - tUL - iMinds
Expertise Centre for Digital Media
Wetenschapspark 2
3590 Diepenbeek, Belgium

Abstract

Recent advances in machine learning have enabled the recognition of high-level categories of materials with a reasonable accuracy. With these techniques, we can construct a per-pixel material labeling from a single image. We observe that groups of high-level material categories have distinct chromaticity distributions. This fact can be used to predict the range of the absolute chromaticity values of objects, provided the material is correctly labeled. We explore whether these constraints are useful in the context of the intrinsic images problem. This paper describes how to leverage material category identification to boost estimation results in state-of-the-art intrinsic images datasets.

1 Introduction

Recent deep learning techniques can predict material labels at every pixel using a combination of a sliding convolutional neural network and a fully connected conditional random field [8]. An example of such a technique, in which Bell *et al*. [4] train their classifier on a dataset with millions of images, predicts the presence of one of their 23 predefined material categories. With their method, accuracies of around 80% are achievable. These high-confidence predictions are in line with findings that low-level cues can be reliably utilized to detect clusters of materials with roughly similar reflectance behavior [14]. It is important to notice that such methods are readily usable on any image and thus make these predictions very broadly applicable.

We can use this aforementioned classification method to identify the dominant material category in an image. We made the observation that the pixels which comprise the material of an object have interesting chromaticity values. Certain material categories have distinct footprints in the YUV color space. Figure 1 gives an intuitive explanation of this principle. Using these distinct footprints and a correctly predicted material category, we design a prior to constrain the chromaticities of the observed objects. We feel that the chromaticity values are strongly correlated with the reflectance estimation. In this paper, we will conduct experiments to show that there is in fact such a relation. Given such a relation, our new prior can be used to link observed chromaticity values to reflectance values of objects.

This prediction is useful in the context of intrinsic image decomposition [9, 11], where reflectance is one of the three recovered factors. Successful recovery of the reflectance values
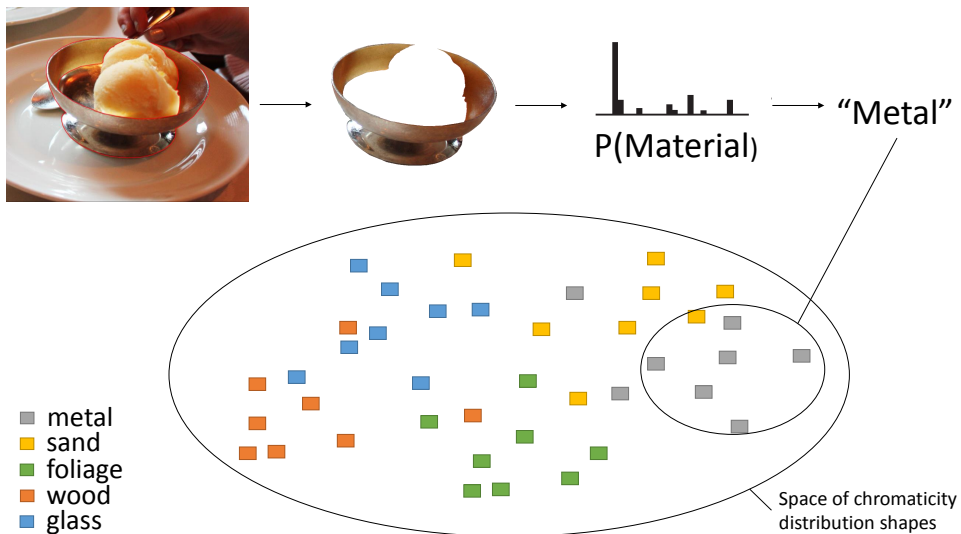
Figure 1: Using recent deep learning techniques, we segment objects in areas of homogeneous materials. We assign a material label to each pixel in these regions. Areas with different materials form islands within the space of chromaticity distributions. The union of all materials distribution covers the entire chromaticity space. If the category is known, it makes sense to use this knowledge to properly constrain the chromaticity values and place more stringent priors on segmented objects parts.

with the help of our chromaticity prior in turn enables better disentanglement of the different intrinsic factors. It is precisely this entanglement that makes intrinsic image decomposition such a hard and long-standing problem [17].

The specific contributions of this paper are:

1. The observation that areas with different materials form distinct subsets within the space of chromaticity distributions.

2. A prior that relates a chromaticity distribution, associated with a specific material category, with possible reflectance values of an object.

3. A novel combination of existing techniques: we used material classification to construct a material-specific chromaticity prior. This is useful for intrinsic image decomposition to improve the estimation of reflectance values.

2 Chromaticity distribution of materials

Existing material databases often are small, artificial datasets with ground truth reflectance, geometry and illumination. The MIT intrinsic image dataset [7], for instance, contains only a limited set of materials that are mostly diffuse. The dataset does not reflect the diversity of materials in the real world. Larger databases, such as OpenSurfaces [2] are unordered collections of images from the real world. The OpenSurfaces dataset contains hundreds of

thousands of images and represents a much wider range of materials. Figure 2 shows how the OpenSurfaces dataset covers a much larger range within the YUV color space and therefore more accurately models chromaticity values of materials in the real world.

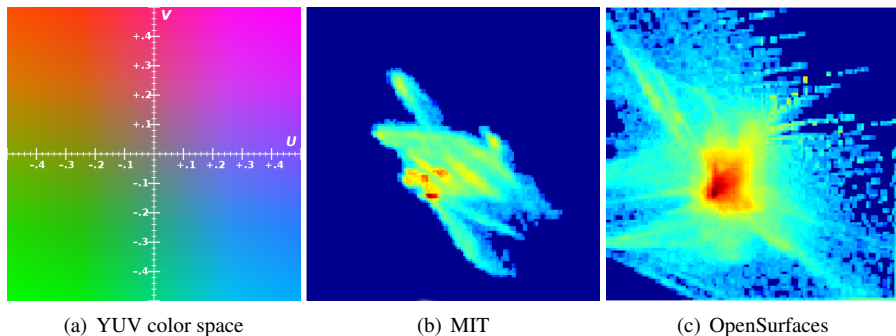(a) YUV color space   (b) MIT   (c) OpenSurfaces

Figure 2: Heatmap plots of chromaticities in YUV color space from random samples of training datasets. Blue represents UV values that are the least frequent, red those that are the most frequent. The YUV color space on the left is for reference. The MIT dataset spans only a small subspace of possible chromaticities. The OpenSurfaces dataset is a more accurate representation of real-world materials.

In the absence of knowledge about the material category in the scene, the OpenSurfaces dataset gives the preferred prediction of chromaticity values. Previous color constancy methods [6] have attempted to cancel out the effects of shading and illumination by transforming the input image into different color spaces. While it is true that the OpenSurfaces dataset covers the entire range of chromaticity values, we have observed that individual materials form clusters of roughly the same chromaticity profile. These clusters have significant overlap in the actual chromaticity values, but their individual distributions have a reasonably unique shape. Figure 3 shows a selection of materials from the OpenSurfaces dataset and their respective chromaticity distributions.

We argue that, when the material category is known, the OpenSurfaces dataset does not constrain the chromaticity values as tightly as possible. A more appropriate solution requires a method tailored towards a specific material. Prior research on image statistics suggests that there is a lot of extra information in natural images that can be exploited [5, 10, 14]. For example, a prior constructed with the OpenSurfaces dataset correctly accounts for the fact that most diffuse objects favor smooth pastel shades over sharp neon tints. However, there are material categories that tend to exhibit sharp highlights and vibrant colors, such as metal and glass. Modeling chromaticities of these materials with the same prior overly generalizes training data by modeling the most common chromaticity values as pastel shades. This might be true in an absolute sense, but subgroups of materials can have very different statistics.

3 Chromaticity Prior

To construct the chromaticity prior, we train our method on samples from segmented photos in the OpenSurfaces material database. We train a separate prior for each of the material categories in the dataset. To keep the processing tractable, we take a limited amount of uniformly distributed samples per training image. We have experimented with the number of
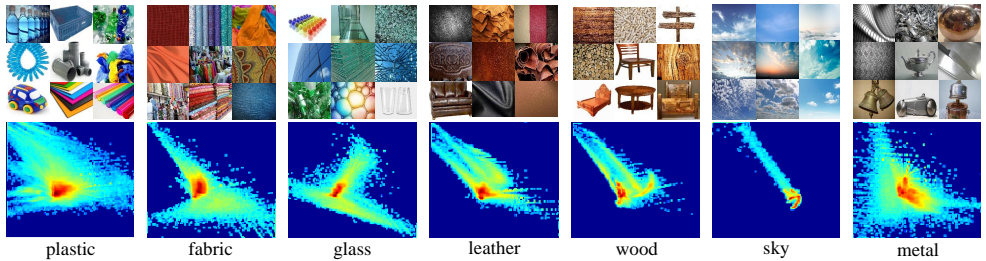
Figure 3: The chromaticities in the YUV color space plotted in heatmap colors, from random samples of different material categories. Blue represents UV values that are the least frequent, red those that are the most frequent. The plots show that various subgroups of materials have different characteristics. Plastic has a much wider range of chromaticity values than sky. Wood spans a limited range of unsaturated colors, while metal has quite a few outliers due to strong specular reflections.

samples and found 2000 per image satisfactory. With more samples results were virtually identical. With these samples, a 3D histogram H of log-RGB chromaticity values is constructed. To fit a non-normalized negative log-likelihood cost function to the training data, we perform a standard density estimation using the L-BFGS optimization algorithm:

$$\underset{P}{\text{minimize}} \quad <P,H> + \log\left(\sum_i \exp(-P_i)\right) + S(P) \tag{1}$$

where P is a 3D probability distribution function (PDF) describing the cost, H is the aforementioned histogram. S is a smoothing term to make the estimation piecewise smooth, by incurring higher costs for larger values of the Jacobian. This is equivalent to Equation 12 of Barron and Malik [1]. Because the smoothing term is isotropic, the log-RGB chromaticity values of both the histogram and test data need to be decorrelated. To this end, the samples of the training data are arranged in a $3 \times n$ matrix $A = AA^T$ and the eigenvalue decomposition is computed: $A = PDP^T$. Here, D is a diagonal matrix. Finally, the whitening transformation is constructed as $W = PD^{-1/2}P^T$.

During inference, we calculate the cost function as the sum of negative log-likelihood values P of the observed chromaticity value at every pixel:

$$cost(C) = \sum_i P(WC_i) \tag{2}$$

where WC_i represents the whitened chromaticity value at the i-th pixel in the input image. Values queried from P are trilinearly interpolated for continuity purposes. Intuitively this cost function assigns a greater penalty to chromaticity values outside the range of the currently detected material category. Chromaticity values that match the material category are more readily accepted as a solution.

Although there is no direct link between chromaticity and reflectance, we can observe in Table 1 that when the material matches that of the object, the chromaticity prior returns the lowest cost. For brevity, we only include the most important materials from OpenSurfaces in the table. Overall, it can be observed that the materials with the best score match very well to the content in the images (Figure 5). Once again, we witness the broad characterization of the material categories. The banana is detected as foliage, but 'food' is not too far behind

in terms of cost. In the next section, we will investigate if it is plausible to use our prior for intrinsic image decomposition.

	Plastic	Fabric	Glass	Leather	Wood	Metal	Paper	Food	Foliage
Basketball	**0.4708**	0.5040	0.5168	0.4937	0.4908	0.5234	0.5190	0.4968	0.4984
Banana	0.2111	0.2276	0.2993	0.2947	0.2169	0.2241	0.2202	0.2062	**0.1922**
Cow	0.1992	0.1920	0.1995	**0.1772**	0.1872	0.2031	0.1986	0.1861	0.1920
Controller	**0.1909**	0.2311	0.2130	0.2009	0.2248	0.2352	0.2397	0.2297	0.2220
Colgate	**0.3597**	0.3718	0.3796	0.3657	0.3672	0.3798	0.3788	0.3691	0.3681
Guitar	0.2539	0.2455	0.2541	0.2366	**0.2291**	0.2604	0.2493	0.2384	0.2183
Hamper	0.5512	0.5361	0.5495	0.5266	**0.4644**	0.5507	0.5478	0.5318	0.5256
Plant	0.3136	0.3090	0.3157	0.3030	0.3024	0.3163	0.3162	0.3055	**0.2967**
Paper bag	0.2792	0.2723	0.3795	0.2707	0.2992	0.2809	**0.2692**	0.2892	0.2730
Van	0.3571	0.3973	0.3502	0.3745	0.3929	**0.3329**	0.3682	0.4320	0.3828

Table 1: Material-specific chromaticity prior cost returned for certain object and material combinations. The best score in every row is in bold. Notice how the materials with the best score match very well to the real material in the images (Figure 5).

4 Results

We would like to apply our chromaticity prior to more interesting applications such as intrinsic image decomposition [9, 11]. Recent work in intrinsic image decomposition using the OpenSurfaces database is described by Bell et al. [3]. As in the original formulation, given an image of a scene, we try to retrieve a set of intrinsic images. These images are all registered with the original image, each describing one intrinsic characteristic. For the purpose of this paper, the recovered characteristics are geometry G, reflectance R and illumination L. Priors are used to estimate the most likely solution for each of these intrinsic factors. The easiest way to test the potential of our new prior, we have integrated it into an existing framework. We have chosen the implementation of Barron and Malik [1] because the source code is readily available.

We use the classifier from Bell *et al.* [4] to determine the dominant high-level material category at each pixel of the input image. This information is used to select the prior that is imposed on the absolute chromaticities. For example, if a group of pixels is determined to contain 'metal', then the prior trained on the subset of images from OpenSurfaces containing metal is used.

In addition to the category, the classifier also provides a confidence measure from its conditional random field (CRF). We only use our specialized priors if the classifier is sufficiently confident about the identity of the material. If the material classification confidence is lower than a specified amount in some image region, it is better not to use the possibly erroneous information and fall back to inference with the prior trained on the entire Open-Surfaces dataset with images from all material categories. The confidence threshold can be trained using a standard cross-validation technique. In our experiments it was determined to be around 75%.

In this section, we will conduct experiments to evaluate our material-specific prior. Every input image is masked off to contain only the objects of interest and is then fed to the estimation algorithm. The test is split up in two parts.

4.1 Results on the MIT dataset

This first experiment tests quantitatively how training on different subsets of materials affects performance on a dataset where ground-truth for the intrinsic factors is available, e.g. the MIT dataset [7]. This ground truth data for geometry, reflectance and illumination is very hard to acquire. We emphasize that the MIT dataset was the only dataset known to us with such ground truth data available. To assemble such a dataset from scratch is a formidable task that requires specialized scanning equipment. We contemplated running experiments on artificially generated ground truth data, but decided against it because it would not be representative of real world scenes. We are fully aware of the limitations of the MIT dataset for comparison purposes, but at least it provides us with a formal way to evaluate our material-specific prior.

Table 2 and Figure 4 contain the results of this experiment. Every row represents the results on a different input image from the MIT dataset. For each test image, we show the mean squared error (MSE) with a prior trained on either all materials or on a specific material. The best matching material (identified by the classifier) is noted between parentheses. The last column reports the improvement due to our method. There are a couple of things that should be remarked.

First, our method gives a consistent boost of $9 - 18\%$. Because we are working in the open source framework of Barron and Malik [1], we can easily put this into perspective by adapting the set of priors that are used in the experiment. For example, disabling their absolute reflectance prior altogether makes for an average decrease in performance of 12%. This demonstrates that putting constraints on reflectances is important, but imposing our material-specific constraints is just as important. We have observed that training on the right material subset gives about the same performance increase as any existing prior from Barron and Malik.

Second, it should be noted that the detected material categories are broad and not necessarily representative of the actual material the object is made of. For example, the 'pear' image is labeled as foliage, while it could also have been labeled as food. In our experience,

	MSE with prior trained on all materials	MSE with prior trained on specific material	Improvement
cup2	0.2583	**0.2248** (paper)	14.92%
deer	0.2848	0.2848 (unknown)	0.00%
frog2	0.3547	**0.2986** (ceramic)	18.77%
paper2	0.3333	**0.3009** (paper)	10.75%
pear	0.3188	**0.2769** (foliage)	15.11%
potato	0.2924	**0.2625** (food)	11.37%
raccoon	0.2755	**0.2425** (plastic)	13.59%
sun	0.3616	**0.3279** (ceramic)	10.28%
teabag1	0.4351	**0.3768** (plastic)	15.48%
squirrel	0.4014	**0.3541** (fur)	13.34%

Table 2: Evaluation of our method on the ground truth reflectance of the MIT dataset. The first column contains the name of the image in the dataset. The second column displays the MSE with a prior trained on all materials. The column of our method contains the MSE and the name of the material that was identified to be the best match in parentheses. The best scores are in bold. Less error is better.
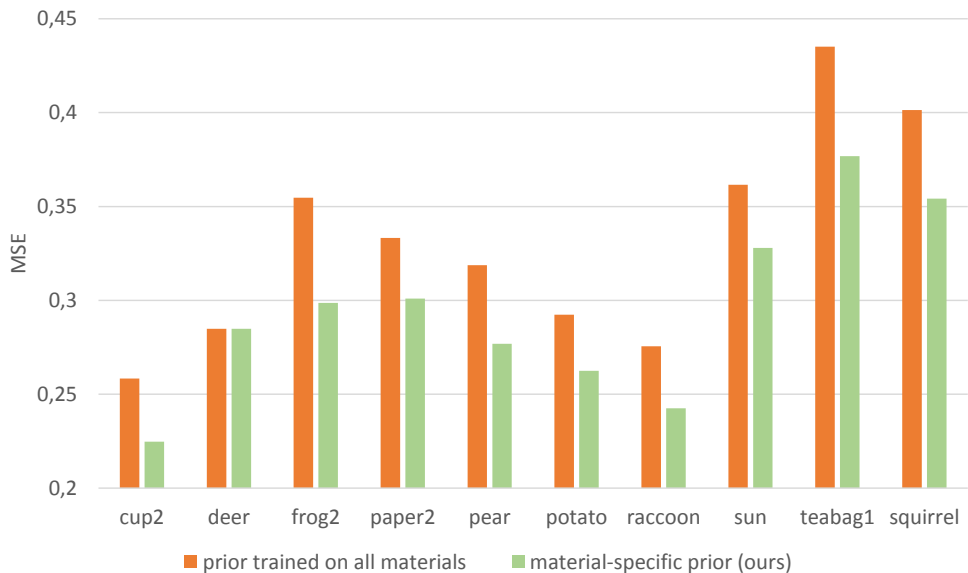
Figure 4: Graph of the error left after optimization. Less is better. Our method of training on specific materials consistently outperforms training on all materials.

it is better to think of the categories as possible reflectance ranges, rather than materials.

Finally, a note on how frequently our material-specific prior is applied. The classifier from Bell et al. [4] has a mean class accuracy of 85.9%. This means that most materials will be correctly identified, while other will be marked as unknown. An example of such a situation is the 'deer' image. In this case, the material classifier could not decisively conclude which material was the best match (marked as unknown in Table 2). In this case we fall back on the prior that was trained on images of all materials.

4.2 Side-by-side visual comparisons

The previous section has demonstrated that using a material-specific prior consistently performs better. However, the images used for evaluation consist of a rather artificial set of objects which span only a limited set of materials. Our second experiment tests how our method performs 'in the wild', on objects with a greater diversity of materials, we use the set of cellphone images that come with the framework of Barron and Malik (see Figure 5). These are just regular images, so there is no ground truth intrinsic data available for comparison. Therefore, it is not possible to apply a quantitative error measurement. Nevertheless, we will show qualitative results to empirically verify that our technique compares favorably with respect to image quality.

Figure 6 provides visual side-by-side comparisons, to show the improvement in the estimated intrinsic factors. We divide the results of the experiment in two columns: the left column contains the estimations generated with a prior trained on all materials, while the right column contains the estimations with the prior trained on the detected material category. Each result consists of the input image, the estimated depth, normals, albedo, shading and illumination. The basketball image has a slightly better separation of albedo and geom-

| plastic | foliage | leather | plastic | plastic |
| wood | wood | foliage | paper | metal |

Figure 5: The images that are used to generate the results. In reading order: Basketball, Banana, Cow, Controller, Colgate, Guitar, Hamper, Plant, Paper bag and Van. The label of the most dominant material category is included. Images courtesy of Barron and Malik [1].

etry when trained with our specialized prior. This is best observable in the normal map and shading images, where the black groove in the center of the ball is more visible. With the banana image, our prior results in a smoother geometric approximation, whereas the original prior models the banana overly bumpy. The cow's albedo is slightly better approximated with our prior. The difference between the priors when applied to the controller image is striking: our prior results in a much better separation between shading and albedo. The old prior models the albedo as shading effects, by incorrectly adjusting the normals. Notice how our optimized albedo and shading end up in a much better minimum. The colgate and guitar images show moderate improvements in the recovered geometry. The hamper, plant, paper bag and van images show a more significant improvement in the recovered geometry. The bumps on the hamper are more pronounced, the edges of the plant and pot are more detailed and the recovered geometric folds of the paper bag align better with the input image. The van shows a significant overall improvement in the recovered detail.

5   Conclusion

This paper has described how knowledge of the high-level material category can be leveraged to predict the distribution of chromaticities in an image. We have shown that groups of materials have similar chromaticity distributions and that this fact can be used to impose stricter priors on the absolute reflectance values. When the material is correctly predicted, our method provides a consistent boost in the context of the intrinsic image problem. We have demonstrated the increase in performance both numerically and qualitatively. Our method is broadly applicable, because it takes regular images as input.

6   Acknowledgements

Figure 6: Visual comparison of results generated with a prior trained on all materials (left column) and with priors trained on the detected material category (right column). Each result consists of 6 images from left to right: the input image, estimated depth, normals, albedo, shading and lighting. Overall, the material-specific priors enable a better separation of the intrinsic factors and finer details are recovered. Best viewed digitally.

References

[1] Jonathan T Barron and Jitendra Malik. Shape, illumination, and reflectance from shading. *TPAMI*, 2015.

[2] Sean Bell, Paul Upchurch, Noah Snavely, and Kavita Bala. OpenSurfaces: A richly annotated catalog of surface appearance. *ACM Trans. on Graphics (SIGGRAPH)*, 32 (4), 2013.

[3] Sean Bell, Kavita Bala, and Noah Snavely. Intrinsic images in the wild. *ACM Trans. Graph.*, 33(4):159:1–159:12, July 2014. ISSN 0730-0301. doi: 10.1145/2601097. 2601206. URL http://doi.acm.org/10.1145/2601097.2601206.

[4] Sean Bell, Paul Upchurch, Noah Snavely, and Kavita Bala. Material recognition in the wild with the materials in context database. *Computer Vision and Pattern Recognition (CVPR)*, 2015.

[5] David J Field. Relations between the statistics of natural images and the response properties of cortical cells. *JOSA A*, 4(12):2379–2394, 1987.

[6] David A Forsyth. A novel algorithm for color constancy. *International Journal of Computer Vision*, 5(1):5–35, 1990.

[7] Roger Grosse, Micah K. Johnson, Edward H. Adelson, and William T. Freeman. Ground-truth dataset and baseline evaluations for intrinsic image algorithms. In *International Conference on Computer Vision*, pages 2335–2342, 2009. doi: http://dx.doi.org/10.1109/ICCV.2009.5459428.

[8] Xuming He, R.S. Zemel, and M.A. Carreira-Perpindn. Multiscale conditional random fields for image labeling. In *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*, volume 2, pages II–695–II–702 Vol.2, June 2004. doi: 10.1109/CVPR.2004.1315232.

[9] Barrow H.G. and Tenenbaum J.M. Recovering intrinsic scene characteristics from images. *Computer Vision Systems*, 2:3–26, 1978.

[10] Jinggang Huang and D. Mumford. Statistics of natural images and models. In *Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference on.*, volume 1, page 547 Vol. 1, 1999. doi: 10.1109/CVPR.1999.786990.

[11] Edwin H. Land, John, and J. Mccann. Lightness and retinex theory. *Journal of the Optical Society of America*, pages 1–11, 1971.

[12] Ravi Ramamoorthi and Pat Hanrahan. A signal-processing framework for inverse rendering. In *Proceedings of the 28th annual conference on Computer graphics and interactive techniques*, SIGGRAPH '01, pages 117–128, New York, NY, USA, 2001. ACM. ISBN 1-58113-374-X. doi: 10.1145/383259.383271. URL http://doi.acm.org/10.1145/383259.383271.

[13] Daniel L Ruderman and William Bialek. Statistics of natural images: Scaling in the woods. *Physical review letters*, 73(6):814, 1994.

[14] G. Schwartz and K. Nishino. Visual material traits: Recognizing per-pixel material context. In *Computer Vision Workshops (ICCVW), 2013 IEEE International Conference on*, pages 883–890, Dec 2013. doi: 10.1109/ICCVW.2013.121.