Texture Classification with Minimal Training Images

Alireza Tavakoli Targhi
CVAP, KTH, Stockholm
att@kth.se

Jan-Mark Geusebroek
University of Amsterdam
jm.geusebroek@gmail.com

Andrew Zisserman
University of Oxford
az@robots.ox.ac.uk

Abstract

The objective of this work is classifying texture from a single image under unknown lighting conditions. The current and successful approach to this task is to treat it as a statistical learning problem and learn a classifier from a set of training images, but this requires a sufficient number and variety of training images.

We show that the number of training images required can be drastically reduced (to as few as three) by synthesizing additional training data using photometric stereo. We demonstrate the method on the PhoTex and ALOT texture databases. Despite the limitations of photometric stereo, the resulting classification performance surpasses the state of the art results.

1. Introduction

Material classification from single images has received extensive theoretical and experimental treatment [5, 7, 8, 15, 18, 19]. Several material databases have been created [1, 2, 3, 8], providing materials under multiple illumination directions. Since the appearance of a surface texture is highly dependent on the illumination direction, statistical learning techniques have been applied to capture this appearance variation from a sufficient number of training examples.

Material recognition as a 3D texton modeling problem was introduced by Leung and Malik [18], triggered by the availability of the CURET collection [8]. Varma and Zisserman [19] improved upon their method and classification performance by constructing a rotational invariant filter set and using multiple training samples per material. Broadhurst [4] continued their work and replaced the texton based classifier with a multivariate Gaussian classifier, further boosting performance on the CURET collection.

Apart from the above appearance based methods, a physical model for texture classification may be used. Chantler and co-workers [6] use photometric stereo to derive a method for material classification invariant under illuminant tilt or, similarly, material rotation. By estimating surface normals, they derive an invariant from surface properties rather than from image properties. We continue on this line of research and combine both physical and statistical models, where we aim at classification of materials from single images taken under arbitrary illumination direction and viewpoint, and combine both physical and statistical models. We improve upon the experimental results of [6] by obtaining perfect recognition rates for two relevant datasets.

The central question posed in this paper is the following: *can we reduce the number of training images (taken under different illumination directions) and reach the same or even a better classification rate?* If we can, then the necessity for a large number of training images, and the labor involved in producing these, can be avoided. We show using a quite simple model, Lambertian photometric stereo [20], that additional training examples can be generated starting from a small set of original images. Despite the shortcomings of the generation method, these synthesized training images are sufficient to surpass the classification performance of methods trained on the original real images.

This idea of generating additional training images to improve classification is not novel, though using a physical model for texture is novel to the best of our knowledge. In other domains, for example, additional training images have been generated in the case of eigenspace representations for object recognition [16], for template-based shape matching of pedestrians [11], and for feature point matching in tracking [**?**]. Recent successes in face recognition rely on the physical modeling of the surface geometry [17]. We show similar methodology for a completely different recognition problem.

2. Augmenting by Rendered Data

The state-of-the-art in classification of material textures is the method proposed by Broadhurst [4]. The classifier takes the marginal responses of the MR8 filter

978-1-4244-2175-6/08/$25.00 ©2008 IEEE

bank [19] and builds a multivariate Gaussian classifier which takes into account the variance between the filter responses of each material class. We reproduce his classification rates of 98% on the CURET database for 61 material samples, using the same setup as [19] of 46 training and 46 test images per sample. This rate differs slightly from the reported rate of over 99% in [4], a difference we attribute to small variations in the construction of the MR8 filter set.

We aim to improve on the results of Broadhurst [4] by including rendered data in the training set. We apply photometric stereo to reconstruct the material surface from the training samples, and render views with previously unseen illumination (and viewing) directions to augment the learning data.

2.1. Surface Reconstruction

Photometric stereo uses several images of the same view of a surface (or object) captured under different illumination directions to determine the local surface orientation and albedo at each pixel. The traditional photometric stereo method [20] assumes that surface reflectance behaves according to Lambert's law where the intensity I of a pixel varies as

$$I = \rho\mu\mathbf{n}\mathbf{L} \tag{1}$$

where ρ is the albedo which represents how much light is reflected in the form of diffuse reflection, μ is the light source intensity, $\mathbf{n} = (n_x, n_y, n_z)$ is the unit surface normal vector, and $\mathbf{L} = (cos(\tau)sin(\sigma), sin(\tau)cos(\sigma), cos(\sigma))^T$ is the light source direction. We choose the coordinate system in which the image plane is parallel to the $x - y$ plane and the z axis coincides with the viewing direction. Given three or more images, $I = [I_1, I_2, \ldots, I_m]$ obtained under light sources $L = [L_1, \ldots, L_m]$, one can invert (1) to solve for the surface normal and albedo at every pixel. See the textbook of Forsyth and Ponce [10] for a more elaborate treatment of the method. Preferably one uses many more than three light sources and Singular Value Decomposition to robustly recover the surface normal and albedo [21].

2.2. Material Rendering

We render images by applying Lambert's law (1) to the computed normal and albedo map. Furthermore, for values $I < 0$, we clip the result to $I = 0$, effectively implementing self-shadows or attached shadows. More complex photometric effects, like cast shadows, highlights and inter-reflections [10], are not taken into account in this simplistic model. Examples of generated images are shown in Fig. 1. Note that, due to the normalization of the images in [19], the representation is unaffected by an affine transformations of the intensities ($I \rightarrow \alpha I + \beta$) constant across the image.

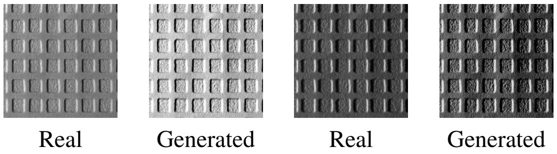


Figure 1: Examples of generated images from the PhoTex database using photometric stereo. Images are rendered after surface reconstruction using 3 illuminations of a paper surface (material ACD). The first pair shows the images with light slant direction $45°$ and tilt $0°$, and the second pair with slant $75°$ and tilt $0°$.

3. Experiments

The **PhoTex** database [3] consists of images of surface textures which are observed from a constant viewpoint for different illumination directions. We follow [9] and select the same 20 materials from the PhoTex database. Each material contains 40 samples taken under different slant and tilt light directions. From the 40 illumination directions available, we use one half as the test set, such that illumination directions are equally spaced and maximally cover the hemisphere. The other half being the training set *T1*. Furthermore, we randomly draw subsets *T2* of 3 images from training set T1, this being the minimal number of images needed for photometric stereo. Note that all random draws are repeated 5 times to yield average performance numbers.

The **ALOT** database [1] consists of 250 materials, see Fig. 2 for examples. The creators of this data set systematically varied viewing angle and illumination angle in order to capture the sensory variation in texture recordings. This collection is similar in spirit to the CURET collection [8]. However, for CURET, photometric stereo cannot be applied due to the absence of 3 (or more) images acquired under similar viewpoint but varying illumination. The acquisition setup for the ALOT [1] is very similar to the ALOI collection of objects [12], see the respective websites [1] for technical details on the setup. For ALOT there are 8 high quality images available, from cameras *c2* and *c4*, with varying illumination direction and similar, but not identical, viewpoint. Since photometric stereo requires the same viewpoint, it is necessary to introduce a registration step to perfectly align them – a potential further source of error in the generated training images. Here, the zero

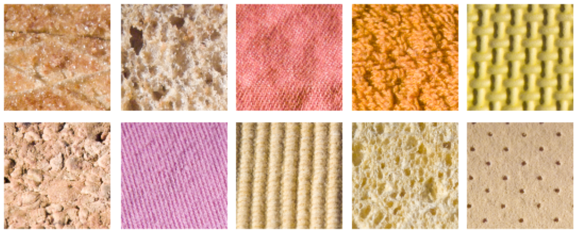


Figure 2: Sample materials from ALOT [1]. In reading order, with ALOT class number between brackets: tea-wafers (9), brown bread (26), cotton (43), terry cloth (48), punched plastic (56); cork (57), cotton (60), ribbed cotton (64), sponge (176), and chamois (196).

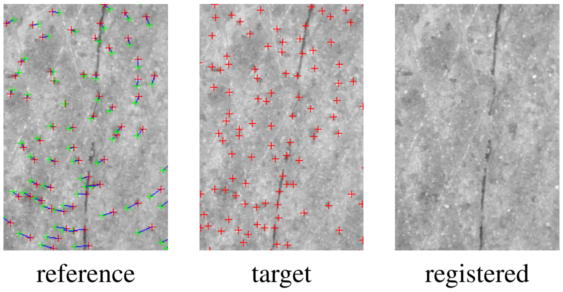
degrees views for the two cameras are used as the training set and for the photometric stereo reconstruction, and the 60 and 120 degree views are used as the test set. The test set then consists of 20 images per material. Again we select a training set *T1*, which here consists of 8 images per material, and a minimal training set *T2* as a random sample of 3 images from T1.

3.1. Image Registration for ALOT images

Due to small misalignments between the two cameras, the viewpoints are not perfectly (pixel accuracy) identical. For materials with not too much depth variation, the distortion can be well approximated by a planar rotation and translation between the two views. Hence, we apply image homography registration[1] [14] to align the images between the two cameras. To minimize the effects of shadows on the calibration procedure, we choose the condition with all five lights turned on as the calibration pair. The other images are aligned using the estimated homography.

We applied the Harris corner detector [13] and subsequently correlation matching to find matching sets of points between images. A RANSAC sampling strategy is applied to deal with noise and outliers, yielding a robust estimate of the correspondences between cameras c2 and c4 for each material. The final maximum likelihood estimate of the homography matrix is computed using singular value decomposition [14]. An example of the resulting correspondences is given in Fig. 3.

The registration method is only valid for near planar surfaces. For surfaces with large relief other registration methods (such as dense stereo) need to be used. Therefore, we have selected materials with limited depth variation, see Fig. 2.



reference target registered

Figure 3: Example of registration of ALOT images using a planar homography. The correspondences between the reference and target are superimposed on the reference image.

	Train		Test	Classification %	
	Real	Rendered		Real	Rendered
T1	20	20	20	96.00	72.00
T2	3	3	20	82.75	68.25

Table 1: Classification performance for the real training data versus rendered exemplars of the training images.

3.2. Results

We will first test how well the rendered data mimics the real images. Therefore, we compare the performance using the original Photex training data with the performance when substituting the training images with rendered data. The experiment demonstrates how accurately one can render the original image data from photometric stereo using our simple Lambertian photometric model. Table 1 gives the results. As expected, the photometric stereo reconstruction together with Lambertian rendering of synthetic images is far from ideal. Despite this, the simple Lambertian model is good enough to achieve a far above chance level categorization, which for 20 material classes is at 5%. This is even the case when only three training images are available. Hence, we expect improved categorization performance when augmenting the real training images with such simply rendered images.

We next test how far generated data aids in classification, using photometric stereo to *augment* the training set. Here, we take the original training data and augment it with generated views of random illumination directions. Adding more and more data is expected to improve recognition rate until saturation. Result are given in Fig. 4. The graphs start at the baseline for the real training images of T1 (20 for PhoTex and 8 for ALOT) and T2 (3 images), respectively, and show performance when adding rendered images to the training data.

[1] Software from Peter Kovesi:http://www.csse.uwa.edu.au/∼pk

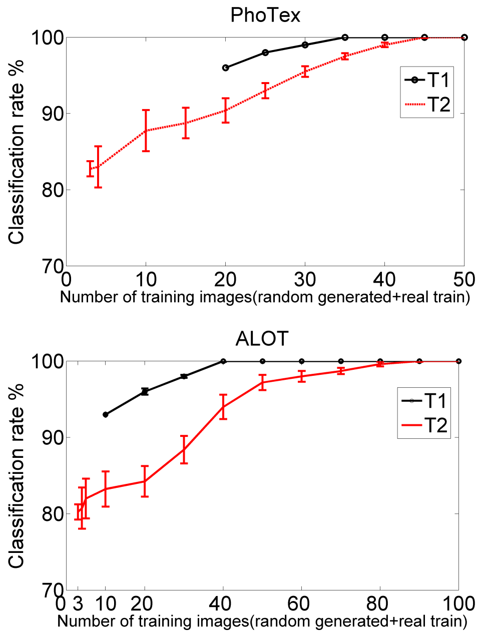


Figure 4: Results for augmenting training data, showing performance improvement as a function of the number of augmented images.

For the PhoTex dataset, the initial points of the curves in Fig. 4 correspond to the results given in Table 1 for the training data. The results for both PhoTex and ALOT show a consistent performance increase when adding more data, and saturates at perfect recognition. Regarding the three-images case (T2, red line), standard deviation is not too large for the initial (random) choice of real training images; then increases when adding Lambertian rendered images of random illumination directions; to decrease again as performance gets closer to 100%. Note that our results improve upon the state of the art for PhoTex [6], while we are doing a harder classification task. Even when having as few as 3 training images, materials are perfectly classified when augmenting the data to a total of 45 images for PhoTex and 90 for ALOT.

4. Conclusions

Interestingly, and despite the considerable body of prior art, our work shows that there is still room to improve in the learning of texture classifiers. Using photometric stereo to obtain a physical model of the texture, in combination with Lambertian rendering to augment the training data, classification performance increased considerably. Indeed, a perfect classification performance was obtained from as few as three original images of each material class for state of the art datasets. Note that these datasets contain non-lambertian materials; there are specularities and unmodelled cast shadows in the images. Apparently, Lambert's law is already descriptive enough to capture many aspects of material surface reflectance.

References

[1] ALOT. www.science.uva.nl/∼mark/ALOT.

[2] KTH-TIPS2. www.nada.kth.se/cvap/databases/kth-tips.

[3] PhoTex. www.cee.hw.ac.uk/texturelab/database/photex.

[4] R. E. Broadhurst. Statistical estimation of histogram variation for texture classification. *Texture*, 2005.

[5] B. Caputo, E. Hayman, and P. Mallikarjuna. Class-specific material categorisation. *ICCV*, 2005.

[6] M. Chantler, M. Petrou, A. Penirsche, M. Schmidt, and G. McGunnigle. Classifying surface texture while simultaneously estimating illumination direction. *IJCV*, 2005.

[7] O. G. Cula and K. J. Dana. Compact representation of bidirectional texture functions. *CVPR*, 2001.

[8] K. J. Dana, B. van Ginneken, S. K. Nayar, and J. J. Koenderink. Reflectance and texture of real-world surfaces. *ACM Transactions on Graphics*, 1999.

[9] O. Drbohlav and M. J. Chantler. Illuminant-invariant texture classification using single training images. *Texture05*, 2005.

[10] D. A. Forsyth and J. Ponce. *Computer Vision: A Modern Approach*. Prentice Hall Professional Technical Reference, 2002.

[11] D. M. Gavrila and J. Giebel. Virtual sample generation for template-based shape matching. *CVPR*, 2003.

[12] J. M. Geusebroek, G. J. Burghouts, and A. W. M. Smeulders. The Amsterdam library of object images. *IJCV*, 2005.

[13] C. Harris and M. Stephens. A combined corner and edge detection. *Proceedings of The Fourth Alvey Vision Conference*, 1988.

[14] R. Hartley and A. Zisserman. *Multiple view geometry in computer vision*. 2004.

[15] E. Hayman, B. Caputo, M. Fritz, and J.-O. Eklundh. On the significance of real-world conditions for material classification. *ECCV*, 2004.

[16] P. Jain, K. Rao, and C. Jawahar. Computing eigen space from limited number of views for recognition. *ICVGIP*, 2006.

[17] R. Jenkins and A. M. Burton. 100% accuracy in automatic face recognition. *Science*, 2008.

[18] T. Leung and J. Malik. Representing and recognizing the visual appearance of materials using three-dimensional textons. *IJCV*, 2001.

[19] M. Varma and A. Zisserman. A statistical approach to texture classification from single images. *IJCV*, 2005.

[20] R. J. Woodham. Photometric method for determining surface orientation from multiple images. *Optical Engineering*, 1980.

[21] A. Yuille, D. Snow, R. Epstein, and P. Belhumeur. Determining generative models of objects under varying illumination: Shape and albedo from multiple images using svd and integrability. *IJCV*, 1999.