Classifying Textures when Seen from Different Distances

X. Lladó* and M. Petrou

Centre for Vision, Speech, and Signal Processing.
University of Surrey, Guildford, GU2 7XH, UK.
E-mail: {llado@eia.udg.es,m.petrou@eim.surrey.ac.uk}

Abstract

The purpose of this work is to analyse what happens to the surface information when the image resolution is modified. We deduce how the same surface appears if seen from different distances. Using 4-source Colour Photometric Stereo, which provides the surface shape and colour information, a method for predicting how surface texture looks like when changing the distance of the camera is presented. We demonstrate this technique by classifying textured surfaces seen from different distances than the textured surfaces in the database.

1. Introduction

The main motivation for this paper is the problem of description of multicoloured surfaces invariant to the geometry. Recognition of 3-dimensional surface textures from 2-dimensional images is difficult. The 2-dimensional texture in the image, the *image* texture, is produced by variation in both surface reflectance and surface relief. The latter two constitute the *surface* texture. While the reflectance properties are intrinsic to the surface, the surface relief produces a pattern of shadings that depends strongly on the direction of the illumination [2]. Thus, the *image* texture created by a 3D *surface* texture changes drastically with the imaging geometry.

This paper uses *Colour Photometric Stereo* (CPS), as described in [1] and [6], to compute the detailed shape and colour of a rough surface when seen by a camera at the zenith of the surface. We then assume that in a database of textures, we have all the information concerning the surface texture constructed from the photometric stereo set. We assume that we are given the image of one of these textures captured by a camera at a different (longer) distance and with unknown direction of illumination. From the information in the database, we predict how each surface would look like when seen from the new distance, for various directions of illumination. Thus we create a "virtual" database of image textures against which we compare the unknown image texture in order to classify it. Recognition of the texture allows us also to guess the approximate orientation of the illumination under which the image was captured. The image texture classifier we use is based on the co-ocurrence matrices [4].

2. Prediction process

2.1. Image prediction

We shall start by considering two grids referring to the pixels of two images of the same surface, captured from two different distances. One of them must correspond to the higher resolution image and it must be finer than the other. Let us indicate by indices ij a pixel of the coarse grid. This pixel is made up from several pixels of the fine resolution grid, some of which contribute to it only partially. Let us for the moment ignore by how much each pixel of the fine resolution contributes to pixel ij of the coarse resolution, and let us simply say that "superpixel" ij corresponds to a tile of size $K \times L$ of fine resolution pixels. We shall refer to the pixels of the coarse resolution as "superpixels" and the term "pixel" will be used only for the fine resolution pixels. Each superpixel may be thought of as representing a surface patch characterised by a particular gradient vector $(p_{ij}, q_{ij}, 1)^T$ and a particular reflectance function $\rho_{ij}(\lambda)$. The superpixel will have intensity I^u_{ij} in the coarse resolution image, $u = 1, 2, 3$ or 4, each corresponding to a different direction of the illumination.

Each superpixel corresponds to a tile of pixels. We wish to keep track of the superpixel to which a pixel contributes. So, we shall give to every pixel three sets of indices: one tells us to which tile it belongs, one tells us where about in the tile it is, and one tells us its location in the fine resolution grid. Let us indicate by indices mn the position of pixels in

* On leave of absence from Computer Vision and Robotics Group, Institute of Informatics and Applications, University of Girona, 17071, Girona, Spain.

the fine resolution grid. So, a pixel that contributes to superpixel ij will have indices $ij;kl\ mn$, where $k = 1, 2, ..., K$ and $l = 1, 2, ..., L$. Any other quantity associated with pixel $ij;kl\ mn$ will be indicated by the same notation as for superpixel ij. That is, pixel $ij;kl\ mn$ corresponds to a surface patch with gradient vector $(p^{mn}_{ij;kl}, q^{mn}_{ij;kl}, 1)^T$ and a reflectance function $\rho^{mn}_{ij;kl}(\lambda)$. Our problem is to predict I^u_{ij}, for a given direction of illumination u, given $\rho^{mn}_{ij;kl}(\lambda)$, $p^{mn}_{ij;kl}$ and $q^{mn}_{ij;kl}$ for all values of i, j, k and l. The values of $\rho^{mn}_{ij;kl}(\lambda)$, $p^{mn}_{ij;kl}$ and $q^{mn}_{ij;kl}$ have been computed from four images by Colour Photometric Stereo. We shall not go into details of this now as they have been published elsewhere [1, 6]. Although the CPS scheme we use can deal with non-Lambertian surfaces, we assume here that the surface we are dealing with is Lambertian.

If the sensitivity of the sensor is $\mathcal{S}(\lambda)$, the spectral distribution of the incident light is $\mathcal{L}(\lambda)$, and the reflectance funcion of the surface patch projected on superpixel ij is $\mathcal{R}(\lambda, N_{ij}, u)$, where N_{ij} is the normal to the surface vector and u the direction of illumination vector, then the intensity value registered by the sensor at superpixel ij is:

$$I^u_{ij} = \int \mathcal{S}(\lambda)\mathcal{L}(\lambda)\mathcal{R}(\lambda, N_{ij}, u)d\lambda \quad (1)$$

$\mathcal{R}(\lambda, N_{ij}, u)$ could be separated in a geometric component $\mathcal{G}(N_{ij}, u)$ and a surface material component, $\rho_{ij}(\lambda)$. Assuming we are dealing with a Lambertian surface,

$$\mathcal{G}(N_{ij}, u) = \frac{N_{ij} \cdot u}{|N_{ij}||u|} \quad (2)$$

where N_{ij} is $(p_{ij}, q_{ij}, 1)^T$, so:

$$I^u_{ij} = \int \mathcal{S}(\lambda)\mathcal{L}(\lambda)\mathcal{G}(N_{ij}, u)\rho_{ij}(\lambda)d\lambda \quad (3)$$

Superpixel ij is made up from several pixels each of which may have its own reflectance function, and its own orientation. So, $\mathcal{G}(N_{ij}, u)\rho_{ij}(\lambda)$ is really the sum of several such factors, one for each pixel that contributes to the superpixel. Then

$$I^u_{ij} = \int \mathcal{S}(\lambda)\mathcal{L}(\lambda) \sum_{k,l=1}^{K,L} \mathcal{G}(N^{mn}_{ij;kl}, u)\rho^{mn}_{ij;kl}(\lambda)d\lambda \quad (4)$$

By exchanging the order of integration and summation, we obtain:

$$I^u_{ij} = \sum_{k,l=1}^{K,L} \int \mathcal{S}(\lambda)\mathcal{L}(\lambda)\mathcal{G}(N^{mn}_{ij;kl}, u)\rho^{mn}_{ij;kl}(\lambda)d\lambda \quad (5)$$

Note that this formula is quite general: it allows us to predict the value of superpixel ij from the information we have on its constituent pixels, even when seen by a different sensor, under illumination with different spectral distribution and different orientation than those under which the original images, from which the surface information was extracted, were captured.

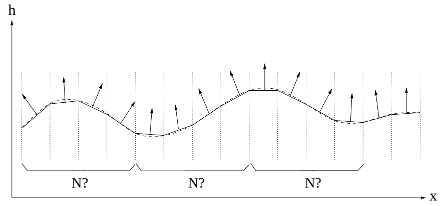

Figure 1. A surface is approximated by many flat facets in fine resolution. In coarse resolution these are replaced by large flat facets. What is the gradient vector of a large facet as a function of the gradient vectors of the small facets it replaces?

We shall restrict ourselves here to the case where the sensor and light source are the same. If we assume that the unknown illumination direction is the same as one of the illumination directions in the original data used by the photometric stereo, then this equation collapses to a trivial one:

$$I^u_{ij} = \sum_{k,l=1}^{K,L} I^{u;mn}_{ij;kl} \quad (6)$$

This approach allows us to go straight to predicting how an image of the surface would look like when imaged under different conditions, from those under which the images in the database were imaged.

Alternatively, we may try to predict first how the surface itself would be approximated in a different resolution from the original one and then predict the image it would create.

2.2. Surface prediction

As our objective is to describe surface information we need to understand what happens with the surface shape information if the distance of the camera is changed. Figure 1 illustrates this problem and permits us to formulate the following question: what will the normal vectors be if the distance of the camera is changed leading to a new image in which every pixel is the union of several old pixels? This question is answered by deriving the relationship between the normal vectors when they are calculated in different image resolutions. The proposed strategy in order to recover the normal vectors is based on the prediction of the planar patch by which the surface is approximated locally in the coarse resolution. We want to reconstruct the facet of the superpixel by using the information of the facets recovered in the fine resolution.

To predict the planar patch of the superpixel we compute the average plane in the LSE sense passing through the patches of the fine resolution. That means it is necessary to know the height $f(x, y)$ at each point of the surface (*height map*). Using the height information and knowing the facets which contribute to superpixel ij, the average plane can be computed. Consequently, the normal vector of the recovered plane will be the normal vector of the superpixel.

If we define a surface by $(x, y, f(x, y))$ where $f(x, y)$ is the height at point (x, y), the normal as a function of (x, y) is

$$N(x, y) = \frac{1}{\sqrt{p^2 + q^2 + 1}}(p, q, 1)^T \tag{7}$$

where the partial derivatives give us the values of the gradient vector $p = \frac{\partial f}{\partial x}$ and $q = \frac{\partial f}{\partial y}$. To recover the height map, we determine $f(x, y)$ from measured values of the unit normal [5, 3]. The partial derivative gives the change in surface height with a small step in either the x or the y direction. That means we can get the surface by summing these changes in height along some path. For example, the surface at (u, v) can be reconstructed by starting at $(0, 0)$, summing the y-derivative along the line $x = 0$ to the point $(0, v)$, and then summing the x-derivative along the line $y = v$ to the point (u, v). This is the next integration path

$$f(u, v) = \int_0^v \frac{\partial f}{\partial y}(0, y)dy + \int_0^u \frac{\partial f}{\partial x}(x, v)dx + c \tag{8}$$

where c is the constant of integration, which represents the unknown height of the surface at the starting point (in our experimental results we have chosen $c = 0$). Any other set of paths would work as well, though it is best to use many different paths and average, so as to reduce the error in the derivative estimates. If different paths yield very different values, obviously the surface is not integrable between the two chosen points. In other words, the described prediction method works well if the surface is continuous. When the surface presents discontinuities, i.e. places where its derivatives do not exist, it is not integrable and the above method leads to incorrect normal vectors. The photometric stereo surface reconstruction yields also the set of points where the recovery of the gradient vectors is impossible. In general these are places which are in shadow in more than one of the four images used in the photometric stereo set. Such points are likely to occur at places where the surface has deep "ravines" and they may be associated with the places where the surface is not differentiable and continuous. Thus, when we reconstruct the surface we stop the piecewise integration at these boundaries. The surface shape we reconstruct and the image intensity we predict consist of image patches and not full images.

3. Experimental results

The proposed method was tested with several photometric sets consisting of four images each, obtained by placing the camera at a distance along the zenith of the surface. Other images of the same surfaces were captured from a longer distance for 4 different illumination directions. These are the images we shall want to classify, using as reference images those of the photometric stereo sets.

The photometric database consists of seven surfaces. First row of figure 2 shows one of the photometric set images for four of the surfaces. From the information in the database, we predict how each surface looks like when seen from the longer distance, for 4 different directions of illumination. Hence we create a "virtual" database of 28 image textures against which we compare the unknown image texture in order to classify it. The first image in the third row of figure 2 shows the results of our predictions with a particular direction of illumination. Note that our method can not predict the intensity values for the pixels which contain a discontinuity in the surface because the normal vector can not be computed there. We flag these points with white grey value.

We start by obtaining a representative feature vector for each texture image in the "virtual" database, using the co-ocurrence matrices. The co-ocurrence matrices were implemented in an isotropic way for fixed distance d, computing two of the most typical features, contrast and homogeneity, for 60 different values of d. The pixels labeled as discontinuity points are not used to compute the co-ocurrence matrices. Among all the computed features, those which could discriminate between the different classes best were chosen. These turned out to be contrast at distance 29, and homogeneity at distance 2. After that we build a classifier for this set of features, which calculates the feature vector for the unknown image texture, and assigns it to one of the classes of the "virtual" database. Second row of figure 2 shows examples of the unknown images we want to classify.

As these images are large, we produce from each one of them 9 subimages to be used for testing. Thus, we had in all $7 \times 9 \times 4 = 252$ test images of size 94×94 each (7 different surfaces, 4 different directions of illumination, 9 subimages). The classification rate when using the method of section 2.1 was 71.47%.

In two alternative experiments we tried to classify the same patterns without going through the prediction stage, but just using features constructed from the original high resolution images. In this case we only classified correctly 21.83% of patterns. In a third experiment, we used

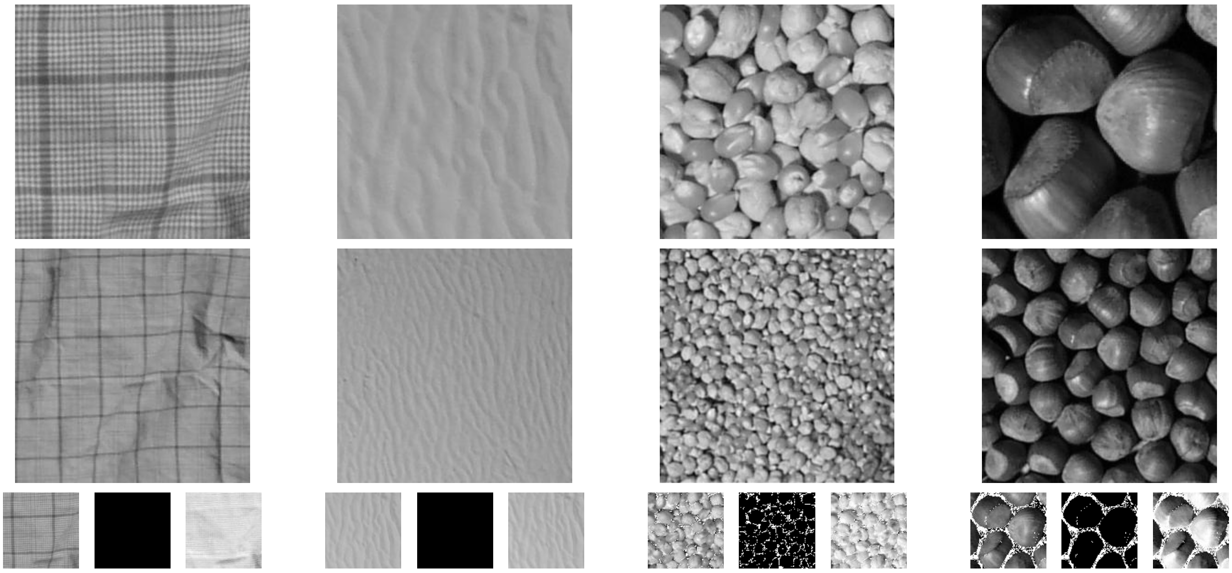

Figure 2. Four of the surfaces used in the classification. First row: one image of the photometric stereo set, second row: images captured from a longer distance, third row: image predictions (section 2.1), detected discontinuities, surface shape rendered (section 2.2).

again features constructed from the original images but we scaled up the distance d used for the construction of the co-ocurrence matrix according to the change in image resolution. In this case the recognition rate of the classifier was 32.94%.

If we use the method of section 2.2 to predict the surface orientation, we may construct our virtual database by using the predicted surface gradient vectors, the average reflectance function for each surface tile and assuming that our sensor has narrow sensitivity (i.e. $\mathcal{S}(\lambda)$ is a delta function) while the illuminant is white (i.e. $\mathcal{L}(\lambda)$ is a constant). In this case the recognition rate was 64.64%.

4. Conclusions

We presented a general framework for recognising textures when seen from different distances. The 4-source CPS has been used in order to obtain the reflectance and the surface shape information of the surface from a close by distance. The proposed method allows one to predict how the texture will look like when seen by a different sensor and under different imaging geometry with an illuminant of different spectral properties. It is based on the assumption of Lambertian surfaces, but it can easily be generalised to other types of surface. The method has been validated using real sets of images in the context of texture recognition when the test data have been captured from different distances than those in the database.

5. Acknowledgments

This work was supported by UdG grant BR00/03, and EPSRC grant GR/M73361. The photometric stereo images used were collected by G. McGunnigle and M.J. Chantler in Heriot-Watt University, UK.

References

[1] S. Barsky and M. Petrou. Colour photometric stereo: Simultaneous reconstruction of local gradient and colour of rough textured surfaces. *In Proceedings 8th International Conference on Computer Vision*, pages II: 600–605, 2001.

[2] M. Chantler. Why illuminant direction is fundamental to texture analysis. *IEE Proceedings in Vision, Image and Signal Processing*, 142(4):199–206, August 1995.

[3] D. Forsyth and J. Ponce. *Computer Vision - A modern approach*. Prentice Hall, Upper Saddle River, N.J., 2002.

[4] R. Haralick, K. Shanmugan, and I. Dunstein. Textural features for image classification. *IEEE Transactions on Systems, Man, and Cybertinetics*, 3(6):610–621, 1973.

[5] R. Klette and K. Schlüns. Height data from gradient fields. *SPIE Proceedings Machine Vision Applications, Architectures, and Systems Integration*, 2908:204–215, November 1996.

[6] M. Petrou and S. Barsky. Shadows and highlights detection in 4-source colour photometric stereo. *In Proceedings International Conference on Image Processing*, Stereoscopic and 3-dimensional Image Processing II. Thessaloniki, October 2001.