

# Two-image comparison under different illumination conditions

Ondrej Drbohlav and Mike Chantler  
School of Mathematical and Computer Sciences  
Heriot-Watt University  
Edinburgh, UK

## Abstract

We present a new theory that shows that two images of the same object taken under two different light directions can be made virtually identical by filtering each image by a directional derivative filter. The direction and magnitude of the derivative is generally different for each image, and depends on illumination directions used. The method requires the object surface to be of uniform Lambertian reflectance and a shallow relief, and the light directions used to be sufficiently inclined from the surface macro-normal. For a specific case when the surface consists of spherical patches, the two images can be made identical, for *any* two light directions used provided that none of them is perpendicular to the viewing axis. We provide some simple experiments which illustrate the validity of this theory.

## 1 Introduction

Handling changes in an image induced by changes in illumination has been the topic of extensive research. Object recognition and texture classification have been the main fields which have contributed to this subject. The changes in an image due to illumination are known to affect the recognition performance dramatically; in the area of face recognition, for example, it was shown that the intra-class variance due to illumination is greater than the inter-class variance due to the change of individual [4]. Our long-term goal is to develop algorithms which would be able to recognise three-dimensional textures captured under different illumination conditions. Such goal implies specific requirements which we discuss below. Our paper presents the novel theory which will enable to construct such algorithms.

We have structured our review as follows. We name two specific properties of the image comparison method which we would like to meet, and discuss their meaning on examples from related work. Finally we review works that presented research that is closest to the work presented here, and then we state the contribution of the paper.

**Two image comparison.** Our goal is to deliver a method which would be able to produce a measure of likeliness that *two* given images are images of the same object captured under different illumination conditions. To describe this requirement in the language of object classification, we assume that the training set for each class consists of only *one* image, which is an image of an object under certain illumination. Such an exemplar image is being compared with a query image. This discriminates our method from methods which

use large training sets, like the method of the *illumination cone* [2, 4], or methods in texture recognition ([8] and [10], to name two) which require several images obtained under different illumination conditions to be present in the training set.

**No joint inter-image features.** Some methods for comparing two images under varying illumination rely strongly on that the two images are *pixel-wise spatially registered*. This means that both the camera and the object are fixed in between changing the illumination, and the co-located pixels in the two images observe the same object point. An example of this approach is that of Chen, Belhumeur and Jacobs [3] which employs *joint* probability of image gradient directions, or that of Jacobs, Belhumeur and Basri [6] which uses the ratio of the two images and employs the fact that such ratio is ‘simpler’ when the two images come from the same object than if they come from two different objects.

In contrast, we require our method to be extendable to the case when the images are not registered. This is mainly motivated by that we want to use the method for relating images of two different instances of a three-dimensional texture (i.e. two realisations of the same texture) which are obviously mis-registered (and even can not be registered, as they only share the surface statistics but are not of the same geometry). This implies the need to deliver a method which makes two images of the same object *virtually identical* (or, of the *same image statistics* in the just discussed case of texture instances), as opposed to *similar*.

There are several works which are close to the ours but do not satisfy the above requirements. Besides the ones already discussed above, a recent work of Osadchy, Lindenbaum and Jacobs [9] achieves illumination quasi-invariance on smooth surfaces using the whitening approach. The assumptions of this method are that the surface is Lambertian and of uniform reflectance, that the surface is of shallow relief and that the illumination direction is sufficiently inclined from the surface macro-normal. However, the method actually requires the images to be spatially registered because rather than increasing similarity between the images of the *same* object, it increases the dissimilarity between images of *different* objects. In addition, it needs to have the filter trained on a number of images, and the filter shape depends on surface roughness.

Our method, as for the surface and light properties, requires the same assumptions as just discussed work [9], namely smooth surfaces of uniform Lambertian reflectance and of shallow relief, and illumination directions which are sufficiently inclined from the surface macro-normal. However, it is able to make the images taken under different illumination conditions almost identical, and therefore, makes possible to relax the requirement of image spatial registration. But even when the images are spatially registered, our method brings true *two-image* comparison because it does not require any training on multiple images, and can be used in the same form for surfaces with arbitrary roughness.

The method presented is based on filtering each of the two images by a linear filter corresponding to directional derivative. Importantly, the filters are generally *different* for each of the images, and depend on the two illumination directions used.

The paper is structured as follows. Section 2 first gives the necessary notations and concepts, and then presents our new theory for image comparison. Section 3 presents simple experiments on real data which illustrate the performance of the method, and Section 4 concludes the paper.

## 2 Theory

### 2.1 Notations and Concepts

**Directional derivatives.** Having a function  $f = f(x, y)$ , the partial derivatives with respect to  $x$  and  $y$  are denoted  $f_x$  and  $f_y$ , respectively. A directional derivative with respect to the vector  $\mathbf{d} \in \mathbb{R}^2$  is denoted  $f_{\mathbf{d}}$ , and its relation to the gradient  $\nabla f = (f_x, f_y)^T$  is  $f_{\mathbf{d}} = \mathbf{d} \cdot \nabla f$ . The vector  $\mathbf{d}$  is called the *directional derivative vector*. Note that while *letter* subscripts denote directional derivatives, *integer* subscripts index the components of a vector:  $\mathbf{a} = (a_1, a_2, \dots, a_n)$ .

**Camera and Coordinate system.** The camera is assumed to be orthographic with square pixels, and the world coordinate system is defined as follows: the  $x$  and  $y$  axes are given by the camera plane, and the optical axis gives the  $z$  axis direction, with the positive direction pointing towards the camera. The  $z$  axis is assumed to be vertical.

**Projected vector.** A *projected* vector is given by projecting a three-vector onto the camera plane. Projected vectors are denoted by a hat. The projected vector  $\hat{\mathbf{a}}$  of a three-vector  $\mathbf{a} = (a_1, a_2, a_3)^T$  has components

$$\hat{\mathbf{a}} = (a_1, a_2)^T. \quad (1)$$

**Surface parametrisation.** It is assumed that the surface can be represented by a height function  $z = z(x, y)$  parametrised by the camera plane axes  $x$  and  $y$ . The function is assumed to be  $C^2$  continuous.

**Surface differential entities.** Having a surface height function  $z = z(x, y)$ , the height gradient  $\mathbf{g}$  and its components  $p, q$  are defined as

$$\mathbf{g} = \nabla z = \begin{pmatrix} z_x \\ z_y \end{pmatrix} \stackrel{\text{def}}{=} \begin{pmatrix} p \\ q \end{pmatrix}. \quad (2)$$

The Hessian of the height function  $z$  is

$$\mathbf{H} = \begin{bmatrix} z_{xx} & z_{xy} \\ z_{yx} & z_{yy} \end{bmatrix} = \begin{bmatrix} p_x & p_y \\ q_x & q_y \end{bmatrix} \stackrel{\text{def}}{=} \begin{bmatrix} a & s \\ s & b \end{bmatrix} \quad (3)$$

Note that the Hessian is symmetric ( $z_{yx} = z_{xy}$ ) which follows from the assumed  $C^2$  continuity. This fact is referred to as *integrability* and plays a crucial role in the development of the theory presented.

**Relationship between the height gradients and surface normals.** Given the surface gradient  $\mathbf{g} = (p, q)^T$ , the unit normal vector at that point is (see for example [5])

$$\mathbf{n} = \frac{1}{\sqrt{1 + p^2 + q^2}} \begin{pmatrix} -p \\ -q \\ 1 \end{pmatrix}. \quad (4)$$

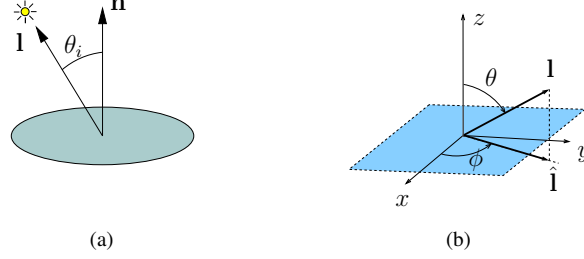


Figure 1: (a) Reflectance geometry: the angle of incidence  $\theta_i$ . (b) Slant and tilt of a light source  $\mathbf{l}$ : slant  $\theta$  is the angle between the camera axis and  $\mathbf{l}$ . Tilt  $\phi$  is the angle between the  $x$ -axis and the projected light  $\hat{\mathbf{l}}$ .

## 2.2 Linear filtering approach

In this Section, we will show how to make the images produced under two different light directions virtually identical by filtering each of them by a (different) linear filter.

For Lambertian objects [7], the reflectance at a surface patch is characterised by *albedo* which represents the amount of light which is scattered from the surface back into the air, and is denoted  $\rho$ . The intensity  $i$  at a pixel observing the patch is then

$$i = \rho \sigma \cos \theta_i, \quad (5)$$

where  $\theta_i$  is the angle of incidence (see Fig. 1(a)) and  $\sigma$  is the light source intensity. In this article, we will be dealing with uniform albedo surfaces, and without loss of generality we set  $\rho = 1$ . The above equation can then be rewritten as

$$i = \sigma \mathbf{l}^T \mathbf{n} = (\sigma \mathbf{l})^T \mathbf{n} = \mathbf{s}^T \mathbf{n}, \quad (6)$$

where  $\mathbf{l}$  and  $\mathbf{n}$  are the unit light and normal vectors, respectively, and the vector  $\mathbf{s} = \sigma \mathbf{l}$  will be called the *scaled light* throughout this paper. Now, expressing the normal  $\mathbf{n}$  as in Eq. (4) gives

$$i = \mathbf{s}^T \frac{(-p, -q, 1)^T}{\sqrt{1+p^2+q^2}} = \frac{-s_1 p - s_2 q + s_3}{\sqrt{1+p^2+q^2}}. \quad (7)$$

Forming the derivative of image intensity with respect to  $x$  gives

$$\begin{aligned} i_x &= \frac{-s_1 p_x - s_2 q_x}{\sqrt{1+p^2+q^2}} - \frac{1}{2} \frac{-s_1 p - s_2 q + s_3}{\sqrt{1+p^2+q^2}^3} (2pp_x + 2qq_x) = \\ &= \frac{-(s_1, s_2)(a, s)^T}{\sqrt{1+p^2+q^2}} - i \frac{(p, q)(a, s)^T}{1+p^2+q^2} = \frac{-\hat{\mathbf{s}}^T(a, s)^T}{\sqrt{1+p^2+q^2}} - i \frac{-\hat{\mathbf{n}}^T(a, s)^T}{\sqrt{1+p^2+q^2}}, \end{aligned} \quad (8)$$

where  $\hat{\mathbf{n}}$  is the projected surface normal  $\hat{\mathbf{n}} = (-p, -q)^T / \sqrt{(1+p^2+q^2)}$  (cf. Eq. (4)). The other component of intensity gradient is, similarly,

$$i_y = \frac{-\hat{\mathbf{s}}^T(s, b)^T}{\sqrt{1+p^2+q^2}} - i \frac{-\hat{\mathbf{n}}^T(s, b)^T}{\sqrt{1+p^2+q^2}}. \quad (9)$$

The intensity gradient  $\nabla i$  can thus be written in a compact form as

$$\nabla i = \frac{\mathbf{H}}{\sqrt{1+p^2+q^2}}(-\hat{\mathbf{s}} + i\hat{\mathbf{n}}) = \tilde{\mathbf{H}}(-\hat{\mathbf{s}} + i\hat{\mathbf{n}}), \quad (10)$$

where  $\tilde{\mathbf{H}} = \mathbf{H}/\sqrt{1+p^2+q^2}$  is the local surface Hessian scaled by the denominator term.

We now consider two images of the surface, one illuminated with light  $\mathbf{s}$  and the other with light  $\mathbf{t}$ . Denoting the intensity observed under the first and second light  $i(\mathbf{s})$  and  $i(\mathbf{t})$ , respectively, the respective gradients are

$$\nabla i(\mathbf{s}) = \tilde{\mathbf{H}}(-\hat{\mathbf{s}} + i(\mathbf{s})\hat{\mathbf{n}}), \quad \nabla i(\mathbf{t}) = \tilde{\mathbf{H}}(-\hat{\mathbf{t}} + i(\mathbf{t})\hat{\mathbf{n}}). \quad (11)$$

Making the directional derivative of the first image with respect to direction  $\hat{\mathbf{t}}$  and of the second image with respect to direction  $\hat{\mathbf{s}}$  gives

$$i_{\hat{\mathbf{t}}}(\mathbf{s}) = -\hat{\mathbf{t}}^T \tilde{\mathbf{H}}\hat{\mathbf{s}} + i(\mathbf{s})\hat{\mathbf{t}}^T \tilde{\mathbf{H}}\hat{\mathbf{n}}, \quad i_{\hat{\mathbf{s}}}(\mathbf{t}) = -\hat{\mathbf{s}}^T \tilde{\mathbf{H}}\hat{\mathbf{t}} + i(\mathbf{t})\hat{\mathbf{s}}^T \tilde{\mathbf{H}}\hat{\mathbf{n}}. \quad (12)$$

The first terms  $(-\hat{\mathbf{t}}^T \tilde{\mathbf{H}}\hat{\mathbf{s}}$  and  $-\hat{\mathbf{s}}^T \tilde{\mathbf{H}}\hat{\mathbf{t}})$  are equal for both images because the Hessian is symmetric due to normal field integrability. The second terms are generally different. If the result of filtering is required to give a virtually identical image, the second terms must be guaranteed to be small compared with the first terms, i.e.

$$|\hat{\mathbf{s}}^T \tilde{\mathbf{H}}\hat{\mathbf{t}}| \gg \max(|i(\mathbf{t})\hat{\mathbf{s}}^T \tilde{\mathbf{H}}\hat{\mathbf{n}}|, |i(\mathbf{s})\hat{\mathbf{t}}^T \tilde{\mathbf{H}}\hat{\mathbf{n}}|). \quad (13)$$

The general necessary conditions for these inequalities to hold are the following:

- a) **Shallow relief.** The projected normal should be of small length ( $\|\hat{\mathbf{n}}\| \ll 1$ ) which scales down the magnitude of the terms required to be small in Eq. (13).
- b) **Non-vertical illumination.** This is an obvious necessity as if, say,  $\mathbf{s}$  is vertical then  $\hat{\mathbf{s}} = \mathbf{0}$  and the symmetric term  $\hat{\mathbf{s}}^T \tilde{\mathbf{H}}\hat{\mathbf{t}}$  would vanish.

As stated, these conditions are necessary but they are not sufficient because  $\hat{\mathbf{s}}^T \tilde{\mathbf{H}}\hat{\mathbf{t}}$  can still vanish in special configurations of the two lights and the Hessian matrix  $\tilde{\mathbf{H}}$ . As an example, consider the case when the Hessian is a scaled identity matrix and the two projected light directions are perpendicular to each other. In such case, the symmetric term  $\hat{\mathbf{s}}^T \tilde{\mathbf{H}}\hat{\mathbf{t}}$  is zero. Whether this is a problem or not depends on the geometry of the surface imaged. The magnitude of the second terms is constrained by the two conditions, and thus even several points within the image in which the first term has large magnitude should be sufficient to make the images virtually identical.

Figure 2 illustrates the theory. The surface is a synthetically generated 3D texture of shallow relief. The mean normal inclination from the vertical direction is  $11^\circ$ . It is illuminated by a light source of  $40^\circ$  slant (see Fig. 1(b) for description of slant and tilt). As shown, filtering each image by a different derivative filter brings the two images very close together.

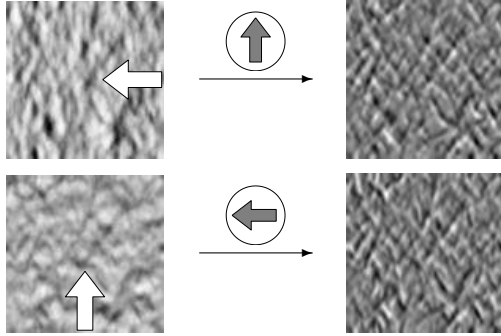


Figure 2: Results of directional derivative filtering for synthetically generated 3D texture. The surface is illuminated by lights with tilt  $0^\circ$  and  $-90^\circ$ , respectively (illumination direction is indicated by white arrows). Making derivatives in reciprocal directions (grey arrows) results in images which are very similar.

### 2.3 Exact fit: spherical patches

In the previous Section, conditions were identified which are necessary for making the method applicable to a general surface geometry. In this Section, we show that for a very specific case of *spherical* surface patches, there exists a pair of directional derivative filters which makes two images taken under different illumination conditions *equal*. It turns out that for a spherical surface our approach can output *exactly* the same images for *any* two illuminations (i.e. even when they are close to the camera axis), provided that none of them is perpendicular to the camera axis. This can be shown as follows. For a sphere of radius  $R$ , the height map is

$$z = \sqrt{R^2 - x^2 - y^2}, \quad (14)$$

and the normals are clearly

$$\mathbf{n} = \frac{1}{R} \begin{pmatrix} x \\ y \\ z \end{pmatrix}. \quad (15)$$

This means that the intensity observed while illuminating the sphere by a light source  $\mathbf{s}$  is

$$i = \frac{1}{R} \mathbf{s}^T(x, y, z)^T \quad (16)$$

and the image derivatives are then

$$i_x = \frac{1}{R}(s_1 + s_3 p), \quad i_y = \frac{1}{R}(s_2 + s_3 q), \quad (17)$$

and the image gradient is now (cf. Eq. (10))

$$\nabla i = \frac{1}{R} (\hat{\mathbf{s}} + s_3 \mathbf{g}). \quad (18)$$

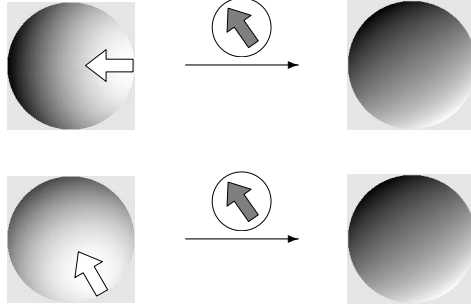


Figure 3: The result of directional derivative filtering of an illuminated spherical cap. The cap is illuminated with lights of tilts  $0^\circ$  and  $-60^\circ$ , respectively (indicated by white arrows). By filtering the images with directional derivatives as explained in Section 2.3 (grey arrows), the two images can be made identical. The slants used were  $40^\circ$  and  $25^\circ$  (top and bottom image, respectively) and the cap angular radius was  $45^\circ$ .

Now, considering the sphere illuminated from two different light directions  $\mathbf{s}$  and  $\mathbf{t}$ , and taking the directional derivatives w.r.t. two general directions  $\mathbf{u}$  and  $\mathbf{v}$ , gives

$$\boxed{i_{\mathbf{u}}(\mathbf{s}) = \frac{1}{R} \mathbf{u}^T (\hat{\mathbf{s}} + s_3 \mathbf{g}) \quad i_{\mathbf{v}}(\mathbf{t}) = \frac{1}{R} \mathbf{v}^T (\hat{\mathbf{t}} + t_3 \mathbf{g})} \quad (19)$$

As the surface gradient  $\mathbf{g}$  varies along the surface, the only way to make the second terms  $s_3 \mathbf{u}^T \mathbf{g}$  and  $t_3 \mathbf{v}^T \mathbf{g}$  equal is to take (provided that both  $s_3$  and  $t_3$  are non-zero)

$$\mathbf{u} = t_3 \mathbf{d}, \quad \mathbf{v} = s_3 \mathbf{d}, \quad (20)$$

with (so far) arbitrary vector  $\mathbf{d} \in \mathbb{R}^2$ . Requiring the first terms to be equal as well leaves us with a linear homogeneous equation for  $\mathbf{d}$ :

$$t_3 \mathbf{d}^T \hat{\mathbf{s}} = s_3 \mathbf{d}^T \hat{\mathbf{t}} \iff \mathbf{d}^T (t_3 \hat{\mathbf{s}} - s_3 \hat{\mathbf{t}}) = 0, \quad (21)$$

which constrains  $\mathbf{d}$  always unless  $t_3 \hat{\mathbf{s}} - s_3 \hat{\mathbf{t}}$  vanishes. But the case when this vanishes corresponds to the case when the lights  $\mathbf{s}$  and  $\mathbf{t}$  are the same (possibly differing in magnitude). In such circumstances, any  $\mathbf{d}$  will make images look the same.

We thus showed that if the scene imaged consists of spherical patches then (not depending on patch radii) two images of such scene taken under two different illumination directions can be made identical by filtering both images with two directional derivative filters. The filters differ only in magnitude but not in orientation for this case. The just stated is true for any two light directions used unless one of the light sources is perpendicular to the optical axis. This result is demonstrated in Fig. 3 which shows a sphere illuminated from two general directions (left image pair) and the images filtered by directional derivative vectors as explained in this Section (right image pair). The filtered images exhibit exact agreement.

## 2.4 Methods

The theory presented in the previous Sections showed that the directional derivative filters to be applied are dependent on the light directions used for acquiring the images. In practise, the light directions are not known, and there is a need to *estimate* the directional derivative vectors for the two filters.

Within this introductory paper, we deal only with the case of spatially registered images. Having images  $i(\mathbf{s})$  and  $i(\mathbf{t})$  taken under two unknown illumination directions  $\mathbf{s}$  and  $\mathbf{t}$ , we compute the image gradients  $\nabla i(\mathbf{s})$  and  $\nabla i(\mathbf{t})$  and then solve for two vectors  $\mathbf{u}, \mathbf{v} \in \mathbb{R}^2$  such that

$$\mathbf{u}^T \nabla i(\mathbf{s}) = \mathbf{v}^T \nabla i(\mathbf{t}) \quad (22)$$

holds. Every pixel produces one equation (22), and the resulting system is clearly overdetermined. The system is solved in a least-square sense using SVD. The solution gives the directional derivative vectors  $\mathbf{u}$  and  $\mathbf{v}$ .

## 3 Experiment

We performed basic tests of the theory on publicly available PhoTex database [1] of 3D textures. The aim of the experiment was to judge whether our filtering approach is feasible for making the images taken under different illumination conditions look similar. We tested the approach on 20 textures and here we present representative examples (see Figure 4).

For each class, we selected a reference image which was taken under certain illumination condition (slant was  $45^\circ$  and tilt was  $30^\circ$ ; this choice was somewhat arbitrary). Then we took images of the same surface taken under different illuminations, applied the SVD-based method for estimating the directional derivative vectors as described in Section 2.4, and displayed the filtered images. Overall, the match of the filtered images is very good, as can be seen from Fig. 4. The match is especially impressive for cases when the perceptual difference of the raw images is high (class AAJ).

Within this paper, we did not address the case of spatially unregistered images although the basic principle presented enables to extend the method for that case. In our preliminary experiments on texture classification, we used a more complex algorithm which searches for a pair of directional derivative vectors which make image *statistics* of the filtered images as close as possible. The full implementation of this approach is the topic for future work.

## 4 Conclusion

This paper presented a new theory necessary to develop novel algorithms for object and texture recognition, classification and retrieval under varying illumination. To the best of our knowledge, it is the only method which is able to make two images captured under two illumination directions look virtually identical, while not requiring large training datasets. The price paid for these advantages is a rather limiting set of assumptions, and the fact that if illumination directions are not known, the directional derivative vectors have to be estimated. As shown, in case of spatially registered images this involves a simple



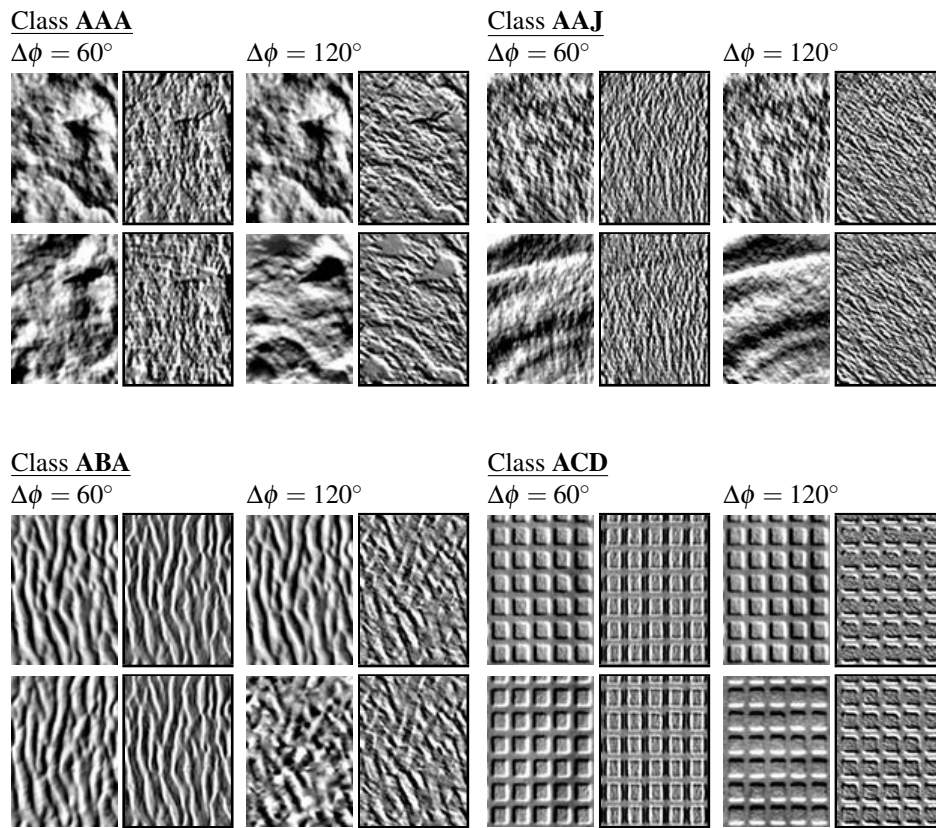


Figure 4: Results for four classes in the PhoTex database (AAA, AAJ, ABA and ACD). The number  $\Delta\phi$  above each quadruple of images is the difference in illumination tilts used for capturing the two images. The layout of the quadruple is the same as in Fig. 2 (the non-framed images are therefore the raw ones, while the framed images are the filtered ones). Overall, the filtering brings the raw images very close together, despite large differences in illumination conditions. Note that although the raw images for Class AAJ, for example, are perceptually quite different, the match of the filtered images is excellent. The slants used for acquiring the raw images were  $45^\circ$  and  $60^\circ$  (top and bottom ones, respectively.) The raw images are normalised for display purposes.

algorithm for solving an overdetermined system of linear homogeneous equations in four unknowns.

Interestingly, it was shown that when the imaged scene consists of spherical surface patches (with arbitrary, varying radii), the two images can be made *identical* under any two illumination directions, provided that none of them is perpendicular to the viewing axis.

## Acknowledgements

O. Drbohlav is supported by the Marie Curie Intra-European Fellowship No. 506053 (PhoCal) within the 6th European Community Framework Programme.

## References

- [1] PhoTex database. Texture lab, Heriot-Watt University, Edinburgh, UK. Available on-line at <http://www.cee.hw.ac.uk/texturelab/database/Photex/>.
- [2] P. N. Belhumeur and D. J. Kriegman. What is the set of images of an object under all possible lighting conditions. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pages 270–277, 1996.
- [3] H. Chen, P. Belhumeur, and D. Jacobs. In search of illumination invariants. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, volume 2, pages 254–261, 2000.
- [4] A. S. Georghiades, D. J. Kriegman, and P. N. Belhumeur. Illumination cones for recognition under variable lighting: Faces. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pages 52–59, 1998.
- [5] B. K. P. Horn and M. J. Brooks. *Shape from shading*. The MIT Press, 1989.
- [6] D. W. Jacobs, P. N. Belhumeur, and R. Basri. Comparing images under variable illumination. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pages 610–617, 1998.
- [7] J. H. Lambert. *Photometria sive de mensura et gradibus luminis, colorum et umbra*. Augustae Vindelicorum, Basel, 1760.
- [8] T. Leung and J. Malik. Representing and recognizing the visual appearance of materials using three-dimensional textons. *International Journal of Computer Vision*, 43(1):29–44, 2001.
- [9] M. Osadchy, M. Lindenbaum, and D. Jacobs. Whitening for photometric comparison of smooth surfaces under varying illumination. In *Proc. European Conference on Computer Vision*, pages 217–228, 2004.
- [10] M. Varma and A. Zisserman. Classifying materials from images: to cluster or not to cluster? In *Texture 2002 : The 2nd International Workshop on Texture Analysis and Synthesis*, pages 139–144, 2002.