Surface Dependent Representations for Illumination Insensitive Image Comparison

Margarita Osadchy, *Member*, *IEEE*, David W. Jacobs, *Member*, *IEEE*, and
Michael Lindenbaum, *Member*, *IEEE*

Abstract—We consider the problem of matching images to tell whether they come from the same scene viewed under different lighting conditions. We show that the surface characteristics determine the type of image comparison method that should be used. Previous work has shown the effectiveness of comparing the image gradient direction for surfaces with material properties that change rapidly in one direction. We show analytically that two other widely used methods, normalized correlation of small windows and comparison of multiscale oriented filters, essentially compute the same thing. Then, we show that for surfaces whose properties change more slowly, comparison of the output of whitening filters is most effective. This suggests that a combination of these strategies should be employed to compare general objects. We discuss indications that Gabor jets use such a mixed strategy effectively, and we propose a new mixed strategy. We validate our results on synthetic and real images.

Index Terms—Image comparison, Illumination, Gaussian random surface,Whitening.

---

1 INTRODUCTION

IMAGE comparison is central to computer vision tasks such as tracking and object recognition. Lighting variation significantly affects surface appearance and makes image comparison difficult. For this reason, many approaches have been suggested for building representations of images that are insensitive to lighting variation. However, the relative advantages of these different representations are often not clear. In this paper, we show that two different classes of representations are appropriate for two quite different classes of scenes. By understanding these two situations, we are able to suggest new illumination-insensitive representations, including a hybrid representation that can be effective for both classes of objects.

We will discuss two classes of surfaces that we call *nonisotropic* and *isotropic*. These terms refer to local properties of a scene. By a nonisotropic surface, we will mean a surface whose properties change rapidly in one direction and slowly (or not at all) in another. This includes, for example, surfaces that have large curvature in just one direction, such as a cylinder, or surfaces with discontinuities in orientation or albedo. In contrast, an isotropic surface will be one in which variation is more or less similar in both directions. The isotropic surfaces we will discuss in this paper are smooth, i.e., their surface properties vary slowly. An isotropic surface with identical, abrupt changes in both directions is less common.

- *M. Osadchy is with the Computer Science Department, University of Haifa, Mount Carmel, Haifa 31905, Israel. E-mail: rita@cs.haifa.ac.il.*
- *D.W. Jacobs is with the Department of Computer Science and UMIACS, A.V. Williams Building, University of Maryland, College Park, MD 20742. E-mail:djacobs@cs.umd.edu.*
- *M. Lindenbaum is with the Computer Science Department, Technion, Haifa 32000, Israel. E-mail: mic@cs.technion.ac.il.*

Chen et al. [7] have already demonstrated that the direction of the image gradient is relatively insensitive to lighting variations for nonisotropic surfaces, because it is approximately aligned with the direction of greatest surface change over a range of lighting conditions. This representation has been suggested by many other authors, (e.g., ([21], [41], [7], [8], [11]). We will show that a number of other image representations are equivalent to, or closely related to the image gradient dirction and, therefore, also most suitable for nonisotropic surfaces (some of this material was first presented in [35]). Specifically, we show that methods based on the image gradient direction are equivalent to normalized correlation in the limit, when the correlation window is small. We also demonstrate that comparison of the gradient direction is equivalent to comparison using the normalized outputs of oriented derivative filters (as done, for example, by [40]). We will briefly show a weaker relationship between gradient direction comparisons and those using histogram equalization or mutual information. These relationships help us to understand when these representations are appropriate and also to predict that these representations will lead to similar matching performance.

Isotropic surfaces offer a different challenge, because the direction of the gradient depends more on lighting than on surface anisotropies (an earlier version of our work on these surfaces was presented in [34]). We begin by providing a simplified, analytic model of isotropic surfaces. We use this to consider image properties produced by isotropic, Lambertian objects with a single point source. While this model is simplified, we believe it captures some of the most significant effects of lighting and provides insight into those aspects of image comparison, which we can test empirically.

Specifically, we use results from signal detection theory to show that, for this model, we should apply a whitening filter to the images before comparison. Whitening does not produce a representation that is less sensitive to lighting variation, but rather provides a more statistically valid image comparison by allowing us to compare quantities that are more independent. Our analysis demonstrates that

standard filters, such as the Laplacian of the Gaussian, provide effective representations for isotropic surfaces because they tend to whiten images. We also show that, for a specific domain, we can learn a whitening filter that is more effective than generic whitening filters.

Given that two quite different comparison methods succeed in different domains, it is natural to consider combining them into one, complete method. In Section 5, we begin this endeavor. We present some simple approaches to combining whitening and gradient direction methods. Perhaps the most elegant is using a jet of oriented, second derivative operators, which combines gradient direction effects that are due to normalization with whitening effects from the second derivative operator. We analyze the effectiveness of this method. We also show that Gabor jets can be understood also as a simple method of combining the two. These combined methods are all somewhat simple and there is clearly room for more effective methods to be developed in future work.

Our main contribution is a better understanding of how the properties of image surfaces influence the effectiveness of different approaches to image comparison. This helps us to understand when existing representations will be effective and to see close connections between some of these methods that have not previously been pointed out. This can provide insight for practitioners who are seeking the best representation for their domain and also suggests new approaches to lighting insensitive image matching.

2 BACKGROUND

In this section, we will begin by discussing some prominent effects of variation in lighting and camera parameters. We can divide these into two sets: The first and simplest are additive, multiplicative, or other monotonic changes in intensity, while the second are the more compex variations caused by changes in lighting direction. We will then review past approaches to lighting insensitive comparisons. Most of these are geared to handling the first set of variations. In addition, the direction of image gradient has been shown to have some insensitivity to changes in lighting direction.

Image intensities arise from the interaction of camera properties, light, shape and materials. In image matching, we seek to compare images in a way that is sensitive to differences in shape and material, which are usually related to object identity, and as insensitive as possible to lighting effects and camera characteristics, which are independent of this identity.

Changes in camera parameters affect image intensities significantly, but largely independently of whatever scene properties produced them. Some main effects include additive (offset) or multiplicative (gain) changes in intensity. Also, gamma correction can alter intensities more drastically, but still monotonically.

Changes in light intensity also have a multiplicative effect on image intensities. This is also independent of scene geometry. Changes in lighting direction, however, can interact with scene geometry in complex ways. In the simplest case of complex Lambertian objects, a change in lighting direction causes some scene points to appear brighter as the light strikes them more directly, while other scene points become dimmer as the light becomes more oblique. This means that the variation in intensities can be far from monotonic. Locally, changes in lighting direction can also affect the direction of the image gradient. As an example, consider a white, Lambertian sphere, illuminated by a directional source of light. The point on the sphere facing the light will be brightest, if it is visible in the scene. The image gradient at other points on the sphere will point in the direction of this brightest point and, so, will be completely determined by lighting direction.

Chen et al. [7] show that no image properties are invariant to the effects of lighting variation on 3D scenes, even for the case of Lambertian objects with no shadows. However, they also show through a statistical analysis that gradient direction can be insensitive to variations in lighting direction. Essentially, they demonstrate that when a surface is nonisotropic, greater curvature or variation in material properties in one direction tends to cause greater variation in image intensities in that direction. To discuss this, it will be useful to distinguish the direction of the image gradient from the polarity-free direction, i.e., the direction modulo π. For example, either side of a roof may appear darker, depending on the direction of the light. Therefore, for a roof shape, the polarity of the image gradient is quite sensitive to the lighting direction. However, for most lighting directions, the image gradient points in a direction perpendicular to the roof ridge. We describe this situation by saying that the polarity-free direction of the image gradient is insensitive to lighting direction.

We will give short shrift to other lighting effects, such as shadows, interreflections, and specularities. These effects can be extremely complex; for example, a mirror can produce any image and a shadow can have any shape. Little work on image matching has explicitly accounted for these effects. However, we note that there is a good deal of work on using cues such as color or polarized light to identify specularities (see Oren and Nayar [33] for a survey), and this could be of value in image matching. We also note that one effect of cast shadows is to make some regions of an image darker, and that methods of image comparison based on local descriptions, normalized for intensity, may be insensitive to such effects.

Since lighting and camera variation often have an additive and multiplicative effect on image intensities, it is natural to cope with illumination by normalizing the mean and variance of image intensities prior to comparison. *Normalized cross-correlation* follows this normalization with correlation, and is a standard way to manage the effects of lighting change (e.g., [22]). Brunelli and Poggio [6], for example, use normalized correlation for face recognition, and Nayar and Bolle [32] normalize using ratios for lighting invariant recognition of planar objects. Normalization may be performed on the entire image, or it may be applied to a small window. Small windows are used for recognition by, for example, Brunelli and Poggio, and are commonly used for image matching in stereo or tracking (see, e.g., Trucco and Verri [46]). Inspired by the retinex lightness constancy algorithm [29], others have also locally normalized images by dividing intensities by a low frequency signal that estimates slowly varying lighting (e.g., [19]). However, it has not been claimed that normalization can compensate for changes in lighting direction for images of complex, 3D scenes.

A related approach is to transform image intensities to make the image histogram constant (histogram equalization, (e.g., [17]) or to give it a specific profile (histogram specification, e.g., [38]). Kittler et al. [26] compare a number

of normalization methods, including histogram equalization, for face recognition. These methods are invariant to any monotonic change in intensities throughout the image, though not to changes in lighting direction.

Many approaches apply normalization to the output of multiscale oriented filters instead of raw pixel intensities (e.g., [30], [40], [18], [28], [42]). A vector of Gabor filters applied to a common location has been called a *jet* (e.g., [30]), and we will use that terminology for other filters as well. By using filters that integrate to zero, such as derivative of Gaussians or Gabor filters, these jets become invariant to additive changes in the image. Normalizing the total magnitude of all filter responses produces invariance to multiplicative image changes. This invariance to offset and gain is often one motivatation for the use of multiscale oriented filters (e.g., [49], [50], [10], [5], [24]).

Methods based on image gradients have been widely applied to images of nonisotropic surfaces. Edge detection is a classic method appropriate when a scene contains discontinuities in shape or albedo since this generally leads to discontinuities in the image. However, the magnitude of image gradients can be sensitive to lighting, making edges sensitive to lighting direction and intensity (see, e.g., [43]).

Therefore, a number of authors have proposed using the direction of the image gradient alone for image comparison ([3], [4], [21], [41], [12], [7], [8], [11]). In some cases, the motivation for this has been that the gradient direction is also invariant to changes in offset and gain in the image (e.g., [8], [11]) and indeed to any monotonic change in image intensity ([41]). As mentioned above, Chen et al. [7] provide a statistical analysis to show that the direction of gradient is also insensitive to changes in lighting direction for nonisotropic scenes.

There are many possible ways to compare images using gradient direction. Perhaps the simplest is to compute the sum of squares of the differences between two gradient direction images. We will call this comparison "DIRP" and the polarity-free version "DIR," for short. Chen et al. also suggest a somewhat superior method, which relies on the distribution of direction differences and uses the statistics of image pairs that come from the same scene, under different lighting conditions.

Finally, we note that there has been much work on lighting insensitive object recognition that uses cues outside the scope of this paper, such as 3D knowledge of the scene (e.g., [44], [16], [2]), color (e.g., [45], [15]), and multiple images (e.g., [31], [51]).

3   IMAGE COMPARISON OF SMOOTH ISOTROPIC SURFACES

As described above, the image gradient directions associated with nonisotropic surfaces are relatively insensitive to lighting variation. Smooth, isotropic surfaces offer a much different challenge, with no known quasi-invariants. We approach these surfaces by considering one image as a model and any other image of the same object, associated with a different illumination, as a corrupted version of it. The difference between the model and the corrupted version is considered an error. Then, we focus not on insensitivity to lighting but on finding good ways to measure the magnitude of its effect.

Our first observation is that the common sum-of-squared-differences (SSD) measure is not adequate. SSD is justified by the (usually implicit) assumption that the difference between a model and its corrupted version is white Gaussian noise. In this case, matching an image to the model that minimizes the SSD is a maximum likelihood (ML) decision. This is clearly not the case for the difference between two smooth images, where the gray-level values, both in the images themselves and in the difference image, are highly correlated.

In this section, inspired by the work of [27], we use a simple model of smooth surfaces as Gaussian random fields to help us understand these correlations. For these surfaces, illumination changes lead to difference images that are colored Gaussian noise. Then, an (ML) decision can be made by minimizing the Mahalanobis distance between the images [9, Chapter 2]. Equivalently, we can decorrelate the difference image by a *whitening* operator before evaluating its energy (the whitened SSD) [47, Chapter 4.3]), which turns out to be preferable for our purposes.

3.1   A Covariance Model of Smooth Surfaces

To whiten a surface's images, its covariance structure must be known. We shall first derive some properties of the covariance associated with smooth surfaces. To that end, we model smooth surfaces as locally *approximately planar*, with the surface specified as a random height function relative to the tangent plane. This way, both the surface height and the associated normals make small random perturbations about a common value. Such a surface will be smooth when nearby surface heights and normals are highly correlated. In the following derivation, we use the Gaussian random field and the Lambertian imaging model as in [27], which aims at estimating the illumination direction from texture images. This goal is quite different from ours; we aim to characterize the covariance of the image for whitening.

3.1.1   A Covariance Model for Images of an Approximately Planar Lambertian Surface

Locally, a smooth surface is modeled as approximately planar, with small perturbations relative to the tangent plane. A local coordinate system $(\mathbf{e}_x, \mathbf{e}_y, \mathbf{e}_z)$ is specified so that the $\mathbf{e}_z$ is normal to the plane and $\mathbf{e}_x, \mathbf{e}_y$ lie on it in arbitrary directions. The perturbation is specified as a (smooth) function $h(\bar{x}) = h(x, y)$, which is an instance of an isotropic Gaussian random field [1]. Specifically, we assume that the random field is homogeneous (stationary), associated with zero mean, $E[h(x, y)] = 0$, and with Gaussian autocorrelation

$$C_S(\bar{x}_i, \bar{x}_j) = \sigma_S^2 e^{-\|\bar{x}_i - \bar{x}_j\|^2/\beta^2}.$$

The surface is modeled as Lambertian, with uniform albedo ρ; (see [48] for a generalized Gaussian model with varying albedo).

The imaging model is as follows: The surface is illuminated by a distant light source from the direction $cos\theta(cos\phi \mathbf{e}_x + sin\phi \mathbf{e}_y) + sin\theta \mathbf{e}_z$. That is, the lighting is characterized by its elevation θ relative to the tangent plane and by its azimuth ϕ relative to $\mathbf{e}_x$. Let I_0 be the normal irradiance of the source at the surface. The view direction is normal to the surface (that is, $\mathbf{e}_z$). We assume that θ is large enough and the object's relief is shallow enough so that the effect of shadows and interreflections is negligible. Then, the irradiance at the point $\bar{x} = (x, y)$ is

$$I(x,y) = \rho I_0 \frac{sin\theta - cos\theta(h_x cos\phi + h_y sin\phi)}{(1 + h_x^2 + h_y^2)^{1/2}}, \qquad (1)$$

where h_x, h_y are the partial derivatives $\frac{\partial h(x,y)}{\partial x}, \frac{\partial h(x,y)}{\partial y}$, respectively. For shallow smooth surface, where $h_x, h_y << 1$,

$$I(x,y) \approx \rho I_0 \big[sin\theta - cos\theta(h_x cos\phi + h_y sin\phi) \big]. \qquad (2)$$

The resulting irradiance is also a random field and with the approximation (2) it is Gaussian as well.

Proposition 1 (Lambertian Image as Gaussian Random Field). *Let $I(x)$ be the irradiance (image) associated with the Gaussian random surface and the imaging model, described above. Then, $I(x)$ is (approximately) a Gaussian random field, with expected value, variance, and autocorrelation given by:*

$$E[I(x)] = \rho I_0 sin\theta$$

$$\sigma_I^2 = \rho^2 I_0^2 cos^2\theta \, \frac{2\sigma_S^2}{\beta^2}$$

$$C_I(\bar{x}_i, \bar{x}_j) = \rho^2 I_0^2 cos^2\theta \, \frac{2C_S(\bar{x}_i, \bar{x}_j)}{\beta^2} \qquad (3)$$

$$\left[1 + \frac{2}{\beta^2} \|\bar{x}_i - \bar{x}_j\|^2 cos^2(\phi - arg(\bar{x}_i - \bar{x}_j)) \right],$$

where $\bar{x}_i, \bar{x}_j$ are two points where the random field is specified, and $arg(\bar{x})$ indicates the angle between $\bar{x}$ and $\mathbf{e}_x$.

The proof is simple and relies on basic properties of random fields; see the Appendix. Interestingly, the image associated with an isotropic surface is not isotropic; the degree of anisotropy is inversely proportional to β, which indicates the smoothness of the surface. For rough surfaces the image is highly anisotropic and [27] use this to determine the lighting direction, based on different expressions than the ones we derive. However, we also have:

Proposition 2 (Isotropy and multiplicative covariance properties). *For smooth surfaces and short distances, the image autocorrelation is approximately isotropic and constant up to an illumination dependent multiplicative factor.*

Proof. For smooth surfaces and short distances, $\beta >> \|\bar{x}_i - \bar{x}_j\|$, implying that the second term in the autocorrelation is negligible. □

This property is important for whitening because, as we shall see below, it implies that the whitening procedure may be carried out efficiently. To demonstrate the multiplicative property, we conducted the following simple experiment: We took a high resolution image of a real, approximately Lambertian sphere, illuminated by a point source. We divided the image into 50×50 pixel patches and calculated the covariance in every patch, for all pixel pairs contained in 7×7 neighborhoods. Fig. 1 shows the estimated covariance as a function of the distance. Each curve represents a different patch. The plots confirm that covariance in different patches differs by a multiplicative factor.

3.1.2 A Covariance Model of a Smooth Surface

We assume that a smooth surface locally (but not necessarily globally) behaves as an approximately planar one. When we move across the smooth surface, the tangent plane surface normal changes, but the (distant) light source direction stays the same. For a Lambertian surface, this is
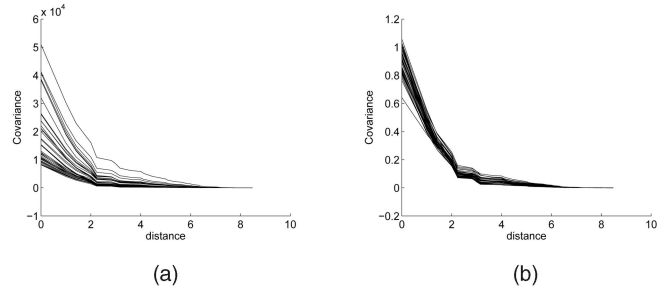

Fig. 1. Covariance estimates for different patches of a real object: (a) Nonnormalized covariances differ by multiplicative factors and (b) covariances which are normalized by the variance are almost the same for all angles.

the same as keeping the surface normals constant but changing the light source direction. According to Proposition 2, this doesn't change the structure of the autocorrelation function, but only changes the scale factor.

Note that the covariance varies when the viewing direction does not coincide with the normal. Therefore, we do not expect the covariance to have the same structure for regions where the object's normal deviates significantly from the camera direction. For smooth objects, the fraction of such image regions is small. Moreover, such regions are often close to the object boundary and therefore not isotropic.

3.2 The Whitening Technique

Consider two images specified by the simple imaging model described above, and associated with the same surface and with two illumination sources $\mathbf{s}_1, \mathbf{s}_2$. Note that by linearity, the difference between the two images is just another "image" obtained by taking the inner product between the normals of the same surface and an imaginary illumination direction $\mathbf{s}_1 - \mathbf{s}_2$. Note that this image may take negative values. Therefore, the difference image may also be modeled as a Gaussian random field or colored Gaussian noise.

In principle, a colored Gaussian noise signal (or image) I with a covariance matrix $C = E[II^T]$ can be decorrelated by multiplying it by a matrix W, the rows of which are the scaled eigenvectors of C, $\frac{1}{\sqrt{\lambda_i}}\mathbf{e}_i$. ($\mathbf{e}_i$ is the ith eigenvector of C and λ_i is the corresponding eigenvalue.) That is, $E[(WI)(WI)^T]$ is the identity matrix. Calculating the covariance matrix and its eigenstructure is, however, computationally prohibitive and requires an unreasonable number of examples. Fortunately, if the signal's dependency is local and stationary, whitening may be carried out much more easily by modeling the signal as an autoregressive (AR) process [23, Chapter 6]. An 1D sequence I_n is called an AR process of order p if it can be generated as the output of the recursive causal linear system

$$I_n = \sum_{k=1}^{p} a_k I_{n-k} + \varepsilon_n, \forall n, \qquad (4)$$

where ε_n is white noise. The term $\bar{I}_n = \sum_{k=1}^{p} a_k I_{n-k}$ in (4) is the best linear mean squared (MS) predictor of I_n, based on the previous p samples. Given a random sequence (with possible dependencies), an AR model can be fitted using SVD to estimate the overdetermined parameters a_k that
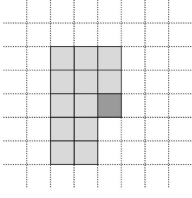
Fig. 2. A typical 2D "causal" neighborhood used in our experiments. The darkest pixel is the one where the gray-level is estimated from the gray-levels in the other dark pixels.

minimize the empirical MS prediction error $\sum_n (I_n - \bar{I}_n)^2$. For Gaussian signals, the prediction error sequence $\varepsilon_n = I_n - \bar{I}_n$ is white, implying that a simple convolution with the filter $W' = (1, -a_1, \ldots, -a_p)$ is a whitening process for I.

To decorrelate the difference image we use a 2D AR model with a "causal" neighborhood model (see [23, Chapter 6]), as illustrated in Fig. 2. We use a "causal" neighborhood (as in the 1D case) because a noncausal neighborhood would not lead to decorrelation [23].

Note that scaling all the gray levels by the same factor would give a correlation function that is the same up to a multiplicative constant. This is essentially what happens when the angle between the tangent plane normal and the illumination direction changes. Fortunately, this does not change either the fitted AR coefficients or the resulting whitening filter, implying that it can be space invariant.

The whitening filter depends on the image statistics. Intuitively, for smoother images, the correlation is larger and decorrelating it requires a wider filter. For images that are not so smooth, the decorrelation may be done over a small range and the filter looks very much like the Laplacian, known to have some whitening effect [39]. This is not surprising because a rough image may be characterized as a moving average over a noise image with a small filter, and such averaging may be inverted with a Laplacian. Therefore, for rougher images, we do not expect performance better than that of an alternative procedure using the Laplacian. As we shall see later, the performance difference is significant for smooth objects.

The whitening method works in principle for any surface. It is more attractive for smooth and isotropic surfaces where illumination insensitive descriptions are not known. It is also more accurate, in this case, because the multiplicative covariance property (Proposition 2) depends on β being large.

Our theory is limited to the case of a class of surfaces drawn from the same random process. In any real application, we expect to deal with surfaces that have varying degrees of smoothness; even a single surface may have smoother and rougher parts. This means the optimal whitening filter will depend on the specific surface, which cannot be known when we are comparing images. We deal with this pragmatically by learning a whitening filter based on images of a set of similar surfaces. This is appropriate when we are matching images within a specific domain. Theoretically, we cannot expect such a filter to perfectly whiten all images of interest, but we show experimentally that it leads to an effective method of image comparison for smooth, isotropic surfaces.

3.3 Whitening Images from Different Objects

Discriminating between smooth objects using their images is difficult not only because of illumination induced changes but also because of the smoothness itself. This smoothness implies that if images of two different smooth objects are similar at a single pixel, they are likely to be similar in a substantial neighborhood about that pixel. For this reason, it is more likely that images of different smooth objects will be more similar than images of different rough objects. Whitening, as a high pass filter, makes images of smooth objects less correlated and, hence, more distinctive.

Decorrelating the models is important in the design of digital communication signals. It is formally proven that for the lowest error rate, the correlation between each pair of signals should be as low as possible. For two signals, the lowest correlation is -1 and the optimal signals are antipodal. For a large number of signals, such correlations between all signal pairs are not possible and the best we can get are decorrelated signals [47, Chapter 4.2]; (see a related discussion in [25]).

Intuitively, the type of change caused by whitening may be understood as follows: Let S denote a 3D surface, with normals $\hat{N}_{i,j}$ and albedos $\rho_{i,j}$. Denote $N_{i,j} \equiv \hat{N}_{i,j}\rho_{i,j}$. Let $I_{1,i,j} = N_{i,j}^T s_1$ be one of its images, associated with illumination vector s_1. Let W be a whitening filter, represented discretely as a matrix with elements $W_{k,l}$; $-n \leq k, l \leq n$. Applying this filter to the image I_1, we get the output image $\mathcal{I}_1$:

$$\mathcal{I}_{1,i,j} = \sum_{k=-n}^{n} \sum_{l=-n}^{n} W_{k,l} I_{1,i-k,j-l}.$$

Define now a new surface, S, such that its scaled surface normals are

$$\mathcal{N}_{i,j} = \sum_{k=-n}^{n} \sum_{l=-n}^{n} W_{k,l} N_{i-k,j-l}.$$

By linearity, the whitened images may be considered as images of the "whitened" surface S. While the original normals are highly correlated, the whitened normals will be white noise, with randomized directions and sizes. The randomization of sizes is analogous to changing the smooth surface by splattering it with gray paint in random locations, which makes the surface visually distinctive from other surfaces. Of course, whitening does not make signals different; it just makes explicit the differences that already exist.

While a white noise image is fully correlated with itself, the correlation of such an image with any nonzero translation of it is zero. It is even more unlikely that two unrelated objects will be highly correlated after whitening. We have tested this observation empirically on images of smooth real objects (see Section 6.2 for a description). We took 91 pairs of images (normalized to unit energy), where every pair is associated with different objects and the same illumination, and computed their inner products and correlations (i.e., inner products after average substraction) before and after whitening. Fig. 3 shows the resulting
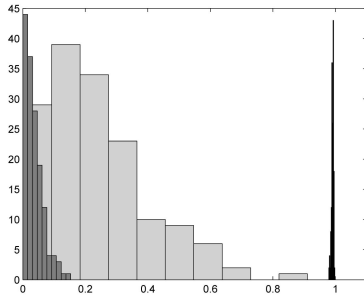
Fig. 3. Distribution of correlations between images of different objects before whitening (black on the right), after removing average (light gray), and after whitening (dark gray).

distributions. As expected, the images are highly correlated before whitening and are much less correlated after whitening. Note that simply removing the average does result in some decorrelation but is much less effective than the proposed whitening.

4 IMAGE COMPARISON OF NONISOTROPIC SURFACES

In this section, we focus on image comparison methods for nonisotropic surfaces. We will show that methods based on normalization are closely related to the direction of gradient. We will consider normalization of local windows of pixels or of the output of oriented filters that compute the derivative of a Gaussian. We will also make brief comments about comparisons based on histogram equalization and mutual information.

4.1 Normalized Correlation in Small Windows

First, we discuss normalized correlation between windows with linear intensity patterns. This analysis is only relevant for small windows.

Proposition 3. *For windows with intensity that is a linear function of image coordinates, normalized correlation computes the cosine of the difference in gradient direction.*

Proof. Normalized correlation starts with subtraction of the mean from each window. Assume I_1 and I_2 are zero mean windows with locally linear intensity. We can choose the x axis so that the y component of ∇I_1 is 0. With these assumptions, without loss of generality, $I_1 = ax, I_2 = bx + cy$. Then, the corresponding gradients are $a\hat{x}$ and the $b\hat{x} + c\hat{y}$, where $\hat{x}$ and $\hat{y}$ are unit vectors in the x and y directions. The angle between them, θ, satisfies

$$\cos\theta = \frac{ab}{a\sqrt{b^2 + c^2}} = \frac{b}{\sqrt{b^2 + c^2}}. \quad (5)$$

The correlation between I_1 and I_2 is:[1]

$$\int_{-1}^{1}\int_{-1}^{1}(abx^2 + acxy)dxdy = \frac{4}{3}ab.$$

1. The size of the integration interval is chosen arbitrarily because its magnitude will be canceled by normalization.

To compute the normalization factors corresponding to I_1 and I_2, we use

$$\|I_1\|^2 = \int_{-1}^{1}\int_{-1}^{1} a^2 x^2 dxdy = \frac{4}{3}a^2,$$

$$\|I_2\|^2 = \int_{-1}^{1}\int_{-1}^{1}(b^2x^2 + c^2y^2 + 2bcxy)dxdy = \frac{4}{3}(b^2 + c^2).$$

Then, the normalized correlation is

$$\frac{I_1 \cdot I_2}{\|I_1\|\|I_2\|} = \frac{4ab}{3}\sqrt{\frac{3}{4a^2}}\sqrt{\frac{3}{4(b^2 + c^2)}} = \frac{b}{\sqrt{b^2 + c^2}} = \cos\theta.$$

$\square$

This demonstrates that with small windows, normalized correlation is similar to DIRP because, like the square of the difference in angle, the cosine function is also monotonic in this difference and changes more rapidly as the difference increases (up to $\frac{\pi}{2}$).

4.2 Oriented Derivatives of Gaussians

Many authors have remarked that jets of multiscale, oriented filters provide some insensitivity to lighting variation. We now analyze a simple version of these approaches using a vector of oriented derivative of Gaussian filters at one scale. We show that comparing the output of these filters effectively computes the cosine of the difference in the direction of gradient of the two images. We call this method, which is sensitive to polarity, "DOGP." With a slight variation, it computes a similar quantity that is invariant to polarity, which we call "DOG." Let $D_\theta(\vec{x})$ denote the directional derivative of $G \star I$ in the direction θ, where G denotes a Gaussian filter, $\star$ denotes convolution and θ indicates the angle relative to the x axis. Define the vector of the output of these filters at $\vec{x}$: $D(\vec{x}) = (D_{\theta_1}(\vec{x}), D_{\theta_2}(\vec{x}), \ldots D_{\theta_k}(\vec{x}))$. Typically, $k = 8$ with equal spacing. We will also consider a common, polarity insensitive variation in which the absolute value of the filter output is used: $D_a(\vec{x}) = (|D_{\theta_1}(\vec{x})|, |D_{\theta_2}(\vec{x})|, \ldots |D_{\theta_k}(\vec{x})|)$. The resulting vectors are normalized prior to comparison. The simplest comparison is correlation, in which case we compute $\frac{D_1 \cdot D_2}{\|D_1\|\|D_2\|}$.

Suppose that the direction of the gradient at $\vec{x}$ is along the x axis (this will be without loss of generality when our analysis moves to the continuous domain) and the magnitude of the gradient is M_x. Then, $D_\theta(\vec{x}) = M_x \cos(\theta)$ and we have

$$D(\vec{x}) = M_x\left(\cos(0), \cos\left(\frac{2\pi}{k}\right), \ldots \cos\left(\frac{2(k-1)\pi}{k}\right)\right).$$

That is, $D(\vec{x})$ is a vector that discretely samples the cosine function, scaled by M_x. If we compute derivative of Gaussians at a point, $\vec{y}$, in another image, at which the direction of the gradient is α and its magnitude is M_y, we have

$$D(\vec{y}) = M_y\left(\cos(-\alpha), \cos\left(\frac{2\pi}{k} - \alpha\right), \ldots \cos\left(\frac{2(k-1)\pi}{k} - \alpha\right)\right).$$

TABLE 1
Coefficients Obtained by Expanding Each Function in Terms of Legendre Polynomials

Coefficient→ Function↓	1	2	3	4	5	6	7
$\frac{2}{\pi}[(\frac{\pi}{2} - \alpha)\cos(\alpha) + \sin(\alpha)]$	1.146	-.179	.007	.020	-.002	-.0003	0
$\frac{2}{11}\cos(2\alpha) + \frac{9}{11}$	1.157	-.180	0	.022	0	-.0007	0

DOGP compares jets by computing $\frac{D(\vec{x}) \cdot D(\vec{y})}{\|D(\vec{x})\|\|D(\vec{y})\|}$. To analyze this, it is useful to approximate the discretely sampled cosine with a continuous function. So, we take

$$\|D(\vec{x})\| \approx M_x \sqrt{\int_0^{2\pi} (\cos(\theta))^2 d\theta} = M_x \sqrt{\pi}.$$

Similarly, $\|D(\vec{y})\| \approx M_y \sqrt{\pi}$ and, therefore,

$$\frac{D(\vec{x}) \cdot D(\vec{y})}{\|D(\vec{x})\|\|D(\vec{y})\|} \approx \frac{1}{\pi} \int_0^{2\pi} cos(\theta)cos(\theta - \alpha)d\theta = cos(\alpha).$$

This is the same comparison measure that normalized correlation performs when a small window is used.

Next, we consider what happens when we take the absolute values of filter outputs. Since a derivative of Gaussian oriented in the direction θ produces a result with the same magnitude as one oriented toward $\theta + \pi$, we only apply filters in a range of directions from 0 to π. We obtain

$$\frac{D_a(\vec{x}) \cdot D_a(\vec{y})}{\|D_a(\vec{x})\|\|D_a(\vec{y})\|} \approx \frac{2}{\pi} \int_0^{\pi} |cos(\theta)||cos(\theta - \alpha)|d\theta.$$

We can assume, without loss of generality, that $0 \leq \alpha \leq \frac{\pi}{2}$. We obtain

$$\frac{2}{\pi} \int_0^{\pi} |cos(\theta)||cos(\theta - \alpha)|d\theta =$$
$$\frac{2}{\pi} \left(\frac{\pi}{2}\cos(\alpha) + \frac{\sin(\alpha)}{2} - \alpha \cos(\alpha) + \frac{\cos(\alpha)\sin(2\alpha)}{2} \right.$$
$$\left. - \frac{\sin \alpha \cos(2\alpha)}{2} \right) = \frac{2}{\pi} \left[\left(\frac{\pi}{2} - \alpha \right) \cos(\alpha) + \sin(\alpha) \right].$$

This first equality is obtained by breaking the integral into intervals of constant sign. The second equality follows from trigonometric identities.

In fact, $\frac{2}{\pi}[(\frac{\pi}{2} - \alpha)\cos(\alpha) + \sin(\alpha)] \approx \frac{2}{11}\cos(2\alpha) + \frac{9}{11}$. To show this, we can expand the two functions with Legendre polynomials. The coefficients of the two functions in this basis are shown in Table 1. They are almost identical. The two functions are plotted in Fig. 4.

Therefore, DOG essentially compares image gradients by taking the cosine of twice the difference in angle. This comparison is insensitive to the polarity of this difference since it is periodic with a frequency of π. Within this range, it is monotonic with the difference in angle and qualitatively identical to the comparison method in DIR.

Note that these functions are a continuous approximation of the discrete functions actually computed. However, for image matching, we have verified that a discrete version with $k = 8$ produces numbers that generally differ by two percent or less from a continuous version (to compute this,

we numerically integrate with a very large k), and performs matching identically with $k = 8$ and larger values of k.

Note that by using a set of eight filters, we obtain a highly redundant feature set. However, if we compare these jets by taking inner products, we can see that this is a discrete approximation of a comparison of a continuous function of the gradient. A nonredundant set of two oriented filters would provide a poor approximation of this continuous function, while eight orientations provide an excellent approximation.

4.3 Histogram Equalization and Mutual Information

We now briefly discuss the connection between comparisons based on the direction of the gradient and those based on histogram equalization (HE) or mutual information (MI). First, consider histogram equalization that consists of a monotonic transformation of pixel intensities. HE does not change the shape of isoluminant contours in the image. So, two images are identical after histogram equalization only if they have identical isoluminant contours. The direction of the gradient is always orthogonal to the isoluminant contours, implying also that the gradient directions are identical at all locations. HE also does not alter the polarity of a gradient because it transforms intensities monotonically. Therefore, whenever two images are identical after HE, the original images will be judged identical by DIRP.

A similar relationship exists between mutual information and DIR. Given an image A, a new image B will maximize mutual information with A whenever any set of pixels that have identical intensities in A all have a single (possibly different) intensity in B. Therefore, B has maximum mutual information with A only if it has identical isoluminant contours and, therefore, identical gradient directions with A.

The converse relationships need not hold. Two isoluminant contours may have identical intensities in one image but not in the other. This will lead to images with identical gradient directions that HE and MI will judge to be different. Intuitively, gradient direction, HE, and MI all measure the similarity in the shape of isoluminant contours.
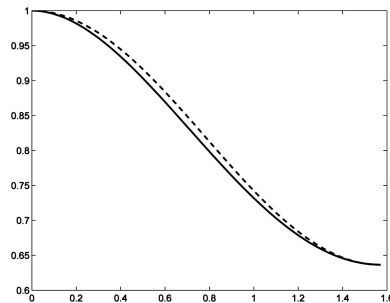

Fig. 4. The solid line shows the comparison value calculated by DOG ($\frac{2}{\pi}[(\frac{\pi}{2} - \alpha)\cos(\alpha) + \sin(\alpha)]$). The dashed line shows that this is very close to $\frac{2}{11}\cos(2\alpha) + \frac{9}{11}$.

However, HE and MI depend on global properties of the image, such as whether distant points have identical intensities, while DIR is more local. Which is better will depend on the domain, but we expect DIR to be preferable when the effects of lighting vary throughout an image.

5 APPROACHES TO COMBINING

So far, we have focused on two types of methods: those based on Whitening and those equivalent to gradient-direction-based methods. We have shown that Whitening is effective on isotropic surfaces while the gradient direction is effective on nonisotropic surfaces. We therefore expect to improve performance on mixed surfaces by combining these methods. In this section, we consider some simple, intuitive methods of combining Whitening with the direction of the gradient.

The simplest method of combining is to sum the outputs of two filters with proper normalization. Our experiments on mixed surfaces show (see Section 6.4) that this simplest combination of Whitening and direction of gradient gives better results than either Whitening or direction of gradient alone.

A more elegant way of combining these two methods is by defining a jet of oriented whitening filters. We showed that the Laplacian can be used for Whitening without learning specific to a class of surfaces. So, a practical method of combining is to use a *jet of oriented second derivatives* (JOSD). Specifically, let $L_\theta(\vec{x})$ denote the result at position $\vec{x}$ of convolving an image, I, with a second derivative of a Gaussian filter, in which the second derivative is taken in the direction θ. By varying θ, we produce a set of filters and by convolving an image with these filters we obtain a vector of values at each pixel: One for each orientation. We form a jet of the magnitudes of these outputs, $L(\vec{x}) = (|L_{\theta_1}(\vec{x})|, \ldots |L_{\theta_k}(\vec{x})|)$. As with other jets, we compare points in two images by computing the jets J and I at corresponding points and taking their normalized correlation: $\frac{J \cdot I}{\|J\|\|I\|}$.

JOSD is an effective combining method because 1) as a jet, it is invariant to offset, gain, and polarity and, so, to lighting changes at discontinuities, 2) normalized oriented filters should produce different responses on nonisotropic surfaces of different orientations, 3) as an approximation of Whitening, JOSD is effective on isotropic surfaces.

One can see a connection between the JOSD and Gabor jets [14], [30]. The Gabor filter is complex. The real component of the Gabor is the product of a cosine and a Gaussian. For an orientation of zero, this component is symmetric about the y-axis and is called the even component of the Gabor. The imaginary part of the Gabor is the product of a sine and a Gaussian, and is called the odd component. The magnitude of the complex Gabor filter is taken when forming the output of Gabor filters into a jet. This captures the amplitude of the image in a band of frequencies, but discards the phase.

It has been noted that the odd components of the Gabor jet are quite similar to an oriented derivative of a Gaussian. Fig. 5 plots a comparison of the two functions. This similarity suggests that a jet composed of only the odd components of a
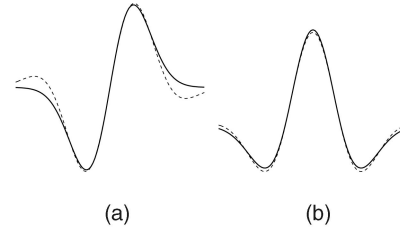

(a)   (b)

Fig. 5. (a) A cross section of the odd component of a Gabor filter compared to a difference of Gaussian. The Gabor is shown as a dashed line, while the difference of Gaussian is solid. (b) The even component of a Gabor filter (dashed) compared to a second derivative of a Gaussian.

Gabor will behave similarly to a jet of oriented derivatives of Gaussians. Therefore, it will yield similar results to those obtained by gradient direction comparison. We will call image comparison based on odd Gabors GO. Our experiments indeed show that the behavior of GO and DOG are very similar qualitatively. Similarly, a jet of even Gabor components (we will call it GE) behaves much like JOSD presented above. While we do not analyze the differences between the even components of Gabors and oriented second derivatives, their similarity is clear from Fig. 5b.

We can therefore see that Gabor jets contain elements of two different methods, one which compares images based on the direction of the gradient and the second which is itself similar to a combination of Whitening and a method sensitive to the direction of the image's second derivative. We now examine the method by which the odd and even Gabor components may be combined.

By taking the magnitude of the output of a complex Gabor filter prior to taking the inner product between two jets, we are extracting the amplitude of the response and discarding the phase. The motivation for this is to achieve quasi-invariance to small deformations in shape (phase has been used separately for fine alignment of images [50]). However, it is not clear that discarding phase should help achieve illumination invariance. We conjecture that Gabor jets are illumination insensitive largely because they combine even and odd components of Gabor jets, which separately produce illumination insensitivity. To test this conjecture, we experimentally compare the performance of Gabor jets (GJ) and a method we call GO + GE. GO + GE separately compares the even and odd components of Gabor jets, and then adds the resulting comparison measures. The details of these experiments are given in Section 6.4, and the results can be seen in Fig. 11. In brief, we compare GJ with GO + GE on both isotropic and nonisotropic objects. In every case, GO + GE outperforms GJ, except when the performance of both measures is quite poor.

6 EXPERIMENTS

The experiments are designed to test our main claims:

1. Different representations provide illumination insensitivity to different types of surfaces. Specifically,

- Direction of gradient and related methods based on local normalization are more effective on nonisotropic surfaces.
- These methods are equivalent and are expected to give the same results.
- Whitening methods are more effective on smooth isotropic surfaces.

2. Combining surface dependent representations is beneficial to illumination insensitive image comparisons.

Direct measurement and analysis of the illumination insensitivity of different representations is not straightforward because different objects are associated with different degrees of change. Moreover, different representations use different units (e.g., gray-levels and directions), which makes comparison difficult. Finally, a constant representation (e.g., an image where all gray-levels are zeros) would be highly stable but, of course, useless. Therefore, we chose to compare the representations, indirectly, using a matching experiment with ensembles of objects. A highly illumination insensitive representation will maintain similarity of images of the same object, while creating dissimilarities between images of other objects. Therefore, it will give more accurate matching.

6.1 Recognition Methods

We applied the following recognition scheme as an application of our approach to image comparison. For a given collection of objects, we took reference images of each object under the same lighting conditions. Then, we took query images of these objects with different lighting directions. A query image was compared to each reference image and matched to the one that minimized the appropriate comparison measure. We have tested the following comparison measures.

LOG. We filtered each image with a Laplacian of Gaussian, normalized the filtered images to unit length, and measured the SSD between the query and each reference image.

Whitening. For every set of images, we learned a whitening filter as a 2D causal filter that minimizes the MS prediction error. The size of the filter varied according to the smoothness of the images. The whitening filter was trained on the difference images obtained by subtracting reference images from the corresponding images associated with the same object under varying illumination. During testing (on different images), we whitened each image, normalized the whitened images to unit length, and measured the SSD between the query and each reference image.

DIR. We first smoothed all images with a Gaussian. We defined the direction of the gradient of the smoothed image, I, as

$$r(\vec{x}) = mod\left(atan\left(\frac{\partial G \star I}{\partial y} \middle/ \frac{\partial G \star I}{\partial x}\right), \pi\right).$$

Discarding the polarity of an edge in order to compare the directions of gradients in two images, r_I, r_J, we took $min((r_I - r_J)^2, mod(\pi - |r_I - r_J|, \pi)^2)$.
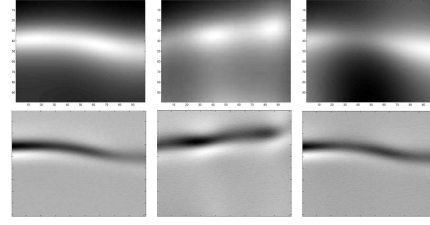

Fig. 6. A typical triplet from the smooth synthetic experiment. Top row: original images. Bottom row: whitened images. The first and third column correspond to the same surface, while the second column corresponds to a different surface.

DOG, GO, GE, GO + GE, GJ, JOSD. These methods use jets of oriented first derivatives of Gaussians (DOG), the odd (GO), or even (GE) components of Gabor Jets (GJ), the sum of these two (GO + GE) or jets of oriented second derivatives (JOSD). All these methods involve a Gaussian, selected so that the amount of smoothing is the same as with DIR.

NC. We perform normalized correlation in every 3×3 window and use its absolute value to discard polarity. This allows a fair comparison with DIR, which is also polarity insensitive. Normalized correlation is an increasing function of the match quality, between zero and one, so we use 1—(absolute value of the normalized correlation) as a distance function, denoted "NC." To integrate the results over the whole image, we computed the magnitude of the vector that contains these absolute normalized correlations.

6.2 Data Sets

Since the Whitening method requires training, we divided image sets used in whitening experiments into training and test sets.

6.2.1 Synthetic Very Smooth Set

Every scene was created as a sum of random harmonic functions, with fixed amplitudes but random directions and phases. This provides an ensemble of images with similar statistical properties. These were rendered as Lambertian surfaces with point sources.

The training set included 2,000 images with a fixed illumination, deviating 67.5 degrees from the frontal direction. Since the synthetic images are very smooth, we trained a large whitening filter of 265 coefficients inside a 23×23 window. The test was conducted on 5,000 triplets of images. Two of each triplet were reference images produced from random scenes illuminated by the same nearly frontal illumination. The third was a query image synthesized from the first scene, with a different illumination varying over a hemisphere with deviation up to 67.5 degrees from the frontal. A typical triplet is shown in Fig. 6.

6.2.2 Real Isotropic Set

We created 18 real, smooth objects from clay (Fig. 7). Even though some parts of some objects were nonisotropic, we consider these to be typical of real objects that are relatively smooth and isotropic. We illuminated the objects by a single light source moving along a half circle, so that its distance from the object was roughly fixed. We used a camera placed vertically above the object, and took 14 images of every
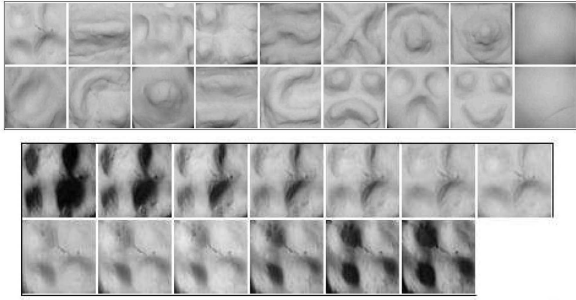
Fig. 7. Examples of images from the "real isotropic set." The top two rows show all the objects from the data set captured under frontal lighting. The bottom two rows show the lighting variation in the set, with images of one object.

object with different lighting directions at angles in the range $[-65, 65]$ degrees to the vertical axis. One image of each object, associated with a nearly vertical (frontal) illumination, was chosen as the reference image.

A single whitening filter was constructed as follows: We randomly picked two objects and trained the whitening filter on the difference between reference images and corresponding images associated with the same object and six other illuminations—12 images in total (two objects x six illuminations). We learned the whitening filter as a 2D causal filter with 25 coefficients inside 7×7 windows. (This size filter produced best results for the given domain). All images of the 18 objects except the reference images were used as query images (234 images).

6.2.3 Real Nonisotropic Set

For nonisotropic objects, we used the Yale database [7], which contains 20 objects with abrupt changes in albedo and shape. The database consists of 63 images of each object with lighting direction varying over a hemisphere with maximum deviation of 90 degrees from the frontal (see [16] for more details on the lighting configuration). In our experiments, we used 90×90 pixel subwindows from each image. Fig. 8 shows the subwindows from images of all objects taken under frontal illumination.

The whitening filter was trained on 20 difference images associated with two objects and 10 illuminations. Varying the size of the whitening filter from a 3×3 to a 7×7 window had no real effect on recognition performance. This is not surprising because these objects are relatively rough. As in the previous set, one image of each object, associated with frontal illumination, was chosen as the reference image. The rest of the images were used as queries.

6.2.4 Mixed Set

The mixed set contains 27 clay objects with different degrees of isotropy (Fig. 10). We took 32 images of each object with light varying over a quarter of a sphere. The light was placed every 60 degrees (four positions) at 8 latitudes with the highest latitude as frontal light. The lowest latitude was 86 degrees from the frontal.[2]

2. The results, described in Fig. 11, are plotted for illumination angles in the range $[0, 65]$, which are available for all data sets.

Fig. 8. Examples of images from the "nonisotropic real set." Top: fragments of all objects captured under frontal lighting. Bottom: lighting variation in the set with images of one object.

6.3 Testing Surface Dependent Representations

To test the contention that different types of surfaces require different representations, we applied Whitening and the Laplacian of Gaussian (expected to be good representations for isotropic surfaces), and DIR, DOG, and normalized correlation (expected to be good representations of nonisotropic surfaces) in the task of object recognition under varying illumination on the 1) synthetic very smooth set, 2) real isotropic set, and 3) real nonisotropic set.

In all our experiments, DOG and NC behaved almost identically to DIR (this was also shown analytically in Sections 4.1 and 4.2). Therefore, we do not show plots for the DOG and NC methods. The experiments (Fig. 9) show that Whitening outperforms other methods on isotropic surfaces. LOG performs quite well on real isotropic surfaces. However, on very smooth surfaces (synthetic images), its performance degrades because its size is insufficient to handle the high correlations between the gray-levels present on surfaces of this type. DIR is not as good as Whitening and LOG on isotropic objects, but outperforms them on nonisotropic objects. Whitening and LOG are almost identical on nonisotropic objects. This is to be expected since many objects in this set contain abrupt changes in albedo or shape; thus, the correlation in these objects is short range and the whitening filter is similar to the Laplacian.

The above results support our conjecture that DIR and other methods based on normalization (DOG, NC, etc.) are best for image comparisons of nonisotropic objects, while Whitening and LOG are best for images produced by isotropic surfaces.

6.4 Testing Combined Methods

We tested the combined methods discussed in the paper (simple combining, JOSD, GJ, and GO + GE) on three sets of real objects: nonisotropic, isotropic, and mixed.

First, we see that even though the combined methods perform differently on different sets, in general, they are better than surface dependent methods, specifically Whitening and DIR. Fig. 12 shows specific examples of objects from different sets where one or even both of the
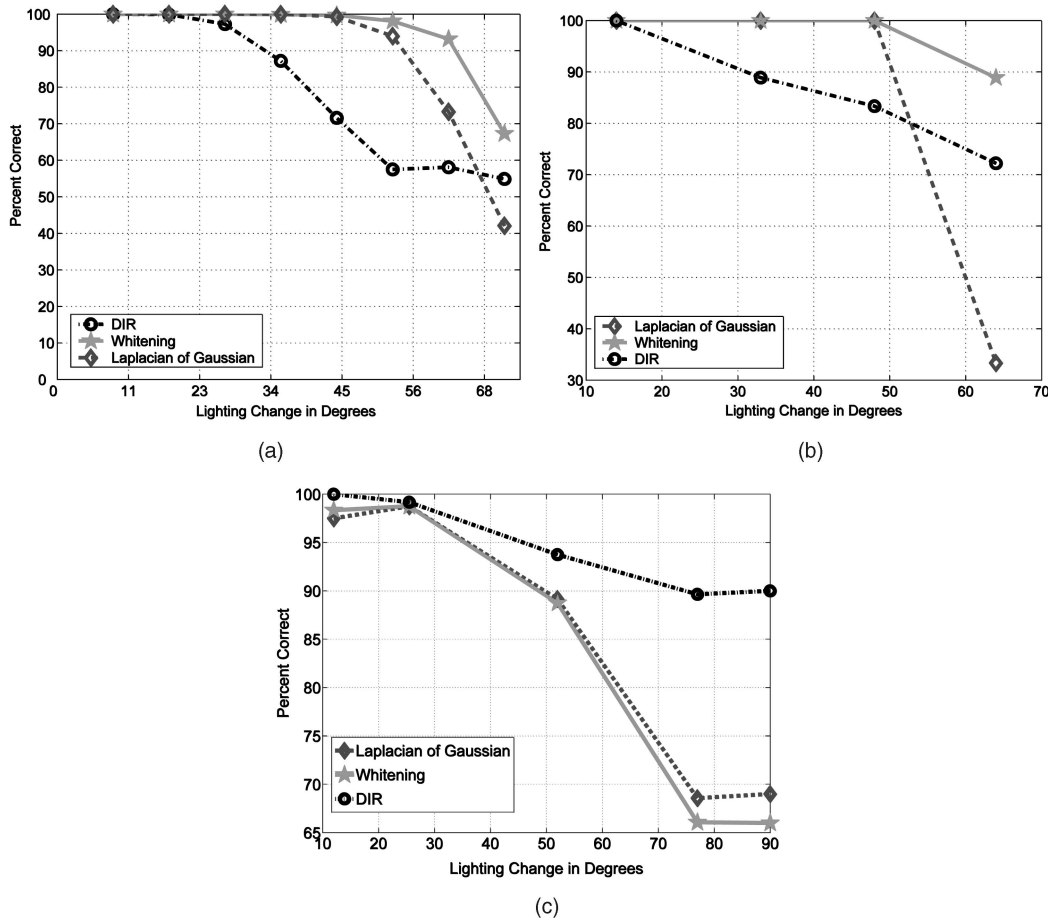
(a)

(b)

(c)

Fig. 9. Experimental results of surface dependent representations: (a) smooth synthetic objects, (b) smooth real objects, and (c) rough real objects.

surface dependent methods fail to recognize the object while the combined method JOSD succeeds.

Next, we point out that GO + GE performs very similarly to GJ. We interpret this as support for our conjecture that Gabor jets work well under variable lighting because of the individual performance of their constituents GE and GO, and not because of the particular way in which GJ combines these two filters. We noted earlier that qualitatively GO is
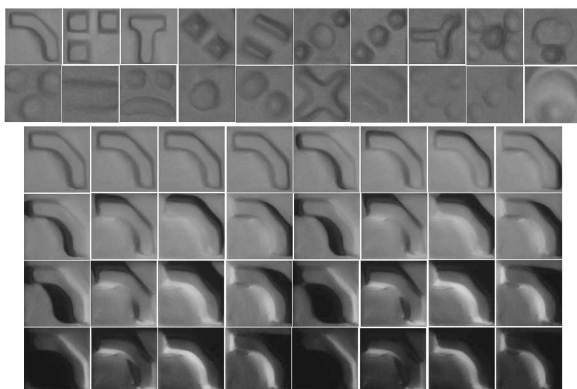

Fig. 10. Examples of images from the "mixed set." Top: fragments of all objects captured under frontal lighting. Bottom: lighting variation in the set with images of one object.

very similar to DOG and GE is very similar to JOSD. We verified the posited similarity by comparing the performance of these methods on the mixed set (Fig. 11d).

7 CONCLUSIONS

We have discussed two classes of illumination insensitive representations and shown that they are effective when applied to two different classes of surfaces. To clarify this further, we have shown that a number of existing methods are actually equivalent to comparisons based on the image gradient direction, which is known to be effective for nonisotropic surfaces, and we have shown that isotropic surfaces can be effectively dealt with by a different strategy based on whitening the image. Experiments clearly show that which method is best depends on the surface type. This leads to a better understanding of existing algorithms, as well as to more effective methods of whitening when we can learn a filter based on the relevant class of surfaces.

Our analysis also suggests that even more effective representations will combine these two strategies. Such mixed approaches can work effectively on a wider range of surfaces. We demonstrate that Gabor jets seem to employ such a mixed strategy, which helps to explain their effectiveness at illumination insensitive recognition, and we develop
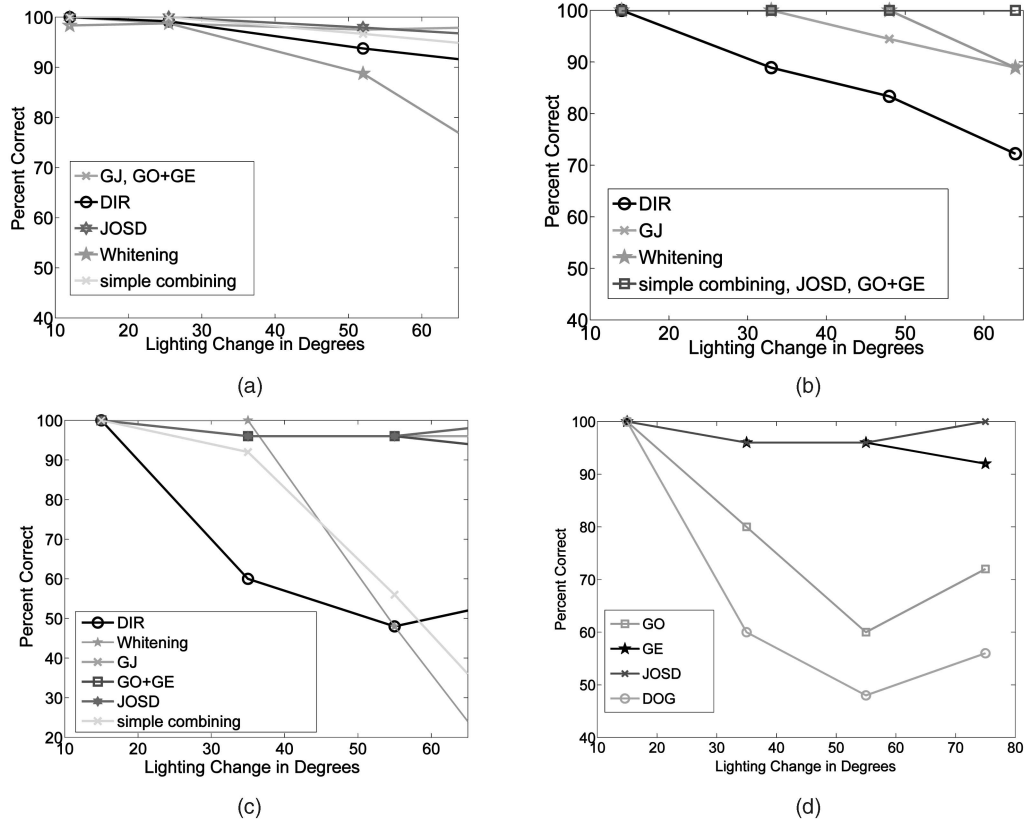
Fig. 11. Experimental results of the combined methods compared against surface dependent representations. All objects are real. (a) Nonisotropic. (b) Isotropic. (c) Mixed. (d) Compares the performance of GE against JOSD and GE against DOG in the mixed set.

new mixed strategies as well. The last few years have seen dozens of papers on illumination insensitive image matching; we hope that a deeper understanding of these methods will assist practitioners in choosing the most effective methods for their problem and help lay the groundwork for improved and more widely applicable methods.

APPENDIX

PROOF OF PROPOSITION 1

Proof. The derivatives of Gaussian fields are Gaussian field themselves and, so is any linear function of them. Therefore, $I(x, y)$ is a Gaussian random field. Its expected value is

$$E[I(x,y)] \approx E\big[\rho I_0\big[sin\theta - cos\theta(h_x cos\phi + h_y sin\phi)\big]\big] = \rho I_0 sin\theta$$

because, by homogeneity, $E[h_x] = E[h_y] = 0$. As for the covariance,
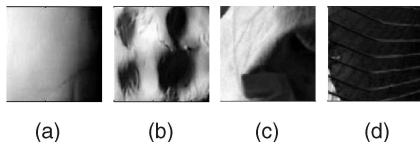

Fig. 12. Examples of some images where the combined methods outperform surface dependent representations. (a) Whitening works, DIR fails. (b) DIR works, Whitening fails. (c) Both Whitening and DIR fail. JOSD works on all four images.

$$C_I(\bar{x}_i, \bar{x}_j) = E[(I(\bar{x}_i) - E[I(\bar{x}_i)])(I(\bar{x}_j) - E[I(\bar{x}_j)])]$$
$$= \rho^2 I_0^2 cos^2\theta \, (cos^2\phi E[h_x(\bar{x}_i)h_x(\bar{x}_j)]$$
$$+ sin^2\phi E[h_y(\bar{x}_i)h_y(\bar{x}_j)]$$
$$+ sin\phi cos\phi E[h_x(\bar{x}_i)h_y(\bar{x}_j) + h_x(\bar{x}_j)h_y(\bar{x}_i)]).$$
(6)

The expected values of the surface derivative products are easily calculated by

$$E[h_x(\bar{x}_i)h_x(\bar{x}_j)] = \frac{\partial^2 C_S(\bar{x}_i, \bar{x}_j)}{\partial x_i \partial x_j}$$
$$= C_S(\bar{x}_i, \bar{x}_j)\frac{2}{\beta^2}\left(1 - \frac{2}{\beta^2}(x_i - x_j)^2\right),$$
$$E[h_y(\bar{x}_i)h_y(\bar{x}_j)] = \frac{\partial^2 C_S(\bar{x}_i, \bar{x}_j)}{\partial y_i \partial y_j}$$
$$= C_S(\bar{x}_i, \bar{x}_j)\frac{2}{\beta^2}\left(1 - \frac{2}{\beta^2}(y_i - y_j)^2\right),$$
(7)
$$E[h_x(\bar{x}_i)h_y(\bar{x}_j)] = \frac{\partial^2 C_S(\bar{x}_i, \bar{x}_j)}{\partial x_i \partial y_j}$$
$$= -C_S(\bar{x}_i, \bar{x}_j)\left(\frac{2}{\beta^2}\right)^2 (x_i - x_j)(y_i - y_j)$$
$$= E[h_y(\bar{x}_i)h_x(\bar{x}_j)].$$

Inserting these terms in (6) and expressing $x_i - x_j$ and $y_i - y_j$ as $\|\bar{x}_i - \bar{x}_j\|cos\psi$ and $\|\bar{x}_i - \bar{x}_j\|sin\psi$, respectively, gives

$$C_I(\bar{x}_i, \bar{x}_j) = \mathcal{C}\left[1 - \frac{2}{\beta^2}\left(cos^2\phi(x_i - x_j)^2 + sin^2\phi(y_i - y_j)^2\right.\right.$$
$$\left.\left. + 2sin\phi cos\phi(x_i - x_j)(y_i - y_j)\right)\right]$$
$$= \mathcal{C}\left[1 - \frac{2}{\beta^2}\|\bar{x}_i - \bar{x}_j\|^2\left(cos^2\phi cos^2\psi + sin^2\phi sin^2\psi\right.\right.$$
$$\left.\left. + 2sin\phi cos\phi sin\psi cos\psi\right)\right]$$
$$= \mathcal{C}\left[1 - \frac{2}{\beta^2}\|\bar{x}_i - \bar{x}_j\|^2(cos\phi cos\psi + sin\phi sin\psi)^2\right]$$
$$= \mathcal{C}\left[1 - \frac{2}{\beta^2}\|\bar{x}_i - \bar{x}_j\|^2 cos^2(\phi - \psi)\right],$$

$$(8)$$

where $\mathcal{C} = \rho^2\,I_0^2\,cos^2\theta\left(\frac{2C_S(\bar{x}_i, \bar{x}_j)}{\beta^2}\right)$. The variance σ_I^2 is $C_I(\bar{x}_i, \bar{x}_i)$. □

ACKNOWLEDGMENTS

REFERENCES

[1]   R.J. Adler, *The Geometry of Random Fields.* John Wiley and Sons, 1981.
[2]   R. Basri and D. Jacobs, "Lambertian Reflectance and Linear Subspaces," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 25, no. 2, pp. 218-233, Feb. 2003.
[3]   M. Bichsel, "Strategies of Robust Object Recognition for the Automatic Identification of Human Faces," PhD thesis, no. 9467, ETH Zurich, 1991.
[4]   R. Brunelli, "Estimation of Pose and Illuminant Direction for Face Processing," MIT AI Memo 1499, 1994.
[5]   C. Bregler and J. Malik, "Learning Appearance Based Models: Mixtures of Second Moment Experts," *Advances in Neural Information Precessing Systems 9,* pp. 845-851, 1997.
[6]   T. Brunelli and T. Poggio, "Face Recognition: Features versus Templates," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 15, no. 10, pp. 1042-1062, Oct. 1993.
[7]   H. Chen, P. Belhumeur, and D. Jacobs, "In Search of Illumination Invariants," *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* vol. I, pp. 254-261, 2000.
[8]   T. Cootes and C. Taylor, "On Representing Edge Structure for Model Matching," *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* vol. I, pp. 1114-1119, 2001.
[9]   R.O. Duda and P.E. Hart, and D.G. Stork, *Pattern Classification.* Wiley, 2001.
[10] P. Felzenszwalb and D. Huttenlocher, "Pictorial Structures for Object Recognition," to appear.
[11] A. Fitch, A. Kadyrov, W. Christmas, and J. Kittler, "Orientation Correlation," *Proc. British Machine Vision Conf.,* vol. 1, pp. 133-142, 2002.
[12] W. Freeman, D. Anderson, P. Beardsley, C. Dodge, M. Roth, C. Weissman, W. Yerazunis, H. Kage, K. Kyuma, Y. Miyake, and K. Tanaka, "Computer Vision for Interactive Computer Graphics," *IEEE Computer Graphics and Applications,* vol. 18, no. 3, May/June 1998.
[13] D. Forsyth and J. Ponce, *Computer Vision: A Modern Approach.* Prentice-Hall, 2003.
[14] D. Gabor, "Theory of Communication," *J. IEEE,* vol. 93, pp. 429-457, 1946.
[15] T. Gevers and A.W.M. Smeulders, "Color Constant Ratio Gradients for Image Segmentation and Similarity of Textured Objects," *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* vol. 1, pp. 18-25, 2001.

[16] A.S. Georghiades, P.N. Belhumeur, and D.J. Kriegman, "From Few to Many: Generative Models for Recognition under Variable Pose and Illumination," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 23, no. 6, pp. 643-660, June 2001.
[17] R. Gonzalez and R. Woods, *Digital Image Processing.* Anderson-Wesley, 1992.
[18] H. Greenspan, S. Belongie, R. Goodman, P. Perona, S. Rakshit, and C. Anderson, "Overcomplete Steerable Pyramid Filters and Rotation Invariance," *Proc. Conf. Computer Vision and Pattern Recognition,* pp. 222-228, 1994.
[19] R. Gross and V. Brajovic, "An Image Preprocessing Algorithm for Illumination Invariant Face Recognition," *Proc. Fourth Int'l Conf. Audio and Video-Based Biometric Person Authentication,* pp. 10-18, 2003.
[20] R.M. Haralick and L.G. Shapiro, *Computer and Robot Vision.* Addison-Wesley, 1992.
[21] D. Hond and L. Spacek, "Distinctive Descriptions for Face Processing," *Proc. Eighth British Machine Vision Conf.,* vol. 1, pp. 320-329, 1997.
[22] B. Jähne, *Digital Image Processing.* Springer-Verlag, 1995.
[23] A.K. Jain, *Fundamentals of Digital Image Processing.* Prentice Hall, 1989.
[24] F. Jiao, S. Li, H. Shum, and D. Schuurmans, "Face Alignment Using Statistical Models and Wavelet Features," *Proc. Conf. Computer Vision and Pattern Recognition,* 2003.
[25] D. Keren, M. Osadchy, and C. Gotsman, "Antifaces: A Novel, Fast Method for Image Detection," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 23, no. 7, pp. 747-761, July 2001.
[26] J. Kittler, Y. Li, and J. Matas, "On Matching Scores for LDA-Based Face Verification," *Proc. 11th British Machine Vision Conf.,* vol. 1, pp. 42-51, 2000.
[27] J.J. Koenderink and S.C. Pont, "Irradiation Direction from Texture," *J. Optical Soc. of Am. A,* vol. 20, no. 10, pp. 1875-1882, 2003.
[28] J.J. Koenderink and A. van Doorn, "Representation of Local Geometry in the Visual System," *Biological Cybernetics,* vol. 55, pp. 367-375, 1987.
[29] E. Land and J. McCann, "Lightness and Retinex Theory," *J. Optical Soc. of Am.,* vol. 61, no. 1, pp. 1-11, 1971.
[30] M. Lades, J.C. Vortbrüggen, J. Buhmann, J. Lange, C. von der Malsburg, R.P. Würtz, and W. Konen, "Distortion Invariant Object Recognition in the Dynamic Link Architecture," *IEEE Trans. Computers,* vol. 42, pp. 300-311, 1993.
[31] S. Narasimhan, V. Ramesh, and S. Nayar, "A Class of Photometric Invariants: Separating Material from Shape and Illumination," *Proc. Int'l Conf. Computer Vision,* pp. 1387-1394, 2003.
[32] S.K. Nayar and R.M. Bolle, "Reflectance Based Object Recognition," *Int'l J. Computer Vision,* vol. 17, no. 3, pp. 219-240, Mar. 1996.
[33] M. Oren and S.K. Nayar, "A Theory of Specular Surface Geometry," *Int'l J. Computer Vision,* vol. 24, no. 2, pp. 105-124, 1997.
[34] M. Osadchy, M. Lindenbaum, and D. Jacobs, "Whitening for Photometric Comparison of Smooth Surfaces under Varying Illumination," *Proc. European Conf. Computer Vision,* pp. 217-228, 2004.
[35] M. Osadchy, D. Jacobs, and M. Lindenbaum, "On the Equivalence of Common Approaches to Lighting Insensitive Recognition," *Proc. IEEE Int'l Conf. Computer Vision,* vol. 2, pp. 1721-1726, 2005.
[36] M. Osadchy, D. Jacobs, and R. Ravamoorthi, "Using Specularities for Recognition," *Proc. IEEE Int'l Conf. Computer Vision,* vol. 2, pp. 1512-1519, 2003.
[37] A. Papoulis, *Probability, Random Variables, and Stochastic Processes,* third ed. McGraw Hill, 1991.
[38] P.J. Phillips and Y. Vardi, "Efficient Illumination Normalization of Facial Images," *Pattern Recognition Letters,* vol. 17, pp. 921-927, 1996.
[39] W. Pratt, *Digital Image Processing,* first ed. Wiley, 1978.
[40] R. Rao and D. Ballard, "An Active Vision Architecture Based on Iconic Representations," *Artificial Intelligence,* pp. 461-505, 1995.
[41] S. Ravela and C. Luo, "Appearance-Based Global Similarity Retrieval of Images," *Advances in Information Retrieval,* 2000.
[42] C. Schmid and R. Mohr, "Local Grayvalue Invariants for Image Retrieval," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 19, no. 5, pp. 530-535, May 1997.
[43] A. Shashua, "On Photometric Issues in 3D Visual Recognition from a Single 2D Image," *Int'l J. Computer Vision,* vol. 21, nos. 1/2, pp. 99-122, 1997.

[44] A. Shashua and T. Riklin-Raviv, "The Quotient Image: Class-Based Re-Rendering and Recognition with Varying Illuminations," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 23, no. 2, pp. 129-139, Feb. 2001.

[45] D.A. Slater and G. Healey, "The Illumination-Invariant Recognition of 3D Objects Using Local Color Invariants," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 18, no. 2, pp. 206-210, Feb. 1996.

[46] E. Trucco and A. Verri, *Introductory Techniques for 3-D Computer Vision.* Prentice-Hall, 1998.

[47] H.L. VanTrees, *Detection, Estimation, and Modulation Theory,* Part I, Wiley, 1965.

[48] M. Varma and A. Zisserman, "Estimating Illumination Direction from Textured Images," *Proc. Conf. Computer Vision and Pattern Recognition,* vol. I, pp. 179-186, 2004.

[49] L. Wiskott, "Labeled Graphs and Dynamic Link Matching for Face Recognition and Scene Analysis," PhD thesis, 1995.

[50] L. Wiskott, J. Fellous, N. Kruger, and C. van der Malsburg, "Face Recognition by Elastic Graph Matching," *Intelligent Biometric Techniques in Fingerprint and Face Recognition,* pp. 355-396, 1999.

[51] L. Wolff and J. Fan, "Segmentation of Surface Curvature with a Photometric Invariant," *J. Optical Soc. of Am. A,* vol. 11, no. 11, pp. 3090-3100, 1994.

Margarita Osadchy received the PhD degree with honors in computer science in 2002 from the University of Haifa, Israel. From 2001 to 2004, she was a visiting research scientist at the NEC Research Institute. During 2004-2005, she was a postdoctoral fellow in the Department of Computer Science at the Technion—Israel Institute of Technology. In 2005, she joined the Department of Computer Science at the University of Haifa, where she is an assistant professor. Her research has focused on computer vision and machine learning, especially in the areas of object and event recognition. She is a member of the IEEE.

David W. Jacobs received the BA degree from Yale University in 1982. From 1982 to 1985, he worked for Control Data Corporation on the development of data base management systems and attended graduate school in computer science at New York University. From 1985 to 1992, he attended MIT, where he received the MS and PhD degrees in computer science. From 1992 to 2002, he was a research scientist and then a senior research scientist at the NEC Research Institute. In 1998, he spent a sabbatical at the Royal Institute of Technology (KTH) in Stockholm. Since 2002, he has been an associate professor of computer science at the University of Maryland, College Park. His research has focused on human and computer vision, especially in the areas of object recognition and perceptual organization. He has also published articles in the areas of motion understanding, memory and learning, and computational geometry. He and his coauthors received honorable mention for the best paper award at CVPR 2000. He also coauthored a paper that received the best student paper award at UIST 2003. He is a member of the IEEE.

Michael Lindenbaum received the BSc, MSc, and DSc degrees in the Department of Electrical Engineering at the Technion, Israel, in 1978, 1987, and 1990, respectively. From 1978 to 1985, he served in the Israel Defense Forces (IDF). He did his postdoctoral work at the NTT Basic research Labs in Tokyo, Japan, and from 1991, he has been with the Department of Computer Science, Technion. He was also a consultant to Hewlett-Packard Laboratories, Israel, and spent a sabbatical in NEC, New Jersey (in 2001). He served on several committees of computer vision conferences, coorganized the IEEE Workshop on Perceptual Organization in Computer Vision, and was an associate editor of *Pattern Recognition* and *Pattern Recognition Letters*. He has worked in digital geometry, computational robotics, learning, and various aspects of computer vision and image processing. Currently, his main research interest is computer vision, and especially statistical analysis of object recognition and grouping processes. He is a member of the IEEE and the IEEE Computer Society.

▷ **For more information on this or any other computing topic, please visit our Digital Library at** www.computer.org/publications/dlib.