Stochastic Models for Capturing Image Variability

Anuj Srivastava

We review a recent result in modeling lower order (univariate and bivariate) probability densities of pixel values resulting from bandpass filtering of images. Assuming an object-based model for images, a parametric family of probabilities, called Bessel K forms, has been derived [1]. This parametric family matches well with the observed histograms for a large variety of images (video, range, infrared, etc.) and filters (Gabors, Laplacian Gaussian, derivatives, etc). The Bessel parameters relate to certain characteristics of objects present in an image and provide fast tools either for object recognition directly or for an intermediate (pruning) step of a larger recognition system. Examples are presented to illustrate the estimation of Bessel forms and their applications in clutter classification and object recognition.

Introduction

Why is there a recent emphasis on explicit probability models for image pixels? One explanation lies in the growing appreciation for the variability exhibited by the

1053-5888/02/$17.00©2002IEEE

images and the realization that exact mathematical models may not be feasible. Stochastic modeling of large dimensional systems has been applied successfully many times in other areas. For instance, in statistical physics, precise models for energetics of a large ensemble of particles are deemed intractable and a statistical framework is frequently used. The key to a statistical approach is to have probability models that capture the essential variability and yet are tractable. Since any statistical technique for image analysis requires probability models, the development of such probability models becomes a necessary first step.

Classical statistics do provide tools for model building and density estimation but their application to image modeling is difficult due to the large dimensionality of the observation space. Even the smallest images from current digital cameras are 128×128 pixels implying estimation of a density on the space $\mathbb{R}^{16K}$. Therefore, one is faced with the simultaneous tasks of model building and dimension reduction. A popular idea in the vision literature has been to first reduce dimensions using purely numerical considerations and then impose probability models on the reduced data. By not involving any physical consideration on the imaged objects, or any contextual knowledge, the images are treated as elements of a vector space and one seeks a low-dimensional vector subspace (or its basis) that best represents those numbers (under some chosen criterion). Principal components [2], independent components [3], [4], sparse coding [5], Fisher's discriminant [6], Fourier transforms, wavelet transforms [7], and many other representations are all instances of this idea. The main advantage of such linear projections is that they are computationally cheap. However, a lack of physical or contextual information leads to a limited performance, particularly in recognition of objects from their cluttered images. Furthermore, it seems likely that the space generated by real images is a curved manifold and cannot be simply "approximated" globally by a low-dimensional vector space. As a simple example, consider the set of images generated by imaging a rotating three-dimensional (3-D) object, at several orientations and positions with respect to the camera. It is easy to see that this set does not form a vector space, i.e., not all images can be generated as linear combinations of other images. More formally, as described in [8], the image spaces associated with 3-D objects can be specified by the orbits of similarity groups acting on their 3-D templates, followed by the projections on image planes. These projected orbits form curved manifolds of same dimensions as the group of similarity transformations. Therefore, the projections to linear vector spaces can only qualify as local approximations or perhaps as preliminary models for a larger sequential inference system. We have pursued the latter philosophy in this article.

Mallat [9], Mumford et al. [10], Wainwright et al. [11], Field [12], [13], and others have extensively studied empirical distributions of images taken from large databases and have discovered certain interesting patterns. They have established that image statistics under common representations, such as wavelets or subspace bases (PCA, ICA, Fisher's etc.), point to non-Gaussian distributions. For example, a popular mechanism for decomposing images locally, in space and frequency, using wavelet transforms leads to coefficients that are quite non-Gaussian. The histograms display heavy tails and sharp cusps at the median. It is imperative that any probability model adopted for image analysis should explain such observed phenomena.

How can we build probability models that account for these non-Gaussian patterns? One idea is to use non-parametric estimators, of the underlying density functions, such as the histograms or the kernel estimators (see [14] for density estimation) to characterize the original image(s). This approach has been deployed successfully by [15]-[17] and others in the context of homogeneous textures. Another, more ambitious, idea is to build analytical models that can explain and capture these phenomena. If these probabilities can be stated in a convenient parametric form, then the resulting analysis can be simplified considerably as compared to working with the full non-parametric forms. Towards that goal Portilla et al. [18] and Wainwright et al. [11] have investigated a family of probabilities, called Gaussian scale mixtures, that provide convenient scale-space representations of natural images. Mallat [9] proposed the use of generalized Gaussian density function. Lee et al. [19] have proposed an occlusion model, called the dead leaves model, to explain image statistics and their scale invariance. Donoho et al. [20] argue that non-Gaussianity and correlation across scales point to the "ubiquity of edges" in the images. They have proposed stochastic models based on image transformations, called curvelet transforms, that capture and represent edges better than the conventional Fourier or wavelet transforms. We should point out that although Markov random fields do provide analytical probability models for the images (refer to [21]), their role in explaining the observed non-Gaussian behavior has been very limited.

In recent papers Grenander et al. [1], [22] have proposed an analytic form, called Bessel K form, to model the observed one-dimensional (1-D) and two-dimensional (2-D) histograms of images. These forms are analytical and parametric, and hence efficient, and they very closely match the observed behavior of non-Gaussianity. In this article, we review the fundamental idea behind its construction, demonstrate its application to images of different modalities (video, infrared, range), and suggest applications in clutter classification and infrared face recognition.

Explaining Non-Gaussianity of Images

The strongest evidence of non-Gaussianity (of images) comes from the observed histograms: heavier tails than Gaussian (implying larger kurtosis), sharp cusps at the

center, and higher correlations at different scales. Bessel representations explain this phenomena via a fundamental hypothesis that images are made up of objects. This hypothesis is also consistent with a bigger goal in image understanding: detection and recognition of objects of interest in given images. Since the models are object based, the model parameters point to the objects present in the images, and the resulting algorithms directly perform object recognition.

To illustrate the role of object-based modeling in non-Gaussian statistics, consider the following example. For an image I let us study the histogram of differences in the values of the (horizontally) neighboring pixels. If two neighboring pixels have equal or similar intensities then their difference is small and it adds to the histogram bin containing zero (which also happens to be the median value). On the other hand, if the two neighboring values are quite different, as is the case for vertical edges, then the difference is high and it adds to the histogram at the tails. In general, regions with horizontal homogeneity add to the the central peak and the sharp, distinct edges map to the tails. It implies that for images with large objects with smooth, homogeneous foregrounds and sharp, distinct edges, the difference-histograms will have a sharp central peak and heavy tails. On the other hand, images with lots of blurred objects in the scene will have difference-histograms that are close to Gaussian (through central limit theorem). Object-based models for images allow for both the homogeneous regions and the edges in an image and hence can explain the observed non-Gaussianity. Object-based models also support the findings of high correlations across scales. Additionally, these models integrate well with more detailed representations (e.g., deformable templates) of images using physical (3-D) models for objects [22].

We start by describing the model. Let $\mathcal{G}$ be the space of 2-D profiles of different objects that can be present in a scene. Their occurrence in image is according to a marked point process with points $\{z_i \in \mathbb{R}^2 | i = 1,\ldots,n\}$. The ith object is centered at z_i, and has profile g_i chosen randomly from $\mathcal{G}$ and weighted by $a_i \sim N(0,1)$. The image is formed using a weighted superposition according to

$$I(z) = \sum_{i=1}^{n} a_i g_i(z - z_i). \tag{1}$$

Of course, the physics (of the imaging device) favors occlusion in the model, and not a superposition, but the choice of superposition is a necessary simplification for deriving the analytic probabilities for images obtained via linear filtering.

Bessel K Forms for Modeling Marginals

Our first objective is to derive an explicit analytical form for the univariate density of the filtered images. Given an image I and a bank of filters $\{F^{(j)}, j = 1,2,\ldots,J\}$, we compute, for each filter $F^{(j)}$, a filtered image $I^{(j)} = I * F^{(j)}$, where * denotes the 2-D convolution operation. Possible filters are:

▲ *Gabor:* A Gabor filter is a bandpass filter with a Gaussian kernel centered around a specific wavenumber (see [23] for details). For an angle $\theta \in [0,2\pi)$ a Gabor filter is given by

$$F(z) = \exp\left(-\frac{1}{2\sigma^2}(z_\theta(1)^2 + z_\theta(2)^2)\right)\exp\left(-j\frac{2\pi z_\theta(1)}{\sigma}\right),$$

where

$$z_\theta = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}\begin{bmatrix} z(1) \\ z(2) \end{bmatrix},$$

and σ denotes the resolution (scale) associated with the filter.

▲ *Laplacian Gaussian:* Another filter suggested by Marr [24] to model early vision is the Laplacian Gaussian filter whose operation on I is given by $F(z) = (G * \Delta)(z)$ where G is a Gaussian kernel

$$G(z) = \exp\left(\frac{-\|z\|^2}{2\sigma}\right),$$

and Δ is the Laplacian operator:

$$\Delta = \frac{\partial^2}{\partial z(1)^2} + \frac{\partial^2}{\partial z(2)^2}.$$

▲ *Gradient:* These filters are used to detect the horizontal and the vertical edges:

$$F(z) = \frac{\partial}{\partial z(1)} \text{ or } \frac{\partial}{\partial z(2)}.$$

▲ In addition, one can utilize a wide variety of filters: neighborhood operators, steerable filters, interpolation filters, and so on. Each filter selects and isolates certain features present in the original image.

Applying 2-D convolution to filter both sides of (1), we obtain a model for the filtered image:

$$I^{(j)}(z) \equiv (I * F^{(j)})(z) = \sum_i a_i g_i^{(j)}\left(\frac{1}{\rho_i}(z - z_i)\right), \tag{2}$$

where $g_i^{(j)} = F^{(j)} * g_i$. The conditional density of $I^{(j)}(z)$, given the Poisson points $\{z_i\}$, the scales $\{\rho_i\}$, and the profiles $\{g_i\}$, can be shown to be normal with mean zero and variance u, where $u \equiv \sum_i (g_i^{(j)}(1/\rho_i(z - z_i))^2$. Let the random variable u be a scaled-Gamma random variable with density function given by

$$\frac{1}{c\Gamma(p)}\exp\left(\frac{-u}{c}\right)\left(\frac{u}{c}\right)^{(p-1)}.$$

Under these conditions, the density function of the random variable $I^{(j)}(z)$ has been shown to be [1]: for $p > 0$, $c > 0$,

$$f(x;p,c)=\frac{1}{Z(p,c)}|x|^{p-0.5}K_{(p-0.5)}\left(\sqrt{\frac{2}{c}}|x|\right), \tag{3}$$

where K is the modified Bessel function and Z is the normalizing constant given by $Z(p,c)=\sqrt{\pi}\Gamma(p)(2c)^{0.5p+0.25}$. Let $\mathcal{D}$ be the space of all such densities: $\mathcal{D}=\{f(x;p,c)\,|\,p>0,c>0\}$. We refer to the elements of $\mathcal{D}$ as Bessel K forms and the parameters (p,c) as Bessel parameters. p is called the shape parameter and c is called the scale parameter. The elements of $\mathcal{D}$ have the following properties. i) They are symmetric and unimodal with the mode at zero. ii) The Bessel K forms are leptokurtic, i.e., the tails are heavier as compared to a normal curve with the same variance. iii) A Bessel K form is a specific kind of normal variance-mean mixture [25] with the mixing variable being a scaled Gamma random variable. iv) The family of Bessel K forms is infinitely divisible. However, if I_1 and I_2 are independent with densities $f(x;p_1,c_1)$ and $f(x;p_2,c_2)$, respectively, with $c_1\neq c_2$, the density of $a_1I_1+a_2I_2$ $(a_1,a_2\in\mathbb{R})$ may not be a Bessel K form. For the bivariate forms studied later, the density associated with $a_1I_1+a_2I_2$ has been approximated by an appropriate Bessel K form. However, a rigorous performance analysis comparing the observed 2-D histograms to the approximated Bessel forms remains to be performed. v) A Bessel K forms is square-integrable only for $p>0.25$. This property limits our later derivation of L^2-metric to the Bessel K forms with p-values larger than 0.25.

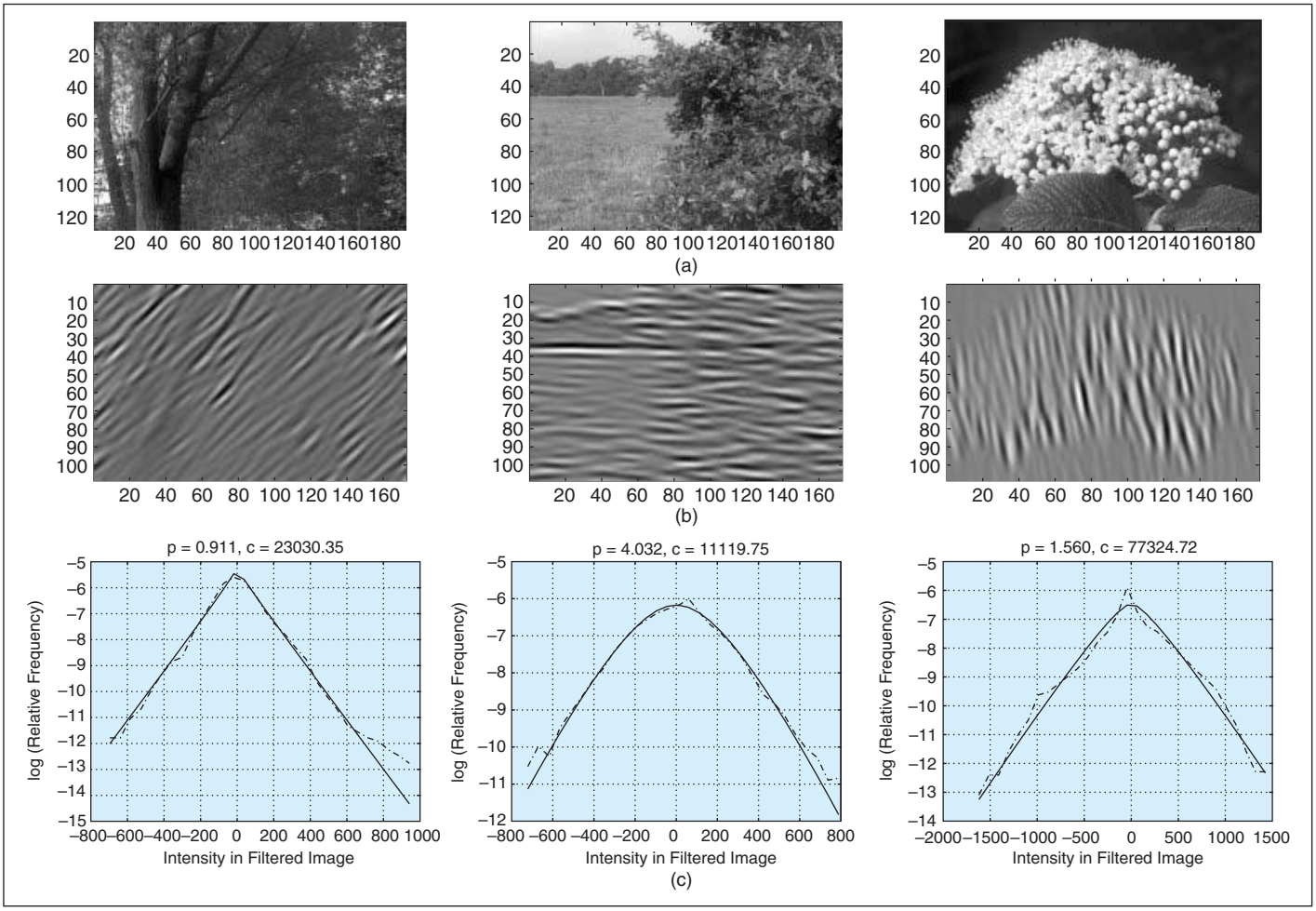
Estimation of Bessel K Forms

As described in [1], p and c can be estimated from the observed data using:

$$\hat{p}=\frac{3}{SK(I^{(j)})-3},\qquad \hat{c}=\frac{SV(I^{(j)})}{\hat{p}}, \tag{4}$$

where SK is the sample kurtosis and SV is the sample variance of the pixel values in $I^{(j)}$. The computational task of estimating the marginal density is that of computing the second and the fourth cumulants of the filtered image.

We list some estimation results. Shown in Fig. 1(a) are some images taken from the van Hateren database [26]; (b) displays the images after filtering by a Gabor filter each, with the filter having an arbitrary orientation and scale, and(c) plots the marginal densities of $I^{(j)}$. The observed densities (histograms) are plotted in broken lines and the estimated Bessel K forms ($f(x;\hat{p},\hat{c})$) are plotted



▲ *1. (a) Images, (b) their Gabor components, and (c) the marginal densities. The observed densities are drawn in broken lines and the estimated Bessel K forms are drawn in solid lines.*

in solid lines. The y-axis is plotted on a log scale to study the shapes of the densities.

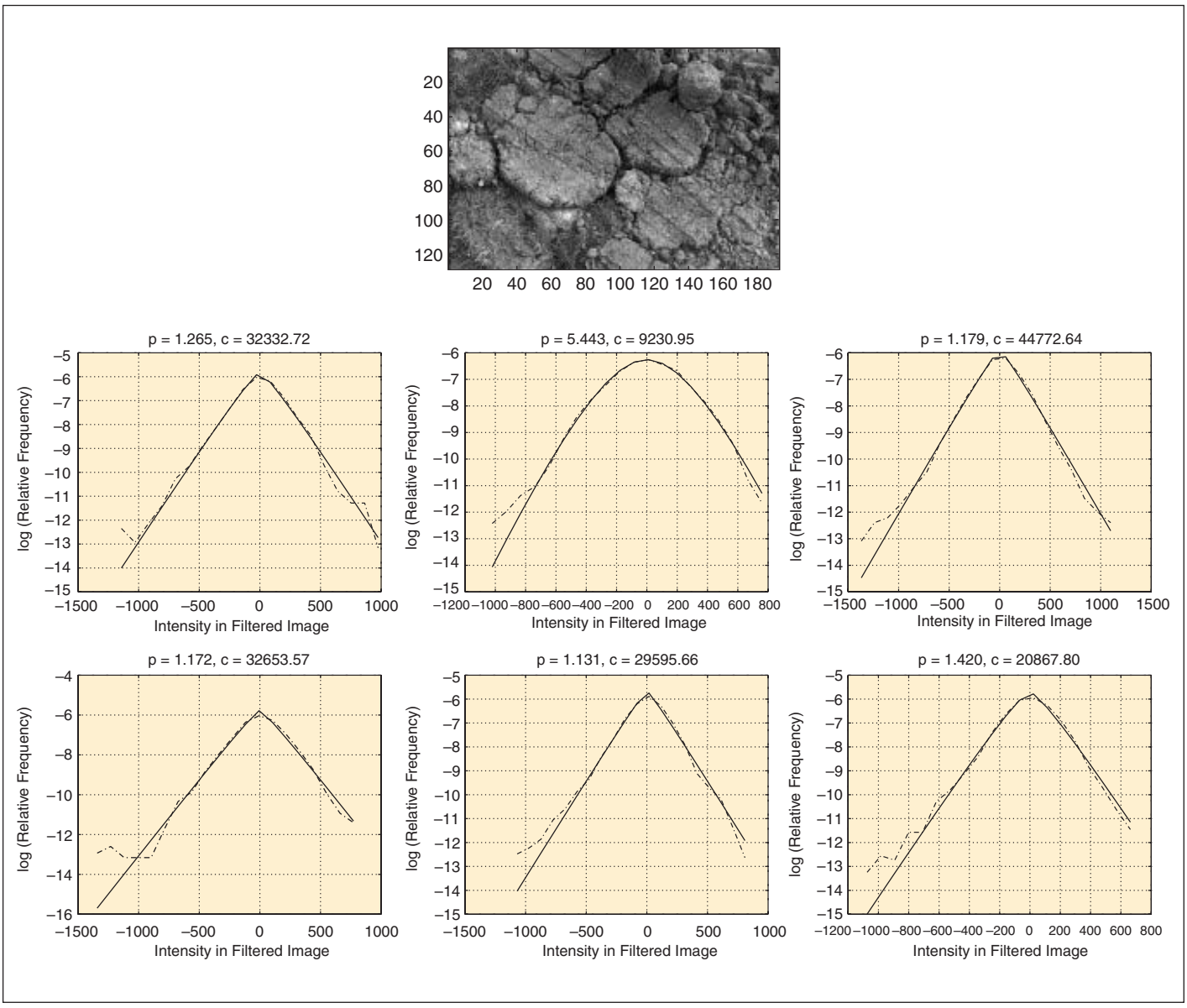
For the image shown in the top panel of Fig. 2, we have estimated Bessel K forms for the images filtered by different Gabor filters. The lower panels plot the marginals for different filter orientations ($\theta = 0, 30, 60, 90, 120$, and $150°$) while keeping the scale fixed. Again, the broken line plots the observed and the solid line plots the estimated densities.

Fig. 3 shows three examples of estimation for filtering by Laplacian-Gaussian filters; (a) shows the natural images from van Hateran database and (b) shows the corresponding estimated Bessel K forms. Fig. 4 shows estimation results for three infrared face images when filtered by Gabor filters. These results suggest the role of Bessel K forms in modeling images beyond the visual spectrum. Shown in Fig. 5 are some examples of range images taken from the Brown range database of Lee and Huang. The three images shown in (a) are filtered using some Gabor filters and the resulting densities are plotted in (b).

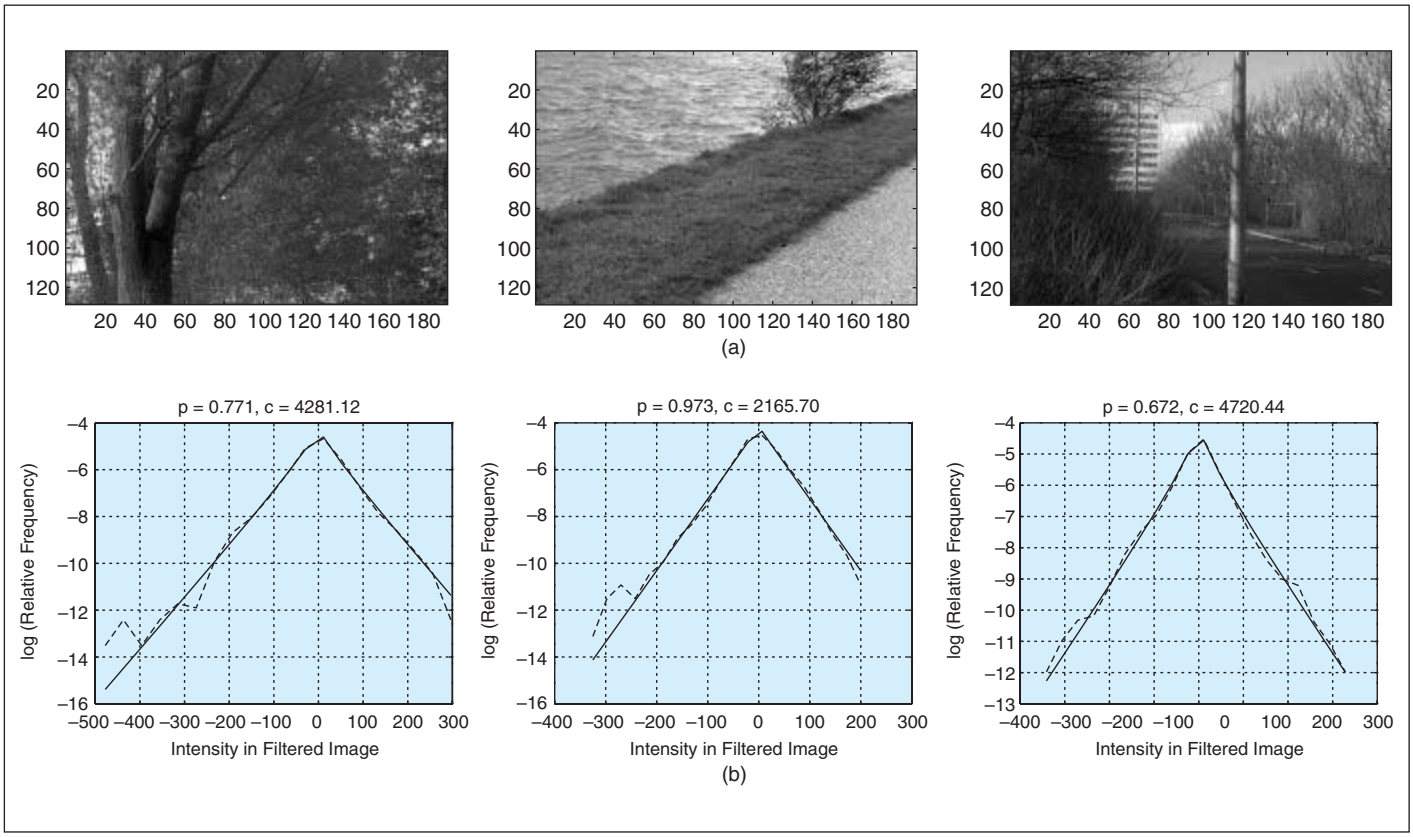
The moment-based estimator given in (4) is suspect to outliers and can be replaced by more robust estimation techniques that account for the outliers. A simple idea is to consider a fraction (say 1%) of the tails as outlier and discard it in parameter estimation. Another idea is to relate p and c to the quartiles of $f(x; p, c)$ and use the observed quartiles to estimate p and c.

Performance Analysis of Bessel K Forms

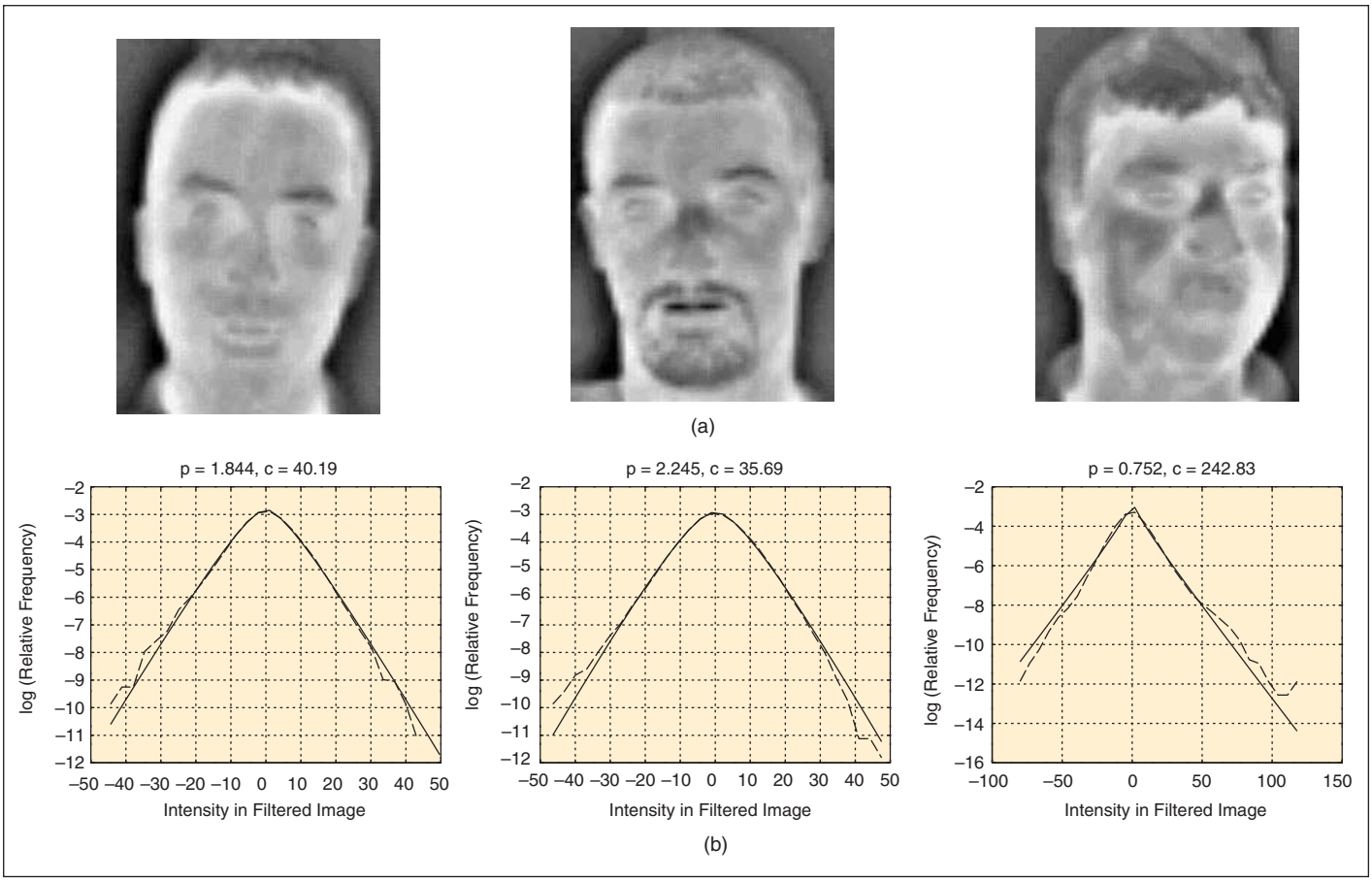
To quantify the performance in modeling observed histograms by estimated Bessel K forms, a number of quantities can be used and we have chosen the Kullback-Leibler (KL) divergence. For any two density functions f_1 and f_2, the KL divergence between f_1 and f_2 is defined as [27]:



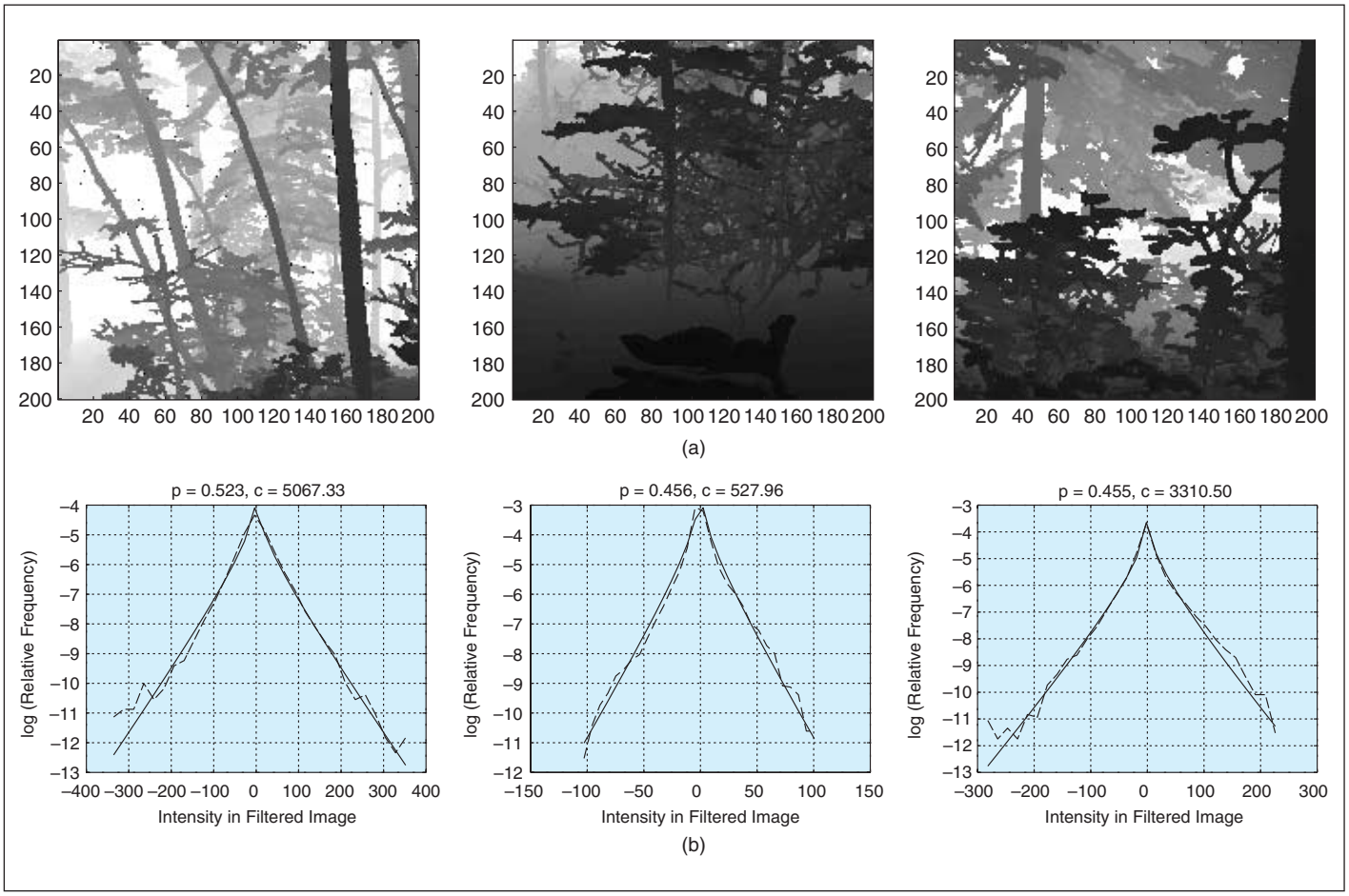
▲ *2. Plots of observed and estimated marginals (on a log scale) of the filtered versions of the image in top panel. Middle panels depict the marginals for different filter orientations: 0, 30, 60, 90, and 150° at a fixed scale.*



▲ *3. Estimated Bessel K forms for natural images (a) when filtered by Laplacian Gaussian filters.*



▲ *4. (a) IR face images filtered by arbitrary Gabor filters and their (b) observed and estimated marginal densities.*



▲ *5. (a) Range images of a forest and (b) corresponding observed and estimated marginal densities.*

$$D(f_1 \| f_2) = \int_{IR} \log\left(\frac{f_1(x)}{f_2(x)}\right) f_1(x)dx.$$

We have computed it by it at the center points of the histogram bins. To relate to the KL-divergence values for our application, we start with some examples. Shown in Fig. 6 are four plots, each containing a pair of densities: observed and estimated, and the KL divergence between them.

Using KL divergence for evaluating the matches between the observed and the estimated densities, we have computed the performance over two larger databases. In each case, for a large combination of images and filters drawn randomly, we have averaged the KL divergence over thousands of resulting filtered marginals. The first database is made up of 300 natural video images downloaded from van Hateran natural image database, and the second database is made up of 220 $\mathbb{R}$ face pictures. Shown in Fig. 7 are the convergence plots of the average KL divergence, plotted against the sample size; (a) is for the natural video images with a limiting value of 0.0719 while (b) is for the infrared images with a limiting value of 0.0479. A comparison of these values with the examples in Fig. 6 underscores the degree of match between the observed histograms and the estimated Bessel K forms.

Bessel Parameters Point to Objects

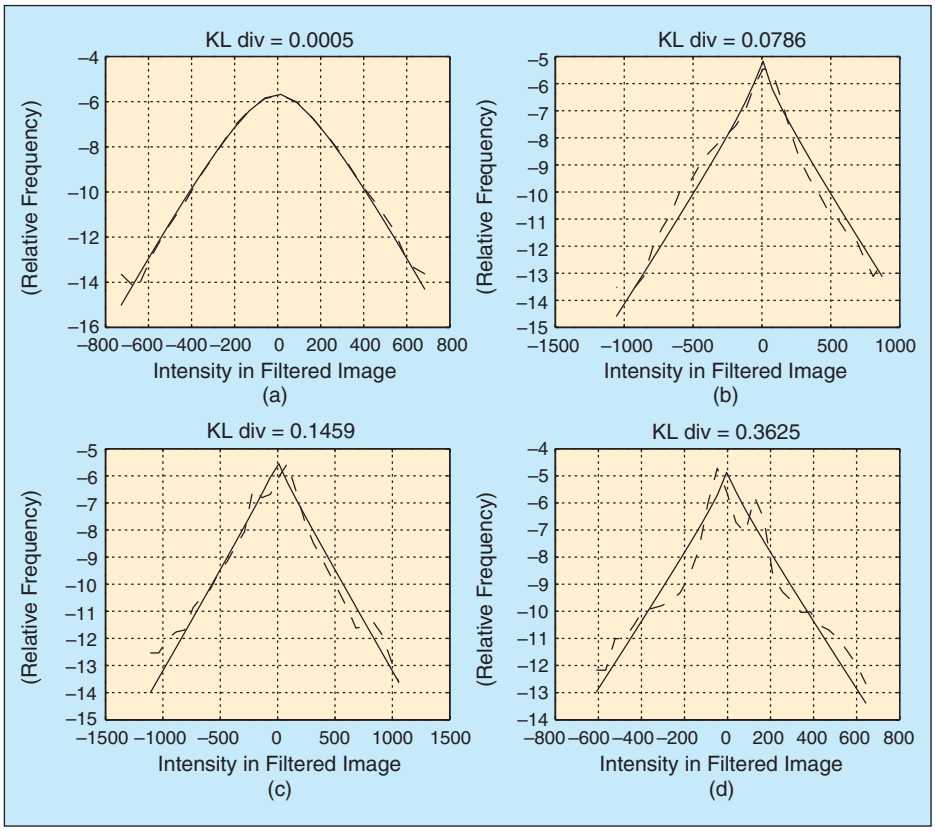
Since Bessel K forms are derived using object-based models, the Bessel parameters can be related to the objects present in the image. Of the two Bessel parameters, p value characterizes objects, and their extracted features, to some extent. To understand this relation, simplify by restricting to only one type of object being present in the image, i.e., in (1) set $g_i = g$ for all is. For the homogeneous Poisson model on z, the kurtosis of g is given by

$$\kappa = \frac{\int_{\mathbb{R}^2} g^4(z)dz}{\left(\int_{\mathbb{R}^2} g^2(z)dz\right)^2},$$

and it quantifies the distinctness of g with respect to its immediate background. As described in [22], if λ is the intensity of the Poisson placement process then the shape parameter of an image made of object g satisfies

$$p < 1 \Rightarrow \lambda < 6\kappa.$$

What does this relation imply? If a filter F is applied to an image I to extract some feature and the resulting intensity values are used in the estimation of p and c, then resulting p will depend upon two factors: i) distinctness



▲ *6. Examples of KL divergence: the divergence values for the four plots are (a) 0.0005, (b) 0.0786, (c) 0.1459, and (d) 0.3625, respectively.*

(quantified by κ) and ii) frequency of occurrence (quantified by λ) of that feature in that image. Objects with sharper, distinct edges have low p value (for a low λ) while scenes with lots of objects (large λ) will have large p values. For example, Fig. 8 shows a variation of p value when the images are filtered for extracting vertical edges ($\theta = 90$). The top row shows images with increasing frequency of vertical edges in going from left to right. Correspondingly, the estimated p value shows an increase (0.19, 0.71, and 2.0).

If an image I is filtered by J filters (to extract J distinct features), then $p^{(1)},\ldots,p^{(J)}$ characterize the distinctness and frequencies of those features. For a large value of J, this information may prove sufficient to characterize objects present in I.

Bivariate Probability Models

These univariate Bessel K forms can also be extended to specify multivariate densities of the filtered images. The basic idea, suggested by Grenander, is model all linear combinations of the pixels by Bessel K forms and then invoke the Cramer-Wold device [28] which states that specification of densities of all linear combinations of a number of random variables specifies uniquely their joint density. However, note that the set of Bessel K forms is not closed under convolution and the linear combinations may not have densities that are Bessel K forms. Our imposition of Bessel K forms on the linear combinations is only an approximation whose performance remains to be carefully investigated.

Let I_1 and I_2 be two pixels whose joint density we are interested in. For instance, these could be neighboring pixels in a filtered image, or pixels in two different filtered images. For $a_1, a_2 \in \mathbb{R}$, consider that random variable $I \equiv a_1 I_1 + a_2 I_2$. We will impose a Bessel K form as the probability density of this random variable and study the results. That is, I is modeled by a Bessel K form with a characteristic function:

$$\Phi(a_1, a_2; w) = \left(\frac{1}{1 + 0.5 w^2 c(a_1, a_2)} \right)^{p(a_1, a_2)},$$

where c and p can be estimated from the data as follows. The characteristic function of Bessel K forms is studied in [1]. Define $\mu_{i,j}$ as the moments:

$$\mu_{i,j} = \iint I_1^i I_2^j f(I_1, I_2) dI_1 dI_2, \quad 0 \le i, \quad j \le 4.$$

Similar to (4), the Bessel parameters for I can be estimated by:

$$p = \frac{3}{\text{kurtosis}(I) - 3}, \quad c = \frac{\text{variance}(I)}{p},$$

where the two required cumulants of I are given by

$$\text{variance}(I) = \sum_{k=0}^{2} \binom{2}{k} a_1^k a_2^{2-k} \mu_{k,2-k},$$

and

$$\text{kurtosis}(I) = \frac{\sum_{k=0}^{4} \binom{4}{k} a_1^k a_2^{4-k} \mu_{k,4-k}}{\sum_{k=0}^{2} \binom{2}{k} a_1^k a_2^{2-k} \mu_{k,2-k})^2}.$$

Given the eight required moments

$$\mu_{0,2}, \mu_{1,1}, \mu_{2,0}, \mu_{4,0}, \mu_{3,1}, \mu_{2,2}, \mu_{1,3}, \mu_{0,4},$$

the variance and the kurtosis of I, and therefore the parameters p and c are easily computed. Having obtained Φ,

we approximate the joint density of I_1 and I_2 by a function $f(I_1, I_2)$ that is related to Φ according to:

$$\Phi(a_1, a_2; w) = \int\int \exp^{(-j(a_1 I_1 + a_2 I_2)w)} f(I_1, I_2) dI_1 dI_2 .$$

For $w = 1$, and interpreting a_1, a_2 as the frequency variables, we obtain an estimate of the joint density by computing

$$\frac{1}{4\pi^2}\int\int \exp(j(a_1 I_1 + a_2 I_2))\Phi(a_1, a_2; 1) da_1 da_2 . \quad (5)$$

Shown in Fig. 9 is an example of this bivariate density estimation. For the picture shown in the top panel, we estimate the joint density of two filtered versions of I, resulting from two Gabor filters at the same scale but with orientations 0 and 15°. The left panels show the mesh and the contour plots of the estimated function $f(I_1, I_2)$ while the right panels show the observed bivariate density.

Metrics and Applications

We have chosen to represent images via the Bessel parameters of their filtered versions. One distinct advantage of having such analytical forms is the resulting theoretical framework for image analysis. For instance, we would like to be able to compare images by directly comparing their respective Bessel parameters. An analytical form can be very useful in the sense that it obviates the need to compute the histograms.

To quantify the difference between any two Bessel K forms, we have chosen the L^2-metric on $\mathcal{D}$. It is possible that other metrics, such as the KL divergence or the L^1 metric, may prove more useful in certain situations. Since we are restricting ourselves to only $\mathcal{D}$, and not the full set of probability density functions, we suggest that many of these choices will provide similar results, especially if the task is classification or hypothesis pruning. A parametric expression for the metric will be much faster to evaluate than the KL divergence evaluated using the histograms. The main drawback of choosing L^2 is that Bessel K forms are not in L^2 for $p < 0.25$. In the case of natural images, the p-values are mostly larger than 0.25, while for images of objects with sharp, well-defined edges, p can sometimes be below 0.25. In cases where $\hat{p} < 0.25$ for a image-filter combination, we can choose one of following: i) drop that filter, ii) approximate p (perhaps badly) by $0.25 + \varepsilon$, and then compute the L^2-metric, or iii) compute the L^2-metric numerically using the quadrature integration.

For $f(x; p_1, c_1)$ and $f(x; p_2, c_2)$ in $\mathcal{D}$, the L^2-metric

$$d(p_1, c_1, p_2, c_2) = \sqrt{\int_x (f(x; p_1, c_1) - f(x; p_2, c_2))^2 dx}$$

is given by: for $p_1, p_2 > 0.25$, $c_1, c_2 > 0$,

$$\left(\frac{\Gamma(0.5)}{2\sqrt{2\pi}}\left(\frac{\mathcal{G}(2p_1)}{\sqrt{c_1}} + \frac{\mathcal{G}(2p_2)}{\sqrt{c_2}} - \frac{2\mathcal{G}(p_1 + p_2)}{\sqrt{c_1}}\left(\frac{c_1}{c_2}\right)^{p_2}\mathcal{F}\right)\right)^{\frac{1}{2}}, \quad (6)$$

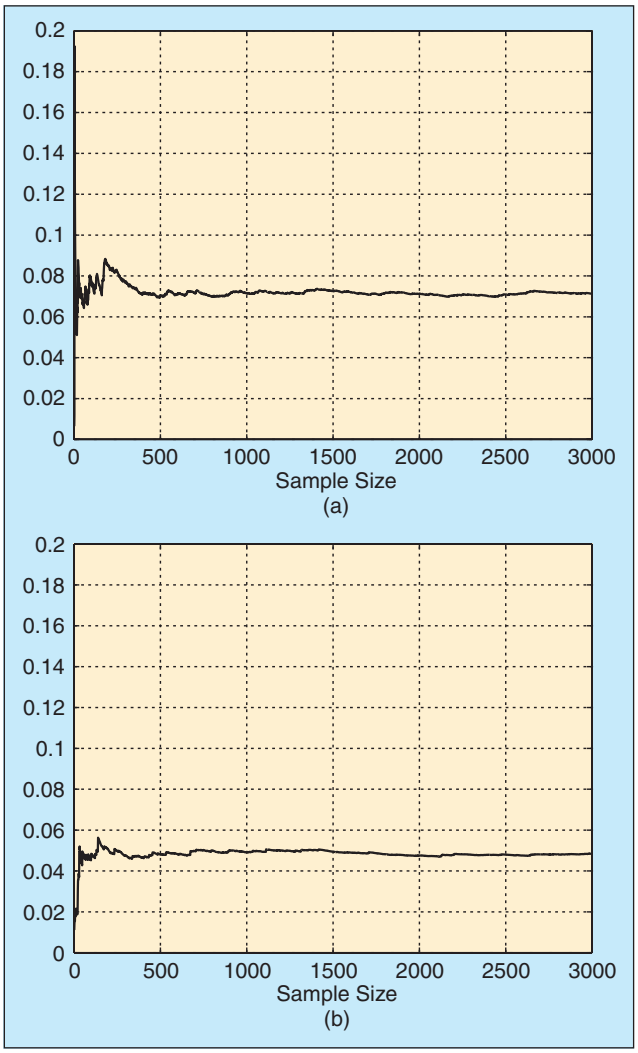
Image statistics under common representations, such as wavelets or subspace bases, point to non-Gaussian distributions.

where $\mathcal{G}(p) = (\Gamma(p - 0.5)/\Gamma)(p)$ and

$$\mathcal{F} = F\left((p_1 + p_2 - 0.5), p_2; p_1 + p_2; 1 - \frac{c_1}{c_2}\right).$$

F is the hypergeometric function. A derivation of this metric is presented in the appendix of [22].

This equation provides a metric between two Bessel K forms. It can be extended to a pseudo-metric on the image space as follows. For any two images, I_1 and I_2, and the filters $F^{(1)}, \ldots, F^{(J)}$, let the parameter values be given



▲ *7. Convergence of average KL-divergence between the observed and the estimated densities as the sample size increases. (a) van Hateran database of natural images and (b) FSU IR face database.*

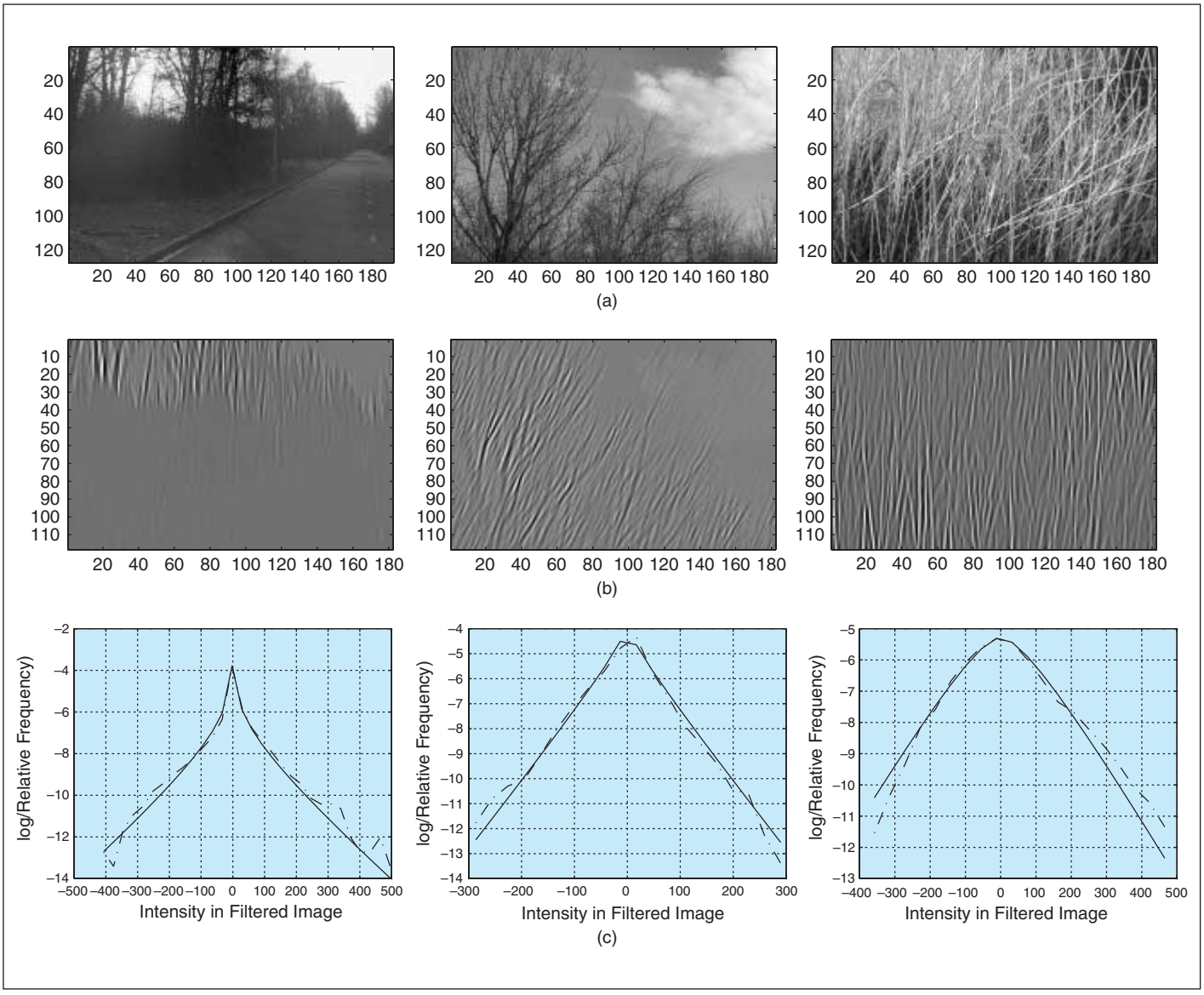
by: $(p_1^{(j)}, c_1^{(j)})$ and $(p_2^{(j)}, c_2^{(j)})$, respectively, for $j = 1,2,\ldots,J$. Then, a pseudo-metric between two images, is defined as

$$d_I(I_1, I_2) = \sqrt{\left(\sum_{j=1}^{J} d(p_1^{(j)}, c_1^{(j)}, p_2^{(j)}, c_2^{(j)})^2 \right)}. \quad (7)$$

Note that d_I is not a proper metric on the image space because two different images can have $d_I = 0$ between them. Also, d_I is dependent upon the choice of filters. The problem of selecting an optimal set of filters for solving a particular application, e.g., face recognition, is an interesting one although it is not pursued here. In this article, we have considered a fixed bank of filters consisting of Gabors, Laplacian, Gaussian, and the derivative filters. Additionally, it has been observed that different filtered versions of the same image are often highly correlated, and therefore, the Euclidean norm d_I may not be appropriate. In such cases, another choice such as the *max* of all components may be pursued.

Clutter Classification

An important application of this Bessel K representation is in the classification of clutter for automated target recognition (ATR) scenarios. In ATR, targets are imaged in cluttered environments, and some knowledge of clutter type, whether it is grass, buildings, trees, or roads, can help improve the performance in target recognition. We utilize the Bessel K forms to represent images and employ the metric defined in (7) to classify the clutter types from their images. We will demonstrate the strength of this model in the context of natural clutter classification. For a simple illustration, consider the images of natural clutter shown in Fig. 10. Using 27 small-scale Gabor filters ($J = 27$), for nine different orientations at three scales each,

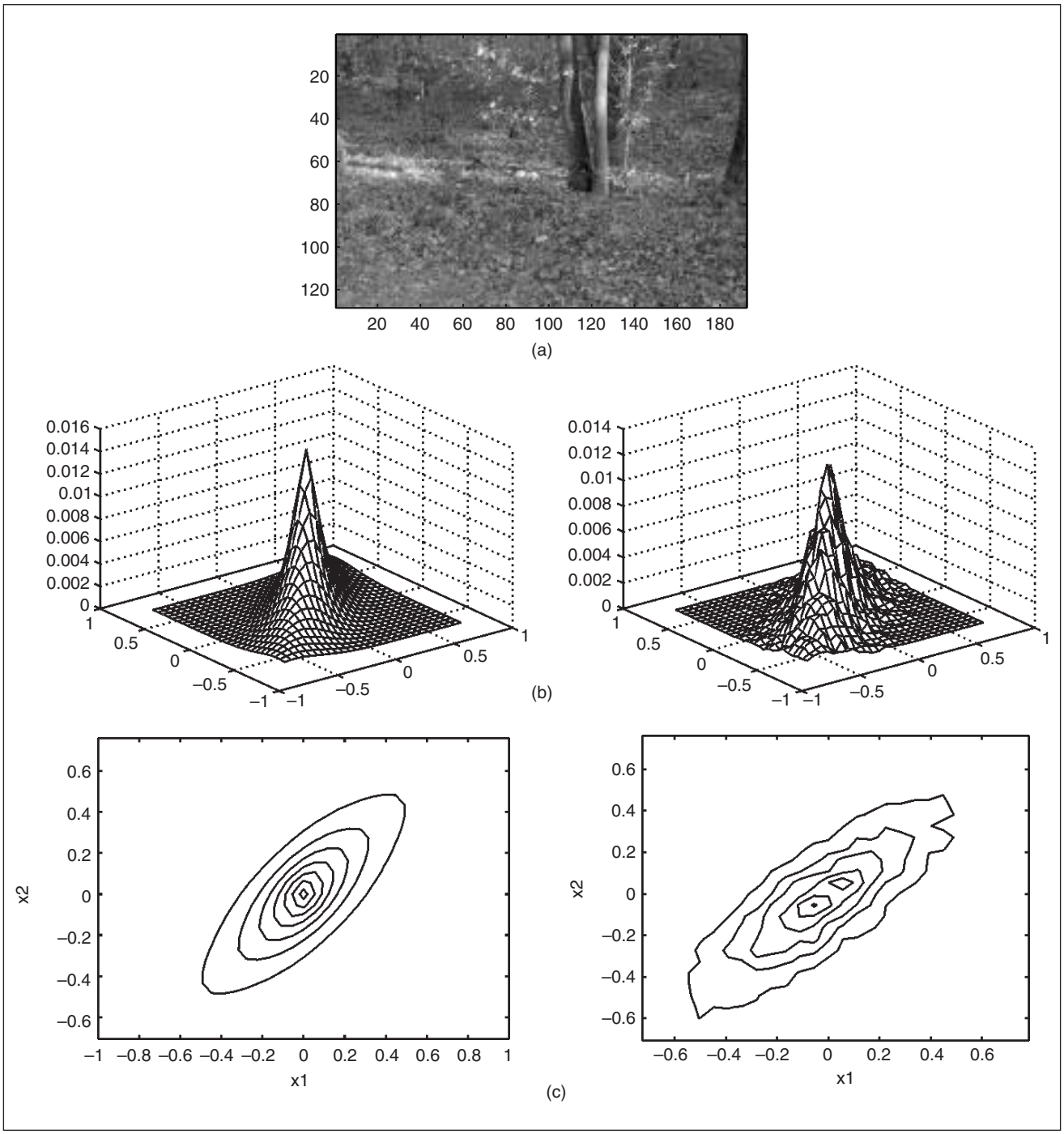


▲ *8. Variation of p-values for extracting edges: (a) the original images, (b) the filtered images, and (c) the densities (log-scale). The estimated p-values are: 0.19, 0.71, and 2.00, respectively.*

we have computed the pairwise distances d_I s. Shown in the bottom panel is the dendrogram clustering of images based on d_I. Dendrogram is a tool in MATLAB for clustering observations in a high-dimensional space when their pairwise distances are given. The clustering in Fig. 10 implies that I_5 and I_6 are clustered close together, I_1 and I_2 are clustered together, and finally I_3 and I_4 are clustered together. This clustering result can be applied to a classification problem as follows: Let the images I_1, I_3, and I_5 be the already classified images, and the remaining three form a set of test images (to be classified). Based on this clustering, I_2 will be classified in the same class as I_1, and so on.

ℝ *Face Recognition*

As another application of Bessel K forms, consider the problem of recognizing people from the frontal ℝ images of their faces. A natural way to proceed is to develop 3-D models of the faces and their thermal states and search over these physical representations for given test images. A facial surface can be represented by a dense polygonal



▲ *9. Bivariate probability densities: For the image shown in (a), the estimated (left) and observed (right) bivariate densities plotted as meshes (b) and contours (c).*

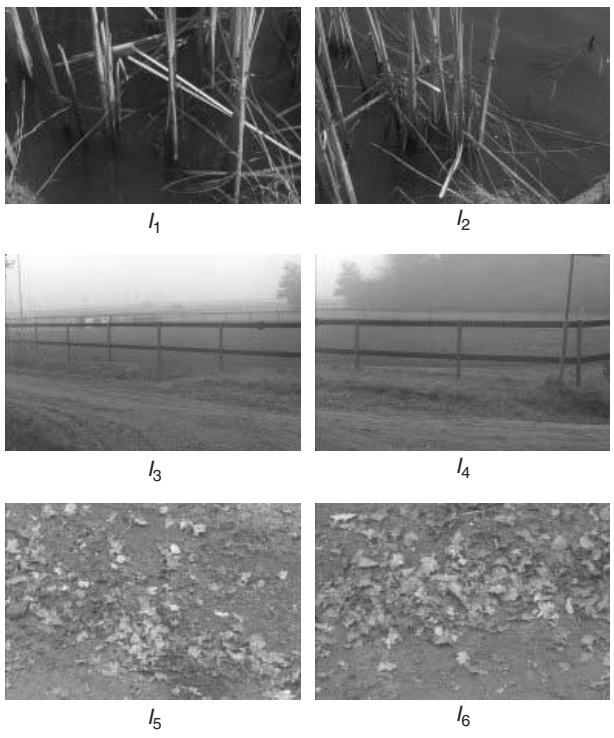
mesh and the thermal state can be denoted by a scalar (temperature) field on this mesh. In general, recognition of objects from their observed images corresponds to the selection of hypothesis (mesh template) in presence of the nuisance variables (pose, thermal state) [29]. This hypothesis selection is performed using detailed models involving physical geometries, thermal variables, pose, and motion [8], [30], [31]. Given an image, the task of searching over all possible 3-D templates is demanding and can benefit from a pruning that places significant probability only on a small subset of the possible targets. We will demonstrate the use of Bessel representations in pruning hypotheses for given face images.

Let $\mathcal{A}$ be the set of all possible objects that can be present in an image, and let S be the group of nuisance variables such as pose, translation, and thermal variables. Define a probability mass function on $\mathcal{A}$ according to: for an $\alpha \in \mathcal{A}$

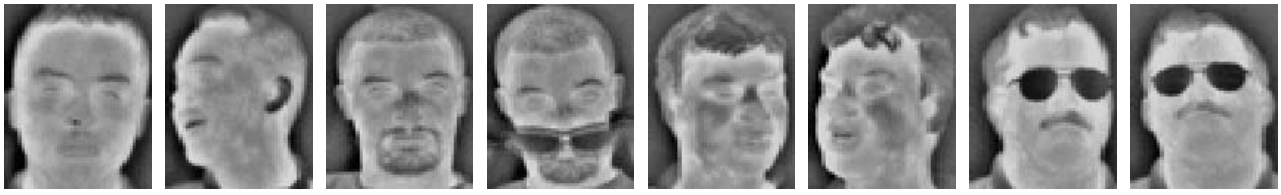
$$P(\alpha|I) = \frac{1}{Z}\exp\left(-\min_{s\in S}\sum_{j=1}^{J} d(p_{obs}^{(j)}, c_{obs}^{(j)}, p_{\alpha,s}^{(j)}, c_{\alpha,s}^{(j)})^2)/D)\right), \tag{8}$$

where Z is a normalizer and D controls our confidence (analogous to the temperature in simulated annealing) in this probability. Here $(p_{obs}^{(j)}, c_{obs}^{(j)})$ are the estimated parameters for the filtered image $I^{(j)}$, and $(p_{\alpha,s}^{(j)}, c_{\alpha,s}^{(j)})$ are the estimated Bessel parameters for the filter $F^{(j)}$ and the target α rendered at the nuisance variable $s \in S$. Note that $(p_{\alpha,s}^{(j)}, c_{\alpha,s}^{(j)})$ can be precomputed offline for all $\alpha \in A$, $j \in \{1,2,\ldots,J\}$, and a finite subset of S.

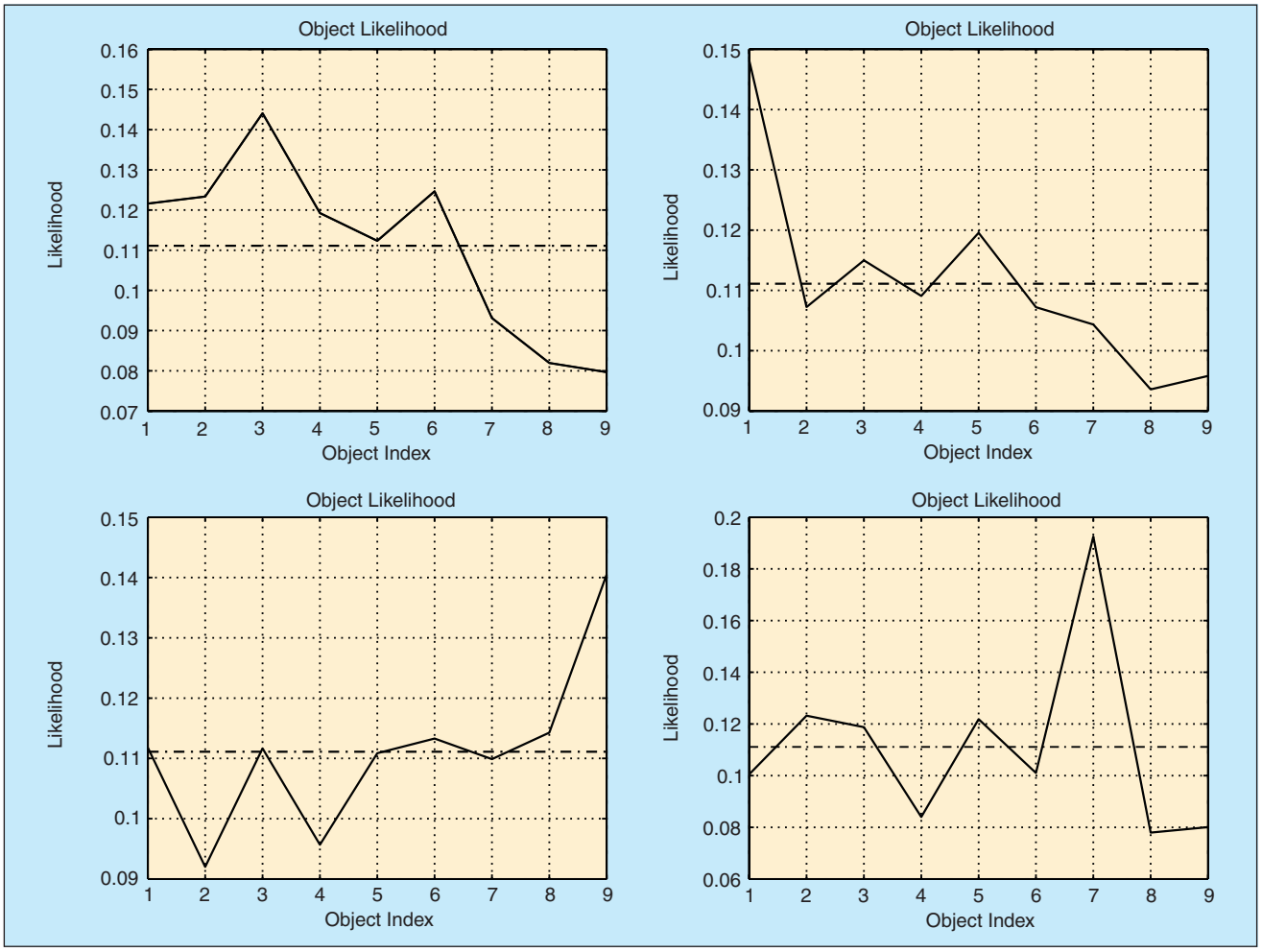
Our experiment used a database of 180 images, 20 pictures each for nine people, some of which are shown in Fig. 11. We divided the database into nonoverlapping training and test sets. Part of the images are used for training and the remaining are used for testing. We have used $J = 39$ filters, including the gradient filters, the Laplacian of Gaussian filters, and the Gabor filters. For each image of the object α at the pose s in the training set, we estimate $(p_{\alpha,s}^{(j)}, c_{\alpha,s}^{(j)})$, for each filter $F^{(j)}$. Then, given a test image I, the estimated parameters $(p_{obs}^{(j)}, c_{obs}^{(j)})$ are used to compute the probability $P(\alpha|I)$ according to (8). Shown in Fig. 12 are the plots of $P(\alpha|I)$ versus α (at $D = 0.3$) for four different images I in the test set. If the goal is to prune the hypothesis set A, then all objects with $P(\alpha|I)$ larger than some threshold, say 0.111, can be short listed. As an ex-



▲ *10. For the images shown on the left, a dendrogram clustering plot using Bessel K forms is shown.*



▲ *11. Examples from the FSU IR face database.*



▲ *12. Plots of $P(\alpha|I)$ for three images. For thresholding at 0.111, only the subjects above the dotted line are short listed.*

ample, the plot in top left is for an image I of the third subject ($\alpha = 3$) and the plot in top right corresponds to $\alpha = 1$. For the latter case, a shortlisting by thresholding leaves only three possible hypotheses, as compared to the original nine hypotheses.

Principal component analysis [2] has been used widely to derive a low dimensional representation of face images with application in recognition using $\mathbb{R}$ images [32]. More recently, independent component analysis [3] has become popular as a dimension reduction tool that can incorporate high-order statistical structures. To demonstrate the effectiveness of our representation, we have performed a comparison of the Bessel forms with the eigen and independent face representations using the nearest neighbor classifier. Table 1 shows the recognition performance of the three methods. For details of this experiment and an analysis of results, please refer to [33]. These experiments indicate that the Bessel K forms result in the best recognition rate among the three methods under general test conditions.

Table 1. Recognition Performance.

Test/Training Ratio	Eigen Races	Independent Faces	Bessel K Forms
1:1	90.48%	89.52%	97.14%
3:1	87.42%	88.05%	91.20%
7:1	80.21%	70.05%	83.96%

Conclusions

This article studies the application of Bessel K forms in modeling the uni- and bi-variate density functions of the filtered image pixels. For a collection of filters (Gabor, Laplacian of Gaussian, and derivative filters), experimental results demonstrate a close fit between the observed histograms and the estimated parametric form. These parametric forms can be used in image understanding: classification of images or recognition of objects in the images.

Acknowledgments

Parts of this research were performed in collaboration with Prof. U. Grenander of Brown Univ. and Prof. X. Liu of Florida State Univ. We thank X. Liu, C. Hesher, and B. Thomasson for some of the data/results presented here. We also thank the producers of the van Hateren database,

the Brown range image database, and the FSU infrared face database for making these images public. This research was partially supported by the grants ARO DAAD19-99-01-0267, NMA 201-01-2010, and NSF DMS 0101429.

Anuj Srivastava is an Assistant Professor in Department of Statistics at Florida State University (FSU). He obtained his Ph.D. in electrical engineering from Washington University in 1996 and was a visiting researcher in Division of Applied Mathematics at Brown University during 1996-1997. He directs the Laboratory for Computational Vision at FSU . His research interests lie in the areas of statistical signal processing and image analysis, statistics on manifolds, and computational statistics.

References

[1] U. Grenander and A. Srivastava, "Probability models for clutter in natural images," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 23, pp. 424-429, Apr. 2001.

[2] M. Kirby and L. Sirovich, "Application of the Karhunen-Loeve procedure for the characterization of human faces," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 12, no. 1, pp. 103-108, 1990.

[3] P. Comon, "Independent component analysis, A new concept?," *IEEE Signal Processing Mag.,* vol. 36, no. 3, 1994.

[4] A.J. Bell and T.J. Sejnowski, "The independent components of natural scenes are edge filters," *Vision Research*, vol. 37, no. 23, pp. 3327-3338, 1997.

[5] B.A. Olshausen and D.J. Field, "Sparse coding with an overcomplete basis set: A strategy employed by v1?," *Vision Research*, vol. 37, no. 23, pp. 3311-3325, 1997.

[6] P.N. Belhumeur, J.P. Hepanha, and D.J. Kriegman, "Eigenfaces vs. fisherfaces: Recognition using class specific linear projection," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 19, no. 7, pp. 711-720, 1997.

[7] S. Mallat, "Multiresolution approximations and wavelet orthonormal basis of $l^2(r)$," *Trans. Amer. Math. Soc.*, vol. 315, pp. 69-87, 1989.

[8] A. Srivastava, M.I. Miller, and U. Grenander, "Bayesian automated target recognition," *Handbook of Image and Video Processing.* San Diego, CA: Academic, pp. 869-881, 2000.

[9] S.G. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 11, pp. 674-693, July 1989.

[10] D. Mumford, "Empirical investigations into the statistics of clutter and the mathematical models it leads to," Lecture for the review of ARO Metric Pattern Theory Collaborative, Brown Univ., Providence, RI, 2000.

[11] M.J. Wainwright, E.P. Simoncelli, and A.S. Willsky, "Random cascades on wavelet trees and their use in analyzing and modeling natural images," *Appl. Comput. Harmon. Anal.*, vol. 11, pp. 89-123, 2001.

[12] D.J. Field, "Relations between the statistics of natural images and the response properties of cortical cells," *J. Opt. Soc. Amer.*, vol. 4, no. 12, pp. 2379-2394, 1987.

[13] B.A. Olshausen and D.J. Field, "Natural image statistics and efficient coding," *Netw.: Computation Neural Syst.*, vol. 7, 1996.

[14] B.W. Silverman, *Density Estimation for Statistics and Data Analysis*. London, U.K.: Chapman & Hall, 1985.

[15] D.J. Heeger and J.R. Bergen, "Pyramid-based texture analysis/synthesis," in *Proc. SIGGRAPHS*, Los Angeles, CA, 1995, pp. 229-238.

[16] S.C. Zhu, Y.N. Wu, and D. Mumford, "Minimax entropy principles and its application to texture modeling," *Neural Comput.*, vol. 9, no. 8, pp. 1627-1660, Nov. 1997.

[17] C. Chubb, J. Econopouly, and M.S. Landy, "Histogram contrast analysis and the visual segregation of iid textures," *J. Opt. Soc. Am. A*, vol. 11, pp. 2350-2374, 1994.

[18] J. Portilla and E.P. Simoncelli, "A parametric texture model based on join statistics of complex wavelet coeeficients," *Int. J. Comput. Vision*, vol. 40, no. 1, pp. 49-70, 2000.

[19] A.B. Lee and D. Mumford, "Occlusion models for natural images: A statistical study of scale-invariant dead leaves model," *Int. J. Comput. Vision*, vol. 41, no. 1, 2, 2001.

[20] D.L. Donoho and A.G. Flesia, "Can recent innovations in harmonic analysis 'explain' key findings in natural image statistics," *Netw.: Computation Neural Syst.*, vol. 12, no. 3, pp. 371-393, 2001.

[21] G. Winkler, *Image Analysis, Random Fields and Dynamic Monte Carlo Methods*. Berlin, Germany: Springer, 1995.

[22] A. Srivastava, X. Liu, and U. Grenander, "Universal analytical forms for modeling image probability," *IEEE Trans. Pattern Anal. Machine Intell.*, to be published.

[23] B. Jahne, H. Haubecker, and P. Geibler, *Handbook of Computer Vision and Applications,* vol. 2. San Diego, CA: Academic, 1999.

[24] D. Marr, *VISION: A Computational Investigation into the Human Representation and Processing of Visual Information*. San Francisco, CA: Freeman, 1982.

[25] O. Barndorff-Nielsen, J. Kent, and M. Sorensen, "Normal variance-mean mixtures and z distributions," *Int. Statist. Rev.*, vol. 50, pp. 145-159, 1982.

[26] H. van Hateren, "Natural stimuli collection: A public database of natural images."Available: http://hlab.phys.rug.nl/

[27] T.M. Cover and J.A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.

[28] P. Billingsley, *Probability and Measure*. New York: Wiley, 1995.

[29] U. Grenander, A. Srivastava, and M.I. Miller, "Asymptotic performance analysis of Bayesian object recognition," *IEEE Trans. Inform. Theory*, vol. 46, no. 4, pp. 1658-1666, 2000.

[30] A. Lanterman, M. Miller, and D. Snyder, "General Metropolis-Hastings jump diffusions for automatic target recognition in infrared scenes," *Opt. Eng.*, vol. 36, no. 4, pp. 1123-1137, 1997.

[31] M.L. Cooper, U. Grenander, M.I. Miller, and A. Srivastava, "Accommodating geometric and thermodynamic variability for forward-looking infrared sensors," *Proc. SPIE, Algorithms for Synthetic Aperture Radar IV*, vol. 3070 pp. 162-172, Apr. 1997.

[32] R. Cutler, "Face recognition using infrared images and eigenfaces," 1996. Available: http://www.cs.umd.edu/\~rgc/face/face.htm

[33] A. Srivastava, X. Liu, B. Thomasson, and C. Hesher, "Spectral probability models for infrared images and their applications to ir face recognition," in *Proc. Workshop Computer Vision Beyond Visual Spectrum*, Kauai, HI, 2001.