

Analytical Image Models and Their Applications

Anuj Srivastava¹, Xiuwen Liu², and Ulf Grenander³

¹ Department of Statistics, Florida State University, Tallahassee, 32306

² Department of Computer Science, Florida State University, Tallahassee, FL 32306

³ Division of Applied Mathematics, Brown University, Providence, RI 021912

Abstract. In this paper, we study a family of analytical probability models for images within the spectral representation framework. First the input image is decomposed using a bank of filters, and probability models are imposed on the filter outputs (or spectral components). A two-parameter analytical form, called a **Bessel K form**, derived based on a generator model, is used to model the marginal probabilities of these spectral components. The Bessel K parameters can be estimated efficiently from the filtered images and extensive simulations using video, infrared, and range images have demonstrated Bessel K form's fit to the observed histograms. The effectiveness of Bessel K forms is also demonstrated through texture modeling and synthesis. In contrast to numeric-based dimension reduction representations, which are derived purely based on numerical methods, the Bessel K representations are derived based on object representations and this enables us to establish relationships between the Bessel parameters and certain characteristics of the imaged objects. We have derived a pseudo-metric on the image space to quantify image similarities/differences using an analytical expression for L^2 -metric on the set of Bessel K forms. We have applied the Bessel K representation to texture modeling and synthesis, clutter classification, pruning of hypotheses for object recognition, and object classification. Results show that Bessel K representation captures important image features, suggesting its role in building efficient image understanding paradigms and systems.

Keywords: Image features, spectral analysis, Bessel K forms, clutter classification, object recognition.

1 Introduction

In the last few decades, statistical approach has become one of the dominating methods for computer vision and image understanding. Central to the promising success of statistical techniques in image understanding are efficient probability models for the observed images. To realize statistical inference algorithms efficiently, dimension reduction is required due to the high dimensionality of the image space. There are in general two approaches for dimension reduction: purely numerical, non-physical methods and physical-based methods. Purely numerical methods, which are closely related to the bottom-up approach for computer

vision and image understanding, use one of the many techniques for dimension reduction by treating images as elements of a vector space and seeking a low-dimensional subspace that best represents those numbers under some chosen criteria. Principal components [12], independent components [4,3], sparse coding [16], Fisher’s discriminant [2], local linear embedding [19], and many other statistical learning algorithms are all instances of this idea. The main advantage is the computational efficiency and the main drawback is knowledge deficiency. Lack of physical or contextual information leads to a limited performance, specially in challenging situations. In the second approach, which is closely related to the top-down approach for image understanding, images are characterized by the physical characteristics of the objects and the resulting physical variables are used to analyze images. An example of this idea is the *deformable template* theory [6] where images are studied through the transformations that match the templates to the observations. One drawback is that they are computationally expensive to implement, since they require synthesis of hypothesized images for image analysis.

In this paper, we study a framework that provides some interactions between the numeric-based and the template-based approaches. Consider a deformable template representation of the imaged objects, as laid out in [9,21]. The basic idea is that images are made up of objects, and their variability can be represented by physical variables. Using 3D models of objects, all occurrences of these objects can be generated using similarity transformations. 3D scenes containing these transformed objects lead to 2D images via occlusion and projection. To build probability models on I , we seek analytical forms that retain some physical considerations, although not as explicitly as the template approach. We replace 3D templates by their 2D profiles (called **generators**) and denote them as g ’s. Let \mathcal{G} be the space of all possible generators associated with all objects, imaged from all angles. Random translation of 3D objects in a scene is modeled by random placements and scalings of g ’s in an image.

Each object contributes to the pixel value $I(z)$ according to $a_i g_i(\frac{1}{\rho_i}(z - z_i))$. Here $z \in W \equiv [0, L] \times [0, L]$ is a variable for pixel location, $g_i : W \mapsto \mathbb{R}_+$ is a generator of a randomly chosen object, $\rho_i \in [0, L]$ is a random scale, and $a_i \in \mathbb{R}$ is a random weight associated with g_i , which is drawn from the generator set \mathcal{G} according to some measure dG . The image formation is now modeled by:

$$I(z) = \sum_i^n a_i g_i\left(\frac{1}{\rho_i}(z - z_i)\right), \quad z, z_i \in W, \quad a_i \in \mathbb{R}, \quad \rho_i \in [0, L]. \quad (1)$$

Since g_i ’s are assumed unknown, the related variables n , ρ_i ’s and z_i ’s are also indeterminable. We aim to derive probability models on I by implicitly incorporating their variability.

Motivated by a growing understanding of animal vision, a popular strategy has been to decompose images into their spectral components using a family of bandpass filters. Similarly, our probability model on I will be through its spectral representation. In the context of texture modeling, the marginal distributions are often chosen as sufficient statistics as the frequencies of values in

the filtered images are relevant and the location information is discarded [11,10, 25]. Simoncelli et al. [18] have suggested using the lower order statistics (mean, variance, skewness, kurtosis) to specify the marginal densities of the wavelet coefficients of the images. Wainwright et al. [22] have studied a family of Gaussian mixtures, for different mixing densities, for modeling the observed histograms. Lee and Mumford [13] have presented a model for capturing the statistics in the images of leaves.

Using a physical model for image formation, we have extended a two-parameter probability model [8], to a full spectrum of bandpass filters and arbitrary images, called **Bessel K forms** [20]. We demonstrate the success of Bessel K forms in modeling the spectral components for video, infrared (IR), and range images of natural and artificial scenes and use the models for hypothesis pruning and object classification.

This paper is organized as follows. Section 2 applies Bessel K forms to model spectral components of images and associates the estimated Bessel K parameters with the observed shapes. Section 3 derives an L^2 -metric on the Bessel K forms and on the image space, while Section 4 applies these metrics to texture modeling and synthesis, clutter classification, hypothesis pruning, and object classification. Section 5 concludes the paper.

2 Analytical Probability Models for Image Spectra

Given an image I and a bank of filters $\{F^{(j)}, j = 1, 2, \dots, K\}$, we compute, for each filter $F^{(j)}$, a filtered image $I^{(j)} = I * F^{(j)}$, where $*$ denotes the 2D convolution operation. In this paper, we mainly use Gabor and Laplacian of Gaussian filters and do not address the issue of filter selection to best accomplish a specific task. Other filters can also be used as long as the resulting marginals are: (i) unimodal with the mode at zero, (ii) symmetric around zero, and (iii) are leptokurtic, i.e. their kurtosis is larger than that of a Gaussian random variable with the same variance.

2.1 Analytical Models

Applying 2D convolution to both sides of (1), we obtain a spectral component

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

The *conditional* density of $I^{(j)}(z)$, given the Poisson points $\{z_i\}$, the scales $\{\rho_i\}$, and the profiles g_i 's, is normal with mean zero and variance u , where $u \equiv \sum_i (g_i^{(j)} (\frac{1}{\rho_i} (z - z_i)))^2$. Under this model and assuming u to be a scaled-Gamma random variable, the density function of $I^{(j)}(z)$ has been shown to be [8]: for $p > 0$, $c > 0$,

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

where K is the modified Bessel function¹, Z is the normalizing constant given by $Z(p, c) = \sqrt{\pi}\Gamma(p)(2c)^{0.5p+0.25}$, and Γ is the gamma function. 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 the **Bessel K forms** and the parameters (p, c) as the **Bessel parameters**. The elements of \mathcal{D} are symmetric and unimodal for the mode at zero. For $p = 1$, $f(x; p, c)$ is the density of a double exponential. In general, it is the p^{th} convolution power (for any $p > 0$) of a double exponential density. Therefore, it is unimodal with the mode at $x = 0$. For the same reason, it is symmetric around zero. One limitation of the Bessel K forms is that they are square-integrable only for $p > 0.25$. This property is due to the choice of Gamma density for u ; as a result, the L^2 -metric of the Bessel K forms is applicable when p -values are larger than 0.25.

As shown in [8], p and c can be estimated using moment estimator, which is given by

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

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 moments of the filtered image. This moment estimator is sensitive to outliers and more robust estimators such as maximum likelihood estimator or robust estimation techniques can be used if needed.

We illustrate estimation results for a variety of images. Shown in the top panels of Fig. 1 are some images taken from the Groningen database. The middle panels display their specific filtered forms (or the spectral components) for Gabor filters chosen at arbitrary orientations and scales, and the bottom panels plot the marginal densities. On a log scale, the observed densities (histograms) are plotted in solid lines with dots and the estimated Bessel K forms ($f(x; \hat{p}, \hat{c})$) are plotted in solid lines.

Fig. 2(a) shows estimation results for two IR face images when filtered by Gabor filters. These results suggest the role of Bessel K forms in modeling images beyond the case of video images of natural scenes.

Shown in Fig. 2(b) are two examples of estimating marginal densities for the case of range images taken from the Brown range database. The images shown in top panels are filtered using Gabor filters and the resulting densities are plotted in the bottom panels.

A distinctive advantage of Bessel K representation, compared to numerical-based low-dimensional representations, is that Bessel K parameters can be related to physical characteristics of objects consisting of the observed image. Theoretically, the physical characteristics of the imaged objects, and the filter used in generating a spectral component, should dictate the resulting Bessel K

¹ Defined as

$$K_\nu(xy) = \frac{\Gamma(\nu + 0.5)(2y)^\nu}{\Gamma(0.5)x^\nu} \int_0^\infty \frac{\cos(xz)}{(z^2 + y^2)^{\nu+0.5}} dz,$$

for $\Re(\nu) > -0.5$, $x > 0$, and $|\arg(y)| < \frac{\pi}{2}$.

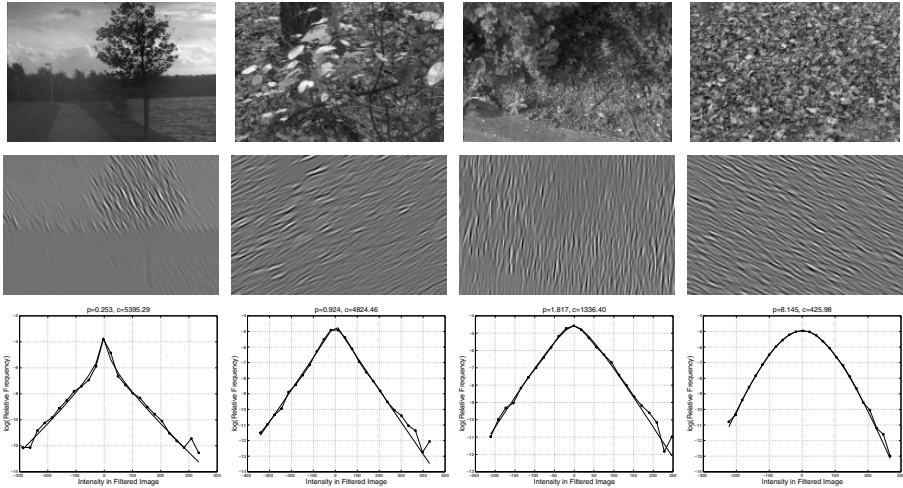


Fig. 1. Images (top panels), their Gabor components (middle panels), and the marginal densities (bottom panels). The observed densities are drawn in solid lines with dots and the estimated Bessel K forms are drawn in solid lines.

form. Since c is essentially a scale parameter relating to the range of pixels values in I , its role is not as important as p in image understanding.

Under certain assumptions of the generator model given in (2), specifically when $a_i \sim N(0, 1)$, ρ_1 is fixed to be 1.0, and all the $g_i \equiv g$ (i.e. a fixed generator), then

$$p = \frac{1}{\frac{\kappa}{3\lambda} - 1}, \text{ where } \kappa = \frac{\left(\int_W g(z_1)^4 dz_1\right)}{\left(\int_W g(z_1)^2 dz_1\right)^2}. \quad (5)$$

This equation provides an important relationship between a generator g and the parameter p . According to (5), $p < 1$ occurs when $\lambda < \frac{\kappa}{6}$. If the generator g has sharp, distinct boundaries (i.e. κ is larger) then the p value is small unless the frequency of occurrence (λ) is large. Specifically, if a filter $F^{(j)}$ is used to extract a particular feature (e.g. oriented edges, junctions, bands, etc.) from the image I , then p is dictated by the **distinctness** (κ) and the **frequency of occurrence** (λ) of that feature in the image. For example, shown in Fig. 3 is 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 (0.31, 0.77, 1.45, and 2.73). Summarizing the relation between p and κ , we have:

$$\text{If } \begin{cases} 0 < \lambda < \kappa/6 & \text{then } p < 1 \\ \kappa/6 < \lambda < \kappa/3 & \text{then } p > 1 \end{cases} .$$

2.2 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 choose the Kullback-

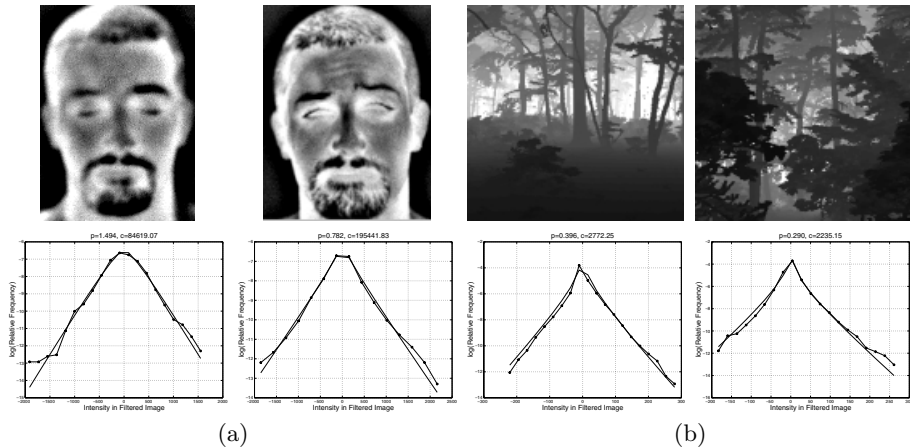


Fig. 2. Observed and estimated marginal densities (bottom panels) for (a) the IR face images (top panels) (b) range images of a forest of randomly chosen Gabor filters.

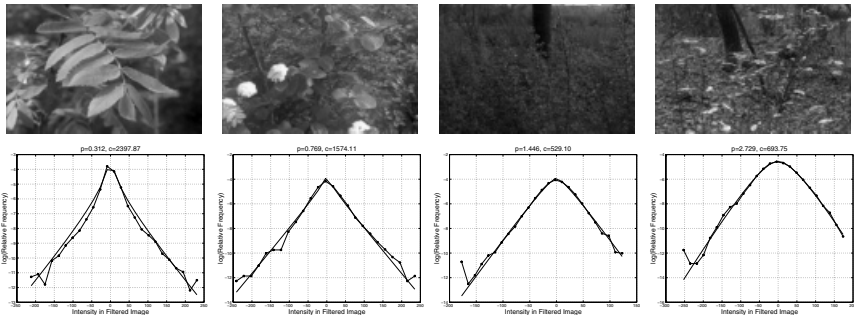


Fig. 3. Variation of p -values for extracting vertical edges ($\theta = 90$). Top panels are the original images, middle panels are the filtered images, and the bottom panels are the densities (log-scale). The estimated p -values are: 0.31, 0.77, 1.45, and 2.73, respectively.

Leibler (KL) divergence. For any two density functions f_1 and f_2 , the divergence is defined as the quantity: $KL(f_1, f_2) = \int_{\mathcal{R}} f_1(x) \log(f_1(x)/f_2(x)) dx$. We have computed it by discretizing at the center points of the histogram bins. To evaluate match between the observed and the estimated densities, we have computed the KL divergence for two large 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 Groningen natural image database, and the second database is made up of 220 IR face pictures. Shown in Fig. 4 are the convergence plots of the average KL divergence, plotted against the sample size. The top plot is for the natural video images with a limiting value of 0.0657 while the bottom plot is for the infrared images with a limiting value of

0.0478. A comparison of these values underscores the degree of match between the observed histograms and the estimated Bessel K forms.

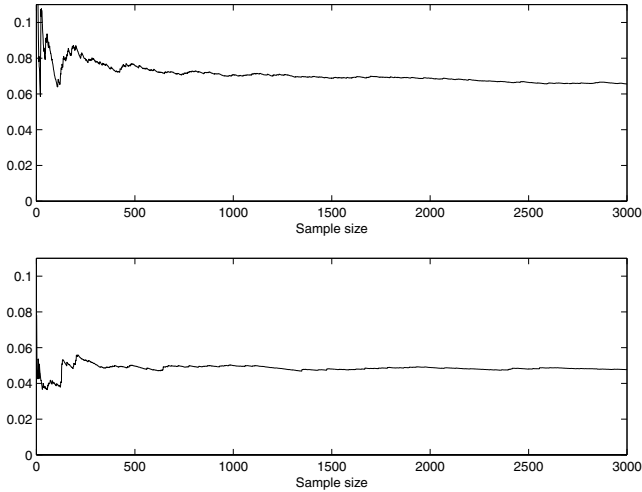


Fig. 4. Convergence of average KL-divergence between the observed and the estimated densities as the sample size increases. The top plot is for the Groningen database of natural images and the bottom plot is for the FSU IR face database.

3 Pseudo-Metrics for Comparing Images

We have chosen to represent images via the Bessel parameters of their spectral components. One distinct advantage, of having such analytical forms for the marginals of the spectral components, is the resulting theoretical framework for image analysis. To quantify the distance between two Bessel K forms, we have chosen the L^2 -metric on \mathcal{D} . It is possible that other metrics, such as the Kullback-Leibler 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 pdfs, we suggest that many of these choices will provide similar results, specially if the task is classification or hypothesis pruning. The main constraint of choosing L^2 is that Bessel K forms are not in L^2 for $p < 0.25$. In cases where the estimated $p < 0.25$, we can choose one of following: (i) drop that filter, (ii) approximate p (perhaps badly) by $0.25 + \epsilon$, 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 is $d(p_1, c_1, p_2, c_2) = \sqrt{\int_x (f(x; p_1, c_1) - f(x; p_2, c_2))^2 dx}$. Given $p_1, p_2 > 0.25$, $c_1, c_2 > 0$, this metric can be computed using the following closed form expression,

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

where $\mathcal{G}(p) = \frac{\Gamma(p-0.5)}{\Gamma(p)}$, $\mathcal{F} = F((p_1 + p_2 - 0.5), p_2; p_1 + p_2; 1 - \frac{c_1}{c_2})$, F is the hypergeometric function, and Γ is the gamma function. The derivation involves evaluation of the integral for the L^2 -metric. For a proof, see [20]. It should be noted that the metric is symmetric with respect to parameters (p_1, c_1) and (p_2, c_2) even though it does not appear that way.

Equation (6) provides a metric between two Bessel K forms, or between two spectral marginals. 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 $F^{(1)}, \dots, F^{(K)}$, let the parameter values be given by: $(p_1^{(j)}, c_1^{(j)})$ and $(p_2^{(j)}, c_2^{(j)})$, respectively, for $j = 1, 2, \dots, K$. Then, the L^2 -distance, between the spectral representations of the two images, is defined as:

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

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. It has been established in the literature that different spectral components of the same images are often correlated, and therefore, this Euclidean form may not be appropriate. In such cases, another choice such as the *max* of all components may be pursued.

4 Applications of Bessel K Representations

Now we present some examples of applying these Bessel K formulations and the resulting metric to image understanding problems. We have selected examples from: (i) texture modeling and synthesis, (ii) clutter classification, (iii) hypothesis pruning, and (iv) object classification.

4.1 Texture Modeling and Synthesis

It has been shown [25,24] that homogeneous textures can be characterized sufficiently using their spectral responses. Because Bessel K forms provide a low-dimensional and analytical representation for spectral components, we use texture synthesis to further verify the effectiveness of Bessel K representation.

Given an observed image and K filters, we model the given texture by its marginals, represented by the estimated Bessel K densities. Besides using fewer numbers to represent a histogram, the density can be sampled to generate a histogram of any number of bins. An advantage of this model is that it can be

verified systematically through generating images with matched histograms. As in [24], we impose a Gibbs-like distribution on the image space by:

$$P(I) = \frac{1}{Z} \exp\left(-\sum_{j=1}^K \sum_{i=1}^{L^{(j)}} \|H(I^{(j)})(i) - f(z_i; p^{(j)}, c^{(j)})\|^2/T\right),$$

where $H(I^{(j)})$ denotes the histogram of the j -th filter response of I , $L^{(j)}$ is the number of bins, and z_i is the center value of i -th bin of the j -th histogram, T is a parameter corresponding to temperature, and Z is a normalizing constant. Here $p^{(j)}, c^{(j)}$ are estimated using an observed image. The texture synthesis is then to generate typical samples from $P(I)$. Here a Gibbs sampler [24] is used to generate the following examples while other sampling algorithms can also be used. Figure 5 shows three examples, where the histogram of the intensity filter is used directly as it does not satisfy the assumptions of Bessel K forms and for all other filters their estimated Bessel K densities are used. It is evident from these examples that Bessel K forms capture the perceptually important characteristics of textures. The basic elements are synthesized well as well as the global patterns.

4.2 Clutter Classification

An important application of this Bessel K representation is in the classification of clutter for ATR (automated target recognition) scenarios. In particular, given an observed image of a target, imaged in a cluttered environment, one would like to characterize the clutter to the extent that it improves the ATR performance. Some knowledge of clutter type, whether it is grass, buildings, trees, or roads, can help improve the task of target recognition. In this section, we utilize the Bessel K forms to represent the image spectra, 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. Consider the images of natural clutter shown in Fig. 6. For a simple illustration, let the images in the top row be training images that are already classified, and the bottom row be images that are to be classified. Using 27 small-scale Gabor filters ($K = 27$), for nine different orientations at three scales each, we have computed the pairwise distances d_I 's.

Using the nearest neighbor approach, and the metric d_I one can perform clutter classification. To illustrate the classification of clutter types, we have plotted a clustering chart in the left panel of Fig. 7 using the dendrogram function in matlab. This function generates a clustering tree for points in image space when their pairwise distances are given. For comparison we run clustering program using an Euclidean metric on a principal subspace of the image space. We extracted non-overlapping patches of size 20×30 from the original images, performed principal component analysis (PCA) in \mathbb{R}^{600} , and retained only first 40 components. Images are then projected onto this linear subspace to compute coefficients and the resulting pairwise Euclidean distances.

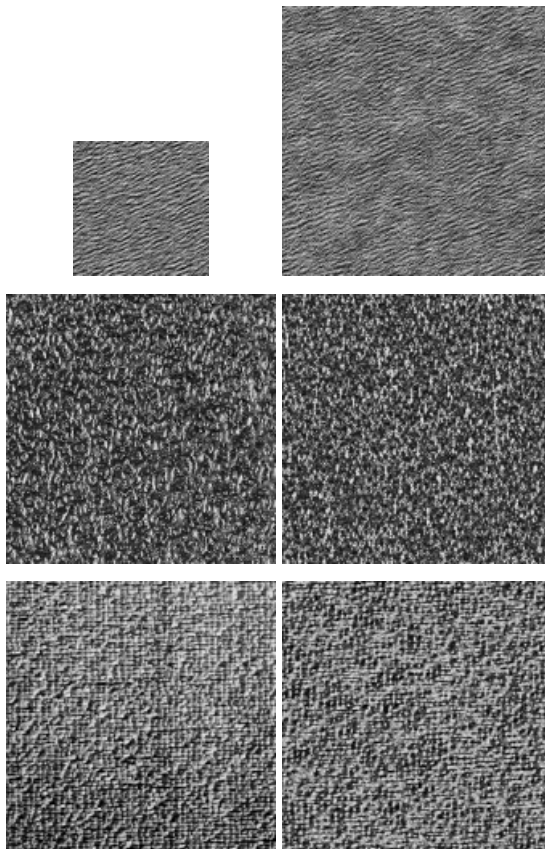


Fig. 5. Texture synthesis examples using Bessel K forms. In each row, the left shows the observed texture and the right a typical synthesized texture.

4.3 Hypothesis Pruning

The Bessel K forms also prove useful in pruning the hypothesis set in target recognition. Recognition of objects from their observed images corresponds to the selection of hypothesis in presence of the nuisance parameters [9]. As stated in Section 1, this hypothesis selection is often performed using detailed models involving physical shapes, texture, pose and motion [21,7,9]. Such methods are based on low- and high-dimensional deformations of targets' templates in order to match their synthesized images with the observed images. The deformations capture the variability in pose, motion, illumination, etc. and form the set of nuisance parameters, call it S , for hypothesis selection; they typically are computationally expensive to implement. Given an image, the task of searching over all possible templates is demanding and can benefit from a pruning that places significant probability only on a small subset.

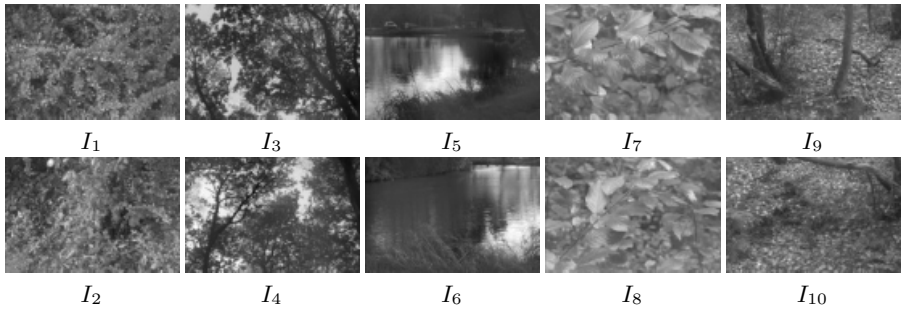


Fig. 6. Ten natural images from the Groningen database: top row are the training images and bottom row are the test images.

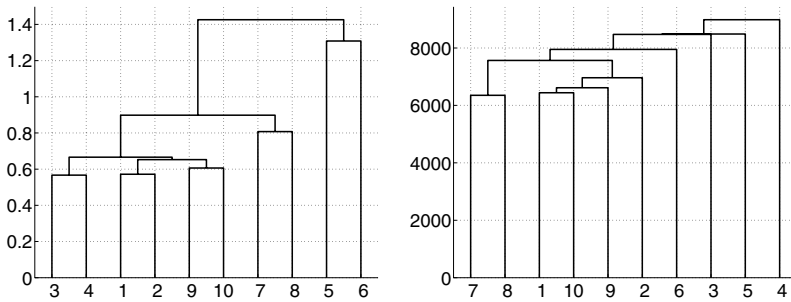


Fig. 7. Dendrogram clustering of images in Fig. 6 using d_I (left panel) and using an Euclidean metric on PCA (right panel). The labels on the horizontal axis correspond to the ones in Fig. 6 and the vertical axis shows the corresponding distance.

Let \mathcal{A} be the set of all possible objects. Define a probability mass function on \mathcal{A} according to:

$$P(\alpha|I) = \frac{\exp\left(-\min_{s \in S} (\sum_{j=1}^K d(p_{obs}^{(j)}, c_{obs}^{(j)}, p_{\alpha,s}^{(j)}, c_{\alpha,s}^{(j)})^2)/T\right)}{\sum_{\alpha'} \exp\left(-\min_{s \in S} (\sum_{j=1}^K d(p_{obs}^{(j)}, c_{obs}^{(j)}, p_{\alpha',s}^{(j)}, c_{\alpha',s}^{(j)})^2)/T\right)}, \quad (8)$$

where T controls our confidence (analogous to the temperature in Gibbs' energies) in this probability. Here $(p_{obs}^{(j)}, c_{obs}^{(j)})$ are the estimated parameters for the image I and filter $F^{(j)}$, and $(p_{\alpha,s}^{(j)}, c_{\alpha,s}^{(j)})$ are the estimated 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 pre-computed offline for all $\alpha \in \mathcal{A}$, $s \in S$, and $j \in \{1, 2, \dots, K\}$.

To illustrate this idea, consider the following experiment. Shown in Fig. 8(a) are some sample images of objects from the Columbia object image library (COIL) [17]. This database consists of 72 images of each 100 objects, taken at five degree separation in azimuth, and has been widely used in testing object recognition algorithms. In this experiment, we have divide 7200 images into non-

overlapping training and test sets. Some of the images are used as training and the remaining for testing, similar to the work presented in [17]. We have used $K = 39$ 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. 9 are the plots of $P(\alpha|I)$ versus α (for $T = 0.5$) for

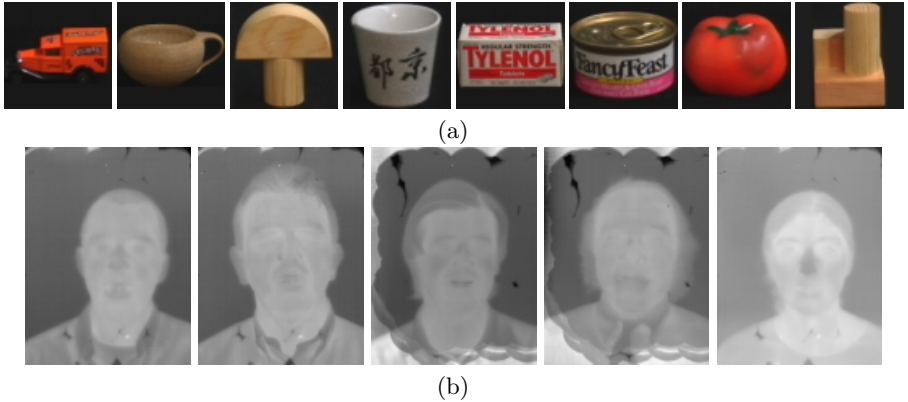


Fig. 8. Sample images from (a) COIL image dataset and (b) Equinox long wave infrared image dataset.

six different images I in the COIL database. All the objects with probabilities larger than some threshold, say 0.01, can be shortlisted for detailed hypothesis testing. As an example, the plot in left corresponds to an image of $\alpha = 1$. In the short-listing by thresholding, we are left with only 14 possible hypothesis, a significant reduction from 100. The middle plot displays the worst case of the whole experiment and still short-lists 35 objects. The right plot displays a best case, where the probability mass function on \mathcal{A} is focused on one object and the hypothesis pruning gives a unique solution.

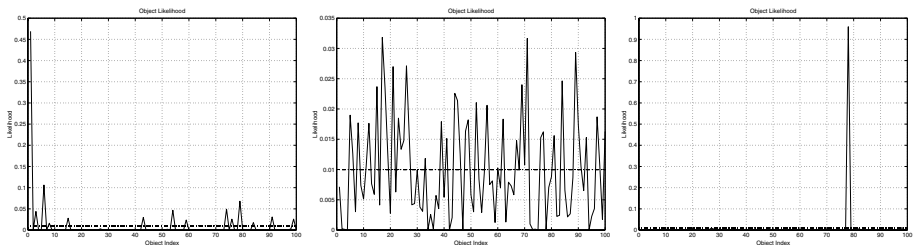


Fig. 9. Plots of $P(\alpha|I)$ versus α for six test images in the COIL database. The test images are of objects α_1 , α_{15} , and α_{78} , respectively, for arbitrary orientations. Dotted lines suggest a threshold level for pruning.

4.4 Object Classification Using Bessel K Representation

Because Bessel K representation defines a distance measure between any images, it can also be used for object classification similar to purely numerical-based representations. One distinctive advantage of Bessel K representation is that it puts probability mass on images that may have large distance in the image-induced Euclidean space. In other words, the probability model on images has multiple modes in the image space, resulting in a more effective representation. Here we have used $P(\alpha|I)$ for object recognition and have compared our results with some other recently proposed procedures: principal component analysis (PCA), independent component analysis (ICA), support vector machines (SVM), and SNoW. Pontil and Verri [17] have applied SVM (Support Vector Machines) method to 3D object recognition and have tested it on a subset of the COIL-100 dataset with half for training and the other half for testing. As pointed out by Yang et al. [23], this dense sampling of training views simplifies the recognition problem. Hence, we have presented recognition results for different training to test ratios in splitting the COIL database. The number of components selected is such that complexity remains similar to that of Bessel representations. As Table 1 summarizes that Bessel representations, in addition to being analytic and parametric, mostly outperform these other methods. The significant performance decrease of the case with 4 training views per object is due to the nearest-neighbor classifier used, which does not generalize well; we expect significant improvement using more sophisticated classifiers, which will be investigated further.

Table 1. Correct recognition rate for the full COIL-100 dataset using PCA, ICA and Bessel forms

Training/test per object	PCA	ICA	SNoW [23]	SVM [23]	Bessel Forms
36 / 36	98.58%	98.47%	95.81%	96.03%	99.89%
18 / 54	96.67%	96.52%	92.31%	91.30%	99.00%
8 / 64	87.23%	87.91%	85.13%	84.80%	92.44%
4 / 68	75.82%	76.03%	81.46%	78.50%	78.65%

The Bessel K form has also been applied to a large infrared face dataset generated by Equinox company². Because the dataset is still under development, here we use a subset consisting of 63 subjects with a total of 3,893 long-wave infrared images of faces. Fig. 8(b) shows some examples in the dataset. Tab. 2 shows the recognition results using the nearest neighbor classifier, which are almost perfect under all the conditions for this large dataset.

5 Conclusion

We have applied Bessel K forms to model the probability densities of the filtered marginals. The estimated parametric forms are shown to match well with the

² Available at <http://www.equinoxsensors.com/products/HID.html>

Table 2. Correct recognition rate with different training images of the Equinox dataset.

Total training/ test images	First correct	First two correct	First three correct
1948 / 1945	99.95%	100.0%	100.0%
993 / 2900	99.97%	100.0%	100.0%
527 / 3366	99.97%	100.0%	100.0%
343 / 3550	99.94%	99.97%	99.97%

observed histograms for a variety of images: video, IR, and range, for gradient, Gabor and Laplacian of Gaussian filters. Given the assumptions behind this construction, we expect this model to perform well in other imaging modalities such as MRI, PET, and radar imaging. We have used L^2 metric on the set of Bessel forms (restricted to $p > 0.25$) to derive a pseudo-metric on the image space. This metric can be used for, among other things, clutter classification and object recognition. Although the performance of Bessel representations in challenging object recognition situations remains to be tested, their ability to prune possible hypotheses, to feed to a more detailed model, seems promising.

Acknowledgments. This research was supported in part by the grants ARO DAAD19-99-1-0267, NMA 201-01-2010, and NSF DMS-0101429. The images used in the experiments are taken from the Groningen image database, the COIL database, Brown range database, Equinox IR face dataset, and FSU IR face database. We are grateful to the producers of these databases for making them public. We also thank Prof. J. Sethuraman for some useful discussions on this paper.

References

1. O. Barndorff-Nielsen, J. Kent, and M. Sorensen, "Normal variance-mean mixtures and z distributions," *International Statistical Review*, vol. 50, pp. 145–159, 1982.
2. P. N. Belhumeur, J. P. Hepanha, and D. J. Kriegman, "Eigenfaces vs. fisherfaces: Recognition using class specific linear projection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19(7), pp. 711–720, 1997.
3. A. J. Bell and T. J. Sejnowski, "The "independent components" of natural scenes are edge filters," *Vision Research*, vol. 37(23), pp. 3327–3338, 1997.
4. P. Comon, "Independent component analysis, a new concept?" *Signal Processing, Special issue on higher-order statistics*, vo. 36(4), pp. 287-314, 1994.
5. I. S. Gradshteyn and I. M. Ryzhik, *Table of Integral Series and Products*, Academic Press, 2000.
6. U. Grenander, *General Pattern Theory*, Oxford University Press, 1993.
7. U. Grenander, M. I. Miller, and A. Srivastava, "Hilbert-schmidt lower bounds for estimators on matrix lie groups for atr," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20(8), pp. 790–802, 1998.
8. U. Grenander and A. Srivastava, "Probability models for clutter in natural images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23(4), pp. 424-429, 2001.

9. U. Grenander, A. Srivastava, and M. I. Miller, "Asymptotic performance analysis of bayesian object recognition," *IEEE Transactions of Information Theory*, vol. 46(4), pp. 1658–1666, 2000.
10. D. J. Heeger and J. R. Bergen, "Pyramid-based texture analysis/synthesis," In *Proceedings of SIGGRAPH*, pp. 229–238, 1995.
11. B. Julesz, "A theory of preattentive texture discrimination based on first-order statistics of textons," *Biological Cybernetics*, vol. 41, pp. 131–138, 1962.
12. M. Kirby and L. Sirovich, "Application of the karhunen-loeve procedure for the characterization of human faces," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12(1), pp. 103–108, 1990.
13. Ann B. Lee and David Mumford, "Occlusion models for natural images: A statistical study of scale-invariant dead leaves model," *International Journal of Computer Vision*, vol. 41, pp. 35–59, 2001.
14. M. I. Miller, A. Srivastava, and U. Grenander, "Conditional-expectation estimation via jump-diffusion processes in multiple target tracking/recognition," *IEEE Transactions on Signal Processing*, vol. 43(11), pp. 2678–2690, 1995.
15. D. Mumford, "Empirical investigations into the statistics of clutter and the mathematical models it leads to," *A lecture for the review of ARO Metric Pattern Theory Collaborative*, 2000.
16. B. A. Olshausen and D. J. Field, "Sparse coding with an overcomplete basis set: A strategy employed by V1?" *Vision Research*, vol. 37(23), pp. 3311–3325, 1997.
17. M. Pontil and A. Verri, "Support vector machines for 3d object recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20(6), pp. 637–646, 1998.
18. J. Portilla and E. P. Simoncelli, "A parametric texture model based on joint statistics of complex wavelets," *International Journal of Computer Vision*, vol. 40(1), pp. 49–71, 2000.
19. S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, pp. 2323–2326, 2000.
20. A. Srivastava, X. Liu, and U. Grenander, "Universal analytical forms for modeling image probabilities," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, in press, 2002.
21. A. Srivastava, M. I. Miller, and U. Grenander, "Bayesian automated target recognition," *Handbook of Image and Video Processing, Academic Press*, pp. 869–881, 2000.
22. M. J. Wainwright, E. P. Simoncelli, and A. S. Willsky, "Random cascades on wavelet trees and their use in analyzing and modeling natural images," *Applied and Computational Harmonic Analysis*, vol. 11, pp. 89–123, 2001.
23. M. H. Yang, D. Roth, and N. Ahuja, "Learning to recognize 3d objects with SNoW," In *Proceedings of the Sixth European Conference on Computer Vision*, vol. 1, pp. 439–454, 2000.
24. S. C. Zhu, X. Liu, and Y. N. Wu, "Statistics matching and model pursuit by efficient MCMC," *IEEE Transactions on Pattern Recognition and Machine Intelligence*, vol. 22, pp. 554–569, 2000.
25. S. C. Zhu, Y. N. Wu, and D. Mumford, "Minimax entropy principles and its application to texture modeling," *Neural Computation*, vol. 9(8), pp. 1627–1660, 1997.