Universal Analytical Forms for Modeling Image Probabilities

Anuj Srivastava * Xiuwen Liu † Ulf Grenander ‡

Abstract

Seeking universal probability models for image representations, we employ a spectral approach where the images are decomposed using several bandpass filters, and probability models on the filter outputs (or spectral components) are imposed. We apply a (two-parameter) analytical form, introduced in [9] and called a **Bessel K form**, for modeling the marginal probabilities of these spectral components, and demonstrate their fit to the observed histograms for video, IR, and range images. A relationship between the Bessel parameters and certain characteristics of the imaged objects is established. Using an expression for L^2-metric on the set of Bessel K forms, we suggest a pseudo-metric on the image space for quantifying image similarities/differences. Some applications, including clutter classification and pruning of hypotheses for target recognition, are presented.

Index terms: image statistics, spectral analysis, Bessel K forms, clutter classification, target recognition.

1 Introduction

Statistical techniques for image analysis and understanding require efficient probability models for the observed images. Given the tremendous variability associated with the imaged objects, detailed (e.g. 3D deformable templates) models may not be feasible in modeling "all possible targets and clutter objects". Therefore, one seeks a balance by designing low-level, coarse representations for modeling the pixel patterns, representations that are tractable and yet capture significant image variation. Here we study tractable, coarse probability models that can form building blocks for a larger image understanding system. Since the image space is very high-dimensional, a direct modeling of the joint probabilities is not possible, even if a large number of observations are provided, and some method for reducing dimensions is required. There are two general reductionist approaches presented in the literature: (i) parameterize the probability densities using certain (low-dimensional) physical parameters (relating to the imaged objects), or (ii) perform dimension reduction via purely numerical, non-physical approaches.

In the first approach, relating to high-level vision, images are characterized by the physical characteristics of the objects (such as shapes, textures, reflectance, illumination, and motion). These quantities are modeled mathematically, within some acceptable approximation, and the resulting physical variables are used to analyze images. Probability models on images now consist of: (i) probability models on these physical variables and (ii) the sensing models. An example of this idea is the *deformable template* theory ([7]) where images are studied through the transformations

*Department of Statistics, Florida State University, Tallahassee, 32306
†Department of Computer Science, Florida State University, Tallahassee, FL 32306
‡Division of Applied Mathematics, Brown University, Providence, RI 021912

that match the templates to the observations. These models are detailed and they capture sufficient variability to discriminate well even in the challenging situations (cluttered scenes, low SNR, distant images, etc.). One drawback is that they are computationally expensive to implement, since they require synthesis of hypothesized images for image analysis. Also, there is a possibility of finding new objects in the image whose physical descriptors may not be known beforehand. The second idea, relating to low-level vision, involves one of many techniques that reduce the dimensions using purely numerical considerations. That is, 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 subspace that best represents those numbers (under some chosen criterion). Principal components [14], independent components [5, 3], sparse coding [19], Fisher's discriminant [2], local linear embedding [22], 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.

1.1 Models for Image Analysis

An important idea is to develop an adaptive strategy that balances these two levels of inferences. In this paper, we study a framework that provides some interaction between the pixel-based and the template-based inferences, and is capable of shifting between the two depending on the available resources, both computational and informational. Consider a deformable template representation of the imaged objects, as laid out in the papers [10, 23]. The basic idea is that images are made up of (images of) objects, and their variability can be represented by physical variables. Using 3D models of objects (including polygonated surfaces, textures, and reflectance functions), 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. What probability models on I (an image) can result from this model? As stated earlier, there are two possibilities:

1. First, the template approach where the set of possible objects is small, and the 3D models are available for all objects. Also, the tools for synthesizing images of these objects under all transformations are assumed available. Under these assumptions, a probability distribution on I can be written explicitly in terms of a probability on the transformation space and the set of objects. Object recognition is now solved on their product space [23, 17]. Inference procedures are laid out in the papers [8, 10, 17].

2. Second, for general problems in image understanding, 3D models can not be pre-stored for all objects, and furthermore, the transformations may not be identifiable in all conditions. Therefore, the representations tend to be less explicit and the probability of I is motivated through empirical studies [18], and not via the physical parameters.

We seek general models that retain some physical considerations, although not as explicitly as the template approach. Pursuing the second case, we replace 3D templates by their 2D profiles (call them **generators**) and denote them as g's. g's are the views (signatures, profiles) of randomly chosen objects, taken from random poses. 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 will be 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, $g_i : W \mapsto I\!R_+$ is a generator of a randomly chosen

object, $\rho_i \in [0, L]$ is a random scale, and $a_i \in I\!R$ is a random weight associated with g_i. g_i's are assumed to be drawn from the generator set $\mathcal{G}$ according to some measure dG. The image formation is now modeled by the equation:

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

Assume: (i) the weights a_i's are *i.i.d.* standard normal, (ii) the scales ρ_i's are i.i.d. uniform on the interval $[0, L]$, (iii) the locations z_i's as samples from a 2D, homogeneous Poisson process, with intensity λ on a compact set $W \subset I\!R^2$, and (iv) a_i's, g_i's, z_i's, and ρ_i's are all assumed independent of each other. Since g_i's are assumed unknown, the related variables n, ρ_i's and z_i's are also indeterminable. However, 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. If certain low-dimensional statistics of these filtered components are found sufficient, then a significant reduction is achieved. Zhu et al. [26] have shown that the marginal distributions of spectral components, obtained using a collection of filters, completely characterize homogeneous textures. The choice of histograms as sufficient statistics implies that only the frequencies of occurrences of (pixel) values, in the filtered images, are relevant and the location information is discarded [11, 27, 13]. Chubb et al. [4] also advocate the use of histogram in texture representation. The focus in these papers has been to model homogeneous textures but we apply spectral analysis to a general setting of image understanding. Simoncelli et al. [21] 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. [24] have studied a family of Gaussian mixtures, for different mixing densities, for modeling the observed histograms. Lee et al. [15]. have presented a model for capturing the statistics in the images of leaves.

Using a physical model for image formation, we have proposed a family of two-parameter probability densities [9], called Bessel K forms, to model the horizontal and the vertical derivatives of an image. In this paper, we now apply an extended version of this model to a full spectrum of bandpass filters and arbitrary images. The two parameters, associated with this family, will form a sufficient statistic for a spectral component, denoting a significant reduction in the representation. The parameters depend only on the variance and the kurtosis of the filtered image pixels, thereby implying a simple estimation procedure. The main results presented here are: (i) we demonstrate the success of Bessel K forms in modeling the spectral components for video, IR, and range images of natural and artificial scenes, (ii) we derive an analytical expression for computing the L^2-metric, on the Bessel family, that leads to a pseudo-metric on image space, and (iii) use the Bessel K forms (and the pseudo-metric) as a tool for clutter classification and for pruning possible hypothesis set for recognition of objects from their images. The last claim is based on: (i) motivating the model in Eqn. 1 by relating it the 3D deformable template representation, (ii) relating Bessel parameters to certain physical characteristics of the imaged objects, and (iii) an example of pruning pruning hypothesis using the COIL database. In addition, we will present an asymptotic approximation of the Bessel K form that can potentially simplify the inference procedures.

This paper is laid out 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 clutter classification and target recognition.

2 Probability Models for Image Spectra

We start with some notation. Given an image I and a bank of filters $\{F^{(j)}, j = 1, 2, \ldots, K\}$, we compute, for each filter $F^{(j)}$, a filtered image $I^{(j)} = I * F^{(j)}$, where $*$ denotes the 2D convolution operation. As an example, a Gabor filter is a bandpass filter with a Gaussian kernel centered around a specific wavenumber (see [12] for details). For a rotation $\theta \in S^1$ on the unit circle, a Gabor filter is given by: $R_\theta(z) \circ \left(\exp(-\frac{1}{2\sigma^2}(z(1)^2 + z(2)^2)) \exp(-j\frac{2\pi z(1)}{\sigma})\right)$, where σ denotes the resolution associated with the filter and R_θ is the 2×2 rotation matrix. Another filter suggested by Marr [16] to model human vision is the Laplacian Gaussian filter whose operation on I is given by $(G * \Delta)I$ where G is a Gaussian kernel and Δ is the Laplacian operator. In addition to these filters, 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. In this paper, we do not address the issue of selecting filters to best accomplish a specific task. Instead, we will assume an arbitrary choice of filters as long as the resulting spectral components have marginals that are: (i) unimodal with the mode at zero, (ii) symmetric around zero, and (iii) are leptokurtic, i.e. their kurtosis are more than that of a Gaussian random variable with the same variance. Additionally, we want the filters such that the resulting representation is computationally efficient. This is true for filters with smaller bandwidths.

2.1 Analytical Models

Applying 2D convolution to both sides of Eqn. 1, we obtain a spectral component

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

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 $u \equiv \sum_i (g_i^{(j)}(\frac{1}{\rho_i}(z - z_i))^2$. One departure here, from the model used in [9], is that the generators are now randomly selected and are included at random scales ρ_i's in Eqn. 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 [9]: for $p > 0$, $c > 0$,

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

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 the **Bessel K forms** and the parameters (p, c) as the **Bessel parameters**. The elements of $\mathcal{D}$ have the following properties:

1. They 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.

2. The kurtosis of a Bessel K form relates to the term $3 + \frac{\text{var}(u)}{(E[u])^2}$. Hence, the Bessel K forms are leptokurtic (the tails are heavier as compared to a normal curve with the same variance).

3. A Bessel K form is a specific kind of normal variance-mean mixture where the mixing variable is scaled Gamma with parameters p and c. It becomes a special case of a larger family of self-reciprocal normal variance mixtures as described by Barndorff-Nielson et al. [1]. This connection opens the possibility of a larger family, namely the *generalized hyperbolic distributions*, to be used in modeling image spectra, if needed. In addition to the shape and the scale, this family allows for a location parameter, skewness, and different rates (of exponential decay) on the two tails. The significance here for image analysis is that more filters, beyond the ones that lead to symmetric and zero-mean histograms, can also be included in analysis.

4. The family of Bessel K forms is infinitely divisible, i.e. any random variable in this family can be written as a sum of two independent random variables from this family. 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 $a_1 I_1 + a_2 I_2$ $(a_1, a_2 \in I\!R)$ may not be a Bessel K form but in certain conditions can be approximated by $f(x; p, c)$ where

$$p = \frac{(a_1^2 p_1 c_1 + a_2^2 p_2 c_2)^2}{a_1^4 p_1 c_1^2 + a_2^4 p_2 c_2^2}, \quad \text{and} \quad c = \frac{a_1^4 p_1 c_1^2 + a_2^4 p_2 c_2^2}{a_1^2 p_1 c_1 + a_2^2 p_2 c_2} \ .$$

5. A Bessel K forms is square-integrable only for $p > 0.25$. This property is due to the choice of Gamma density for u and it limits our later derivation of L^2-metric to the Bessel K forms with p-values larger than 0.25.

How to estimate the Bessel K parameters for a given filtered image? As described in [9], p and c can be estimated using the equations

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

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 moments of the filtered image. We illustrate some estimation results for a variety of images.

- Shown in the top panels of Figure 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 broken lines and the estimated Bessel K forms ($f(x; \hat{p}, \hat{c})$) are plotted in solid lines.

- For the image shown in the top panel of Figure 2, we have estimated Bessel K forms for different Gabor filters. The middle panels plot the marginals for different filter orientations ($\theta = 30, 60, 90$, and 150 degrees) while keeping the scale fixed at $\sigma = 4.0$, and the bottom panels are for different filter scales ($\sigma = 1, 2, 3$, and 5) keeping the orientation fixed at $\theta = 150$ degrees.

- Figure 3 shows four examples of estimation for filtering by Laplacian-Gaussian filters. The top panels show the natural images from Groningen database and the bottom panels show the corresponding estimated Bessel K forms.

- Figure 4 shows estimation results for three infrared(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.

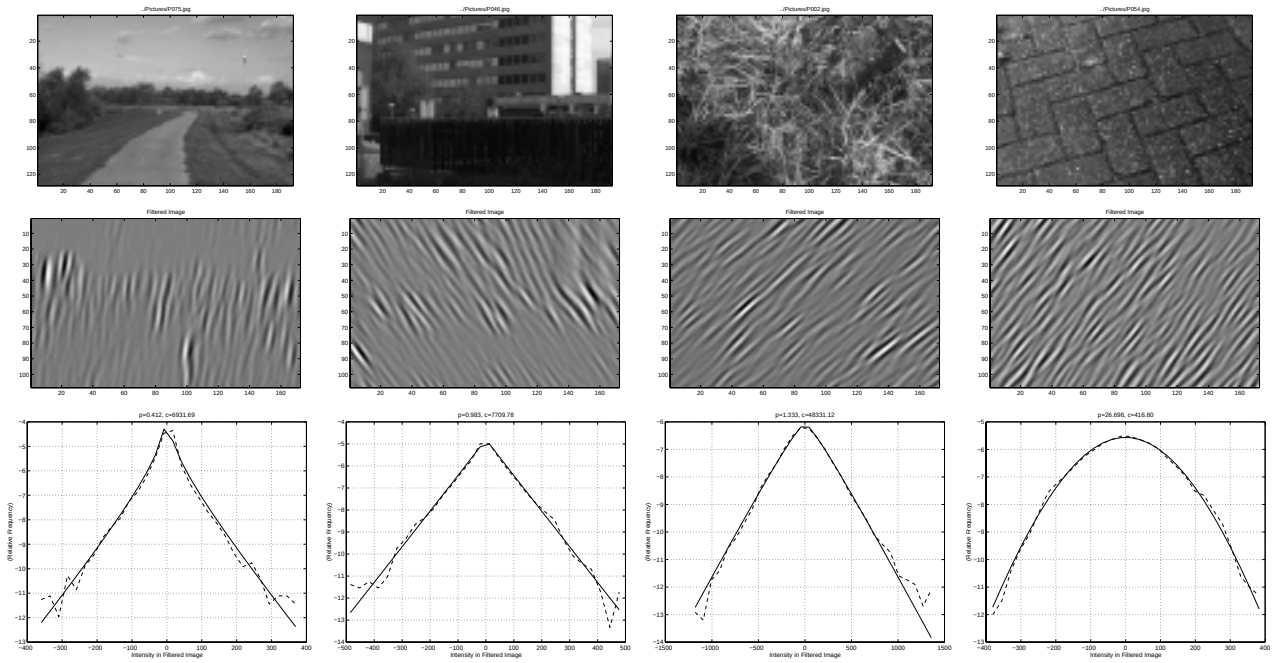

Figure 1: Images (top panels), their Gabor components (middle panels), and the marginal densities (bottom panels). The observed densities are drawn in broken lines and the estimated Bessel K forms are drawn in solid lines.

- Shown in Figure 5 are some examples of estimating marginal densities for the case of range images taken from the Brown range database. The three images shown in top panels are filtered using Gabor filters and the resulting densities are plotted in the bottom panels.

Since the estimation of parameters is based on moments, it may be preferable to use robust estimation techniques to account for the outliers. A simple idea is to consider a fraction (say one percent) of the tail as outlier and discard it in parameter estimation. Using this idea we have found an improvement in estimation performance for the cases where p is quite small ($p < 0.1$). 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.

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-Leibler (KL) divergence. For any two density functions f_1 and f_2, the divergence is defined as the quantity:

$$D(f_1, f_2) = \int_{I\!R} \log(\frac{f_1(x)}{f_2(x)} f_1(x) dx \ .$$

We have computed it by discretizing at the center points of the histogram bins. To illustrate KL divergence values for our application, we start with some examples. Shown in Figure 6 are six plots, each containing a pair of densities: observed and estimated, and the KL divergence between them. These six cases show a decreasing match, in going from top left to bottom right. The

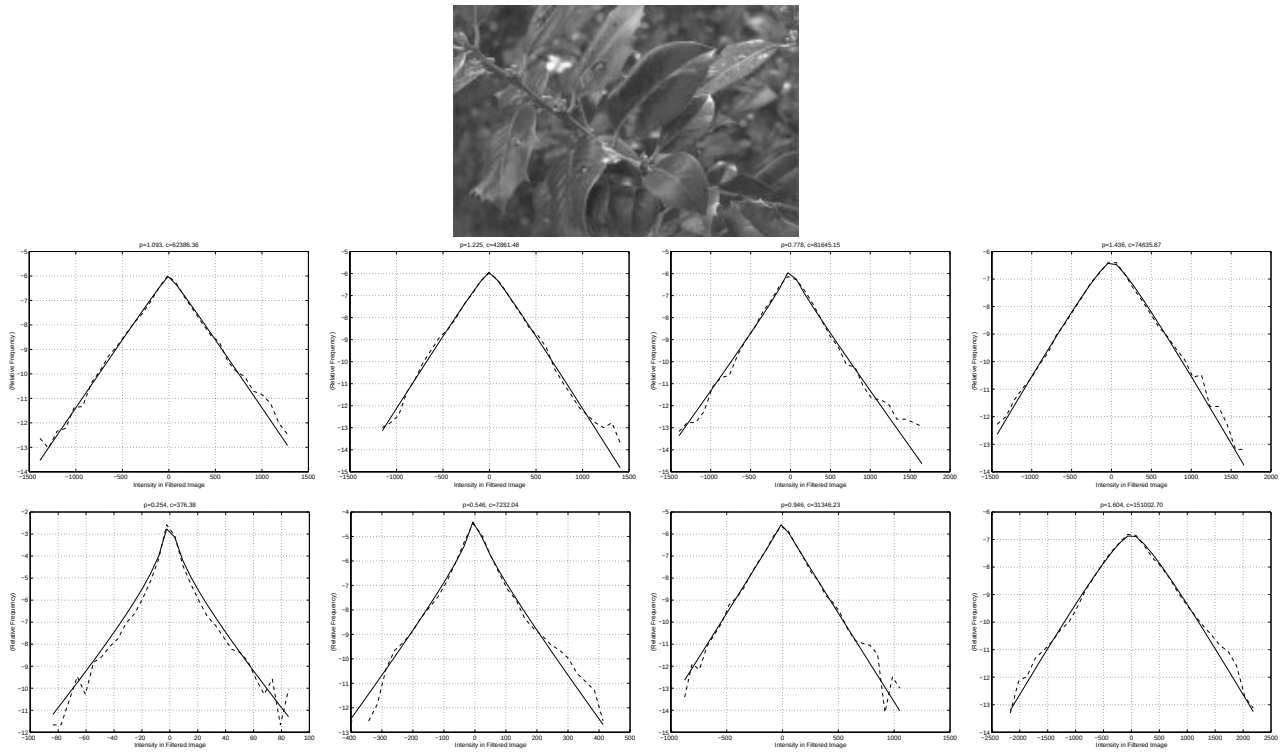

Figure 2: Plots of observed and estimated marginals (on a log scale) of the spectral components of a given image (top panel). Middle panels depict the marginals for different filter orientations: 30, 60, 90, and 150, while the bottom panels are for different filter scales: 1, 2, 3, and 5.

KL divergence, in that order, is found to be 0.0005, 0.0086, 0.0786, 0.1459, 0.2260, and 0.3625, respectively.

Using KL divergence for evaluating match 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 Groningen natural image database, and the second database is made up of 220 IR face pictures. Shown in Figure 7 are the convergence plots of the average KL divergence, plotted against the sample size. The left plot is for the natural video images with a limiting value of 0.0719 while the right plot is for the infrared images with a limiting value of 0.0479. A comparison of these values with the examples in Figure 6 underscores the degree of match between the observed histograms and the estimated Bessel K forms.

2.3 Relating the Shape Parameter to the Imaged Objects

Before we present some applications of these Bessel K forms, we deal with an interesting question. How do the Bessel K parameters p and c, estimated for a spectral component of an image, relate to the objects that are present in that image. The physical characteristics of the imaged objects, and the filter used in generating a spectral component, should dictate the resulting Bessel K form. Since c is essentially a scale parameter relating to the range of pixels values in I, its role is not as

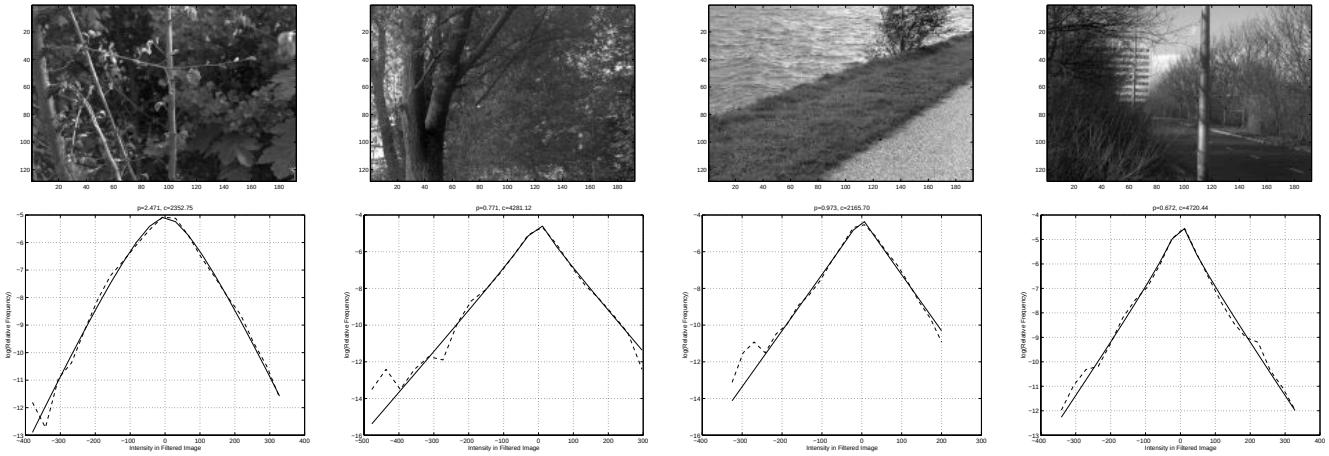

Figure 3: Estimated Bessel K forms for natural images (top panels) when filtered by Laplacian Gaussian filters.

important as p. In the following analysis, we outline a relationship between the imaged generators and the estimated shape parameter. We will perform the analysis for the pixel $I(z)$ as composed of the generators $\{g_i(\frac{1}{\rho_i}(z - z_i))\}$ (Eqn. 1), although the analysis remains same for any spectral component $I^{(j)}(z)$ composed of the generators $\{g_i^{(j)}(\frac{1}{\rho_i}(z - z_i))\}$ (Eqn. 2). Let the characteristic function of the random variable a_i' be given by $\phi(\omega)$. (Later we assume a specific ϕ by choosing a_i's to be standard normal.) The conditional characteristic function of $I^{(j)}(z)$, given the Poisson points $\{z_i\}$'s, the scales $\{\rho_i\}$'s, and the generators $\{g_i\}$'s, is:

$$\Psi_I(\omega|\{z_i\}, \{\rho_i\}, \{g_i\}) = \prod_{i=1}^{n} \phi(\omega g_i(\frac{1}{\rho_i}(z - z_i))) \ ,$$

using the i.i.d. nature of a_i's. Integrating out the uniform, independent placement of z_i's, for a given n, we obtain the conditional characteristic function:

$$\Psi_I(\omega|n, \{g_i\}, \{\rho_i\}) \propto \prod_{i=1}^{n} \left(\int_W \phi(\omega g_i(\frac{1}{\rho_i}(z - z_i)))dz_i \right) \ .$$

Similarly, integrating out the scales we get:

$$\Psi_I(\omega|n, \{g_i\}) \propto \prod_{i=1}^{n} \left(\int_W \int_0^L \phi(\omega g_i(\frac{1}{\rho_i}(z - z_i)))d\rho_i dz_i \right) \ .$$

Now integrate out the random selection of the generators g_i's. As stated earlier, each g_i is drawn independently from some generator space $\mathcal{G}$ according to some measure dG. This gives,

$$\Psi_I(\omega|n) \propto \left(\int_{\mathcal{G}} \int_W \int_0^L \phi(\omega g_1(\frac{1}{\rho_1}(z - z_1)))d\rho_1 dz_1 dG_1 \right)^n \ .$$

The last step is to integrate with respect to the Poisson random variable n:

$$\Psi_I(\omega) \quad \propto \quad \sum_{n=0}^{\infty} \left(\frac{\exp(-\lambda)\lambda^n}{\text{fact}(n)} \left(\int_{\mathcal{G}} \int_W \int_0^L \phi(\omega g_1(\frac{1}{\rho_1}(z - z_1)))d\rho_1 dz_1 dG_1 \right)^n \right)$$

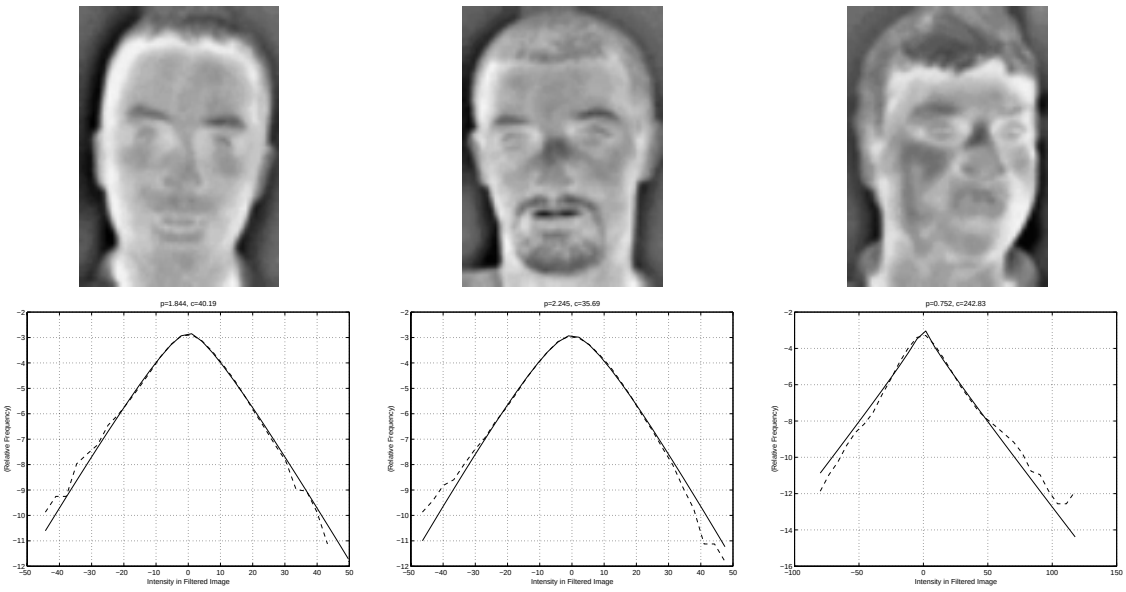

Figure 4: Observed and estimated marginal densities (bottom panels) for the IR face images (top panels) and arbitrary Gabor filters.

$$= \quad \exp\left(\lambda(\int_{\mathcal{G}}\int_{W}\int_{0}^{L}\phi(\omega g_1(\frac{1}{\rho_1}(z-z_1)))d\rho_1 dz_1 dG_1 - 1)\right) \ .$$

WLOG, we can substitute $z = 0$. Furthermore, assume that the generators $g \in \mathcal{G}$ are all even functions. Then,

$$\Psi_I(\omega) \propto \exp\left(\lambda(\int_{\mathcal{G}}\int_{W}\int_{0}^{L}\phi(\omega g_1(\frac{z_1}{\rho_1}))d\rho_1 dz_1 dG_1 - 1)\right) \ . \tag{5}$$

To find the cumulants of I, we use the relation: $\mu_k = \frac{d^k \log(\Psi_I(\omega))}{d\omega^k}|_{\omega=0}$, The two cumulants that we need are:

$$\mu_2 = \lambda\phi''(0)\left(\int_{\mathcal{G}}\int_{W}\int_{0}^{L} g_1(\frac{z_1}{\rho_1})^2 d\rho_1 dz_1 dG_1\right), \quad \mu_4 = \lambda\phi^{(iv)}(0)\left(\int_{\mathcal{G}}\int_{W}\int_{0}^{L} g_1(\frac{z_1}{\rho_1})^4 d\rho_1 dz_1 dG_1\right) \ .$$

The kurtosis of $I(z)$, according to this model, is given by:

$$\text{kurtosis}(I) = \frac{\mu_4}{\mu_2^2} = \frac{\phi^{(iv)}(0)\left(\int_{\mathcal{G}}\int_{W}\int_{0}^{L} g_1(\frac{z_1}{\rho_1})^4 d\rho_1 dz_1 dG_1\right)}{\lambda\left(\phi''(0)\left(\int_{\mathcal{G}}\int_{W}\int_{0}^{L} g_1(\frac{z_1}{\rho_1})^2 d\rho_1 dz_1 dG_1\right)\right)^2} \ .$$

Comparing this with an earlier result that $p = \frac{3}{\text{kurtosis}(I)-3}$, derived in [9], we get

$$p = \frac{1}{\frac{\kappa}{3\lambda} - 1}, \quad \text{where} \ \ \kappa = \frac{\phi^{(iv)}(0)\left(\int_{\mathcal{G}}\int_{W}\int_{0}^{L} g_1(\frac{z_1}{\rho_1})^4 dz_1 dG_1\right),}{(\phi''(0))^2\left(\int_{\mathcal{G}}\int_{W}\int_{0}^{L} g_1(\frac{z_1}{\rho_1})^2 d\rho_1 dz_1 dG_1\right)^2} \ . \tag{6}$$

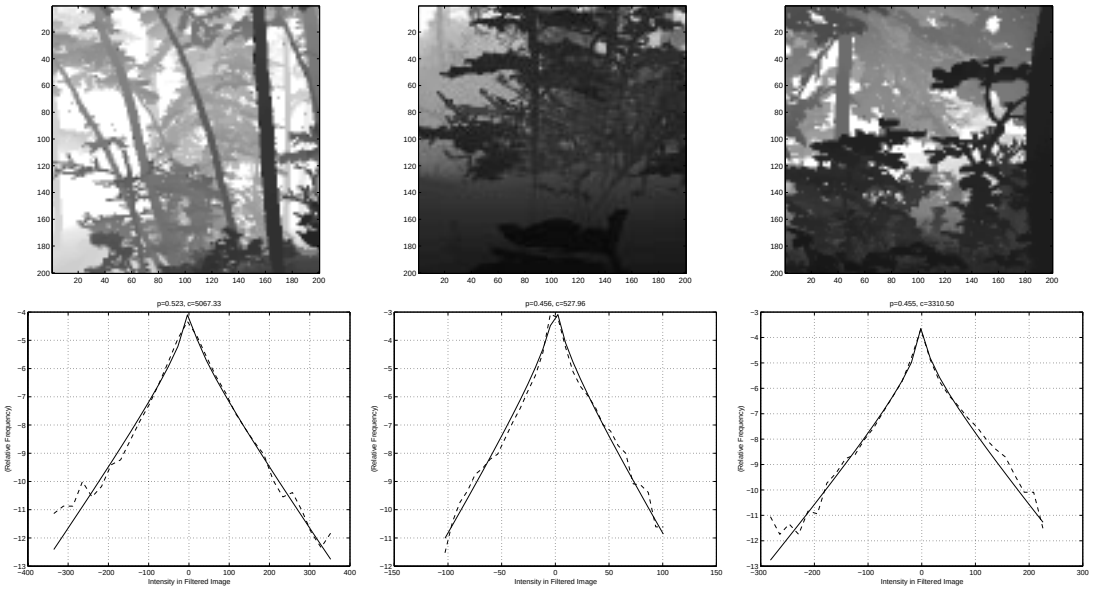

Figure 5: Top panels: range images of a forest. Bottom panels: corresponding observed and estimated marginal densities.

Remark 1: One can replace g_1 by $g_1^{(j)}$ to obtain this relation for any spectral component of I. 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 κ simplifies to $\kappa = \frac{\left(\int_W g(z_1)^4 dz_1\right)}{\left(\int_W g(z_1)^2 dz_1\right)^2}$. This equation provides an important relationship between a generator g and the parameter p. According to Eqn. 6, $p < 1$ 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 Figure 8 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 (0.26, 0.73, 1.21, and 3.74). Summarizing the relation between p and κ, we have:

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

2.4 Asymptotic Approximation of Bessel K Forms

Although the proposed Bessel K forms model the observed histograms very well, their functional form is not easy to work with. Following the discussion in [1], it is possible to approximate the tails of Bessel K forms using a gamma density as follows. For large values of x, the modified Bessel function $K_\rho(x)$ can be approximated by the function $\sqrt{\frac{\pi}{2x}}\exp(-x)$ uniformly over compact sets of ρ. Since small values of ρ imply a large tail, this approximation holds well for a large part of the domain in cases where p is small. The approximation of the Bessel K form can be stated as

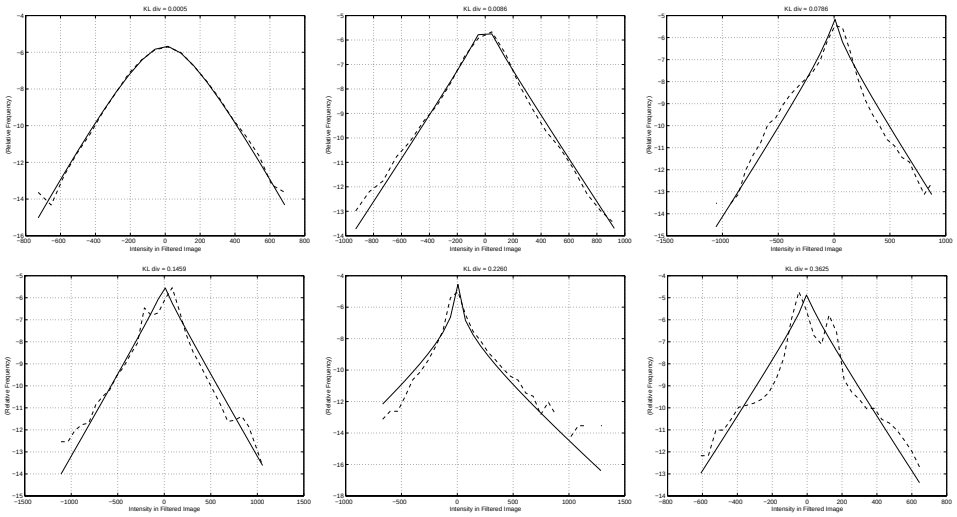

Figure 6: Examples of Kullback-Leibler divergence: the divergence values for the six plots are 0.0005, 0.0086, 0.0786, 0.1459, 0.2260, and 0.3625, respectively.

follows:. The tails of a Bessel K form with the parameters p and c can be approximated well by the function:

$$\tilde{f}(x;p,c) = \frac{1}{\Gamma(p)}(\frac{2}{c})^{p/2}|x|^{p-1}\exp(-\sqrt{\frac{2}{c}}|x|) \ . \tag{7}$$

This result is useful as it provides a simpler expression although only in an asymptotic setting. For $p > 1$, the maxima of $\tilde{f}(x;p,c)$ is attained at the points $\pm(p-1)\sqrt{\frac{2}{c}}$, and therefore this approximation holds only for $p << 1.0$. To illustrate this approximation on filtered marginals, shown in Figure 9 are three examples. The top panels show original images and the bottom panels show the log densities for arbitrary Gabor filters. The observed histograms are plotted in broken lines (- - -), the Bessel K forms $f(x;,p,c)$ are plotted in dotted lines (-.-) , and the asymptotic approximations $\tilde{f}(x;p,c)$ are plotted in solid lines.

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. For instance, we would like to be able to compare images by directly comparing their respective Bessel parameters. An analytical form is very useful in the sense that we do not need to compute the full densities.

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 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.

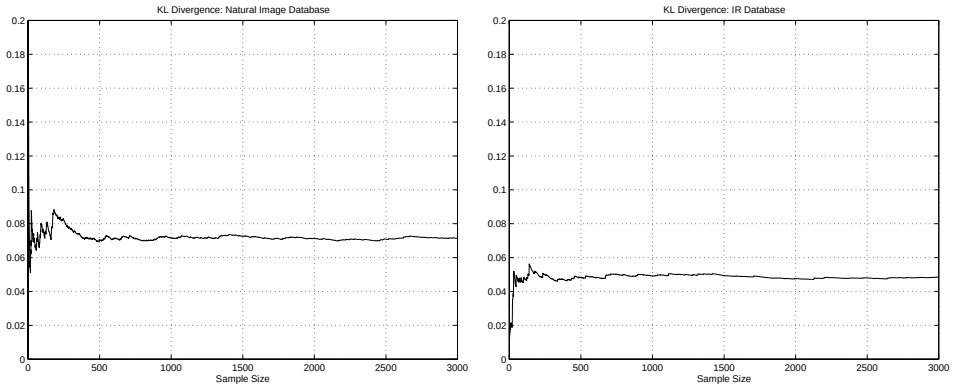

Figure 7: Convergence of average KL-divergence between the observed and the estimated densities as the sample size increases. The left plot is for the Groningen database of natural images and the right plot is for the FSU IR face database.

Remark 2: 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 $d(p_1, c_1, p_2, c_2) = \sqrt{\int_x (f(x; p_1, c_1) - f(x; p_2, c_2))^2 dx}$. This metric can be computed in a closed form, under certain restrictive conditions, as follows.

Theorem 1 *The L^2-distance between the two Bessel K densities, parameterized by (p_1, c_1) and (p_2, c_2), respectively, is given by: for $p_1, p_2 > 0.25$, $c_1, c_2 > 0$,*

$$d(p_1, c_1, p_2, c_2) = \left(\frac{1}{2\sqrt{2\pi}} \Gamma(0.5) \left(\frac{\mathcal{G}(2p_1)}{\sqrt{c_1}} + \frac{\mathcal{G}(2p_2)}{\sqrt{c_2}} - \frac{2\mathcal{G}(p_1 + p_2)}{\sqrt{c_1}} (\frac{c_1}{c_2})^{p_2} \mathcal{F} \right) \right)^{\frac{1}{2}} , \tag{8}$$

where $\mathcal{G}(p) = \frac{\Gamma(p-0.5)}{\Gamma(p)}$ and $\mathcal{F} = F((p_1 + p_2 - 0.5), p_2; p_1 + p_2; 1 - \frac{c_1}{c_2})$ (F is the hypergeometric function). If $c_1 = c_2 = c$, then the metric becomes

$$(\frac{1}{2\sqrt{2\pi}} \frac{\Gamma(0.5)}{\sqrt{c}} (\mathcal{G}(2p_1) + \mathcal{G}(2p_2) - 2\mathcal{G}(p_1 + p_2)) .$$

Proof: Please refer to the appendix. Shown in Figure 10 are some examples of this metric. Each plot shows two Bessel K forms with the parameters (p_1, c_1) and (p_2, c_2). In these examples $p_1 = 0.5$ and $c_1 = 10.0$ are kept fixed, while p_2, c_2 are varied to demonstrate different cases. The distances increase in the plots going from left to right.

Theorem 1 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 $F^{(1)}, \ldots, F^{(K)}$, let the parameter values be given by: $(p_1^{(j)}, c_1^{(j)})$ and $(p_2^{(j)}, c_2^{(j)})$, respectively, for $j = 1, 2, \ldots, K$. Then, the L^2-distance, between the spectral representations of the two images, is defined as:

$$d_I(I_1, I_2) = \sqrt{\left(\sum_{j=1}^{K} d(p_1^{(j)}, c_1^{(j)}, p_2^{(j)}, c_2^{(j)})^2 \right)} . \tag{9}$$

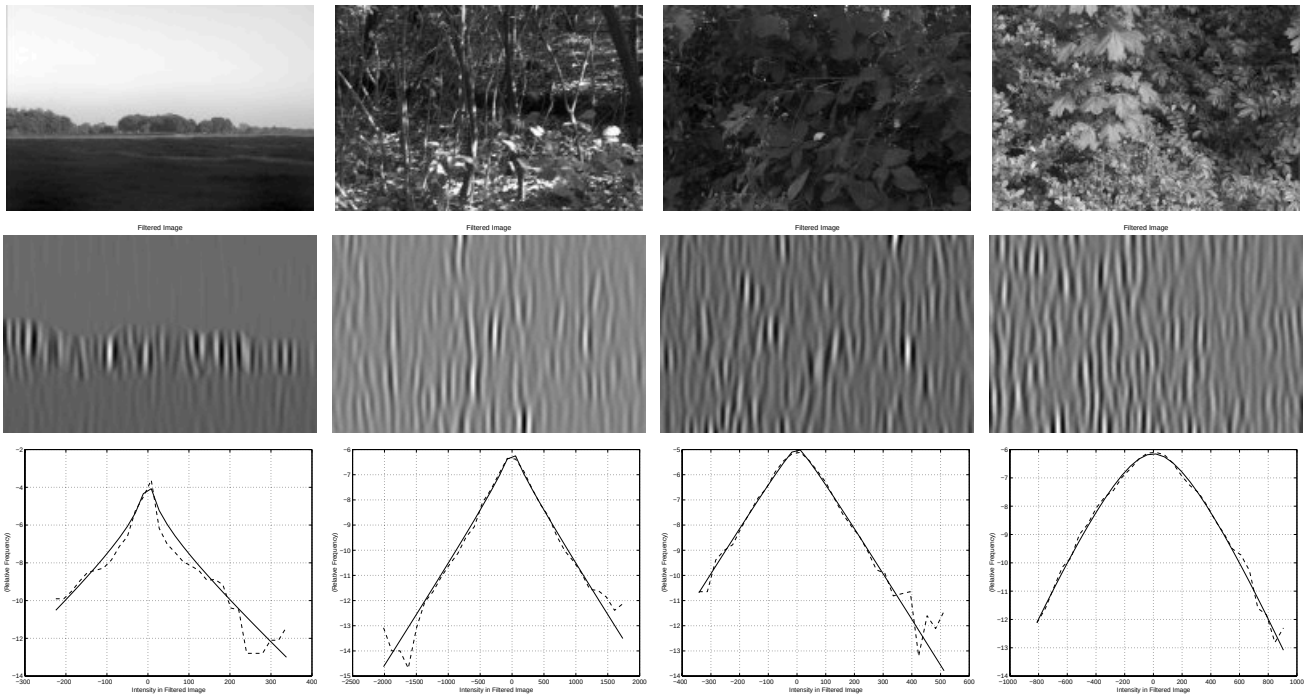

Figure 8: Variation of p-values for extracting vertical edges ($\theta = 90$). Top panels are the original images, middle panels are the filtered images, and the bottom panels are the densities (log-scale). The estimated p-values are: 0.26, 0.73, 1.21, and 3.74, respectively.

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 imager 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 Application 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) clutter classification, and (ii) target recognition.

4.1 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 Eqn. 9 to classify the clutter types from their images. We will demonstrate the strength of this model in the context

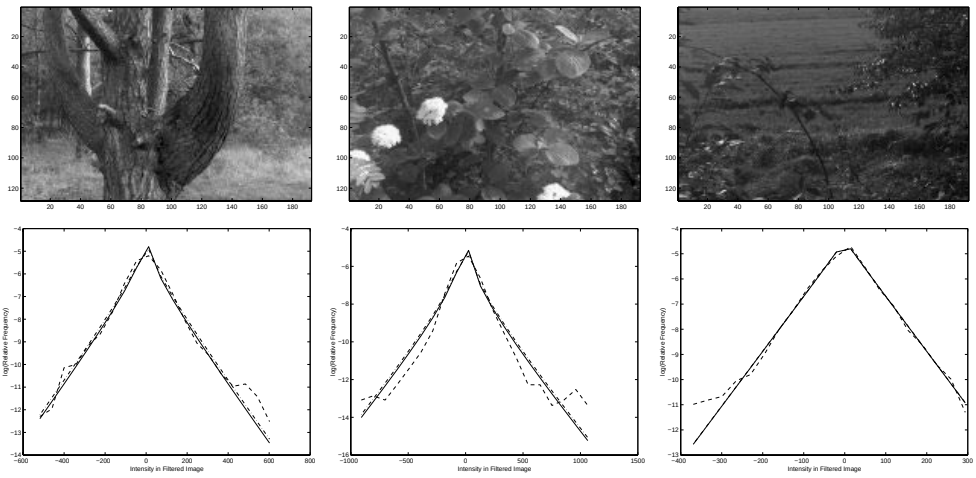

Figure 9: Asymptotic approximations of Bessel forms: For the images shown in top panels, the bottom panels plot the observed histogram (- - -), Bessel K form (-.-.-), and the approximation (solid line).

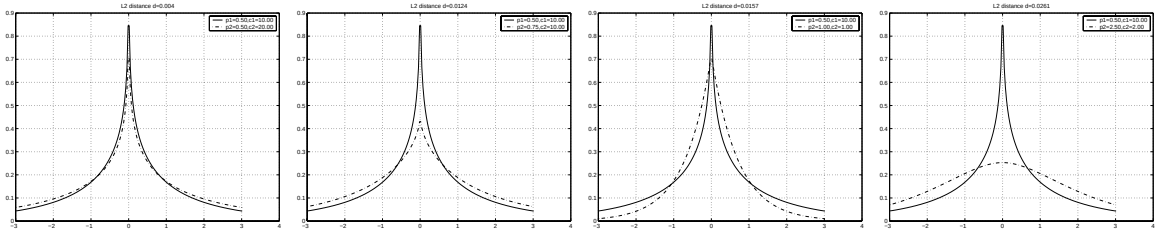

Figure 10: Examples of the L^2-metric on $\mathcal{D}$: (i) $p_2 = 0.5$, $c_2 = 20$, $d = 0.063$, (ii) $p_2 = 0.75$, $c_2 = 10$, $d = 0.114$, (iii) $p_2 = 1.0$, $c_2 = 1.0$, $d = 0.125$, (iv) $p_2 = 2.5$, $c_2 = 2.0$, $d = 0.161$. $p_1 = 0.5$ and $c_1 = 10.0$ are held constant.

of natural clutter classification. Consider the images of natural clutter shown in Figure 11. 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 Figure 12 using the dendrogram function in matlab. This function generates a clustering tree for points in image space when their pairwise distances are given. The clustering of I_1 with I_2, I_3 with I_4, and so on, demonstrates the success of this representation and the metric chosen. This results suggest a role for Bessel representations in clutter classification. For comparison we run clustering program using a 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.

Shown in Figure 13 is another example of natural images with the clustering, based on d_I, as

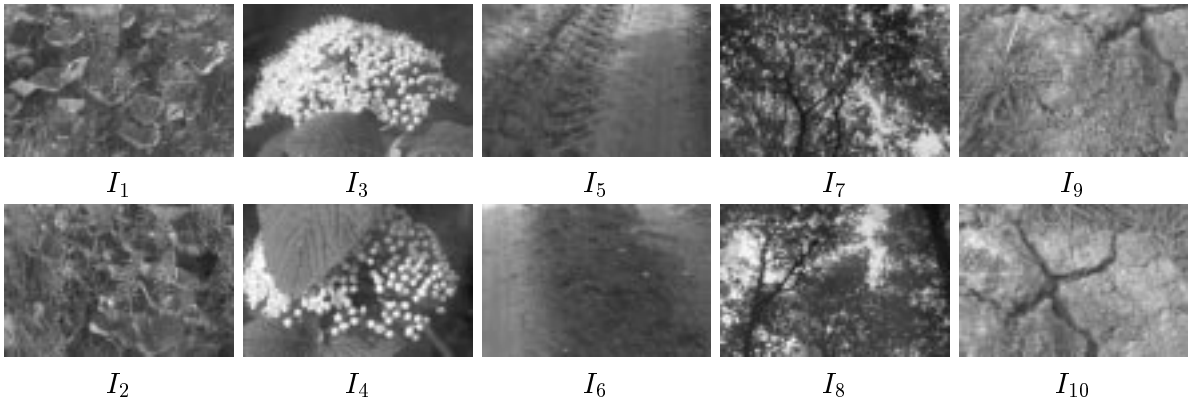

Figure 11: Ten natural images from the Groningen database: top row are the training images and bottom row are the test images.

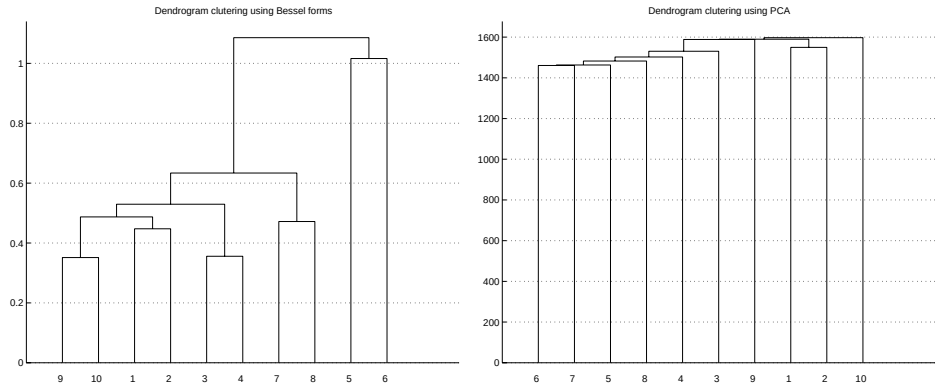

Figure 12: Dendrogram clustering of images in Figure 11 using d_I (left panel) and using a Euclidean metric on PCA (right panel).

shown in the right panel of Figure 12.

4.2 Pruning Hypotheses for Target Recognition

The Bessel K forms can 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 [10]. As stated under Case 1 in Section 1.1, this hypothesis selection is often performed using detailed models involving physical shapes, texture, pose and motion [23, 8, 10]. 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.

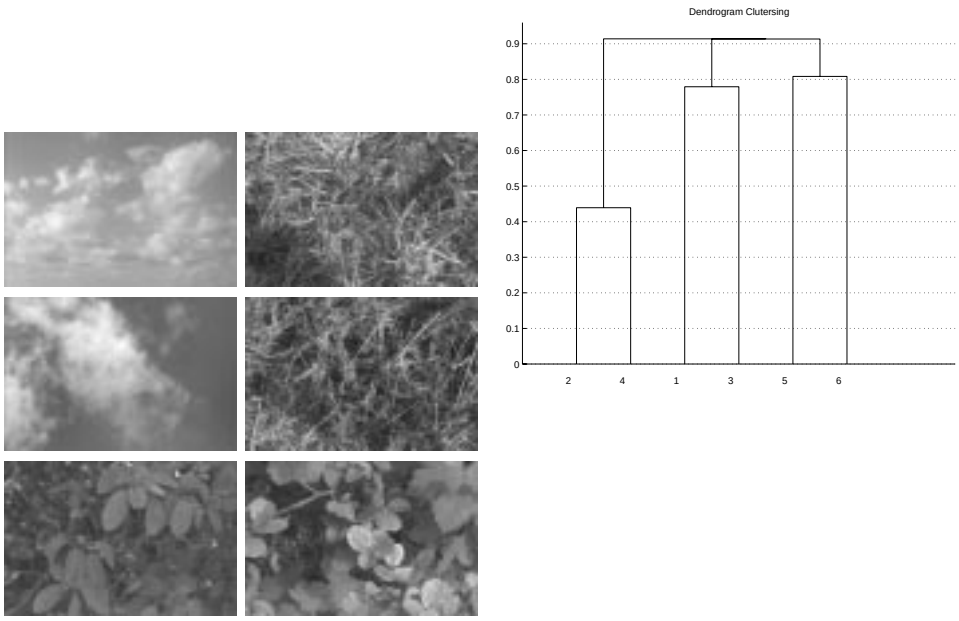

Figure 13: For the images shown in left, a dendrogram clustering plot using Bessel K forms is shown in the right panel.

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

$$P(\alpha|I) = \frac{\exp\left(-\min_{s\in S}(\sum_{j=1}^{K} d(p_{obs}^{(j)}, c_{obs}^{(j)}, p_{\alpha,s}^{(j)}, c_{\alpha,s}^{(j)})^2)/D\right)}{\sum_{\alpha'}\exp\left(-\min_{s\in S}(\sum_{j=1}^{K} d(p_{obs}^{(j)}, c_{obs}^{(j)}, p_{\alpha',s}^{(j)}, c_{\alpha',s}^{(j)})^2)/D\right)}\ , \tag{10}$$

where D 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 all $\alpha \in \mathcal{A}$, $s \in S$, and $j \in \{1, 2, \ldots, K\}$.

To illustrate this idea, consider the following experiment. Shown in Figure 14 are some sample images of objects from the Columbia object image library (COIL) [20]. 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 [20]. 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 Eqn. 10. Shown in Figure 15 are the plots of $P(\alpha|I)$ versus α (for $D = 0.5$) for 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 top 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. Bottom middle plot displays the worst case of the whole experiment and still short-lists 35 objects.

Figure 14: Sample images of objects from COIL image database.

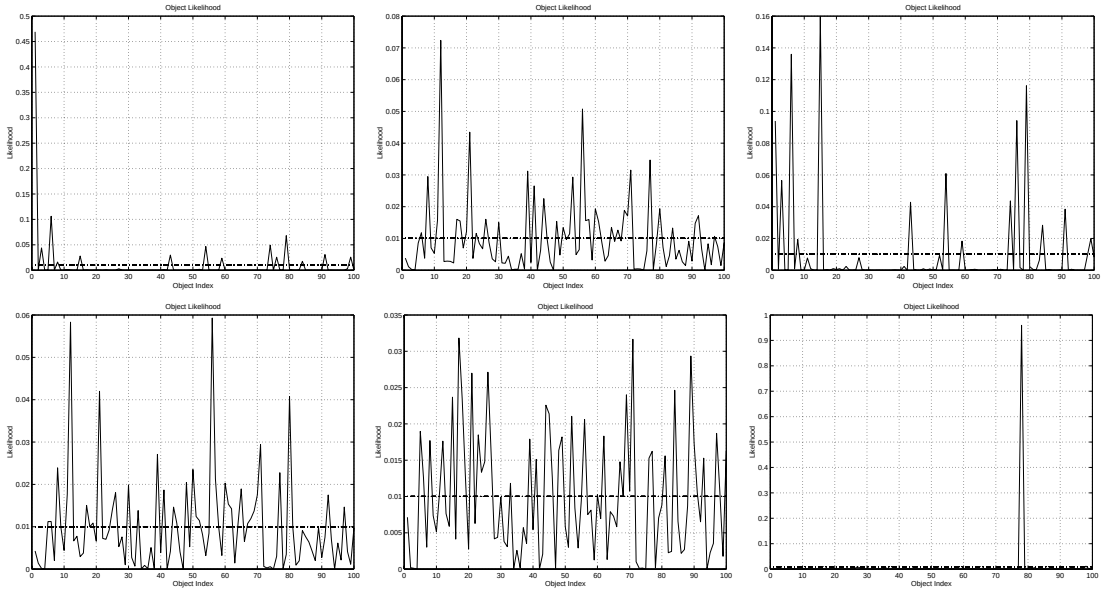

Figure 15: Plots of $P(\alpha|I)$ versus α for six test images in the COIL database. The test images are of objects α_1, α_{12}, α_{15}, α_{12}, α_{15}, and α_{78}, respectively, for arbitrary orientations. Dotted lines suggest a threshold level for pruning.

To support the use of Bessel K models in hypothesis pruning, we have actually used $P(\alpha|I)$ for object recognition and have compared results with some other recently proposed procedures: principal component analysis (PCA), independent component analysis (ICA), support vector machines (SVM), and SNoW. Pontil and Verri [20] 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. [25], 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.

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

Training/test per object	PCA	ICA	SNoW [25]	SVM [20]	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%

4.3 Texture Synthesis

To further illustrate the strength of Bessel representations, we present some examples of texture synthesis. For homogeneous textures, it is possible to completely characterize them using their spectral responses. As described in [27, 26], choosing the filtered marginals for several filters as sufficient statistics leads to a Gibbs' distribution on the image space. This points to a natural Gibbs' type MCMC sampling method to generate high probability images from the probability model. Zhu et al. have utilized the observed marginals to characterize the Gibbs' measure and then derived MCMC algorithms to sample from it. Using the same sampling scheme, except the observed marginals are now replaced by the **estimated Bessel K forms**, we have generated high probability samples on the image space. The Gibbs' distribution on the image is space is given by:

$$f_I(I) = \frac{1}{Z} \exp(-\sum_{j=1}^{K} \lambda_j \| H(I^{(j)}) - f(x; p^{(j)}, c^{(j)}) \|^2) \ ,$$

where H denotes the observed histogram and Z is a normalizing constant. We have used the same procedure as in [26] to estimate the most likely values of the Lagrangean weights λ_j and have applied them to characterize f_I. It must be noted that the marginal density of the image intensity I is **not included** in the representation; only the Bessel K forms for the Gabor spectral components (for $K = 39$) are involved. Shown in the top panels of Figure 16 are the given texture images used to estimate the Bessel K parameters, and shown in the bottom panels are the corresponding samples from a distribution based on the estimated parameters.

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 observed histograms for a variety of images: video, IR, and range, for gradient, Gabor and Laplacian 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 target 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. Bessel parameter p is related to the distinctness and the frequency of occurrence of the filtered characteristics of imaged objects.

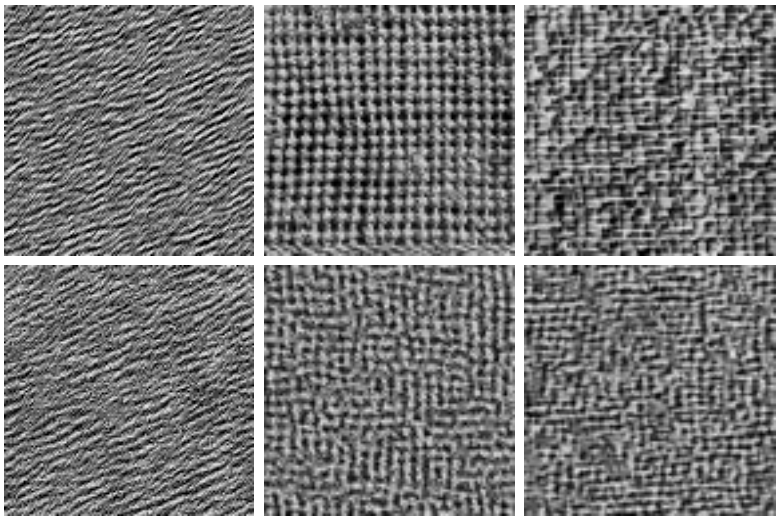

Figure 16: Top row: observed images of the textures. Bottom row: synthesized images using the Bessel K densities.

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, 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*, 50:145–159, 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*, 19(7):711–720, 1997.

[3] A. J. Bell and T. J. Sejnowski. "The "independent components" of natural scenes are edge filters". *Vision Research*, 37(23):3327–3338, 1997.

[4] C. Chubb, J. Econopouly, and M. S. Landy. Histogram contrast analysis and the visual segregation of iid textures. *J. Opt. Soc. Am. A*, 11:2350–2374, 1994.

[5] P. Comon. Independent component analysis, a new concept? *Signal Processing, Special issue on higher-order statistics*, 36(3), 1994.

[6] I. S. Gradshteyn and I. M. Ryzhik. *Table of Integral Series and Products, Editor: Alan Jeffrey.* Academic Press, 2000.

[7] U. Grenander. *General Pattern Theory.* Oxford University Press, 1993.

[8] U. Grenander, M. I. Miller, and A. Srivastava. Hilbert-schmidt lower bounds for estimators on matrix lie groups for atr. *IEEE Transactions on PAMI*, 20(8):790–802, 1998.

[9] U. Grenander and A. Srivastava. Probability models for clutter in natural images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(4), April 2001.

[10] U. Grenander, A. Srivastava, and M. I. Miller. Asymptotic performance analysis of bayesian object recognition. *IEEE Transactions of Information Theory*, 46(4):1658–1666, 2000.

[11] D. J. Heeger and J. R. Bergen. Pyramid-based texture analysis/synthesis. In *Proceedings of SIGGRAPHS*, pages 229–238, 1995.

[12] B. Jahne, H. Haubecker, and P. Geibler. *Handbook of Computer Vision and Applications, Volume 2*. Academic Press, 1999.

[13] B. Julesz. A theory of preattentive texture discrimination based on first-order statistics of textons. *Biological Cybernetics*, 41:131–138, 1962.

[14] 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*, 12(1):103–108, 1990.

[15] 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*, 41(1,2), 2001.

[16] D. Marr. *VISION: A computational Investigation into the Human Representation and Processing of Visual Information*. W. H. Freeman and Company, New York, 1982.

[17] 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*, 43(11):2678–2690, November 1995.

[18] 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.

[19] B. A. Olshausen and D. J. Field. Sparse coding with an overcomplete basis set: A strategy employed by v1? *Vision Research*, 37(23):3311–3325, 1997.

[20] M. Pontil and A. Verri. Support vector machines for 3d object recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(6):637–646, 1998.

[21] J. Portilla and E. P. Simoncelli. A parametric texture model based on join statistics of complex wavelets. *International Journal of Computer Vision*, 40(1), 2000.

[22] S. T. Roweis and L. K. Saul. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290:2323–2326, 2000.

[23] A. Srivastava, M. I. Miller, and U. Grenander. Bayesian automated target recognition. *Handbook of Image and Video Processing, Academic Press*, pages 869–881, 2000.

[24] M. J. Wainwright, E. P. Simoncelli, and A. S. Willsky. Random cacaeds on wavelet trees and their use in analyzing and modeling natural images. *Applied and Computational Harmonic Analysis*, 11:89–123, 2001.

[25] 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*, volume 1, pages 439–454, 2000.

[26] 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*, 22:554–569, 2000.

[27] S. C. Zhu, Y. N. Wu, and D. Mumford. "Minimax entropy principles and its application to texture modeling". *Neural Computation*, 9(8):1627–1660, November 1997.

A Appendix

A.1 Proof of Theorem 1

To establish the theorem, we will need the integral formula ([6] page 676, Eqn. 4):

$$\begin{aligned}\int_0^\infty x^{-\lambda} K_\mu(ax) K_\nu(bx) dx &= \frac{2^{-2-\lambda} a^{-\nu+\lambda-1} b^\nu}{\Gamma(1-\lambda)} \Gamma(\frac{1-\lambda+\mu+\nu}{2}) \Gamma(\frac{1-\lambda-\mu+\nu}{2}) \\ &\quad \Gamma(\frac{1-\lambda+\mu-\nu}{2}) \Gamma(\frac{1-\lambda-\mu-\nu}{2}) \\ &\quad F(\frac{1-\lambda+\mu+\nu}{2}, \frac{1-\lambda-\mu+\nu}{2}; 1-\lambda; 1-\frac{b^2}{a^2}) \ , \end{aligned} \tag{11}$$

where $\Re(a+b) > 0$, $\Re(\lambda) < 1 - |\Re(\mu)| - |\Re(\nu)|$. F is the hypergeometric function; it is an infinite series in its last argument.
To derive the L^2-metric, we start with its square:

$$d(p_1, c_1, p_2, c_2)^2 = \int_x f(x; p_1, c_1)^2 dx + \int_x f(x; p_2, c_2)^2 dx - 2 \int_x f(x; p_1, c_1) f(x; p_2, c_2) dx \ .$$

Consider these terms one by one, starting with the first term:

$$\begin{aligned}&\frac{1}{Z(p_1, c_1)^2} \int_x x^{2p_1-1} K_{(p_1-0.5)}(\sqrt{\frac{2}{c_1}} x) K_{(p_1-0.5)}(\sqrt{\frac{2}{c_1}} x) dx \\ &= \frac{1}{Z(p_1, c_1)^2} 2^{-3} (2c_1)^{p_1} \frac{\Gamma(2p_1 - 0.5)}{\Gamma(2p_1)} \Gamma(p_1)^2 \Gamma(0.5) = \frac{\sqrt{2}}{4\pi} \Gamma(0.5) \left(\frac{1}{\sqrt{c_1}} \frac{\Gamma(2p_1 - 0.5)}{\Gamma(2p_1)} \right) \ ,\end{aligned}$$

using the integral formula Eqn. 11. Similarly, the second term becomes $\frac{\sqrt{2}}{4\pi} \Gamma(0.5) \left(\frac{1}{\sqrt{c_2}} \frac{\Gamma(2p_2 - 0.5)}{\Gamma(2p_2)} \right)$. Substituting for the integral in the cross term gives:

$$\begin{aligned}&\frac{1}{Z(p_1, c_1) Z(p_2, c_2)} \int_x x^{p_1+p_2-1} K_{(p_1-0.5)}(\sqrt{\frac{2}{c_1}} x) K_{(p_2-0.5)}(\sqrt{\frac{2}{c_2}} x) dx \\ &\quad = \frac{1}{Z(p_1, c_1) Z(p_2, c_2)} 2^{-3} 2^{p_1} c_1^{p_2} (\sqrt{\frac{2}{c_1}})^{-p_1+0.5} (\sqrt{\frac{2}{c_2}})^{p_2-0.5} \frac{\Gamma(p_1 + p_2 - 0.5)}{\Gamma(p_1 + p_2)} \\ &\qquad \Gamma(p_1) \Gamma(p_2) \Gamma(0.5) F((p_1 + p_2 - 0.5), p_2; p_1 + p_2; 1 - \frac{c_1}{c_2}) \\ &\quad = \frac{\sqrt{2}}{4\pi} \Gamma(0.5) \left(\frac{1}{\sqrt{c_1}} (\frac{c_1}{c_2})^{p_2} \frac{\Gamma(p_1 + p_2 - 0.5)}{\Gamma(p_1 + p_2)} \mathcal{F} \right) \ .\end{aligned}$$

Combining these three terms, the result in Eqn. 8 follows. It should be noted that the metric is symmetric in the parameters (p_1, c_1) and (p_2, c_2), even though it does not appear that way from the expression in Eqn. 8. The condition associated with the formula Eqn. 11 implies that p_1 and p_2 must be greater than 0.25.