Statistical hypothesis pruning for identifying faces from infrared images

Anuj Srivastava[a,*], Xiuwen Liu[b]

[a]*Department of Statistics, Statistical Consulting Center, Florida State University, Tallahassee, FL 32306-4330, USA*
[b]*Department of Computer Science, Florida State University, Tallahassee, FL 32306-4330, USA*

Abstract

A Bayesian approach to identify faces from their IR facial images amounts to testing of discrete hypotheses in presence of nuisance variables such as pose, facial expression, and thermal state. We propose an efficient, low-level technique for hypothesis pruning, i.e. shortlisting high probability subjects from given observed image(s). (This subset can be further tested using some high-level model for eventual identification.) Hypothesis pruning is accomplished using wavelet decompositions (of the observed images) followed by analysis of lower-order statistics of the coefficients. Specifically, we filter infrared (IR) images using bandpass filters and model the marginal densities of the outputs via a parametric family that was introduced by Grenander and Srivastava [IEEE Trans. Pattern Anal. Mach. Intell. 23 (2001) 424]. IR images are compared using an L^2-metric between the Marginals computed directly from the parameters. Results from experiments on IR face identification and statistical pruning are presented.
© 2003 Elsevier Science B.V. All rights reserved.

Keywords: Infrared image analysis; Nighttime face identification; Bessel K forms; Image statistics; Hypothesis selection

1. Introduction

Automated identification of people from their facial images has become an active area of research in recent years. In particular, infrared (IR) face identification is an area of great potential with applications in both civilian and military domains. Past research in IR image analysis has severely been restricted due to limited public availability of large datasets. But recent access to cheap, hand-held IR cameras has addressed this issue, and has made IR image analysis an area of growing activity (see Ref. [24] for a discussion). While face identification using visible spectrum images is very well discussed in the literature (see for example Refs. [3,13,23] and references therein), the use of IR in face identification has seen only limited research with a few exceptions [6,25]. In addition to identification, IR imaging has recently been utilized for learning the emotional state of a person [22].

An important issue is: What probability models efficiently and sufficiently describe facial images for use in face identification? Similar to the visible spectrum image analysis, statistical inferences are central to IR image analysis and understanding. Keep in mind that images are elements of a very high-dimensional space, and they have to be projected down (either linearly or non-linearly) to smaller dimensions before any statistical model can be imposed. There are two broad categories of models proposed in the literature: in each case images are considered as realizations of a spatial process on some domain in $\mathbb{R}^2$.

1. *Low-level models* directly analyze the image pixels to extract features of interest and impose probabilities on the features. If the image domain is a finite, uniform grid in $\mathbb{R}^2$, then the image space is a finite-dimensional vector space and can be projected linearly to a low-dimensional vector space. Choices of linear projections have included principal components (PCA) [17], independent components (ICA) [2,4], linear discriminant (LDA) [1], sparse coding [21], etc. and they provide bases of analyzing images through their coefficients. Instead, if the images are considered as functions on a continuous, rectangular domain in $\mathbb{R}^2$, then harmonic analysis provides several bases for dimension reduction. Fourier [8], wavelets [20], ridgelets [7], filtering operators, etc. are some examples. In view of the locality of objects in images, wavelet bases have become important in image analysis. In particular, it is common to use Gabor wavelets to decompose observed images simultaneously in space and frequency. Since these representations are non-physical, they retain only limited knowledge of object shapes, geometries, placements, etc.

However, their advantage lies in the efficiency of their implementation.

2. *High-level models* compare observed images with the images synthesized from 3D templates of known objects, to find the best match and to impose probabilities on (template) transformations. These models use physical principles to capture the image variations. Even though these representations capture object shapes, and are robust enough for object recognition, their applications have been rather limited. Given tremendous variability associated with the imaged scenes, high-level models may not be feasible in modeling all possible objects present in a scene.

In IR face identification, the goal is to identify facial shape or some geometrical features, a task that can be accomplished by using high-level models. But to obtain computational efficiency, one aims to utilize a combination of low-level and high-level techniques to achieve a desired performance level. In this paper, we present a low-level technique to analyze given IR facial images and to generate a set of high-probability subjects. The selected subjects can then be tested under a high-level system for face identification although that step is not explored here in this paper. To start with, we motivate and explain the goals of this paper.

1.1. Why infrared face recognition?

Most frequently used sensors for performing face identification are the visible spectrum cameras that collect static or video sequences of scenes containing people. Although visible spectrum images, in general, provide sufficient resolution, sensitivity and discrimination to identify people, their application in designing automated face identification systems has seen only limited success. One major difficulty lies in the camera sensitivity towards nuisance variables such as the nature of illumination shining on the imaged face. Visible spectrum images change drastically depending upon the location, color, and other properties of the incident light. On the other hand, IR cameras provide a measure of thermal emissivity from the facial surface and are relatively stable to the illumination variability. IR cameras operate in the wavelengths ranging from 0.78 to 24 μm, and provide images for identification even in dark environments. Continuing advances in IR camera technology has resulted in cheap, hand-held IR cameras, both cooled and uncooled, that can provide robust, high-resolution pictures in all regions of IR spectrum.

We qualitatively illustrate the stability of IR face images, by studying some images of a subject in widely different thermal states: in outdoor breezy conditions, indoor in a warm room, after a workout, etc. To generate large variability, we even forced some artificial thermal states by wiping the face with a wet cloth or having long exposure to a full sun. Shown in Fig. 1 is a sample of these pictures that demonstrate the ability of IR imaging to provide sufficient resolution and discrimination, and yet be robust to change in thermal states. Top panels show some pictures taken under different conditions, while bottom panels display some artificial thermal states forced to generate a larger variability. Qualitatively, it follows that the variability in these pictures is much less than the variability in visible spectrum images caused by illumination differences. For a formal quantitative analysis of invariance of IR images to incident illumination, refer to Ref. [29].

Limitations of IR cameras, such as their inability to capture visible spectrum (colors) or having resolutions less than visible spectrum cameras, should also be acknowledged. Perhaps a more comprehensive system should include both visible and IR spectrum cameras to avoid individual pitfalls and to perform identification as joint statistical inference on multi-sensor data.

Fig. 1. IR pictures of a subject's face under widely varying thermal conditions. Bottom panels show artificial thermal states generated to study large variability.

1.2. What is statistical hypothesis pruning?

A comprehensive way to characterize face images is to develop 3D models of the faces, including the facial meshes and thermal emissivities, and search over these physical representations given the observed images [15]. As shown in Fig. 2, a facial surface can be represented by a dense polygonal mesh (left panel) and the thermal state can be denoted by texture mapping a thermal field (middle panel) on this mesh (right panel) [28]. In this setup, face identification is performed using detailed models on physical geometries, thermal emissivities, relative pose and motion [5,18]. The main idea is to find transformations of 3D templates such that their synthesized images best match the observed images. The transformations form a set of nuisance variables for hypothesis selection.

Let $\mathscr{A}$ be the discrete index set of all possible subjects that can be present in an image. Given an image I, or a collection of images, our goal is to detect and identify subject(s) present in that image. Bayesian identification corresponds to selecting a subject with the largest posterior probability, i.e. the recognized identity is given by $\hat{\alpha} = \text{argmax}_{\alpha \in \mathscr{A}} P(\alpha|I)$, where $P(\alpha|I)$ is the posterior probability: $P(\alpha|I) = P(I|\alpha)P(\alpha)/P(I)$, for $\alpha \in \mathscr{A}$ and I is an image. Here $P(I|\alpha)$ is the likelihood function of α for a fixed image I and is obtained as an integral over the set S of all nuisance variables (such as pose, thermal variation, motion, etc. as described in Refs. [5,12]) according to: $P(I|\alpha) = \int_S P(I|\alpha, s)P(s|\alpha)\gamma(ds)$, where $\gamma(ds)$ is a reference measure on S. This setup is motivated by Grenander's *deformable template* theory [9] and has been applied to the analysis of IR images of military targets in Refs. [5,18]. In general, the product space $S \times \mathscr{A}$ is huge and it is computationally expensive to implement a strategy that integrates over all of $S \times \mathscr{A}$ to find $\hat{\alpha}$. Instead, it is useful to prune the set $\mathscr{A}$ to only a small subset $\mathscr{A}_0$ of high probability subjects, using an efficient low-level technique, and apply the high-level estimation only to the set $S \times \mathscr{A}_0$. We term the process of selecting the subset $\mathscr{A}_0$, for a given observation I, as *statistical hypothesis pruning*. It is accomplished by defining a new probability $\tilde{P}(\alpha|I)$ that is faster to compute and satisfies the condition that high probability α's under $\tilde{P}(\alpha|I)$ also have high probabilities under $P(\alpha|I)$. In other words, choose a distribution $\tilde{P}$, and constants γ_1, γ_2 such

that for all images I, we have:

$$\{\alpha \in \mathscr{A}|P(\alpha|I) \geq \gamma_1\} \subseteq \{\alpha \in \mathscr{A}|\tilde{P}(\alpha|I) \geq \gamma_2\}.$$

In this paper we do not study the use of $\mathscr{A}_0$ as an input to a deformable template method for face identification. We refer to the existing literature [5,10,18] for such studies.

1.3. Strategies for hypothesis pruning

How to find a probability $\tilde{P}(\alpha|I)$ that can be used for statistical hypothesis pruning? Requirement of efficiency dictates the use of a low-level technique. There are a number of techniques that extract low-level features from images and image analysis then becomes a statistical analysis of these extracted features.

Motivated by a growing understanding of early human vision, a popular strategy has been to decompose images into their wavelet coefficients using a family of Gabor wavelets. Once the images are decomposed the image statistics can be studied via the statistics of the coefficients. So the next question is: what probability models sufficiently characterize the behavior of wavelet coefficients of images? Zhu et al. [30] have shown that the (univariate) histograms of the wavelet coefficients, obtained using a collection of wavelets at different scales and orientations, capture sufficient local features to characterize homogeneous textures. The choice of histograms as sufficient statistics implies that only the frequencies of the coefficients are relevant and their location information is discarded [14]. To gain efficiency, the histograms have been successfully replaced by a family of parametric densities, as described in Refs. [11,27]. The three parameters of the density completely characterize the marginal density of wavelet coefficients, denoting a significant reduction in the representation, and are shown to depend only on mean, variance and kurtosis of the coefficients, thereby implying a simple estimation procedure. $\tilde{P}$ will be defined using a metric on the space of these parametric densities, and the parametric nature of these representations will make the computation of $\tilde{P}$ efficient.

The experiments described later are based on the following databases. FSU face dataset consists of images taken by a Raytheon PalmIR PRO thermal camera that operates in $7-14 \mu m$ spectral range and generates images of size 320×240 in 8-bit BMP format at a (typical) sensitivity level of 100 mK. Shown in Fig. 9 are some examples; this database consists of 20 frontal IR images of each of the nine subjects, at varying angles and facial expressions. In addition, we have also used the Equinox database of IR face images, and IR images provided at the CVBVS'01 web-site, courtesy of Honeywell Laboratories.

This paper is organized as follows. Section 2 describes the process of wavelet decomposition of images and utilizes a parametric probability to model the marginal densities of the wavelet coefficients. Section 3 uses an analytical expression for L^2 metric between these probabilities to
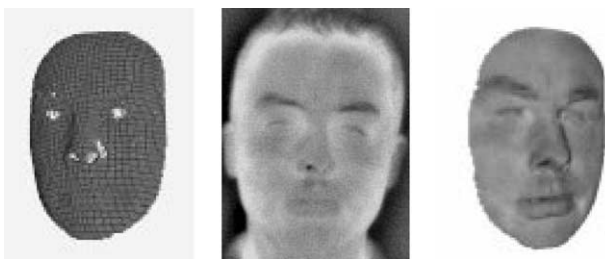

Fig. 2. 3D model for IR face images: the surface geometry (left panel) is registered to the thermal texture (middle panel) and results in a 3D model of IR face (right panel).

impose a pseudo-metric on IR image space. Section 4 defines a pruning probability $\tilde{P}$ and describes experiments to illustrate these ideas for IR face identification.

2. Models for wavelet decompositions of images

Our choice of low-level image representation is based on linear bandpass filtering and wavelet decomposition of the observed images. One reason for performing wavelet decomposition is that wavelet coefficients better lend to statistical modeling than the original images. Before we present details, we introduce some notation. Given an IR image I and a bank of filters $\{F^{(j)}, j = 1, 2, ..., J\}$, we compute, for each filter $F^{(j)}$, a filtered image $I^{(j)} = I * F^{(j)}$, where $*$ denotes the 2D convolution operation. Elements of $I^{(j)}$s are called the wavelet coefficients of image I. Let the pixel location be denoted by $z = [z_1 \; z_2] \in \mathbb{R}^2$. Following is a listing of the filters used in this paper:

1. *Derivative filters*. Pixel values in the directional derivative of an image I are given by:

$$a_1 \frac{\partial I}{\partial z_1} + a_2 \frac{\partial I}{\partial z_2}$$

   for $a_1, a_2 \geq 0$, $a_1^2 + a_2^2 = 1.0$. Set $a_1 = 0$ for extracting the horizontal edges, or $a_2 = 0$ for extracting the vertical edges.

2. *Gabor filters*. These are bandpass filters with Gaussian kernels centered around specific wavenumbers (see Ref. [16] for details). For an angle θ, a Gabor filter is given by:

$$F_{\sigma,\theta}(z) \equiv \exp(-(\tilde{z}_1^2 + (\tilde{z}_2)^2)/2\sigma^2)\exp\left(-j\frac{2\pi\tilde{z}_1}{\sigma}\right),$$

$$\begin{bmatrix} \tilde{z}_1 \\ \tilde{z}_2 \end{bmatrix} = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \end{bmatrix},$$

   where σ denotes the scale/resolution associated with the filter.

3. *Laplacian of Gaussian filter*. Higher order spatial derivatives of an image also form important features for image analysis. As an example, the Laplacian of an image preceded by a Gaussian smoothing makes a useful filter. In this case, the filtered image is obtain by the operation $\nabla(I * G)$ where

$$\nabla = \frac{\partial^2}{\partial z_1^2} + \frac{\partial^2}{\partial z_2^2},$$

   and

$$G = \exp\left(-\left(\frac{z_1^2 + z_2^2}{2\sigma^2}\right)\right).$$

In addition to the above-mentioned filters, one can also utilize other filters: neighborhood operators, steerable filters, interpolation filters, and so on. Each class of filter selects and isolates certain features present in the original image. In this paper, we do not address the issue of optimal filter selection in the context of face identification although it is an interesting problem by itself. Instead, we will assume an arbitrary choice of filters. Please refer to Ref. [19] for a discussion on selecting optimal filters for a specific application.

Statistical modeling of images has been of great focus in last few years, and a major discovery has been the non-Gaussianity of image representations. (A review of recent results in statistical modeling of images is presented in Ref. [26].) As an example, wavelet coefficients of images are found to have high kurtosis, heavy tails, and sharp cusps at the median. Several probability models have been proposed to explain this phenomenon, and in recent papers [11,27] we have proposed a three-parameter family of probability densities to model the univariate density of wavelet coefficients. The starting point of this derivation is a (high-level) physical model which assumes images to be a superposition of views (or 2D profiles) of arbitrary targets from arbitrary poses, and at arbitrary thermal states. Under this model, and certain other simplifying assumptions, the resulting density function of a wavelet coefficient is given by:

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

for $p > 0$, $c > 0$,

where K is a modified Bessel function of third kind and Z is the normalizing constant given by $Z(p, c) = \sqrt{\pi}\Gamma(p) \times (2c)^{0.5p+0.25}$. $\mu = 0$, implies a two parameter family, a case that was studied in the original papers. These densities are referred to as the *Bessel K forms* and the parameters (μ, p, c) as the *Bessel parameters*. Bessel K forms are unimodal, symmetric, and leptokurtic.

An image representation scheme is shown in Fig. 3. Each image I is decomposed (via filtering) into J sets of wavelet coefficients, and the marginal density of each component is modeled using a three-parameter Bessel form. So the original image I is now represented by $3J$ parameters $\{(\mu^{(1)}, p^{(1)}, c^{(1)}), ...(\mu^{(J)}, p^{(J)}, c^{(J)})\}$. We utilize a moment-based estimation procedure to estimate μ, p and c :

$$\hat{\mu}^{(j)} = \mathrm{SM}(I^{(j)}), \qquad \hat{p}^{(j)} = \frac{3}{\mathrm{SK}(I^{(j)})},$$

$$\hat{c}^{(j)} = \frac{\mathrm{SV}(I^{(j)})}{\hat{p}^{(j)}}, \tag{1}$$

where SM is the sample mean, SV is the sample variance and SK is the sample kurtosis of the pixel values in the filtered image $I^{(j)}$. The computational task of estimating the marginal density is that of estimating
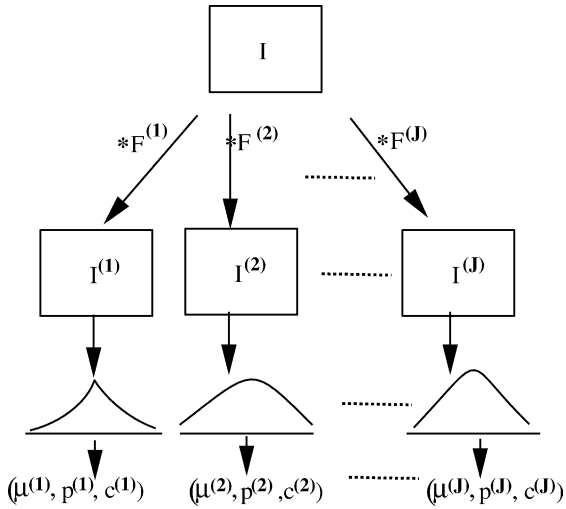
Fig. 3. Representation scheme: I is decomposed into J wavelet coefficients using filters $F^{(1)}, ..., F^{(j)}$. For each component, the marginal density is modeled by a Bessel K form parameterized by the triple $(\mu^{(j)}, p^{(j)}, c^{(j)})$.

the first, second and fourth moments only. Next we illustrate these Bessel K forms through some examples. Shown in the top panels of Fig. 4 are some sample IR images from FSU face database. The middle row displays the filtered images (or the wavelet coefficients) for arbitrary Gabor filters (not shown). Bottom panels plot the marginal densities of $I^{(j)}$ on a log scale: the observed densities or histograms are plotted in the marked lines and the estimated Bessel K forms $(f(x; \hat{\mu}, \hat{p}, \hat{c}))$ are plotted in plain lines. We have also

tested the Bessel K forms for other databases. Shown in Fig. 5 top panels are some additional examples, and in bottom panels are the associated marginal densities: both estimated and observed, each on a log scale. The first two pictures are from Honeywell dataset and the last two are from Equinox dataset.

Fig. 6 displays another example of estimating wavelet densities. For the IR image shown in the leftmost panel, we have estimated its marginal densities for a number of wavelet coefficients corresponding to Gabor filters with different orientations and scales. The estimated densities, along with the observed densities, are shown in the remaining six panels. It must be emphasized that Bessel K family models wavelets coefficients of images in general, and not just the face images or IR images. For different types of images, the Bessel parameters take on different values. Shown in Fig. 7 are three IR images of an indoor environment that can form background clutter for face identification. Shown in the lower panels are the estimated Bessel K forms of the filtered versions of these images, when filtered by arbitrary Gabor filters.

In our experiments, we have found a remarkable fit between the observed and the estimated marginals, for a large set of filtered IR images. To quantify the difference between an observed histogram and the estimated Bessel form, we utilize the Kullback-Leibler (KL) divergence on the space of univariate densities. It should be noted that KL is not a proper distance on the space of univariate densities but it does quantify the difference between any two densities
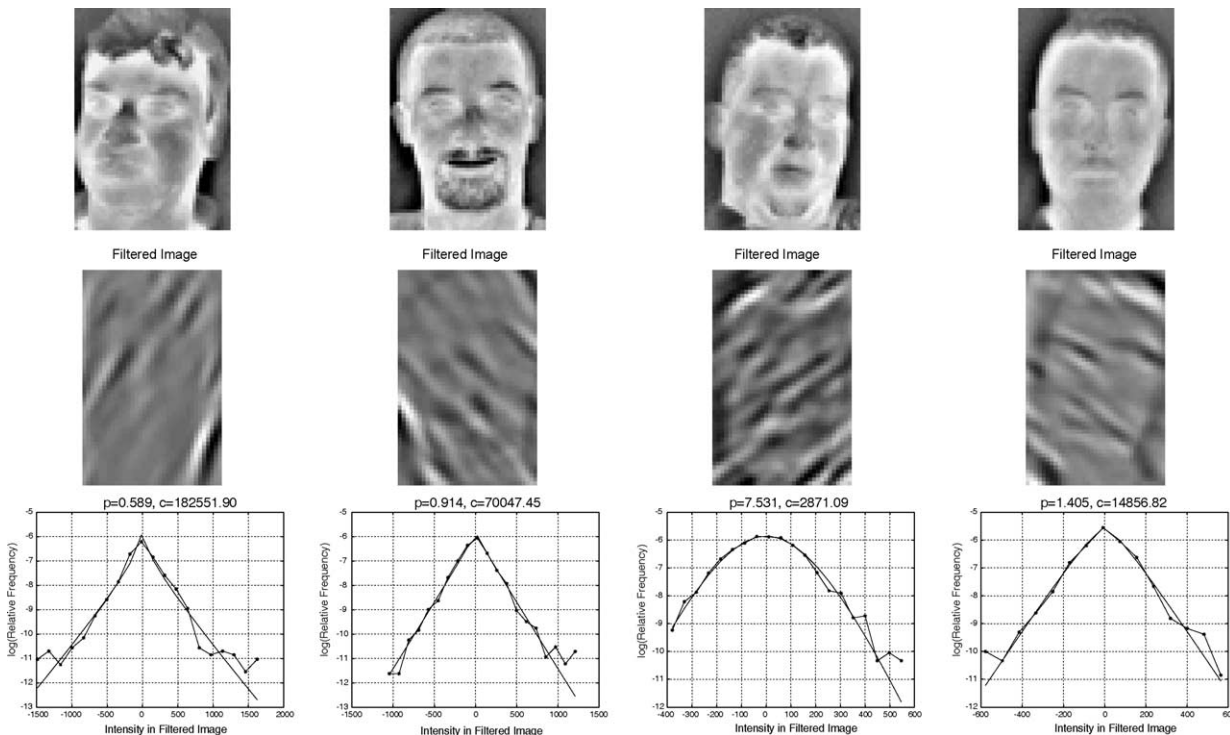

Fig. 4. Estimation of Bessel K forms for IR images: the top row shows some IR images, second row shows the filtered images for arbitrary Gabor filters, and the bottom row shows the estimated marginal densities (plain lines) drawn over the observed marginal densities (marked lines), both on a log scale.
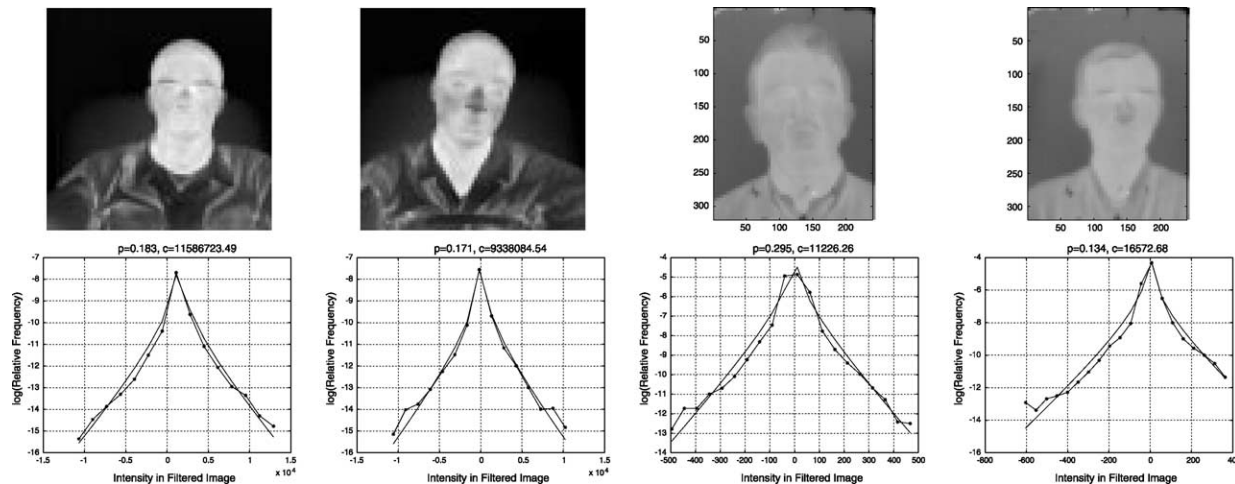
Fig. 5. Estimated Bessel K forms for sample images from Honeywell and Equinox datasets. Top panels: IR images, middle panels: filtered images for arbitrary Gabor filters, and bottom panels: observed (marked lines) and estimated (plain lines) marginal densities.

to some extent. Let $f_{\text{obs}}(x)$ be an observed marginal density and $\hat{f} \equiv f(x; \hat{\mu}, \hat{p}, \hat{c})$ be an estimated Bessel form, then the KL-divergence is computed as:

$$KL(\hat{f}, f_{\text{obs}}) = \int_{\mathbb{R}} f(x; \hat{\mu}, \hat{p}, \hat{c}) \log\left(\frac{f(x; \hat{\mu}, \hat{p}, \hat{c})}{f_{\text{obs}}(x)}\right) dx.$$

We have computed this quantity for a large combination of IR images (from FSU face dataset) and Gabor filters: using 180 face images and 180 Gabor filters (fixed scale but different orientations each one degree apart). Using Monte Carlo sampling to randomly select an image and a filter, this quantity was averaged over 3000 realizations and the average value is found to be 0.0503. The convergence plot is shown in the left panel of Fig. 8. To understand this value, shown on the right side are four

examples of computing KL-divergence: each plot shows an observed density, the corresponding estimated density, and the KL-divergence for that pair (listed on the top). The number 0.0503 underscores the goodness of fit obtained in modeling observed histograms via estimated Bessel K forms.

Since the Bessel K forms are derived using an image generation model, it is possible to provide a physical interpretation to the shape parameter p. As detailed in Ref. [27], p relates to the characteristics (shape, frequency of occurrence, etc.) of the imaged objects as seen in $I^{(j)}$, depending upon the choice of filter $F^{(j)}$. For instance, if the filter is an edge-detector, for a certain orientation, then p relates to the sharpness and the number of edges oriented in that direction, in the image
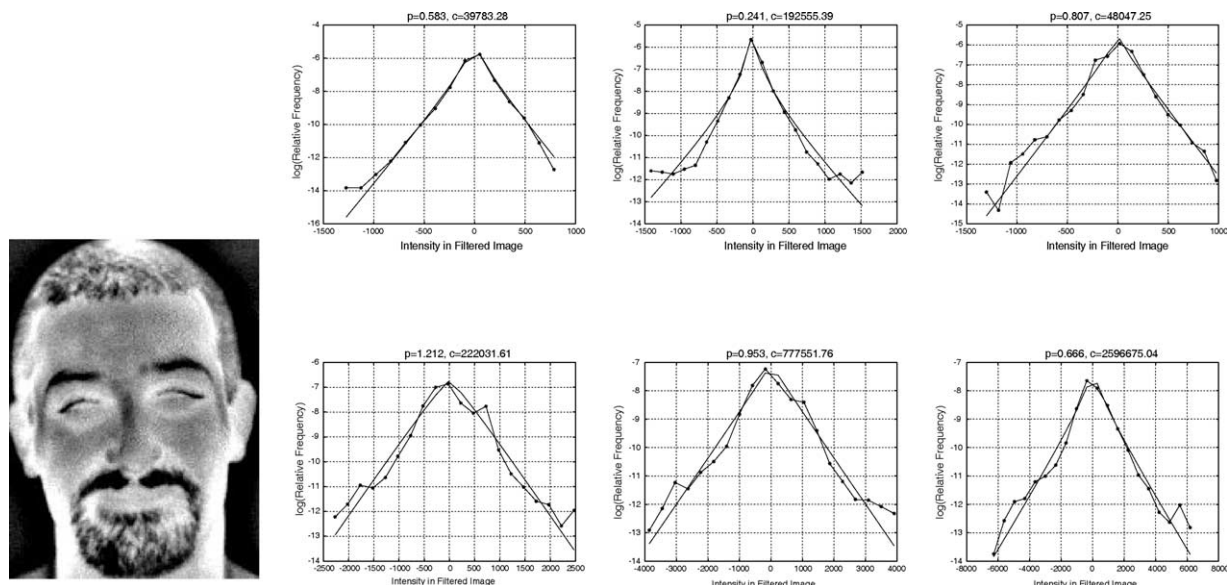

Fig. 6. Estimated marginal densities of several wavelet decompositions of the same image (left panel). Different wavelet coefficients correspond to different Gabor filter orientations and scales.
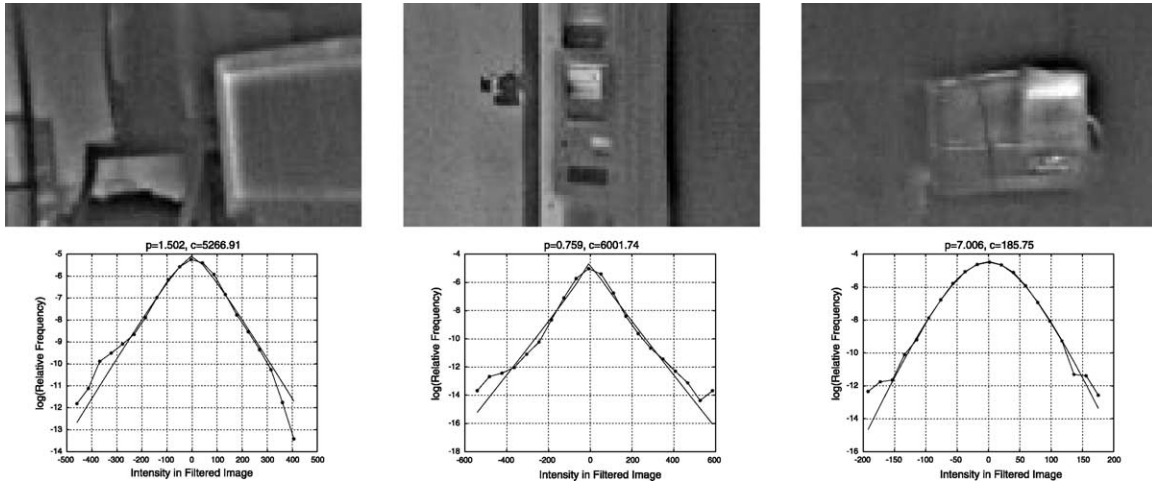
Fig. 7. Estimated Bessel K forms model the wavelet coefficients of background clutter images.

I. For a suitably large number of filters, the set of Bessel K parameters can characterize the contents of *I* to a reasonable extent.

3. Pseudo-metrics for comparing IR images

One distinct advantage of having analytical (parametric) probability models, is the resulting theoretical framework for image analysis. For instance, we would like to be able to compare IR images by directly comparing their respective Bessel parameters without ever computing the histograms.

A closed form expression of a metric involving only the parameters can be very efficient.

To quantify the difference between two Bessel K forms, we have chosen the L^2-metric. Other metrics, such as the Kullback-Leibler divergence, Renyi divergence, earth mover's metric, or the L^1 metric, can also be used. Although L^2 is a common choice for search/optimization problems, our choice is purely based on the convenience of a simple expression. The main drawback of this choice is that the Bessel K forms are not in L^2 for $p < 0.25$ and therefore the metric is not applicable to those cases. In cases where the estimated p values fall below 0.25, we map it back to L^2 by replacing its p value by $0.25 + \epsilon$ for a small $\epsilon > 0$.
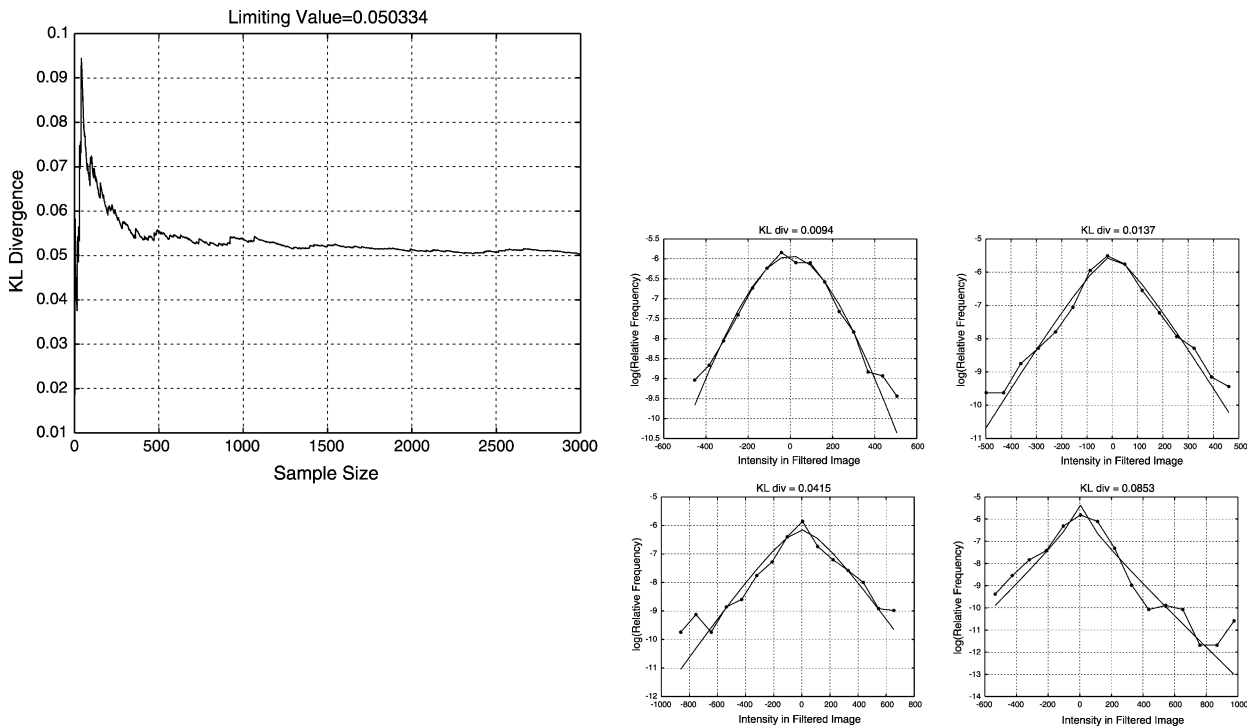

Fig. 8. Left panel: convergence of Monte Carlo averaging of KL-divergence as the sample size increases. Right panel: examples of KL-divergence between estimated and observed densities. Left to right the values are: 0.0004, 0.0137, 0.0415, and 0.0853, respectively.
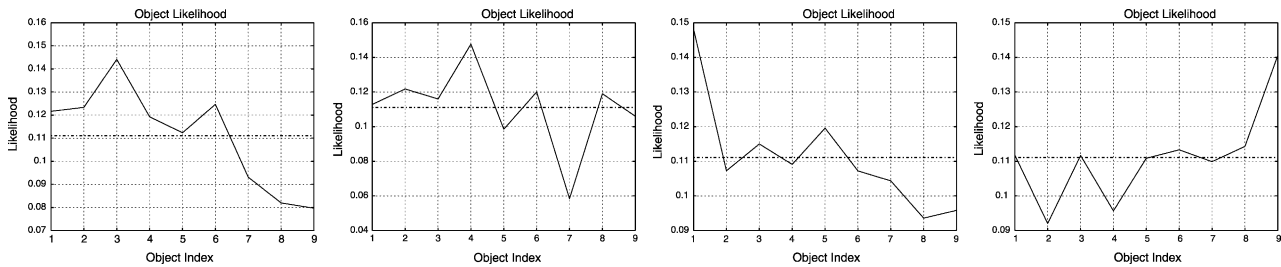
Fig. 9. Example images from FSU IR face database.

Fig. 10. Plots of $\tilde{P}(\alpha|I)$ computed for images drawn from the test set. For thresholding at the uniform probability, only the subjects above the dotted line are shortlisted.

For any two Bessel K forms $f(x; p_1, c_1)$ and $f(x; p_2, c_2)$, the L^2-metric can be defined as: $d(p_1, c_1, p_2, c_2) = \sqrt{\int_{\mathbb{R}} (f(x; p_1, c_1) - f(x; p_2, c_2))^2 \, dx}$. Here we have assumed $\mu_1 = \mu_2 = 0$ as is often the case for the filters used in this paper. This metric can be computed in a closed form, under certain restrictive conditions, as follows. In Ref. [27], the L^2-distance between the two Bessel K forms, parameterized by (p_1, c_1) and (p_2, c_2), respectively, has been shown to be: for $p_1, p_2 > 0.25, c_1, c_2 > 0$,

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

(2)

where $\mathscr{G}(p) = \Gamma(p - 0.5)/\Gamma(p)$ and $\mathscr{F} = F((p_1 + p_2 - 0.5), p_2; p_1 + p_2; 1 - \frac{c_1}{c_2})$ (F is the hypergeometric function).

This result provides a metric to compare two Bessel K forms, or two wavelet marginals. It can be extended to a pseudo-metric on the image space as follows. For any two IR images, I_1 and I_2, and the filters $F^{(1)}, ..., F^{(J)}$, let the corresponding parameter values be given by: $(p_1^{(j)}, c_1^{(j)})$ and $(p_2^{(j)}, c_2^{(j)})$, respectively, for $j = 1, 2, ..., J$. Then, the L^2-distance between wavelet representations of the two images is defined as:

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

It may be useful to choose instead:

$$d_I(I_1, I_2) = \max_j d(p_1^{(j)}, c_1^{(j)}, p_2^{(j)}, c_2^{(j)}).$$

In either case, the value of $d_I(\cdot, \cdot)$ depends upon the choice of filters used in representing the images. Also, $d_I(I_1, I_2)$ is not a proper metric on the image space since it is possible to have two different images with d_I equal to zero. For any two images that have wavelet coefficients with identical marginals, the resulting distance will become zero.

4. Pruning for IR face recognition

Given an image I, we seek to prune the elements of $\mathscr{A}$ using wavelet decomposition of I and the resulting Bessel K

Table 1
Recognition performance of eigen faces, independent faces, and Bessel K forms

Test/training ratio	Eigen faces (%)	Independent faces (%)	Bessel forms (%)
Correct to be the closest			
1:1	90.48	89.52	97.14
3:1	87.42	88.05	91.20
7:1	80.21	70.05	83.96
Correct within the closest two			
1:1	94.29	96.19	99.05
3:1	92.45	91.20	96.23
7:1	87.70	83.42	91.98

Fig. 11. Sample images from Equinox database.

representations. We assume the training data to be images of each of the subject $\alpha \in \mathcal{A}$ at some sampled values of nuisance variables (pose, facial expressions, thermal state, etc.). Let the sample set be denoted by $S_0 \subset S$; the training set consists of $|\mathcal{A}| \times |S_0|$ number of images. Select a bank of filters $F^{(1)}, F^{(2)}, ..., F^{(J)}$ that are found useful for face recognition application. For each of the training image and each filter in the filter bank, we can pre-compute the Bessel parameters $\{(p_{\alpha,s}^{(j)}, c_{\alpha,s}^{(j)}) : j = 1, 2, ..., J, \alpha \in \mathcal{A}, s \in S_0\}$ as part of the training procedure. The training algorithm is as follows.

Algorithm 1. Training algorithm

1. For each $\alpha \in \mathcal{A}, s \in S_0$, and each filter in the filter bank, compute the filtered image $I^{(j)} = I(\alpha, s) * F^{(j)}$.
2. Estimate the Bessel parameters $\{p_{\alpha,s}^{(j)}, c_{\alpha,s}^{(j)} : j = 1, 2, ..., J\}$ using Eq. (1).

For a test image I, let its Bessel representation, under the same bank of filters, be given by $\{(p_{\text{obs}}^{(j)}, c_{\text{obs}}^{(j)}) : j = 1, 2, ..., J\}$. Define a probability mass function on $\mathcal{A}$ according to:

$$\tilde{P}(\alpha|I) = \frac{1}{\tilde{Z}} \exp\left(-\min_{s \in S_0}\left(\sum_{j=1}^{J} d(p_{\text{obs}}^{(j)}, c_{\text{obs}}^{(j)}, p_{\alpha,s}^{(j)}, c_{\alpha,s}^{(j)})^2\right)/T\right) \quad (3)$$

where $\tilde{Z}$ is the normalizer and T controls our confidence (analogous to the temperature in simulated annealing) in this probability. Use of min operator implies a choice of nearest neighbor distance in defining $\tilde{P}$.

Algorithm 2. Test algorithm

1. Compute $I^{(j)} = I * F^{(j)}$, and compute the Bessel parameters $p_{\text{obs}}^{(j)}, c_{\text{obs}}^{(j)}$, for all j, using Eq. (1).
2. Using Eq. (2), compute the metric $d(p_{\text{obs}}^{(j)}, c_{\text{obs}}^{(j)}, p_{\alpha,s}^{(j)}, c_{\alpha,s}^{(j)})$ for all $\alpha \in \mathcal{A}$ and $s \in S_0$. Compute the probability mass function specified in Eq. (3).
3. For a threshold $0 < \nu < 1$, define the set of pruned hypotheses: $\mathcal{A}_0 = \{\alpha \in \mathcal{A} : \tilde{P}(\alpha|I) \geq \nu\}$.

To illustrate this idea, we describe some experimental results. We have divided the face database into non-overlapping training and test sets. Some of the images are used as training and the remaining are used for

testing. We have used $J = 39$ filters, including the gradient filters, the Laplacian of Gaussian filters, and the Gabor filters (Fig. 9).

For the FSU database, Fig. 10 shows the plots of $\tilde{P}(\alpha|I)$ versus α (at $T = 0.3$) for six different images I in the test set. All the subjects that have $\tilde{P}(\alpha|I)$ larger than some threshold, say the uniform probability $\nu = 1/|\mathcal{A}|$, are shortlisted to form $\mathcal{A}_0$. For instance, the first plot is for an image I of the third subject ($\alpha_{\text{true}} = 3$) and the third plot corresponds to $\alpha_{\text{true}} = 1$. For the latter case, a shortlisting by thresholding leaves only three possible hypotheses, as compared to the original nine. Choice of ν reflects a balance between the desired amount of pruning and the probability of incorrect pruning.

To illustrate the strength of these representations, we have also used them for face identification using the nearest-neighbor criterion. For $J = 39$ filters, each training image is represented by 78 parameters. For comparison, we have also computed identification performance using eigen- and independent-component analysis. For eigen and independent faces, we first compute the eigen/independent basis using the training faces and then we project each training image on the eigen/independent basis. The difference between two images is quantified by the Euclidean distance between their projections. To study identification performance under different conditions we vary the number of training faces per subject. For both the eigen and independent representations, we have used 78 coefficients, wherever possible, in order to make the representation complexity comparable to that of Bessel K forms. We have obtained identification results for two databases: FSU database and a subset of Equinox database.

1. *FSU face database.* Table 1 shows the identification performance of the three methods for the FSU face

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

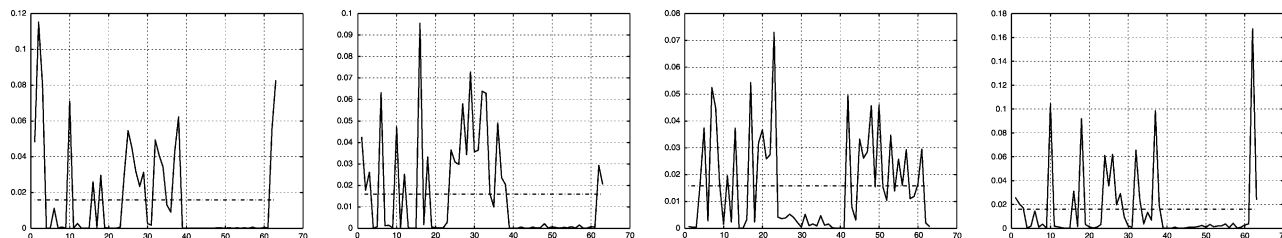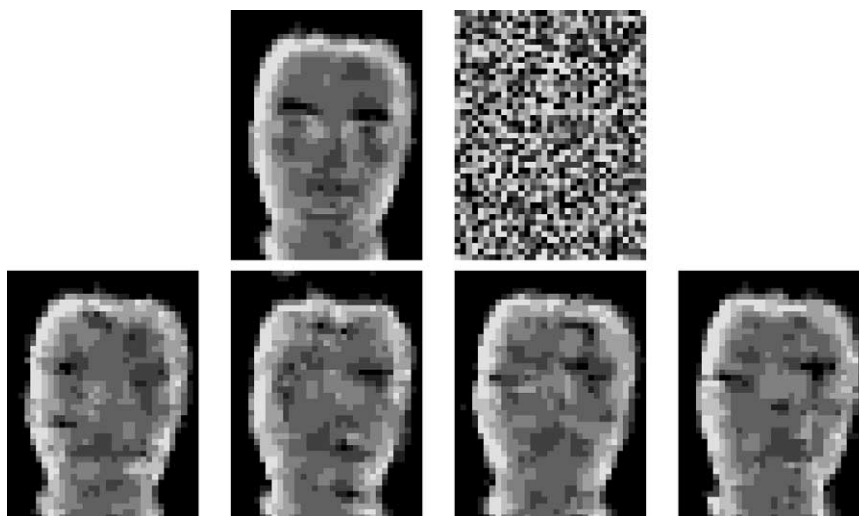| Training/test images	First correct (%)	First two correct (%)
1948/1945	99.95	100.0
993/2900	99.97	100.0
527/3366	99.97	100.0
343/3550	99.94	99.97

Fig. 12. Some example plots of $\tilde{P}$ for the Equinox dataset.

Fig. 13. The top panel shows the input image (left) and the initial white noise image (right). Lower panels show synthesized face images obtained by matching marginal distributions of filtered images.

database. Bessel K forms have the best identification rate among the three methods under all the conditions. This is because that Bessel K forms capture the perceptual characteristics of IR faces through a number of filters and they are less sensitive to deformations due to the filtering stage.

2. *Equinox IR face database*. The Bessel K form has also been applied to a large IR face dataset generated by Equinox.[1] We have used a subset consisting of 63 subjects with a total of 3893 long-wave infrared (LWIR) images of faces. Fig. 11 shows some sample pictures from that dataset. Recognition results are presented in Table 2. High level of identification performance can be attributed to the fact that the images used here were all reasonably well aligned. The variation in head pose and placement were relatively less as compared to the FSU database. Plots of $\tilde{P}$ (at $T = 0.05$) are shown in Fig. 12.

5. Image synthesis using Bessel K forms

An important question in using univariate densities of the wavelet coefficients for face identification is: Can these marginals impose sufficient constraints on a global pattern like a face where the geometric structures are considered

important for identification? We believe the answer is yes, and we demonstrate that by carrying out a face synthesis experiment by matching the marginal distributions of filtered images using a Gibbs sampler. More specifically, given an input image (such as the one shown in the top left panel of Fig. 13), we calculate its wavelet marginals using a bank of $J = 79$ filters. Denote the marginal densities by: $f(x; p_1, c_1), f(x; p_2, c_2), ..., f(x; p_J, c_J)$. The next step is to generate a random image I that has the same marginal densities under the similiar bank of filters. Using a Gibbs sampler for random sampling, we start with a white noise image (top right panel) as the initial condition. Additionally, to avoid local minima in the sampling process, we assume that face images provide boundary conditions for sampling. The bottom panels show some samples from the Gibbs sampler. These examples suggest that important geometric features, such as the global shape of the face and eyes, are captured by the marginal distributions. This also justifies why Bessel K representation yields a good performance for statistical hypothesis pruning.

6. Conclusion

We have applied an analytical probability model for wavelet coefficients, presented in Refs. [11,27] to the task of

---

[1] Available at http://www.equinoxsensors.com/products/HID.html.

statistical hypothesis pruning in IR face identification. Each IR face image is decomposed into its wavelet coefficients and the marginal densities of the coefficients are modeled as three-parameter Bessel K forms. An IR image can now be represented by the set of Bessel parameters denoting a significant reduction in representation. To quantify image differences, we have derived an L^2-metric on the space of Bessel K forms and have demonstrated its application to IR face identification. We have also compared the effectiveness of Bessel K forms with principal and independent component representations for face identification (Algorithm).

Acknowledgements

References

[1] P.N. Belhumeur, J.P. Hepanha, D.J. Kriegman, Eigenfaces vs. fisherfaces: recognition using class specific linear projection, IEEE Transactions on Pattern Analysis and Machine Intelligence 19 (7) (1997) 711–720.

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

[3] R. Chellappa, C.L. Wilson, S. Sirohey, Human and machine recognition of face: a survey, Proceedings of the IEEE 83 (5) (1995) 705–740.

[4] P. Comon, Independent component analysis, a new concept?, Signal Processing, Special Issue on Higher-order Statistics 36 (3) (1994).

[5] M.L. Cooper, U. Grenander, M.I. Miller, A. Srivastava, Accommodating geometric and thermodynamic variability for forward-looking infrared sensors, Proceedings of the SPIE, Algorithms for Synthetic Aperture Radar IV, vol. 3070, 1997.

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

[7] D.L. Donoho, A.G. Flesia, Can recent innovations in harmonic analysis explain key findings in natural image statistics, Network: Computation in Neural Systems 12 (3) (2001) 317–393.

[8] D.J. Field, Relations between the statistics of natural images and the response properties of cortical cells, Journal of Optical Society of America 4 (12) (1987) 2379–2394.

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

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

[11] U. Grenander, A. Srivastava, Probability models for clutter in natural images, IEEE Transactions on Pattern Analysis and Machine Intelligence 23 (4) (2001) 424–429.

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

[13] P.W. Hallinan, G.G. Gordon, A.L. Yuille, P. Giblin, D. Mumford, Two- and Three-dimensional Patterns of Face, A.K. Peters, 1999.

[14] D.J. Heeger, J.R. Bergen, Pyramid-based texture analysis/synthesis, Proceedings of the SIGGRAPHS, 1995, pp. 229–238.

[15] C. Hesher, A. Srivastava, G. Erlebacher, PCA of range images for facial recognition, Proceedings of the 2002 International Conference on Imaging Science, Systems, and Technology, 2002.

[16] B. Jahne, H. Haubecker, P. Geibler, Handbook of Computer Vision and Applications, vol. 2, Academic Press, New York, 1999.

[17] M. Kirby, L. Sirovich, Application of the karhunen-loeve procedure for the characterization of human faces, IEEE Transactions on Pattern Analysis and Machine Intelligence 12 (1) (1990) 103–108.

[18] A. Lanterman, M. Miller, D. Snyder, General Metropolis-Hastings jump diffusions for automatic target recognition in infrared scenes, Optical Engineering 36 (4) (1997) 1123–1137.

[19] X. Liu, D. Wang, Texture classification using spectral histograms, IEEE Transactions on Image Processing, for in forthcoming 2003.

[20] S.G. Mallat, A theory for multiresolution signal decomposition: the wavelet representation, IEEE Transactions on Pattern Analysis and Machine Intelligence 11 (1989) 674–693.

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

[22] I. Pavlidis, J. Levine, P. Baukol, Thermal imaging for anxiety detection, Proceedings of the IEEE Workshop on Computer Vision Beyond the Visible Spectrum, Hilton Head, 2000, pp. 104–109.

[23] A. Pentland, Looking at people: sensing for ubiquities and wearable computing, IEEE Transactions on Pattern Analysis and Machine Intelligence 22 (1) (2000) 107–119.

[24] F. Prokoski, History, current status, and future of infrared identification, Proceedings of the IEEE Workshop on computer Vision Beyond the Visible Spectrum, Hilton Head, 2000.

[25] A. Selinger, D.A. Socolinsky, Appearance-based facial recognition using visible and thermal imagery: a comparative study, www.equinoxsensors.com 2001.

[26] A. Srivastava, A.B. Lee, E.P. Simoncelli, S.-C. Zhu, On advances in statistical modeling of natural images, Journal of Mathematical Imaging and Vision 18 (2003) 17–33.

[27] A. Srivastava, X. Liu, U. Grenander, Universal analytical forms for modeling image probability, IEEE Transactions on Pattern Analysis and Machine Intelligence 28 (9) (2002) 1200–1214.

[28] A. Srivastava, B. Thomasson, S.R.F. Sims, A regression model for IR image prediction, Proceedings of the Aerosense: ATR XI 4379 (2001) 176–186.

[29] L.B. Wolff, D.A. Socolinsky, C.K. Eveland, Quantitative measurement of illumination invariance for face recognition using thermal infrared imagery, CVPR workshop on Computer Vision Beyond Visual Spectrum, 2001.

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