Image Modelling by Linear Composition of Shared Shiftable Bases

Anonymous CVPR submission

Paper ID WZ

Abstract

This article proposes a statistical model for image patches of object shapes, in the form of additive composition of a set of linear bases selected from a large dictionary, such as Gabor wavelets at different locations, orientations and scales. The model has the following three features in terms of the selected bases. Sparsity: only a small number of salient bases should be selected to represent any image patch with small error. Commonality: the selected bases for image patches of the same object category should be as common as possible, so that these image patches can be modelled by shared bases. Shiftability: the shared bases are allowed to locally shift their locations, orientations and scales within limited range to represent each individual image patch of an object category. This model can be used to select the bases and learn their compositions for a training sample of image patches. The computation can be accomplished using what we call the shared sketch by shiftable bases. We show several experiments to illustrate the model and the algorithm.

1. Introduction

1.1. Motivation and foundation

Pattern theory as advocated by Grenander [2] and Mumford [5] postulates a generative model in the form of $p(W \mid \Theta)$ and $p(\mathbf{I} \mid W, \Theta)$, where $\mathbf{I}$ is the image data, W is the interpretation of $\mathbf{I}$ in terms of what is where, and Θ denotes parameters (including structural parameters) in the model. Θ can be learned from training images (with or without the corresponding W). With the learned Θ, computing W for a given image $\mathbf{I}$ can be guided by $p(W \mid \mathbf{I}, \Theta)$, which tells us what value of W gives the most plausible explanation of $\mathbf{I}$.

A number of generative models for various vision tasks have been proposed in the literature. However, unlike the HMM in speech recognition, there has not been a generic modelling scheme that can serve as the common foundation for different vision tasks. Instead, it is often preferred to target $p(W \mid \mathbf{I}, \Theta)$ directly in supervised training, without modelling $p(W \mid \Theta)$ and $p(\mathbf{I} \mid W, \Theta)$ explicitly, especially when W is simple, such as $W \in \{\text{face, non-face}\}$ in face detection.

While methods targeting $p(W \mid \mathbf{I}, \Theta)$ directly have had considerable successes, it is still desirable to have an image model with $p(\mathbf{I} \mid W, \Theta)$. Even though such a generative model may not synthesize realistic $\mathbf{I}$ or provide efficient compression of $\mathbf{I}$, it can still provide a context for explaining the image data so that the *explain-away competition* can be carried out to select the most plausible interpretation of a single image, or to discover the most plausible interpretation of a set of images for unsupervised learning.

In our opinion, the image model that comes the closest to being the foundation for further developments is the sparse coding model of Olshausen and Field [6], which is a simple linear additive model that is built directly on the raw image intensities. The model is of the form $\mathbf{I} = \sum_i c_i \Gamma_i + \epsilon$, where $\{\Gamma_i, i = 1, ..., N\}$ is a dictionary of linear bases, c_i are their coefficients, and ϵ is the error. The size of the dictionary N can be many folds larger than the dimensionality of $\mathbf{I}$. The key principle is *sparsity*. That is, for each typical natural image $\mathbf{I}$, only a small number of c_i should be significantly different from 0, or in other words, only a small number of bases should be selected from $\{\Gamma_i\}$, in order to represent $\mathbf{I}$ with small error ϵ. Formally, one can express the sparsity principle by a regularity function of $\{c_i\}$ such as l_1 norm that encourages sparsity, or by assuming that $\{c_i\}$ follow some probability distribution that concentrates most of its probability mass around 0, but has heavy tails to account for occasionally large values [3]. Olshausen and Field [6] were able to learn a dictionary of localized, elongate and orientated linear bases from natural image patches using this model. These bases resemble Gabor wavelets. Given a dictionary of linear bases, the matching pursuit algorithm of Mallat and Zhang [4] can be used to sequentially select a small number of bases for representing an input image.

Sparsity principle is important for modelling purpose, because it is easier to model low-dimensional structures than high-dimensional ones. However, sparsity alone is clearly inadequate for modelling image patches of different object categories. In this article, we add two more features

to the linear sparse coding model, and develop a statistical model for image patches of objects.

1.2. Commonality and shiftability

The two additional features are commonality and shiftability.

1) *Commonality*. For image patches of each category, we want the bases selected for these image patches to be as common as possible, so that these image patches can be modelled by shared bases. In terms of probability modelling, we want each selected base to have a very high probability to be turned on. Of course, for different object categories, it is preferred that they do not share many common bases. This feature was inspired by the work of Viola and Jones [8] on adaboost method for classification. The weak classifiers selected by their method are in the form of projecting the image patches onto Harr bases and thresholding the projection coefficients. The selected Harr bases are used to characterize all the images, and they are selected to maximally tell the face images apart from non-face images. Our method can be consider a generative version of this scheme, where the shared bases are selected to maximally explain all the input image patches. Our method can be easily extended to modelling multiple object categories in either supervised or unsupervised learning, where each category has its own shared set of common bases.

2) *Shiftability*. Deformation is common in object shapes, even within the same object category and the same pose. The selected bases at fixed locations, orientations and scales may not give optimal representations to all the image patches, even they are reasonably well aligned. To account for deformation, we must allow the shared bases to shift their locations, orientations and scales within limited range when representing each individual image patch. This feature was inspired by the work of Riesenhuber and T. Poggio [7] on HMAX model, which is a hierarchical bottom-up model for object recognition. In their model, the complex cells in V1 are assumed to compute the local maxima of Gabor filter responses relative to the shift of locations and scales. Such a maximum pooling mechanism makes the responses of the complex cells invariant to limited deformation as well as changes in scale and pose etc. In our method, this maximum pooling is incorporated in our shared sketch algorithm for fitting multiple training images simultaneously, where the shared bases can shift to the maximally tuned locations, orientations and scales in representing each individual image patch.

Figure 1 illustrates commonality and shiftability. There are three 82×164 training image patches of cars. The dictionary of linear bases are Gabor wavelets at 12 different orientations with a fixed scale. These Gabor wavelets can be centered at any pixel within the domain of the image patches. In this figure, each Gabor wavelet is represented symbolically by a bar of the same position, orientation and length. In the first row, the left figure displays all the 61 selected bases. The right figure displays the bases that are shared by all the three images. The selected Gabor wavelets have little overlaps between them and they are well connected, so the symbolic representation is essentially a "sketch" or a line-drawing of the input images. The right figure in the first row can be considered the common sketch "averaged" over all the three training images. Clearly, most of the selected bases displayed in the left figure of the first row are common to all the three images.

For the rest three rows, the left figure displays the input training image, and the right figure displays the bases that are actually used to represent the input image on the left. Most of these bases are shifted versions of the shared bases displayed in the right figure of the first row. Clearly, these shared bases can shift their locations and orientations to represent each individual image.

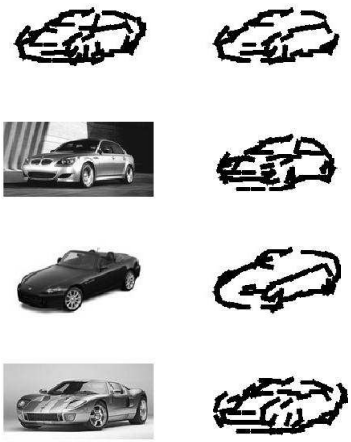

Figure 1. Top row: the left figure displays all the 61 selected bases, the right figure displays the bases shared by all the three images. 2nd to 4th rows: the left figure displays the 82×164 training image, and the right displays the bases used for representing the image on the left. A Gabor wavelet is represented symbolically by a bar at the same location, with the same orientation and length.

It is worth noting that the first training image has strong edges in the background. However, these edges are not shared by the other two images. So no Gabor wavelets are selected to represent them. Also, the car in the second image has a moderately different pose than the other two cars. But it can still be represented by the shared bases after shifting. Figure 2 shows another example for three training images of horses.

The rest of the article is organized as follows. Section 2 gives the technical details of the model and the algorithm. Section 3 describes a number of experiments. Section 4 concludes with a brief discussion.

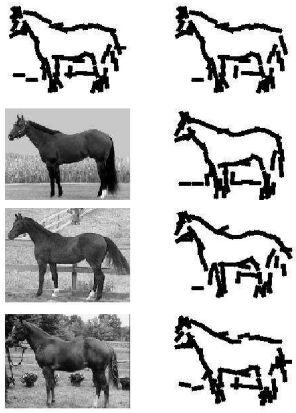

Figure 2. The size of the training images is 118×157. 67 bases are selected in total. See the caption of Figure 1 for explanation.

2. Model and Algorithm

2.1. Gabor bases: edges and spectrum

A Gabor function [1] is of the following form:

$$G(x) = \frac{1}{2\pi\sigma_1\sigma_2} \exp\{-\frac{1}{2}(\frac{x_1^2}{\sigma_1^2} + \frac{x_2^2}{\sigma_2^2})\}e^{ix_1}, \quad (1)$$

where $x = (x_1, x_2) \in \mathbb{R}^2$, and $i = \sqrt{-1}$. We can translate, rotate, and dilate the function $G(x)$ of (1) to obtain a general form of Gabor wavelets: $G_{y,s,\theta}(x) = G(\tilde{x}/s)/s^2$, where $\tilde{x} = (\tilde{x}_1, \tilde{x}_2)$, $\tilde{x}_1 = (x_1 - y_1)\cos\theta - (x_2 - y_2)\sin\theta$, $\tilde{x}_2 = (x_1 - y_1)\sin\theta + (x_2 - y_2)\cos\theta$. The central frequency of $G_{y,s,\theta}$ is $(\cos\theta/s, \sin\theta/s)$.

For an image $\mathbf{I}(x)$, the projection coefficient of $\mathbf{I}$ onto $G_{y,s,\theta}$ or the filter response is $r_{y,s,\theta} = \langle \mathbf{I}, G_{y,s,\theta}\rangle = \int \mathbf{I}(x)G_{y,s,\theta}(x)dx$. The local energy $|r_{y,s,\theta}|^2$ is large if there is an edge or bar structure at y with scale s and orientation θ. The marginal average of local energy $\mathrm{E}[|r_{y,s,\theta}|^2]$ within an image region estimates the average power spectrum around the central frequency of $G_{y,s,\theta}$.

2.2. Model: composing shared and shiftable bases

To model a sample of training images $\{\mathbf{I}_m, m = 1, ..., M\}$, we can select a small set of bases $\{B_k, k = 1, ..., K\}$ from a large dictionary $\{\Gamma_i, i = 1, ..., N\}$, such as Gabor bases at different locations, orientations and scales, $K \ll N$. In order to represent each image $\mathbf{I}_m$ using the selected bases $\{B_k\}$, we need to define two sets of variables to account for commonality and shiftability.

Representing commonality: $\alpha_m = (\alpha_{m,k}, k = 1, ..., K)$. $\alpha_{m,k} = 1$, if B_k is used to represent $\mathbf{I}_m$, i.e., B_k is turned on for $\mathbf{I}_m$. $\alpha_{m,k} = 0$, if B_k is not used to represent $\mathbf{I}_m$, i.e., B_k is turned off for $\mathbf{I}_m$.

Representing shiftability: $\delta_m = (\delta_{m,k}, k = 1, ..., K)$. $\delta_{m,k}$ is defined only if $\alpha_{m,k} = 1$, and it denotes the shifting of B_k in location, orientation and scale in representing $\mathbf{I}_m$. As to the range of $\delta_{m,k}$, the base B_k can shift its location along its normal direction within a range of a small number of pixels, and for each shifted location, the base can also shift its orientation within a small range of angles. We may also allow the base to change its scale. For notational convenience, we may simply denote $B_{k+\delta_{m,k}}$ as the base that is actually used to represent $\mathbf{I}_m$.

Representing linear composition: given $\{B_k\}$ and $\{\alpha_m, \delta_m\}$,

$$\mathbf{I}_m = \sum_{\alpha_{m,k}=1} c_{k,m} B_{k+\delta_{m,k}} + \epsilon. \quad (2)$$

Let $c_m = (c_{m,k}, k = 1, ..., K)$. In order to model $\{\mathbf{I}_m\}$, we need to select $\{B_k\}$, and model $\{\alpha_m, \delta_m, c_m\}$.

Let's start from the simplest possible model, and then generalize it later on.

Modelling commonality: $\alpha_{m,k} \sim$ Bernoulli (a) independently across $k = 1, ..., K$, where a is the probability that B_k is turned on for $\mathbf{I}_m$. a can be close to 1.

Modelling shiftability: $[\delta_{m,k}|\alpha_{m,k} = 1]$ follows a uniform distribution over Δ. Δ denotes the allowed range of shift.

Modelling orthogonal composition: We assume that the bases $\{B_{k+\delta_{m,k}}, \alpha_{m,k} = 1\}$ that are used to represent $\mathbf{I}_m$ have little overlap in spatial domain, or if they do overlap in spatial domain, they have little overlap in frequency domain, i.e., in orientation and scale. Thus they have little correlations between themselves, and we may assume that these bases are orthogonal to each other.

For clarity, we use matrix notation in what follows. Suppose $\mathbf{I}_m$ is defined on a lattice D with $|D|$ pixels. We can vectorize $\mathbf{I}_m$ as a $|D| \times 1$ vector. For every Gabor basis Γ_i in the dictionary $\{\Gamma_i, i = 1, ..., N\}$, we can vecotrize it according to the same order of vectorization. We normalize every Γ_i to have unit l_2 norm, so $\|\Gamma_i\|^2 = 1$. We normalize each $\mathbf{I}_m$ to have unit marginal variance, so $\|\mathbf{I}_m\|^2 = |D|$. This is a sensible thing to do to filter out the effect of overall lighting variation.

Let $\mathbf{B}_m = (B_{k+\delta_{m,k}}, \alpha_{m,k} = 1)$, i.e., $\mathbf{B}_m$ is a $|D| \times K_m$ matrix whose columns are $B_{k+\delta_{m,k}}$ with $\alpha_{m,k} = 1$, where $K_m = \sum_k \alpha_{m,k}$. Note that $\mathbf{B}_m$ is completely determined by α_m, δ_m, and vice versa. Let $R_m = \mathbf{B}_m'\mathbf{I}_m$ be the $K_m \times 1$ vector of projection coefficients or filter responses of $\mathbf{I}_m$ on the bases $B_{k+\delta_{m,k}}$ that are used to represent $\mathbf{I}_m$.

Let $\bar{\mathbf{B}}_m$ be an $|D| \times (|D| - K_m)$ matrix whose columns are orthonormal and also orthogonal to all the column of $\mathbf{B}_m$. The columns of $\bar{\mathbf{B}}_m$ are the residual dimensions, and $\bar{R}_m = \bar{\mathbf{B}}_m'\mathbf{I}_m$ are the residual error. $\|\bar{R}_m\|^2 = |D| - \|R_m\|^2$.

We model the components of R_m to have independent uniform distributions over $[r_0, r_0 + b]$, where r_0 is a threshold for the bases to be turned on. Given R_m, $\bar{R}_m$ have uni-

form distribution over the sphere $\Omega = \{\bar{R}_m : \|\bar{R}_m\|^2 = |D| - \|R_m\|^2\}$, so $p(\bar{R}_m \mid R_m) = 1/|\Omega|$. According to the equipartition principle in information theory, if $|D| - K_m$ is large, this uniform distribution is equivalent to assuming that the components of $\bar{R}_m$ follows $\mathrm{N}(0, \sigma_m^2)$ independently, where $\sigma_m^2 = (|D| - \|R'_m\|^2)/(|D| - K_m)$.

The distribution $p(\mathbf{I}_m \mid \mathbf{B}_m)d\mathbf{I}_m = p(R_m)p(\bar{R}_m \mid R_m)dR_m d\bar{R}_m$. The dimensions are matched, and the Jacobian is 1 because of the orthogonality. Moreover,

$$\log p(\bar{R}_m \mid R_m) = -\frac{1}{2}(|D| - K_m)\log(2\pi e\sigma_m^2).$$

If $|D|$ is much large than both K_m and $\|R_m\|^2$, which is the case in object recognition where $\mathbf{I}_m$ can be a large image, and $\mathbf{B}_m$ only models a small patch of it, against a large background in $\mathbf{I}_m$ of unit marginal variance, then

$$-(|D| - K_m)\log(\sigma_m^2) \approx \|R_m\|^2.$$

Likelihood function: Let $\mathbf{B} = (B_k, k = 1, ..., K)$ denote all the selected bases. Assuming the above approximation to be exact, then the joint distribution is:

$$p(\mathbf{I}_m, \mathbf{B}_m \mid \mathbf{B}) = \frac{1}{2}\|R_m\|^2 + \lambda K_m,$$

where

$$\lambda = \log(a/(1-a)) - [\log(|\Delta|b) - \log(2\pi e)/2],$$

where the first term is the award for commonality, and the second term is the cost for coding the shift and coefficient of a selected base versus leaving it to residual dimensions.

$p(\mathbf{I}_m \mid \mathbf{B})$ can be obtained by integrating out α_m and δ_m. Using $\sum_m \log p(\mathbf{I}_m \mid \mathbf{B})$ as the likelihood, we can learn $\mathbf{B}$ from a sample of training images $\{\mathbf{I}_m, m = 1, ..., M\}$ of a certain object category. After learning $\mathbf{B}$, we can use $p(\mathbf{I} \mid \mathbf{B})$ to get the likelihood that a testing image $\mathbf{I}$ belongs to the same category.

Usually, $p(\mathbf{I}_m, \mathbf{B}_m \mid \mathbf{B})$ is highly peaked, so that we can estimate α_m and δ_m accurately at their posterior modes that maximizes $p(\mathbf{I}_m, \mathbf{B}_m \mid \mathbf{B})$. So $p(\mathbf{I}_m \mid \mathbf{B})$ can be approximated by $p(\mathbf{I}_m, \mathbf{B}_m \mid \mathbf{B})$ by plugging in estimated $\mathbf{B}_m$.

2.3. Algorithm: shared sketch by shiftable bases

We develop a shared sketch algorithm for selecting bases $\mathbf{B}$ and their shifted versions $\{\mathbf{B}_m\}$. Similar to the matching pursuit algorithm of Mallat and Zhang [4], the shared sketch algorithm is a greedy one. At each step, it selects a base from the dictionary, and attempt its shifted versions on all the input images simultaneously. If a shifted version is used to represent an image, then this shifted version inhibits all the other overlapping bases from representing the same image. For two bases Γ_i and Γ_j, they are not overlapping if their correlation $\langle \Gamma_i, \Gamma_j \rangle$ is below a predefined threshold (e.g., .01).

The following is a detailed description of the algorithm. We use the notation i to label the bases in the dictionary $\{\Gamma_i, i = 1, ..., N\}$, and we use k to label the selected bases $\{B_k, k = 1, ..., K\}$. i and k run through two distinct sets.

Step 0: *Initialization and thresholding*. For $i = 1$ to N, for $m = 1$ to M, compute $r_{m,i} = \langle \mathbf{I}_m, \Gamma_i \rangle$. If $|r_{m,i}| < r_0$, set $r_{m,i} = 0$. Let $k = 1$.

Step 1: *Attempt shared shiftable fitting for all candidate bases*. For $i = 1$ to N, for $m = 1$ to M, do the following. For $\delta_{m,i} \in \Delta$, if all $r_{m,i+\delta_{m,i}} = 0$, set $\alpha_{m,i} = 0$. Otherwise, let $\hat{\delta}_{m,i}$ be the one with the maximum $r_{m,i+\delta_{m,i}}$, and set $\alpha_{m,i} = 1$.

Step 2: *Select the best fitting base for shared sketch*. For $i = 1$ to N, compute

$$L_i = \frac{1}{M}\sum_{m:\alpha_{m,i}=1}\left[\frac{1}{2}r^2_{m,i+\hat{\delta}_{m,i}} + \lambda\right].$$

Let j be the base such that $L_j \geq L_i$ for $i = 1, ..., N$. Then let $B_k = \Gamma_j$. For $m = 1$ to M, if $\alpha_{m,j} = 1$, let $\alpha_{m,k} = 1$, let $\delta_{m,k} = \hat{\delta}_{m,j}$, and let $r_{m,k} = r_{m,j+\hat{\delta}_{m,j}}$ (note that k and j belong to two distinct sets).

Step 3: *Inhibit overlapping bases*. For $m = 1$ to M, if $\alpha_{m,k} = 1$, do the following. For $i = 1$ to N, if Γ_i overlaps with $B_{k+\delta_{m,k}} = \Gamma_{m,j+\hat{\delta}_{m,j}}$, set $r_{m,i} = 0$.

Step 4: *Stopping criterion*. If L_j is below a pre-defined threshold, then let $K = k$, stop. Otherwise, let $k = k + 1$, and go back to Step 1.

The threshold on L_j corresponds to a prior distribution on K, which prefers small value.

2.4. Nonorthogonality and scale-specific sketch

Before going to experimental results, we would like to explain two important issues.

Nonorthogonality: For nonorthogonal $\mathbf{B}_m$, let $R_m = \mathbf{B}'_m\mathbf{I}_m$ be the $K_m \times 1$ vector of projection coefficients. The projection of $\mathbf{I}_m$ on the subspace spanned by the bases in $\mathbf{B}_m$ is $\mathbf{J}_m = \mathbf{B}_m(\mathbf{B}'_m\mathbf{B}_m)^{-1}R_m = \mathbf{B}_m C_m$. $\mathbf{J}_m$ is the reconstructed image with C_m being the $K_m \times 1$ vector of least squares reconstruction coefficients, $C_m = (\mathbf{B}'_m\mathbf{B}_m)^{-1}R_m$. $\|\mathbf{J}_m\|^2 = R'_m(\mathbf{B}'_m\mathbf{B}_m)^{-1}R_m$.

Let $\bar{R}_m = \bar{\mathbf{B}}_m\mathbf{I}_m$. Recall that $\bar{\mathbf{B}}_m$ consists of orthonormal columns that are orthogonal to the bases in $\mathbf{B}$. Given R_m, $\bar{R}_m$ have uniform distribution over the sphere $\Omega = \{\bar{R}_m : \|\bar{R}_m\|^2 = |D| - \|\mathbf{J}_m\|^2\}$, which is equivalent to independent $\mathrm{N}(0, \sigma_m^2)$, with $\sigma_m^2 = (|D| - \|\mathbf{J}'_m\|^2)/(|D| - K_m)$. Then

$$p(\mathbf{I}_m \mid \mathbf{B}_m) = f(R_m)f(\bar{R}_m \mid R_m)|\det(\mathbf{B}'_m\mathbf{B}_m)|^{1/2},$$

where $f(R_m)$ is the distribution of the responses R_m, and

$|\det(\mathbf{B}'_m\mathbf{B})|^{1/2}$ is the Jacobian of the linear change of variable $(R'_m, \bar{R}'_m)' = (\mathbf{B}_m, \bar{\mathbf{B}}_m)'\mathbf{I}_m$. Thus

$$p(\mathbf{I}_m, \mathbf{B}_m \mid \mathbf{B}) = \lambda K_m + \frac{1}{2}\left[R'_m(\mathbf{B}'_m\mathbf{B}_m)^{-1}R_m + \log|\det(\mathbf{B}'_m\mathbf{B}_m)|\right].$$

This can be used to modify the shared sketch algorithm. When attempting to add a new base Γ_i to $\mathbf{B}_m$ in the algorithm, and compute the changes in $R'_m(\mathbf{B}'_m\mathbf{B}_m)^{-1}R_m$ and $\det(\mathbf{B}'_m\mathbf{B}_m)$, we can perform a Gram-Schmidt orthogonalization of Γ_i by $\mathbf{B}_m$. This step can be naturally incorporated into the shared sketch algorithm as a form of soft inhibition that replaces the hard inhibition in the Step 3 of the original algorithm.

Scale-specific sketch: In a given image, different patterns may appear at different scales, and these patterns may be organized into hierarchical whole-part relationships. It is therefore desirable to perform separate sketches at different scale ranges or frequency bands, and then organize them into hierarchical relationships, instead of performing a single sketch using Gabor bases across all the scales. Let F be the range of frequencies covered by the Gabor bases within a relatively narrow range of scales. Then these Gabor bases are trying to fit the band-pass image within the frequency band F. Specifically, let $\mathbf{I}_m = \sum_\omega \hat{\mathbf{I}}_m(\omega)\exp(i\omega x)$, where $\hat{\mathbf{I}}_m$ is the discrete Fourier transform of $\mathbf{I}_m$. Then these Gabor bases are trying to fit $\mathbf{I}_m(F) = \sum_{|\omega|\in F}\hat{\mathbf{I}}_m(\omega)\exp(i\omega x)$, while neglecting all the rest of the frequency components outside F.

Recall that we normalize the input image $\mathbf{I}_m$ to unit marginal variance. That is, for an input $\mathbf{I}_m$, we compute $S^2 = \|\mathbf{I}_m\|^2/|D|$, and then change $\mathbf{I}_m$ to $\mathbf{I}_m/S$. This amounts to normalizing $r_{m,i} = \langle \mathbf{I}_m, \Gamma_i\rangle/S$. Because the band-pass $\mathbf{I}_m(F)$ is the image that is being fitted, it is more reasonable to normalize $\mathbf{I}_m(F)$. Specifically, $S^2(F) = \|\mathbf{I}_m(F)\|^2/|D| = \sum_{|\omega|\in F}|\hat{\mathbf{I}}_m(\omega)|^2/|F|$, where $|F|$ is the number of frequency components in F, and $S^2(F)$ is the average spectrum of $\mathbf{I}_m$ within F. So we should normalize $r_{m,i} = \langle \mathbf{I}_m, \Gamma_i\rangle/S(F)$. The average spectrum $S(F)$ can be estimated by the average of $\{|\langle \mathbf{I}_m, \Gamma_i\rangle|^2\}$ for all those candidates Γ_i within this frequency band F.

In this band-pass setting, the dimensionality of $\mathbf{I}_m(F)$ is $|F|$ with its $|F|$ Fourier components $\hat{\mathbf{I}}_m(\omega)$. After projecting $\mathbf{I}_m$ onto the selected orthogonal bases $\mathbf{B}_m$ to get $R_m = \mathbf{B}'_m\mathbf{I}_m$, there are $|F| - K_m$ dimensions left for $\bar{\mathbf{B}}_m$, and $\bar{R}_m = \bar{\mathbf{B}}'_m\mathbf{I}$ belongs to the sphere $\Omega = \{\bar{R}_m : \|\bar{R}_m\|^2 = |F| - \|R_m\|^2\}$. The argument in the previous subsection still follows through.

3. Experiments

4. Discussion

References

[1] J. Daugman, "Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters," *Journal of Optical Society of America*, 2, 1160-1169, 1985. 3

[2] U. Grenander, *General Pattern Theory*, Oxford Univ Press, 1993. 1

[3] M. S., Lewicki and B. A. Olshausen, "Probabilistic framework for the adaptation and comparison of image codes," *Journal of the Optical Society of America*, 16(7): 1587-1601, 1999. 1

[4] S. Mallat and Z. Zhang, "Matching pursuit in a time-frequency dictionary," *IEEE Transactions on Signal Processing*, 41, 3397-415, 1993. 1, 4

[5] D. B. Mumford, "Pattern theory: a unifying perspective," *Proceedings of 1st European Congress of Mathematics*, Birkhauser-Boston, 1994. 1

[6] B. A. Olshausen and D. J. Field, "Emergence of simple-cell receptive field properties by learning a sparse code for natural images," *Nature*, 381, 607-609, 1996. 1

[7] M. Riesenhuber and T. Poggio, "Hierarchical models of object recognition in cortex," *Nature Neuroscience*, 2 , 1019-1025 (1999). 2

[8] P. A. Viola and M. J. Jones, "Robust real-time face detection," *International Journal of Computer Vision*, 57(2), 137-154 2004. 2