# Duality between Foveatization and Multiscale Local Spectrum Estimation

Rafael Navarro, Javier Portilla

Instituto de Óptica Serrano 121, 28006 Madrid, Spain.

Antonio Tabernero

Facultad de Informática (UPM) 28660 Boadilla del Monte, Madrid, Spain.

## ABSTRACT

In this work we demonstrate the relationship existing between two important issues in vision: multi-scale local spectrum analysis, and log-polar foveatization. We show that, when applying a continuous set of self-similar (rotated and scaled) band-pass filters to estimate the local spectrum at a given point of attention of the image, the inverse Fourier transform of this local spectrum is a log-polar foveated version of the original image at that position. Both the local spectrum and its associated foveated image can be obtained through log-polar warping of the spectral/spatial domain followed by a conventional invariant low-pass filtering and the corresponding inverse warping. Furthermore, the low-pass filters in the warped space and frequency domains are mirror versions of each other. Thus, filters with mirror symmetry under the log-polar warping are *self-dual*, and make the foveatization process commute with the Fourier transform. Nevertheless, in order to implement a fovea that can be easily moved across the image, it is preferable to use a fixed bank of steerable filters, instead of applying log-polar warpings with different centers. Using low-pass scalable filters we have implemented a real-time moving fovea. We believe that a dual finite spatial/spectral local representation of images could be a very powerful tool for many visual tasks, which could benefit from a dual explicit representation in space and spatial frequency, as well as from the rotation and scale invariance naturally achieved in both domains.

## 1. INTRODUCTION

One common procedure in either biological and artificial (electronic) visual processing is to analyze the frequency contents of images. Since images are non-stationary signals, frequency content generally changes as one moves the attention point. For this reason, frequency analysis is more useful when it is localized in space. Classic methods for the local estimation of the frequency spectrum such as the sliding-window or short-time Fourier analysis (STFA),[1] localize the signal by a window before computing its Fourier transforming. The spectrogram[2] or the Gabor decomposition of a signal[3] are well-known examples of this approach. More recently, multi-scale methods[4,5] have been proposed as a better solution to the local estimation of the frequency spectrum.

The local spectrum of an image is naturally represented in the frequency domain. Conversely, it has the corresponding representation in the spatial domain, its inverse Fourier transform, denoted hereafter as the local image. This will resemble the original image locally around the point of attention, as the local spectrum contains information gathered around there.

Fig. 1 illustrates the concept of local image and its relationship with local spectrum estimation. The original image on the left panel contains three pure frequencies. The other two panels show the local images associated to a STFA method (center) and to a multi-scale method (right). In the STFA method the local image is simply a windowed version of the original. Here the window size is constant for all frequencies, and in this particular example, the size of the window is well adapted to the middle frequency. However, it is not large enough to capture the lowest frequency, while being unnecessarily large for the highest frequency component.

The local image displayed in the right panel (inverse Fourier transform of the continuous multi-scale local spectrum of the original at the center) illustrates the way multi-scale methods solve this problem, by using a window size inversely proportional to each frequency, and, thus, adapted to the frequency to be estimated. We can see how the



Part of the IS&T/SPIE Conference on Human Vision and Electronic Imaging III
San Jose, California • January 1998 • SPIE Vol. 3299 • 0277-786X/98/$10.00

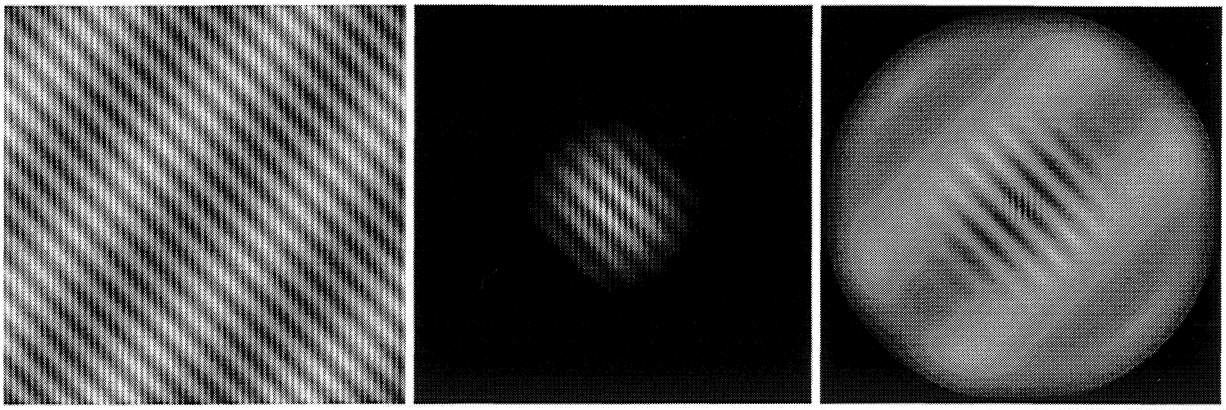


**Figure 1.** Original image containing three frequencies (left) and the corresponding local signals obtained with conventional STFA (center) and with a multi-scale method (right).

highest frequency component disappears rapidly as we move away from the center, while the middle frequency is attenuated more gradually, and in the far periphery only the lowest frequency remains visible.

The study of biological visual systems has revealed how a multi-scale local spectrum estimation is performed, through cells whose receptive fields (RF), having band-pass characteristics, are spatially localized.[6] This property has often been schematically modeled with banks of self-similar filters or channels.[7] In these simple models the RFs would have a spatial support inversely proportional to their tuning frequency, with profiles resembling those shown in the right panel of Fig. 1.

Another major feature of the human visual system (already incorporated to some modern electronic imaging devices) is its strongly inhomogeneous sampling of the retinal image, with a fine sampling at the fovea, which becomes progressively coarser towards the periphery.[8] This is basically the same effect that we can observe in the local signal (Fig. 1, right). Foveatization seems to be the trade-off of the visual system to meet two contradictory goals, namely, high resolution and a wide visual field. This solution seems also desirable in artificial vision, where it is crucial to minimize the amount of information to be processed.[9,10]

The above example shows how the local signal acquires a much richer significance in multi-scale methods as opposed to the simple windowing applied in STFA, allowing us to link these two important features of the visual system: (1) multi-scale local spectrum estimation and (2) foveatization, that, to our knowledge, have been so far treated as independent facts in the literature. Here, we will show that there is an intrinsic connection between them: if the local spectrum is estimated using rotated–scaled filters, we shall proof that the corresponding local image (its inverse Fourier transform) is a foveated version of the original. Conversely, a local spectrum estimation at the origin corresponds to a "foveatization" of the global spectrum. This means that the low frequencies around the DC will be similar in both the global and local spectra (which makes sense, as they have a large support in the local image), whereas the higher frequencies will be progressively smoothed (which corresponds to a higher localization in the spatial domain).

Therefore, foveatization can be realized as a multiscale spectral analysis, but directly performed in the space domain, and *viceversa*, multiscale frequency analysis is equivalent to a foveatization, now implemented in the Fourier domain. We call this property *duality* of the local spectrum in both domains, or equivalently *duality* of foveatization in both domains. The theoretical analysis is strongly simplified under a log-polar warping of both domains, which again reminds us of the biological findings (foveatization and octave-based frequency analysis), and helps us to get the intuitive idea of duality. We believe that this result could have relevant consequences in the analysis of both biological and artificial visual systems. The above considerations are graphically illustrated in Fig. 2, showing the relationship between the global and local images in both domains.

Given a filter applied in one domain, we can obtain the expression of its *dual filter*, to be applied in the conjugate domain. Furthermore, it is possible to design *self-dual* filters, i.e., filters that have the same expression in both domains, making the Fourier transform commute with the low-pass variant filtering.

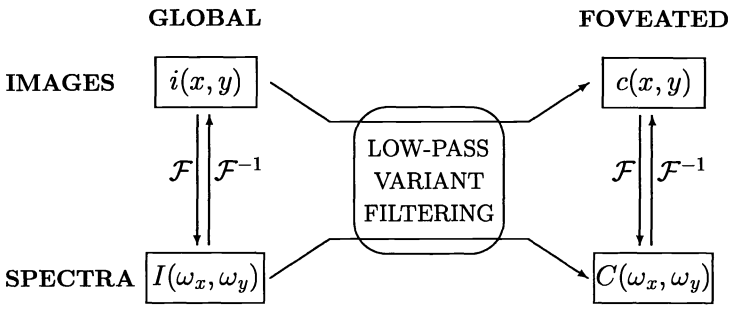



**Figure 2.** Relationship between the global (left) and foveated (right) images in both domains using a multi-scale method. The variant low-pass filtering can be the same if self-dual kernels are used.

## 2. THEORY

The common method for estimating the local spectrum of an image $i(\bar{x})$ at a given point of attention (origin) is to perform a series of inner products with a set of chosen functions $\{h_{\bar{\omega}_i}(\bar{x})\}$:

$$C(\bar{\omega}_i) =< i(\bar{s}), h_{\bar{\omega}_i}(\bar{x}) >= \iint_{\mathcal{R}^2} d\bar{s}\, i(\bar{s})\, h^*_{\bar{\omega}_i}(\bar{x}), \tag{1}$$

The analysis filters for this particular task, $\{h_{\bar{\omega}_i}(\bar{s})\}$, should be band–pass functions centered at $\bar{\omega}_i$ in the frequency domain and around the origin in the space domain. Since we want to stress the idea of a continuos local spectrum, we will will substitute $\bar{\omega}$ for $\bar{\omega}_i$. In Eq. 1 $C(\bar{\omega})$ is obtained from the image, but we can obtain it in the frequency domain as well, through Parseval's identity:

$$C(\bar{\omega}) = \iint_{\mathcal{R}^2} d\bar{\lambda}\, I(\bar{\lambda})\, H^*_{\bar{\omega}}(\bar{\lambda}), \tag{2}$$

where $H_{\bar{\omega}}(\bar{\lambda})$ and $I(\bar{\lambda})$ are the Fourier transforms of $h_{\bar{\omega}}(\bar{x})$ and $i(\bar{x})$, respectively.

The corresponding analysis windows in the Fourier domain $\{H_{\bar{\omega}}(\bar{\lambda})\}$ will average the values of the spectrum of the signal around the frequency $\bar{\omega}$ Therefore, by moving across frequencies to obtain $C(\bar{\omega})$, we are performing a low–pass filtering of the global spectrum $I(\bar{\lambda})$. Consequently, the local spectrum at $\bar{x} = 0$ can be interpreted either as a *band pass filtering* of the signal at the origin, or as a *low pass filtering* of its spectrum. It is this latter view we would like to dwell further on. Note that in Eq. (2), the filtering kernel $H_{\bar{\omega}}(\bar{\lambda})$ is not necessarily expressible as $H(\bar{\omega} - \bar{\lambda})$, so that the low pass filtering will not be invariant in general. This is the main difference between STFA methods that use a constant size window (and can be interpreted as invariant filtering of the spectrum), and multi-scale methods, that involve a variant filtering.

Inspired by these multi-scale methods, we consider here the case where the filtering kernel, $H_{\bar{\omega}}(\bar{\lambda})$, is obtained by scaling and rotating a prototype filter $P(\bar{u})$. In terms of $(\omega, \Omega)$ and $(\lambda, \Phi)$, the polar coordinates of $\bar{\omega}$ and $\bar{\lambda}$ respectively:

$$H_{\bar{\omega}}(\bar{\lambda}) = \frac{1}{\omega^2} P(\frac{\lambda}{\omega}, \Omega - \Phi). \tag{3}$$

The width of the filter scales linearly with the radial center frequency, as in multi-resolution and wavelet schemes, and rotates with the angular center frequency. However, contrarily to other multi-resolution methods,[4,5] this set of kernels have constant integrated response (due to the $\omega^2$ factor), instead of constant energy, following our interpretation of a local average of the global spectrum. The prototype filter $P(\bar{u})$ should be low-pass with its maximum around $\bar{u} = 1$. For convenience, we will impose $P(\bar{u})$ to be real and positive, like most standard low-pass filters. Furthermore, we will restrict ourselves to filters with support in one half-plane of the frequency domain (the equivalent of analytic filters for 1D signals). This latter condition is not strictly necessary, but it is important to interpret the associated local image as a foveated version of the original. Finally, $P(\bar{u})$ must also verify some integral conditions to be discussed below.

## 2.1. Local image: Duality and dual filters

Applying the kernel of Eq. 3, we obtain our local spectrum $C(\bar{\omega})$ as a variant low-pass filter of $I(\bar{\omega})$:

$$C(\omega,\Omega) = \iint_{\mathcal{R}^2} I(\lambda,\Phi)\,\frac{1}{\omega^2}\,P(\frac{\lambda}{\omega},\Omega-\Phi). \tag{4}$$

Let us now consider the inverse Fourier transform of the above expression, what we call the local signal $c(\bar{x})$, and relate it to the the original signal, obtaining that

$$c(\bar{x}) = \iint_{\mathcal{R}^2} d\bar{s}\, i(\bar{s})\, \widetilde{H}^{*}_{\bar{x}}(\bar{s}). \tag{5}$$

Thus, the variant filtering performed in the frequency domain is maintained as another similar variant filtering in the space domain with a *dual* kernel, $\widetilde{H}_{\bar{x}}(\bar{s})$. It is easy to demonstrate that the space-domain kernel $\widetilde{H}_{\bar{x}}(\bar{s})$ is related to the frequency-domain $H_{\bar{\omega}}(\bar{\lambda})$, through a double Fourier transform (plus one axis reflection):

$$\widetilde{H}_{\bar{x}}(\bar{s}) = (\frac{1}{2\pi})^2 \iint_{\mathcal{R}^2} d\bar{\omega}\, e^{-i\bar{\omega}\bar{x}} \iint_{\mathcal{R}^2} d\bar{\lambda}\, e^{i\bar{\lambda}\bar{s}}\, H_{\bar{\omega}}(\bar{\lambda}), \tag{6}$$

Therefore we can use either $H_{\bar{\omega}}(\bar{\lambda})$ or $\widetilde{H}_{\bar{x}}(\bar{s})$ for estimating the local spectrum, depending on the domain of application ($\bar{x}$ or $\bar{\omega}$) we choose.

In the particular case when the frequency domain kernel $H_{\bar{\omega}}(\bar{\lambda})$ is given by Eq. 3, the expression of its dual kernel can be obtained through Eq. 6. Considering again polar coordinates for the space variables $\bar{x}$ $(r,\theta)$, and $\bar{s}$ $(s,\varphi)$:

$$\widetilde{H}_{\bar{x}}(\bar{s}) = \frac{1}{s^2} P(\frac{r}{s}.\theta-\varphi). \tag{7}$$

The dual kernel in the space domain has the same expression as in the frequency domain $H_{\bar{\omega}}(\bar{\lambda})$ (Eq.3), except for the role of the dummy radial variables ($\lambda$ in frequency, $s$ in space), that has been interchanged. Although different in general, both kernels are low-pass filters that broaden as we move further away from the origin. Therefore, a similar variant low-pass filtering is applied in both domains. This is a significant difference with respect to STFA methods, where the operations in the conjugate domains are quite different: windowing in the spatial domain vs. convolution in the frequency domain.

Now we can explain better why we had restricted ourselves to prototype filters with only half-plane support. In theory, nothing prevents us from using a symmetric filter to obtain a local spectrum estimate. However, because of the duality, essentially the same filtering is performed in the spatial domain. Therefore, the local signal at a given position $\bar{x}$ will average values from both $\bar{x}$ and $-\bar{x}$. Prototype filters with half-plane support provide localized kernels in the spatial domain, allowing us to interpret the resulting local signal as a foveated image.

In order to illustrate these concepts let us consider the following prototype filter:

$$P(\omega,\Omega) = \begin{cases} \omega^2\,\cos^2(\Omega)\,\exp^{-\frac{\omega^2}{2}} & \text{for } \Omega \in [0,\pi/2]\cup[3\pi/2,2\pi] \\ 0 & \text{for } \Omega \in (\pi/2,3\pi/2) \end{cases} \tag{8}$$

This frequency/orientation selective filter has been obtained by eliminating the second lobule of $\cos^2(\theta)$ in the Fourier transform of the second derivative (along $x$) of a Gaussian, $\omega_x^2 \exp(-\omega^2/2)$, commonly used in image analysis applications. Both the original frequency kernel (left) and its dual one (right) are displayed in Fig. 3.

Additional contraints on the prototype filter are necessary to guarantee that Eqs. 4 and 5 are well defined everywhere, and that the local image and local spectrum given by those equations form a meaningful Fourier pair. These conditions[11] are that both variant kernels $H_{\bar{\omega}}(\bar{\lambda})$ and $\widetilde{H}_{\bar{x}}(\bar{s})$ be integrable. Translating those conditions onto the prototype filter $P(\bar{u})$ we obtain that

$$\iint_{\mathcal{R}^2} d\bar{u}\, P(\bar{u}) < \infty \qquad \text{and} \qquad \iint_{\mathcal{R}^2} d\bar{u}\, \frac{P(\bar{u})}{|\bar{u}|^2} < \infty, \tag{9}$$

which indicate a sufficient fast decay at both the origin and $\infty$. Additionally it would be convenient that both integrals would be normalized to 1. This would guarantee[11] that both the value at the origin of the image and its DC component are preserved in the local signal and local spectrum respectively.

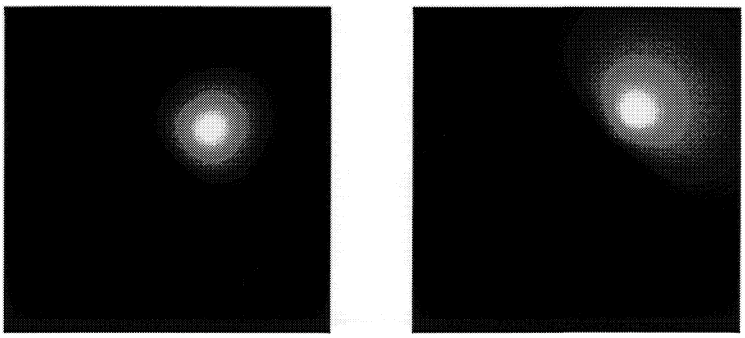


**Figure 3.** 2D kernel (left) derived from the prototype filter of Eq. 8 and its corresponding dual kernel (right).

## 2.2. Log-polar warping and the 2D scale transform

Logarithmic transformations naturally appear associated to scaling, such as in wavelets,[5] or the Cohen's scale transform.[12] The log-polar warping appears as a natural extension for 2D images, turning a rotating-scaling variant filtering into a simple invariant convolution. Scaling by $\omega$ and rotation by $\Omega$ become displacements by $\log\omega$ and $\Omega$ respectively. The warping must include a normalization factor to keep the energy constant:

$$f(r,\theta) \quad \text{with } r \in \mathcal{R}^+,\ \theta \in [0,2\pi] \implies \hat{f}(\hat{r},\theta) = e^{\hat{r}} f(e^{\hat{r}},\theta) \quad \text{with } \hat{r} \in \mathcal{R},\ \theta \in [0,2\pi] \tag{10}$$

Octave (log) filters[7] and log-polar warping[13] of the retinal image are widely used in modeling the visual system, based on the experimental findings about the distribution of photoreceptors in the retina[8] and the profile of receptive fields of simple cortical neurons.[6]

Since the frequency (Eq. 4) and spatial (Eq. 5) filtering involve scaled and rotated filters, both can equally benefit from a log-polar warping. Let us first consider the local spectrum: In the new warped domain it is simply obtained as an invariant low-pass filtering of the global warped spectrum:

$$\hat{C}(\hat{\omega},\Omega) = e^{\hat{\omega}}\, C(e^{\hat{\omega}},\theta) = \int_0^{2\pi} d\Phi \int_{-\infty}^{\infty} d\hat{\lambda}\ \hat{I}(\hat{\lambda},\Phi)\, \hat{P}(\hat{\lambda}-\hat{\omega},\Omega-\Phi), \tag{11}$$

Similarly, the local or foveated image can be realized as an invariant low-pass filtering, now applied in the (warped) spatial domain to the original image:

$$\hat{c}(\hat{r},\theta) = e^{\hat{r}}\, c(e^{\hat{r}},\theta) = \int_0^{2\pi} d\varphi \int_{-\infty}^{\infty} d\hat{s}\ \hat{i}(\hat{s},\varphi)\, \hat{P}(\hat{r}-\hat{s},\theta-\varphi), \tag{12}$$

The above expressions reflect clearly the total duality of the processing to obtain either the local spectrum or the local image from the corresponding global representations, through an invariant convolution with the warped prototype filter $\hat{P}$. The only difference is that in the spatial domain (12) $\hat{P}(\hat{r},\theta)$ is applied, while when estimating the local spectrum (11) we use $\hat{P}(-\hat{\omega},\Omega)$.

Consequently, there is a duality between (1) foveatization (a variant low-pass filtering of an image) and (2) multi-scale multi-orientation local spectrum estimation (a variant low-pass filtering of the global spectrum). Both features have been found in biological visual systems, and now appear to be the two components of a Fourier pair, and therefore, they can be realized as two different aspects of a common fact. Foveatization could be interpreted as a direct natural consequence of local frequency analysis and conversely. This duality is more clearly stated under log-polar warping, where a simple mirror symmetry exists between the kernels used in both domains. Again, the log-polar mapping is present in the visual system, as an approximate model of foveatization, or in the octave based visual channels.

### 2.2.1. Scale and rotation invariant representation of images

Once the foveated image has been expressed as an invariant convolution in the warped log-polar domain, we can take advantage of the translation-invariant properties of the Fourier transform. Then, given an image $i(r,\theta)$ we define $D_i(a,k)$ as the Fourier transform of its log-polar warped representation $\hat{i}(\hat{r},\theta)$,

$$D_i(a,k) = \frac{1}{2\pi}\int_0^{2\pi} d\theta\, e^{-ik\theta} \int_{-\infty}^{\infty} d\hat{r}\, e^{-ia\hat{r}}\ \hat{i}(\hat{r},\theta) \quad \text{with } k \in \mathcal{Z}. \tag{13}$$

Note that since the dependency of $\hat{i}(\hat{r},\theta)$ in $\theta$ is $2\pi$ periodic, we will analyze it through its Fourier series. Accordingly, the first variable $a$ in $D_i(a,k)$, is continous, while the second $k$ is discrete. Expressing Eq. 13 as a function of the original signal $i(r,\theta)$ we obtain:

$$D_i(a,k) = \frac{1}{2\pi}\int_0^{2\pi} d\theta\, e^{-ik\theta} \int_0^{\infty} dr\, e^{-ia\log r}\, i(r,\theta) = \iint_{\mathcal{R}^2} d\bar{r}\, i(\bar{r})\, \{\frac{1}{2\pi r} e^{-i(a\log r + k\theta)}\} =< i(r,\theta), \gamma_{ak}(r,\theta) >, \tag{14}$$

this being a generalization in 2D of Cohen's 1D scale transform.[12] In the same way that the 1D scale transform is invariant under scaling, the modulus of this 2D transform is invariant under either rotation or scaling of the image.

This transform can be realized as the projection of the image $i(r,\theta)$ onto a set of functions $\gamma_{a,k}(r,\theta)$, whose radial part had been proposed before in Ref..[14]) Since these functions are orthogonal, the corresponding inverse transform is simply:

$$i(r,\theta) =< D_i(a,k)\,,\, \gamma^*_{a,k}(r,\theta) >= \frac{1}{2\pi r}\sum_k e^{ik\theta} \int_{-\infty}^{\infty} da\, D_i(a,k)\, e^{ia\log r}. \tag{15}$$

Although there are other alternative proposals[15] for applying the scale transform to images, our transformation (Eq. 14) has the advantage of separating the radial and angular dependencies in the signal, applying a 1D scale transform only to the radial variable, the only one affected by a global change of scale. The Fourier analysis on the angular dependency is only a tool to accomplish invariance under rotations.

Applying the convolution theorem, we can transform Eq. 12 into the product of the 2D *scale-rotation transforms* of the image and the prototype filter,

$$D_c(a,k) = D_i(a,k)\, D_P(a,k), \tag{16}$$

and, therefore, a foveated image can be interpret as an invariant filtering in the scale domain.

## 2.3. Self-Dual filters

We have shown how the variant filtering is similar in both domains as it relies on the same low–pass function $P(\bar{u})$. However, the twist in the dummy variables of the corresponding kernels prevents a complete duality between both domains. However, it is easy to show that by carefully choosing $P(\bar{u})$ we can obtain *self-dual* kernels, where the two kernels of a dual pair are identical. The condition for self-duality is immediate in the log domain (Eqs. 12 and 11): the log-warped prototype filter must be even in its radial variable

$$\hat{P}(s,\theta) = \hat{P}(-s,\theta). \tag{17}$$

Consequently, given any 2D function $f(x,\theta)$, defined in $\mathcal{R}^+ \times [0,2\pi]$, we can obtain a prototype filter generating self-dual kernels as:

$$P(r,\theta) = \frac{1}{r} f(|\log r|,\theta). \tag{18}$$

The only additional condition on $f(x,\theta)$ is that it must decay fast enough when $x \to \infty$, so that the resulting $P(s,\theta)$ verifies the integral conditions of Eq. 9.

Another important motivation behind the use of self–dual filters is that, since both kernels are the same, their integrals would also be identical. That was precisely the condition for the preservation of both the origin and the DC component of an image in its local representations. When using self-dual filters that condition is trivially ensured

by simply using a normalized $P(\bar{u})$. Fig. 4 displays several self-dual kernels centered at different frequencies and orientations (rotated and scaled versions of each other), generated from $f(x,\theta) = \exp -(k_x x^2) \exp -(k_\theta \theta^2)$. Choosing the values $k_x = 16.0$ and $k_\theta = 9.0$ we obtain kernels with a bandwidth of about 0.6 octaves in frequency and $45^o$ in orientation. As the width of filters shown in Fig. 4 increases with the distance to the origin, the corresponding dual kernels, being identical, will produce a "foveatization" effect when applied to an image.

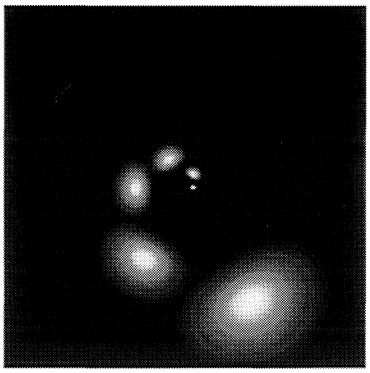


**Figure 4.** Six self-dual 2D filtering kernels centered at different frequencies and orientations.

An interesting property is that it is very simple to obtain a self-dual fovea through non self-dual filters.[11] The idea is to average the two dual filters, so the equivalent one be symmetric. The procedure consists of: (1) applying the desired filter in the frequency domain to obtain the local spectrum, and compute the inverse Fourier transform to obtain one foveated image; (2) applying the same filter directly in the space domain to obtain a second foveated image (these two foveated images will be different unless the filter is self-dual); (3) then, all we need is to average both foveated images. The result is a self-dual fovea, whose associated local spectrum will also be self-dual.

Fig. 5 illustrates the existing duality with an example. The upper panels display the original image (left) and its corresponding global spectrum (right). Their associated local representations are shown in the lower panels. On the left we can appreciate the foveated local image, while the local spectrum is on the right. The center of the original image, the point of attention, was preserved in the local image, while the low pass filtering effect increases progressively as we move away from the origin. The estimated local spectrum is also a foveated version of the global spectrum (in both cases the modulus is displayed in a log intensity scale). The fine details have been preserved around the origin (DC), while being progressively lost towards the periphery. In this example we have applied a set filters similar to those of Fig. 4, but having a smaller bandwidth, about 0.25 octaves, to obtain a larger fovea.

## 3. IMPLEMENTATION THROUGH STEERABLE/SCALABLE FILTERING

In this section we present a practical implementation of a log-polar fovea and discuss some potential applications of the duality local spectrum/image.

The use of scaled filters to compute the local spectrum/image involves a variant filtering, which is highly expensive computationally when implemented directly. In Section 2.2 we saw that the log-polar warping facilitated the mathematical analysis, but practical foveatization through this method have two important drawbacks. First, the warping itself requires a preliminary variant low-pass filtering or interpolation, in order to avoid aliasing, which must be taken into account when applying the filter in the log-domain to obtain the desired global low-pass filtering. Another problem is that there is no efficient way to move the attention point, and the whole foveatization process has to be repeatedfor each location of the fovea. Next we propose some alternative techniques that overcome these problems and have some other interesting features.

### 3.1. Sampling and interpolating the local spectrum/image

Since both the local spectrum (at the origin) and the local image are variant low-pass filtered versions of the corresponding global signals they can be subsampled without (significant) loss of information. Their finite size and

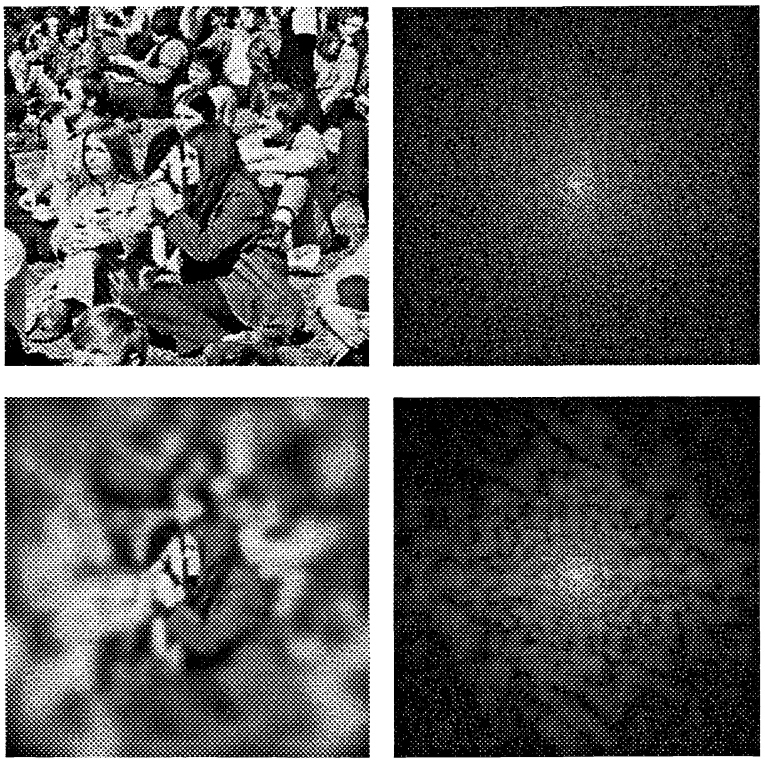


**Figure 5.** Left: Original image (above) and "foveated" image (below) obtained by filtering with self-dual kernels. Right: Moduli of the corresponding Fourier transforms (in a log intensity scale), that is, global spectrum (above) and local spectrum at the origin (below).

bandwidth guarantee that only a finite number of samples are needed (however, this demands some changes with respect to the theory in the way sampling and interpolation are carried out, as in the treatment of the singularity in the center of the fovea).

Our idea is to apply the steerable or deformable filtering model,[16–19] to achieve a continuous local spectrum estimation from a discrete one. Linearity of the Fourier transform allows us to apply this method to the spatial domain in the same way.

Practical filtering schemes used for local spectrum estimation are based on the application of a finite number of filters, each one associated to to a scale and an orientation. This can be seen as equivalent to sampling a continuous local spectrum estimated using a continuum of filters. On the other hand, we have shown that the continous local spectrum is a log-polar low-pass filtered version of the original spectrum when using self-similar filters and the attention point lays at the origin. Can we sample the local spectrum without significant loss of information? And, if that is possible, how can we recover it from a discrete number of samples? In general, an exact recovery is not possible using a finite number of filters, due to the lack of periodicity of scale,[18,19] but we can obtain an arbitrarily good recovery by adding more samples to the representation. As a practical rule, we have verified that a good log-polar sampling requires a density around twice as much as the one we would expect from the spectral bandwidth of the filters. For example, multi-scale filtering schemes use a bandwidth typically around 1 octave in scale and around 40 degrees in orientation, with a sampling density of one filter per octave in scale and about 45 degrees in orientation. For covering 4 octaves, they require, thus, around 16 (4x4) filters. However, a good representation of the continuous local spectrum would require in this case around 64 (2x4x2x4) basis filtersh. The computation of the optimal interpolating functions from a given prototype filter and spectral location of the samples can be carried out by minimizing the squared error between the desired filters and a linear combination of the basis filters.[20,21]

Here, we have implemented a bank of 8 (scales) x 8 (orientations) polar-separable filters, covering 5 octaves and each one having a bandwidth of 1 octave and 40 degrees. The filters have been designed with a new method[22] based on minimizing the interpolation error for a given bandwidth and sampling lattice. Figure 6(a) shows the interpolating functions obtained in this case and the fidelity achieved in the interpolation as a function of the scale (in decibels). We have applied this bank of filters to obtain the dual foveated image corresponding to the (discrete) local spectrum of the image at each point, as a practical demonstration of (1) the existing duality between local spectrum and foveated image, and (2) our sampling/interpolation method based on steerable filters. Figure 6(b) shows a diagram of this procedure. Figure 7 displays the results of applying this method to the original image (a), the resulting local spectrum showing the location of the 8x8 basis filters (b), and the resulting foveated image (c). Panel (d) is explained below.

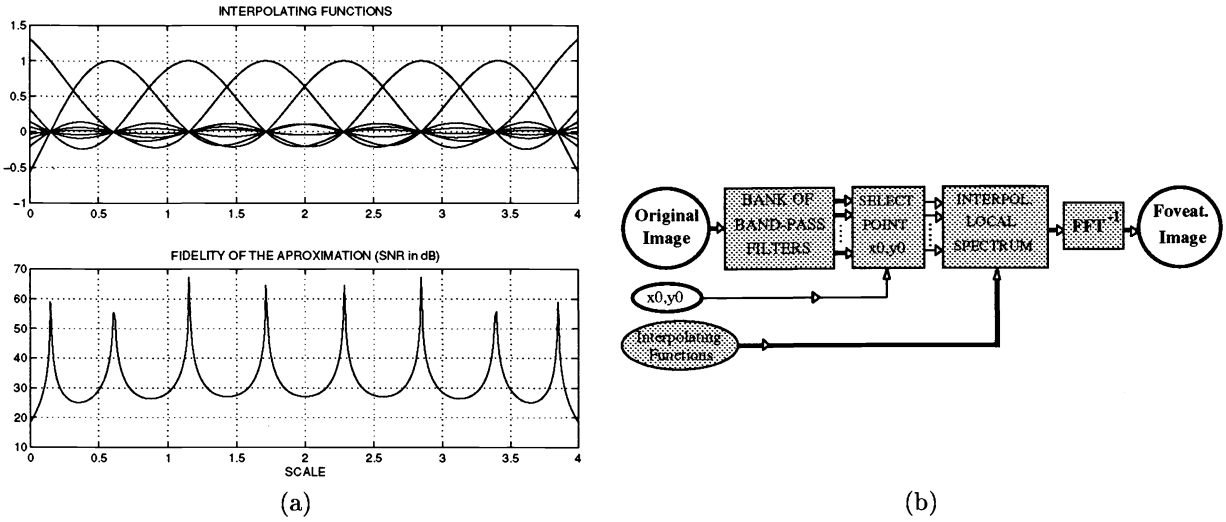



**Figure 6.** (a) Interpolating function of the filters, in scale (upper) and fidelity in the reconstruction as a function of the scale in dB (lower). (b) Diagram of the method used to obtain the foveated image.

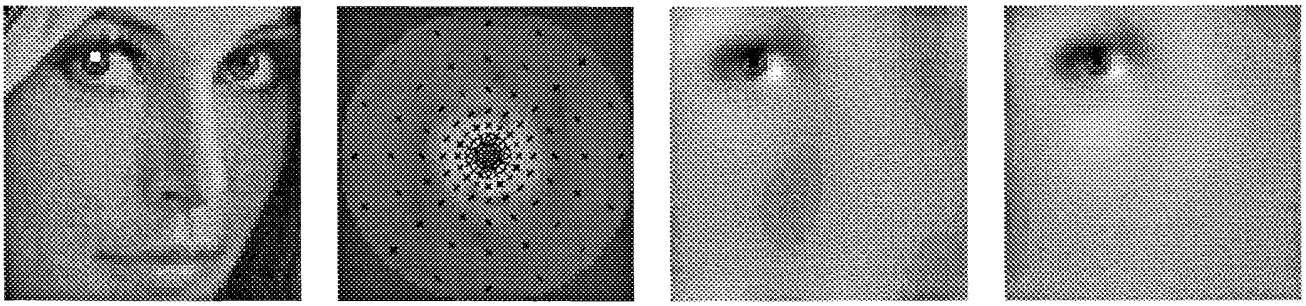



**Figure 7.** (a) Original image; the white dot is the attention point. (b) Modulus of the interpolated spectrum from the 8x8 spectral samples (indicated by black crosses). (c) Fovea obtained through inverse Fourier transform of (b). (d) A similar fovea obtained by filtering directly in the spatial domain.

### 3.2. Efficient implementation using filters with circular symmetry

In contrast to the log-warping method, steerable filtering permits to change the point of attention (O) without doing additional filtering, as we have seen. However, in the above implementation, each move of O required a new interpolation and an inverse Fourier transform. In addition, a high number of filters were required. To solve the first

problem we can take advantage of the duality between local spectrum and foveated image, which allows us to apply directly the filters in the spatial domain. The set of outputs of these spatially shifted low-pass filters (non symmetric in general) correspond now to a sampled version of the foveated image (similarly to the samples in Fig.7(b)). For obtaining the continuous fovea we only need to interpolate these outputs at each attention point, as before. An important advantage of working in the spatial domain is that the number of filters can be drastically reduced by using circular symmetric kernels. In this case, all the masks corresponding to the same scale are shifted version of each other and, therefore, they can be substituted by just one filter. Furthermore, the Fourier transform of these low-pass filters overlap each other in the frequency domain much more than the original masks do in the space domain. This allows us to steer them in the latter domain using a much smaller set of filters than before (the number of required filters is reduced in a factor about 2N, being N the number of orientations considered).

This method has two minor disadvantages, though: (1) the filters are not self-dual (self-duality could be achieved using the technique explained in Section 2.3, but at the cost of doubling computation); (2) we no longer have a direct correspondence between outputs of the filters and samples of the fovea/local spectrum. Nevertheless, self-duality, although desirable, is not a critical requirement in practice for most applications, since most nice properties of duality do not rely on it. The lack of correspondence between samples and outputs of the filters, on the other hand, only affects to those applications that need an explicit expression of the discrete fovea or local spectrum (in these cases the spatial sampling within each scale should be done a posteriori).

To obtain the fovea of Fig.7(d), we have used 5 1-octave Gaussian kernels, distributed by octaves, plus a "null" filter to cover the attention point. With this scheme, the average fidelity in the interpolation was higher than 30 dB in the 5 octave range considered. We can see that it is very similar to the one obtained through inverse Fourier transform of the local spectrum. Differences mainly come from using different kernels and the wrap-around effect in (c).

Figure 8 shows the diagram of the space-domain foveatization. A look up table (LUT) provides the interpolation masks obtained from the interpolating functions (for a detailed explanation, see Ref.[21]). Now, moving the fovea only requires 1) shifting a small number of fixed masks (one for each scale); 2) doing the corresponding point-wise multiplications; and 3) adding the outputs together into a single image. The efficiency of this method allows to perform several foveatizations per second in low-resolution images (64x64 or 128x128) using general purpose hardware and non-optimized software.

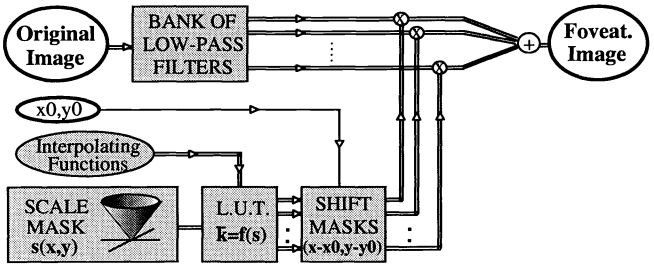



**Figure 8.** Diagram showing the foveatization method using scalable low-pass filters directly in the spatial domain.

### 3.3. Example of application: an active vision system

We have applied the above foveatization method to implement a simple algorithm that sequentially looks for the "interesting" points of an image, and composes a *multi-foveated* image with the corresponding foveas. Figure 9 shows a scheme of this algorithm. Each new attention point is placed over the area with the highest *edge content.* This is done by applying a sobel edge detector to the difference between the original image and the current multi-foveated image (initially set to zero), and then finding the maximum of a low-pass version of the result. In this example we have favoured the location of the foveas in the central part of the image by applying a spatial window to the output of the low pass filter (with a value of 1 at the center that decays to 0.5 at the edges of the image).

Figure 10 shows the results obtained using Gaussian filters of 0.5 octaves bandwidth, after 3 search steps (a), and after 12 steps (b). For each case, the attention points are overlaid on the original (left side) together with the

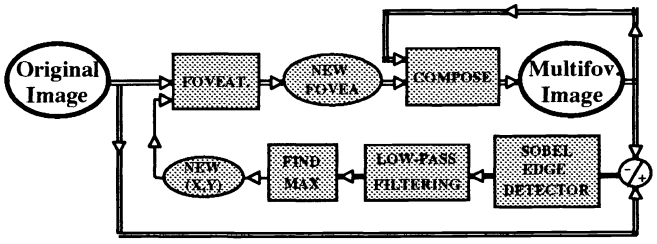
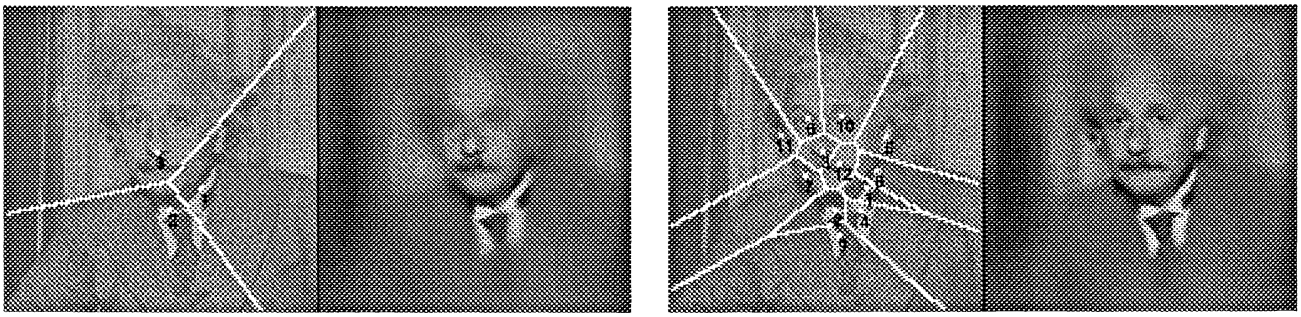



**Figure 9.** Diagram of the automatic multi-foveatization method.

resulting areas (Voronoy cells) used for composing the multi-foveated images (right side). Overlaid numbers refer to the sequential order of selection.



(a) (b)

**Figure 10.** Results of the sequential foveatization (a) after 3 steps, and (b) after 12 steps. See text for details.

This algorithm is robust and behaves in a strikingly similar way to the human visual system, considering its simplicity. Its application to dynamic compression seems promising, with the additional advantage of being free of aliasing artifacts, contrarily to other methods.[9,10]

### 3.4. Towards a discrete dual representation

Apart from errors associated to the discretization process, the sampled versions of the local spectrum and local image carry the same information. Moreover, as they formed a Fourier pair in the continuous case, a linear relationship between them remains after sampling. Thus, the local spectrum can be written as the product of the vector formed with the samples of the fovea by a square matrix, whose rows act as discrete filters in a log-polar grid. We believe that this kind of vectorial representation could be a powerful tool for many visual tasks, which would benefit from the explicit representation of the local image in both domains, as well as from an interesting family of invariant operators that can be derived from it. To the well-known invariances to position and orientation obtained by overcomplete pyramids,[20,23] our representation adds scale invariance, made possible by the high density sampling of this variable in the local spectrum. Moreover, the invariant representations discussed in the continuous case (Section 2.2.1), also apply in the discrete case. In contrast with the former invariance, the latter is phase sensitive, and thus, provides an information related with the shapes of the objects of the image, but independent to their size and orientation. We believe that this dual representation will be a powerful tool for many image processing and computer vision tasks, as it shares the discriminating power of multi-scale space/spatial-frequency methods with the robustness and velocity of foveated vision, taking advantage of two sources of complementary invariant features.

## ACKNOWLEDGEMENTS

This work has been supported by the Spanish Comisión Interministerial de Ciencia y Tecnología (CICYT) under grant TIC-97/325

## REFERENCES


1. J.B. Allen and L.R. Rabiner, "A unified approach to short-time Fourier analysis and synthesis", Proc. of the IEEE **65**, pp. 1558–1564, 1977.
2. R. Koenig, H.K. Dunn, and L.Y. Lacy, "The sound spectrograph", J. Acoust. Soc. Am. **18**, pp. 19–49, 1946.
3. D. Gabor, "Theory of communications", J. Inst. Elect. Eng. **93**, pp. 429–457, 1946.
4. S.G. Mallat, "A theory for multiresolution signal decomposition: the wavelet transform", IEEE Trans. on Patt. Anal. and Mach. Intell.**11**, pp. 674–693, 1989.
5. I. Daubechies, Ten Lectures on Wavelets, SIAM, Philadelphia, PA, 1992.
6. D.A. Pollen and S.F. Ronner, "Visual cortical neurons as localized spatial filters", IEEE Trans. on Sys., Man, and Cyber.**13**, pp. 907-916, 1983.
7. J.G. Daugman, "Spatial visual channels in the Fourier plane", Vision Res., **24**, pp. 891–910, 1984.
8. C.A. Curcio, K.R. Sloan, R.E. Kalina, and A.E. Hendrikson, "Human photoreceptor topography", Journal of Comp. Neurology **292**, pp. 497–523, 1990.
9. P. Kortum and W. Geisler, "Implementation of a foveated image coding system for image bandwidth reduction", Proc. SPIE **2657**, pp. 350–360, San Jose, California, 1996.
10. H. Yamamoto, Y. Yeshurun, M.D. Levine, "An active foveated vision system: attention mechanisms and scan path convergence measures", Comp. Vision and Image Understanding **63**, pp. 50–65, 1996.
11. A. Tabernero, J. Portilla, and R. Navarro, "Duality between the local spectrum of a signal and its inverse Fourier transform, the local signal". Instituto de Optica, Technical Report 53.
12. L. Cohen, "The scale representation", IEEE Trans. on Sig. Proc.**41**, pp. 3275–3293, 1993.
13. C. Braccini, G. Gambardella, and G. Sandini, "A signal theory approach to the space and frequency variant filtering performed by the human visual system", Signal Processing **3**, pp. 231–240, 1981.
14. M. Michaelis and G. Sommer, "A Lie group-approach to steerable filters", Pattern Recognition Letters **16**, pp. 1165–1174, 1995.
15. G. Cristobal and L. Cohen, "Scale in images", SPIE Proc. Advanced Signal Processing, vol. 2846, pp. 251–261, 1996.
16. W.T. Freeman and E.H. Adelson, "The design and use of steerable filters", IEEE Trans. on Patt. Anal. and Mach. Intell.**13**, pp. 891–906, 1991.
17. E.P. Simoncelli, W.T. Freeman, E.H. Adelson, and D. Heeger, "Shiftable multiscale transforms", IEEE Trans. on Inf. Theory**38**, No. 2, pp. 587–607, 1992.
18. P. Perona, "Deformable kernels for early vision", IEEE Trans. on Patt. Anal. and Mach. Intell.**17**, pp. 488-499, 1995.
19. P.C. Teo and Y. Hel-Or, "A computational group theory approach to steerable functions", Standford University, CA, Tech. Rep. STAN-CS-TN-96-33, April 1996.
20. H. Greenspan, S. Belongie, R. Gooodman, P. Perona, S. Rakshit, and C.H. Anderson, "Overcomplete steerable pyramid filters and rotation invariance", in Proc. of the IEEE Conf. on Comp. Vis. and Patt. Recognition, Seattle, pp. 222–228, 1994.
21. J. Portilla and R. Navarro, "Efficient method for space-variant low-pass filtering", in Proc. of the SNRFAI, pp. 287–292, Barcelona, Spain, 1997.
22. J. Portilla, Instituto de Optica, Technical Report to appear.
23. W. Beil, "Steerable filters and invariance theory", Patt. Rec. Lett. **15**, pp. 453–460, 1994.