Sparse Feature Maps in a Scale Hierarchy

Per-Erik Forssén and Gösta Granlund

Computer Vision Laboratory, Department of Electrical Engineering
Linköping University, SE-581 83 Linköping, Sweden

Abstract. This article describes an essential step towards what is called a view centered representation of the low-level structure in an image. Instead of representing low-level structure (lines and edges) in one compact feature map, we will separate structural information into several feature maps, each signifying features at a *characteristic phase*, in a specific scale. By characteristic phase we mean the phases 0, π, and $\pm\pi/2$, corresponding to bright, and dark lines, and edges between different intensity levels, or colours. A lateral inhibition mechanism selects the strongest feature within each local region of scale represented. The scale representation is limited to maps one octave apart, but can be interpolated to provide a continous representation. The resultant image representation is *sparse*, and thus well suited for further processing, such as pattern detection.

Keywords: sparse coding, image representation, view centered representation, edge detection, scale hierarchy, characteristic phase

1 Introduction

From neuroscience we know that biological vision systems interpret visual stimuli by separation of image features into several retinotopic maps [5]. These maps encode highly specific information such as colour, structure (lines and edges), motion, and several high-level features not yet fully understood. This feature separation is in sharp contrast to what many machine vision applications do when they synthesize image features into objects. We have earlier discussed these two approaches, which are called *view centered*, and *object centered* image representations [8]. This report describes an attempt to move one step further towards a view centered representation of low level properties.

As we move upwards in the interpretation hierarchy in biological vision systems, the cells within each feature map tend to be increasingly selective, and consequently the high level maps tend to employ more sparse representations [3]. There are several good reasons why biological systems employ sparse representations, many of which could also apply to machine vision systems.

Sparse coding tends to minimize the activity within an over-complete feature set, whilst maintaining the information conveyed by the features. This leads to representations in which pattern recognition, template storage, and matching are made easier [3]. Compared to compact representations, sparse features convey

more information when they are active, and contrary to how it might appear, the amount of computation will not be increased significantly, since only the *active* features need to be considered.

Most feature generation procedures employ filtering in some form. The outputs from these filters tell quantitatively more about the filters used than the structures they were meant to detect. We can get rid of this excessive load of data, by allowing only certain phases of output from the filters to propagate further. These characteristic phases have the property that they give invariant structural information rather than all the phase components of a filter response.

The feature maps we generate describe image structure in a specific scale, and at a specific phase. The distance between the different scales is one octave (i.e. each map has half the center frequency of the previous one.) The phases we detect are those near the *characteristic phases*[1] 0, π, and $\pm\pi/2$. Thus, for each scale, we will have three resultant feature maps (see figure 1).

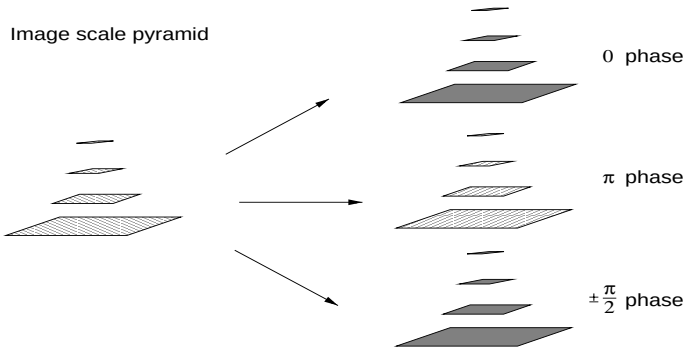

Fig. 1. Scale hierarchies.

This approach touches the field of scale-space analysis pioneered by Witkin [1]. See [2] for a recent overview. Our approach to scale space analysis is somewhat similar to that of Reisfield [4]. Reisfield has defined what he calls a *Constrained Phase Congruency Transform* (CPCT), that maps a pixel position and an orientation to an energy value, a scale, and a symmetry phase (0, π, $\pm\pi/2$, or none). We will instead map each image position, at a given scale, to three complex numbers, one for each of the characteristic phases. The argument of the complex numbers indicates the dominant orientation of the local image region at the given scale, and the magnitude indicates the local signal energy when the phase is near the desired one. As we move away from the characteristic phase, the magnitude will go to zero. This representation will result in a number of complex valued images that are quite sparse, and thus suitable for pattern detection.

[1] We will define the concept of characteristic phase in a following section.

2 Phase from Line and Edge Filters

For signals containing multiple frequencies, the phase is ambiguous, but we can always define the *local phase* of a signal, as the phase of the signal in a narrow frequency range.

The local phase can be computed from the ratio between a band-pass filter (even, denoted f_e) and it's quadrature complement (odd, denoted f_o). These two filters are usually combined into a complex valued *quadrature filter*, $\boldsymbol{f} = f_e + \boldsymbol{i} f_o$ [6]. The real and imaginary parts of a quadrature filter correspond to line, and edge detecting filters respectively. The local phase can now be computed as the argument of the filter response, $\boldsymbol{q}(x)$, or if we use the two real-valued filters separately, as the four quadrant inverse tangent; $\arctan(q_o(x), q_e(x))$.

To construct the quadrature pair, we start with a discretized lognormal filter function, defined in the frequency domain.

$$R_i(\rho) = \begin{cases} e^{-\frac{\ln^2(\rho/\rho_i)}{\ln 2}} & \text{if } \rho > 0 \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

The parameter ρ_i determines the peak of the lognorm function, and is called the center frequency of the filter. We now construct the even and odd filters as the real and imaginary parts of an inverse discrete Fourier transform of this filter.[2]

$$f_{e,i}(x) = \mathrm{Re}(\mathrm{IDFT}\{R_i(\rho)\}) \tag{2}$$

$$f_{o,i}(x) = \mathrm{Im}(\mathrm{IDFT}\{R_i(\rho)\}) \tag{3}$$

We write a filtering of a sampled signal, $s(x)$, with a discrete filter $f_k(x)$ as $q_k(x) = (s * f_k)(x)$, giving the response signal the same indices as the filter that produced it.

3 Characteristic Phase

By *characteristic phase* we mean phases that are consistent over a range of scales, and thus characterize the local image region. In practise this only happens at local magnitude peaks of the responses from the even, and odd filters.[3] In other words, the characteristic phases are always one of 0, π, and $\pm\pi/2$.

Only some occurrences of these phases are consistent over scale though (see figure 2). First, we can note that band-pass filtering always causes ringings in

[2] Note that there are other ways to obtain spatial filters from frequency descriptions that, in many ways produce better filters [7].

[3] A peak in the even response will always correspond to a zero crossing in the odd response, and vice versa, due to the quadrature constraint.

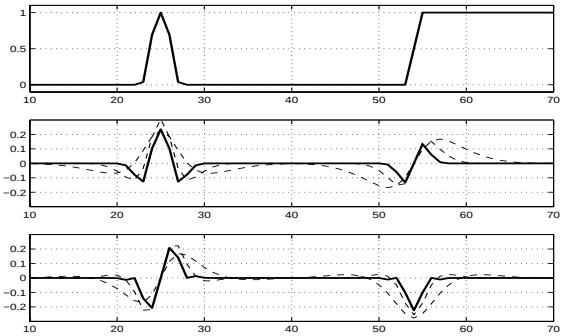

Fig. 2. Line and edge filter responses in 1D.
Top: A one-dimensional signal.
Center: Line responses at $\rho_i = \pi/2$ (solid), and $\pi/4$ and $\pi/8$ (dashed)
Bottom: Edge responses at $\rho_i = \pi/2$ (solid), and $\pi/4$ and $\pi/8$ (dashed)

the response. For isolated line and edge events this will mean one extra magnitude peak (with the opposite sign) at each side of the peak corresponding to the event. These extra peaks will move when we change frequency bands, and consequently they do not correspond to characteristic phases. Second, we can note that each line event will produce one magnitude peak in the line response, and two peaks in the edge response. The peaks in the edge response, however, are not consistent over scale. Instead they will move as we change frequency bands. This phenomenon can be used to sort out the desired peaks.

4 Extracting Characteristic Phase in 1D

Starting from the line and edge filter responses at scale i, $q_{e,i}$, and $q_{o,i}$, we now define three *phase channels*:

$$p_{1,i} = \max(0, q_{e,i}) \tag{4}$$
$$p_{2,i} = \max(0, -q_{e,i}) \tag{5}$$
$$p_{3,i} = \mathrm{abs}(q_{o,i}) \tag{6}$$

That is, we let $p_{1,i}$ constitute the positive part of the line filter response, corresponding to 0 phase, $p_{2,i}$, the negative part, corresponding to π phase, and $p_{3,i}$ the magnitude of the edge filter response, corresponding to $\pm\pi/2$ phase.

Phase invariance over scale can be expressed by requiring that the signal at the next lower octave has the same phase:

$$p_{1,i} = \max(0, q_{e,i} \cdot q_{e,i-1}/a_{i-1}) \cdot \max(0, \mathrm{sign}(q_{e,i})) \tag{7}$$
$$p_{2,i} = \max(0, q_{e,i} \cdot q_{e,i-1}/a_{i-1}) \cdot \max(0, \mathrm{sign}(-q_{e,i})) \tag{8}$$
$$p_{3,i} = \max(0, q_{o,i} \cdot q_{o,i-1}/a_{i-1}) \tag{9}$$

The first max operation in the equations above will set the magnitude to zero whenever the filter at the next scale has a different sign. This operation will reduce the effect of the ringings from the filters. In order to keep the magnitude near the characteristic phases proportional to the local signal energy, we have normalized the product with the signal energy at the next lower octave $a_{i-1} = \sqrt{q_{e,i-1}^2 + q_{o,i-1}^2}$. The result of this operation can be viewed as a phase description at a scale in between the two used. These channels are compared with the original ones in figure 3.

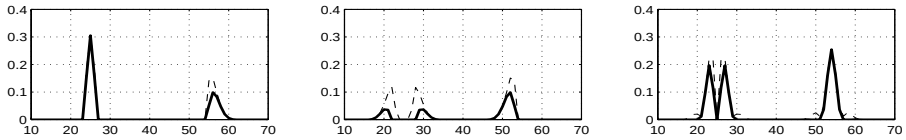

Fig. 3. Consistent phase in 1D. ($\rho_i = \pi/4$)
$p_{1,i}$, $p_{2,i}$, $p_{3,i}$ according to equations 4-6 (dashed), and equations 7-9 (solid)

We will now further constrain the phase channels in such a way that only responses consistent over scale are kept. We do this by inhibiting the phase channels with the complementary response in the third lower octave:

$$c_{1,i} = \max(0, p_{1,i} - \alpha \operatorname{abs}(q_{o,i-2})) \tag{10}$$
$$c_{2,i} = \max(0, p_{2,i} - \alpha \operatorname{abs}(q_{o,i-2})) \tag{11}$$
$$c_{3,i} = \max(0, p_{3,i} - \alpha \operatorname{abs}(q_{e,i-2})) \tag{12}$$

We have chosen an amount of inhibition $\alpha = 2$, and the base scale, $\rho_i = \pi/4$. With these values we sucessfully remove the edge responses at the line event, and a the same time keep the rate of change in the resultant signal below the Nyquist frequency. The resultant characteristic phase channels will have a magnitude corresponding to the energy at scale i, near the corresponding phase. These channels are compared with the original ones in figure 4.

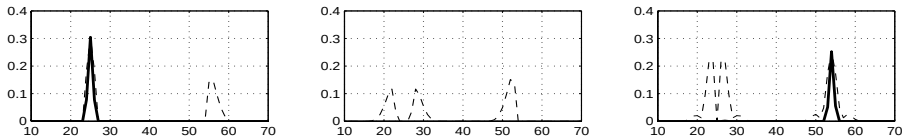

Fig. 4. Phase channels in 1D. ($\rho_i = \pi/4$, $\alpha = 2$)
$p_{1,i}$, $p_{2,i}$, $p_{3,i}$ according to equations 4-6 (dashed), and equations 10-12 (solid).

As we can see, this operation manages to produce channels that indicate lines and edges without any unwanted extra responses. An important aspect of this operation is that it results in a gradual transition between the description of a signal as a line or an edge. If we continuously increase the thickness of a line, it will gradually turn into a bar that will be represented as two edges.[4] This phenomenon is illustrated in figure 5.

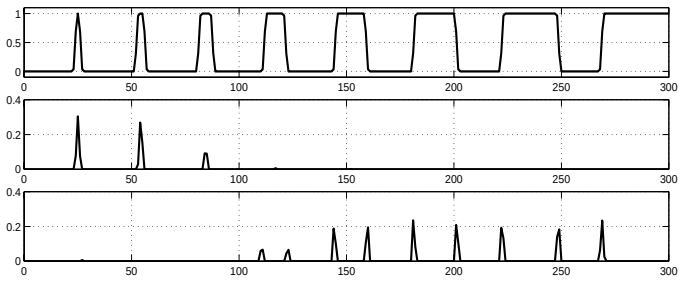

Fig. 5. Transition between line and edge description. ($\rho_i = \pi/4$)
Top: Signal Center: $c_{1,i}$ phase channel Bottom: $c_{3,i}$ phase channel.

5 Local Orientation Information

The filters we employ in 2D will be the extension of the lognorm filter function (equation 1) to 2D [6]:

$$\mathrm{F_{ki}}(\mathbf{u}) = \mathrm{R_i}(\rho)\,\mathrm{D_k}(\hat{\mathbf{u}}) \tag{13}$$

Where

$$\mathrm{D_k}(\hat{\mathbf{u}}) = \begin{cases} (\hat{\mathbf{u}} \cdot \hat{\mathbf{n}}_k)^2 & \text{if } \mathbf{u} \cdot \hat{\mathbf{n}}_k > 0 \\ 0 & \text{otherwise} \end{cases} \tag{14}$$

We will use four filters, with directions $\hat{\mathbf{n}}_1 = (\,0\;1\,)^t$, $\hat{\mathbf{n}}_2 = (\,\sqrt{0.5}\;\sqrt{0.5}\,)^t$, $\hat{\mathbf{n}}_3 = (\,1\;0\,)^t$, and $\hat{\mathbf{n}}_4 = (\,\sqrt{0.5}\;-\sqrt{0.5}\,)^t$. These directions have angles that are uniformly distributed modulo π. Due to this, and the fact that the angular function decreases as $\cos^2\varphi$, the sum of the filter-response magnitudes will be orientation invariant [6].

[4] Note that the fact that both the line, and the edge statements are low near the fourth event (positions 105 to 125) does not mean that this event will be lost. The final representation will also include other scales of filters, which will describe these events better.

Just like in the 1D case, we will perform the filtering in the spatial domain:

$$(f_{e,ki} * p_{ki})(\mathbf{x}) \approx \mathrm{Re}(\mathrm{IDFT}\{\mathrm{F_{ki}}(\mathbf{u})\}) \tag{15}$$

$$(f_{o,ki} * p_{ki})(\mathbf{x}) \approx \mathrm{Im}(\mathrm{IDFT}\{\mathrm{F_{ki}}(\mathbf{u})\}) \tag{16}$$

Here we have used a filter optimization technique [7] to separate the lognorm quadrature filters into two approximately one-dimensional components. The filter $p_{ki}(\mathbf{x})$, is a smoothing filter in a direction orthogonal to $\hat{\mathbf{n}}_k$, while $f_{e,ki}(\mathbf{x})$, and $f_{o,ki}(\mathbf{x})$ constitute a 1D lognorm quadrature pair in the $\hat{\mathbf{n}}_k$ direction.

Using the responses from the four quadrature filters, we can construct a *local orientation* image. This is a complex valued image, in which the magnitude of each complex number indicates the signal energy when the neighbourhood is locally one-dimensional, and the argument of the numbers denote the local orientation, in the *double angle representation* [6].

$$\boldsymbol{z}(\mathbf{x}) = \sum_k a_{ki}(\hat{n}_{k1} + \boldsymbol{i}\hat{n}_{k2})^2 = a_{1i}(\mathbf{x}) - a_{3i}(\mathbf{x}) + \boldsymbol{i}(a_{2i}(\mathbf{x}) - a_{4i}(\mathbf{x})) \tag{17}$$

where $a_{ki}(\mathbf{x})$, the signal energy, is defined as $a_{ki} = \sqrt{q_{e,ki}^2 + q_{o,ki}^2}$.

6 Extracting Characteristic Phase in 2D

To illustrate characteristic phase in 2D, we need a new test pattern. We will use the 1D signal from figure 5, rotated around the origin (see figure 6). The image has also been degraded with a small amount of Gaussian noise. The signal to noise ratio is 10 dB.

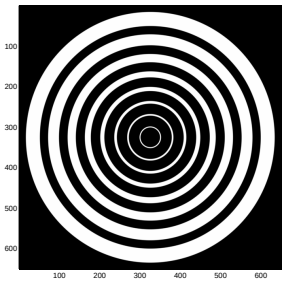

Fig. 6. A 2D test pattern. (10 dB SNR)

When extracting characteristic phases in 2D we will make use of the same observation as the local orientation representation does: Since visual stimuli can

locally be approximated by a simple signal in the dominant orientation [6], we can define the *local phase* as the phase of the dominant signal component.

To deal with characteristic phases in the dominant signal direction, we first synthesize responses from a filter in a direction, $\hat{\mathbf{n}}_z$, compatible with the local orientation.[5]

$$\hat{\mathbf{n}}_z = (\,\mathrm{Re}(\sqrt{\boldsymbol{z}})\;\mathrm{Im}(\sqrt{\boldsymbol{z}})\,)^t \tag{18}$$

The filters will be weighted according to the value of the scalar product between the filter direction, and this orientation compatible direction.

$$w_k = \hat{\mathbf{n}}_k^t \hat{\mathbf{n}}_z \tag{19}$$

Thus, in each scale we synthesize one odd, and one even response projection as:

$$q_{e,i} = \sum_k q_{e,i,k}\,\mathrm{abs}(w_k) \tag{20}$$

$$q_{o,i} = \sum_k q_{o,i,k} w_k \tag{21}$$

This will change the sign of the odd responses when the directions differ more than π, but since the even filters are symmetric, they should always have a positive weight. In accordance with our findings in the 1D study (equations 7-9, 10-12), we now compute three phase channels, $c_{1,i}$, $c_{2,i}$, and $c_{3,i}$, in each scale.

The characteristic phase channels are shown in figure 7.[6] As we can see, the channels exhibit a smooth transition from describing the white regions in the test pattern (see figure 6) as lines, and as two edges. Also note that the phase statements actually give the phase in the dominant orientation, and not in the filter directions, as was the case for CPCT [4].

7 Local Orientation and Characteristic Phase

An orientation image can be be gated with a phase channel, $c_n(\mathbf{x})$, in the following way:

$$\boldsymbol{z}_g(\mathbf{x}) = \begin{cases} 0 & \text{if } c_n(\mathbf{x}) = 0 \\ \dfrac{c_n(\mathbf{x}) \cdot \boldsymbol{z}(\mathbf{x})}{\|\boldsymbol{z}(\mathbf{x})\|} & \text{otherwise} \end{cases} \tag{22}$$

[5] Since the local orientation, $\boldsymbol{z}$, is represented with a double angle argument, we could just as well have chosen the opposite direction. Which one of these we choose does not really matter, as long as we are consistent.

[6] The magnitude of lines this thin can be difficult to reproduce in print. However, the magnitudes in this plot *should* vary just like in figure 5.

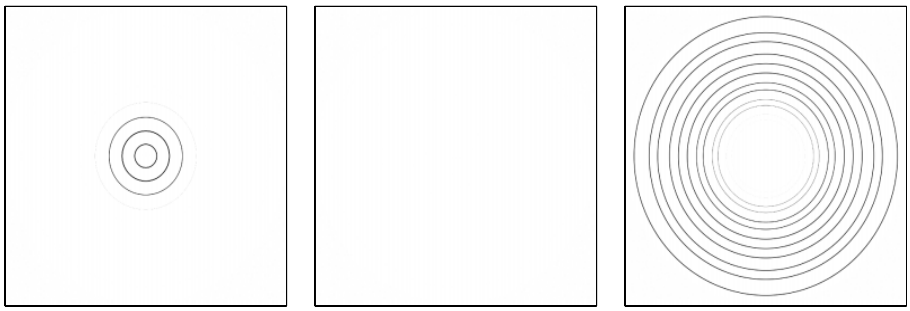

Fig. 7. Characteristic phase channels in 2D. ($\rho_i = \pi/4$)
Left to right: Characteristic phase channels $c_{1,i}$, $c_{2,i}$, and $c_{3,i}$, according to equations 10-12 ($\alpha = 2$).

We now do this for each of the characteristic phase statements $c_{1,i}(\mathbf{x})$, $c_{2,i}(\mathbf{x})$, and $c_{3,i}(\mathbf{x})$, in each scale. The magnitude of the result is shown in figure 8. Notice for instance how the bridge near the center of the image changes from being described by two edges, to being described as a bright line, as we move through scale space.

8 Concluding Remarks

The strategy of this approach for low-level representation is to provide sparse, and reliable statements as much as possible, rather than to provide statements in all points.

Traditionally, the trend has been to combine or merge descriptive components as much as possible; mainly to reduce storage and computation. As the demands on performance are increasing it is no longer clear why components signifying different phenomena should be mixed. An edge is something separating two regions with different properties, and a line is something entirely different.

The use of sparse data representations in computation leads to a mild increase in data volume for separate representations, compared to combined representations.

Although the representation is given in discrete scales, this can be viewed as a conventional sampling, although in scale space, which allows interpolation between these discrete scales, with the usual restrictions imposed by the sampling theorem. The requirement of a good interpolation between scales determines the optimal relative bandwidth of filters to use.

9 Acknowledgements

The work presented in this paper was supported by WITAS, the Wallenberg laboratory on Information Technology and Autonomous Systems, which is greatfully acknowledged.

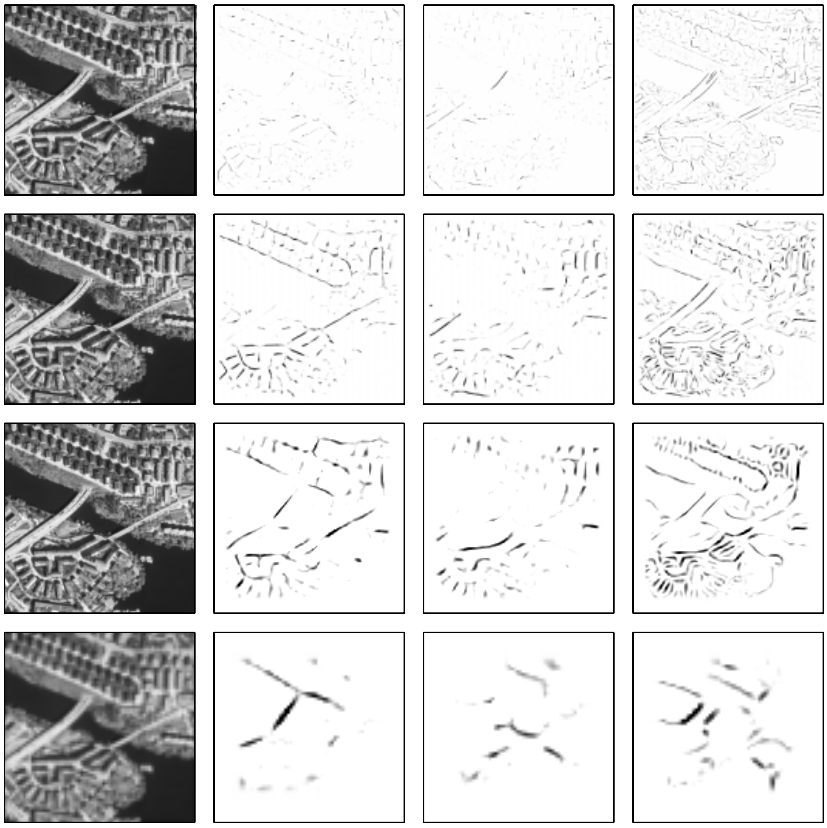

Fig. 8. Sparse feature hierarchy. ($\rho_i = \{\pi/2, \pi/4, \pi/8, \pi/16\}$)

References

1. Andrew Witkin. (1983) *Scale-space filtering.* In Proc. International Joint Conference on Artificial Intelligence, Karlsruhe, 1983.
2. Tony Lindeberg. (1994) *Scale-space Theory in Computer Vision.* Kluwer Academic Publishers 1994. ISBN 0792394186
3. Field, D. J. (1994) *What is the goal of sensory coding?* Neural Computation, 6:559-601, 1994.
4. Reisfeld, D. (1996) *The Constrained Phase Congruency feature detector: simultaneous localization, classification, and scale determination.* Pattern Recognition letters 17(11) 1996 pp. 1161-1169
5. Bear, M. F. et al. (1996) *Neuroscience: Exploring the Brain.* Williams & Wilkins, ISBN 0-683-00488-3
6. Granlund, G. H., Knutsson, H. (1995) *Signal Processing for Computer Vision.* Kluwer Academic Publishers, ISBN 0-7923-9530-1
7. Knutsson, H., Andersson, M., Wiklund J. (1999) *Advanced Filter Design.* Proceedings of SCIA, 1999.
8. Granlund, G. (1999) *The complexity of vision.* Signal Processing 74, pp 101-126.