On Witkin Range of Phase Congruency

Anonymous CVPR submission

Paper ID 648

Abstract

This article identifies a scale manifestative concept in low-level vision, which we call Witkin range of phase congruency, and proposes a simple method for calculating this image feature. This concept is similar to the range of stability in Witkin's scale space filtering, but we define it in terms of the phase congruency among Gabor-type wavelets of different frequencies. The Witkin range of phase congruency leads to a representational and computational scheme for combining image information from multiple scales. In particular, it adds two new dimensions to the traditional edge representation produced by Canny edge detector, namely, the width and sharpness of the edge point. As a result, it combines the edge representation and region representation into an edged-region representation. In addition, this concept unifies two ubiquitous classes of visual phenomena, namely, geometric structures and stochastic textures, in a scale manifestative framework, which can account for the continuous transition from structures to textures in the process of image scaling or zooming. We illustrate our method by a number of experiments on natural images.

1. Introduction

Scale is one of the most important issue in vision. Visual phenomena in natural scenes can appear at a wide range of scales in images, because of the variabilities in object sizes, viewing distances, and camera resolution. Therefore, a meaningful interpretation of a natural image must be either scale invariant or scale manifestative. "Scale invariant" means that the interpretation will stay invariant under image scaling. "Scale manifestative" means that the interpretation has explicit scale parameters that follow simple and explicit transformations under image scaling.

There have been a number of multi-scale theories in vision, most notably, the scale space theory [21, 8] and the multi-resolution wavelet analysis [11]. It has been a common sense that we need to combine information from multiple scales, mainly because some visual phenomena such as edges can persist over a range of scales.

The scale persistency has long been observed. For instance, Marr [12] proposed the "coincidence assumption," which holds that only those features that spatially coincident at all scales are meaningful. Witkin, in his paper on scale space filtering [21], investigated the persistency of local maxima of Gaussian derivatives of 1D signal over scales, and explicitly identified the stability ranges of these local maxima in scale space. These ranges can then be translated into flat intervals as basic elements for representing the 1D signal. Witkin's idea has been extended to 2D by Lindeberg [8, 9] and other researchers. But the behavior of the maxima of Gaussian derivatives in 2D is much more complicated than 1D, so that tracing maxima in 2D can be difficult.

Parallel to scale space theory, the coincidence over scales has been extensively studied in the context of phase congruency of Fourier transform or Gabor wavelet transform. Morrone et al. [14] observed that image features appear at locations where the Fourier components of the image at different frequencies are maximally in phase. A phase congruency function is defined to measure the agreement among the phases at each position, and this function turns out to be equal to the local energy of the image [19, 16]. Kovesi [6] developed a computational method for phase congruency feature detection in 2D image, using log-Gabor wavelets to compute local phases and energies. The phase congruency function is defined on each pixel. Unlike Witkin's scale space filtering, it does not involve tracing over scales or frequencies, and is therefore simpler to implement in 2D.

The phase congruency function is elegant in terms of its relationship with image energy. However, it is not scale manifestative, in the sense that it does not tell us the range of frequencies over which the phases are congruent. In this paper, we introduce Witkin's idea of range of stability to the framework of phase congruency, and replace the phase congruency function by the phase congruency range or what we call Witkin range.

The Witkin range gives us a more informative description of image features. For instance, for an image structure such as an edge, the Witkin range can be translated into geometric scale parameters of the cross-section profile perpendicular to the edge elongation. In particular, the high frequency end of the Witkin range tells us the sharpness of the transition of the image intensity across the edge, whereas the low frequency end of the Witkin range tells us the width or breadth of the two flat regions on the two sides of the edge. In other words, the Witkin range enables us to not only detect the edge curves, but also recognize the edged-regions, so that we can form a representation that combines both edge-based concept and region-based concept.

In natural scenes, there are two ubiquitous classes of visual phenomena. One is geometric structures that can be represented by lines and regions. The other is stochastic textures that are often characterized by some feature statistics. Although these two types of patterns often appear distinctively different, they are actually intrinsically connected: the same group of objects can be perceived as either geometric structures or stochastic textures depending on the viewing distance and camera resolution. Due to this scaling connection, it is natural to believe that the visual system must estimate scale parameters explicitly, and trace the change of the scale parameters over the image scaling process that can be caused by the change of viewing distance. We shall show that the Witkin range provides us with crucial scale information for describing large scale geometric structures and small scale stochastic textures.

The Witkin range of phase congruency can be useful for edge representation, edge-based object recognition, tracking and matching, and texture recognition. It also sheds light on low-level vision theories such as sparse coding, meaningful alignment, and natural image statistics.

2. Background

2.1. Witkin stability range

A key motivation for Witkin [21] to propose his theory of scale space filtering is to combine visual information across different scales. In particular, he studied the stability of the spatial locations of local maxima of Gaussian derivatives of the image data over scales.

See Fig. 1 for an example. A 1D signal is taken from a slice of the image in Fig. 1.a. Let's denote this 1D signal by $u(x)$. Let $G_\sigma(x)$ be a Gaussian kernel function (or density function) centered at 0 with standard deviation σ. Let $u_\sigma = u * G_\sigma$ be the convolution of $u(x)$ with $G_\sigma(x)$. Fig. 1.c displays $u_\sigma(x)$ for a sample of scales $\sigma > 0$. For each $u_\sigma(x)$, we can find the local maxima of its first derivative $\partial u_\sigma(x)/\partial x$, or the zero-crossings of its second derivative $\partial^2 u_\sigma(x)/\partial x^2$. Fig. 1.b plots the contours of these zero-crossings in the scale space. Clearly, the zero-crossings persist over a range of scales, until two zero-crossings merge into a singular point. The range of persistence or stability depends on the widths of the underlying intervals. As a matter of fact, one can recover these intervals based on the stability ranges.

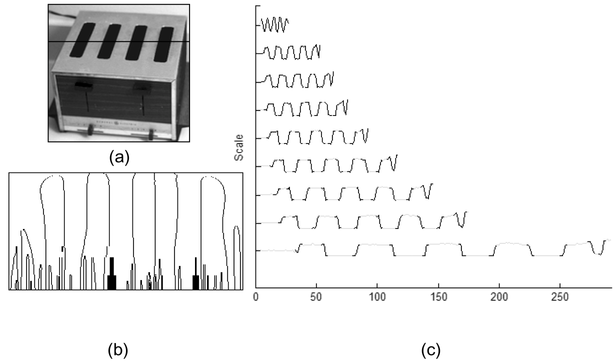


Figure 1. (a) A 1D signal is obtained as a horizontal slice of the toaster image. (b) The contour plot of the zero-crossings of the second derivatives in the joint spatial-scale domain. (c) The 1D signal at multiple resolutions, obtained by convolving the signal with Gaussian kernels and sub-sampling the signal.

Lindeberg [8, 9] applied similar ideas to 2D images. But the behavior of the maxima of Gaussian derivatives in 2D is much more complicated than 1D, so that tracing maxima in 2D can be difficult.

2.2. Phase congruency function

In contrast to the scale space filtering based on local derivative operators, the phase congruency theory started from global Fourier transform. Morrone et al. [14] observed that for a signal $u(x)$, the feature points correspond to those points where the Fourier waves at different frequencies have congruent phases. Specifically, let

$$u(x) = \int A(\omega)\cos(\omega x + \phi_\omega)d\omega \qquad (1)$$

be the Fourier representation of $u(x)$. The phase of frequency ω at a point x is $\omega x + \phi_\omega \bmod 2\pi$. Those x where $\omega x + \phi_\omega$ are congruent across ω are considered feature points. For instance, the top plot of Fig. 2 shows a periodic step function. The bottom plot displays several of its Fourier components $A(\omega)\cos(\omega x + \phi_\omega)$ (see eqn. (1)). Clearly, the edge points and the center points of the intervals correspond to those points where the Fourier waves of different frequencies are in phase.

A phase congruency function is defined as follows:

$$\varphi(x) = \max_{\phi \in [0, 2\pi]} \frac{\int A(\omega)\cos(\omega x + \phi_\omega - \phi)\hat{G}(\omega)d\omega}{\int A(\omega)\hat{G}(\omega)d\omega}, \qquad (2)$$

where $\hat{G}(\omega)$ is a window function in the frequency domain, e.g., a Gaussian kernel around a certain central frequency. Suppose the maximum is achieved at $\phi = \bar{\phi}(x)$. $\bar{\phi}(x)$ may be interpreted as the average phase across the frequencies

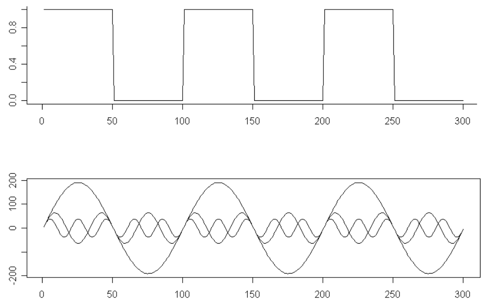


Figure 2. A periodic step function and its several Fourier components. The waves at different frequencies are in phase at edge points and the center points of the intervals.

covered by the window $\hat{G}(\omega)$, and $\varphi(x)$ measures the variation of phases within this window.

The phase congruency function (2) has an elegant connection with the local energy [19]. Specifically, let

$$G * u(x) = \int \hat{G}(\omega) A(\omega) \exp\{i(\omega x + \phi_\omega)\} d\omega,$$

where G is the complex filter, which consists of a pair of filters of quadrature phase, whose Fourier transform is $\hat{G}(\omega)$. Then $\varphi(x) = |G * u(x)|$ is the local energy, and $\bar{\phi}(x) = \arg[G * u(x)]$ is the local phase. So the points of maximum phase congruency correspond to points of maximum local energy.

Kevosi [6] defines a phase congruency function by pooling the information from a bank of log-Gabor filters at different scales and orientations, in the same spirit as function (2). Unlike scale space filtering, the phase congruency function is defined for each pixel without tracing local maxima.

2.3. Gabor filters and edge detection

A biologically motivated class of image elements are Gabor wavelets [2], which are rotated, dilated, and translated copies of the following Gaussian modulated sine and cosine waves [7]

$$G(x) \propto \exp\{-\frac{1}{8}(4x_1^2 + x_2^2)\}(e^{i\kappa x_1} - e^{\kappa^2/2}), \qquad (3)$$

where $x = (x_1, x_2)$. Let's denote a rotated, dilated, and translated copy of (3) by $G_{x,\omega,\theta}$, where x is the center, ω is the frequency of the sine and cosine waves, and θ is the orientation. We normalize the Gabor wavelets over scale so that $G_{x,s\omega,\theta} = sG_{x,\omega,\theta}$, in order to maintain

$$\langle f, G_{x,\omega,\theta} \rangle = \langle f_s, G_{x,s\omega,\theta} \rangle, \qquad (4)$$

where $f_s(x) = f(sx)$ is the scaled version of f. The Gabor filters can be replaced by other zero-mean filter pairs that form Hilbert transforms of each other.

The Gabor filters can be used as edge detectors [20]. For the dictionary of Gabor elements $\{G_{x,\omega,\theta}\}$, at each frequency ω, and at each pixel x, we find the optimal orientation $\hat{\theta} = \arg\max_\theta |\langle u, G_{x,\omega,\theta} \rangle|^2$, where $|\langle u, G_{x,\omega,\theta} \rangle|^2 = \langle u, G^{(0)}_{x,\omega,\theta} \rangle^2 + \langle u, G^{(1)}_{x,\omega,\theta} \rangle^2$, with $G^{(0)}_{x,\omega,\theta}$ and $G^{(1)}_{x,\omega,\theta}$ being cosine and sine components of $G_{x,\omega,\theta}$ respectively, and $\langle\rangle$ denoting inner product. Let $\hat{A} = |\langle u, G_{x,\omega,\hat{\theta}} \rangle|$, $\hat{\phi} = \arctan[\langle u, G^{(0)}_{x,\omega,\hat{\theta}} \rangle / \langle u, G^{(1)}_{x,\omega,\hat{\theta}} \rangle]$ be the magnitude (or energy) and phase at the maximal orientation respectively. We can write

$$[\nabla_\omega u](x) = (A_\omega(x), \theta_\omega(x), \phi_\omega(x)) = (\hat{A}, \hat{\theta}, \hat{\phi}) \qquad (5)$$

as a generalized version of the ubiquitous gradient operator ∇u.

A point (x, ω) is an edge-ridge point if

$$A_\omega(x) \geq A_\omega(x + t(\sin\theta_\omega(x), \cos\theta_\omega(x))), |t| < d, \qquad (6)$$

i.e., $A_\omega(x)$ is maximal along the normal direction of the orientation $\theta_\omega(x)$ within a neighborhood of length $2d$ [1]. Edges and ridges can be discriminated by the corresponding phases [20].

3. Witkin range of phase congruency

In this section, we explain the basic idea of the Witkin range. We also explain that it can be translated into an edged-region representation of the image. After that, we give a precise definition of the Witkin range, and illustrate the edged-region representation by some examples.

3.1. Edged-region representation

Fig. 3 illustrates a fundamental observation. The plot on top shows a horizontal slice of an image of vertical bar $u(x)$. The second and the third plots display the magnitude and phase of $\nabla_\omega u(x)$ on this slice, where each curve corresponds to a frequency ω. It is evident that an edge-ridge point x (i.e., a local maximum in magnitude curves) can exist over a range of frequencies $(\omega_0(x), \omega_1(x))$, and within this range, the phase and orientation of $\nabla_\omega u(x)$ remains constant for $\omega \in (\omega_0(x), \omega_1(x))$. For edge points, the magnitude of $\nabla_\omega u(x)$ also remains constant (subject to discretization error). In the bottom plot of Fig. 3, we trace the two edge points over scale $s \propto 1/\omega$. At a certain frequency or scale, the two edge points merge into a ridge point.

Here comes the foundation of this paper: for an edge point x, the range $\omega \in (\omega_0(x), \omega_1(x))$ in frequency domain can be translated into spatial domain parameters about the cross-section profile of the edge. Specifically, the profile of an edge is along the direction that is perpendicular to the edge elongation, and it can be modeled by a step function blurred by a Gaussian kernel [3], whose bandwidth reflects

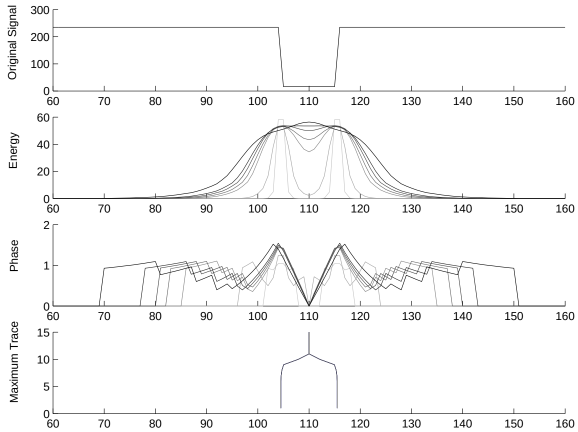
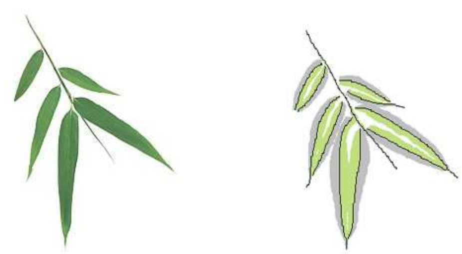


Figure 3. Constancy of positions and phases of local energy maxima across frequencies. The top plot depicts a horizontal slice of an image of a vertical bar. The next two plots show local energy and phase, where each curve corresponds to a frequency. The bottom plot shows that the maxima corresponding to two edge points merge into one ridge point over frequencies or scales.

(a) input image (b) edged-regions (c) ridged-region

Figure 4. Image interpretation by edged- and ridged- regions.



(a) input image (b) edged-regions

Figure 5. Edged-region representation.

the sharpness of the edge. For the Witkin range, the high frequency end $\omega_1(x)$ tells us how sharp the intensity transition is across the edge. The low frequency end $\omega_0(x)$ tells us how wide the two flat pieces of the step function can extend. As x runs on the one dimensional edge curve, the resulting cross-section profile sweeps an edged-region with the edge curve being the mid-axis.

See Fig. 4 and Fig. 5 for illustrations of edged-region representation. An edged-region is composed of two segments of smooth sub-regions, colored by grey and green respectively in the above two figures. The two sub-regions are separated by an edge curve. The widths or breadths of the two sub-regions are decided by the low frequency end $\omega_0(x)$ of the Witkin range. The sharpness of the segmentation is decided by the high frequency end $\omega_1(x)$. That is, by combining the Gabor edge information across frequencies, we essentially perform a local image segmentation, where the Gabor filters at different frequencies explore the two sub-regions being segmented by the edge. This enables us to not only detect the edge, but also recognize the edged-region. The edged-region representation combines both the edge concept and the region concept, which are two most prominent representations in low- and mid- level vision.

Viewed in frequency domain, an edged-region is a composition of Gabor wavelets across the Witkin range of frequencies, or an edged-region spans a range of frequencies.

The Witkin range transforms in a simple way during the scaling process. When we zoom out the image by a factor of s, the Witkin range (ω_0, ω_1) will be scaled to $(s\omega_0, s\omega_1)$. So the edged-region becomes thinner and sharper by a factor of s.

An interesting observation is that, due to finite resolution of the image, the Witkin range $(s\omega_0, s\omega_1)$ will eventually go beyond the frequency limit of the camera resolution as s increases, and the edged-region will be shredded and leaked out. This can explain the transition from geometric structures to stochastic textures, as we will study later.

3.2. Definition of Witkin range of phase congruency

In 2D images, tracing the edge points can be difficult. Recall that the phase congruency function (2) is defined for each pixel without tracing. Similarly, we can also define the Witkin range of phase congruency without tracing the edge point. The following is our version of definition.

The Witkin range is defined on scale-maximum edge points. An edge point $(\hat{x}, \hat{\omega})$ is a scale-maximum edge point if 1) $(\hat{x}, \hat{\omega})$ is an edge point at frequency $\hat{\omega}$ in the sense of inequality (6); 2) it is also a local maxima in scale or frequency domain:

$$A_{\hat{\omega}}(\hat{x}) \geq A_{\omega}(\hat{x}),\ \forall \omega \in (\hat{\omega} - \delta, \hat{\omega} + \delta), \tag{7}$$

i.e., a small neighborhood of $\hat{\omega}$. 3) Its phase $\phi_{\hat{\omega}}(\hat{x})$ is dominated by sine component of the Gabor filter. The scale-maximum edge points transform in a simple way during image scaling.

For a scale-maximum edge point $(\hat{x}, \hat{\omega})$, we define its Witkin range (ω_0, ω_1) as a continuous range of frequencies ω around $\hat{\omega}$ so that $\nabla_{\omega} u(\hat{x})$ has almost constant magnitude, phase and orientation. Specifically, let

$$\Omega = \{\omega : \nabla_{\omega} u(\hat{x}) \in \partial(\nabla_{\hat{\omega}} u(\hat{x}))\}, \tag{8}$$

where $\partial(\nabla_{\hat{\omega}} u(\hat{x}))$ is a small neighborhood of $\nabla_{\hat{\omega}} u(\hat{x})$. Recall that $\nabla_\omega u(x) = [A_\omega(x), \theta_\omega(x), \phi_\omega(x)]$, i.e., magnitude (energy), orientation, and phase, see eqn. (5). In our implementation, $\nabla_\omega u(\hat{x}) \in \partial(\nabla_{\hat{\omega}} u(\hat{x}))$ if

$$\begin{aligned} A_\omega(\hat{x})/A_{\hat{\omega}}(\hat{x}) &\geq f_A, \\ |\theta_\omega(\hat{x}) - \theta_{\hat{\omega}}(\hat{x})| &\leq \epsilon_\theta, \\ |\phi_\omega(\hat{x}) - \phi_{\hat{\omega}}(\hat{x})| &\leq \epsilon_\phi, \end{aligned} \tag{9}$$

that is, the magnitude $A_\omega(\hat{x})$ should be within a factor (e.g., $f_A = .8$) of $A_{\hat{\omega}}(\hat{x})$, the orientation and phase should be close to those of $(\hat{\omega}, \hat{x})$ (e.g., $\epsilon_\theta = \pi/12$, and $\epsilon_\phi = \pi/6$). Then

$$\begin{aligned} \omega_0 &= \max\{\omega \leq \hat{\omega}, \omega \notin \Omega\}, \\ \omega_1 &= \min\{\omega \geq \hat{\omega}, \omega \notin \Omega\}. \end{aligned} \tag{10}$$

We can translate (ω_0, ω_1) to Witkin width $s_0 \propto 1/\omega_0$, and Witkin sharpness $s_1 \propto 1/\omega_1$. The proportion factor can be chosen so that when applied to a bar structure with two parallel edges (see Fig. 3), the Witkin width should agree with the half-width of the bar. It is clear that this proportion factor depends on f_A, ϵ_θ, and ϵ_ϕ in (9).

Our definition of Witkin width generalizes the traditional definition of width for bar structures to any geometric structures. For instance, Fig. 6 shows the edged-region plots of a triangle and a circle. Here we only plot the darker subregion of the edged-region. Specifically, at each edge point $\hat{x}$, we plot a black bar (1 pixel wide) of length s_0, i.e., the Witkin width of this edge point. The bar is perpendicular to the edge elongation, and extends to the darker segment of the edged-region. Then the bars for all the edge points make up the darker sub-region of the edged-region.

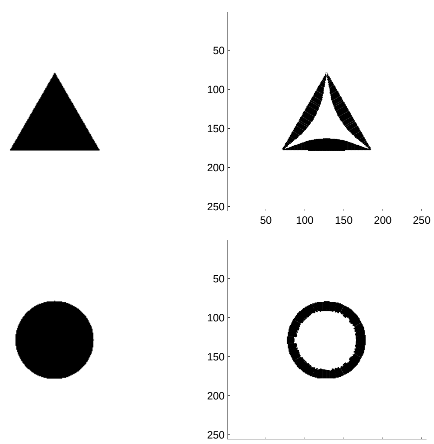


Figure 6. Edged-region representations of triangle and circle. Only the dark sides of the edges are plotted.

Fig. 7, Fig. 8, and Fig. 9 show the edge-region representations of three natural scene images. Note that in our implementation, there is an upper bound on the scale of the Gabor filters (or equivalently, a lower bound on the frequency), so there is an upper bound on Witkin width. That is why at some points, the Witkin edges are not wide enough.

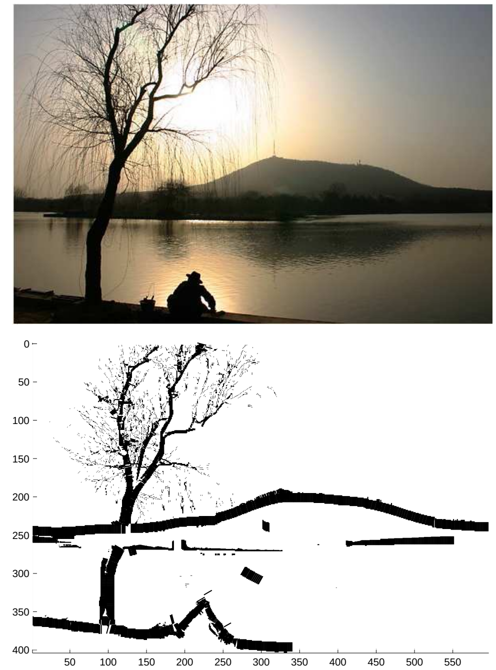


Figure 7. A natural scene image and its edged-region representation.

The above definition of Witkin range is clearly scale manifestative. If we scale the image by a factor, then the Witkin width and sharpness should scale in the same way, as long as they are above the camera resolution.

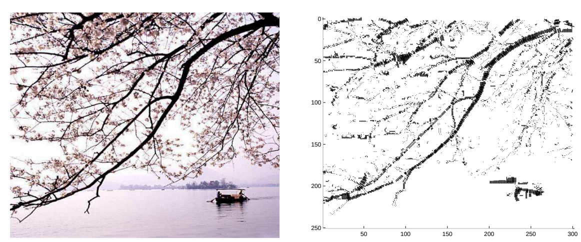


Figure 8. A natural scene image and its edged-region representation.

3.3. Ridged-region representation

The top plot of Fig. 3 displays the cross-section profile of a bar structure. A key point is that a bar structure is not only described by the width of the central flat interval, but also the widths of two flat wings on the two sides of the central interval. So a ridged-region should have three segments, corresponding to the central piece and the two

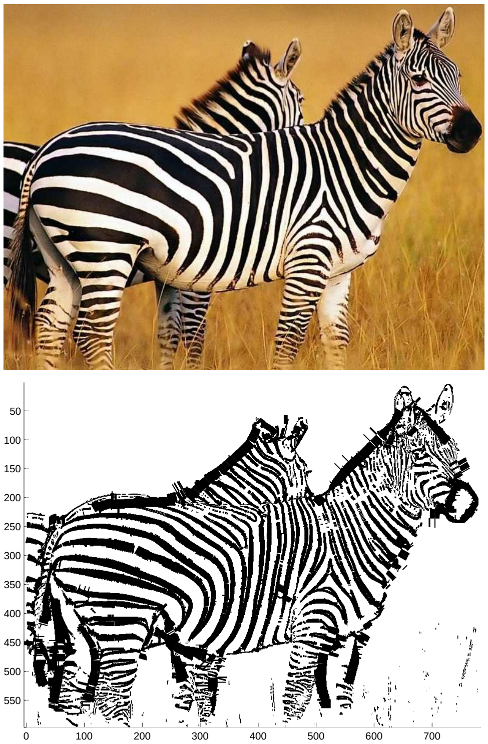


Figure 9. A natural scene image and its edged-region representation.

wings respectively. For an ridge point, we can also define a Witkin range that can be translated to the width of the central piece and the width of the two flat wings. Due to space limit, we shall not elaborate on this. We just want to point out that the width of the central piece is intimately related to the Witkin width of the two parallel edges of the bar structure. But the Witkin width of an edge is a far more general definition, because many edges are not edges of a bar structure. For instance, the base line of the triangle in Fig. 4, or the edge on the circle in Fig 6, or the shorter side of a rectangle.

The width of the edge structure and the width of the two flat wings of the bar structure go beyond the important work of Lindeberg on scale selection for edges and bars [10]. It appears that Lindeberg's scales correspond to the sharpness of the edge and the width of the central piece of the bar.

4. Unifying structures and textures

Geometric structures and stochastic textures are often treated separately in computer vision. Structures are usually obtained by edge detection and image segmentation, while textures are mostly characterized by feature statistics such as histograms of filter responses [4]. However, structures and textures are intrinsically connected by image scaling. Fig. 10 displays a sequence of images of an ivy wall of leaves taken at increasingly far distances. At near distance, the geometric shapes of individual leaves are perceptible. But as the viewing distance increases, the image becomes more complex and the individual shapes become imperceptible, and the image can only be described by a collective texture summary.

Figure 10. A sequence of ivy wall images taken at increasingly far distances.

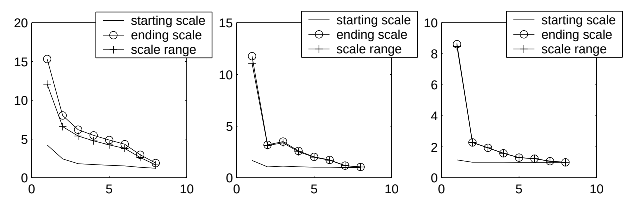


Figure 11. Conditional expectations of Witkin width, sharpness, and range over scale. Plain curve shows starting scale or Witkin sharpness, circled curve shows ending scale or Witkin width, and crossed curve shows the difference between starting and ending scales, or the size of Witkin range. Each plot is conditioned on an interval of magnitudes of scale-maximum edge points. The magnitudes of the plots are in increasing order from left to right.

This suggests that geometric structures and stochastic textures should be treated in a unified framework. The distinction between structures and textures is an artificial one, because the transition from structures to textures is a continuous process caused by the continuous image scaling or zooming. It is therefore desirable to have a scale manifestative quantity to trace this transition.

It has been a mystery how human being perceives the wide variety of texture patterns. Julesz [5], in his study of human texture perception over nearly three decades, proposed two famous conjectures. The first conjecture is about texture statistics, and Julesz proposed co-occurrence statistics of image intensities. The second conjecture is about textons, which are considered as basic elements for texture perception. We believe that image scaling holds the secret to this puzzle, and scale manifestative quantities must be a crucial ingredient in texture perception.

This can be seen even more evidently in Fig. 12, where the texture surfaces of flower leaves and pebbles are slant surfaces that appear in perspective. Clearly, our perception of the texture surfaces is not homogenous and we perceive a gradual change over distance.

The Witkin range of phase congruency can be used to trace the transition from structure to texture, as well as the change of texture information over distance.

We pool the following scale manifestative texture statistics. Let $(\hat{x}, \hat{\omega})$ be a randomly selected scale-maximum edge point. Let A be the magnitude of $\nabla_{\hat{\omega}} u(\hat{x})$. Let (s_1, s_0) be its Witkin sharpness and width defined in previous section. We experiment with the following three statistical properties. $E[s_0|A]$, $E[s_1|A]$, and $E[s_0 - s_1|A]$, that is, the conditional expectations of Witkin sharpness, width, and range. These three statistics can be estimated as follows. We divide the range of A into several intervals. We collect the scale-maximum edge points whose magnitudes fall into each interval, and then estimate the conditional expectations for this interval by corresponding averages.

Fig. 11 shows the change of conditional expectations over the scaling process. We choose 8 images of the ivy wall taken at increasingly far distances. We divide the magnitudes of scale-maximum edge points into three intervals. In Fig. 11, the three plots correspond to the three intervals of magnitudes in increasing order, from left to right. We can see that as the viewing distance increases, the Witkin width, sharpness and range decrease in general.

These plots also trace the transition from structures to textons and texture statistics. At near distance, we see relatively large edged-regions. As the distance increases, the edged-regions become smaller. This roughly correspond to the texton regime in Julesz's second conjecture. If the viewing distance increases still further, the edged-regions will be smaller than the pixel resolution, and they will be shredded and leaked out. Then there are no significant alignments among filter responses, and we may just pool some marginal statistics from the filter responses [4], since the joint patterns have been largely destroyed by image scaling. This roughly corresponds to the regime of texture statistics in Julesz's first conjecture.

The Witkin width also indicates the size of the neighborhood that we should use to pool the texture statistics. The larger the Witkin width, the larger the local window should be for spatial pooling.

The above statistics are also crucial for perceiving slant surfaces such as those in Fig. 12. Fig. 13 show the change of conditional expectations of Witkin width, sharpness, and range over vertical axis of the two images. It is possible that such statistics can be used for re-constructing 3D information of slant texture surfaces.

Figure 12. Slant texture surfaces of flower leaves and stones.

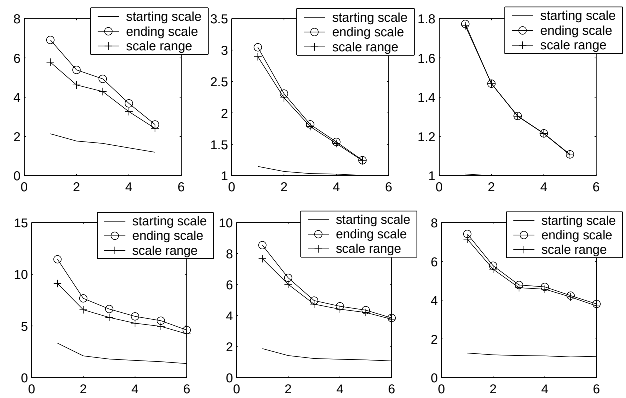


Figure 13. Conditional expectations of Witkin width, sharpness and range over the vertical axis in the two images in Fig 12.

5. Discussion

5.1. Contributions and open ends

The following are contributions of this article.

1) Identify and define the Witkin range for phase congruency, as a substitute for scale space tracing and phase congruency function.

2) Define the Witkin width of edge point, and propose the edged-region (as well as ridged-region) representation that combines both edge concept and region concept.

3) Study geometric structures and stochastic textures in a unified scale manifestative framework, and define a set of scale manifestative texture statistics.

The following are two major open ends of our work.

1) The current version of Witkin range may not be theoretically or empirically superior to other possible alternatives. We hope this work will stimulate more researchers to

experiment with this concept and search for better versions.

2) There are other image structures such as roofs, ramps, as well as topological structures such as corners and junctions. The geometric scale parameters of these structures should also be estimated based on similar ideas.

5.2. Potential applications

1) Edge feature. Our method is not in competition with Canny [1] or other edge detection methods. Instead, it equips each edge point with two important scale parameters, namely, sharpness and width.

2) Edge-based object recognition. For instance, for an object like a tree, the Witkin width is useful for identifying tree trunk, branches and twigs, without resorting to sophisticated region-based analysis.

3) Tracking and matching. In real life, objects can change distances from the viewer rapidly, e.g., a ball is coming, a dog is running away, or the scene outside the window of a moving train. The changes of Witkin ranges help us perceive the change of viewing distances.

4) Texture recognition and shape from texture.

5.3. Connections to other vision theories

1) Sparse coding. Olshausen and Field [15] proposed sparse coding as a strategy for V1. Our work suggests that the sparse coding elements are compositions of phase-congruent Gabor wavelets or edged-regions.

2) Meaningful alignment. Moisan, Desolneux, and Morel [13] proposed meaningful alignment as a statistical principle for perceptual grouping. Our work can be considered as identifying meaningful alignment over scales or in frequency domain.

3) Natural image statistics. Portilla and Simoncelli [17] proposed a class of joint statistics of filter responses to characterize texture patterns. The Witkin range can be considered a scale explicit characterization of the joint distribution. Ruderman and Bialek [18] studied the scaling of image statistics of natural scenes. The Witkin range statistics are worth of being investigated.

References

[1] J. Canny. "A computational approach to edge detection", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, **8**:679–698, 1986. 3, 8

[2] 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

[3] J.H. Elder and S. W. Zucker, "Local scale control for edge detection and blur estimation", *IEEE Trans. PAMI*, vol. 20, no. 7, 699-716, 1998. 3

[4] D. J. Heeger and J. R. Bergen, "Pyramid Based Texture Analysis/Synthesis", *Computer Graphics Proc.*, pp. 229-238, 1995. 6, 7

[5] B. Julesz, "Textons, the elements of texture perception and their interactions", *Nature*, Vol. 290, pp. 91-97, 1981. 6

[6] P. Kovesi, "Image Features from Phase Congruency", *Videre: Journal of Computer Vision Research*, Vol 1, 1999. 1, 3

[7] T. S. Lee, "Image Representation Using 2D Gabor Wavelets", *IEEE Trans. PAMI*, 10, 959-971, 1996. 3

[8] T. Lindberg, "Effective Scale: A Natural Unit for Measuring Scale-Space Lifetime," *IEEE-PAMI*, v.15 n.10, p.1068-1074, 1993. 1, 2

[9] Lindeberg, "Detecting salient blob-like image structures and their scales with a scale-space primal sketch: A method for focus-of-attention", *International Journal of Computer Vision*, vol. 11, pp. 283–318, 1993. 1, 2

[10] Lindeberg, "Feature detection with automatic scale selection". *International Journal of Computer Vision*, vol 30, number 2, pp. 77–116, 1998. 6

[11] S. Mallat, "A Theory of Multiresolution Signal Decomposition: the Wavelet Representation", *IEEE Trans. PAMI*, 11(7):674-693, 1989. 1

[12] D. Marr, *Vision*, W. H. Freeman and Company, 1982. 1

[13] L. Moisan, A. Desolneux, and J.-M. Morel, "Meaningful Alignments", *Int'l J. Computer Vision*, vol. 40, no. 1, pp. 7-23, 2000. 8

[14] M. C. Morrone, J. Ross, D. C. Burr and R. A. Owens. Mach bands are phase dependent. Nature, 324(6094):250-253, 1986. 1, 2

[15] B. A. Olshausen and D. J. Field, "Emergence of Simple-cell Receptive Field Properties by Learning a Sparse Code for Natural Images," *Nature*, Vol. 381, pp. 607-609, 1996. 8

[16] P. Perona and J. Malik. Detecting and localizing composite edges in images. Procs. ICCV, Osaka, 1991. 1

[17] J. Portilla and E.P. Simoncelli, "A parametric texture model based on joint statistics of complex wavelet coefficients". *Int'l Journal of Computer Vision*, 40(1):49-71, October, 2000. 8

[18] D.L. Ruderman and W. Bialek, "Statistics of Natural images: Scaling in the Woods", *Phy. Rev. Lett*, 73, 1994. 8

[19] S. Venkatesh and R. Owens, "An energy feature detection scheme," *International Conference on Image Processing*, 553-557, 1989. 1, 3

[20] Z. Wang and M. Jenkin, "Using complex Gabor filters to detect and localize edges and bars," in *Advances in Machine Vision*, 32, 151-170, 1992. 3

[21] A. Witkin, "Scale-space filtering", In Proc. International Joint Conference on Artificial Intelligence, Karlsruhe, 1983. 1, 2