International Journal of Computer Vision 31(2/3), 159–168 (1999)

The Structure of Locally Orderless Images

JAN J. KOENDERINK AND ANDREA J. VAN DOORN
Helmholtz Instituut, Universiteit Utrecht, Buys Ballot Laboratory, Princetonplein 5, 3584CC Utrecht, The Netherlands
j.j.koenderink@phys.uu.nl

Abstract. We propose a representation of images in which a global, but not a local topology is defined. The topology is restricted to resolutions up to the extent of the local region of interest (ROI). Although the ROI's may contain many pixels, there is no spatial order on the pixels within the ROI, the only information preserved is the histogram of pixel values within the ROI's. This can be considered as an extreme case of a textel (texture element) image: The histogram is the limit of texture where the spatial order has been completely disregarded. We argue that locally orderless images are ubiquitous in perception and the visual arts. Formally, the orderless images are most aptly described by three mutually intertwined scale spaces. The scale parameters correspond to the pixellation ("inner scale"), the extent of the ROI's ("outer scale") and the resolution in the histogram ("tonal scale"). We describe how to construct locally orderless images, how to render them, and how to use them in a variety of local and global image processing operations.

Keywords: scale space, histograms, segmentation, transparency

1. Introduction

In practice one often uses the histograms of pixel values for regions of interest (ROI's) of *limited size*. This is because the scope of images often includes a variety of objects that one would prefer to handle separately rather than pooled (*e.g.*, the blue sky, foliage, a human face, . . . , in a typical holiday snapshot). In such cases the histogram becomes a function of position and one really deals with a *histogram–valued image*. Such images depend on the resolution of the original image (*e.g.*, pixel size), on the size of the regions of interest over which the histograms are evaluated and on the bin–width of the histogram. Thus these are really very complicated objects that are drawn from (at least) a three–parameter family of such entities. One wonders whether this family *as a family* might have interesting and possibly useful structure?

Some interesting features of such representations are already intuitively evident. Suppose we replace local image structure with local histograms. This essentially means that we discard the precise localization of individual image elements, that is the *order* (or *topology*) at the fine level whereas one retains it at the coarse level. In this sense histogram images are "locally orderless images". Such images are interesting because—although the spatial resolution has been much reduced—*some* vestige of the original high resolution has been retained because the tonal resolution has been fully preserved. For instance images consisting of very thin black and white stripes are different from gray images (these would be indiscriminable with local averaging (blurring) instead of disordering) but horizontal cannot be discriminated from vertical stripes. In this sense the locally orderless images are like spatial distributions of local *texture* patches. Indeed, the histogram is arguably the simplest texture description, the one where the spatial character of the texture is fully (instead of only partly as in most textural representations) suppressed.

Many of the impressionists paintings are reminiscent of locally orderless images when we disregard the *touches* that show up the artist's facture rather than reveal information concerning the depicted scene (Homer 1964). That is to say, at the "pixel" level the detail reveals the brush strokes, rather than the leaves of the foliage of some depicted tree. Yet differently colored brush strokes *do* reveal information at the level of resolution of the leaves, even though this information is not of a *spatial* nature. Had the artist chosen to paint uniform areas filled with the average brush stroke color the painting would not merely *look* different, but actually contain *less information* despite the fact that no spatial structure was ever painted at the scale of the brush strokes. Such images can easily be "read" by the human observer but appear not to be used in conventional image processing. (Perhaps one should remember that at their time of origination such paintings were considered extremely "difficult", offensive even: They were considered as possible causes of miscarriage in pregnant women.)

Instances of locally orderless *perceptions* are quite frequently encountered in various contexts. Cases of *amblyopia* ("lazy eye") have been described (Hess 1982) in which the observer is able to distinguish fine black and white stripes from uniform gray but cannot distinguish vertical from horizontal stripes or read text at a similar level of resolution. This condition has been termed "tarachopia" or scrambled vision. It seems likely that the peripheral visual field of *normal* observers has a similar locally orderless structure (Metzger 1975), and so has the central visual field for finest details (Helmholtz 1866). That such cases are *typical* of normal perception, rather than the exception, has been forcefully argued on phenomenological grounds by Ruskin (1857 and 1873), for instance (Ruskin 1873):

> *"Go to the top of Highgate Hill on a clear summer morning at five o'clock, and look at Westminster Abbey. You will receive an impression of a building enriched with multitudinous vertical lines. Try to distinguish one of these lines all the way down from the next to it: You cannot. Try to count them: You cannot. Try to make out the beginning or end of any of them: You cannot. Look at it generally, and it is all symmetry and arrangement. Look at it in its parts, and it is all inextricable confusion."*

2. Construction of locally orderless images

Locally orderless images are characterized by a rather large number of parameters. In order to avoid confusion we construct locally orderless images in a number of intuitively natural steps, thus introducing the parameters in a manner that elucidates their meaning. We begin the discussion with a small number of initial notions:

— a *scene* has the potential to yield a point observation when probed with a detector. When we apply the detector at many positions we obtain an *image* of the scene. We will denote the scene $\mathcal{S}(\mathbf{r})$;

— a *point operator* is a linear detector with weight function

$$G_0(\mathbf{r};\sigma) = \exp(-\mathbf{r}\cdot\mathbf{r}/2\sigma^2)/(\sigma\sqrt{2\pi})^d. \qquad (1)$$

Here the parameter σ defines the *resolution* or size of the point operator. The number d denotes the dimension of the scene domain. We will denote the image

$$I(\mathbf{r};\sigma) = \mathcal{S}(\mathbf{r}) \otimes G_0(\mathbf{r};\sigma). \qquad (2)$$

(The infix operator denotes convolution.)

— an *aperture* (also called window, or region of interest (ROI)) is a spatially distributed weight

$$A(\mathbf{r};\mathbf{r_0},\alpha) = \exp(-(\mathbf{r}-\mathbf{r_0})\cdot(\mathbf{r}-\mathbf{r_0})/2\alpha^2). \qquad (3)$$

(Notice that the maximum value is unity whereas the total weight evaluates to $(\alpha\sqrt{2\pi})^d$. We set the amplitude, rather than the total weight, to unity because the aperture function represents a "soft" characteristic (zero or unity) function.)

Notice that we don't consider certain *images*, but rather certain *scenes* as the fundamental entities: Images are *operationally derived* from scenes whereas scenes are (implicitly and necessarily incompletely) *revealed* through images. More specifically, "scenes" are not to be confused with data or observations, they are categorically different. A scene simply is the potential to be observed but only actual observations are data. In a way "the scene is never seen" (it forever remains an "ideal" or "imaginary" entity, whose potential to yield novel data is never exhausted), we may only know (see) images of scenes. Yet one customarily identifies the (ideal!) scene with "reality". Images are then taken to reveal certain aspects—defined through our understanding of the nature of the imaging process—of reality. In a sense an image is an answer to a question we pose to nature, its *potential* meaning is defined through our understanding of the measurement process, its *actual* meaning through the specific probing. This explains how reality can be simultaneously meaningless in general, yet meaningful to a given observer, with different meanings to different observers—without any contradiction.

Images of scenes are necessarily of finite resolution. This is because the total amount of data obtained from any measurement is always finite. When observed with point operators of size σ we say that the resolution or "inner scale" (the notion of "outer scale" will be introduced later) of the image is σ. We may observe the scene at many levels of resolution, even *simultaneously*. (See figures 1, 2, and 3.) The one–parameter family of images with resolution as parameter is known as *scale space* (Koenderink and van Doorn 1978, Koenderink 1984, Lindeberg 1994, Florack 1997). Of course no observer is able to sample more than some finite region of scales, every image is embedded in "mystery" on all sides. With mystery we mean that one can always imagine a wider scope revealing novel external structure or a microscope that would reveal novel internal structure. The term is due to Ruskin (1857).

Although we write "$S(\mathbf{r})$" for the scene's spatial structure this has to be taken *cum grano salis*. We do in no way imply that S should be regarded as some (known) function of position! The notation is a purely symbolic one. For given $\mathbf{r}$ (even this notion is problematic since infinite precision is implied) it is not the case that some number (or vector, or record, *v.i.*) "$S(\mathbf{r})$" exists. All that can be measured is a sample via some operator (of finite size). It is not so much that S is possibly not differentiable or discontinuous or otherwise nasty or ill defined, but rather that we don't see how such notions might even apply. Fortunately we can be certain that the *images* are C^∞ (even analytic) though. The images have finite resolution, whereas the notion of resolution doesn't even *apply* to the scene. Thus the expression $\mathcal{S}(\mathbf{r}) \otimes G_0(\mathbf{r}; \sigma)$ is purely symbolic and not intended as a recipe for actual calculation: The ob-

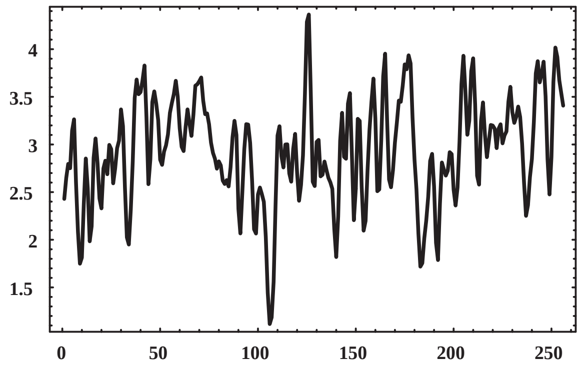

Fig. 1. A random signal. The signal was created as a list of 512 samples from a normal distribution with unit variance and mean adjusted so as to render the signal non–negative. In the scale space representation (figure 2), like here, we have chopped off 128 samples from head and tail. This avoids artifacts due to the blurring in the presence of hard image boundaries.

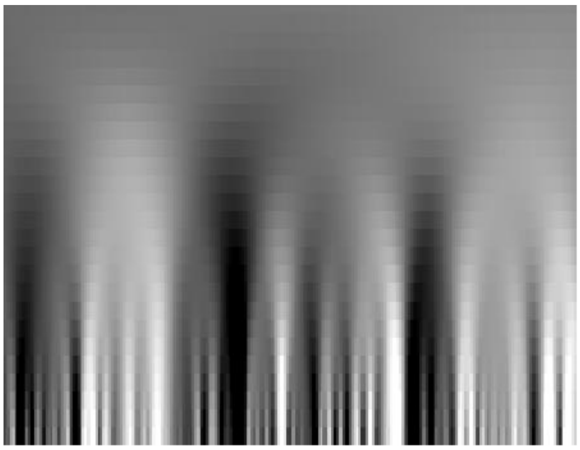

Fig. 2. The scale space based on the signal depicted in figure 1. The signal is progressively coarsegrained as one gets higher in this picture. The bottom row represent the original (unblurred) signal. Any row represents a valid representation of the signal, it should be understood that *any* signal is necessarily limited in resolution.

servation is the *output of a detector*, not the *result of a calculation*. Indeed, no one would know how to substitute $\mathcal{S}(\mathbf{r})$ in the expression. A good reference on these issues is the recent monograph by Florack (1997).

Notice that we write the point operator G_0 with suffix zero. This is because it is often useful to observe higher order properties of the scene. This can be done via derivative operators (Koenderink and van Doorn 1992). For instance, the first order directional derivatives of the point operator yield "edge operators" that let us observe spatial gradients of the scene. All that is written here applies equally well to such derivative images. Notice that derivatives are not com-

Fig. 3. Plot of the modes of the blurred versions of the signal shown in figure 1. Notice how blurring progressively (upwards in the graph) simplifies the signal, that is to say, modes are destructed—not created—as blurring is increased. This is the "causal structure" of scale space. Higher dimensions are slightly more complicated, though not essentially different. Notice that the scale space structure induces a natural hierarchy in the set of modes.

puted but observed. Indeed, we have no idea of how to conceive of the derivatives of a scene other than via observation.

We assume that the observation of the scene introduces a natural *spatial order* or topology. For instance, when a CCD–camera is used we typically obtain observations ordered according to a rectangular matrix. In that case the dimension $d = 2$. Notice that the order is due to the observation, in this case the structure of the CCD–chip and its relation to the optics. The observation also defines the nature of a single datum. The observation may be a scalar (flux collected in a CCD pixel, preferably converted to the irradiance at a position on the image plane or radiance in a direction of the field of view), a vector (as with a RGB–color–CCD camera) or a "property list" or record (list of incommensurable data as often occurs in medical image processing or remote sensing applications). The nature of the datum is a *convention*: Some physical change—say charge built up in a CCD element—is interpreted as a sign of something else—say irradiance at the surface of the chip. In the example the nature of the datum is defined in the manual of the CCD camera.

The apertures are to be considered as similar to fuzzy characteristic functions. Thus they will not be applied as regular weights on the images as such, they may weigh the influence of parts of the image on the results of computations limited by the apertures. Simple examples will be offered later. That one cannot simply take $A(\mathbf{r}; \mathbf{r_0}, \alpha)\, I(\mathbf{r}; \sigma)$ as a "local subimage" is clear from the fact that then the aperture would be permitted to affect pixel values, which obviously makes no sense. For instance, a uniform image would yield a non–uniform sub–image which is of course preposterous.

2.1. *Histograms of images*

"Histograms" are usually introduced in the context of discrete data structures. Here one of our primary aims is to define the concept for continuous (not pixellated) images. The reason is that the "pixels" are only crutches and essentially irrelevant to the image. The pixellation has to do with the structure of image *representation*, not at all with the structure of images *per se*. For a pixellated image $I(\mathbf{r}; \sigma)$ the value of the histogram for the "bin" $(i, i + \Delta i]$ is the number of pixels with values in the half–open interval $i < I \leq i + \Delta i$ divided by the bin–width Δi (the division is convenient to obtain a "density" independent of the bin–width). In the continuous case it is natural to use the area of image regions for which $i < I \leq i + \Delta i$ instead of the number of pixels. A further normalization may be performed by dividing by the total available area or the total number of pixels. This makes the histogram independent of the—accidental—image size.

First we consider the 1D case. We are led to try the *Ansatz*:

$$H(i; \sigma, \beta) = \frac{1}{\sqrt{2\pi}} \int_{\text{full image}} \mathrm{e}^{-\frac{(I(\mathbf{r};\sigma)-i)^2}{2\beta^2}}\, d\mathbf{r}, \qquad (4)$$

where the parameter β represents the "bin–width". We have chosen this particular normalization factor in order to calibrate the method: We naturally require that the method yields the same results as the discrete method in case of linear isophotes, that is to say, for the image

$$I(\mathbf{r}; \sigma) = gx, \qquad (5)$$

(g the magnitude of the linear gradient) and the bin–width b we would expect $H = b/g$. Doing the integral 4 for the linear gradient image 5, we indeed find this expected result, namely β/g. Thus we are led to identify the parameter β with the "bin width".

In two dimensions this method essentially measures the area of a strip extended along the isophote $I(\mathbf{r}; \sigma) = i$ of width $\beta/\|\nabla I(\mathbf{r}; \sigma)\|$. Indeed, the discrete alternative would be the area of the strip between the isophotes at i and $i + \Delta i$.

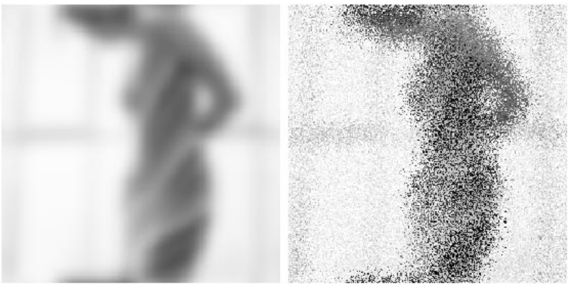

Fig. 4. Image of a figure in a striped dress in front of an illuminated window. (Original shown in figure 8 *left*.) The image subtends 256x256 pixels. Shown are a blurred version (left) and a disordered rendering (right). In both cases the blurring (or scrambling) is over 16 pixels.

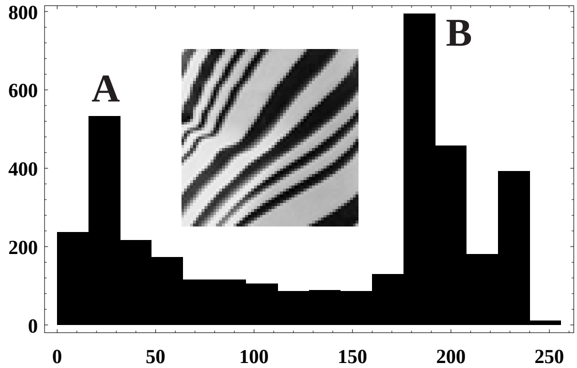

Fig. 5. A region of interest from the middle of the striped dress (figure 4 and figure 8 *left*) and the corresponding histogram. Modes A and B relate to the dark and light stripes of the dress.

The image (simply obtained from the image $I(\mathbf{r};\sigma)$ through a nonlinear gray scale transformation)

$$e^{-\frac{(I(\mathbf{r};\sigma)-i)^2}{2\beta^2}} \tag{6}$$

is itself of much intrinsic interest. It can be regarded as an especially apt definition of "soft isophote", a definition that reveals more than simply a geometrical locus. In addition one obtains a measure of the "width" of the isophote. This image specifies "how much" of any fiducial value is present at any location, it reveals the spatial distribution of the "stuff" that is in a given bin of the histogram. Obviously we need consider a one–parameter family (parameter β) of such images (in the case of scalar images): One such image for each value of β. (Figures 4 through 7.)

In many cases one would like to find the histogram of some *region of interest* instead of the whole image. Indeed, what is a "whole" image anyway? Picture frames are necessarily *arbitrary*. An easy way to do this is to simply *crop* the image. However, this will certainly introduce non–causal effects ("spurious resolution") and thus is not to be recommended if the spatial distribution of histograms is of interest. The unique way in which spurious resolution can be avoided is scale space. It is in fact an easy matter to build a histogram scale space, that is to say, a structure for which the "image" defined by any individual "bin" respects the scale space structure: One simply applies the aperture functions to the "raw bin–image" (equation 6). Thus one obtains:

$$H(i;\mathbf{r_0},\sigma,\beta,\alpha) = \frac{1}{2\pi\alpha^2}\int_{\text{whole space}} A(\mathbf{r};\mathbf{r_0},\alpha)e^{-\frac{(I(\mathbf{r};\sigma)-i)^2}{2\beta^2}}\,d\mathbf{r} \tag{7}$$

Here the parameter α measures the "size of the region of interest". We call it the *outer scale*, whereas the parameter σ ("resolution") denotes the *inner scale*. The division by $2\pi\alpha^2$ avoids the trivial dependence on the size of the region of interest. For instance, a constant function yields a histogram value of unity for the corresponding bin. For the image $i_0 + gx$ the value for the bin i_0 is $1/\sqrt{1+(\alpha g/\beta)^2}$, that is nearly $\beta/g\alpha$ when the region of interest is much broader than the isophote width, *i.e.*, the expected value (see discussion in previous section).

Notice that in expression 7 the aperture weighs the density of contribution to the bin, rather than the pixel values themselves. It is effectively used as a fuzzy characteristic function. The (dimensionless) number $\chi = (\alpha/\sigma)^2$ denotes the "logon content" or total number of degrees of freedom of the ROI. It measures the cardinality of the local set of orderless pixels. The logon content χ conveniently characterizes the structure up to an overall scaling and size. The parameter β denotes the "bin–width", that is the resolution in the domain of the sampled value (*e.g.*, radiance). When these values are not scalars and are not property lists it is necessary to select some suitable positive definite quadratic form for the bin: In principle this construction applies to *arbitrary image and value domains*. For instance, if the value is a color, we may use our favorite color metric.

This construction leads to a similar result as obtained by Griffin (1997) who in an abstract fashion applies a scalespace kernel to the direct product of the image domain and the value domain. Griffin (1997) does not introduce the notion of finite resolution in the image domain though, that is to say, he introduces our parameter α as the resolution parameter (our parameter σ does not figure in his formalism). Thus his notion of

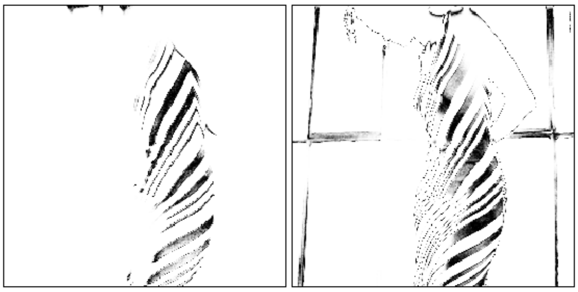

Fig. 6. The density of pixelvalues from modes A (right) and B (left) (see figures 4 and 5). Notice that some of the background pixels (*e.g.*, the window bars) happen to lie in these modes too.

spatial resolution corresponds to our notion of the size of the region of interest. Whereas it has some formal advantage (at least in terms of mathematical elegance) to apply a generalized scale space kernel to the direct product of image and value space, we believe that the present approach yields greater heuristic power and is more immediately suited to image processing applications. The "raw bin image" may actually be computed in parallel by special hardware and thus the whole calculation can be pipelined in a relatively conventional manner.

One may understand the structure of equation 7 in the following way: Think of the image as an orderless collection of "pixels ("infinitesimal squares" say). Then the integral is simply a summation over pixels regardless of their order, one needn't bother with any (spatial) image structure. This sum is simply a *Parzen estimator* (Parzen 1962) of a probability density function, in this case the histogram of pixel values. Each pixel contributes a Gauss function of width β (the bin width) centered at its pixel value to the histogram. Thus the histogram is nothing but a Gaussian blurred version of a sum of unit delta pulses, one for each pixel, at its pixel value. This shows immediately that the histogram obtained in this manner is an element of a *scale space,* with "spatial variable" i and scale parameter β. This view complements that of interpreting equation 7 as a method that essentially measures the total area of the strip centered on the isophote $I(\mathbf{r}; \sigma) = i$ and width equal to $\beta / \|\nabla I(\mathbf{r}; \sigma)\|$ weighted with the aperture function.

2.2. *The locally orderless image*

We prefer to think of $H(i; \mathbf{r_0}, \sigma, \beta, \alpha)$ as a probability density on the I–domain (the i–dependence) that varies from place to place (the $\mathbf{r_0}$–dependence) and represents

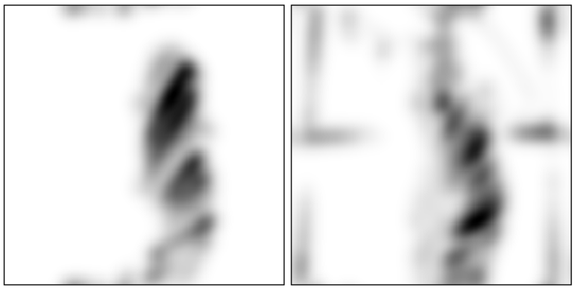

Fig. 7. 'The density of pixelvalues from modes A (right) and B (left) blurred over 16 pixels (see figures 4, 5 and 6). Whereas in figure 6 the individual stripes could be resolved, here a given location may correspond to *both* a light and a dark stripe.

the scene for certain fixed resolution and scope parameters (α a scope parameter, σ and β resolution parameters). Alternatively, one may conceive of an ordered pile of superimposed images, each layer representing the probability density of some fixed value of i. (Essentially the picture presented in figure 7.) One may also think of a graph $i(\mathbf{r_0})$ that has been "drawn with a fuzzy pencil" over the pixel plane.

Notice that the variation with $\mathbf{r_0}$ is a true spatial variation (thus on the scale α the images are ordered, in the sense that we have a topology). However, on the scale σ (it makes no sense to consider even finer scales, since such information has already been lost at observation time) the images are *orderless*. Essentially, at any position we see a mere set (orderless collection) of χ independent samples at resolution σ, the histogram represents all we may know about this set (how many pixels are of radiance i, but not their mutual positions). This is why we speak of $H(i; \mathbf{r_0}, \sigma, \beta, \alpha)$ as a *locally orderless image* of the scene. Such entities occupy the limbo between the *images* $I(\mathbf{r})$ proper and the *histograms* (in the conventional, global sense) $H(i; \sigma, \beta)$ computed over the full image such as they are in common use.

Notice also that $H(i; \mathbf{r_0}, \sigma, \beta, \alpha)$ for fixed i, σ and β is a scale–space with spatial variable $\mathbf{r_0}$ and scale–parameter α. When one considers $H(i; \mathbf{r_0}, \sigma, \beta, \alpha)$ for fixed $\mathbf{r_0}$, σ and α one has another scale–space with i as "spatial variable" and β (the bin–width) as scale–parameter. Of course this is all based upon the image $I(\mathbf{r}; \sigma)$ which is a scale–space with spatial–variable $\mathbf{r}$ and scale–parameter σ. Thus the locally orderless images are elements of three quite distinct but intimately related scale–spaces.

The essential structure of $H(i; \mathbf{r_0}, \sigma, \beta, \alpha)$ is summarized by the equations:

$$\triangle_{(\mathbf{r})} I(\mathbf{r}; \sigma) = \frac{\partial I(\mathbf{r}; \sigma)}{\partial \frac{\sigma^2}{2}}, \tag{8}$$

based upon $\mathcal{S}(\mathbf{r})$ through $I = \mathcal{S} \otimes G_0$,

$$H(i; \mathbf{r_0}, \sigma, \beta, \alpha) = \frac{A(\mathbf{r}; \mathbf{r_0}, \alpha)}{2\pi\alpha^2} \otimes R(\mathbf{r}, i; \sigma, \beta), \tag{9}$$

with $R(\mathbf{r}, i; \sigma, \beta) = \mathrm{e}^{-\frac{\pi(I(\mathbf{r};\sigma)-i)^2}{\beta^2}}$,

$$\frac{\partial^2 H(i; \mathbf{r_0}, \sigma, \beta, \alpha)}{\partial i^2} = \frac{\partial H(i; \mathbf{r_0}, \sigma, \beta, \alpha)}{\partial \beta}, \tag{10}$$

fixed $\mathbf{r_0},\ \sigma,\ \alpha$,

$$\triangle_{(\mathbf{r_0})} H(i; \mathbf{r_0}, \sigma, \beta, \alpha) = \frac{\partial H(i; \mathbf{r_0}, \sigma, \beta, \alpha)}{\partial \alpha}, \tag{11}$$

fixed $i,\ \sigma,\ \beta$.

Here equation 8 expresses the dependence on the scene and the spatial scale space structure (notice that the "scale parameter" is $\sigma^2/2$), equation 9 expresses the relation of the locally orderless image to the primary image, equation 10 expresses the scale space structure of the histogram and equation 11 the scale space structure of the spatial variation of the histogram. Taken together the equations describe the dependence of the locally orderless image on the scene for given scale parameters, but—more importantly—they capture the structure of the locally orderless images *as a family* as a function of the variables $\{\mathbf{r_0}, i\}$ and parameterized by $\{\alpha, \beta, \sigma\}$. Further investigation of the structure will have to depart from this. The triple scale space structure of course allows one to apply all tools developed in the context of scale space in a variety of ways.

The differences with the approach proposed by Griffin (1997) are mainly due to the distinction we make between the spatial resolution parameter σ and the parameter that describes the extent of the regions of interest α. This difference is an important one though since the local histograms are greatly influenced by the magnitude of the spatial resolution. Just think of a treetop in early fall where the foliage is made up of leaves of very different colors: When the spatial resolution suffices to resolve individual leaves the local histograms will reflect this multicolored appearance from green over yellow to reddish. However, when the resolution only suffices to resolve major leaf clusters the histograms reflect merely the color of the mixture, perhaps a yellowish green.

3. Natural operations on locally orderless images

Any local operation that doesn't depend on the local spatial order is a natural operation on locally orderless images. Examples include such well known operations as blurring (histograms replaced with the means), median filters (histograms replaced with the medians), "saliency finders" (*e.g.*, histograms replaced with the variance), *etc.*

More interesting operations actually exploit the *structure* of the histogram. For instance, one might extract the "modes" of the histogram. This can be done by computing the partial order of maxima (peaks inside peaks next to peaks, …). A convenient (and well defined) method is to subdivide the domain at the minima, starting with the lowest, until each segment contains only a single maximum. Typically one will then apply some specific criterion to discard all but the "major modes" and put these in a list with their position and possibly other characteristics. In case only the positions are retained (the i–values) one obtains a *multiply valued* image where one has one or possibly more (a variable number) values at any location. That this seemingly odd representation might actually be rather intuitive can be illustrated by way of an example: Consider a nearly bare treetop against the sky. In the area of the treetop we find pixels of sky–value and pixels of branch– or foliage–value,thus the histogram is likely to have (at least) two major modes. One mode belongs to the sky, the other to the tree. At certain locations one finds both sky and tree modes, thus the locally orderless image gives one the opportunity to retain *both*. This nicely fits the perception of "the sky seen through the bare tree". There exist many instances of this general character in most natural scenes. This particular representation has been rather elegantly described by Griffin (1997) and was first proposed by Noest and Koenderink (1990).

An example of transparent segments is shown in figure 8. Notice that the segment belonging to the light stripes of the dress extensively overlaps with that of the dark stripes of the dress: Here the stripes are effectively unlocalized on the fine scale, so *every* part of the dress region belongs *simultaneously* to the light and to the dark stripes.

Another—very simple—example of such transparent segments is illustrated in figures 9 and 10. It is an extreme example because both segments are extended over the full image. The (1D) image is a sinusoid with

Fig. 8. Original image (left) of a figure in a striped dress in front of an illuminated window. Thresholding the blurred density of histogram stuff (see figure 7) yields segments corresponding to mode A (the black stripes, right) and B (the light stripes, middle) (see figure 5). Notice that these segments almost totally overlap: The segments are regions of *both* light and dark stripes. This is the transparancy effect discussed in the text.

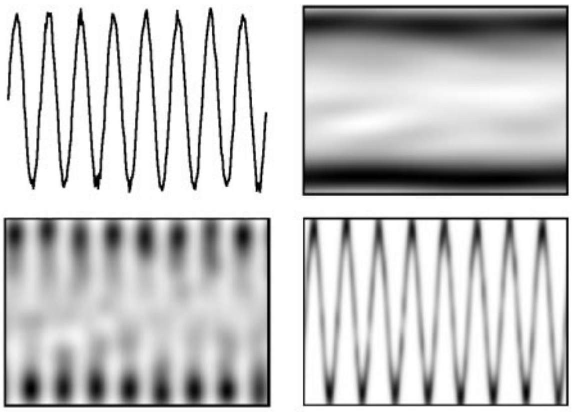

Fig. 9. Three different representations of the signal illustrated in the top–left panel. The signal is a sinusoid with additive Gaussian noise. For the other panels the resolution (σ) is held fixed, whereas the bin–width (β) and the diameter of the ROI (α) are varied. Vertical columns in these pictures are local histograms, the horizontal dimension is simply the image domain. Increasing values are mapped on darker tones. Notice that in the bottom–right panel the sinusoid is resolved, though much of the Gaussian noise has been lost due to the blurring. In the panel on the top–right the ROI is too large to be able to resolve the undulations of the sinusoid. Yet the locally orderless image structure allows one to detect that there are two distinct image segments *transparently overlapping in space* of very different pixel value. Here one sees the layers of extrema of the sinusoid without resolving the spatial structure (like seeing the sky through the branches of a bare tree). The panel on the bottom–left is an in between case. Though the actual structure remains mysterious, it is clear that there is significant, periodic spatial modulation, and that the image is composed of intertwined light and dark patches.

additive Gaussian noise. The histogram of the full image is strongly bimodal, the majority of pixels is either very light or very dark, with relatively fast transition regions in between. When the undulations of the sinusoid cannot be resolved (α exceeds the wavelength) but the tonal variations can be resolved (σ much less than the wavelength), these modes appear when β is much less than the amplitude of the sinusoid. When the spatial resolution suffices (α much less than a wavelength), the

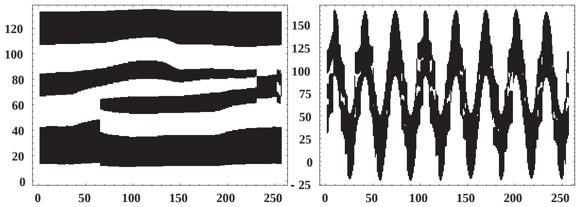

Fig. 10. A plot of the modes of the histograms shown in figure 9. The left panel corresponds to the top–right panel in figure 9, and the right panel to the bottom–right panel in figure 9. The transparent nature of the segments is particularly clear in the lefthand panel, whereas the righthand panel obviously shows the spatially resolved structure. Though spatially unresolved, the orderless image manages to retain the fact that the signal predominantly dwells at two distinct pixel values.

variation of local histograms nicely follows the undulation and the signal is largely resolved. In intermediary situations one obtains unimodal local histograms that vary wildly with position.

Of course, it is not altogether trivial to distinguish such modal segments (although rather often it *may* be trivial) because the value of the mode typically will be a (slowly varying) function of location. Then one can use the continuous nature of the modal segments (the value of the mode varies gradually from point to point whereas different modes are at quite distinct levels) to help define them. In most cases of interest it will be sufficient to set a number of thresholds defining ranges for the various major mode values. The segmentation then is simply a (multiple) thresholding operation.

The possibility to retain multiple "transparent" segments at any location is clearly one of the more interesting properties of locally orderless images.

4. Rendering of locally orderless images

One may generate (approximate) instances of locally disordered images by drawing pixel values at random from a distribution specified by the local histogram. The fashion image (figure 11) was generated in this manner. In practice such instances (there exist huge numbers of them) will look the same to the casual observer, though close scrutiny will of course reveal multitudinous differences. The deterministic spatial structure is at a scale given by α, and blurring, squinting, looking through the eye–lashes or "screwing up the eyes" will indeed reveal this. This is quite similar to "dithered" renderings of continuous scale images on a mere black–and–white device. (See for instance Ulichney 1987 and 1988.) The essentially spatially

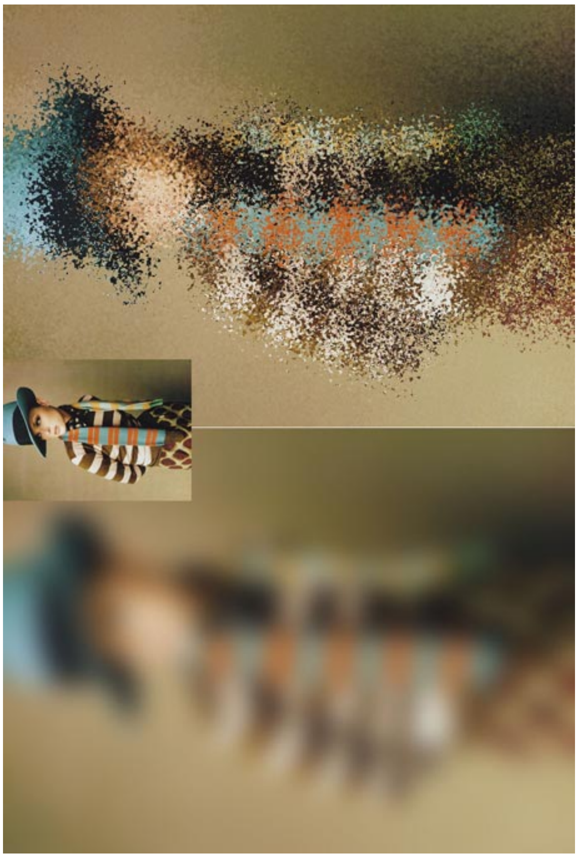

Fig. 11. A fashion image (small inset) blurred (left) and disordered (right) by the same amount. Notice that the spatial resolutions of the blurred and the disordered image are indeed similar, but that the disordered image has retained pixel values that are lost in the blurred image where they were averaged out. Though "unsharp", the disordered rendering has thus retained a vestige of image structure at the original scale. Aesthetically (as well as information technically) the disordered rendering is much to be preferred over the blurred image, in fact it is not unlike impressionist renderings.

random microstructure yields additional scene information though: This information has to do with the local textural qualities of the scene. For instance, the blue sky will appear uniform whereas foliage will appear grainy. Though the grains don't reveal any leaves it is still the case that the average luminance, the highlight luminance (glints of sunlight on the top of the leaves) and the shadow luminance (underside of leaves, cast shadow of one leaf on another) are evident from the rendering. Whereas the grain is not informative of spatial structure as such, it *does* specify the spatial scale of the major luminance variations. As a consequence, the local orderless structure at least partially specifies surface properties of the scene. Thus rendered orderless images are not unlike paintings or drawings where foliage has been "suggested" yet no single leaf has been painted or drawn. The apparent "realism" of many such paintings as compared with many computer graphics efforts (where most surfaces seem to have been made of plastic) suggests that rendering locally orderless images might be an interesting and rewarding endeavor. Notice that the histograms do not require precise spatial models, though the essential physics should indeed be modeled: For instance, the local histograms will depend on surface properties, the local light field, and the viewing direction. One can't simply "texture–map" them.

5. Conclusion

As we have shown the orderless images are elements of three quite distinct but intimately connected scale–spaces. One of these scale spaces is the conventional image scale–space where the space variable is the image coordinate and the scale–parameter the conventional resolution. Another scale–space is that of the histograms, the "space variable" is the pixel value and the scale parameter is the bin–width (called "imprecision" by Griffin 1997). Here the "soft"–bins prevent spurious resolution of the histogram structure. The third is the scale–space of the spatial variation of the histograms (called "scale" by Griffin 1997). The image of values of each bin is an element of the scale–space where image location is the spatial variable, and the width of the region of interest is the scale–parameter. The number χ characterizes the spatial structure, it specifies the number of orderless pixels within a region of interest. The bin-width β is the other parameter that specifies the orderless image, it applies to the resolution in the domain of pixel values. Whenever there exists a fiducial value (such as the total range for a finite domain) one may also turn it into a dimensionless number.

Notice that the formalism generalizes without further ado to image domains of arbitrary dimension. Especially dimension 3 is of great practical interest of course. Notice that one need not restrict oneself to spherically symmetric ROI's: Because scale space naturally factors due to the fact that the Gaussian is separable in Cartesian coordinates one may contemplate more general structures. Because of this latter fact it is also not necessary that the image domain be Euclidean. This is important when not all dimensions are mutually commensurable as is often the case in practice.

In the case of vector valued images, or even worse, the case of property list or record valued images, the "histograms" become densities on higher dimensional spaces. In the case of property lists these spaces have no Euclidean, but only an affine structure. The separability of the Gaussian kernel again lets us handle such cases naturally and with ease: No essential changes to the structure described here are required, only straightforward generalizations. Such cases are common in general image processing (color) and in medical images where a variety of properties is often associated with any given "voxel".

References

Florack, L. M. J. 1997. *Image Structure.* Dordrecht: Kluwer.

Griffin, L. D. 1997. Scale–imprecision space. *Image and Vision Computing*, 15: 369–398.

Helmholtz, H. 1866. *Handbuch der physiologischen Optik*. Hamburg and Leipzig: Voss.

Hess, R. 1982. Developmental sensory impairment: Amblyopia or tarachopia? *Human Neurobiology*, 1: 17–29.

Homer, W. I. 1964. *Seurat and the science of painting*. Cambridge, Mass.: The M.I.T. Press.

Koenderink, J. J. and van Doorn, A. J. 1978. Invariant features of contrast detection: An explanation in terms of self–similar detector arrays. *Biological Cybernetics*, 30: 157–167.

Koenderink, J. J. 1984. The structure of images. *Biological Cybernetics*, 50: 363–370.

Koenderink, J. J. and van Doorn, A. J. 1992. Generic neighborhood operators. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14: 597–605.

Lindeberg, T. 1994. *Scale–space theory in computer vision*. Boston, Mass.: Kluwer.

Metzger, W. 1975. *Gesetze des Sehens*. Frankfurt a.M.: Verlag Waldemar Kramer.

Noest, A. J. and Koenderink, J. J. 1990. Visual coherence despite transparency or partial occlusion. *Perception*, 19: 384.

Parzen, E. 1962. On estimation of a probability density function and mode. *Annual Mathematical Statistics*, 33: 1065–1076.

Ruskin, J. 1900. *Elements of drawing* (first ed. 1857). Sunnyside, Orpington: George Allen.

Ruskin, J. 1873. *Modern Painters* (Vol. I). Boston: Dana Estes & Company.

Ulichney, R. A. 1987. *Digital halftoning.* Cambridge, Mass.: The M.I.T. Press.

Ulichney, R. A. 1988. Dithering with blue noise. *Proc. IEEE*, 76: 56–79.