

Continuous Dimensionality Characterization of Image Structures [★]

Michael Felsberg

Computer Vision Laboratory, Dept. EE, Linköping University, Sweden.

Sinan Kalkan ^{*}

BCCN, University of Göttingen, Germany.

Norbert Krüger

Cognitive Vision Group, Univ. of Southern Denmark, Denmark.

Abstract

Intrinsic dimensionality is a concept introduced by statistics and later used in image processing to measure the dimensionality of a data set. In this paper, we introduce a continuous representation of the intrinsic dimension of an image patch in terms of its local spectrum or, equivalently, its gradient field. By making use of a cone structure and barycentric co-ordinates, we can associate three confidences to the three different ideal cases of intrinsic dimensions corresponding to homogeneous image patches, edge-like structures and junctions. The main novelty of our approach is the representation of confidences as prior probabilities which can be used within a probabilistic framework. To show the potential of our continuous representation, we highlight applications in various contexts such as image structure classification, feature detection and localisation, visual scene statistics and optic flow evaluation.

Key words: intrinsic dimensionality, feature extraction and classification

[★] The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 215078 DIPLECS and from the CENIIT project CAIRIS. This work has further been supported by the EC Grant Drivscio (IST-FP6-FET-016276-2) and the Danish National Project NISA. However, this paper does not necessarily represent the opinion of the European Community, and the European Community is not responsible for any use which may be made of its contents.

^{*} Corresponding author. Bunsenstr. 10, BCCN, University of Göttingen, 37073, Göttingen, Germany. Tel: +49 551 5176509 Fax: +49 551 5176449

Email addresses: mfe@isy.liu.se (Michael Felsberg),

1 Introduction

In this paper, we propose a continuous representation of *intrinsic dimensionality* for the purpose of image processing and demonstrate its usefulness by briefly describing the results of a number of applications of our representation in different fields of computer vision. First formulations of our representation of intrinsic dimensionality has been published in the two conference papers [12, 27]. Extended descriptions of the applications of these have been published at conferences [11, 24], a journal [23] or are under preparation. This paper presents now for a first time a compact description of the continuous iD representation which (although based on ideas formulated in [12, 27]) has now matured by (1) its use in a number of contexts which underline the relevance of our representations, (2) a proper theoretical derivation of the triangular and continuous structure of the intrinsic dimensionality, (3) by new theoretical investigations leading to an appropriate formalism to deal with image noise, and (4) a thorough comparison with and embedding in the existing literature.

The main novelty of our approach is the representation of continuous confidences in a compact interval by means of an adaptive soft-threshold function, such that all three cases add up to one. Hence, the intrinsic dimensionality can be used as a prior in a Bayesian framework of processing, where the subsequent processing steps evaluate the data in terms of conditional probabilities given a particular intrinsic dimensionality. This is also the main difference to classical work on edge and corner detection, see [20, 15, 3, 39, 33, 25] for a selection of (mostly tensor-based) methods, where a hard classification into edge, corner, or no structure is made.

1.1 The Intrinsic Dimensionality of Images in the Literature

The intrinsic dimensionality (iD) is a well known concept from statistics which can be defined as follows (from [5], p. 314):

a data set in d dimensions is said to have an *intrinsic dimensionality* equal to d' if the data lies entirely within a d' -dimensional subspace.

The term itself goes back to the late sixties [42]. It has been used in the context of statistical data analysis, see e.g. [18]. More recent work from the field of machine learning, more concretely manifold learning, often defines the intrinsic dimension as the dimension of the underlying manifold. In [32] statistical

sinan@bccn-goettingen.de (Sinan Kalkan), norbert@mip.sdu.dk (Norbert Krüger).

methods are used to estimate the intrinsic dimension of manifolds embedded in high-dimensional spaces and the approach is related to PCA and previous local methods for dimensionality reduction, e.g. [7]. In [9], a differential geometry setting is used to estimate the intrinsic dimensionality of high-dimensional spaces using entropic graphs. Other work relating images and intrinsic dimensionality often consider the manifold of images or image patches [40,31] rather than the intrinsic dimensionality *within* a single image. The latter approach to intrinsic dimensionality of images was introduced by Zetsche and Barth [44]. In this paper, intrinsic dimensionality was obtained by applying the definition from [5] to the spectrum of an image patch, i.e., the Fourier transform of a neighborhood. The three possible intrinsic dimensionalities in images are defined according to their local spectrum [26] (see also Fig. 1(a)):

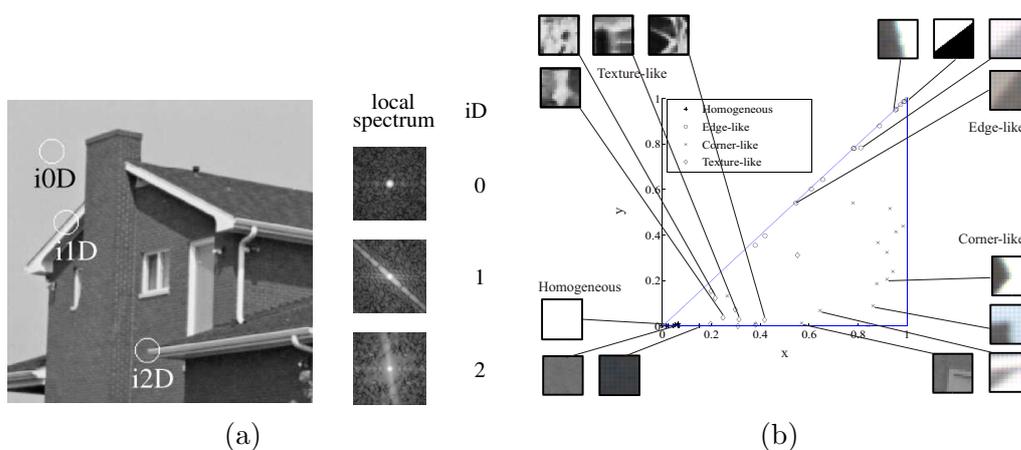


Fig. 1. (a) On the definition of intrinsic dimensionality. In the image on the left, three neighborhoods with different intrinsic dimensionalities are indicated. The other three images show the local spectra of these neighborhoods. (b) Different image structures covering homogeneous patches, edges, junctions and textures and their position in the iD triangle.

i0D It is concentrated in the origin, i.e., the neighborhood is constant.

i1D It is concentrated in a line through the origin, i.e., the neighborhood is varying in one direction. These signals are also called *simple signals* [17].

i2D It is neither concentrated in the origin, nor in a line.

Typical examples for i1D neighborhoods are edges, lines, and sinusoids, whereas corners, junctions, line ends, and spots are instances of i2D neighborhoods. This is also related to the approach described in [40], where the authors consider two subspaces of the manifold spanned by local image patches (20x20): the explicit manifold, generated by a basis of iconic image primitives as edges, ridges, and junctions, and the implicit manifold, represented by a set of features and typically containing textures.

A similar approach to the manifold of image patches, albeit at smaller size (3x3), has been followed in [31]. Here, the authors transform the DC-free, normalized patch (a 7D vector) to a vector in the 7-sphere S^7 using a DCT-based technique. The 7-sphere is then tessellated into a number of Voronoi cells and the cells with the highest probability establish a kind of sparse coding base for natural images. Both methods [40,31] consider the manifold of image patches.

In contrast to that, we define the intrinsic dimensionality *within an image* as the dimension of the local spectrum or, equivalently, the *manifold of the gradient field*. Furthermore, in our approach we do not restrict the intrinsic dimensionality to discrete choices, i.e. integer numbers, but allow a continuous transitions similar to the fractal or capacity dimension. The space of dimensionality itself is not just a real number, as it is equipped with a 2D topology.

There are many possible ways to estimate the intrinsic dimensionality of an image. One possibility is to average the outer product of gradients as it is done for the structure tensor [4, 15]. The structure tensor has by definition two non-negative eigenvalues. The two eigenvalues span one quadrant in 2D space. Rotating this quadrant around its diagonal results in a cone-shaped 3D subspace. Another method combines the local magnitude with a statistical measure of local orientation variance as been used in a precursor of this paper (see [27]).

1.2 The Need for an Extended Definition of the Intrinsic Dimensionality

The main use of the intrinsic dimensionality lies in the adaptive control of other, mostly higher level image processing or computer vision methods. Consider for instance motion estimation. Due to the aperture problem, only the normal flow vector can be estimated in i1D regions. In i0D regions, no estimation is possible, and in i2D regions, most estimation methods will fail because they apply a model which assumes i1D structures (for a detailed analysis see section 4.4 and [23]). Whereas some methods can make use of the covariance matrix of the measurements to adapt to the intrinsic dimensionality [34], some measurement problems require totally different methods for different dimensionalities, and measurements for different intrinsic dimensionalities have to be computed with different algorithms (see, e.g., [2]).

Besides motion estimation, there are further image processing and computer vision tasks where either the quality of the result or the underlying model depend on the iD of the data. For instance, in model-fitting, one would like to choose a model based on those dimensions where the image data is non-

trivial (e.g. fitting lines or planes in 3D); in 3D correspondence search, one would like to reduce the number of candidates by the iD (see for example [28]); and for i2D structures, a different post processing needs to be performed to detect the position of a feature than for i1D structures (see section 4.1). Furthermore, the statistics of underlying depth structures differs significantly with the intrinsic dimensionality (see section 4.3), giving further indication for the need of adaptive higher level processing schemes depending on the iD.

When it comes to measuring the iD, one notices that there are no or only very few 'pure' i0D, i1D, or i2D signals in natural images. Furthermore, as soon as we take noise into account, the mathematically strict definition above becomes useless. Noise is i2D, and most signals contain noise. Hence, in general image neighborhoods are i2D. But how to distinguish between noise and i2D image structures? Estimating the noise level is possible (e.g. [14]), but the transition between noise and i2D image structures remains fuzzy. Apparently, the definition of iD in terms of discrete states is unsuitable. We therefore propose a continuous definition of iD in terms of a *bounded iD cone* which can be mapped canonically to a triangle. Fig. 1(b) shows different image structures being mapped to different areas in the triangle in which the 'pure' cases of homogeneous patches and edges are mapped to the left corner and upper right corner respectively while non-pure cases become mapped to the inner part of the triangle. This triangle representation will then allow for the definition of 3 confidences expressing the similarity to each of the 'pure' i0D, i1D and i2D structures.

Our continuous understanding of intrinsic dimensionality allows for a flexibility in higher level processing. For example, instead of deciding for one dimensionality and selecting the further processing based on this decision, it is possible to combine all three cases based on statistical methods (see, e.g., [41]). This strategy avoids instabilities caused by the early decision about the iD. Furthermore, the confidences for the different intrinsic dimensionalities can be used as weights in post processing schemes (see section 4.1).

To propose a more appropriate model of iD than the discrete one is the main purpose of this paper. This is essentially done in Sect. 2. In Sect. 3, we consider more practical aspects like the parameterization, the soft-thresholding, and the estimation of intrinsic dimensionality. In Sect. 4, we briefly describe some applications of the continuous understanding of intrinsic dimensionality. We conclude the paper with an outlook in Sect. 5.

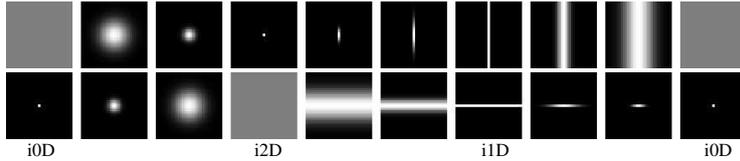


Fig. 2. Representing the continuous intrinsic dimensionality in a 1D periodic space for the example of different Gaussian functions. Upper row: spatial domain. Bottom row: frequency domain.

2 Modelling the Intrinsic Dimensionality

In this section, we elaborate our new modelling of intrinsic dimensionality by three observations, each of them leading to new aspects of the model. For these observations, we consider 2D Gaussian functions to illustrate different cases of intrinsic dimensionality. The family of 2D Gaussian functions contains all three cases of iD (impulse, impulse-line, and constant signal) and continuous variations between these three cases. Although we solely use this single signal family for illustrating the iD , one should notice that all local signals map somewhere in our representation and that there are infinitely many other signals of the same iD . We used Gaussian functions for illustration purposes, since this family represents all possible 2D signals concerning the iD , i.e., any 2D signal corresponds to a certain Gaussian function w.r.t. the intrinsic dimensionality.

2.1 The Intrinsic Dimensionality is Non-Discrete

The $i0D$ Gaussian function has infinite variance, i.e., it is a constant function. The $i1D$ Gaussian function has infinite variance in one orientation and zero variance in the perpendicular orientation, i.e., it is an impulse line. The $i2D$ Gaussian function has zero variance, i.e., it is an impulse.

Apparently, there exist a continuous range of Gaussian functions between a constant function and an impulse line, between the impulse line and the impulse, and the impulse and the constant function. Therefore, the intrinsic dimensionality should not be defined as discrete states, but as a *continuous space*. In Fig. 2, a periodic 1D space is used to represent the intrinsic dimensionality, including the three discrete states from the classical model.

2.2 The Space of iD Has More Than One Dimension

The Gaussian functions in Fig. 2 are either 1D Gaussian functions projected into two dimensions, 1D Gaussian functions on an impulse line, or 2D isotropic

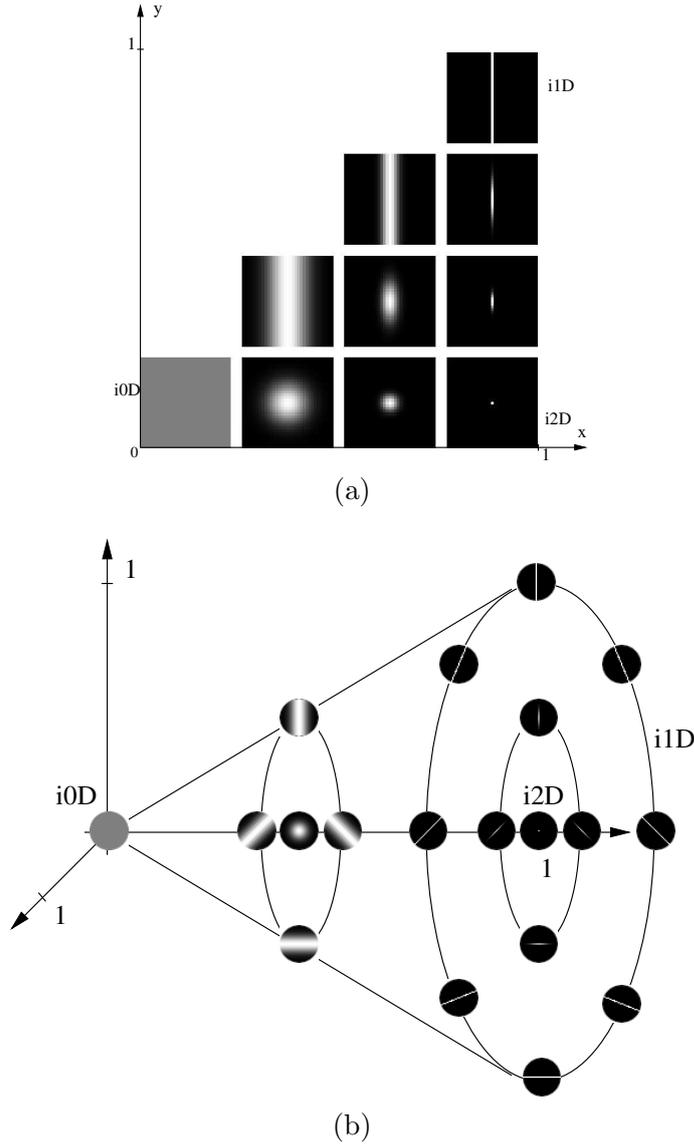


Fig. 3. (a) Representing the intrinsic dimensionality as points in a 2D space of triangular shape for the example of different Gaussian functions. The x -coordinate corresponds to some monotonic function of the inverse radial variance. The y -coordinate corresponds to some monotonic function of the anisotropy. (b) The cone model for the space of intrinsic dimensionality illustrated with the example of different Gaussian functions.

Gaussian functions. The anisotropic, non-degenerated 2D Gaussian functions, however, are not yet contained in the model. The latter do not lie on the periodic 1D space from Fig. 2, but *in between* the cases on the periodic space. This implies that the space of iD cannot be one-dimensional as implicitly expressed in the discrete counting $i0D$, $i1D$ and $i2D$. Anisotropic Gaussian functions (with fixed orientation) form a finite $2D$ continuous space with three vertices. Hence, the space of intrinsic dimensionality with fixed orientation has the shape of a triangle, see Fig. 3(a).

The contour of the triangle corresponds to the periodic 1D space introduced further above. The three corners correspond to the three discrete states in the classical model. The triangular structure allows for changing continuously (and linearly) from any particular case of intrinsic dimensionality to any other case.

2.3 The Space of iD is a Cone

A constant or isotropic function does not include any orientation information. Anisotropic Gaussian functions, however, include 2D orientation information. Therefore, we can distinguish different triangles like the one from Fig. 3(a) by different orientations of the reference coordinate system. The orientations cover again a continuous range. All triangles have a common edge which lies between the $i0D$ case and the $i2D$ case, which are both unoriented. The space of iD is hence obtained by rotating the iD triangle around the $i0D$ - $i2D$ edge. The resulting object is a *cone* in 3D space, see Fig. 3(b).

If the triangle is not rectangular, the resulting object is a double-cone. However, double cones and cones fall into the same equivalence class under topology-preserving mappings. The only formal difference lies in a strictly monotonic change of parameterization. Hence, without loss of generality, we consider a simple cone in the following.

In the cone model, the *ideal* $i0D$ case and the *ideal* $i2D$ case are both represented by points, i.e., subspaces of dimension zero, whereas the *ideal* $i1D$ case is represented by a circle, i.e., a subspace of dimension one. The background of the different dimensions is already given by the definition of iD in terms of subspace dimensions. There exist respectively one zero-dimensional and one two-dimensional subspace of 2D space. The 1D subspaces, however, form a 1D family of subspaces, parametrized by the orientation of the subspace.

Although both are represented by points, there is a fundamental difference between the ideal $i0D$ case and the ideal $i2D$ case: whereas the former covers only the constant signals, the latter includes a huge variety of different signals. Correspondingly, the non-ideal $i0D$ signals are spread in a much smaller volume than the non-ideal $i2D$ signals. This fact is well covered by the model, where the volume around the $i0D$ vertex is much smaller than the volume around the $i2D$ vertex

3 Practical Aspects

Knowing the topological structure of iD space is a first step to make use of it. However, the space has to be parametrized and these parameters have to be estimated in order to apply the model in practice. These two issues are the subject of this section.

3.1 Barycentric Coordinates

As already indicated in Fig. 3(a) and 3(b), we choose, in a first step, the coordinate ranges to lie between 0 and 1. Within our triangular representation, a suitable parameterization for the formalization of the confidences that indicate the iD of the local signal are *barycentric coordinates* [8]. In a second step, the second coordinate becomes a complex number in order to represent every point in the cone.

We set the vertices to $(x, y) = (0, 0)$ for the i0D case, to $(x, y) = (1, 1)$ for the i1D case, and to $(x, y) = (1, 0)$ for the i2D case. Any point in the triangle is represented by its three barycentric coordinates $(c_{i0D}, c_{i1D}, c_{i2D})$ where each coordinate can be interpreted as the likelihood for the corresponding iD case, because each c_{ikD} for $k = 0, 1, 2$ is in the range of $[0, 1]$ and the confidences sum up to one: $\sum_{k=0}^2 c_{ikD} = 1$. The barycentric coordinates correspond to the areas of the opposite triangles (see Fig. 4, left) and are obtained by the formulas:

$$c_{i0D} = 1 - x, \quad c_{i1D} = y, \quad c_{i2D} = x - y . \quad (1)$$

For a point (x, y) in the triangle, the second step results in the coordinate y multiplied by the complex double angle representation [17] of the local orientation:

$$z = y \exp(i2\theta) , \quad (2)$$

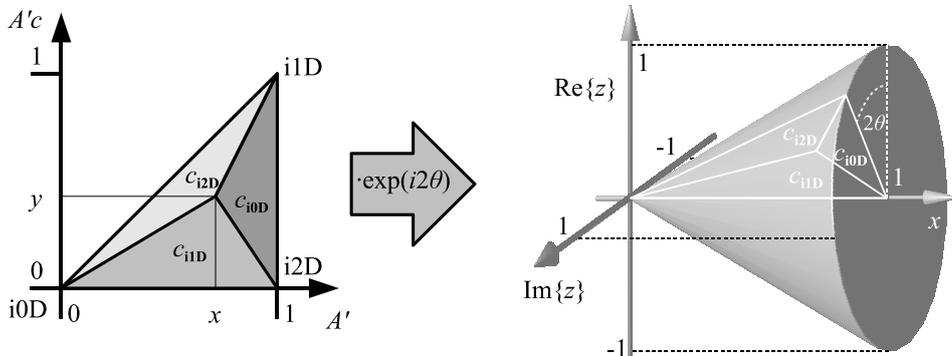


Fig. 4. Barycentric coordinates in a triangle (left) and in a cone (right).

where θ represents the local orientation. This modification leads to a parameterization of the iD cone, see Fig. 4, right.

If we want to extract the barycentric coordinates of a point in the iD triangle from the coordinates in the iD cone, we must project the 3D point onto the 2D slice which has the same orientation 2θ , see the white triangle in Fig. 4. This means we set $y = |z| \in [0, x]$ in (1).

3.2 Soft Thresholding

The iD cone and the space of structure tensors are both cone-shaped, but whereas the former is bounded, the latter is infinite, and can therefore not be used for a parameterization with barycentric coordinates. Still, we can use the structure tensor to estimate the iD cone by splitting the former into a magnitude part and a normalized tensor, mapping the magnitude to $[0, 1]$ by a monotonic function (soft threshold function), and multiplying the modified magnitude and the normalized tensor.

The soft threshold function is a non-linear mapping $g : \mathbb{R}^+ \rightarrow [0, 1] : m \mapsto g(m)$, similar to the activation function known from neural networks. We derive the soft threshold function using the conditional probabilities for the classes 'noise' C_{noise} and 'signal structure' C_{struct} , such that the soft threshold is the posteriori probability of membership of class 'signal structure' [5], Sect. 3.1.2:

$$\begin{aligned} g(m) &= P(C_{\text{struct}}|m) \\ &= \frac{p(m|C_{\text{struct}})P(C_{\text{struct}})}{p(m|C_{\text{struct}})P(C_{\text{struct}}) + p(m|C_{\text{noise}})P(C_{\text{noise}})} \\ &= \left(1 + \frac{p(m|C_{\text{noise}})P(C_{\text{noise}})}{p(m|C_{\text{struct}})P(C_{\text{struct}})}\right)^{-1}. \end{aligned} \quad (3)$$

What remains is to estimate the likelihoods for m and the class priors from image data. This is done by adjusting the parameters of a mixture of exponential distributions to the squared gradient magnitude.

In 2D the squared magnitude of the gradient response to noise is χ_2^2 distributed if we assume additive Gaussian noise [14]. As a result, the likelihood for C_{noise} is modelled according to:

$$p(m|C_{\text{noise}}) = \frac{1}{\mu_{\text{noise}}} \exp\left(-\frac{m}{\mu_{\text{noise}}}\right). \quad (4)$$

For natural images, the distribution of (squared) gradient magnitudes is typi-

cally described as *long-tailed* in the literature [37] and sometimes modeled as a Weibull distribution [16]¹. The Weibull distribution with $\gamma = 1$ is the same exponential distribution (4) as the χ_2^2 distribution. Hence, we approximate the squared gradient magnitude distribution for noisy images by a mixture of two exponential distributions.

The mixture parameters are obtained in a similar way as for a Gaussian mixture, using a variant of the EM algorithm, see e.g. [5], Sect. 2.6. The parameters are obtained by maximizing the likelihood or, equivalently, the log-likelihood:

$$(P_1^*, P_2^*, \mu_1^*, \mu_2^*) = \arg \max_{P_1, P_2, \mu_1, \mu_2} \sum_{n=1}^N \log(p_{\text{mixt}}(m_n)), \quad (5)$$

$$p_{\text{mixt}}(m) = \frac{P_1}{\mu_1} \exp\left(-\frac{m}{\mu_1}\right) + \frac{P_2}{\mu_2} \exp\left(-\frac{m}{\mu_2}\right), \quad (6)$$

where m_n , $n = 1 \dots N$ are samples of the squared gradient magnitude drawn from image data and $P_1, P_2 > 0$ sum to one: $P_1 + P_2 = 1$. Without loss of generality we assume that

$$P(C_{\text{noise}}) = P_1, \quad (7)$$

$$P(C_{\text{struct}}) = P_2 = 1 - P_1, \quad (8)$$

$$\mu_{\text{noise}} = \mu_1, \quad (9)$$

$$\mu_{\text{struct}} = \mu_2, \quad (10)$$

which is achieved in practice by initially selecting $P_1 > P_2$. With a similar derivation as in [5], Sect. 2.6, the equations which need to be iterated in the EM algorithm read

$$P_j^{\text{new}} = \frac{1}{N} \sum_{n=1}^N P^{\text{old}}(j|m_n), \quad (11)$$

$$\mu_j^{\text{new}} = \frac{\sum_{n=1}^N P^{\text{old}}(j|m_n)m_n}{\sum_{n=1}^N P^{\text{old}}(j|m_n)}, \quad (12)$$

where $P^{\text{old}}(j|m_n) = p^{\text{old}}(m_n|C_j)P_j^{\text{old}}/p_{\text{mixt}}^{\text{old}}(m_n)$. The algorithm typically converges within 10-20 iterations.

Plugging the obtained mixture into the soft threshold function (3) results in:

$$g(m) = \left(1 + \frac{P_1\mu_2}{P_2\mu_1} \exp\left(m\left(\frac{1}{\mu_2} - \frac{1}{\mu_1}\right)\right)\right)^{-1}, \quad (13)$$

¹ Note that the squared magnitude is addressed on p. 335 in this reference.

where the exponential is of negative kind ($\mu_{\text{noise}} = \mu_1 < \mu_{\text{struct}} = \mu_2$), i.e. we obtain a logarithmic sigmoid function for the soft-threshold, see [5], Sect. 3.1.3. The computation of the soft threshold is illustrated in Fig. 5.

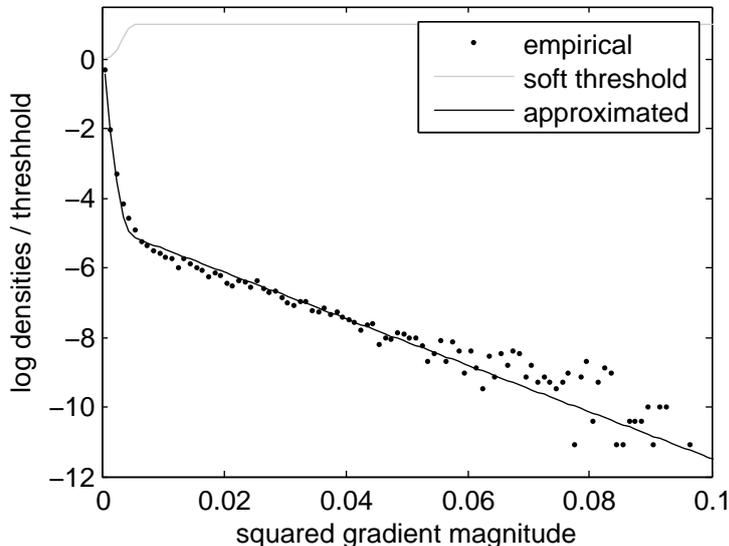


Fig. 5. The three plots show the empirical distribution of the squared gradient magnitude for a natural image (dots), the approximation by a mixture of two χ_2^2 distributions (black), and the final soft thresholding function (grey).

3.3 Estimation of iD

Due to the consisting embedding of the iD space, it is not necessary to estimate the structure tensor to begin with. Instead, we can directly map the gradients to the surface of the iD cone and do the averaging in the iD cone instead of the tensor. Instead of the gradient, we can use any other method for estimating the edge-ness or line-ness of a structure. We do not need an estimator for $i2D$ structures, since the averaging in the iD -cone model will result in such an estimate. The convexity of the model assures that we remain always inside the cone.

Consider for instance two measurements with perpendicular orientations but the same, large magnitude of the gradient. Averaging them will lead to a vanishing z -coordinate, such that we get a result close to $(1, 0)$, i.e., close to the $i2D$ vertex. This is appropriate, since two perpendicular and large gradients in the same neighborhood imply some kind of corner or junction.

When estimating the local orientation in areas with multiple orientation, one can either choose to combine the output of large, regularized linear filters, e.g. Gabor filter outputs, or to apply non-linear smoothing of small filter out-

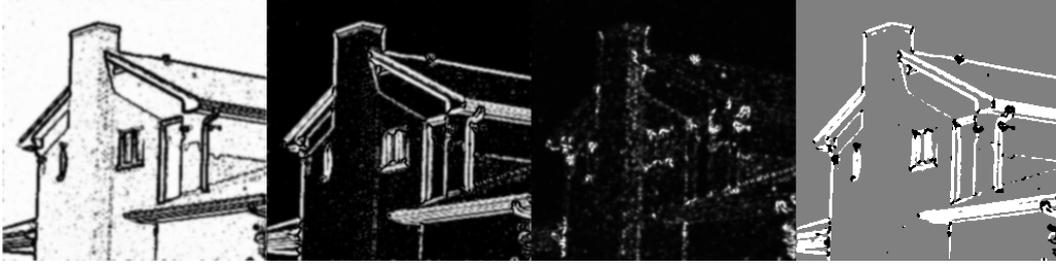


Fig. 6. Barycentric coordinates for the image in Fig. 1(a), black means zero and white means one. From left to right: c_{i0D} , c_{i1D} , c_{i2D} and highest confidence marked in gray, white and black for $i0D$, $i1D$ and $i2D$, respectively.

puts, e.g. centralized differences. For a constant scale of the effective overall operator, it has been shown that as much regularization as possible should be moved to the non-linear smoothing, i.e. the linear filter should be as small as possible [22], Sect. 3.4, which has also been confirmed in [10]. Hence, we use Scharr filters [43] in what follows, since this is the smallest isotropic gradient filter without displacement of the origin. Haar wavelets or 2-tap differences are smaller but lead to slightly larger errors repeating the experiment from [22].

For an image point $\mathbf{u} = (u_1, u_2)$, the proposed estimation method works as follows:

- (1) *Image gradient*: Extract (complex) gradient data $f = \sqrt{m(\mathbf{u})} \exp(i\theta(\mathbf{u}))$, $m(\mathbf{u})$ being the squared magnitude and $\theta(\mathbf{u})$ the orientation at pixel \mathbf{u} .
- (2) *Magnitude normalization and double angle representation*: Convert the gradient data to soft-thresholded double angle representation $d(\mathbf{u}) = g(m(\mathbf{u})) \exp(i2\theta(\mathbf{u}))$, $g(\cdot)$ being the soft threshold function (13), i.e., referring to Sect. 3.1, we have $x = y = g(m(\mathbf{u}))$.
- (3) *Cone representation*: Set the cone coordinates $\mathbf{c}(\mathbf{u}) = (c_1, c_2, c_3) = (|d|, \text{Re}\{d\}, \text{Im}\{d\})$
- (4) *Averaging*: Average the cone coordinates locally: $\mathbf{c}'(\mathbf{u}) = \sum_{\mathbf{i}} w_{\mathbf{i}} \mathbf{c}(\mathbf{i})$ where \mathbf{i} runs over the neighborhood of \mathbf{u} , and $w_{\mathbf{i}}$ is the two-dimensional Gaussian filter with appropriate σ .
- (5) *Triangle representation*: $(x^\Delta(\mathbf{u}), y^\Delta(\mathbf{u})) = (c'_1, \sqrt{(c'_2)^2 + (c'_3)^2})$
- (6) *Normalization of y values*: Since the cone is defined in polar coordinates and we integrate over the angle, we have to multiply with the Jacobean (the radius), i.e., $(\hat{x}(\mathbf{u}), \hat{y}(\mathbf{u})) = (x^\Delta(\mathbf{u}), y^\Delta(\mathbf{u})^2/x^\Delta(\mathbf{u}))$.
- (7) *Barycentric coordinates*: Extract barycentric coordinates from (\hat{x}, \hat{y}) according to (1).

The barycentric coordinates obtained from the described algorithm applied to the image in Fig. 1(a) are given in Fig. 6. The robustness of the method is illustrated by adding different amount of noise to the Lena image, see Fig. 7.

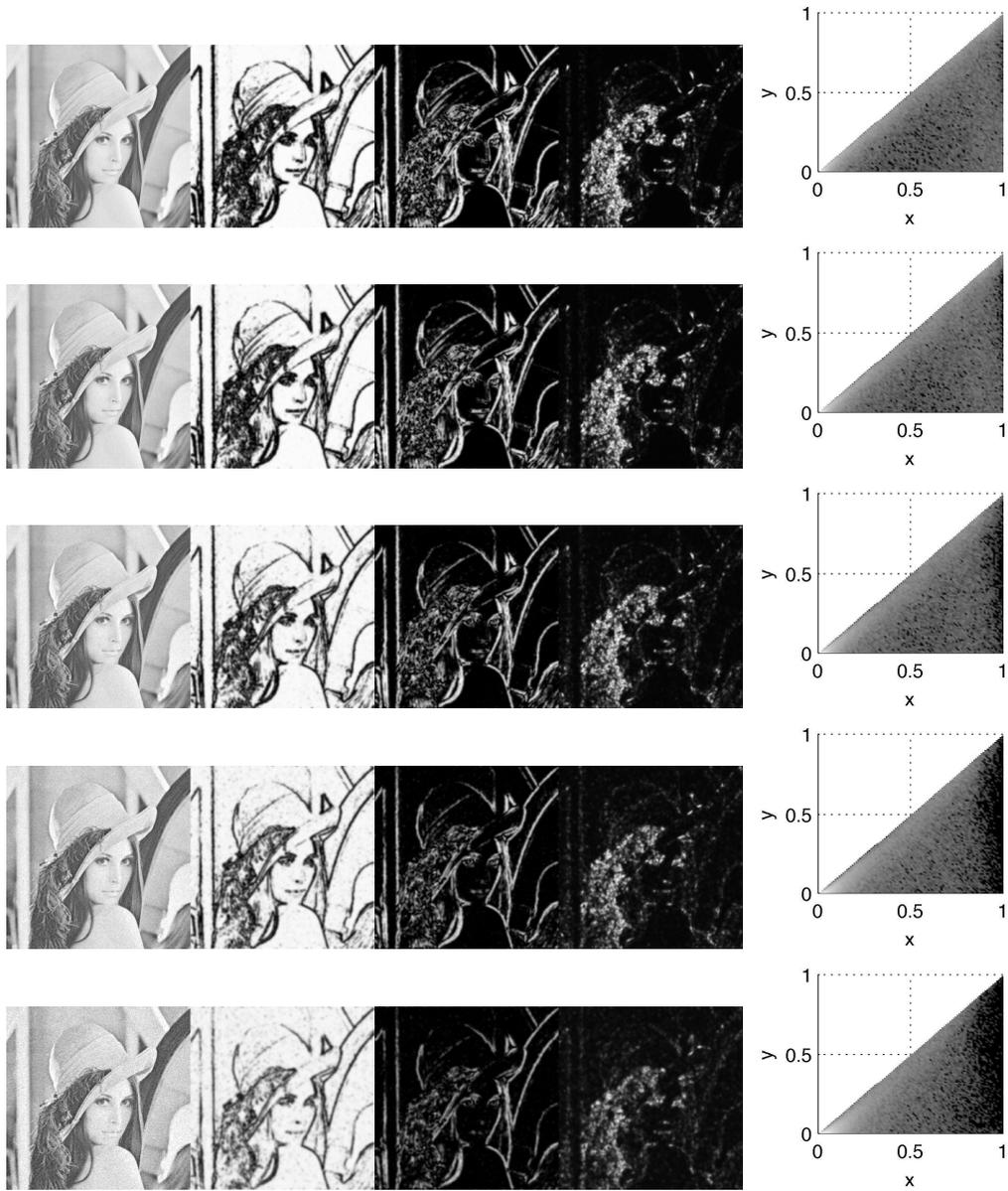


Fig. 7. The Barycentric iD coordinates for different amounts of noise. From left to right: noisy image, i0D, i1D, and density plot. From top to bottom: Peak signal to noise ration (PSNR) 40dB, 37dB, 33dB, 30dB, and 27dB.

4 Applications

The continuous understanding of intrinsic dimensionality and the arrangement of image structures in a triangle (or cone) has been used in a number of applications. Firstly, we have shown that certain classes of image structures (homogeneous patches, edges, corners and textures) become reflected in different areas of the iD-triangle, and therefore, the triangular representation is well suited to distinguish these structures (see section 4.1). Secondly, the approximate position of junctions can be characterized by the maximum of the intrinsic two-dimensional confidences (see section 4.1 and [11]). We further improve this positioning using a model-based approach based on our continuous understanding of intrinsic dimensionality (see section 4.1). Thirdly, we have made use of the triangular representation for an investigation of the relation of image structures to the underlying depth structures (section 4.3 and [24]). Finally, we have shown that the distribution of the quality of optic flow estimation reflects the distribution of image structures within the iD triangle (see section 4.4 and [23]).

4.1 Edges, Junctions and Textures

Fig. 1(b) shows the arrangement of certain sub-structures in the iD triangle for at least ten samples of each sub-structure that were manually chosen. Some image patches are displayed with a pointer to their position in the iD triangle. The figure shows that different sub-structures cover different sub-areas in the triangle: homogeneous image patches are at the i0D corner while edge-like structures are organized along a stripe along the hypotenuse of the triangle. We also see that the 'ideal' homogeneous patch as well as the ideal edge are mapped to the appropriate corners of the triangle². Textures cover an area with medium x and considerable y values. Finally, junctions show high x and low y values. There is good distinguishability but also some overlap, especially between junctions and textures.

The i2D confidences are natural indicators of junction structures (see Fig. 6-right). However, it is known that the positioning of junctions by local energy based operators without model knowledge is biased (see, e.g., [39]); for example, corners tend to become positioned at the inner part (see Fig. 8). One way to address this problem is to integrate model knowledge into the detection algorithm (see, e.g., [15,39]). A corner can be understood as intersection of edges and the point of intersection gives its position [36]. Our cone representation allows us to integrate such knowledge in an efficient way: We have defined a

² Note that the concept of an 'ideal' junction has not been defined in a complete sense in the literature.

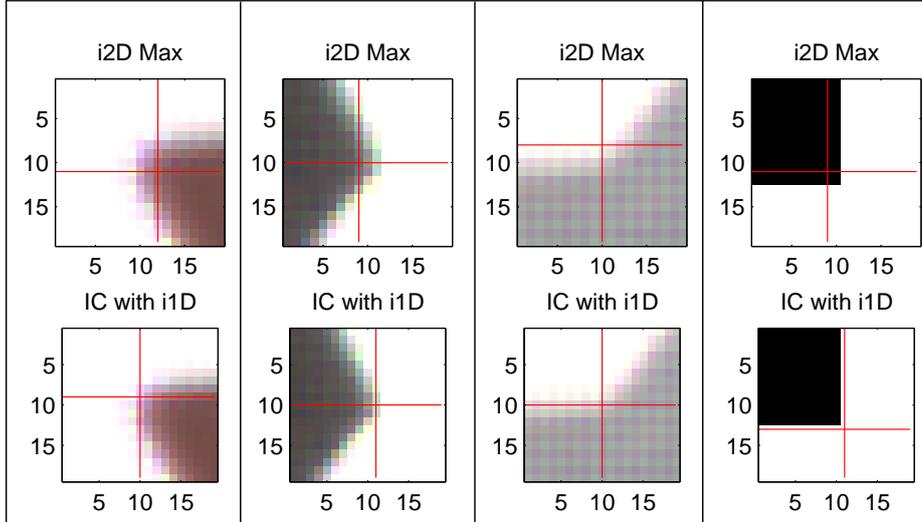


Fig. 8. From top to bottom: Location of the corners as found by i2D confidences (top), and weighted voting scheme (bottom).

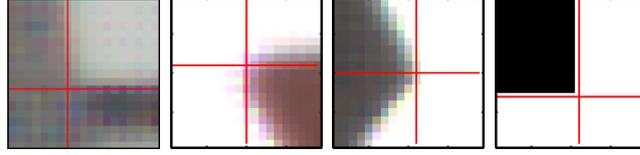
voting procedure that determines the positions of junctions at the intersection of lines in the vicinity. Basically, each point in a local neighbourhood in combination with the orientation computed at that pixel determines a line. The intersections of these lines results in a new position estimate for the corner (see Fig. 8). In our scheme, each of these lines acts as a weighted vote for the true position of the corner. Weighting is done according to the i1D confidence which expresses the likelihood that the signal represents an edge or a line:

$$ic(\mathbf{p}_c) = \int [c_{i1D}(\mathbf{p})]^2 \left[1 - \frac{d(l^{\mathbf{p}}, \mathbf{p}_c)}{d(\mathbf{p}, \mathbf{p}_c)} \right] d\mathbf{p}, \quad (14)$$

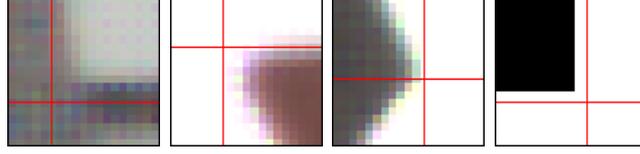
where \mathbf{p}_c is the center of the window; $l^{\mathbf{p}}$ is the line going through pixel \mathbf{p} with a slope defined according to the pixel orientation; $d(l^{\mathbf{p}}, \mathbf{p}_c)$ is the distance between $l^{\mathbf{p}}$ and \mathbf{p}_c ; and, $d(\mathbf{p}, \mathbf{p}_c)$ is the distance between \mathbf{p} and \mathbf{p}_c . The $ic()$ function is similar to the $\mathcal{R}()$ function of [36] and $S()$ function of [13], and it can be interpreted as a specific case of the Hough transform [21].

4.2 Advantage of Using Continuous Measure

To show the advantage of using a continuous measure in a specific example, we modified the $ic()$ function in (14) so that it combines the votes only from pixels whose i1D confidence is highest. The results, shown in Fig. 9, demonstrate that using continuous measures produces better results, especially when the junction is formed by noisy, texture-defined or low-contrast edges.



(a) $ic()$ with continuous i1D.



(b) $ic()$ with discrete labeling.

Fig. 9. Using continuous measures (a) and discrete measures (b) for the intersection consistency defined in (14). For noisy, texture-defined or low-contrast edges, using continuous measures is advantageous.

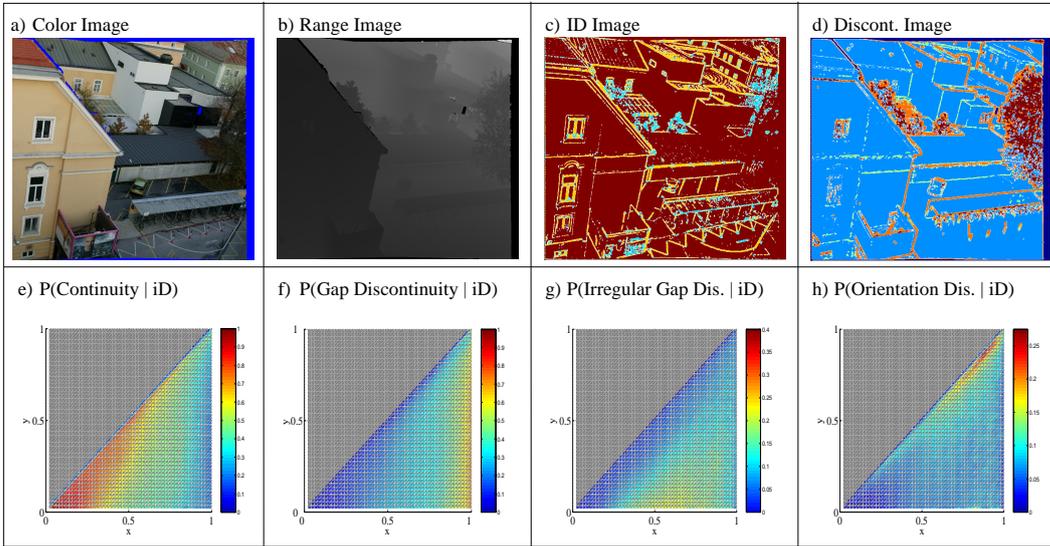


Fig. 10. [Figure to be printed in color] a) Colour image and b) the corresponding range map measured by a laser range finder. c) Maxima of intrinsic dimension confidences (brown: c_{0D} is maximal. yellow: c_{1D} is maximal. blue: c_{2D} is maximal). d) Extraction of different 3D structure (blue: continuous surfaces. green: orientation discontinuities. orange: depth discontinuities. brown: irregular depth discontinuities). e-h) Conditional probabilities of the respective 3D structures given the intrinsic dimension. e) continuous surfaces. f) depth discontinuities. g) irregular depth discontinuities. h) orientation discontinuities. Note that the ranges of (e-h) are not the same; we have not used the same scale in order to make the distributions more visible.

4.3 The Relation of Image Structures to Depth

By making use of our continuous iD representation, we investigated how local 3D structure is reflected in local image structure [24]. Using a set of 20 colour

images with known 3D ground truth³ (see Fig. 10(a,b)), we extracted 4 classes of 3D structures from the 3D ground truth: (1) Continuous surfaces (indicated as blue in Fig. 10(d)) are areas in which there are no depth discontinuities, and the underlying 3D structure does not vary or varies smoothly; (2) At orientation discontinuities (indicated in green in Fig. 10(d)), there exist two surfaces with rather different angle that meet in 3D without creating a depth discontinuity; (3) At gap discontinuities (indicated in orange in Fig. 10(d)), there exists significant depth discontinuities; and, finally (4), irregular gap discontinuities (indicated in brown in Fig. 10(d)) are observed when there is frequent change of depth (e.g., at trees or fences).

In Fig. 10(e-h), the empirical conditional probabilities for the respective 3D structures given the intrinsic dimensionality are shown. Homogeneous image structures reflect continuous surfaces.⁴ Also, most textures and many edges reflect continuous surfaces. In particular high contrast edges are caused by orientation discontinuities (see Fig. 10(h)). High contrast structures (located at the i1D-i2D side of the triangle) are to a high degree caused by gap discontinuities.

4.4 Evaluation of Optic Flow Estimates

We also investigated the quality of optic flow estimation. It is known that for certain image structures, there occur specific problems. For homogeneous image patches, basically no reliable flow estimates can be obtained, while for edges, we have to face the aperture problem. Using our continuous understanding of iD, we investigated these dependencies more closely in [23].

Fig. 11(b) shows a combined error measurement for optic flow estimation using the Nagel algorithm [35] on a set of 11 sequences with known ground truth (see Fig. 10(a,b)). The performance varies significantly with the intrinsic dimensionality. There is rather bad performance on homogeneous image areas and edges, and better performance for textures and junctions. The effect of the aperture problem becomes visible when we look at the deviation of the estimate from the true normal flow (11(c) which is very low at edges, and hence can be used for reliable line-correspondence finding (see, e.g., [29]).

³ The 3D ground truth was recorded by a laser range finder. On average, the images have a resolution of 1140x1000 pixels.

⁴ This was already stated (see, e.g., [19]) as 'no news is good news' and has been used in surface reconstruction/interpolation studies (see [6, 19]).

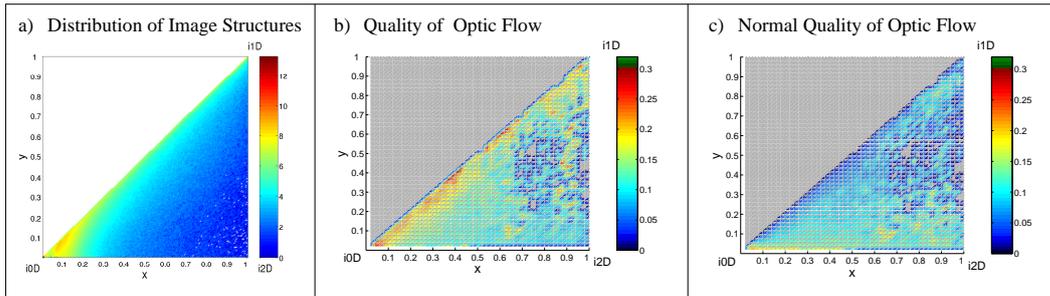


Fig. 11. [Figure to be printed in color] a) Logarithm of the distribution of image structures according to their intrinsic dimensionality. b) and c) Statistical evaluation of Nagel’s optic-flow estimation algorithm. b) Mean error of optic flow expressed by a combined measure evaluating angle and magnitude (also used in [1]) depending on the intrinsic dimensionality. c) Mean difference to normal flow depending on the intrinsic dimensionality.

5 Summary and Discussion

We introduced a new continuous representation for the intrinsic dimensionality of local image structures that is based on a cone representation which can be mapped in a straight forward manner to a triangle. We could demonstrate the usefulness of this representation for the distinction of different signal structure classes, for the junction detection and positioning problem, for the statistical investigation of local image structures to underlying depth structures as well as for the analysis of optic flow estimation algorithms.

The investigations described in this section make also the limits of local signal processing explicit. For example, it is not possible to uniquely distinguish junctions from other structures based on the information in a local image patch only. There are continuous transitions between these structures, and for this reason, final decisions should be postponed until later stages of the processing. Local descriptors such as the intrinsic dimensionality confidences, however, give good indications of the kind of local image structure and the post-processing that should be applied to it. In this way, it can be used as an initializer of processes taking also context information into account (see, e.g., [30, 38]).

Acknowledgement

The authors would like to thank Klas Nordberg, Gerald Sommer, Joachim Weickert and Florentin Wörgötter for fruitful discussions on the topic. We would also like to thank Riegl UK Ltd. (www.riegl.co.uk) for their supply of colour images with known ground truth.

References

- [1] J. Barron, D. Fleet, S. Beauchemin, Performance of optical flow techniques, *International Journal of Computer Vision* 12 (1) (1994) 43–77.
- [2] P. Bayerl, H. Neumann, Disambiguating visual motion by form–motion interaction — a computational model, *International Journal of Computer Vision* (Special Issue).
- [3] J. Bigün, Local symmetry features in image processing, Ph.D. thesis, Linköping University, Sweden, dissertation No 179, ISBN 91-7870-334-4 (1988).
- [4] J. Bigün, G. H. Granlund, Optimal orientation detection of linear symmetry, in: *Proceedings of the IEEE First International Conference on Computer Vision*, London, Great Britain, 1987, report LiTH-ISY-I-0828, Computer Vision Laboratory, Linköping University, Sweden, 1986.
- [5] C. M. Bishop, *Neural Networks for Pattern Recognition*, Oxford University Press, New York, 1995.
- [6] R. M. Bolle, B. C. Vemuri, On three-dimensional surface reconstruction methods, *IEEE Trans. Pattern Anal. Mach. Intell.* 13 (1) (1991) 1–13.
- [7] J. Bruske, G. Sommer, Intrinsic dimensionality estimation with optimally topology preserving maps, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 20 (5) (1998) 572–575.
- [8] H. S. M. Coexeter, *Introduction to Geometry*, 2nd ed., Wiley & Sons, 1969.
- [9] J. Costa, A. O. Hero, Geodesic entropic graphs for dimension and entropy estimation in manifold learning, *IEEE Transaction on Signal Processing* 52 (8) (2004) 2210–2221.
- [10] M. Felsberg, P.-E. Forssén, H. Scharf, Channel smoothing: Efficient robust smoothing of low-level signal features, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 28 (2) (2006) 209–222.
- [11] M. Felsberg, G. Granlund, POI detection using channel clustering and the 2D energy tensor, in: 26. DAGM Symposium Mustererkennung, Tübingen, 2004.
- [12] M. Felsberg, N. Krüger, A probabilistic definition of intrinsic dimensionality for images, in: 25. DAGM Symposium Mustererkennung, 2003.
- [13] W. Förstner, A framework for low level feature extraction, in: *ECCV '94: Proceedings of the third European conference on Computer Vision (Vol. II)*, Springer-Verlag New York, Inc., Secaucus, NJ, USA, 1994.
- [14] W. Förstner, Image preprocessing for feature extraction in digital intensity, color and range images, in: *Proceedings of the International Summer School on Data Analysis and the Statistical Foundation of Geomatics, Lecture Notes on Earth Sciences*, Springer, 1998.

- [15] W. Förstner, E. Gülch, A fast operator for detection and precise location of distinct points, corners and centres of circular features, in: ISPRS Intercommission Workshop, Interlaken, 1987.
- [16] J.-M. Geusebroek, The stochastic structure of images, in: Scale Space and PDE Methods in Computer Vision, vol. 3459 of LNCS, 2005.
- [17] G. H. Granlund, H. Knutsson, Signal Processing for Computer Vision, Kluwer Academic Publishers, Dordrecht, 1995.
- [18] P. Grassberger, I. Procaccia, Measuring the strangeness of strange attractors, *Physica D* 9 (1983) 189–208.
- [19] W. E. L. Grimson, Surface consistency constraints in vision, *Computer Vision, Graphics and Image Processing* 24 (1) (1983) 28–51.
- [20] C. G. Harris, M. Stephens, A combined corner and edge detector, in: 4th Alvey Vision Conference, 1988.
- [21] P. V. C. Hough, Methods and means for recognizing complex patterns, U.S. Patent 3,069,654, Dec. 18.
- [22] B. Johansson, Low level operations and learning in computer vision, Ph.D. thesis, Linköping University, Sweden, SE-581 83 Linköping, Sweden, dissertation No. 912, ISBN 91-85295-93-0 (December 2004).
- [23] S. Kalkan, D. Calow, F. Wörgötter, M. Lappe, N. Krüger, Local image structures and optic flow estimation, *Network: Computation in Neural Systems* 16 (4) (2005) 341–356.
- [24] S. Kalkan, F. Wörgötter, N. Krüger, Statistical analysis of local 3d structure in 2d images, in: IEEE Conference on Computer Vision and Pattern Recognition, 2006.
- [25] U. Köthe, Integrated edge and junction detection with the boundary tensor, in: Proc. of 9th Intl. Conf. on Computer Vision, Nice, vol. 1, 2003.
- [26] G. Krieger, C. Zetsche, Nonlinear image operators for the evaluation of local intrinsic dimensionality, *IEEE Transactions on Image Processing* 5 (6) (1996) 1026–1041.
- [27] N. Krüger, M. Felsberg, A continuous formulation of intrinsic dimension, in: British Machine Vision Conference, 2003.
- [28] N. Krüger, M. Felsberg, An explicit and compact coding of geometric and structural information applied to stereo matching, *Pattern Recognition Letters* 25 (8) (2004) 849–863.
- [29] N. Krüger, F. Wörgötter, Statistical and deterministic regularities: Utilisation of motion and grouping in biological and artificial visual systems, *Advances in Imaging and Electron Physics* 131 (2004) 82–147.

- [30] N. Krüger, F. Wörgötter, Multi-modal primitives as functional models of hyper-columns and their use for contextual integration, Proc. the 1st International Symposium on Brain, Vision and Artificial Intelligence 19-21 October, 2005, Naples, Italy, Lecture Notes in Computer Science, Springer.
- [31] A. B. Lee, K. S. Pedersen, D. Mumford, The nonlinear statistics of high-contrast patches in natural images, *International Journal of Computer Vision* 54 (1-3) (2003) 83–103.
- [32] E. Levina, P. Bickel, Maximum likelihood estimation of intrinsic dimension, in: *Advances in Neural Information Processing Systems 17*, MIT Press, 2005, pp. 777–784.
- [33] F. Mokhtarian, R. Suomela, Robust image corner detection through curvature scale space, *IEEE Trans. Pattern Analysis and Machine Intelligence* 20 (12) (1998) 1376–1381.
- [34] M. Mühlich, R. Mester, Subspace methods and equilibration in computer vision, in: *Proc. 12th Scandinavian Conference on Image Analysis*, 2001.
- [35] H.-H. Nagel, W. Enkelmann, An investigation of smoothness constraints for the estimation of displacement vector fields from image sequences, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 8 (1986) 565–593.
- [36] L. Parida, D. Geiger, R. Hummel, Junctions: detection, classification and reconstruction, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 20 (7) (1998) 687–698.
- [37] J. Portilla, V. Strela, J. Wainwright, E. P. Simoncelli, Image denoising using scale mixtures of Gaussians in the wavelet domain, *IEEE Trans. Image Processing* 12 (11) (2003) 1338–1351.
- [38] N. Pugeault, F. Wörgötter, N. Krüger, Multi-modal scene reconstruction using perceptual grouping constraints, in: *Workshop on Perceptual Organization in Computer Vision*, *IEEE CVPR*, 2006.
- [39] K. Rohr, Recognizing corners by fitting parametric models, *International Journal of Computer Vision* 9 (3) (1992) 213–230.
- [40] K. Shi, S. C. Zhu, Mapping the ensemble of natural image patches by explicit and implicit manifolds, *Computer Vision and Pattern Recognition (CVPR)*.
- [41] E. Sudderth, A. Ihler, W. Freeman, A. Willsky, Nonparametric belief propagation, in: *Proc. Conf. Computer Vision and Pattern Recognition*, 2003.
- [42] G. V. Trunk, Representation and analysis of signals: statistical estimation of intrinsic dimensionality and parameter identification, *General System* 13 (1968) 49–76.
- [43] J. Weickert, H. Scharr, A scheme for coherence-enhancing diffusion filtering with optimized rotation invariance, *Journal of Visual Communication and Image Representation*, Special Issue On Partial Differential Equations In Image Processing, *Computer Vision, And Computer Graphics* (2002) 103–118.

- [44] C. Zetsche, E. Barth, Fundamental limits of linear filters in the visual processing of two dimensional signals, *Vision Research* 30.