Optic Flow Statistics and Intrinsic Dimensionality

Authors: Sinan Kalkan[1], Dirk Calow[2], Michael Felsberg[3], Florentin Wörgötter[1], Markus Lappe[2], and Norbert Krüger[4]

[1] Psychology, University of Stirling, Scotland
{sinan,worgott}@cn.stir.ac.uk

[2] Psychology, University of Münster, Germany
{calow,mlappe}@psy.uni-muenster.de

[3] Electrical Eng., Linköping University, Sweden
mfe@isy.liu.se

[4] Computer Sci., Aalborg University Esbjerg, Denmark
nk@cs.aue.auc.dk

Abstract

Different kinds of visual sub-structures can be distinguished by the intrinsic dimensionality of the local signals. The concept of intrinsic dimensionality has been mostly exercised using discrete formulations. A recent work (Krüger and Felsberg, 2003; Felsberg and Krüger, 2003) introduced a continuous definition and showed that the inherent structure of the intrinsic dimensionality has essentially the form of a triangle. The current study work analyzes the distribution of signals according to the continuous interpretation of intrinsic dimensionality and the relation to orientation and optic flow features of image patches. Among other things, we give a quantitative interpretation of the distribution of signals according to their intrinsic dimensionality that reveals specific patterns associated to established sub-structures in computer vision. Furthermore, we link quantitative and qualitative properties of the distribution of optic-flow error estimates to these patterns.

1 Introduction

The intrinsic dimension (for convenience, will be called iD) (see, e.g., (Zetzsche and Barth, 1990; Felsberg, 2002)) has proven to be a suitable descriptor that distinguishes sub-structures (homogeneous pathces, edges or junctions). The association of iD to an image patch has been done mostly by a discrete classification (Zetzsche and Barth, 1990; Felsberg, 2002; Jähne, 1997). A *continuous* definition of iD has been recently given in (Krüger and Felsberg, 2003; Felsberg and Krüger, 2003). There, it has also been shown that *the topological structure of iD essentially has the form of a triangle* which is spanned by two axes corresponding to origin and line variance. In this paper, we will use this continuous definition to investigate the structure of the distribution of signals in natural images according to their iD. More specifically, we will show that;

- D0 i0D signals split into two clusters. One peak corresponding to saturated or dark patches and a Gaussian-shaped cluster corresponding to image noise at homogeneous but unsaturated/non–black image patches.
- D1 For i1D signals, there exists a concentration of signals in a stripe-shaped cluster corresponding to high origin variance (high amplitude) and low line variance. This reflects the importance of an orientation criterion that is based on local amplitude and orientation information (see, e.g., (Princen et al., 1990)).
- D2 In contrast to the i0D and i1D cases, there exists no such thing like a corner cluster for i2D signals in the distribution of local signals which indicates that it is rather difficult to formulate a local criterion to detect corners in natural images.

Thus, the continuous formulation of iD allows for a more precise characterization of established sub-structures in terms of their statistical manifestation in natural images. As a consequence, properties and inherent problems of classical computer vision algorithms can be reflected, and the limits of local signal processing can be made explicit.

In general, studies (e.g. (Barron et al., 1994; Mota and Barth, 2000)) that analyze optic flow estimation at different iD acknowledge that;

- A0 Optic flow estimates at homogeneous image patches is unreliable due to the fact that the lack of structure makes it impossible to find correspondences in consecutive frames.
- A1 Only normal flow can be computed for edge-like structures (aperture problem).
- A2 Only at i2D structures can true optic flow estimates be computed.

In this paper, we investigate these claims more closely. We will show that the continuous formulation of iD allows for a more quantitative investigation and characterization of the quality of optic flow estimation depending on local signal structures. More specifically, we will show that;

- Q0 There exist significantly more horizontal and vertical structures in natural images. However, the strength of this dominance depends crucially on the iD. Furthermore, we show that the distribution of orientation is directly reflected in the distribution of the estimates of optic-flow directions.
- Q1 Optic flow can be estimated reliably by looking at the normal flow in the stripe-shaped cluster in the i1D signal domain. This suggests that edges are a strong source of reliable information when the aperture problem is taken into account.
- Q2 The quality of optic flow estimation is higher for i2D signals. However, in analogy to the lack of a cluster for i2D signals, there exists a continuous signal domain (covering also sub-areas of i0D and i1D signals) for which a higher quality in the optic flow estimation can be achieved. The increase of the quality, on the other hand, is only slight which suggests that the role of i2D structures for motion estimation might be not as important as suggested by some authors (see, e.g., (Mota and Barth, 2000)). Another possibility is that the specific optic flow algorithm that we have chosen gives sub-optimal results at i2D signals.

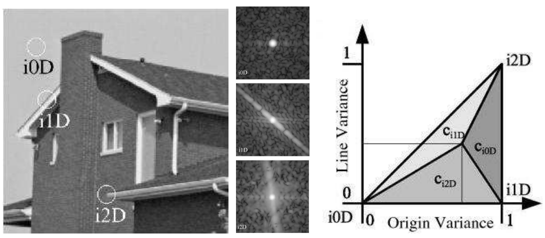
These results support the above-mentioned statements (A0)-(A2). However, by making use of a continuous understanding of iD, these statements can be made quantitatively more specific in terms of (1) characterization of sub-areas for which they hold and (2) their strength. Our analysis suggests a strong relationship between the distribution of the signals in the continuous iD space and properties of optic flow estimation.

The outline of the paper is as follows: In section 2, we introduce the concept of iD in its continuous definition, and we shortly describe our multi-modal image representation. In section 3, we investigate the distribution of natural image patches according to their iD and discuss its consequences. In section 4, we look at the statistics of orientation in natural images while in section 5, we look at the statistics of optic flow and the error of optic flow estimation.

2 Intrinsic Dimensionality and Multi-Modal Primitives

Intrinsic Dimensionality: The iD was introduced by (Zetzsche and Barth, 1990) and was used to formalise a *discrete distinction* between edge-like and junction-like structures. This corresponds to a classical interpretation of local image signals in computer vision. A large variety of edge and corner extraction algorithms have been developed over the last 20 years, and their role in artificial as well as biological systems has been discussed extensively (Koenderink and Dorn, 1982). The three kinds of signals have quite specific characteristics and problems that have been addressed in different contexts.

Recently, it has been shown that the topological structure of the iD must be understood as a triangle[1] that is spanned by two measures: origin variance and line variance. The origin variance describes the deviation of the energy from a concentration at the origin while the line variance describes the deviation from a line structure (see figure 1b,c).



(a) (b) (c)

Figure 1: Illustration of iD. **(a)** Three image patches for three different iDs. **(b)** The local spectra of the patches in (a), from top to bottom: i0D, i1D, i2D. **(c)** The topology of iD.

The triangular topological structure of the iD is counter-intuitive, in the first place, since it realizes a two-dimensional topology in contrast to a linear one-dimensional structure that is expressed in the discrete counting 0, 1 and 2. More importantly, as shown in (Krüger and Felsberg, 2003; Felsberg and Krüger, 2003), this triangular interpretation allows for a *continuous formulation* of iD in terms of 3 confidences assigned to each discrete case. This is achieved by first computing two measurements of origin and line variance.

[1] See (Krüger and Felsberg, 2003; Felsberg and Krüger, 2003).

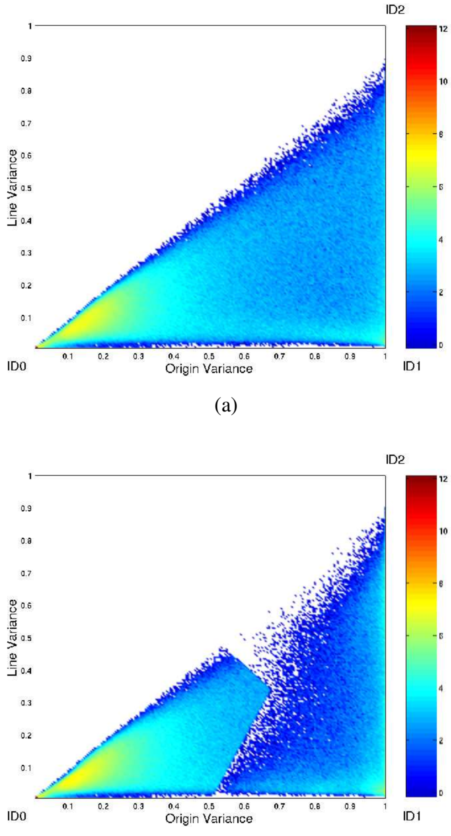
These two measurements define a point in the triangle (figure 1c). The bary-centric coordinates (see, e.g., (Coxeter, 1969)) of this point of the triangle directly lead to a definition of three confidences that add up to one. These three confidences reflect the volume of the 3 areas constituted by the points (figure 1c).

Multi–modal Primitives: In our research over the last 5 years (Krüger et al., 2004b), we have designed a novel image representation using multi-modal Primitives. They are condensed representations of local image structures using visual attributes. These Primitives are a functional abstraction of hyper-columnar structures found in the first cortical area of the human visual system (Hubel and Wiesel, 1969). The relevant attributes for this paper are: *Orientation:* The local orientation is computed using a new filter approach, the monogenic signal (Felsberg and Sommer, 2001). *Optic Flow:* After some comparison (Jäger, 2002), we decided to use the well-known optic flow technique (Nagel and Enkelmann, 1986).

3 Statistics of Intrinsic Dimensionality

We use a set of 7 natural sequences with 10 images each[2]. We compute 831727 Primitives in total. For the processing of each Primitive, we compute a measure for the origin and line variance[3]. This corresponds to one point in the triangle of figure 1c. Hence, taking all Primitives into account, we can display the distribution of these points in this triangular structure. This distribution is shown in figure 2a. As there exist large differences in the histogram, the logarithm is shown.

The distribution shows two main clusters. The peak close to the origin corresponds to low origin variance. It is visible that most of the signals that have low origin variance have high line variance. These correspond to nearly homogeneous image patches. Since the orientation is almost random for such homogeneous image patches, it causes high line variance. There is also a small peak existent at position $(0, 0)$ that corresponds to saturated/black image patches. The other cluster is for high origin variance signals with low line variance, corresponding to edge-like structures. The form of this cluster is a small horizontal stripe rather than a peak. There is a smooth decrease while approaching to the i2D area of the triangle. That means that there exists no cluster for corner-like structures like the ones for homogeneous image patches or edges. Along the origin variance axis, a continuous gap is observed. This gap suggests that there are no signals with zero variance. This is due to the fact that there is at least noise included in the image which causes some line variance.



(b)

Figure 2: Logarithmic plot of the distribution of iD. **(a)** The distribution when the positions are not modified according to iD. **(b)** The distribution when the positions are modified according to iD (See text for details).

We also see from the figure that there are far more i0D signals than i1D or i2D signals. So, we have much more homogeneous structures than edge-like or junction-like structures. Besides, it is clear that there are more i1D structures than i2D structures in natural images. The percentages of i0D, i1D and i2D structures are %86, %11 and %3, respectively. For computing these percentages, the iD of a Primitive is determined by taking the iD which has the highest confidence.

The results described above are taken from Primitives for which the position has been regularly sampled on a hexagonal grid. However, the position of an edge or a corner must be defined according to the internal structure of the signal. For example, we want an edge-like Primitive to be placed

[2]A representative of these sequences are at `http://www.cn.stir.ac.uk/~sinan/sequences/`

[3]For details see (Krüger and Felsberg, 2003).

directly on the edge; or, for a corner, in which a certain number of lines intersect, we want to have the Primitive placed on the intersection. We can achieve this positioning by making use of the local amplitude information in the image depending on the iD[4]. Note also that features such as phase and the optic flow depends on this positioning. When we determine the position of the Primitives for edges and corner-like structures accordingly, we get the distribution shown in figure 2b. It is qualitatively similar to the distribution achieved with regular sampling. However, since the position is determined depending on the local amplitude, there is a shift towards positions with higher amplitude that constitute the gaps at the border between i0D and the i1D and I2D signals. In the later stages of our analysis, we adopted this positioning.

4 Distribution of Orientation Depending on Intrinsic Dimensionality

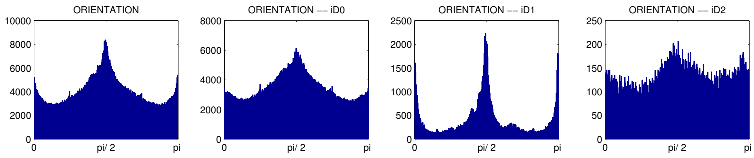


Figure 3: Orientation distribution depending on iD. The first image shows the total distribution.

The distribution of the orientation on the image sequences and the quantitative differences depending on the iD of the Primitives are shown in figure 3. Here, we define a signal to be i*n*D (where *n* is 0, 1 or 2) if the associated confidence is the highest. We see that orientation of i1D signals shows strong peaks at horizontal and vertical structures (i.e., for the values 0, $\pi/2$ and π). This has been already known (see, e.g., (Krüger, 1998)). However, the statistics in figure 3 for the i0D and i2D cases show the quantitative dependency of this statement for signals with different iD: These peaks are much weaker for the i0D and i2D case. Indeed, neither for a completely homogeneous image patch nor for a corner, the concept of orientation is defined at all. However, the continuous formulation of iD prevails that as dominance of horizontal and vertical orientations can be found also for i0D and i2D signals. This means that orientation is a meaningful concept also for some non-i1D signals. This also stresses the advantages of a continuous understanding of iD.

5 Optic Flow and Intrinsic Dimensionality

The distribution of magnitude and direction of the optic flow vectors is shown in figure 4. The quantitative errors in calculation of optic flow are shown in 5a-f using three different measurements.

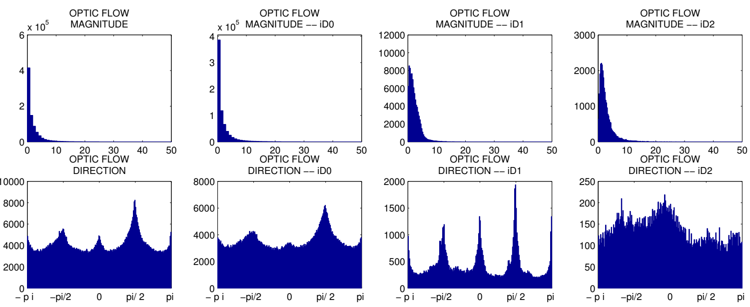
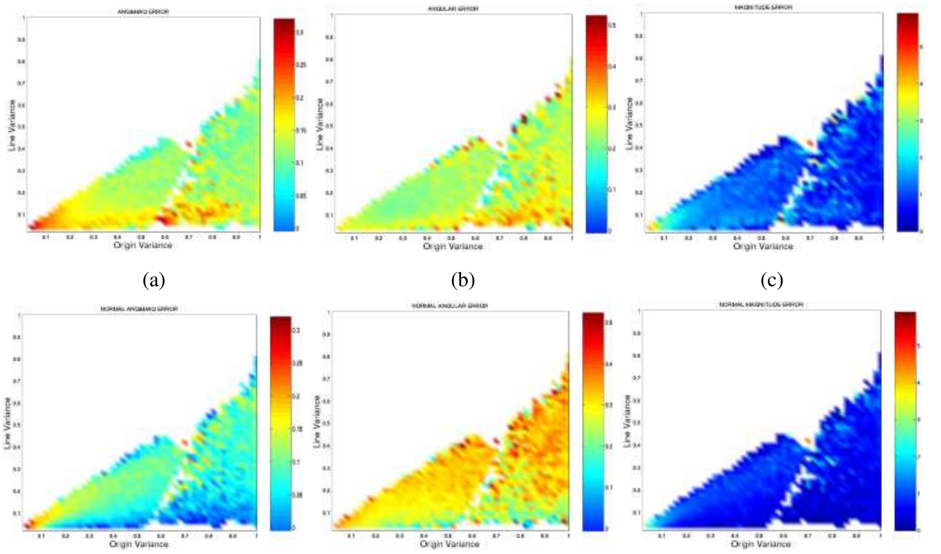


Figure 4: Optic flow distributions depending on ID. The left two images show total distributions.

The distribution of direction varies significantly with the iD. The distribution of the direction for i1D signals directly reflects the statistics of orientation. Since only the normal flow can be computed for pure i1D signals, the dominance of horizontal and vertical orientations (see section 4) leads to peaks at horizontal and vertical flows. The statistics of the true flow can be expected to be nearly homogeneous since in the sequences, a translational forward motion is dominant that leads to an isotropic flow field (see, e.g., (Lappe et al., 1999)). The fact that basically there exits a direct quantitative equivalence of the statistics of line structures, and the statistics of optic flow directions reflects the seriousness of the aperture problem. In contrast, the distribution of direction of optic flow vectors of i0D and i2D signals is nearly homogeneous.

[4]For details, see (Krüger et al., 2004a).



(a) (b) (c)

(d) (e) (f)

Figure 5: Optic flow qualities depending on iD. Color bars show the error values for corresponding colors of corresponding graphs. **(a)-(c)** Errors with the optic flow ground truth. **(d)-(f)** Errors with the projection of the ground truth over the normal vectors of the Primitives. **(a,d)** Errors computed using a mixture of magnitudes and angles. **(b,e)** Errors computed using angles. **(c,f)** Errors computed using magnitudes.

We now analyze the errors of the optic flow estimation depending on the iD. For this, we need to compare the computed flow with a ground truth. We used for our investigations the Brown Range Image Database (brid), a database of 197 range images collected by Ann Lee, Jinggang Huang and David Mumford at Brown University (Huang et al., 2000)[5].

The error of optic flow calculation is displayed in a histogram over the iD triangle (figure 5). For calculating the error, three different error functions were used: the angular error (figure 5b,e); the magnitudal error (figure 5c,f); and a mixture of both (figure 5a,d)[6].

Since for purely i1D signals, the aperture problem allows for the computation of the normal flow only, we have also computed errors with the projection of the ground truth over the normal vectors of the Primitives. For this, we first compute the normal vector of the Primitive using its orientation information; the ground truth is projected over this vector, and the error is computed between the optic flow of the Primitive and this projected ground truth.

The combined error computed using the original ground truth (figure 5a) is high for signals close to the i0D corner of the triangle as well as on the horizontal stripe from the i0D and i1D corner. In the other parts, there is a smooth surface that slightly decreases towards the i2D corner. This is in accordance with the notion that optic flow estimation at corner-like structures is more reliable than for edges and homogeneous image patches (A2). However, in figure 5a, it becomes obvious that the area where more reliable flow vectors can be computed is very broad and covers also traditional i0D and i1D signals. Furthermore, the decrease of error is rather slight which points to the fact that the quality of flow computation is limited in these areas, as well.

Since the current study makes use of a specific optic flow algorithm (Nagel and Enkelmann, 1986), it remains unclear whether the rather slight decrease of the error for i2D signals is a general

[5]The range images are recorded with a laser range-finder. Each image contains 44×1440 measurements with an angular separation of 0.18 degree. The field of view is 80 degree vertically and 259 degree horizontally. The distance of each point is calculated from the time of flight of the laser beam, where the operational range of the sensor is $2 - 200$m. The laser wavelength of the laser beam is $0.9\mu m$ in the near infrared region. Thus, the data of each point consist of 4 values, the distance, the horizontal angle and the vertical angle in spherical coordinates and a value for the reflected intensity of the laser beam. A representative of these sequences are at http://www.cn.stir.ac.uk/∼sinan/sequences/withgt/. The knowledge about the 3D data structure allows for a simulation of a moving camera in a scene and can therefore be used to estimate the correct flow for nearly all pixel positions of a frame of an image sequence.

[6]Angular error: $e(\mathbf{u},\mathbf{v}) = acos(\frac{\mathbf{u}\cdot\mathbf{v}}{|\mathbf{u}|.|\mathbf{v}|})$. Magnitudal error: $e(\mathbf{u},\mathbf{v}) = abs(|\mathbf{u}| - |\mathbf{v}|)$. Mixture error (Barron et al., 1994): $e(\mathbf{u},\mathbf{v}) = acos(\frac{\mathbf{u}\cdot\mathbf{v}+1}{(\mathbf{u}\cdot\mathbf{u}+1)(\mathbf{v}\cdot\mathbf{v}+1)})$.

property of optic flow estimates or is a specific property of the applied algorithm. Indeed, in the functional to be minimized in (Nagel and Enkelmann, 1986), the orientation is used to steer a smoothing (see also (Alvarez et al., 2000)). However, this orientation is ill–defined for i2D signals. Another conclusion could be (when we acknowledge that corners are important for human motion estimation) that the concept of a corner for humans is more sophisticated than simply measuring line and orientation variance by a *local* operator. These are research issues that have to be addressed in future works.

For the combined error computed using the normal ground truth (figure 5d), a different picture occurs. The error is very low for a horizontal stripe from the middle point between the i0D and i1D corners to the i1D corner. When compared to figure 5a, this figure reflects the strength of the aperture problem. On the other hand, it also shows the quality of optic flow estimation when the aperture problem is taken into account. The information for such signals can be of great importance, since for example, constraints for global motion estimation can be defined on line correspondences[7], i.e., correspondences that only require normal flow. In figure 5, the angular and magnitude errors are also displayed. They reflect a similar behaviour, especially visible for the angular error.

References

Alvarez, L., Weickert, J., and Sanchez, J. (2000). Reliable estimation of dense optical flow fields with large displacements. *International Journal of Computer Vision*, 39:41–56.

Barron, J., Fleet, D., and Beauchemin, S. (1994). Performance of optical flow techniques. *International Journal of Computer Vision*, 12(1):43–77.

Coxeter, H. (1969). *Introduction to Geometry (2nd ed.)*. Wiley & Sons.

Felsberg, M. (2002). *Low-Level Image Processing with the Structure Multivector*. PhD thesis, Institute of Computer Science and Applied Mathematics, Christian-Albrechts-University of Kiel.

Felsberg, M. and Krüger, N. (2003). A probablistic definition of intrinsic dimensionality for images. *Pattern Recognition, 24th DAGM Symposium*.

Felsberg, M. and Sommer, G. (2001). The monogenic signal. *IEEE Transactions on Signal Processing*, 49(12):3136–3144.

Huang, J., Lee, A., and Mumford, D. (2000). Statistics of range images. *CVPR*, pages 1324–1331.

Hubel, D. and Wiesel, T. (1969). Anatomical demonstration of columns in the monkey striate cortex. *Nature*, 221:747–750.

Jäger, T. (2002). Interaktion verschiedener visueller Modalitäten zur stabilen Extraktion von objektrepräsentationen. *Diploma thesis (University of Kiel)*.

Jähne, B. (1997). *Digital Image Processing – Concepts, Algorithms, and Scientific Applications*. Springer.

Koenderink, J. and Dorn, A. (1982). The shape of smooth objects and the way contours end. *Perception*, 11:129—173.

Krüger, N. (1998). Collinearity and parallelism are statistically significant second order relations of complex cell responses. *Proceedings of I&ANN 98*.

Krüger, N. and Felsberg, M. (2003). A continuous formulation of intrinsic dimension. *Proceedings of the British Machine Vision Conference*.

Krüger, N., Felsberg, M., and Wörgötter, F. (2004a). Processing multi-modal primitives from image sequences. *Fourth International ICSC Symposium on ENGINEERING OF INTELLIGENT SYSTEMS*.

Krüger, N., Lappe, M., and Wörgötter, F. (2004b). Biologically motivated multi-modal processing of visual primitives. *The Interdisciplinary Journal of Artificial Intelligence and the Simulation of Behaviour*, 1(5).

Krüger, N. and Wörgötter, F. (2004). Statistical and deterministic regularities: Utilisation of motion and grouping in biological and artificial visual systems. *Advances in Imaging and Electron Physics*, 131.

Lappe, M., Bremmer, F., and van den Berg, A. V. (1999). Perception of self-motion from visual flow. *Trends in Cognitive Sciences*, 3:329–336.

Mota, C. and Barth, E. (2000). On the uniqueness of curvature features. *Proc. in Artificial Intelligence*, 9:175–178.

Nagel, H.-H. and Enkelmann, W. (1986). An investigation of smoothness constraints for the estimation of displacement vector fields from image sequences. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8:565–593.

Princen, J., Illingworth, J., and Kittler, J. (1990). An optimizing line finder using a Hough transform algorithm. *Computer Vision, Graphics, and Image Processing*, 52:57–77.

Rosenhahn, B. (2003). *Pose Estimation Revisited (PhD Thesis)*. Institut für Informatik und praktische Mathematik, Chrsitian–Albrechts–Universität Kiel.

Zetzsche, C. and Barth, E. (1990). Fundamental limits of linear filters in the visual processing of two dimensional signals. *Vision Research*, 30.

[7] See, e.g., (Rosenhahn, 2003; Krüger and Wörgötter, 2004).