



Local Mutual Information for Dissimilarity-Based Image Segmentation

Lionel Gueguen, Santiago Velasco-Forero, Pierre Soille

► To cite this version:

Lionel Gueguen, Santiago Velasco-Forero, Pierre Soille. Local Mutual Information for Dissimilarity-Based Image Segmentation. *Journal of Mathematical Imaging and Vision*, 2014, 48 (3), pp.625-644. hal-01110199

HAL Id: hal-01110199

<https://hal.science/hal-01110199>

Submitted on 28 Jan 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Local mutual information for dissimilarity-based image segmentation

Lionel Gueguen¹, Santiago Velasco-Forero³,
and Pierre Soille²

the date of receipt and acceptance should be inserted later

Abstract Connective segmentation based on the definition of a dissimilarity measure on pairs of adjacent pixels is an appealing framework to develop new hierarchical segmentation methods. Usually, the dissimilarity is fully determined by the intensity values of the considered pair of adjacent pixels, so that it is independent of the values of the other image pixels. In this paper, we explore dissimilarity measures depending on the overall image content encapsulated in its local mutual information and show its invariance to information preserving transforms. This is investigated in the framework of the connective segmentation and constrained connectivity paradigms and leads to the concept of dependent connectivities. An efficient probability estimator based on depth functions is proposed to handle multi-dimensional images. Experiments conducted on hyper-spectral and multi-angular remote sensing images highlight the robustness of the proposed approach.

version as of January 24, 2013

1 Introduction

A digital image can be represented by a undirected simple finite graph whose node set matches the set of picture elements (laying in an Euclidean space) and whose edge set corresponds to the set of unordered pairs of adjacent picture elements. The segmentation of an image can be defined as its partition into path-connected

This research was funded by the JRC Specific Programme of European Commission's Seventh Framework Programme for Research and Technological Development (FP7). It was undertaken under the work programme of the Geo-Spatial Information Analysis for Security and Stability action, Global Security and Crisis Management unit, Institute for the Protection and Security of the Citizen. This study was performed whilst L. Gueguen was with the Joint Research Centre of the European Commission.

¹DigitalGlobe Inc., R&D Departement, Longmont CO, U.S.A.

²European Commission, Joint Research Centre, Ispra (VA), Italy

²Fraunhofer Institute for Industrial Mathematics (ITWM), Kaiserslautern, Germany

components given an underlying adjacency graph. Therefore the number of possible segmentations of an image containing n pixels is equal to its Bell number B_n if the complete graph is considered. For non-complete graphs such as the usual 4- or 8-adjacency graphs, the number of possible segmentations is smaller than B_n but is still intractable even for small values of n . In this context, segmentation methods delivering a hierarchy of nested connected components represented by a tree rather than a unique partition allow adaptability while remaining computationally tractable.

Several studies proved that a unique hierarchy of nested segmentations can be obtained through a parametric connectivity relation defined along paths of adjacent pixels of the image grid. A common connectivity relation states that two pixels of a gray-scale image are connected if there exists a path of pixels linking these pixels such that absolute differences (or any other dissimilarity measure) between successive pixels of this path (that is, weights of the edges of the path) do not exceed a given threshold value. This idea has been introduced in image processing in [Nagao et al., 1979], and further extended in numerous segmentation methods [Arbelaez and Cohen, 2006]. A hierarchy of fine to coarse partitions is simply obtained by gradually increasing the dissimilarity threshold value. The single linkage clustering method, described earlier in pattern recognition [Sibson, 1973], is using this property to encode efficiently the nested partitions into a dendrogram. Single-linkage is known to suffer from leakage through elements of transitional values. In the case of images, leakage occurs at pixels with intermediate values at the boundary of two homogeneous regions in the case of images. To counter this effect, the framework of constrained connectivity was introduced in [Soille, 2008] to reject the formation of connected components violating an internal homogeneity constraint such as an upper bound on the range of values of the connected component.

The diversity existing in the hierarchy of partitions depends strongly on the choice of the underlying dissimilarity measure. Most dissimilarity measures computed between two adjacent pixels are fully defined by the values of these pixels. This is the case for the L_m norm. However, in some applications, measures driven by the values of additional pixels may be of interest. For example, the dissimilarity presented in [Soille, 2011] takes the maximum between the L_1 norm and the point-wise minimum of the gradients by erosion and dilation calculated for each pair of adjacent pixels. Hence, this dissimilarity measure between two adjacent pixels involves not only the values of these pixels but also those of their neighbourhood. Image connectivities driven by the image content go beyond the consideration of a limited neighbourhood by taking into account statistics computed for the whole image definition domain (or within a subdomain large enough to estimate these statistics). By doing so, the resulting connectivity relations adapts automatically to the image information content. This idea was recently followed in [Gueguen and Soille, 2011] and led to the notion of frequent and dependent connectivities. With frequent connectivity, the dissimilarity assigned to the edge linking two adjacent pixels is inversely proportional to the frequency of co-occurrence of the values of these pixels. With dependent connectivity, their dissimilarity is inversely proportional to their statistical dependence.

In this paper, we extend our initial results [Gueguen and Soille, 2011] and detail the presentation of the local mutual information as a measure of dependence between the values of adjacent pixels. The local mutual information is based on the joint probability of adjacent pixel values and enables the definition of a degree

of dependence between them. This metric is inspired from the mutual information measuring the dependence between random variables [Cover and Thomas, 1991], which was used successfully in image registration [Viola and Wells, 1997, Maes et al., 1997, Hermosillo et al., 2002, Karacali, 2007] and for template-based object tracking [Panin and Knoll, 2008, Sun and Hoogs, 2010]. We highlight the invariance of the local mutual information to any differentiable bijective transform applied pixelwise to the full image. This property is of main interest for developing segmentation techniques that are invariant to unknown information preserving transforms. Having the capacity to define a joint probability between pixels values, the local mutual information is used to define a dissimilarity measure and its corresponding hierarchy of nested connected components. Additionally, a homogeneity constraint defined in terms of the component entropies is proposed to extend dependent connectivity to the constrained connectivity framework (i.e., by defining the highest non-horizontal cut of the dependent connectivity dendrogram such that every segment of the resulting partition satisfies the proposed homogeneity constraint). Furthermore, the local mutual information is adapted to handle multi-band images. Indeed as the dimension (number of bands) increases, the estimation of the joint probability distribution becomes less robust and requires the use of proper tools. The depth function method [Zuo and Serfling, 2000, Zuo, 2003] is used as an estimator of joint probabilities, as it avoids the assumption of a distribution shape, while maintaining a high approximation power for the Gaussian distribution. Based on this probability estimation method, an approximation of the local mutual information is proposed. Experiments are conducted on simulated data as well as hyper-spectral and multi-angular remote sensing images. They highlight the practical relevance of the depth function and the local mutual information for segmentation purposes. The experiment on simulated data permits to assess the robustness of the depth function, thus the robustness of the local mutual information to unknown joint probability distributions. After this, the experiment on a hyper-spectral image shows that the dependence measure suffers less from the leakage effect than the usual norm based dissimilarity. The last experiment is performed with multi-angular multi-spectral images. As the acquisition angle deviates from nadir, the illumination acquired by the sensor is modified in an uncontrolled way, which can be modelled by an unknown bijective transform. The results show that the segmentations based on depth function and local mutual information are less sensitive to varying acquisition angles.

The paper is structured as follows. The concept of local mutual information and its analytical formulation are presented in Sec. 2 while demonstrating its invariance to bijective transforms and analysing its sensitivity to additive noise. The notion of dependent connectivity and its extension to the constrained connectivity framework are developed in Sec. 3. Section 4 describes the depth function as a probability estimator enabling the approximation of the local mutual information in high dimensions. Experiments are conducted on simulated as well as hyper-spectral and multi-angular images in Sec. 5. Conclusions are drawn in Sec. 6.

2 Local Mutual Information

The local mutual information is increasingly used in the literature for language processing [Church and Hanks, 1990, Bouma, 2009], finance modeling [Cherubini

et al., 2004], and change detection [Gueguen and Datcu, 2009, Gueguen et al., 2011]. The concept of local mutual information and its analytical formulation are recalled in Sec. 2.1 and Sec. 2.2. We then demonstrate its invariance to bijective transforms and analyse its sensitivity to additive noise in Sec. 2.3 and Sec. 2.4 respectively.

2.1 Definition

The *mutual information* I is defined between two random variables X, Y thanks to their joint probability distribution $p_{\mathbf{X}, \mathbf{Y}}$ [Cover and Thomas, 1991]:

$$I(\mathbf{X}; \mathbf{Y}) = \int p_{\mathbf{X}, \mathbf{Y}}(\mathbf{x}, \mathbf{y}) \log \frac{p_{\mathbf{X}, \mathbf{Y}}(\mathbf{x}, \mathbf{y})}{p_{\mathbf{X}}(\mathbf{x})p_{\mathbf{Y}}(\mathbf{y})} d\mathbf{x}d\mathbf{y}, \quad (1)$$

where $p_{\mathbf{X}}(\mathbf{x})p_{\mathbf{Y}}(\mathbf{y})$ is the product of their marginal distributions:

$$p_{\mathbf{X}}(\mathbf{x}) = \int p_{\mathbf{X}, \mathbf{Y}}(\mathbf{x}, \mathbf{y}) d\mathbf{y}, \quad (2)$$

$$p_{\mathbf{Y}}(\mathbf{y}) = \int p_{\mathbf{X}, \mathbf{Y}}(\mathbf{x}, \mathbf{y}) d\mathbf{x}. \quad (3)$$

The mutual information measures the quantity of shared information by the variables, and assesses their statistical dependence. However, this measure is global and recent work proposed to consider the local mutual information [Church and Hanks, 1990, Winter et al., 1997] to attribute a dependence measurement to each realization.

Let \mathbf{X}, \mathbf{Y} be two random variables linked by their joint probability distribution $p_{\mathbf{X}, \mathbf{Y}}$ and let \mathbf{x}, \mathbf{y} be two realizations of the joint random variable. The *local mutual information* i is a function of the realizations expressed by [Gueguen et al., 2011]:

$$i(\mathbf{x}; \mathbf{y}) = \log \frac{p_{\mathbf{X}, \mathbf{Y}}(\mathbf{x}, \mathbf{y})}{p_{\mathbf{X}}(\mathbf{x})p_{\mathbf{Y}}(\mathbf{y})}, \quad (4)$$

$$= \log \frac{p_{\mathbf{X}|\mathbf{Y}}(\mathbf{x} | \mathbf{y})}{p_{\mathbf{X}}(\mathbf{x})}, \quad (5)$$

$$= \log \frac{p_{\mathbf{Y}|\mathbf{X}}(\mathbf{y} | \mathbf{x})}{p_{\mathbf{Y}}(\mathbf{y})}. \quad (6)$$

The expectation of the local mutual information gives back the mutual information:

$$I(\mathbf{X}; \mathbf{Y}) = \mathbb{E}_{\mathbf{X}, \mathbf{Y}} [i(\mathbf{X}; \mathbf{Y})]. \quad (7)$$

Informally, the local mutual information compares the probability of observing \mathbf{x} and \mathbf{y} together with the probability of observing \mathbf{x} and \mathbf{y} by chance. In the framework of Bayesian test of independence [Palomar and Verdu, 2008], the local mutual information is understood as a test statistic to decide if the realizations \mathbf{x} and \mathbf{y} are dependent or independent with respect to a known joint probability distribution $p_{\mathbf{X}, \mathbf{Y}}$. This metric can be normalized as described in [Bouma, 2009].

While the mutual information is a symmetric function defined between two random variables, the local mutual information is not necessarily symmetric. Indeed, the local mutual information is symmetric if and only if the joint probability distribution is itself symmetric: $i(\mathbf{x}; \mathbf{y}) = i(\mathbf{y}; \mathbf{x}) \Leftrightarrow p_{\mathbf{X}, \mathbf{Y}}(\mathbf{x}, \mathbf{y}) = p_{\mathbf{X}, \mathbf{Y}}(\mathbf{y}, \mathbf{x})$.

2.2 Local Mutual Information of Gaussian Variables

Let assume two multidimensional random variables \mathbf{X} and \mathbf{Y} , such that the composite $\mathbf{Z} = [\mathbf{X}, \mathbf{Y}]$ follows a Gaussian probability distribution $\mathcal{N}(\mu_{\mathbf{Z}}, \Sigma_{\mathbf{Z}})$. By marginalisation, the variables \mathbf{X} and \mathbf{Y} follow the Gaussian distributions $\mathcal{N}(\mu_{\mathbf{X}}, \Sigma_{\mathbf{X}})$ and $\mathcal{N}(\mu_{\mathbf{Y}}, \Sigma_{\mathbf{Y}})$, respectively. The functions parameters are linked by:

$$\mu_{\mathbf{Z}} = [\mu_{\mathbf{X}}, \mu_{\mathbf{Y}}], \quad (8)$$

$$\Sigma_{\mathbf{Z}} = \begin{pmatrix} \Sigma_{\mathbf{X}} & \Sigma_{\mathbf{X}\mathbf{Y}} \\ \Sigma_{\mathbf{Y}\mathbf{X}} & \Sigma_{\mathbf{Y}} \end{pmatrix}, \quad (9)$$

where $\Sigma_{\mathbf{X}\mathbf{Y}} = \Sigma_{\mathbf{Y}\mathbf{X}}^T$ are the cross correlation matrices. The variable \mathbf{W} following the equivalent joint independent distribution $p_{\mathbf{X}} \times p_{\mathbf{Y}}$ is also a Gaussian variable $\mathcal{N}(\mu_{\mathbf{W}}, \Sigma_{\mathbf{W}})$ such that:

$$\mu_{\mathbf{W}} = \mu_{\mathbf{Z}}, \quad (10)$$

$$\Sigma_{\mathbf{W}} = \begin{pmatrix} \Sigma_{\mathbf{X}} & 0 \\ 0 & \Sigma_{\mathbf{Y}} \end{pmatrix}. \quad (11)$$

Proposition 1 *The local mutual information of any realization $\mathbf{z} = [\mathbf{x}, \mathbf{y}]$ of the Gaussian variable $\mathbf{Z} = [\mathbf{X}, \mathbf{Y}]$ is expressed by :*

$$i_G(\mathbf{x}; \mathbf{y}) = \frac{1}{2}(\mathbf{z} - \mu_{\mathbf{Z}})^T (\Sigma_{\mathbf{W}}^{-1} - \Sigma_{\mathbf{Z}}^{-1})(\mathbf{z} - \mu_{\mathbf{Z}}) - \frac{1}{2} \log \frac{\det \Sigma_{\mathbf{Z}}}{\det \Sigma_{\mathbf{W}}}. \quad (12)$$

This formulation is essentially proposed by [Theiler and Perkins, 2006] while disregarding the scaling and constant factors, and is also referred to by [Cherubini et al., 2004]. The local mutual information of Gaussian variables is illustrated in Figs. 1–2. It can be interpreted as a hyperbolic paraboloid shaped like a saddle, where the saddle point is localised at the mean $\mu_{\mathbf{Z}}$. The realization dependence decreases as the points get further from the maximum correlation hyperplane passing by the mean and defined by the covariance matrix $\Sigma_{\mathbf{Z}}$. On the contrary, the realization dependence increases along this hyperplane as the realizations get further from the mean $\mu_{\mathbf{Z}}$.

2.3 Invariance to Bijective Transform

The local mutual information is invariant to any bijective transform applied marginally [Gueguen et al., 2012]. Indeed a deterministic transform should not change the statistical dependence of the transformed realizations.

Proposition 2 *Let b be a bijective and differentiable function defined between two multidimensional random variables \mathbf{X} and \mathbf{Z} , such that the realizations are deterministically linked by $\mathbf{z} = b(\mathbf{x})$. Then, for any third random variable \mathbf{Y} and any of its realizations \mathbf{y} , we have $i(b(\mathbf{x}); \mathbf{y}) = i(\mathbf{x}; \mathbf{y})$.*

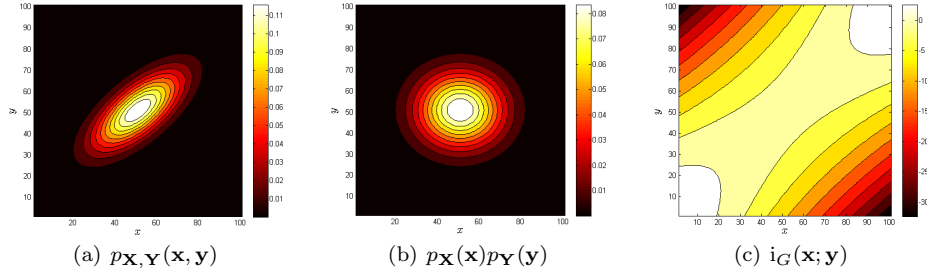


Fig. 1 The joint probability distribution of a 2-dimensional Gaussian variable is represented in (a) with its corresponding independent joint distribution in (b). The corresponding local mutual information expressed in (12) is depicted in (c).

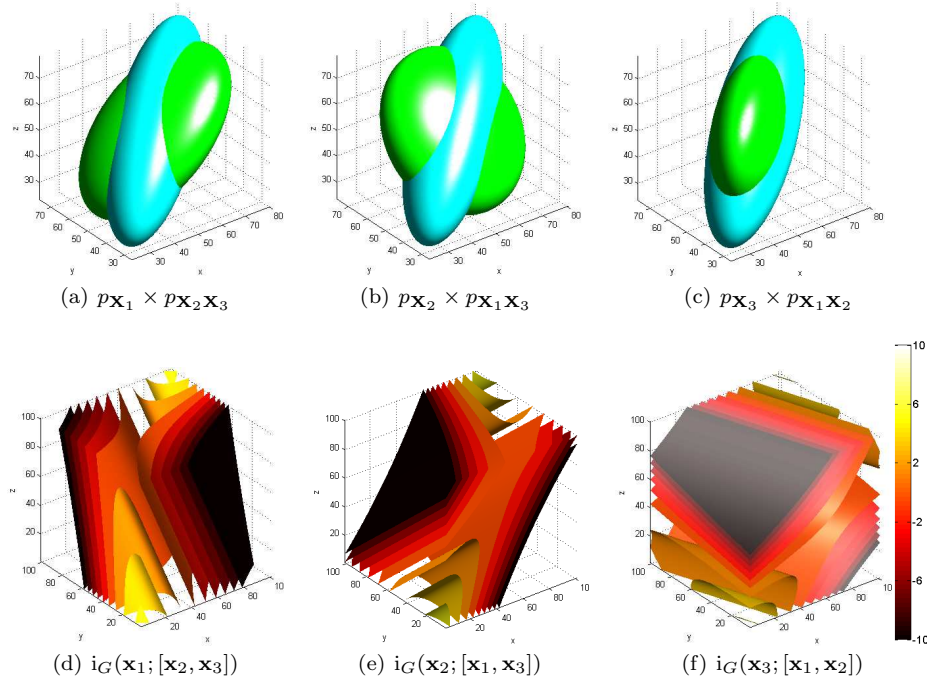


Fig. 2 The top row (a)–(c) represents the isosurfaces of the joint distribution $p_{\mathbf{X}_1 \mathbf{X}_2 \mathbf{X}_3}$ in blue superimposed on the corresponding possible joint independent distributions displayed in green. The bottom row (d)–(f) represents several isosurfaces of the respective local mutual information function as defined by (12).

Proof The change of variable condition states that for any bijective differentiable function b the probability distributions are linked by:

$$p_{\mathbf{Z}}(\mathbf{z})d\mathbf{z} = p_{\mathbf{X}}(\mathbf{x})d\mathbf{x}, \quad (13)$$

$$p_{\mathbf{Z}}(\mathbf{z}) = p_{\mathbf{X}}(b^{-1}(\mathbf{z})) \left| \det \left(\frac{d\mathbf{x}}{d\mathbf{z}} \right) \right|, \quad (14)$$

where $\frac{d\mathbf{x}}{d\mathbf{z}}$ is the Jacobian of the inverse of b evaluated at \mathbf{z} , and its determinant is not null because b is bijective. The same equivalence holds for any conditional distribution $p_{\mathbf{Z}|\mathbf{Y}}(\mathbf{z} | \mathbf{y}) = p_{\mathbf{X}|\mathbf{Y}}(b^{-1}(\mathbf{z}) | \mathbf{y})|\det(\frac{d\mathbf{x}}{d\mathbf{z}})|$, and the local mutual information of the couple $(b(\mathbf{x}), \mathbf{y})$ with respect to the joint distribution $p_{\mathbf{Z}, \mathbf{Y}}$ is expressed by:

$$i(b(\mathbf{x}); \mathbf{y}) = \log \frac{p_{\mathbf{Z}|\mathbf{Y}}(\mathbf{z} | \mathbf{y})}{p_{\mathbf{Z}}(\mathbf{z})}, \quad (15)$$

$$= \log \frac{p_{\mathbf{X}|\mathbf{Y}}(b^{-1}(\mathbf{z}) | \mathbf{y})|\det(\frac{d\mathbf{x}}{d\mathbf{z}})|}{p_{\mathbf{X}}(b^{-1}(\mathbf{z}))|\det(\frac{d\mathbf{x}}{d\mathbf{z}})|}, \quad (16)$$

$$= \log \frac{p_{\mathbf{X}|\mathbf{Y}}(\mathbf{x} | \mathbf{y})}{p_{\mathbf{X}}(\mathbf{x})}, \quad (17)$$

$$= i(\mathbf{x}; \mathbf{y}). \quad (18)$$

□

With the same reasoning, the property holds for the second variable \mathbf{y} . In particular, the local mutual information is invariant to any invertible linear transform applied independently on each variable $\mathbf{z} = A\mathbf{x}$, $\mathbf{w} = B\mathbf{y}$, such that $i(\mathbf{x}; \mathbf{y}) = i(A\mathbf{x}; B\mathbf{y}) = i(\mathbf{z}; \mathbf{w})$.

Such a property is of particular interest in situations where the original data are transformed by an unknown or non computable function which is deterministic and bijective. The practical relevance of this property will be demonstrated for the segmentation of multi-angular satellite images in Sec. 5.3.

2.4 Local Mutual Information of Noisy Data

While deterministic transforms do not impact the local mutual information, the presence of noise decreases realization dependences. Assuming an additive noise model, the local mutual information can be analytically linked to the noise free local mutual information.

Proposition 3 *Let \mathbf{X} and \mathbf{Z} be two multidimensional variables, such that $\mathbf{Z} = \mathbf{X} + \mathbf{N}$ and \mathbf{N} is a noise component. Then, for any third random variable \mathbf{Y} and any of its realizations \mathbf{y} , we have $i(\mathbf{z}; \mathbf{y}) = \log \frac{(p_{\mathbf{X}, \mathbf{Y}} * p_{\mathbf{N}})(\mathbf{z}, \mathbf{y})}{(p_{\mathbf{X}} * p_{\mathbf{N}})(\mathbf{z})p_{\mathbf{Y}}(\mathbf{y})}$, where $*$ is the convolution operator.*

Proof Given the probability distributions of \mathbf{X} and \mathbf{N} and knowing that \mathbf{N} is independent of \mathbf{X} as it is an additive noise component, the distribution of \mathbf{Z} is expressed by:

$$p_{\mathbf{Z}}(\mathbf{z}) = \int p_{\mathbf{X}}(\mathbf{z} - \mathbf{n})p_{\mathbf{N}}(\mathbf{n})d\mathbf{n} = (p_{\mathbf{X}} * p_{\mathbf{N}})(\mathbf{z}). \quad (19)$$

The same equality can be derived for the joint probability $p_{\mathbf{Z}, \mathbf{Y}}(\mathbf{z}, \mathbf{y}) = (p_{\mathbf{X}, \mathbf{Y}} * p_{\mathbf{N}})(\mathbf{z}, \mathbf{y})$.

□

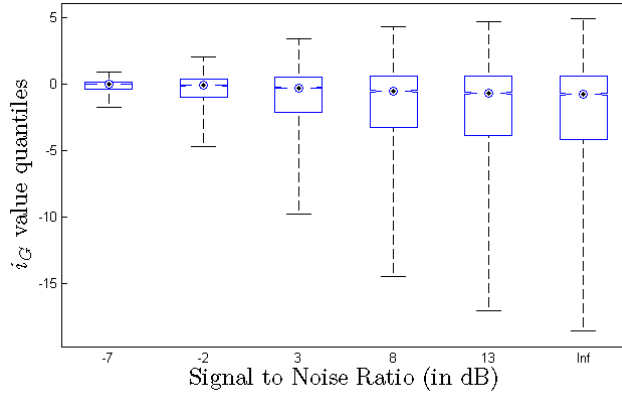


Fig. 3 The $\{0, 0.25, 0.5, 0.75, 1\}$ -quantiles of the $i_G(\mathbf{z}; \mathbf{y})$ values are represented in the form of boxes for various signal to noise ratios, where $\mathbf{Z} = \mathbf{X} + \mathbf{N}$.

The convolution by the distribution of the additive noise spreads the probability in the dimension of \mathbf{X} and thus decreases the dependence of the noisy observations with respect to the third random variable \mathbf{Y} .

In the case of Gaussian random variables, the noise increases the variance of the right member while keeping constant the cross correlation matrices. Therefore, as the noise variance increases to infinity, the noisy random variable \mathbf{Z} becomes independent of any \mathbf{Y} which reduces the dynamic of their local mutual information. This effect is illustrated in Fig. 3, where the quantiles of the values $i_G(\mathbf{z}; \mathbf{y})$ are represented for various Signal to Noise Ratio. In this illustration, the initial random variables \mathbf{X} and \mathbf{Y} are Gaussian, and the first one is affected by an additive Gaussian noise, $\mathbf{Z} = \mathbf{X} + \mathbf{N}$. As expected, the range of the $i_G(\mathbf{z}; \mathbf{y})$ values tends to 0 as the noise variance increases (which is equivalent to a decreasing signal to noise ratio).

3 Dependent Connectivity Based Segmentation

The notion of dissimilarity based connective segmentation, widely described in the literature [Nagao et al., 1979, Serra, 1988, 2006, Meyer and Maragos, 2000, Soille, 2008, 2011], is recalled in Sec. 3.1. A local mutual information based dissimilarity, which was presented at ISMM 2011 [Gueguen and Soille, 2011], is detailed in Sec. 3.2 while introducing the concept of dependent connectivity based segmentation. Novel properties of the dependent connectivity are discussed with regard to the local mutual information invariance. This connectivity relation is then extended to the framework of constrained connectivity in Sec. 3.3.

3.1 Dissimilarity Based Connective Segmentation

Let \mathbf{f} be a digital image and $\mathcal{G} = (V, E)$ an undirected simple finite graph with vertex set V matching the image pixels and edge set E consisting of unordered

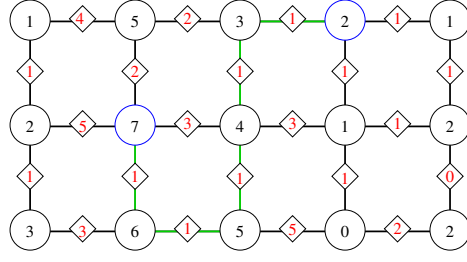


Fig. 4 A 5×3 image seen as a graph $\mathcal{G} = (V, E)$, with circled vertex values and boxed edge values. This example shows a path P in green between the blue vertices.

pairs of vertices indicating the adjacency relations between the image pixels (for example, 4 or 8-connected graphs for 2D images on the square lattice, 6 or 18-connected graphs for 3D images on the cubic lattice, and Delaunay graph for images sampled at irregularly spaced points). A path P between two pixels p and q in \mathcal{G} is a sequence of $n > 1$ pixels $\langle p = p_1, \dots, p_n = q \rangle$ such that any two successive pixels of the sequence are adjacent: $\{p_i, p_{i+1}\} \in E$ for $1 \leq i < n$. The image \mathbf{f} assigns to each element of V (that is, the pixels), a vector of values. Such a graph based representation is exemplified in Fig. 4.

The partition of a graph $\mathcal{G} = (V, E)$ into connected components relies on a function associating a weight w to its edges. This function can be viewed as a measure of the degree of dissimilarity between adjacent vertices. For grey level images, the most common dissimilarity is the absolute difference. A dissimilarity measure involving a larger neighbourhood to prevent chaining through transitions while favouring it within homogeneous regions is presented in [Soille, 2011]. In this paper, we consider dissimilarity measures obtained through a dissimilarity function defined for any pair of value vectors.

Definition 1 ([Gueguen and Soille, 2011]) Let \mathcal{I} be the space of image value vectors. A dissimilarity function indexed by the intensity values of the image \mathbf{f} is defined as any function d from $\mathcal{I} \times \mathcal{I} \rightarrow \mathbb{R}_0^+$ such that $d(\mathbf{f}(p), \mathbf{f}(q)) = d(\mathbf{f}(q), \mathbf{f}(p))$, where $p, q \in V$ are two pixels of the image. This latter property is imposed by the symmetry property of dissimilarity measures.

The most natural dissimilarity functions are obtained by considering the L_m norm of the difference of the input value vectors: $d^{L_m}(\mathbf{f}(p), \mathbf{f}(q)) = \|\mathbf{f}(p) - \mathbf{f}(q)\|_m$. Typical choices for m are 1, 2, or ∞ . Another common choice for multispectral images is the spectral angular distance [Kruse et al., 1993].

Given a dissimilarity function d , the weight of an edge $\{p, q\}$ of the adjacency graph underlying an image \mathbf{f} is denoted by w_d : $w_d(\{p, q\}) = d(\mathbf{f}(p), \mathbf{f}(q))$. Given a dissimilarity threshold α , two distinct pixels p and q of an image \mathbf{f} are d_α -connected if there exists a path going from p to q such that the weight w_d between any two successive pixels of this path does not exceed the value of the dissimilarity threshold level α . In addition, to ensure the reflexivity property of an equivalence relation, a pixel is always said to be d_α -connected to itself.

Definition 2 ([Gueguen and Soille, 2011]) Let an image \mathbf{f} be represented by a graph \mathcal{G} , where the edge weights are given by a dissimilarity function $d(\mathbf{f}(p), \mathbf{f}(q)) = w_d(\{p, q\})$, $(p, q) \in E$. The *connected component* $d_\alpha\text{-CC}(p)$ of a pixel p is the set

of pixels which are d_α -connected to this pixel:

$$d_\alpha\text{-CC}(p) = \{p\} \cup \{q \mid \exists \mathcal{P} = \langle p = p_1, \dots, p_n = q \rangle, \\ d(\mathbf{f}(p_i), \mathbf{f}(p_{i+1})) \leq \alpha, \forall 1 \leq i < n\}. \quad (20)$$

The d_α -connectivity being an equivalence relation, it induces a unique partition of the image \mathbf{f} support into maximal connected regions being the d_α -connected components [Soille, 2008]. We denote the resulting partition by $\Pi_{\mathbf{f}}^d(\alpha)$. This approach for segmentation was put forward implicitly in [Nagao et al., 1979] for the segmentation of multispectral images using the L_1 norm (see also [Baraldi and Parmiggiani, 1996]). For grey level images, the L_m norm boils down to the absolute difference that is denoted by d^A . The corresponding connected components are called quasi-flat zones [Meyer and Maragos, 1999] or lambda-flat zones [Zanoguera and Meyer, 2002] in mathematical morphology. Because the dissimilarity value between adjacent pixels of a connected component can be arbitrarily large, the terminology α -connected components was suggested in [Soille, 2008, 2011]. Rather than using the notion of equivalence relation and path-based connectivity, Serra [2006] proposes a more general approach to image segmentation based on the lattice theory and that does not require the definition of paths. With this theory, the existence of a maximum partition is secured if and only if the homogeneous classes form a connection based on the so-called connective criterion. The resulting segmentation was called a connection in [Serra, 2006] and lately referred to as a connective segmentation [Ronse and Serra, 2008, Ronse, 2008]. The Boolean connective criterion underlying the d_α -connectivity is detailed in [Gueguen and Soille, 2011, Ouzounis and Soille, 2011]. Hence, the d_α -connectivity turns out to be a connection and therefore yields a connective segmentation (maximum partitioning).

A fundamental property of the d_α -connected components of a pixel is that they form an ordered sequence (hierarchy) when increasing the dissimilarity threshold value α [Soille, 2008]:

$$d_{\alpha_1}\text{-CC}(p) \subseteq d_{\alpha_2}\text{-CC}(p), \forall \alpha_1 \leq \alpha_2. \quad (21)$$

This hierarchy is at the root of the greedy algorithm by Kruskal [Kruskal, 1956] for solving the minimum spanning tree problem and at the very basis of the dendrogram representation of the single linkage clustering [Gower and Ross, 1969]. Such a hierarchy can be efficiently encoded in a α -tree [Ouzounis and Soille, 2012].

Definition 3 Let the dissimilarities of a 2-dimensional image \mathbf{f} be summarised by the *edge map* $\mathcal{E}_{\mathbf{f}}^d$, which represents the maximum dissimilarity between a pixel p and its right and bottom neighbours p_r and p_b respectively:

$$\mathcal{E}_{\mathbf{f}}^d(p) = \max \{d(\mathbf{f}(p), \mathbf{f}(p_r)), d(\mathbf{f}(p), \mathbf{f}(p_b))\}. \quad (22)$$

The edge map is essentially a tool that enables the visual analysis of a dissimilarity computed for a specific image, see experiments in Sec. 5.

3.2 Dependent Connectivity

When the dimension of the data increases, the volume of the space increases so fast that the available data becomes sparse. One aspect of the curse of dimensionality is the distance concentration, which denotes the tendency of distances between all pairs of points in high-dimensional data to become almost equal and all pairwise vectors are approximately perpendicular [Radovanović et al., 2010]. Therefore, for multivariate images, the discrimination power of dissimilarity measures based on vector distances or spectral angles decreases with the number of dimensions.

This motivated us to develop alternative dissimilarities taking into account co-occurrence statistics and led to the notion of frequent and dependent connectivities introduced in [Gueguen and Soille, 2011]. While frequent connectivity aggregates preferentially frequent pixel pairs, dependent connectivity privileges the pixel pairs showing a high statistical dependence. The local mutual information is used to estimate this dependence so that the pairs of pixels having a high local mutual information are first connected. This latter approach was shown to be more effective than frequent connectivity because the frequency is highly related to the distance between pixel values [Gueguen and Soille, 2011].

Let \mathbf{X}, \mathbf{Y} be two random variables defined on the product space $\mathcal{I} \times \mathcal{I}$, and any pair of adjacent pixel value $(\mathbf{f}(p), \mathbf{f}(q)) = (\mathbf{x}, \mathbf{y})$, $\{p, q\} \in E$ is a realization of the joint variable (\mathbf{X}, \mathbf{Y}) . Such an image model is widely used in the literature for characterising textures by the computation of grey level co-occurrence matrices [Haralick et al., 1973]. Assuming that any pair of adjacent pixel values $(\mathbf{f}(p), \mathbf{f}(q)) = (\mathbf{x}, \mathbf{y})$ follows an identical distribution $p_{\mathbf{X}, \mathbf{Y}}$, it follows that $p_{\mathbf{X}, \mathbf{Y}}$ is symmetric, such that $p_{\mathbf{X}, \mathbf{Y}}(\mathbf{x}, \mathbf{y}) = p_{\mathbf{X}, \mathbf{Y}}(\mathbf{y}, \mathbf{x})$. The dependence of adjacent pixel value vectors is then expressed by $i(\mathbf{f}(p); \mathbf{f}(q)) = i(\mathbf{x}; \mathbf{y})$, which is a symmetric function.

Definition 4 ([Gueguen and Soille, 2011]) Let the dependent dissimilarity function of an image \mathbf{f} be defined in between adjacent pixel values $(\mathbf{f}(p), \mathbf{f}(q))$, $\{p, q\} \in E$ by:

$$d^D(\mathbf{f}(p), \mathbf{f}(q)) = \max_{\mathbf{x}, \mathbf{y}} \{i(\mathbf{x}; \mathbf{y})\} - i(\mathbf{f}(p); \mathbf{f}(q)). \quad (23)$$

The dissimilarity is symmetric because of the symmetry of the joint probability distribution and is nonnegative. When the dissimilarity is 0 the image values have the maximum statistical dependence, while a greater value accounts for statistical independence.

Given a dissimilarity threshold α , the induced dependent connected components are denoted by d_α^D -CC. They also form an ordered sequence of segmentations when increasing the value α because of the property of hierarchies expressed by (21).

A toy example (see Fig. 5(a)) is considered to compare the behaviour of dependent connectivity with respect to the connectivity based on absolute difference dissimilarity. This image contains two homogeneous parts altered by small fluctuations. The left part is composed of 0s and 1s, while the right part is composed of 5s and 6s. The statistics for computing the dependent connectivity are estimated from the toy example. The dependent connectivity edge weights (d^D) are rescaled between 0 and 10, and displayed in Fig. 5(b). The corresponding edge weights obtained by absolute difference (d^A) are given in Fig. 5(a). The connected compo-

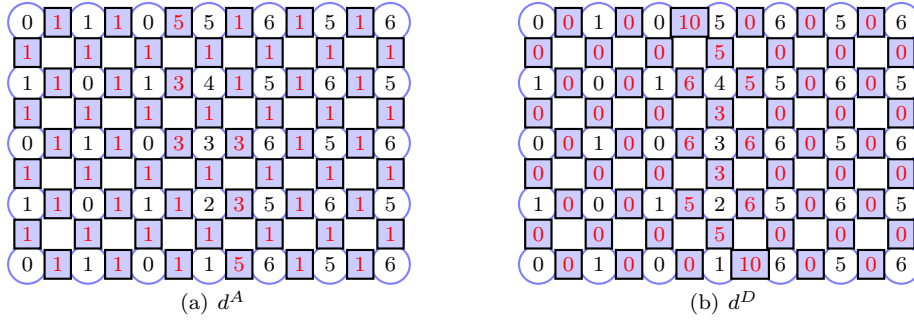


Fig. 5 A toy image is represented by a graph, where the nodes (circles) represent the pixel values. The dissimilarity values are given in the square boxes. a) The absolute difference based dissimilarity is represented. b) The local mutual information based dissimilarity is represented. The local mutual information is estimated from the image pixel values and the resulting dissimilarity d^D is linearly rescaled between 0 and 10, for a better visibility.

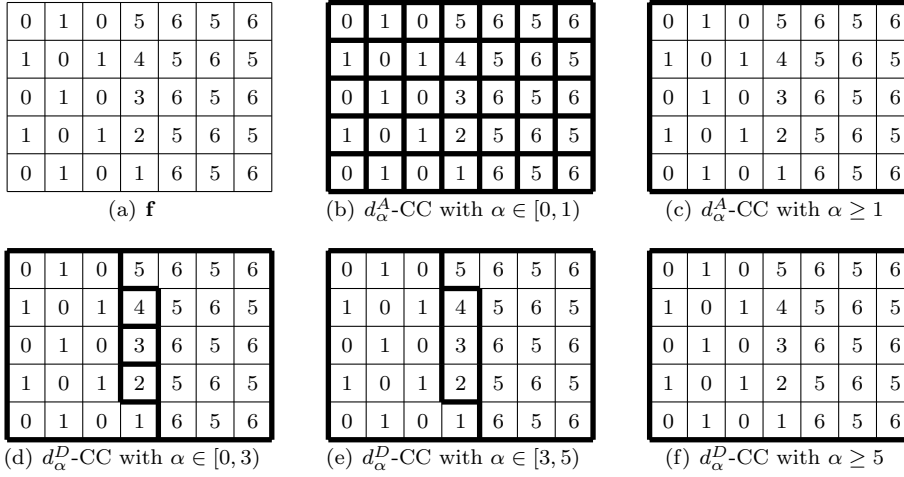


Fig. 6 Experiment on synthetic data showing a transition between two homogeneous regions. a) The image values. The connected components obtained with the absolute difference dissimilarity d^A are given in (b) and (c). The connected components obtained with the local mutual information d^D are given in (d), (e) and (f). Both dissimilarities d^A and d^D are represented in Fig. 5 for the current image.

nents derived from both dissimilarities are then represented in Fig. 6. We observe that the possible connected components generated by d^A are either the individual pixels or the full image support, while the image contains two homogeneous parts. Such situation is due to the leakage effect through the ramp separating these two parts [Soille and Grazzini, 2009, Soille, 2011]. Because the dependent connectivity takes into account the statistical dependences, it succeeds to retain these two image regions, while isolating the ramp, contrary to connected components based on the absolute difference dissimilarity. This toy example shows how the depen-

dent connectivity prevents the leakage effect while producing relevant connected components.

Proposition 4 *If the image values \mathbf{f} are systematically transformed by a bijective differentiable transform $b : \mathcal{I} \rightarrow \mathcal{I}$, the dependent dissimilarity is unchanged thanks to the invariance property (18), and $d^D(b(\mathbf{f}(p)), b(\mathbf{f}(q))) = d^D(\mathbf{f}(p), \mathbf{f}(q))$. Therefore, the d_α^D -CCs computed independently from both images \mathbf{f} and $b(\mathbf{f})$ will be strictly equivalent so that $\Pi_{\mathbf{f}}^{d^D}(\alpha) = \Pi_{b(\mathbf{f})}^{d^D}(\alpha)$.*

For instance, the segmentations shown in Figs. 6d–f are invariant to bijective differentiable transforms applied to the input image. That is, the dissimilarity threshold values do not need to be tuned accordingly, contrary to the dissimilarities based on L_m norms. Such a property is of main interest when processing remote sensing images which undergo radiometric transforms according to atmospheric conditions, sun spectral irradiance and on-board sensor corrections. In most cases, these radiometric transforms are unknown, but one can assume that they are bijective differentiable transforms. In addition, such a property is very useful when fusing bands of different physical dimensions (e.g., spectral and digital surface model), since any adjustment of the band variances by a linear transform (such as principal component analysis [Pearson, 1901], maximum noise fraction transform [Green et al., 1988], or independent component analysis [Comon, 1994]) is a bijective differentiable transform.

3.3 Extension to Constrained Connectivity

Constrained connectivity was introduced in [Soille, 2008] to generate partitions consisting of maximal d_α -connected components satisfying a series of constraints. The most natural constraint regards a condition on the homogeneity of the intensity values of the pixels belonging to each connected component. The simplest homogeneity measure is the range of these values. For instance, one may look for the coarsest partition into d_{α_i} -connected components such as the dissimilarity threshold α_i does not exceed a given threshold value α and the total range (difference between the maximum and minimum value) of each connected component does not exceed a given threshold value denoted by ω .

A constrained connected component of a pixel p is denoted by $d_{(\alpha, \omega)}\text{-CC}(p)$ and is defined with respect to a dissimilarity threshold α and a total range ω [Soille, 2008]:

$$d_{(\alpha, \omega)}\text{-CC}(p) = \max \{ d_{\alpha_i}\text{-CC}(p) \mid \alpha_i \leq \alpha \text{ and } R(d_{\alpha_i}\text{-CC}(p)) \leq \omega \},$$

where the function R returns the difference between the maximum and minimum value of the pixels belonging to its argument. Accordingly, the resulting image partition is denoted by $\Pi_{\mathbf{f}}^d(\alpha, \omega)$. Note that while the criterion underlying the notion of d_{α_i} -connected components is a connective criterion (see Sec. 3.1), the criterion underlying the notion of $d_{(\alpha, \omega)}$ -connected components is not connective [Ouzounis and Soille, 2011]. Ronse [2011] calls *constrained connective segmentation* this approach of segmentation combining connective and constrained criteria.

For multi-channel images, the total range constraint consists of a vector of scalar constraints, one for each channel [Soille, 2008]. The generalisation of this approach to arbitrary constraints expressed in terms of logical predicates is presented

in [Soille, 2007, Soille and Grazzini, 2009, Soille, 2011, Ouzounis and Soille, 2012]. In the case of images with a very high number of dimensions such as hyperspectral images, rather than considering a total range constraint on all channels of the input image, the n first channels obtained after some transformation space such as the principal component or maximum noise fraction transformations [Green et al., 1988] may be considered. This approach is used in the experiments on hyperspectral images in Sec. 5.2 where the total range constraint is applied to the first principal component.

Dependent connectivity generalises directly to constrained connectivity, and any constraint can be applied to the d^D -CCs. However, contrary to what happens with dissimilarities based on L_m norms, the total range constraint detailed above is not directly related to the dissimilarity measure based on local mutual information. One way to address this problem is to define a homogeneity measure expressed in terms of dissimilarities. This led to the notion of strongly connected component or components whose connectivity index do not exceed a given threshold value [Soille, 2008]. Nevertheless, it is possible to design a homogeneity measure directly related to the local mutual information by using entropy measurements. This latter approach is put forward hereafter.

The notion of entropy is well established in physics to measure the disorder of a system, and an entropy based objective function was successfully used in [Zhang et al., 2003, Guigues et al., 2006] in order to evaluate a segmentation. From this perspective, the entropy of a connected component can be used to constrain the segmentation, such that low entropy components are privileged because they have less disorder.

Definition 5 Given a connected component d_α -CC(p) composed of the pixel values $\{\mathbf{f}(p_1), \dots, \mathbf{f}(p_n)\}$, its (Shannon)entropy $H(d_\alpha$ -CC(p)) is defined:

$$H(d_\alpha$$
-CC(p)) = $-\frac{1}{n} \sum_{i=1}^n \log p_{CC}(\mathbf{f}(p_i)), \quad (24)$

where $\{\mathbf{f}(p_1), \dots, \mathbf{f}(p_n)\}$ are independent and identically distributed with the probability distribution p_{CC} .

In practice, p_{CC} is estimated from the pixel values of d_α -CC(p). The entropy H generalises the total range R measure, since values lying in a small range interval get mechanically a low entropy, while the opposite is not always true.

A differential entropy constraint criterion is proposed hereafter in order to obtain the bijective transform invariance. In general, the entropy can be thought as the Kullback-Leibler divergence between the considered distribution and the uniform probability distribution.

Definition 6 The differential entropy HD of a connected component d_α -CC(p) is obtained by comparing the distribution p_{CC} of pixel values inside the component with the global pixel values distribution $p_{\mathbf{f}}$:

$$HD(d_\alpha$$
-CC(p)) = $-\frac{1}{n} \sum_{i=1}^n \log \frac{p_{CC}(\mathbf{f}(p_i))}{p_{\mathbf{f}}(\mathbf{f}(p_i))}. \quad (25)$

Thanks to the proposition 2, this constraint is invariant to any bijective and differentiable function assuming a perfect estimate of p_{CC} , while still privileging the components being more organised than the full image. In summary, constraint connectivity based on the local mutual information and the differential entropy provides a unique partition of any image which is invariant to deterministic transforms. Nevertheless, this constraint is not tested in the experimental section because of its high computational complexity, and it is substituted by a constraint on the range of the pixel values in the component.

4 Probability Distribution Estimation

When considering multichannel images the problem of joint and marginal probability distribution estimation becomes cumbersome due to the explosion of dimensions. Assuming a Gaussian distribution, only the estimation of the mean vector and the covariance matrix are necessary to compute the local mutual information (12). Nevertheless, in case this assumption does not hold true, the estimation fails. Thus, a projection depth function based method for estimating multivariate distributions is presented, such that it is robust to the number of dimensions without assuming an a priori distribution shape.

The projection depth function definition is recalled in Sec. 4.1. It is shown to be equivalent to the Mahalanobis distance for Gaussian random variables in Sec. 4.2 and is therefore an approximation of the logarithm of centered probability distributions. This property is exploited in Sec. 4.3 to define the depth function based local mutual information, which is an original approximation method of the considered metric for multidimensional signals.

4.1 Projection depth function

The basic concept of a projection depth function was firstly introduced in [Donoho and Gasko, 1992] and subsequently developed in [Zuo and Serfling, 2000, Zuo, 2003]. It defines the measures of centrality for a vector \mathbf{x} with respect to a multivariate distribution or a multivariate data cloud (\mathbf{X}) as the worst case outlierness with respect to the one-dimensional scale functional in any one-dimensional projection. Formally, the *projection depth function* df for a vector \mathbf{x} according to a data cloud $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_k]$ is defined as follows [Donoho and Gasko, 1992]:

$$df(\mathbf{x}; \mathbf{X}) = \sup_{\mathbf{u} \in \mathbb{S}^{k-1}} \frac{|\mathbf{u}^T \mathbf{x} - \text{Med}(\mathbf{u}^T \mathbf{X})|}{\text{MAD}(\mathbf{u}^T \mathbf{X})}, \quad (26)$$

where Med is the median, MAD is the median absolute deviation and $\mathbb{S}^{k-1} = \{\mathbf{u} \in \mathbb{R}^k \mid \|\mathbf{u}\| = 1\}$ is the k -dimensional sphere. An illustrative example is provided in Fig. 7.

Robust estimators of localisation Med and variability MAD are included in (26) because they are not unduly affected by outliers [Zuo, 2003]. Projection depth has been used to robust multivariate classification [Cui et al., 2008], classification of function data [Cuevas et al., 2007], and ordering for performing morphological transforms on multivariate images [Velasco-Forero and Angulo, 2012, 2011]. The

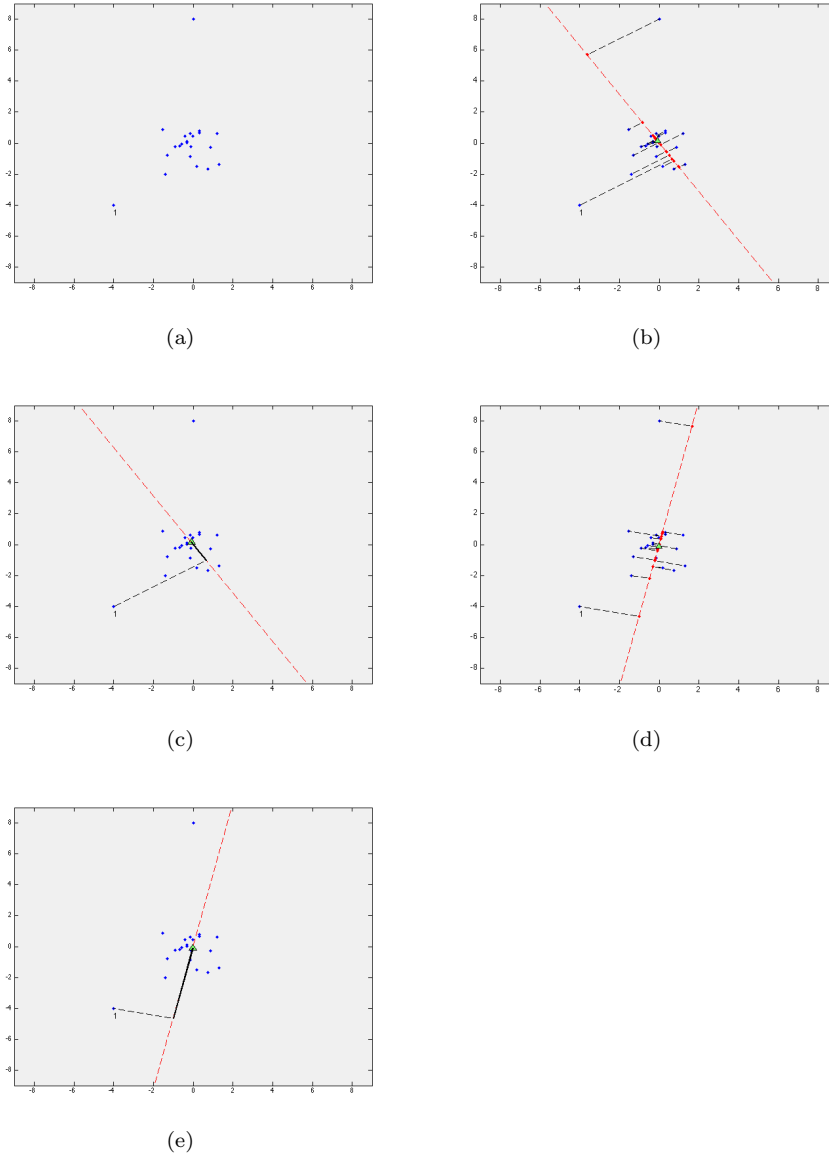


Fig. 7 a) Scatterplot of 22 bi-dimensional vectors. b) Given a random vector (\mathbf{u}) (red line), vectors in (\mathbf{X}) are projected $\mathbf{u}^T \mathbf{X}$ (Red points). Median on the projection ($Med(\mathbf{u}^T \mathbf{X})$) is represented by a triangle. c) For a given vector \mathbf{x} , in this case labeled by "1", its value in the projection depth is the distance in the projection $|\mathbf{u}^T \mathbf{x} - Med(\mathbf{u}^T \mathbf{X})|$ (solid black line). In the definition the denominator $MAD(\mathbf{u}^T \mathbf{X})$ is a scale normalization. d) and e) show the computation for a second random projection. Note that definition (26) takes the maximum of those "univariate eccentricities".

projection depth function is a nonnegative mapping satisfying the following properties [Liu, 1990] [Zuo and Serfling, 2000]:

1. $df(\mathbf{A}\mathbf{x} + \mathbf{b}; \mathbf{A}\mathbf{X} + \mathbf{b}) = df(\mathbf{x}; \mathbf{X})$ holds for any random vector \mathbf{x} in \mathbb{R}^k , any $k \times k$ nonsingular matrix \mathbf{A} , and any $\mathbf{b} \in \mathbb{R}^k$. That invariance to affine transformations means that the depth of a vector $\mathbf{x} \in \mathbb{R}^k$ should not depend on the underlying coordinate system or, in particular, on the scales of the underlying measurements;
2. $df(\theta; \mathbf{X}) = \inf_{\mathbf{x} \in \mathbb{R}^k} df(\mathbf{x}; \mathbf{X})$ holds for any \mathbf{X} having centre θ . That means, for any distribution having a unique 'centre', the depth function should attain its minimum value at this centre;
3. $df(\mathbf{x}; \mathbf{X}) \geq df(\theta + \alpha(\mathbf{x} - \theta); \mathbf{X})$ holds for any \mathbf{X} having a deepest point θ and any $\alpha \in [0, 1]$, i.e., as a point $\mathbf{x} \in \mathbb{R}^k$ moves away from the deepest point along any fixed ray through the centre, the depth at \mathbf{x} should decrease monotonically;
4. $df(\mathbf{x}; \mathbf{X}) \rightarrow \infty$ as $\|\mathbf{x}\| \rightarrow \infty$, for each \mathbf{X} , i.e., the depth of a point \mathbf{x} should approach to infinity as its norm approaches infinity.

Note that $df(\mathbf{x}; \mathbf{X})$ cannot be calculated in practice, because it requires the analysis for an infinite set of random projections. Our approach follows the suggestion in [Zuo, 2006], replacing the supremum in (26) by a maximum over a finite number of randomly chosen projections, obtaining a random approximation to the random projection depth. The same argument have been used in [Cuesta-Albertos and Nietos-Reyes, 2008] for other types of statistical depth functions. Thus, we calculate an approximate value of $df(\mathbf{x}; \mathbf{X})$ by using q random projections uniformly distributed in \mathbb{S}^{k-1} as follows:

$$df(\mathbf{x}; q, \mathbf{X}) = \max_{\mathbf{u} \in \mathbb{U}} \frac{|\mathbf{u}^T \mathbf{x} - \text{Med}(\mathbf{u}^T \mathbf{X})|}{\text{MAD}(\mathbf{u}^T \mathbf{X})}, \quad (27)$$

where $\mathbb{U} = \{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_q\}$ with $\mathbf{u}_i \in \mathbb{S}^{k-1}$. Clearly, if $q \rightarrow \infty$ then $df(\mathbf{x}; q, \mathbf{X}) \rightarrow df(\mathbf{x}; \mathbf{X})$.

4.2 Equivalence With Gaussian Distribution

When Gaussian variables are considered, the depth function admits an analytical formulation expressed in terms of Mahalanobis distance. This analytical formulation shows that the depth function approximate the logarithm of centered probability distributions up to a normalization scalar.

For \mathbf{A} a positive definite matrix, and \mathbf{z} a given vector, and \mathbf{u} a non zero arbitrary vector, the following equality holds [Johnson and Wichern, 2007, p. 65]:

$$\sup_{\mathbf{u} \neq 0} \frac{(\mathbf{u}^T \mathbf{z})^2}{\mathbf{u}^T \mathbf{A} \mathbf{u}} = \mathbf{z}^T \mathbf{A}^{-1} \mathbf{z}. \quad (28)$$

Proposition 5 *Let $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n]$ be a i.i.d. random sample of size n , where $\mathbf{X} \sim \mathcal{N}(\mu, \Sigma)$, then the depth function can be expressed by the Mahalanobis distance:*

$$c_1 df(\mathbf{x}; \mathbf{X})^2 = (\mathbf{x} - \mu)^T \Sigma^{-1} (\mathbf{x} - \mu), \quad (29)$$

with $c_1 = (\Phi^{-1}(3/4))^2 \approx .454936$, where Φ^{-1} is the inverse Gaussian cumulative distribution function.

Proof Since $\mathbf{X} \sim \mathcal{N}(\mu, \Sigma)$, we can say that $\mathbf{u}^T \mathbf{X}$ is symmetric about $\mathbf{u}^T \mu$, then $\text{Med}(\mathbf{u}^T \mathbf{X}) = \mathbf{u}^T \mu$ for any $\mathbf{u} \in \mathbb{R}^k$. Additionally, normalising by the norm of \mathbf{u} , we constrain the search-space to random projections being in the k -dimensional sphere as follows:

$$df(\mathbf{x}; \mathbf{X})^2 = \sup_{\mathbf{u} \neq 0} \left\{ \frac{|\mathbf{u}^T \mathbf{x} - \text{Med}(\mathbf{u}^T \mathbf{X})|^2}{\text{MAD}^2(\mathbf{u}^T \mathbf{X})} \right\}, \quad (30)$$

$$= \sup_{\mathbf{u} \in \mathbb{S}^{k-1}} \left\{ \frac{(\mathbf{u}^T (\mathbf{x} - \mu))^2}{\mathbf{u}^T \Sigma \mathbf{u}} \left(\frac{\sigma(\mathbf{u}^T \mathbf{X})}{\text{MAD}(\mathbf{u}^T \mathbf{X})} \right)^2 \right\}. \quad (31)$$

Multiplying the numerator and denominator terms by $\sigma^2(\mathbf{u}^T \mathbf{X})$, the variance of $\mathbf{u}^T \mathbf{X}$, we obtain:

$$df(\mathbf{x}; \mathbf{X})^2 = \sup_{\mathbf{u} \in \mathbb{S}^{k-1}} \left\{ \frac{(\mathbf{u}^T (\mathbf{x} - \mu))^2}{\mathbf{u}^T \Sigma \mathbf{u}} \left(\frac{\sigma(\mathbf{u}^T \mathbf{X})}{\text{MAD}(\mathbf{u}^T \mathbf{X})} \right)^2 \right\},$$

where the second term is constant and it is the ratio between MAD and the standard deviation σ for a Gaussian distribution [Sachs, 1984, p. 253], i.e., $\text{MAD}(\mathbf{X})/\sigma(\mathbf{X}) = \Phi^{-1}(3/4)$. Thus, the square of the projection depth function can be expressed as:

$$df(\mathbf{x}; \mathbf{X})^2 = \max_{\mathbf{u} \in \mathbb{S}^{k-1}} \left\{ \frac{(\mathbf{u}^T (\mathbf{x} - \mu))^2}{\mathbf{u}^T \Sigma \mathbf{u}} \right\} \frac{1}{(\Phi^{-1}(3/4))^2},$$

and by using (28):

$$df(\mathbf{x}; \mathbf{X})^2 (\Phi^{-1}(3/4))^2 = (\mathbf{x} - \mu)^T \Sigma^{-1} (\mathbf{x} - \mu).$$

□

4.3 Approximate Local Mutual Information

In the absence of a priori information, the sample covariance matrix is a natural candidate in the case of small dimensionality but no longer performs very well for moderate or large dimensionality [Johnstone, 2001]. On the other hand, application of connectivity by using vector distance is affected by the curse of dimensionality [Bellman, 1961], so that robust estimation methods are required. Exploiting the depth function, an approximation of the local mutual information is proposed hereafter.

Proposition 6 *Given a set of realizations $\{\mathbf{z}_i = (\mathbf{x}_i, \mathbf{y}_i)\}$ following a Gaussian distribution $\mathcal{N}(\mu_{\mathbf{Z}}, \Sigma_{\mathbf{Z}})$, the local mutual information is exactly expressed with the depth function from (12) and (29):*

$$\begin{aligned} i_G(\mathbf{x}; \mathbf{y}) &= -\frac{1}{2}(\mathbf{z} - \mu_{\mathbf{Z}})^T \Sigma_{\mathbf{Z}}^{-1}(\mathbf{z} - \mu_{\mathbf{Z}}) + \frac{1}{2}(\mathbf{x} - \mu_{\mathbf{X}})^T \Sigma_{\mathbf{X}}^{-1}(\mathbf{x} - \mu_{\mathbf{X}}) \\ &\quad + \frac{1}{2}(\mathbf{y} - \mu_{\mathbf{Y}})^T \Sigma_{\mathbf{Y}}^{-1}(\mathbf{y} - \mu_{\mathbf{Y}}) - \frac{1}{2} \log \frac{\det \Sigma_{\mathbf{Z}}}{\det \Sigma_{\mathbf{X}} \det \Sigma_{\mathbf{Y}}}, \end{aligned} \quad (32)$$

$$\begin{aligned} &= -\frac{c_1}{2} \left(df(\mathbf{z}; \mathbf{Z})^2 - df(\mathbf{x}; \mathbf{X})^2 - df(\mathbf{y}; \mathbf{Y})^2 \right) \\ &\quad - \frac{1}{2} \log \frac{\det \Sigma_{\mathbf{Z}}}{\det \Sigma_{\mathbf{X}} \det \Sigma_{\mathbf{Y}}}. \end{aligned} \quad (33)$$

Definition 7 Let $\{\mathbf{z}_i = (\mathbf{x}_i, \mathbf{y}_i)\}$ be a set of realizations of the random variable $\mathbf{Z} = (\mathbf{X}, \mathbf{Y})$, we define the *depth function based local mutual information* by:

$$i_{df}(\mathbf{x}; \mathbf{y}) = -\frac{c_1}{2} \left(df(\mathbf{z}; \mathbf{Z})^2 - df(\mathbf{x}; \mathbf{X})^2 - df(\mathbf{y}; \mathbf{Y})^2 \right). \quad (34)$$

Such a function is an approximation up to a constant of the true local information of Gaussian variables. In the case where the random variables are not Gaussian, the depth function still captures the structure of the "centred" distribution, so that $i_{df}(\mathbf{x}; \mathbf{y})$ remains a measure of dependence. The proposed depth function based local mutual information allows us to define a dependent dissimilarity function following (23). It can be viewed as an adaptive connective criterion. The approximation power of (34) is assessed in Sec. 5.1.

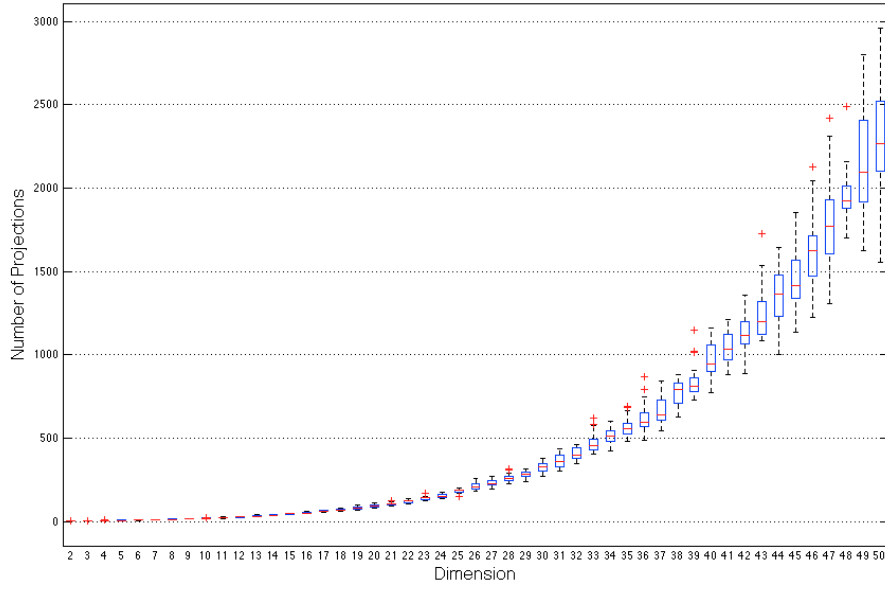
5 Experiments

The approximation of the depth function based mutual information is tested with simulated data in Sec. 5.1. Then two experiments are conducted with remote sensing images in Secs. 5.2 and 5.3 respectively. The first experiment assesses the effectiveness of the proposed dissimilarity to segment hyperspectral image. The second one assesses the invariance property of the proposed local mutual information by processing multi-angular very high resolution images.

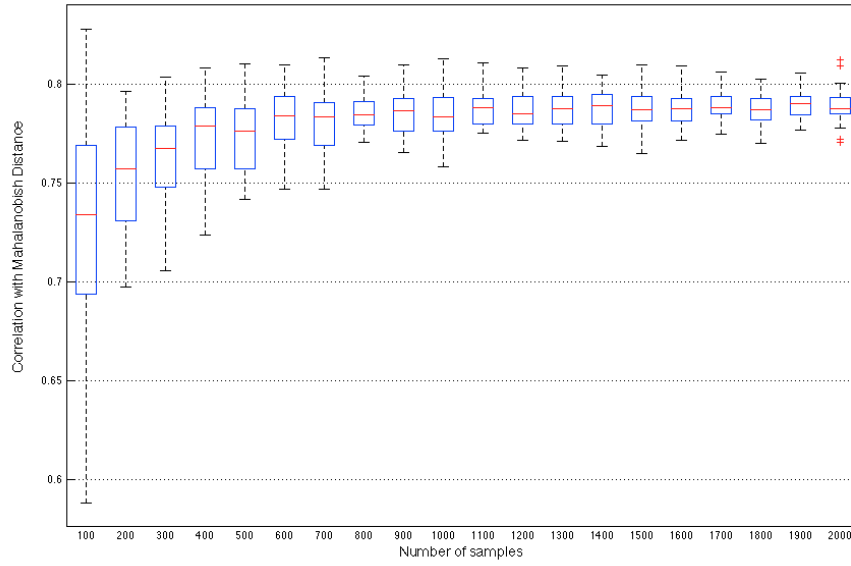
5.1 Simulation of Depth Function Based Local Mutual Information

As expressed by (34), the depth function based local mutual information i_{df} tends to the local mutual information of Gaussian variables i_G , but it does not require the estimation of the covariance matrix $\Sigma_{\mathbf{Z}}$ nor the mean vector $\mu_{\mathbf{Z}}$. The curse of dimensionality affects the approximation requiring more realizations when the dimension increases, such that more projections are required to obtain a better approximation. This fact is illustrated in Fig. 8(a), where several sets of realizations are drawn from k -dimensional Gaussian distributions $\mathbf{Z} \sim \mathcal{N}(\mu_{\mathbf{Z}}, \Sigma_{\mathbf{Z}})$. The minimum number of projections to obtain a significant local mutual information approximation is displayed with respect to the number of dimensions. As expected, the number of required projections explodes as the number of dimension increases. The optimal number of random projections is still an open question and it is beyond the scope of this paper. The second important issue is the number of realizations available for the estimation. In Fig. 8(b), the approximation power of i_G to the true i_{df} is assessed with different sets of Gaussian realizations. As the number of samples increases the approximation has a lower variance, until reaching a normalised correlation of 0.8. For example, 1000 samples are sufficient for a estimating the local mutual information in a 20-dimensional space.

Finally, the robustness to non Gaussian random variables is assessed by synthesising realizations from a mixture of Gaussians. To illustrate the performance of the depth function based local mutual information, we consider random realizations from mixtures of two Gaussian distributions $\mathbf{Z} \sim (1 - \beta)\mathcal{N}(\mu_0, \Sigma_Z) + \beta\mathcal{N}(\mu_1, \Sigma_Z)$, with different values of the mixing rate $\beta = \{0.05, 0.1, \dots, 0.3\}$. Both means μ_0 and μ_1 are selected to be separated by a distance equal to maximum variance of



(a)



(b)

Fig. 8 a) Minimum number of projections to get a correlation coefficient $\rho(i_{df}(\mathbf{x}; \mathbf{y}), i_G(\mathbf{x}; \mathbf{y})) > 0.8$, given a set of 5000 realizations $\{z_i = (x_i, y_i)\}$ following a Gaussian distribution $\mathcal{N}(\mu_{\mathbf{Z}}, \Sigma_{\mathbf{Z}})$. The figure shows the number of projections calculated 50 times for different values of the dimension k . b) Number of samples n vs $\rho(i_{df}(\mathbf{x}; \mathbf{y}), i_G(\mathbf{x}; \mathbf{y}))$, in 25 realizations from a k -variate Gaussian distribution $\mathbf{Z} \sim \mathcal{N}(\mu_{\mathbf{Z}}, \Sigma_{\mathbf{Z}})$, where the dimension is $k = 20$ and the number of projections is 2000.

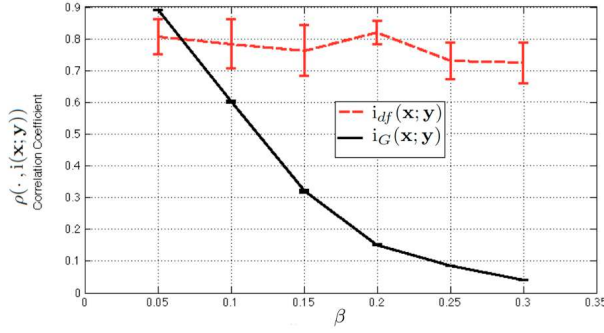


Fig. 9 The data are synthesised thanks to a mixture of Gaussian: $\mathbf{Z} \sim (1 - \beta)\mathcal{N}(\mu_0, \Sigma_{\mathbf{Z}}) + \beta\mathcal{N}(\mu_1, \Sigma_{\mathbf{Z}})$, where the dimension is $k = 4$ and the number of realizations is $n = 5000$. The correlation coefficients $\rho(i_{df}(\mathbf{x}; \mathbf{y}), i(\mathbf{x}; \mathbf{y}))$ and $\rho(i_G(\mathbf{x}; \mathbf{y}), i(\mathbf{x}; \mathbf{y}))$ are plotted for various values of the mixture rate β . 25 tests are performed for each definition and β .

$\Sigma_{\mathbf{Z}}$. As the analytic formulation of the distribution is available, the true local mutual information function $i(\mathbf{x}; \mathbf{y})$ can be computed and considered as a reference. Then, both estimates of the local mutual information i_{df} and i_G are compared to the reference i by computing the normalised correlations. The results are given in Fig. 9, where the robustness of the depth function is emphasised with respect to the Gaussian assumption. As the mixing rate increases, the Gaussian local mutual information fails in approximating correctly the dependence, while the depth function based measure succeeds in maintaining a robust approximation.

5.2 Hyperspectral Image Segmentation

The effectiveness of the proposed dissimilarities is illustrated on a hyperspectral image. This image covers an urban area including the university of Pavia, and was acquired by the ROSIS-03 optical sensor. The image has spatial dimensions of 610 by 340 pixels, with a spatial resolution of 1.3m per pixel and with spectral coverage of 115 bands ranging from 0.43 to $0.86\mu m$. The 12 most noisy channels are usually removed when experimenting with this data set [Plaza et al., 2006] so that our experiments are conducted on resulting 103-band image. The image and the corresponding edge maps are depicted in the top row of Fig. 10.

The thresholds $\alpha, \alpha_d, \alpha_g$ are selected such that the segmentations (see Figs. 10(f)–(h)) have the same number of connected components, i.e. 10^4 components. The results show that this criterion by itself tends to connect all the pixels creating a large connected component. In addition, the number of connected components with respect to the connectivity threshold is analysed in Fig. 11. While the L_1 norm based dissimilarity produces segmentations where the number of components drops immediately, the other dissimilarities produce segmentations where the number of components decreases smoothly in a first stage, before a sharper decrease. This analysis shows that more segmentation possibilities are obtained with the local mutual information moving from finer to coarser segmentations.

By considering a unique homogeneity predicate, segmentations (see Figs. 10(j)–(l)) are derived from the three dissimilarities where homogeneous components are

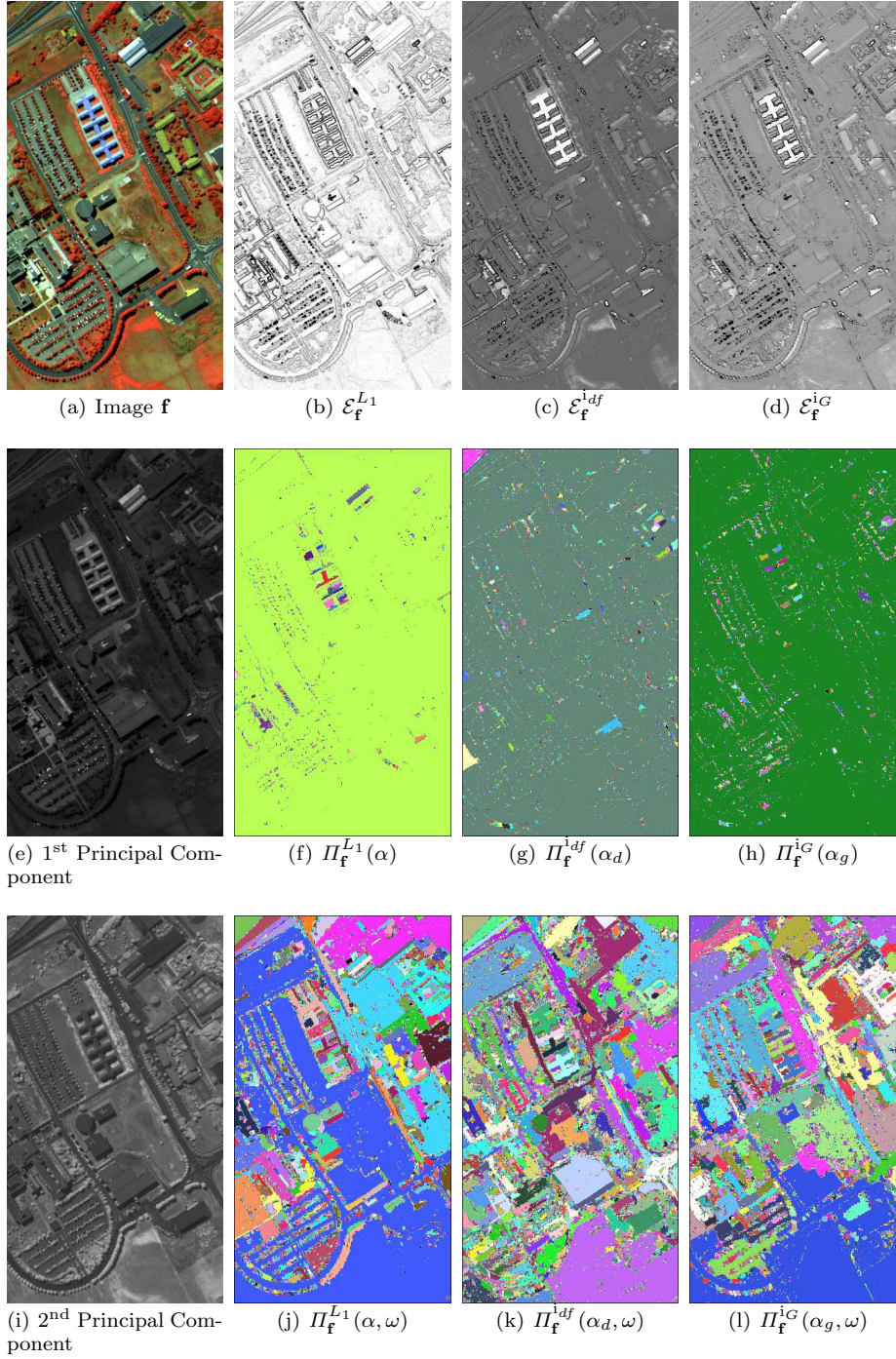


Fig. 10 a) The hyperspectral image is represented with the bands [80,70,30]. The edge maps of the three dissimilarities L_1 , i_{df} , i_G are represented in (b), (c), (d), respectively. e) The principal component with the highest variance is displayed. The unconstrained segmentations are displayed in (f)–(h). Finally, the constrained segmentations derived from the dissimilarities (b)–(d) are displayed in (j)–(l) by adopting a random colour for each connected component. The constraints are the same for the three segmentations, by using a common threshold $\omega = 140$ on the range of the first principal component rescaled in the interval $[0, 255]$.

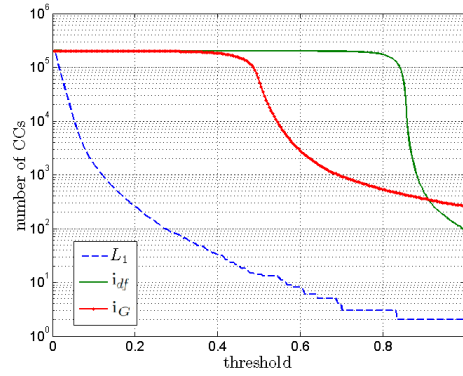


Fig. 11 The number of connected components with respect to the connectivity threshold is depicted. The three dissimilarities are considered and their thresholds are normalised in the range $[0, 1]$.

retained. The considered homogeneity criterion exploits the principal component (see Fig. 10(e)) derived by PCA [Pearson, 1901] from the hyperspectral image. The predicate sets the maximum range ω between the pixels values of the principal component belonging to a connected component. Due to the chaining effect [Duda et al., 2000, p. 554], the absolute difference based segmentation produces large components absorbing several objects of interest on the ground even though this effect is reduced by using the ω constraint. On the contrary, the local mutual information based segmentations mitigates these effects while producing relevant connected components. On the other hand, the local mutual information prevents the connection of pixels having values close to the global average spectral response. This phenomenon can be understood by analyzing the function $i_G(\mathbf{x}; \mathbf{y})$ (see Fig. 1(c)) along the line $\mathbf{x} = \mathbf{y}$, where the local mutual information increases for equal values $\mathbf{x} = \mathbf{y}$ away from the distribution center. Therefore, this property prevents the formation of connected components that would be obtained when considering the L_1 norm.

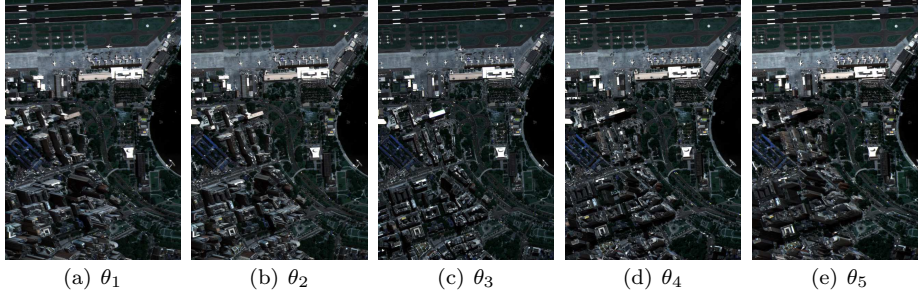
5.3 Multi-Angular Multi-Spectral Optical Images

In this experiment, the dissimilarities are computed from a multi-angular sequence of multi-spectral WorldView-2 images, where atmospheric spectral distortions increase with the off-nadir view angle [Longbotham et al., 2011]. Such distortions may be modelled by unknown bijective transforms of pixel spectral values from one image to the other [Gueguen et al., 2012], while ignoring the geometrical distortions. Therefore, the proposed dissimilarities should be invariant to these unknown bijective transforms and this property is investigated in this section.

A set of WorldView-2 multi-sequence images is considered for this experiment. This data set is composed of 5 Ortho Ready Standard WorldView-2 multi-angular acquisitions, including a 16 bit panchromatic image and a multispectral 8-band image. These images represent a subpart of the city of Rio in Brazil, and were acquired on January 2010. The 5 acquisition angles are summarised in table 1. The corresponding image sequence is depicted in Fig. 12, by displaying the red,

Table 1 The five acquisition configurations $\{\theta_1, \dots, \theta_5\}$ are summarised by the satellite azimuth and elevation, resulting in a decreasing sequence of the in off-nadir view angle.

	θ_1	θ_2	θ_3	θ_4	θ_5
Satellite Azimuth	12.5deg	12.2deg	9.4deg	193.7deg	193.2deg
Satellite Elevation	44.7deg	56.0deg	81.3deg	59.8deg	44.6deg
Off-nadir View Angle	39.2deg	29.8deg	7.6deg	-26.7deg	-39.5deg

**Fig. 12** The multi-angular sequence of multi-spectral images, having a size of $513 \times 768 \times 8$ and a resolution of 1.85m, is represented (true colour composition). The five images acquisitions parameters are summarised in table 1. Credit DigitalGlobe 2010.

green, and blue colour compositions.

In this experiment the multispectral images are analysed and the dissimilarities based on the L_1 norm and the local mutual informations i_{df} and i_G are considered. For notation simplicity, $\mathcal{E}_{\theta_i}^d$ denotes the edge map of the image θ_i where edges are computed thanks to the dissimilarity function d (22). Moreover, $\Pi_{\theta_i}^d(\alpha, \omega)$ denotes the segmentation of the image θ_i composed of the constrained connected components obtained with a dissimilarity threshold α and a range constraint threshold ω . In this case, the constraint is given by the range of the first principal component inside each component. In the described experiment, the various thresholds α are selected to obtain an equivalent number of connected components and ω is strictly the same for all segmentations. The edge maps and segmentations are depicted in Figs. 13-15 for the 3 dissimilarities, respectively.

By visual inspection of Fig. 13, as expected the edge maps obtained with the L_1 norm vary in contrast along the off-nadir view angle. While the most nadir image θ_3 produces sharp edges, the off-nadir edge maps $\mathcal{E}_{\theta_1}^{L_1}$, $\mathcal{E}_{\theta_5}^{L_1}$ exhibit lower contrast, thus less sharp edges. Accordingly, coarser segmentations are obtained as the image gets further from nadir acquisition, assuming a constant dissimilarity threshold α and a constant constraint threshold ω . This instability is noticeable in the series of segmentation displayed in Fig. 13. On the contrary the edge maps produced from the local mutual information based dissimilarity highlight stable contrast across angle (see Figs. 14-15). This behaviour can be explained by the invariance to unknown bijective transforms (proposition 4). Only the visual analysis of resulting segmentations indicates that i_{df} is less impacted by the acquisition angle than i_G , as the segmentation sequence based on i_{df} (see Fig. 14) appears more stable than the one based on i_G (see Fig. 15).

To quantitatively assess the impact of the acquisition off-nadir view angle on the dissimilarity measurements, the various edge maps are compared using the

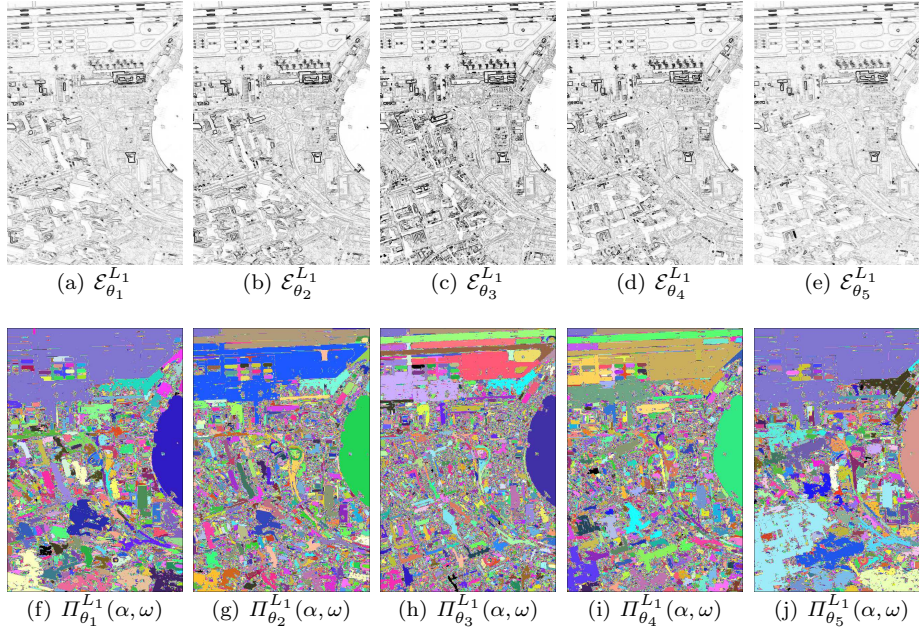


Fig. 13 The first image row contains the edge maps of the five successive multi-angular images, where the dissimilarity is expressed by the L_1 norm. The second image row contains the constrained segmentations obtained with a constant dissimilarity threshold α and a common constraint threshold $\omega = 150$.

Pearson's correlation coefficient which is expressed by

$$\rho(A, B) = \frac{\text{cov}(A, B)}{\sqrt{\text{var}(A)\text{var}(B)}}.$$

Note that this metric is impacted by the geometric distortions such as perspective distortions introduced by varying off-nadir view angles. The Pearson's correlation coefficients calculated between pairs of edge maps derived from the corresponding pair of off-nadir angles and using the dissimilarity measurements L_1 , i_{df} , and i_G are gathered in table 2. For a given dissimilarity measurement, the correlation coefficients between two edge maps decrease with the absolute difference of the off-nadir view angles θ_i . For example, the highest correlation is observed for all 3 dissimilarities for the image pair corresponding to the smallest angular difference, i.e., $9.6^\circ = |\theta_1 - \theta_2|$. A comparison between the values obtained for each dissimilarity measurement confirms that the depth function based local mutual information i_{df} produces the most similar edge maps, as the correlation coefficients are the highest for all image pairs. Moreover, the results show that the Gaussian local mutual information i_G is less stable than the L_1 norm, since the corresponding correlation coefficients are smaller. This observation shows that the depth function produces a closer estimate of the joint distribution than the Gaussian assumption, resulting in a higher stability of the subsequent edge maps.

In order to compare the multi-angular segmentations, we use the asymmetric normalised partition similarity. Having two segmentations $\Pi_A = \{C_i^A\}$ and

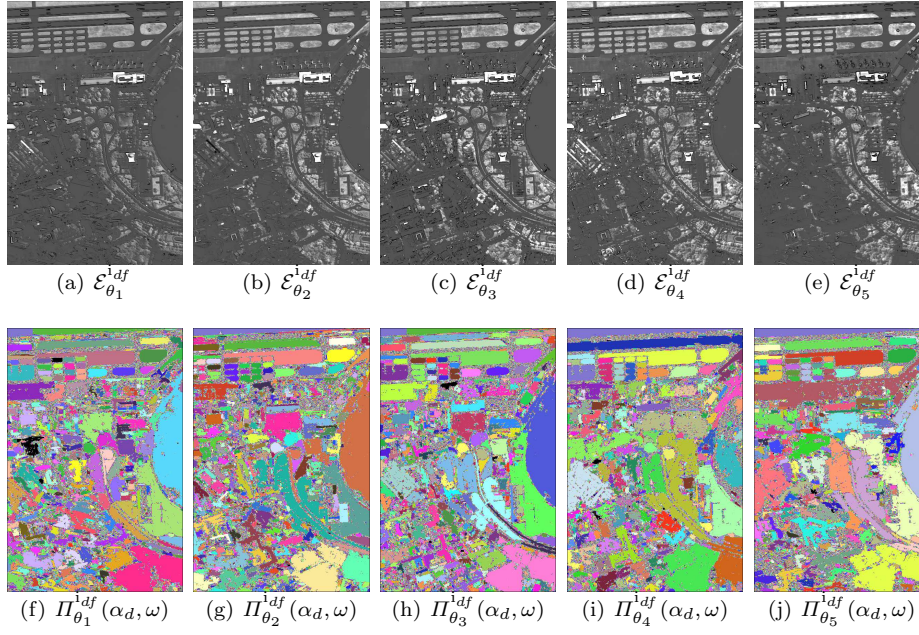


Fig. 14 The first image row contains the edge maps of the five successive multi-angular images, where the dissimilarity is expressed by the depth function based local mutual information i_{df} . The second image row contains the constrained segmentations obtained with a constant dissimilarity threshold α_d and a common constraint threshold $\omega = 150$. α_d is selected to obtain an equivalent number of CC as the one obtained with the L_1 norm and α .

$\Pi_B = \{C_j^B\}$ compound of disjoint sets covering the image grid set, the directional matching Γ between the segmentation sets is obtained by maximising the Jaccard metric \mathcal{J} [Jaccard, 1901]:

$$\Gamma^{\Pi_A \Rightarrow \Pi_B}(C_i^A) = \arg \max_{D \in \{C_j^B\}} \mathcal{J}(C_i^A, D), \quad (35)$$

$$\mathcal{J}(C_i^A, C_j^B) = \frac{|C_i^A \cap C_j^B|}{|C_i^A \cup C_j^B|}. \quad (36)$$

Definition 8 The asymmetric normalised partition similarities \mathcal{SJ} is defined as the weighted average of the Jaccard indices, giving more weights to larger components:

$$\mathcal{SJ}(\Pi_A \Rightarrow \Pi_B) = \frac{\sum_i |C_i^A| \mathcal{J}(C_i^A, \Gamma^{\Pi_A \Rightarrow \Pi_B}(C_i^A))}{\sum_i |C_i^A|}. \quad (37)$$

Comparable weighted partition similarity measurements are investigated in [Meila, 2003]. Identical partitions have a similarity \mathcal{SJ} equals to 1. If one segmentation is finer than the other one $\Pi_A \subseteq \Pi_B$, then $\mathcal{SJ}(\Pi_A \Rightarrow \Pi_B) \leq \mathcal{SJ}(\Pi_B \Rightarrow \Pi_A)$.

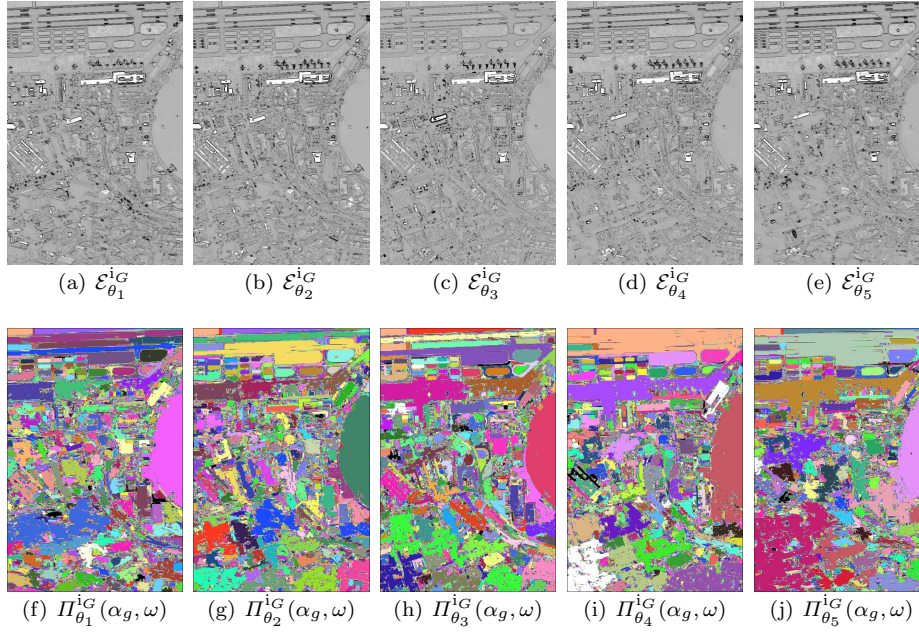


Fig. 15 The first image row contains the edge maps of the five successive multi-angular images, where the dissimilarity is expressed by the Gaussian local mutual information i_G . The second image row contains the constrained segmentations obtained with a constant dissimilarity threshold α_g and a common constraint threshold $\omega = 150$. α_g is selected to obtain an equivalent number of CC as the one obtained with the L_1 norm and α .

Definition 9 A *symmetric* normalised partition similarity is obtained by taking the average of the asymmetric normalised partition similarities:

$$\mathcal{SJ}(\Pi_A, \Pi_B) = \mathcal{SJ}(\Pi_B, \Pi_A) = \frac{\mathcal{SJ}(\Pi_A \Rightarrow \Pi_B) + \mathcal{SJ}(\Pi_B \Rightarrow \Pi_A)}{2}.$$

This metric is computed between the unconstrained segmentations obtained from the multi-angular sequence of images, selecting a constant dissimilarity threshold for all angles. The results are collected in table 3. It shows that the local mutual informations produce segmentations which are more stable than the ones obtained using the L_1 norm, despite the smaller cross correlation of the Gaussian mutual information. In addition, we observe that the depth function based local mutual information gives the segmentations which are the most similar. An analysis of the asymmetric normalised partition similarities also revealed that the segmentations obtained for the depth function based local mutual information are more balanced than those obtained for the other two dissimilarities. Indeed, the mean of the absolute difference of the asymmetric normalised partition similarities obtained for all pairs of multi-angular images is equal to 2.59, 2.38, and 3.34 for L_1 , i_{df} , and i_G respectively.

The numbers of connected components produced by the unconstrained segmentations Π performed on the multi-angular sequence are shown in Fig. 16. These numbers are given for the three dissimilarities and are represented with respect

Table 2 Pearson’s coefficients ρ obtained between the edge maps \mathcal{E} computed from the multi-angular sequence of off-nadir view angles θ_i . The highest normalised correlation coefficients among the 3 dissimilarities L_1 , i_{df} , and i_G are highlighted in bold.

$\rho(A, B)$	$\mathcal{E}_{\theta_1}^{L_1}$	$\mathcal{E}_{\theta_2}^{L_1}$	$\mathcal{E}_{\theta_3}^{L_1}$	$\mathcal{E}_{\theta_4}^{L_1}$	$\mathcal{E}_{\theta_5}^{L_1}$
$\mathcal{E}_{\theta_1}^{L_1}$	1.00	0.79	0.57	0.51	0.50
$\mathcal{E}_{\theta_2}^{L_1}$	-	1.00	0.67	0.54	0.51
$\mathcal{E}_{\theta_3}^{L_1}$	-	-	1.00	0.60	0.52
$\mathcal{E}_{\theta_4}^{L_1}$	-	-	-	1.00	0.75
$\mathcal{E}_{\theta_5}^{L_1}$	-	-	-	-	1.00
$\rho(A, B)$	$\mathcal{E}_{\theta_1}^{i_{df}}$	$\mathcal{E}_{\theta_2}^{i_{df}}$	$\mathcal{E}_{\theta_3}^{i_{df}}$	$\mathcal{E}_{\theta_4}^{i_{df}}$	$\mathcal{E}_{\theta_5}^{i_{df}}$
$\mathcal{E}_{\theta_1}^{i_{df}}$	1.00	0.86	0.73	0.66	0.63
$\mathcal{E}_{\theta_2}^{i_{df}}$	-	1.00	0.77	0.69	0.65
$\mathcal{E}_{\theta_3}^{i_{df}}$	-	-	1.00	0.75	0.67
$\mathcal{E}_{\theta_4}^{i_{df}}$	-	-	-	1.00	0.85
$\mathcal{E}_{\theta_5}^{i_{df}}$	-	-	-	-	1.00
$\rho(A, B)$	$\mathcal{E}_{\theta_1}^{i_G}$	$\mathcal{E}_{\theta_2}^{i_G}$	$\mathcal{E}_{\theta_3}^{i_G}$	$\mathcal{E}_{\theta_4}^{i_G}$	$\mathcal{E}_{\theta_5}^{i_G}$
$\mathcal{E}_{\theta_1}^{i_G}$	1.00	0.66	0.44	0.39	0.37
$\mathcal{E}_{\theta_2}^{i_G}$	-	1.00	0.55	0.43	0.37
$\mathcal{E}_{\theta_3}^{i_G}$	-	-	1.00	0.47	0.38
$\mathcal{E}_{\theta_4}^{i_G}$	-	-	-	1.00	0.62
$\mathcal{E}_{\theta_5}^{i_G}$	-	-	-	-	1.00

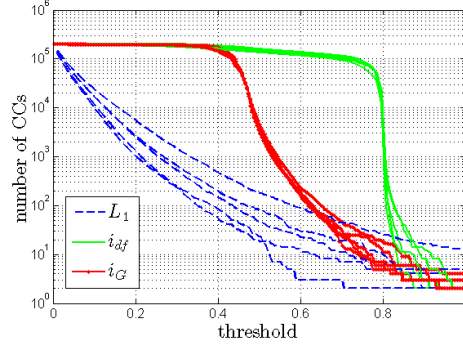


Fig. 16 The numbers of connected components resulting from the unconstrained segmentations obtained from the multi-angular image sequence are represented. The three dissimilarities are considered and their thresholds are normalised in the range $[0, 1]$.

to the normalized dissimilarity thresholds. This analysis highlights the variability (with respect to the angle) of the number of connected components in the hierarchies induced by the considered dissimilarities. The results confirm the higher stability of the dependent connectivities across the view angles.

In summary, these experiments show that the depth function based local mutual information is less sensitive to unknown spectral distortions such as those occurring for multi-angular views of the same scene.

Table 3 Symmetric normalised partition similarities \mathcal{SJ} obtained between the unconstrained segmentations Π performed on the multi-angular sequence. The dissimilarity thresholds α, α_d, m_g are selected to obtain equivalent numbers of connected components for the most nadir image θ_3 . The highest partition similarities among the 3 dissimilarities are highlighted in bold.

$\mathcal{SJ}(A, B)$	$\Pi_{\theta_1}^{L_1}(\alpha)$	$\Pi_{\theta_2}^{L_1}(\alpha)$	$\Pi_{\theta_3}^{L_1}(\alpha)$	$\Pi_{\theta_4}^{L_1}(\alpha)$	$\Pi_{\theta_5}^{L_1}(\alpha)$
$\Pi_{\theta_1}^{L_1}(\alpha)$	1.0	0.27	0.18	0.19	0.22
$\Pi_{\theta_2}^{L_1}(\alpha)$	-	1.0	0.27	0.24	0.19
$\Pi_{\theta_3}^{L_1}(\alpha)$	-	-	1.0	0.28	0.17
$\Pi_{\theta_4}^{L_1}(\alpha)$	-	-	-	1.0	0.21
$\Pi_{\theta_5}^{L_1}(\alpha)$	-	-	-	-	1.0

$\mathcal{SJ}(A, B)$	$\Pi_{\theta_1}^{idf}(\alpha_d)$	$\Pi_{\theta_2}^{idf}(\alpha_d)$	$\Pi_{\theta_3}^{idf}(\alpha_d)$	$\Pi_{\theta_4}^{idf}(\alpha_d)$	$\Pi_{\theta_5}^{idf}(\alpha_d)$
$\Pi_{\theta_1}^{idf}(\alpha_d)$	1.0	0.56	0.49	0.38	0.44
$\Pi_{\theta_2}^{idf}(\alpha_d)$	-	1.0	0.52	0.39	0.44
$\Pi_{\theta_3}^{idf}(\alpha_d)$	-	-	1.0	0.40	0.46
$\Pi_{\theta_4}^{idf}(\alpha_d)$	-	-	-	1.0	0.46
$\Pi_{\theta_5}^{idf}(\alpha_d)$	-	-	-	-	1.0

$\mathcal{SJ}(A, B)$	$\Pi_{\theta_1}^{iG}(\alpha_g)$	$\Pi_{\theta_2}^{iG}(\alpha_g)$	$\Pi_{\theta_3}^{iG}(\alpha_g)$	$\Pi_{\theta_4}^{iG}(\alpha_g)$	$\Pi_{\theta_5}^{iG}(\alpha_g)$
$\Pi_{\theta_1}^{iG}(\alpha_g)$	1.0	0.33	0.28	0.28	0.24
$\Pi_{\theta_2}^{iG}(\alpha_g)$	-	1.0	0.33	0.28	0.26
$\Pi_{\theta_3}^{iG}(\alpha_g)$	-	-	1.0	0.31	0.26
$\Pi_{\theta_4}^{iG}(\alpha_g)$	-	-	-	1.0	0.32
$\Pi_{\theta_5}^{iG}(\alpha_g)$	-	-	-	-	1.0

6 Conclusion

The dependent connectivity is presented as an alternative connection measure to perform image segmentation. The dependence measure is based on the local mutual information between adjacent pixel values. This measure is extended to multi-dimensional images by the introduction of the depth function being robust in high dimensions. This dependence measure is shown to be invariant to any differentiable bijective transform, such that the induced image segmentations have the same invariance property. Experiments with hyper-spectral image show the interests of such dissimilarity to produce segmentations which are less sensitive to the leakage effect. Other experiments with multi-angular multi-spectral images highlight the higher stability of the proposed approach with respect to unknown bijective transforms.

References

- P. Arbelaez and L. Cohen. A metric approach to vector-valued image segmentation. *International Journal of Computer Vision*, 69:119–126, 2006. URL <http://dx.doi.org/10.1007/s11263-006-6857-5>.
- A. Baraldi and F. Parmiggiani. Single linkage region growing algorithms based on the vector degree of match. *IEEE Transactions on Geoscience and Remote Sensing*, 34(1):137–148, January 1996. doi: 10.1109/36.481899.

- R. Bellman. *Adaptive Control Processes: A Guided Tour*. Princeton University Press, 1961.
- G. Bouma. Normalized (pointwise) mutual information in collocation extraction. In C. Chiarcos, E. Eckart de Castilho, and M. Stede, editors, *Proceedings of the Biennial GSCL Conference 2009*, pages 31–40, Potsdam, Germany, Sept. 2009. Gunter Narr Verlag. URL <http://www.ling.uni-potsdam.de/~gerlof/docs/npmi-pfd.pdf>.
- U. Cherubini, E. Luciano, and W. Vecchiato. *Copula Methods in Finance*. John Wiley & Sons, 2004.
- K.W. Church and P. Hanks. Word association norms, mutual information, and lexicography. *Comput. Linguist.*, 16:22–29, March 1990. URL <http://dl.acm.org/citation.cfm?id=89086.89095>.
- P. Comon. Independent component analysis: a new concept. *Elsevier Signal Processing*, 36(3):287–314, 1994.
- T.M. Cover and J.A. Thomas. *Elements of information theory*. John Wiley & Sons, New York, 1991.
- J.A. Cuesta-Albertos and A. Nietos-Reyes. The random Tukey depth. *Computational Statistics and Data Analysis*, 52:4979–4988, 2008.
- A. Cuevas, M. Febrero, and R. Fraiman. Robust estimation and classification for functional data via projection-based depth notions. *Computational Statistics*, 22(3):481–496, 2007.
- X. Cui, L. Lin, and G. Yang. An extended projection data depth and its application to discrimination. *Communication in Statistics - Theory and Methods*, 37(14):2276–2290, 2008.
- D. Donoho and M. Gasko. Breakdown properties of location estimates based on halfspace depth and projected outlyingness. *Annals of Statistics*, 20(4):1803–1827, 1992.
- R. Duda, P. Hart, and D. Stork. *Pattern Classification*. Wiley Interscience, 2000.
- J. Gower and G. Ross. Minimum spanning trees and single linkage cluster analysis. *Applied statistics*, 18(1):54–64, 1969.
- A. Green, M. Berman, P. Switzer, and M. Craig. A transformation for ordering multispectral data in terms of image quality with implications for noise removal. *IEEE Transactions on Geoscience and Remote Sensing*, 26(1):65–74, 1988. doi: 10.1109/36.3001.
- L. Gueguen and M. Datcu. Mixed information measure: Application to change detection in earth observation. In *MultiTemp 2009: The Fifth International Workshop on the Analysis of Multi-temporal Remote Sensing Images*, Connecticut, USA, June 2009.
- L. Gueguen and P. Soille. Frequent and dependent connectivities. In *Proc. of 10th Int. Symp. on Mathematical Morphology*, volume 6671 of *Lecture Notes in Computer Science*, pages 120–131. Springer-Verlag, 2011. doi: 10.1007/978-3-642-21569-8_11.
- L. Gueguen, P. Soille, and M. Pesaresi. Change detection based on information measure. *IEEE Transactions on Geoscience and Remote Sensing*, 49(11):4503–4515, November 2011. doi: 10.1109/TGRS.2011.2141999.
- L. Gueguen, S. Velasco-Forero, and P. Soille. Edge extraction by statistical dependence analysis: application to multi-angular WorldView-2 series. In *IEEE International Geoscience and Remote Sensing Symposium*, pages 3447–3450, Munich, Germany, July 2012. doi: 10.1109/IGARSS.2012.6350679.

- L. Guigues, J.-P. Cocquerez, and H. Le Men. Scale-sets image analysis. *International Journal of Computer Vision*, 68(3):289–317, 2006. doi: 10.1007/s11263-005-6299-0.
- R. Haralick, K. Shanmugam, and I. Dinstein. Textural features for image classification. *IEEE Transactions on Systems, Man, and Cybernetics*, 3(6):610–621, November 1973. doi: 10.1109/TSMC.1973.4309314s.
- G. Hermosillo, C. Chefd’Hotel, and O. Faugeras. Variational methods for multimodal image matching. *International Journal of Computer Vision*, 50:329–343, 2002. doi: 10.1023/A:1020830525823.
- P. Jaccard. Distribution de la flore alpine dans le bassin des dranses et dans quelques regions voisines. *Bulletin de la Societe Vaudoise des Sciences Naturelles*, 37:241–272, 1901.
- R.A. Johnson and D.W. Wichern. *Applied Multivariate Statistics Analysis*. Prentice-Hall International Inc., 2007.
- I.M. Johnstone. On the distribution of the largest eigenvalue in principal components analysis. *Annals of Statistics*, 29:295–327, 2001.
- B. Karacali. Information theoretic deformable registration using local image information. *International Journal of Computer Vision*, 72:219–237, 2007. doi: 10.1007/s11263-006-8704-0.
- F. Kruse, A. Lefkoff, J. Boardman, K. Heidebrecht, A. Shapiro, P. Barloon, and A. Goetz. The spectral image processing system (SIPS)—Interactive visualization and analysis of imaging spectrometer data. *Remote Sensing of Environment*, 44(2–3):145–163, May–June 1993. doi: 10.1016/0034-4257(93)90013-N.
- J. Kruskal. On the shortest spanning subtree of a graph and the traveling salesman problem. *Proc. Am. Math. Soc.*, 7(1):48–50, 1956.
- R. Liu. On a notion of data depth based on random simplices. *Annals of Statistics*, 18(1):405–482, 1990.
- N. Longbotham, W. Emery, C. Bleiler, C. Chaapel, C. Padwick, and F. Pacifici. Spectral classification of WorldView-2 multi-angle sequence. In *IEEE International Geoscience & Remote Sensing Symposium*, Vancouver, Canada, July 2011.
- F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens. Multimodality image registration by maximization of mutual information. *Medical Imaging, IEEE Transactions on*, 16(2):187–198, april 1997. doi: 10.1109/42.563664.
- M. Meila. Comparing clusterings by the variation of information. In *Proc. 16th Annual Conference on Computational Learning Theory and 7th Kernel Workshop, COLT/Kernel*, volume 2777 of *Lecture Notes in Artificial Intelligence*, pages 173–187, 2003. doi: 10.1007/978-3-540-45167-9_14.
- F. Meyer and P. Maragos. Morphological scale-space representation with levelings. *Lecture Notes in Computer Science*, 1682:187–198, September 1999.
- F. Meyer and P. Maragos. Nonlinear scale-space representation with morphological levelings. *Journal of Visual Communication and Image Representation*, 11(3):245–265, 2000. doi: 10.1006/jvci.1999.0447.
- M. Nagao, T. Matsuyama, and Y. Ikeda. Region extraction and shape analysis in aerial photographs. *Computer Graphics and Image Processing*, 10(3):195–223, 1979.
- G. Ouzounis and P. Soille. Pattern spectra from partition pyramids and hierarchies. In *Proc. of 10th Int. Symp. on Mathematical Morphology*, volume 6671 of *Lecture Notes in Computer Science*, pages 108–119. Springer-Verlag, 2011. doi:

- 10.1007/978-3-642-21569-8_10.
- G. Ouzounis and P. Soille. The alpha-tree algorithm. Technical Report 25500 EN, Joint Research Centre, European Commission, September 2012.
- D.P. Palomar and S. Verdu. Lantum information. *IEEE Trans. Information Theory*, 54(3):964–975, March 2008.
- G. Panin and A. Knoll. Mutual information-based 3D object tracking. *International Journal of Computer Vision*, 78:107–118, 2008. doi: 10.1007/s11263-007-0083-7.
- K. Pearson. On lines and planes of closest fit to systems of points in space. *Philosophical Magazine*, 2(6):559–572, 1901.
- A. Plaza, J.A. Benediktsson, J. Boardman, J. Brazile, L. Bruzzone, G. Camps-Valls, J. Chanussot, M. Fauvel, P. Gamba, A. Gualtieri, J.C. Tilton, and G. Trianni. Advanced processing of hyperspectral images. In *Geoscience and Remote Sensing Symposium, 2006. IGARSS 2006. IEEE International Conference on*, pages 1974–1978, August 2006. doi: 10.1109/IGARSS.2006.511.
- M. Radovanović, A. Nanopoulos, and M. Ivanović. Hubs in space: Popular nearest neighbors in high-dimensional data. *J. Mach. Learn. Res.*, 11:2487–2531, 2010.
- C. Ronse. Partial partitions, partial connections and connective segmentation. *Journal of Mathematical Imaging and Vision*, 32(2):97–125, October 2008. doi: 10.1007/s10851-008-0090-5.
- C. Ronse. Idempotent block splitting on partial partitions. Part II: Non-isotone operators. *Orders*, 28(2):307–339, 2011. doi: 10.1007/s11083-010-9190-0.
- C. Ronse and J. Serra. Fondements algébriques de la morphologie. In L. Najman and H. Talbot, editors, *Morphologie Mathématique 1: Approches Déterministes*, chapter 2, pages 49–96. Lavoisier, Paris, 2008.
- L. Sachs. *Applied Statistics: A Handbook of Techniques*. New York: Springer-Verlag, 1984.
- J. Serra. Mathematical morphology for Boolean lattices. In J. Serra, editor, *Image Analysis and Mathematical Morphology. Volume 2: Theoretical Advances*, chapter 2, pages 37–58. Academic Press, 1988.
- J. Serra. A lattice approach to image segmentation. *Journal of Mathematical Imaging and Vision*, 24:83–130, 2006.
- R. Sibson. SLINK: An optimally efficient algorithm for the single-link cluster method. *The Computer Journal*, 16:30–34, 1973.
- P. Soille. On genuine connectivity relations based on logical predicates. In *Proc. of 14th Int. Conf. on Image Analysis and Processing, Modena, Italy*, pages 487–492. IEEE Computer Society Press, September 2007.
- P. Soille. Constrained connectivity for hierarchical image partitioning and simplification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(7):1132–1145, July 2008.
- P. Soille. Preventing chaining through transitions while favouring it within homogeneous regions. In *Proc. of ISMM 2011*, volume 6671 of *Lecture Notes in Computer Science*, 2011.
- P. Soille and J. Grazzini. Constrained connectivity and transition regions. In *Proc. of 9th Int. Symp. on Mathematical Morphology*, volume 5720 of *Lecture Notes in Computer Science*, pages 59–69, 2009. doi: 10.1007/978-3-642-03613-2_6.
- Z. Sun and A. Hoogs. Image comparison by compound disjoint information with applications to perceptual visual quality assessment, image registration and tracking. *International Journal of Computer Vision*, 88:461–488, 2010. doi:

- 10.1007/s11263-010-0316-z.
- J. Theiler and S. Perkins. Proposed framework for anomalous change detection. In *ICML Workshop on Machine Learning Algorithms for Surveillance and Event Detection*, pages 7–14, 2006.
- S. Velasco-Forero and J. Angulo. Mathematical morphology for vector images using statistical depth. In *Proc. of 10th Int. Symp. on Mathematical Morphology*, volume 6671 of *Lecture Notes in Computer Science*, pages 355–366. Springer-Verlag, 2011. doi: 10.1007/978-3-642-21569-8_31.
- S. Velasco-Forero and J. Angulo. Random projection depth for multivariate mathematical morphology. *IEEE Journal of Selected Topics in Signal Processing*, 6(7):753–763, nov. 2012. ISSN 1932-4553. doi: 10.1109/JSTSP.2012.2211336.
- P. Viola and W. Wells. Alignment by maximization of mutual information. *International Journal of Computer Vision*, 24:137–154, 1997. doi: 10.1023/A:1007958904918.
- A. Winter, H. Maitre, N. Cambou, and E. Legrand. Entropy and multiscale analysis: a new feature extraction algorithm for aerial images. In *Acoustics, Speech, and Signal Processing, 1997. ICASSP-97., 1997 IEEE International Conference on*, volume 4, pages 2765–2768, April 1997. doi: 10.1109/ICASSP.1997.595362.
- F. Zanoguera and F. Meyer. On the implementation of non-separable vector levelings. In *Proceedings of VIIth Int. Symp. on Mathematical Morphology*, pages 369–377, Sydney, Australia, 2002. Commonwealth Scientific and Industrial Research Organisation. URL http://www.cmis.csiro.au/ismm2002/proceedings/PDF/33_zanoguera.pdf.
- H. Zhang, J. Fritts, and S. Goldman. An entropy-based objective evaluation method for image segmentation. *Proc. SPIE (International Society for Optics and Photonics)*, 5307(1):38–49, 2003. doi: 10.1117/12.527167.
- Y. Zuo. Projection-based depth function and associated medians. *Annals of Statistics*, 31(5):1460–1490, 2003.
- Y. Zuo. Multidimensional trimming based on projection depth. *Annals of Statistics*, 34(5):2211–2251, 2006.
- Y. Zuo and R. Serfling. General notions of statistical depth function. *Annals of Statistics*, 28(2):461–482, 2000.



Lionel Gueguen received the engineering degree in telecommunications and the MS degree in signal and image processing from the Ecole Nationale Supérieure des Télécommunications de Bretagne, Brest, France, in 2004. He received a PhD degree in signal and image processing from Ecole Nationale Supérieure des Télécommunications Paris, in 2007. He worked on "Joint Information Extraction and Compression of Satellite Image Time Series" under the supervision of Mihai Datcu, and was a member of the Center Of Competence on Information Extraction and Image Understanding for Earth Observation. From 2007 to 2009, he worked as a post-doc researcher at Orange Labs, France, on radio measurements processing for cognitive radio applications. From July 2009 to June

2012, he worked with IPSC, Support to External Security Unit, ISFEREA Action at the Joint Research Center of the European Commission. During his three years with JRC, he worked in particular on tents counting and change analysis of refugee camps, roofless building detection and the Global Human Settlement Layer production. In September 2012, he joined DigitalGlobe Inc. as a R&D scientist to study and develop Image Information Mining techniques applied to electro-optical satellite images.



Santiago Velasco-Forero was born in Bogotá, Colombia. He received a degree in Statistics from the National University of Colombia, Bogotá Campus, M.Sc. in Mathematics in University of Puerto Rico, Mayaguez Campus, and Ph.D. in mathematical morphology and image processing at Ecole des Mines de Paris, France. His research interests are image processing, multivariate statistics, computer vision, pattern recognition and mathematical morphology and their application to biotechnology and remote sensing. Currently, he is a postdoc researcher at ITWM - Fraunhofer Institute in Kaiserlautern, Germany.



Pierre Soille received the engineering degree in 1988 from the Université catholique de Louvain (Belgium), the doctorate degree in 1992 from the same university in collaboration with the Ecole des Mines de Paris (France), and the habilitation degree in 1997 from the Université Montpellier II (France). During the period 1992-1999, he pursued research on image analysis and processing with the CSIRO Mathematical and Information Sciences (Sydney, Australia), the Centre de Morphologie Mathématique of the Ecole des Mines de Paris (Fontainebleau, France), the Pattern Recognition Department of the Fraunhofer-Institut IPK (Berlin, Germany), the Ecole des Mines d'Alès and EERIE (Nîmes, France), and the Biotechnology and Biological Sciences Research Council (Silsoe, U.K.). Since December 1999, he is with the Joint Research Centre of the European Commission (Ispra, Italy). With the former Space Applications Institute and then the Institute for Environment and Sustainability, he contributed to the development of numerous advanced methodologies and algorithms for the generation of enhanced pan-European geospatial data sets. With the Institute for the Protection and Security of the Citizen, he is now focusing on automatic information retrieval from very high resolution satellite image data.