INTRINSIC DIMENSIONALITY ESTIMATION AND DIMENSIONALITY REDUCTION THROUGH SCALE SPACE FILTERING

Konstantinos Karantzalos

Laboratoire de Mathematiques Appliquees aux Systemes (MAS)
Ecole Centrale de Paris, Chatenay-Malabry, France
konstantinos.karantzalos@ecp.fr
http://www.mas.ecp.fr/vision/Personnel/karank/Demos.html

ABSTRACT

Dimensionality reduction techniques are designed to exploit the fact that most high-dimensional datasets from the real world do not uniformly fill the hyperspaces in which they are represented but instead their distributions usually concentrate to nonlinear manifolds of lower intrinsic dimensions. However, when these techniques are applied directly to the initial degraded and noisy data, the assumptions on the possible statistical separation of real world classes do not, in the general case, hold. In this paper, we argue that scale space filtering, by denoising and simplifying effectively the initial dataset, ameliorate the way the properties of our observations are been encoded, strengthening, thus, the assumptions on the possible statistical separation of real world classes. Experimental results on real hyperspectral datasets demonstrate that appropriate vector-valued scale space filtering significantly contributes to the intrinsic dimension estimation and dimensionality reduction of high dimensional datasets.

Index Terms— Multidimensional Data, Hyperspectral Imaging, Manifold Learning, Mathematical Morphology, Levelings, Anisotropic Diffusion, Image Simplification, Denoising

1. INTRODUCTION

The basic goal of hyperspectral data processing, in almost all application areas, is to classify or identify objects. One would expect that, as the number of hyperspectral bands increases, the accuracy of classification should, also, increase. However, this is not the case [1] and in particular, the reported average classification accuracy of remote sensing imagery is about 73% and it has not changed significantly in recent years. Therefore, although the hyperspectral imaging market is rapidly increasing there still remain several challenges, regarding their multidimensional data processing. Denoising high dimensional data and optimally reducing their dimensionality are two major ones [2].

On the one hand, the natural variability of the material spectra, noise, physical disturbances and degradation added by the transmission media and the sensor system, reduce the discrimination of the different structures in hyperspectral imagery and diminish the accuracy of subsequent segmentation and classification processes. The increased significance of smaller spatial and spectral variations among pixels implies, also, that smaller amounts of noise are now likely to have a bigger impact on the results extracted from this kind of imagery. Even thought any denoising process has a significant impact on the accuracy of the results, many studies do not use any strict optimizing criteria when selecting the appropriate smoothing methods, thus, negatively affecting the outcome of subsequent analysis [3].

On the other hand, dimensionality reduction [4], which is a transformation into a meaningful representation of reduced dimensionality, is a crusial step at most high-dimensional data processing procedures. Both linear and nonlinear techniques ([5] and the references therein), which may discard some bands which contain valuable information or project and blur data into a low-dimensional subspace, are actually a trade-off between making the problem simpler and losing on classification accuracy [2, 6, 7]. The assumptions on the possible statistical interpretation/ separation of terrain classes do not, in the general case, hold when these methods are applied directly to the initial degraded and noisy hypercube an not to an elegantly simplified version of it.

In this paper, we argue that vectorial scale space filtering ameliorates intrinsic dimensionality estimation and dimensionality reduction procedures. Firstly, by choosing appropriate nonlinear scale space representations noise can be removed and at the same time the data are efficiently simplified. Secondly, these elegantly simplified hypercubes by describing in a more distinct way the spectral and spatial signatures of our observations, form more appropriate versions of the initial/raw data for estimating their intrinsic dimensionality. Moreover, the computation of complex lower dimensional manifolds -either with linear/nonlinear, local or global approaches- is more efficient since the variance of the embedded hyperspace is more compact. Experimental results on real data and the performed qualitatively and quantitative as-

978-1-4244-3298-1/09/$25.00 ©2009 IEEE DSP 2009

sessment, demonstrate the potentials of advanced scale space filtering for the tasks of intrinsic dimensionality estimation and dimensionality reduction.

2. SCALE SPACE ON THE HYPERCUBE

Lets denote with $I : \Omega \subset \mathcal{R}^d \rightarrow \mathcal{R}^N$ a hyperspectral image with a normalized hyperspectrum of N spectral channels. During our experiments, vector-valued Anisotropic Diffusion Filtering (ADF) and Anisotropic Morphological Levelings (AML) [8], [9] were employed and are briefly detailed in this section.

Excluding atmospheric effects which are tackled during a specific atmospheric correction stage, the dark or photon shot noise and the readout noise, which appears as uncorrelated high-frequency variations in the spatial and spectral space without forming a coherent structure, is what a filtering procedure should be able to address [8]. However, unconstrained spatial smoothing is not desirable and in addition, spectral resolution and band adjacency are, usually, high enough to assume that the spectral vector is a good approximation to the spectral signature of the pixel, i.e the mixture of the spectral signatures of the objects within the pixel plus atmospheric, scatter and radiometric effects. Last but not least, in the spatial directions all the desired properties of the 2D levelings must be retained. To sum up a sophisticated vectorial leveling formulation should retain all its 2D properties for the spatial directions and at the same time respect gross variations among adjacent spectral signatures and only suppress the broad spectral variations (spike-like features).

Towards this end, levelings' construction mechanism is formulated in a way to include a comparison with the adjacent spectral signatures:

$$f \wedge (\delta g_s \vee \delta' g_c) \leq g \leq f \vee (\varepsilon g_s \wedge \varepsilon' g_c) \qquad (1)$$

where δg_s denotes an extensive marker in the spatial axis and $\delta' g_c$ an extensive marker in the spectral one (the anti-extensive operators εg are equally defined). The spatial g_s marker acts as in the 2D case ensuring an elegant simplification in the spatial neighborhood of a pixel and the spectral g_c accounts for the spike-like features by enforcing its relevant operators (δ' and ε') to have a much broader effect. Under this framework and employing always a marker function h for levelings' construction the process is decomposed and the spectral and spatial spaces are treated differently according to the posed constrains. Rephrasing Equation (1) and in a unique parallel step we have that:

$$g = \Lambda(f, h) = (f \wedge (\delta h_s \vee \delta' h_c)) \vee (\varepsilon h_s \wedge \varepsilon' h_c) \qquad (2)$$

Hyperspectral data can be viewed like any video data, where the wavelength corresponds to time or like MRI volumes in medical imaging, where wavelength corresponds to another spatial axis. Instead of defining the stack of a hyperspectral image as $I : \Omega \subset \mathcal{R}^d \rightarrow \mathcal{R}^N$, where N is the number

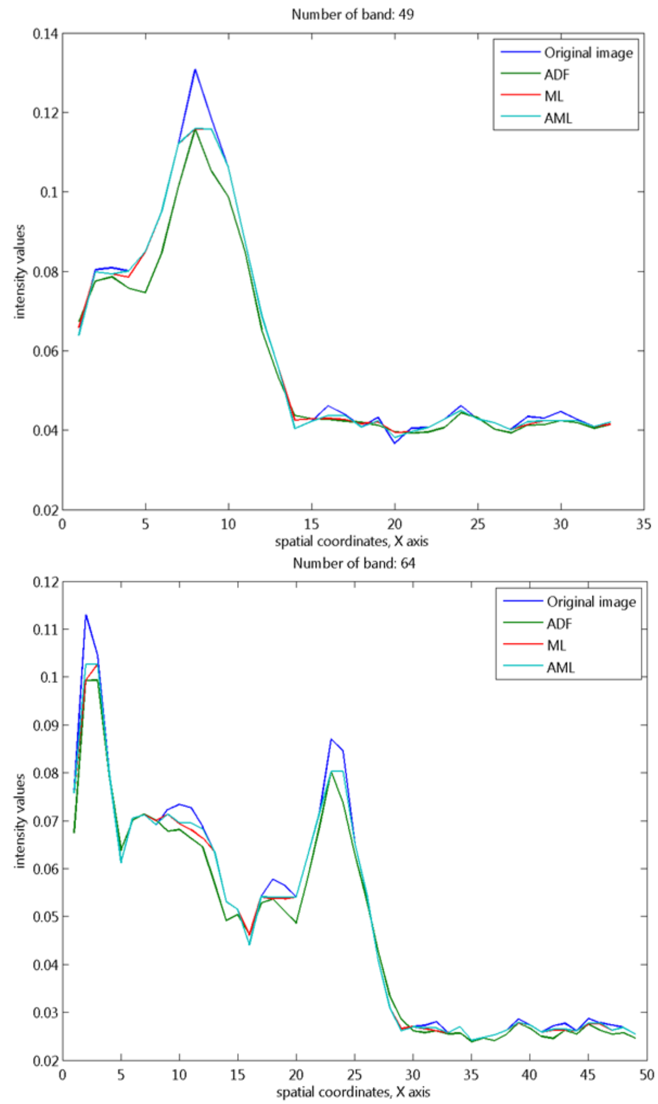

Fig. 1. Spatial simplification: Comparing the filtering result of ADF, ML (channel by channel process) and AML. Two line plots with the cross-sections along the y-axis of the different filters are shown. AML did simplify the initial image by enlarging and creating new flat zones and at the same time followed more constantly and closely original image's intensity values and variation. AML did retain all its elegant 2D properties.

of spectral channels and $I = (I_1(x, y), ..., I_N(x, y)) \in \mathcal{R}^N$, a hypercube can be defined, also, as a 3D function $\mathcal{I} : \Omega \subset \mathcal{R}^3 \rightarrow \mathcal{R}$, where $\mathcal{I}(x, y, z) = I_z(x, y)$).

Following this notation, multiscale levelings can be constructed when the initial (reference) hypercube $\mathcal{I}$ is associated with a series of marker functions $\{h_1, h_2, ..., h_n\}$ -all h are increasingly smoother hypercubes in $\mathcal{R}^3$. The constructed

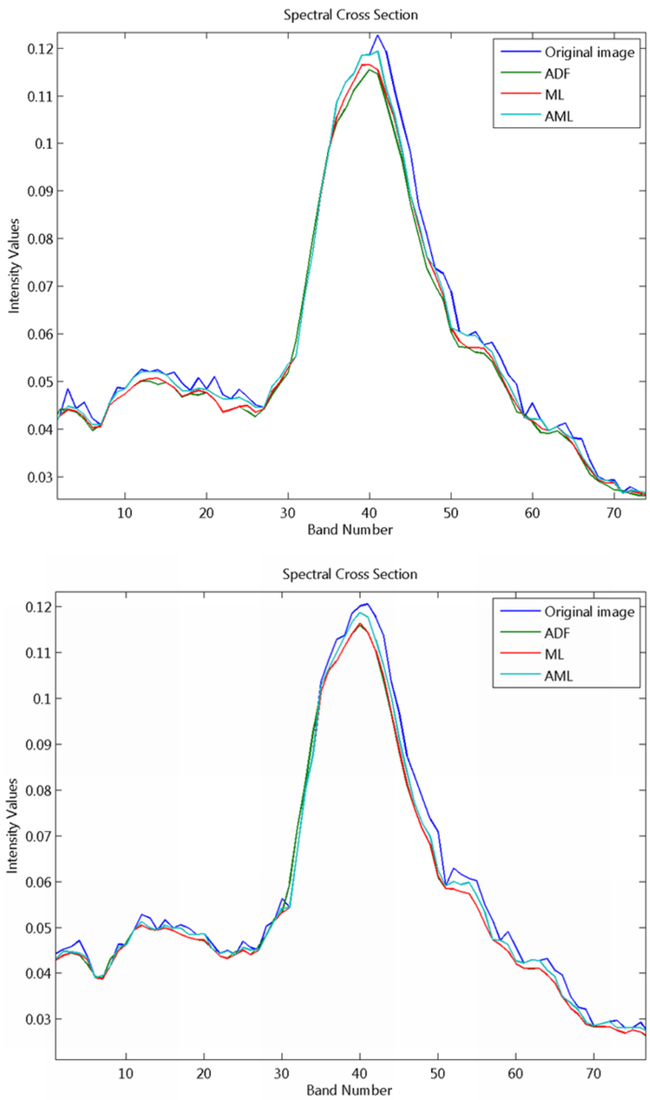

Fig. 2. Spectral simplification: Comparing the filtering result of ADF, ML and AML. Two line plots with the cross-sections along the spectral axis of the different filtered hypercubes are shown. The proposed AML did surpassed broad spectral variations (spike-like features) among adjacent spectral signatures and at the same time followed more constantly the initial intensity.

levelings are respectively

$$g_1 = \mathcal{I},\;\; g_2 = \Lambda(g_1, h_1),\;\; g_3 = \Lambda(g_2, h_2), \\ g_4 = \Lambda(g_3, h_3)\;, ...,\; g_n = \Lambda(g_{n-1}, h_{n-1}) \tag{3}$$

A series g_n of simpler and simpler hypercubes, with fewer and fewer smooth zones are produced forming a 4D scale space with $g : \Omega \subset \mathcal{R}^4$ and $g(x, y, z, n) = g_n(x, y, z)$. Anisotropic diffused markers were chosen, since they have proven to be effective for scalar images [10]. In addition,

Table 1. Quantitative Evaluation for ADF and AML methods

Test Data	Filter Type	Quantitative Measures		
		RMSE	NMSE	SSIM
HYDICE dataset	ADF	0.012	0.009	0.996
	ML	0.009	0.004	0.998
	AML	0.006	0.002	0.999

since levelings are highly constrained by the type of the marker used [11], only those markers who are fully suitable for hyperspectral imagery were appropriate for our case. The recent formulations of [8] provide a suitable diffusion framework which respects the special characteristics of hyperspectral data by separating the elegant vector-valued diffusion approach of [12] in the spatial and spectral space.

Following such a formulation, ADF and AML can efficiently denoise and smooth data. In particular, as it is demonstrated at Figures 1 and 2, where cross sections along the spatial y-axis and the spectral axis are presented, ADF smoothed strongly the data -but created some intensity shifts, in contrast to AML which simplified the data but kept a closer relation with the initial hypercube intensity values. In addition, one can observe that even thought all the compared filters did not displace edges, AML almost everywhere stayed closer to the initial hypercube. AML simplified the image in the spatial directions by enlarging or creating new flat zones (leveled regions with constant intensity values), retaining all its 2D scale space properties. In the spectral direction it accounted for large intensity variations (spike-like features) and at the same time stayed close to the initial hypercube values. The above observations can be further confirmed by the performed quantitative evaluation (Table 1). In all cases, the AML resulted to the lower RMSE and NMSE values and to the larger structural similarity with the original image (SSIM). For the quantitative evaluation the standard RMSE and NMSE measures -which give a quantitative sense for the extent of variation between the intensity values of the compared images- were employed along with the recently proposed complementary quality measure of SSIM [13].

By employing such a scale space filtering and by producing elegantly simplified hypercubes, estimating the intrinsic dimensionality of high-dimensional datasets can be more efficient.

3. ESTIMATING THE INTRINSIC DIMENSIONALITY

The intrinsic dimensionality of data is the minimum number of parameters that is required to account for the observed properties of the data. Knowledge about the intrinsic dimensionality can be used to set input parameters at the dimension reduction algorithms. In addition, nearest neighbor searching algorithms can also profit from a good dimension estimate. Several techniques have been proposed in order to estimate the intrinsic dimensionality d of a dataset X ([14–17] and the references therein). There are estimators that are based on the analysis of local properties of the data like the correlation dimension estimator, the nearest neighbor dimension estimator and the maximum likelihood estimator and estimators that are based on the analysis of global properties like the eigenvalue-based estimator, the packing number estimator and the geodesic minimum spanning tree estimator. In addition, the estimation is computed either by projection or geometric procedures [16].

Projection procedures explicitly construct a mapping, and usually measure the dimension by using some variants of principal component analysis. Another general scheme in the family of projection techniques is to turn the dimensionality reduction algorithm from an embedding technique into a probabilistic, generative model, and optimize the dimension as any other parameter by using cross-validation in a maximum likelihood setting. On the other hand, methods based on the geometric properties do not require any explicit assumption on the underlying data model or the input parameters. Most of the geometric methods use the correlation dimension from the family of fractal dimensions due to the computational simplicity of its estimation.

During our experiments two estimators were employed. The Maximum Likelihood Estimator (MLE), which is a local approach and the global Geodesic Minimum Spanning Tree (GMST). Similar to the correlation dimension and the nearest neighbor dimension estimator, the MLE [17] estimates the number of datapoints covered by a hypersphere with a growing radius r. However, it does so by modelling the number of datapoints inside the hypersphere as a Poisson process. The GMST estimator is based on the observation that the length function of a geodesic minimum spanning tree is strongly dependent on the intrinsic dimensionality d. The GMST is the minimum spanning tree of the neighborhood graph defined on the dataset X. The length function of the GMST is the sum of the Euclidean distances corresponding to all edges in the geodesic minimum spanning tree. The intuition behind the dependency between the length function of the GMST and the intrinsic dimensionality d is similar to the intuition behind local intrinsic dimensionality estimators.

ADF and AML scale space filtering was applied to a number of hyperspectral datasets and the MLE and the GMST intrinsic dimensionality estimators followed. Datasets from i) the HySpex VNIR-1600 airborne sensor (©Norsk Elektro Optikk A/S) with 160 channels (400-1000nm), ii) the CASI-1500 airborne sensor (©ITRES) with 36 channels (380-1050nm), iii) the HYDICE sensor with 210 bands over the range 0.4-2.5 microns and iv) the EOS-1 Hyperion (©USGS) spaceborne sensor with 220 channels were available. Throughout the evaluation procedure the compared ADF was the same with the one that was used for the construction of the AML and each scale n was derived after three iterations t.

Table 2. Intrinsic Dimensionality Estimation

	MLE	GMST	Mean
Initial HC	10.6	8.3	9
Noisy HC (speckle 2%)	42.3	24.3	33
ADF at Noisy HC	21.9	13.5	18
AML at Noisy HC	33.4	13.5	23
ADF at Initial HC	8.5	7.5	8
AML at Initial HC	9.6	6.2	8

As it is demonstrated in Table 2, the estimated intrinsic dimension of HYDICE's dataset with the MLE method is 11 and 8 with the GMST. When the same initial hypercube (HC) is contaminated with a speckle noise then the estimated mean dimensionality significantly increases at 33. By applying ADF and AML scale space filtering a decrease of about 30% and 50% is achieved. Smoothing the noisy hypercube with the ADF method resulted into an estimation of about 18 (the mean value). Simplifying with the AML approach a decrease of about 30% has been also achieved. In both cases the same filtering scale scale (n=5) was used. Furthermore, when applying ADF and AML at the initial dataset with a smaller scale (n=2), then both methods agree that dataset's intrinsic dimensionality is 8 (mean value).

4. REDUCING THE DIMENSIONALITY

Having estimated the intrinsic dimensionality of the available datasets their transformation into a meaningful representation of reduced dimensionality follows. Ideally, the reduced representation has a dimensionality that corresponds to its intrinsic one which is the minimum number of parameters needed to account for the properties our observations [5, 18]. Lets denote with D the dimensionality of a dataset X described by a matrix $mxnxD$. X which consists of mxn datavectors, has an intrinsic dimensionality d with $d < D$ and often $d << D$. In geometric terms, intrinsic dimensionality means that the points in dataset X are lying on or near a manifold with dimensionality d that is embedded in the D-dimensional space. A dimensionality reduction technique transforms dataset X into a new dataset Y with dimensionality d, while retaining the geometry of the data as much as possible.

Apart from the traditional methods of Principal Component Analysis (PCA) and Multidimensional Scaling, other linear and nonlinear have been proposed like the ISOMAP, the Kernel PCA (KPCA), the Diffusion Maps, the Multilayer Autoencoders (MA), the Locally Linear Embedding, the Laplacian Eigenmaps, the Hessian Eigenmaps, the Local Space Tangent Analysis and the Semidefinite Embedding [4, 5]. Most of these techniques are based on the intuition that data lies on or near a complex low-dimensional manifold that is embedded in the high-dimensional space. Ideally, the target dimensionality is set equal to the intrinsic dimensionality of the dataset.

For our experiments we employed the PCA, the ISOMAP, the KPCA and the MA techniques. Briefly, the PCA constructs a low-dimensional representation that describes as much of the variance in the data as possible. This is done by finding a linear basis of reduced dimensionality for the data, in which the amount of variance in the data is maximal. The ISOMAP [19] resolves the problem by attempting to preserve pairwise geodesic (or curvilinear) distances between the datapoints of X. KPCA is the reformulation of traditional linear PCA in a high-dimensional space that is constructed using a kernel function [20]. Finally, MA are feed-forward neural networks with an odd number of hidden layers [21]. The target dimensionality of the low-dimensional data representation was specified by the MLE and GMST estimators.

The quantitative measures presented in Table 3, demonstrate that the use of ADF and AML scale space filtering ameliorated the dimensionality reduction procedure. The initial HC had a standard deviation (SD) of about 0,021 and a variance of about 0,0002, while the noisy HC (speckle 2%) had a SD of about 0,414 and a variance of about 0,2332. Having already (from the previous section) estimated the intrinsic dimensionality of the simplified noisy HC (Table 2: ADF and AML to noisy HC cases) the dimensionality reduction methods were applied. Table 3 presents the resulted SD and variance values. It is clear that after the application of ADF and AML an important improvement on the statistical measures was achieved. For example, the Noisy HC had an overall SD of about 0,023 which was increased at 0,269 when it was simplified by the AML and reduced to a lower representation by the PCA method. Same conclusions are derived by observations made to the other experimental cases when a noisy datasets was simplified with the ADF and AML approaches.

Furthermore, when comparing the variance of the reduced initial hypercube (IHC) with the one from the reduced simplified hypercube (ADF at the IHC and AML at the IHC) an increase has been also taken place. The dimensionality reduction has been facilitated by the vectorial scale space filtering, The elegantly simplified hypercubes encode in a more compact way world's spectral and spatial characteristics, forming, thus, more appropriate versions/represantations for estimating complex lower dimensional manifolds.

Table 3. Quantitative Evaluation for the Noisy Hypercube (NHC) and the Initial Hypercube (IHC)

Before the Dimensionality Reduction			
	NHC	ADF at the NHC	AML at the NHC
SD	0.023	0.021	0.023
Variance	3.1E-05	1.9E-05	2.6E-05

After the Dimensionality Reduction			
	IHC	ADF at the IHC	AML at the IHC
SD	0.36	0.41	0.42
Variance	0.203	0.220	0.245

5. CONCLUSIONS

We have introduced a framework for an efficient intrinsic dimensionality estimation and dimensionality reduction. By choosing appropriate nonlinear scale space representations, noise was removed and data were elegantly simplified. The simplified hypercubes described in a more distinct way the spectral and spatial signatures of our observations, forming appropriate versions for adequately estimating the intrinsic dimensionality of our initial/raw data. In addition, the computation of complex lower dimensional manifolds was more efficient, since the variance of the embedded hyperspace was more compact.

Experimental results on real data, along with the performed qualitatively and quantitative assessment, demonstrated that scale space filtering, by denoising and simplifying effectively the initial dataset, ameliorated the way the properties of our observations were encoded, strengthening, thus, the assumptions on the possible statistical separation of real world classes.

6. REFERENCES

[1] G.G. Wilkinson, “Are remotely sensed image classification techniques improving? Results of a long term trend analysis,” in *IEEE Workshop on Advances in Techniques for Analysis of Remotely Sensed Data*, Greenbelt, Maryland, 2003, pp. 27–28.

[2] A. Plaza, J. A. Benediktsson, J. Boardman, J. Brazile, L. Bruzzone, G. Camps-valls, J. Chanussot, M. Fauvel, P. Gamba, A. Gualtieri, M. Marconcini, J. Tilton, and G. Trianni, “Recent advances in techniques for hyperspectral image processing,” *Remote Sensing of Environment,(to appear)*, 2009.

[3] C. Vaiphasa, “Consideration of smoothing techniques

for hyperspectral remote sensing," *International Journal of Photogrammetry and Remote Sensing*, vol. 60, pp. 91–99, 2006.

[4] S. Theodoridis and K. Koutroumbas, *Pattern Recognition*, 4th Edition, Elsevier Academic Press, 2003.

[5] L.J.P. van der Maaten, E.O. Postma, and H.J. van den Herik, "Dimensionality reduction: A comparative review.," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2007 (submitted).

[6] H. Brunzell and J. Eriksson, "Feature reduction for classification of multidimensional data," *Pattern Recognition*, vol. 33, pp. 1741–1748, 2000.

[7] Andrew Webb, *Statistical Pattern Recognition*, John Wiley and Sons Ltd., UK, 2002.

[8] J. Martin-Herrero, "Anisotropic diffusion in the hypercube," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, pp. 1386–1398, 2007.

[9] K. Karantzalos, "A 4D morphological scale space representation for hyperspectral imagery," *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XXXVII, pp. 127–132, 2008, ISSN 1682-1750.

[10] K. Karantzalos, D. Argialas, and N. Paragios, "Comparing morphological levelings constrained by different markers," in *ISMM, G.Banon, et al. (eds), Mathematical Morphology and its Applications to Signal and Image Processing*, 2007, pp. 113–124.

[11] F. Meyer and P. Maragos, "Nonlinear scale-space representation with morphological levelings," *Journal of Visual Communication and Image Representation*, vol. 11, pp. 245–265, 2000.

[12] David Tschumperle and Rachid Deriche, "Vector-valued image regularization with pdes: A common framework for different applications," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 4, pp. 506–517, 2005.

[13] Z. Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, pp. 600–612, 2004.

[14] C. M. Bishop, M. Svensen, and C.K.I. Williams, "Gtm: The generative topographic mapping," *Neural Computation*, vol. 10, no. 1, pp. 215–235, 1998.

[15] E. Chavez, G. Navarro, R. Baeza-Yates, and J. Marroqun, "Searching in metric spaces," *ACM Computing Surveys*, vol. 33, no. 3, pp. 273–321, 2001.

[16] B. Kegl, "Intrinsic dimension estimation using packing numbers," in *Advances in Neural Information Processing Systems*, MIT Press, Cambridge, MA, 2003, vol. 15, pp. 681–688.

[17] E. Levina and P.J. Bickel, "Maximum likelihood estimation of intrinsic dimension," in *Advances in Neural Information Processing Systems*, MIT Press, Cambridge, MA, 2004, vol. 17.

[18] K. Fukunaga, *Introduction to Statistical Pattern Recognition*, Academic Press Professional, Inc., San Diego, CA, USA., 1990.

[19] J.B. Tenenbaum, "Mapping a manifold of perceptual observations," in *Advances in Neural Information Processing Systems*, MIT Press, Cambridge, MA, 1998, vol. 15, pp. 682–688.

[20] B. Scholkopf, A.J. Smola, and K.R. Muller, "Nonlinear component analysis as a kernel eigenvalue problem," *Neural Computation*, vol. 10, no. 5, pp. 1299–1319, 1998.

[21] G.E. Hinton and R.R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, 2006.