SUPERVISED MULTISPECTRAL IMAGE SEGMENTATION WITH POWER WATERSHEDS

Johannes Jordan, Elli Angelopoulou

Pattern Recognition Lab, University of Erlangen-Nuremberg, Germany
johannes.jordan@cs.fau.de, elli.angelopoulou@cs.fau.de

ABSTRACT

In recent years, graph-based methods have had a significant impact on image segmentation. They are especially noteworthy for supervised segmentation, where the user provides task-specific foreground and background seeds. We adapt the power watershed framework to multispectral and hyperspectral image data and incorporate similarity measures from the field of spectral matching. We also propose a new data-driven graph edge weighting. Our weights are computed by the topological information of a self-organizing map. We show that graph weights based on a simple L_p-norm, as used in other modalities, do not give satisfactory segmentation results for multispectral data, while similarity measures that were specifically designed for this domain perform better. Our new approach is competitive and has an advantage in some of the tested scenarios.

Index Terms— Multispectral imaging, Image Segmentation, Distance measurement, Distance Learning, Self organizing feature maps

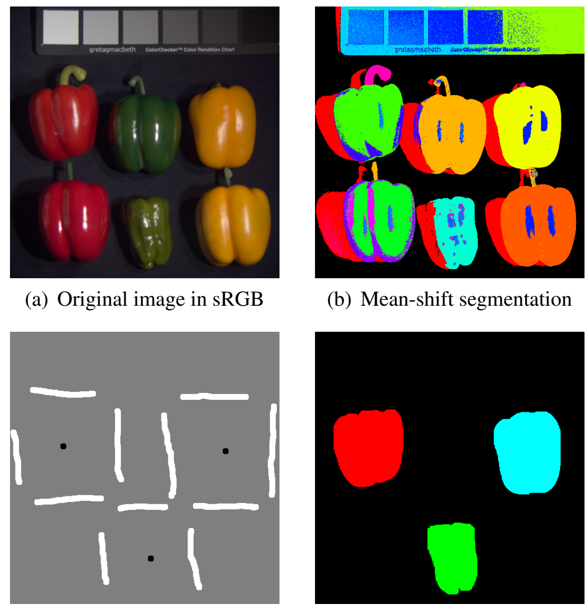
1. INTRODUCTION

In an interactive analysis framework, input from the user is an important prior to the segmentation tasks. As the user explores the multispectral data step-by-step, she may want to compare the spectra of specific objects in the scene or examine reflectance properties of a certain area in detail. Such a segmentation replaces tedious manual labeling of this area. In Fig. 1 we show the results of an unsupervised segmentation on a test image using the popular mean shift algorithm [1] next to those of a manually seeded segmentation. We believe that with the ability to guide the algorithm, the user can obtain a segmentation that is much more helpful for a specific analysis task.

Many approaches for automated multispectral image segmentation exist. Often general clustering methods are applied, e.g. k-means, or mean shift. Others are based on mixture models and/or minimum spanning tree clustering, e.g. [2], or have a hierarchical design. These methods have in common that they strive for a global segmentation, often in combination with classification, being unsuitable for us.

2. GRAPH-CUT SEGMENTATION

In the recent years, several image segmentation algorithms based on graph theory were introduced. In 1997, Shi and Malik proposed the unsupervised normalized cuts algorithm [4]. In 2006, Leo Grady introduced random walks to image segmentation [5]. Both create a graph with a node for each pixel and edges between pixels in a local spatial proximity, whereas the edge weights are based on pixel dissimilarity. While the normalized cuts try to find a minimum cut of the graph that separates foreground and background, Grady's method computes the likely destinations of a random walker. A third paradigm to graph-based image segmentation are watersheds, where the intensity image is considered as a topographic relief [6].



(a) Original image in sRGB (b) Mean-shift segmentation (c) User-provided seed points (d) Power watershed segmentation

Fig. 1. Illustration of the difference between unsupervised clustering and seed-based segmentation on example multispectral image [3].

In 2009, Couprie et al. introduced the power watershed framework, which integrates the aforementioned paradigms into a common mathematical framework [7]. This framework becomes the basis for developing a set of state-of-the-art supervised segmentation methods for multispectral and hyperspectral data.

A graph consists of a pair $G = (V, E)$ with vertices $v \in V$ and edges $e \in E \subseteq V \times V$. An edge, e, spanning two vertices, v_i and v_j, is denoted by e_{ij}. Each pixel is associated with a node and the nodes are connected locally via a 4-connected lattice in our implementation. Each edge has a real value assigned to it, called the weight, w_{ij}, of the corresponding edge e_{ij}. How edge weights are determined is presented in detail in Section 3.

According to [7], in a two-class segmentation, we compute the probability x for any pixel to belong to the foreground or background class as follows:

$$x = \underset{x}{\operatorname{argmin}} \sum_{e_{ij} \in E} w_{ij}^p \, |x_i - x_j|^q + \\ \sum_{v_i} w_{Fi}^p |x_i|^q + \sum_{v_i} w_{Bi}^p \, |x_i - 1|^q \ , \\ \text{s.t. } x(F) = 1, \ x(B) = 0 \ , \tag{1}$$

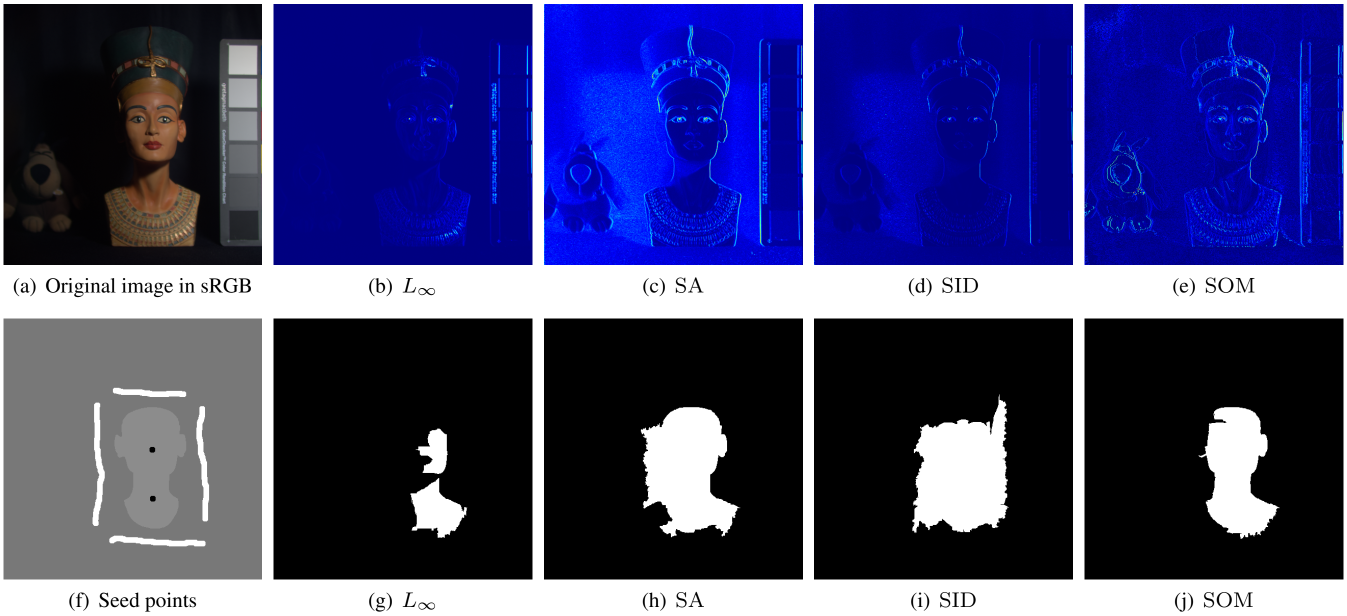


Fig. 2. Gradients according to different similarity measures in x-direction (top) and their segmentation results (bottom) on *statue* image.

where F and B are the foreground and background seeds, respectively. Then, a simple threshold leads to the binary segmentation s with $s_i = 1$ (foreground pixel) if $x_i \geq \frac{1}{2}$, 0 (background pixel) else. Based on the selection of parameters p and q, this minimization matches the graph cuts, random walker, or shortest paths algorithms. With $p = \infty, q \geq 1$, the power watershed algorithm is obtained.

To adjust this framework to the domain of multispectral or hyperspectral data, we need to address the selection of edge weights.

3. EDGE WEIGHTING

So far the power watershed framework was applied on 2D intensity images as well as 3D medical data [7]. In the original formulation, edge weights were defined as

$$w_{ij} = \exp\left(-\beta(\nabla I)^2\right) , \qquad (2)$$

where ∇I is the normalized gradient of the image I.

For color images, the authors relied on the L_2 norm, or the L_∞ norm (also known as the Chebyshev distance) instead of ∇I. This gives reasonable results when computed in the 3-dimensional space formed by the R, G, B triplets. With multispectral or hyperspectral images, we obtain a high-dimensional space in a typical range of 7 to 200 dimensions. It is well-known that the L_p norm is an ill-suited metric for higher dimensions. In this paper we examine two solutions to this problem. First, we examine several similarity measures that were designed especially for the multispectral domain. Second, we propose a novel data-driven similarity measure.

3.1. Spectral Mapping measures

Comparison of spectral vectors is a key part in the task of spectral mapping, where an observed spectrum is to be mapped to known material spectra. Several similarity measures have been designed for this task (see [8, 9]). Since a common goal of supervised segmentation is to discriminate specific objects in a scene, a similarity measure that discerns materials is a fitting replacement of ∇I for edge weights.

- A popular measure of spectral similarity is the spectral angle (SA) that was defined for the spectral angle mapper [10]. It captures the angle between two spectra, disregarding pure intensity changes. For two spectra $\mathbf{x}$ and $\mathbf{y}$ it is defined as

$$\mathrm{SA}(\mathbf{x}, \mathbf{y}) = \cos^{-1}\left(\frac{\langle \mathbf{x}, \mathbf{y}\rangle}{\|\mathbf{x}\|_2 \cdot \|\mathbf{y}\|_2}\right) . \qquad (3)$$

- The spectral information divergence (SID) is another measure that is often applied and was shown to work best for a certain scenario [8]. It is based on the Kullback-Leibler information measure and models a spectral vector $\mathbf{x}$ of length N as a random variable $\mathbf{p}^{(\mathbf{x})}$ with $\mathbf{p}^{(\mathbf{x})}_{1\leq l\leq N} = \frac{x_l}{\sum_{k=1}^{N} x_k}$. Then

$$\mathrm{SID}(\mathbf{x}, \mathbf{y}) = \sum_{l=1}^{N} \mathbf{p}_l^{(\mathbf{x})} \log \frac{\mathbf{p}_l^{(\mathbf{x})}}{\mathbf{p}_l^{(\mathbf{y})}} + \sum_{l=1}^{N} \mathbf{p}_l^{(\mathbf{y})} \log \frac{\mathbf{p}_l^{(\mathbf{y})}}{\mathbf{p}_l^{(\mathbf{x})}} . \qquad (4)$$

- A combination of SA and SID was proposed by Du et al. in 2004 to enhance the spectral discriminatory probability [10]. It exists in two versions:

$$\mathrm{SIDSAM}_1(\mathbf{x}, \mathbf{y}) = \mathrm{SID}(\mathbf{x}, \mathbf{y}) \cdot \sin\left(\mathrm{SA}(\mathbf{x}, \mathbf{y})\right) , \qquad (5)$$

$$\mathrm{SIDSAM}_2(\mathbf{x}, \mathbf{y}) = \mathrm{SID}(\mathbf{x}, \mathbf{y}) \cdot \tan\left(\mathrm{SA}(\mathbf{x}, \mathbf{y})\right) . \qquad (6)$$

  As the difference in the results of SIDSAM_1 and SIDSAM_2 is negligible, we only report results for SIDSAM_1.

- Robila et al. proposed the Normalized Euclidean Distance [8],

$$\mathrm{NED}(\mathbf{x}, \mathbf{y}) = \sqrt{\sum_{l=1}^{N} \left(\frac{x_l}{\overline{x}} - \frac{y_l}{\overline{y}}\right)^2} , \qquad (7)$$

  that is invariant to scalar multiplication and stays in the range $[0, 1]$ regardless of the number of bands N.

3.2. Data-driven measure

We introduce a new data-driven spectral similarity measure based on the concept of topological learning. A self-organizing map (SOM) is an artificial neural network that converts the nonlinear statistical relationship between high-dimensional data into simpler geometric relationships. In other words, the SOM provides a topological representation of the spectral vector distribution of a multispectral image. In the example case of a 2-D SOM (also 3-dimensional SOMs were tested), it consists of a 4-connected lattice of so-called model vectors $\mathbf{m}_i$ with location $L(\mathbf{m}_i)$. Each spectral vector is assigned the corresponding model vector that is nearest neighbor in L_2. In 2011, we showed that by using the topological distance inside the SOM, we obtain valuable differential information between adjoining pixels [11]. The topological distance is given by the L_2 distance between two model vectors in the SOM lattice. We can also use this information directly as a similarity measure,

$$\mathrm{SOM}(\mathbf{x},\mathbf{y}) = \left\| L(\operatorname*{argmin}_i \|\mathbf{x},\mathbf{m}_i\|_2) - L(\operatorname*{argmin}_j \|\mathbf{y},\mathbf{m}_j\|_2) \right\|_2 .$$

As this measure is well-suited for edge detection, it is a natural extension to also use it as a similarity measure for graph-cut based segmentation methods. The particular strengths of a SOM-based similarity measure are its flexibility (it can be fine-tuned) and its generalizability, as it adapts itself to the present data.

Fig. 2 shows example gradient maps in the x-direction by computing the similarity of each pixel with the pixel to its right. This is half the data that is used for computing edge weights in a 4-connected lattice resp. to (2).

4. EVALUATION

We compare the different edge weightings L_∞, SA, SID, SIDSAM, NED and SOM on the CAVE multispectral image database [3]. This is a set of high-quality multispectral images that depict objects of different materials in a laboratory setting. Images have a spatial resolution of 512×512 pixels and cover the spectral range of $400\,nm - 700\,nm$ in 31 bands. No ground-truth segmentation is available for this data set. Thus we hand-labeled certain objects in nine images from the dataset that cover a good variety in difficulties. Labeling was performed within the *Gerbil* framework [12], using its powerful visualization capabilities. In total we have 32 segmentation tasks. We also provide seed point input for the specific tasks: We place foreground seeds in the form of a circle with a 5-pixel radius in the center of each object. Background seeds are placed as hand-drawn lines on the top, left, bottom and right of each object with a distance of 20 to 40 pixels to the object contour. See Fig. 1(c), Fig. 2(f), Fig. 3(b) for examples. The seeds mimic a typical usage scenario.

4.1. Benchmark

Four different algorithms are tested that are included in the framework of Couprie et al.. First is the graph cut based on a maximum spanning forest computation, `mfs`. Second is the power watershed algorithm with $q = 2$, `pw`, that includes a random walker. The other two algorithms are based on these, however they employ geodesic reconstruction, `msfg` and `pwg`, respectively. Parameters of these algorithms are kept to the implementation defaults of [7], except for the new edge weighting. For the SOM similarity, a 32×32 SOM is trained on 15% of the image pixels (drawn randomly). Training takes less than 5 seconds (included in reported running times).

Measure	Algo.	Prec.	Recall	F_1-score	Time
NED	`pwg`	0.908	0.971	**0.929** $\pm$ 0.073	6.0
SA	`pw`	0.896	0.978	**0.928** $\pm$ 0.067	6.8
SA	`msf`	0.892	0.978	**0.926** $\pm$ 0.065	0.7
SA	`pwg`	0.891	0.979	**0.923** $\pm$ 0.071	6.8
NED	`pw`	0.911	0.962	**0.919** $\pm$ 0.089	1.0
NED	`msf`	0.907	0.962	**0.917** $\pm$ 0.091	1.0
SOM	`pwg`	0.897	0.925	**0.898** $\pm$ 0.074	18.6
SOM	`pw`	0.897	0.925	**0.898** $\pm$ 0.074	18.6
SOM	`msf`	0.894	0.903	**0.886** $\pm$ 0.057	9.6
SID	`msf`	0.846	0.973	**0.867** $\pm$ 0.249	2.8
SIDSAM	`msf`	0.819	0.972	**0.849** $\pm$ 0.245	3.4
SID	`pwg`	0.888	0.872	**0.837** $\pm$ 0.210	8.5
SID	`pw`	0.889	0.872	**0.837** $\pm$ 0.210	8.6
SIDSAM	`pw`	0.944	0.764	**0.793** $\pm$ 0.259	10.9
SIDSAM	`pwg`	0.947	0.763	**0.791** $\pm$ 0.266	10.7
L_∞	`pw`	0.937	0.588	**0.659** $\pm$ 0.214	8.9
L_∞	`pwg`	0.923	0.587	**0.655** $\pm$ 0.217	8.9
L_∞	`msf`	0.927	0.538	**0.598** $\pm$ 0.291	0.9

Table 1. Average performance of measure/algorithm combinations.

From the difference between ground-truth and segmentation result we compute precision, p, (the probability that a machine-generated foreground pixel is a true foreground pixel), recall, r, (the probability that a true foreground pixel is detected) and F_1-score, $F_1 = 2\frac{p \cdot r}{p+r}$. These quantities are suitable for our task as they do not depend on image size, but only the number of machine-generated foreground pixels and true foreground pixels.

In Table 1 the average performance over all images is reported. Average running time (in seconds) includes training once per image for SOM. As the geodesic reconstruction had little to no effects even on the single results, `mfsg` is omitted. In Table 2, results per-image are presented for each similarity measure in its overall best performing algorithm combination. All experimental data including images, ground-truth, seed points and segmentation results can be accessed on the web at `http://www5.cs.fau.de/research/data/msseg/`.

4.2. Discussion

It can be observed that the power watershed algorithm performs well for our application. The use of an edge weighting specific to our domain is important. This is indicated by the bad performance of L_∞, which is a good choice for RGB images. The most successful similarity measures SA and NED perform similarly, both qualitatively and quantitatively. They share the property of scale invariance, but are prone to noise in dark image regions. This explains their considerably bad performance on the *statue* image shown in Fig. 2. Here, the foreground object is not well separated from the background due to being partly self-shadowed. SA (see Fig. 2(c)) and NED respond more strongly to noise in the background than to the object boundary in the shadowed region. This is a case where the SOM similarity measure draws an advantage from being trained on the specific image.

The SID measure and its combinations with SA, $\mathrm{SIDSAM}_{1,2}$, do not perform as well in our experiment. While results are often on a par with SA, they are not reliable, as revealed by the high standard error. The measure poses specific problems to power watersheds, which makes the less powerful graph cut algorithm perform better.

	balloons / 4	*statue* / 1	*food* / 4	*lemons* / 2	*peppers* / 3	*feathers* / 6	*flowers* / 5	*toys* / 2	*balls* / 5
L_∞	$0.97 \pm .04$	0.51	$0.72 \pm .24$	$0.46 \pm .40$	$0.50 \pm .35$	$0.82 \pm .09$	$0.50 \pm .29$	$0.48 \pm .04$	$\mathbf{0.97} \pm .02$
SA	$0.97 \pm .03$	0.84	$\mathbf{0.99} \pm .01$	$\mathbf{0.99} \pm .01$	$\mathbf{0.99} \pm .01$	$0.83 \pm .17$	$\mathbf{0.94} \pm .04$	$\mathbf{0.94} \pm .05$	$0.86 \pm .17$
SID	$0.96 \pm .03$	0.21	$0.98 \pm .01$	$\mathbf{0.99} \pm .00$	$\mathbf{0.99} \pm .01$	$\mathbf{0.89} \pm .06$	$0.90 \pm .13$	$0.92 \pm .02$	$0.96 \pm .05$
SIDSAM	$0.96 \pm .03$	0.21	$0.98 \pm .01$	$0.95 \pm .05$	$\mathbf{0.99} \pm .01$	$0.82 \pm .17$	$0.89 \pm .13$	$\mathbf{0.94} \pm .04$	$0.90 \pm .12$
NED	$0.97 \pm .03$	0.77	$0.98 \pm .01$	$\mathbf{0.99} \pm .00$	$\mathbf{0.99} \pm .01$	$\mathbf{0.89} \pm .07$	$0.87 \pm .15$	$\mathbf{0.94} \pm .05$	$0.96 \pm .03$
SOM	$\mathbf{0.98} \pm .01$	**0.91**	$0.94 \pm .05$	$0.98 \pm .00$	$0.95 \pm .05$	$0.83 \pm .17$	$0.89 \pm .07$	$0.77 \pm .03$	$0.83 \pm .13$

Table 2. Average F_1-scores per edge weighting method on each image (number of segmentation tasks next to image name).

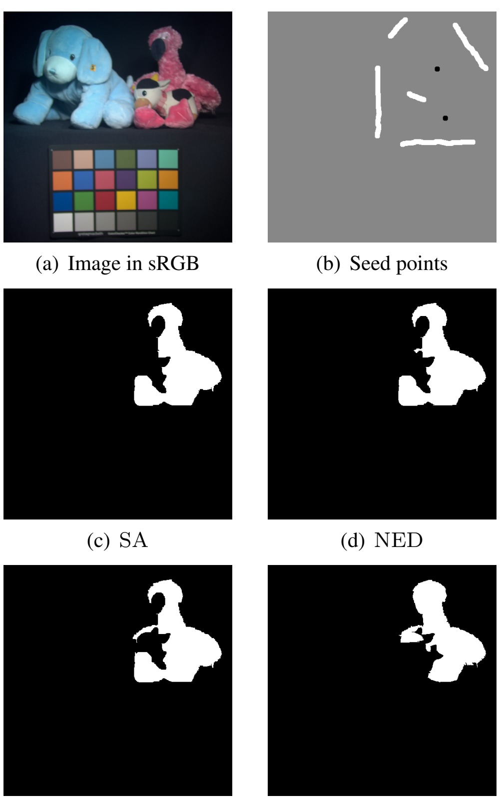
The SOM-based similarity did not perform best in this experiment. However, the results show that the SOM did indeed adapt well to the application, purely based on the data presented to it in training, while the other similarity measures tested where specifically designed for material discrimination based on spectrum. The SOM can be further adapted to a specific scenario by changing parameters of training and read-out, while the other measures are fixed. Typically, the right measure needs to be chosen based on application domain [9] while the SOM approach is general by design.

Fig. 3 depicts one of the more challenging tasks in the benchmark due to the object being partially occluded. Here SIDSAM performs best. SA and NED produce very similar results.

5. CONCLUSIONS

In this paper we introduce supervised segmentation to multispectral and hyperspectral images. We adapt the power watershed framework by incorporating similarity measures known from the field of spectral matching as well as a novel data-driven approach to edge weighting. Our results show that the straight-forward attempt of using the Chebyshev distance does not yield satisfactory results. Both proposed adaptations improve the segmentation performance significantly.

The implementation of this method was integrated into the *Gerbil* multispectral analysis framework [12] and will be released as free software at `http://gerbil.sf.net`. In future work, we deem it worthwhile to further examine data-driven measures for this application as well as the bad performance of the `pw`, SID combination.



(a) Image in sRGB
(b) Seed points
(c) SA
(d) NED
(e) SIDSAM
(f) SOM

Fig. 3. Example segmentation results on *toys* image, one of two tasks.

References

[1] B. Georgescu, I. Shimshoni, and P. Meer, "Mean shift based clustering in high dimensions: A texture classification example," in *ICCV*, 2003, pp. 456–463.

[2] S.K. Pal and P. Mitra, "Multispectral image segmentation using the rough-set-initialized EM algorithm," *Trans. on Geoscience and Remote Sensing*, vol. 40, no. 11, pp. 2495–2501, Nov 2002.

[3] F. Yasuma, T. Mitsunaga, D. Iso, and S. K. Nayar, "Generalized Assorted Pixel Camera: Post-Capture Control of Resolution, Dynamic Range and Spectrum," *IEEE Transactions on Image Processing*, vol. 99, Mar. 2010.

[4] J. Shi and J. Malik, "Normalized cuts and image segmentation," *TPAMI*, vol. 22, pp. 888–905, 1997.

[5] Leo Grady, "Random walks for image segmentation," *TPAMI*, vol. 28, pp. 1768–1783, 2006.

[6] J. Cousty, G. Bertrand, L. Najman, and M. Couprie, "Watershed cuts: Minimum spanning forests and the drop of water principle," *TPAMI*, vol. 31, no. 8, pp. 1362–1374, 2009.

[7] C. Couprie, L. Grady, L. Najman, and H. Talbot, "Power watershed: A unifying graph-based optimization framework," *TPAMI*, vol. 33, no. 7, pp. 1384–1399, 2011.

[8] S.A. Robila and A. Gershman, "Spectral matching accuracy in processing hyperspectral data," in *International Symposium on Signals, Circuits and Systems*. IEEE, 2005, vol. 1, pp. 163–166.

[9] A. Gutiérrez-Rodríguez, M. Medina-Pérez, J. Martínez-Trinidad, J. Carrasco-Ochoa, and M. García-Borroto, "New dissimilarity measures for ultraviolet spectra identification," *Advances in Pattern Recognition*, pp. 220–229, 2010.

[10] Y. Du, C.I. Chang, H. Ren, C.C. Chang, J.O. Jensen, and F.M. D'Amico, "New hyperspectral discrimination measure for spectral characterization," *Optical Engineering*, vol. 43, no. 8, pp. 1777–1786, Aug. 2004.

[11] J. Jordan and E. Angelopoulou, "Edge detection in multispectral images using the n-dimensional self-organizing map," in *ICIP*, 2011, pp. 3181–3184.

[12] J. Jordan and E. Angelopoulou, "Gerbil - A Novel Software Framework for Visualization and Analysis in the Multispectral Domain," *VMV 2010: Vision, Modeling and Visualization*, pp. 259–266, 2010.