

# Quantitative Evaluation Metrics for Superpixel Segmentation

Dylan Stewart<sup>a</sup>, Alina Zare<sup>a</sup>, and J. Tory Cobb<sup>b</sup>

<sup>a</sup>Electrical and Computer Engineering, University of Florida, Gainsville, FL, USA

<sup>b</sup>Naval Surface Warfare Center, Panama City Division, Panama City, FL, USA

## ABSTRACT

Superpixel segmentation methods have been found to be increasingly valuable in image processing and analysis. Superpixel segmentation approaches have been used as a preprocessing step for a wide variety of image analysis tasks such as full scene segmentation, automated scene understanding, object detection and classification, and have been used to reduce computation time during these tasks. While many quantitative evaluation metrics have been developed in the literature to analyze traditional image segmentation and clustering results, these metrics have not been used or adapted to quantitatively evaluate superpixel segmentations. In this paper, multiple superpixel segmentation algorithms are applied to synthetic aperture sonar (SAS) imagery and the results are evaluated using cluster validity indices that have been adapted for superpixel segmentation. Both cluster validity metrics that rely only on internal measures as well as those that use both internal and external measures are considered. Results are shown on a synthetic aperture sonar (SAS) data set.

**Keywords:** Superpixels, synthetic aperture sonar, segmentation, cluster validity, environmental

## 1. INTRODUCTION

Superpixels have been a growing research topic since Ren and Malik developed a formal definition over a decade ago.<sup>1</sup> Although not specifically called superpixels, multiple algorithms were introduced in the late 1980's and early 1990's that segmented images into superpixel-like regions.<sup>2,3</sup> More recently, superpixels have been used as a method to separate images into common small groups of pixels to assist in complex computer vision problems.<sup>4-18</sup> Specific applications include problems such as computer aided surgery,<sup>19</sup> hyperspectral image classification,<sup>20</sup> or, in our case, segmenting SAS imagery<sup>21</sup> to name a few. Within this scope, many authors agree on a set of rules for superpixels. Specifically the need for connected sets of pixels belonging to a superpixel, every pixel containing a label for a specific superpixel, boundaries preserved by the superpixels, superpixels being compact, generated efficiently, and of controlled number.<sup>22</sup>

Qualitative methods of evaluating superpixels mainly focus on choosing visually appealing boundaries. However, qualitative methods do not allow for a mechanism to systematically compare superpixel algorithms and segmentations across large sets of imagery. Yet, assigning quantitative values to measure unsupervised superpixel methods is much of an open question. Quantitative image segmentation and superpixel evaluation metrics in the literature generally rely on having ground truth boundaries with which to evaluate against.<sup>22-28</sup> In our large data set of SAS imagery, there is little to no ground truth to use for superpixel segmentation evaluation. Thus, metrics that are variance based (and do not require ground truth boundaries), such as Intra-Cluster Variation<sup>26</sup> and Explained Variation,<sup>29</sup> are more applicable. However these metrics have thus far only been applied towards standard visual color image segmentation (and not SAS segmentation).<sup>22,25,27</sup> In this paper, we present quantitative superpixel evaluation metrics and investigate their use in SAS image superpixel segmentation.

Superpixel segmentation approaches have been extensively applied to color and gray-scale standard visual imagery. Within this study, we compare implementations of Normalized Cuts,<sup>21</sup> SLIC,<sup>30</sup> and TurboPixels (TP)<sup>24</sup> to segment complex SAS image scenes. We evaluate these algorithms and their resulting segmentations on a large set of SAS imagery using new quantitative metrics that have been adapted from traditional unsupervised clustering for use with superpixels.<sup>31,32</sup>

---

Further author information: (Send correspondence to A. Zare)

A. Zare: E-mail: azare@ece.ufl.edu

## 2. SUPERPIXEL ALGORITHMS

As shown by Cobb and Zare,<sup>21</sup> an adaptation of Normalized Cuts can be successful in segmenting SAS imagery, and the implementation produces visually appealing segmentations. In this paper, the standard Normalized Cuts segmentation approach (applied to the grayscale intensity values from the SAS imagery) is compared to the SLIC<sup>30</sup> and TurboPixels<sup>24</sup> superpixel segmentation algorithms using the proposed quantitative evaluation metrics.

Within our study of SAS imagery, existing superpixel methods (and the attendant features used by them for standard visual imagery) were found not optimal or applicable to raw SAS pixel values without preprocessing to address the noise and speckle commonly seen in SAS imagery. In order to make these superpixel algorithms successful with our SAS dataset, we first despeckle the imagery via median filtering, and then normalize the log of the imagery such that the grayscale values fall in the range (0,1).

## 3. QUANTITATIVE METRICS

Quantitative methods are needed to fully evaluate superpixel methods, especially within an unsupervised framework. In this work, we adapted the Dunn's and Davies-Bouldin cluster validity indices to measure superpixel validity.<sup>31,32</sup> Both of these indices incorporate internal and external cluster properties when evaluating a clustering result. The external measure used considers the distance between each pair of cluster centers and the internal measure considers the diameter of each superpixel.

With the traditional Dunn's and Davies-Bouldin indices, every cluster is considered to be distinct from all others. However, in the case of superpixel segmentation, it is expected that there may be disjoint regions in the image that contain similar pixel values (and, thus, the external measure between these disjoint regions would be small). Thus, we adapt the Dunn's and Davies-Bouldin indices to be more applicable to superpixel segmentation by computing the external measures only between neighboring superpixels (as opposed to between all pairs of clusters/superpixels). While the dissimilarity between superpixels and diameter can be measured with various different distances metrics, we use Euclidean distance for our calculations in this paper. These adapted approaches, Dunn-SP and DB-SP, are outlined in Algorithms 1 and 2, respectively.

---

**Algorithm 1** Dunn-SP Index

---

**Input:** SAS Image and Superpixel Labels

**Output:** Dunn Index for Each Superpixel

- 1: **for** every superpixel  $k$  **do**
  - 2:   Calculate dissimilarity between two neighboring superpixels  $C_i$  and  $C_j$  as  $d(C_i, C_j) = \min_{x \in C_i, y \in C_j} d(x, y)$  using distance between centers of each neighboring superpixel
  - 3:   Calculate diameter of a superpixel  $\text{diam}(k) = \max_{x, y \in k} d(x, y)$
  - 4:   Assign each pixel the Dunn-SP index of its superpixel  $D_k = \min_{j \in N_k} \frac{d(C_k, C_j)}{\max_{m \in N_k} \text{diam}(C_m)}$ , where  $N_k$  are the neighboring superpixels of superpixel  $k$
  - 5: **end for**
- 

**Algorithm 2** DB-SP Index

---

**Input:** SAS Image and Superpixel Labels

**Output:** Davies-Bouldin Index for Each Superpixel

- 1: **for** every superpixel  $k$  **do**
  - 2:   Calculate dissimilarity between two neighboring superpixels  $C_i$  and  $C_j$  as  $d(C_i, C_j) = \min_{x \in C_i, y \in C_j} d(x, y)$  using the difference between centers of each neighboring superpixel.
  - 3:   Calculate diameter of a superpixel  $\text{diam}(k) = \max_{x, y \in k} d(x, y)$
  - 4:   Assign each pixel the DB-SP index of it's superpixel  $DB_k = \max_{j \in N_k} \frac{\text{diam}_k + \text{diam}_j}{\min_{d(C_k, C_j)}},$  where  $N_k$  are the neighboring superpixels of superpixel  $k$
  - 5: **end for**
-

We also consider the variance of a superpixel's intensity values as an internal-only metric,  $V$ , where for superpixel  $k$  with  $N$  pixels where  $\mu_k$  is the mean of superpixel  $k$ :

$$V_m = \sum_{i=1, i \in k}^N \frac{(x_i - \mu_k)^2}{N} \quad (1)$$

While the Dunn's and Davies-Bouldin indices are combinations of internal and external measures, we found it necessary to also evaluate superpixels with a solely internal measure because superpixels are generally an oversegmentation of an image scene. Thus, it is likely that neighboring superpixels contain similar texture and pixel information. Intuitively, lower variance within a superpixel is considered desireable as this indicates that all pixels within the superpixel are similar to each other in intensity.

#### 4. EXPERIMENT AND RESULTS

All superpixel algorithms were applied to a large set of diverse SAS images containing various seabed types and textures and evaluated using the proposed quantitative metrics.

In the first experiment, each superpixel segmentation algorithm was applied to a set of 10 SAS images repeatedly to estimate from 100 to 600 superpixels per image. Each result was evaluated using all three proposed superpixel segmentation metrics. All superpixel metrics were scaled by dividing the maximum respective index for each set to allow for comparison across metrics and highlight similarities and novelties that may be seen across indices. Fig. 1 displays the results from this experiment (as well as running time for each of the superpixel segmentation approaches with increasing number of superpixels).

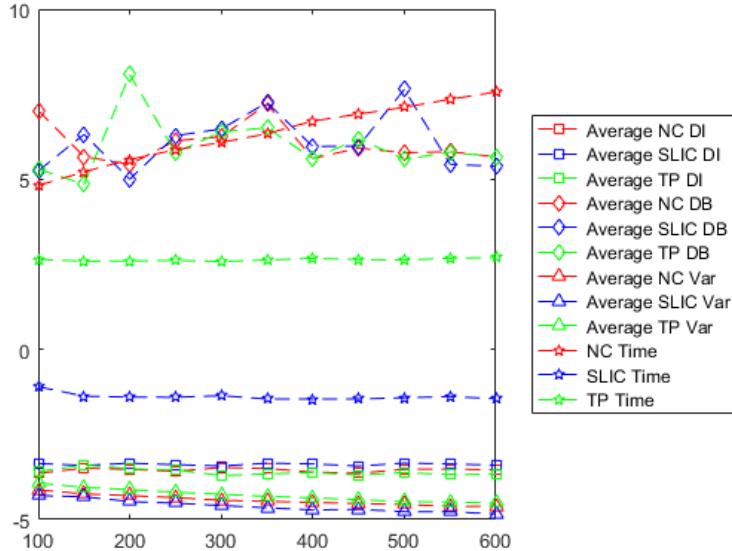


Figure 1: Log of Average (across 10 SAS images) Superpixel Index vs. Number of Superpixels. Time results are also shown as log of the result. Higher Dunn's, lower Davies-Bouldin, and lower variance metrics indicate good segmentation results.

Within Figure (1) we show the average of each superpixel indice for our subset of 10 images across different numbers of superpixels. This experiment helps to illustrate the characteristics of each evaluation metric. By examining the plot, it can be seen that the variance measure decreases (and, thus, improves) with the number of superpixels. This is to be expected as the size of each superpixel decreases, all segmentation methods are more capable of grouping only similarly valued pixels together. The Dunn’s index is relatively flat across the number of superpixels. As shown in Alg. 2, the Dunn’s index is the ratio between the minimum distance between neighboring superpixel centers over the maximum diameter of all superpixels in the neighborhood. The relatively flat Dunn’s index results from the fact that at a low number of superpixels both the numerator and denominator of the Dunn’s index is likely to be large, both the distance between neighboring superpixels are large as well as the deviation within a superpixel. In contrast, at a larger number of superpixels both the numerator and denominator will be small as the variance within a superpixel is likely to decrease and neighboring superpixels will be more similar as the segmentation becomes oversegmented. The Davies-Bouldin indices show the most variance across different numbers of superpixels, however, the over the entire plot trend is generally flat.

In the following results, each algorithm was initialized to have 300 superpixels. Table 1 shows the average and standard deviation found for each metric using each superpixel segmentation algorithm. The quantitative value of all three metrics are quite similar when considering the average and standard deviation of each. The Davies-Bouldin indices are similar across the methods with SLIC having a slightly lower average index value. This can be attributed to SLIC producing slightly more superpixels than the other methods, causing decreases in the diameters of superpixels in the neighborhood. Similarly, Dunn’s index for SLIC is slightly higher for the same reason.

Table 1: Quantitative superpixel results for 10 experiments using 300 superpixels

Superpixel Metric	Normalized Cuts	SLIC	TurboPixels
Dunn’s Index	$0.03 \pm 0.04$	$0.04 \pm 0.06$	$0.03 \pm 0.05$
DB Index	$0.36 \pm 0.10$	$0.34 \pm 0.10$	$0.37 \pm 0.10$
Average Variance	$0.01 \pm 0.00$	$0.01 \pm 0.00$	$0.01 \pm 0.00$

In Table 2, running time between the methods is compared. Normalized Cuts, in general, takes longer due to large eigenvalue problem that must be solved in its implementation. SLIC is generally the fastest in running time.

Table 2: Quantitative superpixel results for 10 experiments using 300 superpixels

Superpixel Metric	Normalized Cuts	SLIC	TurboPixels
Running Time [seconds]	$469.0 \pm 26.5$	<b><math>0.3 \pm 0.1</math></b>	$13.0 \pm 0.4$

In Table 3, we consider the worst results of a superpixel for each metric. This highlights results in which a superpixel is poorly formed. As can be seen, according to two of the metrics, the worst superpixels formed by SLIC outperform the worst superpixels formed by the other two superpixel segmentation approaches according to Dunn’s and the DB indices whereas TurboPixels does the best in terms of the variance metric.

Table 3: Worst superpixel results for 10 experiments using 300 superpixels

<b>Superpixel Metric</b>	<b>Normalized Cuts</b>	<b>SLIC</b>	<b>TurboPixels</b>
Dunn's Index	$1.7 \times 10^{-6}$	$1.3 \times 10^{-5}$	$8.9 \times 10^{-7}$
DB Index	0.46	<b>0.10</b>	1
Average Variance	1	0.80	<b>0.74</b>

Fig. (2) and Fig. (3) are two example results on a full SAS image. Both figures show the original image, the segmentation results from each algorithm, and the resulting index value for each pixel across all three methods. Large values for Dunn's index and small values for the Davies-Bouldin index highlight superpixels that have low internal variance and are distinct from their neighbors. Superpixels with high variance indicate that they should be split into separate superpixels.

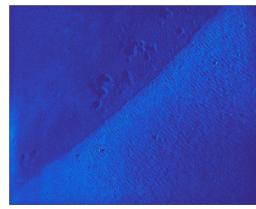
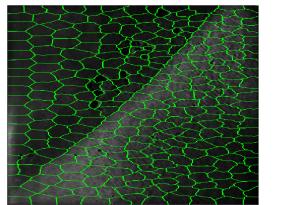


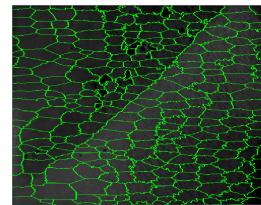
Figure 2: Image 1



(a) NCuts Segmentation



(b) SLIC Segmentation



(c) TP Segmentation



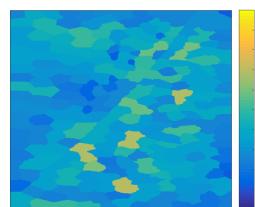
(d) NCuts Dunn's Indices



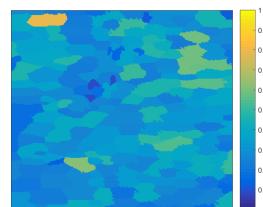
(e) SLIC Dunn's Indices



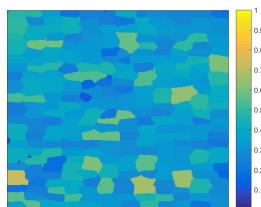
(f) TP Dunn's Indices



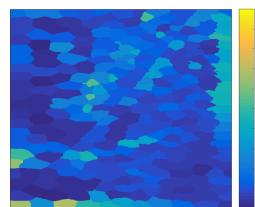
(g) NCuts DB's Indices



(h) SLIC DB's Indices



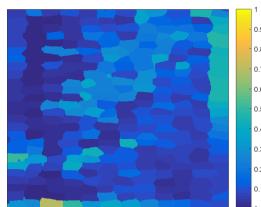
(i) TP DB's Indices



(j) NCuts Variance



(k) SLIC Variance



(l) TP Variance

Fig. 2 - Image 1 Normalized Cuts, SLIC, and TurboPixels Segmentation and Index Results

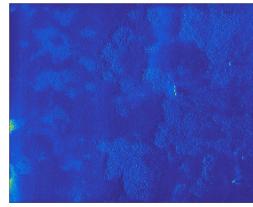
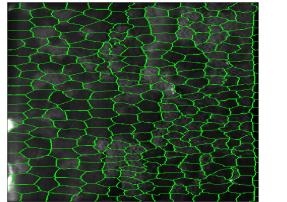
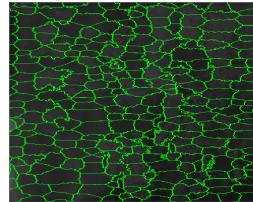


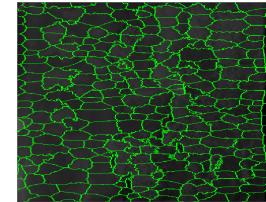
Figure 3: Image 2



(a) NCuts Segmentation



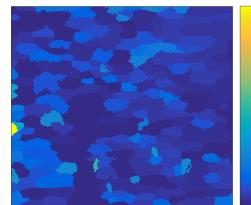
(b) SLIC Segmentation



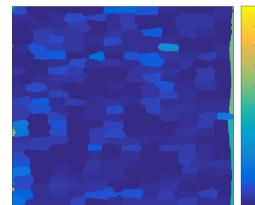
(c) TP Segmentation



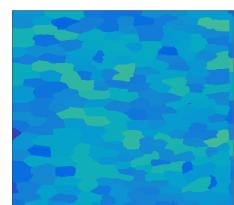
(d) NCuts Dunn's Indices



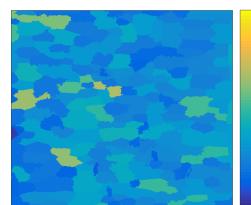
(e) SLIC Dunn's Indices



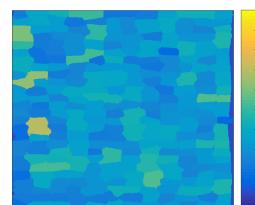
(f) TP Dunn's Indices



(g) NCuts DB's Indices



(h) SLIC DB's Indices



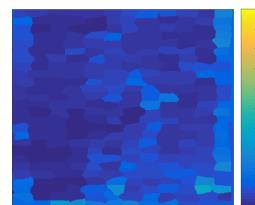
(i) TP DB's Indices



(j) NCuts Variance



(k) SLIC Variance



(l) TP Variance

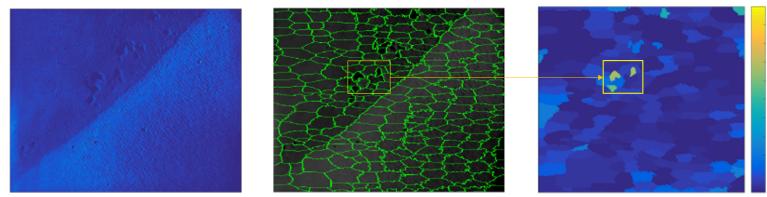
Fig. 3 - Image 2 Normalized Cuts, SLIC, and TurboPixels Segmentation and Index Results

In order to further understand the reasoning behind the mechanics of each index we consider the histogram of values for good cases and bad cases of each measure in Fig. 4-6. In the bounding box of Fig. 4(a), one

superpixel is shown to have a high Dunn's index value and one of its neighbors have a low Dunn's index values. To understand why, consider the histogram of intensity values for both superpixels. The "good" (high Dunn's index) superpixel has a very compact feature histogram whereas the "bad" (low Dunn's index) superpixel has a large spread and is not considered by the index to be a compact. We see a similar behavior when considering the Davies-Bouldin index in detail in Fig. 5. In Fig. 6, the variance metric is considered and superpixels with small spread have a low (desireable) variance.

Figure 4: Dunn's Index Explanation

(a) Dunn's Index Example



(b) Histogram of a good superpixel within the bounding box (high Dunn's index) and a bad neighbor (low Dunn's Index)

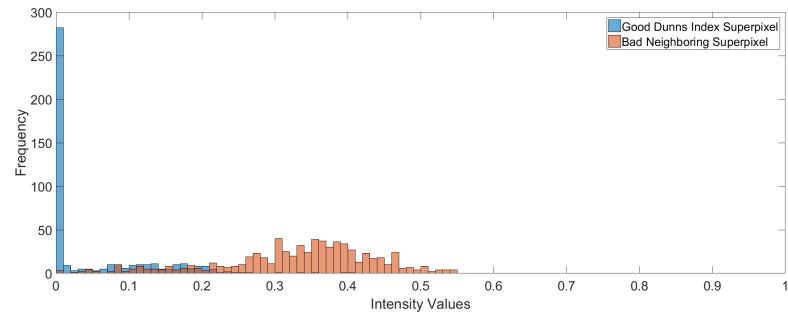
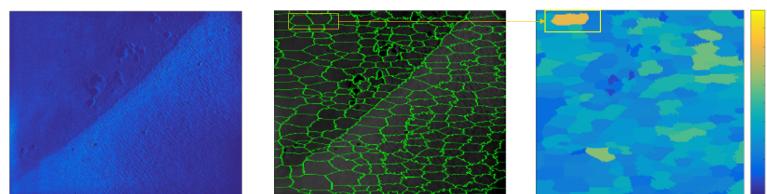


Figure 5: Davies-Bouldin Explanation

(a) Davies-Bouldin Index Example



(b) Histogram of a good superpixel within the bounding box (low DB index) and bad neighbor (high DB index)

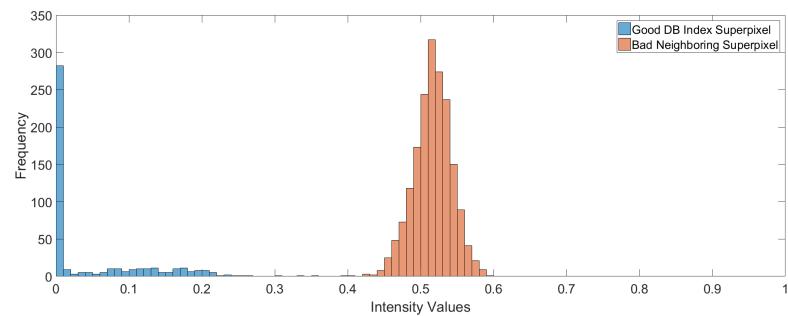
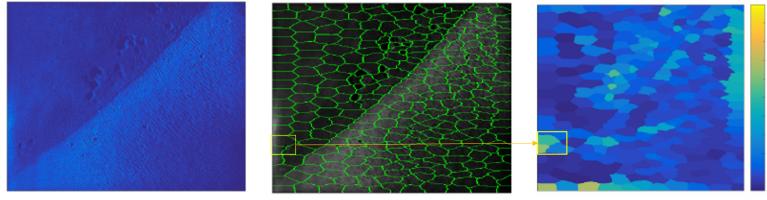
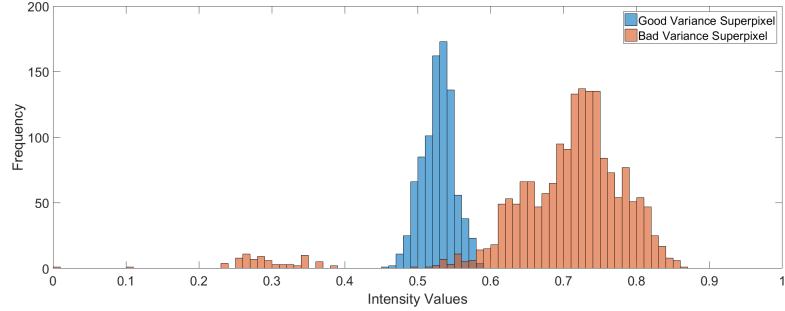


Figure 6: Variance Explanation

(a) Variance Index



(b) Histogram of a good superpixel within the bounding box (low variance) and bad superpixel (high variance)



## 5. SUMMARY

In this paper, we proposed three quantitative metrics to evaluate unsupervised superpixel segmentation for application to SAS imagery. For superpixels whose edges which fall along distinct boundaries, the Dunn’s and Davies-Bouldin indices are found to be the most informative. For an oversegmentation of the scene, variance is the most informative measure.

## ACKNOWLEDGMENTS

The authors graciously thank the Office of Naval Research, Code 321, for funding this research. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the Office of Naval Research.

## REFERENCES

- [1] X. Ren and J. Malik, “Learning a classification model for segmentation,” in *9th IEEE International Conf. on Comp. Vision*, **1**, pp. 10–17, Oct. 2003.
- [2] R. Mester and U. Franke, “Statistical model based image segmentation using region growing, contour relaxation and classification,” *SPIE Symposium on Visual Communications and Image Processing* , pp. 616–624, 1988.
- [3] B. Marcotegui and F. Meyer, “Bottom-up segmentation of image sequences for coding,” *Annals of Telecommunications* **52**, pp. 397–407, July 1997.
- [4] S. W. et al, “Superpixel tracking,” *International Conference on Computer Vision* , pp. 1323–1330, 2011.
- [5] L. Y. et al, “Robust superpixel tracking,” *IEEE Transactions on Image Processing* **23**, pp. 1639–1651, Apr 2014.
- [6] K. Y. et al, “Efficient joint segmentation, occlusion labeling, stereo and flow estimation,” *European Conference on Computer Vision* , pp. 756–771, 2014.
- [7] S. H. et al, “Supercnn: A superpixelwise convolutional neural network for salient object detection,” *International Journal of Computer Vision* **115**, pp. 330–344, Dec 2015.

- [8] F. P. et al, "Saliency filters: contrast based filtering for salient region detection," *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 733–740, 2012.
- [9] G. S. et al, "Improving an object detector and extracting regions using superpixels," *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3721–3727, 2013.
- [10] P. R. et al, "Generating object segmentation proposals using global and local search," *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2417–2424, 2014.
- [11] van den Bergh et al, "Online video seeds for temporal window objectness," *International Conference on Computer Vision*, pp. 377–384, 2013.
- [12] M. L. et al, "Discrete-continuous depth estimation from a single image," *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 716–723, 2014.
- [13] S. G. et al, "Multi-class segmentation with relative location prior," *International Journal of Computer Vision* **80**, pp. 300–316, Dec 2008.
- [14] C. Lerma and J. Kosecka, "Semantic segmentation with heterogeneous sensor coverages," *IEEE International Conference on Robotics and Automation*, pp. 2639–2645, 2014.
- [15] A. Geiger and C. Wang, "Semantic segmentation with heterogeneous sensor coverages," *German Conference on Pattern Recognition*, pp. 183–195, 2015.
- [16] S. G. et al, "Indoor scene understanding with rgbd images: bottom-up segmentation, object detection and semantic segmentation," *International Journal of Computer Vision* **112**, pp. 133–149, 2015.
- [17] D. L. et al, "Holistic scene understanding for 3d object detection with rgbd cameras," *International Conference on Computer Vision*, pp. 1417–1424, 2013.
- [18] J. L. et al, "Patch match filter: efficient edge-aware filtering meets randomized search for fast correspondence field estimation," *International Conference on Computer Vision and Pattern Recognition*, pp. 1854–1861, 2013.
- [19] P. B. et. al, "Eikonal based region growing for superpixels generation: application to semi-supervised real time organ segmentation in ct images," *IRBM* **35**, pp. 20–26, Feb 2014.
- [20] K.-S. K. et al, "Improved simple linear iterative clustering superpixels," *2013 IEEE International Symposium on Consumer Electronics (ISCE)*, pp. 259–260, 2013.
- [21] J. Cobb and A. Zare, "Boundary detection and superpixel formation in synthetic aperture sonar imagery," *Proceedings of the Institute of Acoustics*, pp. 211–221, 2014.
- [22] D. S. et al, "Superpixels: An evaluation of the state-of-the-art," *Computer Vision and Image Understanding*, 2016.
- [23] R. A. et al, "Slic superpixels compared to state-of-the-art superpixel methods," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **34**, pp. 2274–2281, 2012.
- [24] A. L. et al, "Turbopixels: Fast superpixels using geometric flows," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 2290–2297, 2009.
- [25] P. Neubert and P. Protzel, "Superpixel benchmark and comparison," *Forum Bildverarbeitung*, 2012.
- [26] W. Benesova and M. Kottman, "Fast superpixel segmentation using morphological processing," *Conference on Machine Vision and Machine Learning*, 2014.
- [27] A. S. et al, "Measuring and evaluating the compactness of superpixels," *International Conference on Pattern Recognition*, pp. 930–934, 2012.
- [28] D. M. et al, "Learning to detect natural image boundaries using local brightness, color, and texture cues," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **26**, pp. 530–549, 2004.
- [29] A. M. et al, "Superpixel lattices," *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–8, 2008.
- [30] R. A. et al, "Slic superpixels," *Technical Report, cole Polytechnique Fdrale de Lausanne*, 2010.
- [31] J. Bezdek and N. Pal, "Cluster validation with generalized dunn's indices," *Artificial Neural Networks and Expert Systems*, pp. 190–194, 1995.
- [32] D. Davies and D. Bouldin, "A cluster separation measure," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **PAMI-1**, pp. 224–227, 1979.