

Improving Superpixel Boundaries Using Information Beyond the Visual Spectrum

Keith Sullivan
Exelis, Inc.
McLean, VA USA

Wallace Lawson, Donald Sofge
Naval Research Laboratory
Washington, DC USA

Abstract

Superpixels enable a scene to be analyzed on a larger scale, by examining regions that have a high level of similarity. These regions can change depending on how similarity is measured. Color is a simple and effective measure, but it is adversely affected in environments where the boundary between objects and the surrounding environment are difficult to detect due to similar colors and/or shadows. We extend a common superpixel algorithm (SLIC) to include near-infrared intensity information and measured distance information to help oversegmentation in complex environments. We demonstrate the efficacy of our approach on two problems: object segmentation and scene segmentation.

1. Introduction

Superpixels transform an image from a relatively meaningless collection of pixels to small regions with similar characteristics. From the perspective of an autonomous mobile robot, superpixels have a wide range of possible applications including finding navigable paths [6, 13, 5] and localizing objects [4]. In order to accomplish tasks such as these, the superpixel boundaries should closely align with the contours of the objects in the scene. These contours can sometimes be quite difficult to find, particularly in outdoor environments which provide lots of places for an object to be obscured, either intentionally or unintentionally. Additionally, the object might be hidden by shadows, or partially buried. Outdoor environments can also provide a great deal of clutter, sometimes due to natural textures like long grass and leaves. Indoor environments present their own challenges with objects of similar color, harsh lighting (especially in scenes containing windows), and occlusions. Figure 1 shows typical indoor and outdoor environments that are difficult to oversegment. In Figure 1(a), the foreground is difficult to oversegment since it is washed out, while in Figure 1(b), difficulties arise from the matching color of the munitions and grass.



(a) Indoor



(b) Outdoor

Figure 1. Typical indoor and outdoor scenes that are difficult to accurately oversegment.

One promising way of finding superpixels is with an algorithm known as SLIC (*simple linear iterative clustering*) [1]. SLIC operates by clustering together pixels that are similarly colored over small regions of the image. By clustering in this manner, superpixels can be computed ef-



Figure 2. An example of how SLIC is sensitive to lighting conditions in unstructured environments. Note that the superpixels do not adhere to the bottom of the black box, nor to the leaves in the foreground.

ficiently and accurately. Prior work has shown that SLIC superpixels can be computed in real time via a GPU [14] (which makes SLIC attractive for autonomous robotics).

Like other superpixel algorithms, SLIC is sensitive to lighting conditions and clutter: detecting boundaries is challenging in shadows, both for inside and outdoor scenes. For example, looking closely at Figure 2, one notices that the superpixels do not adhere to all the boundaries of the leaves, nor the bottom boundary of the black case. Examples like these are typical for autonomous robots, which tend to be close to the ground, thus producing a low perspective. The low perspective ensures the bottoms of objects will be in deep shadow and/or partially buried in the environment. To facilitate detecting object boundaries we augment the SLIC algorithm to include information from beyond the visual spectrum.

The near infrared (NIR) spectrum is a likely candidate for gathering additional information for oversegmentation for four reasons. First, NIR images preserve object boundaries (e.g., the shape of objects) similar to RGB images. Second, NIR images show consistent pixel values across a single material and are not affected by color patterns. Thus, changes in intensity are not affected by minor color or lighting variations in the object. Third, in the near infrared spectrum different materials have different intensities due to the different absorption and reflectance properties of the material [11]. For example, chlorophyll [10] and metal strongly reflect near-IR light while concrete does not reflect nearly as strongly. Fourth, NIR sensors are becoming more prevalent, and are easily carried by small autonomous robots. Our approach ISLIC (Intensity-SLIC) combines information from the near-IR spectrum with RGB information to improve superpixel boundaries. In particular, we modify the distance

calculation within SLIC to combine NIR information with RGB information.

2. Related Work

The use of superpixels in both computer vision and robotics algorithms is quite common. One approach to finding superpixels is an approach known as normalized cuts [18]. Normalized cuts operates by building an affinity matrix, then segmenting along the boundaries where there is the least amount of affinity. While this produces a segmentation that is highly spatially coherent, it is an expensive and time consuming operation.

Levinshtein et al. [7] proposed turbo pixels as a more efficient way to produce plausible, uniformly sized superpixels. The algorithm operates by placing initial seeds uniformly around the image, then growing these regions over time. Although it is efficient, is difficult to extend the geometric constraints to consider additional information such as intensity and distance information.

The SLIC superpixel algorithm also addresses the speed of superpixel computation, but with a simple approach that produces good results and is easy to extend [1]. The basic approach is to cluster local pixels according to their color but SLIC does not enforce cluster compactness or uniform pixel size. It excels in terms of both speed and performance on testing sets.

Many have proposed to extend superpixels with depth information. This was used in the context of scene understanding by Ren et al. [15]. Weikersdorfer et al. [22] proposed depth adaptive superpixels, which computes superpixels in a manner similar to SLIC, but depth is used in addition to color.

There has been minimal work extending superpixel algorithms to non-visual data. A recent exception is Masoudifar et al. who extended the UCM superpixel algorithm to hyperspectral images [9]. They use principal component analysis to combine all the hyperspectral channels of a single pixel before performing feature extraction and superpixel construction. Thompson et al. [20, 21] used a graph-based approach to cluster pixels of similar values in hyperspectral images.

Turning to the use of NIR data in object identification and classification, several authors have previously explored the use of lidar (using the laser as a NIR illumination source) to classify different types of domains. Sullivan et al. classified different types of terrain using an RGB-D approach combining color image segmentation with classification from a Hokuyo [19]. Similarly, Wurm et al. developed a classifier for distinguishing between grass and concrete using laser intensity [23]. Kirchner et al. used laser intensity to classify materials in an industrial setting [8].

Using passive NIR images, Salamati et al. [11, 17, 16] have examined combined NIR and RGB images for material

classification and semantic segmentation by extracting various features (e.g., SIFT) and applying a classifier to identify either material or the semantic label. Our work differs by relying on the intrinsic nature of the oversegmentation algorithm rather than computing additional features from the images before classification. We feel that our approach is more applicable to the real-time constraints imposed by autonomous robotics.

3. Method

Our superpixel method is an extension of the SLIC superpixel algorithm developed Achanta et al. [1]. SLIC adapts a localized k -means clustering with a weighted distance metric. Clusters are initialized by sampling along a grid (cell size $S = \sqrt{N/k}$ where N is the number of pixels in the image) at roughly equally spaced intervals. During clustering, pixels that are within a $2S \times 2S$ region centered around each cluster center are examined, rather than comparing with all cluster centers as a traditional k -means clustering. Once each pixel is associated with a cluster center, the cluster centers are updated to the mean of all pixels belonging to that cluster. SLIC repeats these steps until either the residual error falls below a threshold, or for a fixed number of iterations.

SLIC depends on a special distance metric D that computes the distance between the cluster center and a given pixel. The distance metric combines color distance d_c and spatial proximity distance d_s which presents a problem. Since SLIC operates in the CIELAB color space (where nearby colors are visually similar), possible values for a pixel's color $[lab]^T$ are bounded while a pixel's position $[xy]^T$ depends on the size of the image. Thus simply clustering on the five-dimensional vector $labxy$ will cause inconsistencies in the superpixels. For example, in large superpixels, spatial proximity will dominate color, resulting in superpixels which do not adhere to image boundaries. The converse is true for smaller superpixels. Thus, the individual distances are normalized before combining them into a single measure D using the normalization constants N_c and N_s :

$$d_c = \sqrt{(l_j - l_i)^2 + (a_j - a_i)^2 + (b_j - b_i)^2} \quad (1)$$

$$d_s = \sqrt{(x_j - x_i)^2 + (y_j - y_i)^2} \quad (2)$$

$$D = \sqrt{\left(\frac{d_c}{N_c}\right)^2 + \left(\frac{d_s}{N_s}\right)^2}. \quad (3)$$

The maximum spatial distance corresponds to the maximum sampling interval $N_s = S = \sqrt{N/K}$. Normalizing color is tricky since color distances change from image to image. Thus, the SLIC authors replace N_c by a constant

m . Varying m and S allows the user to control the trade-off between color similarity and spatial similarity. The final distance equation becomes

$$D = \sqrt{\left(\frac{d_c}{m}\right)^2 + \left(\frac{d_s}{S}\right)^2}. \quad (4)$$

ISLIC modifies Equation 4 to include NIR intensity information and (possibly) distance information. We include intensity information by including an additional term in the distance calculation d_i , assuming that materials with a similar intensity are actually similar. Like above, we include a normalization constant to balance the effect of intensity distance with color and spatial distances.

$$d_i = \sqrt{(\text{intensity}_i - \text{intensity}_j)^2} \quad (5)$$

$$D' = \sqrt{\left(\frac{d_c}{m}\right)^2 + \left(\frac{d_s}{S}\right)^2 + \left(\frac{d_i}{N_i}\right)^2}. \quad (6)$$

If distance information is available (e.g., from a laser or Kinect), then we include it to help distinguish object boundaries, especially in deep shadow or where the object's color matches the background. For example, across the edge of an object we expect a significant jump in distance as we transition from the object to the background. We do not include the distance information z_i with the spatial distance term since we assume the distance is measured in some physical unit (e.g., meters) rather than pixels. Again, we normalize the distance by N_z to balance the effects of distance with color, spatial proximity, and intensity. Thus ISLIC's final distance function is:

$$d_z = \sqrt{(z_i - z_j)^2} \quad (7)$$

$$D'' = \sqrt{\left(\frac{d_c}{m}\right)^2 + \left(\frac{d_s}{S}\right)^2 + \left(\frac{d_z}{N_z}\right)^2 + \left(\frac{d_i}{N_i}\right)^2}. \quad (8)$$

4. Experiments

We conducted two sets of experiments to show how adding NIR information into the SLIC distance calculation can improve boundary accuracy. In the first experiment we used a scanning lidar as a NIR illumination source to detect 10 objects in various indoor and outdoor environments with the goal of accurately detecting the object contours. In the second set of experiments we shifted to detecting contours at the scene level in both indoor and outdoor scenes using passively collected NIR images.



Figure 3. Objects used in our experiments: (left to right, top to bottom) orange box, hard hat, black box, small munition, backpack, sweatshirt, tool box, video camera case, medium munition, large munition.



Figure 4. Four environments used in our experiments: (left to right, top to bottom) grass, upland forest, rainforest, and painted concrete.

4.1. Object Segmentation

The first set of experiments sought to quantify the benefit of including lidar information when computing superpixels for object segmentation. Figure 3 shows the 10 different objects we placed in four different environments (see Figure 4). We chose these four environments due to their significant clutter and shadows which challenge traditional superpixel algorithms.

4.1.1 Methodology

For this set of experiments we included intensity and distance information collected from a 3D point cloud.¹ Since

¹We use a Hokuyo UTM-30 laser scanner, but any laser scanner that returns intensity information would work.

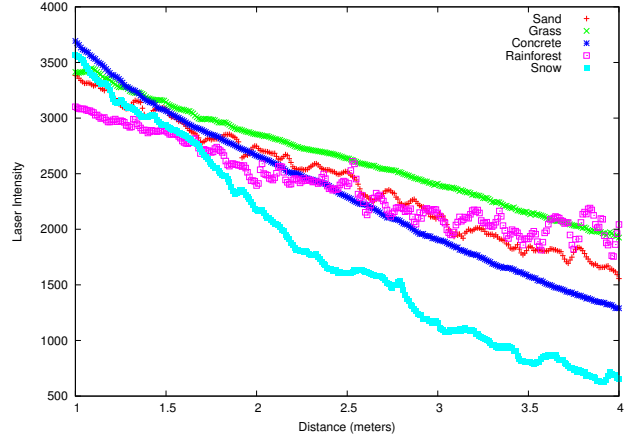


Figure 5. Average spectral intensity values for different outdoor materials. Each curve is the average of 1,000,000 – 3,000,000 data points collected over a variety of incidence angles.

our laser point clouds are sparse (~ 25000 points) compared to the camera image, we first project the point cloud into the coordinate frame of the image, and then apply a median filter (with radius β) to fill-in non-zero values. With this denser laser cloud, we now compute our two new terms for Equation 4. The first term, d_z , uses the distance to the superpixel as determined by the point cloud.

The second term, d_i , is the spectral intensity value of the superpixel. Most modern laser scanners return intensity values in addition to range and bearing, and we exploit that different materials have different spectral reflectance values which are range dependent. Figure 5 shows the average intensity values for several common outdoor terrains. For example, metal and chlorophyll have higher reflectance values than sand and concrete.

Since the laser terms are highly dependent on the point cloud, we compute the normalization constants N_i and N_z based on the input data. For both constants we determine the maximum intensity and distance values, and then scale N_i and N_z by $0 < \alpha \leq 1$.

4.1.2 Results

In our first experiment we placed the objects in each of the four environments (one indoor, three outdoor). For each object and environment we hand-labeled the actual object contours, and error is measured using the average distance from each of the superpixels to the object contour. Note that this inherently favors superpixels that are compact and that naturally align themselves to the contours of the object.

We evaluated the performance of SLIC, SLIC + depth component, and ISLIC. The results are shown in Figure 6. These results were established using parameters that were determined experimentally and vary depending on the environment.

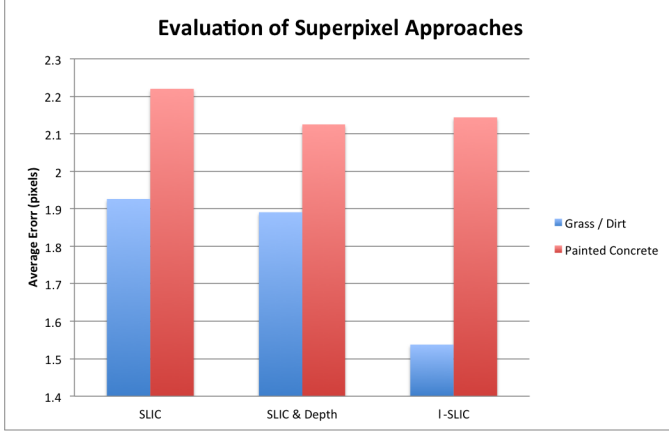


Figure 6. Evaluating performance of SLIC, SLIC with depth measures only, and ISLIC. The Grass/Dirt bar was averaged across the three outdoor environments.

In the indoor environment, where lighting was good, laser reflectance does not significantly improve the results since the bottom of the object was easily separable from the background via color. However, the depth information is useful to properly adhere to the contour along the top of the object; this is especially true when the object is similarly colored to the background.

In the outdoor environments, including the depth distance does not provide a significant improvement over the SLIC algorithm. While this is a bit of a surprising result, it is likely due to the fact that the top of the object has a significantly different color than the naturally colored environment. The results when adding the laser reflectance in are quite striking as they significantly improve performance.

The first experiment used empirically chosen values for the superpixel spacing (S). We performed a second experiment to further analyze the sensitivity of ISLIC to this superpixel spacing. The results of this experiment are in Figure 7. As we increase the distance between the superpixels (i.e., decreasing the number of superpixels), the distance from the object contour naturally increases.

What is interesting is that distances for ISLIC and SLIC increase at approximately the same rate, yet ISLIC dominates the performance of SLIC in this measure. Further, this difference becomes even more pronounced as the number of superpixels decreases.

The performance of ISLIC is sensitive to parameter selection, and automated selection of these parameter values is a topic of ongoing research. Additionally, performance is affected by the transform between the laser point cloud and the RGB image: sensor misalignment results in superpixel edges which do not follow object contours.



Figure 7. Comparing SLIC and ISLIC, varying the size of the superpixel. The graph shows the distance from the superpixel contour to the object contour.

4.2. Scene Segmentation

In our second set of experiments, we sought to examine how including infrared information helps superpixel boundaries in scene segmentation. We used RGB and near-infrared images from the RGB-NIR Database [2]. All images are 1024x768 resolution and the RGB and NIR images were registered using a RANSAC algorithm. We used the indoor (56 image pairs) and urban (57 image pairs) categories of the database for our experiments. Ground truth was obtained by hand labeling major object boundaries for each image. Figure 8 shows some example RGB and NIR images from both the indoor and urban scenes along with the output from ISLIC.

4.2.1 Methodology

We used the NIR image to get intensity information for each pixel. Since the RGB and NIR images are already registered and the NIR image is dense, we did not need to perform any preprocessing on the images. In these experiments, since the NIR images are 8-bit, we knew the range of possible values for each pixel, so we manually set the normalization constant N_i .

4.2.2 Results

Our performance metrics were boundary accuracy [12, 7] and superpixel length. Boundary accuracy computes the percentage of ground truth boundaries that are within 2 pixels of a superpixel boundary. However, boundary accuracy favors long, complex superpixels: in the limit, if each superpixel contained only one pixel, the segmentation would achieve perfect boundary accuracy. Thus, we introduce another metric, called *length*, which measures the length of the superpixel boundaries. The length metric quantifies how



Figure 8. Example images from the RGB-NIR Database. The top row is the RGB image, the middle row is the corresponding NIR image, and the bottom row is the output from ISLIC. The first two columns are from the indoor category while the second two columns are from the urban category.

“noisy” the superpixel boundaries are: cleaner superpixels result in a lower length value, which translates to sharper object boundaries (i.e., Occam’s razor). Cleaner superpixels help with later image processing algorithms such as object identification.

For these experiments we did not use distance information, so we ignored that term in Equation 4. The normalization constants, m and N_i , were chosen experimentally for a given value of S . In the first experiment, we fixed $m = 10$ and $N_i = 10$, and computed the average boundary accuracy and length metrics for both the indoor and urban categories. Figure 9 shows the results. Interestingly, while NIR by itself performs poorly, adding NIR information decreases the length metric without affecting boundary accuracy. This is especially true for small number of superpixels. In other words, adding NIR information results in accurate, clean superpixels.

Figure 10 shows an illustration of how including NIR information results in cleaner superpixels for both the indoor and urban categories. The first column shows the original image, the second column shows the results from SLIC, and the third column shows the results from ISLIC. For both the indoor and outdoor scenes, the boundary accuracy stayed approximately the same (less than 1% difference) but the length metric changed significantly when including NIR information: for indoor it dropped from 195807 to 175092, and for the outdoor image it dropped from 95994 to 86932.

5. Conclusions

We presented an extension of the SLIC superpixel algorithm that incorporates near infrared information into the distance calculation. By including non-visual information, our approach better detects object boundaries, especially in deep shadow and when the object and background are similar in color. Our approach performs well with both active and passive NIR images.

ISLIC inherits SLIC’s speed due to it searching in small $2S \times 2S$ boxes around each pixel rather than exhaustively querying all superpixels. During our experiments, we noticed that ISLIC’s distance computation did not significantly increase runtime. Thus, ISLIC is applicable to real-time use on autonomous robots, especially those equipped with GPUs since ISLIC is clearly parallelizable.

ISLIC is an attractive first step in an image processing pipeline for robotics since it can reduce the computational overhead of subsequent algorithms. For example, instead of applying Chamfer matching to the entire image, we could apply Chamfer matching to just those superpixel edges where the superpixel has certain properties (color, laser reflectance, etc.)

However, ISLIC is sensitive to parameter selection, particularly the normalization constants. We are currently researching methods to automatically set these parameters based on incoming sensor information. Additionally, we are researching methods to optimize the normalization constants (e.g., hill-climbing, stochastic search), and investi-

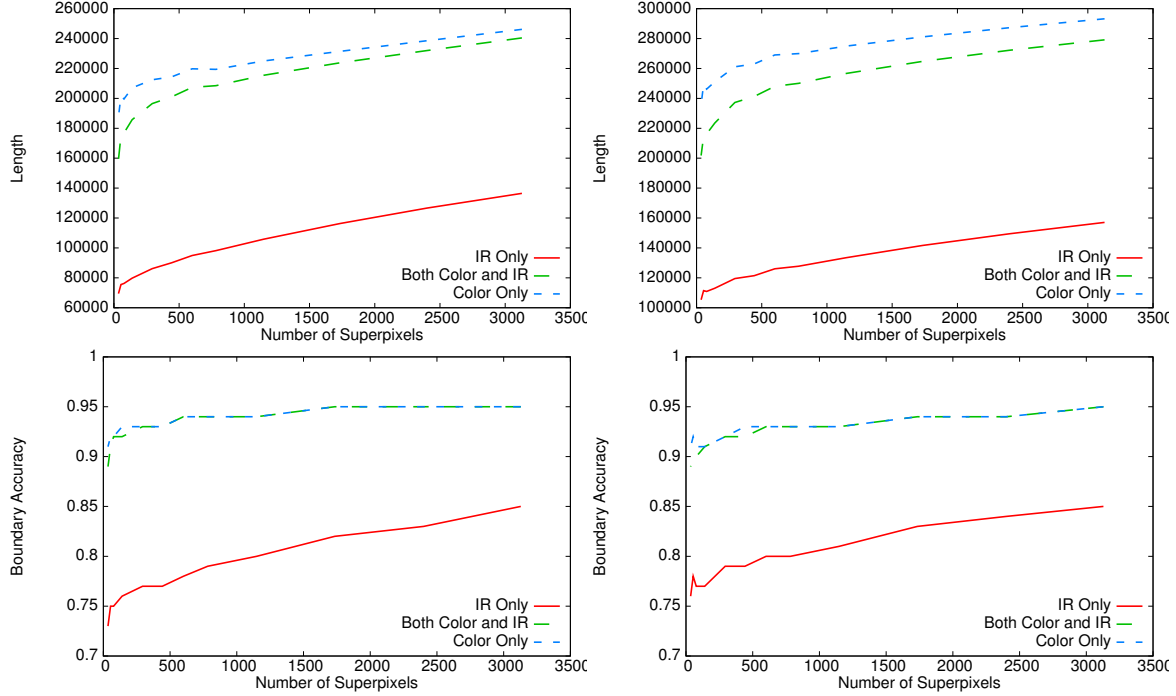


Figure 9. Result for indoor (left) and urban (right) categories. The top row is length and the bottom row is boundary accuracy.

gating how well these optimized constants transfer between environments.

Our generalization of SLIC to the non-visual spectrum opens several avenues for future work. One option is incorporating hyperspectral and/or thermal information. These non-visual imagery techniques could allow SLIC to be used in visually challenging environments such as nighttime and firefighting. Another avenue for future work in incorporating auditory information. High-frequency sonar has been used for vegetation identification [3] and could increase the environments where SLIC is applicable.

References

- [1] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk. SLIC superpixels compared to state-of-the-art superpixel methods. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(11):2274–2282, Nov 2012.
- [2] M. Brown and S. Süsstrunk. Multispectral SIFT for scene category recognition. In *Proceedings of Computer Vision and Pattern Recognition Conference (CVPR)*, 2011.
- [3] P. Caspers, A. Leonessa, and R. Müller. Eigenbeam analysis of the diversity in bat biosonar beampatterns. *The Journal of the Acoustical Society of America*, 135(4), 2014.
- [4] B. Fulkerson, A. Vedaldi, and S. Soatto. Class segmentation and object localization with superpixel neighborhoods. In *Proceedings of the International Computer Vision Conference (ICCV)*, pages 670–677, Sept 2009.
- [5] I. Jebari and D. Filliat. Color and depth-based superpixels for background and object segmentation. *Procedia Engineering*, 41(0):1307 – 1315, 2012.
- [6] D. Kim, S. M. Oh, and J. Rehg. Traversability classification for UGV navigation: a comparison of patch and superpixel representations. In *Proceedings of IEEE International Conference on Intelligent Robots and Systems (IROS)*, pages 3166–3173, Oct 2007.
- [7] A. Levinstein, A. Stere, K. Kutulakos, D. Fleet, S. Dickinson, and K. Siddiqi. TurboPixels: Fast superpixels using geometric flows. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(12):2290–2297, Dec 2009.
- [8] D. Liu, G. Dissanayake, and N. Kirchner. Surface type classification with a laser range finder. *IEEE Sensors Journal*, 2009.
- [9] P. Massoudifar, A. Rangarajan, and P. Gader. Superpixel estimation for hyperspectral imagery. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshop on Perception Beyond the Visible Spectrum*, 2014.
- [10] R. Myneni, F. Hall, P. Sellers, and A. Marshak. The interpretation of spectral vegetation indexes. *IEEE Transactions on Geoscience and Remote Sensing*, 33(2):481–486, Mar 1995.
- [11] S. S. Neda Salamati, C. Fredembach. Material classification using color and NIR images. In *Proceedings of the Color Imaging Conference (CIC)*, 2009.
- [12] P. Neubert and P. Protzel. Compact watershed and preemptive SLIC: On improving trade-offs of superpixel segmentation algorithms. In *Proceedings of the International Conference on Pattern Recognition (ICPR)*, pages 996–1001, Aug 2014.
- [13] C. Rasmussen and D. Scott. Shape-guided superpixel grouping for trail detection and tracking. In *Proceedings of IEEE*



Figure 10. An example of the length metric. The top row is from the indoor category while the bottom row is from the urban category. The first column shows the original image, the second column shows the standard SLIC algorithm, and the third column shows the ISLIC algorithm.

- International Conference on Intelligent Robots and Systems (IROS)*, pages 4092–4097, Sept 2008.
- [14] C. Y. Ren and I. Reid. gSLIC: a real-time implementation of SLIC superpixel segmentation. Technical report, University of Oxford, Department of Engineering Science, 2011.
 - [15] X. Ren, L. Bo, and D. Fox. RGB-(D) scene labeling: Features and algorithms. In *Proceedings of Computer Vision and Pattern Recognition Conference (CVPR)*, pages 2759–2766, June 2012.
 - [16] N. Salamati, D. Larlus, G. Csurka, , and S. Süsstrunk. Semantic image segmentation using visible and near-infrared channels. In *Proceedings of European Computer Vision Conference*, 2012.
 - [17] N. Salamati and S. Süsstrunk. Material-based object segmentation using near-infrared information. In *Proceedings of the Color Imaging Conference (CIC)*, 2010.
 - [18] J. Shi and J. Malik. Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8):888–905, Aug 2000.
 - [19] K. Sullivan, W. Lawson, and D. Sofge. Fusing laser reflectance and image data for terrain classification for small autonomous robots. In *Proceedings of International Conference on Control, Automation, Robotics and Vision*, 2014.
 - [20] D. Thompson, R. Castano, and M. Gilmore. Sparse superpixel unmixing for exploratory analysis of CRISM hyperspectral images. In *Proceedings of the Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing*, pages 1–4, Aug 2009.
 - [21] D. Thompson, L. Mandrake, M. Gilmore, and R. Castano. Superpixel endmember detection. *IEEE Transactions on Geoscience and Remote Sensing*, 48(11):4023–4033, Nov 2010.
 - [22] D. Weikersdorfer, D. Gossow, and M. Beetz. Depth-adaptive superpixels. In *Proceedings of International Conference on Pattern Recognition (ICPR)*, pages 2087–2090, Nov 2012.
 - [23] K. Wurm, R. Kummerle, C. Stachniss, and W. Burgard. Improving robot navigation in structured outdoor environments by identifying vegetation from laser data. In *Proceedings of IEEE International Conference on Intelligent Robots and Systems (IROS)*, pages 1217–1222, Oct 2009.