

A discrepancy measure for segmentation evaluation from the perspective of object recognition

Jian Yang, Yuhong He, John Caspersen, Trevor Jones

Version Post-print/Accepted Manuscript

Citation (published version) Jian Yang, Yuhong He, John Caspersen, Trevor Jones. A discrepancy measure for segmentation evaluation from the perspective of object recognition. ISPRS Journal of Photogrammetry and Remote Sensing. Volume 101, 2015. doi: <https://doi.org/10.1016/j.isprsjprs.2014.12.015>

DOI: <https://doi.org/10.1016/j.isprsjprs.2014.12.015>

Copyright/License



This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License. To view a copy of this license, visit [Creative Commons BY NC ND 4.0 License](https://creativecommons.org/licenses/by-nc-nd/4.0/).

How to cite TSpace items

Always cite the **published version**, so the author(s) will receive recognition through services that track citation counts, e.g. Scopus. If you need to cite the page number of the **author manuscript from TSpace** because you cannot access the published version, then cite the TSpace version **in addition to** the published version using the permanent URI (handle) found on the record page.

This article was made openly accessible by U of T Faculty.
Please [tell us](#) how this access benefits you. Your story matters.

1 **A discrepancy measure for segmentation evaluation from the perspective of object recognition**

2
3 *Jian Yang^{1,2,*}, Yuhong He², John Caspersen³, Trevor Jones⁴*

4
5 *¹ Department of Geography, University of Toronto, 100 St. George Street, Toronto, ON M5S 3G3,*
6 *Canada.*

7 *² Department of Geography, University of Toronto Mississauga, 3359 Mississauga Rd North,*
8 *Mississauga, ON L5L 1C6, Canada.*

9 *³ Faculty of Forestry, University of Toronto, 33 Willcocks Street, Toronto, ON M5S 3B3, Canada.*

10 *⁴ Ontario Forest Research Institute, Ministry of Natural Resources, 1235 Queen Street East, Sault Ste*
11 *Marie, ON, P6A 2E5, Canada.*

12
13 ** Corresponding author. Tel.: +1 416 978 3375.*

14 *Email address: jiangeo.yang@mail.utoronto.ca (J. Yang).*

24 **Abstract**

25 Image segmentation is one of the most important components within the framework of object
26 based image analysis. Segmentation evaluation thus plays a critical role in controlling the quality of object
27 based image analysis workflow. Among a variety of segmentation evaluation methods and criteria,
28 discrepancy measurement is believed to be the most useful and commonly employed in many applications.
29 However, existing measures have largely ignored the importance of object recognition in segmentation
30 evaluation. In this study, a new discrepancy measure of Segmentation Evaluation Index (*SEI*) is proposed
31 to overcome this limitation by redefining the corresponding segment using the condition of two-side 50%
32 overlap instead of one-side 50% overlap that has been commonly used. The effectiveness of *SEI* is further
33 investigated using the schematic segmentation cases and remote sensing images. Results demonstrate that,
34 the proposed *SEI* outperforms the other two existing discrepancy measures in terms of object recognition
35 accuracy and identification of detailed segmentation difference.

36
37 **Keywords:** *image segmentation, quantitative evaluation, discrepancy measure, object recognition,*
38 *corresponding segment, Segmentation Evaluation Index (SEI)*

47 **1 Introduction**

48 As is widely accepted, image segmentation is the first, fundamental, and critical step for object
49 based image analysis. It is a process of partitioning the entire image into a number of non-overlapping
50 segments for the subsequent object recognition, image classification, or information extraction. In the past
51 decades, a variety of popular segmentation algorithms have been proposed, such as watershed
52 segmentation (Li et al., 2010; Li & Xiao, 2007; Vincent & Soille, 1991), mean-shift segmentation
53 (Comaniciu & Meer, 2002), and the fractal net evolution approach (Baatz & Schäpe, 2000; Benz et al.,
54 2004). Nevertheless, it still remains challenging and problematic to integrate a comprehensive framework
55 for segmentation evaluation (Ryherd & Woodcock, 1996; Shandley, Franklin & White, 1996).

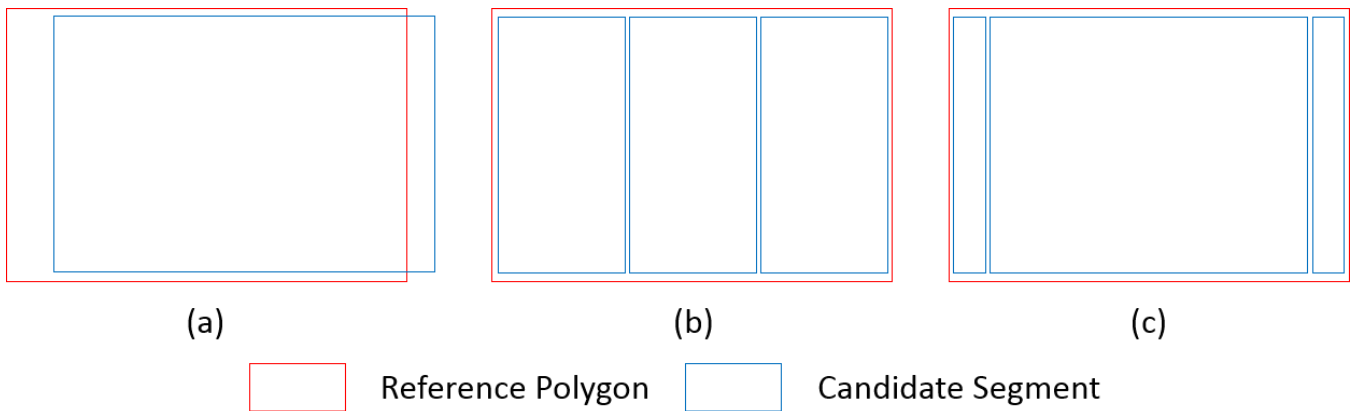
56 In general, the methods of segmentation evaluation include visual inspection, quantitative
57 evaluation, and indirect evaluation (e.g., classification accuracy) (Li et al., 2011). Because human eyes
58 are considered as a strong and experienced source, visual inspection is a commonly-used evaluation
59 method although it might be subjective and qualitative (Pesaresi & Benediktsson, 2001; Zhang, Fritts &
60 Goldman, 2008). On the other hand, classification accuracy can imply the quality of image segmentation
61 in some way, but it is not a direct indication of the segmentation accuracy (Kim, Madden & Warner, 2009;
62 Zhang et al., 2005). Despite some popularity of visual inspection and indirect evaluation, quantitative
63 evaluation has attracted growing attention for segmentation evaluation through providing more detailed
64 information (e.g., over-segmentation and under-segmentation).

65 Quantitative evaluation methods can be categorized into three types: analytical, empirical
66 goodness, and empirical discrepancy, among which empirical discrepancy has been proven the most
67 effective and widely used (Zhang, 1996). To assess image segmentation, the evaluation based on empirical
68 discrepancy measures the dissimilarity between a reference polygon and a corresponding segment. A few
69 discrepancy measures were first proposed to quantify their overlapping area (Carleer, Debeir & Wolff,

70 2005; Clinton et al., 2010; Lucieer & Stein, 2002; Möller, Lymburner & Volk, 2007; Weidner, 2008; Zhan
71 et al., 2005). These overlapping area based measures, later categorized as geometric discrepancy (Liu et
72 al., 2012), are able to characterise the non-overlapping area of a reference polygon and a corresponding
73 segment (Fig. 1a), however, they require that a reference polygon should only have one corresponding
74 segment. In reality, it is very common for a reference polygon to have a number of corresponding segments
75 (Fig. 1b). The prerequisite of one reference to only one corresponding segment is thus almost impossible
76 to satisfy when using the geometric discrepancy measures. The flaw in the geometric discrepancy
77 measures has been recently addressed using newly proposed indices which consider arithmetic
78 discrepancy (Liu et al., 2012; & Yang, Li and He, 2014). Arithmetic discrepancy is defined as the number
79 of corresponding segments for any reference polygon (Liu et al., 2012).

80 In general, the discrepancy measures that are proposed by previous studies label a candidate
81 segment as the corresponding segment of a reference polygon when the overlapping area is over 50% of
82 **EITHER** the reference polygon **OR** the candidate segment (thereafter defined as one-side 50% overlap)
83 (Clinton et al., 2010; Liu et al., 2012; Yang, Li & He, 2014). However, the use of one-side 50% overlap
84 in the discrepancy measures would not work in many segmentation cases, due to the ignorance of object
85 recognition in segmentation evaluation. Using two segmentation results for a polygon as an example (Fig.
86 1b and 1c), all segments (in blue) in the two cases can be considered as corresponding segments for the
87 reference polygon (in red) because the overlapping area is over 50% (in fact 100%) of each candidate
88 segment and the one-side 50% overlap prerequisite for the discrepancy measures is thus satisfied.
89 Consequently, the two segmentation results for the polygon are exactly the same in terms of geometric
90 discrepancy (i.e. the non-overlapping areas are the same) and arithmetic discrepancy (i.e. three segments
91 corresponding to one reference polygon are the same in both cases). However, it is obvious that the
92 candidate segments in Fig. 1c contribute a higher segmentation quality than those in Fig. 1b because the

93 largest segment in Fig. 1c can better recognize the reference polygon. This example clearly demonstrated
94 that the current discrepancy measures cannot efficiently recognize objects as a result of only utilizing one-
95 side 50% overlap. To differentiate these two cases, a new discrepancy (thereafter defined as object-
96 recognized discrepancy) is needed to identify whether or not a corresponding segment is able to well
97 recognize a reference object. Differing from geometric and arithmetic discrepancy, the new discrepancy
98 will incorporate the object recognition information for a correctly recognized object through the
99 prerequisite that the overlapping area between a reference polygon and the candidate segment has to be
100 more than 50% of **BOTH** the reference polygon **AND** the candidate segment (Lamar, McGraw & Warner,
101 2005).



102

103 Fig. 1. An illustration of discrepancies between a reference polygon and candidate segments, including
104 geometric discrepancy (a), arithmetic discrepancy (b), and object-recognized discrepancy (c).

105

106 As far as we know, no study has yet incorporated the information of object recognition into the
107 discrepancy measures, in spite of the fact that object recognition is a most important criterion when
108 measuring the discrepancy between a reference object and a corresponding segment. It is thus of great
109 necessity to identify if the reference object is correctly recognition before a detailed discrepancy
110 measurement is calculated. In order to address this knowledge gap, we propose a new discrepancy measure

111 to evaluate the quality of image segmentation which takes into account geometric discrepancy, arithmetic
112 discrepancy, and object-recognized discrepancy. Schematic segmentation cases and remote sensing
113 images are used to examine the performance of the proposed index in comparison with the other two
114 existing discrepancy measures (Liu et al., 2012; Yang, Li & He, 2014).

115

116 **2 Discrepancy measures**

117 Since object recognition is one of the most important objectives of image segmentation, this study
118 adopted the two-side 50% overlap as the prerequisite to identify the corresponding segment for a reference
119 polygon. Under this context, the reference polygon without any corresponding segment will be considered
120 as an omitted or missing object in the process of segmentation evaluation.

121 Several discrepancy measures that have recently been proposed based on the one-side 50% overlap
122 of corresponding segments include Potential Segmentation Error (*PSE*), Number-of-Segments Ratio (*NSR*)
123 and Euclidian Distance 2 (*ED2*) (Liu et al., 2012).

$$\begin{aligned} PSE &= \frac{\text{area}(S - R)}{\text{area}(R)} \\ NSR &= \frac{|m - v|}{m} \\ ED2 &= \sqrt{PSE^2 + NSR^2} \end{aligned} \quad (1)$$

124
125 where R and S are the datasets of reference polygons and corresponding segments while m and v are the
126 numbers of reference polygons and corresponding segments. A *PSE* value of zero implies that there is no
127 under-segment, and a *NSR* value of zero means an optimal one-to-one relationship between the reference
128 polygons and corresponding segments.

129 Although *ED2* considers both geometric and arithmetic matches, the difficulty of *PSE* and *NSR*
130 normalization can result in the exaggeration of over-segmentation at finer scales when *NSR* overwhelms
131 *PSE* (Yang, Li & He, 2014). In addition, the compensation effect caused by the coexistence of one-to-

132 many over-segmentation and many-to-one under-segmentation can result in invalid *NSR*. For instance,
 133 since the one-side 50% overlap fails to ensure that there is at least a candidate segment corresponding to
 134 any one reference polygon, the over-segmentation of other reference objects can compensate this effect
 135 on the *NSR* even if there is an omitted or missing reference object. Thus, Yang, Li and He (2014)
 136 developed the local metrics of OverSegmentation 2 (*OS2*), UnderSegmentation 2 (*US2*), and Euclidian
 137 Distance 3 (*ED3*).

$$\begin{aligned}
 OS2 &= \sum_i \sum_j \left(1 - \frac{area(r_i \cap s_j)}{area(r_i)}\right), s_j \in S \\
 US2 &= \sum_i \sum_j \left(1 - \frac{area(r_i \cap s_j)}{area(s_j)}\right), s_j \in S \\
 ED3 &= \sum_i \sum_j \sqrt{\frac{\left(1 - \frac{area(r_i \cap s_j)}{area(r_i)}\right)^2 + \left(1 - \frac{area(r_i \cap s_j)}{area(s_j)}\right)^2}{2}}, s_j \in S
 \end{aligned} \tag{2}$$

139 where r_i is an arbitrary element of R and s_j is the corresponding segment, as an element of S .

140 Despite the merits of *ED2* and *ED3* as discrepancy measures, both of them ignore the accuracy of
 141 object recognition due to the use of one-side 50% overlap of corresponding segments. In this study, a new
 142 measure of Segmentation Evaluation Index (*SEI*) was developed from *ED3* using the corresponding
 143 segments identified by the two-side 50% overlap.

$$\begin{aligned}
 SEI_{local}(i) &= \begin{cases} \sqrt{\frac{\left(1 - \frac{area(r_i \cap s_i)}{area(r_i)}\right)^2 + \left(1 - \frac{area(r_i \cap s_i)}{area(s_i)}\right)^2}{2}}, s_i \in S_{2S} \\ 1, s_i \notin S_{2S} \end{cases} \\
 SEI &= \frac{1}{n} \sum_{i=1}^n SEI_{local}(i)
 \end{aligned} \tag{3}$$

145 where $SEI_{local}(i)$ denotes the local SEI for the i th reference polygon and n denotes the number of reference
146 polygons. r_i is the i th reference polygon while s_i is the corresponding segment, as an element of the two-
147 side 50% overlap of corresponding segment dataset S_{2S} .

148 Under the prerequisite of two-side 50% overlap, it is not possible to identify more than one
149 candidate segment corresponding to a given reference polygon. When a candidate segment correctly
150 recognizes an object, SEI_{local} is able to measure the discrepancy between the corresponding segment and
151 the reference polygon. A SEI_{local} value of zero indicates perfect-segmentation, neither over-segmentation
152 nor under-segmentation. If there is no candidate segment corresponding to a given reference polygon, the
153 value of SEI_{local} is defined as one, indicating that it is an omitted or missing object. Given the SEI_{local}
154 values of all reference polygons, the value of SEI is ranging from zero to one. Apparently, a higher value
155 of SEI suggests an overall lower quality of image segmentation.

156 In this study, both schematic segmentation cases and remote sensing images were utilized to
157 examine the performance of SEI for discrepancy measurement, compared with the mentioned two metrics
158 of $ED2$ and $ED3$.

159

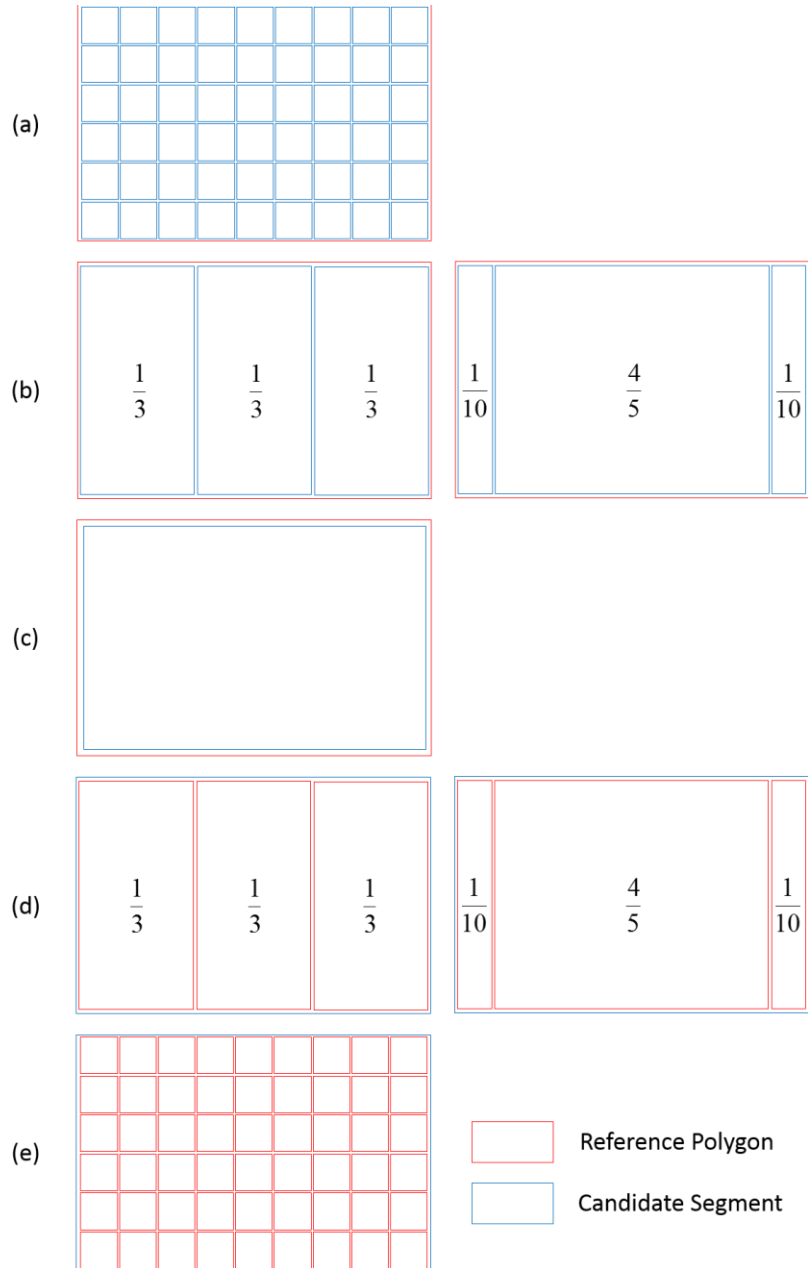
160 **3. Implementations**

161 **3.1: Schematic segmentation cases**

162 The proposed index (SEI) was first validated using five basic schematic cases (Fig. 2) including
163 absolute over-segmentation, one-to-many over-segmentation, one-to-one perfect-segmentation, many-to-
164 one under-segmentation, and absolute under-segmentation. It is worth noting that the absolute over-
165 segmentation represents the case that a reference polygon is divided into infinite segments while the
166 absolute under-segmentation is for the case that infinite reference polygons are belong to a segment. When
167 segmenting remote sensing images, the absolute over-segmentation does not exist because the smallest

168 unit of an image is a pixel. Meanwhile, a finite number of objects within a remote sensing image make
 169 the absolute under-segmentation impossible. Despite the inexistence of these two extreme cases in an
 170 image, this study took both of them into consideration to enable a complete set of schematic segmentation
 171 cases for validating the SEI.

172



173

174 Fig. 2. Schematic representation of the schematic segmentation cases: absolute over-segmentation (a),
 175 one-to-many over-segmentation (b), one-to-one perfect-segmentation (c), many-to-one under-
 176 segmentation (d), and absolute under-segmentation (e).

177

178 Table 1. Discrepancy measure calculation for absolute over-segmentation (AbOver), one-to-many over-
 179 segmentation (Over), one-to-one perfect-segmentation (Perfect), many-to-one under-segmentation (Und),
 180 and absolute under-segmentation (AbUnd).

	AbOver	Over (left)	Over (right)	Perfect	Und (right)	Und (left)	AbUnd
<i>PSE</i>	0	0	0	0	0	0	0
<i>NSR</i>	Infinite	2	2	0	0.67	0.67	1
<i>ED2</i>	Infinite	2	2	0	0.67	0.67	1
<i>OS2</i>	1	0.47	0.47	0	0	0	0
<i>US2</i>	0	0	0	0	0.47	0.47	1
<i>ED3</i>	1	0.47	0.47	0	0.47	0.47	1
<i>SEI</i>	1	1	0.14	0	0.71	1	1

181

182 The *ED2*, *ED3*, and *SEI* for the schematic segmentation cases are calculated and listed in Table 1.
 183 All the discrepancy measures (*ED2*, *ED3*, and *SEI*) are successful in characterizing the three extreme cases
 184 including absolute over-segmentation, one to one perfect segmentation, and absolute under-segmentation,
 185 although the ranges of values are not the same (Table 1). For instance, the value of *ED2* is infinite for the
 186 absolute over-segmentation representation whereas both *ED3* and *SEI* are one. Further, the value of *ED2*
 187 is not completely symmetrical: ranging from zero to infinite in the over-segmentation case while from
 188 zero to one in the under-segmentation case. This result indicates that both *SEI* and *ED3* can characterize

189 the three extreme cases perfectly, while *ED2* is too sensitive for over-segmentation at finer scales but less
190 sensitive for under-segmentation at coarser scales, which is also consistent with the results in Witharana
191 and Civco (2014). In other words, it is very likely that *ED2* amplifies the effectiveness of over-
192 segmentation in contrast to that of under-segmentation (Yang, Li & He, 2014).

193 The proposed *SEI* is able to show advantages over the other two measures when evaluating the
194 segmentation in more common-seen cases (i.e. one-to-many over-segmentation and many-to-one under-
195 segmentation), demonstrated by below examples. The over-segmentation example in Fig. 2b where a
196 reference polygon (in Red) is divided into three segments (in Blue) in two segmentation cases (left and
197 right), but the segments vary in area proportion between the cases. The right segmentation case is
198 apparently better than the left one from the perspective of object recognition, because the largest segment
199 that occupies 4/5 of the reference polygon in the right case is able to represent it, whereas none of segments
200 on the left recognizes the reference polygon. However, neither *ED2* nor *ED3* is able to identify the
201 difference because they conclude the same value for both cases (a value of 2 for *ED2* and 0.47 for *ED3*,
202 Table 1). That is because the one-side 50% overlap allows all the three segments in these two cases
203 becoming the corresponding segments when calculating *ED2* and *ED3*. On the contrary, the proposed *SEI*,
204 which uses the two-side 50% overlap as the prerequisite, overcomes the limitation of *ED2* and *ED3*, and
205 produced a value of 1 (an indication of the worse quality) for the left segmentation case and a value of
206 0.14 (an indication of the better quality) for the right segmentation case. In the under-segmentation cases
207 (Fig. 2d left and right), only the *SEI* is sensitive enough to detect the dissimilarity of two under-
208 segmentation cases. The *ED2* and *ED3* show no differences in their values between two cases (i.e. 0.67
209 for *ED2* and 0.47 for *ED3* in both cases). Differently, the *SEI* (the value of 1) for the left under-
210 segmentation case is higher than that (0.71) for the right under-segmentation cases. The difference of *SEI*
211 values is mainly attributed from object recognition.

212 In addition, the *SEI* value of 0.14 in the over-segmentation case (Fig. 2b right) is lower than that
213 of 0.71 in the under-segmentation case (Fig. 2d right). This is expected because one object out of one is
214 recognized in the over-segmentation case (Fig. 2b right) while only one object out of three is recognized
215 in the under-segmentation case (Fig. 2d right). It is thus reasonable that the former one represents a higher
216 quality of image segmentation.

217 We can thus conclude that the proposed index of *SEI* outperforms *ED2* and *ED3* based on the
218 investigation of schematic segmentation cases. Further examination is conducted using the remote sensing
219 images from both urban and natural forested areas.

220

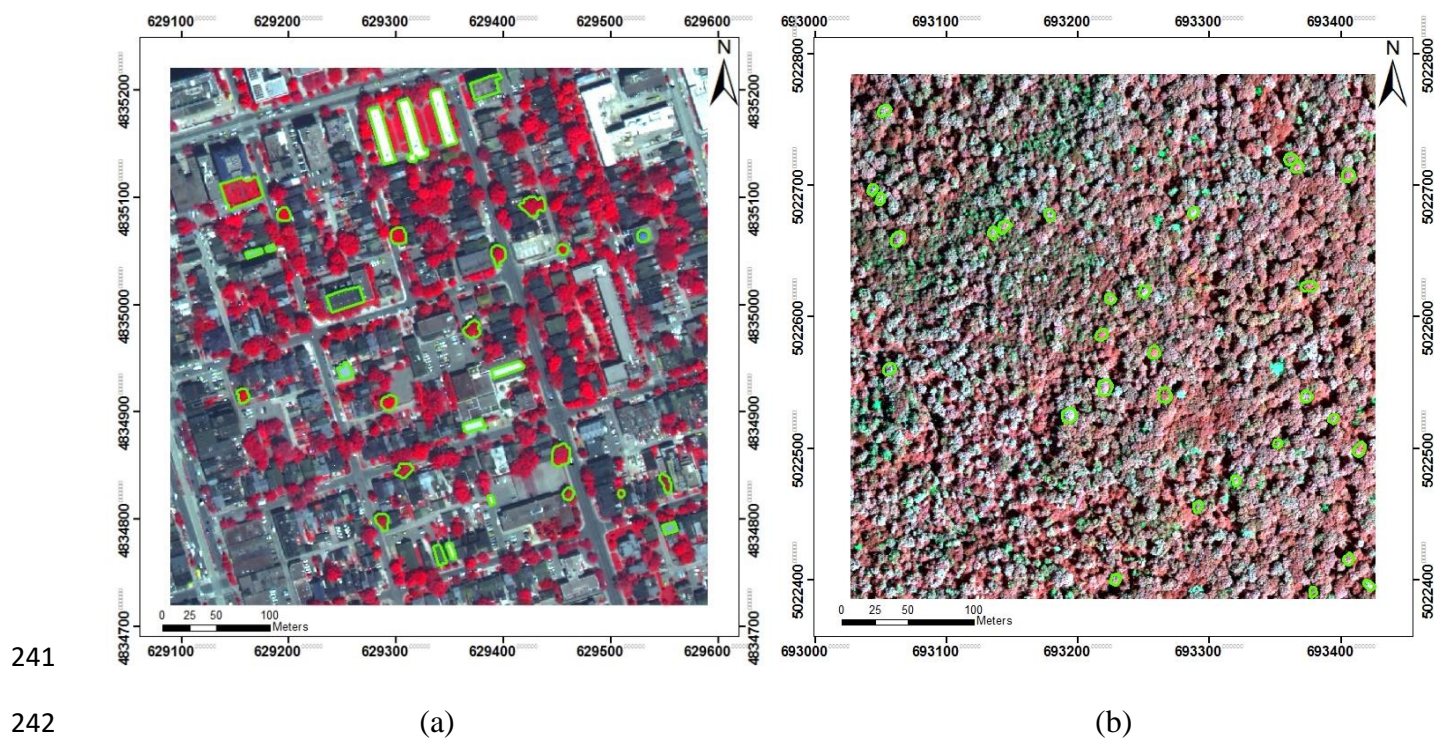
221 **3.2 Experiment 2: remote sensing images**

222 3.2.1 Image data preprocessing

223 Two image datasets were adopted to examine the effectiveness of the proposed *SEI* in
224 segmentation evaluation. One is a WorldView-2 satellite image acquired from the dense urban area in
225 downtown Toronto, Ontario, Canada on June 2nd of 2011, and the other is an ADS40 aerial image collected
226 from the deciduous forested area in Haliburton Forest, Ontario, Canada by Ontario Ministry of Natural
227 Resources during the summer of 2007. The images from urban and forest land were selected because
228 image segmentation has been typically applied in these two types of areas.

229 The WorldView-2 dataset contains a panchromatic band and eight multispectral bands, consisting
230 of coastal (400-450nm), blue (450-510nm), green (510-580nm), yellow (585-625nm), red (630-690nm),
231 red edge (705-745nm), near infrared 1 (770-895nm), and near infrared 2 (860-900nm). The pan-sharpened
232 image with the spatial resolution of 0.5 m was obtained using the Gram-Schmidt procedure (Brower &
233 Laben, 2000) implemented in the ENVI software package. Meanwhile, the multispectral aerial image
234 contains four multispectral bands with the spatial resolution of 0.4 m, including blue, green, red, and near

235 infrared. In order to investigate the performance of *SEI*, two subsets of 1000×1000 pixels were clipped
236 from the pan-sharpened WorldView-2 satellite (Fig. 3a) and ADS40 aerial (Fig. 3b) images, respectively.
237 Particularly, the urban scene describes a typical built-up area with trees, grass and many high-rise
238 buildings, and the forest scene characterizes a deciduous stand with different types of broad-leaved trees.
239 The diversity of land cover objects makes it suitable to verify the robustness and transferability of the
240 proposed *SEI* in segmentation evaluation.



243 Fig. 3. Pan-sharpened WorldView-2 satellite image of downtown Toronto with a band combination of
244 near infrared 1, red, and green as R, G, and B (a), and ADS40 aerial image of Haliburton Forest with a
245 band combination of near infrared, red, and green as R, G, and B (b). Both images are under WGS84 UTM
246 coordinate system. Green polygons represent the manually digitized reference objects.

247

248 3.2.2 Image segmentation

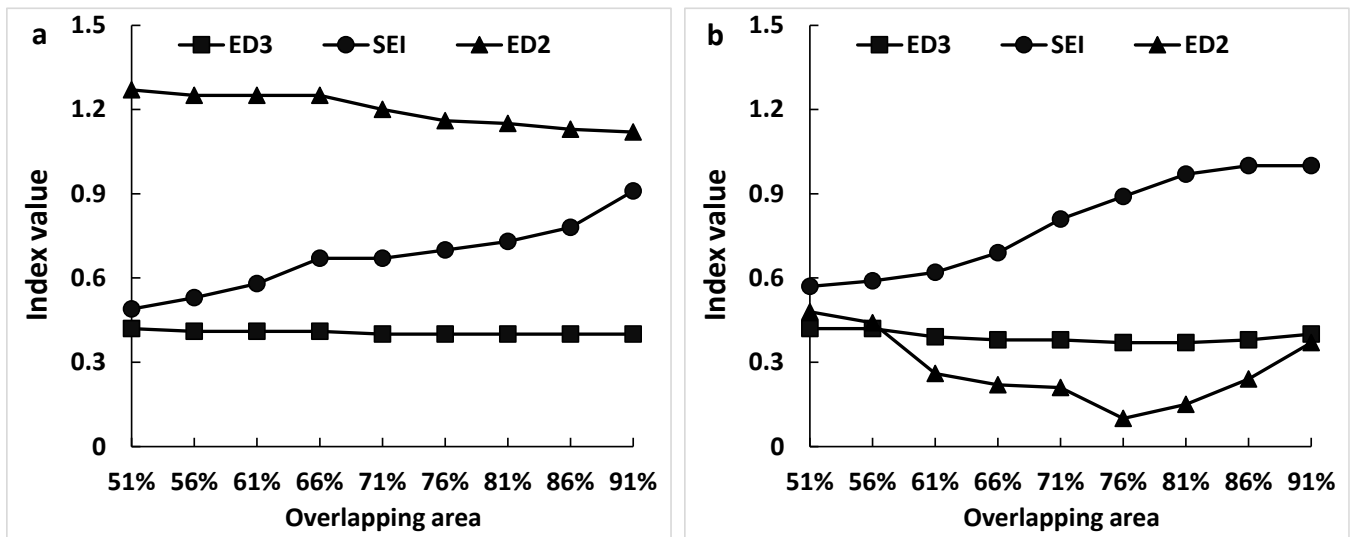
249 This study employed the BerkeleyImageSeg software package (BIS; <http://www.imageseg.com>)
250 to produce the appropriate segmentation for the above two remote sensing images. The program utilized
251 the region merging algorithm proposed by Benz et al. (2004) to partition the input image into segments
252 based on three parameters of threshold, shape (0.1-0.9), and compactness (0.1-0.9). While a higher
253 threshold indicates the larger size of segments, the higher shape and compactness result in the more round
254 and smooth segments, respectively. After a series of trial-and-error parameter selection, we chose to use
255 100, 0.5, and 0.5 as threshold, shape, and compactness for the segmentation of pan-sharpened WorldView-
256 2 satellite image, and 15, 0.9, and 0.9 as threshold, shape, and compactness for the segmentation of ADS40
257 aerial image. In comparison with the urban scene, the forest scene required a finer threshold due to the
258 smaller size and less distinct boundary of tree crowns. Moreover, the highest shape and compactness
259 ensured the good roundness and smoothness of tree crowns in the forest scene.

260 3.2.3 Segmentation evaluation

261 For segmentation evaluation, a total of 30 reference polygons was randomly delineated for each
262 scene (Fig. 3). Particularly, the reference objects in the urban scene include vegetation and impervious
263 rooftops while the reference objects in the forest scene consist of individual deciduous tree crowns. As
264 mentioned above, the proposed *SEI* considers two-side 50% overlaps while *ED2* and *ED3* only consider
265 one-side 50% overlap as the condition to identify corresponding segments. In order to observe the
266 performance of *SEI* for discrepancy measurement, this study adjusted the overlapping area from 51% to
267 91% with an interval of 5% and calculated all three measures (the *SEI*, *ED2* and *ED3* in Fig. 4). Since a
268 larger overlapping area indicates a higher criterion of segmentation evaluation, fewer candidate segments
269 can become the corresponding segments of a given set of reference polygons. Theoretically, an ideal
270 discrepancy measure should be able to capture this tendency, the quality of image segmentation
271 substantially decreasing as the overlapping area raises.

272 The results in Fig. 4 indicated that only *SEI* is able to capture the decreased quality of image
 273 segmentation as the overlapping area raises in both urban and forest scenes. In the urban scene (Fig. 4a),
 274 the decreasing tendency of *ED2* or *ED3* implies an increasing quality of image segmentation, which is not
 275 correct but expected because both measures would detect lower number but higher quality of
 276 corresponding segments as the overlapping areas increases, at an expense of ignoring reference polygons
 277 without any corresponding segments. Addressing the issue in the *ED2* and *ED3*, the proposed *SEI* uses
 278 one to represent an omitted or missing object (i.e. no corresponding segments) and thus accurately
 279 characterizes the tendency with the increasing value from 0.49 to 0.91. In the forest scene (Fig. 4b), *ED2*
 280 or *ED3* failed to identify the variation of segmentation quality. A V-shape reversal of *ED2* value emerges
 281 around the overlapping area of 76%. At that point, the number of corresponding segments determined by
 282 the one-side overlap is just equal to the number of reference polygons. So the *NSR* value of zero makes
 283 the *ED2* value the least there.

284



285

286 Fig. 4. Discrepancy measures versus the increasing overlapping area in both urban (a) and forest
 287 (b).

288 In summary, the proposed *SEI* not only takes the object recognition into consideration as indicated
289 in the schematic cases, but also provides valid discrepancy measurement which is not possible through
290 the utilization of the other two indices.

291

292 **4 Conclusions**

293 When evaluating the quality of image segmentation, it is of great significance to measure the
294 dissimilarity between a reference polygon and a corresponding segment using the appropriate metric.
295 However, a higher priority should be assigned to object recognition by determining whether a reference
296 object is correctly recognized before discrepancy measurement. In this study, a new index of *SEI* was
297 developed through the utilization of the corresponding segments defined by the condition of two-side 50%
298 overlap. Both the schematic segmentation cases and remote sensing images were employed to investigate
299 the performance of *SEI*. The results showed that the proposed *SEI* was more effective to consider object
300 recognition accuracy and to identify detailed segmentation difference, in comparison with the existing
301 discrepancy measures of *ED2* and *ED3*,

302 Although the newly proposed *SEI* was only examined by an urban area and a forest area, we believe
303 that it can be commonly applied to evaluate image segmentation for different types of landscapes based
304 on the verification of schematic segmentation cases. Additionally, the discrepancy measure of *SEI* is also
305 promising as a supervised method for selecting the optimal segmentation scale in object based image
306 analysis, which will be further implemented in the future work.

307 **Acknowledgements**

308 Support for this study was provided by NSERC Discovery Grant to Dr. Yuhong He in the
309 University of Toronto, the Ontario Forest Research Institute (OFRI), and the Ministry of Natural
310 Resources (MNR), Canada.

311 **References**

- 312 Baatz, M., & Schäpe, A. (2000). Multiresolution segmentation: an optimization approach for high
313 quality multi-scale image segmentation. *Angewandte Geographische Informationsverarbeitung*
314 *XII*, 12-23.
- 315 Benz, U.C., Hofmann, P., Willhauck, G., Lingenfelder, I., & Heynen, M. (2004). Multi-resolution,
316 object-oriented fuzzy analysis of remote sensing data for GIS-ready information. *ISPRS Journal*
317 *of Photogrammetry and Remote Sensing*, 58, 239-258.
- 318 Brower, B.V., & Laben, C.A. (2000). Process for enhancing the spatial resolution of multispectral
319 imagery using pan-sharpening. U.S. Patent 6 011 875, Jan. 4, 2000.
- 320 Carleer, A., Debeir, O., & Wolff, E. (2005). Assessment of very high spatial resolution satellite image
321 segmentations. *Photogrammetric Engineering and Remote Sensing*, 71, 1285-1294.
- 322 Clinton, N., Holt, A., Scarborough, J., Yan, L., & Gong, P. (2010). Accuracy assessment measures for
323 object-based image segmentation goodness. *Photogrammetric Engineering and Remote Sensing*,
324 76, 289-299.
- 325 Comaniciu, D., & Meer, P. (2002). Mean shift: A robust approach toward feature space analysis. *IEEE*
326 *Transactions on Pattern Analysis and Machine Intelligence*, 24, 603-619.
- 327 Kim, M., Madden, M., & Warner, T.A. (2009). Forest type mapping using object-specific texture
328 measures from multispectral IKONOS imagery: Segmentation quality and image classification
329 issues. *Photogrammetric Engineering and Remote Sensing*, 75, 819-829.
- 330 Lamar, W.R., McGraw, J.B., & Warner, T.A. (2005). Multitemporal censusing of a population of
331 eastern hemlock (< i> Tsuga canadensis</i> L.) from remotely sensed imagery using an
332 automated segmentation and reconciliation procedure. *Remote Sensing of Environment*, 94, 133-
333 143.

334 Li, D., Zhang, G., Wu, Z., & Yi, L. (2010). An edge embedded marker-based watershed algorithm for
335 high spatial resolution remote sensing image segmentation. *IEEE Transactions on Image*
336 *Processing, 19*, 2781-2787.

337 Li, P., Guo, J., Song, B., & Xiao, X. (2011). A multilevel hierarchical image segmentation method for
338 urban impervious surface mapping using very high resolution imagery. *IEEE Journal of Selected*
339 *Topics in Applied Earth Observations and Remote Sensing, 4*, 103-116.

340 Li, P., & Xiao, X. (2007). Multispectral image segmentation by a multichannel watershed - based
341 approach. *International Journal of Remote Sensing, 28*, 4429-4452.

342 Liu, Y., Bian, L., Meng, Y., Wang, H., Zhang, S., Yang, Y., Shao, X., & Wang, B. (2012). Discrepancy
343 measures for selecting optimal combination of parameter values in object-based image analysis.
344 *ISPRS Journal of Photogrammetry and Remote Sensing, 68*, 144-156.

345 Lucieer, A., & Stein, A. (2002). Existential uncertainty of spatial objects segmented from satellite sensor
346 imagery. *IEEE Transactions on Geoscience and Remote Sensing, 40*, 2518-2521.

347 Möller, M., Lyburner, L., & Volk, M. (2007). The comparison index: A tool for assessing the accuracy
348 of image segmentation. *International Journal of Applied Earth Observation and Geoinformation,*
349 *9*, 311-321.

350 Pesaresi, M., & Benediktsson, J.A. (2001). A new approach for the morphological segmentation of high-
351 resolution satellite imagery. *IEEE Transactions on Geoscience and Remote Sensing, 39*, 309-
352 320.

353 Ryherd, S., & Woodcock, C. (1996). Combining spectral and texture data in the segmentation of
354 remotely sensed images. *Photogrammetric Engineering and Remote Sensing, 62*, 181-194.

355 Shandley, J., Franklin, J., & White, T. (1996). Testing the Woodcock-Harward image segmentation
356 algorithm in an area of southern California chaparral and woodland vegetation. *International*
357 *Journal of Remote Sensing*, 17, 983-1004.

358 Vincent, L., & Soille, P. (1991). Watersheds in digital spaces: an efficient algorithm based on immersion
359 simulations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13, 583-598.

360 Weidner, U. (2008). Contribution to the assessment of segmentation quality for remote sensing
361 applications. *International Archives of Photogrammetry, Remote Sensing and Spatial*
362 *Information Sciences*, 37, 479-484.

363 Witharana, C., & Civco, D.L. (2014). Optimizing multi-resolution segmentation scale using empirical
364 methods: Exploring the sensitivity of the supervised discrepancy measure Euclidean distance 2
365 (ED2). *ISPRS Journal of Photogrammetry and Remote Sensing*, 87, 108-121.

366 Yang, J., Li, P., & He, Y. (2014). A multi-band approach to unsupervised scale parameter selection for
367 multi-scale image segmentation. *ISPRS Journal of Photogrammetry and Remote Sensing*,
368 Accepted.

369 Zhan, Q., Molenaar, M., Tempfli, K., & Shi, W. (2005). Quality assessment for geo - spatial objects
370 derived from remotely sensed data. *International Journal of Remote Sensing*, 26, 2953-2974.

371 Zhang, H., Fritts, J.E., & Goldman, S.A. (2008). Image segmentation evaluation: A survey of
372 unsupervised methods. *Computer Vision and Image Understanding*, 110, 260-280.

373 Zhang, Q., Pavlic, G., Chen, W., Fraser, R., Leblanc, S., & Cihlar, J. (2005). A semi-automatic
374 segmentation procedure for feature extraction in remotely sensed imagery. *Computers &*
375 *Geosciences*, 31, 289-296.

376 Zhang, Y.J. (1996). A survey on evaluation methods for image segmentation. *Pattern Recognition*, 29,
377 1335-1346.

