ISPRS Journal of Photogrammetry and Remote Sensing 66 (2011) 473–483

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

journal homepage: www.elsevier.com/locate/isprsjprs

# Unsupervised image segmentation evaluation and refinement using a multi-scale approach 采用多尺度方法对无监督图像分割进行评价和细化

Brian Johnson *, Zhixiao Xie

*Department of Geosciences, Florida Atlantic University, 777 Glades Rd., Boca Raton, FL 33431, United States*

ARTICLE INFO

*Article history:*
Received 25 April 2010
Received in revised form
6 December 2010
Accepted 10 February 2011
Available online 5 March 2011

*Keywords:*
Object-based image analysis
Multi-scale segmentation
Image segmentation evaluation
Under-segmentation
Over-segmentation

ABSTRACT 全局内切和块内异质性度量，加权方差和莫兰指数

In this study, a multi-scale approach is used to improve the segmentation of a high spatial resolution (30 cm) color infrared image of a residential area. First, a series of 25 image segmentations are performed in Definiens Professional 5 using different scale parameters. The optimal image segmentation is identified using an unsupervised evaluation method of segmentation quality that takes into account global intra-segment and inter-segment heterogeneity measures (weighted variance and Moran's I, respectively). Once the optimal segmentation is determined, under-segmented and over-segmented regions in this segmentation are identified using local heterogeneity measures (variance and Local Moran's I). The under- and over-segmented regions are refined by (1) further segmenting under-segmented regions at finer scales, and (2) merging over-segmented regions with spectrally similar neighbors. This process leads to the creation of several segmentations consisting of segments generated at three different segmentation scales. Comparison of single- and multi-scale segmentations shows that identifying and refining under- and over-segmented regions using local statistics can improve global segmentation results.

一旦确定了最优分割，利用局部异质性度量(方差和Local Moran's I)识别该分割中的欠分割和过分割区域)。 通过（1） 在更精细的尺度上进一步分割欠分割区域，(2) 将过分割区域与光谱相似的邻域合并，对欠分割区域和过分割区域进行了细化。参考：过分割区域合并，欠分割区域继续分割。这一过程导致创建几个分段，由在三个不同的分段尺度上生成的分段组成。 单尺度和多尺度分割的比较表明，利用局部统计识别和细化欠分割和过分割区域可以提高全局分割结果。



## 1. Introduction

High spatial resolution Earth observation imagery obtained from satellite (IKONOS, Quickbird, GeoEye-1, WorldView-2, etc.) and airborne sensors has become increasingly available in recent years. However, traditional pixel-based image analysis methods, which extract information using statistics of single pixels, are often not well suited for this high-resolution imagery (Blaschke and Strobl, 2001; Benz et al., 2004). This is due to the fact that a single pixel often represents only a small part of a classification target (e.g. tree canopy, building rooftop, road) in high-resolution images. When image classification is performed, the high degree of spectral variability found within land cover classes caused by shadows, gaps in tree canopy, etc., reduces the statistical separability between classes, resulting in low classification accuracy (Yu et al., 2006). 基于像素->基于对象

One alternative that has been able to outperform the pixel-based approach to image analysis is the "object-based" approach (Thomas et al., 2003; Blaschke et al., 2004; Yu et al., 2006). One key step in object-based image analysis (OBIA) is to segment an image into relatively homogeneous regions called segments or "image objects" (Benz et al., 2004). The statistics for these image objects are then used for image analysis (e.g. image classification) rather than the statistics of individual pixels. Using image objects as the basic analysis units has a number of benefits, including the reduction of within-class spectral variability and the ability to incorporate spatial and contextual information such as size, shape, texture, and topological relationships (e.g. containment and adjacency) (Blaschke et al., 2004; Benz et al., 2004). For additional information regarding object-based image analysis, readers are encouraged to refer to Blaschke (2010).

In most image segmentation software packages, users need to set one or more parameters that influence the average size of segments produced by the segmentation. Changing these parameters will change the size of segments produced by the segmentation (i.e. the segmentation scale), allowing for an image to be segmented at many different scales. Some studies have involved the use of a single segmentation scale (Chabrier et al., 2006; Espindola et al., 2006; Yu et al., 2006; Moller et al., 2007; Neubert et al., 2008; Kim et al., 2008, 2009). Other studies have employed a multi-scale approach, creating a hierarchy of several single-scale segmentations (Hay et al., 2003; Blaschke et al., 2004; Trias-Sanz et al., 2008; Xie et al., 2008; Zhou and Troy, 2009). 解决多尺度中最佳尺度选择问题

For remote sensing studies, this ability to choose segmentation scale(s) is useful but challenging (Dorren et al., 2003; Kim et al., 2009). One difficulty is that, prior to performing a segmentation, there is no way to really know what parameters will produce a good result (i.e. segments that closely resemble ground features). A segmentation performed using one set of parameters may produce good results for a homogeneous scene, but not work well for a more

* Corresponding author. Tel.: +1 561 297 3250; fax: +1 561 297 2745.
*E-mail addresses:* bjohns53@fau.edu, bjohns53@hotmail.com (B. Johnson).

0924-2716/$ – see front matter. Published by Elsevier B.V. on behalf of International Society for Photogrammetry and Remote Sensing, Inc. (ISPRS).
doi:10.1016/j.isprsjprs.2011.02.006

无监督评价 涉及使用一些质量指标/准则对多个图像分割进行评分和排序，这些标准通常与人类对什么是良好分割的感知一致(Chabrier等人，2006年)。



heterogeneous scene of the same spatial resolution. A second challenge is that, since landscapes typically consist of different types of land cover that vary in size (e.g. trees, roads, and buildings), it may not be possible to properly segment all features in a scene using a single segmentation scale, so over-segmentation (producing too many segments) or under-segmentation (producing too few segments) often occurs. In a recent study, Liu and Xia (2010) found that classification accuracy decreased considerably when over- or under-segmentation occurred(especially under-segmentation) because (a) the spatial, textual, and contextual information extracted from over-segmented objects was not very useful for classification, and (b) under-segmented objects held little value since they contained more than one land cover class. These issues related to over- and under-segmentation are the main disadvantages of using only one segmentation scale for remote sensing images, and why using a multi-scale approach may often be preferable. The main difficulty with the multi-scale approach is that the segmentation scales are typically selected based on extensive knowledge of the study area, and are more or less subjective.

Since segmentation quality has been shown to have an impact on image classification accuracy (Dorren et al., 2003; Kim et al., 2009), research that deals with evaluating image segmentation quality to identify good segmentation parameters has become a topic of interest in remote sensing. Most of these evaluation methods fall into one of three categories: visual, supervised, or unsupervised.

Visual methods involve the user(s) identifying parameters that produce high-quality image segmentations by visually comparing multiple segmentations. This method has been the most widely-used method of assessing image segmentation quality (Zhang et al., 2008). One problem with the visual method is that it can be highly subjective, as different people may have a different idea about what is the best segmentation (Paglieroni, 2004). This process can also be time and labor intensive, since several segmentations need to be inspected in detail to identify the best one(s).

Supervised evaluation methods, also called empirical discrepancy methods, involve comparing multiple image segmentations with a ground truth, or reference digitization that is created by the user for at least part of the image (Zhang, 1997). This comparison involves computing a measure of dissimilarity between the image segmentations and the reference digitization (Chabrier et al., 2006). Dissimilarity measures can be based on differences in color, location, size, or shape (Abeyta and Franklin, 1998; Carleer et al., 2005; Moller et al., 2007; Neubert et al., 2008; Clinton et al., 2010). The segmentation most similar to the reference digitization is determined to be the optimal segmentation. Since different dissimilarity measures can indicate wildly different segmentations as "best", it may be necessary to consider the results of multiple dissimilarity measures when choosing an optimal segmentation (Clinton et al., 2010). Most studies employing supervised methods used them to evaluate the quality of single-scale segmentations (Paglieroni, 2004; Carleer et al., 2005; Chabrier et al., 2006; Moller et al., 2007; Neubert et al., 2008). However, Trias-Sanz et al. (2008) compared different hierarchical, multi-scale segmentations using two reference segmentations, one containing compulsory edges and one with optional edges. The main disadvantage of the supervised approach is that creating a reference digitization can be difficult, subjective, and time-consuming (Zhang et al., 2008). Creating a reference digitization for a large image or for many different images may not be feasible.

Unsupervised evaluation methods, or empirical goodness methods, allow segmentation quality to be assessed quantitatively without the need for a reference digitization or a detailed visual comparison of multiple image segmentations (Chabrier et al., 2006), making it more time and labor efficient and less subjective than the visual and supervised methods. Unsupervised evaluation methods involve scoring and ranking multiple image segmentations using some quality criteria, which are typically established in agreement with human perception of what makes a good segmentation (Chabrier et al., 2006). One widely-accepted definition of what constitutes a good segmentation is one in which: (i) regions are uniform and homogeneous,(ii) regions are significantly different from neighboring regions,(iii) regions have a simple interior without many holes, and (iv) regions have boundaries that are simple, not ragged, and spatially accurate (Haralick and Shapiro, 1985). However, for highly-textured and natural images, only the first two criteria can realistically be applied (Zhang et al., 2008). So, for remote sensing images, a good segmentation should, in theory, maximize intra-segment homogeneity and inter-segment heterogeneity. For this reason, most unsupervised evaluation methods involve calculating intra-segment and inter-segment heterogeneity measurements for each segment, and then aggregating these values into a global value (Chabrier et al., 2006). The global intra- and inter-segment values are then combined to assign an overall "goodness" score to the segmentation (Zhang et al., 2008).

In remote sensing, relatively few studies have incorporated unsupervised segmentation evaluation methods. Stein and De Beurs (2005) used complexity metrics to quantify the semantic accuracy of image segmentations for two Landsat images. Chabrier et al. (2006) tested and compared six different evaluation criteria and three different algorithms for segmenting radar and multi spectral aerial images. Espindola et al. (2006) measured intra-segment homogeneity using the weighted variance of the near infrared (NIR) band and measured intra-segment heterogeneity using a spatial autocorrelation measure, Global Moran's *I*, for the NIR band as well. The two values were normalized and combined to identify optimal segmentation parameters in an urban area. In a similar study, Kim et al. (2008, 2009) computed unweighted variance (NIR band) and Global Moran's *I* (NIR band) and graphed each separately to compare multiple segmentations and identify optimal parameters for segmenting forest stands. The segmentation level where variance began to level off and Moran's *I* was lowest was found to be the optimal segmentation. Radoux and Defourny (2008) used a combination of unsupervised (normalized post-segmentation standard deviation) and supervised (border discrepancy) methods to evaluate segmentation results in a rural area. The unsupervised evaluation methods used in these studies all involved computing global quality scores for image segmentations. However, since some intra- and inter-segment heterogeneity measures (e.g. variance and spatial autocorrelation) can be calculated for individual segments, this information may also be useful for assessing local segmentation quality.

In this study, a global unsupervised segmentation evaluation method similar to that of Espindola et al. (2006) and Kim et al. (2008) is used for scoring and ranking multiple image segmentations of a heterogeneous residential area. We chose to use an unsupervised evaluation method because it does not require a reference digitization or expert knowledge of the landscape, and because intra- and inter-segment heterogeneity information may be useful for assessing local segmentation quality (i.e. local over- and under-segmentation) as well as global segmentation quality. Unlike previous studies, we (1) validate that the unsupervised evaluation method used is a good indicator of segmentation quality by scoring and ranking a reference digitization as well, (2) test the use of three spectral bands (NIR, red, and green) for calculating "goodness" scores, rather than just the NIR band, and (3) refine the optimal single-scale segmentation using local heterogeneity statistics. The optimal single-scale segmentation is refined by (a) further segmenting under-segmented regions at finer scales, and (b) merging over-segmented regions with spectrally similar adjacent regions that are also over-segmented. The proposed methods may be able to overcome some of

参考

参考

方差开始趋于平缓，Moran的I值最低的分割水平被发现是最优分割。

多光谱

->superpixel 此处重心仍然还是放在分割，而不是像合并。本文未使用超像素。

本文未在一开始使用超像素分割，所以不存在一开始就过分割或欠分割的情况，只是普通的分割，所以做法从用方差+莫兰指数的加权准则去识别整张图的过分割区域还是欠分割区域变成全部过分割后，用这个准则去识别、合并过分割过分的地方，至于未在过分割下被过度错分的地区。。。如何保留。但都是为解决多尺度的最优尺度问题，或摆脱多尺度限制下的超像素分割。

(2)如果扩展到高光谱，波段可否特定选择几个，NIR、green和red扩展到更多的有代表性的几个。步骤(3)属于使用局部块的特征来微调最优分割尺度。

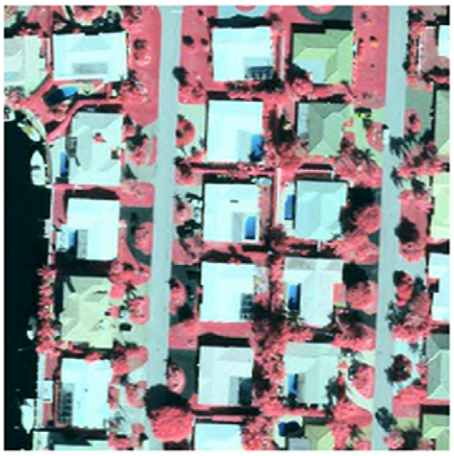


**Fig. 1.** 30 cm resolution color infrared image of the study area (NIR, red, and green spectral bands). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

the limitations of traditional single-scale segmentation methods (many segments are under- or over-segmented) and multi-scale methods (somewhat subjective, reliant on expert knowledge of the study area).

解决问题

## 2. Study area and data

For this study, a 30 cm resolution, color infrared (CIR) digital aerial orthoimage (flight date: Dec 31, 2008) of Broward County, Florida, USA, was obtained from the Broward County Property Appraiser. The digital orthoimage contained information for the NIR, red, and green spectral bands. A 150 m × 150 m (500 × 500 pixel) subset of a residential area, shown in Fig. 1, was chosen as the study area because it contains many different types of land cover within a small area(buildings, light pavement, dark pavement, trees, grass, hedges, shadow, and water). The relatively small size of the scene allowed for a reference digitization, used for testing the unsupervised evaluation method, to be created in a reasonable amount of time.

## 3. Methods

### 3.1. Image segmentation

本文的分割步骤与超像素分割不同

其他分割参数包括：（1）用于分割的波段及其相对权重；（2）颜色/形状和平滑/紧凑权重。

Image segmentation was performed in Definiens Professional 5 using the "Multi-resolution Segmentation" algorithm. The segmentation procedure starts with one-pixel objects and merges similar neighboring objects together in subsequent steps until a heterogeneity threshold, set by a "scale parameter", is reached (Benz et al., 2004). Other segmentation parameters include: (1) which bands to use for the segmentation and their relative weights, and (2) color/shape and smoothness/compactness weights. For additional information regarding the segmentation algorithm, readers are encouraged to refer to Benz et al. (2004).

For this study, all spectral bands were used and given equal weight for image segmentation. A series of segmentations was performed using 25 different scale parameters, ranging from 10 to 250 at an interval of 10, so that the landscape could be analyzed at many different scales. The remaining segmentation parameters (color/shape, smoothness/compactness) were left at their default values. Mean spectral values and variance for each spectral band were calculated for all segments. All of the segmented images were then exported as polygon shape files for further analysis in ArcGIS 9.3. For an overview of the methods used in this study, see Fig. 2.

在本研究中，所有光谱带都被使用，并被赋予相同的权重进行图像分割。利用25个不同尺度的参数进行了一系列的分割，从10到250不等，间隔为10，这样就可以在许多不同的尺度上对景观进行分析。其余的分割参数（颜色/形状、平滑/紧凑）留在它们的默认值。计算了各谱带的平均谱值和方差。

### 3.2. Identifying optimal image segmentation scale

识别最优图像分割尺度

To evaluate segments generated at each scale, global intra- and inter-segment "goodness measures" were calculated. The optimal image segmentation scale was defined as the scale that maximized intra-segment homogeneity and inter-segment heterogeneity. At this optimal scale segments should, on average, be internally uniform and significantly different from their neighbors. These criteria fit with what is generally accepted as a good segmentation for natural images (Zhang et al., 2008).

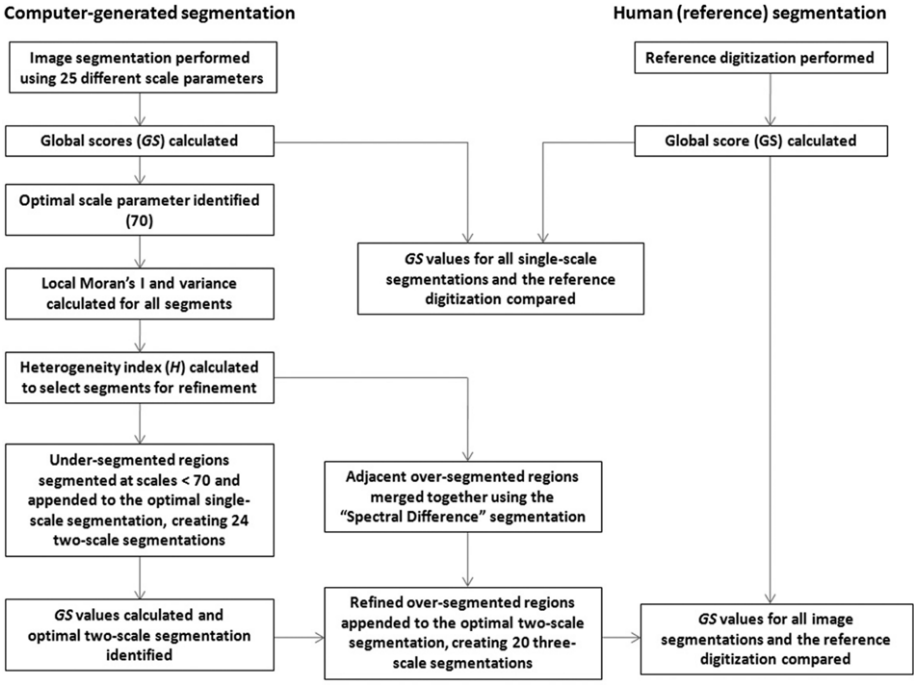



**Fig. 2.** Overview of the methods used in this study.

The global intra-segment goodness measure that we used was variance, weighted by each segment's area, calculated in Eq. (1) as:

块内的goodness度量采用的是方差，由每个块的面积加权

$$w\mathrm{Var} = \frac{\sum_{i=1}^{n} a_i * v_i}{\sum_{i=1}^{n} a_i} \quad (1)$$

低方差->块内更均一

where $v_i$ is the variance and $a_i$ is the area of segment $i$. Variance was chosen as the intra-segment goodness measure because segments with low variance should be relatively homogeneous. Weighted variance ($w$Var) was used for the global calculation so that large segments have more of an impact on global calculations than small ones.

The inter-segment global goodness measure used was Global Moran's $I$, a spatial autocorrelation metric which measures, on average, how similar a region is to its neighbors (Fotheringham et al., 2000). We chose to use Moran's $I$ because it is a reliable indicator of statistical separation between spatial objects (Fotheringham et al., 2000), and was found to be a good indicator of segmentation quality in previous segmentation evaluation studies (Espindola et al., 2006; Kim et al., 2008). For this study, Global Moran's $I(MI)$ was calculated using the formula in Eq. (2):

$$MI = \frac{n\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(y_i-\bar{y})(y_j-\bar{y})}{\sum_{i=1}^{n}(y_i-\bar{y})^2\left(\sum_{i\neq j}\sum w_{ij}\right)} \quad (2)$$

低莫兰指数->相关性低->两个块差异大->更利于图像分割

goodness measures:方差和莫兰指数

where $n$ is the total number of regions, $w_{ij}$ is a measure of the spatial proximity, $y_i$ is the mean spectral value of region $R_i$, and $\bar{y}$ is the mean spectral value of the image. Each weight $w_{ij}$ is a measure of the spatial adjacency of regions $R_i$ and $R_j$. In this study, only adjacent regions (i.e. regions sharing a boundary) were considered for the Moran's $I$ calculation, so if regions $R_i$ and $R_j$ are neighbors, $w_{ij} = 1$, otherwise, $w_{ij} = 0$. Low Moran's $I$ values indicate high inter-segment heterogeneity, which is desirable for an image segmentation.

For all image segmentations, both goodness measures were calculated for each of the three spectral bands. To allow for the intra-segment and inter-segment goodness measures to be considered equally, both were rescaled to a similar range (0–1) using the normalization formula in Eq. (3):

为了使分段内和分段间的goodness度量被平等地考虑，两者都被重新scale到一个相似的范围(0,1)，使用方程中的归一化公式

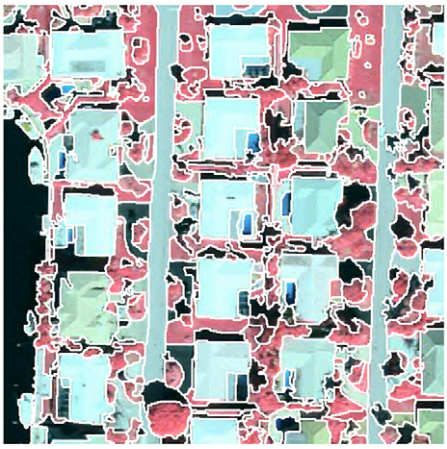
$$(X - X_{\min})/(X_{\max} - X_{\min}) \quad (3)$$

where $X_{\min}$ and $X_{\max}$ are the minimum and maximum value of weighted variance or Moran's $I$ for a spectral band. The normalization equation caused segmentations with low weighted variance or Moran's $I$ to have normalized values relatively close to zero. Since low variance and low spatial autocorrelation are desirable, low normalized values indicate better results. To assign an overall "Global Score" ($GS$) to each image segmentation, the normalized weighted variance and Moran's $I$ values were combined using the formula in Eq. (4).

$$GS = V_{\mathrm{norm}} + MI_{\mathrm{norm}} \quad (4)$$

where $V_{\mathrm{norm}}$ is the normalized weighted variance and $MI_{\mathrm{norm}}$ is the normalized Moran's $I$. This calculation is similar to the method used by Espindola et al. (2006). In Espindola et al. 's study, values were only calculated for the NIR band. However, since there may be useful data in the other bands as well, we calculated the $GS$ values for each spectral band individually, and used the average $GS$ of the three bands [(NIR $GS$ + red $GS$ + green $GS$)/3]to identify the best single-scale segmentation. Using the average $GS$ allowed for all three bands to have an equal influence on the segmentation evaluation. The optimal segmentation was identified as the one with the lowest average $GS$, because at this level there is the lowest combined weighted variance and spatial autocorrelation for the three spectral bands.

高光谱波段如何选择出几个有代表性的进行最优单个尺度切割选择最优



超像素分割：空间上近邻、光谱上相似，而这种分割也可以起到类似的效果。到底哪种更好。

**Fig. 3.** Reference digitization, created in ArcGIS 9.3, overlaid on the study area image. Segments represent individual land cover features (trees, hedges, grass, water, building rooftops, shadow, light pavement, dark pavement, boats). Segment boundaries are shown in white. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

### 3.3. Reference digitization 参考资料数字化

A reference digitization of the entire image was also created manually in ArcGIS 9.3. Pixels in the study area image were assigned to the reference polygon that they had the greatest area inside of, and mean values and weighted variance for all three spectral bands were calculated for each reference polygon. The reference digitization, shown in Fig. 3, was used to test the unsupervised segmentation evaluation method used in this study. In theory, the reference digitization should score very well (low $GS$) since expert knowledge of the study area was required to create it. If the reference digitization does not receive a good score, the evaluation method may not be effective for judging segmentation quality.

### 3.4. Refining under-segmented regions 细化欠分割区域 继续分割

The optimal single-scale segmentation was defined as the scale at which the lowest average $GS$ was achieved. However, even at this optimal scale, upon a visual inspection it was clear that many segments were still under- or over-segmented. To refine the optimal single-scale segmentation by reducing under-segmentation, local intra- and inter-segment heterogeneity statistics were used to identify under-segmented regions and segment them again at finer scales.

Regions considered for further segmentation were those with high intra- and inter-segment heterogeneity. These segments are likely to contain more than one type of land cover because they are heterogeneous internally, and they should not be over-segmented already since they are very different from their neighbors. To identify the regions that needed to be segmented at smaller scales, local intra- and inter-segment heterogeneity statistics were calculated for each segment. Variance was used as the intra-segment heterogeneity statistic, and the inter-segment heterogeneity statistic chosen was Local Moran's $I$ (Anselin, 1995), a decomposed form of Moran's $I$ that measures spatial autocorrelation for each segment, calculated using the formula in Eq. (5):

$$I_i = z_i \sum_{j\neq i} w_{ij} z_j \quad (5)$$

用原先Moran's I的分解形式来测量每个块的空间自相关性。对莫兰指数的小改进

where the $z_i$, $z_j$ are in deviations from the mean. Only neighboring segments were considered for calculations, so $w = 1$ for neighboring segments, and $w = 0$ for all other segments. We chose to use variance and Local Moran's $I$ to measure local intra- and

zi, zj偏离平均值？

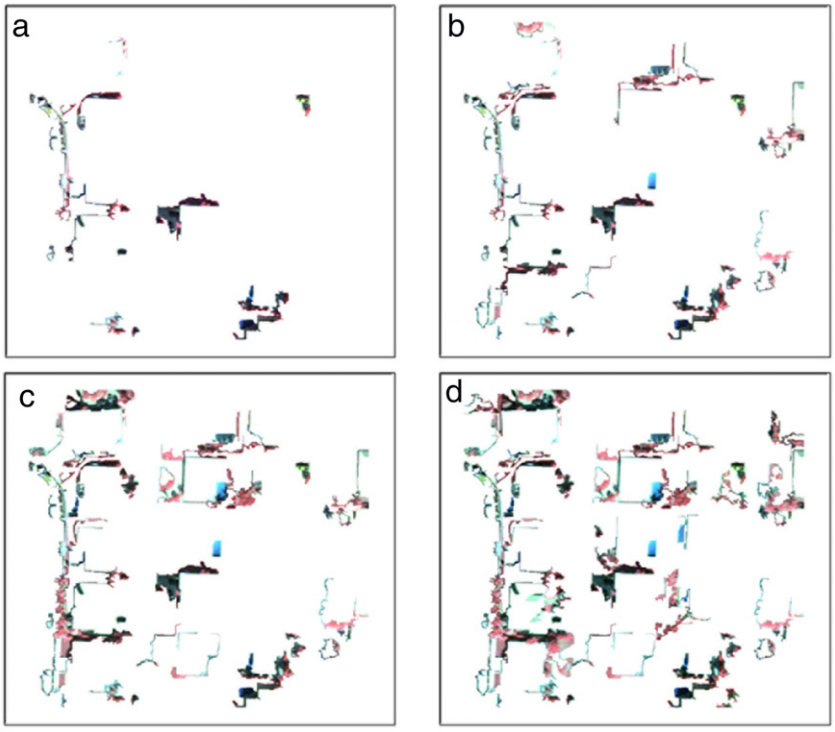



**Fig. 4.** Extracted under-segmented image regions with heterogeneity ($H$) thresholds of 5% (a), 10% (b), 15% (c) and 20% (d).

inter-segment heterogeneity because they were quite similar to the global measures used to calculate the *GS* values.

Variance and Local Moran's $I$ values were normalized for each spectral band using the normalization equation from Eq. (3), and a heterogeneity index ($H$) was created to assign a heterogeneity value to each segment, using the formula in Eq. (6):

$$H = (n\text{Var} - nMI)/(n\text{Var} + nMI) \quad (6)$$

where $n$Var is the normalized variance and $nMI$ represents normalized Moran's $I$ values. 如果高光谱，这里还是简单的求平均吗？

$H$ values range from −1 to 1, with values close to 1 indicating high variance and low Moran's $I$. To combine the information from all three spectral bands, average three-band $H$ values were also calculated [(NIR $H$ + red $H$ + green $H$)/3]. High average $H$ values indicate segments with high overall intra- and inter-segment heterogeneity. Visually inspecting segments with high average $H$ values, we found that they tended to consist of a mixture of land cover with different spectral characteristics, causing them to be heterogeneous internally, and significantly different from their neighbors. Several thresholds were tested for selecting regions with the highest average $H$ values, because it was unknown before-hand which threshold would produce the best results. For this study we tested four different thresholds, the top 5%, 10%, 15%, 20% of segments). Polygon boundaries of these selected segments were used to extract the regions in the image that needed further segmentation. The extracted image regions, shown in Fig. 4, were then re-segmented at finer scales (i.e. using smaller scale parameters) than that which was used for the original segmentation. In this case, the optimal single-scale segmentation scale was achieved using a scale of 70, so the extracted image regions were segmented at scales of 60, 50, 40, 30, 20, and 10. The new segments were then merged back with the remaining segments from the single-scale segmentation using the "Merge" tool in ArcGIS 9.3. As detailed later, this process led to the creation of 24 two-scale image segmentations. *GS* values were calculated for each of the two-scale segmentations in order to compare the scores to those calculated for the single-scale segmentations.

### 3.5. Refining over-segmented regions 细化/调整 过度分割区域

To reduce over-segmentation, average $H$ values were also used to select the least heterogeneous segments from the optimal single-scale segmentation. These segments should be quite homogeneous internally, and very similar to their neighbors. This high degree of similarity both within itself and with neighboring segments indicates the segment and its neighbor(s) are likely to be parts of the same ground feature. For this study, we tested four different $H$ value thresholds for selecting these homogeneous segments (lowest 5%, 10%, 15%, and 20% of segments), and the polygon boundaries for these segments were used to extract the regions in the image that needed to be merged with similar neighbors. These extracted image regions are shown in Fig. 5.

Adjacent over-segmented regions were combined using the "Spectral Difference" algorithm in Definiens Professional 5. This algorithm merges neighboring segments together according to their mean layer intensity values. Segments are merged if the difference in layer mean intensities is below the value specified in the user input "Spectral Difference" parameter, and is used to refine existing segmentation results by merging regions with similar values produced by previous segmentations (Definiens, 2006). For this study, all three spectral bands were given equal weight for the Spectral Difference segmentation, and five different Spectral Difference parameters were tested for merging segments (10, 20, 30, 40, and 50). All neighboring over-segmented regions were merged when the Spectral Difference parameter of 50 was used, so we stopped performing spectral difference segmentations at this point. The refined segments were combined with the remaining segments from the optimal two-scale segmentation using the "Merge" tool in ArcGIS 9.3., which led to the creation of several three-scale image segmentations. *GS* values were calculated for all of the three-scale segmentations and compared with those for the single- and two-scale segmentations to see which segmentation performed best.

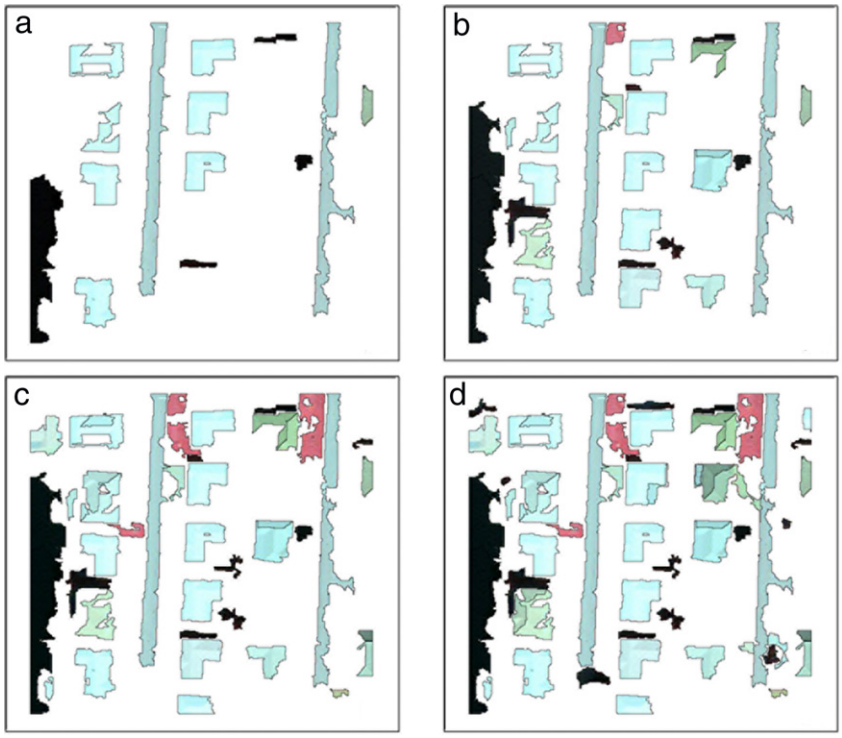



**Fig. 5.** Extracted over-segmented image regions with heterogeneity ($H$) thresholds of 5% (a), 10% (b), 15% (c) and 20% (d). Polygon boundaries are shown as black lines.

## 4. Results and discussion

识别最优图像分割尺度

### 4.1. Identifying the optimal image segmentation scale

As the scale parameter increased for the single-scale image segmentations, weighted variance tended to increase and Global Moran's $I$ tended to decrease for each spectral band, as shown in Table 1. This means that as the average size of segments increased, segments tended to consist of more dissimilar pixels, and became less similar to neighboring segments. These results were similar to results from the Espindola et al. (2006) study, which also observed an increasing trend in global variance, and a generally decreasing trend in Moran's $I$ as the average size of segments increased. However, our findings are somewhat different than those from the study conducted by Kim et al. (2008), which found that variance tended to level off, and Global Moran's $I$ was lowest at the optimal segmentation scale. This may have been because the study area investigated by Kim et al. (2008) was a more homogeneous forest scene, in which the spectral values for the different types of land cover were quite similar. Our study and the study conducted by Espindola et al. (2006) both investigated more heterogeneous urban areas, where different types of land cover can have very different spectral characteristics. In these more heterogeneous environments, variance will likely continue to increase as segmentation scale increases, and Moran's $I$ will continue to decrease until segments become large enough to contain a mixture of many different types of land cover. Once all segments become mixed, spatial autocorrelation should start to increase again. The segmentation scale at which this happens is larger than any of the segmentation scales investigated in this study, and will not occur until the image is significantly under-segmented.

Comparison of $GS$ values for all single-scale segmentations showed that the image segmentation performed using a scale parameter of 70 had the lowest average $GS$(0.623), as shown in Fig. 6. $GS$ values for all three spectral bands were also lowest using the scale parameter of 70, as shown in Table 1. These results

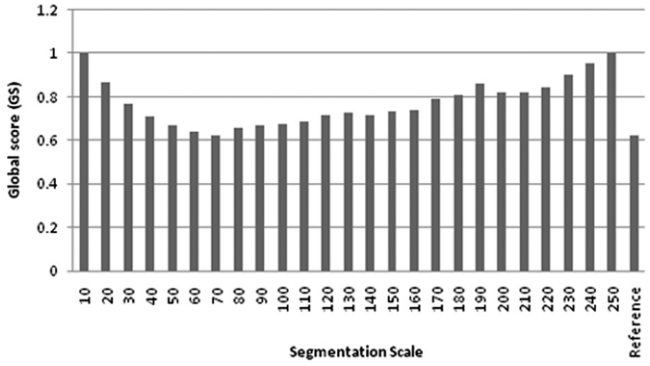



**Fig. 6.** Average $GS$ values for all single-scale segmentations and the reference digitization.

indicate that for the segmentation performed using the scale parameter of 70, segments are, on average, most homogeneous internally and most different from their neighbors. Since that fits with the widely-accepted definition of what constitutes a good segmentation, this segmentation was chosen as the optimal single-scale segmentation.

To assess whether or not the $GS$ values used to evaluate the image segmentations are actually a good indicator of segmentation quality, the $GS$ values calculated for the reference digitization were compared with $GS$ values for the image segmentations. Since manual digitization of an image by professionals is generally accepted to be more accurate than automated segmentations, the reference digitization should have a good score if the $GS$ is an accurate indicator of segmentation quality. When we compared the $GS$ values of the individual spectral bands and the three-band average $GS$ values calculated for the image segmentations with the reference digitization's $GS$ values, we found that for the NIR band, the reference digitization performed better than all of the image segmentations. This suggests that the $GS$ values for the NIR band were a good indicator of segmentation quality. The average three-band $GS$[(NIR $GS$ + green $GS$ + red $GS$)/3] of

在超像素的切割尺度上，会与如下的尺度定义不同。(相反)



**Table 1**
Weighted variance ($w_{Var}$), Moran's $I$($MI$), normalized variance ($V_{norm}$), normalized Moran's $I$($MI_{norm}$), and global scores ($GS$) for all single-scale segmentations and the reference digitization. Values for the optimal single-scale segmentation and the reference digitization are highlighted.

| Scale parameter | NIR band | | | | | Red band | | | | | Green band | | | | | Three-band average |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $w_{Var}$ | $MI$ | $V_{norm}$ | $MI_{norm}$ | $GS$ | $w_{Var}$ | $MI$ | $V_{norm}$ | $MI_{norm}$ | $GS$ | $w_{Var}$ | $MI$ | $V_{norm}$ | $MI_{norm}$ | $GS$ | $GS$ |
| **10** | 98.713 | 0.705 | 0.000 | 1.000 | 1.000 | 69.898 | 0.762 | 0.000 | 1.000 | 1.000 | 68.798 | 0.747 | 0.000 | 1.000 | 1.000 | 1.000 |
| **20** | 211.643 | 0.498 | 0.070 | 0.781 | 0.850 | 148.139 | 0.567 | 0.056 | 0.819 | 0.875 | 143.012 | 0.552 | 0.060 | 0.816 | 0.876 | 0.867 |
| **30** | 310.434 | 0.324 | 0.131 | 0.596 | 0.727 | 218.239 | 0.412 | 0.107 | 0.675 | 0.782 | 204.452 | 0.399 | 0.110 | 0.671 | 0.781 | 0.763 |
| **40** | 410.837 | 0.215 | 0.192 | 0.481 | 0.674 | 281.147 | 0.296 | 0.152 | 0.568 | 0.720 | 262.678 | 0.289 | 0.157 | 0.567 | 0.724 | 0.706 |
| **50** | 500.507 | 0.138 | 0.248 | 0.400 | 0.647 | 336.302 | 0.200 | 0.191 | 0.479 | 0.671 | 314.315 | 0.195 | 0.198 | 0.478 | 0.677 | 0.665 |
| **60** | 586.076 | 0.079 | 0.301 | 0.337 | 0.637 | 398.686 | 0.118 | 0.236 | 0.403 | 0.639 | 364.964 | 0.115 | 0.239 | 0.403 | 0.642 | 0.640 |
| **70** | 659.185 | 0.010 | 0.346 | 0.264 | 0.610 | 461.527 | 0.058 | 0.281 | 0.347 | 0.629 | 419.729 | 0.054 | 0.284 | 0.346 | 0.629 | 0.623 |
| **80** | 746.789 | 0.012 | 0.400 | 0.266 | 0.665 | 508.834 | 0.039 | 0.315 | 0.330 | 0.645 | 463.459 | 0.041 | 0.319 | 0.333 | 0.652 | 0.654 |
| **90** | 813.454 | −0.012 | 0.441 | 0.241 | 0.682 | 552.744 | 0.009 | 0.347 | 0.303 | 0.649 | 507.833 | 0.014 | 0.355 | 0.307 | 0.662 | 0.664 |
| **100** | 879.371 | −0.046 | 0.481 | 0.205 | 0.687 | 622.840 | −0.028 | 0.397 | 0.268 | 0.665 | 567.522 | −0.025 | 0.403 | 0.270 | 0.673 | 0.675 |
| **110** | 935.261 | −0.064 | 0.516 | 0.185 | 0.701 | 681.994 | −0.064 | 0.440 | 0.235 | 0.675 | 616.680 | −0.057 | 0.443 | 0.240 | 0.683 | 0.686 |
| **120** | 1021.536 | −0.077 | 0.569 | 0.172 | 0.741 | 740.208 | −0.090 | 0.481 | 0.211 | 0.692 | 680.152 | −0.081 | 0.494 | 0.217 | 0.711 | 0.715 |
| **130** | 1071.944 | −0.104 | 0.600 | 0.144 | 0.744 | 778.241 | −0.101 | 0.509 | 0.200 | 0.709 | 711.913 | −0.097 | 0.520 | 0.203 | 0.722 | 0.725 |
| **140** | 1108.385 | −0.127 | 0.623 | 0.119 | 0.741 | 810.165 | −0.143 | 0.532 | 0.161 | 0.693 | 739.866 | −0.136 | 0.542 | 0.166 | 0.708 | 0.714 |
| **150** | 1171.997 | −0.151 | 0.662 | 0.094 | 0.756 | 866.219 | −0.166 | 0.572 | 0.140 | 0.712 | 788.004 | −0.156 | 0.581 | 0.146 | 0.728 | 0.732 |
| **160** | 1208.923 | −0.168 | 0.685 | 0.076 | 0.760 | 916.948 | −0.192 | 0.608 | 0.116 | 0.724 | 832.606 | −0.179 | 0.617 | 0.125 | 0.742 | 0.742 |
| **170** | 1275.985 | −0.178 | 0.726 | 0.065 | 0.791 | 1001.896 | −0.196 | 0.669 | 0.112 | 0.781 | 912.089 | −0.179 | 0.681 | 0.125 | 0.806 | 0.793 |
| **180** | 1278.761 | −0.183 | 0.728 | 0.059 | 0.787 | 1038.764 | −0.196 | 0.696 | 0.112 | 0.808 | 945.101 | −0.177 | 0.708 | 0.127 | 0.835 | 0.810 |
| **190** | 1426.239 | −0.167 | 0.819 | 0.077 | 0.895 | 1089.976 | −0.206 | 0.733 | 0.103 | 0.836 | 997.106 | −0.187 | 0.750 | 0.118 | 0.868 | 0.866 |
| **200** | 1436.872 | −0.211 | 0.825 | 0.030 | 0.855 | 1118.203 | −0.262 | 0.753 | 0.051 | 0.804 | 1015.723 | −0.253 | 0.765 | 0.055 | 0.820 | 0.826 |
| **210** | 1474.846 | −0.223 | 0.849 | 0.018 | 0.866 | 1138.979 | −0.295 | 0.768 | 0.021 | 0.789 | 1039.350 | −0.279 | 0.784 | 0.030 | 0.814 | 0.823 |
| **220** | 1542.992 | −0.231 | 0.891 | 0.009 | 0.899 | 1183.407 | −0.306 | 0.800 | 0.010 | 0.810 | 1075.274 | −0.292 | 0.813 | 0.018 | 0.831 | 0.847 |
| **230** | 1667.168 | −0.228 | 0.967 | 0.012 | 0.979 | 1258.599 | −0.316 | 0.854 | 0.001 | 0.855 | 1145.084 | −0.301 | 0.870 | 0.010 | 0.879 | 0.904 |
| **240** | 1720.435 | −0.227 | 0.974 | 0.013 | 0.987 | 1352.043 | −0.311 | 0.921 | 0.005 | 0.926 | 1217.196 | −0.297 | 0.928 | 0.013 | 0.941 | 0.960 |
| **250** | 1763.400 | −0.239 | 1.000 | 0.000 | 1.000 | 1462.211 | −0.317 | 1.000 | 0.000 | 1.000 | 1306.295 | −0.311 | 1.000 | 0.000 | 1.000 | 1.000 |
| *Reference* | 817.138 | −0.167 | 0.443 | 0.076 | 0.519 | 778.468 | −0.148 | 0.509 | 0.156 | 0.665 | 743.669 | −0.149 | 0.545 | 0.153 | 0.698 | 0.628 |

此10-250的25个尺度不是分割的块的数量

而是average size

两尺度结果>单尺度结果 多尺度

0.623 was very close to the reference $GS$ value of 0.628, which indicates a generally good result as well. However, for the red and green bands, the reference digitization had significantly higher $GS$ values than many of the image segmentations, as shown in Table 1. To identify some possible reasons why this occurred, we investigated the statistics for several of the polygons from the reference digitization. We found that for the red and green spectral bands, building rooftops, which made up many of the larger segments in the reference digitization, tended to have higher variance in spectral values for the red and green bands than for the NIR band. This occurred because shaded pixels within the rooftops had much lower brightness values than non-shaded pixels for the red and greed bands, while for NIR band the difference in brightness was not as great. For this reason, the weighted variance calculated was relatively high for both the red and green bands in the reference digitization, and affected the $GS$ values for these two bands. Further investigation of the red and green spectral bands showed that different types of vegetation (trees, grass, hedges) had very similar spectral values. Since different types of vegetation were all segmented separately in the reference digitization, there was a higher degree of spatial autocorrelation between neighboring vegetation segments in the red and green bands of the reference digitization, which also affected the $GS$ values calculated for these two bands. According to these findings, in future studies it may be better to use just the NIR band for the segmentation evaluation, or at least weight it more heavily. For this study, we continue to report the average three-band $GS$ values, but also report $GS$ values for individual spectral bands as well.

### 4.2. Refined under-segmented regions 细化欠分割区域

Of the two-scale segmentations, the best results were achieved when 20% of segments (56 out of the 280 total segments) with the highest $H$ values from the optimal single-scale segmentation were re-segmented at a scale of 50. Results for each of the two-scale segmentations are shown Table 2. The average $GS$ value for the optimal two-scale segmentation was 0.595, which was better than the optimal single-scale value of 0.623. In fact, the average $GS$ values for all three spectral bands from the optimal two-scale segmentation were better than those from the optimal single-scale segmentation. Visually comparing the optimal single- and two-scale segmentations, it is also clear that the two-scale method was better able to segment smaller objects of interest (e.g. small trees, hedges) without causing additional over-segmentation of larger objects, such as buildings, as shown in Fig. 7.

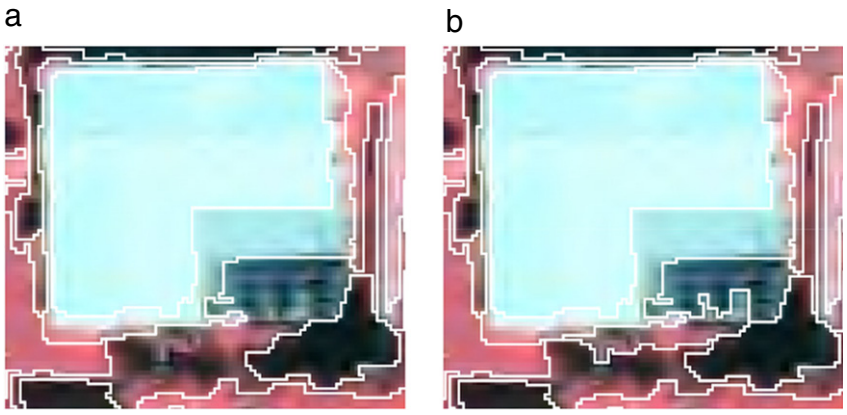
The average $GS$ of the optimal two-scale segmentation (0.595) was also lower than that of the reference digitization (0.628). However, when $GS$ values for each spectral band were investigated individually, we found that the NIR $GS$ for the reference digitization (0.519) was still much better than that of the two-scale digitization (0.578), but the $GS$ values of the other two bands caused the reference segmentation to have a higher average $GS$. Since the reference digitization is expected to outperform the computer-generated segmentations, this again suggests that for future studies it may be better to use the NIR band alone for the purpose of segmentation evaluation, or at least weigh it more heavily than the green and red bands. For our study, since the segmentation with the lowest average $GS$ value was also the one with the lowest $GS$ value for the NIR band, using the average $GS$ instead of using that of the NIR band alone did not affect our results.

### 4.3. Refined over-segmented regions 细化/合并过分割区域

For the three-scale segmentations, the lowest average $GS$ was achieved when the least heterogeneous 20% of segments (56 segments) from the single-scale segmentation were put through a region merging process in Definiens Professional 5 using a "Spectral Difference" parameter of 30 or 40 (both produced identical results). In fact, the $GS$ values for all three spectral bands were lowest for this segmentation. Results for all of the three-scale segmentations are shown in Table 3. The procedure for combining the three different groups of segments (original segments, refined

**Table 2**
*GS* values for all two-scale segmentations. Under-segmented regions from the optimal single-scale segmentation were segmented again at finer scales. The optimal two-scale segmentation is highlighted. Reference and optimal single-scale segmentation statistics are also shown and highlighted to allow for comparison. Best results were obtained when 20% of segments from the single-scale segmentation were further segmented using a scale parameter of 50.

| % of segments selected for further segmentation | Scale parameter | NIR band global score (*GS*) | Red band global score (*GS*) | Green band global score (*GS*) | Three-band average global score (*GS*) |
|---|---|---|---|---|---|
| **5**% | **10** | 0.920 | 0.8445 | 0.852 | 0.872 |
| | **20** | 0.721 | 0.665 | 0.674 | 0.687 |
| | **30** | 0.668 | 0.641 | 0.649 | 0.653 |
| | **40** | 0.637 | 0.623 | 0.629 | 0.630 |
| | **50** | 0.624 | 0.621 | 0.625 | 0.623 |
| | **60** | 0.606 | 0.623 | 0.623 | 0.618 |
| **10**% | **10** | 0.941 | 0.873 | 0.881 | 0.898 |
| | **20** | 0.723 | 0.677 | 0.687 | 0.695 |
| | **30** | 0.649 | 0.633 | 0.629 | 0.637 |
| | **40** | 0.624 | 0.612 | 0.620 | 0.619 |
| | **50** | 0.607 | 0.606 | 0.613 | 0.609 |
| | **60** | 0.597 | 0.613 | 0.616 | 0.609 |
| **15**% | **10** | 0.976 | 0.900 | 0.905 | 0.927 |
| | **20** | 0.743 | 0.706 | 0.713 | 0.720 |
| | **30** | 0.646 | 0.645 | 0.652 | 0.647 |
| | **40** | 0.614 | 0.612 | 0.620 | 0.615 |
| | **50** | 0.601 | 0.608 | 0.614 | 0.608 |
| | **60** | 0.587 | 0.619 | 0.621 | 0.609 |
| **20**% | **10** | 0.961 | 0.890 | 0.897 | 0.916 |
| | **20** | 0.734 | 0.705 | 0.710 | 0.717 |
| | **30** | 0.623 | 0.637 | 0.643 | 0.634 |
| | **40** | 0.595 | 0.601 | 0.611 | 0.602 |
| | **50** | 0.578 | 0.600 | 0.608 | 0.595 |
| | **60** | 0.582 | 0.618 | 0.623 | 0.607 |
| **0**% (*Single-scale segmentations*) | **70** | 0.610 | 0.629 | 0.629 | 0.623 |
| | *Reference* | 0.519 | 0.665 | 0.698 | 0.628 |




两尺度的分割效果优于单尺度

**Fig. 7.** Subset of the image showing the optimal single-scale segmentation (a) and the optimal two-scale segmentation (b). Polygon boundaries are shown in white. Using multiple scales allowed for vegetated and dark pavement regions to be better segmented without causing the building to be over-segmented. The building shown here is the top, center building in the study area image shown in Fig. 1.

under-segmented regions, refined over-segmented regions) to create the final three-scale segmentations is shown in Fig. 8. The average *GS* value for the optimal three-scale segmentation was 0.574, which was significantly better than both the optimal single-scale (0.623) and two-scale (0.595) scores. When the *GS* values for each spectral band in the optimal three-scale segmentation were inspected individually, all three were found to have scores better than those than in the optimal single-scale and two-scale segmentations. Visual inspection of region boundaries before and after the Spectral Difference segmentation, shown in Fig. 9, also indicates an improvement in segmentation quality. The *GS* values for the three-scale segmentation were better than those of the reference digitization for all but the NIR spectral band, which had a *GS* value of 0.568 compared to the 0.519 score of the reference digitization. So, once again, the *GS* values for the red and green bands caused the reference digitization's average *GS* to be higher than that of the three-scale segmentation, and the *GS* for the NIR band was the best indicator of segmentation quality.

### 4.4. Importance of input data in segmentation

光谱i+纹理作为输入

We found the spectral information from the digital orthoimagery to be useful for image segmentation. However, it may also be beneficial to incorporate non-spectral information, such as image texture, as input layers for segmentation (Ryherd and Woodcock, 1996). In recent studies, Lucieer and Stein (2005) used texture images created from a LiDAR digital surface model as input for segmentation in order to identify coastal landform objects, and

**Table 3**
*GS* values for all three-scale segmentations. Neighboring over-segmented regions from the optimal single-scale segmentation were merged using the "Spectral Difference" algorithm in Definiens Professional. Results for the optimal three-scale segmentation are highlighted. There were no adjacent segments when the 5% *H* threshold was used, so no Spectral Difference segmentation was performed for this threshold. We stopped performing the Spectral Difference segmentation once all neighboring segments had been merged.

| % of Segments selected for further region merging | "Spectral difference" parameter | NIR band global score (*GS*) | Red band global score (*GS*) | Green band global score (*GS*) | Three-band average global score (*GS*) |
|---|---|---|---|---|---|
| **10**% | **10** | 0.579 | 0.595 | 0.603 | 0.592 |
| | **20** | 0.576 | 0.590 | 0.596 | 0.587 |
| **15**% | **10** | 0.579 | 0.595 | 0.603 | 0.592 |
| | **20** | 0.575 | 0.587 | 0.594 | 0.585 |
| | **30** | 0.573 | 0.583 | 0.592 | 0.583 |
| **20**% | **10** | 0.577 | 0.593 | 0.603 | 0.591 |
| | **20** | 0.573 | 0.582 | 0.591 | 0.582 |
| | **30** | 0.568 | 0.572 | 0.584 | 0.574 |
| | **40** | 0.568 | 0.572 | 0.584 | 0.574 |
| | **50** | 0.571 | 0.578 | 0.594 | 0.581 |

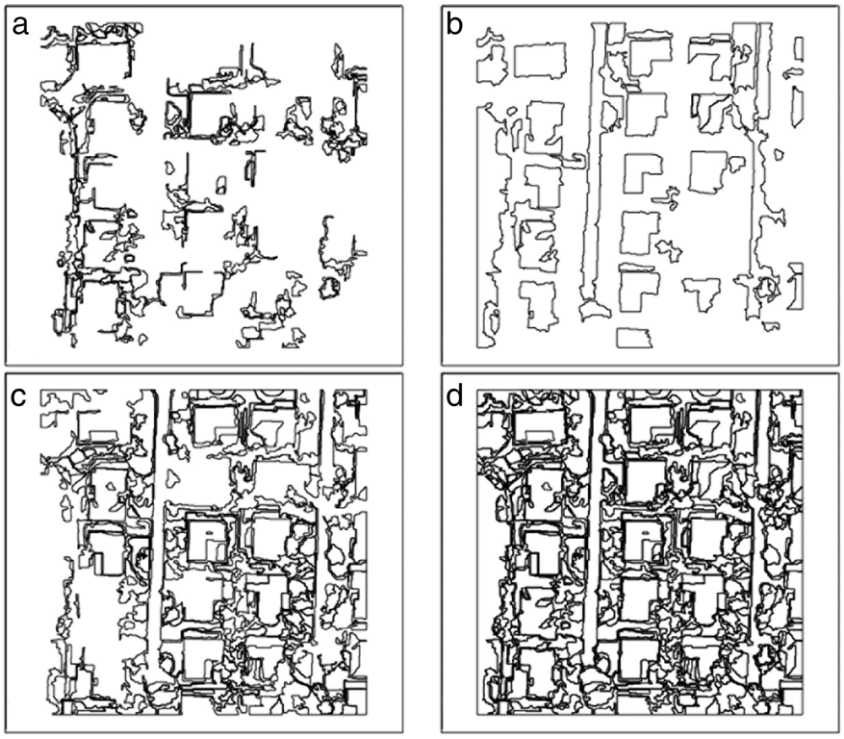



**Fig. 8.** Refined under-segmented regions (a) and over-segmented regions (b) merged with the remaining regions from the optimal single-scale segmentation using the "Merge" tool in ArcGIS 9.3 (c). The result is a segmented image consisting of segments generated at three different scales (d).

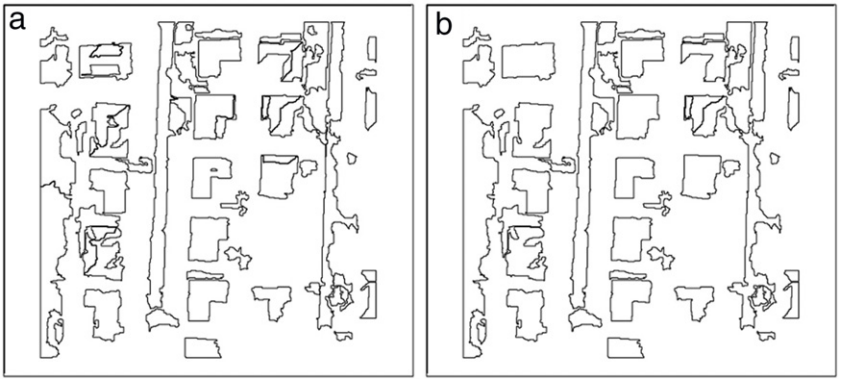



**Fig. 9.** Over-segmented regions, defined in this case as the 20% of segments from the optimal single-scale segmentation with the lowest heterogeneity (*H*) values, (a) and the same regions after refinement using a "Spectral Difference" parameter of 30 or 40 (b). Canal (left-bottom polygon) and building rooftops better represent real-world objects in the refined segmentation.

Kim et al. (2010) found that incorporating texture and topographic variables for segmentation in addition to spectral information led to higher accuracy for vegetation classification than the use of spectral information alone. For future research, it may be interesting to incorporate additional non-spectral information as input for image segmentation to see if segmentation quality improves.

## 5. Conclusion

In this study, an unsupervised image segmentation evaluation method similar to those used in other recent studies (Espindola et al., 2006; Kim et al., 2008, 2009) was employed to score and rank the quality of several image segmentations of a residential area. We also created and computed *GS* values for a reference digitization in order to validate that this segmentation evaluation method is a good indicator of image segmentation quality. Since the reference digitization received a good score (low *GS*), we deemed this evaluation method a reliable indicator of segmentation quality. We tested the use of all three spectral bands for the segmentation evaluation, but found that the NIR band was most useful for evaluation purposes. In our study, using the three-band average *GS* value rather than just using the NIR *GS* value alone did not really have an impact on which segmentations were chosen as the optimal segmentations, since the single-, two-, and three-scale segmentations with the lowest average *GS* values also had the lowest *GS* values for the NIR band. However, for future studies, it may be more useful to use the NIR band alone for segmentation evaluation, or to increase its weight in calculations.

考虑到这些局部统计数据的异质性指数(H)能够识别在最优单尺度图像分割中被欠分割和过分割的区域。

We also found that refining an image segmentation using local statistics (Local Moran's *I* and variance) can improve global segmentation quality. A heterogeneity index (*H*), which took into account these local statistics, was able to identify regions that were under-segmented and over-segmented in the optimal single-scale image segmentation. These regions were refined by segmenting under-segmented regions at finer scales and merging over-segmented regions with spectrally similar neighbors that were also over-segmented. The methods used in this study were able to overcome some of the limitations of traditional single-scale segmentation approaches (many features are under- or over-segmented) and multi-scale methods (scales chosen are somewhat subjective and reliant on expert knowledge of the study area).

We selected *H* thresholds ranging from 5%–20% of segments to be used for refining under- and over-segmented regions, but results may continue to improve if additional thresholds are considered. The final segmented image was created by combining segments generated using three different segmentation scales, but the methods from this study can be used progressively to incorporate additional scales by further segmenting/merging the under- and over-segmented regions in the optimal three-scale segmentation, and so on. Using additional scale parameters, modifying other segmentation parameters (depending on the software used for segmentation), and incorporating non-spectral information may improve segmentation quality as well.

In future studies, we will test how well the methods developed work in a more homogeneous natural area, and a larger urban area containing additional types of land use and land cover. In addition, the ultimate goal of improving image segmentation quality is typically to obtain a more accurate image classification. In this paper, we focused only on improving segmentation quality. We worked under the assumption that, as past studies have found (Dorren et al., 2003; Kim et al., 2009), improvement in segmentation quality will result in higher classification accuracy because higher quality segments have more accurate statistics (mean spectral values, size, shape, etc.). However, in the future we plan to compare image classification accuracy before and after segmentation refinement in order to quantify the impact that segmentation refinement has on classification accuracy.

## Acknowledgements

We would like to thank the Broward County Property Appraiser's office for providing us with the imagery of our study area, and the anonymous reviewers for their detailed and constructive comments.

## References

Abeyta, A., Franklin, J., 1998. The accuracy of vegetation stand boundaries derived from image segmentation in a desert environment. Photogrammetric Engineering & Remote Sensing 64 (1), 59–66.
Anselin, L., 1995. Local indicators of spatial association: LISA. Geographical Analysis 27, 93–115.
Benz, U.C., Hofmann, P., Willhauck, G., Lingenfelder, I., Heynen, M., 2004. Multiresolution, object-oriented fuzzy analysis of remote sensing data for GIS-ready information. ISPRS Journal of Photogrammetry and Remote Sensing 58 (3–4), 239–258.
Blaschke, T., 2010. Object based image analysis for remote sensing. ISPRS Journal of Photogrammetry and Remote Sensing 65 (1), 2–16.
Blaschke, T., Burnett, C., Pekkarinen, A., 2004. New contextual approaches using image segmentation for object-based classification. In: De Meer, F., de Jong, S. (Eds.), Remote Sensing Image Analysis: Including the Spatial Domain. Kluwer Academic Publishers, Dordrecht, pp. 211–236.
Blaschke, T., Strobl, J., 2001. What's wrong with pixels? Some recent developments interfacing remote sensing and GIS. GIS—Zeitschrift für Geoinformations Systeme 14 (6), 12–17.
Carleer, A., Debeir, O., Wolff, E., 2005. Assessment of very high spatial resolution satellite image segmentations. Photogrammetric Engineering & Remote Sensing 71 (11), 1285–1294.
Chabrier, S., Emile, B., Rosenberger, C., Laurent, H., 2006. Unsupervised performance evaluation of image segmentation. EURASIP Journal on Applied Signal Processing 1–12.
Clinton, N., Holt, A., Scarborough, J., Gong, P., 2010. Accuracy assessment measures for object based image segmentation goodness. Photogrammetric Engineering & Remote Sensing 76 (3), 289–299.
Definiens, , 2006. Definiens Professional 5 Reference Book. Definiens AG, München, Germany.
Dorren, L., Maier, B., Seijmonsbergen, A., 2003. Improved Landsat-based forest mapping in steep mountainous terrain using object-based classification. Forest Ecology and Management 183 (1–3), 31–46.
Espindola, G., Camara, G., Reis, I., Bins, L., Monteiro, A., 2006. Parameter selection for region growing image segmentation algorithms using spatial autocorrelation. International Journal of Remote Sensing 27 (14), 3035–3040. 模仿的文章
Fotheringham, A., Brunsdon, C., Charlton, M., 2000. Quantitative Geography: Perspectives on Spatial Analysis. Sage Publications, California.
Haralick, R., Shapiro, L., 1985. Image segmentation techniques. Computer Vision, Graphics, and Image Processing 29 (1), 100–132.
Hay, G.J., Blaschke, T., Marceau, D.J., Bouchard, A., 2003. A comparison of three image-object methods for the multiscale analysis of landscape structure. ISPRS Journal of Photogrammetry and Remote Sensing 57 (5–6), 327–345.
Kim, M., Madden, M., Warner, T., 2008. Estimation of optimal image object size for the segmentation of forest stands with multispectral IKONOS imagery. In: Blaschke, T., Lang, S., Hay, G. (Eds.), Object-Based Image Analysis, Spatial Concepts for Knowledge-Driven Remote Sensing Applications. Springer, Heidelberg, Berlin, New York, pp. 291–307. 模仿的文章
Kim, M., Madden, M., Warner, T., 2009. Forest type mapping using object-specific texture measures from multispectral IKONOS imagery: segmentation quality and image classification issues. Photogrammetric Engineering & Remote Sensing 75 (7), 819–830.
Kim, M., Madden, M., Xu, B., 2010. GEOBIA vegetation mapping in Great Smoky Mountains National park with spectral and non-spectral ancillary information. Photogrammetric Engineering & Remote Sensing 76 (2), 137–149.
Liu, D., Xia, F., 2010. Assessing object-based classification: advantages and limitations. Remote Sensing Letters 1 (4), 187–194.
Lucieer, A., Stein, A., 2005. Texture-based landform segmentation of LiDAR imagery. International Journal of Applied Earth Observation and Geoinformation 6 (3–4), 261–270.
Moller, M., Lymburner, L., Volk, M., 2007. The comparison index: a tool for assessing the accuracy of image segmentation. International Journal of Applied Earth Observation and Geoinformation 9 (3), 311–321.
Neubert, M., Herold, H., Meinel, G., 2008. Assessing image segmentation quality—concepts, methods and application. In: Blaschke, T., Lang, S., Hay, G. (Eds.), Object-Based Image Analysis, Spatial Concepts for Knowledge-Driven Remote Sensing Applications. Springer, Heidelberg, Berlin, New York, pp. 769–784.
Paglieroni, D., 2004. Design considerations for image segmentation quality assessment measures. Pattern Recognition 37 (8), 1607–1617.
Radoux, J., Defourny, P., 2008. Quality assessment results devoted to object-based classification. In: Blaschke, T., Lang, S., Hay, G. (Eds.), Object-Based Image Analysis, Spatial Concepts for Knowledge-Driven Remote Sensing Applications. Springer, Heidelberg, Berlin, New York, pp. 255–271.
Ryherd, S., Woodcock, C., 1996. Combining spectral and texture data in the segmentation of remotely sensed images. Photogrammetric Engineering & Remote Sensing 62 (2), 181–194.

Stein, A., De Beurs, K., 2005. Complexity metrics to quantify semantic accuracy in segmented Landsat images. International Journal of Remote Sensing 26 (14), 2937–2951.

Thomas, N., Hendrix, C., Congalton, R.G., 2003. A comparison of urban mapping methods using high-resolution digital imagery. Photogrammetric Engineering& Remote Sensing 69 (9), 963–972.

Trias-Sanz, R., Stamon, G., Louchet, J., 2008. Using colour, texture, and hierarchical segmentation for high-resolution remote sensing. ISPRS Journal of Photogrammetry and Remote Sensing 63 (2), 156–168.

Xie, Z., Roberts, C., Johnson, B., 2008. Object-based target search using remotely sensed data: a case study in detecting invasive exotic Australian Pine in south Florida. ISPRS Journal of Photogrammetry and Remote Sensing 63 (6), 647–660.

Yu, Q., Gong, P., Chinton, N., Biging, G., Kelly, M., Schirokauer, D., 2006. Object-based detailed vegetation classification with airborne high spatial resolution remote sensing imagery. Photogrammetric Engineering & Remote Sensing 72 (7), 799–811.

Zhang, Y., 1997. Evaluation and comparison of different segmentation algorithms. Pattern Recognition Letters 18 (10), 963–974.

Zhang, H., Fritts, J., Goldman, S., 2008. Image segmentation evaluation: a survey of unsupervised methods. Computer Vision and Image Understanding 110 (2), 260–280.

Zhou, W., Troy, A., 2009. Development of an object-based framework for classifying and inventorying human-dominated forest ecosystems. International Journal of Remote Sensing 30 (23), 6343–6360.