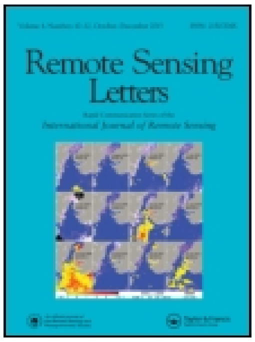

Remote Sensing Letters

ISSN: 2150-704X (Print) 2150-7058 (Online) Journal homepage: http://www.tandfonline.com/loi/trsl20

BASI: a new index to extract built-up areas from high-resolution remote sensing images by visual attention model

Zhenfeng Shao, Yingjie Tian & Xiaole Shen

Remote Sensing Letters, 2014
Vol. 5, No. 4, 305–314, http://dx.doi.org/10.1080/2150704X.2014.889861

BASI: a new index to extract built-up areas from high-resolution remote sensing images by visual attention model

Zhenfeng Shao, Yingjie Tian, and Xiaole Shen*

State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan, China

(*Received 3 October 2013; accepted 25 January 2014*)

A built-up areas saliency index (BASI) to extract built-up areas is proposed in this article. The proposed method is designed to be especially effective for dealing with high-resolution remote sensing images and complex scenes. Due to the complexity of built-up areas in high-resolution images, visual attention model based on textural feature is applied for the calculation of BASI. To highlight built-up areas in complex scenes, we present an improved signal processing method to describe the textural feature of built-up areas, which is used as the low-layer feature of the visual attention model. Comparison studies and experimental results demonstrate the accuracy and robustness of BASI for discrimination between built-up and non-built-up areas from satellite and aerial high-resolution remote sensing data.

1. Introduction

The information of built-up areas is critical for the analysis of land planning, disaster evaluation, urban monitoring and demographic census. For the purpose of high efficiency and low cost, built-up areas extraction based on remote sensing images is becoming an important method for urban research.

Many of built-up areas extraction methods to date have been proposed based on index, for example normalized difference built-up index (NDBI) (Zha, Gao, and Ni 2003), index-based built-up index (IBI) (Xu 2008) and improved NDBI differencing algorithm (Varshney 2013). Those methods utilize only spectral information to extract built-up areas from remote sensing images. However, the methods based on spectral information are preferably suitable for medium or coarse spatial resolution images but less effective when dealing with high-resolution images.

With the advancement of the spatial resolution of remote sensing images, the detail information of ground objects is richer, and the spectrum varies drastically. The phenomena of same object with different spectrums and same spectrum for different objects prevail. Meanwhile, the spectral and structural complexities of buildings will be further enhanced. For high-resolution remote sensing images, some indexes have also been proposed to extract built-up areas. These researches bring in texture, edge or structure characteristics to improve the performance of built-up areas extraction. Pesaresi, Gerhardinger, and Kayitakire (2008) proposed texture-derived built-up presence index (PanTex). Built-up areas extraction was accomplished based on image textural feature as a result of salient textural feature in high-resolution remote sensing images. Grey-level co-occurrence matrix (GLCM) was first employed to compute multi-directional texture, and then built-up areas were extracted by PanTex, which

*Corresponding author. Email: shenxiaole@whu.edu.cn

came from result based on fuzzy rule-based composition of the multi-directional texture. Later, they proposed an improved PanTex which NDVI was exploited to make PanTex enhanced, and extraction results were optimized by morphological filtering (Pesaresi and Gerhardinger 2011). However, PanTex is appropriate for remote sensing images with 5 m resolution, for example the images of Satellite Pour l'Observation de la Terre-5 (SPOT-5) satellite, rather than higher resolution (Pesaresi, Gerhardinger, and Kayitakire 2008).

The improving spatial resolution of remote sensing images makes it more difficult to extract built-up areas because of their spatial complexity and spectral confusion. To overcome this issue, the visual attention mechanism was brought in the process of built-up areas extraction in this article. The most important function of visual attention mechanism is to rapidly and effectively locate interesting regions and targets from the complex visual scene (Itti and Koch 2001). Currently, there has been wide use of Region of Interest (ROI) extracting applications based on visual attention mechanism (Gao, Bi, and Yang 2011; Hu et al. 2013; Wang et al. 2013; Zhang et al. 2013). Hu et al. (2013) proposed a method for detection of salient structure textures and objects in remote sensing images based on saliency index. This method generated saliency index based on the edge density and spatial evenness and detected salient structure textures and objects by segmenting the saliency map.

For a long time, the mainstream visual attention model focused on psychology and neurophysiology. For example, Itti's model (Itti, Koch, and Niebur 1998) uses the characteristics of colour, intensity and orientations, forming the central-surround difference through spatial scale created by Gaussian pyramid, and finally generates the visual saliency map by normalization and linear combinations. Currently, many studies have shown that human perception system is influenced by the further processing of statistical characteristics of surrounding natural signals. From the perspective of information theory, the distribution of information contained in natural environment is uneven. However, the visual attention process can quickly locate the area with more information (Kadir and Brady 2001). In other words, the area which contains more information arouses more visual attention. At present, information theory-based visual attention model has been proposed and verified on biological plausibility by experiments (Lee and Yu 2000; Kadir and Brady 2001; Fritz et al. 2004; Bruce and Tsotsos 2005, 2009). The key point of visual attention model based on information theory is the expression of information. And the conclusion has been drawn that self-information is more suitable to the visual attention process (Bruce and Tsotsos 2005, 2009). In this article, the visual attention mechanism is described from the perspective of information theory, and the self-information is used to calculate the saliency of the built-up areas and finally generates a saliency map. The saliency map can be seen as an indicator of built-up areas.

2. Identification of built-up areas based on texture-derived visual attention model

Although built-up areas in high-resolution remote sensing images are complex, the textural characteristics of built-up areas are remarkable and make built-up areas attract visual attention. We deem that built-up areas have characteristics as follows: (1) Built-up areas are clustered in remote sensing images. (2) Buildings in the same built-up areas usually have similar structures and shapes. (3) Shades are always next to buildings. It results in periodical bright–dark changes. (4) Buildings are usually located regularly and have relatively similar directions, so the texture of built-up areas has noteworthy directional feature compared with other land covers. Thus the textural feature of built-up areas is a prominent feature and causes the built-up areas to be visual salient object in high-resolution remote sensing images.

In this article, built-up areas are regarded as an interesting target of the visual search task. And the visual attention model is used to quickly find out the interesting target from the complex visual scene. Visual attention can be regarded as a selection mechanism which processes a rich stream of visual data to orientate rapidly towards objects of interest by reducing the amount of erroneous visual data (Borji and Itti 2010). And the function of visual attention model is to simulate and explain the visual attention behaviours from computational perspective. According to the information theory-based visual attention model, visual attention selects the most informative parts of a scene and discards the rest by way of localized saliency computation. In this article, the localized saliency computation quantifies the information of visual features by means of the self-information (Bruce and Tsotsos 2005).

Due to the outstanding texture of built-up areas in high-resolution remote sensing images, the textural feature of built-up areas is used as the major low-layer feature of the visual attention model. Textural feature is one of the most important information of remote sensing images and often employed for the segmentation, classification and information extraction approaches. It is difficult to define texture mathematically in spite of its perceptual knowledge. In practice, texture is measured according to its applications. Currently there are four types of texture measurements: statistical methods, structural methods, model-based methods and signal processing methods (Zhang and Tan 2002). Physiologically, human visual system analyses textural image by two steps. Image is first decomposed by frequency and direction. Then image features can be extracted by selective filtering. The signal processing methods are aiming to imitate human visual process of texture analysis. They employ filter approaches to generate subband for different frequencies and directions, and then local energy function is used for the integration and smoothing of subbands to get textural features of image (Randen and Husoy 1999). In this article, a signal processing method is used for identifying textural features of built-up areas. And the nonsampled contourlet transform (NSCT) (Da Cunha, Zhou, and Do 2006) is chosen as the filter in the procedure of the signal processing method. This is because that the result of the NSCT is a multi-scale, multi-direction and shift-invariant image decomposition, which is appropriate for describing the characteristic of built-up areas.

3. Built-up areas saliency index

On the basis of the visual attention model, we propose a new built-up areas saliency index (BASI). In the calculation, the textural feature vectors of built-up areas are obtained first. We utilize NSCT to extract raw textural feature of remote sensing images and further analyse the textural features of built-up areas. Finally we compute BASI based on the visual attention model which is defined by self-information.

The calculation of BASI is as follows:

Step 1: textural feature separation

NSCT is applied to generate multi-scale and multi-direction subbands $f_{s,d}$, where s is the scale and d is the direction.

Step 2: calculation of local textural energy

Local textural energy is calculated for each subband. The window size is set as $(2n+1)\times(2n+1)$, $E_{s,d}$ is the obtained raw textural feature.

$$e_{s,d}(x,y) = \frac{1}{(2n+1)^2} \sum_{i=x-n}^{x+n} \sum_{j=y-n}^{y+n} \left| f_{s,d}(i,j) \times g(i,j) \right| \tag{1}$$

$$E_{s,d}(x,y) = \frac{1}{(2n+1)^2} \sum_{i=x-n}^{x+n} \sum_{j=y-n}^{y+n} \left| \left| f_{s,d}(i,j) \times g(i,j) \right| - e_{s,d}(x,y) \right|^2 \tag{2}$$

where (x,y) represents the pixel position, $g(i,j)$ is a Gaussian mask, $g(i,j) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(i-x)^2+(j-y)^2}{2\sigma^2}}$, where σ is the standard deviation.

Step 3: built-up areas feature enhancement

Buildings are usually located regularly and have relatively similar directions. So the texture of built-up areas possesses noteworthy directional feature compared with other land covers. Intuitively, the texture signals of built-up areas are both strong in the vertical directions. Therefore the textural feature vector $\boldsymbol{T}$ is proposed to emphasize the texture which has strong textural features in the vertical directions. The length of vector $\boldsymbol{T}$ is $(S \times D)/2$, where S is the total number of scales and D is the total number of directions. Vector $\boldsymbol{T}$ is composed of elements $t_{s,d}$, and the element $t_{s,d}$ can be obtained by the following equation:

$$t_{s,d}(x,y) = E_{s,d}(x,y) \times E_{s,d^{\perp}}(x,y) \tag{3}$$

Where s is the scale, d the direction and $d^{\perp}$ the direction perpendicular to d.

The process of Steps 1–3 is shown in Figure 1.

Step 4: independent components analysis

In this work, the independent components analysis (ICA) is applied to textural feature vector T based on two considerations: (1) Dimensionality reduction can reduce redundant information. (2) It is very difficult to estimate the joint probability of multidimensional T, however, if the components are independent, the joint probability equals the product of their probabilities.

The result of applying ICA to $\boldsymbol{T}$ is represented by feature vector $\boldsymbol{V}(\boldsymbol{V} = (v_1, v_2, v_3, \cdots))$.

Step 5: likelihood estimation

When dealing with images, $\boldsymbol{w}$ represents the feature detected from image and is treated as data points in the spatial and textural feature domain, which can be expressed by $\boldsymbol{w} = (i, j, \boldsymbol{V})$. (i,j) represents the pixel position, and $\boldsymbol{V}$ is obtained by Step 4. $P(\boldsymbol{w})$ is the joint probability of w in a defined local neighbourhood. The neighbourhood of position (i,j) is defined as $D_{(i,j)} = \{(x,y) | x \in (i-R, i+R), y \in (j-R, j+R)\}$, where R is the radius of the local neighbourhood. It is indicated that with the higher similarity between $\boldsymbol{w} = (i, j, \boldsymbol{V})$ and the feature of the others points in $D_{(i,j)}$, $P(\boldsymbol{w})$ will increase, and vice versa. There are many methods for evaluating $P(w)$, such as nonparametric and histogram density estimate (Bruce 2005; Bruce and Tsotsos 2005).

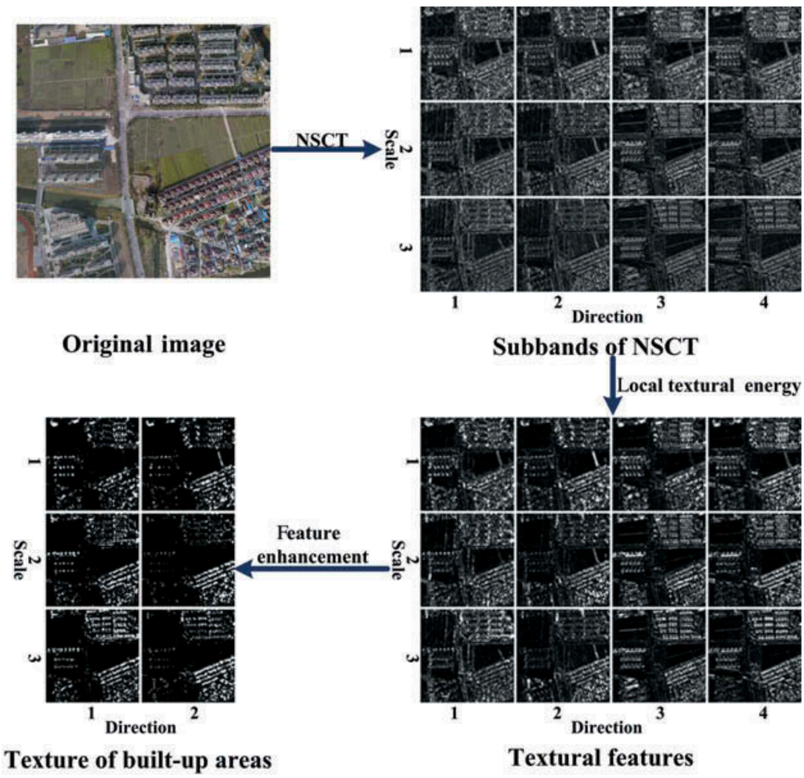

Figure 1. The flow chart of textural measurement for built-up areas.

Because the components of vector $\boldsymbol{V}$ are independent, the joint probability $P(\boldsymbol{w})$ can be estimated easily by Equation (4).

$$P(\boldsymbol{w}) = \prod_{i} P(v_{\mathrm{i}}) \tag{4}$$

Step 6: calculation of self-information

In this article, the saliency is defined based on self-information of image in a local neighbourhood. In a probabilistic event, the self-information $I(\boldsymbol{w})$ depends on the probability $P(\boldsymbol{w})$. When the event indeed occurred, it can be inferred from Equation (5) that the smaller its probability, the larger the self-information.

$$I(\boldsymbol{w}) = -\log(P(\boldsymbol{w})) \tag{5}$$

The value of self-information corresponding to each pixel in a defined neighbourhood is taken as BASI. The BASI describes the presence of built-up areas, and then the position and the range of built-up areas can be obtained by threshold segmentation to the saliency map.

4. Experiments and analysis

4.1 Data set

(1) Aerial images of Taizhou city
Taizhou city, located in Jiangsu Province, China, was selected as the study area to conduct this research. Different land uses, such as urban area, suburban and rural areas, can be found in this study area. Aerial images (spatial resolution: 0.1 m) of Taizhou city were selected as test data to demonstrate the performance of the built-up areas extraction method.
(2) ZY-3 satellite images of Wuhan city
ZY-3 is a Chinese high-resolution imaging satellite launched in January 2012. The experimental data were acquired from ZY-3 in 2012 with a spatial resolution of 2.1 m. The study site is in Wuhan city, located in Hubei Province, China.

4.2 Comparative method

In this article, a series of comparisons are performed between PanTex and BASI. PanTex is also a method using index to extract built-up areas. The index of PanTex is based on fuzzy rule-based composition of anisotropic textural co-occurrence measures by the GLCM. Based on the salient textural feature of built-up areas in high-resolution remote sensing images, the structural characteristics of built-up areas (textural index) were exploited which is calculated using GLCM contrast measures. To calculate the index, the textural indexes in different displacements are standardized and integrated by fuzzy ∩ operator. To validate the effect of two methods in extracting built-up areas, the indices calculated by the proposed method and PanTex were both extracted by Otsu segmentation method (Otsu 1979).

4.3 Experiment I

The built-up areas extraction results detected by PanTex and BASI are shown as Figure 2. Original aerial images, manual delineated built-up areas, the index map and results of built-up areas conducted by PanTex and BASI are shown. Figure 2 presents the extraction results based on three different scenes from the aerial images to validate the effect in various complicated environments. Rows 1, 2 and 3 show three scenes which are urban areas, suburban areas and rural areas, respectively.

Table 1 lists five statistical measures to evaluate the results between PanTex and BASI. The five statistical measures are the user's accuracy of built-up areas, the producer's accuracy of built-up areas, the total accuracy (Congalton 1991), the quantity disagreement and the allocation disagreement (Pontius and Millones 2011). The user's accuracy refers to the probability that a pixel labelled as a certain category in the images actually represents that category on the ground, while the producer's accuracy is indicative of the probability that a certain category of an area on the ground is classified as such. The quantity disagreement is the amount of difference between the reference data and a comparison result that is due to the less than perfect match in the proportions of the categories. The allocation disagreement is the amount of difference between the reference data and a comparison result that is due to the less than optimal match in the spatial allocation of the categories.

As shown in the table, the user's accuracy of BASI above 83.77% is higher than the user's accuracy of PanTex obviously. The producer's accuracy of BASI above 93.26% is consistent with this characteristic. Compared with PanTex, the improvement of total accuracy achieved by

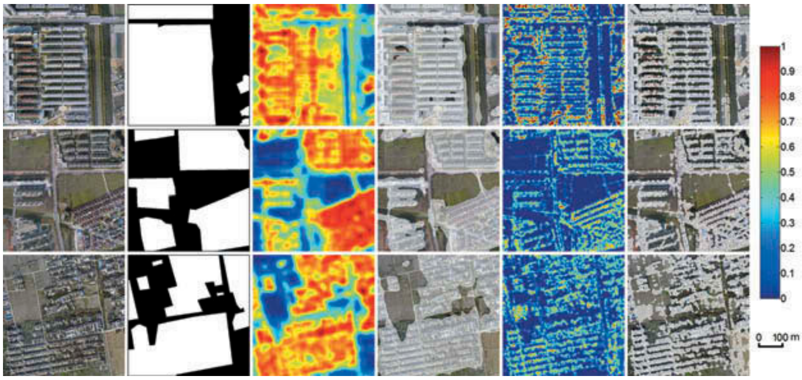

Figure 2. Extraction results of different scenes.

Table 1. Accuracy of BASI and PanTex.

		User's accuracy (%)	Producer's accuracy (%)	Total accuracy (%)	Quantity disagreement (%)	Allocation disagreement (%)
Urban	BASI	93.14	93.26	90.64	8.96	0.47
	PanTex	81.66	34.07	49.37	10.53	40.10
Suburban	BASI	83.77	95.13	87.64	5.49	6.33
	PanTex	77.57	45.37	64.07	13.92	22.01
Rural	BASI	90.40	98.61	92.51	1.71	5.72
	PanTex	75.96	47.07	57.17	18.81	24.02

BASI are 41.27%, 23.57%, and 35.34% for the three images. Moreover, allocation disagreement and quantity disagreement of BASI are both less than the disagreement of PanTex for the three images. Quantity disagreement of BASI is less than 8.96% for the three images, and the allocation disagreement of BASI is less than 6.33% for the three images.

4.4 Experiment II

To further validate the universality of the proposed method, 32 remote sensing images from ZY-3 and aerial data sets with different resolutions and scene complexities were selected. Figure 3 shows the reference images and the extraction results of 32 images. Figure 3(*a*) represents the reference ground truths obtained by manual delineation, Figure 3(*b*) represents the extraction results by BASI, and Figure 3(*c*) represents the extraction results by PanTex. Figure 4 shows the quantity disagreement and the allocation disagreement for extraction results of BASI and PanTex. The two components of disagreement are stacked to show the total disagreement. And the information about total accuracy is also conveyed, since the sum of the total accuracy and total disagreement is 100%.

It is apparent from Figure 4 that, for all 32 images, the total disagreements of BASI are smaller than those of PanTex. The average quantity disagreement of the 32 images by BASI is 9.99%, which is below the average quantity disagreement by PanTex (14.07%). The average allocation disagreement of the 32 images by BASI is 2.54%, which is below the average allocation disagreement by PanTex (22.37%). Every allocation disagreement of the 32 images by BASI is smaller than 8.96%, whereas the allocation disagreement by PanTex ranges from 7.44% to 42.00%.

Figure 3. (*a*) The reference ground truths of 32 images. (*b*) Extraction results of 32 images by BASI. (*c*) Extraction results of 32 images by PanTex.

The improvement of accuracy benefits from the multi-scale texture expression used in BASI. However, the GLCM employed in PanTex define texture by distribution of two pixel values which have some certain spatial relationships between each other. Because there are much more details in high-resolution images, built-up area is not the only land type which has obvious textural feature in these images. So it is not an ideal way to define

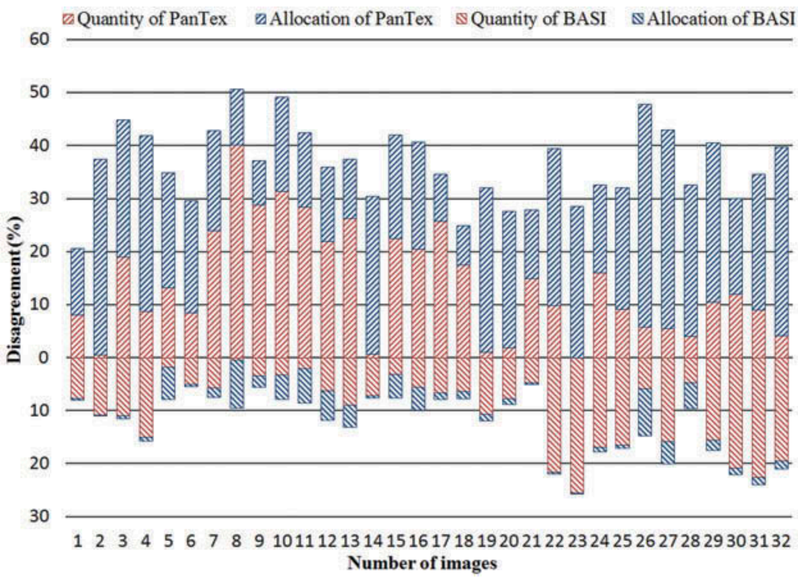

Figure 4. Quantity disagreement and allocation disagreement for BASI and PanTex.

the textural feature of built-up areas by pixels-pair in high-resolution images. The texture measurement of PanTex has some limitation in multi-scale texture expression. The texture expression method used in BASI is also a multi-direction method. The multi-direction texture expressions are employed to highlight the built-up areas' characteristics in texture, and some land types are eliminated such as road, which only has a single texture direction, farmland and water which have even textural feature in different directions. For these reasons, BASI is an ideal method for built-up areas extraction in complex environment.

5. Conclusion and future work

A new built-up areas extraction index based on visual attention model is introduced in this article for high-resolution remote sensing images. The key point of this method is to construct saliency map of built-up areas in the use of multi-scale and multi-direction textural features and visual attention mechanism. Then threshold segmentation on the saliency map is followed, and building regions are extracted. BASI simulates the human visual attention process and constructs saliency map based on information theory. The textural feature is very important to reflect the characteristics of built-up areas in high-resolution remote sensing images. The experiments show that BASI has better performance than PanTex in complex scenes.

In some special situations that buildings are in dispersive distribution, performance of BASI need to improve in some sense. The fusion of multi features such as spectrum, shape, corner may be considered in further research in the hope of promotion of the extraction ability.

Funding

This work was supported by the National Basic Research Program of China [grant number 2010CB731800]; National Science & Technology Specific Projects [grant numbers 2012YQ16018505, 2013BAH42F03]; National Natural Science Foundation of China [grant number 61172174]; Program for New Century Excellent Talents in University [grant number NCET-12-0426].

References

Borji, A., and L. Itti. 2010. “State-of-the-Art in Visual Attention Modeling.” *IEEE Transactions on Pattern Analysis and Machine Intelligence* 35: 185–207.

Bruce, N., and J. K. Tsotsos. 2009. “Saliency, Attention, and Visual Search: An Information Theoretic Approach.” *Journal of Vision* 9: 1–24.

Bruce, N. D. B. 2005. “Features that Draw Visual Attention: An Information Theoretic Perspective.” *Neurocomputing* 65–66: 125–133.

Bruce, N. D. B., and J. K. Tsotsos. 2005. “Saliency Based on Information Maximization.” In *Nineteenth Annual Conference on Neural Information Processing Systems*, edited by Y. Weiss, B. Schölkopf, and J. C. Platt, Whistler, BC, December 5–10, 155–162. Cambridge, MA: MIT Press.

Congalton, R. G. 1991. “A Review of Assessing the Accuracy of Classifications of Remotely Sensed Data.” *Remote Sensing of Environment* 37: 35–46.

Da Cunha, A. L., J. P. Zhou, and M. N. Do. 2006. “The Nonsubsampled Contourlet Transform: Theory, Design, and Applications.” *IEEE Transactions on Image Processing* 15: 3089–3101.

Fritz, G., C. Seifert, L. Paletta, and H. Bischof. 2004. “Attentive Object Detection Using an Information Theoretic Saliency Measure.” In *Second International Workshop on Attention and Performance in Computational Vision*, edited by L. Paletta, J. K. Tsotsos, E. Rome, and G. Humphreys, Prague, May 15, 136–143. Berlin: Springer.

Gao, L. N., F. K. Bi, and J. A. Yang 2011. “Visual Attention Based Model for Target Detection in Large-Field Images.” *Journal of Systems Engineering and Electronics* 22: 150–156.

Hu, X. Y., J. J. Shen, J. Shan, and L. Pan. 2013. “Local Edge Distributions for Detection of Salient Structure Textures and Objects.” *IEEE Geoscience and Remote Sensing Letters* 10: 466–470.

Itti, L., and C. Koch. 2001. “Computational Modelling of Visual Attention.” *Nature Reviews. Neuroscience* 2: 194–203.

Itti, L., C. Koch, and E. Niebur. 1998. “A Model of Saliency-Based Visual Attention for Rapid Scene Analysis.” *IEEE Transactions on Pattern Analysis and Machine Intelligence* 20: 1254–1259.

Kadir, T., and M. Brady. 2001. “Saliency, Scale and Image Description.” *International Journal of Computer Vision* 45: 83–105.

Lee, T. S., and S. X. Yu. 2000. “An Information-Theoretic Framework for Understanding Saccadic Eye Movements.” In *13th Annual Conference on Neural Information Processing Systems*, edited by S. A. Solla, T. K. Leen, and K. Müller, Denver, CO, November 29–December 4, 1999. 834–840. Cambridge, MA: MIT Press.

Otsu, N. 1979. “A Threshold Selection Method from Gray Level Histogram.” *IEEE Transactions on System, Man and Cybernetics* SMC-9: 62–66.

Pesaresi, M., and A. Gerhardinger. 2011. “Improved Textural Built-up Presence Index for Automatic Recognition of Human Settlements in Arid Regions with Scattered Vegetation.” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 4: 16–26.

Pesaresi, M., A. Gerhardinger, and F. Kayitakire. 2008. “A Robust Built-up Area Presence Index by Anisotropic Rotation-Invariant Textural Measure.” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 1: 180–192.

Pontius Jr, R. G., and M. Millones. 2011. “Death to Kappa: Birth of Quantity Disagreement and Allocation Disagreement for Accuracy Assessment.” *International Journal of Remote Sensing* 32: 4407–4429.

Randen, T., and J. H. Husoy. 1999. “Filtering for Texture Classification: A Comparative Study.” *IEEE Transactions on Pattern Analysis and Machine Intelligence* 21: 291–310.

Varshney, A. 2013. “Improved NDBI Differencing Algorithm for Built-up Regions Change Detection from Remote-Sensing Data: An Automated Approach.” *Remote Sensing Letters* 4: 504–512.

Wang, X., Q. Lv, B. Wang, and L. M. Zhang. 2013. “Airport Detection in Remote Sensing Images: A Method Based on Saliency Map.” *Cognitive Neurodynamics* 7: 143–154.

Xu, H. 2008. “A New Index for Delineating Built-up Land Features in Satellite Imagery.” *International Journal of Remote Sensing* 29: 4269–4276.

Zha, Y., J. Gao, and S. Ni. 2003. “Use of Normalized Difference Built-up Index in Automatically Mapping Urban Areas from TM Imagery.” *International Journal of Remote Sensing* 24: 583–594.

Zhang, J. G., and T. N. Tan 2002. “Brief Review of Invariant Texture Analysis Methods.” *Pattern Recognition* 35: 735–747.

Zhang, L. B., H. Li, P. F. Wang, and X. C. Yu. 2013. “Detection of Regions of Interest in a High-Spatial-Resolution Remote Sensing Image Based on an Adaptive Spatial Subsampling Visual Attention Model.” *Giscience & Remote Sensing* 50: 112–132.