



## Article

# A New Stereo Pair Disparity Index (SPDI) for Detecting Built-Up Areas from High-Resolution Stereo Imagery

Feifei Peng <sup>1,2,\*</sup>, Jianya Gong <sup>3,4</sup>, Le Wang <sup>5</sup>, Huayi Wu <sup>4</sup> and Pengcheng Liu <sup>1,2</sup>

<sup>1</sup> Key Laboratory for Geographical Process Analysis & Simulation of Hubei Province, Central China Normal University, Wuhan 430079, China; liupc@mail.ccnu.edu.cn

<sup>2</sup> College of Urban and Environmental Sciences, Central China Normal University, Wuhan 430079, China

<sup>3</sup> School of Remote Sensing and Information Engineering, Wuhan University, Wuhan 430079, China; gongjy@whu.edu.cn

<sup>4</sup> State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing (LIESMARS), Wuhan University, Wuhan 430079, China; wuhuayi@whu.edu.cn

<sup>5</sup> Department of Geography, The State University of New York, Buffalo, NY 14261, USA; lewang@buffalo.edu

\* Correspondence: feifpeng@mail.ccnu.edu.cn; Tel.: +86-27-6786-1441

Received: 30 March 2017; Accepted: 15 June 2017; Published: 20 June 2017

**Abstract:** Within-class spectral variation and between-class spectral confusion in remotely sensed imagery degrades the performance of built-up area detection when using planar texture, shape, and spectral features. Terrain slopes and building heights extracted from auxiliary data, such as Digital Surface Models (DSMs) however, can improve the results. Stereo imagery incorporates height information unlike single remotely sensed images. In this study, a new Stereo Pair Disparity Index (SPDI) for indicating built-up areas is calculated from stereo-extracted disparity information. Further, a new method of detecting built-up areas from stereo pairs is proposed based on the SPDI, using disparity information to establish the relationship between two images of a stereo pair. As shown in the experimental results for two stereo pairs covering different scenes with diverse urban settings, the SPDI effectively differentiates between built-up and non-built-up areas. Our proposed method achieves higher accuracy built-up area results from stereo images than the traditional method for single images, and two other widely-applied DSM-based methods for stereo images. Our approach is suitable for spaceborne and airborne stereo pairs and triplets. Our research introduces a new effective height feature (SPDI) for detecting built-up areas from stereo imagery with no need for DSMs.

**Keywords:** built-up area; stereo image; disparity; height feature; Stereo Pair Disparity Index (SPDI)

## 1. Introduction

Accurately identifying built-up areas is an essential task for government agencies facing the complex and ever changing demands of city planning [1–3]. Built-up areas are dynamic environments, and thus must be periodically monitored [4]. Built-up areas are usually obtained from remotely sensed imagery using planar texture, shape, and spectral features [5–7], such as the Pantex [8,9]. Within-class spectral variation and between-class spectral confusion in remotely sensed imagery however, degrades built-up detection performance [10]; nevertheless, performance can be increased by using height information from Light Detection and Ranging (LiDAR) data or stereo imagery. Stereo imagery incorporates both planar and height information. There is no need for auxiliary height data (e.g., LiDAR data) when detecting built-up areas from stereo imagery. Moreover, stereo imagery is widely available. It is routinely collected by many satellites with stereo observation capabilities [11], such as the IKONOS, WorldView-1~4, QuickBird, GeoEye-1, Cartosat-1/2, Pleiades-HR, ALOS-PRISM, and ZY-3 satellites. Stereo images acquired by these satellites can be used

to detect built-up areas. Terrain slopes and building heights are calculated from stereo-extracted digital surface models (DSMs). These measurements are then used to improve built-up area results obtained by using planar features extracted from Digital Orthophoto Maps (DOMs); themselves produced from stereo images [10,12–14]. There is no specialized height feature or index for detecting built-up areas at this time. This study introduces an effective new way to extract height features from stereo imagery to obtain high-accuracy built-up area identification results.

### 1.1. Planar Features Used for Detecting Built-Up Areas

In remotely sensed imagery, spectral, shape, and texture features extracted from built-up areas differ from those extracted from non-built-up areas [15]. Built-up areas have a unique spectral response in remotely sensed imagery. Several of these spectral features have been proposed to detect built-up areas from moderate-resolution multispectral imagery. In [16], a Normalized Difference Built-up Index (NDBI) is derived from band 4 and band 5 of Landsat TM imagery, and used together with a re-coded Normalized Difference Vegetation Index (NDVI) to obtain built-up areas. In [17], the NDBI is improved by using a semiautomatic segmentation approach, achieving lower commission error. In [18], the Soil Adjusted Vegetation Index (SAVI), the Modified Normalized Difference Water Index (MNDWI), and the NDBI represent three major components of urban areas; vegetation, water, and built-up land, and are used to construct the Index-based Built-up Index (IBI). The IBI significantly enhances the built-up area detection performance from Landsat ETM+ imagery, effectively suppressing background noise. A comparison of these built-up spectral indexes shows that the IBI achieves notably higher accuracy [19].

Several shape or local features can be extracted from high-resolution panchromatic imagery for detecting built-up areas. Many man-made objects (e.g., buildings) in built-up areas have abundant angular points and regular shapes. In [20], SIFT keypoints [21] are represented as vertexes of a graph, while spatial distance and intensity values lead to edges of the graph. The graph is then used to achieve built-up areas from high-resolution panchromatic imagery using a novel multiple subgraph matching method. In [22], local feature points obtained using Gabor filtering form a spatial voting matrix. The matrix is then used to detect built-up areas from an aerial or satellite image using an optimum decision-making approach. In [23], feature points obtained from aerial images using an improved Harris detector are used to detect built-up areas with an improved orientation-sensitive voting technique. In [24], corners obtained from multiple satellite images using an improved Harris detector are used to detect built-up areas with two components: a likelihood-function-based approach for obtaining candidate results, and an unsupervised clustering algorithm for obtaining final results. In [25], a density map for a satellite image is created using a sparse Harris corner voting approach, and then the map is segmented by Cauchy graph embedding optimization to obtain built-up areas. Line segments extracted from remotely sensed imagery can also be used to detect built-up areas, with higher detection accuracy [15]. In [26], a set of measures based on straight lines (e.g., length, contrast, orientation, periodicity, and location) are used to assess land development levels in high-resolution panchromatic satellite images, obtaining urban areas using a parametric or nonparametric classifier. In [27], straight line segments extracted from satellite imagery are represented as vertexes of a graph, while their spatial relationships are encoded as edges of the graph. A set of graph measures are then used to extract suburban-style residential areas.

Several texture features can be extracted from high-resolution multispectral or panchromatic imagery for detecting built-up areas. Man-made objects (e.g., buildings) in built-up areas have abundant texture shapes. In [8], a texture-derived built-up presence index (Pantex) is calculated from a panchromatic image based on fuzzy rule-based composition of the Gray-Level Co-occurrence Matrix (GLCM). The Pantex can be used alone with a simple threshold, and can be used together with a morphological Differential Attribute Profile (DAP) to obtain built-up areas [6,8,9]. The Morphological Building/Shadow Index (MBI/MSI) [28] can be used to detect built-up areas from high-resolution multispectral imagery. In [15], a comparison of several features shows that the Pantex, the histogram

of oriented gradients, lacunarity, and textons are very consistent predictors for classifying formal and informal neighborhoods. Texture features can be used together with spectral and shape features to achieve accurate and robust built-up area detection in different high-resolution multi-sensor images [15,29]. In [30], multivariate textures are separately extracted using a multivariate variogram, and then combined with spectral information to obtain built-up areas from multispectral imagery.

These planar spectral, shape, and texture features cannot be easily adopted to achieve high-accuracy built-up area detection from high-resolution stereo imagery. The performance of built-up area detection using these planar features is degraded by within-class spectral variation and between-class spectral confusion in remotely sensed imagery, especially in the high-resolution panchromatic band [10]. Stereo imagery obtained by some sensors (e.g., Cartosat-1, and WorldView-1) usually has only the panchromatic band acquired with an off-nadir viewing angle [11]. For the raw image having an off-nadir viewing angle, the built-up area detection performance of these features may degrade, since they initially apply to nadir or near-nadir remotely sensed imagery [6]. For the DOM generated from stereo imagery, the performance of these features may also become degraded as a result of resampling processing during DOM production [11]. Incorporating auxiliary information such as height, might improve building detection results.

### *1.2. Built-Up Area Detection Methods Using Height Information*

The data sources of height information used for detecting built-up areas from remotely sensed imagery can be divided into two categories: auxiliary data, and the image itself. The auxiliary data may contain terrain height alone (e.g., STRM-DEM), or contain both terrain and object heights (e.g., LiDAR data, DSMs). In [31,32], terrain slopes were calculated from STRM-DEM, and then used to improve the built-up area detection results obtained by spectral and textural features. The performance gain was low, since building heights were not used for built-up area detection. Terrain slopes and building heights were calculated from LiDAR data [33,34] or open DSMs (e.g., ALOS World 3D-30m [35]), and then used to improve the built-up area detection results. The performance gain was higher due to the use of building heights. However, the accuracy of calculated terrain slopes and building heights is compromised when separating terrain and buildings [36,37], and in turn, influences the accuracy of built-up area detection. Moreover, the auxiliary data must be acquired around the same time as image acquisition. Otherwise, built-up area detection accuracy will be compromised.

Height information used for detecting built-up areas can be obtained from remotely sensed imagery itself, such as images containing building shadows, or stereo imagery. In [28,38], buildings and their shadows are detected using MBI and MSI, and then used to detect built-up areas from general remotely sensed imagery. Stereo imagery consists of two or three images having different viewing angles, and thus can be used to produce height information (e.g., DSMs, and disparity images) by image matching. Existing methods for detecting built-up areas from stereo imagery usually first create DSMs and DOMs from the stereo imagery. Then, terrain slopes and building heights extracted from a DSM are used to improve built-up area results obtained by using planar features extracted from a DOM [12,39]. The built-up area identification results cannot be superimposed on the raw stereo imagery for display during evaluation [40–42]. These methods do not take full advantage of images with different viewing angles of stereo imagery, since these images are only used to produce DSMs and DOMs. The same built-up area looks different in images collected from different viewing angles. The disclosure of the difference and the relationship of the same built-up area among these images will be helpful for obtaining high-accuracy built-up areas, as our research shows.

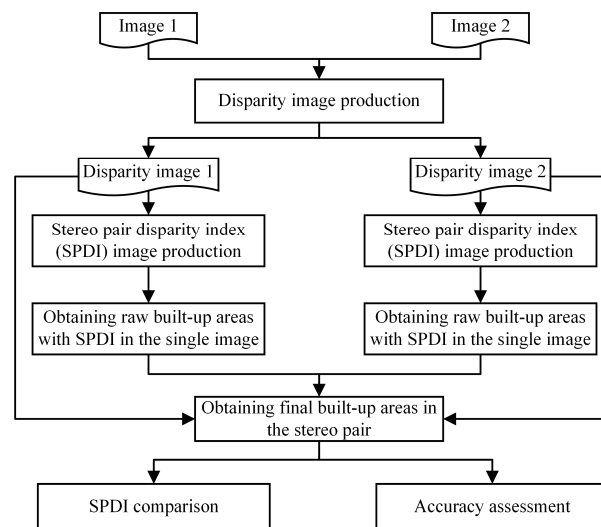
### *1.3. Objective of This Paper*

In this study, a new Stereo Pair Disparity Index (SPDI) effective for indicating built-up areas is calculated from stereo-extracted disparity information. Furthermore, a new method of detecting built-up areas from high-resolution stereo imagery based on the SPDI is proposed that establishes the relationship between images having different viewing angles, using disparity information.

High-accuracy built-up area identification results are obtained from these images, and the results can be superimposed on images for evaluation and verification in production work.

## 2. Methodology

The flowchart of the proposed method of detecting built-up areas from a stereo pair using disparity information is shown in Figure 1. Two disparity images are first produced from a stereo pair. An SPDI image is generated from a disparity image, and then used to detect raw built-up areas for the corresponding raw image in the stereo pair. For a stereo pair, the raw built-up area identification results for the two constituent images are obtained separately, and thus may be inconsistent. The final consistent built-up areas derived from these intermediate results are realized by linking the images using disparity information. These built-up area identification results are evaluated with ground truth data, and a SPDI comparison between built-up and non-built-up areas.



**Figure 1.** Flowchart of the proposed method for detecting built-up areas from a stereo pair.

### 2.1. Disparity Image Production

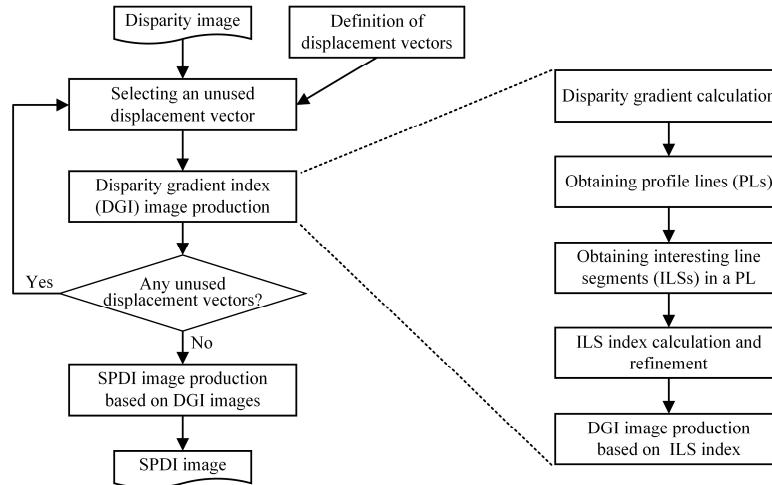
Two disparity images are produced from a stereo pair as follows. An epipolar stereo pair is generated from the raw stereo pair [43,44]. One epipolar image is designated as the reference image, while the other is considered as a secondary image. For a pixel in the former, its homonymous pixel in the latter is obtained in the same row using the semi-global matching method [45,46]. When pixels fail to match, they are filled using a fast marching inpainting method [47]. A disparity image that can be superposed on the reference image is created. If the reference image and a secondary image are exchanged, then the disparity image for the other image is also obtained. In this way, two disparity images in epipolar projection are generated, and can be superposed on the two epipolar images, respectively. Furthermore, for two homonymous pixels belonging to above-ground objects (e.g., buildings) in the two disparity images, one pixel has a larger disparity value than the surrounding area in one of the disparity images, while the other pixel has a smaller disparity value in the other, complementary disparity image [45]. For the latter disparity image, the disparity value of each pixel is negated, creating a new disparity image.

### 2.2. Stereo Pair Disparity Index (SPDI) Image Production

An SPDI image is produced from a disparity image. The flowchart of SPDI image production is shown in Figure 2. In a disparity image, the gradient usually changes around the boundary of buildings. For a pixel in the disparity image, the SPDI is calculated based on gradients around the target pixel. Since the gradient is relevant to the specified displacement vector, a series of displacement



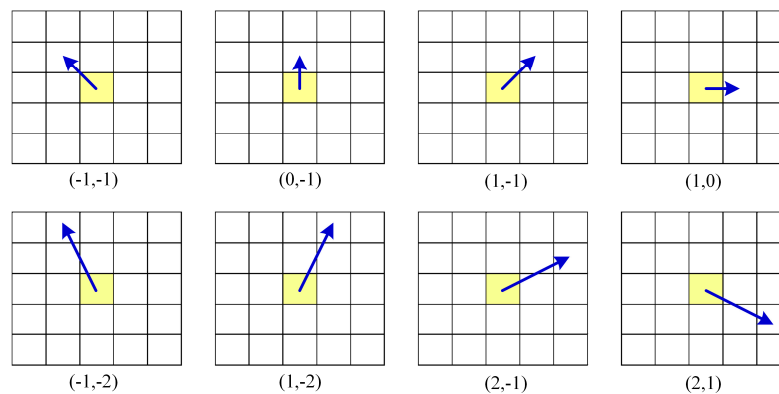
vectors are first defined. A Disparity Gradient Index (DGI) image is created from the disparity image based on a specified displacement vector. The DGI image production procedure is shown in the right part of Figure 2, and will be described in detail in the following subsection. Several DGI images are created for the same disparity image with the defined displacement vectors in turn. Finally, an SPDI image is generated from these DGI images.



**Figure 2.** Flowchart of stereo pair disparity index (SPDI) image production.

### 2.2.1. Definition of Displacement Vectors

Eight displacement vectors used for disparity gradient calculation and DGI image production are illustrated in Figure 3. The vector value  $(\Delta x, \Delta y)$  is given below each displacement vector. The upper four displacement vectors have the same displacement size (one), while the lower four displacement vectors also have the same displacement size (two). If a displacement vector is rotated by 180 degrees, the new displacement vector is equivalent to the raw displacement vector in the gradient calculation. Therefore, only half of all potential displacement vectors of a displacement size are necessary for analysis.



**Figure 3.** Eight displacement vectors used for producing a disparity gradient index (DGI) image.

### 2.2.2. DGI Image Production Based on a Displacement Vector

The procedure for producing a DGI image from a disparity image using a specified displacement vector is as shown in the right part of Figure 2. The disparity gradient value of a pixel  $(i, j)$  is calculated with a specified displacement vector  $(\Delta x, \Delta y)$  as Equation (1).

$$DG(i, j) = d(i + \Delta x, j + \Delta y) - d(i, j), \quad (1)$$

A disparity gradient image is created by calculating the gradient for all pixels in the disparity image. Then, the disparity gradient image is decomposed into numerous Profile Lines (PLs) using the specified displacement vector. For each PL, a number of Interesting Line Segments (ILSs) are obtained. The index of each ILS on the PL is calculated, and refined using adjacent PLs. Finally, a DGI image is created based on the index of ILSs on all PLs.

#### Obtaining PLs from a Disparity Gradient Image

Numerous PLs are generated by decomposing the disparity gradient image with the specified displacement vector  $(\Delta x, \Delta y)$ . The equation of a PL is determined as Equation (2).

$$\Delta y \times x + \Delta x \times y + c = 0, \quad (2)$$

where,  $c$  is the parameter of the equation. The range of  $c$  is determined by the four corners in the image, whose coordinates are  $(0, 0)$ ,  $(0, col - 1)$ ,  $(row - 1, 0)$ , and  $(row - 1, col - 1)$ , respectively. Using the four corners, four  $c$  values are obtained, and labeled as  $c_{cor1}$ ,  $c_{cor2}$ ,  $c_{cor3}$ , and  $c_{cor4}$ . Then, the union of these PLs is determined based on the four  $c$  values, as Equation (3).

$$ULines = \{L_k | L_k : \Delta y \times x + \Delta x \times y + c_k = 0, c_{min} \leq c_k \leq c_{max}\}, \quad (3)$$

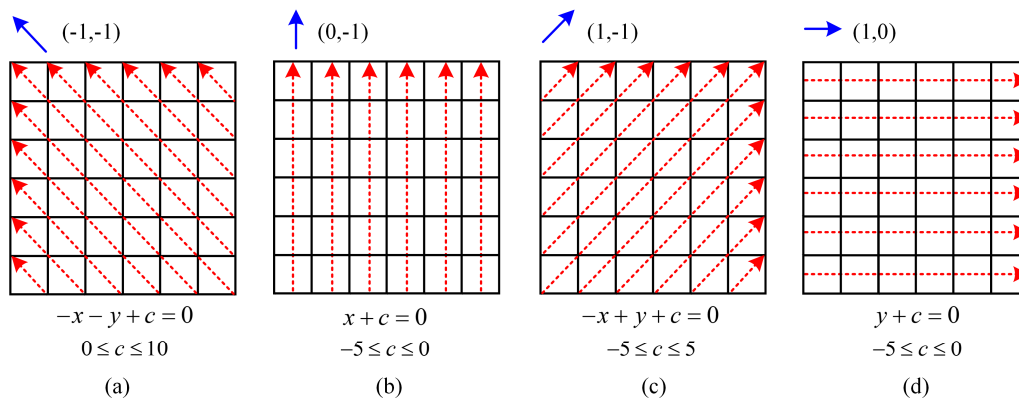
where,  $c_{max} = \max(c_{cor1}, c_{cor2}, c_{cor3}, c_{cor4})$ ,  $c_{min} = \min(c_{cor1}, c_{cor2}, c_{cor3}, c_{cor4})$ . The total number of the obtained PLs is calculated as Equation (4).

$$numL = 1 + c_{max} - c_{min}, \quad (4)$$

For each PL, its points are determined by its PL equation. The first point  $(x_{k0}, y_{k0})$  in the  $k$ th PL is decided by the specified  $c_k$  as in Equation (2). Then, other points on the PL are obtained with the PL equation. The point union of the PL is as Equation (5).

$$UPoints(k) = \{(i, j) | (i, j) = (x_{k0} + \Delta x \times p, y_{k0} + \Delta y \times p), 0 \leq i \leq col - 1, 0 \leq j \leq row - 1, p \in N\}, \quad (5)$$

Figure 4 illustrates how PLs are obtained from a disparity gradient image. The Figure 4 a–d show PLs obtained from the same disparity gradient image using four same sized displacement vectors. The blue arrow is the displacement vector, while the red dotted lines with directional arrows are PLs corresponding to the given equations.



**Figure 4.** Using different displacement vectors to obtain numerous profile lines (PLs) from the same disparity gradient image (DGI). (a–d) Using displacement vectors  $(-1, -1)$ ,  $(0, -1)$ ,  $(1, -1)$ , and  $(1, 0)$ , respectively.

### Obtaining ILSs in a PL

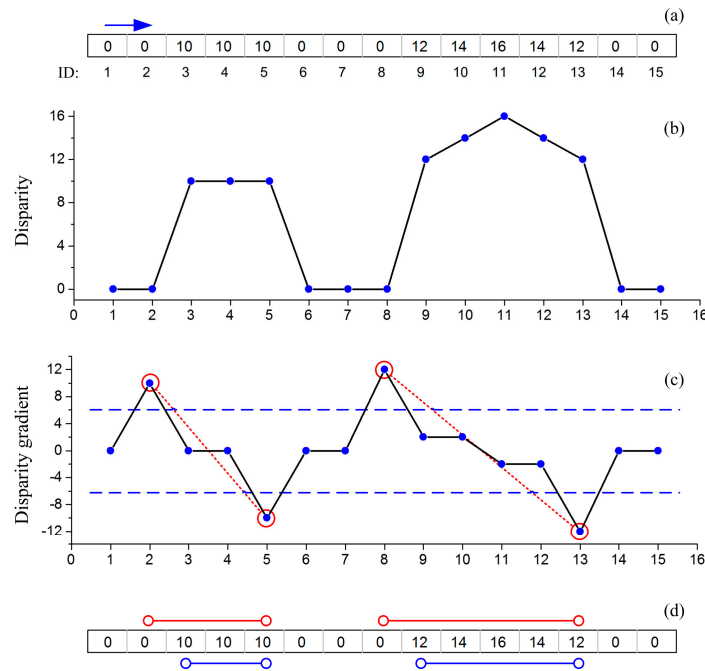
The pixels with large gradient changes are chosen as interesting points in order to obtain the ILS. The interesting points on the  $k$ th PL are obtained as Equation (6).

$$\text{UItpoints}(k) = \{(i, j) | (i, j) \in \text{Upoints}(k), \text{abs}(DG(i, j)) \geq tg\}, \quad (6)$$

where,  $tg$  is the threshold of absolute disparity gradient value, and set referring to height. The disparity value of a pixel indicates its height [48], as Equation (7).

$$d = \frac{B}{H} \times \frac{1}{L} \times h, \quad (7)$$

where,  $B/H$  is the base-height-ratio of a stereo pair that is used to produce the disparity image [49,50],  $L$  is the ground sample distance. A disparity  $d$  is obtained with a threshold height  $h$  (e.g., 3 m) through Equation (7), and then used as  $tg$ . The setting of  $tg$  determines the interesting points. The interesting point number decreases with the increase of  $tg$ . In Figure 5c, four interesting points are obtained when  $tg$  is set at 6, which corresponds to the two blue dotted lines.



**Figure 5.** Obtaining Interesting Line Segments (ILSs) in a Profile Line (PL). (a) PL with disparity values. (b) Disparity profile. (c) Profile of disparity gradient calculated with the displacement vector (1, 0). (d) PL with obtained ILSs.

Several ILSs are acquired by matching the selected interesting points on each PL. Two interesting points on a PL are associated with an above-ground object (e.g., building). For the two points, one disparity gradient value is positive, while the other is negative. These two matched points determine a line segment, as shown in Figure 5c. The red circles indicate the interesting points obtained by thresholds illustrated by the dashed blue lines, while a red dotted line between two interesting points illustrates the match between them. There are several matched line segments on each PL. The union of the matched line segments on the  $k$ th PL is as Equation (8).

$$\text{UMatchedLines}(k) = \{ml(w) = [(i1, j1), (i2, j2)] | (i1, j1) \in \text{UItpoints}(k), (i2, j2) \in \text{UItpoints}(k), \text{grad}(i1, j1) > 0, \text{grad}(i2, j2) < 0\} \quad (8)$$

These line segments, however, still rely on the specified displacement vector used for gradient calculation. Matched line segments are refined with the specified displacement vector separately, producing the final ILSs. The union of the ILSs obtained on the  $k$ th PL is as Equation (9).

$$\text{UItsLines}(k) = \{ \text{Itsline}(w) = [(i1 + \Delta x, j1 + \Delta y), (i2 + \Delta x - 1, j2 + \Delta y - 1)] \mid \text{ml}(w) = [(i1, j1), (i2, j2)], \text{ml}(w) \in \text{UMatchedLines}(k) \} \quad (9)$$

As shown in Figure 5d, the two red lines are the raw matched line segments, while the two blue lines are the final ILSs refined with the specified displacement vector. In this way, ILSs on all PLs are obtained, in turn.

#### ILS Index Calculation and Refinement

For each obtained ILS, the ILS index is calculated based on the length and the disparity difference with the surroundings. The influence of ILS length is calculated as Equation (10).

$$p(\text{length}) = \begin{cases} e^{\frac{\text{length}}{tl1}-1}, & \text{length} < tl1 \\ 1, & tl1 \leq \text{length} \leq tl2, \\ e^{1-\frac{\text{length}}{tl2}}, & \text{length} > tl2 \end{cases} \quad (10)$$

where, the ILS *length* is calculated from the two end points  $(x1, y1)$  and  $(x2, y2)$ , as  $\text{length} = \sqrt{(x2 - x1)^2 + (y2 - y1)^2}$ . The unit of the two parameters,  $tl1$  and  $tl2$ , is pixels. The two parameters are set based on the range of building lengths. If  $tl1$  is set as a very small value while  $tl2$  is set as a very large value,  $p(\text{length})$  may be overestimated. In contrast, if  $tl1$  is set as a very large value while  $tl2$  is set as a fewer large value,  $p(\text{length})$  may be underestimated. The influence of the disparity difference with the surroundings is calculated as follows. The difference *difD* between the average disparity value of the ILS  $\text{AvgD}_{ILS}$  and the disparity  $d_{sp}$  of a surrounding point in the PL is calculated as  $\text{difD} = \text{AvgD}_{ILS} - d_{sp}$ . The average disparity value of the ILS is calculated as  $\text{AvgD}_{ILS} = \sum_{i=1}^{pn} \frac{d(i)}{pn}$ , where  $pn$  is the total number of the points of the ILS,  $d(i)$  is the disparity value of the  $i$ th point of the ILS. Then, the influence of the disparity difference *difD* of the surrounding point on the ILS index is calculated as Equation (11).

$$p(\text{difD}) = \begin{cases} 0, & \text{difD} < tg \\ 1, & tg \leq \text{difD} \leq tg2, \\ e^{1-\frac{\text{difD}}{tg2}}, & \text{difD} > tg2 \end{cases} \quad (11)$$

where, the two parameters,  $tg$  and  $tg2$ , have no unit, since they measure difference between two disparity values whose unit is pixels [48]. They can be set according to the range of disparity values of buildings, as calculated from building heights in the scene covered by the stereo imagery assisted by Equation (7);  $tg$  is the parameter for obtaining interesting points as in Equation (6). If  $tg$  is set as a very small value while  $tg2$  is set as a very large value,  $p(\text{difD})$  may be overestimated. In contrast, if  $tg$  is set as a very large value while  $tg2$  is set as a fewer large value,  $p(\text{difD})$  may be underestimated. Moreover, the two points surrounding the two end points of an ILS usually lie on the profile line. For an ILS, the final influence of the disparity difference  $p(\text{difD}_{Final})$  on the ILS index is calculated as (12).

$$p(\text{difD}_{Final}) = \min(p(\text{difD}_1), p(\text{difD}_2)), \quad (12)$$

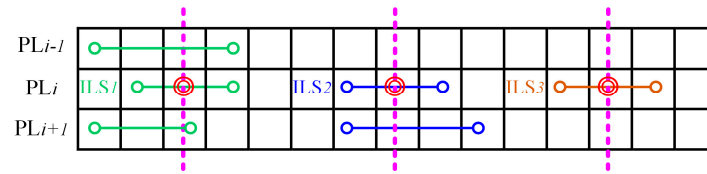
where,  $p(\text{difD}_1)$  and  $p(\text{difD}_2)$  are the influence of two surrounding points, respectively. Finally, the ILS index is calculated as (13).

$$\text{ILS\_Index} = p(\text{length}) \times p(\text{difD}_{Final}), \quad (13)$$

After calculating the index of ILSs on all Profile Lines (PLs), the index of each ILS on a PL is refined with the help of the two adjacent PLs. The ILSs in built-up areas are gathered, with corresponding ILSs in the two adjacent PLs. For an ILS, there is a line that passes through the middle point of the ILS, is perpendicular to the PLs, and intersects with an ILS in an adjacent PL. If an ILS does not meet these criteria, its DGI value is updated as Equation (14).

$$ILS\_Index = ILS\_Index \times 0.5, \quad (14)$$

For instance in Figure 6,  $ILS_1$ ,  $ILS_2$ , and  $ILS_3$  are three ILSs to be refined on the  $PL_i$ .  $PL_{i+1}$  and  $PL_{i-1}$  are the two adjacent PLs. The red double circle is the middle point of the three ILSs. The purple dashed lines are perpendicular to the PLs. The index of  $ILS_3$  needs to be updated, while the indexes for  $ILS_1$  and  $ILS_2$  need no updating.



**Figure 6.** Interesting line segment (ILS) index refinement with adjacent profile lines (PLs).

#### DGI Image Production Based on ILS Index

A DGI image is created based on the index of ILSs on all PLs. The DGI value of a pixel  $(i, j)$  is assigned the index  $ILS(w)$  of the ILS to which the pixel belongs, as Equation (15).

$$DGI(i, j) = \begin{cases} ILS\_Index(w), & \text{if } (i, j) \in ILS(w) \\ 0, & \text{else} \end{cases}, \quad (15)$$

A pixel which does not belong to an ILS is assigned 0. Thus, a DGI image is created for the disparity image using the specified displacement vector. For a disparity image, eight DGI images are created with, eight corresponding defined displacement vectors.

#### 2.2.3. SPDI Image Production Based on DGI Images

For a disparity image, an SPDI image is created from the eight generated DGI images. The SPDI value of a pixel is calculated from DGI values of the pixel in the eight DGI images. Mathematically, it is calculated as Equation (16).

$$SPDI(i, j) = \sum_{t=1}^8 \frac{(DGI(i, j)_t)}{8}, 0 \leq SPDI(i, j) \leq 1, \quad (16)$$

where,  $DGI(i, j)_t$  is the DGI value calculated with the  $t$ th displacement vector. The SPDI values of all pixels in the disparity image are calculated in this way, creating the SPDI image. The SPDI values range from 0 to 1, and indicate the probability that a pixel belongs to a built-up area. Pixels with SPDI values closer to 1 are more likely to belong to built-up areas. The SPDI image is in epipolar projection, and can be superimposed on the disparity image. For a stereo pair, two SPDI images are produced from its two disparity images.

#### 2.3. Obtaining Raw Built-Up Areas with SPDI in a Single Image

In a disparity image, an area where pixels with high SPDI values are gathered is likely to be a built-up area. The threshold between high and low SPDI values is determined by using a boxplot [51] as reference. The SPDI values; greater than 0, are sorted from the least to the greatest,



creating a ranking list. Then, the threshold  $Spdi\_b$  is automatically determined as in Equation (17) without the need for human intervention.

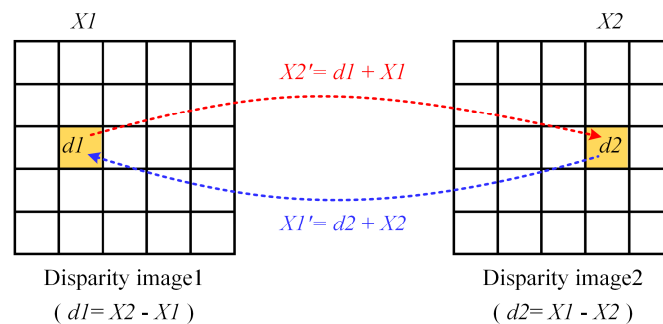
$$Spdi\_b = \min_{b=0}^1 \{b | Q1(b) - (Q3(b) - Q1(b)) \times 1.5 > 0\}, \quad (17)$$

where,  $b$  is a value used to create a sublist from the ranking list by picking out SPDI values greater than the value  $b$ . The first quartile  $Q1(b)$  is the 25th percentile in the sublist, while the third quartile  $Q3(b)$  is the 75th percentile. The threshold  $Spdi\_b$  is the minimum of all  $b$  values meeting the requirements of equation Equation (17).

Once the threshold is determined, the pixels whose SPDI values are greater than the threshold are selected as pixels belonging to built-up areas. Based on these pixels, built-up areas are identified using the Voronoi-based method [52]. Pixels with few neighboring pixels are first deleted, and the remaining pixels are triangulated, and are used to create a border polygon. The border polygon is used as a boundary for built-up area identification results as determined from the disparity image. The built-up area identification results are also results for the epipolar image used as the reference image when producing the disparity image.

#### 2.4. Obtaining Final Built-Up Areas in a Stereo Pair

Built-up area results for an epipolar stereo pair are derived from the raw results obtained separately from the two corresponding epipolar images. For the two epipolar images, the raw results must be consistent since they cover the same scene, instead they may in fact be inconsistent. Considering that the areas supported by both images are likely to be true built-up areas, the raw results of the two images are fused to obtain consistent and credible results. However, these results cannot be directly superimposed given the disparity between the homonymous points, and thus cannot be immediately fused unless they are aligned by transforming the results from one image into the other image with the help of disparity images. The transformation need not be applied to all pixels in the results, but the transformation must be applied to the boundary pixels in the secondary image, determining new boundary pixels in the reference image. The transformation between homonymous points is illustrated in Figure 7.  $d1$  and  $d2$  are disparity values of the homonymous points. The values  $X1$  and  $X2$  are the raw column numbers, while  $X1'$  and  $X2'$  are the calculated column numbers. Then, the aligned raw results in the two images are overlaid to determine consistent and credible built-up area results for the epipolar stereo pair.



**Figure 7.** Location transformation between homonymous points in an epipolar stereo pair.

Final built-up area identification results in geographical projection are created for the raw stereo pair. Results for the epipolar stereo pair cannot be immediately superposed on the raw stereo pair; as the epipolar stereo pair is in epipolar projection, while the raw stereo pair is in geographical projection. Results for the raw stereo pair are obtained when results for the epipolar stereo pair are transformed into a geographical projection [43,44].

## 2.5. SPDI Comparison

SPDI values are evaluated through the comparison of SPDI value distribution between built-up and non-built-up areas. In an SPDI image, when an SPDI value is specified as the threshold, the pixels whose SPDI values are no greater than the threshold are determined. The percentages of the pixels belonging to built-up and non-built-up areas are calculated as in Equations (18) and (19), respectively.

$$PerBU_{spdi\_t} = \frac{|\cup(i,j)|0 < SPDI(i,j) \leq spdi\_t, (i,j) \in BuArea|}{TN}, \quad (18)$$

$$PerNBU_{spdi\_t} = \frac{|\cup(i,j)|0 < SPDI(i,j) \leq spdi\_t, (i,j) \in NbuArea|}{TN}, \quad (19)$$

where,  $spdi\_t$  is the threshold, which ranges from 0 to 1, and  $TN$  is the total number of pixels in the SPDI image. The terms  $BuArea$  and  $NbuArea$  refer to built-up and non-built-up areas, respectively. The pixels having zero (0) SPDI value are not included in calculations; as some pixels having zero (0) SPDI values may be in non-built-up areas, or may be in built-up areas (e.g., gaps between buildings). Then, the percentage of the pixels belonging to built-up and non-built-up areas is compared between them. During the comparison, several thresholds  $spdi\_t$  can be specified.

SPDI values are empirically evaluated in terms of effectiveness, reliability, and consistency. SPDI effectiveness, reliability, and consistency rather than the SPDI percentage itself are used to compare the accuracy of built-up area results for different stereo pairs. SPDI effectiveness is measured by the difference between SPDI percentages in built-up areas and in non-built-up areas. SPDI effectiveness increases with an increase of the difference. SPDI reliability is measured by the difference between SPDI percentages for results and for ground truth data. SPDI reliability declines with an increase of the difference. SPDI consistency is measured by the percentage difference between two SPDI images of a stereo pair. SPDI consistency declines with an increase of the difference. Robust SPDI consistency means that SPDI is not influenced by the image acquisition angles. SPDI effectiveness and reliability are calculated for a single image, while SPDI consistency is calculated for a stereo pair. For stereo pairs covering different scenes, the built-up area identification results obtained from them are not comparable through the use of the SPDI percentage. The SPDI percentage varies with the percentage of built-up areas and the percentage of building foot print areas in a scene, and thus is related to a specific scene. In contrast, SPDI effectiveness, reliability, and consistency are unrelated to a specific scene; as they use the difference between SPDI percentages in built-up areas and in non-built-up areas.

## 2.6. Accuracy Assessment

The accuracy must be assessed in a stereo pair as a whole rather than in a single image. For a stereo pair, its accuracy  $Accuracy_{pair}$  is calculated based on the accuracy that is separately calculated in its two images,  $Accuracy_1$  and  $Accuracy_2$ , as Equation (20).

$$Accuracy_{pair} = \frac{Accuracy_1 \times \cos(LA_1) + Accuracy_2 \times \cos(LA_2)}{\cos(LA_1) + \cos(LA_2)}, \quad (20)$$

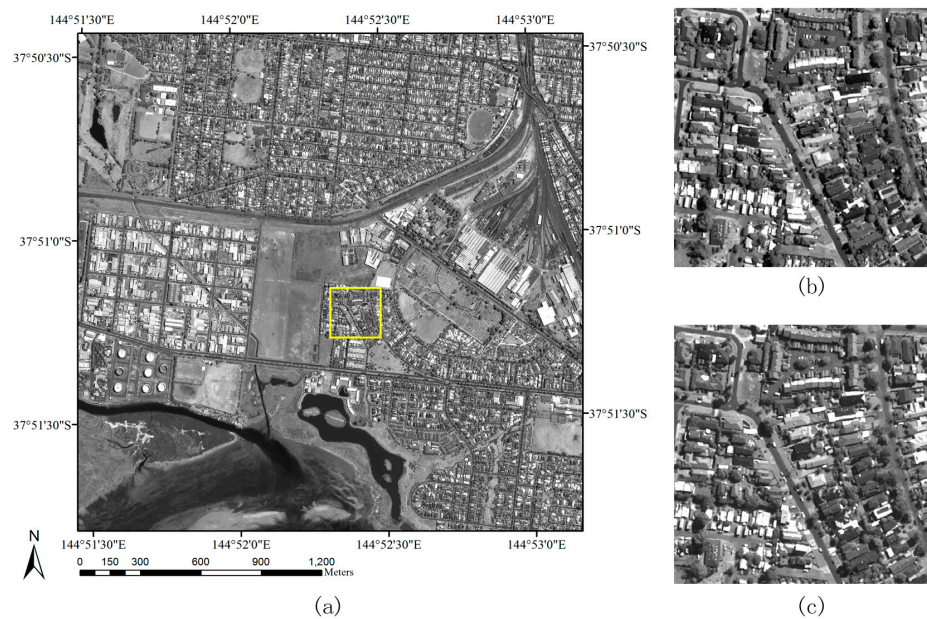
where,  $LA_1$  and  $LA_2$  are off-nadir looking angles of the two images, respectively. The accuracy in the equation can be detection percentage, branch factor [53,54], or kappa coefficient [9]. A large detection percentage or kappa coefficient, or a small branch factor means that built-up area identification results have high accuracy.

## 3. Results

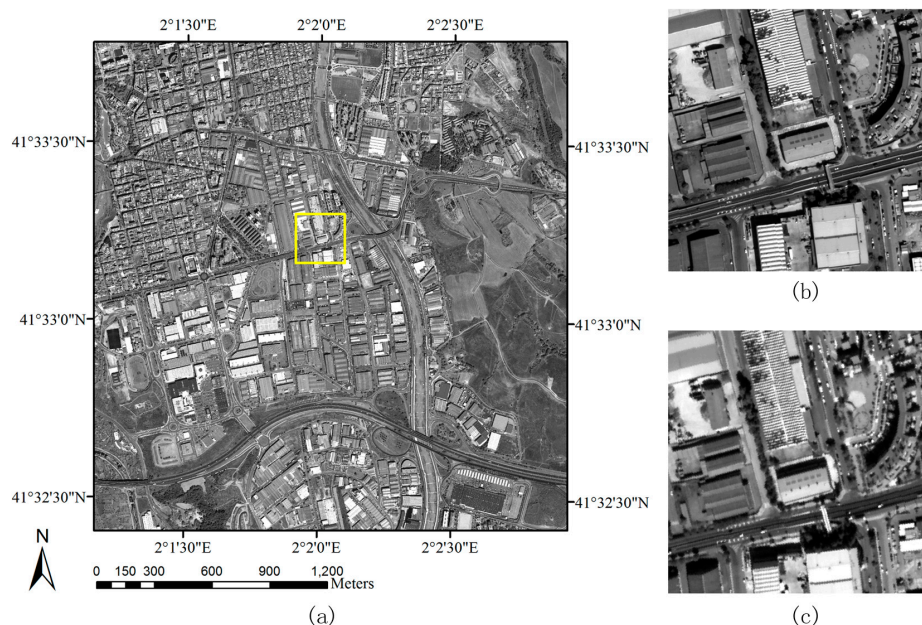
### 3.1. Study Data

Two stereo pairs acquired separately by the Pleiades-1A and WorldView-1 satellite sensors were used in our experiments, as shown in Figures 8 and 9. The first scene covered by stereo

pair 1 is located in Williamstown, Melbourne, Australia, as shown in Figure 8. The second scene covered by stereo pair 2 is located in Terrassa, Catalonia, Spain, as shown in Figure 9. The two scenes differ from each other over the shape and structure of built-up areas. In the first scene, built-up areas consist of high-density small-size residential buildings, medium-sized factory buildings, and chimneys. There are high-density tall trees adjacent to built-up areas, as shown in the top left corner of Figure 8. In the second scene, built-up areas consist of high-density small-size residential buildings and large-sized industrial and commercial buildings. The characteristics of the two stereo pairs also differ from each other, as shown in Table 1.



**Figure 8.** Images of stereo pair 1. (a) Image 1 of the stereo pair as a whole. (b,c) A detail view of the yellow rectangle seen in (a) in Image 1 and in Image 2, respectively.



**Figure 9.** Images of stereo pair 2. Descriptions of the content are the same as those in Figure 8. (a) Image 1 of the stereo pair as a whole. (b,c) A detail view of the yellow rectangle seen in (a) in Image 1 and in Image 2, respectively.

**Table 1.** Characteristics of two stereo pairs.

| Characteristics              | Stereo Pair 1    |          | Stereo Pair 2  |          |
|------------------------------|------------------|----------|----------------|----------|
|                              | Image 1          | Image 2  | Image 1        | Image 2  |
| Sensor name                  | Pleiades-1A      |          | Worldview-1    |          |
| Acquisition date             | 25 February 2012 |          | 29 August 2008 |          |
| Acquisition time             | 00:25:32         | 00:26:27 | 10:51:12       | 10:50:27 |
| Radiometric resolution       | 8 bits           | 8 bits   | 11 bits        | 11 bits  |
| Ground sample distance (GSD) | 0.5 m            | 0.5 m    | 0.51 m         | 0.66 m   |
| Sun azimuth                  | 54.0°            | 53.7°    | 152.8°         | 152.5°   |
| Sun elevation                | 49.6°            | 49.7°    | 55.0°          | 54.9°    |
| Satellite azimuth            | 2.7°             | 185.0°   | 256.2°         | 7.6°     |
| Satellite elevation          | 73.9°            | 73.8°    | 85.9°          | 55.0°    |
| Off-nadir view angle         | 16.2°            | 16.5°    | 4°             | 33°      |
| Convergence angle            | 32.3°            |          | 36.7°          |          |
| Base-height-ratio            | 0.57             |          | 0.77           |          |

### 3.2. Experimental Setting

Our built-up area identification method was compared with a method for single remotely sensing images, and two methods for stereo images. The method based on a planar feature (Gabor) for single images [22] was chosen for comparison because of its outstanding performance. The two methods for stereo images use a height feature alone (nDSM) [36], and a combination of height and planar features [13], respectively. Our method was fused at the decision level with a method using a planar feature (Gabor) [22] in order to be fairly compared with the method [13]. The ground truth of built-up areas in our study was obtained by manually outlining the built-up area boundary from the two stereo pairs.

The implementation of our method was divided into two parts. The first part concerning epipolar image production, and the transformation between the epipolar and geographical projection was implemented with ENVI 5.1 and IDL 8.3. The second part concerning SPDI calculations, and built-up area identification was implemented with Visual Studio 2010 and OpenCV 2.4.4.

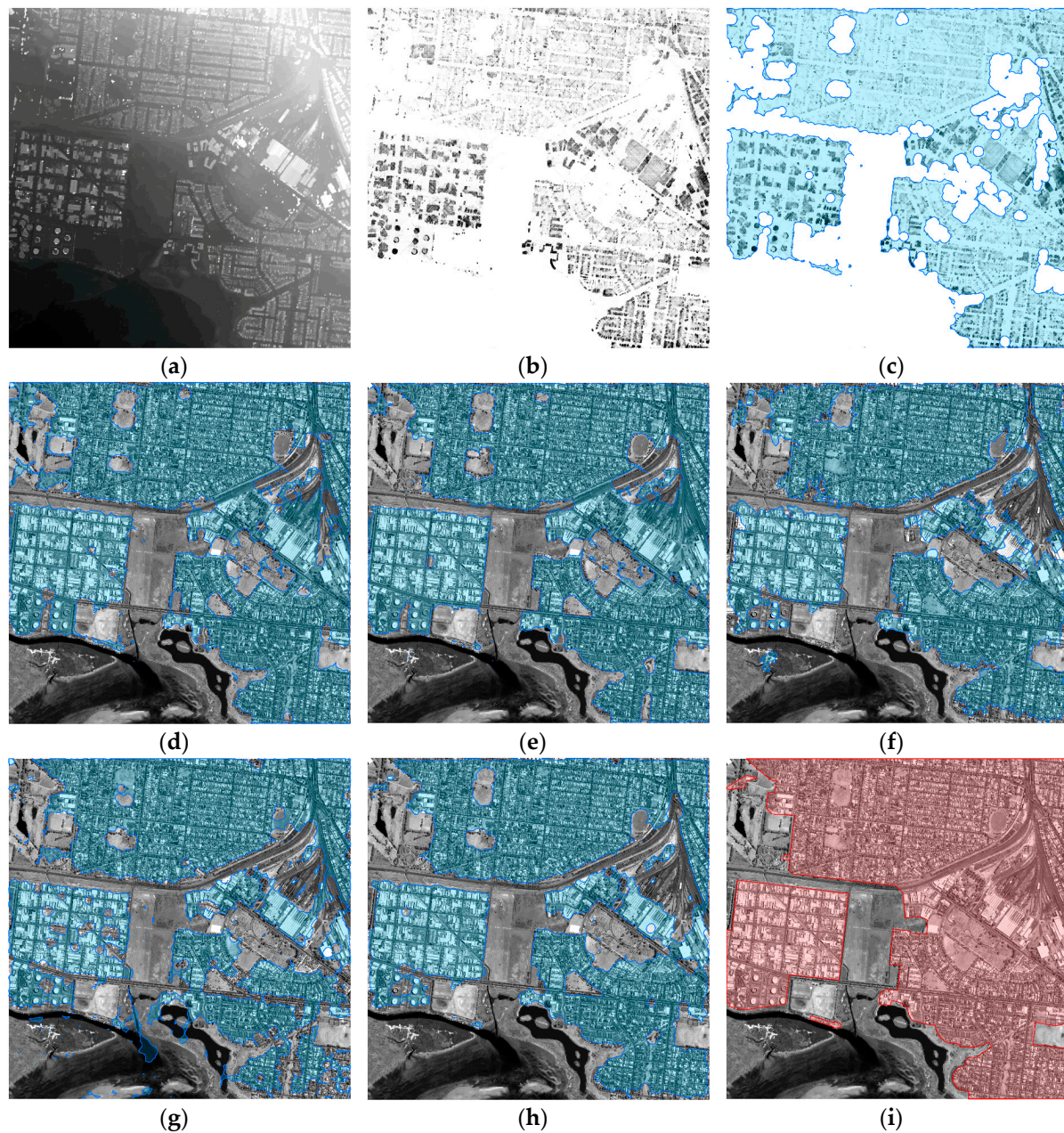
The experimental parameters for the two stereo pairs were set according to the scenes in the study areas. As can be seen from Figures 8 and 9, most buildings in stereo pair 1 are small and short with little disparity between the roof and footprint, while most buildings in stereo pair 2 are larger and higher. Therefore, parameters related to building footprint size  $t/1$  and  $t/2$  for stereo pair 2 were greater than that for stereo pair 1. The two parameters were set to 2 pixels and 60 pixels respectively for the stereo pair 1, but 6 pixels and 80 pixels for the stereo pair 2. Moreover, parameters related to disparity threshold  $tg$  and  $tg2$  for stereo pair 2 were also greater than that for stereo pair 1. The two parameters were set to 20 and 160 respectively for the stereo pair 1, but 50 and 300 for the stereo pair 2.

### 3.3. Results and Accuracy

The built-up area identification results from the two stereo pairs are shown in subfigures a–e of Figures 10 and 11. Light pixels indicate large disparity values in subfigure a, while dark pixels indicate large SPDI values in subfigure b of Figures 10 and 11. Areas filled with pale blue color in subfigures c and d of Figures 10 and 11 are SPDI-based built-up area results superposed on SPDI image and Image 1 of the stereo pair, respectively. As can be seen from subfigures a–c of Figures 10 and 11, pixels of buildings usually have larger disparity values than the surroundings, have large SPDI values, and are in built-up areas. Some pixels of above-ground non-buildings (e.g., trees) also have larger disparity values than the surroundings and large SPDI values. For instance, trees in the upper-left corner of Figure 10 and the upper-right corner of Figure 11 have large SPDI values, resulting in false built-up area identification results as shown in subfigure d of Figures 10 and 11. The method consumed about 420 s for stereo pair 1, and 360 s for stereo pair 2, which was comparable with the time of manually delineating the ground truth. Besides, the built-up area identification results are improved



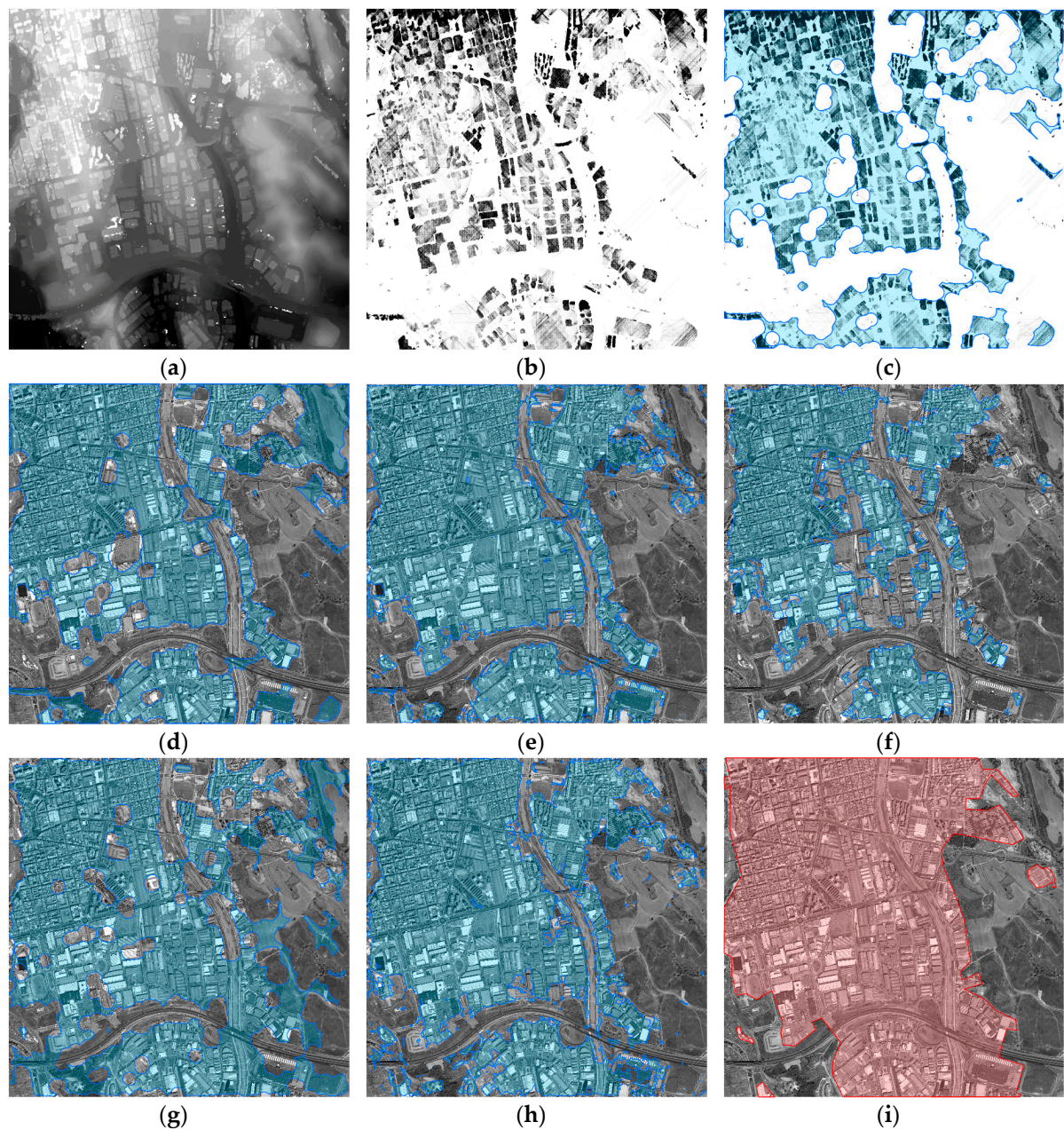
when SPDI is combined with a planar feature (Gabor) during built-up area detection, as shown in subfigure e of Figures 10 and 11.



**Figure 10.** Results for stereo pair 1. (a) Disparity image. (b) SPDI image. (c,d) SPDI-based built-up area results superposed on SPDI image and Image 1 of the stereo pair, respectively. (e) Built-up area results obtained using the combination of SPDI and Gabor. (f–h) Built-up area results obtained using the methods [22], [36], and [13], respectively. (i) Ground truth map.

Built-up area results obtained from a single image using the method [22] are shown in subfigure f of Figures 10 and 11. The used single image was the Digital Orthophoto Map (DOM) produced from the corresponding stereo pair used in our study. Some building pixels may not all be in the identified built-up areas. Especially in stereo pair 2, some large buildings with flat, single color roofs were not identified as built-up areas. The identified built-up areas were more fragmentary than our method, especially for stereo pair 2.





**Figure 11.** (a–i) Results for stereo pair 2. Descriptions of the content are the same as those in Figure 10.

Built-up area results obtained from stereo pairs using the methods [36], and [13] are shown in subfigures g and h of Figures 10 and 11, respectively. There were some false built-up areas for the method [36] using a height feature alone, such as areas in the bottom center part of subfigure g in Figure 10 and the right part of subfigure g in Figure 11. These false built-up areas were much more than that obtained with our method based on SPDI. Moreover, built-up area results were improved by the method [13] using a combination of height and planar features, as shown in subfigure h of Figures 10 and 11. The results were very similar to those obtained by our method of combining SPDI and Gabor features. However, our method obtained better results in some areas, such as dense train roads in the right center part of subfigure e in Figure 10. The built-up area identification results obtained by our method and these compared methods were a subset of the ground truth of built-up areas, not including areas comprised of wide roads with poor texture.

Table 2 shows the accuracy results of five built-up area identification methods. For stereo pair 1, our two methods, SPDI alone, and SPDI and Gabor for stereo images, achieved a higher detection percentage and kappa coefficient than that for single images [22]. When compared with the methods [36] and [13] for stereo images, our two methods also achieved higher accuracy. Among the five methods, our two methods achieved the highest detection percentage (0.84 and 0.86) and kappa coefficient (0.71 and 0.75), with a low branch factor (0.04 and 0.03). For stereo pair 2, our two methods also achieved good accuracy results. The method [22] for single images, achieved the lowest branch factor, but also lowest detection percentage and kappa coefficient. Our SPDI-based achieved a litter high branch factor (0.12), while the method based on nDSM [36] achieved a much higher branch factor (0.18). When height and planar features were combined to detect built-up areas, the branch factor was significantly reduced by our method and the method [13], achieving highest detection percentage (0.81 and 0.79) and kappa coefficient (0.87 and 0.85) among five methods. On the whole, the accuracy of stereo pair 1 was higher than the accuracy of stereo pair 2.

**Table 2.** Accuracy results of five built-up area identification methods.

| Study Data    | Accuracy             | SPDI | SPDI & Gabor | Method [22] | Method [36] | Method [13] |
|---------------|----------------------|------|--------------|-------------|-------------|-------------|
| stereo pair 1 | detection percentage | 0.84 | 0.86         | 0.75        | 0.70        | 0.76        |
|               | branch factor        | 0.04 | 0.03         | 0.02        | 0.04        | 0.01        |
|               | kappa coefficient    | 0.71 | 0.75         | 0.63        | 0.55        | 0.66        |
| stereo pair 2 | detection percentage | 0.69 | 0.81         | 0.56        | 0.71        | 0.79        |
|               | branch factor        | 0.12 | 0.04         | 0.01        | 0.18        | 0.04        |
|               | kappa coefficient    | 0.76 | 0.87         | 0.69        | 0.74        | 0.85        |

### 3.4. SPDI Comparison

As can be seen from b of Figures 10 and 11, SPDI values in built-up and non-built-up areas vary significantly. Most of the pixels with high SPDI values were in built-up areas, while the pixels in non-built-up areas rarely had high SPDI values. For instance, rectangular and complex buildings, and circular chimneys in built-up areas in Figure 10 had high SPDI values. In built-up areas, some pixels of buildings, and gaps between buildings had zero (0) SPDI value. In non-built-up areas, some pixels of non-buildings, such as tall trees and highway overpasses, had high SPDI values.

For the two stereo pairs, SPDI comparisons are shown in Figures 12 and 13, respectively. In the legend, “BU” and “NBU” indicate the SPDI percentage in built-up and non-built-up areas, respectively. “Truth” and “Result” indicate the SPDI percentage in ground truth and identification results, respectively. “1” and “2” indicate the SPDI percentage in Image 1 and Image 2, respectively. The values on the horizontal axis are the specified thresholds for calculation of the SPDI percentage. Ten thresholds were specified for SPDI comparisons, ranging from 0.1 to 1.0 with a step of 0.1. For each threshold, the percentages of the SPDI values were calculated for the true and identified built-up and non-built-up areas, respectively for two images, for a total of four bins and eight values.

As can be seen from Figures 12 and 13, SPDI percentages in identified built-up areas were larger than that in non-built-up areas at ten thresholds. The sum of percentage in built-up and non-built-up areas was not 100% even when the threshold is the maximum SPDI value (1.0). This is because pixels having zero (0) SPDI value were not included in the calculation. The maximum difference between SPDI percentages in identified built-up areas and in non-built-up areas was 25.2% for stereo pair 1, and 35.0% for stereo pair 2; thus, the SPDI has robust effectiveness because of these large differences. The maximum difference between SPDI percentages for the identified built-up area results and for the ground truth data was 0.5% for stereo pair 1, and 3.7% for stereo pair 2; thus, the SPDI has robust reliability because of these small differences. The maximum difference between SPDI percentages in Image 1 and in Image 2 was 0.7% for stereo pair 1, and 1.7% for stereo pair 2; thus, the SPDI has robust consistency because of these small differences. The maximum percentage of areas identified as non-built-up and the percentage differences were lower in stereo pair 1 than stereo pair 2. Therefore,

areas identified as built-up in stereo pair 1 were more accurate than those in stereo pair 2. This aligns with the accuracy comparison of the two stereo pairs.

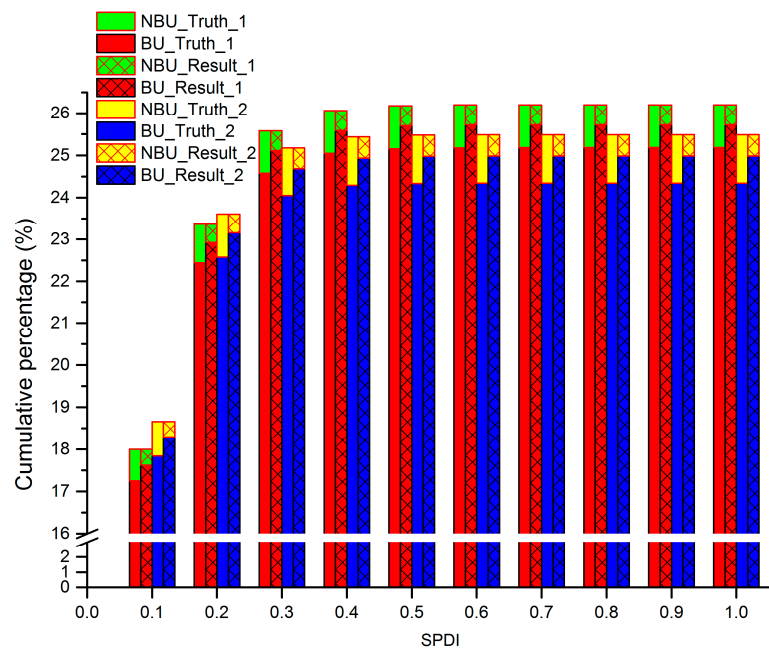


Figure 12. SPDI comparison for stereo pair 1.

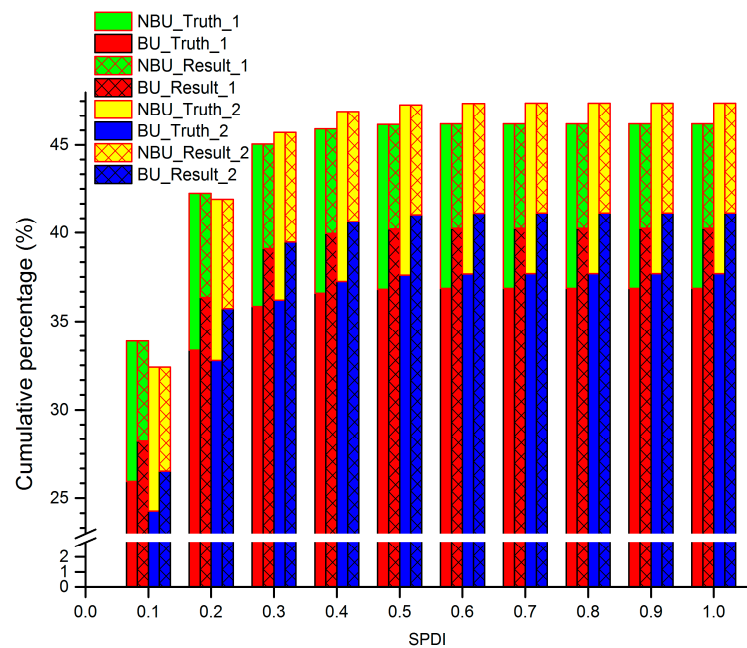


Figure 13. SPDI comparison for stereo pair 2.

## 4. Discussion

### 4.1. Performance Analysis of SPDI Indicating Built-Up Areas in Stereo Images

The proposed SPDI is a new effective height feature indicating built-up areas in stereo imagery. As shown in the experimental results for two stereo pairs covering different scenes with diverse urban settings, the SPDI effectively differentiates between built-up and non-built-up areas. The SPDI delivers a robust effectiveness measure, since the difference between SPDI percentages in built-up



areas and in non-built-up areas was very large. Given that the difference between SPDI percentages for results and for ground truth data was very small, the SPDI delivers a robust reliability measure. The percentage difference between two SPDI images of a stereo pair was very small, and therefore the SPDI delivers a robust consistency measure. In future research, the SPDI will be extensively tested on more stereo images covering different scenes in more diverse urban settings (e.g., build shape, spatial pattern, land use), including forested areas within or adjacent to urban areas, and hilly areas with dense and high slopes. In these circumstances, it may be very challenging to calculate correct disparity gradient values, detect meaningful interesting line segments, and obtain unaffected SPDI values. In these instances, planar shape, texture, and spectral features (e.g., Gabor [22]) can be used together with the SPDI so as to achieve higher accuracy of built-up areas.

The SPDI still has some limitations, and will be improved in the future. As can be seen from Figures 10 and 11, the SPDI performance seems to vary with land cover or land use types. Some pixels of large buildings have zero (0) SPDI values, since some meaningful interesting points are missed or mismatched during SPDI calculation. Some pixels for non-building objects (e.g., tall trees, and highway overpasses) have high SPDI values, since some meaningless interesting points were detected and then mismatched during SPDI calculation. SPDI values of these kinds of pixels will be improved by determining more appropriate parameters for SPDI calculations. They can be set based on the range of building heights and lengths in a specified scene, or set based on the heights and lengths of most buildings in a region to apply to all scenes. These pixels likely have little effect on identifying built-up areas, since only neighboring pixels with high SPDI values are significant in the process. Furthermore, the computational efficiency and automation of SPDI calculation will be promoted.

The SPDI has some advantages compared with existing planar and height features used for detecting built-up areas. The SPDI does not suffer from the influence of within-class spectral variation and between-class spectral confusion in remotely sensing imagery [10], which degrades the performance of built-up area detection when using planar texture, shape, and spectral features (e.g., Gabor [22]). The SPDI calculation considers the disparity gradient unlike terrain slopes, building heights or nDSM [34]. Moreover, the SPDI calculation needs no orthorectification processing on raw images of stereo imagery. Thus, the SPDI can be integrated with planar features extracted from raw images to improve built-up area detection accuracy. The SPDI calculation does not make assumptions about object shapes, and therefore the SPDI is suitable for both regular and irregular objects. The SPDI could be extended to detect individual buildings with their footprint borders.

#### *4.2. Performance Analysis of Stereo Images for Built-Up Area Detection*

Stereo images have an advantage over single remotely sensed images when they are used to detect built-up areas. Stereo images contain inherent height information unlike single images. Experimental results for two stereo pairs show that our methods achieve higher accuracy of built-up area identification results than the traditional method [22] for single images. In single images, within-class spectral variation and between-class spectral confusion degrades built-up detection performance when using planar features [10] (e.g., Gabor [22]). The performance can be increased by using height information achieved from auxiliary height data (e.g., LiDAR data). The auxiliary data must be acquired around the same time as image acquisition. Otherwise, built-up area detection accuracy will be compromised. In contrast, stereo imagery incorporates both planar and height information. There is no need to require auxiliary height data when detecting built-up areas from stereo imagery. For both high-resolution stereo images and single images, built-up area identification results may not include areas comprised of wide roads with poor texture. A combination of height and planar features could improve the accuracy of built-up area identification results.

Our study introduces a new method of detecting built-up areas from stereo images using stereo-extracted disparity information. There is no need to produce DSMs, which are required by existing methods of detecting built-up areas from stereo images [13,36]. Experimental results for two stereo pairs show that our methods achieve higher accuracy of built-up area identification results than

two other widely-applied DSM-based methods [13,36]. The disparity information is used not only to calculate SPDI values, but also to align raw inconsistent built-up areas detected separately from two images of a stereo pair, achieving consistent and reliable built-up area results. Given the significance of disparity information for our study, the accuracy of stereo-extracted disparity images needs to be evaluated if the ground truth data are available. Disparity images can be generated with the state-of-art epipolar image production method and image matching method.

Our method applies to stereo pairs and triplets acquired by spaceborne or airborne sensors. Although our method is designed for stereo pairs, it is also suitable for stereo triplets by establishing two stereo pairs with a conjunct image. Our method uses a pair-wise matching method for stereo pairs, and can be improved with multiview matching [55] so as to be more suitable for stereo triplets. Our method also applies to heterotype stereo images comprising of two or three images acquired by different sensors or at different time. However, our method may achieve a lower accuracy of built-up area identification results for heterotype stereo images than homotype stereo images, since heterotype stereo images have a lower accuracy of image matching [49].

## 5. Conclusions

A new Stereo Pair Disparity Index (SPDI) for indicating built-up areas in high-resolution stereo images was calculated from stereo-extracted disparity information. As shown in the experimental results for two stereo pairs covering different scenes with diverse urban settings, the SPDI effectively differentiates between built-up and non-built-up areas. The SPDI does not suffer from the influence of within-class spectral variation and between-class spectral confusion, unlike planar texture, shape, and spectral features. Further, a new method of detecting built-up areas from stereo pairs is proposed based on the SPDI, using disparity information to establish the relationship between two images of a stereo pair. As shown in the experimental results for two stereo pairs, our method achieves higher accuracy built-up area results from stereo images than the traditional method for single images, and two other widely-applied DSM-based methods for stereo images. Our approach is suitable for spaceborne and airborne stereo pairs and triplets. Our research introduces a new effective height feature (SPDI) for detecting built-up areas from stereo imagery with no need for DSMs. In the future, we will determine more appropriate parameters in SPDI calculation, improve automation and computational efficiency, and combine more planar features to achieve higher accuracy of built-up area identification results, with more extensive testing on more stereo images covering different scenes in more diverse urban settings.

**Acknowledgments:** This work was supported by the National Key Basic Research and Development Program of China under Grant 2012CB719906, the Open Research Fund of State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing under Grant 16E03, and the Self-determined Research Funds of CCNU from the Colleges' Basic Research and Operation of MOE under Grant CCNU16A05048. Special thanks are given to the data providers for the provision of the stereo datasets: Airbus Defense and Space for the Pleiades-1A imagery, and Digital Globe for the Worldview-1 imagery. The authors would also like to thank Beril Sirmacek for providing her code for built-up area detection as a comparison, Stephen C. McClure for language assistance, and the editors and anonymous reviewers for their very constructive comments.

**Author Contributions:** Feifei Peng designed and performed the experiments, and wrote the manuscript. Jianya Gong provided the original idea for this study. Le Wang revised the manuscript. Huayi Wu and Pengcheng Liu contributed to the discussion of the experiment results. All authors read and approved the final manuscript.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Shi, K.; Huang, C.; Yu, B.; Yin, B.; Huang, Y.; Wu, J. Evaluation of NPP-VIIRS night-time light composite data for extracting built-up urban areas. *Remote Sens. Lett.* **2014**, *5*, 358–366. [[CrossRef](#)]
2. Miyazaki, H.; Iwao, K.; Shibasaki, R. Development of a new ground truth database for global urban area mapping from a gazetteer. *Remote Sens.* **2011**, *3*, 1177–1187. [[CrossRef](#)]



3. Goldblatt, R.; You, W.; Hanson, G.; Khandelwal, A. Detecting the boundaries of urban areas in India: A dataset for pixel-based image classification in Google Earth Engine. *Remote Sens.* **2016**, *8*, 634. [[CrossRef](#)]
4. Wulder, M.A.; Coops, N.C. Satellites: Make earth observations open access. *Nature* **2014**, *513*, 30–31. [[CrossRef](#)] [[PubMed](#)]
5. Gamba, P.; Stilla, U.; Juergens, C.; Maktav, D. Foreword to the special issue on human settlement monitoring using multiple earth observation data. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2012**, *5*, 1071–1075. [[CrossRef](#)]
6. Pesaresi, M.; Huadong, G.; Blaes, X.; Ehrlich, D.; Ferri, S.; Gueguen, L.; Halkia, M.; Kauffmann, M.; Kemper, T.; Lu, L.; et al. A global human settlement layer from optical HR/VHR RS data: Concept and first results. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2013**, *6*, 2102–2131. [[CrossRef](#)]
7. Gamba, P. Human settlements: A global challenge for EO data processing and interpretation. *Proc. IEEE* **2013**, *101*, 570–581. [[CrossRef](#)]
8. Pesaresi, M.; Gerhardinger, A.; Kayitakire, F. A robust built-up area presence index by anisotropic rotation-invariant textural measure. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2008**, *1*, 180–192. [[CrossRef](#)]
9. Pesaresi, M.; Ehrlich, D.; Caravaggi, I.; Kauffmann, M.; Louvrier, C. Toward global automatic built-up area recognition using optical VHR imagery. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2011**, *4*, 923–934. [[CrossRef](#)]
10. Yan, W.Y.; Shaker, A.; El-Ashmawy, N. Urban land cover classification using airborne LiDAR data: A review. *Remote Sens. Environ.* **2015**, *158*, 295–310. [[CrossRef](#)]
11. Peng, F.; Wang, L.; Gong, J.; Wu, H. Development of a framework for stereo image retrieval with both height and planar features. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2015**, *8*, 800–815. [[CrossRef](#)]
12. Miyazaki, H.; Xiaowei, S.; Iwao, K.; Shibasaki, R. An automated method for global urban area mapping by integrating Aster satellite images and GIS data. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2013**, *6*, 1004–1019. [[CrossRef](#)]
13. Tong, X.; Liu, S.; Weng, Q. Geometric processing of quickbird stereo imageries for urban land use mapping: A case study in Shanghai, China. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2009**, *2*, 61–66. [[CrossRef](#)]
14. Pandey, A.; Kumar, A.; Jeyaseelan, A.T. Urban built-up area assessment of Ranchi township using Cartosat-1 stereopairs satellite images. *J. Indian Soc. Remote Sens.* **2013**, *41*, 141–155. [[CrossRef](#)]
15. Graesser, J.; Cheriadat, A.; Vatsavai, R.R.; Chandola, V.; Long, J.; Bright, E. Image based characterization of formal and informal neighborhoods in an urban landscape. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2012**, *5*, 1164–1176. [[CrossRef](#)]
16. Zha, Y.; Gao, J.; Ni, S. Use of normalized difference built-up index in automatically mapping urban areas from TM imagery. *Int. J. Remote Sens.* **2003**, *24*, 583–594. [[CrossRef](#)]
17. He, C.; Shi, P.; Xie, D.; Zhao, Y. Improving the normalized difference built-up index to map urban built-up areas using a semiautomatic segmentation approach. *Remote Sens. Lett.* **2010**, *1*, 213–221. [[CrossRef](#)]
18. Xu, H. A new index for delineating built-up land features in satellite imagery. *Int. J. Remote Sens.* **2008**, *29*, 4269–4276. [[CrossRef](#)]
19. Varshney, A.; Rajesh, E. A comparative study of built-up index approaches for automated extraction of built-up regions from remote sensing data. *J. Indian Soc. Remote Sens.* **2014**, *42*, 659–663. [[CrossRef](#)]
20. Sirmacek, B.; Unsalan, C. Urban-area and building detection using SIFT keypoints and graph theory. *IEEE Trans. Geosci. Remote Sens.* **2009**, *47*, 1156–1167. [[CrossRef](#)]
21. Qi, K.; Wu, H.; Shen, C.; Gong, J. Land-use scene classification in high-resolution remote sensing images using improved correlatons. *IEEE Geosci. Remote Sens. Lett.* **2015**, *12*, 2403–2407.
22. Sirmacek, B.; Unsalan, C. Urban area detection using local feature points and spatial voting. *IEEE Geosci. Remote Sens. Lett.* **2010**, *7*, 146–150. [[CrossRef](#)]
23. Kovacs, A.; Sziranyi, T. Improved Harris feature point set for orientation-sensitive urban-area detection in aerial images. *IEEE Geosci. Remote Sens. Lett.* **2013**, *10*, 796–800. [[CrossRef](#)]
24. Tao, C.; Tan, Y.; Zou, Z.; Tian, J. Unsupervised detection of built-up areas from multiple high-resolution remote sensing images. *IEEE Geosci. Remote Sens. Lett.* **2013**, *10*, 1300–1304. [[CrossRef](#)]
25. Li, Y.; Tan, Y.; Deng, J.; Wen, Q.; Tian, J. Cauchy graph embedding optimization for built-up areas detection from high-resolution remote sensing images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2015**, *8*, 2078–2096. [[CrossRef](#)]

26. Unsalan, C.; Boyer, K.L. Classifying land development in high-resolution panchromatic satellite images using straight-line statistics. *IEEE Trans. Geosci. Remote Sens.* **2004**, *42*, 907–919. [CrossRef]
27. Unsalan, C.; Boyer, K.L. A theoretical and experimental investigation of graph theoretical measures for land development in satellite imagery. *IEEE Trans. Pattern Anal. Mach. Intell.* **2005**, *27*, 575–589. [CrossRef] [PubMed]
28. Huang, X.; Zhang, L. Morphological building/shadow index for building extraction from high-resolution imagery over urban areas. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2012**, *5*, 161–172. [CrossRef]
29. Iannelli, G.; Lisini, G.; Dell, F.; Feitosa, R.; Costa, G.; Gamba, P. Urban area extent extraction in spaceborne HR and VHR data using multi-resolution features. *Sensors* **2014**, *14*, 18337–18352. [CrossRef] [PubMed]
30. Zhang, J.; Li, P.; Wang, J. Urban built-up area extraction from landsat TM/ETM+ images using spectral information and multivariate texture. *Remote Sens.* **2014**, *6*, 7339–7559. [CrossRef]
31. Chen, J.; Chen, J.; Liao, A.; Cao, X.; Chen, L.; Chen, X.; He, C.; Han, G.; Peng, S.; Lu, M.; et al. Global land cover mapping at 30 m resolution: A POK-based operational approach. *ISPRS J. Photogramm. Remote Sens.* **2015**, *103*, 7–27. [CrossRef]
32. Ban, Y.; Jacob, A.; Gamba, P. Spaceborne SAR data for global urban mapping at 30 m resolution using a robust urban extractor. *ISPRS J. Photogramm. Remote Sens.* **2015**, *103*, 28–37. [CrossRef]
33. Yu, B.; Liu, H.; Wu, J.; Hu, Y.; Zhang, L. Automated derivation of urban building density information using airborne LiDAR data and object-based method. *Landsc. Urban Plan.* **2010**, *98*, 210–219. [CrossRef]
34. Wu, B.; Yu, B.; Wu, Q.; Yao, S.; Zhao, F.; Mao, W.; Wu, J. A graph-based approach for 3D building model reconstruction from airborne LiDAR point clouds. *Remote Sens.* **2017**, *9*, 92. [CrossRef]
35. JAXA. ALOS global digital surface model “ALOS world 3D-30m” (AW3D30). Available online: <http://www.eorc.jaxa.jp/ALOS/en/aw3d30/index.htm> (accessed on 12 March 2017).
36. Peng, F.; Gong, J.; Wang, L.; Wu, H.; Yang, J. Impact of building heights on 3D urban density estimation from spaceborne stereo imagery. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2016**, *XLI-B3*, 677–683. [CrossRef]
37. Yu, S.; Yu, B.; Song, W.; Wu, B.; Zhou, J.; Huang, Y.; Wu, J.; Zhao, F.; Mao, W. View-based greenery: A three-dimensional assessment of city buildings’ green visibility using floor green view index. *Landsc. Urban Plan.* **2016**, *152*, 13–26. [CrossRef]
38. Turlapaty, A.; Gokaraju, B.; Qian, D.; Younan, N.H.; Aanstoos, J.V. A hybrid approach for building extraction from spaceborne multi-angular optical imagery. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2012**, *5*, 89–100. [CrossRef]
39. Duan, Y.; Shao, X.; Shi, Y.; Miyazaki, H.; Iwao, K.; Shibasaki, R. Unsupervised global urban area mapping via automatic labeling from Aster and Palsar satellite images. *Remote Sens.* **2015**, *7*, 2171–2192. [CrossRef]
40. Rau, J.; Jhan, J.; Hsu, Y. Analysis of oblique aerial images for land cover and point cloud classification in an urban environment. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 1304–1319. [CrossRef]
41. Suliman, A.; Zhang, Y. Development of line-of-sight digital surface model for co-registering off-nadir VHR satellite imagery with elevation data. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2015**, *8*, 1913–1923. [CrossRef]
42. Longbotham, N.; Chaapel, C.; Bleiler, L.; Padwick, C.; Emery, W.J.; Pacifici, F. Very high resolution multiangle urban classification analysis. *IEEE Trans. Geosci. Remote Sens.* **2012**, *50*, 1155–1170. [CrossRef]
43. Wang, M.; Hu, F.; Li, J. Epipolar resampling of linear pushbroom satellite imagery by a new epipolarity model. *ISPRS J. Photogramm. Remote Sens.* **2011**, *66*, 347–355. [CrossRef]
44. Oh, J.; Lee, W.H.; Toth, C.K.; Grejner-Brzezinska, D.A.; Lee, C. A piecewise approach to epipolar resampling of pushbroom satellite images based on RPC. *Photogramm. Eng. Remote Sens.* **2010**, *76*, 1353–1363. [CrossRef]
45. Hirschmuller, H. Stereo processing by semiglobal matching and mutual information. *IEEE Trans. Pattern Anal. Mach. Intell.* **2008**, *30*, 328–341. [CrossRef] [PubMed]
46. Alobeid, A.; Jacobsen, K.; Heipke, C. Comparison of matching algorithms for DSM generation in urban areas from IKONOS imagery. *Photogramm. Eng. Remote Sens.* **2010**, *76*, 1041–1050. [CrossRef]
47. Telea, A. An image inpainting technique based on the fast marching method. *J. Graph. Tools* **2004**, *9*, 25–36. [CrossRef]
48. Morgan, G.L.K.; Jian-Guo, L.; Hongshi, Y. Precise subpixel disparity measurement from very narrow baseline stereo. *IEEE Trans. Geosci. Remote Sens.* **2010**, *48*, 3424–3433. [CrossRef]

49. Choi, S.Y.; Kang, J.M.; Shin, D.S. A comparison of accuracies of the RPC models: Homo- and heterotype stereo pairs of Geoeye and Worldview images. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* **2012**, *I-4*, 65–69. [[CrossRef](#)]
50. Jing, L.; Yong-she, S.; Jian-rong, W.; Jun-feng, Y. The research and design of the base-height ratio for the three linear array camera of satellite photogrammetry. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2008**, XXXVII Pt B1, 757–760.
51. Williamson, D.F.; Parker, R.A.; Kendrick, J.S. The box plot: A simple visual method to interpret data. *Ann. Intern. Med.* **1989**, *110*, 916–921. [[CrossRef](#)] [[PubMed](#)]
52. Yan, H.; Weibel, R. An algorithm for point cluster generalization based on the Voronoi diagram. *Comput. Geosci.* **2008**, *34*, 939–954. [[CrossRef](#)]
53. Lin, C.; Nevatia, R. Building detection and description from a single intensity image. *Comput. Vis. Image Underst.* **1998**, *72*, 101–121. [[CrossRef](#)]
54. Tarantino, E.; Figorito, B. Extracting buildings from true color stereo aerial images using a decision making strategy. *Remote Sens.* **2011**, *3*, 1553. [[CrossRef](#)]
55. Zhang, L.; Gruen, A. Multi-image matching for DSM generation from IKONOS imagery. *ISPRS J. Photogramm. Remote Sens.* **2006**, *60*, 195–211. [[CrossRef](#)]



© 2017 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).