

## Research Article

# A Subpixel Matching Method for Stereovision of Narrow Baseline Remotely Sensed Imagery

Ning Ma,<sup>1,2</sup> Peng-fei Sun,<sup>1</sup> Yu-bo Men,<sup>2</sup> Chao-guang Men,<sup>2</sup> and Xiang Li<sup>2</sup>

<sup>1</sup>College of Computer Science and Information Engineering, Harbin Normal University, Normal University Road 1, Harbin 150001, China

<sup>2</sup>College of Computer Science and Technology, Harbin Engineering University, Nantong Street 145, Harbin 150001, China

Correspondence should be addressed to Chao-guang Men; menchaoguang@hrbeu.edu.cn

Received 24 August 2016; Revised 9 January 2017; Accepted 26 January 2017; Published 23 February 2017

Academic Editor: Ram Avtar

Copyright © 2017 Ning Ma et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In this paper, an accurate and efficient image matching method based on phase correlation is proposed to estimate disparity with subpixel precision, which is used for the stereovision of narrow baseline remotely sensed images. The multistep strategy is adopted in our technical frame; thus the disparity estimation is divided into two steps: integer-pixel prematching and subpixel matching. Firstly, integer-pixel disparity is estimated by employing a cross-based local matching method. Then the relationship of corresponding points is established under the guidance of integer-pixel disparity. The subimages are extracted through selecting the corresponding points as the center. Finally, the subpixel disparity is obtained by matching the subimages utilizing a novel variant of phase correlation approach. The experiment results show that the proposed method can match different kinds of large-sized narrow baseline remotely sensed images and estimate disparity with subpixel precision automatically.

## 1. Introduction

Stereovision is an advanced task in remote sensing and photogrammetry [1]. The aim of stereovision is to estimate the disparity through matching two or more images of same scene in different views and extract digital elevation model (DEM) through the disparity [2]. Intuitively, the disparity represents the displacement vectors between corresponding pixels that horizontally shift from the left image to the right image [3]. The stereovision of narrow baseline remotely sensed imagery is a new research hotspot for stereovision in recent years [4]. Automatic subpixel image matching is the essential technique for narrow baseline stereovision [5]. We will brief review the stereovision with different baseline and subpixel image matching method in the remaining part of this section, respectively.

*1.1. Principle of Stereovision.* The process of DEM extracting can be represented by a stereovision model [6], as shown in Figure 1.  $B$  and  $H$  express the baseline and the altitude of a satellite.  $L$  and  $R$  are the left and right images captured by

the satellite from different views;  $h$  represents the height of ground object to be measured.  $A$  and  $B$  are the 3D points of the real world scene, which are projected to the three 2D locations  $a$ ,  $a'$ , and  $b'$  in  $L$  and  $R$ . The 2D points of  $A$  and  $B$  in  $L$  are coincident; thus the disparity  $d$  of the two corresponding points  $a$  and  $a'$  equals the distance between  $a'$  and  $b'$  in  $R$ .  $D$  represents the mapping of  $d$  in the 3D scene; according to the similar triangles of geometry principle, the equation  $D/B = h/(H - h)$  is established. Assume that the ground sample distance (GSD) of the sensed image is  $G$  (meters/pixel); the height  $h$  of ground object can be generated:  $h = dG/(B/(H - h))$ . Because the altitude of satellite is much larger than the height of ground object, that is,  $H \gg h$ , the height  $h$  can be approximated as  $h \approx dG/(B/H)$ . It can be known that the precision of  $h$  is proportional to the precision of disparity and GSD and is inversely proportional to the  $B/H$  ratio. In most real sensors,  $H$  and  $G$  are limited and fixed; therefore, only two aspects can be used to enhance the precision of  $h$ : (1) Improve the precision of disparity to subpixel level [7]; (2) increase the length of baseline, that is, increasing the  $B/H$  ratio. In traditional stereovision of remotely sensed imagery,

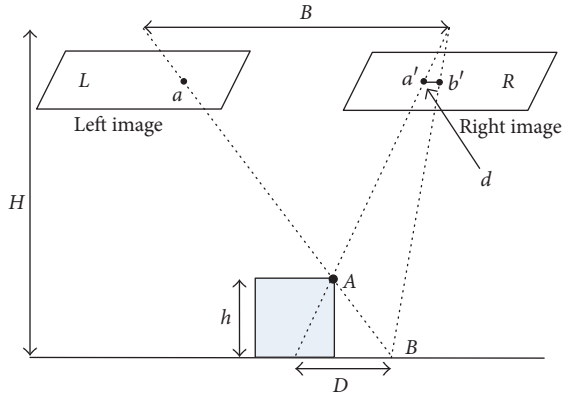


FIGURE 1: Stereovision model.

$B/H$  ratio is usually set to about 1, called stereovision of wide baseline remotely sensed imagery. Even if the disparity precision is only up to the integer-pixel level, the DEM can be extracted accurately.

### 1.2. Comparison of Wide Baseline and Narrow Baseline Stereo.

The stereovision of wide baseline remotely sensed imagery can not perform well in urban areas with dense buildings [8]. The captured process of the wide baseline remotely sensed image pair is shown in Figure 2(a). Due to the long baseline and large viewing angle, there is a large occlusion area and large geometric distortion between the images. In addition, the interval time of two imaging processes is long; the illumination condition has changed greatly. The above interference factors will lead to a high false matching rate in stereovision of wide baseline remotely sensed imagery. Therefore, it requires a lot of manual and postcorrection to extract DEM [9]. In order to solve this problem of DEM extraction in urban areas, stereovision of narrow baseline remotely sensed imagery technique has emerged and gradually become a new research focus in stereovision and remote sensing [10, 11]. The captured process of the narrow baseline remotely sensed image pair is shown in Figure 2(b). Due to the short baseline and small viewing angle, there is a small occlusion area and little geometric distortion between images. When  $B/H \rightarrow 0$  (i.e., a very narrow baseline), then occlusions will be minimized, essentially because each image view becomes more geometrically similar. In addition, the interval time of two imaging processes is short; it can be considered that the illumination condition is similar to the same time. Due to above advantages, the stereovision of narrow baseline remotely sensed imagery is suitable for extracting DEM in the urban areas [8]. Two sets of simulated remotely sensed image pairs provided by Beijing Institute of Space Mechanics and Electricity are shown in Figure 3. Figure 3(a) is simulated wide baseline remotely sensed image pair with a  $B/H$  ratio of 1, and Figure 3(b) is simulated narrow baseline remotely sensed image pair with a  $B/H$  ratio of 0.05. From the comparison of Figures 3(a) and 3(b), we can see that the occlusion range and the geometric distortion of narrow baseline image pair are less, and there is a higher degree of similarity between left and right image. It is beneficial to

improve the accuracy of image matching; thus fully automatic DEM extraction can be achieved through stereovision of narrow baseline remotely sensed imagery.

However the reduction of  $B/H$  ratio will inevitably lead to the reduction of precision for DEM extracting. If the stereovision of narrow baseline is going to extract the DEM as precise as the stereovision of wide baseline, the disparity needs to achieve  $1/T$  pixel precision, where  $T$  is the ratio of wide baseline and narrow baseline [6]. For example, when  $B/H$  is 0.05, the precision of disparity needs to achieve  $1/20$  pixels to satisfy the DEM extraction. Therefore, the key technique for the stereovision of narrow baseline remotely sensed imagery is the high precision subpixel image matching [5, 7].

### 1.3. A Survey of Subpixel Image Matching Method.

Automatic subpixel image matching is one of the most essential techniques in stereovision. It can be divided into three major categories: interpolation based method, fitting based method, and phase correlation method [12]. Interpolation based method includes the original image interpolation method and the matching cost interpolation method. The matching cost interpolation method is the representative algorithm, which combines the advantages of high efficiency and accuracy [13, 14]. The initial cost volume is interpolated by various interpolation function, and the subpixel disparity is estimated by searching the extremum of interpolated cost volume. At present, this method is usually used for the auxiliary system of driverless vehicles, robot navigation, and unmanned aerial vehicle. The original image interpolation method significantly increases the computing load, limits the possible precision to the chosen upsampling rate, and also may introduce interpolation artifacts [5]. The fitting based method generally postprocess the cost volume or disparity plane by fitting method to estimate the subpixel disparity. The cost volume fitting method fits the peak neighborhood of cost volume into a parabola; then the subpixel disparity is estimated by searching the extreme of parabola [15, 16]. The disparity plane fitting method models the disparity field by segmentation constraints, and subpixel disparity is obtained by fitting disparity plane [17–20]. The fitting based method is efficient, but the disparity precision is low. Phase correlation method provides high efficiency and accuracy via fast Fourier transform and other supplementary approaches under ideal conditions. In general, the main peak location of the inverse Fourier transform of the normalized cross-power spectrum was interpolated with the parabolic, Gaussian, and sinc functions to get the subpixel disparity [21–24].

In this paper, a subpixel image matching method based on phase correlation is proposed to estimate the disparity with subpixel precision for stereovision of narrow baseline remotely sensed imagery. The rest of this paper is organized as follows. The principle of phase correlation is introduced briefly in Section 2. A novel improved phase correlation method with subpixel precision is described in detail in Section 3. The complete algorithm is described in Section 4. Experimental evaluation is presented in Section 5. Finally, the paper is concluded in Section 6.

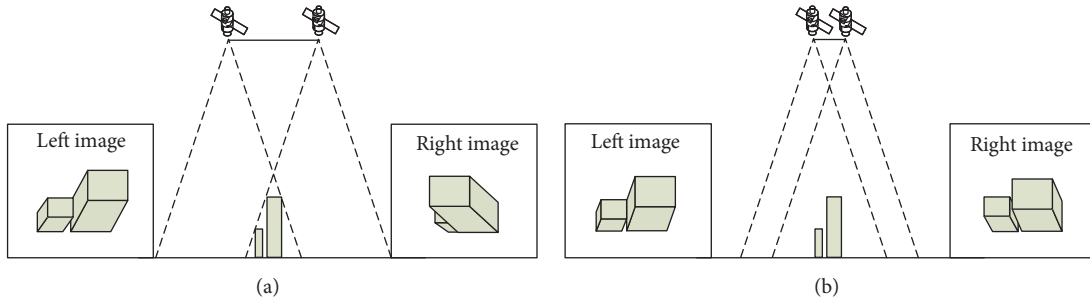


FIGURE 2: The captured process of the remotely sensed image pair. (a) Traditional wide baseline stereo configuration and (b) narrow baseline stereo configuration.

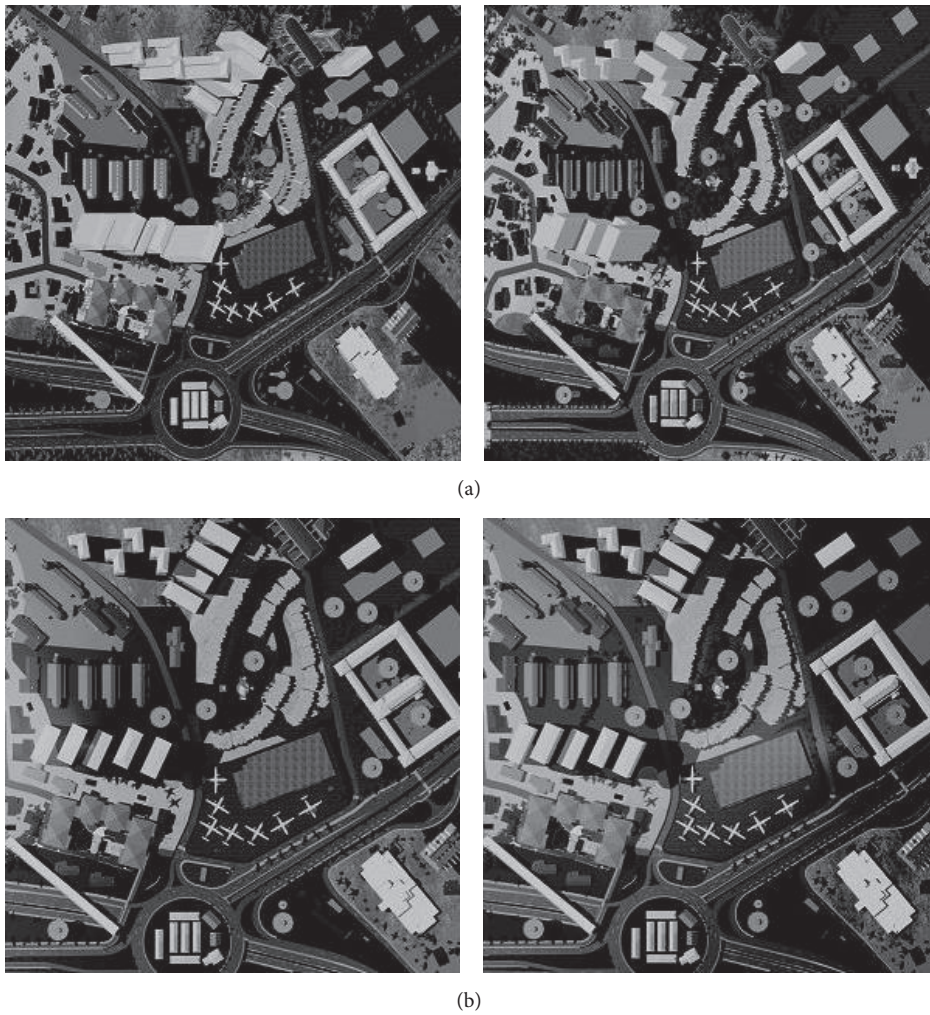


FIGURE 3: Two sets of simulated remotely sensed image pairs provided by Beijing Institute of Space Mechanics and Electricity. (a) Simulated wide baseline remotely sensed image pair with a  $B/H$  ratio of 1 and (b) simulated narrow baseline remotely sensed image pair with a  $B/H$  ratio of 0.05.

## 2. Phase Correlation

Phase correlation method is based on the well-known Fourier-domain shift property. This states that a translation between two images in the spatial domain will be expressed

in the frequency domain as a linear phase difference between their Fourier transform [5]. Let  $f_1$  and  $f_2$  be two images with translation relationship, the size of image is  $M \times N$ ,  $f_1 = f(x, y)$ ,  $f_2 = f(x + \Delta x, y + \Delta y)$ , and  $(\Delta x, \Delta y)$  are horizontal and vertical translations, respectively. Let  $F_1(u, v)$

and  $F_2(u, v)$  denote the 2D discrete Fourier transform (DFT) of two images:

$$F_1(u, v) = \frac{1}{MN} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x, y) \exp\left(-j2\pi\left(\frac{ux}{M} + \frac{vy}{N}\right)\right), \quad (1)$$

$$F_2(u, v) = F_1(u, v) \exp\left(-j2\pi\left(\frac{u\Delta x}{M} + \frac{v\Delta y}{N}\right)\right),$$

where  $x$  and  $y$  are spatial domain variables,  $x = 0, 1, 2, \dots, M-1$ ,  $y = 0, 1, 2, \dots, N-1$ .  $u$  and  $v$  are frequency domain variables,  $u = 0, 1, 2, \dots, M-1$ ,  $v = 0, 1, 2, \dots, N-1$ . The normalized cross-power spectrum  $Q(u, v)$  of two images is calculated as

$$Q(u, v) = \frac{F_1(u, v) F_2^*(u, v)}{|F_1(u, v) F_2^*(u, v)|} = \exp\left(-j2\pi\left(\frac{u\Delta x}{M} + \frac{v\Delta y}{N}\right)\right), \quad (2)$$

where  $F_2^*(u, v)$  is complex conjugate of  $F_2(u, v)$ . The phase correlation  $PC(x, y)$  of two images is calculated by 2D inverse discrete Fourier transforms (IDFT) of  $Q(u, v)$ :

$$PC(x, y) = F^{-1}(Q(u, v)) = \delta(x + \Delta x, y + \Delta y), \quad (3)$$

where  $F^{-1}(\cdot)$  is IDFT function.  $\delta(x + \Delta x, y + \Delta y)$  is a 2D pulse function which exists as peak at position  $(\Delta x, \Delta y)$ . The peak location  $(\Delta x, \Delta y)$  corresponds to the translation between images.

Phase correlation is a common method to estimate the translation between images. Only phase information of images is utilized; thus it is almost not affected by noise and radiation difference and more robustness than spatial domain method. But the digital image is a discrete function whose Fourier transform is a discrete transform, so only the integer-pixel translation ( $\text{int}(\Delta x), \text{int}(\Delta y)$ ) can be obtained by searching the peak location.

### 3. Improved Phase Correlation with Subpixel Precision

To estimate the subpixel translation, various algorithms have been proposed. In [23], Foroosh et al. achieved the subpixel estimation by approximating the Dirichlet function derived from the normalized cross-power spectrum. Nagashima et al. propose a fitting method to enhance the matching precision from integer-pixel level to subpixel level, which fits a curve surface by using the neighborhood data around phase correlation peak [21]. However, these methods only take into account the peak and its neighborhood; the distribution of phase correlation function is neglected; thus, the precision of translation estimation is low in stereovision of remotely sensed imagery. An improved phase correlation method used for subpixel estimation is designed in this section. The flow chart is shown in Figure 4. Due to the periodicity of DFT, an image can be considered to "wrap around" at an edge;

therefore, discontinuities, which are not supposed to exist in real world, occur at every border in 2D DFT computation [21]. A 2D Hanning window function is applied to reduce the effect of discontinuities at image borders. The 2D Hanning window function is defined as

$$w(x, y) = \frac{1 + \cos(\pi x/M)}{2} \frac{1 + \cos(\pi y/N)}{2}. \quad (4)$$

The subpixel translation between two images can be obtained by downsampling high resolution images with integer-pixel translation. Therefore, the phase correlation of two images can be approximated by 2D sinc function:

$$PC(x, y) \approx \frac{\sin(\pi(x + \Delta x))}{\pi(x + \Delta x)} \frac{\sin(\pi(y + \Delta y))}{\pi(y + \Delta y)}. \quad (5)$$

In theory, sinc function can approximate the distribution of phase correlation in ideal condition. However the phase correlation peak is decreased in remotely sensed images matching due to the interference of random noise, image aliasing, edge effects, and other influences. Therefore, the sinc function can not describe the distribution of phase correlation accurately. In our method, the sinc function is improved by introducing phase correlation coefficient:

$$PC(x, y) \approx \alpha \frac{\sin(\pi(x + \Delta x))}{\pi(x + \Delta x)} \frac{\sin(\pi(y + \Delta y))}{\pi(y + \Delta y)}, \quad (6)$$

where  $\alpha$  is the phase correlation coefficient,  $\alpha \leq 1$ . When the translation is integer-pixel and there is no interference factor,  $\alpha = 1$ ; when the translation is subpixel or there are interference factors,  $\alpha < 1$ . Due to this improved phase correlation method is applied to the stereovision of narrow baseline remotely sensed imagery, and the image pair is treated by epipolar rectification before matching. Thus, the phase correlation  $PC(x, y)$  can be separated in spatial domain, and a 1D function expression is given which only exists as translation in horizontal direction:

$$PC_x(x) \approx \alpha \frac{\sin(\pi(x + \Delta x))}{\pi(x + \Delta x)}. \quad (7)$$

The peak position  $(\Delta x, 0)$  is the translation which is expected to be estimated. To locate the high precision subpixel peak position, a peak evaluation method based on uniformly spaced sampling is proposed. Assume that  $p$  is the real peak position of sampled-data, as shown at the bottom left corner of Figure 4, and  $p - k$  and  $p + k$  are a pair of points located  $k$  pixels away from  $p$ . The phase correlation values  $PC_x(p - k)$ ,  $PC_x(p)$ , and  $PC_x(p + k)$  of three sample points used for peak evaluation are satisfying the following relationship:

$$\alpha \frac{\sin(\pi(p - k + \Delta x))}{\pi(p - k + \Delta x)} = PC_x(p - k) \quad (8a)$$

$$\alpha \frac{\sin(\pi(p + \Delta x))}{\pi(p + \Delta x)} = PC_x(p) \quad (8b)$$

$$\alpha \frac{\sin(\pi(p + k + \Delta x))}{\pi(p + k + \Delta x)} = PC_x(p + k). \quad (8c)$$

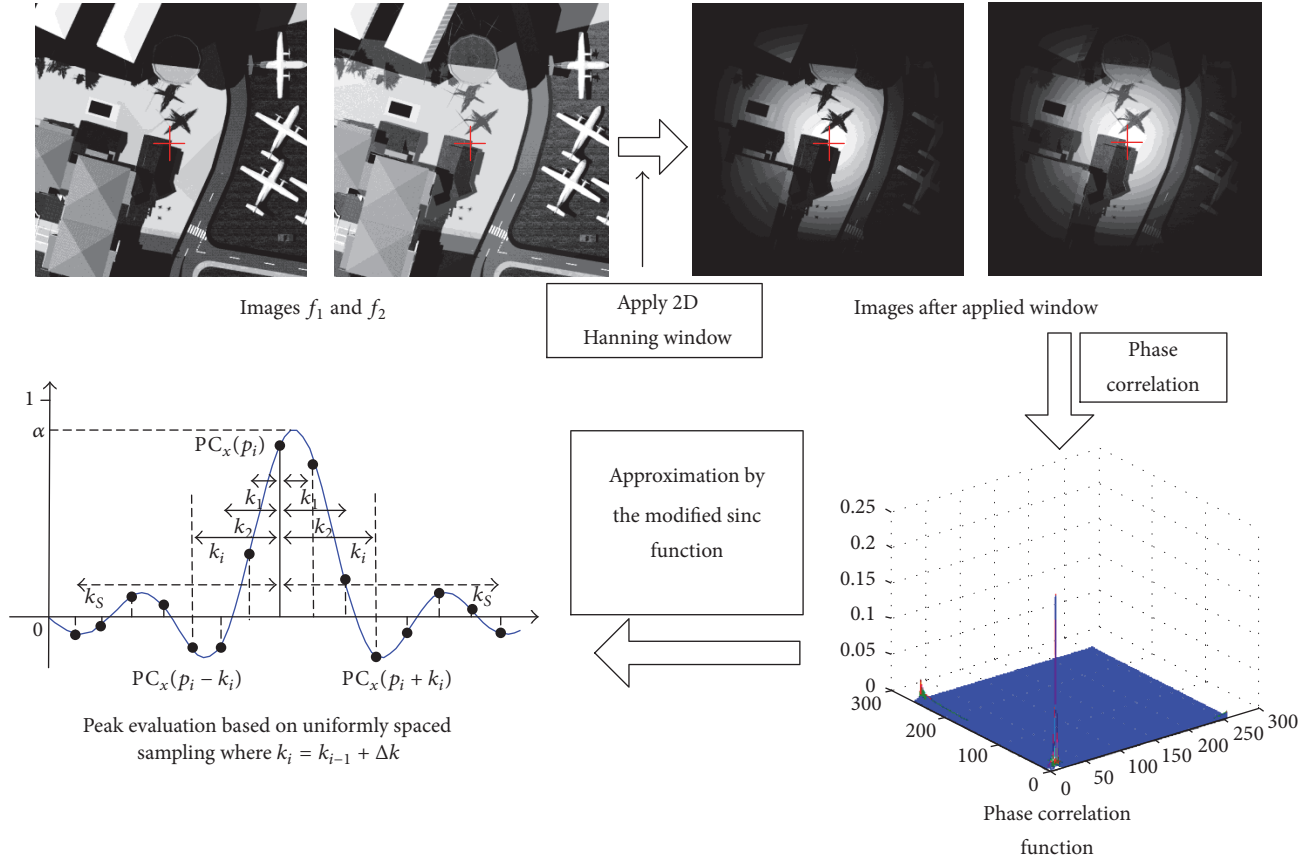


FIGURE 4: Flow chart of improved phase correlation image matching method.

Add (8c) to (8a); the above equation group can be rewritten as

$$\frac{\alpha}{\pi} (\sin(\pi(p + \Delta x))) = PC_x(p)(p + \Delta x) \quad (9a)$$

$$\begin{aligned} \frac{\alpha}{\pi} (\sin(\pi(p - k + \Delta x)) + \sin(\pi(p + k + \Delta x))) \\ = PC_x(p - k)(p - k + \Delta x) \end{aligned} \quad (9b)$$

$$+ PC_x(p + k)(p + k + \Delta x).$$

Equation (9b) can be rewritten by using the difference product formula:

$$\frac{\alpha}{\pi} (\sin(\pi(p + \Delta x))) = PC_x(p)(p + \Delta x) \quad (10a)$$

$$\begin{aligned} \frac{\alpha}{\pi} (2 \sin(\pi(p + \Delta x)) \cos(\pi k)) \\ = PC_x(p - k)(p - k + \Delta x) \\ + PC_x(p + k)(p + k + \Delta x). \end{aligned} \quad (10b)$$

Add (10a) to (10b); the equation group ((10a) and (10b)) can be rewritten as

$$\begin{aligned} PC_x(p - k)(p - k + \Delta x) + PC_x(p + k)(p + k + \Delta x) \\ = 2 \cdot PC_x(p)(p + \Delta x) \cos(\pi k). \end{aligned} \quad (11)$$

The above equation can be expressed as a linear expression of  $\Delta x$ :

$$\begin{aligned} (PC_x(p - k) + PC_x(p + k) - 2 \cdot \cos(\pi k) PC_x(p)) \\ \cdot \Delta x \\ = 2 \cdot p \cdot \cos(\pi k) PC_x(p) - (p - k) PC_x(p - k) \\ + (p + k) PC_x(p + k). \end{aligned} \quad (12)$$

Through (12), the horizontal translation  $\Delta x$  with subpixel precision can be estimated:

$$\Delta x = \frac{2 \cdot p \cdot \cos(\pi k) PC_x(p) - (p - k) PC_x(p - k) + (p + k) PC_x(p + k)}{PC_x(p - k) + PC_x(p + k) - 2 \cdot \cos(\pi k) PC_x(p)}. \quad (13)$$

To reduce the influence of noise, the peak evaluation is performed by selecting multiple sets of uniformly spaced sampled-data which are all the data on the  $x$  axis, and the least square is applied to solve the optimal translation. Assume that  $S$  sets of uniformly spaced sampled-data  $(p_i - k_i, p_i, p_i + k_i)_{i \in S}$  are selected for peak evaluation;  $p_i, k_i$  are determined as follows:  $p_1 = p_2 = \dots = p_S = \arg \max_x PC_x(x)$ ,  $k_1 = 1$ ,  $k_2 = k_1 + \Delta k, \dots, k_S = k_{S-1} + \Delta k$ , ( $\Delta k = 1$ ). Equation (12) can be rewritten as

$$\begin{bmatrix} a(p_1, k_1) \\ a(p_2, k_2) \\ \vdots \\ a(p_S, k_S) \end{bmatrix} \cdot \begin{bmatrix} \Delta x \\ \Delta x \\ \vdots \\ \Delta x \end{bmatrix} = \begin{bmatrix} b(p_1, k_1) \\ b(p_2, k_2) \\ \vdots \\ b(p_S, k_S) \end{bmatrix}, \quad (14)$$

where  $a(p_i, k_i)_{i \in S} = PC_x(p_i - k_i) + PC_x(p_i + k_i) - 2 \cdot \cos(\pi k_i) PC_x(p_i)$ , and  $b(p_i, k_i)_{i \in S} = 2 \cdot p_i \cdot \cos(\pi k_i) PC_x(p_i) - (p_i - k_i) PC_x(p_i - k_i) + (p_i + k_i) PC_x(p_i + k_i)$ . The objective of (14) is  $\min(\sum_{i=1}^S |b(p_i, k_i) - a(p_i, k_i) \cdot \Delta x|^2)$ . Assume that  $A = [a(p_1, k_1), a(p_2, k_2), \dots, a(p_S, k_S)]^T$  and  $B = [b(p_1, k_1), b(p_2, k_2), \dots, b(p_S, k_S)]^T$ , then the horizontal translation  $\Delta x$  with subpixel precision can be obtained by solving the least square linear equation:

$$\Delta x = (A^T A)^{-1} A^T B. \quad (15)$$

#### 4. Complete Subpixel Image Matching Algorithm

The input of our algorithm is rectified narrow baseline remotely sensed image pair, and output is the disparity with subpixel precision. Because the stability of phase correlation method is poor when the disparity search range is large [9], therefore, a multistep strategy is adopted in our technical frame, and the disparity estimation is divided into two steps: integer-pixel prematching and subpixel matching. The complete framework of our algorithm is shown in Figure 5. Firstly, integer-pixel disparity is estimated by employing an effective cross-based matching algorithm [25]. Then relationship of corresponding points is established under the guidance of integer-pixel disparity. The subimages are extracted through selecting the corresponding points as center. Finally, subpixel disparity is obtained by matching the subimages utilizing the improved phase correlation proposed in Section 3.

Assume that a narrow baseline remotely sensed image pair consists of left image  $L$  and right image  $R$ , which can be expressed as a classic stereo model  $L(x, y) = R(x + d(x, y), y)$ . The disparity  $d(x, y)$  is used to describe the geometric position variation between  $L$  and  $R$ , which consists of two parts, integer-pixel disparity  $d_z(x, y)$  and subpixel disparity  $d_s(x, y)$ ; that is,  $d(x, y) = d_z(x, y) + d_s(x, y)$ . The details of our subpixel image matching algorithm are described as follows from step (a) to step (g).

(a) Assume  $(x_0, y_0)$  is the pixel which is to be matched. In the first step, the cross-based support window of  $(x_0, y_0)$  is defined. Firstly, an upright cross is established for  $(x_0, y_0)$ . It consists of two orthogonal line segments: horizontal segment

$\varphi_H(x_0, y_0)$  and vertical segment  $\varphi_V(x_0, y_0)$ . A quadruple  $\{h^-, h^+, v^-, v^+\}$  denotes the left, right, up, and bottom arm length, respectively, and the length of each arm is determined by searching for the extreme point in that direction based on the color similarity. The upright cross skeleton is defined as

$$\begin{aligned} \varphi_H(x_0, y_0) &= \{(x, y) \mid x \in [x_0 - h^-, x_0 + h^+], y = y_0\}, \\ \varphi_V(x_0, y_0) &= \{(x, y) \mid x = x_0, y \in [y_0 - v^-, y_0 + v^+]\}. \end{aligned} \quad (16)$$

Then the cross-based support window  $\varphi(x_0, y_0)$  is defined based on the upright cross skeleton.  $\varphi(x_0, y_0)$  is a combination of each horizontal segment  $\varphi_H(x, y)$  where  $(x, y)$  traverses over the vertical segment  $\varphi_V(x_0, y_0)$ . The formula of  $\varphi(x_0, y_0)$  is defined as

$$\varphi(x_0, y_0) = \bigcup_{(x, y) \in \varphi_V(x_0, y_0)} \varphi_H(x, y). \quad (17)$$

(b) The matching cost of  $(x_0, y_0)$  is calculated. Firstly, the initial matching cost  $C_{x, y, d_z}$  of each pixel in the support window is calculated:

$$\begin{aligned} C_{x, y, d_z} &= (1 - \beta) \cdot \min(|L(x, y) - R(x + d_z, y)|, \tau_1) \\ &+ \beta \\ &\cdot \min(|\nabla_x L(x, y) - \nabla_x R(x + d_z, y)|, \tau_2), \end{aligned} \quad (18)$$

where  $\nabla_x$  is horizontal gradient,  $\beta$  balances the color and gradient terms, and  $\tau_1, \tau_2$  are truncation values. Then the matching cost  $C_{x_0, y_0, d_z}^A$  of  $(x_0, y_0)$  is aggregated:

$$C_{x_0, y_0, d_z}^A = \sum_{(x, y) \in \varphi(x_0, y_0)} C_{x, y, d_z}. \quad (19)$$

(c) The integer-pixel disparity  $d_z(x_0, y_0)$  of  $(x_0, y_0)$  is estimated by winner-takes-all strategy:

$$d_z(x_0, y_0) = \arg \min_{d_z \in S_d} C_{x_0, y_0, d_z}^A, \quad (20)$$

where  $S_d$  is the searching range of disparity.

(d) The relationship of corresponding points is established under the guidance of integer-pixel disparity. Assume that the integer-pixel disparity of pixel  $(x_0, y_0)$  in  $L$  is  $d_z(x_0, y_0)$ ; the corresponding point of  $(x_0, y_0)$  in  $R$  is  $(x_0 + d_z(x_0, y_0), y_0)$ .

(e) The subimages  $L_s(x_0, y_0)$  and  $R_s(x_0, y_0)$  with same size are extracted through selecting the corresponding points as the center. Because the selection of the two subimages is based on the guidance of the integer-pixel disparity, thus there is only a small range of horizontal translation between  $L_s(x_0, y_0)$  and  $R_s(x_0, y_0)$ .

(f) The subpixel horizontal translation  $\Delta x$  between  $L_s(x_0, y_0)$  and  $R_s(x_0, y_0)$  is estimated by using the improved phase correlation method proposed in Section 3. The  $\Delta x$  is

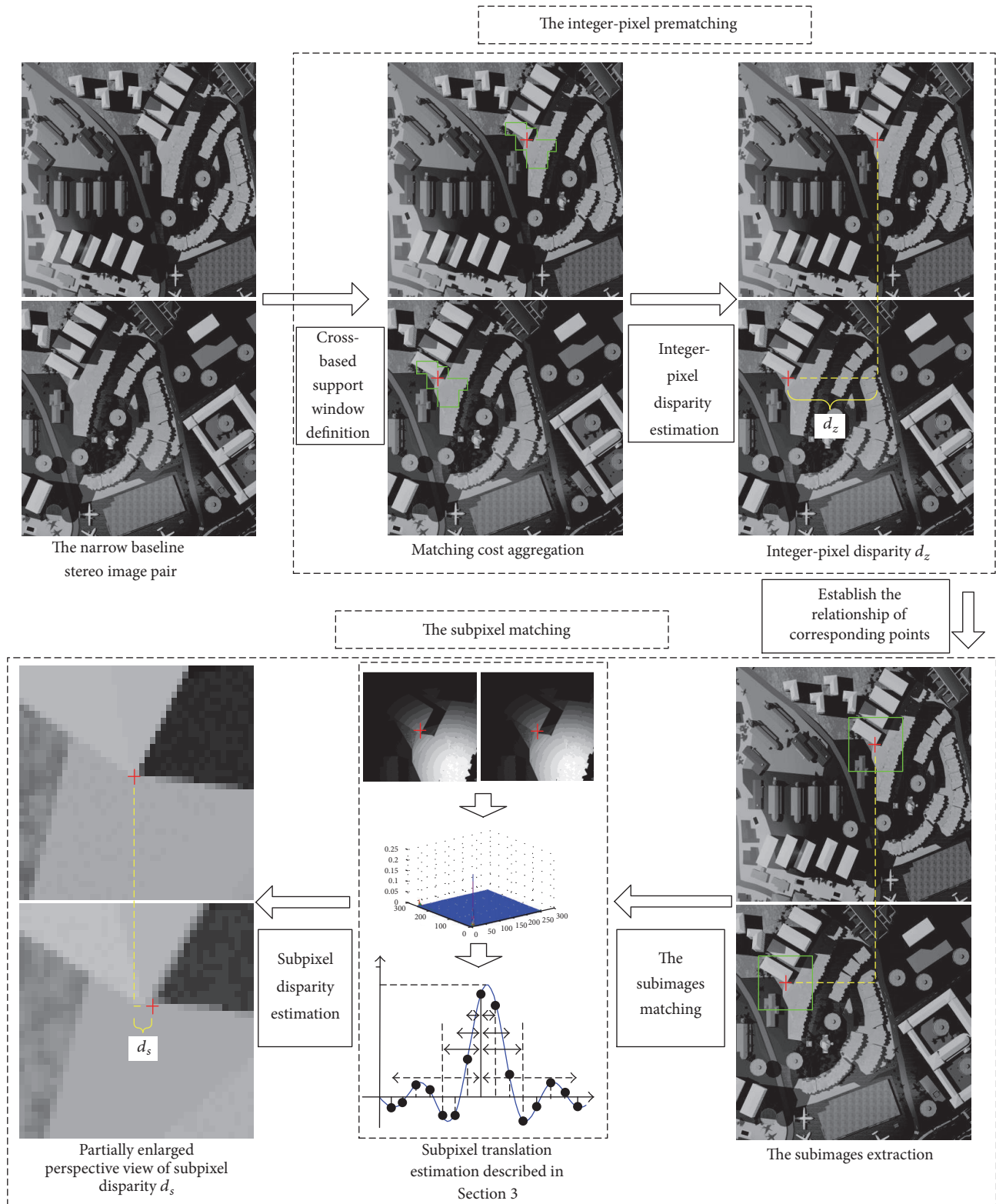


FIGURE 5: The complete framework of our subpixel image matching algorithm.

TABLE 1: General information of the test image pairs.

Test number	Data source	Image size	GSD (m)	Location	True disparity (pixels)
1	Pleiades	7590 × 6510	0.5	USA: Washington	8.738
2	SPOT-5	2050 × 2100	2.5	China: Beijing	8.738
3	SPOT-6	2015 × 1950	1.5	China: Beijing	8.738
4	WorldView-3	2000 × 2000	0.3	Spain: Madrid	8.738
5	GeoEye-1	3000 × 3000	0.4	Holland: Amsterdam	8.738
6	GF-2	4500 × 4750	0.8	China: Karamay	8.738

equivalent to the subpixel disparity  $d_s(x_0, y_0)$  of the center pixel  $(x_0, y_0)$ .

(g) Finally, the disparity result is obtained by combining the integer-pixel disparity and the subpixel disparity; that is,  $d(x_0, y_0) = d_z(x_0, y_0) + d_s(x_0, y_0)$ .

## 5. Experiments and Analysis

We test the proposed subpixel matching method for stereovision of narrow baseline remotely sensed imagery on a standard personal computer with Intel(R) Core(TM) i7 CPU, and the algorithm is implemented utilizing VS2010+OpenCV. The results analysis includes two aspects: precision analysis and complexity analysis. Constant parameter settings are used for all experiments. The parameters of this method are set as  $\beta = 0.11$ ,  $\tau_1 = 7/255$ , and  $\tau_2 = 2/255$ .

*5.1. Test on Simulated Narrow Baseline Remotely sensed Images Generated by Translation.* To evaluate the matching precision of proposed method, six urban scenes of narrow baseline remotely sensed image pairs, including Pleiades, SPOT-5, SPOT-6, WorldView-3, and GF-2, are used to perform image matching. For each image pair, the original image is utilized as the reference image (often called the left image), and the original image is moved 8.738 pixels along the  $x$ -axis as the target image (called the right image). Thus the true disparity is known to be 8.738 pixels for all the matching points of each image pair. Table 1 shows the general information of the test image pairs used in these experiments.

Figure 6 shows the distribution of matched points for the test narrow baseline remotely sensed image pairs. Because the image size is different for each image pair, the step size of grid is different too. The step sizes of grid in Tests 1–6 are set to  $540 \times 540$  pixels,  $160 \times 160$  pixels,  $180 \times 180$  pixels,  $170 \times 170$  pixels,  $300 \times 300$  pixels, and  $380 \times 380$  pixels.

In order to evaluate the matching precision of our method quantitatively, interpolation based method [14], fitting based method [15], and traditional phase correlation method [23] are used to compare with our method. Root mean square error (RMSE) and mean error of matched points are used as the assessment indices to evaluate these methods. Figure 7 shows quantitative evaluation results for the four subpixel image matching methods. Here the traditional phase correlation is called traditional PC for short. Figure 8 shows statistical results of the quantitative evaluation. Figure 8(a) shows statistical results where the RMSE is less than or

equal to 0.05 pixels. Figure 8(b) shows statistical results where the RMSE is greater than 0.05 pixels and less than 0.1 pixels. Figure 8(c) shows statistical results where the RMSE is greater than or equal to 0.1 pixels. From the experimental results, one can see that the two PC methods have a high matching precision, and the precision is better than that of interpolation method and fitting method. In addition, the matching precision is related to the spatial resolution of the images. High spatial resolution corresponds to a high matching precision. Because the peak evaluation method based on uniformly spaced sampling is designed to fit the phase correlation function, our method outperforms the traditional PC method, about 66% matched points of our method with RMSE less than or equal to 0.05 pixels.

*5.2. Test on Simulated Narrow Baseline Remotely sensed Images.* To evaluate the precision of DEM extraction by using the disparity estimated by our method, a simulated narrow baseline remotely sensed image pair provided by Beijing Institute of Space Mechanics and Electricity is used to perform image matching. The simulated narrow baseline remotely sensed image pair is generated by SE-WORKBENCH software. The attached parameters file of the image pair not only provides the true disparity, but also provides the true elevation of targets and buildings in the scene. Therefore, the precision of disparity and elevation estimated by our method can be evaluated carefully. Table 2 shows the general information of the test image pair used in these experiments. The  $B/H$  ratio of the image pair is 0.05, and spatial resolution is 0.3 meters/pixel. Overlap ratio of left and right images is 60%, and the overlap area includes 22 targets and 40 buildings. Figure 9 shows the simulated narrow baseline remotely sensed image pair and distribution of targets and buildings. Figure 9(a) is the image pair, Figure 9(b) is the distribution of targets, and Figure 9(c) is the distribution of buildings. The targets and buildings are marked with red numbers. Figure 10 shows the matching results. Figure 10(a) is the distribution of matched points for the buildings. Figure 10(b) is the distribution of matched points for the targets.

To evaluate the matching precision of our method quantitatively, interpolation based method [14], fitting based method [15], and traditional phase correlation method [23] are used to compare with our method. RMSE of disparity and mean elevation error are used as the assessment indices to evaluate the precision of four methods, where the elevation



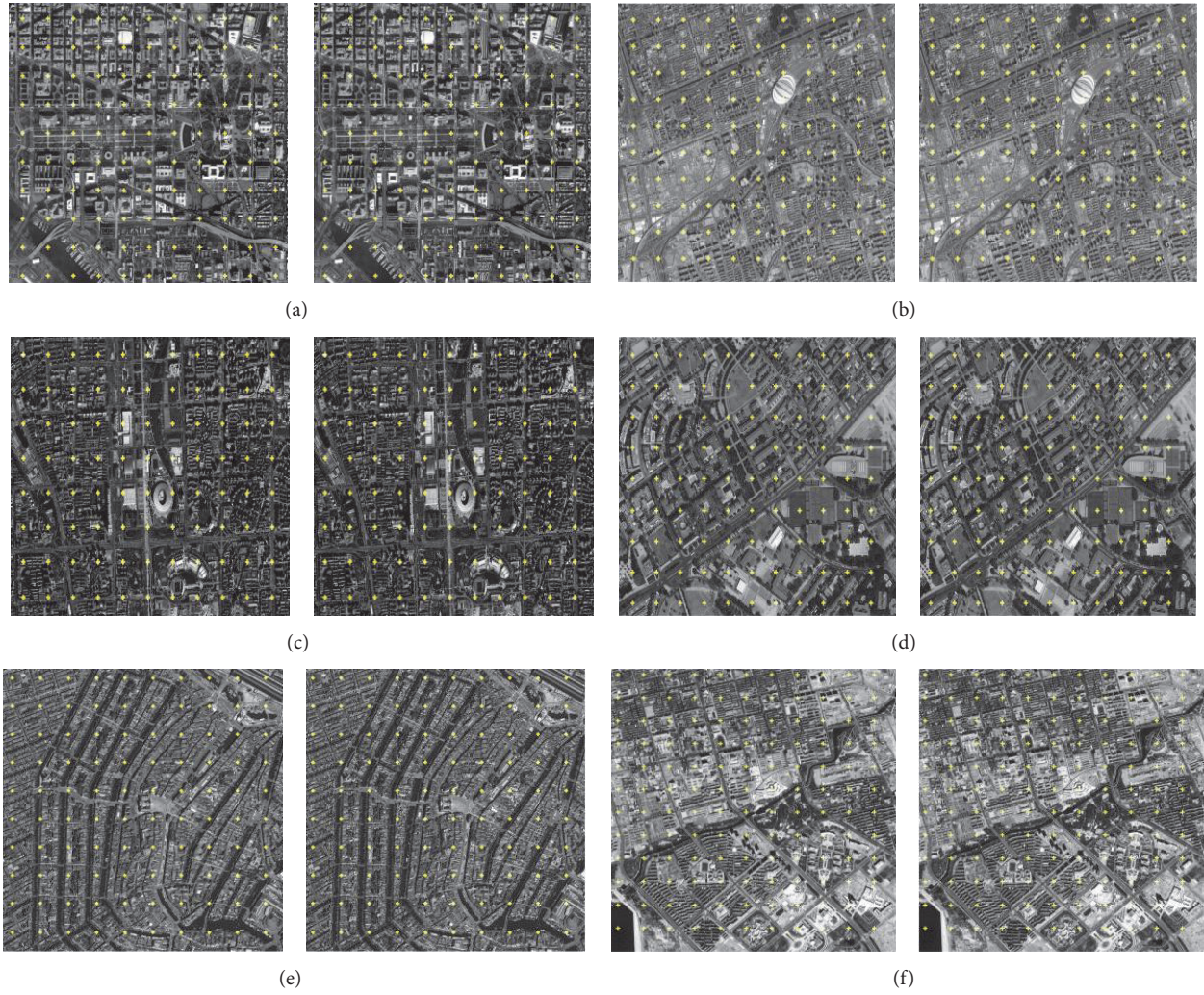


FIGURE 6: The distribution of matched points for the test narrow baseline remotely sensed image pairs. (a-f) are the results of Pleiades, SPOT-5, SPOT-6, WorldView-3, GeoEye-1, and GF-2 image pairs, respectively.

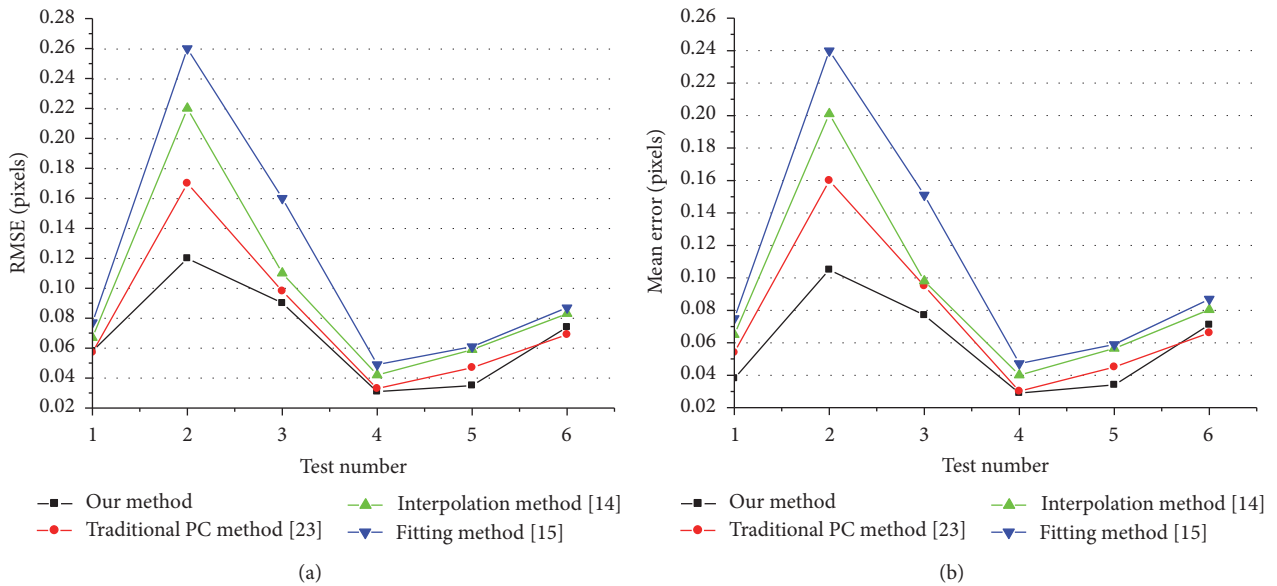


FIGURE 7: The quantitative evaluation results for the four subpixel image matching methods. (a) RMSE and (b) mean error.

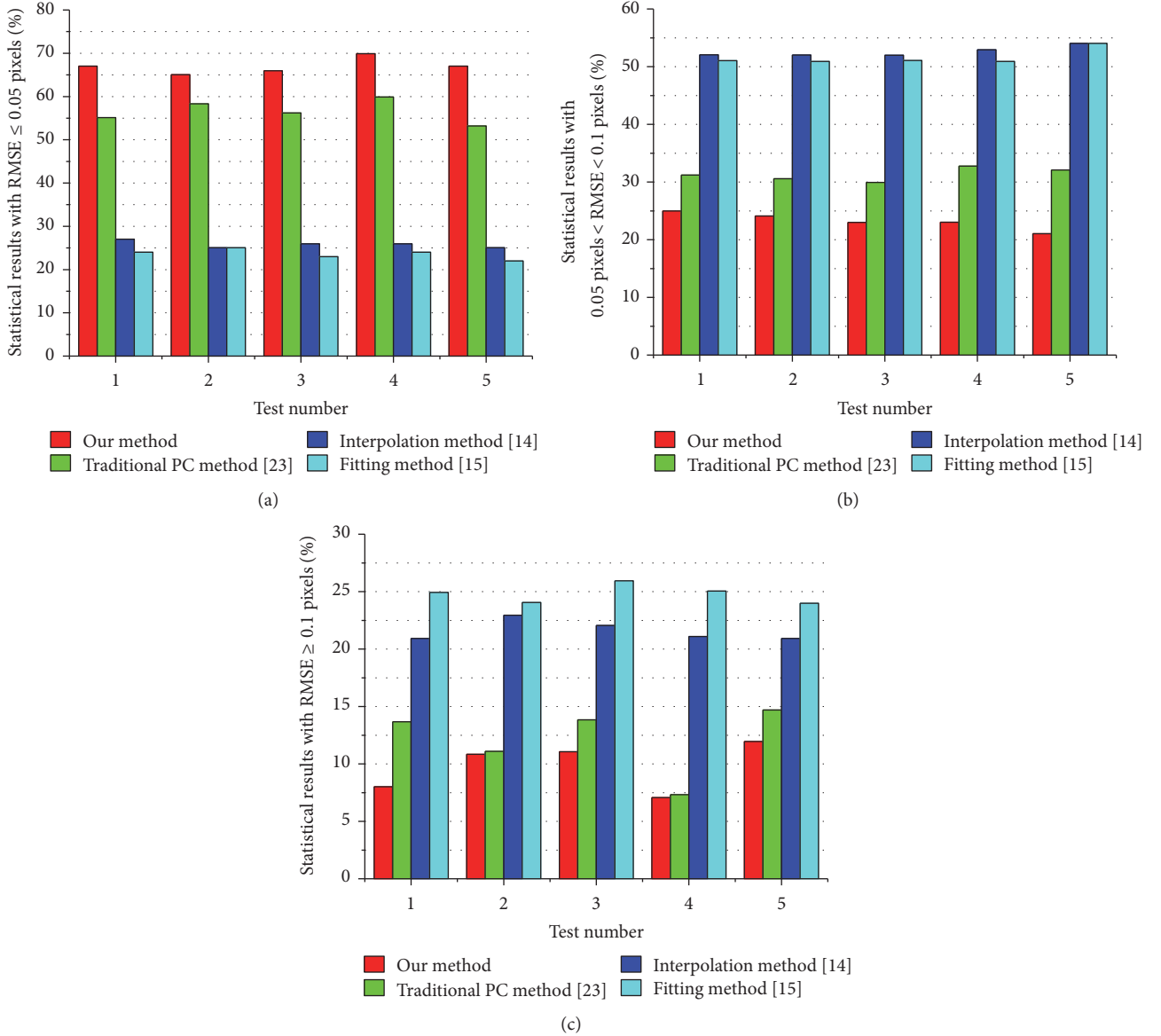


FIGURE 8: The statistical results of the quantitative evaluation. (a) Statistical results of RMSE less than or equal to 0.05 pixels; (b) statistical results of RMSE greater than 0.05 pixels and less than 0.1 pixels; and (c) statistical results of RMSE greater than or equal to 0.1 pixels.

error is the absolute difference between the true and the extracted elevation. The simulated narrow baseline remotely sensed image pair is matched 300 times by each subpixel image matching method. Figures 11 and 12 show the statistical results of four methods for targets and buildings, respectively. The results show that the RMSE and the mean elevation error obtained by our method are lower than the other three methods for most targets and buildings. Table 3 shows the average values of RMSE and mean elevation error. The true average elevation of 22 targets is 67.818 m, and the true average elevation of 40 buildings is 112.353 m. According to the statistical results in Table 3, the mean elevation error of targets estimated by our method is 0.303 m, and the mean elevation error of buildings estimated by our method is 0.406 m. Such precision can satisfy the requirements for

stereovision of narrow baseline remotely sensed imagery; therefore, it is proved that the image matching method based on phase correlation proposed in this paper has practical application value.

**5.3. Computational Complexity.** Assume that there are  $N_1$  points needed to be matched. In integer-pixel prematching step of our algorithm, the complexity of constructing the cross-based support window for each point is  $O(1)$ , and the total complexity for all points of this prematching step is  $O(N_1 \cdot d)$ , where  $d$  is the disparity range.

In the subpixel matching step, the core algorithm is based on the phase correlation; therefore, the complexity is unrelated to the disparity range but related to the size of subimage. Assume that there are  $N_2$  pixels in the extracted subimage.

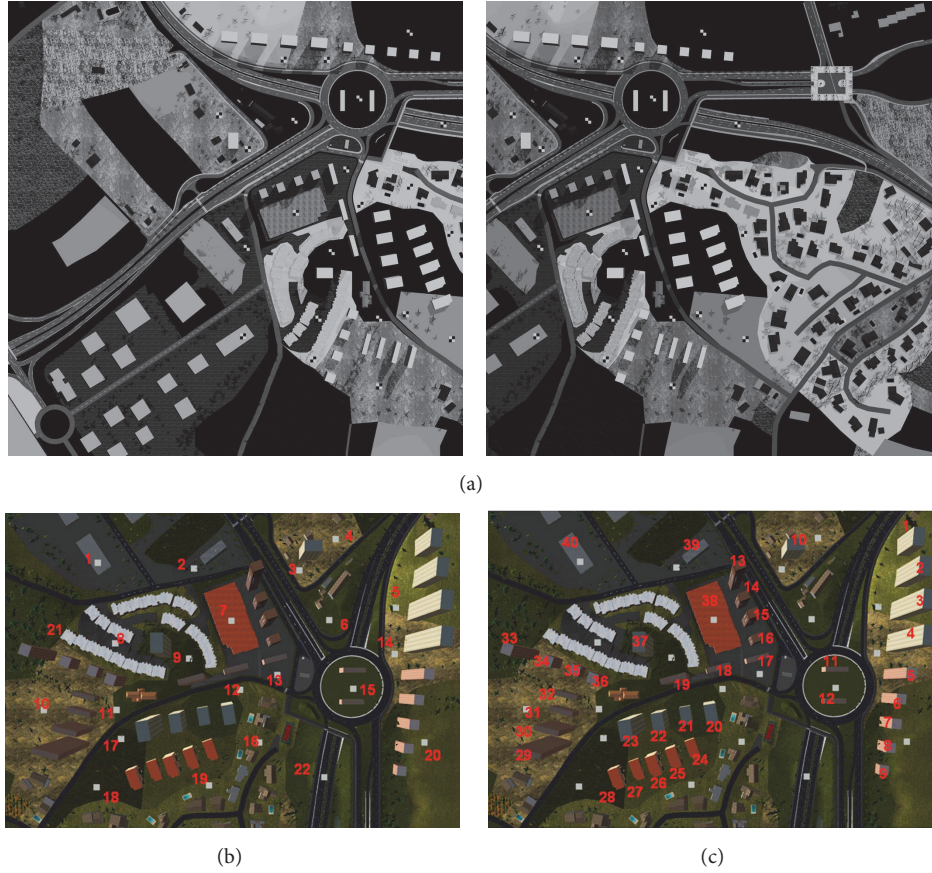


FIGURE 9: The simulated narrow baseline remotely sensed image pair and distribution of targets and buildings. (a) The image pair with a  $B/H$  ratio of 0.05. (b) The distribution of targets. (c) The distribution of buildings.

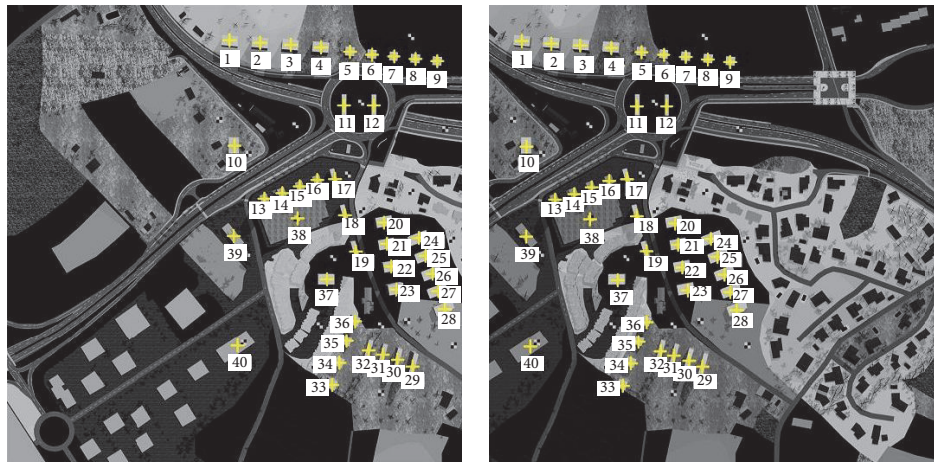
The complexity of fast DFT is  $O(N_2 \cdot \log N_2)$ , the complexity of Hanning window is  $O(N_2)$ , the complexity of normalized cross-power spectrum is  $O(N_2)$ , and the complexity of IDFT of normalized cross-power spectrum is  $O(N_2 \cdot \log N_2)$ . The peak evaluation is a least square fitting operation for  $S$  sets of uniformly spaced sampled-data essentially. The minimum value of  $S$  is 1; that is, the peak position is located at the most left or right of sampled-data. However, the maximum value of  $S$  is  $((\sqrt{N_2} - 1)/2)$ ; this means the peak position is located at the center of sampled-data; thus  $S \leq ((\sqrt{N_2} - 1)/2)$ . The complexity of subpixel peak position evaluation utilizing the least square method is  $O(((\sqrt{N_2} - 1)/2)^2)$ , which is equivalent to  $O(N_2)$ . The complexity of a single point matching in subpixel matching step is  $O(N_2 \cdot \log N_2)$ ; thus the complexity of total points is  $O(N_1 \cdot N_2 \cdot \log N_2)$ . The complexity of complete image matching algorithm is  $O(N_1 \cdot d) + O(N_1 \cdot N_2 \cdot \log N_2)$ . Because the complexity of subpixel matching is higher than that of integer-pixel prematching, the complexity of our algorithm is  $O(N_1 \cdot N_2 \cdot \log N_2)$ .

In order to clearly illustrate the complexity of our algorithm, we compare with the classical normalized cross correlation (NCC) algorithm [2]. Assume that the support window of NCC is  $\varphi$ ; in the practical image matching,  $\varphi \approx N_2 \ll N_1$ . When  $\log N_2 \approx d$ , the complexity of our method

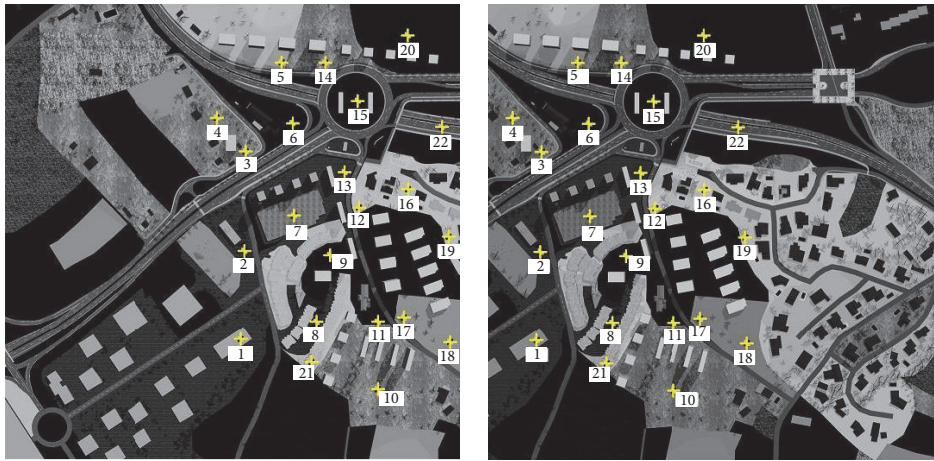
is similar to that of NCC. Generally, the size of subimage is  $30 \times 30 \sim 40 \times 40$  pixels; the number of pixels in subimage is  $900 \sim 1600$ ; thus the value range of  $\log N_2$  is  $9.814 \sim 10.644$ . The disparity range of narrow baseline stereo image pair is usually within 20 pixels. In our experiment, the disparity range is 17 pixels;  $\log N_2 \approx d$  is established; therefore, the complexity of our method is similar to that of the classical NCC method.

## 6. Conclusion

Through the analysis of narrow baseline remotely sensed imagery stereovision, a subpixel image matching approach based on improved phase correlation is proposed. The cross-based local matching method is employed for prematching the image pair, and the obtained integer-pixel disparity reduces the search range of subpixel matching. A high precision subpixel matching step is implemented under the guidance of integer-pixel disparity. A peak evaluation method based on uniformly spaced sampling is designed to improve the precision of disparity estimation. The experimental results show that our method has superior performance on precision and efficiency, and it is beneficial to the stereovision of narrow baseline remotely sensed imagery.

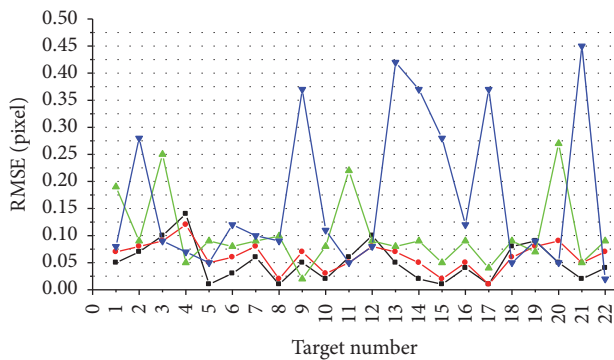


(a)

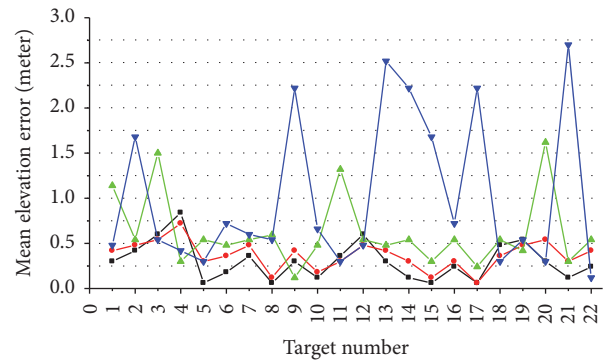


(b)

FIGURE 10: The matching results. (a) The distribution of matched points for the buildings. (b) The distribution of matched points for the targets.



(a)



(b)

FIGURE 11: The statistical results of targets. (a) Comparison of RMSE. (b) Comparison of mean elevation error.

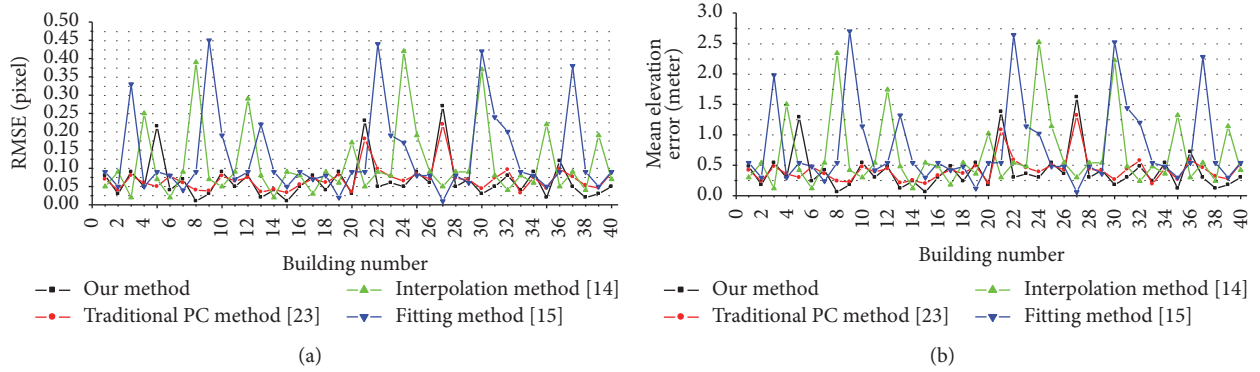


FIGURE 12: The statistical results of buildings. (a) Comparison of RMSE. (b) Comparison of mean elevation error.

TABLE 2: The general information of the test image pair.

Target number	Elevation (meter)	True disparity (pixel)	Building number	Elevation (meter)	True disparity	Building number	Elevation (meter)	True disparity (pixel)
1	73	12.167	1	111.42	18.570	21	97.3	16.217
2	65	10.833	2	131.54	21.923	22	131.33	21.888
3	63	10.500	3	147.13	24.522	23	165.92	27.653
4	65	10.833	4	153.77	25.628	24	134.85	22.475
5	67	11.167	5	123.43	20.572	25	85.14	14.190
6	61	10.167	6	104.77	17.462	26	92.28	15.380
7	67	11.167	7	98.22	16.370	27	71.51	11.918
8	82	13.667	8	88.81	14.802	28	83.5	13.917
9	74	12.333	9	75.07	12.512	29	169.2	28.200
10	72	12.000	10	98.46	16.410	30	131.22	21.870
11	75	12.500	11	79.79	13.298	31	98.42	16.403
12	65	10.833	12	72.94	12.157	32	98.42	16.403
13	60	10.000	13	155.81	25.968	33	154.57	25.762
14	64	10.667	14	111.81	18.635	34	115.5	19.250
15	63	10.500	15	133.81	22.302	35	105.21	17.535
16	61	10.167	16	100.81	16.802	36	133.07	22.178
17	76	12.667	17	92.21	15.368	37	157.59	26.265
18	64	10.667	18	120.38	20.063	38	66.3	11.050
19	62	10.333	19	162.39	27.065	39	72.5	12.083
20	66	11.000	20	145.29	24.215	40	72.44	12.073
21	92	15.333	—	—	—	—	—	—
22	55	9.167	—	—	—	—	—	—

TABLE 3: The RMSE and mean elevation error of four methods.

Algorithms	Targets		Buildings	
	RMSE (pixel)	Mean elevation error (meter)	RMSE (pixel)	Mean elevation error (meter)
Interpolation method [14]	0.103	0.619	0.114	0.681
Fitting method [15]	0.169	1.012	0.133	0.797
Traditional PC method [23]	0.061	0.368	0.071	0.422
<b>Our method</b>	<b>0.051</b>	<b>0.303</b>	<b>0.068</b>	<b>0.406</b>

## Competing Interests

The authors declare no conflict of interests.

## Authors' Contributions

Ning Ma proposed the idea of the method. Ning Ma and Yubo Men performed the validation experiment of the method. Ning Ma, Chaoguang Men, and Xiang Li wrote this paper. All the authors read and approved the final paper.

## Acknowledgments

This work was supported by National Natural Science Foundation of China (Grant nos. 11547157 and 61100004), Natural Science Foundation of Heilongjiang Province of China (Grant no. F201320), and Harbin Municipal Science and Technology Bureau (Grant no. 2014RFQXJ073).

## References

- [1] T. Long, W. Jiao, G. He, and Z. Zhang, "A fast and reliable matching method for automated georeferencing of remotely-sensed imagery," *Remote Sensing*, vol. 8, no. 1, article 56, 2016.
- [2] J. Delon and B. Rougé, "Small baseline stereovision," *Journal of Mathematical Imaging and Vision*, vol. 28, no. 3, pp. 209–223, 2007.
- [3] C. C. Pham and J. W. Jeon, "Domain transformation-based efficient cost aggregation for local stereo matching," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 23, no. 7, pp. 1119–1130, 2013.
- [4] D. Fan, E. Shen, L. Li et al., "Small baseline stereo matching method based on phase correlation," *Journal of Geomatics Science and Technology*, vol. 30, pp. 154–157, 2013.
- [5] G. L. K. Morgan, J. G. Liu, and H. Yan, "Precise subpixel disparity measurement from very narrow baseline stereo," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 9, pp. 3424–3433, 2010.
- [6] J.-L. Bian, C.-G. Men, and X. Li, "A fast stereo matching method based on small baseline," *Journal of Electronics and Information Technology*, vol. 34, no. 3, pp. 517–522, 2012.
- [7] N. Sabater, J. M. Morel, and A. Almansa, "How accurate can block matches be in stereo vision?" *SIAM Journal on Imaging Sciences*, vol. 4, pp. 472–500, 2011.
- [8] L. Igual, J. Preciozzi, L. Garrido et al., "Automatic low baseline stereo in urban areas," *Inverse Problems and Imaging*, vol. 1, pp. 318–348, 2007.
- [9] E. Shen, D. Fan, and X. Sun, "Smallbaseline stereo matching method based on SGM and phase correlation," *Journal of China University of Mining and Technology*, vol. 44, no. 1, pp. 183–188, 2015.
- [10] N. Sabater, G. Blanchet, L. Moisan, A. Almansa, and J.-M. Morel, "Review of low-baseline stereo algorithms and benchmarks," in *Image and Signal Processing for Remote Sensing XVI*, 783005, 12, Toulouse, France, October 2010.
- [11] T. Arai and A. Iwasaki, "Fine image matching for narrow baseline stereovision," in *Proceedings of the 32nd IEEE International Geoscience and Remote Sensing Symposium (IGARSS '12)*, pp. 2336–2339, Munich, Germany, July 2012.
- [12] J. Inglada, V. Muron, D. Pichard, and T. Feuvrier, "Analysis of artifacts in subpixel remote sensing image registration," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, no. 1, pp. 254–264, 2007.
- [13] V.-C. Miclea, C.-C. Vancea, and S. Nedevschi, "New sub-pixel interpolation functions for accurate real-time stereo-matching algorithms," in *Proceedings of the 11th IEEE International Conference on Intelligent Computer Communication and Processing (ICCP '15)*, pp. 173–178, IEEE, Cluj-Napoca, Romania, September 2015.
- [14] I. Haller and S. Nedevschi, "Design of interpolation functions for subpixel-accuracy stereo-vision systems," *IEEE Transactions on Image Processing*, vol. 21, no. 2, pp. 889–898, 2012.
- [15] C. Stentoumis, L. Grammatikopoulos, I. Kalisperakis, and G. Karras, "On accurate dense stereo-matching using a local adaptive multi-cost approach," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 91, pp. 29–49, 2014.
- [16] J. Žbontar and Y. Le Cun, "Computing the stereo matching cost with a convolutional neural network," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '15)*, pp. 1592–1599, Boston, Mass, USA, June 2015.
- [17] C. Shi, G. Wang, X. Yin, X. Pei, B. He, and X. Lin, "High-accuracy stereo matching based on adaptive ground control points," *IEEE Transactions on Image Processing*, vol. 24, no. 4, pp. 1412–1423, 2015.
- [18] N. Jiang, Y. Qu, and Y. Li, "Fast sub-pixel accuracy stereo image matching based on disparity plane," in *Proceedings of the International Conference on Optical Instruments and Technology: Optoelectronic Imaging and Processing Technology*, Beijing, China, May 2015.
- [19] F. Besse, C. Rother, A. Fitzgibbon, and J. Kautz, "PMBP: Patch-Match Belief Propagation for correspondence field estimation," *International Journal of Computer Vision*, vol. 110, no. 1, pp. 2–13, 2014.
- [20] S. Xu, F. Zhang, X. He, X. Shen, and X. Zhang, "PM-PM: PatchMatch with potts model for object segmentation and stereo matching," *IEEE Transactions on Image Processing*, vol. 24, no. 7, pp. 2182–2196, 2015.
- [21] S. Nagashima, T. Aoki, T. Higuchi, and K. Kobayashi, "A sub-pixel image matching technique using phase-only correlation," in *Proceedings of the International Symposium on Intelligent Signal Processing and Communications (ISPACS '06)*, pp. 701–704, IEEE, Yonago, Japan, December 2006.
- [22] H. S. Stone, M. T. Orchard, E.-C. Chang, and S. A. Martucci, "A fast direct Fourier-based algorithm for subpixel registration of images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 39, no. 10, pp. 2235–2243, 2001.
- [23] H. Foroosh, J. B. Zerubia, and M. Berthod, "Extension of phase correlation to subpixel registration," *IEEE Transactions on Image Processing*, vol. 11, no. 3, pp. 188–199, 2002.
- [24] X. Tong, Y. Xu, Z. Ye et al., "An improved phase correlation method based on 2-D plane fitting and the maximum kernel density estimator," *IEEE Geoscience and Remote Sensing Letters*, vol. 12, no. 9, pp. 1953–1957, 2015.
- [25] J. Lu, K. Shi, D. Min, L. Lin, and M. N. Do, "Cross-based local multipoint filtering," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '12)*, pp. 430–437, IEEE, Providence, RI, USA, June 2012.



# Hindawi

Submit your manuscripts at  
<https://www.hindawi.com>

