# scientific reports

OPEN

# A discovery about the positional distribution pattern among candidate homologous pixels and its potential application in aerial multi-view image matching

Ka Zhang[1,2,3,4,5✉], Wen Xiao[6,7], Yehua Sheng[1,2,3,4✉], Junshu Wang[1,2,3,4,7], Shan Zhang[1,2] & Longjie Ye[1,2]

In aerial multi-view photogrammetry, whether there is a special positional distribution pattern among candidate homologous pixels of a matching pixel in the multi-view images? If so, can this positional pattern be used to precisely confirm the real homologous pixels? These problems have not been studied at present. Therefore, the study of the positional distribution pattern among candidate homologous pixels based on the adjustment theory in surveying is investigated in this paper. Firstly, the definition and computing method of pixel's pseudo object-space coordinates are given, which can transform the problem of multi-view matching for confirming real homologous pixels into the problem of surveying adjustment for computing the pseudo object-space coordinates of the matching pixel. Secondly, according to the surveying adjustment theory, the standardized residual of each candidate homologous pixel of the matching pixel is figured out, and the positional distribution pattern among these candidate pixels is theoretically inferred utilizing the quantitative index of standardized residual. Lastly, actual aerial images acquired by different sensors are used to carry out experimental verification of the theoretical inference. Experimental results prove not only that there is a specific positional distribution pattern among candidate homologous pixels, but also that this positional distribution pattern can be used to develop a new object-side multi-view image matching method. The proposed study has an important reference value on resolving the defects of existing image-side multi-view matching methods at the mechanism level.

Extracting information of ground objects such as shape, position, size and relationship from images is a hot research topic in the fields of geosciences and computer vision[1–4]. Especially in geographic information science, aerial photogrammetry has become the main technique of 3D geospatial information collection because of its advantages of low-cost, large coverage and high efficiency[5,6]. In the process of 3D geospatial information extraction based on aerial photogrammetry, multi-view image matching is a key step. Homologous pixels of the same object in multiple images obtained from this step are fundamental for other technical steps such as automatic aerial triangulation[7,8], production of geographic information products[9–11] and 3D scene reconstruction[12–14].

[1]Key Laboratory of Virtual Geographic Environment (Nanjing Normal University), Ministry of Education, 1# WenYuan Road, Nanjing 210023, People's Republic of China. [2]School of Geography, Nanjing Normal University, 1# WenYuan Road, Nanjing 210023, People's Republic of China. [3]Jiangsu Center for Collaborative Innovation in Geographical Information Resource Development and Application, 1# WenYuan Road, Nanjing 210023, People's Republic of China. [4]State Key Laboratory Cultivation Base of Geographical Environment Evolution (Jiangsu Province), 1# WenYuan Road, Nanjing 210023, People's Republic of China. [5]Key Laboratory of Urban Land Resources Monitoring and Simulation, Ministry of Natural Resources, Shenzhen 518034, People's Republic of China. [6]School of Engineering, Newcastle University, Newcastle upon Tyne NE1 7RU, UK. [7]These authors contributed equally: Wen Xiao and Junshu Wang. ✉email: zhangka81@126.com; shengyehua@njnu.edu.cn

There are two basic problems in multi-view image matching: one is the matching measure computation, and the other is the searching scope identification for confirming the homologous pixel in each searching image[15–17]. The matching measure is the quantitative basis for judging the real homologous pixel from candidate pixels in the searching scope, which affects the accuracy and robustness of multi-view image matching. The searching scope of the homologous pixel determines how many pixels will be used as candidates in the matching process, affecting the reliability and computational efficiency. The existing multi-view image matching studies mainly focus on the following aspects: how to enhance the robustness of matching measure computation using pixels' image-side grayscale or feature information; how to use various restriction strategies to reduce the searching scope of homologous pixel; and how to optimize the matching process.

In the aspect of matching measure computation, commonly used methods can be divided into two categories based on the type of pixels' attributes utilized in the computing process: grayscale-based models and feature-based models. The grayscale-based matching measure models use pixels' image-side grayscale values in red, green and blue channels to compute the similarity between two matching windows[18]. The following functions are frequently used in similarity computation, such as sum of squared differences (SSD)[19], normalized cross correlation (NCC)[20], zero mean normalized cross correlation (ZNCC)[21], sum of absolute differences (SAD)[22], and arithmetic distance[23]. The feature-based matching measure models will firstly perform a feature transform on images, such as scale invariant feature transform (SIFT)[24], speeded up robust features transform (SURF)[25], Harris transform[26], Rank transform[27], Census transform[28], contourlet and wavelet transform[29,30], then use obtained feature vectors as attribute values to compute the matching measure. The grayscale-based matching measure models have advantages of being simple computation and easy to realize. But they cannot work well in poorly or repetitively textured regions of the image such as water bodies and building roofs, and they are also sensitive to factors such as geometric distortions of images, changes of light intensity and contrast. In comparison, the feature-based matching measure models have strong adaptability to the changes of geometry and radiation among images. But their matching results heavily depend on the quality of feature extraction, which is more time-consuming.

In the aspect of defining the searching scope of homologous pixels, current research emphasis is on how to use various restriction strategies to reduce the searching scope and decrease the total number of candidate pixels. Frequently-used restriction strategies include pyramid hierarchy constraint[31], epipolar line constraint[32], prior object-space knowledge constraint, such as known elevation or DEM[33,34], and spatial relation constraint, such as disparity continuity[35], plane homography[36] and geometric invariance[37,38]. Among abovementioned restriction strategies, except that the epipolar line constraint is generally workable, other constraints are only suitable to specific scenes or can be used in certain conditions. For example, the constraint of disparity continuity is not valid in disparity fault regions of ground objects (such as edges of buildings). Constraints of plane homography and spatial invariance may be invalid due to factors such as shielding relations among objects and geometric distortions of objects in the image. Furthermore, these restriction strategies can easily lead to transmission and accumulation of mismatched results, and the quality of one pixel's matching result may be restricted by that of its neighbor pixels.

In the aspect of realization and optimization of matching process, according to the different optimizing ways, image matching methods can be divided into two types: local matching and global matching[39]. Local matching methods use the support window centered by the matching pixel to compute the matching cost, and use the strategy of "winner take all—WTA" to find the homologous pixels with the minimal cost. Local methods have advantages in computation efficiency, so can be used for real-time processing. But they mainly treat the matching problem from a local perspective, and are more sensitive to local characteristics of the image, hence can easily generate mismatches in the regions with depth discontinuity and occlusion. On the contrast, global matching methods use energy minimization to transform the image matching problem into the discrete optimization problem under the constraint of disparity continuity. Energy minimization methods such as dynamic programming algorithm[40], belief propagation algorithm[41,42], graph cut algorithm[43] and Markov Random Field algorithm[44] are often used. The overall matching results of these methods are generally good. However, these methods normally need to set many parameters with non-physical meanings, and they have high computational cost and poor real-time performance.

In addition, with the booming development of deep learning and its extensive applications in the field of geosciences[45], deep neural networks are also applied in some stages of image matching to improve the efficiency and accuracy of existing matching methods, such as the learning of matching cost[46], the learning of the direct process from stereopair to disparity map[47,48]. However, the methods based on neural networks still do not make any substantive improvements in the existing matching algorithms in term of the principle. Besides, they have the following obvious shortages: the matching performance depends on sufficient training data, and the trained neural network is often difficult to obtain the same matching effect when applied to unfamiliar scenes.

Overall, existing image matching methods can all be summarized as image-side methods. They only use image-side information (such as grayscale values, feature vectors, relations among pixels, etc.) of pixels in the matching window to compute the matching measure and to restrict the searching scope of homologous pixels. But the image-side information is not unique or invariable. When this information is used alone, it is difficult to construct one certain distinctive feature to accurately express the similarity between pixels, and especially difficult to precisely confirm the fine searching scope of homologous pixels. Therefore, existing matching methods cannot overcome the problem of low robustness of the matching measure caused by the phenomenon of "different spectrum of the same object" or "different objects of the same spectrum" in the image, nor the problem of the change of spatial relation among pixels caused by object occlusions and perspective distortions. When these methods are applied in difficult matching regions with repetitive or week textures, discontinuous disparity and geometric deformation, they manifest drawbacks of low efficiency and poor reliability.
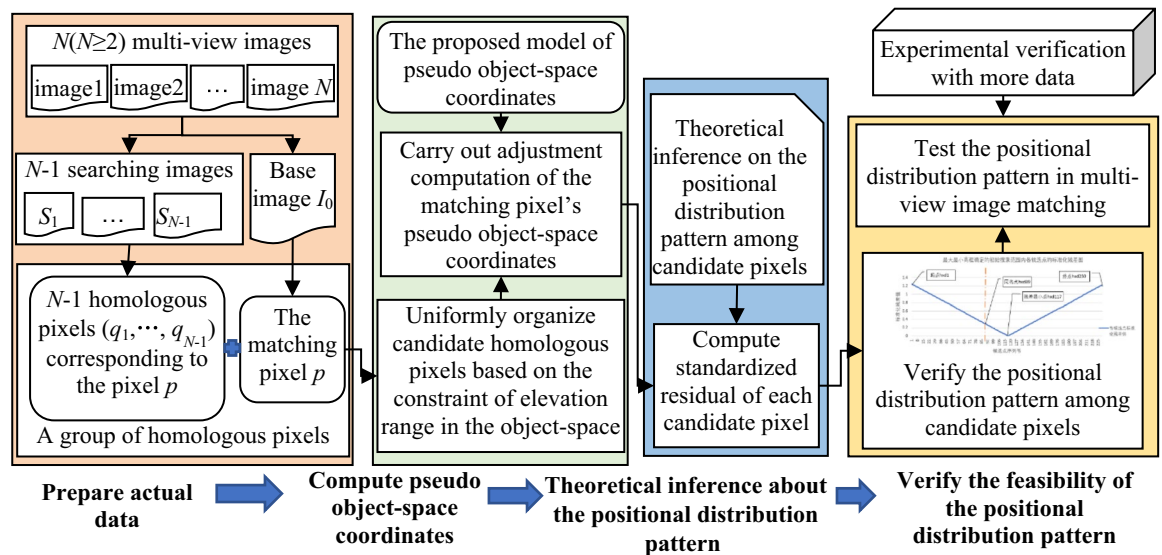
**Figure 1.** The general flow diagram of the proposed research.

Furthermore, existing research paid little attention to following questions: whether there is a certain positional distribution pattern among candidate homologous pixels of the matching pixel in multi-view images? How to quantitatively describe this positional distribution pattern? Can this pattern be used to further refine the searching scope of candidate homologous pixels, or even to exclusively confirm the homologous pixels during multi-view image matching? Thorough investigation of these questions will prospectively enhance the accuracy, efficiency and robustness of multi-view image matching from the aspect of matching mechanism, and it has the utmost significance in resolving shortcomings of existing image matching methods.

Therefore, this paper proposes a new conception of pseudo object-space coordinates and its computing model, which transforms the positional distribution pattern among candidate homologous pixels of a matching pixel into a problem of surveying adjustment. On this basis, a new method of studying the positional distribution pattern among candidate homologous pixels based on surveying adjustment theory is established, and actual aerial multi-view images are used to carry out experiments for verifying the proposed method. It is found that there is a specific corresponding relationship between the standardized residuals of candidate homologous pixels and the positional distribution pattern among these pixels, and this positional distribution pattern can be used to develop a new multi-view image matching method without utilizing the image-side grayscale information.

The remainder of the paper is organized as follows. "Method" section details the principle of the proposed method of the positional distribution pattern among candidate homologous pixels. "Experiment and discussion" section presents detailed experimental results and discussions. Finally, conclusions and potential future work are presented in "Conclusion" section.

## Method

**General idea.** To study the positional distribution pattern among candidate homologous pixels of the matching pixel in multi-view images, the first thing is to transform the problem of seeking homologous pixels into a problem of surveying adjustment. As all know, in photogrammetry, although the image-space results of multi-view image matching are a group of homologous pixels identified from many candidate pixels in multiple target images corresponding to the matching pixel in the reference image, the final aim is to compute the object-space 3D geospatial coordinates of the ground object using image-space 2D row and column coordinates of the homologous pixels. Therefore, considering the fact that the ground object corresponding to the homologous pixels has unique object-space 3D coordinates, this paper proposes a new concept of pseudo object-space coordinates to express the object-space information computed with all candidate pixels referring to the conception of pseudo distance in GNSS data processing. Based on this new concept, the paper transforms the problem of multi-view image matching into a problem of surveying adjustment of pseudo object-space coordinates according to the theories of surveying error adjustment.

The diagram of the proposed method is shown in Fig. 1. Firstly, experimental data are prepared as follows: $N$ images with known interior and exterior orientation parameters ($N \geq 2$ is the total number of multi-view images), and one group of real homologous pixels in these images (one pixel is used as the matching pixel). Secondly, based on the proposed model of pseudo object-space coordinates, adjustment of the matching pixel's pseudo object-space coordinates is carried out. Thirdly, according to the adjustment results of pseudo object-space coordinates, the standardized residual of each candidate pixel in the searching scope of homologous pixels is computed. Fourthly, based on the fact that errors in observations will strongly distort the adjustment results, the standardized residual is used as a quantitative index to depict the positional feature of each candidate pixel, and a theoretical inference about the positional distribution pattern among candidate homologous pixels based on the surveying adjustment theory is hypothesized. Lastly, the experimental data are used to draw the curve of the
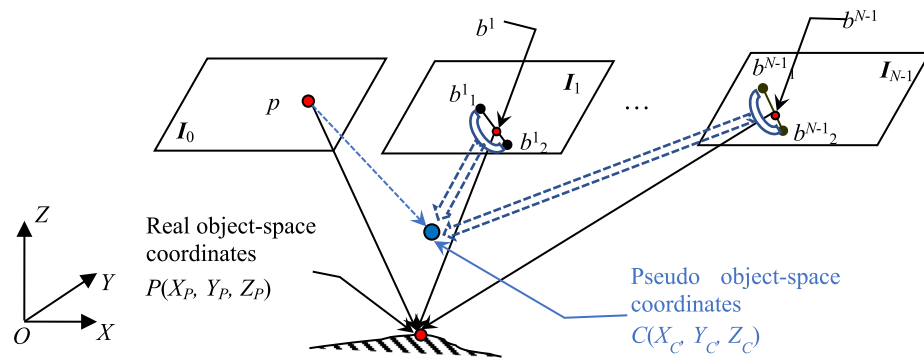
**Figure 2.** The relation between pixel's real object-space 3D coordinates and its pseudo object-space coordinates.

relationship between the position of each candidate pixel and its standardized residual for verifying the inference, and to further test the application effect of the positional distribution pattern in multi-view image matching.

To describe the subsequent content conveniently, for $N$ ($N \geq 2$) multiple images with known interior and exterior orientation parameters obtained by automatic aerotriangulation software, one of $N$ images used to pick the matching pixel is named as the base image $I_0$, and the remaining $N-1$ images as searching images $I_1, \ldots, I_k, \ldots, I_{N-1}$ ($k = 1, 2, \ldots, N-1$). Projection centers of the base image and searching images are denoted as $S_0$ and $S_k$, and respective exterior orientation parameters are denoted as $(X_0, Y_0, Z_0, \varphi_0, \omega_0, \kappa_0)$ and $(X_k, Y_k, Z_k, \varphi_k, \omega_k, \kappa_k)$. The row and column numbers of the pixel are denoted as $(r, c)$, and corresponding image plane coordinates are $(x, y)$.

### Definition and computation of pseudo object-space coordinates.

The definition of proposed pseudo object-space coordinates is as follows: for a matching pixel $p$ in the base image $I_0$, the 3D coordinates, which are calculated with the pixel $p$ and all candidate pixels in the searching scope of its homologous pixels in the $N-1$ searching images, are defined as the pseudo object-space coordinates of the pixel $p$. Seen from the definition, the pseudo object-space coordinates have the form of 3D coordinates $(X, Y, Z)$, but are not true 3D coordinates of ground object corresponding to the pixel $p$, hence named as pseudo coordinates.

The sketch for displaying the relation between a pixel's real object-space 3D coordinates and its pseudo object-space coordinates is shown in Fig. 2. In this figure, the point $p$ represents the matching pixel in the base image, and its real homologous pixels in $N-1$ searching images are denoted as $[b^1, \ldots, b^{N-1}]$. Two endpoints of the searching range of homologous pixel in the $k$th searching image are marked as $b^k_1$ and $b^k_2$ (that are on the epipolar line, and the method of confirming which can be referred to in "Analysis of the positional distribution pattern among candidate homologous pixels based on the adjustment of pseudo object-space coordinates"), and the $j$th candidate pixel on this epipolar line is labeled as $b^k_j$. Then, the coordinates $(X_P, Y_P, Z_P)$ computed by the pixel $p$ and its real homologous pixels $[b^1, \ldots, b^{N-1}]$ are real object-space 3D coordinates, and the coordinates $(X_C, Y_C, Z_C)$ computed by the pixel $p$ and all its candidate pixels $\{[b^1_1, \ldots, b^1_j, \ldots, b^1_2], \ldots, [b^{N-1}_1, \ldots, b^{N-1}_j, \ldots, b^{N-1}_2]\}$ in $N-1$ searching images are pseudo object-space coordinates. In addition, the dashed arrow pointing to the pseudo coordinates is used to represent the participated matching pixel and its candidate homologous pixels, and the hollow arc on each searching image is used to represent all candidate pixels.

To compute the pseudo object-space coordinates, the method of multi-view forward intersection in photogrammetry is used. For the matching pixel $p$ and its $M_t = \sum_{k=1}^{N-1} M_k$ candidate homologous pixels in $N-1$ searching images ($M_k$ is the total number of candidate homologous pixels in the $k$th searching image $I_k$), the pseudo object-space coordinates of pixel $p$ is computed by the following steps: firstly, the error equation matrix used to solve the pseudo object-space coordinates is formed as $\mathbf{V} = \mathbf{AX} - \mathbf{L}$ according to the principle of multi-view forward intersection. In this error equation matrix, $\mathbf{V}$ is the residual matrix of $2(M_t + 1)$ image plane coordinates (its dimensions are $2(M_t + 1)$ rows $\times$ 1 column); $\mathbf{X} = [dX, dY, dZ]^T$ is the coordinate corrections matrix of undetermined pseudo object-space coordinates (its dimensions are 3 rows $\times$ 1 column); $\mathbf{A}$ and $\mathbf{L}$ are the coefficient matrix and the constant matrix (their dimensions are $2(M_t + 1)$ rows $\times$ 3 columns and $2(M_t + 1)$ rows $\times$ 1 column respectively), and detailed formulas for computing the elements in each row of $\mathbf{A}$ and $\mathbf{L}$ can be referred to in any photogrammetry textbooks[49]. Secondly, using the principle of least squares, the error equation matrix is solved by equation $\mathbf{X} = (\mathbf{A}^T\mathbf{A})^{-1}(\mathbf{A}^T\mathbf{L})$ to obtain the corrections $(dX, dY, dZ)$ of the initial coordinates $(X^0, Y^0, Z^0)$. Then, the pseudo object-space coordinates $(X, Y, Z)$, of the pixel $p$, are obtained by add the corrections: $X = X^0 + dX$, $Y = Y^0 + dY$ and $Z = Z^0 + dZ$.

To be noted, although the computing of pseudo object-space coordinates is based on the principle of multi-view forward intersection, their computing processes are not identical. In traditional multi-view forward intersection, only one candidate pixel in each searching image is used in the computing process, and this process needs to run iteratively. However, in the computing of pseudo object-space coordinates, all candidate homologous pixels in the searching range of each searching image corresponding to the matching pixel are used, and corrections of undetermined coordinates only need to be calculated once. The goal of computing pseudo object-space coordinates is not to solve the real object-space 3D coordinates of the matching pixel, but to analyze the positional distribution pattern among candidate homologous pixels according to the adjustment results of pseudo object-space coordinates.
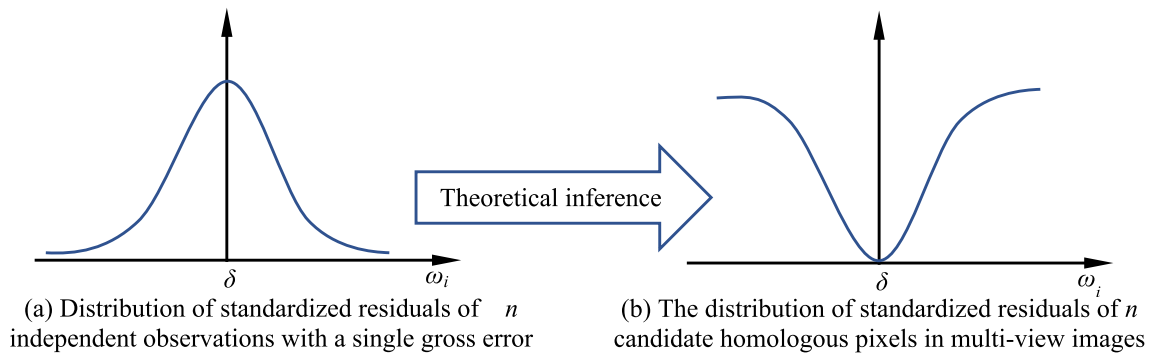
4

(a) Distribution of standardized residuals of $n$ independent observations with a single gross error

(b) The distribution of standardized residuals of $n$ candidate homologous pixels in multi-view images

**Figure 3.** The theoretical inference on the positional distribution pattern among candidate homologous pixels in multi-view images according to the surveying adjustment theory.

**Analysis of the positional distribution pattern among candidate homologous pixels based on the adjustment of pseudo object-space coordinates.** Seen from above, for a matching pixel in the base image, there is a certain correlation between the computation of its pseudo object-space coordinates and the multi-view image matching of confirming its homologous pixels. That is, they both need the participation of all candidate homologous pixels in the searching range of each searching image. So, the problem of multi-view image matching can be transformed into a surveying adjustment problem of pseudo object-space coordinates to analyze the positional distribution pattern among candidate homologous pixels in the multi-view images.

*Theoretical inference on the positional distribution pattern among candidate homologous pixels.* According to the reliability theory of surveying adjustment system[50], for $n$ independent observations with a single gross error, the standardized residual $\omega_i$ ($i = 1,2,…,n$) of each observation after adjustment obeys the normal distribution with mean value $\delta$ and variance 1, and its distribution is in the shape of a Gaussian bell (shown in Fig. 3a). If this theory is applied to the adjustment process of the matching pixel's pseudo object-space coordinates, and the group of real homologous pixels in all candidate pixels corresponding to the matching pixel is seen as one "special" gross error, then the problem of multi-view image matching can be regarded as an "special" surveying adjustment problem using $n$ independent observations with a single gross error. Therefore, the following theoretical inference about the positional distribution pattern among candidate homologous pixels in multi-view images is made: after adjustment of pseudo object-space coordinates, the standardized residuals of all candidate homologous pixels will also obey one certain distribution pattern, which is in the form of an opposite shape of Fig. 3a, like an approximate V-shaped distribution (shown in Fig. 3b).

*Computation of standardized residuals for candidate homologous pixels.* Based on the above theoretical inference, to analyze the positional distribution pattern among candidate homologous pixels of a matching pixel, their standardized residuals should be firstly calculated. The computing process is described as follows: (1) For matching pixel $p$, the searching range of its candidate homologous pixels in each searching image is identified under the constraint of a predefined elevation range in the object-space; (2) For different searching ranges in different searching images, their lengths are refined to a uniform value for uniformly assigning independent candidate homologous pixels in different searching images to different groups. (3) The adjustment computation of pseudo object-space coordinates of the matching pixel is carried out to obtain the standardized residual of each group of candidate homologous pixels.

Determining candidate homologous pixels for the matching pixel based on the constraint of elevation range in object space. As shown in Fig. 4, for matching pixel $p$ in base image $I_0$, its candidate homologous pixels in each searching image are determined by the following method under the constraint of object-space approximate elevation range.

① According to a priori knowledge interactively input by the operator, the approximate object-space minimum and maximum elevations $Z_{min}$ and $Z_{max}$ are assigned to the base image. It is not necessary to be accurate, just to ensure that the range $[Z_{min}, Z_{max}]$ includes the real elevation of the matching pixel $p$. Then the ground point $P$ corresponding to the pixel $p$ must locate between the line segment $P_{min}P_{max}$ along the projection line $S_0P$. Utilizing the collinearity equation (shown in Eq. (1)), object-space plane coordinates $(X_{min}, Y_{min})$ and $(X_{max}, Y_{max})$ of the points $P_{min}$ and $P_{max}$ can be figured out by the image plane coordinates of pixel $p$ and elevations $Z_{min}$ and $Z_{max}$,

$$\begin{cases} X = X_0 + (Z - Z_0)\frac{a_1^0 x + a_2^0 y - a_3^0 f}{c_1^0 x + c_2^0 y - c_3^0 f} \\ Y = Y_0 + (Z - Z_0)\frac{b_1^0 x + b_2^0 y - b_3^0 f}{c_1^0 x + c_2^0 y - c_3^0 f} \end{cases} \tag{1}$$

where, $f$ is the focal length in the camera's interior orientation parameters; coefficients ($a_1^0$, $a_2^0$, $a_3^0$; $b_1^0$, $b_2^0$, $b_3^0$; $c_1^0$, $c_2^0$, $c_3^0$) are nine direction cosines in the rotation matrix determined by the exterior rotational angles ($\varphi_0$, $\omega_0$, $\kappa_0$) of the base image.
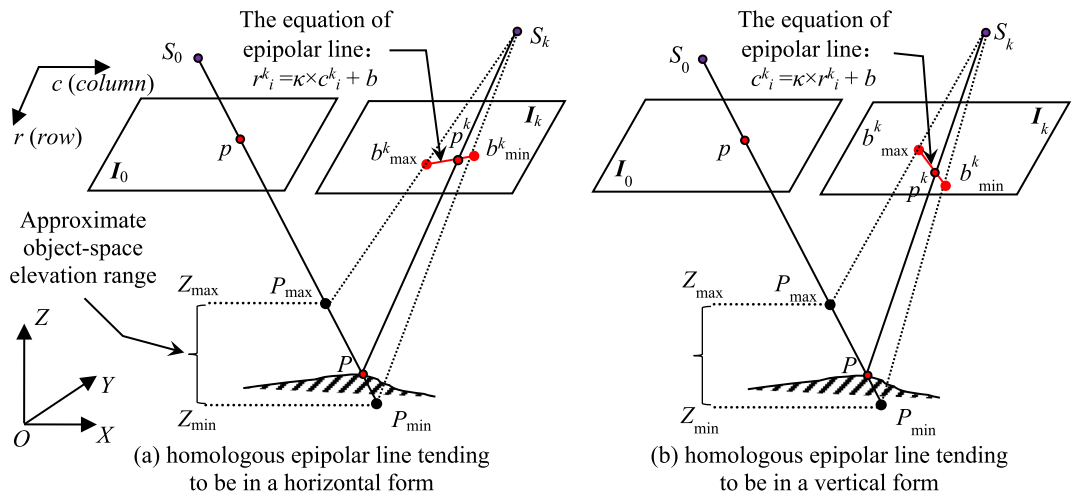
5

**Figure 4.** The sketch of determining candidate homologous pixels based on the object-space elevation range.

② Utilizing the collinearity equation (shown in Eq. (2)), the object-space points $P_{min}(X_{min}, Y_{min}, Z_{min})$ and $P_{max}(X_{max}, Y_{max}, Z_{max})$ are re-projected onto the $k$th searching image $I_k$ to obtain the corresponding pixels $b^k_{min}(x^k_{min}, y^k_{min})$ and $b^k_{max}(x^k_{max}, y^k_{max})$. The homologous pixel $p^k$ of the pixel $p$ in the image $I_k$ must locate on the line segment $b^k_{min}b^k_{max}$, which is actually a segment of the homologous epipolar line. The segment $b^k_{min}b^k_{max}$ is the searching range of the homologous pixel $p^k$ in the image $I_k$, and all $M_k$ pixels located on $b^k_{min}b^k_{max}$ are considered as candidate homologous pixels. Next, according to the camera's interior orientation parameters, row and column numbers $(r^k_{min}, c^k_{min})$ and $(r^k_{max}, c^k_{max})$ of the points $b^k_{min}$ and $b^k_{max}$ are figured out. Then the linear equation of epipolar line $b^k_{min}b^k_{max}$ can be defined, based on which the row and column numbers $(r^k_i, c^k_i)$ of the $i$th candidate pixel $p^k_i$ in the image $I_k$ are obtained ($i = 1, 2, …, M_k$).

$$\begin{cases} x = -f\frac{a_1(X-X_k)+b_1(Y-Y_k)+c_1(Z-Z_k)}{a_3(X-X_k)+b_3(Y-Y_k)+c_3(Z-Z_k)} \\ y = -f\frac{a_2(X-X_k)+b_2(Y-Y_k)+c_2(Z-Z_k)}{a_3(X-X_k)+b_3(Y-Y_k)+c_3(Z-Z_k)} \end{cases} \tag{2}$$

where, coefficients $(a_1, a_2, a_3; b_1, b_2, b_3; c_1, c_2, c_3)$ are nine direction cosines in the rotation matrix determined by the exterior rotational angles $(\varphi_k, \omega_k, \kappa_k)$ of the $k$th searching image $I_k$; $(X_k, Y_k, Z_k)$ are object-space 3D coordinates of the projection center of the image $I_k$.

In Fig. 4, there are two cases of the form of epipolar line in each searching image, and the linear equation of each case is confirmed as follows. In following description, $int()$ is the integral function, $abs()$ is the absolute value function, $min()$ is the minimum function, and $max()$ is the maximum function.

a. If $abs(c^k_{max} - c^k_{min}) \geq abs(r^k_{max} - r^k_{min})$, the epipolar line tends to be horizontal (shown in Fig. 4a), and its linear equation is $r^k_i = \kappa \times c^k_i + b$. Where, $\kappa$ and $b$ are respectively the slope and intercept of the line, and are computed by equations $\kappa = (r^k_{max} - r^k_{min})/(c^k_{max} - c^k_{min})$ and $b = r^k_{max} - \kappa \times c^k_{max}$. The range of values of variable $c^k_i$ is $[min(c^k_{max}, c^k_{min}), max(c^k_{max}, c^k_{min})]$, and the total number of candidate homologous pixels is computed by equation $M_k = int(abs(c^k_{max} - c^k_{min}) + 1)$.

b. If $abs(c^k_{max} - c^k_{min}) < abs(r^k_{max} - r^k_{min})$, the epipolar line tends to be vertical (shown in Fig. 4b), and its linear equation is $c^k_i = \kappa \times r^k_i + b$. Corresponding slope and intercept of the line are computed by equations $\kappa = (c^k_{max} - c^k_{min})/(r^k_{max} - r^k_{min})$ and $b = c^k_{max} - \kappa \times r^k_{max}$. The range of values of variable $r^k_i$ is $[min(r^k_{max}, r^k_{min}), max(r^k_{max}, r^k_{min})]$, and the total number is computed by equation $M_k = int(abs(r^k_{max} - r^k_{min}) + 1)$.

**Uniformly grouping candidate homologous pixels in searching images.** As seen from above description, for the matching pixel $p$, the total number $M_k$ of its candidate homologous pixels in each searching image are different. Furthermore, different candidate pixels in different searching images are independent from each other, and it is difficult to construct any relations among these pixels utilizing their serial numbers. Thus, to analyze the distribution pattern among candidate homologous pixels, the total number of candidate homologous pixels in each searching image is unified by the following method, which can uniformly assign independent candidate pixels to different groups (the sketch is shown in Fig. 5).

① According to the number of candidate homologous pixels in each searching image, the image with the maximal number is used as the main searching image (like image $I_1$ in Fig. 5), and remaining $N-2$ images are named as vice searching image $I_t$ ($t = 2, …, N-1$). The total number of candidate homologous pixels in the main searching image is labeled as $M_z$, and its value is computed by the equation $M_z = max(M_k | k = 1, 2,…, N-1)$.

② For the $i$th candidate pixel $b^1_i(r^1_i, c^1_i)$ on the homologous epipolar line $b^1_1b^1_{Mz}$ in the main searching image ($i = 1, 2,…, M_z$), the candidate object-space coordinates $(X_i, Y_i, Z_i)$ are firstly figured out with the pixels $p$ and $b^1_i$. Next, using the collinearity equation and the linear equation of homologous epipolar line $b^t_1b^t_{Mz}$ in the $t$th vice searching image $I_t$, the point $P_i$ is re-projected onto the image $I_t$ to yield the candidate pixel $b^t_i$ corresponding
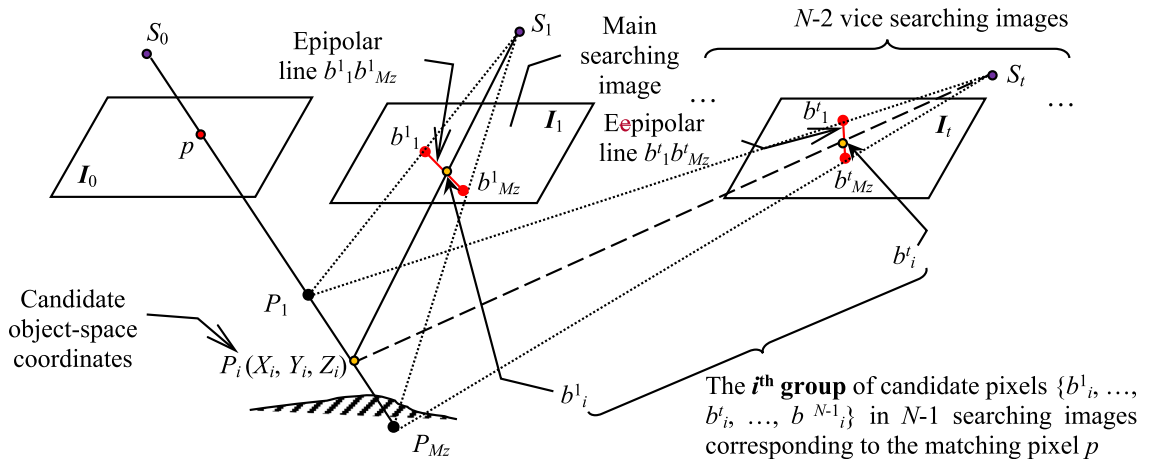
**Figure 5.** The sketch of uniformly grouping candidate homologous pixels in multi-view images.

to the pixel $b^1_i$. Then these pixels $\{b^1_i, \ldots, b^t_i, \ldots, b^{N-1}_i\}$ are organized as the $i$th group of candidate homologous pixels of the pixel $p$.

③ After all candidate pixels in the main searching image are processed by the step 2, the number of candidate homologous pixels corresponding to the matching pixel $p$ in each searching image will have the same value $M_z$, and the candidate pixels with the same serial number belong to the same group. Therefore, the original independent candidate pixels in all searching images is unified into the following $M_z$ groups: $\{b^1_1, b^2_1, \ldots, b^{N-1}_1\}$, $\{b^1_2, b^2_2, \ldots, b^{N-1}_2\}$, $\ldots$, $\{b^1_i, b^2_i, \ldots, b^{N-1}_i\}$, $\ldots$, $\{b^1_{Mz}, b^2_{Mz}, \ldots, b^{N-1}_{Mz}\}$ (the group serial number $i$ is from 1 to $M_z$).

To be noted, after completing the unifying process of candidate pixels, except for the main searching image, there are repetitive pixels in $M_z$ candidate pixels $\{b^t_1, b^t_2, \ldots, b^t_{Mz}\}$ in the $t$th vice searching image $I_t$, and repetitive pixels must be neighbor pixels with consecutive serial numbers.

**Computing standardized residual for each group of candidate homologous pixels based on surveying adjustment theory.** According to the surveying adjustment theory, utilizing relevant information during the adjustment of pseudo object-space coordinates, the standardized residual of each candidate homologous pixel can be figured out and used to analyze the positional distribution pattern among candidate pixels. For $M_z$ groups of candidate homologous pixels in searching images: $\{b^1_1, b^2_1, \ldots, b^{N-1}_1\}$, $\ldots$, $\{b^1_{Mz}, b^2_{Mz}, \ldots, b^{N-1}_{Mz}\}$ corresponding to the matching pixel $p$, the standardized residual $\omega_i$ of the $i$th group of candidate pixels $\{b^1_i, b^2_i, \ldots, b^{N-1}_i\}$ is computed as follows:

① Removing repetitive candidate homologous pixels in the searching image. For the $M_z$ candidate pixels $\{b^k_1, \ldots, b^k_j, \ldots, b^k_{Mz}\}$ ($j = 1, \ldots, M_z$) in the $k$th searching image $I_k$ ($k = 1, \ldots, N-1$), starting from the second candidate pixel, each pixel is judged whether it repeats with the previous adjacent pixel one by one. If yes, it will be marked as a repetitive pixel. After all $M_z$ candidate pixels are processed by the repeatability judgement, $\overline{M}_k(\overline{M}_k \leq M_z)$ new candidate pixels $\{q^k_1, \ldots, q^k_{j'}, \ldots q^k_{\overline{M}_k}\}$ ($j' = 1, \ldots, \overline{M}_k$) without repetitive pixels are obtained, and the sum of the matching pixel $p$ and its candidate homologous pixels is $M$: $M = (1 + \sum_{k=1}^{N-1} \overline{M}_k)$. Furthermore, the correspondence relation between the original serial number $j$ and the new serial number $j'$ is known.

② Constructing the error equation matrix of all $M$ pixels. According to the computing of pseudo object-space coordinates in "Definition and computation of pseudo object-space coordinates", for the matching pixel and its $\overline{M}_k$ candidate homologous pixels in each searching image, the error equation matrix (shown in Eq. (3)) for solving the pseudo object-space coordinates of the pixel $p$ can be formed.

$$\mathbf{V}_{2M\times1} = \mathbf{A}_{2M\times3}\mathbf{X}_{3\times1} - \mathbf{L}_{2M\times1} \tag{3}$$

Detailed forms of matrixes $\mathbf{A}_{2M\times3}$ and $\mathbf{L}_{2M\times1}$ are shown in Eq. (4). Where, the superscript of each element ($I_0, I_1, \ldots, I_{N-1}$) represents the label of the image, and the number from 1 to $\overline{M}_1$ or $\overline{M}_{N-1}$ in the subscript of each element represents the serial number of candidate homologous pixel in different searching image.

$$
\left\{
\begin{aligned}
&\mathbf{V}_{2M\times1} = \left[v_x^{I_0}, v_y^{I_0}, v_{x,1}^{I_1}, v_{y,1}^{I_1}, \cdots, v_{x,M'_1}^{I_1}, v_{y,M'_1}^{I_1}, \cdots, v_{x,1}^{I_{N-1}}, v_{y,1}^{I_{N-1}}, \cdots, v_{x,\overline{M}_{N-1}}^{I_{N-1}}, v_{y,\overline{M}_{N-1}}^{I_{N-1}}\right]^T \\
&\mathbf{X}_{3\times1} = \left[dX\ dY\ dZ\right]^T \\
&\mathbf{A}_{2M\times3} =
\begin{bmatrix}
a_{11}^{I_0} & a_{12}^{I_0} & a_{13}^{I_0} \\
a_{21}^{I_0} & a_{22}^{I_0} & a_{23}^{I_0} \\
a_{11,1}^{I_1} & a_{12,1}^{I_1} & a_{13,1}^{I_1} \\
a_{21,1}^{I_1} & a_{22,1}^{I_1} & a_{23,1}^{I_1} \\
& \vdots & \\
a_{11,\overline{M}_1}^{I_1} & a_{12,\overline{M}_1}^{I_1} & a_{13,\overline{M}_1}^{I_1} \\
a_{21,\overline{M}_1}^{I_1} & a_{22,\overline{M}_1}^{I_1} & a_{23,\overline{M}_1}^{I_1} \\
& \vdots & \\
a_{11,1}^{I_{N-1}} & a_{12,1}^{I_{N-1}} & a_{13,1}^{I_{N-1}} \\
a_{21,1}^{I_{N_k}} & a_{22,1}^{I_{N_k}} & a_{23,1}^{I_{N_k}} \\
& \vdots & \\
a_{11,\overline{M}_{N-1}}^{I_{N-1}} & a_{12,\overline{M}_{N-1}}^{I_{N-1}} & a_{13,\overline{M}_{N-1}}^{I_{N-1}} \\
a_{21,\overline{M}_{N-1}}^{I_{N-1}} & a_{22,\overline{M}_{N-1}}^{I_{N-1}} & a_{23,\overline{M}_{N-1}}^{I_{N-1}}
\end{bmatrix}
, \mathbf{L}_{2M\times1} =
\begin{bmatrix}
x^{I_0} - \left(x^{I_0}\right)^0 \\
y^{I_0} - \left(y^{I_0}\right)^0 \\
x_1^{I_1} - \left(x_1^{I_1}\right)^0 \\
y_1^{I_1} - \left(y_1^{I_1}\right)^0 \\
\vdots \\
x_{M'_1}^{I_1} - \left(x_{\overline{M}_1}^{I_1}\right)^0 \\
y_{M'_1}^{I_1} - \left(y_{\overline{M}_1}^{I_1}\right)^0 \\
\vdots \\
x_1^{I_{N-1}} - \left(x_1^{I_{N-1}}\right)^0 \\
y_1^{I_{N-1}} - \left(y_1^{I_{N-1}}\right)^0 \\
\vdots \\
x_{\overline{M}_{N-1}}^{I_{N-1}} - \left(x_{\overline{M}_{N-1}}^{I_{N-1}}\right)^0 \\
y_{\overline{M}_{N-1}}^{I_{N-1}} - \left(y_{\overline{M}_{N-1}}^{I_{N-1}}\right)^0
\end{bmatrix}
\end{aligned}
\right.
\tag{4}
$$

③ Computing the reliability matrix. Utilizing the coefficient matrix $\mathbf{A}$, the constant matrix $\mathbf{L}$, and the weight matrix $\mathbf{P}_{ll}$ of observations (as all observations in this paper are image plane coordinates of pixels and are obtained independently, the weight matrix is a unit matrix $\mathbf{E}$), the reliability matrix $\mathbf{R}$ can be derived by the following equation.

$$
\mathbf{R}_{2M\times2M} = \left(\mathbf{P}_{ll}^{-1} - \mathbf{A}\left(\mathbf{A}^T\mathbf{P}_{ll}\mathbf{A}\right)^{-1}\mathbf{A}^T\right)\mathbf{P}_{ll}
\tag{5}
$$

④ Resolving the error equation matrix. Based on the principle of least squares, the error equation matrix shown in Eq. (3) is resolved to obtain the coordinates correction matrix $\mathbf{X}$, the residual matrix $\mathbf{V}$ of all observations and the root mean square error of unit weight $\sigma_0$.

$$
\left\{
\begin{aligned}
X &= \left(\mathbf{A}^T\mathbf{P}_{ll}\mathbf{A}\right)^{-1}\left(\mathbf{A}^T\mathbf{P}_{ll}\mathbf{L}\right) \\
V &= AX - L \\
\sigma_0 &= \sqrt{\mathbf{V}^T\mathbf{P}_{ll}\mathbf{V}/(2M-3)}
\end{aligned}
\right.
\tag{6}
$$

⑤ Computing the standardized residual $\omega_n$ of the $n$th candidate pixel in all $M$ pixels ($n \in [1, M]$). As there are two observations of one pixel at the $x$ and $y$ coordinate directions, the standardized residual of each candidate pixel is computed by taking the average of standardized residuals of observations $x$ and $y$, as shown in Eq. (7).

$$
\omega_n = \left(\frac{abs\left((\mathbf{V}_{ll})_{2\boldsymbol{n},0}\right)}{\sqrt{(\mathbf{R})_{2\boldsymbol{n},2\boldsymbol{n}}} \times \sigma_0/\sqrt{(\mathbf{P}_{ll})_{2\boldsymbol{n},2\boldsymbol{n}}}} + \frac{abs\left((\mathbf{V}_{ll})_{2n+1,0}\right)}{\sqrt{(\mathbf{R})_{2n+1,2n+1}} \times \sigma_0/\sqrt{(\mathbf{P}_{ll})_{2n+1,2n+1}}}\right)/2
\tag{7}
$$

⑥ Computing the standardized residual $\omega(b_j^k)$ of the $j$th candidate pixel $b_j^k$ in the $k$th searching image $I_k$. Known from the repeatability judgement of the step (1), serial number $j$ of pixel $b_j^k$ corresponds to serial number $j'$ in the new candidate pixels without repetitive pixels. In addition, the total number of the matching pixel and its candidate pixels in $(k-1)$ searching images processed by the repetitiveness assessment is $M':M' = (1 + \sum_{i'=1}^{k-1} \overline{M}_{i'})$. So, $\omega(b_j^k)$ is computed as follows.

$$
\omega\left(b_j^k\right) = \omega_{M'+j'}
\tag{8}
$$

⑦ Computing the standardized residual $\omega_i$ of the $i$th group of candidate homologous pixels. As the $i$th group of candidate pixels consist of $N-1$ candidate pixels, $\omega_i$ is computed by taking the average of standardized residuals of $N-1$ candidate pixels, as shown in Eq. (9).

$$
\omega_i = \frac{\sum_{k=1}^{N-1} \omega\left(b_i^k\right)}{N-1}
\tag{9}
$$

(a) Experimental multi-view images captured in Nanjing, including some of the homologous pixels.



(b) Experimental multi-view images in ISPRS benchmark dataset, including some of the homologous pixels.
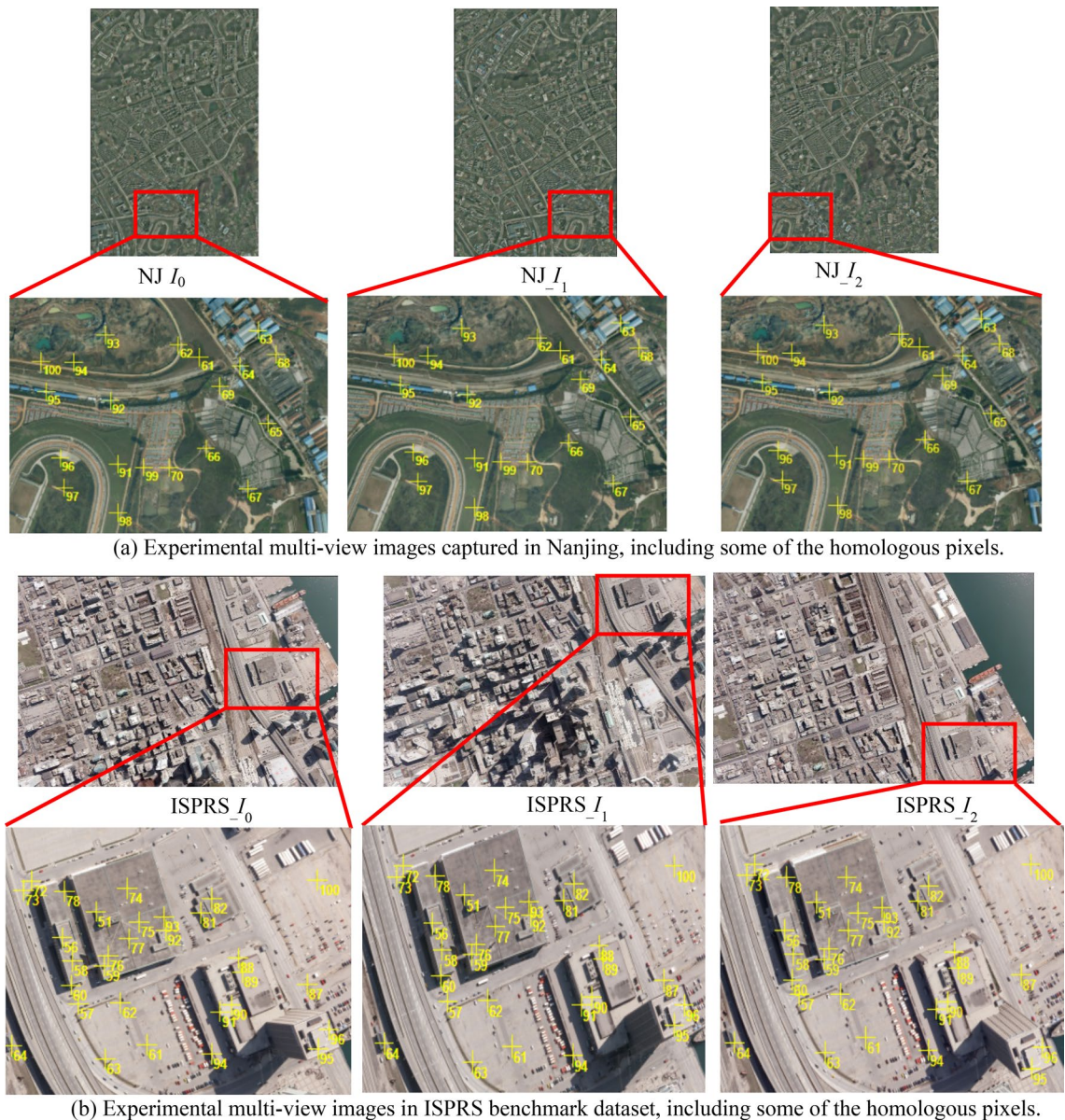
**Figure 6.** Two types of experimental aerial images (Images in sub-figure (**a**) were captured by our partner Jiangsu Digitalland Technology Corporation, Ltd., Co. with an UltraCam metric digital camera in Nanjing China in November 2011, and images in sub-figure (**b**) are internationally public benchmark data in Toronto Canada provided by ISPRS Working Group III/4: http://www2.isprs.org/commissions/comm3/wg4).

## Experiment and discussion

**Experimental data.** To verify the proposed approach, experiments were carried out using two sets of actual aerial multi-view images. Figure 6a shows three aerial images of Nanjing, China (captured by Jiangsu Digitalland Technology Corporation, Ltd., Co., using an UltraCam metric digital camera in November 2011), and Fig. 6b shows three aerial images of Toronto, Canada (public data in the ISPRS benchmark dataset, obtained through the ISPRS official website: http://www2.isprs.org/commissions/comm3/wg4). For the two datasets, the base image and two searching images are respectively labeled as $NJ\_I_0$, $NJ\_I_1$, $NJ\_I_2$ and $ISPRS\_I_0$, $ISPRS\_I_1$, $ISPRS\_I_2$, the approximate elevation ranges of the region covered by these images are respectively [– 28 m, 188 m] and [– 20 m, 210 m], the forward overlap of these images is around 60%, and their interior and exterior orientation parameters are all known (detailed parameters are listed in Tables 1 and 2). In addition, 100 groups of actual homologous pixels were semi-automatically selected by an image-side multi-view matching method based on NCC[10] for each experimental data, and were used as ground truth for the following two experiments (Part of matching pixels in the base image and their homologous pixels in the two searching images are depicted in Fig. 6. The yellow cross represents their positions, and the number beside the cross is the serial number). Image-space row and column numbers and object-space 3D coordinates of these homologous pixels are respectively labeled as $(r^k_j, c^k_j)$ and $(X_j, Y_j, Z_j)$ ($k = 0$, $k = 1$, $k = 2$ is respectively the serial number of the based image, the first and the second searching images; and $j = 1, 2, …, 100$).

| Image type | Image-space resolution [pixel] | Focal length [mm] | The position of principal point [pixel] | | Pixel size [mm] |
|---|---|---|---|---|---|
| | | | Row | Column | |
| Nanjing aerial images | 11,310 columns × 17,310 rows | 70.5 | 8634 | 5654 | 0.006 |
| ISPRS benchmark images | 11,500 columns × 7500 rows | 101.4 | 3769 | 5749 | 0.009 |

**Table 1.** Interior orientation parameters of experimental aerial images.

| Image type and tagID | | WGS-84 coordinates of projection centres | | | exterior rotational angles | | |
|---|---|---|---|---|---|---|---|
| | | $X$[m] | $Y$[m] | $Z$[m] | $\varphi$ [rad] | $\omega$[rad] | $\kappa$[rad] |
| Nanjing aerial images | NJ_$I_0$ | 397,361.04579 | 3,552,859.53939 | 3718.27647 | − 0.008873 | 0.015662 | − 0.014234 |
| | NJ_$I_1$ | 396,163.27625 | 3,552,863.48668 | 3718.31694 | − 0.007112 | 0.009475 | − 0.011622 |
| | NJ_$I_2$ | 398,558.69302 | 3,552,849.68689 | 3718.66453 | − 0.008784 | 0.015934 | − 0.017909 |
| ISPRS benchmark images | ISPRS_$I_0$ | 631,179.232 | 4,834,062.777 | 1636.415 | 0.001505 | 0.000377 | − 1.751372 |
| | ISPRS_$I_1$ | 630,789.997 | 4,834,064.034 | 1632.821 | 0.002248 | 0.000534 | − 1.748176 |
| | ISPRS_$I_2$ | 631,567.432 | 4,834,062.247 | 1640.148 | 0.000473 | 0.000852 | − 1.749056 |

**Table 2.** Exterior orientation parameters of experimental aerial images.

**Experimental design.**    Utilizing above data, two experiments were designed as follows: experiment **I** was used to verify the theoretical inference on the positional distribution pattern among candidate homologous pixels, and experiment **II** was used to further test whether the positional distribution pattern can be applied in multi-view image matching.

Experiment **I**: For each matching pixel in the base image, firstly its initial set of $m$ groups of candidate homologous pixels in the two searching images are confirmed and their standardized residuals computed, and the standardized residual curve is depicted to illustrate the positional distribution form among candidate homologous pixels. Secondly, if the curve of the initial set of candidate homologous pixels obeys an approximate V-shaped form as the theoretical inference, the following process is iteratively performed:

Using the valley point (the point with the minimum standardized residual) on the curve as the boundary, divide the initial set of candidate homologous pixels into two subsets, and take the subset including the real homologous pixels as a new initial set of candidate pixels (supposing that it has $n$ groups of pixels). ② If $n \neq m$, compute standardized residuals and draw distribution curve for the new set of $n$ groups of candidate pixels again. ③ Judge whether the distribution curve of the new set of candidate pixels still obeys the approximate V-shaped form, if yes, perform step ① again; if not, or the total number of pixels in the new set do not change, stop the iteration.

According to the characteristic of V-shaped curve, on the standardized residual distribution curve obtained in each iteration above, the point with the minimum standardized residual should be located near the middle position of all pixels. Furthermore, since two points cannot form a V-shaped curve hence will stop the iteration, the total number $n$ of candidate homologous pixels at the last iteration in each searching image should be two (one of which is the real homologous pixel). This value of $n$ is much smaller than that of initial $m$. As the total number of candidate homologous pixels decreases gradually in the iterative process, it is considered as the iterative refinement of the searching range of homologous pixels based on the positional distribution pattern among candidate pixels. In this iterative refinement process, for each searching image, the output results of each iteration before the last are the row and column numbers of two endpoints of the searching range and the position of the real homologous pixel on either the left or right side of the pixel with the minimum standardized residual, and the output of the last iteration is two candidate pixels.

Experiment **II**: According to the iterative refinement of the searching range of homologous pixels in experiment **I**, the total number $n$ of candidate homologous pixels corresponding to the matching pixel in the $k$th searching image $I_k$ is two. So the average value of the row and column numbers ($r_i$, $c_i$) of these $n$ candidate pixels is used as the row and column numbers ($r^k$, $c^k$) of the final homologous pixel in the image $I_k$: $r^k = \frac{\sum_{i=1}^n r_i}{n}$; $c^k = \frac{\sum_{i=1}^n c_i}{n}$, and this process of determining the homologous pixel is considered as a new object-side multi-view matching based on the positional distribution pattern among candidate homologous pixels. The reason for adopting Average operation rather than Max or Min operation is as follows: either the Max operation or the Min operation includes the real homologous pixel, the accuracy of this operation for different matching pixel is not guaranteed, and its image-space precision of the homologous pixel is 1 pixel or 0. Whereas, the Average operation can ensure that the image-space precision is 0.5 pixel.

Therefore, for the 100 matching pixels in experimental images, the new object-side multi-view matching was firstly carried out to obtain the row and column numbers ($r^k_{j,O}$, $c^k_{j,O}$) ($k = 1, 2$; $j = 1, 2,\ldots, 100$) of homologous pixels in each searching image. Then based on the image-space matching results, object-space 3D coordinates ($X_{j,O}, Y_{j,O}, Z_{j,O}$) were figured out according to the principle of multi-view forward intersection. Secondly, utilizing obtained image-space/object-space matching results and corresponding ground truth, image-space row error

$\Delta r^k_j$ and column error $\Delta c^k_j$ and object-space plane error $\Delta P_j$ and elevation error $\Delta E_j$ of matching pixels were computed by equations: $\Delta r^k_j = r^k_{j,O} - r^k_j$; $\Delta c^k_j = c^k_{j,O} - c^k_j$, $\Delta P_j = \sqrt{\left(X_{j,O} - X_j\right)^2 + \left(Y_{j,O} - Y_j\right)^2}$; $\Delta E_j = Z_{j,O} - Z_j$. The curves of image-space and object-space matching errors of all matching pixels were respectively drawn. Thirdly, to verify that the object-side multi-view matching method does not depend on image-side grayscale information, original searching images were randomly adjusted in grayscale, brightness and hue with Photoshop (mainly used to simulate condition such as strong/weak light, repetitive/weak texture, etc.), and were used again in object-side multi-view matching. Image-space and object-space errors ($\Delta r^k_j$, $\Delta c^k_j$) and ($\Delta P_j$, $\Delta E_j$) of matching results were also recalculated. Lastly, to compare the object-side matching method with traditional image-side matching method, image-side matching based on NCC was carried out for the adjusted searching images to obtain matching results ($r^k_{j,I}$,$c^k_{j,I}$) and ($X_{j,I}$, $Y_{j,I}$, $Z_{j,I}$), and corresponding image-space and object-space errors ($\Delta r'^k_j$, $\Delta c'^k_j$) and ($\Delta P'_j$, $\Delta E'_j$) were also computed.

**Experimental results.**   The proposed method of analyzing the positional distribution pattern among candidate homologous pixels in multi-view images was realized using the Visual C#.NET 2015 program language, and tested with real aerial images according to the above experimental design. The computer configurations for algorithm verification are: ThinkPad T580 laptop, Win10 64-bit operating system, Intel CORE i7 CPU with 1.99 GHz, and 8G RAM. In the following results, all displayed images were generated by our own program, and all statistical graphs were generated by Microsoft Excel 2016.

*Results of experiment I.*   Using the mode of multi-view images $I_0$-$I_1$-$I_2$, the experiment **I** were carried out for the two types of experimental images. For the 100 matching pixels in Nanjing and ISPRS experimental images respectively, corresponding standardized residual distribution curves of initial *m* groups of candidate homologous pixels in the initial searching range are shown in Fig. 7a. After processed by the iterative refinement of the searching range, standardized residual distribution curves before the last iteration are shown in Fig. 7b, and the residual distribution curves at the last iteration are shown in Fig. 7c.

In Fig. 7, the horizontal axis represents the serial number of each group of candidate pixels (the maximal total number of candidate pixels for each of the 100 matching pixels was used as the boundary), the vertical axis represents the standardized residual of each group of candidate pixels, and the third axis perpendicular to the horizontal axis in the plane represents the serial number of each searching range during the iterative refinement of the searching range (the searching range at the *i*th iteration was labeled as "range *i*").

Need to be noted, to effectively display the standardized residual distribution form of each groups of candidate pixels in each searching range for all matching pixels during the iterative refinement process, the standardized residual corresponding to each group serial number in Fig. 7 is the average of standardized residuals of candidate pixels with this group serial number. These average standardized residuals can better express the whole positional distribution form among candidate homologous pixels during the iterative refinement process.

Furthermore, to clearly demonstrate the positional distribution form among candidate homologous pixels for singular pixel, two matching pixels located on building facade and ground from the two experimental data were taken as examples, and corresponding experimental results were displayed in detail in Figs. 8 and 9. In these figures, the red line in each searching image is the homologous epipolar line defined by the approximate object-space elevation range (that is the initial searching range); and in each sub-figure of the standardized residual distribution curve, each blue dot represents a group of candidate pixels; serial numbers of the first group of candidate pixels, last group of candidate pixels, the group of candidate pixels with the minimum residual and the group of real homologous pixels are all marked with note boxes; and the vertical red dotted line indicates the position of the group of real homologous pixels.

*Results of the experiment II.*   According to the design of the experiment **II**, part of the 100 matching pixels and corresponding image-space experimental results of the new object-side multi-view matching method are shown in Fig. 10(sub-figure (a) shows positions of partial matching pixels in the base image and corresponding real homologous pixels in the two searching images, and sub-figure (b) shows image-space positions of homologous pixels obtained by the object-side multi-view matching method.).

After the grayscale information of Nanjing and ISPRS searching images are adjusted, image-space experimental results of the object-side multi-view matching method for the same matching pixels are respectively shown in Fig. 11a,c. As a contrast, corresponding experimental results of an image-side NCC-based matching method are respectively shown in Fig. 11b,d.

In addition, to clearly display comparative experimental results of the object-side matching method and traditional image-side matching method for original and grayscale adjusted images of the two types of experimental data, corresponding statistical curves of errors ($\Delta r^k_j$, $\Delta c^k_j$) and ($\Delta P_j$, $\Delta E_j$) of the object-side matching results, and ($\Delta r'^k_j$, $\Delta c'^k_j$) and ($\Delta P'_j$, $\Delta E'_j$) of the image-side NCC-based matching results, are respectively shown in Figs. 12 and 13. In these figures, sub-figures (a) and (b) respectively show statistical curve of image-space errors ($\Delta r^k_j$, $\Delta c^k_j$) and object-space errors ($\Delta P_j$, $\Delta E_j$) of object-side multi-view matching results for original and grayscale adjusted experimental images, and sub-figure (c) shows corresponding statistical curve of image-space errors ($\Delta r'^k_j$, $\Delta c'^k_j$) and object-space errors ($\Delta P'_j$, $\Delta E'_j$) of the image-side NCC-based matching results for grayscale adjusted experimental images.

Furthermore, to intuitively display the quantitative effect of the application of positional distribution pattern among pixels in multi-view image matching, the rule that the image-space row and column errors of one matching result are both less than or equal to 1 pixel ($\Delta r \leq 1$ &$\Delta c \leq 1$) is used as the judgment criterion of matching accuracy. The quantitative comparison results of matching accuracy, image-space and object-space errors for
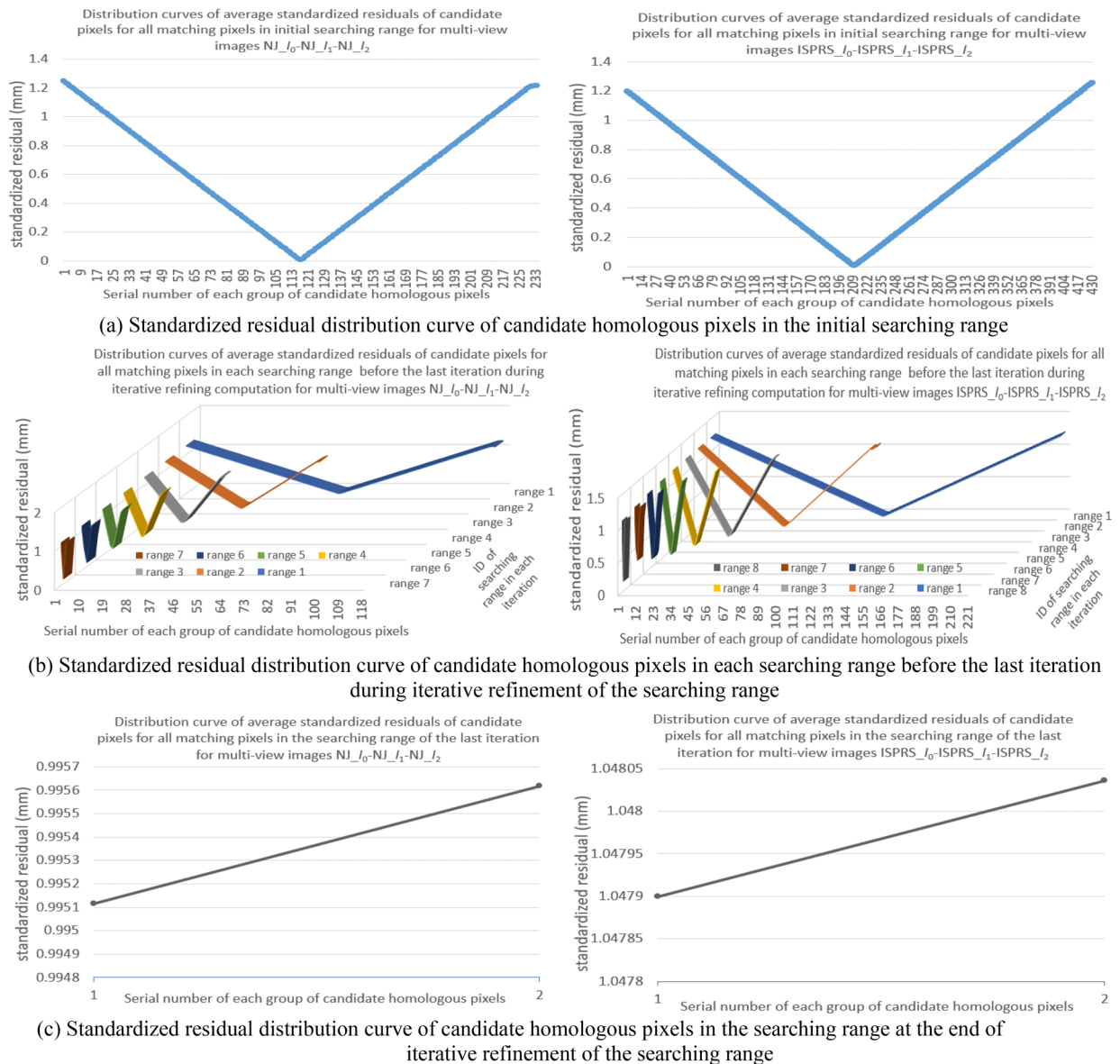
(a) Standardized residual distribution curve of candidate homologous pixels in the initial searching range



(b) Standardized residual distribution curve of candidate homologous pixels in each searching range before the last iteration during iterative refinement of the searching range



(c) Standardized residual distribution curve of candidate homologous pixels in the searching range at the end of iterative refinement of the searching range

**Figure 7.** The experimental results of positional distribution form among candidate homologous pixels for all matching pixels in Nanjing and ISPRS multi-view images (left figures and right figures are the results for images NJ_$I_0$-NJ_$I_1$-NJ_$I_2$ and images ISPRS_$I_0$-ISPRS_$I_1$- ISPRS_$I_2$ respectively, and all these figures were created by Microsoft Excel 2016).

object-side and image-side multi-view matching methods are listed in the Table 3. In addition, the running time of the object-side matching method for the two types of experimental data is respectively 125.126 s and 909 s. As the total number of candidate pixels in the initial searching range and the number of iterative refinement process are not identical, the running time for the two types of experimental data are different. Although the efficiency is not high, it is not the emphasis of this study, which is whether the positional distribution pattern is valid for confirming the exact homologous pixels in multi-view matching and robust to the distortion of image-side grayscale information. In future work, we will try to adopt acceleration techniques such as parallel computing, GPU acceleration and code optimization in the algorithm implementation process to improve running efficiency of our method.

**Analysis and discussion.** Seen from above experimental results in Figs. 7, 8, 9 of experiment I, for any matching pixel in the base image of Nanjing multi-view images or ISPRS benchmark images (no matter where it is located), its candidate homologous pixels in the searching range on each searching image obey the following positional distribution pattern: First, when the total number of candidate homologous pixels is more than two, the standardized residual distribution curve of all candidate pixels presents an approximate V-shaped form. Second, on the V-shaped distribution curve, the valley point with the minimum residual basically locates at the
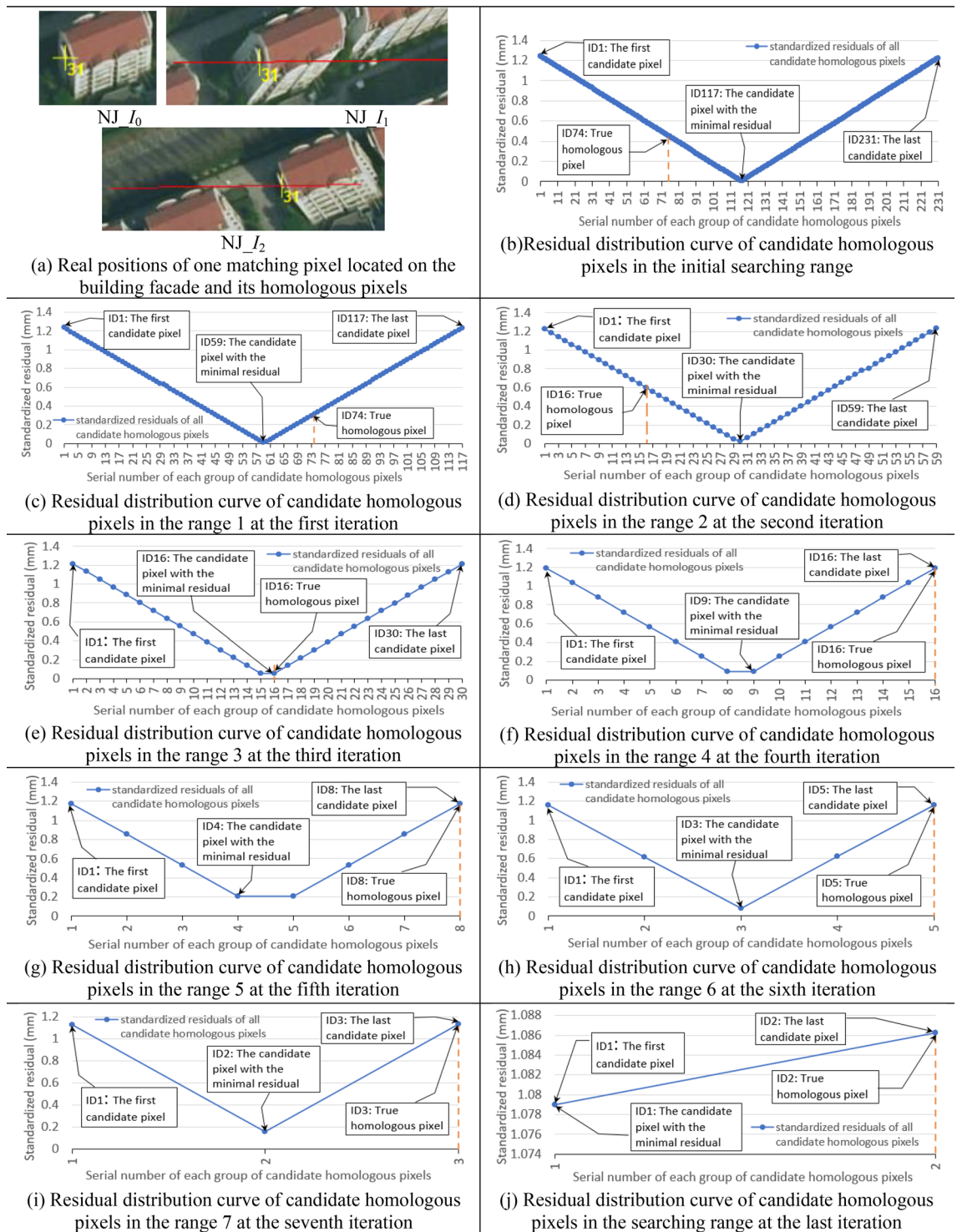
**Figure 8.** The experimental results of positional distribution form among candidate homologous pixels for one matching pixel located on the building facade in Nanjing experimental images NJ_$I_0$-NJ_$I_1$-NJ_$I_2$ (Graphs in (b) to (j) were created by Microsoft Excel 2016).

middle of all candidate pixels, and the real homologous pixel lies on either the left or right of the valley point (including this valley point).

Utilizing the V-shaped distribution pattern among candidate homologous pixels, the iterative refinement of searching range in the searching image can be realized. After refinement, the total number of candidate homologous pixels in each searching image will be reduced from initial hundreds (depending on the size of the initial searching range) to final two adjacent pixels containing the real homologous pixel. However, for different images
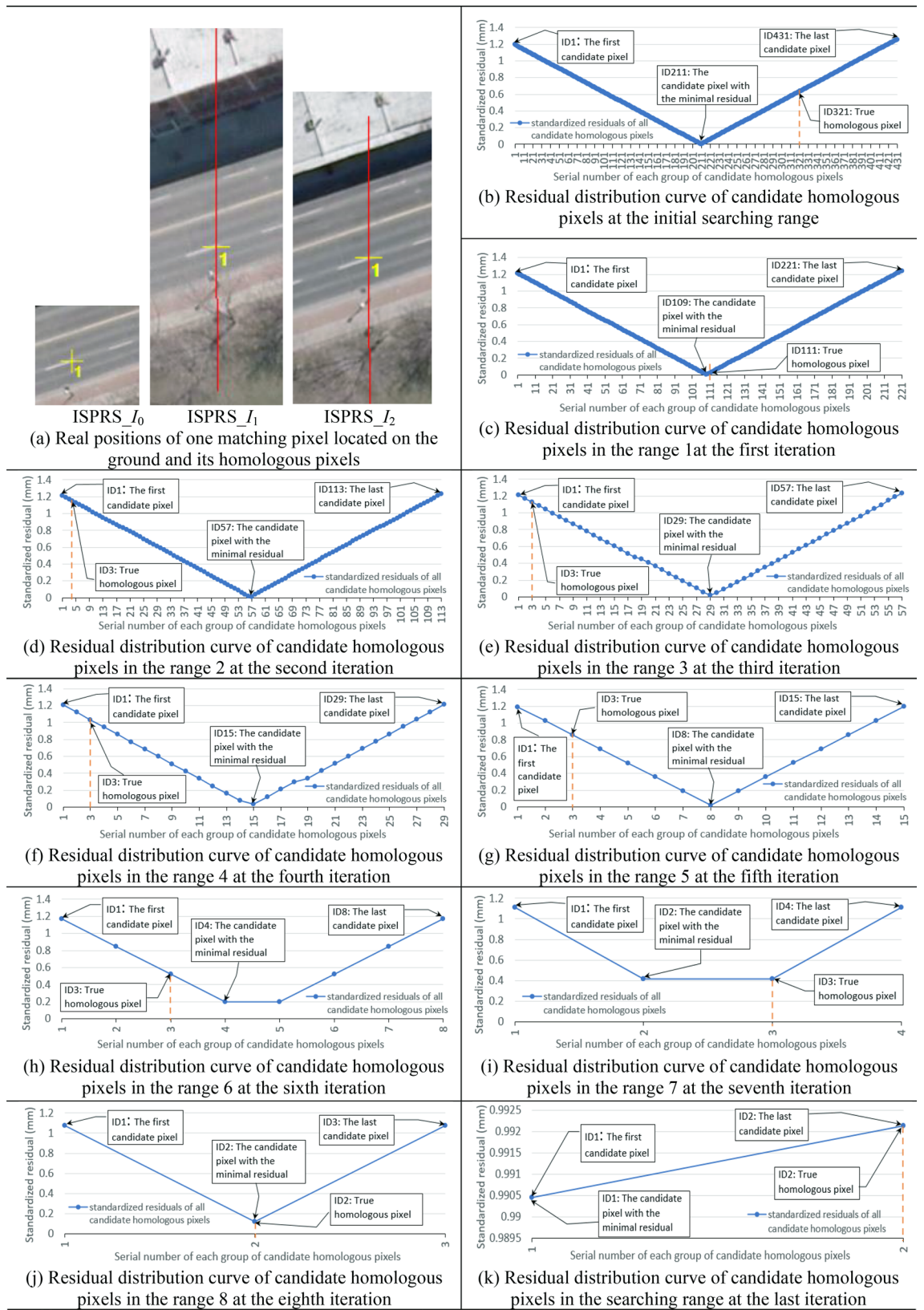
(a) Real positions of one matching pixel located on the ground and its homologous pixels

(b) Residual distribution curve of candidate homologous pixels at the initial searching range

(c) Residual distribution curve of candidate homologous pixels in the range 1 at the first iteration

(d) Residual distribution curve of candidate homologous pixels in the range 2 at the second iteration

(e) Residual distribution curve of candidate homologous pixels in the range 3 at the third iteration

(f) Residual distribution curve of candidate homologous pixels in the range 4 at the fourth iteration

(g) Residual distribution curve of candidate homologous pixels in the range 5 at the fifth iteration

(h) Residual distribution curve of candidate homologous pixels in the range 6 at the sixth iteration

(i) Residual distribution curve of candidate homologous pixels in the range 7 at the seventh iteration

(j) Residual distribution curve of candidate homologous pixels in the range 8 at the eighth iteration

(k) Residual distribution curve of candidate homologous pixels in the searching range at the last iteration

**Figure 9.** The experimental results of positional distribution form among candidate homologous pixels for one matching pixel located on the ground in ISPRS experimental images ISPRS_$I_0$-ISPRS_$I_1$-ISPRS_$I_2$ (Graphs in (b) to (k) were created by Microsoft Excel 2016).
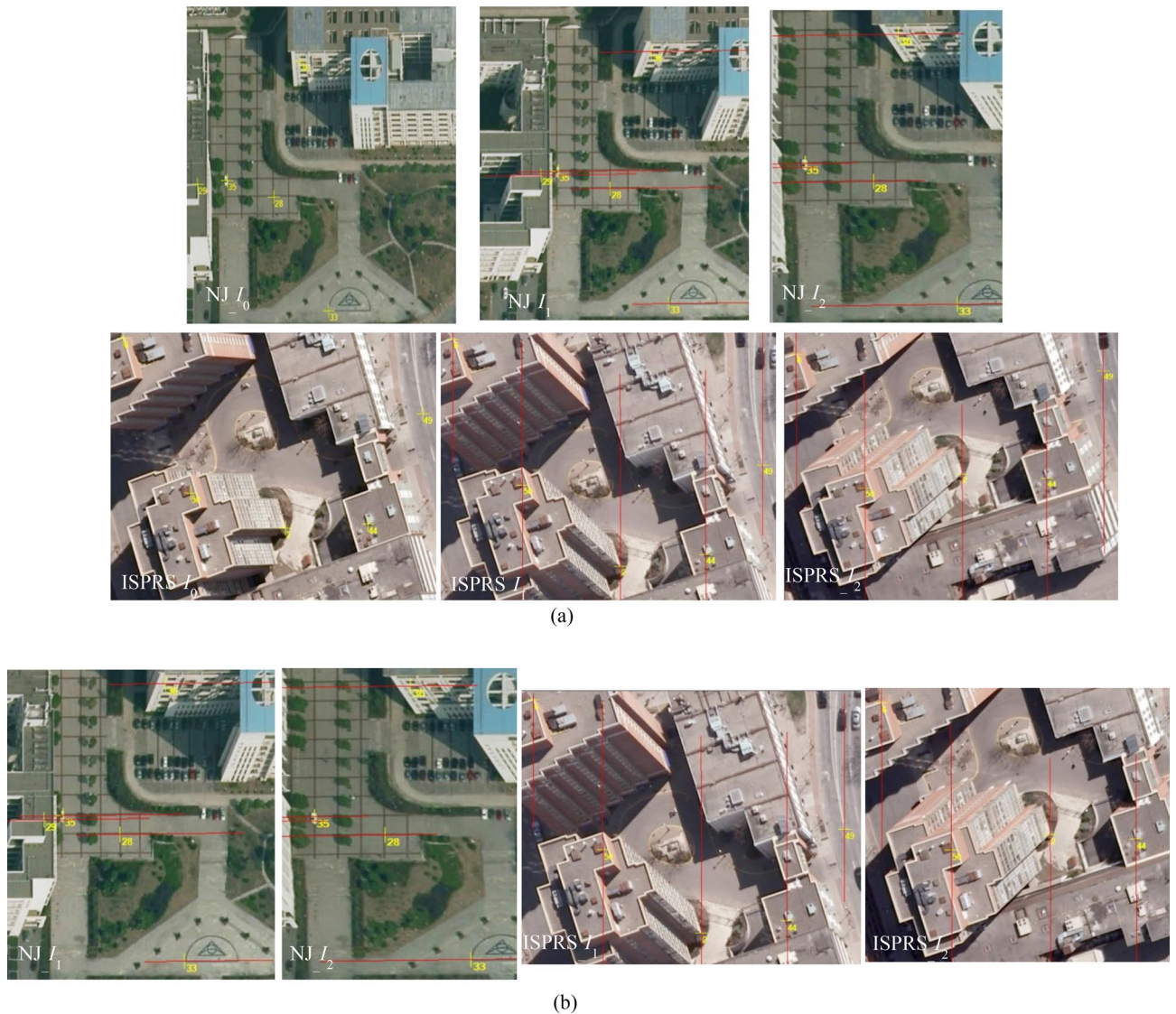
**Figure 10.** The image-space experimental results of object-side multi-view matching method based on positional distribution pattern among candidate homologous pixels for original Nanjing and ISPRS images: NJ_$I_0$-NJ_$I_1$-NJ_$I_2$ and ISPRS_$I_0$-ISPRS_$I_1$-ISPRS_$I_2$ (All displayed images in this figure were generated by our own program realized using the Visual C#,NET 2015). (**a**) The real positions of partial matching pixels in the base image $I_0$ and their homologous pixels in two searching images $I_1$ and $I_2$ for the two types of experimental data (local area). (**b**) Homologous pixels obtained by object-side multi-view image matching in two searching images for the two types of experimental data (local area)

or different matching pixels in the same image, the iterative times and the total number of candidate pixels in each iteration may be different.

Furthermore, positional distribution pattern among candidate homologous pixels, which is verified in experiment I, can be used to develop a new object-side multi-view image matching method. This new method does not need to construct 2D image-side matching windows and use the grayscale or feature information of pixels in the windows to compute the matching similarity at all, but only needs the standardized residual of each candidate homologous pixel obtained by the adjustment of pseudo object-space coordinates of the matching pixel. Seen from above experimental results in Figs. 10, 11, 12, 13 and Table 3 of experiment II, as the object-side matching method does not depend on the grayscale information of the image, its matching results are identical to original results when matching images have nonlinear variation in the grayscale (even the eyes can no longer distinguish objects in the images). As a contrast, the matching accuracy of the traditional image-side matching method will decrease greatly in this condition. The reasons for choosing image-side NCC-based matching method for comparison are as follows. Firstly, the NCC-based method can obtain better matching result than SAD or SSD-based method for images with grayscale distortion. Secondly, the experimental matching pixels and their homologous pixels are obtained by the NCC-based method interactively. If the new object-side matching
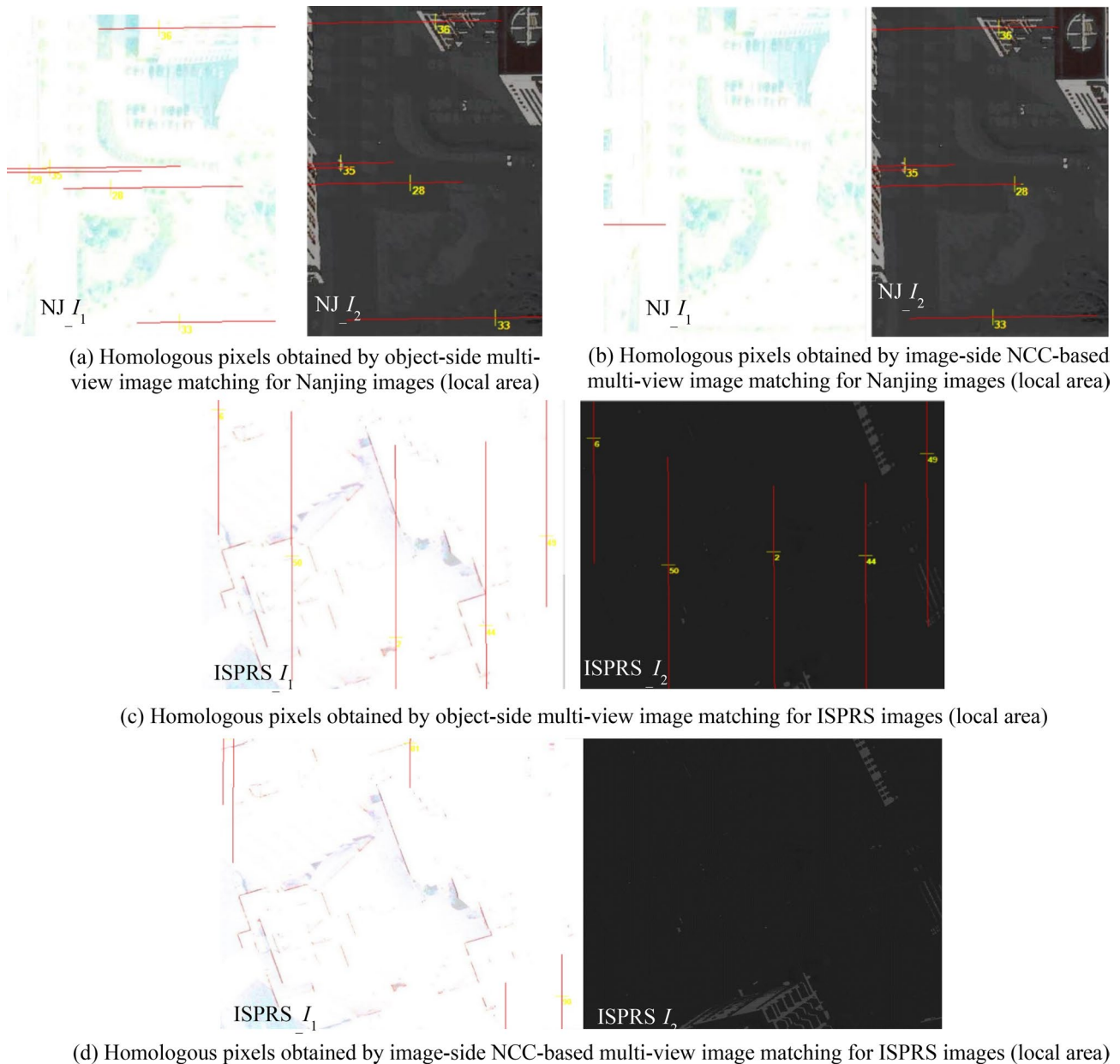
(a) Homologous pixels obtained by object-side multi-view image matching for Nanjing images (local area)

(b) Homologous pixels obtained by image-side NCC-based multi-view image matching for Nanjing images (local area)

(c) Homologous pixels obtained by object-side multi-view image matching for ISPRS images (local area)

(d) Homologous pixels obtained by image-side NCC-based multi-view image matching for ISPRS images (local area)

**Figure 11.** The comparative image-space experimental results of object-side and image-side multi-view matching methods for grayscale adjusted Nanjing and ISPRS images: $NJ\_I_0$-$NJ\_I_1$-$NJ\_I_2$ and $ISPRS\_I_0$-$ISPRS\_I_1$-$ISPRS\_I_2$ (All displayed images in this figure were generated by our own program realized using the Visual C#,NET 2015).

method is more robust than NCC-based method under the circumstance of gray-scale distortion, it can strongly prove the validity of the proposed method.

In addition, the quantitative experimental results in Table 3 show that the new object-side matching method can obtain a matching error of less than or equal to 0.5 pixel in the image space, a plane error of less than 0.4 m and an elevation error of less than 0.5 m in the object space. These object-space errors have met the plane accuracy requirement for large scale digital terrain mapping in China, as well as the elevation accuracy requirement of regions with a terrain slope greater than 2°, according to the Chinese surveying and mapping industry standard of "Digital products of fundamental geographic information 1:500, 1:1000 and 1:2000 digital line graphs" (numbered as CH/T 9008.1-2010) (In this industry standard, for 1:500 digital line graphs, depending on the types of terrain, the limitations of the plane error are 0.6 m (terrain slope < 6°) and 0.8 m (terrain slope ≥ 6°), and the limitations of the elevation error are 0.4 m (terrain slope < 2°), 0.8 m (2° ≤ terrain slope < 6°) , 1.0 m (6° ≤ terrain slope < 25°) and 1.4 m (terrain slope ≥ 25°)).

(a) Statistical curve of image-space errors and object-space errors of object-side multi-view matching results for original Nanjing images NJ_$I_0$-NJ_$I_1$-NJ_$I_2$



(b) Statistical curve of image-space errors and object-space errors of object-side multi-view matching results for grayscale adjusted Nanjing images NJ_$I_0$-NJ_$I_1$-NJ_$I_2$



(c) Statistical curve of image-space errors and object-space errors of image-side NCC-based multi-view matching results for grayscale adjusted Nanjing images NJ_$I_0$-NJ_$I_1$-NJ_$I_2$

**Figure 12.** The comparative results of statistical curve of image-space errors (left column) and object-space errors (right column) of different multi-view matching results for original and grayscale adjusted Nanjing images (All these graphs were created by Microsoft Excel 2016).

## Conclusion

Considering whether there is a specific positional distribution pattern among candidate homologous pixels of the matching pixel in multi-view images, this paper proposes a new conception of pseudo object-space coordinates and its adjustment computation model, and a new method of studying the positional distribution pattern based on the surveying adjustment theory. Actual aerial images were used to carry out experiments and verify the proposed method. Through experiments, it has been found: for candidate homologous pixels in the searching range corresponding to the matching pixel in multi-view images, using the standardized residual of each candidate pixel obtained by the adjustment of pseudo object-space coordinates as the indicator, their image-space positions appear a special V-shaped distribution pattern. Furthermore, this positional distribution pattern can be used to realize the iterative refinement of the searching range of the homologous pixel, which can reduce candidate homologous pixels from initial hundreds to final two.

In addition, based on the special positional distribution pattern among candidate homologous pixels, this paper designs a full object-side multi-view image matching process. In this process, the construction of 2D image-space matching windows and the computation of matching similarity between matching windows with image-side grayscale information are all not needed. It only uses the positional distribution pattern among
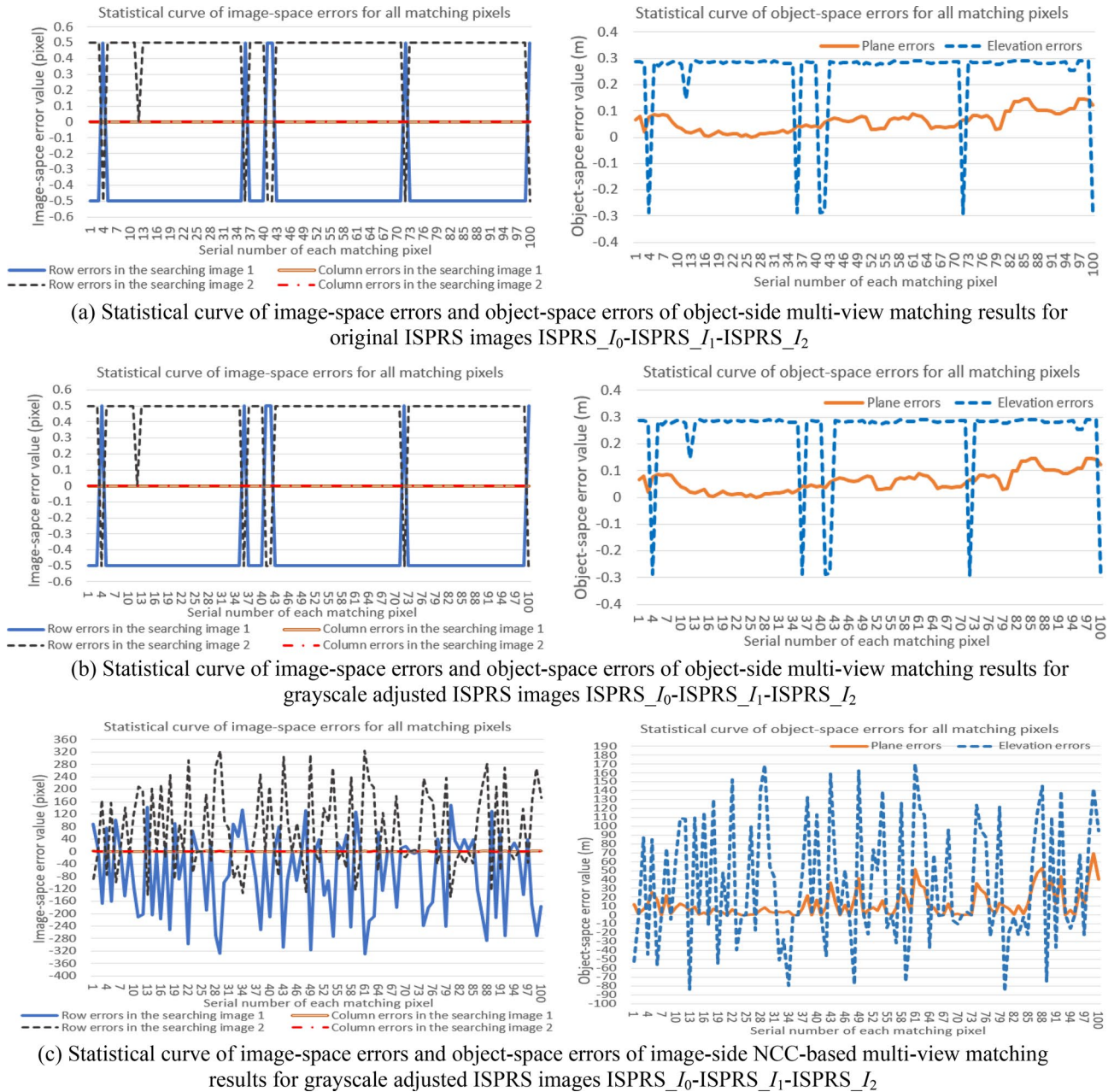
(a) Statistical curve of image-space errors and object-space errors of object-side multi-view matching results for original ISPRS images ISPRS_$I_0$-ISPRS_$I_1$-ISPRS_$I_2$



(b) Statistical curve of image-space errors and object-space errors of object-side multi-view matching results for grayscale adjusted ISPRS images ISPRS_$I_0$-ISPRS_$I_1$-ISPRS_$I_2$



(c) Statistical curve of image-space errors and object-space errors of image-side NCC-based multi-view matching results for grayscale adjusted ISPRS images ISPRS_$I_0$-ISPRS_$I_1$-ISPRS_$I_2$

**Figure 13.** The comparative results of statistical curve of image-space errors (left column) and object-space errors (right column) of different multi-view matching results for original and grayscale adjusted ISPRS images (All these graphs were created by Microsoft Excel 2016).

candidate pixels to iteratively confirm the homologous pixels in multi-view images corresponding to the matching pixel. Whereas, the image-side grayscale information is a required input for traditional region-based matching methods. Thus, the proposed object-side matching method has an obvious distinction from traditional matching methods. The experimental results verify that the object-side matching method can achieve an image-space matching accuracy of 0.5 pixel, and an object-space positioning accuracy that meets the precision requirement of large-scale digital mapping in China.

To our knowledge, this is the first report of positional distribution pattern among candidate homologous pixels in multi-view images in the field of photogrammetry, and the proposed research method on this pattern is expected to provide a new thought for resolving the problems that existing image-side multi-view image matching methods face, i.e., low reliability of matching results in difficult matching regions (such as large geometric distortion, obvious radiometric difference, weak texture or repetitive texture, etc.) at a mechanism level. Thus, the research of this paper has great significance on enriching the existing theories and methods of multi-view image matching, and has potential application value in improving the computational efficiency and accuracy of dense matching for large-scale multi-view images.

| Image type | | The total number of matching pixels | Matching method | Matching accuracy (%) | Image-space error of matching results (pixel) | | Object-space error of matching results (m) | |
|---|---|---|---|---|---|---|---|---|
| | | | | | The range of $\Delta r$ | The range of $\Delta c$ | The range of $\Delta P$ | The range of $\Delta E$ |
| Original images | $NJ\_I_0$-$NJ\_I_1$-$NJ\_I_2$ | 100 | Image-side method | Used as the ground truth | | | | |
| | | | Object-side method | 100 | [− 0.5, 0] | [− 0.5, 0.5] | (0.02, 0.4) | (− 0.5, − 0.2) |
| | $ISPRS\_I_0$-$ISPRS\_I_1$-$ISPRS\_I_2$ | | Image-side method | Used as the ground truth | | | | |
| | | | Object-side method | 100 | [− 0.5, 0.5] | [0, 0] | (0, 0.15) | (− 0.3, 0.3) |
| Searching images adjusted in gray-scale | $NJ\_I_0$-$NJ\_I_1$-$NJ\_I_2$ | 100 | Image-side method | 36 | [− 2, 3] | [− 174, 177] | [0, 86) | (− 78, 163.3) |
| | | | Object-side method | 100 | [− 0.5, 0] | [− 0.5, 0.5] | (0.02, 0.4) | (− 0.5, − 0.2) |
| | $ISPRS\_I_0$-$ISPRS\_I_1$-$ISPRS\_I_2$ | | Image-side method | 18 | [− 330, 325] | [− 1, 2] | [0, 70) | (− 86.4, 170.3) |
| | | | Object-side method | 100 | [− 0.5, 0.5] | [0, 0] | (0, 0.15) | (− 0.3, 0.3) |

**Table 3.** Quantitative comparison results of object-side and image-side multi-view matching methods in image matching for two types of experimental images.

However, the study object of this paper is still limited to central projective aerial images, and further study is needed in the following two aspects in the future. First, the proposed method should be applied to more types of multi-view images, for example, push-broom multiline space images and close-range images. Second, morphological characteristics (such as slope, length, etc.) of the standardized residual distribution curve of candidate pixels should be further studied to design a quantitative index for accurately indicating whether the real homologous pixel is located to the left side or right side of the valley point on the distribution curve. If this quantitative index can be found, a new object-side multi-view image matching method based on the positional distribution pattern of pixels, and does not depend on the image quality, can be really realized.

# References

1. Vogiatzis, G. & Hernández, C. Video-based, real time multi-view stereo. *Image Vis. Comput.* **29**, 434–441 (2011).
2. Carrasco, P. L. N., Bonin-Font, F., Campos, M. M. & Codina, G. O. Stereo-vision graph-SLAM for robust navigation of the AUV SPARUS II. *IFAC-PapersOnLine* **48–2**, 200–205 (2015).
3. Jiang, S., Jiang, W. S., Huang, W. & Yan, L. G. UAV-based oblique photogrammetry for outdoor data acquisition and offsite visual inspection of transmission line. *Remote Sens.* **9**(278), 1–25 (2017).
4. Oliveira, R. A., Tommaselli, A. M. G. & Honkavaara, E. Generating a hyperspectral digital surface model using a hyperspectral 2D frame camera. *ISPRS J. Photogramm.* **147**, 345–360 (2019).
5. Colomina, I. & Molina, P. Unmanned aerial systems for photogrammetry and remote sensing: A review. *ISPRS J. Photogramm.* **92**, 79–97 (2014).
6. Shan, J. *et al.* Toward a unified theoretical framework for photogrammetry. *Geo-spat. Inf. Sci.* **23**(1), 75–86 (2020).
7. Li, G. Y. *et al.* ZY-3 block adjustment supported by glas laser altimetry data. *Photogramm. Rec.* **31**(153), 88–107 (2016).
8. Yuan, X. X., Chen, S. Y., Yuan, W. & Cai, Y. Poor textural image tie point matching via graph theory. *ISPRS J. Photogramm.* **129**, 21–31 (2017).
9. Zhang, Y. J. *et al.* Fully automatic generation of geoinformation products with Chinese ZY-3 satellite imagery. *Photogramm. Rec.* **29**(148), 383–401 (2014).
10. Zhang, K., Sheng, Y. H. & Lv, H. Y. Extraction of road traffic sign information based on vehicle-borne mobile photogrammetric system. *Photogramm. Rec.* **30**(150), 187–210 (2015).
11. Noh, M. J. & Howat, I. M. Automatic relative RPC image model bias compensation through hierarchical image matching for improving DEM quality. *ISPRS J. Photogramm.* **136**, 120–133 (2018).
12. Cheng, L., Gong, J. Y., Li, M. C. & Liu, Y. X. 3D building model reconstruction from multi-view aerial imagery and lidar data. *Photogramm. Eng. Remote Sens.* **77**(2), 125–139 (2011).
13. Toschi, I. *et al.* Oblique photogrammetry supporting 3D urban reconstruction of complex scenarios. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.* **42–1**, 519–526 (2017).
14. Rupnik, E., Pierrot-Deseilligny, M. & Delorme, A. 3D reconstruction from multi-view VHR-satellite images in MicMac. *ISPRS J. Photogramm.* **139**, 201–211 (2018).
15. Gruen, A. Development and status of image matching in photogrammetry. *Photogramm. Rec.* **27**(137), 36–57 (2012).
16. Remondino, F. *et al.* State of the art in high density image matching. *Photogramm. Rec.* **29**(146), 144–166 (2014).
17. Hartmann, W., Havlena, M. & Schindler, K. Recent developments in large-scale tie-point matching. *ISPRS J. Photogramm.* **115**, 47–62 (2016).
18. Luo, G. F. *et al.* Fuzzy matching based on gray-scale difference for quantum images. *Int. J. Theor. Phys.* **57**(8), 2447–2460 (2018).
19. Banks, J. & Corke, P. Quantitative evaluation of matching methods and validity measures for stereo vision. *Int. J. Robot. Res.* **20**(7), 512–532 (2001).
20. Heo, Y. S., Lee, K. M. & Lee, S. U. Robust stereo matching using adaptive normalized cross-correlation. *IEEE Trans. Pattern Anal.* **33**(4), 807–828 (2011).

21. Mattoccia, S., Tombari, F. & Stefano, L. D. Reliable rejection of mismatching candidates for efficient ZNCC template matching. In *Proceedings of 2008 15th IEEE International Conference on Image Processing* 849–852 (IEEE, 2008). https://doi.org/10.1109/ICIP.2008.4711888.

22. Manap, N. A. & Soraghan, J. J. Disparity refinement based on depth image layers separation for stereo matching algorithms. *J. Telecommun. Electron. Comput. Eng.* **4**(1), 51–64 (2012).

23. Li, B. *et al.* Image matching based on two-column histogram hashing and improved RANSAC. *IEEE Geosci. Remote Sens.* **11**(8), 1433–1437 (2014).

24. Lowe, D. G. Distinctive image features from scale invariant keypoints. *Int. J. Comput. Vis.* **60**(2), 91–110 (2004).

25. Bay, H., Ess, A., Tuytelaars, T. & Gool, L. V. Speeded-up robust features (SURF). *Comput. Vis. Image Und.* **110**(3), 346–359 (2008).

26. Gueguen, L. & Pesaresi, M. Multi scale Harris corner detector based on differential morphological decomposition. *Pattern Recogn. Lett.* **30**(1), 544–557 (2009).

27. Demetz, O., Hafner, D. & Weickert, J. Morphologically invariant matching of structures with the complete rank transform. *Int. J. Comput. Vis.* **113**(3), 220–232 (2015).

28. Yang, Y. Z. *et al.* An efficient stereo matching algorithm based on four-moded Census transform for high-resolution images. *3D Res.* **9**(3), 1–12 (2018).

29. Zhang, K., Sheng, Y. H. & Lv, H. Y. Stereo matching cost computation based on nonsubsampled contourlet transform. *J. Vis. Commun. Image Res.* **26**, 275–283 (2015).

30. Ye, Y. X. *et al.* A local phase based invariant feature for remote sensing image matching. *ISPRS J. Photogramm.* **142**, 205–221 (2018).

31. Rosnell, T. & Honkavaara, E. Point cloud generation from aerial image data acquired by quadrocopter type micro unmanned aerial vehicle and a digital still camera. *Sensors* **12**(1), 453–480 (2012).

32. Yue, Q., Ma, C. W. & Cao, X. Q. A robust stereo matching algorithm based on improved epipolar distance. *Optik* **130**, 91–98 (2017).

33. Ling, X. *et al.* An image matching algorithm integrating global SRTM and image segmentation for multi-source satellite imagery. *Remote Sens.* **8**(672), 1–19 (2016).

34. Zhang, K., Sheng, Y. H., Wang, M. Z. & Fu, S. X. An enhanced multi-view vertical line locus matching algorithm of object space ground primitives based on positioning consistency for aerial and space images. *ISPRS J. Photogramm.* **139**, 241–254 (2018).

35. Laraqui, M., Saaidi, A., Mouhib, A. & Abarkan, M. Dense matching for multi-scale images by propagation. *Multimed. Tools Appl.* **77**(17), 22923–22952 (2018).

36. Sun, Y. *et al.* Line matching based on planar homography for stereo aerial images. *ISPRS J. Photogramm.* **104**, 1–17 (2015).

37. Hu, H. *et al.* Reliable spatial relationship constrained feature point matching of oblique aerial images. *Photogramm. Eng. Remote Sens.* **81**(1), 49–58 (2015).

38. Zhao, W. H., Yan, L. & Zhang, Y. S. Geometric-constrained multi-view image matching method based on semi-global optimization. *Geo-spat. Inf. Sci.* **21**(2), 115–126 (2018).

39. Scharstein, D. & Szeliski, R. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *Int. J. Comput. Vis.* **47**, 7–42 (2002).

40. Hirschmüller, H. Stereo processing by semiglobal matching and mutual information. *IEEE Trans. Pattern Anal.* **30**(2), 328–341 (2008).

41. Hu, T. *et al.* Dense corresponding pixel matching between aerial epipolar images using an RGB-belief propagation algorithm. *IEEE Geosci. Remote Sens.* **9**(6), 1036–1040 (2012).

42. Barzigar, N., Roozgard, A., Cheng, S. & Verma, P. SCoBep: dense image registration using sparse coding and belief propagation. *J. Vis. Commun. Image Res.* **24**(2), 137–147 (2013).

43. Yang, Y., Liu, Q. & Liu, H. Dense depth image synthesis via energy minimization for three-dimensional video. *Signal Process.* **112**, 199–208 (2015).

44. Taniai, T., Matsushita, Y., Sato, Y. & Naemura, T. Continuous 3D label stereo matching using local expansion moves. *IEEE Trans. Pattern Anal.* **40**(11), 2725–2739 (2018).

45. Reichstein, M. *et al.* Deep learning and process understanding for data-driven Earth system science. *Nature* **566**, 195–204 (2019).

46. Yang, M. L. & Lv, X. B. Learning both matching cost and smoothness constraint for stereo matching. *Neurocomputing* **314**(7), 234–241 (2018).

47. Kendall, A., Martirosyan, H., Dasgupta, S. & Henry, P. End-to-end learning of geometry and context for deep stereo regression. In *Proceedings of 2017 IEEE International Conference on Computer Vision* 66–75 (IEEE, 2017). https://doi.org/10.1109/ICCV.2017.17.

48. Yu, W. *et al.* Hierarchical semantic image matching using CNN feature pyramid. *Comput. Vis. Image Und.* **169**, 40–51 (2018).

49. Wolf, P. R., Dewitt, B. A. & Wilkinson, B. E. *Elements of Photogrammetry with Application in GIS* 4th edn. (McGraw-Hill Education-Europe Press, New York, 2014).

50. Wilcox, R. R. *Introduction to Robust Estimation and Hypothesis Testing* (Academic Press, Cambridge, 2011).

## Acknowledgements

## Author contributions

K.Z. and Y.H.S. deigned research; K.Z. performed research; K.Z., Y.H.S., and J.S.W. analyzed data; S.Z. and L.J.Y. aided experiments; and K.Z. and W.X. wrote the paper.

## Competing interests

The authors declare no competing interests.

## Additional information

**Correspondence** and requests for materials should be addressed to K.Z. or Y.S.

**Reprints and permissions information** is available at www.nature.com/reprints.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.