

Article

Building Façade Recognition Using Oblique Aerial Images

Xiucheng Yang, Xuebin Qin, Jun Wang, Jianhua Wang, Xin Ye and Qiming Qin *

Institute of Remote Sensing and Geographic Information System, Peking University, Beijing 100871, China; E-Mails: xiuchengyang@163.com (X.Y.); qxbsdnd@163.com (X.Q.); wangjun.cau@foxmail.com (J.W.); wjhdg@163.com (J.W.); lanlang524@126.com (X.Y.)

* Author to whom correspondence should be addressed; E-Mail: qmqin@pku.edu.cn; Tel.: +86-10-6275-4674.

Academic Editors: Devrim Akca, Richard Gloaguen and Prasad S. Thenkabail

Received: 18 March 2015 / Accepted: 7 August 2015 / Published: 18 August 2015

Abstract: This study proposes a method to recognize façades from large-scale urban scenes based on multi-level image features utilizing a recently developed oblique aerial photogrammetry technique. The method involves the use of multi-level image features, a bottom-up feature extraction procedure to produce regions of interest through monoscopic analysis, and then a coarse-to-fine feature matching strategy to characterise and match the regions in a stereoscopic model. Feature extraction from typical urban Manhattan scenes is based on line segments. Windows are re-organised based on the spatial constraints of line segments and the homogeneous structure of the spectrum. Façades as regions of interest are successfully constructed with a remarkable single edge and evidence from windows to get rid of occlusion. Feature matching is hierarchically performed beginning from distinctive facades and regularly distributed windows to the sub-pixel point primitives. The proposed strategy can effectively solve ambiguity and multi-solution problems in the complex urban scene matching process, particularly repetitive and poor-texture façades in oblique view.

Keywords: multi-level feature; coarse-to-fine strategy; spatial relationship; oblique aerial photogrammetry; façade information

1. Introduction

Information about the building façade is key in the field of building modeling, landmark recognition, navigation and scene understanding and other outdoor urban environment related applications. The

façade is always extracted based on ground data from on-board cameras [1] or terrestrial laser scanners [2]. Airborne platforms have the advantages of low cost, high efficiency, wide coverage and extensive applicability. The oblique aerial images captured by recently developed airborne oblique photogrammetry [3,4] simultaneously acquire both rooftop and façade information and thus provide numerous advantages for façade-related applications in remote sensing compared with traditional systems in the vertical imaging view.

Since the development of airborne oblique photogrammetry, the visualization and texture mapping of façade information based on aerial oblique images [5–7] has been applied extensively, but recognition-based applications have only been explored. On one hand, facade could be segmented and confirmed in a 2D image space. Lin and Nevatia [8] detected façades utilizing line segment detection and perpetual organization based on a single oblique image. Perpetual organization, which is a common approach in the detection of rectangular building roofs, appeared to be suitable but was actually unfeasible: the oblique view leading to remarkable occlusion and feature detection due to inconspicuous spectral differences between adjacent facades makes perpetual organization invalid. The edge between façade and ground surface is usually shielded and the vertical edges between the adjacent façades are hard to detect owing to spectral similarity. On the other hand, more researchers attempted to consider facade in 3D stereo vision space. Xiao *et al.* [9] extracted buildings and reconstructed a simple building model based on recognizing vertical façade planes. The 2D characteristics of remarkable linear horizontal and vertical structures and the 3D characteristics of significant changes in height in a pair of input images were used to implement the façade reconstruction. The façade was hypothetically generated from a monocular image and verified with a stereo image pair. Nyaruhuma *et al.* [10] verified building outlines in 2D cadastral datasets based on determining the spatial location of the façade. Five factors, including line match, line direction, correlation coefficient, sift match and building edge ratios, were employed to confirm the locations of the vertical façades in 3D space based on the 3D point cloud data from stereo matching. Meixner and Leberl [11] detected a façade mapped in 3D object space based on a 3D point cloud using vertical aerial photography (greater than 20°) instead of utilizing various auxiliary data. Zebedin *et al.* [12] adopted an image optimization-based method to ascertain the positions of façades in Digital Surface Model, which was employed to initialize the hypotheses and describe the 3D information of the façades.

The currently continuous imaging with large overlaps makes stereo matching based methods feasible for façade information extraction. Matching techniques can be divided into two categories: area-based matching and feature-based matching. Area-based matching is a pixel-wise dense matching technique in which the centre of a small window is matched by statically comparing windows of the same size in the reference image and target image. Feature-based matching measures the comparability of an obvious feature based on an invariance principle; that is, the extraction, description and measurement of the feature determines the matching results. Feature matching is divided into multiple levels based on the information content of the “interesting” parts of images: local low-level features (points and lines), regional mid-level features, such as regularly shaped structures, and regional high-level features that describe a specific object.

The oblique aerial photography technique produces a new data resource for façade-related issues, but it has to develop adaptive matching algorithms to address the specific characterizations of large-scale urban oblique aerial imagery, such as depth discontinuities, occlusions, shadows, low texture and

repetitive pattern. Terrain discontinuities invalidate the surface constraints. Required information is lost in areas with occlusions, and shadows can result in spectrum information confusion with normal regions. The repeated structures may produce multiple peaks, which result in high error matching probabilities. Areas with poor texture are vulnerable to mismatching or non-matching. These problems inevitably result in uncertainty and ambiguity and thus render area-based and local feature matching techniques problematic and invalid in oblique images that cover large areas.

Fortunately, the remarkable linear structure of a building façade makes both the corner points and straight lines stand out. A building façade that is based on a Cartesian coordinate system and whose prominent structures are orthogonal with one another [13] allows for the easy application of geometric constraints on the line primitives; that is, lines should be either parallel or orthogonal to one another. If the classical line features that are obtained from windows or façade edges can be sorted and re-grouped, the processing of individual line segments becomes the processing of groups of line segments; thus, more geometric information is available for disambiguation in the recognition and matching process. Meanwhile, several studies have used a compromise that involves close cooperation between monoscopic and stereoscopic analyses to first produce areas of interest in each image and then characterize and reconstruct these areas into stereo images. Noronha and Nevatia [14] constructed 3D models of rectilinear buildings from multiple aerial images. Hypotheses for rectangular roof components were generated by hierarchically grouping lines in the images and then matching them in successive stages.

In summary, our aim is to recognize and obtain the spatial information of building facades, allowing to deal with widely covered areas, which is one of the biggest challenges in using oblique aerial photogrammetry. We mine image feature information ranging from simple to complex structures. Moreover, the façade and the microstructures are effectively detected and matched step by step. We first generate façade (or other objects of interest) hypotheses and then verify them over wide areas. We mine the image feature information ranging from simple to complex structures. The façade and microstructures are effectively detected and matched step by step, and a novel three-layer approach is used for façade reconstruction using information from simple to complex image structures. Simple structures are specific local features in the image itself, such as points or edges that are directly detected by a corner- or linear-feature-detecting algorithm. Complex structures include mid- or high-level regional features, such as windows and façades that are constructed from low-level features.

The remainder of the paper is structured as follows. The proposed feature extraction and matching methodology is introduced in Section 2. Experimental performance evaluations are presented in Section 3, and Section 4 provides the conclusions of the study.

2. Methodology

The methodology is presented schematically in Figure 1. Experimental data are described in Section 2.1. The proposed approach consists of a multi-level feature extraction procedure (from low-level local features to high-level regional features), which is introduced in Section 2.2, and a coarse-to-fine hierarchical feature matching procedure (backwards from regional features to local features), which is presented in Section 2.3.

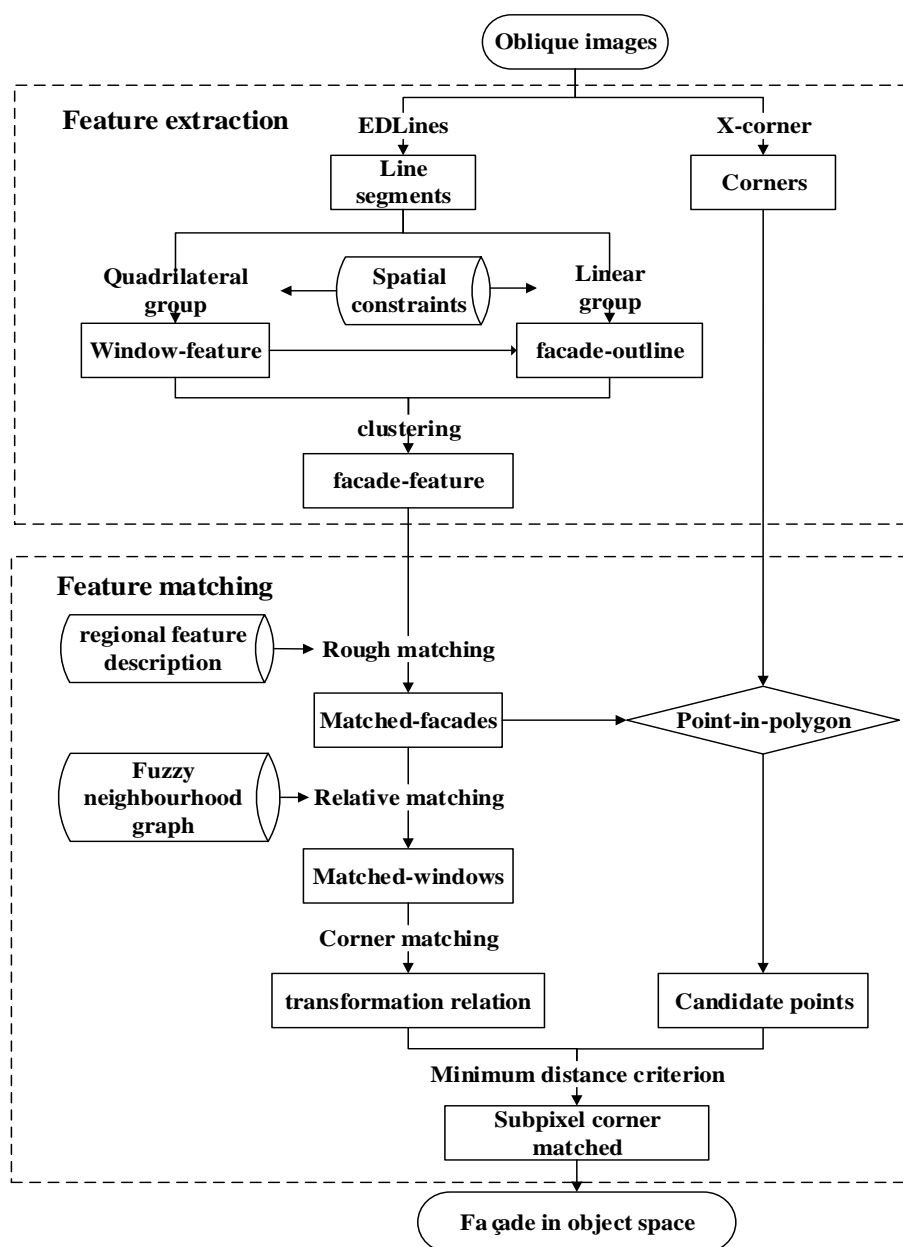


Figure 1. Workflow of façade recognition in the image space and localization in the object space.

The feature extraction was performed in three parts. First, straight-line segments were extracted as low-level features from oblique images using a fast and accurate line segment detector. Second, the discrete line segments were grouped as parallelogram window regional features (mid-level features) using spatial constraint analysis with fuzzy production rules and a “take the best” strategy. Third, the high-level features (façade regional features) were created by combining the existing window regional features, including both candidates and fakes, and an outline re-organized by a set of segments because of image noise, occlusions and deficiencies in the line extraction algorithm. Assuming that the façade is rectangular or is composed of several rectangles, the most robust edge of the façade, the border between the rooftop and the façade, was reorganized by a line linking process. The other three vertical corner lines were ignored. Thus, the spatial cluster succeeded in grouping the windows and extended the edge to the integrated outline. X-corners, and especially the angular points of windows, are widely distributed in façade scenes and serve as high-precision correspondences in the final matching process to fit the 3D façade information.

The line segments were constructed into the window regional features and the façade regional features, which decreased the location accuracy but improved the distinguishability, so the backwards coarse-to-fine strategy is feasible and applicable.

The feature matching was based on three phases in which the accuracy increases with each procedure. In the initial matching phase, we employed the regional feature description to evaluate the similarity of the façade regional features. The façades with high similarity were defined as seeds and set as reference objects. The relative location information, which was parameterized by the angle and distance eigenvectors between the non-matched and reference façades, was used to “densify” the correspondences via graph propagation. In the next matching phase, an adjacent matrix of the window network graph was weighted to authentically express the distribution of windows in the façade, and an iterative improvement was applied to acquire a rough matching result. The matched window regional feature that was formed by the original line segments has pixel-level accuracy compared to the former approximate regions. A third matching phase was implemented to improve the precision of the geometric measurements so that sub-pixel point features could be detected and matched. The initial match reduces the search range to the façade patch, and the next match obtains a set of correspondences to estimate the transform model between each façade pair. A spatial distance measure was defined to express the degree of matching between the sub-pixel points in the reference image and the points transformed from them in the search image, and the minimum distance rule with a threshold limitation was adopted to determine if a correspondence exists. Moreover, several improved strategies, such as symmetric processing and Random Sample Consensus (RANSAC) elimination, were applied.

2.1. Materials

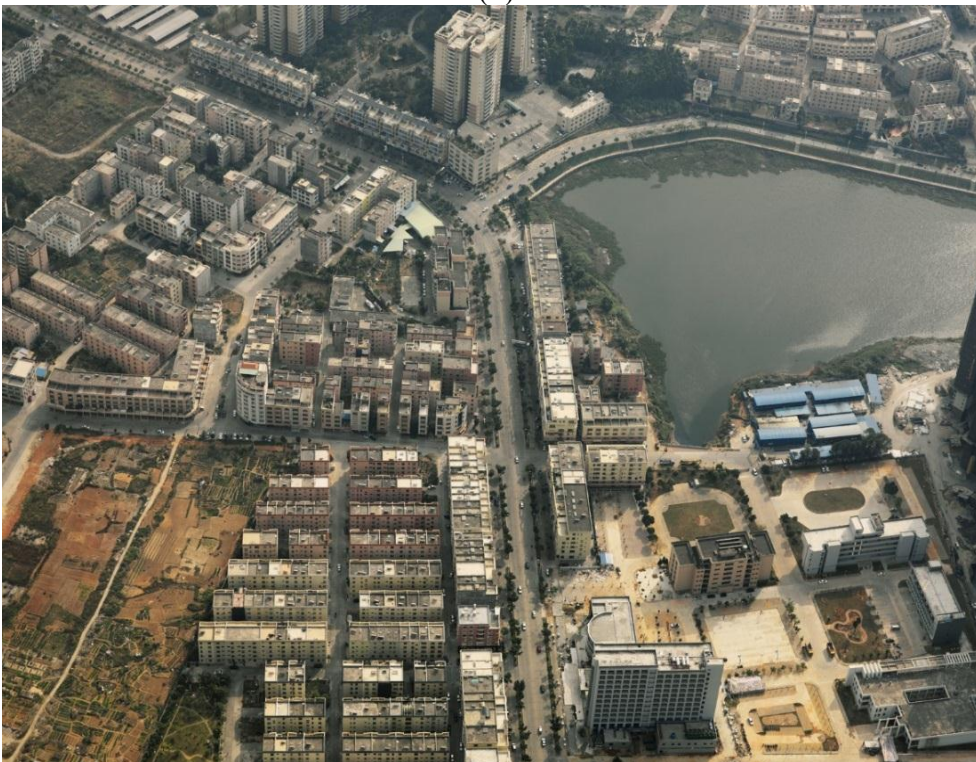
Oblique images were captured by SWDC-5 oblique aerial photogrammetry system composed of five Hasselblad H3D cameras [15] (Figure 2), one of them is vertically oriented and others are tilted with an angle of 45°. There is also a navigation and positioning system (composed of IMU and GPS) installed on the cameras, so the angle of tilt and the exterior orientation of images can be obtained when the camera is imaging. The ground resolution of the images is approximately 10 cm when the flying height is approximately 850 m.



Figure 2. SWDC-5 imaging equipment.



(a)



(b)

Figure 3. Images of the study area. The red manually drawn boxes depict the potential façades to be processed. Image (a) is used as the reference image (the study area is shown by the green box); and image (b) is used as the search image. (a) Original reference image; (b) Original search image.

The proposed method was conducted in the area in Yangjiang, China, using the successive oblique images from the camera in left orientation (Figure 3). Omitting some façades that are almost invisible because of occlusion or the view angle, 127 façades from 108 buildings should be detected and matched based on the proposed approach. Both qualitative and quantitative evaluation criteria are provided based on the manually counted number of façades and the assumption that the flat façades are vertical.

2.2. Feature Detection

An important issue in feature-based image matching is how to extract features to effectively describe the original image. This section presents a multi-level feature extracting approach. Line segments, which are the lowest level feature, were directly extracted previously. At the next level, the window regional feature was constructed using more information about the interrelationships. At the highest level, the façade regional feature was generated as the object of interest for the reconstruction.

2.2.1. Line Segment Extraction

The EDLines algorithm, which was proposed by Akinlar and Topal [16], was applied to detect the straight line segments. The algorithm is suitable for automatic high-precision matching because the linear time line segment detector requires no parameter tuning, produces continuous, clean and accurate results, and controls false linear features well. Moreover, the algorithm is more robust against scale and view variations, in which the scale in an image varies and considerable perspective distortion is present in an oblique image, than the typical Hough transformation and LSD [17].

2.2.2. Window Regional Feature Grouping

A window is usually characterised by a quadrilateral shape, which indicates certain spatial constraints for the line segments. As shown in Figure 4, due to the short lengths of the window edges, the edges are well detected in only two cases (existing and absent). In other words, the case in which an edge is divided into a series of fragments can be neglected. A particular solution may directly reorganize several straight line segments into windows. Fuzzy production rules were used to construct candidate sets, and a “take the best” strategy was used to determine the optimal set to build the discrete line segments into a quadrangle that is representative of a middle-level feature structure in complex scenes using spatial constraints, such as direction, distance and topological relations.

For each line segment, the adjacent and parallel edges were searched simultaneously and then integrated. Prior to this procedure, several pre-processes were used to reduce the dimension of the segments. The search space was diminished to segments that belong to the potential blocks with the lowest grey values segmented by the K-Means algorithm (three categories in this paper) and expanded by the dilation of morphology, which attempts to increase the efficiency and eliminate pseudo segments. The detailed implementation of the grouping is discussed in the following paragraphs, and the processing flow is shown in Figure 5.



Figure 4. Line primitives extracted with the EDlines detector.

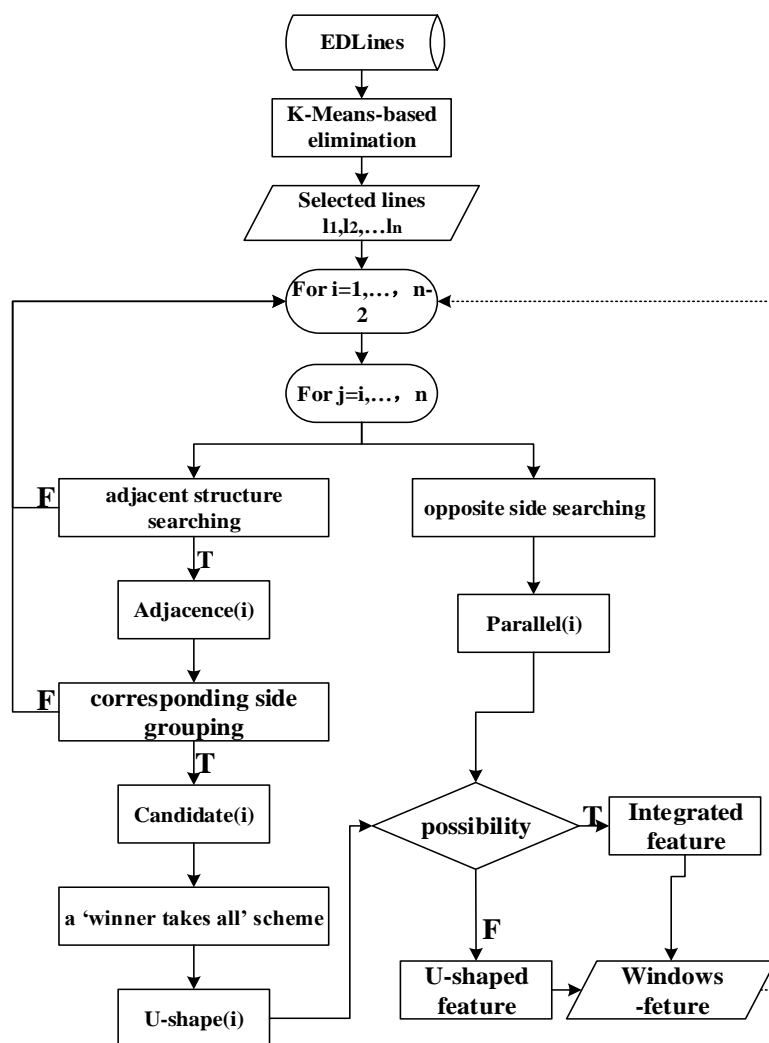


Figure 5. Flowchart of the window feature grouping process.

Searching for Candidate Adjacent Segments

For each reference line segment l_i , we may obtain several candidate segments (Figure 6), which are defined as set A_i and are judged by distance (δ_d) and direction (δ_a) constraints based on Equation (1).

Let α denote the angle between adjacent edges of the windows.

$$\text{If } \begin{cases} \text{pro}_{l_i}^{l_j} < \delta_d \\ |\text{dir}_{l_i}^{l_j} - \alpha| < \delta_a \end{cases}, \text{ then } l_j \in A_i \quad (1)$$

where $\text{pro}_{l_i}^{l_j}$ represents the minimum distance between the endpoints of l_i and l_j , $\text{dir}_{l_i}^{l_j}$ represents the angle between the lines, and α is a variable that is estimated based on the position on the image plane.

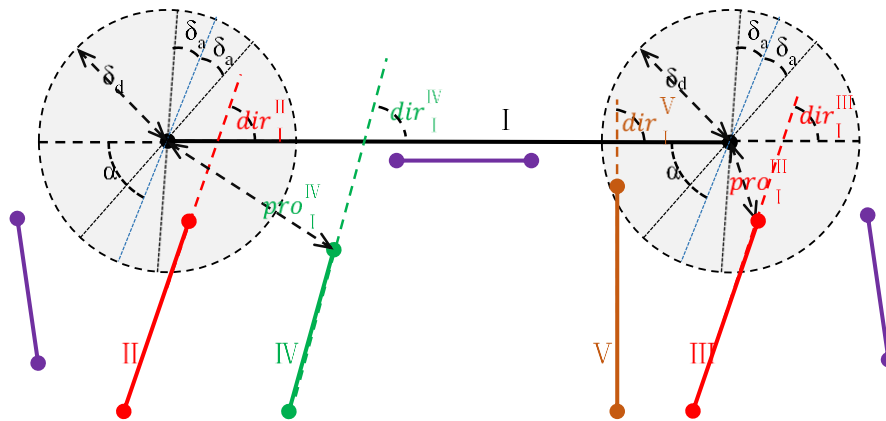


Figure 6. Finding adjacent line segments. The red line segments *II* and *III* are adjacent to line segment *I* because the end points are within the adjacent regions and the angles are within the ranges. Line segments *IV* (green) and *V* (orange) are either too distant or at too high an angle to include, and the purple segments are excluded because they do not meet both conditions. Thus, set $A_I = \{II, III\}$ is obtained.

Determining the Best U-shaped Structure

The candidate segments in A_i were screened to a set C_i consisting of certain candidate subsets via intersection judgment and direction (δ_a) constraints (Equation (2)) (the situation only containing one element is directly excluded) to constitute a U-shaped feature following Equation (2).

For arbitrary $l_a, l_b \in A_i$

Let $\text{midpt}(a)$ denote midpoint of l_a , $\text{midpt}(b)$ denote midpoint of l_b , $l_{(\text{midpt}(a), \text{midpt}(b))}$ denote the straight line joining $\text{midpt}(a)$ and $\text{midpt}(b)$.

$$\text{If } \begin{cases} \text{intersect}(l_{(\text{midpt}(a), \text{midpt}(b))}, l_i) = \text{FALSE} \\ \text{dir}(l_{(\text{midpt}(a), \text{midpt}(b))}, l_i) < \delta_a \end{cases}, \text{ then } \{l_a, l_b\} \subset C_i \quad (2)$$

where, intersection judgment ($\text{intersect}(l_{(\text{midpt}(a), \text{midpt}(b))}, l_i)$) involves guaranteeing the ipsilateral, and the direction constraint eliminates the pairs with low accuracy. Figure 7 shows the phenomena excluded in the rules.

After several subsets were selected, the best one was determined through the “winner takes all” scheme. For each pair $\{l_{N1}, l_{N2}\}$ belonging to C_i , we can evaluate the score with the following formula and select the lowest (smaller than the threshold) structure as the optimal one for constructing the ultimate U-shaped structure. In the processing, the obtained U-shaped structure is indicated as $U_i = \{l_{j1}, l_{j2}\}$.

$$w_{l_{N1}}^{l_{N2}} = |length_{l_i}/length_{l_{N1 \leftrightarrow N2}} - 1| + dir_{l_i}^{l_{N1 \leftrightarrow N2}}/(\pi/2) \quad (3)$$

where, $l_{N1 \leftrightarrow N2}$ denotes the imaginary parallel edge produced by connecting the endpoints of adjacent edges l_{N1}, l_{N2} ; $length_{l_i}$ represents the length of l_i .

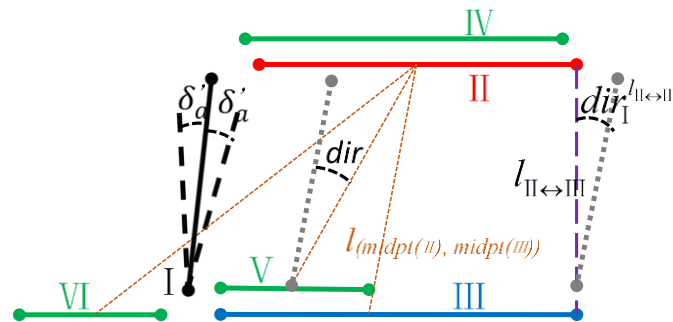


Figure 7. Producing the candidate subsets. Segment II belongs to set A_i , and the procedure is to judge whether a relevant subset that includes II exists or not. Segment IV is disregarded because it is close to the same endpoint. Segment V is excluded because of its overlarge angular deviation, and Segment V is rejected for intersection judgment. Segment III remains for weight calculation according to the imaginary edge $l_{II \leftrightarrow III}$.

Search for Candidate Parallel Edges

Based on actual cases, each reference line segment l_i should have a parallel line. Thus, parallel set P_i was generated using a distance (δ'_d) and orientation (δ_a) constraint (Equation (4)) as follows:

$$\text{If } \begin{cases} dist_{l_i}^{l_k} < \delta'_d \\ dir_{l_i}^{l_k} < \delta_a \end{cases}, \text{ then } l_k \in P_i \quad (4)$$

Generating the Windows Regional Feature

The final procedure determines whether the optimal structure exists. A probability analysis (Equation (5)) based on a pre-existing U shape for each element of set P_i was applied to select the qualified element, which was then weighted according to Equation (6). Consequently, the segment with the highest score serves as the remainder and results in an integrated window feature.

For arbitrary $l_0 \in P_i$,

If

$$\begin{cases} pro_{l_{j1}}^{l_0} < \delta_d \\ pro_{l_{j2}}^{l_0} < \delta_d \end{cases} \quad (5)$$

then

$$w_{l_o} = \frac{pro_{l_{j1}}^{l_o}}{length(line_{j1})} + \frac{pro_{l_{j2}}^{l_o}}{length(line_{j2})} + \frac{dir_{l_i}^{l_o}}{\pi/2} \quad (6)$$

However, set P_i may be empty before or after the probability analysis, in which case the window feature is constructed by directly closing the U-shaped polygon.

Parameter Setting in the Processing

In the processes described above, a set of loose thresholds should be selected; otherwise, the probability of missing correct candidates will increase. In our implementation, the thresholds were empirically set to $\delta_d = 10$, $\delta_d' = 100$ and $\delta_a = 10$. The quantity of candidate segments is more important than the quality because the “take the best” strategy significantly selects the optimal segments. A large set of feasible candidates is also helpful.

2.2.3. Selection and Linear Grouping of Façade Edges

The façade edge is usually divided into several broken line segments (Figure 4). We adopted the four organization criteria proposed by Izadi and Saeedi [18] and combined them with local spectral similarity around the line segments to restore the entire edge. We selected the wall–roof edge from the set of linked lines based on the window regional features using the following two steps.

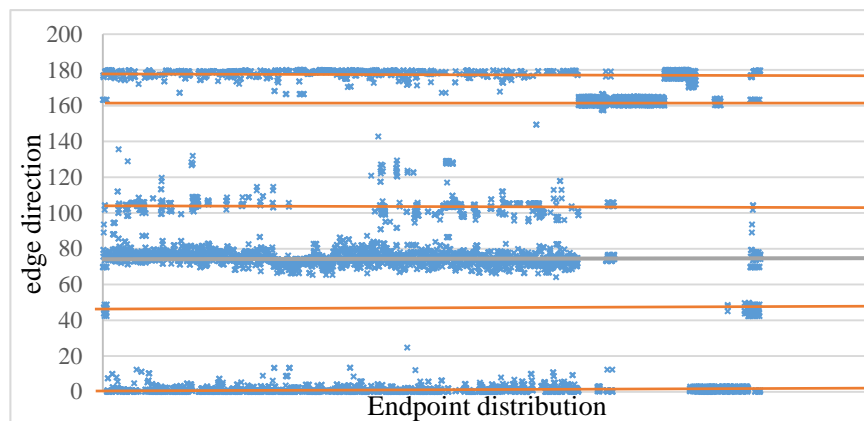


Figure 8. Directional clustering of a sample image. Six main directions are present. The value of approximately 80° is considered to represent the vertical edges. The value of approximately 100° , which may be closer to 90° , is considered to represent the horizontal edge because of the much smaller number of samples.

Judging Edge Directions through Cluster Analysis

A certain number of directions ($\theta_1, \theta_2, \theta_3, \dots, \theta_n$, where $|\theta_1 - 90| = \min|\theta - 90|$), which represent the vertical or horizontal directions in the object space, can be obtained by a directional cluster analysis of the windows (Figure 8). Considering the geometric properties of perspective projection imaging, the vertical wall edges are mapped into near-vertical directions in oblique images. The near-vertical direction, which includes most of the samples (defined as θ_1 , generally $|\theta_1 - 90| < 10$), corresponds to

vertical edges that are ignored by the process; the others correspond to horizontal edges that are retained to extract the wall–roof edges. Clustering is expected to reduce the search space and limit false extractions.

Façade Linear Feature Organization

The extraction problem is now reduced to the remaining horizontal segments. We assume that the longer the line is, the higher the probability that it is a façade edge. Based on the image geometry, the windows that belong to a certain façade in the object space remain below the wall–roof edge in the image space. The extraction traverses the lines that remain after the process given by the following rules.

- i. If no window regional feature is located between two straight lines, the shorter one is eliminated.
- ii. If no window regional feature is located below the straight line within a certain range, the line is eliminated.
- iii. If many windows are located above the straight line within a certain range, the line is eliminated.

After this process, most of the wall–roof edges can be distinguished properly, yet the fragmentation problem still exists because of large gaps that were neglected in the linear grouping. The following rules are intended to identify collinear straight lines.

- i. If two lines partially overlap, the lines are replaced by a new line that extends the longer line to fully cover the shorter one (Figure 9 (left)).
- ii. If two lines are nearly collinear but are far apart from each other and there is another line in the neighbourhood that overlaps both lines, the lines are replaced by a new line that connects the farthest endpoints (Figure 9 (right)).



Figure 9. Examples of connection processing.

At this point, the edges between the roof and the wall have been restored, although some may be mixed together. Fortunately, this effect is negligible.

2.2.4. Façade Regional Feature Construction Based on Plane Sweeping Methods

The façade regional feature, which is a hypothetical façade, was constructed by incorporating both the wall–roof edges and the window regional features. Each wall–roof edge was swept downward in the clustering direction θ_1 until it encountered another straight line or the border of the image. This procedure was used to create a prototype of a corresponding hypothetical façade that contains all of the windows that had been passed by. For each prototype, the vertical edge direction θ_1 was revised (to θ) by clustering based on the window edges that belong to the façade. The windows with large deviations were eliminated. Ultimately, the façade feature was generated by the wall–roof edges, the direction θ and the farthest window.

2.3. Feature Matching and Reconstruction

The regions of interest for the façade were produced by monoscopic analyses. Next, we characterised and reconstructed the areas. The hierarchical structure, high-level façades of interest, mid-level line set of the windows and low-level primitives of the line segments or points provided several options for feature-based matching. The distinctive façades of interest contain considerable information, but the precision is too low to be reconstructed in the object space. The windows, or the groups of interconnected line segments, are distributed in a regular pattern. However, significant comparability, repeatability and low textures are observed. Hence, distinguishing the correspondences is difficult. Sub-pixel point primitives are desirable and indispensable in man-made scenes but lack a distinctive appearance that can be distinguished by a conventional strategy in cases of repetitive patterns and low-texture environments.

A new hierarchical coarse-to-fine feature-based matching approach is thus proposed to promote the comprehensive utilisation of multi-level image features. The scheme begins with coarse matching of the façade regional features to reduce the search space into separate sub-images. Coarse matching of the window regional features is then implemented to establish transform matrixes between the façade pairs. Finally, fine matching of the sparse point features is implemented to restore the 3D spatial information of the façade.

2.3.1. Coarse Matching Based on the Façade Features

Matching the façade features is the basis of our approach to accurate and robust façade spatial locational information reconstruction in the object space. A region description method was adopted to minimize the influence of inaccurate façade edges; it consists of Euler's number (EN), the ratio of the areas of the windows to the façade (AR) and the grey histogram (H). A seed propagation solution was used in which the matched pair extends from high similarity to low similarity along the topological graph generated by the coordinates of the façade's centroid ($c(x, y)$). The feature similarities of EN and AR were calculated by $|1 - EN_1/EN_2|$ ($=E$) and $|1 - AR_1/AR_2|$ ($=R$), respectively, and that of the histogram was measured by the sum of the chi-square distribution ($d_{chi-square}$) and intersection ($d_{intersection}$) [19]. The smaller the value is, the higher the similarity is. The highly similar façades were defined as seeds, and the others were retained to determine if they match.

$$\begin{aligned} d_{chi-square}(H_1, H_2) &= 1 - \sum_i \frac{(H_1(i) - H_2(i))^2}{H_1(i) + H_2(i)} \\ d_{intersection}(H_1, H_2) &= \sum_i \min(H_1(i), H_2(i)) \end{aligned} \quad (7)$$

Generating and Matching Seeds

The seed façades, which are the relatively sparse correspondences and are defined by high similarity, are the bases of the subsequent propagating processes. Thus, the accuracy of the pre-matcher should significantly outweigh the quantity to avoid initial errors; that is, only the most reliable façades should be utilized rather than trying to match the maximum number. For one façade f_i^l in the reference image (l), the similarity $s(f_i^l, f_j^r)$ ($=E + R$) with each façade f_j^r of the target image (r) was calculated. The highest three were set as matching candidates $p_i = \{f_{i1}^r, f_{i2}^r, f_{i3}^r\}$. An element is a matched feature and defined as a seed (d) only if two conditions have been satisfied: the similarity $s(f_i^l, f_j^r)$ of the element

should be greater than the threshold α , the histogram similarity between the element and the reference façade f_i^l is the highest among the three candidates and greater than the threshold β . N matching pairs $\{\vec{d}_n\}_{n=1}^N$ exist after the traversal, and the possibility of a mismatch cannot be avoided. As a classic means to resolve the problem, the RANSAC algorithm was employed to remove the outliers by fitting the underlying fundamental matrix.

Matching Entire Façades

The seed façades were used as base stations to describe the other façade features by mutual geometric relationships that were described by the angle and distance creating eigenvector $\{\vec{\alpha}_n, \vec{\rho}_n\}_{n=1}^N$ instead of the previous descriptions, which were limited by inaccurate locations of interest such that the similarity (s_i^j and H) is actually indistinguishable and invalid. Perspective rectification based on the seeds was performed to reduce the distortion from geometric differences between views.

The “densifying” of the correspondences across the entire area is illustrated below. The newly constructed eigenvector measures the similarity along the topological neighbourhood graph that was constructed by the Delaunay triangulation algorithm according to the centroid coordinates. The propagation, instead of the traversals, is expected to utilize the relative relationship constraints to ensure the matching accuracy.

2.3.2. Coarse Matching Based on the Window Features

The primitive match generates the corresponding segmented façade; thus, the subsequent match can focus attention on interesting parts as sub-images that are independent of the background, although the matched façades cannot obtain precise geographic position information. Point primitives, particularly the rich corner features, seem to be effective and feasible. However, either the repeated structure of the windows or the imprecise elimination of the fundamental matrix causes problems in the point matching because the appearance is not distinctive, and there is no constraint to eliminate the vagueness. In contrast, the groups of line segments (the aforementioned window regional feature) naturally match such that more regional information is available for disambiguation. Although the spectral and geometric characteristics may be similar and indistinguishable and uniqueness becomes a problem, the spatial position and distribution characteristics are well determined. Therefore, the window regional features were quantified by the spatial positions relative to the façades and bridged to build a topological neighbourhood graph.

Construction of a Window Topological Graph

The *a-priori* knowledge of the grid structure indicates that the windows are generally distributed transversely and longitudinally. Therefore, the neighbourhood graph was generated by connecting the centroids in two directions that were obtained by the directional clustering. The graph is represented by the logical matrix $M_{m,n}(\zeta)$ that was inspired by the regular structure, where a binary variable $X_{i,j}(\zeta)$ indicates whether window ζ is present on the node in row i ($1, 2, \dots, m$) and column j ($1, 2, \dots, n$).

Fuzzy Weighting

Two weights were introduced to indicate the spatial position of the window in the façade while avoiding mismatching that is characterized by multi-valued mapping and ambiguity in binary patterns, especially when too few windows are present and they are scattered in space. In other words, the fuzzy adjacent matrix $\mu_{m,n}(\zeta)$ is expected to enhance the uniqueness and validation of the matching. For window ζ , we let ω_r^ζ and ω_c^ζ indicate the longitudinal and horizontal positions that correspond to the row and column in the graph, respectively. The distances from the centroid of the window ζ ($c(\zeta)$) to the façade edges δ_t^ζ , δ_l^ζ and δ_r^ζ correspond to the top, left and right edges, respectively. The average length ε of the vertical edges of the windows was used to normalize the parameters to weaken the perspective projection effect. Hence, the weights were evaluated as $\omega_r^\zeta = \delta_t^\zeta/\varepsilon$ and $\omega_c^\zeta = \delta_l^\zeta/\delta_r^\zeta$. However, ω_c^ζ will be miscalculated and result in notable mismatches if the endpoints of the vertical edges exhibit significant dislocation. Thus, to estimate the degree of dislocation, the difference μ between the matched façades was calculated by $\mu = |\text{ratio}_1(\text{length}) - \text{ratio}_2(\text{length})|$, where $\text{ratio}_1(\text{length})$ and $\text{ratio}_2(\text{length})$ are the ratios of the length of the matched wall–roof edge to the seed façades. When $\mu > \delta$ ($\delta = 0.01$ in this study), the matched façade is replaced by the closest eligible one to change the weight ω_c^ζ to ω'_c^ζ .

Iterated Matching

Given the weighted matrix $\mu_{m,n}(\zeta)$, the next objective is to determine the correspondences. The row and column were matched as separate units that are equivalent to the corresponding node between the graphs. An iterative strategy was used to eliminate mismatches and modify the initial result while avoiding the influence of viewpoint changes and position errors of the façade edges. The weights of row ($\overline{\omega_r}$) and column ($\overline{\omega_c}$) are the averages of ω_r^ζ and ω_c^ζ of all of the relevant elements, respectively. The match was initialized by the least distance rule and stopped when the iteration was accomplished.

Thus, for a matched façade π , the relevant precise sub-pixel X-corners $A_\pi = \{\alpha_1, \alpha_2, \dots, \alpha_m\}$ in the reference image can be shifted to $B_\pi = \{\beta_1, \beta_2, \dots, \beta_m\}$ in the search image according to ϕ^π , and the correspondence between the corners $\Gamma_\pi = \{\gamma_1, \gamma_2, \dots, \gamma_n\}$ of the search image is determined. For each element $\beta_i (\leftarrow \alpha_i) \in B_\pi$, the nearest γ_j is pre-defined as the undetermined correspondence, and a threshold is used to control the error; thus, a match set $\psi = \{\{\alpha_1, \gamma_1\}, \{\alpha_2, \gamma_2\}, \dots, \{\alpha_p, \gamma_p\}\}$ is obtained. Conversely, either can be regarded as the reference image, so another match set $\psi' = \{\{\alpha'_1, \gamma'_1\}, \{\alpha'_2, \gamma'_2\}, \dots, \{\alpha'_p, \gamma'_p\}\}$ can be obtained as well. Consequently, only the pairs that belong to both sets remain.

2.3.3. Fine Matching Based on X-Corner Features

The corresponding windows represent the match between the lines and the X-corners that are naturally obtained by intersections. The intersections are lower in accuracy and fewer in quantity than in direct corner detection based on image grey information by calculating the curvature and gradient, such as the Forstner, Harris and SUSAN methods. However, the matched intersections can be used as approximations for the refined corner feature to avoid mismatches in fine matching that are caused by repetitive patterns and low texture if the camera geometry is estimated in advance.

In this study, the Shi–Tomasi operator was used to extract well-defined X-corner feature points [20]. A calculation mechanism related to the peak value position was then used to acquire sub-pixel accuracy

from the pixel-level corners to meet the demands of the measurement [21,22]. Point-by-point matching between two views is generally a time-consuming task. However, the search space may be reduced by orders of magnitude in the area of the façade of interest. The approximate transformation formula ϕ^π between façades in binocular views was calculated from the initial corresponding intersections using the least squares fitting method. The RANSAC algorithm was used to remove outliers in the fitting procedure. The outliers originated from the incorrectly matched windows and the imprecise intersections. An approximate nearest neighbour search and a symmetric strategy were applied to allow fast and stable matching.

2.3.4. Spatial Localization of the Façade

POS data combined with ground control points were employed to generate 3D sparse points using a space intersection, which was shown to be reasonable in the comparative studies of Sukup *et al.* [23]. Once the 3D-structured corners are recovered in the object space, the façades, which are assumed to be planes, can easily be reconstructed by interpolation, and the locations in the object space can be determined. Two direction vectors (the normal vector \mathbf{n} and tangent vector \mathbf{t}) can be calculated to quantitatively describe the information of the measurements based on 3D linear fitting, which also evaluates the precision of the matching.

3. Experimental Results and Evaluation

The proposed façade recognition approach was implemented in C++. The results of the 2D façade detection in the image space and the 3D reconstructed information in the object space are presented and evaluated in the following sections.

3.1. Failure of Conventional Approach

In the field of feature matching and three-dimensional information fitting, SIFT-based matching algorithm and epipolar-line constraints based method is commonly accepted. Scale Invariant Feature Transform (SIFT) features are widely used to detect and describe point features in image matching. However, the areas of windows and façades are characterized as poor-texture, which make it difficult to detect SIFT features robustly. Just as shown in Figure 10, there are few corresponding SIFT features (green cross). Epipolar-line constraint is widely utilized in the point matching method, and it can reduce the image search space to epipolar line from the whole image area on the basis of the robust shi-tomasi corner operator. On one hand, a certain number of originally corresponding points are necessary to build the epipolar line. On the other hand, it is difficult to distinguish the one-to-one corresponding points because of the repetitive structure of windows (just as shown in Figure 11).

Therefore, the proposed multi-level features are developed to resolve the repetitive structure and poor texture phenomenon which is special and remarkable in façade-based large-scale urban oblique aerial imagery.

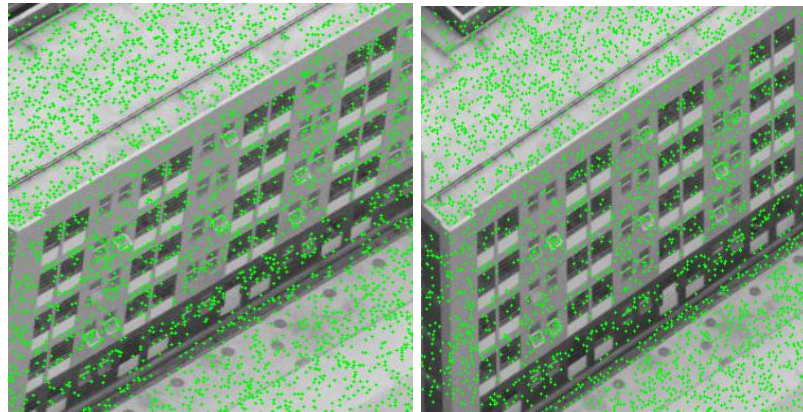


Figure 10. SIFT features detected in binocular vision images.

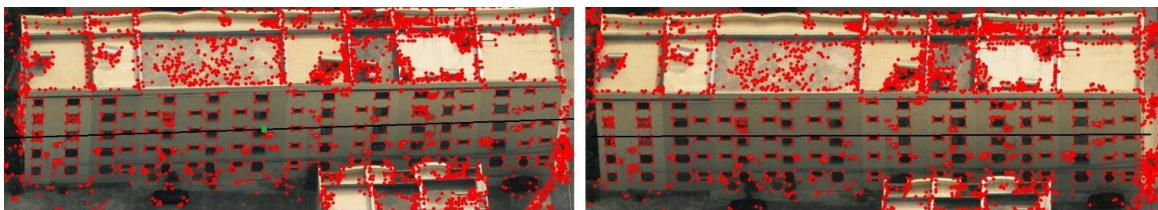


Figure 11. Epipolar-based shi-tomasi point matching. For the corner (green) in reference/left image, the search space of the corresponding point is reduced to the epipolar line (black line in the search/right image). However, there may exist several candidate matching points because of repetitive and poor-texture characterization of the façade area.

3.2. Evaluation of Regional-Feature Detection

The results of detecting façades and windows from the reference image and the search image are shown in Figure 12.



(a)

Figure 12. Cont.



(b)

Figure 12. Results of the regional feature extraction in the image space. The green boxes show the window features, and the red boxes show the façade features. The orange quadrangles are false windows that were eliminated during the processing. (a) Results of reference image; (b) Results of search image.

(1) The quadrilaterals show the constructed window regional features, and the orange boxes represent false window features that were eliminated by subsequent spatial clustering. The green boxes show the final window features (Figure 12). The window features were extracted well. The spatial distribution is reasonable; most of the visible façades that contain windows were successfully extracted using a certain number of window features (Table 1), so it is feasible to detect façades based on windows. False features are identified by the K-Means spectral clustering analysis to select candidate line segments; thus, K-Means spectral clustering improves façade detection. However, many windows are missing, mainly because of occlusion or small size (Figure 13). Specifically, self-occlusion and emerged occlusion make several windows difficult to image and thus result in feature absence. Most of the small windows are missing because the ED line algorithm operates with minimum length limit (if the image patch is approximately 3000×3000 , then the minimum length of the line segments is approximately 15 pixels). Nevertheless, the quality is more important than the quantity. The missing windows, which serve as transitional features that help to generate candidate façades and perform fine matching, can usually be ignored as long as a certain number of windows are present in one façade.

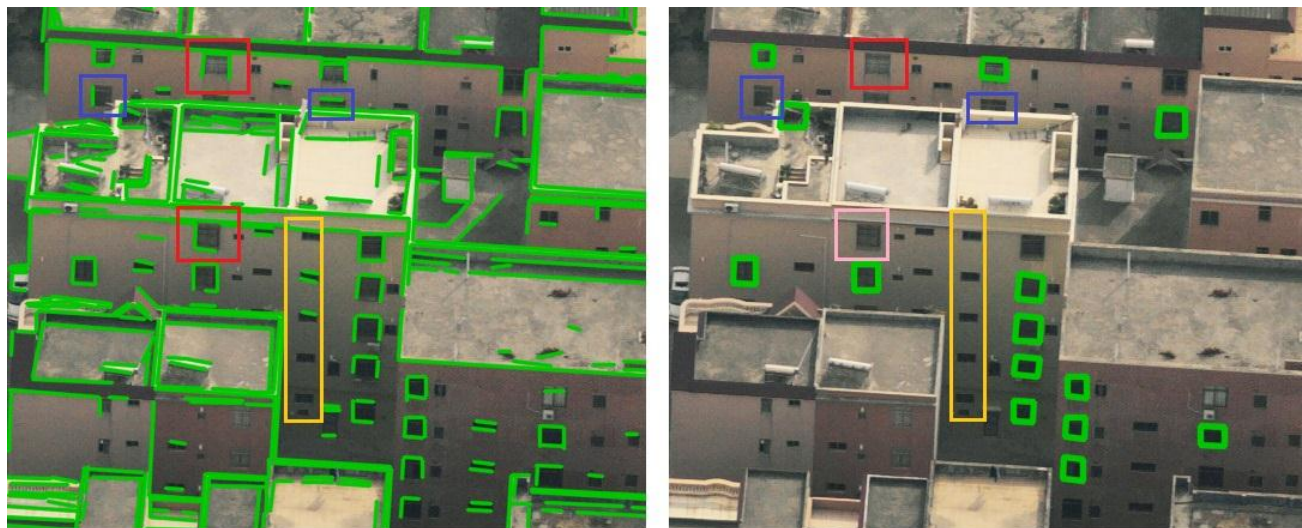


Figure 13. Failed window regional feature extraction. The windows enclosed in red boxes are examples of self-occlusion, and the windows enclosed in blue boxes are examples of emerged occlusion; both lead to feature absence. The windows enclosed in orange boxes are examples of small windows, which also lead to feature absence.

Table 1. Façade statistics of the corresponding window regional feature extraction ratio.

	>50%	>25%	>0%	Failure (No)
reference	73/70	23/22	22/21	7/14
target	57/52	23/20	37/34	10/21

Note: For the first three columns, the number before the virgule denotes the quantity of façades that meet the corresponding extraction ratio, and the number after it represents the number of façades constructed. For the last column, the number before the virgule denotes the quantity of façades for which no windows are detected, and the number after it represents the number of façades not constructed.

(2) Figure 12 shows examples in which the façade boundaries are overlaid on the input images. The detected façades are divided into four types: completely correct detections, partially correct detections, missing detections and incorrect detections. A feature is regarded as completely correct (Figure 14a) if the wall–roof edge and the vertical edges are properly located even though the bottom of the façade may not be imaged because of occlusion. However, in several cases, either inaccurate endpoint locations or an incorrect distribution of windows causes incompleteness (Figure 14b). Nevertheless, the detection is considered to be partially correct because the façade is constructed based on a poorly defined area of interest. Under some circumstances, the windows or wall–roof edge may not be extracted; this results in a missing façade (Figure 14c). Moreover, false wall–roof edges and false window regional features may meet the criteria, which leads to an incorrectly detected façade.

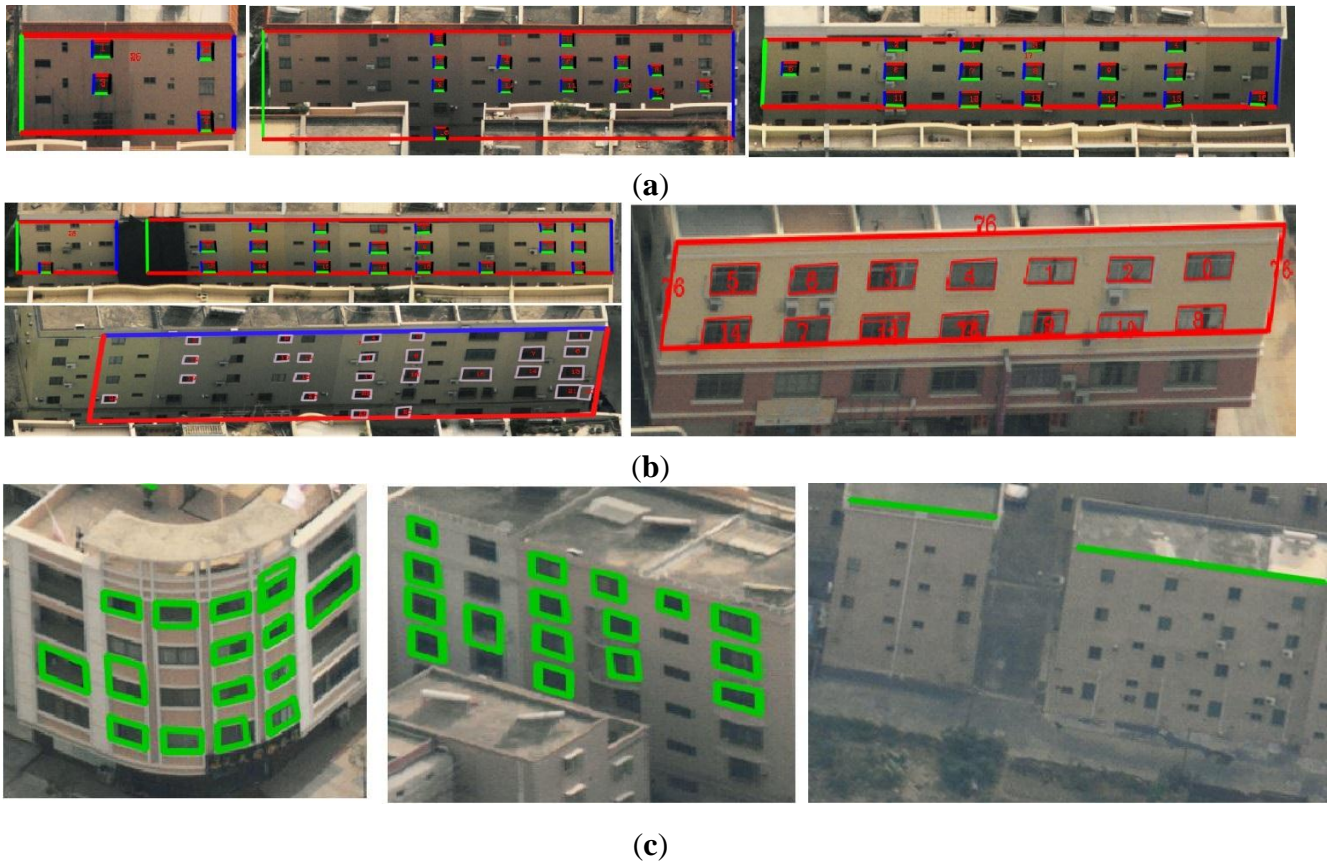


Figure 14. Three main types of extracted façades. **(a)** Classical types of completely correct detections, which are similar to reality. The left image shows the ideal case. Other objects are partially blended because of occlusion (middle), and part of the façade is not included in the regional feature because the non-obvious window features cause occlusion (right); **(b)** Several types of non-integrity. Deficiencies or splitting are observed due to non-integrity of the wall-roof edge (left), and the bottom façade is missing due to the failure to detect the window regional features (right); **(c)** Main types of missing façades. Arc-shaped façades (left) and a non-distinctive edge between the wall and the rooftop (middle) are observed. Nearly all of the window features are not extracted from a façade (right).

Table 2 shows the accuracy assessment at the object level. To quantify the façade extraction results, we adopted three frequently used metrics [24,25]. *Precision* indicates the extent to which the detected façades are at least partially real. *Recall* is a measure of the omission error. *Overall accuracy* and *F1-score* are composite metrics that consider both correctness and completeness. These metrics are computed as follows:

$$\begin{aligned}
 precision &= \frac{N_{TP}}{N_{TP} + N_{FP}} \times 100\% \\
 recall &= \frac{N_{TP}}{N_{TP} + N_{FN}} \times 100\% \\
 Overall\ Accuracy &= \frac{N_{TP}}{N_{TP} + N_{FP} + N_{FN}} \times 100\% \\
 F_1 &= \frac{2 \times precision \times recall}{precision + recall} \times 100\%
 \end{aligned} \tag{8}$$

where N_{TP} , N_{FP} and N_{FN} represent the number of true positives (*i.e.*, completely and partially correct extracted façades), false positives (*i.e.*, incorrectly extracted façades) and false negatives (*i.e.*, missing façades), respectively.

Table 2. Quantitative assessment of the proposed façade regional feature detection method.

Image	No. of TP	No. of FP	No. of FN	Precision	Recall	Overall Accuracy	F1-score
Reference	112	3	15	97.3%	88.2%	86.2%	92.5%
Search	106	4	21	96.4%	83.5%	80.9%	89.5%

The accuracy of the boundary delineation was qualitatively analysed at the pixel level. The errors mainly originate from the partially correct detections and dislocations of the wall–roof edge. The partially correct detection indicates the existence of the boundary but ignores the inaccuracy. Several line features are present along the wall–roof edge because of eaves or other linear objects (Figure 15). However, only the one with the highest weight remained, which indicates that the weighting schemes lead to different choices. Therefore, the “non-integrity” and non-uniqueness make it problematic to perform the reconstruction based on the façade features and perform façade matching based on edges.

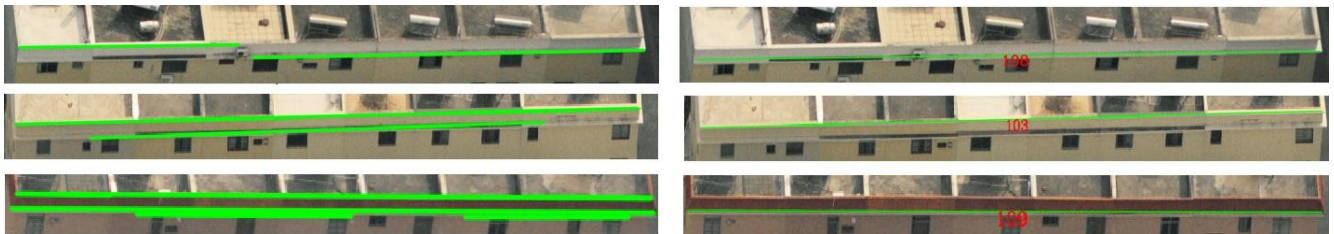


Figure 15. Non-unique locations of wall–roof edges because of the selection rules.

3.3. Evaluation of Feature Matching

After extracting the features (*e.g.*, the façade and window regional features and X-corners), we established multi-level feature matches step by step. After determining which correspondences passed on to the next stage for further processing, it is necessary to quantify the performance of the matches. A confusion matrix can be constructed to represent the dispositions of the test set, including a series of common metrics. Commonly used definitions, including counting the number of true and false matches and match failures, were adopted [26]. Then, these numbers were normalised into unit rates by following Equation (9).

TP: true positives, *i.e.*, the number of correct matches;

FN: false negatives, *i.e.*, the number of the matches not found;

FP: false positives, *i.e.*, the number of incorrect matches;

TN: true negatives, *i.e.*, the number of non-matches that were correctly rejected.

$$\text{TPR: true positive rate, } \text{TPR} = \frac{TP}{TP+FN} = \frac{TP}{P}$$

$$\text{FPR: false positive rate, } \text{FPR} = \frac{FP}{FP+TN} = \frac{FP}{N}$$

$$\text{PPV: positive predictive value, } \text{PPV} = \frac{TP}{TP+FP} = \frac{TP}{P_i}$$

$$\text{ACC: accuracy, } \text{ACC} = \frac{TP+TN}{P+N}$$

(9)

(1) The façades were matched by combining the procedures of seed façade selection and propagation. For all of the detected façades, we chose several candidate seeds (25 in the study area) with region-based similarity measures and then removed the outliers (three in the study area) using the RANSAC algorithm (Figure 16). Due to the optimal solution of the seeds' selection, only the outstanding seeds served as base stations, which were exactly matched. The others were then examined to determine if correspondence existed (Figure 17). The results of the façade regional feature matching are evaluated by the confusion matrix and common performance metrics (Table 3). The high accuracy demonstrates that the process gave an effective solution to the problematic inaccuracy of the partially correct detection. Indeed, precise matching is also the basis of effective post-processing. Moreover, although the façade matching rate is 73.2%, the actual rate reaches 81.5% when façades that belong to different buildings are considered.



Figure 16. Selection and matching of seed façades based on regional descriptions after the outliers were removed using RANSAC. The blue and white circles denote the candidate seed façades based on the regional description. The blue circles are the final seed façades, and the white circles are considered to be outliers and were removed by the RANSAC algorithm. The other nodes are the façades that have not yet been matched.

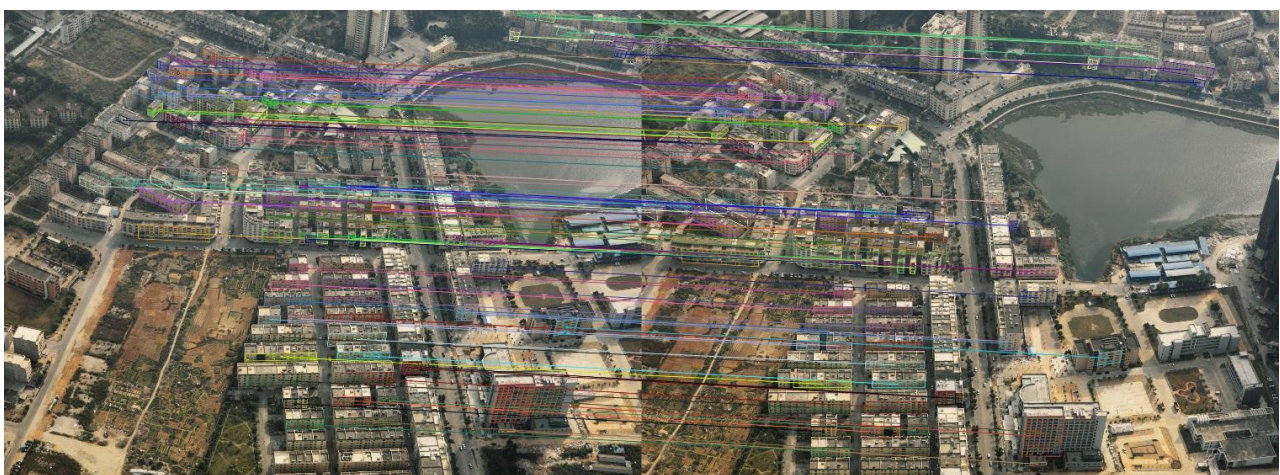


Figure 17. Results of the façade regional feature matching based on mutual geometric relationships with the seed façades. The correspondences are linked with lines.

Table 3. Confusion matrix and common performance metrics calculated from façade matching.

	Correct Matches	Correct Non-Matches	Predicted Positive/Negative	
Predicted matches	TP = 93	FP = 0	P' = 93	PPV = 100%
Predicted non-matches	FN = 3	TN = 26	N' = 29	
Actual positive/negative	P = 96	N = 26	Total = 122	
	TPR = 96.9%	FPR = 0		ACC = 97.5%

(2) In the matched façades, the windows were matched well utilizing the fuzzy topological neighbourhood graph and the iterative process. The reference and target images contained 828 and 702 windows, respectively, of which correspondences were found for 497 (Figure 18). The relevant confusion matrix and common performance metrics (Table 4) provide a quantitative evaluation of the window regional feature matching.

**Figure 18.** Corresponding windows linked by lines.**Table 4.** Confusion matrix and common performance metrics calculated from window matching.

	Correct Matches	Correct Non-Matches	Predicted Positive/Negative	
Predicted matches	TP = 431	FP = 4	P' = 435	PPV = 99.1%
Predicted non-matches	FN = 66	TN = 536	N' = 602	
Actual positive/negative	P = 497	N = 540	Total = 1037	
	TPR = 86.7%	FPR = 0.7%		ACC = 93.2%

(3) Once the windows were matched, which means that there are series of corresponding intersections, the approximate transformation formula ϕ^π between the corresponding façades can be evaluated separately using the RANSAC algorithm. Then, the refined X-corners in the sample area are matched (Figure 19), which not only improves the accuracy from the pixel-level to the sub-pixel but also increases the number of matching points by almost three-fold. Moreover, façade regions of interest reduce the search space by 2×3 orders of magnitude. Approximately, 10^5 X-corners (697,617 and 623,051) are detected in each of the perspectives, whereas only 1010^3 are contained in the corresponding façades.

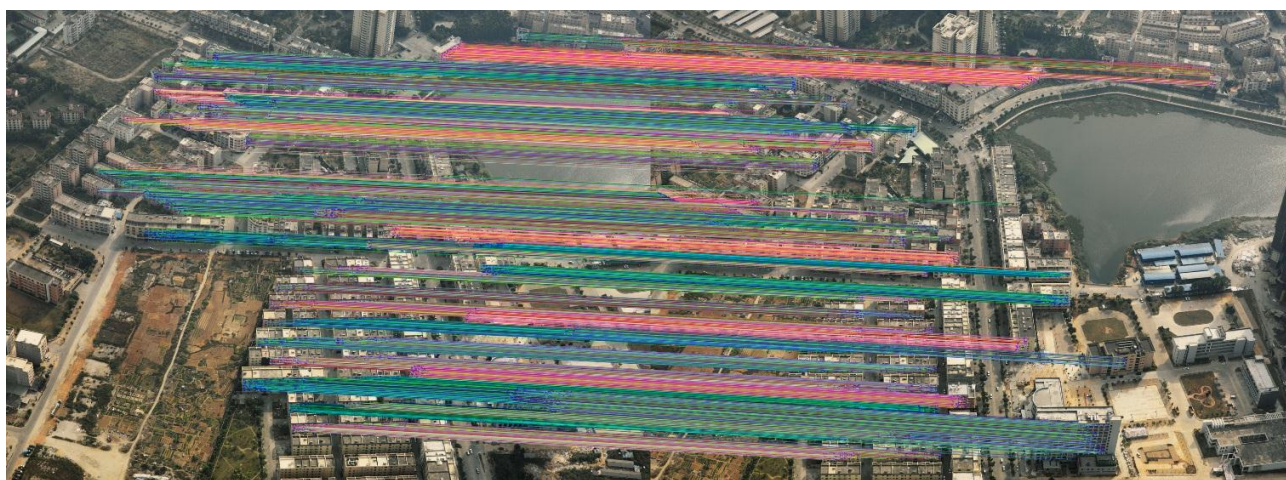


Figure 19. The final results of matching X-corners connected by lines.

Ultimately, two vectors, a normal vector \mathbf{n} and tangent vector \mathbf{t} , were fit to express the façade plane, which was also used to quantitatively verify the accuracy of the proposed hierarchical matching approach.

Table 5 compares the results from different corresponding features, sub-pixel X-corners and pixel-wise intersections. The fitted parameters are evaluated with the hypothetical standard of the vertical plane, in which the classified statistics are conducted based on the absolute difference. The fitted vectors based on the X-corners are far more precise, which justifies the reliable and necessary procedure of moving backwards from regional features to local features.

Table 5. Statistics of vectors fitted by X-corners and intersections, respectively.

		<0.001(%)	<0.01(%)	<0.1(%)	<1(%)	<2(%)
X-corner	\mathbf{n}	59.5	32.7	7.8	0	0
	\mathbf{v}	61.2	26.9	11.9	0	0
Intersection	\mathbf{n}	8.3	20.4	23.1	38.8	9.4
	\mathbf{v}	10.3	15.2	28.0	27.1	20.4

4. Conclusions

We addressed the problem of large-scale object recognition and spatial localization of urban façades from oblique aerial images. The approach can effectively segment façade areas from oblique aerial images and obtain precise 3D spatial position information.

Multi-level features were extracted using a bottom-up approach. The low-level line segments are well detected and require the construction of an object feature. The mid-level windows are obtained by organization and clustering algorithms and play a transitional role in both façade extraction and restoration. The high-level façades are constructed through a new plane sweeping procedure and serve as areas of interest. The test achieved approximately 90% of the comprehensive *F1-score*, which indicates that the proposed feature-extracting approach achieved promising results.

The newly proposed coarse-to-fine hierarchical matching approach exhibits refined sub-pixel performance based only on image information. Although most façades are characterized by an uncertain outline, we conclude that the seed propagation algorithm is fairly insensitive; the experimental area is almost completely correct (ACC = 97.5%). Transitional matching was conducted effectively (ACC = 93.2%)

on the window features to avoid ambiguity and mismatching caused by repetitive patterns and poor textures if the corresponding point primitives are searched directly. Finally, the intersections succeeded in building the transformation relationship between the façade pairs, and the sub-pixel level corner primitives alternatively fit the 3D vectors of the façade information.

The experiment regards remarkable window features as a transition connection; that is, the validity lies in the existence and extraction of windows that are part of the façade. The approach fails in cases in which there is no evidence of regular windows in the façade. Additionally, we grouped the line primitives into windows but ignored the robust points because of the lack of spatial relationship constraints. In the future, we plan to comprehensively utilize both line segments and X-corners to improve the window extraction rate, which may improve the matching effect.

The experiments were conducted in binocular stereo vision. The study area can also be obtained in more than two successive images, so a multi-view strategy may compensate for the insufficient or unstable extraction of regional features and improve the matching rate. In the case of N views, images can be defined as C_N^2 binocular tuples; thus, several features may be matched in a certain pair even though they may not appear in all of the images. This would improve the extraction and matching ratio. The images used in this study were captured in one direction. Better results may have been obtained if the images were captured in four directions (N, S, E and W).

Acknowledgments

Thanks to China Survey for providing data. This work was supported by the “National High-Tech Research and Development Program of China (2012AA121305)”.

Author Contributions

Xiucheng Yang proposed methodology to main multi-level image features to recognize building façade, designed the algorithm and performed the experiments, and finally wrote this manuscript. Xuebin Qin helped to complete the whole methodology and algorithm. Jun Wang proposed to replace conventional Line Directors with EDLines algorithm and proposed the organization method. Jianhua Wang and Xin Ye modified and improved the algorithm and experiment design. Qiming Qin organized and directed this research.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Kang, Z.; Zhang, L.; Zlatanova, S.; Li, J. An automatic mosaicking method for building façade texture mapping using a monocular close-range image sequence. *ISPRS J. Photogramm. Remote Sens.* **2010**, *65*, 282–293.
2. Truong-Hong, L.; Laefer, D.F. Octree-based, automatic building façade generation from LiDAR data. *Comput.-Aided Des.* **2014**, *53*, 46–61.

3. Jurisch, A.; Mountain, D. Evaluating the viability of Pictometry® imagery for creating models of the built environment. In *Computational Science and Its Applications–ICCSA 2008*; Springer Berlin Heidelberg: Berlin, Germany, 2008; pp. 663–677.
4. Wang, Y.; Schultz, S.; Giuffrida, F. Pictometry's proprietary airborne digital imaging system and its application in 3D city modelling. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2008**, *37*, 1065–1066.
5. Frueh, C.; Sammon, R.; Zakhor, A. Automated texture mapping of 3D city models with oblique aerial imagery. In Proceedings of the 2nd International Symposium on 3D Data Processing, Visualization and Transmission, 2004 (3DPVT 2004), Thessaloniki, Greece, 9 September 2004; IEEE: Piscataway, NJ, USA, 2004; pp. 396–403.
6. Grenzdörffer, G.J.; Guretzki, M.; Friedlander, I. Photogrammetric image acquisition and image analysis of oblique imagery. *Photogramm. Rec.* **2008**, *23*, 372–386.
7. Wang, M.; Bai, H.; Hu, F. Automatic texture acquisition for 3D model using oblique aerial images. In Proceedings of the First International Conference on Intelligent Networks and Intelligent Systems 2008 (ICINIS'08), Wuhan, China, 1–3 November 2008; pp. 495–498.
8. Lin, C.; Nevatia, R. 3-D descriptions of buildings from an oblique view aerial image. In Proceedings of the International Symposium on Computer Vision 1995, Coral Gables, FL, USA, 21–23 November 1995; pp. 377–382.
9. Xiao, J.; Gerke, M.; Vosselman, G. Building extraction from oblique airborne imagery based on robust façade detection. *ISPRS J. Photogramm. Remote Sens.* **2012**, *68*, 56–68.
10. Nyaruhuma, A.P.; Gerke, M.; Vosselman, G.; Mtalo, E.G. Verification of 2D building outlines using oblique airborne images. *ISPRS J. Photogramm. Remote Sens.* **2012**, *71*, 62–75.
11. Meixner, P.; Leberl, F. Characterizing building façades from vertical aerial images. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2010**, *38*, 98–103.
12. Zebedin, L.; Klaus, A.; Gruber, B.; Karner, K. Façade reconstruction from aerial images by multi-view plane sweeping. In Proceedings of the ISPRS Commission III—Photogrammetric Computer Vision PCV '06, Bonn, Germany, 20–22 September 2006.
13. Zhong, B.; Xu, D.; Yang, J. Vertical corner line detection on buildings in quasi-Manhattan world. In Proceedings of the 2013 20th IEEE International Conference on Image Processing (ICIP), Melbourne, VIC, Australia, 15–18 September 2013; pp. 3064–3068.
14. Noronha, S.; Nevatia, R. Detection and modeling of buildings from multiple aerial images. *IEEE Trans. Pattern Anal. Mach. Intell.* **2001**, *23*, 501–518.
15. GEO-VISION. Available online: http://www.jx4.com/Products/cp/SWDC_5shuzihangkongqingxiesheyinyi/20140411/471.html (accessed on 12 August 2015).
16. Akinlar, C.; Topal, C. EDLines: A real-time line segment detector with a false detection control. *Pattern Recognit. Lett.* **2011**, *32*, 1633–1642.
17. Zhang, Y.; Liu, Y.; Zou, Z. Comparative study of line extraction method based on repeatability. *J. Comput. Inf. Syst.* **2012**, *8*, 10097–10104.
18. Izadi, M.; Saeedi, P. Three-dimensional polygonal building model estimation from single satellite images. *IEEE Trans. Geosci. Remote Sens.* **2012**, *50*, 2254–2272.

19. Schiele, B.; Crowley, J.L. Object recognition using multidimensional receptive field histograms. In Proceedings of the 4th European Conference on Computer Vision Cambridge, UK, 15–18 April 1996; Springer Berlin Heidelberg: Berlin, Germany, 1996; pp. 610–619.
20. Shi, J.; Tomasi, C. Good features to track. In Proceedings of the 1994 IEEE Computer Society Conference on Computer Vision and Pattern Recognition 1994 (CVPR'94), Seattle, WA, USA, 21–23 June 1994; pp. 593–600.
21. Lucchese, L.; Mitra, S.K. Using saddle points for subpixel feature detection in camera calibration targets. In Proceedings of the 2002 Asia-Pacific Conference on Circuits and Systems 2002 (APCCAS'02), Denpasar, Indonesia, 28–31 October 2002; IEEE: Piscataway, NJ, USA, 2002; Volume 2, pp. 191–195.
22. Chen, D.; Zhang, G. A new sub-pixel detector for x-corners in camera calibration targets. In Proceedings of 13th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision, Plzen, Czech Republic, 31 January–4 February 2005.
23. Sukup, J.; Meixner, P.; Sukup, K. Using PixaView technology—Testing measurement accuracy in oblique photography. *GeoInformatics* **2009**, *12*, 12–14.
24. Fawcett, T. An introduction to ROC analysis. *Pattern Recognit. Lett.* **2006**, *27*, 861–874.
25. Ok, A.O. Automated detection of buildings from single VHR multispectral images using shadow information and graph cuts. *ISPRS J. Photogramm. Remote Sens.* **2013**, *86*, 21–40.
26. Szeliski, R. *Computer Vision: Algorithms and Applications*; Springer Science & Business Media: Berlin, Germany, 2010.