3D Building Model Reconstruction from Multi-view Aerial Imagery and Lidar Data

Liang Cheng, Jianya Gong, Manchun Li, and Yongxue Liu

Abstract
A novel approach by integrating multi-view aerial imagery and lidar data is proposed to reconstruct 3D building models with accurate geometric position and fine details. First, a new algorithm is introduced for determination of principal orientations of a building, thus benefiting to improve the correctness and robustness of boundary segment extraction in aerial imagery. A new dynamic selection strategy based on lidar point density analysis and K-means clustering is then proposed to identify boundary segments from non-boundary segments. Second, 3D boundary segments are determined by incorporating lidar data and the 2D segments extracted from multi-view imagery. Finally, a new strategy for 3D building model reconstruction including automatic recovery of lost boundaries and robust reconstruction of rooftop patches is introduced. The experimental results indicate that the proposed approach can provide high quality 3D models with high-correctness, high-completeness, and good geometric accuracy.

Introduction
Accurate and timely updated 3D building model data are very important for urban planning, communication, transportation, tourism, and many other applications. Aerial photogrammetry has been and still is one of the preferred ways to obtain three-dimensional information of the Earth’s surface (Brenner, 2005), and it appears to provide the economic means to acquire truly three-dimensional city data (Foerstner, 1999). Being very well understood and delivering accurate results, its major drawback is that the current state of automation in 3D building model reconstruction is still surprisingly low (Suveg and Vosselman, 2004). Airborne lidar technology has also been an important way for the derivation of 3D building models. A laser scanning system is able to provide data of directly measured three-dimensional points, which is beneficial to improving the automation level in the 3D reconstruction processes. However, the quality of the building models derived from lidar data is restricted by the ground resolution of lidar data. In general, lidar data has reached a one-meter level of spatial resolution, but it is still much lower than the aerial imagery regularly used in photogrammetry with a centimeter level of spatial resolution. It is difficult to obtain building models with highly precise geometric position and fine details only using lidar data. It is clear that photogrammetry and lidar technology are fairly complementary for 3D building reconstruction, and their integration can lead to more accurate and complete products, and a higher automation level of processes (Ackermann, 1999; Ballsavias, 1999).

Some typical research associated with 3D building reconstruction by integrating imagery and lidar data have been reviewed in the following section; however, for most of the existing approaches, it is difficult to determine highly accurate and detailed building models. Because most of the existing approaches did not strive for reconstruction of highly accurate 3D building boundaries, their common idea is to determine approximate building boundaries from lidar data, then to refine them using image information. Its main limitation is the unstable quality of the approximate boundaries determined by lidar data. The quality of these boundaries is significantly influenced by the quality of lidar data filtering processing and irregular point spacing. Meanwhile, the process of robust reconstruction of building roof patches based on lidar data is still very difficult. Therefore, a novel approach for the reconstruction of 3D building models with precise geometric position and fine details by integrating multi-view aerial imagery and lidar data is proposed. It consists of three main steps: 2D building boundary extraction, 3D building boundary determination, and 3D building model reconstruction.

Related Works
The research of 3D building reconstruction from imagery and lidar data has been increasingly focused. Schenk and Csató (2002) proposed a method on the establishment of a common reference frame between lidar data and aerial imagery by utilizing sensor-invariant features, such as breaklines and surface patches for a richer and more complete surface description. McIntosh and Krupnik (2002) merged edges detected and matched in aerial images and the digital surface model produced from airborne lidar data to facilitate the generation of an accurate surface model, which provides a better representation of surface discontinuities. Rottensteiner and Briese (2003) detected building regions and roof planes from lidar data, and then grouped these roof planes to create polyhedral building models, in which the shapes of the roof plane boundaries were improved by an overall adjustment integrating aerial imagery information. Nakagawa

Liang Cheng, Manchun Li, and Yongxue Liu are with the Department of Geographical Information Science, Nanjing University, Nanjing, Jiangsu, 210093, China (lcheng@nju.edu.cn).

Jianya Gong is with the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan, Hubei, 430079, China.

Photo: Photogrammetric Engineering & Remote Sensing
0099-1112/11/7702–0125/$3.00/0
© 2011 American Society for Photogrammetry and Remote Sensing
and Shibasaki (2003) improved the location and shape of an initial 3D model in horizontal direction by projecting the model to the nadir image of TLS (Three Linear Scanner) imagery and modified the 3D model in vertical direction according to the intersection results of the TLS imagery. Ma (2004) derived roof planes of a building using a clustering algorithm and performed the refinement of a building boundary by incorporating aerial stereo imagery. A method based on the theory of Dempster-Shafer for the detection of buildings in densely built-up urban areas was presented by the fusion of first and last pulse laser scanner data and multi-spectral images (Rottensteiner et al., 2005). Zhang et al. (2005) presented a coarse-to-fine strategy for 3D building model generation and texture mapping by the integration of three video image sequences (two oblique views of building walls and one vertical view of building roofs), a coarse two-dimensional map, and lidar data. An approach for 3D building reconstruction automatically based on aerial CCD imagery and sparse laser scanning sample points was presented by You and Zhang (2006). Sohn and Dowman (2007) collected rectilinear lines around building outlines by integrating a data-driven and model-driven manner, then accomplished a full description of building outlines by merging convex polygons, which were obtained as a building region hierarchically divided by the extracted lines using the Binary Space Partitioning (BSP) tree. Some other related approaches were introduced by (Mcintosh and Krupnik, 2002; Stamos and Allen, 2002; Seo, 2003; Frueh and Zakhor, 2004; Hu, 2007; Sampath and Shan, 2007; Sohn and Dowman, 2008; Habib et al., 2008).

Many studies have been conducted for 3D building model reconstruction only using imagery or only using lidar data. Some detailed reviews of techniques for the automatic 3D reconstruction of buildings using aerial images were made by (Remondino and El-Hakim, 2006). Some typical approaches associated with the 3D building reconstruction using lidar data were also presented (Maas and Vosselman, 1999; Sthole and Vosselman, 2004; Brenner, 2005; Vosselman and Kessels, 2005; Vestri, 2006). Meanwhile, 2D map data, including 2D GIS and urban planning maps, has been incorporated with lidar data or aerial imagery for 3D building reconstruction by many studies (Haala and Brenner, 1999; Vosselman and Dijkman, 2001).

Studies on the extraction of building boundaries are also focused in this research, because it is a crucial and difficult step in 3D building model reconstruction. The automatic extraction of a building boundary from an image has been a research issue for decades. Mohan (1989) described an approach based on perceptual grouping for detecting and describing 3D objects in complex images and applied the method to detect and describe complex buildings in aerial images. McGlone (1994) extracted building boundaries by calculating vanishing points using the Gaussian Sphere technique to detect horizontal and vertical lines. Caselles et al. (1997) introduced a scheme for the detection of object boundaries based on active contours evolving in time according to intrinsic geometric measures of the image. A detailed review of techniques for the automatic extraction of buildings from aerial images was made by Mayer (1999). In addition, many studies on boundary extraction using lidar data have been reported. To eliminate noise effects and obtain building boundaries with precise geometric positions, some researchers used the minimum description length (MDL) method to regularize the ragged building boundaries (Weidner and Forstner, 1995; Suveg and Vosselman, 2004). Zhang et al. (2006) used the Douglas-Peucker algorithm to remove noise in a footprint, and then adjusted the building footprints based on estimated dominant directions. Sampath and Shan (2007) performed building boundary tracing by modifying a convex hull formation algorithm, then implemented boundary regularization by a hierarchical least squares solution with perpendicularity constraints. Other related studies associated with boundary extraction using lidar point clouds were also introduced (Vosselman and Kessels, 2005; Brenner, 2005).

2D Building Boundary Extraction

Lidar Data Processing
In order to obtain the segmented building points from raw lidar data, the first process is usually to separate the ground points from non-ground points. Numerous algorithms have been developed to separate ground points from non-ground points. Sithole and Vosselman (2004) made a comparative study of eight filters and pointed out that the filters estimating local surfaces were found to perform best. So the linear prediction algorithm proposed by Kraus and Pfeifer (1996) is used for deriving a bare DEM from raw lidar data. Non-ground points can be identified by comparing the bare DEM and the raw lidar data. In a dataset that contains only non-ground points, the region-growing algorithm based on a plane-fitting technique proposed by Zhang et al. (2006) is used, which finally leads to the segmented building points. Figure 1b illustrates the segmented lidar points belonging to a building. It is reported that the omission and commission errors of determined buildings by the algorithm are 10 percent and 2 percent, respectively. It indicates that the segmentation algorithm worked well to identify most building measurements.

Building Image Generation
To retrieve the building features of interest from a very high-resolution aerial image, a building image is first generated to reduce the complexity of processes. The complexity of processes comes at first, followed by other vision problems from the data acquisition: the image is a projection of the real 3D object, therefore introducing deformation and multi-facess view of the object. Shadows appear more or less, depending on the sun position and height of the buildings; occlusions are frequent in very dense areas. Also, the aerial image is mostly composed of roads, buildings, etc, which have all similar types of primitives. In a building image, only one building is covered and non-building features are removed (Figure 1d). Based on the lidar points belonging to each building (Figure 1b), the building image would be generated one building by one building automatically. Since the orientation parameters of the aerial images are known, lidar points would be directly projected onto the aerial images by means of the collinearity equation shown as Figure 1c. A bounding rectangle (BR) of the building can be created based on the projected lidar points. The BR would be enlarged with a threshold to ensure all the boundaries of a building in the aerial image can be fully covered. A raster image is generated by interpolating the projected lidar data in Figure 1c, and then a buffering zone can be created. Figure 1d is the building image by cutting and filtering Figure 1a using the BR and the buffering zone, respectively. For automatic generation of the building images, based on ArcGIS® software, we use ArcObjects to implement the above algorithm. The subsequent processes will be based on the building image, and its complexity will be substantially reduced.

Line Segment Extraction

Principal Orientation Determination
Most buildings are rectangular whose boundaries have perpendicular structures with two rectilinear axes. These two perpendicular axes have the most dominant contexture and
can be treated as two major principal axes (Rau and Chen, 2003a). In this study, a new algorithm is proposed for determining two principal orientations of a building. The principal orientations can be accurately and robustly determined by a hierarchical determination of building orientations, which consists of two steps: (a) rough principal orientation estimation, and (b) precise principal orientation determination.

Based on the segmented building points (projected on the aerial images, shown as Figure 1c), rough principal orientations of a building can be estimated by analyzing the lidar points belonging to the building. For this purpose, an algorithm is used for determining two major directions of a building, which includes two following steps. (a) according to the segmented building lidar points, the edge points are detected by a region-grow method, and (b) based on the edge points, a least-square method proposed by Chaudhuri and Samal (2007) is used to calculate the two major directions of a building. Considering the various geometric shapes of the buildings, a value range of rough principal orientation is constructed by a threshold (e.g., 5 degrees) for the following precise principal orientation determination.

The precise principal directions are determined by finding maximum values in the accumulative array from a Hough Transformation that falls within estimated ranges of rough principal directions. The Hough Transform can transfer an object from image space to parameter space. The principal directions of a building in an image can be detected by analyzing the accumulative array from a Hough transformation. In general, there are only two principal directions for a building. Based on the rough principal orientation constraints, principal orientation determination consists of the following steps (Figure 2 shows the flow chart of the proposed algorithm):

Step A: selecting a building image;
Step B: applying the Hough transformation on the image;
Step C: finding the maximum value \( M \) in the whole accumulative array;
Step D: setting a threshold \( T = M \cdot t \) (the initial value of \( t \) is set to 0.9), and keeping those cells in the accumulative array with a value greater than the threshold \( T \);
Step E: selecting the range of one rough principal direction;
Step F: adding the counting numbers with the same \( \theta \) value in the range. If all the accumulative values equal 0, then decreasing the value of \( t \) and going back to step c, if \( t \) is greater than 0; if \( t \) equals 0, the whole processing has failed. If some accumulative values are greater than 0, go to next step;
Step G: selecting the range of the other rough principal direction and go to Step E. If both rough principal directions are processed, go to Step H;
Step H: detecting a peak in each of the two ranges. Each peak refers to a principal orientation of a building.

![Figure 1. A building image generation: (a) An aerial image, (b) The segmented lidar points, (c) The projected lidar points and BR (white), and (d) The building image.](image-url)
Considering the influence of the image quality, perspective distortions, and other possible errors, there may exist small angular discrepancies between parallel lines. So a value range for a principal orientation is constructed by a threshold (e.g., 2 degrees) for the subsequent extraction of line segments.

**Line Segment Extraction**

The Edison detector (Meer and Georgescu, 2001) is used as the edge detector in this study. The line segments are then extracted using Hough Transformation with the principal orientations priority. We use the Hough Transform because it is a sound method for detecting straight lines in digital images. Also we improve the Hough Transformation by setting the searching priority for peak detection according to the principal orientations. The Adaptive Cluster Detection (ACD) algorithm (Risse, 1989) is a classical dynamic method to detect straight lines based on the Hough Transformation. A modified ACD algorithm is briefly described as follows: (a) search for the location of the maximum accumulation, (b) determine the candidate points that contribute to the peak, (c) analyze the continuity for those candidate points to construct a line segment and to determine its two terminal positions, (d) remove those candidate points’ contribution from the parameter space, and (e) proceed until the value in accumulative arrays are smaller than the threshold. Peak detection in the accumulative array is first performed on the principal orientations, which would ensure that a few line segments with weak signals on the principal orientations (possible boundary segments) can be extracted effectively. This strategy avoids missing as many boundary details as possible, which is very significant for determining the complete final models.

**Boundary Segment Selection**

In this section, the extracted line segments will be automatically separated into two sets: boundary segments and non-boundary segments. A dynamic strategy based on lidar point density analysis and K-means clustering for accurate boundary selection is introduced here. Plate 1a is a building image; the extracted line segments are shown in Plate 1b. Based on the extracted line segments, the boundary segment selection is performed as follows.

**Boundary Candidate Selection**

Two rectangle boxes with a certain width are generated along two orthogonal directions of a boundary segment which is invoked by the method proposed by Sohn and Dowman (2007). Plate 1c shows two kinds of rectangular boxes (one is pink; the other is cyan) created for each segment. If no lidar points can be found in either box, the line segment is removed because the line segment is far from a building. If lidar points are found in both boxes and the density values of the two boxes are almost equal, the line segment is removed because the line segment surrounded by lidar points should locate on the rooftops. A value range (1 to 10 times the lidar point spacing) is applied to determine the optimal width of a rectangle and we suggested that 3 to 5 times can be selected as the threshold. The use of too small values may cause that boundary segments to be regarded as non-boundary segments; the reason is that a too small box may contain nothing (no lidar points) even if this box lies inside of a building due to the irregular spatial distribution of lidar points. On the other hand, the use of too large thresholds may cause many non-boundary segments to be regarded as building segments. Because the density values of the two boxes with too large a size may be almost equal, they have a different number of the lidar points. The remained line segments are considered as boundary candidate segments, as shown in Plate 1d.

**Accurate Boundary Selection**

The remaining line segments are grouped according to their angles and distances. For example, Plate 1e shows that three candidate boundary segments are organized as one group. For the group, two rectangular boxes are also generated to each segment. Focusing on the left segment, the lidar point density of one box can be calculated by Equation 1; the difference of lidar point density of two boxes can be calculated by Equation 2. In Plate 1e, the differences of lidar point density for the three segments (left, middle, right) are calculated, namely $d_1,d_2,d_3$, respectively.

\[
\text{Lidar Point Density (LPD)} = \frac{\text{the number of Lidar points in the box}}{\text{area of the box}}
\]  

\[
\text{Difference of Lidar Point Density} = \text{abs(LPD}_{\text{Left Box}} - \text{LPD}_{\text{Right Box}}) \tag{2}
\]

Adding $d_1,d_2,d_3$ ($d_1 = 1.11,d_2 = 3.65,d_3 = 0.72$) into a dataset, the dataset is defined by Equation 3:

\[
L = \{[d_k]|k = 0,...,m\} \tag{3}
\]

where $d_k$ refers to the difference of lidar point density corresponding to the segment $k$. The K-means clustering algorithm (with $K = 2$) is applied to divide the dataset into two new datasets: a dataset with large values, and a dataset with small values. So two new datasets are created: one dataset with $d_1 = 1.11,d_3 = 0.72$, the other dataset with $d_2 = 3.65$. Generally, the difference of lidar point density corresponding to an accurate boundary is larger than that
Plate 1. Boundary segment selection: (a) A building image, (b) The extracted line segments (red), (c) Two rectangle boxes (pink and cyan) for each segment, (d) The remained line segments after boundary candidate selection (red), (e) Three candidate segments organized as one group extracted from a small local region (the yellow box in (d)), and (f) The extracted segment after accurate boundary selection.
corresponding to an inaccurate boundary. So the boundary corresponding to the dataset with large values (d_s) is the one with more precise position in the group. The accurate boundary segment is selected, illustrated in Plate 1f. In some cases, if the dataset contains several large values or if the K-means clustering cannot produce true results, all of segments in one group will be selected as the accurate segments.

3D Building Boundary Determination

After the boundaries are extracted, the line-based photogrammetric intersection is often used for 3D building boundary determination. The main disadvantage of the method is that the accuracy of the determined 3D boundaries would be largely influenced by the intersection angle of the two interpretation planes and is lower than the accuracy of lidar data in the vertical direction.

A method for 3D boundary determination is introduced, which contributes to advance the geometric accuracy of the 3D boundary segments. To overlap the extracted boundary segments and lidar data, lidar points are directly projected onto the aerial images by means of the collinearity equation. Focusing on one endpoint of a boundary segment, we collect the lidar points around it and detect the edge points (the points on the building edges), and then find the nearest one among these edge points. The height value Z of the endpoint is assigned by the height value of the nearest edge point. For each endpoint, a ray is generated starting from the exposure station of the camera and toward itself. The intersection between the ray and the horizontal plane at the height value Z defines the corresponding 3D point in object space of the endpoint. The detailed calculation method is defined as Equation 4. Figure 3a shows a building covered by four aerial images. Based on the method above, each aerial image can provide a dataset of 3D boundaries. Different datasets are illustrated by different gray levels and line styles in Figure 3b.

\[
\begin{bmatrix}
  x - x_0 \\
  y - y_0 \\
  -f
\end{bmatrix} = \lambda M
\begin{bmatrix}
  x - x_L \\
  y - y_L \\
  z - z_L
\end{bmatrix}
\]

where \( \lambda \) is the projection coefficient, \( M = M_x M_y M_z \).

\[
\begin{bmatrix}
  u \\
  v \\
  w
\end{bmatrix} = M^T
\begin{bmatrix}
  x - x_0 \\
  y - y_0 \\
  -f
\end{bmatrix}
\]

Rearranging, with known Z

\[
X = X_L + (Z - Z_L) \frac{u}{w}
\]

\[
Y = Y_L + (Z - Z_L) \frac{v}{w}
\]

Specifically, this formula describes the perspective transformation between the 3D point \((X, Y, Z)\) in the object space and the 2D image pixel \((x, y)\) in the image space using explicit interior and exterior orientation parameters. The interior orientation parameters include principal point offsets \((x_0, y_0)\) and focal length \(f\); the exterior orientation parameters consist of perspective center \((X_L, Y_L, Z_L)\) and rotation angels \(\omega, \varphi, \kappa\). The rotation matrix \(M_t\) is determined by the three rotations angles.

3D boundary segment refinement becomes possible thanks to multi-view images. The refinement processes include the following three steps.

Step 1: Grouping the Segments: All the 3D segments are projected to the 2D X-Y plane. On the 2D plane, two segments are regarded as parallel if the angle between them is less than a threshold (e.g., 4 degrees). All segments are grouped according to their angles and distances. A value range (from 1 degree to 10 degrees) is used to determine the angle threshold and we suggest that 4 degrees can be selected as the optimal threshold. The use of too small values may cause that many parallel segments will be regarded as unparallel segments. Meanwhile, the use of too large values may cause that many unparallel segments will be regarded as parallel segments.

Figure 3. 3D boundary determination: (a) A building covered by four aerial images, (b) Four sets of 3D boundary segments shown by different gray levels and line styles, and (c) The refined results of 3D boundary segments.
Step 2: Overlapping the Segments and Lidar Data on the 2D Plane: In one group, the segments near to the boundary are selected using the boundary segment selection algorithm above.

Step 3: Regarding One Group, All Endpoints of the Remaining Segments are Used to Construct a New Segment: In those endpoints, there may exist inliers and outliers. In general, a least squares method may produce a line with a bad fit to the inliers because it will fit both inliers and outliers. RANdom SAMpling Consensus (RANSAC) (Fischler and Bolles, 1981), on the other hand, can produce a line which is only computed from the inliers. Therefore, RANSAC is here selected to construct the new segment, which can in general lead to the more robust results in comparison to a least squares method. Figure 3c shows the refinement results from Figure 3b. The use of multi-view imagery improves the geometric accuracy of the derived 3D boundaries.

3D Building Model Reconstruction

For 3D rooftop line determination, the region-growing algorithm based on a plane-fitting technique proposed by Zhang et al. (2006) is used to detect 3D rooftop patches. We make an improvement on this algorithm. In the original algorithm, a plane is constructed based on the points in the category using a least squares fit. The improvement is to use RANSAC algorithm instead of least squares method to construct a plane. Regarding the segmented lidar points, the region-growing processes are repeated to construct all rooftop planes of a building. 3D rooftop line segments are determined based on intersection of the 3D rooftop planes. The processes of region-growing segmentation rely on the selection of seed points. Different strategies on seed-point selection can sometimes result in differences in areas of individual patches with a building footprint. The lidar measurement density also has a big effect on detection results, and it is still difficult to fully automatically extract some complex rooftop patches. Therefore, manual editing would be still needed, but only on rooftop line determination, not on boundary determination. In many cases, 20 to 30 percent rooftop lines needed to be manually edited.

In this section, an improved method is proposed for 3D building model reconstruction from the determined 3D boundary segments and 3D rooftop segments. To reconstruct 3D models using these 3D segments, the major problem is that the topology relationships of the segments are lost. The SMS (Split-Merge-Shape) method proposed by Rau and Chen (2003b) is an effective solution to automatically recover the topology relationships between these 3D segments. An improved SMS method based on lidar data is proposed here, including two main improvements: automatic recovery of lost boundary segments, and robust reconstruction of 3D roof patches.

Lost Boundary Recovery

The overall processes include six steps: data pre-processing, splitting a building, detecting lost boundaries, locating lost boundaries, reconstructing lost boundaries, and merging into a building.

Step 1: Data Preprocessing. Preprocessing is the same as the SMS method, including collinear processing, orthogonal processing, dangle removal, and dangle snapping.
polygons with large lidar density (Label 2), and polygons with middle lidar density (Label 3). If there are no lost boundary segments, only two types of polygons will exist: the one with large density and the one with small density. So, we can identify that there exist few lost boundary segments in the polygon of Label 3.

Step 4: Locating Lost Boundaries: In the polygon (Label 3), many rectangle boxes with a certain width (about two times lidar point spacing) can be generated along the direction of one side, as shown in Figure 5a. The lidar point density of these boxes can be calculated. The same processes are conducted along the perpendicular direction, as Figure 5b. The position where the lidar point density changes dramatically is where the lost boundary should be located, as shown in Figure 5c.

Step 5: Reconstructing Lost Boundaries: After the edge lidar points (bold points in Figure 5d) are detected, the RANSAC algorithm is used to construct a new segment, which is the lost segment.

Step 6: Merging into a Building: After the lost boundary segments are reconstructed, there are only two types of polygons: polygons with large lidar density and the others with small lidar density. The K-means clustering algorithm is employed again to divide a dataset of lidar point density into two sets. All polygons with larger density are merged into a new large polygon. The outline of the new polygon is the complete building boundary, shown in Figure 4b.

Robust Rooftop Reconstruction

A complete boundary (Figure 6b) is derived from the discontinuous boundary segments (Figure 6a) by using the lost boundary recovery method above. Figure 6c consists of the recovered complete building boundary and 3D rooftop lines. The topology between all these boundaries and rooftop lines can be recovered using the above “splitting-merging” processes, which would generate many closed rooftop polygons shown in Figure 6d. Focusing on a 2D rooftop polygon, the corresponding 3D patch can be reconstructed by collecting and analyzing lidar points belonging to it. As we know, there often exist many structures, such as water towers, funnels, and pipelines on the top of buildings. In general, the lidar points on these top structures would not be removed in the filtering processes, shown as Figure 6e. Due to these lidar points, the 3D
rooftop patches reconstructed by a least squares method would have a low geometric accuracy in the vertical direction. When reconstructing 3D models using the RANSAC algorithm, these lidar data on the top of structures can be automatically removed, which will bring 3D models with higher accuracy, especially in vertical direction. Figure 6f shows the reconstructed 3D model.

**Evaluation**

**Data Set**

An experimental area covered by aerial images and lidar data is used to test the performance and applicability of the approach, shown in Figure 7. It covers an area of 1,500 m by 540 m. The lidar data have average point spacing of 1.1 m, with a horizontal accuracy of 30 cm and a vertical accuracy of 15 cm. The lidar points with different height values are shown as a different shade of gray. Each aerial image is 3,056 pixels by 2,032 pixels with a spatial resolution of 5 cm, and shown by a rectangle in Figure 7. The orientation parameters of the aerial images are known. The experimental area contains many buildings with different orientations, different structures, and different texture conditions of rooftop. Most of buildings have complex geometric shapes. Image texture conditions are also different, including simple texture, highly repeated texture, and complex texture. The complex texture conditions are formed because the trees are close to the buildings.

In addition, there are existing true orthophotos and 3D building models constructed manually in the experimental area, which would be taken as reference data to evaluate the quality of the 3D building models reconstructed by the proposed approach. The true orthophotos have a horizontal accuracy of 50 cm. Terrasolid software, a software package for processing lidar and images, was used to construct 3D building models. TerraModeler, a tool of Terrasolid software, is mainly used for 3D building model reconstruction. Orthophotos as a reference are required to detect 3D building shapes, breaklines, and other features for 3D building model reconstruction.

**Overlay of Imagery and Lidar Data**

The aerial images and lidar data were directly overlaid by means of the collinearity equation. Seventeen distinct lidar points (e.g., corner points of a building) that are distinguishable with its neighbors are selected to test the accuracy of data overlay. The offset values between the projected lidar points and the corners in the aerial images are checked. Root Mean Square Error (RMSE) of 3.1 pixels and 3.0 pixels on the X-axis and Y-axis, respectively, are achieved. The accuracy of data registration was satisfactory in this study, and the subsequent processes can continue.

**Visual Appearance**

Figure 8 illustrates the 3D building models reconstructed by the proposed approach. To visually observe the quality of the 3D building models, we compared the models with the reference data including 3D models constructed by
manual operation, true orthophotos, and aerial images. Figure 9 illustrates four buildings’ detailed 3D models derived by the proposed approach and by manual operation, respectively. Three pairs of circles in Figure 9 demonstrate the detailed differences between the two kinds of models (e.g., the missed corners and the missed small structures). The results of observation demonstrate that the reconstructed 3D models provide a reliable appearance, and most of them have a high coincidence with the reference data, which means the proposed approach can lead to accurate and reliable 3D building models. In subsequent analysis, the quantitative evaluation will be conducted at two regions (Region 1 and 2 in Figure 8) in the experimental area.

Correctness and Completeness Comparison
In the correctness and completeness analysis of 3D models, we focused on the evaluation of correctness and completeness on the boundaries of these 3D models. The boundaries in the true orthophotos were taken as ground truth. For one 3D boundary, to check the distance and angle between the ground truth and itself, we overlap its 3D model and the true orthophotos. If the angle is smaller than a threshold and the distance is smaller than a threshold, then the boundary segment is considered as a true one; otherwise, it is considered as a wrong one. If a boundary in the reference data did not be detected by our approach, we call it a mis-detected segment. The correctness and completeness are calculated by Equation 5.
where \( TP \) is the number of true segments, \( FP \) is the number of wrong segments, and \( FN \) is the number of mis-detected segments. We used ten pixels as the distance threshold because of the complexity of the whole processes during 3D building model reconstruction, which includes registration of imagery and lidar data, 2D building boundary extraction, 3D building boundary determination, and 3D building model reconstruction. Any errors will be propagated from each step to the final result. For example, in Step One (Registration of Imagery and Lidar Data), the RMSE of 3.1 pixels and 3.0 pixels on the \( X \)-axis and \( Y \)-axis, respectively, are estimated. Considering more errors derived in the other process steps and errors in the lidar data and aerial images, we selected a relatively large number (ten pixels; 50 cm) as the distance threshold. For 3D building model reconstruction from multi-source datasets, however, it is not a loose threshold. In the 3D Building Boundary Determination Section, we use 4 degrees as a threshold for

![Figure 9. Comparison on the appearance of the 3D building models reconstructed by our approach (top) and manual operation (down): (a) and (b) No difference between the two kinds of 3D models, and (c) and (d) Three pairs of circles show the differences between the two kinds of 3D models.](image)
parallel segment determination. Considering that more errors may be produced in the 3D Building Model Reconstruction Section, here we select a larger number (5 degrees) as the angle threshold. Because we focused on the quality of the reconstructed 3D boundaries, a boundary reconstruction approach only based on lidar data introduced by Sampath and Shan (2007) is used, i.e., a lidar-approach, for further checking the performance of the proposed approach. The key idea of this lidar-approach is a hierarchical regularization approach: initially, relatively long line segments are extracted from the building boundary and their least squares solution is determined, assuming that these long line segments lie on two mutually perpendicular directions. Then, all the line segments are included to determine the least squares solution, using the slopes of the long line segments as weighted approximations. We compare the correctness and completeness of the building boundaries derived by the proposed approach and by the lidar-approach. Table 1 gives the results of correctness and completeness of 3D boundaries from the two approaches at Region 1 and 2, respectively. In the Region 1, the correctness of the boundaries derived by the proposed approach and the lidar-approach are 98 percent and 85 percent, respectively; the completeness of the boundaries derived by the proposed approach and the lidar-approach are 91 percent and 82 percent, respectively. In the Region 2, the correctness of the boundaries derived by the proposed approach and the lidar-approach are 92 and 82 percent, respectively; the completeness of the boundaries derived by the proposed approach and the lidar-approach are 89 percent and 87 percent, respectively. From Table 1, it is clear that the proposed algorithm leads to a very high correctness and completeness, and the proposed approach can lead to boundaries with much higher quality than the lidar-approach. By overlapping the boundaries reconstructed by the two approaches, Figure 10 illustrates their detailed differences: (a) some details may be missed by the lidar-approach, shown by the A labels in Figure 10, (b) some artifacts, especially some fake corners, may be created by the lidar-approach, shown by the B labels in Figure 10, and (c) building boundary segments may be shifted as an offset distance, shown by the C labels in Figure 10.

**Geometric Accuracy Comparison**

In this section, we focus on the estimation of horizontal accuracy of boundaries of these 3D models. The manually constructed 3D models cannot be taken as a reference data for geometric accuracy evaluation in vertical direction, because these 3D models have a reliable appearance, but don’t have a high geometric accuracy. The estimation is conducted at Region 1 and 2 in the experimental area. Figure 11 shows the overlay of the 3D boundaries and the true orthophotos. Some correct boundary segments are selected for the estimation of the boundaries’ accuracy. Focusing on a boundary segment, we measure the distance from the midpoint of the segment to its reference point (the corresponding point in the true orthophotos). According to the statistics values of the selected boundary segments, the mean absolute error (MAE), RMSE, and maximum error (MAX) are calculated for checking the accuracy of final boundaries.

<table>
<thead>
<tr>
<th>Test area</th>
<th>Approach</th>
<th>Dataset</th>
<th>Correctness</th>
<th>Completeness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Region 1</td>
<td>Our approach</td>
<td>Multi-view aerial images</td>
<td>98%</td>
<td>91%</td>
</tr>
<tr>
<td></td>
<td>lidar-approach</td>
<td>(5 cm) + lidar data (1.1 m)</td>
<td>85%</td>
<td>82%</td>
</tr>
<tr>
<td>Region 2</td>
<td>Our approach</td>
<td>Multi-view aerial images</td>
<td>92%</td>
<td>89%</td>
</tr>
<tr>
<td></td>
<td>lidar-approach</td>
<td>(5 cm) + lidar data (1.1 m)</td>
<td>82%</td>
<td>87%</td>
</tr>
</tbody>
</table>

Table 1. The Correctness and Completeness of the 3D Building Boundaries Reconstructed by Two Approaches

Figure 10. Comparison on the correctness and completeness of the 3D building boundaries reconstructed by our approach (black) and lidar-approach (gray): A: the missed details, B: some artifacts, and C: shifted boundary segments.)
respectively. In the Region 2, the RMSE, MAE, and MAX of the boundaries derived by the proposed approach are 0.23 m, 0.20 m, and 0.50 m, respectively; the RMSE, MAE, and MAX of the boundaries derived by the lidar-approach are 0.39 m, 0.30 m, and 0.98 m, respectively. Figure 12 illustrates the boundaries reconstructed by our approach (left side) and lidar-approach (right side) for Region 1 and 2. The short lines (orthogonal to boundaries) refer to errors of the boundaries, which are shown as vectors given the length and orientation. Each error line has been enlarged 20 times. From Table 2 and Figure 12, it is clear that boundaries determined by our approach have a higher geometric accuracy than by lidar-approach, which also proves that the approach of data integration can effectively advance the geometric accuracy of 3D models.

Conclusions
To automatically obtain 3D building models with precise geometric position and fine details, a new approach integrating multi-view imagery and lidar data is proposed in this study. In comparison to the lidar-approach, the 3D building models derived by the proposed approach have a higher correctness, completeness, and geometric accuracy, which means that the proposed approach can lead to higher quality products. The main contributions of this study are as follows: (a) the data integration strategy of multi-view imagery and lidar data is systematically studied for automatic reconstruction of 3D building models, (b) a new algorithm for the determination of principal orientations of a building is proposed, thus benefiting to improve the correctness and robustness of boundary segment extraction in aerial imagery, (c) a new dynamic selection strategy based on lidar point density analysis and K-means clustering is proposed for separating boundary and non-boundary segments, and (d) a new strategy for 3D building model reconstruction including “detect-locate-reconstruct” method and lidar-based robust reconstruction processes is proposed, which reduces the complexity of 3D reconstruction and improves the quality of the final 3D models.

However, it is still difficult to automatically reconstruct 3D building models with very complex structures (e.g., curve structures and very complex rooftop structures), where further investigation is needed. In addition, the proposed approach still requires further improvement in order to achieve more efficient generation of 3D building models.

Acknowledgments
This work is supported by the National Natural Science Foundation of China (Grant No. 41001238), the National 863 project, the National 973 project (Grant No. 2006CB701304), the Foundation of Innovative Research Groups of the National Natural Science Foundation of China (Grant No. 40721001). Sincere thanks are given for the comments and contributions of anonymous reviewers and members of the Editorial team.

The experimental results are listed in Table 2. In Region 1, the RMSE, MAE, and MAX of the boundaries derived by the proposed approach are 0.20 m, 0.14 m, and 0.64 m, respectively; the RMSE, MAE, and MAX of the boundaries derived by the lidar-approach are 0.43 m, 0.35 m, and 1.13 m.

Table 2. The Errors of the 3D Building Boundaries Reconstructed by Two Approaches

<table>
<thead>
<tr>
<th>Test area</th>
<th>Approach</th>
<th>Dataset</th>
<th>MAE(m)</th>
<th>RMSE(m)</th>
<th>MAX(m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Region 1</td>
<td>Our approach</td>
<td>Multi-view aerial images + lidar data(1.1 m)</td>
<td>0.14</td>
<td>0.20</td>
<td>0.64</td>
</tr>
<tr>
<td></td>
<td>Lidar-approach</td>
<td>lidar data (1.1 m)</td>
<td>0.35</td>
<td>0.43</td>
<td>1.13</td>
</tr>
<tr>
<td>Region 2</td>
<td>Our approach</td>
<td>Multi-view aerial images + lidar data(1.1 m)</td>
<td>0.20</td>
<td>0.23</td>
<td>0.50</td>
</tr>
<tr>
<td></td>
<td>Lidar-approach</td>
<td>lidar data (1.1 m)</td>
<td>0.30</td>
<td>0.39</td>
<td>0.98</td>
</tr>
</tbody>
</table>
Figure 12. Comparison of geometric accuracy of the 3D building boundaries reconstructed by our approach (left side) and lidar-approach (right side); the short lines (orthogonal to boundaries) refer to errors of the boundaries, which are shown as vectors given the length and orientation; each error line has been enlarged 20 times: (a) Region 1, our approach (left side) versus the lidar-approach (right side), and (b) Region 2, our approach (left side) versus the lidar-approach (right side).

References


Ma, R., 2004. Building Model Reconstruction from Lidar Data and Aerial Photographs, Ph.D. dissertation, The Ohio State University, Columbus, Ohio, 166 p.


(Received 03 September 2009; accepted 23 April 2010; final version 25 May 2010)