THE PHOTOGRAMMETRIC RECORD

The Photogrammetric Record 31(153): 29–50 (March 2016)
DOI: 10.1111/phor.12141

AUTOMATIC KEYLINE RECOGNITION AND 3D RECONSTRUCTION FOR QUASI-PLANAR FAÇADES IN CLOSE-RANGE IMAGES

CHANG LI (lcshaka@126.com)

Key Laboratory for Geographical Process Analysis & Simulation, Hubei Province, and College of Urban and Environmental Science, Central China Normal University, China

YONGJUN ZHANG (zhangyj@whu.edu.cn)
ZUXUN ZHANG (zhangzx@cae.cn)

School of Remote Sensing and Information Engineering, Wuhan University, China

Abstract

Critical keylines, such as concave and convex edges of a building façade, can be lost in photogrammetric recognition procedures. To solve this problem and to reconstruct quasi-planar 3D façades automatically and precisely, a set of algorithms and techniques for the automatic recognition of lines and 3D reconstruction is proposed. This includes: (1) a procedure for line-segment matching that satisfies the spatial requirements of a 3D scene based on "global independence" and "local dependence"; (2) a technique of generalised point bundle block adjustment combined with spatial line constraints (in the form of virtual observations) to control the propagation of error; and (3) the methods of perceptual organisation, plane fitting and plane–plane intersection are suggested to acquire the critical keylines corresponding to concave and convex building edges. Experimental results show that these new algorithms are feasible and applicable to recognition and 3D reconstruction. Recommendations for recognition methods are provided depending on whether or not a priori topological relationships are available between the planes under consideration.

KEYWORDS: generalised point photogrammetry, line recognition, line-segment matching, plane fitting, quasi-planar façades, 3D reconstruction

INTRODUCTION

THREE-DIMENSIONAL RECONSTRUCTION OF BUILDINGS is an important component of a digital city model (Baillard and Zisserman, 1999; Brenner, 2000; Zhang et al., 2005; Khoshelham et al., 2010), and building façades are a significant component of street landscapes. However, the automatic reconstruction of 3D objects from images remains a classic problem in computer vision and digital photogrammetry (Schindler and Bauer, 2003; Remondino et al., 2008; Haala and Kada, 2010). Recently, the interdisciplinary field of photogrammetric computer vision has emerged that involves fusing photogrammetry and computer vision to

support fully automated 3D modelling and mapping tasks using visual and electromagnetic sensors (Förstner, 2002; Zhang, 2004). Photogrammetric computer vision is becoming increasingly more popular, as evidenced by related research. Thus, the methods and techniques of the fields of photogrammetry and computer vision can be integrated to enhance both the automation and accuracy of 3D reconstruction. Typically, airborne and satellite imagery are used to extract texture information on the upper surfaces of 3D cartographic objects such as the ground surface and buildings (Baillard and Zisserman, 1999; Suveg and Vosselman, 2004; Barazzetti and Scaioni, 2009; Khoshelham et al., 2010; Huang et al., 2013; Rottensteiner et al., 2014). However, near-nadir photography from airborne and spaceborne platforms cannot acquire detailed textural information regarding building façades because of the limitations in the camera's viewing direction. Thus, a terrestrial (land-based) platform generally needs to be used for generating 3D building-façade models. Moreover, the processing of close-range images acquired by such terrestrial platforms is more challenging than that of aerial and satellite images because of the large geometric distortions among overlapping images, including partial occlusions, image rotation and viewpoint changes. These problems are addressed in this paper.

As well as planar surfaces, buildings possess both concave and convex components; therefore, recognising such edges is a key task in 3D building reconstruction. A concave edge has an external angle less than 180° (for example, in a recess); in a convex edge this angle is greater than 180° (as in a bay window). Line-segment matching can be used to address a broad range of scenes that include concave and convex edges. However, these key edges are not always matched or recognised because of noise and differences in line-extraction and line-matching performance. In the example shown in Fig. 1, the line matching was successful (Bay et al., 2005), but the regions marked with pecked ellipses (key edges) were not recognised. In addition, current algorithms for line matching do not always acquire both conjugate (corresponding) endpoints of a line segment (also see Fig. 1); this outcome is inconvenient for 3D reconstruction. Thus, the question of how to optimally and automatically recognise these "lost" concave and convex edges, together with matching the endpoints of line segments, is investigated in this paper.

The question of how to express discrete spatial features (such as point clouds and 3D lines) as a structure with topological relationships is an important issue in 3D reconstruction and modelling. Common and traditional methods of scene representation are always based on triangulated irregular networks (TINs) (Pollefeys et al., 2004). However, a 3D surface that is generated using a TIN always contains many "blurry" edges. In a TIN-based model, the model surface is divided into many small triangular facets, and the corresponding 3D scene on each triangular facet has a continuous surface. When an image contains a large planar region, the triangular facets are extended into a polygon that encompasses the entire plane, forming a *plane-based model* or a *piecewise planar model* (Werner and Zisserman,

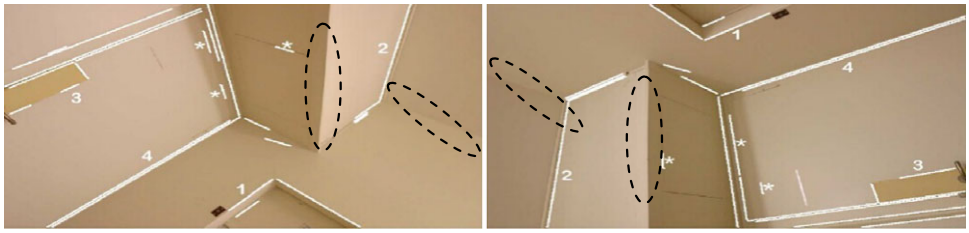


Fig. 1. Results for the recognition of line segments (Bay et al., 2005). Key concave and convex edges, indicated by the pecked ellipses, were not recognised.

2002; Bodis-Szomoru et al., 2014). Hence, a plane-based model can be applied to the 3D reconstruction of building façades to improve efficiency. In addition, such façades can often be regarded as roughly planar or forming a vertical building (Kang et al., 2010). Therefore, this paper focuses on the 3D modelling of quasi-planar façades based on a 3D wireframe surface model (3DWSM).

Motivated by the aforementioned studies and concerns, this paper has three main objectives:

(1) To offer an approach to image matching and shape recognition based on a particular philosophical principle, namely, the image-matching algorithm must satisfy the objective requirements for the spatial distribution of a 3D scene. This paper investigates an algorithm for line-segment matching, based on the authors' original research, which is improved and extended into a more generic technique for line-segment matching.
(2) To extend the scope of generalised point photogrammetry. A method of generalised point bundle block adjustment is proposed which is combined with spatial line constraints in the form of virtual observations to control the propagation of error in image rectification with large oblique angles using two vanishing points. (*Large oblique angle* implies the camera axis is significantly non-orthogonal relative to the building façade.)
(3) To improve the efficiency, precision and reliability of keyline recognition. The methods of perceptual organisation (coplanarity clustering), plane fitting and plane–plane intersection are proposed for the acquisition of critical keylines corresponding to concave and convex edges. The problem of missing key edge lines caused by failures of line extraction or matching is efficiently overcome.

A flowchart of the proposed methods is presented in Fig. 2 and will be elaborated upon in the remainder of the paper.

Stereomatching of Line Segments

Image matching can be a bottleneck in modern photogrammetric technology and is a classically difficult problem in 3D reconstruction. Area-based matching (ABM) and

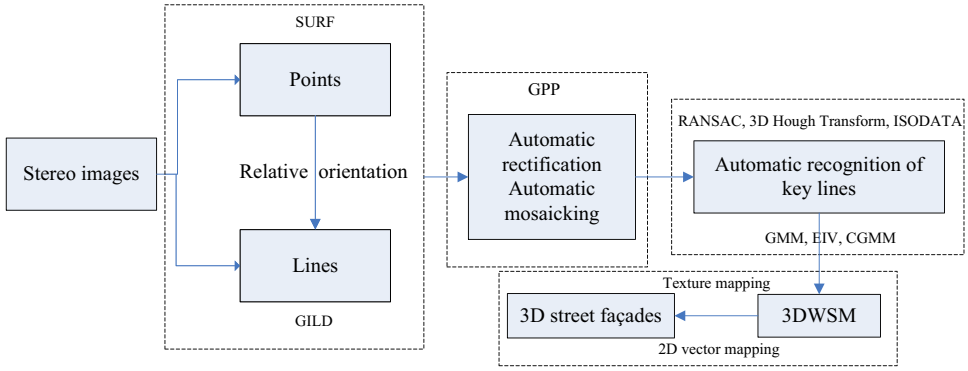


Fig. 2. Flowchart of the proposed methods for 3D façade reconstruction. Abbreviations are explained in the text.

feature-based matching (FBM) are classic types of image-matching algorithms (Gruen, 2012). Point descriptors include the scale-invariant feature transform (SIFT) (Lowe, 2004), speeded-up robust features (SURF) (Bay et al., 2008), principal components analysis SIFT (PCA-SIFT), gradient location-orientation histogram (GLOH) and binary robust invariant scalable keypoints (BRISK). Line descriptors include the mean standard-deviation line descriptor (MSLD) (Wang et al., 2009) and the line band descriptor (LBD) (Zhang and Koch, 2013). However, because the texture of a neighbourhood is not sufficiently rich, a linear feature description is less clear and robust than a point description (Fan et al., 2012). Homography is commonly used for line matching (Schmid and Zisserman, 1997, 2000), but it has certain limitations. When a scene is not planar or approximately planar, such algorithms are invalid. For aerial images, Ok et al. (2012) have proposed a new pairwise approach for line segments that are nearly parallel to the epipolar line. However, this approach may be not suitable for application to close-range stereo-images because of discontinuities and variations in parallax that occur in such imagery, which can change the topological relationships between objects.

The essence of image matching, as presented by Zhang (2007), is that the spatial distribution of a 3D environment is neither continuous nor smooth, yet local areas are. Thus, such an environment represents a *unity of opposites*, namely, global independence and local dependence (GILD). Therefore, Zhang (2007) proposed an independence rule for image matching in accordance with this principle. Furthermore, in the digital photogrammetry grid (DPGrid) system (Zhang et al., 2011), an innovative system for image matching based on this GILD rule was implemented. It is well known that a line segment contains more information than a point. In many state-of-the-art approaches to point matching (such as semi-global matching (Hirschmüller, 2005), dynamic programming and optimisation techniques based on graph cuts) the global minimum of an energy function is identified through local scanning and matching. In contrast, the proposed GILD strategy consists of globally matching potential candidates and then locally eliminating the ambiguity or outliers in candidate sets using line-segment continuity constraints. The advantage of this strategy is that omission-error matching, caused by an inappropriate (for example, too small) search space and occlusions (such as at the endpoints of a line segment), is reduced through globally independent matching. Furthermore, commission-error matching, caused by repetitive textures, similar scene structure and so on, is simultaneously reduced through locally dependent matching. In addition, complexly optimised searches in the solution space are avoided in the proposed GILD approach, thereby increasing computational efficiency.

Based on the GILD approach, an improved and more generic technique for line-segment matching has been developed by incorporating the procedures outlined in (1) to (4) below.

(1) Relative Orientation (RO) using FBM

Points are incorporated to assist in line-segment matching. According to the definition of epipolar geometry, corresponding (conjugate) points must be located along an epipolar line. Hence, relative orientation, performed by using the well-known SURF descriptor and random sample consensus (RANSAC) (Fischler and Bolles, 1981), can be performed to determine orientation parameters and transform a 2D matching problem into a 1D matching problem. In RO, five points from the matched results are randomly sampled to iteratively estimate and find the optimal parameters of the coplanarity model using a voting scheme.

(2) Line-Segment Extraction

This is implemented, on the basis of the authors' previous research (Li et al., 2009), using the following four steps:

(a) The image is pre-processed by Wallis filtering that is used to enhance image contrast and reduce the noise.
(b) The Canny (1986) operator is used to detect edges in the image.
(c) Feature grouping (perceptual organisation) and line fitting with hypothesis testing are utilised to merge short line segments.
(d) An adaptive least-squares matching algorithm (Ackermann, 1984; Gruen, 1985) is used to acquire the higher precision lines.

(3) Generating Candidate Line Segments using ABM and Global Independence

Let the endpoints of the target line segment, after its automatic extraction, be p and q (Fig. 3(a)). The sets of candidate endpoints for the line segment are $Q = \{q_1, q_2, \ldots, q_n\}$ and $P = \{p_1, p_2, \ldots, p_m\}$, which are generated by computing the global normalised cross correlation (NCC) to obtain local peak points. The matching processes for p and q are independent. The correlation coefficient is calculated based on the NCC, which is a classic ABM method. There are $n \times m$ possible line segments. In Fig. 3(a), eq and ep are the epipolar lines corresponding to q and p. $\{q_1, p_1\}$, $\{q_1, p_2\}$, $\{q_2, p_1\}$ and $\{q_2, p_2\}$ are the candidate line segments. Note the following five properties:

(a) Independence must be maintained at the endpoints of the matched line segment because the spatial distribution of a 3D object is not globally continuous or smooth.
(b) The search window must be sufficiently large to capture the corresponding endpoints and reduce omission-error matching.
(c) The candidate endpoints for a line segment must be generated by matching in the vicinity of the epipolar line, not just directly on the epipolar line, because of:
    (i) error due to uncalibrated radial distortion; and
    (ii) residual errors in the relative orientation affecting the propagated epipolar line.

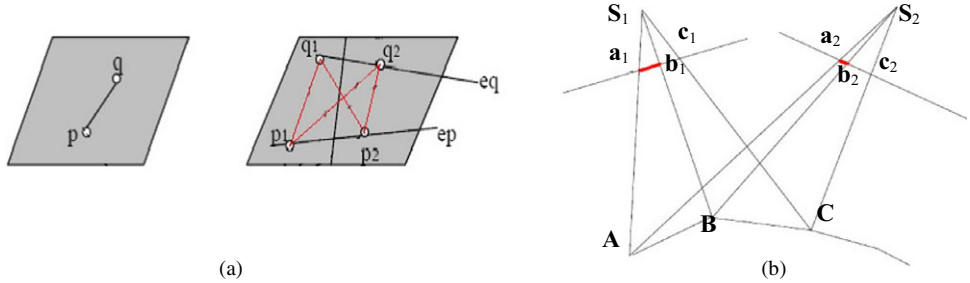


FIG. 3. Sketch of line-segment matching: (a) matching to generate candidate line segments (ep and eq are epipolar lines for p and q); (b) geometric length deformation between corresponding line segments.

(d) The geometric distortion among close-range stereo-images should first be corrected using the following affine-transformation model to compensate for the effects of image shifts and rotations:

$$x_R = a_0 + a_1 x_L + a_2 y_L, \quad y_R = b_0 + b_1 x_L + b_2 y_L \tag{1}$$

where (x_L, y_L) and (x_R, y_R) are the pixel coordinates in the left and right images, respectively. The coefficients a_0, a_1, a_2, b_0, b_1 and b_2 can be calculated from the conjugate points obtained by the relative orientation.

(e) Self-adaptive matching for occluded endpoints, which can effectively reduce omission errors, is implemented by:
(i) reserving endpoints with high NCC values;
(ii) abandoning endpoints with low NCC values that may be occluded; and
(iii) searching for new endpoints with high NCC values along the line segment.

(4) Eliminating the Ambiguity in Candidate Line Segments based on Local Dependence

The spatial distribution of a 3D object is locally continuous and smooth, in a similar way to a line segment. Therefore, continuity restrictions (local dependence) can be imposed during line-segment matching to eliminate ambiguities in the set of candidate line segments. In this paper, relaxation matching, which can control commission errors, is proposed to ensure local dependence by computing the probability for each candidate line segment based on Bayes' theorem. The basic algorithm can be described by the following two procedures:

(a) *Resampling*. For each pair consisting of the target line segment and one of the candidate line segments, resample one of the line segments to eliminate the length difference between them. In Fig. 3(b), the lengths of the corresponding line segments a_1b_1 and a_2b_2 are not equal: $|a_1b_1| \neq |a_2b_2|$. Thus, the shorter of the two line segments between the target and the candidate must first be found as follows:

$$Min\_Length\{|qp|, |q_i p_j|\}, \quad (i = 1, .., n; j = 1, .., m).$$

Then, taking the length of the shorter line segment as a reference, select windows which are 7 pixels wide in the direction orthogonal to the line segment. Next, resample the window of the longer line segment to a new window that is equal to the size (length and width) of the shorter line segment. Because of depth discontinuities and differences in silhouettes, different viewpoints will yield different backgrounds on the non-façade side of the line. Therefore, it is necessary to process the two sides separately; it then seems appropriate to choose the larger of the two similarity scores.

(b) *Matching*. Match the line segment based on probabilistic relaxation. This matching method has been successfully applied in the VirtuoZo digital photogrammetric system (Zhang et al., 1992). Based on the features of a line segment, a new compatibility coefficient is defined as follows:

$$C(i,j;k,l) = C_{ijkl} = \mathrm{NCC}\{resample\{(i,j),(k,l)\}\} \tag{2}$$

where (i, j) is the window for the target line segment qp in the left (reference) image, (k, l) is the window for the candidate line segment $q_i p_j$ $\{i = 1, \ldots, n; j = 1, \ldots, m\}$ in the right image, $resample\{(i, j), (k, l)\}$ represents the resampling window generated in the previous step (4a) and NCC represents the similarity-measuring approach. By measuring

the probability of the existence of a connectivity region between the endpoint and starting point of the line segment and then calculating whether the target and candidate line segments are compatible, the line segment can be matched under the constraint of local dependence (continuity restrictions). C_{ijkl} plays an important role in the relaxation-matching approach. Local consistency in line-segment matching can be achieved based on Bayes' theorem using an iterative scheme (Zhang et al., 2000; Zhang and Gruen, 2006):

$$P_{(a_i=w_j)}^{(n+1)} = \frac{P_{(a_i=w_j)}^{(n)} Q_{(a_i=w_j)}^{(n)}}{\sum_{\lambda\in\Omega} P_{(a_i=\lambda)}^{(n)} Q_{(a_i=\lambda)}^{(n)}} \Rightarrow P^{(n+1)}(i,j) = \frac{P^{(n)}(i,j)Q^{(n)}(i,j)}{\sum_{s=1}^{m_j} P^{(n)}(i,s)Q^{(n)}(i,s)} \tag{3}$$

and

$$Q^{(n)}(i,j) = P^{(n-1)}(i,j)\left(c_0 + c_1 \sum_{j\in N_i}\sum_{l=1}^{n_j} C_{ijkl} P^{(n-1)}(k,l)\right) \tag{4}$$

where c_0 and c_1 are relaxation coefficients set to 0 and 1, respectively; n is the number of iterations; and $Q^{(n)}(i, j)$ represents the probability of the existence of a connecting region from the starting point to the endpoint of the line segment in the nth iteration. When a candidate line segment with a matching probability higher than 0·9 is found, or when the maximum number of iterations is reached, the iterative calculation is terminated. At this time, the candidate line segment with the highest matching probability is considered to be the corresponding line segment.

Note that the proposed GILD approach differs from the method by Schmid and Zisserman (2000) as follows: (a) line-segment extraction; (b) similarity measurement; (c) matching of corresponding endpoints; (d) matching strategy for short-range and long-range motion; and (e) self-adaptive matching. Moreover, image pyramids and least squares matching (Ackermann, 1984; Gruen, 1985) can also be utilised to obtain matching results with high reliability and precision.

Automatic Image Rectification and Automatic Mosaicking with Generalised Point Photogrammetry

With terrestrial (land-based) imagery, large oblique angles lead to substantial perspective geometric distortions of building-façade textures. However, images with such tilted camera axes cannot be easily rectified and mosaicked without measured ground control points (GCPs) determined by traditional surveying methods. Therefore, rectification of the raw image based on the following proposed method is necessary to acquire a visually appealing texture.

Camera Calibration

In computer vision, vanishing points have been used for camera calibration (Caprile and Torre, 1990; Hartley and Zisserman, 2003). In generalised point photogrammetry (GPP) (Zhang and Zhang, 2004; Zhang et al., 2008), six equations, involving three interior orientation parameters (the principal point (x_0, y_0) and the principal distance f) and three exterior direction parameters (φ, ω and κ) can be computed based on vanishing points

located in three orthogonal directions relative to three limitless points along the X, Y and Z axes. Let X, Y and Z extend to infinity. The collinear equations of the vanishing points can be acquired as follows:

$$\begin{aligned} x_{X\infty} &= x_0 - f \cdot a_1/a_3, \quad y_{X\infty} = y_0 - f \cdot a_2/a_3 \\ x_{Y\infty} &= x_0 - f \cdot b_1/b_3, \quad y_{Y\infty} = y_0 - f \cdot b_2/b_3 \\ x_{Z\infty} &= x_0 - f \cdot c_1/c_3, \quad y_{Z\infty} = y_0 - f \cdot c_2/c_3 \end{aligned} \tag{5}$$

where ($x_{X\infty}$, $y_{X\infty}$), ($x_{Y\infty}$, $y_{Y\infty}$) and ($x_{Z\infty}$, $y_{Z\infty}$) are the coordinates of the intersections; φ, ω and κ are the angular elements corresponding to the coefficients of the rotation matrix (a_i, b_i and c_i, i = 1, 2, 3). The interior orientation parameters can be acquired using equation (5). Since the solution of (x_0, y_0) is unstable, (x_0, y_0) is approximately regarded as the centre coordinates of the image. Moreover, the interior orientation parameters (x_0, y_0 and f) are fixed during the solution processing. However, distortion parameters should be solved by pre-calibration.

Automatic Image Rectification and Mosaicking

The object in this study is a quasi-planar façade; indeed, façades are often approximately planar and vertical. Thus, two groups of lines parallel to the X and Y axes (corresponding to two vanishing points) in the object space are generally available on building façades, but the vanishing point in the depth direction (Z) cannot always be detected when the façade resembles a wall in the X–Y plane with little Z component; this is a common problem in terrestrial façade images. Furthermore, even if a vanishing point in Z can be detected in a quasi-planar façade scene, its accuracy is very low and it generally cannot be used because the extracted line segments corresponding to the Z vanishing point are always short and poorly distributed spatially over the façade scene (Li et al., 2011). Hence, the application of two constraints (in the X and Y directions) as control conditions for calculating the orientation parameters was proposed by Kang et al. (2010). However, the vanishing point in the depth orientation (Z) is not geometrically restricted. Thus, to allow for automatic rectification of a close-range image with large oblique angles, a more stable and rigorous algorithm is proposed called generalised-point bundle block adjustment combined with additional spatial line constraint. This method was developed based on the works cited above by applying object-space compensation in the Z direction. The steps of this procedure are as follows:

(1) Select matching line segments and group them in the vanishing-point direction as observations for subsequent adjustment.
(2) Acquire the initial values of the orientation parameters via the processing reported by Kang et al. (2010).
(3) Implement generalised-point bundle block adjustment combined with spatial line constraints via alternate convergence.

Typically, six parameters are used to describe a line segment:

$$\begin{cases} X = X_0 + t \cdot \cos\alpha \\ Y = Y_0 + t \cdot \cos\beta \quad (t_0 \le t \le t_1) \\ Z = Z_0 + t \cdot \cos\gamma \end{cases} \tag{6}$$

where (X_0, Y_0, Z_0) is the starting point, (α, β, γ) is the angular direction of the line, t is the distance from a particular point on the line to the starting point, and t_0 and t_1 define the

limits of parameter t. After corresponding line segments are matched and orientation parameters are gained in the previous step, 3D line segments are calculated by space intersection and parameterised using equation (6). Then, equation (6) can be substituted into the collinearity equations for a stereopair. Using two stereo-images (left and right images) as an example, after adding the constraint condition in the Z (depth) direction, the following is obtained:

$$\begin{cases} x_L - x_{L0} = -f_L \dfrac{a_{L1}(X_0 + t_0 \cdot \cos\alpha - X_{LS}) + b_{L1}(Y_0 + t_0 \cdot \cos\beta - Y_{LS}) + c_{L1}(Z_0 - Z_{LS})}{a_{L3}(X_0 + t_0 \cdot \cos\alpha - X_{LS}) + b_{L3}(Y_0 + t_0 \cdot \cos\beta - Y_{LS}) + c_{L3}(Z_0 - Z_{LS})}, |\theta| \geq 45^\circ \\ y_L - y_{L0} = -f_L \dfrac{a_{L2}(X_0 + t_0 \cdot \cos\alpha - X_{LS}) + b_{L2}(Y_0 + t_0 \cdot \cos\beta - Y_{LS}) + c_{L2}(Z_0 - Z_{LS})}{a_{L3}(X_0 + t_0 \cdot \cos\alpha - X_{LS}) + b_{L3}(Y_0 + t_0 \cdot \cos\beta - Y_{LS}) + c_{L3}(Z_0 - Z_{LS})}, |\theta| < 45^\circ \\ x_R - x_{R0} = -f_R \dfrac{a_{R1}(X_0 + t_1 \cdot \cos\alpha - X_{RS}) + b_{R1}(Y_0 + t_1 \cdot \cos\beta - Y_{RS}) + c_{R1}(Z_1 - Z_{RS})}{a_{R3}(X_0 + t_1 \cdot \cos\alpha - X_{RS}) + b_{R3}(Y_0 + t_1 \cdot \cos\beta - Y_{RS}) + c_{R3}(Z_1 - Z_{RS})}, |\theta| \geq 45^\circ \\ y_R - y_{R0} = -f_R \dfrac{a_{R2}(X_0 + t_1 \cdot \cos\alpha - X_{RS}) + b_{R2}(Y_0 + t_1 \cdot \cos\beta - Y_{RS}) + c_{R2}(Z_1 - Z_{RS})}{a_{R3}(X_0 + t_1 \cdot \cos\alpha - X_{RS}) + b_{R3}(Y_0 + t_1 \cdot \cos\beta - Y_{RS}) + c_{R3}(Z_1 - Z_{RS})}, |\theta| < 45^\circ \\ Z_1 - Z_0 = 0 \end{cases} \tag{7}$$

where f_L, x_{L0}, y_{L0} and f_R, x_{R0}, y_{R0} are elements of the interior orientation elements for the left and right images, respectively; X_{LS}, Y_{LS}, Z_{LS} and X_{RS}, Y_{RS}, Z_{RS} are the positional elements of the exterior orientation for the left and right images, respectively; the coefficients of the rotation matrix (a_{Li}, b_{Li} and c_{Li}, $i = 1, 2, 3$) and (a_{Ri}, b_{Ri} and c_{Ri}, $i = 1, 2, 3$) consist of the rotation angles φ_L, ω_L, κ_L and φ_R, ω_R, κ_R for the left and right images, respectively; and Z_0 and Z_1 are the coordinates of the two endpoints of the line segment in the depth direction. Note that the angle θ (defined by the direction of an image line with respect to the image x axis) of the observed line is used to select the equation in the x or y direction; details are provided in previous papers (Zhang and Zhang, 2004; Zhang et al., 2008). The weight matrix is very important to the suggested combined bundle adjustment to ensure that reliable solutions are obtained. Observations that exhibit gross errors (for example, if the Z values of the endpoints are significantly unequal) are removed by the robust estimation technique. Moreover, the weights of all observations are determined based on the a posteriori variance estimation (iterative method with variable weights) presented by Li and Yuan (2002), which is recalculated based on the iterative results obtained during the bundle adjustment.

As a result, the virtual constraints are incorporated into the generalised point bundle adjustment. Then, the orientation parameters (except for the coordinates of the perspective centre) are employed to rectify the stereo-images with large oblique angles. Note that the proposed method is a supplementary method of camera calibration and image rectification under the condition of two vanishing points. If three orthogonal vanishing points can be acquired with high accuracy, then the traditional method can also be used.

After rectifying the images' highly tilted camera axes, the method presented by Kang et al. (2010) is used to automatically retrieve the corresponding images for each building façade from the raw image sequence. This process is performed via the detection of the range variance using the histogram of the projective differences between the corresponding points and lines for each of the façades in the raw images.

Key Edge Recognition

Recognition of non-extracted or unmatched concave and convex critical edge lines is a complicated task, so the following methods are suggested for resolving this problem.

Recognising the Edges of Building Roofs

In contrast to aerial and satellite images, terrestrial images possess the particular feature that the sky appears above the roof of a building. Therefore, a roof can be recognised by segmenting the sky from other objects, in other words performing a single-class or unary classification. Afterwards, the extracted sky data is regarded as the background (the foreground will include building roofs). Even if the roofs are not recognised in a few scenes, most roof edges can still be obtained, thereby reducing the post-processing workload.

This paper presents an algorithm based on parallelepiped classification for roof recognition, which can obtain favourable results but at high computational cost. First, images with large oblique angles must be rectified and mosaicked. Then, the brightness values from each pixel of the red/green/blue (RGB) images are used to produce a 3D RGB mean vector, $\boldsymbol{\mu}_s = (\mu_{s1}, \mu_{s2}, \mu_{s3})$, in which μ_s is the mean value of the training data obtained for the class s (sky). A similar notation is used for class c (cloud) (s and c constitute the two possible classes (unary classification)). Here σ_s and σ_c are the standard deviation of the training data for classes s and c, respectively. Afterwards, on the basis of a 3σ criterion for normal distributions with a decision tree (Quinlan, 1987), a parallelepiped method is used to determine the foreground and background:

$$\left\{\begin{array}{l} \text{if } \Omega_S\{(x,y)|\mathbf{v}_i \in (\mu_{s1} \pm 3\sigma_{s1}, \mu_{s2} \pm 3\sigma_{s2}, \mu_{s3} \pm 3\sigma_{s3})\} \Rightarrow \text{sky} \\ \left.\begin{array}{l} \text{else if } \Omega_C\{(x,y)|\mathbf{v}_i \in (\mu_{c1}, \pm 3\sigma_{c1}, \mu_{c2} \pm 3\sigma_{c2}, \mu_{c3} \pm 3\sigma_{c3})\} \\ \text{if } \Omega_C(x,y) \subset \Omega_S(x,y) \end{array}\right\} \Rightarrow \text{cloud} \\ \text{else background (buildings and occlusions etc.)} \end{array}\right\} \Rightarrow \text{foreground} \quad (8)$$

where $\mathbf{v}_i$ is a pixel vector with an RGB value under the specified classification, and Ω_S and Ω_C are the segmented areas corresponding to the sky and the clouds, respectively. To eliminate noise interference, the topological relationship is constrained to satisfy the requirement that area Ω_C lies inside area Ω_S. After the extraction of lines from the image, several fragmented line segments, including roof edges, are acquired. Then, a broken roof edge can be treated as an initial recognition buffer (area) that can be automatically located by searching and classifying sky along the top of the image. Note that there is no need to recognise the entire sky; the focus is only in the neighbourhood of the roof. Hence, when compared with the Gaussian mixture model and support vector machine, the computational efficiency and reliability of roof recognition are considerably improved by the suggested method due to the reduction in the area subjected to image segmentation. Furthermore, because the sky's boundary is not always the roof boundary of the façade, a few human–computer interactions may be necessary for post-editing of roof boundaries.

Finally, after the image areas are identified as either foreground or background in the classification based on the parallelepiped method and a decision tree, the edges can be easily detected using a Canny (1986) operator such that the edge coordinates of the building roof are obtained by tracing the roof edges.

Key Edges for Building Recognition

As illustrated in Fig. 1, it is difficult to entirely avoid the problem of "missing" concave and convex edges. A method to recognise missing edges is proposed as follows:

(1) The matched conjugate lines are intersected to obtain spatial 3D lines.
(2) These spatial lines are clustered for the classification of coplanar lines by combining RANSAC, the 3D Hough transform (Overby et al., 2004) and the iterative self-organising data analysis technique algorithm (ISODATA) (used for 3D geometric pattern recognition). RANSAC can be employed to detect planes quickly and accurately (Tarsha-Kurdi et al., 2007). Hence, to begin, three points from two random parallel line segments are selected; the parameters of the corresponding plane are calculated as:

$$x\cos\theta\cos\varphi + y\cos\ \theta\ sin\,\varphi + z\sin\theta = \rho(\theta,\varphi) \tag{9}$$

where a plane $\Pi \in \Re^3$ is uniquely defined by a triplet (θ, φ, d), $\theta \in \{0, 2\pi\}$ and $\varphi \in \{-\pi/2, \pi/2\}$ denote the two angles (azimuth and elevation) associated with the spherical representation of the plane's unit-length normal vector, and $d \geq 0$ denotes the distance from the origin of the coordinate system to the plane. When the triplet (θ, φ, d) is calculated for a 3D plane, information related to any other 3D plane constitutes noise and outliers that should be eliminated using RANSAC.

However, because of the shortcomings of the Hough transform, in the initial space segmentation some geometrically coplanar planes that are actually discontinuous planes may be mistakenly clustered. Therefore, this paper proposes the application of ISODATA to the initial results for the final space segmentation. In ISODATA, the standard deviation within each cluster and the distances between cluster centres are calculated. Clusters are split if one or more standard deviations are greater than a user-defined threshold; conversely, they are merged if the distance between clusters is smaller than a user-defined threshold; thus, the application of this procedure can overcome the shortcomings of the 3D Hough transform.
(3) The utilisation of spatial coplanar lines is suitable for fitting spatial planes. Three methods are considered to determine which yields the best fitting results, which directly affect the accuracy of edge recognition:

(a) *Gauss–Markov Model (GMM).* The Gauss–Markov theorem (Plackett, 1950) states that, in a linear regression model in which the errors are expected to be zero, are uncorrelated and have equal variances, the best linear unbiased estimator of the coefficients is the ordinary least squares estimator. Suppose that the equation of a plane is

$$ax_i + by_i + cz_i + 1 = 0 \tag{10}$$

where $\mathbf{p}_i = (x_i,\ y_i,\ z_i)^{\mathrm{T}}$ represents the 3D coordinates of a point on the plane and $\mathbf{n} = (a,\ b,\ c)$ is the normal vector to the plane. The n points are composed of line segments, which is computed using an iterative method with variable weights. This is a weighted least squares solution.

(b) *Constrained GMM (CGMM).* Usually, constraints can be added to the fitting of the plane to achieve more optimal fitting performance. For example, in Fig. 4, suppose that $\mathbf{n}_1 = (a_1,\ b_1,\ c_1)$, $\mathbf{n}_2 = (a_2,\ b_2,\ c_2)$ and $\mathbf{n}_3 = (a_3,\ b_3,\ c_3)$ are the normal vectors of planes 1, 2 and 3, respectively. The three spatial equations for points i, j and k are:

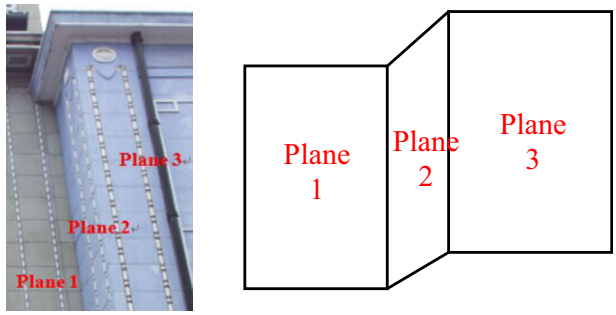


FIG. 4. Topological relationship among planes.

$$\begin{cases} a_1x_i + b_1y_i + c_1z_i + 1 = 0 \\ a_2x_j + b_2y_j + c_2z_j + 1 = 0 \\ a_3x_k + b_3y_k + c_3z_k + 1 = 0. \end{cases} \tag{11}$$

Suppose that the constraints are $n1 \perp n2$, $n2 \perp n3$ and $n1 /\!/ n3$. Then:

$$\begin{cases} a_1a_2 + b_1b_2 + c_1c_2 = 0 \\ a_2a_3 + b_2b_3 + c_2c_3 = 0 \\ a1/a3 = b1/b3 = c1/c3. \end{cases} \tag{12}$$

The combined adjustment model consists of equations (11) and (12), which are also computed using the iterative method with variable weights (Li and Yuan, 2002). This is a constrained Gauss–Markov model (CGMM) or a constrained adjustment model.

(c) *Errors-in-Variables (EIV) Model*. A model of this type is a regression model that accounts for measurement errors in the independent variables, such that observational errors in both the dependent and independent variables are considered (Markovsky and van Huffel, 2007). Such a model is introduced to improve the accuracy of plane fitting. The EIV model for the plane equation (9) or (10) can be written as follows:

$$\arg\min_{\mathbf{E}_A, \mathbf{e}_b} \|[\mathbf{E}_A \mathbf{e}_b]\|_F, (\mathbf{A} - \mathbf{E}_A)\mathbf{X} = \mathbf{b} - \mathbf{e}_b \tag{13}$$

where $[\mathbf{E}_A \ \mathbf{e}_b]$ is the augmented matrix in which $\mathbf{E}_A$ and $\mathbf{e}_b$ are adjacent, $\mathbf{b}$ is the closure-error vector and $\|\cdot\|_F$ is the Frobenius norm. According to the geometric interpretation of principal component analysis (PCA), this approach also yields results equivalent to those of an EIV model. Data centralisation is applied to the point set, which consists of observations of the two endpoints of line segments. Singular value decomposition (SVD) is used to solve the equation of the plane. To remove the impact of gross errors, RANSAC should be applied.

(4) The intersections of the clustered spatial planes yield the critical lines corresponding to concave and convex edges.

Experiments and Analysis

Several sets of close-range digital images were used for the experiments. Three of the investigated datasets, which were captured to cover various building façades, are shown in Fig. 5. The images were acquired using two hand-held non-metric digital cameras (Kodak Professional DCS Pro SLR/n and Sony Cybershot 5·0 Mpixels) along both residential and commercial streets, as shown in Fig. 5. The image size for datasets 1 and 2 (from the Kodak camera) is 1000 × 1500 pixels, with a pixel size of 0·025 mm. The image size for dataset 3 (from the Sony camera) is 1000 × 750 pixels, with a pixel size of 0·035 mm. The images in these datasets were compressed to avoid the necessity of processing large raw images. The proposed algorithms were implemented and the 3D model visualised using Visual C++ and OpenGL (Windows 7 Professional; 64 bit IBM ThinkPad X200; Intel Core 2 Duo CPU P8600 @ 2·40 GHz; 4·00 GB RAM). The experimental results for the matching, rectification, mosaicking and recognition of the images, together with the 3D reconstruction of the building façades, are illustrated and discussed below.

Image Matching for Line Segments

Corresponding (conjugate) points were obtained using SURF. Then, the relative orientation was performed for the detection of outliers using RANSAC and for the calculation of epipolar lines. Five conjugate points are sufficient to solve for the five relative orientation parameters; however, it is suggested that detecting and matching approximately 20 to 30 correct SURF points in the four corner areas of a stereopair to enhance the matching speed. It is noted that SURF descriptor matching is an error-prone process, so the initial statistical sample of matched SURF points should be larger than 20 to 30. Sigma naught (σ_0) for the relative orientation was approximately a quarter of the physical pixel size. Afterwards, the geometric distortion between the close-range stereo-images was corrected using equation (1).

Fig. 6 shows the line-segment matching results for the three images with numbers 33 to 35. To ensure the acquisition of reliable results, only line segments longer than 40 pixels were retained for matching in the line-extraction step. Through repeated testing, the NCC value was set to 60% for the generation of candidate line segments, and the search area was expanded by 3 to 4 pixels in the direction orthogonal to the epipolar line to reduce matching omission errors caused by the uncertainty (error) in the epipolar line and uncalibrated camera parameters. Moreover, an image pyramid was constructed using a 3-pixel mean filter (thus a 3 × 3 array of pixels was degraded to a single coarse pixel) to enhance reliability and to reduce commission errors in matching. Approximately 200 successfully matched conjugate line segments were identified for each image pair using the

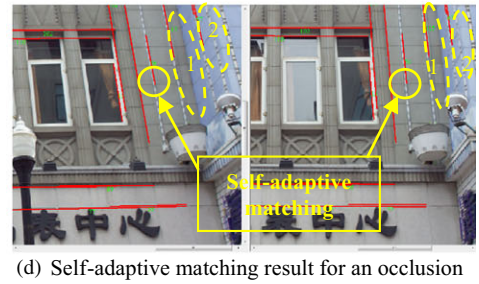
(a) Dataset 1 Kodak (b) Dataset 2 Kodak (c) Dataset 3 Sony

Fig. 5. Close-range images acquired with: (a) and (b) a Kodak Professional DCS Pro SLR/n camera; and (c) a Sony Cybershot 5·0 Mpixel camera.

(a) Line matching between images 33 and 34

(b) Line matching between images 34 and 35

(c) Occlusion of a line segment



(d) Self-adaptive matching result for an occlusion

FIG. 6. Line-segment results for dataset 1 (sequence of images 33 to 35).

proposed GILD strategy. Although a large number of densely matched line segments were not obtained, a sufficient number of matched lines were still available to calculate the plane parameters for 3D reconstruction.

All corresponding line segments for the two overlapping stereopairs shown in Fig. 6 were randomly distributed throughout the overlap areas, with a few outliers. Furthermore, the overall accuracy of the matching results was higher than 97%, which shows the commission error is well controlled. Regarding the computational efficiency for a stereopair, the proposed GILD method of line-segment matching required less than 5 s using the hardware, software and data size specified above. The proposed method cannot only match conjugate line segments but can also successfully match the two endpoints of a line segment. In Figs. 6(c) and (d), self-adaptive matching for occluded endpoints is demonstrated in the circular areas which reduces omission-error matching; however, two key edges were not matched (marked as pecked ellipses 1 (concave) and 2 (convex) on the right of Fig. 6(d)) because the corresponding line segments were missed during the automatic extraction process.

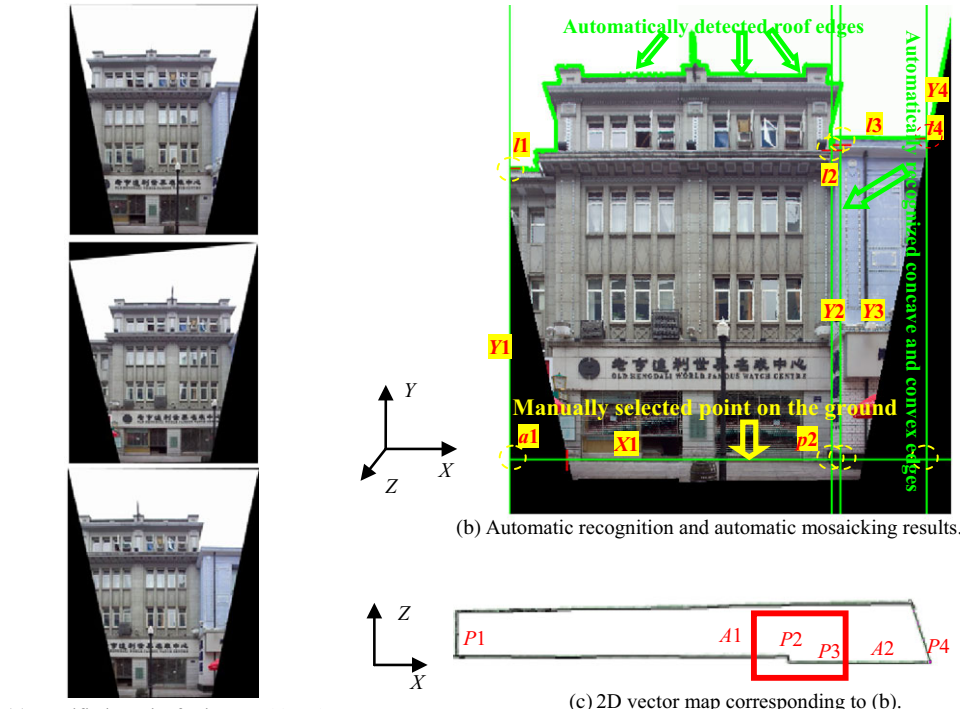
Image Rectification and Mosaicking for Images with Large Oblique Angles

The line segments were extracted and grouped according to the angle histogram; the results were added to the interpretation plane, which describes the geometric relationship between parallel lines in object space and the corresponding lines in image space. Then, the vanishing-point coordinates were determined based on the appropriately weighted intersection of the lines in the image space. Using two constraints (in the X and Y directions) as control conditions (Kang et al., 2010), together with equation (5), the orientation parameters were calculated. The stereopair consisting of images 34 and 35 was chosen to demonstrate the effectiveness of the improved method (namely, generalised point bundle block adjustment combined with spatial line constraints). The significance level α that was used for hypothesis testing was set to 0·05 for the iterative method with variable weights. The corresponding interior and exterior orientation parameters were calculated in the experiments. Moreover, the spatial-intersection results (3D start and endpoints (X_1, Y_1, Z_1) and (X_2, Y_2, Z_2)) prior to adopting the proposed methodology showed that, in the depth direction, $Z_1 \neq Z_2$. However, after the application of the proposed adjustment method, the Z coordinates were the same ($Z_1 = Z_2$). The new orientation parameters for the images were adjusted and utilised to rectify and mosaic the images, as seen in Figs. 7(a) and (b).

Key Edge Recognition

As shown in Fig. 7(b), the edges of the building roofs were detected and recognised using the suggested parallelepiped method. Different training samples were selected for different images depending on the various spectra of the sky and clouds that were obtained with changes in the environment. The RGB values ($\mathbf{v}$ in equation (8) with these 8-bit images) for datasets 1 and 2 were set to (251 ± 4, 252 ± 3, 254 ± 1), and for dataset 3 they were set to (236 ± 19, 248 ± 7, 252 ± 3); this was because of the occlusion of the sky and the scattering by atmospheric particles.

For key edges in building recognition, the initial segmentation results (object-space line segments intersected by conjugate line segments with interior and exterior parameters) were clustered using the Hough transform. This was extended with RANSAC to acquire both the initial cluster centres and the number of clusters for the next step in the classification. Subsequently, ISODATA was applied to complete the final space segmentation. During this



(a) Rectified results for images 33 to 35

(b) Automatic recognition and automatic mosaicking results.

(c) 2D vector map corresponding to (b).

FIG. 7. Texture space and corresponding 2D vector space.

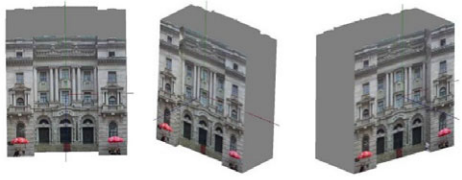
step, clusters may be split: the Hough transform can mistakenly cluster some geometrically coplanar planes that are, in reality, discontinuous planes. For example, the missing concave and convex keylines indicated in Fig. 6(d) were recognised as a result of the proposed plane intersection calculations; they are labelled $l1$, $l2$, $l3$ and $l4$ in Fig. 7(b). The plane parameters for planes 1 and 2, as shown in Fig. 4, were calculated using the three proposed methods, namely, RANSAC + GMM, RANSAC + EIV and RANSAC + CGMM. The angles between the two planes and the mean error in the back-projection of the image are presented in Table I. The RANSAC + CGMM approach yielded the best recognition results among the three methods in terms of both the angles and the mean error; RANSAC + EIV also yielded good results.

Thus, if the topological connections between planes are known (for example, the vertical or parallel relationships depicted in Fig. 4), then RANSAC + CGMM is recommended for recognising key edges; otherwise, RANSAC + EIV is the preferred choice. This difference is apparent in the three datasets depicted in Fig. 8. In Figs. 8(a) and (b), for identifying the lines corresponding to the concave and convex corners of a building, RANSAC + CGMM is suggested in the case of datasets 1 and 2. However, RANSAC + EIV is suggested for dataset 3 (Fig. 8(c)) because no a priori information regarding the topological relationship between the adjacent planes is available. For the building boundary with the ground, a point on this boundary should be manually selected to determine the Y coordinate in Fig. 7(b); after image rectification, the line segments corresponding to the X vanishing point are parallel to the x direction in the image space. Subsequently, the 3D model of the façade can be

TABLE I. Angles between two building planes and the mean error using different algorithms.

Algorithm	*RANSAC + GMM*	*RANSAC + EIV*	*RANSAC + CGMM*
Angles between planes	86° 31′ 37″	88° 14′ 15″	90°
Mean error	0·8 pixel	0·6 pixel	0·5 pixel

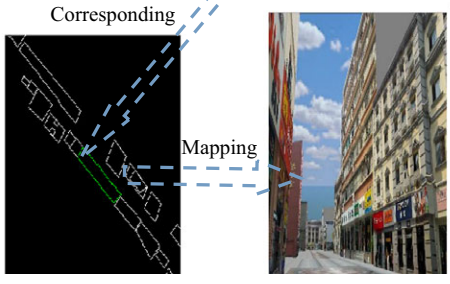
(a) Results of the 3D reconstruction for dataset 1



(b) Results of the 3D reconstruction for dataset 2

(c) Results of the 3D reconstruction for dataset 3

(d) 3D model (relative coordinates)



(e) 2D vector

(f) 3D model (absolute coordinates)

(g) Point cloud and surface reconstruction based on a TIN

FIG. 8. 3D reconstructions created using a wireframe surface model and textures.

reconstructed using a 3DWSM (wireframe model). Finally, the surface model in the depth (Z) direction can be reconstructed readily and reproducibly in a virtual reality environment or as a 2D vector map along the Z direction.

Compared with a traditional TIN based on point clouds for 3D interpretation (Fig. 8(g)), the proposed method produces clearer contour edges without the blurriness observed in Fig. 8(a). Lafarge and Mallet (2012) note that a hybrid representation that combines meshes (irregular elements) and geometric primitives (regular structures) can generally provide highly accurate modelling results. However, these authors report that this algorithm is not optimal when the vertical accuracy of input points is poor and the point density is weak (typically with low-resolution digital surface models (DSM)). Unfortunately, the input data in the current paper are stereopairs rather than multiview stereo-images, so a dense point cloud may not be available. Therefore, the suggested 3DWSM approach is more suitable than a TIN for 3D representation. The façades of datasets 1 to 3 that are shown in Figs. 8(a) to (c) were generated from 54, 8 and 31 images, respectively. Finally, the image textures were mapped onto three 3D models of street façades for the three datasets. Approximately 100 façades from several streets were reconstructed to test the proposed approaches. Figs. 8(d) to (f) illustrate the transformation of a 3D model from a relative coordinate system into an absolute coordinate system by using a 2D vector map and applying a scale factor λ:

$$\lambda = |A1P2|/|a1p2| = \sqrt{[(Z_{A1} - Z_{P2})^2 + (X_{A1} - X_{P2})^2]/[(y_{a1} - y_{p2})^2 + (x_{a1} - x_{p2})^2]} \quad (14)$$

where lines $a1p2$ and $A1P2$ can be identified in Figs. 7(b) and (c), respectively. For this final step, it is proposed that shape contexts are employed (Belongie et al., 2000, 2002; Ling and Jacobs, 2007) for shape matching between a 2D vector map and the projection of a 3D model in the Z–X plane. However, it is essential to employ human–computer interactions when ambiguities arise in the shape matching.

Conclusions

On the basis of photogrammetric computer vision, a set of algorithms and techniques for keyline recognition and 3D reconstruction of quasi-planar façades has been proposed in this paper. Recognising the edges of building façades is a key step in 3D reconstruction. The classic problem of line-segment matching can be overcome by applying the GILD principle; in this manner, good experimental results can be achieved. Furthermore, plane fitting and spatial intersection are proposed to compensate for missing concave and convex edges of a façade. The accuracy of plane fitting is a critical factor that affects edge-recognition accuracy. Strategies for plane fitting are proposed, depending on whether the topological relationships between the planes are known or unknown.

Finally, a 3D model of a building façade can be reconstructed by mapping the rectified texture space into the model space. Moreover, a 3D model in absolute coordinates can be obtained if the corresponding 2D vector map is known. Otherwise, a 3D model is obtained in a relative coordinate system, which is similar to the results obtained in computer vision.

In the 3D reconstruction of a planar façade, the test results for the proposed methods and techniques illustrate several advantages:

(1) automation was achieved based on a set of matching and recognition methods;
(2) accuracy was improved by applying the GPP model and the proposed adjustment;

(3) reliability was enhanced using 3DWSM, yielding a result without blurring; and
(4) efficiency was achieved by applying a high-speed automatic processing algorithm.

The approach presented in this paper, along with that reported in the paper by Kang et al. (2010), is applicable for the 3D modelling of urban streets. The method demonstrates high efficiency in built-up areas. Future work will concentrate on more meticulous aspects of 3D reconstruction, such as integrating volumetric models with 3DWSM, improving the edge-extraction and edge-recognition performance in the case of more complex and high-rise buildings, and applying computer-aided design (CAD) for high-level 3D building modelling.

Acknowledgements

The authors thank the National Natural Science Foundation of China (NSFC) (grant nos. 41101407, 41571434 and 41322010); the Natural Science Foundation of Hubei Province (grant nos. 2014CFB377 and 2010CDZ005), China; and self-determined research funds of Central China Normal University (CCNU) from the college's basic research and operation of MOE (grant nos. CCNU15A02001) for supporting this work.

References

Ackermann, F., 1984. Digital image correlation: performance and potential application in photogrammetry. *Photogrammetric Record*, 11(64): 429–439.

Baillard, C. and Zisserman, A., 1999. Automatic reconstruction of piecewise planar models from multiple views. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Fort Collins, Colorado, USA. 2: 559–565.

Barazzetti, L. and Scaioni, M., 2009. Automatic orientation of image sequences for 3D object reconstruction: first results of a method integrating photogrammetric and computer vision algorithms. *Proceedings of 3D-ARCH*, Trento, Italy. 25–28.

Bay, H., Ferrari, V. and van Gool, L., 2005. Wide-baseline stereo matching with line segments. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, San Diego, California, USA. 1: 329–336.

Bay, H., Ess, A., Tuytelaars, T. and van Gool, L., 2008. Speeded-up robust features (SURF). *Computer Vision and Image Understanding*, 110(3): 346–359.

Belongie, S., Malik, J. and Puzicha, J., 2000. Shape context: a new descriptor for shape matching and object recognition. *Neural Information Processing Systems Conference*, Cambridge, UK. 7 pages.

Belongie, S., Malik, J. and Puzicha, J., 2002. Shape matching and object recognition using shape contexts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(4): 509–522.

Bodis-Szomoru, A., Riemenschneider, H. and vanGool, L., 2014. Fast, approximate piecewise-planar modeling based on sparse structure-from-motion and superpixels. *IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, Ohio, USA. 469–476.

Brenner, C., 2000. Towards fully automatic generation of city models. *International Archives of Photogrammetry and Remote Sensing*, 33(B3): 85–92.

Canny, J., 1986. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(6): 679–698.

Caprile, B. and Torre, V., 1990. Using vanishing points for camera calibration. *International Journal of Computer Vision*, 4(2): 127–139.

Fan, B., Wu, F. and Hu, Z., 2012. Robust line matching through line–point invariants. *Pattern Recognition*, 45(2): 794–805.

Fischler, M. A. and Bolles, R. C., 1981. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6): 381–395.

Förstner, W., 2002. Computer vision and photogrammetry – mutual questions: geometry, statistics and cognition. *Bildteknik/Image Science, Swedish Society for Photogrammetry and Remote Sensing*. 151–164.

Gruen, A., 1985. Adaptive least squares correlation: a powerful image matching technique. *South African Journal of Photogrammetry, Remote Sensing and Cartography*, 14(3): 175–187.

Gruen, A., 2012. Development and status of image matching in photogrammetry. *Photogrammetric Record*, 27(137): 36–57.

Haala, N. and Kada, M., 2010. An update on automatic 3D building reconstruction. *ISPRS Journal of Photogrammetry and Remote Sensing*, 65(6): 570–580.

Hartley, R. and Zisserman, A., 2003. *Multiple View Geometry in Computer Vision*. Cambridge University Press, Cambridge, UK. 560 pages.

Hirschmüller, H., 2005. Accurate and efficient stereo processing by semi-global matching and mutual information. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2: 807–814.

Huang, H., Brenner, C. and Sester, M., 2013. A generative statistical approach to automatic 3D building roof reconstruction from laser scanning data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 79: 29–43.

Kang, Z., Zhang, L., Zlatanova, S. and Li, J., 2010. An automatic mosaicking method for building facade texture mapping using a monocular close-range image sequence. *ISPRS Journal of Photogrammetry and Remote Sensing*, 65(3): 282–293.

Khoshelham, K., Nardinocchi, C., Frontoni, E., Mancini, A. and Zingaretti, P., 2010. Performance evaluation of automated approaches to building detection in multi-source aerial data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 65(1): 123–133.

Lafarge, F. and Mallet, C., 2012. Creating large-scale city models from 3D-point clouds: a robust approach with hybrid representation. *International Journal of Computer Vision*, 99(1): 69–85.

Li, C., Wu, H., Hu, M. and Zhou, Y., 2009. A novel method of straight-line extraction based on Wallis filtering for the close-range building. *IEEE Asia Pacific Conference on Postgraduate Research in Microelectronics & Electronics*. 290–293.

Li, C., Zhang, Z. and Zhang, Y., 2011. Evaluating the theoretical accuracy of error distribution of vanishing points. *Acta Geodaetica et Cartographica Sinica*, 40(30): 393–396.

Li, D. and Yuan, X., 2002. *Error Processing and Reliability Theory*. Wuhan University Press, Wuhan, China. 417 pages.

Ling, H. and Jacobs, D. W., 2007. Shape classification using the inner-distance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(2): 286–299.

Lowe, D. G., 2004. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2): 91–110.

Markovsky, I. and van Huffel, S., 2007. Overview of total least-squares methods. *Signal Processing*, 87(10): 2283–2302.

Ok, A. O., Wegner, J. D., Heipke, C., Rottensteiner, F., Soergel, U. and Toprak, V., 2012. Matching of straight line segments from aerial stereo images of urban areas. *ISPRS Journal of Photogrammetry and Remote Sensing*, 74: 133–152.

Overby, J., Bodum, L., Kjems, E. and Iisøe, P. M., 2004. Automatic 3D building reconstruction from airborne laser scanning and cadastral data using Hough transform. *International Archives of Photogrammetry and Remote Sensing*, 35(B3): 296–301.

Plackett, R. L., 1950. Some theorems in least squares. *Biometrika*, 37(1–2): 149–157.

Pollefeys, M., van Gool, L., Vergauwen, M., Verbiest, F., Cornelis, K., Tops, J. and Koch, R., 2004. Visual modeling with a hand-held camera. *International Journal of Computer Vision*, 59(3): 207–232.

Quinlan, J. R., 1987. Simplifying decision trees. *International Journal of Man-Machine Studies*, 27(3): 221–234.

Remondino, F., el-Hakim, S. F., Gruen, A. and Zhang, L., 2008. Turning images into 3-D models. *IEEE Signal Processing Magazine*, 25(4): 55–65.

Rottensteiner, F., Sohn, G., Gerke, M., Wegner, J. D., Breitkopf, U. and Jung, J., 2014. Results of the ISPRS benchmark on urban object detection and 3D building reconstruction. *ISPRS Journal of Photogrammetry and Remote Sensing*, 93: 256–271.

Schindler, K. and Bauer, J., 2003. Towards feature-based building reconstruction from images. *11th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision*, Plzen, Czech Republic. 126–132.

Schmid, C. and Zisserman, A., 1997. Automatic line matching across views. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, San Juan, Puerto Rico. 666–671.

Schmid, C. and Zisserman, A., 2000. The geometry and matching of lines and curves over multiple views. *International Journal of Computer Vision*, 40(3): 199–233.

Suveg, I. and Vosselman, G., 2004. Reconstruction of 3D building models from aerial images and maps. *ISPRS Journal of Photogrammetry and Remote Sensing*, 58(3–4): 202–224.

TARSHA-KURDI, F., LANDES, T. and GRUSSENMEYER, P., 2007. Hough-transform and extended RANSAC algorithms for automatic detection of 3D building roof planes from lidar data. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Systems*, 36(3/W52): 407–412.

WANG, Z., WU, F. and HU, Z., 2009. MSLD: a robust descriptor for line matching. *Pattern Recognition*, 42(5): 941–953.

WERNER, T. and ZISSERMAN, A., 2002. Model selection for automated architectural reconstruction from multiple views. *British Machine Vision Conference*, Cardiff, UK. 53–62.

ZHANG, J., ZHANG, Z., KE, T., ZHANG, Y. and DUAN, Y., 2011. Digital photogrammetry grid (DPGRID) and its application. *ASPRS Annual Conference*, Milwaukee, Wisconsin, USA. 8 pages.

ZHANG, L. and GRUEN, A., 2006. Multi-image matching for DSM generation from IKONOS imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, 60(3): 195–211.

ZHANG, L. and KOCH, R., 2013. An efficient and robust line segment matching approach based on LBD descriptor and pairwise geometric consistency. *Journal of Visual Communication and Image Representation*, 24(7): 794–805.

ZHANG, Y., ZHANG, Z., ZHANG, J. and WU, J., 2005. 3D building modelling with digital map, lidar data and video image sequences. *Photogrammetric Record*, 20(111): 285–302.

ZHANG, Z., 2004. Digital photogrammetry and computer vision. *Geomatics and Information Science of Wuhan University*, 29(12): 1035–1039.

ZHANG, Z., 2007. From digital photogrammetry workstation (DPW) to digital photogrammetry grid (DPGrid). *Geomatics and Information Science of Wuhan University*, 32(7): 565–571.

ZHANG, Z., ZHANG, J., WU, X. and ZHANG, H., 1992. Global image matching with relaxation method. *Proceedings of the International Colloquium on Photogrammetry, Remote Sensing and Geographic Information Systems*, Wuhan, China. 175–188.

ZHANG, Z., ZHANG, J., LIAO, M. and ZHANG, L., 2000. Automatic registration of multi-source imagery based on global image matching. *Photogrammetric Engineering & Remote Sensing*, 66(5): 625–629.

ZHANG, Z. and ZHANG, J., 2004. Generalized point photogrammetry and its application. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 35(B5): 77–81.

ZHANG, Z., ZHANG, Y., ZHANG, J. and ZHANG, H., 2008. Photogrammetric modeling of linear features with generalized point photogrammetry. *Photogrammetric Engineering & Remote Sensing*, 74(9): 1119–1127.

Résumé

Des lignes caractéristiques comme les arêtes concaves et convexes de la façade des bâtiments peuvent être perdues lors de la restitution photogrammétrique. Pour résoudre ce problème et reconstruire de manière automatique et précise des façades 3D quasi-planes, un ensemble d'algorithmes est proposé pour la reconnaissance automatique des lignes et la reconstruction 3D. Il inclut: (1) une procédure pour l'appariement de segments de lignes satisfaisant des exigences spatiales d'indépendance globale et de dépendance locale dans une scène 3D; (2) une technique de compensation par faisceaux généralisée assortie de contraintes sur les lignes (sous la forme d'observations virtuelles) pour contrôler la propagation de l'erreur; et (3) les méthodes d'organisation perceptuelle, calage de plan et intersection plan–plan sont suggérées pour l'acquisition des lignes caractéristiques correspondant aux arêtes concaves et convexes des bâtiments. Les résultats expérimentaux montrent que ces algorithmes sont viables et peuvent être appliqués pour la reconnaissance et la reconstruction 3D. Des recommandations sont exprimées pour les méthodes de reconnaissance, selon que l'on dispose ou non de relations topologiques entre les plans considérés.

Zusammenfassung

Wichtige Linien, wie etwa konkave oder konvexe Kanten einer Gebäudefassade, können bei einer photogrammetrischen Objekterkennung verlorengehen. Um dies zu vermeiden und quasi-planare 3D Fassaden automatisch und exakt zu rekonstruieren wird hierzu ein Satz von Algorithmen und Techniken vorgeschlagen. Darin sind enthalten: (1) eine Prozedur für Liniensegmentzuordnung, die den räumlichen Anforderungen einer 3D Szene genügt und auf "globaler Unabhängigkeit" und "lokaler Abhängigkeit" beruht; (2) einer Technik zur

generalisierten, punktbasierten Bündelblockausgleichung, kombiniert mit räumlichen Linienbedingungen (in Form von virtuellen Beobachtungen), um die Fehlerfortpflanzung zu kontrollieren; und (3) den Methoden zur Wahrnehnumgsorganisation, Ebenenanpassung und Ebenenschnitten, die vorgeschlagen werden, um wichtige Linien, die konkave und konvexe Gebäudekanten darstellen zu erfassen. Die experimentellen Ergebnisse zeigen, dass diese neuen Algorithmen geeignet und anwendbar für die Erkennung und 3D Rekonstruktion sind. Es werden Empfehlungen für Erkennungsmethoden gegeben, in Abhängigkeit davon, ob vorab topologische Beziehungen bei den betrachteten Ebenen verfügbar sind oder nicht.

Resumen

Líneas críticas, tales como aristas cóncavas y convexas de la fachada de un edificio, se pueden perder en los procedimientos de reconocimiento fotogramétricos. Para resolver este problema y reconstruir fachadas 3D casi planas de forma automática y precisa, se proponen un conjunto de algoritmos y técnicas para el reconocimiento automático de líneas y la reconstrucción 3D. Esto incluye: (1) un procedimiento para la adaptación de la línea del segmento que satisfaga las necesidades de espacio de una escena 3D basado en la "independencia global" y la "dependencia local"; (2) un ajuste de bloque generalizado con restricciones lineales (en forma de observaciones virtuales) para controlar la propagación de error; y (3) para adquirir las líneas críticas correspondientes a aristas cóncavas y convexas de construcción se sugieren los métodos de organización perceptual, montaje plano y la intersección de planos. Los resultados experimentales muestran que estos nuevos algoritmos son factibles y aplicables al reconocimiento y la reconstrucción 3D. Se proporcionan recomendaciones para los métodos de reconocimiento en función de la disponibilidad a priori de las relaciones topológicas entre los planos que se consideran.

摘 要

重要的关键线,例如建筑物立面的凹凸边缘,在摄影测量的识别处理中往往会丢失。为了解决这个丢失问题,并且为了自动、精确重建三维拟平面的立面,提出了一套自动识别线段和三维重建的算法和技术,包括:(1)一个满足三维场景空间分布客观规律的、基于"全局独立性"和"局部约束性"的线段匹配处理算法;(2)一个联合空间线约束(采用虚拟观测方程)的广义点光束法平差技术,用以控制误差的传播;(3)一套感知编组、平面拟合和"面-面"交会的方法,用来获取重要的关键线对应的建筑物凹凸边缘。实验结果表明,这些新算法可行且能应用于线特征识别和三维重建。而且,对于是否已知平面之间的先验拓扑关系,给出了不同的识别建议和策略。