ISPRS Journal of Photogrammetry and Remote Sensing 65 (2010) 282–293

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

journal homepage: www.elsevier.com/locate/isprsjprs

An automatic mosaicking method for building facade texture mapping using a monocular close-range image sequence

Zhizhong Kang [a,*], Liqiang Zhang [b], Sisi Zlatanova [c], Jonathan Li [d]

[a] Department of Geodesy and Geomatics, School of Land Science and Technology, China University of Geosciences, 100083 Beijing, China
[b] Research Center for Remote Sensing and GIS, School of Geography, Beijing Normal University, 100875 Beijing, China
[c] OTB Research Institute for Housing, Urban and Mobility Studies, Delft University of Technology, 2628 BX Delft, The Netherlands
[d] Department of Geography, Faculty of Environmental Studies, University of Waterloo, 200 University Avenue West, Waterloo, Ontario, Canada N2L 3G1

ARTICLE INFO

Article history:
Received 3 September 2007
Received in revised form
28 November 2009
Accepted 29 November 2009
Available online 16 December 2009

Keywords:
Terrestrial imaging
Digital camera calibration
Mosaicking
Building facade
Texture mapping

ABSTRACT

This paper presents an automatic mosaicking method for generating building facade textures from a monocular close-range digital image sequence. The process begins with the computation of the camera parameters (except the coordinates of the projective center), which are determined by combining vanishing point geometry with constraints of a straight line bundle as well as prior information of parallel lines in object space. The raw images are later rectified for the purpose of eliminating their salient geometric distortion. Next, automatic retrieval of the relevant image segment is implemented using the detecting range variance by means of the histogram of projective differences between the corresponding points for each of the facades from the raw image sequence. A strip model of the least-squares adjustment, which is similar to the strip block adjustment in aerial triangulation, is employed to determine the spatial alignment of each of the image segments in order to generate the facade textures from the relevant image segments. Afterwards, the entire building facade texture is mosaicked by ortho-image generation. Two refining strategies are proposed to optimize the mosaic result. One is refining the mosaic region where corresponding points are difficult to match but plenty of horizontal lines are available, and the constraint of corresponding horizontal lines is introduced to implement this process. The other strategy is to refine the unsatisfactory mosaic region by densifying the corresponding points by means of the spatial alignment of the relevant image segment computed by the strip method. The experimental results indicate that this method is widely applicable and compares well with other reported approaches with regard to automation level and applicability, for uncalibrated images as well as images with large geometric distortions.

1. Introduction

Three-dimensional (3D) spatial models of urban environments are useful in a variety of applications ranging from urban planning, virtual tourism, and heritage protection to training and simulations for urban disaster and emergency scenarios (e.g., Zlatanova et al., 2005, 2006). A number of algorithms for the extraction of 3D geometric building models have been reported. These algorithms can be categorized as either image- or range-based approaches. Image-based methods (e.g., Grün, 2000; Suveg and Vosselman, 2004) extract the object geometry using stereo or motion techniques. For a comprehensive review of image-based 3D modeling, please refer to Remondino and El-Hakim (2006). Range-based approaches (e.g., Stamos and Allen, 2000; Zhao and Shibasaki, 2003; Dold and Brenner, 2004; Frueh et al., 2005) directly capture the 3D geometric information of an object, based on costly (at least currently) active sensors, and can provide a highly detailed and accurate representation of most shapes. When it comes to visualizing virtual city models, which is a key point in most applications, building texturing is essential for realistic rendering. Digital images acquired by remote sensing techniques from terrestrial, airborne, and spaceborne platforms are currently data sources important for texture mapping. Airborne and satellite imagery has been used for the extraction of texture information of the top surface of 3D cartographic objects (e.g., terrain and buildings). For the production of fly-through or walk-through 3D views to assist environmental planning and disaster management, detailed texture information on building facades is also required and can be acquired by terrestrial imaging techniques. However, the

* Corresponding address: School of Land Science and Technology, China University of Geosciences, 29 Xueyuan Road, Haidian District, 100083 Beijing, China. Tel.: +86 10 8232 2371; fax: +86 10 8232 1807.
E-mail address: zzkang@cugb.edu.cn (Z. Kang).

0924-2716/$ – see front matter
doi:10.1016/j.isprsjprs.2009.11.003

processes of image searching, preprocessing, and mapping are currently still inefficient because of the massive numbers of terrestrial images involved. As a result, the extraction of building facade textures is a prohibitively time-consuming and labor-intensive task. An alternative, fast, and low-cost option is urgently needed and a number of studies towards this problem have been reported.

Brenner and Haala (1998) have proposed an interactive scheme that utilizes projective transformation for the fast production of building facade textures. The parameters of the projective transformation are determined by a minimum number of four points in 3D world coordinates on a plane (in 3D space), along with their corresponding image coordinates. Coorg and Teller (1999) reported a variety of algorithms for extracting vertical facade textures from controlled close-range imagery. This method relies on controlled imagery, i.e., in photogrammetric terms, the exterior orientation of each image is assumed to be known, and was demonstrated on a complex dataset of nearly 4000 high-resolution digital images of a small (200 m^2) office park. Haala and Kada (2005) have presented a method of mapping terrestrial panoramic images to the facades of existing 3D building models. For geometric processing of the panoramic scenes, the exterior orientation is determined from control points, which are measured manually. Rau et al. (2006) proposed an approach which integrates GPS, GIS, and photogrammetry for multi-face texture mapping of a 3D building model. By means of integrating the GIS graphic interface and GPS information about the camera location, a large number of pictures can be managed efficiently. In total, more than 1400 terrestrial photos for 300 polyhedrons were utilized and about 40 man-hours were spent in texture mapping by the reported method.

The acquisition of video is both convenient and rapid. Furthermore, an image sequence can be extracted from video to reconstruct a 3D geometric and texture model of a building efficiently. Zhang et al. (2005) reported an approach for rapidly creating a textured 3D model of a building by combining helicopter-based video, Lidar data, and a 2D vector map. However, there were some limitations due to the helicopter-based video used. One was the high cost for the acquisition of video from the helicopter, and the other was that the weather and the height limit in urban areas tend to result in low resolution and large distortions of facade textures. As opposed to helicopter-based video, ground-level video is usually captured more quickly and cheaply, and the resolution of the textures obtained is higher. Making use of these benefits, Tsai et al. (2006) developed a system to generate and map (near) photo-realistic texture attributes onto 3D building models using terrestrial video sequences. This system required human interaction to identify the initial four tie points (e.g., four roof and ground corners on the image) for photogrammetric corrections, and moreover their coordinates were obtained from digital terrain models and building layout CAD files, which are not always available.

For large and tall buildings, several images are typically required to cover the facade of the entire building due to the limitations of selecting the imaging sensor position for image acquisition in built-up areas. This raises the issue of image mosaics. At present, existing commercial image mosaicking software packages have mainly been developed to process either satellite and aerial images, such as ESRI China (Beijing)'s eYaMosaic and Supresoft's ImageXuite, or digital camera images to create the virtual reality panoramic images, such as Panorama Factory, MGI PhotoVista, and Canon PhotoStitch. All these have satisfactory performance when processing uncalibrated images with small geometric distortion. However, images acquired at the ground level have large obliquity due to the attitude and direction of the digital imaging sensors, which leads to considerable geometric distortion. Therefore, alternative image mosaicking methods that could handle such geometric distortion are needed.

In order to mosaick multiple images to generate complete facade textures, photogrammetric correction techniques are regarded as having addressed the geometric issues by establishing relationships between texture images and models. One way to accomplish this is to acquire the camera parameters with additional equipment (Spann and Kaufman, 2000; Varshosaz, 2004; Luhmann and Tecklenburg, 2005; Frueh et al., 2005) or existing geometric data sources, such as building layout CAD files and digital terrain models (Tsai et al., 2006). A number of algorithms have been developed to reconstruct the original or related camera parameters, projection geometry, and pose, including using correlations of overlapped images (Coorg and Teller, 2000), vanishing points (Caprile and Torre, 1990; Cipolla et al., 1999; Guillou et al., 2000; Lee et al., 2002), or vision-based modeling (Kumar et al., 2000). Another approach for the photogrammetric correction of raw texture images is to register a group of correlated images to a common image space (Kim et al., 2003; Tsai et al., 2005). Additional equipment, or existing geometric data sources, allow adequate accuracy of photogrammetric correction, which can ensure continuity in the geometry during the mosaicking process. However, these are not available at all times.

Unwanted objects, occlusions, or obstacles are often present in the texture data, reducing the realism of the generated 3D models. Therefore, relevant research on occlusion-free texturing has been reported in several studies (Bohm, 2004; Zhang and Kang, 2004; Varshosaz, 2004; Frueh et al., 2005). In addition, several algorithms for building facade interpretation from facade textures, particularly for windows, have been proposed for the purpose of refining 3D facade models (Lee and Nevatia, 2004; Mayer and Reznik, 2007).

This paper is improved and extended from two conference papers (Kang et al., 2007a,b) and presents a novel automatic mosaicking algorithm for building facade texture mapping using a monocular close-range digital image sequence. The image acquisition is optimized for the capture of a ground-level video or image sequence for the purpose of practicability in terms of speed, cost efficiency, flexibility for the scene under investigation, and the texture resolution from the raw data. Camera calibration is of utmost importance for mosaicking, and we therefore implement it in a two-step fashion in order to attain adequate accuracy. Vanishing point geometry is employed to determine the digital camera's parameters (except lens distortion and the coordinates of the projective center) by means of a least squares adjustment from uncontrolled images. Constraints are imposed in this adjustment system, in the form of a straight line bundle as well as prior information of parallel lines in object space. Before mosaicking, we automatically retrieve the relevant image segment, which is implemented via detecting range variance by means of the histogram of the projective differences between corresponding points for each of the facades in the raw image sequence. As mentioned earlier, geometric discontinuities may occur in the mosaicking process after image rectification, and thus the least-squares adjustment inspired by the strip block adjustment in aerial triangulation is introduced in the mosaicking; moreover, two refining strategies are developed in order to eliminate the geometric discontinuity and create a seamless image mosaic. In this stage, the interior parameters are fixed and the angular orientation parameters are restricted to vary within a small range. As a result of the adjustment system for the mosaicking, six exterior orientation parameters ensure the least squares of projective differences between corresponding point pairs. The automation comprises image registration, rectification, retrieving, and mosaicking, and allows high efficiency even in handling a large dataset.

We start by describing the image acquisition process, including the image sensors employed, the platforms used, and the

(a) Dataset (1) Sony IP-7 at 50 m.

(b) Dataset (2) Sony IP-7 at 10 m.

(c) Dataset (3) Kodak Pro at 10 m.

Fig. 1. Acquired image sequences.

requirements for image acquisition (Section 2). Section 3 describes an algorithm for camera calibration based on the vanishing point geometry and constraints derived from it, as well as parallel lines in object space. In Section 4, automatic retrieving of the relevant image segment using only image information from the raw image sequence for each of the facades is described. We then present a strip model of the least-squares adjustment, which is inspired by the strip block adjustment in aerial triangulation, and then present two refining strategies for the purpose of creating a seamless image mosaic. Section 5 describes a variety of image mosaic experiments on large real datasets designed to validate our approach. Section 6 concludes this paper.

2. Data acquisition

In this paper, three sets of image and video sequences are processed: (1) a video sequence obtained by a hand-held digital camcorder (Sony IP-7) on a moving van, with a distance between the photographic center and the corresponding building of about 50 m, as in Fig. 1(a); (2) a video sequence obtained using the hand-held Sony IP-7 on a moving wheelchair to adapt the acquisition technique to the narrow pedestrian street with the distance between the photographic center and the building of interest reduced to 10 m, as in Fig. 1(b); and (3) an image sequence taken by a hand-held digital camera (KODAK PROFESSIONAL DCS Pro SLR/n) along the pedestrian street, as in Fig. 1(c). The image sizes in datasets (1) and (2) are 720 pixels × 480 pixels. In dataset (1), a single image can cover almost the whole facade texture in the Y-axis direction. However, the range of the facade texture covered by a single image becomes much smaller in dataset (2), as the distance between the photographic center and the building decreases from 50 m to 10 m. The image size of dataset (3) is 4500 pixels × 3000 pixels and is enough to cover the whole facade texture in the Y-axis direction with a single image. It is not necessary to process the imagery with such a large image size; therefore, the images in dataset (3) are compressed and the image size is reduced to 1500 pixels × 1000 pixels. To fix the interior parameter while taking the photos, the focal length is set to infinity and the automatic focus function is turned off.

3. Image rectification

For terrestrial imagery, large oblique angles lead to salient geometric distortions of facade textures, and rectification of the raw image is therefore necessary to acquire a visually appealing texture. The key to image rectification is how to compute camera

parameters (except the coordinates of the projective center). The algorithm presented by Kang and Tan (2005), which uses curvature variance, is employed to calibrate for lens distortion. This algorithm is independent of the focal length and uses the straight lines in the image, which are not necessarily parallel. This section proposes an algorithm based on vanishing point geometry and the geometric constraints derived from prior information of the parallel object lines in order to compute the other camera parameters (except the coordinates of the projective center) from the uncontrolled images. Among these camera parameters, the angular orientation parameters will be refined and the coordinates of the projective center will be computed in the next section. Vanishing point geometry is employed by means of a least squares adjustment. Constraints are imposed on the adjustment system in the form of a straight line bundle, as well as prior information of parallel lines in object space, to obtain high accuracy. Although this algorithm was briefly introduced in Kang et al. (2007a), comprehensive details are presented here, and all formulas used in the method are deduced.

3.1. The computation of vanishing point

Straight lines are extracted from the image by means of the Log calculator, and those lines parallel to the X- and Y-axis in object space are later used to compute the two vanishing points. Vanishing point $V(x_V, y_V)$ is the intersection point of a converging set of straight lines which are parallel with each other in object space. Accordingly, each straight line, e.g., line ij, belonging to this set should pass through the vanishing point, which results in Eq. (1) as follows:

$$d = (y_j - y_i)\frac{(x_V - x_i)}{s_{iV}} - (x_j - x_i)\frac{(y_V - y_i)}{s_{iV}} \tag{1}$$

where d is the distance of point j to line iV, and s_{iV} is the distance between point i and vanishing point v.

Afterwards, we determine two vanishing points utilizing an adjustment system formed using Eq. (1), and from these two points the third vanishing point, as well as the initial values of the digital camera (DC) angular parameters and the focal length (Caprile and Torre, 1990).

3.2. Calibration of camera parameters

Fig. 2 illustrates the geometry between the three vanishing points in the orthogonal directions and the image orientation parameters, which are drawn from the vanishing point as well as the pinhole projection geometry of terrestrial imagery and can be parameterized as:

$$\left.\begin{aligned} x_{X\infty} &= x_0 + f\cot\varphi\sec\omega\cos\kappa - f\tan\omega\sin\kappa \\ y_{X\infty} &= y_0 - f\cot\varphi\sec\omega\sin\kappa - f\tan\omega\cos\kappa \\ x_{Y\infty} &= x_0 + f\sec\omega\csc\omega\sin\kappa - f\tan\omega\sin\kappa \\ y_{Y\infty} &= y_0 + f\sec\omega\csc\omega\cos\kappa - f\tan\omega\cos\kappa \\ x_{Z\infty} &= x_0 - f\tan\varphi\sec\omega\cos\kappa - f\tan\omega\sin\kappa \\ y_{Z\infty} &= y_0 + f\tan\varphi\sec\omega\sin\kappa - f\tan\omega\cos\kappa \end{aligned}\right\} \tag{2}$$

where $(x_{X_\infty}, y_{X_\infty})$, $(x_{Y_\infty}, y_{Y_\infty})$ and $(x_{Z_\infty}, y_{Z_\infty})$ denote the coordinates of the vanishing points of the three orthogonal coordinate axes, respectively, f is the focal length, (x_0, y_0) is the principal point coordinates, and φ, ω, k are the three angular orientation parameters.

Substitution of the relevant terms of Eq. (1) into these completes the observation equations used to compute the camera parameters, except for lens distortion and the coordinates of the projective center. A model of the adjustment system of observations and parameters is later formed utilizing the constraint of a straight lines bundle to acquire suitable accuracy. As the initial values of the relevant parameters are decomposed from the vanishing points, the accuracy is not always high, which can be attributed to the possible low quality of the images and errors of line extraction. As a result, this adjustment system has a small convergent radius and is not robust.

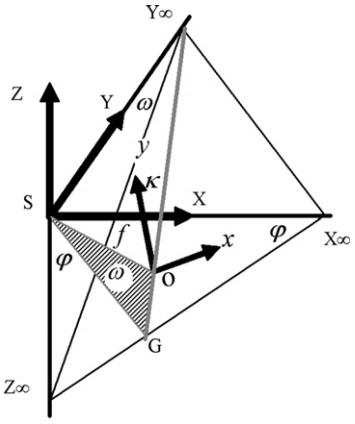


Fig. 2. Geometry between the vanishing points and orientation parameters.

To address this problem, a constraint is imposed in this adjustment system, in the form of *a priori* information of parallel lines in object space, in order to strengthen the control over the parameter computation. In the following, we describe this constraint in more detail.

Since two groups of lines parallel to the X- and Y-axes in object space are generally available on building facades, the following conditions can be derived as follows:

In the X direction:

$$-f\frac{b_1x_i + b_2y_i - b_3f}{c_1x_i + c_2y_i - c_3f} = -f\frac{b_1x_j + b_2y_j - b_3f}{c_1x_j + c_2y_j - c_3f}. \tag{3}$$

In the Y direction:

$$-f\frac{a_1x_m + a_2y_m - a_3f}{c_1x_m + c_2y_m - c_3f} = -f\frac{a_1x_n + a_2y_n - a_3f}{c_1x_n + c_2y_n - c_3f}, \tag{4}$$

where:

(x_i, y_i), (x_j, y_j) are the coordinates of the two end points of a line parallel to the X-axis,

(x_m, y_m), (x_n, y_n) are the coordinates of the two end points of a line parallel to the Y-axis,

f is the focal length, and

$\begin{bmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{bmatrix}$ denotes the rotation matrix.

Eqs. (3) and (4) can be control conditions for the calculation of the orientation parameters, and from them a model of the adjustment system combined with the constraints of the straight lines bundle and prior information of parallel lines in object space is deduced. As imposed by the rigid geometric constraints, this adjustment system has the characteristics of a large convergent radius and high stability.

As a result of the above least squares adjustment, the orientation parameters (except the coordinates of the projective center) are used to rectify the terrestrial image, as shown in Fig. 3. The extracted lines are automatically grouped into horizontal and vertical lines in object space using an algorithm that combines the angle histogram with the geometric constraint of the normal vector to the interpretation plane proposed by Kang and Tan (2005).

4. The automatic mosaic of facade texture

As mentioned in the Section 1, factors such as image quality, resolution, geometric distortion, feature extraction and correspondence error, and deficiencies in the mathematic model can degrade the accuracy of photogrammetric correction, which may lead to geometric discontinuities in the mosaicking. As seen in

Fig. 3. Rectified image.

Fig. 1(a), the image sequence is taken along a linear route with the distance between digital camera and the facade of interest being relatively long (about 50 m), so we can safely assume that the difference in the Z coordinates between neighboring photographic stations is very small. In addition, the focal length is set to infinity and the automatic focus function is turned off during the image acquisition. As a result, after the facade textures are rectified onto the plumb plane, the geometric discrepancies between adjacent rectified images appear very small. Zhang and Kang (2004) presented a strategy combining the correlation coefficient with geometric constraints between neighboring images to acquire a mosaic point pair, by which only the displacements d_x and d_y in the X-axis and the Y-axis, respectively, are determined to complete the mosaicking process. This method performs well on relatively distant terrestrial imagery, but not if the image is taken too close to the facade (about 10 m) (Fig. 1(b) and (c)), as the geometric discontinuity cannot be eliminated perfectly by the rectification process. Moreover, the shorter the distance is between the photographic station and the building facade of interest, the smaller the range of facade texture covered by a single image is, so the whole facade texture in the Y-axis direction cannot be covered by a single image. Hence, the mosaic process is required in both the X-axis and Y-axis directions to acquire the whole facade texture. Obviously, the method of selecting the mosaic point pair is not applicable in this situation. Therefore, a strip model of the least-squares adjustment, which is inspired by the strip block adjustment in aerial triangulation, is employed to determine the spatial alignment of each of the image segments. Afterwards, the entire building facade texture is mosaicked by the method of ortho-image generation. While this mosaicking strategy was briefly presented in Kang et al. (2007a), relevant models are parameterized in this paper. Furthermore, two new refining strategies are proposed in order to eliminate the geometric discontinuity and obtain a seamless image mosaic. Before mosaicking, we automatically retrieve the relevant image segment, which uses only the image information. More details about this set of algorithms are given below.

4.1. Automatic image retrieving

To recover a building facade texture efficiently, the raw image sequence should be automatically divided into segments corresponding to each building facade. Rau et al. (2006) and Varshosaz (2004) managed terrestrial images by resorting to additional instruments, such as a GPS or a motor-driven theodolite. Unfortunately, those additional instruments are not available in most cases. We use the method presented in Kang et al. (2007a) to automatically retrieve corresponding images for each building facade from the raw image sequence. This is done via the detecting range variance by means of the histogram of projective differences between corresponding points for each of the facades from the raw image sequence.

4.2. Automatic mosaicking

The raw image sequence is divided into a number of image segments, each of which corresponds to a building facade with a dominant plane. Therefore, the spatial alignment of each of the image segments will be recovered when the corresponding rays of points on each image intersects on the projective plane by adjusting the exterior orientation parameters (EOPs) of each image. The method of generating an ortho-image is later employed to create the entire facade texture according to this spatial alignment.

First, the projective relationship can be parameterized as follows based on the projective formula of an image point on the projective plane:

$$\left.\begin{aligned} &-(f+Z_{S1})\frac{a_{11}x_1+a_{12}y_1-a_{13}f}{c_{11}x_1+c_{12}y_1-c_{13}f}+X_{S1}\\ &\quad=-(f+Z_{S2})\frac{a_{21}x_2+a_{22}y_2-a_{23}f}{c_{21}x_2+c_{22}y_2-c_{23}f}+X_{S2}\\ &-(f+Z_{S1})\frac{b_{11}x_1+b_{12}y_1-b_{13}f}{c_{11}x_1+c_{12}y_1-c_{13}f}+Y_{S1}\\ &\quad=-(f+Z_{S2})\frac{b_{21}x_2+b_{22}y_2-b_{23}f}{c_{21}x_2+c_{22}y_2-c_{23}f}+Y_{S2} \end{aligned}\right\} \tag{5}$$

where

f is the focal length,

(X_{Si}, Y_{Si}, Z_{Si}) are the coordinates of the projective center S_i,

$\begin{bmatrix} a_{i1} & a_{i2} & a_{i3} \\ b_{i1} & b_{i2} & b_{i3} \\ c_{i1} & c_{i2} & c_{i3} \end{bmatrix}$ denotes the rotation matrix of image p_i, and

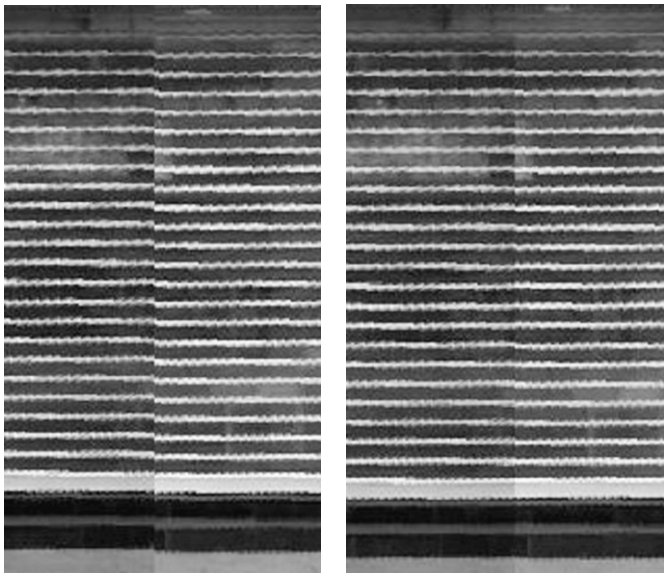
(x_i, y_i) denote the coordinate of image point a_i $(i = 1, 2)$.

Eq. (5) depicts only an ideal instance because of the existence of orientation parameter error, which results in projective differences between projections of the corresponding point pairs on the projective plane. To calculate this projective difference, Eq. (5) is transformed to (6):

$$\left.\begin{aligned} \Delta X &= -(f+Z_{S1})\frac{a_{11}x_1+a_{12}y_1-a_{13}f}{c_{11}x_1+c_{12}y_1-c_{13}f}+X_{S1}\\ &\quad+(f+Z_{S2})\frac{a_{21}x_2+a_{22}y_2-a_{23}f}{c_{21}x_2+c_{22}y_2-c_{23}f}-X_{S2}\\ \Delta Y &= -(f+Z_{S1})\frac{b_{11}x_1+b_{12}y_1-b_{13}f}{c_{11}x_1+c_{12}y_1-c_{13}f}+Y_{S1}\\ &\quad+(f+Z_{S2})\frac{b_{21}x_2+b_{22}y_2-b_{23}f}{c_{21}x_2+c_{22}y_2-c_{23}f}-Y_{S2} \end{aligned}\right\}. \tag{6}$$

According to the theory of optical projection, the corresponding rays of each image can intersect each other on the projective plane by adjusting the EOPs of each image. Namely, the least-squares adjustment system formed on the basis of Eq. (6) is employed to calculate the EOPs, which makes all the coordinate differences between corresponding projective points least squares. The mathematic model is described in more detail below.

In this adjustment system, the interior orientation parameters are regarded as known parameters, and the angular orientation parameters are restricted to vary in a small range, as they were calibrated in Section 3. Therefore, each image has six unknown parameters, i.e., exterior orientation parameters. We introduce the following formula depicting the displacement of projecting points



(a) Before refining. (b) After refining.

Fig. 4. The refined result using the constraint of corresponding horizontal lines.

on the projective plane caused by the small angle (Konecny and Lehmann, 1984) into the adjustment system for the purpose of model simplicity:

$$\left.\begin{aligned} \mathrm{d}X &= \mathrm{d}b_x + \frac{X}{H}\mathrm{d}b_z + \left(H + \frac{X^2}{H}\right)\mathrm{d}\varphi + \frac{XY}{H}\mathrm{d}\omega - Y\mathrm{d}\kappa \\ \mathrm{d}Y &= \mathrm{d}b_y + \frac{Y}{H}\mathrm{d}b_z + \frac{XY}{H}\mathrm{d}\varphi + \left(H + \frac{Y^2}{H}\right)\mathrm{d}\omega + X\mathrm{d}\kappa \end{aligned}\right\}. \quad (7)$$

Then, we have the observation equation as:

$$\left.\begin{aligned} v_{\Delta X} &= \mathrm{d}X_1 - \mathrm{d}X_2 - (X_1^0 - X_2^0) \\ v_{\Delta Y} &= \mathrm{d}Y_1 - \mathrm{d}Y_2 - (Y_1^0 - Y_2^0) \end{aligned}\right\} \quad (8)$$

where:

$$\mathrm{d}X_i = \mathrm{d}b_{xi} + \frac{X_i^0}{H}\mathrm{d}b_{zi} + \left(H + \frac{X_i^{0^2}}{H}\right)\mathrm{d}\varphi_i + \frac{X_i^0 Y_i^0}{H}\mathrm{d}\omega_i - Y_i^0\mathrm{d}\kappa_i$$

$$\mathrm{d}Y_i = \mathrm{d}b_{yi} + \frac{Y_i^0}{H}\mathrm{d}b_{zi} + \frac{X_i^0 Y_i^0}{H}\mathrm{d}\varphi_i + \left(H + \frac{Y_i^{0^2}}{H}\right)\mathrm{d}\omega_i + X_i^0\mathrm{d}\kappa_i$$

$$X_i^0 = -(f + Z_{Si})\frac{a_{i1}x_i + a_{i2}y_i - a_{i3}f}{c_{i1}x_i + c_{i2}y_i - c_{i3}f} + X_{Si}$$

$$Y_i^0 = -(f + Z_{Si})\frac{b_{i1}x_i + b_{i2}y_i - b_{i3}f}{c_{i1}x_i + c_{i2}y_i - c_{i3}f} + Y_{Si}.$$

Finally, the spatial alignment of each image segment is determined by the least squares adjustment system presented above, by which the method of generating the ortho-image is used to make the whole facade texture.

4.3. Refining mosaic result

As seen in Fig. 4(a), a geometric discontinuity appears because the observation values in the least squares adjustment system for mosaicking are corresponding point coordinates, yet few corresponding points have been tracked. To address this problem, two refining strategies are proposed. One involves refining the mosaic region where corresponding points are difficult to match, but where many of horizontal lines are available. The constraint of corresponding horizontal lines is introduced to implement this process. The other strategy is refining the mosaic region by densifying the corresponding points using the spatial alignment of a relevant image segment.

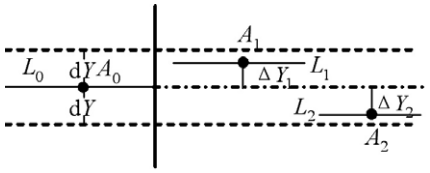


Fig. 5. The included range of corresponding lines.

4.3.1. The constraint of corresponding horizontal lines

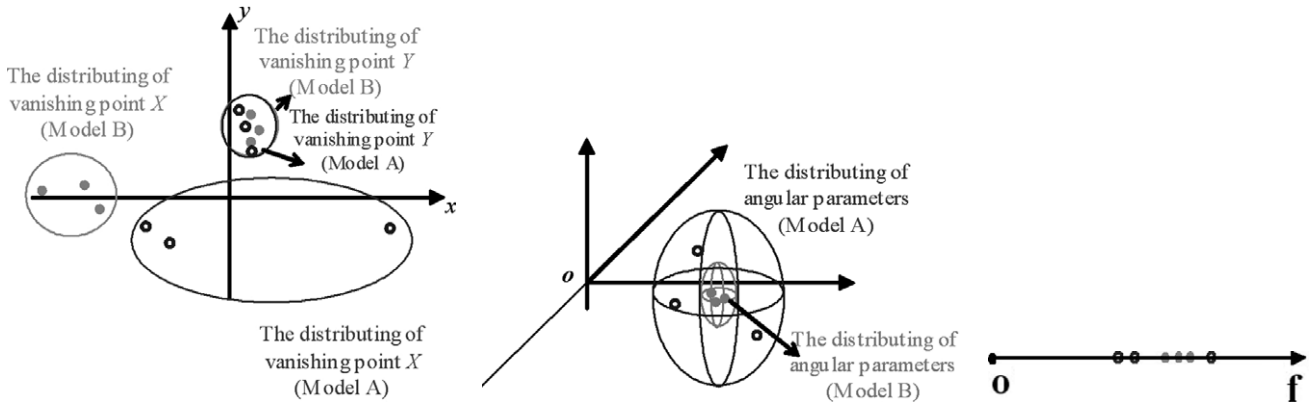
As Fig. 4 illustrates, although no corresponding points are tracked in this area, a number of horizontal lines are available, and thus the question becomes how to match corresponding lines for this case. First, those lines are projected onto the mosaic image. We introduce the so-called included range of the corresponding line for every horizontal line in the left image. As Fig. 5 shows, the included range is a horizontal strip whose center is line L_0 and whose width is twice $\mathrm{d}Y$ in the Y component of the projective difference of the corresponding points near line L_0. Since the lines in Fig. 5 are horizontal, the center A_0 of line L_0 is used to determine the corresponding lines. Therefore, the lines in the right image should be corresponding lines if their center points (A_1 and A_2) are within the included range of L_0. As we can see, the projective differences ΔY_1 and ΔY_2 between line L_0 and L_1, L_2 are equal to the Y coordinate difference between the centers A_0 and A_1, A_2. According to this, the refining strategy for the mosaic result is to minimize ΔY by least squares, which can be added to the least squares adjustment for mosaicking as additional observations. The mosaic result after the refining process is illustrated in Fig. 4(b). The geometric discontinuity disappears after imposing the constraint of corresponding horizontal lines.

4.3.2. Densifying the corresponding points

Fig. 6(a) illustrates another case of a geometric discontinuity due to insufficient corresponding points in an area. This can be attributed to a large geometric distortion of the building superstructure in the raw image, which leads to the difficulty of matching corresponding points on the raw images. Intuitively, the best way to refine the image is to densify the corresponding points. As presented in Section 4.2, the projective difference between corresponding points in the mosaic image becomes very small

(a) Before refining. (b) After refining.

Fig. 6. The refined result by densifying corresponding points.



(a) Vanishing point. (b) Angular orientation parameters. (c) Focal length.

Fig. 7. Convergence distribution (the open symbols denote model A and the solid symbols denote model B).

after the least squares adjustment, which makes densifying the corresponding points in the mosaic image possible.

The region which needs to be densified should be determined first. To do this, a grid is overlaid on the image; the size of the grid depends on the size of the image. Since the image size in Fig. 6 is 1000 pixels × 1500 pixels, a 20 × 20 grid is overlaid so each grid cell size contains 50 pixels × 75 pixels. Those grid cells overlaid in the region of the sky are discarded because it is useless to densify those points. Those grid cells containing few corresponding points should be the regions requiring the densification. Afterwards, the corresponding point tracking process is re-implemented in these regions. The projective differences between the existing corresponding points near or in these regions are used to forecast the positions of potential corresponding points so that more corresponding points can be matched, as in Fig. 6(b). Finally, those corresponding points are added to the least squares adjustment system. After this refining process, a seamless mosaic result is acquired, as illustrated in Fig. 6(b).

5. Experimental results

In this study, experiments were conducted using the three datasets introduced in Section 2 to illustrate the feasibility and the robustness of the developed automatic mosaic algorithms. The images were captured at different times to cover different building facades and exhibit varying geometric distortions.

5.1. Image rectification

The key to rectification is the digital camera calibration, and thus two calibration algorithms of the DC angular parameters were tested on the image sequence. To compare the differences of convergent radius and stability between the model employing the constraint of the straight lines bundle (model A) and the model using the constraint of prior information of parallel lines in object space (model B), the least squares adjustments were implemented on the same image using the results of grouping lines by the three methods: (1) grouping lines by the angle histogram; (2) the angle histogram with the geometric constraint of the normal vector to the interpretation plane (Kang and Tan, 2005); and (3) the initial value acquired manually.

The distribution of the vanishing point coordinates, angular parameters, and focus is shown in Fig. 7 based on the results of the least squares adjustment. The distribution illustrates that model B has a larger convergent radius and higher stability than model A; namely, the values acquired by model B are generally optimal, while those acquired by model A are locally optimal.

5.2. Automatic mosaic experiment

5.2.1. Automatic image retrieving

The experiment was implemented by using the method presented in Kang et al. (2007a) on a 146 m long segment of

(a) ImageSuite.

(b) PhotoStitch.

Fig. 8. Mosaic of dataset (1) obtained using existing software.

Fig. 9. Refined facade texture.

street facade comprising two cross-roads and four connected buildings. Fifty-six images were acquired to cover this segment. The histograms of the projective differences of the corresponding points were generated to track the large range variance from those images. The positions of the peak values exactly indicate the positions in the raw images of the two cross-roads and the one building, which has a large range difference with the others. According to the movement of each facade determined by tracking the large range variance, two image segments were selected from the 58 images and automatically corresponded to the buildings of interest.

5.2.2. Mosaicking

For comparison, image mosaics were produced from the close-range digital camera images using both existing image mosaic software and the algorithm we have presented. First, using two existing software packages, we processed datasets (1) and (3). The results for dataset (1) are shown in Fig. 8. The existing software packages are not applicable for processing images with a large obliquity of the camera posture, because a typical 2D mosaic usually chooses a reference image from the collection without rectifying it and all other images can be warped in the 2D coordinate system of this image by the corresponding homographies.

Since dataset (1) was acquired about 50 m away from the buildings, the difference of the Z coordinates between each photographic station was very small compared to the distance between the photographic station and the building. Therefore, after the rectification process, a mosaic point pair was selected by combining with the correlation coefficient and the geometric constraint (Zhang and Kang, 2004) to implement the mosaic process for dataset (1). The mosaic result shows that the strict control imposed on selecting the mosaic point pair gave a positive result. The more refined facade texture after the texture analysis process is shown in Fig. 9.

Dataset (2) was acquired on a narrow street; therefore, the distance between the photographic station and the building was only about 10 m. As a result, the oblique angle of the image was larger, as was the geometric distortion of the image. First, the mosaic process was implemented for dataset (2) using the same method as for dataset (1). Unfortunately, Fig. 10(a) illustrates that obvious geometric discontinuities appear. These discontinuities are due to nonlinear geometric distortion that remained after rectification, which was not eliminated by the displacements dx and dy in the X-axis and Y-axis directions. The strip method was employed to remove the remaining nonlinear geometric distortion so that the geometric discontinuity disappeared, as shown in Fig. 10(b).

Because the Sony IP-7 camcorder has a small field of view (FOV), the size of the extracted images is only 720 pixels $\times$ 560 pixels. At the projective center near the facade, the texture covered by a single image is small. As a result, both the right-and-left adjacent and up-and-down adjacent images need to be mosaicked to generate the whole texture. Because the strip method is inspired by the strip block adjustment, it cannot only mosaic single strip images, but also multiple strip images. Therefore, the mosaic processes were implemented on both of the single and multiple strip images of dataset (2). The mosaic results of

(a) Select mosaic point pair. (b) Strip method.

Fig. 10. The comparison of mosaic results.

three single strips (top, middle, and bottom) and the three-strip images are illustrated in Figs. 11 and 12, respectively. The results show that the strip method allows us to acquire the entire texture from small FOV images by mosaicking multiple strips of images simultaneously. However small geometric discontinuities still remain in the mosaic image after the least squares adjustment, which makes the mosaicking imperfect. This can be attributed to the low resolution as well as the small FOV, which degrades the accuracy of the camera calibration. Furthermore, the small image size leads to a large number of images being necessary to cover the facades. In dataset (2), 238 images were used to cover a building facade with a length of 82 m. On the computer used to process the images (CPU: Intel Pentium 4 1.6 GHz, RAM: 256MB DDR), the processing time for image rectification was 39 min. In the strip method adjustment system, each image has six unknown parameters; therefore, there are 1428 unknown parameters in total. The time for the adjustment was one minute. Although the process is fully automatic, 40 min in total is too long to process an image mosaic for one building facade.

To solve this problem of low efficiency, it is best to use a camcorder with a large FOV. Presently, high definition camcorders are available in the market, e.g., the SONY HDR-UX1, which has an image size of 1920 pixels × 1080 pixels. Since this kind of camcorder was not available, a KODAK PROFESSIONAL DCS Pro SLR/n digital camera with a wide-angle lens was used. To cover the whole facade texture in the Y-direction with the least number of images, the camera was used in the upright position. The image size was 3000 pixels × 4500 pixels. The raw images were compressed to 1000 pixels × 1500 pixels to reduce the processing time because the resolution of 1.5 million pixels is sufficient to generate the facade texture. Dataset (3) was acquired in this manner. In dataset (3), a building facade 82 m long was covered by a single strip of 36 images, which is much less than in dataset (2). Since the image size of dataset (3) is 4.3 times that of dataset (2), the processing time per image in dataset (3) was longer, by 30 s. As a result, the total processing time was 18 min. There are only 216 unknown parameters in the adjustment system, so the time for adjustment is reduced to 10 s. Compared to dataset (2), the efficiency is highly improved. Moreover, as shown in Fig. 13, the facade texture generated from the 36 images is much more vivid and closer to a seamless mosaic compared to the result shown in Fig. 12.

The site of the experiment was a pedestrian street situated in Wuhan, China, and comprises 28 building facades. The image mosaics for those facades were mapped into a 3D model of this street scene as shown in Fig. 14. As terrestrial close-ranged images have high resolution and are acquired along the street, the walk-through street scene is close to the real view from the street.

6. Conclusions

The experimental results indicate that the proposed automatic mosaicking method is effective and reliable for building facade texture mapping. The experiments of image rectification on uncalibrated images show that the model of least squares adjustment using only the constraint of the straight lines bundle has a small convergence radius and is sensitive to the initial value, which tends to be affected by the quality of the image and the errors of line extraction. After imposing the constraint of prior information of parallel lines in object space, the adjustment model shows a larger convergence radius and higher stability in the calibration of the DC parameters. The histograms of projective differences of the corresponding points showed good performance in automatically retrieving image segments which correspond to

(a) Lower strip.

(b) Middle strip.

(c) Upper strip.

Fig. 11. Mosaic results of single strip.

Fig. 12. Mosaic of three strip images.

each building facade from the raw image sequence. Three datasets were used to test the presented mosaic algorithms, both on video and digital camera image sequences. The results showed that in addition to generating a seamless mosaic, the strip method allows us to acquire the whole texture from large FOV images via single strip images, and also from small FOV images by mosaicking multiple strips simultaneously.

The performed tests revealed several advantages.

- Although the image sequences are acquired at ground level, which makes existing methods of recovering building facade textures inefficient, our method is implemented from data acquisition to image mosaic, in a rapid and highly efficient way based only on image information.

Fig. 13. Whole facade texture.

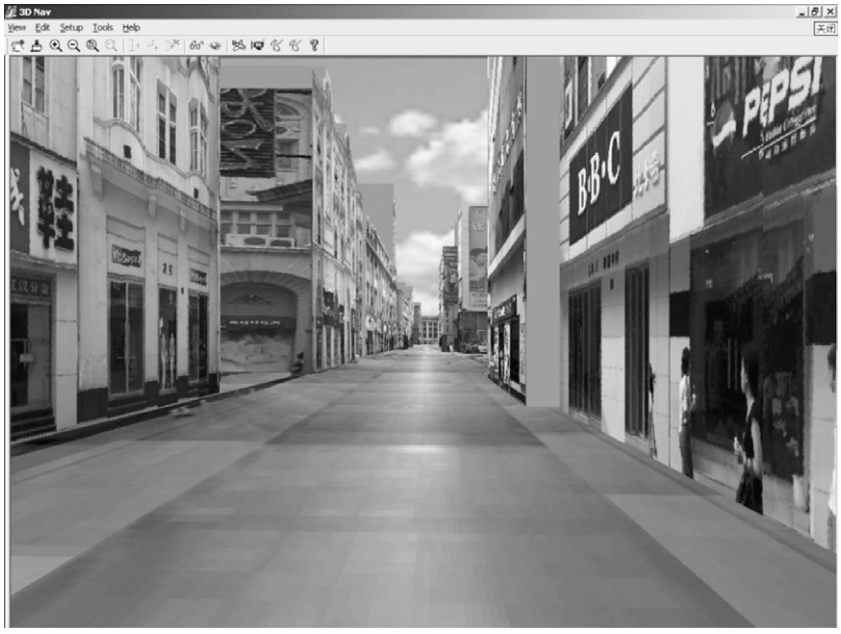


Fig. 14. Image mosaics mapped into the 3D street model.

- Terrestrial close-ranged images have high resolution and are taken along the street; therefore, a realistic textured walk-through 3D scene can be reconstructed.

The approach presented in this paper is applicable for urban streets and exhibits high efficiency in built-up areas. At this stage, it can handle only roughly planar facades, which are common in urban areas. This approach is appropriate for the rapid reconstruction of realistic 3D visualization that can be used in a variety of applications, e.g., urban planning, 3D navigation, virtual tourism, and 3D games. Future work will concentrate on the image mosaic for more complex (non-planar) and high-rise facades and the improvement of availability.

Acknowledgements

The comments of anonymous reviewers for improvement of this paper are gratefully acknowledged. This research was supported by the Natural Science Foundation of China under Grant No. 40801191 and the National High Technology Research and Development Program of China (863 Program) with the serial number 2006AA12Z220.

References

Bohm, J., 2004. Multi-image fusion for occlusion-free facade texturing. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences 35 (Part 5), 867–872.

Brenner, C., Haala, N., 1998. Fast production of virtual reality city models. International Archives of Photogrammetry and Remote Sensing 32 (Part 4), 77–84.

Caprile, B., Torre, V., 1990. Using vanishing points for camera calibration. International Journal of Computer Vision 4 (2), 127–140.

Cipolla, R., Drummond, T., Robertson, D.P., 1999. Camera calibration from vanishing points in images of architectural scenes. In: Proceedings of the 10th British Machine Vision Conference, BMVC 99, Nottingham, UK, 13–16 September, vol. 2, pp. 382–391.

Coorg, S., Teller, S., 1999. Extracting textured vertical facades from controlled close-range imagery. In: Proceedings of Computer Vision and Pattern Recognition Conference, CVPR '99, Fort Collins, Colorado, USA, 23–25 June, pp. 625–632.

Coorg, S., Teller, S., 2000. Spherical mosaics with quaternions and dense correlation. International Journal of Computer Vision 37 (3), 259–273.

Dold, C., Brenner, C., 2004. Automatic matching of terrestrial scan data as a basis for the generation of detailed 3D city models. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences 35 (Part B3), 1091–1096.

Frueh, C., Jain, S., Zakhor, A., 2005. Data processing algorithms for generating textured 3D building facade meshes from laser scans and camera images. International Journal of Computer Vision 61 (2), 159–184.

Grün, A., 2000. Semi-automated approaches to site recording and modelling. International Archives of Photogrammetry and Remote Sensing 33 (Part 5/1), 309–318.

Guillou, E., Meneveaux, D., Maisel, E., Bouatouch, K., 2000. Using vanishing points for camera calibration and coarse 3D reconstruction from a single image. The Visual Computer 16 (7), 396–410.

Haala, N., Kada, M., 2005. Panoramic scenes for texture mapping of 3D city models. International Archives of Photogrammetry and Remote Sensing 36 (Part 5/W8) (on CD-ROM).

Kang, Z., Tan, Y., 2005. Lens distortion calibration based on curvature variance. Geomatics and Information Science of Wuhan University 31 (9), 777–780.

Kang, Z., Zhang, Z., Zhang, J., Zlatanova, S., 2007a. Rapidly realizing 3D visualization for urban street based on multi-source data integration. In: Li, J., Zlatanova, S., Fabbri, A. (Eds.), Geomatics Solutions for Disaster Management. In: Lecture Notes in Geoinformation and Cartography (LNG & C), Springer-Verlag, Berlin, pp. 149–163.

Kang, Z., Zhang, L., Zlatanova, S., 2007b. An automatic mosaicking for building facade texture mapping. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences 36 (Part 4/W45) (on CD-ROM).

Kim, D.H., Yoon, Y.I., Choi, J.S., 2003. An efficient method to build panoramic image mosaics. Recognition Letters 24 (14), 2421–2429.
Konecny, G., Lehmann, G., 1984. Photogrammetrie. Walter de Gruyter, Berlin, pp. 218–221.
Kumar, R., Sawhney, H.S., Guo, Y., Hsu, S., Samarasekera, S., 2000. 3D manipulation of motion imagery. In: Proceedings of International Conference on Image Processing, Vancouver, BC, Canada, 10–13 September, vol. 1, pp. 17–20.
Lee, S.C., Jung, S.K., Nevatia, R., 2002. Integrating ground and aerial views for urban site modeling. In: Proceedings of the 16th International Conference on Pattern Recognition, ICPR'02, Québec City, QC, Canada, 11–15 August, vol. 4, pp. 107–112.
Lee, S.C., Nevatia, R., 2004. Extraction and integration of window in a 3D building model from ground view images. In: Proceedings of 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR'04, Washington, DC, USA, 27 June–2 July, vol. 2, pp. 113–120.
Luhmann, T., Tecklenburg, W., 2005. High-resolution image rectification and mosaicing—A comparison between panorama camera and digital camera. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences 36 (Part 5/W8) (on CD-ROM).
Mayer, H., Reznik, S., 2007. Building facade interpretation from uncalibrated wide-baseline image sequences. ISPRS Journal of Photogrammetry and Remote Sensing 61 (6), 371–380.
Rau, J.Y., Teo, T.A., Chen, L.C., Tsai, F., Hsiao, K.H., Hsu, W.C., 2006. Integration of GPS, GIS and photogrammetry for texture mapping in photo-realistic City modeling. In: Lecture Notes in Computer Science, vol. 4319. pp. 1283–1292.
Remondino, F., El-Hakim, S., 2006. Image-based 3D modelling: A review. The Photogrammetric Record 21 (115), 269–291.
Spann, J.R., Kaufman, K.S., 2000. Photogrammetry using 3D graphics and projective textures. International Archives of Photogrammetry and Remote Sensing 33 (Part B5/1), 748–755.
Stamos, I., Allen, P.K., 2000. 3D model construction using range and image data. In: Proceedings of 2000 Conference on Computer Vision and Pattern Recognition, CVPR 2000, Hilton Head, SC, USA, 13–15 June, pp. 531–536.
Suveg, I., 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.
Tsai, F., Lin, H.C., Liu, J.K., Hsiao, K.H., 2005. Semiautomatic texture generation and transformation for cyber city building models. In: Proceedings of International Geoscience and Remote Sensing Symposium, IGARSS 05, Seoul, Korea, 25–29 July, vol. 7, pp. 4980–4983.
Tsai, F., Chen, C.H., Liu, J.K., Hsiao, K.H., 2006. Texture generation and mapping using video sequences for 3D building models. In: Abdul-Rahman, A., Zlatanova, S., Coors, V. (Eds.), Innovations in 3D Geo Information Systems. In: Lecture Notes in Geoinformation and Cartography (LNG & C), Springer-Verlag, Berlin, pp. 429–438. Part 6.
Varshosaz, M., 2004. Occlusion-free 3D realistic modeling of buildings in urban areas. International Archives of Photogrammetry and Remote Sensing 35 (Part B4), 437–442.
Zhang, Y., Zhang, Z., Zhang, J., Wu, J., 2005. 3D building modelling with digital map, lidar data and video image sequences. The Photogrammetric Record 20 (111), 285–302.
Zhang, Z., Kang, Z., 2004. The rendering of building texture from land-based video. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences 35 (Part B3), 732–738.
Zhao, H., Shibasaki, R., 2003. Special issue on computer vision system: Reconstructing textured CAD model of urban environment using vehicle-borne laser range scanners and line cameras. Machine Vision and Applications 14 (1), 35–41.
Zlatanova, S., Fabbri, A., Li, J., 2005. Geo-information for disaster management: Large scale 3D data needed by Urban Areas. GIM International 19 (3), 10–13.
Zlatanova, S., van Oosterom, P., Verbree, E., 2006. Geo-information supports management of urban disasters. Open House International 31 (1), 62–79.