Reconstruction of structured scenes from two uncalibrated images

Guanghui Wang a,b,*, Hung-Tat Tsui a, Zhanyi Hu b

a Department of Electronic Engineering, The Chinese University of Hong Kong, Shatin, NT, Hong Kong
b National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing 100080, PR China

Received 2 August 2004

Abstract

This paper investigates a practical heuristics method for reconstruction of structured scenes from two uncalibrated images. The method is based on an initial estimation of principal homographies from 2D initial point matches that may contain some outliers, and the homographies are refined recursively by incorporating the supporting matches of both points and lines on principal space surfaces. Then epipolar geometry is recovered directly from the refined homographies and cameras are calibrated from three orthogonal vanishing points and the recovered infinite homography. There are several points of novelty. First, a simple homography-guided method for matching line segments between two views is proposed. Second, under the assumption of zero-skew, the cameras are auto-calibrated with all the four intrinsic parameters varying between the two views. The advantages of the method is that it can build more realistic models with minimal human interactions, and it also allows us to reconstruct more visible surfaces on the detected planes than traditional methods which can only reconstruct the overlapping parts, since the homography provides a one-to-one mapping of points and lines between different views. Extensive experiments with real images illustrate the validity and advantages of the proposed method.

© 2004 Elsevier B.V. All rights reserved.

Keywords: Plane induced homography; Camera calibration; 3D reconstruction; Absolute conic; Structured scene

1. Introduction

The recovery of structure and motion from uncalibrated images is one of the main goals in computer vision community. Significant progress has been made during the past two decades (Hartley and Zisserman, 2000). However, a lot of efforts have been spent on the development of high
accuracy and complete modeling of complex scenes based on matching primitive features, such as points and lines, without regarding to their relationships or context in the scene. In recent years, attentions are focused on using geometric constraints arising from the scenes to optimize the reconstruction, especially for architecture scenes (Wilczkowiak et al., 2002; Szeliski and Torr, 1998; Zhang and Kosecka, 2003). Such constraints may be expressed in terms of parallelism, orthogonality, coplanarity and other special inter-relationship of features.

There are many related works in the literature. Werner and Zisserman (2002a) present a coarse-to-fine strategy to reconstruct buildings. The planar model of the principal scene planes is generated by plane sweeping. Then, this model is used as a basis to estimate more details around the plane. They develop the idea in (Werner and Zisserman, 2002b) to automatically fit a plane and model the perturbation surfaces of architectures. The model is parameterized by combining disparity and gradient extrema and is trained from examples. The above methods are based on a relatively good projective or metric reconstruction of points and lines, which may not be available in some situations.

Baillard and Zisserman (1999) use inter-image homographies to validate and estimate the plane models. They first generate a one-parameter family of half-planes containing a 3D line. They then solve the ambiguity by measuring the similarities of projected images over multiple views. A reconstruction of 3D line segments is needed here. Dick et al. (2000) also start from a feature based reconstruction, where the planar model is initialized from the reconstructed points via RANSAC. Strong geometric constraints, such as perpendicularity and verticality, are explicitly used to optimize the model. In (Bartoli and Sturm, 2003), the geometric relationship between points and planes is described by multi-coplanarity constraints, and a maximum likelihood estimator that incorporates the constraints and structures is used in a bundle adjustment manner. Xu et al. (2000) propose a linear algorithm for camera calibration and reconstruction from two views given homographies of two planes in space. The method calibrates two one-parameter cameras under the assumption of zero-skew, unit aspect ratio and the principal point is at the image center. Some commercial systems, such as Facade (Debevec et al., 1996) and Photobuilder (Cipolla et al., 1999) can produce very realistic results. However, those systems usually require a lot of human interactions.

In this paper, we focus on the reconstruction of structured scenes from uncalibrated images and develop a practical heuristics system, which incorporates the information of points, lines, planes and their geometrical constraints. Unlike other research work reported in the literature, our method is based on 2D point matches that may contain many outliers. We adopt a RANSAC mechanism to detect the principal planar surfaces of the object directly from 2D images. The reconstruction is finally done by combining piecewise planar models of the scene, while only a few user interactions are needed in the process. We have made several contributions in the paper. First, we propose a simple method to match line segments between two views guided by homographies and the homographies are refined in return by incorporating the matching of lines. The infinite homography is also estimated in a similar way. Second, we calibrate the camera with four intrinsic parameters (focal lengths $f_u, f_v$ and principal point $u_0, v_0$) varying between two views under the assumption of zero-skew. Third, the proposed method may reconstruct more visible surfaces of the object as long as they can be seen in one of the images, while traditional method can only reconstruct the overlapping parts of the scene among different views.

The remaining parts of this paper are organized as follows. In Section 2, some preliminaries on homography are reviewed. The plane detection strategy is then elaborated in detail in Section 3. We present the method for recovering epipolar geometry and camera calibration in Section 4. The test results with synthetic data and real images are given in Section 5, followed by some conclusions of this paper in Section 6.

2. Notations and preliminaries

In order to facilitate our discussions in the subsequent sections, some preliminaries on plane
induced homography are presented here. Readers may refer to Hartley and Zisserman (2000) for more details. In this paper, the following notations are used. An image point in homogeneous coordinates is denoted by a bold lower case letter with the third element setting to 1, e.g. $x_i$. $(x_i)$ stands for the $i$th element of vector $x$. A matrix is denoted by a bold upper case letter, e.g. $H$. A feature in the second view is denoted by a feature letter with the third element setting to 1, e.g. $m_i, l_i, e_i$.

Under perspective projection, the projections $m_i$ and $m'_i$ in two images of the same 3D point $X_i$ are related with the epipolar constraint

$$m_i^T F m'_i = 0 \quad \text{Fe} = F^T e' = 0,$$

where, $F$ is the fundamental matrix, which is a $3 \times 3$ rank 2 matrix defined up to scale, $e$ and $e'$ are the corresponding epipoles in the two images. If the point $X_i$ lies on a plane in space, then $m_i$ and $m'_i$ are related with a homography $H$ induced by the plane

$$m'_i \sim H m_i,$$

where, $H$ is in general a full rank $3 \times 3$ matrix defined up to scale. Thus 4 point correspondences in general position (i.e. no three points are coplanar) can uniquely determine the homography. In the case of line correspondences $l_i \leftrightarrow l'_i$, we have

$$l_i \sim H^T l'_i \quad \text{or} \quad l'_i \sim H^{-T} l_i$$

then $H$ can be linearly computed from four such line pairs in general position (i.e. no three of them are concurrent) which correspond to lines on the same plane in space. In particular, if the fundamental matrix is accurately recovered, the homography can be determined from 3 point matches $m_i \leftrightarrow m'_i$ $(i = 1, 2, 3)$ by combining the epipolar constraint as

$$H = A - e' (M^{-1} b)^T,$$

where, $A = [e']^T F, M = [m_1, m_2, m_3]^T, b$ is a 3-vector with $(b)_i = (m'_i \times (Am_i))^T (m'_i \times e')/\|m'_i \times e'\|^2$.

3. Plane detection strategy

The strategy is based on an initial point matching result. We first estimate the principal planar homography by a recursive RANSAC method. Then find the line correspondences guided by the estimated homography and refine the homography by incorporating the feature of both points and lines on the plane.

Seeking correspondances between images is a difficult task. Most of the available methods are found to be error prone for man-made structure scenes due to the ambiguities caused either by large homogeneous regions of texture or repeated patterns in images. Fig. 1 shows two images of the Wadham College of Oxford. Using the correlation and relaxation method (Zhang et al., 1995), we establish 1128 initial matches, as shown in the first view of Fig. 1, where about 13% of the matches are outliers.

3.1. Coarse estimation of homography

Usually, most of the matches lie on several principal planar surfaces in a structured scene. The matches on a world plane are related with a planar homography. We adopt a similar method like Vincent and Lagniere (2001) to estimate the planar homography recursively based on a RANSAC mechanism (Fischler and Bolles, 1981). During iteration, planes are hypothesized from a randomly selected sample of four pairs of point matches (if a reasonable fundamental matrix can be computed from the initial matches, then three pairs of matches are enough to generate the planar homography from Eq. (4)). The plane which is best supported by those point pairs which have not been assigned to previous planes is selected. It should be noted that this method can only detect the principal surfaces of the object with relatively large number of initial matches. Fig. 2(a)–(d) give the result of the plane detection based on the initial matches of Fig. 1. The supporting matches for each plane are shown in different colors with their positions in the first view, while the outliers and the features not associated with the detected planes are eliminated. Totally five planes (which correspond to the left and right wall, the right roof,
and the planes on which the left windows and the right windows lie) are detected automatically.

Two types of mistakes may occur during the process. First, two or more homographies for the same world plane may happen sometimes. This is due to the noise of detected feature points. In our applications, this kind of mistakes can be easily detected and corrected. If two homographies have more than 40% of identical supporting features, it is most likely that they are for the same world plane. Second, an estimated homography may correspond to a virtual world plane. This mistake can be avoided by comparing the number of supporting points for each estimated plane, since there are less supporting points for a virtual plane compared with a principal plane. Nevertheless, this may eliminate some physical planes with less supporting matches as well. In Fig. 2(c), the left
roof is not detected automatically, since there are insufficient initial matches on the plane. The homography induced by this plane is provided by human interaction.

3.2. H-guided line matching

Line matching is a difficult problem since there are no strong disambiguating geometric constraints available and the topological connections between line segments are often lost during segmentation (Schmid and Zisserman, 1997). The existing approaches to line matching can be generally divided into two categories: One is to match individual line segments based on their geometric attributes, such as orientation, length and extent of overlap, or using nearest line strategy (Deriche and Faugeras, 1990). The other one is to match groups of line segments based on the idea of graph-matching (Gros, 1995). Since most of the segments of a structured object lie on the detected principal surfaces, the matching between these line segments can be obtained using the homographies. We apply the Canny edge detector and orthogonal regression algorithm to fit the line segments in images (Schmid and Zisserman, 1997). Fig. 4(a) shows 328 detected line segments of the first view.

Suppose the total numbers of the extracted line segments in image 1 and image 2 are \( n_1 \) and \( n_2 \), respectively. For a segment \( l_j \) in the second view, we can map it into the first view via Eq. (3) as \( \hat{l}_j \sim H_k \hat{l}_j \), where, \( H_k \) is the homography induced by the \( k \)th plane of the object. Then compare some relations between \( \hat{l}_j \) and \( l_i \) according to the following criteria:

\[
\begin{align*}
\text{angle} (l_i, \hat{l}_j) &< \varepsilon_1, \\
\text{dist}^2 (l_i, \hat{l}_j) &< \varepsilon_2, \\
\text{overlap} (l_i, \hat{l}_j) &> \varepsilon_3,
\end{align*}
\]

and adopt a winner-take-all strategy to select the matching candidates correspond to the plane. Here, \( \text{angle}(\bullet, \bullet) \) denotes the smaller angle between two lines, \( \text{dist}^2(\bullet, \bullet) \) denotes the sum of square of the Euclidean distance between the two endpoints of the first line segment to the second line, \( \text{overlap}(\bullet, \bullet) \) denotes the length of overlap of the two line segments. In the experiment, each \( \varepsilon_j \) is a given threshold. We choose \( \varepsilon_1 = 0.20, \varepsilon_2 = 0.18 \). \( \varepsilon_3 \) is selected according to the length of the two segments, we set it to the half length of the shorter segment in practice.

Most of the matching candidates obtained above are correct (i.e. lie on the images of the \( k \)th plane in space). However, a few of them may lie outside the plane (referred as incorrect matches here). For two pairs of matched lines \( l_i \leftrightarrow l'_i \) and \( l_j \leftrightarrow l'_j \), if they correspond to two coplanar lines in space, as shown in Fig. 3, then their intersections \( p_{ij} \) and \( p'_{ij} \) (in homogeneous form) must satisfy the following constraint:

\[
p'_{ij} \sim H_k p_{ij}.
\]

However, the two intersections may not conform to the above equation due to noise even the two lines are coplanar. In practice, we use the following criteria:

\[
d^2(p'_{ij}, H_k p_{ij}) < \varepsilon,
\]

where, \( d^2(\bullet, \bullet) \) denotes the square of the Euclidean distance between \( p'_{ij} \) and the mapped point \( H_k p_{ij} \) (the third element should be normalized to 1),

Fig. 3. The intersections of two coplanar lines between two views conform to the homography constraint.
and ε is a given threshold. For each pair of obtained candidates, we compute its intersection points with the remaining of line pairs as in Fig. 3. If more than 50% of the intersections satisfy (7), then this pair is considered to be correct. Otherwise, eliminate this match from the list.

In the same way, we can obtain all the correct line matches corresponding to all the detected space planes in Section 3.1. Fig. 4(b)-(d) show the line matching results on the six planes, respectively, and Fig. 4(d) gives all the matched segments together in the first view.

3.3. Homography refinement

From the previous discussion, we have obtained the correct matches of feature points and line segments on the detected planes. Suppose there are \( m \) matched points and \( n \) matched lines on the \( k \)-th plane in space. It is well known that line features are more stable and can be more precisely detected than point features. Therefore, the homography can be re-estimated by incorporating the information of both points and lines on the plane as
The above equation system can provide $2(m + n)$ linear constraints on the eight entities of homography. Thus, $H_k$ can be computed by singular value decomposition (SVD). However, this estimation is a least-squares solution and may not have any geometric meaning. Furthermore, we introduce an optimal algorithm to refine the solution by minimizing the following cost function:

$$f(H_k) = \frac{1}{m} \sum_{i=1}^{m} (d^2(m_i', H_k m_i) + d^2(m_i, H_k^{-1} m_i')) + \frac{1}{2n} \sum_{j=1}^{n} (\text{dist}^2(l_j, H_k^T l_j') + \text{dist}^2(l_j', H_k^{-T} l_j)),$$

where, the first term is the normalized geometric re-projection error of the matched points in bilateral directions, the second term is the normalized error of the line matches in the two views, and $\text{dist}^2(\cdot, \cdot)$ has the same meaning as in Eq. (5). The problem can be solved via an iterative method such as Gauss–Newton or Levenberg–Marquardt algorithm (Hartley and Zisserman, 2000), while the previous SVD estimation is taken as the initial value of the iteration.

Using the refined homographies, we may find more supporting point matches and line matches on the detected principal surfaces, and refine the homographies further according to Eqs. (8) and (9). The process is repeated iteratively until there are no new matches found. After retrieving all the homographies, we can also generate the intersection line between each pair of planes, as these lines may not be detected or matched in the previous steps. Suppose $H_1$ and $H_2$ are the homographies induced by two space planes. Let $l \rightarrow l'$ be the projections of the intersection of the two planes. Then,

$$l \sim H_1^T l', \quad l \sim H_2^T l'.$$

After a simple computation, we have

$$(H_2^T H_1)^T l \sim l.$$

Thus the intersection in the first view may be determined from the eigenvector corresponding to the real eigenvalue of matrix $(H_2^{-1} H_1)^T$. Actually, the matrix $H_2^{-1} H_1$ is the mapping from the first image onto itself. It is a planar homology which has a fixed point (vertex) and a line of fixed points (axis) (Hartley and Zisserman, 2000; Van Gool et al., 1998). The transformation has two equal and one distinct eigenvalues. The axis (i.e. the intersection line of the two planes here) is the join of the eigenvectors corresponding to the degenerate eigenvalues. The third eigenvector corresponds to the vertex, which is the epipole in the second view.

According to the intersections of any two adjacent planes and the line matches on each plane, it is easy to obtain the contour of each detected plane. Fig. 4(f) gives an example. In Fig. 4(f), most of the contours are found automatically, while those of some windows are modified by human interactions. Note that although there are no initial matches on the ground plane, we can still estimate its homography from the two intersection lines of the ground plane with the left and the right walls.

Homography provides a one-to-one mapping of points and lines between two views. For any points or lines on the detected planes in one image, we can immediately obtain their correspondences in the other view, even if these correspondences may be occluded or lie outside of the image. Thus we can reconstruct more parts of the object than the traditional structure from motion methods (Hartley and Zisserman, 2000) which can only reconstruct the overlapping parts of two images.

4. Epipolar geometry and camera calibration

4.1. Recovery of epipolar geometry

The epipolar geometry is encapsulated by fundamental matrix $F$, which is recovered from point matches. However, the classical methods to compute the fundamental matrix are unstable when most of the points lie close to one or two planes. Here, we adopt a similar method as Luong and Faugeras (1993) to compute $F$ directly from the refined homographies.

Suppose $H$ is the homography induced by a space plane, and an image point pair $m \leftrightarrow m'$
corresponds to a point on the plane. Then Eq. (1) becomes

\[ m^T F m = (Hm)^T F m = m^T (H^T F) m = 0. \]  \hspace{1cm} (12)

This is true for all points on the plane. Thus \( H^T F \) must be a skew-symmetric matrix, we have

\[ H^T F + F^T H = 0. \]  \hspace{1cm} (13)

Expanding Eq. (13) yields six linear equations on the eight entities of \( F \). So with two or more than two homographies, a SVD approach can be employed to give a least-squares solution of \( F \). Similarly, if we reverse the role of the two views, then a symmetric expression is obtained as

\[ H^{-T} F^T + FH^{-1} = 0. \]  \hspace{1cm} (14)

Eq. (14) provides the same constraints as (13). In practice, we incorporate this equation simultaneously with Eq. (13) so as to obtain a better estimation of the fundamental matrix, and the rank 2 constraint is imposed on \( F \) after computation.

4.2. Camera calibration

The image of absolute conic (IAC) \( \omega = (KK^T)^{-1} \) depends only on the camera calibration matrix \( K \). It is a symmetric matrix with five degrees of freedom defined up to scale. Once \( \omega \) is computed, the intrinsic parameters of the camera can be recovered by Cholesky decomposition of \( \omega^{-1} \).

For many kinds of man-made objects, such as architectures, we can usually obtain three mutually orthogonal pairs of parallel lines from a single image. Thus three orthogonal vanishing points, say \( v_x, v_y, v_z \), can be easily computed from the images (Heuvel, 1998; Liebowitz and Zisserman, 1999; McLean and Kotturi, 1995). Since the vanishing points with perpendicular directions are conjugate with respect to the image of the absolute conic, we have:

\[
\begin{align*}
  v_x^T \omega v_x &= 0, \\
  v_y^T \omega v_y &= 0, \\
  v_z^T \omega v_z &= 0.
\end{align*}
\]  \hspace{1cm} (15)

The above equations can give three linear independent constraints on \( \omega \). Similarly, we can have the constraints on \( \omega' \) in the second view as:

\[
\begin{align*}
  v_x'^T \omega' v_y' &= 0, \\
  v_y'^T \omega' v_z' &= 0, \\
  v_z'^T \omega' v_x' &= 0.
\end{align*}
\]  \hspace{1cm} (16)

Some researchers (such as Liebowitz and Zisserman, 1999; Caprile and Torre, 1990) use these constraints to calibrate the camera from a single view or from two views for varying camera parameters under the assumption of square pixels (i.e. zero-skew and unit aspect ratio). However, the square pixel assumption is much less tenable and may not hold for most off-the-shelf digital cameras.

The three pairs of vanishing points \( v_x \leftrightarrow v_x', v_y \leftrightarrow v_y', v_z \leftrightarrow v_z' \) lie on the plane at infinity, thus we can retrieve the infinite homography \( H_\infty \) from Eq. (4). The absolute conic also lies on the infinite plane whose images in the two views are related with the following equation:

\[ \omega' = H_\infty^{-T} \omega H_\infty^{-1}. \]  \hspace{1cm} (17)

This means that the constraints on the images of absolute conic, such as that given in Eqs. (15) and (16), can be easily transferred from one view to the other via the infinite homography. Eq. (17) can result six linear equations for the entities of \( \omega \) and \( \omega' \). It is proved in (Hartley and Zisserman, 2000) that Eq. (17) can at most give four independent constraints on the images of absolute conic if the two cameras have identical intrinsic parameters (i.e. \( \omega = \omega' \)).

In practice, it is difficult to fix the camera parameters over different views. It can be proved that Eq. (17) can provide five independent constraints on the images of absolute conic when the cameras are with varying parameters and under general motion. The proof is omitted here due to the limitation of space. The result is easy to be verified using MAPLE and Matlab. Therefore, a total of 11 constraints on \( \omega \) and \( \omega' \) can be obtained from Eqs. (15)-(17), while only eight of them are independent. Under the assumption of zero-skew, which is a quite natural and safe assumption for most imaging conditions, we can linearly calibrate two cameras with all the four parameters varying between views. If we have more prior knowledge about the camera, such as known aspect ratio or
principal point, then the cameras can be calibrated in the least-square sense from these constraints. We may also calibrate a general five-parameter-camera model with constant parameters in the two views.

Note: During calibration, degeneration may occur if one of the vanishing points is located at infinity in the images or the rotation between views is about the axes of the image coordinates. Please refer to Zisserman et al. (1998), Sturm (1997), Sturm and Maybank (1999), Hartley and Zisserman (2000) for more details on the degeneration and ambiguities arisen in camera calibration.

4.3. Improvements on estimation of $H_\infty$

(a) From the homographies of two parallel planes, such as those of the left (or right) wall and the left (or right) windows in Fig. 1, we can obtain the intersections $l_\infty \leftarrow l'\infty$ from Eqs. (10) and (11). Then $l_\infty$ and $l'\infty$ must be images of the vanishing line which lies on the plane at infinity.

(b) Let $l_i \leftarrow l'_i$ be a line pair in two views, $H$ be the homography induced by a plane parallel to the line in space. Then the intersection point of the line with the plane is imaged as

$$v_i \sim l \times (H^T l'), \quad v'_i \sim l' \times (H^{-T} l')$$

(18)

$v_i$ and $v'_i$ must be images of the vanishing point of the line and lie on the plane at infinity.

Combining all the correspondences given in (a) and (b), together with the three pairs of orthogonal vanishing points, the infinite homography can be re-estimated from Eq. (8) via SVD, and a further optimization may be obtained from the cost function (9).

If we can obtain the vanishing line $l_\infty$ of a plane and the vanishing point $v$ of the direction which is perpendicular to the plane, then it is easy to verify that $l_\infty = ov$ and $l'_\infty = ov'$. This can add more constraints to the images of absolute conic to obtain a better result for camera calibration.

After retrieving the camera parameters, a standard structure from motion algorithm is employed to reconstruct the whole object in the Euclidean space according to the contours shown in Fig. 4(f). Then a bundle adjustment is applied to optimize the reconstruction.

5. Some experiments

5.1. Camera calibration with simulated data

During simulations, the camera setup for the first view is: $f_u = 1200$, $f_v = 1000$, $s = 0$, $u_0 = 510$, $v_0 = 380$, rotation axis $r_1 = [2, 1, 4]^T$, rotation angle $\alpha_1 = \pi/7$ and translation vector $t_1 = [-5, -10, 200]^T$. For the second view: $f_u = 1100$, $f_v = 950$, $s = 0$, $u_0 = 500$, $v_0 = 390$, rotation axis $r_2 = [2, 5, 1]^T$, rotation angle $\alpha_2 = -\pi/6$ and translation vector $t_2 = [5, 0, 220]^T$.

We randomly select 100 points on each of the three world coordinate planes, and use their images (with added Gaussian noise) to compute the homography of this plane. We select three pairs of parallel lines corresponding to the directions of the three world axes. Each line is composed of 100 evenly distributed points. The Gaussian noise are added to the corresponding image points and the lines are fitted using least square fit from these points, while the vanishing points are computed as the intersections of the parallel lines. In order to provide more statistically meaningful results, we vary the noise level (the standard deviation of the Gaussian noise, unit: pixel) from 0 to 3 in steps of 0.2 during the test, and take 1000 independent tests at each noise level.

We recover the epipolar geometry from the homographies of the three world coordinate planes, and compute the infinite homography according to Eq. (4). Then we calibrate the two cameras under assumption of zero-skew according to the method proposed in Section 4.2. Fig. 5 gives the relative error and standard deviation of the four estimated intrinsic parameters of the first camera. The result for the second camera is similar to the first one and is omitted here. We can see from the simulations that the proposed calibration methods have low relative errors and standard deviations even under higher noise level.

5.2. Test with real images

The first test is two images of the Wadham College of Oxford, as shown in Fig. 1. The image resolution is $1024 \times 768$. From the above discussion, we have recovered the homographies and contours
of the principal planes. Then we use the proposed method to compute the fundamental matrix and calibrate the two cameras. The calibration results are shown in Table 1. We reconstruct the building from the extracted contours of Fig. 4(f) and map the texture to the model. Fig. 6 shows the result from different viewpoints, where the information of the chimneys is added interactively according to the initial matches. There are three mutually orthogonal planes in the scene: the ground plane, the left and the right walls. The reconstructed angle between these planes are 90.36°C, 90.28°C and 89.85°C respectively. The reconstructed angles between the two sets of parallel planes (i.e. the left wall and windows, the right wall and windows) are 0.32°C and 0.18°C respectively.

The second test is on two images of a model house (as shown in Fig. 7) with a resolution of 768 × 576. The number of initial matches found by Zhang et al. (1995) is 549 (as shown in the first view) with about 22% outliers. Using the proposed methods, four principal planes, which correspond to the ground, the roof, the front and the side walls, are detected automatically. A total of 333 pairs of supporting points to these planes are

Table 1
Calibration result of the Wadham College images

<table>
<thead>
<tr>
<th>Test 1</th>
<th>$f_u$</th>
<th>$f_v$</th>
<th>$u_0$</th>
<th>$v_0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Image 1</td>
<td>1093.3</td>
<td>1133.7</td>
<td>363.0</td>
<td>363.0</td>
</tr>
<tr>
<td>Image 2</td>
<td>1095.8</td>
<td>1280.5</td>
<td>320.6</td>
<td></td>
</tr>
</tbody>
</table>

Fig. 5. The relative errors and standard deviations of the four estimated intrinsic parameters.

Fig. 6. Reconstruction results of the Wadham College under different viewpoints with texture mapping.
shown in Fig. 8. Fig. 9 shows 38 pairs of matched line segments. Then we interactively draw the contours of the house from the line matches and intersections of planes. Table 2 lists the calibration results of the two views. The reconstruction result under different viewpoint with texture mapping is shown in Fig. 10. The three reconstructed angles between the ground the two walls are 90.27°, 90.41° and 89.78° respectively. This is very close to the intuitive result that the three planes are mutually orthogonal.

The third test is two images of a church in Valbonne, as shown in Fig. 11. The image resolution is 512 × 768. We use the same method to detect the principal planar surfaces and the contours of these planes. The intermediate results are omitted here. Table 3 lists the calibration results of the
two views. The reconstruction result is given in Fig. 12.

We can see from all the reconstructions that they are largely consistent with the real cases, and seem very realistic. Note that the reconstructed models are not merely the overlapping parts of the images, parts of them are beyond or occluded in one of the images.

6. Conclusions

In this paper, we have developed a practical heuristics method for reconstruction from two uncalibrated images by incorporating the information of points, lines, planes and their geometrical constraints. The method is based on an initial point matches that may contain many outliers. This requirement is easily satisfied even for wide-baseline images. We adopt a RANSAC mechanism to detect the principal planar surfaces of the object from 2D images. This is under the assumption that many initial matches are present on these surfaces. Otherwise the method may fail. The proposed method can be applied to model many structured scenes, especially those contain-
ing several large planar surfaces. However, it is hard to detect the smaller planes automatically. There are several points of novelty as mentioned above. The main contribution is that the proposed method can automatically reconstruct photo-realistic and accurate model of architectures with only minimal human interactions. In future studies, we shall consider the case of 3 or more images and incorporate with other approaches, such as those proposed by Werner and Zisserman (2002a,b). Then more constraints can be obtained and a better result is expected.

Acknowledgments

The work is supported by the Hong Kong RGC Grant CUHK 4378/02E, the National High Technology Development program of China under Grant No. 2002AA135110 and the Natural Science Foundation of China under Grant Nos. 60175009 and 60121302. The authors would like to thank Dr. Huynh for her helpful comments on the paper and thank the Geometry Group of the University of Oxford for providing the test images.

References


