An automatic 3D reconstruction method based upon coplanar and region constraints is proposed for synthesizing piecewise 3D planar models of a real-world scene portrayed in a sequence of digital images. Significantly, the reconstruction process requires no prior knowledge about the scene and the camera parameters. The proposed modeling method comprises three steps. Firstly, the corresponding feature points and lines are extracted from a series of images of the world scene. Secondly, the extracted corresponding points and lines are filtered in accordance with region and coplanar constraints and are used to identify the possible half-planes of real-world planes. Finally, a complete 3D planar model is constructed by identifying the correct half-planes in the world scene enlarging these half-planes to their full extent, and then merging all the extended half-planes which from part of the same world plane. The feasibility of the proposed approach is demonstrated by reconstructing 3D planar models of three real-world scenes containing objects with multiple planar facets.

Keywords: 3D reconstruction; Multiple views; Planar model; Planar homography; Region extension; Building modeling; Real-world scenes modeling

1. INTRODUCTION

The automatic generation of 3D models from digital images is a common requirement in many computer-based event simulation and visualization applications, including flight simulators, games, special effects, visual maps and so forth. Most the man-made objects, such as buildings, vehicles, furniture, and so forth are composed of multiple planar facades or facets, and thus reconstructing the planar facades or facets in a realistic and accurate manner is essential in reproducing a life-like representation of the real-world scene. Many studies on image-based reconstruction have been proposed over the past decade. In general, these systems exploit the geometric properties of the real-world planes, such as vertical facades [1, 2], horizontal planes [3], rectangular structures [4] and symmetry-based structures [5], or utilize a simple 3D wire-frame approach without attempting to verify the actual contents of planar regions [6, 7]. More recently, some novel studies focus on reconstructing the models of entire city scenes [8, 9], while some focus on 3D reconstruction from single image [10, 11].

Recent studies have shown that reconstruction systems based on inter-image homographies provide a particularly powerful technique for reconstructing models of real-world scenes containing planar regions [12-15]. For example, Baillard et al. [12] presented a method for reconstructing 3D models of buildings utilizing an inter-image homography technique and the concept of “half-planes”, namely a family of 3D planes rotating about a single 3D line. In [12], the half-plane was parameterized by the coordinates of the 3D line and the angle of rotation of the plane about the 3D line. Given the
3D borderline of a world plane, the correct half-plane, i.e. the half-plane corresponding to this world plane, was determined by searching through all the possible half-planes and by comparing the similarities of the regions projected by the half-plane between images. However, whilst this approach enhances the quality of the reconstruction results, the exhaustive search of the half-planes incurs a significant computational cost. In order to resolve this problem, the current study proposes a new half-plane parameterization method. Without loss of generality, an assumption is made that each correct half-plane contains at least one feature point, and thus the half-planes can be parameterized by the salient feature points distributed around the 3D line rather than searching every possible angle of rotation about the 3D line. Thus, the magnitude of the search process is reduced since it is necessary only to consider the salient feature points rather than 180 different angles of rotation. The complexity of the search process is further reduced by the imposition of two constraints, namely a coplanar constraint and a region constraint. The region constraint restricts the range over which the feature points are searched, while the coplanar constraint ensure that the selected feature points are coplanar in the 3D world. The imposition of these two constraints ensures that the maximum number of the feature points considered in the search process is far less than 180, i.e. the angles of rotation considered in the approach proposed in [12], and thus the efficiency of the search process is significantly improved.

In practice, the estimated half-plane does not cover the entire region occupied by the corresponding 3D plane in the real world. Furthermore, more than one half-plane may represent the same 3D plane. Thus, in order to extract the complete planar region, the reconstruction method proposed in this study extends the correct half-plane to its maximum extent, and then merges all the half-planes which represent the same world plane. In performing the region-extension process, the search region must first be properly bounded to avoid the requirement for an exhaustive search over the entire image. In [16], an approach was proposed for estimating the boundary between neighboring planes in stereo images by utilizing the characteristics of the fixed line which is a possible borderline of two arbitrary world planes determined by inter-image homographies. The proposed approach commenced by meshing the region around each half-plane with a fine rectangular grid. Then, for each grid rectangle, the fixed line of this grid rectangle and the half-plane was considered as a possible candidate for the borderlines of the plane. A set of candidate fixed lines comprising the fixed lines associated with all grid rectangles was accumulated using the Hough transform, and the correct borderlines were than selected from the Hough space with the highest supporting score. This method provides a convenient means of establishing the most likely boundaries of a half-plane, and is therefore used in the present study to bound the search range for the half-plane region-extension process. In general, region-extension methods extend the region of interest in the outward direction on a pixel-by-pixel basis [12]. However, such an approach incurs a considerable computational cost. Thus, in this study, the region-extension process is based on the use of coplanar feature points rather than pixels. Since the number of feature points within the bounded search region is far less than the total number of pixels, the proposed approach is considerably more efficient than conventional pixel-by-pixel schemes. Once each half-plane has been enlarged to its maximum extent (i.e. the point at which it meets the boundaries of its neighboring planes), the complete region of the corresponding real-world plane is constructed by merging all the extended half-planes which are both adjacent to one another and coplanar.
Before the 3D geometric information contained within a series of real world images can be extracted, it is necessary to determine the geometric relationship between the multiple views. In the reconstruction system proposed in this study, this is achieved by performing the following three-step procedure: (1) identify the correspondences between images; (2) estimate the camera pose; and (3) obtain the 3D lines and points. The corresponding points and lines identified in the first step are used to obtain the 3D lines and points from image sequences, to find the half-planes corresponding to the planes in the real-world scene and to produce the final planar models. The corresponding points are extracted using the Scale-Invariant Feature Transformation (SIFT) algorithm proposed in [17], while the corresponding lines are obtained by applying geometric and photometric constraints [12, 18]. Having identified the corresponding lines and points, the 3D space represented in the multiple views is defined by calibrating the camera [19] and calculating the camera poses [20]. Finally, the 3D positions of the lines and points are obtained from the corresponding lines and points using the 3D geometric information of multiple cameras [21]. Fig. 1 presents a schematic overview of the proposed 3D planar reconstruction system.

The remainder of this paper is organized as follows. Section 2 describes the process of identifying the half-planes of the planar region in the real-world scene from the corresponding lines and points over multiple views. Section 3 describes the half-plane extension and merging procedures performed to produce the final planar models. Section 4 presents the final experimental results. Finally, Section 5 provides some brief concluding remarks.

2. IDENTIFICATION OF HALF-PLANES IN ACCORDANCE WITH COPLANAR AND REGION CONSTRAINTS

This section describes the process of identifying the half-planes in a world scene based upon the corresponding lines and points within a sequence of images of the scene. In order to robustly estimate the planes, the computation process is based on an inter-image homography approach. The computation procedure involves two stages, namely (1) computing the homographies amongst the multiple images of the scene based on corresponding lines and points, and (2) parameterizing and verifying the planar fac-
Baillard et al. [12] proposed a planar reconstruction method in which the half-planes were parameterized by a single 3D line \( L \) and an angle of rotation \( \mu \) with 180 possibilities (see Fig. 2(a)). The half-plane was then verified by computing the intensity similarity of every pixel contained within it. However, this approach incurs a considerable computational cost since it is necessary to search a total of 180 possible angles of rotation. Moreover, when the region within the world plane is smooth and monotonic, the proposed method failed to determine the correct half-plane due to the false high similarity scores produced by the smooth regions. To reduce the complexity of the half-plane search process and to resolve the mismatch problem, the reconstruction method proposed in this study modifies the half-plane parameterization method proposed in [12] by replacing the angle of rotation parameter, \( \mu \), by a 3D feature point \( X \) (note that the use of the single 3D line \( L \) is retained, see Fig. 2(b)). Without loss of generality, an assumption is made that each correct half-plane contains at least one corresponding point in the multiple images. In order to reduce the computation time, the half-planes in the image are obtained by searching only those feature points which satisfy prescribed region and coplanar constraints for each 3D line \( L \). Note that here, the region constraint restricts the range over which the search process is performed, and the coplanar constraint is to limit the selection of candidate half-planes to those half-planes in which all the feature points are coplanar in the 3D world scene.

![Fig. 2](image_url)

**Fig. 2** One-parameter (angle \( \mu \) or point \( X \)) family of half-planes spanned by 3D line \( L \)

(a) half-plane is parameterized by angle \( \mu \);
(b) half-plane is parameterized by feature point \( X \)

### 2.1 Region Constraint

The half-plane search process commences by obtaining a set \( s_r \) of all the corresponding points within a restricted region around 3D line \( L \) in the first image. Note that the restricted region is bounded by a quadrangle extended from \( l_1 \), where \( l_1 \) is the 2D line produced by projecting \( L \) onto the first image. Let \( s_p \) be the set of all pixel points within the restricted quadrangle region \( r(L, W) \), where \( L \) is the length of the restricted region and is equal to the length of \( l_1 \) and \( W \) is the width. In addition, let the long side of the restricted quadrangle region be located along \( l_1 \). The set containing all the corresponding points in the range \( r(L, W) \) is given by:

\[
    s_r = \left\{ s_p \cap s_i \right\},
\]

where \( s_r \) contains all the corresponding points in the first image.

In practice, it is desirable to minimize the amount of information required to establish the possible half-planes in the real-world scene in order to improve the efficiency of the search process. Thus, in this study an assumption is made that if the number of \( s_r \), \( |s_r| \), in the region \( r(L, W) \) is greater than 1, then sufficient information exists to verify the half-plane. Since the value of \( L \) is fixed, the value of \( W \), i.e. the width of the re-
stricted quadrangle region, is simply defined as \( W = n \times L / 5 \), where \( n \) is initially set to a small value such as 1. If the resultant number of members in set \( s_r \) is 0, the value of \( n \) is increased by 1. This process is repeated iteratively until \( |s_r| \geq 1 \) or \( W = L \).

2.2 Coplanar Constraint

Once all the points in \( s_r \) satisfying the region constraint in the first image have been extracted, the set containing the 3D points \( S_r \) is constructed from the points within \( s_r \) in the first image and those within the corresponding set \( s'_r \) in the second image. Having obtained the set \( S_r \), an extraction procedure is performed to identify the set \( S \) containing all of the points which satisfy the coplanar constraint.

Common sense dictates that of all the half-planes spanned by the 3D line \( L \), the more realistic half-planes, i.e. the half-planes which more accurately resemble the real-world plane, are those which contain more coplanar feature points than the others. Thus, for each point \( X_j \) in \( S_r \), the number of points lying on the plane parameterized by \( L \) and \( X_j \) is computed, and the set \( S \) containing feature points whose associated half-planes contain the greatest number of coplanar points is derived in accordance with the following equation based on the 3D line \( L \),

\[
S = \arg \max S_j = \left\{ X_j \left| \left. \sum_{i} d(X_i, \pi(L, X_j)) < \epsilon, \forall X_i \in S_r \right. \right\},
\]

where \( S_j \) represents the set of points coplanar with \( \pi(L, X_j) \) which is the plane parameterized by the 3D line \( L \) and the feature point \( X_j \), and \( d(X, \pi) \) is the function measures the distance from the point \( X \) to the plane \( \pi \). If the distance is smaller than the threshold \( \epsilon \), the point \( X \) is considered as on the plane \( \pi \).

2.3 Verifying Half-Plane Regions

The procedure described above yields a set \( S \) of constrained feature points associated with the 3D line \( L \). The next step in the reconstruction process is to identify the correct half-plane region associated with the line \( L \). For each line \( L \), the possible half-planes of \( L \) are constructed by \( L \) and all the feature points \( X_k \) belonging to the set \( S \), that is, \( \pi(L, X_k) \) (see Fig. 2(b)). In this study, the homography \( H(\pi(L, X)) \) introduced by \( \pi \) between two different views is computed directly from the feature point \( X \) and line \( L \) which specify this half-plane \( \pi \). Let the projection matrix associated with the first view be denoted as \( P_1 = [I | 0] \), and let the projection matrix of any of the other views be denoted as \( P'_i = [A'_i | a'_i] \). Finally, let the planar equation of half-plane \( \pi \) be denoted as \( (\pi', \pi'', \pi''', \pi'''' \pi')^T \). The homography \( H \) can then be expressed as:

\[
H = A' + a' v^T \quad \text{where} \quad v = -\frac{1}{\pi'}(\pi', \pi'', \pi''', \pi'''' \pi')^T.
\]

In order to identify the correct half-plane amongst the set of all possible half-planes, the image intensity similarity of each half-plane region is estimated over the multiple views using the following similarity score function:

\[
\text{Sim}(X_j) = \sum_{i=1}^{n} \sum_{x_j \in \text{ROI}} \text{Cor}^2 \left( x_j, \hat{x}_j \right), \quad \text{where} \quad \hat{x}_j = H' \left( \pi \left( L, X_j \right) \right) x_j.
\]

In computing Eq. (4), the half-plane \( \pi(L, X) \) defined by \( X \) and line \( L \) is given, and the
homography $H^i$ introduced by $\pi$ between the first view and the $i$-th view is known. Meanwhile, the POI (i.e. the Points of Interest in the first view) are obtained by an edge detection method with a low threshold setting in order to make the maximum use of the texture information available in the image. In addition, $Cor(x, x^i)$ is the normalized cross-correlation between point $x$ in the first view and point $x^i$ in the $i$-th view within a local $n \times n$ window. After estimating the score for each feature point $X_k$, the point $X$ with the highest similarity score is considered to be the best solution for the half-plane.

The total number of points in $S$ satisfying both the region and the coplanar constraints is far less than 180. Consequently, the number of half-plane search trials is significantly reduced, and thus the efficiency of the search process is improved. Table 1 presents the pseudo-code for the half-plane search algorithm described in this section.

Table 1 Half-plane search algorithm

<table>
<thead>
<tr>
<th>Objective</th>
<th>Given 2D and 3D corresponding lines and points, find half-planes $\pi$ and corresponding regions $S$ by searching the corresponding points which satisfy both the region and coplanar constraints.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Algorithm</td>
<td>For each line $L$</td>
</tr>
<tr>
<td>1) Region Constraint:</td>
<td></td>
</tr>
<tr>
<td>(A) For $n = 1$–5</td>
<td></td>
</tr>
<tr>
<td>(a) Estimate $s_r = {s_r \cap s_5}$ with $s_r$ in $r(L, W)$ and $W = n \times L$;</td>
<td></td>
</tr>
<tr>
<td>(b) If $</td>
<td>s_r</td>
</tr>
<tr>
<td>(B) If $</td>
<td>s_r</td>
</tr>
<tr>
<td>2) Coplanar Constraint:</td>
<td></td>
</tr>
<tr>
<td>(A) Obtain the set $S_t$ containing the 3D points constructed by $s_t$ and the corresponding set $s_t'$;</td>
<td></td>
</tr>
<tr>
<td>(B) Estimate $S = \arg \max S_r = {X</td>
<td>d(X, \pi(L, X)) &lt; \varepsilon, \forall X \in S_r};$</td>
</tr>
<tr>
<td>(C) If $</td>
<td>S</td>
</tr>
<tr>
<td>3) Verifying Half-Plane Regions:</td>
<td></td>
</tr>
<tr>
<td>(A) For each point $X_k$ in $S$</td>
<td></td>
</tr>
<tr>
<td>(a) Compute the homography $H(\pi(L, X_k));$</td>
<td></td>
</tr>
<tr>
<td>(b) Compute similarity score of half-plane region constructed by $X_k$ and $L$ over views;</td>
<td></td>
</tr>
<tr>
<td>(B) The point $X$ with the highest score is considered to be the best solution for the half-plane;</td>
<td></td>
</tr>
</tbody>
</table>

3. RECONSTRUCTION OF PLANAR MODEL

This section describes the process of reconstructing a planar model of the real-world scene using the half-planes identified by the algorithm presented in Section 2. The process commences by incrementally extending each half-plane within a bounded search region using homography as the homogeneity criterion. The planar model is then constructed by merging all the extended half-planes which are both adjacent to one another and coplanar.

3.1 Defining the Boundaries for the Half-Plane Extension Process

Since in this study, each half-plane is parameterized using a line-point representation, if the corresponding real-world plane contains uneven regions or is partially occluded by other objects, the half-plane identified by the algorithm described in Section 2
may be neither absolutely correct nor entirely complete. Thus, a region-extension method is required to extend the half-planes such that they represent the complete planar regions within the real-world scene. However, before performing the region-extension process, it is necessary to bound the search region in order to prevent the requirement for an exhaustive search over the entire image. In practice, the search region for each half-plane is bounded by the borderline between this plane and its neighboring planes. Let $r_h$ be the projected region on the first image of the 3D half-plane $\pi_h$ identified in Section 2, and let $r_r (r = 1, 2, \ldots, n)$ be the projected region of the 3D half-plane $\pi_r$ nearby $\pi_h$ (see Fig. 3 (a)). If $\pi_r$ is not coplanar with $\pi_h$, a 2D fixed line of $\pi_r$ and $\pi_h$ can be obtained. This line is then considered to be a candidate for the borderline of the half-plane $\pi_h$. Note that the coplanarity between two 3D planes is determined by calculating the angle between their normal vectors. If the angle is close to 0 or 180 degrees, then the two 3D planes are considered as coplanar.

Fig. 3 (a) Experimental image, $r_h$ is the half-plane to be bounded, $r_{1:3}$ are the nearby half-planes, and $l_{1:3}$ are the obtained fixed lines. (b) Image of half-plane $r_h$ and fixed points $sFP$, which is a set of fixed points for each homography $H$. (c) Hough accumulator array (see explanation in the text).

Fig. 4 Geometry of fixed line and fixed point for $\pi_r$ and $\pi_h$

Simon et al. [16] proposed an accumulative method for robustly extracting the borderlines between neighboring planes in multiple views based on the characteristics of the fixed line determined by inter-image homographies. The fixed line of $\pi_r$ and $\pi_h$ is the line which lies on both planes (see Fig. 4). Let $l_r$ be the projection on $I_1$ of the 3D line $L_r$ formed by the intersection of $\pi_r$ and $\pi_h$. If $l_r$ is transformed to $I_2$ using $H_r^{-T}$ and $H_h^{-T}$, where $H_r$ and $H_h$ are the homographies between $I_1$ and $I_2$ introduced by $\pi_r$ and $\pi_h$ respectively, then the result should coincide with the line $l'_r$, which is the projection of $L_r$ on $I_2$ [21]. Therefore the line $l_r$ can be derived as the real eigenvector of $H_r^{-T}H_h^{-T}$ using

$$\lambda l_r = H_r^{-T}H_h^{-T}l_r.$$ (5)
In theory, the fixed line in Eq. (5), can be solved by applying singular value decomposition (SVD) to $H_h^T H_r$. However, extracting the eigenvectors of $H_h^T H_r$ using the SVD approach is problematic, since $H_h^T H_r$ is a non-symmetric matrix [16]. Thus, an alternative method must be found.

In this study, the fixed line in Eq. (5) is found by computing the POI $s_{POI}$ [12] using the Canny edge detector [22] on the region of the image around $r_h$, computing the values of Euclidean distance $||x_i - H_h^{-1} H_r x_i||$ for all $x_i$ in $s_{POI}$, ranking all $x_i$ in order of ascending Euclidean distance, and then selecting the top 1% of the ranked list to form a set of possible fixed points, $s_{FP}$ (see Fig. 3(b)). The Hough transform is performed on $s_{FP}$ yielding a Hough accumulator array $A(\rho, \theta)$ with a Hough resolution of 4 pixels and 0.5 degrees, respectively. Finally, the locations of the local maxima of $A(\rho, \theta)$ are grouped into clusters using a k-means algorithm, and the average locations of the clusters are taken as the parameters of the fixed lines (see Fig. 3(c)). Fig. 3(a) illustrates the corresponding positions of the fixed lines on the experimental image.

3.2 Region Extension Based on Feature Points

Having identified the bounds of the search region for each half-plane, a region-extension process is performed to ensure that the half-planes selected using the algorithm presented in Section 2 cover the entire region occupied by the corresponding real-world 3D planes. Schindler [14] presented a region-extension method for planar reconstruction applications in which it was required that the point correspondences between two views projected by the points on a 3D plane $\pi$ must satisfy the same 2D homography $H_\pi$ introduced by this 3D plane. More specifically, if a region belongs to a real 3D plane, then all of the points in this region are subject to the same homography $H$ between two different views (see Fig. 5). Adopting the same principle, this study treats the half-plane as a seed region, and then obtains the complete planar region by enlarging this seed region using its homography properties as an extension criterion.

![Fig. 5 Detection of plane regions](image)

(a) The images of the plane triangle are related by a homography $H$. (b) Extension of triangle to find the image region that satisfies $H$.

Conventional region-extension methods generally extend the seed region in the outward direction on a pixel-by-pixel basis. However, this is a slow and computationally intensive process since it is necessary to check every possible pixel to determine whether or not it belongs to the seed region. To resolve this problem, the half-plane extension
process performed in this study is based on the feature points contained within the bounded search region rather than all the pixels within the search region.

Section 2 has described the processes involved in obtaining the correct half-plane $\pi_h$ attached to the 3D line $L$ and estimating the homography $H(\pi_h)$ between two different views. Let $\pi_1$ denote the projection of $\pi_h$ in the first view. Furthermore, let $R_1$ be the bounded search region of $\pi_1$ obtained using the method described in Section 3.1. As described above, for reasons of computational efficiency, the extension process performed in this study is based on the corresponding feature points within $R_1$ rather than the pixels in $R_1$. If a corresponding point pair satisfies $H(\pi_h)$, then the 3D point reconstructed by this point pair lies on the 3D plane $\pi_h$. Thus, by testing all the corresponding points in $R_1$ using $H(\pi_h)$, those points which are not coplanar with $\pi_h$ can be filtered out to leave the following set of coplanar feature points:

$$s_h = \{ x_i \ | \ d(\pi^1, H(\pi_h)x_i) < \epsilon, \forall x_i \in \text{feature points in } R_1 \}$$  (6)

Where $x_i^2$ is the point corresponding to $x_i^1$ in view 2, and $d(\pi^2, H(\pi_h)x_i^2)$ is the distance between $x_i^2$ and $H(\pi_h)x_i^1$. Recall that $l_1$ is the 2D line projected by the 3D line $L$ on the first view. In the proposed half-plane extension process, the coplanar feature points in $s_h$ are sorted in ascending order based on their distance from line $l_1$ in view 1 and the seed region $\pi_1$ is extended incrementally by adding one point $x_i^1$ in $s_h$ each time (starting with the nearest point and working toward the furthest point) until all of the points in $s_h$ have been added. Each time a new point is added, the boundaries of the seed region are expanded to take in this new point, and thus a new, larger seed region is obtained (see Fig. 6). Fig. 7 illustrates a real-world example, in which the triangular region corresponds to the projected image $\pi_1$ of the half-plane $\pi_h$, while the region $R_1$ marked by green lines represents the bounded search region for the half-plane extension process. Finally, the blue triangles indicate the feature points which satisfy the homography $H(\pi_h)$ criterion, while the red rhombs indicate those which do not. The pseudo-code for the half-plane extension method is summarized in Table 2.
Table 2 Region-extension algorithm

Objective
Given the region \(\pi_1\), which is the projected image of the 3D half-plane \(\pi_h\) on the first view, and a set of corresponding feature points, extend \(\pi_1\) subject to a homography criterion.

Algorithm
1) Define the bounded search region \(R_1\), using the method presented in Section 3.1
2) For each corresponding point in \(R_1\), filter out those points which are not coplanar with \(\pi_h\) by checking whether a point pair satisfies \(H(\pi_h)\) or not.
3) The remaining feature points form a set \(s_h\):
   \[
   s_h = \{ \mathbf{x}^i \mid \mathbf{d}(\mathbf{x}^i, H(\pi_h)\mathbf{x}^i) < \varepsilon, \forall \mathbf{x}^i \in \text{feature points in } R_1 \} \]
4) Sort the points \(\mathbf{x}^i\) in \(s_h\) in ascending order according to \(d(\mathbf{x}^i, \mathbf{l})\).
5) Extend the seed region \(\pi_1\) incrementally by adding one point \(\mathbf{x}^i\) in \(s_h\) each time until all the points in \(s_h\) have been added.

3.3 Grouping of Extended-Planes and Rendering of Planar Model

In the reconstruction method proposed in this study, the objects within the scene are synthesized in accordance with the information contained within multiple views of the scene. However, the reconstruction results obtained for an object based on the information within any two of the available views may well differ in terms of the boundary of the planar region. Therefore, in the proposed approach, the complete planar regions associated with each object in the world scene are reconstructed by merging the extended half-planes which are both neighboring to one another and coplanar (see Fig. 8).

The reconstruction process commences by using the borderline characteristics \([16, 21]\) described in Section 3.1 to estimate whether or not two neighboring planes are coplanar. Assume that a basic plane is known. If two planes have the same borderline according to this basic plane, the two planes are coplanar. In the reconstruction process proposed in this study, the two planes are judged to be coplanar if the distance in the Hough space of the fixed lines of the two planes is less than 10 for a Hough resolution of 4 pixels and 0.5 degrees.

![Fig. 8 Coplanar extending-planes](image)

In this example, extending-plane \(\pi(L_0)\) is coplanar with extending-plane \(\pi(L_1)\), and thus the two planes are grouped together.

If two neighboring planes are judged to be coplanar, they are merged to form a single plane. For example, suppose that line \(L_0\) in Fig. 8 is a 3D line on the basic plane \(\pi(L_0)\) and \(L_i\) lies on the a different plane \(\pi(L_i)\). A new plane, \(\pi_{new}\), comprising both lines is initially obtained by taking the cross product of \(L_0\) and \(L_i\) as the normal vector. The optimum solution is then obtained by minimizing the distances between the new plane \(\pi_{new}\) and the points \(X_j\) on plane \(\pi(L_0)\) or \(\pi(L_i)\) using the Levenberg Marquardt algorithm. In optimizing the solution, the error function is defined as follows:
where the Euclidean orthogonal distance is computed as the inner product of \( X \) and \( \pi_{\text{new}} \).

Finally, the resulting merged planes and their geometric information are rendered to produce the final 3D reconstructed model. Pseudo-code of the algorithm used to construct the final planar models from the half-planes is summarized in Table 3.

### Table 3: Planar model reconstruction algorithm

<table>
<thead>
<tr>
<th>Objective</th>
<th>Produce planar model from extended half-planes.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Algorithm</td>
<td></td>
</tr>
</tbody>
</table>

2) For each half-plane \( \pi_h \)
   - Let the region \( R = \Phi \);
     - For each image \( I \)
       - Let \( I \) be base image and the region \( R = \Phi \);
         - For each image \( I \)
           - (A) Let \( I \) be second image;
             - (B) Decide the bounded searching region \( B \) for extending \( \pi_h \) using the algorithm described in Table 1;
             - (C) Obtain the extended region of \( \pi_h \) using the algorithm described in Table 2;
             - (D) Update \( R \) by adding new extended region;
         - For each planar region \( R \)
           - \( R = \{ R, R_j \} \);
   3) For each planar region \( R_m \)
     - For each planar region \( R_n \)
       - If \( R_m \neq R_n \) and \( R_m \) is coplanar with \( R_n \),
         - Estimate a new plane \( \pi_{\text{new}} \) from the two planar regions by minimizing the distance function:
           \[ \sum_{x \in (R_m \cup R_n)} d(X, \pi_{\text{new}}) \]
   4) Render the plane models.

### 4. EXPERIMENTAL RESULTS

This section verifies the performance of the proposed reconstruction method using two kinds of image sets, namely a series of images of a stacked arrangement of camera boxes on a table surface and the Aerial Views image sets provided by the Visual Geometry Group [23], the latter ones are the same image sets as those considered by Baillard et al. [12]. The accuracy of the half-planes reconstructed by the proposed method is quantified in terms of the orientation error relative to the corresponding real-world planes and is compared with that of the reconstruction results obtained using the method presented in [12]. In addition, the computational comparison between the proposed method and the method in [12] is also evaluated.

Fig. 9 presents three consecutive images of a stacked box arrangement taken by a common digital camera. Each image has a resolution of 1024×768 pixels, and the camera is calibrated using Yang’s method [5] implemented in the OpenCV library [20]. The fundamental matrices between the three views are estimated via the salient corresponding feature points are then decomposed to obtain the respective camera poses [21].
Fig. 9 Consecutive images of stacked camera boxes

Fig. 10(a) shows the 3D wire-frame model of the boxes constructed from the corresponding line segments extracted from the three images. Fig. 10(b) shows the estimated half-planes associated with some of these 3D line segments. After extending the half-planes and grouping the half-planes which are judged to be coplanar, a total of 7 planes are obtained for the real-world scene, as shown in Fig. 10(d). It is observed that each of the grouped coplanar regions covers the entire region of the corresponding box facet. Fig. 11 presents three views of the rendered 3D model of these 7 planes of the stacked-box with texture mapping. It is observed that all the planar facets which are visible in the original image set are fully reconstructed. However, those which are hidden in the original image set can not be reconstructed due to a lack of information.

The performance of the proposed reconstruction scheme was evaluated in terms of the orientation accuracy of the reconstructed half-planes, i.e. the difference between the intersection angles of the real planes and those of the corresponding reconstructed half-planes. For simplicity, the evaluation experiment considered only orthogonal or parallel plane pairs, i.e. planes characterized by a real-world intersection angle of 90 degrees or 0 degrees. Table 4 compares the orientation errors of five plane pairs in the model constructed using the proposed method with those of the corresponding plane pairs in the model produced using the method proposed by Baillard et al. [12]. The results confirm that the current method yields a more accurate representation of the 3D planes in the real-world scene than the method presented in [12]. Fig. 14(a) shows the indices of the planes of the boxes which are used for orientation accuracy evaluation.
The second set of reconstruction experiments was performed using the Aerial Views I dataset provided by the Visual Geometry Group [23]. The dataset comprises six consecutive images taken by an airplane-mounted camera. The dataset also includes the calibrated camera poses corresponding to each of the six different views. The images in the Aerial Views I dataset are acquired with a resolution of 600×600 pixels. Fig. 12 presents representative images in the dataset, while Fig. 13 shows various views of the planar models constructed from the images in the Aerial Views I datasets. From an observation of Fig. 13, it is clear that all of the 3D planes in the real-world scene are correctly and fully reconstructed other than some of the vertical walls which can not be discerned in the original image sets.

![Fig. 12 Images in the Aerial Views I datasets](image1)

![Fig. 13 Reconstructed 3D model of Aerial Views I dataset observed from different views](image2)

![Fig. 14 The indices of planes used for accuracy evaluation (a) stacked-box (b) Aerial Views I](image3)

The accuracy of the reconstruction result in Fig. 13 was quantified by evaluating the orientation accuracy of 15 plane pairs in the Aerial Views I dataset known to be orthogonal or parallel in the real world (i.e. to have an intersection angle of 90 degrees or 0 degrees). The planes which are used for orientation accuracy evaluation are shown in Fig. 14(b) with indices. Table 5 compares the intersection errors (in degrees) of the corresponding plane pairs in the model constructed by the proposed system with those of the plane pairs in the model constructed using the method presented in [12]. As in the first reconstruction experiment, it is observed that the results obtained using the proposed method are significantly more precise than those obtained using the method proposed by Baillard et al. [12].
The computational cost was evaluated in terms of the number of trials for searching a half-plane, the total computation time for searching all the half-planes, and the time consumed by the feature corresponding detection process. Table 6 shows the results of the computational evaluation of the stacked-box set reconstruction, while Table 7 shows the results of the Aerial Views set I. Both tables present the comparison of the computational cost using our method with that using the method proposed by Baillard et al. [12]. As shown in these tables, the average number of search trials required in our method is significantly reduced. Even for the worst case, the number of trials required in our method is still less than that required in Baillard’s method. From these tables, it is observed that the computation time of half-plane searching by our method is far less than by Baillard’s method (about 90% improvement). These tables also show that even with the extra time required for computing the corresponding points by SIFT, the total time cost in this method is still much less than that in the method [12].

Table 4 Errors of intersection angles between the constructed half-plane pairs for the stacked box image set

<table>
<thead>
<tr>
<th>Plane Pair</th>
<th>Our Method</th>
<th>Baillard’s Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>(A, B)</td>
<td>0.62°</td>
<td>0.49°</td>
</tr>
<tr>
<td>(A, F)</td>
<td>2.90°</td>
<td>2.88°</td>
</tr>
<tr>
<td>(C, B)</td>
<td>13.27°</td>
<td>13°</td>
</tr>
<tr>
<td>(G, E)</td>
<td>1.6°</td>
<td>4.87°</td>
</tr>
<tr>
<td>(E, F)</td>
<td>1.57°</td>
<td>9.64°</td>
</tr>
<tr>
<td>Average</td>
<td>3.79°</td>
<td>6.18°</td>
</tr>
</tbody>
</table>

Table 5 Errors of intersection angles between the constructed half-plane pairs for the Aerial Views I image set

<table>
<thead>
<tr>
<th>Plane Pair</th>
<th>Our Method</th>
<th>Baillard’s Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>(A, B)</td>
<td>3.77°</td>
<td>7.36°</td>
</tr>
<tr>
<td>(A, C)</td>
<td>4.66°</td>
<td>4.59°</td>
</tr>
<tr>
<td>(A, D)</td>
<td>0.88°</td>
<td>4.82°</td>
</tr>
<tr>
<td>(A, E)</td>
<td>4.08°</td>
<td>9.69°</td>
</tr>
<tr>
<td>(A, F)</td>
<td>0.83°</td>
<td>1.76°</td>
</tr>
<tr>
<td>(B, C)</td>
<td>1.10°</td>
<td>9.77°</td>
</tr>
<tr>
<td>(B, D)</td>
<td>3.43°</td>
<td>11.38°</td>
</tr>
<tr>
<td>(B, E)</td>
<td>0.51°</td>
<td>2.54°</td>
</tr>
<tr>
<td>(B, F)</td>
<td>3.02°</td>
<td>6.17°</td>
</tr>
<tr>
<td>(C, D)</td>
<td>4.39°</td>
<td>7.18°</td>
</tr>
<tr>
<td>(C, E)</td>
<td>1.26°</td>
<td>11.42°</td>
</tr>
<tr>
<td>(C, F)</td>
<td>3.98°</td>
<td>6.05°</td>
</tr>
<tr>
<td>(D, E)</td>
<td>3.65°</td>
<td>13.85°</td>
</tr>
<tr>
<td>(D, F)</td>
<td>0.41°</td>
<td>5.22°</td>
</tr>
<tr>
<td>(E, F)</td>
<td>3.39°</td>
<td>8.63°</td>
</tr>
<tr>
<td>Average</td>
<td>2.60°</td>
<td>7.34°</td>
</tr>
</tbody>
</table>

Table 6 The comparison of the computational cost between our method and Baillard’s method for the stacked box image set

<table>
<thead>
<tr>
<th>Set</th>
<th>Our Method</th>
<th>Baillard’s Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Boxes Set</td>
<td></td>
<td></td>
</tr>
<tr>
<td>no. of trials (worst case)</td>
<td>80 trials</td>
<td>180 trials</td>
</tr>
<tr>
<td>no. of trials (in average)</td>
<td>10.8 trials</td>
<td>180 trials</td>
</tr>
<tr>
<td>Total Time for half-plane search</td>
<td>63.73 sec</td>
<td>510.64 sec</td>
</tr>
<tr>
<td>Time for feature points detection</td>
<td>11.95 sec</td>
<td>N/A</td>
</tr>
</tbody>
</table>

Table 7 The comparison of the computational cost between our method and Baillard’s method for the Aerial Views I image set

<table>
<thead>
<tr>
<th>Set</th>
<th>Our Method</th>
<th>Baillard’s Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aerial Views</td>
<td></td>
<td></td>
</tr>
<tr>
<td>no. of trials (worst case)</td>
<td>38 trials</td>
<td>180 trials</td>
</tr>
<tr>
<td>no. of trials (in average)</td>
<td>6.0 trials</td>
<td>180 trials</td>
</tr>
<tr>
<td>Total Time for half-plane search</td>
<td>12.02 sec</td>
<td>116.78 sec</td>
</tr>
<tr>
<td>Time for feature points detection</td>
<td>5.11 sec</td>
<td>N/A</td>
</tr>
</tbody>
</table>
5. DISCUSSION

We have demonstrated with our experiments that our method yields a significant improvement in both the efficiency and the accuracy of reconstructing the 3D planar models when comparing to the method proposed in [12]. The improvement in efficiency is due to our new parameterization scheme of the half-plane. The trials of searching for the correct half-plane are significantly reduced when using our method. As for the accuracy, the original method in [12] usually failed to search for the correct half-plane when the planar region is mostly homogeneous. This is because the similarity scores of image pixels in this region are insensitive to the rotating angle. Therefore, there may be several possible angles with high similarity scores, causing difficulty in defining the correct one.

In the proposed method, we use the set of feature points that satisfy the coplanar constraint as the parameters to identify the correct half-plane. Since a realistic half-plane would contain more coplanar feature points than the others, therefore searching for the half-plane by these coplanar feature points yields more accuracy than by the angles.

The corresponding points and lines play an important role during the reconstruction process. In our approach, the initial rough corresponding points are extracted using the SIFT [17], and the corresponding lines are obtained by the method proposed in [24]. The correctness of these initial corresponding points and lines are not strongly ensured during the detection process. There may exist some erroneous correspondences among them. In our system, the initial corresponding points are firstly used for calculating the camera poses. During the calculating process, a robust estimator - the RANdom SAmple Consensus (RANSAC) algorithm [21] - is used. The RANSAC algorithm is able to cope with a large proportion of the erroneous corresponding points. Therefore, after the calculation of camera poses, the erroneous corresponding points are excluded. Since the camera poses have been obtained, the epipolar geometry between the two cameras can be simply established [21]. We then apply the epipolar constraint to the original rough corresponding lines and eliminate those which violate the epipolar constraint. The remaining corresponding points and lines, which satisfy the geometric relations between the two views, are the final correct correspondences.

Our system can efficiently produce the accurate and detailed piece-wise planar 3D models from a sequence of images without any prior knowledge of the scene and the cameras. However, this system still has some limitations. Firstly, if the proportion of the erroneous corresponding points is too large (more than 50%), the RANSAC algorithm is unable to handle the enormous errors. Consequently, the reconstruction process cannot be accomplished because of the incorrect camera poses. Secondly, since the half-planes are defined by the 3D lines and feature points which are calculated from the captured images. Therefore, if no line segment or no feature point is extracted from the images of the world plane, the corresponding half-plane cannot be obtained. The former usually happens when the boundary between two world planes is unobvious and unobservable; the latter often occurs when the region of the world plane is smooth and texture-less. Finally, this approach may fail to reconstruct the world plane of which the projected im-
age regions are too small. This is because these regions do not have enough content information for calculating the similarity score, and the corresponding feature point is also hard to be detected in these regions.

6. CONCLUSION

This paper has presented a method for reconstructing 3D planar models of real-world scenes based on the information contained within multiple views of the scene. The proposed technique applies the same half-plane concept as that proposed in [12], but parameterizes the half-plane using the corresponding feature points in the multiple views rather than the angle of rotation of the plane about a 3D line. The experimental results have shown that the proposed parameterization scheme yields a significant improvement in the efficiency of the search process performed to identify the candidate half-planes. The search process is further simplified via the imposition of region and coplanar constraints, which limit the region over which the search process is performed and exclude any points which are not coplanar in the 3D world scene, respectively. Having identified the correct half-planes in accordance with the results obtained from a similarity score function, a region-extension process is performed to ensure that the selected half-planes cover the entire region occupied by the corresponding real-world planes. To avoid the requirement for an exhaustive search over the full image when performing the half-plane extension process, the search region is defined in advance for each half-plane using the Hough transform-based plane boundary identification method presented in [16]. Significantly, the half-plane extension process performed in this study is based on the corresponding feature points contained within the bounded search region rather than all the pixels within the search region, and thus the efficiency of the extension process is greatly improved compared to that of conventional pixel-by-pixel extension schemes. Finally, the 3D planar model is constructed by merging all the half-planes belonging to the same world plane, i.e. all the half-planes which are both neighbouring to one another and coplanar. The feasibility of the proposed method has been confirmed by reconstructing 3D planar models of a stacked arrangement of boxes and two urban scenes, respectively. The experimental results have shown that the proposed method is both more efficient and more accurate than the 3D reconstruction system proposed by Baillard et al. [12].

REFERENCES

4. J. Košec̆k and W. Zhang, "Extraction, matching, and pose recovery based on domi-


20. Open Source Computer Vision Library, <sourceforge.net/projects/opencvlibrary/>


Yu-Chih WANG (王煜智) received the B.S. and M.S. degrees in Electrical Engineering from National Cheng Kung University, Tainan, Taiwan, in 2003 and 2005, respectively. He is pursuing the Ph.D. degree in Institute of Computer and Communication, National Cheng Kung University, Tainan, Taiwan. His current research interests include computer vision, image processing and stereo matching.

Kai-Yuan LIN (林凱遠) received the M.S. degree in Electrical Engineering from National Cheng Kung University, Tainan, Taiwan, in 2007. His research interests include image-based 3D reconstruction and computer vision.

Pau-Choo (Julia) CHUNG (詹寶珠) received the B.S. and M.S. degrees in electrical engineering from National Cheng Kung University (NCKU), Taiwan, in 1981 and 1983, respectively, and the Ph.D. degree in electrical engineering from Texas Tech University, Lubbock, in 1991. She is currently the Distinguished Professor with NCKU and an IEEE Fellow.