myjournal manuscript No.
(will be inserted by the editor)

Accurate Constraint-Based Modeling From A Single Perspective View*

Manolis I.A. Lourakis, Antonis A. Argyros

Institute of Computer Science
Foundation for Research and Technology - Hellas
Vassilika Vouton, P.O.Box 1385
GR 711 10, Heraklion, Crete, Greece
Tel.: +30 2810 391716, Fax: +30 2810 391601
e-mail: {lourakis|argyros}@ics.forth.gr

Received: date / Revised version: date

Abstract Recovery of a 3D model from a single image is possible provided that adequate geometric knowledge about the imaged scene is a priori available. This prior knowledge is essential for disambiguating among the infinitely many 3D reconstructions that are compatible with a given 2D image. In practice, single view reconstruction methods employ geometric knowledge in the form of constraints such as coplanarity, parallelism, perpendicularity, etc, that are assumed to be supplied by a user based on his/her interpretation of the scene. Most of the existing methods, however, produce reconstructions that only approximately satisfy the supplied geometric constraints. This paper puts forward a single view reconstruction method which produces reconstructions that accurately satisfy all specified geometric constraints. This is achieved by first obtaining a preliminary reconstruction and then refining it in an extendable, constrained minimization framework. Sample experimental results demonstrate the approach.

1 Introduction

Image-based geometric modeling strives to derive 3D models directly from a set of images [2]. It is an attractive paradigm for photorealistic modeling of geometric objects that has generated considerable interest in related techniques during recent years. This paper focuses on a particular class of such techniques, specifically those dealing with Single View Reconstruction (SVR), whose aim is to create a 3D graphical model corresponding to a scene for which only a single image is available. Due to their use of a very limited amount of input data, SVR techniques usually call for a priori geometric scene knowledge that is supplied in the form of user input.

* This work was partially supported by the EU COOP-CT-2005-017405 project RECOVER.

During the last decade, research in visual geometry has produced several methods for SVR. The *tour into the picture* (TIP) technique of Horry et al. [8] is one of the earliest such methods. Assuming images with one-point perspective, TIP roughly models a scene using an axis-aligned box. Foreground objects are manually modeled as "billboards" by separate polygons. A "spidery mesh" interface facilitates the interactive manipulation of the modeling box as well as its vanishing point, resulting in novel rendered views. The applicability of TIP is limited by the fact that the front and back faces of the employed box should be parallel to the image plane. When applicable, however, TIP produces visually convincing results. Two more flexible methods for SVR are proposed by Liebowitz et al. in [9]. The first is based on measuring the heights of points above a ground plane. To achieve this, however, the vertical projection on the ground plane of each point whose height is to be measured has to be visible in the image. Clearly, this requirement restricts the number of objects that can be reconstructed. The second method reconstructs planes sequentially and necessitates the computation of the vanishing line of each plane being reconstructed. As is the case with all sequential approaches, this second method may suffer from accumulated errors, unless the scene includes a reference ground plane that has visible intersections with all other planes being reconstructed.

Sturm and Maybank [13] develop a method for reconstructing a piecewise planar scene from a single image. Their method relies on the availability of user-provided constraints regarding perpendicularity, parallelism and coplanarity that are used for camera calibration and 3D reconstruction. Compared to the methods of [9], that in [13] is capable of reconstructing planes with arbitrary orientations whose vanishing lines are not known, provided that they share enough common points with already reconstructed planes. Therefore, [13] accepts a wider class of scenes that are amenable to SVR. On the other hand, perpendicularity and parallelism con-

straints are used only for camera calibration and not during reconstruction. Furthermore, coplanarity constraints are only approximately satisfied by the final reconstruction. User-provided geometric knowledge such as coplanarity, distance ratios, and plane angles is also employed by Grossman et al. [5], who describe an algebraic SVR method that employs this knowledge to disambiguate an imaged scene and obtain a unique reconstruction. Inspired by TIP [8], Hoiem et al. [7] propose an automatic method for SVR that models a scene as a collection of several planar billboards. Using statistical learning techniques, planar image regions are labeled into coarse categories depending on their orientation in the scene. Then, using simple assumptions on the relative orientation of regions from different categories, labels are used to derive a pop-up model by cutting and folding. Overall, and despite being limited to outdoor images with a ground plane and vertical planar structures, the method is interesting since it is the first attempt towards fully automatic SVR. All SVR methods briefly reviewed above are restricted to surfaces that are either planar or can be approximated by planes. The work of Zhang et al. [14] addresses the problem of reconstructing free-form curved surfaces by employing a sparse set of user-specified constraints on the local scene shape to formulate a constrained variational optimization problem whose solution yields a smooth 3D surface satisfying the constraints. Evidently, the resulting 3D surface is not necessarily geometrically accurate or even viable. Furthermore, owing to the employment of an orthographic projection model, the method is applicable only to images with limited perspective distortion.

In this paper, we propose a novel geometric approach for reconstructing a piecewise planar scene from a single perspective view and a set of user-supplied geometric constraints. Similar to most of the proposed SVR methods, we choose to model the surface of objects rather than their volume. Thus, we reconstruct planar faces as opposed to polyhedral primitive solids such as the prisms and pyramids employed in [2]. This is because solid primitives are often not fully visible in a single image due to occlusions and field of view limitations, therefore their reconstruction is not possible without considerable generalization. The proposed approach models objects using surface representations to which geometric constraints are added. It is inspired by the work of [13] since the latter is the most flexible of the SVR methods that have been proposed in the literature and builds upon it by accepting a richer repertoire of user-supplied geometric constraints and guaranteeing that the recovered model accurately satisfies all of them. More specifically, starting with an initial reconstruction obtained as in [13], our approach refines it in a constrained nonlinear least squares framework until it exactly adheres to the supplied constraints.

The rest of the paper is organized as follows. Section 2 introduces the notation that is used in the remainder of the paper and reviews some background material. Section 3 describes how the initial reconstruction is obtained and section 4 presents the proposed method for refining it according to geometric constraints. Some implementation details are given in section 5. Experimental results from a prototype implementation are presented in section 6 and the paper concludes with a brief discussion in section 7.

2 Elements of Single View Geometry

2.1 Camera Model

In the following, vectors and arrays appear in boldface and are represented using projective (homogeneous) coordinates [6]. An image point with Euclidean coordinates (x, y) is represented by the homogeneous 3-vector $\mathbf{x} = (x, y, 1)^T$ with ^{T} denoting transposition. Homogeneous vectors that are equal up to a common scale factor are equivalent. Similarly, an image line is represented by a homogeneous 3-vector $\mathbf{l}$ such that $\mathbf{l}^T\mathbf{x} = 0$ for all points $\mathbf{x}$ lying on it.

A pinhole camera is a device that perspectively projects points in space onto a plane. Let the center of projection be the origin of a Euclidean coordinate system and assume that the image plane is defined by $Z = f$, f being the camera's focal length. A point in space with coordinates $\mathbf{X} = (X, Y, Z)^T$ projects on the image point defined as $(\frac{fX}{Z}, \frac{fY}{Z})^T$. Using homogeneous coordinates, perspective projection can be expressed by the following linear mapping between the 3D space and a 2D image [6]:

$$\begin{bmatrix} fX \\ fY \\ Z \end{bmatrix} = \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}. \quad (1)$$

The 3×4 matrix in the right side of the above expression is the *camera projection matrix*, which is usually denoted by $\mathbf{P}$ and written more compactly as $\mathbf{P} = diag(f, f, 1)\,[\mathbf{I}|\mathbf{0}]$, where $diag(f, f, 1)$ is a diagonal 3×3 matrix, $\mathbf{I}$ is the 3×3 identity matrix and $[\mathbf{I}|\mathbf{0}]$ is matrix $\mathbf{I}$ augmented with a fourth column equal to the zero vector.

The previous analysis of perspective projection has assumed that the origin of image coordinates coincides with the principal point (i.e., the intersection of the normal from the center of projection to the image plane) and that pixel units are equal along the two image axes (i.e., pixels are square). In practice, these assumptions are not always convenient and, therefore, they can be relaxed as follows. When the origin of the image plane coordinates is not at the principal point, let the coordinates of the latter be (u_0, u_0). Then, the camera projection matrix can be written as $\mathbf{P} = \mathbf{K}\,[\mathbf{I}|\mathbf{0}]$, where $\mathbf{K}$ is the *intrinsic*

calibration parameters matrix, defined as [6]:

$$\mathbf{K} = \begin{bmatrix} f_u & s & u_0 \\ 0 & f_v & v_0 \\ 0 & 0 & 1 \end{bmatrix}. \quad (2)$$

The parameters f_u and f_v correspond to the focal length expressed in pixel units along the two axes of the image, s is the *skew* parameter and (u_0, v_0) are the coordinates of the image principal point in pixels. Parameter s is related to the angle between the two image axes and is zero for most cameras. Furthermore, the *aspect ratio* $r = \frac{f_v}{f_u}$ for a certain camera is fixed and equal to one in most cases. A camera with zero skew and unit aspect ratio is commonly referred to as a *natural* camera.

2.2 Vanishing Points and Homographies

Vanishing points and planar *homographies* are geometric objects that arise from properties of perspective projection and are of foremost importance in the context of SVR. Assuming an infinite 3D line that is imaged under perspective, a point on it that is infinitely far away from the camera projects to a finite image point known as the vanishing point that depends only on the 3D line's direction and not on its position. Thus, parallel 3D lines share the same vanishing points. In a similar manner, the vanishing points of sets of non-parallel, coplanar 3D lines lie on the same image line, which is known as the *vanishing line* of the underlying plane. Parallel planes share the same vanishing line. The transformation that maps a plane to another under perspective projection (e.g. a scene plane to image mapping) is a general plane-to-plane projective transformation that is known as a homography. A homography that maps the image plane to another one so that it removes the effects of projective distortion is referred to as a *metric rectification homography* [10]. Such a homography allows metric properties of the imaged plane, such as angles, length and area ratios, to be directly measured from its perspective image.

2.3 Intrinsic Camera Calibration

The process of estimating the intrinsic calibration parameters (i.e. interior orientation) of a camera is referred to as (intrinsic) *camera calibration*. Customarily, single view camera calibration is performed by determining the *image of the absolute conic* (IAC). The absolute conic is a special imaginary conic, having the property that its image projection depends on the intrinsic parameters but not on the camera orientation or position. The IAC is itself a conic whose equation is defined by a homogeneous 3×3 symmetric matrix ω given by $\omega = (\mathbf{K}\mathbf{K}^T)^{-1}$. Working with the IAC is more convenient than working with $\mathbf{K}$ since the constraints involved in single view camera calibration are linear in the elements of ω, thus admitting a simple algebraic solution. Suitable such constraints result from pairs of vanishing points corresponding to perpendicular directions [1] and metric rectification homographies [10]. Assuming that enough constraints for a chosen parametrization of the IAC exist, ω can be estimated by solving a homogeneous linear system formed by all available constraints; more details can be found in [11]. Having estimated ω, $\mathbf{K}$ can finally be determined from its Cholesky decomposition.

3 Obtaining an Initial Reconstruction

This section is an overview of the SVR method of Sturm and Maybank [13] that we employ to obtain an initial reconstruction of points and planes that is to be refined later. Planes are represented in the so-called *Hessian normal form* $\mathbf{n}^T\mathbf{X} = -d$, where $\mathbf{n}$ is the unit normal vector and d is the distance of the plane from the origin. It is assumed that vanishing lines of planes have been computed using user-supplied information concerning groups of parallel 3D lines and that the camera has been intrinsically calibrated as outlined in section 2.3.

Consider an image point $\mathbf{x}_i$ that is the projection of an unknown 3D point $\mathbf{X}_i$ that is to be reconstructed. The depth information of point $\mathbf{X}_i$ is lost after its projection on the image plane. Thus, $\mathbf{X}_i$ can lie anywhere along the ray defined by $\mathbf{x}_i$ and the center of projection. Knowledge of the camera calibration matrix $\mathbf{K}$ permits the definition of a parametric representation of this backprojected ray as $\lambda_i \, \mathbf{x}_i^{'}$, where λ_i is a scalar that corresponds to the distance of a point on the ray from the optical center and $\mathbf{x}_i^{'} = \mathcal{N}(\mathbf{K}^{-1}\mathbf{x}_i)$ with $\mathcal{N}(\cdot)$ denoting normalization to unit vector norm. In other words, to reconstruct a 3D point, it suffices to estimate a single parameter λ_i. Camera calibration also facilitates the estimation of a plane's normal $\mathbf{n}$ from its vanishing line $\mathbf{l}$ as [6]:

$$\mathbf{n} = \mathcal{N}(\mathbf{K}^T\mathbf{l}). \quad (3)$$

Thus, for the plane to be fully reconstructed, its only parameter that remains to be determined is its distance d from the origin.

The key observation behind the method of [13] is that the reconstruction of a plane permits the reconstruction of all points on it. Conversely, the reconstruction of at least one or three (depending on whether the normal vector has been estimated or not) points on a plane enables the reconstruction of the latter. Owing to the well-known depth/scale ambiguity, reconstruction from one or more images is possible only up to an unknown overall scale factor. For this reason, the position of the first plane to be reconstructed is determined arbitrarily by setting its parameter d to some value d_0. Having completed the estimation of the parameters of one plane, its intersections with the backprojected rays of all points lying on it allows these points to be reconstructed by determining

their corresponding λ_i from $\lambda_i = -\frac{d}{\mathbf{n}^T \mathbf{x}_i'}$. Then, the reconstructed points that belong to planes that have not yet been reconstructed facilitate the reconstruction of such planes, which in turn allows the recovery of more 3D points and so on. This scheme that alternates between reconstructing points and planes allows the reconstruction of points and planes that are "linked" together by means of common points. Despite them being essential for expanding a reconstruction, common points in the presence of noise cannot simultaneously satisfy the equations of all planes they belong to. This problem is dealt with in [13] by directly estimating a reconstruction which is such that minimizes the sum of squared distances from common points to planes. More specifically, the signed Euclidean distance of point i on a backprojected ray from plane j is given by

$$D_{ij} = \mathbf{n}_j^T \mathbf{x}_i' \lambda_i + d_j. \tag{4}$$

Observing that Eq. (4) is linear in λ_i and d_j, all such distances can be concatenated together leading to a matrix expression of the form $\mathbf{Mr}$, where $\mathbf{M}$ depends on the plane normals $\mathbf{n}_j$ and image projections $\mathbf{x}_i'$ and the vector of unknowns $\mathbf{r}$ consists of the λ_i's and d_j's for all points and planes, respectively. Then, a reconstruction can be computed up to scale by minimizing $||\mathbf{Mr}||$ subject to $||\mathbf{r}|| = 1$. The solution to this minimization problem is the eigenvector of $\mathbf{M}^T\mathbf{M}$ corresponding to its smallest eigenvalue [6]. More details on the exact definition of matrix $\mathbf{M}$ as well as the complete reconstruction algorithm can be found in [13].

Despite its elegance, the method of [13] has the major drawback that it cannot directly incorporate geometric constraints other than coplanarity. Recall, for instance, that the normals $\mathbf{n}_j$ of the various planes are kept fixed to the values computed from their vanishing lines with the aid of Eq. (3). Therefore, even if the user could supply constraints related to the perpendicularity or parallelism of certain planes, they cannot be directly imposed on the reconstruction. Furthermore, no constraints such as length or area ratios, segment angles, etc, can be exploited for increasing the accuracy of the reconstruction. Even the coplanarity constraints are approximately satisfied since the method minimizes the point to plane distances of Eq. (4) rather than demanding them to be exactly zero. All the above contribute to geometric inaccuracies in the reconstruction that manifest themselves as skewed planes and not perfectly parallel and/or perpendicular planes. We have found that such problems are most noticeable after mapping on skewed planes textures with regular patterns such as tiles. In the following section, we propose an approach that allows these shortcomings to be remedied.

4 Refining the Initial 3D Reconstruction

Suppose that n image points and m 3D planes have been identified by the user and that an initial reconstruction of them has been obtained from a single view. Our SVR method of choice for this initial reconstruction is that of [13] presented in section 3. However, as it will soon become clear, our proposed refinement is not tailored to it but can be used with any other SVR method producing a piecewise planar reconstruction. Assume further that the user has supplied his/her prior knowledge of the scene in the form of geometric constraints such as point coplanarity and known plane relative orientations (i.e., dihedral angles). Let $\mathbf{X}_i,\ i = 1 \dots n$ be the reconstructed estimates of 3D points that project on image points $\mathbf{x}_i,\ i = 1 \dots n$. Also, let $\Pi_j,\ j = 1 \dots m$ denote the scene's planes whose initial parameter estimates are given by $\mathbf{n}_j, d_j,\ j = 1 \dots m$ and let $\Pi = \{\Pi_j \,|\, j = 1 \dots m\}$ be the set of all such planes. Finally, let $\mathsf{A} \subseteq \Pi \times \Pi$ be the set of plane pairs (Π_i, Π_j) whose dihedral angles are a priori known and are equal to θ_{ij}. Notice that this set includes parallel and perpendicular plane pairs, since their dihedral angles are equal to $0°$ and $90°$, respectively. The rest of this section explains how can the available geometric constraints be imposed on the initial reconstruction.

The idea is to jointly refine the set of initial point and plane parameter estimates for finding the set of parameters that most accurately predict the locations of the observed n points on the image and, at the same time, satisfy the supplied geometric constraints. Formally, this can be formulated as minimizing the average *reprojection error* with respect to all point and plane parameters subject to the geometric constraints, specifically

$$\min_{\mathbf{X}_i, \mathbf{n}_j, d_j} \sum_{i=1}^{n} d(\mathbf{K}\mathbf{X}_i,\ \mathbf{x}_i)^2 \tag{5}$$

$$\begin{aligned} \text{subject to} \\ d_k &= d_0, \\ \{\mathbf{n}_j^T \mathbf{X}_i + d_j &= 0,\ \mathbf{X}_i \text{ on } \Pi_j\}, \\ \{||\mathbf{n}_j|| &= 1,\ \Pi_j \in \Pi\}, \\ \{\mathbf{n}_i^T \mathbf{n}_j &= \cos(\theta_{ij}),\ (\Pi_i,\ \Pi_j) \in \mathsf{A}\}, \end{aligned}$$

where $\mathbf{K}\mathbf{X}_i$ is the predicted projection of point i on the image (cf. Eq. (1)) and $d(\mathbf{x},\ \mathbf{y})$ denotes the reprojection error defined as the Euclidean distance between the image points represented by the homogeneous vectors $\mathbf{x}$ and $\mathbf{y}$. The first constraint in (5) specifies that the d parameter of some plane k is kept fixed to d_0 so that overall scale remains unchanged. Expressions in curly brackets of the form $\{C, P\}$ denote sets of constraints C defined by the geometric property P.

Clearly, (5) amounts to a non-linear least squares minimization problem under non-linear constraints. It involves 3 unknowns for each 3D point and 4 for each plane, which amount to a total of $3n + 4m$. Image projections are 2D, thus the total number of image measurements defining the average reprojection error equals $2n$. Regarding constraints, each plane introduces one constraint specifying that its normal vector should have unit

norm. Furthermore, each point yields one constraint for each plane on which it lies and the known dihedral angles introduce $|\mathsf{A}|$ additional constraints. In practice, the planes to be reconstructed are "interconnected" with several common points, therefore the number of available constraints plus that of projected image point coordinates to be fitted well exceeds the total number of unknowns. Constraints in (5) model the prior geometric scene knowledge and being hard ones, force a constrained minimizer to exactly satisfy them. In addition, the criterion minimized is not an algebraic but rather a geometric one, therefore it is physically meaningful. Imposing all constraints simultaneously has the advantage of distributing the error to the whole reconstruction, avoiding the error build-up inherent in sequential reconstruction. It should also be noted that other types of geometric constraints such as known length ratios and angles can be incorporated into (5) in a straightforward manner. Finally, the set of minimization unknowns in (5) can be extended to include the parameters of $\mathbf{K}$, allowing intrinsic calibration to be refined as well.

5 Implementation Details

Image line segments that are necessary for detecting vanishing points are defined manually. Maximum likelihood estimates (MLE) of the vanishing points corresponding to imaged parallel line segments are computed with the nonlinear technique suggested in [10]. Vanishing lines of planes are estimated from pairs of vanishing points corresponding to at least two sets of parallel, coplanar lines. Calibration matrix estimates are obtained by combining linear constraints arising from orthogonal vanishing points and metric rectification homographies, as described in [11]. In cases where the available calibration constraints are not enough to employ a natural camera model (i.e., their number is less than three), the principal point is approximated by the image center and only the focal length is estimated. The minimization in (5) is carried out numerically with the aid of the NLSCON constrained non-linear least squares routine [12], which implements a damped affine invariant Gauss-Newton algorithm. Bootstrapped with the initial reconstruction, NLSCON iteratively refines it until it converges to a local minimizer satisfying the specified constraints. Despite them being infeasible with respect to the constraints of (5), we have found experimentally that initial reconstructions computed as described in section 3 are sufficiently close to constrained minimizers, thus facilitating constrained minimization convergence. The Jacobian of the objective function as well as that of the constraints in (5) with respect to the reconstruction parameters that are necessary for the non-linear minimization have been computed analytically with the aid of MAPLE's symbolic differentiation facilities. The recovered reconstructions are saved in the VRML format, which is very convenient for visualizing them by means of virtual walk-throughs and also allows them to be imported by a wide variety of 3D graphics software for further use. To increase realism, textures are first extracted from the original image, then corrected for perspective distortion effects by warping according to their estimated metric rectification homographies and finally mapped on the recovered planar faces. Finally, it is worth mentioning that the analysis of section 4 has assumed that the radial lens distortion in the image is negligible. If this is not the case, the effects of distortion can be corrected by applying a preprocessing technique such as [3].

6 Experimental Results

This section provides experimental results from a prototype C implementation of the proposed method, developed along the guidelines set forth in section 5. The first experiment was carried out with the aid of the high resolution 768×1024 image shown in Fig. 1(a). This image has plenty of parallel lines that can be employed for computing vanishing points. The three mutually orthogonal line sets shown with different colors in Fig. 1(b) were used for estimating three orthogonal vanishing points. Three calibration constraints arise from these points, from which the intrinsic calibration matrix was estimated assuming a natural camera.

Following calibration, the outlines of eight planar regions were marked manually using polylines, which are illustrated along with plane numbers in Fig. 2(a). Polyline vertices specify the points to be reconstructed. Vanishing lines for the walls were estimated using appropriate pairs from the set of the three orthogonal vanishing points. The roof eaves (i.e., planes 6 and 7) were approximated with isosceles trapezoids and their vanishing lines were computed with the aid of the normal bisector of the base sides that is defined by the intersections of trapezoid diagonals and non-parallel sides. The marked polylines specify coplanarity constraints, which combined with plane vanishing lines, permitted the recovery of an initial reconstruction using the method of [13]. This reconstruction was already of a satisfactory quality, as can be verified from the average deviation from 90° of angles that should be right, which amounted to 2.52°. Nevertheless, demanding plane pairs 0-3, 1-2, 1-4, 1-5, 2-4, 2-5 and 4-5 to be parallel and pairs 0-1, 0-2, 0-5, 0-4, 1-3, 2-3 and 3-4 to be orthogonal (see Fig. 2(a)), forced this deviation to become zero, thus yielding a more accurate reconstruction. A sample view of this final reconstruction is shown in Fig. 2(b) and a perspective top view of the wireframe model corresponding to it in Fig. 3. From the latter figure, it is evident that parallel and perpendicular planes have been accurately reconstructed as such. It should also be noted that the initial reconstruction using [13] has preserved plane parallelism due to the fact that identical vanishing lines were specified for all

(a)

(b)

Fig. 1 (a) An image of the campanile of St. Stefano's cathedral in Prato, Italy and (b) the lines employed to detect orthogonal vanishing points shown superimposed on (a) in different colors.

parallel planes. If, however, vanishing lines had been estimated independently for each plane, they would have been different and would result in not perfectly parallel reconstructed planes.

The second experiment employs the 800×600 image shown in Fig. 4(a), on which twelve planes were specified as illustrated by the polylines in Fig. 4(b). Two orthogonal vanishing points were estimated from a pair of parallel line segments lying on walls 0, 2, 6 and 1, 3 (see Fig. 4(b)). Approximating the camera's principal point with the image center and setting the aspect ratio to

(a)

(b)

Fig. 2 (a) The polylines and numbers of the planes to be reconstructed from the image of Fig. 1(a) and (b) a novel view of the reconstructed VRML model. The red pyramid in (b) corresponds to the reconstructed camera.

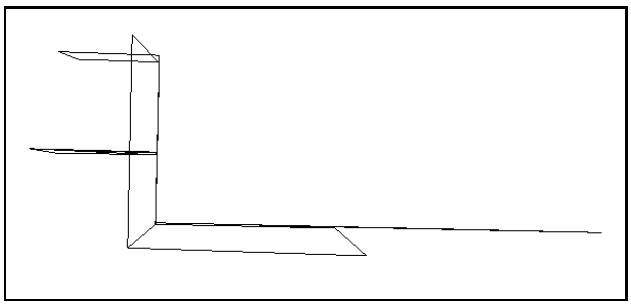
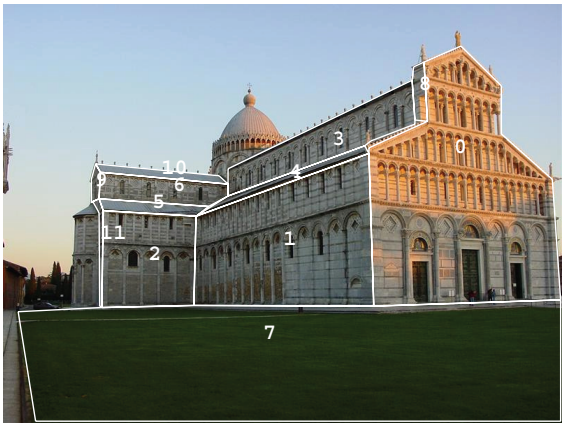


Fig. 3 Top view of the wireframe model corresponding to the reconstruction of Fig. 2(b) with the proposed method that demonstrates the correctness of parallelism and orthogonality relations among reconstructed planes.

(a)



(b)

Fig. 4 (a) An image of the Duomo of Pisa, Italy and (b) the polylines defining planes to be reconstructed and corresponding plane numbers.

one, this pair of vanishing points provides one constraint that suffices to estimate the focal length. Application of the proposed method for imposing plane parallelism and perpendicularity constraints yielded a reconstruction satisfying all constraints and whose right angles reconstruction error was zero. For comparison, this error had a mean value of 4.41° and a standard deviation of 2.95° in the initial reconstruction obtained with [13].

(a)

(b)

Fig. 5 (a),(b) Novel views of the VRML model reconstructed from Fig. 4(a). The hole visible in (a) is caused by self-occlusion in the original input image.

Figures 5(a) and (b) show different views of the textured VRML model reconstructed using the proposed method.

To facilitate a visual comparison between final and initial results, Figs. 6(a) and (b) show top views of the wireframe model reconstructed with the proposed method from slightly different viewpoints. Using similar viewpoints, Figs. 7(a) and (b) show top views of the wireframe model reconstructed using [13]. As it can be confirmed from them, the proposed method has accurately reconstructed 3D planes, preserving their orthogonality and parallelism. This is in contrast with the results of [13], where such constraints are violated. For instance, as can be seen from Figs. 7(a) and (b), right angles between walls 1-2 and 0-1 have not been recovered correctly. Furthermore, it can be verified from Fig. 7(a) that the recovered shapes of the two roofs in the far left part of the image (planes 5 and 10) are inaccurate.

7 Conclusion

This paper has presented a method for constraint-based single view reconstruction. The method starts by obtaining a preliminary reconstruction and then refines it in a constrained minimization framework, which ensures that user-specified constraints are fully exploited. Making a more effective use of user-specified constraints has been demonstrated experimentally to improve the geometrical accuracy of reconstructions and thus contribute to their overall quality. Future work will concern the incorporation into the method of more types of geometric constraints such as length ratios, symmetries, etc and its extension to the reconstruction of curved surfaces such as those of cylinders and spheres.

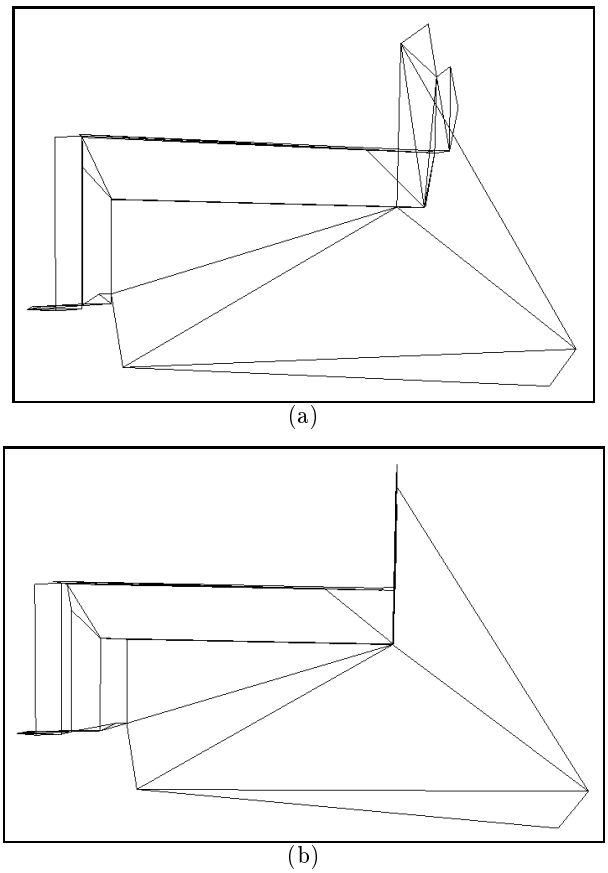


Fig. 6 Different top views of the wireframe model corresponding to the reconstruction of the image in Fig. 4(a) that was obtained with the proposed method.

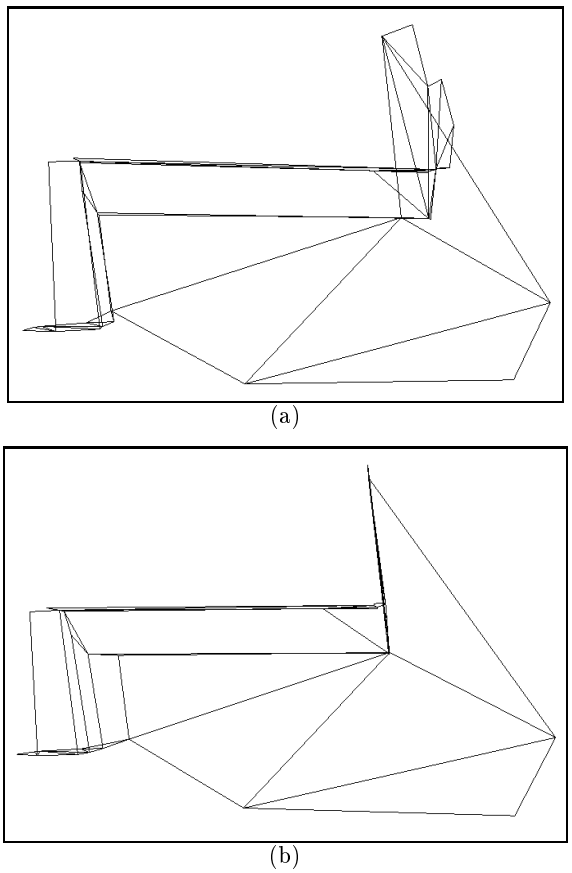


Fig. 7 Top views from viewpoints similar to those in Fig. 6 of the wireframe model corresponding to the initial reconstruction of Fig. 4(a) obtained using [13].

We conclude by mentioning that a well-known drawback of SVR method relates to the fact that the recovered models often lack completeness due to self-occlusions that introduce holes in the reconstruction. Occasionally, such problems can be dealt with by using texture synthesis techniques such as [4].

References

1. Caprile, B., Torre, V.: Using Vanishing Points for Camera Calibration. IJCV **4**(2), 127–140 (1990)
2. Debevec, P., Taylor, C., Malik, J.: Modeling and Rendering Architecture from Photographs: A Hybrid Geometry and Image-Based Approach. In: Proc. of SIGGRAPH'96, pp. 11–20 (1996)
3. Devernay, F., Faugeras, O.: Automatic Calibration and Removal of Distortion from Scenes of Structured Environments. In: SPIE, vol. 2567. San Diego, CA (1995)
4. Efros, A., Freeman, W.: Image Quilting for Texture Synthesis and Transfer. In: Proc. of SIGGRAPH'01, pp. 341–346. ACM Press, New York, NY, USA (2001)
5. Grossmann, E., Ortin, D., Santos-Victor, J.: Single and multi-view reconstruction of structured scenes. In: Proc. of ACCV'02, pp. 228–234 (2002)
6. Hartley, R., Zisserman, A.: Multiple View Geometry in Computer Vision. Cambridge University Press (2000)
7. Hoiem, D., Efros, A., Hebert, M.: Automatic Photo Pop-Up. In: Proc. of SIGGRAPH'05, pp. 577–584. ACM Press, New York, USA (2005)
8. Horry, Y., Anjyo, K., Arai, K.: Tour Into the Picture: Using a Spidery Mesh Interface to Make Animation From a Single Image. In: Proc. of SIGGRAPH'97, pp. 225–232. ACM Press, New York, USA (1997)
9. Liebowitz, D., Criminisi, A., Zisserman, A.: Creating Architectural Models from Images. Computer Graphics Forum **18**(3), 39–50 (1999)
10. Liebowitz, D., Zisserman, A.: Metric Rectification for Perspective Images of Planes. In: Proc. of CVPR'98, pp. 482–488. Santa Barbara, CA (1998)
11. Liebowitz, D., Zisserman, A.: Combining Scene and Auto-calibration Constraints. In: Proc. of ICCV'99, pp. 293–300. Kerkyra, Greece (1999)
12. Nowak, U., Weimann, L.: A Family of Newton Codes for Systems of Highly Nonlinear Equations. Tech. Rep. 91-10, Konrad-Zuse-Zentrum für Informationstechnik Berlin (ZIB) (1991). Available at http://www.zib.de/Publications/abstracts/TR-91-10/
13. Sturm, P., Maybank, S.: A Method for Interactive 3D Reconstruction of Piecewise Planar Objects from Single Images. In: Proc. of BMVC'99, pp. 265–274 (1999)
14. Zhang, L., Dugas-Phocion, G., Samson, J., Seitz, S.: Single View Modeling of Free-Form Scenes. In: Proc. of CVPR'01., vol. 1, pp. 990–997 (2001)