Enforcing Scene Constraints In Single View Reconstruction†

Manolis I. A. Lourakis and 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
email: {lourakis | argyros}@ics.forth.gr

Abstract

Three-dimensional reconstruction from a single view is an under-constrained process that relies critically upon the availability of prior knowledge about the imaged scene. This knowledge is assumed to be supplied by a user in the form of geometric constraints such as coplanarity, parallelism, perpendicularity, etc, based on his/her interpretation of the scene. In the presence of noise, however, most of the existing methods yield reconstructions that only approximately satisfy the supplied geometric constraints. This paper proposes a novel single view reconstruction method that provides reconstructions which exactly satisfy all user-supplied constraints. This is achieved by first obtaining a preliminary reconstruction and then refining it in an extendable, constrained optimization framework.

Categories and Subject Descriptors (according to ACM CCS): I.3.5 [Computational Geometry and Object Modeling]: Geometric algorithms, languages, and systems I.2.10 [Vision and Scene Understanding]: 3D/stereo scene analysis

1. Introduction

Image-based modeling (IBM) constitutes an attractive paradigm for generating photorealistic 3D models directly from a set of images. This work deals with a particular class of IBM methods, namely those concerned with Single View Reconstruction (SVR), whose aim is to create 3D graphical models corresponding to scenes for which only a single perspective image is available. Due to their use of a very limited amount of input data, SVR techniques typically call for a priori geometric scene knowledge that is supplied through user input. Relevant research has resulted into several SVR techniques, e.g. [HAA97, SM99, LCZ99, GOSV02, HEH05]. All such techniques rely on user-provided geometric constraints such as coplanarity, perpendicularity, parallelism, distance ratios and plane angles in order to disambiguate among the infinitely many 3D reconstructions that are compatible with a given 2D image. Most of these techniques, however, do not guarantee that the supplied constraints are respected by the recovered reconstruction. Coplanarity constraints, for instance, often hold only approximately while perpendicularity and parallelism constraints are usually employed only for camera calibration and not enforced during reconstruction.

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. The proposed approach models objects using surface representations to which geometric constraints are added. It complements the work of [SM99], which is one of the most flexible SVR methods proposed in the literature, and improves 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 [SM99], our approach refines it in a constrained nonlinear least squares framework until it exactly adheres to the supplied constraints. The rest of the paper presents some background knowledge in section 2 and section 3 describes the proposed method for enforcing the supplied geometric constraints. Some implementation details are given in section 4, experimental results are presented in section 5 and the paper concludes in section 6.

2. Background

Vectors and arrays are represented using projective (homogeneous) coordinates [HZ00]. Homogeneous objects that are equal up to a scale factor are equivalent. An image point with Euclidean coordinates (x, y) is represented by the homogeneous 3-vector $x = (x, y, 1)^T$ with ^{T} denoting transpo-

† Supported in part by the EU COOP-CT-2005-017405 project RECOVER.

sition. Similarly, an image line is represented by a homogeneous 3-vector l such that $l^T x = 0$ for all points x lying on it. We adopt a camera-centered coordinate system and a *pinhole* camera model, which projects a point in space with coordinates $X = (X,Y,Z)^T$ to the homogeneous image point KX, with K being the *intrinsic calibration* matrix [HZ00]. Planes are represented in the *Hessian normal form* $n^T X = -d$, where n is the unit normal vector and d is the distance of the plane from the origin.

We next provide an overview of the SVR method of Sturm and Maybank [SM99] that is employed to obtain an initial reconstruction of points and planes that is to be refined later. It is assumed that vanishing lines of planes have been estimated from appropriate pairs of vanishing points and that the camera has been intrinsically self-calibrated as explained in [LZ99, HZ00]. Given the calibration, a plane's normal can be estimated from its vanishing line l as $K^T l$. The key observation behind [SM99] is that the reconstruction of a plane permits the reconstruction via backprojection 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. 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 is implemented in [SM99] by a linear technique that involves minimizing the sum of squared distances of points to planes.

3. Enforcing Geometric Constraints

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 [SM99], briefly presented in section 2. 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 X_i be the reconstructed estimates of 3D points that project on image points $x_i,\ i = 1\ldots n$. Also, let $\Pi_j,\ j = 1\ldots m$ denote the scene's planes whose initial parameter estimates are given by $n_j, d_j,\ j = 1\ldots m$ and let $\Pi = \{\Pi_j \,|\, j = 1\ldots m\}$ 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

$$\min_{X_i, n_j, d_j} \sum_{i=1}^{n} d(KX_i,\, x_i)^2, \qquad \text{subject to} \tag{1}$$
$$\begin{aligned} d_k &= d_0, \\ \{n_j^T X_i + d_j &= 0,\ X_i \text{ on } \Pi_j\}, \\ \{||n_j|| &= 1,\ \Pi_j \in \Pi\}, \\ \{n_i^T n_j &= \cos(\theta_{ij}),\ (\Pi_i,\, \Pi_j) \in \mathsf{A}\ \}, \end{aligned}$$

where KX_i is the predicted projection of point i on the image and $d(x,\, y)$ denotes the reprojection error defined as the Euclidean distance between the image points represented by the homogeneous vectors x and y. The first constraint in (1) 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, (1) 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 (1) model the prior geometric scene knowledge and being hard ones, force a constrained minimizer to exactly satisfy all of them, irrespective of their order. In addition, the criterion minimized is not an algebraic but rather a geometric one, therefore it is physically meaningful [HZ00]. 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 (1) in a straightforward manner. Finally, the set of minimization unknowns in (1) can be extended to allow refinement of the intrinsic calibration parameters.

4. 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 [LZ98]. 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 [LZ99]. 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. To relieve the user from having to specify all possible parallel plane pairs, the transitive closure of the parallelism relationship is determined by a graph-based algorithm similar to that of Warshall. Similarly, the transitive closure for the perpendicularity relationship is computed by noting that if plane i is perpendicular to j and j is parallel to k, then i is also perpendicular to k. The minimization in (1) is carried out numerically with the aid of the NLSCON constrained non-linear least squares routine [NW91], 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 (1), we have found experimentally that initial reconstructions computed as described in section 2 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 (1) 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. Recovered reconstructions are saved in the VRML format to aid in their visualization. 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.

5. Experimental Results

This section provides experimental results from a prototype implementation of the proposed method, developed along the guidelines set forth in section 4. The experiment reported here was carried out with the aid of the 800×600 image shown in Fig. 1, on which twelve planes were specified as illustrated by the overlaid polylines. Two orthogonal vanishing points were estimated from two sets of parallel line segments lying on walls 0, 2, 6 and 1, 3. Approximating the camera's principal point with the image center and setting the aspect ratio to one, this pair of vanishing points provides one constraint that suffices to estimate the focal length. Following calibration, the coplanarity constraints corresponding to the marked polylines were exploited to recover an initial reconstruction using [SM99]. In that reconstruction, the absolute deviation from 90° of angles that should be right had a mean value of 4.41° and a standard deviation of 2.95°, indicating certain inaccuracies. Plane parallelism has been preserved due to the fact that identical vanishing lines were specified for all 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. Application of the proposed method for imposing the parallelism of plane pairs 0-2, 0-6, 1-3, 1-8, 1-9, 1-11 and the perpendicularity of 0-1, 0-7, 1-2, 1-7, 2-7, 2-11, 7-9, formulated a minimization problem involving 156 unknowns, 72 image measurements and 142 constraints. The solution of this problem required a few seconds and yielded a reconstruction satisfying all constraints and whose right angles reconstruction error was zero. Figures 2(a) and (b) show different views of the textured VRML model reconstructed using our method.

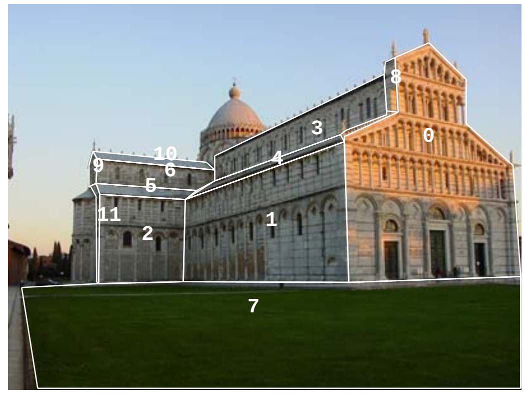
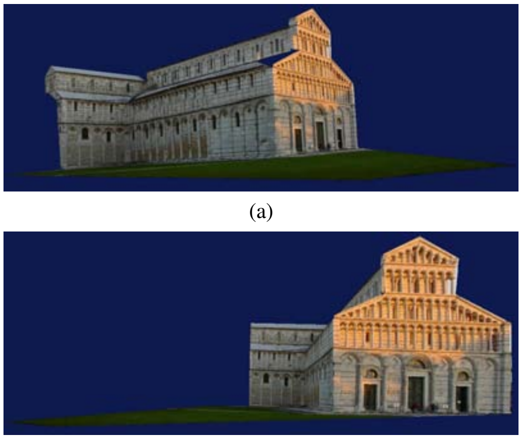


Figure 1: *An image of the Duomo of Pisa, Italy with the polylines defining planes to be reconstructed and their corresponding plane numbers overlaid.*

To facilitate a visual comparison between final and initial results, Figs. 3(a) and (b) show top views of the wireframe model reconstructed with the proposed method from slightly different viewpoints. Using similar viewpoints, Figs. 3(c) and (d) show top views of the wireframe model reconstructed using [SM99]. 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 [SM99], where such constraints are violated. For instance, as can be seen from Figs. 3(c) and (d), right angles between walls 1-2 and 0-1



(a)

(b)

Figure 2: *(a),(b) Novel views of the VRML model reconstructed from Fig. 1. The hole visible in (a) is caused by self-occlusion in the input image.*

have not been recovered correctly. Furthermore, it can be verified from Fig. 3(c) that the recovered shapes of the two roofs in the far left part of the image (planes 5 and 10) are imprecise.

6. Conclusion

This paper has presented a method for constraint-based SVR. The method starts by obtaining a preliminary reconstruction and then refines it in a constrained minimization framework, which ensures that user-specified constraints are accurately satisfied. 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.

References

[GOSV02] Grossmann E., Ortin D., Santos-Victor J.: Single and multi-view reconstruction of structured scenes. In *Proc. of ACCV'02* (2002), pp. 228–234. 1

[HAA97] 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* (New York, USA, 1997), ACM Press, pp. 225–232. 1

[HEH05] Hoiem D., Efros A., Hebert M.: Automatic Photo Pop-Up. In *Proc. of SIGGRAPH'05* (New York, USA, 2005), ACM Press, pp. 577–584. 1

[HZ00] Hartley R., Zisserman A.: *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2000. 1, 2

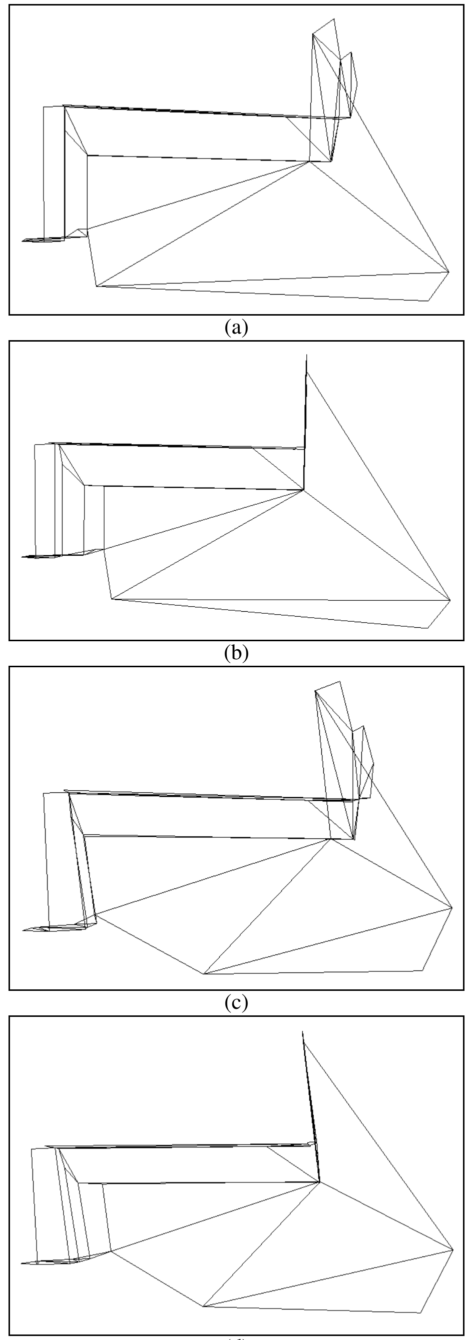
[LCZ99] Liebowitz D., Criminisi A., Zisserman A.: Creating Architectural Models from Images. *Eurographics 18*, 3 (Sep. 1999), 39–50. 1

[LZ98] Liebowitz D., Zisserman A.: Metric Rectification for Perspective Images of Planes. In *Proc. of CVPR'98* (Santa Barbara, CA, Jun. 1998), pp. 482–488. 3

[LZ99] Liebowitz D., Zisserman A.: Combining Scene and Auto-calibration Constraints. In *Proc. of ICCV'99* (Kerkyra, Greece, Sep. 1999), pp. 293–300. 2, 3

[NW91] 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), Dec. 1991. Available at `http://www.zib.de/Publications/abstracts/TR-91-10/`. 3

[SM99] Sturm P., Maybank S.: A Method for Interactive 3D Reconstruction of Piecewise Planar Objects from Single Images. In *Proc. of BMVC'99* (Nottingham, UK, Sep. 1999), pp. 265–274. 1, 2, 3, 4



(a)

(b)

(c)

(d)

Figure 3: *Top views of the wireframe model corresponding to the reconstruction obtained with the proposed method {(a)-(b)} and with that of [SM99] {(c)-(d)}.*