Multi-Level 3D Reconstruction with Visibility Constraints

George T. Chou* **Seth Teller**
Computer Graphics Group
MIT Laboratory for Computer Science
Cambridge, MA 02139

Abstract

In this paper, we examine the problem of 3D reconstruction from a large number of images annotated with camera pose information. Our proposed method involves computing 3D descriptions of the scene at multiple levels of representation (vertex, line segment, and surface) simultaneously. A major problem encountered during bottom-up processing is the uncertainty of low-level feature correspondence, which often results in mis-matches. We observe that visibility constraints from surface-level information can help in resolving ambiguity. This motivates our design of a hypothesis-based 3D correspondence system that integrates information from multiple levels. Experimental results on reconstructing urban scenes are presented.

1 Introduction

Large-scale 3D reconstruction is a challenging problem in computer vision. In this paper, we assume our input images are controlled, i.e. annotated with information on each camera's orientation and position. This allows us to perform reconstruction in a spatially focused manner, where only images viewing a certain 3D region of interest are selected for processing. For flexibility and efficiency, we also want our system to be able to process a large set of photographs taken from very different viewpoints. Most existing reconstruction systems do not have the latter capability. They typically require the input as a continuous image stream for easy tracking of features [4] [6] [7].

On the output side, we seek to produce more than just low-level descriptions such as sparse 3D vertices and line segments. Mid-level descriptions in the form of 3D surfaces are desired. There are two important reasons for this. First, higher representations are more compact and meaningful than low-level representations, making them better suited for visualization and subsequent processing. Second, information from higher levels can provide important constraints on low-level interpretations. In other words, there is the potential for synergy when processes from different levels interact. This aspect of hierarchical processing has been explored by researchers in stereovision [2] [5] [8].

Our construction of 3D surface representation begins with the correspondence of low-level features. In the system, this is achieved by intersecting the back-projections of features, which we call *extrusions*, in the 3D space. Unfortunately, mis-matches occur with 3D correspondence just as in 2D. The solution we proposed earlier, based on comparing the structural properties of image features [1], is inadequate for dealing with large viewpoint changes. Instead, we observe that surface-level descriptions enable us to predict the presence of occlusion, which can be used to eliminate invalid correspondences. This provides the key to resolving mis-matches in our system.

*Funding for this research was provided by the Advanced Research Projects Agency of the Department of the Defense under contract DACA76-97-K-0002.

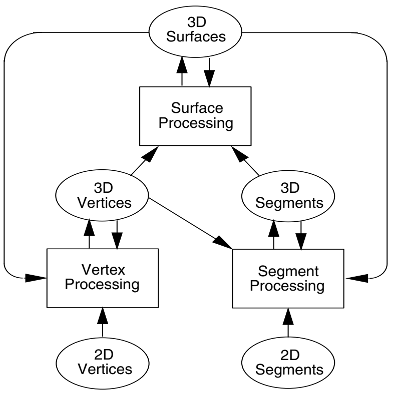

Figure 1: Flow of information in the system

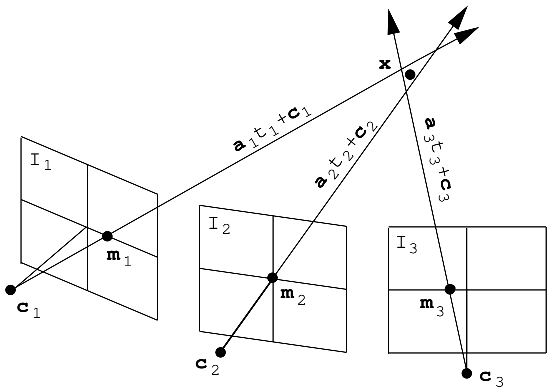

Figure 2: Intersection of vertex extrusions

The flow of information in the system is shown in Figure 1. Low-level and mid-level descriptions are computed simultaneously, so that the latest surface information can be used to correct and refine correspondences made during vertex and line processing. Hypothesis testing provides a convenient method for managing the uncertainty of early-stage correspondences. High-level descriptions (objects, parts) are not addressed in this paper.

2 Basic Operations

To be able to process images with significant viewpoint changes, feature correspondences must be determined in 3D space. This section presents some of the basic 3D operations used in our system.

2.1 Vertex Extrusion Intersection

Suppose we are given the corresponding image positions $\mathbf{m}_i$ of a 3D point $\mathbf{x}$ projected onto a set of images I_i. A vertex extrusion is the 3D ray emanating from the camera center $\mathbf{c}_i$, passing through image vertex $\mathbf{m}_i$, with direction $\mathbf{a}_i$ (see Figure 2). The 3D point $\mathbf{x}$ can be estimated by minimizing the sum of squared distances between the extrusions and the point:

$$D_V(\mathbf{x}) = \sum_i \|(\mathbf{a}_i\ t_i + \mathbf{c}_i) - \mathbf{x}\|^2 \qquad (1)$$

An estimate of $\mathbf{x}$ can be obtained by solving the linear system

$$\begin{aligned} &\sum_i (\mathbf{a}_i\mathbf{a}_i^T - \mathbf{I})^T(\mathbf{a}_i\mathbf{a}_i^T - \mathbf{I})\ \mathbf{x} \\ &= \sum_i (\mathbf{a}_i\mathbf{a}_i^T - \mathbf{I})^T(\mathbf{a}_i\mathbf{a}_i^T - \mathbf{I})\mathbf{c}_i \end{aligned} \qquad (2)$$

Previously, we described a solution based on SVD and back-substitution [1]. The residual of the intersection process is given by $D_V(\mathbf{x})$ in Equation (1), with $t_i = \mathbf{a}_i^T(\mathbf{x} - \mathbf{c}_i)$.

2.2 Segment Extrusion Intersection

For line segments, the computation is more involved. A segment extrusion is the 3D wedge emanating from the camera center $\mathbf{c}_i$, passing through image line segment L_i, and bounded by vectors $\mathbf{v}_{2i-1}$ and $\mathbf{v}_{2i}$ (see Figure 3). The normal of the extrusion plane is given by $\mathbf{e}_i = \mathbf{v}_{2i-1} \times \mathbf{v}_{2i}$.

Our first task is to estimate the 3D line $L : \mathbf{v}t + \mathbf{p}$ at the intersection of extrusions. Its direction $\mathbf{v}$ can be estimated as the eigenvector with the smallest eigenvalue in

$$\mathbf{A}^T\mathbf{A}\ \mathbf{v} = \lambda\ \mathbf{v} \qquad (3)$$

where $\mathbf{A}$ is a matrix whose rows are the normals $\mathbf{e}_i$ of the extrusion wedges. This problem can be solved by SVD.

Next, we estimate a fixed point $\mathbf{p}$ on the 3D line L. First, intersect every extrusion plane $(\mathbf{e}_i, \mathbf{c}_i)$, with the reference plane $(\mathbf{v}, \mathbf{c}_1)$, which has the normal $\mathbf{v}$ and contains the point $\mathbf{c}_1$.

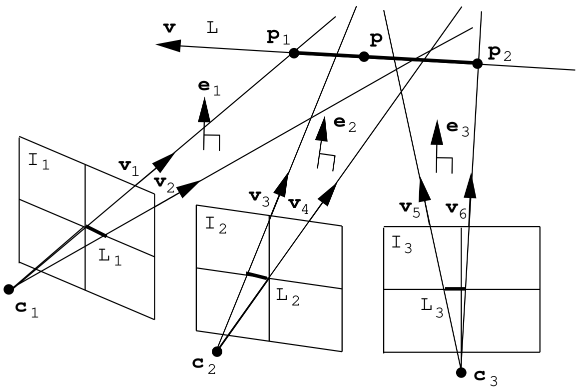

Figure 3: Intersection of segment extrusions

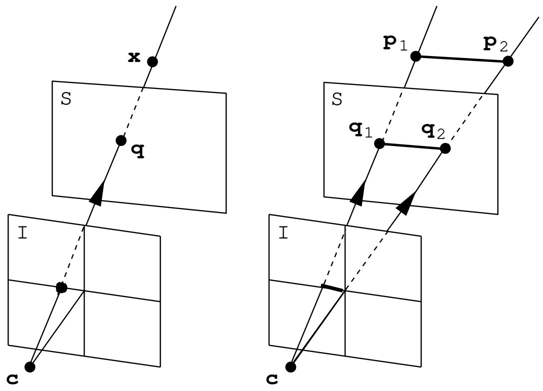

Figure 4: Test of intersection visibility

This produces a set of lines l_i lying on the reference plane. The point $\mathbf{p}$ is estimated to be the 3D point of intersection of lines l_i.

The residual of the intersection process is computed as the sum of squared distances between the extrusion rays and the 3D line L:

$$D_L(\mathbf{v}, \mathbf{p}) = \sum_j \|(\mathbf{v}_j\ s_j + \mathbf{c}_i) - (\mathbf{v}\ t_j + \mathbf{p})\|^2 \quad (4)$$

where $i = \lfloor \frac{j+1}{2} \rfloor$, and

$$s_j = \frac{(\mathbf{v} \times \mathbf{v}_j) \cdot ((\mathbf{c}_i - \mathbf{p}) \times \mathbf{v})}{\|\mathbf{v} \times \mathbf{v}_j\|^2} \quad (5)$$

with the constraint of $s_j \geq 0$, and

$$t_j = \frac{(\mathbf{v} \times \mathbf{v}_j) \cdot ((\mathbf{c}_i - \mathbf{p}) \times \mathbf{v}_j)}{\|\mathbf{v} \times \mathbf{v}_j\|^2} \quad (6)$$

Once the 3D line L is defined, the solution line segment $(\mathbf{p}_1, \mathbf{p}_2)$ can be estimated as follows. First, determine the union of line extrusions, and make sure the union is continuous. Then intersect the two rays bounding the union with the 3D line L. The endpoints $\mathbf{p}_1$ and $\mathbf{p}_2$ are given by $\mathbf{p}_j = \mathbf{v}\ t_j + \mathbf{p}$, where t_j are computed by Equation (6) with $\mathbf{v}_j$ as the bounding rays.

2.3 Planar Surface Estimation

Surface processing is activated whenever there are two or more 3D line segments linked in a chain by their endpoints. Let $\mathbf{x}_i$ be the set of endpoints from segments that define a planar surface. We can estimate the parameters of the plane, its normal $\mathbf{n}$, and a point $\mathbf{p}$ on the plane by finding the best-fitting plane to the points $\mathbf{x}_i$.

We estimate the point $\mathbf{p}$ as the mean of $\mathbf{x}_i$. The normal vector $\mathbf{n}$ is estimated by minimizing the sum of squared distances between the points and the plane:

$$D_S(\mathbf{n}, \mathbf{p}) = \sum_i \mathbf{n}^T(\mathbf{x}_i - \mathbf{p})(\mathbf{x}_i - \mathbf{p})^T\mathbf{n} \quad (7)$$

To solve this minimization problem requires finding the eigenvector $\mathbf{n}$ with the smallest eigenvalue in the following equation:

$$\mathbf{B}\ \mathbf{n} = \lambda\ \mathbf{n} \quad (8)$$

where matrix $\mathbf{B} = \sum_i (\mathbf{x}_i - \mathbf{p})(\mathbf{x}_i - \mathbf{p})^T$. The solution $\hat{\mathbf{n}}$ can be determined by SVD. The residual of the plane fitting process is computed by substituting the estimates of $\mathbf{n}$ and $\mathbf{p}$ into Equation (7).

2.4 Intersection Visibility Test

Feature extrusions can intersect in 3D space by accident, especially when the number of participating extrusions is small (e.g. 2 or 3). Interestingly, we have observed that most accidental intersections occur in violation of a simple visibility constraint – that feature extrusions should not penetrate surfaces to form intersections. Assuming no transparent surfaces in the scene, violation of the visibility constraint can be detected as follows.

For each vertex extrusion, we first compute point $\mathbf{q}$ on the planar surface S, intersected by the extrusion ray (see Figure 4, left). Next we compute d, the distance from camera center $\mathbf{c}$ to surface point $\mathbf{q}$ minus the distance from camera center $\mathbf{c}$ to vertex intersection $\mathbf{x}$:

$$d = \|\mathbf{q} - \mathbf{c}\| - \|\mathbf{x} - \mathbf{c}\| \tag{9}$$

If $d > 0$ (point $\mathbf{x}$ is closer to the camera than $\mathbf{q}$), then there is no violation of the visibility constraint. Otherwise, there may be a violation.

For each segment extrusion, we first compute the line segment $(\mathbf{q}_1, \mathbf{q}_2)$ on the planar surface S, intersected by the extrusion wedge (see Figure 4, right). Next we compute d, the distance from camera center $\mathbf{c}$ to surface segment $(\mathbf{q}_1, \mathbf{q}_2)$ minus the distance from camera center $\mathbf{c}$ to line segment $(\mathbf{p}_1, \mathbf{p}_2)$:

$$d = \|(\mathbf{q}_1, \mathbf{q}_2) - \mathbf{c}\| - \|(\mathbf{p}_1, \mathbf{p}_2) - \mathbf{c}\| \tag{10}$$

If $d > 0$ (the segment $(\mathbf{p}_1, \mathbf{p}_2)$ is closer to the camera than $\mathbf{q}$), then there is no violation of the visibility constraint. Otherwise, there may be a violation.

3 Correspondence Hypothesis

Due to many factors, some already mentioned, early-stage correspondences tend to be uncertain. In this section, we describe a hypothesis-based approach for managing uncertain correspondences.

3.1 Probability Models

Suppose we are given low-level features y from N distinct images and information z on surfaces in the scene. Let H denote the hypothesis that the image features arise from a single 3D object (vertex or line segment). We model the likelihood probability of image features and surfaces given feature correspondence as a product of three factors:

$$p(y, z \mid H) = \alpha\ f_1(y)\ f_2(y)\ f_3(y, z) \tag{11}$$

where α is a normalization factor. The other factors are described below.

The data consistency constraint $f_1(y)$ is a Gaussian function of the intersection error of feature extrusions in 3D:

$$f_1(y) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(\frac{-D(y)/N}{2\sigma^2}\right) \tag{12}$$

where N is the number of features, $D(y)$ is the intersection error as computed by D_V in Equation (1) for vertices, or D_L in Equation (4) for line segments. The standard deviation σ is replaced by σ_v for vertices and σ_l for line segments. This factor favors hypotheses with small intersection error.

The non-accidentalness constraint $f_2(y)$ is a saturation function of the number of features N:

$$f_2(y) = \frac{(N-1)^2}{\alpha^2 + (N-1)^2} \qquad N \geq 1 \tag{13}$$

where α is the half-saturation constant. When N=1, $f_2(y)$ is zero. As N gets large, $f_2(y)$ approaches one. This factor favors hypotheses with many supporting features.

The visibility constraint $f_3(y, z)$ is a product of logistic functions of the relative distances d_i computed in Equation (9) for vertex extrusions, or Equation (10) for segment extrusions:

$$f_3(y, z) = \prod_{i=1}^{N} \frac{1}{1 + \exp(-\beta\ d_i)} \tag{14}$$

where index i loops through the extrusions. Because of the product operation, the entire term goes to zero if there is any d_i<<0, which occurs when extrusion i violates the visibility constraint. Hence, this factor favors hypotheses which uphold the visibility constraint.

3.2 Hypothesis Formation

The posterior probability of feature correspondence given low-level feature information y and mid-level surface information z is

$$p(H \mid y, z) = \frac{p(y, z \mid H)\ p(H)}{p(y, z)} \tag{15}$$

Let $\overline{H}$ denote the null hypothesis that the image features do not correspond. According to *maximum a posteriori* (MAP) decision rule, the correspondence hypothesis H is accepted and $\overline{H}$ rejected if

$$p(H \mid y, z) > p(\overline{H} \mid y, z) \tag{16}$$

Otherwise H is rejected and $\overline{H}$ is accepted. This condition is equivalent to

$$p(y, z \mid H) > p(y, z \mid \overline{H}) \, \frac{1 - p(H)}{p(H)} \tag{17}$$

We assume that $p(y, z \mid \overline{H}) = p(y) \; p(z)$ is a constant, i.e. when there is no correspondence, y and z are conditionally independent and constant. We also assume $p(H)$ to be a constant prior. Then the entire right hand side of Equation (17) reduces to a constant, which we refer to as p_h.

Thus, the criterion for hypothesis formation is

$$p(y, z \mid H) > p_h \tag{18}$$

3.3 Hypothesis Confirmation

Once a hypothesis has been formed, an increase in confidence is usually due to an increase in feature support. Whenever a hypothesis is updated, we test to see if

$$p(y, z \mid H) > p_e \tag{19}$$

where $p_e >> p_h$, and p_e is a threshold controlling the output of elements. If the above condition is true, the correspondence hypothesis is considered highly reliable, and is promoted to a 3D element.

When a hypothesis is promoted into an element, all of its supporting features become committed to it immediately. This means that these features must withdraw their support to other hypotheses. In essence, hypotheses which share common low-level features are in competition for the commitment of the features.

3.4 Hypothesis Rejection

Confidence in a hypothesis may be decreased for three possible reasons: (1) when the number of supporting features is reduced due to the promotion of a competing hypothesis; (2) when the intersection error becomes too large due to the inclusion of new features; or (3) when the hypothesis is found to violate the visibility constraint of a newly formed surface.

The criterion for hypothesis rejection is

$$p(y, z \mid H) < p_h \tag{20}$$

4 Incremental Processing

Our system processes input features incrementally without any ordering constraint. As features are inserted into the system, new hypotheses are generated at a high rate. To control computational complexity, we need to minimize the number of hypotheses generated. This can be achieved by matching new features with existing elements and hypotheses first, before testing them with unmatched extrusions.

The sequence of processing for each new feature inserted into the system is as follows:

1. Construct a feature (vertex/segment) extrusion. If the segment extrusion intersects two vertex elements with its bounding vectors, a segment element is formed directly.
2. Intersect the extrusion with any existing 3D element. If the result passes the condition of Equation (19), commit the feature to the element.
3. Intersect the extrusion with any existing 3D hypothesis. If the result passes the condition of Equation (18), link the feature with the hypothesis. Check for promotion of the hypothesis to an element.
4. Intersect the extrusion with any unmatched extrusion of the same type. If the result passes the condition of Equation (18), form a new 3D hypothesis.

When a 3D hypothesis is promoted to a 3D element, the following sequence is initiated:

1. Commit all supporting extrusions to the new element. Remove existing links between these features with other hypotheses.

Check for rejection of affected hypotheses via the condition of Equation (20). The processing for vertex elements ends here.

2. Check to see if the two endpoints of the segment element are linked to other segment elements. If so, a new surface hypothesis is formed for each connected segment pair.

3. If there are two or more surface hypotheses sharing the same line segment, check to see if the hypotheses can be merged by comparing their plane normals.

When two adjacent surface hypotheses are merged, the following sequence is initiated:

1. Estimate the parameters of the new surface hypothesis based on vertices from line segments of the merging hypotheses. Update the links to supporting segment elements.

2. Check for closure of the surface polygon by following the segment elements. If closed, promote the surface hypothesis to a surface element. For vertical surfaces, closure at the ground level is not required.

3. If a new surface element has formed, check to see if any existing vertex or segment hypothesis violates the surface's visibility constraint. Remove the violating extrusions and update the hypotheses accordingly.

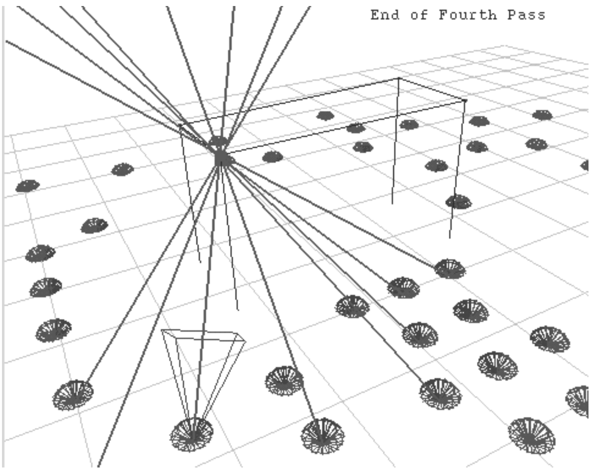

Figure 5: An intersection of vertex extrusions

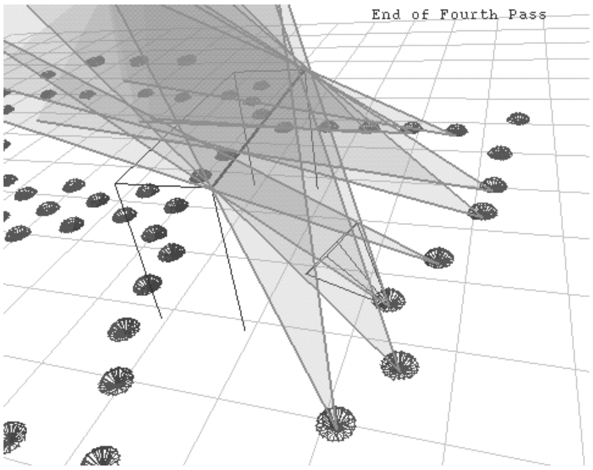

Figure 6: An intersection of segment extrusions

5 Experimental Results

We tested our system on roughly 4,000 RGB photographs collected from over 81 positions in and around Technology Square. Details on the collection process and the pose estimation technique can be found in [3].

Automatic reconstruction results have been obtained for two of the buildings. Figure 5 shows several vertex extrusions intersecting to form a 3D vertex element. Figure 6 shows several segment extrusions intersecting to form a 3D segment element. Figure 7 shows the reconstructed surfaces and segment elements for one of the buildings. Note that the roof surface is formed automatically from the surrounding line segments by the algorithm.

Figure 8 shows the segment hypotheses generated without the visibility constraint. Figure 9 shows the segment hypotheses generated with the visibility constraint. The most visible effect of applying the visibility constraint is that many fewer hypotheses are generated.

We also compiled statistics on the effects of enforcing the visibility constraint. The data displayed in Table 1 represent the state of the system at the end of reconstruction for two of the buildings. There are several things to note in the table:

Visibility Constraint	No. of Elem.	Elem. Extru.	Extru. per Elem.
Off	16	130	8.13
On	16	125	7.81
Visibility Constraint	No. of Hypo.	Hypo. Extru.	Extru. per Hypo.
Off	674	2251	3.34
On	131	493	3.76

Table 1: Effects of the visibility constraint

1. Applying the visibility constraint decreases the average number of supporting Extrusions per Element (EpE) to 7.81 from 8.13. This is consistent with the clipping effect of the constraint.

2. Applying the visibility constraint reduces the total number of hypotheses to 131 from 674. As illustrated in Figures 8 and 9, hundreds of mis-matches that violate the visibility constraint are no longer accepted as hypotheses.

3. Applying the visibility constraint increases the average number of supporting Extrusions per Hypothesis (EpH) to 3.76 from 3.34. Contrast this with the effect on elements (EpE), which was decreased. The key here is that mis-matches which violate the visibility constraint tend to have fewer supporting extrusions than normal hypotheses. When they are eliminated from the pool of hypotheses, EpH is raised as a result.

Although our system performs well on two of the buildings in Tech Square, it has difficulties with the other two buildings. The main problem lies with unreliable feature detection. Due to the inconsistent and highly cluttered nature of feature observations in these cases, a large number of hypotheses are formed, but few elements are generated. It has been suggested that perceptual grouping can help in reducing the effects of inconsistency and clutter [2] [8]. This points to a possible direction for future work.

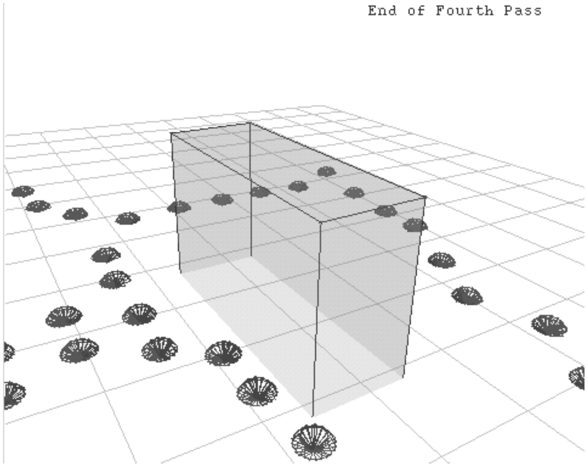

Figure 7: Reconstructed surfaces and segment elements of a building

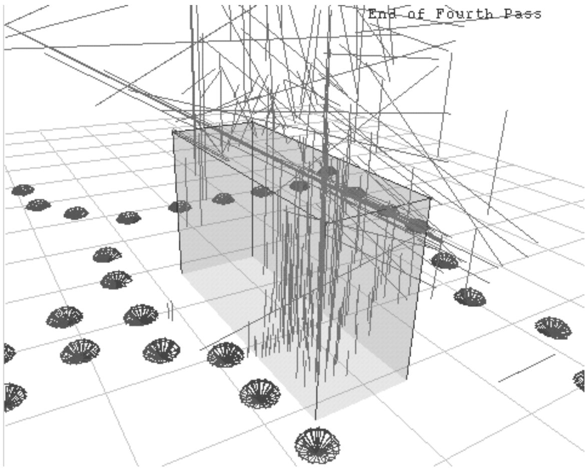

Figure 8: Segment hypotheses generated without the visibility constraint

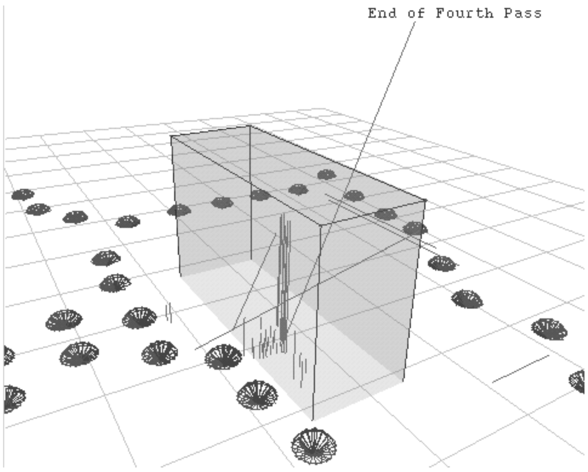

Figure 9: Segment hypotheses generated with the visibility constraint

6 Conclusion

We have presented an automated system for feature-based 3D reconstruction using a large set of photographs annotated with pose information. Descriptions of the scene are computed at the vertex, line segment, and surface levels simultaneously. A hypothesis-based approach has been proposed for modeling uncertainty in low-level feature correspondences. Integration of bottom-up extrusion information with top-down visibility constraint is effective in reducing mis-matches.

References

[1] G. T. Chou and S. Teller, Multi-Image Correspondence using Geometric and Structural Constraints, *Proc. Image Understanding Workshop*, pp. 794-808, 1997.

[2] R. Chung and R. Nevatia, Use of Monocular Groupings and Occlusion Analysis in a Hierarchical Stereo System, *Computer Vision and Image Understanding*, 62(2), pp. 245-268, 1995.

[3] S. Coorg, N. Master and S. Teller, Acquisition of a Large Pose-Mosaic Dataset, *Proc. Conf. on Computer Vision and Pattern Recognition*, pp. 872-878, 1998.

[4] M. Herman and T. Kanade, The 3D Mosaic Scene Understanding System, *Artificial Intelligence*, 30, pp. 289-341, 1986.

[5] H. S. Lim and T. O. Binford. Structural Correspondence in Stereo Vision, *Proc. Image Understanding Workshop*, pp. 794-808, 1988.

[6] L. Matthies, T. Kanade and R. Szeliski. Kalman Filter-based Algorithms for Estimating Depth from Image Sequences, *Intl. Journal of Computer Vision*, 3, pp. 209-236, 1989.

[7] R. Szeliski and S. B. Kang, Recovering 3D Shape and Motion from Image Streams using Non-Linear Least Squares, *Journal of Visual Communication and Image Representation*, 5(1), pp. 10-28, 1994.

[8] V. Venkateswar and R. Chellappa, Hierarchical Stereo and Motion Correspondence Using Feature Groupings, *Intl. Journal of Computer Vision*, 15, pp. 245-269, 1995.