Planar Scene Modeling from Quasiconvex Subproblems

Visesh Chari[1,2], Anil Nelakanti[1,3], Chetan Jakkoju[1], and C.V. Jawahar[1]

[1] Center for Visual Information Technology,
International Institute of Information Technology, Hyderabad, India-500032.
[2] INRIA Rhône Alpes Grenoble, France.
[3] MirriAd Limited, London, UK

Abstract. In this paper, we propose a convex optimization based approach for piecewise planar reconstruction. We show that the task of reconstructing a piecewise planar environment can be set in an L_∞ based Homographic framework that iteratively computes scene plane and camera pose parameters. Instead of image points, the algorithm optimizes over inter-image homographies. The resultant objective functions are minimized using Second Order Cone Programming algorithms. Apart from showing the convergence of the algorithm, we also empirically verify its robustness to error in initialization through various experiments on synthetic and real data. We intend this algorithm to be in between initialization approaches like decomposition methods and iterative non-linear minimization methods like Bundle Adjustment.

1 Introduction and Related Work

In this paper, we describe a convex optimization based approach for piecewise planar reconstruction by optimizing inter-image homographies. This work is motivated by both the recent success of convex optimization based methods in various geometric problems like triangulation, resectioning [1, 2], and the available sophistication in robust estimation of homographies across views [2].

Convex optimization methods have achieved recent success in the estimation of various geometric quantities like homography, pose, 3D point cloud (triangulation) [1, 2] etc., and are even shown to be reasonably robust to noise [2]. There are even works on outlier estimation and removal using convex optimization [3] . On the other hand, there also has been progress on robust estimation of homographies from multiple views of a scene plane [2]. However, even though homographies can also be expressed as a function of the camera pose, and can be decomposed using SVD in a similar manner to fundamental matrices [4, 5], piecewise planar reconstruction as a 3D reconstruction pipeline has not received much attention.

To this extent, we intend to develop an algorithm that can be a useful "bridge" between SVD based initialization methods mentioned above and non-linear optimization methods like Bundle Adjustment (BA). We focus on the iterative reconstruction process, that alternates between optimizing a six parameter camera pose vector for each view, and a four element plane parameter vector for each scene plane, by optimizing over the resulting inter-image homographies.

We make the following contributions in this work. First, we introduce objective functions for producing optimal estimates of pose and plane parameters, along the lines

of [2]. Then, we show how a Branch and Bound (BnB) algorithm may be formulated for the computation of optimal rotation between views [4].

Some of the recently proposed frameworks on L_∞ based quasi-convex cost functions problems form the motivation for our work [1, 6], while some closely related works include projective Bundle Adjustment (pBA) [7] and BA with constraints [8]. However, we differ from these works in the kinds of objective functions minimized (quasiconvex as opposed to non-linear) and in the quantities we optimize (homographies as opposed to 3D points). Recent study of bi-linear problems also has relevance to our work [9] since plane and pose parameters are combined together in a bi-linear form in the expansion of a homography (Equation 1). However, the formulation proposed in [9] requires that the entire set of plane and pose parameters need to be optimized together. Also, estimation of rotation parameters becomes infeasible in such a scenario. Thus we do not resort to a formulation along the lines of [9].

The rest of this paper is organized in the following manner. Section 2 sets the problem of pose estimation in a homographic framework and motivates the need for the use of optimization. Section 3 presents our solution and algorithm details. Experimental analysis on synthetic and real-world sequences are done in Section 4 and finally, we conclude with a discussion on future directions and applications in Sections 5.

2 SVD based Initializations

Let there be m planes in the world, characterized by the parameters $\left[n^1, d^1, \ldots, n^m, d^m\right]$. The jth plane is characterized by the parameters (n^j, d^j), where n^j represents the normal of the plane and d^j represents the perpendicular distance from world origin. Let there be two cameras with external parameters $[\mathbf{I} \mid \mathbf{0}]$ and $[\mathbf{R} \mid \mathbf{t}]$. For simplicity, let us assume that the internal parameters of the cameras are set to identity ($\mathbf{K} = \mathbf{I}$). Thus the homography induced by the jth plane between the two views [10] is given by

$$\mathrm{H^j} = \left[\mathbf{R} \; - \; \frac{\mathbf{t} {n^j}^T}{d^j} \right] \tag{1}$$

Decomposition algorithms for obtaining camera pose and plane normals from homography matrix using Equation 1 are well known [11, 5]. However, since, the process of pose computation from correspondences through the homography matrix involves two SVDs, a theoretical sensitivity analysis of such algorithms is difficult and approximate [12]. Thus it is more advantageous to do an empirical study of the error in the estimation of plane and pose parameters, given noise in image correspondences.

Figures(1a-1c), depict the poor performance of one of the SVD based decomposition algorithms [5]. The experiments consisted of adding increasing amounts of noise to a previously determined set of normalized image correspondences. Homographies obtained after a standard RANSAC routine were then decomposed to obtain estimates of the plane and pose parameters. Variances are plotted against error in pixel coordinates, with a maximum variance of 5 pixels which corresponds to approximately 1% of the image size. As can be seen, translation and normal estimations are adversely affected by image noise. The errors for the other algorithm [11], were similar.

The variances in Figures(1a) plot the error in estimation of rotation parameters when noise is introduced into the system. As is seen, the maximum variation of rotation parameters in the Euler angle space is 6 degrees, for as high as one percent image noise. Comparison with the translation and normal errors, which are as high as 40 degrees in the polar space Figures(1b-1c), show that the decomposition algorithm produces much more robust estimates of rotation than either translation or normal parameters. This explains the greater need for better estimates of translation and normal parameters compared to that of rotation parameters that are much close to the actual values.

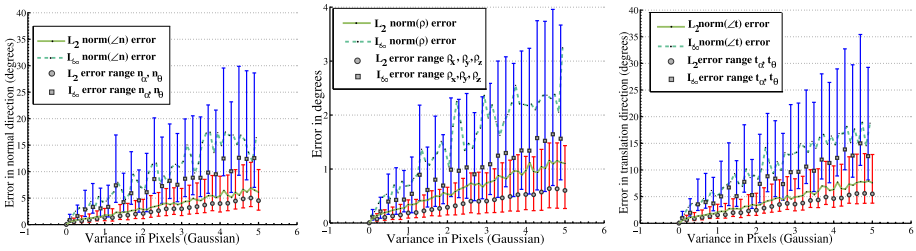


Fig. 1: (a,b,c) Plot the $L2$ and L_∞ errors in the rotation angles, translation direction and normal direction respectively. Also are plotted the maximum error ranges for these quantities. The translation and normal direction errors are computed as Euclidean distances in polar space.

3 Optimization Framework

In this section, we describe our algorithm. First, we start with the simple case when rotation is assumed known, and the rest of the parameters are optimized (Section 3.1). The reason for this is the non-convexity of the orthonormality constraints of the rotation matrix. Since algorithms for estimating the rotation already exist [4], and since we have shown rotation parameters to be robustly recovered from SVD decompositions as compared to other parameters (Figure 1a), we treat rotation separately (Section 3.3). Finally, in order to bring all the SVD decomposition estimates into a single coordinate system, we describe a convex function in Section 3.2.

3.1 Formulation of the Objective Function

We wish to find plane and pose parameters that best fits Equation 1 which is non-linear in terms of quantities $(\mathbf{R}, \mathbf{t}, n^j, d^j)$ that need to be computed. However, observe that when either the plane or the pose parameters are known, Equation 1 is linear in the remaining unknowns. This simple fact is used to define an objective function that measures the geometric distance between the homography computed from plane/pose parameters and the homography estimated from point correspondences. If the homography matrix with varying pose parameters and fixed plane parameters is defined as $\mathcal{H}rt^j = \left[\mathbf{R} - \frac{\mathbf{t}{n_c^j}^T}{d_c^j}\right]$ for the jth plane then the corresponding objective function is

$$\mathcal{F}_{(\mathbf{R},\mathbf{t})} = \sum_{i=1}^{8} \frac{H_i^j}{H_9^j} - \frac{\mathcal{H}rt_i^j}{\mathcal{H}rt_9^j} \tag{2}$$

Similarly, when the plane parameters are allowed to vary fixing pose parameters the homography function is $\mathcal{H}nd^j = \left[\mathbf{d^j} R_c - t_c \mathbf{n^j}^\top\right]$ and the objective function

$$\mathcal{F}_{(\mathbf{n},\mathbf{d})} = \sum_{i=1}^{8} \frac{H_i^j}{H_9^j} - \frac{\mathcal{H}nd_i^j}{\mathcal{H}nd_9^j} \tag{3}$$

(R_c, t_c, n_c^j, d_c^j) are fixed and the optimization runs over free variables denoted by bold letters. There are two important observations to make at this point. Firstly, equations (2, 3) are both linear fractional: both the numerator and denominator are affine in terms of the unknowns. Secondly, it is possible to optimize all parameters by alternatively minimizing Equation 2 and Equation 3 till convergence.

The proposed algorithm is a two step process. An initial estimate of the parameters is acquired using SVD-based decomposition in the first. However, estimates from SVD decomposition in the first step do not all have the same scale factor. Such estimates need to be threaded together and brought down to a common universal scale before carrying out the optimization. This is done by minimizing the difference between various estimates of a single quantity as described in Section 3.2.

Subsequently, in the second step, this estimate is improved in an optimization framework. However, minimizing Equation 2 without enforcing the constraints inherent to a rotation matrix will not lead to a physically valid rotation matrix. Equation 2 fails to be a linear fractional with rotation constraints enforced complicating its minimization. Hence, rotation is handled separately as explained in Section 3.3 and Equation 2 is minimized by varying only the translation as in Step 7 of Algorithm 1.

The optimization takes advantage of the fact that the objective functions are quasiconvex and employs convex optimization techniques at minimizing them. Variables t^i and (n^j, d^j) are minimized in alternating iterations. Optimization of t^i takes into account information from all visible planes. Similarly, optimization for (n^j, d^j) is done with information from all views in which the plane is visible. This two step process ensures the quasiconvexity of the objective functions. The complete method is summarized in Algorithm 1.

Algorithm 1 Complete Algorithm Summarized.

1: Input: Homographies ${}^kH_j$ for $j = 1, \ldots, J$ and $k = 1, \ldots, K$ of plane Π_j between the camera views ${}^kP$ and reference view ${}^0P = [I|0]$.

2: SVD-based decomposition: Decompose ${}^kH_j$ to get ${}^kR_j$, $\frac{{}^kt_j}{{}^kd_j}$, ${}^kn_j$.

3: Initialization: ${}^kR =$ median$_j$ $\{{}^kR_j\}$ and $t =$ median$_j\{{}^kt_j\}$.

4: Set to universal scale: Assume each actual camera translation to be a unit vector in the direction of $\frac{{}^kt}{d_j}$, i.e., $\|{}^kt\| = 1$. Let ${}^kG_j = [{}^kR - \frac{{}^kt{n_j}^T}{{}^kd_j}]$ and ${}^kG_j^s = (g_1, g_2, \ldots, g_9)^T$.

5: Iterative Minimization:

6: $\Sigma_k \Sigma_j \left\{{}^kH_j^s - {}^kG_j^s\right\} \leq \delta$

7: Update $({}^kt)$: $({}^kt) = \arg\min_{{}^kt} \max_{j=1}^{J} \sqrt{\Sigma_i [\frac{{}^jh_i}{{}^jh_9} - \frac{{}^jg_i}{{}^jg_9}]^2} \forall k = 1, \ldots, K.$

8: Update (n_j, d_j): $(n_j, d_j) = \arg\min_{n_j, d_j} \max_{k=1}^{K} \sqrt{\Sigma_i [\frac{{}^kh_i}{{}^kh_9} - \frac{{}^kg_i}{{}^kg_9}]^2} \forall j = 1, \ldots, J.$

3.2 Universal Scale

Each decomposition by the algorithms of Faugeras [11] and Zhang [5] produces estimates of $\{R, t, n\}$ assuming d (perpendicular distance of plane from origin) to be unity. Thus estimates vary by a scale factor and need to be tied down to a single universal scale which in the presence of noise has to be computed using optimization.

Let the solutions of translation obtained by decomposing homography H_i^j be t_i^j . Ideally, the actual translation is $t_i = t_i^j d^j$. Since various estimates of the same quantity must be consistent, we find an $x = \left[t_1, t_2, \ldots, t_k, d^1, d^2, \ldots, d^m\right]^\top$ for which an error $|f(x)|_\infty$ is minimum. $f(x)$ is a vector with elements of the set $\{t_i - t_i^j d^j \mid i \in [1, k], j \in [1, m]\}$ stacked up. Optimal estimates are found by performing the minimization $x^* = \arg\min_x |f(x)|_\infty$.

The considered error function is convex [13], made from the pointwise maximum of the convex function $(t_i - t_i^j d^j)$. An unconstrained optimization in this case could lead to the trivial solution of all zeros for x which is undesirable. To avoid this we *fix* perpendicular distance of anyone of the planes (say, d^1) to unity. This also sets the overall scale of the minimization process.

3.3 Retrieving Rotation

Constraints inherent to rotations and normals like orthonormality constraints of the rotation matrix are non-convex and do fit into a convex framework. Such constraints have been handled in the literature [4, 14] using under estimators and over estimators of the non-convex function with a Branch and Bound algorithm. We, thus, handle rotation separately rather than in the above optimization. We use image coordinates of planes available on the lines of [4] to solve for rotation R_i of the i^{th} view. The objective function to be minimized is

$$\mathcal{F}_{(\mathbf{R}_i, \mathbf{t}_i)} \equiv \mathbf{Find}(R_i, \mathbf{t}_i) \qquad \text{s.t. } \angle(H_i^j \mathbf{x}_1^j, (\mathbf{R}_i - \mathbf{t}_i \frac{{n^j}^T}{d^j})\mathbf{x}_1^j) < \epsilon_{min} \tag{4}$$

which can be alternatively posed as

$$\mathcal{F}_{(\mathbf{R}_i, \mathbf{t}_i)} \equiv \mathbf{Find}(\mathbf{R}_i, \mathbf{t}_i) \qquad \text{s.t. } \angle(H_i^j \mathbf{x}_1^j, \mathbf{R}_i(\mathbf{I} - \mathbf{t}_i \frac{{n^j}^T}{d^j})\mathbf{x}_1^j) < \epsilon_{min} \tag{5}$$

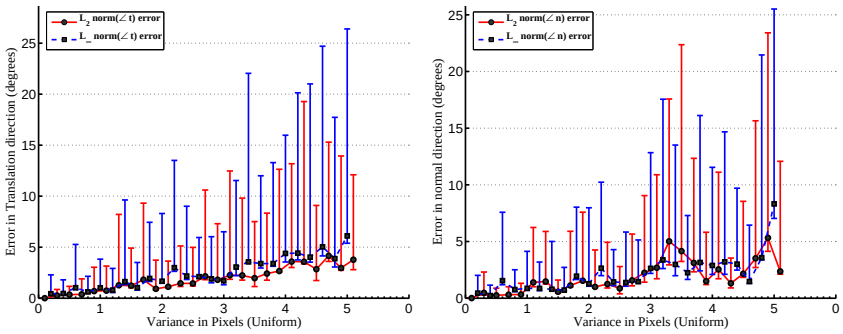
where x_1^j are points from the jth plane in the first view. Arguments of bounds and in general the branching strategy of [4] can now be incorporated into the current framework. The analysis that estimates of rotation from SVD-based methods are more robust than that of translations and normals as noted in Section 2 practically helps the idea of handling rotation separately at a later stage. Figure 3c shows the performance of the objective function described above in the presence of varying noise. The L_2 norm in angular space (roll-pitch-yaw) is plotted against increasing amounts of noise in image pixels.

4 Experimental Analysis

In order to test the proposed algorithm, we have designed experiments using SeDuMi [16] on both synthetic and real-world data. Synthetic data is obtained by generating points on planes and projecting them onto camera matrices. Real world data sets tested include the Oxford Model House, Corridor, and UNC datasets. In all these cases, the real world is assumed to be segmented into planes apriori *i.e.* interest points and hence correspondences computed are assumed to be clustered into planes. However, there are automatic algorithms to achieve such a classification [15].

4.1 Synthetic Data

Generation Random points are generated on the XY-plane which is then re-positioned at a random location. Two random camera matrices are generated and the world points of many such planes are projected using them to generate image points. Gaussian noise of varying standard deviation is added to these image points to create synthetic correspondence data. Homographies are then computed using the RANSAC after normalization [10] which can alternatively be generated by [1]. The generated Homographies are decomposed using Faugeras' and Zhang's algorithms [11, 5] to generate data for both initialization and comparison. Algorithm 1 is then run with this data, to produce our estimate and is compared with the SVD-based algorithms and Bundle Adjustment in the 6-parameter pose space by plotting the euclidean distance between estimated and ground truth values.



(a) Translation (b) Normal

Fig. 2: Plot of L_2 and L_∞ norms of the distance in pose space between estimated and ground truth quantities from Algorithm 1 against increase in variance of Gaussian error in point correspondences. Comparison with the two SVD based methods is shown.

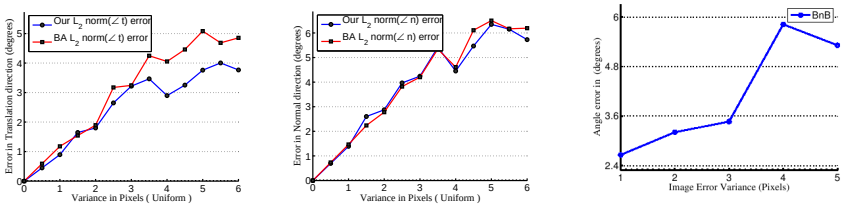
Effect of noise Figures (2a,2b) show the effect of increasing image noise on the accuracy of estimation. Two observations can be made for both translations and normals. First, the average error in the estimation of both parameters is less than 5 degrees even for a 1% error in the image coordinates, which is a considerable amount of error. This justifies the robustness of our algorithm to image noise. The second observation is that the mean errors (averaged for 100 trials) in all these cases are located close to the minimum errors represented by the lower end of the error bar. We can conclude that most

of the estimations center around the mean, with only a few deviating towards the higher end. Another interesting observation is that even the resilience to noise is apparent till about 3 pixel error after which the maximum error in both cases seems to increase. This can be attributed to the fact that after a point the algorithm possibly settles into a local minima because of the inaccurate initialization. However, this is still better than the results of SVD-based methods in Figures 1b, 1c.

Comparison with Bundle Adjustment We empirically compare our algorithm with standard iterative non-linear optimization technique of Bundle Adjustment (BA) [17], which uses Levenberg-Marquardt internally. BA is initialized by the output of the SVD-based approaches similar to ours. This initialization is used to minimize the following error over the normals and the translations

$$(R, t, n_j, d_j) = \arg \min_{{}^kR, {}^kt, n_j, d_j} \sum_k \sum_j \sum_i [\frac{h_i}{h_9} - \frac{x^T A_i x}{\overline{x}^T A_9 \overline{x}}]^2 \tag{6}$$

where, $x = ({}^1R^s, \ldots, {}^KR^s, {}^1t^T, \ldots, {}^Kt^T, n_1^T, \ldots, n_J^T, d_1, \ldots, d_J)$ and A_i is a matrix s.t. $x^T A_i x = g_i$ and $\overline{x}$ is x with the initial SVD estimates of ${}^kR, {}^kt, n_j, d_j$ substituted. The improvement in translations is shown in Fig (3a) and that of normals in Fig (3b). They are shown for varying levels of variance each of which has been tested for 100 trials. They clearly show our algorithm performing better than BA.



(a) Translation (b) Normal (c) Rotation

Fig. 3: (a-b)Plot of L_2 norm of the distance in pose space between estimated and ground truth quantities from Algorithm 1 and Bundle adjustment against increase in variance of Gaussian error in point correspondences.(c) Error in recovery of rotation parameters using the objective function of Section 3.3

Effect of planes and views Figures (4a,4c,4b,4d) show the effect of the number of planes and views on the performance of the algorithm. Contrary to intuition, increasing the number of planes does not seem to have much effect on the accuracy of the estimates of translation parameters. On the other hand, increasing the number of views increases the parameter size, and the accuracy of translation estimates dwindles since the number of planes and hence, measurements is kept constant. In the case of normals, however, increasing the number of views results in a marked improvement in the accuracy of their estimates.

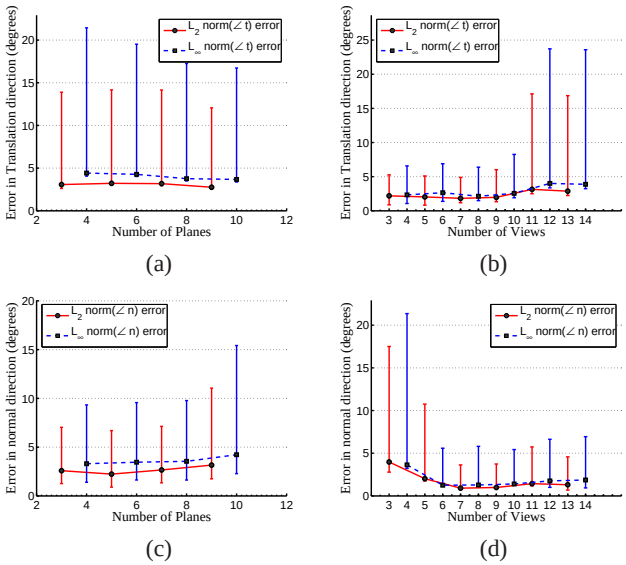


Fig. 4: The above figures plot the effect of planes and views on the accuracy in estimation of the translation and normal parameters. First two figures plot the effect on translations and last two plot the effect on normals. For the experiment with increasing planes, the number of views was kept constant at 10, and that for views, the number of planes was set to be 3.

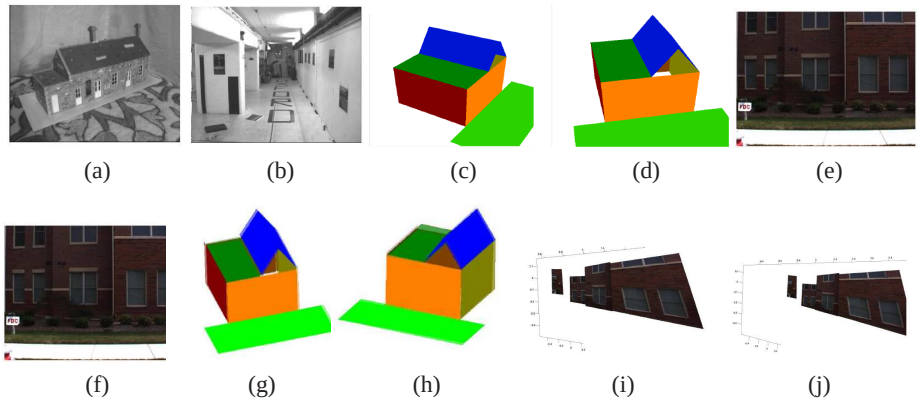
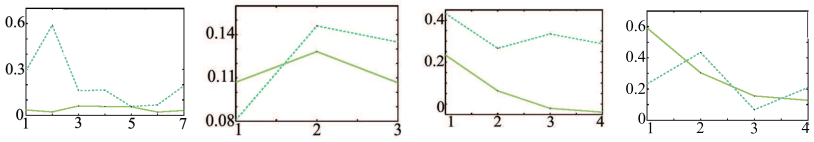


Fig. 5: Sample images of scenes reconstructed using our approach. (House(a), Corridor(b), synthetic(c-d), UNC((e-f))). (g-h) illustrates the accuracy of our reconstruction, the ground truth and reconstructed models are overlapping. (i-j) Texture mapped UNC reconstructions



(a) (b) (c) (d)

Fig. 6: Plots of the L_∞ error between plane and pose parameters with respect to the ground truth, for the House and Corridor sequence. $L2$ error shows similar plots. Y-axis of plots (a),(b),(c) and (d) is the angular error in radians, X-axis of (a) and (c) is the number of views, where as X-axis of (b) and (d) is the number of planes. In the plots (a),(b),(c) and (d), dotted curve represents the Faugeras initialization and other curve represents our approach

4.2 Real Data

In order to test on data from the real-world, we chose two Oxford data sets and the UNC data set. The House, and Corridor data sets (Figures (5a,5b)) are accompanied by correspondences and estimates of the camera matrices, while the UNC data set only comprises camera matrices.

Figures 6a-6b show the comparison between our estimation and that of the decomposition of Faugeras for the Oxford data sets. The L_2 and L_∞ errors between the estimated and ground truth quantities are plotted. In order to compare normals, we took the best estimate of normals from the available decompositions. As can be seen from the plots, estimates of translation from our algorithm are far better than the corresponding algorithm by Faugeras. We found that Zhang's algorithm produces estimates similar to that of Faugeras' algorithm in most cases. The same situation is repeated in the Corridor sequence (Figures 6c-6d), where translation is very accurately obtained. An explanation of why certain plane parameters are "perturbed" by a higher error is that some of the homographies are erroneous and the error in a particularly bad homography is distributed across planes. Finally, the UNC data set (Figures 5i,5j) show the visual accuracy of our reconstruction.

5 Discussion and Conclusion

We proposed a framework that reconstructs piecewise planar scenes in much the same way as Bundle Adjustment for point sets. The algorithm incorporates both multiple planes and views and does not constrain all the planes to be visible in any single view. This makes it a useful bridge between initialization approaches and non-linear minimization methods

The existing framework is not without its drawbacks. Currently, though the objective functions show robustness to noise, it does not work very well in the presence of outliers. Existing literature in convex optimization that handles outliers may be used for this purpose [3]. Similarly, uncertainty of correspondences can also be handled with techniques like [18]. Secondly, constraints *between* planes like orthogonality may help in stabilizing the overall reconstruction [8]. One other issue related to this algorithm

is its practical applicability. Recent results reported in [6, 19] are very relevant to our work and may be used to improve the run time of our algorithm, making it suitable for faster computation required by videos. We believe that our current contribution lays down a useful framework for practically viable optimization over planes, and wish to investigate further into its use for large scale optimization.

References

1. Kahl, F., Henrion, D.: Globally optimal estimates for geometric reconstruction problems. In: ICCV. (2005)
2. Kahl, F.: Multiple view geometry and the l-infinity norm. In: ICCV. (2005)
3. Sim, K., Hartley, R.: Removing outliers using the l-infinity norm. In: CVPR (1). (2006) 485–494
4. Hartley, R., Kahl, F.: Global optimization through searching rotation space and optimal estimation of the essential matrix. In: ICCV. (2007)
5. Zhang, Z., Hanson, A.R.: 3d reconstruction based on homography mapping. In: ARPA. (1996)
6. Fredrik Kahl, Sameer Agarwal, M.K.C.D.J.K.S.B.: Practical global optimization for multiview geometry. In: ECCV. (2006)
7. Mitra, K., Chellappa., R.: A scalable projective bundle adjustment algorithm using the l norm. In: ICVGIP. (2008)
8. Bartoli, A., Sturm, P.: Constrained structure and motion from multiple uncalibrated views of a piecewise planar scene. In: IJCV. (2003)
9. Chandraker, M., Kreigman, D.: Convex optimization for bilinear problems in computer vision. In: CVPR. (2008)
10. Hartley, R., Zisserman, A.: Multiple View Geometry in Computer Vision. Cambridge University Press (2004)
11. Faugeras, O., Lustman, F.: Motion and structure from motion in a piecewise planar environment. In: IJPRAI. (1988)
12. Antonio Criminisi, Ian D. Reid, A.Z.: A plane measuring device. In: Image Vision Comput. (1999) 625–634
13. Boyd, S., Vandenberghe., L.: Convex Optimization. Cambridge University Press, New York, NY, USA, 2004 (2004)
14. C. Olsson, F.K., Oskarsson., M.: Optimal estimation of perspective camera pose. In: ICPR. (2006)
15. Bartoli., A.: A random sampling strategy for piecewise planar scene segmentation. In: CVIU. Volume 105(1). (2007) 42–59
16. Sturm, J.F.: Using sedumi 1.02, a matlab toolbox for optimization over symmetric cones (1999)
17. Bill Triggs, Philip F. McLauchlan, R.I.H.A.W.F.: Bundle adjustment - a modern synthesis. In: Workshop on Vision Algorithms. (1999) 298–372
18. Ke, Q., Kanade., T.: Quasiconvex optimization for robust geometric reconstruction. In: ICCV. (2005)
19. S. Agarwal, N.S., Seitz, S.: Fast algorithms for l-inf problems in multiview geometry. In: CVPR. (2008)