

Triangulating a Plane

Carl Olsson¹, Anders Eriksson²

¹Centre for Mathematical Sciences, Lund University, Sweden

²School of Computer Science, University of Adelaide, Australia

Abstract. In this theoretical paper we consider the problem of accurately triangulating a scene plane. Rather than first triangulating a set of points and then fitting a plane to these points, we try to minimize the back-projection errors as functions of the plane parameters directly. As this is both geometrically and statistically meaningful our method performs better than the standard two step procedure. Furthermore, we show that the error residuals of this formulation are quasiconvex thereby making it very easy to solve using for example standard local optimization methods.

1 Introduction

The use of planes and their homographies has become increasingly important since the introduction of graph cuts for dense stereo reconstruction [3]. Since then a number of methods building on this work has been proposed (e.g. [22,13,5]), all working in a similar fashion. Typically a family of planes are used to represent the scene, and using α -expansion each pixel is classified as belonging to one of the planes in the family. The cost of assigning a pixel to a plane is computed using back-projection between the two cameras via plane-induced homography (see [8]). If a pixel, back-projected from one camera to the other, looks similar to the corresponding pixel in the second camera then the cost of assigning that pixel to the current plane is low. Furthermore, to obtain smooth classifications a standard regularization term is added [3].

In this paper we investigate the problem of determining the family of planes accurately. The typical way of determining a scene plane from stereo correspondences is by first triangulating the image points (using for example [7]) and then fitting a plane to the 3D-points (see [8]). There are two downsides to this approach. First, when we are triangulating the points we are not using the knowledge that all the points are lying on the same plane. Second, and perhaps more importantly, when fitting the plane we are measuring distances in 3D. Hence, we are optimizing a quantity that we cannot observe in our data and therefore may be inaccurate. The latter is a problem in particular if the baseline is small. On the other hand fitting problems with a small baseline is very important since descriptors such as SIFT usually perform much better in this case.

Instead we propose to estimate the plane by minimizing the back-projection errors of the plane-induced-homography. We will show that this method gives a good estimation of the plane. Direct estimation (without computing the 3D points) of a scene plane has been considered before. For example in [11,10] two iterative methods are presented. There are however no guarantees of convergence to the global optimum. In [1] both structure and motion is estimated. Here 3D points are constrained to fulfill

co-planarity constraints exactly, and a bundle adjustment process is employed. In this paper we assume that the cameras are known, but a similar approach (optimizing reprojection error with the points constrained to lie on the unknown plane) can of course be used. Such a method would however have to rely on triangulation of the 3D points for initialization and would only be locally convergent.

In contrast, we show in this paper that when minimizing back-projection errors of the plane-induced-homography, the problem can be globally optimized using convex optimization. In particular we show that the error residuals are affine functions composed with a projection. It was shown in [9] that this type of functions are examples of quasiconvex functions. Since quasiconvexity is preserved under the max-operation these problems exhibit no local minima when we minimize the max-norm of the errors. More recently minimizing the least squares error, using standard Levenberg-Marquardt type procedures, was addressed in [6,19]. It was shown that for the vast majority of problems from this class it is possible to use local methods and verify that the solution is in fact globally optimal. Furthermore, in [21,14,16,17] systematic ways for handling outliers in the data was given.

1.1 Quasiconvex Optimization

In this section we very briefly recall the definition and basic properties of quasiconvex functions. A much more detailed treatment can be found in [18].

When dealing with optimization problems convexity is a very useful property. For example, when minimizing a convex function we are guaranteed that local methods converge to globally optimal solutions. Unfortunately, convexity does hardly ever occur in multiple view geometry problems because projections are in general not convex. More commonly occurring is the slightly weaker notion of quasiconvexity.

Definition 1. *A function f is called quasiconvex on a convex set C if its sublevel sets*

$$S_\mu(f) = \{x \in C; f(x) \leq \mu\} \quad (1)$$

are convex for all $\mu \in \mathbb{R}$.

Figure 1 shows a function that is quasiconvex. It is not convex since it is possible to draw a line between two points on graph, such that the function is above the line. Quasiconvexity is not preserved under addition. It is, however, preserved under the max operation. That is minimizing the maximal error (see equation (4)) is a quasiconvex problem if all the error residuals are quasiconvex. The simplest, and perhaps the most common way of solving such problems is to employ a bisection method. Checking whether there is an x such that $f(x) \leq \mu$ for a fixed μ is a convex problem. Hence, by solving a sequence of convex problems one can find an optimal μ (see [9,12,18]). In [6,19] it was shown that in the vast majority of cases it is also possible to solve the least squares formulation (see equation (7)) using local methods.

2 3D-plane Triangulation

Next we consider the problem of determining a scene plane from two images. We assume that the image data is given as two sets of corresponding points in the two images

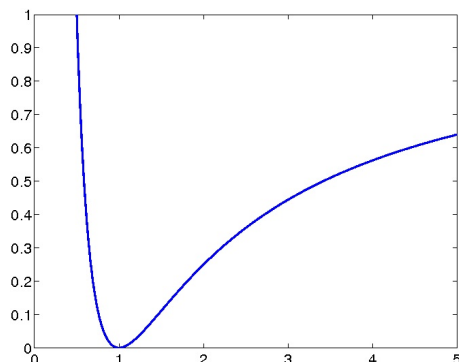


Fig. 1. A quasiconvex but not convex function.

that are known to be projections of points located on the 3D-plane. Let $P_1 = [R_1 \ t_1]$, $P_2 = [R_2 \ t_2]$ be the two known (calibrated) camera matrices, $\{x^i\}$, $\{y^i\}$ be the image coordinates and p the unknown scene plane. When dealing with points, we will use lower case letters to denote points in euclidean coordinates and capital letters to denote its homogeneous coordinates. Hence $\{X^i\}$ and $\{Y^i\}$ are the homogeneous coordinates of $\{x^i\}$ and $\{y^i\}$ respectively.

It is well known that (in a noise free system) there is a homography H_{21} from image 1 to image 2 such that

$$Y^i \sim H_{21}X^i. \quad (2)$$

(Here \sim denotes equality up to an unknown scale factor.) And similar for the inverse homography $H_{12} = H_{21}^{-1}$ from image 2 to image 1

$$X^i \sim H_{12}Y^i. \quad (3)$$

Now, suppose that the system is not noise free. We would like to find the plane (or homography) that gives the smallest back-projection errors in both images. Therefore we formulate the following minimization problem

$$\min_p \max_i R_i(p) \quad (4)$$

where

$$R_i(p) = \max(\|x^i - \Pi(H_{21}Y^i)\|^2, \|y^i - \Pi(H_{21}^{-1}X^i)\|^2) \quad (5)$$

and $\Pi : \{(x_1, x_2, x_3) \in \mathbb{R}^3; x_3 > 0\} \rightarrow \mathbb{R}^2$ is the projection mapping given by

$$\Pi(x) = \begin{pmatrix} x_1/x_3 \\ x_2/x_3 \end{pmatrix}. \quad (6)$$

The constraint $x_3 > 0$ reflects the fact that visible points in the image should be located in front of the cameras. The residual errors are similar to what is used in homography estimation with the \mathbb{L}_∞ norm [9,12], note however that the homography H_{21} depends

on the plane p . In homography estimation it is not possible to use back-projection errors in both images since the inverse of H_{21} cannot be parameterized linearly in terms of the elements in H_{21} , and therefore does not yield a quasiconvex problem. However in this setting this is not a problem, since, as we will show, both H_{21} and its inverse can be parameterized linearly by the plane parameters p .

In (4) and (5) we have used the maximum (squared) residual error as is traditionally done in the \mathbb{L}_∞ -framework. However, recent results (see [6,19]) show that solving the least squares formulation using local methods also works well for the same framework. In both cases we need to parameterize H_{21} and its inverse linearly. The least squares formulation of our problem is

$$\min_p \sum_i R_i(p) \quad (7)$$

where

$$R_i(p) = \|x^i - \Pi(H_{21}Y^i)\|^2 + \|y^i - \Pi(H_{21}^{-1}X^i)\|^2. \quad (8)$$

In this paper we will use both formulations. The latter is solved using local methods with initialization from the former.

2.1 Homography from a Plane

In this section we derive the expression for the homographies H_{12} and H_{21} . For simplicity let us first assume that the camera matrices are of the form $P_1 = [I \ 0]$ and $P_2 = [R \ t]$, and the plane parameters are $p = (a^T, b)^T$ where $a \in \mathbb{R}^3$ and $0 \neq b \in \mathbb{R}$.

Lemma 1. *If $P_1 = [I \ 0]$ and $P_2 = [R \ t]$ then the homography H_{21} can be written*

$$H_{21} = Rb - ta^T. \quad (9)$$

The special case $b = 1$ is proven in [8] and our proof is just a simple extension. It is however essential, since without it, it is not possible to use both H_{12} and its inverse in the error residual, which would prevent us from using symmetric error residuals.

Let x and y be the projections of the point z belonging to p . Then

$$X = P_1Z = [I \ 0]Z, \quad (10)$$

which implies that $Z = (X^T, d)^T$ for some $d \geq 0$. Since z belongs to p we have $0 = p^T Z = a^T X + bd$. And therefore

$$d = -\frac{a^T X}{b}. \quad (11)$$

Now, since Z is the homogeneous coordinates for z we may multiply Z with the scale factor b , assuming b is positive, without changing anything. The fact that this assumption is no restriction will be motivated later on. Therefore we obtain

$$Z = \begin{pmatrix} bX \\ -a^T X \end{pmatrix} \quad (12)$$

Projecting z into image 2 now gives

$$Y = P_2 Z = [R \ t] \begin{pmatrix} bX \\ -a^T X \end{pmatrix} = (Rb - ta^T)X, \quad (13)$$

which proves the statement.

Now for the general case we use a transformation to prove the following.

Corollary 1 *If $P_1 = [R_1 \ t_1]$, $P_2 = [R_2 \ t_2]$ then the homography H_{21} can be written*

$$H_{21} = (b - t_1^T R_1 a) R_2 R_1^T + (R_2 R_1^T t_1 - t_2)(R_1 a)^T. \quad (14)$$

Let

$$T = \begin{bmatrix} R_1^T & -R_1^T t_1 \\ 0 & 1 \end{bmatrix}. \quad (15)$$

Changing coordinates using T we get in the new coordinate system

$$\tilde{P}_1 = P_1 T = [I \ 0] \quad (16)$$

$$\tilde{P}_2 = P_2 T = [R_2 R_1^T \quad -R_2 R_1^T t_1 + t_2] \quad (17)$$

$$\tilde{p} = T^T p = \begin{bmatrix} R_1 a \\ b - t_1^T R_1 a \end{bmatrix}. \quad (18)$$

Substituting (16), (17) and (18) into (9) now yields (14). A similar expression for $H_{21}^{-1} = H_{12}$ is of course obtained by exchanging P_1 and P_2 . Hence H_{21} depends linearly on the plane parameters p and we will write $H_{21}(p)$ to indicate this dependence. Furthermore, we let H_{21}^i denote the i 'th row of H_{21} .

To prove that the error residual $\|x^i - \Pi(H_{21}(p)Y^i)\|^2$ is a quasiconvex function on the set $\mathcal{C} = \{p; H_{21}^3(p)Y^i > 0\}$ we need to prove that the sublevel set

$$\{p \in (C); \|x^i - \Pi(H_{21}(p)Y^i)\|^2 \leq \mu^2, \forall i\} \quad (19)$$

is convex for a fixed μ (see Definition 1, Section 1.1). However since $H_{21}^3(p)Y^i > 0$ it is easy to see that (19) is equivalent to

$$\|(ap, bp)\| \leq \mu cp, \quad (20)$$

where

$$ap = (x_1^i H_{21}^3(p) - H_{21}^1(p))Y^i, \quad (21)$$

$$bp = (x_2^i H_{21}^3(p) - H_{21}^2(p))Y^i, \quad (22)$$

$$cp = H_{21}^3(p)Y^i. \quad (23)$$

and x_j^i denotes the j 'th coordinate of x^i . Since (21)-(23) are linear functions (20) will be a second order cone constraint (see [2]) which is convex. Hence the error residuals are quasiconvex functions (H_{12} is handled in the same way), and the theory from [9,12,21,14,6,19,16,20] extends to this problem as well.

2.2 Chirality

In the proof of lemma 1 we did not motivate b being positive. It is easy to see that if $b < 0$ then there will be a sign change when multiplying X with b , and hence X might not have positive depth in the second camera. Similarly when using 1 we want the quantity $b - R_1^T t_1$ to be positive. This can always be ensured unless the cameras are located on opposite sides of the plane. Let $P_1 = [R_1 \ t_1]$, $P_2 = [R_2 \ t_2]$ and $p = (a^T, b)^T$ as previously. We then have

$$b - a^T R_1^T t_1 = p^T \begin{bmatrix} c_1 \\ 1 \end{bmatrix} \quad (24)$$

$$b - a^T R_2^T t_2 = p^T \begin{bmatrix} c_2 \\ 1 \end{bmatrix} \quad (25)$$

where c_1 and c_2 are the camera centers. Now it is easy to see that if one of these are negative then c_1 and c_2 cannot be on the same side of the plane.

Since the cameras need to see the same points the sign of b is normally not a problem in practice. However, note that in case we for some reason would like to solve a problem where this is not fulfilled one simply multiplies X with $-b$ instead and the same result holds. If we do not know anything about the relative locations of the cameras and the plane we have to test both possibilities. Similar to other multiple view geometry problems (see [9]), it is easily shown that for each choice there is a local minimum.

3 Experiments

In this section we perform some simple experiments to verify the theory and evaluate the quality of the proposed methods. In equation (14) the homography H_{21} is written as a linear expression in a and b . However as 3 parameters is enough for specifying a plane we will choose b to be 1 in all our implementations. Note that we still need to use (14). It is not possible to parameterize both H_{21} and H_{12} linearly using only (9) (with $b=1$).

3.1 Stability of the Proposed Formulation

As we have mentioned before the standard way of fitting a scene plane to image data, is to first compute 3D-points using triangulation and then fit a plane to these points. In our first experiment we compare this approach to the two proposed formulations on synthetically generated data. We use synthetic data since we would like to know the true parameters of the plane that generated the measurements. The setup is as follows: First we placed 30 points randomly on the plane $z = 0$, within the box $-1 \leq x \leq 1$, $-1 \leq y \leq 1$. Then we placed two cameras at a distance roughly 4 from the origin with camera centers fulfilling $z \leq -2$. We then added noise with standard deviation 0.0025 to the image coordinates. Figure 2 shows a typical example of the generated images. We also selected a maximal value for the baseline, that is, the camera centers was placed closer to each other than this maximal value. Figure 3 shows the results for a number of different values of the maximal baseline. To the left is the back-projection error when

measured with the max-norm (5). For each data point in the graph we generated 500 instances of the problem and computed the median of the back-projection error. We chose the median and not the mean since in rare instances the 3D point triangulation method produces a plane that is almost perpendicular to the true plane, resulting in extremely large back-projection errors. Even though we are averaging over a large number of experiments this gives a noisy graph which is difficult to interpret, therefore we use the median instead (for comparison we plotted one of the graphs obtained when using the mean in Figure 2). For the two proposed formulations the median and the mean are very similar.

To the left is the result when the back-projection error is measured using the max-norm (4). As expected the proposed \mathbb{L}_∞ formulation performs best here (regardless of the baseline).

In the middle is the result when the error is measured with the \mathbb{L}_2 -norm (8). Here the proposed \mathbb{L}_2 formulation is the winner. Somewhat surprisingly the proposed \mathbb{L}_∞ formulation performs better than the 3D-point triangulation approach when the baseline is small. The reason is that when the baseline is small distances in the direction of the depth is difficult to observe accurately in the cameras. Hence, the position of the 3D-points in this direction is uncertain. Therefore, using distances in 3D instead of back-projection errors results in a more unstable procedure.

To the right is perhaps the most interesting graph. Here we have plotted the median of the distance from the computed solutions to the true (noiseless solution). If $(a_{true}, 1)$ is the parameters for the true solution the distance is measured as $\|a_{true} - a_{est}\|_2$ where $(a_{est}, 1)$ is the estimated solution with one of the methods. When the baseline is sufficiently large the 3D point triangulation approach is almost as good as the proposed \mathbb{L}_2 formulation however when the baseline becomes smaller it is less stable than the other two methods.

To test the stability with respect to noise we also plotted the same figures when varying the noise level instead of the maximal baseline. The results can be seen in Figure 4. Here the maximal baseline was set to 0.5 and the noise level was varied between 0 and 0.005. The proposed formulations appear to exhibit roughly linear growth in the errors whereas the triangulation approach seem to grow faster.

3.2 Outlier Removal and Estimation

Next we evaluate our method on real data. In real settings the data is often corrupted by outliers. A popular method for removing outliers is RANSAC [4], however, here we will use the approach pioneered by Sim and Hartley [21], and later refined in [17]. This is an iterative method that is guaranteed to remove one outlier in each iteration. The algorithm works by solving the problem

$$\min s \tag{26}$$

$$\text{s.t. } \|(a_i p + a_{i0}, b_i p + b_{i0})\| \leq \mu(c_i p + c_{i0}) + s, \quad \forall i, \tag{27}$$

where $a_i, b_i, c_i, a_{i0}, b_{i0}$ and c_{i0} are constructed from the i 'th measurement (counting backward and forward homographies separately). It is shown in [17] that by removing

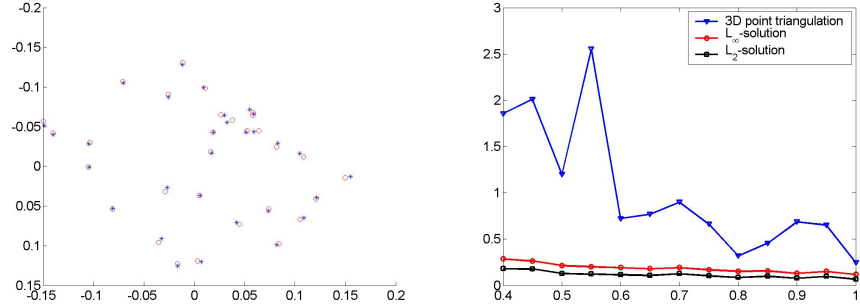


Fig. 2. Left: An example image used in the synthetic experiments. x - the exact image point, o - image point perturbed by noise (noise with std. dev. 0.0025 as in Figure 3). Right: Same result as in Figure 3 (right panel) using the mean instead of the median.

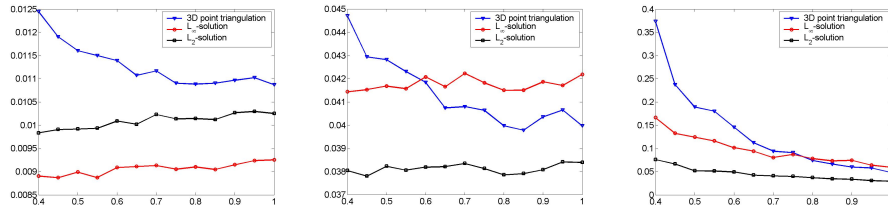


Fig. 3. Results of the synthetic experiments for the three methods. Left: The max-norm back-projection error (5) versus the maximal baseline. Middle: The \mathbb{L}_2 -norm back-projection error (8) versus the maximal baseline. Right: The distance to the true (noise less) solution.

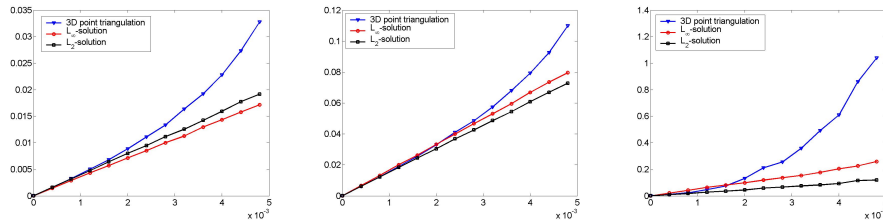


Fig. 4. Results of the synthetic experiments for the three methods. Left: The max-norm back-projection error (5) versus the noise level. Middle: The \mathbb{L}_2 -norm back-projection error (8) versus the noise level. Right: The distance to the true (noise less) solution.

the residuals for which the dual variables y are nonzero we are guaranteed to remove one outlier. This procedure is then iterated until the solution is good enough ($s \leq 0$).

We ran the algorithm on the stereo pair seen in Figure 5. Since the posters located on the wall has similar texture we obtained a lot of mismatches where points on one of the poster matches to points on the other. Using SIFT [15] we determined 678 point correspondences. In order to find a solution with all errors less than 10 pixels 112 iterations was needed. In total 241 image points were discarded. Manual inspection reveals that almost all the discarded cases is either a mismatch or a point not belonging to the dominant plane.

4 Conclusions

In this theoretical paper we have proposed a procedure for triangulating a scene plane. Our method is based on the framework of quasiconvex optimization making it easy to solve with guaranteed optimality. Furthermore, since we have shown that our problem belongs to this class, all the previously developed theory naturally applies to this problem as well. Since the formulation is based on back-projection which is a geometrically meaningful quantity it is also stable with respect to noise and geometry.

Acknowledgments

This work has been funded by the European Research Council (GlobalVision grant no. 209480), the Swedish Research Council (grant no. 2007-6476), the Swedish Foundation for Strategic Research (SSF) through the program Future Research Leaders and the Australian Research Council's Discovery Projects funding scheme (project DP0988439).

References

1. A. Bartoli and P. Sturm. Constrained structure and motion from multiple uncalibrated views of a piecewise planar scene. *International Journal of Computer Vision*, 52(1), 2003. 1
2. S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004. 5
3. Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 23(11):1222–1239, 2001. 1
4. M. A. Fischler and R. C. Bolles. Random sample consensus: a paradigm for model fitting with application to image analysis and automated cartography. *Commun. Assoc. Comp. Mach.*, 24:381–395, 1981. 7
5. Y. Furukawa, B. Curless, S. M. Seitz, and R. Szeliski. Reconstructing building interiors from images. In *Proc. Int. Conf. on Computer Vision*, Kyoto, Japan, 2009. 1
6. R. Hartley and Y. Seo. Verifying global minima for L_2 minimization problems. In *Conf. Computer Vision and Pattern Recognition*, Anchorage, USA, 2008. 2, 4, 5
7. R. Hartley and P. Sturm. Triangulation. *Computer Vision and Image Understanding*, 68(2):146–157, 1997. 1
8. R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2004. 1, 4

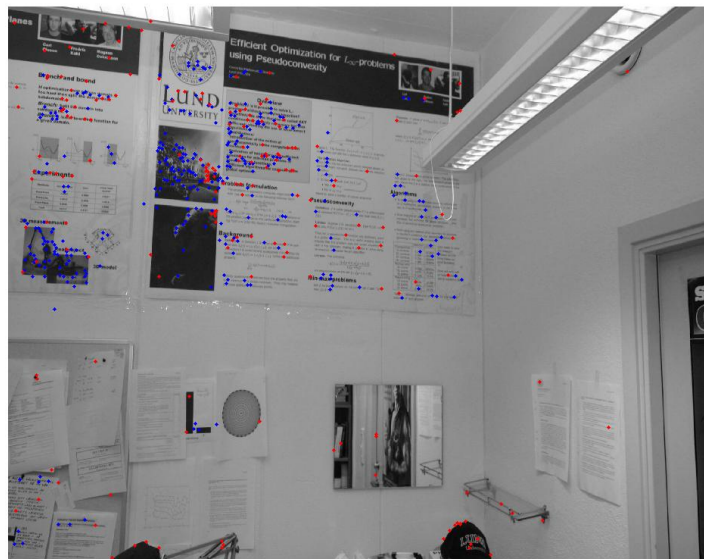


Fig. 5. Resulting estimation for the office stereo pair. Red points - outliers, blue points - inliers.

9. F. Kahl and R. Hartley. Multiple view geometry under the L_∞ -norm. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 30(9):1603–1617, 2008. 2, 3, 5, 6
10. K. Kanatani and H. Niitsuma. Optimal two-view planar scene triangulation. In *Proc. Asian Conf. on Computer Vision*, Queenstown, New Zealand, 2010. 1
11. Y. Kanazawa and K. Kenichi. Direct reconstruction of planar surfaces by stereo vision. *IEICE Transactions on Information and Systems*, E78-D(7), 1995. 1
12. Q. Ke and T. Kanade. Quasiconvex optimization for robust geometric reconstruction. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 29(10):1834–1847, 2007. 2, 3, 5
13. V. Kolmogorov and R. Zabih. Multi-camera scene reconstruction via graph cuts. In *European Conference on Computer Vision*, Copenhagen, Denmark, 2002. 1
14. H. Li. A practical algorithm for L_∞ triangulation with outliers. In *Conf. Computer Vision and Pattern Recognition*. Minneapolis, USA, 2007. 2, 5
15. D. Lowe. Distinctive image features from scale-invariant keypoints. *Int. Journal Computer Vision*, 2004. 9
16. C. Olsson, O. Enqvist, and F. Kahl. A polynomial-time bound for matching and registration with outliers. In *Conf. Computer Vision and Pattern Recognition*, Anchorage, USA, 2008. 2, 5
17. C. Olsson, A. Eriksson, and R. Hartley. Outlier removal using duality. In *Conf. Computer Vision and Pattern Recognition*, San Francisco, USA, 2010. 2, 7
18. C. Olsson and F. Kahl. Generalized convexity in multiple view geometry. *J. Math. Imaging Vis.*, 38(1):35–51, 2010. 2
19. C. Olsson, F. Kahl, and R. Hartley. Projective least-squares: Global solutions with local optimization. In *Conf. Computer Vision and Pattern Recognition*, Miami, USA, 2009. 2, 4, 5
20. Y. Seo and R. Hartley. A fast method to minimize L_∞ error norm for geometric vision problems. In *Int. Conf. Computer Vision*, Rio de Janeiro, Brazil, 2007. 5
21. K. Sim and R. Hartley. Removing outliers using the L_∞ -norm. In *Conf. Computer Vision and Pattern Recognition*, pages 485–492, New York City, USA, 2006. 2, 5, 7
22. O. Woodford, P. Torr, I. Reid, and A. Fitzgibbon. Global stereo reconstruction under second order smoothness priors. *Proc. IEEE Trans Pattern Analysis and Machine Intelligence*, 31:2115–2128, 2009. 1