

MONOCULAR RECTANGLE RECONSTRUCTION

Based on Direct Linear Transformation

Cornelius Wefelscheid, Tilman Wekel and Olaf Hellwich
Computer Vision and Remote Sensing, Berlin University of Technology
Sekr. FR3-1, Franklinstr. 28/29, D-10587, Berlin, Germany

Keywords: Rectangle, Quadrangle, Reconstruction, Direct linear transformation, Single view, Perspective projection.

Abstract: 3D reconstruction is an important field in computer vision. Many approaches are based on multiple images of a given scene. Using only one single image is far more challenging. Monocular image reconstruction can still be achieved by using regular and symmetric structures, which often appear in human environment. In this work we derive two schemes to recover 3D rectangles based on their 2D projections. The first method improves a commonly known standard geometric derivation while the second one is a new algebraic solution based on direct linear transformation (DLT). In a second step, the obtained solutions of both methods serve as seeding points for an iterative linear least squares optimization technique. The robustness of the reconstruction to noise is shown. An insightful thought experiment investigates the ambiguity of the rectangle identification. The presented methods have various potential applications which cover a wide range of computer vision topics such as single image based reconstruction, image registration or camera path estimation.

1 INTRODUCTION

The research field of 3D reconstruction has been studied intensively in the last years. The majority of current reconstruction techniques rely on at least two images from different perspectives to compute the depth of a scene (Hartley and Zisserman, 2003). This topic is referred as structure from motion (Faugeras and Lustman, 1988) which can be solved by the five point algorithm (Nister, 2004) or simultaneous localization and mapping (SLAM) approaches (Davison et al., 2007).

Although recovering 3D information based on one image is mathematically impossible, it has been shown that humans are able to perceive the 3D shape of an object based on monocular images. The example in Fig. 1 shows that the drawing of a Necker cube is perceived as a 3D object rather than an arrangement of lines in 2D space. Prior knowledge about the given scene allows us to interpret this figure correctly as a projection of a known 3D geometry and we do not rely on multiple perspectives. The rectangular structure as well as the symmetry of a cube is used as apriori-information. We implicitly assume that a cube is more likely to see than an arbitrary arrangement of lines. Pizlo (Pizlo, 2008) shows that 3D information obtained from single images in combination with

prior knowledge is more reliable and robust than 3D information from stereo. Psychology has investigated the shape constancy and the shape ambiguity problem (Todd, 2004). The shape constancy problem raises the question whether two different 2D views could be yielded by the same 3D object. The shape ambiguity problem deals with the question whether the same 2D view is induced by either the same or two different 3D objects. For an engineering-oriented investigation both are substantial: (1) The shape constancy problem presumes that a 3D shape can be inferred from monocular images. (2) The shape ambiguity problem forbids to trust a solution that is based on a single image. Among many other geometrical models, rectangular structures are of particular interest. Many objects in our environment are characterized by rectangles such as doors, windows or buildings. The mathematical description as well as the detection is relatively easy which is very important in practice. This work presents two approaches for the reconstruction of a 3D rectangle from its 2D perspective projection on a single image plane. We discuss the problem of shape constancy and shape ambiguity from a perceptual viewpoint and provocatively hypothesize that all perspective distorted quadrangles look like rectangles, which is supported by an experimental setup at the end of this paper. This paper is organized as fol-

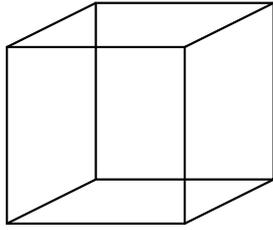


Figure 1: Image of a Necker cube.

lows. In the next section we present related work and emphasize the differences compared to our approach. Two mathematical methods are derived in Section 3. Here, we present an improved and simplified version of the approach presented in (Haralick, 1989) before we introduce a new method based on DLT. The identification problem as well as the accuracy analysis are discussed in Section 4. The proposed methods are evaluated on real world data before the paper concludes with an outlook to future work.

2 PREVIOUS WORK

Using rectangular structures to compute various information such as calibration and orientation of the camera is not new. Different approaches for single image based reconstruction have been presented in the past. All of them rely on several constraints such as parallelism and orthogonality in order to retrieve the missing information (Wilczkowiak et al., 2001). Vanishing points are used to compute the internal and external parameters of a camera (Sturm and Maybank, 1999). The computation tends to become unstable since these points are often placed near infinity. The work presented in (Haralick, 1989) is partly similar to our approach. It presents different derivations handling degenerated scene configurations such as coplanarity. This is not mandatory as it can be seen in Section 3. In contrast to previous efforts, we introduce a novel method which solves the stated problem by using a standard DLT method. In (Delage et al., 2007), Markov random fields (MRF) are used for detecting different planes and edges to form a 3D reconstruction from single image depth cues. In contrast to our work they assume orthogonal planes instead of dealing with the rectangle structure itself. (Micusk et al., 2008) describes an efficient method for detecting and matching rectilinear structures. They use MRF to label detected line segments. This approach enables the detection of rectangles even if the four line segments are not detected accurately. In (Lee et al., 2009), the scene is reconstructed by building hypotheses of intersecting line segments.

3 DERIVATION

We present two methods for reconstructing a rectangle in 3D space. The first method is based on geometric relations while the second one is a new algebraic solution. We assume a calibrated camera in both cases. In this context we are only interested in quadrangles with a convex shape since the projection of a rectangle is never concave. Our primary goal is to compute the orientation and the aspect ratio of a rectangle in 3D space from a perspectively distorted 2D image of a rectangle. This is equivalent to the computation of the extrinsic parameters of the camera, e.g. in the local coordinate system defined by the sides of the rectangle. The secondary goal is to verify that the observed quadrangle is yielded by a rectangle in 3D space and not by any other planar quadrangle. We have to exclude as many non-rectangular quadrangles as possible from further processing early and efficiently. The theoretical aspects of this problem are discussed in Section 4.

3.1 Geometric Method

Fig. 2 shows the arrangement of a 3D rectangle projected onto an image plane and Fig. 3 contains the 2D image representation. For the sake of clarity, we consider a camera that is placed in the origin and looks in Z-direction. $P_1 \dots P_4$ are the corner points of the rectangle and $p'_1 \dots p'_4$ are the corresponding projections in the image plane of the camera. They can be expressed in homogeneous coordinates $\mathbf{p}'_1 \dots \mathbf{p}'_4$. Neglecting the intrinsic camera parameters, the points \mathbf{p}'_i are transformed to \mathbf{P}'_i , which are the corner points of the rectangle's projection in the world coordinate system. They are connected by the edges of the rectangle \mathbf{l}_{12} , \mathbf{l}_{14} , \mathbf{l}_{23} and \mathbf{l}_{34} . Opposing edges intersect at the vanishing points \mathbf{v}_1 and \mathbf{v}_2 . The center point \mathbf{M} is defined as the intersection of the rectangle's diagonals. \mathbf{M}' is the projection of the center point \mathbf{M} . The line defined by \mathbf{v}_1 and \mathbf{M}' intersects the rectangle's edges at its centers \mathbf{P}'_{14} and \mathbf{P}'_{23} , respectively. \mathbf{P}'_{12} and \mathbf{P}'_{34} are defined by the second vanishing point \mathbf{v}_2 . The points \mathbf{P}'_i and the camera rotation angles ω , ϕ and κ are deduced from the corner points \mathbf{P}'_i . The distance d from projection center to rectangle center can be chosen arbitrarily. In the following we derive the computation of a rectangle based on a quadrangle. According to Figs. 2 and 3 we can derive the following simple equations:

$$\mathbf{l}_{ij} = \mathbf{P}'_i \times \mathbf{P}'_j \quad (1)$$

$$\mathbf{M}' = \mathbf{l}_{13} \times \mathbf{l}_{24} \quad (2)$$

$$\mathbf{v}_1 = \mathbf{l}_{12} \times \mathbf{l}_{34} \quad (3)$$

$$\mathbf{v}_2 = \mathbf{l}_{14} \times \mathbf{l}_{23}$$

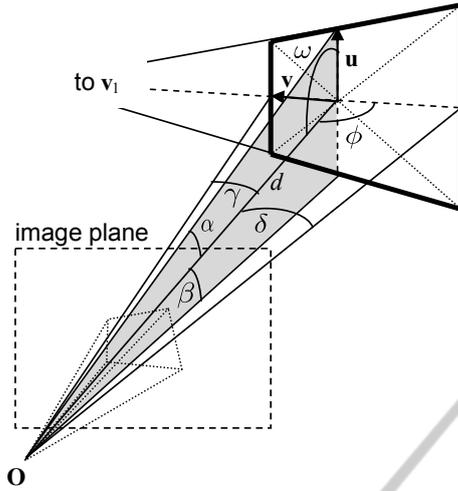


Figure 2: 3D arrangement of camera and 3D rectangle.

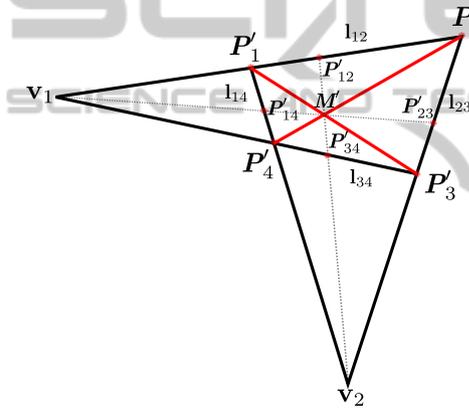


Figure 3: 2D image of a rectangle.

$$\begin{aligned}
 P'_{12} &= (\mathbf{v}_2 \times \mathbf{M}') \times \mathbf{l}_{12} \\
 P'_{34} &= (\mathbf{v}_2 \times \mathbf{M}') \times \mathbf{l}_{34} \\
 P'_{23} &= (\mathbf{v}_1 \times \mathbf{M}') \times \mathbf{l}_{23} \\
 P'_{14} &= (\mathbf{v}_1 \times \mathbf{M}') \times \mathbf{l}_{14}
 \end{aligned} \quad (4)$$

$$\begin{aligned}
 \alpha &= \arccos \frac{P'_{12} \cdot \mathbf{M}'}{|P'_{12}| |\mathbf{M}'|}, \beta = \arccos \frac{P'_{34} \cdot \mathbf{M}'}{|P'_{34}| |\mathbf{M}'|} \\
 \gamma &= \arccos \frac{P'_{14} \cdot \mathbf{M}'}{|P'_{14}| |\mathbf{M}'|}, \delta = \arccos \frac{P'_{23} \cdot \mathbf{M}'}{|P'_{23}| |\mathbf{M}'|}.
 \end{aligned} \quad (5)$$

Given all angles in the presented setup, the description of the points in space is straight forward:

$$|P_{12}| = \frac{2d \sin(\beta)}{\sin(\alpha + \beta)}, |P_{34}| = \frac{2d \sin(\alpha)}{\sin(\alpha + \beta)} \quad (6)$$

$$|P_{14}| = \frac{2d \sin(\delta)}{\sin(\delta + \gamma)}, |P_{23}| = \frac{2d \sin(\gamma)}{\sin(\delta + \gamma)}. \quad (7)$$

P_{ij} are the center points of the edges defined by P_i and P_j :

$$P_{ij} = |P_{ij}| \frac{P'_{ij}}{|P'_{ij}|}. \quad (8)$$

The rectangle is parametrized by the center point M and the spanning vectors \mathbf{u} and \mathbf{v} ,

$$\mathbf{M} = d \frac{\mathbf{M}'}{|\mathbf{M}'|} \quad (9)$$

$$\mathbf{u} = P_{12} - \mathbf{M} \quad (10)$$

$$\mathbf{v} = P_{14} - \mathbf{M}.$$

The corner points of the rectangle can now be calculated as

$$P_1 \dots P_4 = \mathbf{M} \pm \mathbf{v} \pm \mathbf{u}. \quad (11)$$

The equations derived above yield the following formula for ω :

$$\omega = \arctan \frac{2}{\cot(\beta) - \cot(\alpha)}. \quad (12)$$

The formula for ϕ and κ can be derived analogously.

3.2 DLT Method

The second method utilizes the well known DLT in order to compute the parameters of the 3D rectangle. In the following we define a linear system of 15 equations:

$$\begin{aligned}
 P'_1 \times (\mathbf{M} + \mathbf{v} + \mathbf{u}) &= \mathbf{0} \\
 P'_2 \times (\mathbf{M} - \mathbf{v} + \mathbf{u}) &= \mathbf{0} \\
 P'_3 \times (\mathbf{M} - \mathbf{v} - \mathbf{u}) &= \mathbf{0} \\
 P'_4 \times (\mathbf{M} + \mathbf{v} - \mathbf{u}) &= \mathbf{0} \\
 \mathbf{M}' \times \mathbf{M} &= \mathbf{0}.
 \end{aligned} \quad (13)$$

Each term in Eq. 13 yields only two linearly independent equations. We compose a design matrix $(A|B)$ which is solved for $(M_x, u_x, v_x, M_y, u_y, v_y, M_z, u_z, v_z)$ using Singular Value Decomposition (SVD):

$$(A|B) \cdot (M_x, u_x, v_x, M_y, u_y, v_y, M_z, u_z, v_z)^T = \mathbf{0}. \quad (14)$$

The design matrix $(A|B)$ is defined as followed:

$$A = \begin{pmatrix} \mathbf{1}^T & \mathbf{0}^T \\ \mathbf{0}^T & -\mathbf{1}^T \\ 1 & -1 & 1 & \mathbf{0}^T \\ \mathbf{0}^T & -1 & 1 & -1 \\ 1 & -1 & -1 & \mathbf{0}^T \\ \mathbf{0}^T & -1 & 1 & 1 \\ 1 & 1 & -1 & \mathbf{0}^T \\ \mathbf{0}^T & -1 & -1 & 1 \\ 1 & 0 & 0 & \mathbf{0}^T \\ \mathbf{0}^T & -1 & 0 & 0 \end{pmatrix} \quad (15)$$

$$B = \begin{pmatrix} -P'_{1,x} & -P'_{1,x} & -P'_{1,x} \\ P'_{1,y} & P'_{1,y} & P'_{1,y} \\ -P'_{2,x} & P'_{2,x} & -P'_{2,x} \\ P'_{2,y} & -P'_{2,y} & P'_{2,y} \\ -P'_{3,x} & P'_{3,x} & P'_{3,x} \\ P'_{3,y} & -P'_{3,y} & -P'_{3,y} \\ -P'_{4,x} & -P'_{4,x} & P'_{4,x} \\ P'_{4,y} & P'_{4,y} & -P'_{4,y} \\ -M'_x & 0 & 0 \\ M'_y & 0 & 0 \end{pmatrix}, \quad (16)$$

where the subindex x , y or z indicates the coordinate of the corresponding point or vector. Using the presented equations, we always compute parallelograms. We can formulate an additional condition to check if the detected shape is a rectangle:

- The angle between the spanning vectors \mathbf{u} and \mathbf{v} must be perpendicular, so it must hold $\mathbf{u}^T \mathbf{v} = 0$.

The condition is satisfied by rectangles only.

3.3 Optimization

If noise is taken into account we try to find a rectangle which sufficiently approximates the observed quadrangle. We limit the parametrization to eight degrees of freedom in order to assure the orthogonality of the spanning vectors. It turns out to be reasonable to exclude v_x or v_y from the parameter set. Both values v_x or v_y are then computed out of u , $v_{y/x}$, and v_z such that u and v are perpendicular.

$$v_x = -\frac{u_y v_y + u_z v_z}{u_x} \quad (17)$$

Omitting v_z would lead to a division by zero if the rectangle is coplanar to the image plane. This cannot occur for v_x or v_y because the image plane would be perpendicular to the rectangle.

In this case the projection of the rectangle is a line rather than a quadrangle. The computed rectangle is still not optimal because the spanning vectors are only close to be perpendicular. If we are only interested in the orientation and ratio of the rectangle, a parallelogram will already be a good approximation. We minimize the reprojection error defined in Eq. 18 in order to get the optimal solution. P_r and P_q is a set of corner points. The index r represents the back projected rectangle whereas q is the measured quadrangle.

$$\min \sum_{i=1}^4 \text{dist}(P'_{ri}, P'_{qi})^2 \quad (18)$$

This minimization is done using the Levenberg-Marquardt-algorithm. The algorithm shows a good convergence behavior if the obtained parallelogram

parameters are used as seeding points. The optimization as well as the proposed methods are analyzed on synthetic and real world data in the following sections.

4 EXPERIMENTS

In this section we want to discuss the ambiguity of the rectangle identification. Each quadrangle in an image can be perfectly restored to a parallelogram in 3D space. According to the derivation in section 3.1, this reconstruction is unique up to a scale factor. In practice we have to deal with the presence of noise. In this section, we investigate this problem in two experiments: 1. Is it possible to distinguish between a parallelogram and a rectangle? 2. How accurate is the restoration of a noisy rectangle?

We define LE to be the length of the longest edge of the quadrangle and normalize all errors to make them invariant to the image size.

4.1 Identification

If we detect a quadrangle in an image we do not know if this is a projection of a rectangle or just a parallelogram. To examine this problem, we create arbitrary parallelograms by randomly choosing uniformly distributed spanning vectors. Since we do not want to investigate the scaling, \mathbf{u} and \mathbf{v} have a constant length. The ratio between \mathbf{u} and \mathbf{v} is a sample drawn from a uniform distribution between 0.01 and 1.0. We create 10.000 parallelograms in total. A rectangle is computed for each parallelogram minimizing the reprojection error as presented in Section 3.2 and 3.3. The resulting projection error could either be caused by noise or our assumption of measuring a rectangular structure in 3D space is violated. If no prior knowledge of the scene is given it is impossible to identify the specific source of error. This question definitely depends on the accuracy of the quadrangle detector. The results presented in Fig. 4 show that after projective distortion it is not possible to distinguish between a 3D parallelogram and a 3D rectangle. As the cumulative histogram presents, 90 percent of all randomly created parallelograms have a distance of less than 0.05 LE to the closest 3D rectangle. For the sake of clarity we give a numeric example: If the longest edge of a quadrangle is 50 pixels long, and a quadrangle detector has an accuracy of $\sigma = 2.5$ pixels, 90 percent of all parallelograms are misperceived as a rectangle. Based on the geometric appearance only, it is not possible to distinguish reliably between a parallelogram and a rectangle.

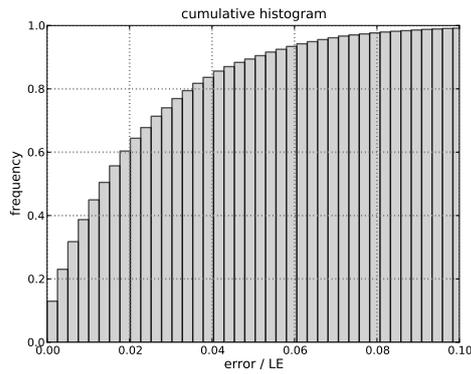


Figure 4: Cumulative histogram of the average error between original and calculated corner points of the quadrangle.

4.2 Accuracy

In contrast to the identification experiments, we evaluate the accuracy of the reconstructed rectangle in the following. The most important quantities are the ratio and the normal vector of the rectangle. Similar to the experiment in 4.1, we randomly create 3D rectangles. Now, the spanning vector \mathbf{v} is perpendicular to \mathbf{u} . We add normally distributed noise to the corner points of the 2D quadrangle given by $\sigma_{px} = \sigma \cdot LE$. We plot the average error of 1000 rectangles for different sigmas between 0.001 and 0.03. The angle and ratio errors with respect to the original rectangle are shown in Fig. 5 and Fig. 6 respectively. As it can be seen in the figures, the error increases almost linearly with the pixel noise. Nevertheless the methods deliver good results even at high noise ratios.

5 RESULTS

In this section we evaluate the described methods on real world data. We choose an object which has been precisely measured by a laser scanner in order to provide ground truth. The cube shown in Fig. 7 contains 27 markers with different orientations. Nine markers are placed on each side. All three planes are perpendicular to each other. The colored markers in the center of each plane are ignored and we get 24 rectangles in total. Fig. 8 and Fig. 9 show the reconstruction of the mentioned object and the corresponding ground truth for a better illustration. In this example the distance is of all rectangles is set to ground truth to find a common scale. Using a calibrated camera (12 mega-pixel), we have taken five images of the object from different perspectives. We want to analyze the aspect ratio and the orientation of each marker.

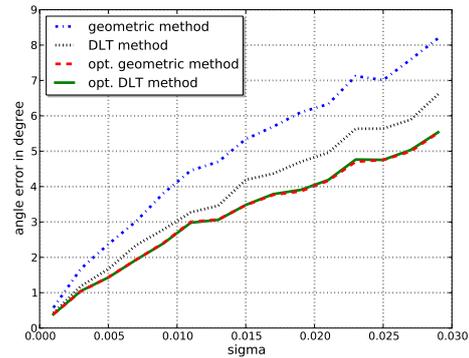


Figure 5: Angle error for different sigmas.

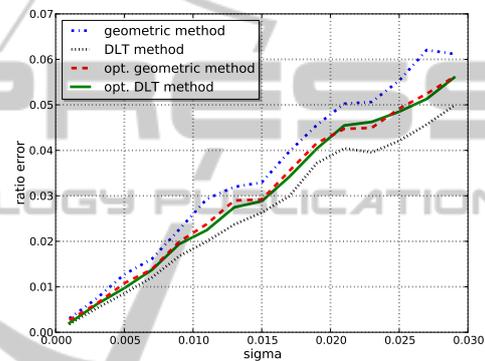


Figure 6: Ratio error for different sigmas.

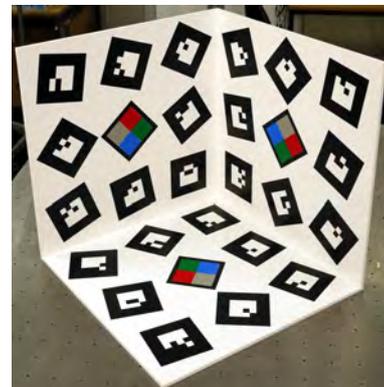


Figure 7: Calibration object used for evaluation.

The corner points are precisely measured and we can directly compare the ratio of each marker to ground truth. For evaluating the angle error, we set the coordinate system to the upper left corner of the first marker. The computed angle should be either zero or 90 degree. The difference to the closest value is defined as reconstruction error. The mean errors as well as the standard deviations are presented in Table 1 and Table 2. The resulting error is relatively

Table 1: Mean and standard deviation of the ratio error.

Ratio	Mean	Std.
Geometric	0.0123	0.0135
DLT	0.0123	0.0135
Opt. geometric	0.0119	0.0128
Opt. DLT	0.0114	0.0123

Table 2: Mean and standard deviation of the angle error in degree.

Angle	Mean	Std.
Geometric	1.503	1.550
DLT	1.502	1.550
Opt. geometric	1.122	1.311
Opt. DLT	1.116	1.308

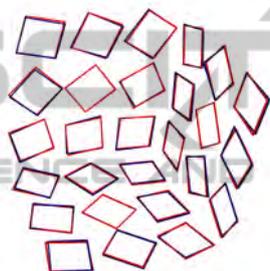


Figure 8: Reconstruction of the image in Fig. 7 marked in blue and the ground truth measured with a laser scanner marked in red.

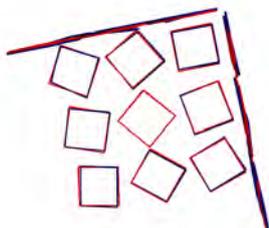


Figure 9: The top view of the reconstruction shows the precise reconstruction. Only small errors can be seen in the ground plane.

low. The evaluation on real world data show the same characteristics as the simulation. In both cases the optimization improves the angle accuracy but shows less effect on the ratio. Regarding the experiment in Section 4.2, we can assume that the quadrangle detector has a high accuracy.

6 CONCLUSIONS

We have presented two methods to compute a 3D rectangle from a 2D quadrangle. The given results show that the methods are stable even when applied

to noisy data. They can be utilized for further applications which could use rectangles as meaningful shapes. These higher order shapes can improve the accuracy in many computer vision tasks e.g. camera calibration, orientation and path estimation. Monocular SLAM algorithms can be significantly improved by using rectangles as landmarks. The initialization of new rectangles from a single view and the estimation of their depth improves the stability of such methods. In future work we will try to derive a method that enables an analytic error propagation from the pixel coordinates to 3D space. This maximizes the benefit of rectangle based models in an extended Kalman filter.

REFERENCES

- Davison, A. J., Reid, I. D., Molton, N. D., and Stasse, O. (2007). MonoSLAM: real-time single camera SLAM. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(6):1052–1067.
- Delage, E., Lee, H., and Ng, A. (2007). Automatic single-image 3d reconstructions of indoor manhattan world scenes. *Robotics Research*, pages 305–321.
- Faugeras, O. D. and Lustman, F. (1988). Motion and Structure From Motion in a Piecewise Planar Environment. *Intern. J. of Pattern Recogn. and Artific. Intellige.*, 2(3):485–508.
- Haralick, R. M. (1989). Determining camera parameters from the perspective projection of a rectangle. *Pattern Recognition*, 22(3):225–230.
- Hartley, R. and Zisserman, A. (2003). *Multiple view geometry in computer vision*. Cambridge University Press New York, NY, USA.
- Lee, D. C., Hebert, M., and Kanade, T. (2009). Geometric reasoning for single image structure recovery. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*.
- Micusik, B., Wildenauer, H., and Kosecka, J. (2008). Detection and matching of rectilinear structures. In *CVPR*. IEEE Computer Society.
- Nister, D. (2004). An efficient solution to the five-point relative pose problem. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(6).
- Pizlo, Z. (2008). *3D shape: its unique place in visual perception*. The MIT Press.
- Sturm, P. and Maybank, S. J. (1999). A method for interactive 3d reconstruction of piecewise planar objects from single images. In *British Machine Vision Conference*, pages 265–274.
- Todd, J. T. (2004). The visual perception of 3d shape. *Trends in Cognitive Sciences*, 8(3):115–121.
- Wilczkowiak, M., Boyer, E., and Sturm, P. (2001). Camera calibration and 3D reconstruction from single images using parallelepipeds. In *International Conference on Computer Vision, Vancouver*, pages 142–148.